Winning Space Race with Data Science

Samuel Pomajevich
03/29/2022

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection
  - Data Wrangling
  - Exploratory Data Analysis with Data Visualization
  - Exploratory Data Analysis with SQL
  - Interactive Visual Analysis with Folium
  - Interactive Visual Analysis with Plotly Dash
  - Machine Learning Predictions
- Summary of all results
  - Exploratory Data Analysis Results
  - Interactive Analytics Results
  - Predictive Analytics Results

Introduction

- Project background and context
  - 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.
- Problems you want to find answers
  - What features determine if the rocket will land successfully?
  - How features when analyzed together, affect the success of landing?
  - What features can our new company use to most efficiently produce successfully landings?

Section 1

Methodology

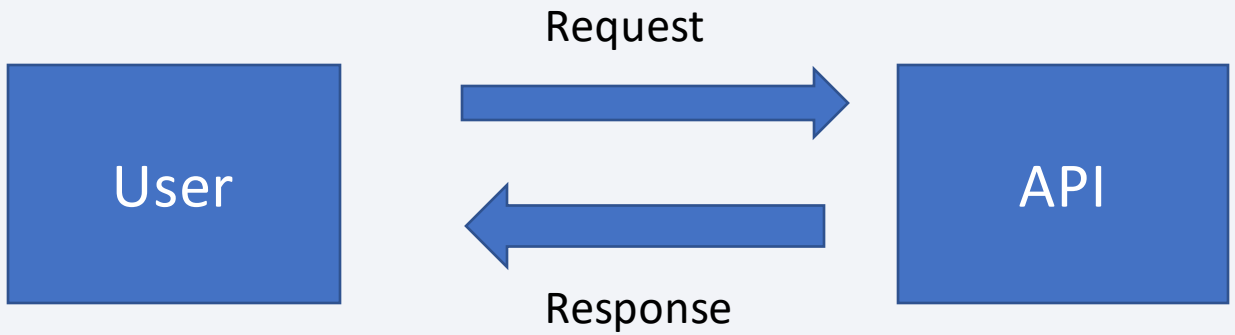
Methodology

Executive Summary

- Data collection methodology:
  - SpaceX API
  - List of Falcon 9 and Falcon Heavy Launches Wikipedia
- Perform data wrangling
  - One-hot encoding was applied to the categorical values
  - The empty values (NaN) were assigned an averaged value
- 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

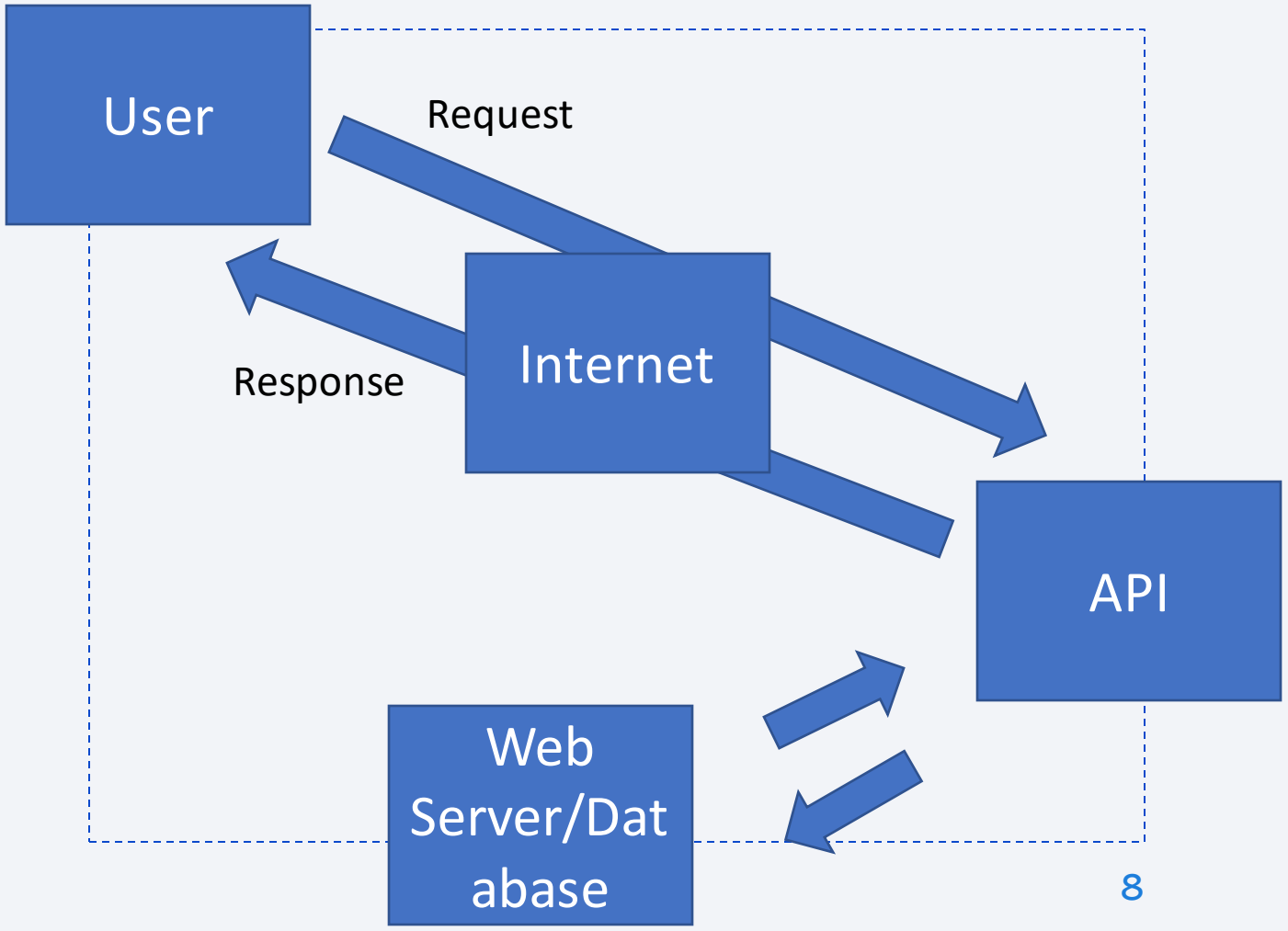
Data Collection

- Using the library "requests" the data was collected using a "GET" request from the SpaceX API.
- The API provided the launch site, payload mass, longitude and latitude of launch, booster version, and multiple other additional features.
- The address of the API used was https://api.spacexdata.com/v4/
- Data was also web scraped from the Wikipedia page: "List of Falcon 9 and Falcon Heavy launches"



