IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Name>
<Date>

Outline

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

Executive Summary

- **Summary of methodologies**

In the current age of rapid technological developments space travel is becoming more and more affordable stimulating demand on commercial travel. SpaceX is one of the most successful companies in this area. The affordability of SpaceX Falcon 9 rocket launches mostly comes from the idea of reusing the first stage. For the competing company SpaceY its important to analyze SpaceX price of each launch. In this work we present Space X information gathering and dashboard creating for our team. We also demonstrate training of a machine learning model and use public information to predict if SpaceX will reuse the first stage, instead of conventional rocket science technical analysis.

- **Summary of all results**

Here we analyze successful Falcon 9 landing on sea and ground. The success rate was determined ~67%. For affordability it's important to analyze landings success. Landing successes depend on multiple factors, like booster versions, orbit types, payload mass, location. Yearly trend for increasing success rate from 2013 to 2020 was observed. Minor dependance of the successful landing from payload and launch site found. We also analyzed success rates for each orbit type. Specifics for success rate dependence on the launch site is also presented. Applicability of Several Machine Learning algorithms were tested, best hyperparameters were chosen.

Introduction

Project Background and context

SpaceX is one of the leading companies in affordable space flights. One reason SpaceX can do this is the rocket launches are relatively inexpensive. SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upwards of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage. Therefore, if we can determine if the first stage will land, we can determine the cost of a launch.

Problems you want to find answers

In this capstone, we will predict if the Falcon 9 first stage will land successfully. SpaceX 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 SpaceX 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. In this lab, you will collect and make sure the data is in the correct format from an API. The following is an example of a successful and launch..

Section 1

Methodology

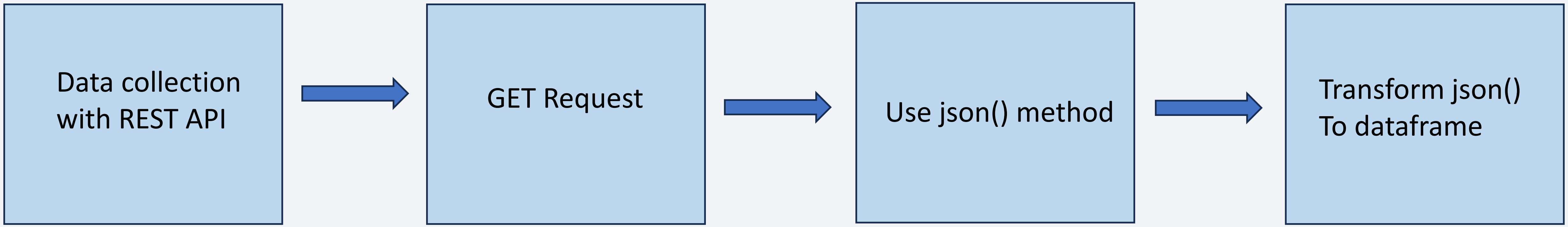
Methodology

Executive Summary

- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- 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

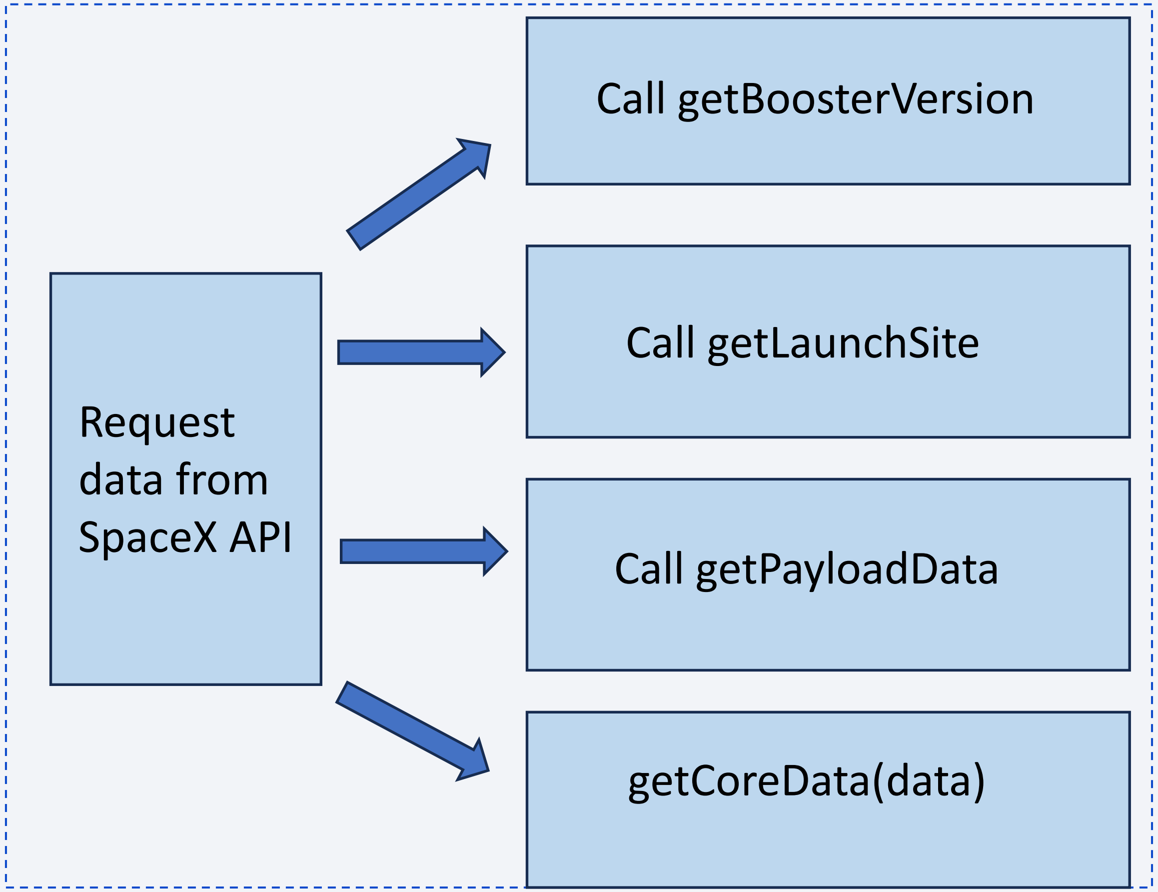
Data Collection

https://github.com/TeyaTopuria/applied-data-science-capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb



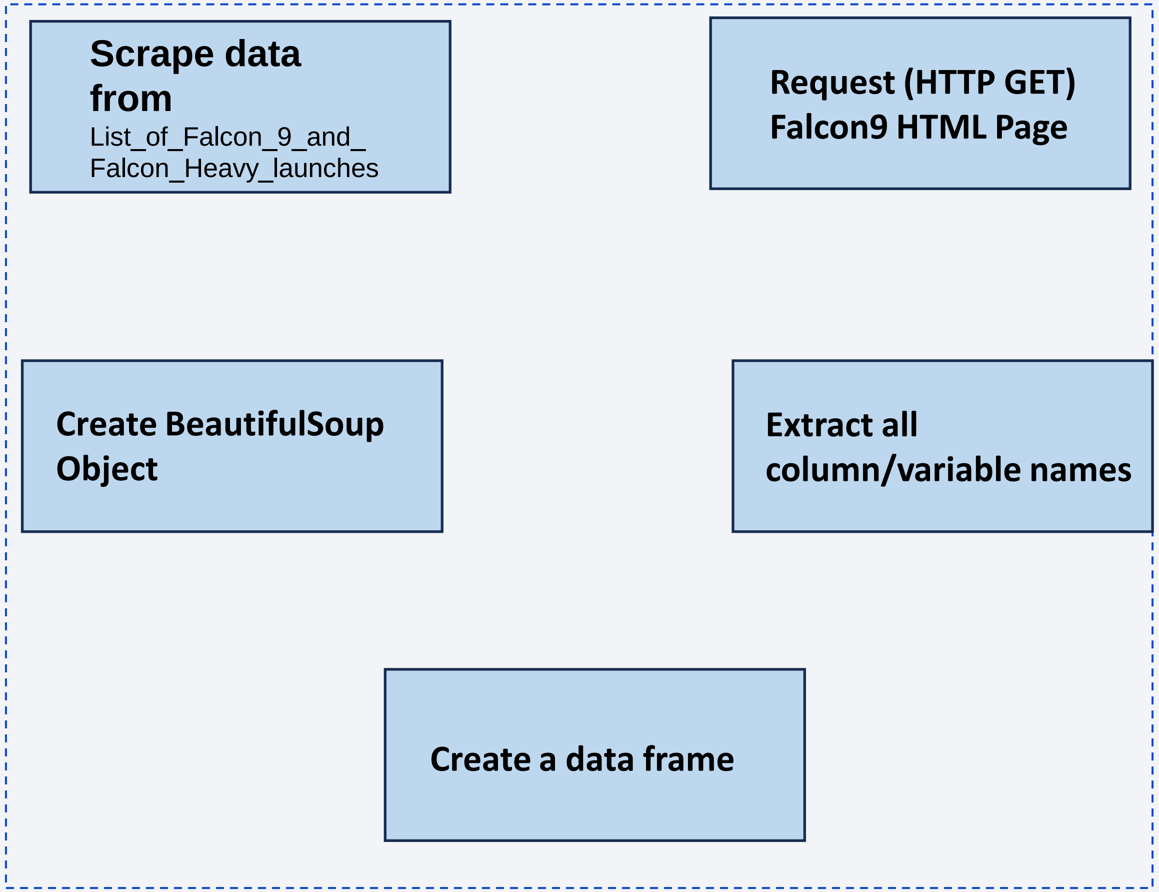
Data Collection – SpaceX API

- https://github.com/TeyaTopuria/applied-data-science-capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

