Universidad Nacional de Misiones. Facultad de Ciencias Exactas, Químicas y Naturales. Secretaría de Investigación y Postgrado. Maestría en Tecnología de la Información

Maestrando
Ing. Eduardo Hugo Bennesch

Modelo de Location Business Analytics en Cloud para pymes del sector de comercialización de insumos y herramientas para la construcción del NEA argentino

Trabajo final de Maestría presentada para obtener el título de "Magíster en Tecnología de la Información"

"Este documento es resultado del financiamiento otorgado por el Estado Nacional, por lo tanto, queda sujeto al complimiento de la Ley N° 26.899".

Director
Dr. Diego Godoy

Co- Director
Mgter. Karina Eckert

Mayo 2023

Esta obra está licenciado bajo Licencia Creative Commons (CC) Atribución-NoComercial-CompartirIgual 4.0 Internacional. https://creativecommons.org/licenses/by-nc-sa/4.0/

rid.unam.edu.ar

UNIVERSIDAD NACIONAL DE MISIONES

Facultad de Ciencias Exactas, Químicas y Naturales

Maestría en Tecnologías de la Información

Trabajo Final de Maestría

Modelo de Location Business Analytics en Cloud para pymes del sector de comercialización de insumos y herramientas para la construcción del NEA Argentino.

Directores

Dr. Diego Godoy (Director)

Mgter. Karina Eckert. (Co-Director)

Autor

Ing. Eduardo Hugo Bennesch

Mayo 2023

Dedicatoria

A mi familia, principalmente a Laura por ser siempre mi norte.

Resumen

Lo único constante es el cambio y las organizaciones lo saben. Por eso están presionadas siempre a desarrollarse en la búsqueda de una adaptación continua; como resultado se encuentran tomando decisiones teniendo en cuenta distintos factores.

En las Pequeñas y Medianas Empresas (PyMEs) es común que la toma de decisiones se realice con información incompleta y basadas en estimaciones personales. En Argentina, estas empresas representan el 96% de las empresas del país generando el 60% del empleo creado por el sector privado.

La industria de la construcción en el Noroeste Argentino (NEA) es una de las principales fuentes de empleo y crecimiento económico en la región. Sin embargo, las pequeñas y medianas empresas (PyMEs) que se dedican a la comercialización de insumos y herramientas para la construcción enfrentan numerosos desafíos como la falta de recursos y la competencia de grandes empresas.

Para abordar estos desafíos, se propone un modelo de *Location Business Analytics* en *Cloud* para PYMEs del sector de comercialización de insumos y herramientas para la construcción del NEA; este modelo aprovechará la tecnología de almacenamiento de datos geográficos en la nube para brindar a las PYMEs una herramienta poderosa y accesible para analizar y mejorar sus operaciones comerciales y de marketing.

Como propuesta de solución a ello se realizó una revisión de los adelantos tanto en materia teórica como practica de las soluciones aplicadas al sector de provisión de materiales, se desarrolló un modelo de explotación de datos geográficos combinados con datos de negocio de una pyme Ficticia cuyo *dataset* se descargó vía web; estos datos fueron cruzados y analizados para luego volcarlos en dos algoritmos de Redes Neuronales y Árbol de Decisión desarrollados en R (*NeuralNetTools* y *Rpart*); la aplicación de estos algoritmos demostraron alto grado de precisión en la devolución de resultados (99,98% para *NeuralNetTools* y 99,63% para *Rpart*) estableciendo una variable de rendimiento en términos de distancia, costos y ganancias por ventas que demostró ser de utilidad para analizar operaciones con clientes.

Para la elección de la API que fue utilizada en el trabajo se desarrolló un estudio comparativo de 5 plataformas de geolocalización; se concluyó que serían más adecuadas las plataformas de *Crowdsourcing* o edición colaborativa de los datos en esta región del mundo resultando el mejor motor de geocodificación *OpenCageData* con una media de error inferior 0,0285kms aplicando Semiverseno y 0,0553kms aplicando Distancia Euclidiana.

Finalmente se realizó el diseño, desarrollo e implementación de un tablero de inteligencia de negocios basado en la geo localización utilizando los datos anteriormente analizados que puede proporcionar a las PYMEs del sector de la construcción en el NEA una solución práctica y efectiva para mejorar su desempeño y competitividad en el mercado. La validación del modelo propuesto fue realizada mediante opinión experta con un puntaje general de 4.48 de 5 puntos (aceptable / muy aceptable).

Este trabajo concluye que el modelo propuesto es viable económicamente, permite aumentar la competitividad y el análisis profundo de variables de negocios que puede ser implementado en el sector estudiado en el corto y mediano plazo sin dificultades.

Palabras Clave

Aprendizaje de Máquina, NEA, PyME, Inteligencia de Localización, Json.

Abstract

The only thing that is constant is change and organizations know this. That is why they are always under pressure to develop in the search for continuous adaptation, as a result of which they find themselves making decisions taking into account different factors.

In Small and Medium Enterprises (SMEs) it is common for decision making to be made with incomplete information and based on personal estimates. In Argentina, these companies represent 96% of the companies in the country and generate 60% of the employment created by the private sector.

The construction industry in Northwest Argentina (NEA) is one of the main sources of employment and economic growth in the region. However, small and medium-sized enterprises (SMEs) engaged in the commercialization of construction inputs and tools face numerous challenges, such as lack of resources and competition from large companies.

To address these challenges, a Location Business Analytics in Cloud model is proposed for SMEs in the NEA construction supplies and tools marketing sector, this model will leverage geographic data storage technology in the cloud to provide SMEs with a powerful and accessible tool to analyze and improve their business and marketing operations.

As a proposed solution to this, a review of the advances in both theoretical and practical solutions applied to the materials supply sector was carried out, and a model was developed to exploit geographic data combined with business data from a fictitious SME whose dataset was downloaded via the web, These data were cross-referenced and analyzed in order to later dump them into two Neural Network and Decision Tree algorithms developed in R (NeuralNetTools and Rpart). The application of these algorithms showed a high degree of accuracy in the return of results (99.98% for NeuralNetTools and 99.63% for Rpart), establishing a performance variable in terms of distance, costs and sales profits that proved to be useful for analyzing operations with customers.

For the selection of the API used in the work, a comparative study of 5 geolocation platforms was carried out, resulting in the most appropriate Crowdsourcing platforms or collaborative editing of data in this region of the world, with the best geocoding engine being OpenCageData, with an average error of 0.0285kms applying Semiverse and 0.0553kms applying Euclidean Distance.

Finally, the design, development and implementation of a geo-location based business intelligence dashboard based on the data analyzed above was carried out, which can provide SMEs in the construction sector in the NEA with a practical and effective solution to improve their performance and competitiveness in the market. The validation of the proposed model and the developed dashboard was validated by expert opinion with an overall rating of 4.48 out of 5 points (acceptable / very acceptable).

This work concludes that the proposed model is economically viable, allows increasing competitiveness and deep analysis of business variables that can be implemented in the studied sector in the short and medium term without difficulties.

Keywords

Machine Learning, NEA, Pyme, Location intelligence, Json

Reconocimientos

A mis directores de Tesis por el incondicional apoyo durante este largo tiempo de superación y aprendizaje.

Índice

Capítulo 1 10

1. Introducción 10

1.1. Marco Metodológico 11

1.2. Planteamiento del Problema 11

1.3. Metodología 12

Capítulo 2 16

2. Marco Teórico 16

2.1. Business Intelligence 16

2.2. Sistemas de Información Geográfica 16

2.3. Location Intelligence **¡Error! Marcador no definido.**

2.4. Recopilación de Requisitos en proyectos BI 20

2.5. Inteligencia Artificial 20

2.6. Aprendizaje de máquina 20

2.7. CRISP-DM 22

2.8. Modelado de Proceso de Negocio (BPMo) 23

2.9. Notación y Modelado de Procesos de Negocio (BPMN) 24

2.10. Power BI 25

Capítulo 3 27

3. Estado del Arte 27

3.1. Introducción 27

3.2. Trabajos Relacionados 27

3.3. Objetivo y Preguntas de Investigación 27

3.4. Reporte de Resultados 28

3.5. Resumen de los hallazgos 33

Capítulo 4 35

4. Análisis y diseño del proyecto 35

4.1. Requerimientos: 35

4.2. Recopilación **¡Error! Marcador no definido.**

Capítulo 5 46

5. Implementación 46

5.1. Contexto y motivación 46

5.2. Trabajo de Campo 46

5.3. Ejecución del proyecto 47

5.4. Pruebas de precisión 47

5.5. Resultados 50
5.6. Discusión 52
5.7. Análisis de los resultados 53
5.8. Sumario de apreciaciones: 53
5.9. Evaluación y Despliegue. 55
Capítulo 6 60
6. Resultados 60
6.1. Pruebas de Geocodificación 60
6.2. Predicciones utilizando aprendizaje de máquina 60
6.3. Evaluación Encuestas 60
6.4. Implementabilidad del modelo 64
Capítulo 7 66
7.1. Conclusiones 66
7.2. Trabajos Futuros 67

Lista de Tablas

Tabla 2: Palabras Clave 27
Tabla 3: resultados de la búsqueda por fuente 28
Tabla 4: Variación en Kms. 51
Tabla 5: Exactitud de geolocalización. 51
Tabla 6: Variación en Kms, Distancia euclidiana. 52
Tabla 7: Exactitud de geolocalización, Distancia euclidiana. 52
Tabla 8: *Preguntas.* 61
Tabla 9: *Escala.* 61
Tabla 10: *Cargos de los encuestados.* 61
Tabla 11: Resultados. 64
Tabla 12: Preguntas de investigación 71

Lista de Figuras

Figura 1: variación interanual ventas minoristas Pyme[2]. 11
Figura 2: Variación interanual venta minorista país sectorizado [2], 12
Figura 3: Diagrama Metodológico. 13

Figura 4: Capas de información en un SIG [15] 17
Figura 5: Modelo de GeoMarketing [14]. 18
Figura 6 : Modelo de procesos [15]. 20
Figura 7 : Ciclo de Vida CRISP-DM [7]. 23
Figura 8: Centro de competencias de inteligencia de negocios [25]. 25
Figura 9: Fases para definir un modelo de negocio [30]. 29
Figura 10: Estructura del mapa lógico [18]. 30
Figura 11: Diagrama de clases BD tabular [18]. 31
Figura 12: Spatial Data Mining [17]. 32
Figura 13: Diagrama BPMN proceso de venta. 37
Figura 14: Mapa de zonas, Ciudad de Posadas [32]. 39
Figura 15: Código de Función llamada a api. 42
Figura 16: Modelo de datos en Power Bi. 43
Figura 17: Arquitectura del Tablero FerreteriaBi. 44
Figura 18: fragmento del código Json para el mapa de capas. 45
Figura 19: Comparativa de Visualización de mapas. 47
Figura 20: Ejemplo Rest API Json. 48
Figura 21: Ejemplo Rest API xml. 48
Figura 22: Error de Nominatim al resolver dirección. 49
Figura 23: Dirección Geolocalizada correctamente. 49
Figura 24: Comparativa – real contra puntos geocodificados. 54
Figura 25: *Página 1 del tablero.* 57
Figura 26: *Página 2 del tablero.* 59
Figura 27: Respuestas RQ1. 62
Figura 28: Respuestas RQ2. 62
Figura 29: Respuestas RQ3. 63
Figura 30: Respuestas RQ4. 63
Figura 31: Respuestas RQ5. 64

Capítulo 1

1. Introducción

El 44% de los empresarios Argentinos toma sus decisiones en base a su intuición [1]**…**

Los avances en las Tecnologías de la Información y de la Comunicación (TIC) están proporcionando más formas alternativas de acceder a la información y a los datos de almacenamiento, así como una mayor accesibilidad a múltiples dispositivos, transformando el Internet de las cosas en un ecosistema digital gigante, revolucionando todos los modelos de negocios [1]; de hecho según Gartner[2] se espera que los ingresos globales en el mercado de software de análisis e inteligencia empresarial (*Business Intelligence*) crezcan de U$$17.05 mil millones en 2017 a U$$29.48 mil millones para 2022[3].

Los constantes desarrollos en tecnología impulsados por la competencia obligan ininterrumpidamente a las organizaciones a realizar cambios constantes.

Según la encuesta realizada en 2018 por la Universidad Siglo 21, antes de tomar una decisión empresarial el 44% no alcanza a realizar una búsqueda profunda de información; el 40% no logra realizar un análisis de la información obtenida y el 39% no usa criterios analíticos rigurosos al momento de optar por una opción[1].

Principales dificultades que afrontan las PyMEs del NEA:

- El nivel de accesibilidad logística es menor que el nacional, con un 11% de empresas con baja accesibilidad contra un 4% de la media nacional[4].
- Los costos de flete terrestre son elevados, no existe infraestructura ferroviaria y los ríos no están convenientemente explotados[5].
- Excesivo cortoplacismo en las operaciones comerciales[1]. .
- Falta de inversión: Con respecto a la tecnología y recursos informáticos la proporción de PyMEs que utilizan estos recursos es baja con un promedio de 12.5%[4].
- Gastos de comercialización: En las empresas de esta zona del país, los gastos de comercialización representan un concepto altamente incidente en la estructura de costos[5].
- Estructura de costos compleja y alta presión tributaria[5].

Todo ello evidencia la necesidad de contar con soluciones que ayuden a mejorar la rentabilidad generando un mayor rendimiento comercial y principalmente lograr eficiencia en los procesos logísticos que inciden fuertemente en la estructura de costos.

También se evidencia que la potencialidad de la ciencia del análisis de datos está prácticamente subexplotada por las PyMEs del NEA, entonces es necesario y urgente proporcionar un cambio de cultura para poder ayudar a transitar este camino junto con la transformación digital.

El desarrollo de aplicaciones de Inteligencia de negocio (por sus siglas en Ingles de Bussines Intelligence, BI) con un enfoque a GIS (Siglas en inglés para sistema de información geográfica), respaldadas por bases de datos geoespaciales son relevantes para el bienestar organizacional y ayudan al proceso de toma de decisiones en la planificación, implementación y monitoreo de acciones estratégicas [8].

En plena era de la información está de más aclarar que ésta es un activo importante y las PyMEs que ya han madurado están llenas de datos. Hasta ahora esos datos eran solo para realizar un control de lo que se ha trabajado, y la creación de un modelo de *Location Business Analyt*ics podrá facilitar la transformación de esta información en variables de análisis para proporcionar soporte a la toma de decisiones fundada en indicadores de precisión.

¿Qué mejoras puede ofrecer este proyecto para las empresas?

- Respuestas a preguntas de negocio críticas en tiempo real [9].
- Generación de perfiles de clientes más precisos [10].
- Identificar puntos críticos para la reducción de costos. [9]
- Posibilidad de representar datos en un formato que es más fácil de entender. [10]
- Ampliación del campo de visión que afecta la toma de decisiones. [9]
- Mejora en el acceso de datos para su representación geográfica.
- Transformación de sistemas GIS en completos tableros de mando. [11]
- Integración de datos privados con datos públicos (Producción, Impositiva, Demográfica, Información de tendencias en redes sociales) [11].

1.1. Objetivo general.

Desarrollar una plataforma para la toma de decisiones en el ámbito comercial para el caso de estudio de una empresa de insumos de construcción

1.2. Objetivos específicos.

- Definir el estado del arte de las arquitecturas/sistemas basadas en *Location Business Analytics.*
- Recopilar los requisitos que tendrá la arquitectura propuesta para el caso de estudio.
- Diseñar una arquitectura para integrar datos de Gestión comercial con información geoespacial.
- Implementar un modelo de *machine learning* para analizar el entrecruzamiento de datos geoespaciales y la información de negocio.
- Desarrollar un tablero en el que se visualice la información y los distintos indicadores resultantes de los requerimientos del negocio.
- Evaluar por medio de encuestas con Expertos y Tomadores de decisiones la utilidad de la arquitectura.

1.3. Marco Metodológico

1.3.1. Planteamiento del Problema

Según el informe de la CAME[4] de Mayo 2022, el sector "***Ferretería, mat. Electicos y mat. Para la construcción***" equivalen al 9.3% de las ventas minoristas de nuestro país y se presenta una disminución del 8.1% respecto a abril (Fig. 2). Además, se evidencia una merma interanual del 1% (Fig. 1), para todo el sector PyME Argentino fue de 3.4% (Fig. 2).

Figura 1**:** *variación interanual ventas minoristas Pyme*[4].

Figura 2: *Variación interanual venta minorista país sectorizado* [4],

La volatilidad de las variables macro y microeconómicas en Argentina obligan a la gerencia empresarial a acelerar los cambios y la toma de decisiones, estas se toman mayormente basadas en la intuición [1] o con herramientas y datos precarios carentes de la precisión que requiere el dinamismo del entorno empresarial Argentino (Tabla 1).

Tabla *1: Problemas de la gerencia PyME* [1].

· Falta de presupuesto.
· Carencia de conocimientos.
· Fuentes de datos numerosas y desordenadas.
· Falta de un equipo que se mantenga actualizado respecto de las tecnologías de analítica.
· Incapacidad para proporcionar a los usuarios finales las herramientas poder obtener valor añadido.
· Desconfianza en la tecnología.
· Resistencia al cambio.

El problema planteado entonces es la necesidad de contar con una solución analítica, accesible, económica y tecnológicamente viable para las PYMEs antes mencionadas que permita automatizar con precisión la detección de oportunidades de negocio y produzca una baja en los costos operativos de ubicación de stock respecto de potenciales compradores, minimizando el esfuerzo logístico.

A su vez, se requiere que la implementación sea transparente y pueda generar confianza a la gerencia empresarial.

1.4. **Metodología**

1.4.1. **Tipo de investigación: Aplicada**

Todos los ítems identificados en este trabajo tienen la finalidad de aplicarse a una solución informática que podrá implementarse de manera general en una o más empresas del sector ferretero que tengan implementado un pequeño sistema de administración de su información de ventas minoristas.

Se desarrollará sobre un *origen de datos* específico pero la estructura de minería y analítica podrá ser adaptada a futuros trabajos de inteligencia de negocios si así se requiriera.

1.4.2. **La elaboración de este trabajo consistirá en 6 etapas bien definidas**

Como se detalla en (Fig. 3), la elaboración de este trabajo tiene el siguiente itinerario metodológico:

- Revisión de la literatura siguiendo las pautas establecidas por Kitchenham [6] para tener una imagen actualizada del estado del arte y visualizar los avances en materia de innovación de negocio aplicada al sector estudiado.
- Elaboración del marco teórico de los conceptos abarcados en el trabajo.
- Elaboración y captura de los requisitos.
- Prueba y selección de herramientas a utilizar.
- Desarrollo del modelo propuesto.
- Recopilación de resultados.
- Desarrollo de conclusiones.

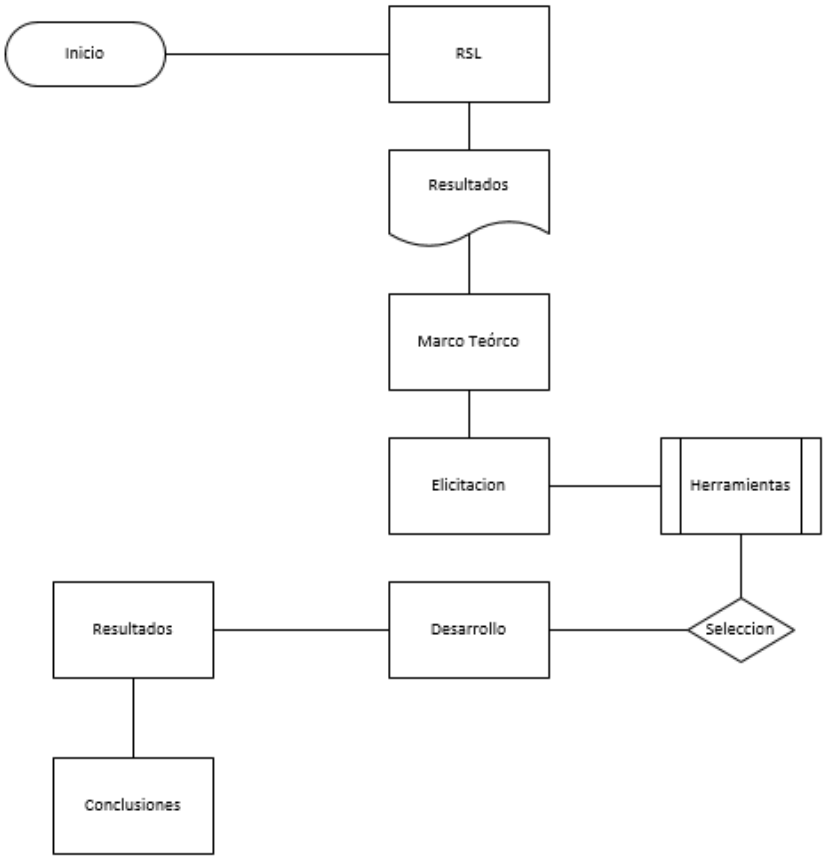

Figura 3: *Diagrama Metodológico.*

Se delineará con notación del Modelo de Proceso de Negocios [7] los procesos y actividades de la empresa seleccionada para llevar a cabo el procesamiento de los requerimientos que luego se utilizaran para desarrollar el tablero de *Location Analytics* y poder elaborar una solución que pueda ser utilizada de manera general por otras empresas similares; esto intenta desarrollar un caso de éxito en la región.

Para cumplir con el objetivo general de este trabajo se desarrollará un tablero de control de *business intelligence* que integre visualizaciones de datos geoespaciales en un mapa de capas GeoJson (que en síntesis es un formato estándar para representar elementos geográficos sencillos) con datos de negocio más la explotación de al menos 1 variable en un modelo de aprendizaje de máquina. Los orígenes de datos fueron alojados en un conjunto de repositorios que incluyen repositorios en la nube[8].

Para la elaboración final del tablero BI & GIS se utilizarán las como base las 5 fases del Ciclo de Vida de CRISP-DM propuesto en[9]: "*Entendimiento del Negocio, Entendimiento de los Datos, Preparación de los Datos, Modelado, Evaluación y Despliegue*", implementadas sobre la base de consultoría que he desarrollado en el sector privado.

Se optó por la plataforma integral Power BI debido a su enfoque innovador para simplificar la gestión de datos y el análisis empresarial. Esta herramienta permite a personas y organizaciones acceder fácilmente a los datos, generar informes de manera sencilla o automatizada, visualizarlos en paneles y compartirlos con un mínimo esfuerzo. Al ser respaldada por Microsoft, una empresa con una sólida reputación y respaldada por evaluaciones independientes como la de Gartner, que la han destacado frente a la competencia, queda claro que Power BI ofrece una oportunidad excepcional para que las instituciones de investigación y los profesionales satisfagan sus necesidades de análisis de datos. [8].

