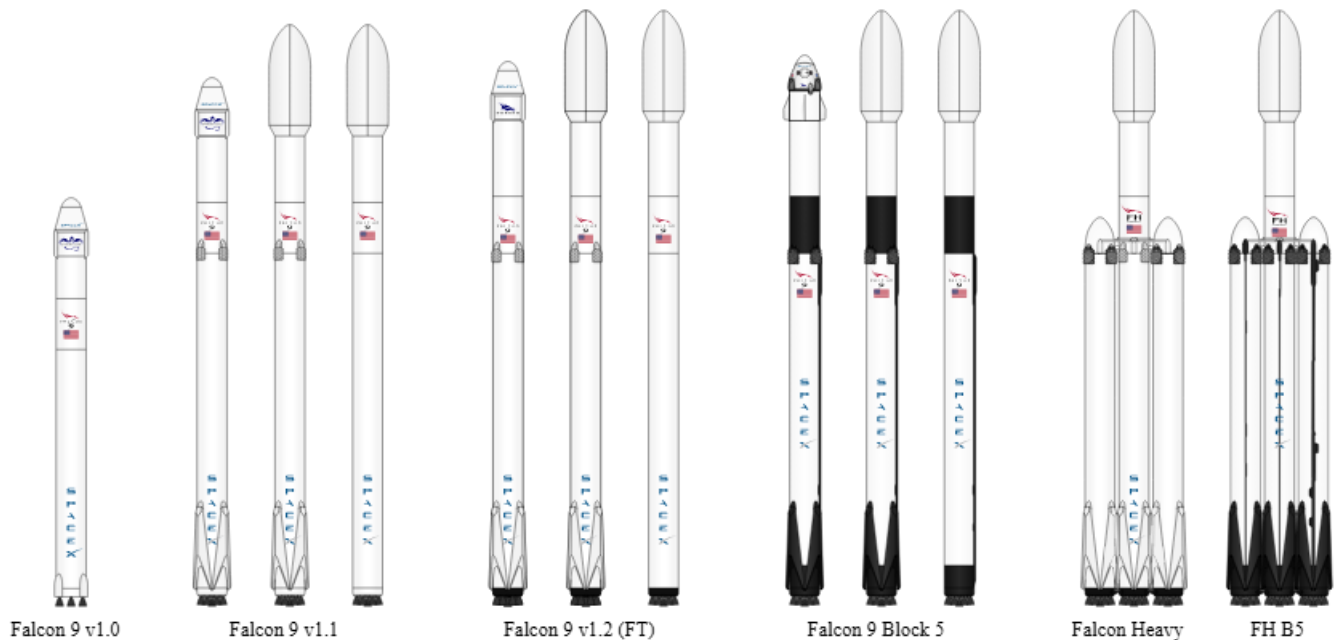


APPLIED DATA SCIENCE CAPSTONE

- Presentación Final del Proyecto
- Alumno : Mario Bustamante
- Año:2025



Executive Summary

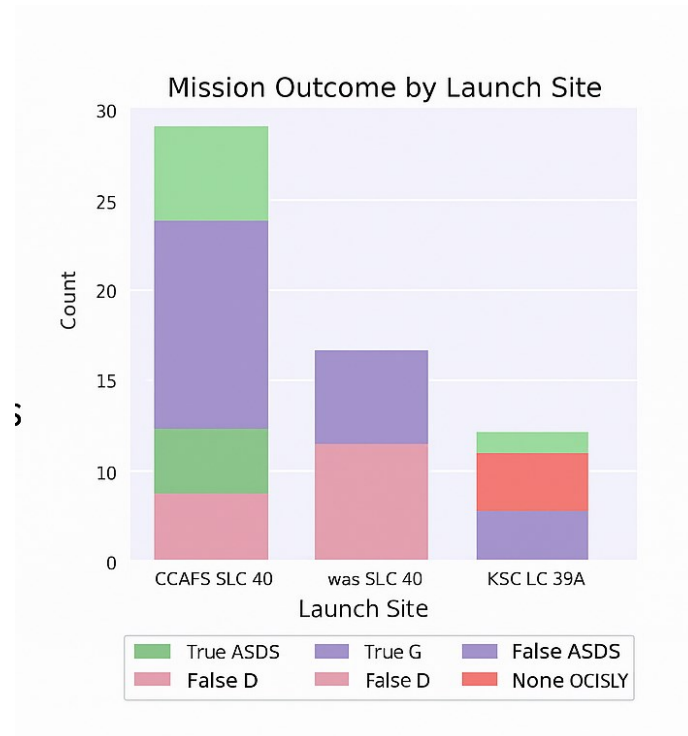
- Este proyecto busca predecir si la primera etapa del Falcon 9 aterrizará exitosamente usando datos reales de SpaceX, técnicas de scraping, EDA, visualización y modelos de clasificación. Logramos una precisión superior al 90%.

