



IBM Developer  
SKILLS NETWORK

# Winning Space Race with Data Science

Alberto Lopez Pereira  
14/07/2024



# Índice

---

- Resumen Ejecutivo
- Introducción
- Metodología
- Resultados
- Conclusión
- Apéndice



# Resumen Ejecutivo

---

## Resumen de metodologías

- Recolección de datos
- Procesamiento de datos
- Análisis exploratorio de datos con visualización de datos
- Análisis exploratorio de datos con SQL
- Construcción de un mapa interactivo con Folium
- Construcción de un panel de control con Plotly Dash
- Análisis predictivo (Clasificación)

## Resumen de todos los resultados

- Resultados del análisis exploratorio de datos
- Demostración de análisis interactivo en capturas de pantalla
- Resultados del análisis predictivo

# Introducción

---

## Antecedentes y contexto del proyecto

SpaceX se ha consolidado como la empresa más exitosa de la era espacial comercial, haciendo asequibles los viajes espaciales. La compañía indica en su sitio web que los lanzamientos del cohete Falcon 9 tienen un costo de 62 millones de dólares, mientras que otros proveedores cobran más de 165 millones de dólares por lanzamiento. Gran parte de este ahorro se debe a la capacidad de SpaceX de reutilizar la primera etapa del cohete. Por consiguiente, si podemos determinar si la primera etapa aterrizará, podremos estimar el costo de un lanzamiento. Basándonos en información pública y modelos de aprendizaje automático, nos proponemos predecir si SpaceX reutilizará la primera etapa.

## Preguntas a responder:

1. ¿Cómo afectan variables como la masa de la carga útil, el sitio de lanzamiento, el número de vuelos y las órbitas al éxito del aterrizaje de la primera etapa?
2. ¿Ha aumentado la tasa de aterrizajes exitosos a lo largo de los años?
3. ¿Cuál es el mejor algoritmo que se puede utilizar para la clasificación binaria en este caso?



Section 1

# Methodology

# Methodología

---

## Metodología de recolección de datos:

- Utilización de la API REST de SpaceX
- Empleo de técnicas de web scraping en Wikipedia

## Realización de data wrangling

- Filtrado de datos
- Tratamiento de valores faltantes
- Aplicación de codificación one-hot para preparar los datos para una clasificación binaria
- Realización del análisis exploratorio de datos (EDA)
  - Mediante visualización
  - Mediante SQL
- Desarrollo de análisis visual interactivo usando Folium y Plotly Dash
- Realización de análisis predictivo mediante modelos de clasificación

# Recolección de datos

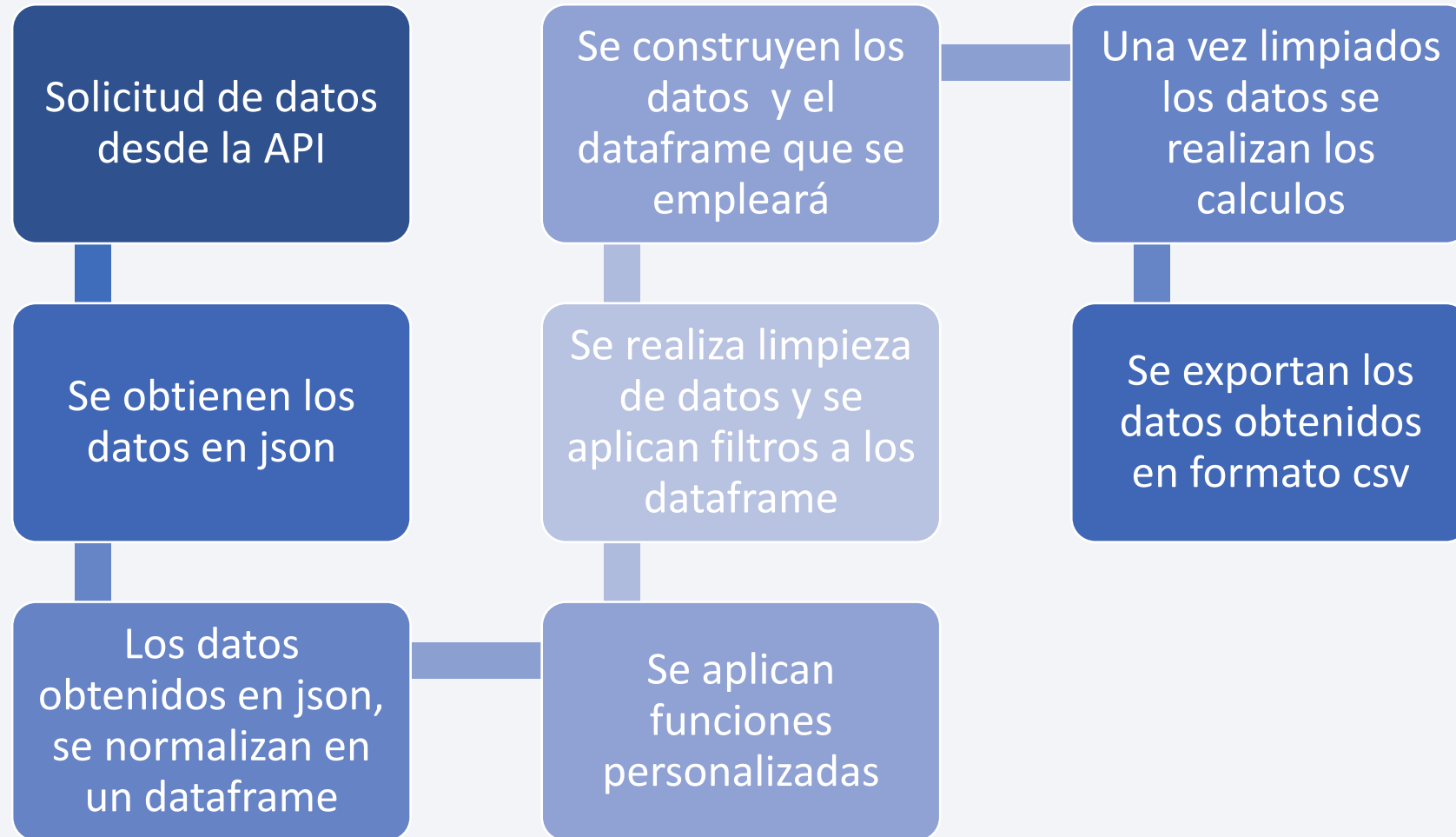
---

La recolección de los datos se realizó desde dos fuentes de datos:

1. API REST desde la URL [api.spacexdata.com](https://api.spacexdata.com) en donde se obtuvo la Información de los lanzamientos. La Información obtenida de cada lanzamiento incluye los campos FlightNumber, Date, BoosterVersion, PayloadMass, Orbit, LaunchSite Outcome, Flights, GridFins, Reused, Legs, LandingPad, Block, ReusedCount, Serial, Longitude y Latitude
2. Web Scraping desde Wikipedia en la URL [https://en.wikipedia.org/w/index.php?title=List\\_of\\_Falcon\\_9\\_and\\_Falcon\\_Heavy\\_launches&oldid=1027686922](https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922) . De este Web Scraping

