IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Prakasam Perugu
25 Feb 2024

Outline

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

Executive Summary

- Summary of methodologies
  - Data collection
  - Data wrangling
  - Exploratory data analysis and Feature engineering
  - Interactive visual analytics and dashboard
  - Machine learning prediction
- Summary of all results
  - Stage 1 landing success is predicted with accuracy ranging 83.33% to 94.44% using different machine learning models

Introduction

- 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 capstone, we will predict if the Falcon 9 first stage will land successfully.

Section 1

Methodology

Methodology

Executive Summary

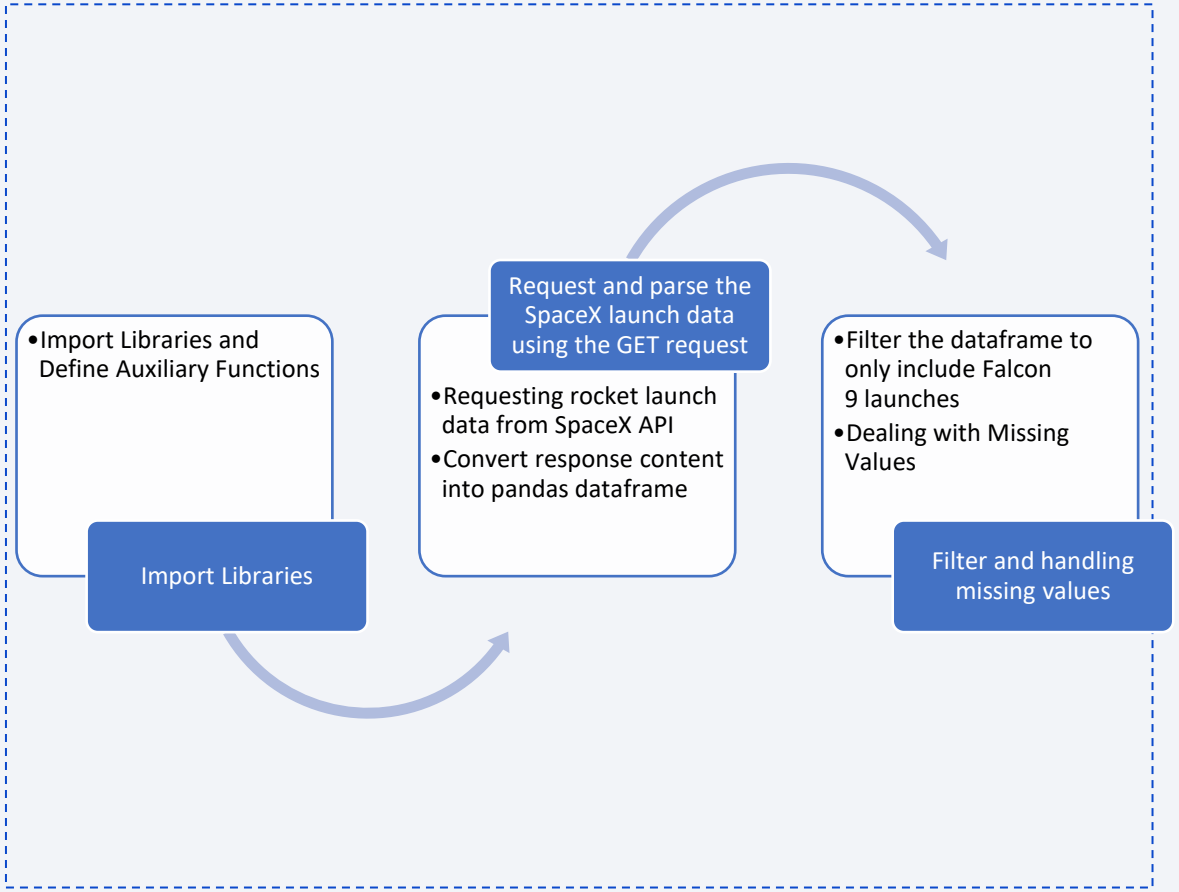
- Data collection methodology:
  - Making a get request to the SpaceX API
  - web scraping to collect Falcon 9 historical launch records from a Wikipedia page
- Perform data wrangling
  - Handled null values, converted outcomes into training labels
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Standardized data, data split for train and test, found best hyper parameters, calculated accuracy score on test data for model evaluation.

Data Collection

- Using API
  - Made get request to the SpaceX API
  - Cleaned the requested data
- Web scrapping using ‘BeautifulSoup’
  - Extracted a Falcon 9 launch records HTML table from Wikipedia
  - Parsed the table and converted it into a Pandas data frame

Data Collection – SpaceX API

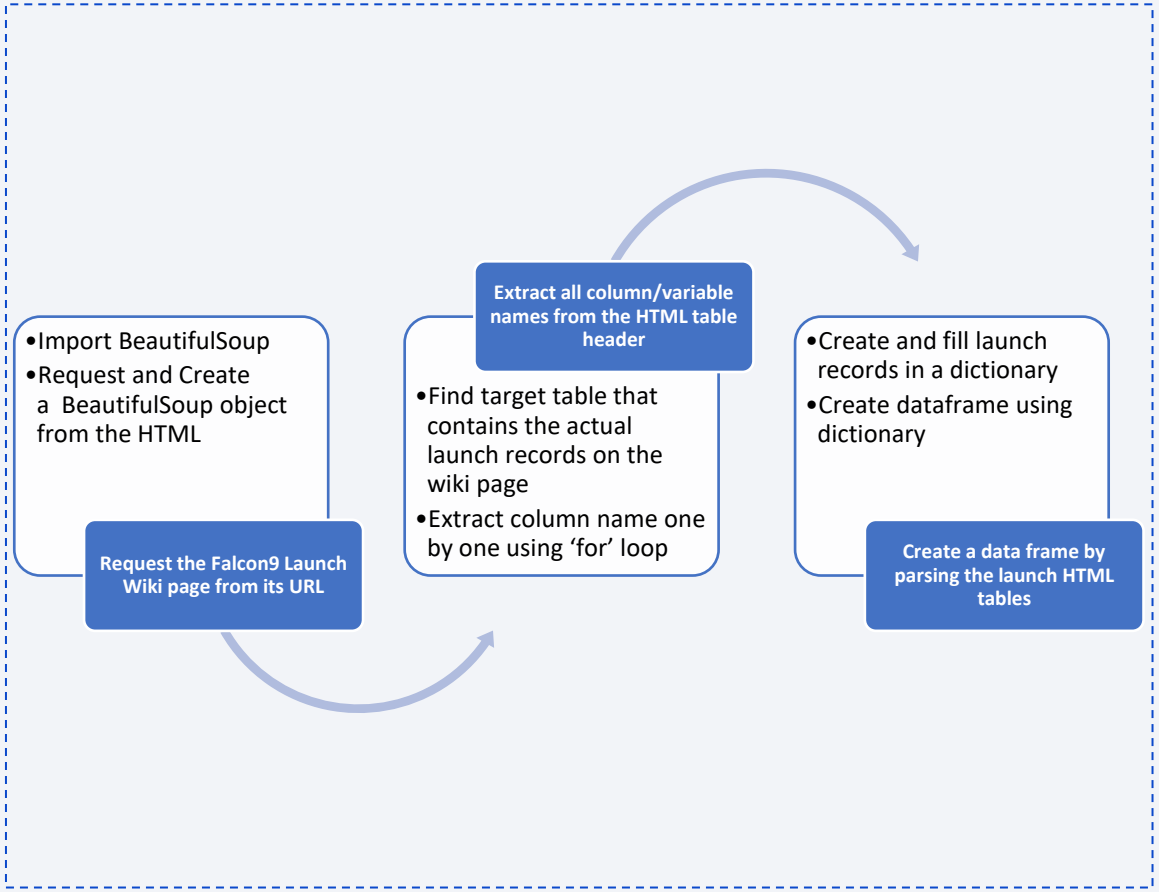
- https://github.com/pprakashsvu238/Projects/blob/main/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

