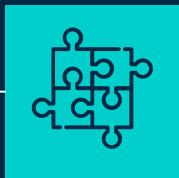


# SMART MODELS TECHNOLOGY COMPANY

*BIENVENIDOS!*

# CONTENIDO



01

ETL

Recorrido por el  
ETL de Google y  
YELP



02

PROCESO DE  
INGENIERÍA

Creación y  
Automatización del  
DataWare House



03

PUNTO DE PARTIDA  
ANALYTICS Y  
MACHINE LEARNING

¿Dónde estamos?  
¿Hacia dónde  
vamos?

The background is a dark blue field decorated with various geometric elements. There are several thin, vertical white lines of varying lengths scattered across the frame. Interspersed among these lines are small squares in three colors: pink, orange, and cyan. Some of these squares are solid, while others are outlined in white. The overall aesthetic is modern and minimalist.

# ETL YELP!



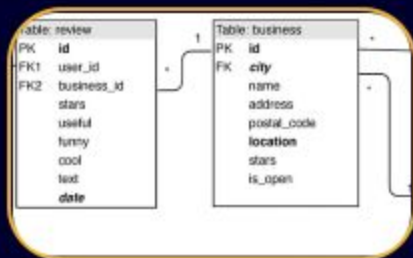
# ETL Grupo de datos Yelp

Descubre cómo realizamos el proceso de extracción, transformación y carga de los datos de Yelp para obtener insights valiosos.





# Extracción de Datos



## Dataset Descargado

Descargamos 5 datasets de Yelp directamente del sitio web.



## Archivos JSON

Cada dataset contenía varios archivos en formato JSON con diferentes atributos.



## Librerías Python

Usamos las librerías pandas y numpy para manipular los datos con eficiencia.



# Transformación de los Datos

1

## Limpiar Datos

Eliminamos valores nulos.  
Eliminamos campos que no eran relevantes en nuestras métricas.

2

## Unir Archivos

Unimos los archivos en un solo conjunto de datos usando las claves comunes, para poder hacer análisis más profundos.

3

## Crear Diccionario

Por cada archivo creamos un diccionario que explicara el significado y el tipo de cada atributo.

4

## Ordenar Datos

Filtramos y ordenamos los datos según las métricas que se querían analizar.



# Cambio de Formato



## JSON a Parquet

Pasamos los archivos de JSON a parquet, un formato binario que permite comprimir y almacenar los datos en columnas Y optimizar memoria.



## Big Data

Manejar grandes volúmenes de datos complejos y heterogéneos fue un éxito gracias al uso eficiente y efectivo de las herramientas y técnicas adecuadas.



## Google Cloud

Elegimos Google Cloud como plataforma en la nube para ofrecer servicios de almacenamiento y análisis de datos.



# Análisis de Datos

## Comportamiento de Clientes

Análisis de preferencias y opiniones de los clientes.

## Rendimiento de Negocios

Análisis de las características y el rendimiento de los negocios.

## Tendencias de Mercado

Análisis de las tendencias y oportunidades del mercado.





# Beneficios del Proceso ETL

- Mejora la calidad, la integridad y la utilidad de los datos.
- Permite integrar y transformar datos de diferentes fuentes en un almacén de datos o en otro sistema de destino.
- Proporciona insights valiosos para fines de inteligencia empresarial, machine learning o informes.



# Conclusiones

## Capacidad de Manejo de Datos

Realizamos un proceso de ETL con éxito demostrando nuestra capacidad para manejar grandes volúmenes de datos complejos y heterogéneos.

## Comunicación Efectiva

Mostramos nuestra habilidad para comunicar los resultados de forma clara y concisa, usando formatos visuales y resumidos.

## Mejora de Servicio

Los insights obtenidos nos permitieron tomar decisiones estratégicas para mejorar nuestro servicio, aumentar nuestra competitividad y satisfacer las necesidades de nuestros usuarios.

The background is a dark blue gradient. It features several thin, vertical white lines of varying lengths scattered across the frame. Interspersed among these lines are small squares in three colors: light blue, pink, and orange. Some squares are solid, while others are outlined. The overall aesthetic is modern and minimalist.

