UNIVERSIDAD NACIONAL DE CAÑETE

FACULTAD DE INGENIERÍA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS

“IMPLANTACIÓN DE LA HERRAMIENTA CLOUD QUICKSIGHT PARA LA VISUALIZACIÓN DE DATOS OBTENIDOS DEL PROYECTO SMART CITIES: CAÑETE PLAN PILOTO CON MEDIDORES ELÉCTRICOS INTELIGENTES”

TESIS

Para Optar el Título Profesional de Ingeniero de Sistemas

AUTOR:

BENI ALEXIS LUYO HUAMÁN
orcid.org/0000-0003-1918-7680

ASESOR:

ALEX ABELARDO PACHECO PUMALEQUE
orcid.org/0000-0001-9721-0730

CAÑETE – PERÚ

2021

UNIVERSIDAD NACIONAL DE CAÑETE

FACULTAD DE INGENIERÍA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS

TESIS

"IMPLANTACIÓN DE LA HERRAMIENTA CLOUD QUICKSIGHT PARA LA VISUALIZACIÓN DE DATOS OBTENIDOS DEL PROYECTO SMART CITIES: CAÑETE PLAN PILOTO CON MEDIDORES ELÉCTRICOS INTELIGENTES"

PRESENTADO POR:

BENI ALEXIS LUYO HUAMÁN
orcid.org/0000-0003-1918-7680

PARA OPTAR EL TITULO PROFESIONAL DE:

INGENIERO DE SISTEMAS

Aprobado Por:

Mg. Amanda Durán Carhuamaca
Presidente

FIRMA DIGITAL
Firmado digitalmente por: DURAN CARHUAMACA Amanda FAU 20491363402 soft
Motivo: En señal de conformidad
Fecha: 17/09/2021 12:45:53-0500

M.Sc. Edwin Roque Tito
Secretario

FIRMA DIGITAL
Firmado digitalmente por: ROQUE TITO Edwin FAU 20491363402 soft
Motivo: En señal de conformidad
Fecha: 20/09/2021 12:56:56-0500

Mg. Marisol Daga Chaca
Vocal

FIRMA DIGITAL
Firmado digitalmente por: DAGA CHACA Marisol FAU 20491363402 soft
Motivo: Soy el autor del documento
Fecha: 17/09/2021 10:33:11-0500

Mg. Alex Abelardo Pacheco Pumaleque
Asesor de Tesis

FIRMA DIGITAL
Firmado digitalmente por: PACHECO PUMALEQUE Alex Abelardo FAU 20491363402 soft
Motivo: Soy el autor del documento
Fecha: 13/09/2021 12:00:23-0500

Dedicatoria:

El siguiente trabajo está dedicado mis familiares, docentes y amigos que me han apoyado a lo largo de mi carrera.

Agradecimientos:

Agradecer al Dr. Igor Aguilar por sus enseñanzas y los ingenieros Cesar Cárdenas y Alex Pacheco por asesorarme en este trabajo.

Resumen

Actualmente los proyectos de Smart Cities están en pleno auge, lo cual no se limita al diseño e implementación de soluciones tecnológicas sino también al proceso de recolección, análisis y visualización de datos, esto genera un papel importante para las herramientas de visualización de datos. Es por ello que esta investigación busca implantar la herramienta Amazon Quicksight para la visualización de datos obtenidos de la simulación Smart Cities: Cañete plan piloto con medidores eléctricos inteligentes. Se aplicó dos fases para que los datos fueran analizados y procesados por medio de la herramienta Quicksight para visualizarlos y conseguir la toma de buenas decisiones en la organización. Los resultados de los indicadores fueron: un aumento en la calidad de datos del 27.39%, aumento de la lectura y visualización de datos del 63.70% y aumento en el porcentaje de datos útiles del 47.30%. Se concluye que la implantación de la herramienta Amazon Quicksight es efectiva para lograr la visualización de datos obtenidos del proyecto Smart Cities Simulation.

Palabras claves: Visualización de Datos, Amazon Quicksight, Smart Cities, Simulación, Universidad Nacional de Cañete.

Abstract

Currently Smart Cities projects are booming, which is not limited to the design and implementation of technological solutions but also to the process of data collection, analysis and visualization, this generates an important role for data visualization tools. That is why this research seeks to implement the Amazon Quicksight tool for the visualization of data obtained from the Smart Cities simulation: Cañete pilot plan with smart electricity meters. Two phases were applied so that the data were analyzed and processed through the Quicksight tool to visualize them and achieve good decision-making in the organization. The results of the indicators were: an increase in data quality of 27.39%, an increase in the reading and visualization of data of 63.70% and an increase in the percentage of useful data of 47.30%. It is concluded that the implementation of the Amazon Quicksight tool is effective to achieve the visualization of data obtained from the Smart Cities Simulation project.

Keywords: Data Visualization, Amazon Quicksight, Smart Cities, Simulation, National University of Cañete.

Índice General

Dedicatoria: 3
Agradecimientos: 4
Resumen 5
Abstract 6
I. Introducción 13
1.1 Formulación del problema 14
Problema general 14
Problemas específicos 14
1.2 Justificación 15
1.3 Formulación hipótesis 15
Hipótesis general 15
Hipótesis especificas 15
1.4 Objetivos de la investigación 16
Objetivo general 16
Objetivos específicos 16
1.5 Contribución 16
II. Marco Teórico 17
2.1 Antecedentes de la investigación 17
A nivel internacional 17
A nivel nacional 20
2.2 Bases Teóricas 25
Smart cities (ciudades inteligentes) 25
Implementación de arquitectura de una Smart Cities 26
Modelo de Smart City 27
Conexión con los retos sociales 28

Smart Grid (red inteligente) 31
Smart Meter (medidor inteligente) 32
AMR 33
AMI 33
Red LTE (4G) 34
Cloud Computing (computación en la nube) 35
¿Qué es cloud computing? 35
Una primera guía de proveedores de Cloud computing 35
Ventajas 37
Desventajas 37
Capas de cloud computing 37
Big Data 38
Técnicas Big Data 40
Minería de datos 40
Proceso KDD 41
Amazon Quicksight 42
Preparación de datos 42
SPICE 43
Análisis de datos 43
Elementos visuales 44
Flujo de trabajo típico de Amazon QuickSight 44
Visualización de datos 45
¿Qué es una visualización? 48
Tipos de visualización de datos 49
Procesamiento de datos 52
¿Qué es el procesamiento de datos? 52
Análisis de datos 52

Definición de términos 53
III. Materiales y Métodos 55
3.1 Materiales e insumos 55
3.2 Metodología 56
Diseño de investigación 56
Método de investigación: Hipotético – Deductivo 56
Tipo de investigación: Explicativa 56
Tipo de investigación: Experimental 56
Tipo de investigación: Aplicada 56
Diseño de investigación: Pre - Experimental 57
Variables y operacionalización 58
Variable independiente 58
Variable dependiente 58
Operacionalización de variables 59
Población 61
Muestra 61
Técnica e instrumentos de recolección de datos 61
Fichaje 61
Ficha de registro 61
Validez y confiabilidad 62
Validez 62
Validez de criterio 62
Validez de contenido 63
Validez de constructo 63
Confiabilidad 64
Método: Test-Retest 64
Técnica: Coeficiente de correlación de Pearson 64

3.3 Metodología de desarrollo del proyecto 68
Etapa 1: Análisis y procesamiento de datos 68
Etapa 2: Visualización y sus formas de compartir 88
IV. Resultados y Discusión 93
4.1 Resultados 93
Porcentaje de calidad de datos 93
Porcentaje de lectura y visualización de datos 95
Porcentaje de datos útiles 97
4.2 Discusión 98
V. Conclusiones 100
VI. Recomendaciones 101
VII. Referencias Bibliográficas 102
VIII. Anexos 107
Presupuesto de la investigación 108
Cronograma de la investigación 109
Modelo de ficha de registro 110
Resultados de instrumento ficha de registro 111
Pre-test 111
Post-test 112
Ejemplo de visualización con datos de simulación 114
Fotografías 114

Índice de figuras

Figura 1: Áreas verticales y horizontales del modelo de Smart City. 28

Figura 2: Retos de la sociedad y sus áreas tecnológicas del modelo de Smart City. 30

Figura 3: Esquema de una smart grid. 32

Figura 4: Crecimiento de Cloud computing a nivel mundial. 36

Figura 5: Capas de cloud computing. 38

Figura 6: Tecnologías Big Data. 39

Figura 7: Proceso KDD. 41

Figura 8: Amazon QuickSight Workflow. 45

Figura 9: Grafica de hipotecas. 48

Figura 10: Ejemplo de cuadro de mando empresarial. 50

Figura 11: Cambios en la dieta alimenticia en la última centuria. 51

Figura 12: Porcentaje de calidad de datos. Fuente: Elaboración propia. 95

Figura 13: Porcentaje lectura y visualización de datos. Fuente: Elaboración propia. 96

Figura 14: Porcentaje de datos útiles. Fuente: Elaboración propia. 97

Figura 15: Fichaje pre calidad de datos. Fuente: Elaboración propia. 111

Figura 16: Fichaje pre interpretación de datos. Fuente: Elaboración propia. ..111

Figura 17: Fichaje pre análisis de datos. Fuente: Elaboración propia. 112

Figura 18: Fichaje post calidad de datos. Fuente: Elaboración propia. 112

Figura 19: Fichaje post interpretación de datos. Fuente: Elaboración propia. 113

Figura 20: Fichaje post análisis de datos. Fuente: Elaboración propia. 113

Figura 21: Retención de muestras vector. Fuente: Elaboración propia. 114

Índice de tablas

Tabla 1. Tabla histórica de hipotecas. Fuente: INE (2011). 47

Tabla 2. Materiales y Equipos 55

Tabla 3. Insumos. 55

Tabla 4. Operacionalización de variables. 60

Tabla 5. Técnicas e instrumentos de recolección de datos. 62

Tabla 6. Validación del indicador - Porcentaje de datos de calidad obtenidos. 63

Tabla 7. Validación del indicador - Porcentaje de lectura y visualización de datos. 63

Tabla 8. Validación del indicador - Porcentaje de datos útiles. 64

Tabla 9. Análisis de confiabilidad. 65

Tabla 10. Confiabilidad del indicador Porcentaje de calidad de datos. 66

Tabla 11. Confiabilidad del indicador Porcentaje de lectura y visualización de datos. 66

Tabla 12. Confiabilidad del indicador Porcentaje de datos útiles. 67

Tabla 13. Margen de referencia del índice de calidad. 94

Tabla 14. Descriptivos según implementación de la herramienta cloud Quicksight. 98

Tabla 15. Matriz de consistencia. 107

Tabla 16. Presupuesto de investigación. 108

Tabla 17. Cronograma de investigación. 109

I. Introducción

Hoy en día más de la mitad de la población mundial vive en ciudades. Para 2050, casi siete de cada diez personas vivirán en zonas urbanas. Las ciudades representan más del 70% de las emisiones mundiales de carbono y entre el 60% y el 80% del consumo de energía. La rápida urbanización ha creado nuevos problemas, como la desigualdad social, la congestión del tráfico y la contaminación del agua y sus problemas de salud conexos. **(ITU, 2018)**

En el Perú, en junio de 2018 el Ministerio de Transporte y Comunicaciones elaboró un plan de recomendaciones para la implementación de políticas para el desarrollo de proyectos smart cities. Dicho plan está orientado a una mejor y ordenada gestión de servicios para los ciudadanos mediante el desarrollo de TIC's: soportes tecnológicos que facilitan el procesamiento, administración y compartición de la información. **(ContentLab, 2019)**

Uno de los subsistemas del modelo de smart city es Smart Grid, que son redes inteligentes que gestionan flujos de energía utilizando tecnologías de información y comunicación. Su desarrollo conjuntamente con el uso de Smart meters o medidores inteligentes, permiten recopilar datos automáticamente para procesarlos, transformarlos en información y llevarlos a la visualización, permitiendo a los usuarios una buena gestión de sus consumos de energía en el día a día, logrando ahorros económicos y preservando el ambiente.

Fran Vázquez de Smart Cities Perú, explica que un requisito básico de las ciudades inteligentes es: "levantar información, hacerle un seguimiento, luego lanzar un reporte anual de los indicadores de sostenibilidad y de inteligencia de una ciudad y recién ahí empezar a diseñar soluciones tecnológicas para resolver los problemas. Esta información debe ser totalmente abierta y utilizada por universidades, empresas, startups para que a partir de ahí se fijen retos", explica. De acuerdo con el autor los datos y la información tienen gran valor para el desarrollo de una smart city. La visualización de datos es muy importante para lograr dicho desarrollo, ya que los datos se generan de forma exponencial, se necesitan procesar y analizar para luego emplear un modo visual de forma rápida y óptima que permitan obtener información de calidad ahorrando tiempo y dinero, permitiendo la toma de buenas decisiones.

El motivo de la investigación es el análisis y visualización de datos sobre el comportamiento de la red LTE con los smart meters. Esto permite que la provincia de Cañete se desarrolle en el ámbito de smart cities para implantar un modelo de visualización de datos utilizando la herramienta cloud Quicksight.

En la provincia de Cañete la recolección de datos sobre consumo eléctrico se realiza de forma manual, un personal de la empresa eléctrica visita cada domicilio por áreas y sectores, registrando los datos de consumo eléctrico de cada medidor convencional con una frecuencia de 30 días. Esto genera que los datos no sean 100% confiables, redundancia de datos, perdida de registros, procesamiento y visualización de datos a destiempo. Una de las herramientas analíticas de visualización más potentes para filtrar y traducir datos en información es la herramienta Amazon Quicksight. Por tal motivo el objetivo de esta investigación es implantar la herramienta cloud Quicksight para la visualización de datos en tiempo real. Estos datos son obtenidos del proyecto Smart Cities Simulation: Cañete plan piloto con medidores eléctricos inteligentes.

1.1 Formulación del problema

Problema general

¿Cómo influye la herramienta cloud Quicksight en la visualización de datos obtenidos del proyecto Smart Cities: Cañete plan piloto con medidores eléctricos inteligentes?

Problemas específicos

- ¿Cómo influye la herramienta cloud Quicksight en el porcentaje de datos de calidad obtenidos del proyecto Smart Cities: Cañete plan piloto con medidores eléctricos inteligentes?
- ¿Cómo influye la herramienta cloud Quicksight en el porcentaje de lectura y visualización de datos obtenidos del proyecto Smart Cities: Cañete plan piloto con medidores eléctrico inteligentes?
- ¿Cómo influye la herramienta cloud Quicksight en el porcentaje de datos útiles obtenidos del proyecto Smart Cities: Cañete plan piloto con medidores eléctricos inteligentes?

1.2 Justificación

Desde el punto de vista social, el siguiente proyecto permitirá integrar a la comunidad del distrito de San Vicente de Cañete a la nueva temática de Smart Cities y Cloud Computing, contribuyendo al desarrollo sostenible de esta localidad.

Desde el punto de vista económico, el proyecto traerá un gran beneficio económico a la empresa proveedora de energía eléctrica ya que se reducirá gastos al tener la recolección de datos automatizada y a los usuarios porque permitirá el acceso a la información en todo momento y desde cualquier lugar vía online para la toma de decisiones correspondientes.

Desde el punto de vista tecnológico, permitirá el acceso a nuevas tecnologías dentro del entorno de Smart Cities contribuyendo a la automatización de varios procesos locales tales como el gobierno electrónico, telemedicina, seguridad ciudadana, medio ambiente y energías renovables.

Desde el punto de vista medio ambiental, este proyecto generará el uso de energías renovables, en base a la información adecuada que se logre obtener y visualizar, ya que el concepto de Smart Cities es lograr una sociedad inteligente con un desarrollo sostenible utilizando energías que no dañen el medio ambiente.

1.3 Formulación hipótesis

Hipótesis general

HG: La implantación de la herramienta cloud Quicksight permitirá visualizar de forma eficiente los datos obtenidos de la simulación Smart Cities: Cañete plan piloto con medidores eléctricos inteligentes.

Hipótesis especificas

- HE1: La herramienta cloud Quicksight aumentará el porcentaje datos de calidad obtenidos de la simulación Smart Cities: Cañete plan piloto con medidores eléctricos inteligentes.
- HE2: La herramienta cloud Quicksight aumentará el porcentaje de lectura y visualización de datos obtenidos de la simulación Smart Cities: Cañete plan piloto con medidores eléctricos inteligentes.

- HE3: La herramienta cloud Quicksight aumentará el porcentaje de datos útiles obtenidos de la simulación Smart Cities: Cañete plan piloto con medidores eléctricos inteligentes.

1.4 Objetivos de la investigación

Objetivo general

Implantar la herramienta cloud Quicksight para la visualización de datos obtenidos de la simulación Smart Cities: Cañete plan piloto con medidores eléctricos inteligentes.

Objetivos específicos

- Determinar cómo influye la herramienta cloud Quicksight en el porcentaje de datos de calidad obtenidos de la simulación Smart Cities: Cañete plan piloto con medidores eléctricos inteligentes.
- Determinar cómo influye la herramienta cloud Quicksight en el porcentaje de lectura y visualización de datos obtenidos de la simulación Smart Cities: Cañete plan piloto con medidores eléctricos inteligentes.
- Determinar cómo influye la herramienta cloud Quicksight en el porcentaje de datos útiles obtenidos de la simulación Smart Cities: Cañete plan piloto con medidores eléctricos inteligentes.

1.5 Contribución

Esta investigación permitirá conocer el porcentaje y obtención de datos válidos para utilízalos en la toma de decisiones sobre el comportamiento de la red LTE con los smart meters y como referencia para fututas investigaciones que contribuyan al desarrollo sostenible de la provincia de Cañete.

II. Marco Teórico

2.1 Antecedentes de la investigación

A nivel internacional

(Medina, 2016): “En este documento se dará a conocer un método para dar solución al decreto 0348 “Por el cual se reglamenta el servicio público de transporte terrestre automotor especial y se adoptan otras disposiciones” mediante el uso de Cloud Computing, en donde se explicara el funcionamiento de los diversos servidores utilizados en el sistema, como trabaja cada uno de ellos de manera independiente y como pueden complementarse unos con otros, se compararan dos escenarios en los cuales se implementara el sistema de maneras distintas, cada servicio se clasificaran según su función (conectividad, almacenamiento o computo), además se expondrán los resultados obtenidos de las pruebas realizadas por el sistema “piloto” y se compararan los costos aproximados que produciría cada escenario dispuesto en la plataforma ofrecida por Cloud Computing.”

Este trabajo nos habla sobre la implementación de un servicio cloud, aplicada a una simulación en un sistema piloto, para explicar el funcionamiento de dicho servicio, complementación con otros y comparar costos de dos escenarios simulados.

(Rodriguez, 2017): “El sector público, en el cumplimiento de sus funciones, genera, gestiona, sufraga y conserva una gran cantidad de información de calidad. Dicha información es potencialmente reutilizable tanto por los ciudadanos como por las empresas. Esa información que obra en poder de las instituciones y administraciones públicas presenta un potencial económico y social enorme a la par que contribuye a la transparencia en la gestión pública y a la generación de confianza por parte de la ciudadanía.

La finalidad de las iniciativas open data consiste en hacer públicos los datos en poder de las administraciones públicas y el sector privado para que puedan ser posteriormente reutilizados con la finalidad de generar nuevos conocimientos y aplicaciones que, en muchos casos, añadirán un valor al dato primario y contribuirán a potenciar la innovación. Este reciente paradigma provee a las empresas de enormes oportunidades para sacar provecho de diversos tipos de

datos que contribuyan a la toma de decisiones acertadas y por ende mayores posibilidades de éxito en un entorno tan competitivo.

La apertura de los datos conlleva beneficios para todos los que participan de ella: para los ciudadanos, para las empresas y para la propia administración. Para las empresas porque a partir de estos datos pueden diseñar y crear nuevos servicios y productos con los que se mejora su productividad y su competitividad; para los ciudadanos porque les permite estar mejor informados respecto a la gestión pública lo que contribuye a que tomen mejor sus decisiones; a la propia administración pública porque el hecho de proporcionar una mayor cantidad de información a sus ciudadanos permite a estos la búsqueda de soluciones a problemáticas para las cuales las instituciones no tienen tiempo ni recursos.

El reto está en procesar cantidades enormes de datos y convertirlas en decisiones inteligentes y oportunas, por eso se hace necesario aprovechar esta importante oportunidad de crear valor al combinar y utilizar con eficacia conjuntos de datos de diversas fuentes.

La Ingeniería Dirigida por Modelos es una propuesta que se ha venido trabajando desde hace varios años por diversos autores que plantean el uso de modelos como eje fundamental en todo el ciclo de vida de un proyecto de software que reduce el tiempo y esfuerzo en el desarrollo. Este es el camino propuesto para la presentación de un metamodelo, materializado en un lenguaje de dominio especifico, que a partir de una fuente de datos abiertos captura, transporta, transforma y analiza datos, y finalmente permite visualizar información que apoya la toma de decisiones."

Esta investigación habla sobre la enorme cantidad de información que genera el sector público y como poder utilizarla para generar nuevos conocimientos para potenciar la innovación, sacando provecho a la buena toma de decisiones.

(Saporíti & Agustín, 2017): "Uno de los principales objetivos de este trabajo de tesis fue el de implementar una herramienta que sirva para analizar los sentimientos respecto a un tema, a través de minería de opiniones. Pensamos a esta herramienta como un potencial producto. Un caso de uso que habíamos pensado era que pudiera ser utilizada por la gerencia de una empresa para conocer qué opina una población sobre ella. Este conocimiento podría servir

como apoyo para tomar decisiones, respecto a sus productos o su publicidad, por ejemplo. En nuestro sistema integramos distintas herramientas y comparamos sus resultados. Los distintos análisis que realizamos en el capítulo 4 sugieren que las tres herramientas manejan distintas formas de detectar opinión. Esto nos lleva a pensar que la elección de un servicio de análisis de opinión puede ser determinante en los resultados que le llegan al usuario, y por tanto a las decisiones que pueda tomar en función de éstos."

El trabajo de investigación trata sobre el análisis de sentimientos en comentarios realizados en las redes sociales, como procesarlos y llevamos a un ambiente grafico para la toma de decisiones.

(Gónzales, 2019): "El procesamiento de datos en tiempo real es un paradigma de los sistemas Big Data que consiste en procesar pequeños lotes de datos con un tiempo de respuesta muy bajo, típicamente milisegundos. El procesamiento y análisis de datos se realiza bajo un flujo continúo llamado streaming, en el cual los datos circulan a través de componentes del sistema que esté implementado. A este enfoque de diseño se le conoce como fast data. Las aplicaciones de procesamiento de datos en streaming tienen usos en áreas multidisciplinares como la detección de fraude, sistemas de recomendación o monitorización de pacientes en el ámbito de la salud, entre otros. Sin embargo, desarrollar este tipo de aplicaciones es un reto tecnológico ya que requiere de amplios conocimientos sobre lenguajes de programación, algoritmos, arquitectura Big Data e infraestructura. Además, tienen unos estrictos requisitos de escalabilidad, rendimiento y estabilidad. En este Trabajo de Fin de Máster (TFM) se presenta una metodología para el desarrollo de aplicaciones de datos en streaming basada en el paradigma del Desarrollo de Software Dirigido por Modelos. Las aplicaciones se diseñan a alto nivel mediante un lenguaje visual de dominio específico, de forma que no es un requisito que los usuarios tengan conocimientos técnicos previos en la materia. Para dar soporte a esta metodología, se ha desarrollado la herramienta open-source Datalyzer. Datalyzer es una plataforma web donde cada usuario tiene un espacio personal para crear, almacenar y ejecutar aplicaciones de forma auto-administrada. El lenguaje de dominio específico que este TFM presenta está integrado en la plataforma en un editor gráfico que contiene una serie de componentes que se

conectan entre sí representando flujos de datos. Datalyzer incorpora una librería de fuentes de datos pre-configuradas y procesadores que se pueden añadir a las aplicaciones. Además, se ofrece un punto de extensión para que los usuarios puedan añadir nuevas fuentes de datos y procesadores personalizados con la creación de modelos que contienen la configuración de los mismos. Una vez completado el proceso de diseño, las aplicaciones se generan de forma autónoma usando un generador de código. La herramienta tiene un panel de control donde se puede monitorizar y gestionar las aplicaciones que se estén ejecutando en Datalyzer. Además, en el panel de control también se pueden visualizar los datos que las aplicaciones procesan mediante diversos gráficos, tablas y otros visualizadores. Finalmente, se evalúa el sistema implementado con pruebas de rendimiento y escalabilidad. También se expone un caso de estudio basado en el catálogo de datos abiertos de la Comunidad de Madrid que contiene información de tráfico y contaminación en tiempo real con el objetivo de generar una aplicación de monitorización y alertas."

Esta investigación habla sobre el cómo procesar datos en tiempo real con el uso de la herramienta Datalyzer, herramienta web open-source desarrollada por el autor para procesar datos y llevarlos a un ambiente de visualización gráfica.

A nivel nacional

(Villanueva, 2015): "El presente proyecto final de Tesis, es un proyecto de aplicación a nivel de prototipo, que brindará una solución de Business Intelligence en el área de proyectos de una empresa bancaria. Esto se realizará a través del uso de tecnología SaaS. Las ventajas de la tecnología SaaS en comparación con las aplicaciones locales son evidentes para las empresas (la mayoría de ellas foráneas), el tiempo de amortización de la inversión es menor, la experiencia del usuario es mejor y sencilla, requiriendo de muy poca capacitación. En tal sentido se abordará el presente proyecto final de tesis, presentado los objetivos, hipótesis y alcance del mismo. En lo que respecta al Estado del Arte, se presentará un marco teórico que permita conocer los conceptos de las herramientas a utilizar (Datawarehouse, Business Intelligence, Tecnología en la nube y SaaS) entre otros.

Posteriormente se construirá el prototipo, para ello se hará inicialmente un diagnóstico situacional de la empresa en estudio, que nos permita tener un punto

de referencia con respecto a los objetivos estratégicos de la empresa, los factores críticos de éxito y el riesgo asociado. Lo cual nos permitirá visualizar si las estrategias están alineadas a los procesos y estos a su vez con las aplicaciones de sistemas de información que trabajan en la empresa de estudio y con énfasis en el área de proyectos. También se hará un análisis de procesos y el estado de la base de datos en la empresa bancaria, que nos permitirá evaluar los inputs para el Data Mart a implementar. Luego se ejecutará un proceso ETL y se hará el análisis multidimensional que nos permitirá tener un Data Mart y afinar el prototipo inicial que correrá en una prueba piloto en la muestra definida, haciendo los ajustes pertinentes en esta etapa."

Esta investigación habla sobre el uso de una herramienta cloud computing de inteligencia de negocios, para mejorar los procesos de una entidad bancaria trabajando con sistemas de información, también para evaluar sus bases de datos para poder implementar un Data Mart.

(Cárdenas Lino, 2016): "El objetivo principal del presente trabajo es describir la simulación y el análisis de desempeño de la red de comunicaciones LTE para infraestructura de medición avanzada (AMI) en las redes eléctricas inteligentes, o Smart Grid por su nombre en inglés. Para ello se elabora primero el estado del arte en temas de infraestructura de medición avanzada y su intervención para recolectar datos ocupando tecnologías de comunicaciones. Se crea un entorno de desarrollo para simulaciones utilizando OMNeT++ donde se implementan y modelan diferentes escenarios donde cohabiten y convergen dispositivos de medición de consumo eléctrico y terminales móviles, asegurando que tengan características de movilidad para cada uno.

Se utiliza la herramienta de simulación en escenarios con datos geo-referenciales reales de la ciudad de Santiago incluidos las posiciones de estaciones base de una empresa de telecomunicaciones, se realizaron diversas simulaciones para validar el comportamiento del simulador, garantizando los detalles técnicos que tiene la red LTE a su vez que se utiliza como base para la elaboración de diversos escenarios optimizados.

Se construye modelos avanzados donde aplicamos investigaciones colaborativas para optimizar posiciones de estaciones base eNodeB, conociendo

la ubicación de los nodos y el área geográfica donde se aplica la optimización. Se aplica el simulador para observar el comportamiento de la red con datos de modelos reales y determinar una cobertura mínima para los smart meters.

El aporte de este trabajo se centra en crear una herramienta que sirve para modelar y simular una red de comunicaciones móviles con clientes de una red de medición avanzada adicional a los clientes móviles tradicionales. Además, se propone continuar la línea de investigación con nuevas tecnologías que contengan un mayor alcance, además con la intervención de nuevas tecnologías computacionales para el desarrollo de Smart Grid Evolucionado."

Esta investigación trata sobre una simulación para observar el comportamiento de la red LTE en las smart grid, creando una herramienta para modelar y simular comunicaciones de móviles.

(Palacios Frisancho, 2017): "Nuestro país uno de los países más reconocidos por su agricultura, llegando a estar entre los primeros puestos en varios productos de exportación, sin embargo, no está libre de perdidas, por diversos aspectos, pudiendo perderse cultivos antes, durante y después de la cosecha, así como también se puede llegar a perder cultivos completos por la falta de prevención y estimación de posibles amenazas.

Junto a ello viviendo en una época de auge tecnológico, en el capítulo I vemos cómo podremos hacer posible esta difícil unión de estos dos aspectos: tecnología y naturaleza, sin alterar el uno al otro y mejorar la productividad en el campo agrícola.

En el capítulo II se desarrolló los conceptos generales sobre la quinua y de manera más amplia y profundo el de Internet del Todo, que es el nuevo concepto revolucionario de las telecomunicaciones, ya que nos dará ahora una nueva visión del futuro, un futuro donde podremos comunicar los campos con las personas sin la necesidad de estar presentes, poder monitorizar de manera remota y dejar que el mismo campo tome sus decisiones, poder prever posibles atentados en contra de un cultivo y lograr un mejor producto final.

Para lo cual en el capítulo III se desarrolló el uso de la plataforma de desarrollo de la empresa Particle, denominado Electron, un dispositivo electrónico que nos

permitirá comunicar los sensores, de donde recolectaremos los datos para luego ser procesados y tenerlos en nuestra Red Fog, para luego poder ser visualizados de manera gráfica por Internet en cualquier dispositivo que se encuentre conectado, finalmente en el capítulo IV vemos los resultados de estas interacciones."

Este trabajo mostrara el uso de la plataforma de desarrollo Electron, la cual permite la recolección de datos mediante sensores para luego tenerlos en la Red Fog donde serán visualizados de manera gráfica a través de internet.

(Salazar, 2017): "El principal objetivo de este trabajo, ha sido la implementación de un Datamart enfocado para el área comercial – Ventas de la empresa Azaleia del Perú, que permita apoyar la toma de decisiones y crecimiento de ventas en el mercado bajo los lineamientos estratégicos de la empresa.

La implementación parte desde el análisis realizado al proceso del área de ventas de la empresa Azaleia del Perú, dónde se pudo evidenciar varios puntos importantes de los cuales resaltan tres. Como primer punto, el manejo de diferentes sistemas que contienen información del área, genera para el usuario carga operativa en la obtención y consolidación de esta. Como segundo punto, se encuentra la dependencia generada con el área de sistemas, debido al mantenimiento que realizan en la estructura de la base de datos para poder obtener la información solicitada por los diferentes usuarios generando un cuello de botella lo cual impacta en tiempo de respuesta en las ventas. Y como tercer punto, la integridad en la información como soporte a la toma de decisiones de la gerencia general.

Como resultado de la investigación se tendrá el diagnóstico y solución a implementar en la empresa Azaleia del Perú, la cual se enfoca en dos puntos relevantes: mejora en la obtención de la información por parte de los usuarios, reduciendo la carga operativa y dependencia al área de tecnología de información y un mejor monitoreo de los indicadores, que permita a la gerencia general identificar patrones en el comportamiento de las ventas, permitiendo obtener respuestas más acertadas basado en la demanda del mercado, para la toma de decisiones."

Esta investigación habla sobre los datos que generan sus áreas y los datos que solicitan los usuarios, tienen en cuanta la integridad de los datos, buscan la solución para mejorar la obtención de datos de sus usuarios para identifica patrones de comportamientos en las ventas para la toma de decisiones.

(Peralta, 2019): “La visualización de grandes cantidades de datos es una de las principales tareas que realiza un analista de datos. En sistemas tradicionales de manejo de datos, registros de enormes conjuntos de datos no pueden ser consultados por su similitud debido a su complejidad, en términos de volumen y multiplicidad. En esta tesis, proponemos un enfoque efectivo para la indexación de millones de elementos, con el propósito de ejecutar simples y múltiples consultas visuales de similitud sobre datos multidimensionales asociadas a una ubicación geográfica. Nuestro enfoque hace uso del método Z-order curve para mapear nuestro conjunto de datos en una alta dimensionalidad a un espacio de una dimensión considerando la similitud entre los datos.

Respaldamos nuestra propuesta mediante la comparación con otros métodos del estado del arte en la literatura, utilizando métricas de preservación de vecindad y analizando las ventajas y desventajas entre estos métodos. Adicionalmente, presentamos un conjunto de resultados usando datos reales de diversas fuentes y analizamos los conocimientos obtenidos a partir de su exploración interactiva.”

Esta investigación habla sobre la generación de consultas visuales simplificadas a partir de la generación de enormes cantidades de datos de una determinada ubicación geográfica.

2.2 Bases Teóricas

Smart cities (ciudades inteligentes)

En la actualidad vivimos en un mundo cambiante, donde la tecnología avanza exponencialmente, hablamos de la era digital.

Con ayuda de la tecnología los países pueden mejorar la planificación, administración y el desarrollo de un gobierno sostenible, permitiendo maximizar oportunidades económicas y minimizar los daños ocasionados al medio ambiente.

Para esto ya estamos hablando de Smart Cities (ciudades inteligentes), las cuales buscan usar las tecnologías para lograr una ciudad sostenible y evitar el uso de energías que contaminen el ambiente. Hablar de Smart cities también abarca temas como:

- Sistema de infraestructura.
- Plataformas inteligentes.
- Espacio urbano.
- Energías renovables.

Para conocer más acerca de Smart Cities veamos los siguientes conceptos:

“Smart City busca por medio de soluciones basadas en TIC un desarrollo sostenible de la ciudad, generando un gran impacto en su calidad de vida, productividad y competitividad”. **(Cluster tecnología, 2011)**

“Una ciudad inteligente es aquella que coloca a las personas en el centro del desarrollo, incorpora Tecnologías de la Información y Comunicación en la gestión urbana y usa estos elementos como herramientas para estimular la formación de un gobierno eficiente que incluya procesos de planificación colaborativa y participación urbana”. **(Bouskela et al., 2016)**

“Son ciudades inteligentes aquellas que disponen de un sistema de innovación y de trabajo en red para dotar a las ciudades de un modelo de mejora de la eficiencia económica y política permitiendo el desarrollo social, cultural y urbano.” **(Grupo Interplataformas de Ciudades Inteligentes, 2015)**

“Nosotros podríamos definir el concepto de inteligencia como capacidad para resolver un problema, y, en consecuencia, definimos a la Ciudad Inteligente (Smart City) a aquella Ciudad que es capaz de afrontar los problemas típicos.” **(Bayod, 2015)**

“Ciudad inteligente es aquella que mediante la incorporación de tecnologías, procesos y servicios innovadores garantiza su sostenibilidad energética, medioambiental, económica y social, para mejorar la calidad de vida de las personas y favorecer la actividad empresarial y laboral”. **(Grupo Interplataformas de Ciudades Inteligentes, 2015)**

De los siguientes conceptos podemos deducir que Smart Cities son aquellas ciudades que hacen el buen uso de las tecnologías con el objetivo de avanzar de manera sostenible cuidando el ambiente.

Implementación de arquitectura de una Smart Cities

Existen algunos lineamientos que se deben aplicar para obtener el mejor resultado en la evolución de una Smart City y así aprovechar las ventajas que nos brinda la era digital.

Actualmente las personas hacen uso indispensable de los smartphone, para uso de las redes sociales mayormente, aprovechando esto, se aplica la recolección de datos para plantear una solución con dispositivos conectados que sean capaces de captar diferentes señales del medio ambiente, comportamientos de la ciudadanía con el fin de obtener información y vincularla a un proceso de analytics, datos abiertos y Big Data para la toma de decisiones rápidas y eficientes para prevenir imprevistos.

También se habla de la comunicación ‘Machine to Machine’ con el uso de dispositivos inteligentes conectados a internet.

Esto nos genera gran cantidad de información y será necesario crear centros de operación para el control, el cual estará implementado con computadoras y aplicaciones las cuales, recolectaran, procesaran y analizaran los datos recolectados por los dispositivos de minería de datos. En estos centros se podrán tomar las decisiones correspondientes a temas de seguridad, al tener monitorizada la ciudad con cámaras, tema de transporte con ayuda de sensores

de movimiento en las calles que ayudara con el control inteligente de los semáforos y temas del ambiente con el uso de energías limpias y renovables para mejorar la calidad de vida.

Es bueno contar con una buena infraestructura de conectividad que soporte la carga masiva de información que pueden generar las aplicaciones, equipos, que garantice la cobertura de toda una ciudad y cada uno de los ciudadanos, con el apoyo de las tecnologías de red de datos como la fibra óptica y redes inalámbricas 3G y 4G – LTE.

Modelo de Smart City

Actualmente existe un modelo que engloba todos los aspectos tecnológicos que están relacionado con las ciudades de un modo equilibrado.

El modelo tiene 4 áreas temáticas:

1. Energía y medio ambiente.
2. Edificios e infraestructuras.
3. Movilidad e intermodalidad.
4. Gobierno y servicios sociales.

