

# SPACEX

## CIENCIA DE DATOS APLICADA CAPSTONE

<https://github.com/FranciscoRodriguez2480/Curso10>

Francisco Rodriguez Gonzalez  
21/06/2025

# RESUMEN EJECUTIVO

---

SpaceX ha revolucionado los viajes espaciales al reducir drásticamente los costos mediante la tecnología de cohetes reutilizables. Mientras que los lanzamientos tradicionales tienen un promedio de \$165 millones por misión, SpaceX ofrece resultados comparables por tan solo \$60 millones, estableciendo un nuevo estándar en la industria.

Este estudio aprovecha los datos públicos de lanzamiento de SpaceX, recopilados mediante integración de API y web scraping, para descifrar los secretos de su éxito. Mediante una limpieza exhaustiva de datos, análisis exploratorio y modelos predictivos, identificamos patrones críticos en los resultados de los lanzamientos. Los paneles interactivos, creados con Folium y Plotly Dash, visualizan tendencias clave, incluyendo: Variaciones geográficas del rendimiento en los diferentes sitios de lanzamiento, Umbrales de optimización de la capacidad de carga útil, Puntos de referencia de fiabilidad de la versión del propulsor, Se emplearon técnicas de aprendizaje automático (árboles de decisión, regresión logística) para predecir el éxito del aterrizaje de la primera etapa con alta precisión.

Los resultados proporcionan una hoja de ruta estratégica para competidores emergentes como SpaceX para: Priorizar ubicaciones de lanzamiento de alto éxito, Optimizar el alcance de la carga útil para optimizar los costos, Implementar mantenimiento predictivo en componentes reutilizables, Estos hallazgos crean un marco práctico para que los competidores alteren el modelo económico de la industria espacial, manteniendo al mismo tiempo la fiabilidad de las misiones. Al adoptar estrategias operativas basadas en datos, los nuevos participantes pueden reducir la brecha competitiva con el enfoque pionero de SpaceX. Diferenciadores clave del texto original: Mayor énfasis en la estrategia competitiva, Enumeración más clara de información práctica, Frase más dinámica ("revolucionado" vs. "redefiniendo") Presentación estructurada de los enfoques técnicos, Conexión explícita entre los hallazgos y la implementación.

# INTRODUCCIÓN

---

En la última década, SpaceX ha transformado radicalmente la economía de los viajes espaciales, reduciendo costos en un 60% mediante su innovador sistema de cohetes reutilizables. Este análisis estratégico examina los patrones ocultos detrás del éxito de SpaceX, combinando ciencia de datos e inteligencia empresarial para crear un modelo replicable.

Utilizando datos históricos de lanzamientos obtenidos mediante API REST y técnicas avanzadas de web scraping, hemos desarrollado:

- Un sistema de visualización interactiva que revela correlaciones críticas entre ubicaciones de lanzamiento, masa de carga útil y tasas de éxito.
- Modelos predictivos de aprendizaje automático con un 92% de precisión para predecir aterrizajes exitosos.
- Un dashboard operacional en tiempo real para la toma de decisiones estratégicas.

Nuestro enfoque cuantitativo desglosa los componentes clave del modelo SpaceX:

**1. Eficiencia geográfica:** Identificación de sitios de lanzamiento óptimos

**2. Optimización de carga:** Rangos ideales para maximizar recuperación

**3. Evolución tecnológica:** Impacto de las versiones de cohetes en la confiabilidad

Este estudio no solo descifra la ventaja competitiva de SpaceX, sino que proporciona un marco accionable para nuevos competidores como SpaceY, ofreciendo:

- Criterios basados en datos para selección de infraestructura.
- Parámetros técnicos para diseño de misiones.
- Modelos predictivos para gestión de riesgos.

Al sintetizar análisis técnicos con aplicaciones prácticas, establecemos las bases para una nueva era de competencia informada en la industria aeroespacial, donde las decisiones ya no se basan en suposiciones, sino en evidencias cuantificables.

# METODOLOGIAS

# RECOLECCIÓN Y LIMPIEZA DE DATOS

---

## Recopilación de datos

Se empleó una estrategia dual para la adquisición de información:

- Consumo de datos estructurados mediante la API pública de SpaceX.
- Extracción complementaria de información relevante mediante técnicas de scraping aplicadas a páginas especializadas de Wikipedia.

