IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Name : Abdul Jaweed

Date : 20 Dec 2022

Outline

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

Executive Summary

In this capstone project, we will predict the SpaceX Falcon 9 first stage will land successfully. If we can determine if the first stage will land, we can determine the cost of a launch. This will be achieved with the use of different machine learning classification algorithms.

The methodology followed will include data collection, data wrangling and pre-processing, (EDA) exploratory data analysis, data visualization and finally, machine learning prediction.

During our investigation, the results of our analysis indicate that there are some features of rocket launches that have a correlation with the success or failure launches.

In the end we conclude that except decision tree all three model perform same accuracy and may be the best machine learning algorithm to for this problem.

Introduction

The main goal of this capstone project is to predict whether the Falcon 9 first stage will land successfully.

Because Space X advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because Space X can reuse the first stage.

Therefore if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

This brings us to our main question that we are trying to answer : For a given set of features about a Falcon 9 rocket launch, will the first stage of the rocket land successfully?

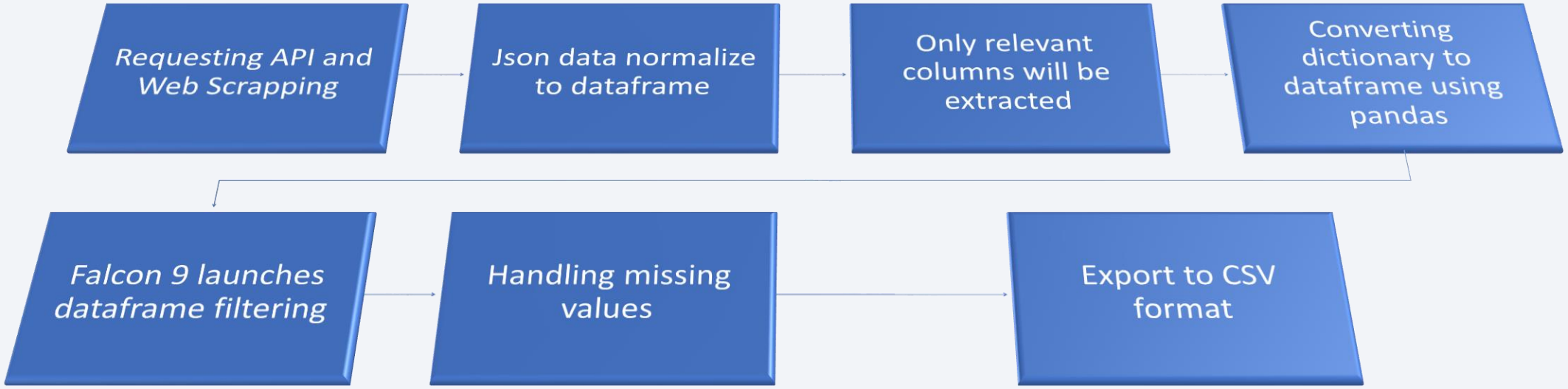
Section 1

Methodology

Methodology

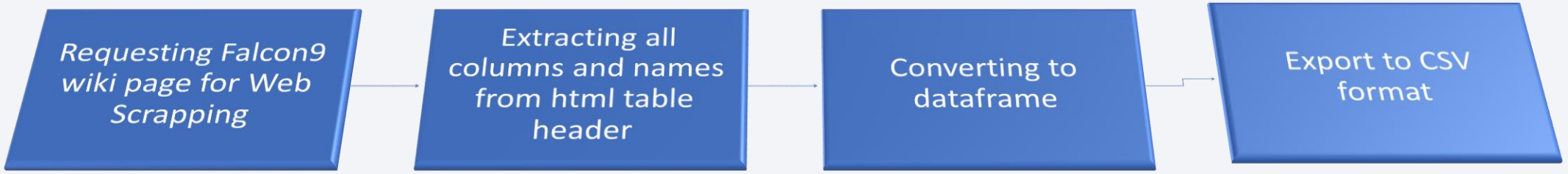
- The data was collected by two methods : 1. SpaceX API and 2. Web scraping.
- Data wrangling or data cleaning was performed to transform and clean data using pandas Python library.
- Using clean data to perform (EDA) exploratory data analysis using SQL queries and python visualization libraries like matplotlib and seaborn.
- Using Folium library to create a map and using Plotly Dash library to create a Dashboard.
- Using Render hosting service to deploy our dashboard.
- Using multiple machine learning classification models to perform a best predictive score.

Data Collection – SpaceX API



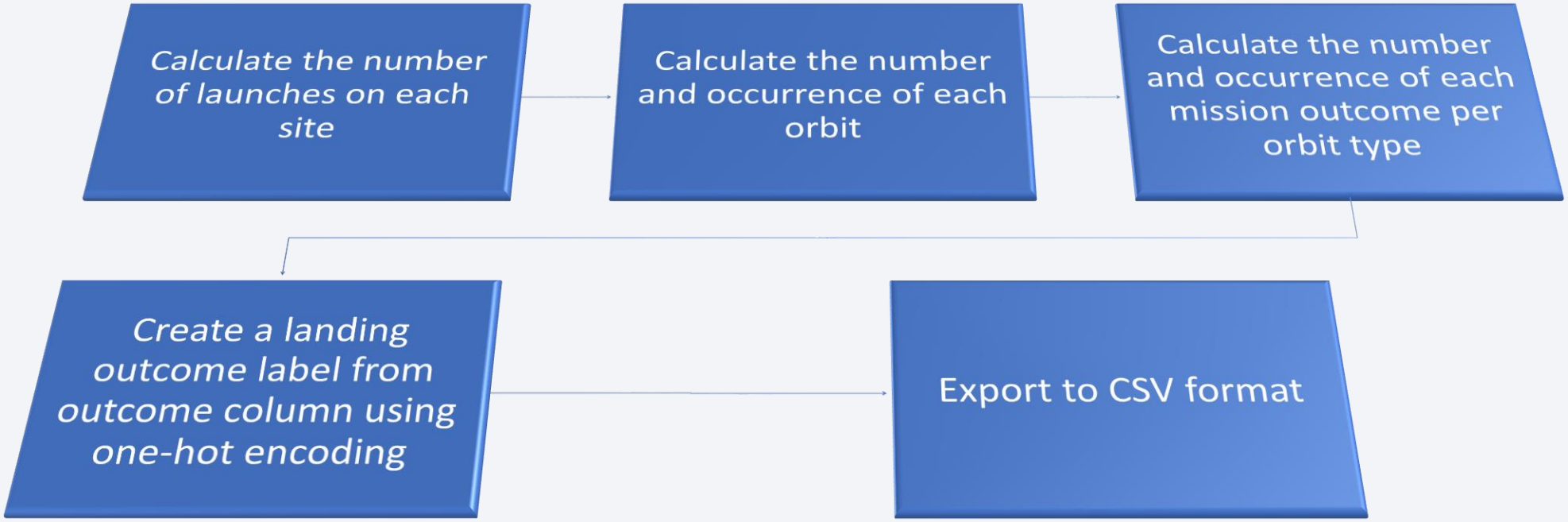
GitHub link : Data Collection API

Data Collection – Web Scraping



GitHub link : Data Collection Web Scraping

Data Wrangling



GitHub link : Data Wrangling

EDA with Data Visualization

- **GitHub link :** EDA with Data Visualization
- **Scatter plots :** Scatter plots were used to represent the relationship between the two variables. Different sets of features were compared such as *Flight* Number vs. Launch Site, Payload vs. Launch Site, Flight Number vs. Orbit Type and Payload vs. Orbit Type.
- **Bar chart :** Bar charts were used to makes it easily to compare values between multiple groups at a glance. The x-axis represents a category and the y-axis represents a discrete value. Bar charts were used to compare the Success Rate for different Orbit Types.
- **Line chart :** Line charts are used to showing the data trends over time. A line chart was used to show Success Rate over a certain number of years.

