



IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Ana María Jiménez Ortega
30-11-2024



INDICE

- Introducción
- Resumen ejecutivo
- Metodología
- Resultados
- Conclusión
- Anexo

Introducción

Con fin de determinar el precio de los lanzamientos de los cohetes fabricados por la empresa Space Y, hemos procedido a usar un conjunto de técnicas de análisis de datos sobre el funcionamiento de los cohetes Falcon 9 de la empresa Xpace X de Allon Musk, nuestra principal competidora y una de las principales empresas en el sector de la fabricación aeroespacial y en la realización de servicios de transportes espaciales.

El Falcon 9 es un cohete que permite su reutilización y, como consecuencia, una reducción de costes y un abaratamiento de los precios a posibles clientes, por lo que el objetivo de este estudio ha sido analizar los datos que proporciona el funcionamiento este cohete, como, por ejemplo, el número de vuelo(**FlightNumber**), el sitio de lanzamiento(**LaunchSite**), el tipo de órbita(**Orbit**), la masa de carga útil(**PayloadMass**) o el número de aterrizaje exitoso o fracasados(**landing_Outcomes**), punto, este último, esencial para predecir si un vuelo va a aterrizar con éxito y con ello permitir la reutilización, posterior, del cohete y determinar el costo del lanzamiento.

Resumen ejecutivo

Para analizar las distintas variables que hace que la empresa Space X pueda ofrecer servicios aeroespaciales a precios bajos con respecto a otras empresas del mismo sector hemos utilizado algunas fuentes de datos que ofrece información sobre el funcionamiento del cohete Falcon 9.

A estas fuentes de datos le hemos aplicado una serie de metodologías, cuyos detalles expondremos en las distintas secciones que componen esta presentación y que nos ha llevado a una serie de conclusiones para determinar, en nuestro caso, que factores de nuestros cohetes puede ayudarnos a poder competir en precio con la empresa de Allon Musk.

Metodología

Methodology

Methodology

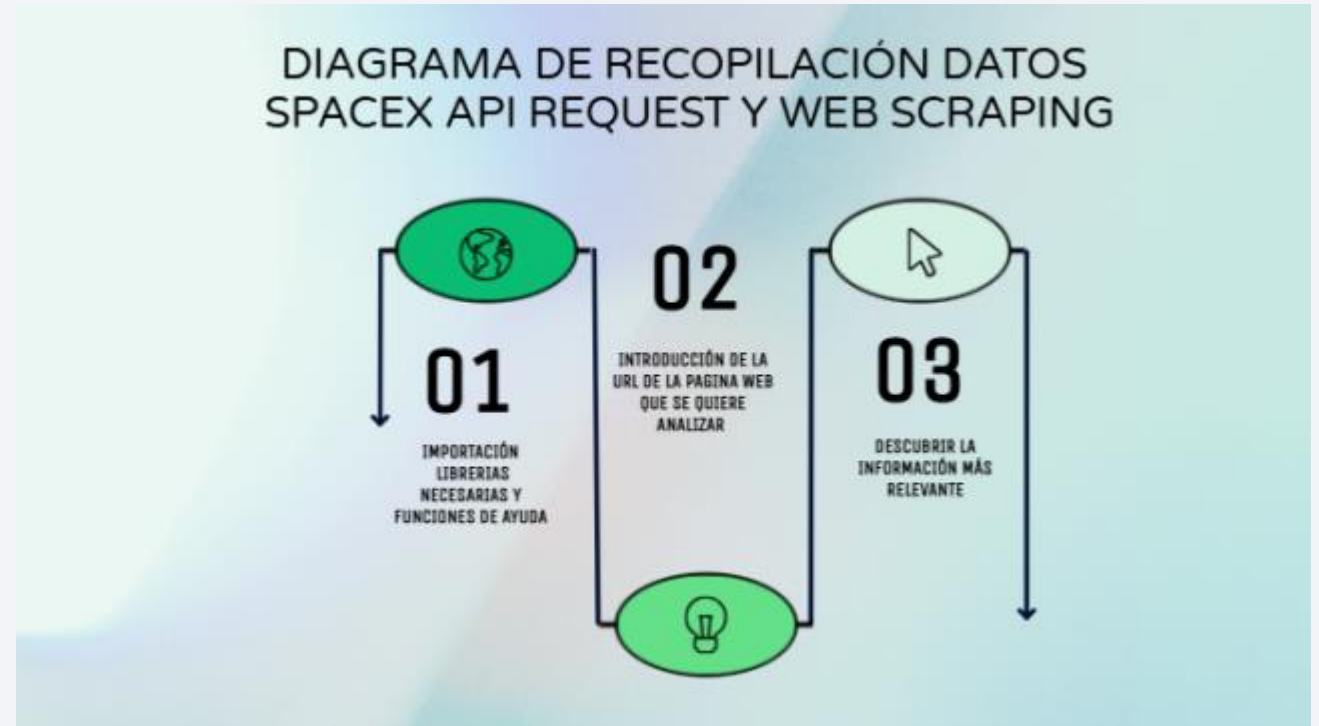
Executive Summary

- Uso de metodología de recolección de datos: usando API con Web Scraping de Wikipedia
- Manipulación de datos con el fin de determinar los datos faltantes ,el tipo de datos y conversión del tipo de dato de la columna “Outcome”
- Análisis exploratorios de datos (EDA) mediante visualización y SQL
- Análisis visuales interactivos con Folium y Plotly Dash
- Análisis predictivos mediante los siguientes modelos de clasificación:
 - Regresión logística
 - KNN
 - SVM
 - Arboles de decision

Recogida de datos

Los conjuntos de datos
son recopilados a través:

- SpaceX API request.
- Web Scraping



Ver anexo

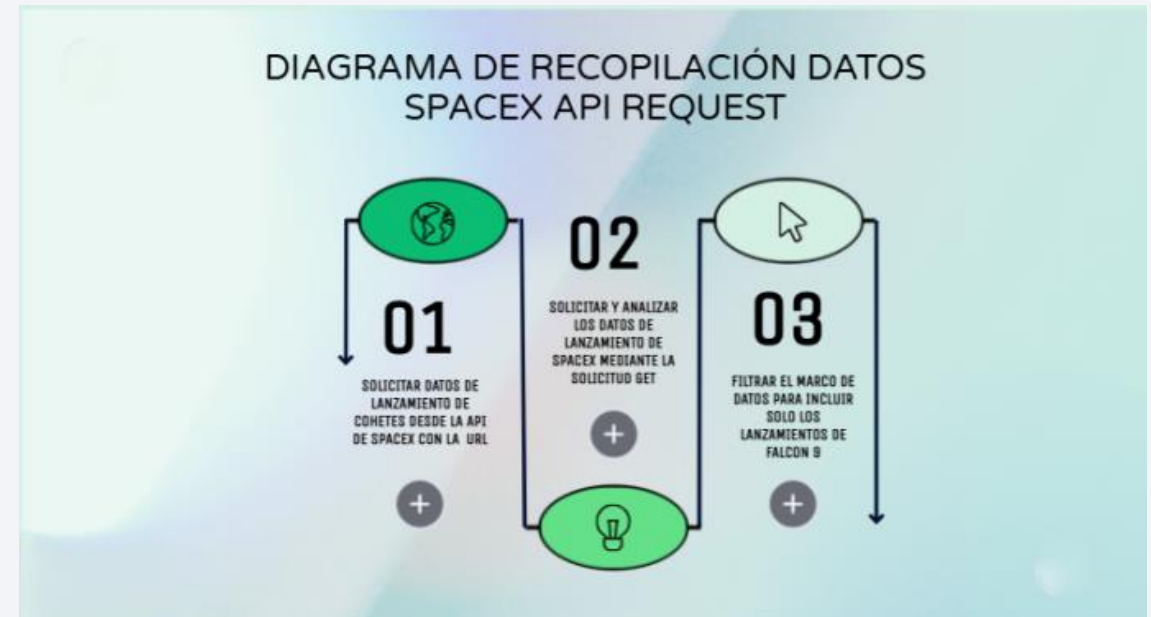
Recogida de datos– SpaceX API

La forma en que se ha realizado la recopilación de datos se proporciona en forma de diagrama de flujo para obtener una descripción general.

