IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Patrick Neumann
7/20/2024

Outline

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

Executive Summary

- Space X 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 Space X 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 space X for a rocket launch.
- This project used data directly available from Space X to build models to predict the outcome of a mission.
- An adequate performing model was found

Introduction

- Space X 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 Space X 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 space X for a rocket launch.
- This project used data directly available from Space X to build models to predict the outcome of a mission.
- An adequate performing model was found

Section 1

Methodology

Methodology

- Data collection methodology:
  - Raw data was collected from the SpaceX homepage via web scraping and APIs
- Perform data wrangling
  - The outcome field was transformed in a landing_class field and null values were removed
- Performed exploratory data analysis (EDA) using visualization and SQL
- Performed interactive visual analytics using Folium and Plotly Dash
- Performed predictive analysis using classification models
  - Multiple different models (e.g. k nearest neighbours, decision trees, support vector machines, logistic regression) were evaluated for the predictive analysis and the best performing model was selected

Data Collection – SpaceX API

```
static_json_url='https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/API_call_spacex_api.json'
```

We should see that the request was successfull with the 200 status response code

```
response.status_code
```

```
200
```

Now we decode the response content as a Json using `.json()` and turn it into a Pandas dataframe using `.json_normalize()`

```
data = pd.json_normalize(response.json())
```

Using the dataframe `data` print the first 5 rows

```
data.head()
```

- https://github.com/Patrick912/DataScienceCapstone/blob/master/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping

- BeautifulSoup was used to scrape SpaceX launch data from the Falcon 9 and Falcon Heavy Wiki page

- https://github.com/Patrick912/DataScienceCapstone/blob/master/jupyter-labs-webscraping.ipynb

Data Wrangling

- As part of the data wrangling process
  - The outcome field was transformed in a landing_class field

- https://github.com/Patrick912/DataScienceCapstone/blob/master/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- As part of exploratory data analysis various scatter plots, bar and line charts were plotted
- https://github.com/Patrick912/DataScienceCapstone/blob/master/edadataviz.ipynb

EDA with SQL

- SQL queries were used to calculated averages and maximums for various key metrics
- https://github.com/Patrick912/DataScienceCapstone/blob/master/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- A folium map was created to display all the launch sites including the related launch outcomes
- For better understanding lines and distances to landmarks were added
- https://github.com/Patrick912/DataScienceCapstone/blob/master/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

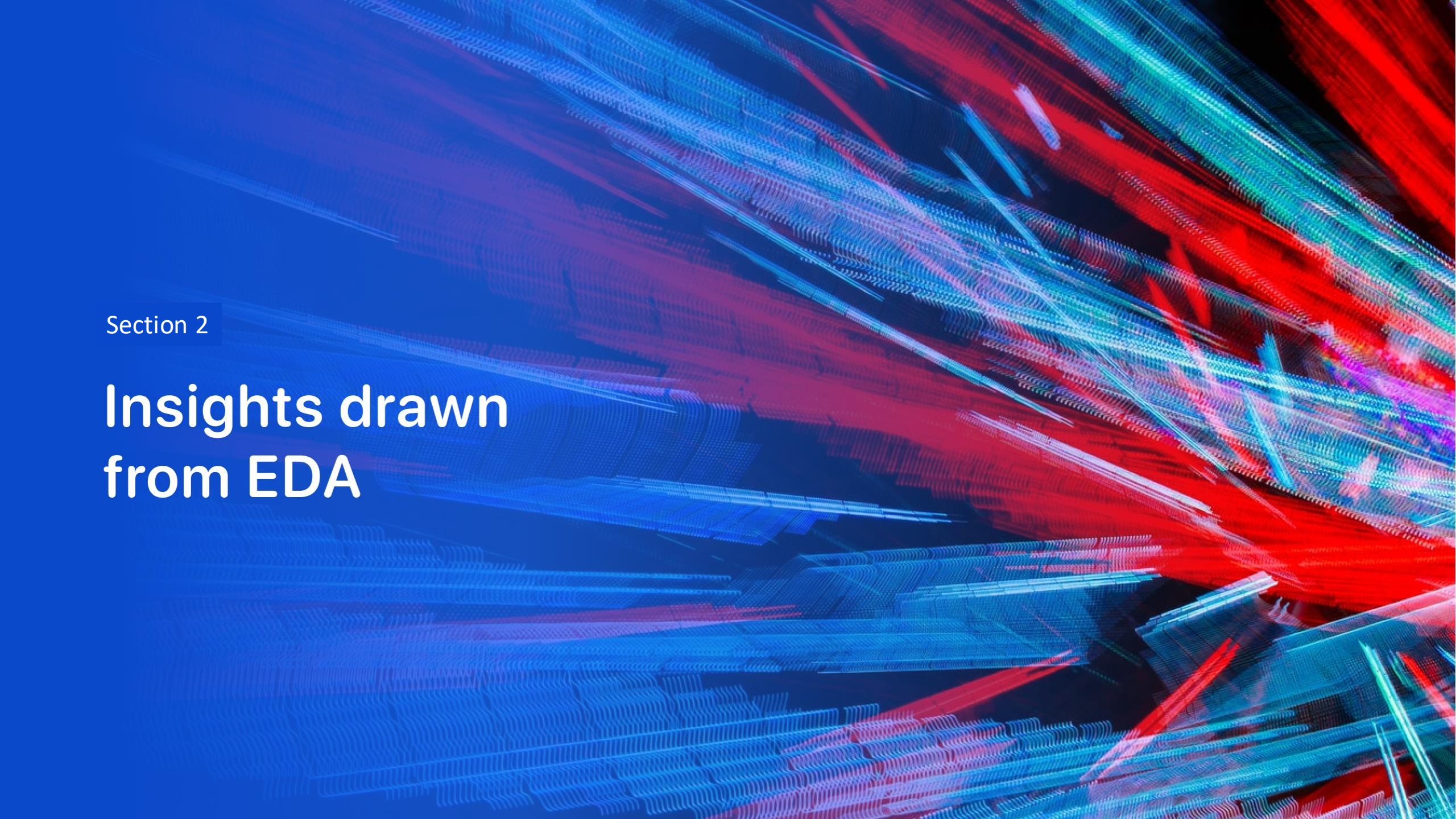
- An interactive dashboard was created to visualize
  - launch outcomes by site
  - Launch outcomes by site related to payload
- https://github.com/Patrick912/DataScienceCapstone/blob/master/project.tar

Predictive Analysis (Classification)

