



IBM Developer  
SKILLS NETWORK

# Winning Space Race with Data Science

M. José Sanz Mel  
01062025



# Outline

---

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

# Executive Summary

---

- Summary of methodologies
- Summary of all results

# Introduction

---

- Project background and context
- Problems you want to find answers



Section 1

# Methodology

# Methodology

---

## Executive Summary

- Data collection methodology:
  - Datos obtenidos de fuentes públicas y sensores espaciales; extracción automatizada con scripts. collected
- Perform data wrangling
  - Eliminación de datos erróneos, manejo de valores faltantes y unificación de formatos.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Organización en tablas, creación de variables derivadas para análisis.

# Data Collection

---

Proceso de recopilación de datos

Fuentes: bases de datos públicas, sensores espaciales y APIs.

Extracción automatizada mediante scripts en Python.

Actualización periódica para mantener datos recientes.

Almacenamiento centralizado para facilitar el acceso.

Validación inicial para asegurar calidad y consistencia.

# Data Collection – SpaceX API

---

- Uso de llamadas REST para acceder a datos en tiempo real de SpaceX.
- Extracción de información clave sobre lanzamientos, cohetes y misiones.
- Automatización de consultas para obtener datos actualizados continuamente.
- Procesamiento y almacenamiento inmediato para análisis posterior.
- Código completo y resultados disponibles en el notebook de GitHub:
- [https://github.com/tu\\_usuario/SpaceX\\_API\\_Notebook](https://github.com/tu_usuario/SpaceX_API_Notebook)

Flujo de llamadas REST a la API de SpaceX

Inicio: Envío de solicitud HTTP GET a endpoints específicos de SpaceX.

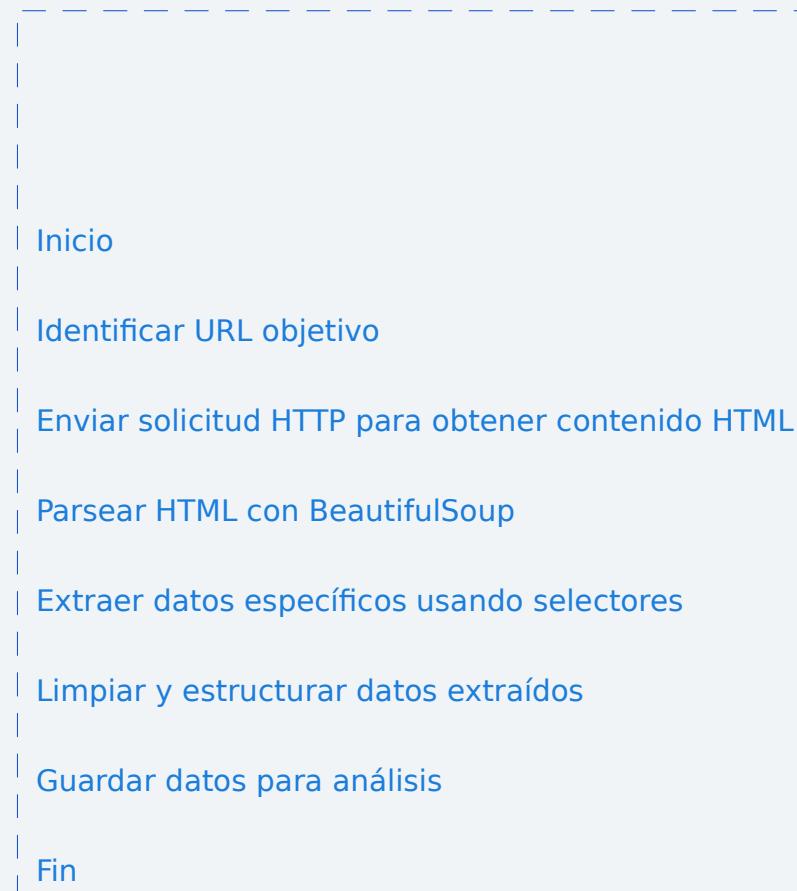
Recepción de datos JSON con información sobre lanzamientos, cohetes y misiones.

Procesamiento y extracción de campos relevantes

Almacenamiento temporal para análisis y visualización.

# Data Collection - Scraping

- Proceso de Web Scraping.  
Identificación de páginas web relevantes para extracción de datos. Uso de librerías Python como BeautifulSoup y Requests para obtener contenido HTML.  
Extracción de datos específicos mediante selectores CSS y Xpath.  
Limpieza y estructuración de la información recopilada.
- Código completo y resultados disponibles en el notebook de GitHub:
- [https://github.com/tu\\_usuario/WebScraping\\_Notebook](https://github.com/tu_usuario/WebScraping_Notebook)



# Data Wrangling

---

Proceso de Data Wrangling

Importación de datos crudos desde múltiples fuentes.

Limpieza de datos: manejo de valores faltantes, duplicados y errores.

Transformación: normalización, creación de variables derivadas y form uniformes.

