IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

James Grajek
September 16, 2023

Outline

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

Executive Summary

- Summary of methodologies

  Data was collected either through web scraping, csv files or using wget to determine if the SpaceX Falcon9 rocket lands successfully.

  Visualizations were used to look at the data using Folium.

- Summary of all results

  EDA Results

  Interactive Analytics

  Predictive Analysis

Introduction

- Project background and context

  SpaceX advertises Falcon 9 rocket with a cost of $62M and can sell at this price because they reuse the first stage of the rocket. If we can determine if the first stage will land successful we can determine the cost of a launch.

- Problems you want to find answers

  Will the first stage of the rocket land successfully?

  What factors can influence a successful landing?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected either through SpaceX Rest API, webscraping, csv files or using wget Perform data wrangling
  - Data was processed using pandas, numpy, machine learning, SQL and Visual techniques.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash

  Markers and measurements were used to visually see our data on a map. We also used Plotly to see our data in pie charts and a scatter plot.
- Perform predictive analysis using classification models

  Classification machine learning models were used to train and test our data.

Data Collection

- Describe how data sets were collected.

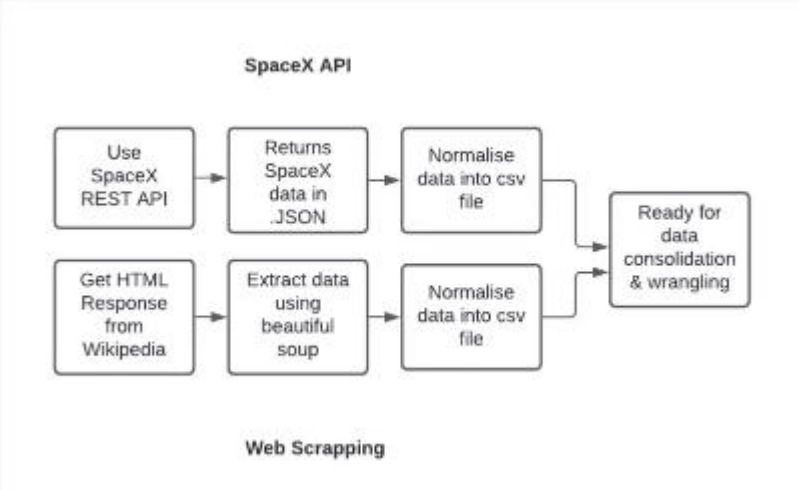
  SpaceX API

  Using BeautifulSoup, webscraping was performed on a wiki page

  Using wget data was analyzed from a site

  I also downloaded a csv file on my local machine to analyze the data in Excel

- You need to present your data collection process use key phrases and flowcharts



Data Collection – SpaceX API

- Data collection with SpaceX REST calls

- https://github.com/jrgrajek/Capstone.git

```
spacex_url="https://api.spacexdata.com/v4/launches/past"

response = requests.get(spacex_url)
```

Decode the response content as a Json then turn it into a Pandas dataframe

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

Clean our data

Store data from requests in lists to then create a new df

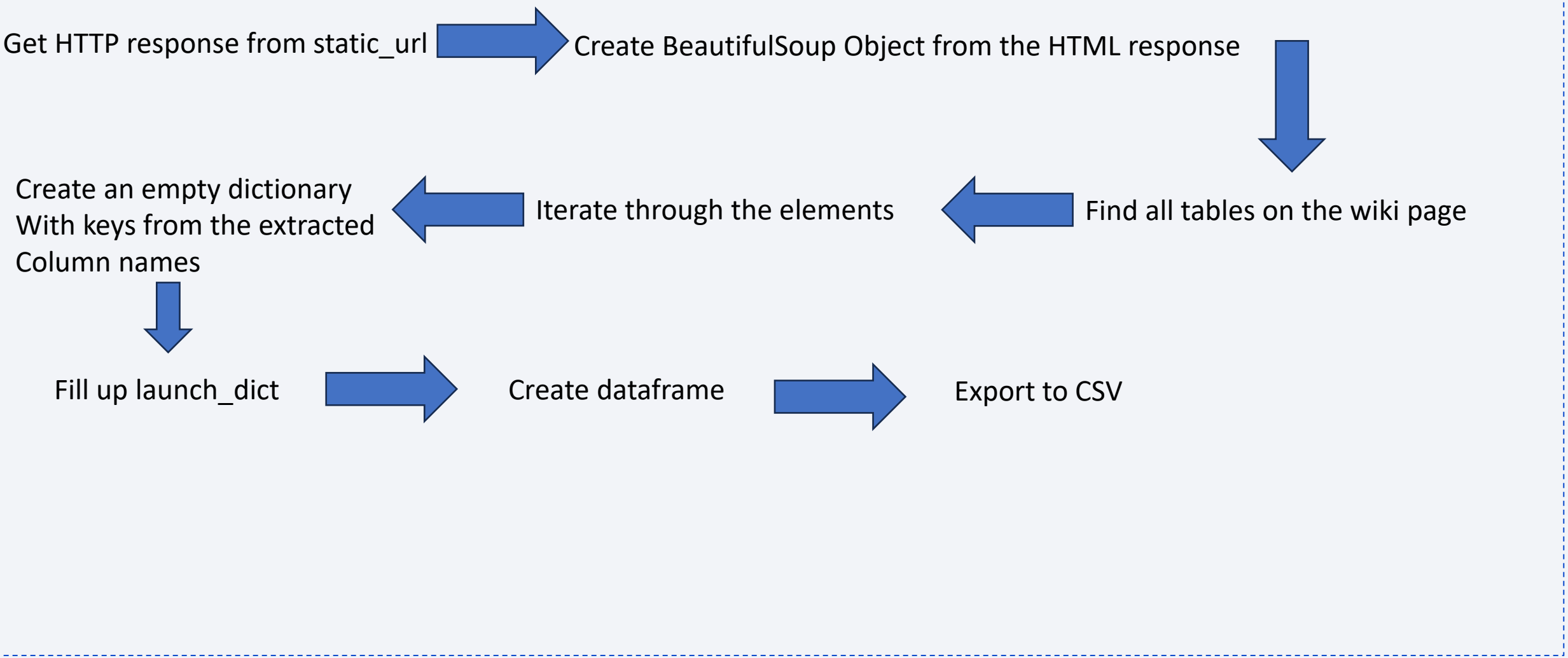
Combine columns into a dictionary from:

getBoosterVersion(data)
getLaunchSite(data)
getPayloadData(data)
getCoreData(data)

Assign List to Dictionary then df → Filter df for only Falcon 9

Data Collection - Scraping

- https://github.com/jrgrajek/Webscraping.git



Data Wrangling

https://github.com/jrgrajek/Data-Wrangling.git

Check for missing values → **Calculate the number of launches on each site**

↓

Calculate the number and occurrence of each orbit

↓

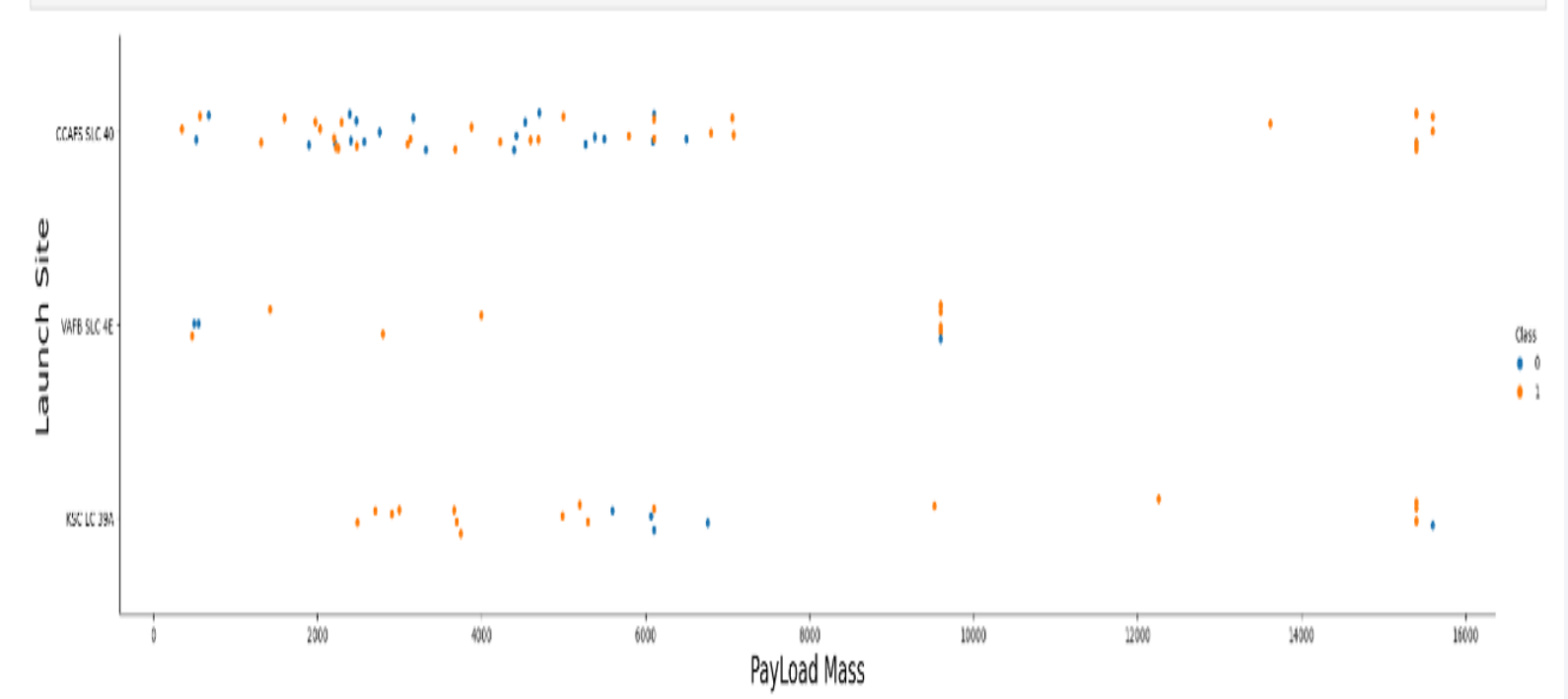
Calculate the number and occurence of mission outcome per orbit type

