



IBM Developer  
SKILLS NETWORK

# Winning Space Race with Data Science

EDWARD ALBERTO ALVAREZ FRANCO  
AGOSTO DE 2024



# Outline

---

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

# Executive Summary

---

- Summary of methodologies
  - Recolección de los datos por dos vías: a través de la SpaceX API <https://api.spacexdata.com/v4/launches/past> y también se obtuvo datos a través de web scraping de la página Falcon 9 and Falcon Heavy Launches Records de Wikipedia.
  - Limpieza y data wrangling básico de los datos recolectados.
  - Análisis exploratorio de los datos (EDA) y visualización de los datos.
  - Análisis predictivo (Machine Learning) con base a los datos obtenidos.
- Summary of all results
  - Fue posible recolectar los datos vía fuentes públicas los cuales luego de la realización de la limpieza de los mismos y la aplicación del data wrangling nos dieron la información que necesitábamos para el análisis.
  - A través de EDA se pudo identificar las características de los datos que mejor pueden predecir los aterrizajes exitosos de los cohetes.
  - Determinamos que el mejor modelo de Machine Learning para predecir el éxito de las misiones de aterrizaje es el algoritmo de árbol de decisiones (decision tree)

# Introduction

---

- Project background and context
  - SpaceX es la compañía que ofrece los mas bajos costos del Mercado en proveer cohetes para transportar carga al espacio, la razón principal de su costo competitivo radica en el hecho que SpaceX reutiliza los cohetes de la primera etapa de sus transportadores. La empresa logra esto al aterrizar controladamente los cohetes de la primera etapa y de esta manera los recupera. Por tal motive es de gran importancia determinar la probabilidad de éxito del aterrizaje de estos cohetes ya que esta información esta directamente relacionada con los costos de operación de la empresa SpaceX.
- Problems you want to find answers
  - Con base en los datos históricos de lanzamientos, cuales son las características principales de los mismos que acompañan el éxito o fracaso del aterrizaje de los cohetes de primera etapa.
  - Qué lugar de lanzamiento brinda los mejores resultados para la empresa SpaceX.



Section 1

# Methodology

# Methodology

---

## Executive Summary

- Data collection methodology:
  - Los datos se obtuvieron de dos fuentes publicas, la API de SpaceX <https://api.spacexdata.com/v4/launches/past> y web scraping de la pagina Falcon 9 and Falcon Heavy Launches Records de Wikipedia
- Perform data wrangling
  - Se eliminaron las columnas que contenían datos irrelevantes para el objetivo del análisis.
  - Los valores numéricos faltantes se remplazaron con la media del total de los valores de su respectiva columna.
  - Se realizó un hot encoding para los modelos de clasificación.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Los datos fueron normalizados y se realizó una división de los mismos en dos grupos, el grupo de Testing y el grupo de Training con los cuales se llevaron a cabo diferentes modelos de clasificación.

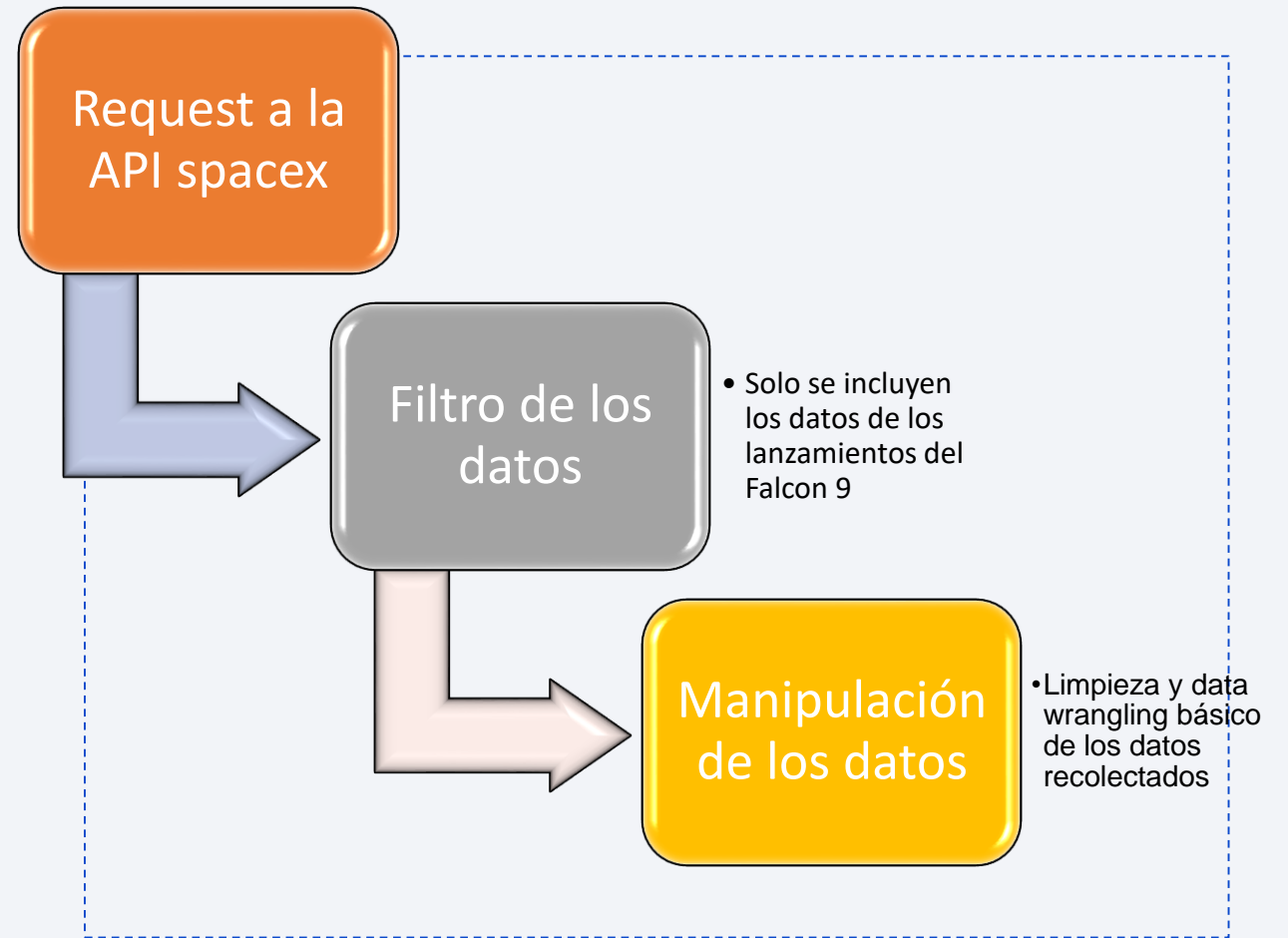
# Data Collection

---

- Los datos se recolectaron a través de la API publica de SpaceX <https://api.spacexdata.com/v4/launches/past> y a través de web scraping de la pagina Falcon 9 and Falcon Heavy Launches Records de Wikipedia