“A estos bloques se adhiere un área horizontal que integra tecnologías que son aplicables a cualquiera de las áreas temáticas verticales, y que comprende las Tecnologías de la Información y la Comunicación, Sensores, Seguridad y Materiales. En total el modelo está compuesto por 5 áreas a su vez contiene varias líneas o aplicaciones sobre las que se agrupan los elementos tecnológicos que se identifican más adelante en el documento”. **(Grupo Interplataformas de Ciudades Inteligentes, 2015)**

El esquema que se muestra a continuación resume el contenido del modelo:

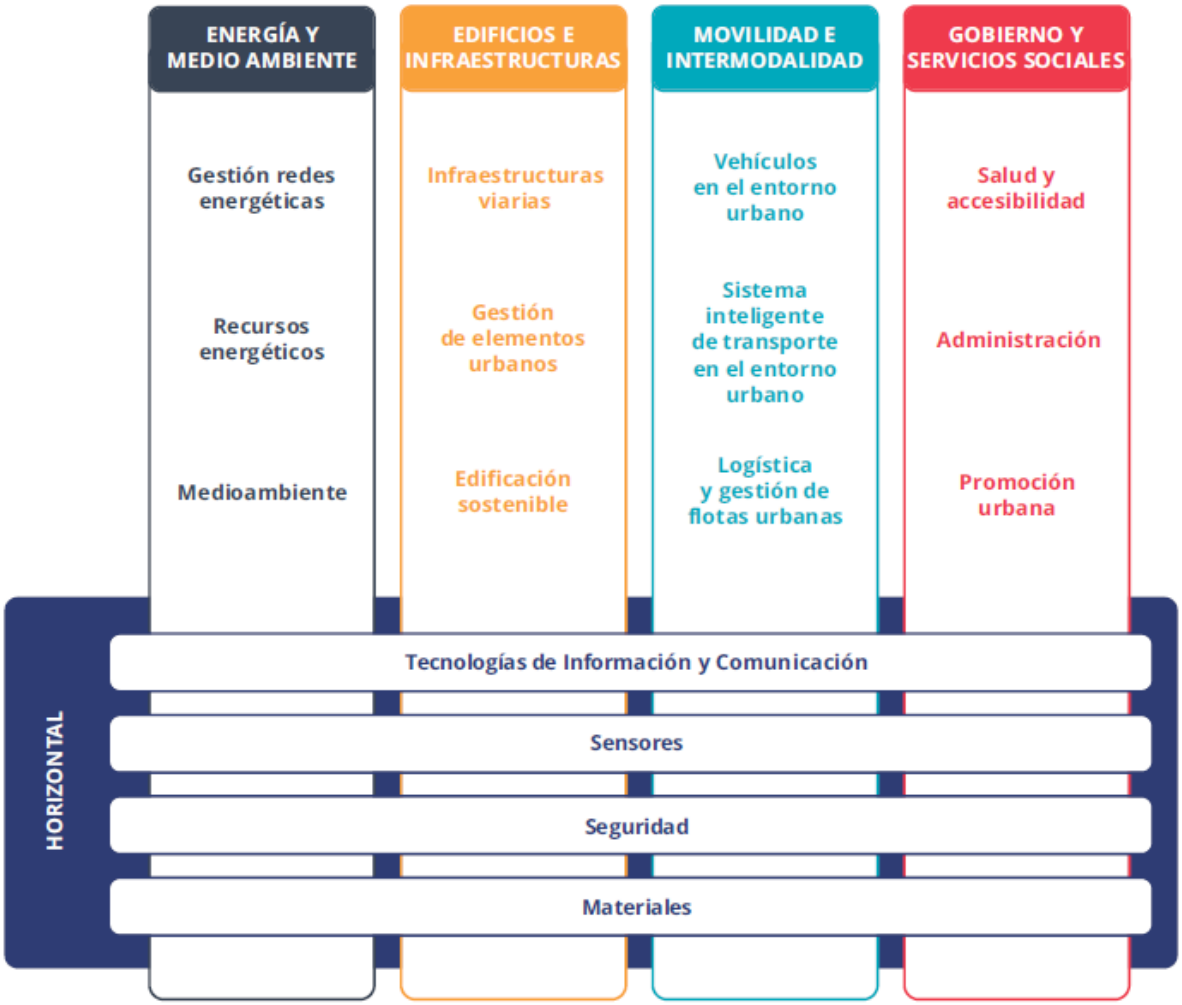


Figura 1: Áreas verticales y horizontales del modelo de Smart City.

Fuente: Smart Cities documento de visión a 2030 (2015:22).

Conexión con los retos sociales

El uso de la tecnología debe prestar servicios a la sociedad con la finalidad de hacer la vida más fácil y permitir el desarrollo de un entorno urbano inteligente para lograr una visión en común que cuenta con parámetros sociales, económicos y ambientales.

La Comisión Europea planteo unos lineamientos para que los proyectos Smart Cities estén encaminados a:

- ✓ Impulsar la demanda del mercado de soluciones inteligentes en la ciudad mediante la concienciación hacia los consumidores acerca de las nuevas

tecnologías y los procesos utilizados en la implementación de soluciones Smart Cities.

- ✓ Apoyar con el uso herramientas de adquisición y planificación de inversión.
- ✓ Promover la participación y aceptación por parte de los ciudadanos.
- ✓ Garantizar las compras públicas conjuntas, transfronterizas vinculadas a la implementación de Smart Cities.

"Una de las prioridades del Programa Horizonte 2020 (H2020) para la Investigación y la Innovación en la Unión Europea está dedicada a los retos de la sociedad, que trata de remarcar la conexión de la investigación con el fin último indicado de mejorar la calidad de vida de las personas. La labor de GICI está alineada con las prioridades políticas y los retos de la Estrategia Europa 2020 reflejadas en este programa, con el fin estimular la investigación e innovación que permitan alcanzar los objetivos políticos de la UE". **(Grupo Interplataformas de Ciudades Inteligentes, 2015)**

Retos sociales definidos por la Comisión Europea:

- ✓ Salud, cambio demográfico y bienestar.
- ✓ Seguridad alimentaria, agricultura y silvicultura sostenibles, investigación marina, marítima y de aguas interiores y bioeconomía.
- ✓ Energía segura, limpia y eficiente.
- ✓ Transporte inteligente, ecológico e integrado.
- ✓ Acción por el clima, medio ambiente, eficiencia de los recursos y materias primas.
- ✓ Europa en un mundo cambiante: sociedades inclusivas, innovadoras y reflexivas.
- ✓ Sociedades seguras: proteger la libertad y la seguridad de Europa y sus ciudadanos.

Es importante mantener informado a la ciudadanía sobre las iniciativas y las actividades que van encaminadas a la construcción de una Smart City, lo que se va desarrollando para que puedan participar en las acciones necesarias y contribuir en el desarrollo de la ciudad en función a sus necesidades e intereses.

Por eso es necesario establecer una estrecha relación entre la ciudadanía y los avances tecnológicos significativos en el camino hacia una Smart City para así poder garantizar que el desarrollo responda a las necesidades reales que demanda la ciudadanía.

En la siguiente imagen podemos observar los retos de la sociedad que responden principalmente a cada área que integra el modelo de Smart Cities, aunque todas las áreas tecnológicas abordan los ocho retos:

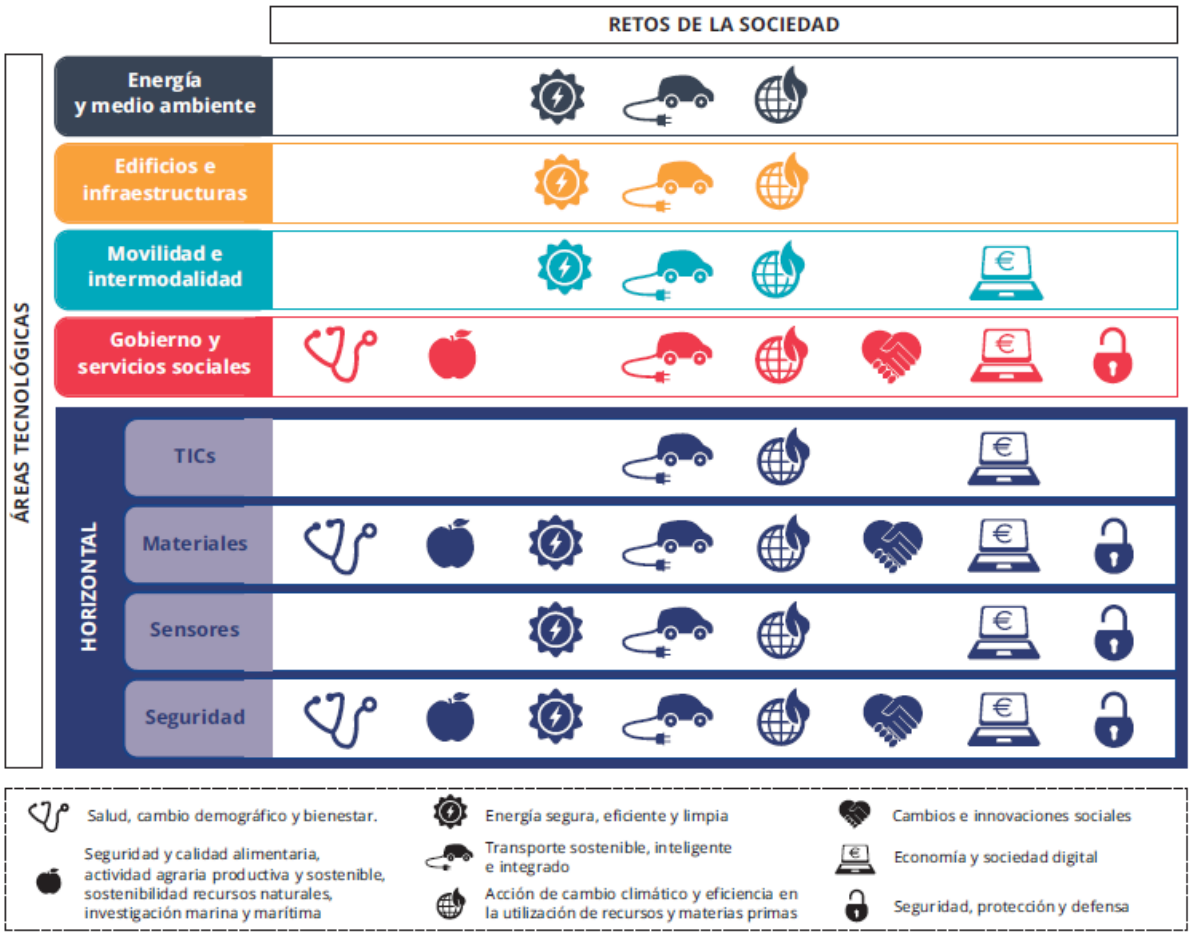


Figura 2: Retos de la sociedad y sus áreas tecnológicas del modelo de Smart City.

Fuente: Smart Cities documento de visión a 2030 (2015:27).

Smart Grid (red inteligente)

Las smart grids o redes eléctricas inteligentes son una forma de gestión eficiente de la electricidad. La Plataforma Tecnológica Europea de smart grids (ETIP), la define como una red eléctrica capaz de integrar de forma inteligente el comportamiento y las acciones de los usuarios conectados en ella (quienes generan electricidad, quienes consumen y quienes realizan ambas acciones) para proporcionar un suministro de electricidad seguro, económico y sostenible. **(López Vásquez, 2016)**

Las smart grids, surgieron como una respuesta a la necesidad de modernizar la red eléctrica, articulando los procesos de control y monitoreo con tecnologías verdes, también conocidas como no contaminantes o ecológicas. Las redes inteligentes permiten que los recursos energéticos distribuidos, por sus siglas en ingles DER se conecten de forma segura a la red. **(Gómez et al., 2018)** Las redes inteligentes son autónomas y mejoran la eficacia y eficiencia en la gestión de energía eléctrica, permitiendo que las empresas de servicios públicos optimicen la infraestructura existente, minimizando la construcción de más plantas de energía **(Kobus et al., 2015)**

Para lograrlo se usan tecnologías de última generación, creando un sistema que permita la comunicación bidireccional de electricidad e información entre el consumidor final y las compañías eléctricas, satisfaciendo sus necesidades, mejorando la eficiencia y buen funcionamiento de la red eléctrica.

Los objetivos de las smart grids son las siguientes:

- Garantizar óptimos niveles de fiabilidad, seguridad y calidad del suministro. Reducir el impacto medioambiental del sistema eléctrico de suministro. Proporcionar a los consumidores mayor información de oferta.
- Permitir a los consumidores formar una parte importante en la optimización del sistema.
- Facilitar y mejorar la conexión y el funcionamiento de los generadores de todo tipo de tamaños y tecnologías.

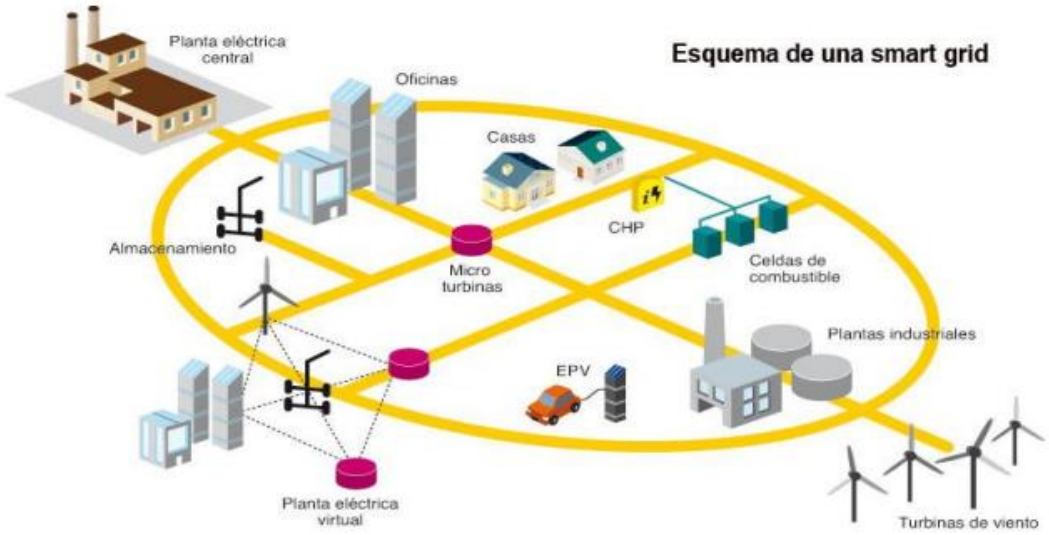


Figura 3: Esquema de una smart grid.

Fuente: Estudio de la situación actual de las Smart Grids (2016:17).

Smart Meter (medidor inteligente)

Un medidor eléctrico inteligente se encarga de medir el consumo detallado de energía, también proporciona otras funcionalidades y/o servicios que los medidores convencionales no. En base a sus funcionalidades, los medidores inteligentes se dividen en dos grupos: AMR, que están habilitados para telemetría y AMI, que están habilitados para controlar tareas de gestión remotamente. Los medidores inteligentes, a diferencia de medidores convencionales, tienen una medida tecnología de control 100% automatizado, remoto y con cifrado de datos. Sin lugar a dudas, los medidores inteligentes cambiar la forma en que gestionamos la energía y los dispositivos móviles con la red de comunicación LTE, lo que nos lleva a imaginar el inicio de las ciudades inteligentes. Además, la seguridad de infraestructura de medición avanzada (AMI) es esencial para ganar la confianza del consumidor y proteger sus datos. Los autores sostienen que los beneficios, como ahorrar en facturas de energía y el costo reducido de la lectura del medidor, son un beneficio incalculable, que transforma el sistema de energía obsoleto en un sistema inteligente. **(Pacheco et al., 2020)**

Podemos afirmar que los smart meters son medidores automatizados que capturan y envían datos sobre el consumo de energía eléctrica a través de la red

de internet, permitiendo a los consumidores poder gestionar de manera responsable el consumo de energía en tiempo real.

Para poder generar un consumo de energía responsable e inteligente por parte de los consumidores, la información de ser accesible, la cual se puede lograr mediante estas tres maneras:

- Publicación online de los datos recolectados por parte de los Operadores de la Red de Distribución, como sucede en el caso de España, donde los clientes deben registrarse en el portal web de su distribuidora para consultar su consumo con un día de desfase
- Existencia de una plataforma única que centralice los datos de la totalidad de los consumidores, como es el caso de Reino Unido.
- Acceso directo al contador por parte del cliente. Para ello, el contador debe tener habilitado un puerto de comunicación accesible, como puede ser wifi o red LAN, diferente al usado por la compañía para la lectura y operación del contador. Este método es el más complejo porque exige un contador más avanzado y la posibilidad de tener acceso físico, pero a la vez es el más potente, supuesto permite la información en tiempo real. **(Structuralia Blog, 2018)**

AMR

Automatic Meter Reading (AMR) cuyo significado en español es la Lectura de Medición Remota, es un término aplicado a la variedad de tecnologías que permiten a las empresas de servicios públicos leer consumos de gas, agua y electricidad con mayor eficiencia. Contempla una comunicación bidireccional entre el medidor y el servidor de datos. Registra la medición de energía mensual o por intervalos de tiempo predefinidos. **(Afinidad Eléctrica, 2008)**

AMI

Un sistema AMI proporciona a las empresas eléctricas las herramientas necesarias para gestionar la demanda y comportamiento de los consumidores, así como para mejorar sus procesos. Asimismo, permitirá que sea posible el monitoreo y control de la eventual producción de energía residencial, principalmente recogida desde sistemas fotovoltaicos instalados en los sectores residenciales y comerciales en general. Comprende funciones de medición,

recolección y análisis de datos de consumo de energía. Se puede utilizar para manejar comandos y datos desde la oficina central hacia los dispositivos. Además, contribuye a las empresas eléctricas a proteger sus ingresos, haciendo su trabajo más fácil y rápido, y ayuda a incrementar la calidad de sus servicios, lo que mejora la calidad de vida de los usuarios. **(Muñoz, 2019)**

Red LTE (4G)

LTE responde a las siglas Long Term Evolution (evolución a largo plazo) y hace referencia a la tecnología de banda ancha inalámbrica que sirve para la transmisión de datos con la finalidad de dar acceso a Internet a los dispositivos móviles. Se emplea para la subida y bajada de datos desde la red de redes a una alta velocidad. Fue desarrollado por la 3GPP y no es más que una evolución del estándar 3G, sin llegar a alcanzar el 4G, pese a que la mayoría de las marcas comerciales de tecnología equiparan el LTE a esta cuarta generación. De hecho, muchos han establecido que, en realidad, el LTE es un 3,9G. La principal ventaja que aportó su llegada fue el incremento de la velocidad de navegación. En concreto, el LTE permite una velocidad máxima de 170 Mbps con dos antenas base y dos en el dispositivo, o de 300 Mbps con cuatro antenas base y cuatro en el equipo. (**SoftwareLab Blog, 2018)**

Esta tecnología posee las siguientes características:

- Desarrollo y despliegue fácil y barato por parte de los operadores, ya que utiliza un protocolo de arquitectura simple, basado en el IP.
- Uso flexible del espectro radioeléctrico, pues es capaz de operar, por el tipo de duplexación, en FDD (bandas pareadas) y en TDD (bandas no pareadas).
- Baja latencia y compatibilidad con otras tecnologías 3GPP.

Cloud Computing (computación en la nube)

Es un modelo de computación que ofrece servicios informáticos a través de internet, de tal forma que los recursos, como: hardware, software, y datos, se pueden ofrecer al cliente bajo demanda.

Cloud computing es una mejor forma de gestionar un negocio. En lugar de ejecutar una aplicación uno mismo, esto se puede hacer desde un centro de datos compartido, de forma rápida y con un menor costo. Un ejemplo seria la plataforma Gmail, donde no se necesita de un servidor ni equipo especializado que controle el funcionamiento ni las actualizaciones. Cuando se utiliza esta aplicación es ejecuta en la nube y solo se tiene que iniciar sesión, personalizar y comenzar a usar.

¿Qué es cloud computing?

Cloud computing no solo es una frase de moda, es un nuevo término que representa un nuevo modelo de informática y que muchos analistas consideran puede ser tan relevante como la propia web y un sinónimo de la propia web. Cloud computing es la evolución de un conjunto de tecnologías que afectan al enfoque de las organizaciones y empresas en la construcción de sus infraestructuras de TI. Al igual que ha ido sucediendo con la evolución de la web y su actual web 2.0, la informática en nube no incorpora nuevas tecnologías, sino que se han unido tecnologías potentes e innovadoras para construir este nuevo modelo y arquitectura de la web. Estas tecnologías que han configurado la nube son variadas, aunque las más notables son: virtualización, almacenamiento físico y en la web, centros de datos, software como servicio, aplicaciones web y sistemas operativos web. **(Joyanes Aguilar, 2012)**

Una primera guía de proveedores de Cloud computing

InformationWeek, uno de los portales tecnológicos de gran impacto mundial, asi como blogs tecnológicos prestigiosos como TechCrunch, revistas especializadas como PC World, PC Magazine, Wired… publican continuos informes relativos a la nube y de las empresas que le proporcionan aplicaciones. Desde Google junto con Amazon e IBM como pioneros, hasta Sun Microsystems, HP, EMC, Oracle… e incluso Microsoft, que ya a finales de junio de 2008 anuncio la creación de una plataforma de Cloud computing preparándose y especializándose en realizar aplicaciones en la nube. **(Joyanes Aguilar, 2012)**

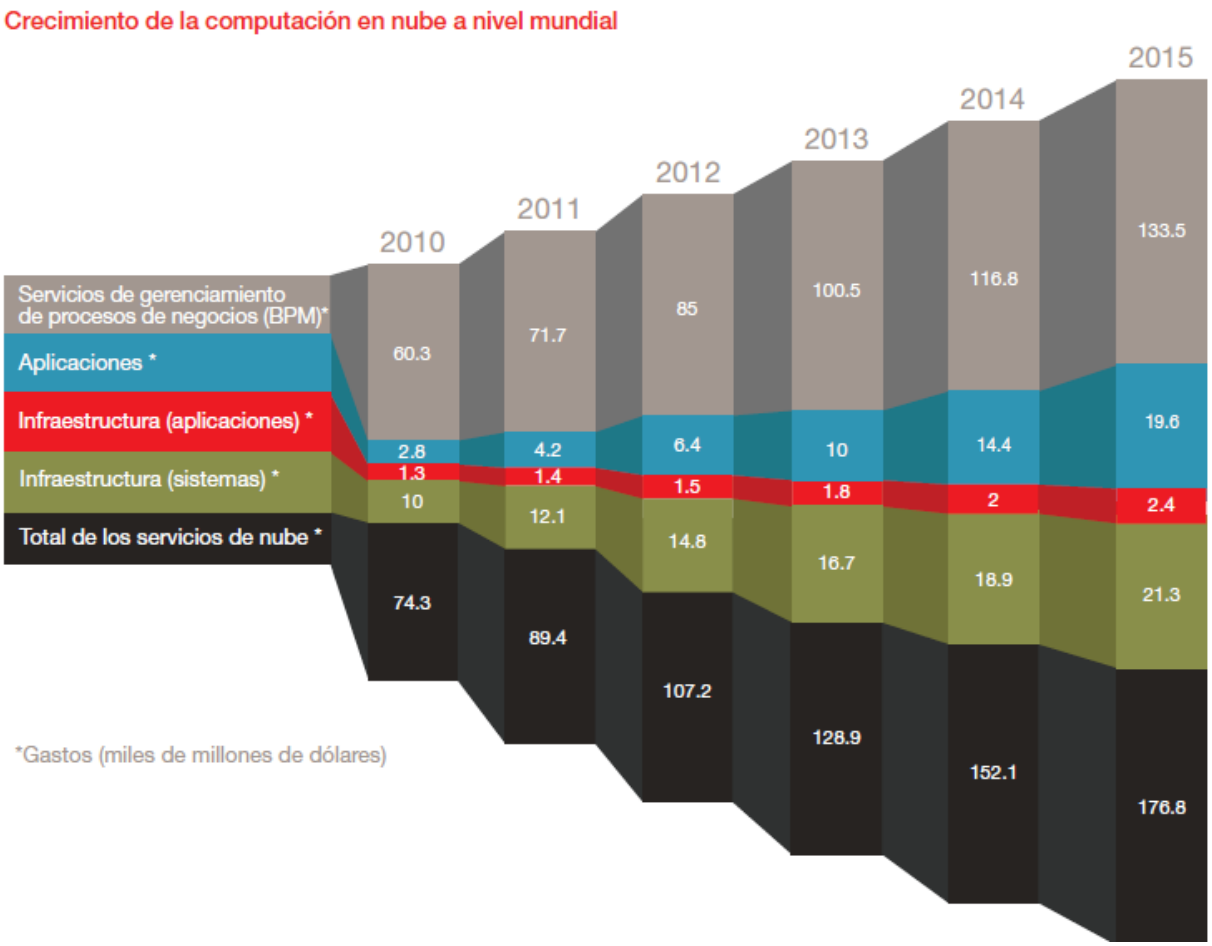


Figura 4: Crecimiento de Cloud computing a nivel mundial.

Fuente: Gartner (2011).

En la actualidad diversas empresas hacen uso de todo tipo de aplicaciones en la nube, debido a la rapidez con la que se puede trabajar, estas aplicaciones tienen menor costo dado que no se paga por la totalidad del personal, productos e instalaciones necesarias para el funcionamiento. Además, las aplicaciones en la nube son más adaptables, más seguras, más fiables y son accesibles en cualquier momento.

Ventajas

- ✓ Disponibilidad del servicio y/o aplicación web las 24 horas, 7 días a la semana y 365 días al año.
- ✓ Abstracción de problemas de mantenimiento, instalación y soporte.
- ✓ Accesibilidad mediante diferentes tecnologías compatibles, tales como: PDA's, móviles, portátiles, netbooks, etc.
- ✓ No saturación de uso del disco duro del ordenador o aplicación que se usa, debido a que solo se necesita navegador web e internet.

Desventajas

- ✓ En alguna ocasión debido a una catástrofe natural o error humano, el servicio quedara fuera de servicio.
- ✓ El almacenamiento de datos privados, fotos, videos.
- ✓ Descontrol del manejo, almacenamiento y uso de la información.
- ✓ Depende mucho de la tecnología que use el servicio, para poder acceder o no a usarlo.
- ✓ Mayor dependencia de proveedores de internet.
- ✓ Posibilidad de que delincuentes cibernéticos vulneren la seguridad del servicio.

Capas de cloud computing

Para entender mejor que es cloud computing debemos conocer sus tres capas de servicio:

- ✓ SaaS (software as a service): software como servicio, directo para su consumo por los usuarios finales. Ejemplo: ERP, CRM, correo electrónico bajo demanda, escritorio virtual, juegos.
- ✓ PaaS (platform as a service): plataforma como servicio para actividades de desarrollo o despliegue de aplicaciones. Ejemplo: servidores web, herramientas de desarrollo, bases de datos, big data.

- ✓ IaaS (infraestructura as a service): infraestructura como servicio para administración de TIC. Ejemplo: máquinas virtuales, equipos de comunicaciones, cortafuegos.

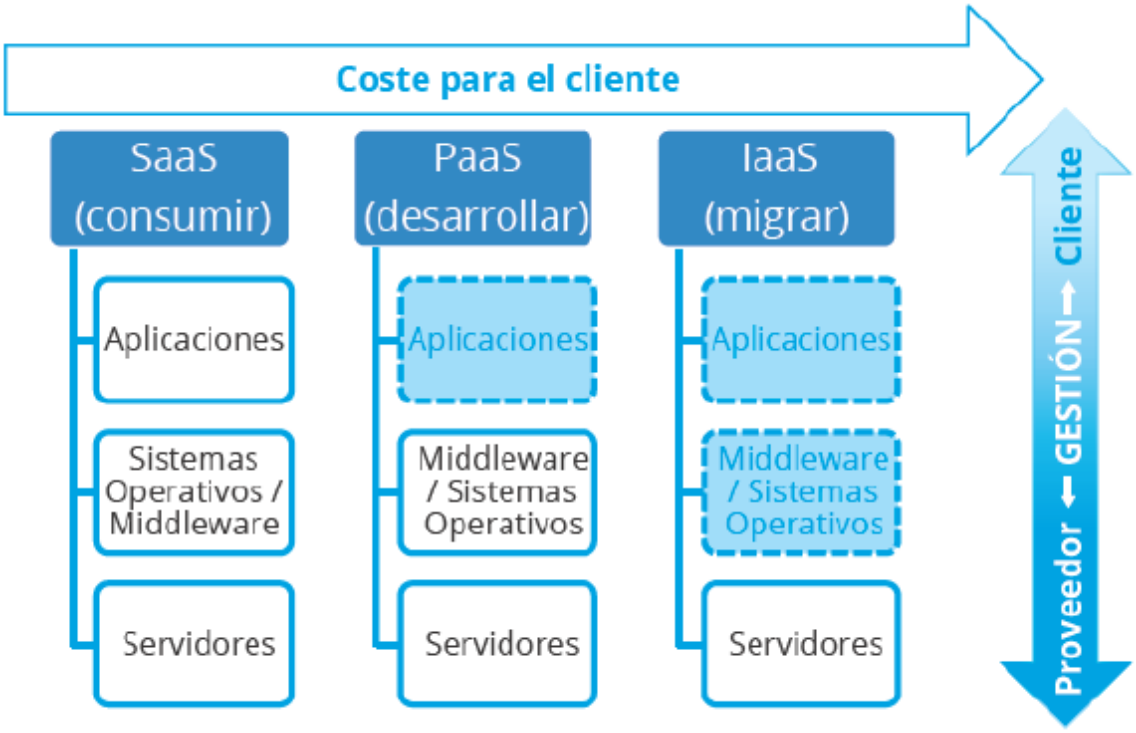


Figura 5: Capas de cloud computing.

Fuente: Cloud computing, guía de aproximación para el empresario (2017).

Big Data

"El crecimiento en el volumen de datos generados por diferentes sistemas y actividades cotidianas en la sociedad ha forjado la necesidad de modificar, optimizar y generar métodos y modelos de almacenamiento y tratamiento de datos que suplan las falencias que presentan las bases de datos y los sistemas de gestión de datos tradicionales. Respondiendo a esto aparece Big Data, término que incluye diferentes tecnologías asociadas a la administración de grandes volúmenes de datos provenientes de diferentes fuentes y que se generan con rapidez". **(Hernández Leal et al., 2017)**

En la actualidad a cada segundo se genera una gran cantidad de datos, tanto estructurados y no estructurados, dado esto, surge la necesidad de utilizar nuevas técnicas de análisis y gestión de estos datos, para esto nace una nueva metodología denominada Big Data.

“Como se ha venido comentando, el tratamiento de Big Data ha exigido el desarrollo de soluciones computacionales que permitan afrontar las necesidades y retos que traen consigo los grandes volúmenes de datos, su variedad de fuentes y la velocidad con que se generan. Como tecnologías de Big Data se clasifican aquellas que dan soporte a la captura, transformación, procesamiento y análisis de los datos, ya sean estructurados, semiestructurados o no estructurados. Seguidamente. Se decide presentar estas tecnologías ya que son software de libre uso y que permite la generación de soluciones de Big Data de acuerdo con las necesidades particulares de un dominio de datos u organización. Cabe aclarar que existen un mayor número de tecnologías que soportan Big Data, tanto libres como propietarias, pero para efectos de este documento se ha acotado de acuerdo con lo anteriormente expuesto y tomando las tecnologías que dieron las bases iniciales al ecosistema Big Data.” **(Hernández Leal et al., 2017)**

Debido a que el Big Data ha demandado la creación de soluciones que permitan afrontar las nuevas necesidades de los grandes volúmenes de datos como la captura, procesamiento y análisis, se ofrece una gran variedad de software libre para poder satisfacer estas necesidades.

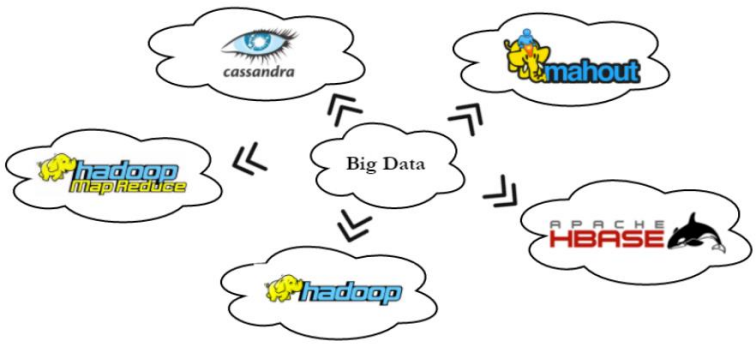


Figura 6: Tecnologías Big Data.

Fuente: Big Data: una exploración de investigaciones, tecnologías y casos de aplicación (2017).

Técnicas Big Data

"En cuanto a técnicas de Big Data, se dará una breve introducción, cabe aclarar que existen diferentes clasificaciones y que muchas de estas técnicas se aplican tanto en soluciones Big Data como en otros enfoques. Para efectos de este documento se describen, sin entrar en detalle, la minería de datos, el aprendizaje máquina (machine learning), el reconocimiento de patrones, los algoritmos genéticos y las reglas de asociación." **(Hernández Leal et al., 2017)**

Minería de datos

"Es la ciencia que se dedica a encontrar las agujas en el pajar del big data, que extrae lo valioso de entre la ingente cantidad de información para obtener lo valioso." **(Fundación Bankinter, 2015)**

Es el proceso de detectar la información procesable dentro de grandes conjuntos de datos, utilizando el análisis matemático para deducir los patrones y tendencias que existen en los datos, normalmente estos patrones no se pueden detectar mediante la exploración tradicional de los datos porque las relaciones son muy complejas o porque existen demasiados datos.

Se compone de cuatro etapas principales:

- ✓ Determinación de los objetivos: delimita los objetivos que el cliente desea, con la orientación del especialista en datamining (minería de datos).
- ✓ Preprocesamiento de los datos: selección, limpieza, enriquecimiento, reducción y transformación de las bases de datos, esta etapa consume el 70% del tiempo total de un proyecto de datamining.
- ✓ Determinación del modelo: se realiza análisis estadísticos de los datos, para luego llevar a cabo la visualización grafica de los mismos para obtener una aproximación.
- ✓ Análisis de resultados: verifica los resultados obtenidos para ver si son coherentes y se coteja con los resultados obtenidos por el análisis estadístico. Luego el cliente determina si son novedosos y si le aportan nuevo conocimiento que le permita tomar buenas decisiones.

Proceso KDD

"El Descubrimiento de conocimiento en bases de datos (kdd, del inglés Knowledge Discovery in Databases) es básicamente un proceso automático en el que se combinan descubrimiento y análisis. El proceso consiste en extraer patrones en forma de reglas o funciones, a partir de los datos, para que el usuario los analice. Esta tarea implica generalmente pre-procesar los datos, hacer minería de datos (data mining) y presentar resultados. Kdd se puede aplicar en diferentes dominios, por ejemplo, para determinar perfiles de clientes fraudulentos (evasión de impuestos), para descubrir relaciones implícitas existentes entre síntomas y enfermedades, entre características técnicas y diagnóstico del estado de equipos y máquinas, para determinar perfiles de estudiantes "académicamente exitosos" en términos de sus características socioeconómicas y para determinar patrones de compra de los clientes en sus canastas de mercado". **(Pereira Timarán et al., 2016)**

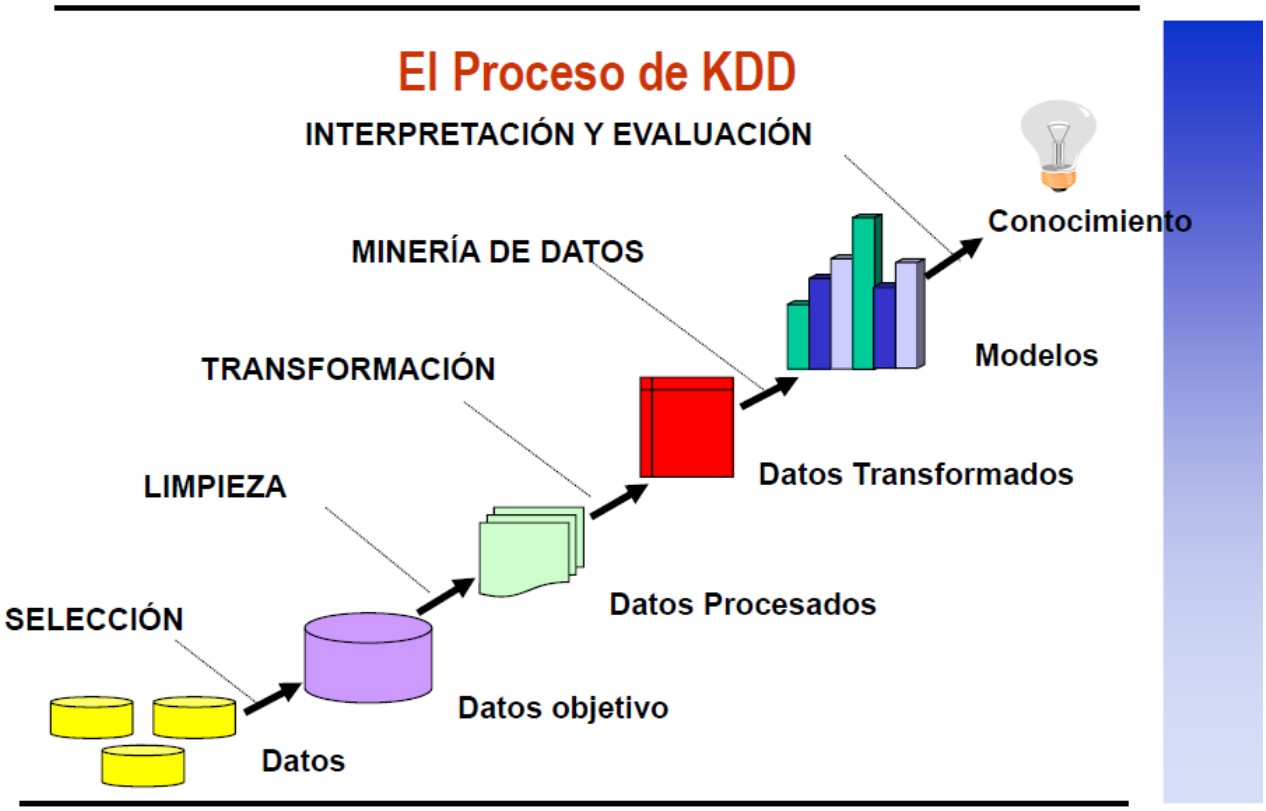


Figura 7: Proceso KDD.

Fuente: Metodología de análisis de datos (2005:5).

Amazon Quicksight

“Amazon QuickSight es un servicio de análisis empresarial que puede utilizar para crear visualizaciones, realizar análisis ad hoc y obtener información útil empresarial a partir de sus datos. Puede detectar automáticamente los orígenes de datos de AWS y trabajar también con sus orígenes de datos. Amazon QuickSight permite a las organizaciones escalar de cientos a miles de usuarios y ofrece un desempeño fiable gracias a su sólido motor en memoria (SPICE).” **(Amazon Inc., 2018)**

Quicksight es una herramienta de análisis de datos altamente escalable, fácil de usar y muy flexible que se ofrece como solución en la nube, en donde podemos crear visualizaciones graficas a partir de datos, realizando un análisis para obtener información útil.