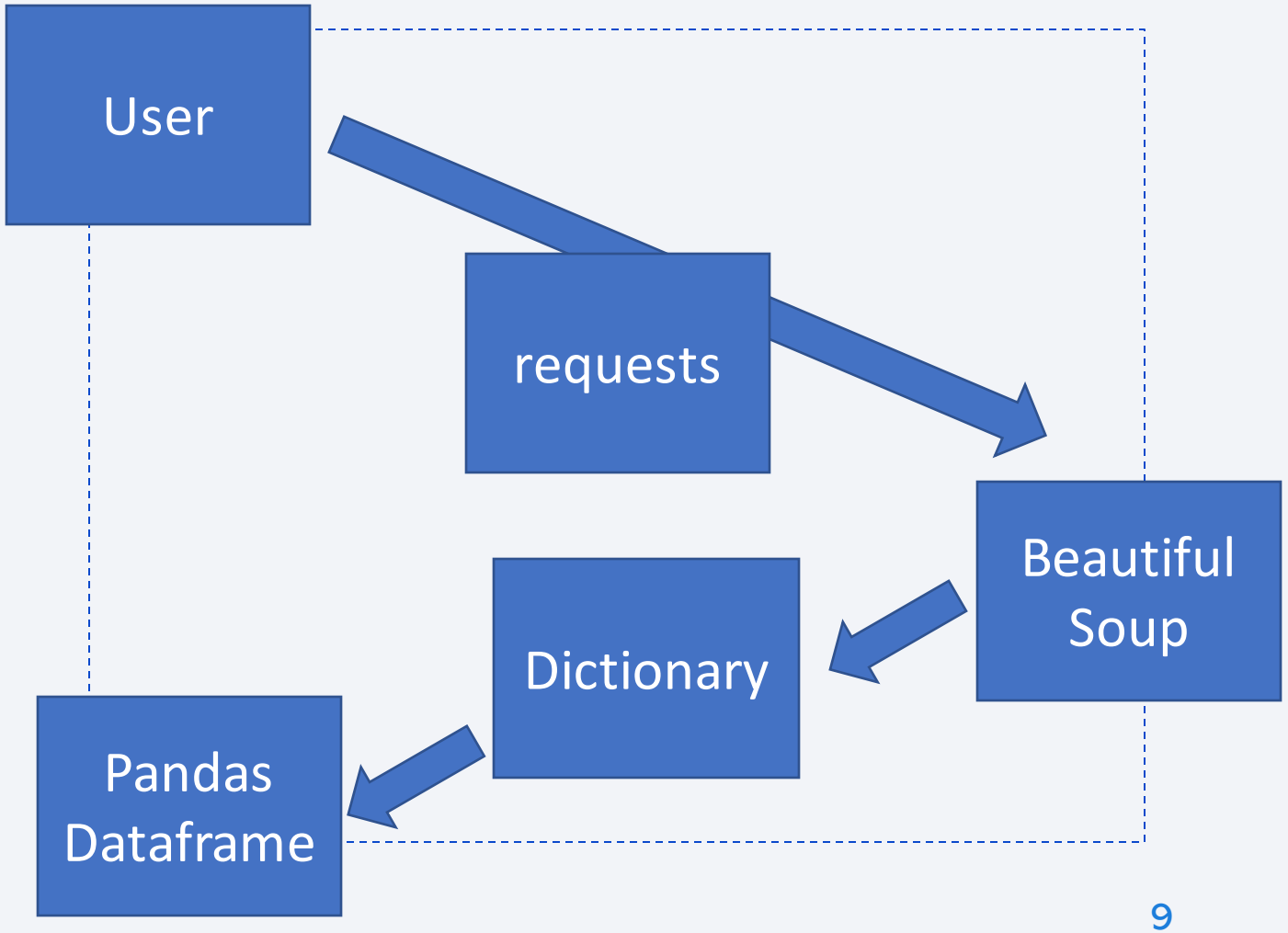
Data Collection – SpaceX API

- Collected data from SpaceX API using GET request from "requests" package.
- Data was converted from json to a pandas dataframe to then clean and wrangle.
- GitHub URL with SpaceX API



Data Collection - Scraping

- Used requests to get eh html of a page.
- Then used BeautifulSoup to parse the html of the Falcon 9 launch Wikipedia page.
- Went on to use a dictionary to organize the data.
- Converted the dictionary into a pandas dataframe.
- GitHub URL to Scraping



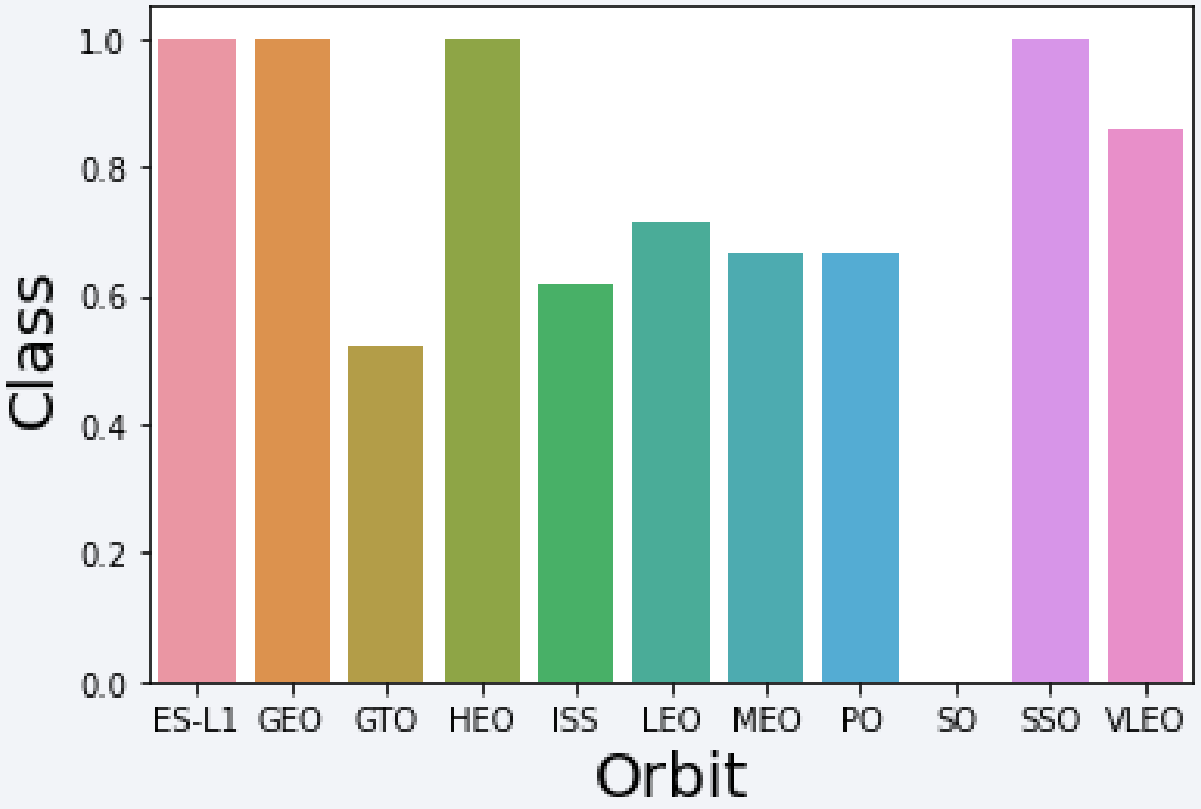
Data Wrangling

- Removed null values from the payload mass data and replaced it with the mean payload mass.
- Created a dataframe column "Class" containing binary values for a successful or failed landing outcome.
- This was completed by classifying the failed outcomes and creating a set of them to then compare to the dataframe.
- GitHub URL Demonstrating Web Scraping

EDA with Data Visualization

- Data was explored using Seaborn, Matplotlib, and the features of the dataframe.
- Composed multiple scatter plots to demonstrate relationship between a feature and a successful landing outcome.
- GitHub URL with EDA with Data Visualization

Success rate of each orbit type of the Falcon 9 rocket.



EDA with SQL

- Used SQL and IBM's DB2 to select and view multiple pieces of the data set.
- Queried the distinct launch sites.
- Queried data launched by NASA (CRS)
- Calculated average payload mass
- Found oldest date of landing success on a ground pad
- GitHub URL with EDA with SQL

Build an Interactive Map with Folium

- Built an Interactive Map with Folium to visualize the launch sites of every launch In our data set.
- Used the longitude and latitude of each launch to mark it on the interactive map with a GREEN marker for success and a RED marker for failure.
- Then calculated the distance to multiple different landmarks such as the coastline, nearest road, etc.
- GitHub URL for Interactive Map with Folium

Build a Dashboard with Plotly Dash

- Built an interactive dashboard using Plotly Dash
- Pie chart showing the percentage of launches by launch site
- Scatter plot with Outcome vs. Payload Mass (kg)
  - This enables an interactive way to examine the relationship between success and the payload mass.
- Plotly Dash enables the user to interact with the data and get a better understanding of the relationships and importance of features.