# Data Collection – SpaceX API

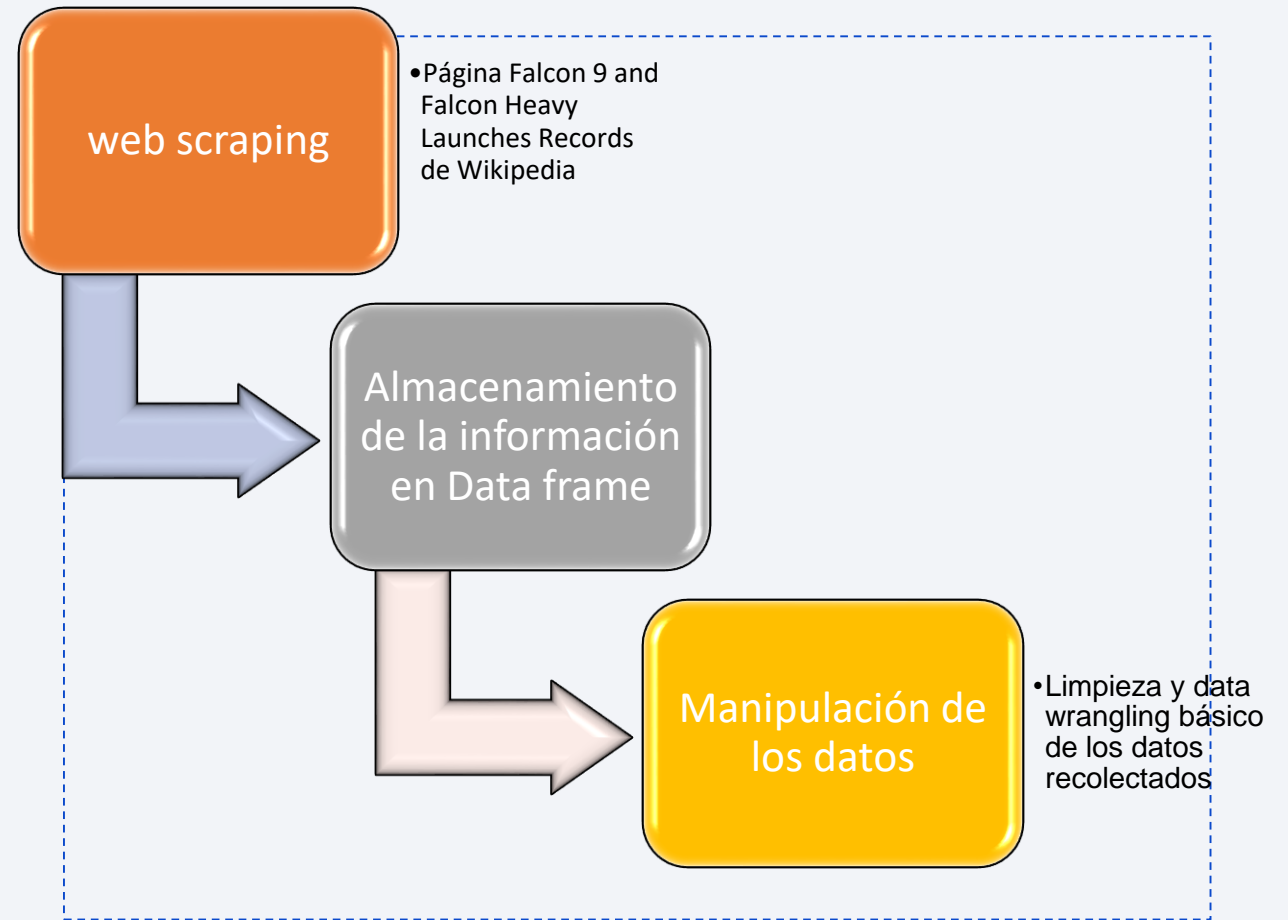
- API pública de SpaceX  
<https://api.spacexdata.com/v4/launches/past>
- GitHub URL of the completed SpaceX API calls notebook:
- <https://github.com/ingedalvarez/Applied-Data-Science-Capstone-IBM/blob/main/jupyter-labs-spacex-data-collection-api.ipynb>





# Data Collection - Scraping

- Colección de los datos de la pagina Wikipedia Falcon 9 and Falcon Heavy Launches Records a través de web scraping
- GitHub URL of the completed web scraping notebook, <https://github.com/ingedalvarez/Applied-Data-Science-Capstone-IBM/blob/main/jupyter-labs-webscraping.ipynb>



# Data Wrangling

---

- Se realizó un Análisis exploratorio de los datos (EDA) con el fin de tener una visión general de la calidad de los datos y ver cuales características necesitaban limpieza y transformación..
- Luego los datos se procesaron para subsanar los datos faltantes.
- Se agregó una nueva columna llamada Class la cual tien datos booleanos donde 1 significa aterrizaje exitoso y 0 aterrizaje fallido.
- Las características categóricas de los datos se codificaron a través de un one hot encode con el fin de tener datos numéricos en estas característica que luego puedan ser usadas en los modelos de Machine Learning.
- GitHub URL of completed data wrangling related notebooks,  
<https://github.com/ingedalvarez/Applied-Data-Science-Capstone-IBM/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb>

# EDA with Data Visualization

---

- Se hizo uso de gráficos de punto para visualizar la relación entre diferentes pares de características del conjunto de datos.
- Se elaboró un grafico de barras para visualizar la relación del porcentaje de éxito de la misión con respecto a la orbita final de la misión
- GitHub URL of completed EDA with data visualization notebook, <https://github.com/ingedalvarez/Applied-Data-Science-Capstone-IBM/blob/main/edadataviz.ipynb>

# EDA with SQL

---

- Muestra 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'
- Muestra la masa total de carga útil transportada por los cohetes lanzados por la NASA (CRS)
- Muestra la masa promedio de la carga útil transportada por el cohete de la versión F9 v1.1
- Enumere la fecha en la que se logró el primer aterrizaje exitoso en la plataforma de tierra.
- Enumere 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
- List the total number of successful and failure mission outcomes.
- Enumere los nombres de las versiones de refuerzo que han transportado la masa máxima de carga útil.
- List the records which will display the month names, failure landing\_outcomes in drone ship ,booster versions, launch\_site for the months in year 2015.
- 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.
- GitHub URL of completed EDA with SQL notebook, [https://github.com/ingedalvarez/Applied-Data-Science-Capstone-IBM/blob/main/jupyter-labs-eda-sql-coursera\\_sqllite.ipynb](https://github.com/ingedalvarez/Applied-Data-Science-Capstone-IBM/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb)

# Build an Interactive Map with Folium

---

- Marcadores que identifican los puntos de lanzamiento de los cohetes.
- Clusters de marcadores los cuales muestran conjuntos relacionados de eventos en las coordenadas marcadas como pueden ser lanzamientos por sitio de lanzamiento.
- Líneas, las cuales se utilizan para medir distancias entre dos puntos, como la distancia de los sitios de lanzamiento a la orilla del mar.

Estos indicadores nos ayudan a entender de una mejor manera los datos, podemos con ellos fácilmente mostrar todos los sitios de lanzamiento junto con los aterrizajes exitosos y fallidos.

- GitHub URL of completed interactive map with Folium map,  
[https://github.com/ingedalvarez/Applied-Data-Science-Capstone-IBM/blob/main/lab\\_jupyter\\_launch\\_site\\_location.ipynb](https://github.com/ingedalvarez/Applied-Data-Science-Capstone-IBM/blob/main/lab_jupyter_launch_site_location.ipynb)