Por lo tanto se entiende que ésta es la plataforma para desarrollar el tablero que incluirá la arquitectura necesaria para desplegar un tablero de *Location intelligence* y aprovechando el amplio grado de conectividad y aceptación de distintas plataformas para poder insertar un modelo de aprendizaje de máquina.

Para determinar qué motor GIS gratuito disponible para el público sería utilizado en el tablero, se llevó a cabo una evaluación de precisión de las diferentes plataformas. Esta evaluación incluyó la generación de geocodificación y geocodificación inversa de direcciones conocidas. Posteriormente, se midieron las distancias a un punto en común y se contrastaron con las distancias reales para evaluar la precisión de cada plataforma.

Para la aplicación de un modelo de aprendizaje de maquina a nuestra arquitectura se probaron 2 lenguajes, ambos con similares conclusiones; sin embargo se terminó eligiendo R por sus bondades respecto a la integración con Power Bi, lo que permitió crear un botón que llame directamente al modelo desarrollado (Árbol de Decisión) y generar automáticamente una respuesta; lo último mencionado se desarrollara con más profundidad en el apartado de desarrollo de este documento.

El proyecto se expuso a un grupo de profesionales informáticos y se elaboró una encuesta para evaluar el desempeño del tablero a manos de expertos.

Este documento consta de:

- Introducción

  Breve descripción del contenido de este trabajo y un avance general del contenido de la investigación, describiendo la estructura del documento.

- Marco Teórico

  Recopilación y análisis de las teorías y enfoques teóricos que validan el encuadre y los procesos de esta investigación.

- Estado del Arte

  Evaluación sistemática y resumen del estado actual del conocimiento en las áreas alcanzadas por este trabajo.

- Análisis y diseño del proyecto

  Sintonización del desarrollo y modelado del proyecto de investigación, detalle de los requisitos, herramientas y enfoques utilizados y descripción de la adaptación de las tecnologías que son utilizadas para la concreción de este trabajo.

- Implementación

  Descripción del proceso de implementación del modelo propuesto.

- Resultados

  Valoración de los resultados obtenidos del proceso de análisis, desarrollo e implementación del modelo propuesto.

- Conclusiones

  Verificación del cumplimiento de la hipótesis propuesta basada en los objetivos planteados y los resultados obtenidos durante el desarrollo de este trabajo.

Toda la información del trabajo más las evaluaciones de los expertos fueron recopiladas y analizadas para llegar a una conclusión final.

Capítulo 2

2. Marco Teórico

2.1. **Business Intelligence**

Desde el punto de vista técnico, [10] define a la inteligencia de negocios como el BI se refiere al proceso de extracción, transformación, gestión y análisis de datos comerciales, con el fin de apoyar la toma de decisiones. Este proceso se basa principalmente en grandes conjuntos de datos, particularmente en el almacén de datos, con la misión de difundir inteligencia o conocimiento en toda la organización, tanto en lo táctico como en lo operativo.

2.1.1. **Proceso BI**

Dentro del proceso de inteligencia de negocios se pueden definir 5 fases:

Fuente de datos

Un sistema de BI es capaz de extraer datos de múltiples fuentes de datos, que representan diferentes unidades de negocio, como marketing, producción, recursos humanos y finanzas. Los datos extraídos deben limpiarse, transformarse e integrarse para su análisis [10].

Análisis de datos

En esta fase se realiza la conversión de los datos en información o conocimiento a través de diferentes técnicas de análisis de datos, como informes, modelado, visualización y minería de datos. Los resultados del análisis de datos ayudan a los gerentes a comprender mejor el medio ambiente y a tomar mejores decisiones [10].

Comprensión del estado del negocio

Una comprensión sólida de la situación actual de toma de decisiones, derivada de un análisis exhaustivo de los datos, es fundamental para la toma de decisiones efectiva. Los sistemas de inteligencia empresarial (BI) deben ser capaces de asistir a los tomadores de decisiones en la elaboración de una visión completa y detallada de sus contextos de toma de decisiones. [10].

Evaluación de riesgos

En [14] se identifican amenazas y oportunidades y, por lo tanto se responden a las mismas en consecuencia. En la actualidad los entornos donde operan las empresas son cada vez más complejos. La toma de decisiones comerciales tiene más probabilidades de sufrir riesgos del entorno externo e interno. Por lo tanto, es importante realizar una acertada evaluación de riesgos para un sistema BI.

Soporte en la decisión

El objetivo final de BI es ayudar a los gerentes a tomar decisiones sabiamente, basándose en los datos comerciales actuales [10].

2.2. **Sistemas de Información Geográfica**

Hay numerosas definiciones para los Sistemas de Información Geográfica (SIG), a modo de resumen [11] define a un SIG como un conjunto de hardware, software y procedimientos para capturar, manejar,

manipular, analizar, modelizar y representar datos georreferenciados, en el intento de abordar desafíos relacionados con la gestión y la planificación.

Las ciencias aplicadas al análisis espacial, en especial la geografía, estudian la información de la superficie buscando distintas capas temáticas (suelo, las ciudades, hidrografía, proliferación humana, carreteras y ferrocarriles, así como cualquier infraestructura desarrollada); detectando distintos tipos de capas de información acerca de una zona determinada (Fig. 4)[12].

Figura 4*: Capas de información en un SIG* [12].

2.2.1. **Geolocalización**

La Geolocalización se basa en la representación de un punto determinado dentro del espacio real. El Sistema de Posicionamiento Global (GPS, del inglés Global Positioning System) es el sistema más utilizado para realizar la localización geográfica de un dispositivo a través de su sistema satelital, tiene la capacidad de localizar un aparato determinado con una precisión que puede llegar a los pocos metros [13].

Existe un gran número de trabajos innovadores, entre ellos MOTEL [16] (*Mobile Technology Enhanced Learning*) que ha desarrollado una plataforma y éste permite relacionar a alumnos e investigadores utilizando dispositivos móviles relacionando los mensajes con coordenadas geográficas, marcados como puntos en un mapa y agregados a la información de los usuarios.

2.2.2. **Geocodificación**

Geocodificar consiste en asignar coordenadas geográficas (latitud, longitud) a puntos del espacio [14], siendo el formato de mayor utilidad el formato de los puntos que se correspondan con direcciones postales.

También se debe mencionar la Geocodificación Inversa que es, tal como su nombre lo indica, a partir de coordenadas geográficas encontrar la ubicación en términos más comprensibles, por ej: Una dirección postal.

2.2.3. **GeoMarketing**

Disciplina que estudia las relaciones existentes entre las estrategias y políticas de Marketing, y el territorio o espacio, donde se ubican las empresas, sus clientes, proveedores y puntos de distribución, también conocida como inteligencia de mercado geográfico, es una metodología que brinda la elaboración de estrategias efectivas para vender más, reducir costos y elevar resultados a través de información que contiene metadatos georreferenciados como se describe en (Fig. 5). Esta disciplina incluye cualquier proyecto de marketing que incorpore inteligencia de ubicación para mejorar la precisión con la que la campaña llegue al consumidor adecuado y en el momento adecuado [15].

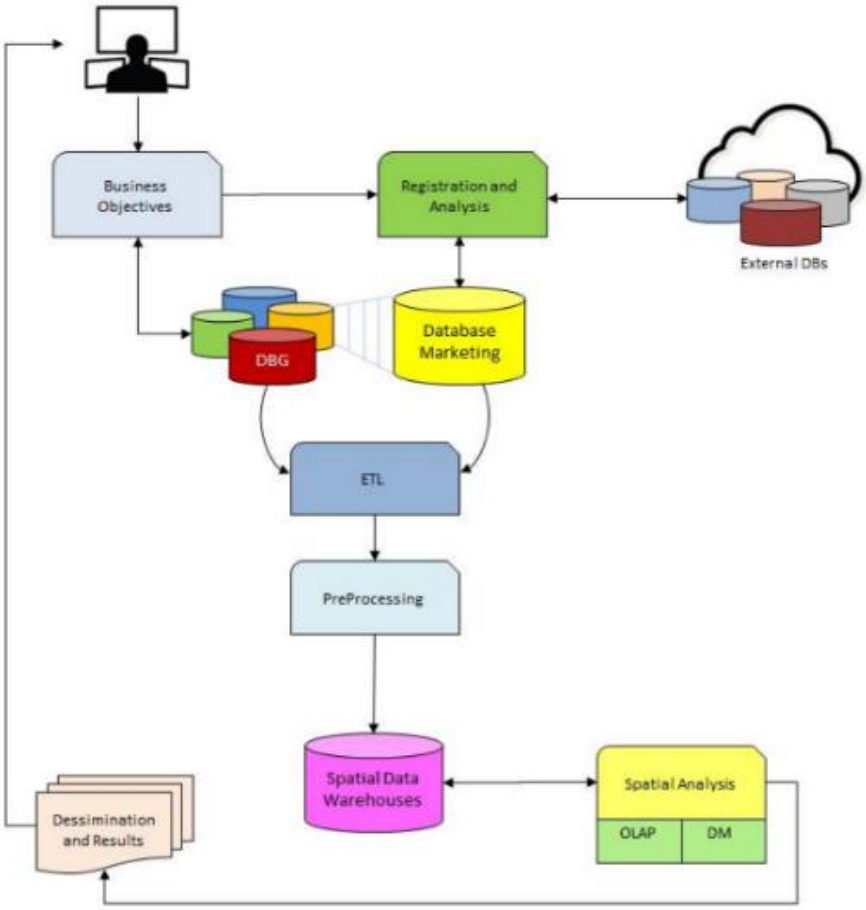

Figura 5: *Modelo de GeoMarketing* [15].

2.3. **Inteligencia de Negocios + Geolocación**

2.3.1. ***Location Data*** **(Datos de ubicación)**

Son datos con una dimensión geográfica. Los datos de ubicación están en todas partes, ya que todas las acciones que ocurren en o cerca de la superficie de la Tierra utilizan aspectos geográficos. Por lo general, se lo conoce como cualquier dato con coordenadas (latitud, longitud y, a veces, altitud), pero también abarca diferentes unidades geográficas agregadas, incluidas direcciones, códigos postales, puntos de referencia, distritos, ciudades, regiones y mucho más [12].

2.3.2. **Tipos de datos de ubicación**

Los tipos de datos geográficos se pueden dividir en dos grandes categorías [12]:

- Datos vectoriales: Esto se representa como puntos, líneas o polígonos. Es probable que los datos se creen digitalizándolos y almacenando información en longitud y latitud. Este tipo de datos es útil para almacenar datos que tienen límites discretos y distintos, como fronteras, parcelas de terreno, calles y puntos de interés.
- Datos *ráster*: almacena información en celdas y, por lo tanto, es adecuado para almacenar datos continuos, como imágenes de satélite, modelos de elevación y otras fotografías aéreas.

2.3.3. **Location Intelligence**

Es el proceso de convertir datos geográficos (espaciales) en conocimientos y resultados comerciales. Cualquier dato con una posición geográfica, ya sea implícita o explícitamente, requiere métodos de preprocesamiento conscientes de la ubicación, visualización, así como métodos analíticos para obtener información de ellos. Por lo tanto, las aplicaciones de inteligencia de ubicación pueden revelar patrones ocultos de relaciones espaciales que no pueden derivarse por otros medios normales. Conduce a una mejor toma de decisiones sobre problemas espaciales, dónde suceden las cosas, por qué suceden en algunos lugares y las tendencias espaciales en el análisis de series temporales. Comprender la dimensión de ubicación de los desafíos actuales en la industria, el comercio minorista, la agricultura, el clima y el medio ambiente puede conducir a una mejor comprensión de por qué las actividades económicas, sociales y ambientales tienden a ubicarse donde están [12].

2.3.4. **Location data intelligence**

Todas las industrias utilizan inteligencia de ubicación. Ayuda a las industrias a comprender qué están haciendo sus clientes, dónde se encuentran, cuál es el entorno geográfico de sus clientes y cuáles son sus intereses. La inteligencia de ubicación normalmente se define como el uso de datos de ubicación con otros atributos para agregar contexto y obtener información, servicios y productos útiles que ayuden a las organizaciones a tomar decisiones efectivas y eficientes. La información derivada a través de la inteligencia de ubicación puede tener una perspectiva comercial y económica, así como una perspectiva ambiental y social [12].

Geo Business Intelligence, o *Location Intelligence*, es una colección de tecnologías que integran *Business Intelligence* con Geo-Localización y se utiliza para implementar sistemas de apoyo a la decisión con el propósito de mejorar la eficiencia de las actividades comerciales en el entorno de servicio. Desde un punto de vista general, *Business Intelligence* tiene como objetivo lograr la Optimización de Procesos y la Reducción de Costos. Básicamente, esto se obtendrá a través de herramientas de análisis de datos y procesos comerciales con el objetivo de obtener los KPI (*Key Performance Indicators*) adecuados de interés, mantenerlos bajo control, y lograr una visualización inmediata de los mismos (informes, cuadros de mando dinámicos, sistemas de análisis histórico, etc.) [16].

Para tomar decisiones de negocio que resulten en términos de: mayor conocimiento de los ingresos, reducción de costes y mayores eficiencias necesitamos un sistema mejorado que amplíe las posibilidades de análisis sobre cualquier tipo de KPI: gráfico, geográfico, documental, digital, financiero, estadístico, etc [16].

Desde una perspectiva comercial, el valor de los mapas y los datos de ubicación es crucial en muchas aplicaciones comerciales. Una mirada rápida a las grandes empresas como Google, Apple, Microsoft y Nokia muestra que cada una de estas empresas tiene sus propios servicios y productos de ubicación y mapas. Hay que destacar la frecuencia con la que utiliza el servicio de ubicación de la API de Google Maps a través de un teléfono. Esto también destaca la importancia de los datos de ubicación, ya que todas estas empresas no harían todo lo posible para tener su propia producción interna de datos de ubicación si no fuera necesario. Las aplicaciones comerciales en datos de ubicación incluyen no solo usos individuales de datos de ubicación, sino también aplicaciones innovadoras que van desde marketing individualizado, vehículos autónomos, logística y transporte hasta atención médica [12].

2.4. **Recopilación de Requisitos en proyectos BI**

Existen herramientas que permiten a los proyectos de Inteligencia de negocios convertir la información disponible dentro de la organización en conocimiento útil para dar soporte a la toma de decisiones. Esto es debido a que los proyectos BI tienen distintos requerimientos que los proyectos de ingeniería de software, por este motivo los procesos de obtención de requisitos asociados al desarrollo de software no son útiles a la hora de ser aplicados en este tipo de proyectos [17].

Es por ello que para la recopilación de requisitos y desarrollo de proyectos de explotación de la información debe utilizarse un modelo más específico.

La capa más compleja dentro de la Ingeniería de Requerimientos es justamente la de recopilación., ya que describir necesitados y expectativas futuras del potencial usuario del sistema agrega un grado más de dificultad al analista del sistema que se está por desarrollar. Además, si el sistema que se va a desarrollar no es transaccional, sino como es el caso de los sistemas BI, un sistema de apoyo a la toma de decisiones, la dificultad se incrementa aún más. En [18] se considera a la toma de decisiones como un proceso más a ser modelado, proponiendo una heurística basada en la Notación de Modelado de Procesos de Negocio (BPMN) como de muestra en (Fig. 6).

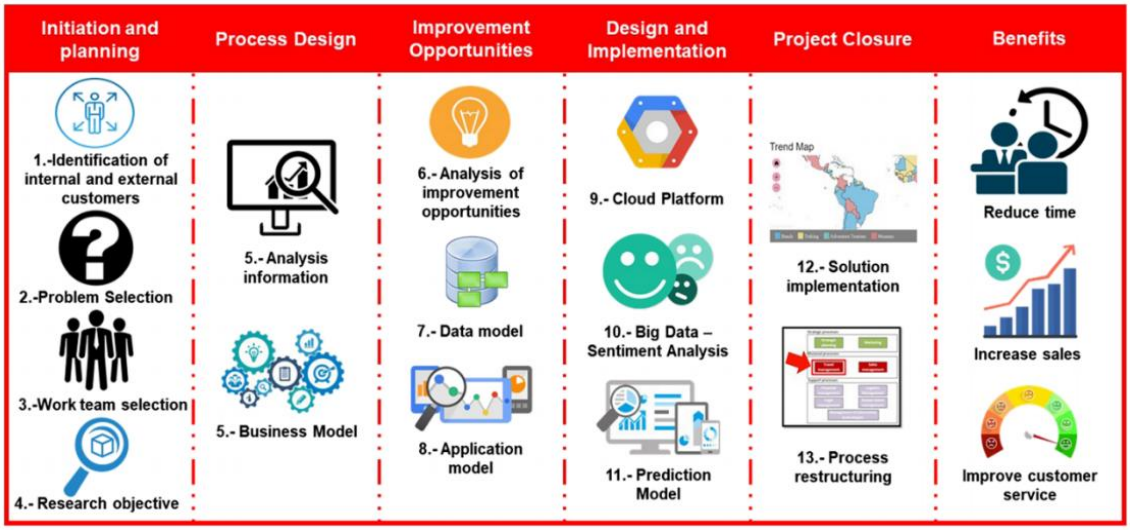

Figura 6 *: Modelo de procesos [18]*

2.5. **Inteligencia Artificial**

La inteligencia artificial se encuentra arraigada en la historia del desarrollo de la inteligencia humana, presentando aspectos que desafían las narrativas teleológicas que sugieren que la IA simbólica surge de manera natural e inevitable como resultado de los esfuerzos a lo largo de los siglos por reducir el razonamiento humano a un formalismo lógico. A lo largo del tiempo, las capacidades cognitivas humanas han sido teorizadas, clasificadas y valoradas de diversas formas. [19].

2.6. **Aprendizaje de máquina**

El aprendizaje de máquina es la subdisciplina de la Inteligencia Artificial (IA) dedicada a estudiar algoritmos y modelos que son utilizados para realizar una tarea sin utilizar instrucciones explícitas, en lugar de basándonos en patrones e inferencias [20].

Partiendo de un conjunto de datos (conocido como “conjunto de entrenamiento”), estos algoritmos implementan modelos matemáticos que se usarán para hacer predicciones sobre nuevos datos a posteriori [20].

Aprendizaje supervisado: El algoritmo construye un modelo matemático partiendo de un conjunto de datos que contiene tanto las entradas como las salidas deseadas [20].

Aprendizaje no supervisado: El algoritmo implementa un modelo matemático partiendo de un conjunto de datos que solo contiene las entradas, y no las salidas deseadas [20].

Aprendizaje por refuerzo: Dado un entorno de reglas y metas a alcanzar, un agente (el programa) recibe un feedback, positivo o negativo dependiendo de la estrategia que use para alcanzar la meta objetivo, sin caer en mínimos locales [20].

2.6.1. **Redes neuronales Artificiales**

Las Redes de Neuronas Artificiales (RNA) son modelos remotamente basados en las redes de neuronas biológicas que constituyen el cerebro. Estos sistemas están compuestos por una colección de neuronas artificiales, conectadas entre sí [20].

Cada conexión permite transmitir información (como si de la señal de una neurona biológica se tratase) entre neuronas. Estas conexiones tienen atribuidos unos pesos, los cuales se ajustan a medida que el sistema va aprendiendo. Dichos pesos simbolizan la intensidad de la conexión. Las neuronas artificiales pueden tener un umbral por el cual sólo pasan las señales en caso de ser superiores a dicho umbral. Además, cada neurona artificial puede procesar la información que recibe [20].

2.6.2. **Árbol de regresión y particionamiento recursivo (rpart)**

Este algoritmo sirve para encontrar la variable independiente que mejor separa los datos en distintos grupos en concordancia con las categorías de la variable objetivo. Se considera la mejor separación que puede expresarse con una regla y a cada regla corresponde un nodo [21].

Su funcionamiento se basa en buscar la variable que mejor separa los datos en grupos, obtener entonces una regla y separar los datos según la misma. Este procedimiento se realiza de manera recursiva hasta que no existe forma de obtener una mejor separación y entonces el algoritmo se detiene. Matemáticamente se expresa en la Ecuación 1 [21].

Ecuación 1*: Fórmula Árbol de regresión y particionamiento recursivo* [21].

$$\sum_{i=1}^{i=n} (y_i - \hat{y}_i)^2$$

2.6.3. **Redes neuronales**

La red de Perceptrón Multicapa (MLP) [19] se compone de varias capas que controlan el traspaso de información entre las capas de entrada y salida; la información viaja en un solo sentido donde un conjunto de datos en la capa de entrada se propagan a través de una o más capas ocultas sucesivamente hasta llegar a la capa final con los datos de salida. Las capas ocultas entre las capas de entrada y salida son los componentes clave de una red neuronal y hacen de nexo para la transferencia de información. Cada capa oculta se compone de nodos que tienen conexiones ponderadas que sirven para definir cuál será la fuerza en el flujo de información entre capas (Ecuación 2). En la práctica se pueden usar capas de sesgo que estén conectadas a capas ocultas y de salida, estas con similares o análogas en términos de intercepción en un modelo de regresión estándar [25].

Ecuación 2: *Algoritmo Red Perceptrón Multicapa* [19].

$$y_k = f_0\left(\sum_h w_{hk}\, f_h\left(\sum_i w_{ih}\, x_i\right)\right)$$

2.6.4. **Cloud Services**

Debido a la rapidez con la que avanzan las TIC (Tecnologías de la Información y las Comunicaciones), la computación en nube se convirtió en un servicio predominante. Las nubes son sistemas informáticos y de almacenamiento de datos que integran diferentes tecnologías para interconectar y gestionar los recursos demandados en los ordenadores distribuidos. Los usuarios pueden utilizar los datos, el almacenamiento y las aplicaciones proporcionadas por los proveedores de servicios en la nube para aprovechar nuevos datos, servicios y trabajos en curso [23].

Las aplicaciones de la computación en nube incluyen estas cuatro características. Primero, tiene el centro seguro y confiable de almacenamiento de datos. En segundo lugar, puede compartir datos entre varios equipos. En tercer lugar, puede permitir a los usuarios utilizar internet de forma infinita [23].

Existen tres capas (modelos de entrega) de computación en la nube [23]:

1. Software como servicio (**SaaS**) [23].
2. Plataforma como servicio (**PaaS**) [23].
3. Infraestructura como servicio (**IaaS**) [23].

2.7. **CRISP-DM**

CRISP-DM es el estándar de facto y un modelo de proceso independiente de la industria para aplicar proyectos de minería de datos; consta de seis fases iterativas desde la comprensión del negocio hasta la implementación. A continuación se describe brevemente la idea principal, las tareas y el resultado de estas fases, según la guía del usuario de CRISP-DM [24].

Entendimiento del negocio

La situación del negocio debe evaluarse para obtener una visión general de los recuses disponibles y necesarios. La determinación del objetivo de la minería de datos es uno de los aspectos más importantes en esta fase. Primero se debe explicar el tipo de minería de datos y los criterios de éxito de la minería de datos. Se debe crear un plan de proyecto obligatorio.

