IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Name>
<Date>

Outline

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

Executive Summary

In this endeavor we employ Webscraping methodologies with API calls. Data cleaning and data prep strategies. SQL queries for data insights, as well as data visualization tools, and interactive dashboards. We utilize machine learning models such as KNN and SVM and logistic regression models to produce the final model with a accuracy of 0.78

Introduction

As of 16 November 2024, rockets from the Falcon 9 family have been launched 408 times, with 405 full mission successes, three failures(SpaceX 2010)(Wikipedia).

Launching Space Rockets is a multibillion dollar endeavor, that is both crutial to progessing science in the public and private sectors.

Methods to reduce cost and wastes of rocket launches are many such as reusable rockets. However prediction methods of successful launches may be possible in order to reduce chances of failure using know features of successful launches.

We aim to find a model to predict successful launch sites.

SpaceX 2018

Section 1

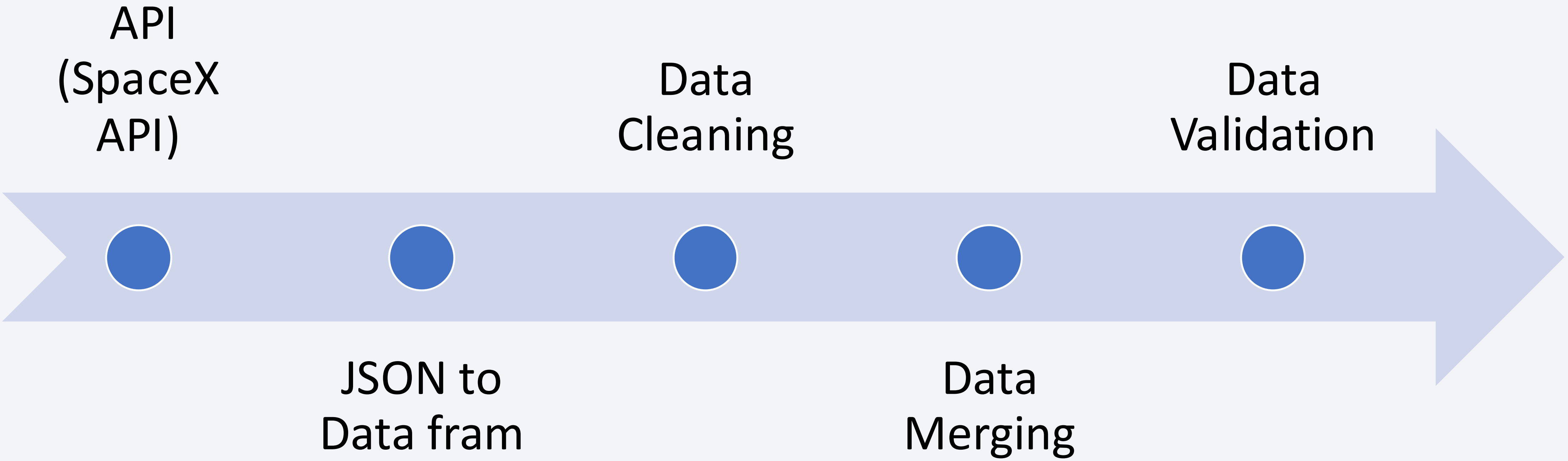
Methodology

Methodology

Executive Summary

- Data collection methodology:
  - The data was sourced from the SPACEX RESTfuL API were separated features were combined into one data set. The data was also web scraped from various website such as Wikipedia launch information on spacex heavy carriers
- Perform data wrangling
  - The data was processed and cleaned for specific carrier such as falcon 9 and null values were removed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - The model was built using a multiple linear regression model using features engineered from launch site characteristicss

Data Collection



Data Collection – SpaceX API

- https://github.com/oedatainsight/Capsptone-Project-SpaceX/blob/master/jupyter-labs-spacex-data-collection-api-v2.ipynb

Data Collection - Scraping

- https://github.com/oedatainsight/Capsptone-Project-SpaceX/blob/master/labs-jupyter-spacex-Data%20wrangling-v2.ipynb

Define Data Need

Identify required fields

Identify URLs to scrape

Request Data from Web Pages

Parse HTML Content

Extract Target Data

Clean Extracted Data

|Save Data to File

Data Wrangling

- https://github.com/oedatainsight/Capsptone-Project-SpaceX/blob/master/labs-jupyter-spacex-Data%20wrangling-v2.ipynb

Define Data Need

Identify required fields

Identify URLs to scrape

Request Data from Web Pages

Parse HTML Content

Extract Target Data

Clean Extracted Data

|Save Data to File

EDA with Data Visualization

- Summarize what charts were plotted and why you used those charts
- We used a combination of Line plots that could clearly show difference over time. Scatter plots that can show differences in trends and between groups, and pie charts for clear comparisons of launch rate between groups

- https://github.com/oedatainsight/Capsptone-Project-SpaceX/blob/master/jupyter-labs-eda-dataviz-v2.ipynb

EDA with SQL

- Using bullet point format, summarize the SQL queries you performed
  - We used SQL queries to find
    - The launch sites
    - Launch sites with CCA
    - Average Payload carried
    - First successful landing achieves
    - Successes in drones
    - Total number of successful and failed missions
    - Version carrying the maximum payloads
    - Counts of landing outcomes in date ranges
  - https://github.com/oedatainsight/Capsptone-Project-SpaceX/blob/master/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- The map objects show the location of the launch sites in America with the magnitude of count being the circles and the frequency of successes being the color. WE also marked the proximity to landmarks such as trains and cities for each of the launch sites as features for success.

- https://github.com/oedatainsight/Capsptone-Project-SpaceX/blob/master/lab-jupyter-launch-site-location-v2.ipynb

Build a Dashboard with Plotly Dash

- We added a pie char to the dashboard for ease and clarity of the proportion of successful launches at each site. We also added a payload VS success fate as that does play a role int eh success rates too.

