



IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Jaime Antonio Escudero Holguin
05 de marzo de 2023



Outline

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

Executive Summary

- En este proyecto final predeciremos si SpaceX Falcon 9 aterrizará de manera exitosa. Con esto podemos determinar el costo de los lanzamientos.
- Las etapas del proyecto son:
 - Data Collection
 - Data Wrangling and Preprocessing,
 - Exploratory Data Analysis,
 - Data Visualization,
 - Machine Learning Prediction.
- Se determina correlación entre los datos.
- el mejor algoritmo para este problema es el árbol de decisiones

Introduction

- El objetivo es predecir si Falcon 9 aterrizará con éxito. SpaceX se propone utilizar cohete varias veces, con una costo final totalmente reducido. El ahorros se deben a la reutilización sus cohetes. Si podemos determinar si la primera etapa aterrizará, podemos determinar el costo de un lanzamiento. Esta información se puede usar para determinar la viabilidad del proyecto y cuantos clientes potenciales tendría.
- La pregunta principal que estamos tratando de responder: para un conjunto dado de características sobre el lanzamiento de los cohetes:
¿el cohete aterrizará con éxito?

Section 1

Methodology

Methodology

Executive Summary (1)

- Data collection methodology:

Los datos se recopilaron a través de dos métodos: solicitar datos de la API de SpaceX y datos de lanzamiento de web scraping de una página de Wikipedia.

- Perform data wrangling

Se realizó una disputa de datos para transformar y limpiar los datos mediante la biblioteca pandas de Python.

- Perform exploratory data analysis (EDA) using visualization and SQL

Con los datos limpios, se realizó un análisis exploratorio de datos utilizando herramientas de visualización como matplotlib de Python y las bibliotecas seaborn, además de responder preguntas mediante consultas SQL.

Methodology

Executive Summary (2)

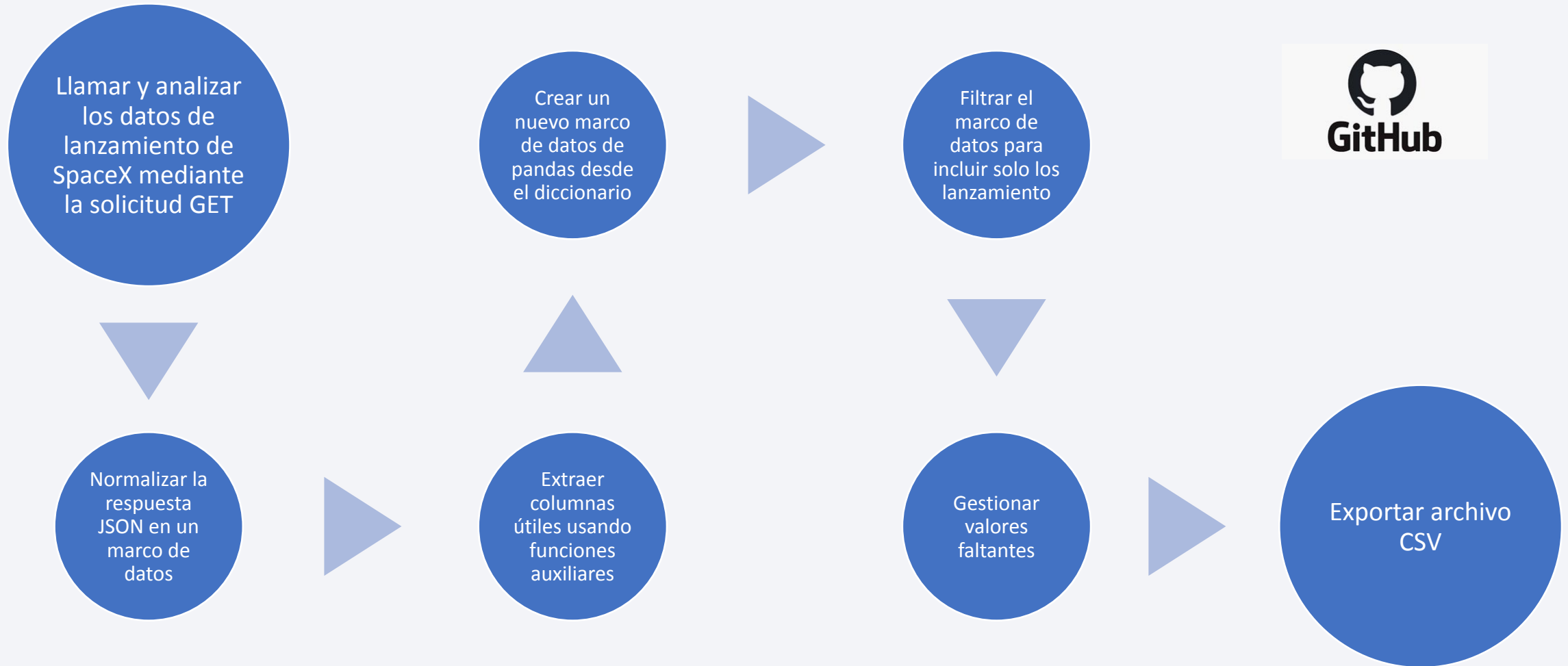
- Perform interactive visual analytics using Folium and Plotly Dash

Los paquetes de visualización interactivos de Python se utilizaron para responder algunas preguntas analíticas. Folium se usó para crear mapas, mientras que Plotly Dash se usó para crear visualizaciones de datos interactivas.

- Perform predictive analysis using classification models

Se utilizaron cuatro modelos diferentes de clasificación de aprendizaje automático para el análisis predictivo. Los modelos que se utilizaron son regresión logística, máquinas de vectores de soporte, k-vecino más cercano y clasificador de árboles de decisión. Cada modelo fue entrenado, ajustado y evaluado para encontrar el mejor.

Data Collection



Data Collection - Scraping



	Flight No.	Launch site	Payload	Payload mass	Orbit	Customer	Launch outcome	Version Booster	Booster landing	Date	Time
0	1	CCAFS	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success\n	F9 v1.0B0003.1	Failure	4 June 2010	18:45
1	2	CCAFS	Dragon	0	LEO	NASA	Success	F9 v1.0B0004.1	Failure	8 December 2010	15:43
2	3	CCAFS	Dragon	525 kg	LEO	NASA	Success	F9 v1.0B0005.1	No attempt\n	22 May 2012	07:44
3	4	CCAFS	SpaceX CRS-1	4,700 kg	LEO	NASA	Success\n	F9 v1.0B0006.1	No attempt	8 October 2012	00:35
4	5	CCAFS	SpaceX CRS-2	4,877 kg	LEO	NASA	Success\n	F9 v1.0B0007.1	No attempt\n	1 March 2013	15:10

Solicite datos de lanzamiento de cohetes desde Wikipedia

Extraiga todos los nombres de columna del encabezado de la tabla HTML

Cree un marco de datos analizando las tablas HTML de lanzamiento

Exportar archivo CSV

Data Wrangling

Calcular el número de lanzamientos
en cada sitio

Calcular el número y ocurrencia de cada órbita

Calcular el número y la ocurrencia del
resultado de la misión por tipo de órbita

Crear una etiqueta de resultado
de aterrizaje a partir de la
columna Resultado utilizando la
codificación one-hot

Exportar a CSV



```
dt.head(5)
```

Number	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flight
1	2010-06-04	Falcon 9	6104.959412	LEO	CCAFS SLC 40	None	None
2	2012-05-22	Falcon 9	525.000000	LEO	CCAFS SLC 40	None	None
3	2013-03-01	Falcon 9	677.000000	ISS	CCAFS SLC 40	None	None
4	2013-09-29	Falcon 9	500.000000	PO	VAFB SLC 4E	False	Ocean
5	2013-12-03	Falcon 9	3170.000000	GTO	CCAFS SLC 40	None	None