Entendimiento de los datos

La recopilación de datos de las fuentes de datos, su exploración, descripción y verificación de la calidad de la datos son tareas esenciales en esta fase. Para hacerlo más concreto; la ya del usuario describe la tarea de descripción de datos mediante el uso de análisis estadístico y la de terminación de atributos y sus cotejos. La selección de datos debe realizarse mediante la definición de criterios de inclusión y exclusión. La mala calidad de los datos se puede manejar limpiando los datos.

Preparación de los datos

Dependiendo del modelo utilizado (definido en la gimen fase) se deben construir los atributos derivados. Para todos estos pasa son posibles diferentes métodos y dependen del modelo.

Modelado

La fase de modelado de datos consiste en seleccionar la técnica modelado, construir el caso de prueba y el modelo. Se pueden utilizar todas las tácticas de minería de datos. En general, la elección depende del problema comercial y de los datos. Más importante es cómo explicar la elección. Para construir el modelo, se deben establecer parámetros específicos. Pan evaluar el modelo; es apropiado evaluar el modelo contra los criterios de evaluación y seleccionar los mejores.

Evaluación

En la fase de evaluación los resultados se comparan con los objetivos comerciales definidos. Por lo tanto, los resultados deben interpretarse y deben definirse acciones adicionales. Otro punto es que se debe revisar el proceso en general.

Despliegue

La fase de implementación se describe de forma general en la guía del usuario. Podría ser un informe final o un componente de software. La guía del usuario describe que la fase de implementación consiste en planificar la implementación, la supervisión y el mantenimiento.

CRISP-DM es un modelo completo y documentado (Fig. 7), en la misma se describen las 5 fases del ciclo de vida. Gracias a sus etapas debidamente organizadas, estructuradas y definidas es posible que un proyecto pueda ser fácilmente entendido o revisado [25]. Aunque el proceso CRISP-DM es independiente de la herramienta elegida.

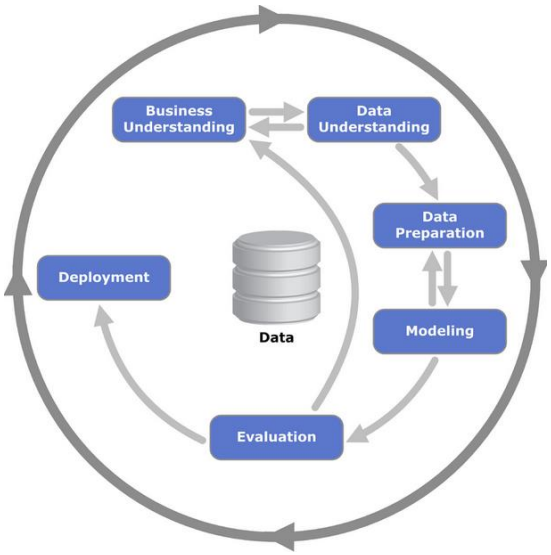

Figura 7 *: Ciclo de Vida CRISP-DM* [25].

2.8. Modelado de Proceso de Negocio (BPMo, siglas en inglés para Business Process Modelling)

El modelado de procesos de negocio surgió con el fin de proporcionar una mejor comprensión de los procesos de negocio en las organizaciones. El resultado del modelado de procesos de negocio es un modelo de proceso, que consiste en un conjunto de modelos de actividad y restricciones de ejecución entre ellos. Por lo general, se ilustra con actividades y eventos que están asociados con los flujos de control. Dichos procesos

se pueden escribir con diferentes lenguajes para el modelado de procesos, también conocidos como técnicas o notaciones. Elegir un lenguaje apropiado requiere consideración, porque no todos los lenguajes de modelado son adecuados para todos los aspectos del proceso [26].

2.9. **Notación y Modelado de Procesos de Negocio (BPMN)**

La Notación y Modelado de Procesos de Negocio (por sus siglas en inglés BPMN, Business Process Model and Notation) es un estándar para el modelado de procesos comerciales que permite graficar adecuadamente los procesos comerciales. El resultado es un *Business Process Diagram* (BPD), que aplica diagramas de flujo para especificar la información y el funcionamiento de los procesos involucrados.

BPMN sirve para respaldar el modelado de procesos comerciales y es útil tanto para usuarios técnicos como para usuarios comerciales ya que la notación es simple e intuitiva pero tiene la capacidad suficiente para representar procesos semánticos complejos [7]. La especificación BPMN 2.0 incorpora semántica de ejecución, mapeo entre gráficos de notación y varios lenguajes de ejecución como el lenguaje de ejecución de procesos comerciales (BPEL).

Entre los usuarios se pueden destacar a los analistas comerciales que crean y refinan los procesos, los desarrolladores responsables de implementarlos y la gerencia comercial que supervisan y administran dichos procesos. Esta notación sirve como un lenguaje común y es de gran utilidad para minimizar la brecha de comunicación que existe entre el diseño y la implementación del proceso comercial [7].

2.9.1. **Modelo de Centro de Competencias de Inteligencia de Negocios (CCBI)**

Dentro del Centro de Competencias de Inteligencia de Negocios se integran los datos, procesos, roles, proveedores, capacitación, entrega y soporte como se visualiza en la Figura 8 [27].

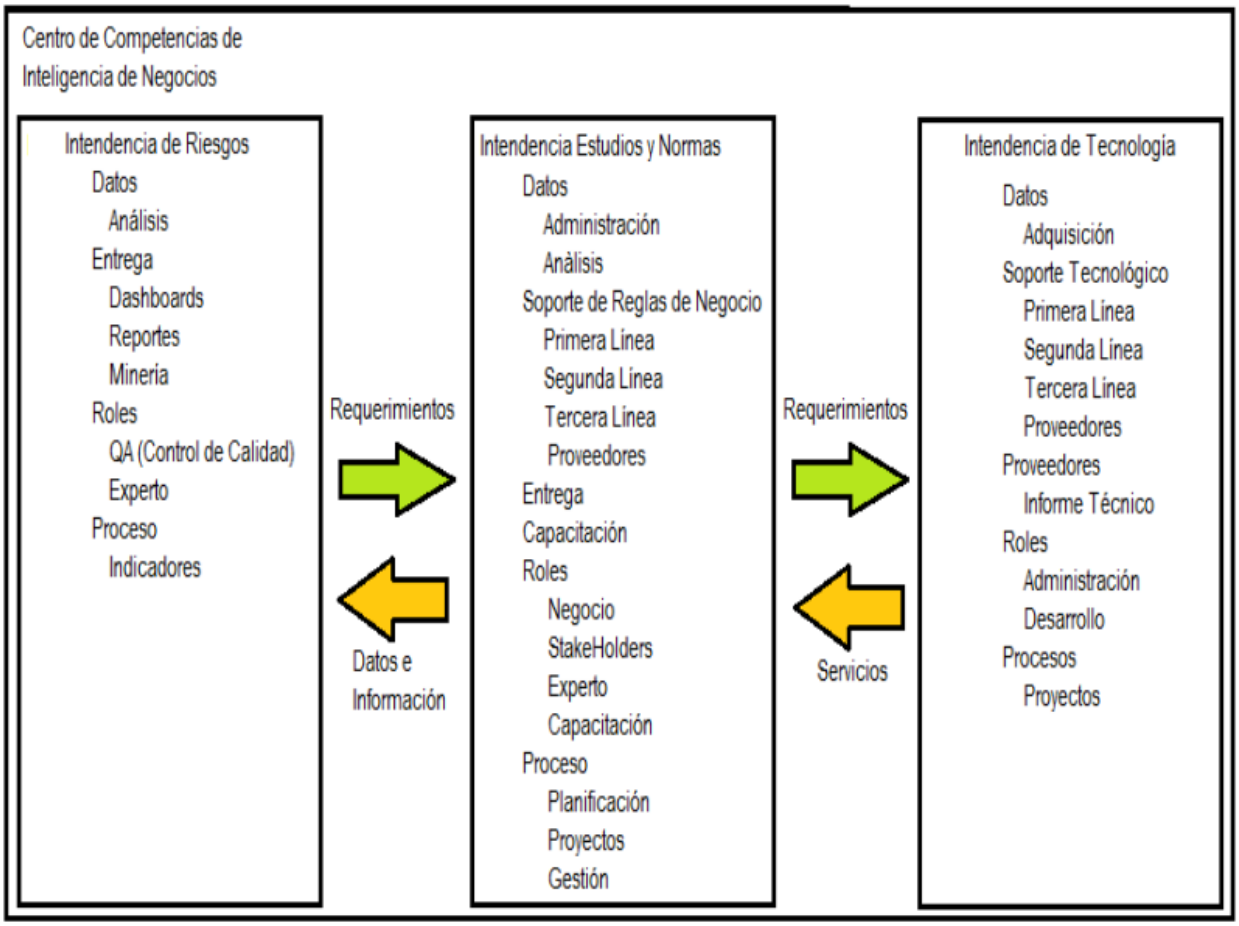

Figura 8: *Centro de competencias de inteligencia de negocios [27].*

2.10. **Power BI**

Power BI es una plataforma en línea que posibilita la exploración, transformación, visualización y compartición de datos, así como la generación de informes y paneles que pueden ser compartidos con otros usuarios dentro de un mismo departamento, organización o incluso con el público en general. Esta herramienta constituye un conjunto integral para el análisis de datos empresariales y la difusión de información a través de informes y paneles interactivos.

La entrada de datos en Power BI puede adoptar diversas formas, incluyendo hojas de cálculo, archivos de texto y bases de datos, entre otros. Estos datos son procesados y transformados por los usuarios para crear conjuntos de datos pertinentes. La etapa de transformación es crucial, ya que permite la corrección de errores, la eliminación de datos redundantes y la preparación de los datos para análisis posterior mediante la normalización y la organización en formas estructuradas adecuadas.

Durante el desarrollo de informes y paneles en Power BI, es fundamental filtrar los datos para incluir únicamente la información relevante, lo que permite focalizarse en los aspectos más significativos de los datos analizados. Este enfoque selectivo optimiza la eficiencia del análisis y la presentación de los resultados. [8].

Por decimoquinto año consecutivo, Gartner ha posicionado a Microsoft como líder de su Cuadrante Mágico 2022 para plataformas de análisis e inteligencia empresarial. También se ha posicionado más a la derecha en Integridad de la Visión y más alto en la Capacidad de Ejecución dentro del cuadrante de Líderes por cuarto año consecutivo [2].

2.11. **CrowdSourcing**

Es un recurso de información proveniente de una amplia gama de individuos, quienes se congregan para abordar problemas colectivamente, ya sea para generar contenido, innovar en productos, hallar soluciones y otras iniciativas similares. Las empresas emplean el crowdsourcing como una estrategia para mejorar sus operaciones. Este término se refiere a la participación masiva de personas que no están afiliadas a una entidad específica o institución, en una colaboración abierta y sin restricciones. Es un modelo de trabajo colaborativo que involucra a diversos participantes externos [15].

2.11.1. **CrowdMapping**

La construcción colaborativa de mapas de información, las bases de datos *Crowdsourcing* alimentadas por una multitud conectada y sus entradas multimedia, en tiempo real y con datos geográficos. Los mapas colectivos brindan información sobre eventos globales como desastres de salud, humanitarios, políticos e incluso naturales [15].

Enfoque sistémico del RLDM (R*etail Location Decision Making* - toma de decisiones de ubicación minorista) basado en datos [28].

Estos entornos son típicamente eficaces en la integración, uso y desarrollo de enfoques tradicionales de RLDM. Esto incluye el uso de datos espaciales más tradicionales (datos de censos, datos de ubicación de competidores, datos de programas de lealtad, etc.) y GIS para realizar análisis de competidores, mapeo de mercado, análisis de áreas comerciales, selección y revisión de sitios, planificación de redes, elaboración de perfiles de clientes y planificación de adquisiciones y fusiones. Su capacidad para participar en decisiones basadas en datos [28].

Capítulo 3

3. Estado del Arte

3.1. Introducción

Se realizó una revisión sistemática de la literatura para evidenciar el estado del arte en los temas que abarca este trabajo, por tal motivo eligió como tema principal para la búsqueda:

"Location Business Analytics en Cloud Aplicada a Pymes".

Dentro del tema principal se identificaron las siguientes palabras clave: *Location Business Analytics, Machine learning, Cloud Services,* PYME y SME.

A continuación se muestra en la Tabla 2 las palabras clave y sus palabras relacionadas en la búsqueda.

Tabla 2: *Palabras Clave*

Palabra Clave	Relacionadas
PYME	SME
Location	GIS, espacial
Business	
Analytics	
Cloud	

3.2. Trabajos Relacionados

[29] *Location Data Analytics in the Business Value Chain: A Systematic Literature Review.*

En [29] se realizó una revisión sistemática de la literatura sobre *Location Data Analytics* en la cadena de valor empresarial. Se tomaron artículos desde el 2014 hasta el 2019 y si bien no se llega a efectivizar conclusiones definitivas sobre los enfoques, la efectividad y las peculiaridades de cada aspecto de los entornos se pudo concluir que las medidas y las estrategias de validación pueden mejorar de manera que se pueda identificar un punto de referencia integral de los estudios, particularmente con la ubicación y el análisis espacial.

3.3. Objetivo y Preguntas de Investigación

Esta RSL se realiza con el objetivo de sintetizar la literatura existente acerca de los avances en *Location Business Analytics* en *Cloud* Aplicada a Pymes. El conocimiento extraído de esta RSL será la base para un mejor entendimiento de los adelantos en materia de investigación para el aporte científico que permita desarrollar un modelo integrado de GIS y *Business Intelligence* que de soporte en la toma de decisiones a la gerencia de pequeñas y medianas empresas (ver ANEXO 1).

3.3.1. Ejecución

Los resultados obtenidos en la búsqueda y los estudios primarios resultantes se detallan en la (Tabla 3).

Tabla 3: resultados de la búsqueda por fuente

Fuente	Artículos encontrados	Articulos restantes	Estudios primarios	Porcentaje por fuente
Science Direct	4	1	0	
Google Académico	292	19	9	84,60%
IEEE	0	0	0	
ACM	0	0	0	
SCOPUS	5	2	0	
MINCYT	0	0	0	
Microsoft Academic	10	5	1	15,40%
Totales	311	27	10	100,00%

3.4. Reporte de Resultados

3.4.1. Tipos de Metodologías y modelos existentes para desarrollar proyectos de BI and GIS

En [30] se desarrolló un proyecto para explotar la información de los datos de hospedajes en la plataforma Airbnb para dar una solución de soporte a la toma de decisiones utilizando metodología CRISP-DM y modelando con IBM SPSS Modeler.

En [31] se propone realizar el proyecto bajo el Ciclo de vida de Kimbal y en cuanto a la gestión del proyecto, se utilizó para cada PEC la metodología especificada en la Guía del PMBOK (Project Management Body of Knowledge), la cual define las siguientes fases en su ciclo iterativo:

1. Iniciación: partiendo de la planificación inicial que figura en el siguiente apartado, se revisa el alcance, recursos y tiempo disponibles para cada PEC.
2. Planificación: se detalla la planificación inicial en actuaciones más concretas para lograr los entregables previstos.
3. Ejecución: llevar a cabo las actuaciones necesarias hasta alcanzar los objetivos.
   Durante todo este proceso se lanzan en paralelo las siguientes fases:
4. Seguimiento y control: para evitar desviaciones sobre lo previsto.
5. Realimentación: permitiendo que las situaciones que se hayan dado durante el desarrollo de una PEC sirvan de aprendizaje para el resto de entregas.

En [32] se optó por un Modelo de Análisis de inteligencia de negocios (BI&A) en Cloud basado en una arquitectura híbrida, preparada para realizar el procesamiento de los datos localmente y en la nube. Esto se debe a que las soluciones BI&A en Cloud son una buena opción en una PYME. Se especifica la importancia de realizar este proceso en forma gradual y utilizar inicialmente un esquema híbrido. El proceso de extracción, transformación y carga de datos (ETL) es el responsable de procesar y transformarlos, según el modelo dimensional elaborado (Fig. 9). Estos datos se almacenan en un "*Cloud Data Warehouse*", cuando se carga toda la información en la nube, esta puede ser consumida por los servicios de BI y *Analytics*.

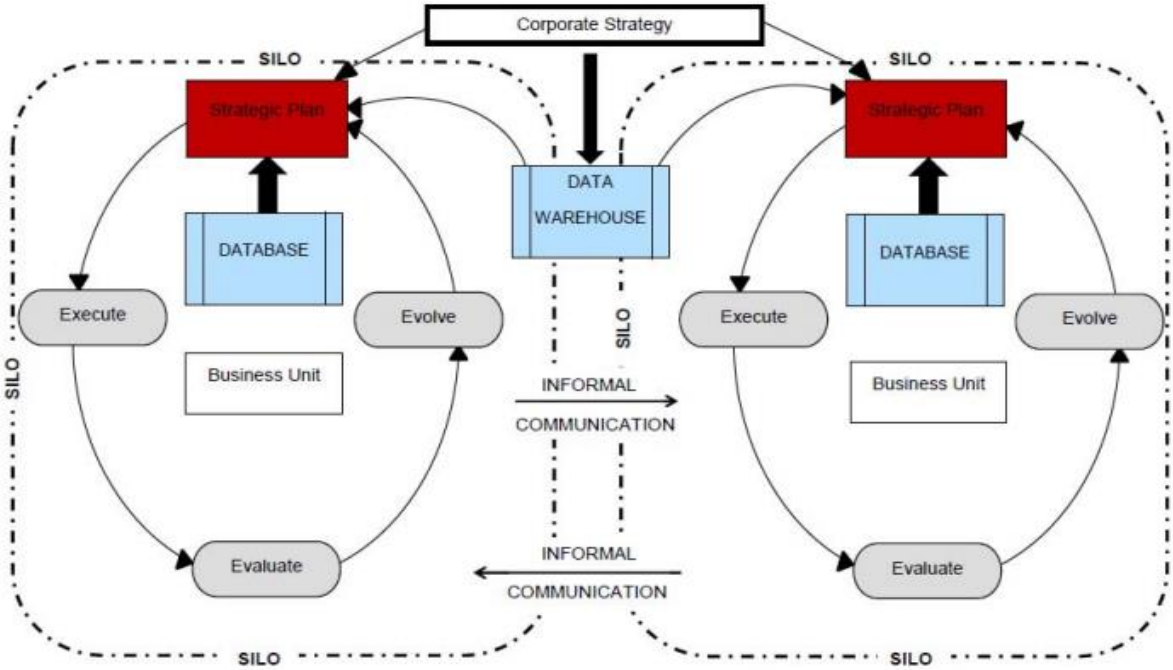

Figura 9: *Fases para definir un modelo de negocio [32].*

En [33] para la implementación de un proyecto Análisis de Sentimientos se propuso un modelo de negocio en un caso donde se presentaron escenarios en los cuales los procesos internos del sector turístico pueden ser utilizados como fuentes de información que permitan analizar las características y comportamientos de los consumidores.

3.4.1.1. Algoritmos de localización

En [34] se utilizan dos algoritmos matemáticos para lograr la máxima precisión en el aproximamiento de los objetos de estudio:

Location Set Covering Model (LSCP), modelo de cobertura máxima definida por la localización, modelado dentro de la programación lineal [35]; este modelo busca ubicar un número mínimo de servicios necesarios para obtener la cobertura obligatoria de toda la demanda [36]. En otras palabras, cada punto de demanda tiene como mínimo un servicio ubicado dentro de un tiempo o distancia establecida o estandarizada.

Maximum Covering Location Problem (MCLP): reconoce una cobertura mandatoria para toda la población en todas las localidades sin importar que tan lejos vivan, el desarrollo de este modelo podría requerir demasiados recursos [34].

3.4.2. Validación de las metodologías o los modelos

En [6] se realizaron 2 tipos de evaluaciones sobre el modelo. Primero se realizó una implementación piloto en una PyME que fue complementada con un análisis económico y en la segunda se realizó una validación por juicio de expertos.

En la prueba piloto, según los indicadores obtenidos, se observó que el periodo de recuperación de la inversión era de 13 meses con un recuperación del 74%, debido al mínimo gasto en hardware adicional y

software permanente que brindan las soluciones en la nube. El beneficio sobre el costo de inversión fue del 174% resaltando el atractivo y los beneficios de este tipo de implementaciones [6].

Se examinó la opinión de 4 expertos en las tecnologías *Cloud Computing*, BI&A para la validación a través de juicio de expertos de la implementación piloto que se realizó [6]. Las valoraciones fueron mayormente positivas.

3.4.3. **Tratamiento de los orígenes de datos**

En [27] se realizó tratamiento de las tablas dado que concluyeron que para obtener una visualización adecuada de los datos es necesario que en nuestras B.D. se encuentren la longitud y latitud de lo que se vaya a mostrar en un mapa. En este aspecto nos pueden ayudar los servicios que nos ofrecen Google y Microsoft.

3.4.4. **Bases de Datos Espaciales y no espaciales.**

Para mejorar la precisión de las metas en proyectos de integración GIS, es necesario identificar la ubicación de los recursos (Fig. 10), de modo que se debe diseñar una base de datos espacial donde cada objeto ocupa una posición georeferenciada en un plano cartográfico [37].

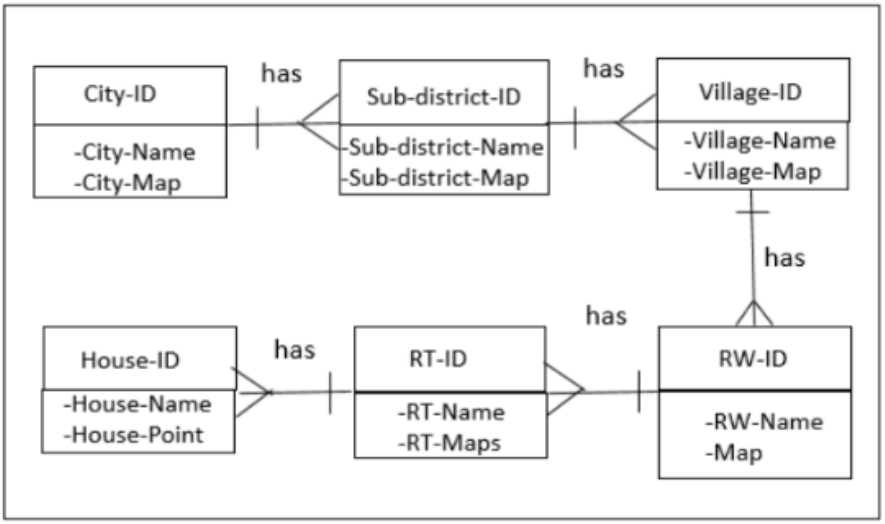

Figura 10: Estructura del mapa lógico [37].