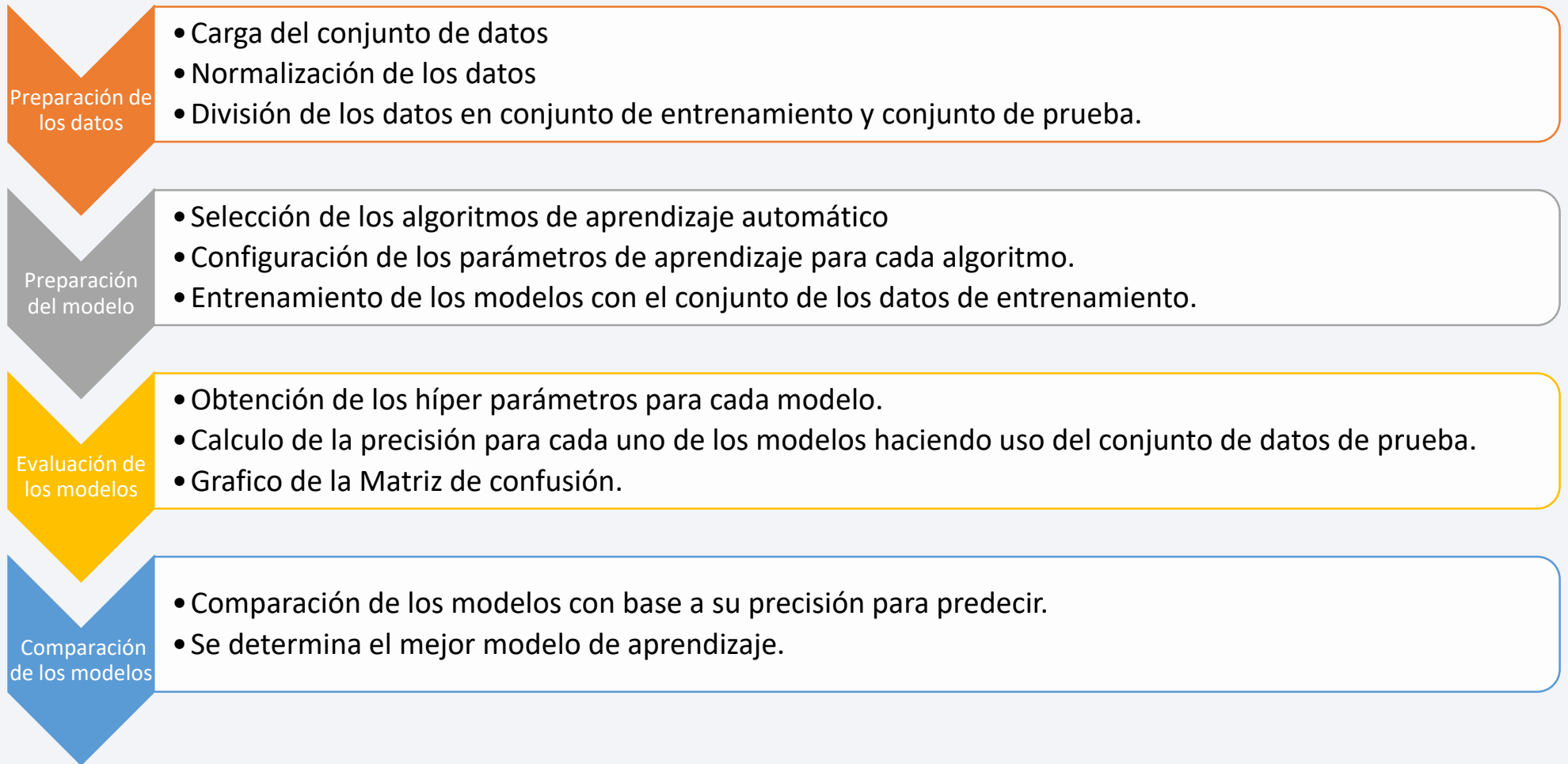


# Build a Dashboard with Plotly Dash

---

- El Dashboard presenta menús dropdown, gráficos de torta y gráficos de dispersión.
  - Los menú dropdown brindan al usuario la posibilidad de escoger los datos a graficar (individualmente o todos a la vez).
  - La grafica de torta muestra el total de los aterrizajes (exitosos o fallidos) por sitio de lanzamiento.
  - Por medio de un control deslizable de rango el usuario puede seleccionar la carga del cohete.
  - EL grafico de dispersión muestra la relación entre el éxito de la mision y su carga útil.
- GitHub URL of completed Plotly Dash lab, <https://github.com/ingedalvarez/Applied-Data-Science-Capstone-IBM/blob/main/Build%20a%20Dashboard%20Application%20with%20Plotly%20Dash%201.pdf>

# Predictive Analysis (Classification)



- GitHub URL of completed predictive analysis lab, [https://github.com/ingedalvarez/Applied-Data-Science-Capstone-IBM/blob/main/SpaceX\\_Machine%20Learning%20Prediction\\_Part\\_5.ipynb](https://github.com/ingedalvarez/Applied-Data-Science-Capstone-IBM/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb)

# Results

---

- Exploratory data analysis results
  - Fecha del primer aterrizaje exitoso fue en 2015.
  - La carga útil promedio de los cohetes es de 3 toneladas.
  - El porcentaje de aterrizajes exitosos ha ido en aumento con el transcurrir de los años.
  - SpaceX posee 4 sitios de lanzamiento de sus cohetes.
- Predictive analysis results
  - El modelo de aprendizaje automático que presenta los mejores resultados contra el conjunto de datos de pruebas es el Clasificador por Árbol de decisiones, aunque la diferencia de desempeño con los otros modelos es insignificante.



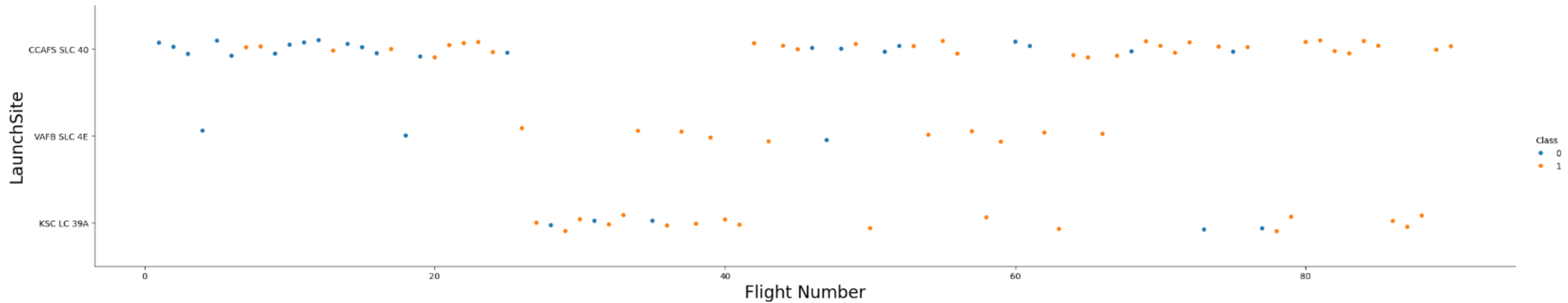
The background of the slide is an abstract composition. It features a dark blue base color. Overlaid on this are numerous diagonal streaks in shades of red and cyan. A faint, light blue grid pattern is also visible, particularly in the lower half of the image. The overall effect is dynamic and technological.

Section 2

# Insights drawn from EDA



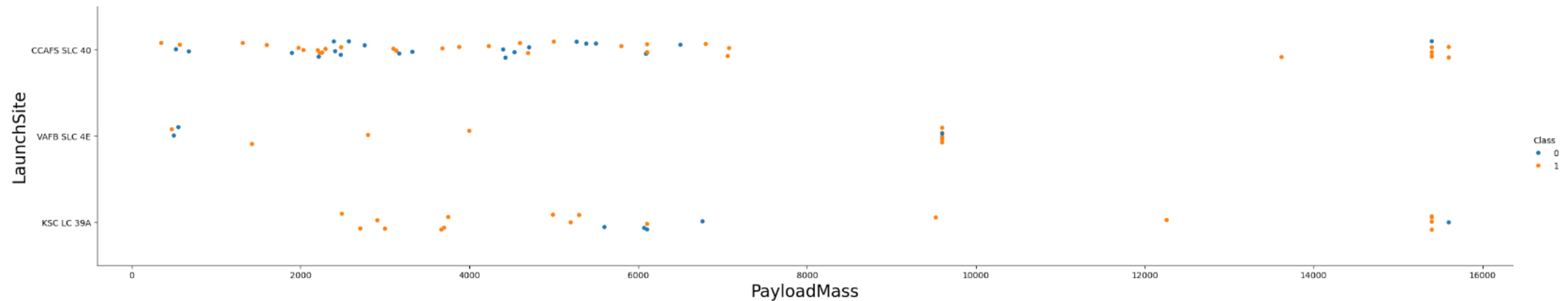
# Flight Number vs. Launch Site



- En el grafico se puede apreciar que la rata de éxito de los aterrizajes ha ido aumentando con los años aunque aun esta lejos de ser del 100%
- EL lugar de lanzamientos mas utilizado por SpaceX es CCAFS SLC 40.