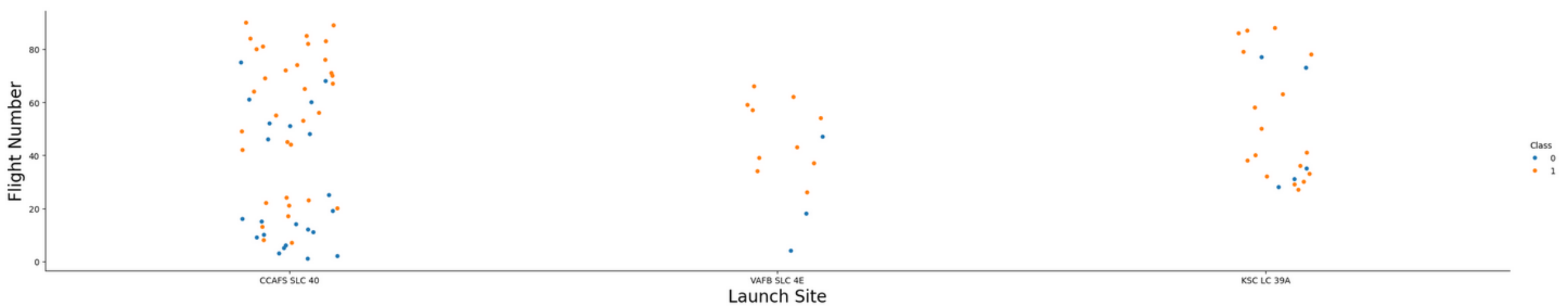
- Summarize how you built, evaluated, improved, and found the best performing classification model
- You need present your model development process using key phrases and flowchart
- https://github.com/Patrick912/DataScienceCapstone/blob/master/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results

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

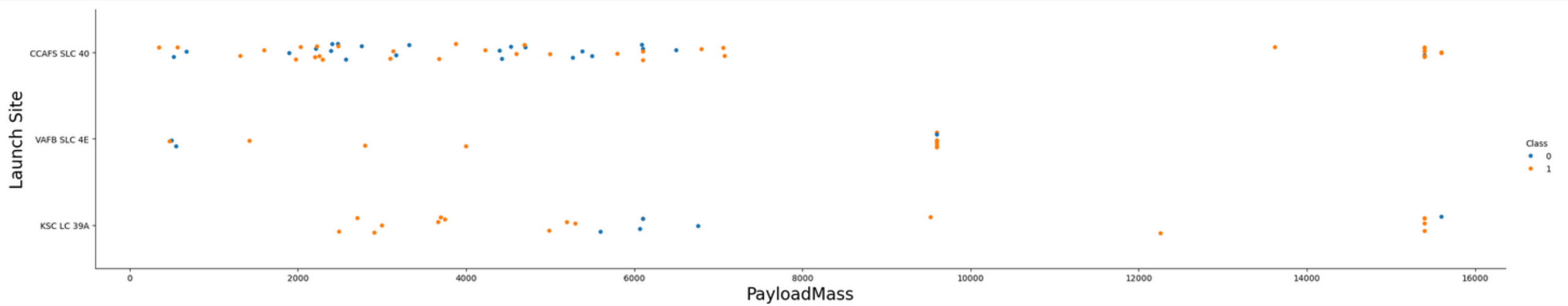

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



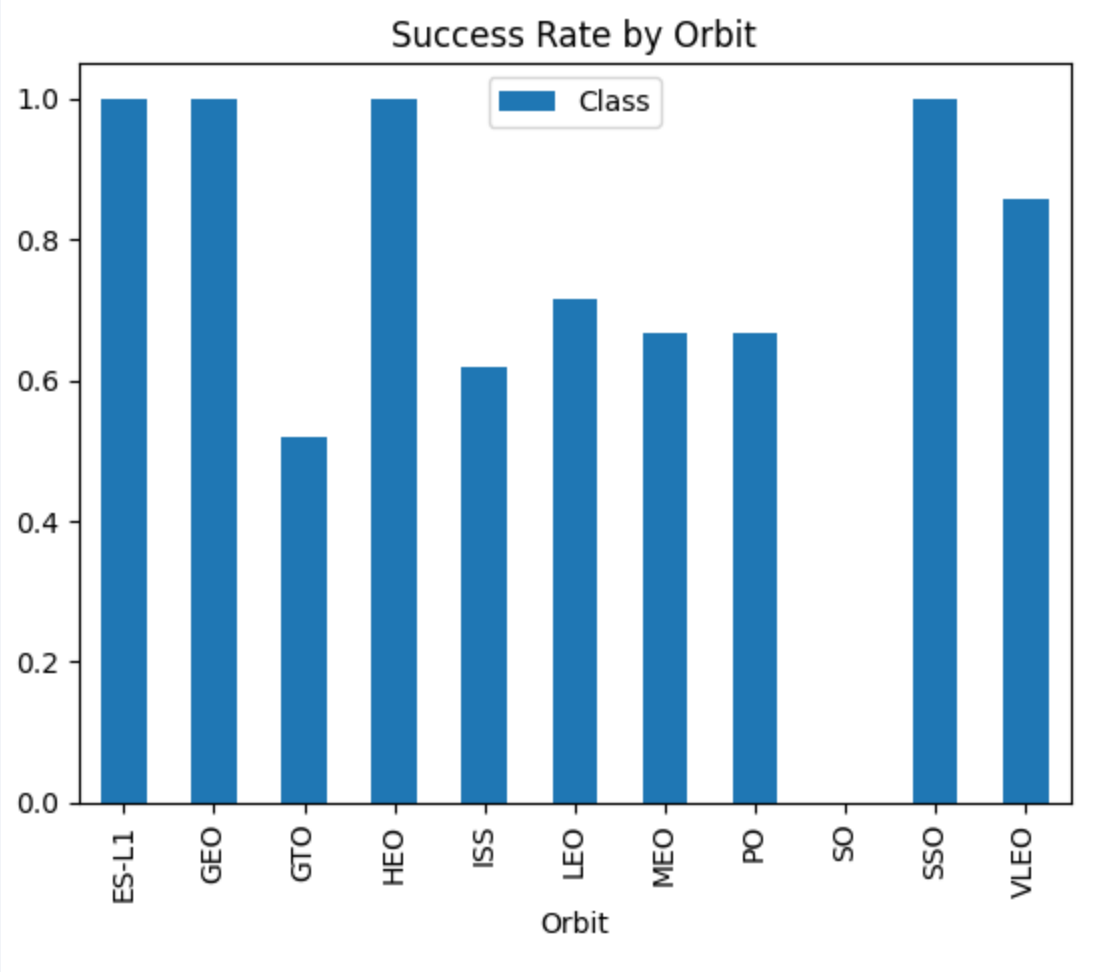
- The higher flight numbers have a higher success rate

