IBM Developer
SKILLS NETWORK

# Winning Space Race with Data Science

Eliel Leandro Alves Junior
18/04/2022

# Outline

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

# Executive Summary

- We have entered another space-age, unlike the first one where the main objective was the dispute between the USA and the USSR to know who would be the first to conquer space. Today, the dispute is no longer ideological but commercial to highlight the companies Virgin Galactic, Blue Origin, and SpaceX is marketing space travel to all people who can afford it. Among these companies, we can highlight SpaceX for its successes. The main reason for SpaceX's success is the reuse of the first stage of the rockets, thus reducing the launch cost by 30%, therefore a significant step toward the commercialization of trips to space. Therefore, in this project, we seek to analyze the success of the first stage landing.

- This work followed the steps: Data Collection through API, Data Collection with Web Scraping, Data Wrangling, Exploratory Data Analysis with SQL, Exploratory Data Analysis with Data Visualization, Interactive Visual Analytics with Folium, and Machine Learning Prediction.

- We can say that the models' results obtained reasonable accuracy within the collected data.

# Introduction

- SpaceX began its journey into space when Falcon 1 successfully launched a payload into orbit in July 2009. They then worked on a new version of the Falcon family, the Falcon 5. However, this new model was canceled as the customer wanted to send a payload that exceeded the payload mass of the falcon 5. Therefore, to meet this customer's need, the Falcon 9 was created.

- The Falcon 9 is a two-stage, liquid-propellant launch vehicle designed to reduce the cost of previous models as much as possible.

- In December 2015, SpaceX successfully launched a payload into orbit on Falcon 9 and managed to recover the first stage in a boost back landing, returning to the launch site. Some time passed, and in mid-September 2017, SpaceX recovered sixteen first stages on land or ships at the launch site. Two of them were stages that flew and were previously recovered.

# Introduction

- With its Falcon 9 flagship rocket, SpaceX reduced the cost per launch by $62 million with the ability to recover stage one. However, recovery from the first stage is not always successful.

- Therefore, in this work, we will be analyzing and testing some models to predict whether the SpaceX Stage One recovery will be successful or not.

Section 1

# Methodology

# Methodology

Executive Summary:

- Data collection methodology:
  - The data were collected through the SpaceX API and scraping the Wikipedia page.
- Perform data wrangling
  - The data has been processed.

- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
- In this step, we split the data into training and test sets and then applied various KNN and Decision Trees models. Then next, we evaluate the model's accuracy score and their best test set parameters.

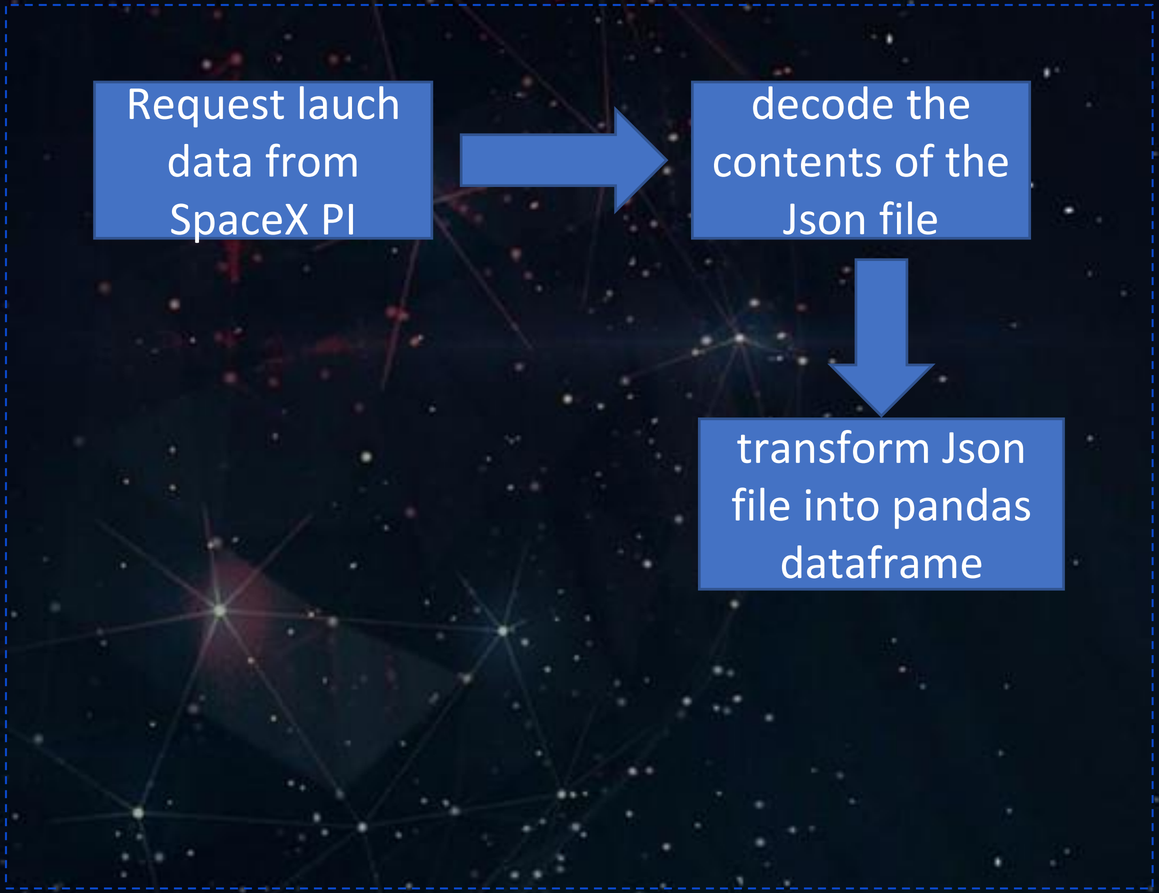
# Data Collection

- Data was collected using the SpaceX API.
- After collecting the data, we decode the contents of the Json file by calling the .json() function and then transforming in a pandas dataframe using .json_normalize() .
- After cleaning the data, we check for missing values and put other values in their place.
- We also use web scraping from the Wikipedia page to find Falcon 9 launch logs with Beautiful Soup.
- Then, we extract the launch records as an HTML table, analyze the table, and put it in a pandas DataFrame for later analysis.

# Data Collection – SpaceX API

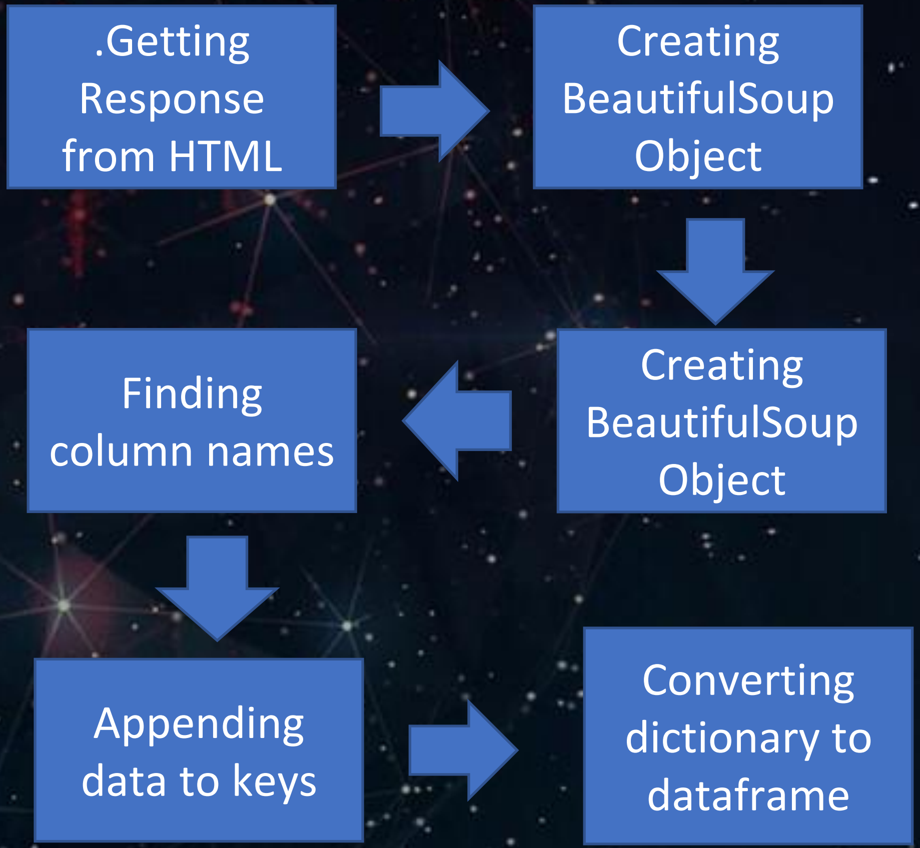
- Through the get request, we access the SpaceX API to Collect Data, Clean and Save it to a Dataframe

