IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Michael
5/13/2025

Outline

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

Executive Summary

- High-level description of project goals: predicting Falcon 9 the first stage landing success to assess launch cost efficiency

Introduction

- SpaceX has revolutionized the aerospace industry by introducing reusable rocket technology. One of the most significant cost-saving measures is the ability to recover and reuse the Falcon 9 rocket's first stage. A successful landing significantly reduces the cost of a launch, bringing it down from approximately $165 million to just $62 million.
- Problems you want to find answers:
- 1. Can we predict whether the first stage of falcon 9 will land successfully based on launch characteristics?
- 2. What is the predicted cost of a SpaceX Launch?
- 3. How can visualizations help decision-makers quickly identify patterns in successful vs failed landings?

Section 1

Methodology

Methodology

Executive Summary

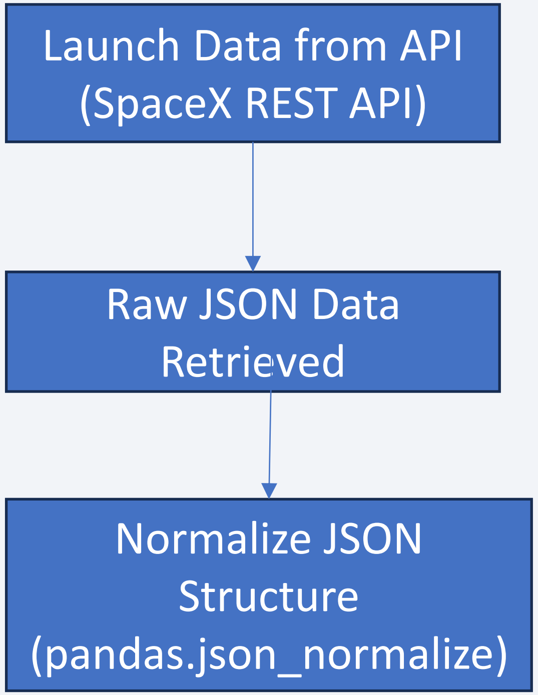
- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

- Data was collected from the SpaceX URL using an API request
- This allowed us to get data straight from the source

Data Collection – SpaceX API

- To begin the analysis, data was collected directly from the publicly available SpaceX RESTful API
- This API provides detailed information about Falcon 9 rocket launches, including mission details, payloads, rocket specifications, launch sites, and landing outcomes
- By retrieving raw JSON responses, data was normalized by using pandas.json_normalize

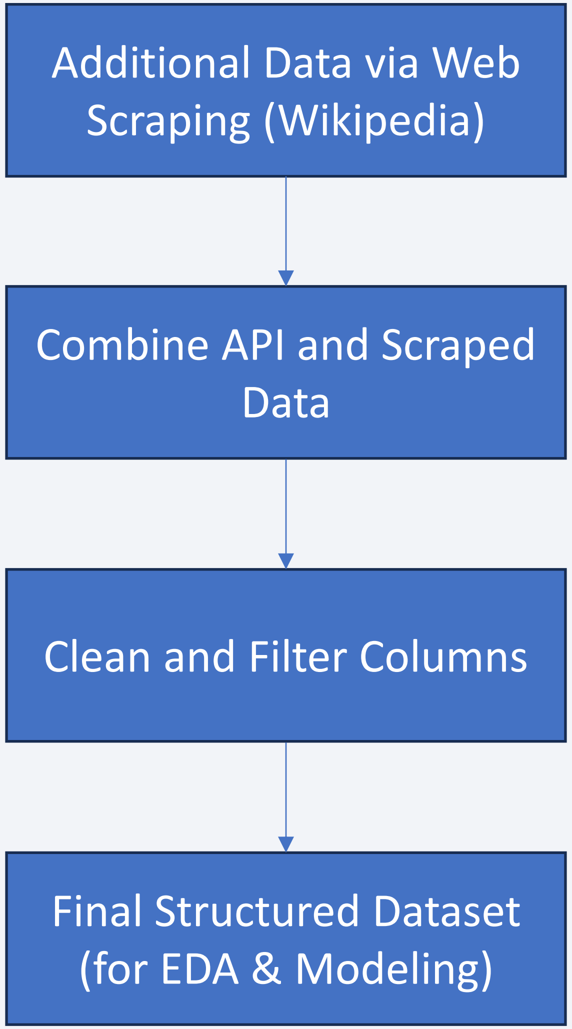


- Link
- https://github.com/poofmagicdragon/Coursera-Data-Science-Submissions/blob/f2295d2a693b0740813ba68f9903065609937ab1/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping

- The scaping process was conducted using Python’s requests library to fetch HTML content and BeautifulSoup to parse and extract relevant information

- Link
  - https://github.com/poofmagicdragon/Coursera-Data-Science-Submissions/blob/f2295d2a693b0740813ba68f9903065609937ab1/jupyter-labs-spacex-data-collection-api.ipynb



Data Wrangling

- Data wrangling was done by converting nested JSON responses into flat tables using json_normalize
- Columns like payload mass, orbit type, and launch site were standardized, and missing values were handled through imputation or removal
- Categorical variables were transformed using one hot encoding

Link

- https://github.com/poofmagicdragon/Coursera-Data-Science-Submissions/blob/a7367941173bd9941c8d4400fcfe2918cd71bb73/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- Graphs we used include:
- A scatter plot Flight number vs Launch Site
- A scatter plot of Payload Mass vs Launch Site
- A bar chart of Class vs Orbit