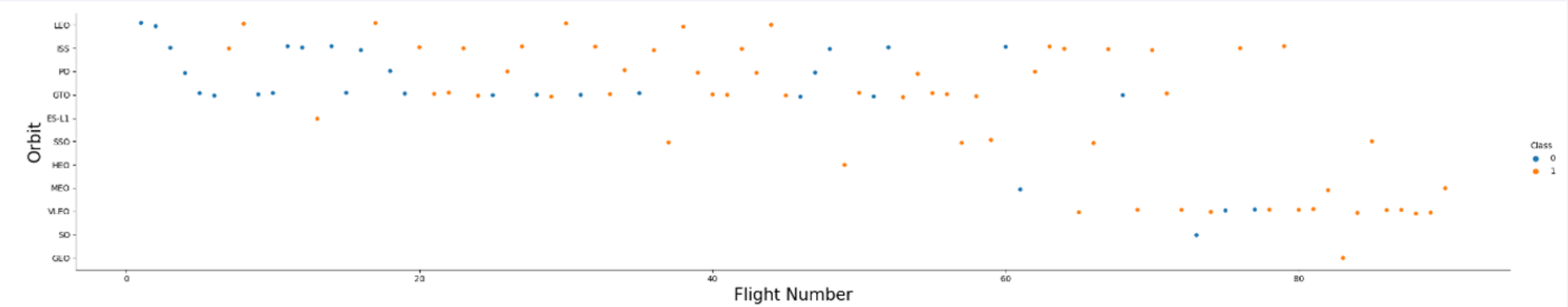
↓

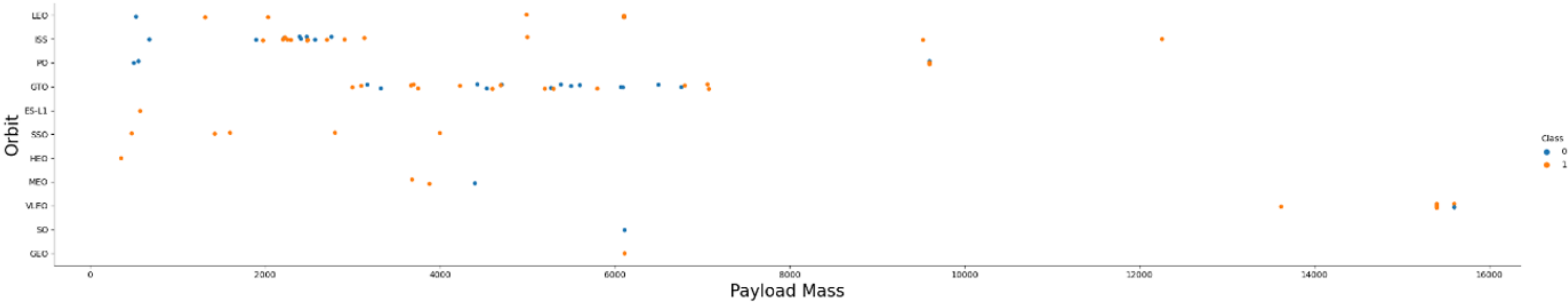
Create a landing outcome label from Outcome column

https://github.com/jrgrajek/Data-Visualization.git

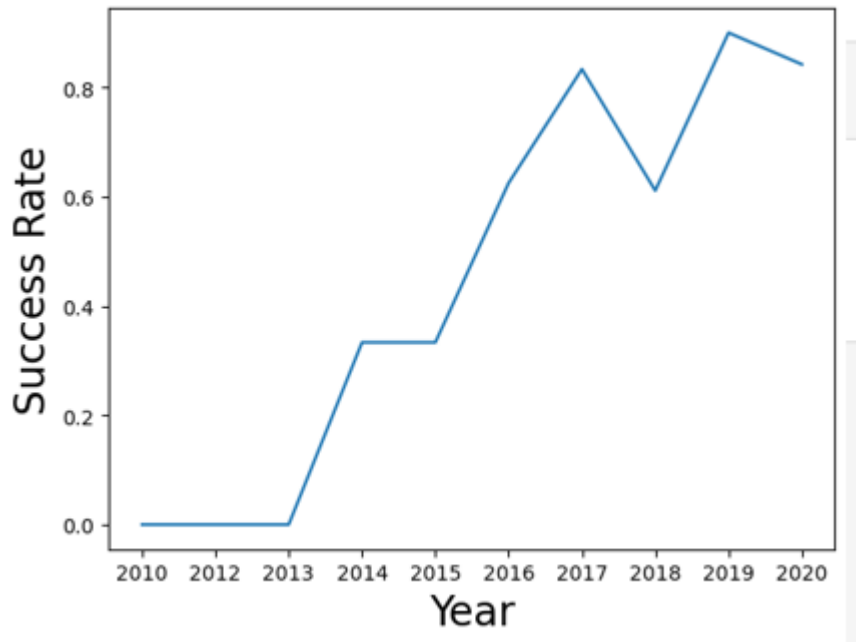
EDA with Data Visualization







Success rate steadily climbing since 2013!!



EDA with SQL

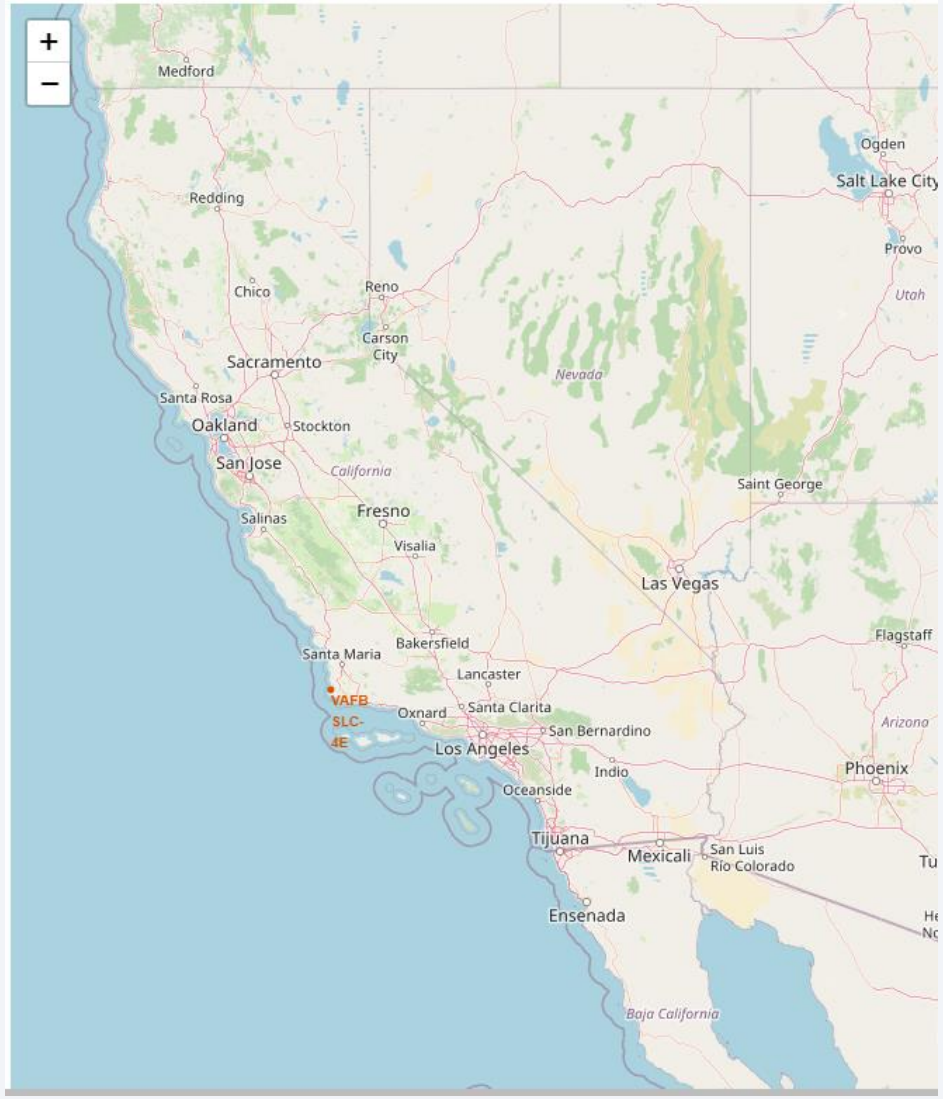
https://github.com/jrgrajek/SQL.git

SQL queries performed

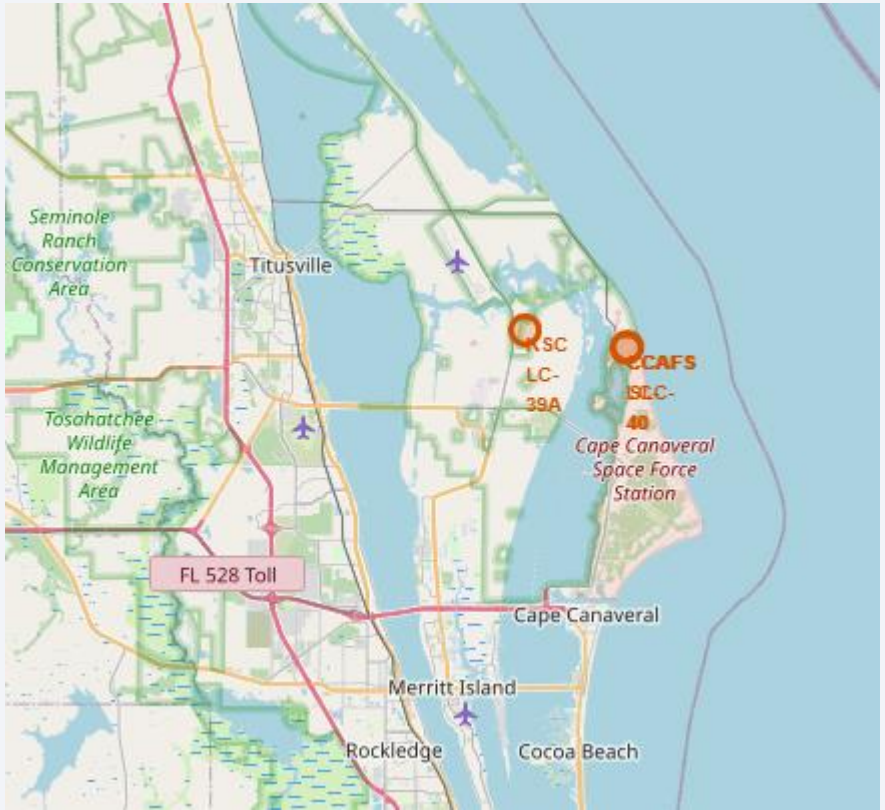
- **Display the names of the unique launch sites in the space mission**
- **Display 5 records where launch sites begin with the string 'CCA'**
- **Display the total payload mass carried by boosters launched by NASA (CRS)**
- **Display average payload mass carried by booster version F9 v1.1**
- **List the date when the first succesful landing outcome in ground pad was achieved**
- **List 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. Use a subquery**
- **List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015**
- **Rank 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**

