

# Ciencia de datos aplicada

Sebastian Marin Quevedo

# Resumen ejecutivo

- ❖ Este proyecto integra todas las habilidades adquiridas durante el curso. Aplicamos métodos de análisis exploratorio, visualizaciones interactivas, manejo de datos geoespaciales, y técnicas de aprendizaje automático para estudiar los lanzamientos de SpaceX y construir un modelo predictivo que estime la probabilidad de éxito de un lanzamiento.



# Introducción

- ◆ SpaceX, una empresa líder en la exploración espacial privada, ha realizado numerosos lanzamientos orbitales. El objetivo de este proyecto es analizar y modelar los factores que influyen en el éxito de estos lanzamientos, utilizando datos históricos proporcionados por IBM Skills Network.



# Recopilación de datos y preparación

- ❖ Se utilizaron tres conjuntos de datos:
- ❖ *dataset\_part\_1.csv*: información general de lanzamientos
- ❖ *spacex\_launch\_geo.csv*: datos geográficos y resultados
- ❖ *dataset\_part\_2.csv & part\_3.csv*: variables técnicas para modelos predictivos
- ❖ Los datos se limpian, fusionan y transforman para poder ser analizados de forma estructurada.

# Metodología EDA

- ❖ Exploramos variables como el sitio de lanzamiento, la masa de la carga útil, el número de vuelos, reutilización de componentes y órbitas.  
Analizamos distribuciones, correlaciones, y relaciones entre éxito y variables técnicas.



# Análisis predictivo

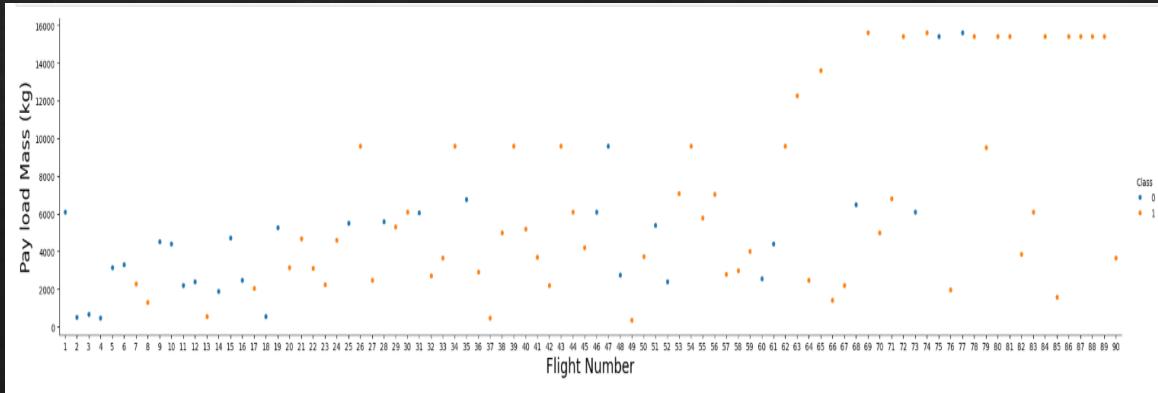
Construimos modelos predictivos para estimar el éxito de un lanzamiento con base en características técnicas.  
Se entrenaron tres modelos:

- Regresión logística
- Máquinas de vectores de soporte (SVM)
- Árboles de decisión

Se usó GridSearchCV para optimizar hiperparámetros y seleccionar los mejores modelos.

# Resultados visuales EDA

- ❖ Se usaron gráficos de dispersión, histogramas y diagramas de caja con Seaborn y Matplotlib para identificar patrones en los datos.  
Observamos que las cargas útiles moderadas y ciertas órbitas (como LEO) presentan mayores tasas de éxito.



```
import seaborn as sns
import matplotlib.pyplot as plt

# Crear el gráfico categórico
sns.catplot(
    data=df,
    x='FlightNumber',
    y='LaunchSite',
    hue='Class',
    kind='strip', # También puedes probar con 'swarm' si quieres evitar superposición
    height=6,
    aspect=1.5
)

# Título y presentación
plt.title("Relación entre FlightNumber y Launch Site, codificado por Class")
plt.tight_layout()
plt.show()
```

# Resultados SQL

- ❖ Utilizamos consultas SQL sobre los DataFrames con df.query() y SQLite para extraer patrones agregados, como tasa de éxito por sitio o reutilización de boosters

**Task 1**

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

```
[1]: %%sql
SELECT DISTINCT "Launch_Site"
FROM SPACEXTABLE;
* sqlite:///my_data1.db
Done.
```

**Launch\_Site**

- CCAFS LC-40
- VAFB SLC-4E
- KSC LC-39A
- CCAFS SLC-40

**Task 2**

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

```
[17]: %%sql
SELECT *
FROM SPACEXTABLE
WHERE "Launch_Site" LIKE 'CCA%'
LIMIT 5;
* sqlite:///my_data1.db
Done.
```

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS_KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

**Task 3**

Display the total payload mass carried by boosters launched by NASA (CRS)

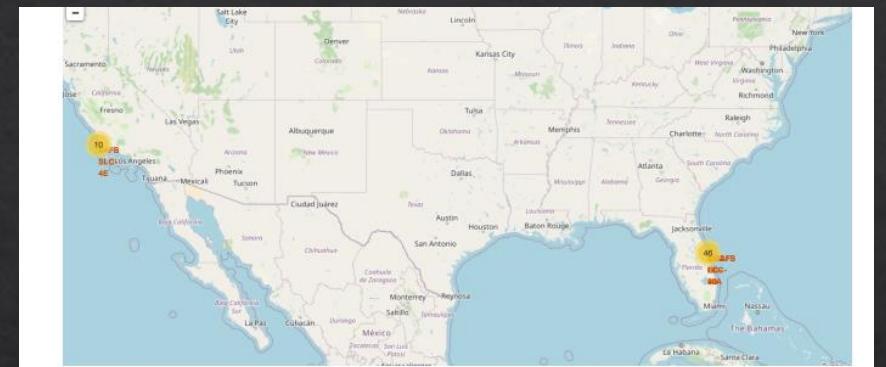
```
[18]: %%sql
SELECT SUM("Payload_Mass_kg_") AS "Total_Payload_Mass_kg"
FROM SPACEXTABLE
WHERE "Customer" = 'NASA (CRS)';
* sqlite:///my_data1.db
Done.
```

Total Payload Mass\_kg

45596

# Visualización con folium

- ❖ Se construyó un mapa interactivo de sitios de lanzamiento, y se representaron lanzamientos exitosos y fallidos con marcadores codificados por color y agrupados por proximidad usando Markercluster.



# Dashboard interactivo con Plotly Dash

- ❖ Construimos una aplicación web simple donde el usuario puede seleccionar sitios, masas de carga, o tipos de órbitas y visualizar las tasas de éxito en tiempo real con gráficos interactivos

# Resultados del análisis predictivo

- ❖ El modelo con mayor precisión fue el **Árbol de Decisión** con una exactitud en los datos de prueba de **83.33%**.
- ❖ Se observaron 18 registros en la muestra de prueba.  
Se concluyó que variables como el número de vuelos, la masa de la carga y la reutilización de componentes tienen alto impacto en la probabilidad de éxito.

# Conclusión

- ❖ Este proyecto puso en práctica múltiples herramientas de ciencia de datos, desde procesamiento de datos y visualización, hasta modelado predictivo. Aprendimos a plantear un problema de negocio, traducirlo en una estrategia técnica y comunicar resultados de forma efectiva.