IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

SAMUEL GITONGA

January 1, 2023 Type your text

Outline

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

Executive Summary

- SpaceX is a private American aerospace manufacturer and space transport services company founded in 2002 by Elon Musk.
- The company has developed the Falcon 1 and Falcon 9 launch vehicles, both designed to be reusable, and the Dragon spacecraft, which is flown into orbit by the Falcon 9 launch vehicle to supply the International Space Station (ISS) with cargo and to ferry astronauts to and from the ISS. SpaceX's ultimate goal is to reduce space transportation costs and enable the colonization of Mars.
- Given mission data such as payload amount and intended orbit, the models developed in this report were able to predict the successful landing of the first-stage stage rocket booster within an accuracy level of 83.3%.

Introduction

- The commercial space era is here, and firms are making space travel accessible to everybody. Among SpaceX's successes is the delivery of spacecraft to the International Space Station. SpaceX is able to accomplish this because rocket launches are very affordable.
- SpaceX offers Falcon 9 rocket launches on its website for 62 million dollars; other companies charge up to 165 million dollars apiece; much of the savings is due to SpaceX's ability to reuse the first stage.
- As a result, if we can predict whether the first stage will land, we can estimate the cost of a launch.
- SpaceY will be able to make more educated bids for rocket launches versus SpaceX using the data science insights and models in this report.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Acquired historical launch data from Open Source REST API for SpaceX and from Wikipedia page ‘List of Falcon 9 and Falcon Heavy Launches’
- Perform data wrangling
  - Explored data to determine the label for training supervised models
  - Created a landing outcome training label from ‘Outcome’ column
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Methodology

Executive Summary

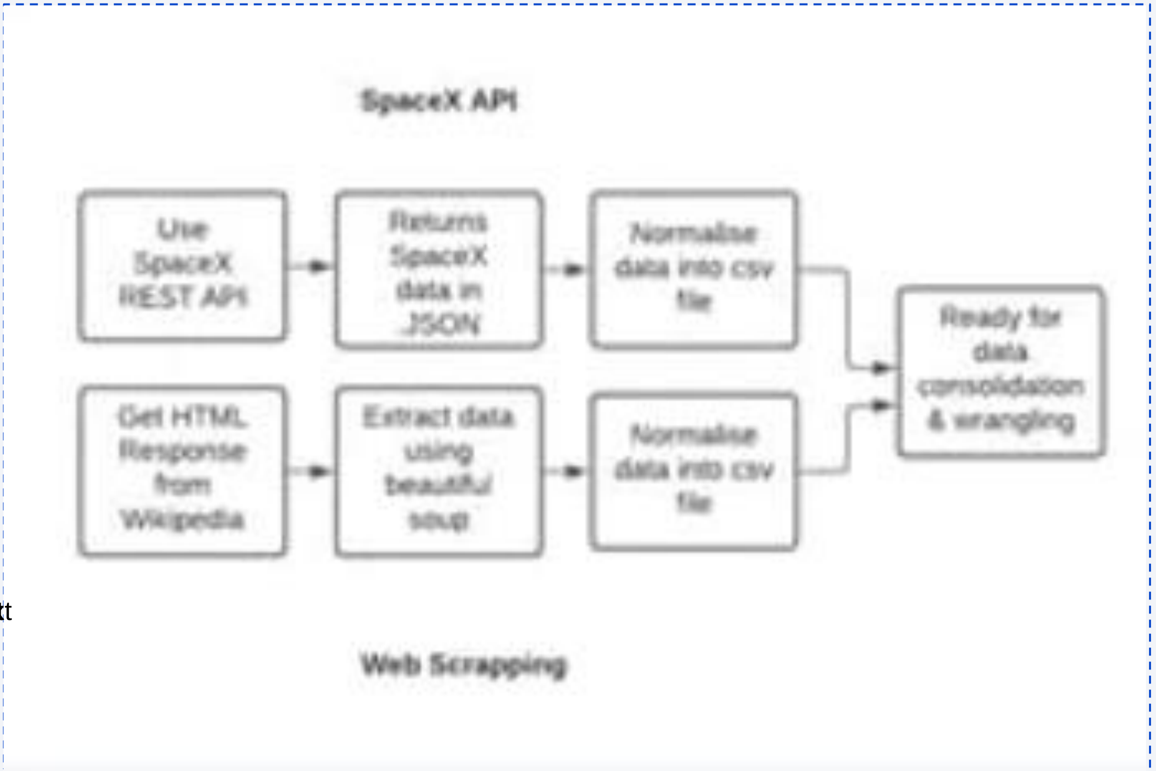
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
    - Created a column for training label ‘Class’
    - Standardized the data and split into training and test set
    - Fit to various model (Logistic Regression, Support Vector Machine, Decision Tree Classifier, K Nearest Neighbors Classifier)
    - Used a cross-validated grid-search to select the best ones for each model
    - Evaluated accuracy of each model

Data Collection – SpaceX API

- Requested and parsed the SpaceX launch data using the GET request
- Filtered the dataframe to only include Falcon 9 launches
- Replaced missing payload mass values from classified mission with mean
- Github: https://github.com/SAMGIT2023/JUPYTER