# Recolección de datos– SpaceX API

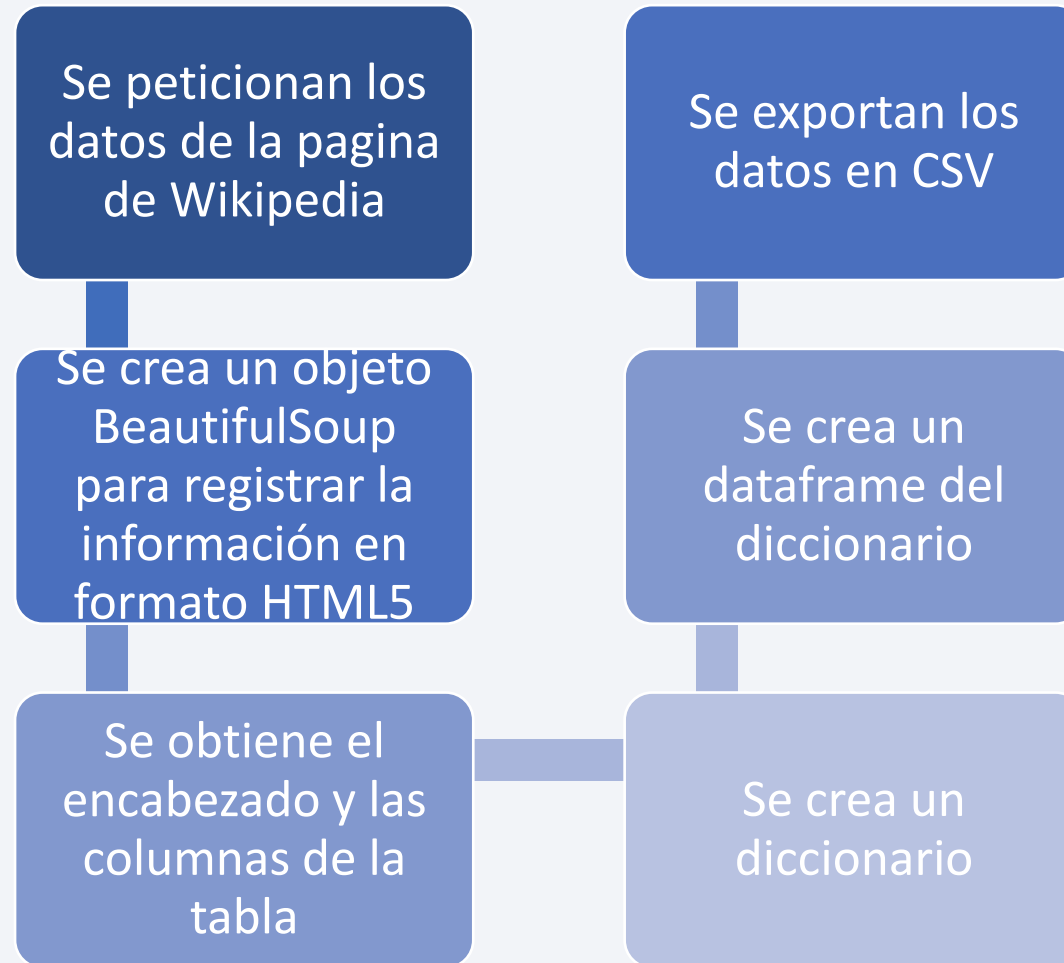
---





# Recolección de datos- Scraping

---



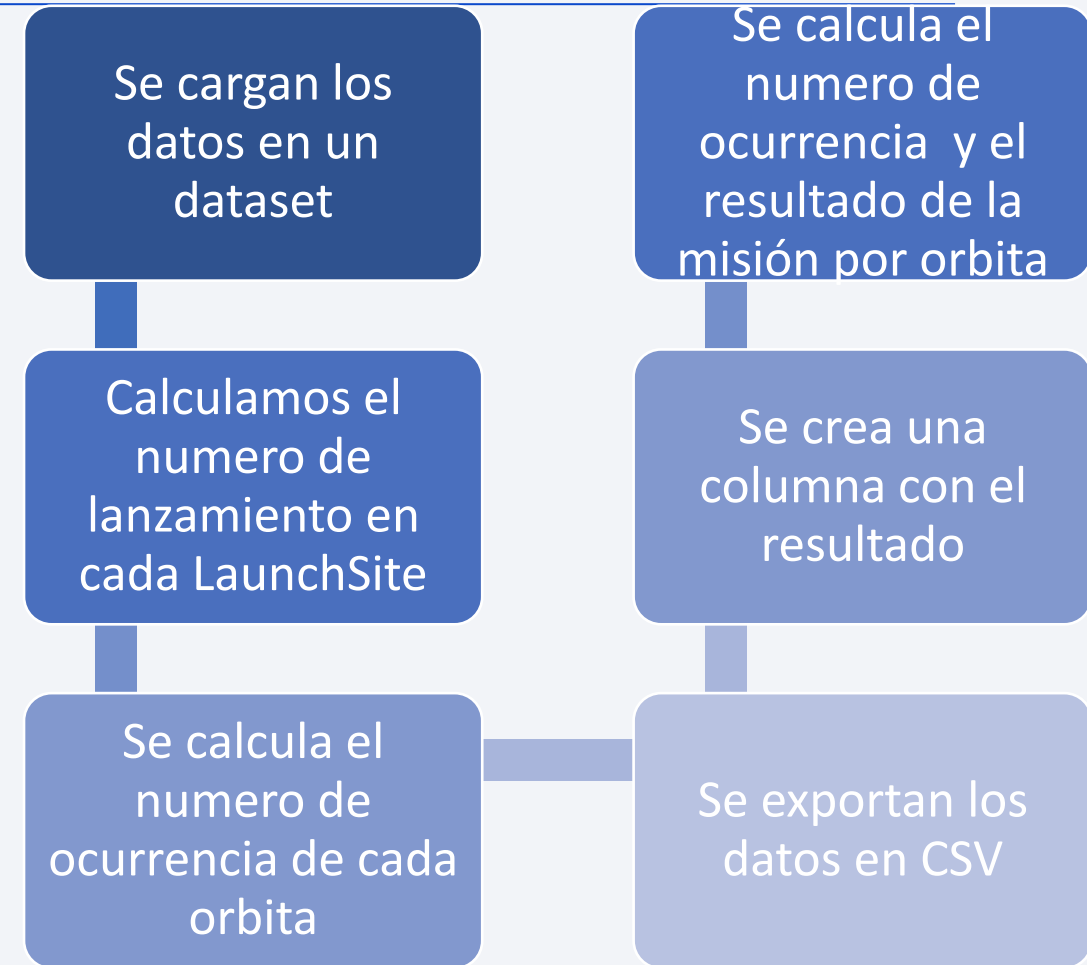
# Data Wrangling

El dataset abarca diversos escenarios de aterrizaje del propulsor, incluyendo éxitos y fracasos en diferentes ubicaciones. Para realizar un análisis exploratorio de estos datos, se realizaron las siguientes tareas:

- Cuantificar los lanzamientos por sitio.
- Determinar la frecuencia y cantidad de cada tipo de órbita.
- Examinar los resultados de las misiones según el tipo de órbita.
- Generar una nueva columna de etiquetas basada en los resultados de aterrizaje.

El objetivo final fue crear etiquetas binarias para entrenamiento: "1" para aterrizajes exitosos y "0" para los fallidos, simplificando así los datos para su uso en modelos predictivos.

Este enfoque permite una comprensión profunda de los patrones de éxito en los aterrizajes y prepara los datos para futuros análisis o modelados de machine learning.

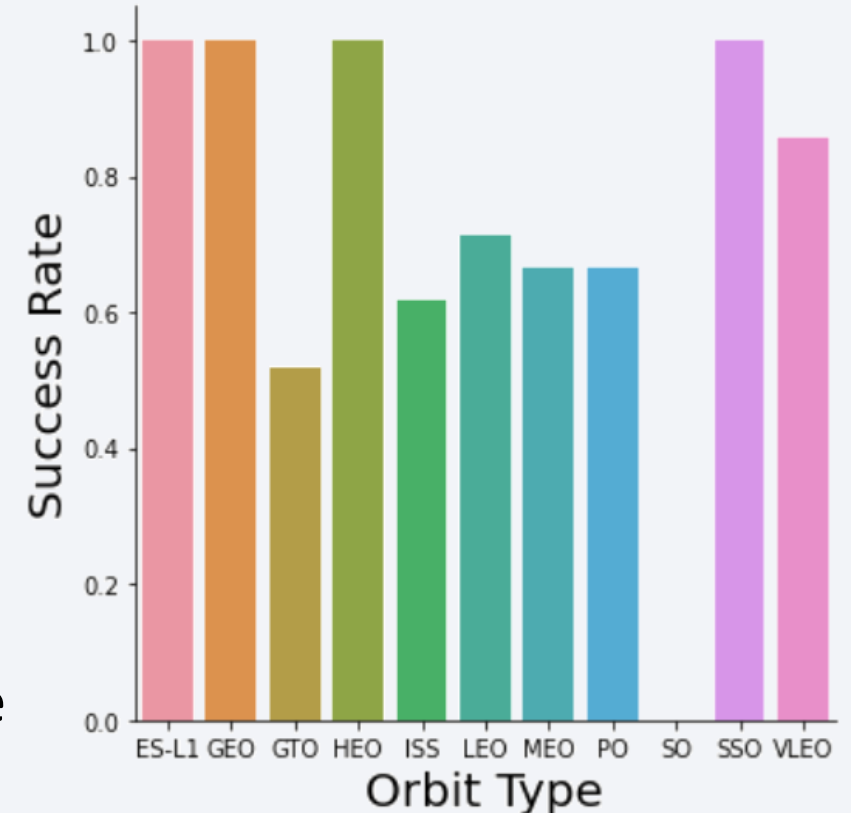


# EDA with Data Visualization

---

En este punto se realizó el EDA y la Visualización de los datos que permitió una preparación de los datos para el análisis. Se realizaron gráficos que representaron:

- Relación entre FlightNumber vs. PayloadMass
- Identificación de la relación entre Flight Number y Launch Site
- Visualización de la relación entre FlightNumber y Orbit type
- Visualización de la relación entre Payload and Orbit type
- Visualización del launch success yearly trend
- Crearon las columnas categoricas



# EDA with SQL

---

Se realizaron las siguientes consultas:

- Mostrar los nombres de los sitios de lanzamiento únicos en la misión espacial.
- Mostrar 5 registros donde los sitios de lanzamiento comienzan con la cadena 'CCA'
- Visualización de la masa total de carga útil transportada por los propulsores lanzados por la NASA (CRS)
- Visualización de la masa de carga útil promedio transportada por la versión de refuerzo F9 v1.1
- Listar la fecha en que se logró el primer aterrizaje exitoso en la plataforma de tierra
- Enumerar los nombres de los propulsores que tienen éxito en naves no tripuladas y tienen una masa de carga superior a 4000 pero inferior a 6000.
- Enumerar el número total de resultados de misiones exitosas y fallidas.
- Listar los nombres de las versiones de refuerzo que han llevado la masa máxima de carga útil
- Enumerar los resultados de los aterrizajes fallidos en naves no tripuladas, sus versiones propulsoras y los nombres de los sitios de lanzamiento para los meses del año 2015.
- Clasificar el recuento de resultados de aterrizaje (como fracaso (barco drone) o éxito (plataforma terrestre)) entre la fecha 2010-06-04 y 2017-03-20 en orden descendente

