IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Mohamad Oghli
21/2/2023

Outline

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

Executive Summary

- First collecting the required data using different resources. Then do some data wrangling to prepare it for the analysis phase. After that we will conduct Exploratory Analysis on the data using SQL, Pandas and Matplotlib libraries. We will also build dashboards and launch sites map to analyze data interactively using Plotly Dash and Folium. Finally, we will do predictive analysis using different machine learning methods to determine Falcon 9 will land successfully

- After conducting the analysis on the dataset, we found that all launch sites near the coast. Also, the more flight number increases the more the success rate increase of Flacon9 launches. And the heavier the mass payload the more likely that Flacon9 will land successfully. At the end of predictive analysis, we found that decision tree is the best Machine Learning method for predicting mission outcome of Flacon9.

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.
- Predict if the Falcon 9 first stage will land successfully.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data collected using REST API of SpaceX launches and Web Scraping of different web resources.
- Perform data wrangling
  - Convert Mission Outcomes into training labels with 1 means the booster successfully landed and 0 it was unsuccessful.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Create Column for the class and standardize the data. Then split into training and test data.
  - Find best Hyperparameter for SVM, Classification Trees and Logistic Regression with best method on test data.

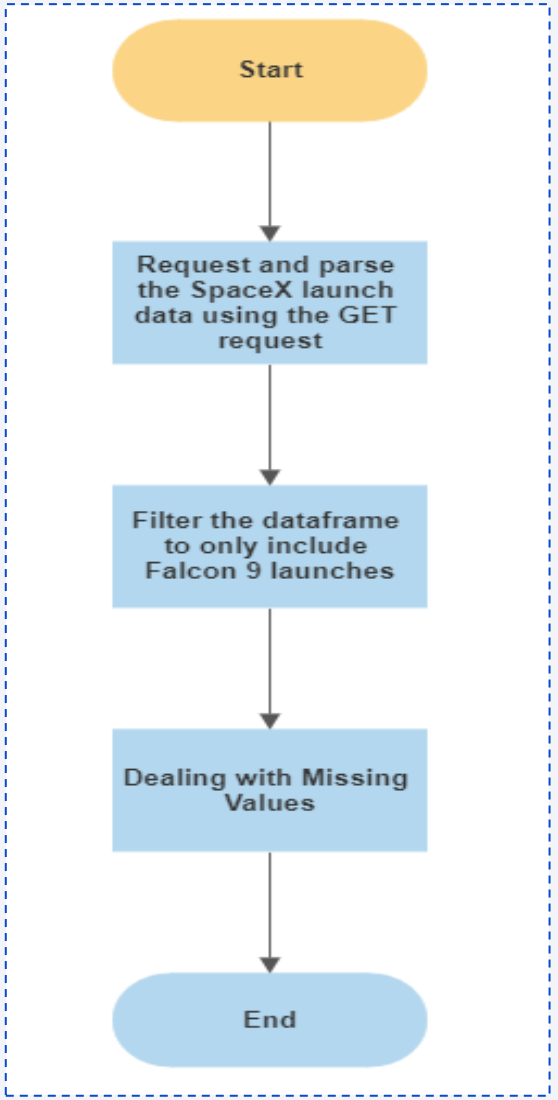
Data Collection

- Data collected using Two main methods:
  - Request REST API of SpaceX launches Program.
  - Web Scraping of different internet web resources mainly List of Falcon 9 and Falcon Heavy launches on Wikipedia.

Data Collection – SpaceX API

- The flowchart demonstrate the process of requesting SpaceX API data, saving it in data frame and manipulating it.

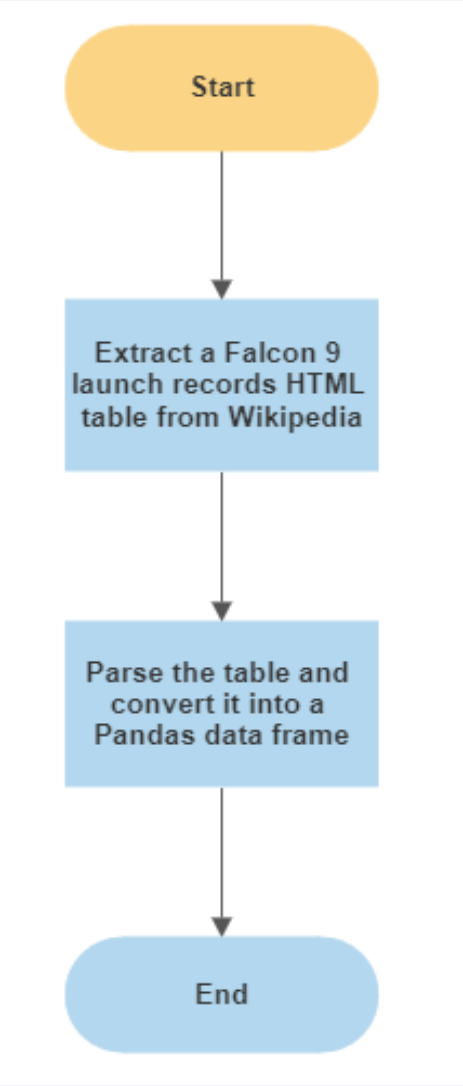
- Completed SpaceX API calls notebook



Data Collection - Scraping

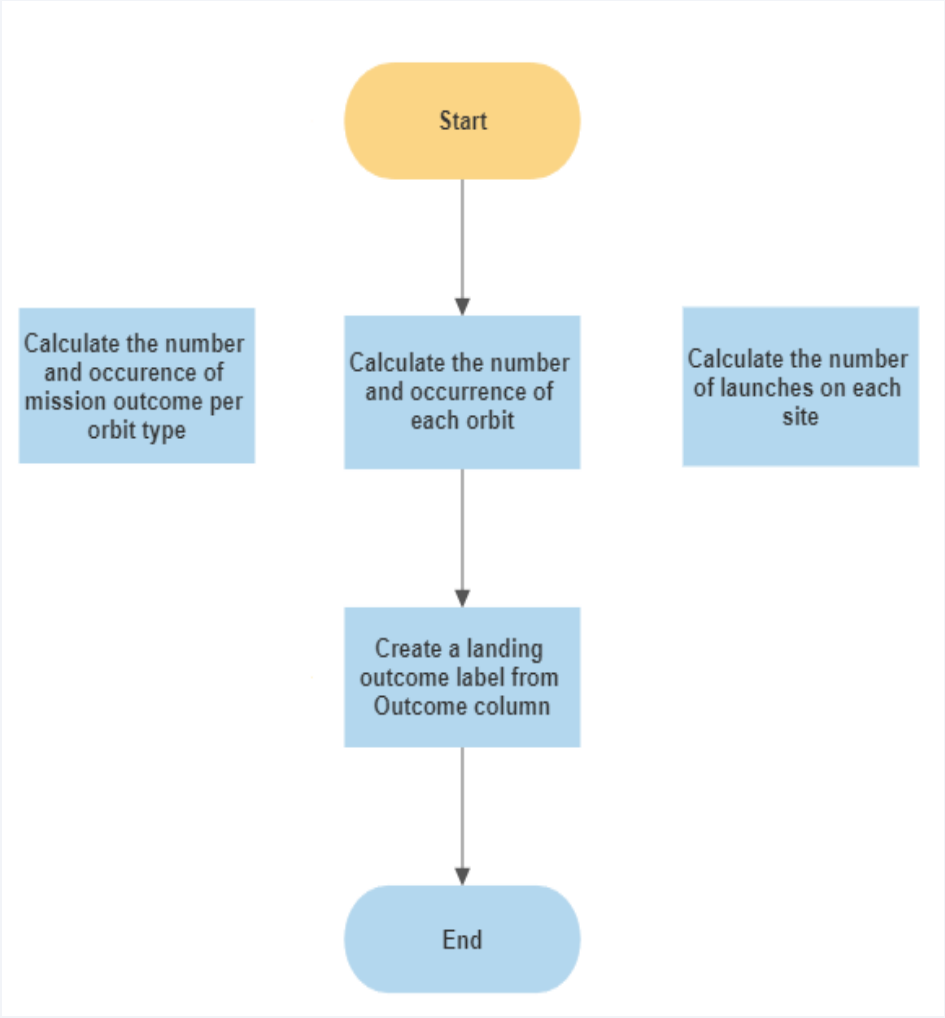
- The flowchart show the process of extracting data from Wikipedia and saving it to data frame.

