IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Muskan Deep Kaur
04/06/2024

Outline

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

Executive Summary

- In our report of SpaceX, we used different tools to find Rocket Name, Payload Mass, Name of Launch Sites being used for better evolution and its corresponding Cores to interpret landing of the Rockets, whether or not the rocket landing was successful or not, what have used during the testing of Rockets i.e. Launch Pads, Legs, Grid fins.

- After interpreting all the results the accuracy for landed rockets were 83% which includes the key factors such as Payload mass and Cores which are crucial for condition of a rocket.

Introduction

- This project is all about fetching SpaceX data to build our new Space company i.e., SpaceY. SpaceX is a famous space agency which build, craft and test space rockets. SpaceX cost only 62 million dollars to launch a rocket while other companies cost 165 million dollars. SpaceX's rocket are built in a way that the core could be taken in use more than one time.
- In the project we want to find out the best key features to help us build our rocket. This is going to be done by knowing how the rocket is build? What special features it have? How to make our rocket launch successfully? These all analysis has been interpreted using the data science methodologies.

Section 1

Methodology

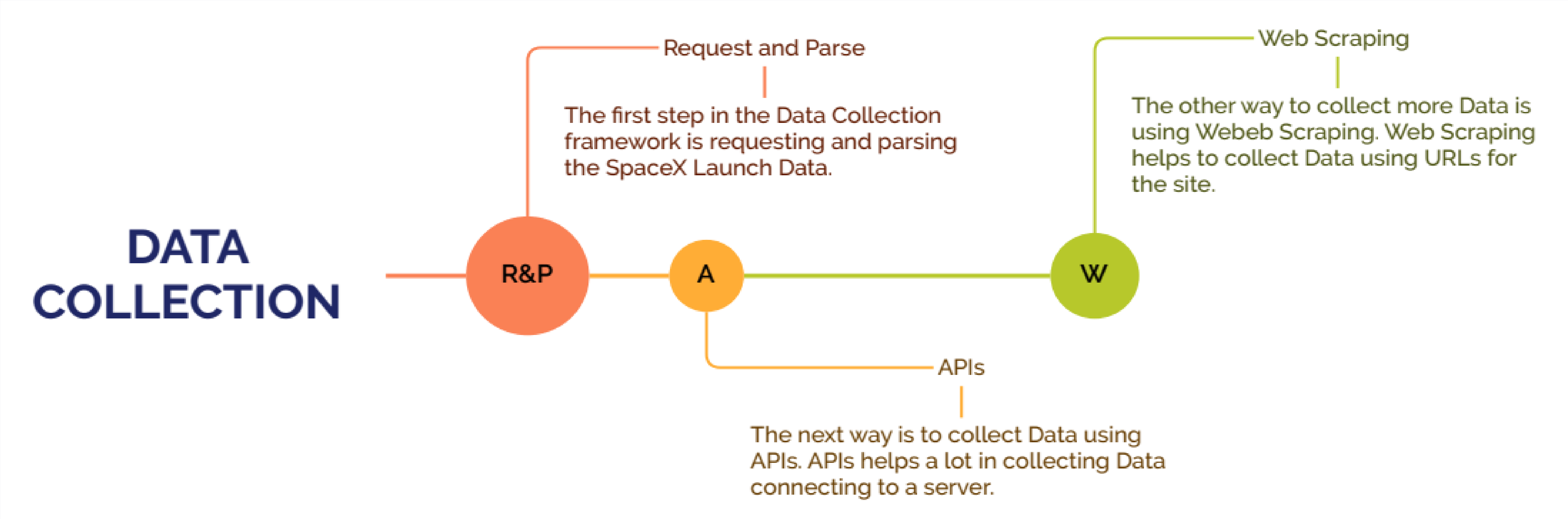
Methodology

Executive Summary

- Data collection methodology:
  - Data for SpaceX has been collected using URL of wiki and many other URLs from where we can collect our Data and requesting from http servers to collect data with building connection with the website.
- Perform data wrangling
  - Using different tools we collect and pre-process the Data to identify the pattern of the Data, to know our Data.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Then used classification models to build our model to predict results more accurately.