- Full notebook is available:. [Github URL to Notebook](#)t



# Data Collection - Scraping

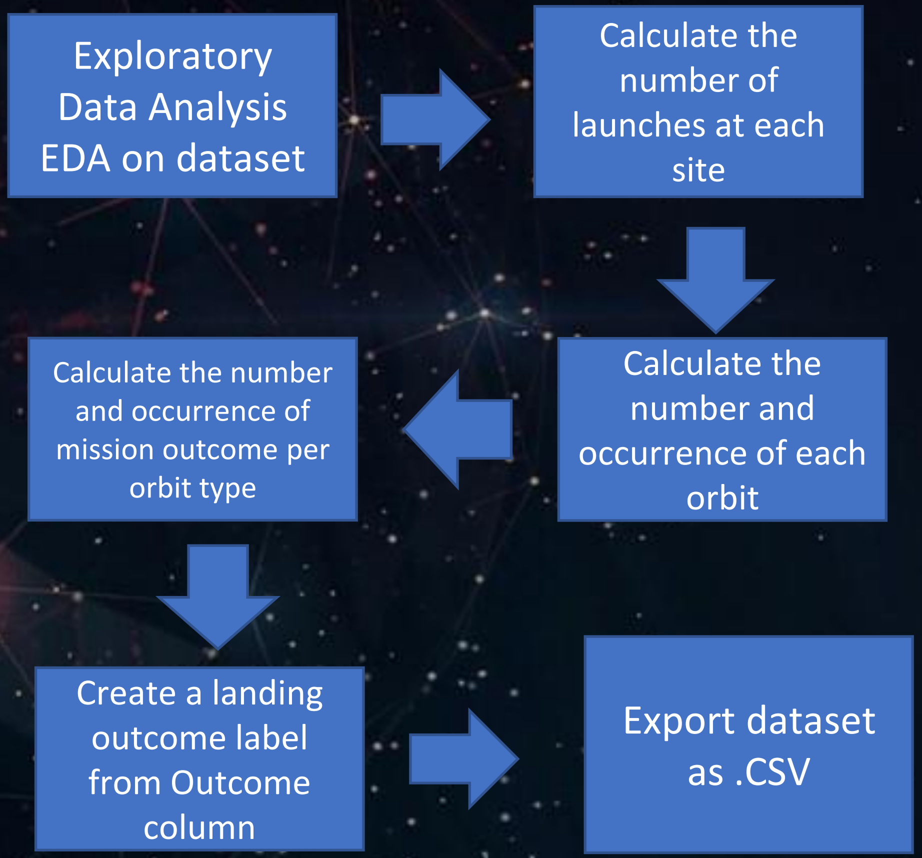
- We use web scraping to collect the Falcon 9 launch logs with BeautifulSoup.
- We study the table and convert it into a data frame.
- Full notebook is available :[Github URL to Notebook]



# Data Wrangling

- In the collected data, there are several cases where the booster was not successful in its landing. In the data we can verify that True RTLS means that successfully landed on a block of earth. For False RTLS it means that the landing result was unsuccessful on a block of land. True ASDS means it has successfully landed on a drone ship and the fake ASDS means it has not successfully landed on a drone ship. Therefore, we convert these results to 1 the ones that landed successfully and 0 means not landed successfully.

Full notebook is available : [GitHub URL to Notebook]



# EDA with Data Visualization

- We visualize the relationship between the flight Site number and release, payload and launch site, the success rate of each orbit type, flight number and orbit type, and the launch booming annual trend.
- Summary of charts:
  - Scatter plot between Flight Number and Launch Site The data were collected through the SpaceX API
  - Scatter plot between Payload Mass and Launch Site
  - Bar chart observing the success rate for Orbit type
  - Scatter plot between Flight Number and Orbit type
  - Scatter plot between Payload Mass and Orbit type
  - Line chart showing the Launch success yearly trend

Full notebook is available :[Github URL to Notebook]

# EDA with SQL

- Performed SQL queries to gather information about the dataset:
- Display the names of the unique launch sites in the space mission
- Display 5 records where launch sites begin with the string ‘CCA.’
- Display the total payload mass carried by boosters launched by NASA (CRS)
- Display average payload mass carried by booster version F9 v1.1
- List the date when the first successful landing outcome in the ground pad was achieved.
- List the names of the boosters which have success in drone ships and have payload mass greater than 4000 but less than 6000
- List the total number of successful and failed mission outcomes
- List the names of the booster_versions which have carried the maximum payload mass.
- List the failed landing_outcomes in drone ships, their booster versions, and launch site names for the year 2015
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the dates 2010-06-04 and 2017-03-20, in descending order

Full notebook is available :[Github URL to Notebook]

# Build an Interactive Map with Folium

- In this step, we managed to Mark all launch locations and place objects on the map such as markers, circles, and lines to mark the success or failure of launches for each site on the folium map.
- We placed color markers to identify which launch. locations to show which were successful or otherwise.
- We calculate distances between a launch site and its surroundings.

Full notebook is available :[Github URL to Notebook]

# Build a Dashboard with Plotly Dash

- We made a Dashboard with Plotly Dash containing the following items:

1. We plot pie charts showing total launches by specific sites.

2. We plot pie charts showing total launches by specific sites.

3. We plot a scatter plot showing the relationship between Result and Payload Mass (Kg) for different booster versions.

Full notebook is available :[Github URL to Notebook]

# Predictive Analysis (Classification)

- Manipulate the data
- Split data into training and testing datasets
- Choose parameters and algorithms for GridSearchCV
- Use the dataset split in the GridSearchCV object and train our dataset.
- Find the accuracy for each model that was used.
- Plot Confusion Matrix
- Find the model that has the best accuracy score.
- Full notebook is available :Github URL to Notebook