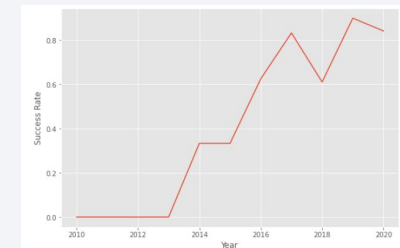
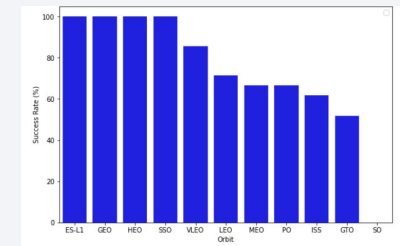
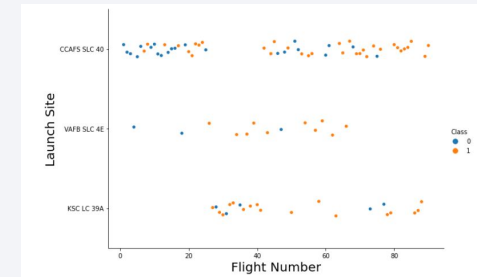
We can use the following line of code to determine the success rate:

```
df["Class"].mean()
```

```
1.0
```

EDA with Data Visualization

- Scatter plots: Se utilizaron gráficos de dispersión para representar la relación entre dos variables. Se compararon diferentes conjuntos de funciones, como el número de vuelo frente al sitio de lanzamiento, la carga útil frente al sitio de lanzamiento, el número de vuelo frente al tipo de órbita y la carga útil frente al tipo de órbita
- Bar chart: los gráficos de barras que se utilizaron facilitan la comparación de valores entre varios grupos de un vistazo. El eje x representa una categoría y el eje y representa un valor discreto. Se usaron gráficos de barras para comparar la tasa de éxito para diferentes tipos de órbita
- Line chart: los gráficos de líneas son útiles para mostrar tendencias de datos a lo largo del tiempo. Se usó un gráfico de líneas para mostrar la tasa de éxito durante un cierto número de años.



EDA with SQL

A continuación se incluye una lista de algunas de las consultas SQL realizadas en el conjunto de datos:
Visualización de los nombres de los sitios de lanzamiento únicos en la misión espacial

- Visualización de 5 registros donde los sitios de lanzamiento comienzan con la cadena 'CCA'
- Visualización de la masa de carga útil total transportada por los propulsores lanzados por la NASA (CRS)
- Visualización del promedio masa de carga útil transportada por la versión de refuerzo F9 v1.1, Enumerar la fecha en que se logró el primer resultado de aterrizaje exitoso en la plataforma de tierra
- Enumerar los nombres de los propulsores que tienen éxito en la nave de drones y tienen una masa de carga superior a 4000 pero inferior a 6000
- Enumerar el número total de éxito y resultados de la misión fallida
- Enumerar los nombres de las versiones de refuerzo que han transportado la masa máxima de carga útil
- Enumerar los resultados de aterrizaje fallidos en naves no tripuladas, sus versiones de refuerzo y los nombres de los sitios de lanzamiento en el año 2015
- Clasificación del conteo de resultados de aterrizaje entre la fecha 2010-06-04 y 2017-03-20, en orden descendente.



Build an Interactive Map with Folium

- Se crearon objetos y se agregaron a un mapa de Folium.
- Se utilizaron objetos de marcador para mostrar todos los sitios de lanzamiento en un mapa, así como los lanzamientos exitosos/fallidos para cada sitio en el mapa.
- Los objetos de línea se utilizaron para calcular las distancias entre un sitio de lanzamiento y sus proximidades.
- Al agregar estos objetos, se encuentran los siguientes patrones geográficos sobre los sitios de lanzamiento positivos:
 - Cerca de los ferrocarriles
 - Cerca de las carreteras
 - Cerca de la costa
 - A cierta distancia de las ciudades



Build a Dashboard with Plotly Dash

- La aplicación de tablero contiene dos gráficos:
- Un gráfico circular que muestra el lanzamiento exitoso de cada sitio. Este gráfico es útil porque puede visualizar la distribución de los resultados de aterrizaje en todos los sitios de lanzamiento o mostrar la tasa de éxito de los lanzamientos en sitios individuales.
- Un gráfico de dispersión que muestra la relación entre los resultados del aterrizaje y la masa de carga útil de diferentes propulsores. El tablero toma dos entradas, a saber, el sitio (s) y la masa de carga útil. Este gráfico es útil ya que puede visualizar cómo las diferentes variables afectan los resultados del aterrizaje,



Predictive Analysis (Classification)

Crear columna para "Class"

Estandarizando los datos

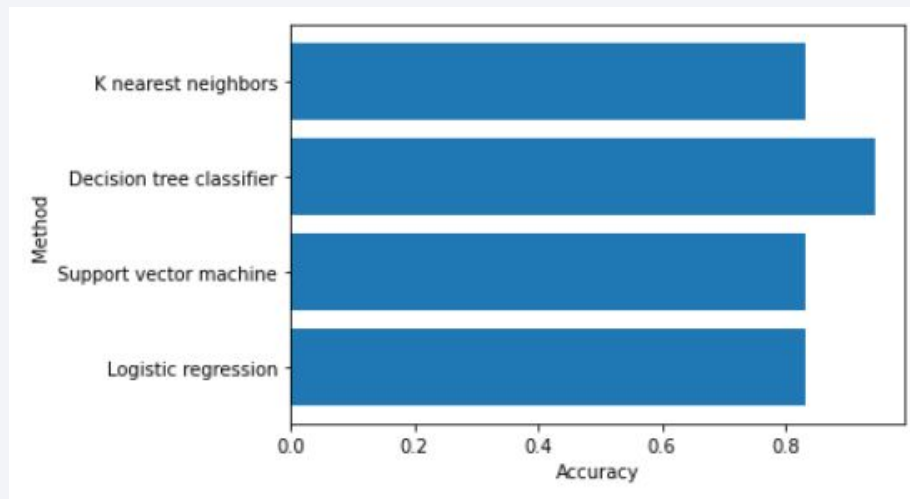
Encuentre el mejor hiperparámetro para SVM, árboles de decisión, K-Nearest Neighbours y regresión logística.

Use datos de prueba para evaluar modelos en función de sus puntajes de precisión y matriz de confusión



Results

- Los resultados del análisis de datos exploratorios revelaron que la tasa de éxito de los aterrizajes del Falcon 9 fue del 66,66 %
- Los resultados del análisis predictivo mostraron que el algoritmo del árbol de decisiones fue el mejor método de clasificación con una precisión del 94 %



