IBM Developer
SKILLS NETWORK

# Winning Space Race with Data Science

By Taiwo Ariyo
August 6, 2024

# Outline

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

# Executive Summary

- Space X is known to cut the cost of rocket launches by reusing the first stage of its rockets, this project aims to predict whether the first stage of a rocket would land by using data such as payload mass, orbit, and launch site.
- This project uses machine learning algorithms like logistical regression, KNN, support vector machines (SVM), decision tree classifiers, and visualization tools to derive a model and make predictions. Results from these predictions showed that at specific LaunchSites, Orbit, and Payload mass, the chances of a successful landing were higher.

# Introduction

- Data has become very pivotal in business development and success. In this project, we utilized rocket launch data from Space X API. We executed some preprocessing on the data and performed queries using SQL to derive insights. After these, we employed visualization techniques such as scatter plots, line plots, and bar charts to understand the relationship between variables like payload mass, orbits, and launch site. This project utilizes machine learning to make predictions about the outcomes of the success or failure of a rocket landing.
- It is important to determine the chances of a successful landing. This is necessary to strategically allocate resources and reduce costs. The results from this project provide us with answers to questions such as: Which location is best to carry out a successful landing? What mass of the payload provides the best chances of a successful landing? What orbit is best for a successful landing?

Section 1

# Methodology

# Methodology

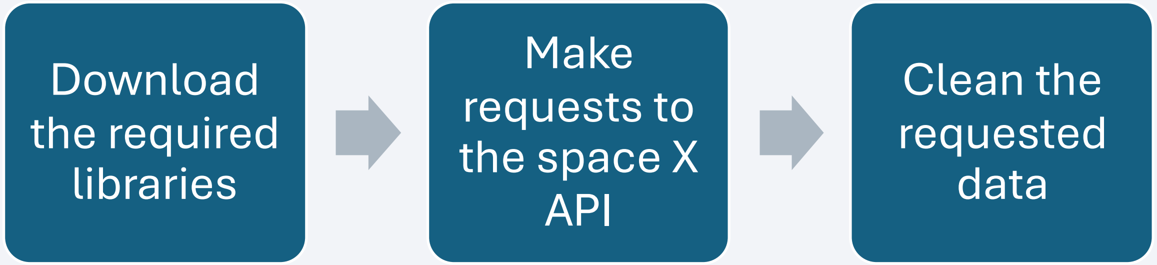
## Executive Summary

- Data collection methodology:
  Helper functions and requests were made to the SpaceX API to collect required data in a tabular format.
- Perform data wrangling:
  Useful information such as Booster Version, Pay Load Mass, Launch site, Coordinates, and so on was extracted from the data and set in columns. Data standardization, cleaning, and dealing with missing values were performed. This enables querying and gaining insights from the data to be easier accomplished.
- Perform exploratory data analysis (EDA) using visualization and SQL:
  Querying the data using SQL enabled us to extract and compare important variables that will aid in our understanding of how variables relate to one another.
- Perform interactive visual analytics using Folium, Plotly, and Dash:
  Utilization of these visualization tools allowed us to inspect variables and determine the effect of one variable over the other. Also, we see how these variables affect our desired outcome. Our desired outcome is a successful rocket landing.
- Perform predictive analysis using classification models:
  Upon standardizing the data, we used Support Vector Machines (SVM) and Classification models like Logistic Regression and Classification trees to make predictions of a successful landing outcome.
- How to build, tune, and evaluate classification models:

To achieve a functioning model, we found the best hyperparameter for the classification model. We split our data into training sets and test sets. The training set is used to train the model while the test set confirms for us that the model is functioning properly. We then perform a model evaluation to find the method that performs best.