- Web scraping Falcon 9 and Falcon Heavy Launches Records from Wikipedia

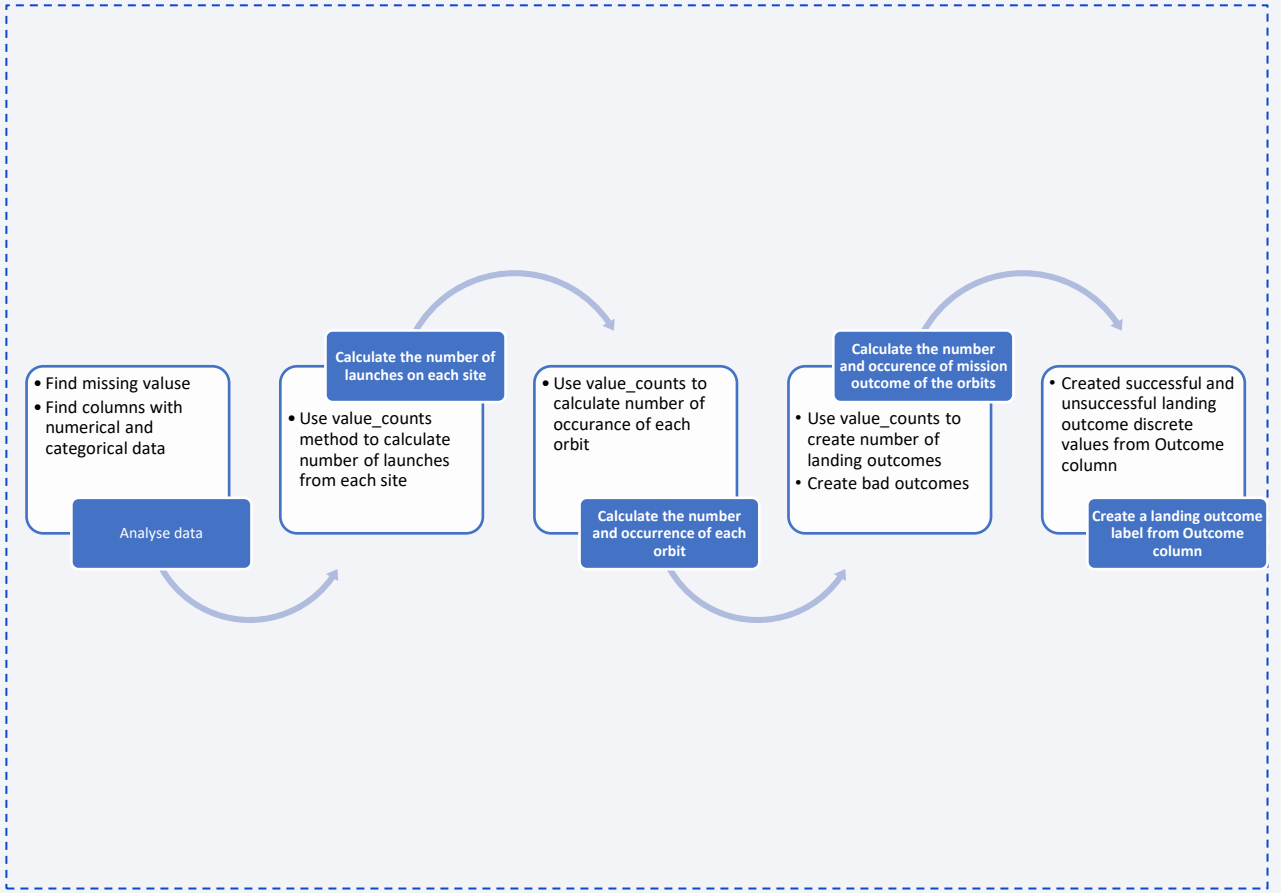
- https://github.com/pprakashsvu238/Projects/blob/main/jupyter-labs-webscraping.ipynb



Data Wrangling

- Performed Exploratory Data Analysis
- Determined training labels

- https://github.com/pprakashsvu238/Projects/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb



EDA with Data Visualization

- Scatter plots, bar charts and line charts are developed for Flight number, payload mass, orbit, launch site, success rate and launching year to visualize and find out influencing features on successful landing

- https://github.com/pprakashsvu238/Projects/blob/main/jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

EDA with SQL

- Connect database

- ***Listed the unique launch sites in the space mission, launch sites starts with CCA***
- ***Queried the total payload mass carried by boosters launched by NASA (CRS)***
- ***Queried average payload mass carried by booster version F9 v1.1***
- ***Queried the date of first successful landing from the ground pad***
- ***Listed 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***
- ***Queried month name with unsuccessful landing outcome***
- ***Rank the count of landing outcomes in a given period***

- https://github.com/pprakashsvu238/Projects/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Circles, markers, marker cluster and lines are added for NASA centre, launch sites, successful/failed launches to find out the successful launch sites.
- Added markers, mouse position, polyline, distances of closed city, coastal line, highway, and railway line to find the importance of launch site with respect to these proximities.

- https://github.com/pprakashsvu238/Projects/blob/main/lab_jupyter_launch_site_location.jupyterlite.ipynb

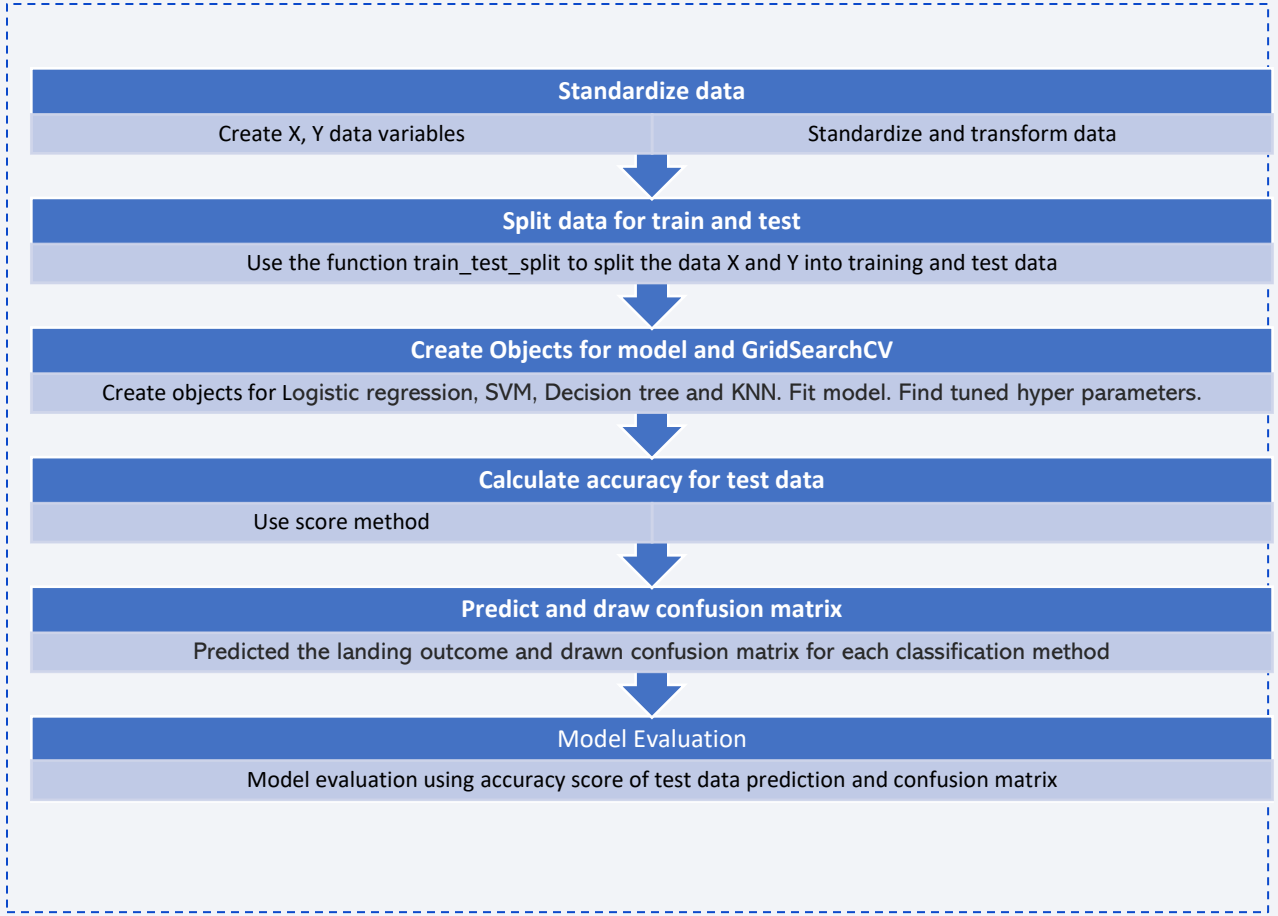
Build a Dashboard with Plotly Dash

- Added a pie chart to show the total successful launches count for all sites
- Added a scatter chart to show the correlation between payload and launch success