- https://github.com/oedatainsight/Capsptone-Project-SpaceX/blob/master/Dashboard_Ploty.py

Predictive Analysis (Classification)

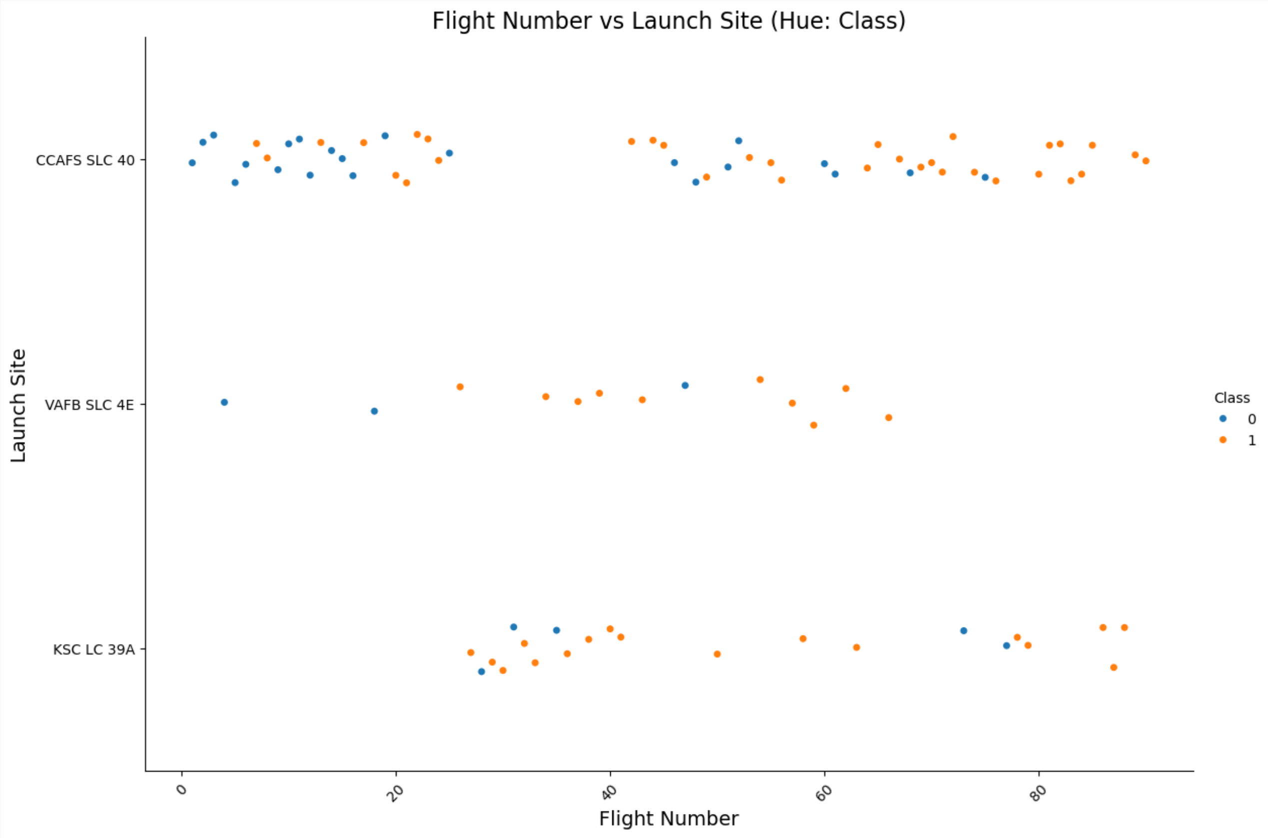
- We made the model using a logistic regression classified to see if we can predict successful launches.
- We utilized SVM, KNN, and logistic regression and tested the accuracy against eachother to determine the most accurate model.
- We could also test by including and removing features to see if we can get a higher R^2 or higher accuracy
- https://github.com/oedatainsight/Capsptone-Project-SpaceX/blob/master/SpaceX-Machine-Learning-Prediction-Part-5-v1.ipynb

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

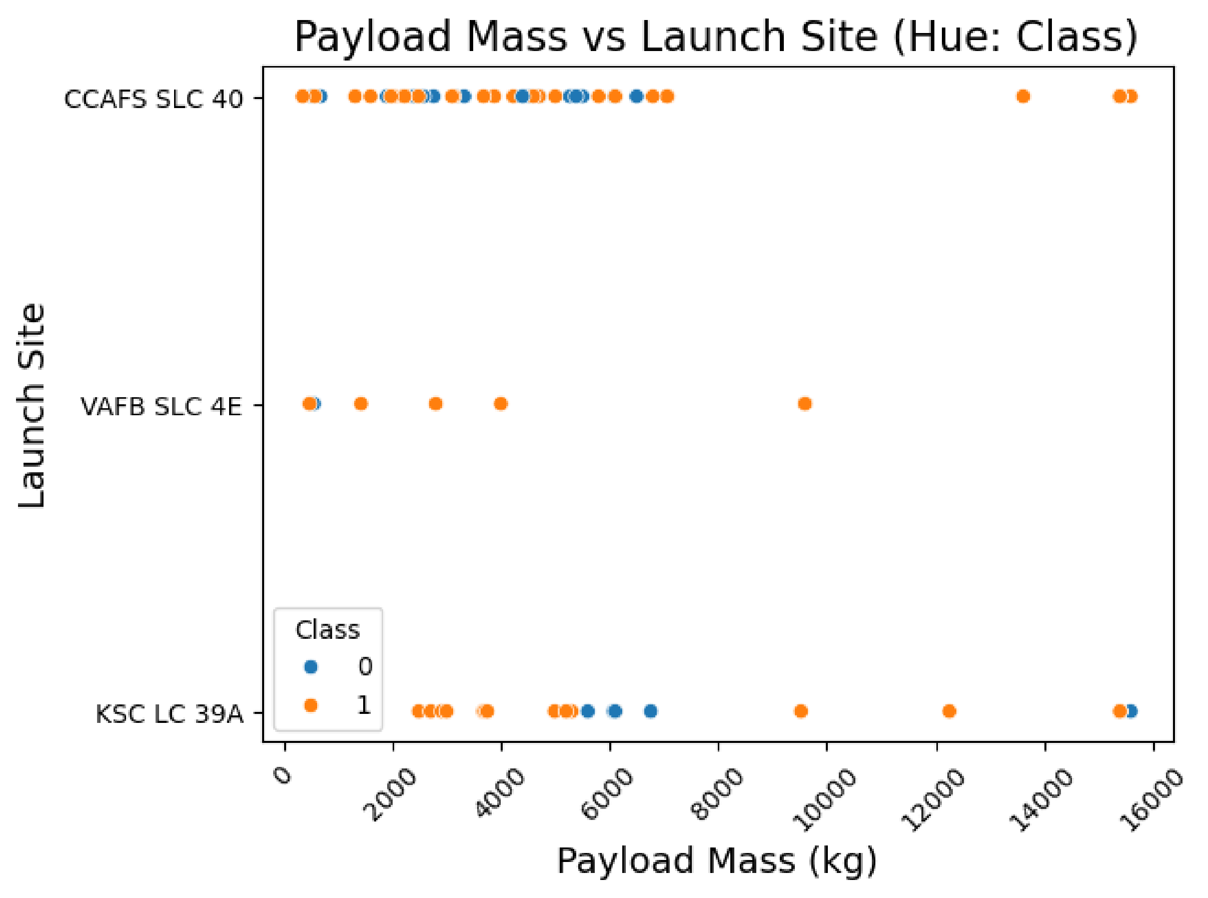
- Show a scatter plot of Flight Number vs. Launch Site
- Show the screenshot of the scatter plot with explanations