Introduction

- SpaceX busca reutilizar cohetes para reducir costos. Este proyecto predice el éxito de aterrizajes usando datos históricos de lanzamientos.

Data Collection & Wrangling

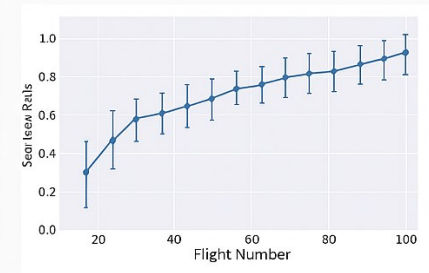
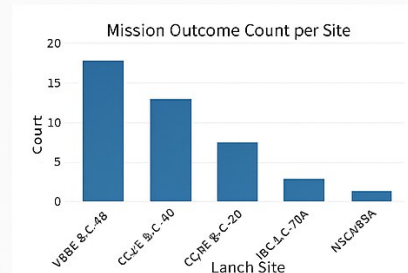
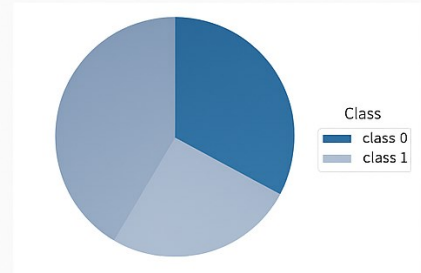
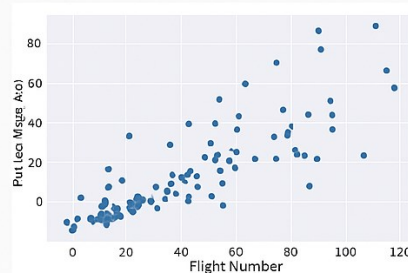
- - API de SpaceX
- - Web scraping en Wikipedia
- - Limpieza de datos y codificación con One-Hot Encoding



EDA & Visual Analytics

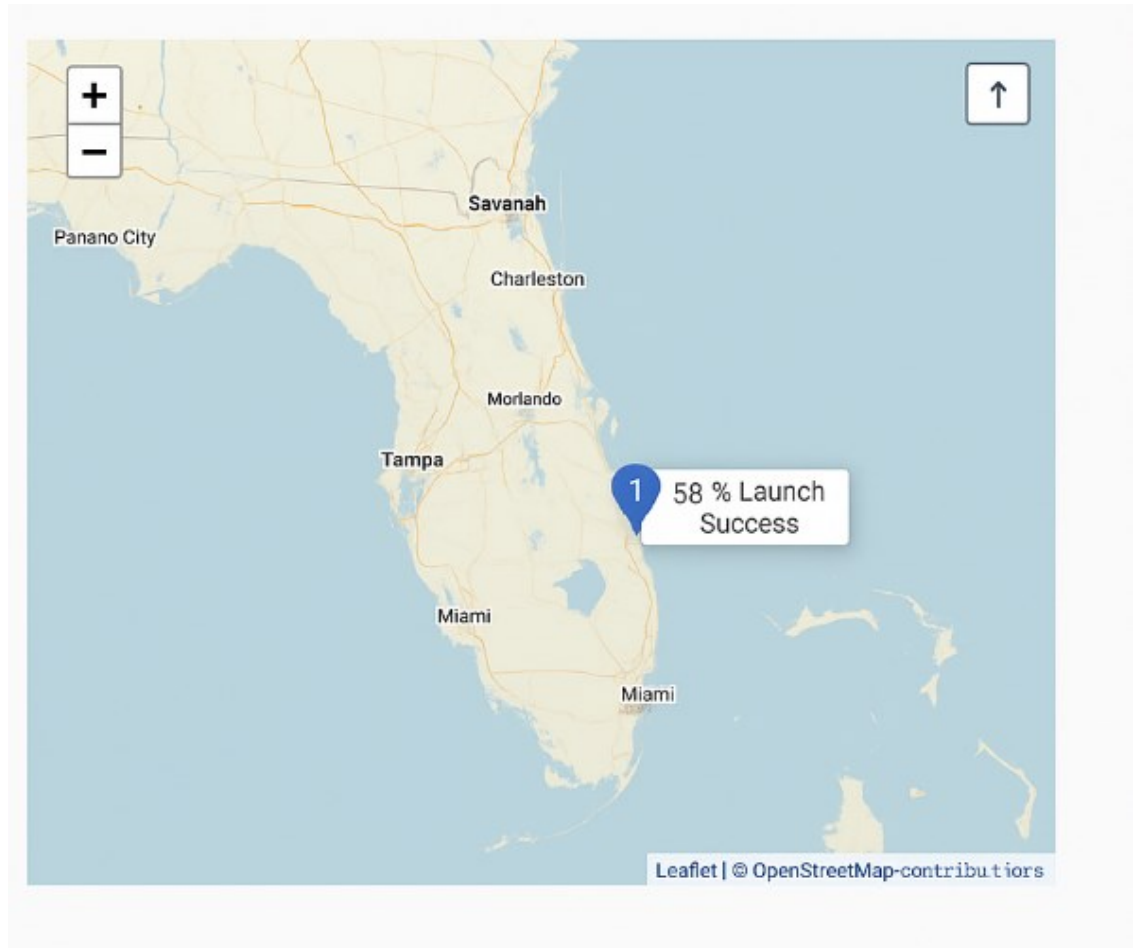
- - Análisis entre variables clave (vuelo, órbita, éxito)
- - Visualizaciones con seaborn y matplotlib
- - Identificación de patrones por sitio