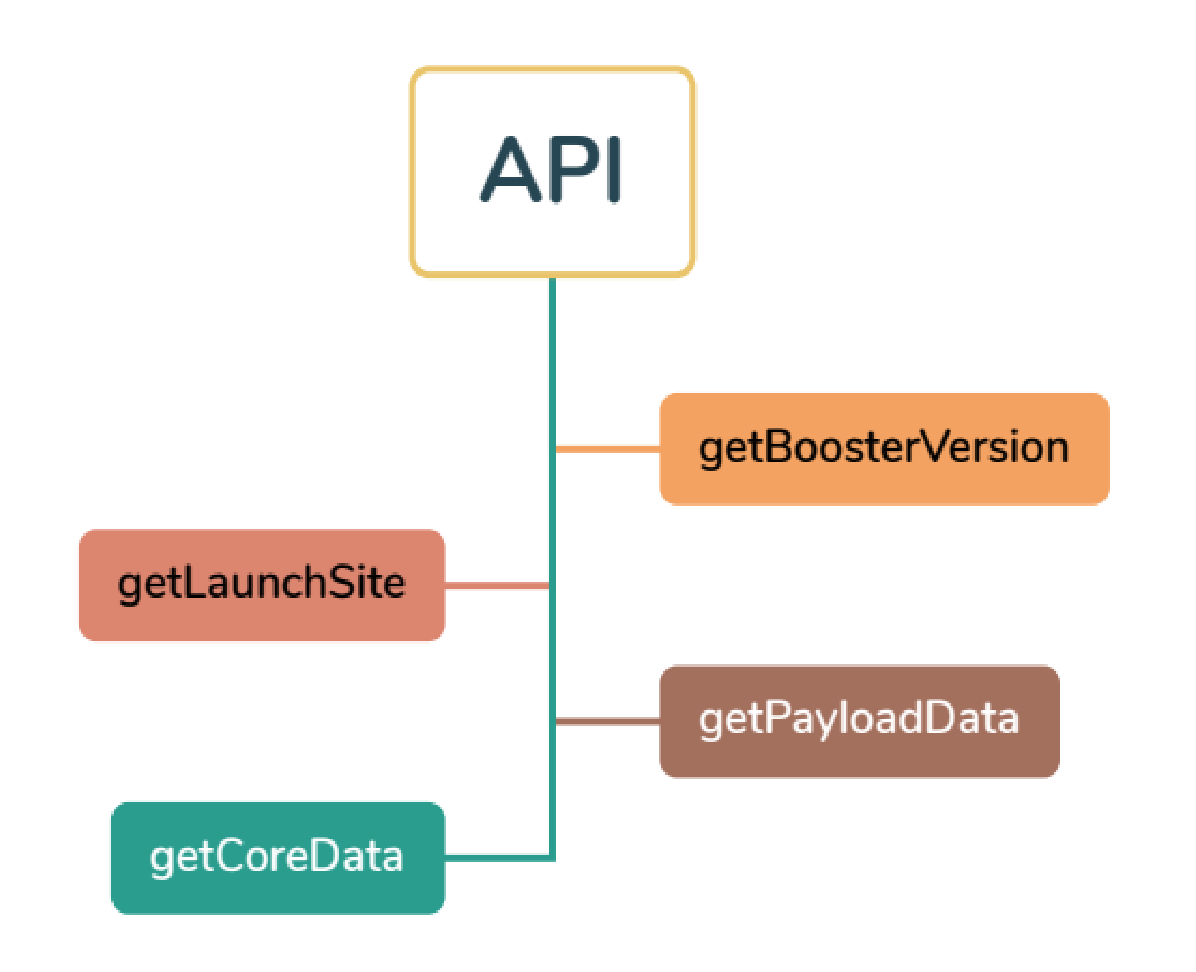
Data Collection

- Data for SpaceX has been collected from URLs pf the websites which provided the Data needed for our analysis i.e. web scraping from wikipedia and also using APIs.



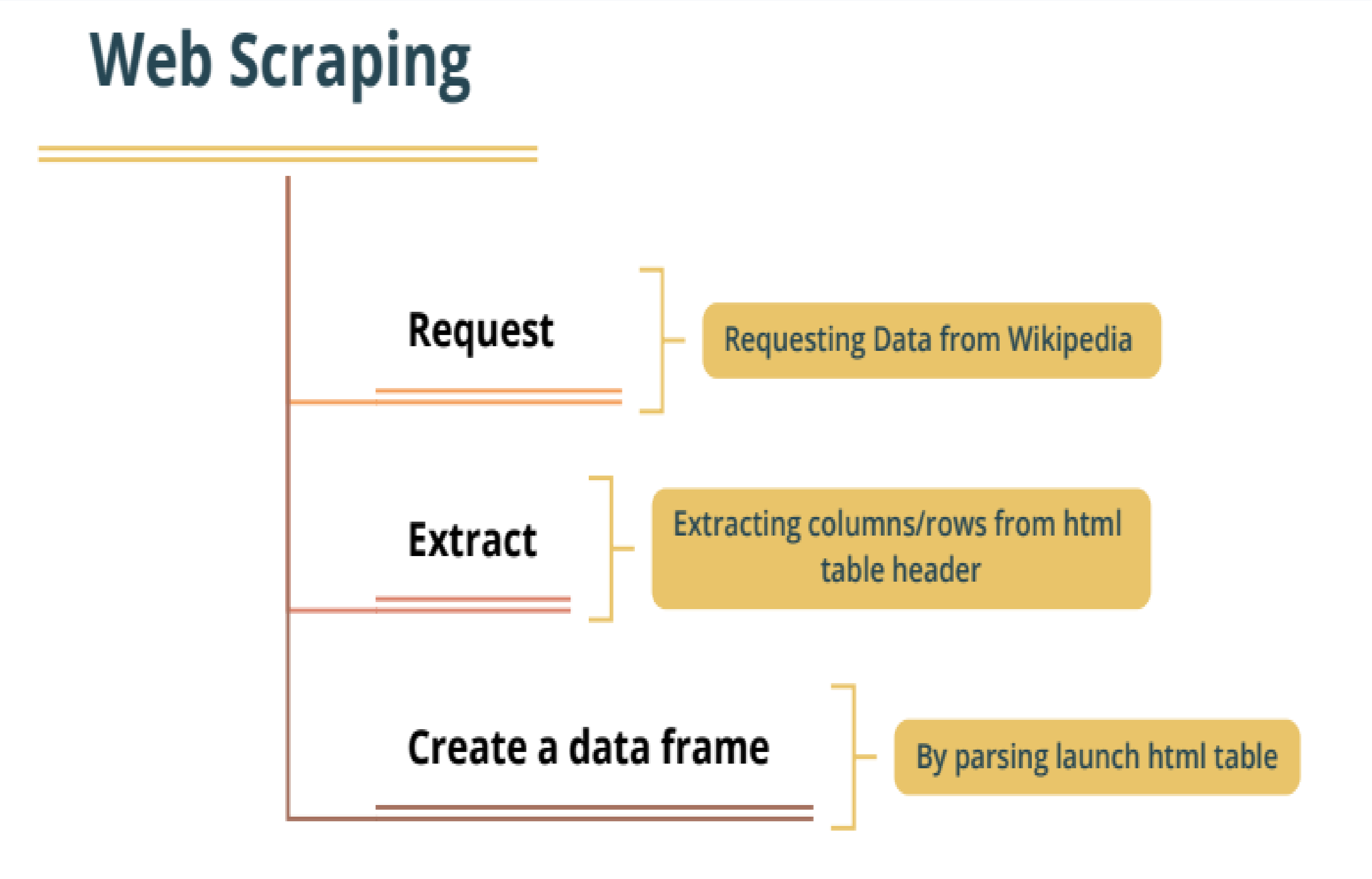
Data Collection – SpaceX API

- We have used APIs such as "getBoosterVersion", "getLanuchSite", "getPayloadData", "getCoreData".
- GitHub URL of the completed SpaceX API calls notebook is: https://github.com/muskandeepkaur/SpaceX_Capstone.git