(Agüero, 2016) la define como: “Se trata de una herramienta de BI (Business Intelligence) en memoria optimizada para analítica, muy rápida y que permite compilar visualizaciones, realizar análisis personalizados y obtener perspectivas empresariales a partir de nuestros datos. Está abierta a herramientas de terceros como Tableau.”

Preparación de datos

“Los conjuntos de datos almacenan cualquier preparación de datos que haya realizado para que pueda volver a utilizarlos en diferentes análisis. La preparación de datos proporciona opciones como, por ejemplo, añadir campos calculados, aplicar filtros y cambiar nombres de campos o tipos de datos. Si está basando el origen de datos en una base de datos SQL, también puede utilizar la preparación de datos para unir tablas. También puede escribir una consulta SQL si desea trabajar con datos de más de una sola tabla.” **(Amazon Inc., 2018)**

La preparación de datos se refiere a tener lista la información para su visualización, se puede aplicar cambios, añadir algunas consultas, filtrar e infinidad de cosas para llevar a cabo una visualización de forma eficiente.

La preparación de datos es el proceso de transformación de los datos sin procesar para usarlos en un análisis. Esto incluye la realización de cambios como los siguientes:

- Filtrado de los datos para que pueda centrarse en lo que le importa.
- Cambio de nombre de los campos para que sean más fáciles de leer.
- Cambio de tipos de datos para que sean más útiles.
- Adición de campos calculados para mejorar el análisis.
- Creación de consultas de SQL para limitar los datos.

SPICE

"SPICE es el motor de cálculo en memoria, en paralelo y ultrarrápido de Amazon Quicksight. SPICE está diseñado para realizar cálculos avanzados y ofrecer datos de forma rápida. El almacenamiento y la capacidad de procesamiento disponibles en SPICE aceleran las consultas analíticas que puede ejecutar con respecto a los datos importados. Mediante SPICE, ahorra tiempo ya que no tiene que recuperar los datos cada vez que cambia un análisis o actualiza un elemento visual." **(Amazon Inc., 2018)**

El motor encargado de realizar cálculos avanzados de forma rápida es SPICE, el cual gracias a su gran capacidad de almacenamiento y procesamiento acelera las consultas analíticas que se le realizan a los datos ingresados, este motor nos beneficia con un gran ahorro de tiempo.

Análisis de datos

"Un análisis de datos es el espacio de trabajo básico para crear e interactuar con elementos visuales que son representaciones gráficas de sus datos. Cada análisis incluye un conjunto de elementos visuales que debe reunir y organizar para sus fines, como un análisis de ventas, un análisis de costos o un seguimiento de indicadores clave de desempeño. Cada análisis puede contener historias, que puede utilizar para guardar una presentación secuencial de iteraciones diferentes del análisis. Esto resulta útil si desea mostrar los cambios a lo largo del tiempo o proporcionar comparaciones visuales de sus datos." **(Amazon Inc., 2018)**

Elementos visuales

“Un elemento visual, también conocido como "visualización de datos", es una representación gráfica de un conjunto de datos que utiliza un tipo de diagrama, gráfico, gráfico o tabla. Todos los elementos visuales comienzan en el modo AutoGraph, que selecciona automáticamente una visualización en función de los campos seleccionados. También puede tomar el control y elegir sus propios elementos visuales. Amazon QuickSight admite una serie de elementos visuales, como gráficos combinados, mapas de calor y árbol, tablas dinámicas y muchos más. O también puede mejorar las visualizaciones mediante la aplicación de filtros, el cambio de colores o la organización de varios elementos en el espacio de trabajo, por nombrar solo algunas opciones.” **(Amazon Inc., 2018)**

Lo elementos visuales son representaciones de un conjunto de datos analizados, los cuales pueden ser gráficos de calor, mapas de calor, tablas dinámicas, etc., en Quicksight existen muchos elementos visuales para escoger por el usuario.

Flujo de trabajo típico de Amazon QuickSight

La primera vez que crea un análisis, el flujo de trabajo típico se parece al siguiente:

1. Añada o cargue un origen de datos y utilícelo para crear un conjunto de datos.
2. (Opcional) Prepare los datos: téngalos preparados para los informes mediante la estandarización de nombres de campo o la adición de cálculos, por ejemplo.
3. Visualice (o cree) un nuevo análisis a partir del conjunto de datos.
4. Elija algunos campos para crear el primer elemento visual en el análisis. Puede utilizar AutoGraph para crear dinámicamente un elemento visual basándose en el número y el tipo de campos elegidos. Como alternativa, puede elegir el tipo de elemento visual que desea utilizar.
5. (Opcional) Realice cambios en el elemento visual si así lo desea (por ejemplo, añada un filtro o cambie el tipo de elemento visual).
6. (Opcional) Añada más elementos visuales al análisis. Puede cambiarles el tamaño y organizarlos en el espacio de trabajo.
7. (Opcional) Capture el análisis en una historia para crear una narrativa sobre algunos aspectos del análisis de datos.

8. (Opcional) Publique el análisis como un panel para compartir información con otros usuarios.

En la siguiente ilustración podemos observar como es el flujo de trabajo de QuickSight:

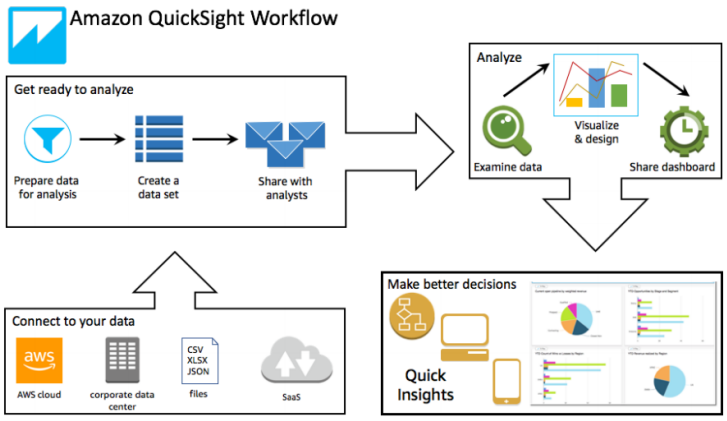


Figura 8: Amazon QuickSight Workflow.

Fuente: Amazon QuickSight: Guía del usuario (2018).

Visualización de datos

"En la actualidad se habla de visualización, como consecuencia de que hay una cierta tendencia a mostrar conjuntos ingentes de datos, elaborados en forma gráfica, cercana y didáctica para que los intérpretes lo entiendan. Se busca poner un cierto orden a la plurisemia de términos que se suelen emplear en el ámbito de las rutinas de producción, separando las visualizaciones como propiedades de los productos gráficos, de las que son de tipo complementario en los relatos visuales no necesariamente periodísticos". **(Valero Sancho, 2014)**

Actualmente el consumo de información se ha multiplicado exponencialmente debido a dos factores, los cuales son: que cada vez se produce más información y que cada vez tenemos más capacidad de acceso a dicha información a través de internet. La capacidad para sacar provecho a esta información está ligada a la capacidad para explotarla y transformarla en información útil.

Según **(Julià Minguillón, 2016)**: "La visualización de datos es un ámbito de conocimiento en constante evolución que últimamente se ha visto impulsado debido a la gran cantidad de datos disponibles, los cuales esperan a ser

analizados, interpretados y contextualizados mediante una narrativa que combina texto, imágenes y otros recursos interactivos, más allá de presentaciones estáticas que usan ciertos elementos gráficos como soporte. El boom de las redes sociales, el uso de dispositivos móviles y la práctica digitalización de todos los servicios (consumo, educación, ocio, etc.) permiten, en la práctica, disponer de datos sobre cualquier actividad humana que opere parcial o totalmente con la tecnología. Así, desde la aparición de internet a mediados de los años noventa, la capacidad de generar, procesar y compartir datos ha ido creciendo exponencialmente, se ha multiplicado por mil cada pocos años y ha obligado a adoptar nuevos prefijos (mega-, giga-, tera-, etc.) que reduzcan las cifras manejadas a cantidades razonables."

Cuando se aplica el procesamiento de datos a dicha información, esto cobra sentido y se trasforman en conocimiento. En el ámbito tecnológico, la explotación de datos está evolucionando en las últimas décadas para diseñar diversos mecanismos de interpretación. Y entre estos mecanismos de explotación el que más resalta es la visualización de datos.

"La visualización de datos es la presentación gráfica de información con dos propósitos. Por un lado, la interpretación y construcción de significado a partir de los datos (es decir, el análisis); y, por otro lado, la comunicación". **(Red.es, 2016)**

Es una herramienta muy potente para descubrir y comprender la lógica que se encuentra detrás de un conjunto datos, también para compartir esta información con otras personas desde un punto de vista objetivo, el dicho: "una imagen vale más que mil palabras" tiene mucha influencia en esta herramienta, porque se entiende que se comunica mucho mejor gráficamente que verbalmente y la representación gráfica está diseñada acorde con los principios formales que rigen la visualización de datos.

Muchas personas suelen estar observando por horas una tabla con números y no entender lo que quiere decir, pero cuando se representa esta información gráficamente la persona la entiende sin problemas y en menos tiempo, veamos un ejemplo sobre datos históricos de hipotecas:

Periodo	Entidad	Nº Hipotecas
2010	Otras entidades	103,055
2008	Otras entidades	122,332
2009	Banca	367,298
2011	Banca	258,667
2008	Banca	408,712
2010	Banca	337,825
2003	Cajas de ahorro	685,464
2011	Cajas de ahorro	279,461
2006	Banca	576,777
2006	Cajas de ahorro	1,016,728
2004	Banca	583,213
2009	Otras entidades	111,066
2005	Banca	594,477
2003	Banca	486,579
2004	Cajas de ahorro	791,349
2003	Otras entidades	119,617
2005	Cajas de ahorro	924,882
2007	Banca	518,638
2008	Cajas de ahorro	683,546
2007	Cajas de ahorro	1,017,548
2011	Otras entidades	86,420
2007	Otras entidades	154,809
2009	Cajas de ahorro	549,558
2010	Cajas de ahorro	475,536
2005	Otras entidades	184,773
2004	Otras entidades	147,843
2006	Otras entidades	202,889

Tabla 1. Tabla histórica de hipotecas. Fuente: INE (2011).

Acá podemos observar gráficamente la anterior tabla:

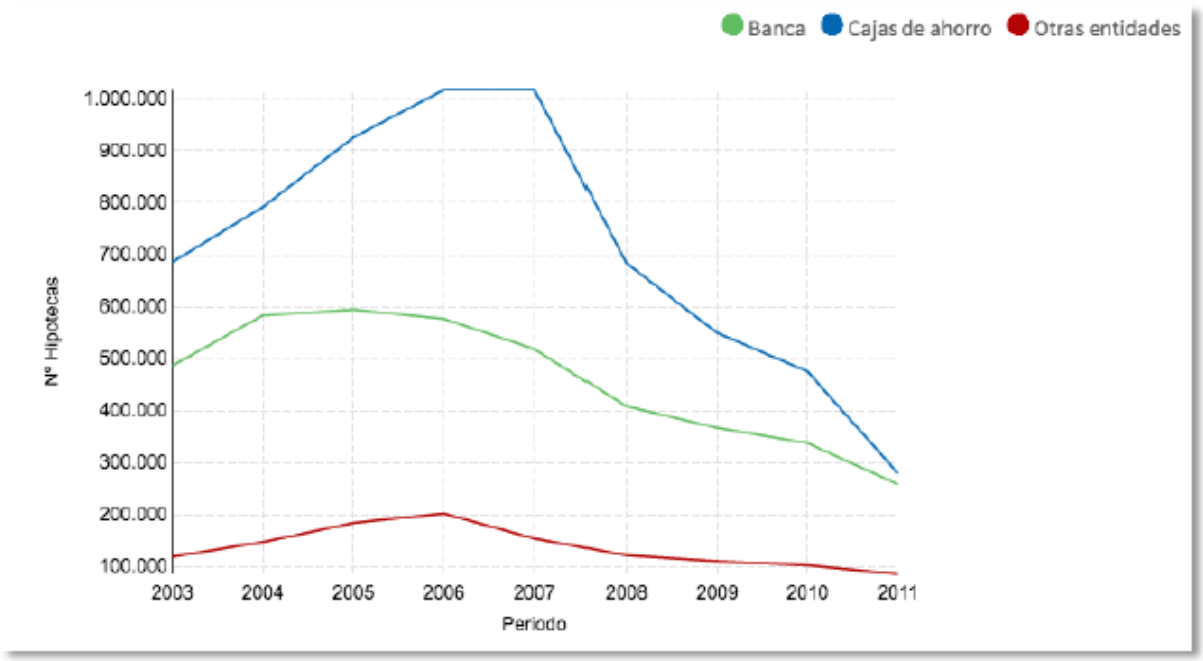


Figura 9: Grafica de hipotecas.

Fuente: INE (2011).

De esto se concluye que datos numéricos no se pueden comunicar correctamente mediante tablas, pero cuando se representan gráficamente se vuelve más visible e inteligente.

¿Qué es una visualización?

“La visualización de la información se está convirtiendo en más que un conjunto de herramientas, tecnologías y técnicas para grandes conjuntos de datos. Está emergiendo como un medio en su propio derecho, con una amplia gama de potencial expresivo ". **(Rodenbeck, 2008)**

La visualización es simplemente un conjunto de herramientas y técnicas para analizar y procesar grandes cantidades de datos.

(Manovich, 2010) nos da la siguiente definición: “A pesar de la creciente popularidad de infovis (una abreviatura común para “visualización de información”), no es así fácil de encontrar una definición que funcione para todo tipo de infovis que se están creando hoy, y al mismo tiempo lo separarían claramente de otros campos relacionados como la visualización científica y el

diseño de información. Así que comencemos con una definición provisional que podemos modificar más adelante. Vamos a definir visualización de información como un mapeo entre datos discretos y visual. representación. También podemos usar diferentes conceptos además de “representación”. Cada una aportando un significado adicional. Por ejemplo, si creemos que un cerebro utiliza una serie de distintas modalidades de representación y cognitivas, podemos definir infovis como un mapeo de otras modalidades cognitivas (como Matemática y proposicional) a una modalidad de imagen.”

Tipos de visualización de datos

“Existen multitud de técnicas y aproximaciones para la visualización según sea la naturaleza del dato de la información. Desde el punto de vista de los datos, especialmente datos estructurados (o semi-estructurados) y su explotación visual, podemos establecer grosso modo la siguiente clasificación de tipos de visualización según complejidad y elaboración de la información.” **(Red.es, 2016)**

En la actualidad las empresas buscan herramientas cloud para la visualización de datos, con la finalidad de poder obtener representaciones graficas de sus datos generados a lo largo del tiempo; estas herramientas o servicios ofrecen infinidad de técnicas para la visualización de datos, de los cuales le mostraremos los siguientes:

- **Elementos básicos de representación de datos**

  Un elemento básico de visualización de datos puede ser un gráfico, un mapa, un KPI, tablas de datos, un grafo, etc. Estos son algunos tipos de visualizaciones básicas:

  - ✓ Gráficas: barras, líneas, columnas, puntos, tree maps, etc.
  - ✓ Mapas: burbujas, coropletas, mapa de calor, etc.
  - ✓ Tablas: anidaciones, dinámicas, drill-down, transiciones, etc.

- **Cuadro de mandos**

  “Un cuadro de mando es una composición compleja de visualizaciones individuales que guardan una coherencia y una relación temática entre ellas. Son ampliamente utilizados en las organizaciones para análisis de conjuntos de variables y toma de decisiones.” **(Red.es, 2016)**

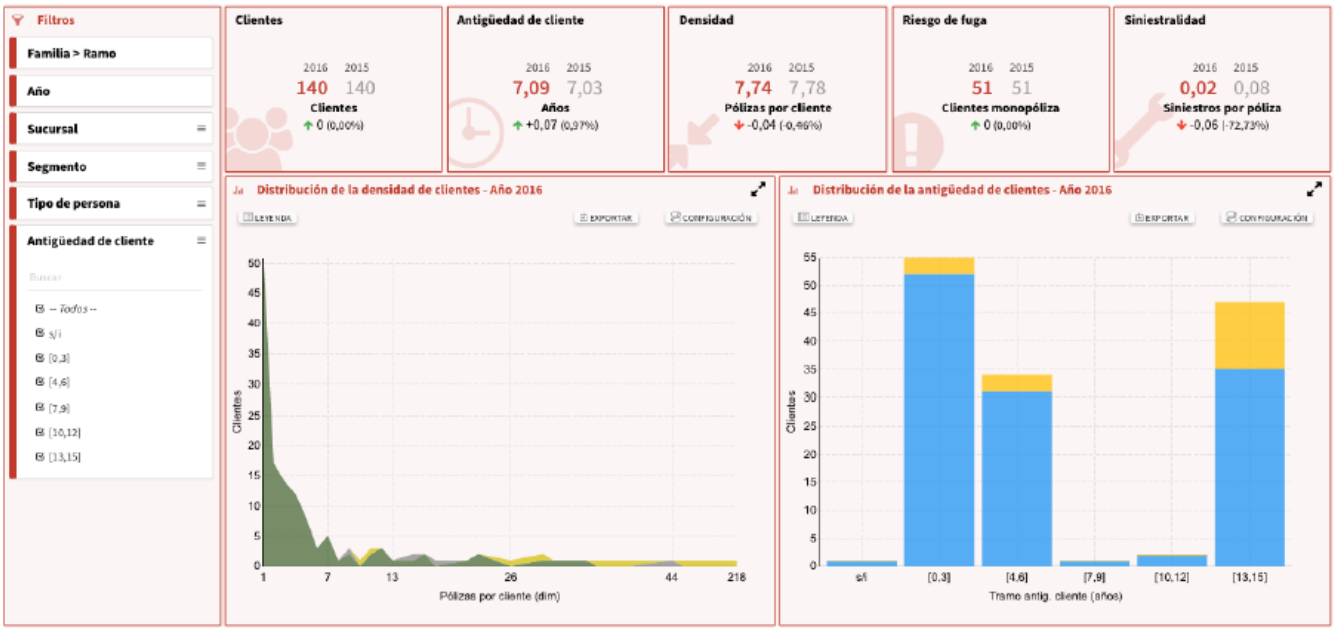


Figura 10: Ejemplo de cuadro de mando empresarial.

Fuente: Red.es (2016).

- **Infografías**

  “Una infografía también es una composición elaborada de visualizaciones que construye, a partir de diferentes elementos, un significado complejo para el usuario. Sin embargo, las infografías no están tan destinadas al análisis de variables sino a la construcción de narrativas a partir de los datos; es decir, las infografías se utilizan para contar “historias”.” **(Red.es, 2016)**

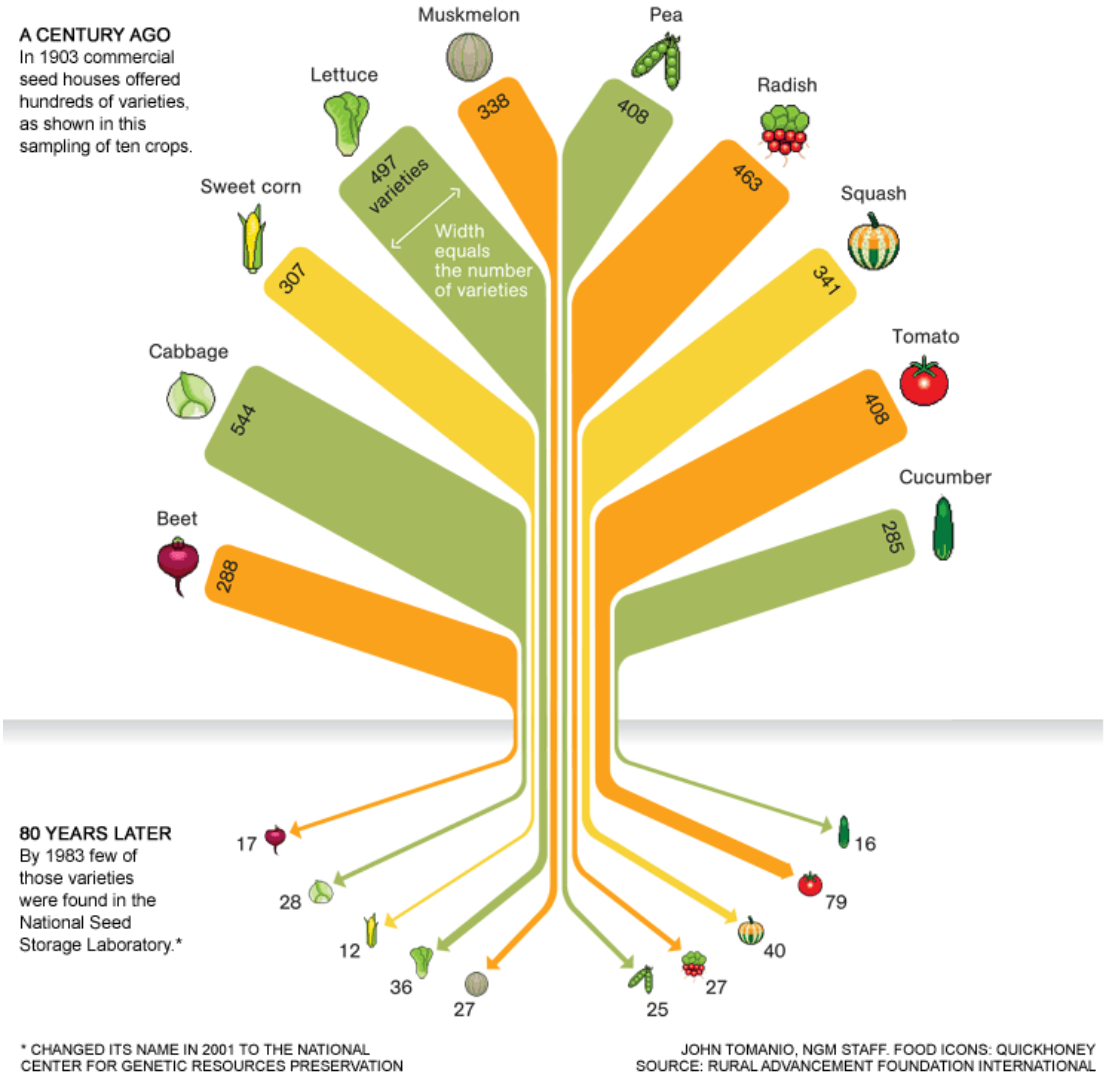


Figura 11: Cambios en la dieta alimenticia en la última centuria.

Fuente: Red.es (2016).

Procesamiento de datos

En estos últimos años se ha venido realizando importantes esfuerzos para desarrollar procedimientos de anotación automática y procesamiento de información, con el objetivo de reducir la intervención del ser humano.

¿Qué es el procesamiento de datos?

"Cada minuto se capturan y recopilan millones de datos acerca de búsquedas, intereses y tendencias de los usuarios en internet, pero, ¿qué representa este conjunto de estadísticas, números y símbolos para una empresa y qué puede hacer con toda esta información? Para dar valor a este conjunto de datos almacenados en la red, ha surgido el procesamiento de datos, cuya misión es procesar grandes cantidades de datos del internet y transformarlos en información valiosa para la toma de decisiones gerenciales". **(Barrera, 2018)**

Es una técnica que consiste en la recolección de datos para entrada, los cuales se evalúan y ordenan para obtener información útil y con valor, que luego será analizado por el usuario final para ayudar a la toma de decisiones o para alguna acción conveniente.

El procesamiento de datos tiene tres etapas que la componen, las cuales son:

- ✓ Entrada: se obtienen los datos y se llevan a un bloque central para ser procesados, estos datos se denominan entrada y se clasifican para hacer el proceso más fácil y rápido.
- ✓ Proceso: se ejecutan las operaciones necesarias para convertir los datos de entrada en información útil, una vez completada se ejecuta la operación de salida, donde se prepara un informe base para la toma de decisiones.
- ✓ Salida: administración del resultado de salida que se define como los procesos necesarios para que la información útil llegue al usuario final.

Análisis de datos

El Análisis de Datos (Data Analysis, o DA) es la ciencia que examina datos con el propósito de sacar conclusiones sobre la información. El análisis de datos es usado en varias industrias para permitir que las compañías y las organizaciones

tomen mejores decisiones empresariales y también es usado en las ciencias para verificar o reprobar modelos o teorías existentes. El análisis de datos se distingue de la extracción de datos por su alcance, su propósito y su enfoque sobre el análisis. Los extractores de datos clasifican inmensos conjuntos de datos usando software sofisticado para identificar patrones no descubiertos y establecer relaciones escondidas. El análisis de datos se centra en la inferencia, el proceso de derivar una conclusión basándose solamente en lo que conoce el investigador. **(Rouse, 2012)**

El término “análisis” ha sido usado por varios proveedores de software de inteligencia de negocios como una palabra de moda que describe varias funciones. Los bancos y las compañías de tarjetas de crédito, por ejemplo, analizan los retiros y los patrones de gasto para prevenir el fraude o robo de identidad. Las compañías de comercio electrónico (Ecommerce) examinan el tráfico en el sitio web o los patrones de navegación para determinar qué clientes son más o menos propensos a comprar un cierto producto o servicio, basándose en compras previas o patrones de visualización. El análisis de datos moderno normalmente usa tableros de información que se basan en flujos de datos en tiempo real. El llamado análisis en tiempo real implica análisis e informes dinámicos basados en los datos que introducidos en un sistema un minuto antes del tiempo actual de uso. **(Rouse, 2012)**

Definición de términos

- **AWS:** Amazon Web Services, plataforma virtual donde se ofrecen servicios de cloud computing. **(Amazon Inc., 2018)**
- **TIC:** Tecnologías de información y comunicación, herramientas para procesar, administrar y compartir información mediante recursos tecnológicos. **(Espacioteca, 2017)**
- **ETIP:** Plataformas europeas de tecnología e innovación, organismo encargado de guiar investigaciones y desarrollo para redes inteligentes. **(INTENSYS4EU, 2020)**
- **AMR:** Automatic Meter Reading, un término aplicado a la variedad de tecnologías que permiten a las empresas de servicios públicos leer consumos de gas, agua y electricidad con mayor eficiencia. **(Afinidad Eléctrica, 2008)**

- **AMI:** Advanced Meter Infraestructure, dispositivo que comprende funciones de medición, recolección y análisis de datos de consumo de energía. **(Afinidad Eléctrica, 2008)**
- **FDD:** Frequency Division Duplex, significa que el radio transmite y recibe datos al mismo tiempo, logrando así alto rendimiento y latencia muy baja. **(Bertenbreiter, 2015)**
- **TDD:** Time Division Duplex, que significa que el radio puede transmitir o recibir datos, pero no realizar ambas acciones al mismo tiempo. **(Bertenbreiter, 2015)**
- **3GPP:** 3rd Generation Partnership Project, es un acuerdo de colaboración en tecnología de telefonía móvil que fue establecido en diciembre de 1998. **(EcuRed, 2019)**
- **Streaming:** Hace referencia al hecho de escuchar música o ver vídeos sin necesidad de descargarlos completos antes de que los escuches o veas. **(DeStreaming, 2019)**
- **LTE:** Long Term Evolution, estándar para la comunicación inalámbrica de alta velocidad. **(SoftwareLab Blog, 2018)**
- **Open – Source:** Código abierto, es un modelo de desarrollo de software basado en la colaboración abierta. **(Levine & Prietula, 2014)**
- **Business Intelligence:** Inteligencia de negocios, conjunto de estrategias, aplicaciones, datos, productos, tecnologías y arquitectura técnicas, enfocados a la administración y creación de conocimiento sobre el medio, a través del análisis de los datos existentes en una organización o empresa. **(Signaturit, 2017)**

III. Materiales y Métodos

3.1 Materiales e insumos

Para el análisis, procesamiento y visualización de datos se necesitaron:

Tabla 2. Materiales y Equipos

Materiales y Equipos	
Materiales	• 2 Paquetes de hojas bond • 1 Kit de útiles de escritorio • 1 Kit de muebles de oficina
Hardware	• 1 Laptop Lenovo V330-15IKB Core i7, 8Gb RAM, 512Gb SSD • 1 Microsoft Kit Desktop 1576 1113 • 1 Monitor LG 27UL500 27'' • 1 Memoria USB Kingston 64Gb • 1 Impresora Multifuncional EPSON L575 • 1 Router Movistar internet
Software	• Amazon Quicksight Edición Standard • Amazon S3 Edición Standard • Microsoft Excel 2016 • IBM SPSS 25 • Navegador Web Google Chrome • Bloc de Notas

Fuente: Elaboración propia (2020).

Tabla 3. *Insumos.*

Insumos	
Datos	• Scalar • Vector

Fuente: Elaboración propia (2020).

3.2 Metodología

Diseño de investigación

Método de investigación: Hipotético – Deductivo

Para **(Valderrama, 2015)**, “el método hipotético-deductivo se inicia con la observación de casos particulares para poder plantear un problema, el cual puede remitir a una teoría a través de un proceso de inducción. Partiendo del marco teórico se formula una hipótesis mediante un razonamiento deductivo que; luego, esta se intenta validar empíricamente.”

En la presente investigación se utilizará el método hipotético-deductivo porque a partir de las observaciones de la problemática que existe en el proceso de visualización de datos se podrá analizar las teorías relacionadas al tema y corroborar las hipótesis.

Tipo de investigación: Explicativa

(Behar, 2008), dice que el tipo de estudio explicativo busca encontrar las causas que ocasionan ciertos fenómenos, busca explicar por qué ocurre un fenómeno orientado a la comprobación de hipótesis y análisis causales.

Esta investigación es de tipo explicativa porque se pretende analizar la influencia de la herramienta cloud Quicksight en las dimensiones de calidad, interpretación y análisis de datos dentro del proceso de visualización de datos obtenidos del proyecto Smart Cities: Cañete plan piloto con medidores eléctricos inteligentes.

Tipo de investigación: Experimental

Para **(Behar, 2008)**, el tipo de estudio es experimental porque se desea comprobar los efectos de una intervención especifica.

Esta investigación es de tipo experimental porque se pretende saber los resultados de haber aplicado la herramienta cloud Quicksight en el proceso de visualización de datos obtenidos del proyecto Smart Cities: Cañete plan piloto con medidores eléctricos inteligentes.

Tipo de investigación: Aplicada

Según **(Valderrama, 2015)**, el tipo de estudio es aplicada porque se aplica inmediatamente sobre una realidad concreta, este tipo de investigación la realizan los egresados de pre y post grado para conocer la realidad social,

económica, política y cultural de su ámbito, para plantear soluciones concretas, reales, factibles y necesarias a los problemas reconocidos.

Esta investigación es de tipo aplicada debido a que la herramienta cloud Quicksight será aplicada al proceso de visualización de datos obtenidos del proyecto Smart Cities: Cañete plan piloto con medidores eléctricos inteligentes.

Diseño de investigación: Pre - Experimental

Según **(Valderrama, 2015)**, el diseño pre-experimental de un grupo consta de 3 etapas:

1. Administrar un pre-test para medir el proceso de visualización de datos (variable dependiente – proceso que realizan de forma manual).
2. Aplicar el tratamiento experimental, Herramienta cloud Quicksight a los sujetos.
3. Administrar un post-test que mida nuevamente el proceso de visualización de datos.

En la investigación se aplicará el diseño de investigación pre-experimental porque se pretende analizar el proceso de visualización de datos antes y después de aplicar la herramienta cloud Quicksight.

$$\mathbf{G \rightarrow O_1 - X - O_2}$$

Donde:

G: grupo experimental con 3 etapas

O_1: pre-test

X: tratamiento experimental

O_2: post-test

Variables y operacionalización

Variable independiente

- Herramienta cloud Quicksight.

Variable dependiente

- Visualización de datos.

Operacionalización de variables

Variable	Definición Conceptual	Definición Operacional	Dimensión	Indicadores
Herramienta cloud Quicksight.	Amazon QuickSight es un servicio de análisis empresarial que puede utilizar para crear visualizaciones, realizar análisis ad hoc y obtener información útil empresarial a partir de sus datos. Puede detectar automáticamente los orígenes de datos de AWS y trabajar también con sus orígenes de datos. Amazon QuickSight permite a las organizaciones escalar de cientos a miles de usuarios y ofrece un desempeño fiable gracias a su sólido motor en memoria (SPICE).” **(Amazon Inc., 2018)**.	La herramienta Quicksight es ofrecida como servicio cloud por la empresa Amazon, que permite analizar datos mediante la visualización para poder representarlos gráficamente logrando un mejor entendimiento de los usuarios y una ayuda en la toma de decisiones.	No aplicable.	No aplicable.
			No aplicable.	

Visualización de datos	“La visualización de datos es un ámbito de conocimiento en constante evolución que últimamente se ha visto impulsado debido a la gran cantidad de datos disponibles, los cuales esperan a ser analizados, interpretados y contextualizados mediante una narrativa que combina texto, imágenes y otros recursos interactivos, más allá de presentaciones estáticas que usan ciertos elementos gráficos como soporte.” **(Julià Minguillón, 2016)**.	La visualización de datos es una técnica que consiste en analizar los datos de entrada para poder representarlos inteligentemente de forma gráfica para lograr el entendimiento de los usuarios finales y como ayuda en las tomas de decisiones de la organización.	Calidad.	Porcentaje de datos de calidad obtenidos
			Interpretación.	Porcentaje de lectura y visualización de datos.
			Análisis.	Porcentaje de datos útiles.

Tabla 4. Operacionalización de variables.

Población

(Carrasco, 2006) indica que, “Es un grupo de objetos que lo forma un determinado espacio dentro del cual se encuentra el problema de la investigación, este posee ciertas características que son más concretas y las comparan con el universo.”

La población de estudio será constituida por el total de simulaciones realizadas en el proyecto de investigación Smart Cities Simulation: Cañete plan piloto con medidores eléctricos inteligentes, que serían un total de 25 simulaciones.

Muestra

Para esta investigación se utilizará la muestra no probabilística o dirigida. De acuerdo con **(Hernández Sampieri et al., 2014)** menciona que es un “subgrupo de la población en la que la elección de los elementos no depende de la probabilidad sino de las características de la investigación”. **(S. Pickers, 2018)** también nos dice: “No sirven para hacer generalizaciones pero sí para estudios exploratorios. En este tipo de muestras, se eligen a los individuos utilizando diferentes criterios relacionadas con las características de la investigación, no tienen la misma probabilidad de ser seleccionados ya que el investigador suele determinar la población objetivo.” La población de este estudio es totalmente homogénea por esta razón es suficiente con 10 simulaciones realizadas.

Técnica e instrumentos de recolección de datos

El autor **(Behar, 2008)** nos menciona que “las técnicas apoyan a verificar la situación problemática que se planteó. Según el tipo de investigación que se desarrolle es que se utilizan unas u otras técnicas, herramientas e instrumentos.”

Fichaje

(Navas, 2014) indica que, “Es un modo de recolectar y almacenar información. Cada ficha contiene una serie de datos extensión variable pero todos referidos a un mismo tema, lo cual le confiere unidad y valor propio.”

Esta técnica permitió recolectar datos de los indicadores sobre la influencia de Quicksight en el proceso de visualización de datos.

Ficha de registro

Según **(Hernández Sampieri et al., 2014)** “Este instrumento sirve para recolectar la información, estos contendrán los valores iniciales que resumirán el proyecto,

dicha información será requerida para el procesamiento de datos, además de ello debe establecerse qué información será la utilizada y cuál es la que se va a divulgar".

Se elaboró una ficha de registro donde se detallarán por cada simulación programada, el porcentaje de los indicadores sobre la influencia de Quicksight en proceso de visualización de datos.

Tabla 5. Técnicas e instrumentos de recolección de datos.

Variable dependiente	Dimensión	Indicador	Técnica	Instrumento
Visualización de datos	Calidad de datos.	Porcentaje de datos de calidad.	Fichaje	Ficha de registro
	Interpretación de datos.	Porcentaje de lectura y visualización de datos	Fichaje	Ficha de registro
	Análisis de datos.	Porcentaje de datos útiles.	Fichaje	Ficha de registro

Validez y confiabilidad

Validez

Según **(Hernández Sampieri et al., 2014)**, "Podemos decir que la validez muestra el grado en el que un instrumento puede medir la variable con la que se está trabajando."

Validez de criterio

(Hernández Sampieri et al., 2014), definen que: "Se contrastan las puntuaciones que dan como resultado luego de aplicar el instrumento con aquella puntuación es que se obtienen de otros criterios que son del tipo externo."

Validez de contenido

(Hernández Sampieri et al., 2014), nos dicen: “Éste refiere al grado que se le puede dar a nuestro instrumento en algún tipo de dominio de contenido dentro de lo que se está midiendo. Éste representa al concepto o a la variable.”

Validez de constructo

(Hernández Sampieri et al., 2014), definen: “Éste busca explicar cómo se vinculan de una manera en la que las mediciones estén correlacionadas de forma teórica.”

Tabla 6. Validación del indicador - Porcentaje de datos de calidad obtenidos.

Expertos	Total
Mg. Pacheco Pumaleque, Alex	85 %
Mg. Cárdenas Lino, César	81%
Mg. Vilca Pizarro, Joel	90%
Promedio	85%

Tabla 7. Validación del indicador - Porcentaje de lectura y visualización de datos.

Expertos	Total
Mg. Pacheco Pumaleque, Alex	85 %
Mg. Cárdenas Lino, César	81%
Mg. Vilca Pizarro, Joel	90%
Promedio	85%

Tabla 8. Validación del indicador - Porcentaje de datos útiles.

Expertos	Total
Mg. Pacheco Pumaleque, Alex	85 %
Mg. Cárdenas Lino, César	81%
Mg. Vilca Pizarro, Joel	90%
Promedio	85%

La validez de nuestro instrumento fue corroborada por los expertos en la materia (juicio de expertos) de cloud computing y visualización de datos. Como se puede observar en las tablas anteriores, el promedio de validez de cada indicador es cercano a 1, lo que significa que el instrumento utilizado es válido.

Confiabilidad

(Hernández Sampieri et al., 2014): "Para medir la confiabilidad de un instrumento es importante señalar el grado en el que la aplicación es repetida a un mismo objeto, verificando si este genera resultados similares o iguales que puedan ser consistentes."