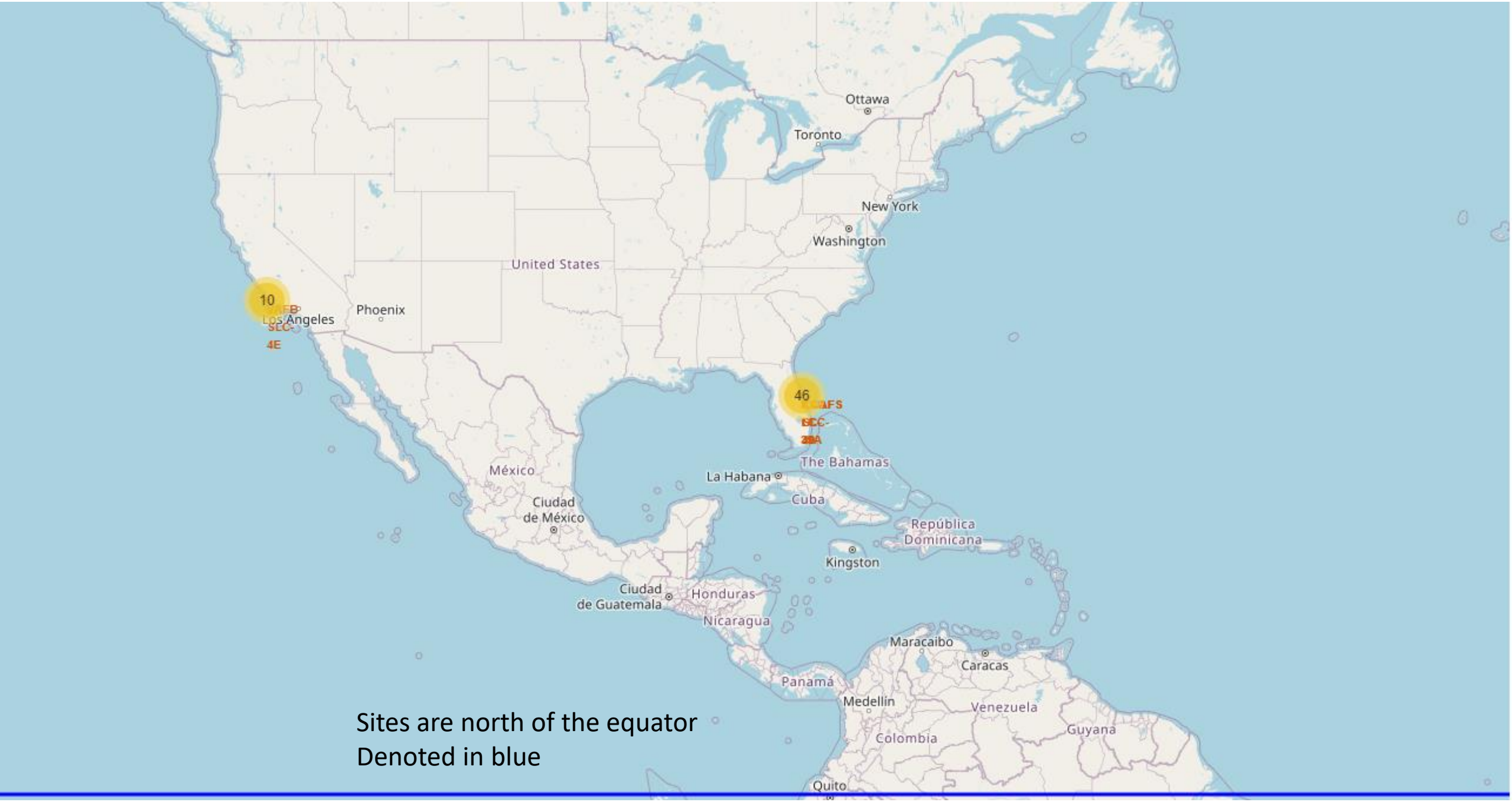
Build an Interactive Map with Folium

https://github.com/jrgrajek/Folium.git

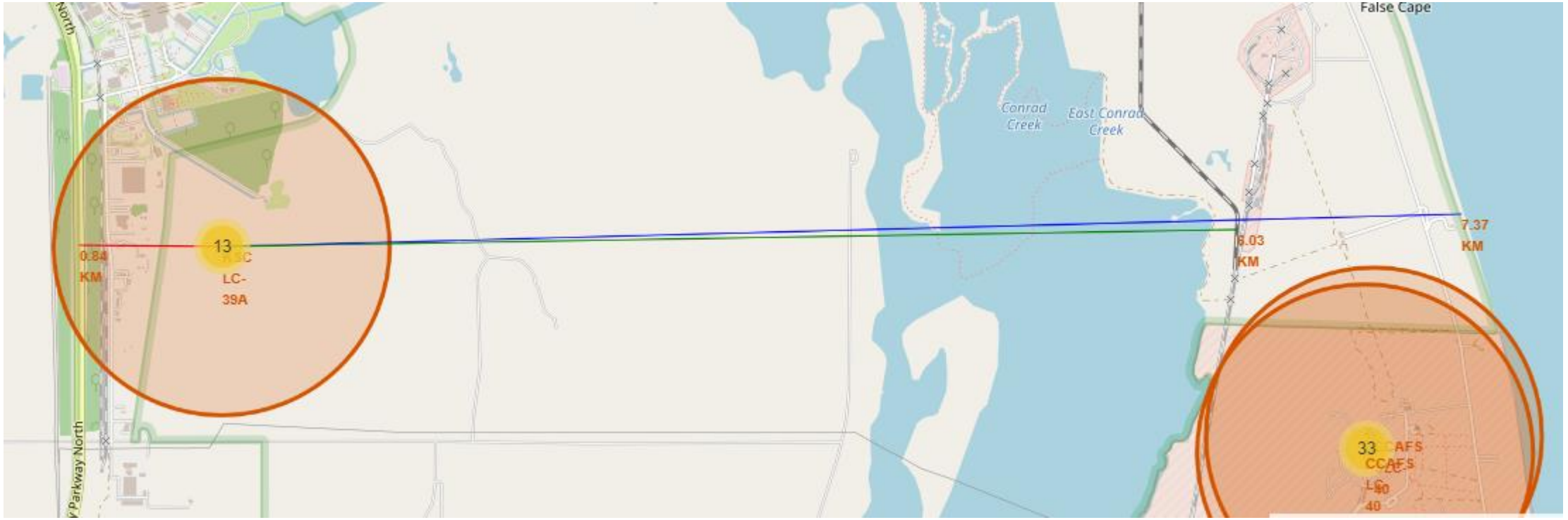


Markers were added to find a great site for building a launch site




Ottawa
Toronto
New York
Washington
United States
10
Phoenix
Angeles
4E
46
The Bahamas
México
La Habana
Cuba
Ciudad
de México
República
Dominicana
Kingston
Ciudad
de Guatemala
Honduras
Nicaragua
Maracaibo
Caracas
Panamá
Medellín
Venezuela
Guyana
Colombia
Quito
Sites are north of the equator
Denoted in blue

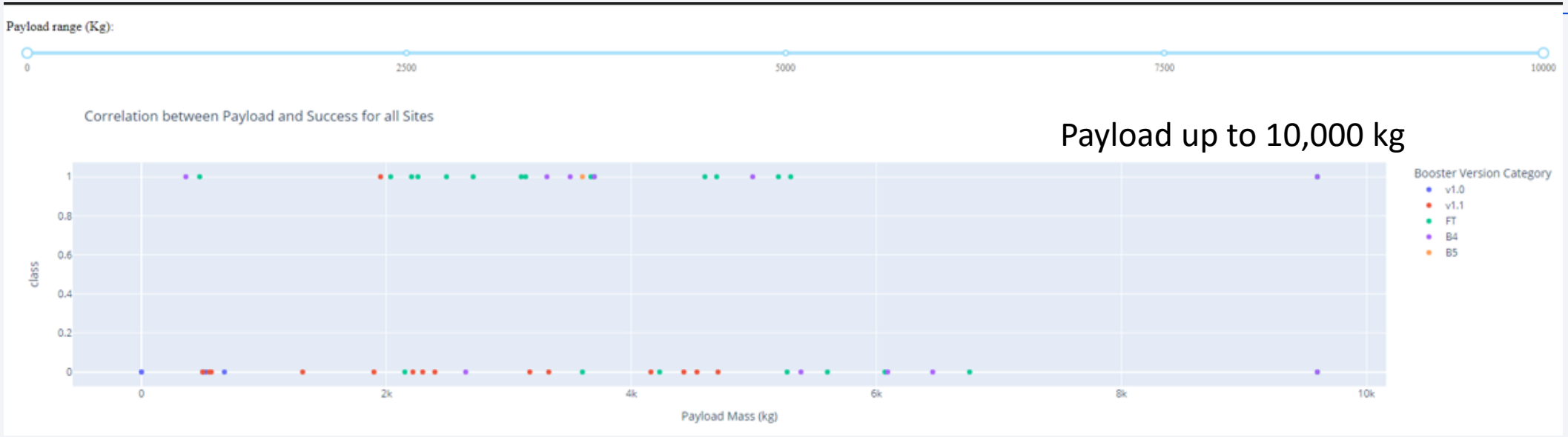
Markers were added denoting Success or Failure of launch by Clicking on the marker