- https://github.com/pprakashsvu238/Projects/blob/main/spacex_dash_app.py

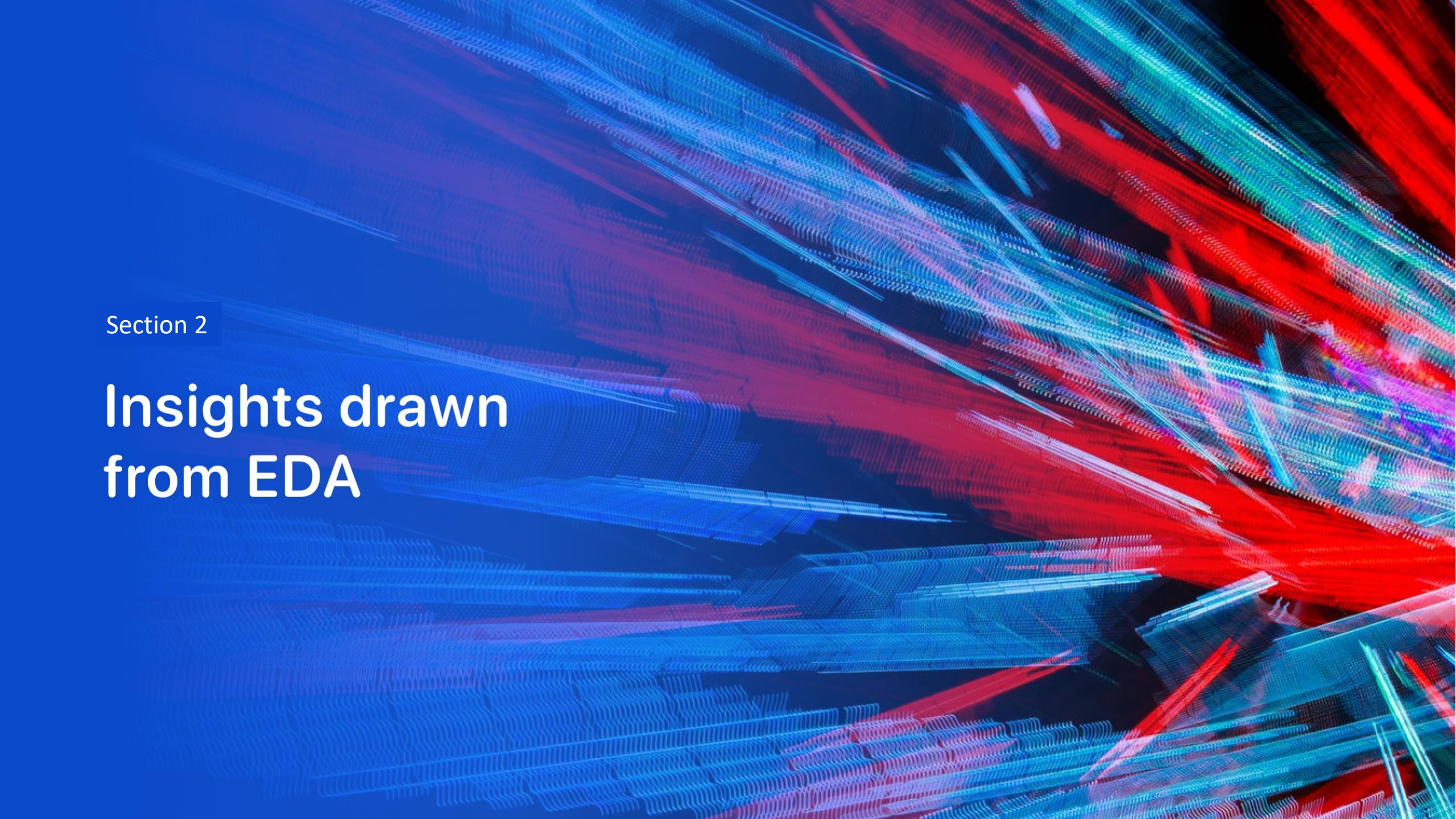
Predictive Analysis (Classification)

- Summary of steps for predicting landing outcome using classification models
  - Created X (data) and Y (landing 'Class') variables. Standardized and transformed data
  - Created train and test samples
  - Objects for logistic regression, SVM, Decision tree and KNN.
  - GridSearchCV object was created and fitted for finding out tuned hyper parameters
  - Accuracy was calculated for each classification method
  - Score for test data was calculated for model evaluation
  - Predicted the landing outcome and drawn confusion matrix for each classification method
- https://github.com/pprakashsvu238/Projects/blob/main/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