EDA with SQL

- **GitHub link : EDA with SQL**

- Displaying the unique launch sites in the space mission
- Displaying first 5 records where launch sites begin with the string ‘CCA’
- Displaying the total payload mass carried by boosters launched by NASA (CRS)
- Displaying average payload mass carried by boasted version F9 v1.1
- Listing the date when the first successful landing outcome in ground pad was achieved

EDA with SQL

- Listing the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000.
- Listing the total number of successful and failure mission outcomes
- Listing the names of the booster versions which have carried maximum payload mass.
- Listing the failed landing outcomes in drone ship, their booster versions and launch site names for in year 2015
- Ranking 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

Build an Interactive Map with Folium

- **GitHub link :** [Building an Interactive Map with Folium]

Objects was created and added to a Folium map. Marker objects were used to show all launch sites on a map as well as the successful/failed launches for each site on the map. Line objects were used to calculate the distances between a launch site to its proximities

- Adding these objects by following geographical patterns about the launch sites are found:
  - Do launch sites keep certain distance away from cities? Yes
  - Are launch sites in close proximity to highways? Yes
  - Are launch sites in close proximity to coastline? Yes
  - Are launch sites in close proximity to railways? Yes

Build a Dashboard with Plotly Dash

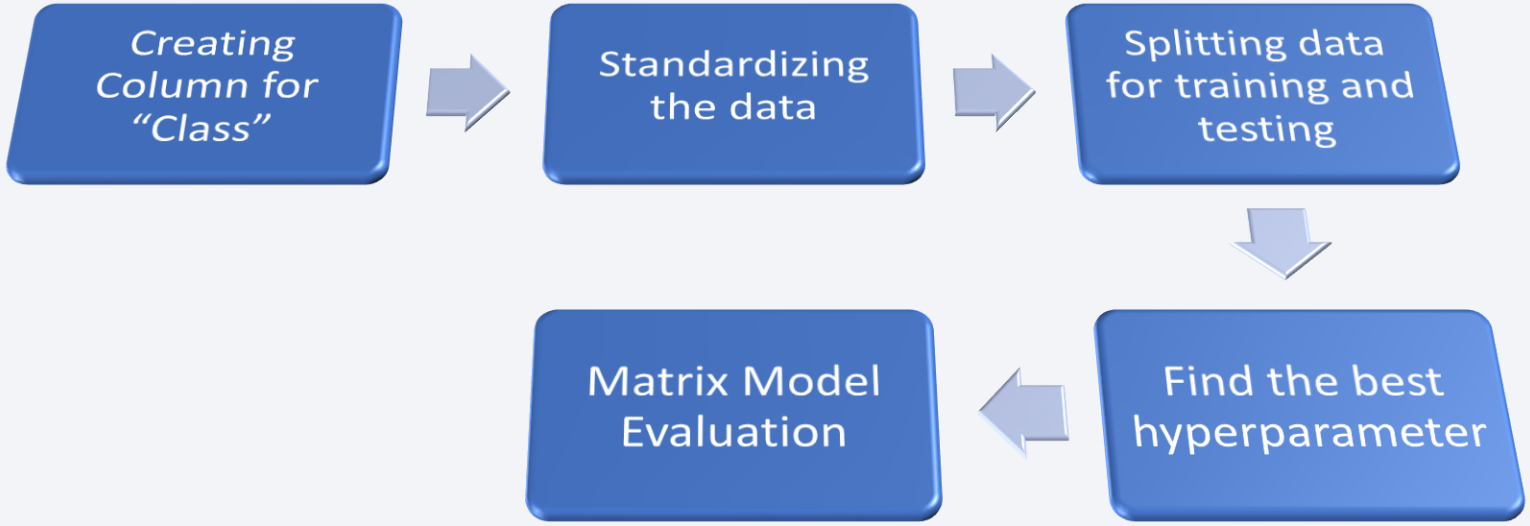
- **GitHub link :** SpaceX Launch Records Dashboard
- **Dashboard link :** SpaceX Launch Records Dashboard

The Dashboard Application contain two charts:

1. Pie-Chart : shows the successful launch by each site.
2. Scatter Plot : shows the relationship between landing outcomes on the payload mass of different boosters.

Bonus-one : I deploy this dashboard on render web hosting service

Predictive Analysis (Classification)

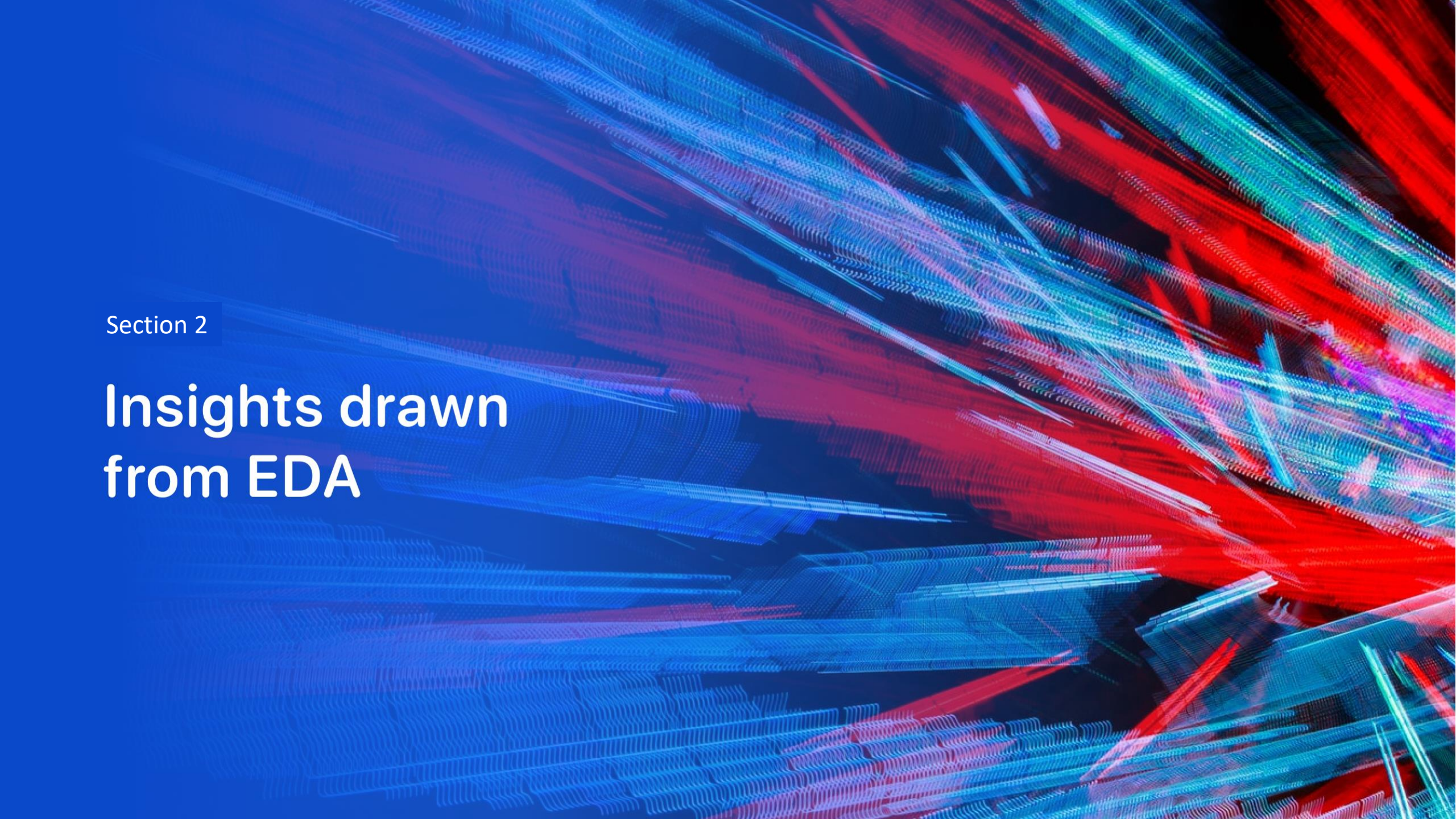


GitHub link : [Space-X Machine Learning Prediction]