Cuando se trabaja con datos poblacionales en [37] se destaca la necesidad del diseño de base de datos tabular donde sea fácil de procesar grandes volúmenes de datos personales (Fig. 11).

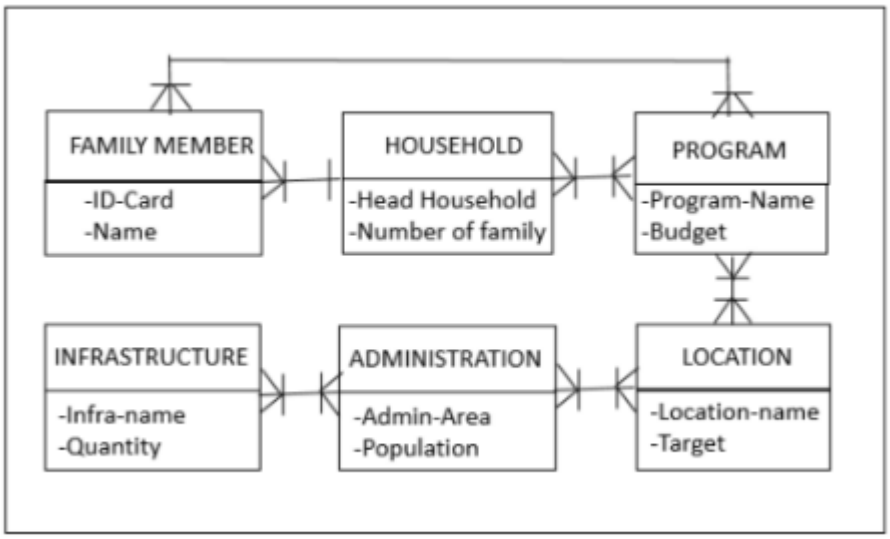

Figura 11: Diagrama de clases BD tabular [37].

En [33] se propone un modelo de datos que permite entender y comunicar la información al resto de la compañía descripto con la siguiente arquitectura:

1. Diccionario de datos: reconoce todos los atributos que forman parte del proceso que deben ser considerados para la herramienta automatizada en desarrollo [33].
2. Modelo lógico y físico: la distribución sistemática donde se muestra la interrelación de cada atributo mencionado en el diccionario de datos, además de mostrar el origen de cada información [33].
3. Análisis y selección del Servicio de *Cloud Computing* - con el fin de identificar y seleccionar un servicio de *Cloud Computing* adecuado para la implementación del modelo [38].

3.4.4.1. **Spatial Data Mining**

Al aplicar analítica predictiva [39] sobre una base de datos espacial podemos inferir comportamientos, modelos, relaciones y estimaciones para desarrollar predicciones de los mismos. Un proceso *data minin*g implementado sobre una base de datos espaciales puede ser muy costoso si el volumen de información a analizar es muy alto [27].

El proceso de descubrir información en la bse de datos espacial es mucho más complejo que el descubrir patrones en la base de datos relacional, debido a que en la parte espacial se deben tener elementos vecinos ya que algún objeto de estos puede tener una influencia significativa sobre los objetos de análisis y nos podrían dar información muy útil [27].

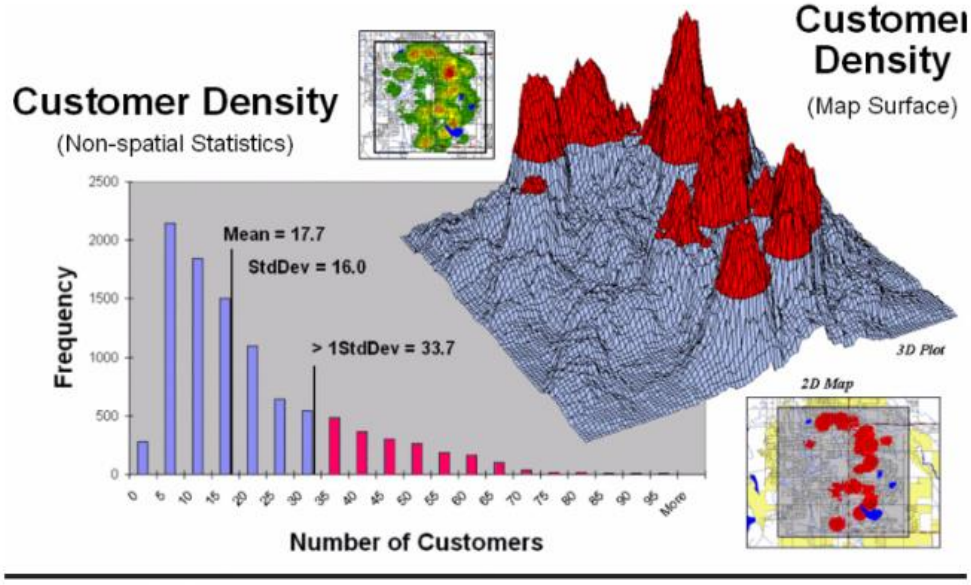

Figura 12: Spatial Data Mining [27].

La detección de tendencias espaciales es otra técnica, significa un cambio regular de uno o más atributos a partir de un objetivo inicial (Fig. 12). La clasificación es otra técnica nos permite asignar un objeto a un determinado grupo basado en los atributos del objeto; uno de los algoritmos utilizados en esta aplicación es el árbol de decisión [27].

3.4.5. Aplicaciones de la integración de BI con GIS en la toma de decisiones en las pequeñas empresas.

La estrategia de GeoMarketing se basa en el uso de tecnologías de mapeo geo-referenciado para identificar patrones de proximidad entre los clientes, proveedores y competidores que interactúan en el ecosistema de un negocio determinado. Estos patrones permiten optimizar las acciones de marketing orientadas tanto a la retención de nuevos clientes como a la fidelización de los existentes. Al utilizar técnicas de análisis espacial y visualización de datos, el GeoMarketing se convierte en una herramienta poderosa para entender las dinámicas de mercado y las interacciones entre los distintos actores que participan en él. La información obtenida a través de estas técnicas permite a las empresas adaptar su estrategia de marketing y tomar decisiones más informadas, lo que se traduce en una mayor eficacia y rentabilidad de sus iniciativas comerciales. En definitiva, el GeoMarketing representa una forma innovadora y efectiva de aprovechar el poder de los datos espaciales para mejorar la competitividad y el rendimiento de cualquier negocio [15].

Tras la implementación del modelo de proceso y gracias al soporte de la visualización del tablero web, la empresa donde se aplicó la solución redujo el tiempo de toma de decisiones en un 60%, con una media de 6 días (rango 3-8 días). Se pudo verificar que los beneficios e indicadores en los procesos utilizando nuevas tendencias y tecnologías aseguran un mayor número de ventas, considerando que con herramientas de análisis masivo de datos [33].

Desde el punto de vista del tratamiento de los datos, la combinación de bases de datos espaciales con bases de datos tabulares en un *Datawareouse* para que luego dicha convergencia pueda ser explotada y visualizada mejora significativamente la velocidad y los procesos de decisión [37].

Integrar en un sistema Inteligencia de negocios con los sistemas de información geográfica nos permite traducir, comunicar y medir los objetivos de las Organización [27].

3.4.6. Ventajas de los sistemas GIS aplicados a la inteligencia de negocios

La combinación de datos espaciales y no espaciales que está diseñada y organizada en detalle y real (sin muestreo) es muy útil en el proceso de toma de decisiones para apoyar la planificación e implementación de programas que requieren la precisión del objeto o destino [37].

En [15] se describe como a través del GeoMarketing se utiliza la ubicación de los clientes, así como la competencia que los rodea, para mapear y, por lo tanto, obtener recursos para asegurar una ventaja competitiva. Entre ellos se encuentra la mayor capacidad para identificar nichos de mercado, oportunidades, perfil público objetivo, comportamiento del consumidor, acciones de competencia y alcance de ventas.

Con el uso de técnicas de modificación genética es posible abordar cuestiones pertinentes como: quién compra, donde comprar y por qué comprar [15].

El mercado ofrece los datos, el comerciante debe entender su significado, lo que lleva a usar GeoMarketing para obtener una ventaja competitiva, ya que la competencia puede no usarlo en el momento actual [15].

3.5. Resumen de los hallazgos

En la actualidad, se dispone de diversas metodologías para abordar el desafío de diseñar soluciones que integren sistemas GIS con inteligencia de negocios. A pesar de las diferencias en los modelos y enfoques, está claro que la combinación de estos sistemas puede aportar importantes beneficios en términos de soporte a la toma de decisiones en la gerencia. Este enfoque representa no solo una ventaja, sino también una oportunidad, especialmente para el sector PyME, que se beneficia de la continua reducción de costos y el advenimiento de numerosas plataformas de corte gratuito.

En el marco de desarrollar procesos específicos de la explotación el trabajo de desarrollar un proceso de recopilación de requisitos de [17] ha sido fundamental.

Los hallazgos encontrados por los distintos autores en este campo de estudio permiten extraer una serie de conclusiones valiosas acerca de las temáticas y metodologías abordadas. A través de un análisis detallado de estas conclusiones, se pueden identificar las mejores prácticas y estrategias para diseñar arquitecturas eficientes que permitan aprovechar al máximo los beneficios de la integración entre GIS y BI.

- Las aplicaciones de BI con un tratamiento GIS y desarrolladas sobre una base de datos tabular son relevantes para ayudar al proceso de toma de decisiones en la planificación, implementación y seguimiento de proyectos mucho más focalizados [11].
- Según Gartner Group, alrededor del 70% al 80% de la información que se puede considerar de relevancia para la toma de decisiones tiene caracterización espacial [7].
- Tanto en [8] como en varios artículos se demostró que el uso de tecnologías como *Cloud Computing* orientado a la pauta de Software Libre puede ser rentable para el sector PyME, dado que el costo de estos servicios está dentro del poder adquisitivo de las pequeñas empresas.

- Se pudo demostrar que una visión de TI aumenta la rentabilidad de una PyME de forma indirecta, ahorrando tiempo y dinero. También proporciona los medios para tomar medidas inmediatas ante una situación inesperada [8].
- La implementación de Inteligencia de Negocios + GIS dentro de la Organización permite desarrollar una mejor relación con los clientes, disponer de servicios eficientes y de alta calidad con el apoyo de la tecnología, bases de datos y sistemas de información y los sistemas de Inteligencia de Negocios + GIS no son inmutables, más bien son dinámicos y deben ser revisados, evaluados y actualizados continuamente [10].

Capítulo 4

4. Análisis y diseño del proyecto

La implementación de sistemas GIS y soluciones de inteligencia de negocios en el sector comercial ha cobrado gran relevancia en los últimos años debido a su capacidad para mejorar la toma de decisiones empresariales. En este sentido, el análisis y diseño de proyectos que integren estas herramientas se convierten en una necesidad para aquellas empresas que buscan obtener una ventaja competitiva en un mercado cada vez más exigente.

El objetivo principal de este trabajo es el desarrollo de una aplicación que combine de manera efectiva un sistema GIS con una solución de inteligencia de negocios, aplicable al sector comercial de venta de insumos y herramientas para la construcción en la ciudad de Posadas, Misiones. Los requerimientos de esta aplicación incluyen la necesidad de diseñar una interfaz gráfica amigable, integrar un sistema de georeferenciamiento, implementar un modelo de datos eficiente, desarrollar herramientas de análisis espacial y generar informes y tableros de inteligencia de negocios claros y concisos. Además, se busca garantizar la seguridad y la integridad de los datos almacenados y proporcionar soporte técnico y capacitación para los usuarios.

Para lograr esto, es necesario comprender a fondo la problemática específica del sector comercial en cuestión, así como también los procesos involucrados en la venta y coordinación de entregas al cliente. En este sentido, se ha diseñado un diagrama en notación BPMN para visualizar de manera clara el flujo de procesos, con el objetivo de optimizar la eficiencia y la satisfacción del cliente en todo el proceso de compra. Con la implementación de esta aplicación, se espera lograr una mejora en la identificación de oportunidades de negocio, así como también en la toma de decisiones informadas y la optimización de procesos empresariales.

4.1. **Requerimientos:**

4.1.1. Requisitos del trabajo

- Lograr un mejor entendimiento de la problemática específica que afecta al sector comercial en cuestión, con el objetivo de identificar oportunidades de negocio que puedan ser aprovechadas.
- Realizar una revisión exhaustiva del estado actual de la literatura en lo que respecta a la aplicación de sistemas de georeferenciamiento en combinación con soluciones de inteligencia de negocios.
- Desarrollar una aplicación que integre de manera efectiva GIS con una solución de inteligencia de negocios, que pueda ser aplicada al sector comercial objeto del estudio.
- Comprobar el impacto de este trabajo y su viabilidad de implementación en la práctica. En cuanto a los requisitos de la aplicación en sí, se deberán tener en cuenta factores como la facilidad de uso, la capacidad de procesamiento de datos, la precisión y la escalabilidad de la solución, entre otros aspectos relevantes.

4.1.2. Requisitos de la aplicación

Los requisitos del desarrollo e implementación del modelo propuesto fueron obtenidos a partid de los requisitos del presente trabajo y en base a trabajos realizados en soporte

de datos para consultoría de negocios durante el periodo 2019-2020 que por cuestiones de privacidad de la información no se consideró pertinente la explicitación de estos.
No es intención de este trabajo profundizar en la obtención de requisitos sino utilizar el negocio de ejemplo para implementar los avances obtenidos durante el proceso de realización del mismo.

- Diseñar una interfaz gráfica amigable que permita a los usuarios acceder y consultar los datos espaciales y los informes generados por la solución de inteligencia de negocios.
- Integrar un sistema de GIS que permita visualizar y analizar los datos espaciales en un mapa interactivo.
- Implementar un modelo de datos que permita integrar y procesar los datos de negocio y espaciales de manera eficiente y efectiva.
- Desarrollar algoritmos y herramientas de análisis espacial que permitan descubrir patrones y relaciones entre los datos de negocio y espaciales.
- Generar informes y tableros que presenten los resultados del análisis de manera clara y concisa, de manera que los usuarios puedan tomar decisiones informadas.
- Garantizar la seguridad y la integridad de los datos almacenados y procesados por la aplicación.
- Desplegar la aplicación en un entorno de producción estable y escalable, para que pueda ser utilizada por un gran número de usuarios de manera simultánea y sin interrupciones.
- Proporcionar soporte técnico y capacitación para que los usuarios puedan aprovechar al máximo las funcionalidades de la aplicación.

4.2. **Recopilación**

Para la recopilación de los requerimientos del proyecto se sigue el ciclo de vida CRISP-DM [10], el cual establece siete etapas claramente definidas para el desarrollo y la implementación de soluciones de inteligencia de negocios en sistemas de almacenamiento y gestión de datos en la nube. Este ciclo de vida es ampliamente utilizado en la industria debido a su enfoque estructurado y sistemático para la resolución de problemas complejos en la gestión de datos.

Las etapas del ciclo de vida CRISP-DM incluyen desde la comprensión del negocio y los objetivos del proyecto, hasta la evaluación de los resultados y la entrega final del sistema. Siguiendo este proceso, podemos asegurarnos de que nuestro proyecto cumpla con los requisitos específicos del negocio y alcance los objetivos previstos de manera eficiente y efectiva.

4.2.1. **Entendimiento del Negocio**

El negocio pertenece al rubro de comercialización de insumos y herramientas para la construcción y cuenta con una ferretería ficticia en Avenida Roque Pérez 1763, Posadas – Misiones.

Con el objetivo de entender mejor los procesos de venta y coordinación de entregas al cliente, se diseñó un diagrama en notación BPMN que ilustra de manera visual el flujo de los procesos involucrados en la venta de productos en mostrador y la posterior coordinación de la entrega al cliente. De esta forma, se busca optimizar la eficiencia y la satisfacción del cliente en todo el proceso de compra como se observa en (Fig. 13).

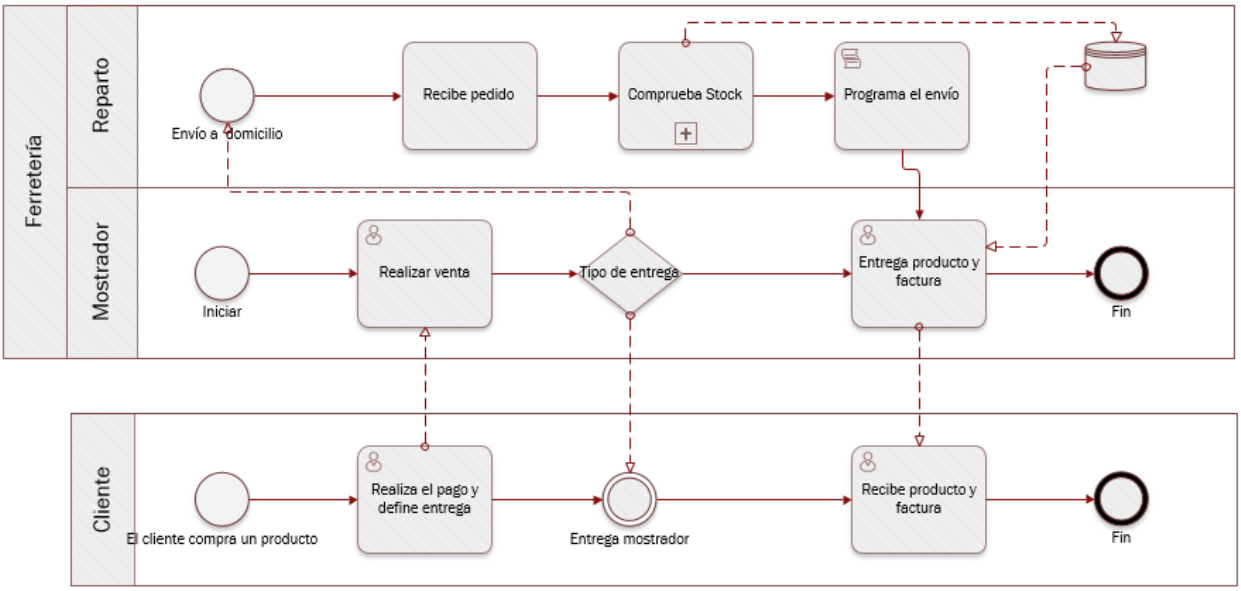

Figura 13: *Diagrama BPMN proceso de venta.*

- Tipo de ventas: Mostrador.
  Las ventas en mostrador de una ferretería suelen ser transacciones rápidas y directas entre el cliente y el vendedor. En general, el cliente llega con una idea clara de lo que necesita y se dirige al mostrador para pedir ayuda al vendedor. El vendedor proporciona información y asesoramiento sobre los productos y, una vez que el cliente ha tomado una decisión, procesa el pago y entrega los productos al cliente. Las ventas en mostrador son una parte fundamental de la operación diaria de una ferretería y requieren un enfoque basado en el servicio al cliente y la eficiencia.

- Tipo de Clientes: Variados, desde clientes ocasionales a grandes clientes comerciales e industriales.
  Los clientes pueden ser muy variados, desde aquellos que compran ocasionalmente para uso personal, hasta grandes clientes comerciales e industriales que requieren grandes cantidades de productos y un servicio personalizado. Los clientes ocasionales suelen llegar al mostrador con una idea general de lo que necesitan y esperan recibir asesoramiento sobre los productos adecuados para su proyecto.
  Por otro lado, los grandes clientes comerciales e industriales, por lo general, tienen necesidades específicas y esperan un alto nivel de servicio y atención personalizada por parte del vendedor. En ambos casos, la interacción comercial entre el cliente y el vendedor en el mostrador es fundamental para el éxito de la venta. El vendedor debe escuchar las necesidades del cliente y brindar una asesoría clara y precisa, ofreciendo opciones de productos y soluciones para el proyecto. El trato personalizado y la atención al detalle son fundamentales para satisfacer las necesidades de los clientes y garantizar su lealtad a la ferretería.

- Productos: gama variada del sector construcción y ferretería.
  Entre estos productos se pueden encontrar materiales de construcción como ladrillos, cemento, hierro, yeso, madera, entre otros. También pueden ofrecer herramientas manuales y eléctricas para la construcción, como taladros, sierras, martillos, y herramientas de medición. Adicionalmente, pueden contar con productos especializados para la construcción, como sistemas de impermeabilización, productos de aislamiento, y elementos de decoración como cerámicas, azulejos y pinturas. Asimismo, pueden ofrecer servicios complementarios, como asesoramiento técnico, entrega a domicilio y financiamiento para sus clientes. En definitiva, una pyme de venta de insumos para la construcción

en el NEA tiene una amplia variedad de productos y servicios para satisfacer las necesidades de sus clientes en el sector de la construcción.

- Tipo de Entrega: por mostrador y envíos a domicilio.
  El proceso de entrega puede variar dependiendo del tipo de entrega que se ofrezca. En el caso de la entrega por mostrador, el cliente suele pagar y recibir sus productos en el momento de la compra, lo que hace que este proceso sea rápido y eficiente. Por otro lado, en la entrega a domicilio, la ferretería suele coordinar con el cliente la fecha y hora de entrega de los productos en su domicilio, asegurándose de que se cumpla con los plazos acordados y que los productos sean entregados en óptimas condiciones. En este tipo de entregas, la distancia entre el punto de salida de envíos y el domicilio medida en km es una variable fundamental que debe tomarse en cuenta al momento de realizar el cálculo de costos y esfuerzo de venta.
  En algunos casos, la ferretería puede contar con un servicio de mensajería o transporte propio o tercerizado para realizar los envíos a domicilio. En cualquier caso, el objetivo de la ferretería es garantizar una entrega segura y oportuna de los productos a sus clientes, ya sea por mostrador o por envío a domicilio, para satisfacer sus necesidades en el sector de la construcción.

4.2.2. **Entendimiento de los Datos**

4.2.2.1. **Orígenes de datos**

Luego de un análisis exhaustivo de las opciones disponibles, se tomó la decisión de utilizar un origen de datos público que incluyera datos ficticios relacionados con una Ferretería. Dicho origen de datos fue descargado de un sitio web confiable, cuya dirección es la siguiente:

https://vdocumento.com/ferreteria-5584510528907.html

Una vez que se contó con el conjunto de datos necesario, se procedió a cruzarlos aleatoriamente con una tabla de zonas que se creó específicamente para este proyecto. Cabe destacar que dicha tabla se basó en la tabla de “Barrios de la ciudad de Posadas”, y se optimizó con el diagrama de zonas de recolección de basura (Fig. 14). Este proceso de integración de datos permitió obtener una fuente de información valiosa y útil para el desarrollo del proyecto de *Business Intelligence* [40].

