IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Vaibhav Mishra
05/02/2022

Outline

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

Executive Summary

The purpose of this presentation is to provide insight on SpaceX launch. Data are obtained through web scrapping from trusted and open source. Then, data is presented in plots, maps and dashboard. Last, by using prediction technique such as logistic regression, support vector machine, decision tree and K nearest neighbors.

Through research, we found most of the launch sites near the coast. Among all launch sites, KSC LC-39A had the highest success rate. We are able top use prediction technique to predict the outcome based on features such as orbit, payload mass and booster version. The accuracy of the prediction is as high as 83%.

Introduction

- To provide insight of SpaceX launch
- To predict outcome of SpaceX launch based on multiple features
- To reduce the cost by visualizing which features affect the success rate the most

Section 1

Methodology

Methodology

Executive Summary

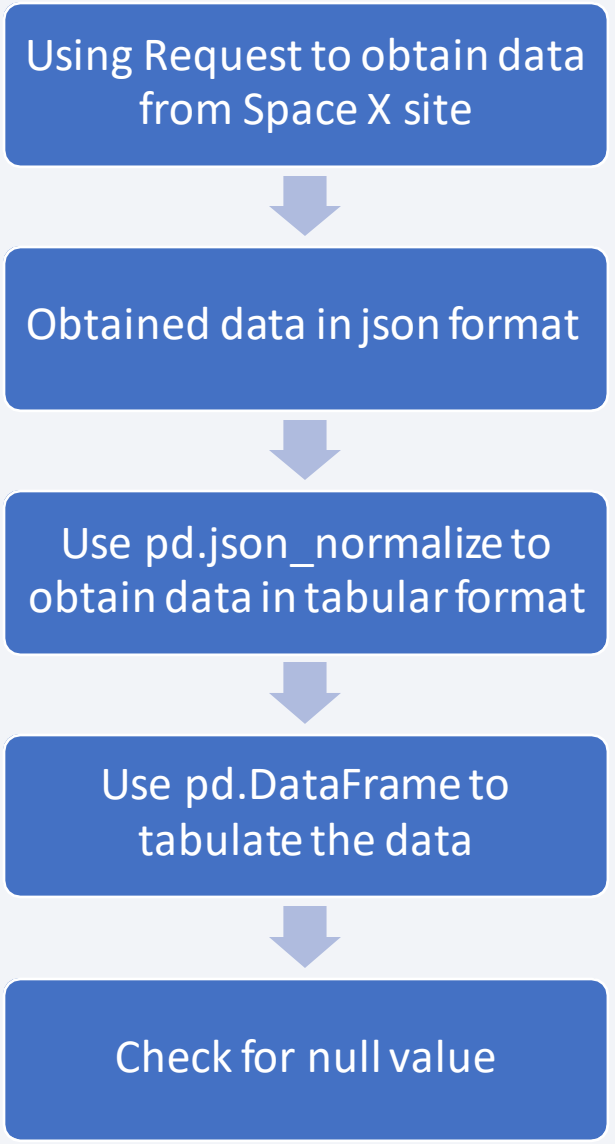
- Data collection methodology:
  - Using Pandas, BeautifulSoup, and other libraries to collect data from trusted and open source online
- Perform data wrangling
  - Check and replace null value, then perform one hot encoding
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Using GridSearchCV to get best hyperparameters for prediction technique, then use SCORE, and confusion matrix to check for accuracy

Data Collection

- By using Pandas library to access data in .csv format. The .csv file is requested from the internet using Request
- Then, obtain additional information by using BeautifulSoap library to perfom web scraping. The additional information is obtained from Wikipedia.

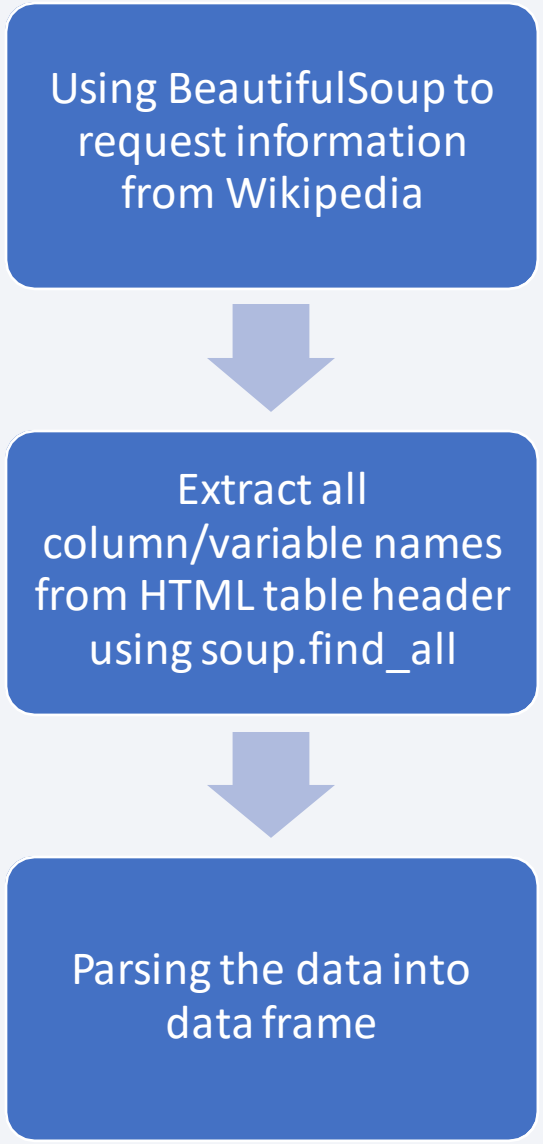
Data Collection - SpaceX API

- Github reference link: https://github.com/VaibhavMishra30/IBM-Data-Science-Capstone-Project/blob/main/Data%20Collection.ipynb



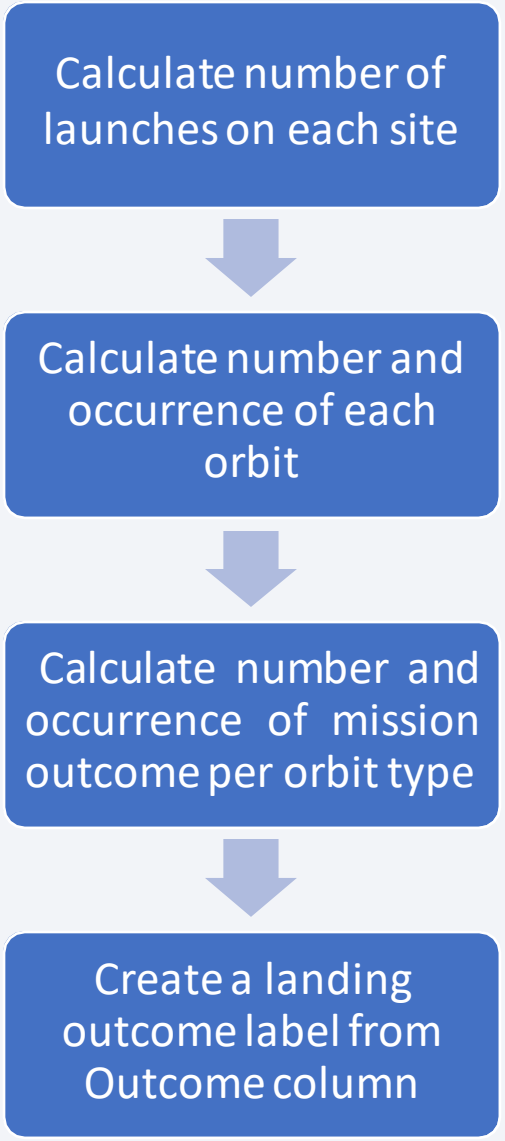
Data Collection - Scraping

- Github reference link: https://github.com/VaibhavMishra30/IBM-Data-Science-Capstone-Project/blob/main/Data%20Collection%20by%20Web%20Scraping.ipynb