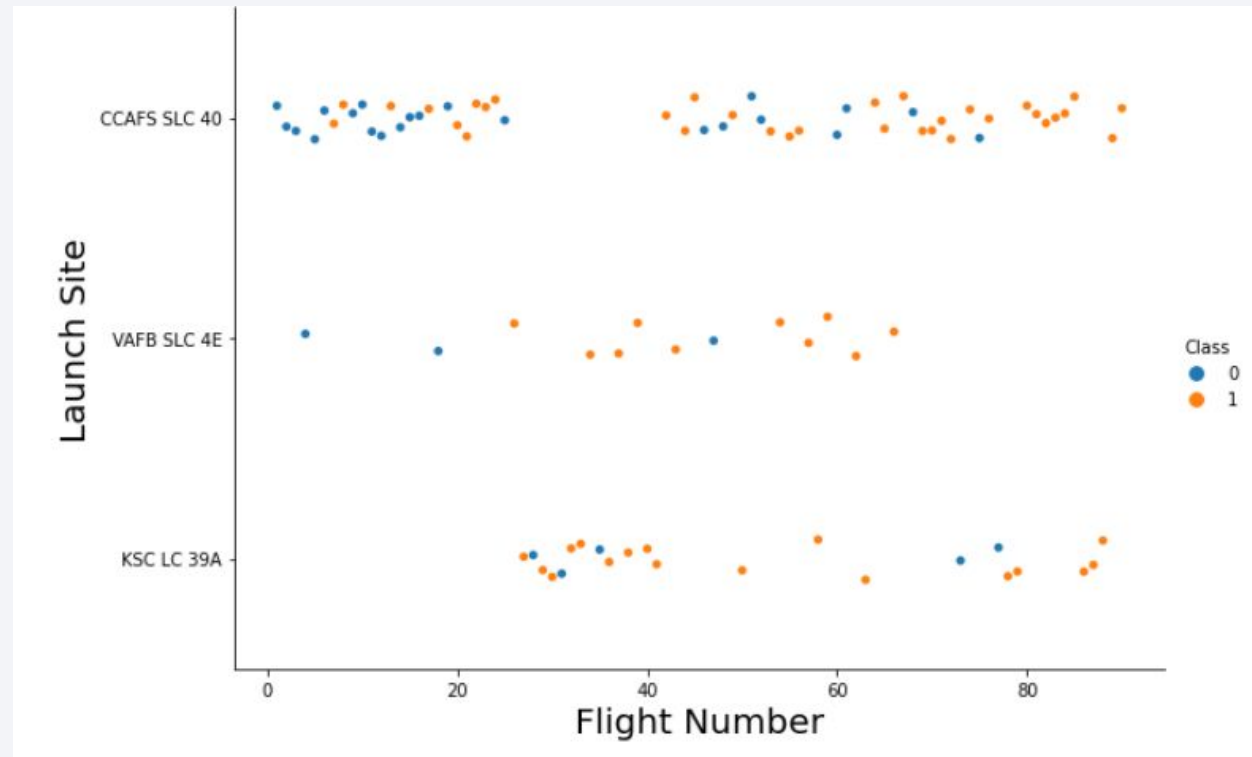
The background of the slide is an abstract composition. It features a solid blue area on the left side, which transitions into a complex pattern of diagonal streaks in shades of blue, red, and cyan on the right. These streaks have a textured, almost woven appearance. Overlaid on this pattern is a faint, light blue grid that recedes into the distance, creating a sense of depth and perspective.

Section 2

Insights drawn from EDA

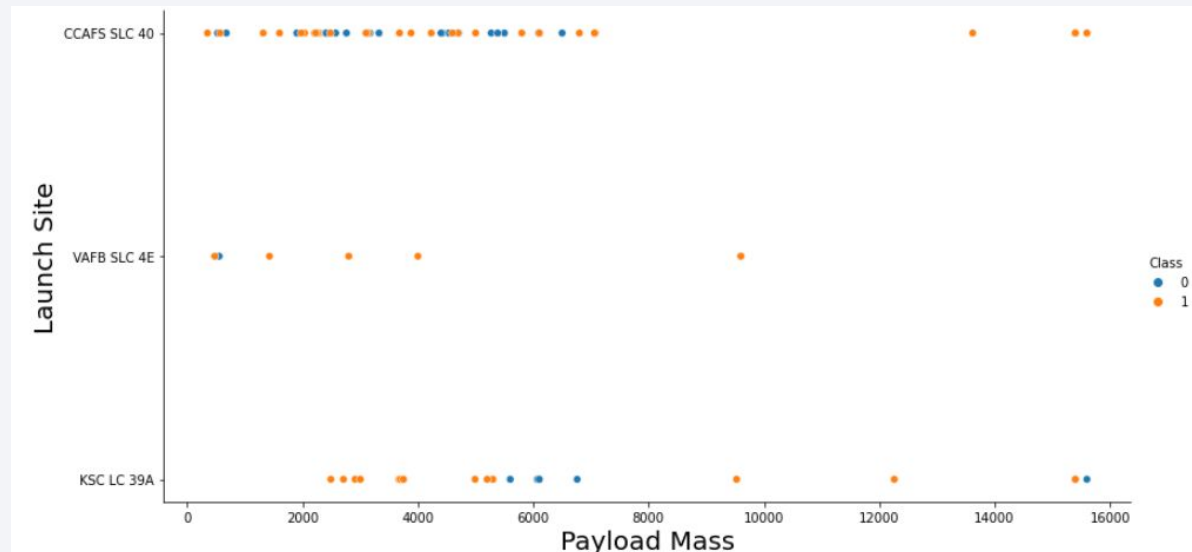
Flight Number vs. Launch Site

- Esta figura muestra que la tasa de éxito de 1 al 20 como prueba y experiencia.
- el lugar de lanzamiento con mayor tasa de éxito es KSC LC-39A a partir del vuelo 40.
- Parece haber un aumento en los vuelos exitosos después del lanzamiento número 40.
- El mayor número de lanzamientos no exitosos se realizó en CCAFS LC-40



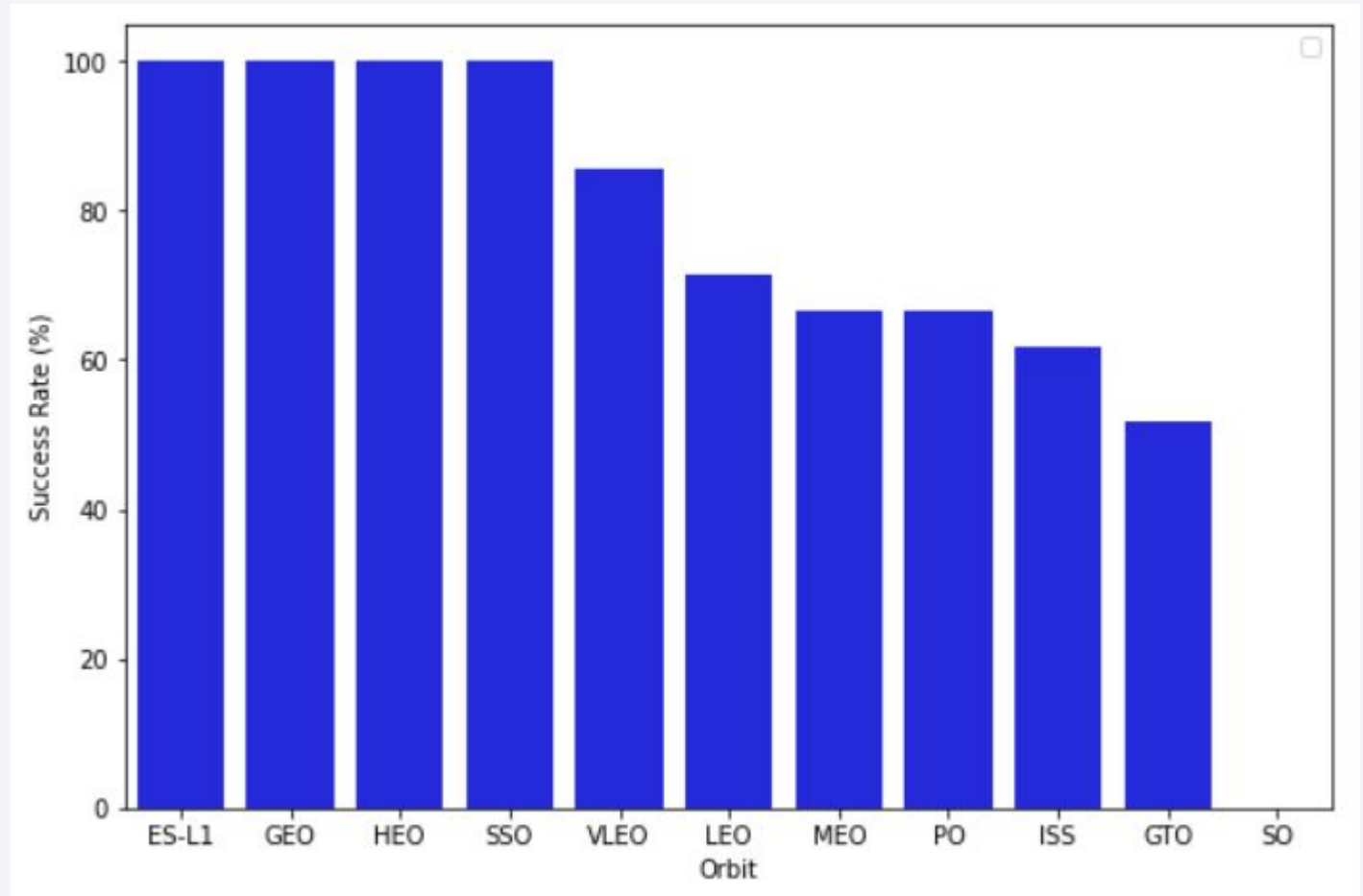
Payload vs. Launch Site

- Para el sitio de lanzamiento VAFB-SLC no se lanzan cohetes con carga pesadas
- No presenta correlación entre la carga y el sitio de lanzamiento.
- El lugar de lanzamiento KSC LC-39A tiene mejor tasa de éxito hasta los 5000 unidades de carga