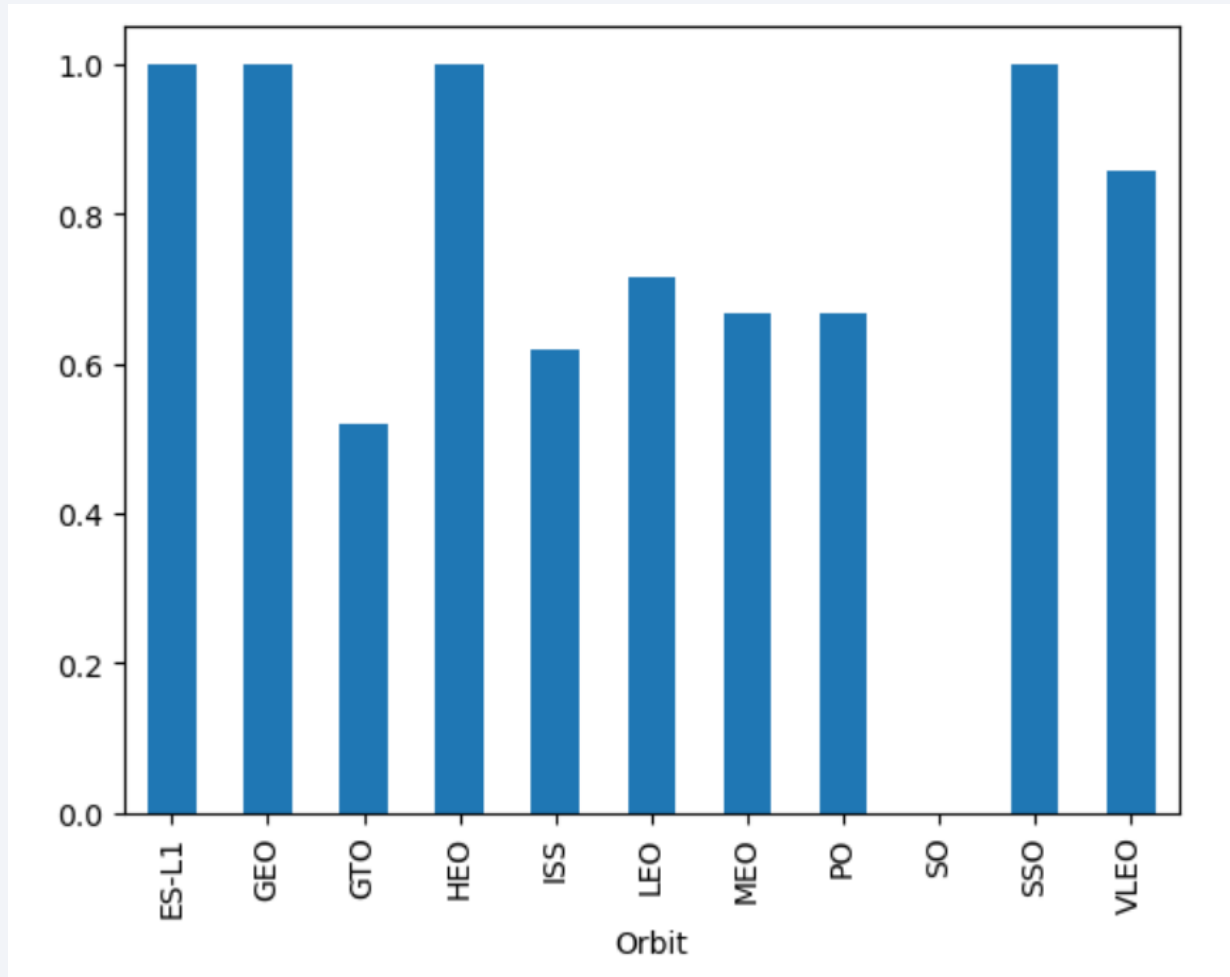
# Payload vs. Launch Site



- La mejor rata de éxito del aterrizaje de los cohetes de primera etapa se presenta en los cohetes con una carga útil superior a las 14 toneladas.
- Solo dos de los tres sitios de lanzamiento tienen la capacidad de lanzar cohetes de mas de 10 toneladas
- La tasa de éxito de aterrizaje de los cohetes de 4 toneladas es del 100%

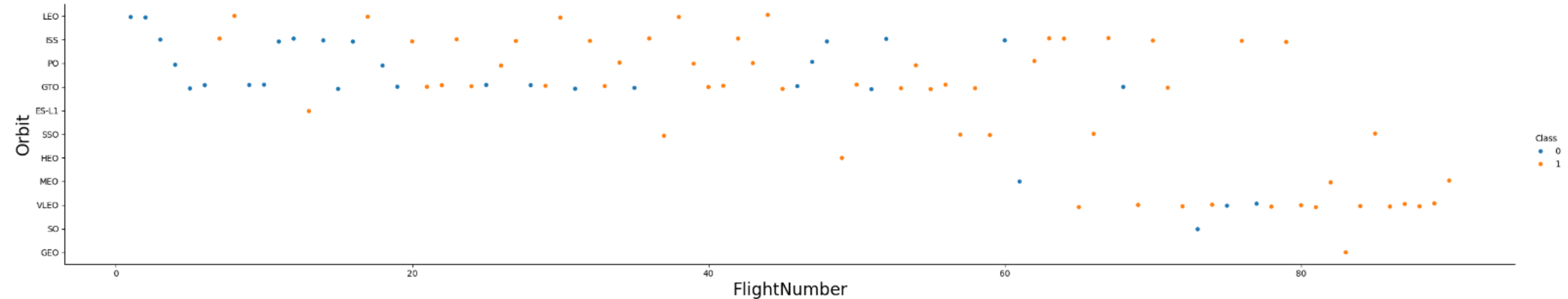
# Success Rate vs. Orbit Type

---



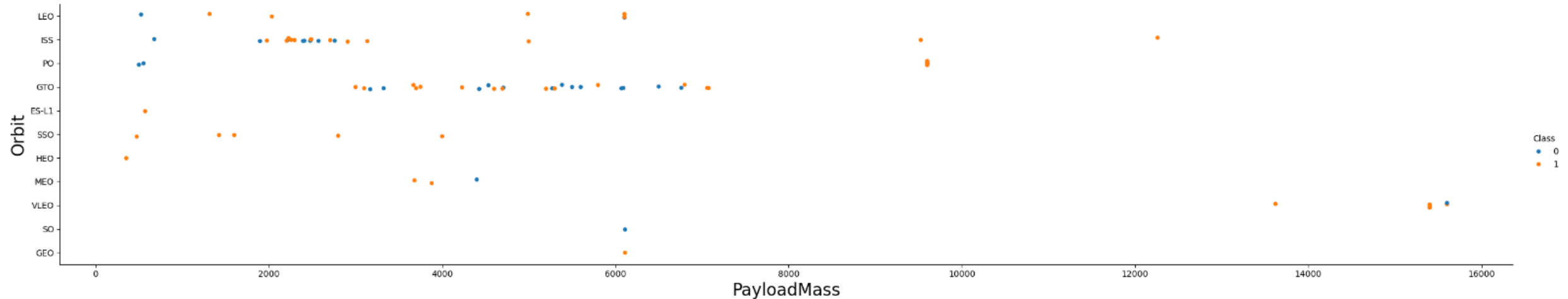
- En el grafico se puede apreciar claramente cuales son las cuatro orbitas con la tasa de éxito mas alta.
- La orbita GTO es la que presenta mas baja rata de éxito de los aterrizajes.

# Flight Number vs. Orbit Type



- Comparado con la gran cantidad de fracasos en los aterrizajes al inicio del Proyecto se puede evidenciar que la tasa de éxito de los aterrizajes se ha ido incrementando con el tiempo.
- Con relación a los últimos lanzamientos del conjunto de datos podemos evidenciar que los lanzamientos se han centrado en dos orbitas: LEO y VLEO.