Predictive Analysis (Classification)

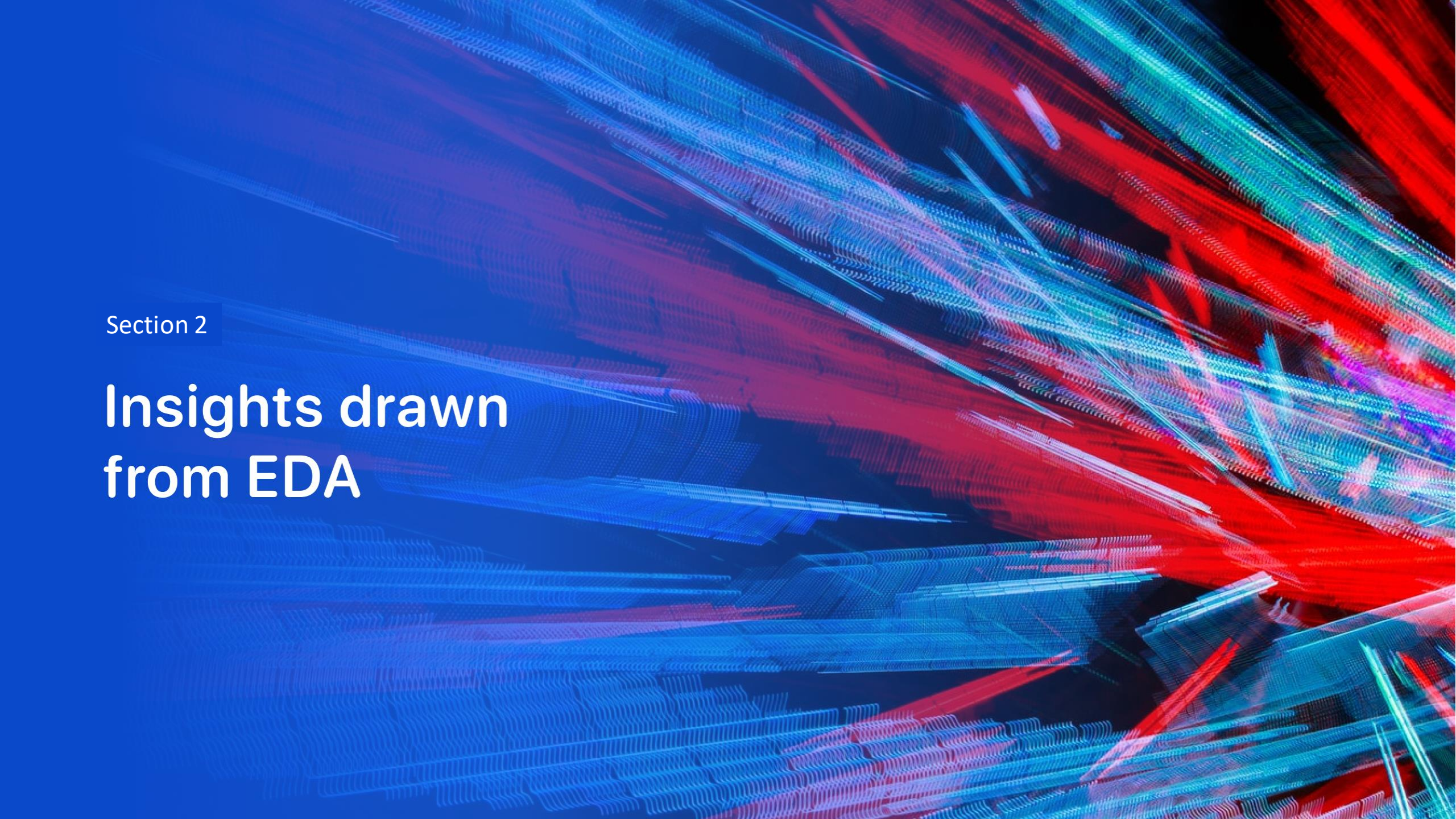
- Loaded data using Numpy and Pandas
- Standardized the data to better represent what is needed
- Split the data into train and test sets
- Used GridSearchCV for multiple types of models including:
  - Logistic regression
  - Support vector machines
  - Decision tree
  - K nearest neighbors
- GitHub URL to Predictive Analysis

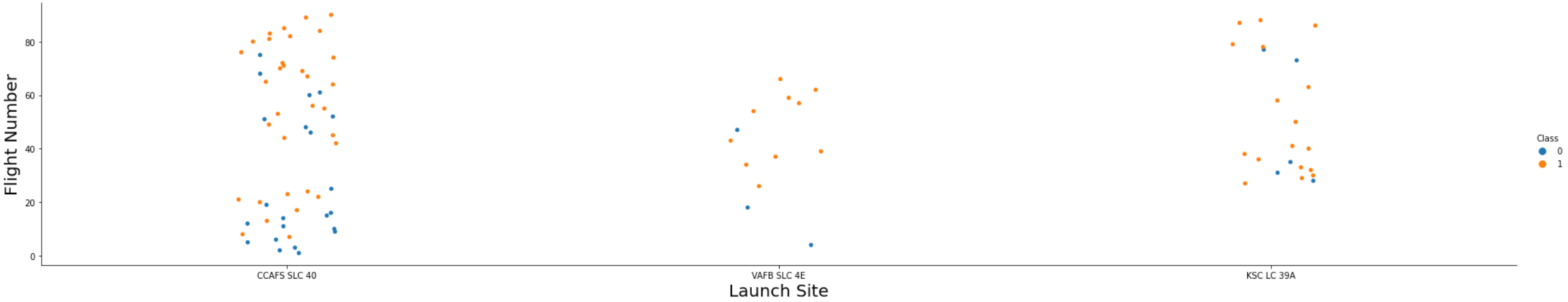
Results

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

Section 2

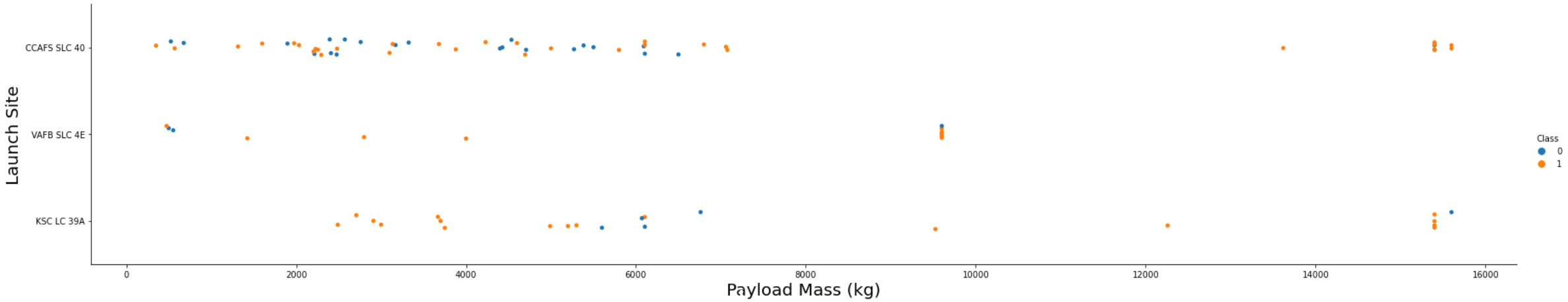
Insights drawn from EDA





Flight Number vs. Launch Site

- From the above graph, it can be seen that KSC LC 39A has most of the higher flight numbers, VMFB SLC 4E had many of the midflight number and CCAFS SLC 40 had early and late flight numbers.