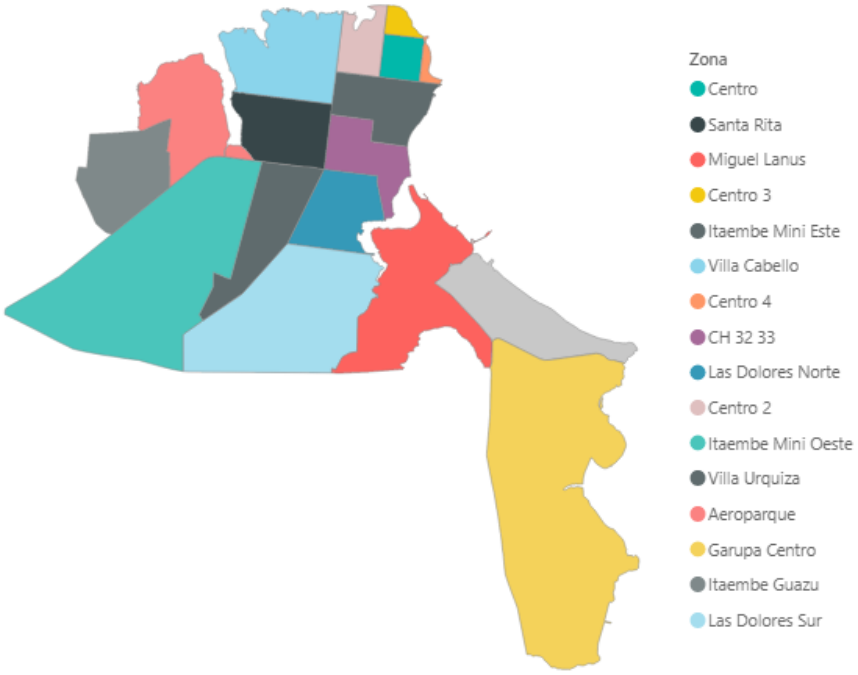

Figura 14: *Mapa de zonas, Ciudad de Posadas* [40].

Se creó una tabla "Geolocalizacion" que cruza los datos de direcciones del *origen de datos* principal con información de API Nominatim de openstreetmap y el cálculo de distancia entre estas coordenadas y las del domicilio de "Casa Central". El aparejamiento de "Zonas" con la tabla principal se realiza mediando la función BUSCARV de la sintaxis DAX (Ecuación 3).

***Ecuación 3**: Busqueda de Clientes por Zona.*

=BUSCARV([@Cliente];Zonas[#Todo];4;FALSO)

Los datos de distancia entre coordenadas se realizaron exportando una nueva tabla desde el Tablero en Power Bi; esta tabla surgió de la consulta entre las gráficas cargadas en el tablero y la función de distancia también desarrollada en el mismo, en este punto hay una doble retroalimentación ya que el tablero utiliza un origen de datos resultante de su mismo proceso en un ciclo de lectura, procedo, exportación y relectura. Cabe destacar que este proceso no es automático.

4.2.3. Refinamiento y Tablas

Para mejorar la calidad y precisión del origen de datos denominado "*FerretariaTesis.xlsx*"; se llevó a cabo un proceso de refinamiento utilizando la herramienta OpenRefine [41] que es una herramienta de código abierto que se utiliza para limpiar, transformar y enriquecer conjuntos de datos. Esta herramienta ofrece una interfaz gráfica intuitiva que permite visualizar y manipular grandes cantidades de datos de manera eficiente. Entre las funciones que ofrece OpenRefine[41] se incluyen la eliminación de duplicados, la corrección de errores de formato, la separación y unión de celdas, la transformación de datos y la realización de operaciones complejas sobre los mismos. Además, permite exportar los datos refinados en diferentes formatos, lo que facilita su integración con otras herramientas de análisis y visualización de datos.

En este proceso se aplicaron diversas acciones para optimizar el conjunto de datos, entre las que se incluyen:

- Se eliminaron los campos nulos.
- Se corrigieron los formatos y el tipo de valores.
- Se corrigió el campo fecha para que sea compatible con el formato de Power Query.
- Se reemplazaron valores inválidos.
- Se eliminaron duplicados y campos sin correlación.

De esta manera, se logró obtener un conjunto de datos más limpio, completo y adecuado para su posterior análisis y utilización en la toma de decisiones. Cabe destacar que el refinamiento de los datos es una tarea fundamental en cualquier proyecto de análisis de datos, ya que permite maximizar la calidad y confiabilidad de la información disponible.

Del archivo resultante se descompuso en 5 hojas de cálculo:

Ventas
Zona
Factura
Fecha
Cliente
Codigo
Producto
Cant
M3
V.bruta
Descuento
V.Neta
Costo

Geolocación
Cliente
Sucursal
Domicilio
Zona
Distancia a Casa Central

Zona
Zona
Idzona

Clientes
Clientes

Casa Central
Sucursal
Domicilio
Zona

4.2.4. **Preparación de los Datos**

Durante el proceso de desarrollo del proyecto, se llevó a cabo la obtención de datos utilizando la herramienta de Power Query. A partir de allí, se procedió a realizar la consulta correspondiente a las tablas previamente referidas y se llevó a cabo la creación y modelado de diversas tablas y columnas según las necesidades del proyecto. Este proceso se llevó a cabo mediante el uso de técnicas y metodologías específicas como el refinamiento de datos utilizando OpenRefine[41], eliminando datos redundantes, aplicando el formato correcto para cada tipo de dato, revisando la indexación de las tablas y las conexiones de todos los orígenes de datos; permitiendo así obtener resultados precisos y satisfactorios en todo momento.

Cabe destacar que la utilización de Power Query resultó ser altamente efectiva y versátil, permitiendo llevar a cabo tareas complejas de manera eficiente y eficaz.

4.2.4.1. **Función de Geocodificación:**

Durante la fase de desarrollo del proyecto, se procedió a la creación y diseño de un algoritmo de consulta y extracción de datos en formato *json* o *xml*, el cual se llevó a cabo utilizando la función "GET" perteneciente al lenguaje de consulta (M) que es el lenguaje que utiliza Power Query para poder conectar, combinar y refinar los orígenes de datos y así dejar una base de datos limpia que permita extraer y analizar la información que se desee. Esta función fue utilizada con el objetivo de llevar a cabo una llamada efectiva a la API de Nominatim, la cual se puede observar detalladamente en (Fig. 15). Posteriormente, se llevó a cabo una tarea esencial que consistió en reemplazar la variable previamente asignada, conocida como "*address*", por las diversas direcciones ya conocidas que se encontraban en la columna correspondiente a los clientes. Dicha tarea se logró gracias a la funcionalidad y versatilidad del algoritmo antes mencionado.

De esta manera, se logró optimizar y mejorar significativamente la eficiencia en el manejo y procesamiento de los datos involucrados en el proyecto en cuestión, permitiendo obtener resultados precisos y satisfactorios en todo momento.

```
let
    AddressToLocate = (address as text) =>
let
    Origen = Json.Document(Web.Contents("https://nominatim.openstreetmap.org/search.php?q="&address&"&format=jsonv2")),
    Origen1 = Origen{0},
    #"Convertido en tabla" = Record.ToTable(Origen1),
    #"Filas filtradas" = Table.SelectRows(#"Convertido en tabla", each ([Name] = "lat" or [Name] = "lon")),
    #"Tabla transpuesta" = Table.Transpose(#"Filas filtradas"),
    #"Encabezados promovidos" = Table.PromoteHeaders(#"Tabla transpuesta", [PromoteAllScalars=true]),
    #"Valor reemplazado" = Table.ReplaceValue(#"Encabezados promovidos",".",",",Replacer.ReplaceText,{"lat"}),
    #"Valor reemplazado1" = Table.ReplaceValue(#"Valor reemplazado",".",",",Replacer.ReplaceText,{"lon"}),
    #"Tipo cambiado" = Table.TransformColumnTypes(#"Valor reemplazado1",{{"lat", type number}, {"lon", type number}})
in
    #"Tipo cambiado"
in
    AddressToLocate
```

Figura 15: Código de Función llamada a api.

En la Tabla “Geolocación” se agregaran 2 columnas que realizaran la invocación a la función personalizada antes descripta.

4.2.4.2. **Modelo de datos en Power BI**

A fin de ilustrar de manera más precisa el proceso de consumo de datos en Power BI, se presenta a continuación una descripción detallada del modelo de datos resultante. Este modelo es el resultado de la ejecución del motor de consulta de Power Query, como se puede observar en (Fig. 16).

El origen de datos se somete a un proceso riguroso para garantizar la calidad y consistencia de los datos, y se transforma en una estructura de tablas optimizada para su uso en el tablero de BI. Este proceso incluye la eliminación de datos redundantes, la corrección de errores y la integración de múltiples fuentes de datos. El resultado final es un modelo de datos sólido y bien estructurado, que permite la exploración y visualización efectiva de la información.

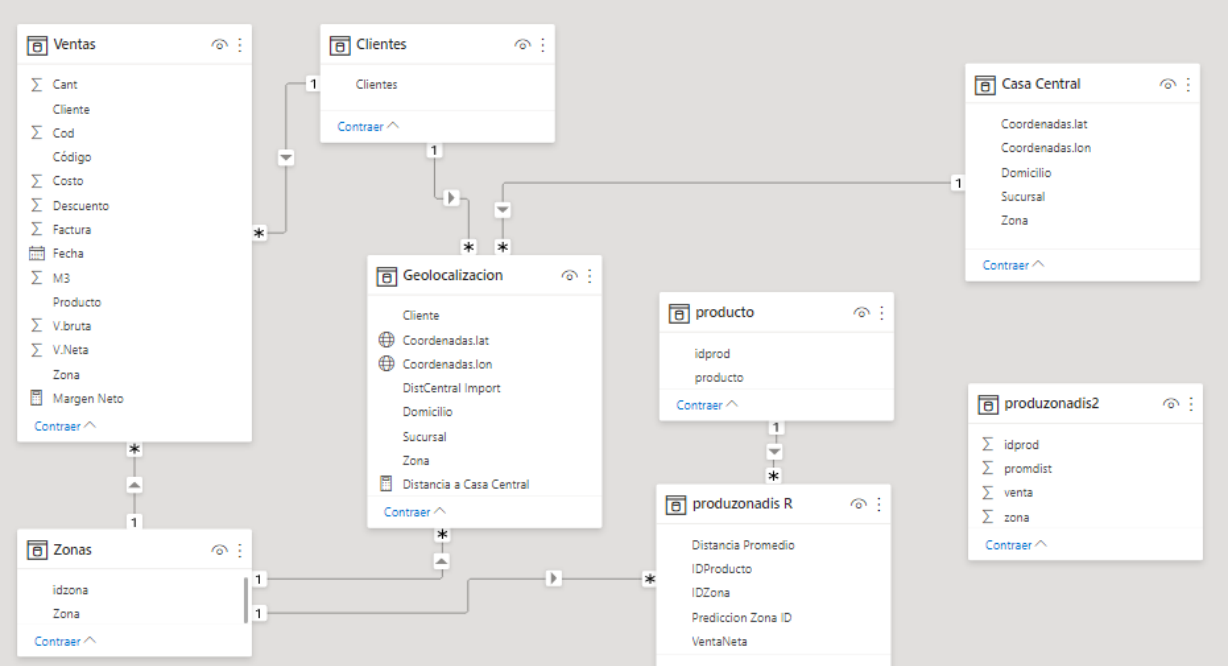

Figura 16: Modelo de datos en Power BI.

4.2.5. **Modelado**

El modelado de una aplicación en Power BI es un proceso crucial para la implementación de un tablero de *Business Intelligence* exitoso. Durante este procedimiento, el motor de consulta de Power Query desempeña una función esencial al cargar en la plataforma las fuentes de datos requeridas, como por ejemplo el archivo denominado "FerretariaTesis.xlsx" y la tabla titulada "Zona.csv". Una vez que se cuenta con estos conjuntos de datos, se procede a crear una estructura tabular específica diseñada para facilitar la visualización y el análisis efectivo de la información. Además, se utiliza la columna "Domicilio" de la tabla Geolocalización para realizar la geocodificación a través de la API de OpenStreetMap, lo que permite calcular la distancia a la sucursal central y medir el esfuerzo / beneficio de cada cliente.

Los datos resultantes se exportan a una nueva tabla "exportados.csv" que luego se utiliza en modelos de aprendizaje automático para determinar las zonas en las que es más conveniente ofertar distintos productos y a qué clientes ofrecer mejores descuentos. Todo este proceso se integra en un tablero de Power Bi configurado para su publicación a través del servicio web de Power BI, lo que permite a los usuarios realizar visualizaciones en tiempo real tanto en la web como en la aplicación de Android.

4.2.5.1. **Arquitectura del tablero en Power BI**

En (Fig. 17) se representa gráficamente la arquitectura del proyecto, que consiste en un tablero de Business Intelligence (BI) desarrollado en Power BI. Es importante destacar que el motor de consulta de Power Query juega un papel fundamental en este proceso, ya que se encarga de subir a la plataforma las Fuentes de Datos "FerretariaTesis.xlsx" y "Zona.csv".

Una vez que estos datos están disponibles, el motor de consulta genera la estructura de tablas que se muestran en (Fig. 16). Estas tablas están cuidadosamente diseñadas para permitir la visualización y el análisis eficiente de la información, lo que contribuye significativamente al éxito del proyecto.

De la columna "Domicilio" de la tabla Geolocalización se genera la consulta a la API de OpenStreetMap para realizar el proceso de geocodificación, luego de este llamado a la API se crea una función para generar 2 nuevas columnas en la tabla con las coordenadas correspondientes a latitud y longitud de cada dirección conocida, luego estas coordenadas geográficas serán utilizadas para realizar el cálculo de distancias a Sucursal central. Con el Origen de datos "Zona.csv" se generara en www.geojson.io una tabla acompañada de un archivo Json que servirá de fuente para levantar un plano de capas y que este coordine los datos con la base de datos de los clientes.

De los datos resultantes dentro del tablero se exporta una nueva tabla "exportados.csv" que contiene los datos de ventas de los clientes sumado al cálculo de distancia a sucursal, esto último es importante y a que es la variable con la cual luego se medirá el esfuerzo / beneficio de cada cliente para determinar a través de los modelos de aprendizaje automática en que zonas será más conveniente ofertar distintos productos ya que clientes ofrecerles mejores descuentos. Estos cálculos se vuelven a insertar en el tablero a través de 2 gráficas "Árbol y Neural Net". El tablero ya configurado se publica a través del servicio web de Power BI que proveerá luego 2 interfaces, Web y App Android lo que permitirá a los usuarios conectados realizar las visualizaciones de los datos del tablero en tiempo real.

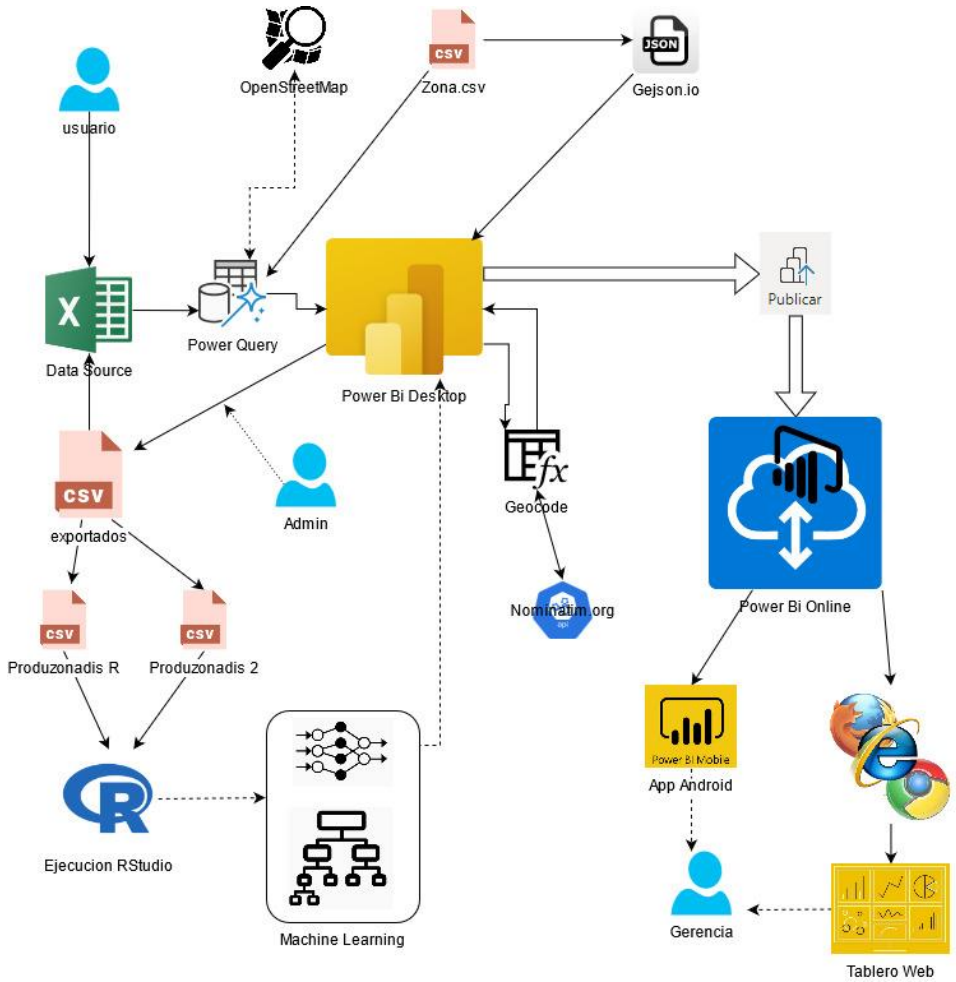

Figura 17: *Arquitectura del Tablero FerreteriaBi.*

4.2.5.2. Tablas auxiliares

Estas tablas se utilizaran para almacenar información adicional que no se encuentra directamente relacionada con la tabla principal de datos, pero que se requiere para enriquecer o mejorar el análisis y la visualización de los datos. Estas tablas pueden contener datos como listas de valores, códigos, descripciones, categorías, fechas y otros elementos que permiten agregar contexto y detalle a los datos principales.

Las tablas auxiliares se crearon en Power BI utilizando el editor de consultas y también mediante la importación de archivos externos, de esta manera estas tablas se pueden utilizar para enriquecer los gráficos, las tablas dinámicas y otros elementos visuales, así como para filtrar y segmentar los datos de manera más precisa y efectiva.

4.2.5.2.1. Capas de GeoJson e integración con el Tablero

A través del uso de la herramienta en línea GeoJson.io, se llevó a cabo la creación de un archivo de datos geoespaciales denominado "Zonas Posadas.geojson". Posteriormente, se procedió a su transformación al formato JSON con el fin de que pudiera ser interpretado y visualizado mediante el visualizador de mapa de capas presente en el tablero.

Es importante destacar que la elaboración de este archivo conllevó el procesamiento de una gran cantidad de información, incluyendo cientos de puntos georeferenciados que se encuentran interconectados en un archivo plano (Fig. 18). De esta manera, se permitió la creación de distintas capas que a su vez cuentan con atributos específicos para cada uno de los puntos geográficos presentes en el archivo.


```
{"type":"Topology","arcs":[[[8689,14757],[889,-100],[34,-86],[-6,-40],[-20,-28]
13751],[-977,129],[107,1003],[57,397],[104,-19],[-14,-40],[135,-10],[9,27],[40,
-15],[5,-23],[-3,-27],[-11,-16],[-14,-20],[-10,-21],[-17,-20],[-7,-20],[-7,-20]
,-20],[9,-5],[12,-2],[8,-4],[18,-6],[3,3],[11,3],[8,1],[6,14],[5,12],[11,11],[1
],[24,11],[26,7],[23,0],[28,-2],[18,-5],[28,-8],[21,-7],[7,1],[75,87],[41,119],
],[-5,11],[0,5],[-34,5],[38,0],[8,9],[13,4],[30,6],[47,6],[121,9],[132,-4],[63,
14658],[-890,102],[79,661],[73,-6],[55,0],[2,8],[17,-8],[8,-1],[13,-3],[7,-1],[
-31],[30,-19],[46,-32],[32,-34],[15,-46],[7,-21],[10,-30],[18,-25],[37,-47],[25
],[6,-15],[2,-10],[1,-16],[-1,-20],[0,-18],[3,-10],[7,-3],[3,-4],[-4,-5],[10,-1
13,-22],[2,-4],[3,-2],[7,-1],[1,-5],[-4,-4],[1,-6],[4,-7],[11,-12],[26,-17],[16
2,-3],[2,-7],[7,-5],[8,-5],[7,-7],[5,-2],[-21,-27]],[[9493,13639],[95,865],[16,
```


Figura 18: *fragmento del código Json para el mapa de capas.*

4.2.6. Fase final: Evaluación y Despliegue

El desarrollo de esta fase final del modelo se concluye en el **Capítulo 5** de este trabajo.

Capítulo 5

5. Implementación

5.1. **Contexto y motivación**

En los sectores comerciales la palabra precisión implica cierto grado de tecnificación, esto sumado al crecimiento y la rápida adopción de sistemas GIS implementados para mejorar los sistemas productivos [42]. Es por ello que el posicionamiento además de ser una disciplina importante en esta era digital también es parte esencial de nuestra vida cotidiana [43]. En los últimos tiempos el sistema de posicionamiento global (GPS) como medio de obtención de datos rápido, eficaz y relativamente barato ha ido aumentando [44].

Dentro de este nuevo concepto de datos aplicados a la producción han aparecido numerosas plataformas digitales de acceso libre y gratuito que ofrecen todo tipo de soluciones e información para las más diversas aplicaciones; entre ellas se destacan las plataformas de crowdsourcing mejor conocidas como Información Geográfica Voluntarias o por sus siglas en inglés (VGI) [45], el ejemplo más popular de VGI que podemos citar es el servicio OpenStreetMap (OSM), una base de datos geográfica mundial de código abierto [45].

En este caso de estudio la observación realizada de los datos geográficos serán analizados cuatro plataformas GIS:

- Google Maps.
- Bing Maps.
- OpenstreetMaps (sus 2 API, Nominatim y OpenCageData).
- Here.

Para el análisis de las distancias se utilizaron dos fórmulas, inicialmente la fórmula del haverseno [46] para calcular la distancia entre los puntos y así determinar la precisión de cada API, luego se aplicó el cálculo de distancia euclidiana entre dos puntos [47]. Los resultados y el análisis de ambas fórmulas aplicadas a cada motor geocodificador fueron volcados en tablas comparativas a fin de que puedan apreciarse las claras diferencias entre los métodos y las plataformas.

Se realizó la combinación de un trabajo de campo junto con un trabajo de desarrollo e implementación de una interfaz de análisis con Lenguaje M de consulta y código DAX en Power Bi para los datos recopilados y la información resultante de este trabajo.

5.2. **Trabajo de Campo**

Se utilizó un GPS Garmin NUVI 755T Car GPS para tomar las coordenadas geográficas de 10 puntos aleatorios de la Ciudad de Posadas cuya dirección (calle y altura) fueron visiblemente fáciles de reconocer. Los datos obtenidos fueron volcados a una planilla en Excel conservando la relación entre las coordenadas tomadas por el GPS y la dirección observada en el punto elegido para que puedan ser consumidos por Power Bi.