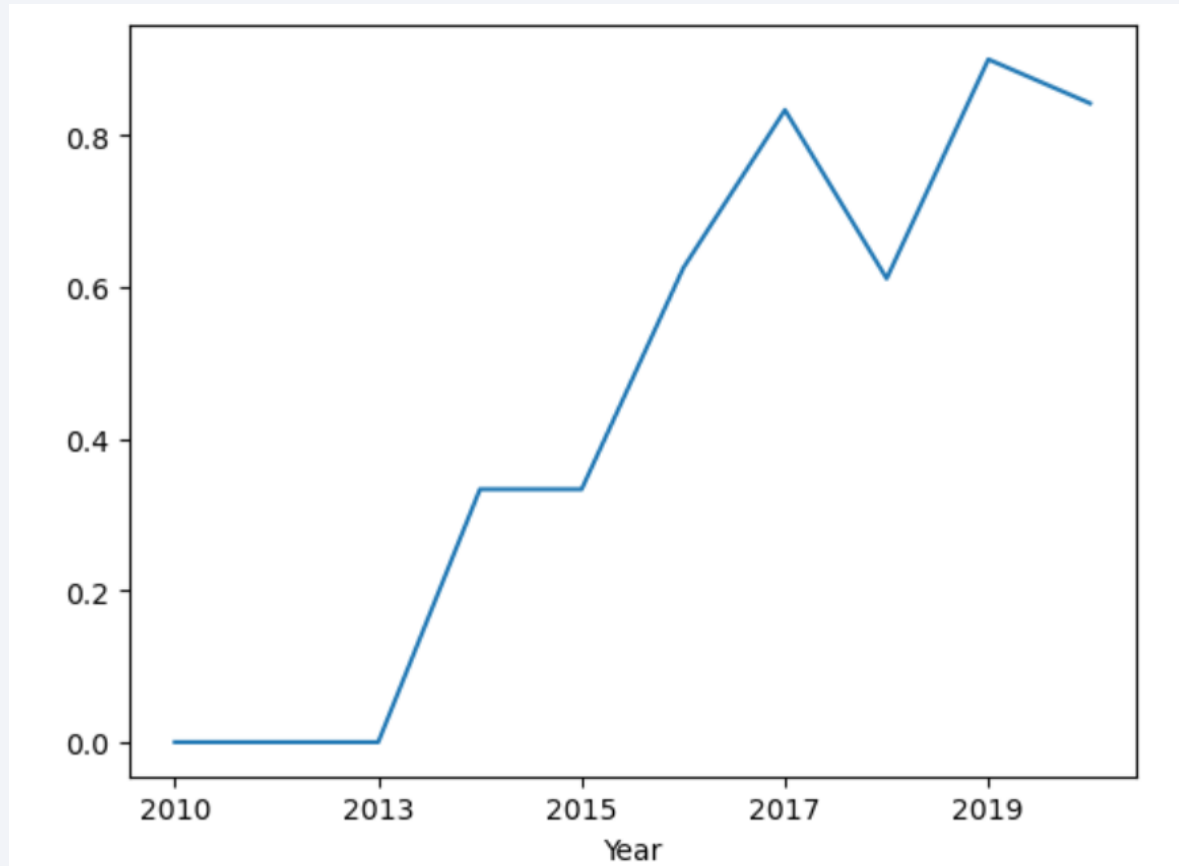
# Payload vs. Orbit Type



- La tasa de éxito de los aterrizajes se acerca al 100% para cargas útiles superior a 10 ton.
- Con respecto a la carga útil la orbita LEO es la que tiene mas éxito en las misiones.

# Launch Success Yearly Trend

---



- A partir del 2015 la tasa de éxito de los aterrizajes de los cohetes ha tenido un incremento alto.
- En el año 2018 cayó la tasa de éxito que venia desde el año 2015.
- Según los datos a partir del 2019 la tasa de éxito de los aterrizajes ha encontrado un techo que no ha podido superar (+/- 80%)



# All Launch Site Names

---

```
%sql select distinct launch_site from SPACEXTABLE;  
* sqlite:///my_data1.db  
Done.
```

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

- Se realizó una consulta select distinct a la base de datos para que esta me devolviera los sitios de lanzamiento sin repetir nombres.

# Launch Site Names Begin with 'CCA'

```
%sql select * from SPACEXTABLE 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
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

---

```
%sql select sum(payload_mass__kg_) as total_payload_mass from SPACEXTABLE where customer = 'NASA (CRS)';
```

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

```
Done.
```

<u>total_payload_mass</u>
---------------------------

45596
-------

- Se realizó una consulta con la función de agregación sum sobre la columna payload\_mass\_kg\_ filtrada por la columna customer = NASA (CRS) la que nos devolvió el valor 45,596 Kg como el total de la carga desplazada al espacio por los lanzamientos que aparecen en el conjunto de datos analizados para el cliente NASA.

# Average Payload Mass by F9 v1.1

---

```
%sql select avg(PAYLOAD_MASS_KG_) as average_payload_mass from SPACEXTABLE where booster_version like '%F9 v1.1%';
```

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

```
Done.
```

<u>average_payload_mass</u>
-----------------------------

2534.6666666666665
--------------------

- Se realizó una consulta con la función agregada avg sobre la columna Payload Mass al cohete versión F9 v1.1 la cual arrojó que la carga útil promedio de este propulsor es de al rededor 2 toneladas y media.

# First Successful Ground Landing Date

---

```
%sql select min(date) as first_successful_landing from SPACEXTABLE where landing_outcome = 'Success (ground pad)';
```

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

```
Done.
```

<u>first_successful_landing</u>
---------------------------------

2015-12-22
------------

- Tras consultar la tabla SPACEXTABLE encontramos que el 2015-12-22 fue la fecha del primer aterrizaje exitoso de SpaceX.



# Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql select booster_version from SPACEXTABLE where landing_outcome = 'Success (drone ship)' and PAYLOAD_MASS__KG_ between 4000 and 6000
```

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

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

- La siguiente fue la consulta usada para determinar las versiones de los propulsores que han tenido un aterrizaje exitoso sobre dron en el mar.
- %sql select booster\_version from SPACEXTABLE where landing\_outcome = 'Success (drone ship)' and PAYLOAD\_MASS\_\_KG\_ between 4000 and 6000;

# Total Number of Successful and Failure Mission Outcomes

---

```
%sql select mission_outcome, count(*) as total_number from SPACEXTABLE group by mission_outcome;
```

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

```
Done.
```

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

Con base en los datos almacenados en la base datos podemos apreciar un porcentaje cercano al 100% en los resultados de las misiones llevada a cabo por los cohetes de SpaceX.

# Boosters Carried Maximum Payload

---

```
%sql select booster_version from SPACEXTABLE where payload_mass__kg_ = (select max(payload_mass__kg_) from SPACEXTABLE);
```

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

```
Done.
```

<b>Booster_Version</b>
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

# 2015 Launch Records

```
%sql select substr(Date, 6,2) as month, date, booster_version, launch_site, Landing_Outcome from SPACEXTABLE where Landing_Outcome = 'Failure (drone ship)'
```

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

Done.

month	Date	Booster_Version	Launch_Site	Landing_Outcome
01	2015-01-10	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)
01	2016-01-17	F9 v1.1 B1017	VAFB SLC-4E	Failure (drone ship)
03	2016-03-04	F9 FT B1020	CCAFS LC-40	Failure (drone ship)
06	2016-06-15	F9 FT B1024	CCAFS LC-40	Failure (drone ship)

- %sql select substr(Date, 6,2) as month, date, booster\_version, launch\_site, Landing\_Outcome from SPACEXTABLE where Landing\_Outcome = 'Failure (drone ship)'
- En el año 2015 todos los intentos de aterrizaje se llevaron a cabo sobre drones en el mar y todos resultaron en intentos fallidos.

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

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

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

Done.

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

- Clasificación de resultados de aterrizaje entre la fecha 2010-06-04 y 2017-03-20, en orden descendente.

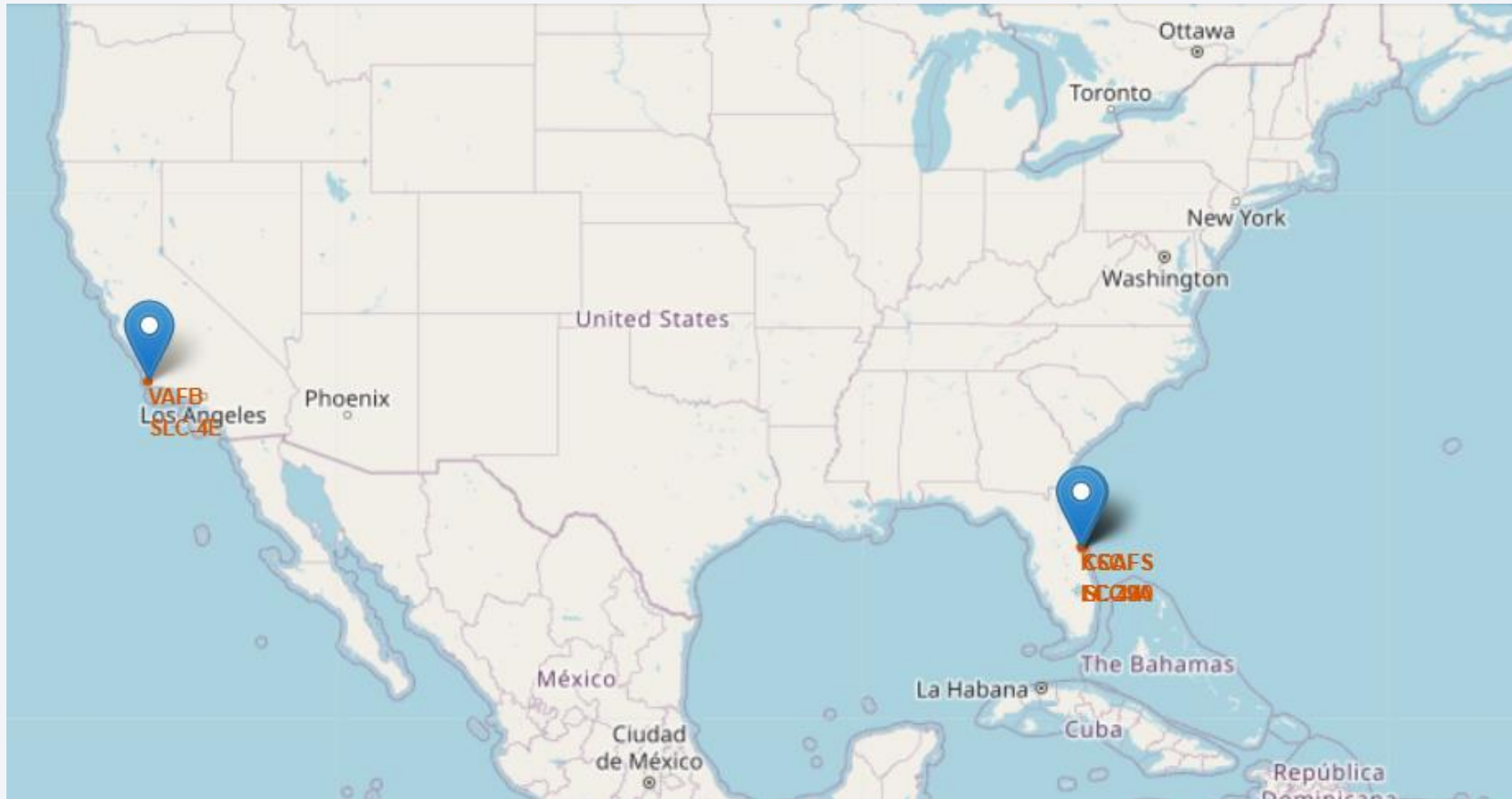
A satellite view of Earth from space, showing the curvature of the planet and city lights at night. The background is a deep blue gradient.