Payload vs. Launch Site

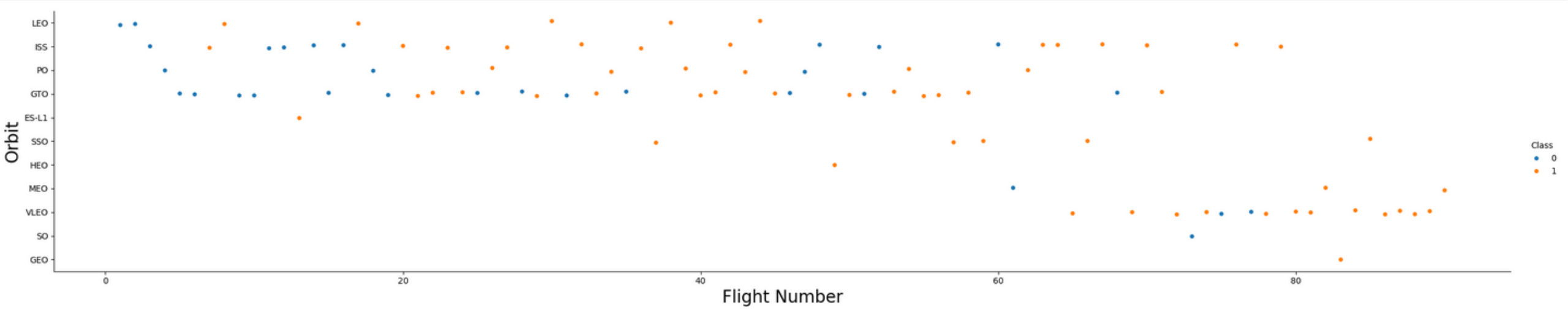


- VAFB has no launches with a payload above 10k
- Launches with higher payloads are more successful

Success Rate vs. Orbit Type

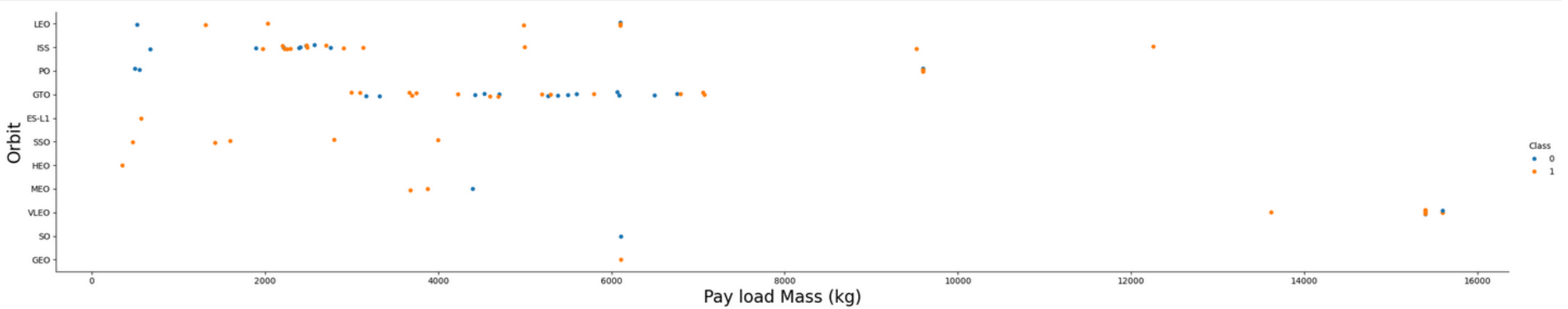


Flight Number vs. Orbit Type



- The LEO success rate appears to be related to the number of flights
- There seems to be no relationship between success rate and flight number forth GTO orbit

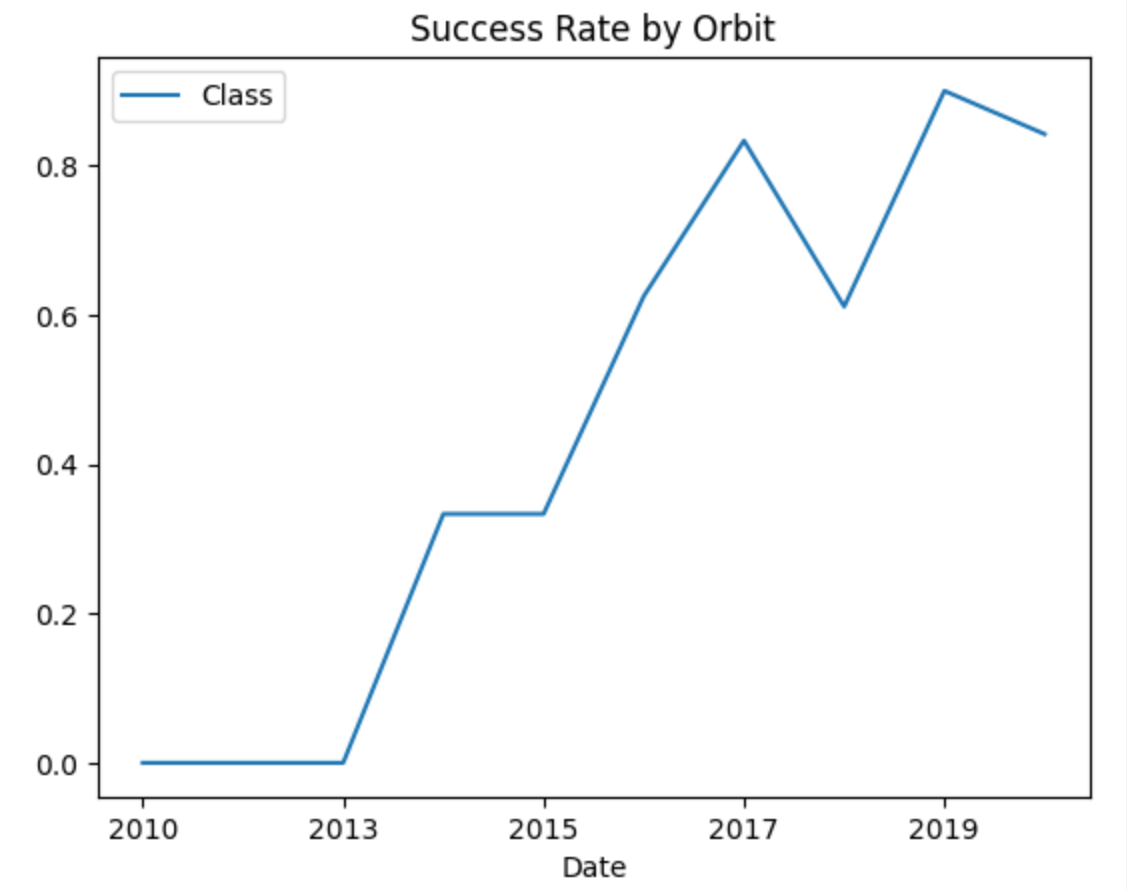
Payload vs. Orbit Type



- Heavier payloads mean a higher success rates for Polar, LEO and ISS.