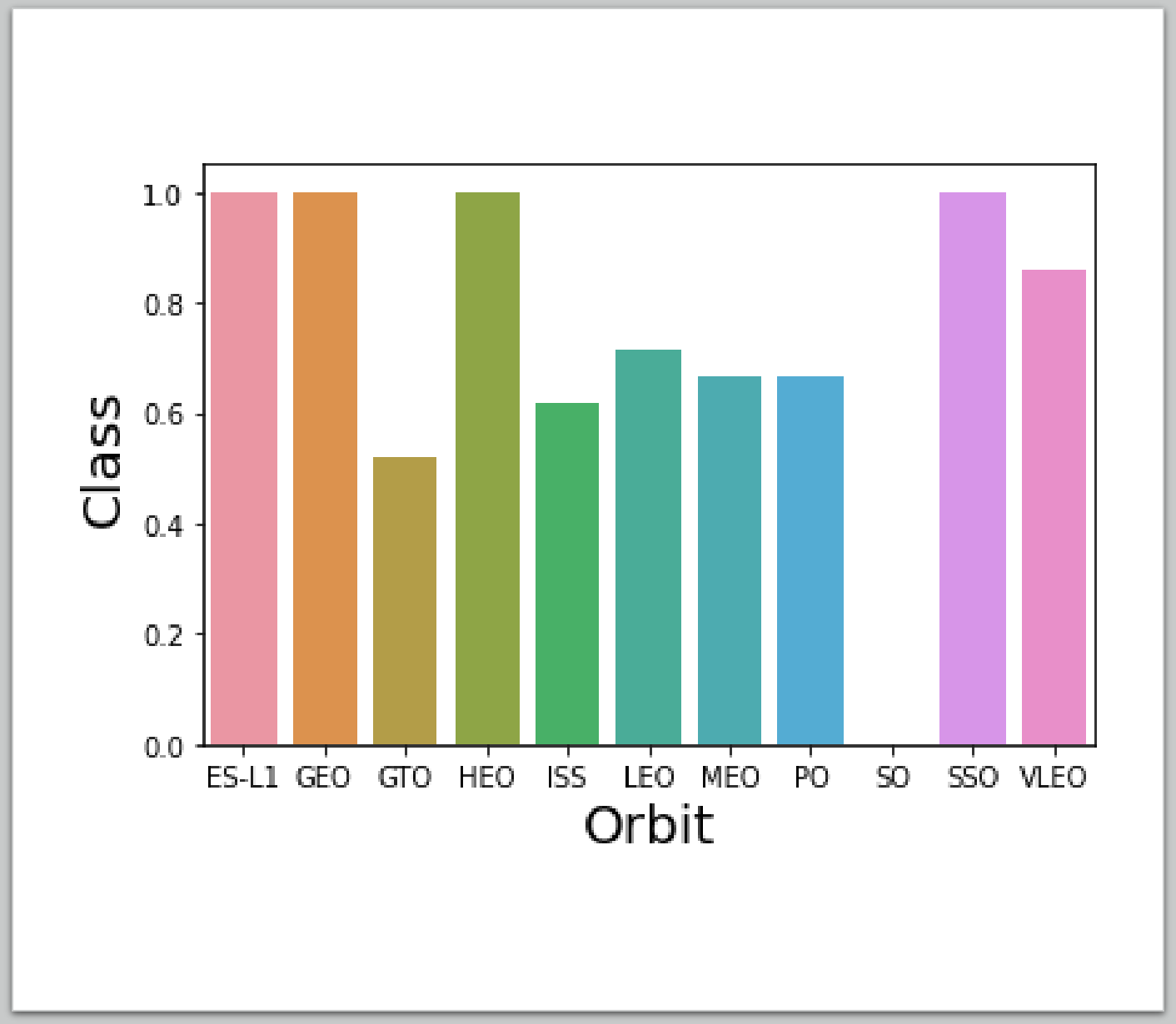


Payload vs. Launch Site

- It can be seen that the higher payload masses had less launches and some launch sites did not launch larger payload mass flights,].

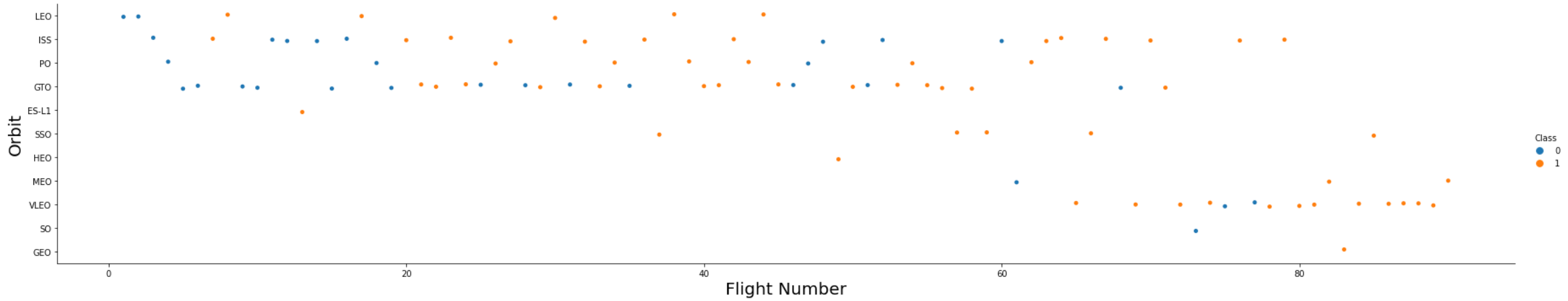
Success Rate vs. Orbit Type

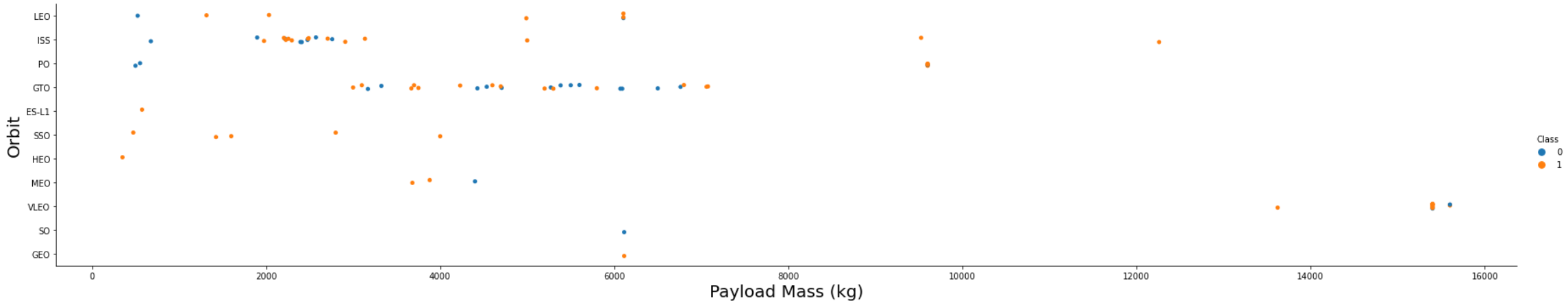
- There are multiple orbit types with an average of 1 for success which is equivalent to 100% success. It may appear that if we have these orbit types, the landing will be successful.