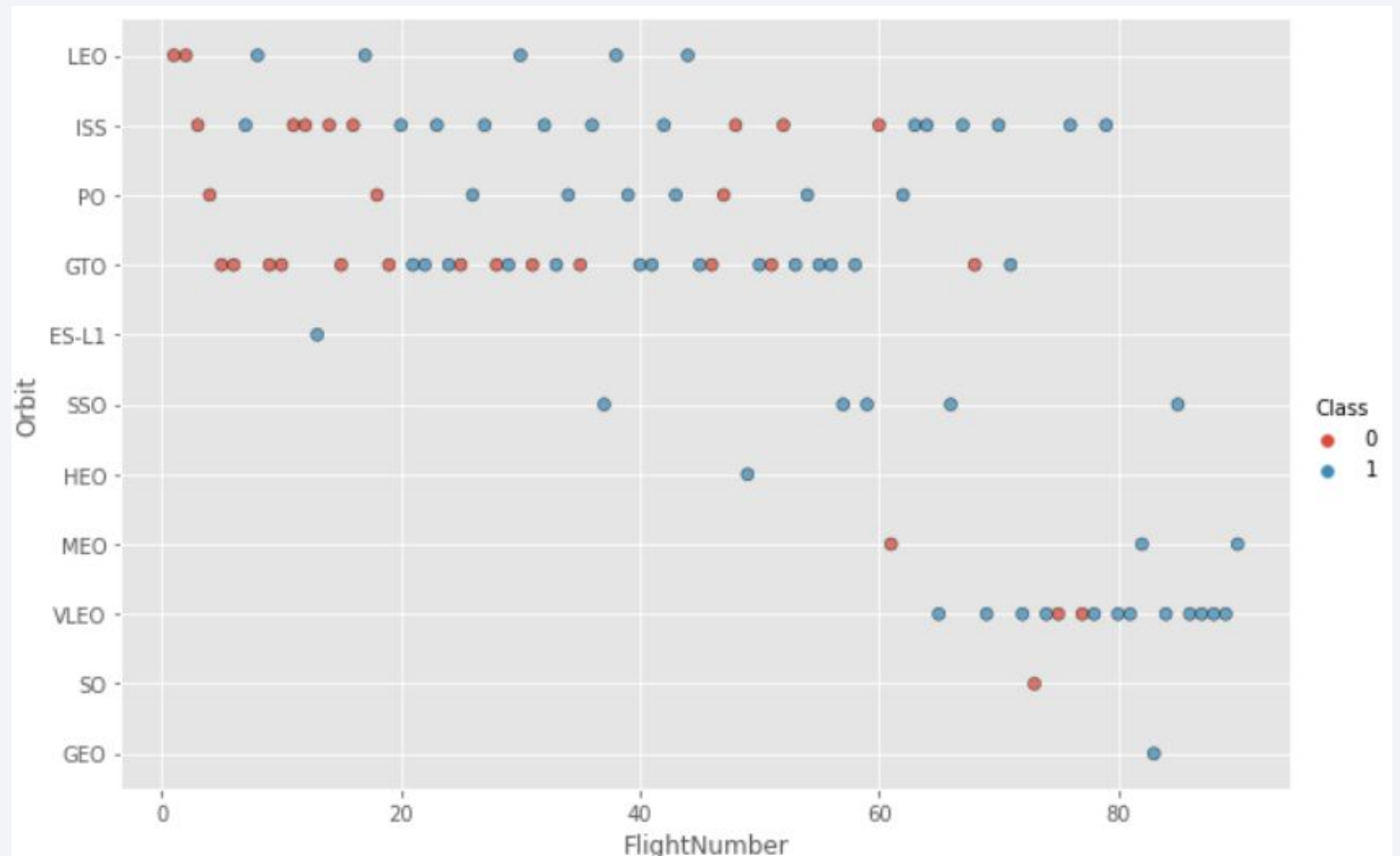
Success Rate vs. Orbit Type

- La órbita SO no tiene datos exitosos.
- Las 4 primeras órbitas tienen una tasa de éxito del 100%
- Para mayor inferencia se debe contrastar con el número de registros para cada órbita.



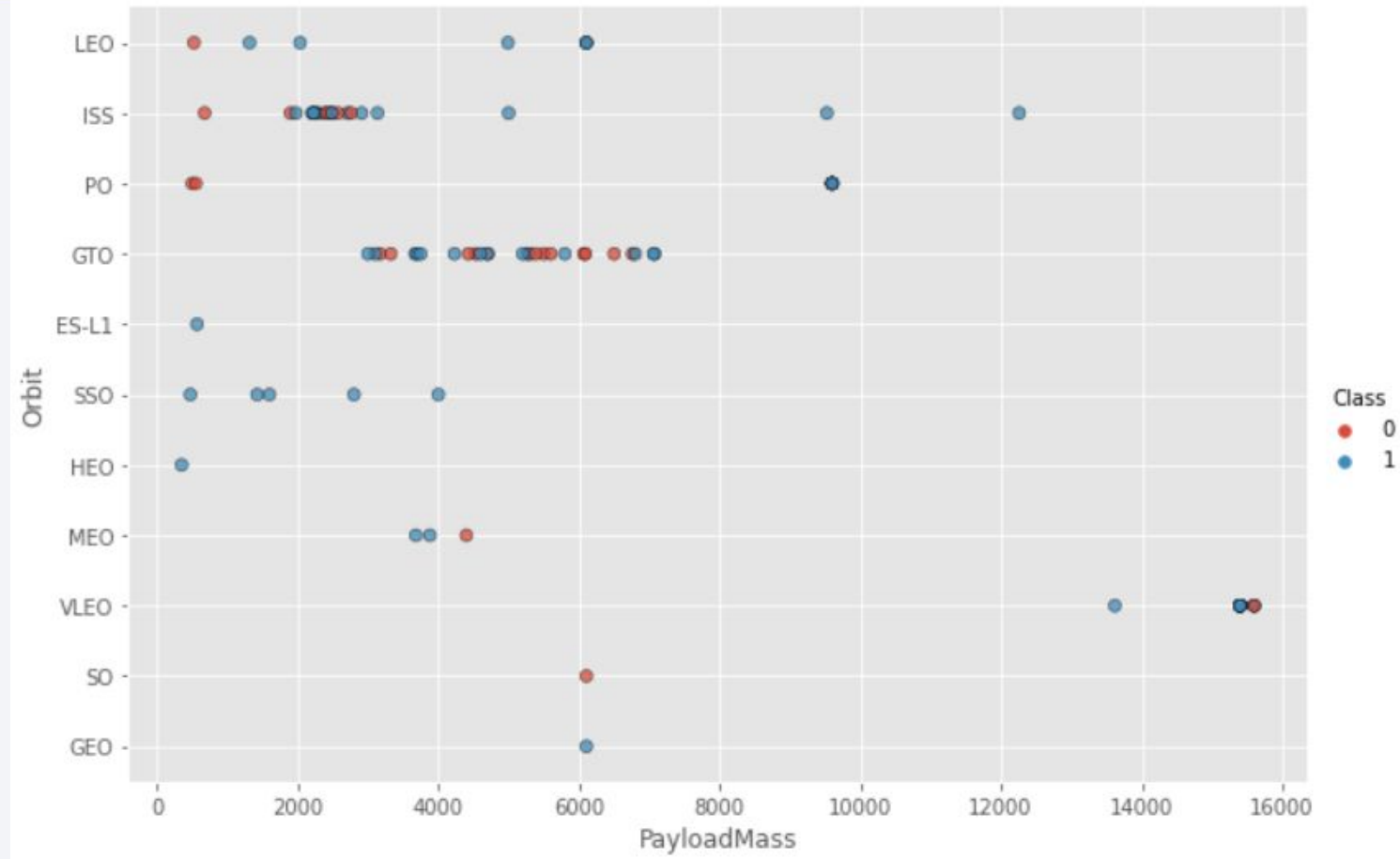
Flight Number vs. Orbit Type

- En la órbita LEO, el éxito se correlaciona positivamente con el número de vuelos.
- la órbita ISS tiene mayor tasa de éxito teniendo en cuenta el número de registros
- La órbita SSO tiene una tasa de éxito del 100%, sin embargo, con menos vuelos que las otras órbitas.
- no existen fracasos después de vuelo 80
-