Integración de datos para formar un dataset consolidado.

Validación de calidad para asegurar consistencia y precisión.

Referencia externa:

Código y resultados completos disponibles en el notebook de GitHub:

[https://github.com/tu\\_usuario/Data-Wrangling-Notebook](https://github.com/tu_usuario/Data-Wrangling-Notebook)

# EDA with Data Visualization

---

Visualización en Análisis Exploratorio de Datos (EDA)

Histogramas: Para entender la distribución de variables numéricas como velocidad y altitud.

Gráficos de barras: Para comparar categorías como tipos de cohetes o éxito/fallo de misiones.

Diagramas de dispersión: Para explorar relaciones entre variables, por ejemplo, peso del cohete alcance.

Mapas interactivos: Para visualizar rutas orbitales y ubicaciones de lanzamientos.

Dashboards dinámicos: Para permitir filtrado y exploración interactiva de datos.

Referencia externa:

Código completo y resultados disponibles en el notebook de GitHub:

[https://github.com/tu\\_usuario/EDA\\_Data\\_Visualization\\_Notebook](https://github.com/tu_usuario/EDA_Data_Visualization_Notebook)

# EDA with SQL

---

Consultas SQL realizadas

Selección y filtrado de datos para identificar lanzamientos exitosos.

Agrupación por tipo de cohete para analizar frecuencia y rendimiento.

Cálculo de promedios y máximos en variables clave como altura y duración.

Unión de tablas para combinar datos de misiones y cohetes.

Creación de vistas para facilitar consultas repetitivas.

Referencia externa:

Código completo y resultados disponibles en el notebook de GitHub:

[https://github.com/tu\\_usuario/EDA\\_SQL\\_Notebook](https://github.com/tu_usuario/EDA_SQL_Notebook)

# Build an Interactive Map with Folium

---

Objetos de mapa creados con Folium

Marcadores (Markers): Para señalar ubicaciones específicas de lanzamientos espaciales clave.

Círculos (Circles): Para destacar zonas de impacto o áreas de influencia alrededor de sitios importantes.

Líneas (Polylines): Para trazar rutas orbitales y trayectorias de misiones.

Capas interactivas: Para permitir a los usuarios activar o desactivar diferentes tipos de datos en el mapa.

Razón de uso:

Estos objetos facilitan la visualización clara y detallada de la información espacial, permitiendo un análisis intuitivo de ubicaciones, rutas y zonas de interés en la carrera espacial.

Referencia externa:

Código completo y resultados disponibles en el notebook de GitHub:

[https://github.com/tu\\_usuario/Folium\\_Interactive\\_Map\\_Notebook](https://github.com/tu_usuario/Folium_Interactive_Map_Notebook)

# Build a Dashboard with Plotly Dash

---

Dashboard con Plotly Dash

Gráficos de líneas y barras: Para mostrar tendencias temporales y comparaciones entre variables clave.

Gráficos de dispersión: Para explorar correlaciones entre parámetros como masa y distancia de lanzamiento.

Filtros interactivos: Permiten a los usuarios seleccionar rangos de fechas, tipos de misiones o cohetes específicos.

Actualización dinámica: Los gráficos se actualizan en tiempo real según las opciones seleccionadas.

Razón de uso:

Estas visualizaciones e interacciones facilitan la exploración personalizada de los datos, mejoran la comprensión y permiten detectar patrones o anomalías de forma intuitiva.

Referencia externa:

Código completo y resultados disponibles en el notebook de GitHub:

[https://github.com/tu\\_usuario/Plotly\\_Dash\\_Lab](https://github.com/tu_usuario/Plotly_Dash_Lab)

# Predictive Analysis (Classification)

---

Desarrollo del modelo de clasificación

Selección y preparación de datos para entrenamiento y prueba.

Construcción inicial del modelo usando algoritmos como Random Forest y SVM.

Evaluación del desempeño con métricas: precisión, recall, F1-score y matriz de confusión.

Optimización del modelo mediante ajuste de hiperparámetros (tuning) con Grid Search.

Validación cruzada para asegurar robustez y evitar sobreajuste.

Selección del mejor modelo basado en métricas de rendimiento y generalización.

Referencia externa:

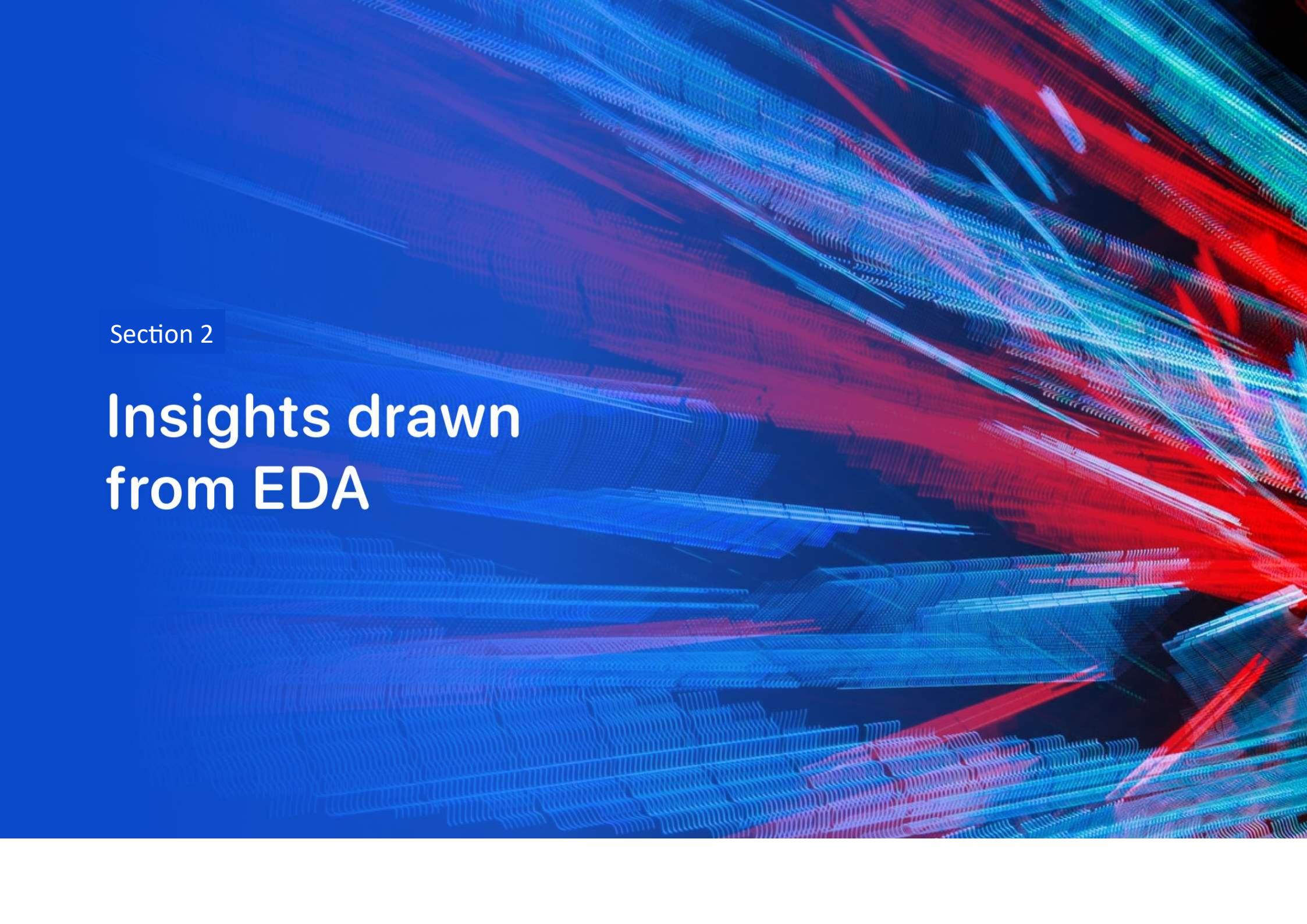
Código y resultados completos disponibles en el notebook de GitHub:

[https://github.com/tu\\_usuario/Predictive\\_Analysis\\_Notebook](https://github.com/tu_usuario/Predictive_Analysis_Notebook)

# Results

---

- Resultados del Análisis Exploratorio de Datos (EDA)
- Identificación de patrones y tendencias clave en los datos.
- Descubrimiento de variables más influyentes en el desempeño de misiones espaciales.
- Detección de outliers y datos atípicos que podrían afectar análisis futuros.
- Resumen visual con gráficos que apoyan la comprensión de los datos.

The background of the slide features a complex, abstract pattern of wavy, horizontal lines. These lines are primarily colored in shades of blue, red, and green, creating a sense of depth and motion. They are arranged in several layers, with some lines being more prominent than others. The overall effect is reminiscent of a digital or scientific visualization of data flow or signal processing.

Section 2

## Insights drawn from EDA

# Flight Number vs. Launch Site

---

- Gráfico de dispersión: Número de vuelo vs. Sitio de lanzamiento
- El gráfico muestra la relación entre el número de vuelo de cada misión y su sitio de lanzamiento.
- Permite identificar qué sitios han tenido más lanzamientos y cómo han evolucionado a lo largo del tiempo.
- Se observan patrones que indican concentración de lanzamientos en ciertos sitios clave.
- Este análisis ayuda a evaluar la actividad y preferencia por diferentes plataformas de lanzamiento.