- https://github.com/TeyaTopuria/applied-data-science-capstone/blob/main/jupyter-labs-webscraping.ipynb



Data Wrangling

Load Space X dataset → Identify percentage of the missing values and column types → Create a landing outcome label → Determine the success rate

https://github.com/TeyaTopuria/applied-data-science-capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

Part of Exploratory Data Analysis was performed using visualization

Scatter Plot for Flight Number/Launch Site Dependency visualization

Bar Chart for success rate of each orbit type

Scatter Plot for Payload Mass/Orbit type Dependency visualization

Scatter Plot for Payload Mass/Launch Site Dependency visualization

Scatter Plot for Flight Number /Orbit type Dependency visualization

Line Plot for Visualize the launch success yearly trend

https://github.com/TeyaTopuria/applied-data-science-capstone/blob/main/edadataviz.ipynb

EDA with SQL

- Display the names of the unique launch sites in the space mission
- Display 5 records where launch sites begin with the string 'CCA'
- Display the total payload mass carried by boosters launched by NASA (CRS)
- Display average payload mass carried by booster version F9 v1.1
- List the date when the first succesful landing outcome in ground pad was acheived.
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- List the total number of successful and failure mission outcomes
- List the names of the booster_versions which have carried the maximum payload mass.
- List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.
- 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.

https://github.com/TeyaTopuria/applied-data-science-capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

Mark all launch Sites on the map **(Circle)**

→

Mark success/fail Launches for each site **(Marker)**

→

Calculate distances on the map **(Mouse Position)**

→

Draw a line between Two points **(PolyLine)**

https://github.com/TeyaTopuria/applied-data-science-capstone/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

Drop Down List

This Option Enables Launch Site Selection

Pie Chart

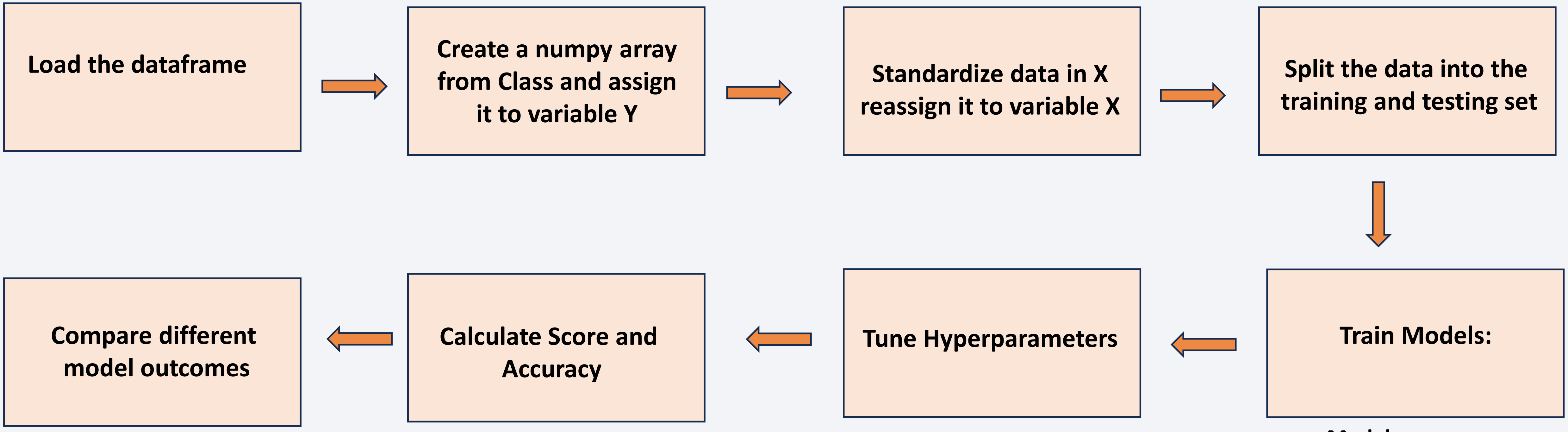
This Option Enables Comparison for Total Successful Launches For Each Site

Scatter Plot

This Option Enables Correlation for Payload And Launch Success

https://github.com/TeyaTopuria/applied-data-science-capstone/blob/main/spacex_dash_app(1)working.py

Predictive Analysis (Classification)



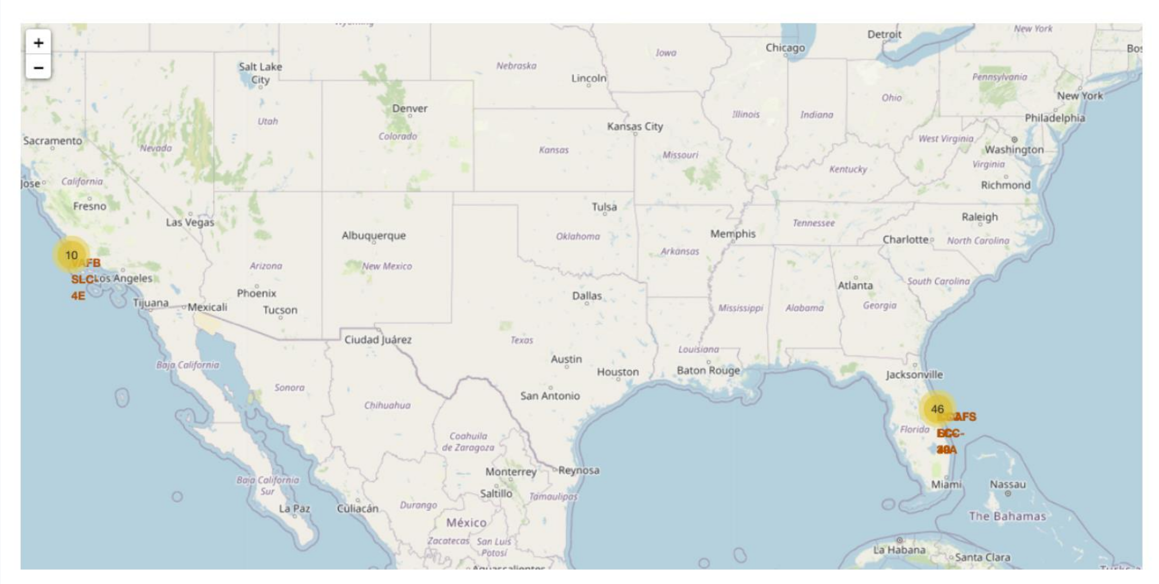
Models:
Logistic Regression
SVM
Decision Tree
KNN_CV

https://github.com/TeyaTopuria/applied-data-science-capstone/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results

Exploratory data analysis results

The success rate was determined ~67%.