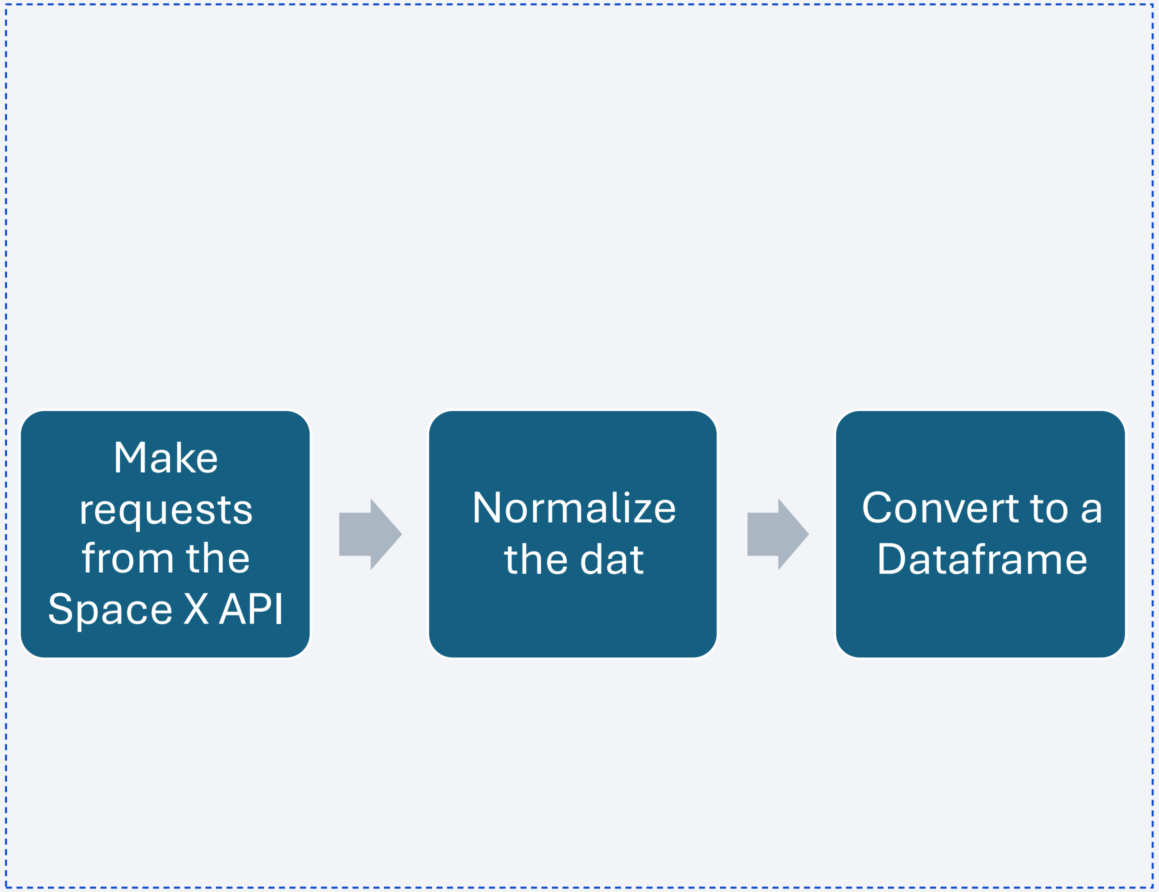
# Data Collection

- Useful data was collected by making API calls to the  Space X API.



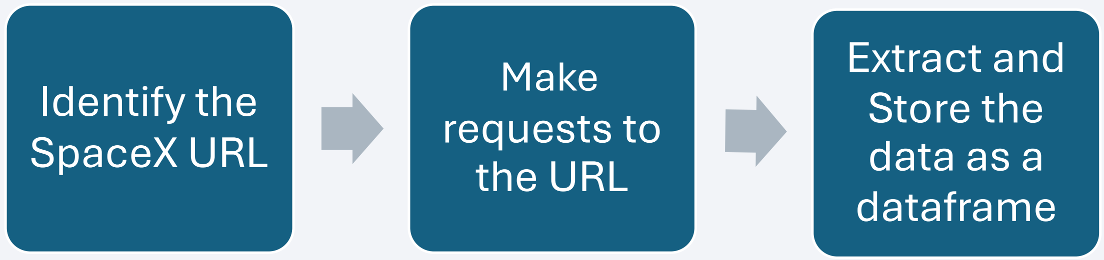
# Data Collection – SpaceX API

- To collect data, we request the SpaceX API and store them in a table using pandas. We also normalize the JSON data.
- Add the GitHub URL of the completed SpaceX API calls notebook (must include completed code cell and outcome cell), as an external reference and peer-review purpose



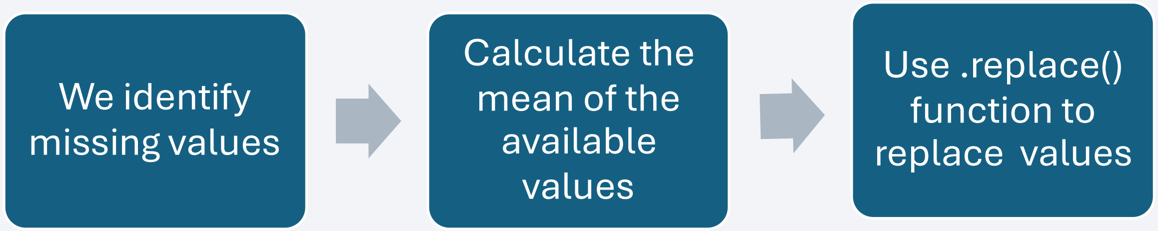
# Data Collection - Scraping

- We require specific variables such as Booster version, payload mass e.t.c for analysis and prediction for this project. Specific requests were made to the SpaceX URL during the data scraping process.
- Add the GitHub URL of the completed web scraping notebook, as an external reference and peer-review purpose



# Data Wrangling

- Data processing included dealing with missing values like the Payload mass and replacing them with the mean. This data-wrangling process included finding specific answers about the rocket launches and creating new columns such as the landing outcome as a one-hot encoded variable.




- Add the GitHub URL of your completed data wrangling related notebooks, as an external reference and peer-review purpose

# EDA with Data Visualization

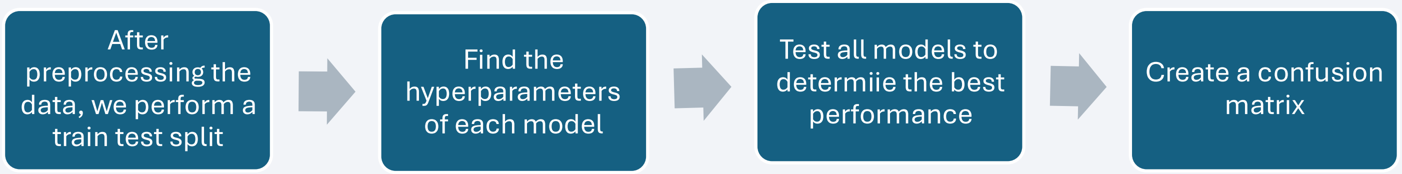
- The exploratory data analysis involved visualizing the launch data with Scatter plots, Bar Charts, and Line charts. This enabled us to see the relationship between one variable to another as well as for comparison purposes.
- Add the GitHub URL of your completed EDA with the data visualization notebook, as an external reference and peer-review purpose