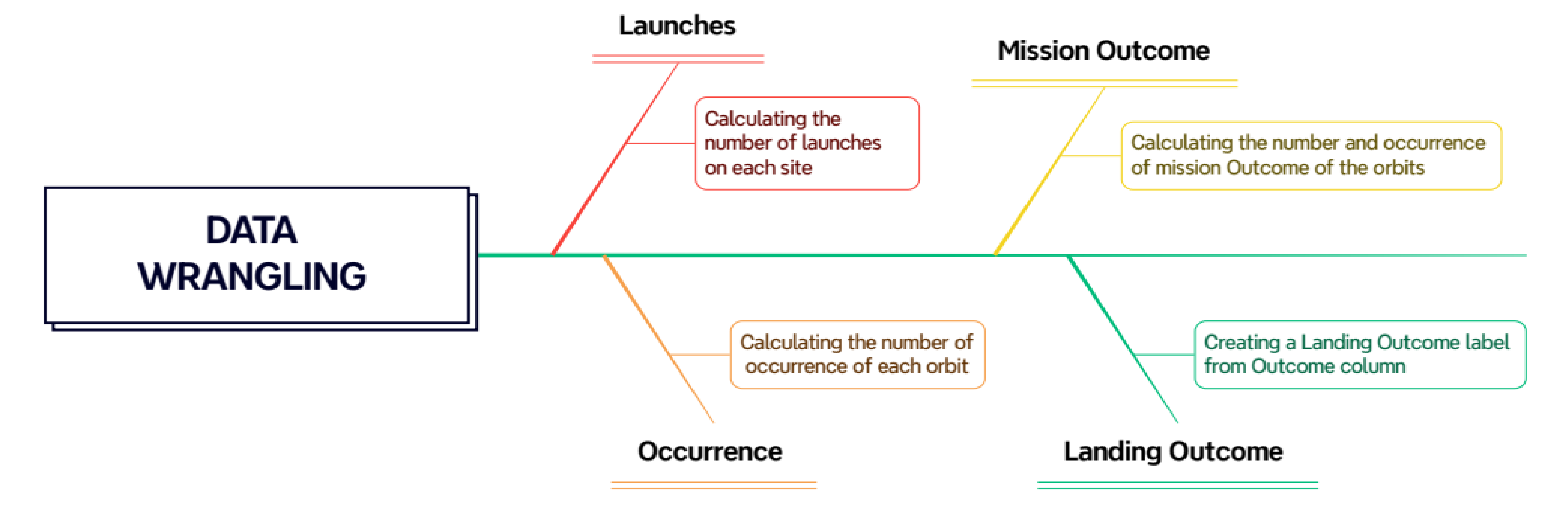
Data Collection - Scraping

- We used web scraping functions to make a request from wikipedia site to collect Data by .get(url) method.
- GitHub URL of the completed web scraping notebook is: https://github.com/muskandeepkaur/SpaceX_Capstone.git



Data Wrangling

- Data Wrangling here is calculating the number of launches on each site, number of occurrence of each orbit type, mission outcome of the orbits, and evaluating the successfully landed rockets.
- GitHub URL of the completed data wrangling notebook is: https://github.com/muskandeepkaur/SpaceX_Capstone.git



EDA with Data Visualization

- In the process of visualizing data many plots are formed to represent relationships between different features and to help better understand the Data and insights from the Data. Plots and charts are the best way to understand Data easily and to think from it more clearly.

- GitHub URL of EDA with data visualization notebook is:

  https://github.com/muskandeepkaur/SpaceX_Capstone.git

EDA with SQL

SQL queries used for EDA are listed below:

- DISTINCT
- LIKE
- LIMIT
- COUNT
- SUM
- MIN
- THEN

Github URL for the EDA by SQL notebook is:

https://github.com/muskandeepkaur/SpaceX_Capstone.git

Build an Interactive Map with Folium

- Folium is used to make our visualizations more interactive, we used markers, clusters, circles to the map to point out the launch sites and where the rocket failed and where it was successfully launched and landed.
- Folium helps to better understand our Data making it more interactive and real, for our audience to hover and click on any site to see details from the map.
- GitHub URL of interactive map with Folium map notebook is:

  https://github.com/muskandeepkaur/SpaceX_Capstone.git

Build a Dashboard with Plotly Dash

- The first step is to add a launch site drop-down input component, then adding Callback function on selected site, adding a Range Slider to selected Payload, and to render the Scatter plot.
- Dashboards make graphs and charts appear more beautiful and precise, it added virtualization to the charts or graphs to display information for a particular part we want to extract.
- GitHub URL of Plotly Dash lab is:

https://github.com/muskandeepkaur/SpaceX_Capstone.git

Predictive Analysis (Classification)

- Prediction from the model has done by Creating arrays, normalizing the Data, splitting Data into train and test, creating logistic regression models, calculating accuracy of the Test and Train Data, creating support machine, Decision tress, and using KNN. All methods helps step by step to predict the model based on Data and evaluate the results.

- GitHub URL of your completed predictive analysis lab is:

  https://github.com/muskandeepkaur/SpaceX_Capstone.git

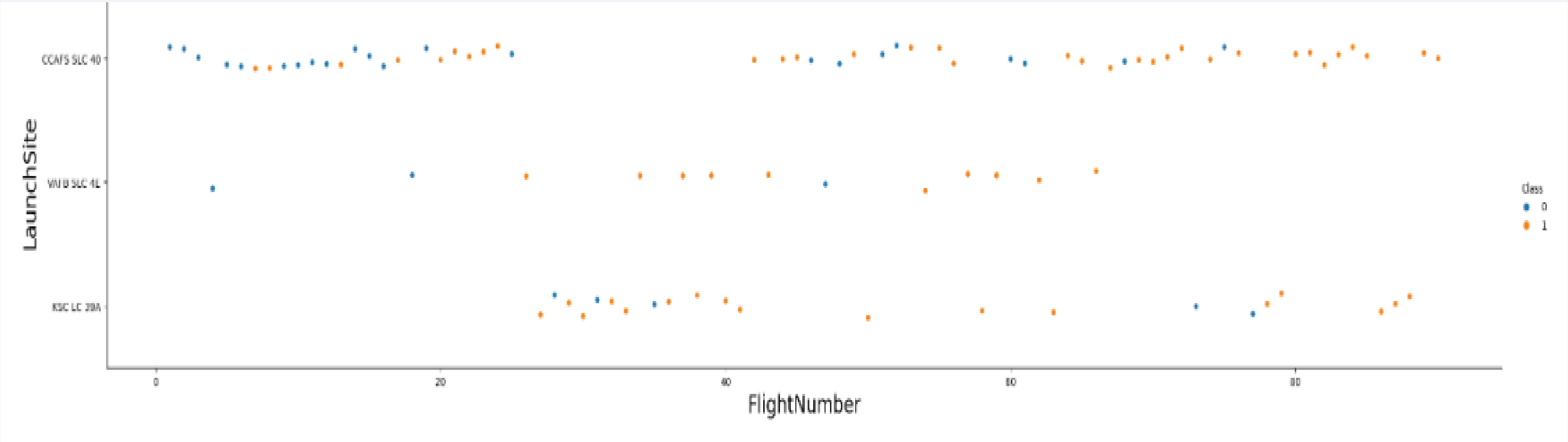
Results

- The core of the rocket is used twice and tested twice with using launching pads for land area and without using launching pads for ocean ship landing.
- The data shows the relationship of the rockets with the core including others features shows that it is viable equally by the production.
- The Training model shows the accuracy of the landing of the Rocket is 83% which is higher than normal rate. While the Rocket could hold upto 3 times reusable and the Buit model is not so expensive.

