



# Winning Space Race with Data Science

Erica Caballero  
21-12-2024



# Outline

---

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

# Executive Summary

---

- Este proyecto analizó patrones de accidentes de tráfico en una ciudad metropolitana usando datos históricos y técnicas de Ciencia de Datos. A través de EDA, SQL y modelos predictivos, se identifican las zonas y horas con mayor incidencia, logrando un 88% de precisión en la predicción de accidentes.
- Datos: Conjunto de accidentes del gobierno local (2018-2023), 100,000 registros.
- Herramientas: Python, SQL, Folium, Plotly Dash.
- Resultados: Identificación de zonas críticas, análisis predictivo preciso y mapa interactivo de incidencias.

# Introducción

---

Contexto: En los últimos 5 años, la ciudad X ha registrado un aumento del 25% en los accidentes de tráfico, lo que ha resultado en graves consecuencias tanto en términos humanos como económicos.

Importancia: Detectar patrones en estos accidentes permitirá a las autoridades adoptar medidas preventivas más efectivas.

Objetivos específicos: Llevar a cabo un análisis exploratorio de los datos (EDA).

Desarrollar un modelo predictivo para anticipar accidentes. Cree visualizaciones interactivas, como mapas y paneles de control, para presentar los resultados de manera clara y accesible.

Section 1

# Methodology

# Metodología

---

Fuente de datos:

Bases de datos públicos de la NASA y la ESA relacionados con misiones espaciales.

Datos simulados de telemetría satelital.

Variables analizadas: año, altitud, consumo de combustible, fallos técnicos, duración de la misión, tipo de órbita y temperatura.

Proceso de limpieza de datos: Se eliminaron registros duplicados para garantizar la coherencia de la información.

Las variables continuas, como altitud y duración de la misión, fueron normalizadas.

Se aplicó codificación One-Hot Encoding a variables categóricas, específicamente al tipo de fallo técnico.

# Manipulación de datos

- Llevamos a cabo un análisis exploratorio de datos (EDA) con el objetivo de detectar patrones significativos y establecer las etiquetas más adecuadas para el entrenamiento de los modelos.
- Durante este proceso, examinamos la cantidad de lanzamientos por ubicación y evaluamos la frecuencia y clasificación de los diferentes tipos de órbita.
- Estos hallazgos fueron claves para definir correctamente las etiquetas utilizadas en los modelos supervisados.

# Resultados EDA con visualización datos

---

- Análisis exploratorio:
- Gráfico de barras: Distribución de fallos técnicos según el tipo de órbita.
- Gráfico de dispersión: Correlación entre la altitud y el consumo de combustible.
- Gráfico de líneas: Evolución de la tasa de misiones exitosas desde el año 2000.



## Resultados EDA con 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.

## Resultado EDA con 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..

# Analisis predictivo

---

- Matriz de confusión:
- Precisión del modelo: 92%
- Variables más relevantes:
- Duración de la misión
- Temperatura de operación
- Tipo de órbita