# EDA with SQL

- jupyter-labs-eda-sql-coursera_sqllite.ipynb

# Predictive Analysis (Classification)

- To make predictions, we utilized classification models like Logistic Regression, K.N.N, S.V.M, and Decision trees classifier. To achieve this, the data was standardized and split into training and test sets using the train_split_test function. Then, we find the best hyperparameter for our models with the GridSearchCV function. The best method is determined using the hyperparameter values and test data. The confusion matrix is displayed to understand the model performance better.




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

# Results

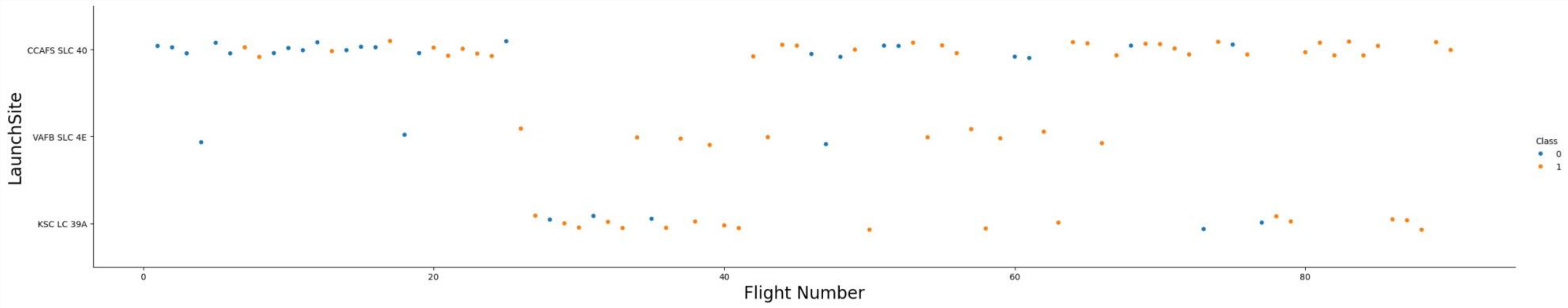
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

# Insights drawn from EDA

# Flight Number vs. Launch Site

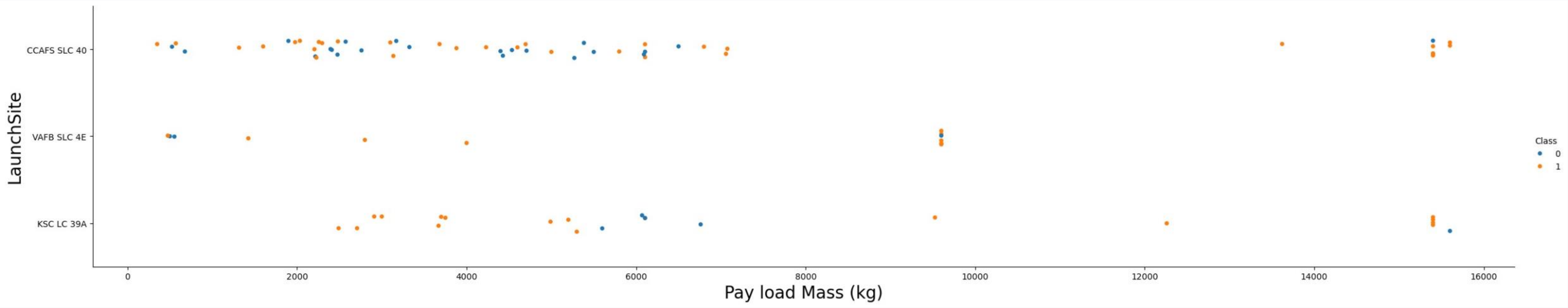
- A scatter plot of Flight Number vs. Launch Site

- 


- From the above data, we can derive that launch sites KSC LC 39A and VAFB SLC 4E have higher successful landings than CCAFS SLC 40.