El enlace para el cuaderno completo se proporciona a continuación:

URL de GitHub de la API de SpaceX completada

(https://github.com/Anajior/Ciencia-de-datos-aplicada-Capstone/blob/main/1_jupyter-labs-spacex-data-collection-api.ipynb)



Recogida de datos - Scraping

La forma en que se ha realizado la recopilación de datos se proporciona en forma de diagrama de flujo para obtener una descripción general.

Enlace al cuaderno completo:

https://github.com/Anajior/Ciencia-de-datos-aplicada-Capstone/blob/main/2_jupyter-labs-webscraping.ipynb



Data Wrangling

El proceso de organización de datos se proporciona en un diagrama de flujo para obtener una visión general.

Enlace al cuaderno completo:

https://github.com/Anajior/Ciencia-de-datos-aplicada-Capstone/blob/main/3_%20labs-jupyter-spacex-Data%20wrangling.ipynb



Ver anexo

EDA con visualización de datos

El análisis de datos exploratorios (EDA) nos permite analizar e investigar conjuntos de datos y resumir sus características principales, a menudo, empleando métodos de visualización de datos.

Para esta metodología utilizado los siguientes gráficos:

- Diagrama de dispersión (**SCATTER PLOT**):

En nuestro caso nos permite buscar la relación entre dos variables, por ejemplo, entre Flight Number and Launch Site, o Payload Mass and Launch Site

- Gráfica de barras (**BAR CHART**)

En nuestro caso nos permite comparar dos variables, como son: “Class” y “Orbit”

- Gráfica de líneas (**LINE CHART**):

En nuestro caso nos permite visualizar tendencias y cambios a lo largo del tiempo.

- Add the GitHub URL of your completed EDA with data visualization notebook, as an external reference and peer-review purpose

EDA con SQL

En este apartado de nuestro análisis exploratorio de datos (EDA) hemos realizado las siguientes consultas SQL:

- 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'
- Mostrar la masa total de la carga útil transportada por los propulsores lanzados por la NASA (CRS)
- Mostrar la masa promedio de la carga útil transportada por el cohete de la versión F9 v1.1
- Indicar la fecha en la que se logró el primer aterrizaje exitoso en la plataforma de tierra
- Enumerar los nombres de los propulsores que tienen éxito en los barcos no tripulados y tienen una masa de carga útil mayor a 4000 pero menor a 6000
- Enumerar el número total de resultados de misiones exitosas y fallidas.
- Enumerar los nombres de las versiones de refuerzo que han transportado la masa máxima de carga útil. Utilice una subconsulta
- Enumerar los registros que mostrarán los nombres de los meses, los resultados de los aterrizajes fallidos en el barco no tripulado, las versiones de los propulsores y el sitio de lanzamiento para los meses del año 2015.
- Clasificar el recuento de resultados de aterrizaje (como Fallo (dron) o Éxito (plataforma de tierra)) entre la fecha del 4 de junio de 2010 y el 20 de marzo de 2017, en orden descendente
- Add the GitHub URL of your completed EDA with SQL notebook, as an external reference and peer-review purpose

Construya un mapa interactivo con Folium

Folium es una biblioteca de Python que nos permite visualizar datos geoespaciales y añadirles marcadores para mostrar los sitios de lanzamiento de SpaceX y sus puntos de referencia más importantes más cercanos como ferrocarriles, carreteras, ciudades y costas, para ello se utilizaron polilíneas para conectarlos.

Para nuestro estudio se añadieron los siguientes marcadores:

- **Folium.Circle:** nos agrega un área de círculo resaltada con una etiqueta de texto en una coordenada específica
- **Folium.Marker:** Nos permite mostrar la ubicación de los sitios de lanzamiento en el mapa. Dentro de estos marcadores utilizamos **Folium.Icon** para distinguir los éxitos (**verde**) de los fracasos(**rojo**) de los lanzamientos.
- **Folium.Polyline:** Muestra elementos lineales en el mapa que, en este caso, nos indica las distancias entre el sitio de lanzamiento y un lugar próximo (carretera, ferrocarril, costa o ciudad)
- Add the GitHub URL of your completed interactive map with Folium map, as an external reference and peer-review purpose

Construya un tablero con Plotly Dash

Plotly Dash es un marco de Python para crear aplicaciones analíticas web, por lo que es ideal para crear aplicaciones de visualización de datos con interfaces de usuario altamente personalizadas.

En nuestro caso hemos utilizado los siguientes tipos de gráficos:

- Circular (**PIE-CHAR**):

Este gráfico nos permite reflejar el porcentaje de éxitos o fracasos en los lanzamientos con relación al lugar donde se producen los mismos.

- Diagrama de dispersión (**SCATTER-CHART**):

Como hemos dicho anteriormente, el diagrama de dispersión nos permite reflejar la relación entre dos variables, en este caso, la relación que existe entre la variable “Masa de carga útil (**Payload Mass (kg)**)” y la variable “Class”.

- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

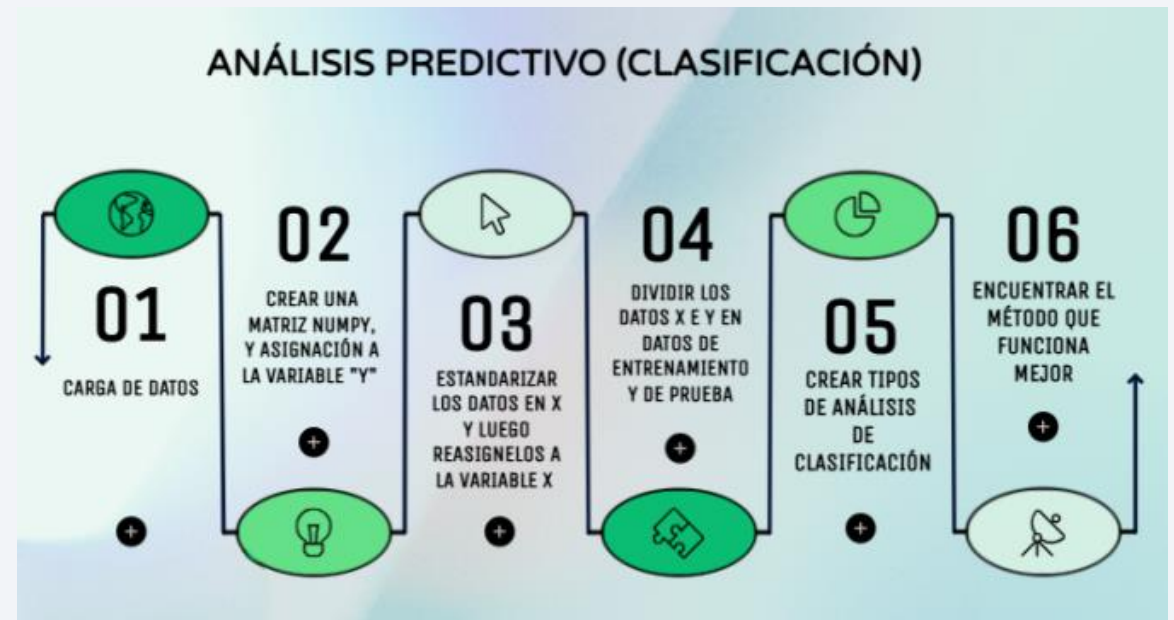
Análisis Predictivo (Clasificación)

En esta última parte de la sección se ha creado un sistema de aprendizaje automático para predecir si la primera fase del Falcon 9 aterriza con éxito.

Para ello se ha seguido los pasos mostrados en el gráfico.