Section 3

# Launch Sites Proximities Analysis

# Locaciones de lanzamiento de SpaceX

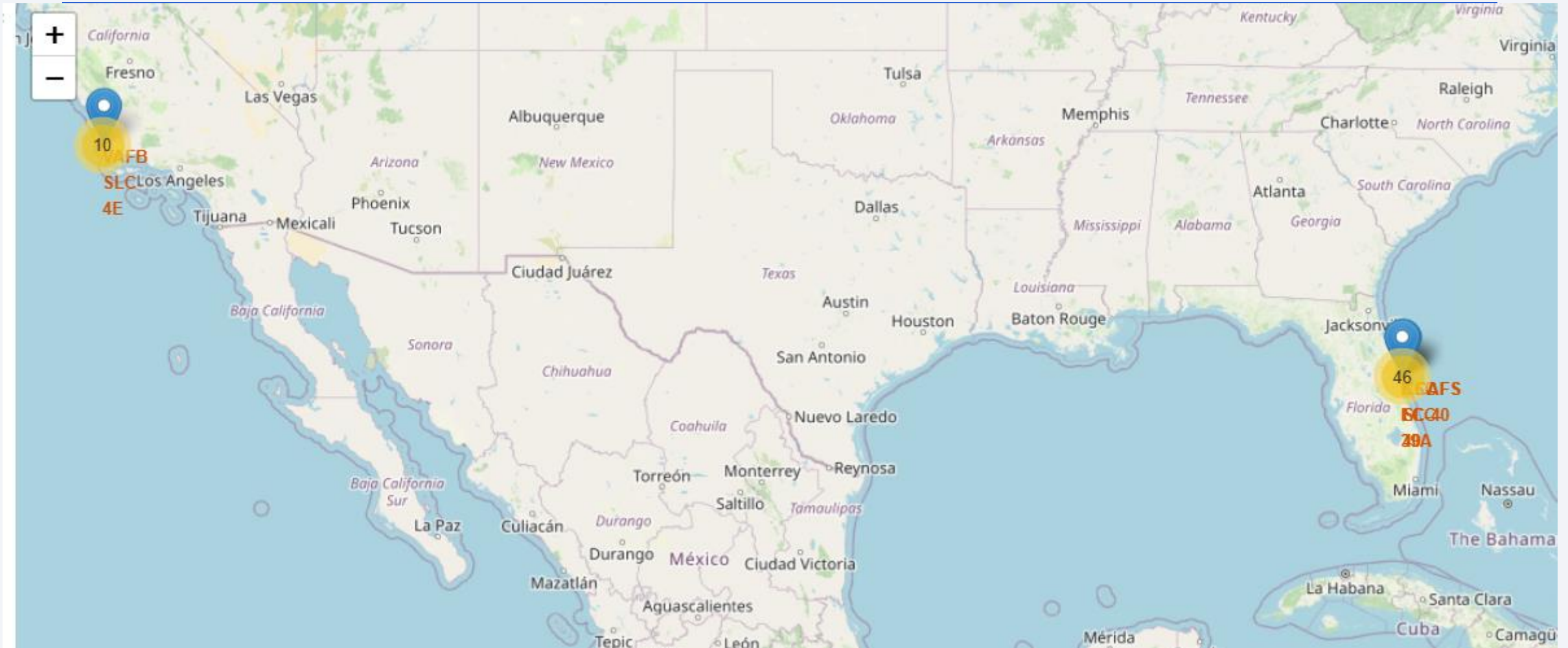
---



- En el mapa se puede apreciar que los lugares de lanzamiento de la empresa SpaceX se encuentran cerca de ambas costas de EEUU.