Data Wrangling

- Github reference link: https://github.com/VaibhavMishra30/IBM-Data-Science-Capstone-Project/blob/main/Data%20Wrangling.ipynb



EDA with Data Visualization

- Plotted charts:
  - Payload Mass (kg) vs Launch Site
  - Launch Site vs Flight Number
  - Success rate of each orbit type
  - Orbit Type vs Flight Number
  - Payload Mass (kg) vs Orbit Type
  - Launch success yearly trend
- Using pd.get_dummies to create dummy variables to categorical columns
- Github reference link: https://github.com/VaibhavMishra30/IBM-Data-Science-Capstone-Project/blob/main/EDA%20with%20Data%20Visualization.ipynb

EDA with SQL

- Using SQL query function, find:
  - Unique launch sites in the space mission
  - 5 Records where launch sites begin with the string 'CCA'
  - Total payload mass carried by boosters launched by NASA (CRS)
  - Average payload mass carried by booster version F9 v1.1
  - Date when the first successful landing outcome in ground pad was achieved
  - Names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
  - Total number of successful and failure mission outcomes
  - Names of the booster_versions which have carried the maximum payload mass
  - Failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
  - 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
- Github reference link: https://github.com/VaibhavMishra30/IBM-Data-Science-Capstone-Project/blob/main/EDA%20with%20SQL.ipynb

Build an Interactive Map with Folium

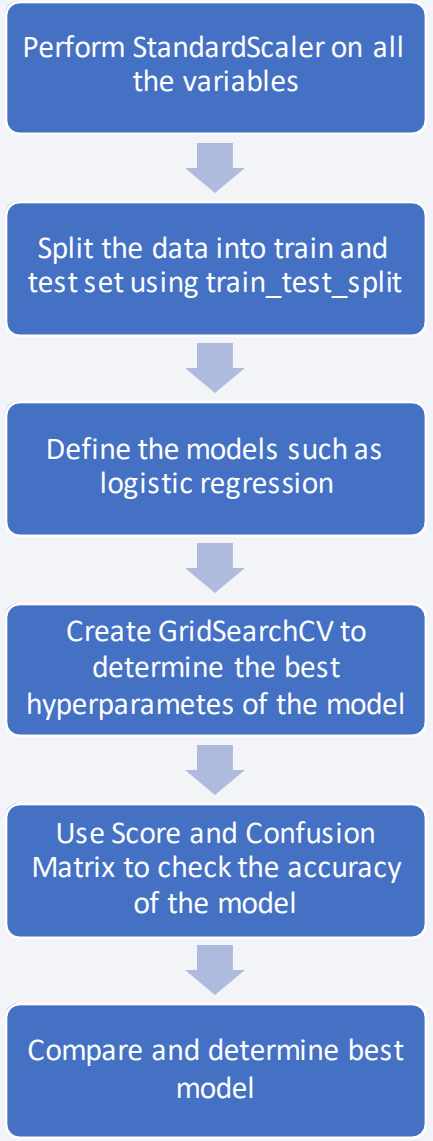
- Mark each launch site on the map, then mark the success / failed launches on the map.
- Calculate the distance between one launch site to its proximities
- Found that normally launch site is nearer to the coast than to the dense population area
- Github reference link: https://github.com/VaibhavMishra30/IBM-Data-Science-Capstone-Project/blob/main/Interactive%20Visual%20Analytics%20with%20Folium%20lab.ipynb

Build a Dashboard with Plotly Dash

- Added interactive dashboard showing the success rate of all launch site, the highest success rate of launch site, and Interactive Plot of Success Rate, Booster Version, and Payload Mass (kg)
- This help user to understand more on the insight
- Github reference link: https://github.com/VaibhavMishra30/IBM-Data-Science-Capstone-Project/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

- Go through model such as logistic regression, decision tree, support vector machine, and K-Nearest neighbors.
- Found that the accuracy for these 4 models are practically the same.
- Github reference link: https://github.com/VaibhavMishra30/IBM-Data-Science-Capstone-Project/blob/main/Machine%20Learning%20Prediction%20lab.ipynb



Results

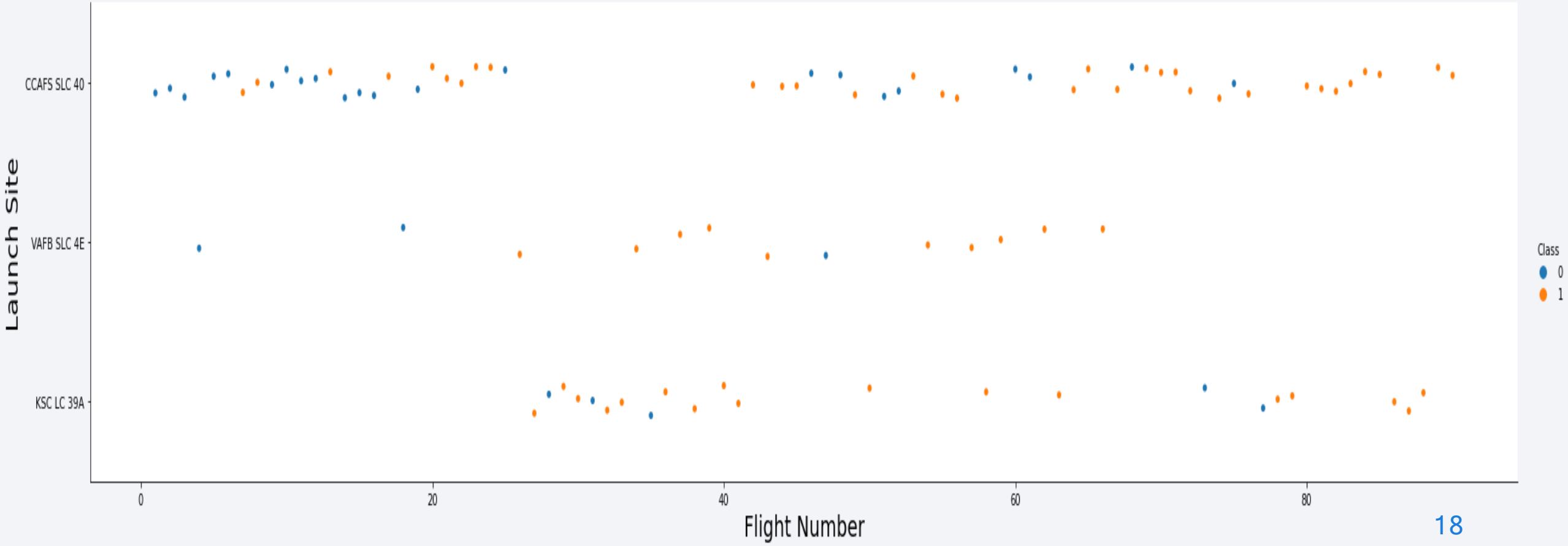
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