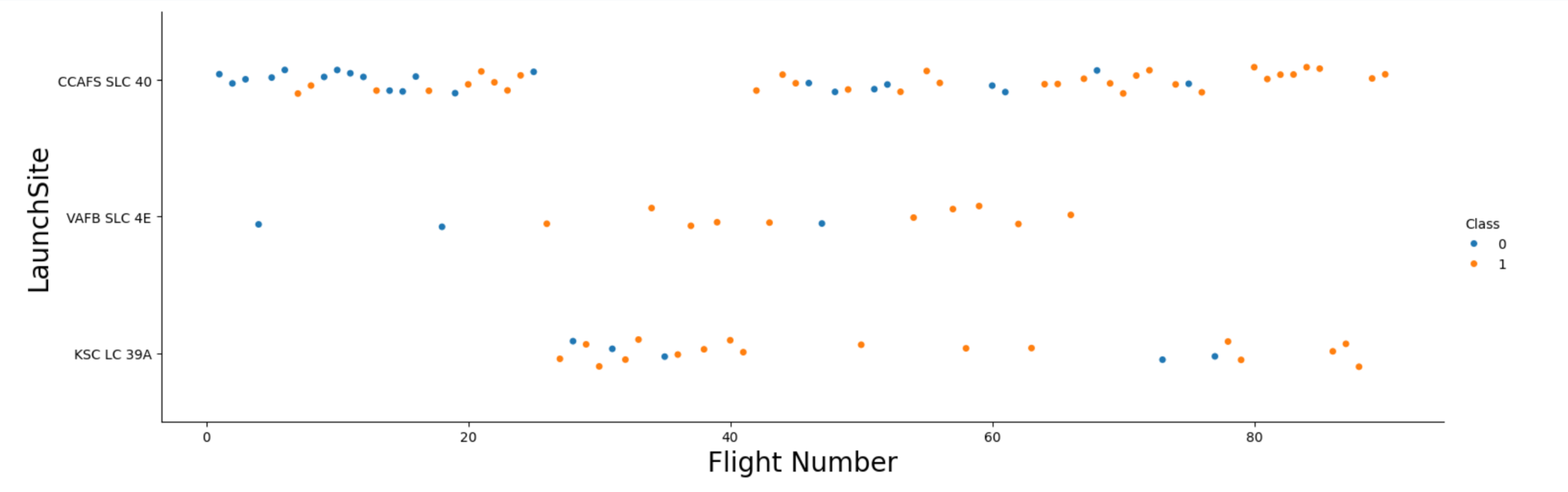
Predictive analysis results

Accuracy score was ~**67%.**
Major confusion matrix problem was **False Positive**

Section 2

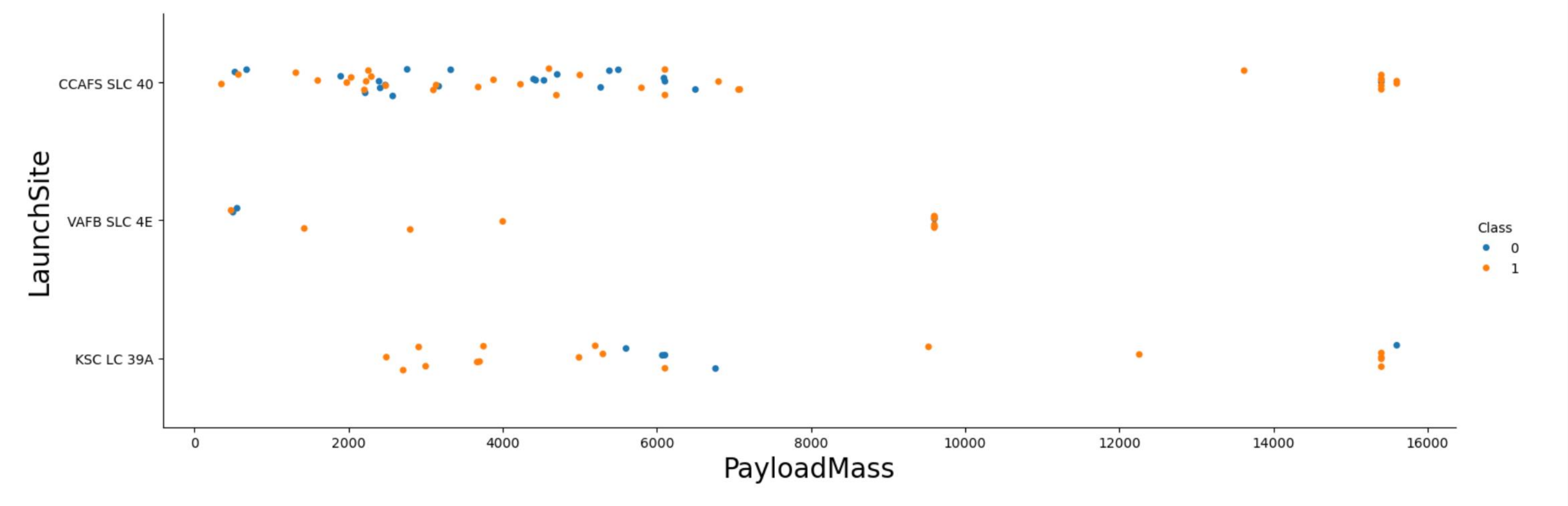
Insights drawn from EDA

Flight Number vs. Launch Site



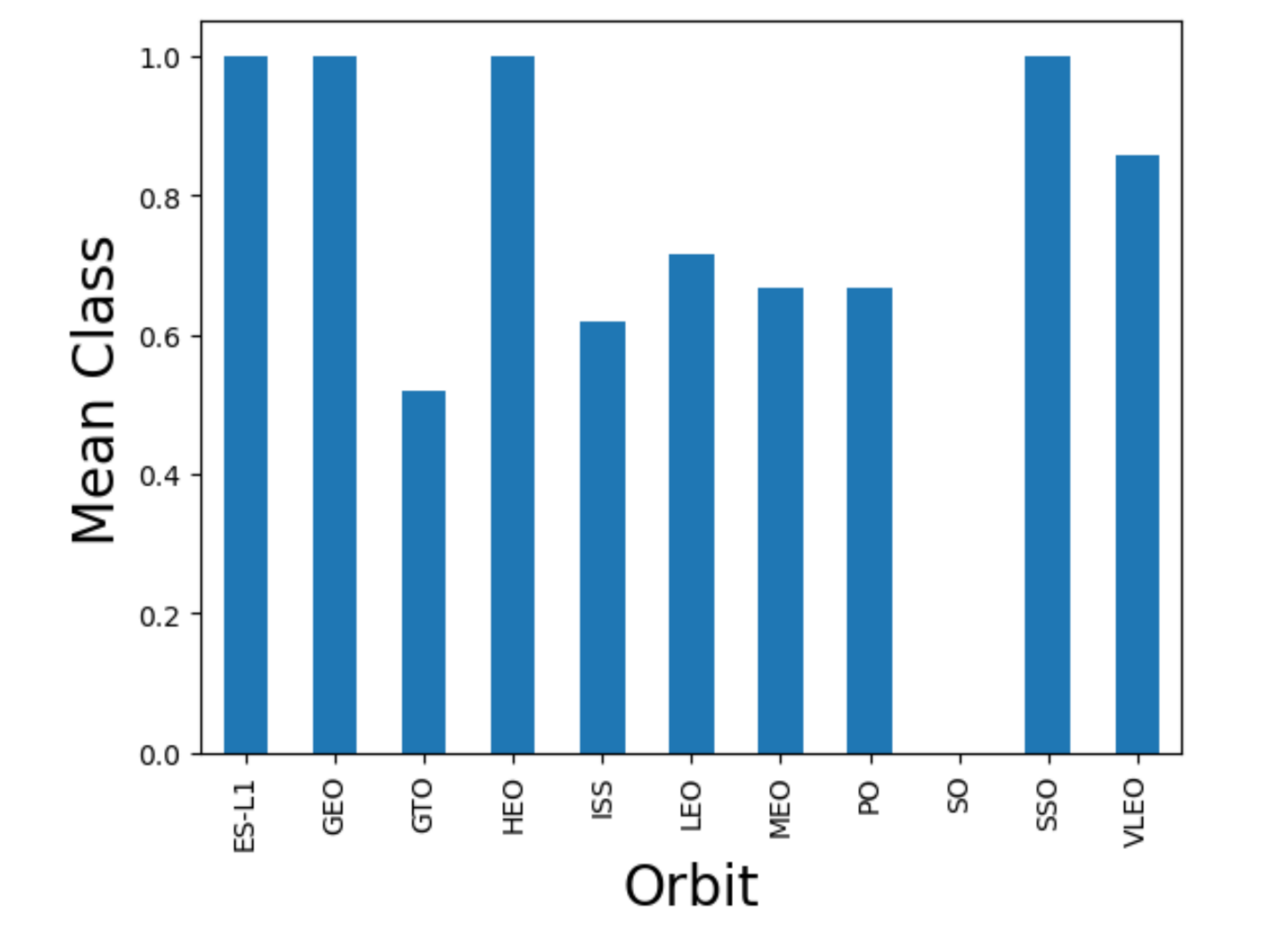
High flight number appear to relate to higher success rates