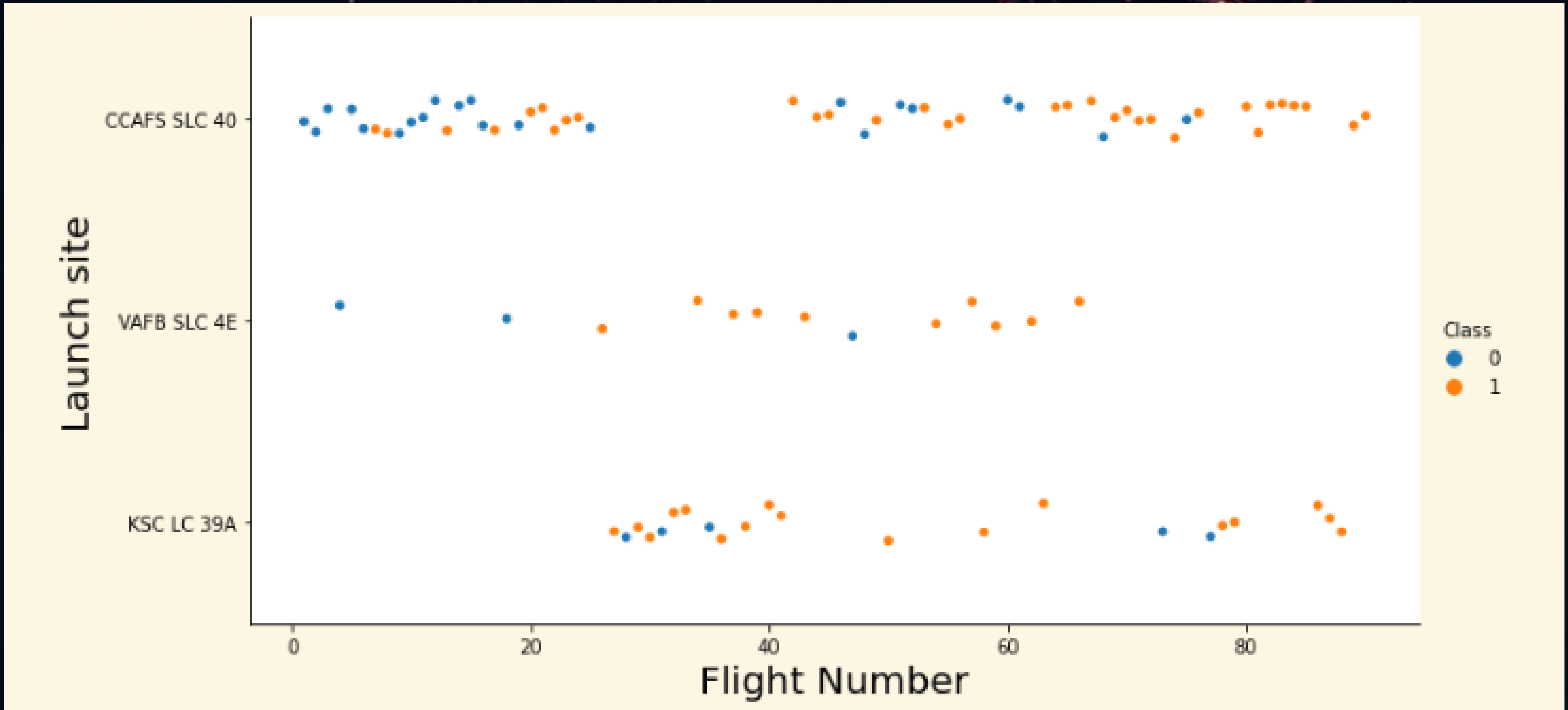
# Results

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

Section 2

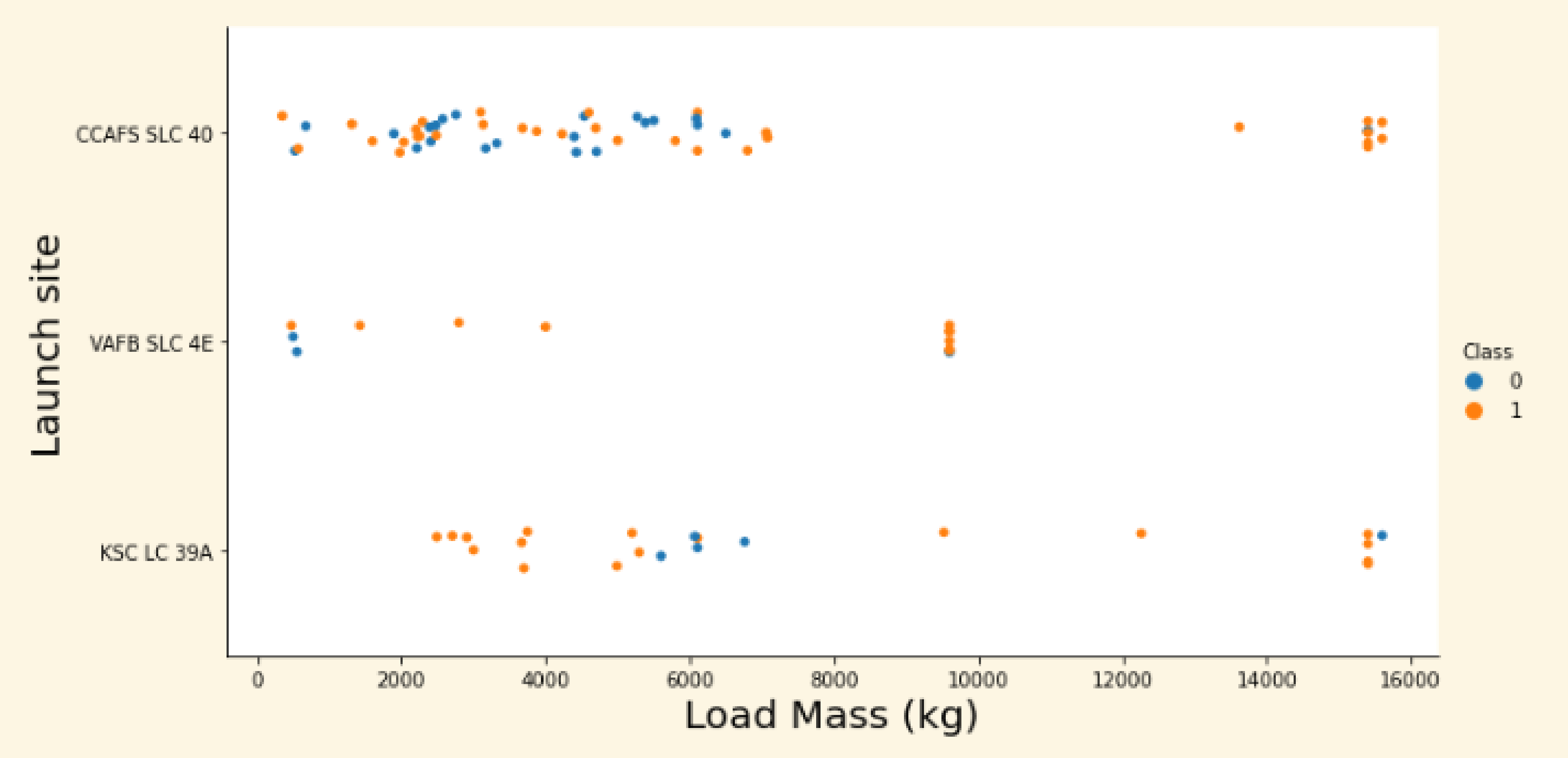
# Insights drawn from EDA

# Flight Number vs. Launch Site



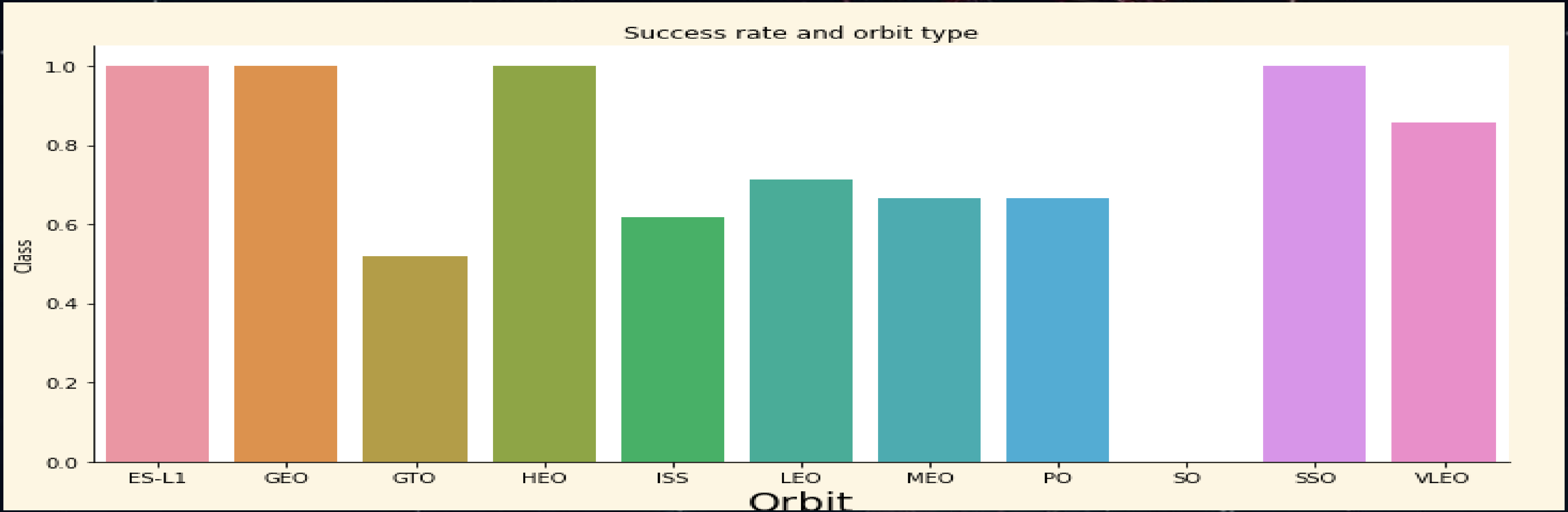

- Analyzing the graph, we can see that the greater the number of flights at a launch site, then the higher the success rate at that launch site.

# Payload vs. Launch Site



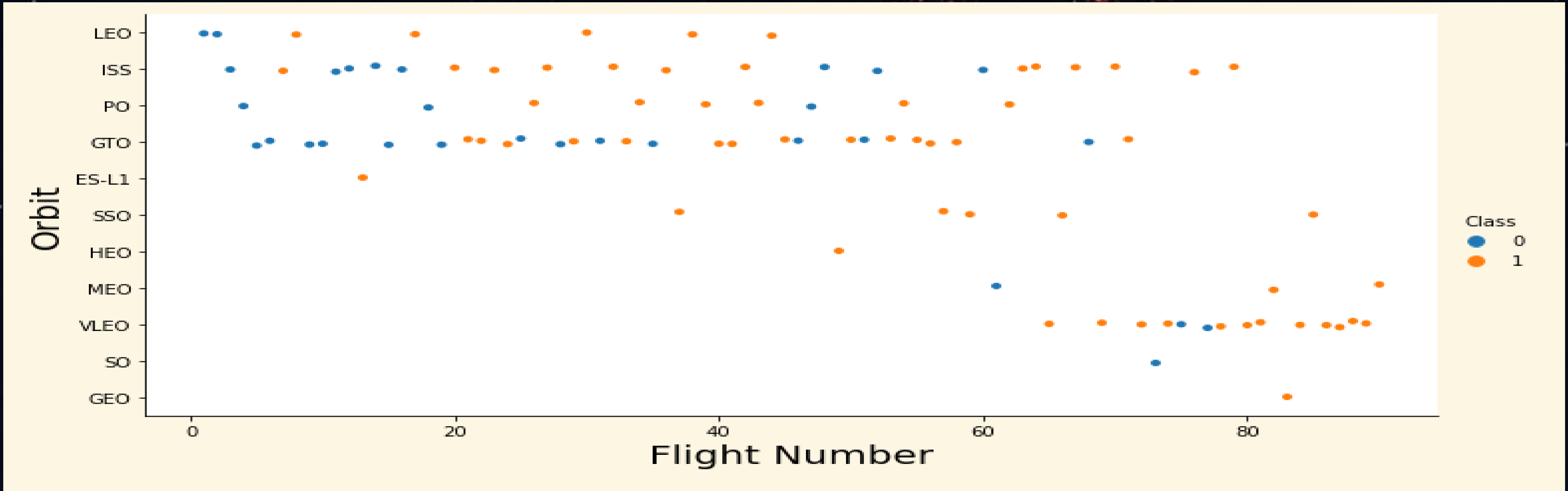

- Analyzing the graph, we can conclude that the higher the payload mass for the CCAFS SLC 40 Launch Site, the higher the success rate of the Rocket.