Results

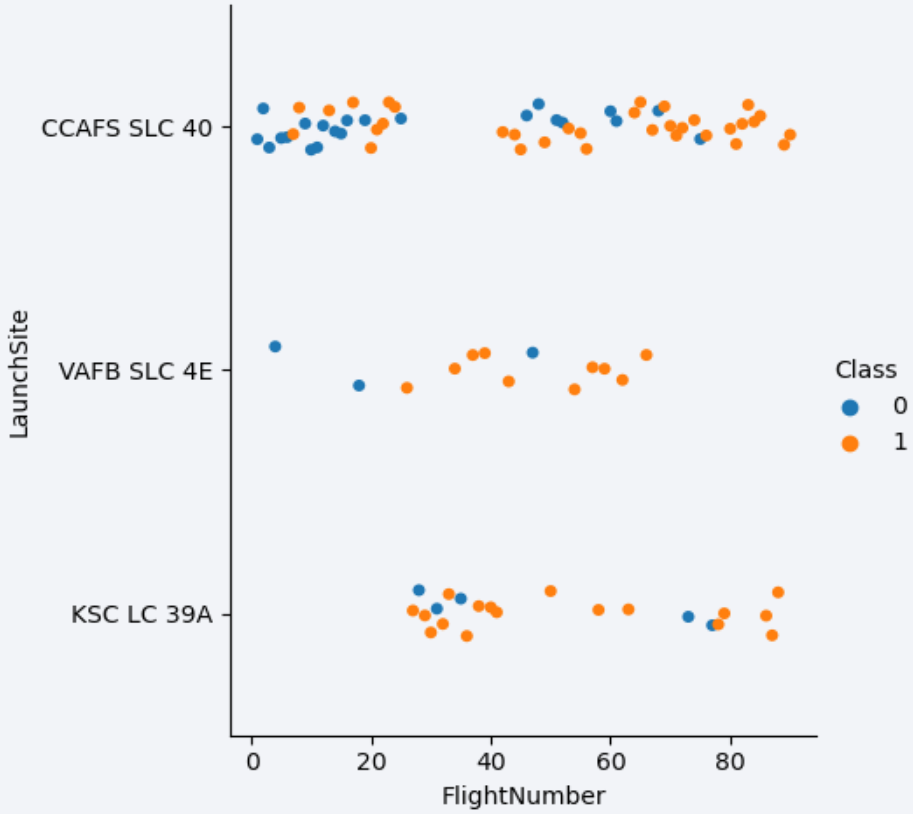
- The results of the (EDA) exploratory data analysis revealed the success rate of the SpaceX Falcon 9 rocket landings is 66%.

- The predictive analysis results showed that the Decision Tree Algorithm is less accuracy and Logistic Regression, Support Vector Machine and K Nearest Neighbors was same accuracy and the classification method accuracy is 83%.


Section 2
Insights drawn from EDA

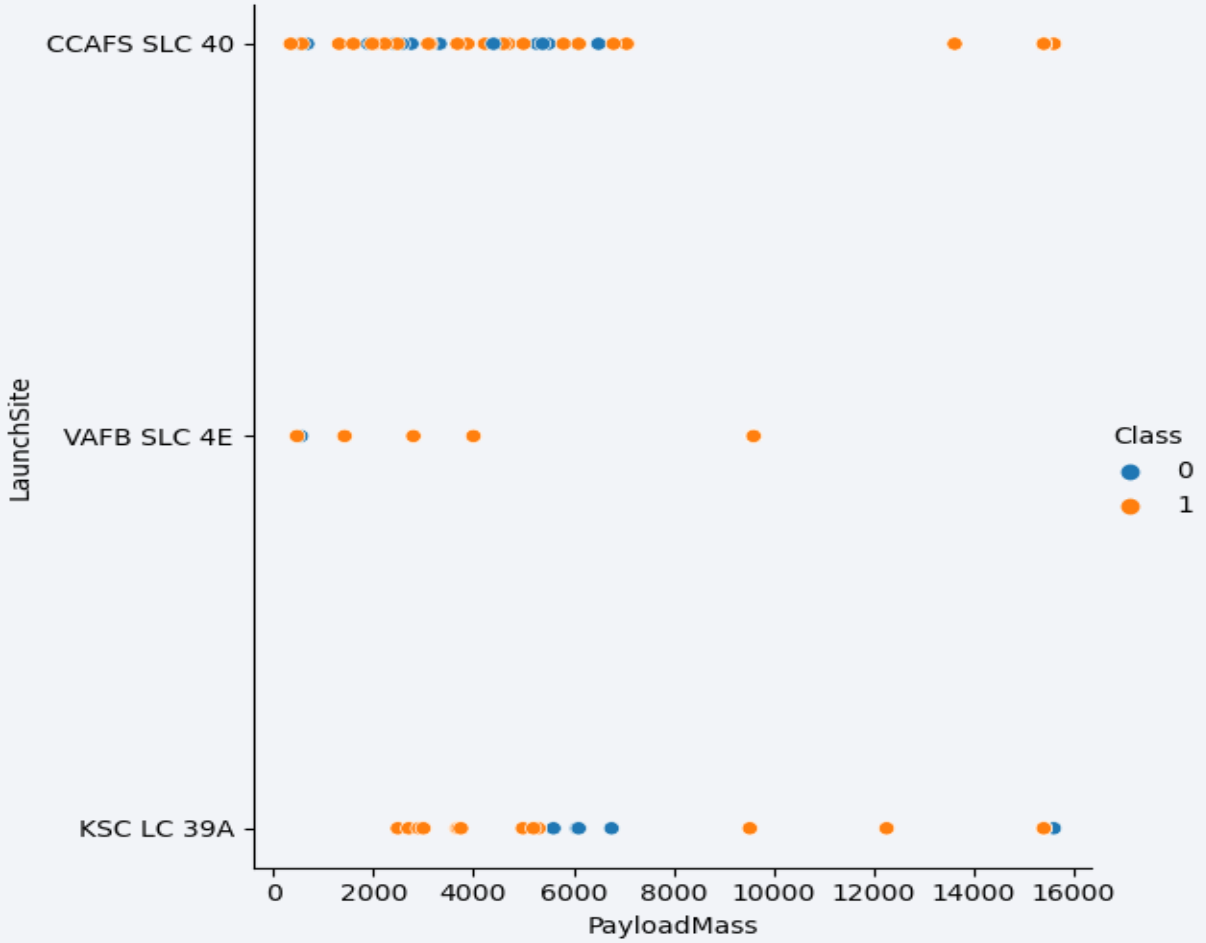
Flight Number vs. Launch Site

- This figure shows that the success rate has increased as the number of flight increased.
- The blue dots represent the unsuccessful launches while the orange dots represent successful launches.
- This figure showing that was more successful after the 40th Launch.



