IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Naveed Ahmed Khatri

30th September 2023

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

- In this capstone project, there were total five modules. Each module served as a final exam of what was already covered in the previous courses of this specialization.
- Data collection from API, Data Wrangling, EDA with Pandas and Matplotlib, SQL, visualization using maps and dashboarding, and machine learning model with implementation, all were included.

Introduction

- It was our goal to 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.

Section 1

Methodology

Data Collection – SpaceX API

- Request data from SpaceX API (rocket launch data)
- Decode response using . json () and convert to a dataframe using . json_normalize
- Request information about the launches from SpaceX API using custom functions
- Create dictionary from the data
- Create dataframe from the dictionary
- Filter dataframe to contain only Falcon 9 launches
- Replace missing values of Payload Mass with calculated .mean()
- Export data to csv file

Data Collection - Scraping

- Request data (Falcon 9 launch data) from Wikipedia
- Create BeautifulSoup object from HTML response
- Extract column names from HTML table header
- Collect data from parsing HTML tables
- Create dictionary from the data
- Create dataframe from the dictionary
- Export data to csv file

EDA with SQL

- Queries Display
- Names of unique launch sites
- 5 records where launch site begins with 'CCA'
- Total payload mass carried by boosters launched by NASA (CRS)
- Average payload mass carried by booster version F9 v1.1.List
- Date of first successful landing on ground pad
- Names of boosters which had success landing on drone ship and have payload mass greater than 4,000 but less than 6,000

EDA with SQL

- Total number of successful and failed missions
- Names of booster versions which have carried the max payload
- Failed landing outcomes on drone ship, their booster version and launch site for the months in the year 2015
- Count of landing outcomes between 201006 04 and 2017 03 20 (desc)

Build an Interactive Map with Folium

- Markers Indicating Launch Sites
- Colored Markers of Launch Outcomes
- Distances Between a Launch Site to Proximities

Build a Dashboard with Plotly Dash

- Dropdown List with Launch Sites
- Pie Chart Showing Successful Launches
- Slider of Payload Mass Range
- Scatter Chart Showing Payload Mass vs. Success Rate by Booster Version

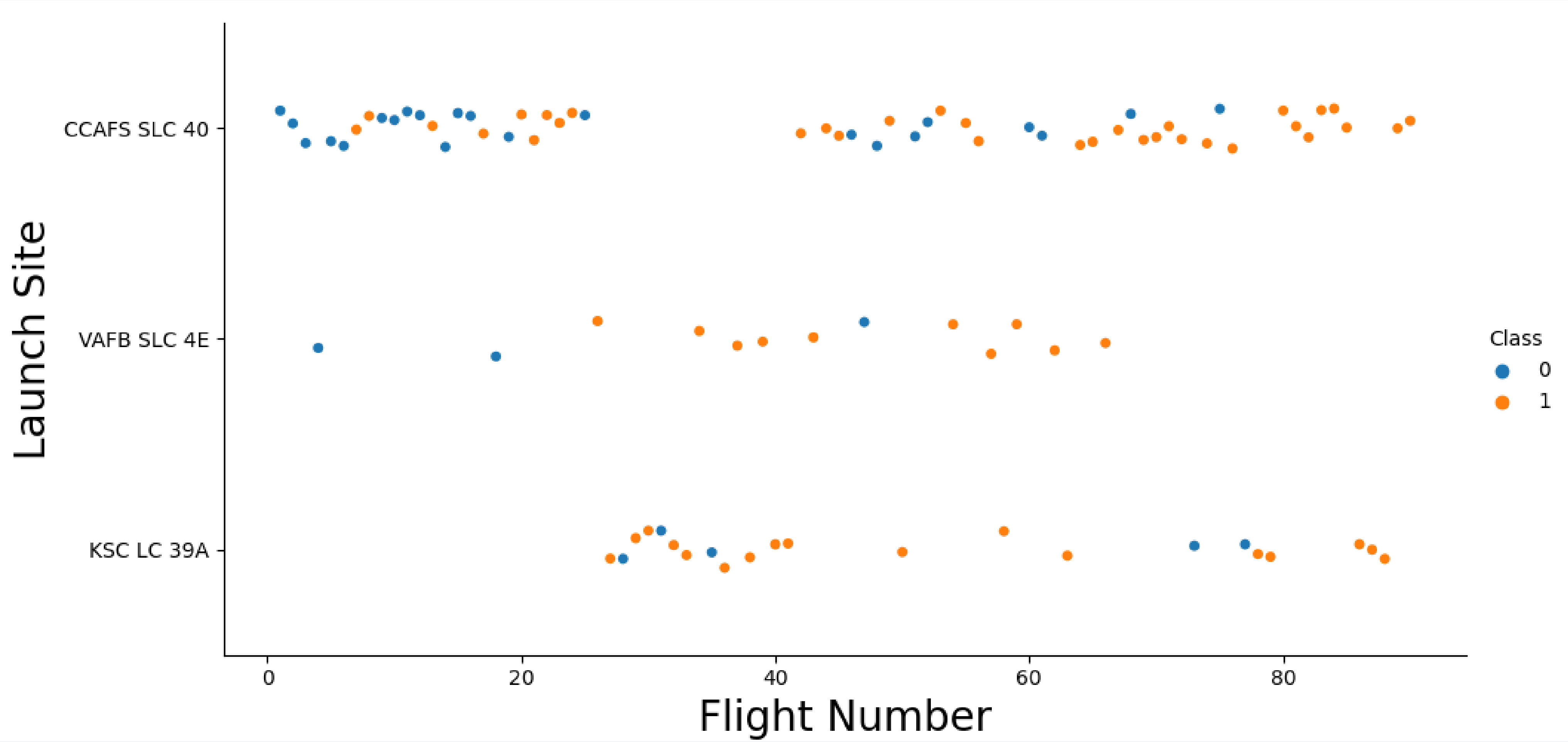
Predictive Analysis (Classification)

- Create NumPy array from the Class column
- Standardize the data with StandardScaler. Fit and transform the data.
- Split the data using train_test_split
- Create a GridSearchCV object with cv=10 for parameter optimization
- Apply GridSearchCV on different algorithms: logistic regression (LogisticRegression()), support vector machine (SVC()), decision tree (DecisionTreeClassifier()), K Nearest Neighbor (KNeighborsClassifier())
- Calculate accuracy on the test data using .score() for all models
- Assess the confusion matrix for all models

