Chatbot de análisis estadístico para la toma de decisiones en pymes. Caso piloto: Empresa PROVETSA

CARRERA DEL RIO MONTESDEOCA MANUEL ALBERTO

MARTÍNEZ ZEA PAOLO FRANCISCO

FACULTAD DE POSTGRADO

MAESTRÍA EN INTELIGENCIA DE NEGOCIOS Y CIENCIA DE DATOS

14 DE JUNIO DE 2025

Chatbot, Análisis estadístico, Pymes, Toma de decisiones, Automatización

Resumen ejecutivo

En el contexto de las pequeñas y medianas empresas (pymes), las decisiones estratégicas suelen estar condicionadas por la falta de personal especializado en análisis de datos, infraestructura tecnológica adecuada y, en muchos casos, por restricciones presupuestarias que dificultan adoptar soluciones analíticas avanzadas. Frente a este escenario, se plantea una solución accesible, funcional y escalable: un chatbot de análisis estadístico orientado a apoyar la toma de decisiones gerenciales y de otros actores clave.

El prototipo, desarrollado en el marco académico y probado con datos simulados de la empresa ficticia PROVETSA, permite la interacción mediante una interfaz conversacional cerrada (Tidio), donde el usuario final elige entre distintos análisis estadísticos según sus necesidades. El uso de este tipo de chatbots responde a una tendencia creciente en la automatización de la comunicación en entornos comerciales (Polo Peña & Frías Jamilena, 2018). Los datos ingresados son enviados a un backend en Python, que se conecta a un archivo Excel que simula una base de datos previamente estructurada y procesada mediante técnicas ETL.

El backend ejecuta tres tipos de análisis: exploratorio descriptivo, regresión lineal y análisis de clúster. Cada uno entrega resultados interpretados automáticamente en un lenguaje accesible para usuarios no técnicos, facilitando decisiones como proyecciones de ventas o segmentación de productos.

Este prototipo constituye una prueba de concepto viable, con potencial de evolución futura mediante técnicas de procesamiento de lenguaje natural o inteligencia artificial, sentando así las bases para democratizar el análisis de datos en entornos empresariales con recursos limitados.

Resumen ejecutivo gráfico

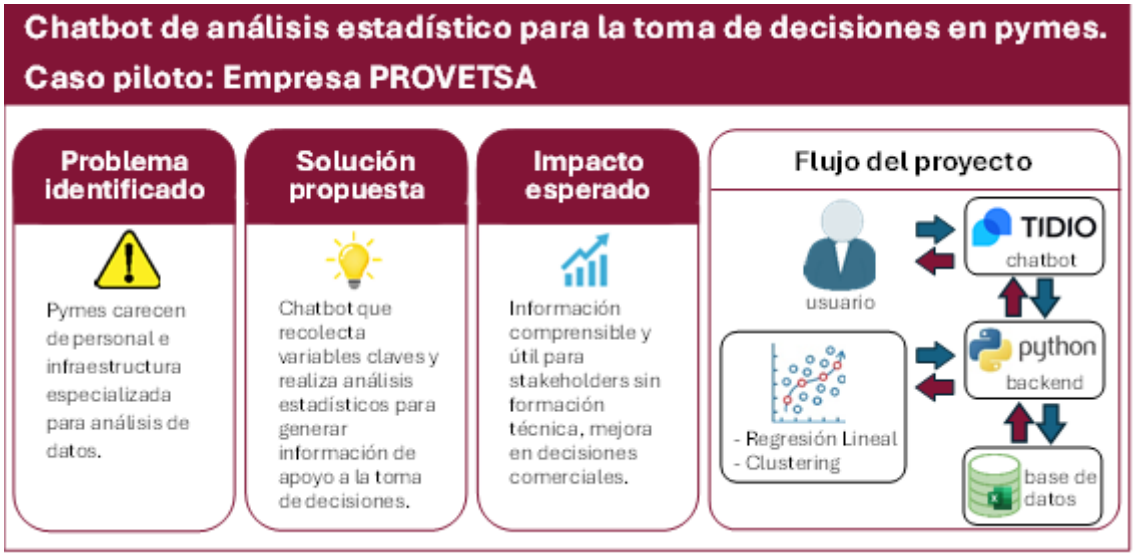


Figura 1

Resumen gráfico del funcionamiento del prototipo: integración entre chatbot, programa en Python y análisis estadístico automatizado. Nota. Elaboración propia.