- Link
- https://github.com/poofmagicdragon/Coursera-Data-Science-Submissions/blob/7fd4d18f79e662c9998f7183dbac8f4f4abc2c7f/edadataviz.ipynb

EDA with SQL

- SQL Queries include:
- Names of unique launch sites
- Total payload mass carried by boosters launched by NASA
- Average payload mass carried by booster version F9 v1.1

- Link
- https://github.com/poofmagicdragon/Coursera-Data-Science-Submissions/blob/7fd4d18f79e662c9998f7183dbac8f4f4abc2c7f/edadataviz.ipynb

Build an Interactive Map with Folium

- Circles are used to highlight the location of the launch site to make it easier to see
- Launch outcomes were added to each site so we can see which ones have the most number of successes

Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard
- Explain why you added those plots and interactions
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model
- You need present your model development process using key phrases and flowchart
- Add the GitHub URL of your completed predictive analysis lab, as an external reference and peer-review purpose

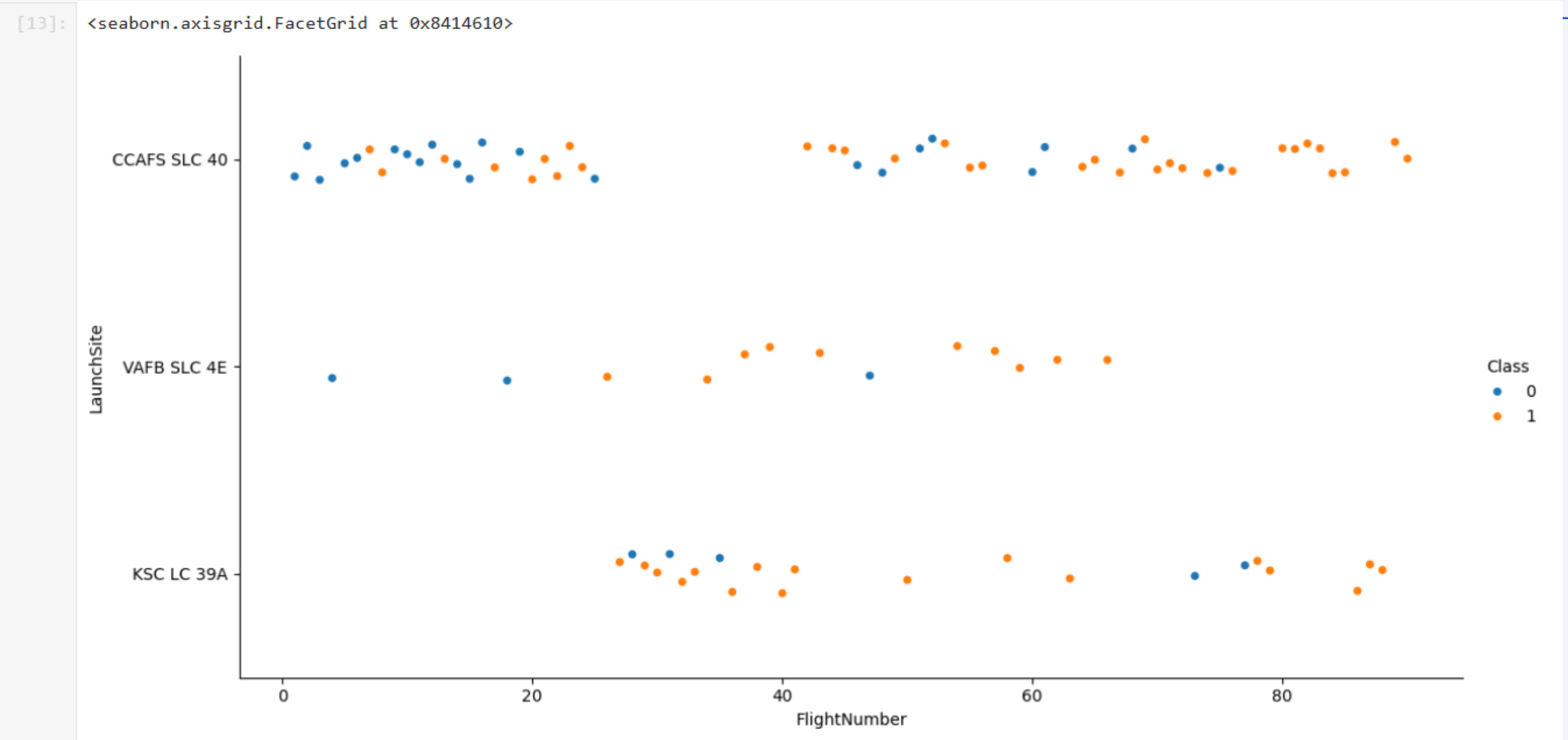
Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