# Success Rate vs. Orbit Type



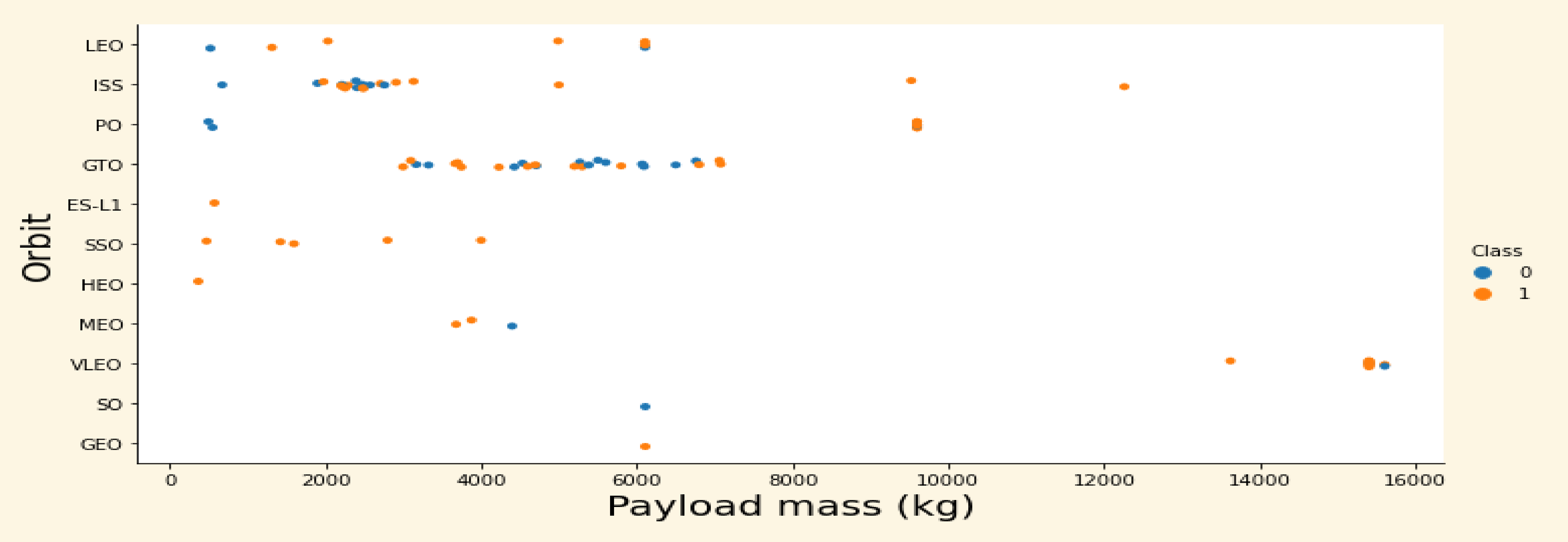

- Through the graph the Orbit ES-L1, GEO, HEO,SSO has the best success

# Flight Number vs. Orbit Type



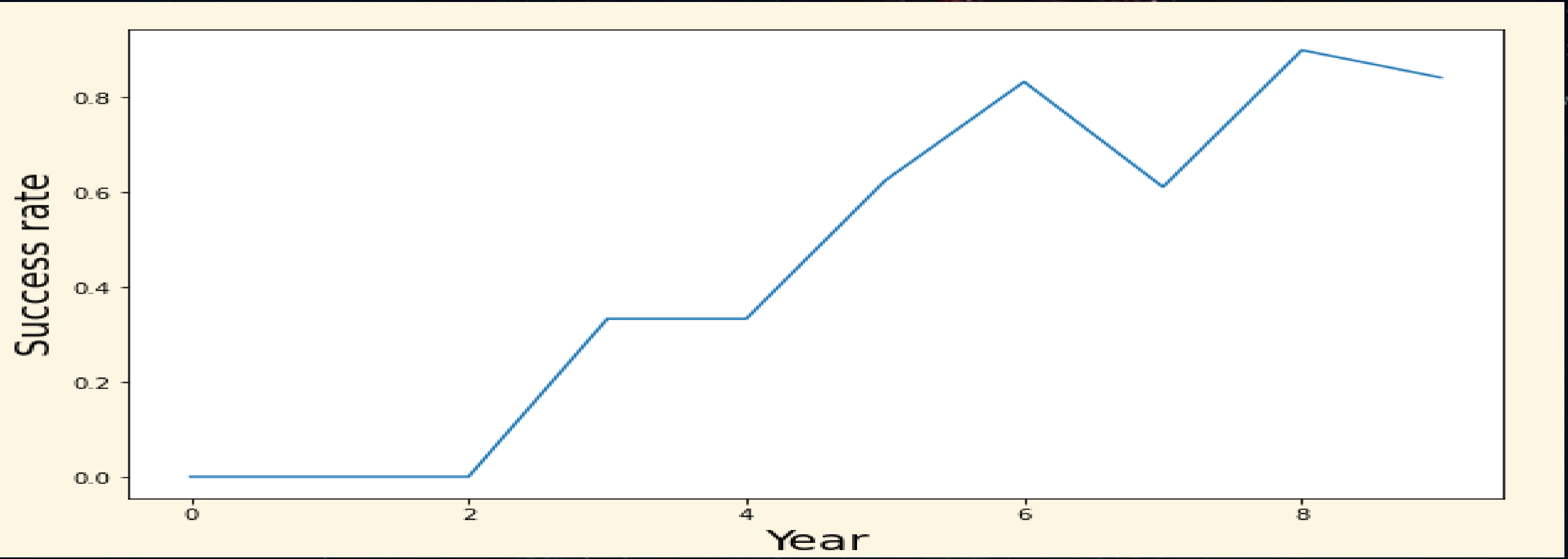

- Looking at the graph, the flight number vs. orbit type. we can find that in LEO orbit, the success is totally related to the number of flights, however in the GTO orbit, it cannot find no relationship between flight number and orbit.

# Payload vs. Orbit Type




- For heavier loads, successful landings are more for PO, LEO and ISS orbits.

# Launch Success Yearly Trend




- As of the year 2013, the success rate continues increasing until 2020.

# All Launch Site Names

- When we use DISTINCT in SQL is to show only unique column values, so find unique locations SpaceX launch

In 14

```
%sql SELECT  DISTINCT LAUNCH_SITE FROM SPACEXTBL;
```

```
 * ibm_db_sa://nzl21889:***@125f9f61-9715-46f9-9399-c8177
.databases.appdomain.cloud:30426/bludb
Done.
```

Out 14

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

# Launch Site Names Begin with 'CCA'

In 207

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

```
 * ibm_db_sa://nzl21889:***@125f9f61-9715-46f9-9399-c8177b21803b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30426/bludb
Done.
```

Out 207

| 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 |

- The above query was used to display 5 records where launch sites start with `CCA`

# Total Payload Mass

- We used the query opposite to find the total payload carried by NASA boosters as 45596.

In 58

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) AS SUM  FROM SPACEXTBL WHERE CUSTOMER LIKE 'NASA (CRS)';
```

```
 * ibm_db_sa://nzl21889:***@125f9f61-9715-46f9-9399-c8177b21803b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30426/bludb
Done.
```

Out 58

| SUM |
|---|
| 45596 |

# Average Payload Mass by F9 v1.1

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) AS AVG  FROM SPACEXTBL WHERE BOOSTER_VERSION LIKE 'F9 v1.1%';
```

```
 * ibm_db_sa://nzl21889:***@125f9f61-9715-46f9-9399-c8177b21803b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30426/bludb
Done.
```

| AVG |
|---|
| 2534 |

- We used the query above to calculate the average mass of cargo carried by F9 booster version v1.1

# First Successful Ground Landing Date

```
In 74   1  %sql SELECT MIN(DATE) AS DATE  FROM SPACEXTBL WHERE LANDING__OUTCOME LIKE 'Success (ground pad)';

 * ibm_db_sa://nzl21889:***@125f9f61-9715-46f9-9399-c8177b21803b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cl
Done.
```

Out 74

| DATE |
|---|
| 2015-12-22 |

- With the query above, we can find the date of first successful land landing result was December 22 2015

# Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql SELECT  BOOsTER_VERSION AS NAME FROM SPACEXTBL WHERE LANDING__OUTCOME LIKE 'Success (drone ship)' AND PAYLOAD_MASS__KG_> 4000 AND
 PAYLOAD_MASS__KG_ < 6000;
```

```
 * ibm_db_sa://nzl21889:***@125f9f61-9715-46f9-9399-c8177b21803b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30426/bludb
Done.
```

| name |
| --- |
| F9 FT B1022 |
| F9 FT B1026 |
| F9 FT B1021.2 |
| F9 FT B1031.2 |

- The above query was used to determine the successful landings with payload mass greater than 4000 but less of 6000

# Total Number of Successful and Failure Mission Outcomes

```
%sql SELECT MISSION_OUTCOME,COUNT(*) AS TOTAL FROM SPACEXTBL GROUP BY MISSION_OUTCOME;
```

```
 * ibm_db_sa://nzl21889:***@125f9f61-9715-46f9-9399-c8177b21803b.c1ogj3sd0tgtu0lqde00.dat
Done.
```

| mission_outcome | total |
|---|---|
| Failure (in flight) | 1 |
| Success | 99 |
| Success (payload status unclear) | 1 |

- We use the query above to find if it was hit or miss for Mission Outcomes

# Boosters Carried Maximum Payload