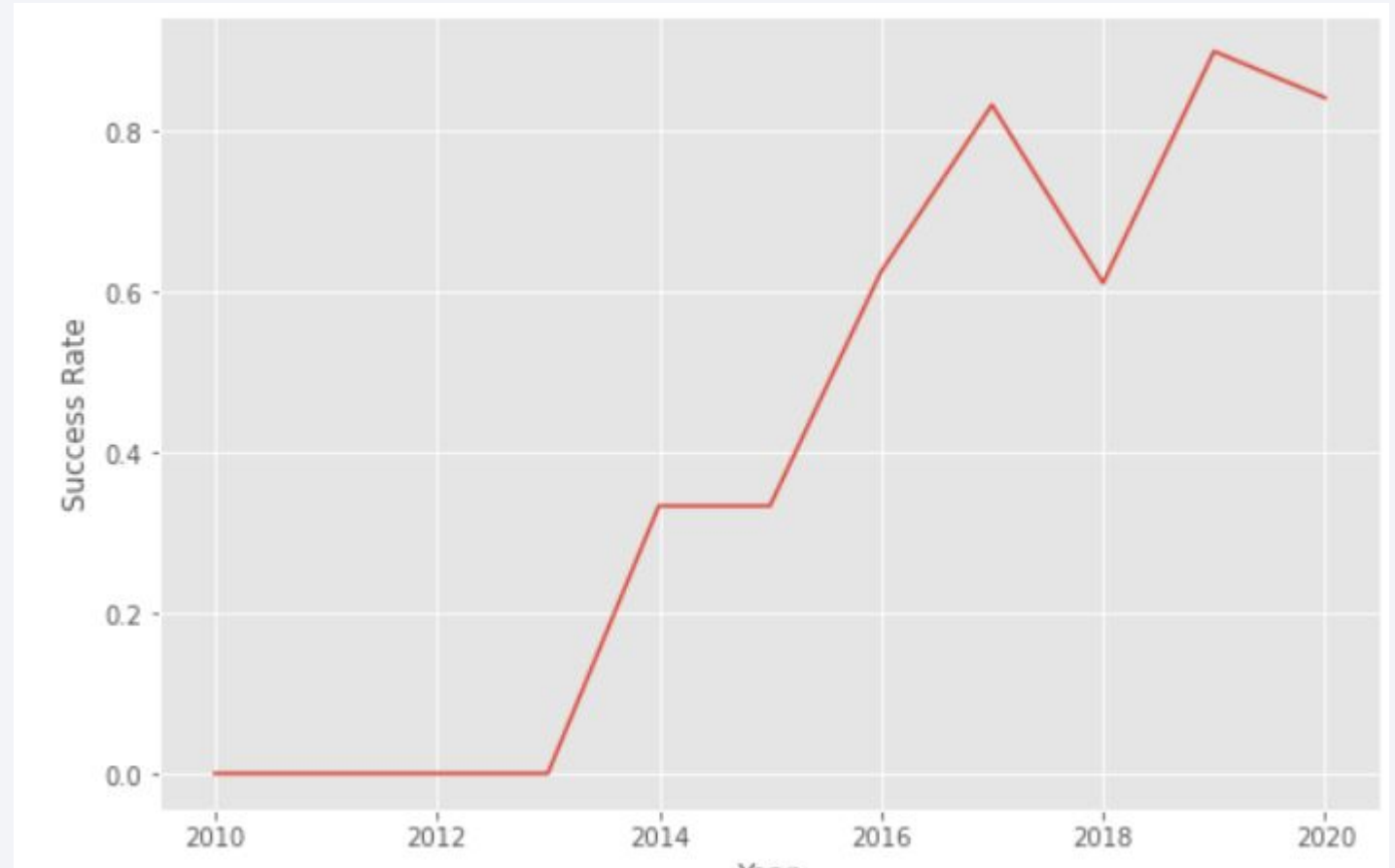
Payload vs. Orbit Type

- PO, LEO e ISS tienen mayor tasa de éxito en cargas mayores a 4000.
- GTO registra ma mayor cantidad de fracasos
- El único fracaso en carga mayor a 8000 está en la órbita VLEO



Launch Success Yearly Trend

- La proyección de éxito mejora con cada año a partir del 2013.
- Se deben verificar las causas del descenso del año 2018.
-



All Launch Site Names

- Se usó la función DISTINCT para la columna Launch_Site.



```
%sql Select distinct Launch_Site From SPACEXTBL
```



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

```
Done.
```

```
Launch_Site
```

```
CCAFS LC-40
```

```
VAFB SLC-4E
```

```
KSC LC-39A
```

```
CCAFS SLC-40
```

Launch Site Names Begin with 'CCA'

- Usamos la función LIKE, LIMIT

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

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS_KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	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)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- Usamos el prefijo SUM y el condicional WHERE

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE CUSTOMER='NASA (CRS)'
```



```
* sqlite:///my_data1.db  
Done.  
SUM(PAYLOAD_MASS__KG_)  
45596
```

Average Payload Mass by F9 v1.1

- Se uso AVG, y el condicional WHERE para columna BOOSTER_VERSION

```
%sql SELECT BOOSTER_VERSION, AVG(PAYLOAD_MASS_KG_) FROM SPACEXTBL WHERE BOOSTER_VERSION='F9 v1.1'
```

* sqlite:///my_data1.db
Done.

Booster_Version	AVG(PAYLOAD_MASS_KG_)
F9 v1.1	2928.4

Booster_Version	PAYLOAD_MASS_KG_
F9 v1.1	3170
F9 v1.1	3325
F9 v1.1	2296
F9 v1.1	1316
F9 v1.1	4535

First Successful Ground Landing Date

- Usamos la función .query y .min
- Se obtiene como resultado 22-12-2015

```
%sql Select "Landing _Outcome", (date), min(substr(Date,7,4)) From SPACEXTBL where "Landing _Outcome"='Success (ground pad)'
```

```
↳ * sqlite:///my_data1.db  
Done.  
Landing _Outcome  Date  min(substr(Date,7,4))  
Success (ground pad) 22-12-2015 2015
```


Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

Usamos varias funciones para hacer diferentes data frames y filtrar lo requerido.