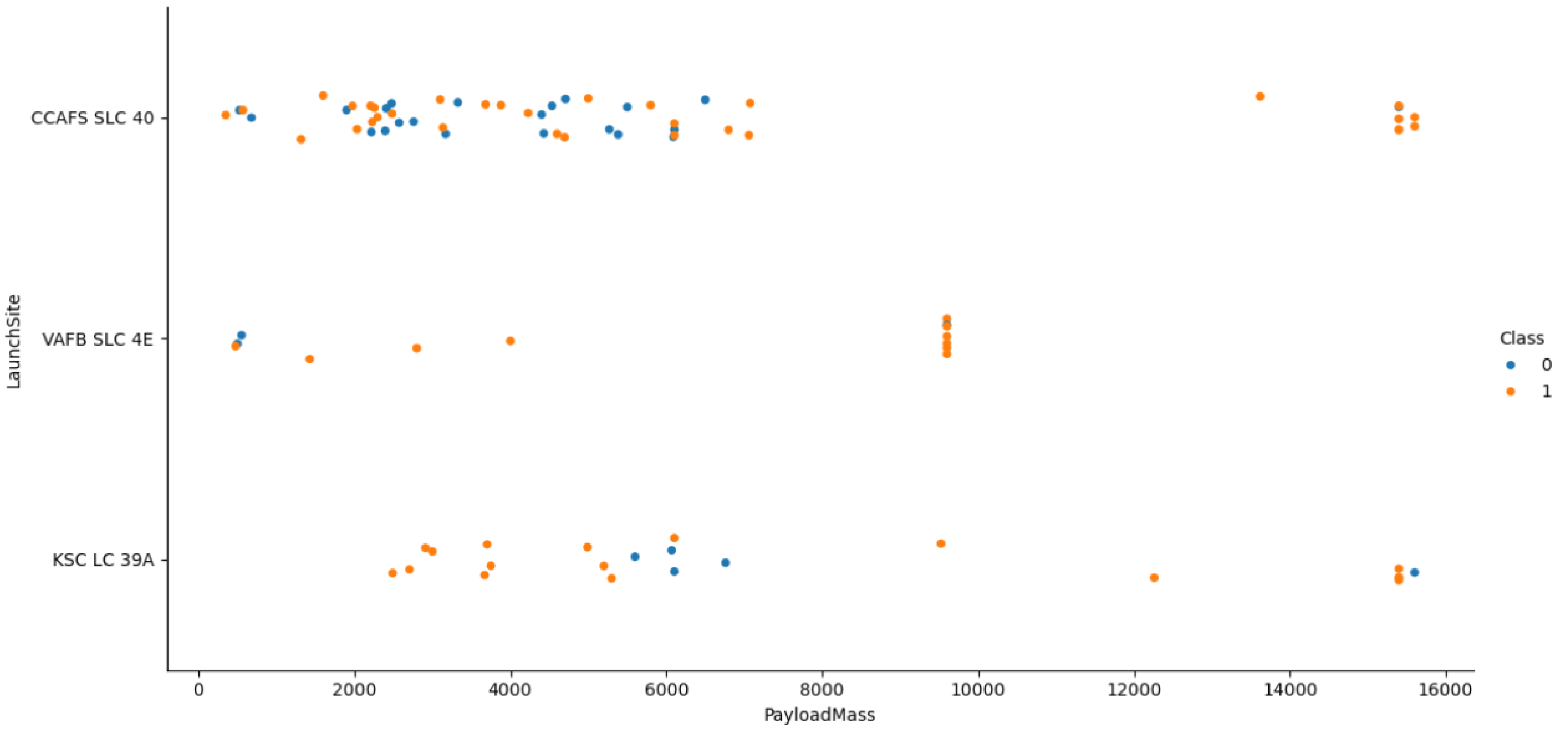
Flight Number vs. Launch Site



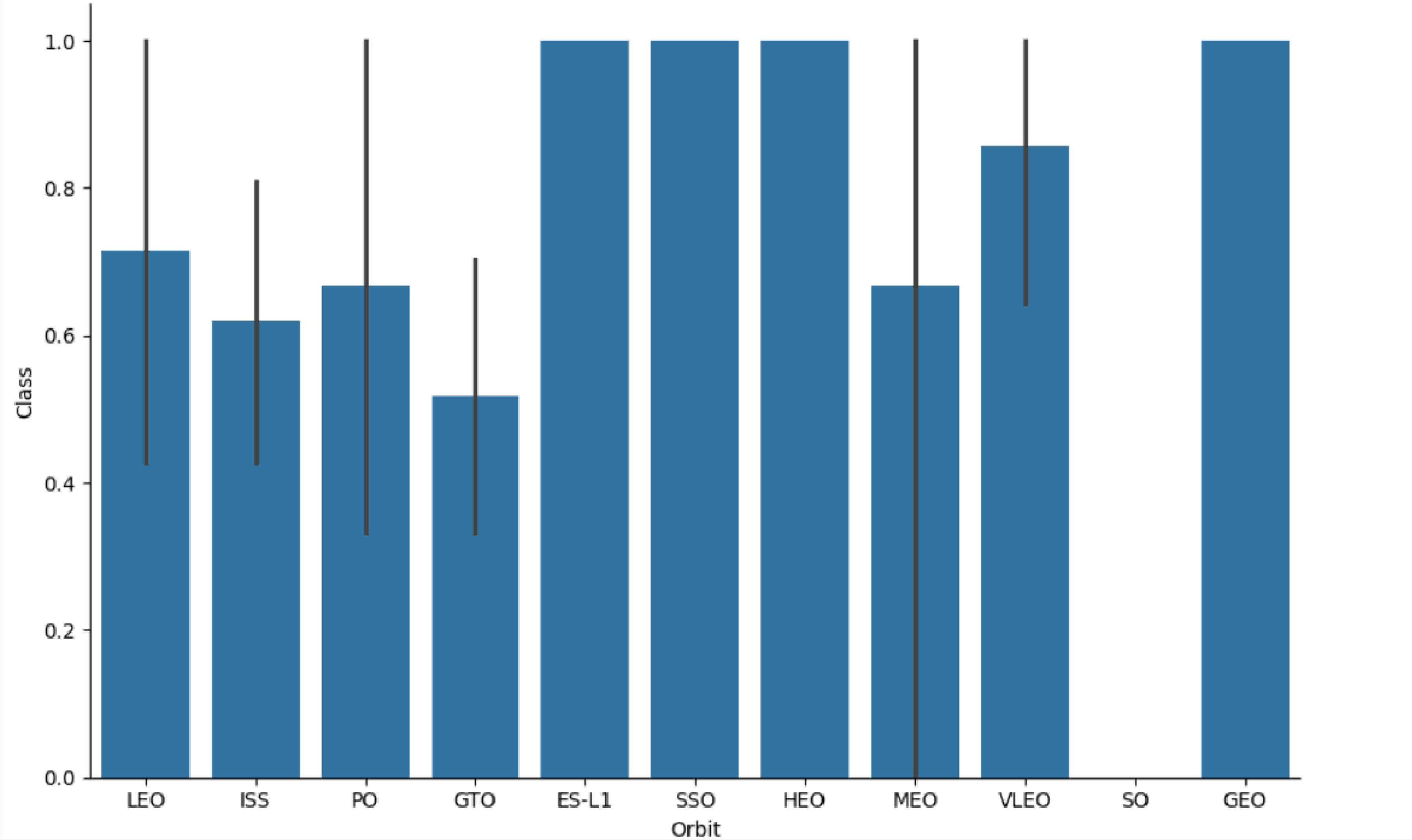
We can see that SLC has two clusters. The first has mostly Class 0 and the second has mostly class 1.