- Completed web scraping notebook



Data Wrangling

- Convert Mission Outcomes into Training Labels with 1 for successful landing and 0 for unsuccessful.
- Completed data wrangling related notebooks



EDA with Data Visualization

- Scatter Plot Chart
  - Relationship between Flight Number and Launch Site.
  - Relationship between Payload and Launch Site.
  - Relationship between FlightNumber and Orbit type.
  - Relationship between Payload and Orbit type.
- Bar Chart
  - Relationship between success rate of each orbit type.
- Line Chart
  - Launch success yearly trend.
- Completed EDA with data visualization notebook

EDA with SQL

- SQL queries
  - Names of the unique launch sites in the space mission.
  - First 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.
  - Records which will display the month names, failure landing outcomes in drone ship ,booster versions, launch sites for the months in year 2015.
  - Count of successful landing outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

- Completed EDA with SQL notebook

Build an Interactive Map with Folium

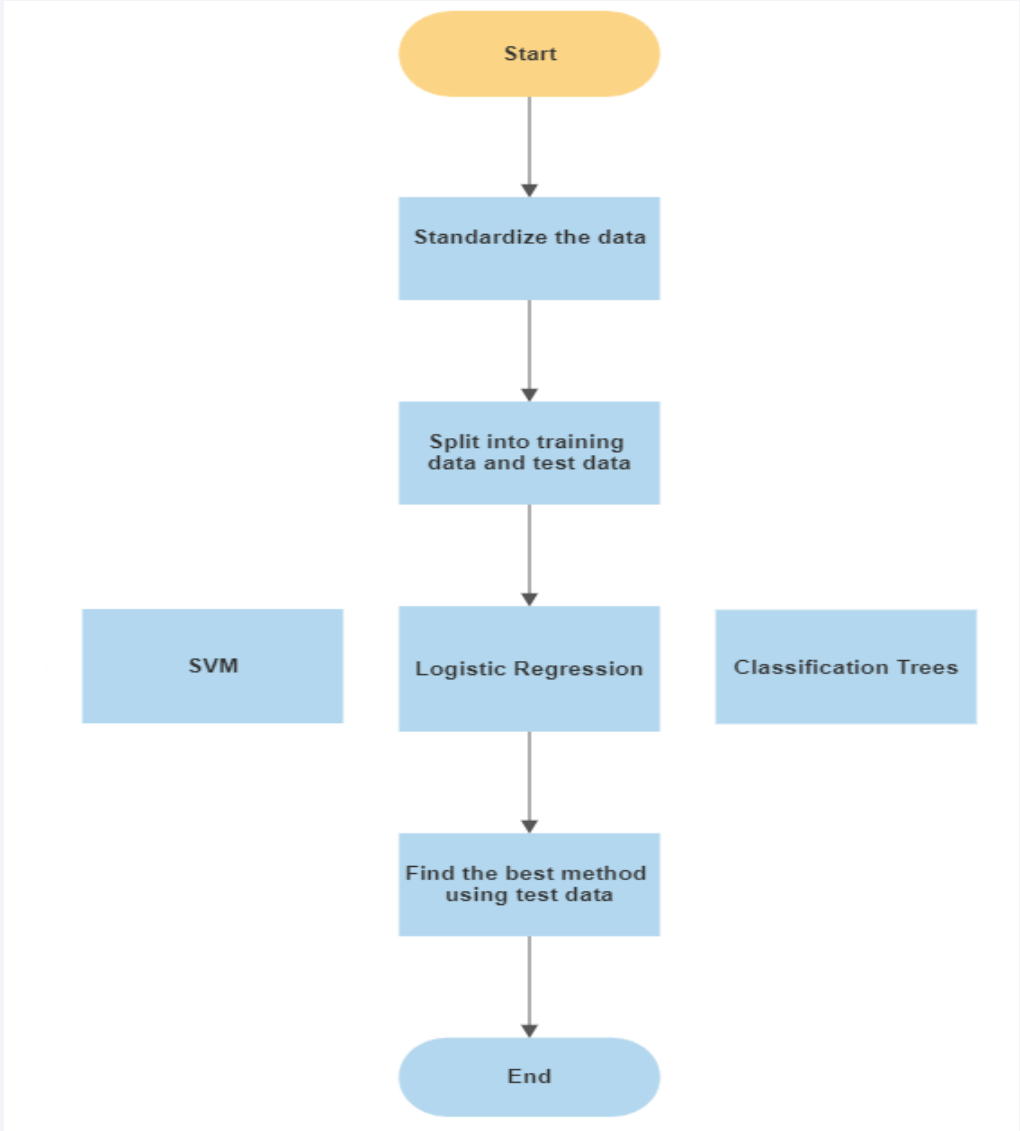
- Added Markers, Circles for all launch sites on the map.
- Added Marker Cluster for success/failed launches for each site on the map.
- Drawing PolyLine marker between a launch site to its proximities.
- Completed interactive map with Folium map

Build a Dashboard with Plotly Dash

- Total successful launches for all sites and success rate for each site using Pie Chart.
- Correlation between payload and launch outcome for All Sites according to booster version using Scatter Plot Chart.
- Completed Plotly Dash lab

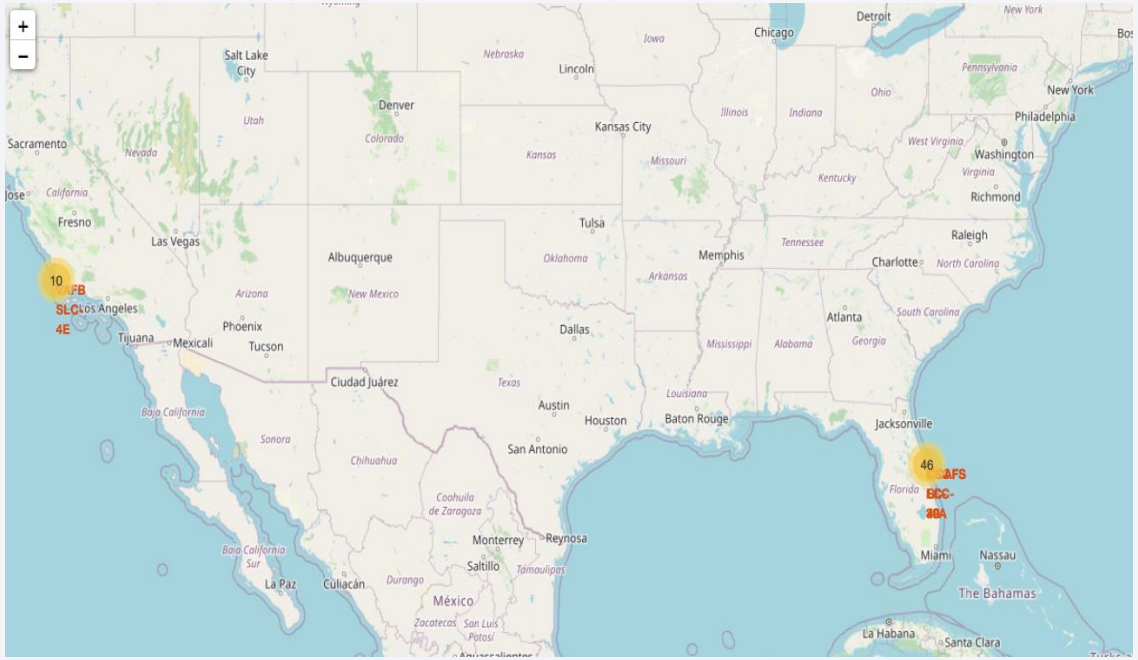
Predictive Analysis (Classification)