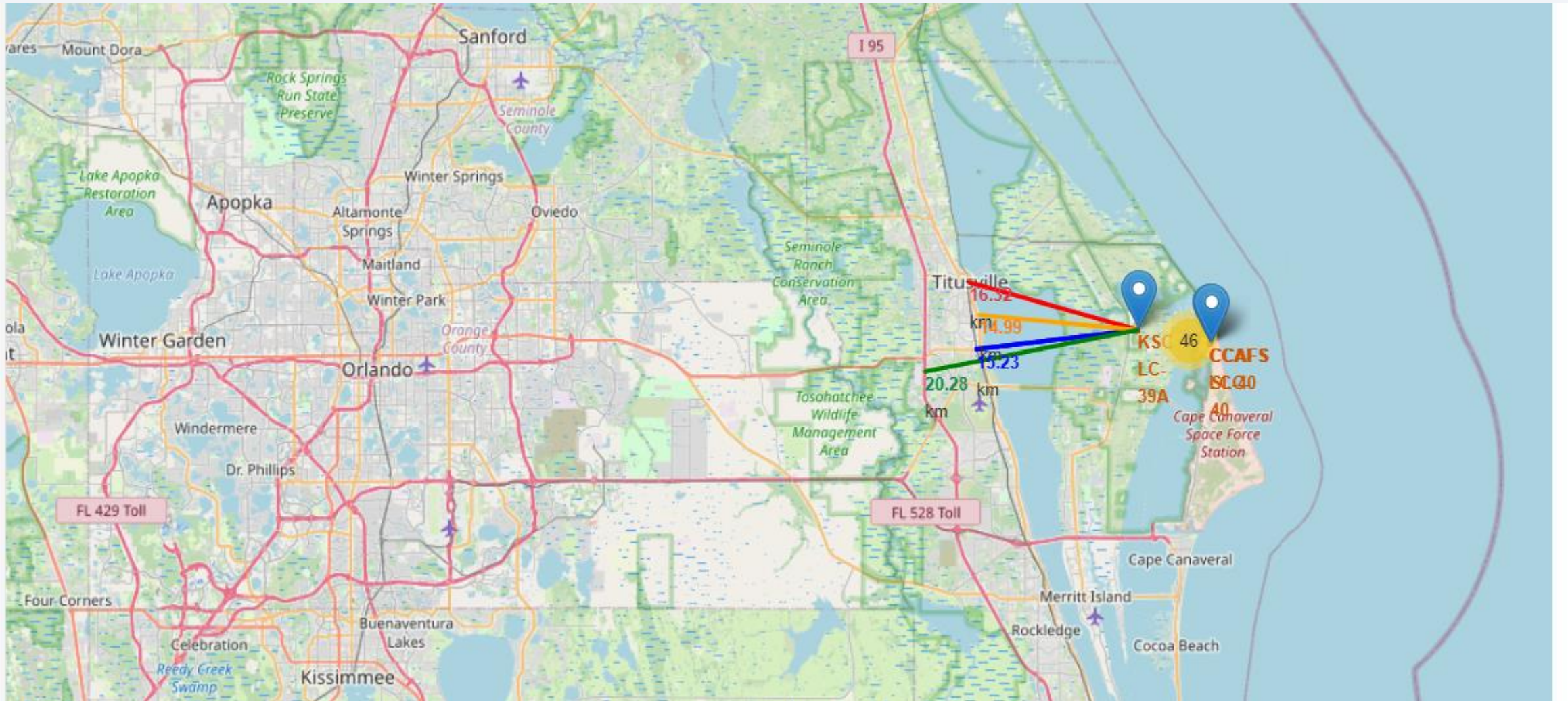


# Lanzamientos exitosos y fallidos por cada sitio de lanzamiento





## Distancia a puntos claves desde los sitios de lanzamiento en Florida







Section 4

# Build a Dashboard with Plotly Dash

# Total misiones exitosas por lugar de lanzamiento

---

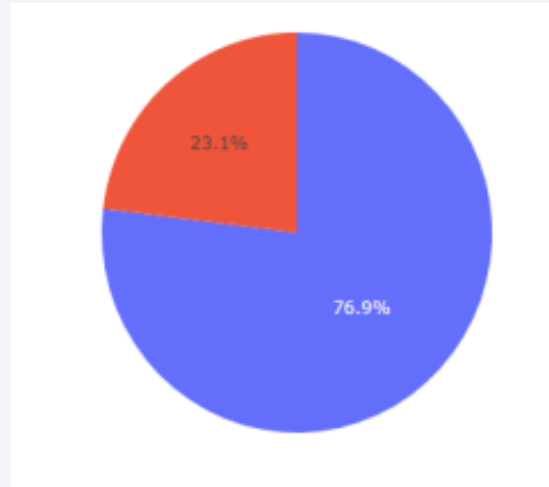
Total Success Launches by Site



- Se observa que el lugar de lanzamiento mas exitoso es KSC LC-39A el cual duplica en exitos al Segundo lugar.

# Relación de éxito y fracaso de las misiones del punto de lanzamiento KSC LC-39A

---

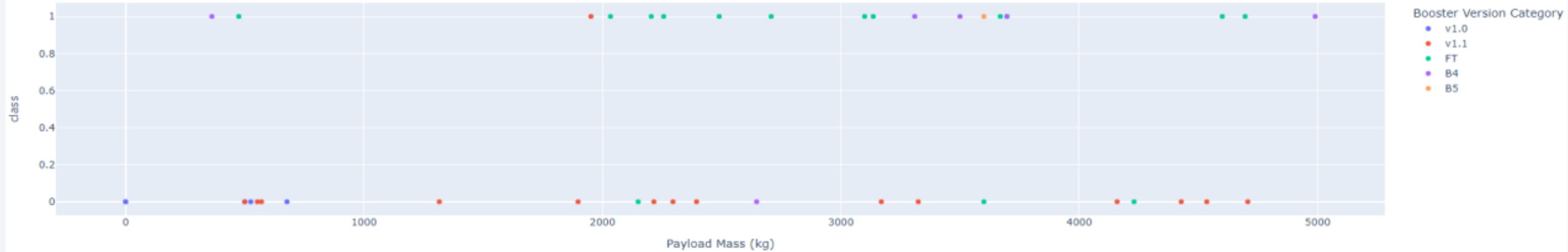


El grafico muestra que menos del 25% de las misiones de aterrizaje de este punto de lanzamiento son no exitosas.

# Ejemplos de gráficos de dispersión de Carga Útil vs Resultado de la misión.

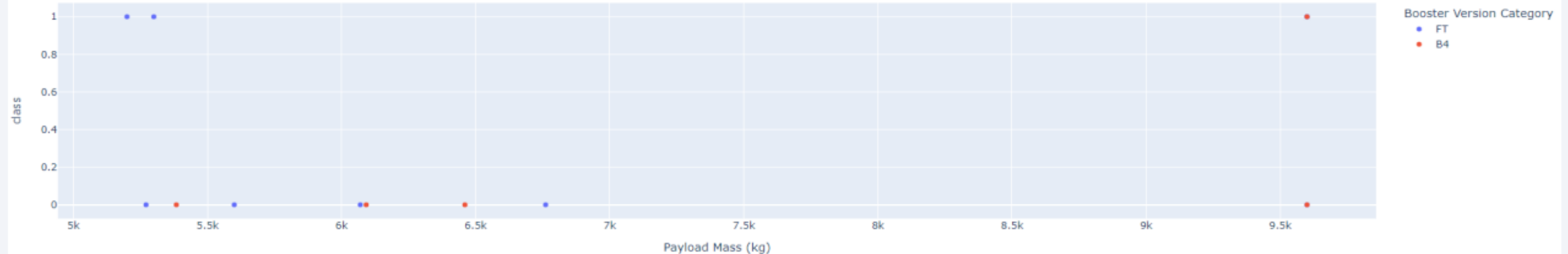
Correlation between Payload and Success for all Sites