Enlace al cuaderno completo:

https://github.com/Anajior/Ciencia-de-datos-aplicada-Capstone/blob/main/8_SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb



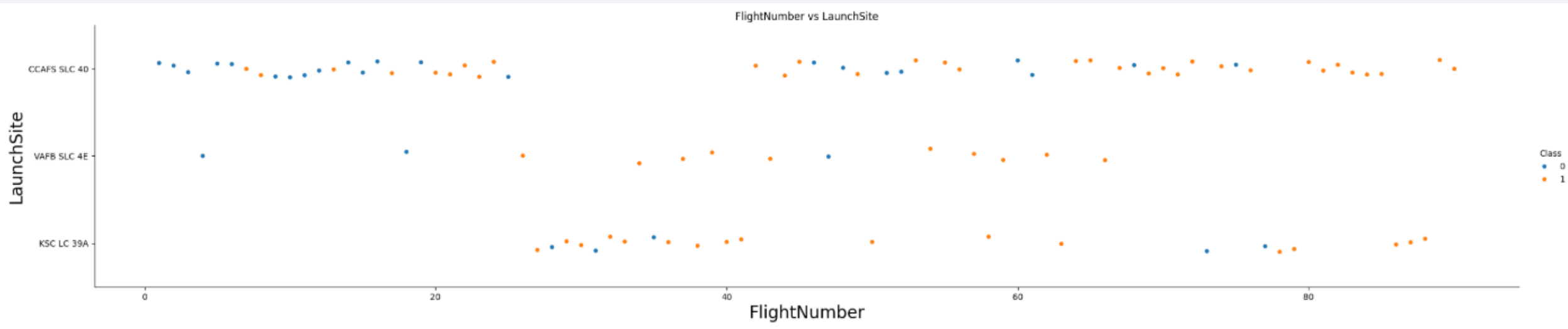
Ver anexo



Perspectivas extraídas de EDA

Insights drawn
from EDA

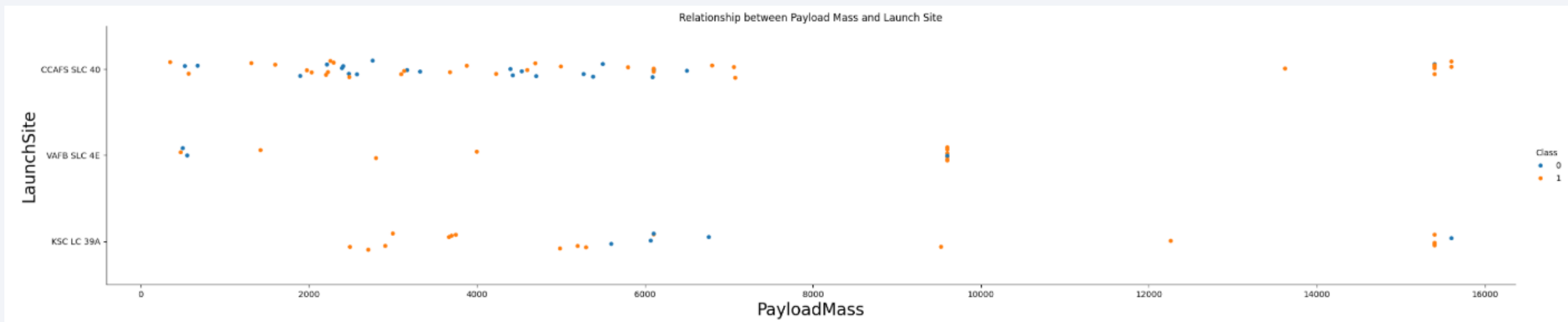
Número de vuelo vs. sitio de lanzamiento



En imagen se representa los éxitos y fracasos de los distintos vuelos realizados desde los diferentes sitios en los que se efectuaron los mismos.

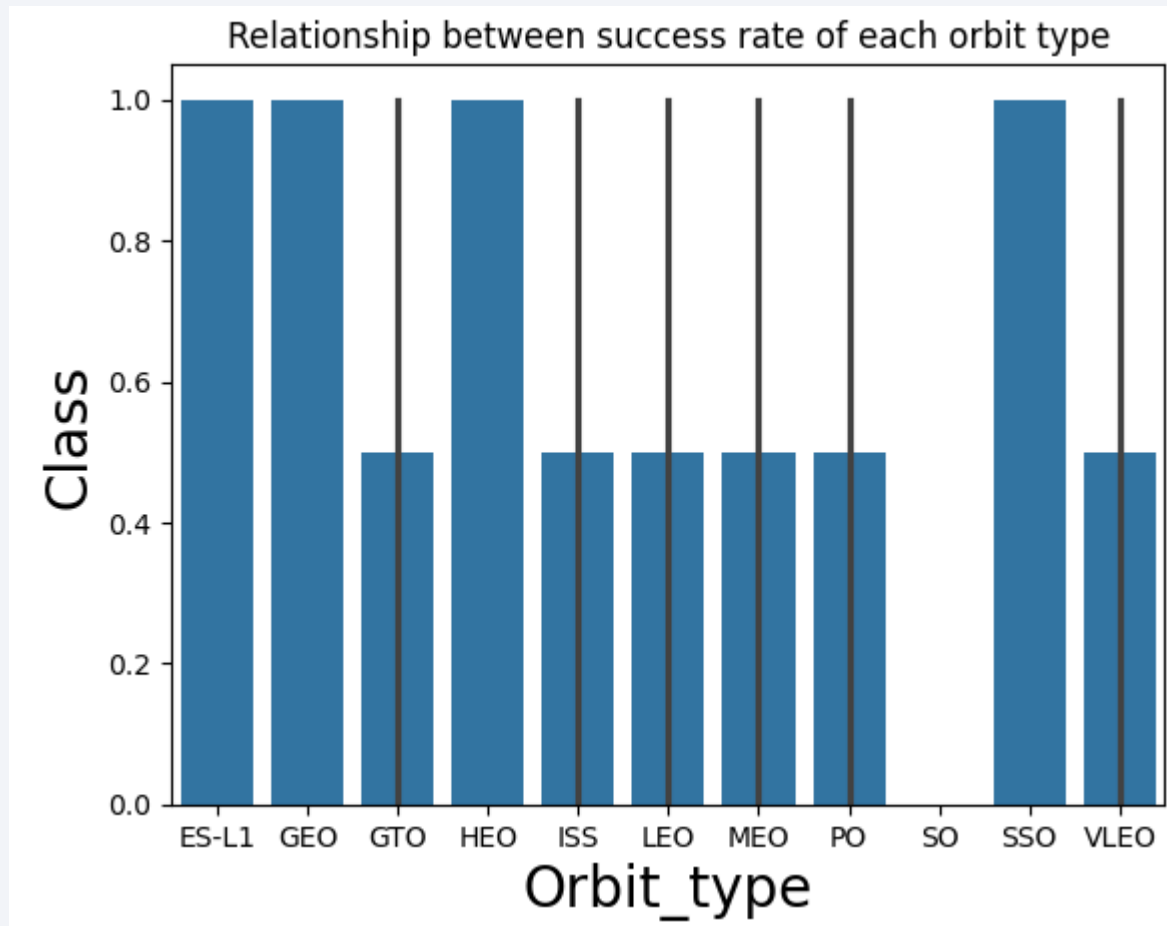
Así podemos ver que desde “CCAFS SLC-40 “ se produjeron mas lanzamientos y resultados positivos (**rojo**), mientras desde “VAFB SLC-4E” se produjeron menos lanzamientos, pero solo dos terminaron en fracaso (**azul**).

Carga útil vs. sitio de lanzamiento



En el grafico observamos los puntos de dispersión de masa de carga útil (PayloadMass) vs lugar de lanzamiento (Lauchsite), y encontramos que para el sitio de lanzamiento VAFB-SLC 4E no hay cohetes lanzados para una masa de carga útil pesada (superior a 10 000).

