

Sprint #2

Grupo 4



"Potenciamos el éxito empresarial mediante soluciones avanzadas de consultoría en Business Intelligence (BI). Nuestro equipo especializado transforma la información en conocimiento estratégico, impulsando la toma de decisiones informadas para alcanzar objetivos empresariales con precisión y eficiencia."

MIEMBROS DEL EQUIPO

Juan Ignacio Berti

- Científico de Datos

jignacio.berti@gmail.com

Agustín Chienello

- Analista de Datos

aguschienello27@gmail.com

M. Soledad García Ortiz

- Ingeniería de Datos

mariasoledad.garciaortiz@gmail.com

Mario Suaza

Mentor

Diccionario de Datos



ORDENES

Nombre columna	Tipo de dato	Descripción
IdOrden	(Int)	Identificador único de orden.
IdCliente	(Int)	Identificador único de cliente por orden
estado	(Str)	Estado de la orden.
fecha_compra	(date)	Fecha en que se realizó la compra.
fecha_aprobacion	(date)	Fecha en que se aprobó la compra.
fecha_envio	(date)	Fecha de envío.
fecha_recepción	(date)	Fecha de entrega.
fecha_recepcion_estimada	(date)	Fecha estimada de entrega.

ARTÍCULOS_ORDEN

IdOrden	(Int)	Identificador único de orden.
articulo	(Int)	Número de artículo de la orden.
id_producto	(Str)	Identificador único de producto.
IdVendedor	(Int)	Identificador único de vendedor.
fecha_limite_entrega	(date)	Fecha de entrega límite de productos.
precio	(float)	Precio del producto.
gasto_envio	(float)	Precio del envío por producto.

PAGOS

Nombre columna	Tipo de dato	Descripción
IdOrden	(Int)	Identificador único de orden.
medio_de_pago	(Str)	Medio de pago por orden
cuotas	(Int)	Cantidad de cuotas en que se pagó
Payment_value	(float)	Gasto por pago

CLIENTES

Nombre columna	Tipo de dato	Descripción
IdClientes	(str)	Identificador único de cliente.
codigo_postal	(str)	Código postal de cliente
IdCiudad	(int)	Identificador único de Ciudad

VENDEDORES

IdVendedor	(Int)	Identificador único de vendedor.
codigo_postal	(Int)	Código postal del vendedor
IdCiudad	(Int)	Identificador único de ciudad

Diccionario de Datos

PRODUCTOS

Nombre columna	Tipo de dato	Descripción
id_producto	(Str)	Identificador único de producto.
IdCategorias	(Int)	Identificador único de categorías.
product_name_lenght	(Int)	Cant. caracteres del nombre del producto.
product_description_length	(Int)	Cant. caracteres de la descripción del producto.

CATEGORIAS

IdCategorias	(Int)	Identificador único de categorías.
Categorias	(Str)	Categorías de los productos

OPINIONES

IdOpinion		Identificador único de opinión
IdOrden	(Int)	Identificador único de orden.
puntaje	(Int)	Puntos al servicio.
titulo	(Str)	Título del mensaje
mensaje	(Str)	Opinión sobre el servicio.
fecha_opinion	(Date)	Fecha en que se realizó lo opinión.
fecha_respuesta	(Date)	Fecha de respuesta a la opinión.