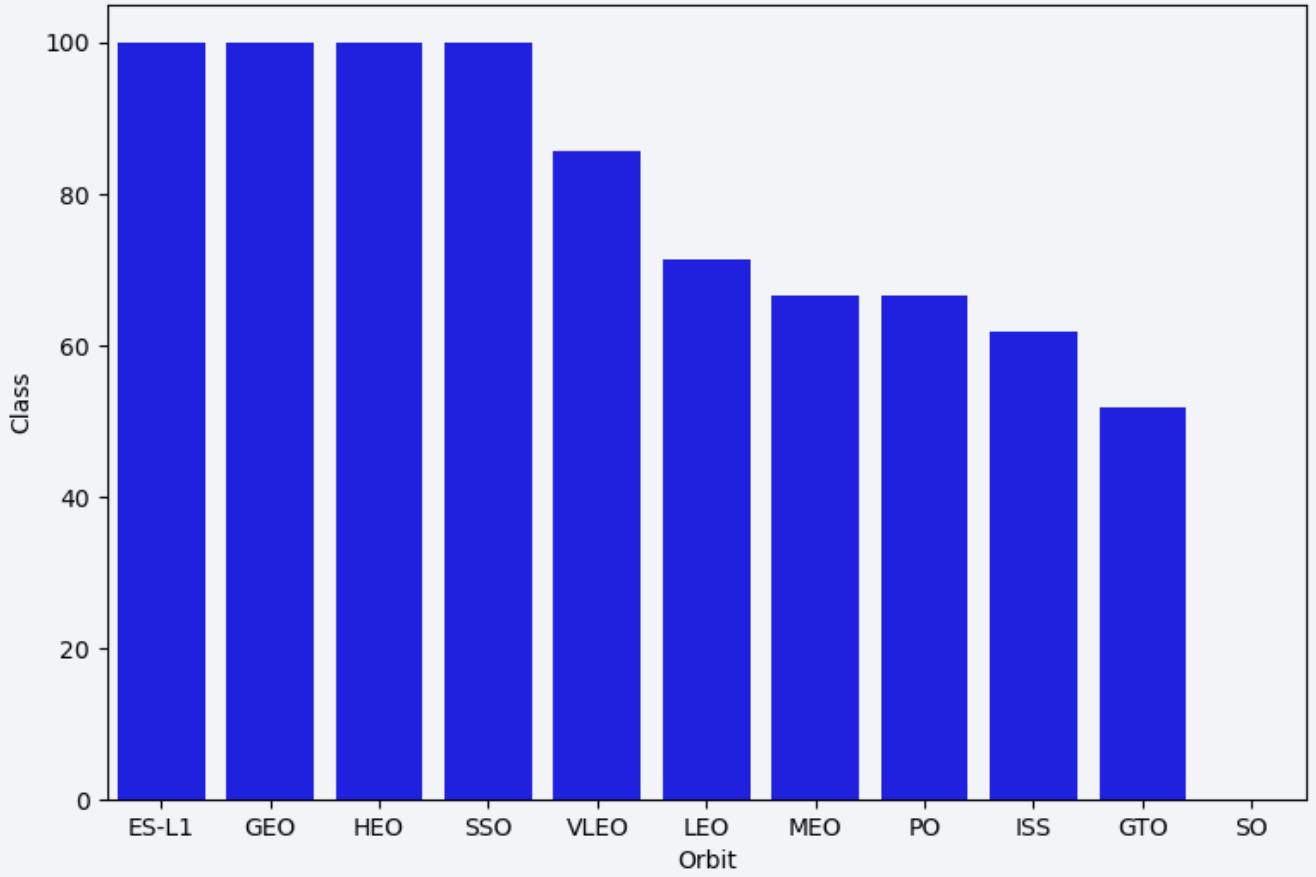
Payload vs. Launch Site

- The blue dots represent the unsuccessful launches while the orange dots represent successful launches.
- CCAFS SLC 40 was a launch site was more rocket launches compare to other.
- VAFB SLC 4E launch site was the less rocket launches and no rocked launched for heavy payload mass
- KSC LC 39A launch site was the more success rate.



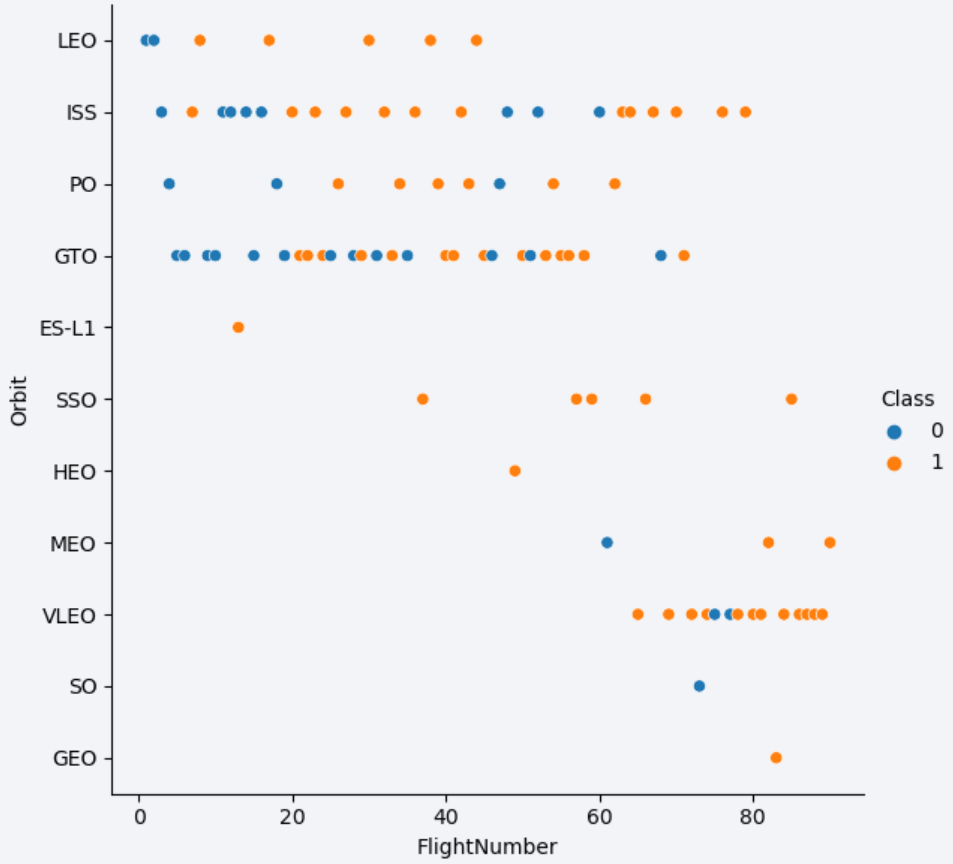
Success Rate vs. Orbit Type

- Orbit ES-L1, GEO, HEO, and SSO has a 100% success rates.
- Orbit did not have any successful launches with a 0% success rate.