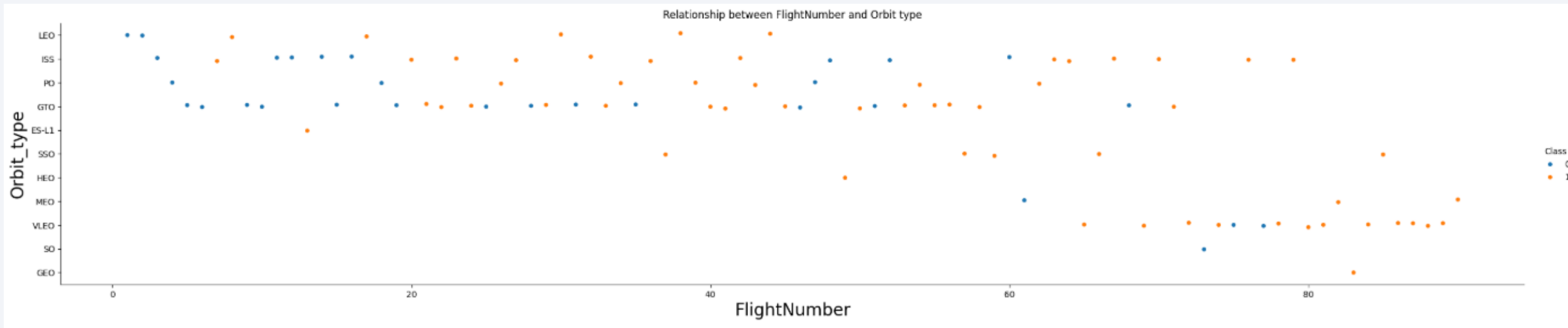
Tasa de éxito frente al tipo de órbita



Este gráfica nos indica qué tipo de órbita ha tenido una mayor tasa de éxito. En este caso, son:

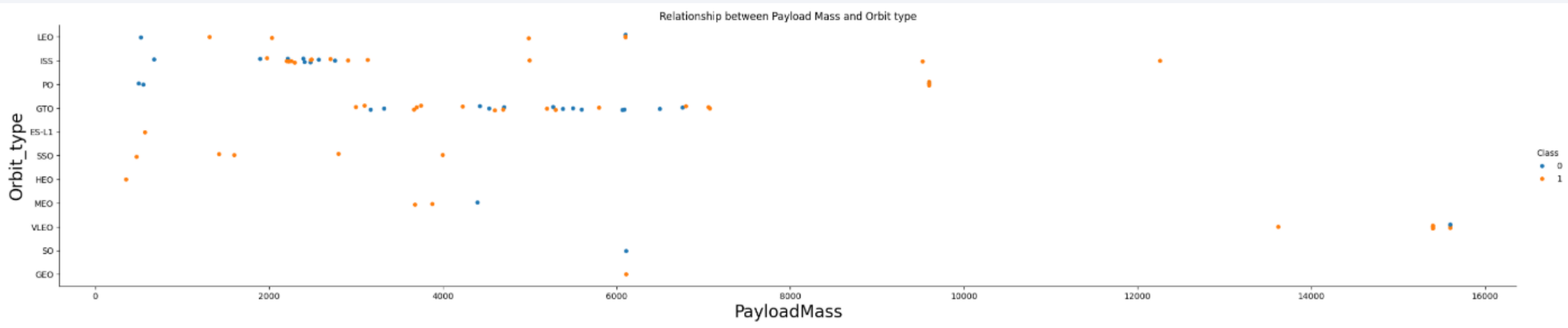
- ES-L1
- GEO
- HEO
- SSO

Número de vuelo vs. tipo de órbita



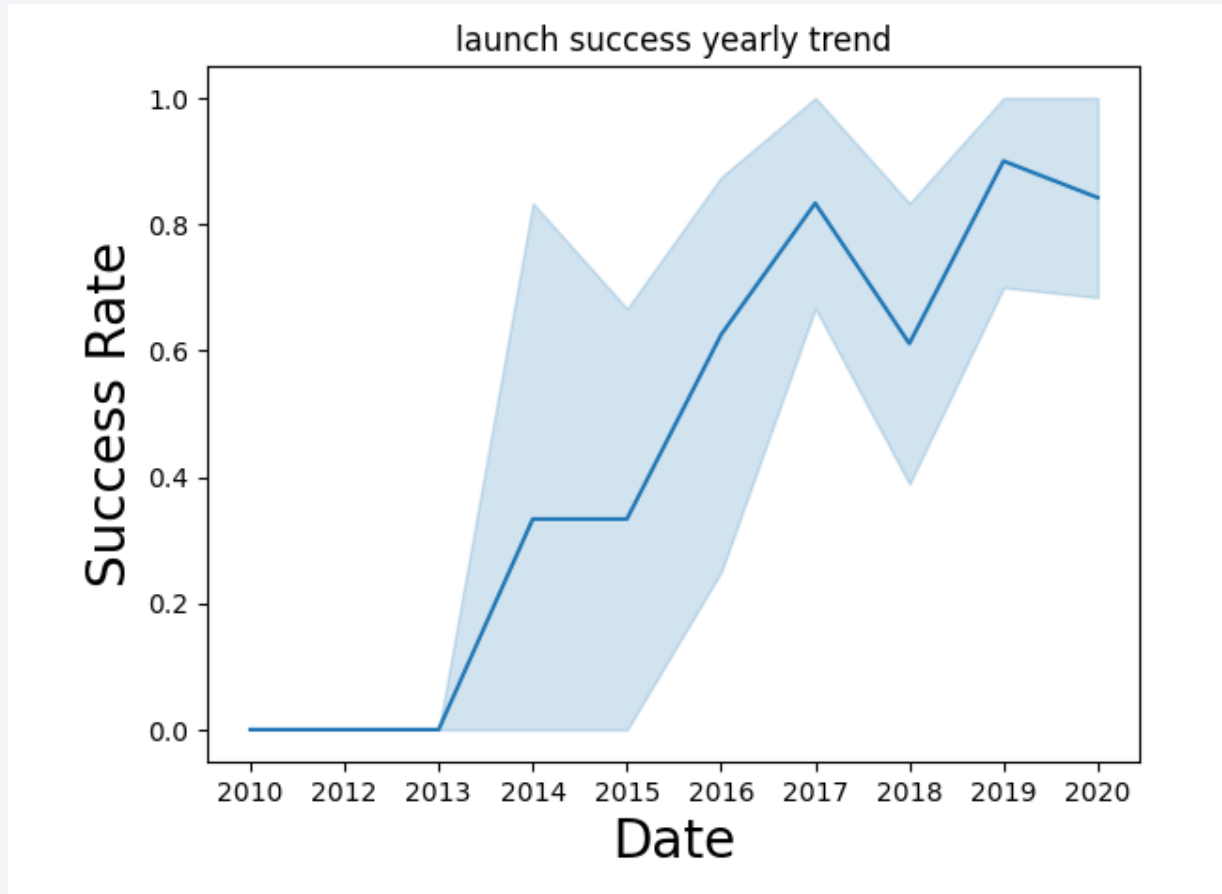
Se puede observar que en la órbita LEO, el éxito parece estar relacionado con el número de vuelos. Por el contrario, en la órbita GTO, no parece haber relación entre el número de vuelos y el éxito.

Carga útil vs. tipo de órbita



Aquí vemos que, con cargas útiles pesadas(Payload), la tasa de aterrizaje exitoso o de aterrizaje positivo es mayor para Polar, LEO e ISS.