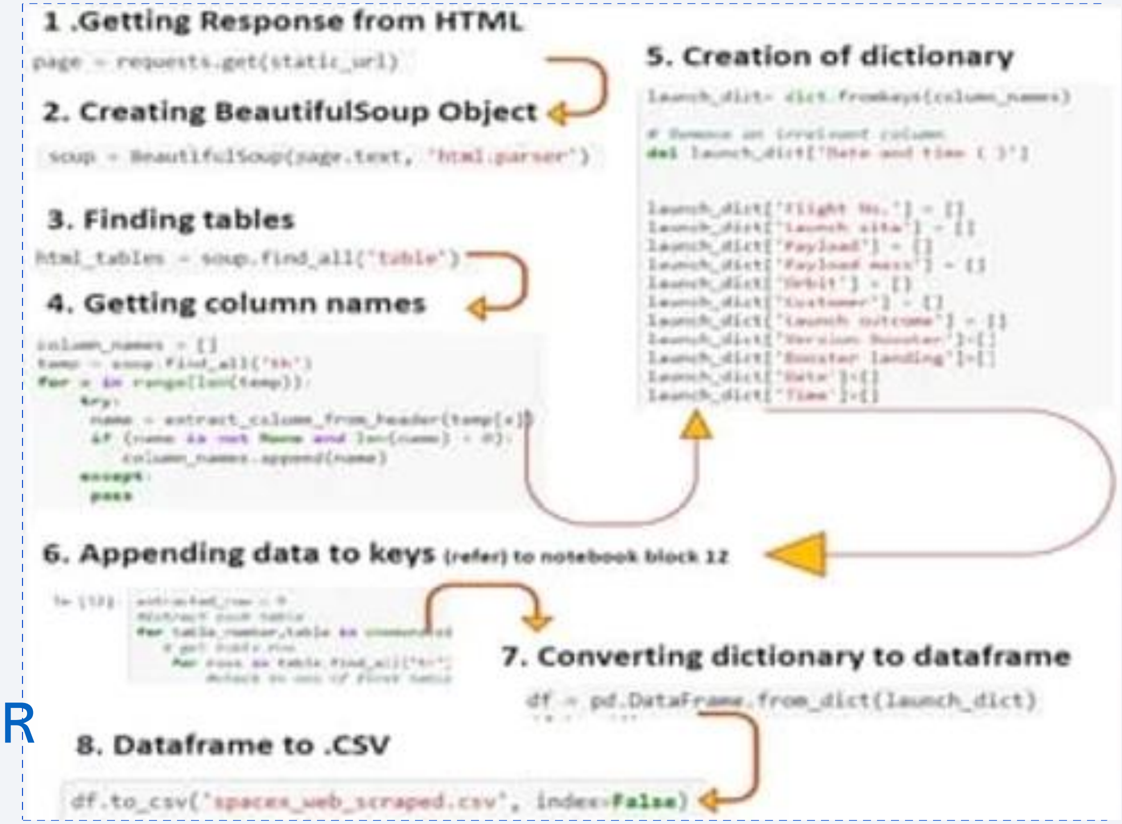
Type your text

Type your text



Data Collection - Scraping

- Acquired historical launch data from Wikipedia page ‘List of Falcon 9 and Falcon Heavy Launches’
- Extracted all column/variable names from the HTML table header
- Parsed the table and converted into Pandas Data Frame
- Github: https://github.com/SAMGIT2023/JUPYTER



Type your text

Type your text

Data Wrangling

- Calculated the number of launches on each site
- Calculated the number and occurrence of each orbit
- Calculated the number and occurrence of mission outcome per orbit type
- Created a landing outcome training label from ‘Outcome’ column
- Handle null values
- Github: https://github.com/SAMGIT2023/JUPYTER

Type your text

Type your text

EDA with Data Visualization

- Read the dataset into a Pandas dataframe
- Used Matplotlib and Seaborn visualization libraries to plot
  - Flight Number vs Payload Mass
  - Flight Number vs Launch Site
  - Payload vs Launch Site
  - Orbit type vs Success Rate
  - Flight Number vs Orbit Type
  - Payload vs Orbit Type
  - Year vs Success Rate
- Github: https://github.com/charuzziiiee/IBM-Data-Science/blob/master/Complete%20the%20EDA%20with%20Visualization%20lab.ipynb

Type your text

EDA with SQL

- Loaded data into an IBM DB2 instance
- Ran SQL queries to display and list information about
  - Launch sites
  - Payload masses
  - Booster versions
  - Mission outcomes
  - Booster landings
- Github: https://github.com/charuzziiiee/IBM-Data-Science/blob/master/Complete%20the%20EDA%20with%20SQL%20lab.ipynb

Build an Interactive Map with Folium

- Used Python Interactive mapping library called Folium
- Marked all launch sites on a map
- Marked the successful/ failed launches for each site on map
- Calculated the distances between a launch site to its proximities
- Github: https://github.com/charuzziiiee/IBM-Data-Science/blob/master/Interactive%20Visual%20Analytics%20with%20Folium%20lab.ipynb

Build a Dashboard with Plotly Dash

- Used Python interactive dashboarding library called Plotly Dash to enable stakeholders to explore and manipulate data in an interactive and real-time way
- Pie chart showing success rate
- Scatter chart showing payload mass vs. landing outcome
- Drop-down menu to choose between all sites and individual launch sites
- Github: https://github.com/charuzziiiee/IBM-Data-Science/blob/master/spacex_dash_app.py

Predictive Analysis (Classification)

- Standardized the data
- Split the data into training data and test data
- Fit the training data to various models (Logistic Regression, Support Vector Machine, Decision Tree Classifier, K Nearest Neighbors Classifier)
- Used cross-validated grid-search to select the best model
- Evaluated accuracy of each model using test set to select the best model
- Github: https://github.com/charuzziiiee/IBM-Data-Science/blob/master/Machine%20Learning%20Prediction%20lab.ipynb

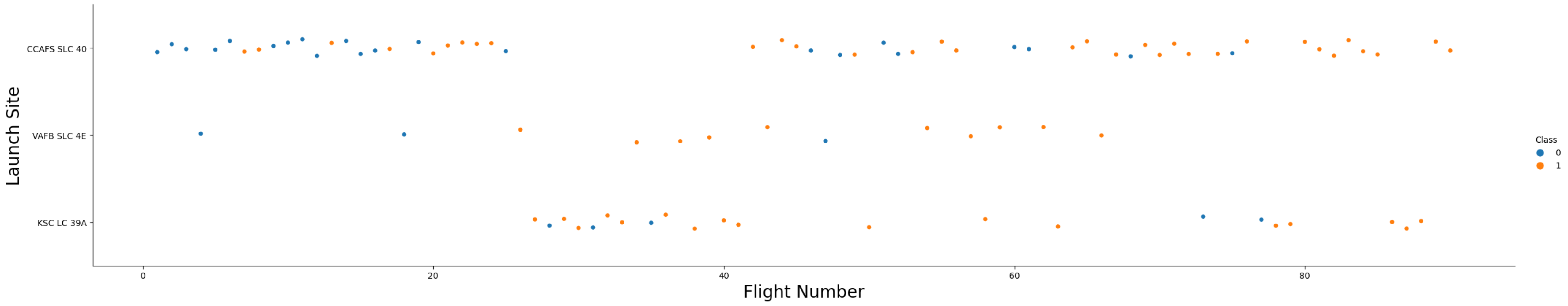
Results

- The SVM, KNN, and Logistic Regression models are the best in terms of prediction accuracy for this dataset.
- Low-weighted payloads perform better than heavier payloads.
- The success rates for SpaceX launches are directly proportional to time and they will eventually perfect the launches.
- KSC LC 39A had the most successful launches from all the sites.
- Orbit GEO, HEO, SSO, ES L1 has the best success rate.

Section 2

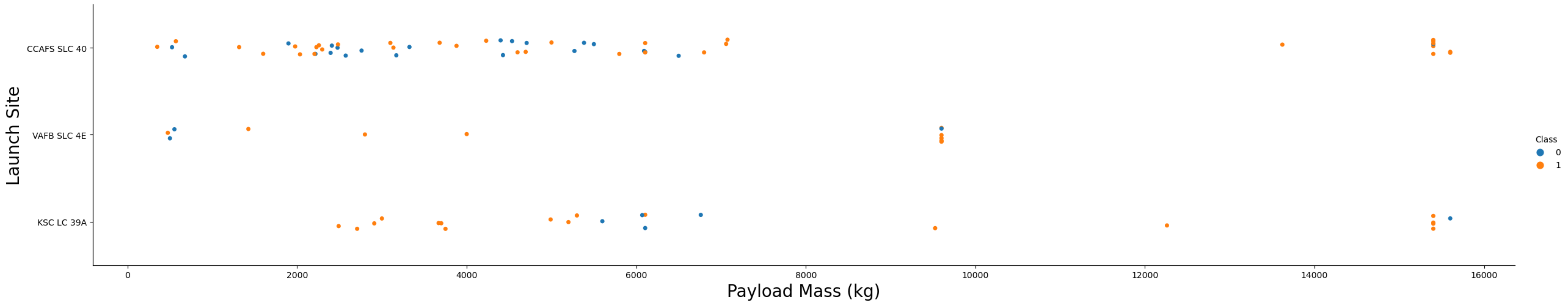
Insights drawn from EDA

Flight Number vs. Launch Site



- Launches from the site of CCAFS SLC 40 are significantly higher than launches from other sites.