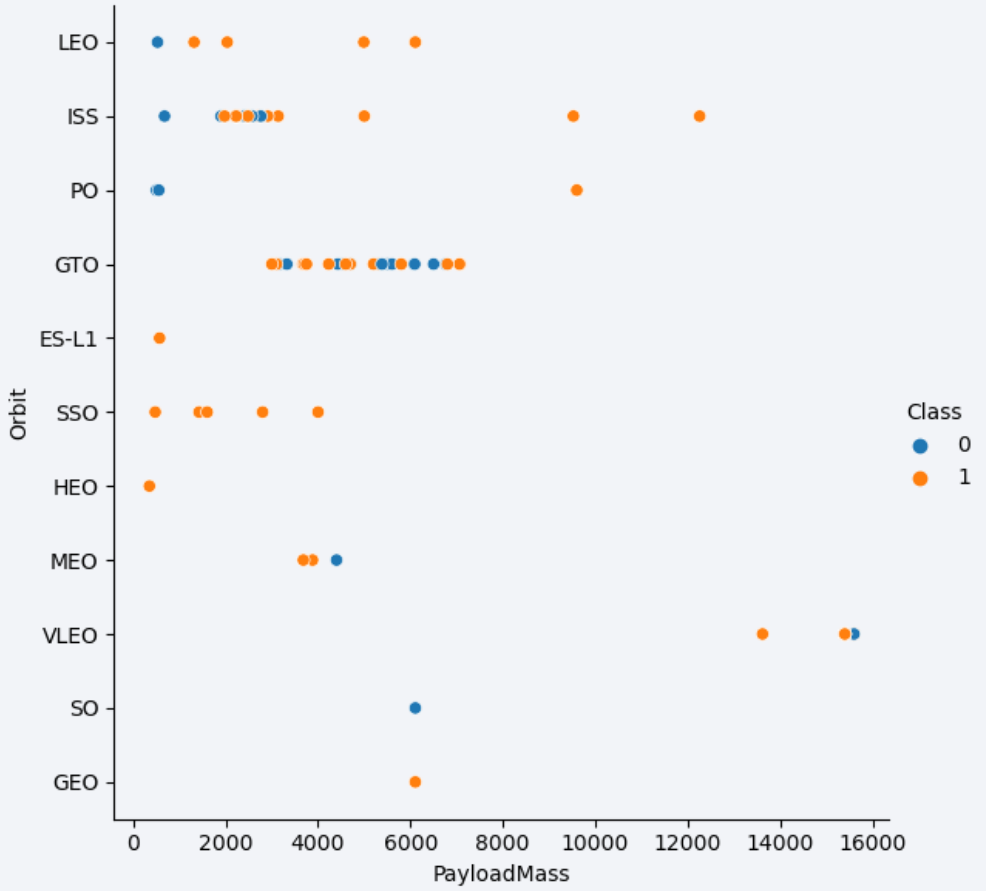
Flight Number vs. Orbit Type

- GEO orbit have only one success rate.
- SO orbit have only one fail.
- ISS orbit have more success rate.
- GTO orbit have combination of both success or fail.



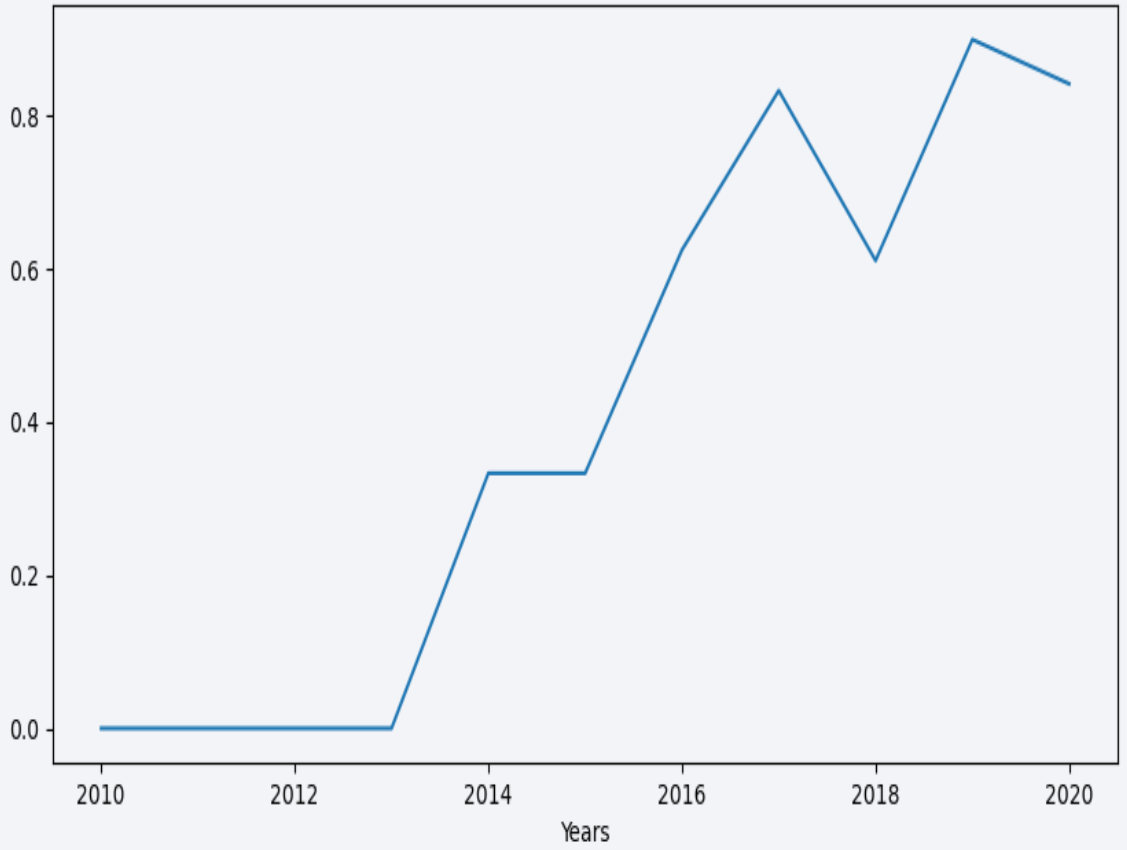
Payload vs. Orbit Type

- As the payloads get heavier, the success rate increases in the PO, SSO, LEO and ISS orbits.
- There seems to be no direct correlation between orbit type and payload mass for GTO orbit as both successful and failed launches are equally present.



Launch Success Yearly Trend

The general trend of the chart shows an increase in landing success rate as the years pass. There is however a dip in 2018 as well as in 2020.



All Launch Site Names

- Display the names of the unique launch sites in the space mission
- The DISTINT clause was used to return only the unique rows from the launch site column.
- The name of the launch sites are CCAFS LC-40, CCAFS SLC-40, KSC LC-39A, VAFB SLC-4E.

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

Launch Site Names Begin with 'CCA'

- Display 5 records where launch sites begin with the string 'CCA'
- The LIMIT and LIKE clauses were used to display only the top five results where the launch site name starts with 'CCA'

launch_site
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40

Total Payload Mass

- Display the total payload mass carried by boosters launched by NASA (CRS).
- The SUM() function was used to the calculate the total payload carried by boosters from NASA from the payload_mass_kg column.

payloadmass
256163

Average Payload Mass by F9 v1.1

- Display average payload mass carried by booster version F9 v1.1
- The WHERE clause was used to filter results so that the calculations were only performed on booster_versions only if they were named “F9 v1.1”

payloadmass
5692

First Successful Ground Landing Date

- List the date when the first successful landing outcome in ground pad was achieved.
- Find the dates of the first successful landing outcome on ground pad.
- The WHERE clause ensured that the results were filtered to match only when the 'landing_outcome' column is 'Success (ground pad)'

1
2010-04-06

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

booster_version
F9 FT B1022
F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes
- The COUNT() function is used to count the number of occurrences of different mission outcomes with the help of the GROUPBY clause applied to the 'mission_outcome' column. A list of the total number of successful and failure mission outcomes returned.

missionoutcomes
44
1

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass
- The MAX() function was used in a subquery to retrieve a list of boosters which have carried the maximum payload mass

boosterversion
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3

2015 Launch Records

- List the failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015.
- The SELECT statement was used to retrieve multiple columns from the table. The YEAR(DATE) function was used to retrieve only those rows with a 2015 launch date.