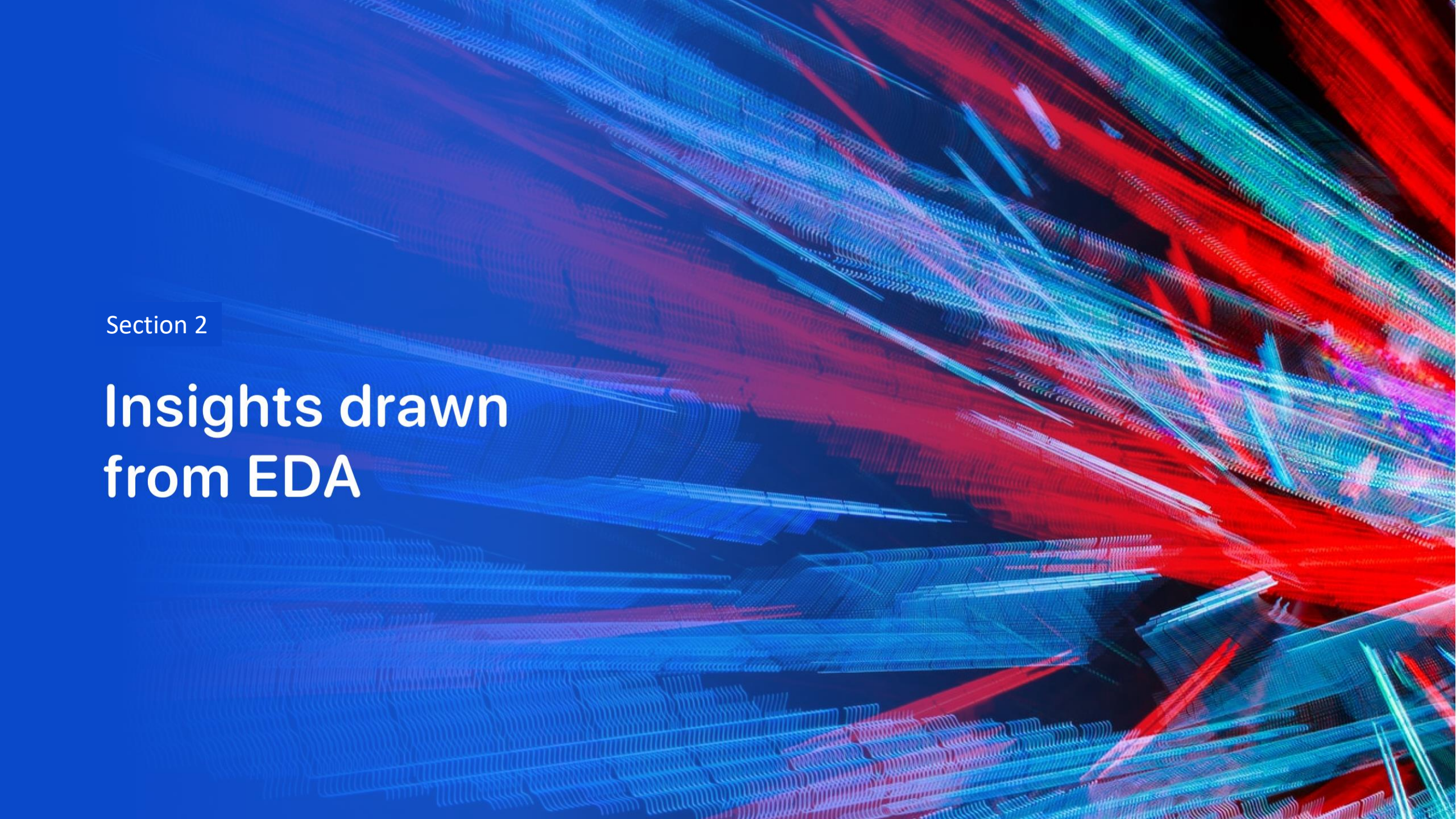
- Main Steps
  - Create a column for the class.
  - Standardize the data.
  - Split into training data and test data.
  - Find best Hyperparameter for SVM, Classification Trees and Logistic Regression.
  - Find the best method using test data.
- [Completed predictive analysis lab]



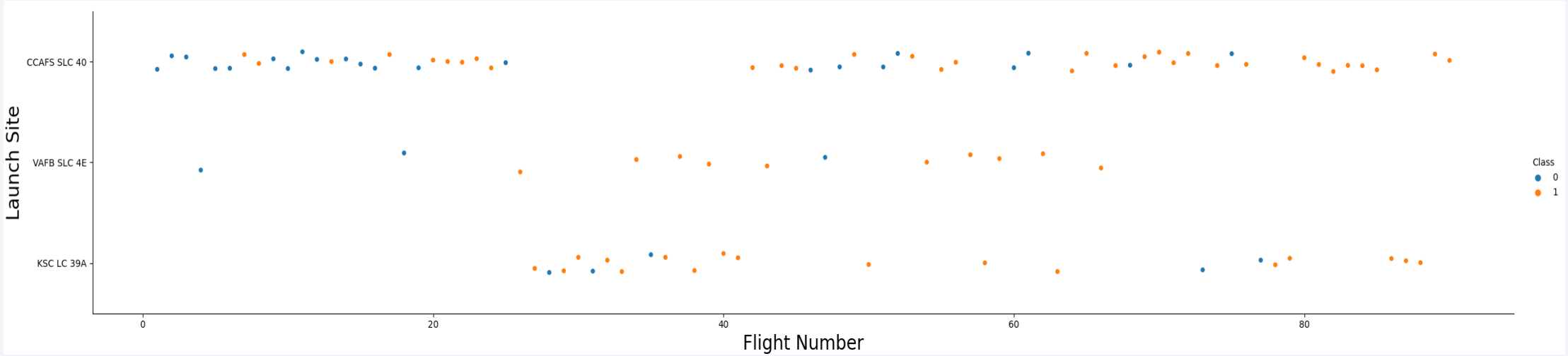
Results

- The more the flight number increase the success rate increase in all launch sites.
- Launch success rate since 2013 kept increasing till 2020.
- ES-L1, HEO and SSO orbits has the highest success rate.
- Decision Tree method has the highest prediction score for landing outcome using Grid Search.



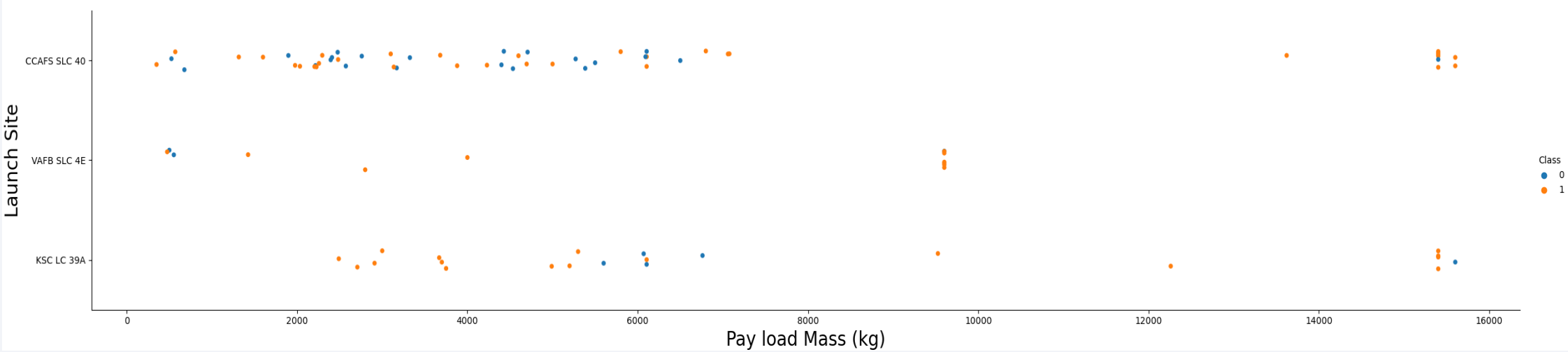

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



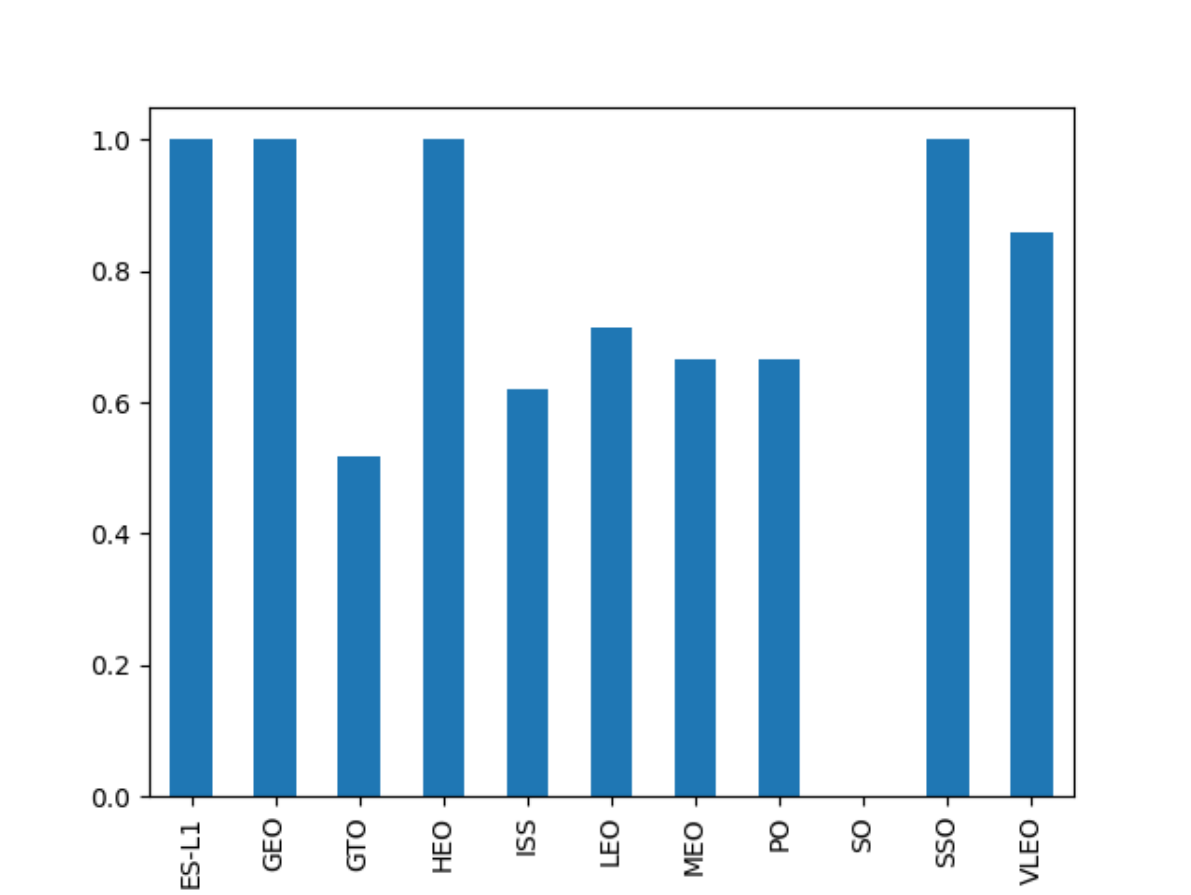
- We can see the more the flight number increase the success rate increase in all launch sites.