```
* sqlite:///my_data1.db
Done.
Booster_Version PAYLOAD_MASS_KG_
F9 FT B1022      4696
F9 FT B1026      4600
F9 FT B1021.2    5300
F9 FT B1031.2    5200
```

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes
- usamos la función COUNT y GROUP BY

```
%sql SELECT Mission_Outcome, COUNT(DATE) FROM SPACEXTBL GROUP BY Mission_Outcome
```

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

Boosters Carried Maximum Payload

- Usamos la función max y doble query

```
%sql SELECT BOOSTER_VERSION, PAYLOAD_MASS__KG_ FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL)
```

```
* sqlite:///my_data1.db
Done.
Booster_Version PAYLOAD_MASS__KG_
F9 B5 B1048.4    15600
F9 B5 B1049.4    15600
F9 B5 B1051.3    15600
F9 B5 B1056.4    15600
F9 B5 B1048.5    15600
F9 B5 B1051.4    15600
F9 B5 B1049.5    15600
F9 B5 B1060.2    15600
F9 B5 B1058.3    15600
F9 B5 B1051.6    15600
F9 B5 B1060.3    15600
F9 B5 B1049.7    15600
```

2015 Launch Records

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

se uso la funcion Substr(#,7,4) para el año

```
%sql SELECT date , "Booster_Version", "Launch_Site" FROM SPACEXTBL WHERE substr(Date,7,4)='2015' AND "Landing _Outcome" = 'Failure (drone ship)';
```

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

```
Done.
```

Date	Booster_Version	Launch_Site
10-01-2015	F9 v1.1 B1012	CCAFS LC-40
14-04-2015	F9 v1.1 B1015	CCAFS LC-40

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

- 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