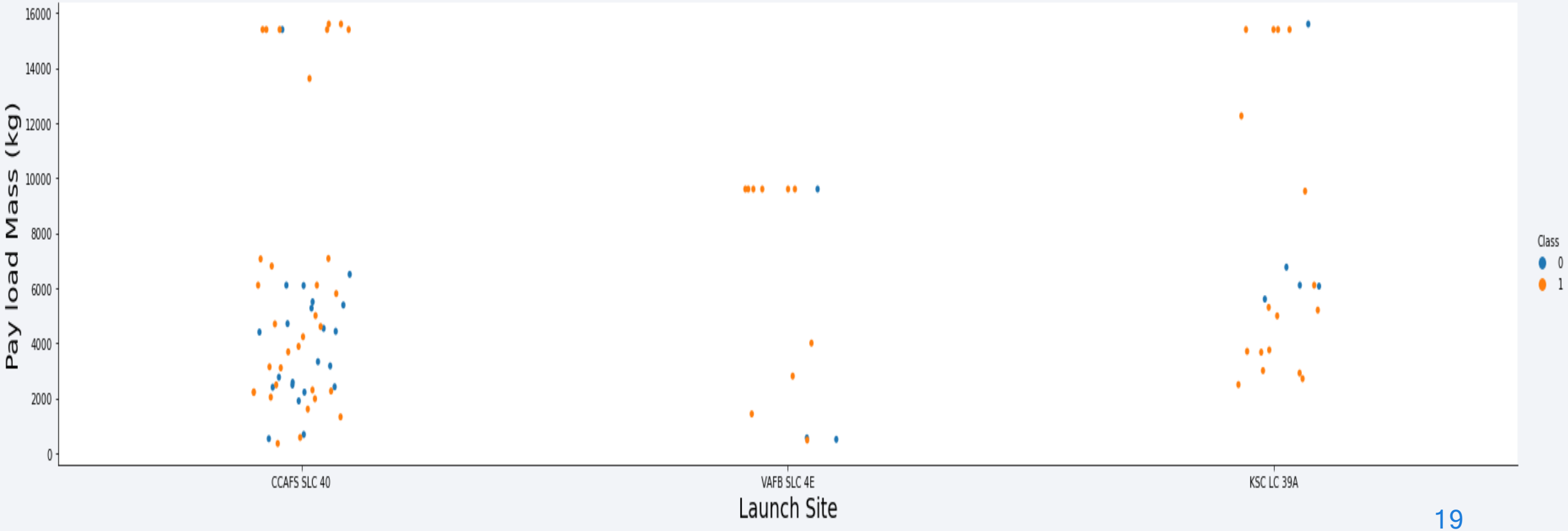
Flight Number vs. Launch Site

Most launches happen at CCAFS SLC 40 launch site



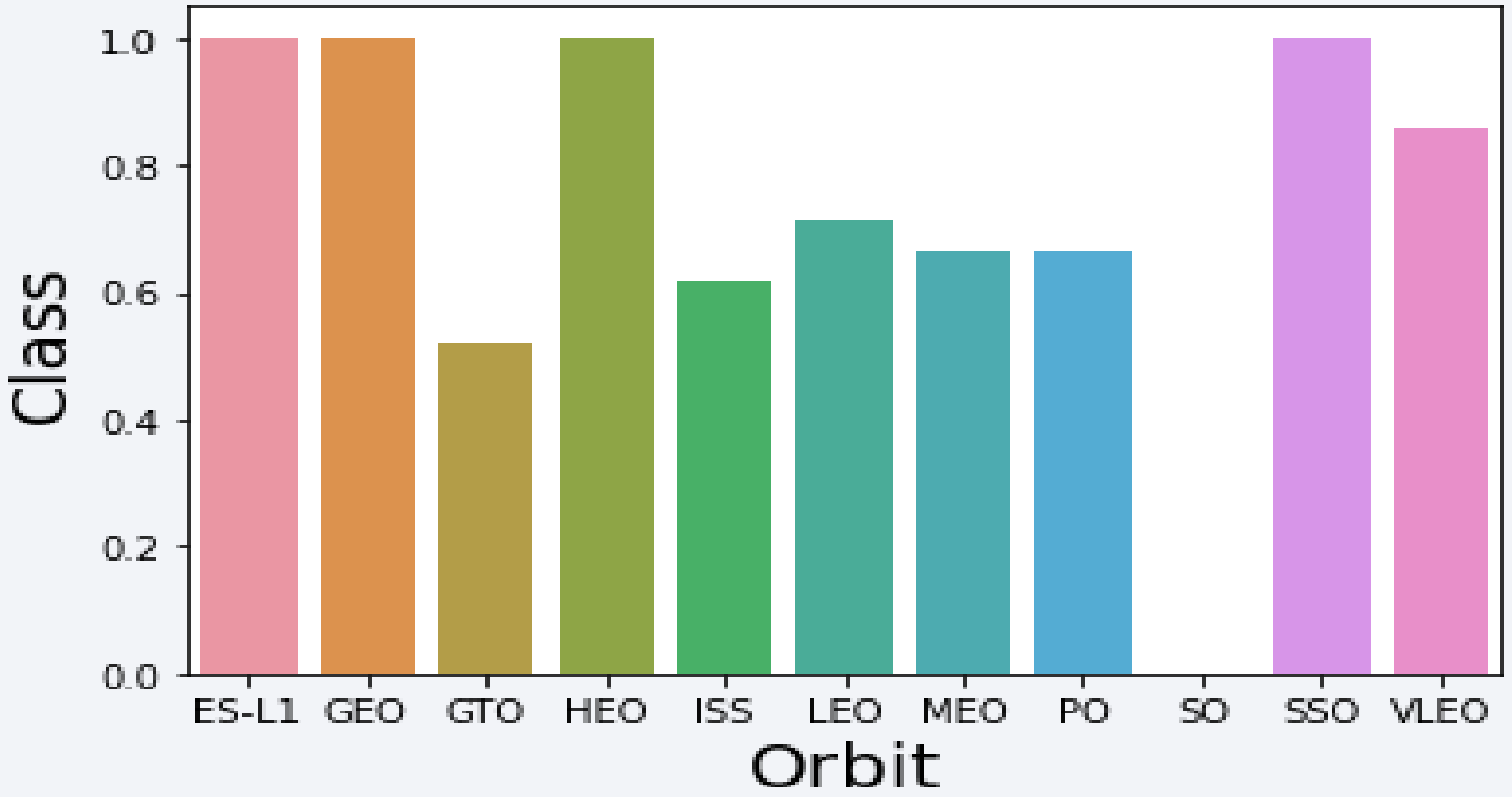
Payload vs. Launch Site

1. Launch site CCAFS SLC 40 carry most payload mass within 0kg to 8000kg
2. No 10000kg and above payload happen at VAFB SLC 4E launch site