```
%sql SELECT  BOOsTER_VERSION AS NAME FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = ( SELECT  MAX(PAYLOAD_MASS__KG_)  FROM SPACEXTBL);
```

```
 * ibm_db_sa://nzl21889:***@125f9f61-9715-46f9-9399-c8177b21803b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30426/bludb
Done.
```

| name |
|---|
| 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 |

- The above query was used to find the Booster version name with a maximum payload

# 2015 Launch Records

```
%sql SELECT  BOOSTER_VERSION,LANDING__OUTCOME, LAUNCH_SITE AS NAME FROM SPACEXTBL WHERE LANDING__OUTCOME LIKE 'Failure (drone ship)' AND DATE LIKE
 '2015-%' ;
```

```
 * ibm_db_sa://nzl21889:***@125f9f61-9715-46f9-9399-c8177b21803b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30426/bludb
Done.
```

| booster_version | landing__outcome | name |
|---|---|---|
| F9 v1.1 B1012 | Failure (drone ship) | CCAFS LC-40 |
| F9 v1.1 B1015 | Failure (drone ship) | CCAFS LC-40 |

- We used the above query to find drone ship crash landing results, their boost versions and launch site names for year 2015

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

```
%sql SELECT LANDING__OUTCOME, COUNT(*) AS TOTAL FROM SPACEXTBL WHERE DATE >= '2010-06-04' AND DATE <= '2017-03-20' GROUP BY LANDING__OUTCOME ORDER
 BY TOTAL DESC;
```