5.3. **Ejecución del proyecto**

Desde Power BI se utilizaron cuatro plataformas para geo localizar y visualizar las direcciones obtenidas, esto permitió conocer la interfaz GIS de cada plataforma y cómo resuelve la geolocalización de cada dirección. La primera comparativa se realizó sobre las coordenadas obtenidas en la (Fig. 19).

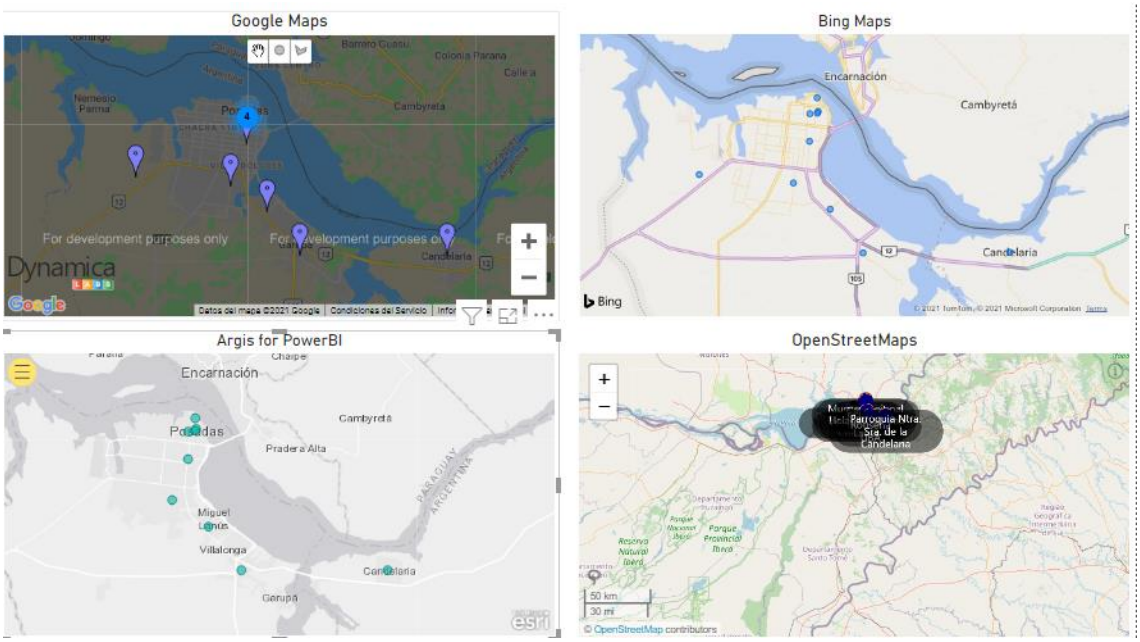

Figura 19: *Comparativa de Visualización de mapas.*

5.4. **Pruebas de precisión**

Se realizó una función de geocodificación de distintas direcciones conocidas de la Ciudad de posadas para medir la variación (en kms) entre la localización real y la resuelta por la API de cada plataforma:

https://api.opencagedata.com/
https://maps.googleapis.com/
http://dev.virtualearth.net
https://nominatim.openstreetmap.org/

Para ello se tomaron diez ubicaciones conocidas con sus respectivas direcciones postales y se anotó la Geolocación con un dispositivo GPS. De estas direcciones postales se realizó la llamada a las API de cada motor antes mencionado para realizar el proceso de Geocodificación, esto es, convertir las direcciones postales en coordenadas geográficas (latitud y longitud). Esto es realizando una llamada a API mediante CURL que devuelve dependiendo el caso un json o un xml. Tanto el json como el xml resultante se procesan con lenguaje de consulta M dentro de Power Query y se realiza el refinamiento de los campos para que puedan ser consumidos por la plataforma Bi, (Fig 20) y (Fig 21).

```
JSON   Datos en bruto   Encabezados
Guardar  Copiar  Contraer todo  Expandir todo  Filtrar JSON
documentation:                  "https://opencagedata.com/api"
licenses:                       [...]
rate:                           {...}
results:
  0:
    annotations:
      DMS:
        lat:                    "27° 21' 53.00100'' S"
        lng:                    "55° 53' 37.11588'' W"
      MGRS:                     "21JXK0941472681"
      Maidenhead:               "GG22bp22sl"
      Mercator:
        x:                      -6222051.909
        y:                      -3149477.995
      OSM:                      {...}
      UN_M49:                   {...}
      callingcode:              54
      currency:                 {...}
      flag:                     "🇦🇷"
      geohash:                  "6g0k247ktyzfw1kv449"
      qibla:                    73.08
      roadinfo:
        drive_on:               "right"
        road:                   "Félix de Azara"
```


Figura 20: *Ejemplo Rest API Json.*


```
−<Location>
  −<Name>
      Calle Félix de Azara 1552, Centro, N3300LQB Posadas, Misiones, Argentina
    </Name>
  −<Point>
      <Latitude>-27.3647714</Latitude>
      <Longitude>-55.8933775</Longitude>
    </Point>
  −<BoundingBox>
      <SouthLatitude>-27.368634117570675</SouthLatitude>
      <WestLongitude>-55.899176532765857</WestLongitude>
      <NorthLatitude>-27.360908682429322</NorthLatitude>
      <EastLongitude>-55.887578467234142</EastLongitude>
    </BoundingBox>
    <EntityType>Address</EntityType>
  −<Address>
      <AddressLine>Calle Félix de Azara 1552</AddressLine>
      <AdminDistrict>Misiones</AdminDistrict>
      <AdminDistrict2>Departamento Capital</AdminDistrict2>
      <CountryRegion>Argentina</CountryRegion>
    −<FormattedAddress>
        Calle Félix de Azara 1552, Centro, N3300LQB Posadas, Misiones, Argentina
      </FormattedAddress>
      <Locality>Posadas</Locality>
      <PostalCode>N3300LQB</PostalCode>
    </Address>
```


Figura 21: *Ejemplo Rest API xml.*

Uno de los inconvenientes que se presentaron con las direcciones obtenidas es el problema de la semántica y la falta de flexibilidad en términos de lenguaje natural de las plataformas, simplemente con pequeños cambios como escribir A. en vez de Avenida o viceversa puede provocar que la API directamente no resuelva la localización; incluso se detectaron problemas con los signos de puntuación. Esto se convirtió

en un desafío ya que para poder realizar la comparativa, las direcciones levantadas debían ser resueltas por las cinco plataformas GIS sin devolver mensaje de error (incluso si las resolvieran erróneamente como se podrá visualizar) en (Fig. 22) y (Fig. 23).

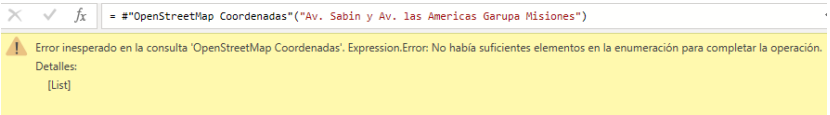

Figura 22: *Error de Nominatim al resolver dirección.*

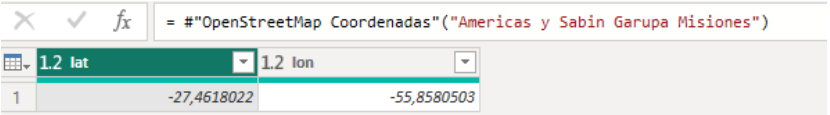

Figura 23: *Dirección Geolocalizada correctamente.*

Una vez realizada la Geocodificación se obtiene dos puntos geográficos relacionados a una dirección postal, la prueba de precisión de API consiste en medir la distancia entre el punto real medido con dispositivo de GPS y el punto resultante de la geocodificación; el valor resultante se tomara como el error medido en kilómetros (km).

Para medir este error se utilizaron dos fórmulas de medición de distancia; el método de cálculo Euclidiano de distancia entre dos puntos que es una función heurística que se obtiene a partir de la distancia directa sin obstáculos con la que se obtiene la longitud de una línea diagonal en un triángulo y la fórmula del Haverseno que es una ecuación que calcula la distancia de un arco entre dos puntos de longitud y latitud [43].

Formula del Haverseno: Para calcular la distancia entre 2 puntos geográficos se utilizó la fórmula del haverseno como se observa en (Ecuación 3):

Ecuación 4: *Calculo Haverseno.*

$$P = \Pi/180$$

$$A = 0{,}5 - \cos((Lat1 - Lat2) * P)/2 + \cos(Lat1 * P) * \cos(Lat2 * P) x (1 - \cos((LNG2 - LNG1) * P))/2$$

$$D = 12742 * \arcsin(\sqrt{A})$$

Donde P es el resultado de dividir Pi para generar un semicírculo y A es la distancia entre 2 puntos sobre un circulo máximo de la esfera. Lo que permite resolver D (distancia) multiplicando el doble del radio terrestre por la tabla inversa del arcoseno de la raíz cuadrada de A.

La implementación de estas formula matemática se realizó a través de Power Bi en lenguaje de fórmulas para el análisis de datos (DAX), primeramente convirtiendo las coordenadas en variables visualizado en (Fig. 24):

```
var Lat1 = MIN( ’ Direcciones Conocidas ’ [ OpenStreetMap  Latitud ] )
var Lng1 = MIN( ’ Direcciones Conocidas ’ [ OpenStreetMap  Longitud ] )

var Lat2 = MIN( ’ Direcciones Conocidas ’ [ Latitud  Real ] )
var Lng2 = MIN( ’ Direcciones Conocidas ’ [ Longitud  r e a l ] )
```

Figura 24: *Conversión de coordenadas a variables.*

Se realiza el cálculo de distancia entre los puntos (Fig. 25):

```
var P = DIVIDE( PI(), 180 )
var A = 0.5 - COS((Lat2-Lat1) * p)/2 +
    COS(Lat1 * p) * COS(lat2 * P) * (1-COS((Lng2- Lng1) * p))/2
var final = 12742  * ASIN((SQRT(A)))
return final
```

Figura 25: *Calculo de distancia.*

De la aplicación del algoritmo con la fórmula del haverseno resulta una tabla con las distancias calculadas entre los puntos tomados con el GPS y los puntos geocodificados con cada API.

Calculo de distancia euclidiana: Se realizó el cálculo de la distancia entre los puntos geográficos tomados con el GPS y los puntos geocodificados aplicando la fórmula en (Ecuación 4):

Ecuación 5: *Calculo de Distancia Euclidiana.*

$$d_E(P,Q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^{n}(p_1 - q_1)^2}$$

La fórmula se implementa en lenguaje de Análisis DAX desarrollando una nueva variable con la distancia calculada (Fig. 26):

```
Distancia Euclidiana = SQRT(POWER( ’Direcciones Conocidas ’[Latitud Real] – ’Direcciones
    Conocidas ’[Latitud Real], 2) + POWER( ’Direcciones Conocidas ’[Longitud real] – ’
    Direcciones Conocidas ’[OpenCageData Longitud], 2)) * 110.574
```

Figura 26: *Código DAX de cálculo Euclidiano.*

5.5. **Resultados**

El resultado del desarrollo de esta aplicación fue la diferencia en kilómetros entre el punto (latitud y longitud) tomado con el GPS y el punto (latitud y longitud) generado a partir de la Geolocación de la dirección asociada al punto tomado con el GPS; con estos datos se elaboró una tabla con las variaciones correspondientes aplicando la fórmula del Semiverseno (Ecuacion 4) a cada API como se detalla en (Tabla 4).

Tabla 4: *Variación en Kms.*

	Google	OpenCageData	OpenStreetMap	Here	Bing
Americas y Sabin Garupa Misiones	0, 005	0, 05	0, 05	2, 32	2, 68
Av. Cocomarola 7205 Posadas Misiones	0, 052	0, 024	1, 147	0, 06	0, 16
Av. Juan Domingo Perón 1915 Posadas Misiones	0, 012	0, 01	0, 01	0, 01	290, 96
Av. República Oriental del Uruguay 3915 Posadas Misiones	0, 004	0, 012	1, 21	0, 03	0, 11
Avenida San Roque González 800 Candelaria Misiones	0, 195	0, 032	0, 032	0, 03	1, 01
Bolivar 2376 Posadas Misiones	0, 009	0, 007	0, 007	0, 02	0, 02
Félix de Azara 1552 Posadas Misiones	0, 026	0, 032	0, 032	0, 02	0, 01
Los Lirios y Las Calandrias Itaembe Guazú Misiones	0, 109	0, 034	0, 034	242, 51	0, 02
Roque González 698 Posadas Misiones	0, 086	0, 079	0, 079	0, 18	0, 18
San Martin 1846 Centro Posadas Misiones	0, 042	0, 005	2, 643	0, 02	146, 15
Total	0, 54	0, 285	5, 244	245, 2	441, 3

Se pudo observar que OpenCageData y Nominatim son las API con menos distancia de error entre los puntos reales y los puntos geolocalizados generados por la API (Tabla 5).

Tabla 5: Exactitud de geolocalización.

	Google	OpenCageData	OpenStreetMap	Here	Bing
Total	0, 54	0, 285	5, 244	245, 2	441, 3
Media	0, 054	0, 0285	0, 5244	24, 52	44, 13
Mediana	0, 034	0, 028	0, 042	0, 03	0,17

Como resultado de la aplicación del cálculo de la distancia euclidiana (Ecuación 7) entre los puntos reales y los geocodificados con cada API se elaboró la (Tabla 6) con las distancias calculadas y la variación en kilómetros del error:

Tabla 6: Variación en Kms, Distancia euclidiana.

	Google	OpenCageData	OpenStreetMap	Here	Bing
Americas y Sabin Garupa Misiones	0,0113	0,1118	0,118	7,4124	5,7435
Av. Cocomarola 7205 Posadas Misiones	0,539	0,264	1,1429	0,0620	0,1596
Av. Juan Domingo Perón 1915 Posadas Misiones	0,0121	0,0104	0,0104	0,0120	301,8277
Av. República Oriental del Uruguay 3915 Posadas Misiones	0,0042	0,0127	1,2056	0,0262	0,1087
Avenida San Roque González 800 Candelaria Misiones	0,1941	0,0320	0,0320	0,0303	1,0078
Bolívar 2376 Posadas Misiones	0,0090	0,0067	0,0067	0,0178	0,0174
Félix de Azara 1552 Posadas Misiones	0,0275	0,0357	0,0357	0,0163	0,0064
Los Lirios y Las Calandrias Itaembé Guazú Misiones	0,4340	0,1542	0,1542	979,1726	0,0938
Roque González 698 Posadas Misiones	0,1721	0,1583	0,1583	20,3425	0,3581
San Martín 1846 Centro Posadas Misiones	0,0422	0,0056	2,9206	0,0201	156,2909
Total	0,9603	0,5528	5,7773	1007,1122	465,6139

La fórmula del Haverseno o Semiverseno puede acarrear errores en la medición incluso superiores al 3% es por ello que se realizó el cálculo utilizando la fórmula de distancia euclidiana que puede dar una precisión arriba de los 0.5km (Tabla 6):

Tabla 7: Exactitud de geolocalización, Distancia euclidiana.

	Google	OpenCageData	OpenStreetMap	Here	Bing
Total	0,9603	0,5528	5,7773	1007,1122	465,6139
Media	0,0960	0,0553	0,5777	100,7112	45,5614
Mediana	0,0349	0,0292	0,01325	0,0283	0,2588

5.6. **Discusión**

Si bien la zona elegida para el estudio está situada en un conglomerado urbano relativamente grande, no hay una actualización continuada en el tiempo de los datos geoespaciales por parte de las plataformas GIS de acceso gratuito, se encuentran muchos faltantes en la información lo que debilita la calidad de los procesos de geolocalización de un punto determinado. En este sentido las plataformas de *Crowdsourcing* tienen cierta ventaja respecto a estos faltantes de datos ya que cuentan con la colaboración de usuarios locales que aportan esos datos.

Aun así, es común encontrar calles y rutas que no tienen información de altura (N° de domicilio) lo que hace imposible a la API resolver ese punto y localizarlo. Esto se evidencia en los grandes saltos en km en alguna dirección; este error generalmente está relacionado con la lejanía al centro urbanístico o conglomeración urbana.

En esta investigación se omitieron los puntos mal resueltos ya que se consideró irrelevante evidenciar esta problemática.

5.7. **Análisis de los resultados**

De estas pruebas se concluyó utilizar OpenCageData ya que fue la Api que demostró mejores métricas de precisión en el proceso de Geocodificación.

Los métodos de medición de distancia entre puntos georreferenciados utilizaron, demostraron ambos funcionar correctamente y en términos de precisión no hubo grandes diferencias, a favor de la formula Euclidiana se debe destacar la simpleza del código y menor costo computacional, especialmente si se trabaja con grandes volúmenes de datos puede ser una ventaja importante. Los errores en la geocodificación se deben a los datos faltantes por parte de las plataformas GIS como se observa en (Tabla 5) y (Tabla 7), en este punto del planeta donde incluso las VGI muestran baches en la información impidiendo a las API resolver algunas direcciones, quedando claro que éstas siguen siendo plataformas con información más detalladas que las demás.

5.8. **Sumario de apreciaciones:**

5.8.1. **Googgle Maps.**

- **Sumario**
  - Positivo:
    - Controles en la interface.
    - Clusteriza.
  - Negativo
    - Sin zoom automático
    - Marca de agua
    - No tiene precisión (requiere coordenadas)
    - Visualización poco atractiva.

5.8.2. **BingMaps.**

- **Sumario**
  - Positivo:
    - Visualización atractiva y sencilla.
    - Zoom automático.
    - Es parte de Power Bi.
  - Negativo:
    - No tiene precisión (requiere coordenadas)

5.8.3. **OpenStreetMaps** (OpenCageData y Nominatim).

- **Sumario**

- Positivo:
  - Gran cantidad de datos.
  - Buen nivel de granularidad.
  - Compatibilidad.
- Negativo:
  - No tiene precisión (requiere coordenadas)
  - Sin zoom automático
  - Pocos controles en el tablero
  - Las etiquetas quedan poco claras y son confusas.

En las pruebas realizadas para determinar la eficacia de varias API geocodificadoras, se pudo observar que OpenCageData y Nominatim demostraron una resolución de ubicación satisfactoria, mientras que Google no logró cumplir con los requisitos de precisión esperados. A pesar de que ambas API pertenecen al motor geocodificador de OpenStreetMap, se decidió optar por https://opencagedata.com/ debido a sus mejores cualidades en términos de compatibilidad con Power Bi.

Tras un análisis exhaustivo de las métricas de precisión de geocodificación y geocodificación invertida, así como la facilidad de gráfica y manejo de datos, se concluyó que OpenCageData fue la mejor opción para el proyecto. Además, su compatibilidad con las herramientas utilizadas en este proyecto demostró ser óptima, lo que asegura una integración eficiente y eficaz en la gestión de datos. En resumen, la elección de OpenCageData como la API geocodificadora para este proyecto se basó en su mejor capacidad para resolver una ubicación, facilidad de manejo y compatibilidad con las herramientas utilizadas en el proceso de análisis de datos.

En (Fig. 27) se tomó un domicilio, en este caso el del Edificio de Facultad de Ciencias Exactas, Químicas y Naturales de la Universidad nacional de Misiones en Felix de Azara 1552 para graficar como resuelve cada API la ubicación del punto geocodificado, el punto azul representa la dirección en el mapa:

Figura 27: *Comparativa – real contra puntos geocodificados.*

5.9. **Evaluación y Despliegue**

5.9.1. **Análisis de distancias con ml (R)**

En base al entrecruzamiento de datos dentro del tablero y al resultado del cálculo de distancia entre los clientes y la sucursal central se exportaron dos tablas en formato csv que luego serían convertidas a xls:

- Produzonadis2.csv con los campos: producto, venta, zona, promdist, idprod.
- Produzonadis3.csv con los campos: venta, zona, promdist, idprod.

Finalmente para el estudio de un modelo de aprendizaje de máquina que sea útil se optó por utilizar produzonadis3.csv ya que el campo alfanumérico "producto" producía errores al correr el algoritmo.

5.9.2. **Leyenda de columnas**

- Venta: suma de ventas netas por producto, por zona en este intervalo de tiempo.
- Zona: zona, barrio donde se desea segmentar el valor de venta.
- Promdist: distancia promedio a sucursal central de los clientes dentro de cada zona.
- Idprod: código de cada producto.

5.9.3. **Construcción del predictor**

Se buscó elaborar un modelo para responder la siguiente pregunta:

¿Qué zonas serán mejores para vender un determinado producto en relación a la ganancia (Venta Neta) con el esfuerzo?

Entendiéndose como esfuerzo a la menor o mayor distancia promedio a nuestra sucursal central desde cada zona.

Para la construcción del predictor se utilizaron 2 modelos que arrojaron resultados.

Con el objetivo de llevar a cabo las tareas encomendadas en este proyecto, se hizo uso de las librerías disponibles en el paquete NeuralNetTools, el cual provee un conjunto de funciones a desarrollar en el contexto de esta investigación. Dentro de estas funciones, se encuentra plotnet, la cual permite trazar un diagrama que facilita la interpretación de la red neuronal, con el fin de evaluar la importancia de la variable. En su gran mayoría, estas funciones requieren la extracción de las ponderaciones del modelo en un formato estandarizado para las distintas clases de objetos que se manejan en la red neuronal en R. Una de estas funciones es la función de ponderaciones neuronales, que se emplea para recuperar las ponderaciones del modelo, independientemente del tipo de modelo que se necesite.

5.9.4. **Tablero**

La elaboración de un tablero de mando en Power BI con información geocodificada combinada con datos de negocio de una pequeña y mediana empresa (pyme) puede proporcionar una serie de ventajas significativas. En primer lugar, al incorporar información geográfica, se pueden visualizar datos de una manera mucho más intuitiva, lo que permite al usuario obtener información valiosa de forma más clara y

rápida. Además, la capacidad de Power BI para realizar análisis de datos sofisticados, como la segmentación de clientes por ubicación geográfica, puede ayudar a la pyme a identificar oportunidades de mercado que podrían haber pasado desapercibidas de otra manera. En resumen, la utilización de Power BI para desarrollar un tablero de mando con información geocodificada y de negocio puede proporcionar una herramienta poderosa para la toma de decisiones informadas y la identificación de oportunidades de negocio (ver código en ANEXO 2).

El tablero se compone de 2 páginas:

- Ventas Posadas.
- Predicción.

La vista principal está desarrollada con visualizaciones de información y 2 visualizadores de control.