Método: Test-Retest

(Hernández Sampieri et al., 2014), Nos dicen que es uno de los procedimientos que se utilizan con más frecuencia para determinar si el coeficiente tiene la estabilidad necesaria y puede ser confiable, este tipo de procedimiento se realiza con un mismo instrumento, y se aplica las veces que sean necesarias a un mismo grupo."

Técnica: Coeficiente de correlación de Pearson

(Riquelme, 2019), nos dice: "El Coeficiente de Correlación de Pearson es una medida de la correspondencia o relación lineal entre dos variables cuantitativas aleatorias. En palabras más simples se puede definir como un índice utilizado para medir el grado de relación que tienen dos variables, ambas cuantitativas.

Teniendo dos variables, la correlación facilita que se hagan estimaciones del valor de una de ellas, con conocimiento del valor de la otra variable.

Este coeficiente es una medida que indica la situación relativa de los sucesos respecto a las dos variables, es decir, representa la expresión numérica que indica el grado de correspondencia o relación que existe entre las 2 variables. Estos números varían entre límites de +1 y -1.”

Tabla 9. Análisis de confiabilidad.

Coeficiente	Relación
0,00 ≤ sig. < 0,20	Despreciable
0,20 ≤ sig. < 0,40	Baja o ligera
0,40 ≤ sig. < 0,60	Moderada
0,60 ≤ sig. < 0,80	Aceptable
0,80 ≤ sig. < 1,00	Muy Alta

Fuente: Valderrama, Santiago (2015)

De la tabla anterior podemos decir:

Si el valor es cercano a 1 entonces el instrumento será confiable y si el valor es menor a 0.6 entonces el instrumento no será poco confiable.

El cálculo del coeficiente de correlación de Pearson se realizó con ayuda del software SPSS 25.

Para el indicador “Porcentaje de calidad de datos” se obtuvo un coeficiente de 0.934 y según lo que nos indica la tabla N°10 la relación del coeficiente de correlación de Pearson es “Muy Alta”.

Tabla 10. Confiabilidad del indicador Porcentaje de calidad de datos.

Correlaciones

		Calidad de datos Test	Calidad de datos Retest
Calidad de datos Test	Correlación de Pearson	1	,934**
	Sig. (bilateral)		,000
	N	10	10
Calidad de datos Retest	Correlación de Pearson	,934**	1
	Sig. (bilateral)	,000	
	N	10	10

**. La correlación es significativa en el nivel 0,01 (bilateral).

Para el indicador “Porcentaje de lectura y visualización de datos” se obtuvo un coeficiente de 1.000 y según lo que nos indica la tabla N°11 la relación del coeficiente de correlación de Pearson es “Muy Alta”.

Tabla 11. Confiabilidad del indicador Porcentaje de lectura y visualización de datos.

Correlaciones

		Lectura y visualización de datos Test	Lectura y visualización de datos Retest
Lectura y visualización de datos Test	Correlación de Pearson	1	1,000**
	Sig. (bilateral)		,000
	N	10	10
Lectura y visualización de datos Retest	Correlación de Pearson	1,000**	1
	Sig. (bilateral)	,000	
	N	10	10

**. La correlación es significativa en el nivel 0,01 (bilateral).

Para el indicador “Porcentaje de datos útiles” se obtuvo un coeficiente de 1.000 y según lo que nos indica la tabla N°12 la relación del coeficiente de correlación de Pearson es “Muy Alta”.

Tabla 12. Confiabilidad del indicador Porcentaje de datos útiles.

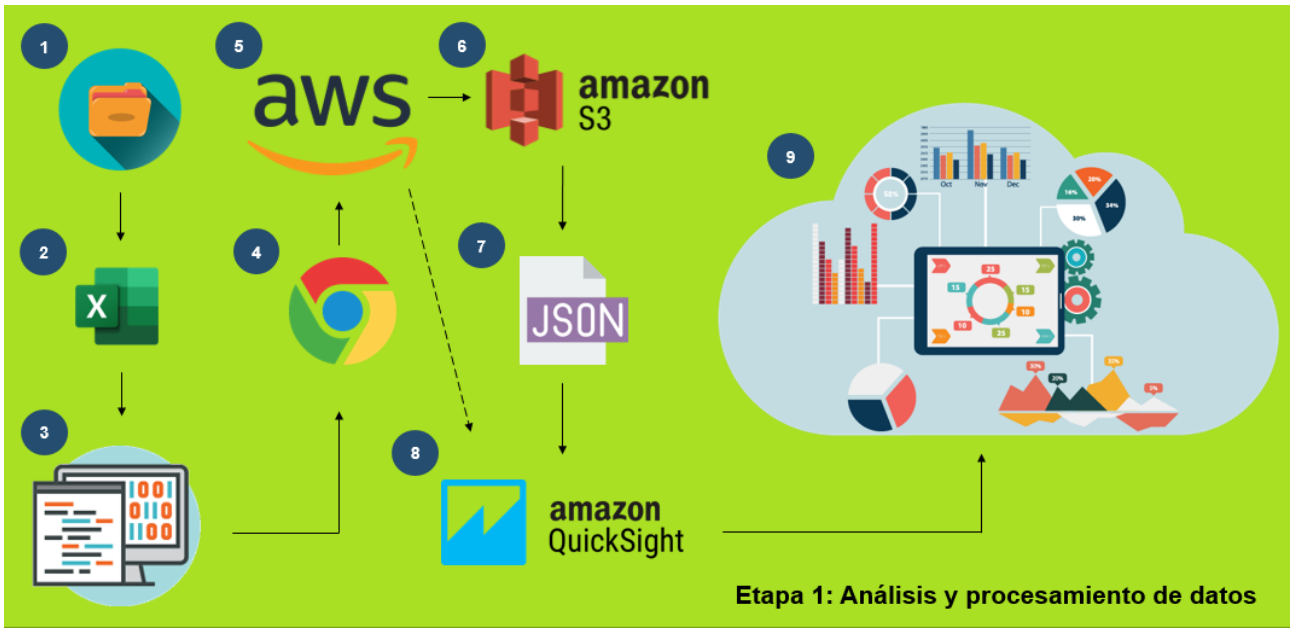
Correlaciones

		Datos útiles Test	Datos útiles Retest
Datos útiles Test	Correlación de Pearson	1	1,000**
	Sig. (bilateral)		,000
	N	10	10
Datos útiles Retest	Correlación de Pearson	1,000**	1
	Sig. (bilateral)	,000	
	N	10	10

**. La correlación es significativa en el nivel 0,01 (bilateral).

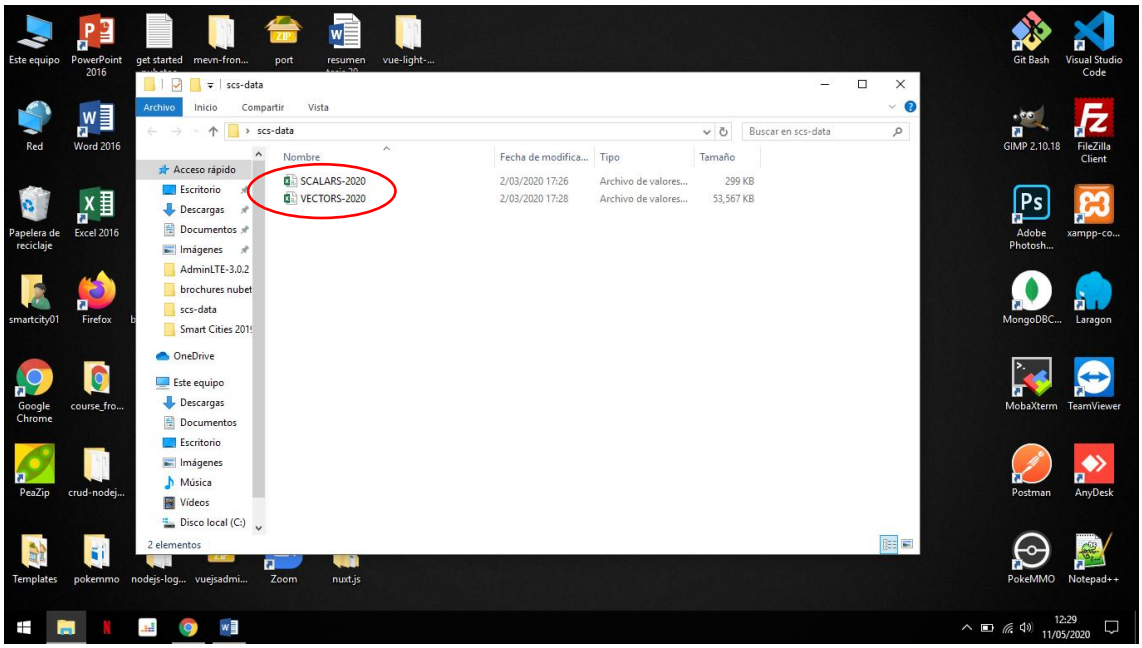
3.3 Metodología de desarrollo del proyecto

La investigación se desarrolló en las siguientes etapas:

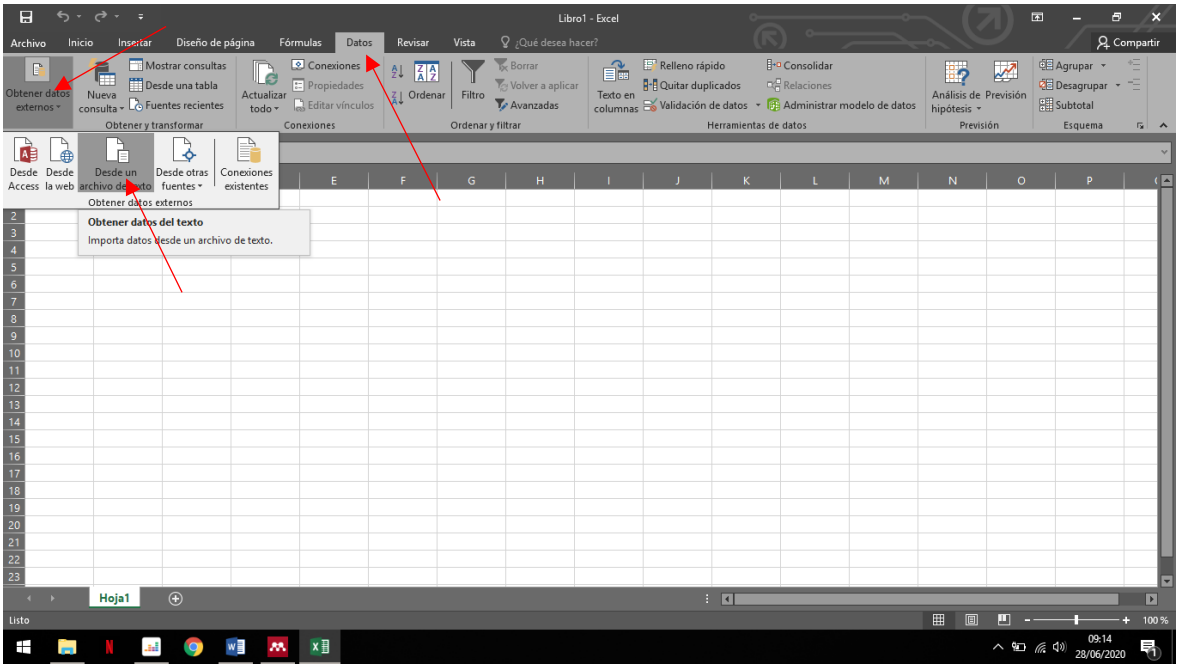


Etapa 1: Análisis y procesamiento de datos

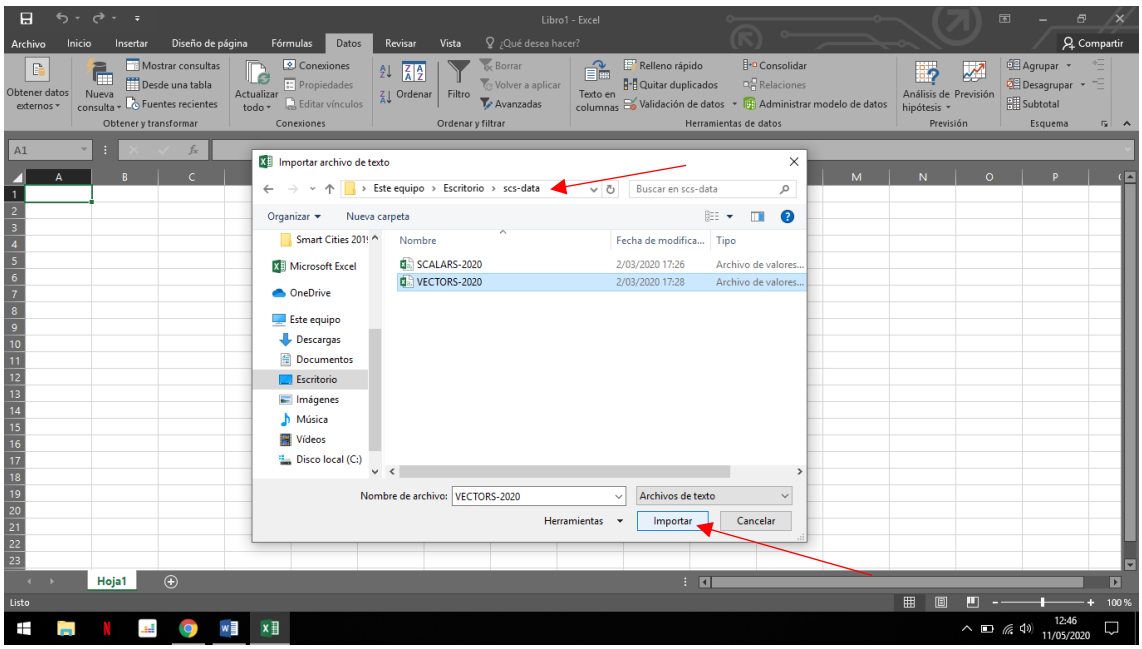
1) El primer paso será obtener los datos en formato CSV de la simulación del proyecto Smart Cities Simulation, estos datos estarán delimitados por comas, por lo que más adelante explicaremos su importación en tablas de Excel para tener más orden. Luego guardamos los datos en una carpeta local de la laptop.



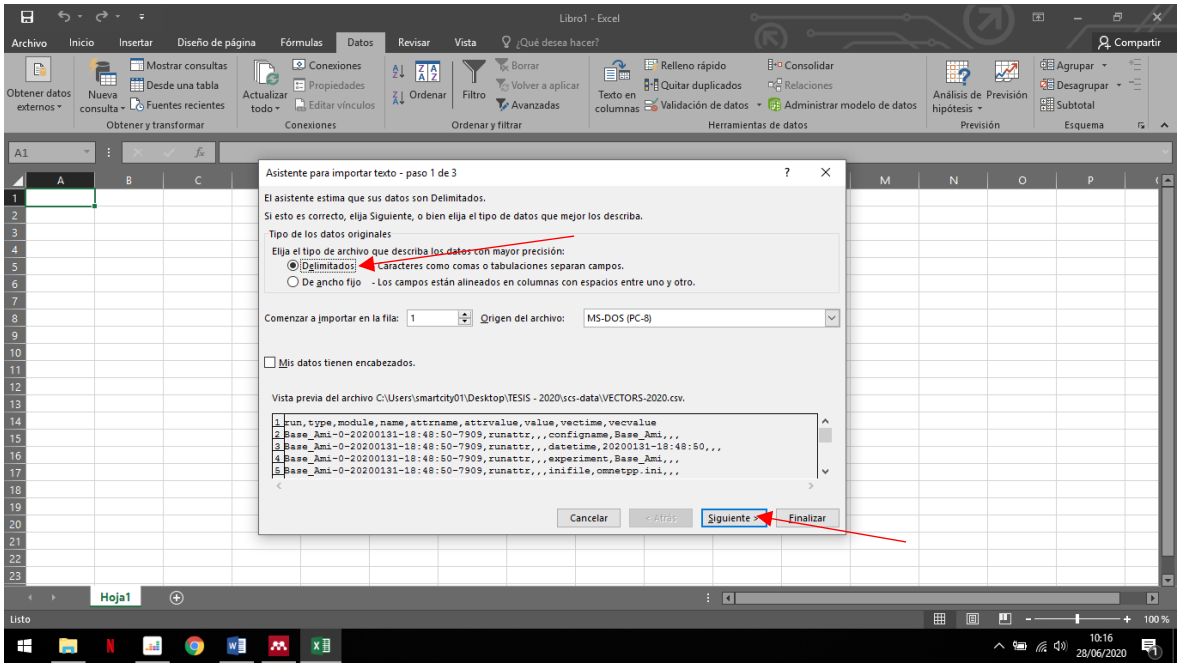
2) Para importar la data, primero damos clic en la sección “Datos”, luego clic en la opción obtener datos externos, y a continuación clic en “Desde un archivo de texto”.



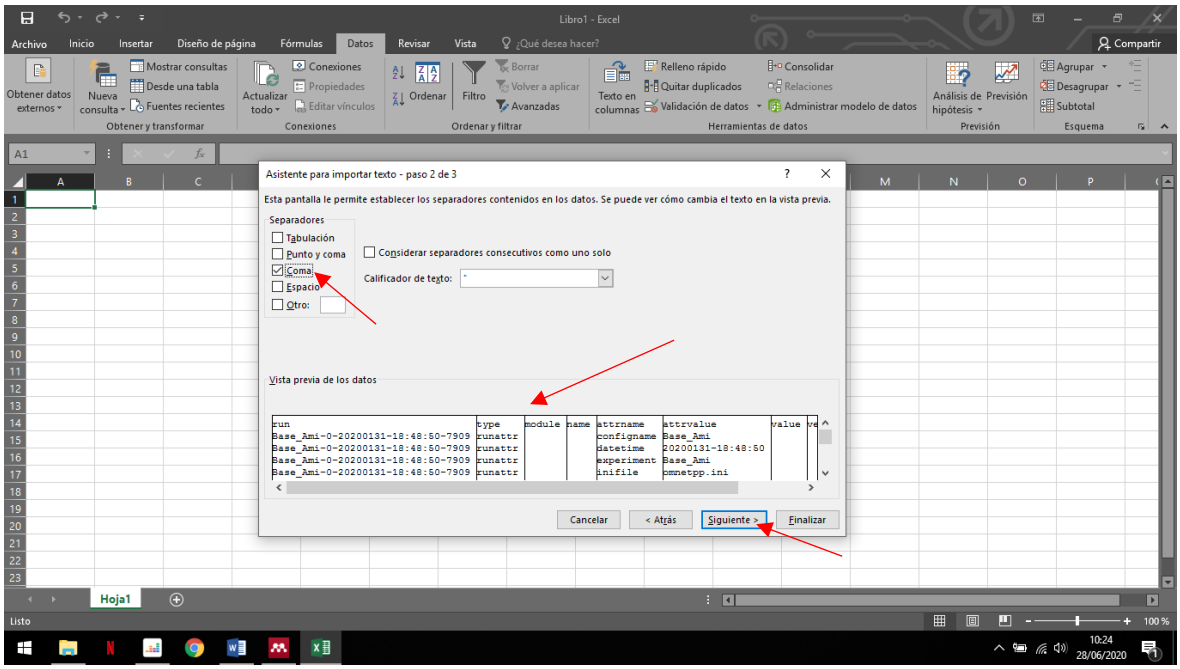
Aparece una ventana emergente donde tendremos que ubicar el directorio donde se encuentra la carpeta donde guardamos los datos, para luego dar clic en importar.



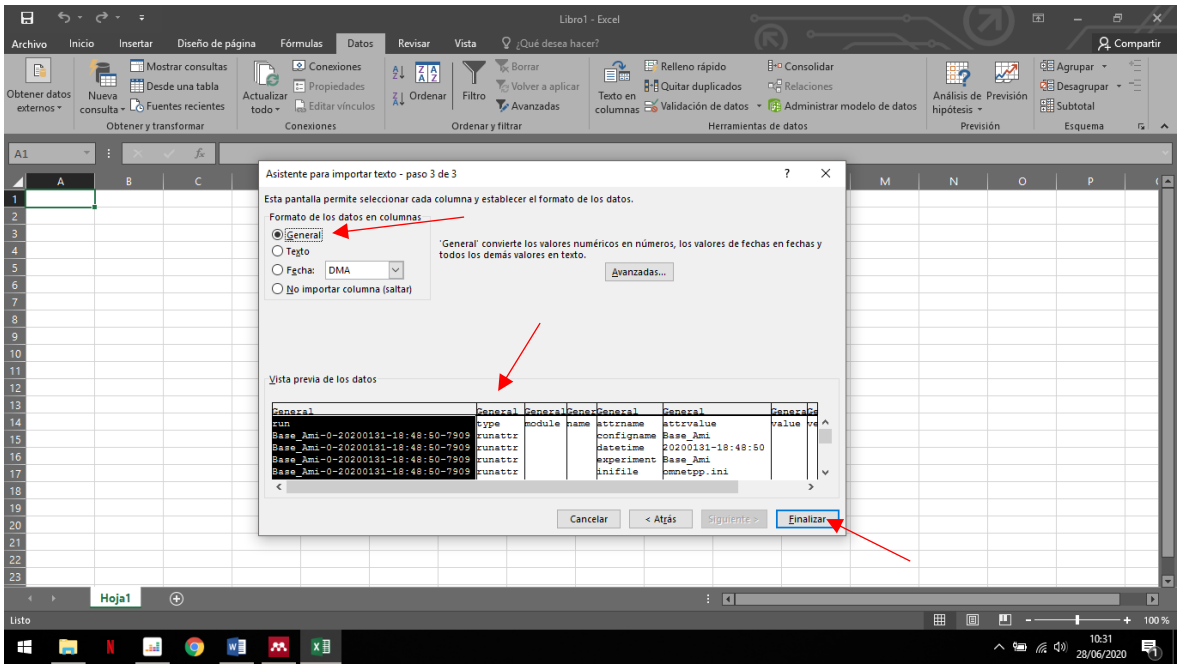
3) Configuramos la importación seleccionando los parámetros del formato csv, primero marcamos el tipo de datos originales, en este caso es “Delimitados”, luego damos en siguiente.



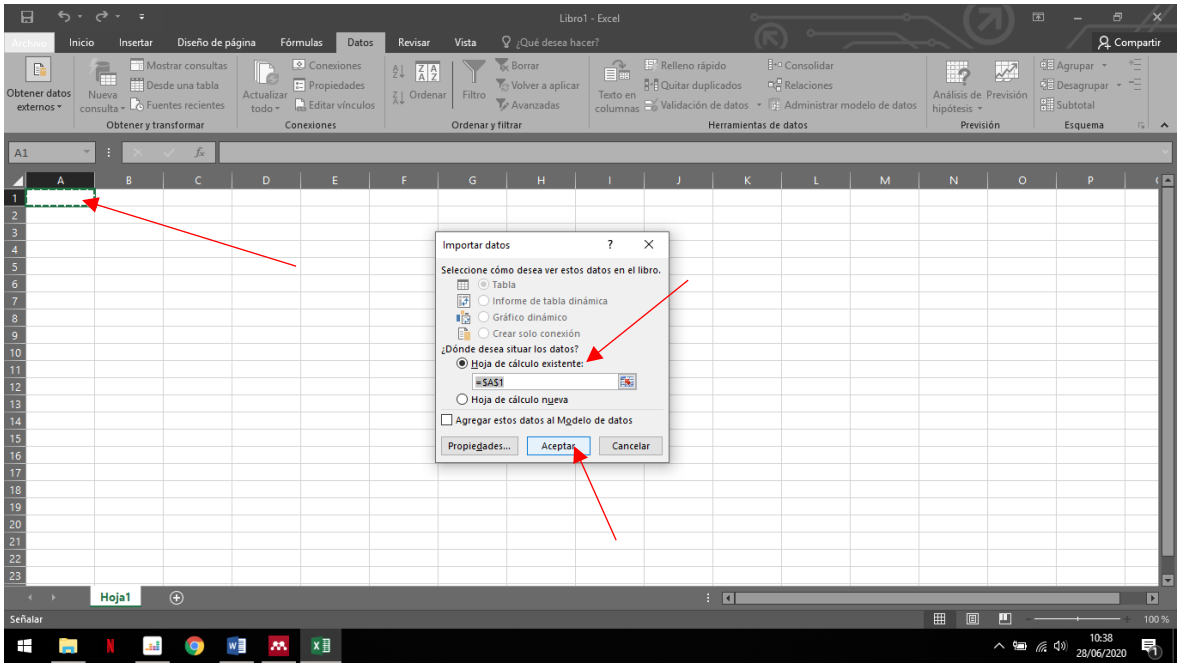
Luego en la sección separadores, marcamos la opción “Coma”, esto hace referencia a que estamos importando un archivo en formato CSV. En la parte inferior de la ventana observamos que se visualiza una vista previa de como querían los datos del archivo en tablas de Excel.



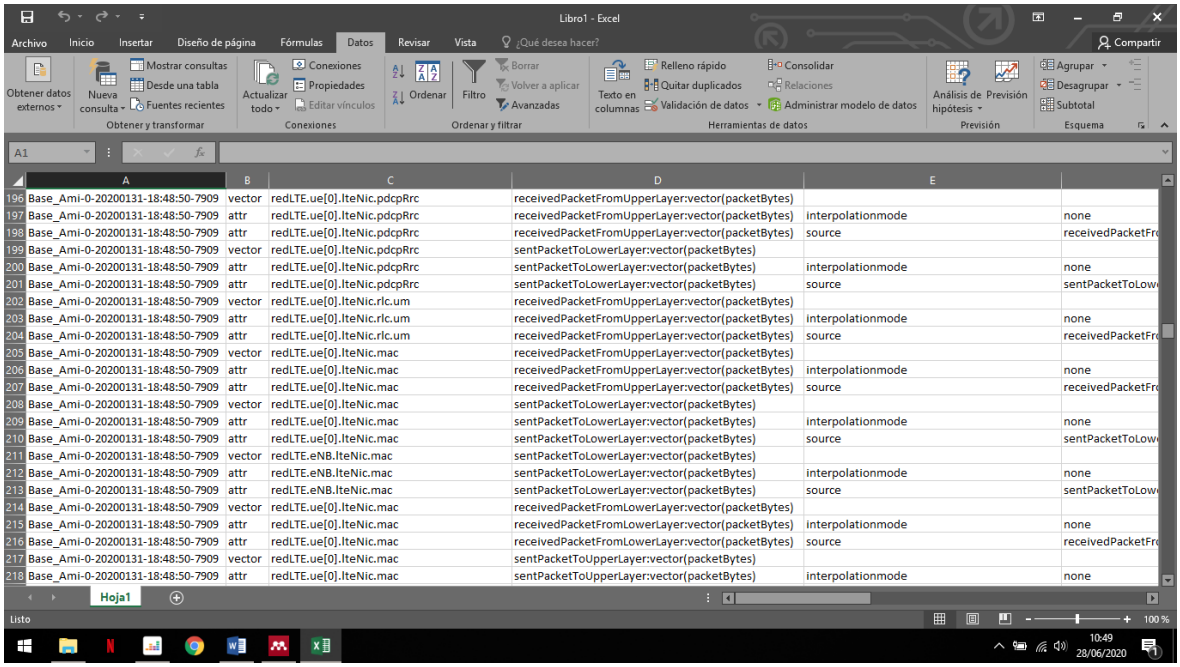
Configuramos la sección de formato de las columnas, marcamos “General” para que automáticamente convierta los valores numéricos en números, los valores de fechas en fechas y lo demás en texto, en la parte inferior podemos observar la asignación del formato a cada columna. Damos clic en finalizar.



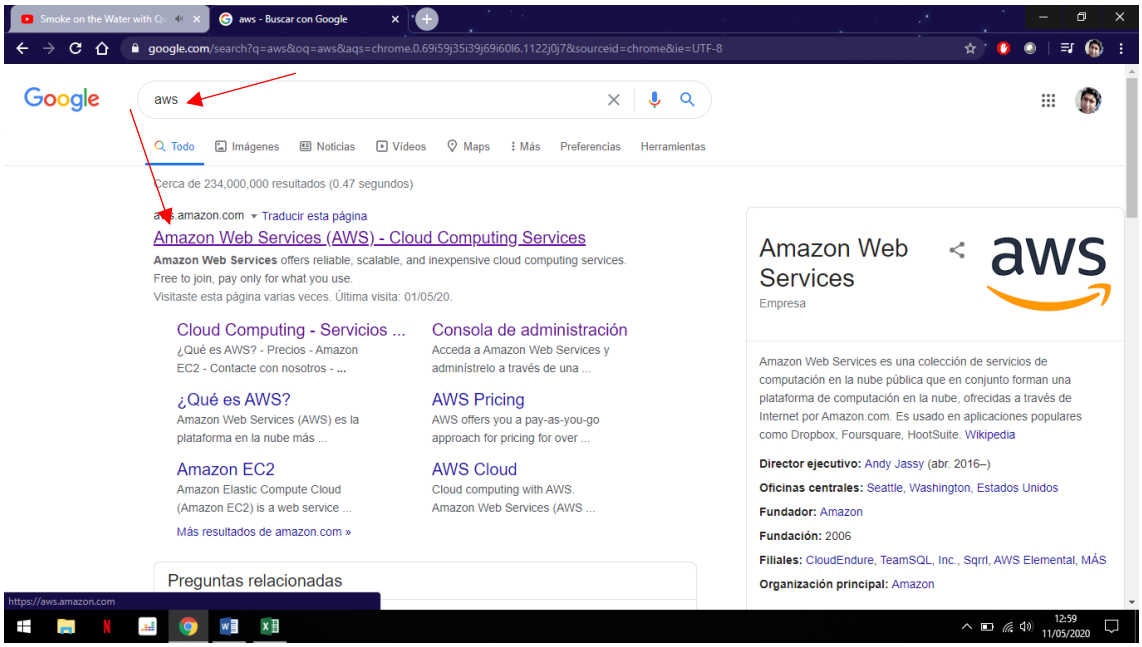
Por ultimo seleccionamos desde donde queremos que se sitúen los datos, en este caso desde la primera cuadricula de la hoja de cálculo y damos clic en aceptar.



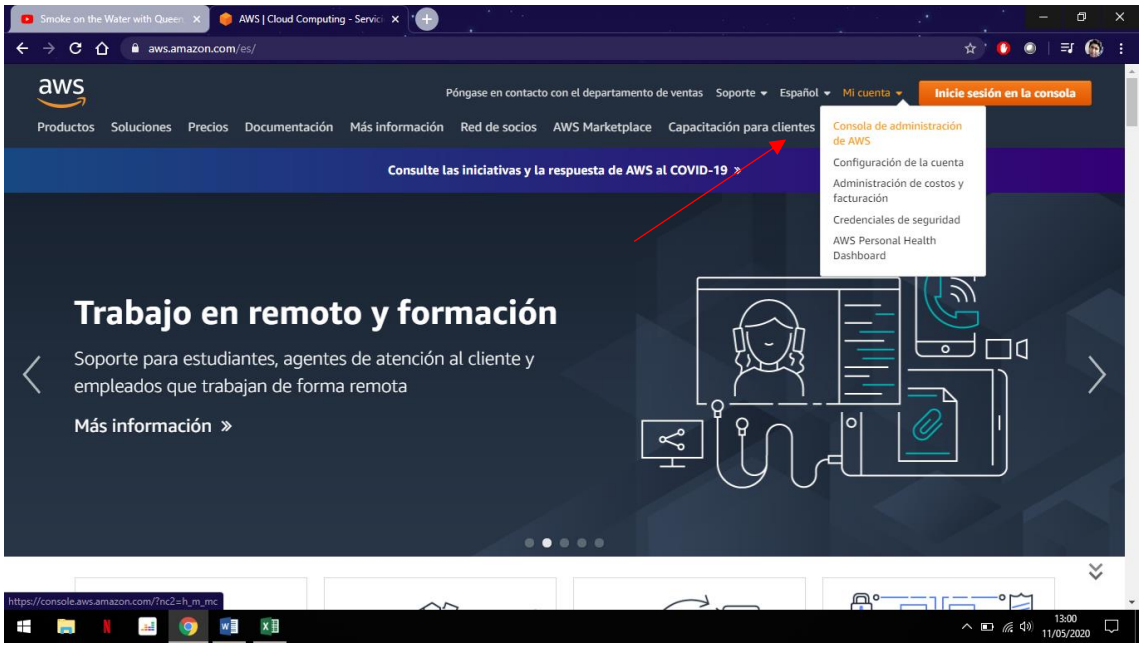
Observamos cómo de formato csv pasa a formato xlsx, así podemos manejar la data de forma ordenarla y eliminar algunos datos que no se utilizaran, una vez todo listo, lo exportamos con el formato csv y lo guardamos en un directorio.



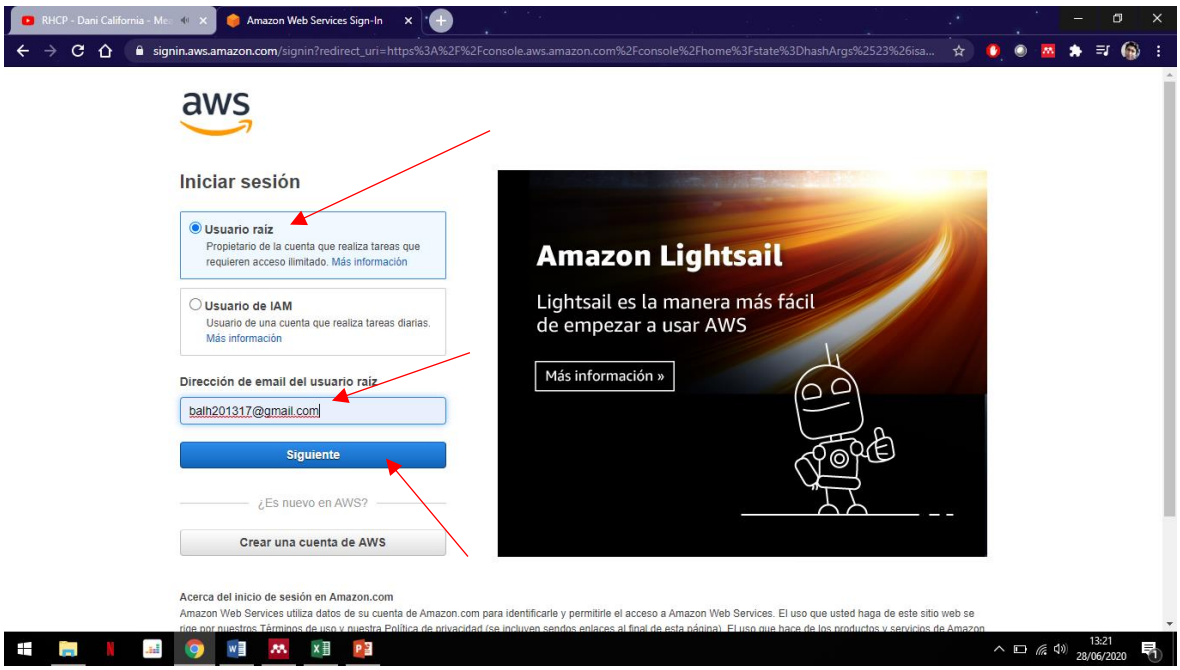
4) Abrimos el navegador Google Chrome para acceder a la página de Amazon Web Services. Escribimos en la barra de búsqueda “aws” y damos clic en el enlace que diga: “Amazon Web Services (AWS)”.



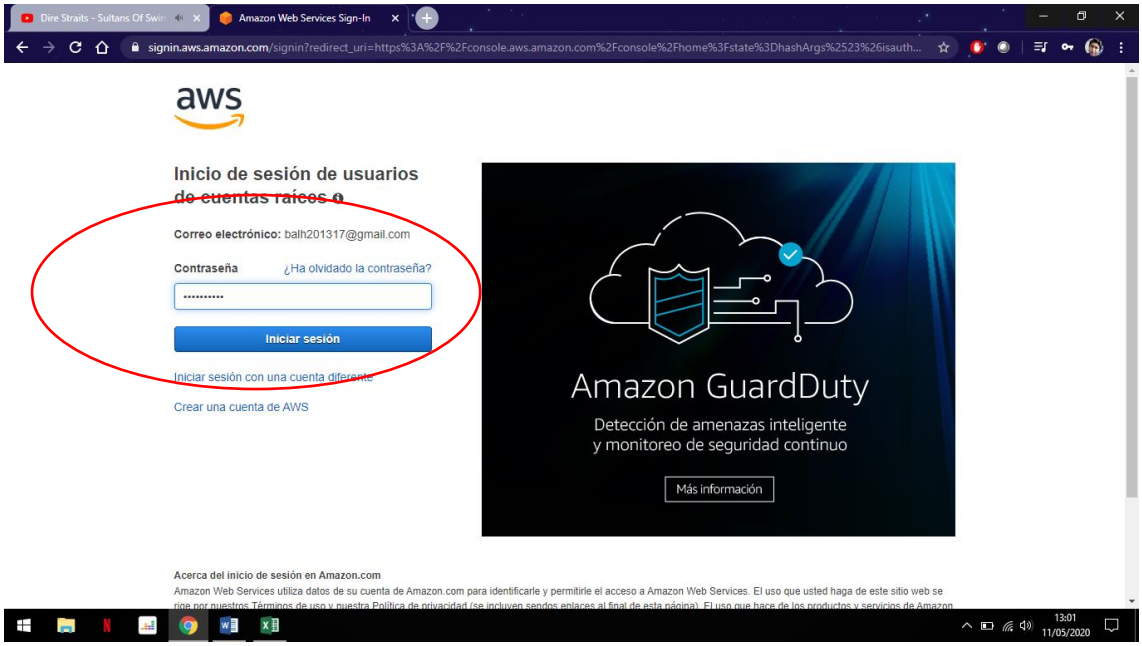
5) Una vez dentro de la página de AWS, vamos a la sección de “Mi cuenta”, desplegamos y hacemos clic en “Consola de administración de AWS”.