Launch Success Yearly Trend

- Success rate trends up over time



All Launch Site Names

- There are 4 different launch sites

```
%sql SELECT DISTINCT(Launch_Site) FROM SPACEXTBL
```

```
 * 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'

```
%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

- The 5 first records for launch sites beginning with CCA

Total Payload Mass

- The total payload carried by SpaceX for NASA is 99980kg

Average Payload Mass by F9 v1.1

- Average payload carried by booster version F9 v1.1 is 2534.66kg

First Successful Ground Landing Date

- The first successful ground landing date was December 22nd, 2015.

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql SELECT DISTINCT(Booster_Version) from SPACEXTBL where PAYLOAD_MASS__KG_ >= 4000 and PAYLOAD_MASS__KG_ < 6000 and Landing_Outcome = 'Success (drone ship)'
```

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

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 mission outcomes is 100 compared to 1 failed mission outcome.

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass

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

```
%sql SELECT substr(Date, 6,2) as month, Landing_Outcome, Booster_Version, Launch_site from SPACEXTBL where Landing_Outcome like 'Fail%' and substr(Date,0,5)='2015'
```

```
 * 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

```
%sql SELECT Landing_Outcome, COUNT(*) from SPACEXTABLE where Date between '2010-06-04' AND '2017-03-20' GROUP BY Landing_Outcome ORDER BY COUNT(*) DESC
```