5.9.4.1. **Información:**

- Mapa de Bing.
  - En el mapa de bing se Visualizan los clientes con su locación geográfica, las cantidades compradas y la distancia a casa central.
- Mapa de capas.
  - Se visualizan las zonas donde se encuentra determinado cliente o producto vendido.
- Venta Promedio.
  - Venta promedio histórica desde que comienza el dataset.
- Productos más vendidos.
  - Gráfica de barras con los productos más vendidos topeados según cantidades vendidas.
- Venta Bruta por Zona.
  - Gráfico de torta de las ventas brutas por cada zona y su aporte porcentual al total de los ingresos.
- Mejores Clientes.
  - Gráfico de barras topeando mejores 3 clientes según ventas netas y a en que zonas comercializan.
- Series de ventas y margen de ganancias.
  - Gráfico de líneas con serie histórica de ventas Brutas, Ventas Netas y se calcula el margen de ganancia en cada serie.
- Margen de ventas.
  - Margen de ventas calculado total histórico.

5.9.4.1.1. **Control:**

- Control de tiempo.
  - Control de intervalo de fechas en el que se desea ver las gráficas.
- Distancia a Casa Central.
  - Control de las distancias existentes de menor a mayor.
  - Puede segmentar según un rango de distancias.

La vista de Predicciones se compone de 2 graficas de visualización de información, 1 grafica de control y 2 gráficos estructura de los modelos en R.

5.9.4.2. **Información:**

- Zonas.
  - Tabla de zonas e id zonas, este grafico depende de la selección de producto.
- Zonas de Predicción.
  - Mapa de capas con la representación de las zonas que aparecen en la tabla de zonas.

5.9.4.2.1. **Control:**

- Producto.
  - Listado de productos, al seleccionar cada producto llama a la función de ML de R que genera el Árbol de Decisión.
  - La selección de este listado se relaciona con las zonas en la tabla zona.

5.9.4.2.2. **Gráficas en R:**

- Árbol en R.
  - Representación del árbol en Rpart.
- Re Neuronal en R.
  - Representación de la Red Neuronal con NeuralNetTools.

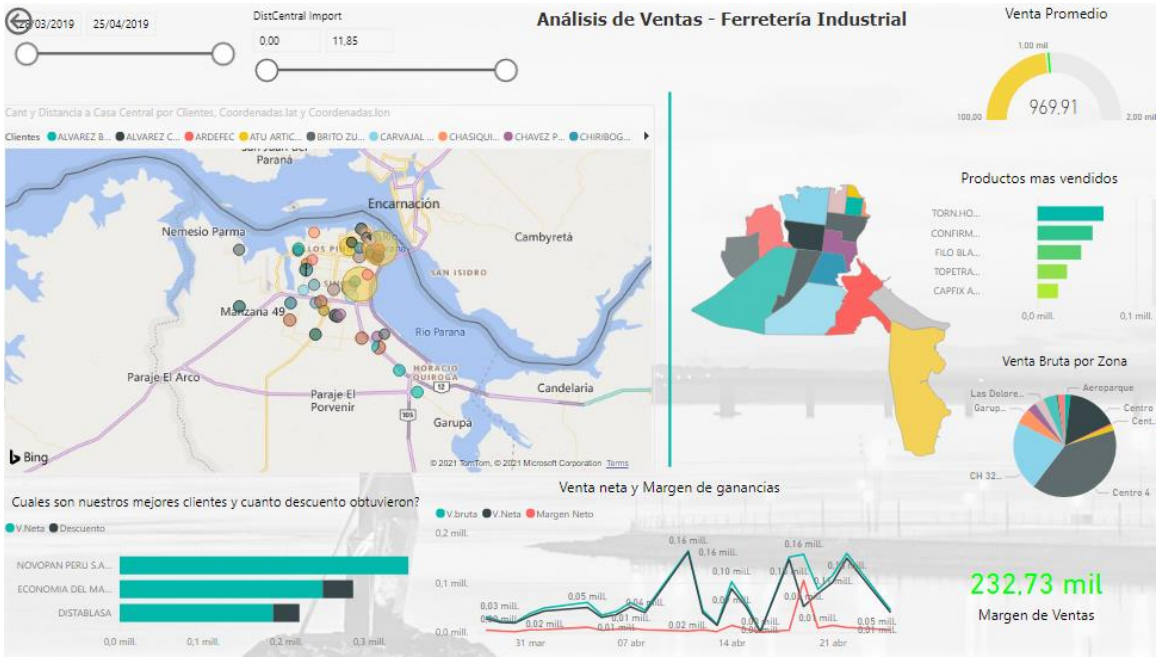

Figura 28: *Página 1 del tablero.*

Al utilizar las tecnologías adecuadas, se puede crear un tablero de mando dinámico que permita la visualización en tiempo real de información de diversos clientes, lo cual resulta sumamente útil para medir el beneficio por esfuerzo de venta que se dedica a cada uno de ellos, es decir, los ingresos de la venta menos el costo de las promociones realizadas dividido la distancia de envío; entendiendo que a mayor distancia será mayor el costo de flete y por lo tanto menor el beneficio. En este sentido, el resultado de la implementación de un tablero de mando bien diseñado y eficiente es una herramienta valiosa para las empresas (Fig. 28), ya que les permite tomar decisiones informadas basadas en datos en tiempo real, así como también monitorear y mejorar sus procesos de ventas. Por lo tanto, contar con un tablero de mando que permita la visualización dinámica de información, y que permita medir el esfuerzo de venta por cliente, se convierte en una ventaja competitiva importante para las empresas que buscan optimizar su rendimiento y mejorar su eficiencia en la gestión de ventas.

El esfuerzo se midió en función a (Ecuación 5):

Ecuación 6*: Cuantificación del Esfuerzo.*

$$Beneficio\ por\ Esfuerzo = \frac{(Venta\ Bruta - \ Promociones)}{Distancia\ a\ Casa\ Central}$$

La realización de un análisis detallado en R, utilizando la técnica de partición de datos (Rpart), permitió la creación de una función altamente efectiva que relaciona los productos ofrecidos por una empresa con las zonas y clientes en donde se ha registrado un menor esfuerzo para concretar las ventas (Fig. 26). Esta valiosa función permite a las empresas identificar oportunidades de mercado y mejorar su desempeño en la gestión de ventas, al permitirles enfocar sus esfuerzos de manera más efectiva en aquellos clientes y zonas en donde se requiere un mayor empuje para cerrar ventas. Asimismo, la creación de esta función en R ofrece la ventaja adicional de contar con un lenguaje de programación poderoso y flexible que facilita la manipulación de grandes volúmenes de datos y la implementación de algoritmos sofisticados, lo que permite a las empresas obtener resultados más precisos y confiables en sus análisis de ventas y estrategias comerciales.

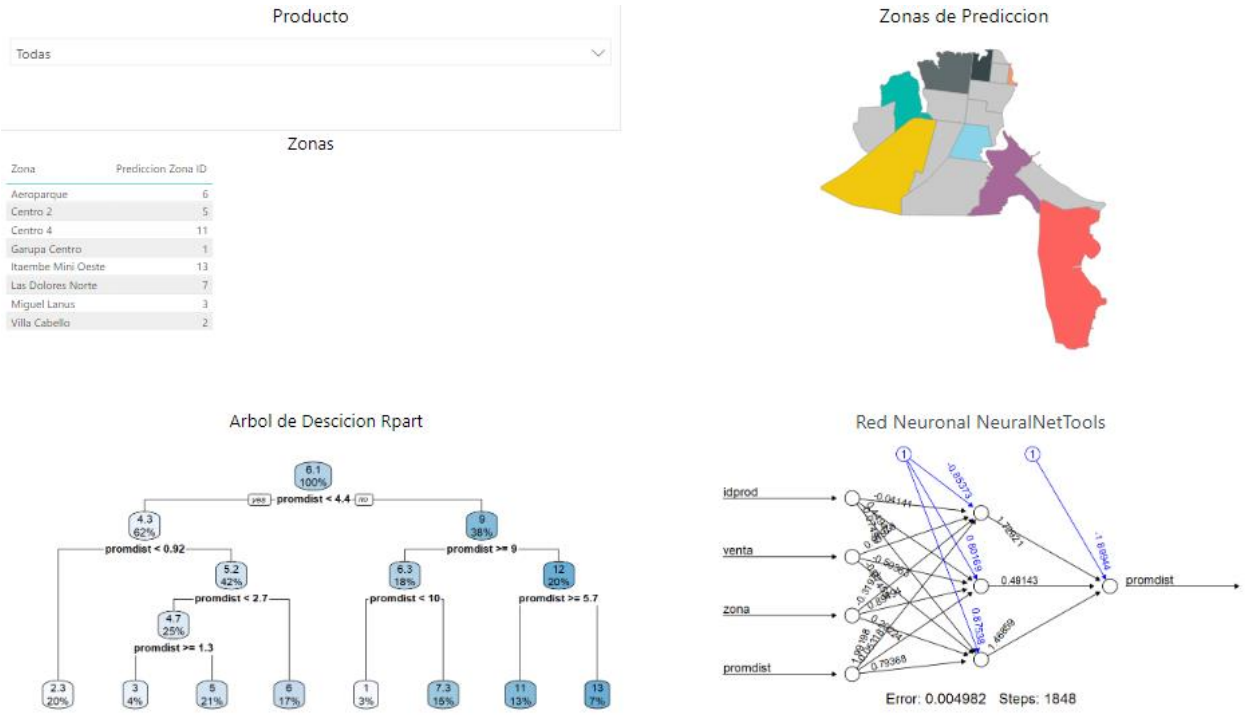

Figura 29: *Página 2 del tablero.*

Capítulo 6

6. Resultados

6.1. Pruebas de Geocodificación

De las pruebas realizadas sobre las APIs de Geolocación, Geocodificación, Geocodificación invertida los resultados nos indican que en zonas geográficas donde la población y desarrollo productivo son relativamente bajos los motores de las soluciones empresariales más conocidas tiene una base de datos pobre en recursos e información; en cambio los motores GIS de las soluciones que se abastecen mediante Crowd Sourcing (Crowd Mapping) de tipo colectivas y gracias a la colaboración de los usuarios tienen mejores resultados, producto de una base de datos enriquecida por la experiencia local.

Tanto OpenCageData como Nominatim pertenecen ambas a la Plataforma OpenStreetmaps que se abastecen del conocimiento colectivo para incrementar la precisión de los datos de referenciamiento, las capas de visualización y las características de las API en el servicio de Sistemas GIS.

Es por ello que cuando se trabaja en el NEA Argentino esta es la plataforma con mejores resultados en comparación a los competidores según el testeo realizado.

6.2. Predicciones utilizando aprendizaje de máquina

6.2.1. Red neuronal:

Utilizando un modelo NeuralNet para predecir zonas con menor esfuerzo de venta.

Precisión: 99,98%.

Pasos: 1848

6.2.2. Árbol:

Utilizando Árbol de Decisión Rpart para predecir zonas con menos esfuerzo de venta la precisión fue del 99.63%.

6.3. Evaluación Encuestas

Para obtener una retroalimentación que valide este proyecto en el ámbito de desarrollo e implementación de servicios basados en el conocimiento, se llevó a cabo una encuesta dirigida exclusivamente a profesionales del sector informático. La encuesta se realizó en línea y se encontraba disponible en el siguiente enlace:

https://forms.gle/GVjpehkcH7SjoHuy9

Una vez completada la encuesta, se proporcionó a los participantes el enlace al tablero final del proyecto para su revisión. El objetivo de la encuesta era conocer las impresiones y evaluaciones de los profesionales sobre nuestro proyecto y para lograr este objetivo.

En (Tabla 8) se muestran las 5 preguntas centradas en evaluar la viabilidad económica, el impacto en la competitividad, la utilidad de las predicciones ofrecidas por el tablero propuesto, la recomendación específica para una PYME del sector de insumos de ferretería y construcción, y la proyección de su relevancia a largo plazo. Se buscó que los encuestados puedan evaluar la viabilidad financiera, el potencial de mejora competitiva, la utilidad práctica de las predicciones, la adecuación al contexto específico de la industria y la sostenibilidad a largo plazo son consideraciones cruciales para tomar decisiones informadas sobre la adopción de esta herramienta de gestión. En última instancia, estas preguntas apuntan a comprender si el tablero propuesto realmente puede ofrecer un valor tangible y sostenible para las PYMEs en el sector del Nea, especialmente en términos de mejora operativa, toma de decisiones estratégicas y posicionamiento en el mercado.

Tabla 8: *Preguntas.*

Código	Pregunta
RQ1	¿Considera económicamente viable la implementación del tablero propuesto en una PYME del Nea?
RQ2	¿Considera que el tablero propuesto incrementa la competitividad de una PYME del Nea?
RQ3	¿Considera de las predicciones en la ventana “Predicciones” constituyen una ayuda importante a la toma de decisiones?
RQ4	¿Recomendaría el tablero propuesto a una PYME del sector de Insumos de ferretería y construcción?
RQ5	¿Considera que el tablero propuesto estará vigente en un mediano o largo plazo?

Se estableció una escala de evaluación para las preguntas realizadas a fin de podes cuantificar los resultados (Tabla 9), la escala va desde “Muy poco” a “Muy aceptable” en una escala del 1 al 5.

Tabla 9: *Escala.*

Escala	Puntaje
Muy poco	1
Poco	2
Regular	3
Aceptable	4
Muy aceptable	5

Se redujo el universo de personas encuestadas a 5 cargos que se consideran críticos a la hora de evaluar e implementar este tipo de soluciones (Tabla 10) preguntando ¿qué tipo de cargos ocupan los encuestados o qué tipo de actividades desarrollan? Y segmentando en base a las respuestas.

Tabla 10: *Cargos de los encuestados.*

Cargo
Analista Universitaria de Sistemas
Jefe de Oficina de Prensa
Analista de Ciberseguridad.
Analista funcional
Informática

6.3.1. **Respuestas**

Las respuestas se compilaron y se representan a continuación en gráficos de torta a fin de facilitar la visualización de las mismas (Figuras 27, 28, 29, 30 y 31).

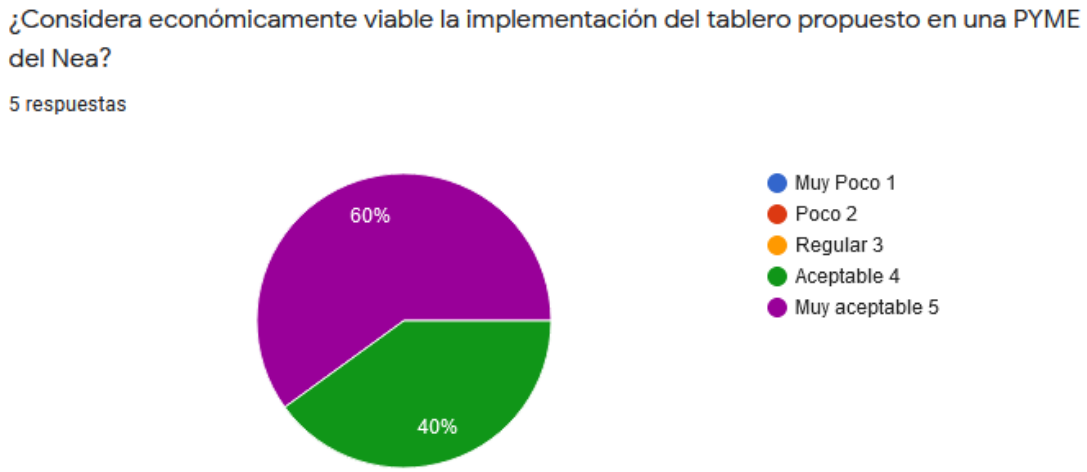

Figura 30: *Respuestas RQ1.*

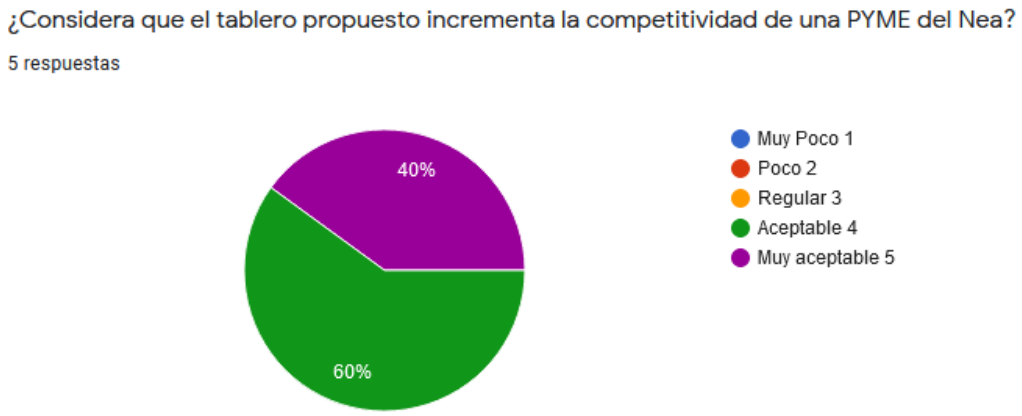

Figura 31: *Respuestas RQ2.*

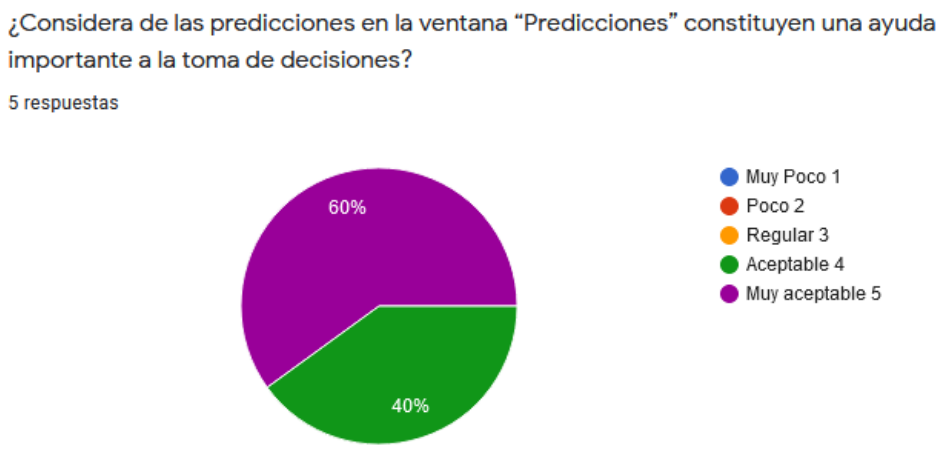

Figura 32: *Respuestas RQ3.*

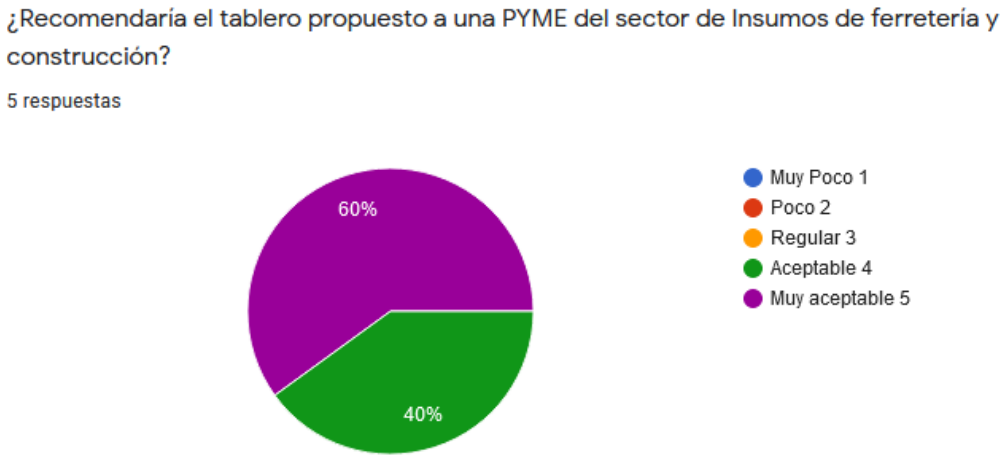

Figura 33: *Respuestas RQ4.*

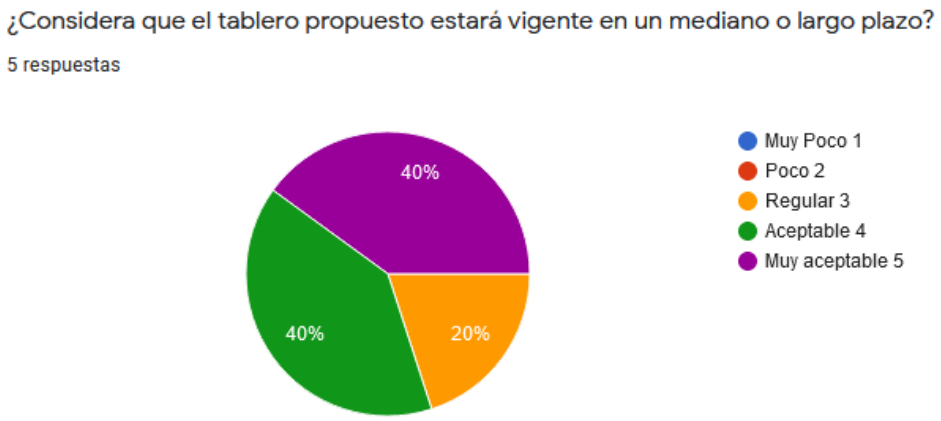

Figura 34: *Respuestas RQ5.*

En base a las respuestas obtenidas se elaboró la Tabla 11 donde se puede observar la matriz de los resultados obtenidos.

Tabla 11: *Resultados.*

Cód.	RQ1	RQ2	RQ3	RQ4	RQ5
E1	5	5	5	5	5
E2	5	4	5	5	5
E3	5	4	5	5	4
E4	4	5	4	4	3
E5	4	4	4	4	4
Promedio	4.6	4.4	4.6	4.6	4.2

6.4. Implementabilidad del modelo

Es importante destacar que la plataforma Power BI, al igual que otras soluciones de inteligencia de negocios disponibles en el mercado, ofrece soporte para una amplia variedad de orígenes de datos, desde archivos de texto plano hasta bases de datos complejas y de gran tamaño. Además, todas las herramientas utilizadas en el desarrollo del modelo son de acceso gratuito y público, lo que representa una ventaja importante para las empresas que buscan implementar soluciones de BI de manera económica.

En cuanto a los costos asociados con la implementación del modelo, el principal ítem de gasto es el de consultoría, soporte técnico, desarrollo del tablero y proceso de obtención de datos, lo que se traduce en horas de servicio. Es importante tener en cuenta que el tiempo de desarrollo del modelo dependerá principalmente del estado de los datos de la empresa, ya que gran parte del trabajo se centrará en depurar y transformar los datos para que sean útiles en la implementación del modelo. De hecho, un origen de datos de mala calidad podría retrasar significativamente el proyecto e incluso ponerlo en riesgo.