Payload vs. Launch Site



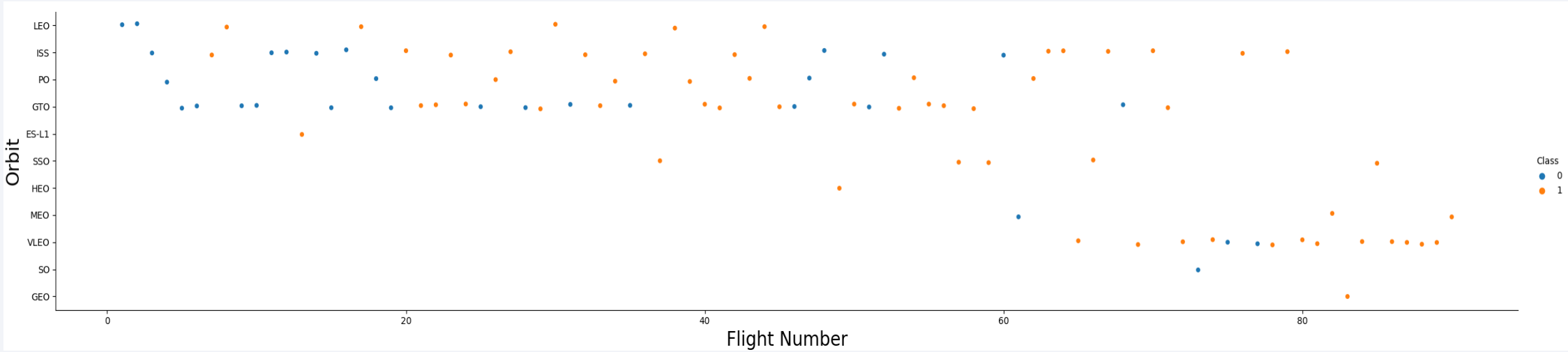
- We can see for the VAFB-SLC launch site there are no rockets launched for heavy payload mass(greater than 10000).

Success Rate vs. Orbit Type



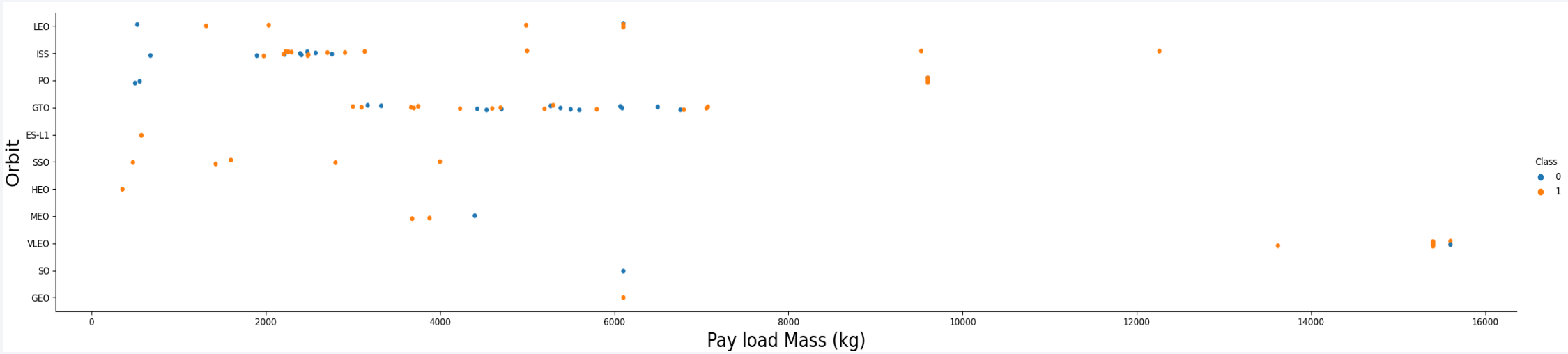
- We can see that ES-L1, HEO and SSO has the highest success rate.

Flight Number vs. Orbit Type



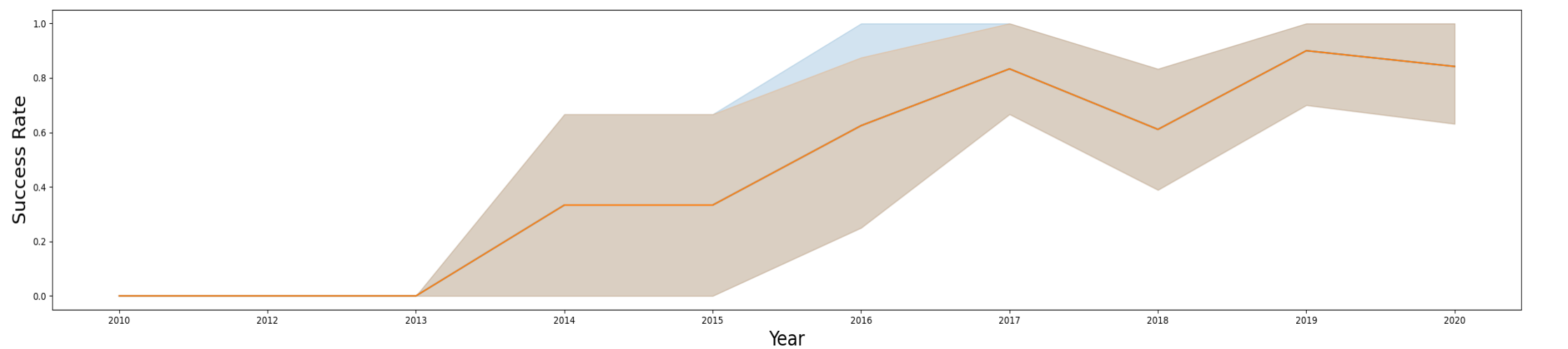
- You can see that in the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

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 success rate since 2013 kept increasing till 2020.

All Launch Site Names

Out[6]:

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

- Unique launch sites in the space mission.

Launch Site Names Begin with 'CCA'

Out[7]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	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)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

- We can see that all of launch sites in LEO orbit and have successful mission outcome.

Total Payload Mass

Out[8]:

Total_Payload_Mass
45596

- Total payload mass carried by boosters launched by NASA (CRS).

Average Payload Mass by F9 v1.1

Out[10]:

Average_Payload_Mass
2928.4

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

First Successful Ground Landing Date

Out[11]:

First_Succesful_Date
01-05-2017

- Date when the first successful landing outcome in ground pad was achieved.

Successful Drone Ship Landing with Payload between 4000 and 6000

Out[13]:

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

- 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

Out[14]:

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

- We can see that most of the missions have succeeded.

Boosters Carried Maximum Payload

Out[15]:

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

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

2015 Launch Records

Out[48]:

substr(Date, 4, 2)	Booster_Version	Launch_Site
01	F9 v1.1 B1012	CCAFS LC-40
04	F9 v1.1 B1015	CCAFS LC-40

- We can see that we have only 2 failed landing outcomes in drone ship for year 2015 and it's in CCAFS LC-40 Launch Site.