Payload vs. Launch Site

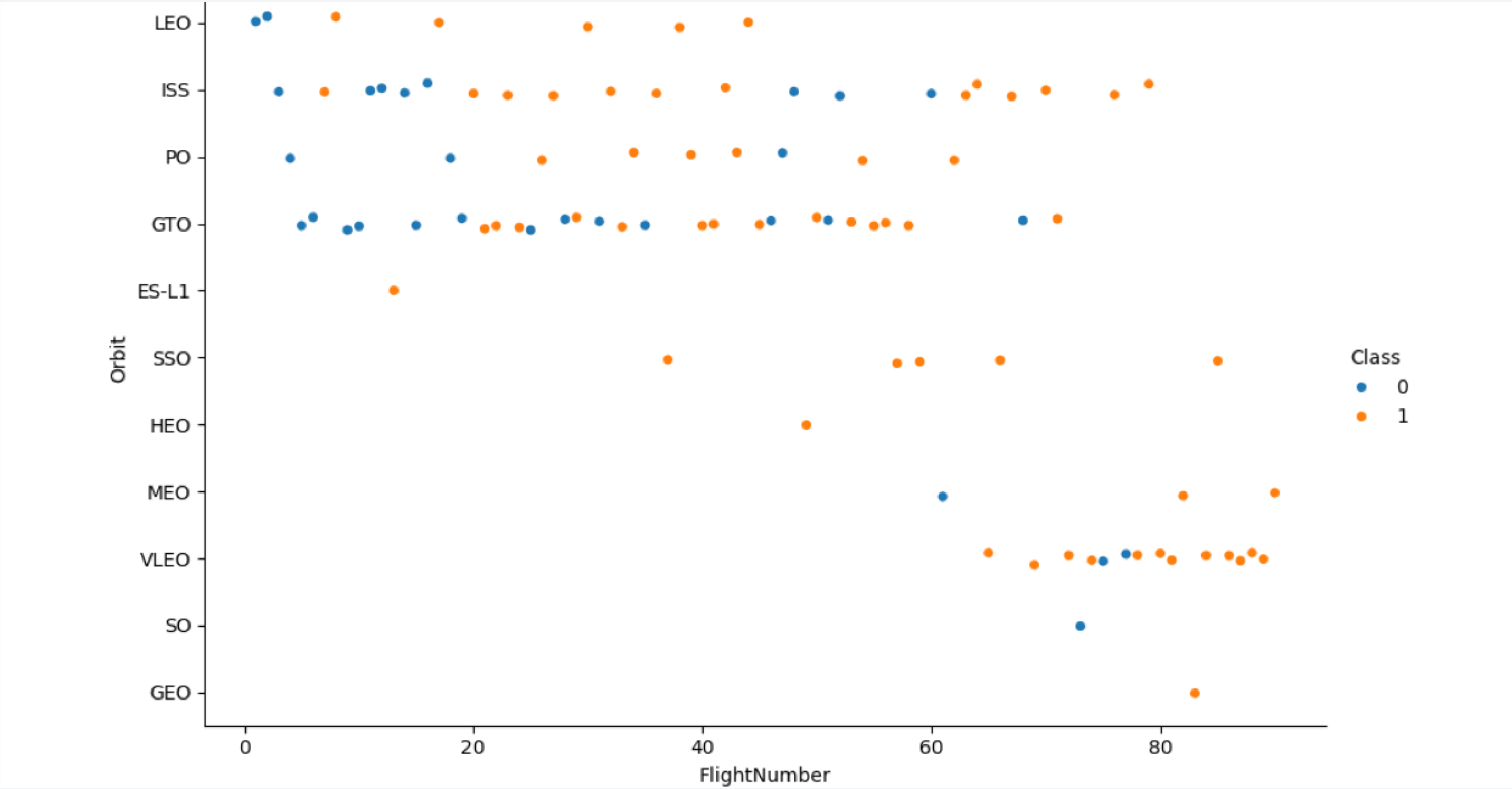
[16]: <seaborn.axisgrid.FacetGrid at 0x851f2d0>