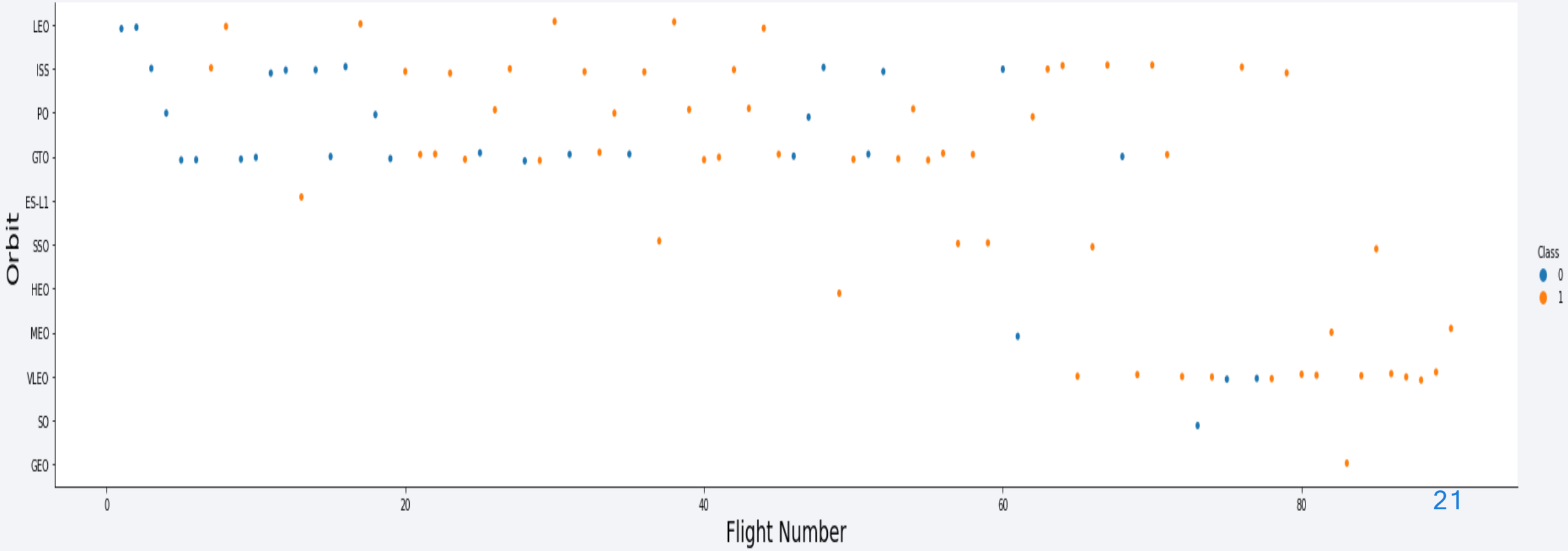
Success Rate vs. Orbit Type

Orbit ES-L1, GEO, HEO, SSO have the highest success rate while orbit SO has the lowest success rate



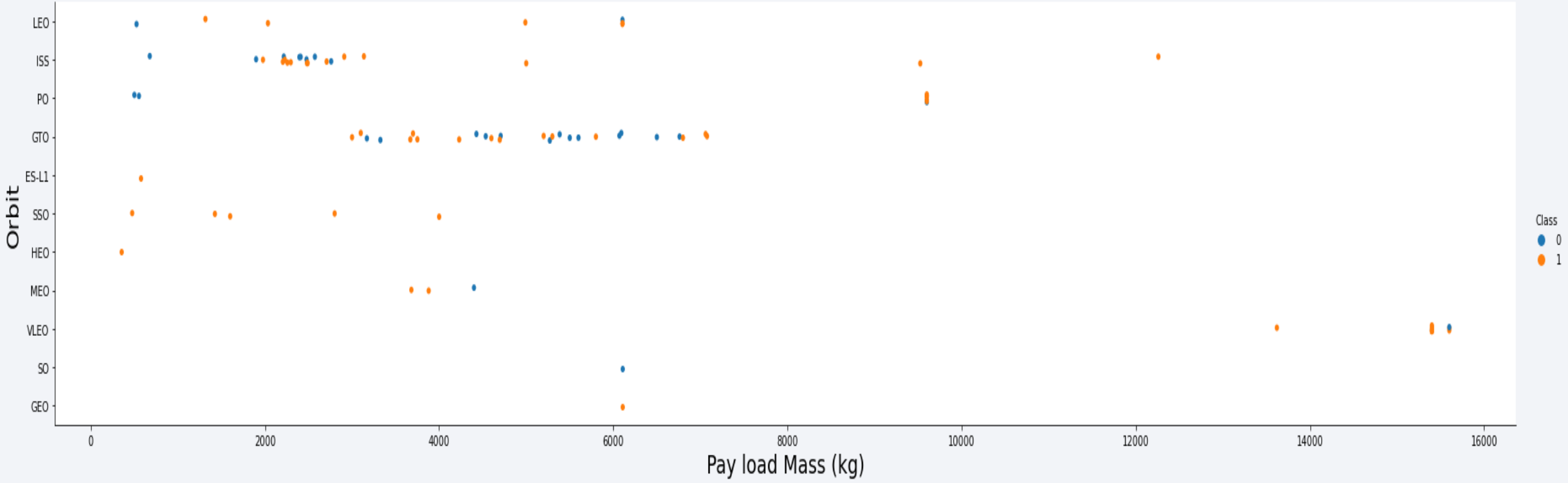
Flight Number vs. Orbit Type

Orbit VLEO has most successful rate, but most of the flights happen at orbit ISS and GTO



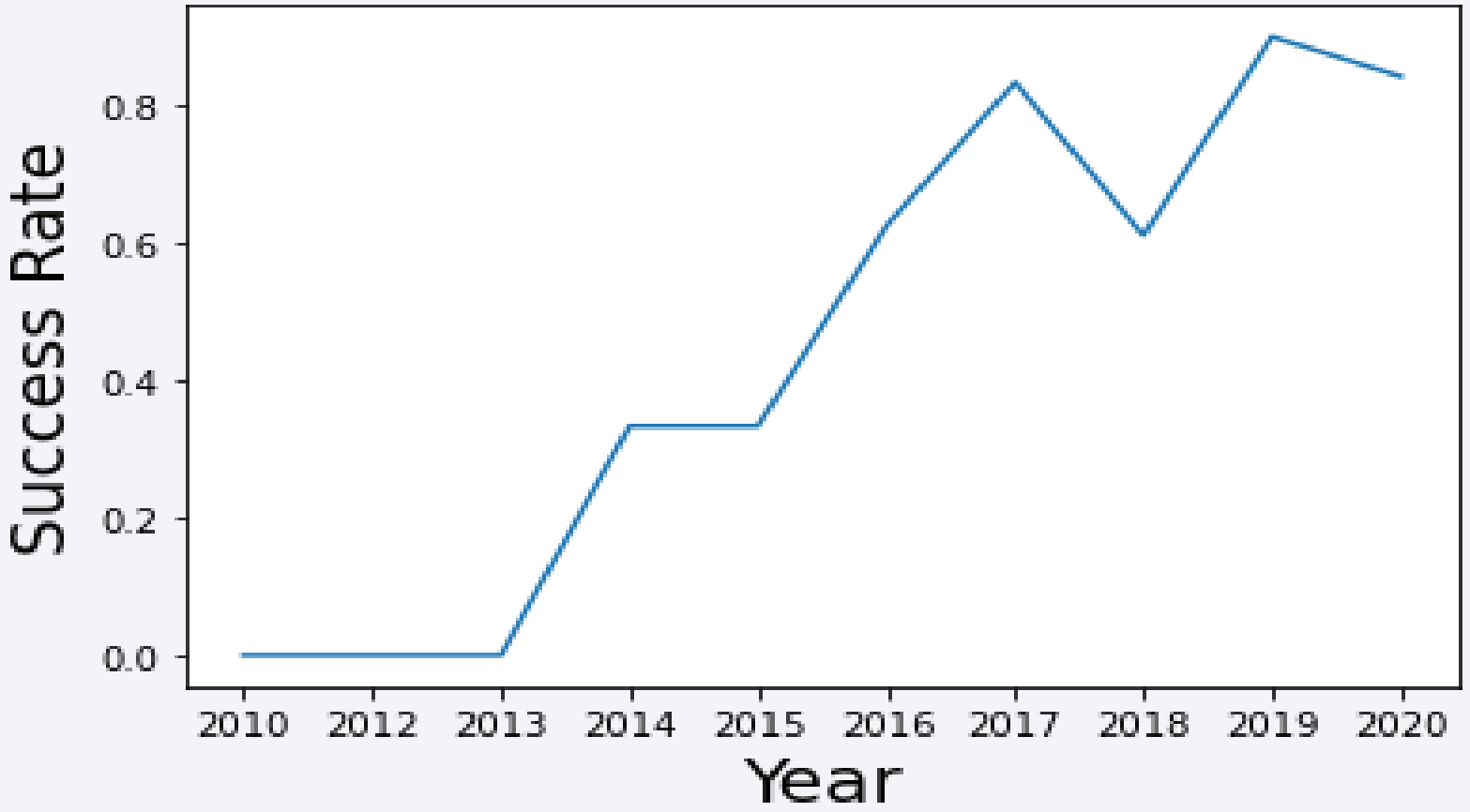
Payload vs. Orbit Type

1. Most common payload mass is within 2000kg to 6000kg.
2. Most payload mass above 10000kg happen in orbit VLEO



Launch Success Yearly Trend

Launch success rate increase every year with a slight decrease happen in year 2018 and 2020