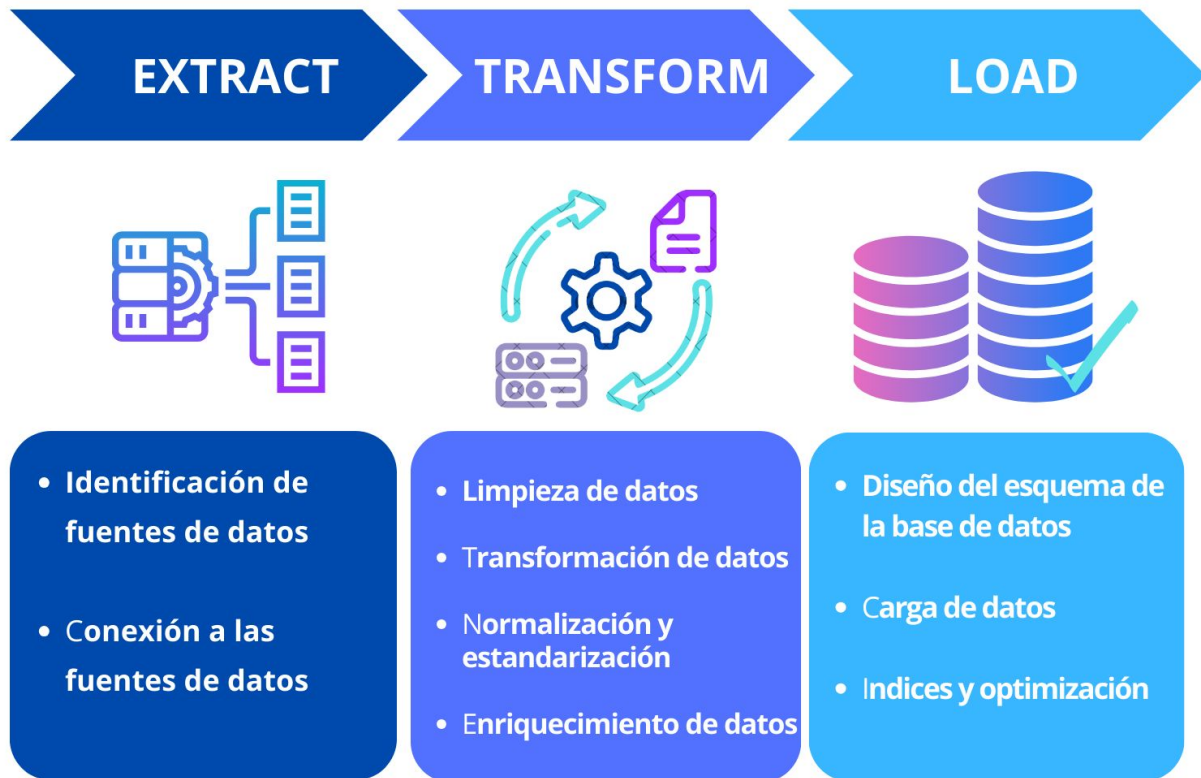
LOCALIZACION

Nombre columna	Tipo de dato	Descripción
codigo_postal	(Int)	Código postal
latitud	(Float)	Latitud
longitud	(Float)	Longitud
IdCiudad	(Int)	Identificador único de ciudad.

CIUDAD

IdCiudad	(Int)	Identificador único de ciudad.
Ciudad	(Str)	Nombre de ciudad.
Estado	(Str)	Nombre de estado.

ETL (Extract, Transform, Load):



