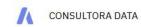


	Propuesta de modelos ML para la empresa TaxiCom 2.0							
Título	Objetivo	Input. Datos de entrada consumidos en BigQuery.	ML	OutPut. Entregable				
Análisis de demanda por hora	Predecir las zonas y horarios con mayor demanda.	(Zonastaxi, taxis_amarillos, taxis_verdes)	Modelos de regresión: Random Forest, XGBoost o LightGBM. Clustering (K-Means, DBSCAN): Para segmentar zonas de alta/baja demanda. Series temporales: ARIMA, SARIMA, o Prophet, para predecir demanda horaria.	Mapa de calor con zonas y horarios de mayor demanda. Predicciones de demanda diaria/semanal para mejorar asignación de recursos. Predicción de consumo de combustible por tipo de vehículo y ruta. Identificación de vehículos menos eficientes y recomendaciones de mejora Reportes de impacto ambiental de taxis y vehículos por tipo de combustible. Mapas de zonas con mayor contaminación generada. Mapa de calor con zonas y horarios de mayor demanda. Predicciones de demanda diaria/semanal para mejorar asignación de recursos. Identificación de vehículos más eficientes y menos contaminantes. Reportes de recomendaciones para reemplazo de vehículos.				
Predicción de profit por zonas	Establecer una predicción de ganancia para taxis amarillos por zona.	Taxis Amarillos		 Mejores zonas de ganancias Predicción en ganancia por cantidad de viajes. 				
Contaminación de los taxis	Correlación entre demanda de taxis por zonas y contaminació n del aire producida por vehículos.			- Gráficos de regresión -				
Sistema de recomendación de zonas	Recomendar zonas de acuerdo a los días de la semana, el mejor horario, ganancia en taxis	Yellow_Tripdata_2024-10.c sv transformed_taxi_zone_me rged_with_locations.csv	XGBoost (XGBRegressor):Hiperparámetros utilizados: n_estimators=200: Número de árboles a construir. max_depth=6: Profundidad máxima de cada árbol. learning_rate=0.1: Tasa de aprendizaje que controla cuánto cambia el modelo en cada	 La api ofrece de acuerdo al día seleccionado la siguiente información La zonas con mayor demanda La hora de dicha demanda Ganancia promedio según la zona, el hora de día y día de la semana La distancia promedio que se recorrerá Cantidad de viajes registrados en la zona 				



	amarillos y para taxis verdes.		iteración. train_test_split (de sklearn):Divide los datos en conjuntos de entrenamiento y prueba para evaluar el rendimiento del modelo. R^2 Score (coeficiente de determinación): Métrica para evaluar la calidad de las predicciones del modelo. Un valor cercano a 1 indica un buen ajuste	- Un url de maps para ir a la zona. - Un gráfico por día con la demanda por zona
Clasificación de vehículos por eficiencia	Clasificar los vehículos en cuatro categorías de eficiencia basadas en su consumo combinado	transformed_Vehicle Fuel Economy Data.csv	- Modelo utilizado: Regresión lineal.	 La API clasifica los vehículos en cuatro categorías de eficiencia basadas en su consumo combinado Un archivo CSV que lista todos los vehículos clasificados por categoría y ordenados de mayor a menor eficiencia. Gráficos visuales para cada métrica clave (e.g., rango, consumo en ciudad) que permiten comparar rápidamente los vehículos más destacados. Predicción del costo basado en los valores promedio del modelo
Calidad de aire en NY	Análisis Histórico y predicción de la contaminació n de aire (Pm 2.5)	sensores 22-24.csv	Primer modelo series de tiempo: Uso del algoritmo Arima Segundo modelo series de tiempo: Uso del algoritmo Prophet	- Graficos y analisis - Predicción a 1 - 3 meses.
	Clustering Basado en Sitiossegún el índice de contaminació n.	MergeValueLocation.scv = location.csv + sensores 22-24	Clustering: Modelos utilizados: Kmeans , regresión lineal.	 Api de agrupación por Índice de contaminación Función que devuelve, con los datos de los sensores y sitios, un mapa con la zona seleccionada en color (según el índice de métricas) y sus datos de contaminación por número de mes, día y hora.
	Agrupar las Boroughs en función de sus característica s clave	Air_quality.csv Boroughts_clusters	 Obtención del número óptimo de clusters con elbow Aplicación del algoritmo k means Análisis de correlación con regresión lineal. 	 Gráficos aleatorios de zonas con variables Gráfico de regresión lineal que muestra la correlación entre tránsito y pm 2.5 Gráfico de las 10 zonas con más correlación respecto a la recta de regresión.

Objetivo	ML	Entregable
Conocer las zonas de alta demanda para optimización del negocio.	Regresión lineal. Kmeans	API de recomendación de zonas en base a la demanda de usuarios
Mostrar el desarrollo temporal de la calidad del aire en función de zonas.	Prophet Arima Clustering: Kmeans, regresión lineal.	Análisis Histórico y predicción de la contaminación de aire (Pm 2.5)
Prever y predecir posibles ganancias en función a zonas de demanda de taxis amarillos y verdes.	Prophet Clustering: Kmeans, regresión lineal.	Gráfico de proyección de ganancias. Función que recomienda las zonas de posibles sobreganancias.
Clasificar vehículos en base a su eficiencia y previsión de gasto.	Regresión lineal. DBSCAN	API de clasificación de vehículos en base a categorías de eficiencia.