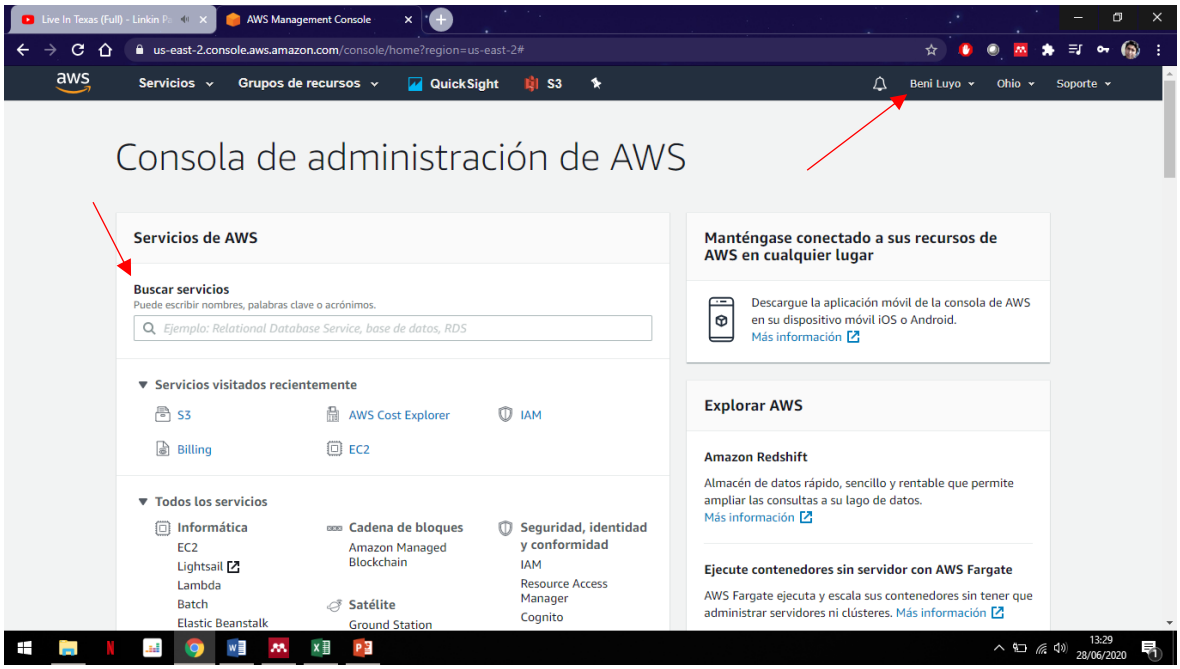
Luego seleccionamos que tipo de usuario somos, Usuario Raíz (Administrador) o Usuario de IAM (usuario no admin con rol especifico asignado). En esta investigación se utilizamos el Usuario Raíz, luego en el input escribimos el correo de la cuenta y damos en siguiente. Ojo que para este punto ya deberíamos tener una cuenta de AWS creada.



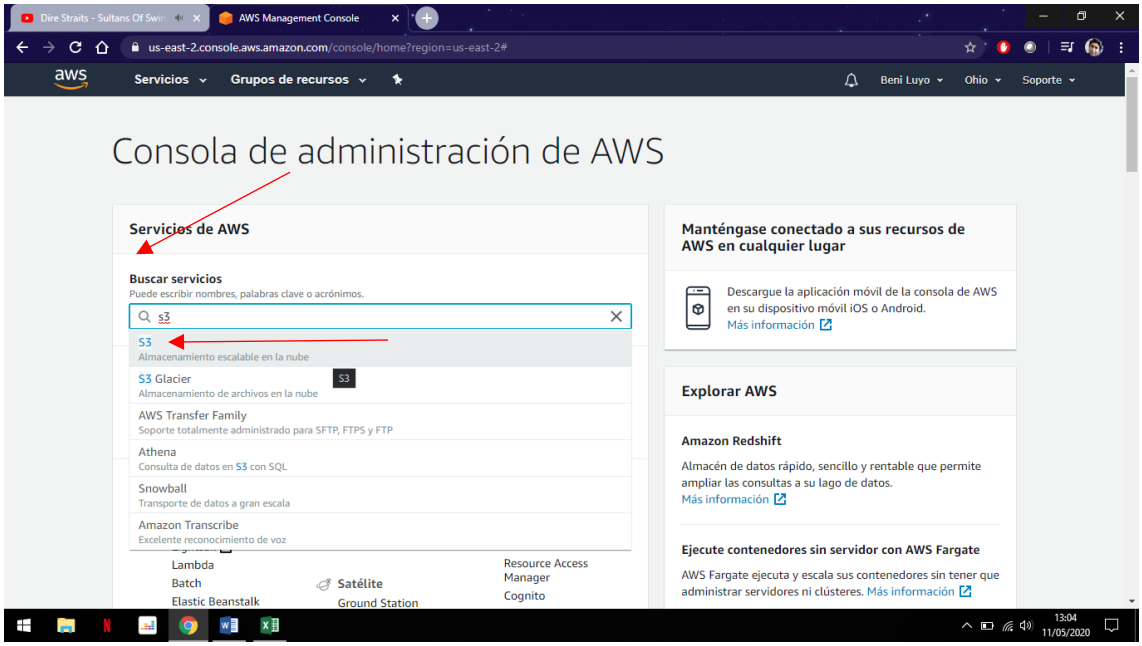
Luego se nos pedirá escribir la contraseña, una vez escrita damos clic en iniciar sesión.



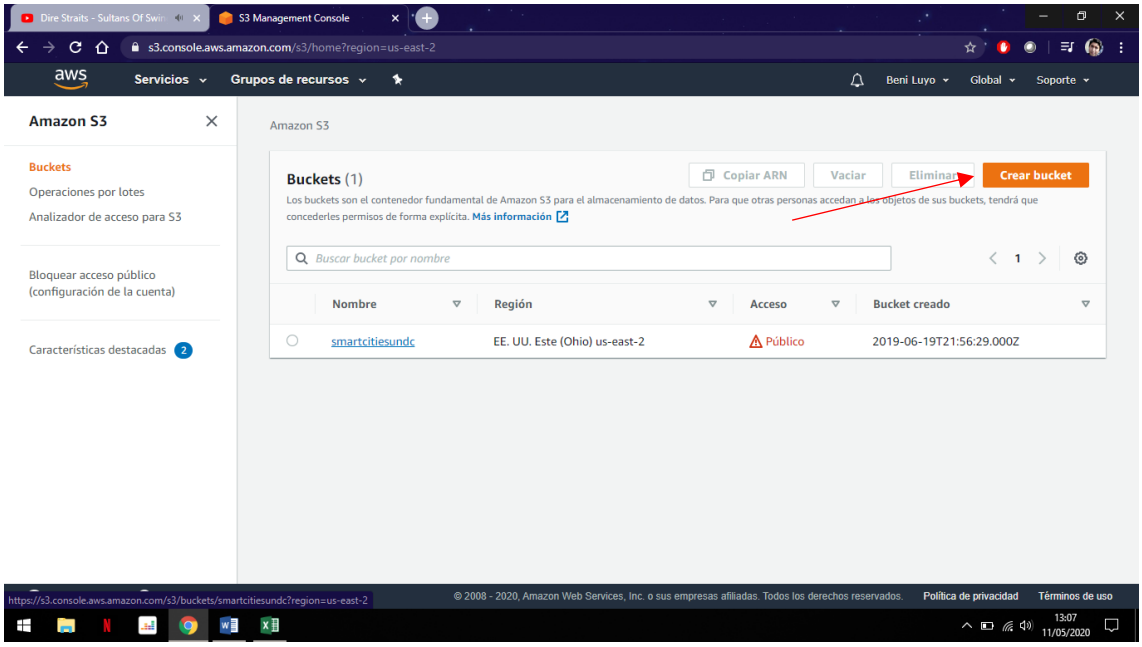
Esta es la consola de administración de servicios brindados por AWS, donde podemos observar una barra de búsqueda con los servicios cloud y en la parte superior algunas funciones, la región donde accedimos y el nombre del usuario que ha ingresado.



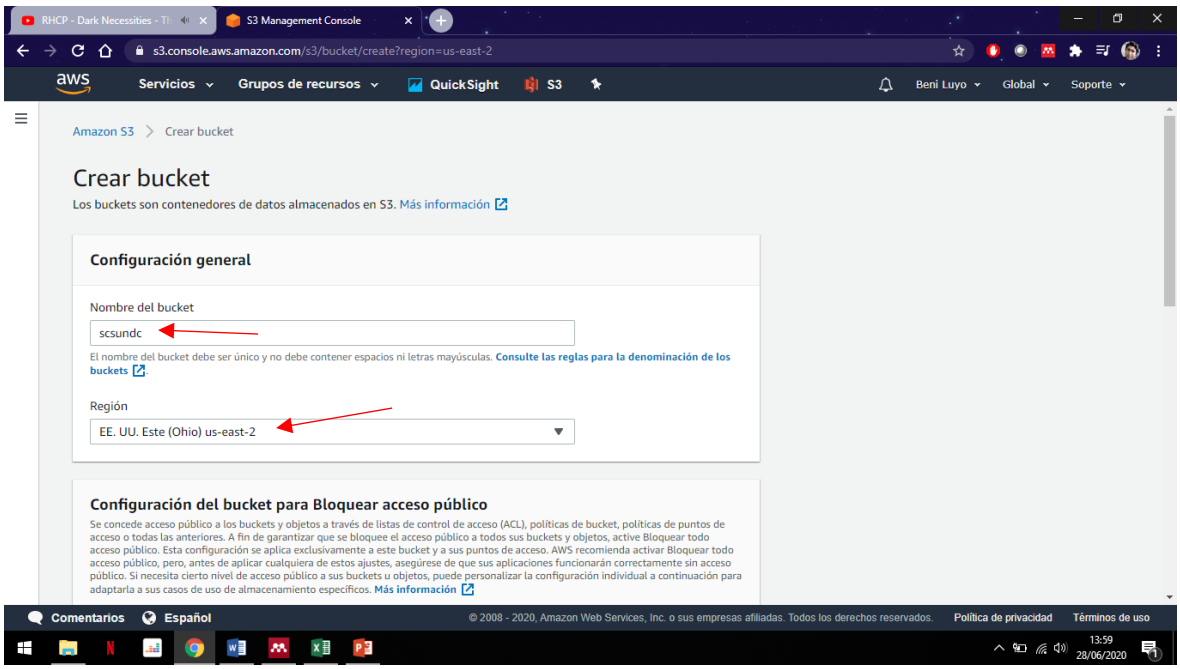
6) En la barra de búsqueda escribimos el servicio de almacenamiento en la nube “S3”, el cual nos permite almacenar archivos mediante buckets. Hacemos clic en S3 y esperamos.



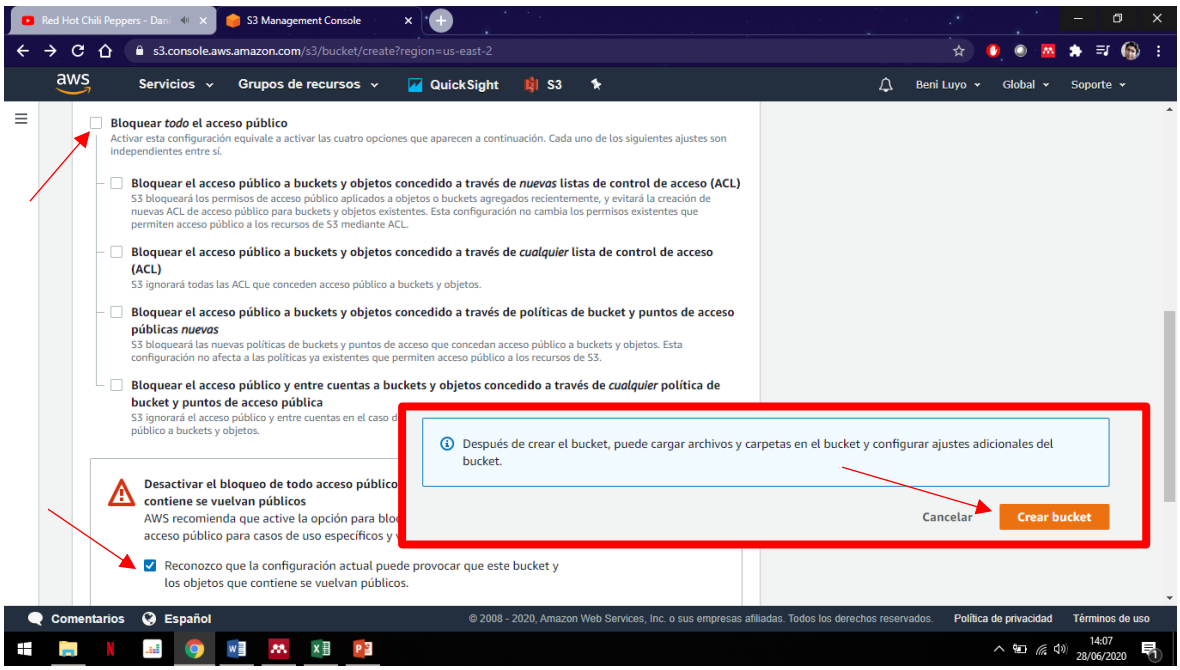
Ahora si no tenemos un bucket crearemos uno, un bucket es un contenedor de objetos donde podemos guardar archivos y cualquier metadato opcional que describa al archivo. hacemos clic en el botón “Crear bucket”.



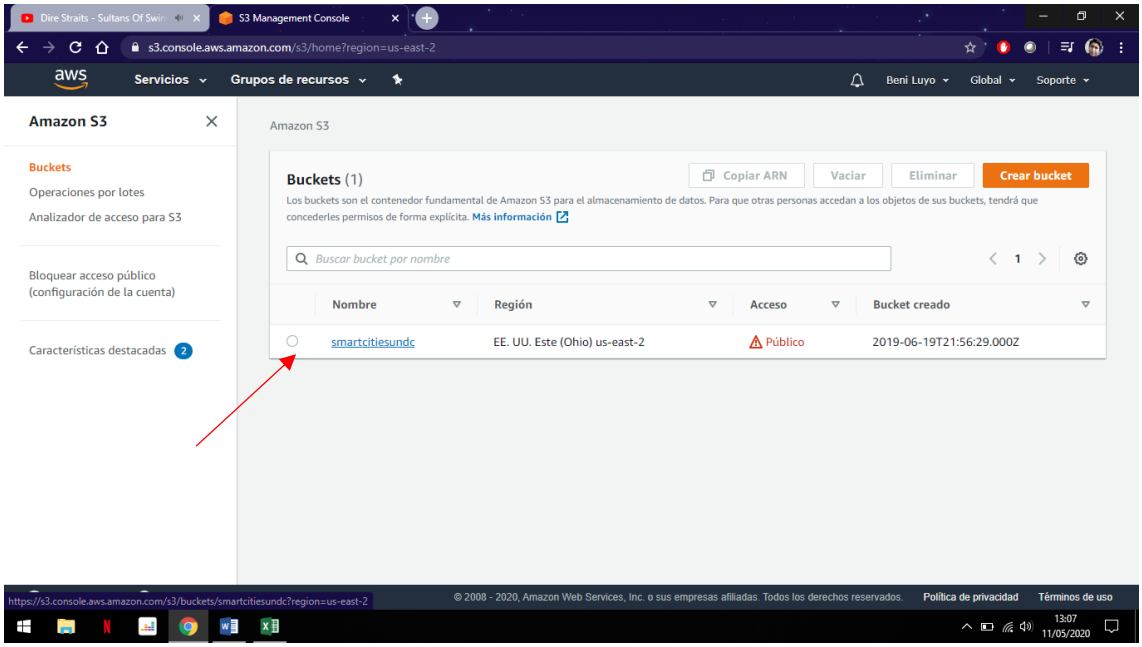
Ahora colocamos un nombre a nuestro bucket y seleccionamos la región, en esta investigación dejamos la región por defecto que es Ohio.



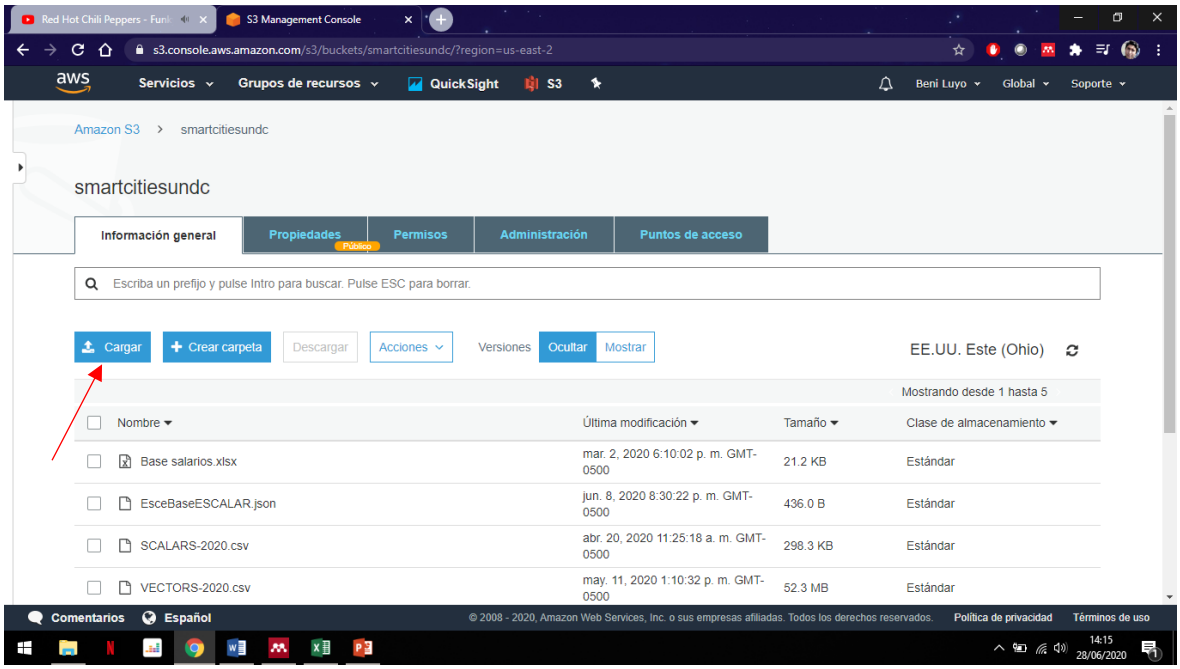
Luego pasamos a configurar los permisos de nuestro bucket, en esta investigación trabajamos con el acceso público activado. Deseleccionamos la opción “Bloquear todo el acceso público”, luego nos aparecerá un recuadro de advertencia de cambio a acceso público, donde seleccionamos la opción “Reconozco que la configuración actual puede provocar que este bucket y los objetos que contiene se vuelvan públicos.” Luego seguimos bajando y damos clic en crear bucket.



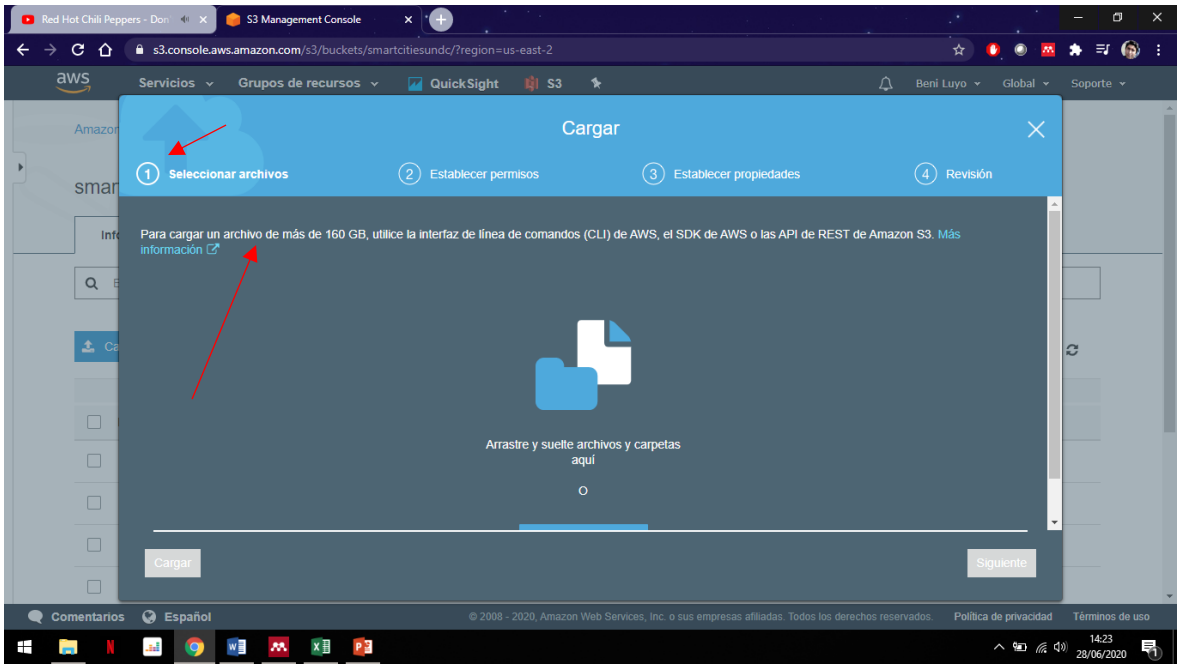
Luego observamos que aparecerá el bucket creado en la lista del recuadro de administración. Seleccionamos el bucket donde se guardará la data para ver las opciones.



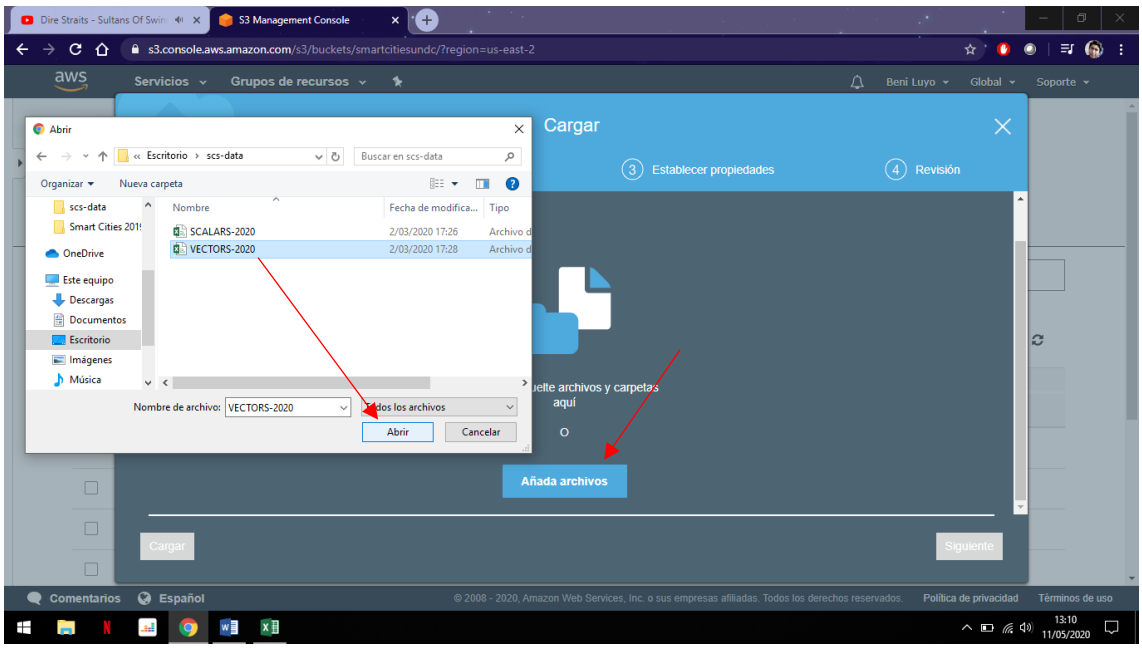
Este apartado nos muestra las diferentes pestañas y configuraciones que podemos realizar con el bucket, tambien nos muestra la lista de archivos que se encuentran almacenados. Ahora para cargar la data, primero hacemos clic en el boton cargar.



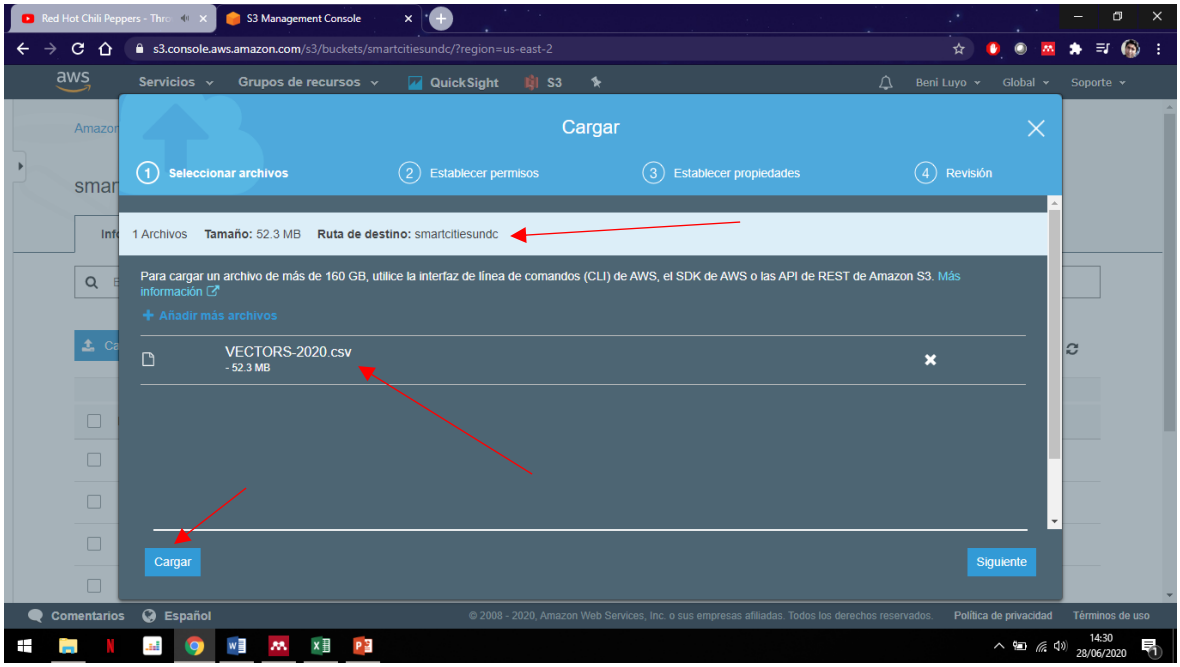
Nos aparecera un ventana emergente, donde observamos pasos parar cargar la data e indicaciones sobre el peso del archivo para subir. En esta investigacion se realizo la carga de manera rapida, omitiendo los pasos para subir archivos.



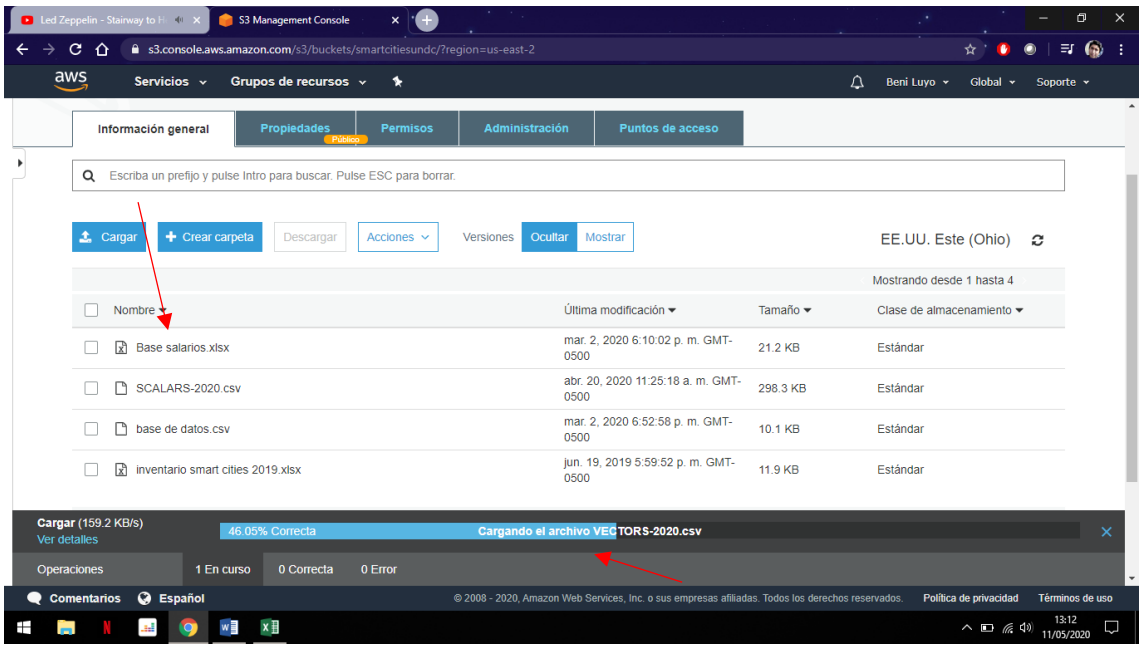
Luego hacemos clic en añadir archivos, aparecera una ventana emergente donde podemos buscamos el directorio en donde se encuentra la data. Seleccionamos el arcihvo, le damos en abrir.



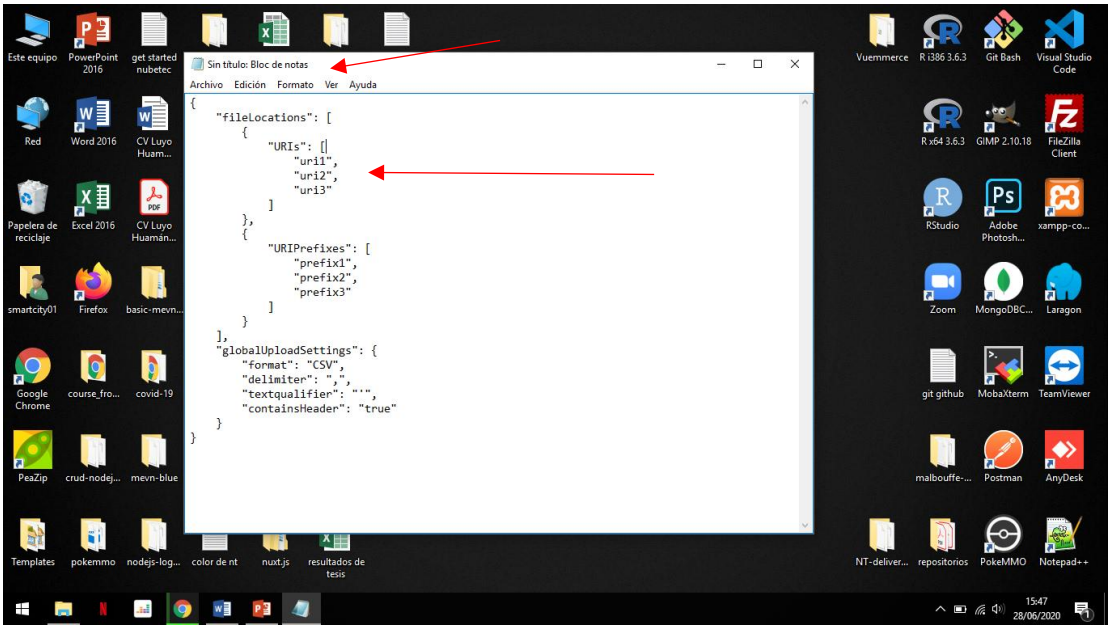
Luego nos aparecera la informacion del tamaño del archivo y ya listo para ser cargado, damos clic en cargar y esperamos.



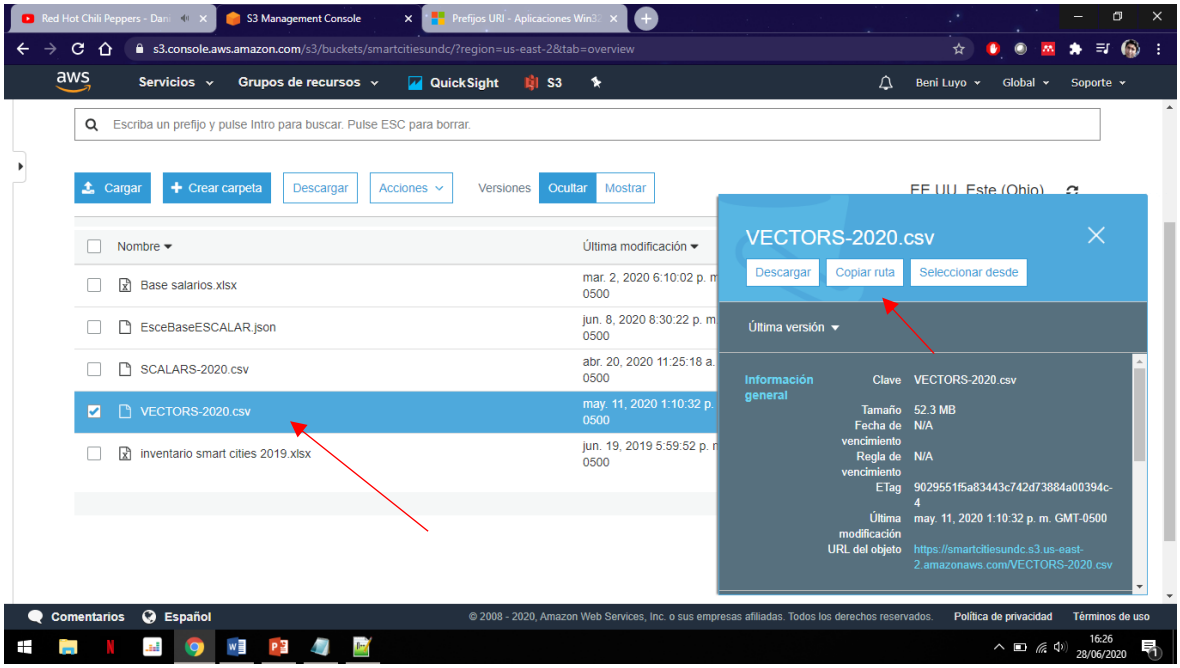
Observamos que en la parte inferior de la pantalla aparece una barra de progreso, que nos muestra el procentaje de carga del archivo, esperamos a que se complete y aparecera en la lista de archivos almacenados.

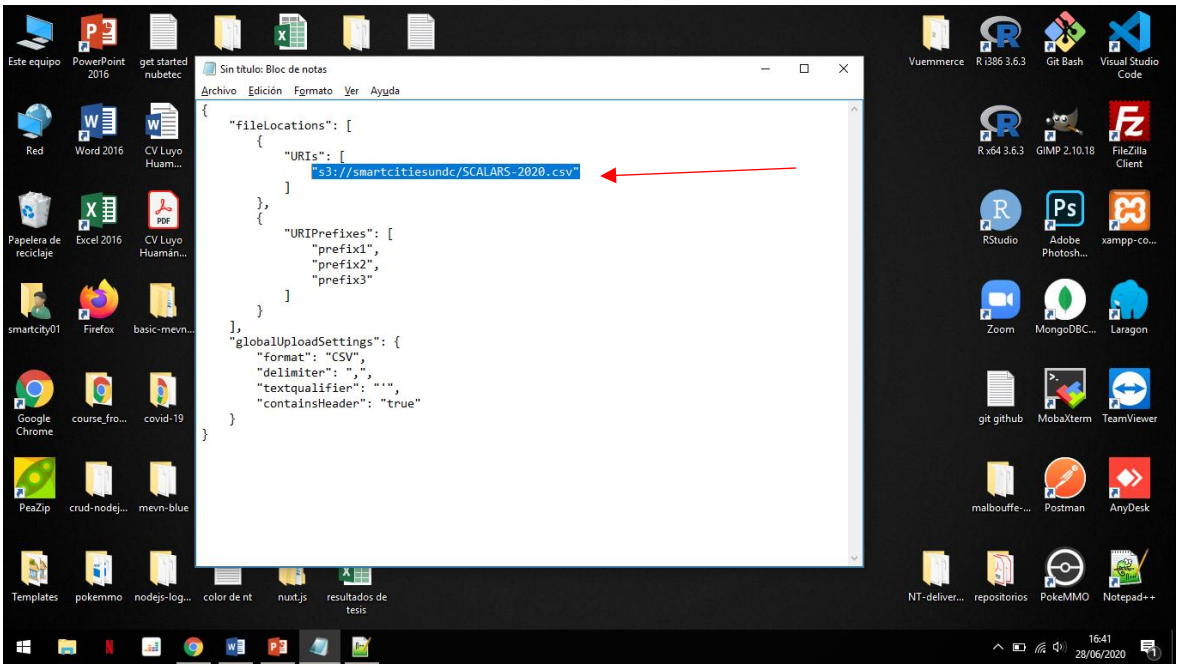


7) Ahora pasaremos a la configuración del archivo de manifiesto json. Este archivo nos permite la conexión entre diversos servicios de AWS, para esta investigación se necesita la comunicación entre los servicios de S3 y Quicksight. Para generarlo, primero copiamos el formato de esta dirección: https://docs.aws.amazon.com/es_es/quicksight/latest/user/supported-manifest-file-format.html, abrimos el bloc de notas y pegamos el formato.

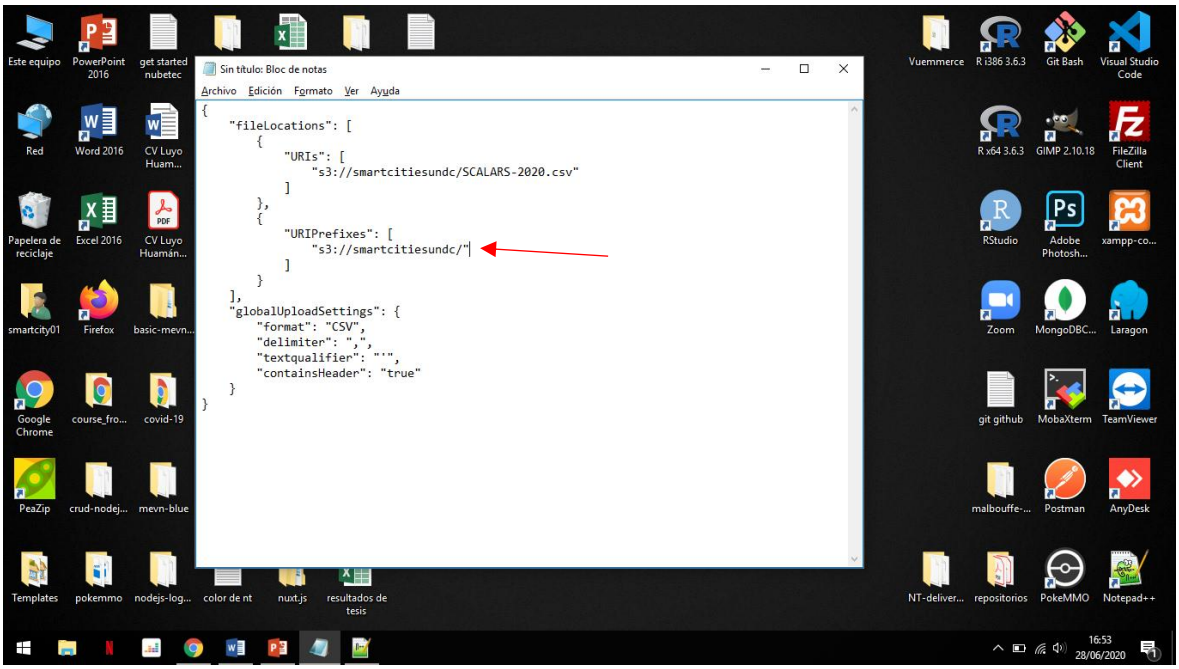


Configuramos las URIs, que son las rutas en donde se encuentra el bucket de S3 y los archivos. En este caso copiamos la ruta del archivo en el bucket de S3.