```
 * ibm_db_sa://nzl21889:***@125f9f61-9715-46f9-9399-c8177b21803b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30426/bludb
Done.
```

| landing__outcome | total |
|---|---|
| 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 |

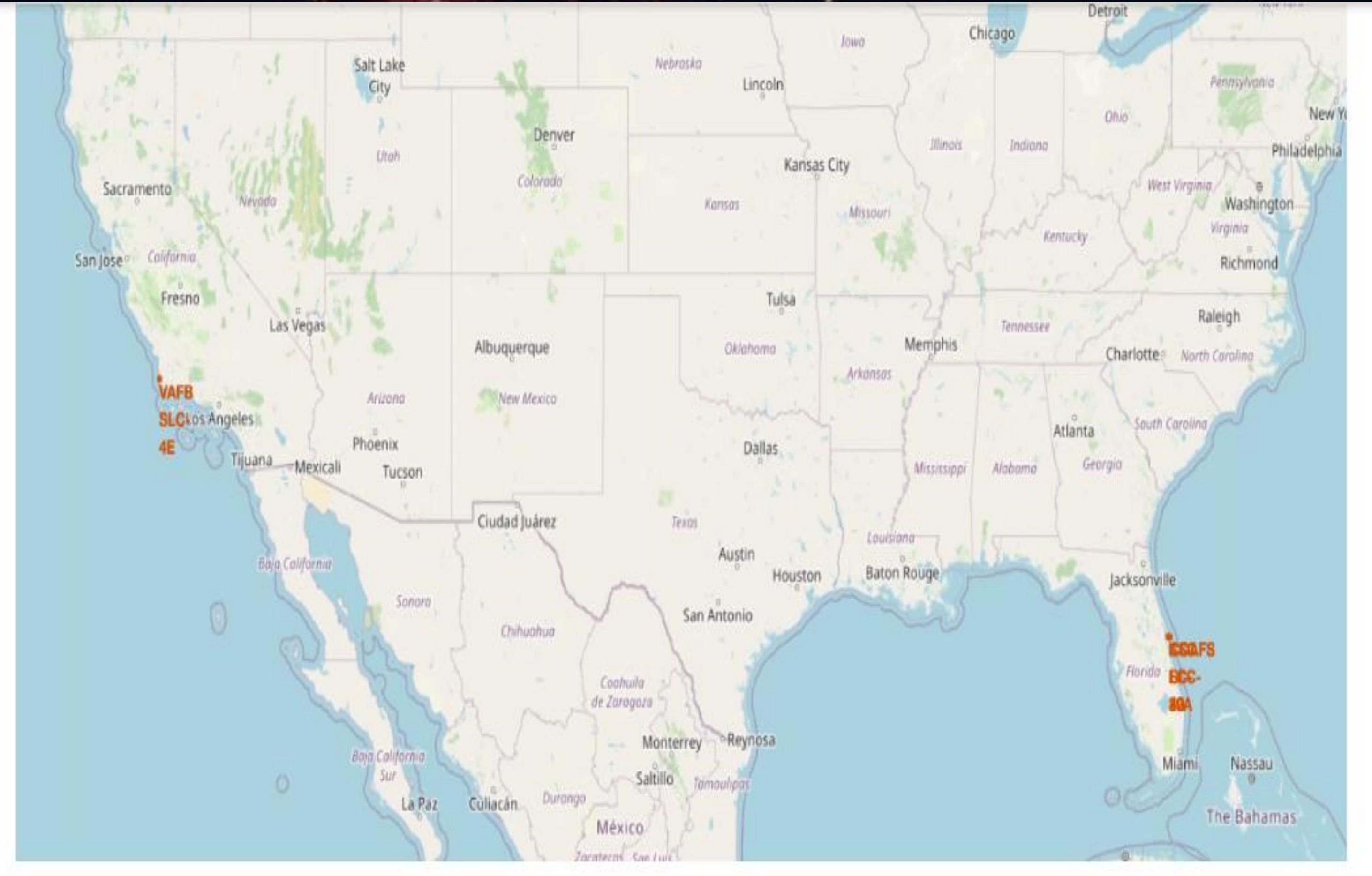
- We use the above query to find the landing results and the landing number in between 2010-06-04 to 2010-03-20 in descending order.

Section 3

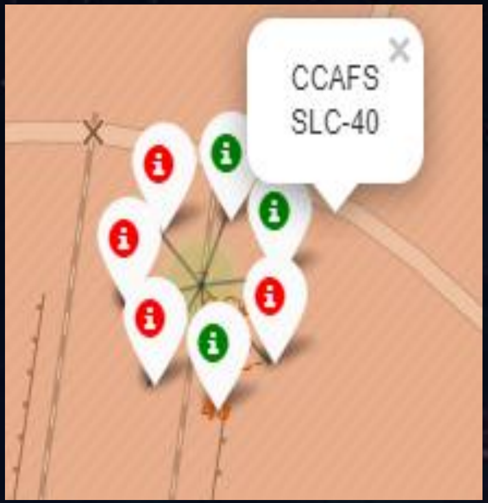
# Launch Sites Proximities Analysis

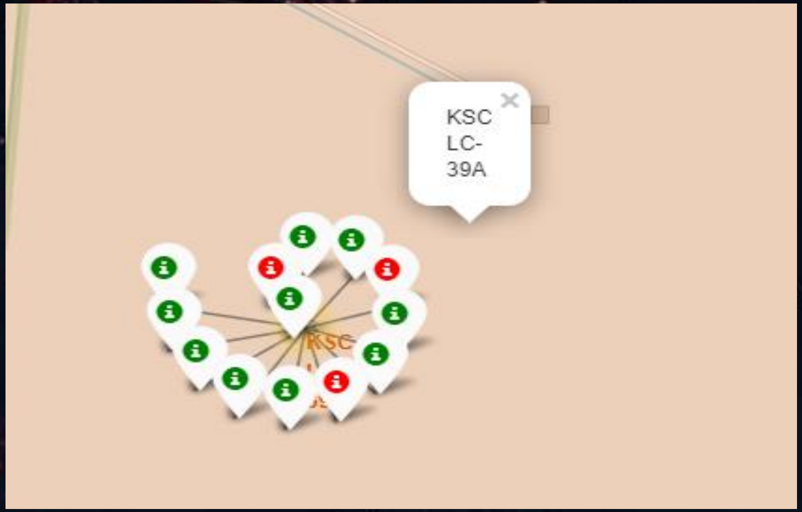
# Map of Launch Site Locations

- SpaceX has a launch site in Southern California and also on the Atlantic coast of Florida

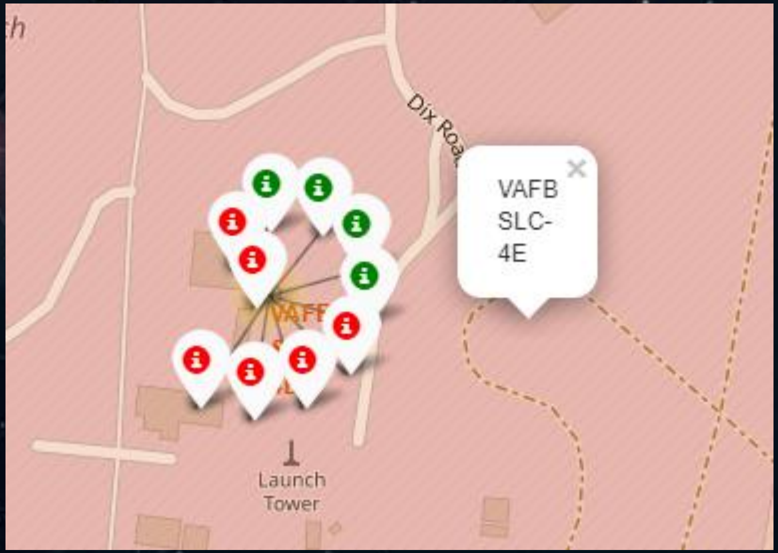


# Markers showing success and failure launch sites

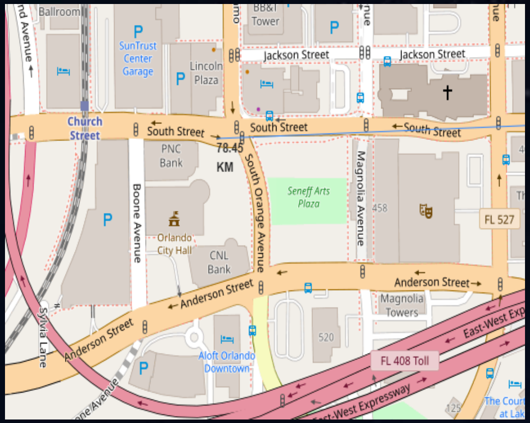



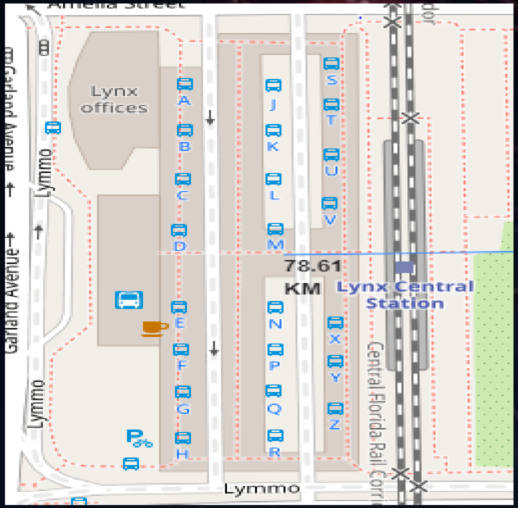



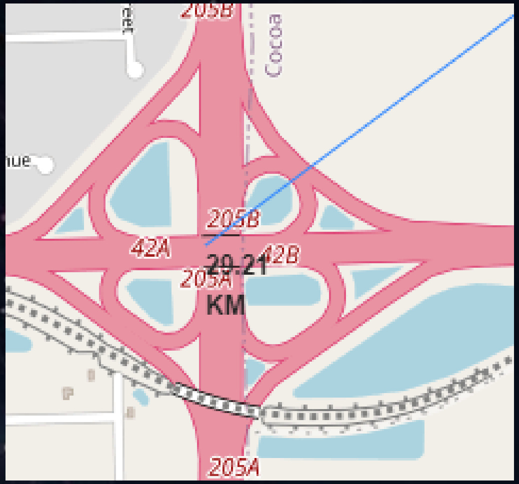

- The first three figures represent the Florida Launch Site, and the fourth figure below represents the California Launch Site. The green marks show successful launches, and the red marks show failures.

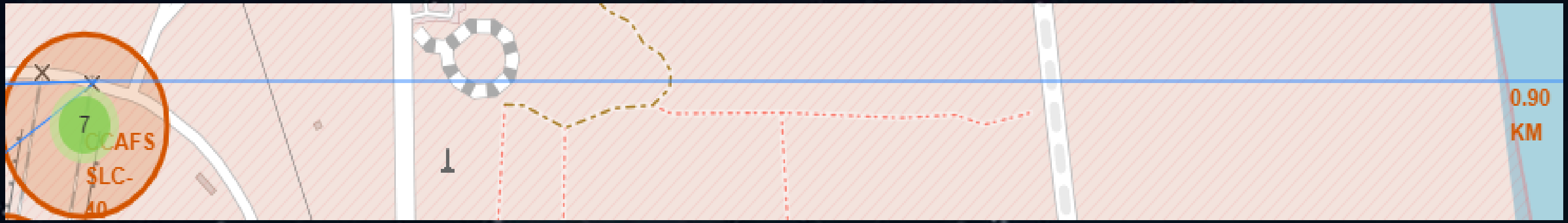


# <Folium Map Screenshot 3>



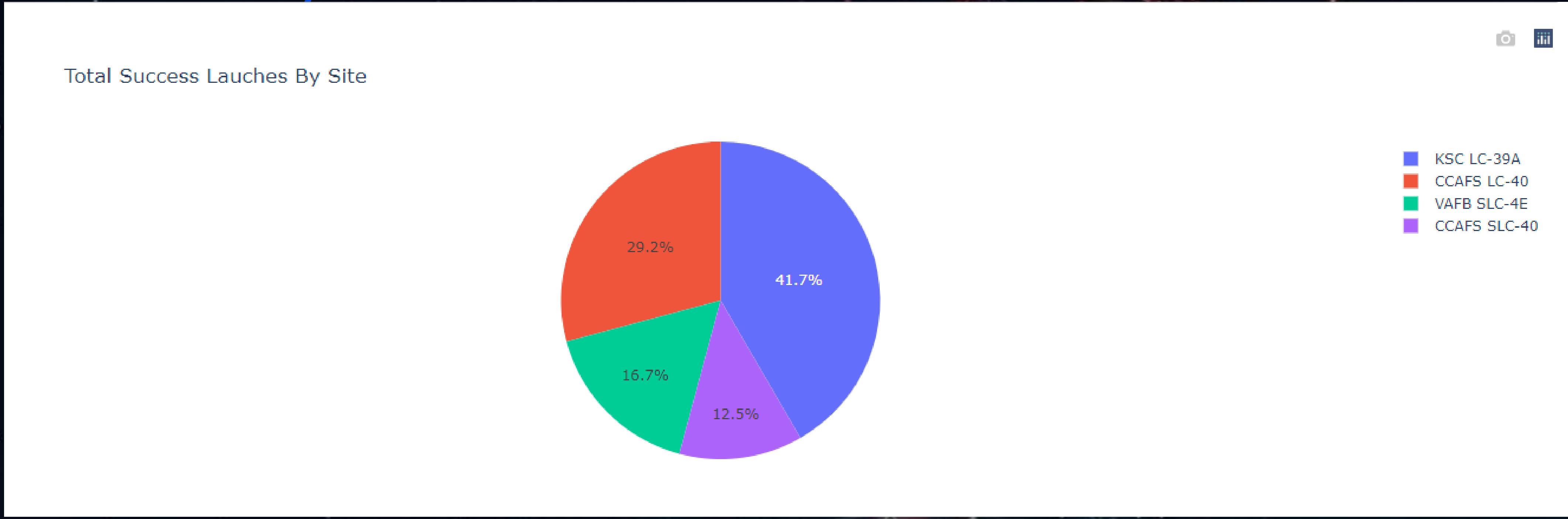




- Each blue line represents the distance to the nearest
  1. Distance coastline
  2. Distance Highway
  3. Distance to city
  4. Distance to Railway Station

Section 4

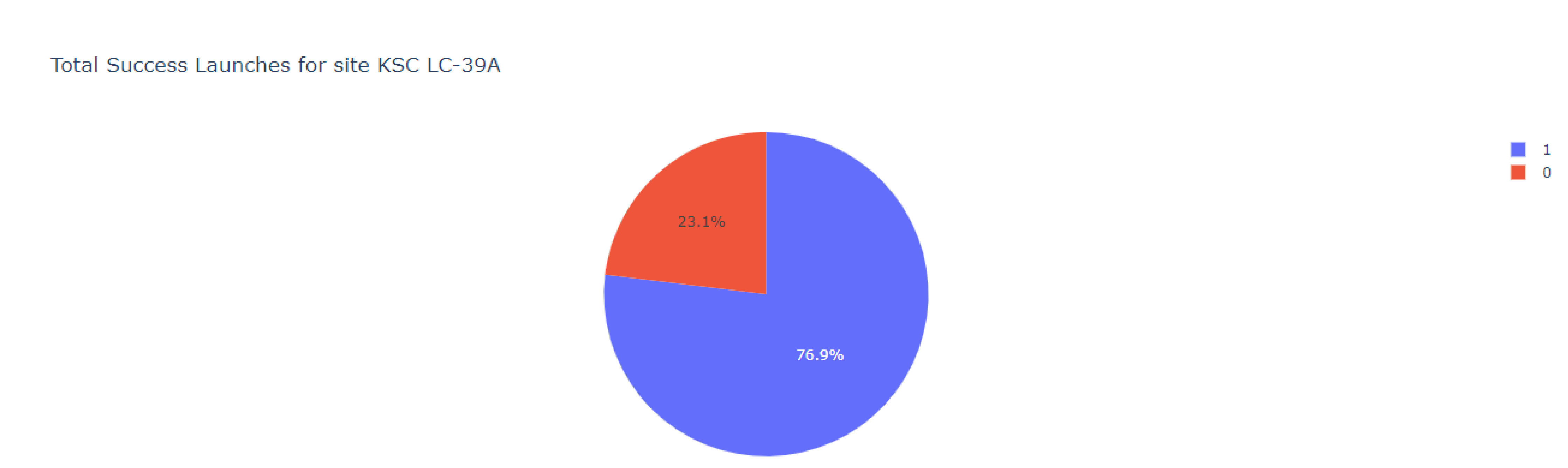