Detalle técnico

1. Fuente de datos

La base de datos utilizada en este proyecto fue desarrollada exclusivamente para fines académicos, elaborada a partir de registros internos de ventas de una empresa del sector veterinario. Con el fin de resguardar la confidencialidad y cumplir con las normativas vigentes en materia de protección de datos personales, los valores fueron ajustados mediante un factor de transformación que mantiene las relaciones estadísticas originales sin revelar cifras reales. Además, todos los nombres, identificadores y códigos han sido debidamente anonimizados.

El archivo se presenta en formato Excel y actúa como una base de datos relacional, compuesta por nueve tablas: Ventas, Cliente, Producto, Línea, Tipo, Zona, Ciudad, Vendedor y un Diccionario de Datos. Estas tablas están interrelacionadas mediante claves primarias y

foráneas, lo que permite realizar análisis integrales y consultas estructuradas (ver Anexo A para el Diccionario de Datos y Anexo B para el Diagrama Entidad-Relación).

Criterios de selección y variables. Se incluyeron variables clave que representan distintas dimensiones del proceso comercial:

- Dimensión temporal: Año, Mes.
- Dimensión geográfica: Zona, Ciudad.
- Dimensión comercial: Código del Vendedor, Código del Producto, Línea, Tipo.
- Dimensión de resultados: Unidades vendidas, Valor total de venta, Precio unitario.

La base contiene más de 10.000 registros, lo cual brinda representatividad suficiente para analizar el comportamiento del negocio a lo largo del tiempo y en distintas zonas geográficas.

Análisis exploratorio. Se aplicaron técnicas de análisis exploratorio de datos, como medidas de tendencia central, dispersión, segmentaciones por categorías y visualizaciones gráficas. Entre los hallazgos más relevantes se identificaron patrones estacionales, productos de alta rotación, ciudades con mejor desempeño y diferencias entre vendedores. Este análisis inicial sirvió para definir las variables más significativas y dar paso a la aplicación de técnicas multivariadas.

2. Técnica

Las técnicas seleccionadas responden al objetivo general del proyecto: facilitar el acceso al análisis estadístico a través de un chatbot conversacional que apoye la toma de decisiones en pymes. Se eligieron tres enfoques complementarios, comúnmente utilizados en el análisis multivariado para entornos empresariales (Hair, Black, Babin & Anderson, 2019), que permiten ofrecer valor desde distintos ángulos:

- Análisis exploratorio: Proporciona una descripción general de las variables seleccionadas para detectar patrones, tendencias o relaciones simples. Esta técnica es

útil para comprender el comportamiento de indicadores como ventas por producto o ingresos mensuales.

- Regresión lineal: Permite analizar la relación entre dos variables y estimar el impacto de una sobre otra. Se seleccionó por su utilidad para establecer relaciones directas —por ejemplo, entre unidades vendidas e ingresos— y facilitar proyecciones sencillas.
- Análisis de clúster (K-means): Agrupa elementos con características similares, como productos o clientes, permitiendo segmentar de forma automática y sin necesidad de etiquetas previas. Esta técnica apoya decisiones estratégicas como la personalización de ofertas o la redefinición del portafolio comercial.

Todas estas técnicas se ejecutan dentro de un programa en Python, desarrollado especialmente para el proyecto, y sus resultados se presentan en un lenguaje accesible y comprensible, orientado a usuarios no especializados en análisis de datos. La implementación se apoyó en bibliotecas ampliamente utilizadas en ciencia de datos, como pandas y NumPy, por su versatilidad para manejar estructuras tabulares (McKinney, 2018).

3. Arquitectura tecnológica y flujo de datos

El prototipo se basa en una arquitectura sencilla pero efectiva, que combina una interfaz conversacional, un motor de análisis local y una base de datos simulada. El flujo de funcionamiento se desarrolla en las siguientes fases:

1. Interacción: El usuario final se comunica con un chatbot configurado en la plataforma Tidio, donde selecciona el tipo de análisis requerido mediante flujos conversacionales cerrados (ver Anexo C).
2. Comunicación con el programa: Las solicitudes del usuario final se envían desde Tidio hacia el programa en Python, que se ejecuta localmente en el equipo y utiliza FastAPI,

un microframework inspirado en Flask que permite implementar servicios web ligeros (Grinberg, 2018), para recibir y procesar los datos solicitados.