Payload vs. Launch Site

- Show a scatter plot of Payload vs. Launch Site

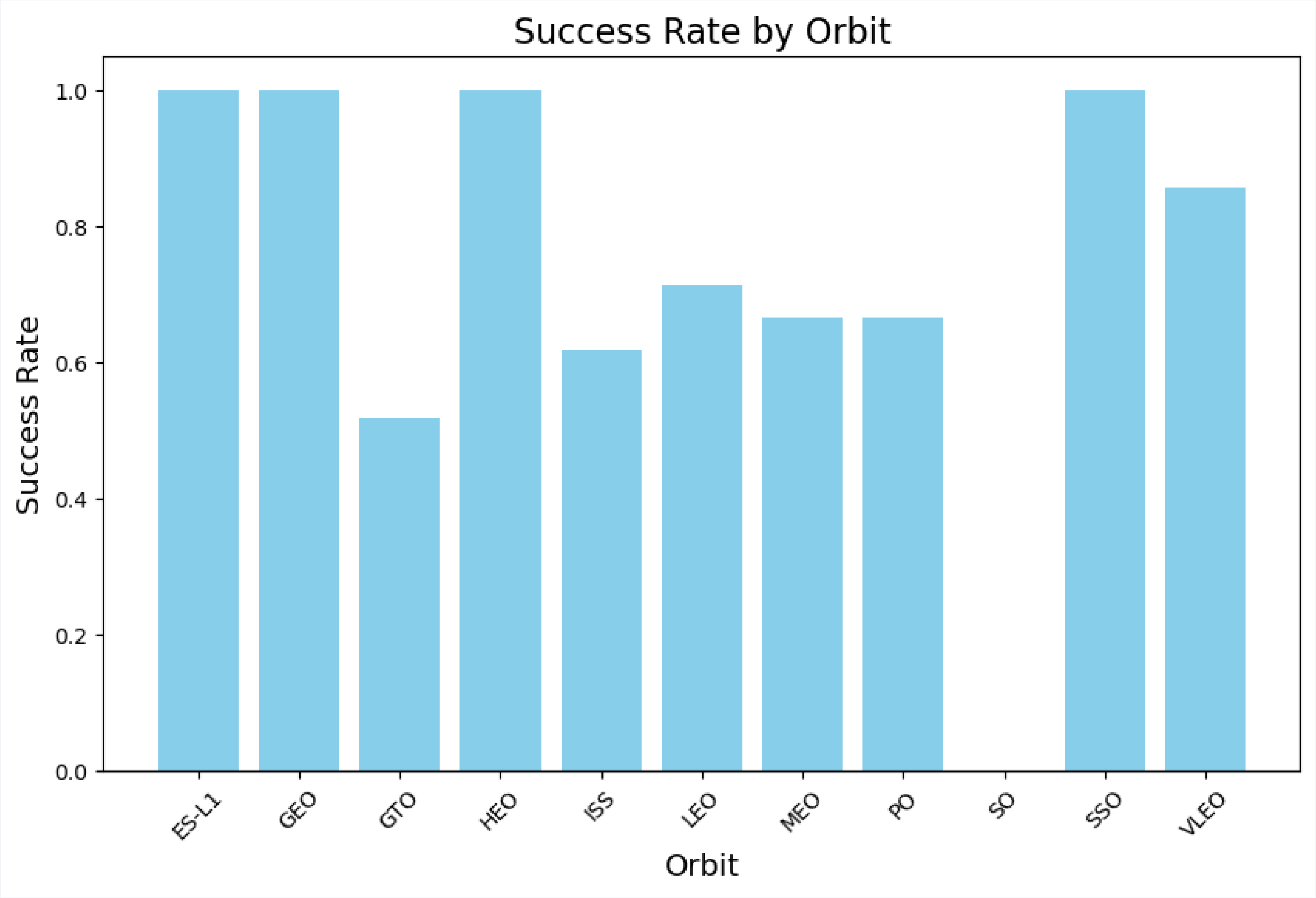
- Show the screenshot of the scatter plot with explanations