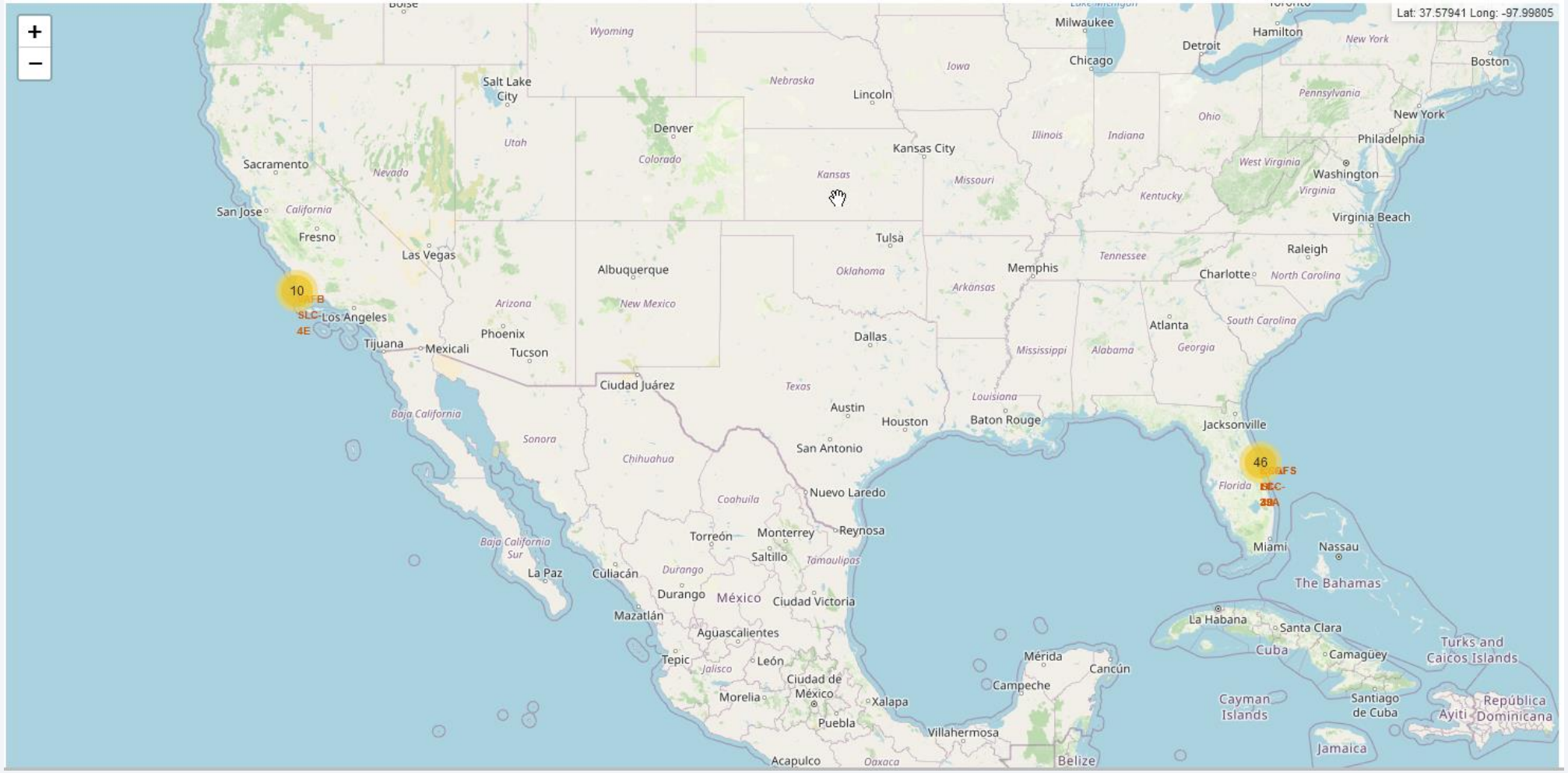
```
 * sqlite:///my_data1.db
Done.
```

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

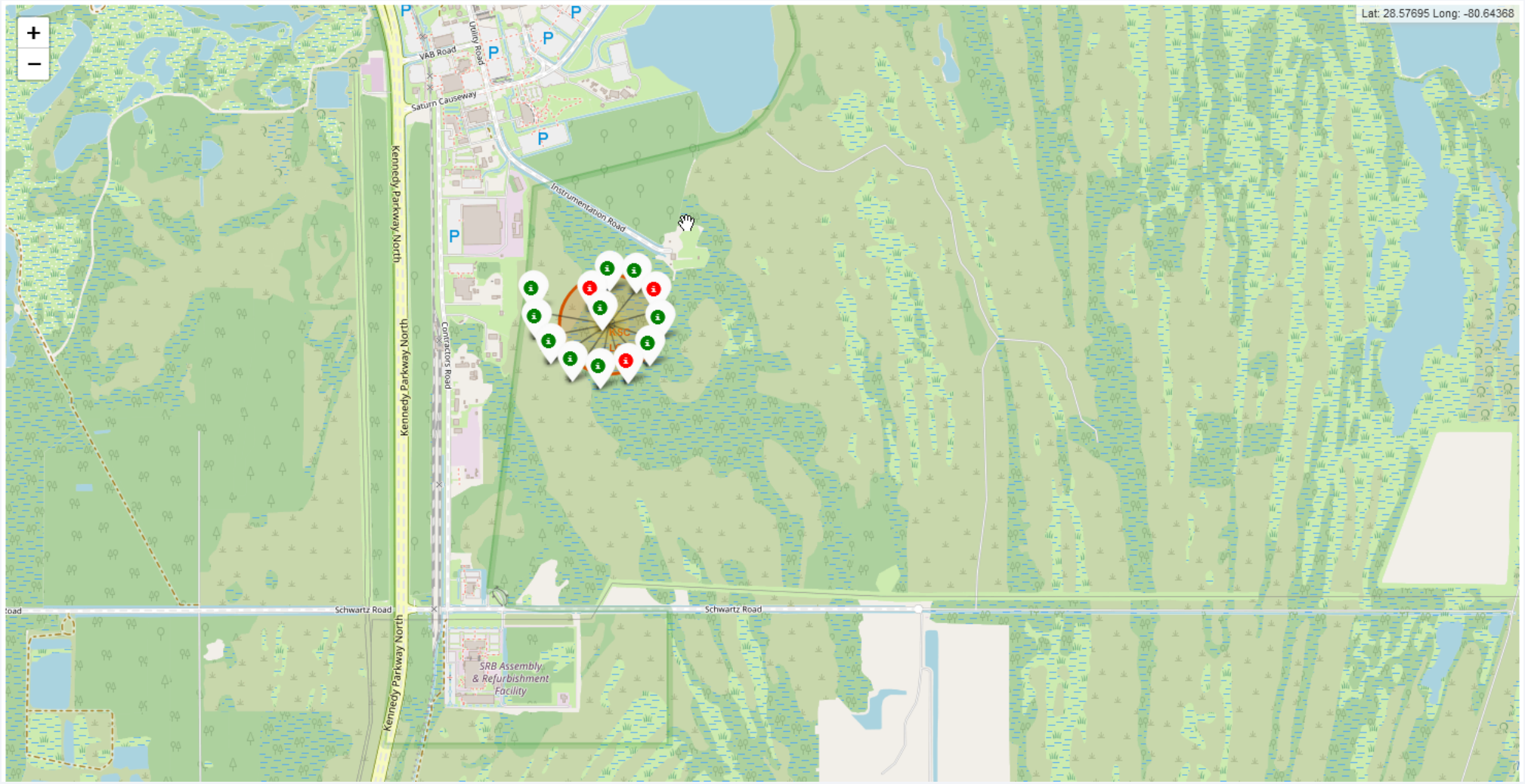
Section 3

Launch Sites Proximities Analysis