1	mission_outcome	booster_version	launch_site
10	Success	F9 v1.1 B1012	CCAFS LC-40
11	Success	F9 v1.1 B1013	CCAFS LC-40
2	Success	F9 v1.1 B1014	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
- COUNT() function was used to count the different landing outcomes.
- The WHERE and BETWEEN clauses filtered the results to only include results between 2010-06-04 and 2017-03-20.
- The GROUPBY clause ensure that the counts were grouped by their outcome. The ORDERBY and DESC clauses were used to sort the results by descending order.

landing__outcome
Success (ground pad)
Success (ground pad)
Success (drone ship)
Success (drone ship)
Failure (drone ship)
Controlled (ocean)
Failure (drone ship)
No attempt
No attempt
No attempt
No attempt
No attempt
No attempt
No attempt
Failure (parachute)

Section 3

Launch Sites Proximities Analysis

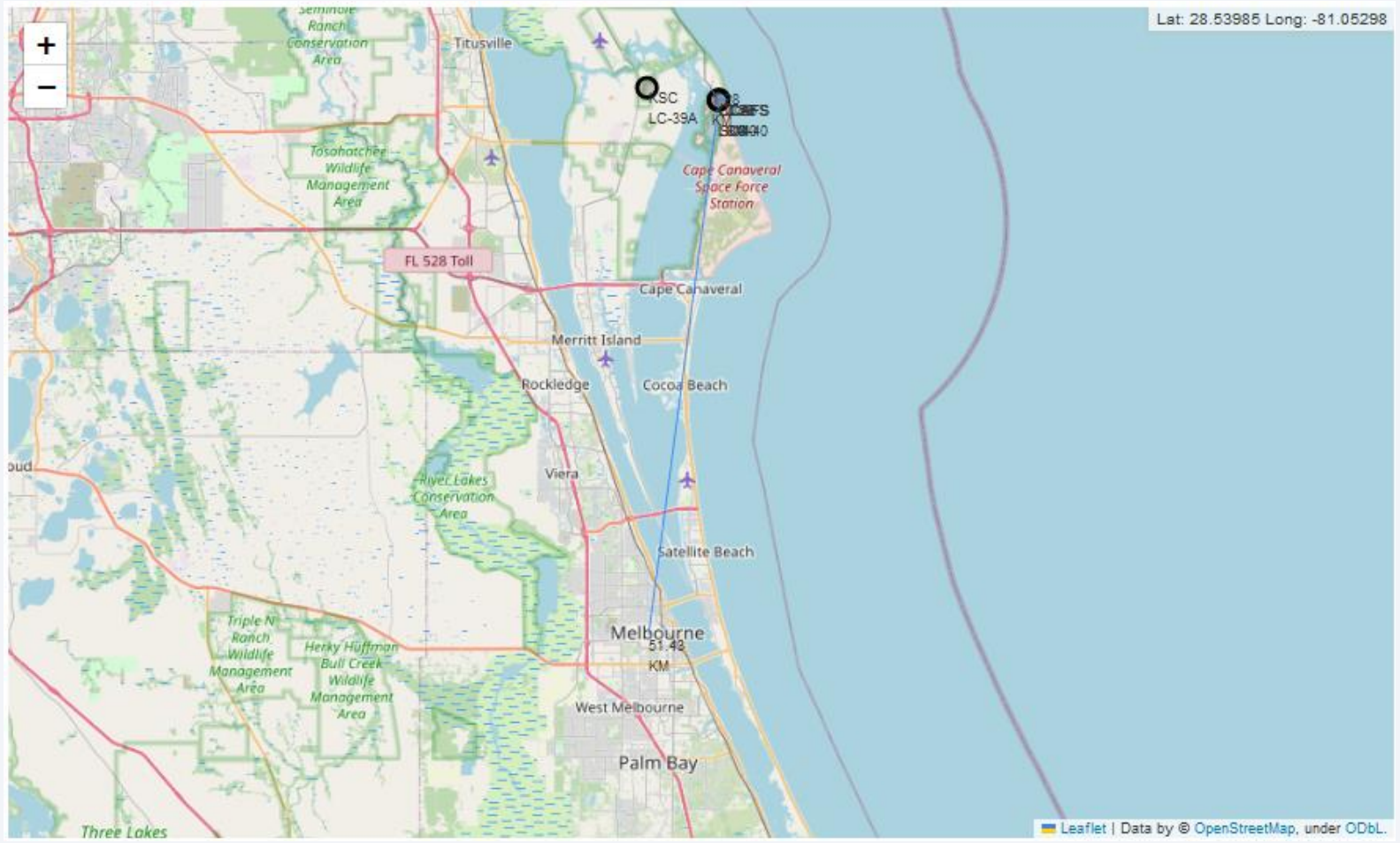
Mark all Launch Sites on a Map

- The yellow markers are indicators of where the locations of all the SpaceX launch sites are situated in the US.
- The launch sites have been strategically placed near the coast

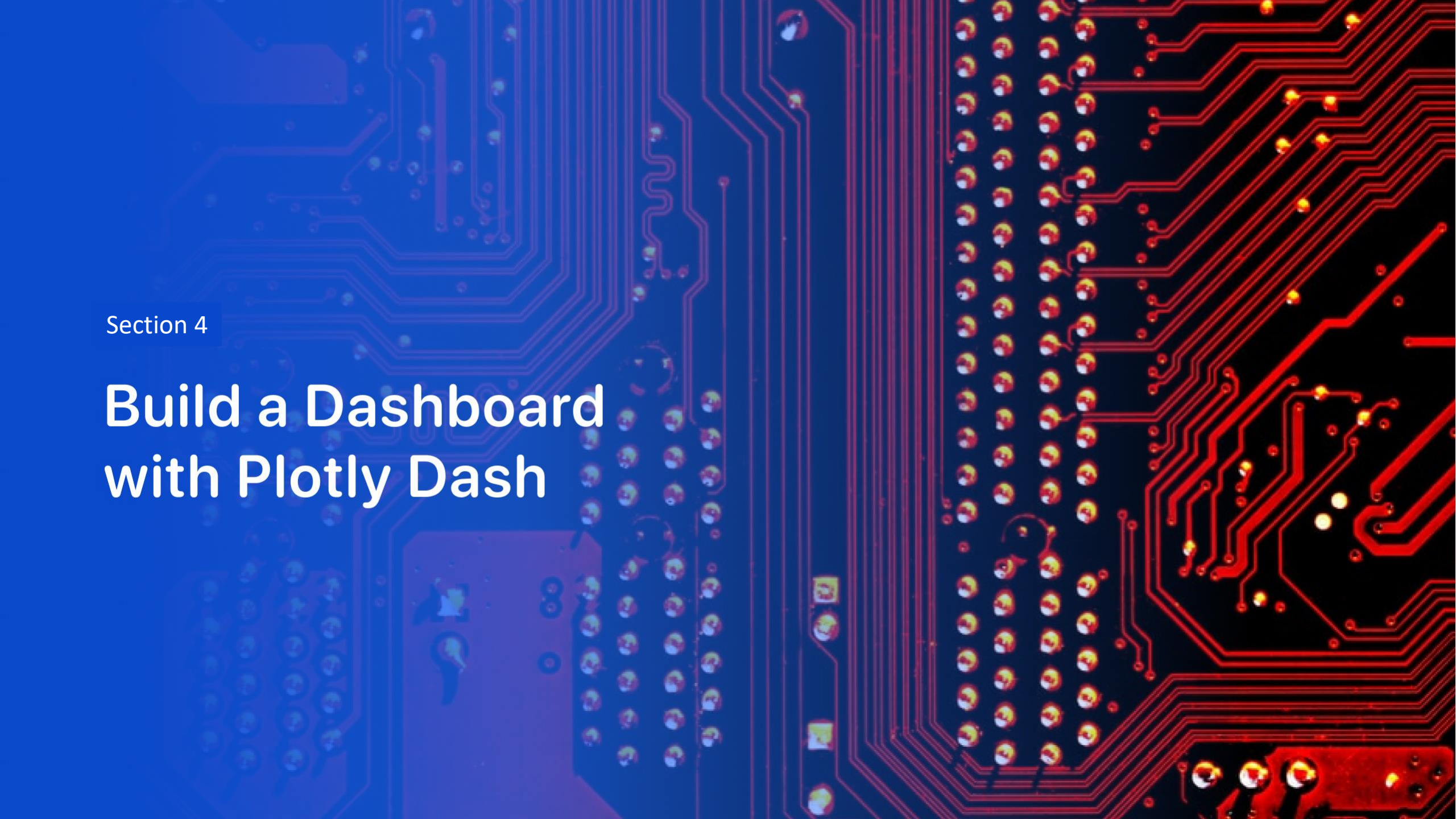
Mark the Success/Failed Launches for each Site on the Map

When we zoom in on a launch site, we can click on the launch site which will display marker clusters of successful landings (green) or failed landing (red).