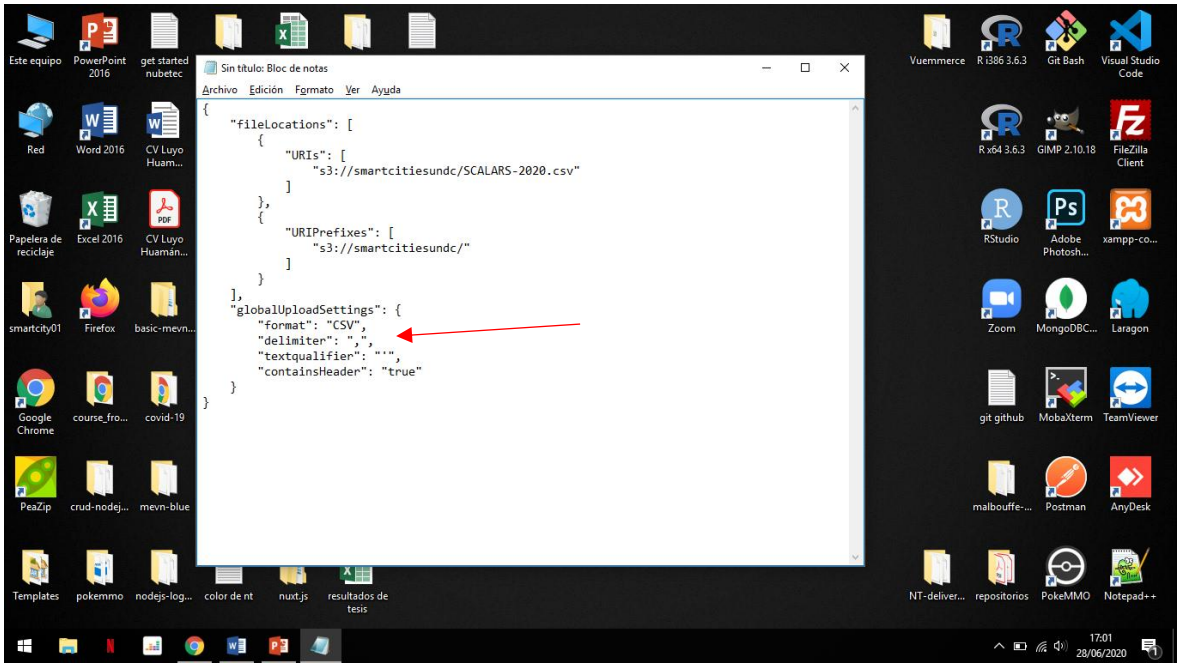




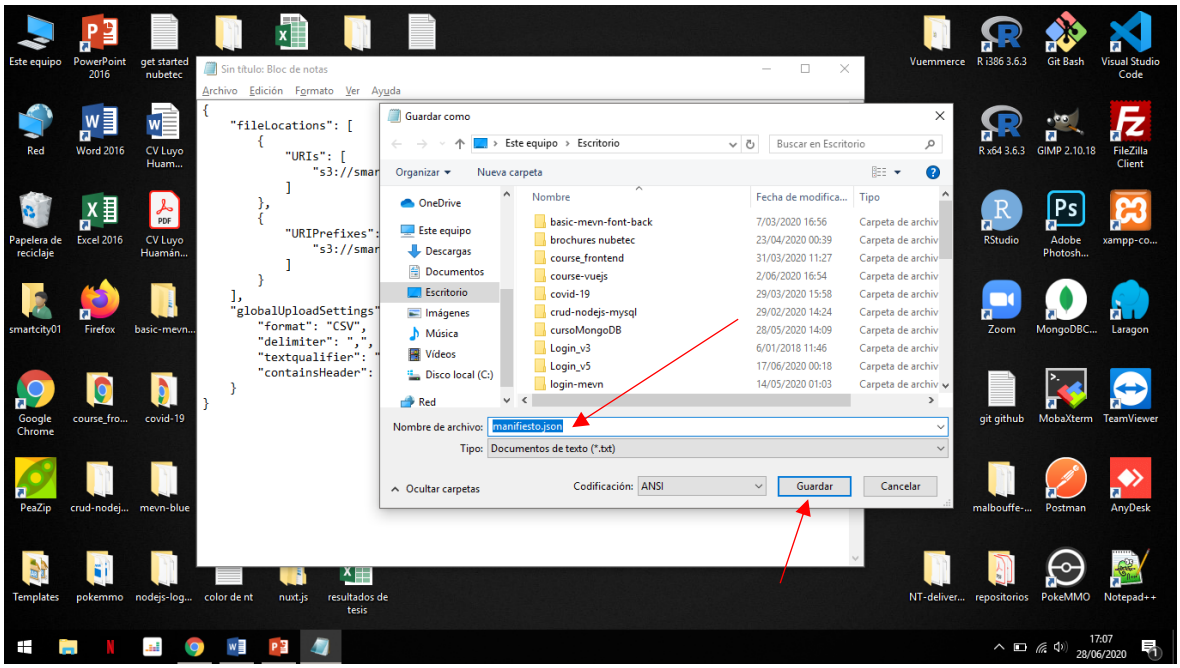
Hacemos el mismo procedimiento para las URIPrefixes, solo que en este caso solo colocamos la URI del bucket.



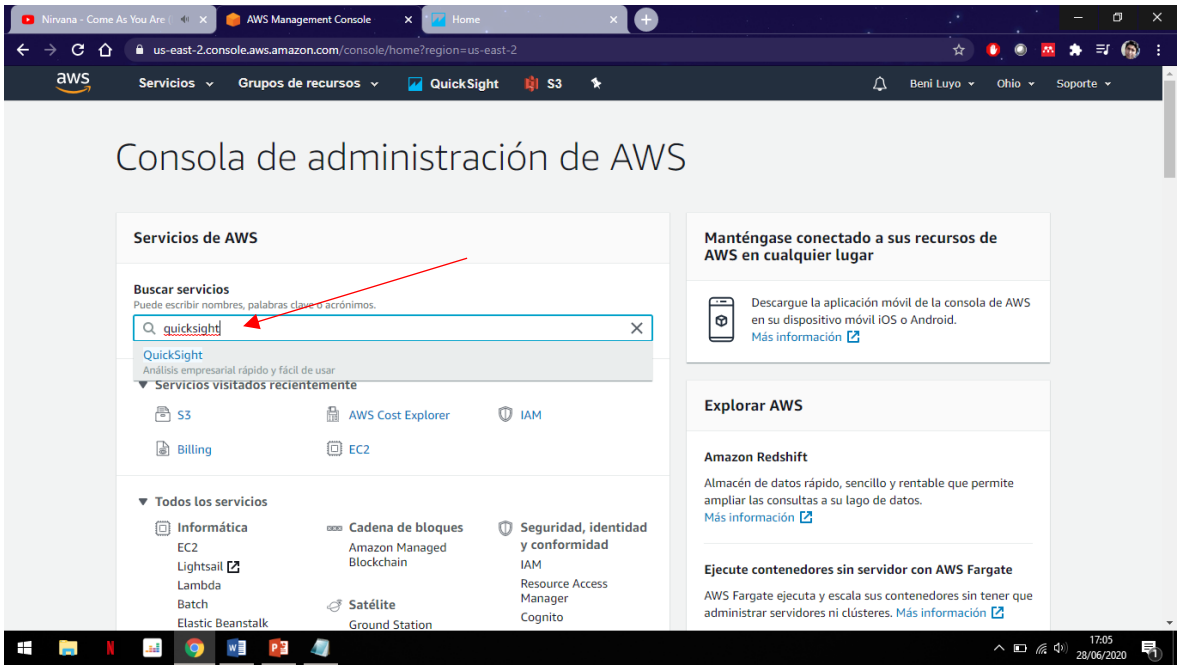
Luego para finalizar configuramos el formato del archivo, su delimitación, el calificador de texto y las cabeceras. En esta investigación guardamos el archivo de datos en formato CSV, delimitado por coma y con cabeceras. En el caso de las cabeceras trabaja con valores booleanos.



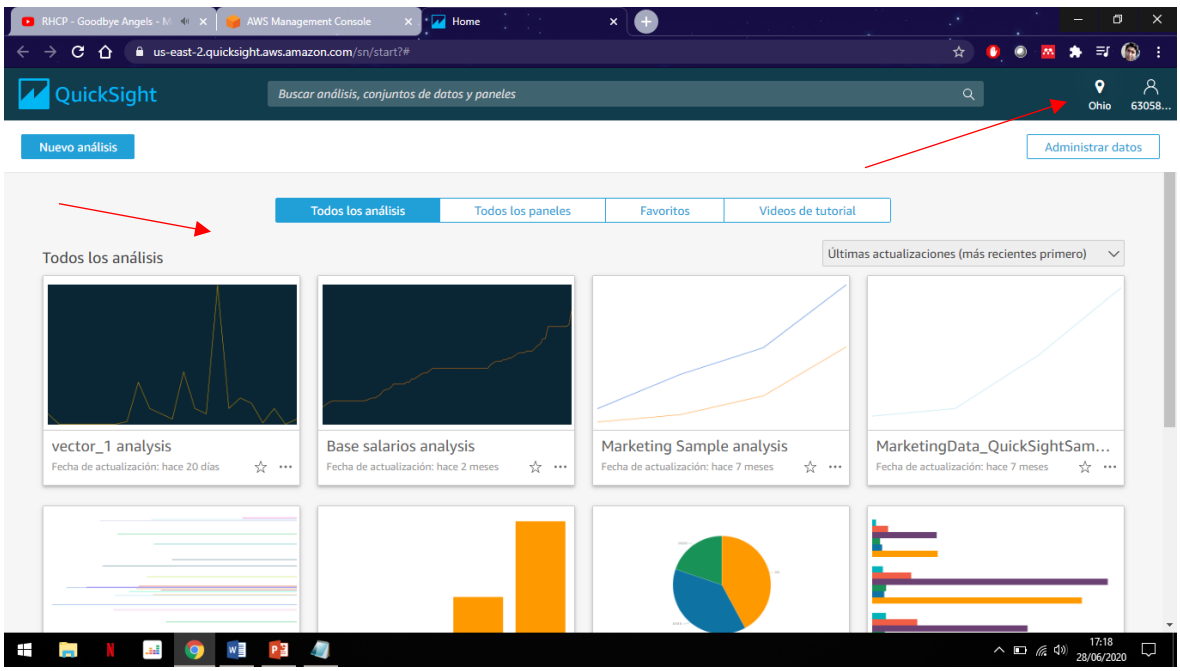
Al final guardamos en archivo ya configurado, escribimos el nombre del archivo con la extensión .json, damos clic en guardar.



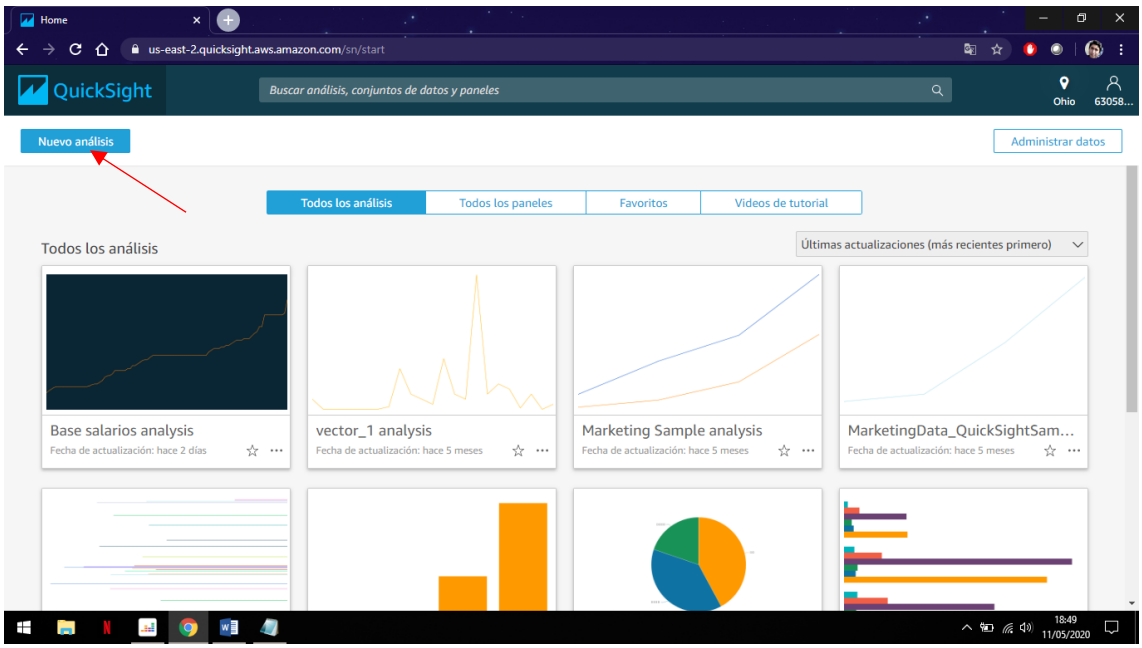
8) Ahora en la consola de administración buscamos la herramienta Quicksight, hacemos clic y vemos que se abrirá una nueva pestaña.



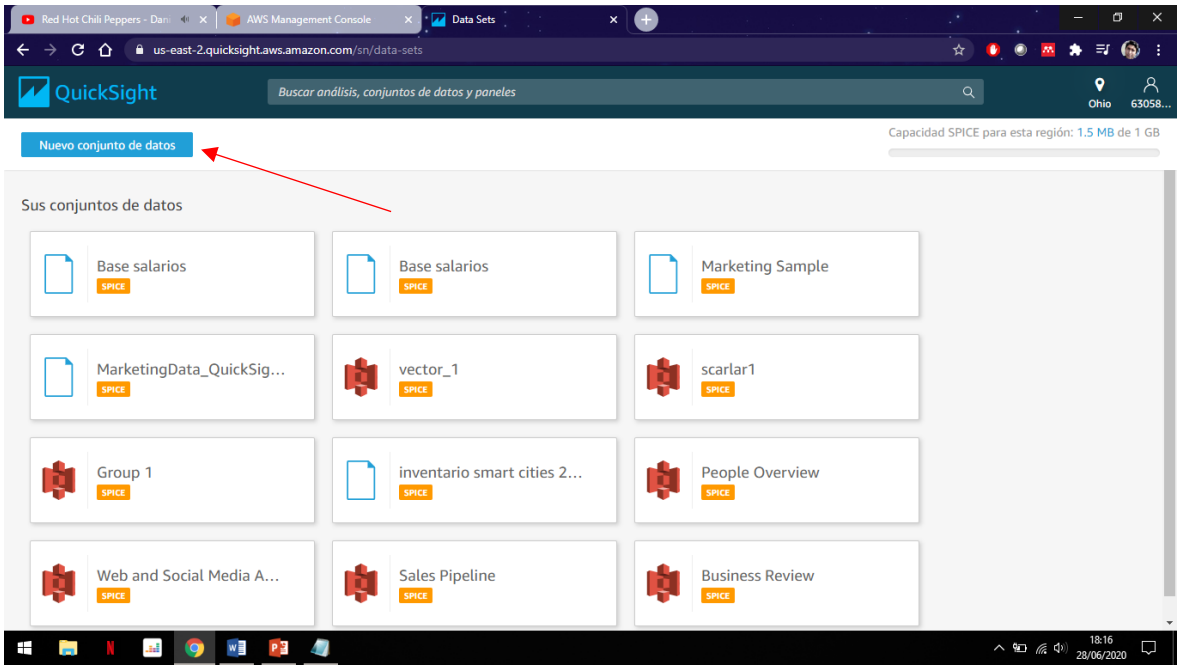
En la nueva pestaña observamos la consola de administración, en la parte superior encontramos la región y el id de usuario, en el recuadro principal observamos las diferentes secciones de insight's creados para la visualización de datos.



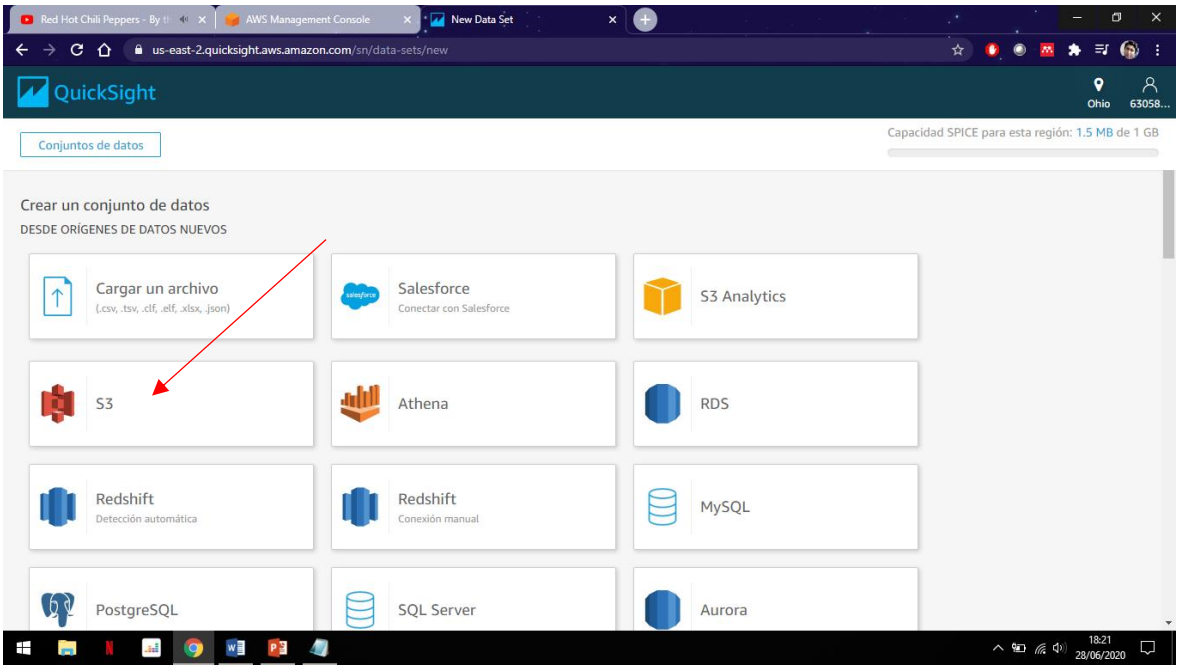
9) Iniciamos un nuevo análisis dando clic a nuevo análisis.



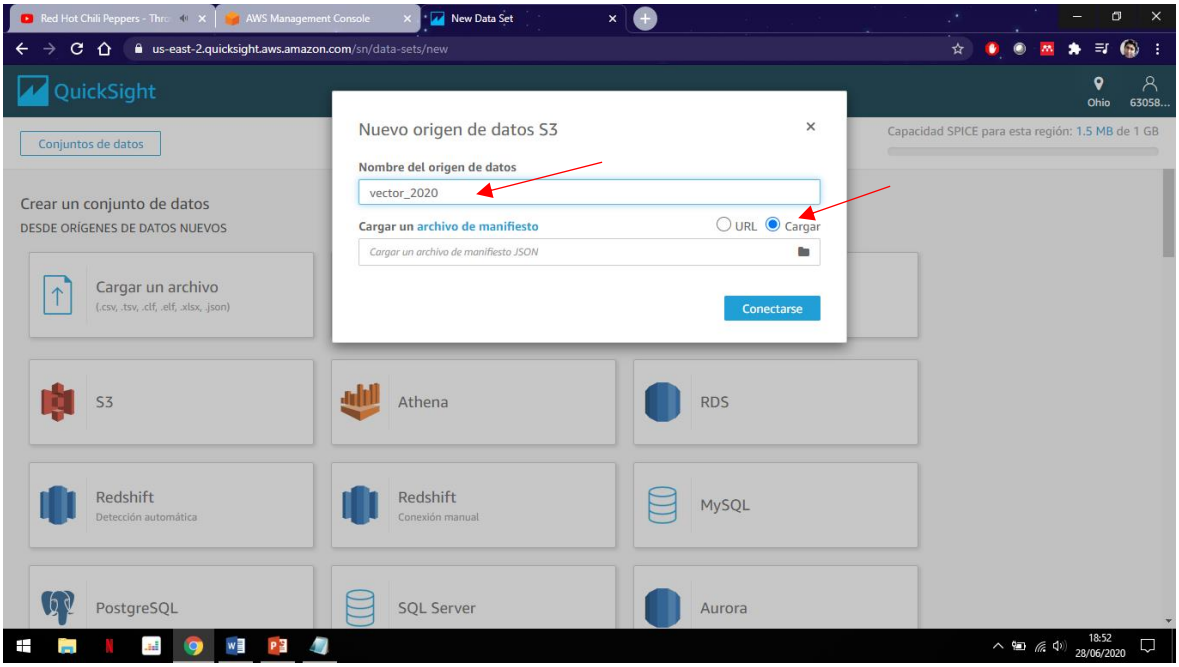
Luego nos aparecerá los conjuntos de datos trabajados y el almacenamiento disponible correspondiente al plan de pago adquirido, en este caso para esta investigación tenemos la prueba gratuita. Hacemos clic en “Nuevo conjunto de datos” para crearlo.



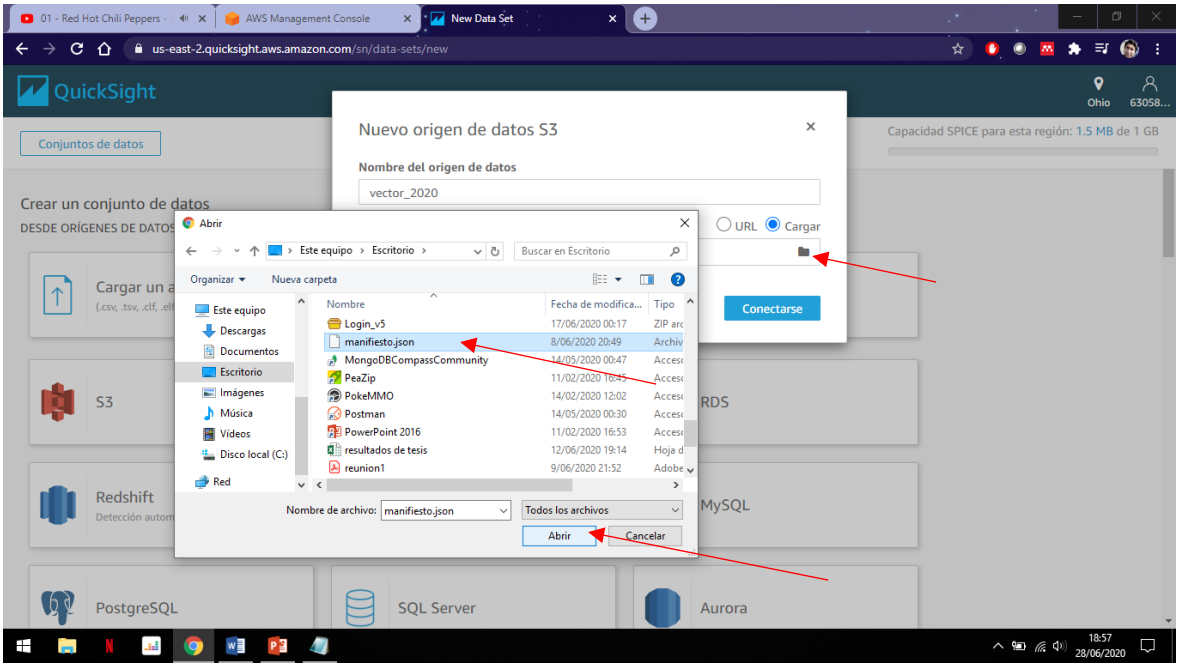
Luego hacemos clic en el botón de S3 para conectarnos y jalar los datos.



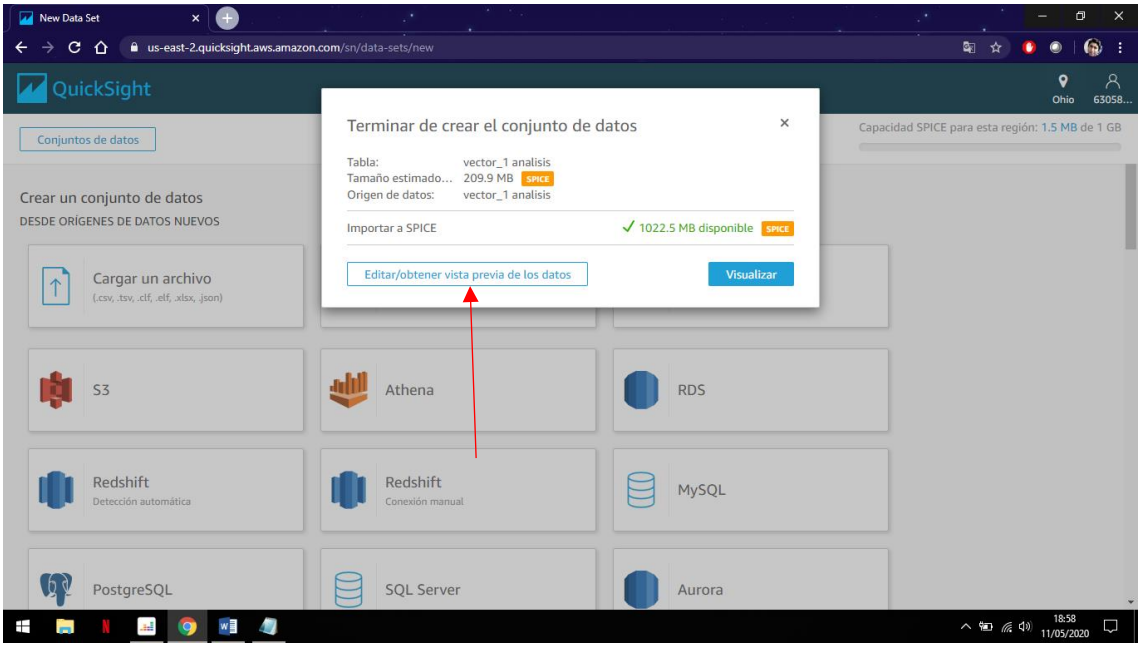
Nos aparecerá una ventana emergente para configurar la conexión con S3 y poder jalar la data guarda en el bucket. Primero colocamos el nombre para el origen de datos y luego cargamos el archivo de manifiesto.json.



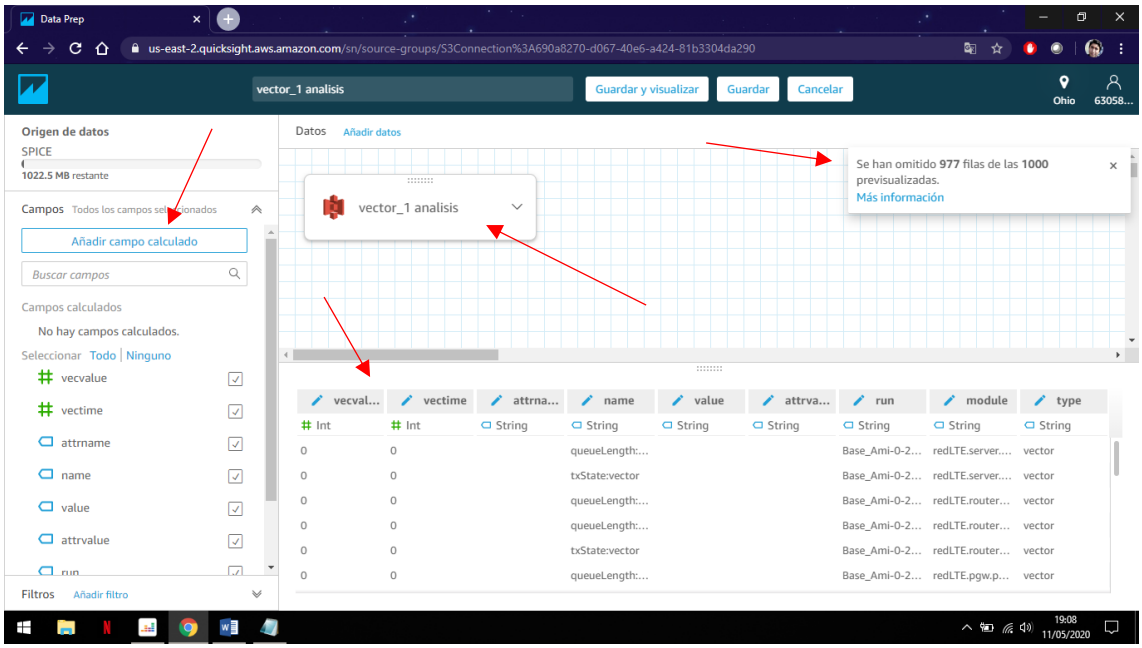
Luego se nos habilitara el icono de directorio para buscar el archivo de manifiesto dentro de nuestro desktop, damos clic, buscamos el archivo, lo seleccionamos y damos clic en abrir. Para finalizar damos clic en “Conectarse”.



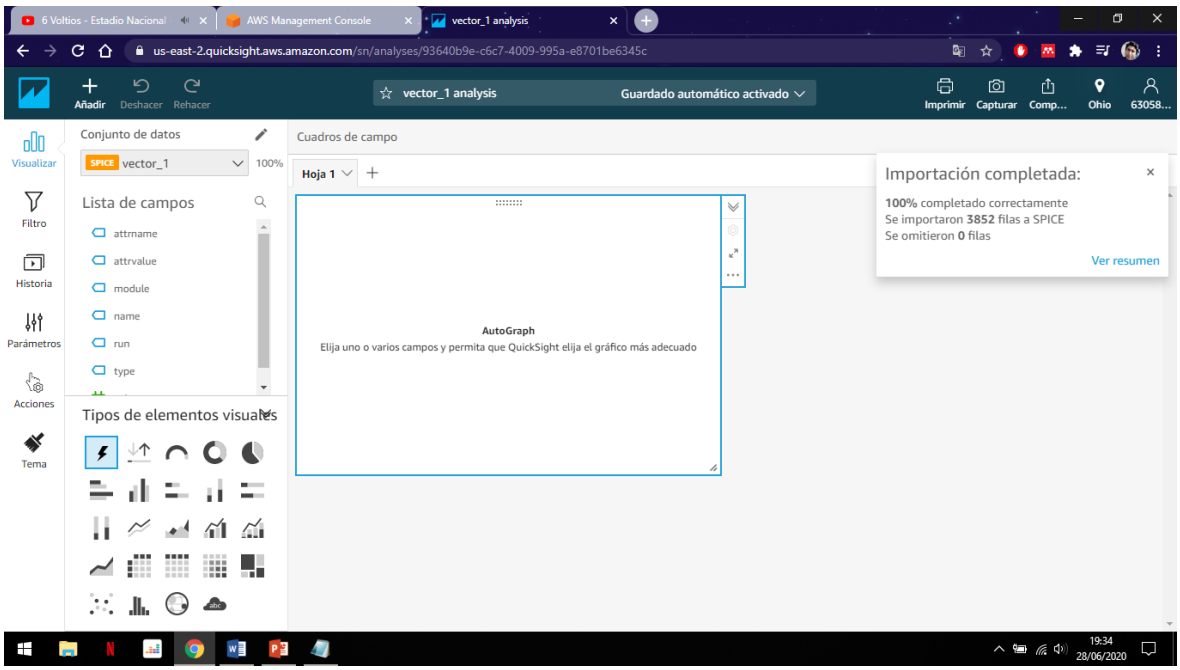
Luego si todo esta correcto aparecerá un aviso, donde haremos clic en editar / obtener vista previa de los datos, para seleccionar los datos que vamos a analizar.



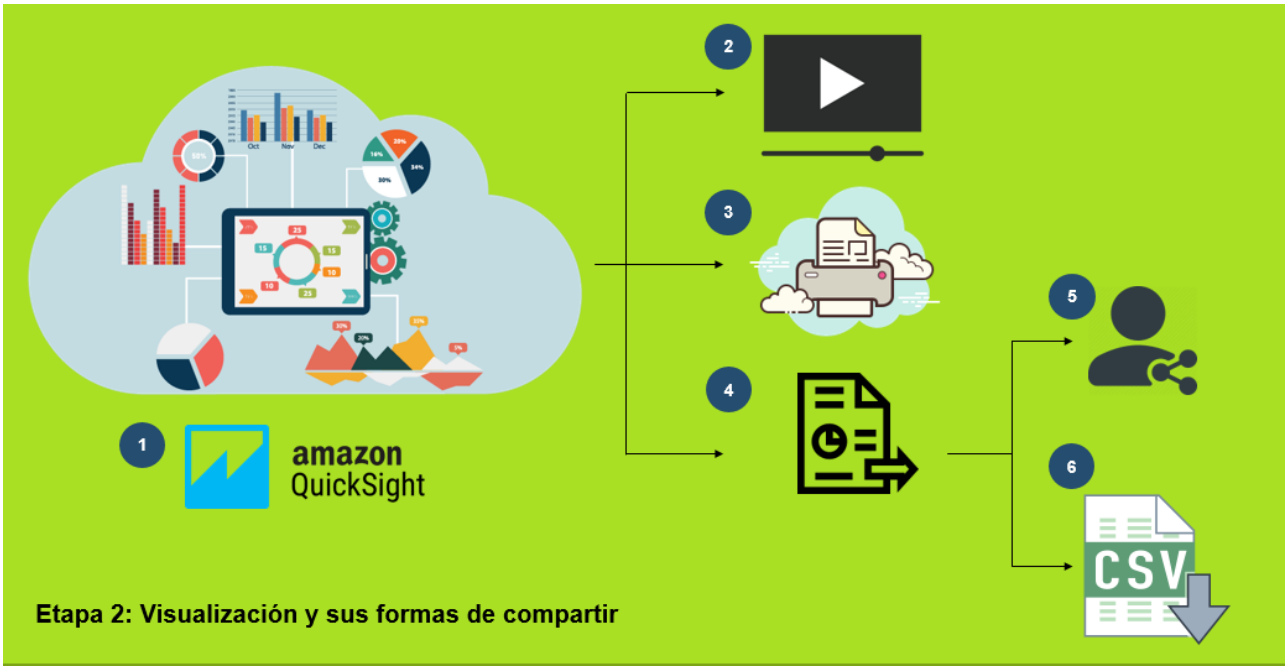
En esta sección podemos hacer el análisis previo a la visualización de datos. Observamos la conexión con S3, las filas importadas y omitidas, las columnas con sus cabeceras y sus valores respectivos, podemos crear algunos campos calculados.



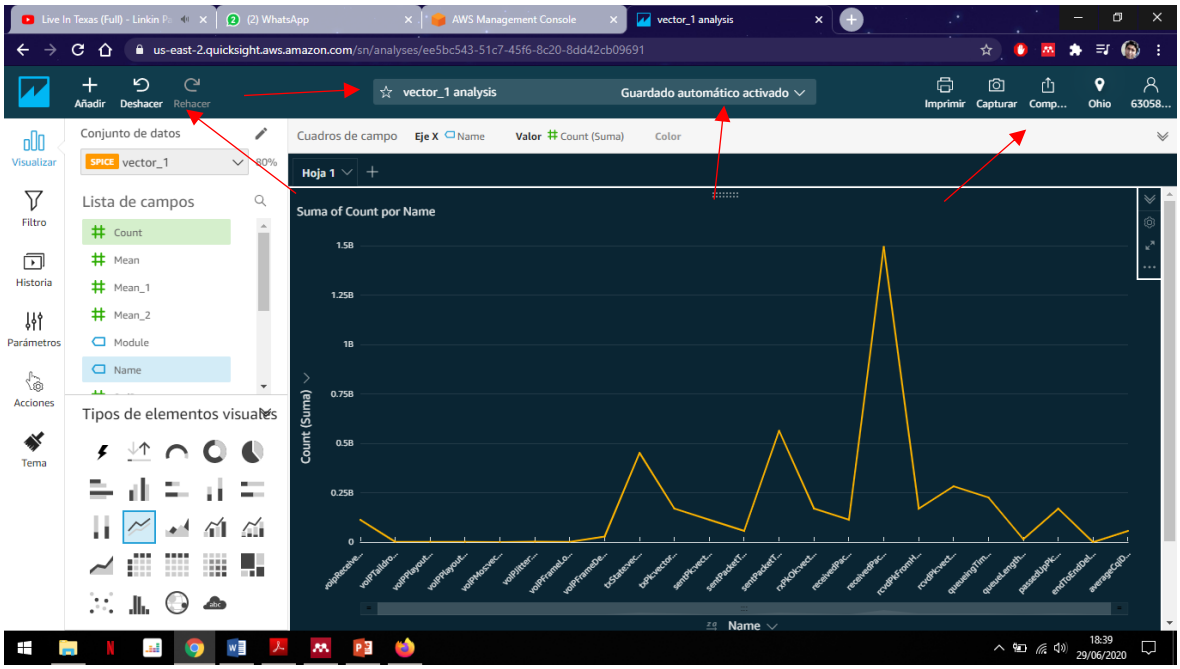
Damos clic en “Guardar y visualizar” para pasar la consola de administración de Quicksight y realizar la visualización de datos mediante las configuraciones que realicemos a nuestro criterio de análisis.