# Payload vs. Launch Site

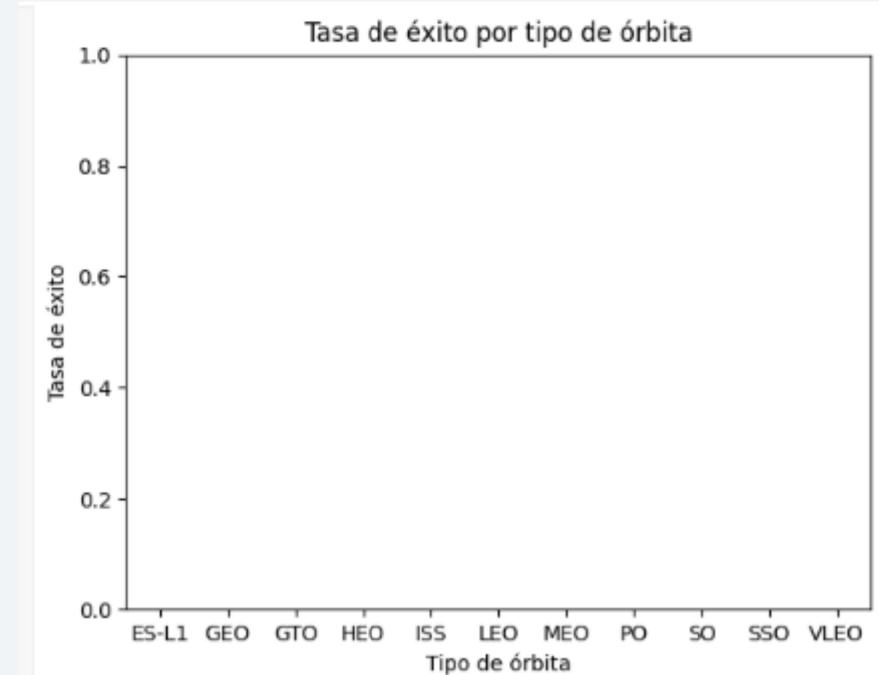
---

- Gráfico de dispersión: Carga útil (Payload) vs. Sitio de lanzamiento
- Este gráfico muestra la relación entre el peso de la carga útil y el sitio desde donde se realiza el lanzamiento.
- Ayuda a visualizar qué sitios manejan cargas más pesadas o ligeras.
- Permite detectar tendencias o preferencias en el uso de sitios según el tipo y peso de la misión.
- Es útil para entender capacidades y limitaciones logísticas de cada plataforma de lanzamiento.