Tendencia anual de éxito de lanzamiento



En esta gráfica podemos ver que la tasa de éxito más alta se produjo en el año 2019

Todos los nombres de sitios de lanzamiento

A la vista de los datos, estos son los nombres de los sitios de lanzamientos (LaunchSite):

CCAFS LC-40

KSC LC-39A

VAFB SLC-4E

Los nombres de los sitios de lanzamiento comienzan con 'CCA'

```
[11]: %sql Select Launch_Site from SPACEXTBL where Launch_Site like ('CCA%') LIMIT 5;
```

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

```
[11]: Launch_Site
```

```
CCAFS LC-40
```

```
CCAFS LC-40
```

```
CCAFS LC-40
```

```
CCAFS LC-40
```

```
CCAFS LC-40
```

La imagen muestra los 5 primeros registros en los que los sitios de lanzamiento comienzan con las letras 'CCA'

Total Masa de carga útil

```
[12]: %sql SELECT sum(PAYLOAD_MASS__KG_) as Totalpayloadmasskg from SPACEXTBL where Customer ="NASA (CRS)";  
* sqlite:///my_data1.db  
Done.  
[12]: Totalpayloadmasskg  
45596
```

La imagen muestra el total de la masa total de la carga útil transportada por los cohetes lanzados por la NASA, es decir, 45596 Kg.

Promedio de la Masa de carga útil de F9 v1.1

```
[13]: %sql SELECT avg(PAYLOAD_MASS_KG_) as Averagepayloadmasskg from SPACEXTBL where Booster_Version = "F9 v1.1";
```

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

```
Done.
```

```
[13]: Averagepayloadmasskg  
2928.4
```

La imagen muestra masa media de la carga útil transportada por el F9 v1.1 , que fue de 2928,4 kg

Primera fecha de aterrizaje en tierra exitosa

```
[14]: %sql SELECT Date FROM SPACEXTBL where Landing_Outcome like '%ground pad%';

* sqlite:///my_data1.db
Done.

[14]:      Date
-----
2015-12-22
```

En la imagen de arriba se puede ver que la primera plataforma terrestre exitosa fue el 22 de diciembre de 2015.

Aterrizaje exitoso de barcos no tripulados con carga útil entre 4000 y 6000

```
[15]: %sql select Booster_Version from SPACEXTBL where Landing_Outcome='Success (drone ship)' and PAYLOAD_MASS__KG_ BETWEEN 4001 and 5999;
```

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

```
Done.
```

```
[15]: Booster_Version
```

```
F9 FT B1022
```

```
F9 FT B1026
```

```
F9 FT B1021.2
```

```
F9 FT B1031.2
```

Esta imagen enumera los nombres de los propulsores que han aterrizado con éxito en un barco no tripulado y tenían una masa de carga útil superior a 4000 pero inferior a 6000

Número total de resultados de misión exitosos y fallidos

```
[16]: %sql select count(*) as TotalMission_Outcome from SPACEXTBL;

* sqlite:///my_data1.db
Done.
[16]: TotalMission_Outcome
-----
          101
```

Esta imagen calcula el número total de resultados de misión exitosos y fallidos

Los propulsores que transportaron una carga útil máxima (Payload)

```
[17]: %sql select Booster_Version from SPACEXTBL where PAYLOAD_MASS_KG_ = (select max(PAYLOAD_MASS_KG_)from SPACEXTBL);
```

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

```
[17]: Booster_Version
```

F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1051.3
F9 B5 B1056.4
F9 B5 B1048.5
F9 B5 B1051.4
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3
F9 B5 B1051.6
F9 B5 B1060.3
F9 B5 B1049.7

En la imagen de arriba se muestra que 12 propulsores han transportado la masa máxima de carga útil de 15600 kg.

Récords de lanzamiento 2015

```
[18]: %sql SELECT Booster_Version,Launch_Site,Landing_Outcome FROM SPACEXTBL WHERE Landing_Outcome = 'Failure (drone ship)' AND Date BETWEEN '2015-01-01' AND '2015-12-31'
```

* sqlite:///my_data1.db
Done.

```
[18]:
```

Booster_Version	Launch_Site	Landing_Outcome
F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

La imagen presenta 2 versiones del Falcon 9 que tuvieron un récord de lanzamiento en el año 2015

Resultados de aterrizaje de clasificación entre 2010-06-04 y 2017-03-20

```
[19]: %sql SELECT Landing_Outcome, COUNT(landing_outcome) AS "Total" FROM SPACEXTBL WHERE (DATE BETWEEN '2010-06-04' AND '2017-03-20') GROUP BY Landing_outcome ORDER BY "Total" DESC;
```

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

```
[19]:
```

Landing_Outcome	Total
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

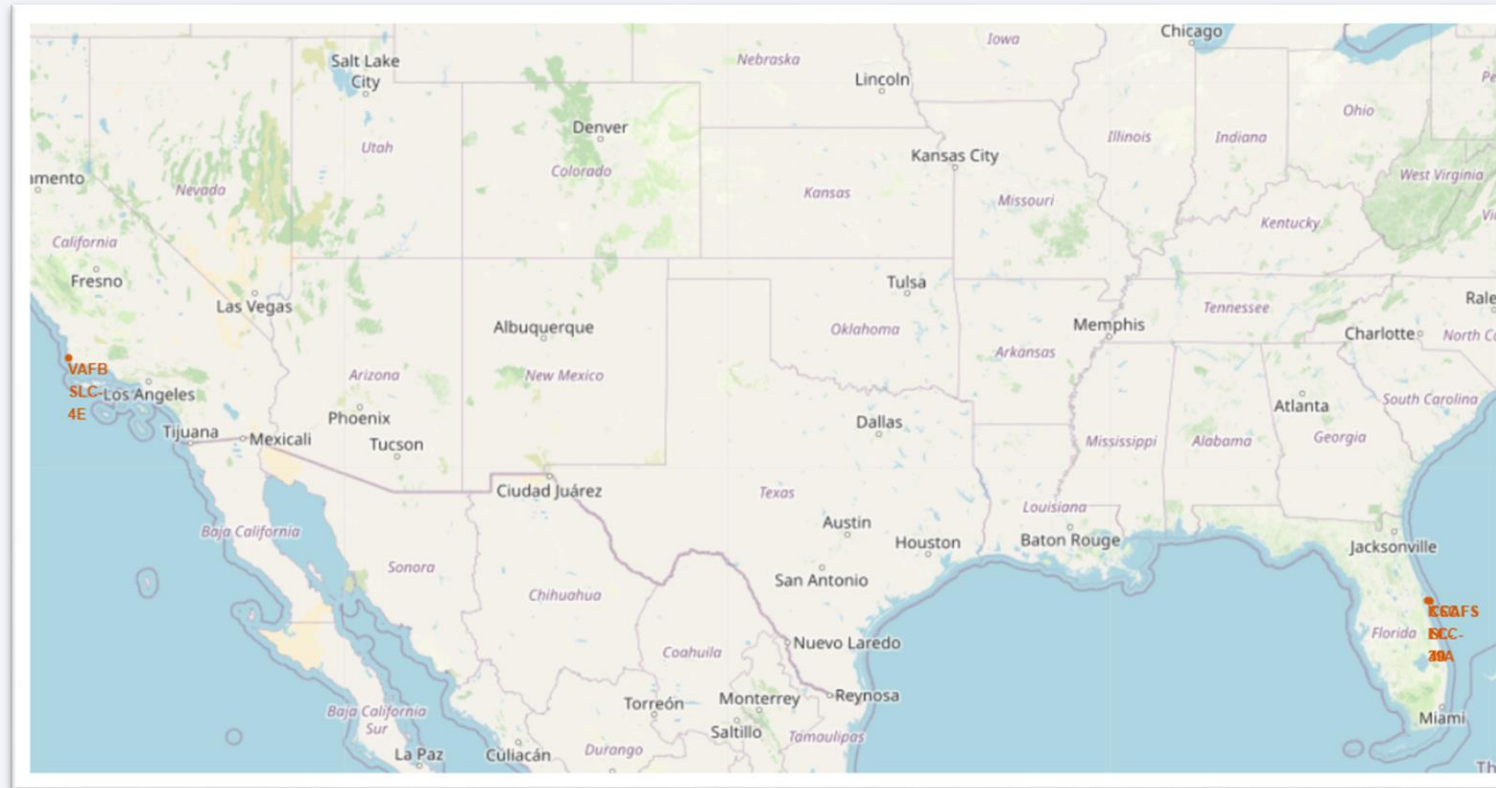
La imagen nos indica que entre el 04-06-2010 y 20-03-2017 se han producido un total de 5 fallos en cohetes no tripulados y 3 exitos en cohetes Plataforma de tierra .