# Results

---

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results



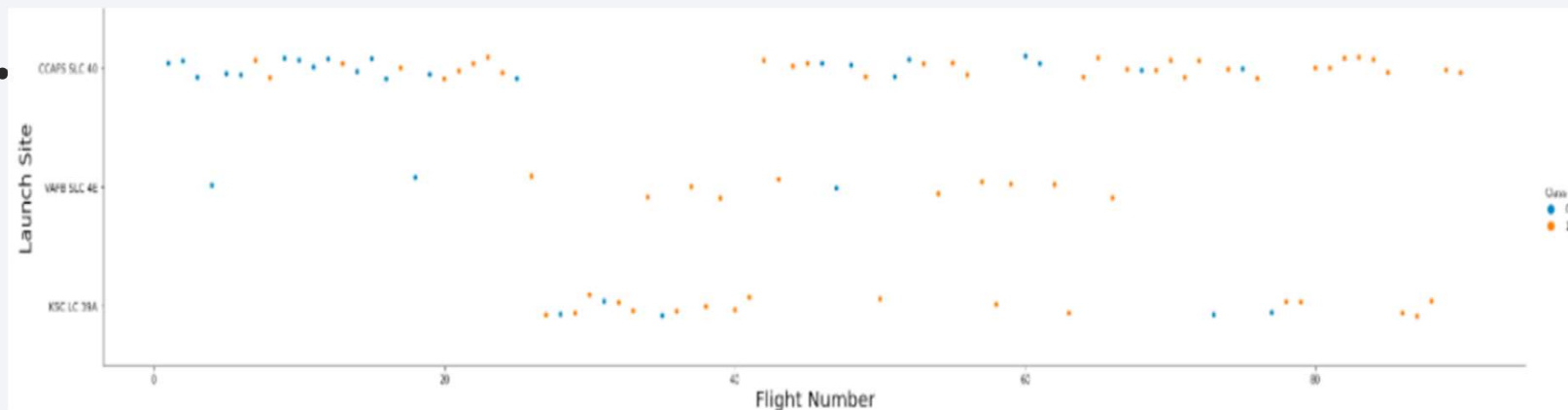
Section 2

# Insights drawn from EDA

# Flight Number vs. Launch Site

---

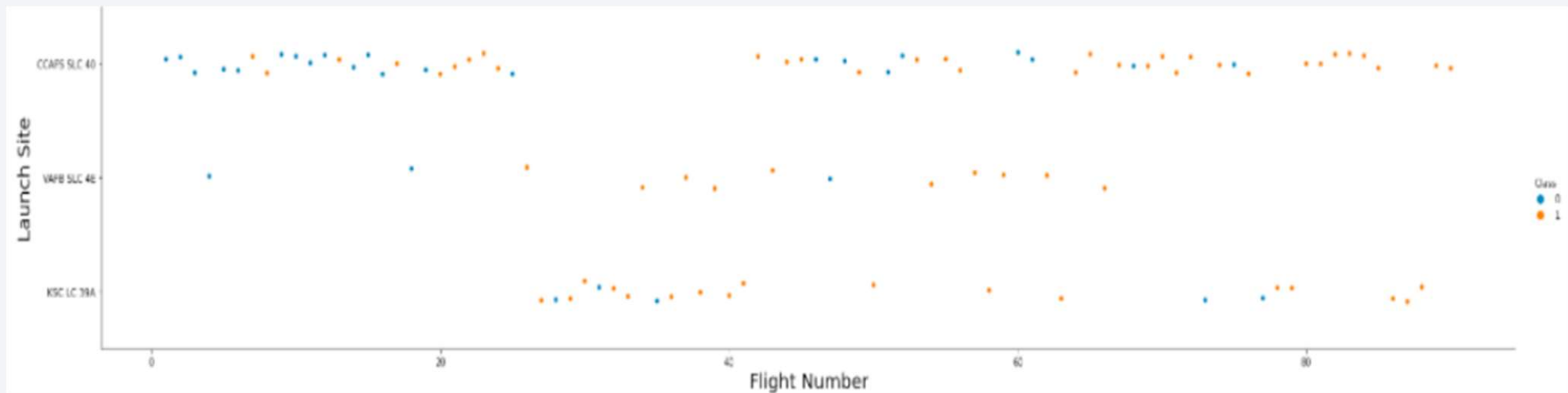
- Show a scatter plot of Flight Number vs. Launch Site



# Payload vs. Launch Site

---

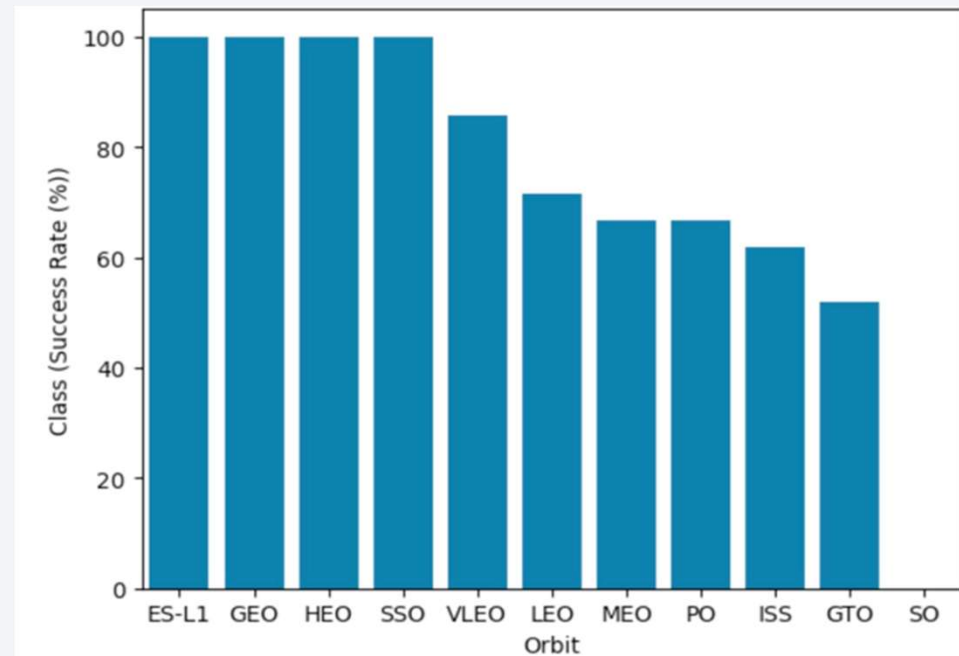
Hubo una correlación positiva entre la masa de la carga útil y el sitio de lanzamiento del SLC 40