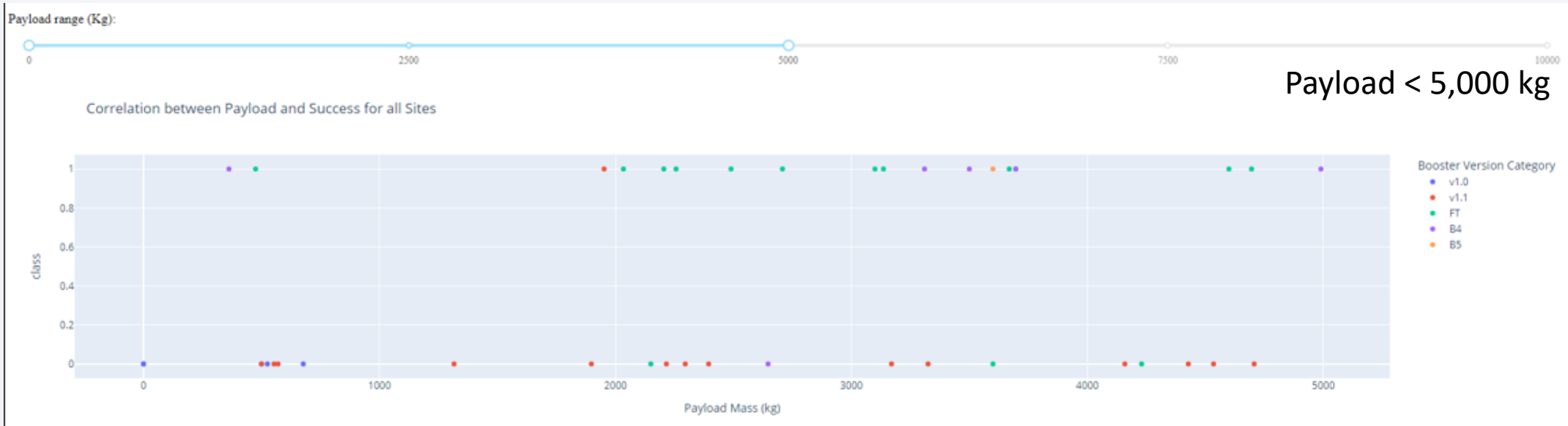


Distance to the coast, railway and nearest highway

Build a Dashboard with Plotly Dash

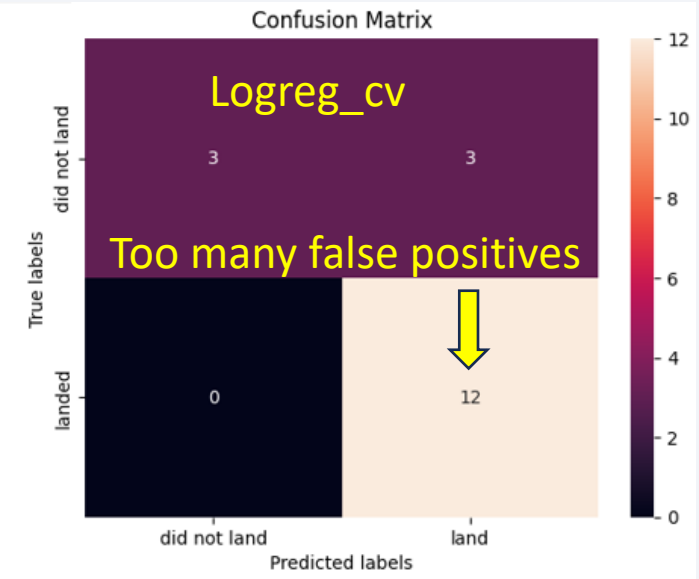
https://github.com/jrgrajek/Plotly.git

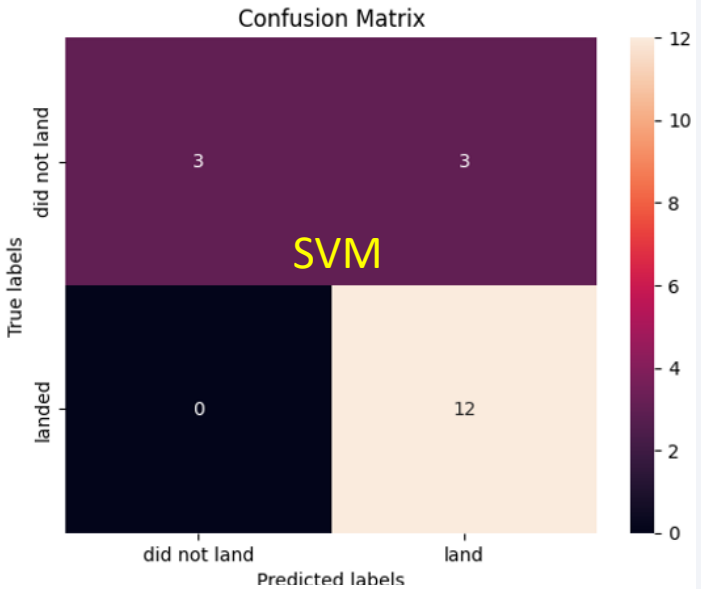


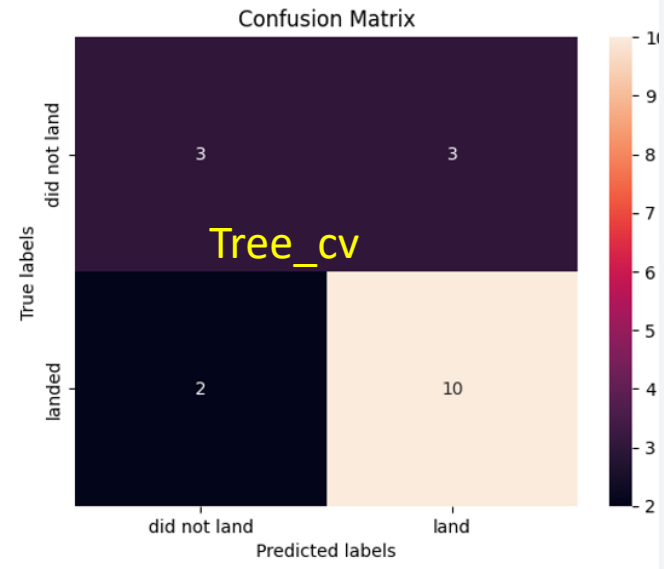


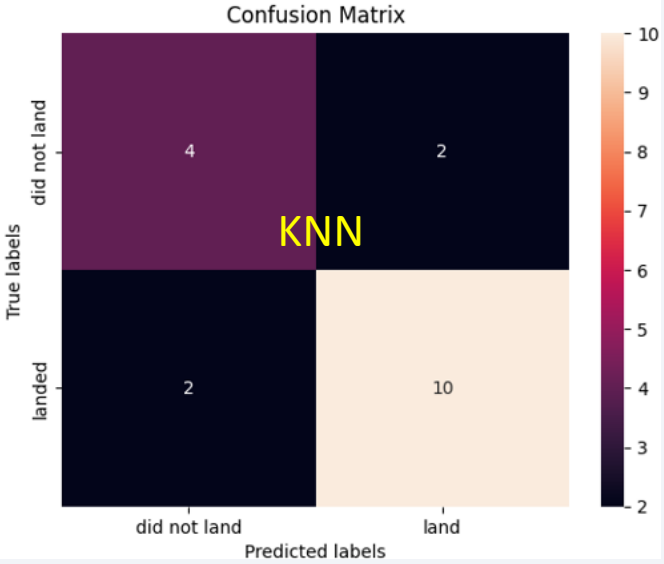
https://github.com/jrgrajek/Machine-Learning-Capstone.git

Predictive Analysis (Classification)









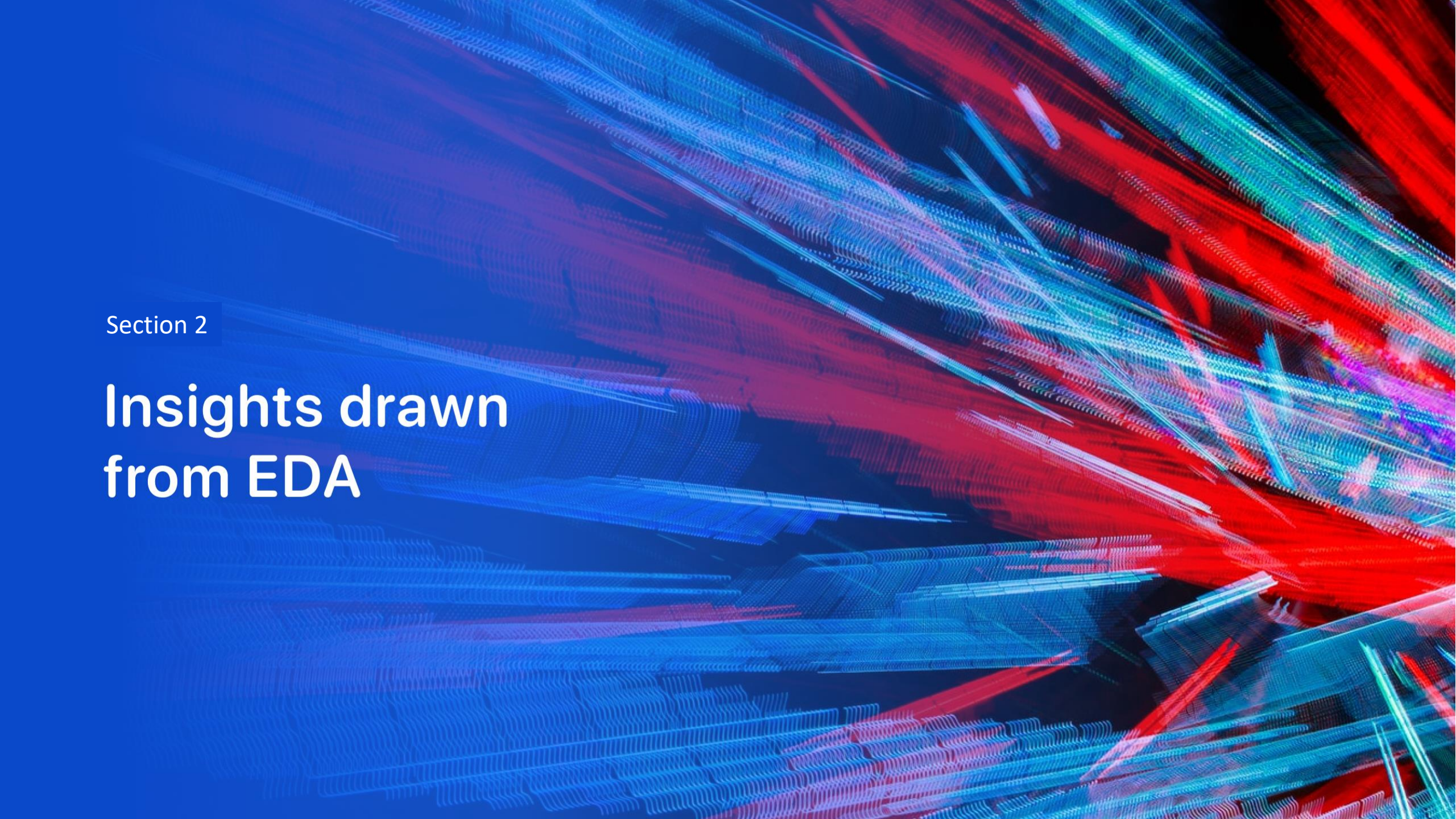
Results

All machine learning models gave similar results in the 80%+ range

Low weighted payloads perform much better than heavier payloads

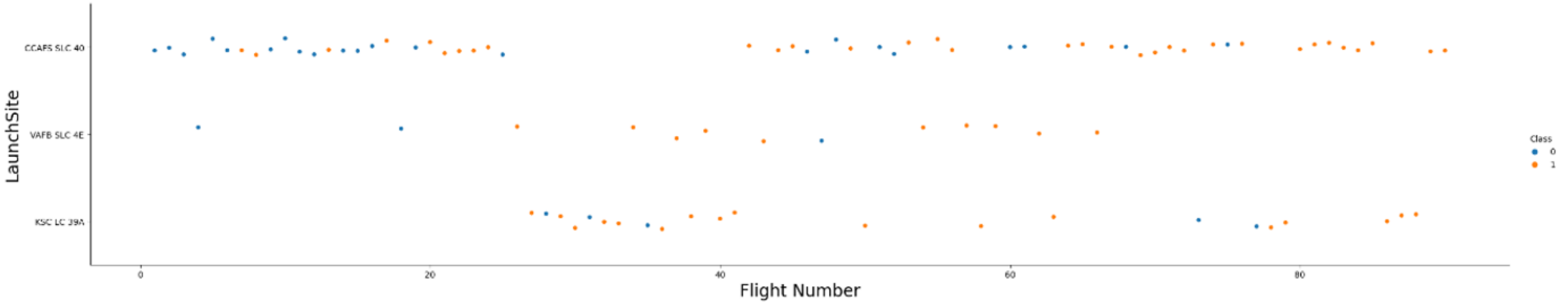
KSC LC 39A had the most successful launches of the sites