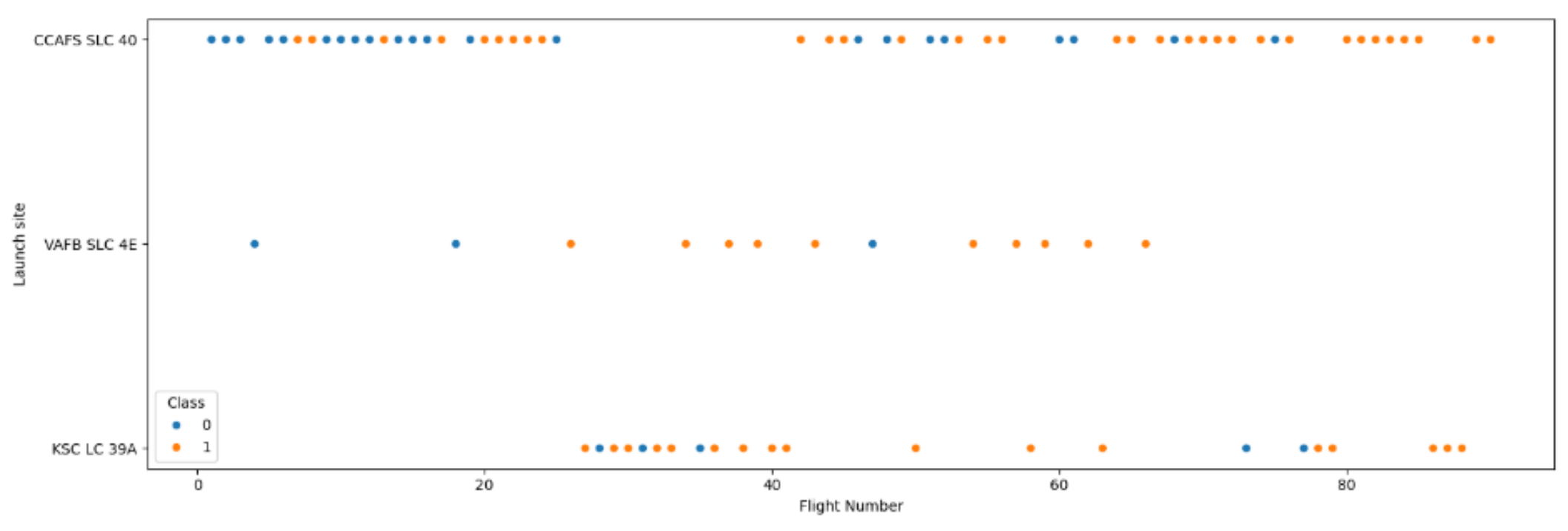


Results

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

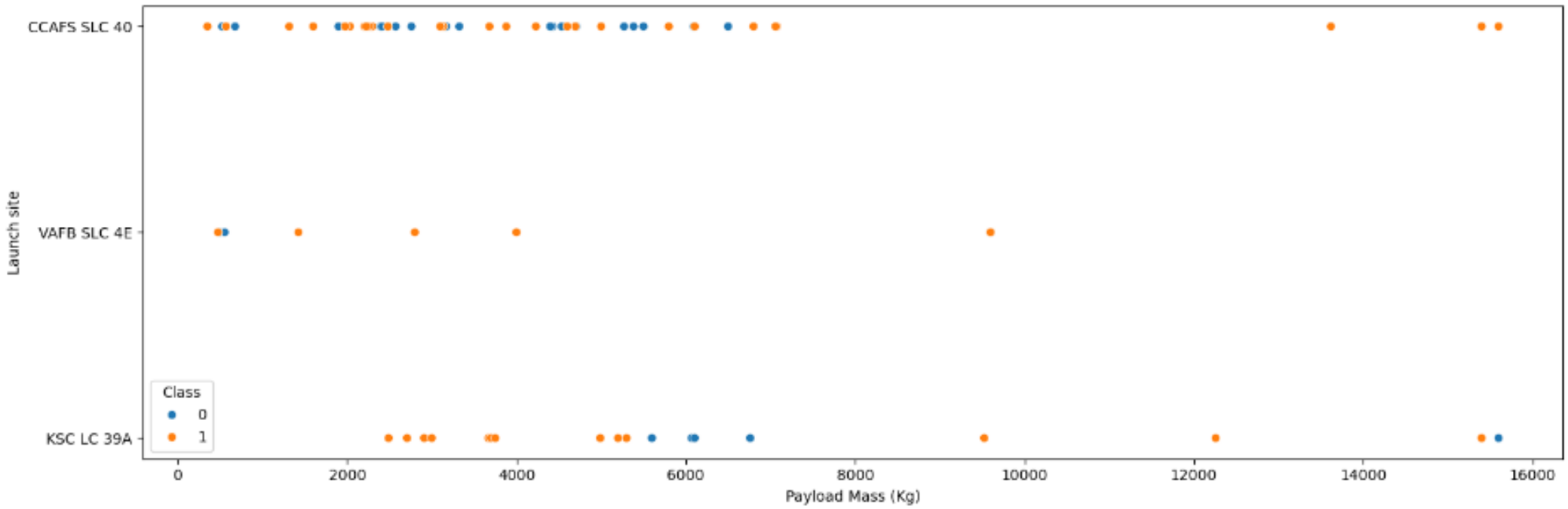

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



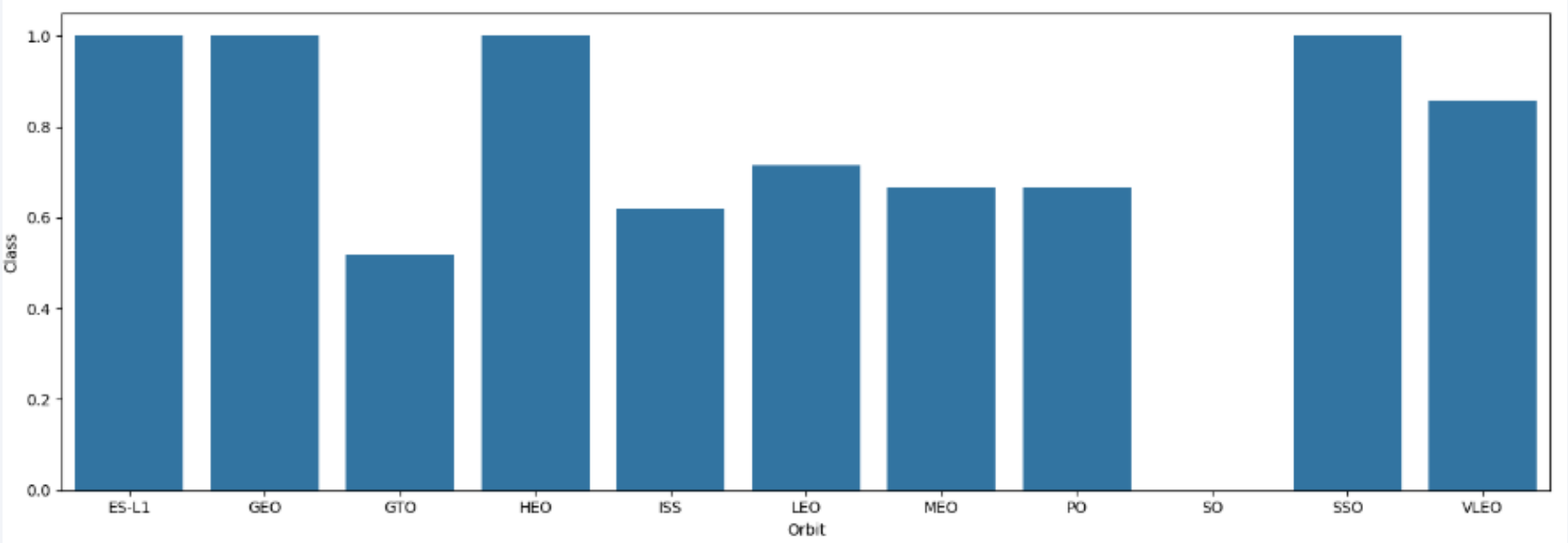
- CCAFS LC-40, has a success rate of 60 %, while KSC LC-39A and VAFB SLC 4E has a success rate of 77%

Payload vs. Launch Site



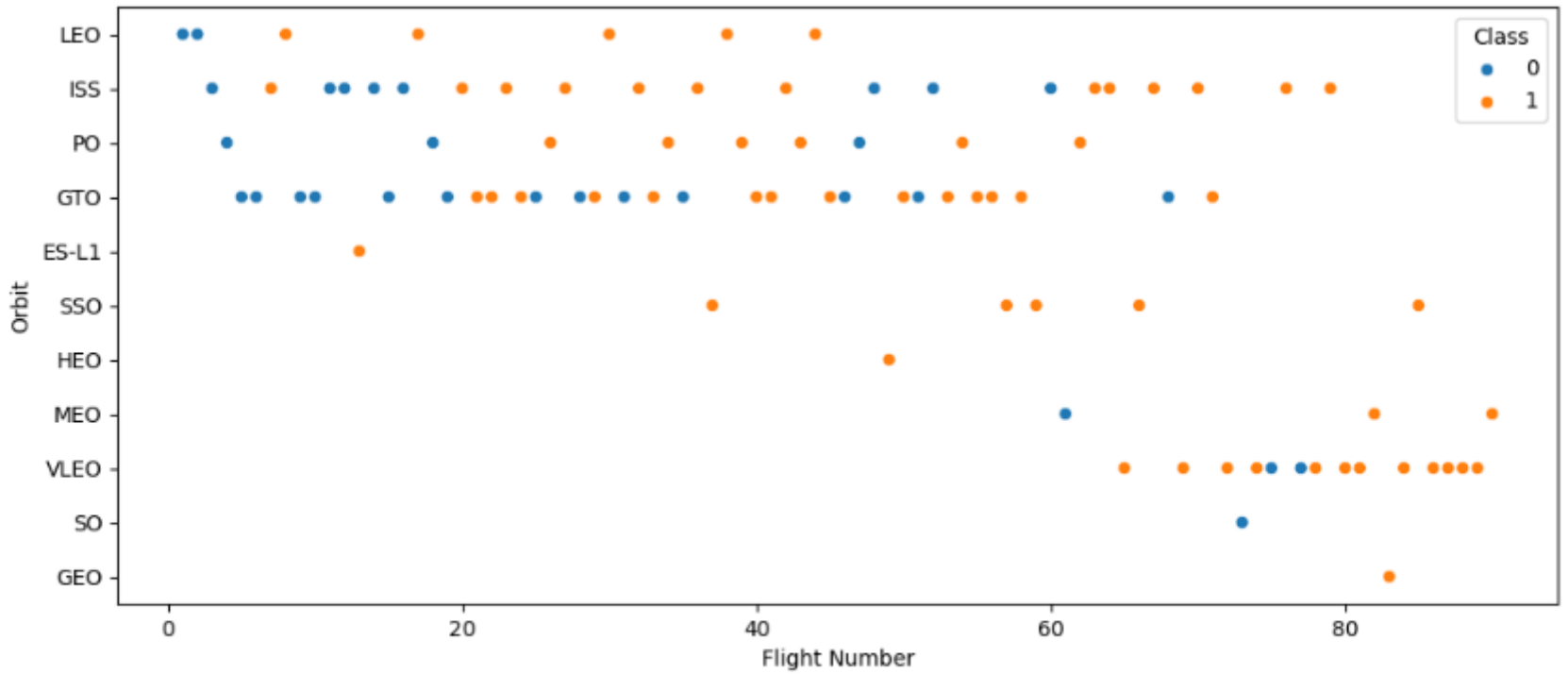
Now if you observe Payload Vs. Launch Site scatter point chart you will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).