Success Rate vs. Orbit Type

- Show a bar chart for the success rate of each orbit type

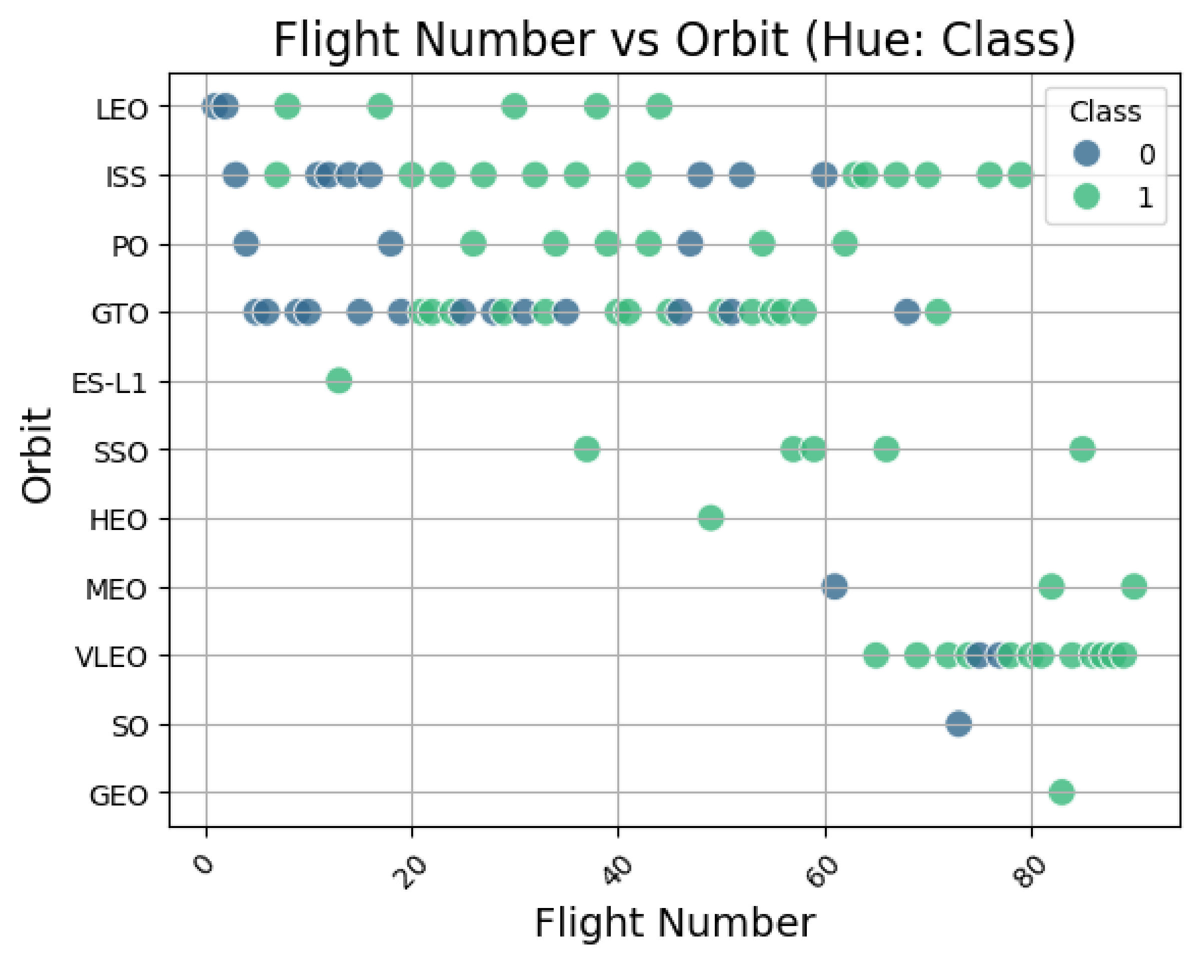
- Show the screenshot of the scatter plot with explanations



Flight Number vs. Orbit Type

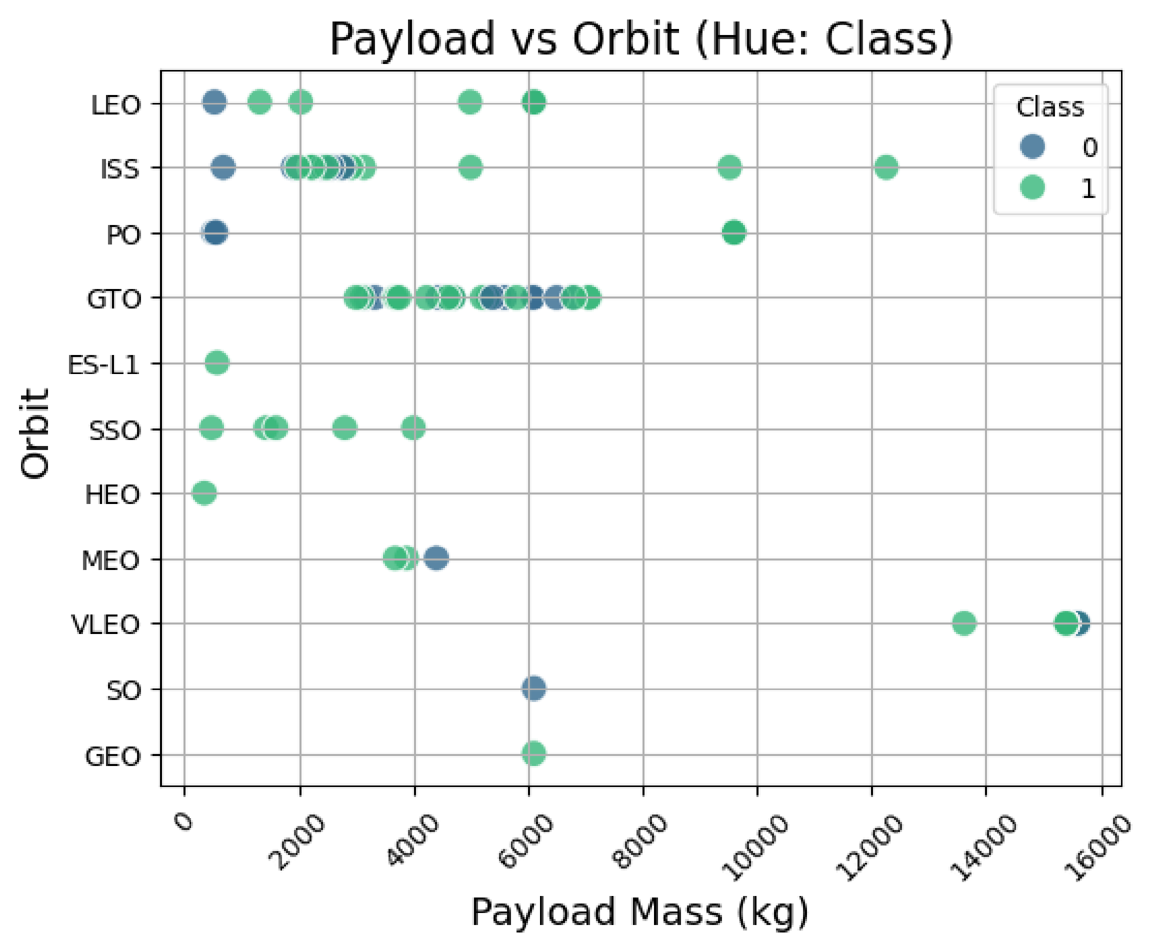
- Show a scatter point of Flight number vs. Orbit type