All Launch Site Names

Total 4 launch sites available as a result of query as shown below

In [8]:
```
%sql select distinct(launch_site) from SPACEXTBL
```

```
 * ibm_db_sa://scy22371:***@1bbf73c5-d84a-4bb0-85b9-ab1a4348f4a4.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:32286/bludb
Done.
```

Out[8]:

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

Launch Site Names Begin with 'CCA'

5 Records where launch sites begin with CCA

In [16]:
```
%%sql
select * from SPACEXTBL
where launch_site like 'CCA%'
limit 5
```

```
 * ibm_db_sa://scy22371:***@1bbf73c5-d84a-4bb0-85b9-ab1a4348f4a4.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:32286/bludb
Done.
```

Out[16]:

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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	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

Total payload carried is 45596kg!

```
%%sql
select sum(payload_mass__kg_) from SPACEXTBL
where customer='NASA (CRS)'
```

```
 * ibm_db_sa://scy22371:***@1bbf73c5-d84a-4bb0-85b9-ab1a4348f4a4.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:32286/bludb
Done.
```

1
45596

Average Payload Mass by F9 v1.1

The average payload mass by F9 v1.1 is 2928kg

In [26]:
```
%%sql
select avg(payload_mass__kg_) from SPACEXTBL
where booster_version='F9 v1.1'
```

```
 * ibm_db_sa://scy22371:***@1bbf73c5-d84a-4bb0-85b9-ab1a4348f4a4.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:32286/bludb
Done.
```

Out[26]:

1
2928

First Successful Ground Landing Date

The first successful landing on the ground pad is on 2015-12-22!

In [30]:
```
%%sql
select min(DATE) from SPACEXTBL
where landing__outcome = 'Success (ground pad)'
```

```
 * ibm_db_sa://scy22371:***@1bbf73c5-d84a-4bb0-85b9-ab1a4348f4a4.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:32286/bludb
Done.
```

Out[30]:

1
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

There are total of 4 booster version carrying payload mass greater than 4000 but less than 6000kg

In [36]:
```
%%sql
select booster_version from SPACEXTBL
where landing__outcome = 'Success (drone ship)' and payload_mass__kg_ between 4000 and 6000
```

```
 * ibm_db_sa://scy22371:***@1bbf73c5-d84a-4bb0-85b9-ab1a4348f4a4.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:32286/bludb
Done.
```

Out[36]:

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

Total Number of Successful and Failure Mission Outcomes

Total number of successful and failure mission outcomes is 101

In [46]:
```
%%sql
select count(mission_outcome) from SPACEXTBL
```

```
 * ibm_db_sa://scy22371:***@1bbf73c5-d84a-4bb0-85b9-ab1a4348f4a4.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:32286/bludb
Done.
```

Out[46]:

1
101

Boosters Carried Maximum Payload

Below show all the booster version carried maximum payload

In [47]:
```
%%sql
select booster_version from SPACEXTBL
where payload_mass__kg_ = (select max(payload_mass__kg_) from SPACEXTBL)
```

```
 * ibm_db_sa://scy22371:***@1bbf73c5-d84a-4bb0-85b9-ab1a4348f4a4.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:32286/bludb
Done.
```

Out[47]:

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

Below show fail landing on drone ship in 2015

In [49]:
```
%%sql
select DATE,landing__outcome,booster_version,launch_site from SPACEXTBL
where landing__outcome = 'Failure (drone ship)' and DATE like '2015%'
```

```
 * ibm_db_sa://scy22371:***@1bbf73c5-d84a-4bb0-85b9-ab1a4348f4a4.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:32286/bludb
Done.
```

Out[49]:

DATE	landing__outcome	booster_version	launch_site
2015-01-10	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
2015-04-14	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

In [53]:
```
%%sql
select DATE,landing__outcome from SPACEXTBL
where DATE between '2010-06-04' and '2017-03-20'
order by DATE desc
```

```
 * ibm_db_sa://scy22371:***@1bbf73c5-d84a-4bb0-85b9-ab1a4348f4a4.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:32286/bludb
Done.
```

Out[53]:

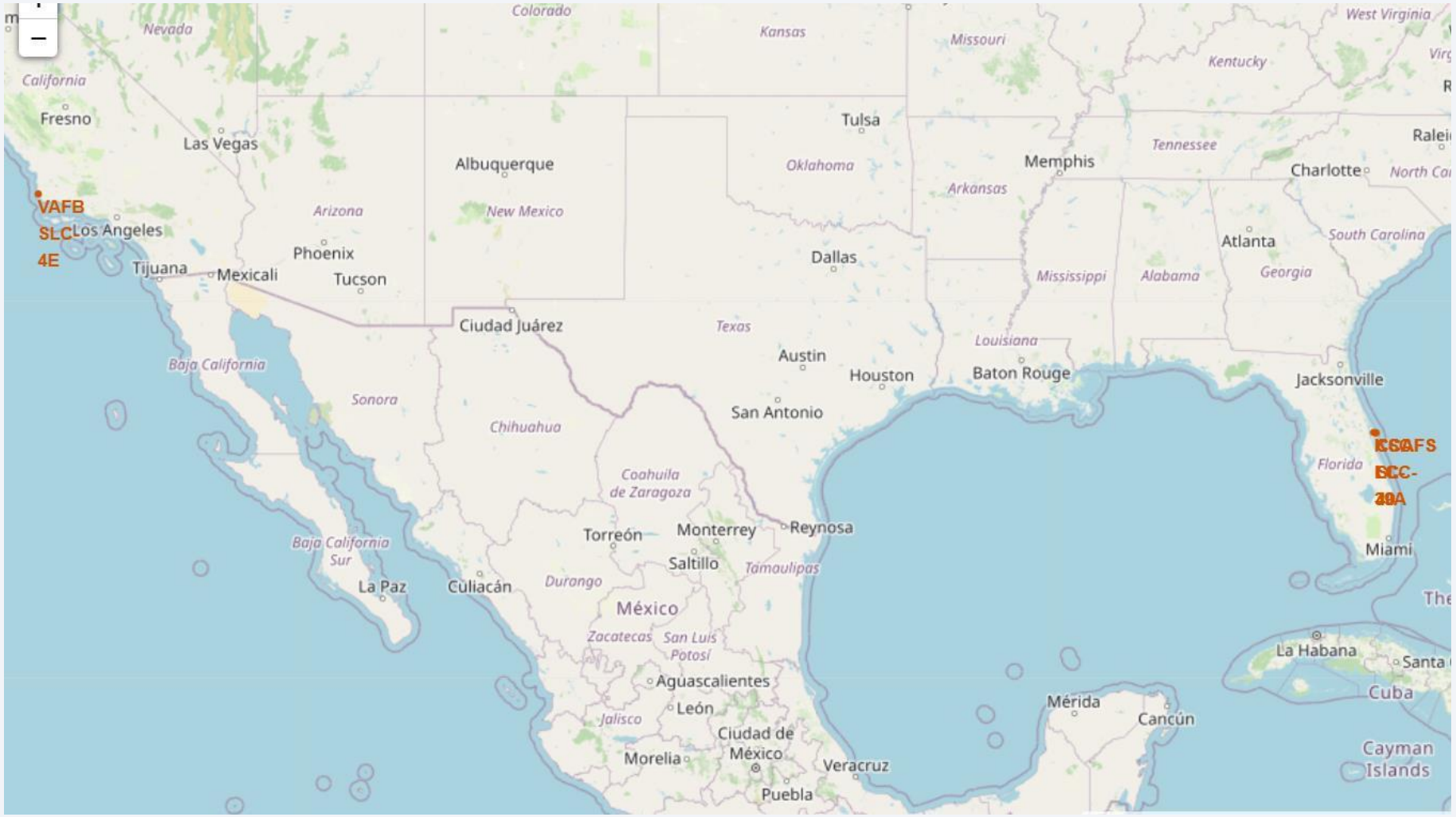
DATE	landing__outcome
2017-03-16	No attempt
2017-02-19	Success (ground pad)
2017-01-14	Success (drone ship)
2016-08-14	Success (drone ship)
2016-07-18	Success (ground pad)
2016-06-15	Failure (drone ship)
2016-05-27	Success (drone ship)
2016-05-06	Success (drone ship)
2016-04-08	Success (drone ship)