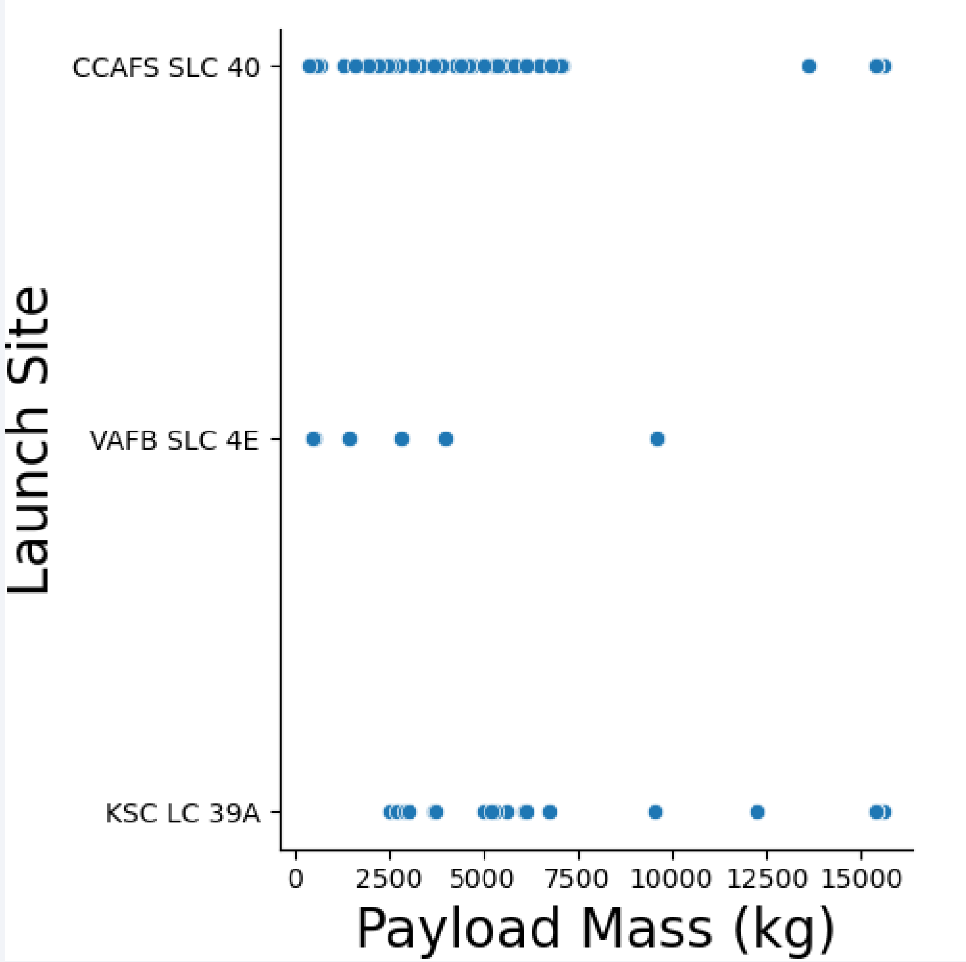
Section 2

Insights drawn from EDA

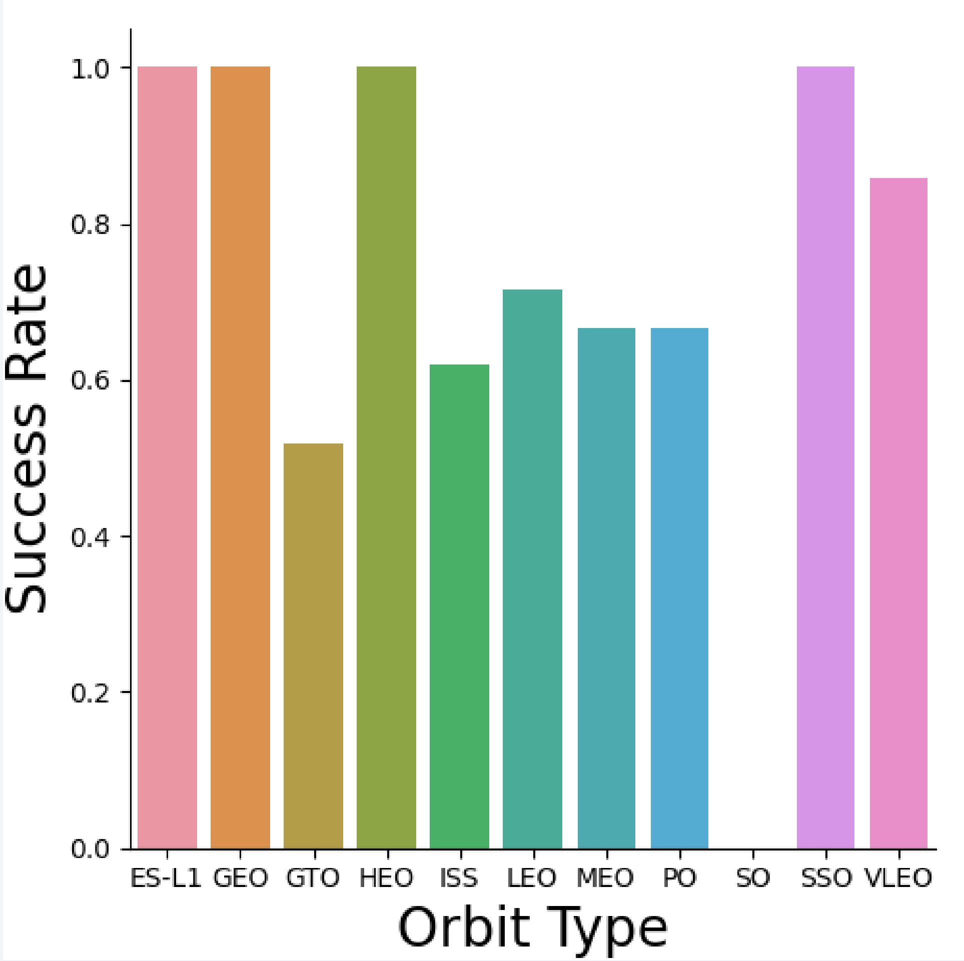
Flight Number vs. Launch Site



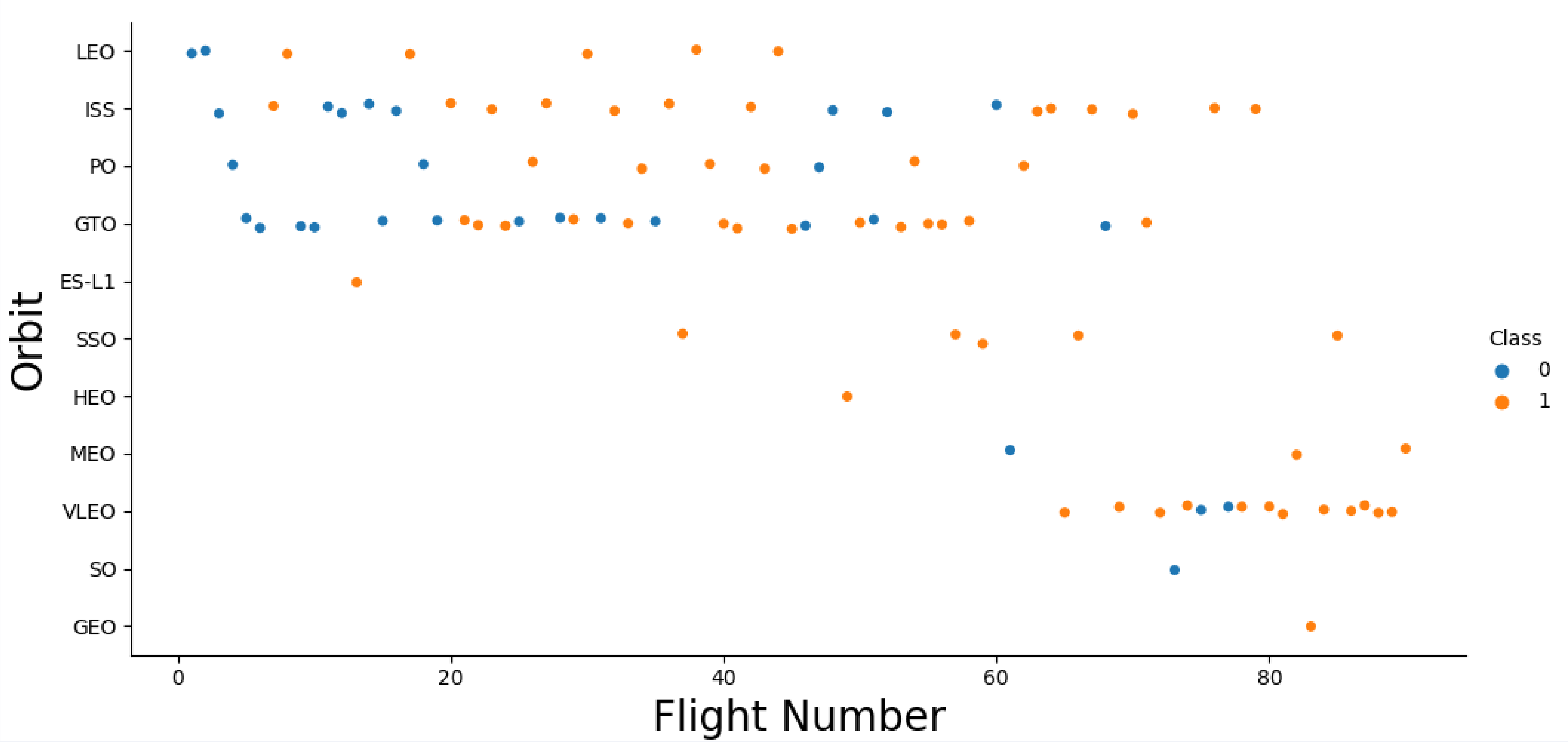
Payload vs. Launch Site



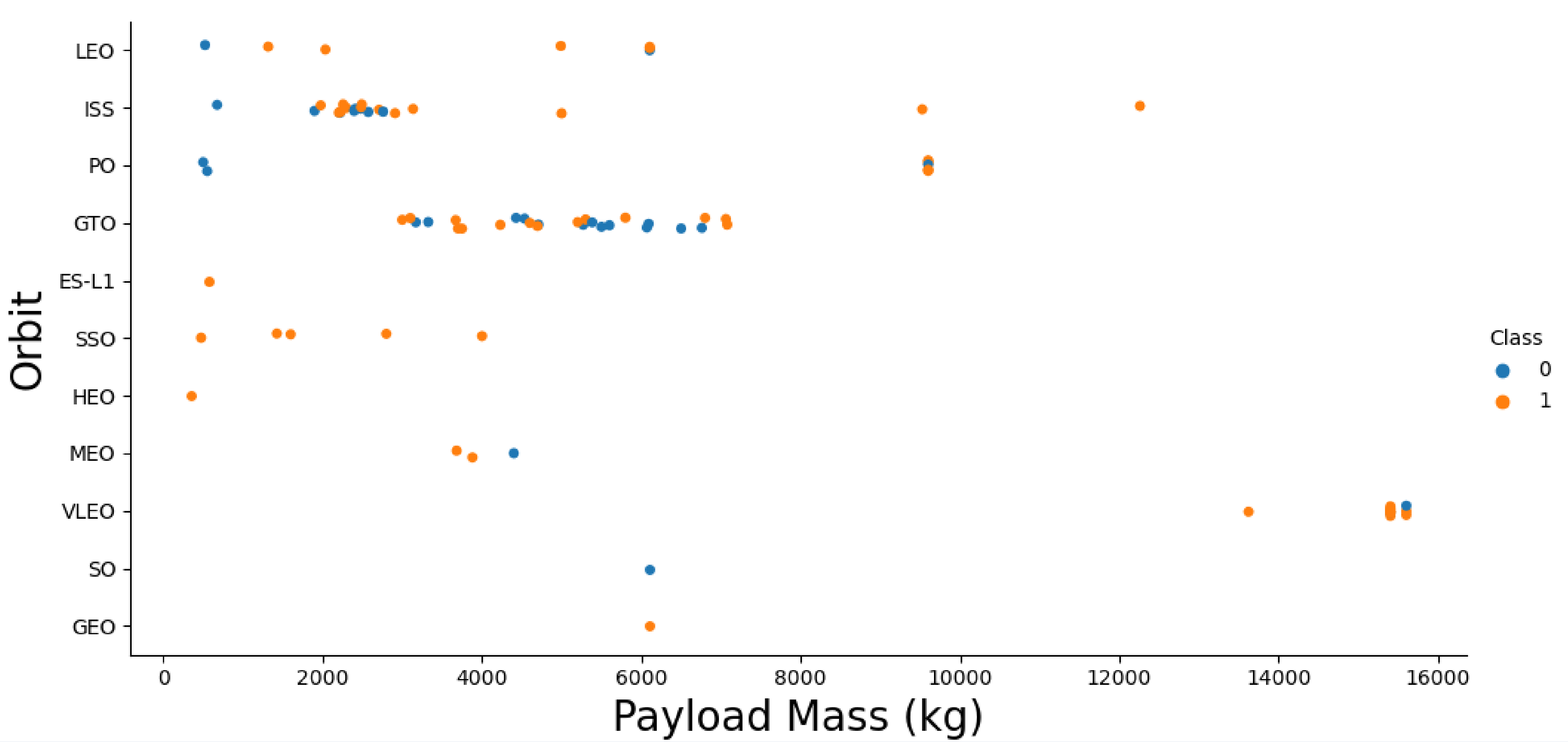
Success Rate vs. Orbit Type



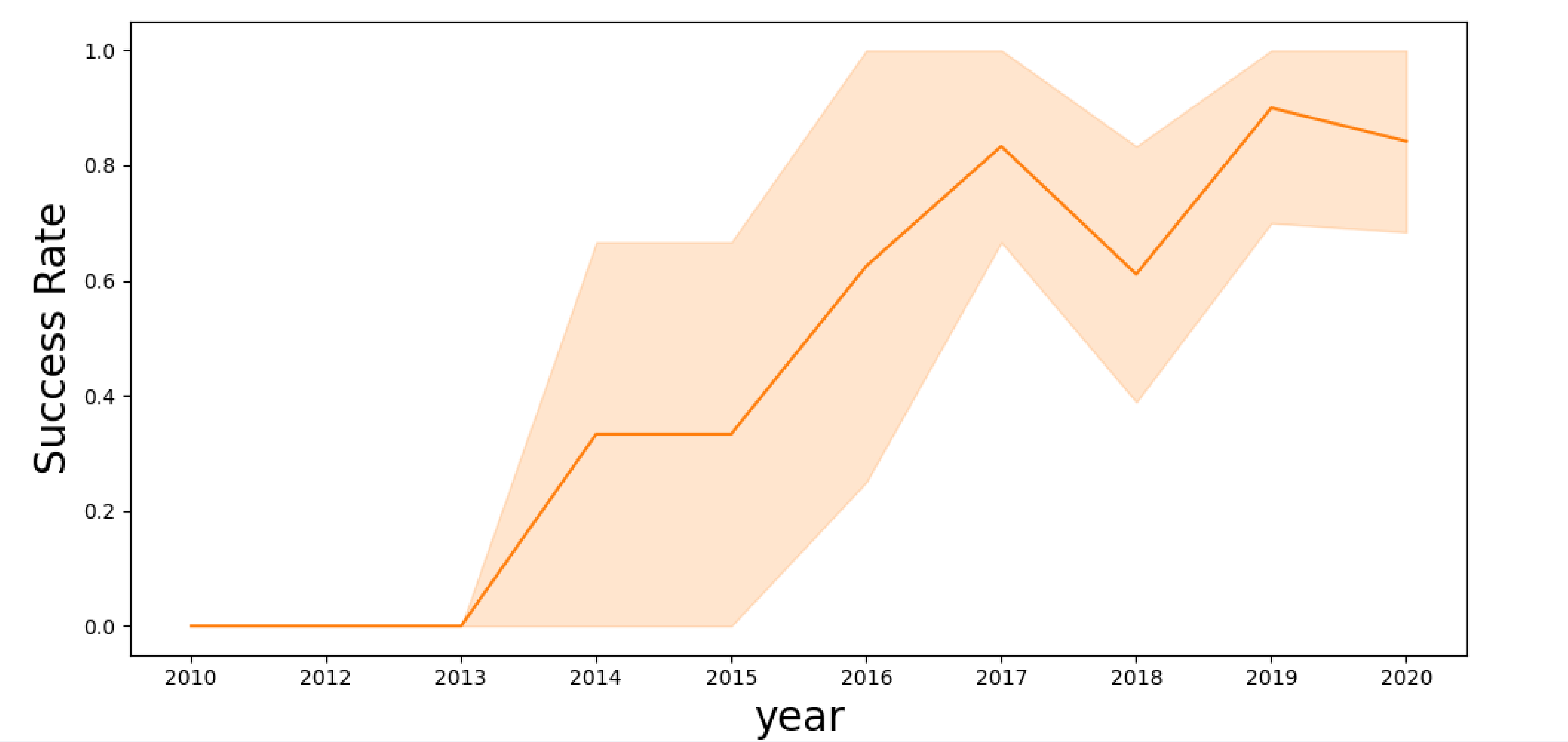
Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

```
[17]: %sql select distinct "Launch_Site" from SPACEXTABLE;

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

[17]:

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

Launch Site Names Begin with 'CCA'

```
[21]: %%sql
SELECT * from SPACEXTABLE where "Launch_Site" like "CCA%" limit 5
```

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

[21]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-08-12	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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