```
%sql SELECT "Landing Outcome", COUNT("Landing Outcome") FROM SPACEXTBL WHERE DATE BETWEEN '04-06-2010'  
AND '20-03-2017' GROUP BY "Landing _Outcome" ORDER BY COUNT("Landing _Outcome") DESC;
```

Landing _Outcome	COUNT("I
Success	20
No attempt	10
Success (drone ship)	8
Success (ground pad)	6
Failure (drone ship)	4
Failure	3
Controlled (ocean)	3
Failure (parachute)	2
No attempt	1

A satellite view of Earth from space, showing the curvature of the planet and city lights at night. The image is a composite of a solid blue background on the left and a satellite photograph of Earth on the right. The Earth's surface is dark, with numerous bright yellow and orange lights representing cities and urban areas. The horizon of the Earth is visible as a thin, curved line separating the dark surface from the deep blue of space.

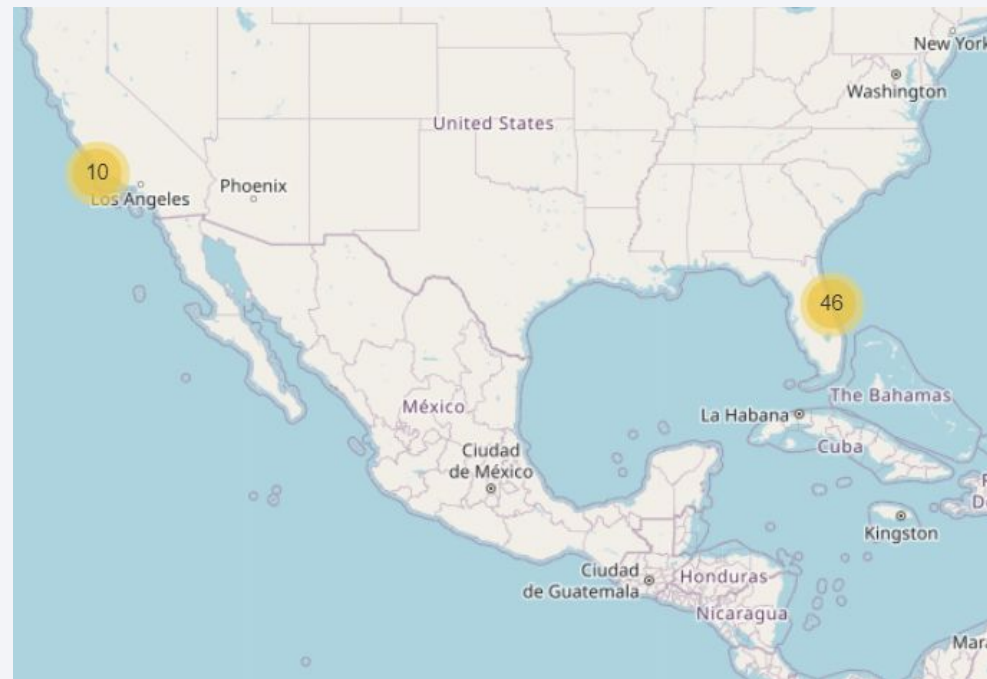
Section 3

Launch Sites Proximities Analysis

Localización de lugares de lanzamiento

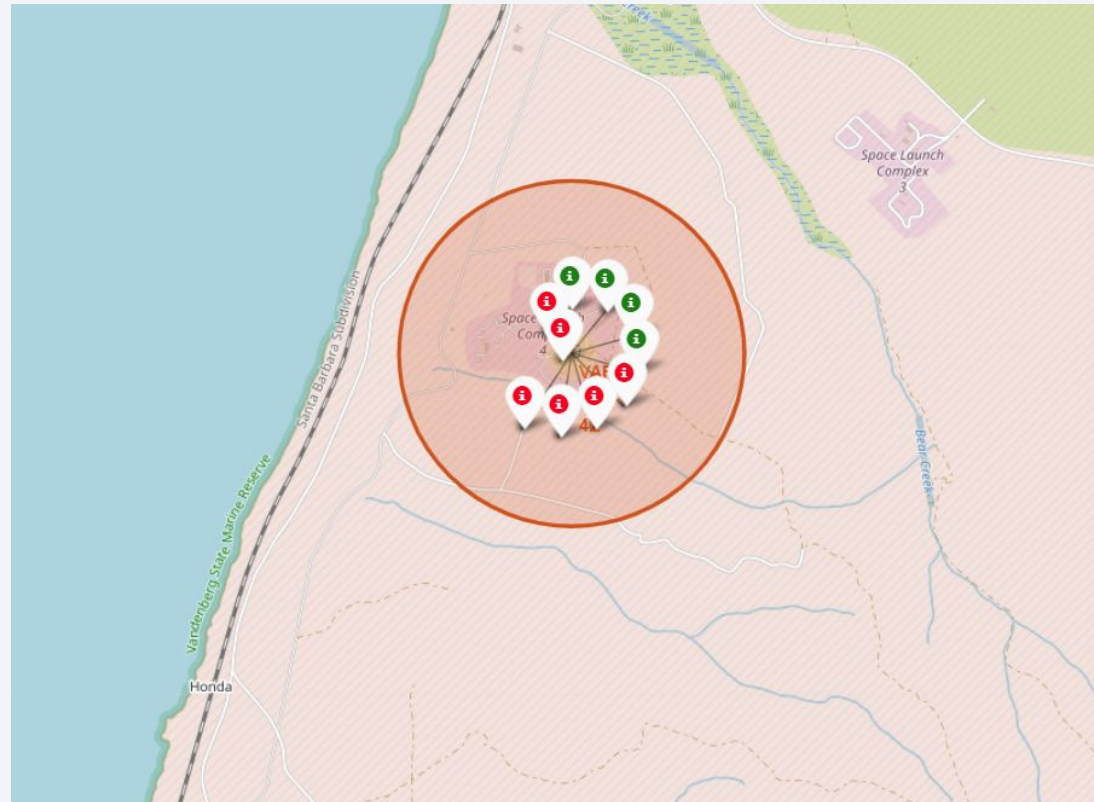
Los marcadores indican la ubicación de los lugares de lanzamiento.

los sitios están dentro de estados unidos y en ambas costas, que significa diversas condiciones climáticas y factores.



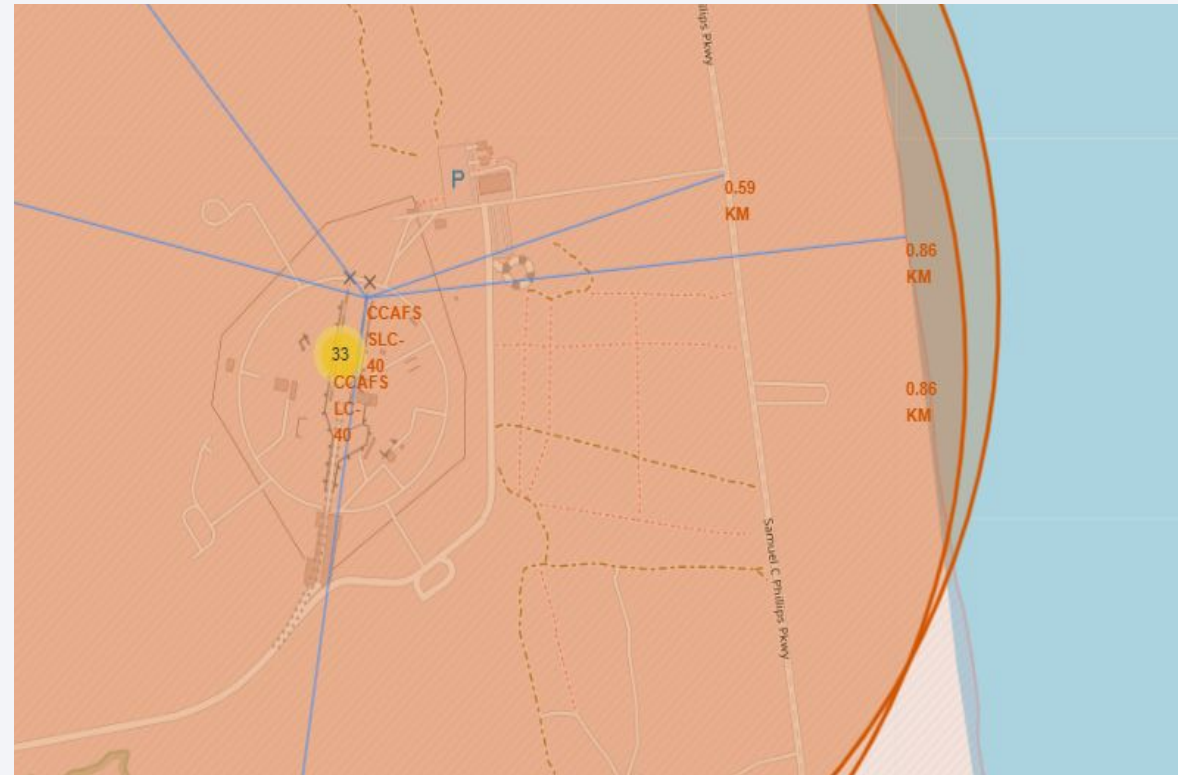
Plataforma costa oeste

- Podemos observar con rojo los lanzamientos fallidos y con verde los exitosos.
- la plataforma



Lugares cercanos al sitio de lanzamiento

- El mapa generado muestra que el sitio de lanzamiento seleccionado está cerca de una carretera para el transporte de personal y equipo.
- El sitio de lanzamiento también está cerca de las costas para realizar pruebas de fallas en el lanzamiento.
- Los sitios de lanzamiento también mantienen cierta distancia de las ciudades.



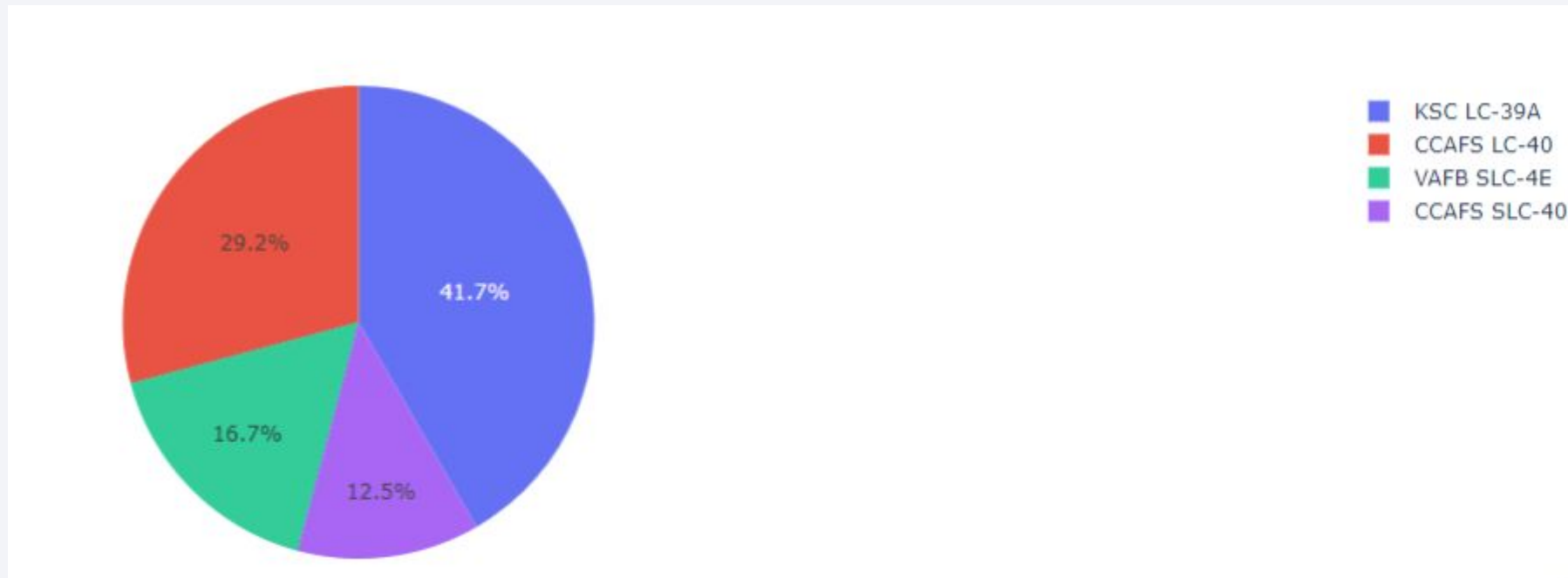


Section 4

Build a Dashboard with Plotly Dash

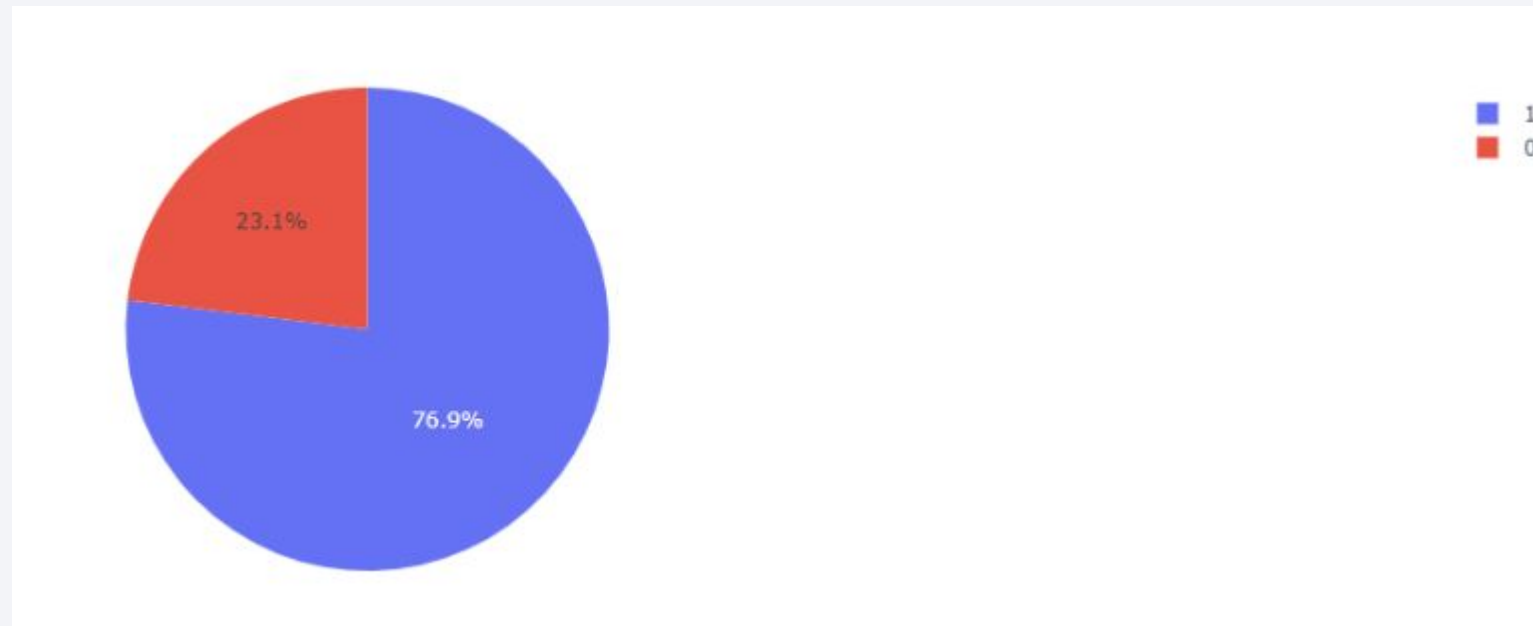
Lanzamientos exitosos por sitio