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

Out[56]:

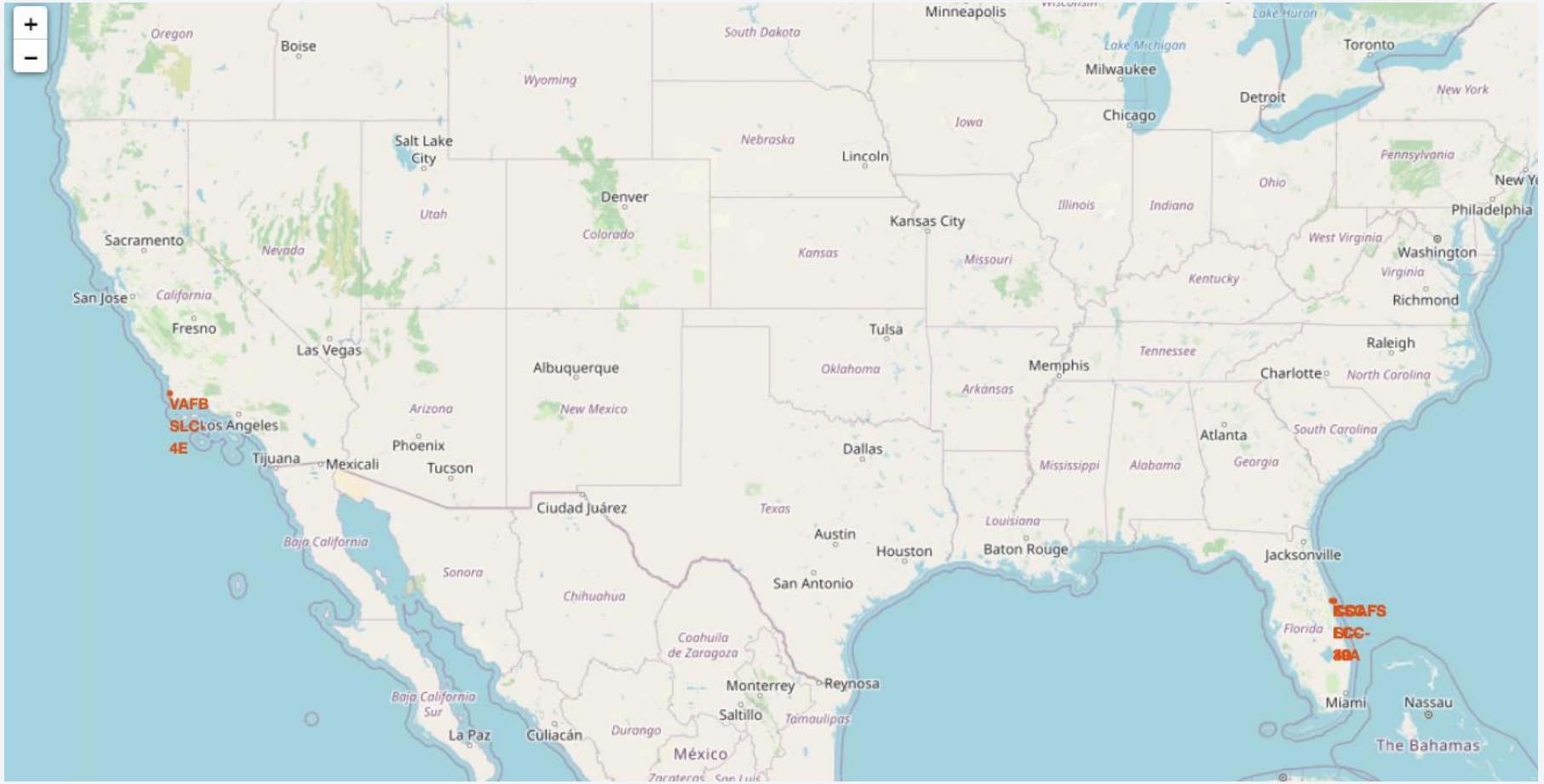
rank	Count_Successful_Landing	Landing _Outcome
1	20	Success
2	8	Success (drone ship)
3	6	Success (ground pad)

- We can see that the highest count of landing success between the date 2010-06-04 and 2017-03-20 is 20.