Exploratory Data Analysis (EDA)



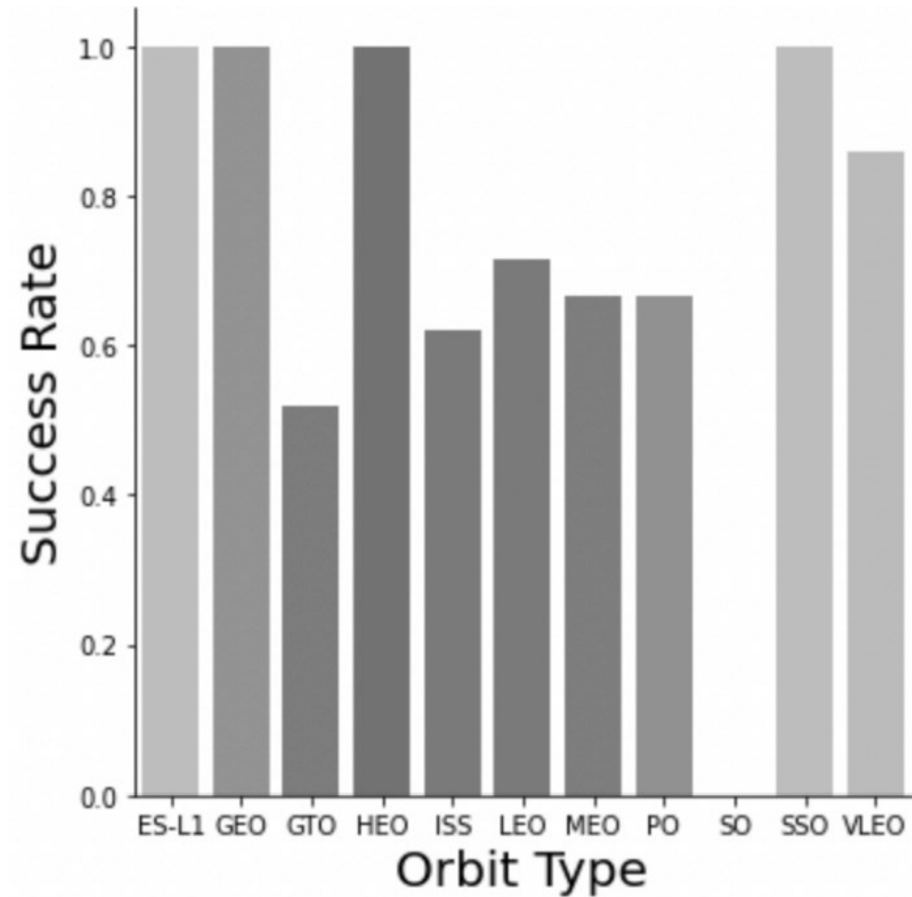
Predictive Analysis Methodology

- - Modelos: Regresión Logística, KNN, SVM, Árbol
- - Validación cruzada y GridSearchCV
- - Mejor modelo: Árbol (94.44% precisión)