A satellite view of Earth at night, showing the curvature of the planet and numerous city lights glowing against the dark blue background of the night sky. The lights are concentrated in coastal areas and major urban centers, creating a pattern of bright yellow and orange spots across the dark blue surface of the Earth.

Análisis de proximidades de sitios de lanzamiento

Launch Sites Proximities Analysis

Ubicaciones de los sitios de lanzamiento

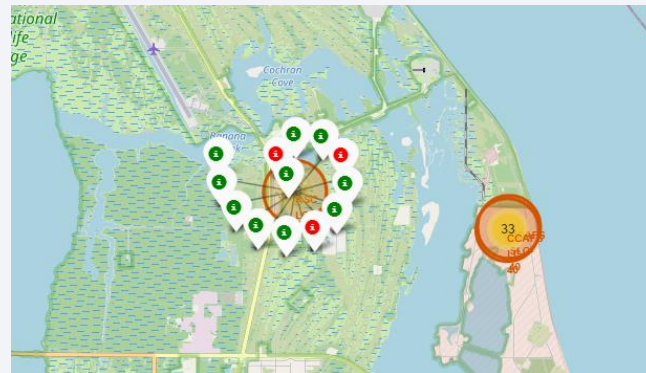


En la imagen aparecen indicados, mediante tres marcadores "Circle", las ubicaciones de los lugares donde tienen lugar los lanzamientos

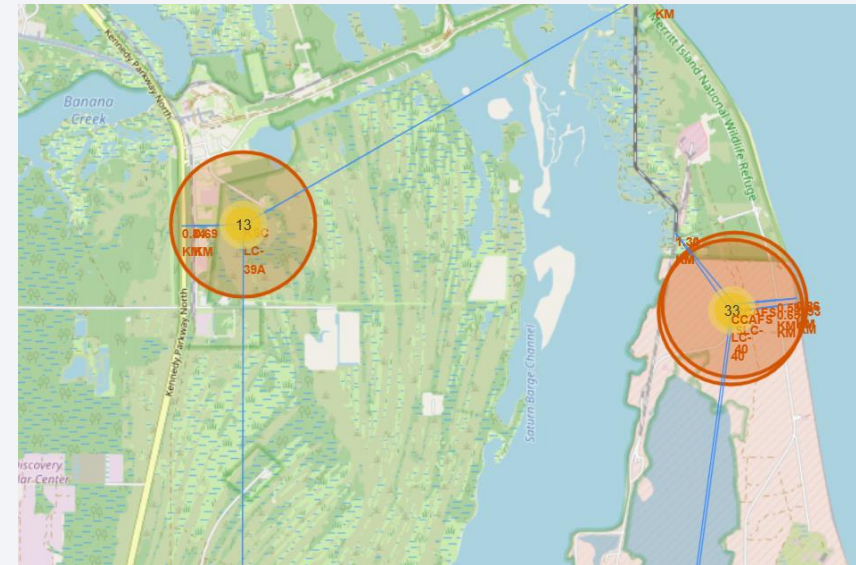
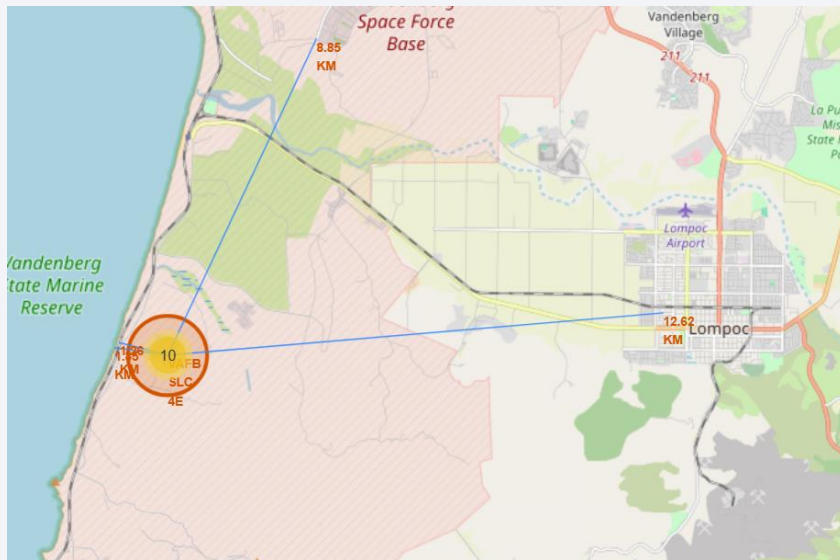
Ubicación de los lanzamientos exitosos/fallidos



El mapa de la imagen superior nos indica el número de lanzamientos de cada uno de los sitios donde sean producido. Mientras que en las dos imagenes inferiores nos indica, mediante iconos de color verde, los lanzamientos que sean realizado con éxito, y rojos, los fallidos.



Distancias entre un sitio de lanzamiento y sus proximidades



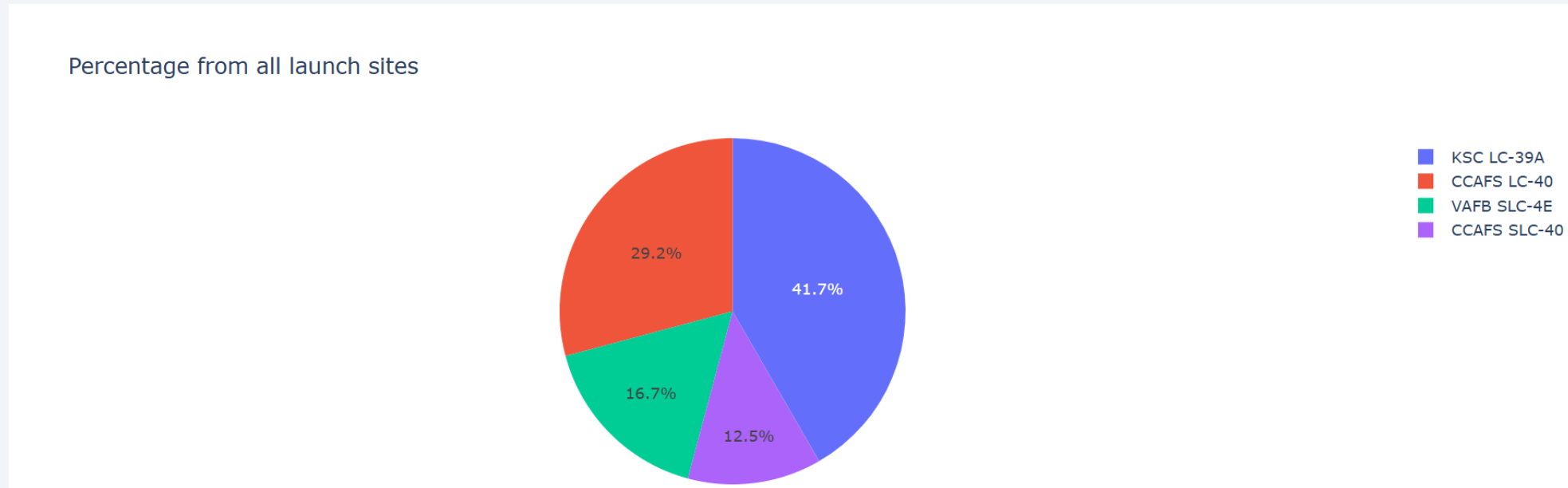
Las dos imágenes nos muestran la distancia entre un sitio de lanzamiento y sus proximidades. Así podemos ver, en la imagen de la izquierda , el sitio de lanzamiento **VAFB SLC-4E** está a 12,62 kilomentros de la ciudad de Lompoc, y a 8,85 kilomentros de la Vandenberg Space Force Base.