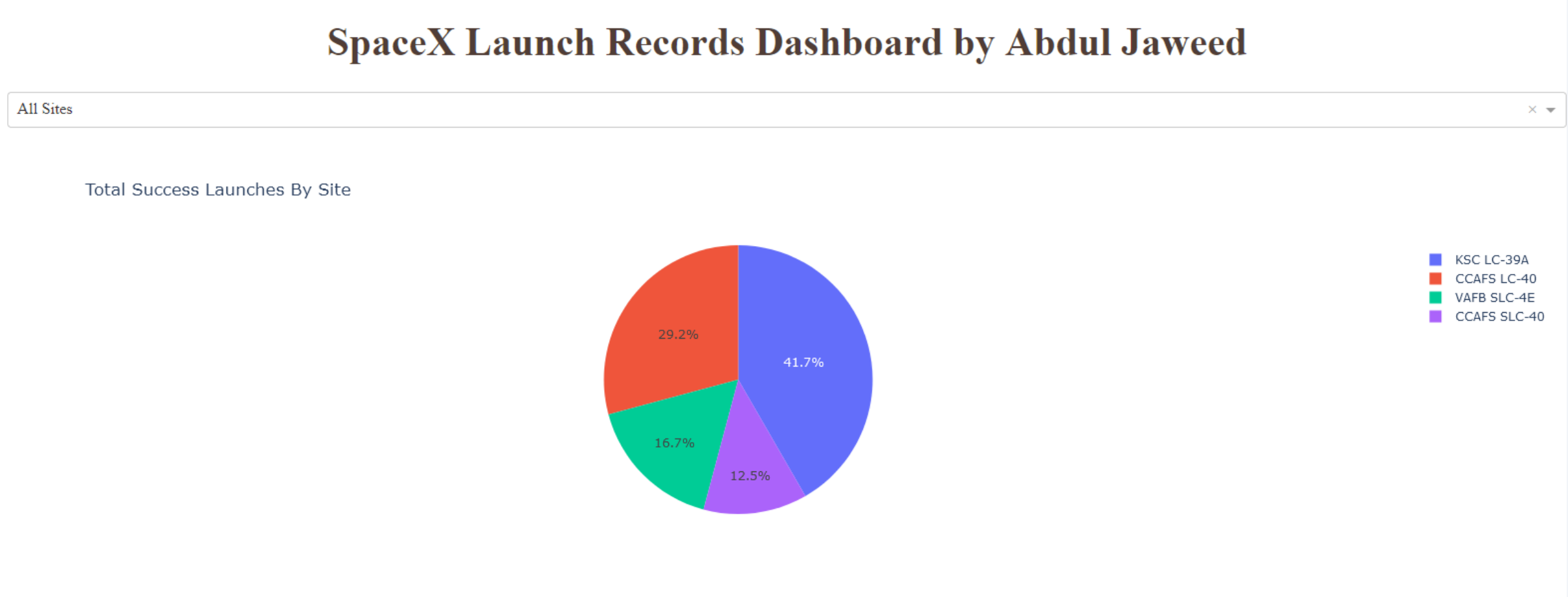
Calculate the Distances Between a Launch Site to its Proximities



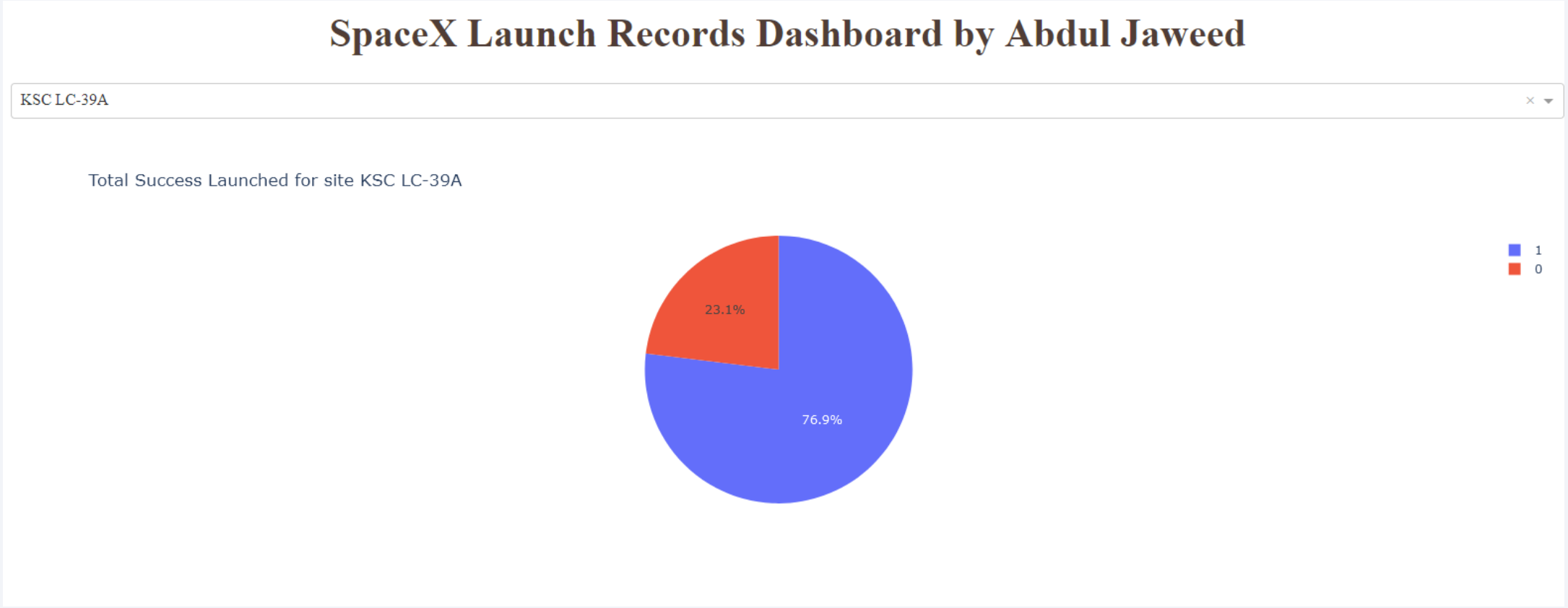
- The generated map shows that the selected launch site is close to a highway for transportation of personnel and equipment. The launch site is also close to the coastlines for launch failure testing.
- The launch sites also maintain a certain distance from the cities. (Can be viewed in notebook).