Section 2

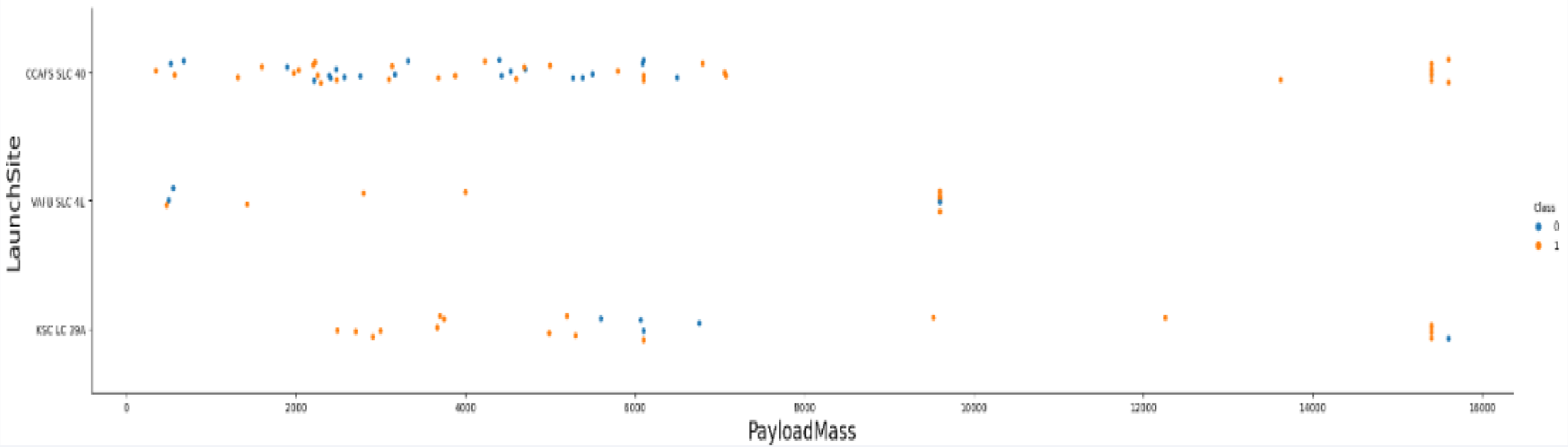
Insights drawn from EDA

Flight Number vs. Launch Site



The Graph shows which sites are most frequently used, any trends of preferences in site selection, operational efficiency of sites, utilization of infrastructure, global search of space exploration efforts, and emerging trends in space launch operations.

Payload vs. Launch Site



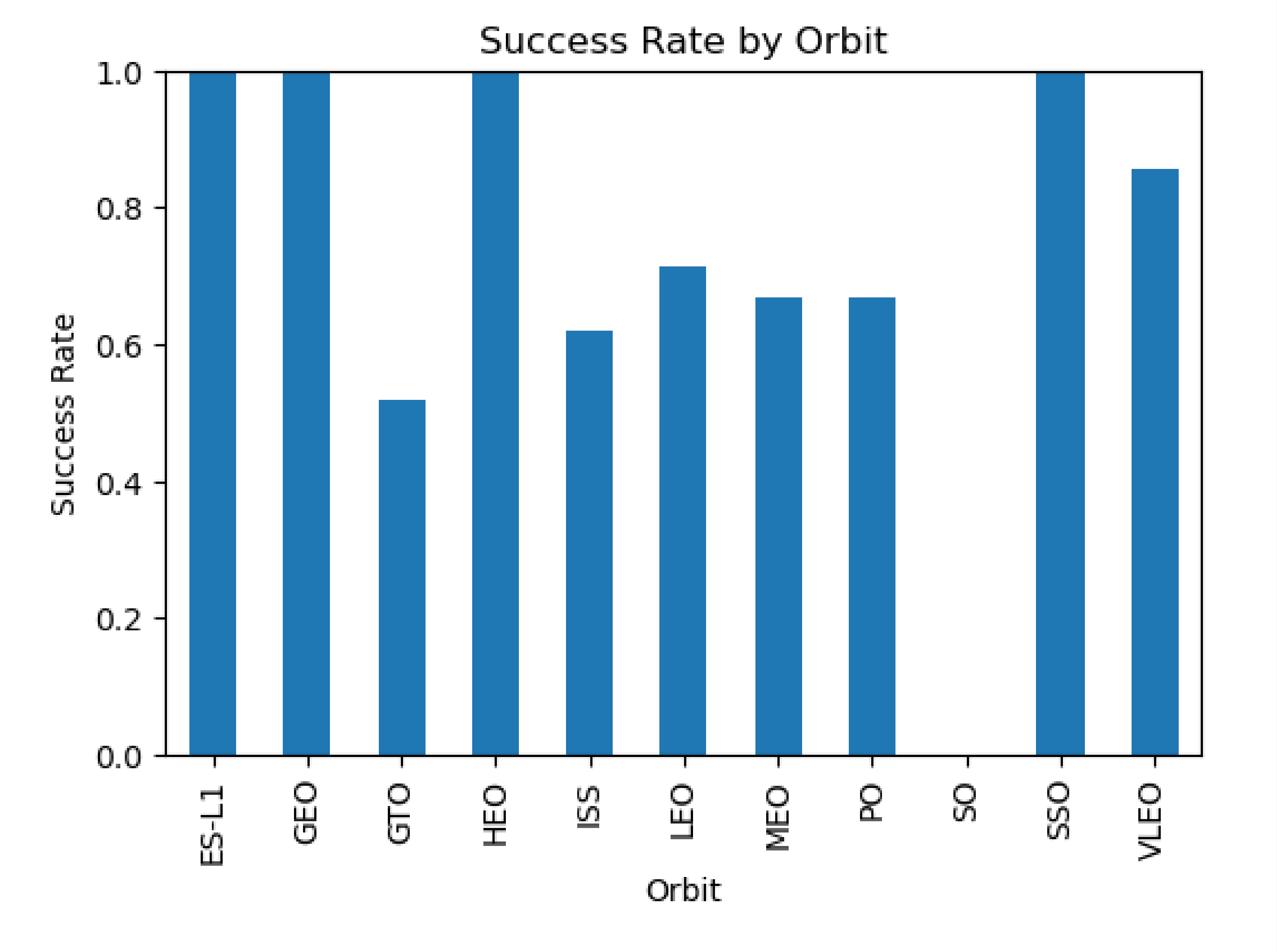
The graph shows:

- Specific Payload types per site and Payload sizes different sites handle
- Changes over time and main uses of each site
- Global Payload handling variation

Success Rate vs. Orbit Type

The graph shows:

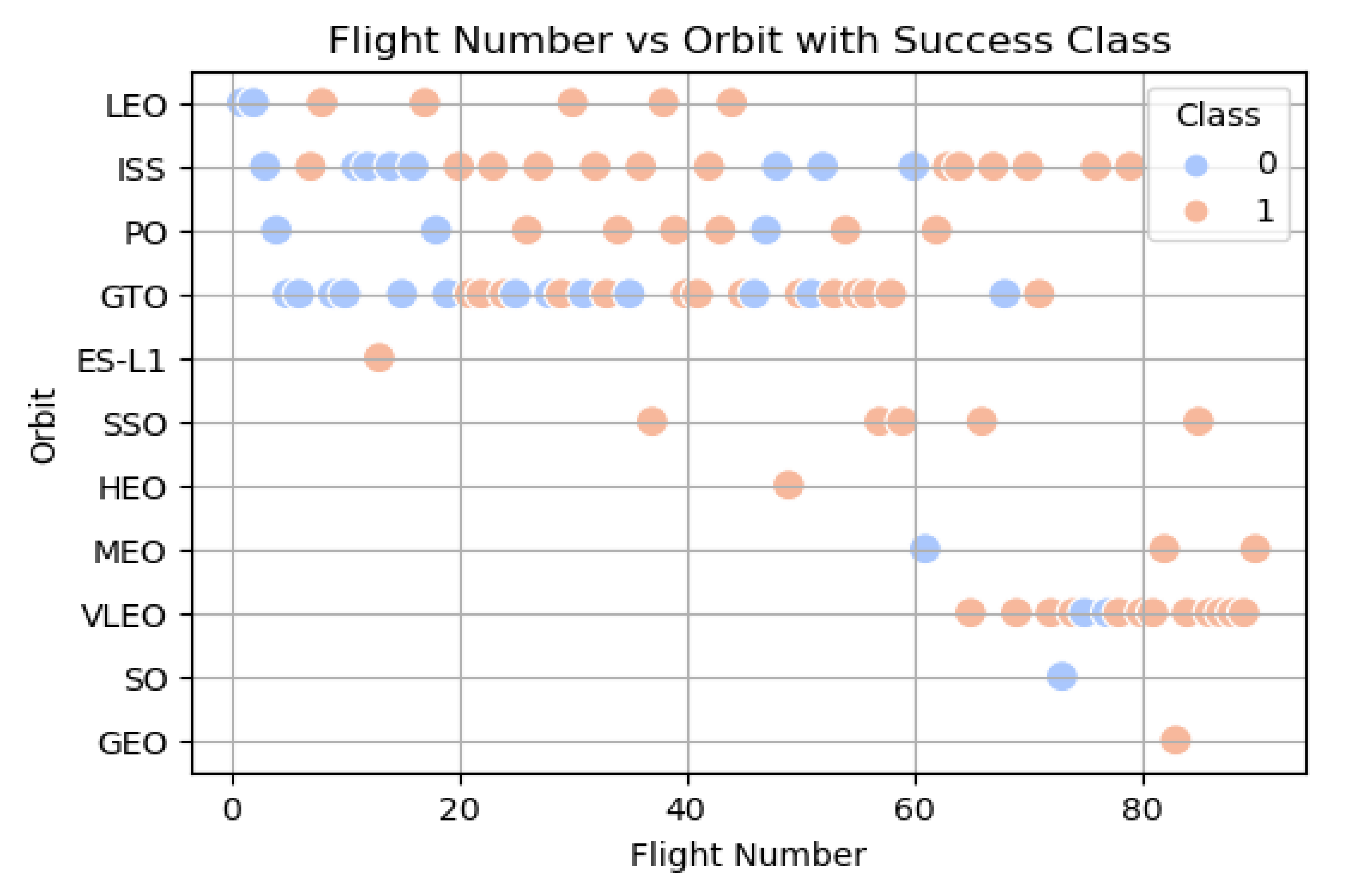
- Which orbit type have higher success rates.
- Orbit type with lower success rates indicating more difficulty.
- Effectiveness of launch systems for different orbits.
- Success rates changes over time for specific orbits.
- Areas needing improvement for certain orbit types.



Flight Number vs. Orbit Type

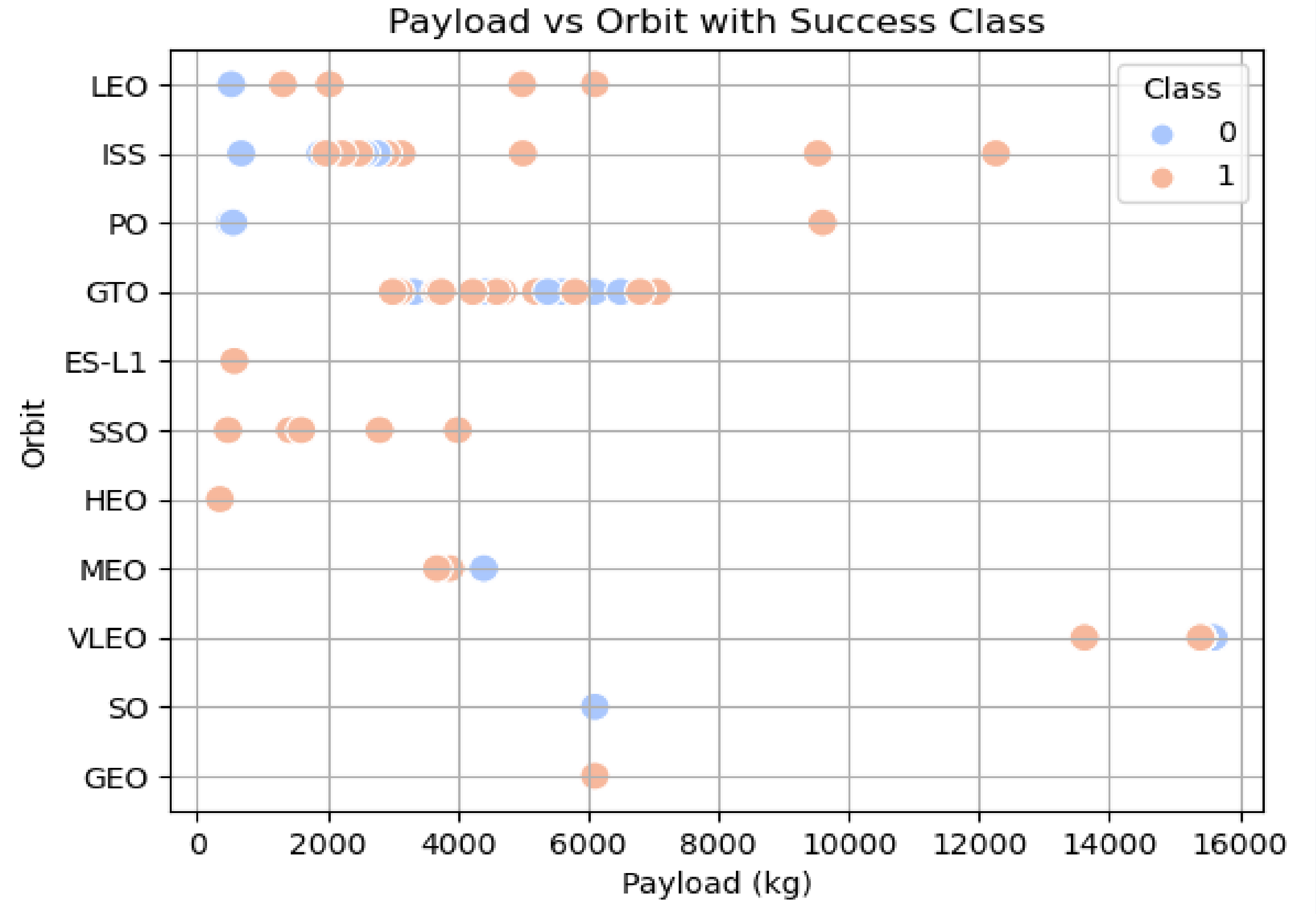
The graph shows:

- How often different orbit types are targeted.
- Popularity of certain orbits over time.
- Which orbits are prioritized by launch providers.
- Variety of missions targeting each orbit type.
- Insights into strategic goals.