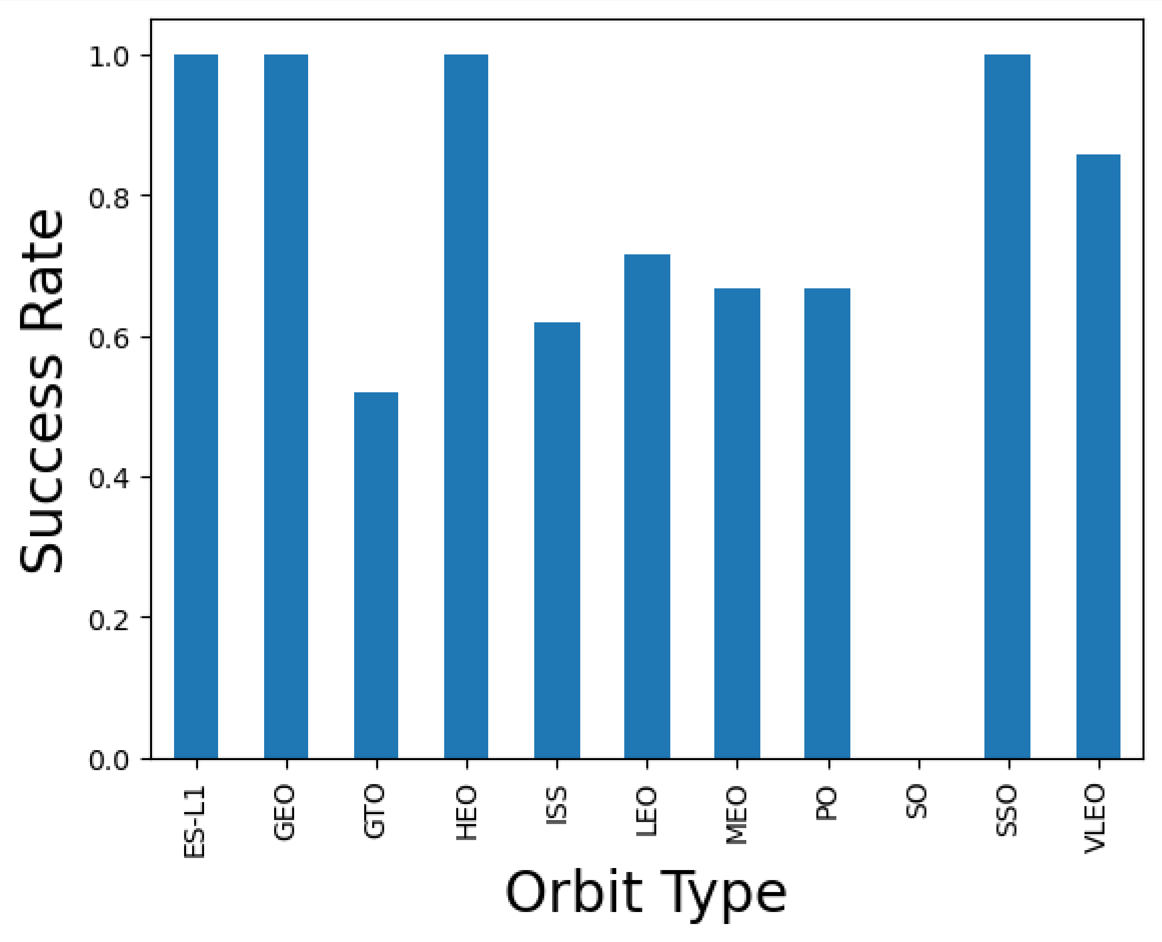
Payload vs. Launch Site



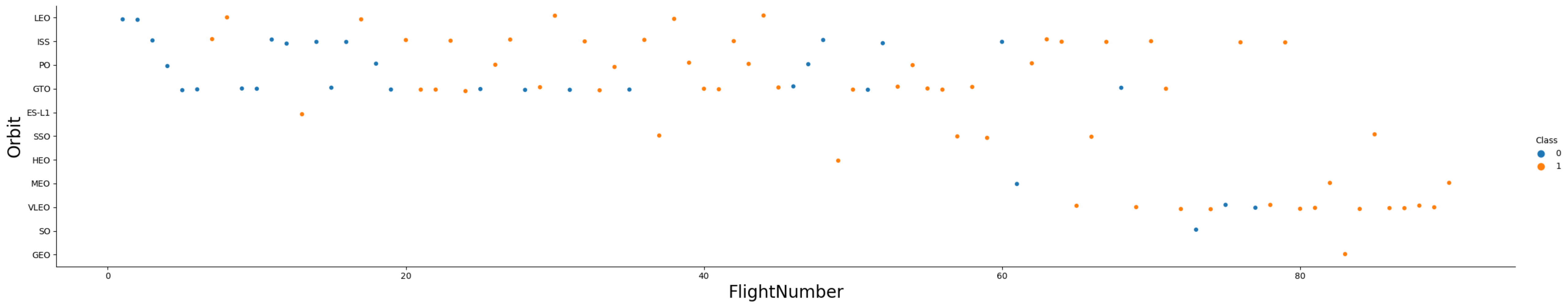
- Majority of Payloads with lower mass have been launched from CCAFS SLC 40.

Success Rate vs. Orbit Type

- The orbit types of ES-L1, GEO, HEO, SSO are among the highest success rate.

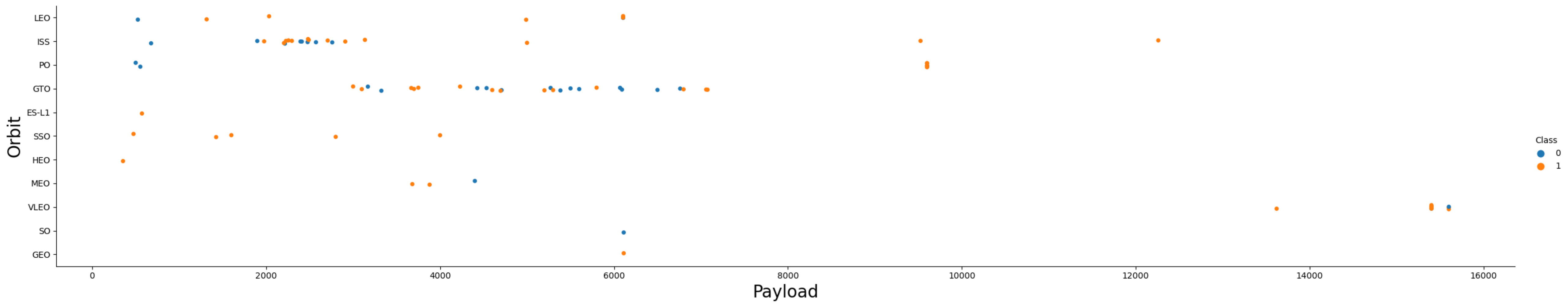


Flight Number vs. Orbit Type



- A trend can be observed of shifting to VLEO launches in recent years.