3. Procesamiento y análisis: El programa accede al archivo Excel, realiza el análisis estadístico correspondiente y genera automáticamente una interpretación del resultado en lenguaje natural.
4. Respuesta: El resultado interpretado se envía nuevamente al chatbot, que lo presenta al usuario final como parte de la conversación.

Para posibilitar esta conexión entre el chatbot en la nube y el programa en ejecución local, se emplea la herramienta *ngrok*, que crea un canal seguro y temporal hacia la aplicación, sin necesidad de realizar configuraciones técnicas avanzadas.

Este enfoque permite validar de manera funcional el prototipo, al tiempo que sienta las bases para futuras versiones más robustas, integradas a bases de datos reales o servicios en la nube, e incluso con capacidades ampliadas mediante inteligencia artificial.

Resultados

La implementación del prototipo integrador se concretó mediante la combinación de un chatbot conversacional desarrollado en la plataforma Tidio y un programa en Python ejecutado localmente, el cual se conecta a un archivo Excel que simula una base de datos estructurada. Esta integración permitió automatizar el proceso de consulta, análisis y entrega de resultados estadísticos en lenguaje comprensible para el usuario.

El chatbot presenta tres opciones principales, cada una vinculada con una técnica estadística distinta:

- **Análisis exploratorio**: En este caso, se configuraron preguntas que permiten al usuario final conocer, por ejemplo, cuáles productos generaron mayores ingresos, cómo varía el promedio de ventas según los meses del año, o cuál es el comportamiento general de

los vendedores. La visualización de estos patrones facilitó una primera lectura del negocio y reveló concentraciones de ventas en productos específicos y estacionalidades marcadas.

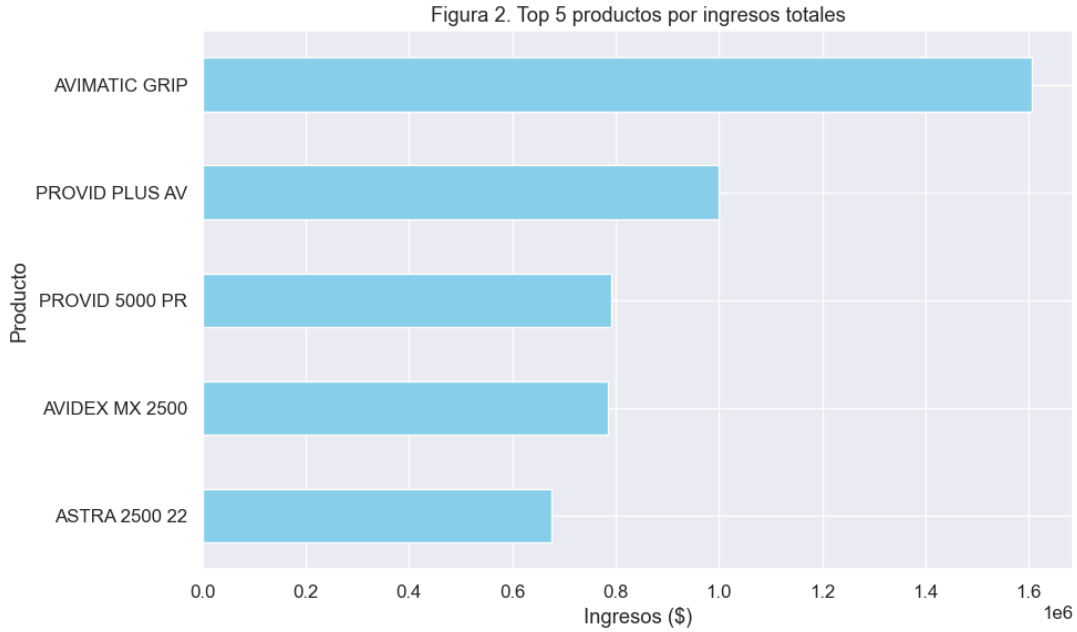


Figura 2

Top 5 productos por ingresos totales. Nota. Elaboración propia.

- **Regresión lineal**: A través de preguntas como "¿Qué relación hay entre unidades vendidas e ingresos?" o "¿Cómo influyen las ventas de un producto en sus ingresos?", el sistema ejecuta modelos que explican dicha relación. Por ejemplo, en el análisis entre unidades vendidas e ingresos, se encontró una relación directa con un coeficiente de determinación (R^2) superior al 80 %, lo que sugiere que las unidades explican gran parte de la variabilidad en los ingresos. Este tipo de análisis permite realizar estimaciones prácticas y comprensibles.