Flight Number vs. Orbit Type

- It can be seen that most early flight numbers reached one group of orbit type
- The later flight numbers were in another orbit type



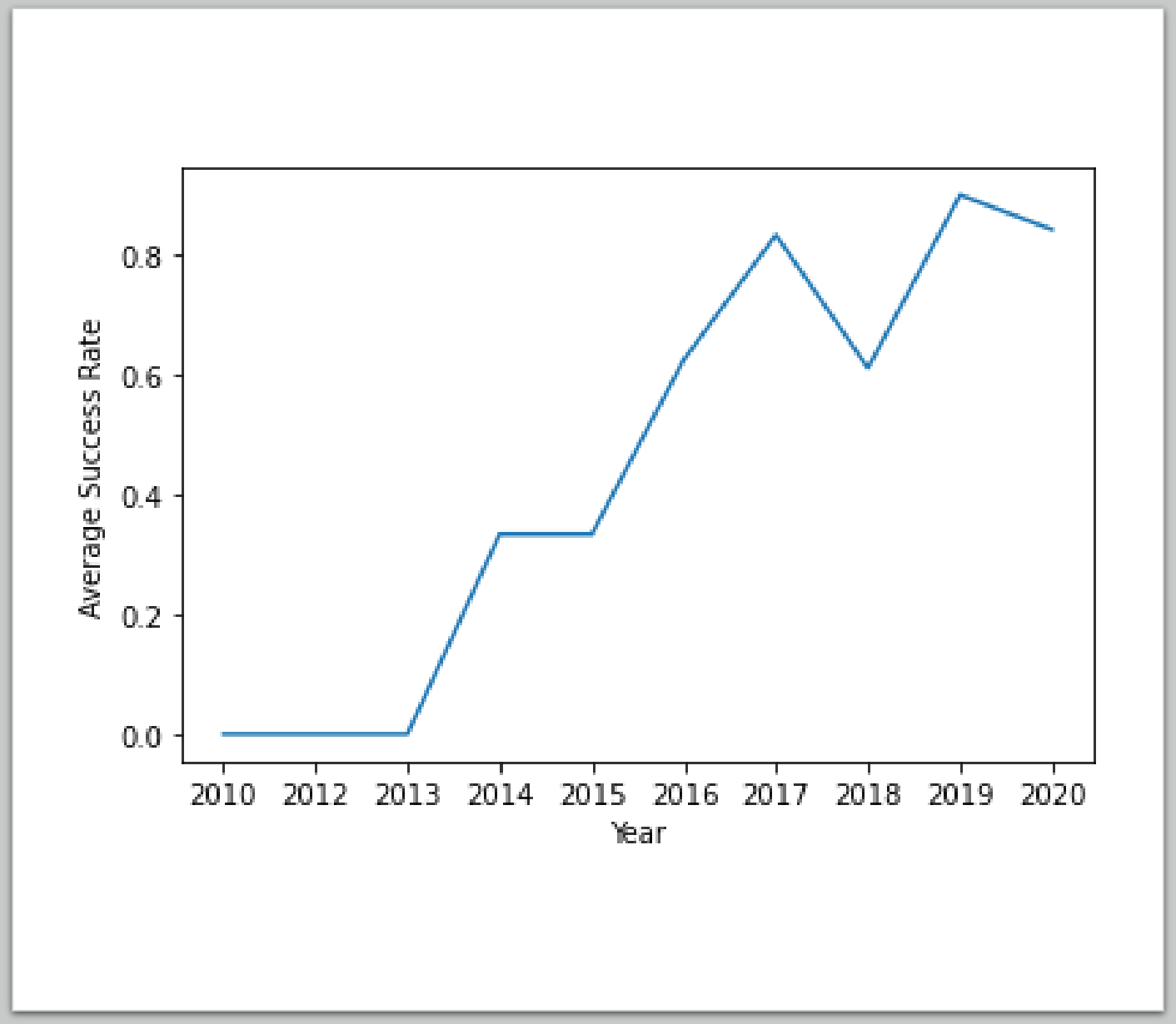


Payload vs. Orbit Type

- Again, we see that there were not many flights with greater payload masses. IT can be seen that some orbit types had none of the greater payload mass flights.

Launch Success Yearly Trend

- It is seen that as time has gone by, the success rate has increased
- The highest success rate was in 2019



All Launch Site Names

SQL Query:

%**sql** SELECT DISTINCT launch_site FROM SPACEXTBL;

Launch Site
CCAFS LC-40
CCAFS SLC-40
KSC LC-39A
VAFB SLC-4E

Launch Site Names Begin with 'CCA'

SQL Query:

```
%sql SELECT * FROM SPACEXTBL
   WHERE launch_site LIKE 'CCA%'
   LIMIT 5;
```

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

Using SQL Queries:

- %**sql** SELECT SUM(payload_mass__kg_) FROM SPACEXTBL WHERE customer = 'NASA (CRS)';

The total payload mass by boosters lanched with NASA (CRS) is 45596 kg.

Average Payload Mass by F9 v1.1

- Using SQL Queries:
  - %**sql** SELECT AVG(payload_mass__kg_) FROM SPACEXTBL WHERE booster_version = 'F9 v1.1';

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

First Successful Ground Landing Date

Using SQL Queries:

```
%sql SELECT MIN(DATE) FROM SPACEXTBL WHERE landing__outcome = 'Success (ground pad)';
```

The first successful ground landing date is December 22, 2015

Successful Drone Ship Landing with Payload between 4000 and 6000

- SQL Query:
- %**sql** SELECT booster_version FROM SPACEXTBL WHERE landing__outcome = 'Success (drone ship)' AND (payload_mass__kg_ BETWEEN 4000 AND 6000);

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

Total Number of Successful and Failure Mission Outcomes

- SQL Query:
- **%sql** SELECT mission_outcome,count(mission_outcome) as COUNT from SPACEXTBL GROUP BY mission_outcomer

mission_outcome	COUNT
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- SQL Query:
- %**sql** SELECT DISTINCT booster_version FROM SPACEXTBL WHERE payload_mass__kg_ = (SELECT MAX(payload_mass__kg_) FROM SPACEXTBL);

booster_version
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
F9 B5 B1051.3
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

2015 Launch Records

- Failed landing on a drone ship in 2015
- %**sql** SELECT landing__outcome, booster_version, launch_site FROM SPACEXTBL WHERE DATE LIKE '2015%' AND landing__outcome = 'Failure (drone ship)';

landing__outcome	booster_version	launch_site
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

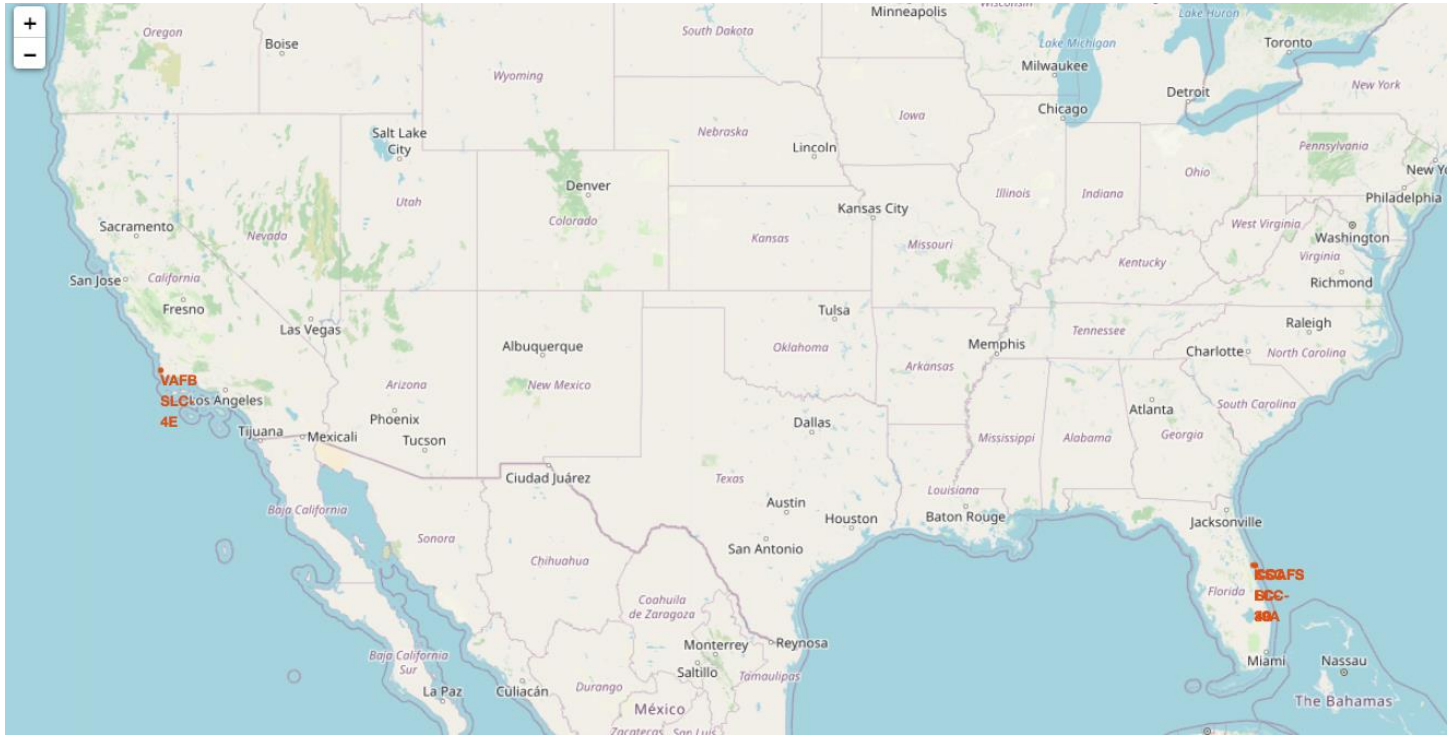
- %**sql** SELECT landing__outcome, COUNT(*) AS COUNT FROM SPACEXTBL WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY landing__outcome ORDER BY COUNT DESC;