Section 3

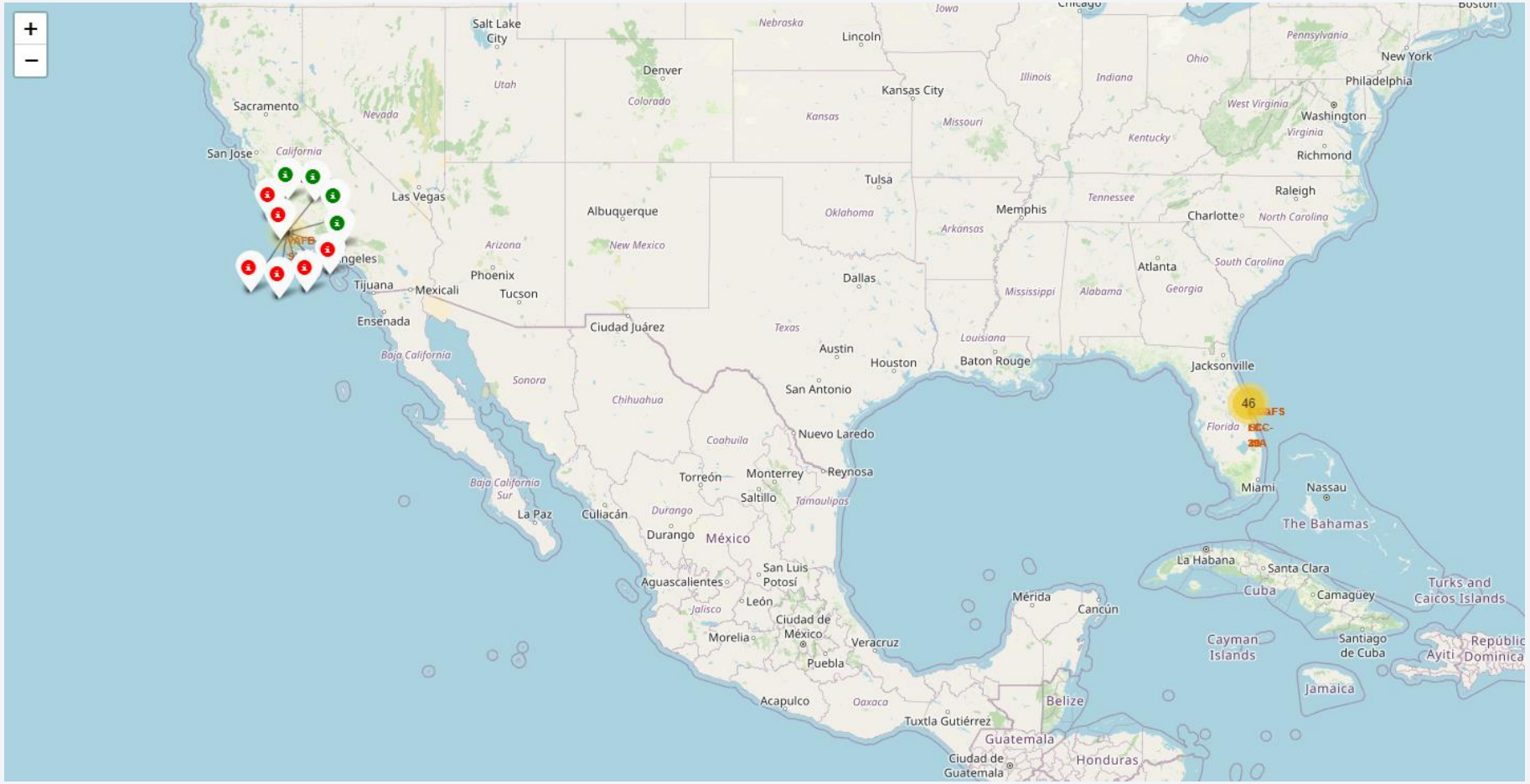
Launch Sites Proximities Analysis

Launch Sites Locations Markers



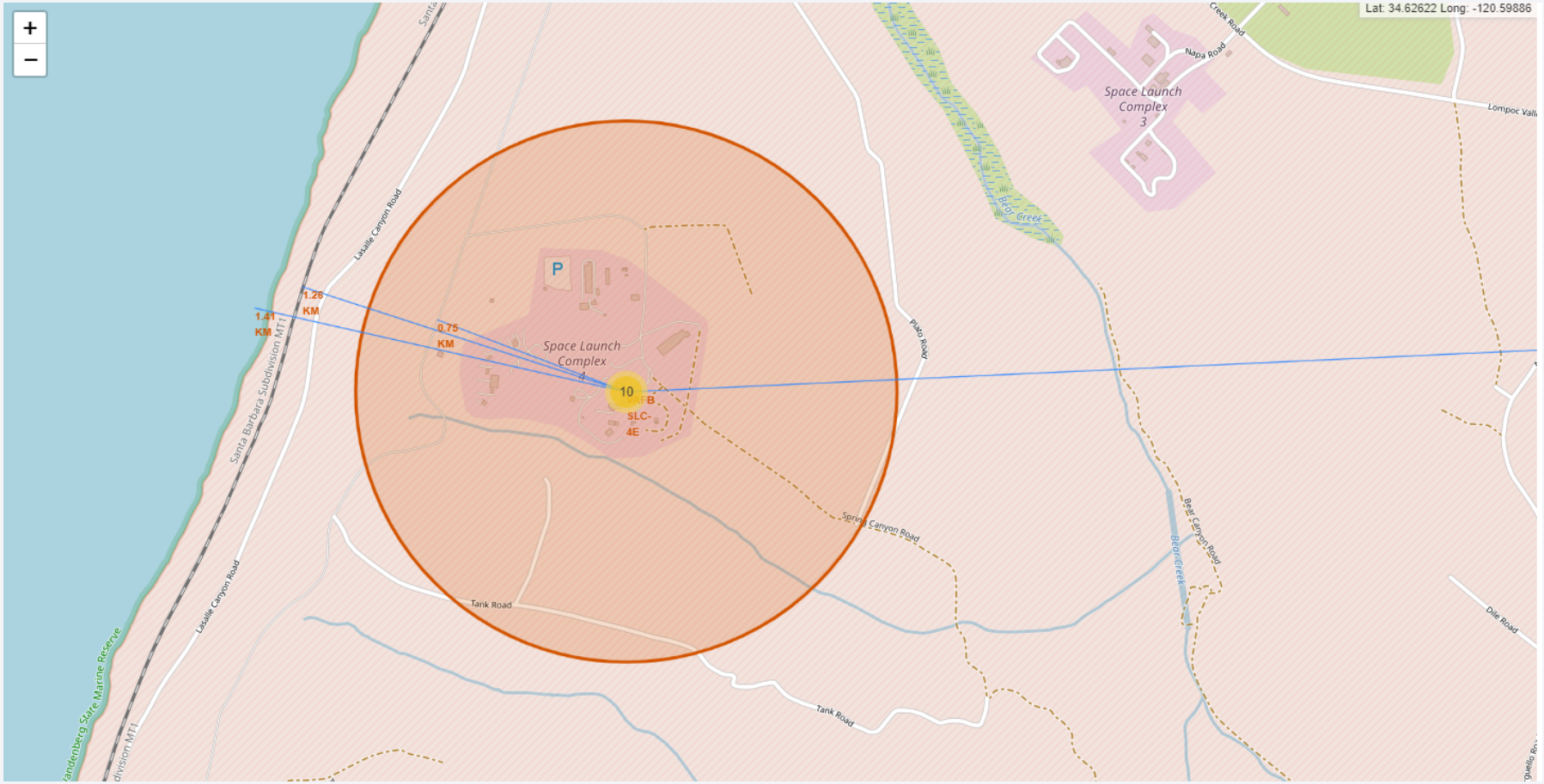
- We can see that all launch sites are near coast.

Launch Outcomes Color Markers

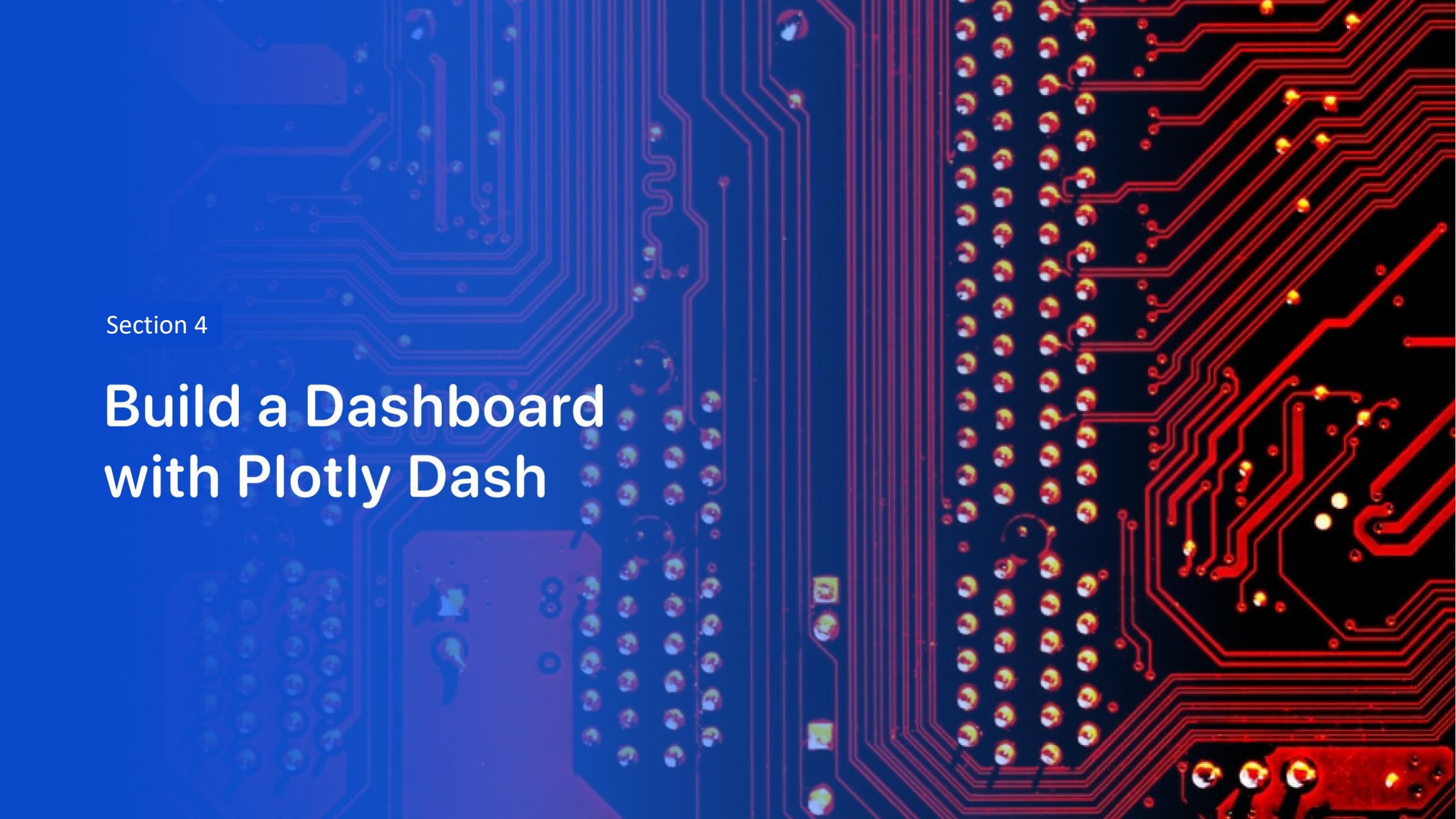


- You can notice that most of launch outcomes on VAFB SLC 4E unsuccessful.

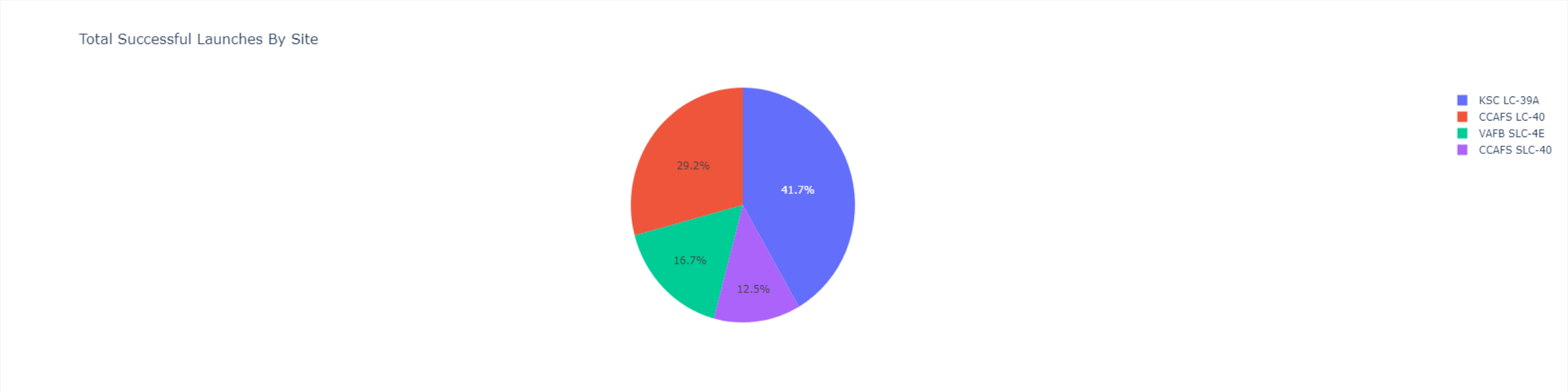
Launch Site Distance to its Proximities Markers



- You can notice that the railway is the nearest to the launch site.