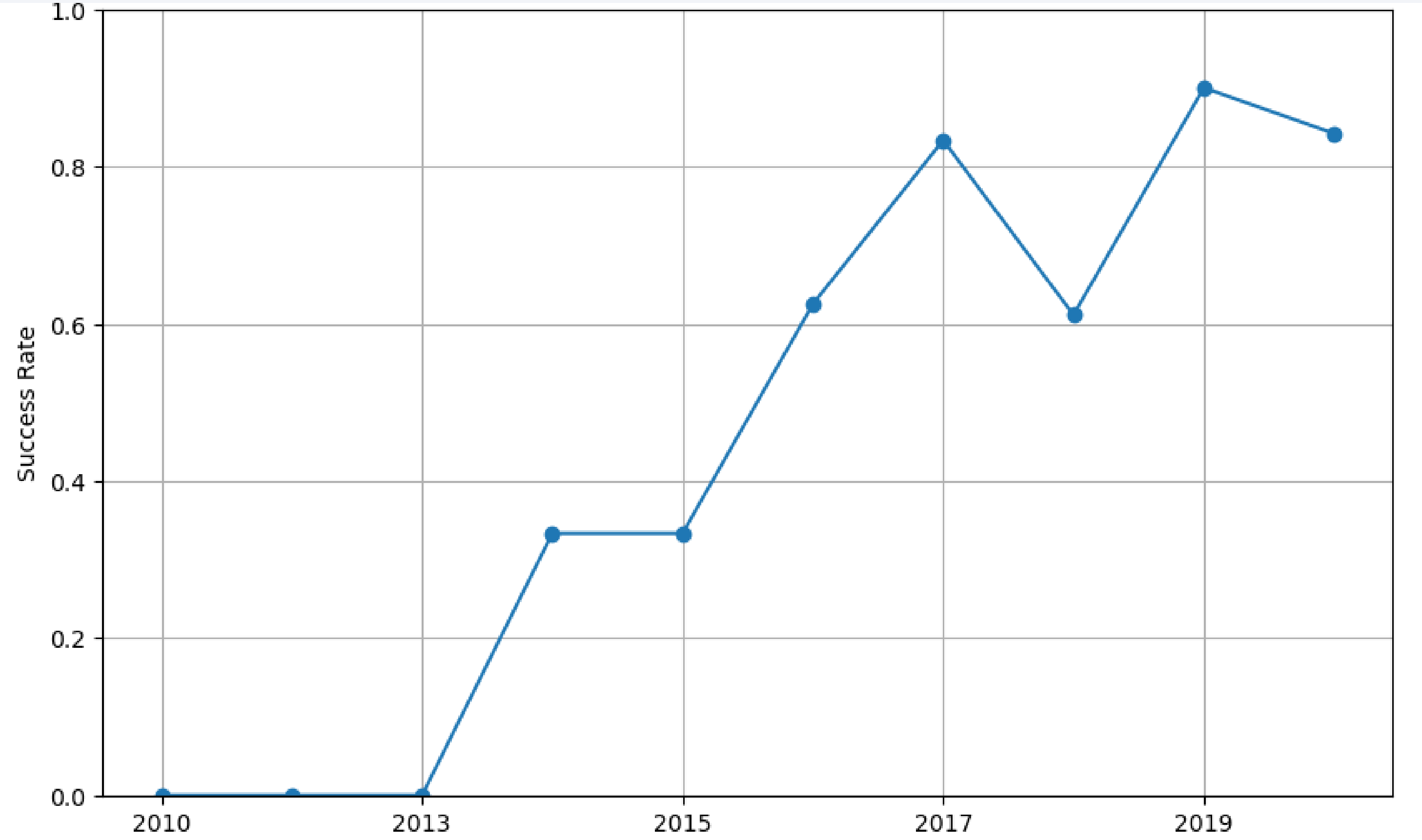
Payload vs. Orbit Type

- The graph shows which
  payload are sent
  to specific orbits, revealing trends in mission objectives. It highlights preferred orbits for
  heavy or specialized payloads and patterns in payload distribution, helping
  to understand strategic priorities and utilization of different orbital environments.



Launch Success Yearly Trend

The graph shows the change in number of the successful launches over time, highlighting improvements in technology and reliability, identifying years with notable success or failures, and revealing overall growth or decline in successful space missions.



All Launch Site Names

Name of the sites could be find out by running the following SQL query:

- SELECT DISTINCT Launch_Site FROM SPACEXTABLE;

The launch site used are:

- CCAFS SLC-40(Cape Canaveral Space Force Station Space Launch Complex 40) - Florida, USA
- KSC LC-39A (Kennedy Space Center Launch Complex 39A) - Florida, USA
- VAFB SLC-4E(Vandenberg Space Force Base Space Launch Complex 4E) - California, USA
- STLS(South Texas Launch Site, also known as Boca Chica) - Texas, USA

Launch Site Names Begin with 'CCA'

The SQL query to find 5 records where launch site begin with 'CCA' is:

```
SELECT * FROM SPACEXTABLE WHERE Launch_Site LIKE 'CCA%' LIMIT 5;
```

The query shows the result with launch_site beginning with 'CCA' with a large table showing features for the result.

Total Payload Mass

The SQL query to find the total Payload Mass is:

SELECT SUM(PAYLOAD_MASS__KG_) AS TotalPayloadMass FROM SPACEXTABLE WHERE Customer = 'NASA (CRS)';

The query helps to find out the Total Payload Mass carried by Boosters launches by NASA which is 45596.