Stack Tecnológico & Flujo de Trabajo

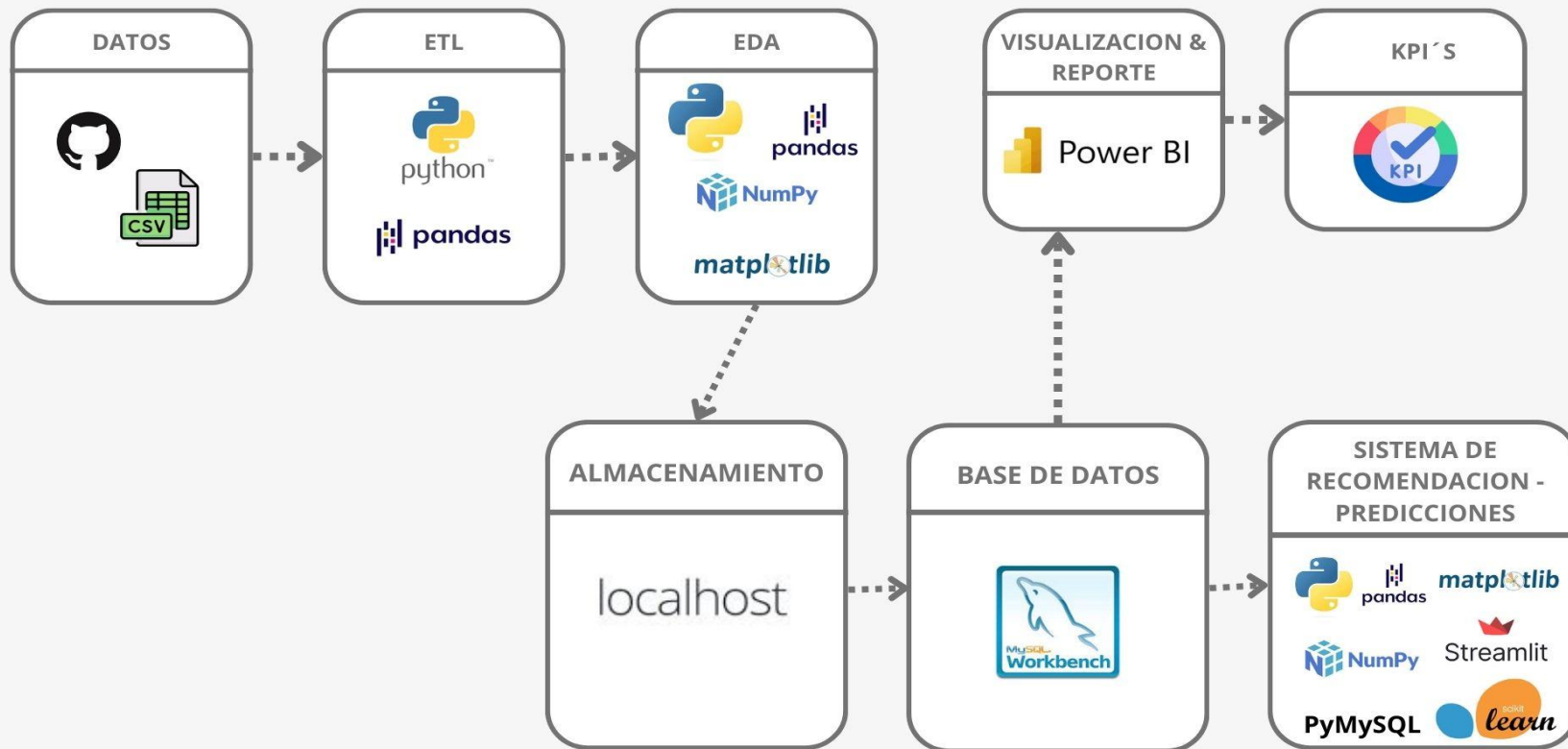