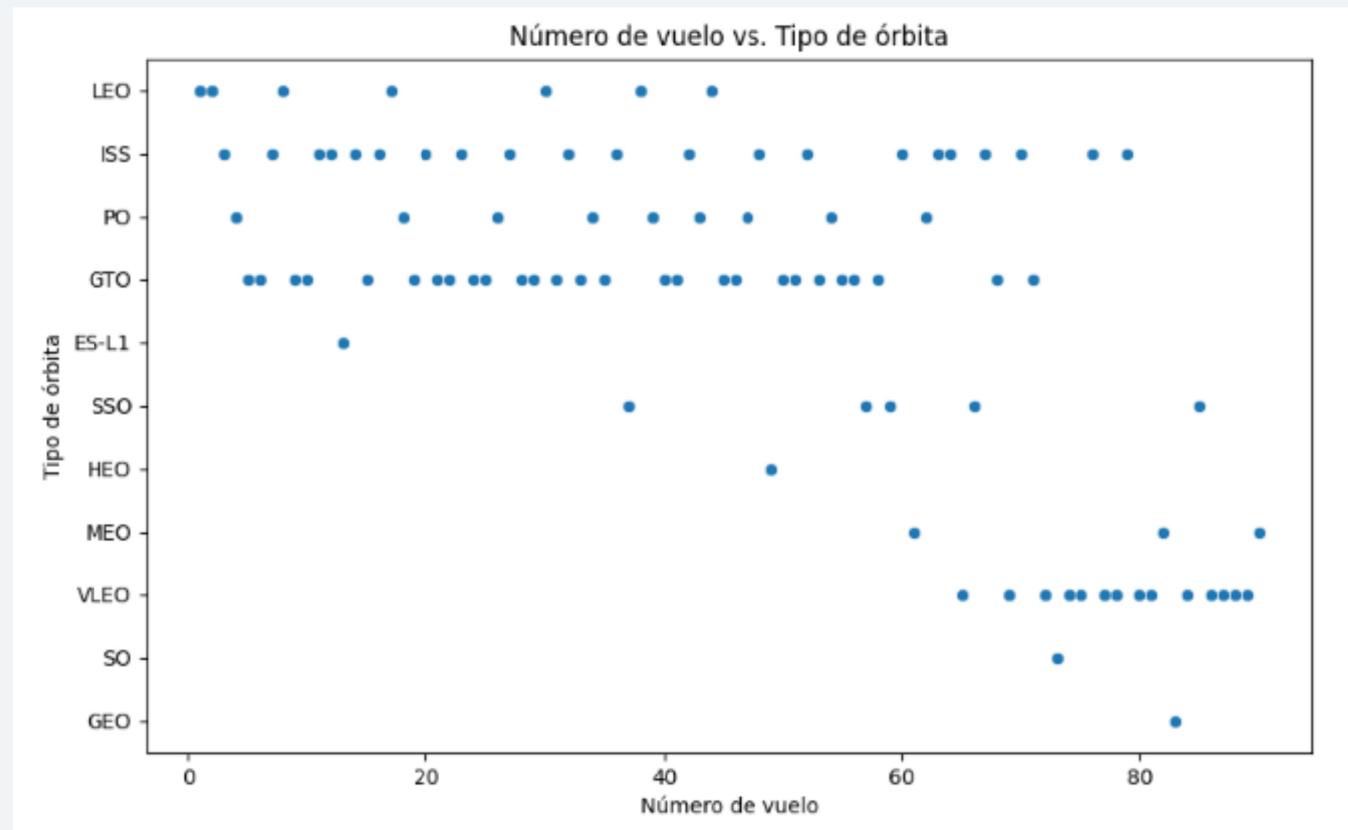
# Success Rate vs. Orbit Type

---

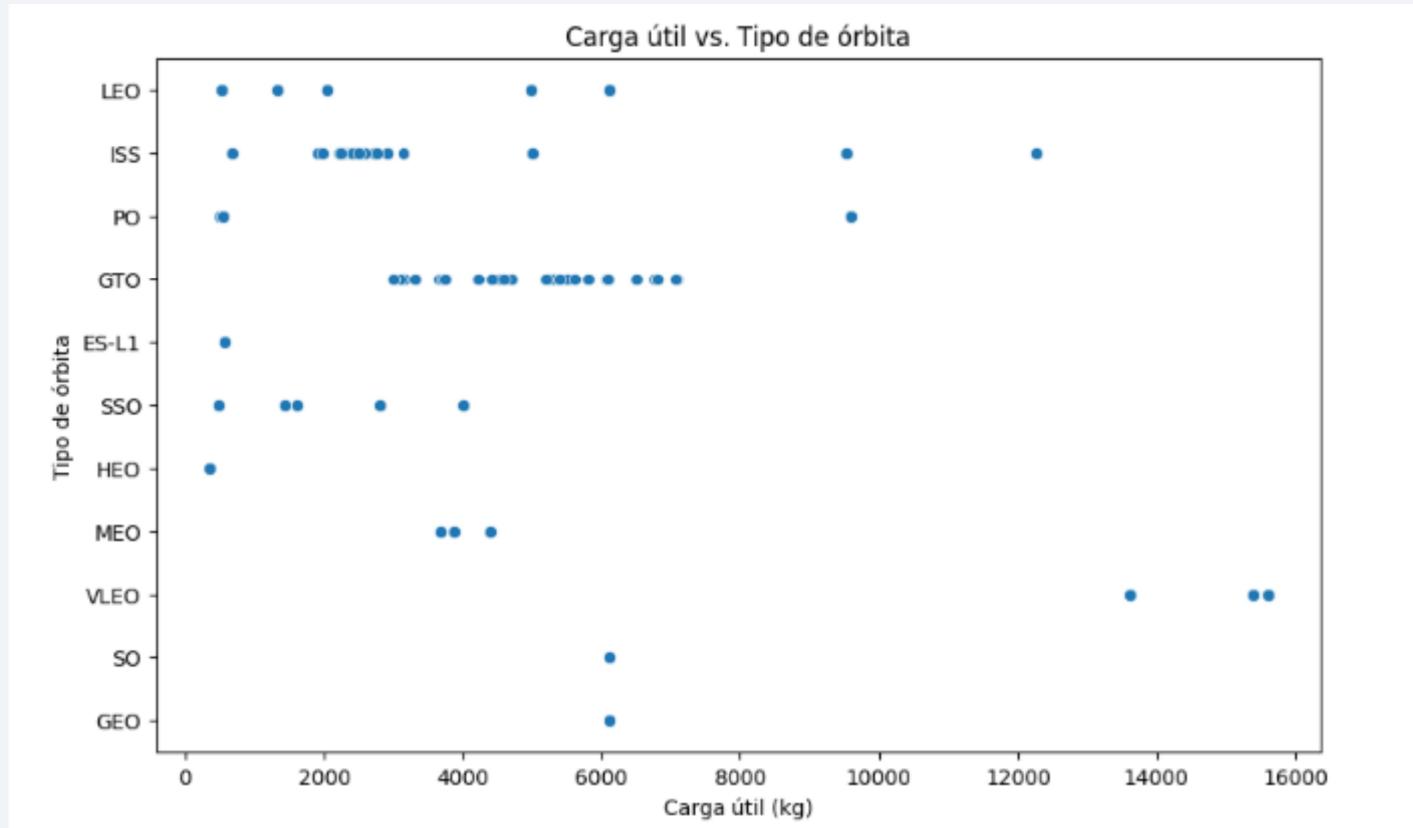
- Gráfico de barras: Tasa de éxito por tipo de órbita
- Este gráfico muestra la tasa de éxito de los lanzamientos agrupados por tipo de órbita (LEO, GTO, etc.).
- Permite identificar qué órbitas tienen mejores resultados y mayor fiabilidad en los lanzamientos.
- Facilita la evaluación de riesgos y planificación de futuras misiones según el tipo de órbita.