```
[22]: %%sql
select sum("PAYLOAD_MASS__KG_") from SPACEXTABLE where Customer = "NASA (CRS)"

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

[22]:

sum("PAYLOAD_MASS__KG_")
45596

Average Payload Mass by F9 v1.1

```
[23]: %%sql
select avg("PAYLOAD_MASS__KG_") from SPACEXTABLE where "Booster_Version" = "F9 v1.1"

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

[23]:

avg("PAYLOAD_MASS__KG_")
2928.4

First Successful Ground Landing Date

```
[33]: %%sql
select min(Date) from SPACEXTABLE where "Landing_Outcome" = "Success (ground pad)"

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

[33]:

min(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

[46]:
```
%%sql
select "Booster_Version" from SPACEXTABLE where "Landing_Outcome" = "Success (drone ship)"
AND "PAYLOAD_MASS__KG_" between 4000 and 6000
```

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

[46]:

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

Total Number of Successful and Failure Mission Outcomes

```
[62]: %%sql
      select count(MISSION_OUTCOME) from SPACEXTABLE GROUP BY MISSION_OUTCOME

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

[62]:

count(MISSION_OUTCOME)
1
98
1
1

Boosters Carried Maximum Payload

```
[63]: %%sql
select "Booster_Version" from SPACEXTABLE
where "PAYLOAD_MASS__KG_" = (select max("PAYLOAD_MASS__KG_") from SPACEXTABLE)

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

[63]:

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

In [38]:

```
%sql SELECT substr(Date,4,2) as month, DATE,BOOSTER_VERSION, LAUNCH_SITE, [Landing _Outcome] \
FROM SPACEXTBL \
where [Landing _Outcome] = 'Failure (drone ship)' and substr(Date,7,4)='2015';
```