# Payload vs. Launch Site

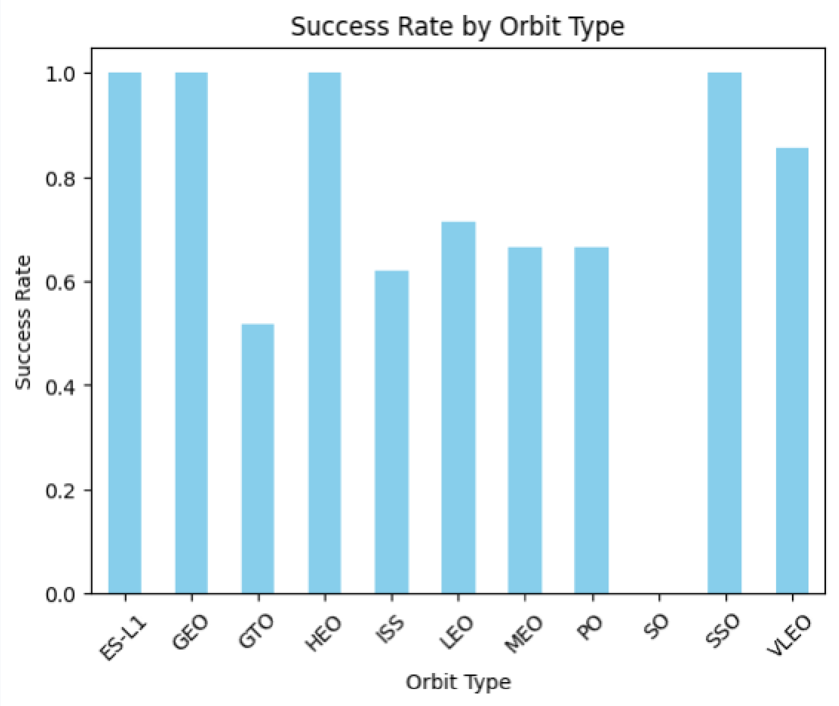
- A scatter plot of Payload vs. Launch Site




- A careful observation of the above chart shows that for launch site VAFB SLC 4E, there are no rockets launched for heavy payload mass(greater than 10000).

# Success Rate vs. Orbit Type

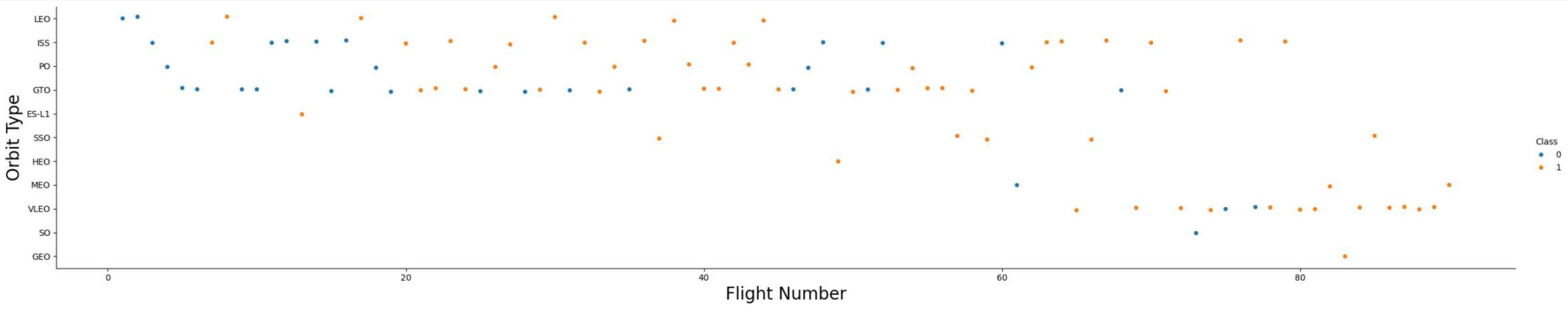
- A bar chart for the success rate of each orbit type

- 


- The bar chart shows the orbits with the highest success rates are ES-11, GEO, HEO, and SSO

# Flight Number vs. Orbit Type

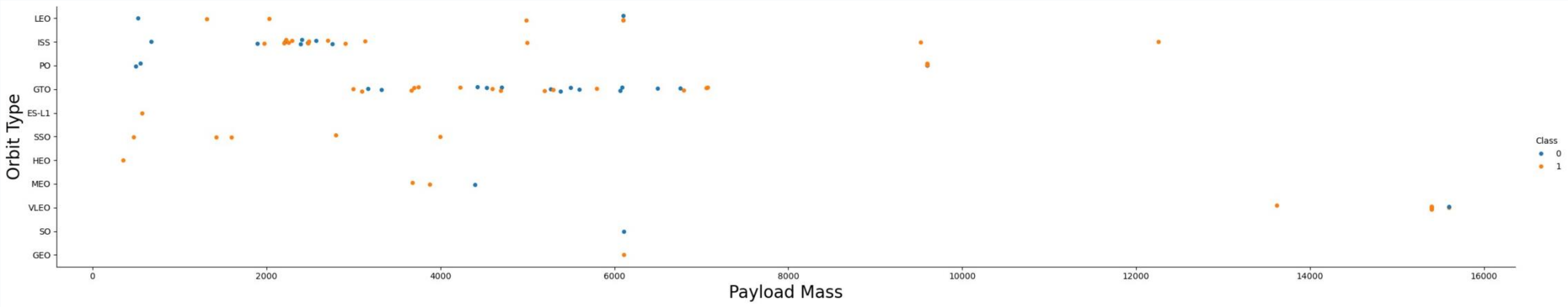
- A scatter point of Flight number vs. Orbit type




- 

- The charts show in the LEO orbit, the chances of success seem to be related to the number of flights. Likewise, in the GTO orbit, we observe no relationship between flight number and the chances of success.