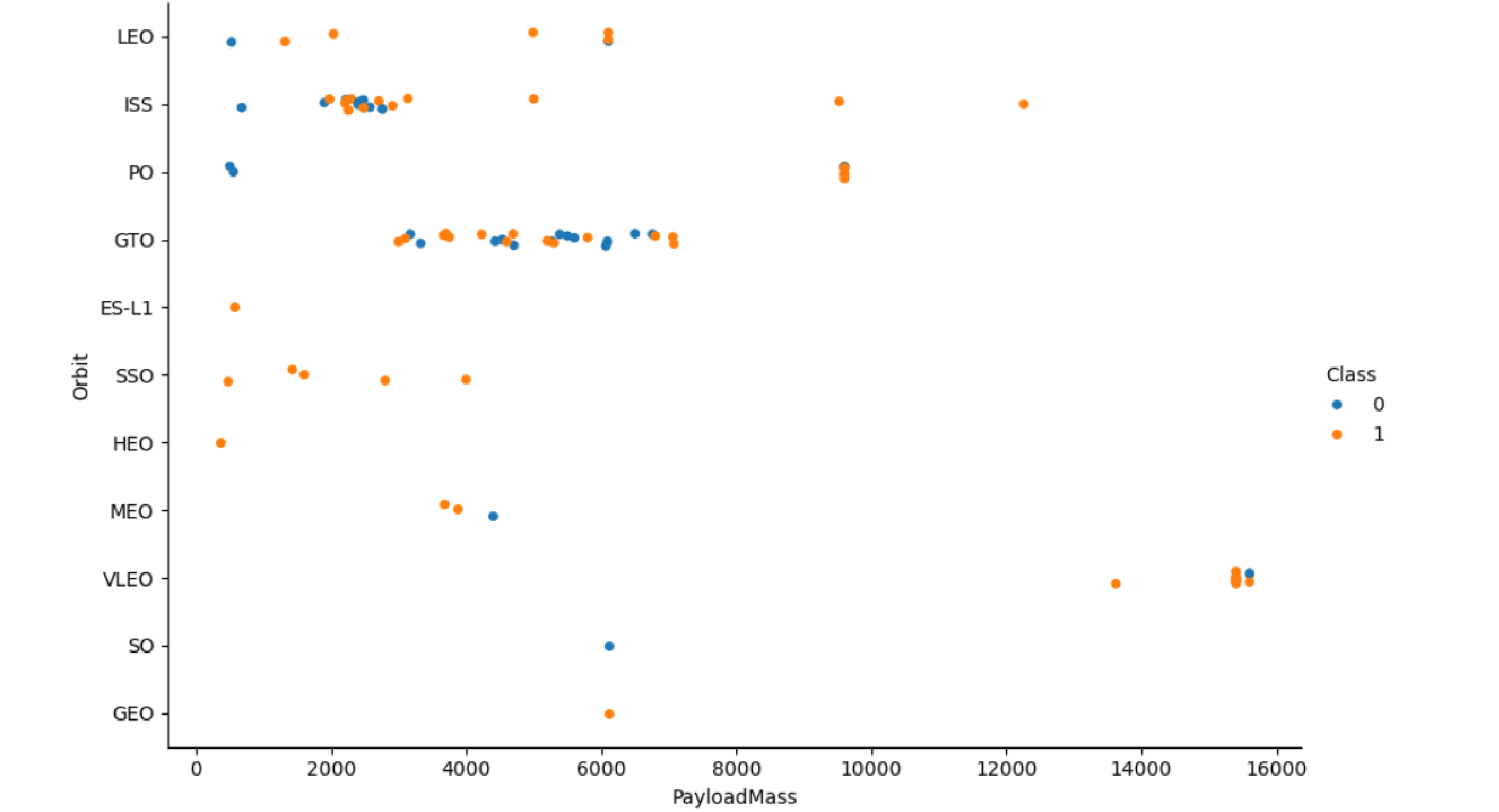
Success Rate vs. Orbit Type



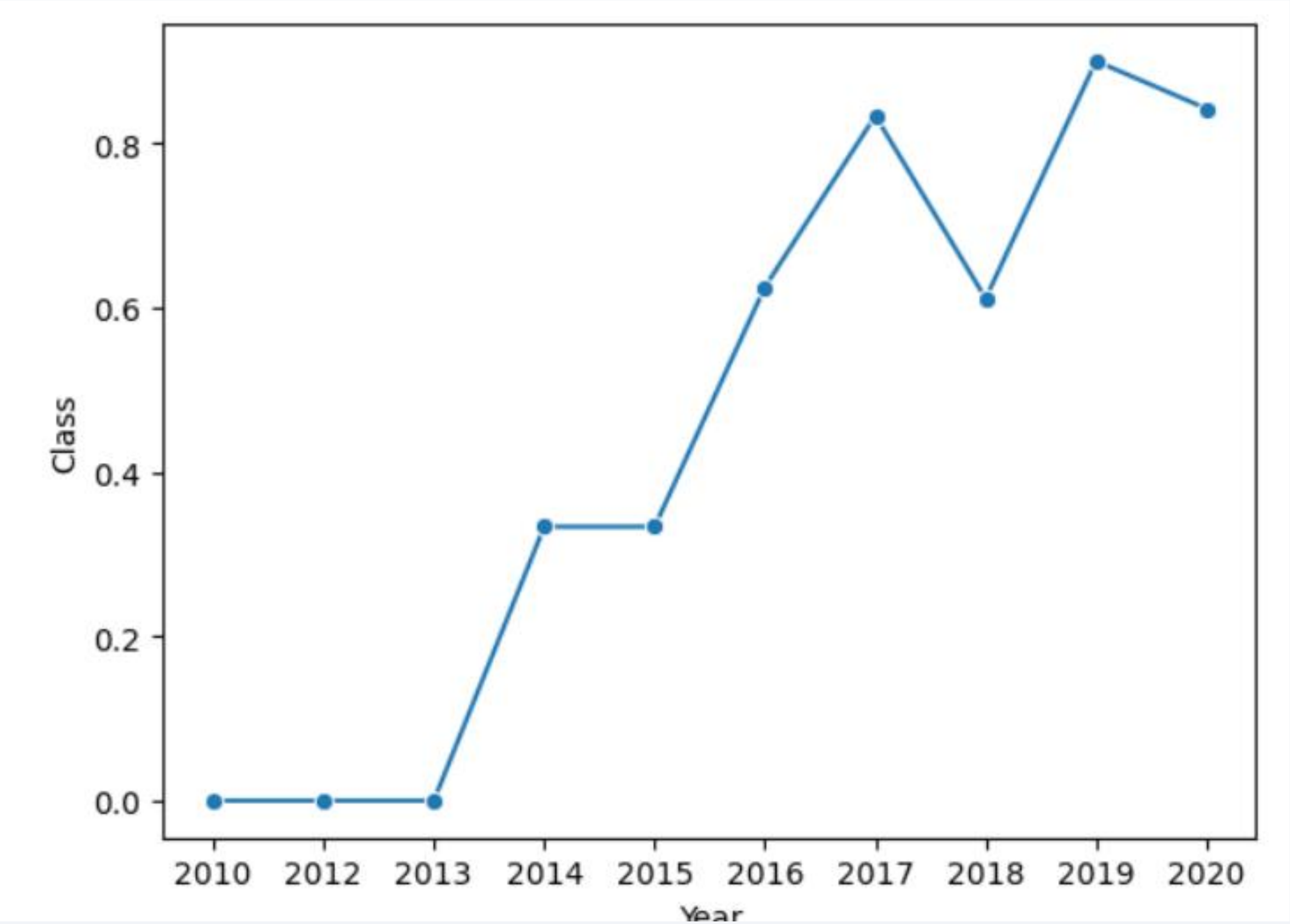
Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

Display the names of the unique launch sites in the space mission

[16]:
```
%sql SELECT DISTINCT(Launch_Site) FROM SPACEXTABLE;
```

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

[16]:

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

Launch Site Names Begin with 'CCA'

Display 5 records where launch sites begin with the string 'CCA'

[26]:
```
%%sql SELECT * FROM SPACEXTABLE
WHERE Launch_Site LIKE 'CCA%'
LIMIT 5;
```

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

[26]:

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

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

[27]:
```
%sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTABLE
```

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

[27]:

SUM(PAYLOAD_MASS__KG_)
619967

Average Payload Mass by F9 v1.1

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

[28]:

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

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

[28]:

AVG(PAYLOAD_MASS__KG_)
340.4

First Successful Ground Landing Date

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

Hint:Use min function

[31]:

```
%sql SELECT Date FROM SPACEXTABLE WHERE Landing_Outcome == 'Success' ORDER BY Date LIMIT 1;
```

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

[31]:

Date