# Build a Dashboard with Plotly Dash

# Pie chart showing percentage of Total successful launches by site




- The graph shows the successful outcome of the launch site. And it also shows that KSC LC 39A is the location of more successful SpaceX launches.

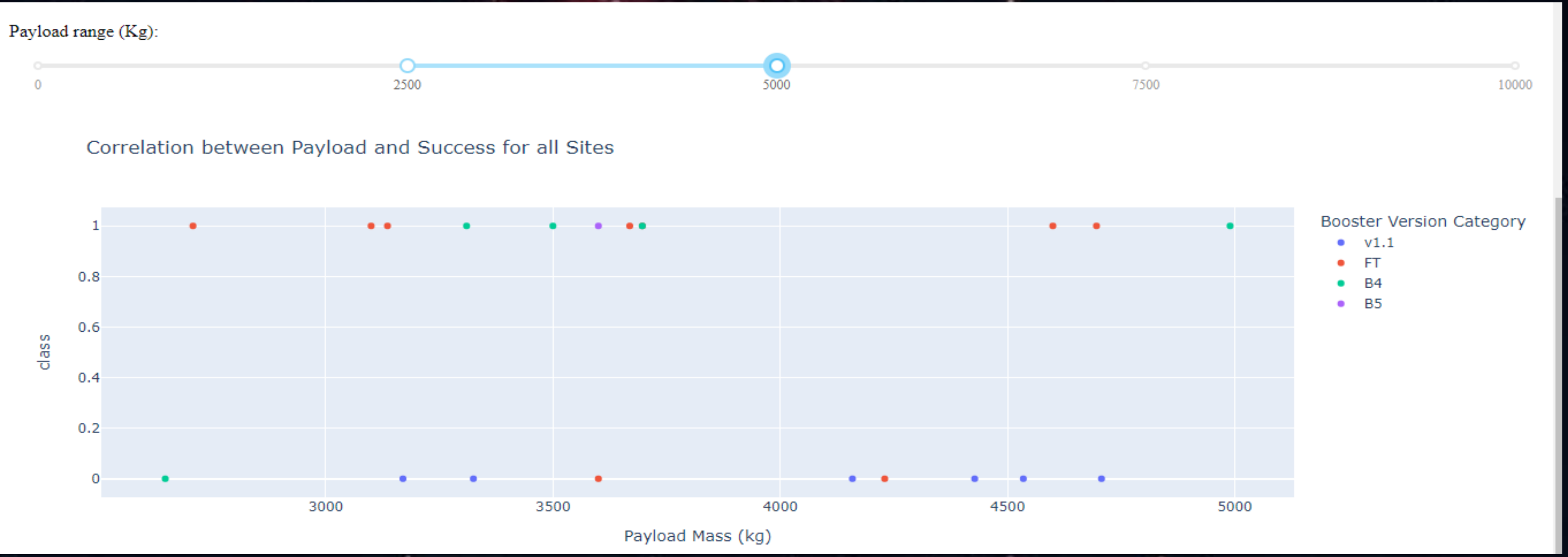
# Most successful launch site

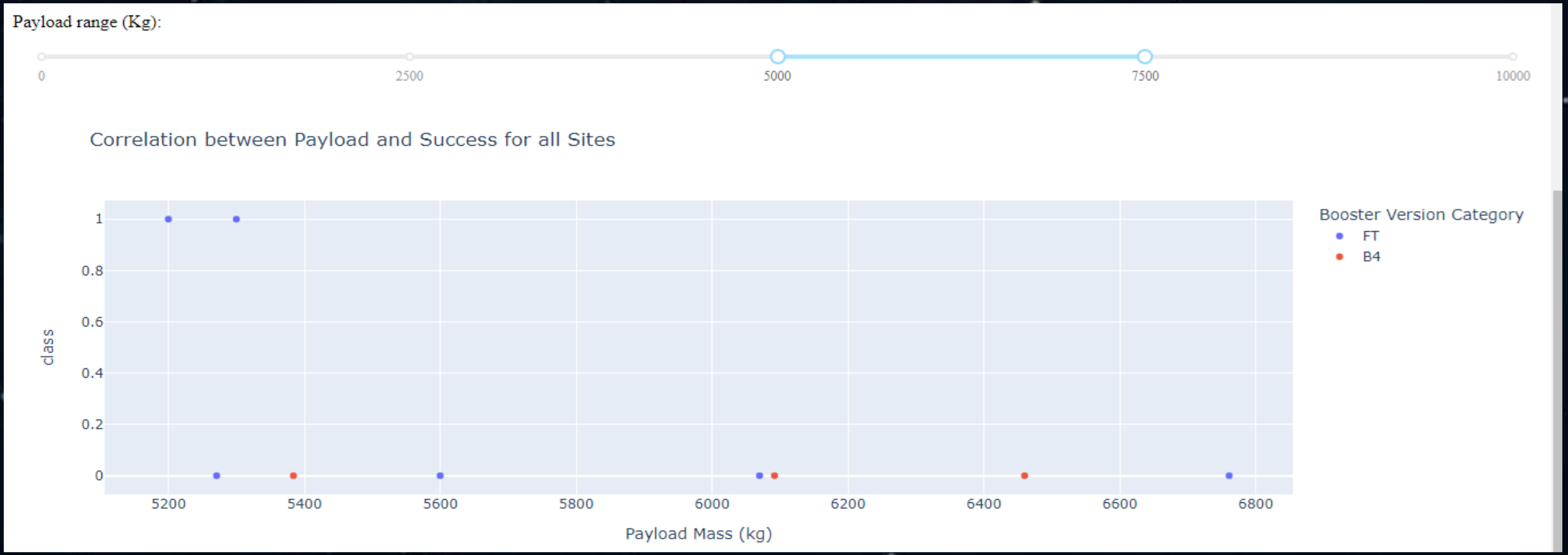



- The graph shows us the successful results for KSCLC-39A. This site has a good percentage of success as a result.

# Payload Mass vs. Launch Outcome Scatter Plot

- The graphs show the relationship between the mass of the load and the launch results per booster version.
- Filtering the mass in the range of 2500 - 5000 kg, we observed that the V1.1 version had more failures, and the FT version was more successful.
- When we increase the load range from 5000 to 7500 kg, for heavier loads, we have two versions, FT and B4, so we can observe that between 5200-5400 kg, the FT version was more successful, but when the load increases, we did not observe any success in both versions.



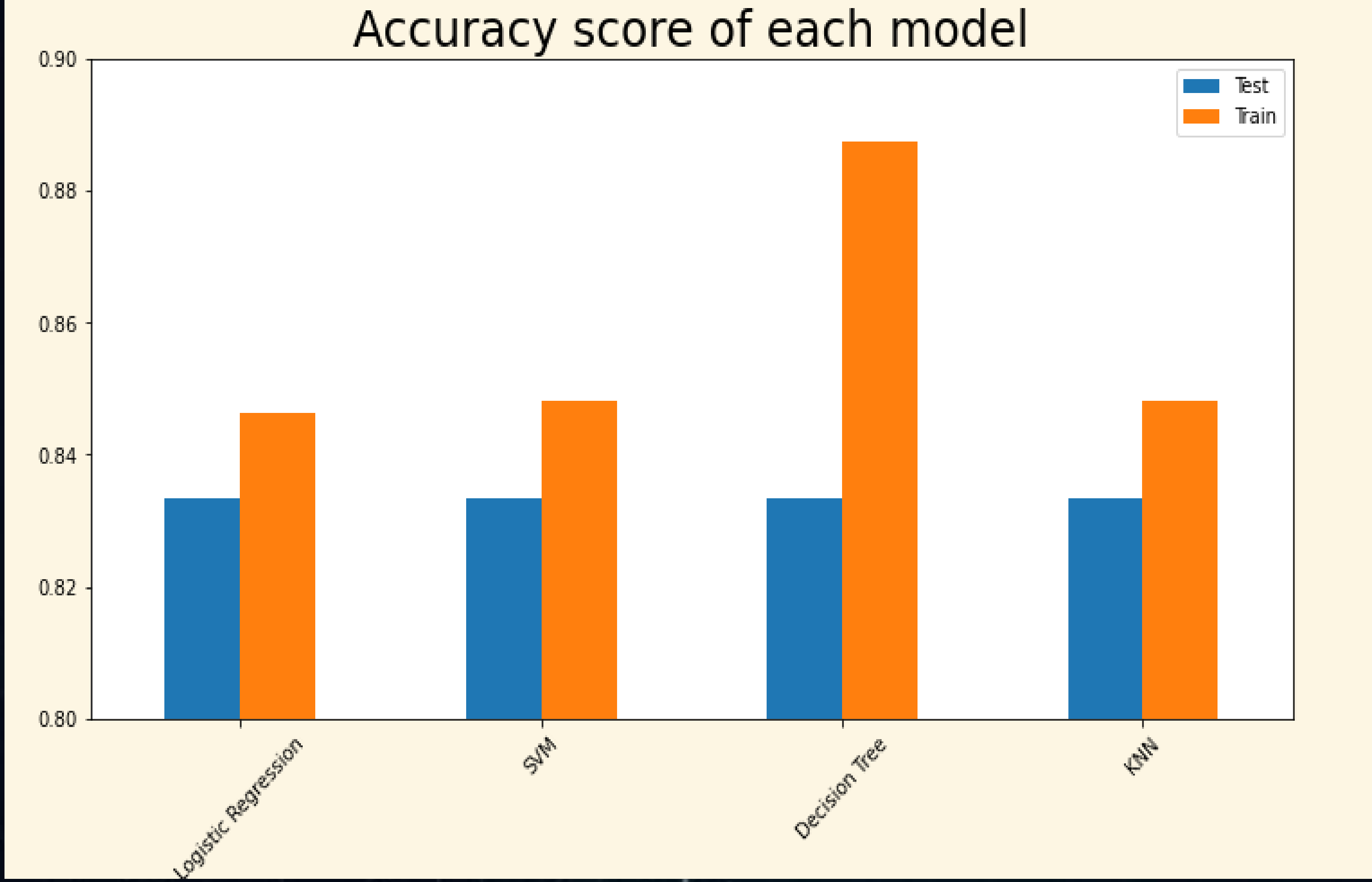



Section 5

# Predictive Analysis (Classification)

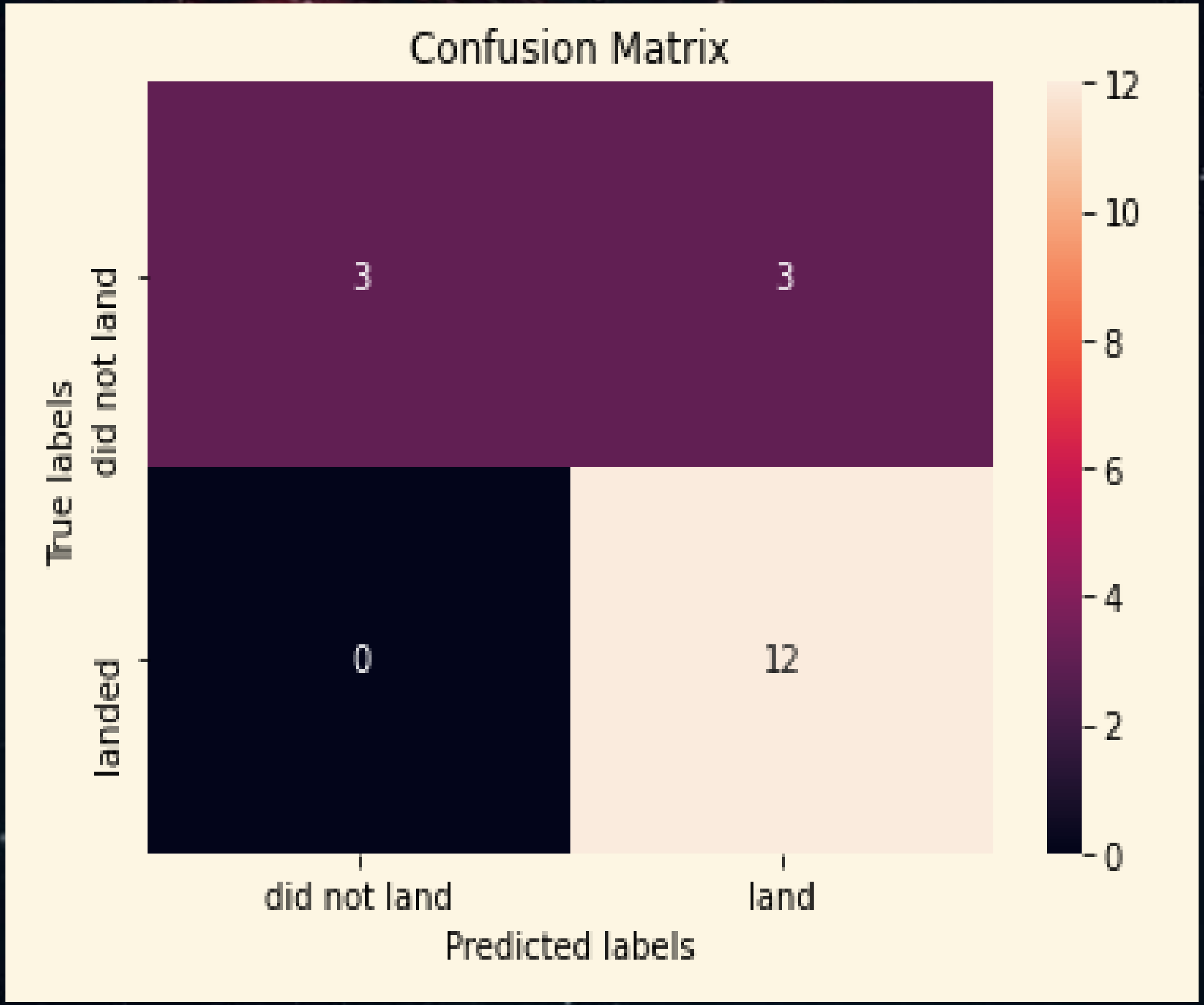
# Classification Accuracy

- The models used presented the same precision value for the test set.
- In the train set, the decision tree model presented a better precision compared to the other models, and this happens because the decision tree makes finding an accuracy using this set, but when using the test set, its accuracy differs very.