# ETL GOOGLE



# ETL Google

Se tomaron los reviews para los estados 'California', 'New York', 'Florida', 'Nevada', e 'Illinois'; damos formato de fecha a la columna 'data', filtramos en busca de Hoteles Bares y Locales Nocturnos.

Se borran las columnas que no usaremos , filtramos y concatenamos los archivos que trae el dataset, después unimos los dos archivos procesados por la columna 'gmap\_id' y dejamos los que coinciden.



# ETL Google

Los datos de Google tienen una dimensión mayor a nivel mundial, en esta ocasión cruzaremos y **complementaremos** las reseñas obtenidas en Yelp de acuerdo a posicionamiento global que ofrece el dataset de google y el **Score de sentimiento** que generamos para nuestros cliente.





■ Aprovechamos el posicionamiento global que nos proporciona **Google**, adicionamos una categorización de sentimiento



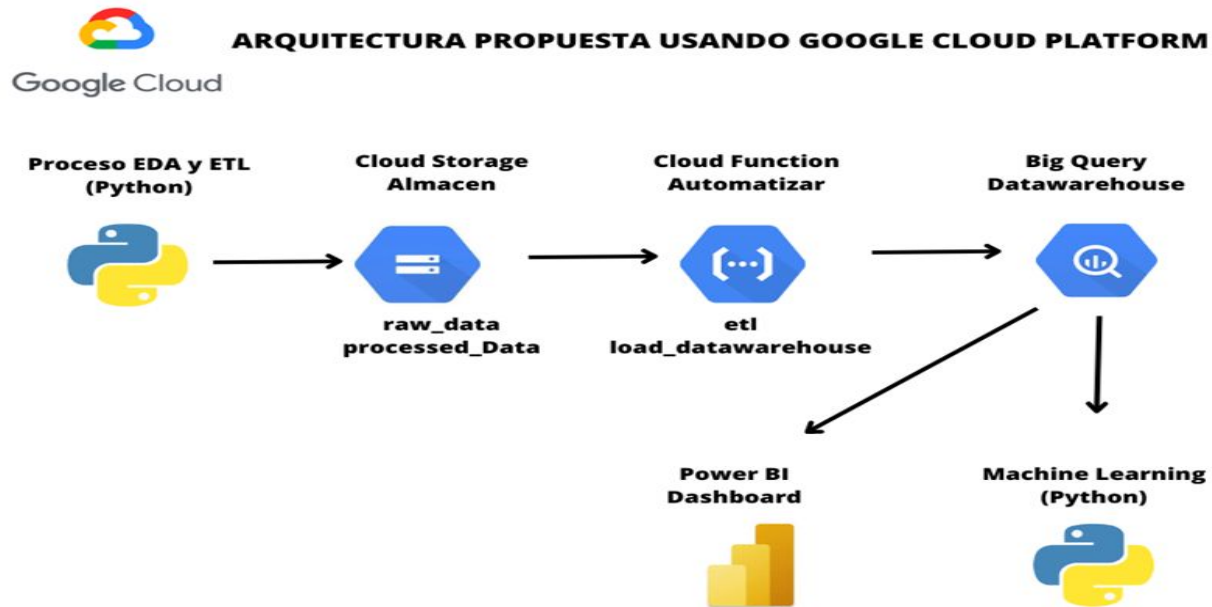
Después de esta **focalización**■ podemos realizar el **análisis** solicitado para ofrecer a nuestro cliente los resultados esperados.

The background is a dark blue field decorated with various geometric elements. It includes numerous small squares in white, light blue, and orange, some of which are solid while others are outlines. Thin white vertical lines of varying lengths are scattered across the composition, creating a sense of depth and structure.