Crear un panel de control con Plotly Dash

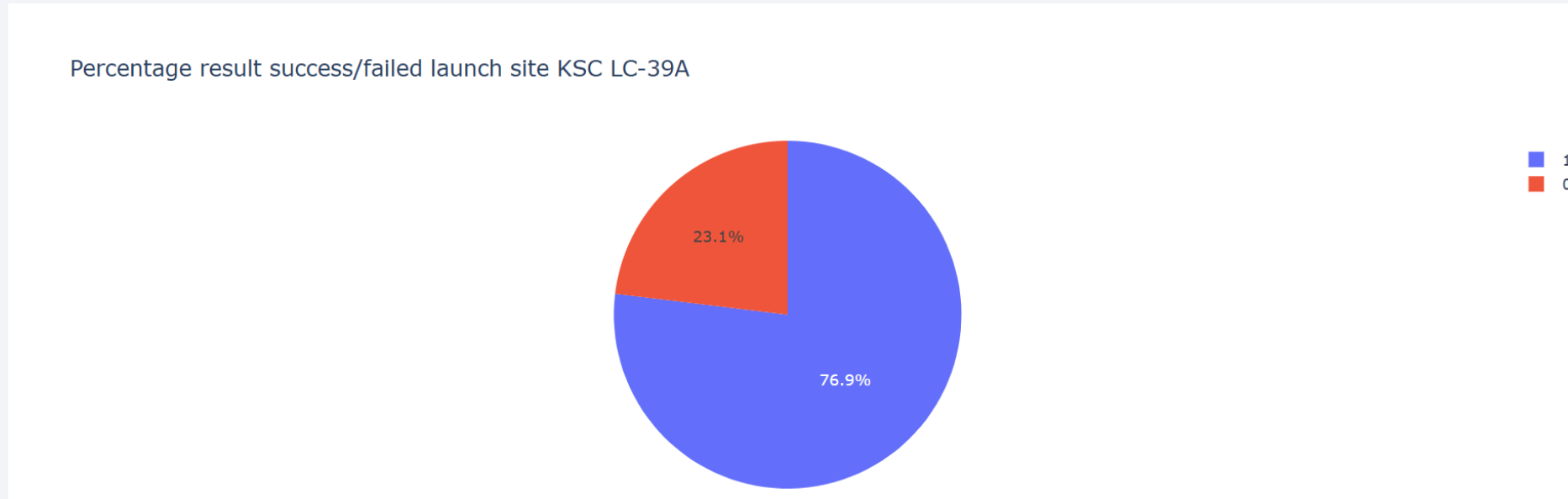
Build a Dashboard with Plotly Dash

Porcentaje de todos los sitios de lanzamiento



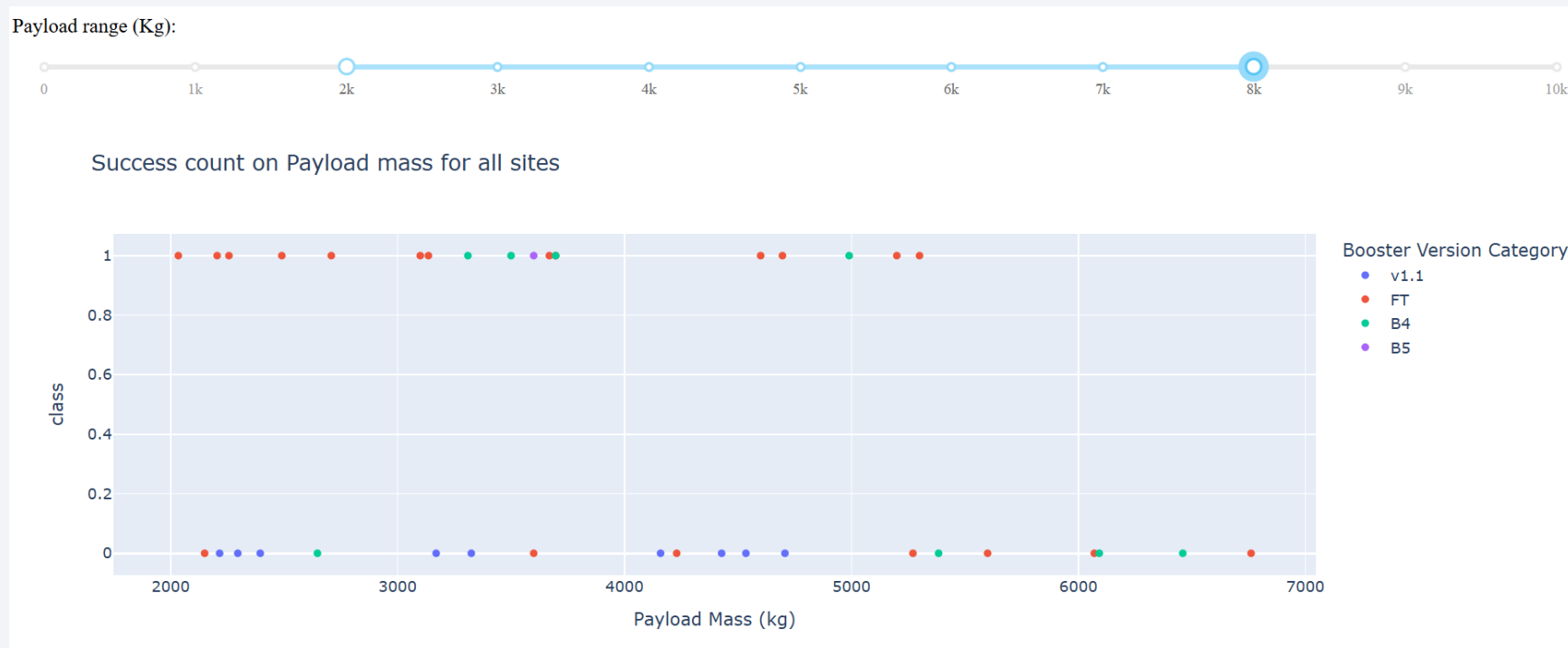
Este gráfico circular nos muestra el porcentaje de lanzamientos por cada uno de los sitios donde sean efectuado, siendo “KSC LC-39A” el lugar donde se produjo un 41,7% de los lanzamientos.

Porcentaje de resultado exitoso/fallido en KSC LC-39A



Este gráfico circular nos muestra el porcentaje de resultado de los lanzamientos en KSC LC-39. En este caso, vemos que el porcentaje de éxito ha sido de un 76,90 %

Recuento de éxitos con una masa de carga útil entre 2000 y 7000 kg para todos los sitios de lanzamientos



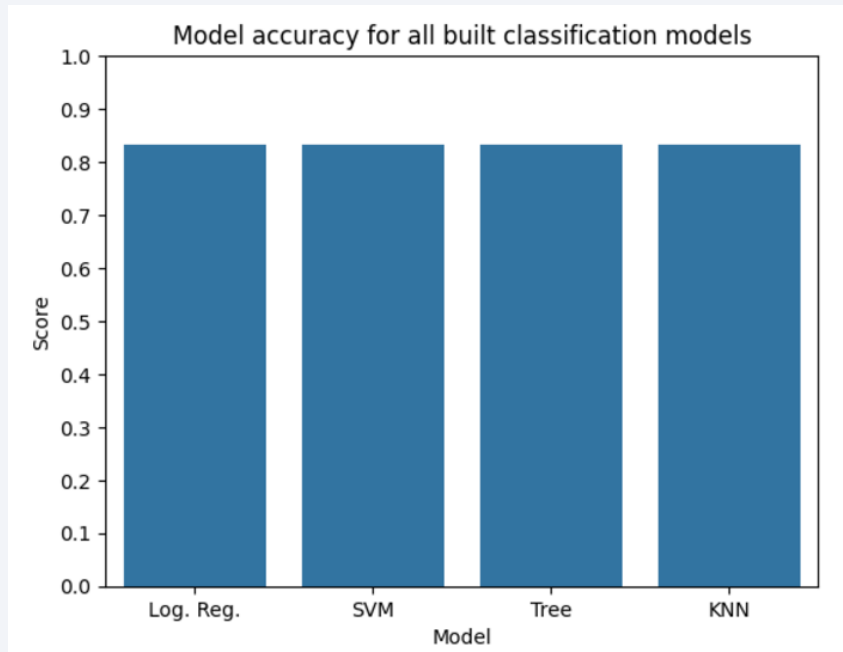
Este diagrama de dispersión nos muestra que, en el caso de un rango de carga útil entre 2000 kg y 7000 kg, la versión del Falcon 9 que más taxa de éxito han tenido ha sido FT.