Section 4

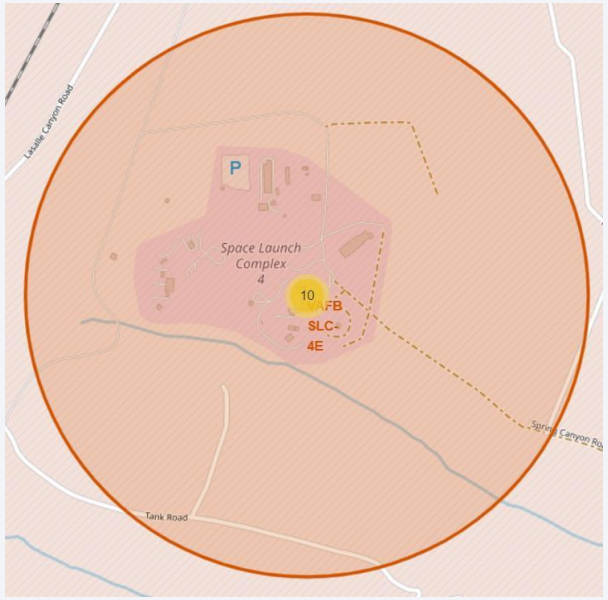
Launch Sites Proximities Analysis

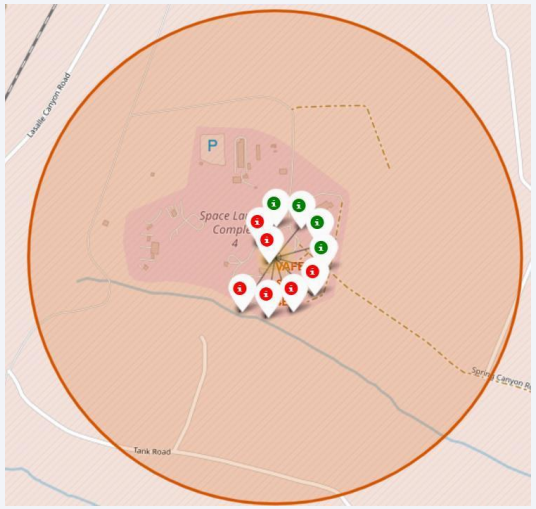
<Folium Map Screenshot 1>

All the launch sites are near to the coast.



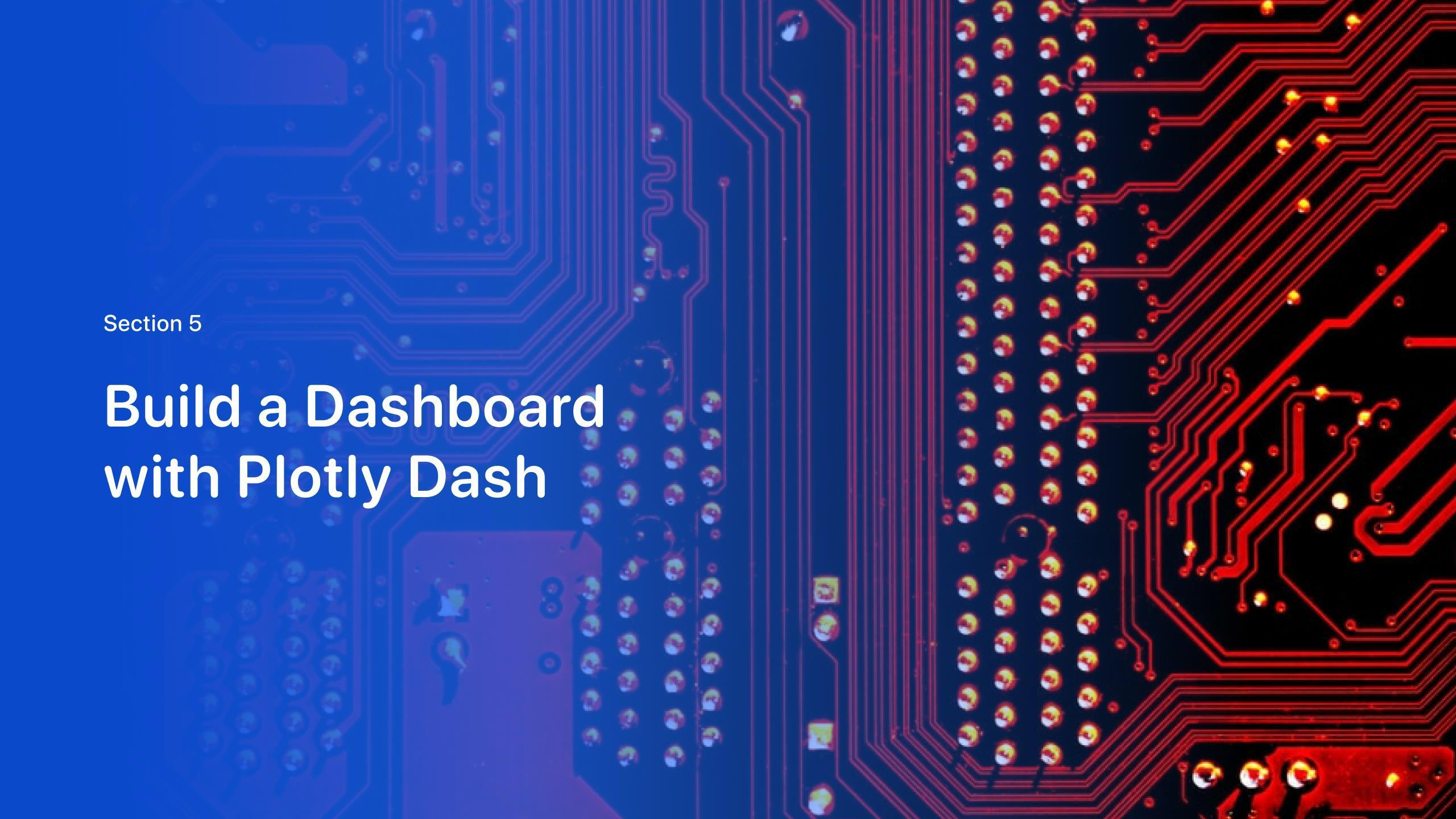
<Folium Map Screenshot 2>





<Folium Map Screenshot 3>

- All the launch site close to the coast and further from dense population to avoid unnecessary damage