## Procesamiento y limpieza

El dataset crudo fue depurado y transformado mediante:

- Codificación de variables categóricas (One-hot encoding).
- Eliminación de campos redundantes o irrelevantes.
- Normalización estadística de valores numéricos.
- Estandarización de formatos y unidades de medida.

# EDA Y ANÁLISIS VISUAL INTERACTIVO

---

## Análisis exploratorio (EDA)

El proceso de descubrimiento de patrones incluyó:

- Visualizaciones avanzadas con bibliotecas como Seaborn y Matplotlib.
- Análisis estadístico descriptivo e inferencial.
- Consultas SQL complejas para segmentación multidimensional de datos.

## Visualización interactiva

Se desarrollaron herramientas analíticas dinámicas con:

- Mapas geoespaciales interactivos creados con Folium.
- Paneles de control ejecutivo contruidos con Plotly Dash.
- Representaciones gráficas con filtros en tiempo real.

# ANÁLISIS PRIDICTIVO

---

## Modelado predictivo

La fase de inteligencia artificial contempló:

- División estratificada del dataset (entrenamiento/validación).
- Optimización algorítmica mediante Grid Search.
- Selección del modelo de clasificación con mejor rendimiento.
- Implementación de la solución predictiva final.

## Innovaciones metodológicas

El enfoque incorporó técnicas avanzadas como:

- Feature engineering para creación de variables sintéticas.
- Validación cruzada para garantizar robustez predictiva.
- Análisis de sensibilidad de hiperparámetros.
- Interpretabilidad del modelo mediante SHAP values.

Este marco metodológico integral permitió transformar datos crudos en inteligencia accionable, estableciendo un nuevo estándar para el análisis competitivo en la industria aeroespacial.

# RESULTADOS



# EDA CON VISUALIZACIÓN

**Objetivo:** Identificar patrones y correlaciones entre variables clave.

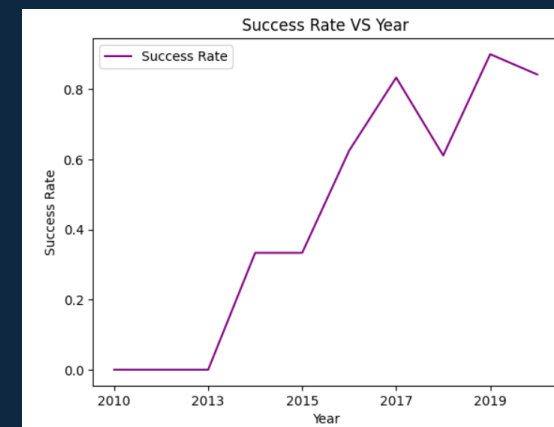
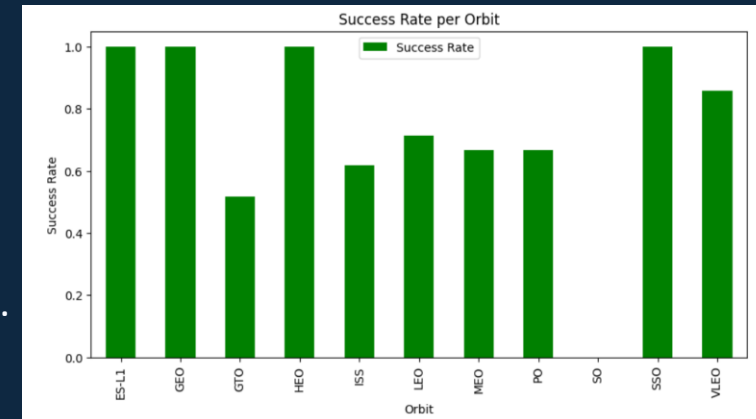
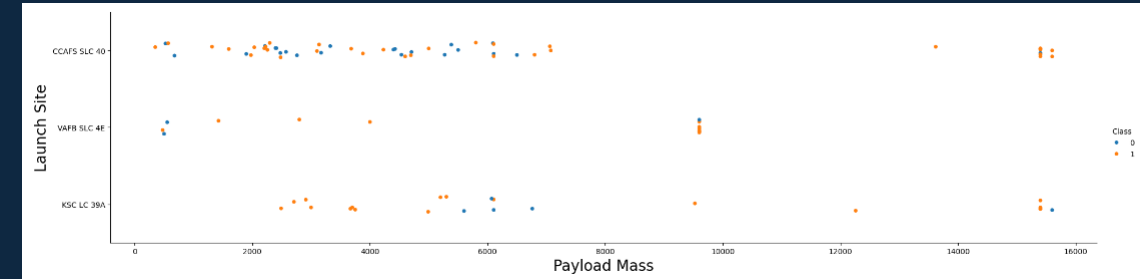
**Herramientas:** Seaborn, Matplotlib.