- Show the screenshot of the scatter plot with explanations



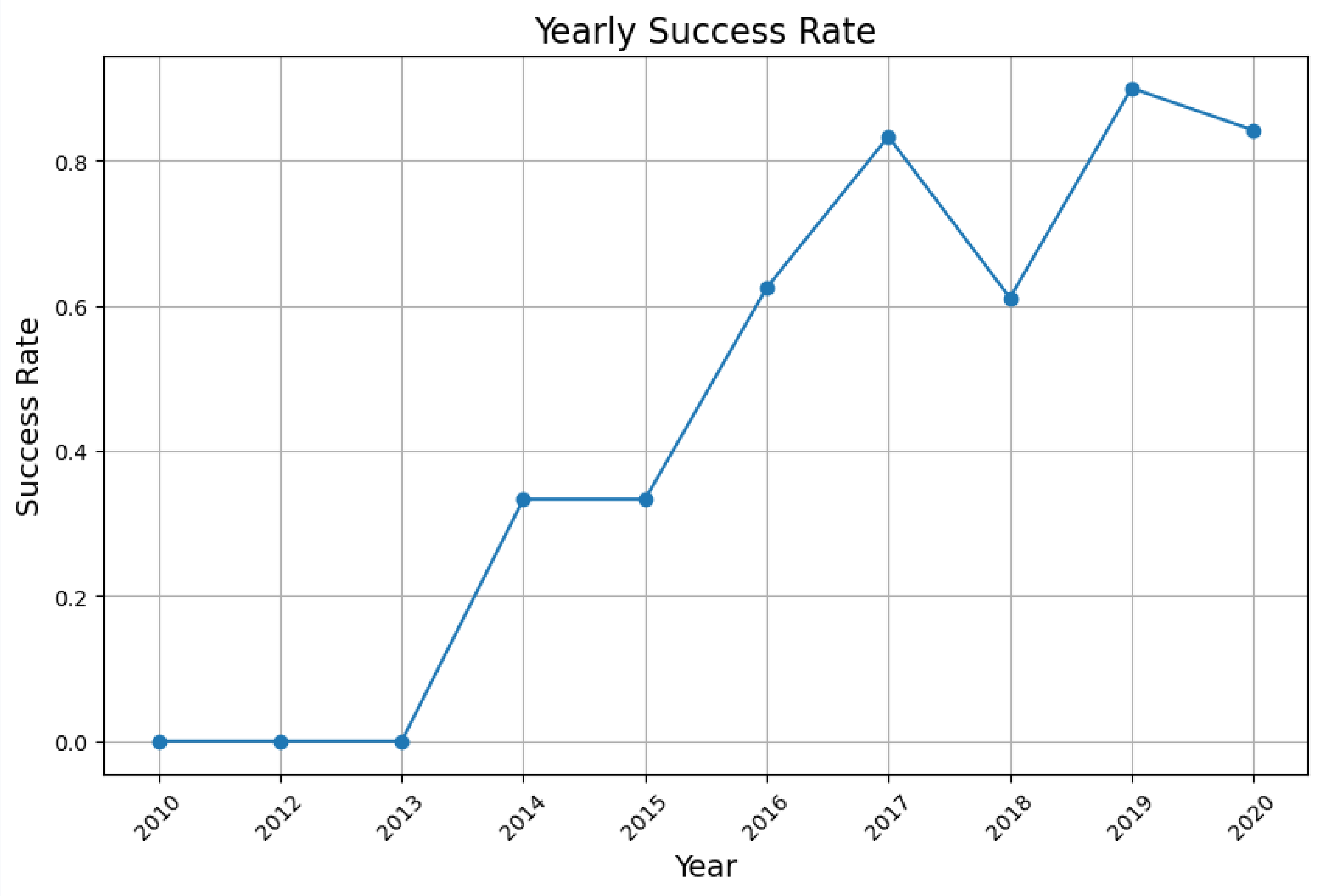
Payload vs. Orbit Type

- Show a scatter point of payload vs. orbit type
- Show the screenshot of the scatter plot with explanations



Launch Success Yearly Trend

- Show a line chart of yearly average success rate
- Show the screenshot of the scatter plot with explanations



All Launch Site Names

- Find the names of the unique launch sites

- Present your query result with a short explanation here
- %sql SELECT DISTINCT "Launch_Site" FROM SPACEXTABLE;

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

Launch Site Names Begin with 'CCA'

Find 5 records where launch sites begin with `CCA`

Present your query result with a short explanation here

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

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

- Calculate the total payload carried by boosters from NASA
- %sql SELECT SUM("PAYLOAD_MASS_KG_") AS TotalPayload FROM SPACEXTABLE WHERE "Customer" LIKE '%NASA%';

- Present your query result with a short explanation here

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
- %sql SELECT AVG("PAYLOAD_MASS_KG_") AS AveragePayload FROM SPACEXTABLE WHERE "Booster_Version" = 'F9 v1.1';
- Present your query result with a short explanation here

First Successful Ground Landing Date

Find the dates of the first successful landing outcome on ground pad

Present your query result with a short explanation here

```
%sql SELECT "Date" FROM SPACEXTABLE
WHERE "Landing_Outcome" = 'Success
(ground pad)' ORDER BY "Date" ASC LIMIT 1;
```

Date
2015-12-22

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
- %sql SELECT "Booster_Version" FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Success (drone ship)' AND "PayloadMass" > 4000 AND "PayloadMass" < 6000;
- Present your query result with a short explanation here

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

- SELECT * SUM("Class") AS Total_Success, COUNT(*) - SUM("Class") AS Total_Failure
- FROM SPACEXTABLE;

- Present your query result with a short explanation here

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass
- %sql SELECT DISTINCT "Booster_Version" FROM SPACEXTABLE WHERE "PayloadMass" = (SELECT MAX("PayloadMass") FROM SPACEXTABLE) LIMIT 20;
- Present your query result with a short explanation h

there were several more but we have truncated the list for clarity

booster_Version
F9 v1.0 B0003
F9 v1.0 B0004
F9 v1.0 B0005
F9 v1.0 B0006
F9 v1.0 B0007
F9 v1.1 B1003
F9 v1.1
F9 v1.1 B1011
F9 v1.1 B1010
F9 v1.1 B1012
F9 v1.1 B1013
F9 v1.1 B1014
F9 v1.1 B1015
F9 v1.1 B1016
F9 v1.1 B1018
F9 FT B1019
F9 v1.1 B1017
F9 FT B1020
F9 FT B1021.1
F9 FT B1022

2015 Launch Records

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
- %sql SELECT "Landing_Outcome", "Booster_Version", "LaunchSite" FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Failure (drone ship)' AND strftime('%Y', "Date") = '2015';

- Landing_Outcome Booster_Version "LaunchSite"
- Failure (drone ship) F9 v1.1 B1012 LaunchSite
- Failure (drone ship) F9 v1.1 B1015

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

- Here we rank the count of landing outcomes between the date 2010-06-04 and 2017-03-20, in descending order
- %sql SELECT "Landing_Outcome", COUNT(*) AS Outcome_Count FROM SPACEXTABLE WHERE "Date" BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY "Landing_Outcome" ORDER BY Outcome_Count DESC;