2018-07-22

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

[33]:
```
%%sql SELECT Booster_Version FROM SPACEXTABLE
WHERE PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000 AND Mission_Outcome == 'Success'
```

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

[33]:

Booster_Version
F9 v1.1
F9 v1.1 B1011
F9 v1.1 B1014
F9 v1.1 B1016
F9 FT B1020
F9 FT B1022
F9 FT B1026
F9 FT B1030
F9 FT B1021.2
F9 FT B1032.1
F9 B4 B1040.1

Total Number of Successful and Failure Mission Outcomes

List the total number of successful and failure mission outcomes

```
%%sql SELECT COUNT(Mission_Outcome) as Successes FROM SPACEXTABLE
WHERE Mission_Outcome == 'Success'
```

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

Successes
98

```
%%sql SELECT COUNT(Mission_Outcome) as Failures FROM SPACEXTABLE
WHERE Mission_Outcome == 'Failure (in flight)'
```

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

Failures
1

Boosters Carried Maximum Payload

```
%%sql SELECT Booster_Version FROM SPACEXTABLE
WHERE PAYLOAD_MASS__KG_ == (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTABLE)
```

```
 * 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

```
%%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 LIKE 'Failure%'
```

```
 * 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

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

```
%%sql SELECT COUNT(Landing_Outcome) FROM SPACEXTABLE
WHERE Date BETWEEN '2010-06-04' AND '2017-03-20'
```