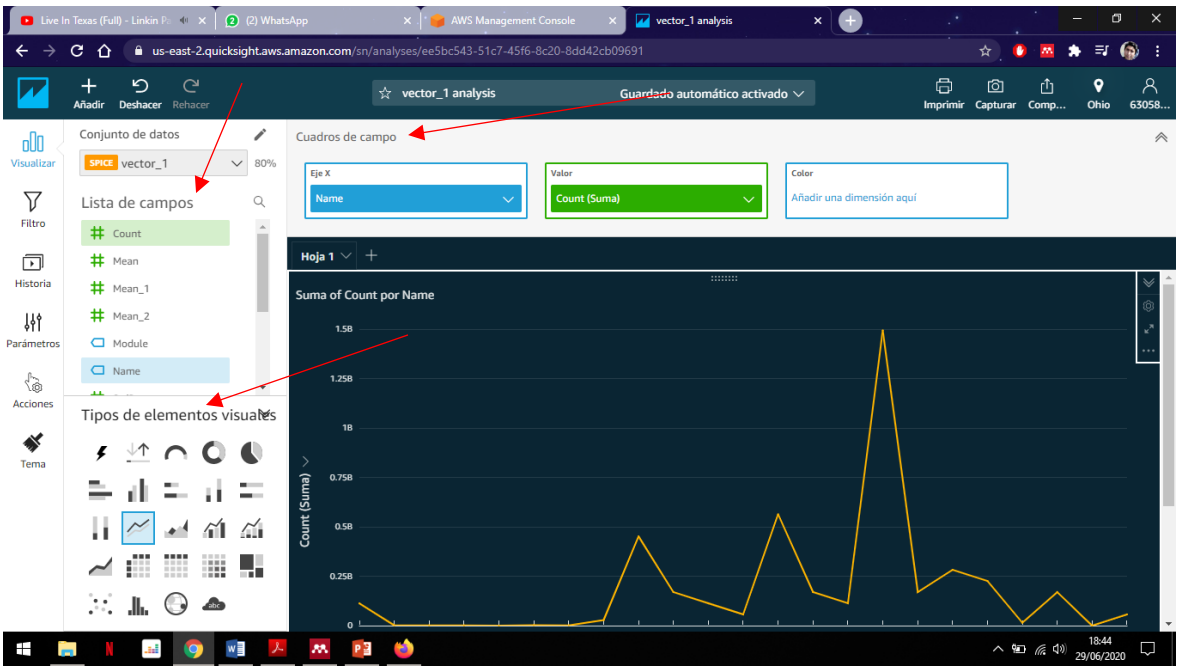
Etapa 2: Visualización y sus formas de compartir



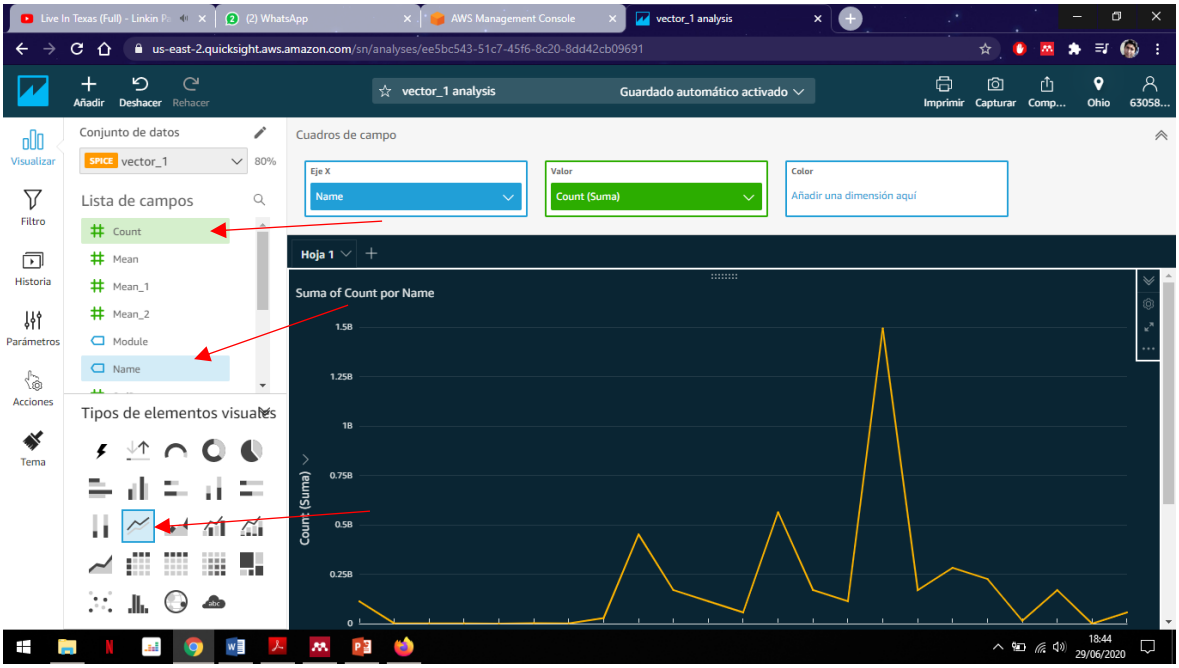
1) Quicksight es una herramienta de business intelligence, que nos ofrece gran variedad de elementos visuales, campos y consultas que podemos crear para poder visualizar de manera óptima la información. En la parte superior observamos el navbar donde encontramos el nombre del análisis de datos, el tipo de guardado (automático por defecto), herramientas y algunas formas de exportar y compartir la visualización de datos que veremos más adelante.



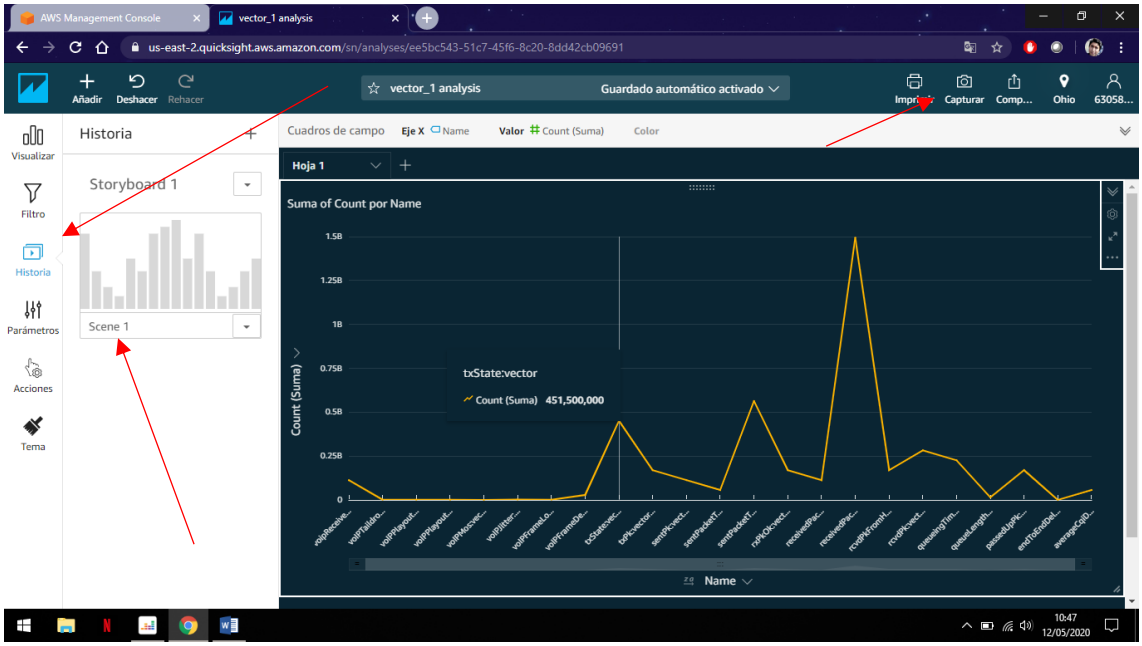
En la sección “Visualizar” tenemos una lista de los campos de datos importados. Están los tipos de elementos visuales, luego en el campo principal observamos la visualización de los datos mediante un gráfico (previamente escogido por el usuario) y también observamos los cuadros de campo, donde podemos seleccionar los campos y colocarlos en el eje que más se acomode en nuestro análisis.

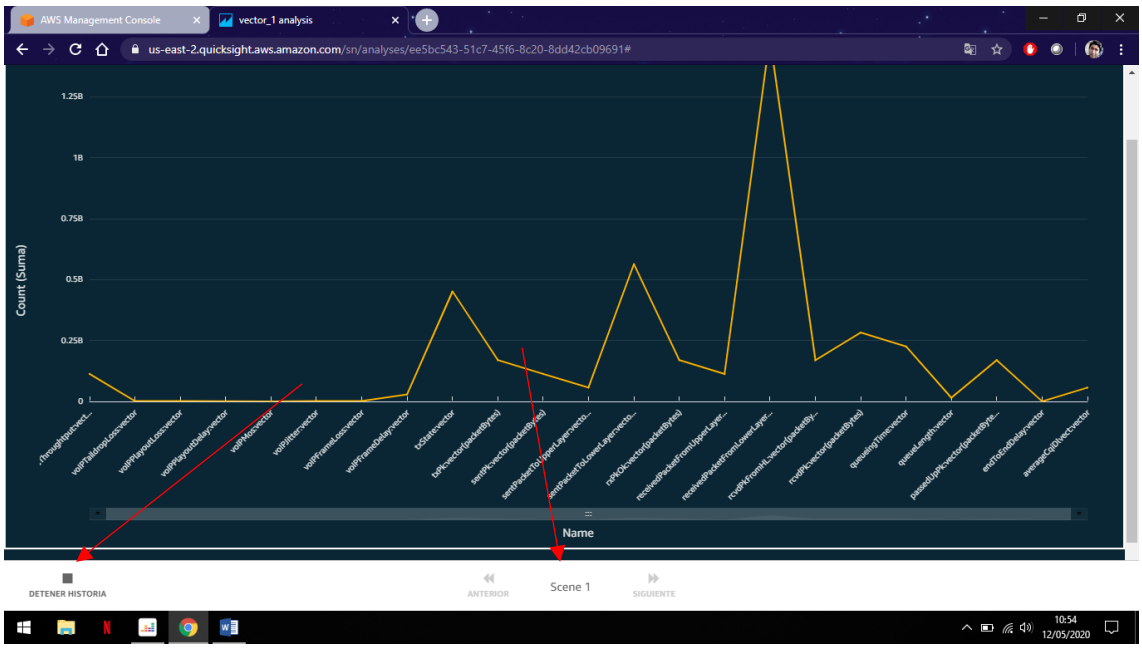


Solo nos enfocaremos en las acciones que realizaremos para esta investigación que es un tema muy amplio explicar toda la herramienta. Para esta investigación escogimos el grafico de líneas y como campos: name (nombre de AMI) y count (cantidad de paquetes recibidos).

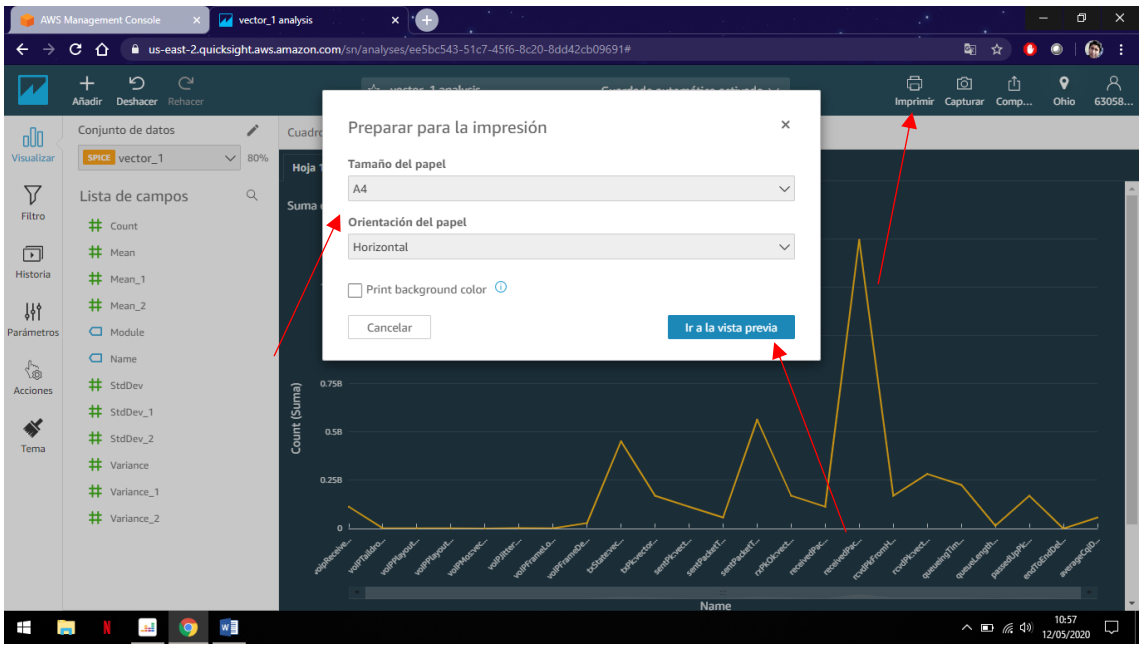


2) Una vez terminado el análisis, podemos capturarlo o pasarlo a un formato multimedia que ofrece Quicksight, lo exporta a la sección Historia guardándolo como Scene 1 como nombre predeterminado, luego se puede cambiar, con esto podemos realizar presentaciones sobre nuestro análisis.

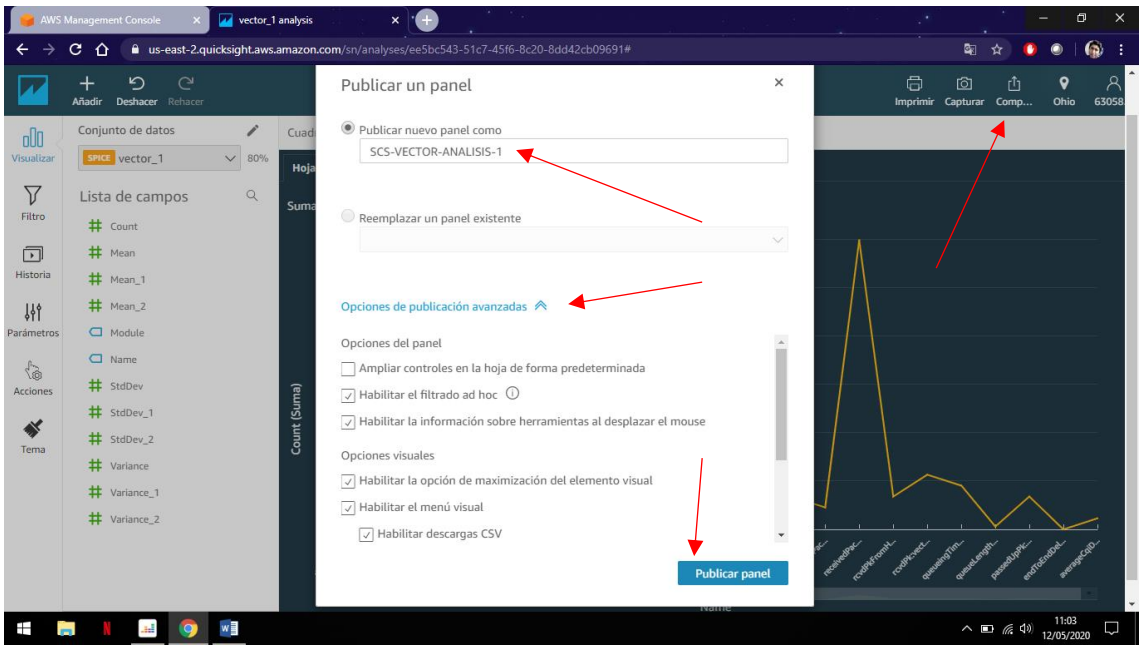




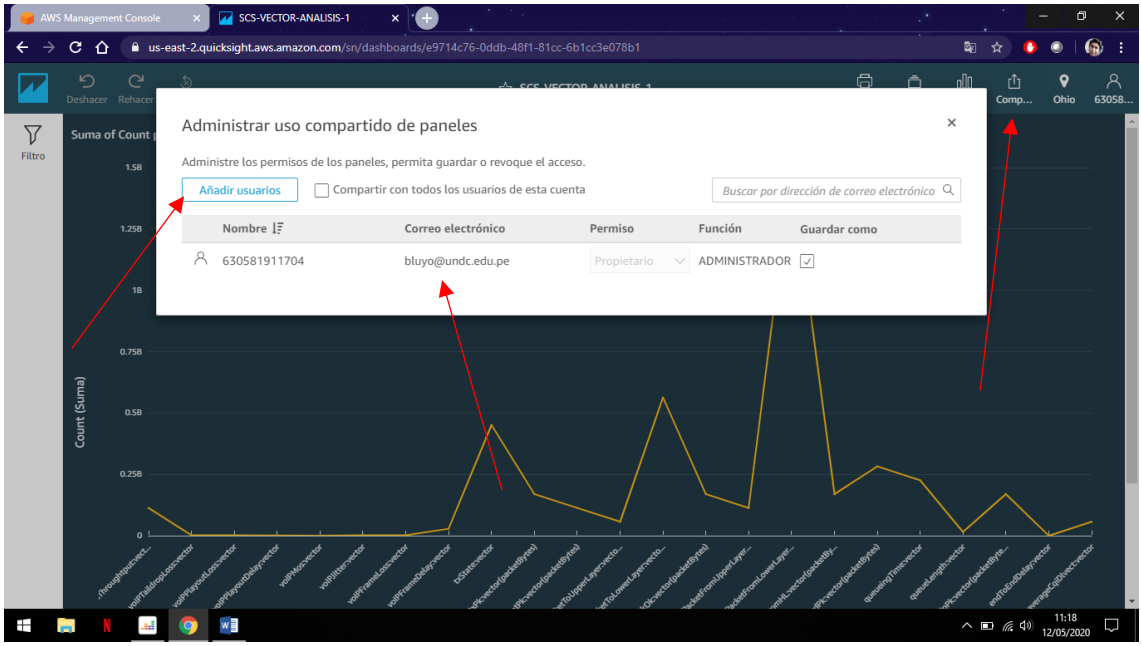
3) También podemos imprimir la gráfica desde la nube, configuramos el tipo de hoja a usar y la orientación. Vamos a vista previa y damos click en imprimir.



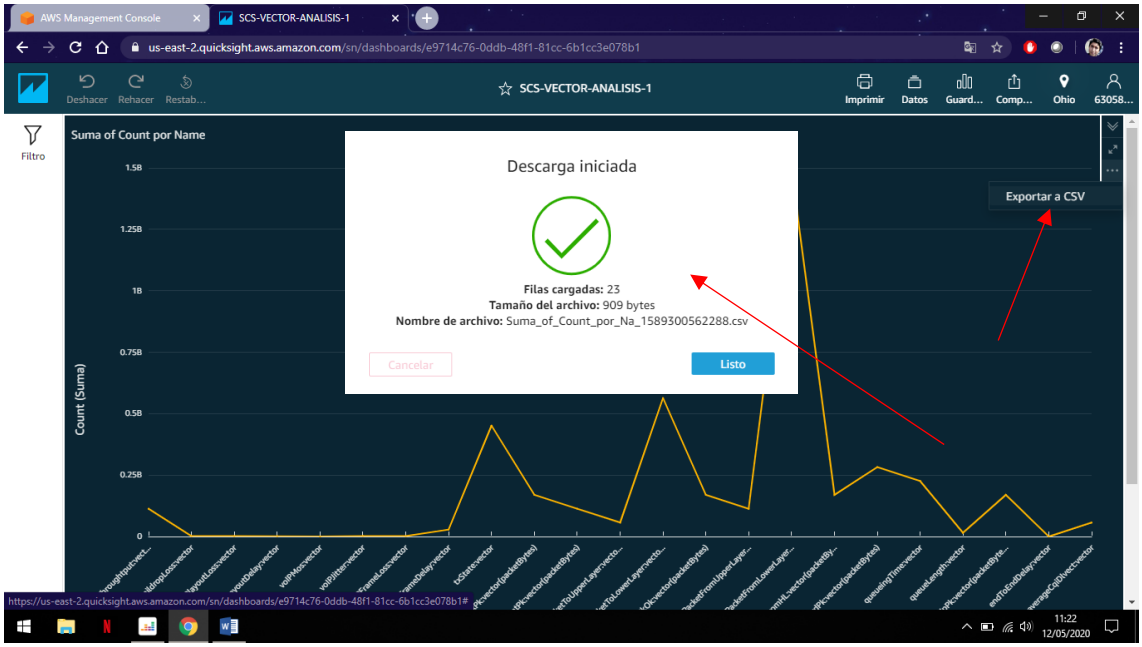
4) Otra forma de compartir es publicando un panel, le damos a compartir, luego en publicar panel, le asignamos un nombre al nuevo panel, configuramos las acciones a realizar y las opciones visuales en opciones avanzadas y le damos a publicar



5) Una vez creado el panel nos aparecerá una ventana para para compartir con otros usuarios Quicksight, si no aparece podemos ir al botón compartir y seleccionar compartir panel, luego solo el Usuario Raíz puede administrar el acceso al panel, delegar permisos, etc. Estos paneles solo se pueden compartir con otros usuarios de AWS, por medidas de seguridad.



6) Podemos descargar el análisis en formato csv, en la parte superior derecha del grafico esta una barra en donde encontraremos la opción de exportar a csv, damos clic, esperamos a que descargue y ubicamos la carpeta en donde se guardó para futuros análisis.



IV. Resultados y Discusión

4.1 Resultados

Porcentaje de calidad de datos

La calidad de datos es solo una percepción, no existen estandarizaciones y es por eso que deben crearse y definirse reglas en función de los objetivos y requerimientos de la organización o negocio.

(PowerData, 2014) nos dice que estas reglas nos deben permitir:

- Identificar problemas de calidad de datos en los activos informacionales del negocio.
- Corregir estas cuestiones.
- Predecir tendencias.
- Contribuir con su aportación a ampliar el alcance del análisis.

Para poder medir la calidad de los datos, en **(PowerData, 2014)** nos dice que existen una serie de medidas cuantitativas que son universales:

- Completitud: es el grado en el que todos los atributos del dato están presentes.
- Validez: representa el ajuste de un valor de datos a su conjunto de valores de.
- Unicidad: la medida en que todos los valores distintos de un elemento de datos aparecen sólo una vez.
- Integridad: tiene que ver con el grado de conformidad con las reglas de relación de datos definidas.
- Precisión: determina en qué medida los datos representan correctamente la verdad sobre un objeto del mundo real o se ajustan a lo establecido por una fuente autorizada.
- Coherencia: representa el grado en que una pieza única de datos contiene el mismo valor a través de múltiples conjuntos de datos.
- Oportunidad: este atributo de la calidad de datos permite conocer si éstos están disponibles cuando se requiere.
- Representación: tiene que ver con el formato, patrón, legibilidad y utilidad de los datos para su uso previsto.

Teniendo en cuenta todo lo anterior, en esta investigación para obtener el porcentaje de calidad de los datos, se debe observar que los datos obtenidos mantengan su integridad, legitimidad y nos muestren datos verdaderos sobre el desempeño de la red 4G LTE en los medidores eléctricos inteligentes, los cuales son: name, type, module, attrname, attrvalue, value, vectime y vecvalue, deben tener relación con sus respectivos valores, de lo contrario no serían datos de calidad ya que serían datos basura.

El valor de la calidad de datos se obtiene mediante la división del número de lecturas por personas sobre el número de simulaciones realizadas multiplicada por 100, para luego promediarlas y lograr el fin de esta investigación.

$$calidad\ de\ datos = \frac{n°\ lectura\ por\ personas}{n°\ de\ simulaciones} * 100$$

Para el n° de lectura por persona se califica en la escala del 1 al 10.

Tabla 13. Margen de referencia del índice de calidad.

Margen de referencia	Índice de calidad
= 100	Excelente
≥ 50	Bueno
≤ 49	Malo

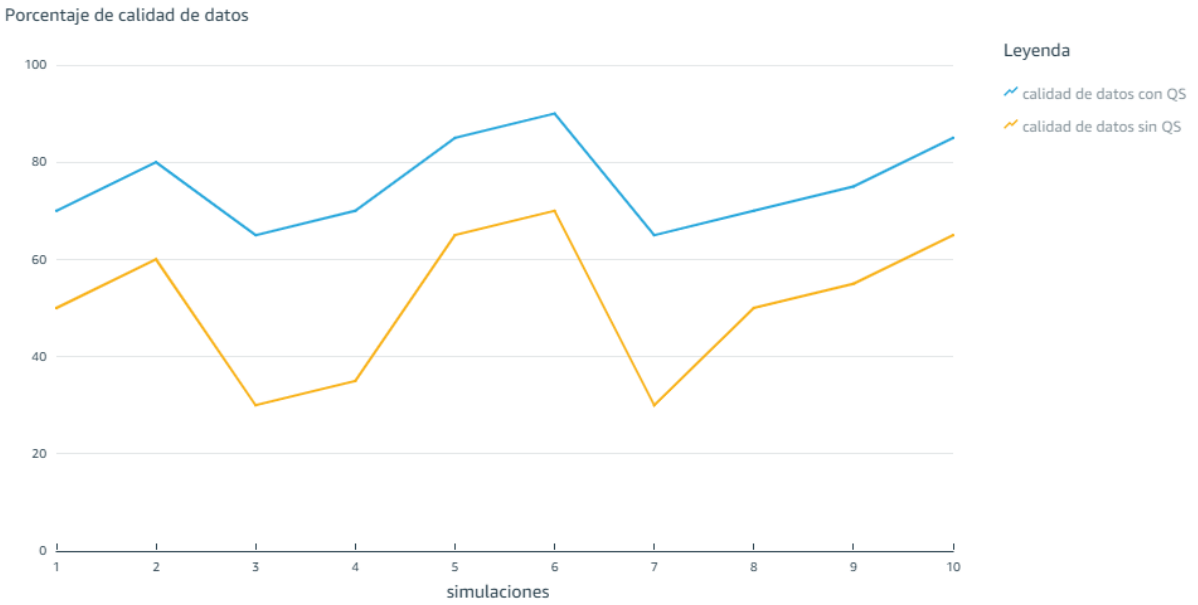


Figura 12: Porcentaje de calidad de datos. Fuente: Elaboración propia.

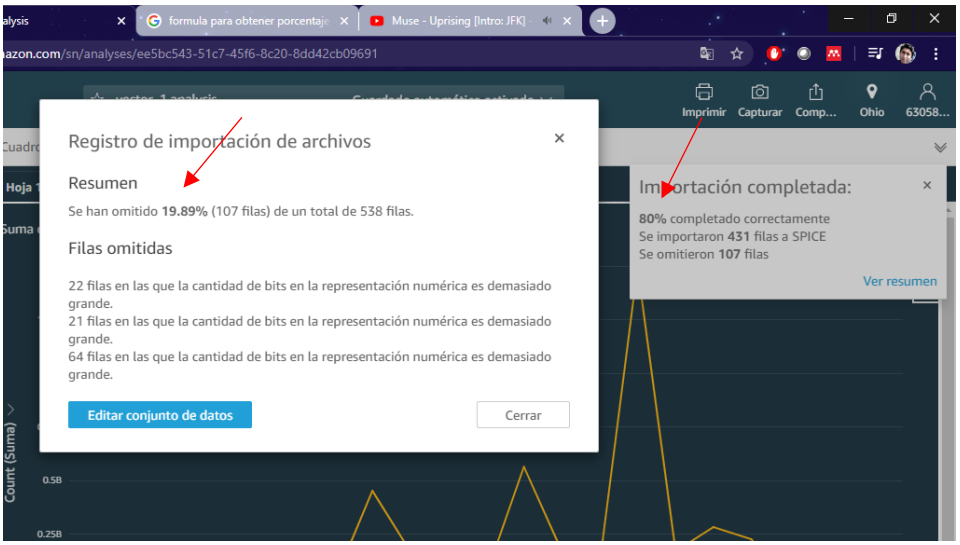
En la figura 12 podemos observar cómo se pudo aumentar el porcentaje de calidad de datos aplicando la herramienta cloud Quicksight, teniendo gran impacto en las 10 simulaciones.

Porcentaje de lectura y visualización de datos

En el pre-test, solo se podrá calcular mediante juicios de expertos, ya que no se aplica la herramienta y por lo tanto solo habrá datos estructurados con bajo porcentaje de lectura y sin visualización gráfica.

En el post-test la misma herramienta Quicksight nos muestra el número total y el porcentaje de filas importadas y leídas por su motor de análisis de datos SPICE, es por esta razón que en el pre-test no es posible obtener este indicador porque no se aplica la herramienta.

En la siguiente imagen podemos observar como la herramienta nos muestra de manera automática el número total y el porcentaje de filas leídas y visualizadas.



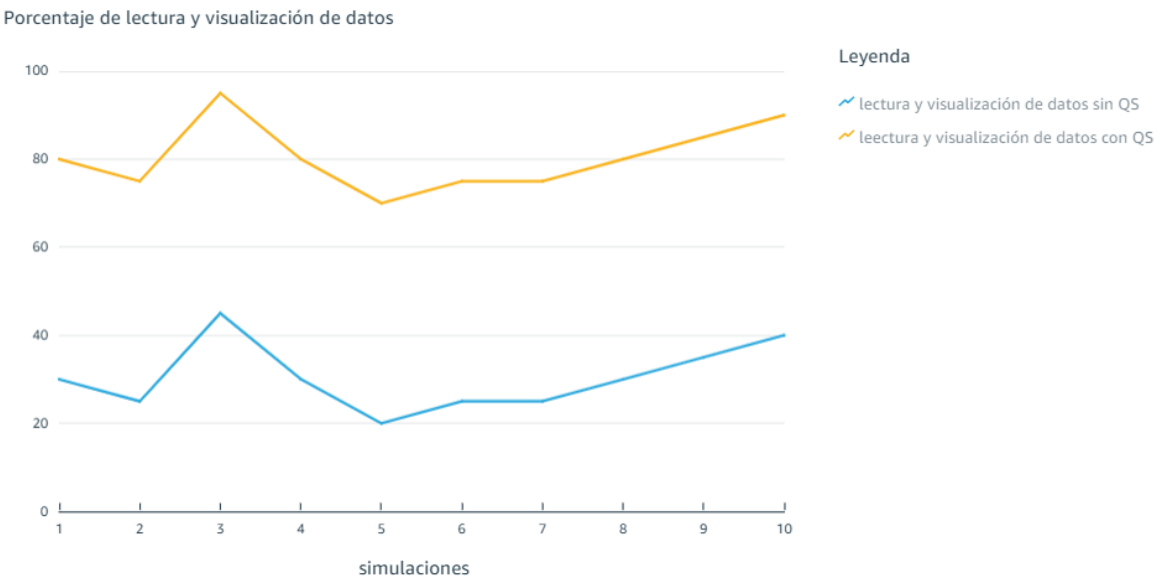


Figura 13: Porcentaje lectura y visualización de datos. Fuente: Elaboración propia.

En la figura 13 podemos observar cómo hubo una mejora en el porcentaje de lectura y visualización de datos, esto se debe a que en el pre test no se aplica la herramienta Quicksight y solo se muestran datos en tablas con bajo índice de lectura y sin visualización gráfica.

Porcentaje de datos útiles

Basándonos en la referencia de **(PowerData, 2014)**, tomamos los objetivos y requerimientos de la empresa para poder obtener el porcentaje de datos útiles, en otras palabras los datos que nos ayudaran a la toma de decisiones, los datos que permitan ver el desempeño de los smart meters mediante la red 4G LTE, name, value, vectime y vecvalue. En el pre-test este porcentaje será bajo porque tendremos toda la data sin procesar y sin visualización, siendo más complicado poder usarla para la toma de decisiones por los expertos.

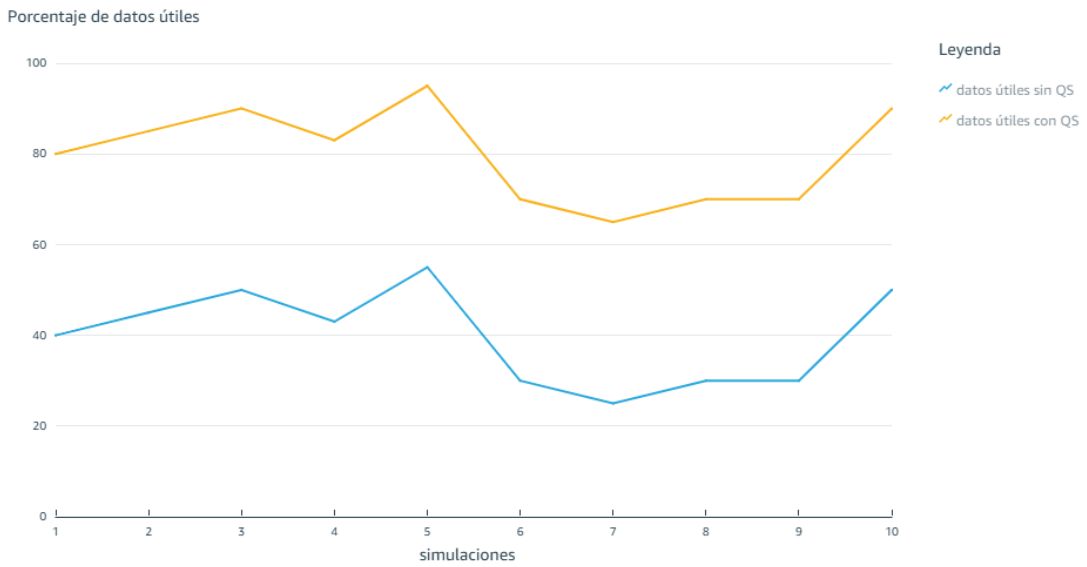


Figura 14: Porcentaje de datos útiles. Fuente: Elaboración propia.

En la figura 14 podemos observar como la herramienta Quicksight tuvo impacto positivo en el porcentaje de datos útiles. Gracias a la visualización, se podrán tomar las mejores decisiones en la organización.

En la siguiente tabla se muestra los resultados del grupo experimental, donde presento un aumento del 27.39% en la media de porcentajes de calidad de datos, se aprecia un aumento del 63.70% en la lectura y visualización de datos. Finalmente, un aumento del 47.30% en el porcentaje de datos útiles.

Tabla 14. Descriptivos según implantación de la herramienta cloud Quicksight.

Indicadores	Control	M	CVM
Porcentaje de calidad de datos	Con Quicksight	75.5	+27.39%
	Sin Quicksight	51	
Porcentaje de lectura y visualización de datos	Con Quicksight	80.5	+63.70%
	Sin Quicksight	30.5	
Porcentaje de datos útiles	Con Quicksight	79.8	+47.30%
	Sin Quicksight	39.8	

Nota: N=10; M = Media; CVM = Coeficiente de variación de medias.

4.2 Discusión

El propósito de esta investigación fue determinar la influencia de la herramienta cloud Quicksight para la visualización de datos obtenidos del proyecto Smart Cities: Cañete plan piloto con medidores eléctricos inteligentes, Estos resultados se contrastaron con anteriores investigaciones citadas anteriormente.

Al aplicar la herramienta Quicksight se logró aumentar de porcentaje de datos de calidad, corroborando mediante encuestas aplicadas a los involucrados en el proyecto Smart Cities que afirmaron la obtención de datos limpios de buena calidad para lograr la visualización, estos resultados son congruentes con las investigaciones de **(Salazar, 2017)** donde evaluó la calidad de información mediante encuestas a usuarios obteniendo un 85% de afirmación en la mejora de la calidad, permitiendo mejorar la toma de decisiones en la empresa aplicada.

Mediante la aplicación de la herramienta cloud Quicksight para medir el porcentaje de lectura y visualización de datos se logró importar 80% de la data con 431 filas para el proceso de visualización de datos, gracias al motor de análisis SPICE limpiando algunos datos nulos y llevándolos a la visualización. Estos resultados son congruentes con los de **(Peralta, 2019)**, el cual logro la representación visual de 210 millones datos, mediante mapas de calor (representación gráfica de datos) utilizando diferentes consultas evaluadas tienen gran importancia al interpretar los datos y que es de vital importancia en la visualización de datos.

Al obtener diferentes insight's con la herramienta Quicksight se logró generar información importante a partir de los datos útiles, esto se debe a que podemos visualizar estos resultados de diferentes formas que nos ayudan a una buena toma de decisiones enfocada a los objetivos y metas de la empresa u organización. Estos resultados son congruentes con los resultados de **(Rodriguez, 2017)** afirmando que la disposición de la información y su visualización soportan la toma de decisiones en la cadena agroindustrial es del 100% cierta.

En relación a la pregunta de investigación: ¿Cómo influye la herramienta cloud Quicksight en la visualización de datos obtenidos de la simulación Smart Cities: Cañete plan piloto con medidores eléctricos inteligentes? La investigación nos permite aseverar que la implantación de Quicksight mejora el nivel de confiabilidad de los datos obtenidos, lo cual ayuda a mejorar el análisis, interpretación y toma de decisiones en base a gráficos estadísticos (insight's) con alto porcentaje de datos de calidad, lectura y visualización y datos útiles.

Finalmente **(Cárdenas Lino, 2016)**, en su investigación logra obtener datos sobre el desempeño de la red LTE en los Smart Meters mediante simulaciones, citando que su investigación seria de ayuda para trabajos futuros con cloud computing, donde se corrobora que si es posible, ya que pudimos llevar a la visualización los datos obtenidos de una simulación similar, con ayuda de la herramienta cloud Quicksight. También habla sobre la seguridad en cloud computing, donde se pudo contrastar que, si es eficiente, ya que los elementos de visualización son privados y para poder acceder a ellos, se necesita una cuenta de AWS con su respectivo permiso otorgado por el creador del elemento visual.

V. Conclusiones

Basándonos en los resultados logrados en la investigación podemos concluir que:

La implantación la herramienta cloud Quicksight incidió positivamente en el proceso de visualización de datos obtenidos de la simulación Smart Cities: Cañete plan piloto con medidores eléctricos inteligentes, porque permitió tomar decisiones acertadas en tiempo real a través de la transformación, interpretación y visualización de información.

Se concluye que la implantación de la herramienta cloud Quicksight aumento el porcentaje de datos de calidad con una media del 27.39%, por lo tanto, se puede afirmar que Quicksight influyo positivamente en el aumento del porcentaje de calidad de datos obtenidos de la simulación Smart Cities: Cañete plan piloto con medidores eléctricos inteligentes.

Se concluye que la implantación de la herramienta cloud Quicksight aumento el porcentaje de datos lectura y visualización de datos con una media del 63.70%, por lo tanto, se puede afirmar que Quicksight influyo de manera positiva en el aumento del porcentaje de lectura y visualización de datos obtenidos de la simulación Smart Cities: Cañete plan piloto con medidores eléctricos inteligentes.