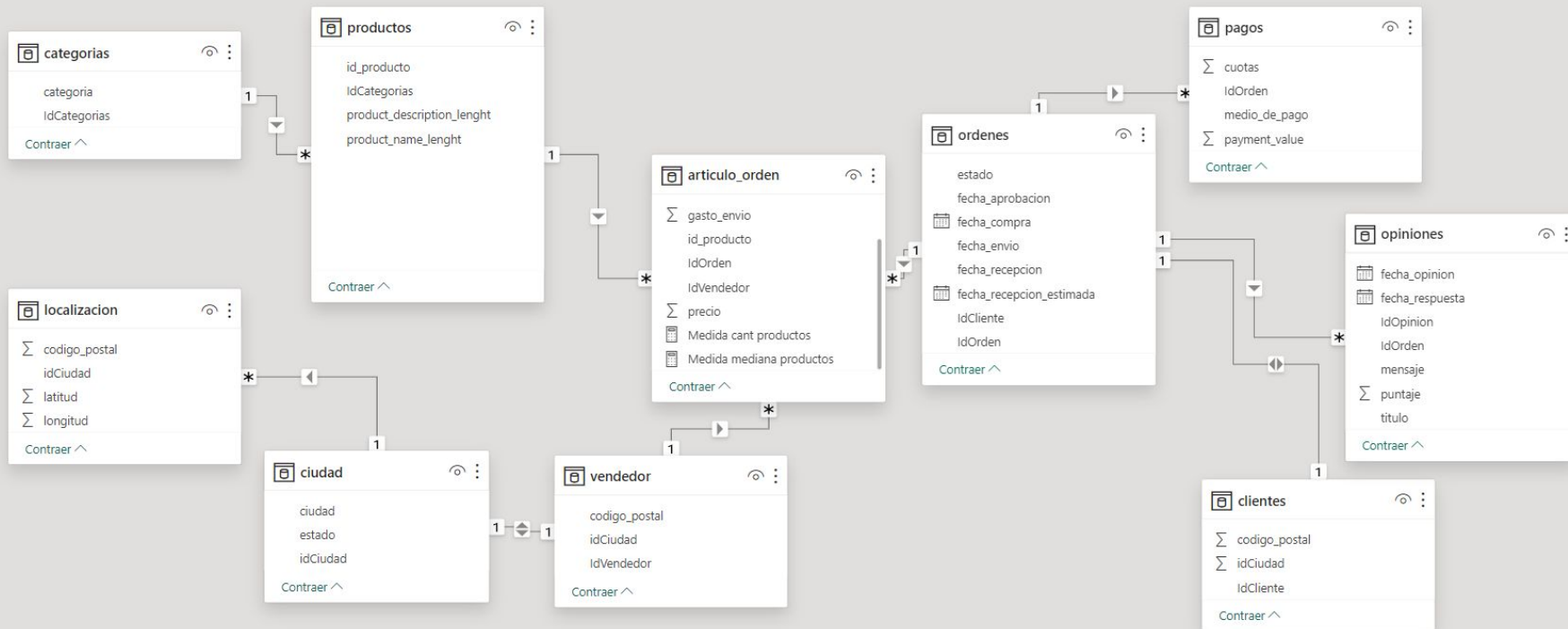