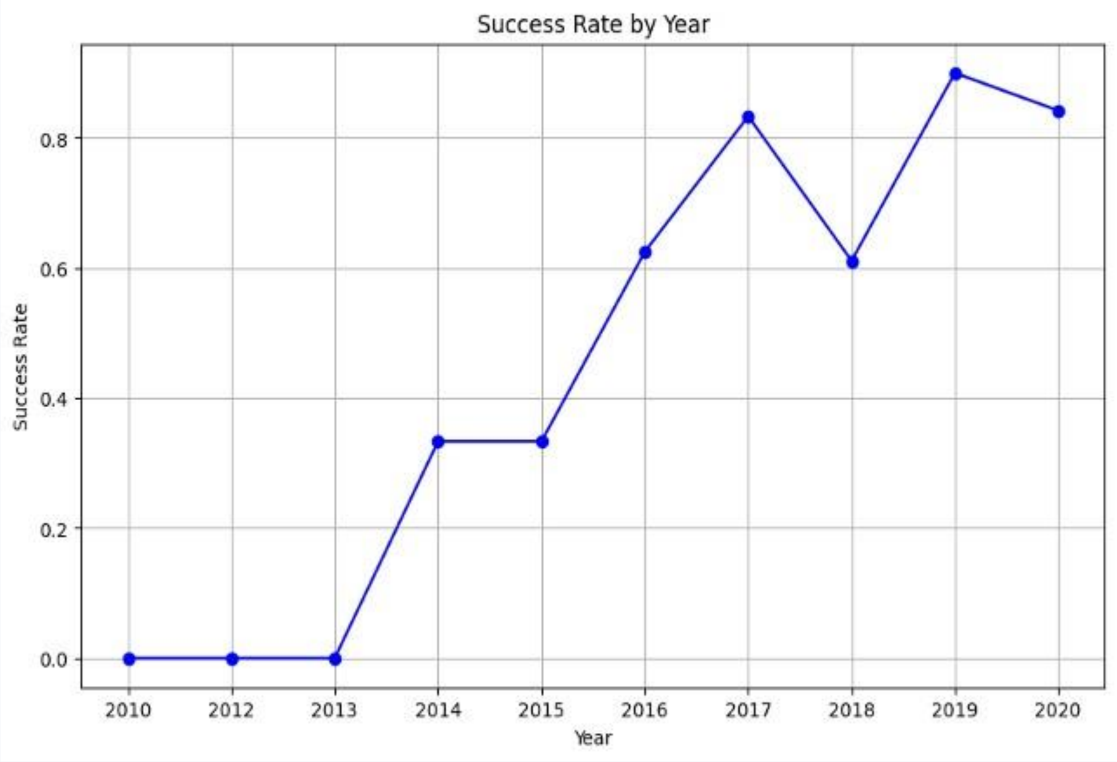
# Payload vs. Orbit Type

- Show a scatter point of payload vs. orbit type
- 


- From the above chart, heavier payloads have a more positive landing rate for orbits LEO, ISS, and Polar. In contrast, the chances of a positive landing for orbit GTO is not affected by payload mass

# Launch Success Yearly Trend

- A line chart of the yearly average success rate




- The line chart shows that the success rate since 2013 kept increasing till 2020

# All Launch Site Names

- To find the names of all unique launch sites, we query the data by stating:
- %sql select distinct Launch_Site from SPACEXTBL
- Therefore, we have names of launch sites such as CCAFS LC-40, VAFB SLC-4E, KSC, LC-39A, and CCAFS SLC-40.

# Launch Site Names Begin with 'CCA'

- To find 5 records where launch sites begin with `CCA` we used the query:
- %sql select *  from SPACEXTBL where Launch_Site like '%CCA%'limit 5
- This query outputs the data in a tabular form.

# Total Payload Mass

- To calculate the total payload carried by boosters from NASA we query the table:
- **%sql** select sum(PAYLOAD_MASS__KG_) from SPACEXTBL where customer='NASA (CRS)’
- The result of the query is 45596kg

# Average Payload Mass by F9 v1.1

- To calculate the average payload mass carried by booster version F9 v1.1 the query is as follows:
- **%sql** select avg(PAYLOAD_MASS__KG_) from SPACEXTBL where Booster_Version like '%F9 v1.1%’
- The result of the query is 2534.6666666666665kg

# First Successful Ground Landing Date

- To find the dates of the first successful landing outcome on ground pad we used the query as follows:
- **%sql** select min(date) from SPACEXTBL where Landing_Outcome like '%ground pad%’
- The query result is 2015-12-22

# Successful Drone Ship Landing with Payload between 4000 and 6000

- To list the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000, we query as follows:
- **%sql** select booster_version from SPACEXTBL where Landing_Outcome like 'Success (drone ship)' AND PAYLOAD_MASS__KG_<6000 AND PAYLOAD_MASS__KG_>4000
- The result of the query is, F9 FT B1022, F9 FT B1026, F9 FT B1021.2, F9 FT B1031.2

# Total Number of Successful and Failure Mission Outcomes

- To calculate the total number of successful and failure mission outcomes the query required is:
- **%sql** select Mission_Outcome,count(*) from SPACEXTBL group by Mission_Outcome
- The query result is:

Failure (in flight)
1
Success
98
Success
11
Success (payload status unclear)

# Boosters Carried Maximum Payload

- To list the names of the booster which have carried the maximum payload mass, we query as follows:
- **%sql** select distinct Booster_Version from SPACEXTBL where PAYLOAD_MASS__KG_=(select max(PAYLOAD_MASS__KG_) from SPACEXTBL)
- The result of our query is: 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