# Interactive Map with Folium

---

## **Marcadores de todos los sitios de lanzamiento:**

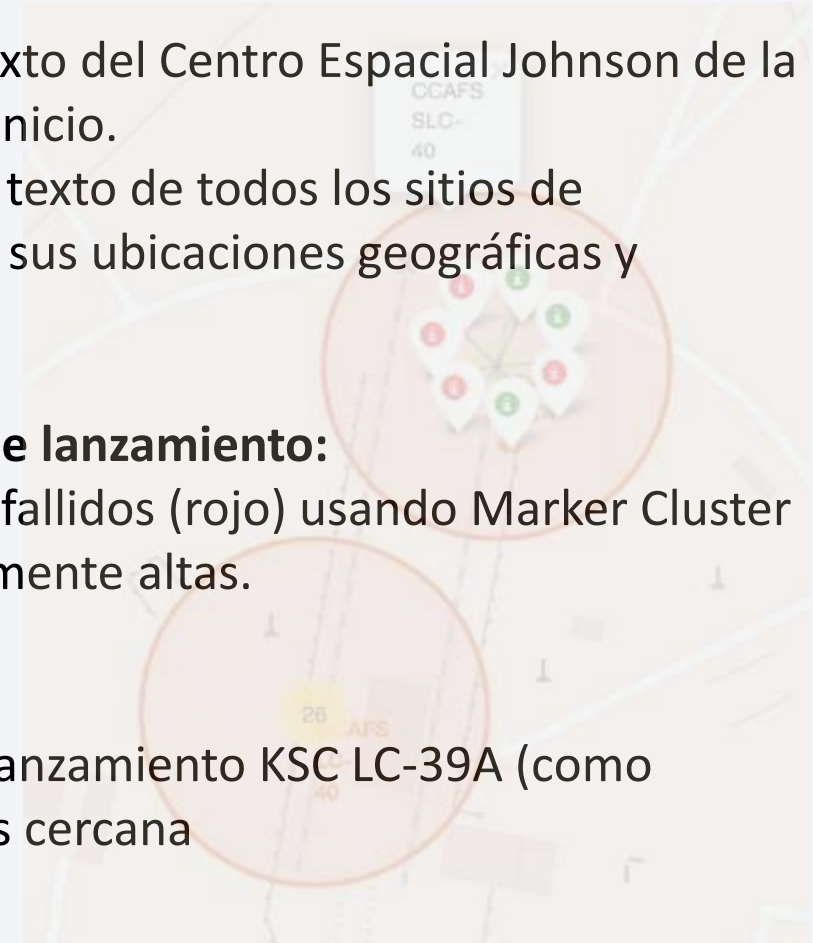
- Se agregó un marcador con círculo, etiqueta emergente y etiqueta de texto del Centro Espacial Johnson de la NASA usando sus coordenadas de latitud y longitud como ubicación de inicio.
- Se agregaron marcadores con círculo, etiqueta emergente y etiqueta de texto de todos los sitios de lanzamiento usando sus coordenadas de latitud y longitud para mostrar sus ubicaciones geográficas y proximidad al ecuador y las costas.

## **Marcadores de colores de los resultados del lanzamiento para cada sitio de lanzamiento:**

- Se agregaron marcadores de colores de lanzamientos exitosos (verde) y fallidos (rojo) usando Marker Cluster para identificar qué sitios de lanzamiento tienen tasas de éxito relativamente altas.

## **Distancias entre un Sitio de Lanzamiento y sus proximidades:**

- Se agregaron líneas de colores para mostrar distancias entre el sitio de lanzamiento KSC LC-39A (como ejemplo) y sus proximidades como ferrocarril, autopista, costa y ciudad más cercana





# Dashboard con Plotly Dash

---

## **Lista desplegable de sitios de lanzamiento:**

- Se agregó una lista desplegable para habilitar la selección del sitio de lanzamiento.

## **Gráfico circular que muestra los lanzamientos exitosos (todos los sitios/cierto sitio):**

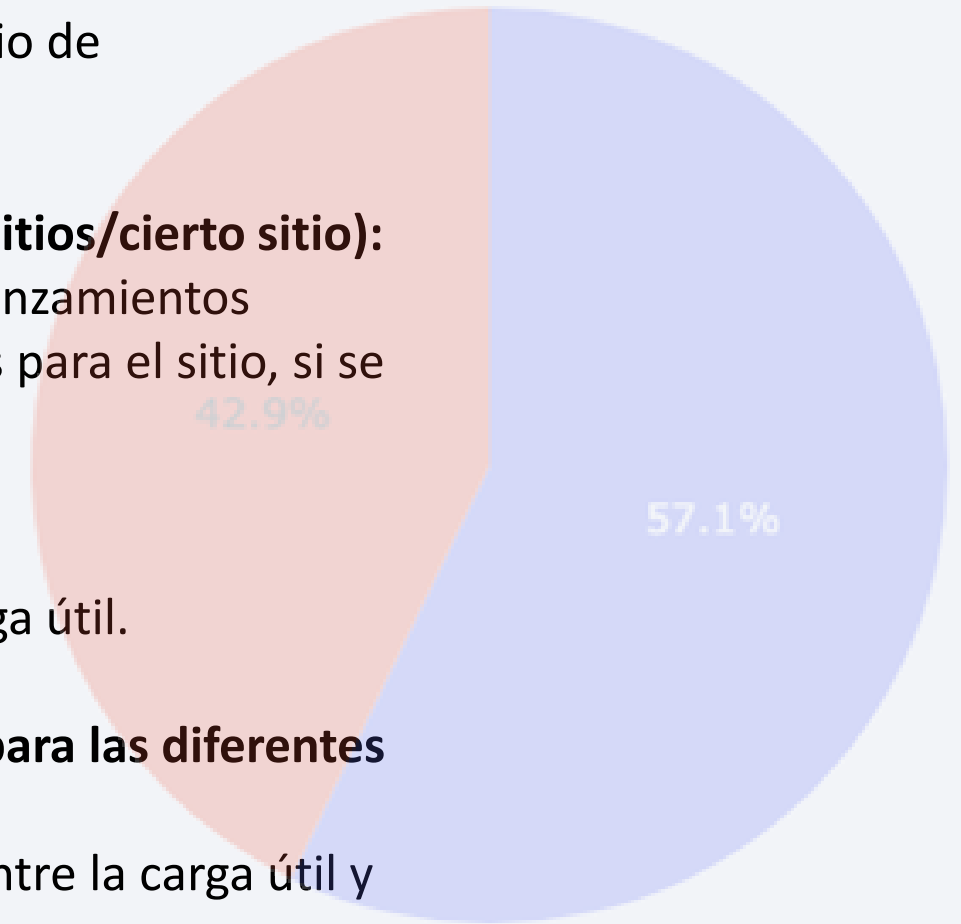
- Se agregó un gráfico circular para mostrar el recuento total de lanzamientos exitosos para todos los sitios y el recuento de éxitos versus fallas para el sitio, si se seleccionó un sitio de lanzamiento específico.

## **Control deslizante del rango de masa de carga útil:**

- Se agregó un control deslizante para seleccionar el rango de carga útil.

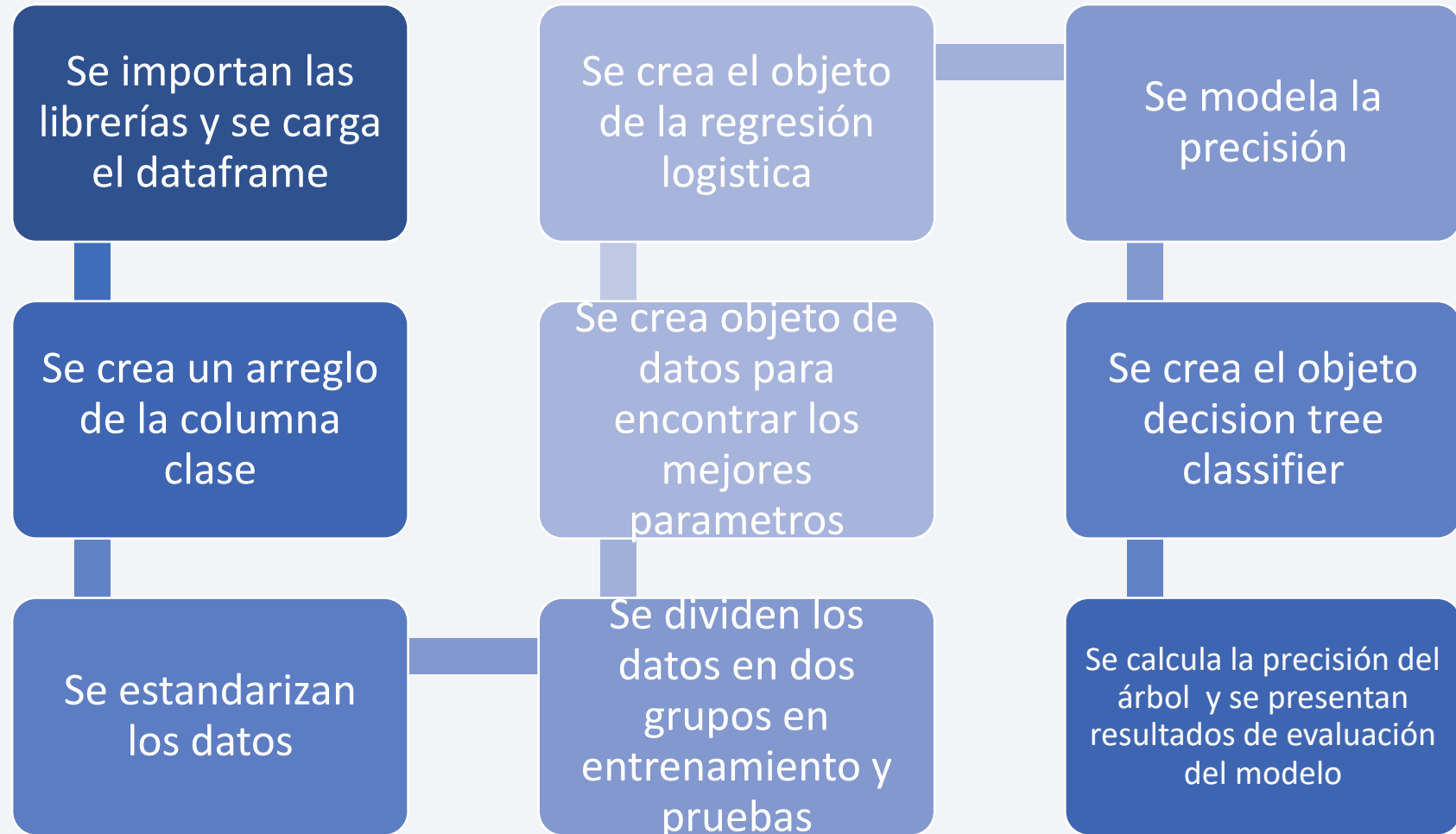
## **Gráfico de dispersión de masa de carga útil frente a tasa de éxito para las diferentes versiones de refuerzo:**