SpaceX has steadily increased their success rate since 2013


Section 2
Insights drawn from EDA

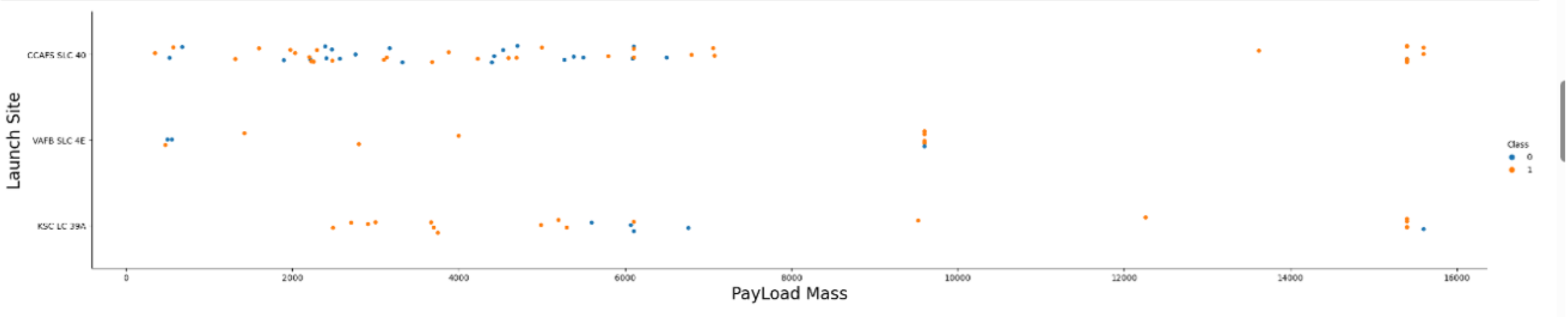
Flight Number vs. Launch Site

- Scatter plot of Flight Number vs. Launch Site



CCAFS SLC 40 have many more launches

Payload vs. Launch Site

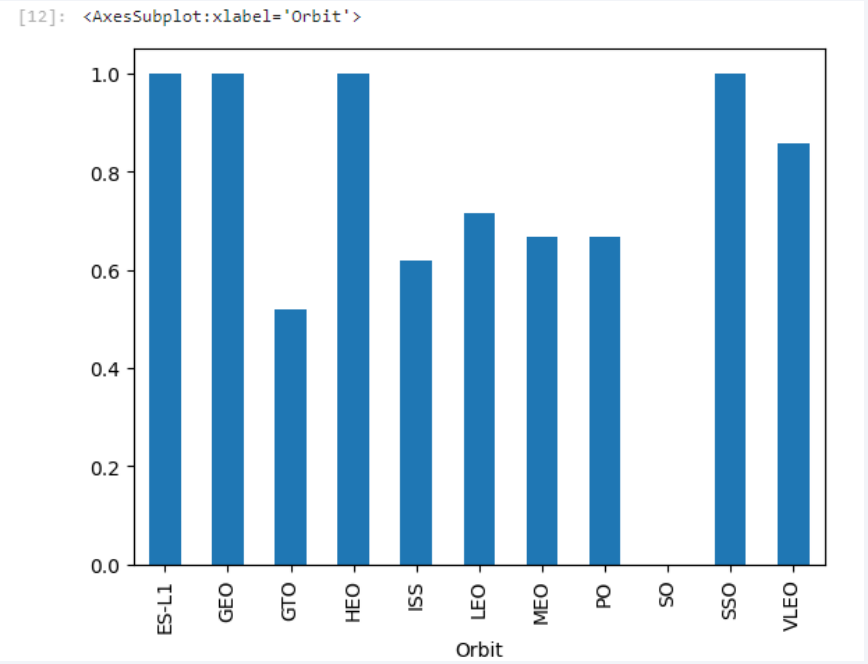


Most launches have Payload Mass < 7500 kg

VAFB-SLC has no payloads > 10,000 kg

Success Rate vs. Orbit Type

[12]: <AxesSubplot:xlabel='Orbit'>



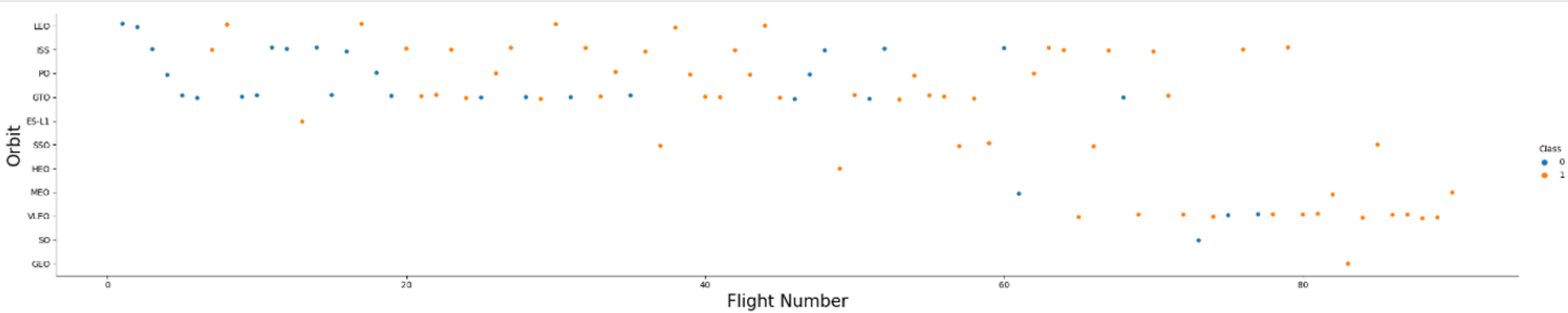
ES-L1, GEO, HEO and SSO have a success rate of 100%

SO has no success

Flight Number vs. Orbit Type

LEO orbit success appears to be related to the # of flights

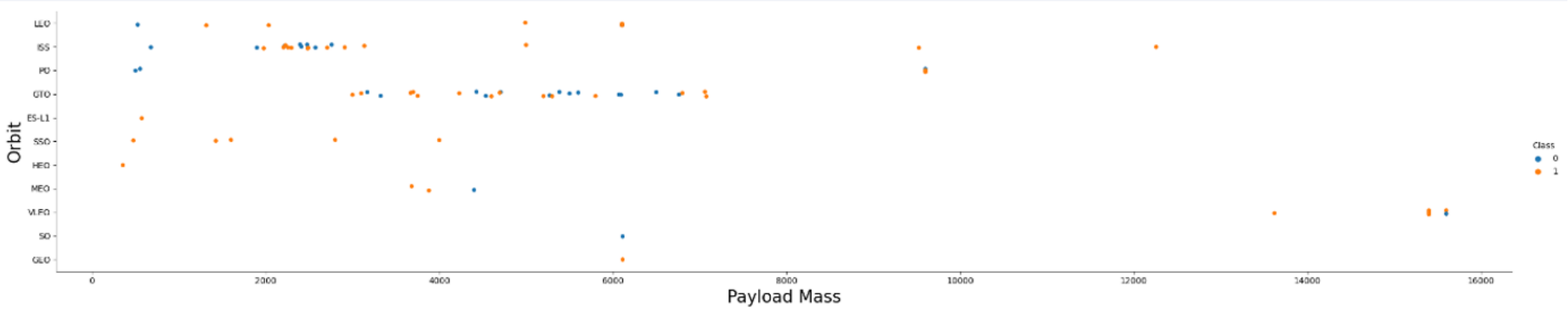
There appears to be no relationship between flight # when in GTO orbit



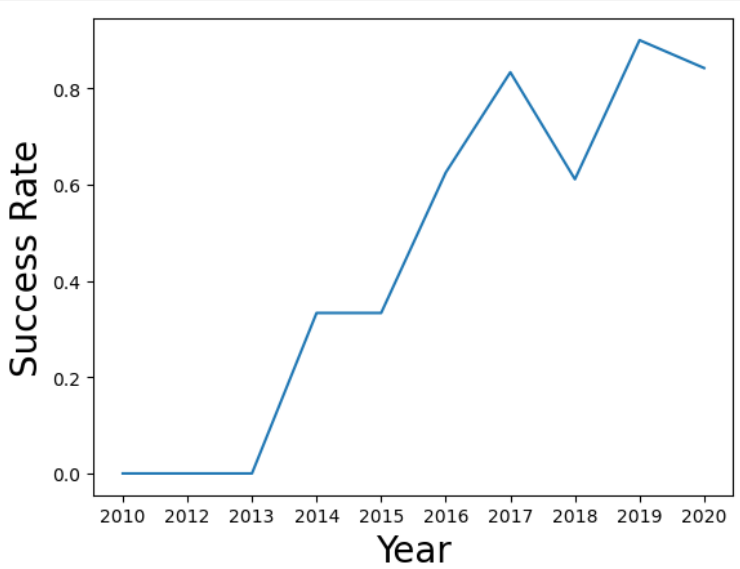
Payload vs. Orbit Type

Polar, LEO and ISS have more success with heavy payloads

GTO can’t be distinguished for heavy payloads



Launch Success Yearly Trend



Success has steadily climbed from 2013 but has since leveled off in recent data

All Launch Site Names

```
%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

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

```
%sql select SUM(PAYLOAD_MASS__KG_) from SPACEXTBL where Customer = 'NASA (CRS)';
```

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

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

Display average payload mass carried by booster version F9 v1.1

```
%sql select AVG(PAYLOAD_MASS__KG_) from SPACEXTBL where Booster_Version = 'F9 v1.1';
```

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

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

```
%sql select MIN(Date) from SPACEXTBL where Landing_Outcome = 'Success (ground pad)'
```

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

MIN(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql SELECT DISTINCT BOOSTER_VERSION FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ BETWEEN 4000 AND 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

```
%sql SELECT MISSION_OUTCOME, COUNT(*) AS TOTAL FROM SPACEXTBL GROUP BY MISSION_OUTCOME ORDER BY MISSION_OUTCOME;
```

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

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

Boosters Carried Maximum Payload

```
%sql select Booster_Version FROM SPACEXTBL where payload_mass__kg_=(select MAX(payload_mass__kg_) from Spacextbl)
```

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

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,4,2) as 'Month Number', substr(Date,7,4) as 'Year', 'Booster_Version', 'Launch_Site' FROM SPACEXTBL
WHERE substr(Date,7,4) = '2015' and 'Landing_Outcome' = 'Failure (drone ship)';
```

Month Number	Year	'Booster_Version'	'Launch_Site'

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

```
%%sql SELECT "LANDING_OUTCOME", COUNT(*) as 'COUNT' FROM SPACEXTBL

WHERE substr(Date,1,4) || substr(Date,6,2) || substr(Date,9,2)

between '20100604' and '20170320' GROUP BY "Landing_Outcome" ORDER BY "COUNT" DESC;
```