Average Payload Mass by F9 v1.1

The SQL query for average payload mass by F9 v1.1 is:

```
SELECT AVG(PAYLOAD_MASS__KG_) AS AveragePayloadMass FROM
SPACEXTABLE WHERE Booster_Version = 'F9 v1.1';
```

This query helps to find out the average Payload Mass by F9 v1.1 which is 2928.4

First Successful Ground Landing Date

The SQL query to find out the date when the first successful landing happened:

SELECT MIN(Date) AS FirstSuccessfulLandingDate FROM SPACEXTABLE WHERE Landing_Outcome = 'Success (ground pad)';

The query shows the result dated 2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

The SQL query to find the successful Drone Ship Landing with Payload between 4000 and 6000 is shown by:

```
SELECT Booster_Version FROM SPACEXTABLE WHERE Landing_Outcome = 'Success (drone ship)' AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000;
```

The query shows the following result for Booster Version:

- F9 FT B1022
- F9 FT B1026
- F9 Ft B1021.2
- F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

SQL query to find out the result is:

```
SELECT Landing_Outcome, COUNT(*) AS TotalCount FROM SPACEXTABLE GROUP BY Landing_Outcome;
```

The result is shown here,

Landing_Outcome	TotalCount
Controlled (ocean)	5
Failure	3
Failure (drone ship)	5
Failure (parachute)	2
No attempt	21
No attempt	1
Precluded (drone ship)	1
Success	38
Success (drone ship)	14
Success (ground pad)	9
Uncontrolled (ocean)	2

Boosters Carried Maximum Payload

We used SQL query to find out
the Boosters carried Maximum Payload.
The result is shown here,

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

2015 Launch Records

The result for the query is shown below for 2015 Records:

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

The result represents the details for 2015 Launch Records.

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 is shown here.

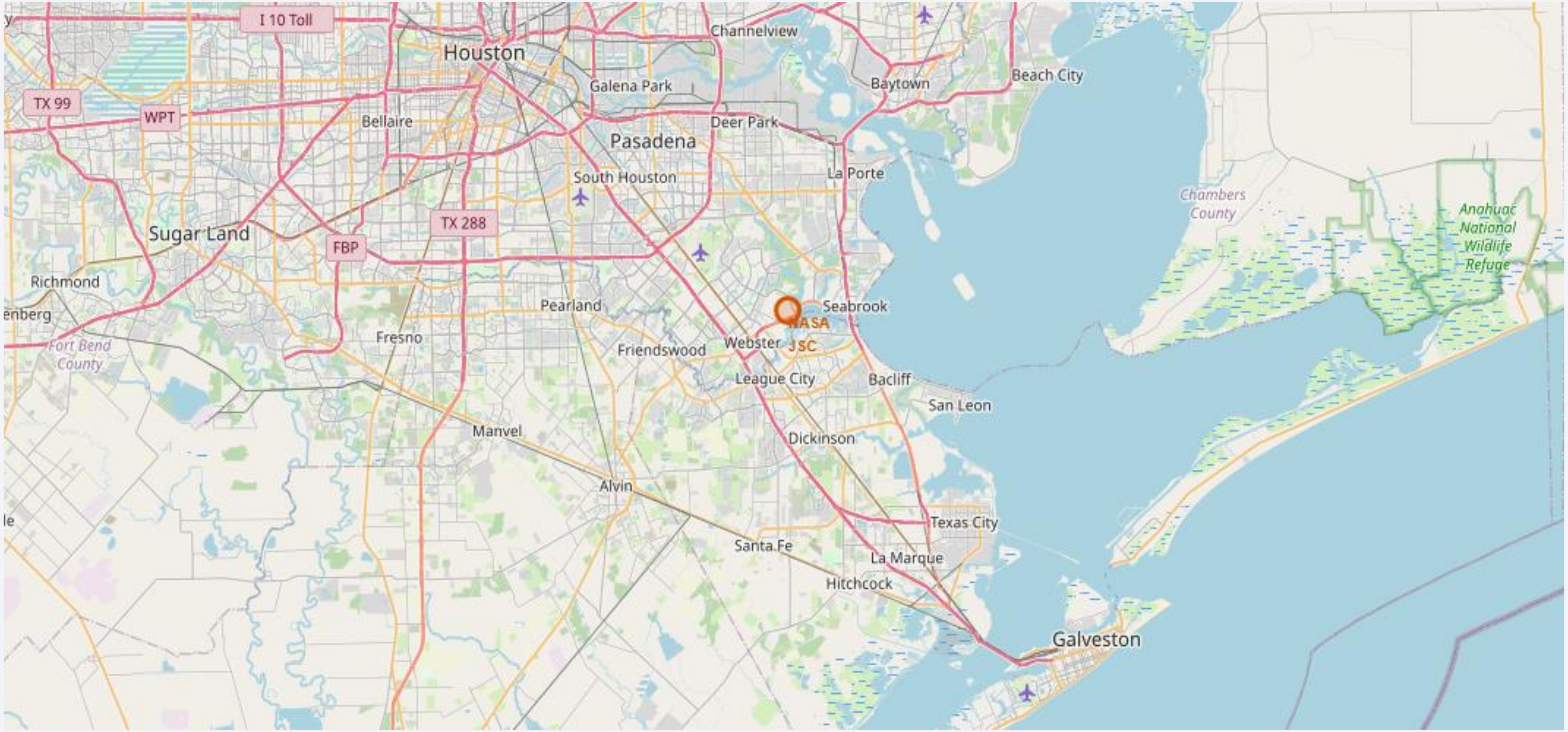
The result clearly shows the success rate for landing is more than failure rate which is a good point.

Landing_Outcome	Outcome_Count
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