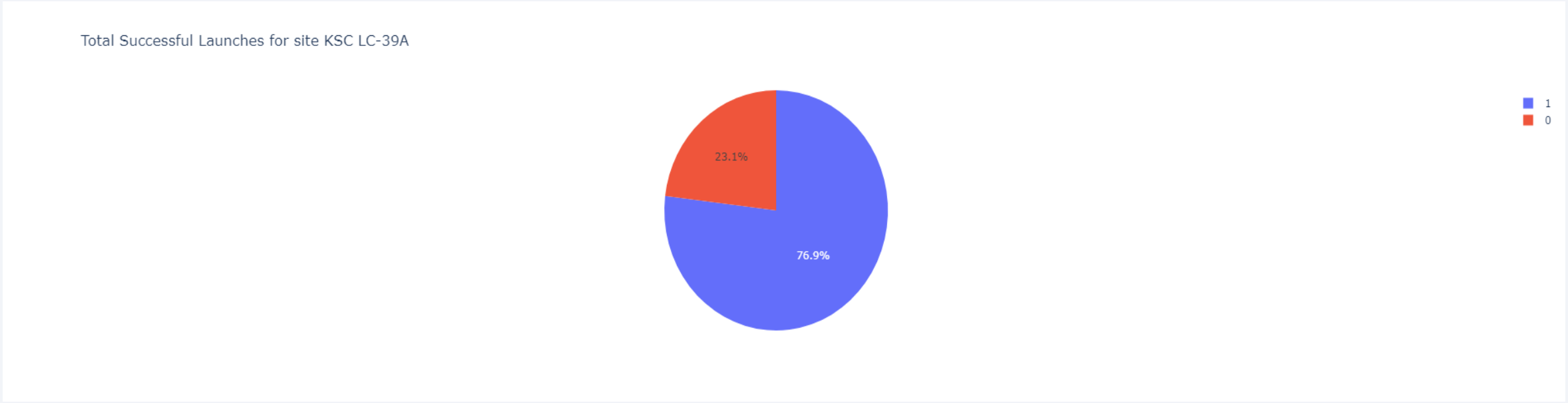

Section 4
Build a Dashboard with Plotly Dash

Total Successful Launches for All Sites



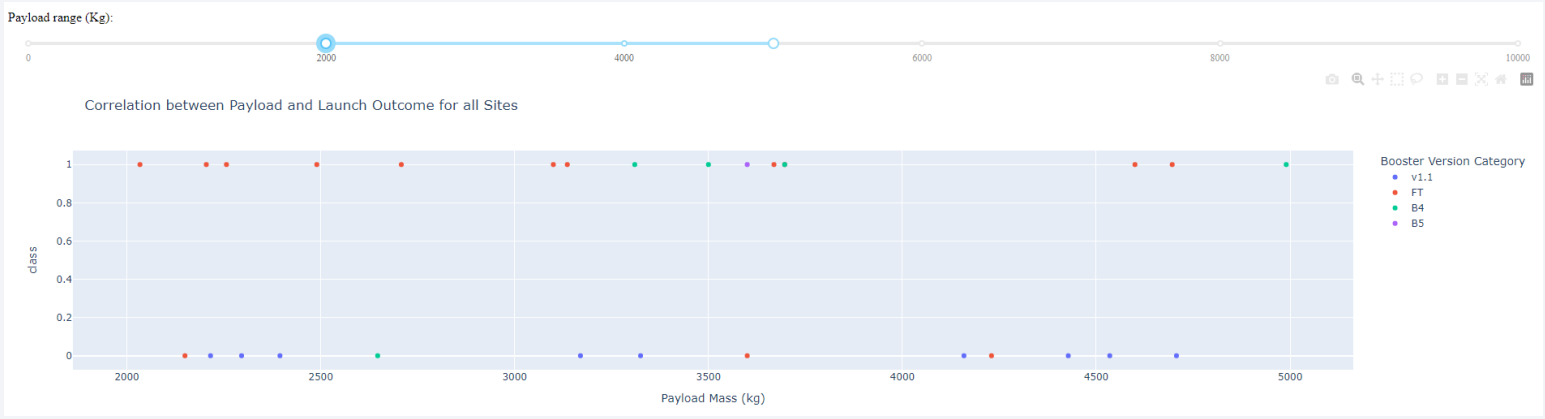
- We notice that KSC LC-39A has the highest success rate for launches.

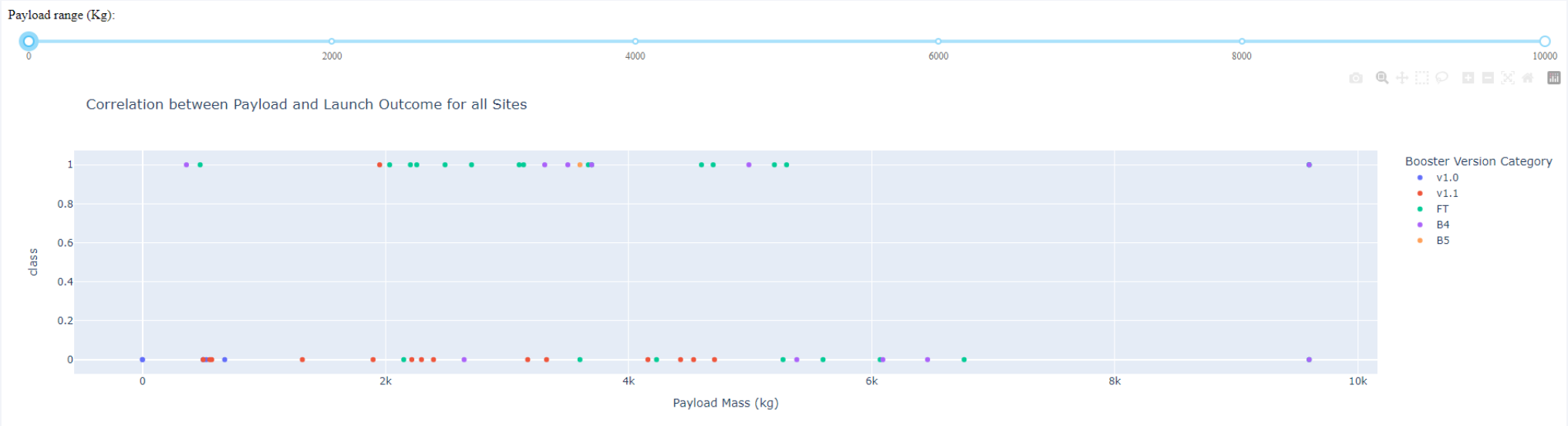
Success Rate for KSC LC-39A Site



- We can see that the success rate is 76.9 percent for the site.