Payload vs. Launch Site



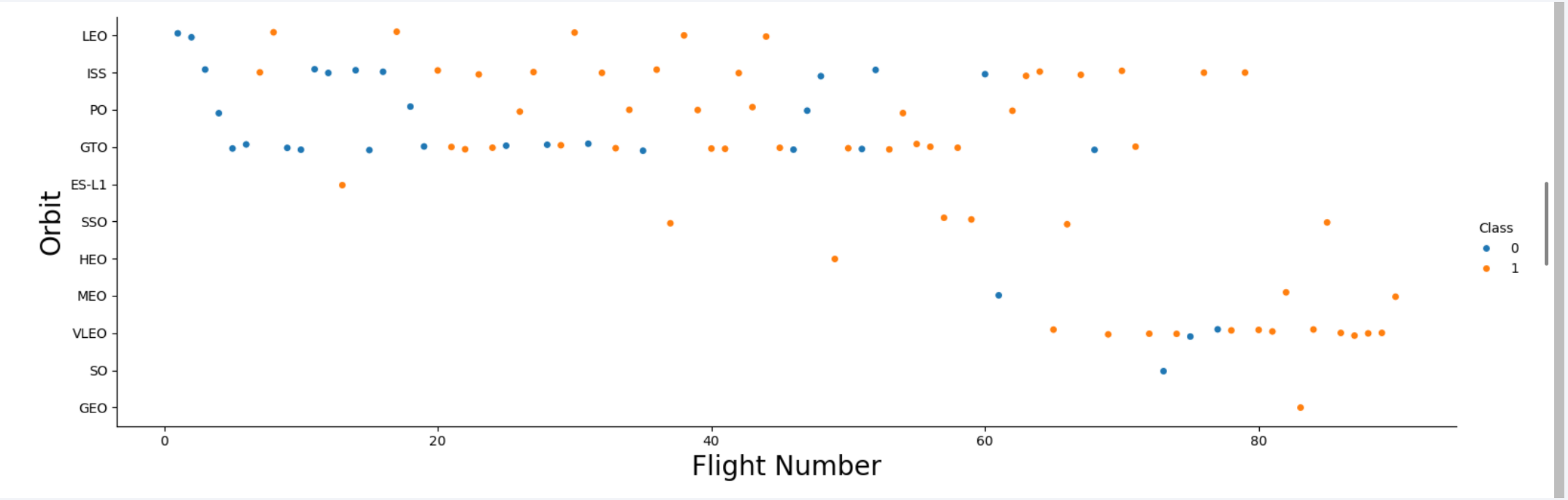
VAFB SLC launch site had no payloads heavier than 10000

Success Rate vs. Orbit Type



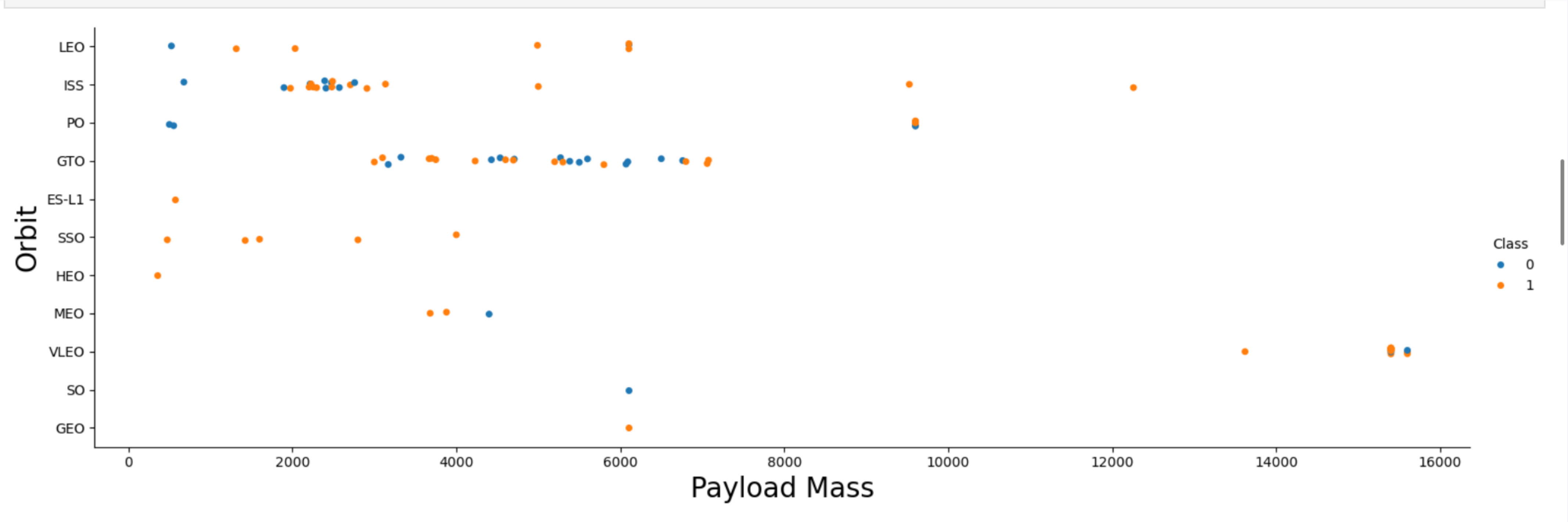
ES-L1, GFO HEO and SSO orbits have highest success rates.

Flight Number vs. Orbit Type



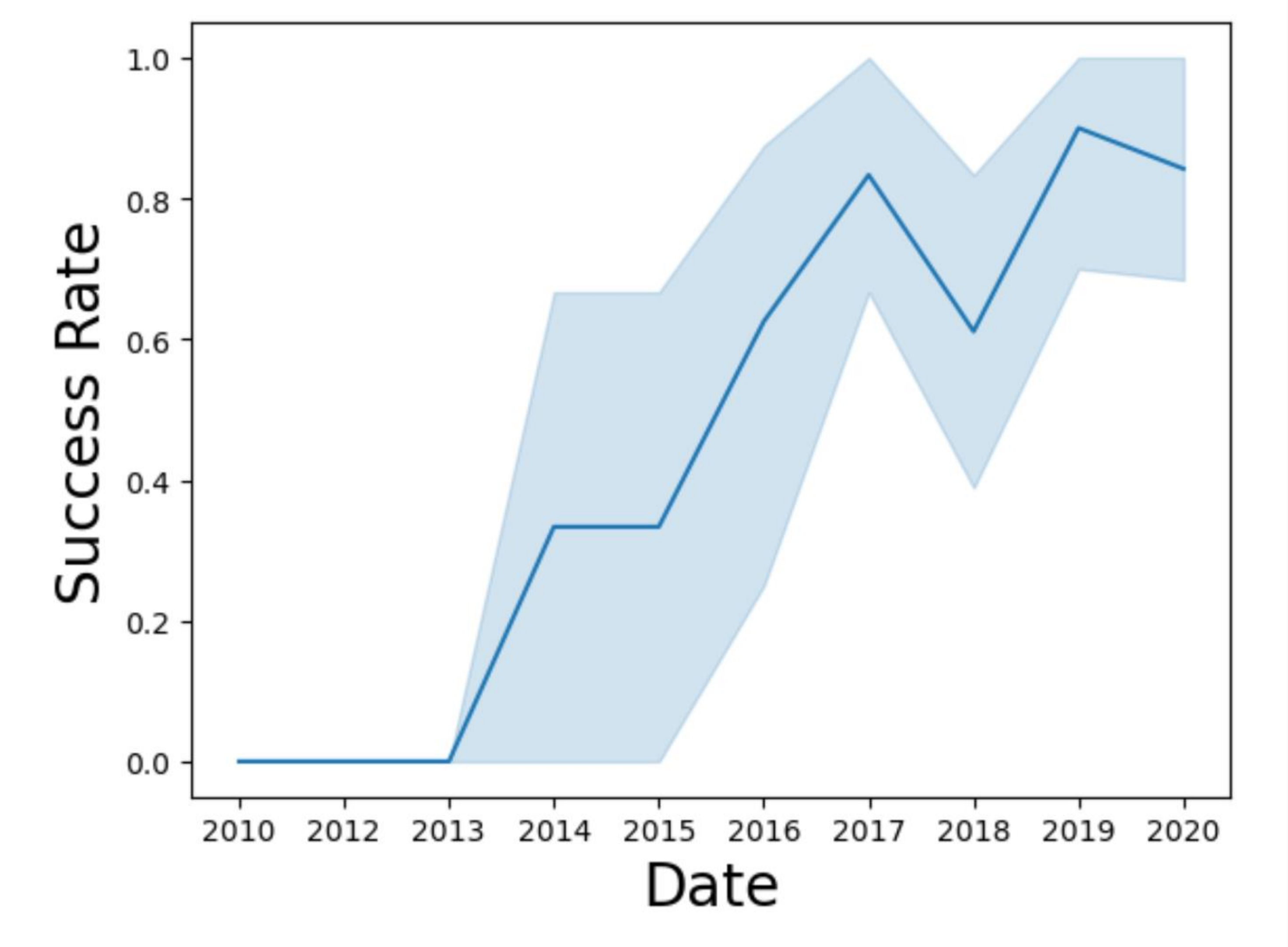
You can observe that in the LEO orbit, success seems to be related to the number of flights. Conversely, in the GTO orbit, there appears to be no relationship between flight number and success.