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

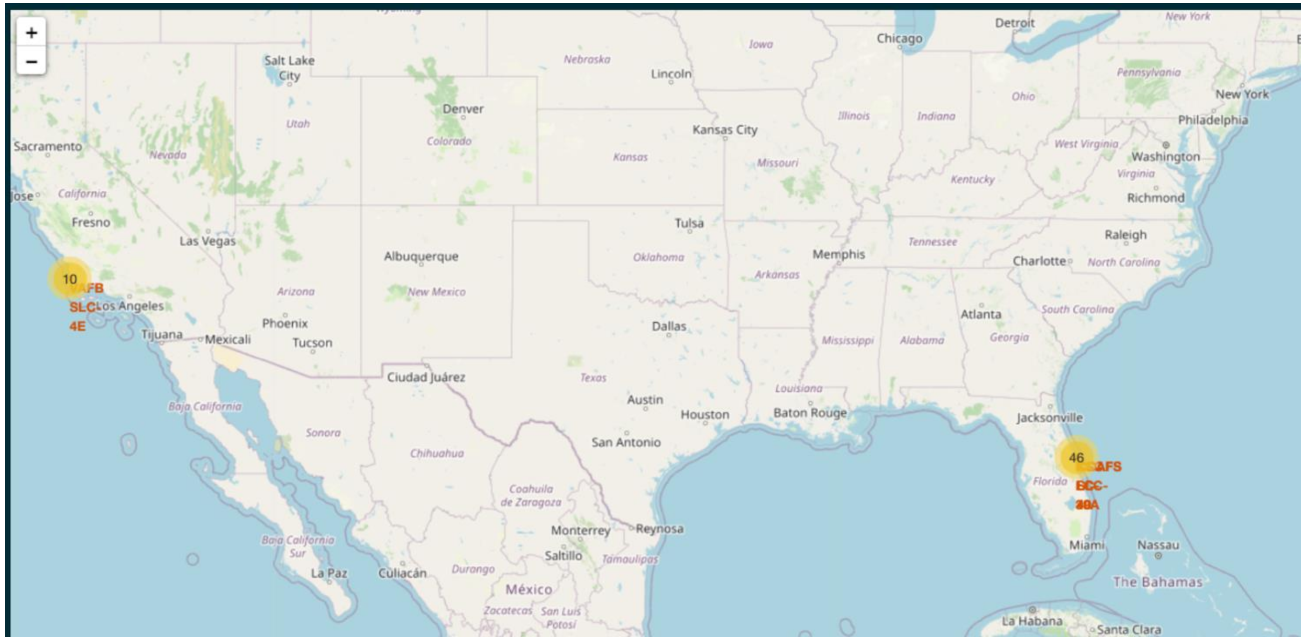
Section 3

Launch Sites Proximities Analysis

Launch Site Analysis

Explore the generated folium map and make a proper screenshot to include all launch sites' location markers on a global map

Explain the important elements and findings on the screenshot



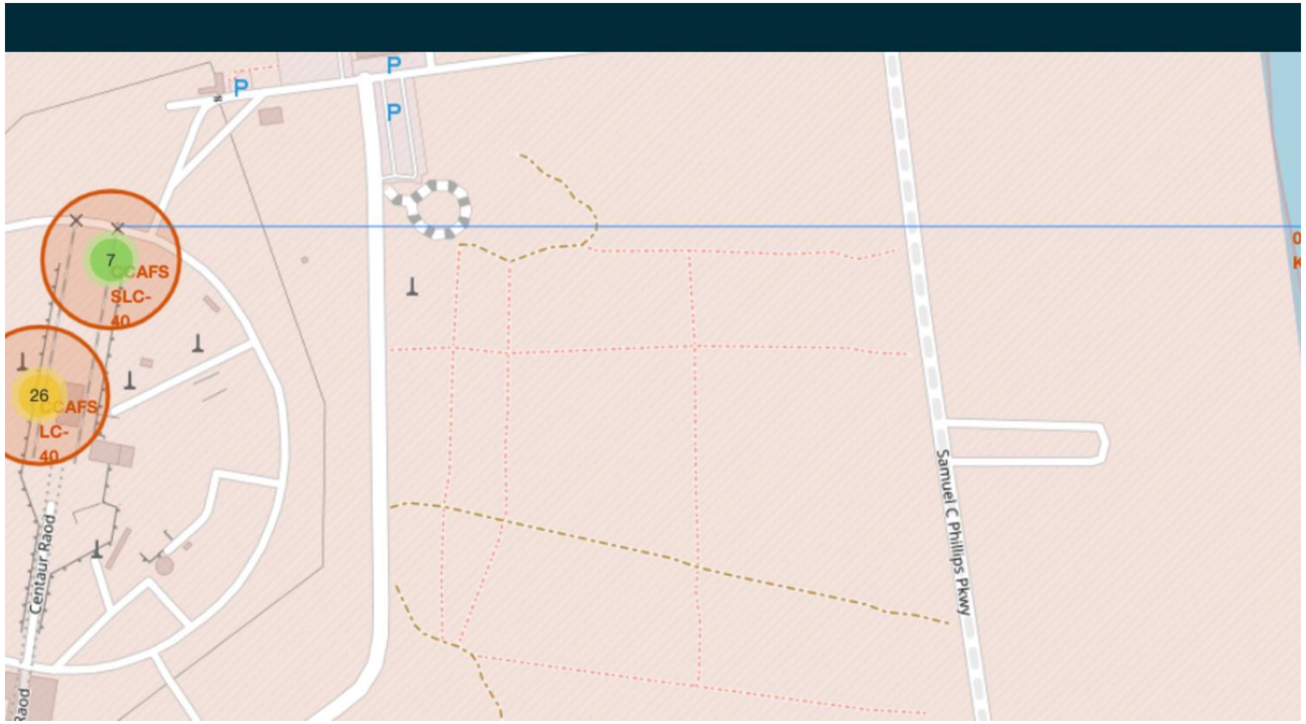
Launch Outcomes per Launch Site

- Explore the folium map and make a proper screenshot to show the color-labeled launch outcomes on the map