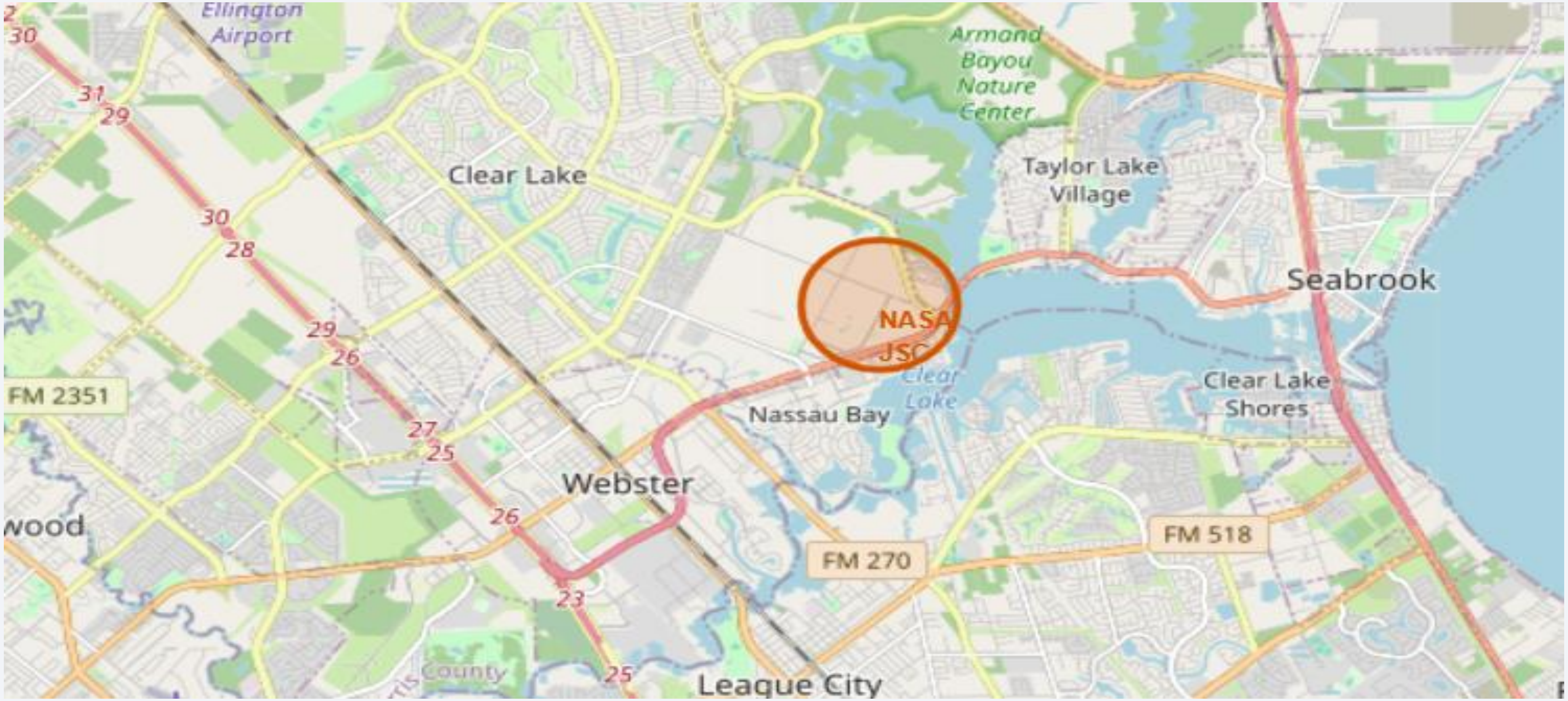
```
 * sqlite:///my_data1.db
Done.
```

COUNT(Landing_Outcome)
31

Section 3

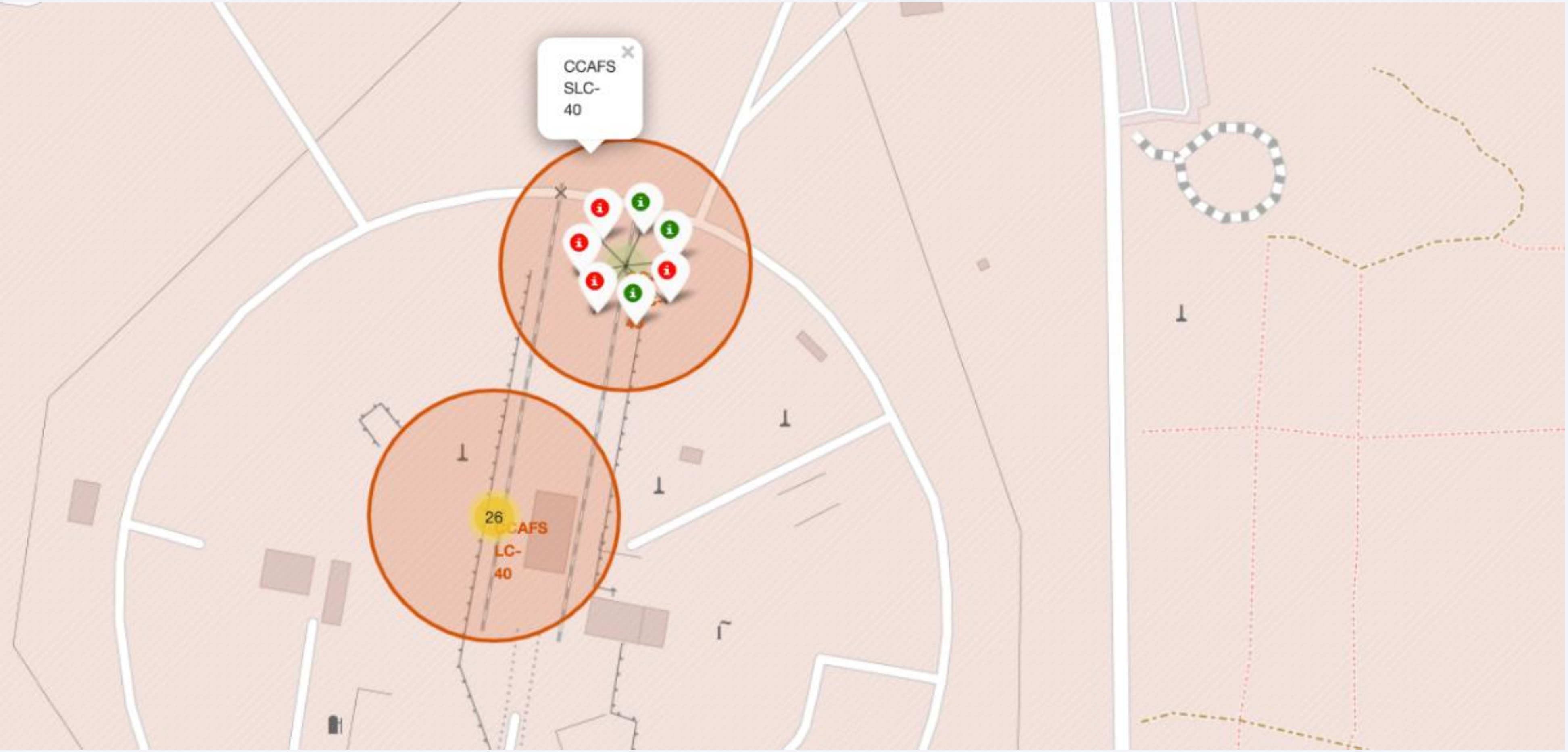
Launch Sites Proximities Analysis

Houston, Texas NASA Johnson Space Center



Number of Launches from Different Stations

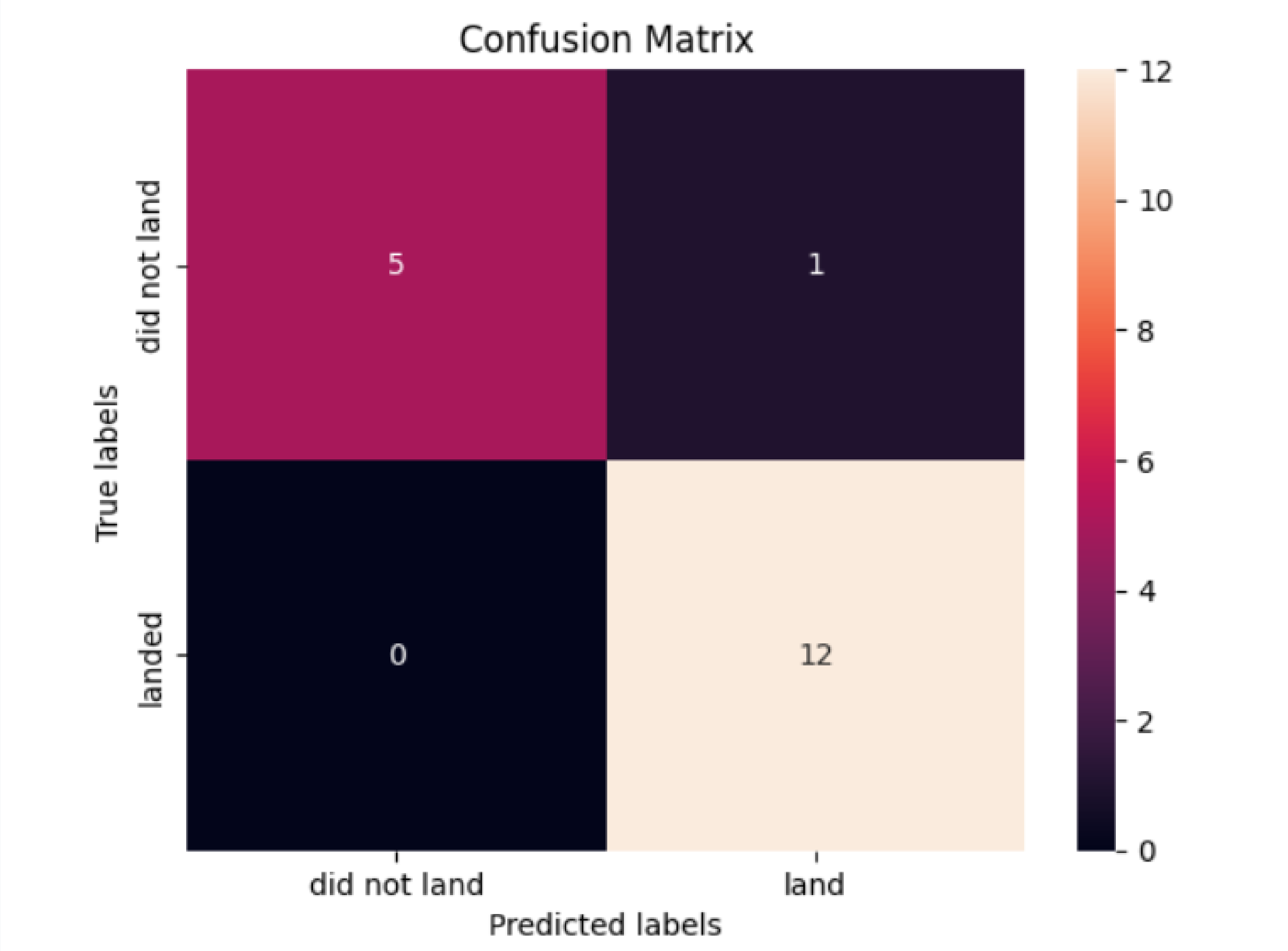
Successes of Launches at Certain Space Stations



Section 4

Predictive Analysis (Classification)

Confusion Matrix



Conclusions

- This project set out to analyze and predict the success of SpaceX Falcon 9 first-stage landings
- From exploratory analysis payload mass had correlations with success rates, and certain orbit types
- We also used machine learning and grid search based on various launch features
- These models offer potential for predicting the likelihood of landing success.

Thank you!