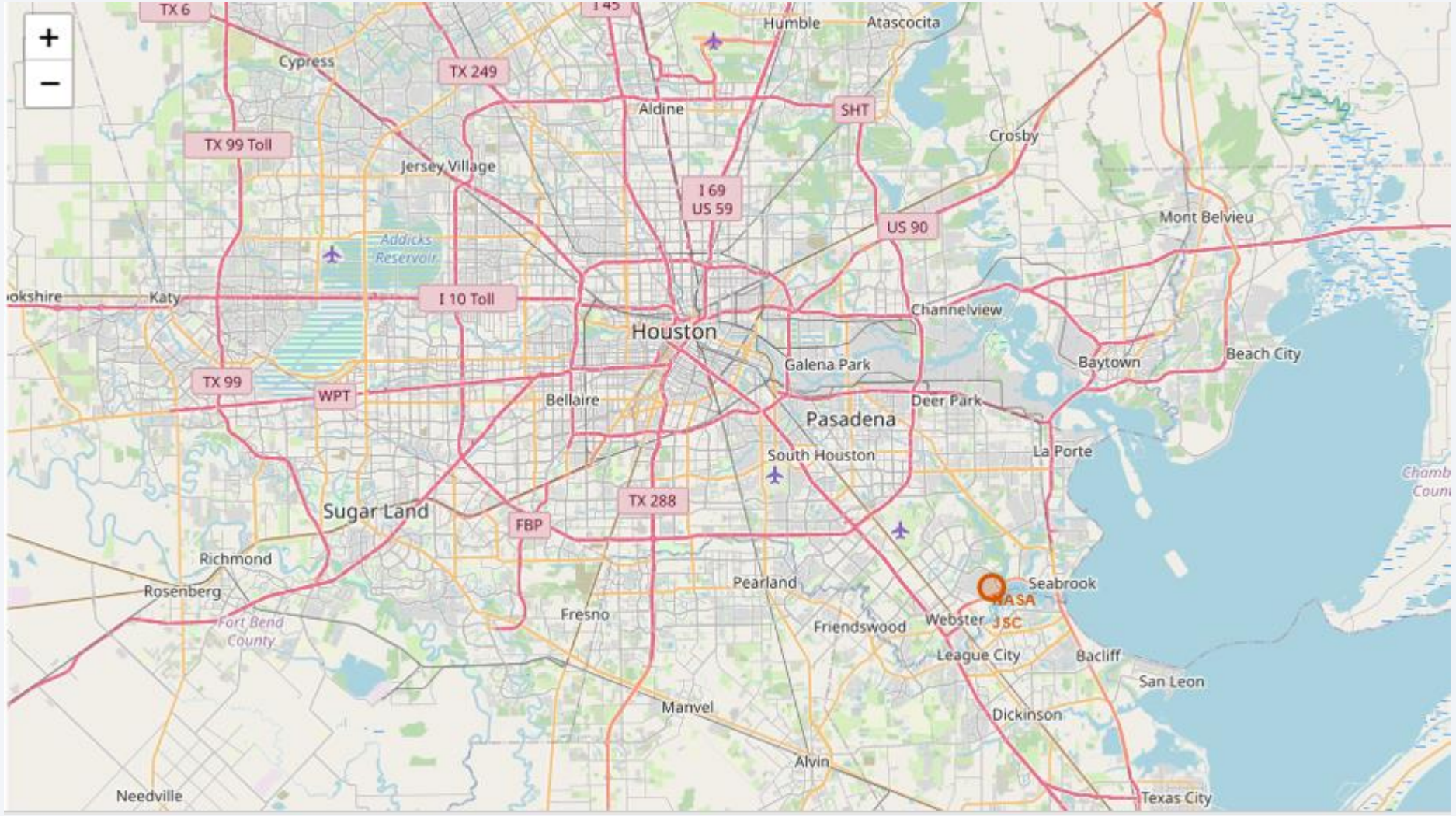
```
 * sqlite:///my_data1.db
Done.
```

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

Section 3

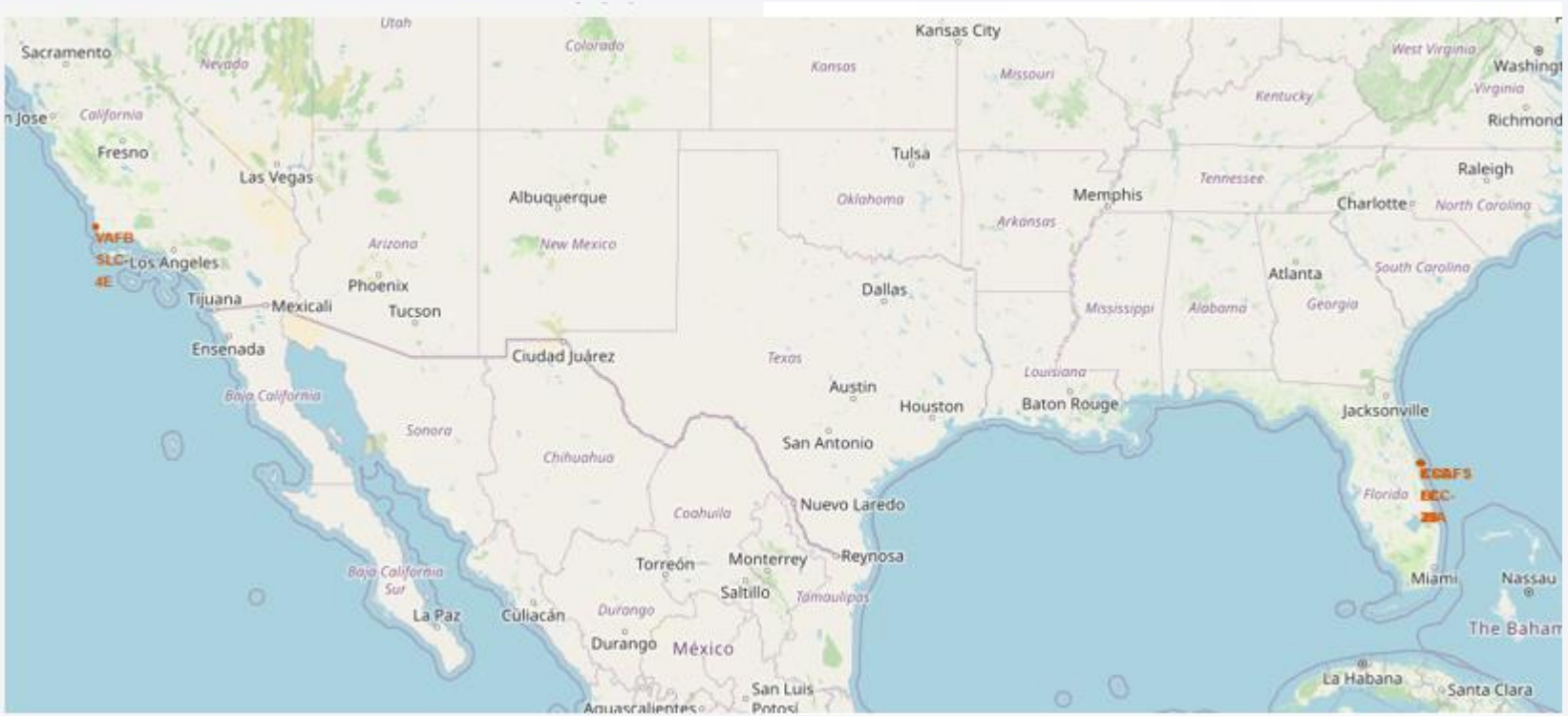
Launch Sites Proximities Analysis

<Folium Map Screenshot 1>

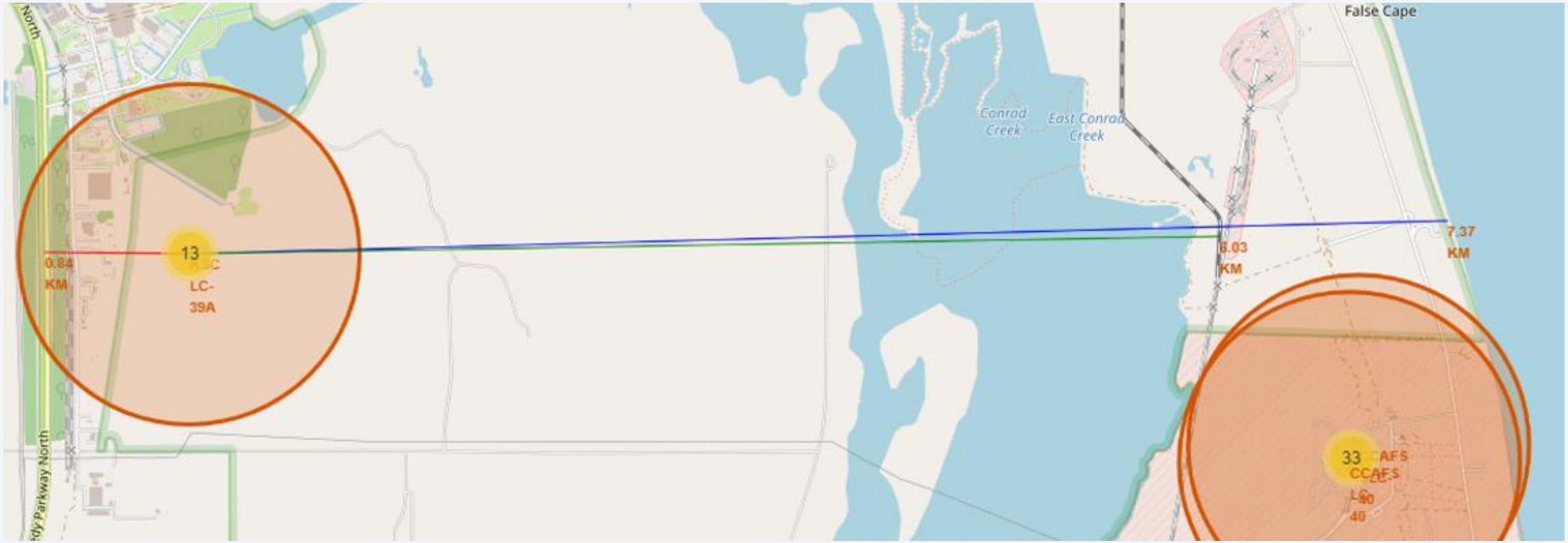


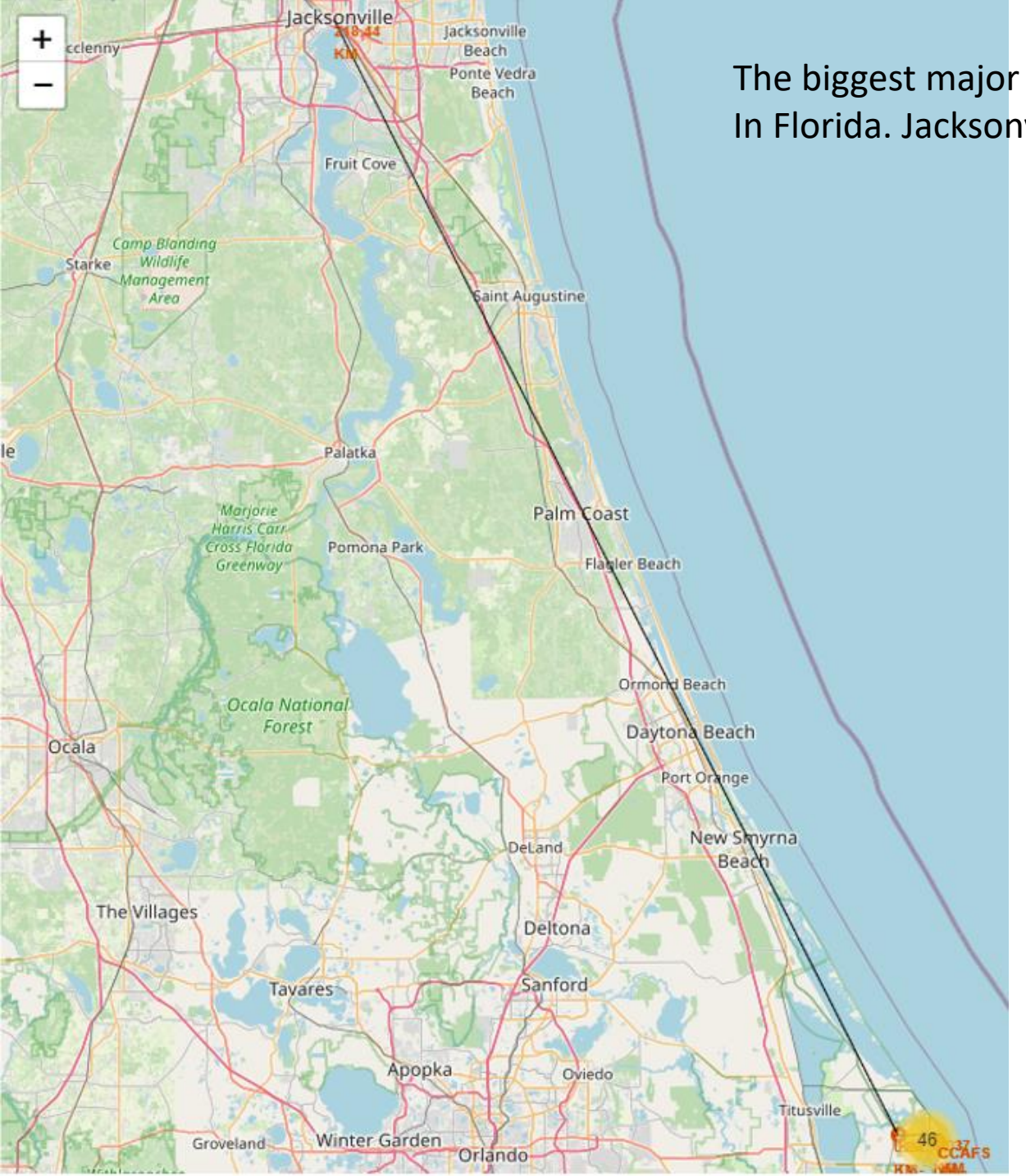
Upon zooming in to The marker for Houston you will Find a small yellow circle