# ARQUITECTURA DE DATOS



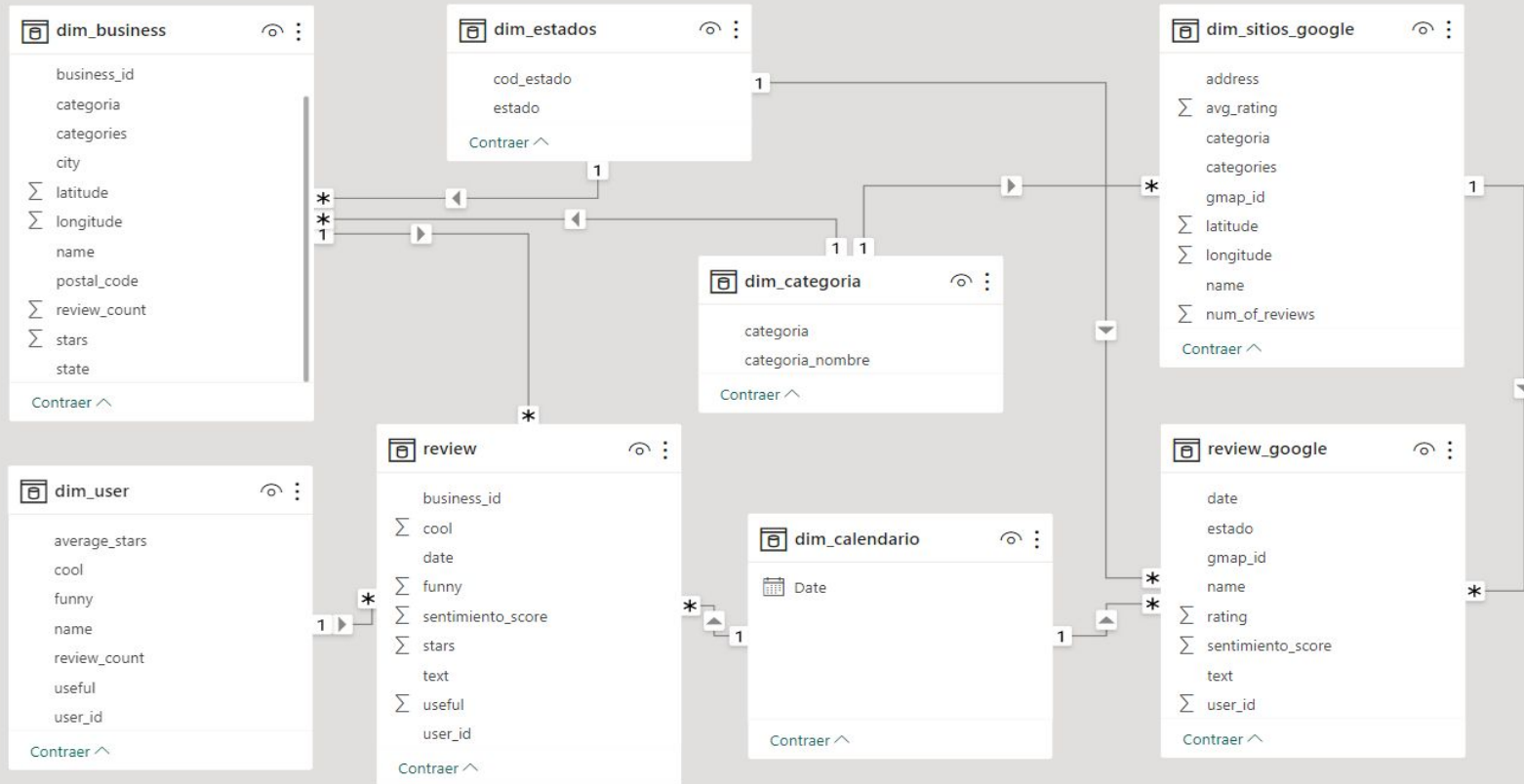
## 02 PROCESO DE INGENIERÍA: Arquitectura propuesta