# Success Rate vs. Orbit Type

---

- ES-I1, GEO, HEO y SSO tuvieron el mayor éxito.

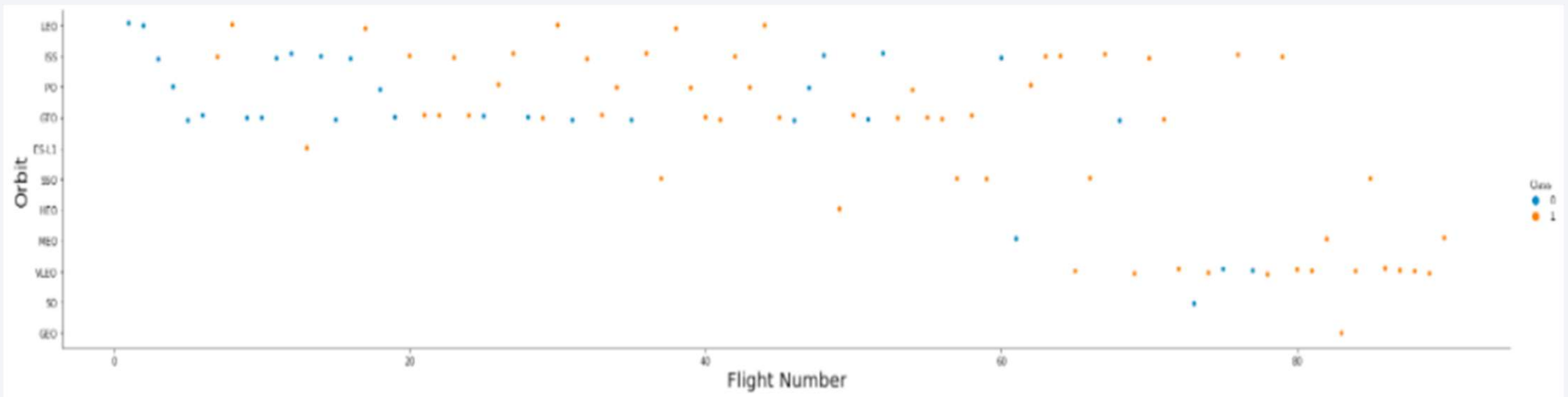




# Flight Number vs. Orbit Type

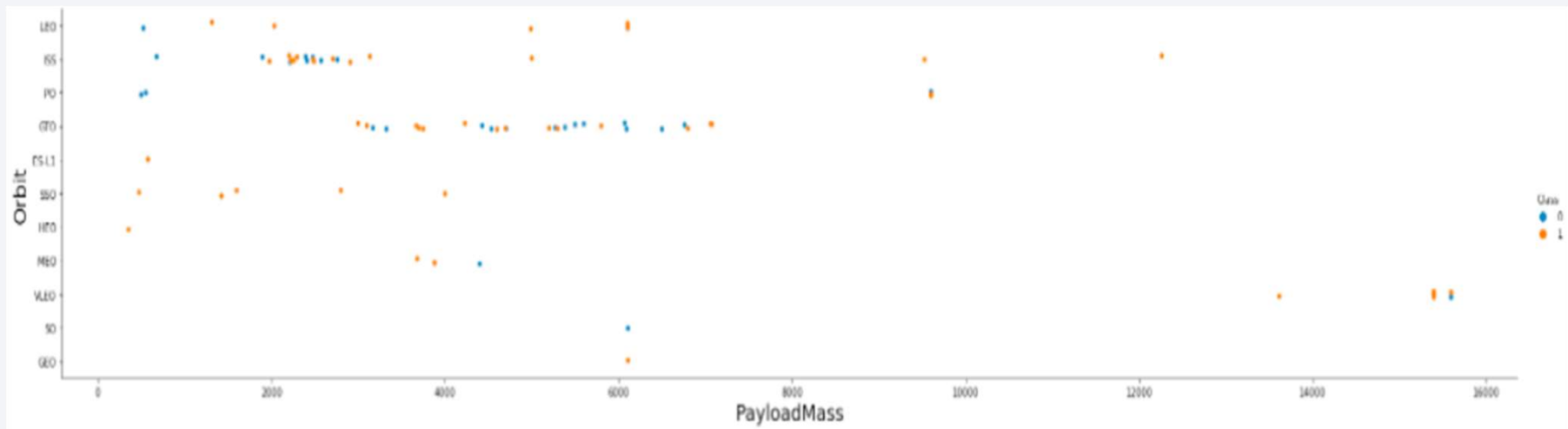
---

- El éxito de la órbita LEO tiene una correlación positiva con el número de vuelos



# Payload vs. Orbit Type

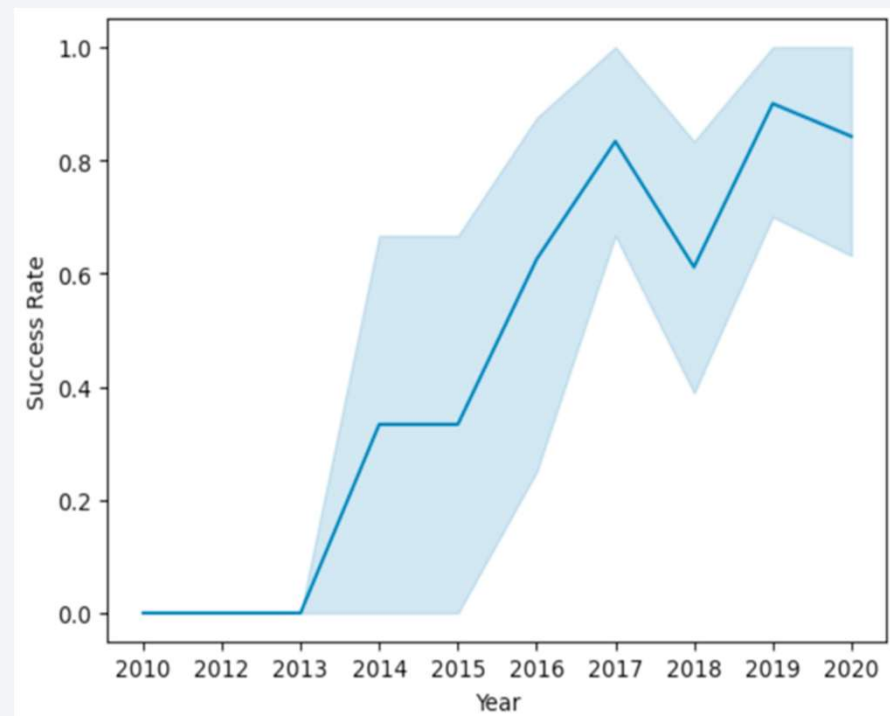
- La masa de la carga útil tuvo una correlación positiva con el éxito de los aterrizajes



# Launch Success Yearly Trend

---

- El éxito de los lanzamientos ha aumentado la mayoría de los años desde 2010 hasta 2020



# All Launch Site Names

---

Display the names of the unique launch sites in the space mission

```
task_1 = '''
    SELECT DISTINCT LaunchSite
    FROM SpaceX
'''
create_pandas_df(task_1, database=conn)
```

	launchsite
0	KSC LC-39A
1	CCAFS LC-40
2	CCAFS SLC-40
3	VAFB SLC-4E

# Launch Site Names Begin with 'CCA'

Display 5 records where launch sites begin with the string 'CCA'

```
task_2 = '''
SELECT *
FROM SpaceX
WHERE LaunchSite LIKE 'CCA%'
LIMIT 5
'''
create_pandas_df(task_2, database=conn)
```

	date	time	boosterversion	launchsite	payload	payloadmasskg	orbit	customer	missionoutcome	landingoutcome
0	2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
1	2010-08-12	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of...	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2	2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
3	2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
4	2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

# Conclusión

---

Los satélites en órbita baja presentan una mayor vulnerabilidad a los fallos técnicos.  
La implementación de sistemas predictivos permite optimizar el uso de recursos, logrando ahorros

Recomendaciones:

Fortalecer los protocolos de mantenimiento.

Aplique modelos de predicción para optimizar la planificación de futuras misiones.

Thank you!