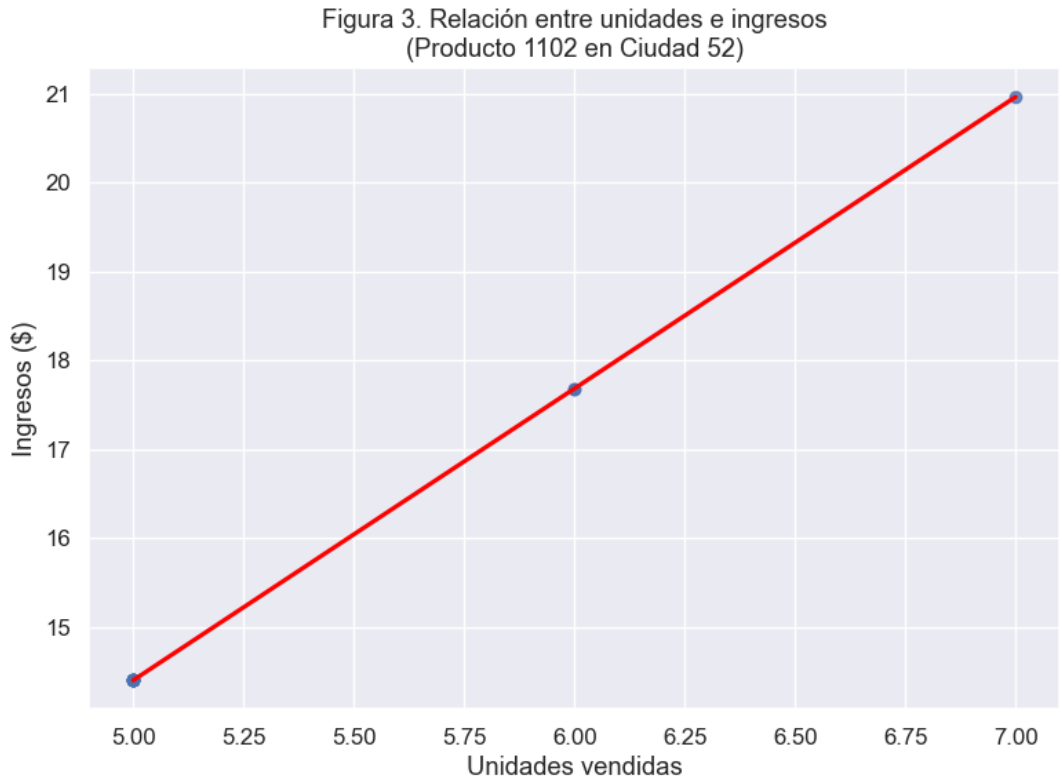


Figura 3

Relación entre unidades vendidas e ingresos para el producto 1102 en la ciudad 52.

Nota. Elaboración propia.

- **Análisis de clúster**: Se configuraron casos como "¿Qué productos se parecen entre sí en función de ventas e ingresos?" o "¿Cómo agrupar a los clientes por comportamiento de compra?". Los resultados segmentaron los elementos analizados en grupos con patrones comunes, lo cual representa una oportunidad para diseñar estrategias diferenciadas o ajustar decisiones de inventario y marketing.