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

Out[38]:

month	Date	Booster_Version	Launch_Site	Landing _Outcome
01	10-01-2015	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	14-04-2015	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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

```
[120]: %%sql
select Date, "Landing_Outcome", count("Landing_Outcome") as myCount from SPACEXTABLE
where Date between '2010-06-04' and '2017-03-20'
group by "Landing_Outcome" order by "myCount" desc
```

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

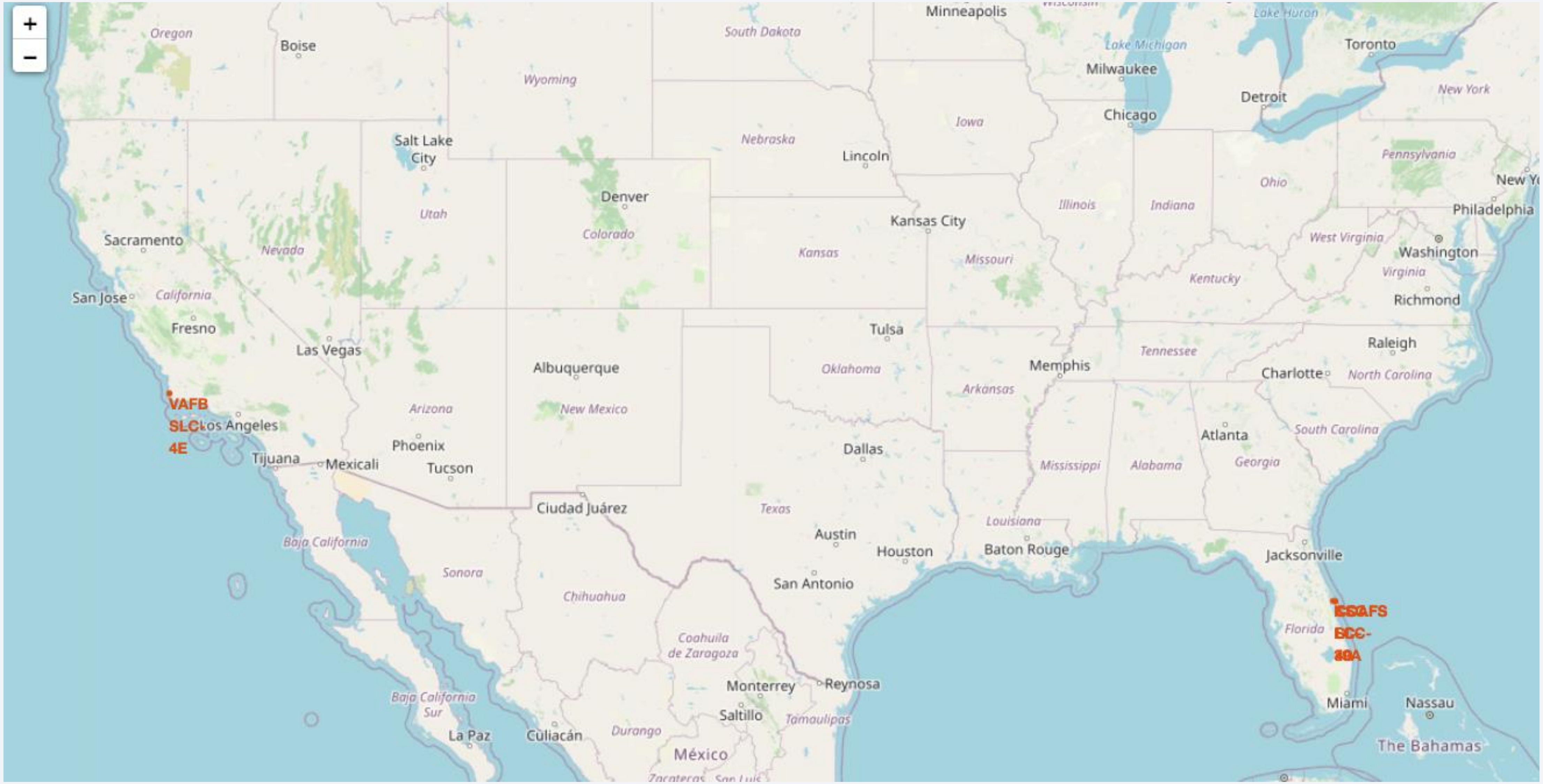
[120]:

Date	Landing_Outcome	myCount
2012-05-22	No attempt	10
2015-12-22	Success (ground pad)	5
2016-08-04	Success (drone ship)	5
2015-10-01	Failure (drone ship)	5
2014-04-18	Controlled (ocean)	3
2013-09-29	Uncontrolled (ocean)	2
2015-06-28	Precluded (drone ship)	1
2010-08-12	Failure (parachute)	1