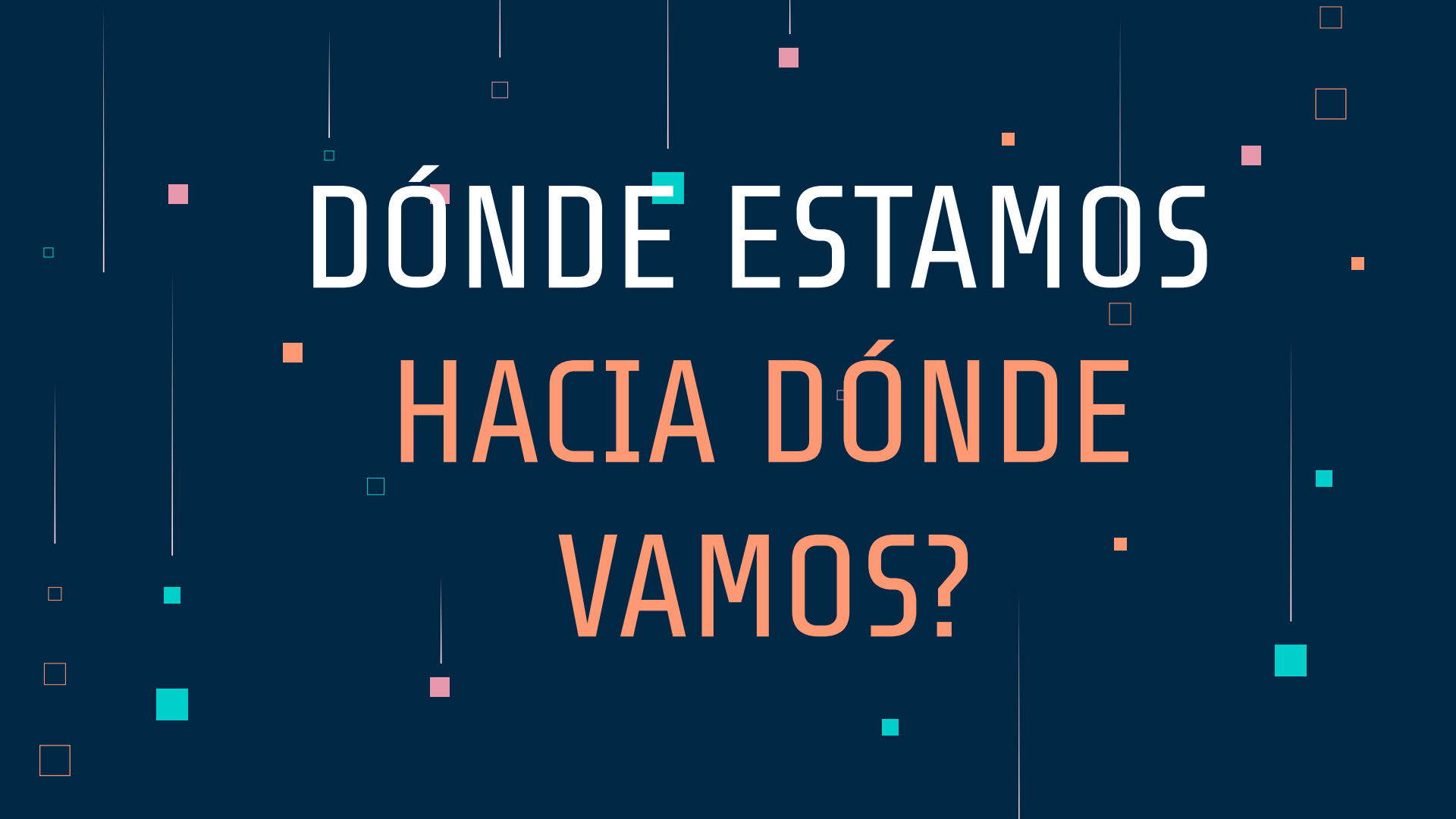
## 02 PROCESO DE INGENIERÍA: Diagrama Datawarehouse



The background is a dark blue field decorated with various geometric elements. It includes numerous small squares in teal, orange, and pink, as well as thin white vertical lines of varying lengths. Some squares are solid, while others are hollow outlines. The overall aesthetic is modern and minimalist.

# CARGA INCREMENTAL



The background is a dark blue field decorated with various geometric elements. There are several thin white vertical lines of varying lengths. Scattered throughout are small squares in three colors: teal, pink, and orange. Some of these squares are solid, while others are just white outlines. The text is centered and consists of three lines: 'DÓNDE ESTAMOS' in white, 'HACIA DÓNDE' in orange, and 'VAMOS?' in orange.

DÓNDE ESTAMOS  
HACIA DÓNDE  
VAMOS?

# NUESTRO PROCESO



# SMART MODELS TECHNOLOGY COMPANY

*GRACIAS!!!!*