- Se agregó un gráfico de dispersión para mostrar la correlación entre la carga útil y el éxito del lanzamiento.



# Predictive Analysis (Classification)

---



# Resultados

---

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



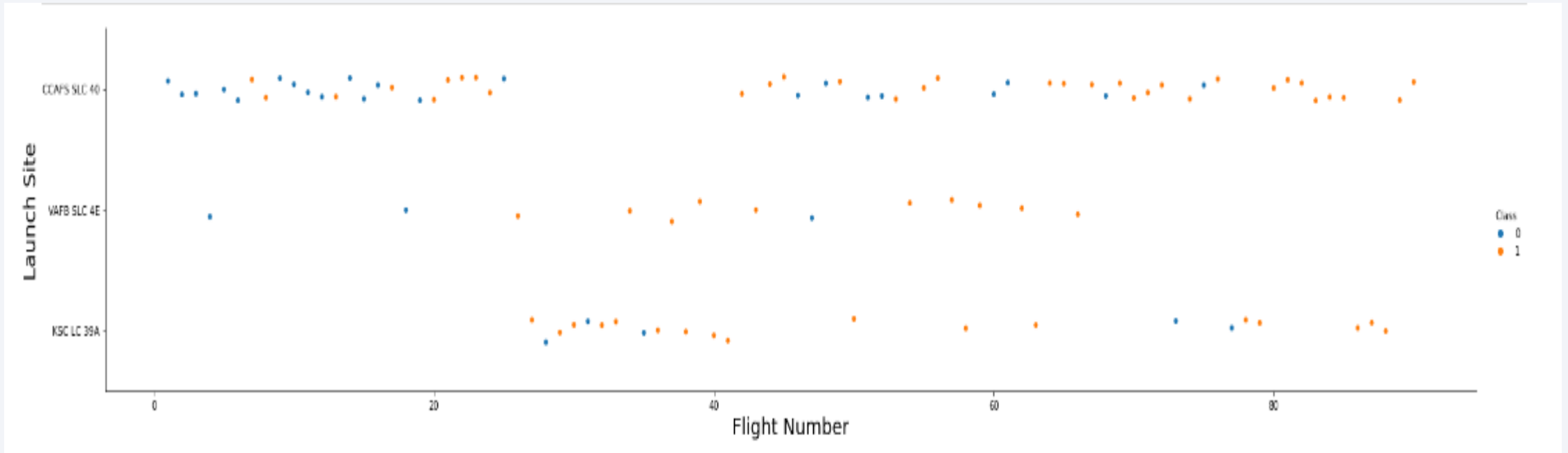
The background of the slide is an abstract composition. It features a dark blue base color. Overlaid on this are numerous diagonal streaks in shades of red and cyan. A faint, light blue grid pattern is also visible, particularly in the lower half of the image. The overall effect is dynamic and technological.

Section 2

# Insights drawn from EDA



# Flight Number vs. Launch Site

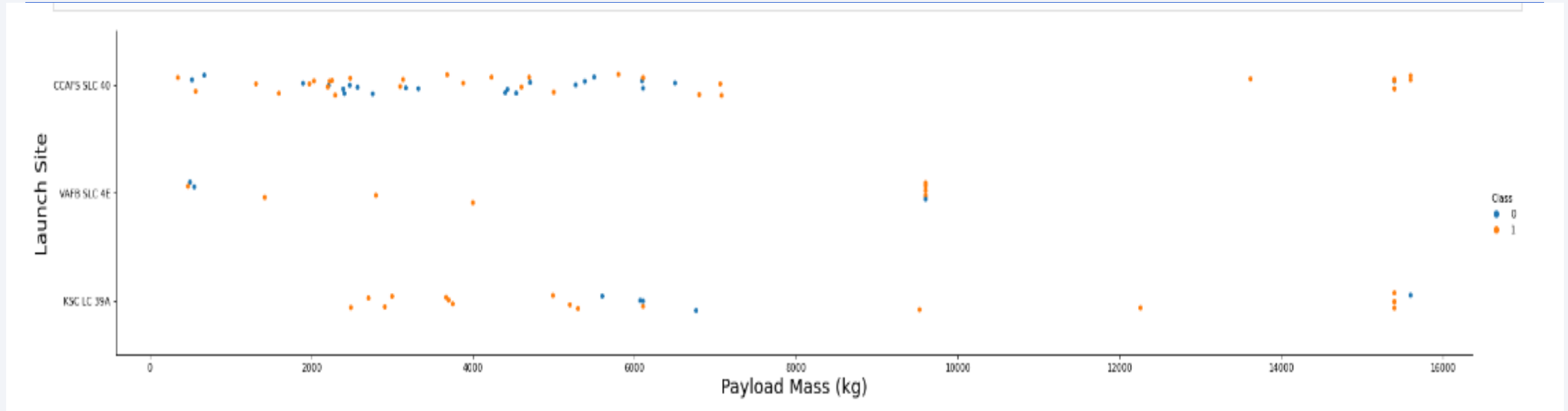


## Explicación:

- Todos los primeros vuelos fallaron, mientras que los últimos vuelos tuvieron éxito.
- El sitio de lanzamiento CCAFS SLC 40 tiene aproximadamente la mitad de todos los lanzamientos.
- VAFB SLC 4E y KSC LC 39A tienen mayores tasas de éxito.
- Se puede suponer que cada nuevo lanzamiento tiene una mayor tasa de éxito



# Payload vs. Launch Site



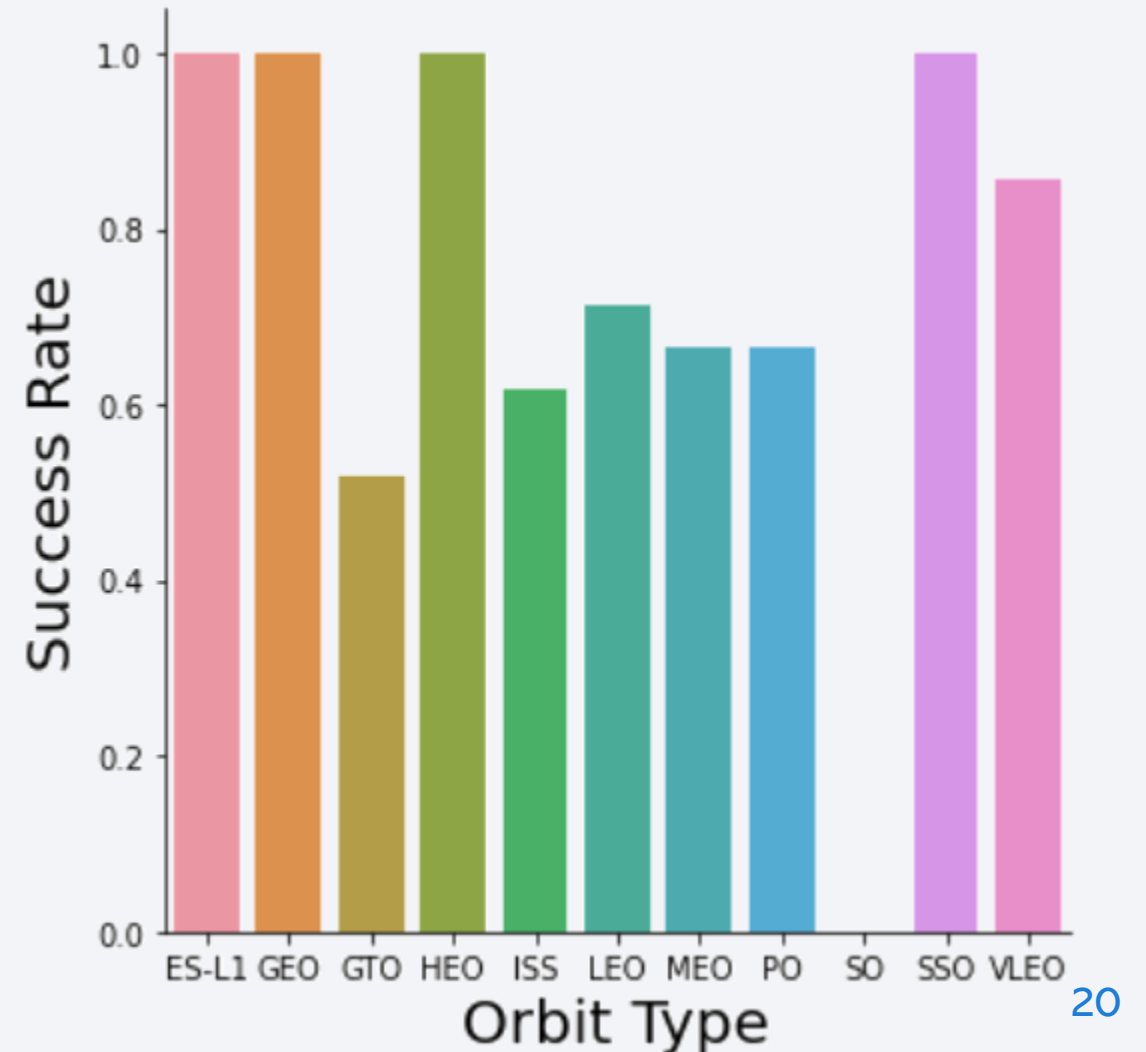
## Explicación:

- Para cada sitio de lanzamiento, cuanto mayor sea la masa de la carga útil, mayor será la tasa de éxito.
- La mayoría de los lanzamientos con una masa de carga útil superior a 7.000 kg tuvieron éxito.
- KSC LC 39A también tiene una tasa de éxito del 100% para una masa de carga útil inferior a 5500 kg

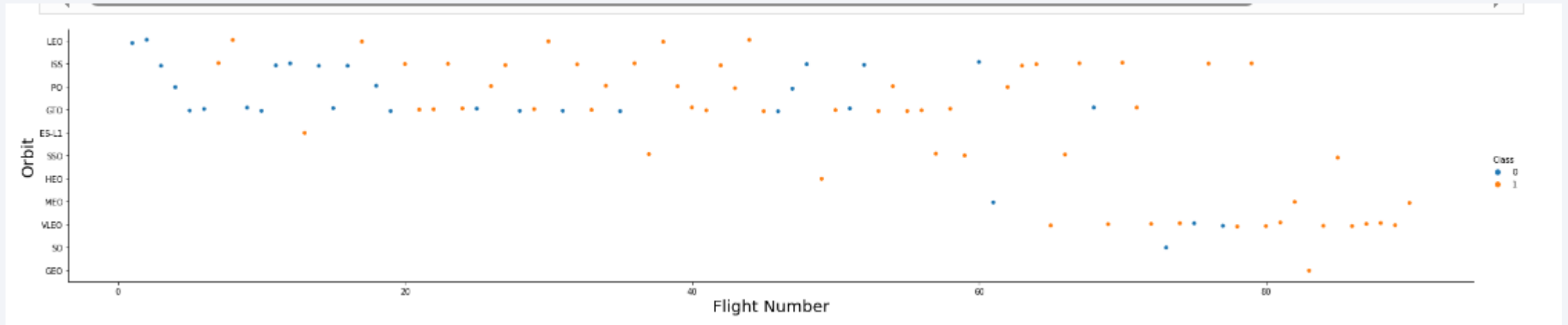
# Success Rate vs. Orbit Type

## Explicación:

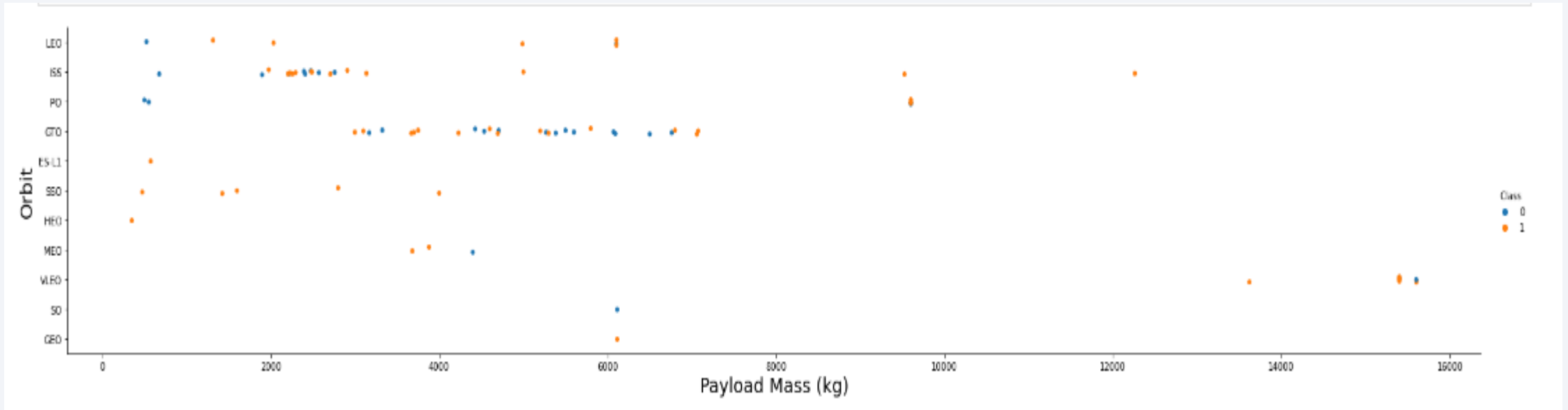
- Órbitas con una tasa de éxito del 100 %:
- -ES-L1, GEO, HEO, SSO
- • Órbitas con tasa de éxito del 0%:
- **Conclusión:**
- Órbitas con tasa de éxito entre 50% y 85%: GTO, ISS, LEO, MEO, PO



# Flight Number vs. Orbit Type



# Payload vs. Orbit Type



- Explicación:
- Las cargas útiles pesadas tienen una influencia negativa en las órbitas GTO y positiva en las órbitas GTO y Polar LEO (ISS).

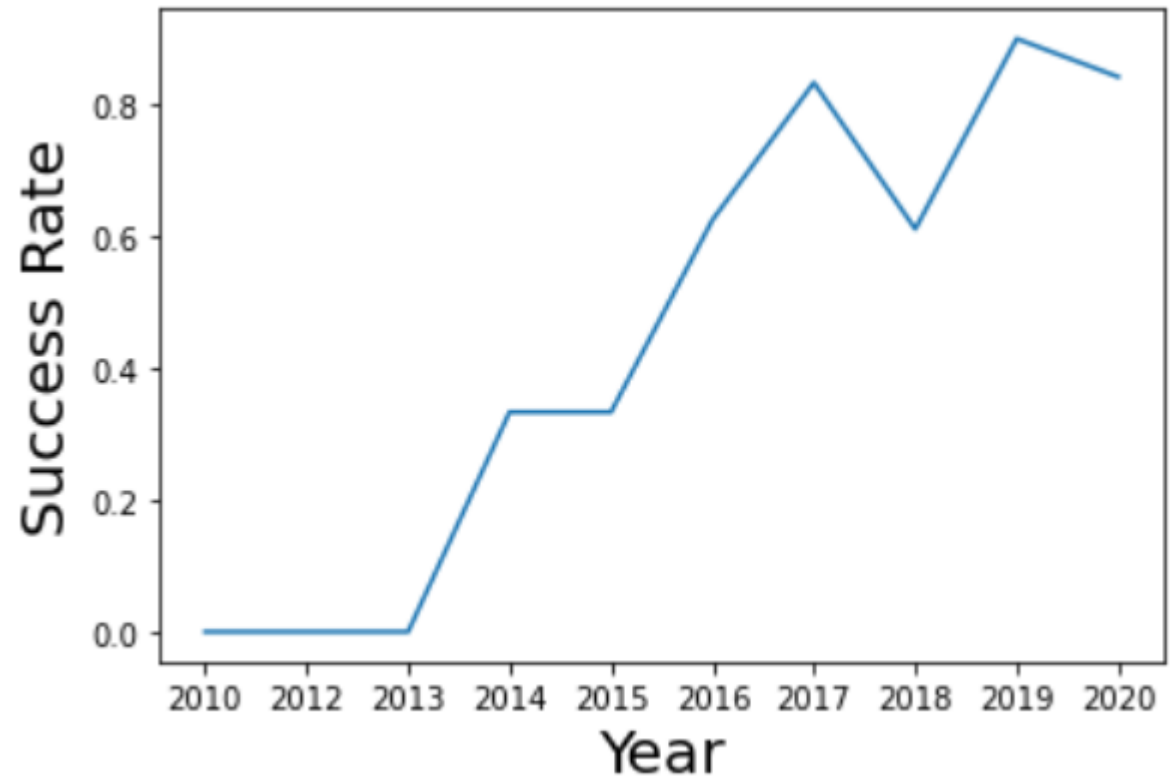
# Launch Success Yearly Trend

---

## Explicación:

La tasa de éxito es creciente desde el año 2013.

En el año 2018 se presentó una disminución en la cantidad de éxitos





# All Launch Site Names

---

## Explicación:

Se muestran los nombres únicos de la columna `launch_site` y se devuelven 4 datos

## Task 1

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

In [23]:

```
%sql SELECT DISTINCT LAUNCH_SITE FROM SPACEXTBL ORDER BY 1;
```

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

Out[23]:

Launch_Site
-------------

CCAFS LC-40
-------------

CCAFS SLC-40
--------------

KSC LC-39A
------------

VAFB SLC-4E
-------------

# Launch Site Names Begin with 'CCA'

## Explicación:

Se muestran todas las columnas de la tabla spacextbl donde launch\_site inicia con CCA, extrayendo 5 registros

### Task 2

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

In [24]:

```
%sql SELECT * FROM SPACEXTBL WHERE LAUNCH_SITE LIKE 'CCA%' LIMIT 5;
```

\* sqlite:///my\_data1.db

Done.