- To list the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015, the query that was used is:
- **%sql** select month(DATE) as month,Landing__Outcome,Booster_Version,Launch_Site from SPACEXTBL where landing__outcome like '%Success%ground%pad%' order by Month
- The result of our query is:

| month | substr(Date,0,5)=='2015' | Landing_Outcome | Booster_Version | Launch_Site |
|---|---|---|---|---|
| 01 | 1 | Failure (drone ship) | F9 v1.1 B1012 | CCAFS LC-40 |
| 04 | 1 | Failure (drone ship) | F9 v1.1 B1015 | CCAFS LC-40 |
| 01 | 0 | Failure (drone ship) | F9 v1.1 B1017 | VAFB SLC-4E |
| 03 | 0 | Failure (drone ship) | F9 FT B1020 | CCAFS LC-40 |
| 06 | 0 | Failure (drone ship) | F9 FT B1024 | CCAFS LC-40 |

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

- To 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 (*) from SPACEXTBL where DATE between '2010-06-04' and '2017-03-20' group by Landing_Outcome

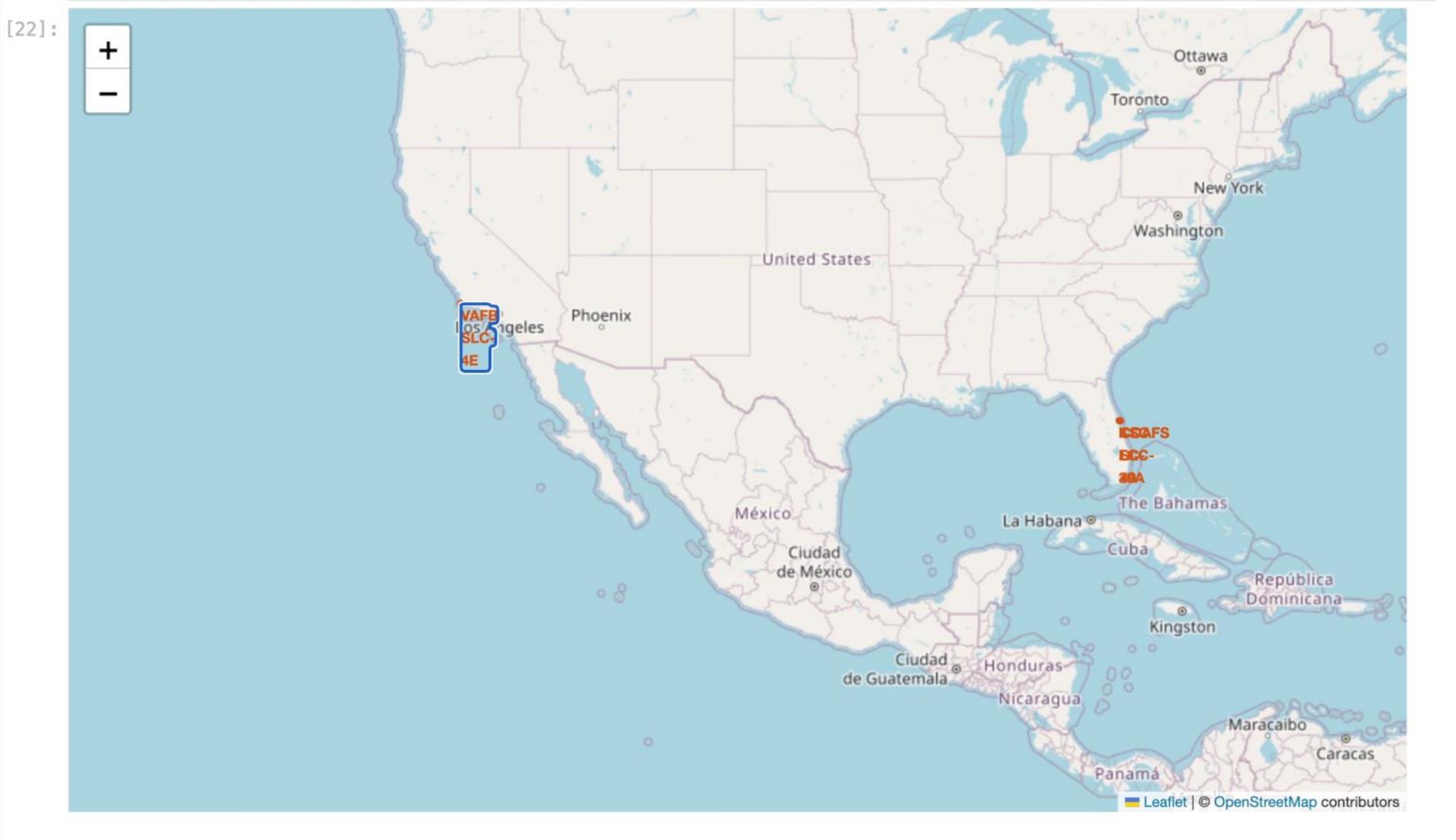
| Landing_Outcome | count (*) |
|---|---|
| Controlled (ocean) | 3 |
| Failure (drone ship) | 5 |
| Failure (parachute) | 2 |
| No attempt | 10 |
| Precluded (drone ship) | 1 |
| Success (drone ship) | 5 |
| Success (ground pad) | 3 |
| Uncontrolled (ocean) | 2 |