Section 4
Build a Dashboard with Plotly Dash

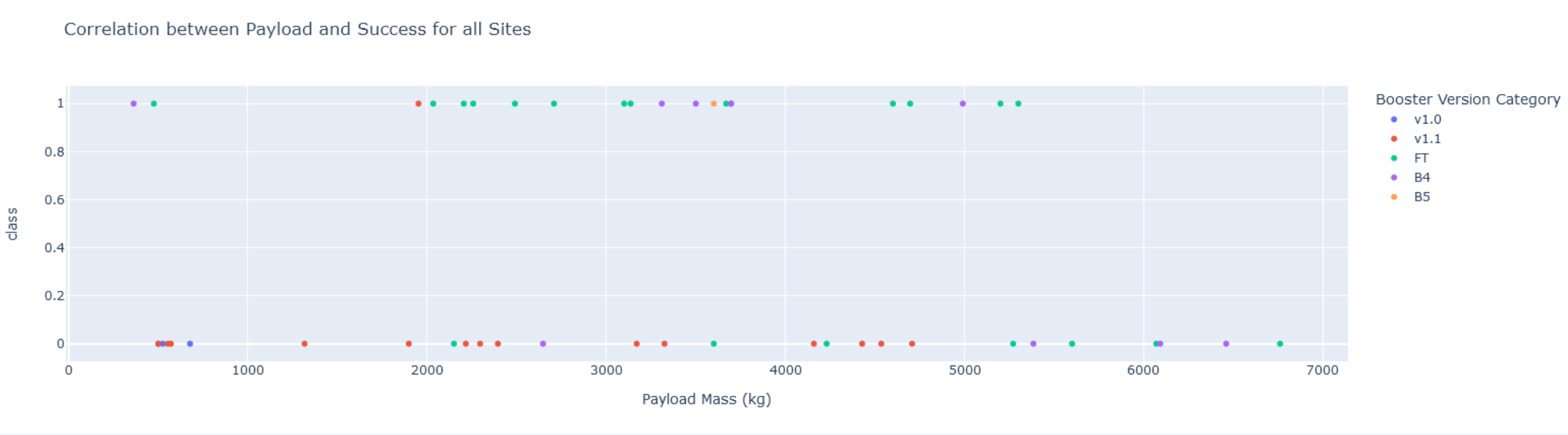
Total Successful Launches by Site



Launch Site with Highest Success Ratio



Correlation Between Payload and Success for all Sites

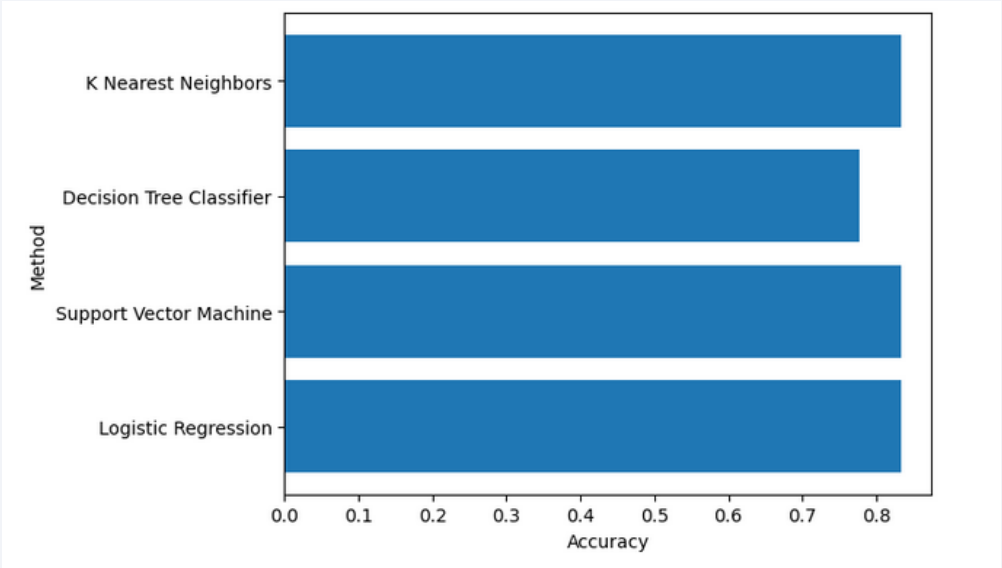


Section 5

Predictive Analysis (Classification)

Classification Accuracy

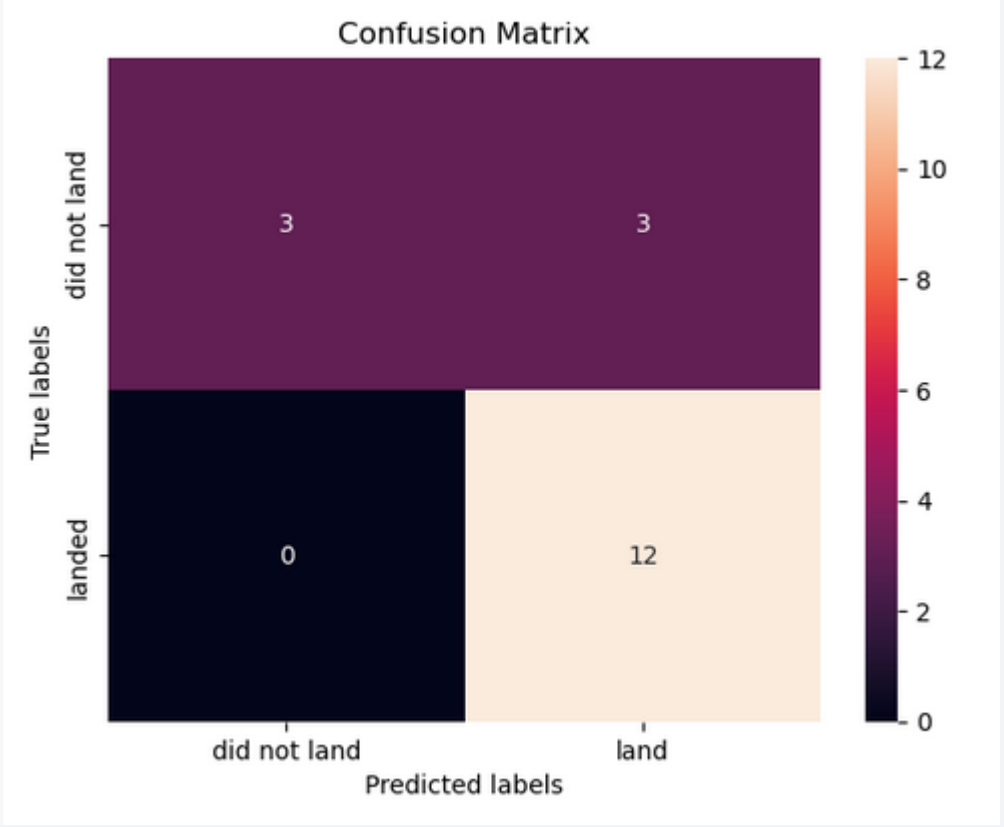
- Except Decision Tree the other three model like K Nearest Neighbors, Support Vector Machine and Logistic Regression Machine Learning Algorithms perform same accuracy 83%.



	method	accuracy
0	Logistic Regression	0.833333
1	Support Vector Machine	0.833333
2	Decision Tree Classifier	0.777778
3	K Nearest Neighbors	0.833333

Confusion Matrix

- The model predicted 12 successful landings when the True label was successful (True Positive) and 3 unsuccessful landings when the True label was failure (True Negative).
- The model also predicted 3 successful landings when the True label was unsuccessful landing (False Positive).
- The model generally predicted successful landings.



Conclusions

- The analysis showed that there is a positive correlation between number of flights and success rate as the success rate has improved over the years.
- The results of the (EDA) exploratory data analysis revealed the success rate of the SpaceX Falcon 9 rocket landings is 66%.
- Orbit ES-L1, GEO, HEO, and SSO has a 100% success rates.
- I deploy this dashboard on render web hosting service for better experience.
- Except Decision Tree the other three model like K Nearest Neighbors, Support Vector Machine and Logistic Regression Machine Learning Algorithms perform same accuracy 83%.

Appendix

- GitHub Repository Link : https://github.com/Abdul-Jaweed/SpaceX-Falcon9-Landing-Prediction

- Dashboard Link : https://spacex-falcon9-landing-prediction.onrender.com/

- Coursera Project Link : https://www.coursera.org/learn/applied-data-science-capstone?specialization=ibm-data-science#syllabus

Thank you!