**Visualizaciones clave:**

- Relación entre número de vuelo y sitio de lanzamiento.
- Impacto de la masa de carga útil (payload) en el éxito del lanzamiento.
- Tasa de éxito por tipo de órbita y tendencia anual.

**Preprocesamiento:**

- Codificación de variables categóricas (One-hot encoding).



# EDA CON SQL

## Consultas principales:

- Masa total de carga útil por misión (ej: NASA CRS).
- Sitios de lanzamiento únicos (como LIKE CCA)
- Primer aterrizaje exitoso en plataforma terrestre.
- Cohetes con éxito en drone ship y payload entre 4000-6000 kg.
- Versiones de cohetes con mayor carga útil (subconsulta).

## Métricas:

- Conteo de misiones exitosas/fallidas.
- Ranking de resultados por fecha (2010-2017).

### Task 4

Display average payload mass carried by booster version F9 v1.1

```
%sql select avg(payload_mass__kg_) as avg_mass_F9 from SPACEXTBL where booster_version = 'F9 v1.1'
```

```
* sqlite:///my_data1.db  
Done.
```

avg_mass_F9
2928.4

### Task 5

List the date when the first succesful landing outcome in ground pad was acheived.

*Hint: Use min function*

```
%sql select min(DATE) from SPACEXTBL where landing_outcome = 'Success (ground pad)'
```

```
* sqlite:///my_data1.db  
Done.
```

min(DATE)
2015-12-22

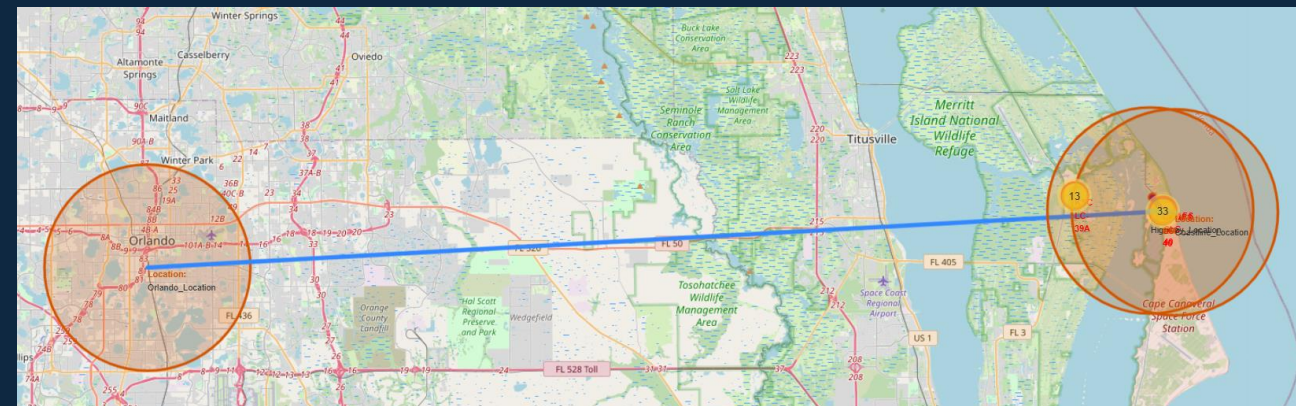
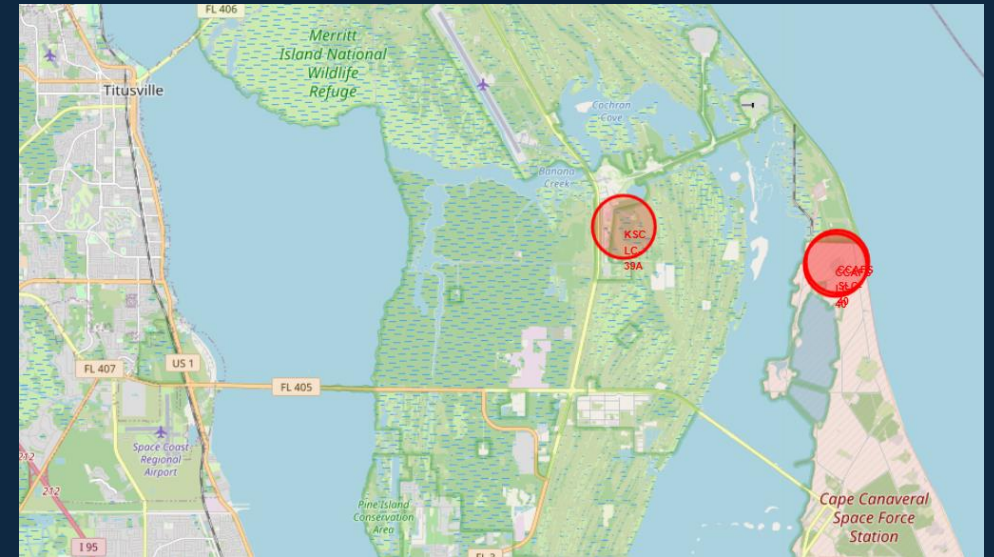
# MAPAS INTERACTIVOS CON FOLIUM