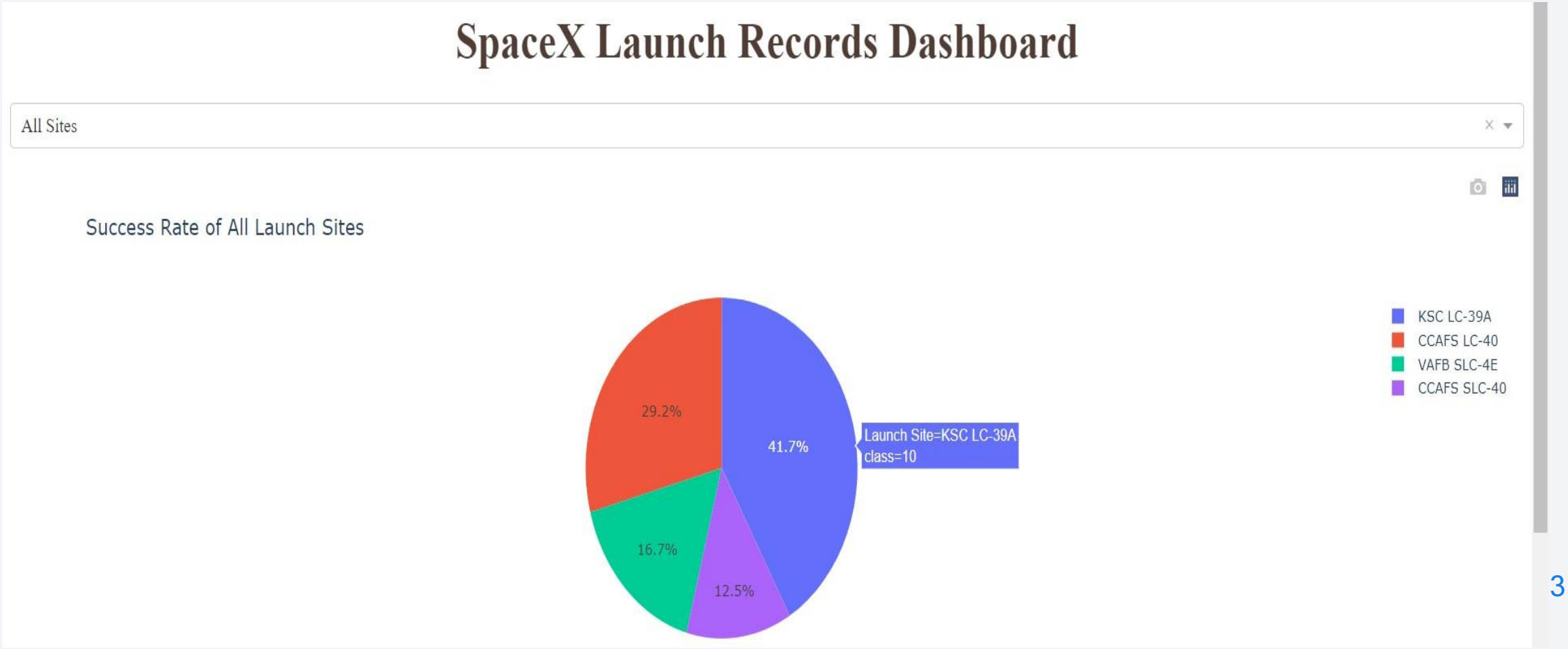


Section 5

Build a Dashboard with Plotly Dash

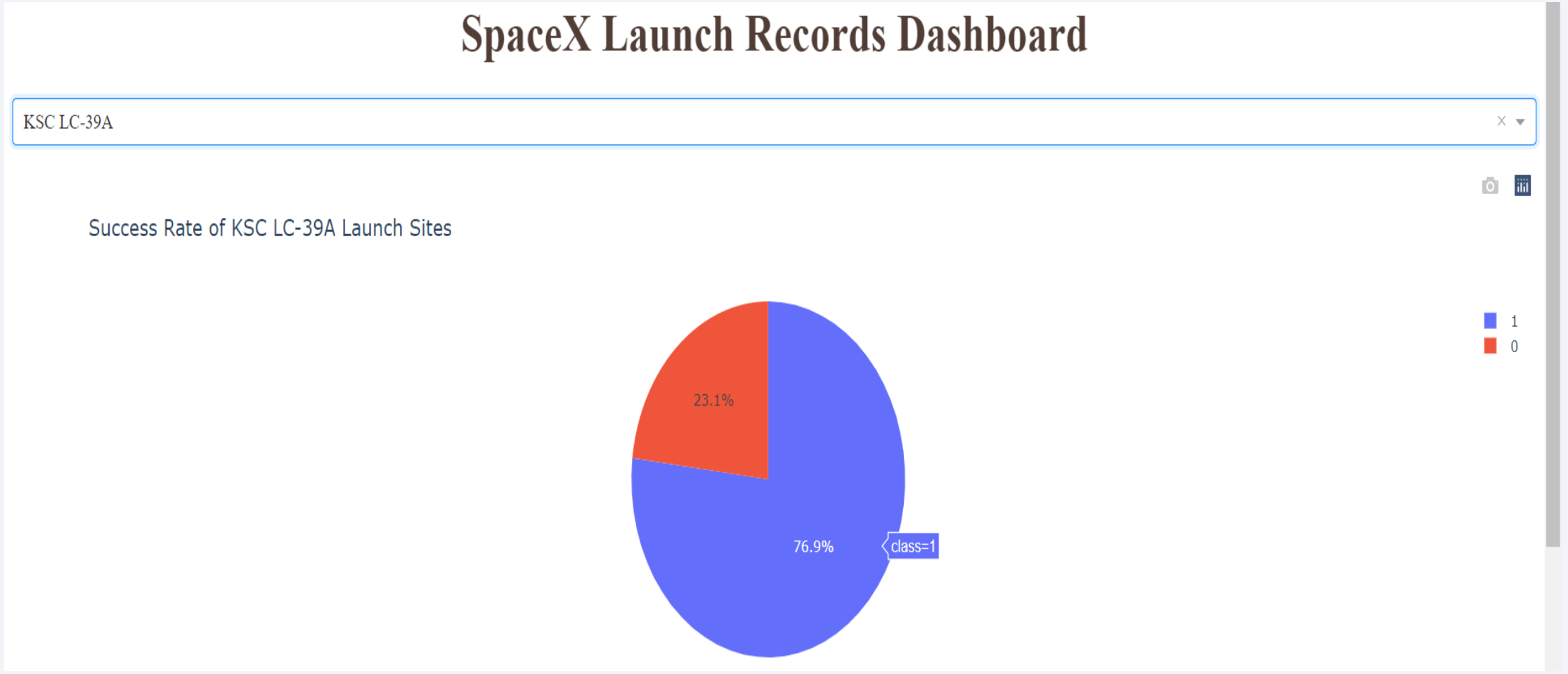
Success Rate of All Launch Site

The largest successful launches happen in launch site – KSC LC-39A. Total 10 successful launches happened here out of total 24 successful launches!