- Explain the important elements and findings on the screenshot

Proximities Calculations to launch Site

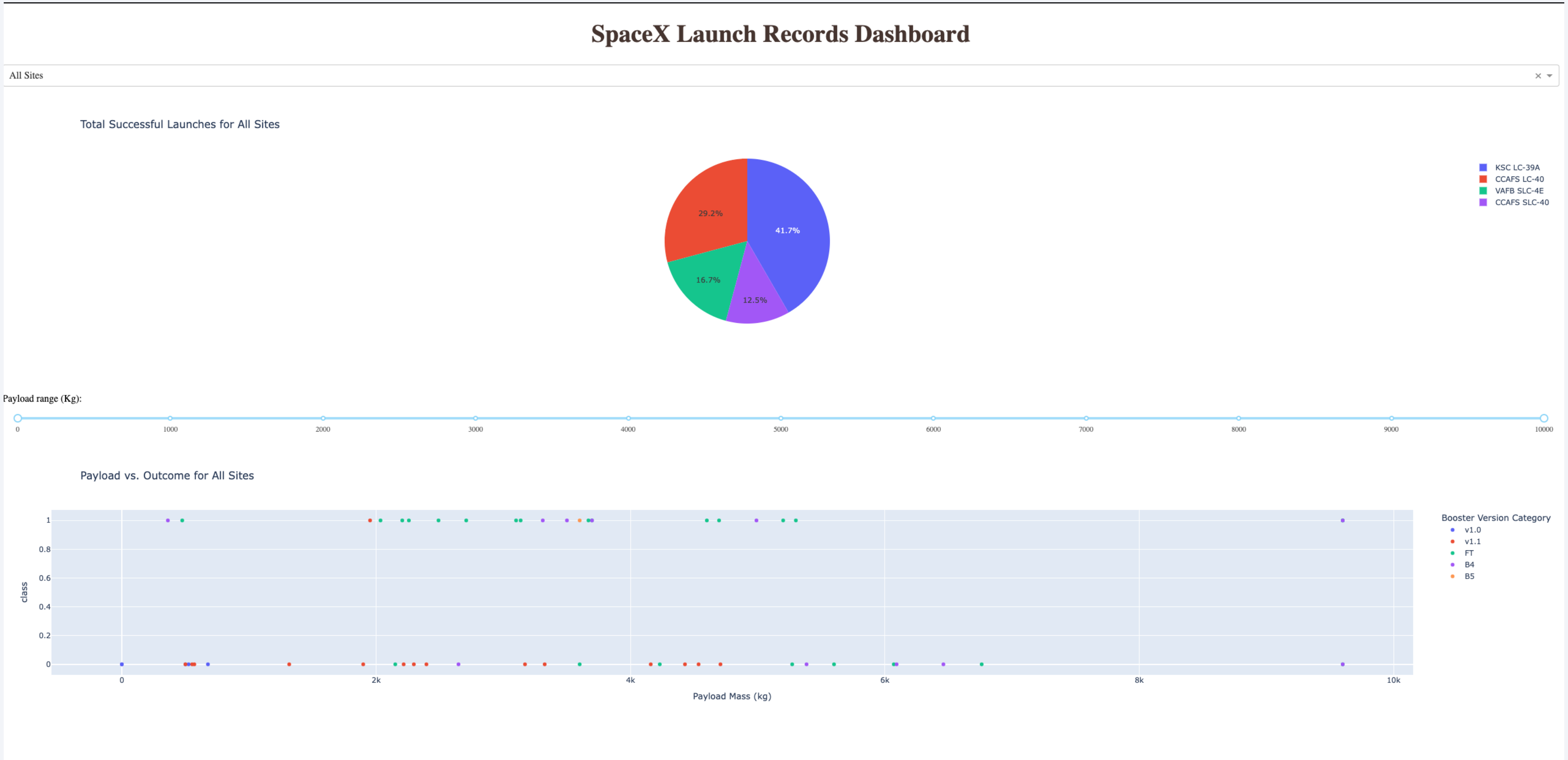
Launch site displayed as proximal to landmarks such as railways and cities as a factor for launch site success