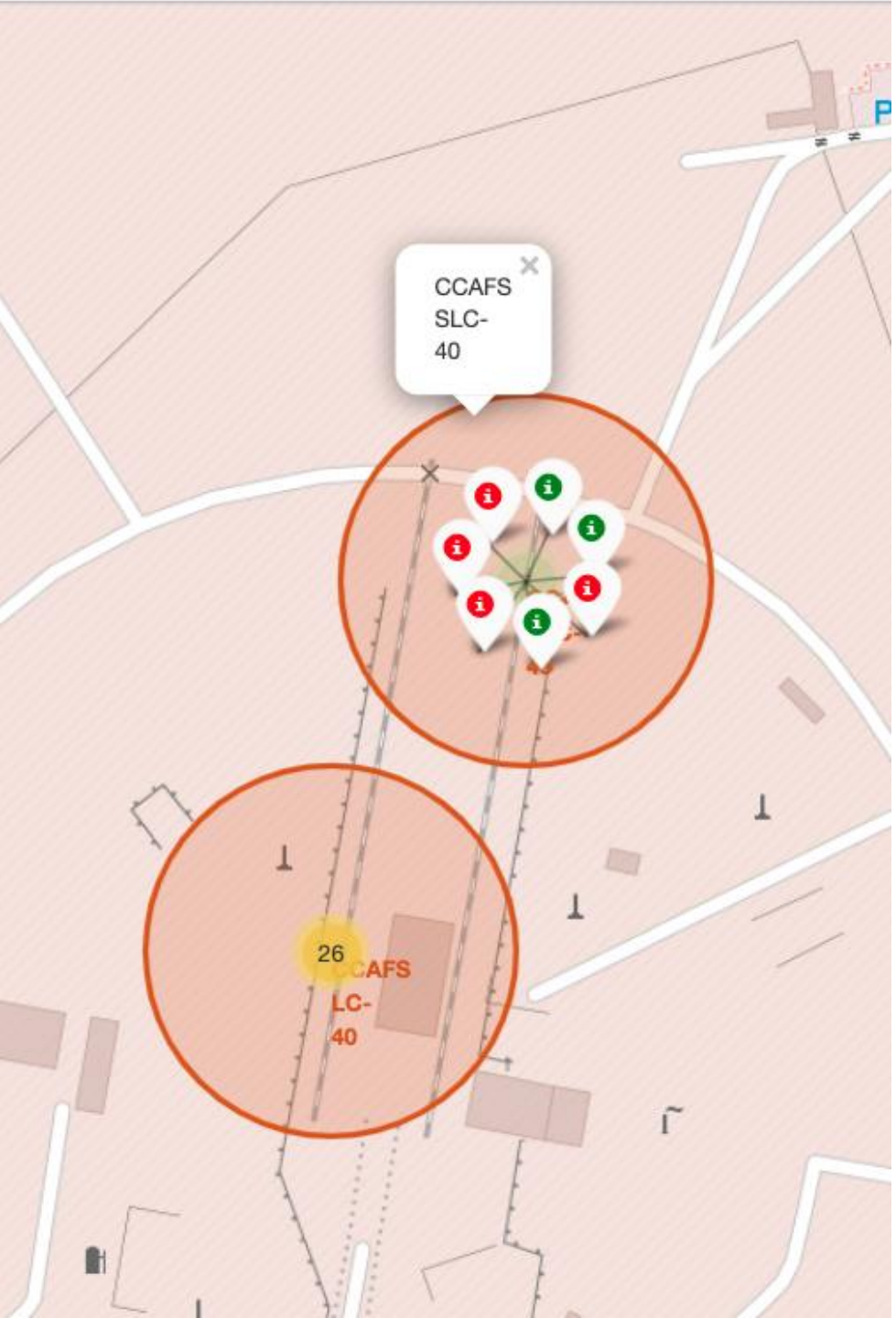
landing__outcome	COUNT
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1