Análisis Predictivo (Clasificación)

Predictive Analysis (Classification)

Classification Accuracy



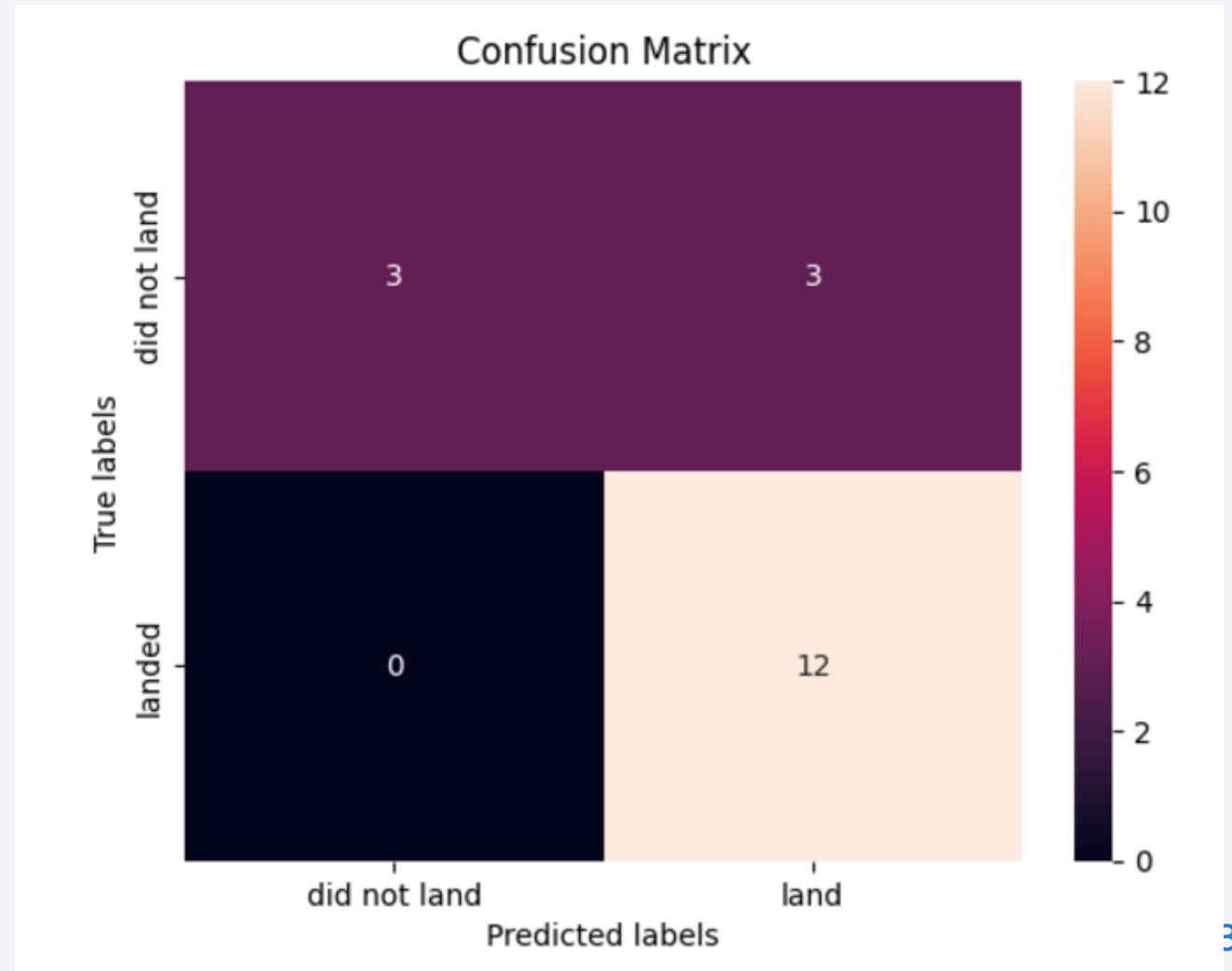
[34]:

	ML Method	Accuracy Score
0	Support Vector Machine (SVM)	0.833333
1	Logistic Regression	0.833333
2	K Nearest Neighbour (KNN)	0.833333
3	Decision Tree	0.833333

En la gráfica se puede ver que todos los métodos tienen una puntuación de precisión idéntica de 0,83333, por lo que decidimos usar la regresión logística para la clasificación.

Confusion Matrix

El gráfico muestra la matriz de confusión del modelo de Regresión Logística que se eligió. El modelo solo falló en predecir con precisión 3 etiquetas.



Conclusiones

Conforme a las métricas que hemos utilizado durante nuestro análisis podemos extraer los siguientes datos:

- De acuerdo con los mapas que nos ha proporcionado la librería de Folium, SpaceX tiene tres sitios de lanzamiento, siendo los que realizan mayor número de lanzamiento los sitiados en la costa de Florida:
 - KSC LC-39A con 13, de los que 10 fueron un éxito.
 - CCAFS SLC-40 con 33, de los que solo 10 fueron, también, un éxito.
- Aplicando la metodología EDA y viendo algunos diagramas de dispersión, podemos decir, por una parte, que, en la relación entre la carga útil masa de carga útil y el resultado de lanzamiento, comprobamos que en CCAFS SLC -40 tuvieron más lanzamientos exitosos con una masa de carga útil de entre 2000 y 7000 kg, y, por otra, que existe una relación entre los distintos tipos de órbitas y la masa de carga útil en los resultados de los lanzamientos efectuados, siendo en este caso, los lanzamientos hacia la órbita GTO, ubicada a 35.786 kilómetros sobre el ecuador de la tierra, donde se ha registrado mayor número de resultados positivos
- Conforme a la información que nos ha proporciona el gráfico de dispersión realizados con Space_dahs, dentro de las distintas versiones del Falcon 9 podemos ver que:
 - Aplicando todos los rangos de masa de carga útil la versión F1 es la que más porcentaje de éxito ha tenido, mientras que la versión v1.1 es la que más fracasos ha sufrido.
 - Aplicando un rango de entre 2000 y 10000 kg, la versión v1.1 sigue teniendo más fracasos
- Por último, en cuanto al análisis predictivo, el resultado es el mismo en los 4 tipos de clasificación que hemos realizado por los [44](#) que hemos elegido el primero de ellos, la regresión logística.

Anexo

- Enlaces Sección 1:

- **RECOPIACIÓN DE DATOS :**

- <https://view.genially.com/674effaa20856261e6a9b53e/horizontal-infographic-diagrams-diagrama-de-recopilacion-de-datos>

- **DATA WRANGLING :**

- <https://view.genially.com/674f702112a0badcdbc60749/horizontal-infographic-diagrams-data-wrangling>

- **ANÁLISIS PREDICTIVO (CLASIFICACIÓN):**

- <https://view.genially.com/67507fb1522fcca000b7b3ce/horizontal-infographic-diagrams-analisis-predictivo>

Thank you!