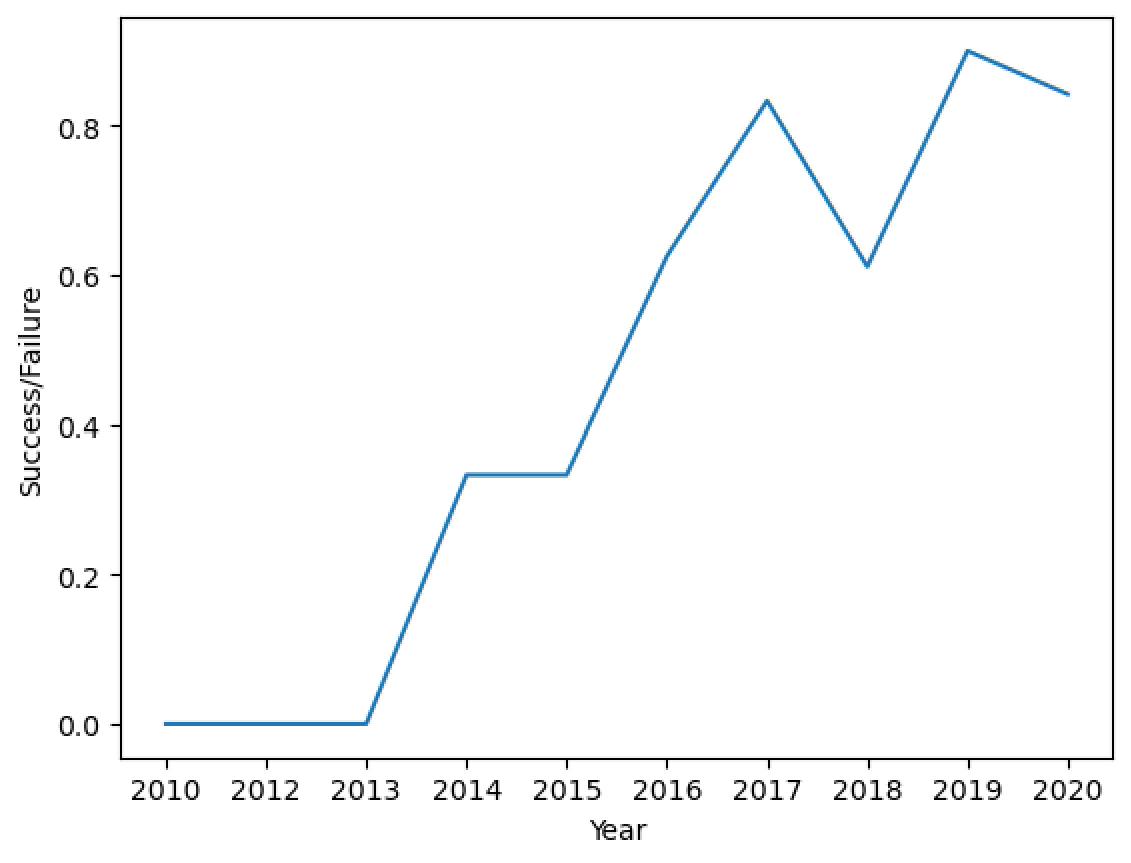
Payload vs. Orbit Type



- There is an association between ISS and Payload at the range around 2000, as well as between GTO and the range of 4000-8000.

Launch Success Yearly Trend

- Launch success rate has increased significantly since 2013 and has been established since 2019, potentially due to technological advancements and lessons learned.



All Launch Site Names

All Launch sites of the Space X program.

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

Launch Site Names Begin with 'CCA'

- 5 Launch Site Names that begin with ‘CCA’

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

Total Payload Mass

```
%%sql
SELECT SUM(PAYLOAD_MASS__KG_)
FROM SPACEXTBL
WHERE Customer = 'NASA (CRS)';
```

```
 * ibm_db_sa://thk90242:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/BLUDB
Done.
```

1
45596

- The total payload carried by boosters from NASA

Average Payload Mass by F9 v1.1

```
%%sql
SELECT AVG(PAYLOAD_MASS__KG_)
FROM SPACEXTBL
WHERE Booster_Version LIKE 'F9 v1.0%';
```

```
 * ibm_db_sa://thk90242:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/BLUDB
Done.
```

1
340

- The average payload mass carried by booster version F9 v1.1

First Successful Ground Landing Date

```
%%sql
SELECT MIN(Date)
FROM SPACEXTBL
WHERE Landing__Outcome = 'Success (ground pad)';
```

```
 * ibm_db_sa://thk90242:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/BLUDB
Done.
```

1
2015-12-22

- The first successful landing outcome on a ground pad

Successful Drone Ship Landing with Payload between 4000 and 6000

- The names of boosters that have successfully landed on a drone ship and had payload mass greater than 4000 but less than 6000

booster_version
F9 FT B1021.1
F9 FT B1023.1
F9 FT B1029.2
F9 FT B1038.1
F9 B4 B1042.1
F9 B4 B1045.1
F9 B5 B1046.1

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes

mission_outcome	total_number
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- The names of the booster which have carried the maximum payload mass

booster_version
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
F9 B5 B1051.3
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

2015 Launch Records

- The failed landing outcomes in drone ships, their booster versions, and launch site names for in the year 2015