Out[24]:

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

# Total Payload Mass

---

## Task 3

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

```
In [25]: %sql SELECT SUM(PAYLOAD_MASS__KG_) AS TOTAL_PAYLOAD FROM SPACEXTBL WHERE PAYLOAD LIKE '%CRS%';
```

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

```
Done.
```

```
Out[25]: TOTAL_PAYLOAD  
111268
```

### Explicación:

- Visualización de la masa total de carga útil transportada por los propulsores lanzados por la NASA (CRS)

# Average Payload Mass by F9 v1.1

---

## Task 4

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

```
In [26]: %sql SELECT AVG(PAYLOAD_MASS_KG_) AS AVG_PAYLOAD FROM SPACEXTBL WHERE BOOSTER_VERSION = 'F9 v1.1';
```

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

```
Done.
```

```
Out[26]: AVG_PAYLOAD  
2928.4
```

### Explicación:

- Muestra la masa de carga útil promedio transportada por la versión de refuerzo F9 v1.1

# First Successful Ground Landing Date

---

## Task 5

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

*Hint: Use min function*

```
In [33]: %sql SELECT MIN(DATE) AS FIRST_SUCCESS_GP FROM SPACEXTBL WHERE Landing_Outcome = 'Success (ground pad)';
```

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

```
Done.
```

```
Out[33]: FIRST_SUCCESS_GP
```

```
2015-12-22
```

## Explicación:

- Listar la fecha en que se logró el primer aterrizaje exitoso en la plataforma de tierra



# Successful Drone Ship Landing with Payload between 4000 and 6000

---

## Task 6

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

In [36]: `%sql SELECT DISTINCT BOOSTER_VERSION FROM SPACEXTBL WHERE PAYLOAD_MASS_KG_ BETWEEN 4000 AND 6000 AND Landing_Outcome = 'Succ`

\* sqlite:///my\_data1.db

Done.

Out[36]: **Booster\_Version**

F9 FT B1022

F9 FT B1026

F9 FT B1021.2

F9 FT B1031.2

### Explicación:

- Enumerar los nombres de los propulsores que tienen éxito en naves no tripuladas y tienen una masa de carga superior a 4000 pero inferior a 6000

# Total Number of Successful and Failure Mission Outcomes

---

## Task 7

List the total number of successful and failure mission outcomes

```
In [37]: %sql SELECT MISSION_OUTCOME, COUNT(*) AS QTY FROM SPACEXTBL GROUP BY MISSION_OUTCOME ORDER BY MISSION_OUTCOME;
```

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

Done.

```
Out[37]:
```

Mission_Outcome	QTY
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

### Explicación:

- Enumerar el número total de resultados de misiones exitosas y fallidas

# Boosters Carried Maximum Payload

## Task 8

List the names of the booster\_versions which have carried the maximum payload mass. Use a subquery

In [38]:

```
%sql SELECT DISTINCT BOOSTER_VERSION FROM SPACEXTBL WHERE PAYLOAD_MASS_KG_ = (SELECT MAX(PAYLOAD_MASS_KG_) FROM SPACEXTBL)
```

\* sqlite:///my\_data1.db

Done.

Out[38]:

Booster_Version
-----------------

F9 B5 B1048.4
---------------

F9 B5 B1048.5
---------------

F9 B5 B1049.4
---------------

F9 B5 B1049.5
---------------

F9 B5 B1049.7
---------------

F9 B5 B1051.3
---------------

F9 B5 B1051.4
---------------

F9 B5 B1051.6
---------------

F9 B5 B1056.4
---------------

F9 B5 B1058.3
---------------

F9 B5 B1060.2
---------------

F9 B5 B1060.3
---------------

### Explicación:

- Listar los nombres de las versiones de propulsores que han llevado la masa máxima de carga útil

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

## Task 10

Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order.

In [49]:

```
%%sql select Landing_Outcome, count(*) as count_outcomes from SPACEXTBL
      where Date between '2010-06-04' and '2017-03-20'
      group by Landing_Outcome
      order by count_outcomes desc;
```

\* sqlite:///my\_data1.db

Done.

Out[49]:

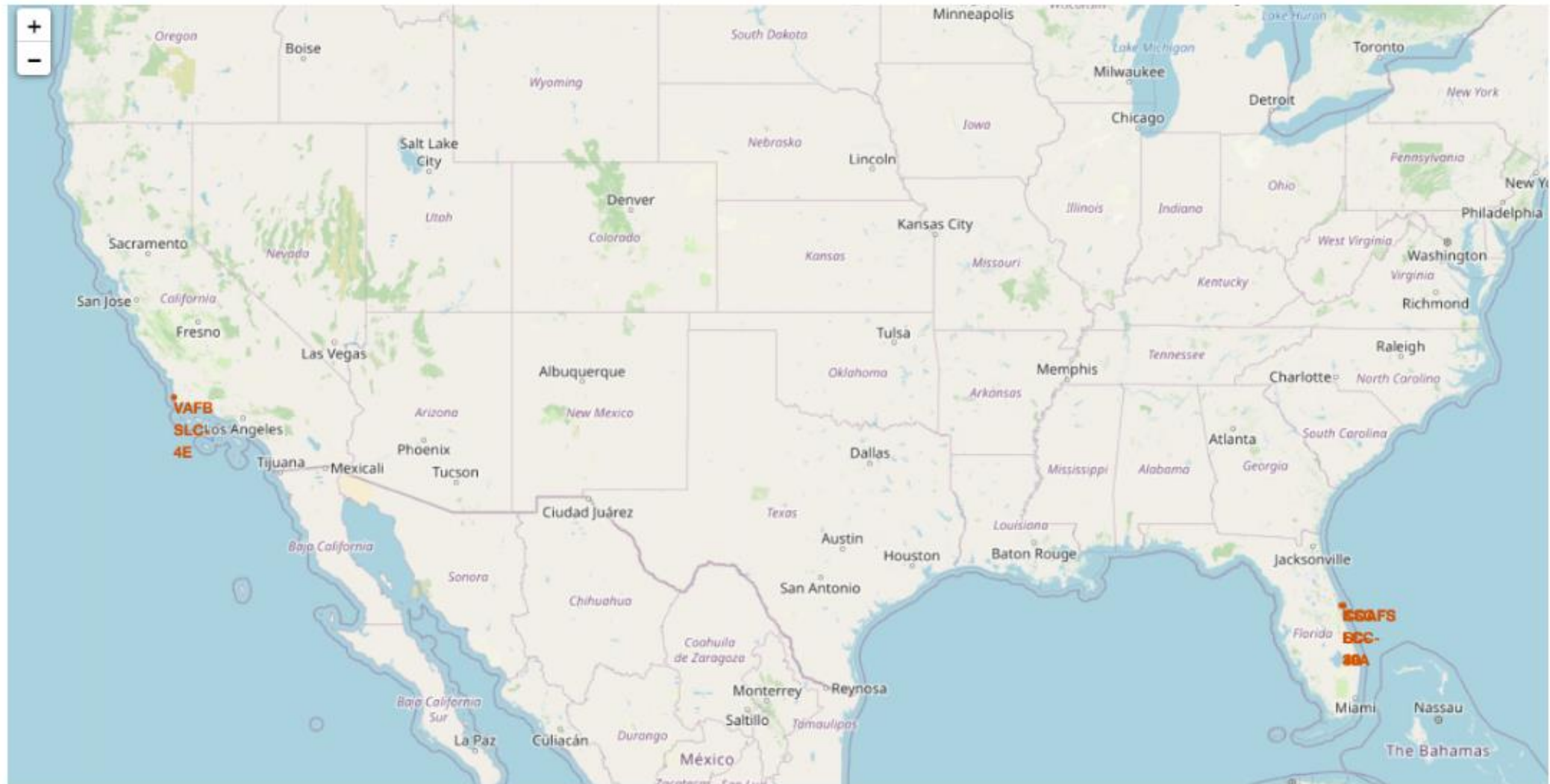
Landing_Outcome	count_outcomes
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

A satellite view of Earth from space, showing the curvature of the planet and city lights at night. The background is a deep blue gradient.