Launch Sites with circle labels and popup labels

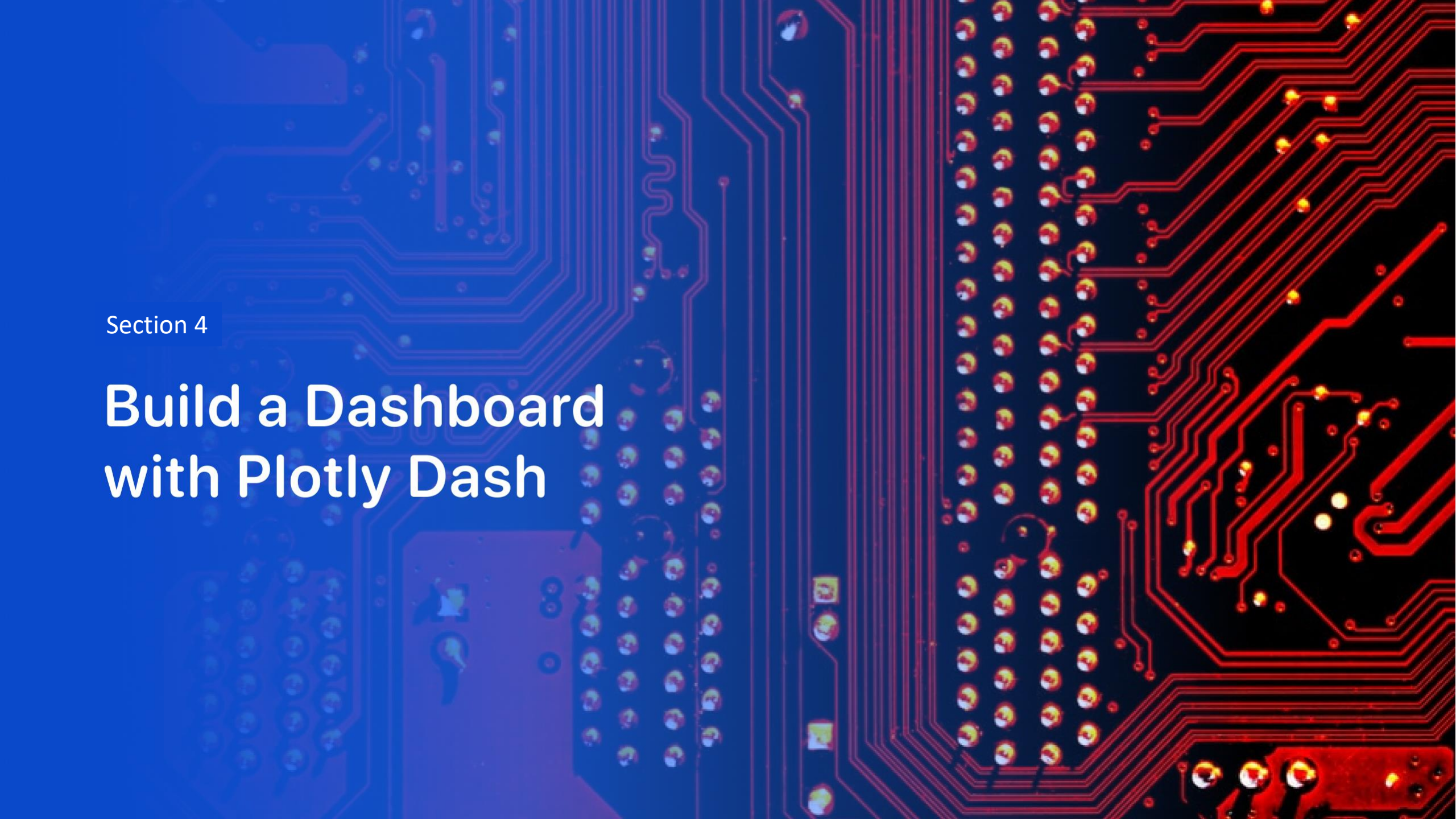


KSC LC-39A with distances to the coast, rail and highway

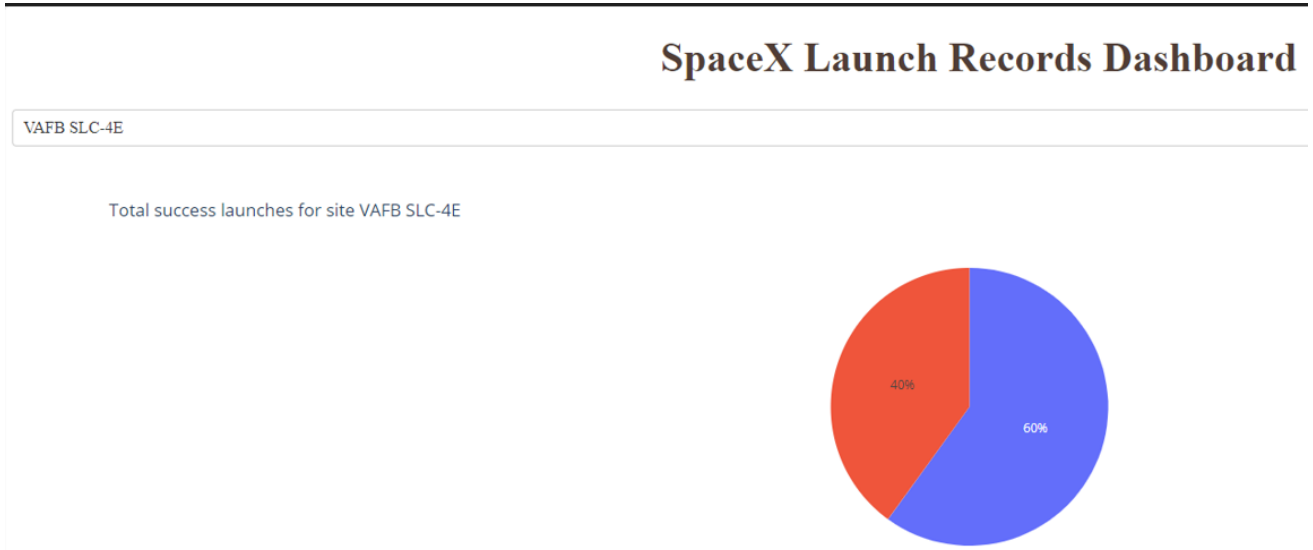


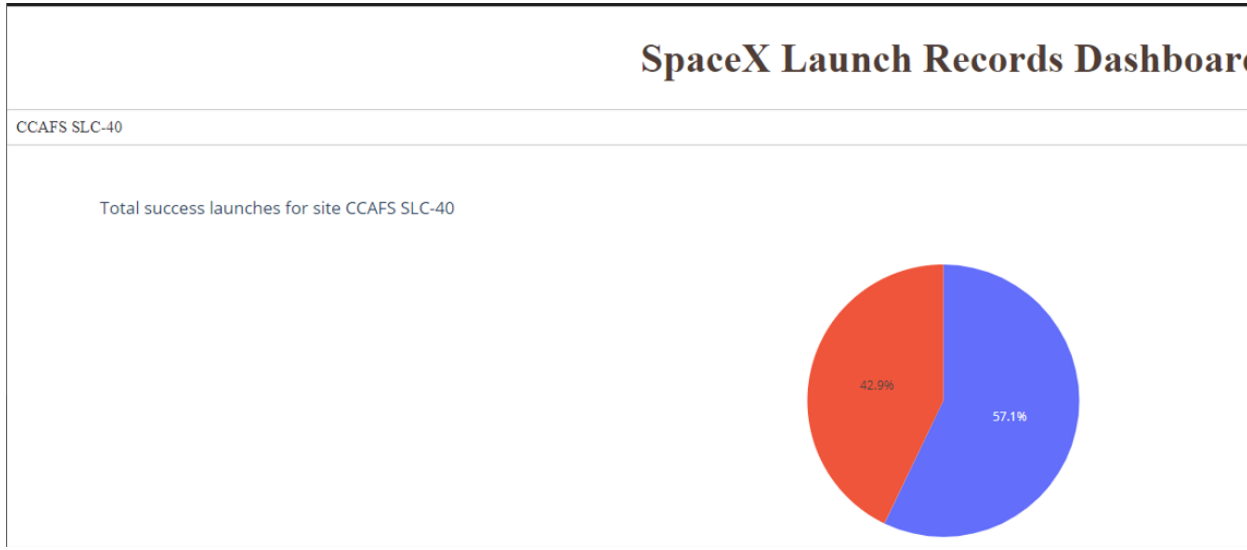


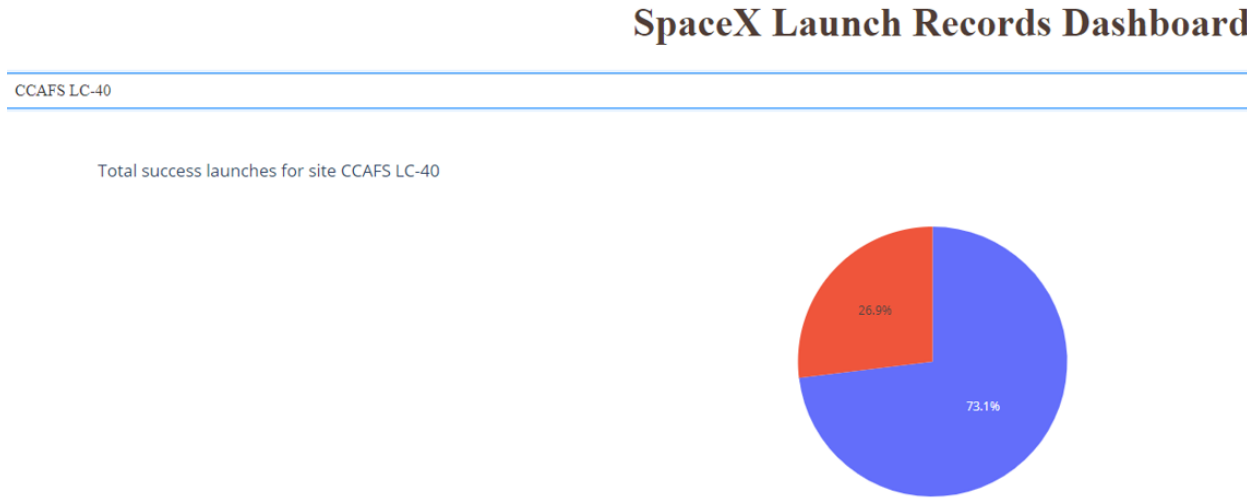
The biggest major city is far from the 2 sites
In Florida. Jacksonville is about 218 km away