En este contexto, es importante destacar que las opiniones recopiladas en la encuesta realizada arrojan un resultado positivo en relación a la percepción que tienen los gerentes de la posibilidad de implementar una solución de BI en el corto y mediano plazo, lo que sugiere un alto grado de interés y compromiso por parte de la empresa en adoptar tecnologías avanzadas para mejorar su rendimiento y eficiencia operativa. En resumen, la implementación del modelo es factible y ofrece un gran potencial de beneficios para las empresas que buscan mejorar su rendimiento y crecer en el mercado.

Capítulo 7

7.1. Conclusiones

Este trabajo expone las principales dificultades que enfrentan las pequeñas y medianas empresas (PyMEs) en la región NEA de Argentina, como el bajo nivel de accesibilidad logística, los costos elevados de flete terrestre, la falta de inversión en tecnología y recursos informáticos, y la compleja estructura de costos y alta presión tributaria. Para abordar estos desafíos, este trabajo propuso una plataforma basada en *Location Business Analytics* para la toma de decisiones comerciales en una empresa de suministros de construcción. La presión por desarrollar e implementar soluciones tecnológicas que maximicen las oportunidades de negocios en las pequeñas empresas es cada vez mayor, por lo tanto las PyMEs que sigan tomando decisiones basadas en la experiencia solamente y no con soporte de datos estarán en condiciones de desventaja frente a sus competidores que sí lo hagan.

Se dedicó buena parte de la investigación a resolver los problemas pertinentes a implementar Geolocación de direcciones de texto plano ya que se concluyó que la clave para combinar información de negocio con información geográfica radicaba en este punto, una vez resuelto este problema fue posible realizar una verdadera combinación de datos entre ambas plataformas (Power BI y OpenCageData) y hacer aprovechable en términos de negocios una solución de *Location Intelligence*.

La combinación de las herramientas de georeferenciamiento con las soluciones de explotación de información de negocio se presenta como una estrategia altamente efectiva para la identificación de nuevas oportunidades comerciales. La metodología que se propuso incluyó la revisión de la literatura, la definición de los requisitos, el diseño y la implementación de una arquitectura para integrar datos de gestión comercial con información geoespacial e implantación de algoritmos de aprendizaje maquina en este proceso de análisis de datos, sumado a la evaluación de la utilidad de la arquitectura mediante encuestas con expertos y tomadores de decisiones. A pesar de las limitaciones en los datos geoespaciales en la zona de estudio, se concluye que la plataforma puede ser útil para mejorar la eficiencia y la rentabilidad de las PyMEs en el NEA de Argentina.

Se evidenció en las pruebas de precisión de APIs de Geocodificación que las plataformas de *Crowdsourcing* demostraron ser más precisas y contar con mejores datos que el resto de las plataformas, siendo ***OpenCageData*** la más precisa con una media de error de 0,0553 kms. Las formulas del Semiverseno y del Calculo Euclidiano de distancia se desatacaron como medio para realizar métricas del error de geocodificación siendo esta ultima la de mayor simpleza del código y menor costo computacional, especialmente si se trabaja con grandes volúmenes de datos.

Otro aspecto importante de este modelo es la posibilidad de combinación de múltiples orígenes de datos, con múltiples plataformas de alojamiento de datos sumados a una gran variedad de herramientas de visualización y diagnóstico. Todo esto sumado a los bajos costos iniciales de implementación y una gran capacidad de escalabilidad del modelo.

La integración de algoritmos *machine learning* con Power BI resultó ser novedoso, con rangos de precisión de 99,98% para las RN y 99.63% en el Árbol de Decisión lo que demuestra fiabilidad al integrar algoritmos desarrollados tanto en Python como en RStudio y encontrar respuestas a preguntas de negocio que la gerencia tradicionalmente respondía por intuición generando nuevos resultados para maximizar el aprovechamiento de nuevas oportunidades comerciales y detectar amenazas.

La capacidad de estas herramientas para identificar patrones y tendencias que maximicen el potencial del negocio puede proporcionar una ventaja competitiva significativa y mejorar sustancialmente la toma de decisiones estratégicas. Por lo tanto, la adopción de estas herramientas puede ser considerada como una

inversión rentable y fundamental para el crecimiento y la prosperidad de cualquier empresa que busque mantenerse a la vanguardia en un mercado cada vez más competitivo.

La validación por opinión experta de la solución desarrollada demuestra principalmente que el modelo es viable de implementar para resolver la problemática planteada por este trabajo con un promedio de respuesta a las RQs de 4,48 sobre 5 que valida aceptablemente el resultado del estudio realizado.

En la actualidad las PyMEs que implementen soluciones de Inteligencia de Negocios aprovechando la integración con Sistemas GIS adquieren una clara ventaja competitiva; sin embargo en un breve lapso de tiempo se volverá un requisito indispensable del mercado y quienes no lo hagan quedaran en una clara desventaja con sus competidores.

7.2. Trabajos Futuros

Como trabajo futuro resta seguir la investigación en torno a las formas de medir el rendimiento de plataformas de geocodificación y su adaptabilidad para combinarse con distintas tecnologías y plataformas destinadas a el análisis de datos comerciales para las PyMEs en este sector de la región.

La utilización de capas desarrolladas en geojson embebidas en aplicaciones GIS resultó de utilidad, se propone investigar otras aplicaciones de capas sobre GIS y sus implicancias en el desarrollo de tableros de negocio.

El trabajo se limitó a un sector comercial específico en un área determinada, se propone investigar la implementación de este trabajo en otros entornos comerciales productivos y también ampliar el área de implementación ya que distintos lugares pueden devolver resultados que aporten novedad respecto de las soluciones propuestas.

. Para futuras investigaciones, es posible no solo continuar con la extensión del modelo, sino también explotar las herramientas y expandir el análisis a todas las redes sociales con el fin de tener mercados de información más precisos y posiblemente nuevos.

Bibliografía

[1] Identidad 21, “Asumir el liderazgo: ¿Cómo toman decisiones de riesgo los empresarios argentinos?,” *Universidad Siglo 21*, 2019. https://identidad.21.edu.ar/asumir-el-liderazgo-como-toman-decisiones-de-riesgo-los-empresarios-argentinos/.

[2] Gartner, “Gartner Delivering Actionable, Objective Insight to Executives and Their Teams,” *Gartner*. 2023, [Online]. Available: https://www.gartner.com/en.

[3] E. Pérez, *Inclusión financiera de las pequeñas y medianas empresas en Colombia*. 2018.

[4] Confederación Argentina de La Mediana Empresa, “CAME | Confederación Argentina de la Mediana Empresa,” *CAME*, 2022. https://www.redcame.org.ar/novedades/12158/las-ventas-minoristas-pymes-cayeron-34-anual-en-mayo.

[5] P. R. Castillo, “Daniel Slutzky. Estructura social agraria y agroindustrial del Nordeste de la Argentina: desde la incorporación a la economía nacional al actual subdesarrollo concentrador y excluyente,” *Mundo Agrar.*, vol. 18, no. 37, p. 051, May 2017, doi: 10.24215/15155994e051.

[6] B. Kitchenham, O. Pearl Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, “Systematic literature reviews in software engineering - A systematic literature review,” *Information and Software Technology*, vol. 51, no. 1. Elsevier, pp. 7–15, Jan. 01, 2009, doi: 10.1016/j.infsof.2008.09.009.

[7] M. von Rosing, S. White, and H. de Man, *The Complete Business Process Handbook Body of Knowledge From Process Modelling to BPM*, vol. 1. 2017.

[8] V. Krishnan, “IR @ INFLIBNET: Research Data Analysis with Power BI,” *INFLIBNET’s Institutional Repos.*, 2017, [Online]. Available: https://ir.inflibnet.ac.in/handle/1944/2116.

[9] C. Shearer, “The CRISP-DM Model The New Blueprint for Data Mining,” *J. data Wareh.*, vol. 5, 2000, Accessed: Apr. 24, 2024. [Online]. Available: https://www.scirp.org/reference/ReferencesPapers?ReferenceID=1592780.

[10] F. Burstein and C. W. Holsapple, *Handbook on Decision Support Systems 2*. Springer Berlin Heidelberg, 2008.

[11] S. Cristobal, “Los Sistemas de Información Geográfica,” *Geoenseñanza*, vol. 11, pp. 107–116, 2006.

[12] A. Hassan and J. Vijayaraghavan, *Geospatial Data Science Quick Start Guide: Effective techniques for performing smarter geospatial analysis using location intelligence*. Packt Publishing, 2019.

[13] J. Fombona Cadavieco and E. Vázquez-Cano, “Posibilidades de utilización de la geolocalización y realidad aumentada en el ámbito educativo,” *Educ. XX1*, vol. 20, no. 2, pp. 319–342, 2017, doi: 10.5944/educXX1.10852.

[14] J. A. Quesada, “COMPARACIÓN DE LAS APLICACIONES DE GOOGLE Y YAHOO PARA LA GEOCODIFICACIÓN DE DIRECCIONES POSTALES CON FINES EPIDEMIOLÓGICOS,” *Rev. Esp. Salud Publica*, vol. 87, no. 2, pp. 201–206, 2013, doi: 10.1590/S1135-57272013000200009.

[15] T. Guarda, M. F. Augusto, and I. Lopes, “Geographic market intelligence as a competitive advantage,” *Iber. Conf. Inf. Syst. Technol. Cist.*, vol. 2019-June, no. June, pp. 19–22, 2019, doi: 10.23919/CISTI.2019.8760856.

[16] M. Angelaccio, A. Basili, B. Buttarazzi, and W. Liguori, “Using geo-business intelligence to improve quality of life,” 2012, doi: 10.1109/estel.2012.6400196.

[17] M. F. -POLLO CATTANEO, “MODELO DE PROCESO PARA ELICITACIÓN DE REQUERIMIENTOS EN PROYECTOS DE EXPLOTACIÓN DE INFORMACIÓN,” *Univ. Nac. LA PLATA Fac. INFORMÁTICA*, p. 278, 2017, [Online]. Available:

https://sedici.unlp.edu.ar/handle/10915/96315.

[18] A. Q. Retamal, V. Vega, J. Gallardo, and C. J. Meneses, “Methodological proposal for agile requirements engineering for sme View project Automatic generation of explanations from Rule-based Expert System and Ontology View project,” 2009.

[19] S. Dick, “Artificial Intelligence,” *Harvard Data Sci. Rev.*, vol. 1, no. 1, Jun. 2019, doi: 10.1162/99608f92.92fe150c.

[20] S. Magaz, “Aprendizaje Máquina y Computación Cuántica,” 2020.

[21] F. K.-H. Chun, M. Graefen, A. Haese, and P. I. Karakiewicz, “Recursive Partitioning,” *An Introd. to Recursive Partitioning Using RPART Routines*, pp. 1–60, 2019, doi: 10.4135/9781412971980.n280.

[22] M. W. Beck, “NeuralNetTools: Visualization and analysis tools for neural networks,” *J. Stat. Softw.*, vol. 85, no. 11, p. 1, 2018, doi: 10.18637/jss.v085.i11.

[23] M. Abdel-Basset, M. Mohamed, and V. Chang, “NMCDA: A framework for evaluating cloud computing services,” *Futur. Gener. Comput. Syst.*, vol. 86, pp. 12–29, 2018, doi: 10.1016/j.future.2018.03.014.

[24] C. Schröer, F. Kruse, and J. M. Gómez, “A systematic literature review on applying CRISP-DM process model,” in *Procedia Computer Science*, 2021, vol. 181, no. 2019, pp. 526–534, doi: 10.1016/j.procs.2021.01.199.

[25] A. Azevedo and M. F. Santos, “KDD , SEMMA AND CRISP-DM : A PARALLEL OVERVIEW,” *IADIS Eur. Conf. Data Min.*, pp. 182–185, 2008.

[26] M. Kocbek, G. Jošt, M. Heričko, and G. Polančič, “Business process model and notation: The current state of affairs,” *Comput. Sci. Inf. Syst.*, vol. 12, no. 2, pp. 509–539, 2015, doi: 10.2298/CSIS140610006K.

[27] C. N. G. Ramírez, “Sistema de Información Geográfica para la inteligencia de Negocios, caso de Estudio en la Superintendencia de Economía Popular y Solidaria,” *Univ. SAN Fr. QUITO*, 2014.

[28] J. M. Aversa, “Spatial Big Data Analytics: The New Boundaries of Retail Location,” *Wilfrid Laurier Univ. Sch. Commons @ Laurier*, p. 2019, 2019, [Online]. Available: https://scholars.wlu.ca/cgi/viewcontent.cgi?article=3257&context=etd.

[29] L. E. Ferro-Diéz, N. M. Villegas, and J. Diáz-Cely, “Location data analytics in the business value chain: A systematic literature review,” *IEEE Access*, vol. 8, pp. 204639–204659, 2020, doi: 10.1109/ACCESS.2020.3036835.

[30] A. M. Báguena, “Data Analytics for Smart cities,” *Dep. en “E.T.S. Ing. Informáticos (UPM),”* 2019, [Online]. Available: https://oa.upm.es/55594/.

[31] A. V. M. Doblado, “Sistema de información para la Geolocalización de los gastos de los empleados aplicando Geomarketing,” *Univ. Oberta Catalunya*, 2019, [Online]. Available: http://hdl.handle.net/10609/97247.

[32] M. E. López Inga and R. M. Guerrero Huaranga, “Cloud Business Intelligence & Analytics model for SMEs retail,” *Ing. Solidar.*, vol. 14, no. 24, pp. 1–17, Jan. 2018, doi: 10.16925/in.v14i24.2157.

[33] G. Zapata, J. Murga, C. Raymundo, F. Dominguez, J. M. Moguerza, and J. M. Alvarez, “Business information architecture for successful project implementation based on sentiment analysis in the tourist sector,” *J. Intell. Inf. Syst.*, vol. 53, no. 3, pp. 563–585, 2019, doi: 10.1007/s10844-019-00564-x.

[34] A. T. Murray, J. Xu, Z. Wang, and R. L. Church, “Commercial GIS location analytics: capabilities and performance,” *Int. J. Geogr. Inf. Sci.*, vol. 33, no. 5, pp. 1106–1130, 2019, doi: 10.1080/13658816.2019.1572898.

[35] C. TOREGAS and C. ReVELLE, “OPTIMAL LOCATION UNDER TIME OR DISTANCE CONSTRAINTS,” *Pap. Reg. Sci.*, vol. 28, no. 1, pp. 133–144, Jan. 1972, doi: 10.1111/j.1435-5597.1972.tb01521.x.

[36] C. Toregas and C. Revelle, “Binary Logic Solutions to a Class of Location Problem,” *Geogr. Anal.*, vol. 5, no. 2, pp. 145–155, Apr. 1973, doi: 10.1111/j.1538-4632.1973.tb01004.x.

[37] S. Adi, S. Heripracoyo, and E. Irwansyah, “GIS-Based Business Intelligence (BI) for People,” *IEEE*, 2019, [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8843729.

[38] D. Sikeridis, I. Papapanagiotou, B. P. Rimal, and M. Devetsikiotis, “A Comparative Taxonomy and Survey of Public Cloud Infrastructure Vendors,” *Distrib. Parallel, Clust. Comput.*, Oct. 2017, doi: https://doi.org/10.48550/arXiv.1710.01476.

[39] A. Dixon, Z. Zhao, J. C. Bohorquez, R. Denney, and N. Johnson, “Statistical physics and modern human warfare,” *Model. Simul. Sci. Eng. Technol.*, vol. 51, pp. 365–396, 2010, doi: 10.1007/978-0-8176-4946-3_14/COVER.

[40] Sitio web institucional de la Municipalidad de Posadas - Misiones., “Municipalidad de Posadas,” *Municipalidad de Posadas*, 2022. https://www.posadas.gov.ar/.

[41] A. Delpeuch, “OpenRefine,” *Google Inc.*, 2021. https://openrefine.org/ (accessed Jul. 28, 2021).

[42] H. Zhang, J. Zheng, G. Dorr, H. Zhou, and Y. Ge, “Testing of GPS Accuracy for Precision Forestry Applications,” *Arab. J. Sci. Eng.*, vol. 39, no. 1, pp. 237–245, Nov. 2014, doi: 10.1007/s13369-013-0861-1.

[43] H. Alkan and H. Celebi, “The Implementation of Positioning System with Trilateration of Haversine Distance,” in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC*, Sep. 2019, vol. 2019-Septe, doi: 10.1109/PIMRC.2019.8904289.

[44] E. Abdi, H. S. Mariv, A. Deljouei, and H. Sohrabi, “Accuracy and precision of consumer-grade GPS positioning in an urban green space environment,” *Forest Sci. Technol.*, vol. 10, no. 3, pp. 141–147, 2014, doi: 10.1080/21580103.2014.887041.

[45] G. Szwoch, “Combining Road Network Data from OpenStreetMap with an Authoritative Database,” *J. Transp. Eng. Part A Syst.*, vol. 145, no. 2, p. 04018085, Nov. 2019, doi: 10.1061/jtepbs.0000215.

[46] T. Vincenty, “Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations,” *Surv. Rev.*, vol. 23, no. 176, pp. 88–93, 1975, doi: 10.1179/sre.1975.23.176.88.

[47] J. P. Snyder, “Map Projections--a Working Manual,” *U.S. Geol. Surv. Prof. Pap. 1395*, 1987.

Anexo 1

El siguiente anexo de adjuntan los resultados del proceso de RSL para el estado del arte de este trabajo, los criterios de búsqueda en este apartado representan únicamente a la sección 3 del trabajo.

7.2.1. Preguntas de investigación:

Tabla 12: *Preguntas de investigación*

Pregunta de investigación	Dimensiones
RQ1: ¿Qué tipo de Metodologías y modelos se utilizan para desarrollar proyectos de BI and GIS?	Tipos de Propuestas: método, conocimiento, herramienta, características.
RQ1.2: ¿Cómo se validan las metodologías o los modelos?	Validación
RQ2: ¿Cómo fue el tratamiento de los orígenes de datos?	Tipo de datos.
RQ3: ¿Cómo puede ayudar la integración de BI con GIS en la toma de decisiones en las pequeñas empresas?	Validación: propuesta, ejemplo, conclusión.
RQ4: ¿Cuáles son las ventajas de los sistemas GIS aplicados a la inteligencia de negocios?	Característica: definiciones.

7.2.2. Cadenas de Búsqueda

Science Direct:

Location Business Analytics AND Cloud AND Pyme

Fecha	**2014-**

Google Scholar y Microsoft Academic:

Location Business Analytics AND Cloud AND Pyme

Desde	**2014**

IEEEXplore:

((("Document Title":Location Business Analytics) AND "Abstract":Cloud) AND "Abstract":Pyme)

Desde	**2014**

ACM Digital Library:

Title:(Location Business Analytics) AND Abstract:(Cloud) AND Abstract:(Pyme)

Fecha	**Publication Date: (01/01/2014 TO**

Scopus:

TITLE(Location AND Business AND Analytics) AND ABS(Cloud) AND ABS(Pyme) AND PUBYEAR > 2014

7.2.3. Fuentes de Búsqueda

Las búsquedas se realizaron en los siguientes repositorios digitales:

- https://scholar.google.es
- https://www.sciencedirect.com
- https://ieeexplore.ieee.org/Xplore/home.jsp
- http://dl.acm.org/
- https://www.elsevier.com/solutions/scopus
- https://www.biblioteca.mincyt.gob.ar/
- https://academic.microsoft.com

7.2.4. Criterios de Exclusión e Inclusión

7.2.4.1. Criterios de exclusión:

- Fecha de publicación del artículo, todas las publicaciones anteriores a 2017.
- Artículos no técnicos (guías, artículos de introducción, editoriales, etc.)

7.2.4.2. Criterios de inclusión:

- Artículos relacionados a GIS aplicado a los negocios
- Aplicación de modelos BI y análisis espacial
- Aplicación de técnicas de Deep Learning en tablas de datos geoespaciales.

Anexo 2

Se incluye la programación utilizada para las funciones en el tablero de Pwer Bi.

Código Fuente Tabla "Ventas":

```
let
    Origen =
Web.Workbook(Web.Contents("https://www.dropbox.com/scl/fi/q6drs6jnu539q2hecy7vc
/FerretariaTesis.xlsx?dl=0&rlkey=hyou5zft65pkda861ichs3u7d"), null, true),
    Ventas_Sheet = Origen{[Item="Ventas",Kind="Sheet"]}[Data],
    #"Encabezados promovidos" = Table.PromoteHeaders(Ventas_Sheet,
[PromoteAllScalars=true]),
    #"Tipo cambiado" = Table.TransformColumnTypes(#"Encabezados
promovidos",{{"Zona", type text}, {"Factura", Int64.Type}, {"Fecha", type date},
{"Cod", Int64.Type}, {"Cliente", type text}, {"Código", type text}, {"Producto",
type text}, {"Cant", Int64.Type}, {"M3", type number}, {"V.bruta", type number},
{"Descuento", type number}, {"V.Neta", type number}, {"Costo", type number},
{"Column14", type any}, {"Column15", type any}, {"Column16", type any}}),
    #"Columnas quitadas" = Table.RemoveColumns(#"Tipo cambiado",{"Column14",
"Column15", "Column16"})
in
    #"Columnas quitadas"
```

Codigo Fuente "Funcion Geolocalizacion":

```
let
    AddressToLocate = (address as text) =>
let
    Origen =
Json.Document(Web.Contents("https://nominatim.openstreetmap.org/search.php?q="&
address&"&format=jsonv2")),
    Origen1 = Origen{0},
    #"Convertido en tabla" = Record.ToTable(Origen1),
    #"Filas filtradas" = Table.SelectRows(#"Convertido en tabla", each ([Name]
= "lat" or [Name] = "lon")),
    #"Tabla transpuesta" = Table.Transpose(#"Filas filtradas"),
    #"Encabezados promovidos" = Table.PromoteHeaders(#"Tabla transpuesta",
[PromoteAllScalars=true]),
    #"Valor reemplazado" = Table.ReplaceValue(#"Encabezados
promovidos",".",",",Replacer.ReplaceText,{"lat"}),
    #"Valor reemplazado1" = Table.ReplaceValue(#"Valor
reemplazado",".",",",Replacer.ReplaceText,{"lon"}),
    #"Tipo cambiado" = Table.TransformColumnTypes(#"Valor reemplazado1",{{"lat",
type number}, {"lon", type number}})
in
```

```
    #"Tipo cambiado"
in
    AddressToLocate
```

Anexo 3

Certificado de Publicación:

Certificado de haber Presentado un artículo en el VI Congreso Iberoamericano de Ciudades Inteligentes (ICSC-CITIES 2023).

El articulo presentado y publicado pertenece a una sección del presente trabajo.

Link descarga certificado:
https://drive.google.com/drive/folders/17nOd0zblgAugpVEYGYGjEWvcnAeqEyuT

Link a publicación:
https://icsc-cities.com/proceedings/2023.pdf