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

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

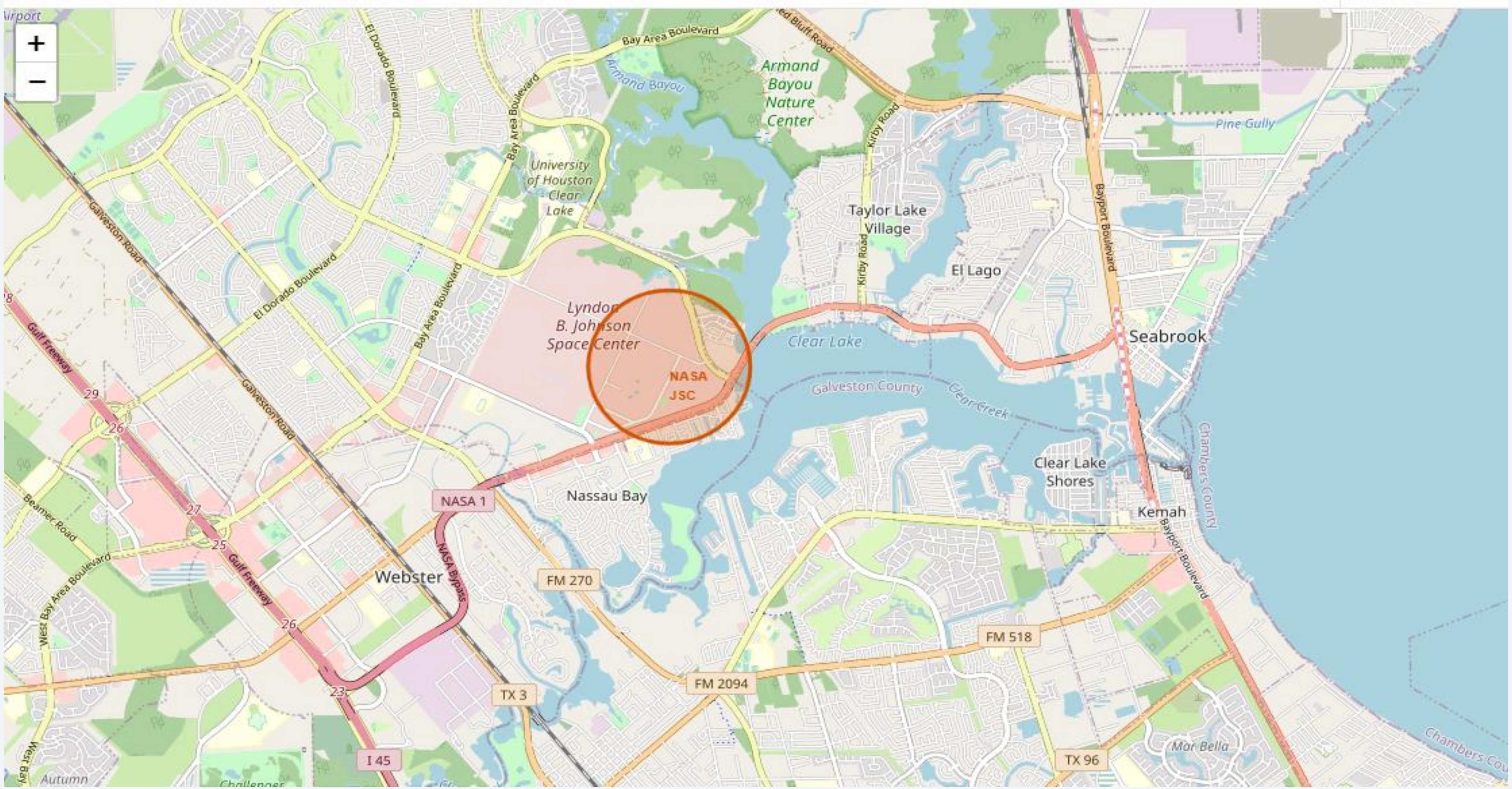
- The count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the dates 2010-06-04 and 2017-03-20, in descending order

landing__outcome	total_number
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