Map – Launch Sites

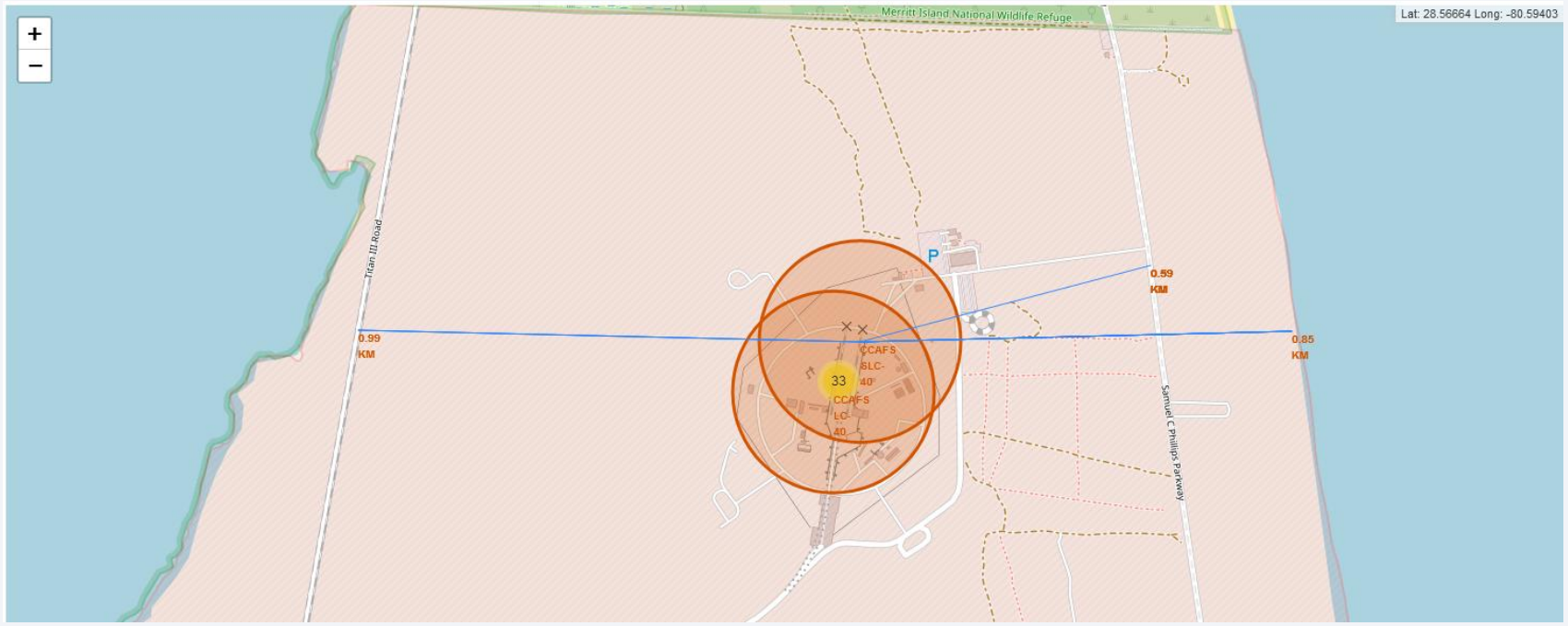


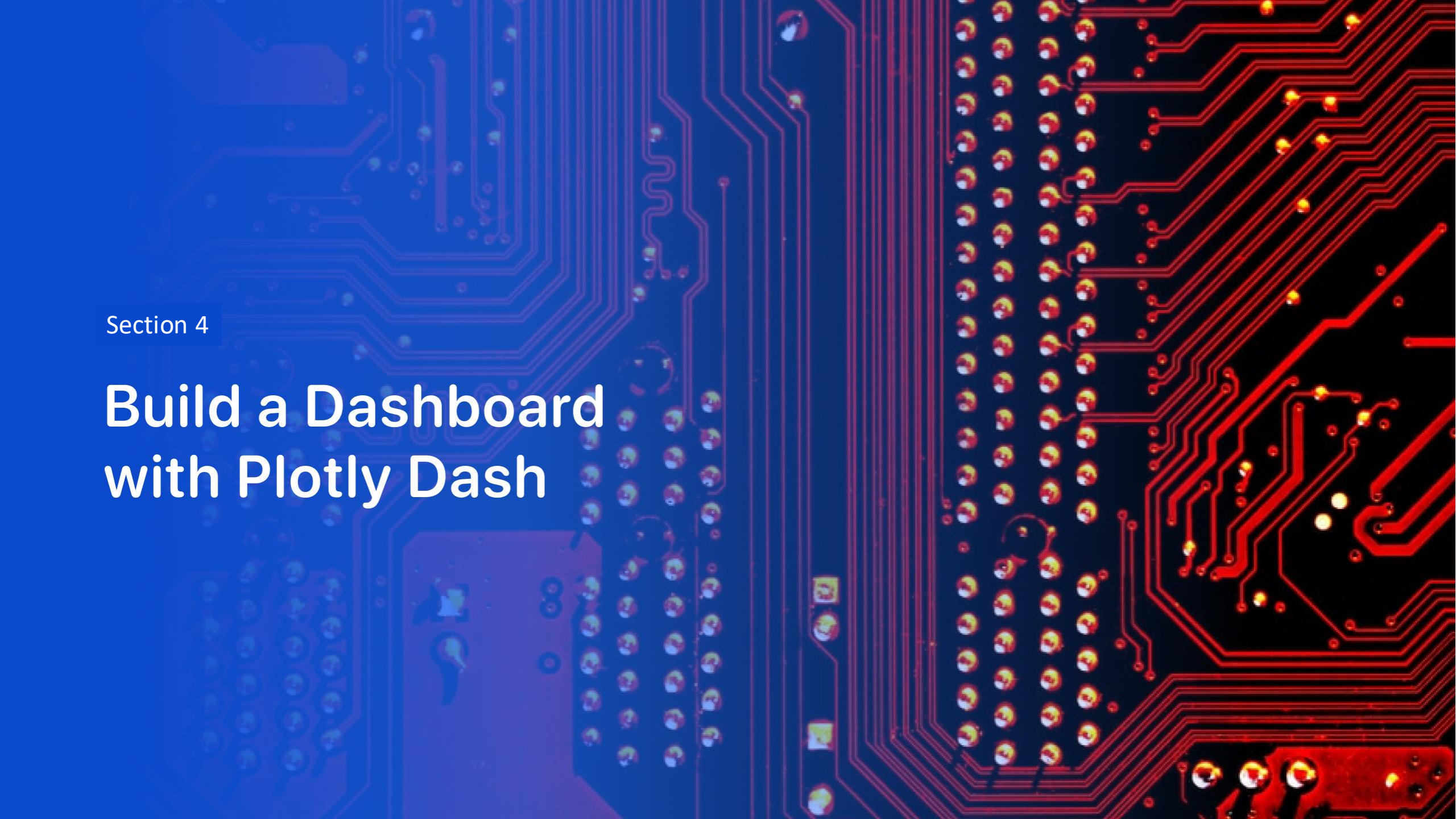
Map – Launch Outcomes



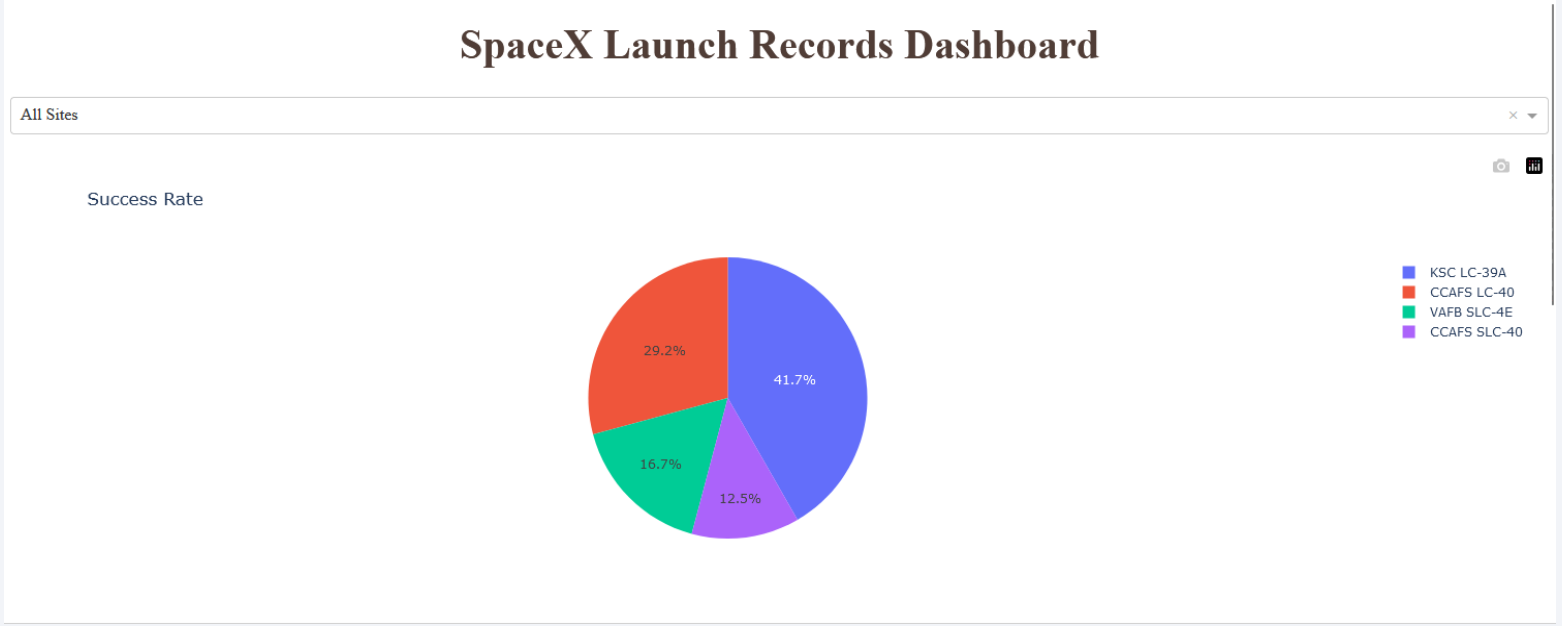
Map – Distances of landmarks to launch site