Section 3
Launch Sites
Proximities Analysis

Launch Locations with Folium

- Locations of all three launch sites in the United States



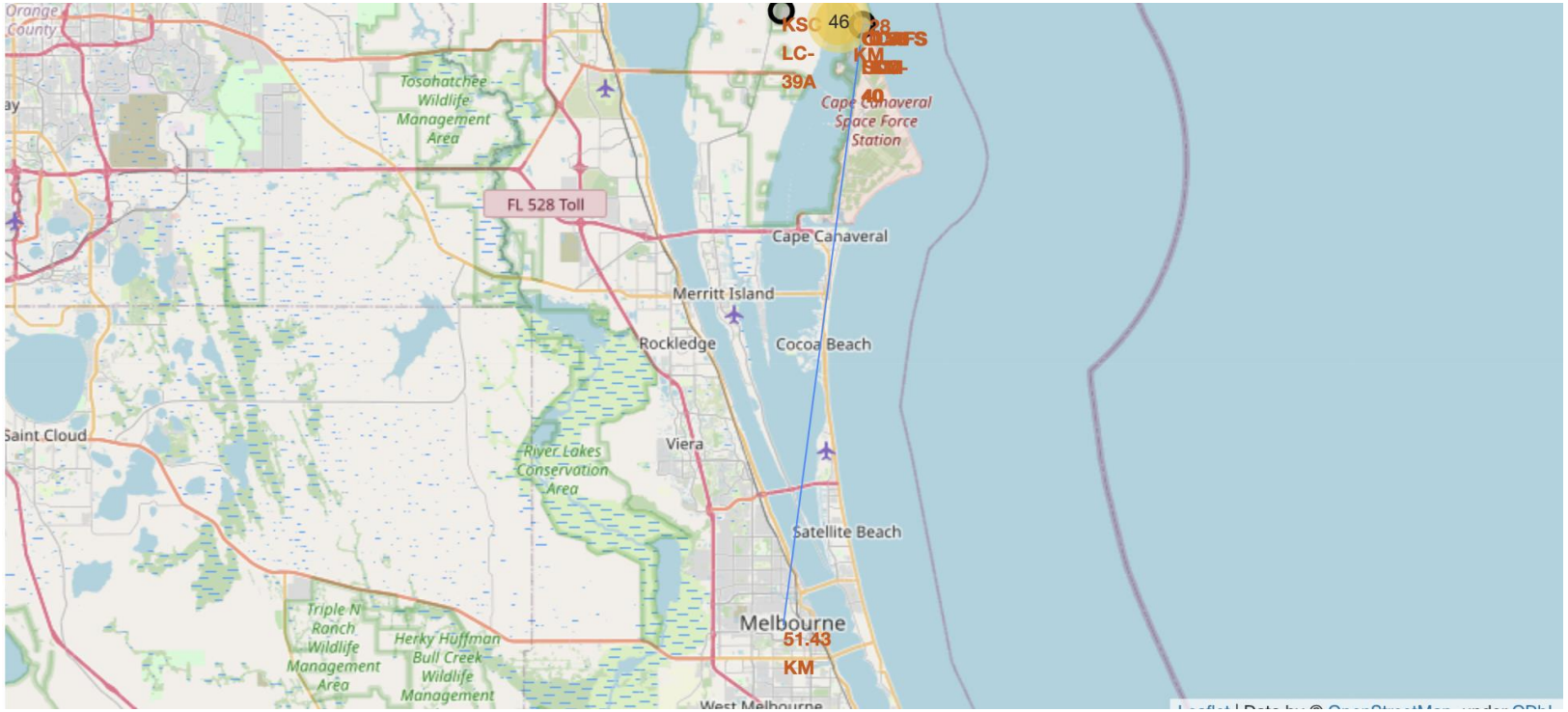


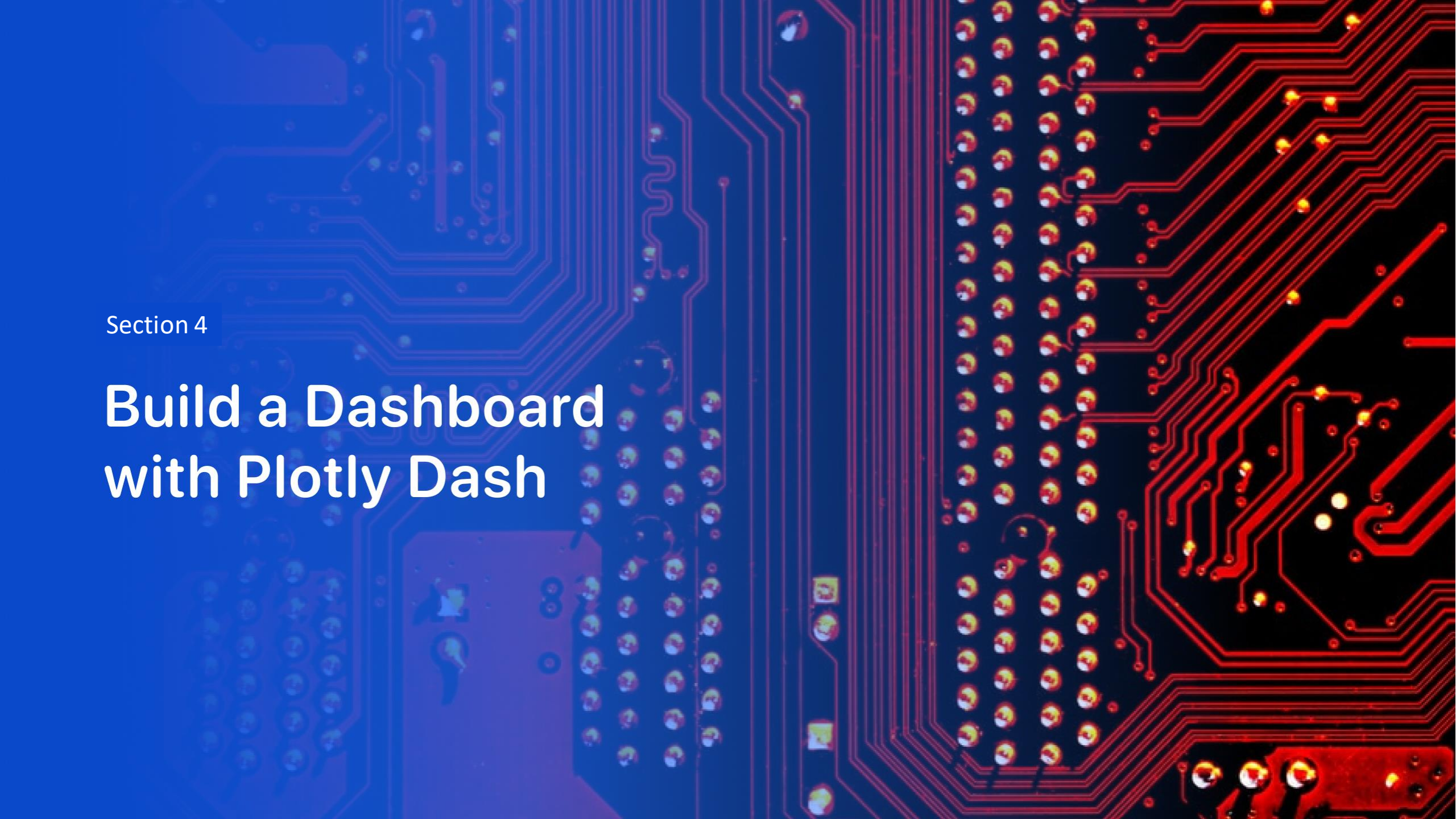
Launch Locations Success/Failures

- Green shows that the launch resulted in a successful landing
- Reg shows that the launch resulted in a successful landing

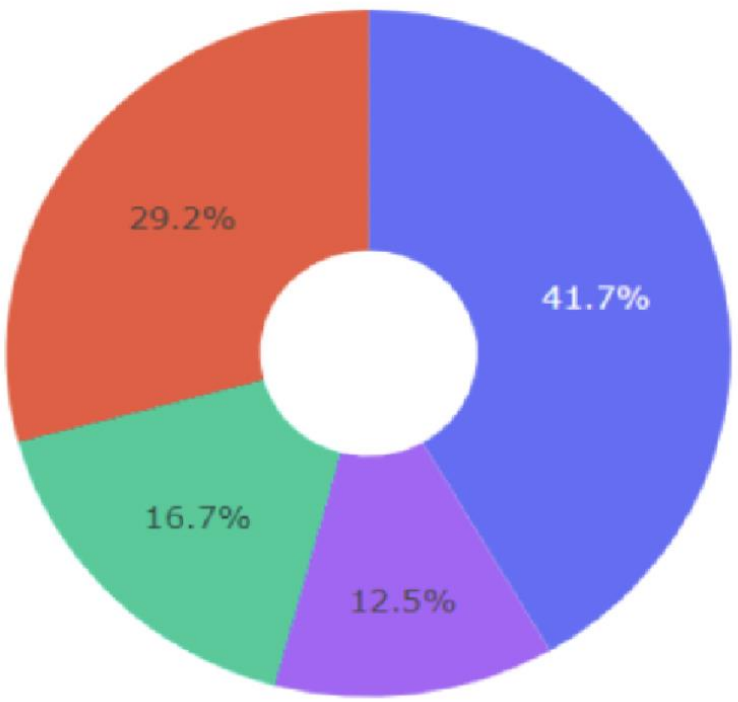
<Folium Map Screenshot 3>

- Measured distance to a nearby city.