Payload vs. Orbit Type



With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.
However, for GTO, it's difficult to distinguish between successful and unsuccessful landings as both outcomes are present.

Launch Success Yearly Trend



The sucess rate since 2013 kept increasing till 2020

All Launch Site Names

40]:

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

Following statement was used for this query

sql SELECT DISTINCT "Launch_Site" from SPACEXTABLE;

Launch Site Names Begin with 'CCA'

]:

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

Following statement was used for this query

%sql SELECT * from SPACEXTABLE where "Launch_Site" like 'CCA%' LIMIT 5;

Total Payload Mass

Display the total payload mass carried by boosters launched by NASA (CRS)

```
%sql SELECT sum(PAYLOAD_MASS__KG_) as total_payload FROM SPACEXTBL WHERE Customer LIKE 'NASA (CRS)%';
```

```
 * sqlite:///my_data1.db
Done.
```

total_payload
48213

Average Payload Mass by F9 v1.1

Display average payload mass carried by booster version F9 v1.1

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) as average_payload FROM SPACEXTBL WHERE "Booster_Version" LIKE 'F9 v1.1';
```

```
 * sqlite:///my_data1.db
Done.
```

average_payload
2928.4

First Successful Ground Landing Date

List the date when the first succesful landing outcome in ground pad was acheived.

Hint:Use min function

```
]: %sql SELECT min(Date) from SPACEXTBL where "Landing_Outcome" = 'Success (ground pad)';
```

```
 * sqlite:///my_data1.db
Done.
```

]:

min(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

```
%sql SELECT DISTINCT "Booster_Version" FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000 AND "Landing_Outcome" = 'Success (drone ship
```

```
 * sqlite:///my_data1.db
Done.
```

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

List the total number of successful and failure mission outcomes

```
%sql SELECT "Mission_Outcome", COUNT(*) AS TOTAL_Outcomes FROM SPACEXTBL GROUP BY "Mission_Outcome" ORDER BY "Mis
```

```
 * sqlite:///my_data1.db
Done.
```

Mission_Outcome	TOTAL_Outcomes
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

```
]: %sql select "Booster_Version" from SPACEXTBL where "PAYLOAD_MASS__KG_"=(select MAX("PAYLOAD_MASS__KG_") from SPACI
```

```
 * sqlite:///my_data1.db
Done.
```

]:

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

List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.

Note: SQLLite does not support monthnames. So you need to use substr(Date, 6,2) as month to get the months and substr(Date,0,5)='2015' for year.

```
%sql select substr(Date, 6,2) as Month, "Landing_Outcome", "Booster_Version", "Launch_Site" from SPACEXTABLE wher
```

```
 * sqlite:///my_data1.db
Done.
```

Month	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

%sql select substr(Date, 6,2) as Month, "Landing_Outcome", "Booster_Version", "Launch_Site" from SPACEXTABLE where substr(Date,0,5)='2015' and "Landing_Outcome"='Failure (drone ship)';

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

```
%sql select "Landing_Outcome", count("Landing_Outcome") as Count_Landing from SPACEXTBL where "Date" between '201
```