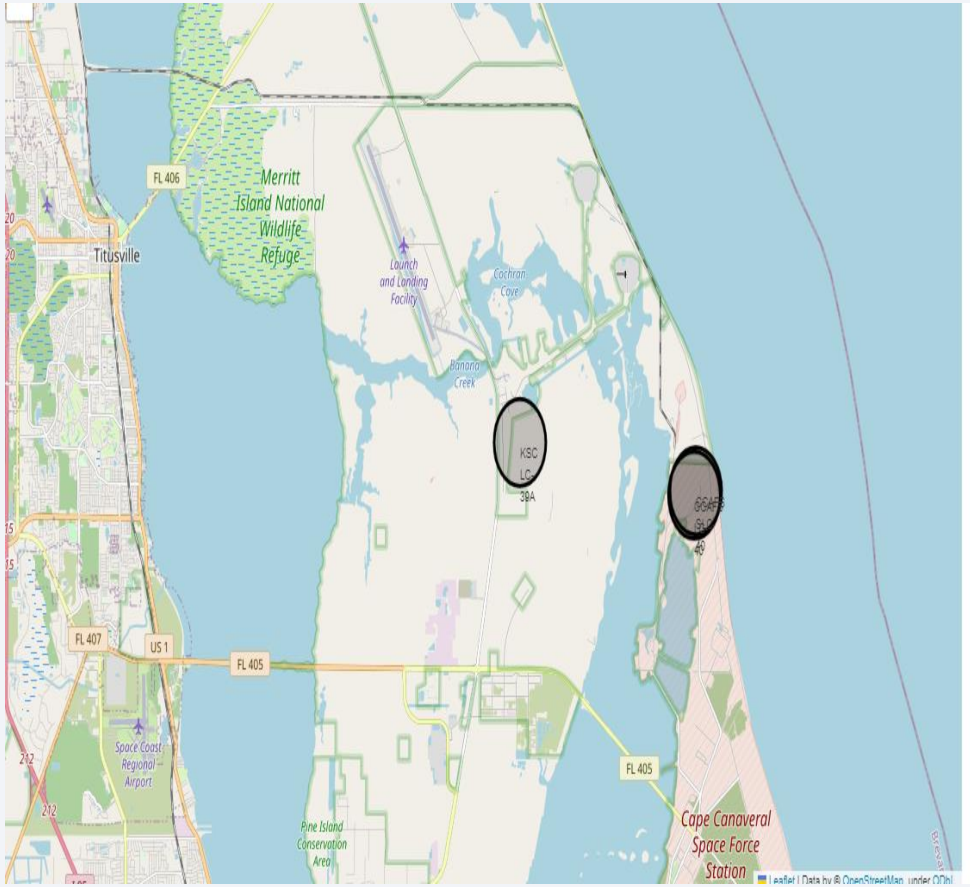
Section 3

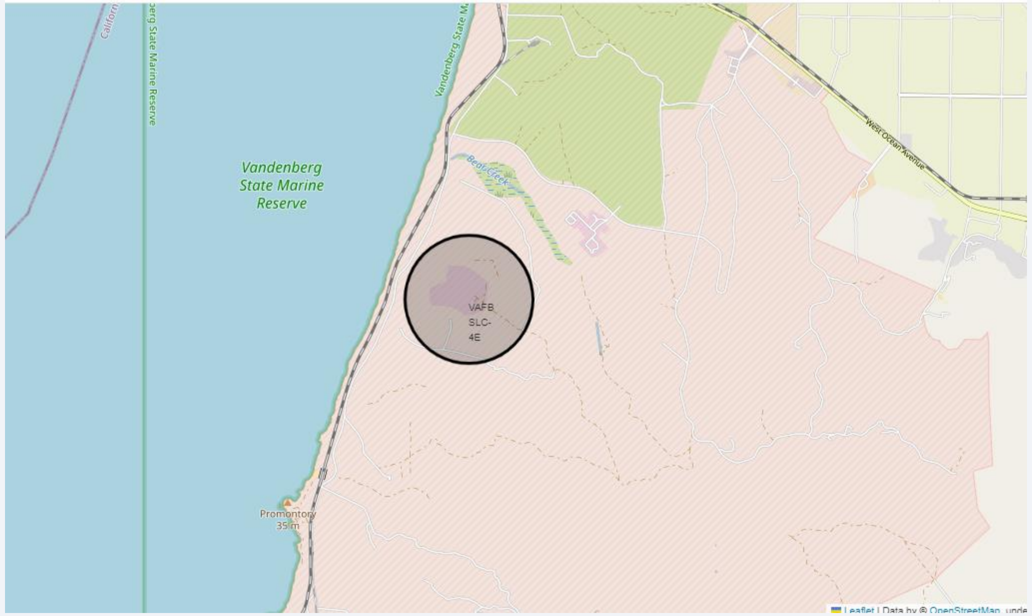
Launch Sites Proximities Analysis

All Launch Sites Marked on a Map

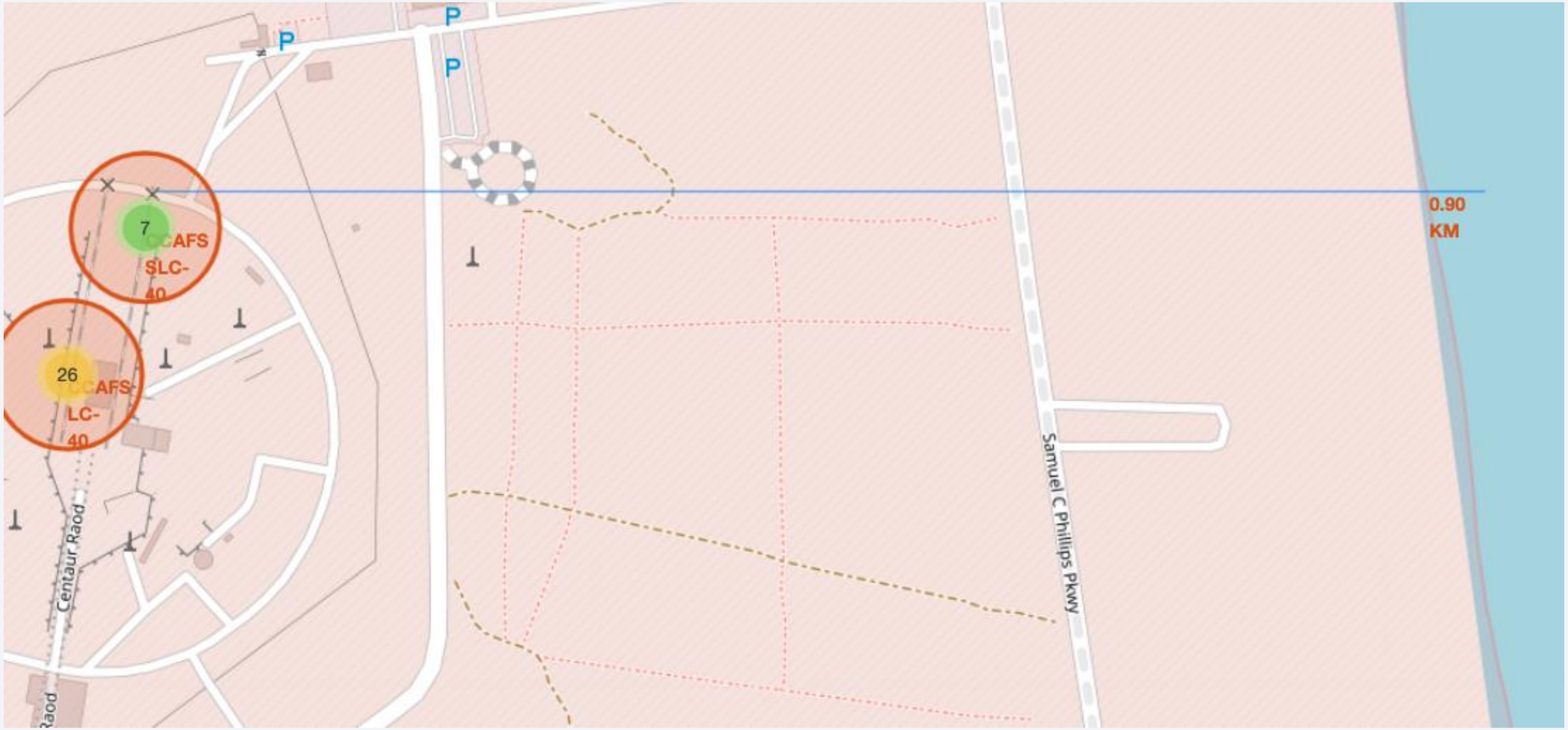


Success/Failed Launches on a Map





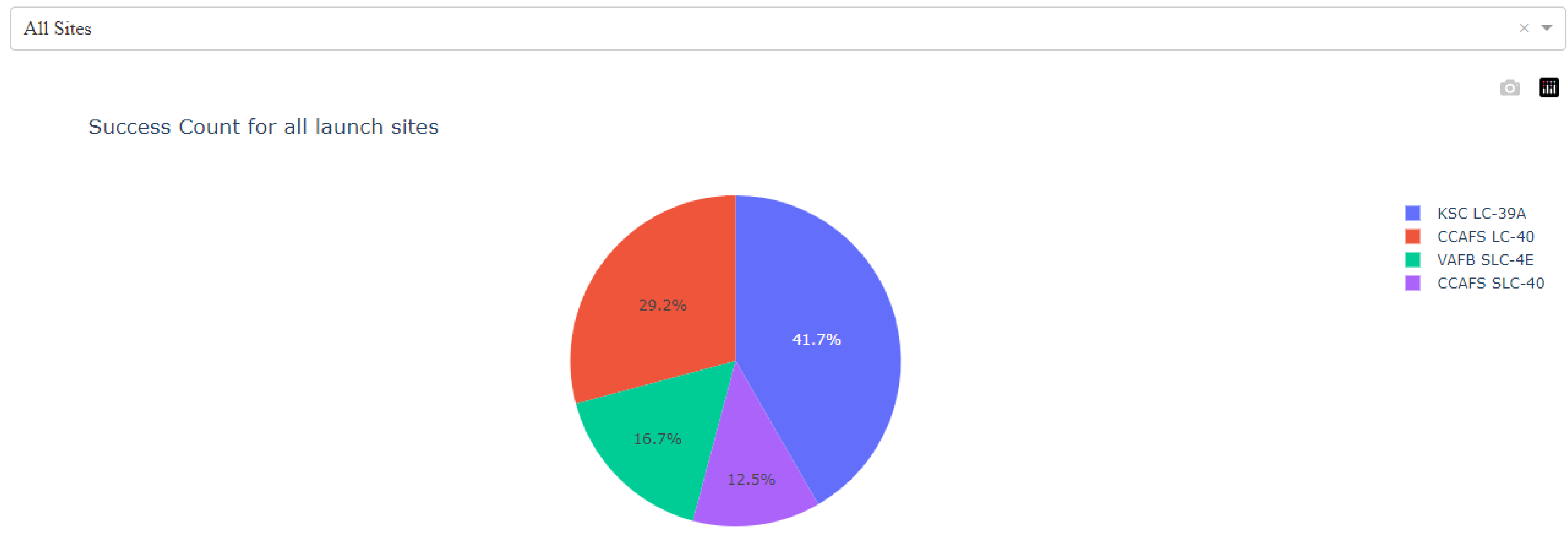
Distance to Coast Line



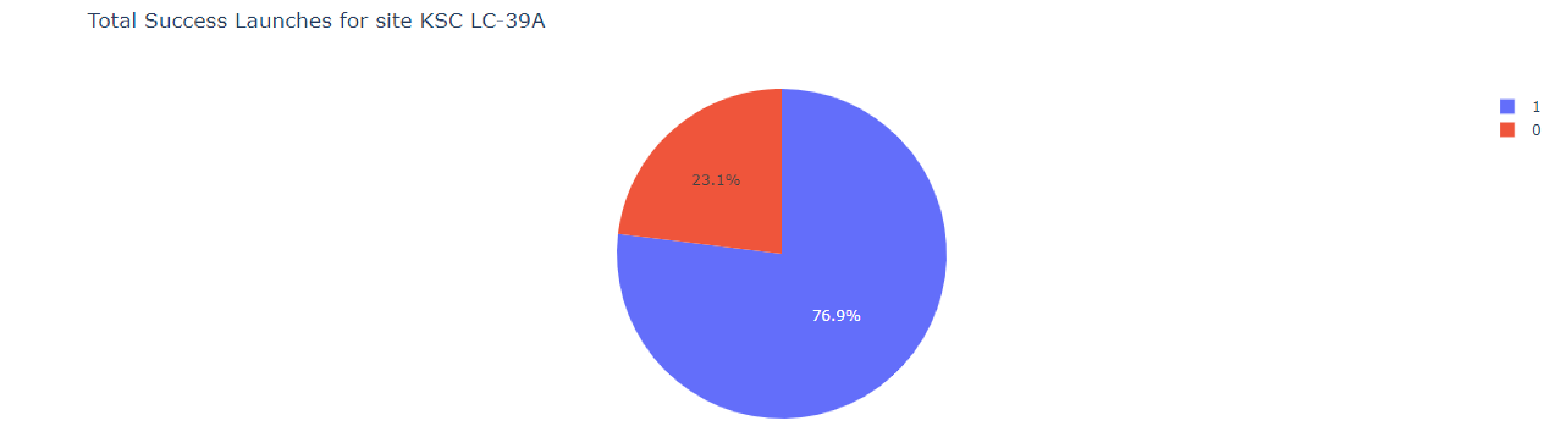
Section 4

Build a Dashboard with Plotly Dash

Success Count for All Launch Sites

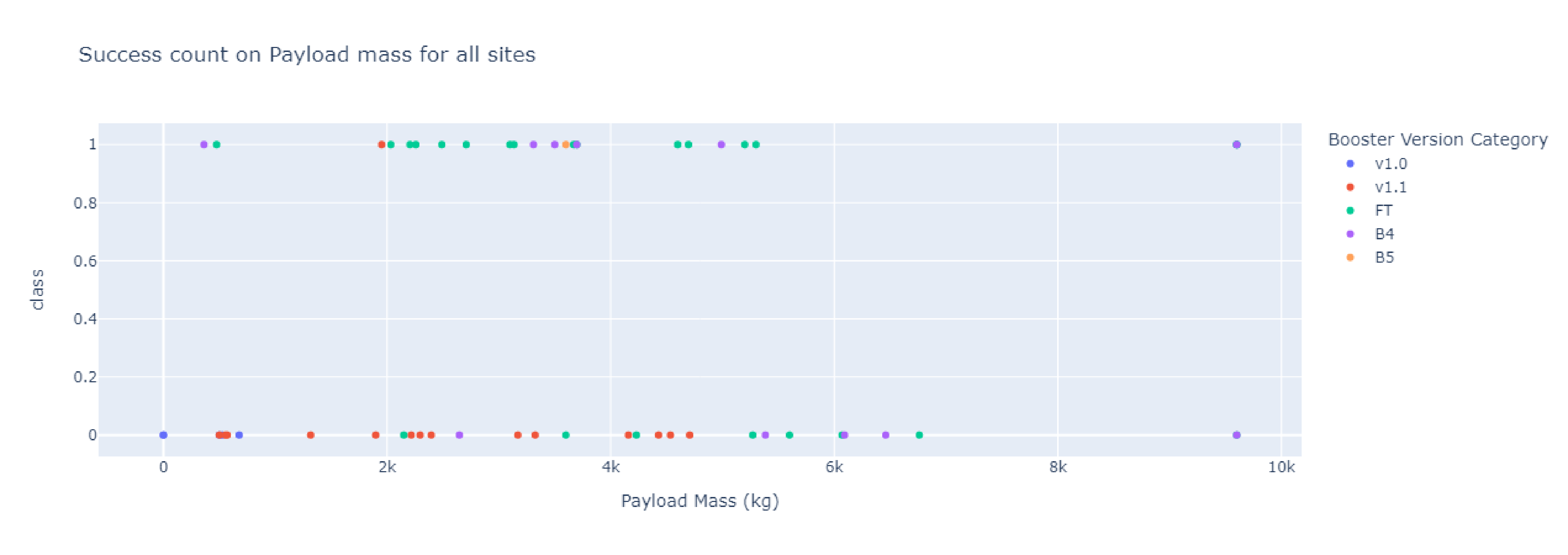


Success Rate by Site



- KSC LC-39A achieved a 76.9% success rate while getting a 23.1% failure rate

Success count on Payload Mass for All Sites

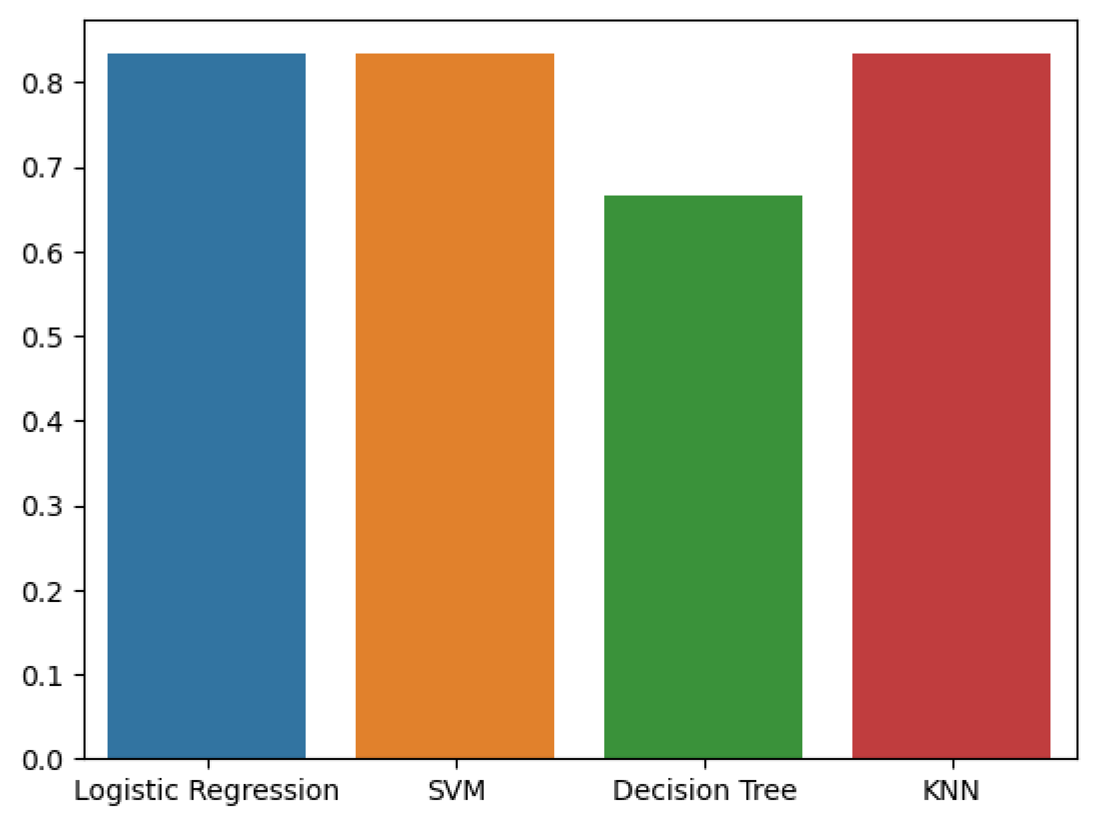


Section 5

Predictive Analysis (Classification)

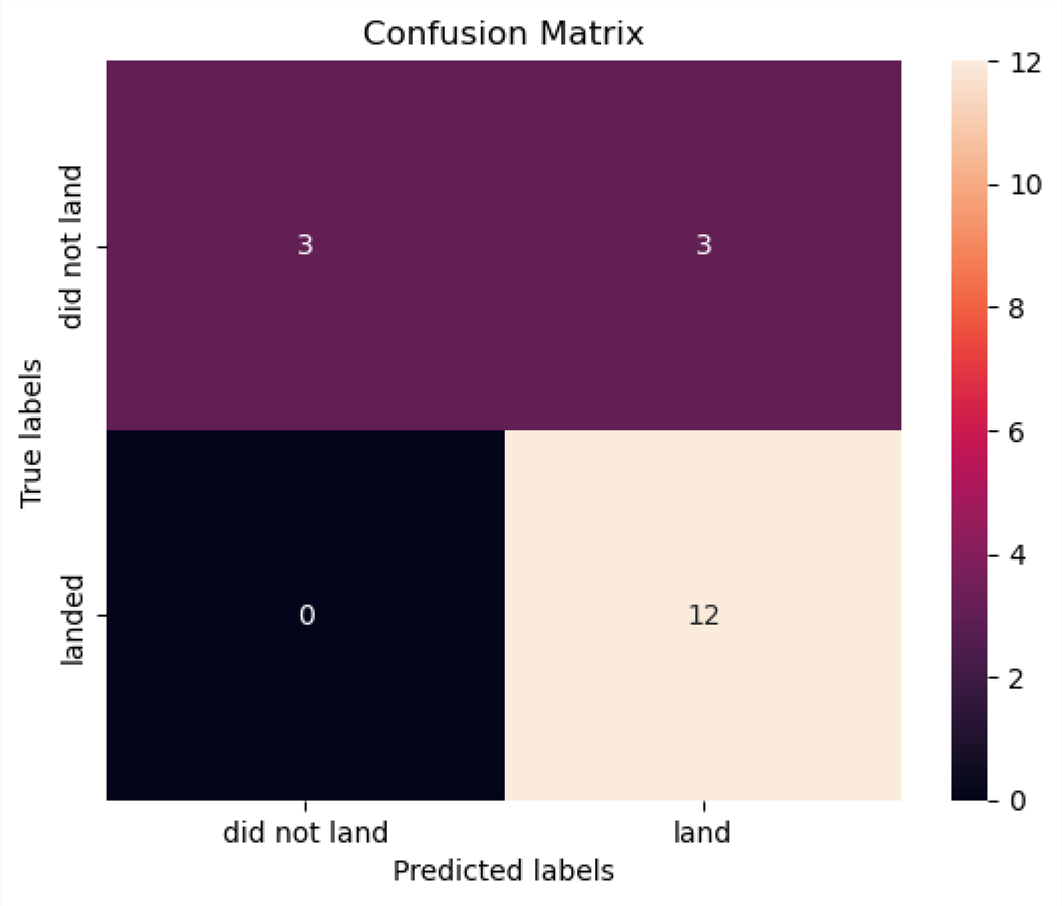
Classification Accuracy