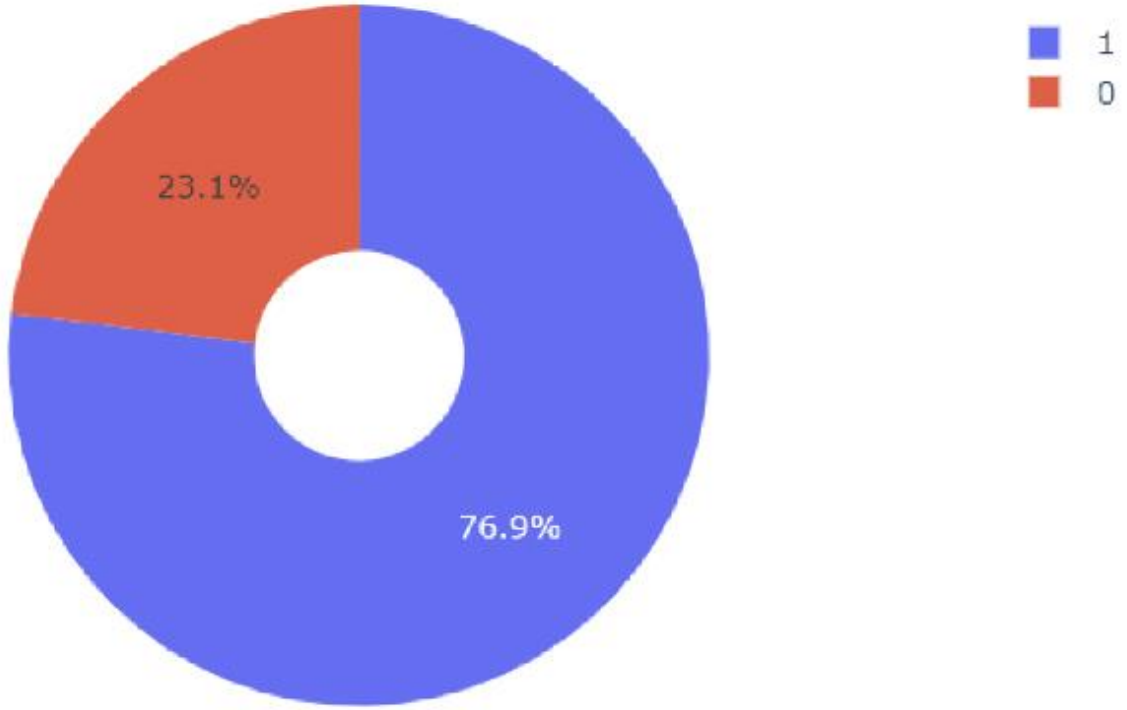

Section 4
Build a Dashboard with Plotly Dash

Total Successful Launches by all Launch Sites



KSC LC-39A has the most successful
launch percentage.

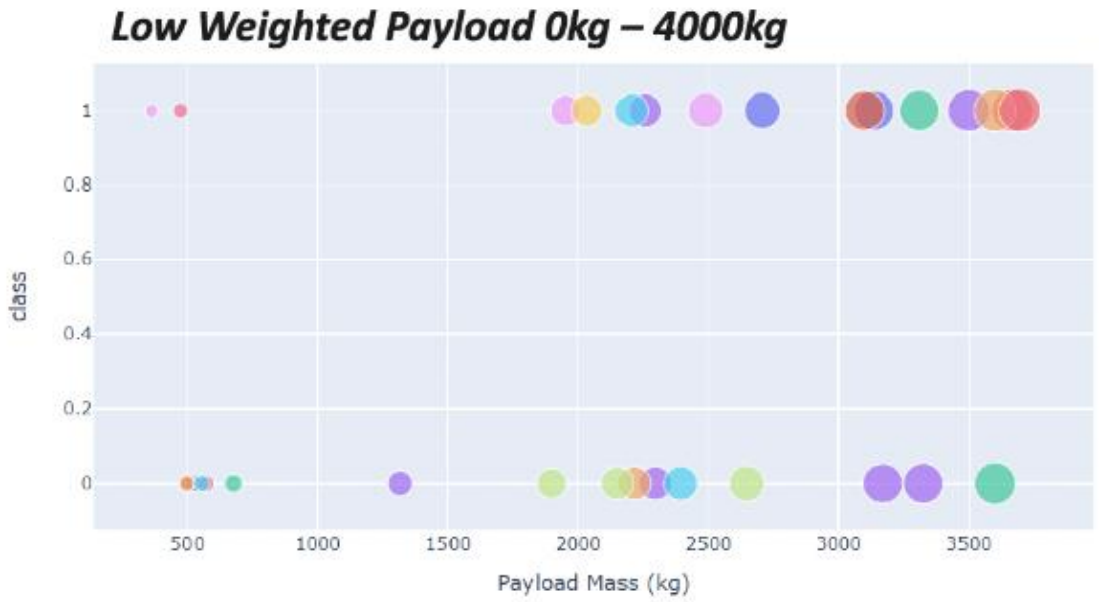
Success Rate of Launch Site: KSC LC-39A

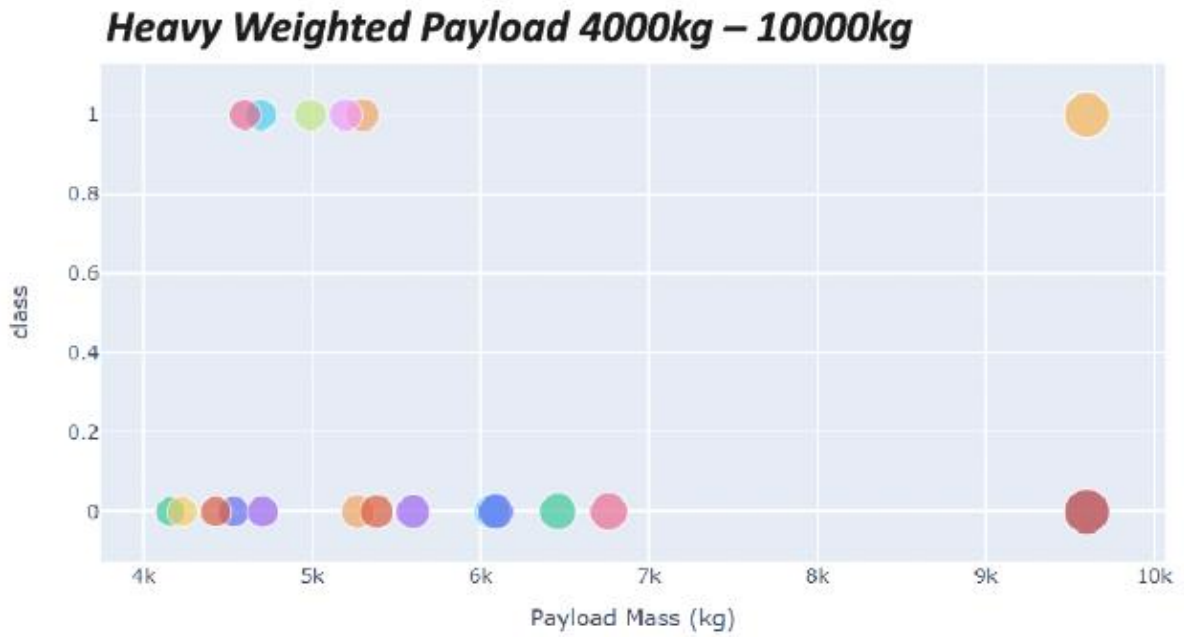


KSC LC-39A achieved a 76.9% success rate while getting a 23.1% failure rate

Payload Mass and Success Rates

- The lighter payload mass flights had a greater success rate





Section 5

Predictive Analysis (Classification)

Classification Accuracy

- Using the "score" method, the accuracy for each predictive method was determined to be the same.

Accuracy of logistic regression method: 0.8333333333333334
Accuracy of support vector machine method: 0.8333333333333334
Accuracy of decision tree method: 0.8333333333333334 Accuracy of k nearest neighbors method: 0.8333333333333334

Confusion Matrix

- The Confusion Matrix shows that our model was fairly accurate.
- The problem looks to be false positives

Conclusions

- Through completing predictive analysis which 4 different methods, all methods had the same accuracy using the "score" method.
- As the years have passed, SpaceX has become more successful in landing their Falcon 9 rockets.
- Lighter payloads had more success than heavier payload.
- KSC LC-39A was the launch site with the most successful landings.

Thank you!