Section 3

# Launch Sites Proximities Analysis

# Mapa de Locaciones de lanzamiento

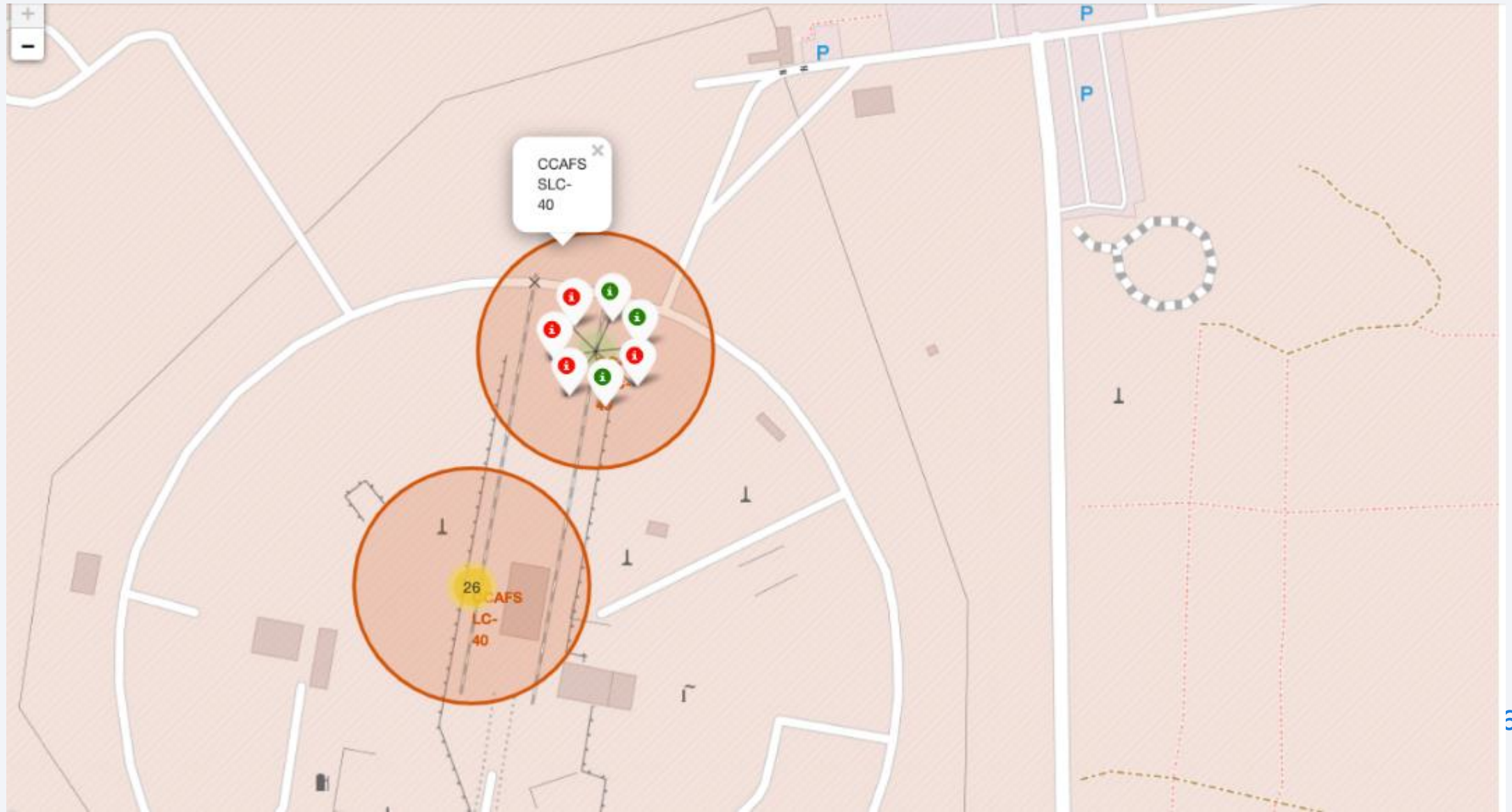








# Label de colores en el mapa





Section 4

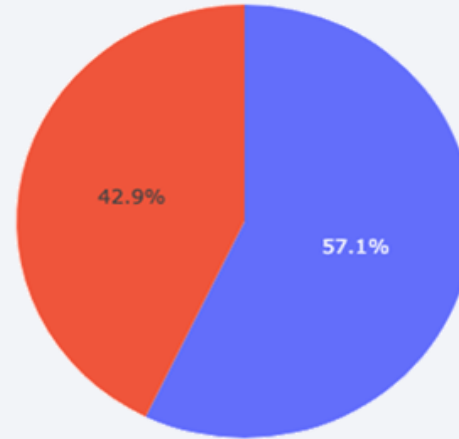
# Build a Dashboard with Plotly Dash

# Dashboard

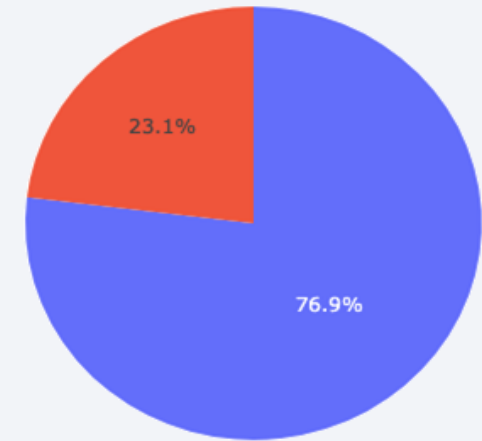
## Explicación:

- Se presentan el total de lanzamientos exitosos por cada sitio

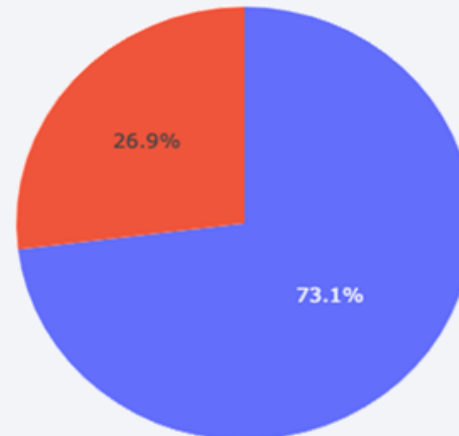
Total Success Launches for Site CCAFS SLC-40



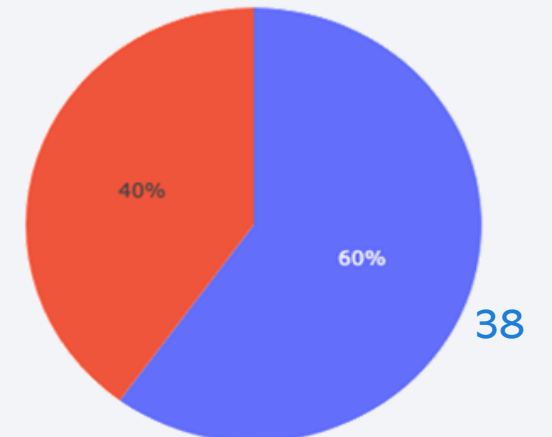
Total Success Launches for Site KSC LC-39A



Total Success Launches for Site CCAFS LC-40



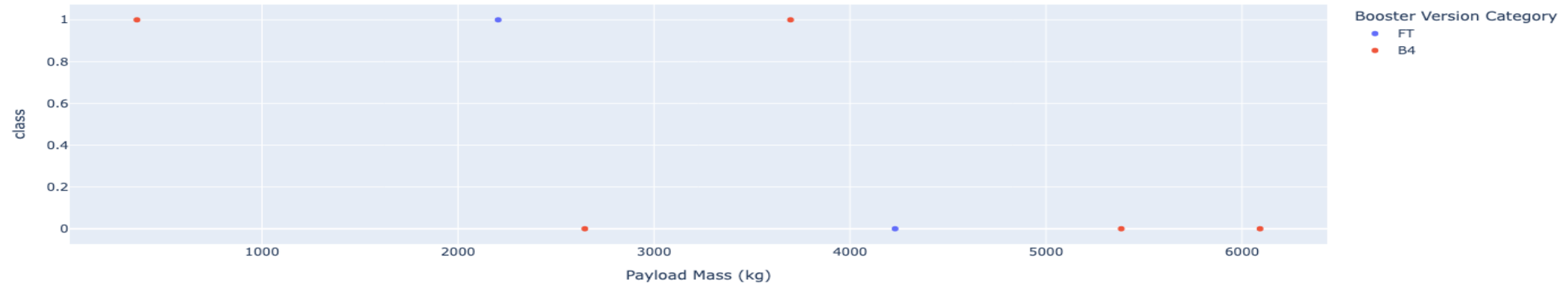
Total Success Launches for Site VAFB SLC-4E