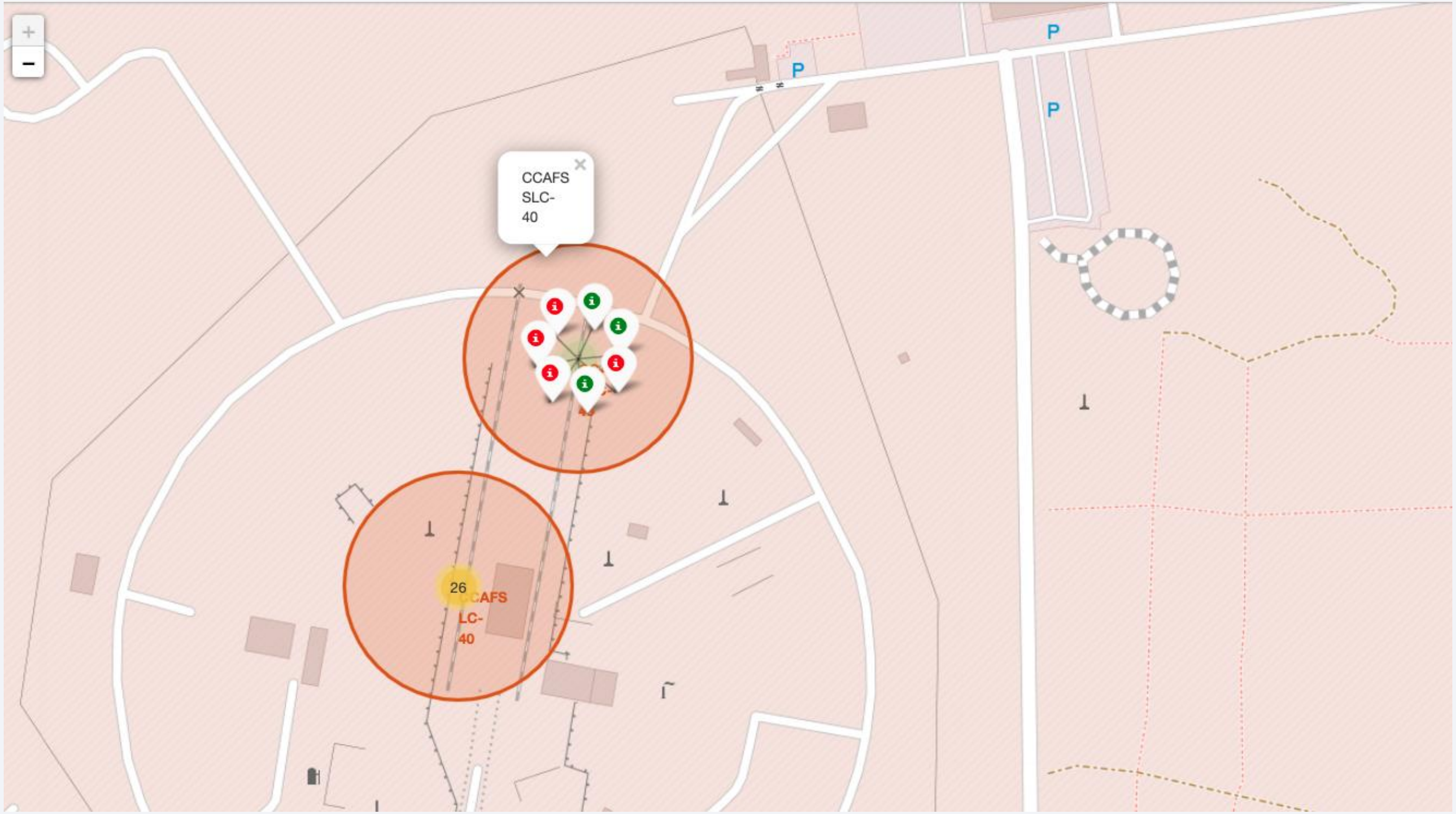
Section 3

Launch Sites Proximities Analysis

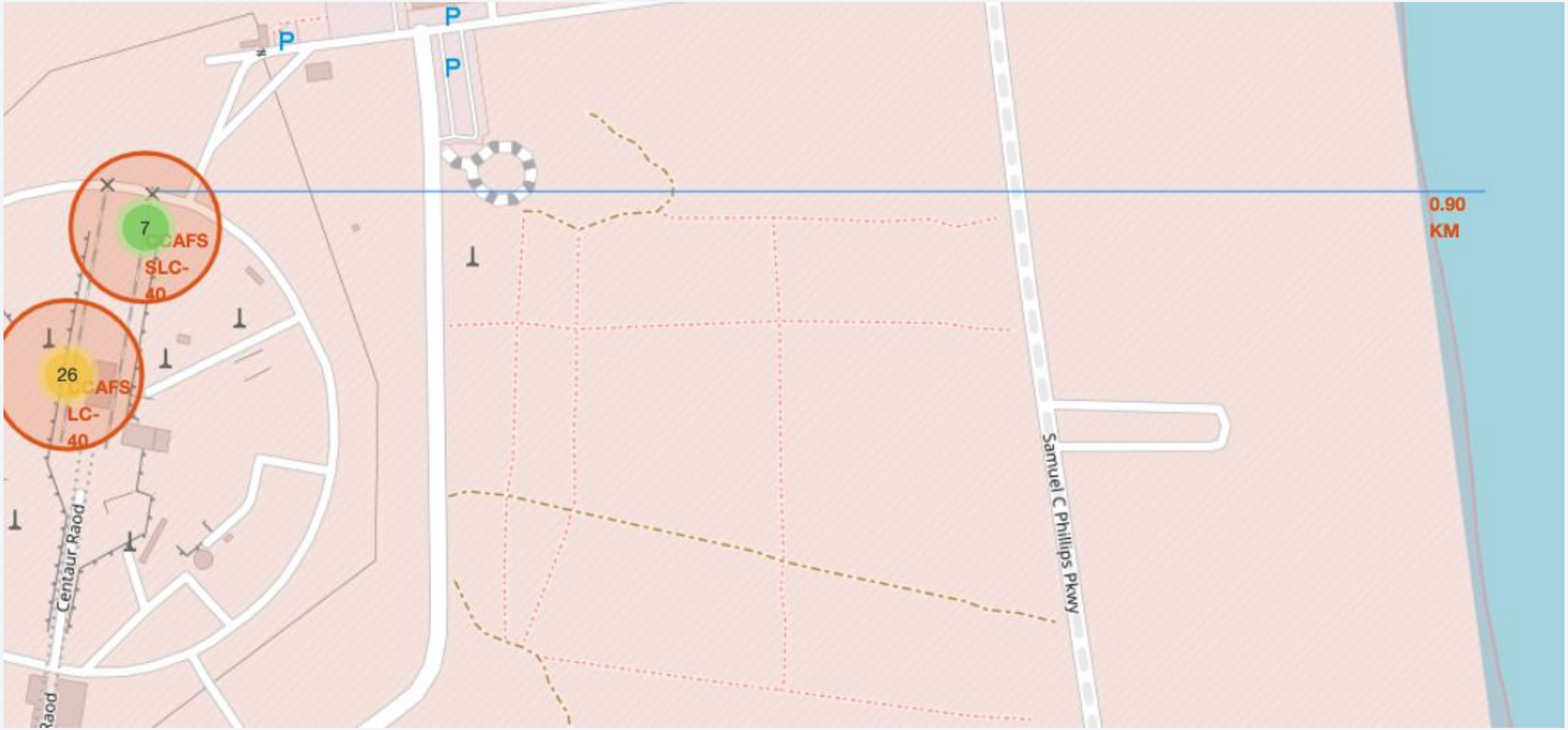
Global Map with all Launch Sites



Color Labeled Launch Outcomes



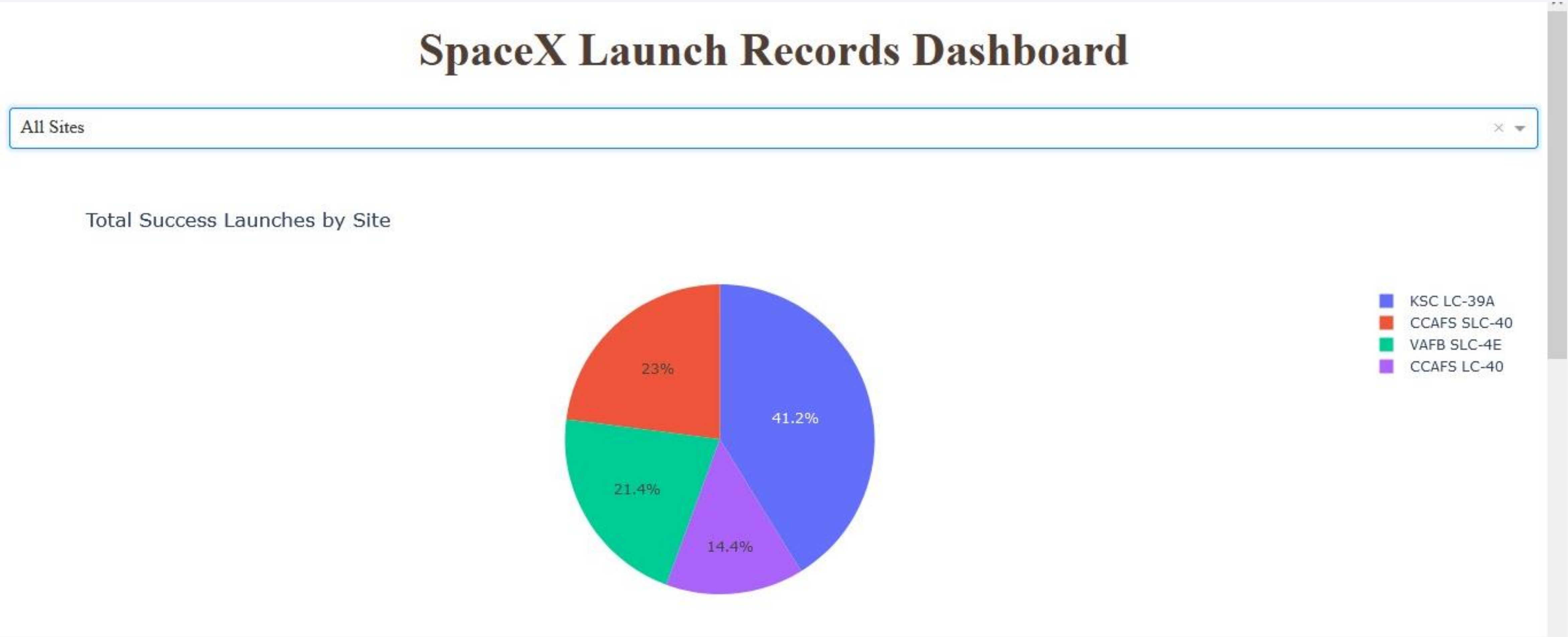
Map with Proximities



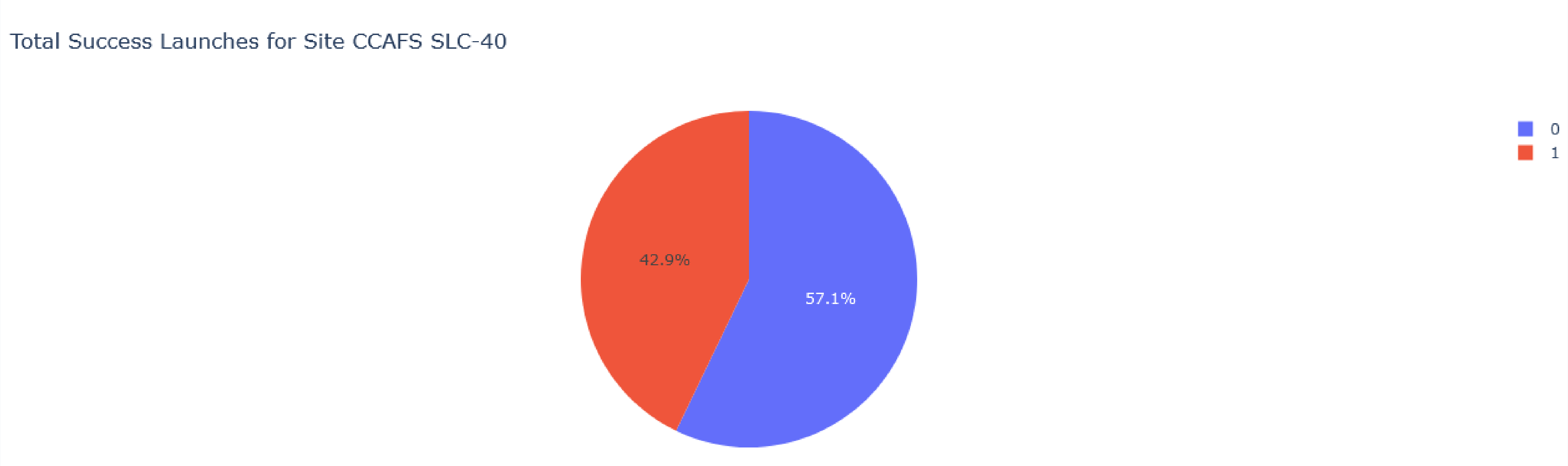
Section 4

Build a Dashboard with Plotly Dash

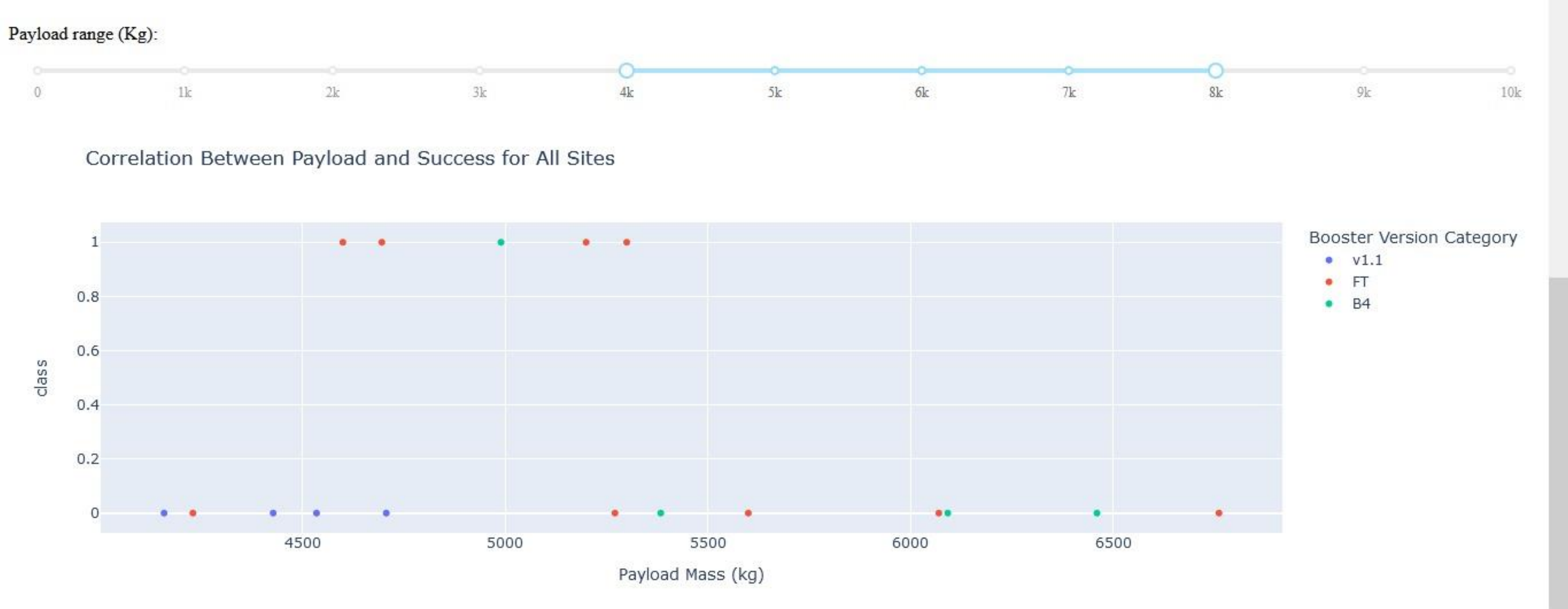
Launch Success Count for All Sites



Launch Site with Highest Launch Success Ratio



Payload vs. Launch Outcome with Range Slider



Section 5

Predictive Analysis (Classification)

Classification Accuracy

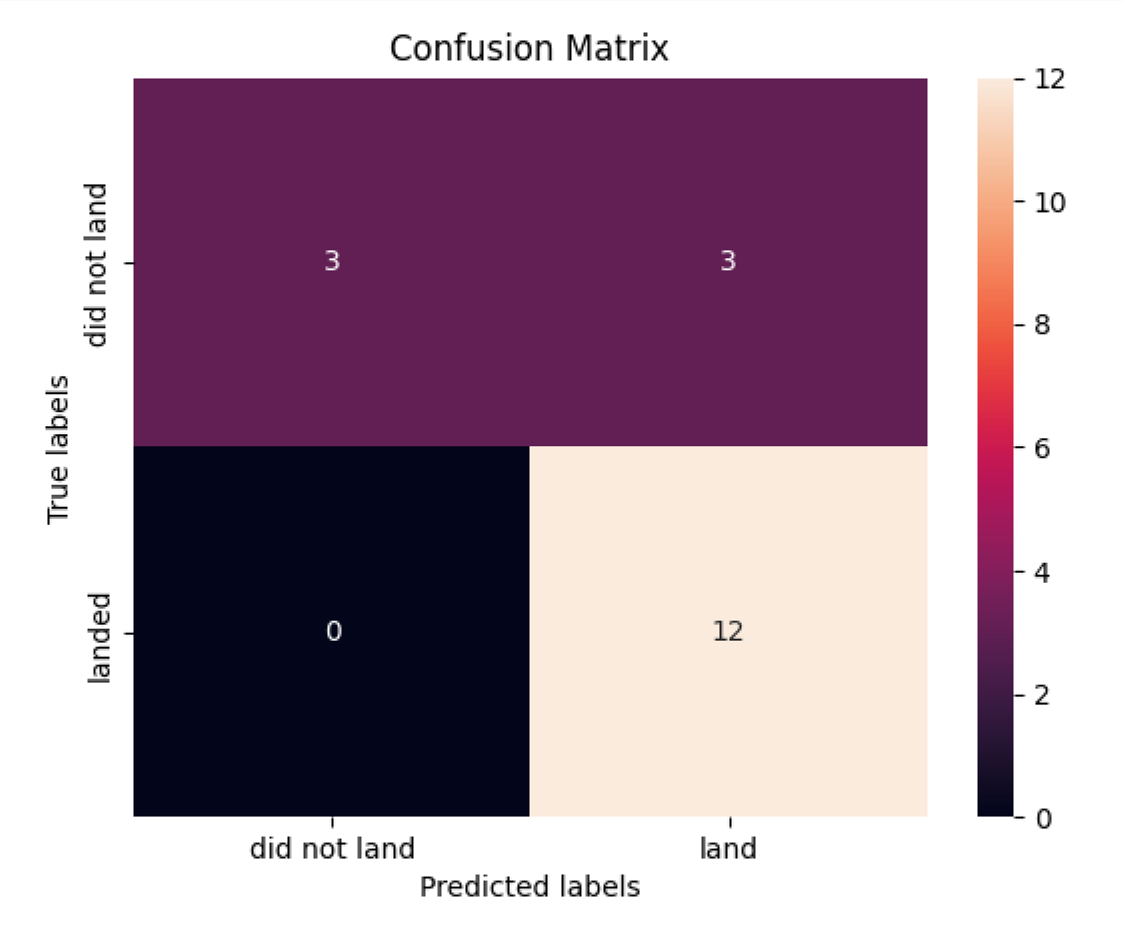
In [30]:

```
accuracy = [svm_cv_score, logreg_score, knn_cv_score, tree_cv_score]
accuracy = [i * 100 for i in accuracy]

method = ['Support Vector Machine', 'Logistic Regression', 'K Nearest Neighbour', 'Decision Tree']
models = {'ML Method':method, 'Accuracy Score (%)':accuracy}

ML_df = pd.DataFrame(models)
ML_df
```

Out[30]:

	ML Method	Accuracy Score (%)
0	Support Vector Machine	83.333333
1	Logistic Regression	83.333333
2	K Nearest Neighbour	83.333333
3	Decision Tree	83.333333

Confusion Matrix



Conclusions

- Model Performance : The models performed similarly on the test set with the decision tree model slightly outperforming
- Equator : Most of the launch sites are near the equator for an additional natural boost due to the rotational speed of earth whic h helps save the cost of putting in extra fuel and boosters
- Coast : All the launch sites are close to the coast
- Launch Success : Increases over time • KSC LC 39A : Has the highest success rate among launch sites. Has a 100% success rate for launches less than 5,500 kg
- Orbits : ES L1, GEO, HEO, and SSO have a 100% success rate
- Payload Mass : Across all launch sites, the higher the payload mass (kg), the higher the success rate

Thank you!