- Puede ver en la gráfica que el sitio KSC LC-39A tiene mayor cantidad lanzamientos exitosos, así como el mayor tasa de éxito de lanzamiento.



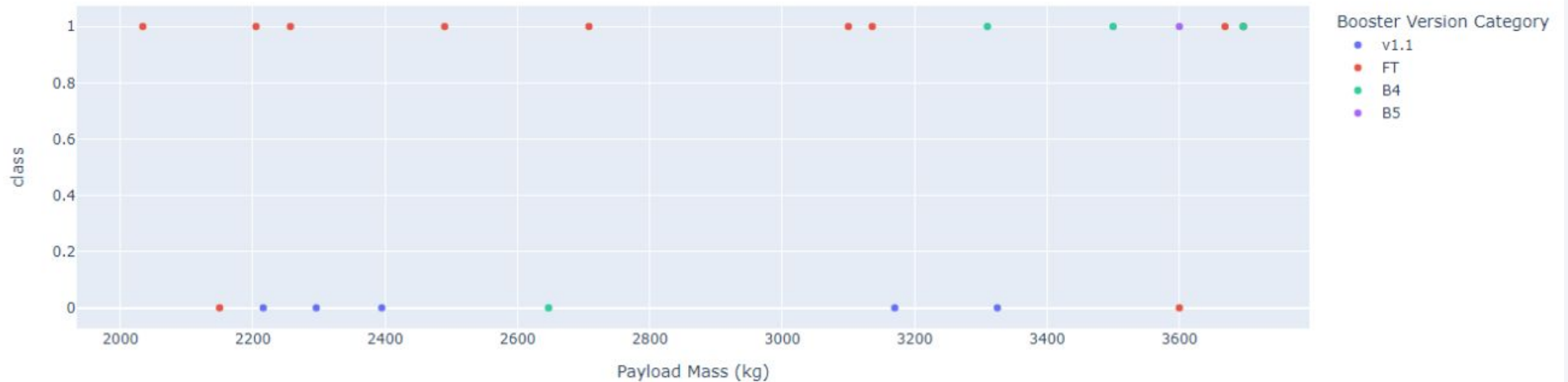
Sitio de lanzamiento con la tasa de éxito más alta

El KSLC-39A tiene la tasa de éxito más alta con un 76,9 %.



Masa de carga útil VS éxito de lanzamiento

El rango de carga útil entre 2000 kg y 4000 kg tiene la tasa de éxito más alta.

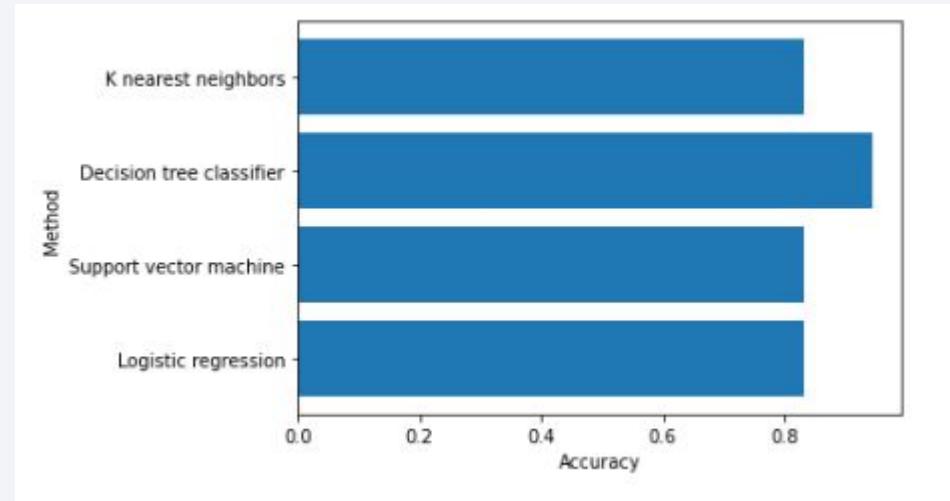


Section 5

Predictive Analysis (Classification)

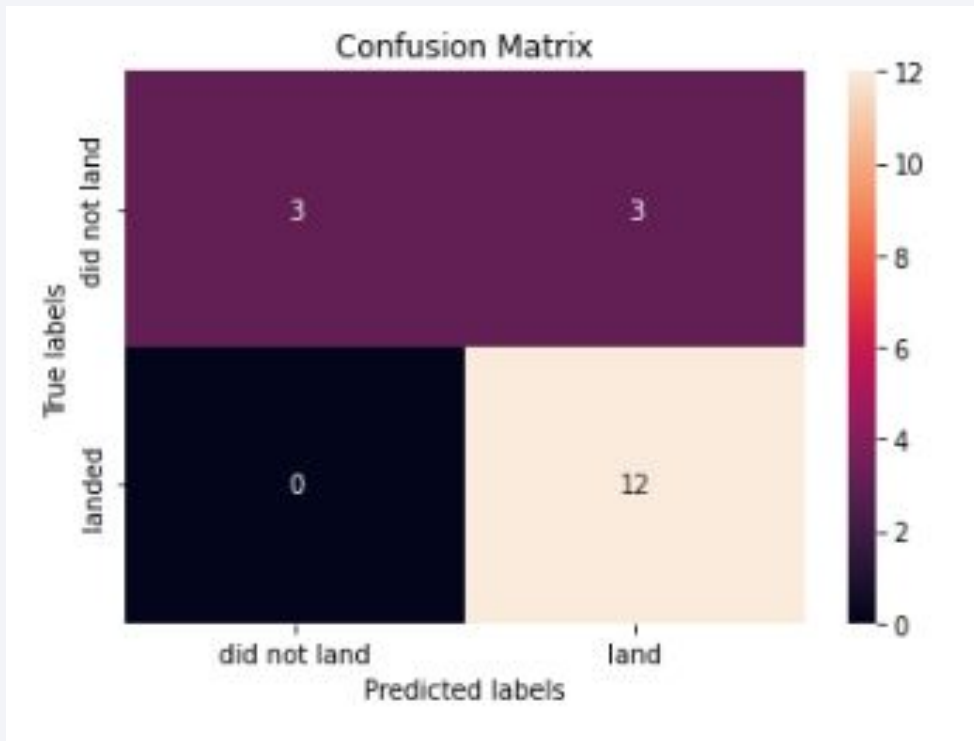
Classification Accuracy

- El clasificador Decision Tree tuvo la mejor precisión con un 94 %.



	method	accuracy
0	Logistic regression	0.833333
1	Support vector machine	0.833333
2	Decision tree classifier	0.944444
3	K nearest neighbors	0.833333

Confusion Matrix



- El modelo predijo 12 aterrizajes exitosos cuando la etiqueta True fue exitosa (True Positive) y 3 aterrizajes fallidos cuando la etiqueta True falló (True Negative).
- El modelo también predijo 3 aterrizajes exitosos cuando la etiqueta Verdadero fue un aterrizaje fallido (Falso Positivo).
- El modelo generalmente predijo aterrizajes exitosos.

Conclusions

- El análisis mostró que existe una correlación positiva entre el número de vuelos y la tasa de éxito, ya que la tasa de éxito ha mejorado a lo largo de los años.
- Hay ciertas órbitas como SSO, HEO donde los lanzamientos fueron más exitosos.
- La tasa de éxito se puede vincular a la masa de la carga útil, ya que las cargas útiles más ligeras generalmente demostraron ser más exitosas que las cargas útiles más pesadas.
- Los sitios de lanzamiento están estratégicamente ubicados cerca de carreteras y vías férreas para el transporte de personal y carga, pero también lejos de las ciudades por seguridad.
- El mejor modelo predictivo para este conjunto de datos es el Clasificador de árboles de decisión, ya que tuvo la mayor precisión con un 94 %.
-

Appendix

- Repositorio



Thank you!