## Low weighted payload (0 – 5000 kg)



Correlation between Payload and Success for all Sites

## Heavy weighted payload (5000 – 10000 kg)

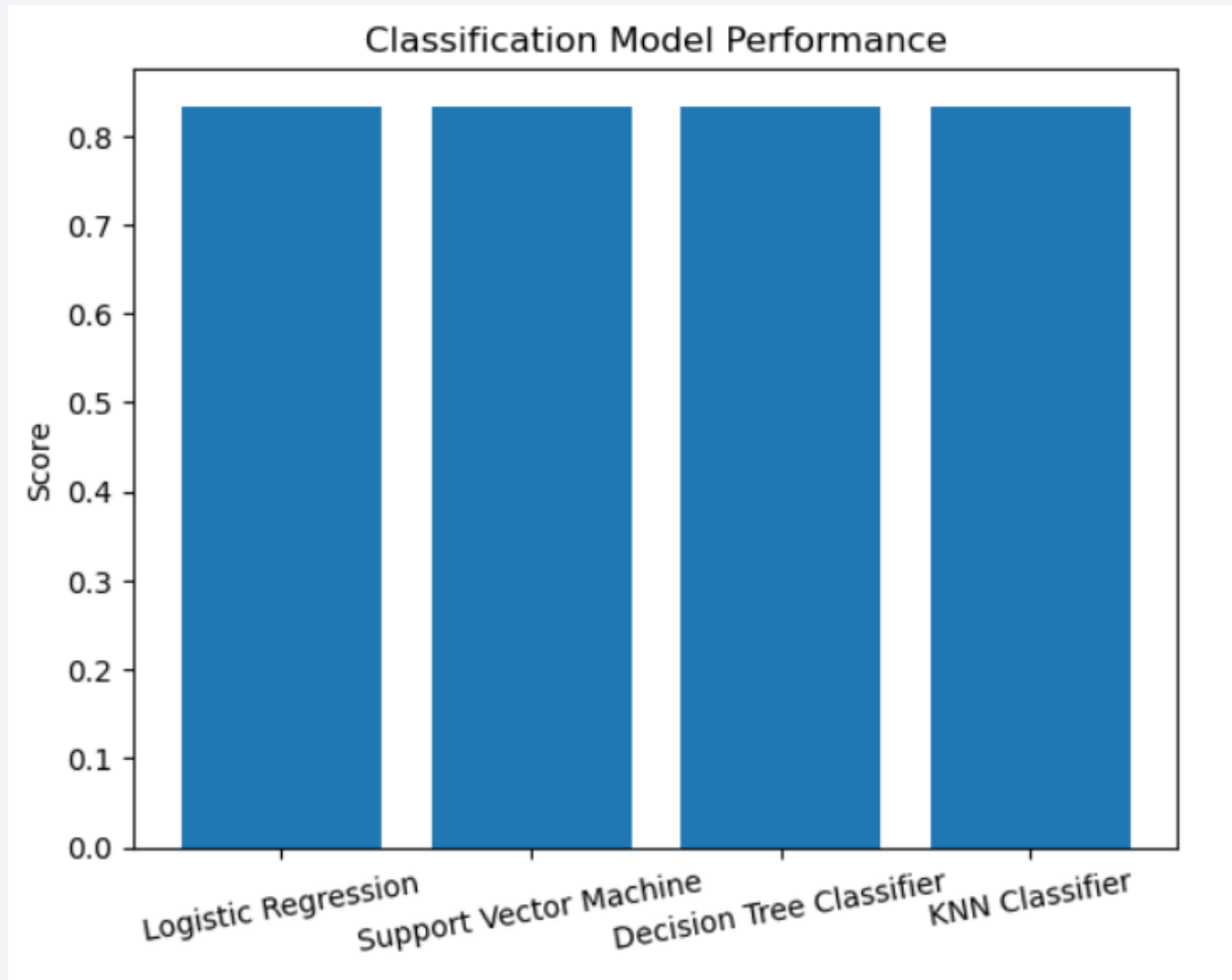


Section 5

# Predictive Analysis (Classification)

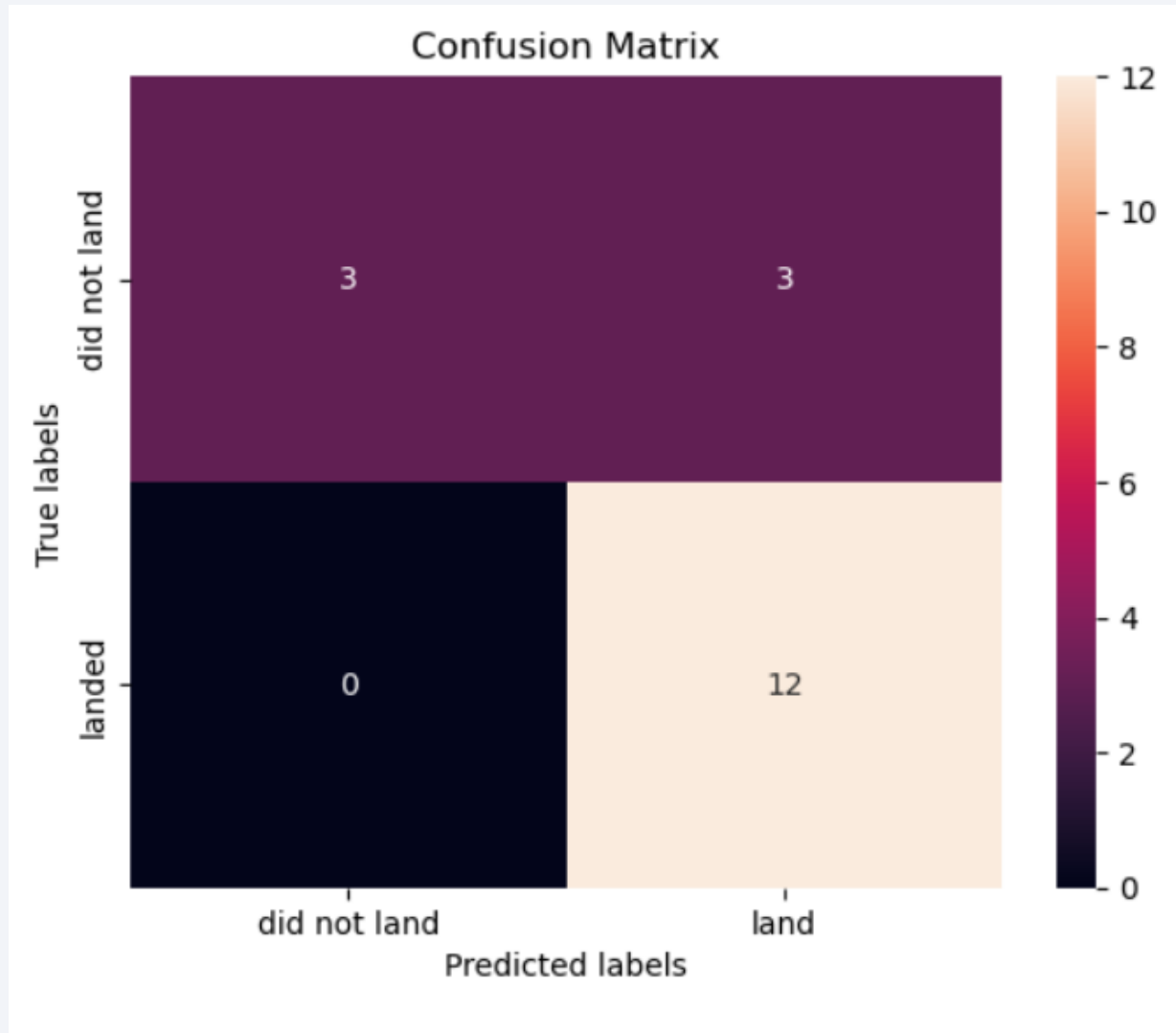


# Classification Accuracy



- Se realizaron 4 modelos de clasificación de aprendizaje automático cuyo desempeño se aprecia en el gráfico de barras mostrado en la imagen.
- Según los datos obtenidos con la evaluación realizado con el conjunto de datos de prueba ninguno de los modelos probados presenta mejor desempeño que los demás.

# Confusion Matrix



- Las matrices de confusión de cada uno de los cuatro modelos de clasificación probados presentaron el mismo comportamiento, en la imagen se puede apreciar una de ellas.



# Conclusions

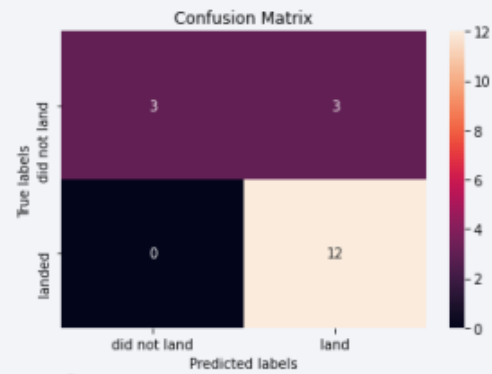
---

- El sitio de lanzamiento con el mayor porcentaje de éxito en la recuperación de los propulsores de primera etapa es KSC LC-39A.
- Las misiones con carga superior a 10 toneladas presentan mayores éxitos de recuperación de los propulsores de primera etapa.
- A medida que ha pasado el tiempo el éxito de los aterrizajes ha aumentado hasta alrededor del 85% punto en el cual no ha sido posible aumentar el porcentaje de éxito de las misiones.
- Cualquiera de los modelos de clasificación utilizados en el análisis puede ser usado para predecir con un alto nivel de precisión si la misión será exitosa o fallida.

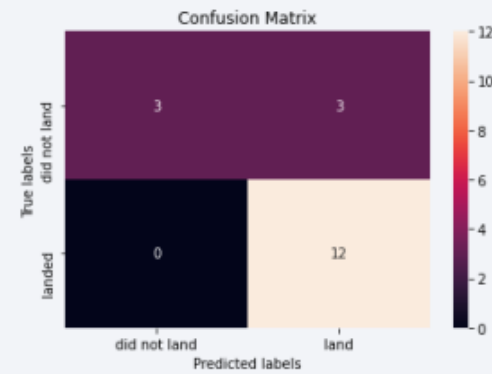
# Appendix

---

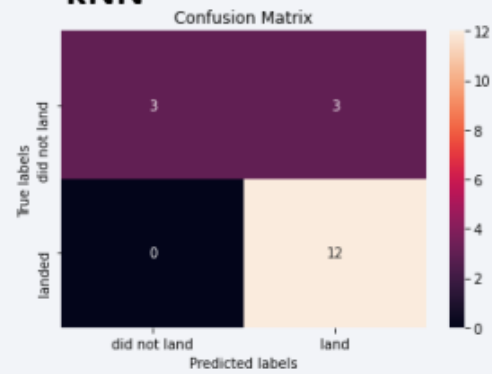
**Logistic regression**



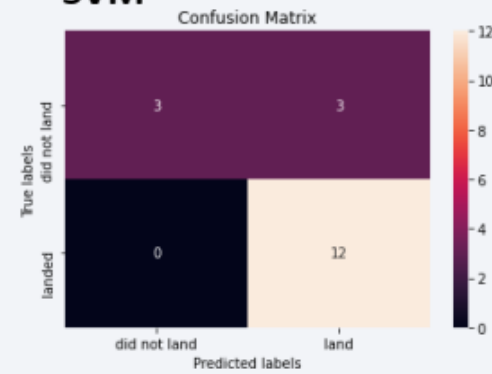
**Decision Tree**



**kNN**



**SVM**



Thank you!