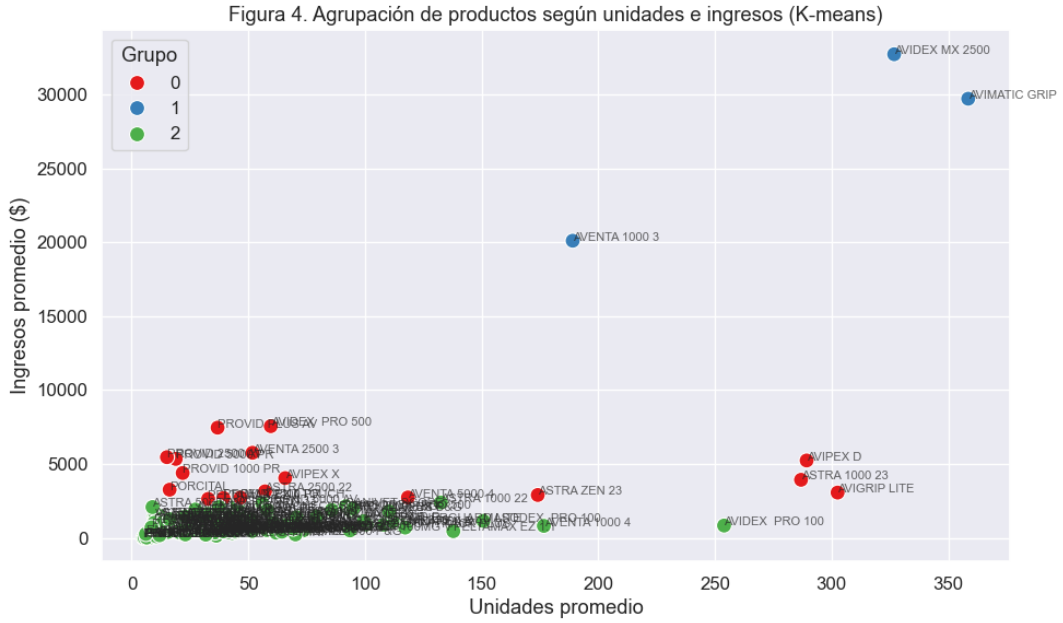


Figura 4

Agrupación de productos según su comportamiento de ventas e ingresos mediante K-means. Nota. Elaboración propia.

En todos los casos, los resultados no se presentan como cifras técnicas, sino como interpretaciones automatizadas en lenguaje natural, diseñadas para que cualquier persona, incluso sin formación en estadística, pueda comprenderlos y tomar decisiones a partir de ellos. Los resultados obtenidos demuestran que el prototipo cumple con los objetivos planteados en este proyecto académico: facilitar el acceso a herramientas de análisis estadístico mediante un entorno conversacional amigable y funcional. Su enfoque modular, la posibilidad de conectarse a fuentes de datos simples como Excel, y su capacidad para interpretar automáticamente los análisis realizados, hacen de esta propuesta una solución viable y escalable para pequeñas y medianas empresas con recursos limitados. Además, sienta las bases para futuras mejoras, como la incorporación de nuevos tipos de análisis, conexión con bases de datos reales o la integración de inteligencia artificial conversacional.

Como evidencia del funcionamiento del chatbot, se incluye una captura de pantalla (ver Anexo D) que muestra una de las interacciones realizadas en la plataforma Tidio. En este ejemplo, el usuario solicita el total de ingresos generados por un vendedor específico en un año determinado, y el sistema responde con la cifra correspondiente, procesada en tiempo real mediante el backend en Python.

Referencias bibliográficas

Grinberg, M. (2018). *Flask web development: Developing web applications with Python* (2nd ed.). O'Reilly Media.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning.

McKinney, W. (2018). *Python for data analysis: Data wrangling with pandas, NumPy, and IPython* (2nd ed.). O'Reilly Media.

Polo Peña, A. I., & Frías Jamilena, D. M. (2018). Chatbots como herramienta de marketing relacional: Conceptualización, retos y oportunidades. *Revista Española de Investigación de Marketing ESIC, 22*(2), 100–113. https://doi.org/10.1016/j.reimke.2018.07.001

Anexo A

Diccionario de Datos

Tabla: Ventas

Campo	Tipo de variable	Descripción
Año	Ordinal	Año en que se realizó la venta.
Mes	Ordinal	Mes en que se realizó la venta.
Código Cliente	Discreta	Identificador del cliente.
Código Producto	Discreta	Identificador del producto.
Código Zona	Discreta	Identificación de la zona asiganada al cliente.
Código Ciudad	Discreta	Identificación de la ciduad del cliente.
Código Vendedor	Discreta	Identificación del vendedor.
Unidades	Discreta	Número de unidades vendidas.
Valores	Continua	Valor total en dóalres de la venta.

Tabla: Cliente

Campo	Tipo de variable	Descripción
Código Cliente	Discreta	Identificador del cliente.
Nombre Cliente	Nominal	Nombre del cliente.

Tabla: Producto

Campo	Tipo de variable	Descripción
Código Producto	Discreta	Identificador del producto.
Nombre Producto	Nominal	Descripción del producto.
Código Línea	Discreta	Identificador de la línea del producto.
Código Tipo	Discreta	Identificación del tipo del producto.
Precio	Continua	Precio unitario del producto en dóalres.

Tabla: Linea

Campo	Tipo de variable	Descripción
Código Línea	Discreta	Identificador de la línea del producto.
Nombre Línea	Nominal	Descripción de la línea del producto.