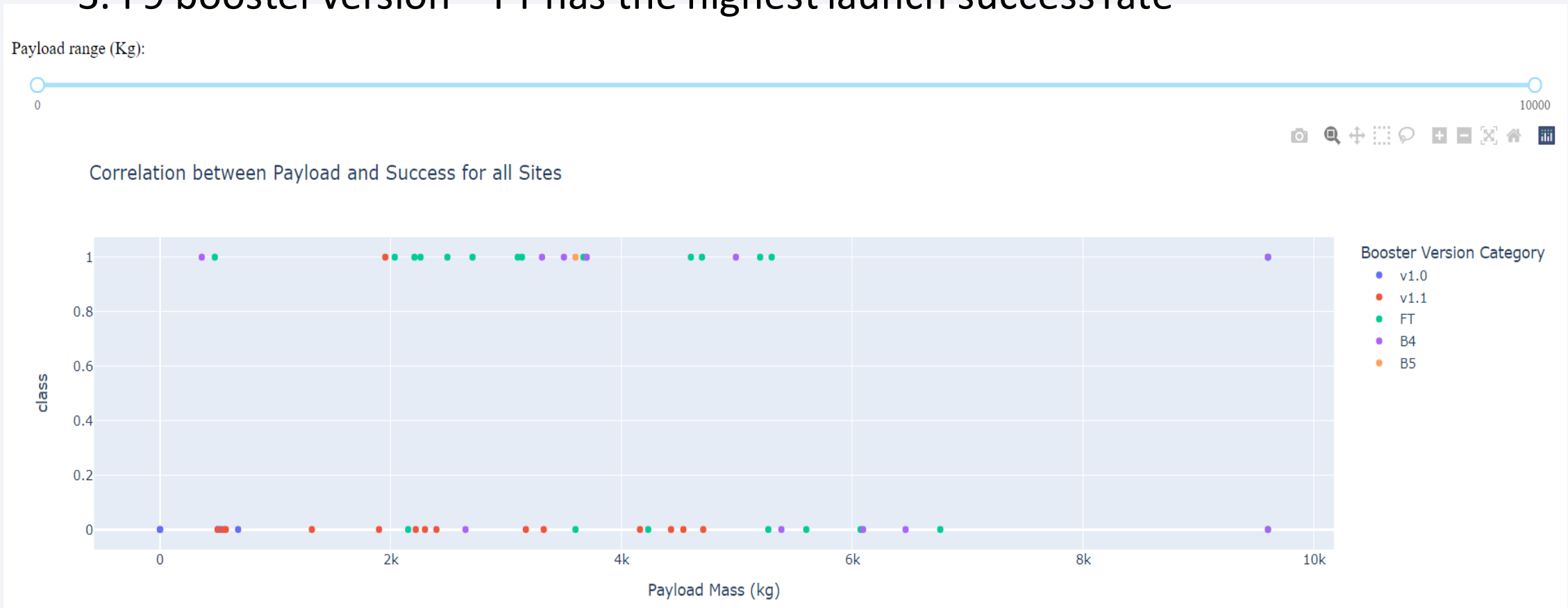
<Success Rate of KSC LC-39A Launch Site

The highest launch success rate is at launch site - KSC LC-39A.
Approximate 77% of launches here are successful.



Interactive Plot of Success Rate, Booster Version, and Payload Mass (kg)

1. Most successful launch carry within 2000kg to 4000kg payload mass.
2. Most unsuccessful launch carry within 4000kg to 7000kg payload mass.
3. F9 booster version – FT has the highest launch success rate



Section 6

Predictive Analysis (Classification)

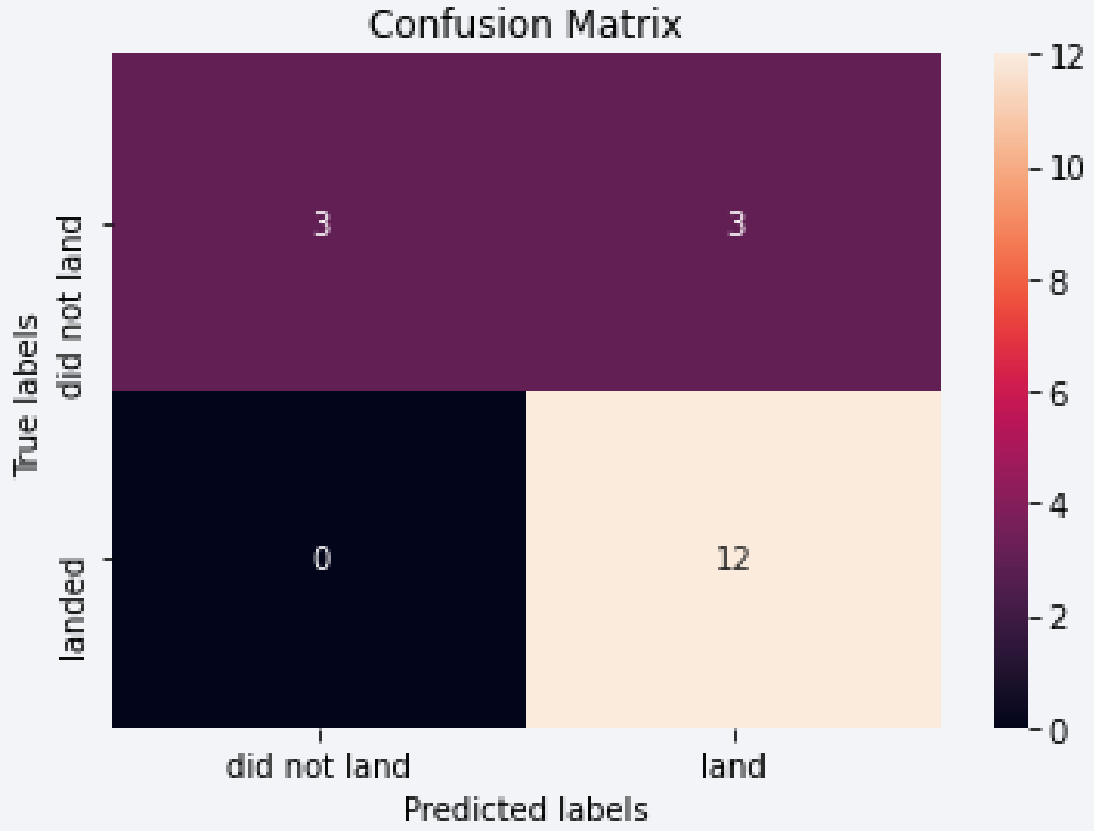
Classification Accuracy

We can see all the prediction techniques used result in practically the same accuracy score.

Prediction Technique	Prediction Accuracy (SCORE method)
Logistic Regression	0.8333
Support Vector Machine (SVM)	0.8333
Decision Tree	0.8333
K Nearest Neighbors (KNN)	0.8333

Confusion Matrix

The confusion matrix from all the prediction technique are the same. They all have 3 mistake prediction which supposed to be did not land but predicted land. Hence, all prediction technique are good to be used in this scenario.



Conclusions

- All the launch sites are nearer to the coast area than to dense population area
- The successful rate appears to be increasing every year proved that Space X is a successful project
- With the model we build, we can predict 83% using all the variable provided.
- The model will help ensure the safety and sustainability of the project
- Finally, all the variables play a role in succeeding the launching and landing of the spaceship

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!