Diagrama Entidad-Relación



EDA (Exploratory Data Analysis):

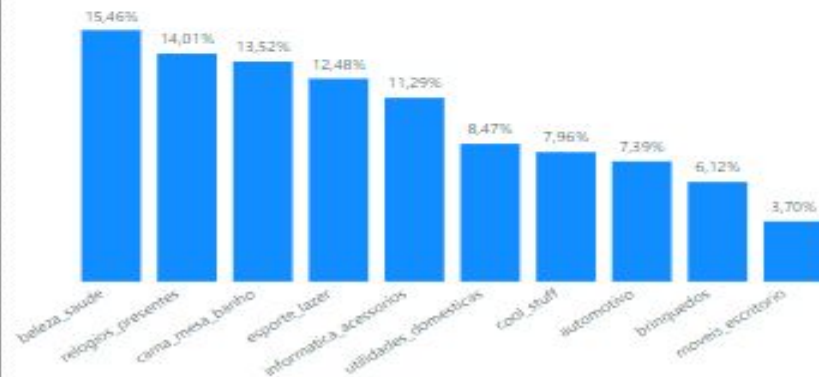
Total de vendas

\$1.415.881.547

Vendas por Ano



TOP 10 Total de vendas por categoria



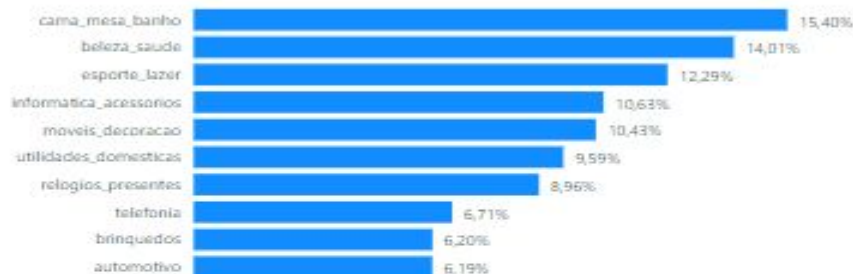
KPI Produtos menos vendidos

\$848.113.678

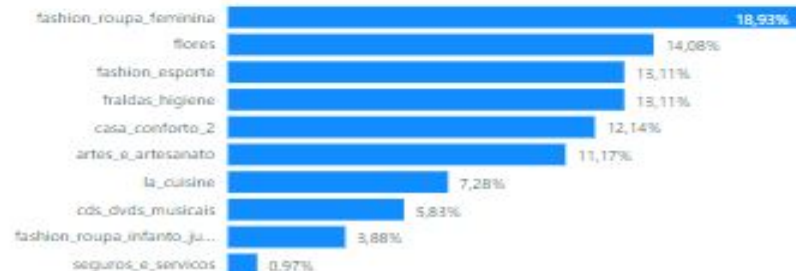
KPI Categorias menos vendidas

\$1.499.896

TOP 10 Categorias más vendidas



TOP 10 Categorias menos vendidas



Distribución geográfica de vendedores vs órdenes

VENDEDORES POR ESTADO



ORDENES POR ESTADO



Métodos de pago utilizados

KPI Órdenes procesadas

103886

Tasa de conversión de pagos

1,04

Promedio de puntaje de clientes

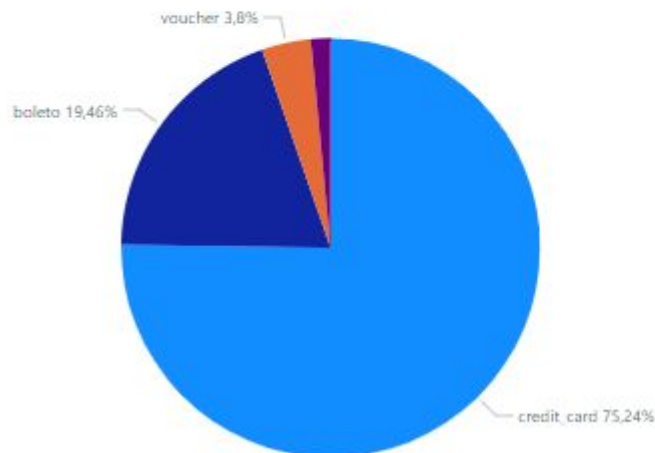
4,09

Promedio de tiempo de entrega
(Dias)