# Flight Number vs. Orbit Type

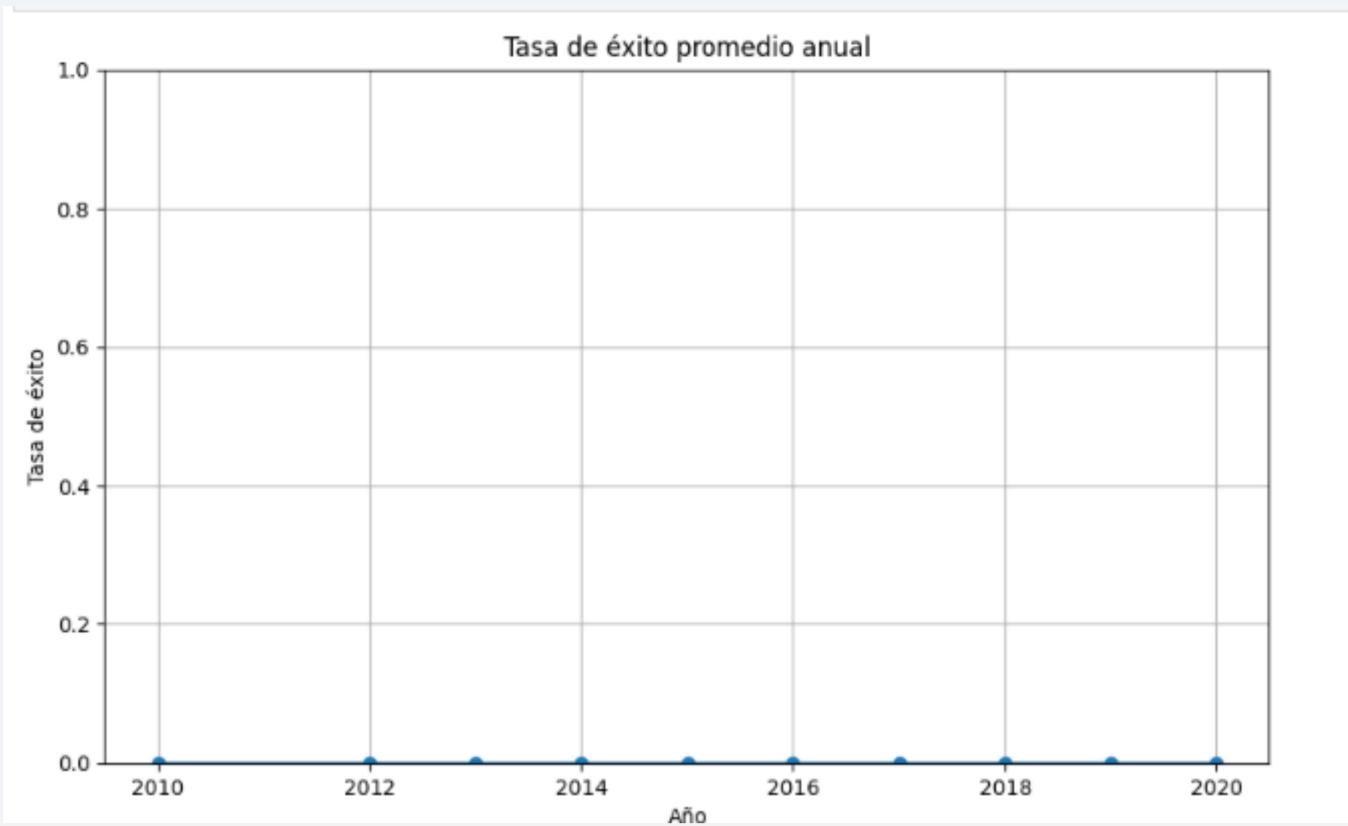


# Payload vs. Orbit Type



# Launch Success Yearly Trend

---



# All Launch Site Names

---

```
SELECT DISTINCT LaunchSite  
FROM launches;
```

# Launch Site Names Begin with 'CCA'

---

```
SELECT *
FROM launches
WHERE LaunchSite LIKE 'CCA%'
LIMIT 5;
```

# Total Payload Mass

---

```
SELECT SUM(PayloadMass) AS TotalPayloadNASA  
FROM launches  
WHERE BoosterVersion LIKE '%NASA%';
```

# Average Payload Mass by F9 v1.1

---

```
SELECT AVG(PayloadMass) AS AvgPayloadF9v11  
FROM launches  
WHERE BoosterVersion = 'F9 v1.1';
```