-

Section 3

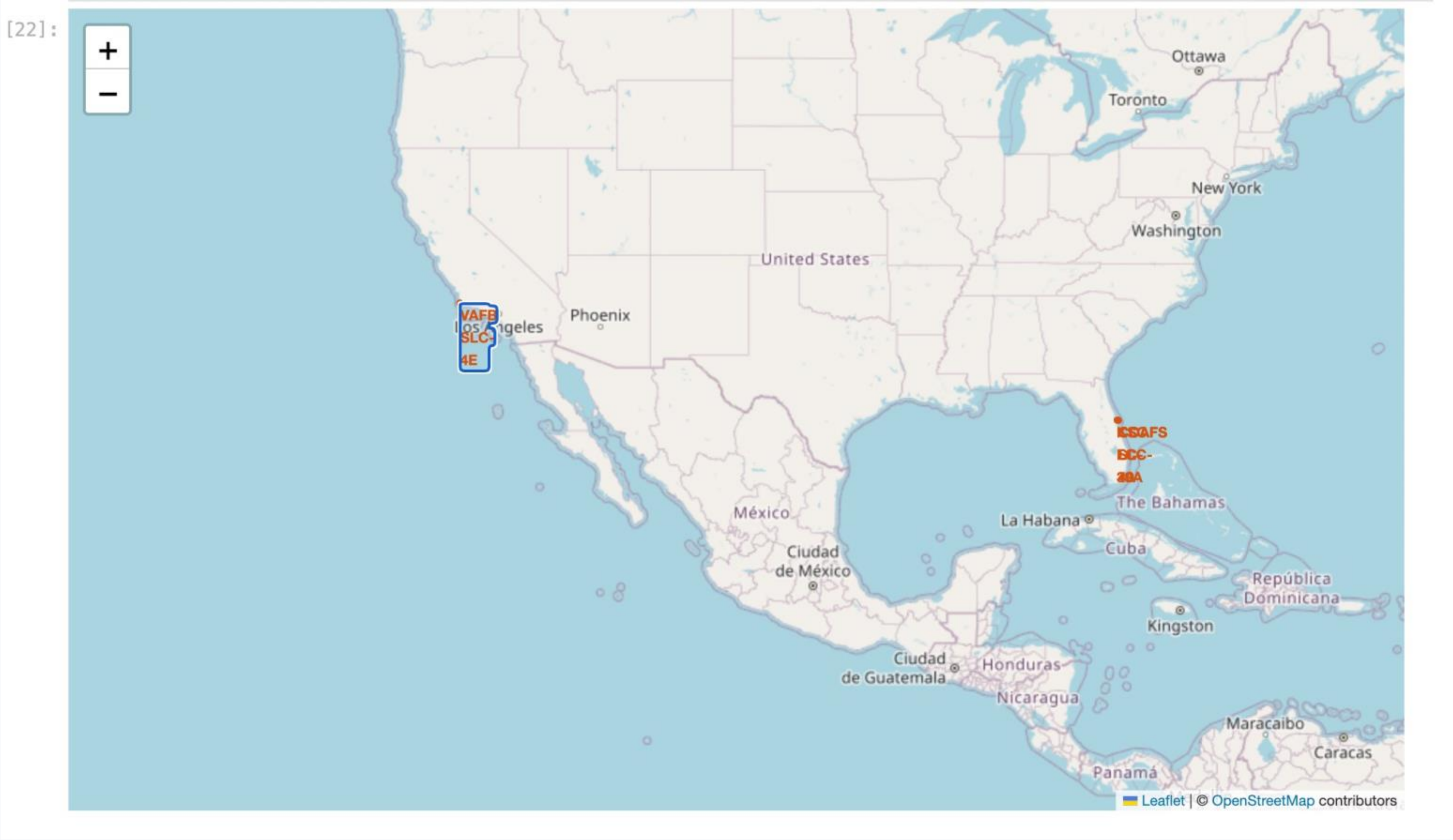
# Launch Sites Proximities Analysis

# Launch sites on a map Map



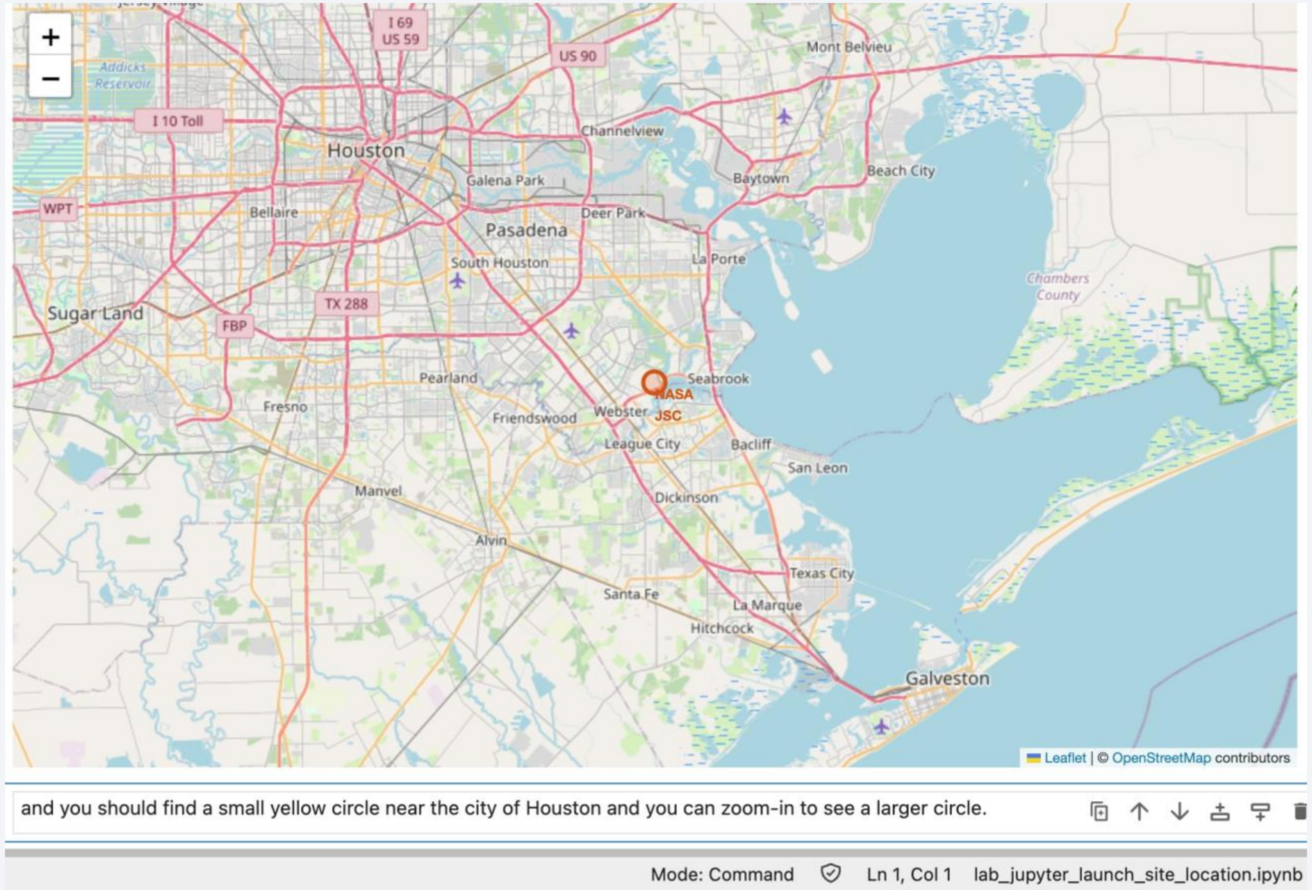

•

# Distances between a launch site to its proximities




•

# Site's location on a map using site's latitude and longitude coordinates



Section 4

# Build a Dashboard with Plotly Dash

Section 5

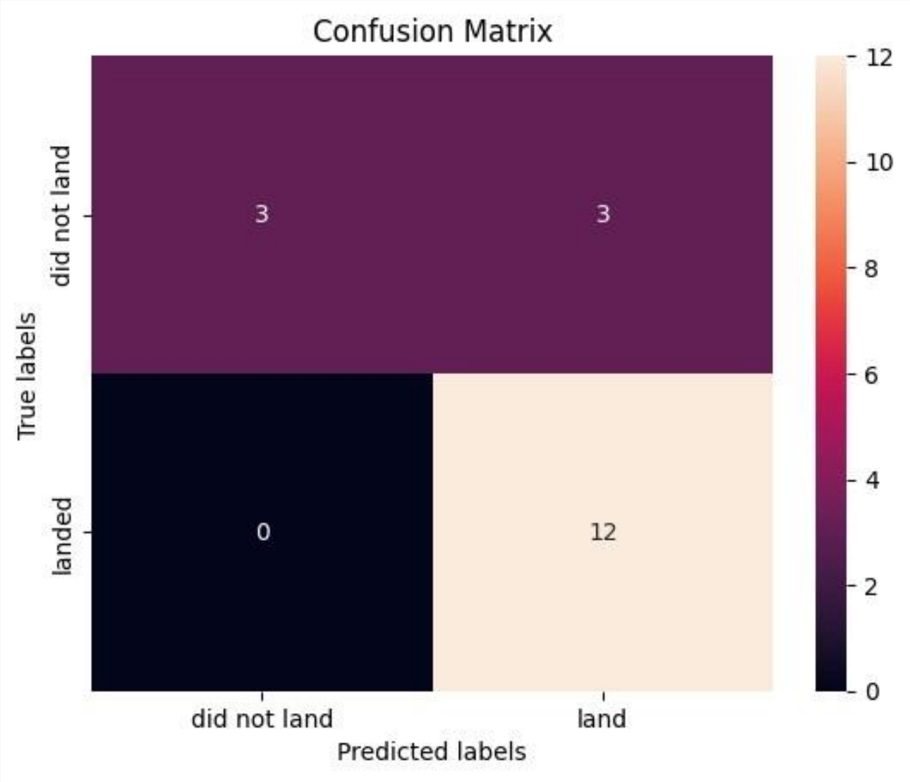
# Predictive Analysis (Classification)

# Classification Accuracy

- Visualize the built model accuracy for all built classification models, in a bar chart

- All models have similar accuracy.

# Confusion Matrix



# Conclusions

- All machine models have very similar accuracy results.
- The chances of success seem to be related to the number of flights. Likewise, in the GTO orbit, we observe no relationship between flight number and the chances of success.
- Heavier payloads have a more positive landing rate for orbits LEO, ISS, and Polar. In contrast, the chances of a positive landing for orbit GTO is not affected by payload masS.

# Thank you!