Tabla: Tipo

Campo	Tipo de variable	Descripción
Código Tipo	Discreta	Identificación del tipo del producto.
Nombre Tipo	Nominal	Descripción del tipo del producto.

Tabla: Zona

Campo	Tipo de variable	Descripción
Código Zona	Discreta	Identificación de la zona asiganada al cliente.
Nombre Zona	Nominal	Nombre de la zona.

Tabla: Ciudad

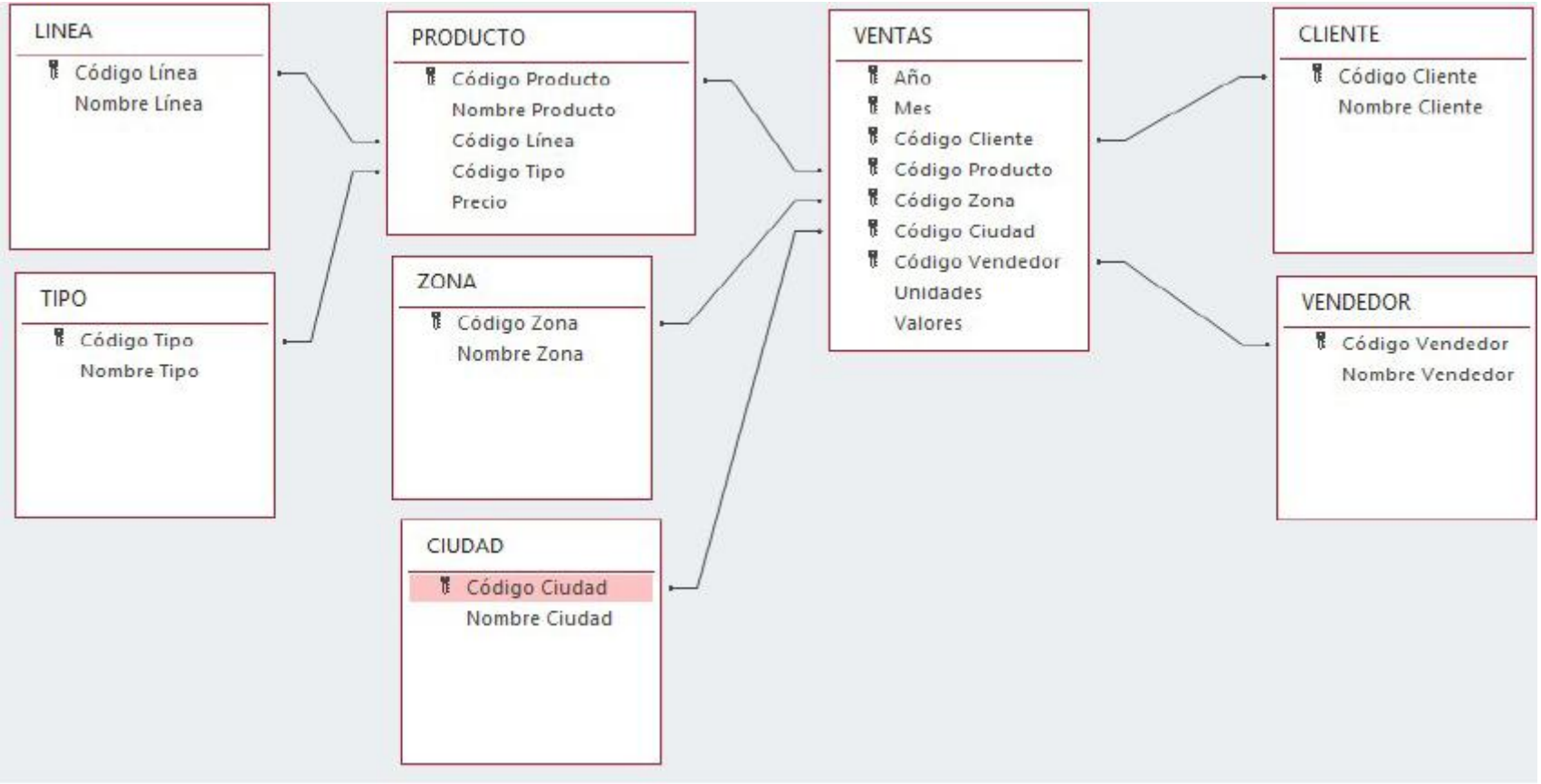
Campo	Tipo de variable	Descripción
Código Ciudad	Discreta	Identificación de la ciduad del cliente.
Nombre Ciudad	Nominal	Nombre de la ciudad.