```
 * sqlite:///my_data1.db
Done.
```

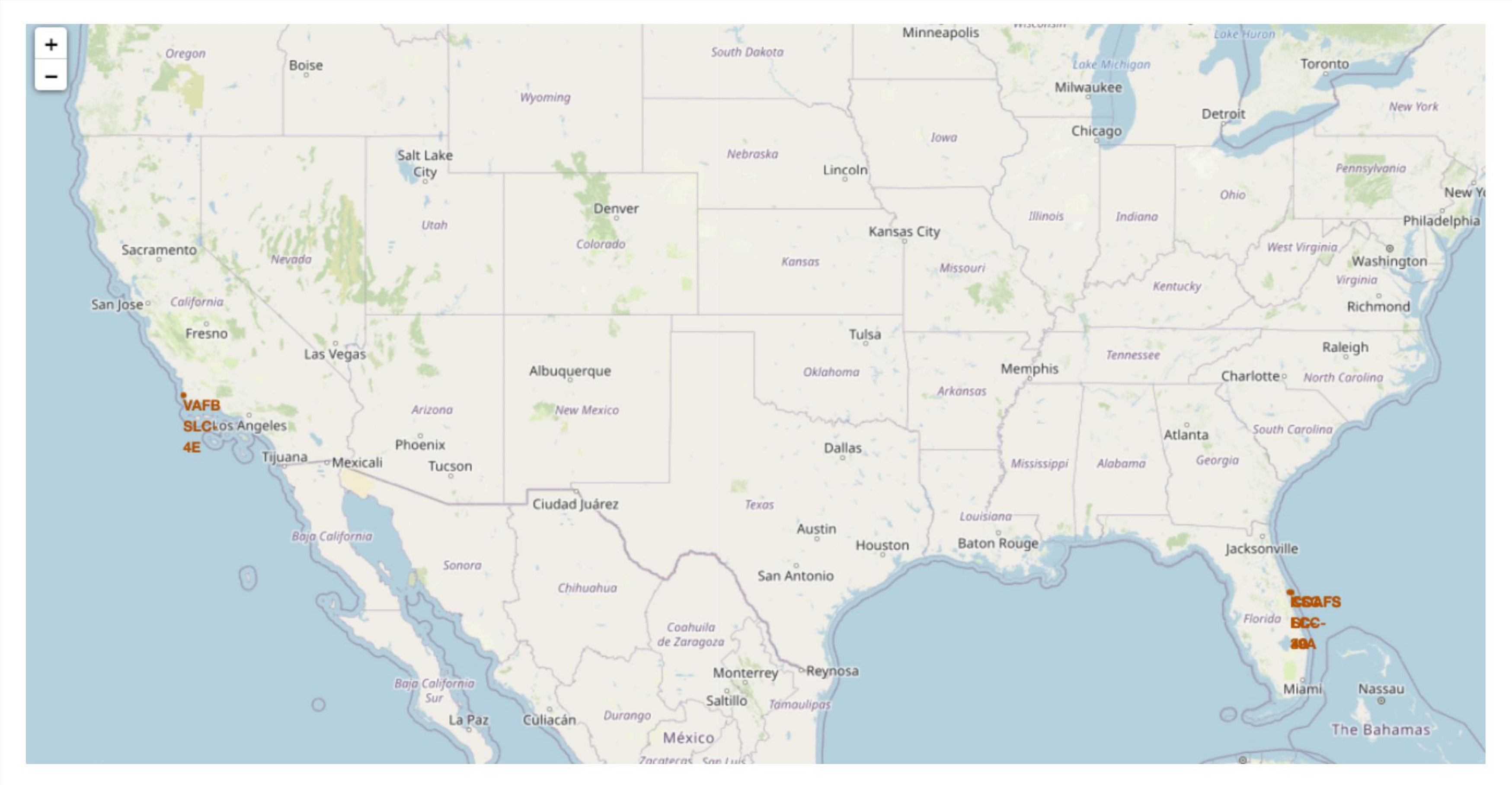
Landing_Outcome	Count_Landing
Controlled (ocean)	3
Failure (drone ship)	5
Failure (parachute)	2
No attempt	10
Precluded (drone ship)	1
Success (drone ship)	5
Success (ground pad)	3
Uncontrolled (ocean)	2

%sql select "Landing_Outcome", count("Landing_Outcome") as Count_Landing from SPACEXTBL where "Date" between '2010-06-04' and '2017-03-20' GROUP BY "Landing_Outcome";

Section 3

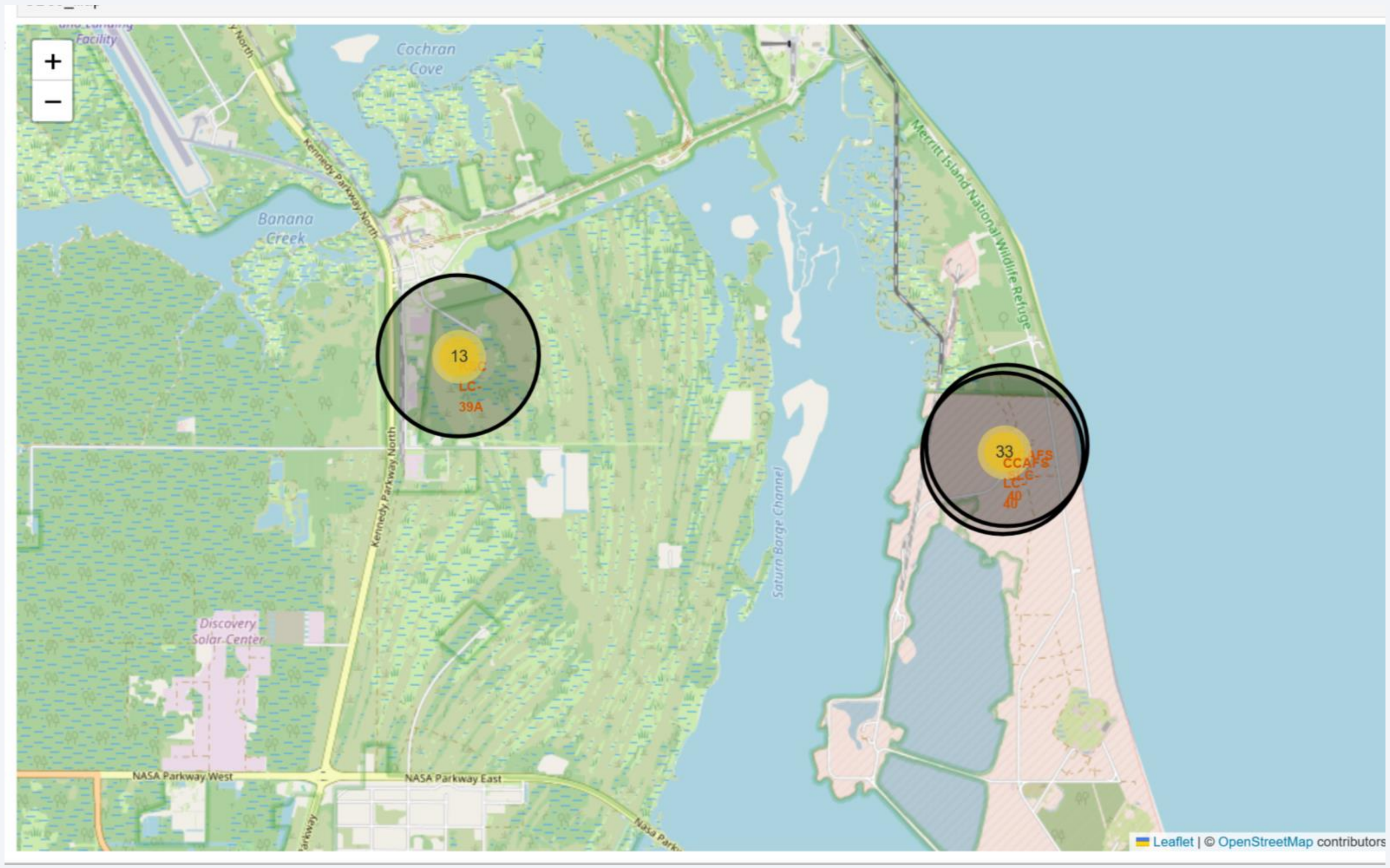
Launch Sites Proximities Analysis

Launch site locations



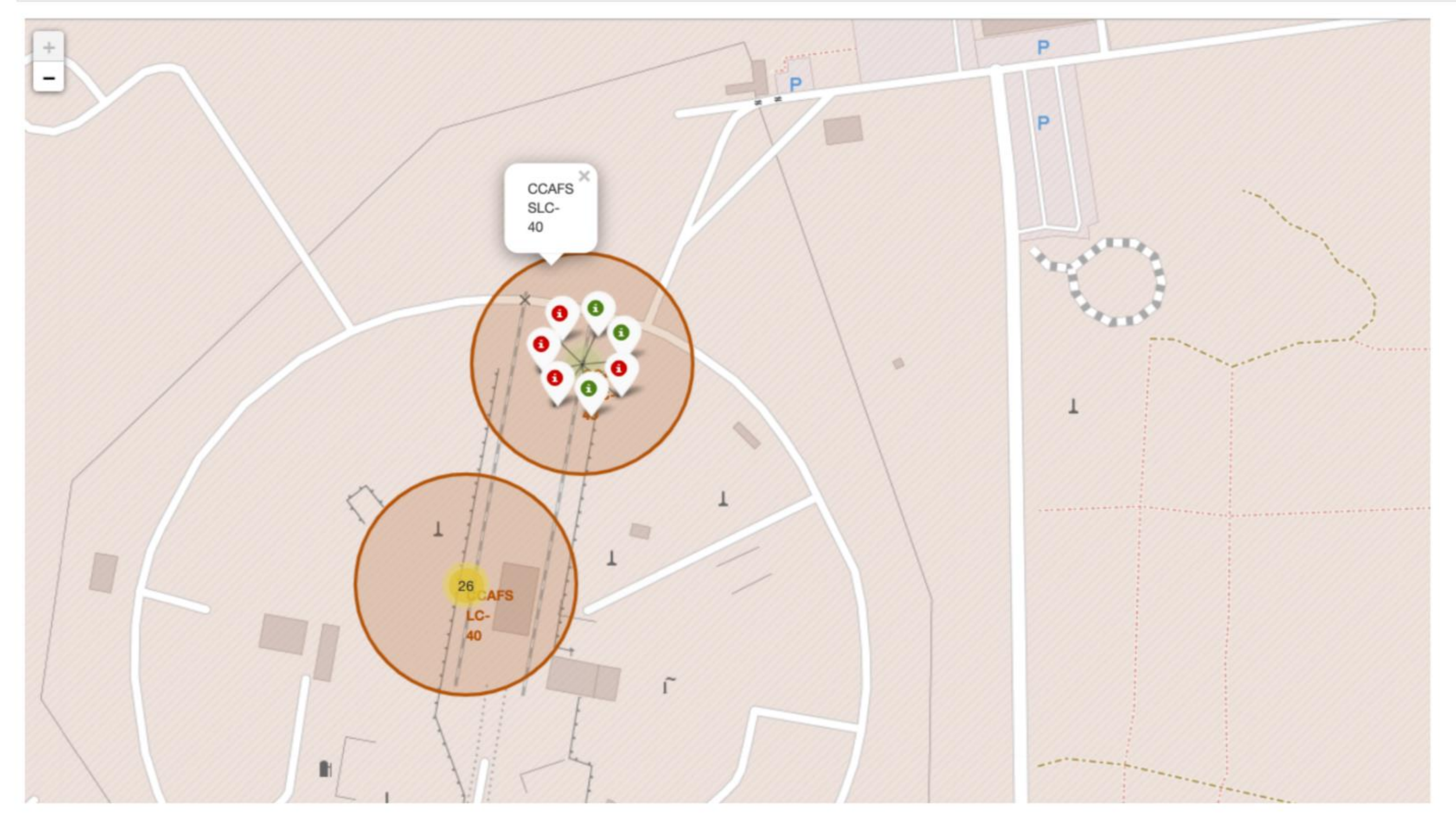
Note that all launch sites are close to equator and close to the coast line