Success Rate vs. Orbit Type



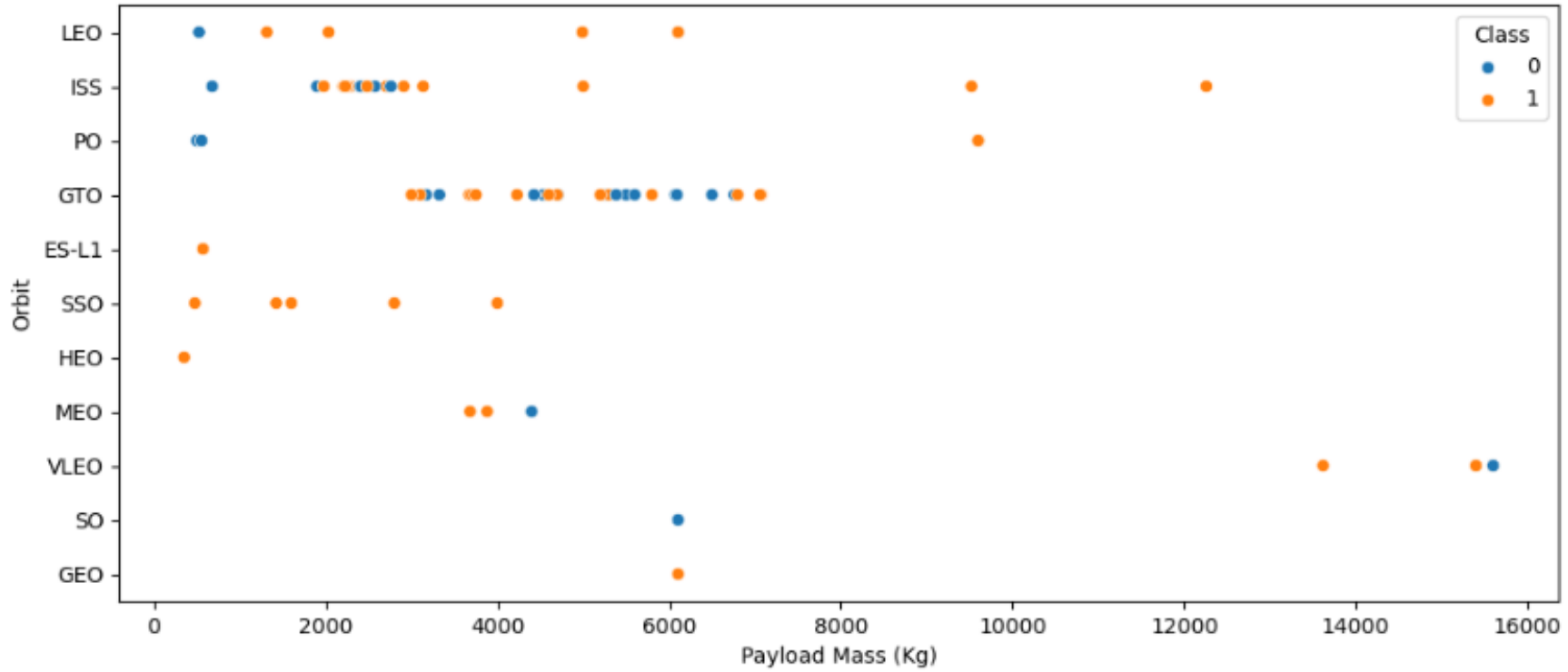
- ES L-1, GEO, HEO and SSO orbits have highest successrate

Flight Number vs. Orbit Type



You should see that in the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

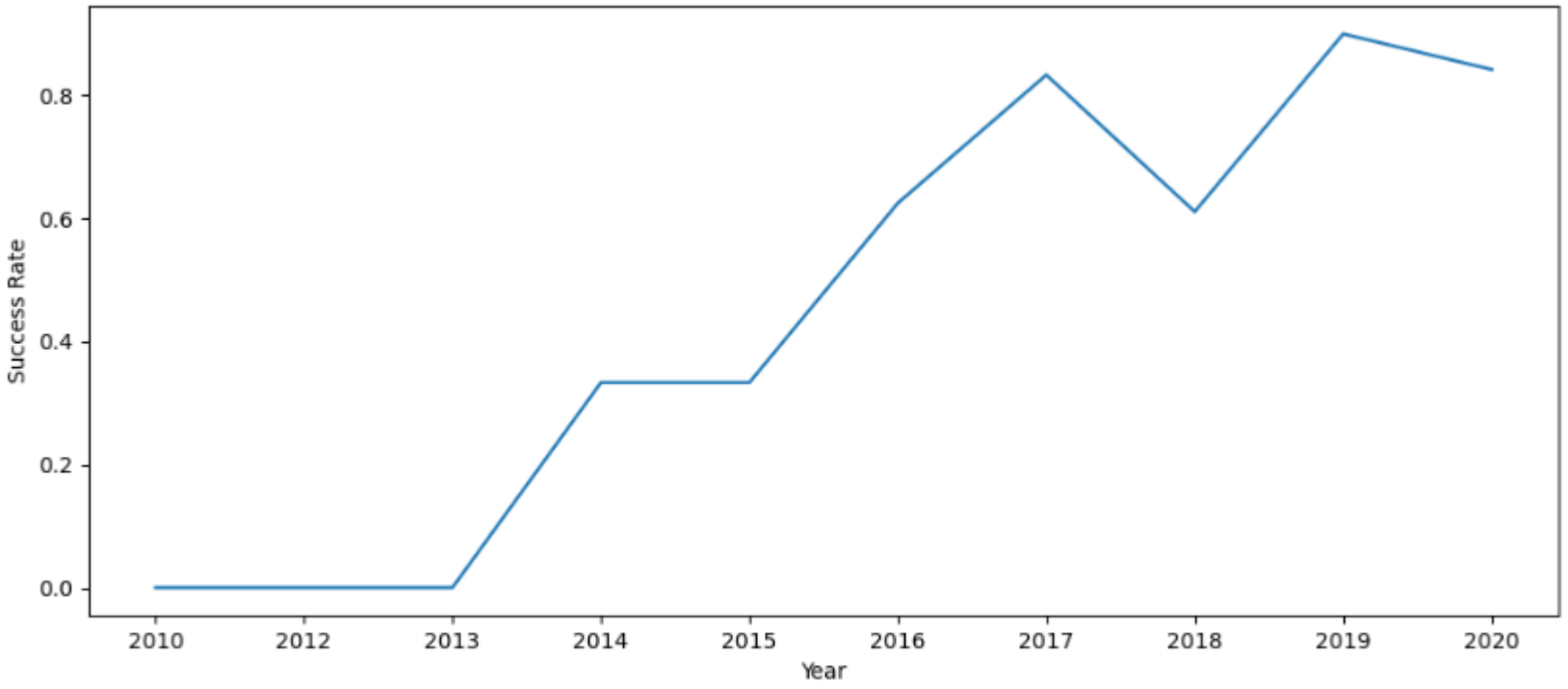
Payload vs. Orbit Type



With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.

However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.

Launch Success Yearly Trend



you can observe that the sucess rate since 2013 kept increasing till 2020

All Launch Site Names

- There are 4 different launch sites

```
%sql SELECT DISTINCT "Launch_Site" FROM SPACEXTABLE;
```

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

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

Launch Site Names Begin with 'CCA'

- Top 5 records where launch sites begin with `CCA`

```
%sql select * from SPACEXTBL where "Launch_Site" like "CCA%" limit 5;
```

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

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

- Total payload carried by boosters from NASA

```
%sql select sum("PAYLOAD_MASS__KG_") as "NASA (CRS) Total Payload_Kg" from SPACEXTBL where "Customer"= "NASA (CRS)";
```

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

NASA (CRS) Total Payload_Kg
45596

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1

```
%sql select avg("PAYLOAD_MASS__KG_") as "Avg Payload by F9 v1.1" from SPACEXTBL where "Booster_Version" like "F9 v1.1%";
```

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

Avg Payload by F9 v1.1
2534.6666666666665

First Successful Ground Landing Date

- The first successful landing outcome on ground pad was on 2015-12-22

```
%sql select min("Date") as "Date", "Landing_Outcome" from SPACEXTBL where "Landing_Outcome"="Success (ground pad)";
```

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

Date	Landing_Outcome
2015-12-22	Success (ground pad)

Successful Drone Ship Landing with Payload between 4000 and 6000

- Names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

```
%sql select "Booster_Version","PAYLOAD_MASS__KG_","Landing_Outcome" from SPACEXTBL where "Landing_Outcome"="Success (drone ship)"
```

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

Booster_Version	PAYLOAD_MASS__KG_	Landing_Outcome
F9 FT B1022	4696	Success (drone ship)
F9 FT B1026	4600	Success (drone ship)
F9 FT B1021.2	5300	Success (drone ship)
F9 FT B1031.2	5200	Success (drone ship)

Total Number of Successful and Failure Mission Outcomes

- Total number of successful and failure mission outcomes

```
%sql select count("Mission_Outcome") as "Total" from SPACEXTBL where "Mission_Outcome" like "Succ%" or "Mission_Outcome" like "Fa
```