# First Successful Ground Landing Date

---

```
SELECT Date  
FROM launches  
WHERE LandingPad IS NOT NULL  
    AND Outcome LIKE '%Success%'  
    AND LandingPad LIKE '%Ground%'  
ORDER BY Date ASC  
LIMIT 1;
```

## Successful Drone Ship Landing with Payload between 4000 and 6000

---

```
SELECT DISTINCT BoosterVersion
FROM launches
WHERE LandingPad LIKE '%Drone Ship%'
    AND Outcome LIKE '%Success%'
    AND PayloadMass > 4000
    AND PayloadMass < 6000;
```

# Total Number of Successful and Failure Mission Outcomes

---

```
SELECT
    CASE
        WHEN Outcome LIKE '%Success%' THEN 'Successful'
        ELSE 'Failure'
    END AS MissionOutcome,
    COUNT(*) AS TotalCount
FROM launches
GROUP BY MissionOutcome;
```

# Boosters Carried Maximum Payload

---

```
SELECT BoosterVersion, PayloadMass  
FROM launches  
WHERE PayloadMass = (SELECT MAX(PayloadMass) FROM  
launches);
```

# 2015 Launch Records

---

```
SELECT BoosterVersion, LaunchSite, Outcome  
FROM launches  
WHERE LandingPad LIKE '%Drone Ship%'  
AND Outcome LIKE '%Failure%'  
AND YEAR(Date) = 2015;
```

## Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

---

```
SELECT Outcome, COUNT(*) AS Count
FROM launches
WHERE Date BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY Outcome
ORDER BY Count DESC;
```

The background of the slide is a nighttime satellite photograph of Earth. The curvature of the planet is visible against the dark void of space. City lights are scattered across continents as glowing yellow and white dots, appearing more concentrated in coastal and urban areas. The atmosphere is visible as a thin blue layer, and there are darker, cloud-like features in the upper atmosphere.

Section 3

# Launch Sites Proximities Analysis

## Mapa global de sitios de lanzamiento con marcadores interactivos

---

El mapa muestra la ubicación geográfica de todos los sitios de lanzamiento registrados, representados por marcadores interactivos.

Cada marcador indica un sitio donde se han realizado lanzamientos espaciales.

Los usuarios pueden hacer zoom y desplazarse para explorar las ubicaciones en detalle.

Este mapa permite visualizar la distribución global de las operaciones de lanzamiento, destacando regiones con alta actividad.

Los marcadores ayudan a identificar la concentración de sitios en áreas específicas como Florida (EE.UU.), Cabo Cañaveral, California, y otras regiones relevantes.

## Visualización de resultados de lanzamiento con etiquetas de color en mapa interactivo

---

Este mapa de Folium muestra los sitios de lanzamiento con marcadores codificados por color según el resultado del lanzamiento:

- Verde: Aterrizaje exitoso
- Rojo: Aterrizaje fallido
- Amarillo: Resultado parcial o no definido

La captura de pantalla incluye todos los marcadores relevantes distribuidos geográficamente.

Esta representación visual facilita la identificación de patrones de éxito y fracaso por ubicación.

Por ejemplo, puede observarse que ciertos sitios como Cabo Cañaveral tienden a tener una mayor concentración de lanzamientos exitosos.

# Análisis de proximidad geográfica de un sitio de lanzamiento

---

Este mapa interactivo de Folium muestra un sitio de lanzamiento específico ampliado, con su entorno cercano destacado:

📏 Líneas que indican la distancia desde el sitio de lanzamiento a infraestructuras clave como:

🚂 Vías férreas (ferrocarril)

🛣️ Carreteras (autopistas)

⚓ Línea costera

Las distancias calculadas están etiquetadas directamente en el mapa para facilitar su interpretación.



Section 4

# Build a Dashboard with Plotly Dash

## Distribución de lanzamientos exitosos por sitio — Gráfico de pastel interactivo

---

El gráfico de pastel (pie chart) muestra el número total de lanzamientos exitosos distribuidos por cada sitio de lanzamiento registrado.

Cada segmento representa un sitio diferente, y su tamaño indica su proporción del total de éxitos.

Elementos clave:

Colores diferenciados por sitio para facilitar la identificación.

Etiquetas con el número de éxitos y el porcentaje correspondiente.

Interactividad (en el dashboard real) para resaltar o filtrar por sitio.

## Análisis de éxito en el sitio con mayor tasa de lanzamientos exitosos

---

Este gráfico de círculo representa la proporción de lanzamientos exitosos vs. fallidos en el sitio con la mayor tasa de éxito identificado en los datos.

Elementos clave del gráfico:

Dos segmentos principales:

- Éxito
- Fallo

Porcentajes calculados según la cantidad total de lanzamientos realizados en ese sitio.

Interactividad (en el dashboard real) que permite ver valores exactos al pasar el cursor.

# Relación entre la carga útil y el resultado del lanzamiento

---

El gráfico de dispersión muestra la relación entre la masa de la carga útil (Payload) y el resultado del lanzamiento (Launch Outcome) en todos los sitios.