Zoom in map with circle markers



Note close proximity of some launch sites

Success/failed labling

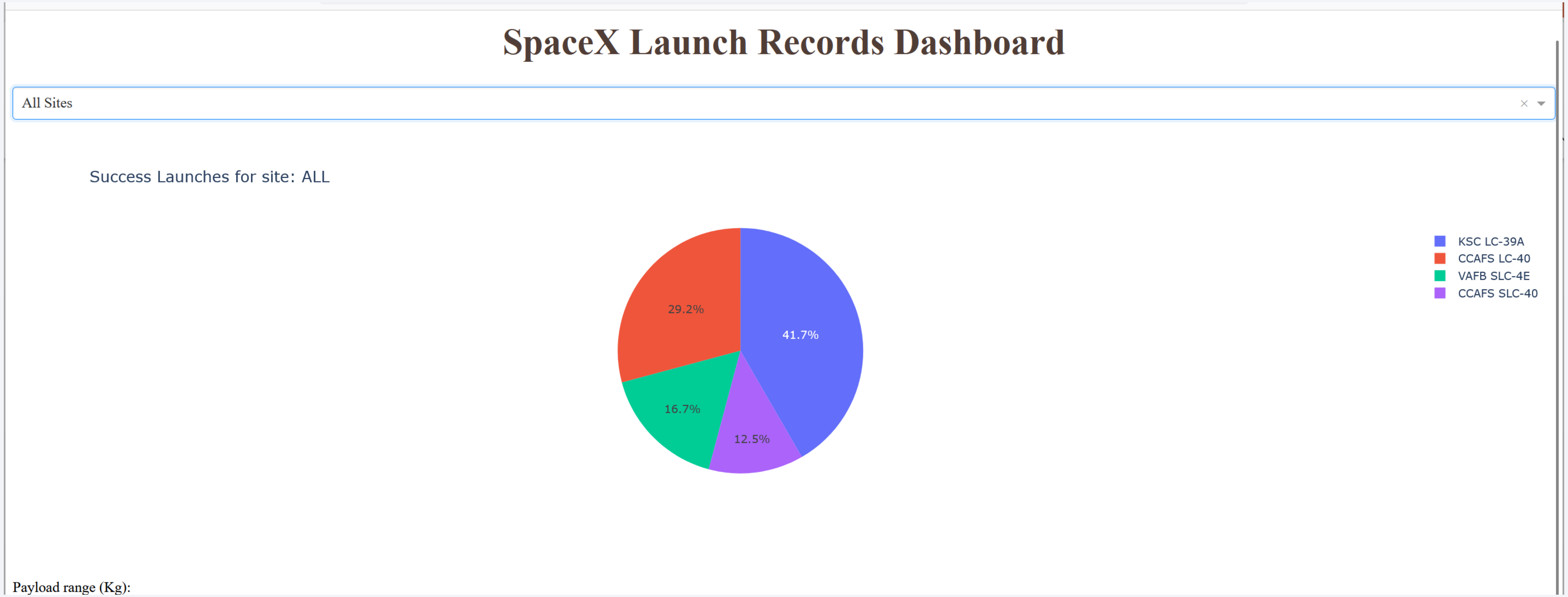


From the color-labeled markers in marker clusters, you should be able to easily identify which launch sites have relatively high success rates

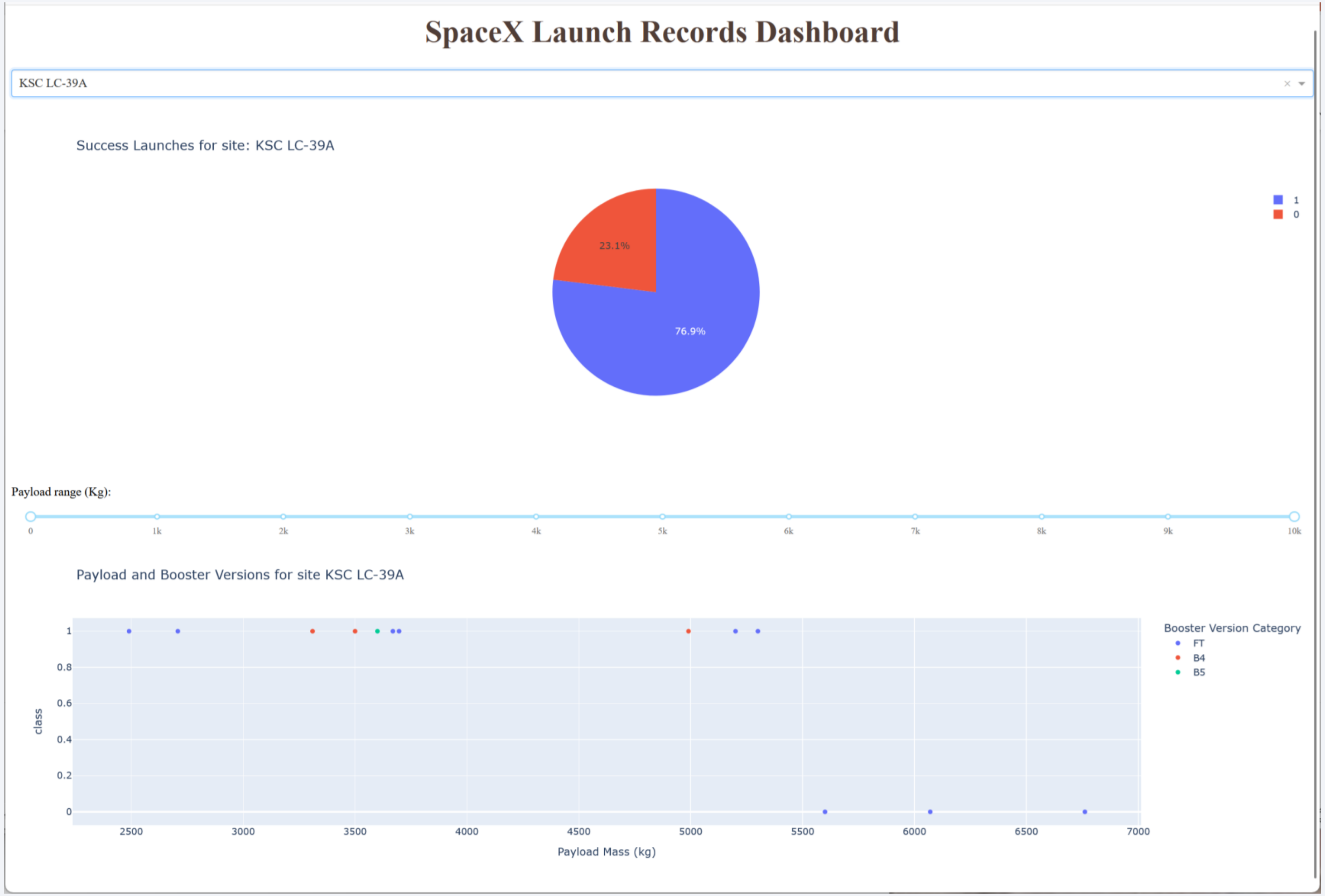
Section 4

Build a Dashboard with Plotly Dash

Piechart for launch success count for all sites

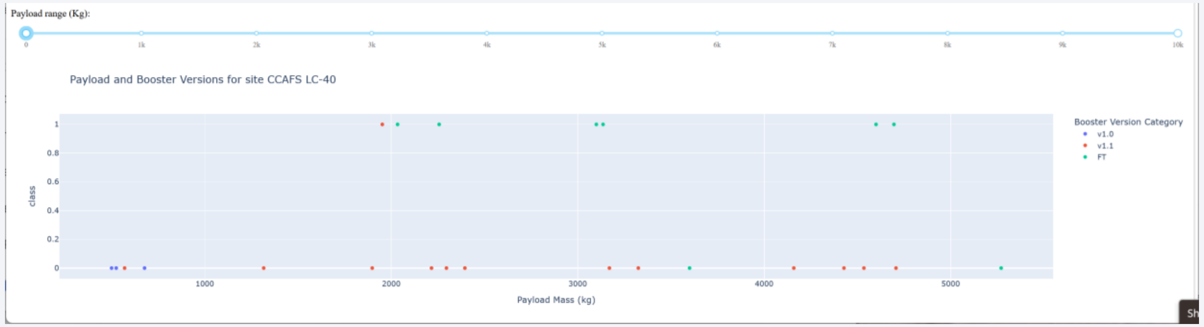


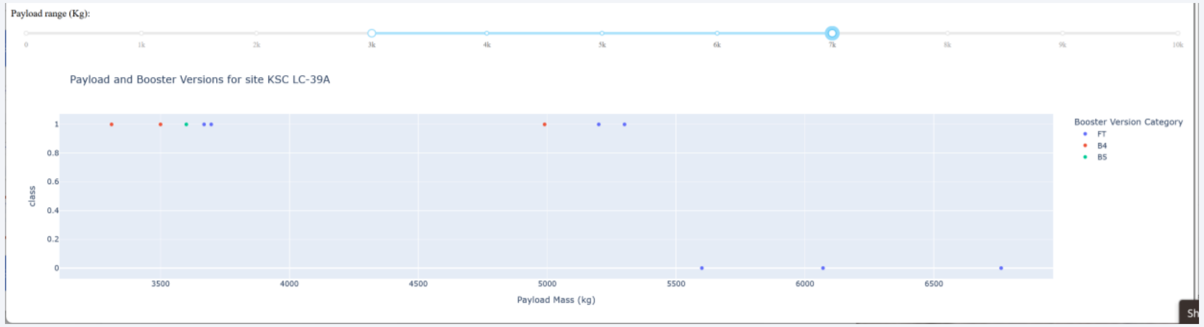
Piechart for the launch site with highest launch success ratio

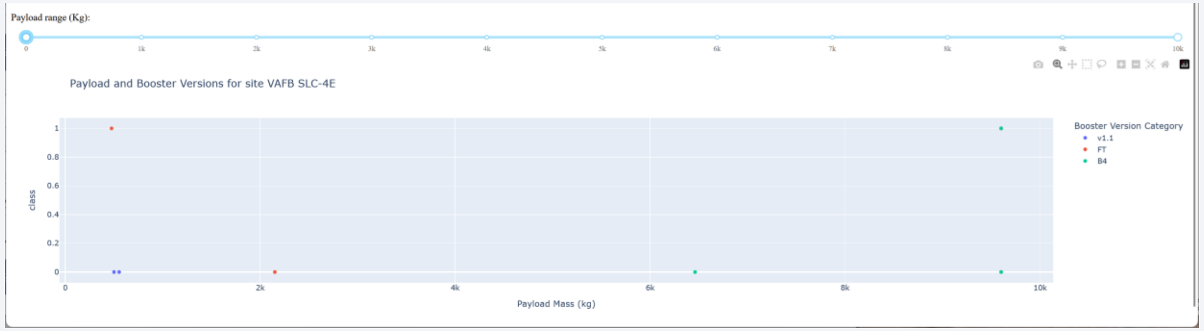


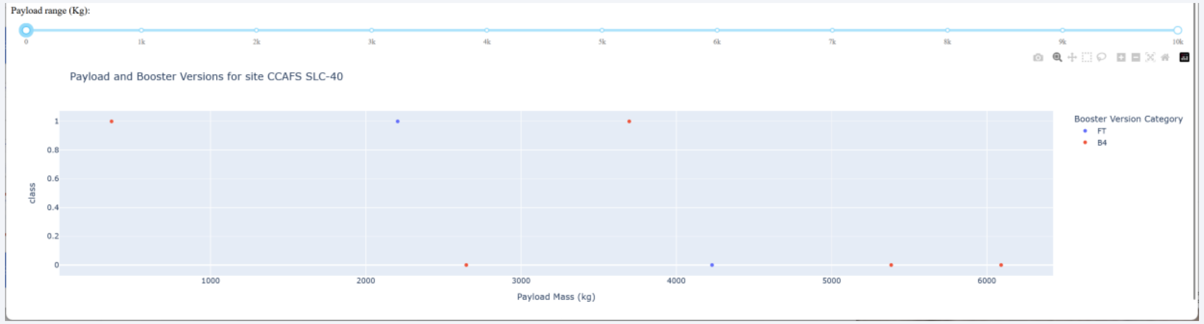