# Confusion Matrix

- All models presented the same confusion matrix. And we can see that the models were able to predict that the first stage did not land. The array has a high sensitivity, thus finding false positives, i.e., unsuccessful landings being marked as successful.



# Conclusions

We can conclude that all classification models, especially the decision tree, obtained good results for the training data, and for the test data, the models obtained the same results.

We also observed that for light payloads, we found better performance than heavier payloads.

SpaceX launch successes are directly proportional to the years they eventually were perfecting the releases.

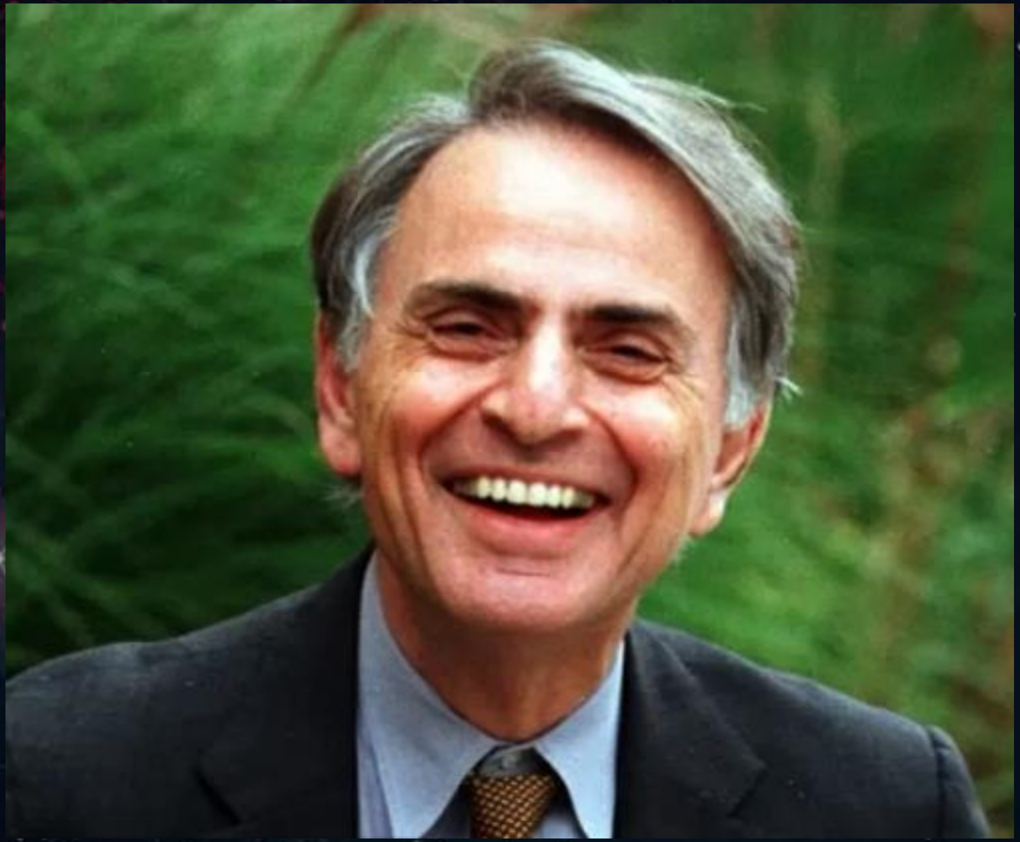
- We can see that the KSC LC-39A had the highest successful launches of all sites.
- Orbit ES-L1, HEO, GEO, and SSO have the best hits.

Therefore, all models used had very high confidence in their prediction.

# References


- NASA. (2011, August) Falcon 9 Launch Vehicle NAFCOM Cost Estimates. https://www.nasa.gov/pdf/586023main_8-3-11_NAFCOM.pdf
- Matthew Weaver, "Welcome back, baby:́ Elon Musk celebrates SpaceX rocket launch - and landing," https://www.theguardian.com/science/2015/dec/22/welcomeback-baby-elon-musk-celebrates-spacex-rocket-launch-and-landingSpaceX,
- "Background on Tonightś Launch,". http://www.spacex.com/news/2015/12/21/background-tonights-launch
- Samantha Masunaga. Elon Musk posts blooper reel of crashing SpaceX rockets. http://www.latimes.com/business/la-fi-spacex-bloopers-20170914-story.html

“Somewhere, something incredible is waiting to be known.” Carl Sagan



# Thank you!