Section 4
Build a Dashboard with Plotly Dash

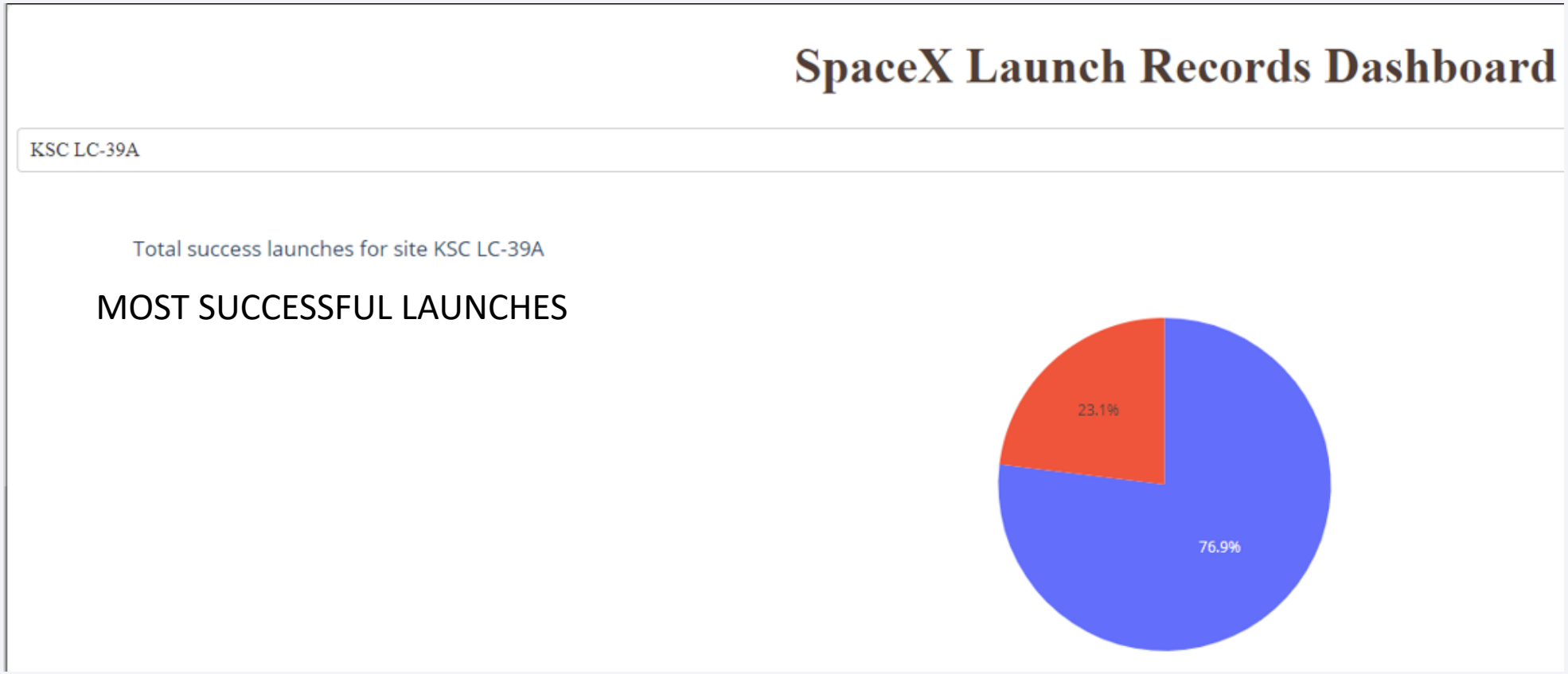
Dashboard by Site using Dash



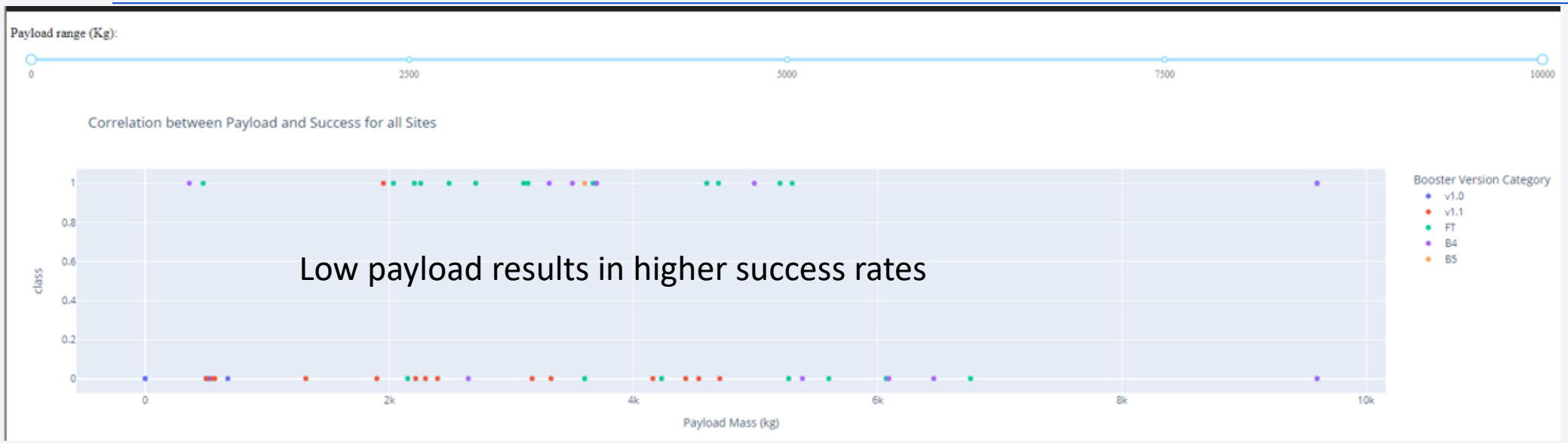


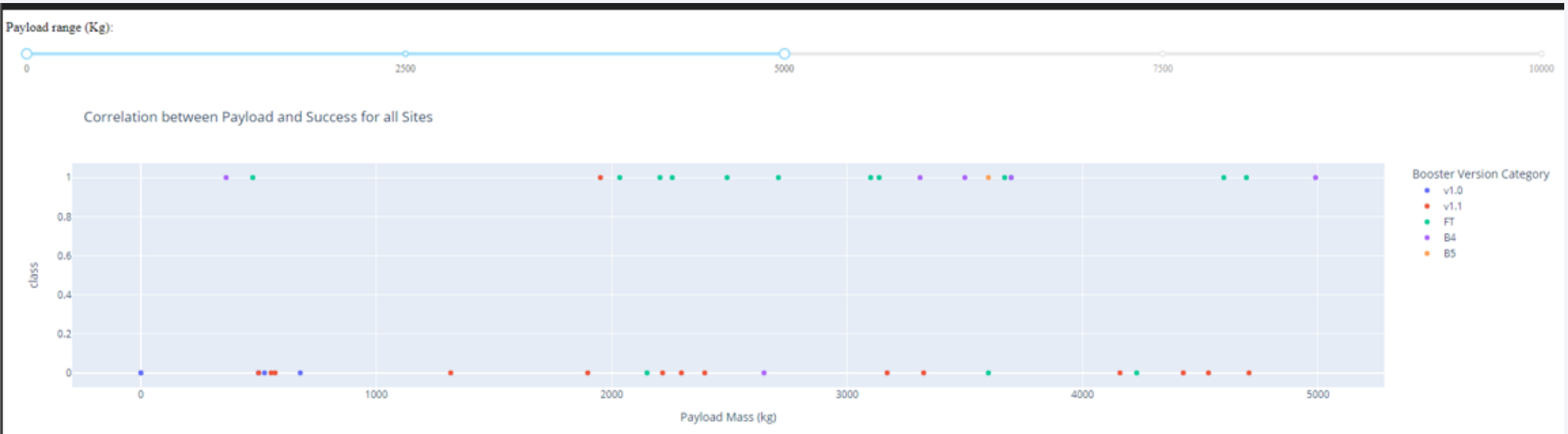


<Dashboard Screenshot 2>



Payload for all sites and Payload for Payload <= 5,000 kg

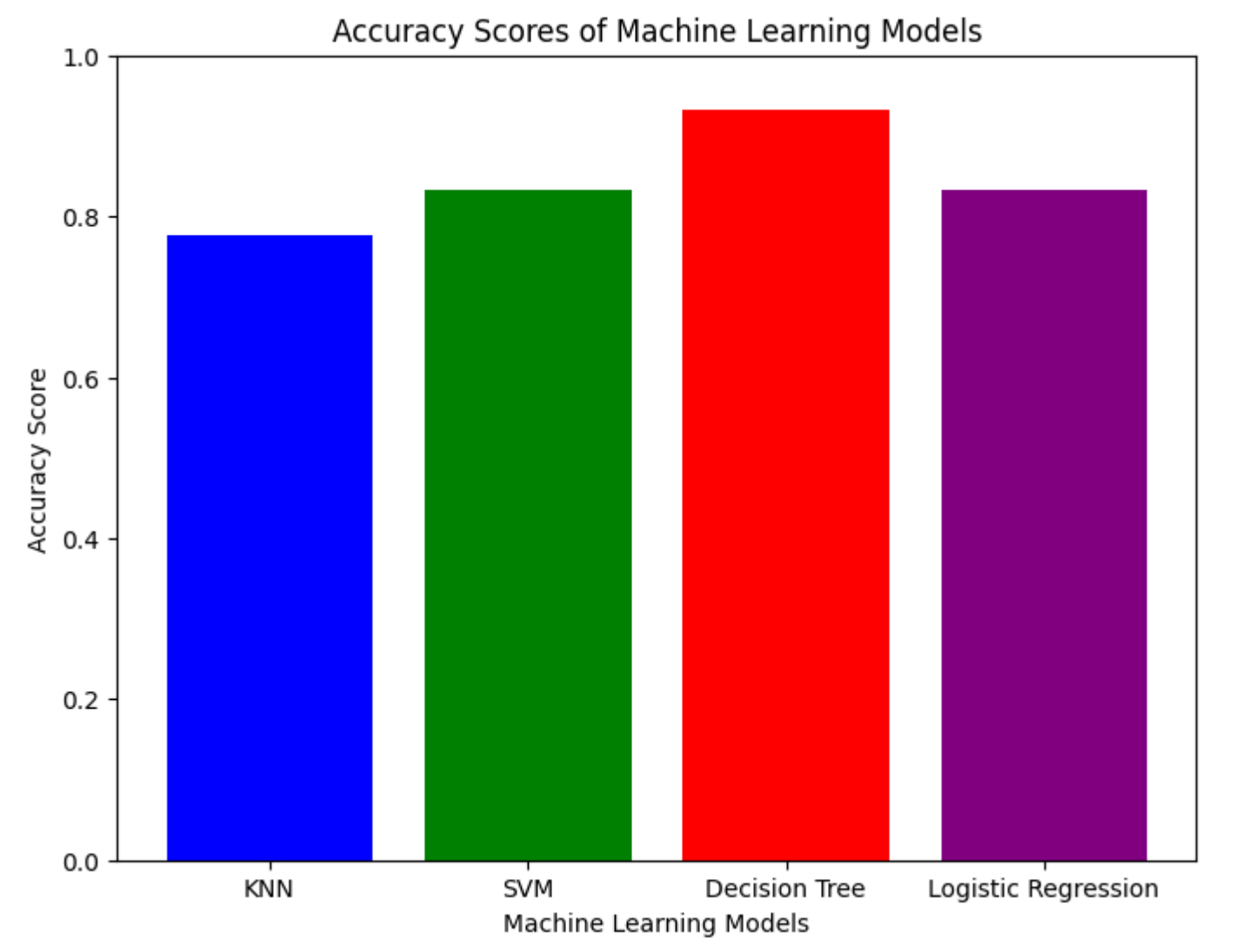




Section 5

Predictive Analysis (Classification)

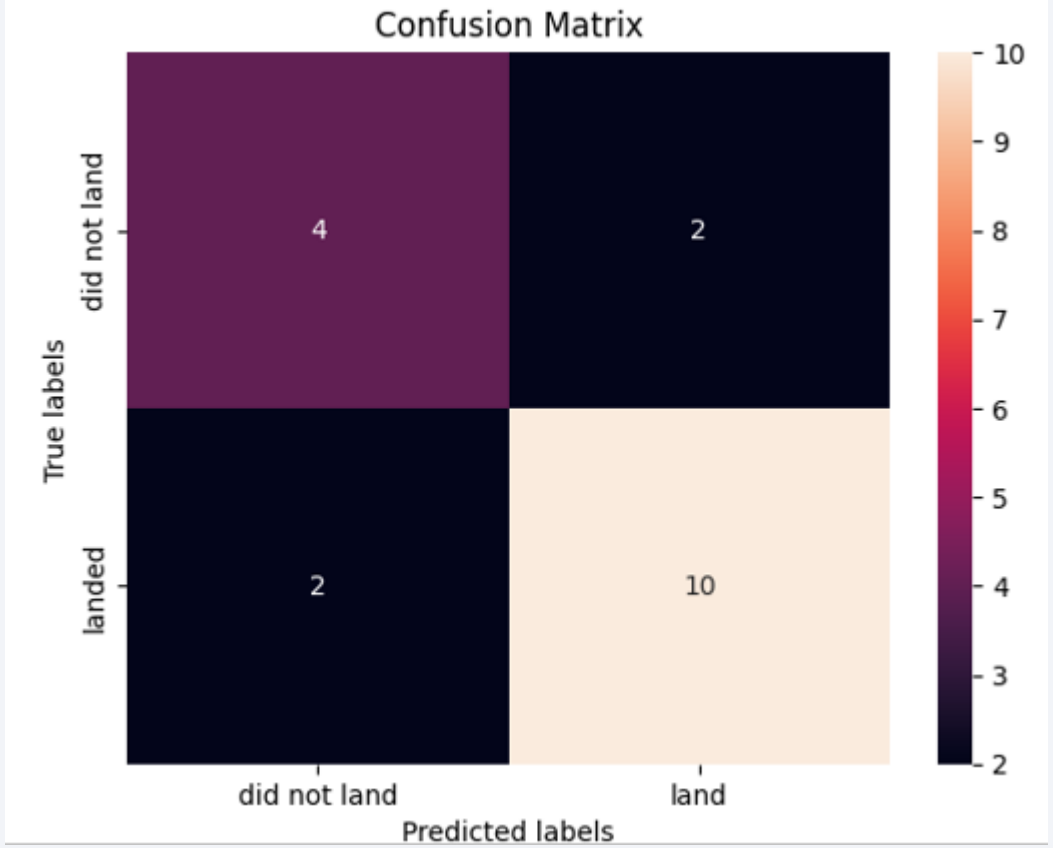
Classification Accuracy



Decision Tree gave the best results but All models were fairly close.

Confusion Matrix

KNN performed best as it had the least false Positives and more true negatives



Conclusions

- Successful landings of the Falcon 9 have increased since 2013 enabling SpaceX to beat their competitors on price as they can reuse part of the first stage.
- Lighter payloads fare better than heavier payloads
- KSC LC-39A had the most successful launches of our sites
- Orbits with the best success rates were: HEO, GEO, ES L1 and SSO

Appendix

- All codes can be found in my Github jgrajek/. I created separate repositories for each module so it would be easy to find the file as the descriptions the lab makes were not always evident.

Thank you!