Launch site KSC LC-39A has a highest launch success ratio

Payload vs. Launch Outcome scatter plot for all sites







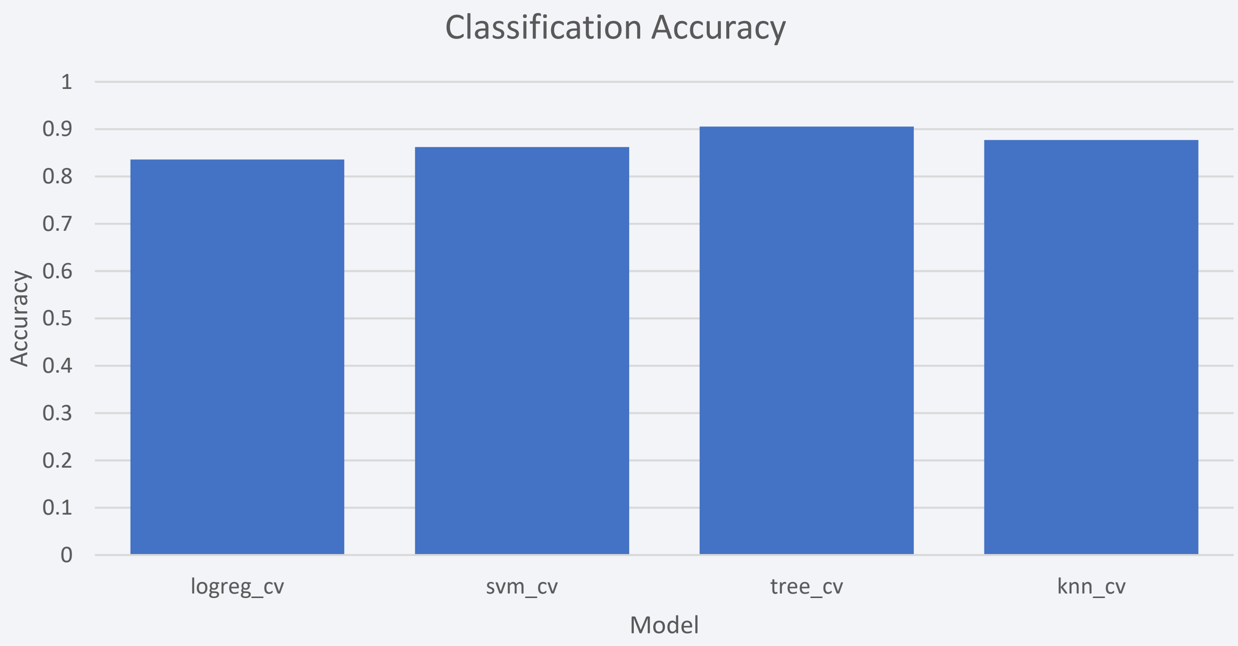


https://github.com/TeyaTopuria/applied-data-science-capstone/blob/main/spacex_dash_app(1)working.py

Section 5

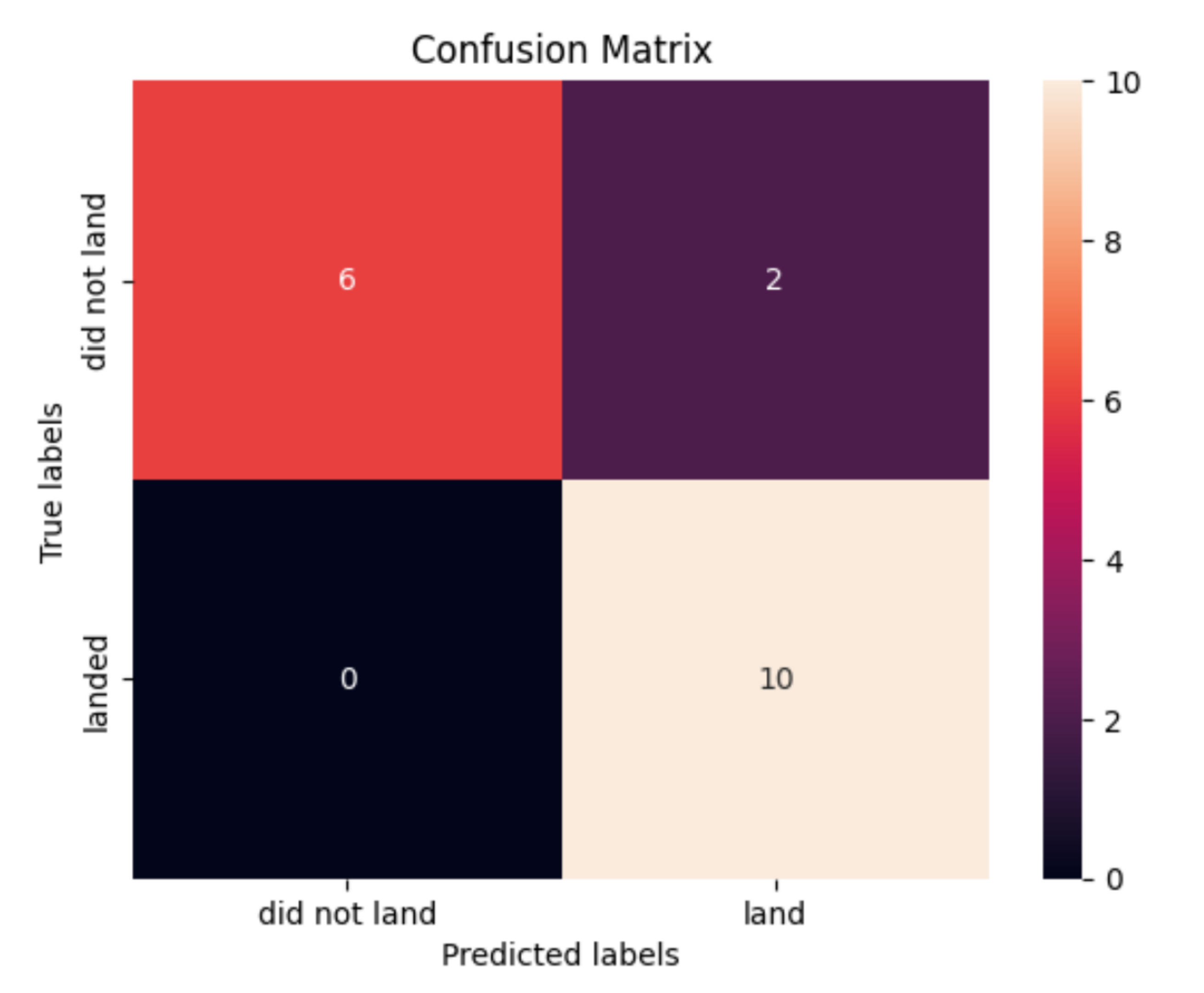
Predictive Analysis (Classification)

Classification Accuracy



Decision Tree has highest accuracy

Confusion Matrix



Here is Decision Tree confusion matrix

Conclusions

- The success rate was determined ~67%
- Decision Tree model has highest accuracy
- Launch site KSC LC-39A has a highest launch success ratio

Thank you!