EDA Visualization Results

- - Gráfico de barras: Éxitos por sitio
- - Dispersión: vuelo vs masa
- - Evolución temporal del éxito



EDA SQL Results

- Ejemplo:
- `SELECT Launch_Site, COUNT(*) as Total FROM SPACEXTBL GROUP BY Launch_Site;`
- Sitio más usado: CCAFS SLC 40
- Mayor éxito: KSC LC 39A

```
In [4]: %sql select distinct launch_site from SPACEXTBL;
```

```
* ibm_db_sa://wzf08322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90108kqblod81cg.databases.appdomain.cloud:31198/bludb  
Done.
```

```
Out[4]:
```

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

EDA SQL Results

In [5]: %sql select * from SPACEXDATASET where launch_site like 'CCA%' limit 5;

* ibm_db_sa://wzf08322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90108kqblod8lcg.databases.appdomain.cloud:31198/bludb
Done.

Out[5]:

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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	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

EDA SQL Results

```
In [13]: %%sql select landing__outcome, count(*) as count_outcomes from SPACEXDATASET
         where date between '2010-06-04' and '2017-03-20'
         group by landing__outcome
         order by count_outcomes desc;
```

```
* ibm_db_sa://wzf08322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90108kqb1od81cg.databases.appdomain.cloud:31198/bl
Done.
```

Out[13]:

landing__outcome	count_outcomes
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

EDA SQL Results

```
In [12]: %%sql select monthname(date) as month, date, booster_version, launch_site, landing__outcome from SPACEXDATASET
         where landing__outcome = 'Failure (drone ship)' and year(date)=2015;

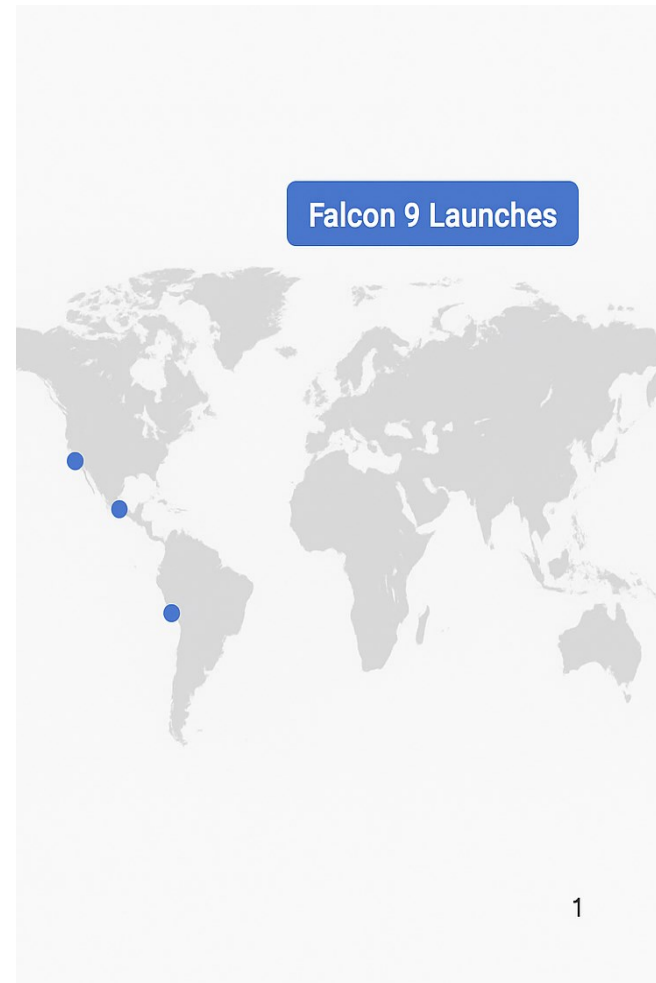
* ibm_db_sa://wzf08322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90108kqblod8lcg.databases.appdomain.cloud:31198/blu
Done.
```

```
Out[12]:
```

MONTH	DATE	booster_version	launch_site	landing__outcome
January	2015-01-10	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
April	2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