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

Total
101

Boosters Carried Maximum Payload

- Names of the booster which have carried the maximum payload mass

```
%sql select "Booster_Version", "PAYLOAD_MASS__KG_" from SPACEXTBL where "PAYLOAD_MASS__KG_" in (select max("PAYLOAD_MASS__KG_") f
```

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

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

2015 Launch Records

- Failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

```
%sql select substr(Date,6,2) as month, "Landing_Outcome", "Booster_Version", "Launch_Site" from SPACEXTBL where substr(Date,0,5)=
```

```
 * 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

- Rank of landing outcomes between the date 2010-06-04 and 2017-03-20, in descending order

```
%%sql select "Landing_Outcome", count("Landing_Outcome") as "Total",
RANK() OVER (order by count("Landing_Outcome") desc) as "RANK"
from SPACEXTBL
group by "Landing_Outcome";
```

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

Landing_Outcome	Total	RANK
Success	38	1
No attempt	21	2
Success (drone ship)	14	3
Success (ground pad)	9	4
Failure (drone ship)	5	5
Controlled (ocean)	5	5
Failure	3	7
Uncontrolled (ocean)	2	8
Failure (parachute)	2	8
Precluded (drone ship)	1	10
No attempt	1	10

Section 3

Launch Sites Proximities Analysis

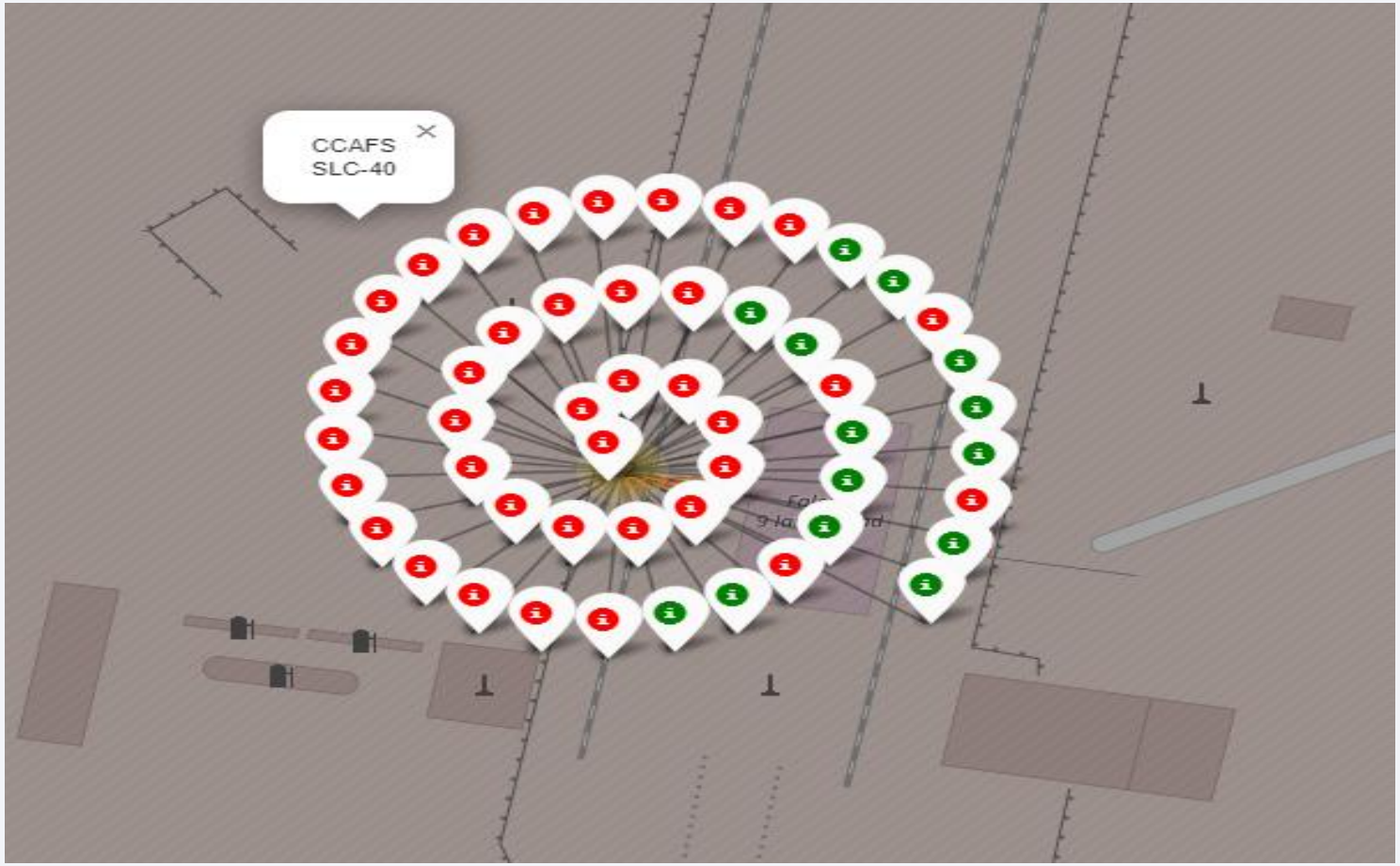
Location of All Launch Sites on Map

All launch sites are on coastal areas in North America

One is on west coast and three are on east coast

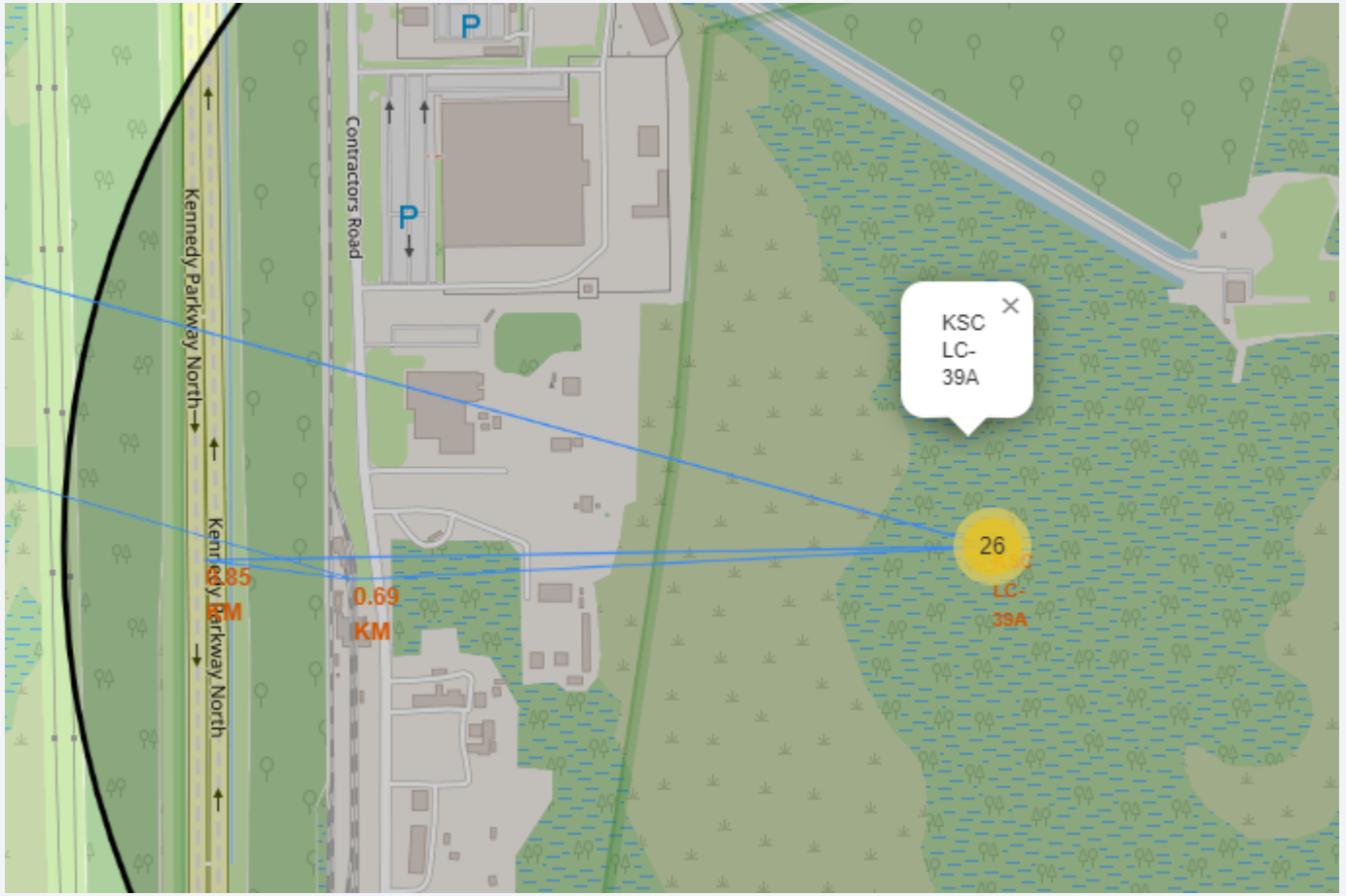
Launch Outcomes

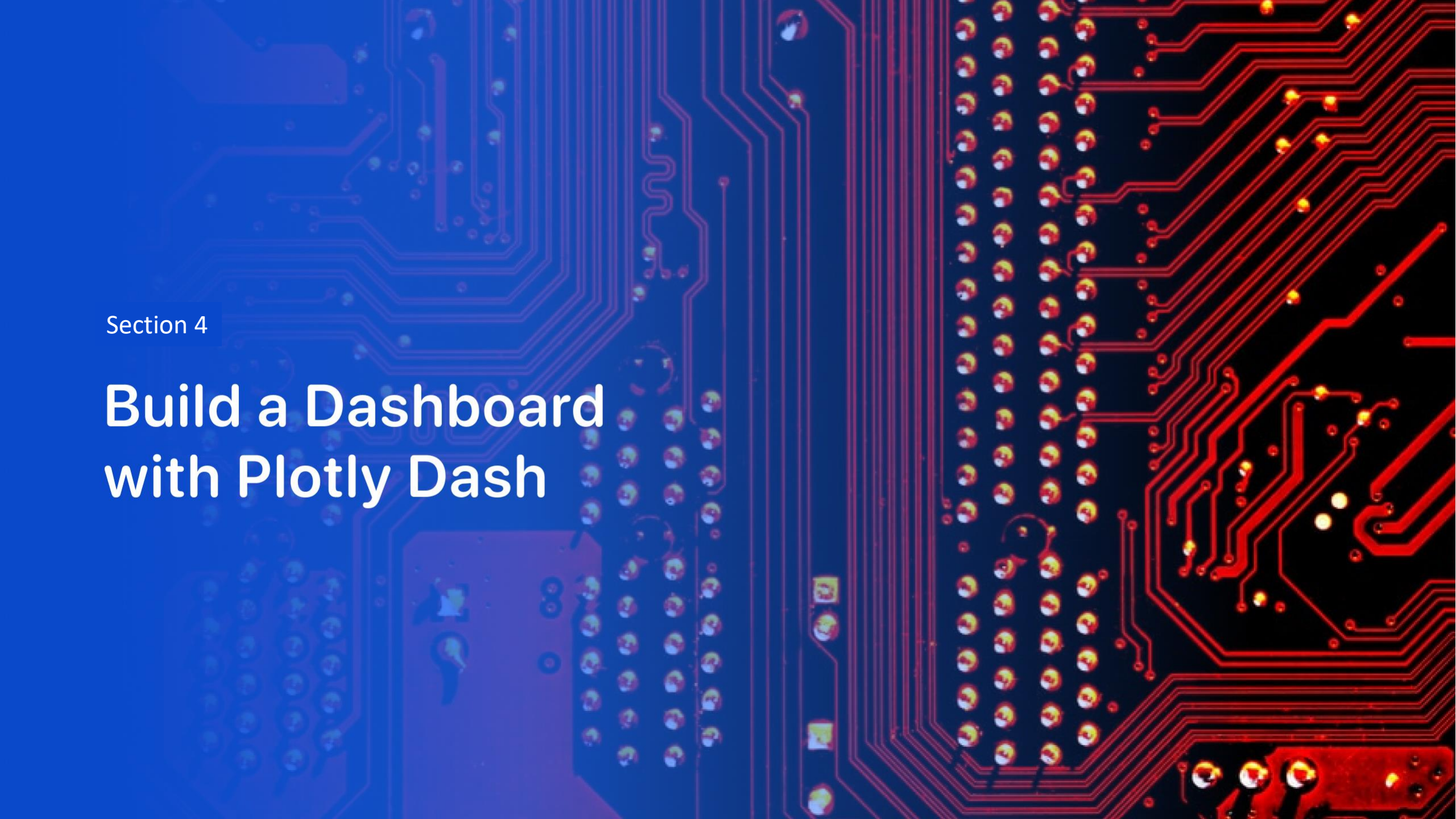
Launch outcomes at CCAFS SLC-40. Red color – Failed, Green color - Successful



Launch Site Proximities

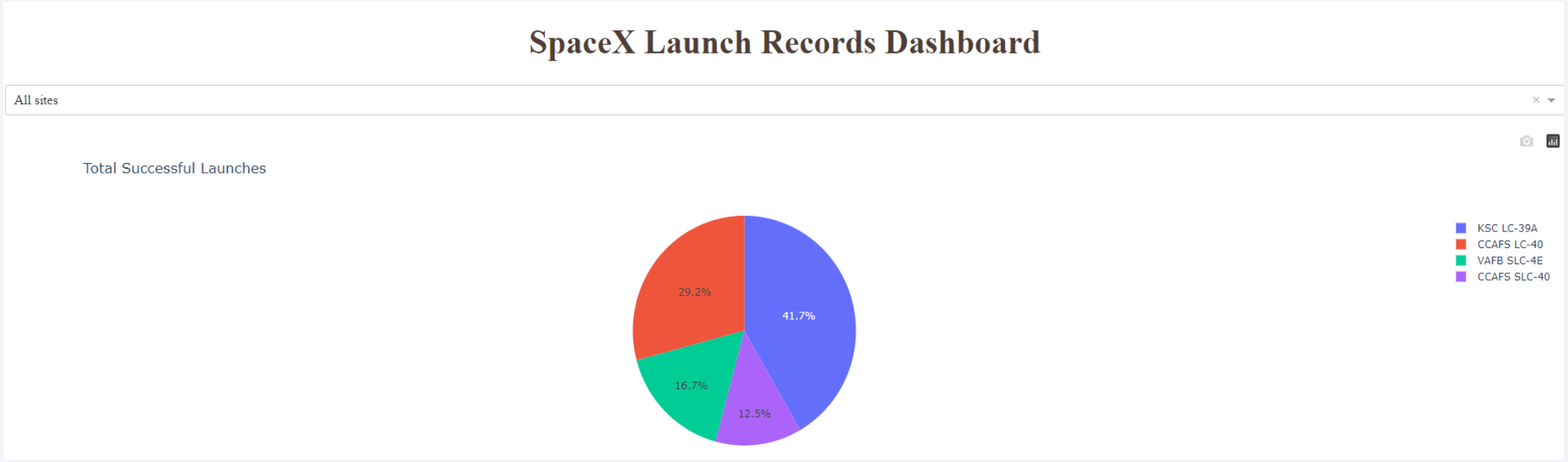
- Launch site is located near to the highway, railway and coastal area, but it is far from city.




Section 4
Build a Dashboard with Plotly Dash

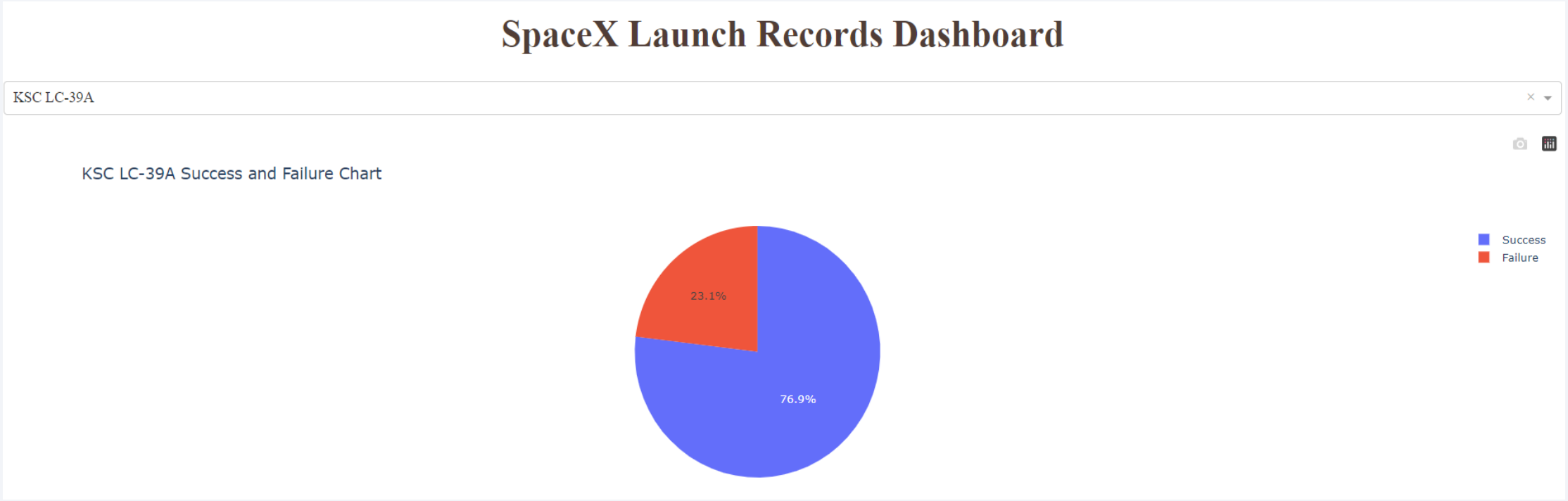
Launch Sites - Proportion of Successful Launches