Section 3

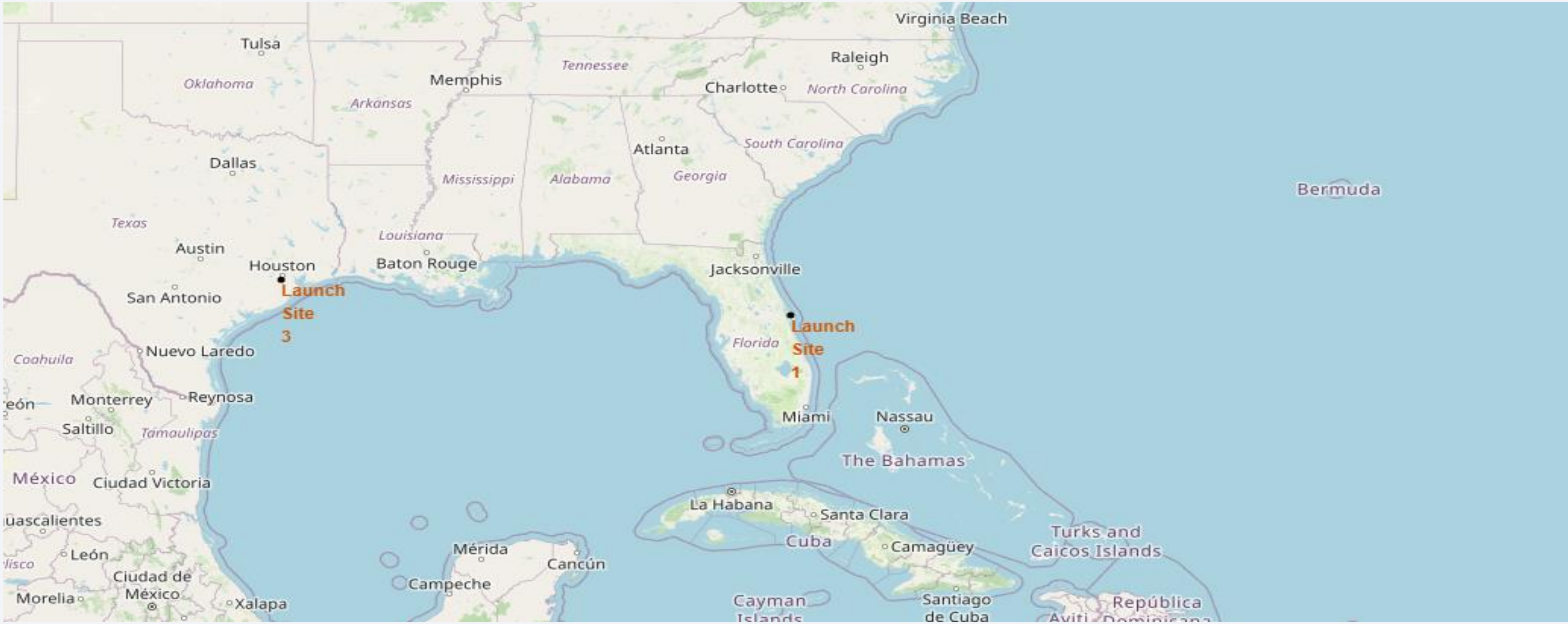
Launch Sites Proximities Analysis

Launch site locations



The map shows marked launch sites on it.

Launch site locations with pop-up details



The map shows marked launch sites with pop up such as launch site1 & 3 here.

Roadway map to determine distance

The map shows roadways and all roots to find out distances between a launch site and its proximities.

Section 4

Build a Dashboard with Plotly Dash

Creating Dashboard

Command to use in creating dashboards are:

```
import dash
import dash_html_components as html
import dash_core_components as dcc
from dash.dependencies import Input, Output
```

With these commands dash boards can be created which make higher-level component slides, graphs, and plots.

Section 5

Predictive Analysis (Classification)

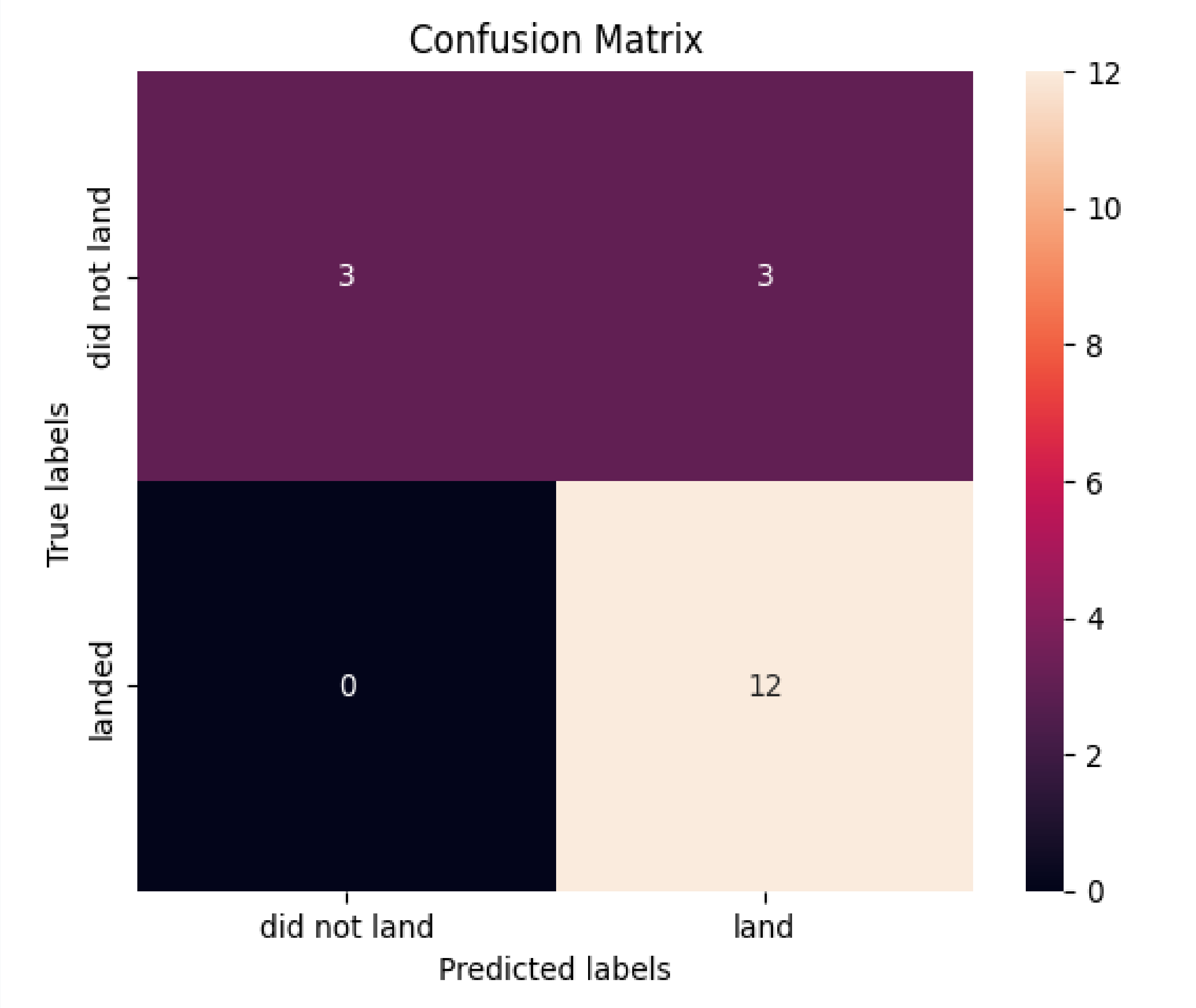
Classification Accuracy

We use logistic regression to find out the accuracy of the Validation Data which in result found to be 84%, then we performed accuracy test on our test data and we found the result to be 84%, which is nearly to the model prediction analytic insights.

Confusion matrixes helps a lot in identifying true prediction value. Using different methods of regression, classification, and decision tress average accuracy to be found is 84% which fits our model well.

Confusion Matrix

This shows the confusion matrix with the best outcome.



Conclusions

- SpaceX built their Rockets on designing and focusing more on Cores and their evaluation.
- Detecting Launch Sites and their working framework, we could deliver our Rocket to parse distance problem for best launch.
- SpaceX have their accuracy of rockets to success around 84% which is good to consider.

Appendix

- Some other projects created during the course are below with GItHub URLs:

  Data Science Ecosystem (https://github.com/muskandeepkaur/Data-Science-Ecosystem.git)

  Extracting Stock Data (https://github.com/muskandeepkaur/Extracting-Stock-Data.git)

  SpaceX Capstone (https://github.com/muskandeepkaur/SpaceX_Capstone.git)

Thank you!