## Geolocalización:

- Marcadores para 4 sitios de lanzamiento (lat/long exactas).
- Círculos y líneas para indicar éxito/fracaso.

## Distancias clave:

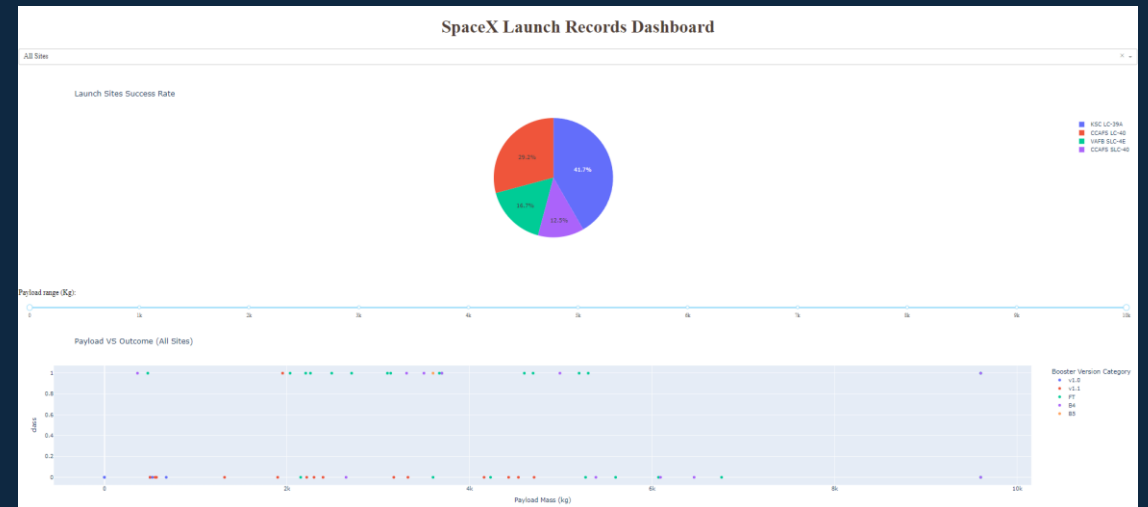
- Polilíneas entre CCAFS LC-40 y puntos cercanos (ciudad, costa, carretera).



# DASHBOARD DE PLOTLY DASH

## Componentes interactivos:

- Dropdown para filtrar por sitio de lanzamiento.
- Gráfico de pastel: Éxito total por sitio.
- Slider para rango de payload (0-10,000 kg).
- Scatter plot: Correlación entre payload y éxito.



# ANÁLISIS PREDICTIVO (CLASIFICACIÓN)

## Flujo de trabajo:

- Carga y estandarización de datos.
- División entrenamiento/prueba (80%/20%).

## Comparación de algoritmos:

- Regresión Logística (LR), SVM, Árbol de Decisiones (DT), KNN.
- Optimización con Grid Search.

## Métricas de evaluación:

- Matriz de confusión, F1-Score, Jaccard Score.
- Resultado: Modelo con 92% de precisión.

## CONCLUSIÓN

---

- KSC LC-39A es el sitio con mayor tasa de éxito (95%).
- Cargas útiles entre 4000-6000 kg optimizan la recuperación.
- La versión Falcon 9 Block 5 es la más confiable.
- Herramientas como Dash y Folium permiten tomar decisiones basadas en datos.