- KSC LC-39A has highest successful launches, whereas CCAFS SLC-40 has least successful launches



Highest Success Rate Launch Site

- KSC LC-39A has 76.9% Success rate, highest among all launch sites



Correlation Between Payload and Success

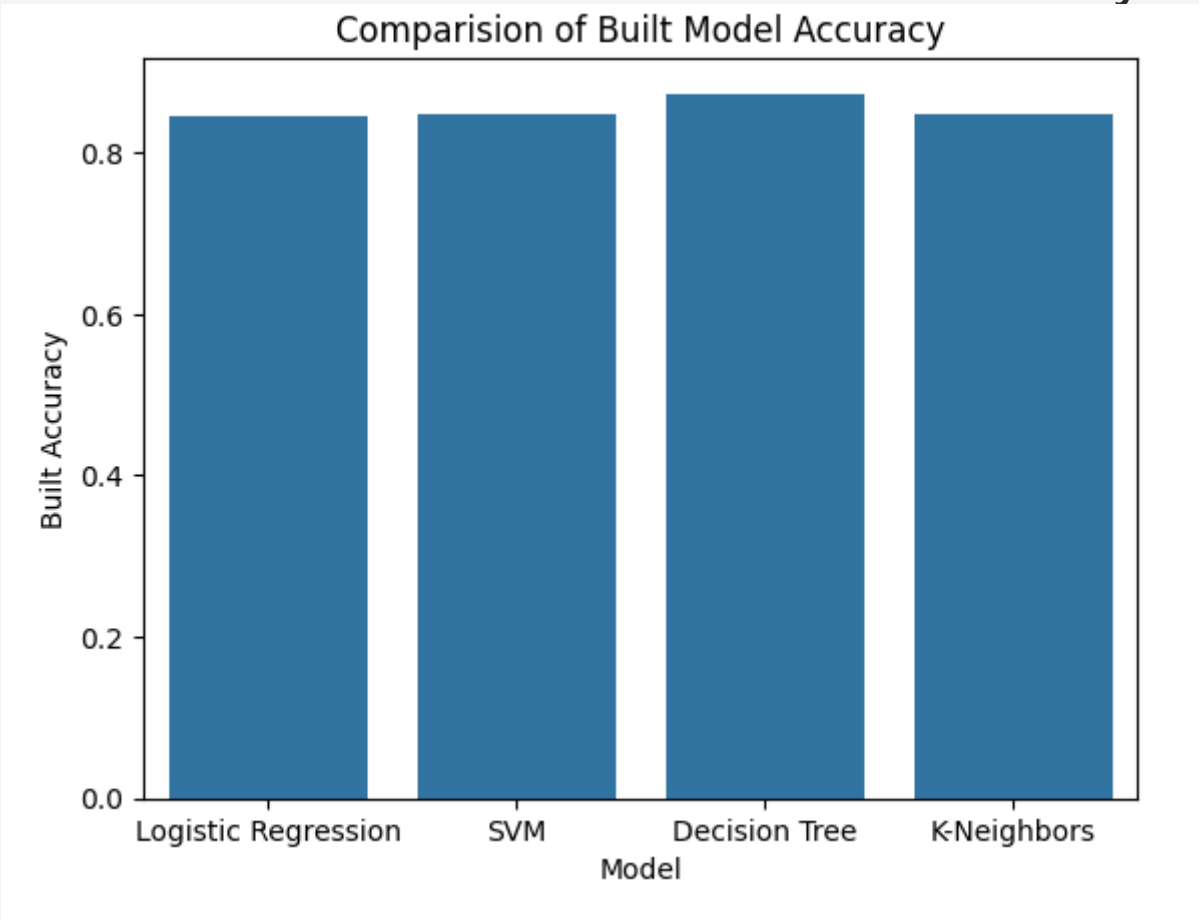
- Boosters range between 6K to 10K has less success rate

Section 5

Predictive Analysis (Classification)

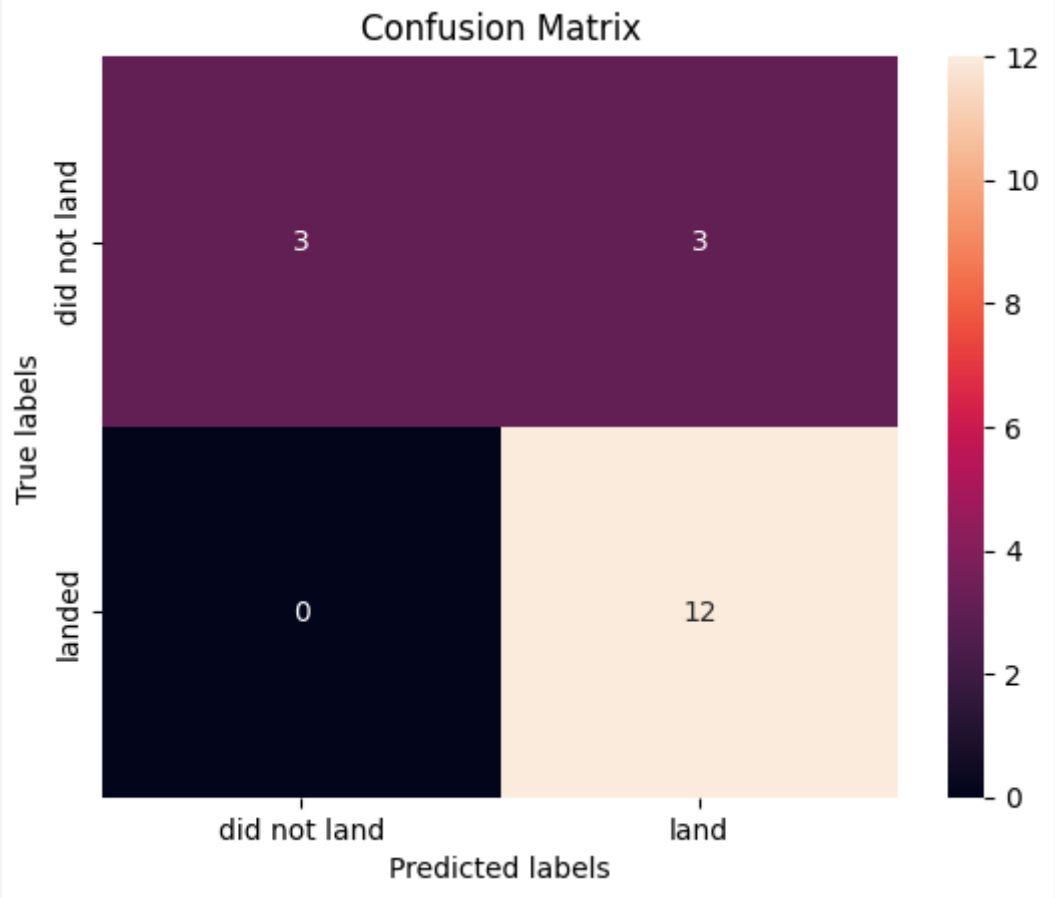
Classification Accuracy

- Decision Tree Classification model has the Built Model Accuracy



Confusion Matrix

- KNN has best Confusion matrix with only a problem of False Positive



Conclusions

- Prediction for stage-1 landing success is about 83%
- KSC LC 39A launch site has highest success rate
- F9 Booster version for range 6K to 10K has low success rate

- In conclusion, SpaceX rockets cost less due to successful landing of stage-1 which can be re-used future rocket launches.

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!