- Logistic Regression, SVM, and KNN have the same accuracy of 83.33% on the test set. While Decision Tree only has 66.67% accuracy on the test set.



Confusion Matrix

- The confusion matrices of the best-performing models are the same.
- The major problem is false positives as evidence by the models incorrectly predicting the 1st stage booster to land in 3 out of 18 samples in the test set.



Conclusions

- The SVM, KNN, and Logistic Regression models are the best in terms of prediction accuracy for this dataset.
- Low-weighted payloads perform better than heavier payloads.
- The success rates for SpaceX launches are directly proportional to time.
- KSC LC 39A had the most successful launches from all the sties.
- Orbit GEO, HEO, SSO, ES L1 has the best success rate.

Appendix

- https://github.com/charuzziiiee/IBM-Data-Science/blob/main/Collecting%20Data.ipynb
- https://github.com/charuzziiiee/IBM-Data-Science/blob/main/Data%20Collection%20with%20Web%20Scraping%20lab.ipynb
- https://github.com/charuzziiiee/IBM-Data-Science/blob/main/Data%20wrangling.ipynb
- https://github.com/charuzziiiee/IBM-Data-Science/blob/master/Complete%20the%20EDA%20with%20Visualization%20lab.ipynb
- https://github.com/charuzziiiee/IBM-Data-Science/blob/master/Complete%20the%20EDA%20with%20SQL%20lab.ipynb
- https://github.com/charuzziiiee/IBM-Data-Science/blob/master/Interactive%20Visual%20Analytics%20with%20Folium%20lab.ipynb
- https://github.com/charuzziiiee/IBM-Data-Science/blob/master/spacex_dash_app.py
- https://github.com/charuzziiiee/IBM-Data-Science/blob/master/Machine%20Learning%20Prediction%20lab.ipynb

Thank you!