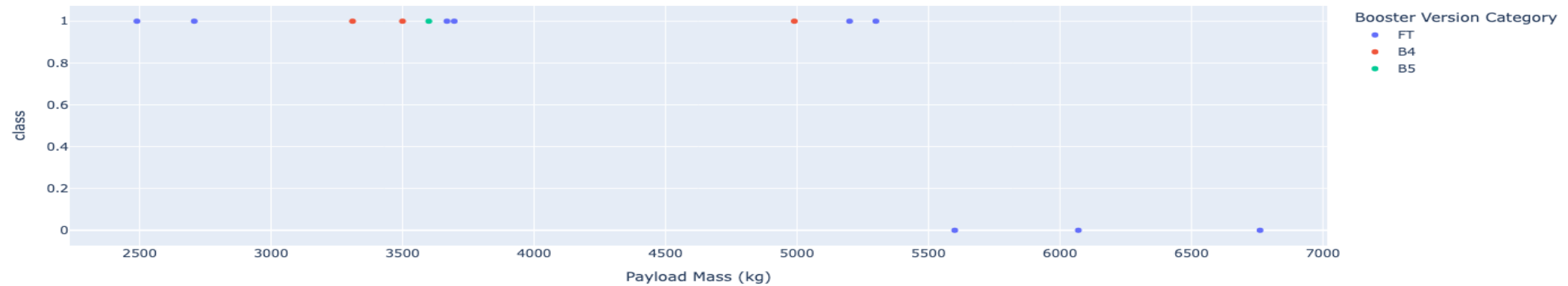


# Gráficos de cargas útiles

Correlation entre Payload and Success por Site CCAFS SLC-40

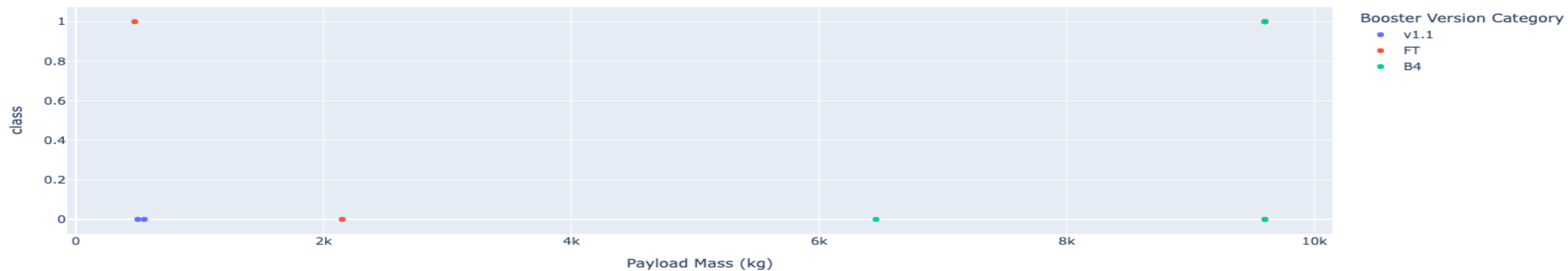


Correlation entre Payload and Success por Site KSC LC-39A

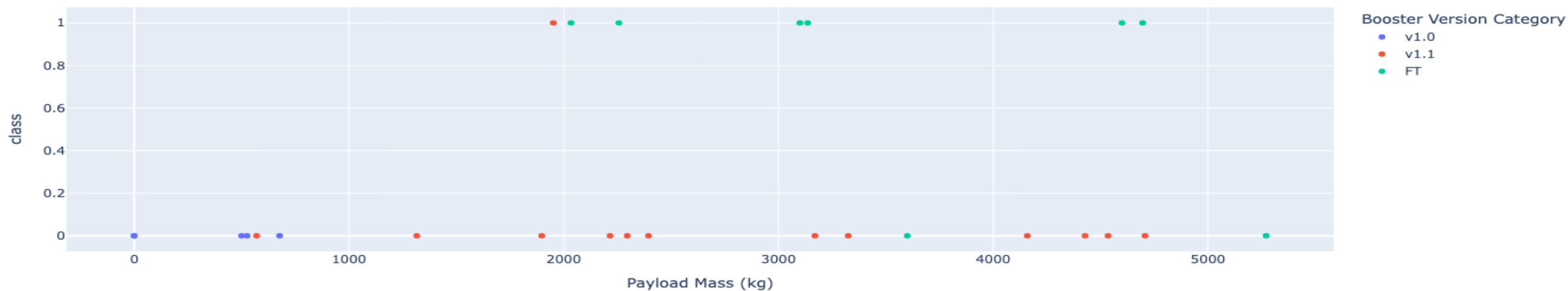


# Graficos de cargas útiles

Correlation entre Payload and Success por Site VAFB SLC-4E



Correlation entre Payload and Success por Site CCAFS LC-40



Section 5

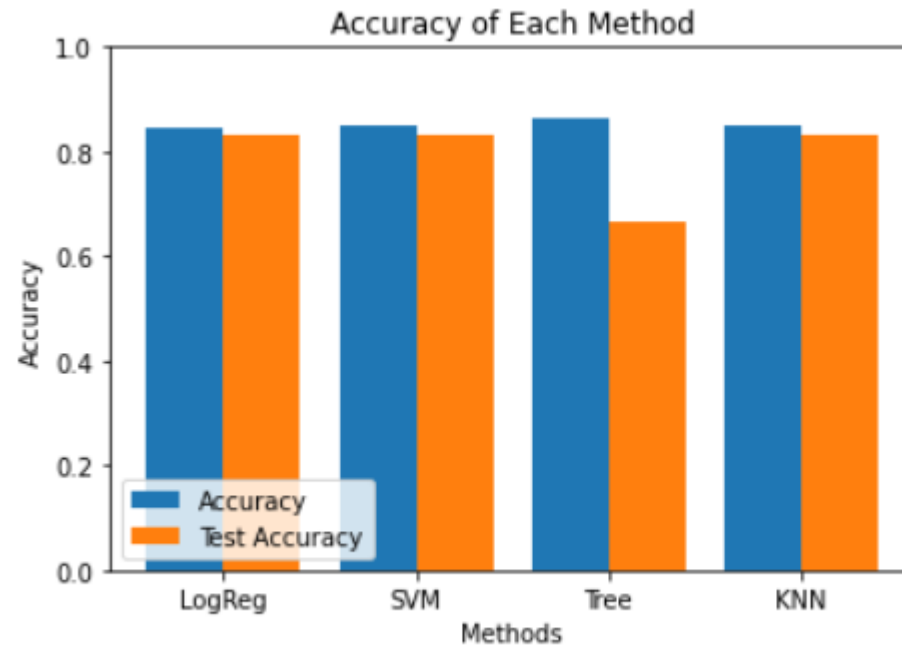
# Predictive Analysis (Classification)



# Classification Accuracy

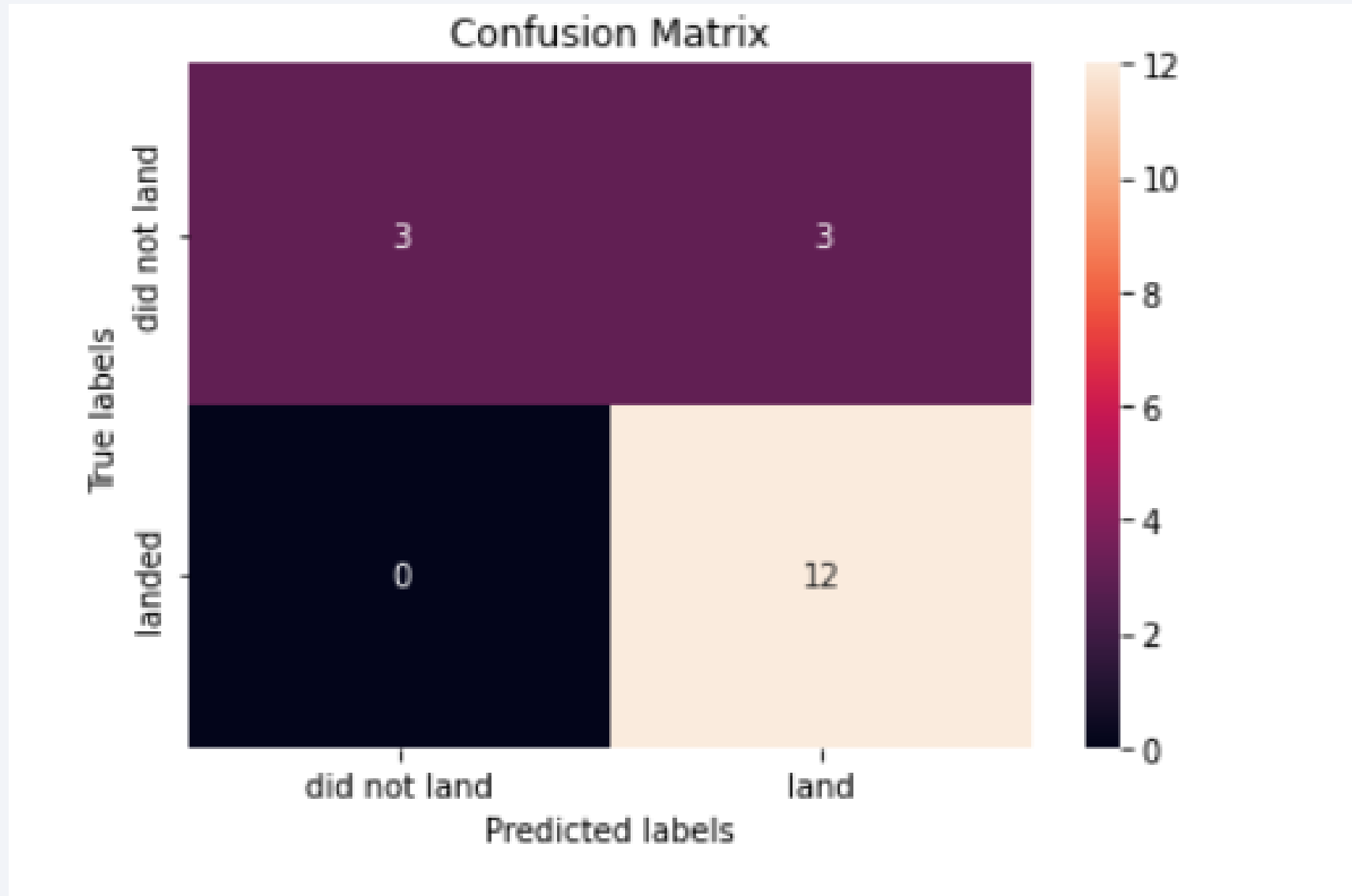
## Explicación:

- Tree Model es el que tiene mejor Accuracy



Model	Accuracy	TestAccuracy
LogReg	0.84643	0.83333
SVM	0.84821	0.83333
Tree	0.86429	0.66667
KNN	0.84821	0.83333

# Confusion Matrix



# Conclusions

---

- Las órbitas ES-L1, GEO, HEO y SSO tienen una tasa de éxito del 100%
- KSC LC-39A es la que tiene la mayor tasa de éxito
- La tasa de éxito crece con los años desde el año 2013, con una disminución en 2017
- El modelo de árbol de decisión es el que tiene la mayor precisión

# Appendix

---

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!