Tabla: Vendedor

Campo	Tipo de variable	Descripción
Código Vendedor	Discreta	Identificación del vendedor.
Nombre Vendedor	Nominal	Nombre del vendedor.

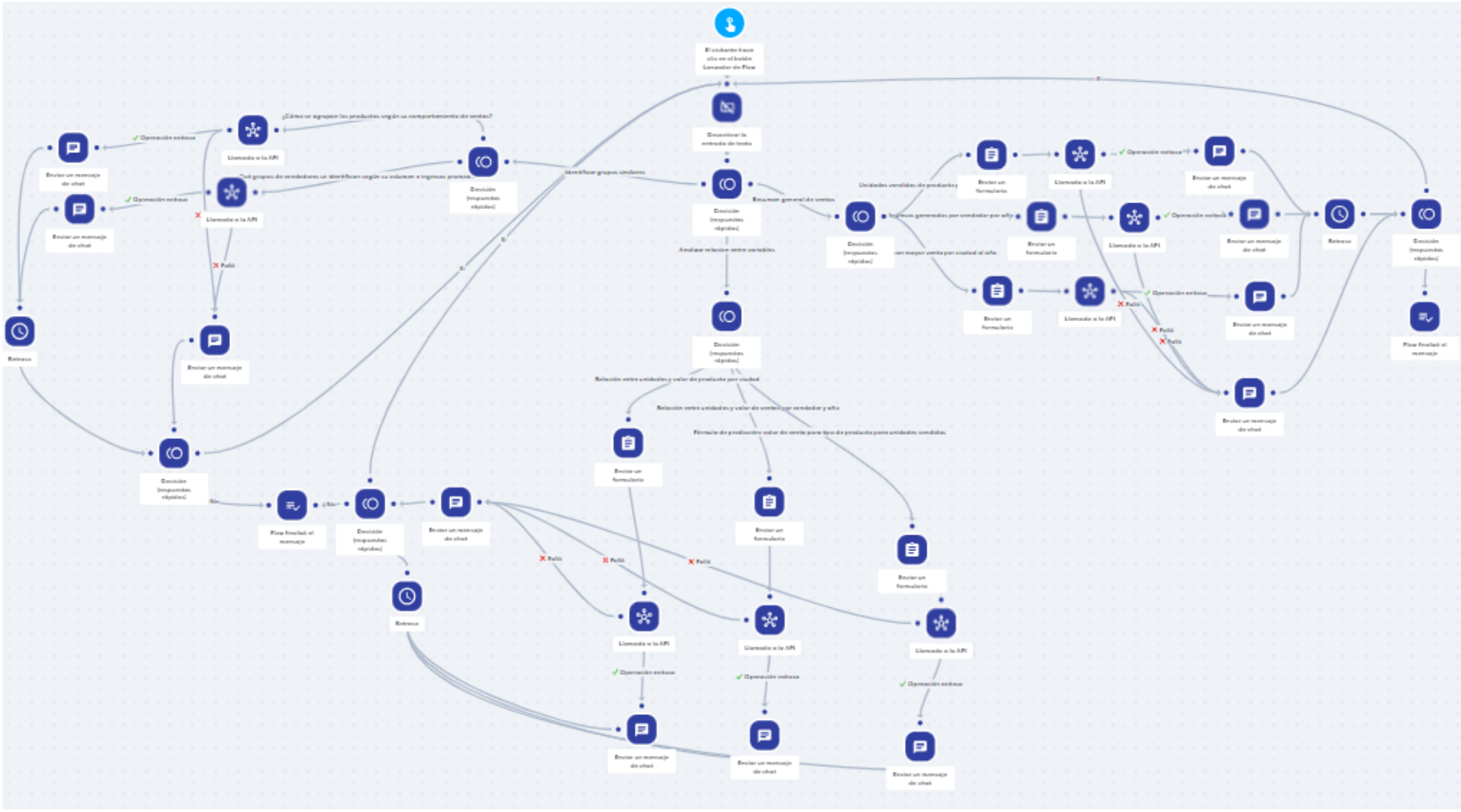
Anexo B

Diagrama Entidad-Relación



Anexo C

Flujo de Tidio



Anexo D

Interacción en el chatbot con consulta de ingresos por vendedor en el año 2024