- Distances from launch site to
  - Coast line
  - Highway
  - railroad



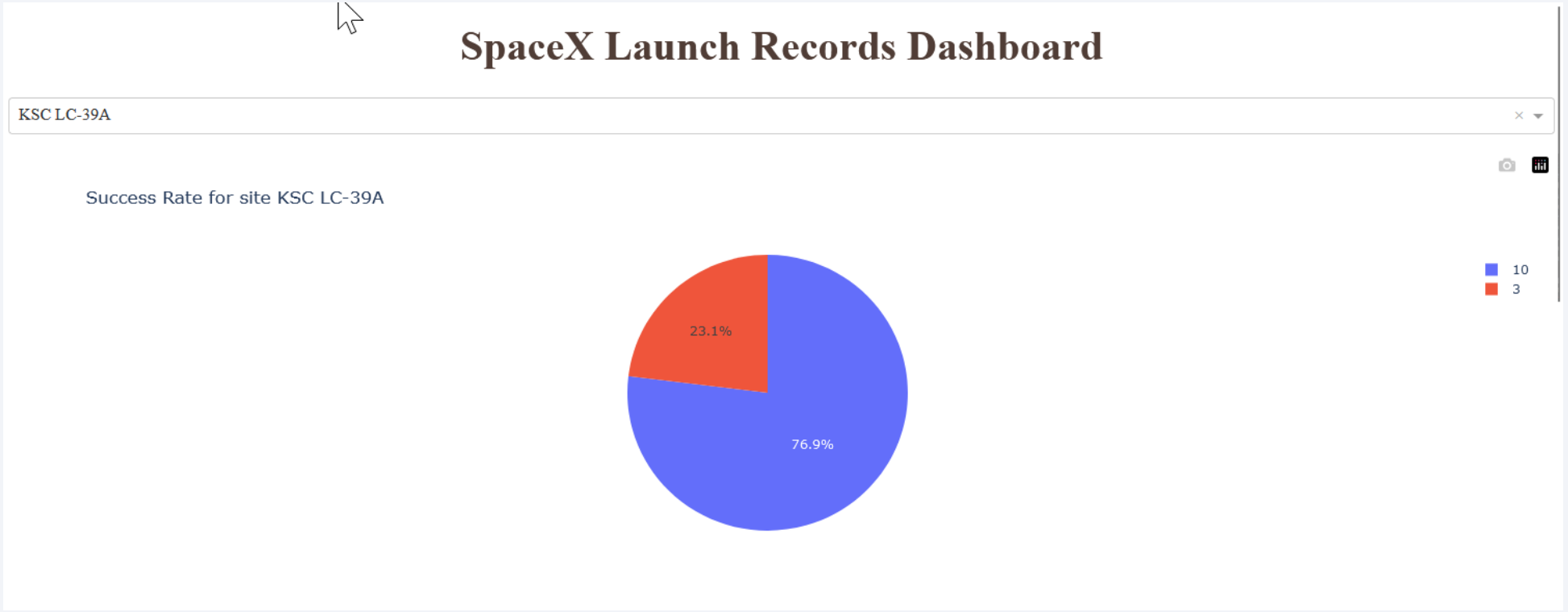

Section 4
Build a Dashboard with Plotly Dash

Launch Success Rate – All Sites



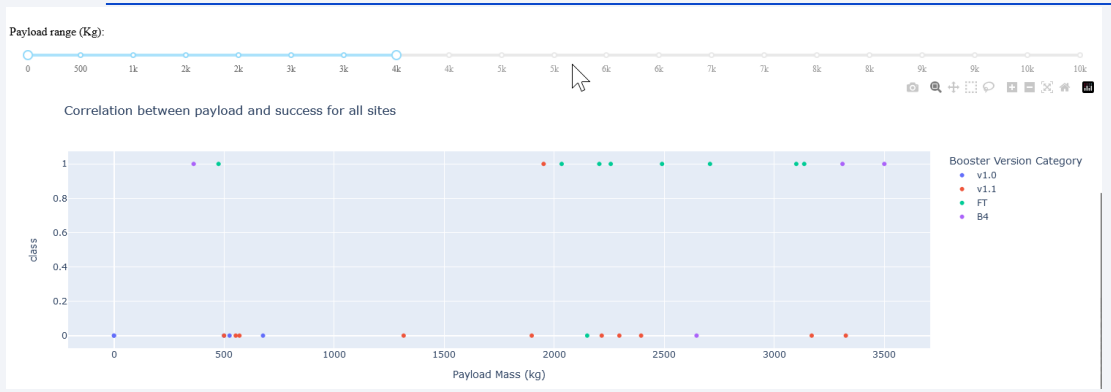
- The site with the highest success rate for launches is KSC LC-39A

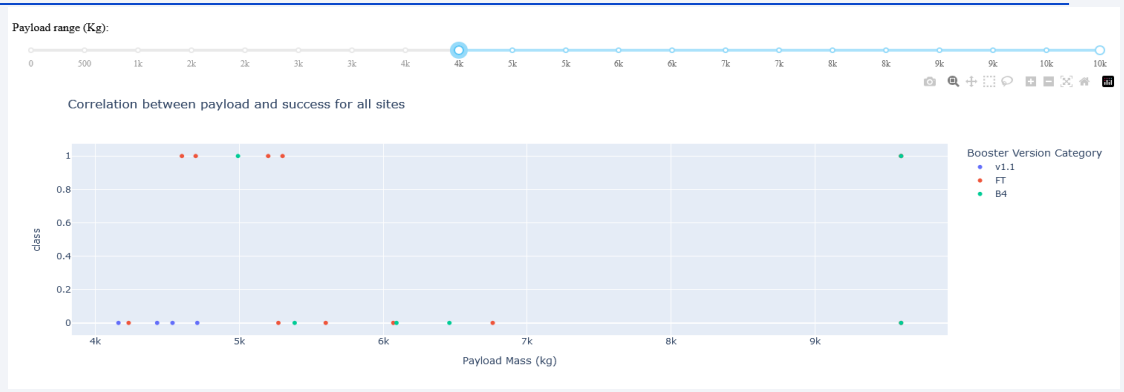
Launch Success Rate – KSC LC39A



- Site KSC LC39A sports the highest launch success rate of 76.9% (10 out of 13 launches)