Se concluye que la implantación de la herramienta cloud Quicksight aumento el porcentaje de datos útiles con una media del 47.30%, por lo tanto, se puede afirmar que Quicksight influyo positivamente en el aumento del porcentaje de datos útiles obtenidos de la simulación Smart Cities: Cañete plan piloto con medidores eléctricos inteligentes.

VI. Recomendaciones

Se sugiere aplicar la herramienta cloud Quicksight a la mayoría de procesos que tenga que ver con la recolección de datos y toma de decisiones, esto permitirá lograr mejoras con ayuda de la visualización de datos.

Para posteriores investigaciones se recomienda utilizar el indicador de porcentaje de lectura y visualización de datos, ya que este indicador permite saber que tan eficiente es la herramienta que se utilizara para la visualización de datos.

Se recomienda seguir investigando sobre nuevas herramientas cloud computing que ayuden en la mejora de procesos con datos, ya que nos ofrecen seguridad y rendimiento óptimo sin tener muchos recursos.

Por parte de lo observado, AWS es una plataforma que sigue en desarrollo, aun no se puede explorar al 100% la herramienta Quicksight, por tal motivo se recomienda seguir con las investigaciones sobre visualización de datos con otras herramientas que nos pueden ofrecer Microsoft Azure, Huawei Cloud o Google Cloud.

VII. Referencias Bibliográficas

Afinidad Eléctrica. (2008). *Telelectura de medidores de energía con tecnologías AMR y AMI*. Afinidad Eléctrica. http://test.afinidadelectrica.com/?p=645

Agüero, J. (2016). *Cloud, serverless y Big Data, protagonistas en AWS Summit 2016 - Paradigma*. Paradigmadigital.Com. https://www.paradigmadigital.com/dev/cloud-serverless-big-data-protagonistas-aws-summit-2016/

Amazon Inc. (2018). Amazon QuickSight: User Guide. *Amazon Web Services*, 564. https://docs.aws.amazon.com/quicksight/latest/user/amazon-quicksight-user.pdf

Barrera, A. (2018). *¿QUÉ ES EL PROCESAMIENTO DE DATOS?* Next_u. https://www.nextu.com/blog/que-es-el-procesamiento-de-datos/

Bayod, E. (2015). Ciudades Inteligentes: Definición y Nivel de CiberVulnerabilidad. *GESI (Grupo de Estudios En Seguridad Internacional)*, *12*, 1. http://dialnet.unirioja.es/servlet/articulo?codigo=5147076

Behar, D. (2008). Metodologia de la investigación. In *Ediciones Shalom*.

Bertenbreiter, S. (2015). *FDD/TDD/HDD y ¿Qué es TDD/HDD Split Frequency?* Laufquen. https://laufquen.com.ar/fddtddhdd-y-que-es-tddhdd-split-frequency/

Bouskela, M., Casseb, M., Bassi, S., De Luca, C., & Facchina, M. (2016). La ruta hacia las Smart Cities: Migrando de una gestión tradicional a la ciudad inteligente. *Banco Interamericano de Desarrollo (BID)*, 1–148.

Cárdenas Lino, C. A. (2016). Simulación y análisis de desempeño de la red LTE para infraestructura de medicion avanzada en smart grid. In *Universidad de Chile*. Universidad de Chile.

Carrasco, S. (2006). *Metodologia de la investigacion cientifica*.

Cluster tecnología, información y comunicación. (2011). *Smart City - Ciudad Inteligente*. Cluster Tecnología, Información y Comunicación. http://www.clustertic.co/site/DescubraelemClusteremTIC/Áreasestratégicas/SmartCityCiudadInteligente.aspx

ContentLab. (2019). *Smart Cities: un medio para alcanzar el desarrollo sostenible*. El Comercio Perú. https://elcomercio.pe/especial/perusostenible/planeta/smart-cities-medio-alcanzar-desarrollo-sostenible-noticia-1994641

DeStreaming. (2019). *¿QUE ES EL STREAMING?* DeStreaming.Es. https://www.destreaming.es/tutoriales/que-es-el-streaming/

EcuRed. (2019). *3GPP*. EcuRed. https://www.ecured.cu/3GPP

Espacioteca. (2017). *Las TIC que son y para que sirven*. Espacioteca. https://espacioteca.com/2017/03/26/las-tic-que-son-y-para-que-sirven/

Fundación Bankinter. (2015). Big Data El poder de los datos. *Future Trends Forum*, 82.

Gómez, V. A., Hernández, C., & Rivas, E. (2018). Visión General, Características y Funcionalidades de la Red Eléctrica Inteligente (Smart Grid). *Información Tecnológica*, *29*(2), 89–102. https://doi.org/10.4067/s0718-07642018000200089

Gónzales, M. (2019). Desarrollo dirigido por modelos para aplicaciones de datos en streaming. In *Universidad Autónoma de Madrid*.

Grupo Interplataformas de Ciudades Inteligentes. (2015). Smart Cities: documento de visión a 2030. *Gici*, *1*(June 2016), 13–127. https://doi.org/10.1007/978-3-319-39595-1

Hernández Leal, E. J., Duque Méndez, N. D., & Moreno Cadavid, J. (2017). Big Data: una exploración de investigaciones, tecnologías y casos de aplicación. *TecnoLógicas*, *20*(39), 15–38. https://doi.org/10.22430/22565337.685

Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, M. del P. (2014). Metodología de la Investigación. In *McGraw - Hill / Interamericana Editores S.A. de C.V.*

INTENSYS4EU. (2020). *ETIP SNET*. ETIP SNET. https://www.etip-snet.eu/about/etip-snet/

ITU. (2018). *Ciudades inteligentes y sostenibles - ITU PP-18 - Plenipotentiary*

Conference. Backgrounder. https://www.itu.int/web/pp-18/es/backgrounder/smart-sustainable-cities#ref1

Joyanes Aguilar, L. (2012). Computación en la Nube: Notas para una estrategia española en cloud computing. *Revista Del Instituto Español de Estudios Estratégicos*, *2*, 24. file:///C:/Users/Usuario/Downloads/Documat-ComputacionEnLaNube-4098278_1.pdf

Julià Minguillón, A. (2016). Introducción a la visualización de datos. *Universitat Oberta de Catalunya*, 38. http://hdl.handle.net/10609/57624

Kobus, C. B. A., Klaassen, E. A. M., Mugge, R., & Schoormans, J. P. L. (2015). A real-life assessment on the effect of smart appliances for shifting households' electricity demand. *Applied Energy*, *147*, 335–343. https://doi.org/10.1016/j.apenergy.2015.01.073

Levine, S. S., & Prietula, M. J. (2014). Open collaboration for innovation: Principles and performance. *Organization Science*, *25*(5), 1414–1433. https://doi.org/10.1287/orsc.2013.0872

López Vásquez, J. C. (2016). Estudio de la situación actual de las Smart Grids. In *Universidad de Cantabria*. Universidad de Cantabria.

Manovich, L. (2010). What is Visualization? *Information Visualization*, *May*, 23.

Medina, S. (2016). Diseño e implementación de un servicio en cloud computing para la visualizacion en linea de transporte escolar. In *Universidad Militar Nueva Granada, Bogotá D.C.* (Issue jul). https://doi.org/10.1017/CBO9781107415324.004

Muñoz, J. E. (2019). *Revista ElectroIndustria - Infraestructura de Medición Avanzada (AMI)*. Eléctro Industria. http://www.emb.cl/electroindustria/articulo.mvc?xid=3406&ni=infraestructura-de-medicion-avanzada-40ami41

Navas, M. (2014). *Técnica del Fichaje*. Calaméo. https://es.calameo.com/books/002591296093f8cec67d5

Pacheco, A., Reyes, O., Reyes, A., Espino, C., Cardenas, C., & Mendoza, E. (2020). Technological proposal of an LTE network simulator with smart meter.

Test Engineering and Management, *83*(13381), 13381–13387.

Palacios Frisancho, G. (2017). Diseño e implementacion de un sistema para el monitoreo de cultivos nativos utilizando internet del todo y redes Fog. In *Universidad Nacional del Altiplano*. Universidad Nacional del Altiplano.

Peralta, E. (2019). Exploración visual basada en similitud de grandes conjuntos de datos multidimensionales georreferenciados. In *Universidad Católica San Pablo*.

Pereira Timarán, S. R., Arteaga Hernández, I., Caicedo Zambrano, S. J., Hidalgo Troya, A., & Alvarado Pérez, J. C. (2016). El proceso de descubrimiento de conocimiento en bases de datos. *Ediciones Universidad Cooperativa de Colombia*, *2016*, 63–86. https://doi.org/dx.doi.org/10.16925/9789587600490

PowerData. (2014). *Calidad de Datos. Cómo impulsar tu negocio con los datos.* Power Data. https://www.powerdata.es/calidad-de-datos

Red.es. (2016). Visualización de datos: definición, tecnologías y herramientas. *Iniciativa Aporta*, *nov*, 51. http://datos.gob.es/sites/default/files/doc/file/informe_herramientas_visualizacion.pdf

Riquelme, M. (2019). *¿Qué es y cómo se interpreta el coeficiente de correlación de Pearson?* Web y Empresas. https://www.webyempresas.com/coeficiente-de-correlacion-de-pearson/

Rodenbeck, E. (2008). 17-10-2008: Presentación de Eric Rodenbeck, “La visualización de información es un medio” - InfoVis: Wiki. *O’Reilly Emerging Technology Conference*. https://infovis-wiki.net/wiki/2008-10-17:_Presentation_of_Eric_Rodenbeck,_%22information_Visualization_Is_A_Medium%22

Rodriguez, L. (2017). “Metamodelo para la integración de datos abiertos aplicado a Inteligencia de Negocios.” In *Universidad de Oviedo*. http://di002.edv.uniovi.es/~cueva/investigacion/tesis/Tesis-LuzAndrea.pdf

Rouse, M. (2012). *Análisis de datos*. TechTarget. https://searchdatacenter.techtarget.com/es/definicion/Analisis-de-Datos

S. Pickers. (2018). *¿Cómo determinar el tamaño de una muestra?* Psyma. https://www.psyma.com/company/news/message/como-determinar-el-tamano-de-una-muestra

Salazar, J. (2017). Implementación de inteligencia de negocios para el área comercial de la empresa Azaleia - Basado en metodología ágil scrun. In *Universidad San Ignacio de Loyola*.

Saporíti, I., & Agustín, J. (2017). Minería de opiniones y visualización de datos aplicables a estudios de mercado. *Universidad Nacional de La Plata*, 1–135. http://sedici.unlp.edu.ar/bitstream/handle/10915/61531/Documento_completo. y Tibaldo, A. Minería de opiniones.pdf-PDFA.pdf?sequence=1

Signaturit. (2017). *¿Qué es Business Intelligence (BI) y qué herramientas existen?* Signaturit. https://blog.signaturit.com/es/que-es-business-intelligence-bi-y-que-herramientas-existen

SoftwareLab Blog. (2018). *¿ Qué es LTE ? Diferencias entre LTE y 4G*. SoftwareLab. https://softwarelab.org/es/lte-4g/

Structuralia Blog. (2018). *El papel de los Smart Meters o contadores inteligentes en el ahorro energético*. Structuralia. https://blog.structuralia.com/el-papel-de-los-smart-meters-o-contadores-inteligentes-en-el-ahorro-energetico

Valderrama, S. (2015). Pasos Para Elaborar Proyectos de Investigacion Cientifica. In *Editorial San Marcos* (p. 495).

Valero Sancho, J. L. (2014). La visualización de datos. *Ámbitos. Revista Internacional de Comunicación*, *jul*, 25. https://www.redalyc.org/articulo.oa?id=168/16832256009

Villanueva, J. (2015). Solución De Business Intelligence Utilizando Tecnología Saas. Caso: Área De Proyectos En Empresa Bancaria – Perú. In *Universidad Nacional de Piura*. https://pirhua.udep.edu.pe/handle/11042/2790

VIII. Anexos

IMPLANTACIÓN DE LA HERRAMIENTA CLOUD QUICKSIGHT PARA LA VISUALIZACIÓN DE DATOS OBTENIDOS DE LA SIMULACIÓN DEL PROYECTO SMART CITIES: CAÑETE PLAN PILOTO CON MEDIDORES ELÉCTRICOS INTELIGENTES

<table>
<tr><th>PROBLEMA</th><th>OBJETIVOS</th><th>HIPÓTESIS</th><th>VARIABLES</th><th>DIMENSIONES</th><th>INDICADORES</th></tr>
<tr>
<td rowspan="3">PROBLEMA GENERAL
¿Cómo influye la herramienta cloud Quicksight en la visualización de datos obtenidos de la simulación Smart Cities: Cañete plan piloto con medidores eléctricos inteligentes?

PROBLEMAS ESPECÍFICOS
✓ ¿Cómo influye la herramienta cloud Quicksight en el porcentaje de datos de calidad obtenidos de la simulación Smart Cities: Cañete plan piloto con medidores eléctricos inteligentes?
✓ ¿Cómo influye la herramienta cloud Quicksight en el porcentaje de lectura y visualización de datos obtenidos de la simulación Smart Cities: Cañete plan piloto con medidores eléctrico inteligentes?
✓ ¿Cómo influye la herramienta cloud Quicksight en el porcentaje de datos útiles obtenidos de la simulación Smart Cities: Cañete plan piloto con medidores eléctricos inteligentes?</td>
<td rowspan="3">OBJETIVO GENERAL
Implantar la herramienta cloud Quicksight para la visualización de datos obtenidos de la simulación Smart Cities: Cañete plan piloto con medidores eléctricos inteligentes.

OBJETIVOS ESPECÍFICOS
✓ Determinar cómo influye la herramienta cloud Quicksight en el porcentaje de datos de calidad obtenidos de la simulación Smart Cities: Cañete plan piloto con medidores eléctricos inteligentes.
✓ Determinar cómo influye la herramienta cloud Quicksight en el porcentaje de lectura y visualización de datos obtenidos de la simulación Smart Cities: Cañete plan piloto con medidores eléctricos inteligentes.
✓ Determinar cómo influye la herramienta cloud Quicksight en el porcentaje de datos útiles obtenidos de la simulación Smart Cities: Cañete plan piloto con medidores eléctricos inteligentes.</td>
<td rowspan="3">HIPÓTESIS GENERAL
La implantación de la herramienta cloud Quicksight permitirá visualizar de forma eficiente los datos obtenidos de la simulación Smart Cities: Cañete plan piloto con medidores eléctricos inteligentes.

HIPÓTESIS ESPECÍFICAS
✓ La herramienta cloud Quicksight aumentará el porcentaje datos de calidad obtenidos de la simulación Smart Cities: Cañete plan piloto con medidores eléctricos inteligentes.
✓ La herramienta cloud Quicksight aumentará el porcentaje de lectura y visualización de datos obtenidos de la simulación Smart Cities: Cañete plan piloto con medidores eléctricos inteligentes.
✓ La herramienta cloud Quicksight aumentará el porcentaje de datos útiles obtenidos de la simulación Smart Cities: Cañete plan piloto con medidores eléctricos inteligentes.</td>
<td>VARIABLE INDEPENDIENTE:
Herramienta cloud Quicksight.</td>
<td>No aplicable.</td>
<td>No aplicable.</td>
</tr>
<tr>
<td rowspan="2">VARIABLE DEPENDIENTE:
Visualización de datos.</td>
<td>Calidad de datos.</td>
<td>Porcentaje de datos de calidad obtenidos.</td>
</tr>
<tr>
<td>Interpretación de datos.</td>
<td>Porcentaje de lectura y visualización de datos.</td>
</tr>
<tr>
<td colspan="3"></td>
<td>Análisis de datos.</td>
<td>Porcentaje de datos útiles.</td>
</tr>
<tr>
<td>MÉTODO
Hipotético - Deductivo</td>
<td>TIPO DE INVESTIGACIÓN
Experimental – Aplicada.</td>
<td>NIVEL DE INVESTIGACIÓN
Explicativa.</td>
<td colspan="2">DISEÑO DE INVESTIGACIÓN
Pre- Experimental.</td>
<td>TECNICAS E INSTRUMENTOS
Fichaje – Ficha de registro.</td>
</tr>
</table>

Tabla 15. Matriz de consistencia.

Presupuesto de la investigación

Concepto	Cantidad	Costo Unitario S/.	Total S/.
Hardware			
Laptop Lenovo V330-15IKB Core i7, 8Gb RAM, 512Gb SSD	1	S/. 6065.00	S/. 6065.00
Microsoft Kit Desktop 1576 1113	1	S/. 95.00	S/. 95.00
Monitor LG 27UL500 27''	1	S/. 773.50	S/. 773.50
Memoria USB Kingston 64Gb	1	S/. 90.00	S/. 90.00
Impresora Multifuncional EPSON L575	1	S/. 957.80	S/. 957.80
Access Point D-Link DIR-615	1	S/. 80.00	S/. 80.00
Software			
Cuenta de AWS	1	S/. 20.00	S/. 20.00
Amazon Quicksight Edición Standard*(1)	1	-	-
Amazon S3 Edición Standard*(2)	1	-	-
Microsoft Office Licencia Digital Permanente	1	S/. 22.00	S/. 22.00
IBM SPSS Statistics 25*(3)	1	-	-
Navegador Web Google Chrome	1	-	-
Bloc de Notas	1	-	-
Materiales			
Paquetes de hojas bond	2	S/. 15.00	S/. 30.00
Kit de útiles de escritorio	1	S/. 618.71	S/. 618.71
Kit de muebles de oficina	1	S/. 1547.50	S/. 1547.50
Cursos Especializados			
Big Data - Utec	1	S/. 2200.00	S/. 2200.00
AWS - Egroject	1	S/. 6000.00	S/. 6000.00
		Total:	**S/. 18499.51**

Tabla 16. Presupuesto de investigación.

*(1) Se utilizó la capa gratuita con un max. de 1Gb de almacenamiento.

*(2) Se utilizó la capa gratuita con un max. de 5Gb de almacenamiento.

*(3) Se utilizó la versión de prueba gratuita de 14 días.

Cronograma de la investigación

No.	ACTIVIDAD	DURACION (días)	FECHA INICIAL	FECHA FINAL
	Preparación de la investigación			
1	Planteamiento del problema y Marco teórico	10	07/01/2019	18/01/2019
2	Viabilidad, Plan de tesis, Análisis de la información	5	21/01/2019	25/01/2019
3	Curso de especialización Utec	6	05/02/2019	21/02/2019
	Acondicionamiento de instalaciones			
4	Evaluar las condiciones del espacio físico del local	1	04/03/2019	04/03/2019
5	Adecuación del local	10	05/03/2019	18/03/2019
6	Instalación de software para desarrollo de tesis	2	25/03/2019	26/03/2019
7	Instalación en ambiente virtual AWS (S3, Quicksight)	2	27/03/2019	28/03/2019
	Ejecución, seguimiento y control			
8	Diseño de pruebas con datos de simulación	60	01/04/2019	31/05/2019
9	Realización de las pruebas con datos de simulación	60	03/06/2019	31/07/2019
10	Curso especialización Egroject y Presentación en UNSA	5	05/08/2019	30/08/2019
11	Análisis de los resultados	60	02/09/2019	31/10/2019
12	Primer informe de Tesis y Ponencia Escuela Naval	30	01/11/2019	29/11/2019
13	Nuevas Pruebas	60	06/01/2020	28/02/2020
14	Segundo informe de Tesis	30	02/03/2020	31/03/2020
	Cierre			
15	Corrección de errores	30	01/04/2020	30/04/2020
16	Pruebas con correcciones	30	04/05/2020	23/05/2020
17	Presentación y revisión de Tesis	2	01/06/2020	03/07/2020
18	Aprobación de Tesis	30	06/07/2020	31/07/2020
19	Sustentación de Tesis	30	03/08/2020	28/08/2020

Tabla 17. Cronograma de investigación.

Modelo de ficha de registro

FICHA DE REGISTRO PRE TEST Y POST TEST

INSTRUCCIONES:

Marca con un aspa (X) una de las alternativas según su criterio que crea usted conveniente.

1	2	3	4	5
Insatisfecho	Poco satisfecho	Indiferente	Satisfecho	Muy satisfecho

DIMENSIÓN: CALIDAD DE DATOS	1	2	3	4	5
1. ¿Cómo considera el la estructura de los datos obtenidos de la simulación?					
2. ¿Cómo considera usted la transformación de dichos datos en insight's o tablas?					
3. ¿Cómo considera usted la disponibilidad de la información?					
4. ¿Cómo considera usted la integridad de los datos?					
DIMENSIÓN: INTERPRETACION DE DATOS					
5. ¿Cómo considera la lectura de datos?					
6. ¿Cómo considera la visualización de información?					
7. ¿Cómo considera el tiempo que demanda la lectura de datos?					
8. ¿Cómo considera el diseño de visualización de datos?					
DIMENSIÓN: ANALISIS DE DATOS					
9. ¿Cómo considera la toma de decisiones con los resultados de datos útiles?					
10. ¿Cómo considera el análisis de datos para obtener datos útiles?					

Resultados de instrumento ficha de registro

Pre-test

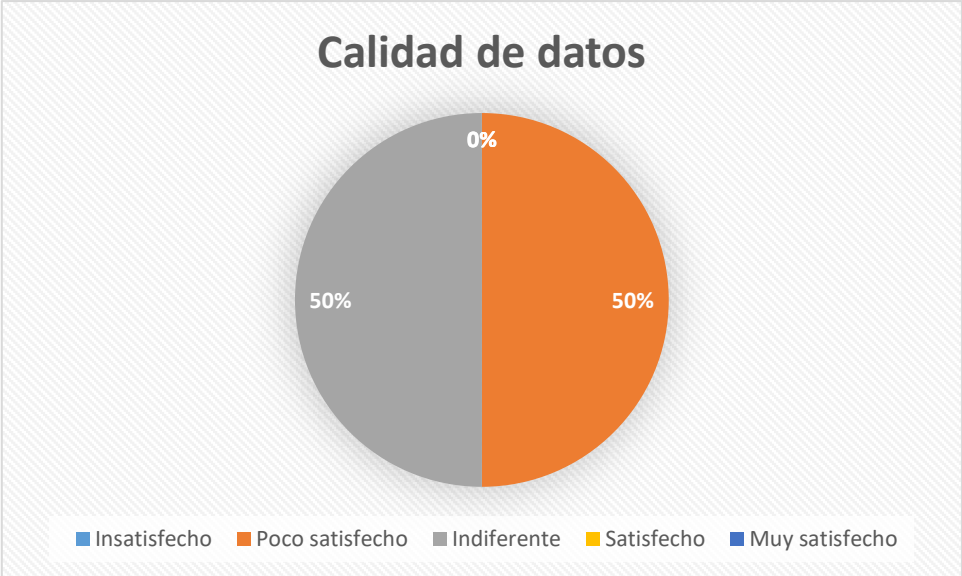


Figura 15: Fichaje pre calidad de datos. Fuente: Elaboración propia.

Después de aplicar el instrumento, observamos los resultados en base a la escala de Likert, tienen un 50% de poco satisfecho y 50% de indiferente, lo que indica que los datos no son de calidad y que no sirven para una toma de decisiones futuras.

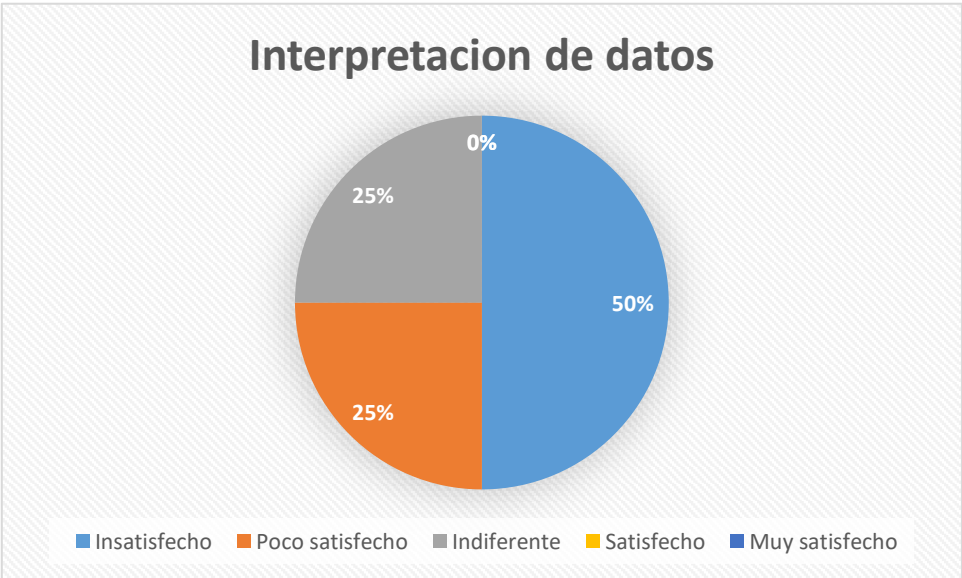


Figura 16: Fichaje pre interpretación de datos. Fuente: Elaboración propia.

Para la dimensión de interpretación de datos, los resultados fueron: un 50% de insatisfecho, 25% de poco satisfecho y 25% de indiferente, lo que indica que los datos necesitan una herramienta para visualizar y analizar.

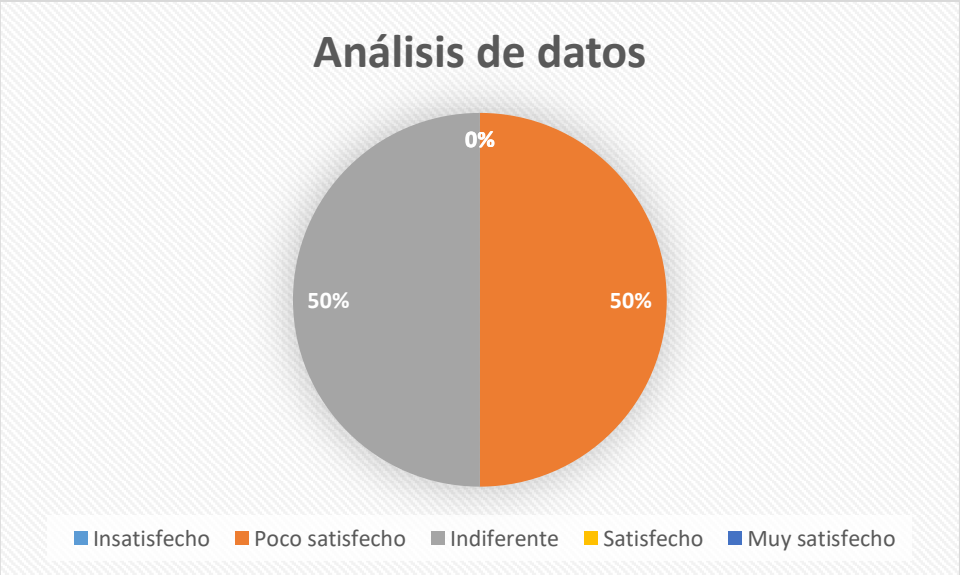


Figura 17: Fichaje pre análisis de datos. Fuente: Elaboración propia.

Para la dimensión de análisis de datos, los resultados fueron: 50% de poco satisfecho y 50% de indiferente. Lo cual nos dice que estos datos no son útiles para una toma de decisiones porque no están analizados ni visualizados.

Post-test

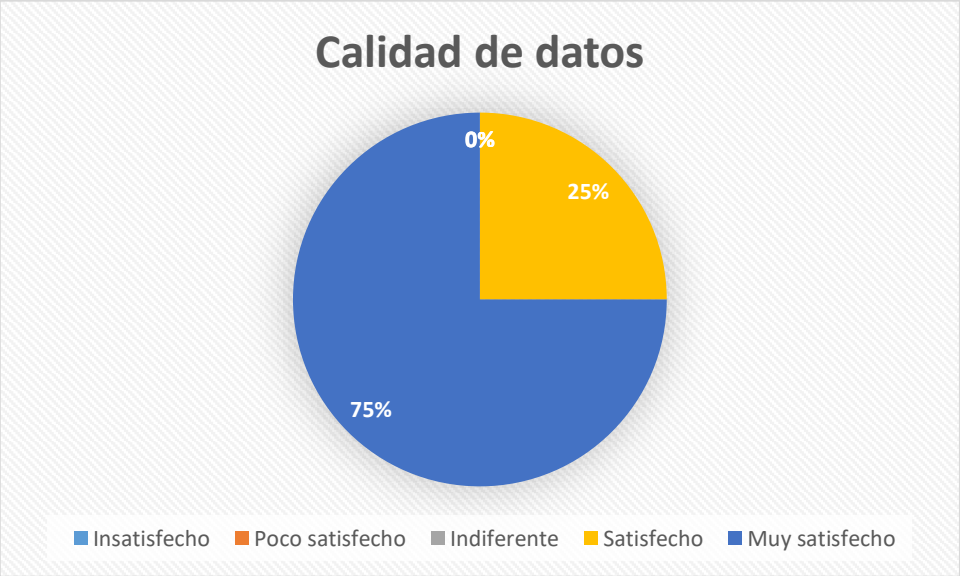


Figura 18: Fichaje post calidad de datos. Fuente: Elaboración propia.

Luego de aplicar la solución, se aplicó un post-test donde observamos: 25% de satisfecho y 75% de muy satisfecho, lo que indica que gracias a la herramienta aplicada los datos tienen buena calidad y están listos para llevarlo a análisis y visualización.

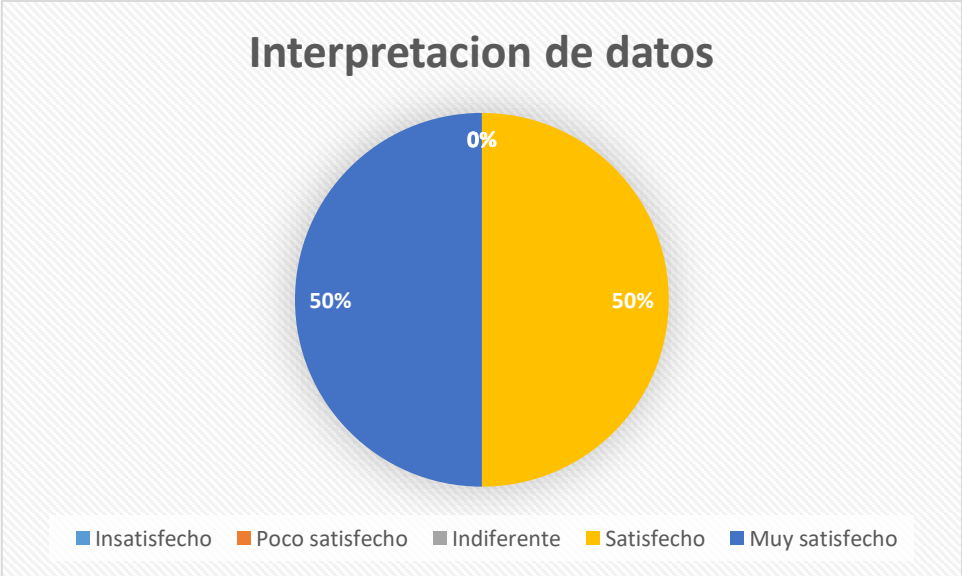


Figura 19: Fichaje post interpretación de datos. Fuente: Elaboración propia.

Para la dimensión de interpretación de datos, se observa: 50% de satisfecho y 50% de muy satisfecho, lo que indica que los datos tienen buena lectura y se interpretan mejor en la visualización con insight's.

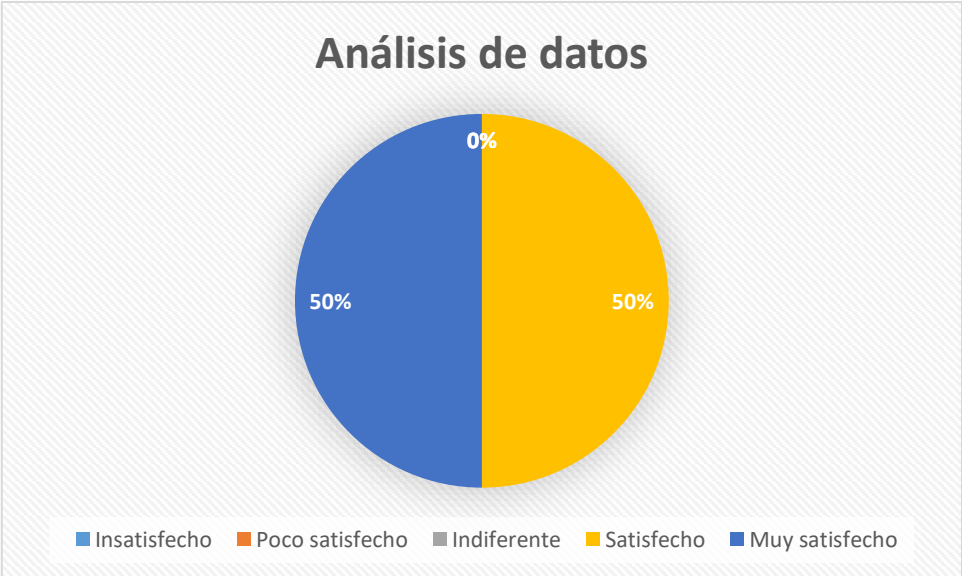


Figura 20: Fichaje post análisis de datos. Fuente: Elaboración propia.

Para la dimensión de análisis de datos tenemos: 50% de satisfecho y 50% de muy satisfecho, lo que indica que se pueden tomar buenas decisiones con los insight's generados.

Ejemplo de visualización con datos de simulación

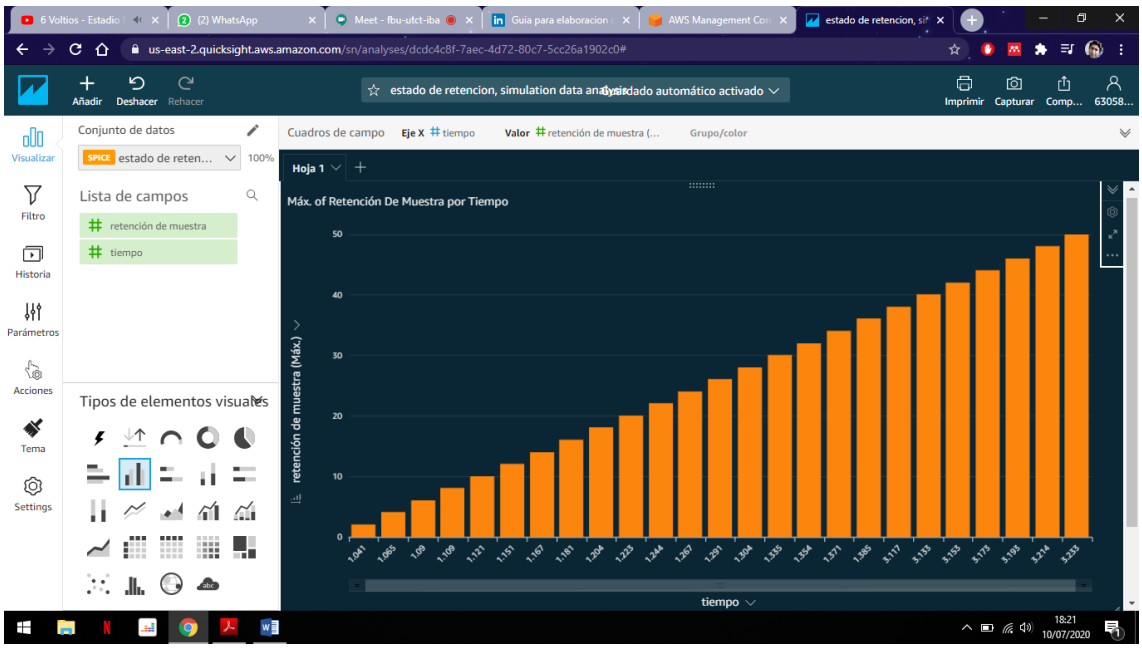


Figura 21: Retención de muestras vector. Fuente: Elaboración propia.

En este ejemplo se observa el tiempo de retención de cada muestra de vector, en este caso fueron 50 y están ordenamos de menor a mayor tiempo.

Fotografías









UNIVERSIDAD NACIONAL DE CAÑETE

Código: F-M01.03.04-GT-003 / Rev. 1 Fecha de Aprobación: 11-11-2019	Revisión: 01

ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS

ACTA DE SUSTENTACIÓN DE TESIS PARA OPTAR EL TÍTULO PROFESIONAL DE INGENIERO DE SISTEMAS

En la ciudad de Cañete, en la Escuela Profesional de Ingeniería de Sistema .de la Facultad de Ingeniería, siendo las 12:00 h horas, del día 26 de agosto de 2021, se reunió a través de una reunión meet: https://meet.google.com/edg-thih-fca, el Jurado conformado por:

Presidente(a) : Mg. Amanda Durán Carhuamaca
Secretario (a) : M.Sc. Edwin Roque Tito.
Vocal : Mg. Marisol Daga Chaca.

Para el acto de la sustentación virtual de la tesis intitulada: "**Implantación de la herramienta Cloud Quicksight para la visualización de datos obtenidos del Proyecto Smart Cities: Cañete Plan Piloto con Medidores Eléctricos Inteligentes**".
para optar al Título Profesional de Ingeniero de Sistemas, por el bachiller

Beni Alexis Luyo Huamán

Después de concluido el acto de sustentación y luego de que el mencionado dio respuesta a las preguntas respectivas, el Jurado Evaluador, declara:

1. ( ) Aprobado, con mención honrosa. La cual amerita su publicación.
2. (**X**) Aprobado, por unanimidad
3. ( ) Aprobado, por mayoría
4. ( ) Desaprobado

Con una calificación promedio de: 16 (dieciséis)

Por tanto, el Bachiller se encuentra expedito (**X**), impedido ( ), para realizar los trámites correspondientes para la obtención del Título Profesional de: Ingeniero de Sistemas

Siendo las 13:00 Horas se dio por terminado el acto de sustentación.

FIRMA DIGITAL Firmado digitalmente por: DURAN CARHUAMACA Amanda FAU 20491363402 soft Motivo: Soy el autor del documento Fecha: 31/08/2021 09:07:08-0500	FIRMA DIGITAL Firmado digitalmente por: ROQUE TITO Edwin FAU 20491363402 soft Motivo: Soy el autor del documento Fecha: 26/08/2021 14:36:59-0500	FIRMA DIGITAL Firmado digitalmente por: DAGA CHACA Marisol FAU 20491363402 soft Motivo: Soy el autor del documento Fecha: 26/08/2021 14:45:11-0500
Presidente de Jurado	Secretario	Vocal