Section 4

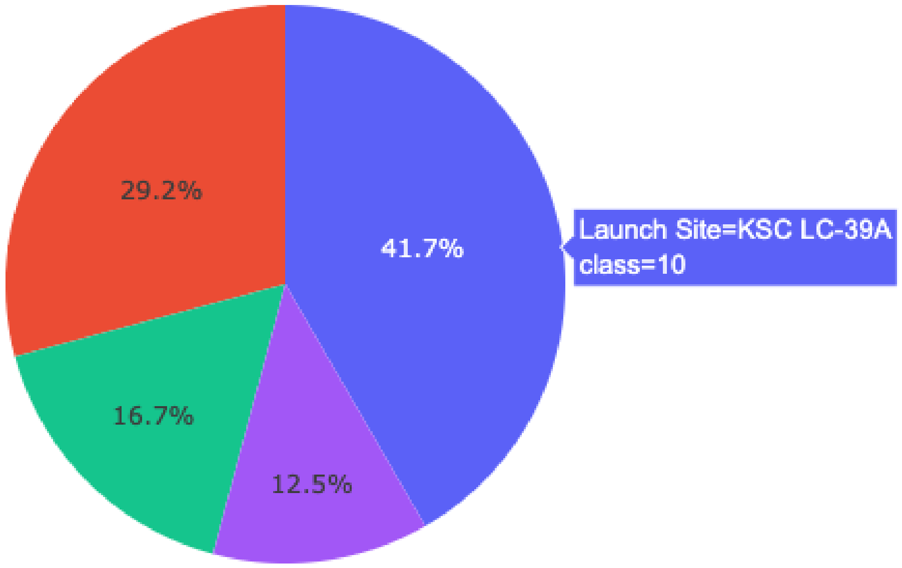
Build a Dashboard with Plotly Dash

Launch Success Counts per Site



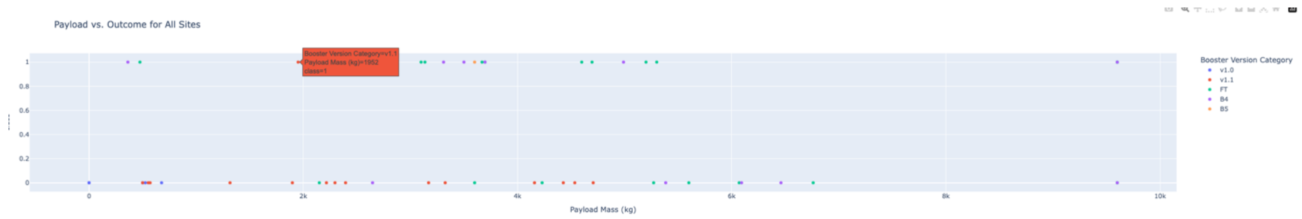
Highest Launch site Ratio

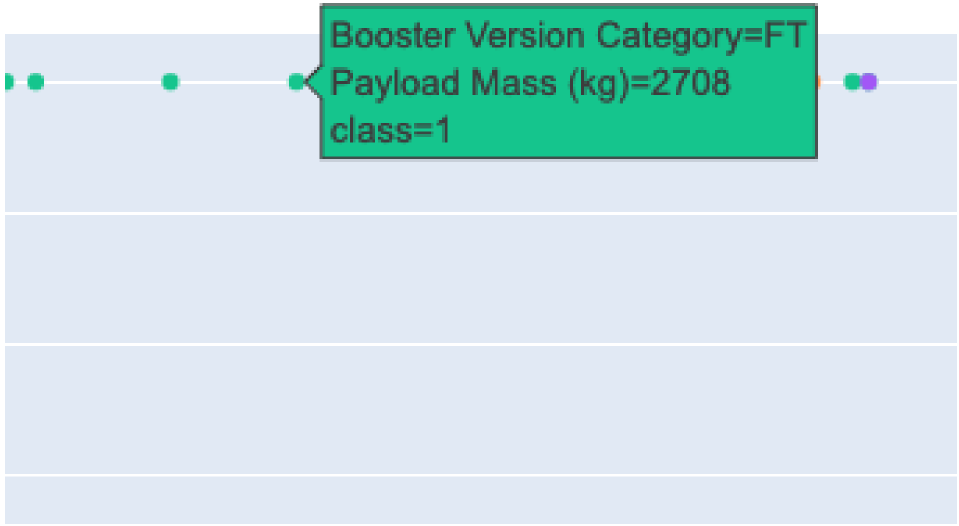
- KSC LC-389A Class 10 with a record of 41.7% launches that were successful in this site



Payload Vs. Launch Outcomes

This is a Dashboard Representation of the Payload vs launch outcomes



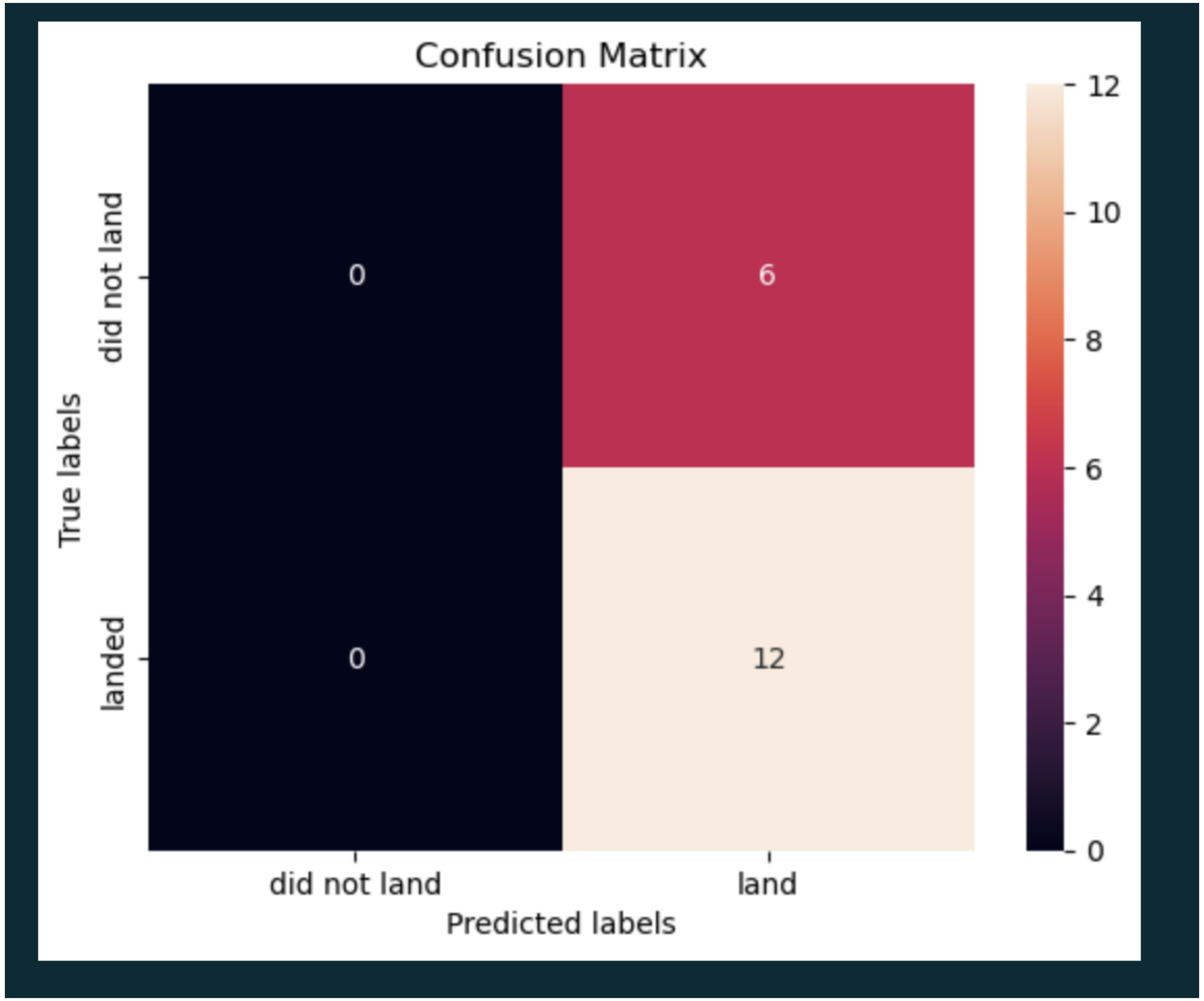


Section 5

Predictive Analysis (Classification)

Confusion Matrix

Show the There is a high miss rate in the confusion matrix when predicted true labels of that that did not land



Conclusions

- Logisitic regression preformed the best to predict successful launches
- SVM was close but took too long to run.
- There may be other factors that contribute to successful launches.
- More research needs to be done and data collected

Appendix

- https://github.com/oedatainsight/Capsptone-Project-SpaceX/blob/master/spacex_launch_geo.csv
- https://github.com/oedatainsight/Capsptone-Project-SpaceX/blob/master/spacex_launch_dash.csv
- https://github.com/oedatainsight/Capsptone-Project-SpaceX/blob/master/Dashboard_Ploty.py

Classification Accuracy

- Visualize the built model accuracy for all built classification models, in a bar chart
- Find which model has the highest classification accuracy

Thank you!