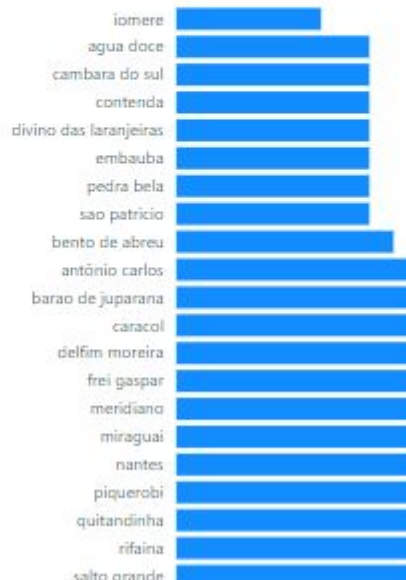
12,50

Medios de pago

credit_card boleto voucher debit_card not_defined

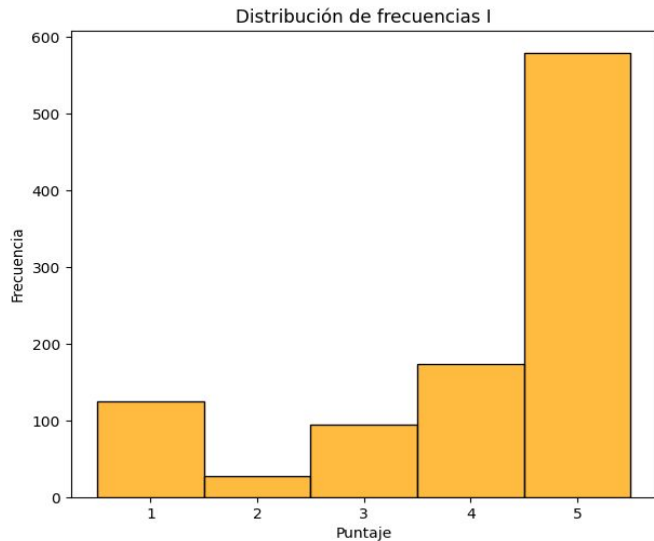


Entrega promedio en dias por ciudad

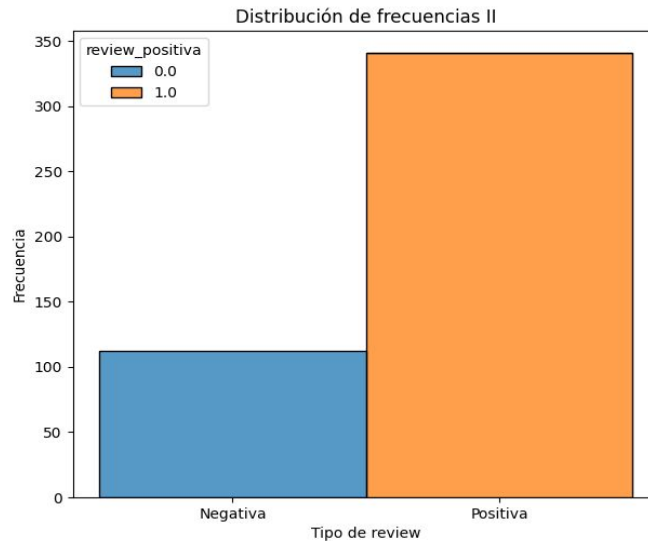


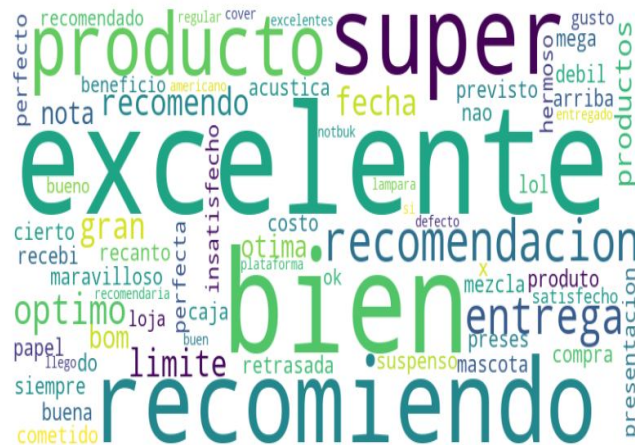
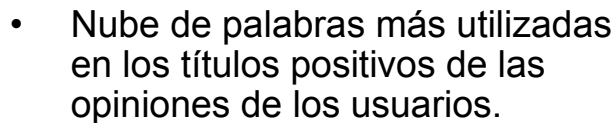
Procesamiento de lenguaje natural

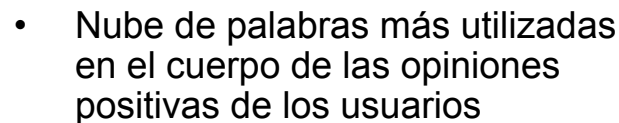
- Histograma de frecuencia del puntaje dado por los usuarios.



- Cantidad de opiniones positivas y negativas







Modelo de Machine Learning



Es una librería distribuida optimizada de gradient boosting diseñada para ser altamente eficiente, flexible y portable. Implementa algoritmos de aprendizaje automático en el marco del Gradient Boosting. XGBoost proporciona un árbol paralelo de refuerzo que resuelve muchos problemas de ciencia de datos de una manera rápida y precisa.

Prototipado



Permite a los desarrolladores transformar scripts de Python en aplicaciones web interactivas con pocas líneas de código.

Streamlit es muy popular entre los científicos de datos por su facilidad de uso y por permitir compartir resultados y modelos de manera interactiva. Soporta una amplia gama de bibliotecas de Python y se integra sin problemas con tecnologías de visualización de datos.

	Actividad	Semana 1					Semana 2				
		Lun 20	Ma 21	Mier 22	Jue 23	Vier 24	Lun 27	Ma 28	Mier 29	Jue 30	Vier 1
Sprint N° 1											
	Definir metodología de trabajo										
	Formulación de KPI'S										
	MVP										
	Selección de tecnologías a utilizar										
	Repositorio Git Hub										
	Redacción de README										
	Presentación										
Sprint N° 2											
	Proceso de ETL										
	Análisis EDA										
	Tabla de Hechos y dimensiones										
	Dashboard										
	Diccionario de datos										
	Stack tecnológico										
	Análisis tabla "opiniones"										
	Panificación modelo ML										
	Presentación										

MUCHAS GRACIAS!

GRUPO 4