Correlation Between Payload and Launch Outcome for All Sites



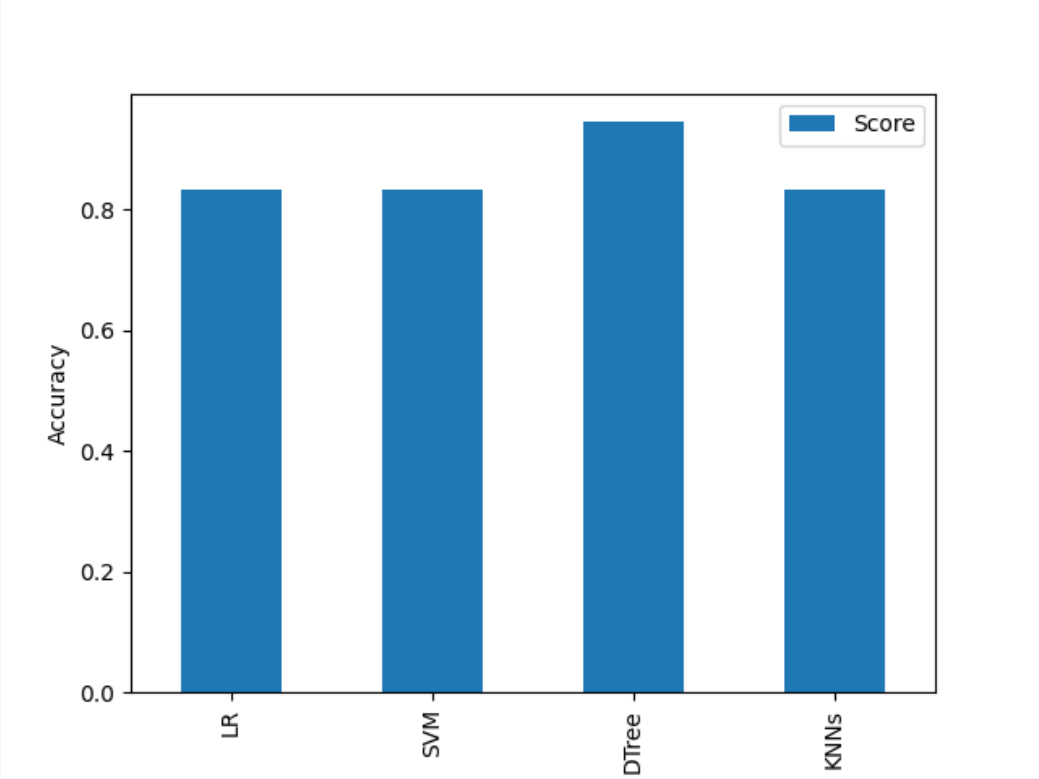


- We notice that FT Booster version has the highest success rate and B4 has the heaviest payload mass.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

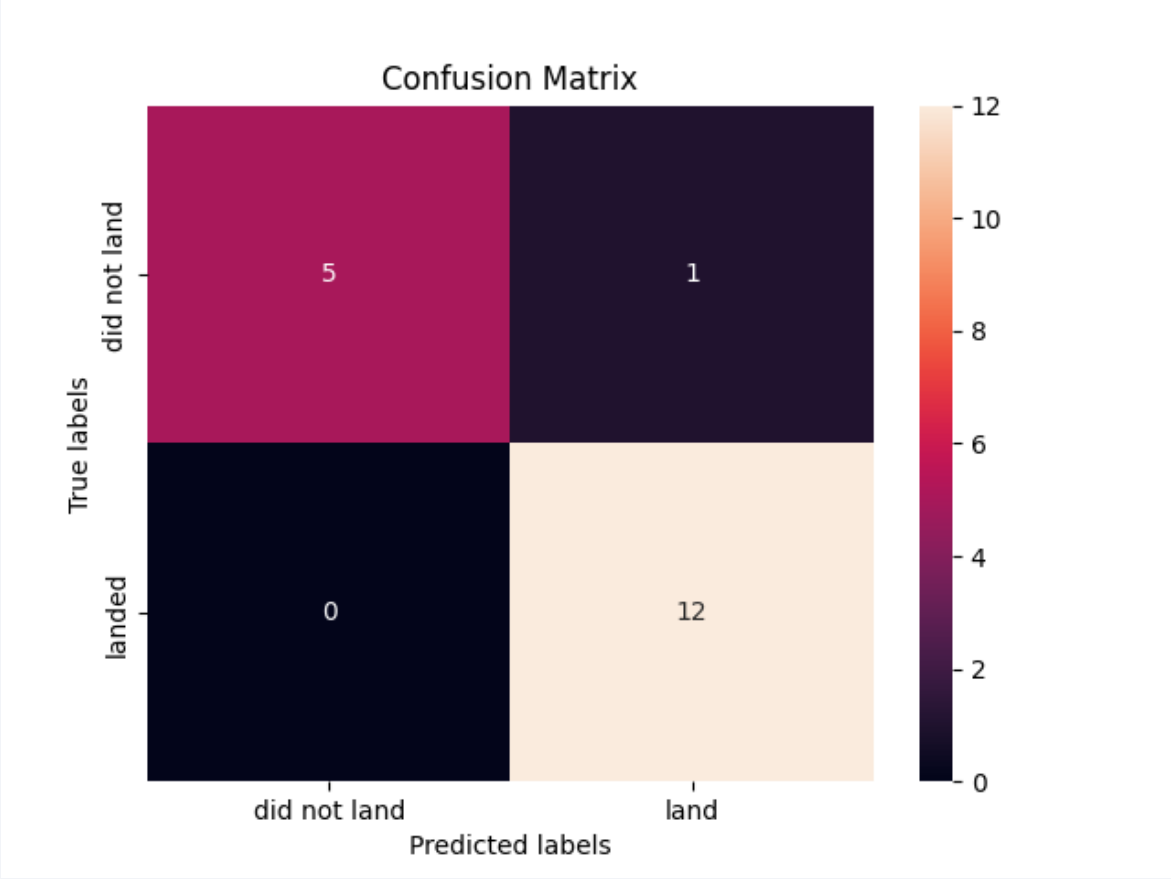


Out[36]:

	Method	Score
0	Logistic Regression	0.833333
1	Support Vector Machine	0.833333
2	Decision Tree	0.944444
3	K Nearest Neighbors	0.833333

- We can see that Decision Tree has the highest score for classification accuracy.

Confusion Matrix



- We notice that only one unsuccessful landing outcome didn't predict correctly by Decision Tree.

Conclusions

- All launch sites are near coast.
- Success rate for Falcon 9 launches since 2013 kept increasing till 2020.
- Most of the missions have succeeded which are 98 in total.
- KSC LC-39A site has the highest success rate for launches.
- Decision Tree method has the best accuracy for predicting landing outcome.

Appendix

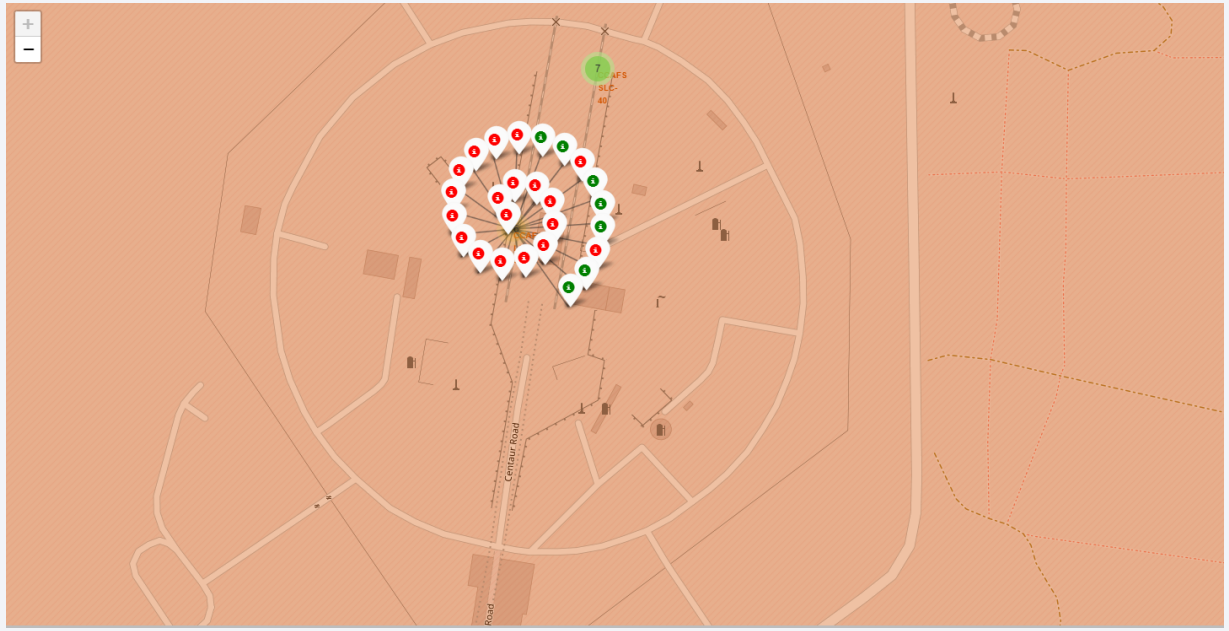
In [56]:

```
%%sql
SELECT
RANK() OVER(ORDER BY COUNT("Landing _Outcome") DESC) AS rank,
COUNT("Landing _Outcome") AS Count_Successful_Landing,
"Landing _Outcome"
FROM SPACEXTBL
WHERE Date BETWEEN "04-06-2010" AND "20-03-2017"
AND "Landing _Outcome" LIKE '%Success%'
Group by "Landing _Outcome";
```

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

Out[56]:

rank	Count_Successful_Landing	Landing _Outcome
1	20	Success
2	8	Success (drone ship)
3	6	Success (ground pad)



- Rank count of successful landing outcomes between the date 04-06-2010 and 20-03-2017 in descending order query.
- Cluster Marker of success/failed launches for CCAFS LC-40 site.

Thank you!