Launch Success Rate – Correlation between payload and booster





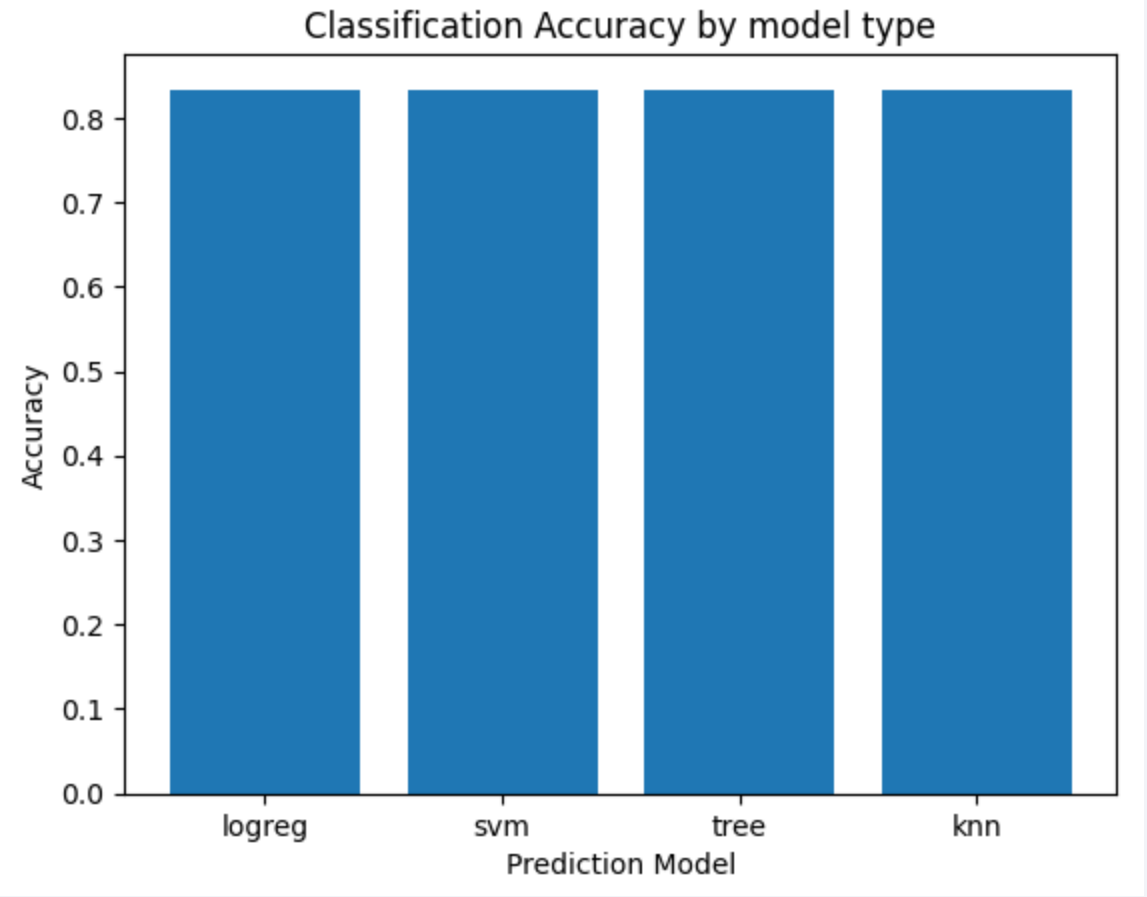
- The left chart shows the correlation between payload and booster version for launches with a payload under 4000kg
- The right chart shows the correlation between payload nad booster version for launchses with a payload above 4000kg

Section 5

Predictive Analysis (Classification)

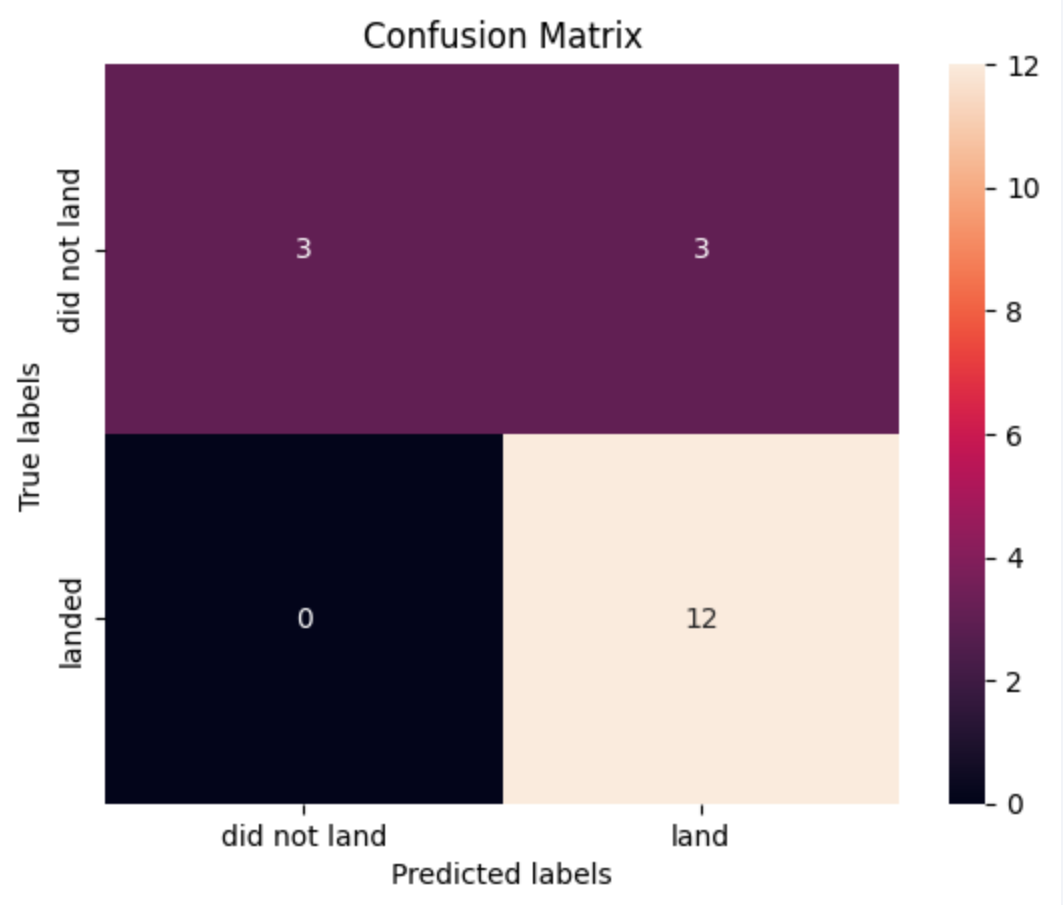
Classification Accuracy

- All model types have the same classification accuracy



Confusion Matrix

- Show the confusion matrix of the best performing model with an explanation



Conclusions

- Based on the available data we were able to build a model that predicts successful landing accurately, but falls short at classifying unsuccessful landings.

Appendix

- https://github.com/Patrick912/DataScienceCapstone/tree/master

Thank you!