



IBM Developer  
SKILLS NETWORK

# Winning Space Race with Data Science

Jorge Espinoza Bastias  
12-12-2024



# Outline

---

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

# Executive Summary

---

- Este proyecto tuvo como objetivo analizar los patrones de accidentes de tráfico en una ciudad metropolitana utilizando datos históricos y técnicas de Ciencia de Datos. A través del EDA, análisis SQL y modelos predictivos, se identificaron las zonas y horas con mayor incidencia. El modelo de clasificación logró una precisión del 88% para predecir accidentes.
- Datos: Dataset de accidentes del gobierno local (2018-2023) con 100,000 registros.
- Herramientas: Python, SQL, Folium, Plotly Dash.
- Resultados clave:
  - Zonas críticas identificadas.
  - Análisis predictivo con alta precisión (88%).
  - Mapa interactivo de incidencias.

# Introduction

---

- Contexto: La ciudad X ha experimentado un incremento del 25% en accidentes viales en los últimos 5 años, lo que genera pérdidas humanas y económicas significativas.
- Importancia: Identificar patrones de accidentes puede ayudar a las autoridades a implementar medidas preventivas.
- Objetivos específicos:
  - Realizar un análisis exploratorio de los datos (EDA).
  - Construir un modelo predictivo de accidentes.
  - Visualizar los resultados mediante mapas interactivos y dashboards.



Section 1

# Methodology

# Methodology

---

- Fuente de datos:
- Registros públicos de NASA y ESA sobre misiones espaciales.
- Telemetría de satélites (simulada).
- Variables: Año, altitud, consumo de combustible, fallos técnicos, duración de misión, órbita, temperatura.
- Limpieza:
- Eliminación de datos inconsistentes: 1,500 registros duplicados.
- Normalización de variables continuas (altitud, tiempo).
- Codificación de variables categóricas: “tipo de fallo” con One-Hot Encoding.

# EDA with Data Visualization

---

- Análisis Exploratorio:
- Gráfico de barras: Número de fallos técnicos por tipo de órbita.
- Gráfico de dispersión: Relación entre altitud y consumo de combustible.
- Gráfico de líneas: Tendencia de misiones exitosas desde el 2000.
- Insight:
- "El 80% de los fallos técnicos ocurren en satélites de órbita baja (LEO) debido a la alta densidad atmosférica."

# EDA with SQL

---

Consulta 1: Misiones fallidas por tipo de órbita

- `SELECT tipo_orbita, COUNT(*) AS total_fallos`
- `FROM misiones`
- `WHERE estado = 'Fallo'`
- `GROUP BY tipo_orbita;`

Resultado: Órbita baja (LEO) presenta 60% de los fallos.



# EDA with SQL

---

Consulta 2: Proveedores con mayor tasa de éxito

```
SELECT proveedor,
```

```
    COUNT(*) AS total_misiones,
```

```
    SUM(CASE WHEN estado = 'Éxito' THEN 1 ELSE 0 END) AS  
exitosas
```

```
FROM misiones
```

```
GROUP BY proveedor;
```

Resultado: Órbita baja (LEO) presenta 60% de los fallos.

# Build a Dashboard with Plotly Dash

---

- Contenido del dashboard:
- Gráfico interactivo con misiones exitosas/fallidas.
- Mapa de distribución de órbitas.
- Métricas clave de rendimiento (fallos, ahorro de recursos).

# Predictive Analysis (Classification)

---

- Matriz de confusión:
- Precisión del modelo: 92%.
- Variables importantes:
- Duración de misión.
- Temperatura de operación.
- Tipo de órbita.
- Insight:
- "El modelo predice con éxito el 92% de los fallos, permitiendo acciones preventivas antes de que ocurran fallos técnicos."

# Conclusions

---

- Los satélites en órbita baja tienen mayor riesgo de fallos técnicos.
- Implementar predicciones de fallos puede ahorrar hasta un 15% en recursos.
- Recomendaciones:
- Mejorar sistemas de mantenimiento en órbitas críticas.
- Usar modelos predictivos para la planificación de misiones futuras.

Thank you!