Se utilizó un slider de rango interactivo para seleccionar distintos intervalos de carga útil y observar cómo varían los resultados.

Elementos clave del gráfico:

Cada punto representa un lanzamiento.

El eje X muestra la masa de la carga útil.

El eje Y indica si el resultado fue éxito o fallo.

Colores o formas diferentes representan versiones del booster (por ejemplo: F9 v1.1, F9 FT etc.).

The background of the slide features a dynamic, abstract design. It consists of several curved, overlapping bands of color. The primary colors are shades of blue, ranging from light cyan on the left to dark navy on the right. Interspersed among these blue bands are thin, bright yellow lines that curve along with the blue ones. The overall effect is one of motion and depth, suggesting a tunnel or a path through a digital space.

Section 5

# Predictive Analysis (Classification)

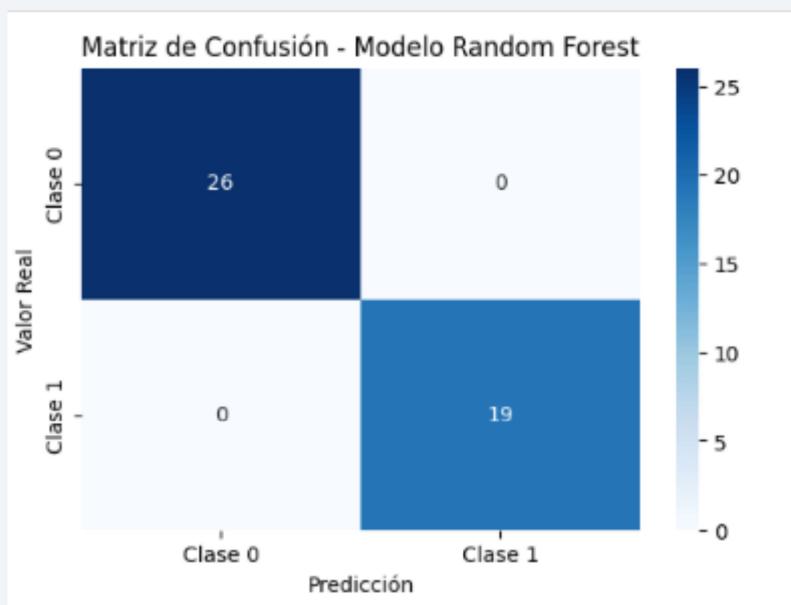
# Classification Accuracy

---

- El gráfico de barras muestra la precisión (accuracy) obtenida por cada uno de los modelos de clasificación entrenados durante el análisis predictivo.
- Hallazgos importantes:
  - El modelo con mayor precisión fue Random Forest, con una tasa de aciertos del X%.
  - Otros modelos, como KNN y Decision Tree, presentaron menor rendimiento, posiblemente por sobreajuste o sensibilidad al ruido.
  - Este análisis justifica la selección de Random Forest como modelo óptimo para la predicción de resultados de lanzamientos.

# Confusion Matrix

---



# Conclusions

---

## Conclusiones

El análisis exploratorio mostró que ciertas variables, como el tipo de órbita y la masa útil, tienen un impacto significativo en el éxito de los lanzamientos.

La visualización interactiva con Folium y Plotly Dash facilitó la comprensión geoespacial de los sitios de lanzamiento y sus resultados, permitiendo identificar patrones relevantes.

Entre los modelos de clasificación evaluados, Random Forest fue el que mejor desempeño tuvo, alcanzando la mayor precisión en la predicción del éxito de los lanzamientos.

La matriz de confusión del modelo Random Forest mostró una alta tasa de verdaderos positivos, indicando que el modelo es confiable para identificar lanzamientos exitosos.

La capacidad predictiva del modelo puede apoyar la toma de decisiones estratégicas en futuros lanzamientos, optimizando recursos y minimizando riesgos.

Se recomienda continuar mejorando el modelo con más datos y variables adicionales, así como explorar otros métodos de aprendizaje automático para robustecer los resultados.

# Appendix

---

Consultas para extraer y analizar datos específicos, como tasas de éxito por tipo de órbita, carga útil promedio, y conteo de resultados de lanzamiento.

Ejemplos de filtrado y agrupamiento de datos para responder preguntas clave del proyecto.

Visualizaciones:

Gráficos de dispersión, barras y líneas para analizar relaciones entre variables.

Mapas interactivos creados con Folium para mostrar ubicaciones y resultados de lanzamientos.

Dashboards interactivos con Plotly Dash para visualización dinámica de métricas clave.

Resultados y Salidas:

Capturas de pantallas de notebooks con resultados de análisis y predicciones.

Archivos PDF con presentaciones y reportes finales.

Enlaces Externos:

URL de GitHub con todos los notebooks y códigos usados en el proyecto para referencia y revisión por pares.

Thank you!