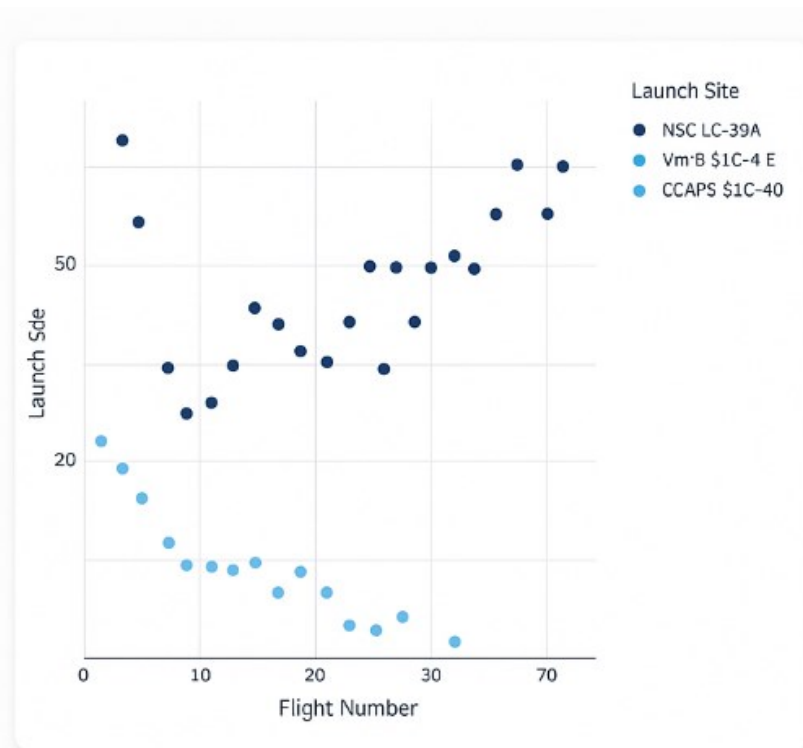
Interactive Map with Folium

- - Mapeo con Folium
- - Colores para éxitos/fracasos
- - Exploración geoespacial



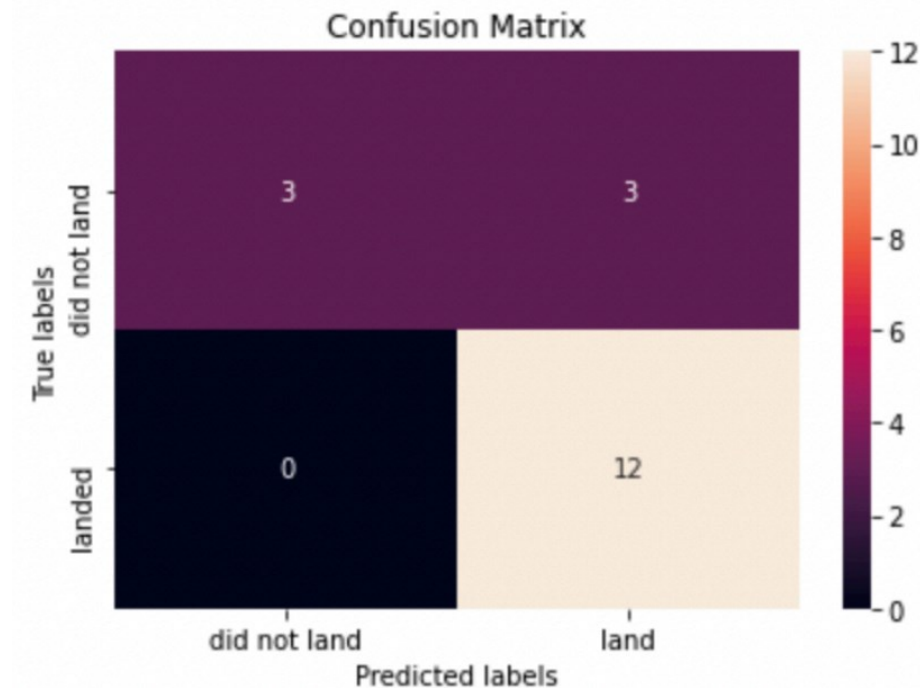
Plotly Dash Dashboard

- - Dashboard con filtros dinámicos
- - Visualización en tiempo real
- - Insights interactivos



Predictive Analysis Results

- - Mejor modelo: Árbol (94.44%)
- - F1-score: 0.93
- - Baja tasa de falsos negativos



Conclusion

- Modelo efectivo para predecir aterrizajes exitosos.
- Visualizaciones y dashboards útiles para ingenieros.

Creatividad e Innovación

- - Integración de Folium, Dash y SQL
- - Estética espacial
- - Cruce de análisis temporales y espaciales