IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Nuri ÖZBEY
05.09.2022

Outline

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

Executive Summary

Summary of methodologies

- Data Collection vith API,
- Data Colllection with Web Scraping
- Exploratory Data Analysis (EDA) with Data Visualization
- EDA with SQL
- Interactive Map with Folium
- Dashboards with Plotly Dash
- Predictive Analysis

Summary of all results

- Exploratory Data Analysis results
- Interactive maps and dashboard
- Predictive results

Introduction

Project background and context

❖ SpaceX advertises Falcon 9 rocket with a cost of 62 million dollars; other providers cost upwards of 165 million.

❖ Much of the savings is because SpaceX can reuse the first stage.

❖ SpaceY wants to determine the price of each launch, training a machine learning model and using public information to predict if SpaceX will reuse the first stage.

❖ Public information indicates a first stage Falcon 9 Booster to cost upwards of 15 million.

Problems you want to find answers

- Determine the price of each launch.
- If SpaceX will reuse the first stage and predict Falcon 9 Booster will land successfully or not.
- The machine learning model can predict the landing success of Falcon 9 Booster.

Section 1

Methodology

Methodology

Executive Summary

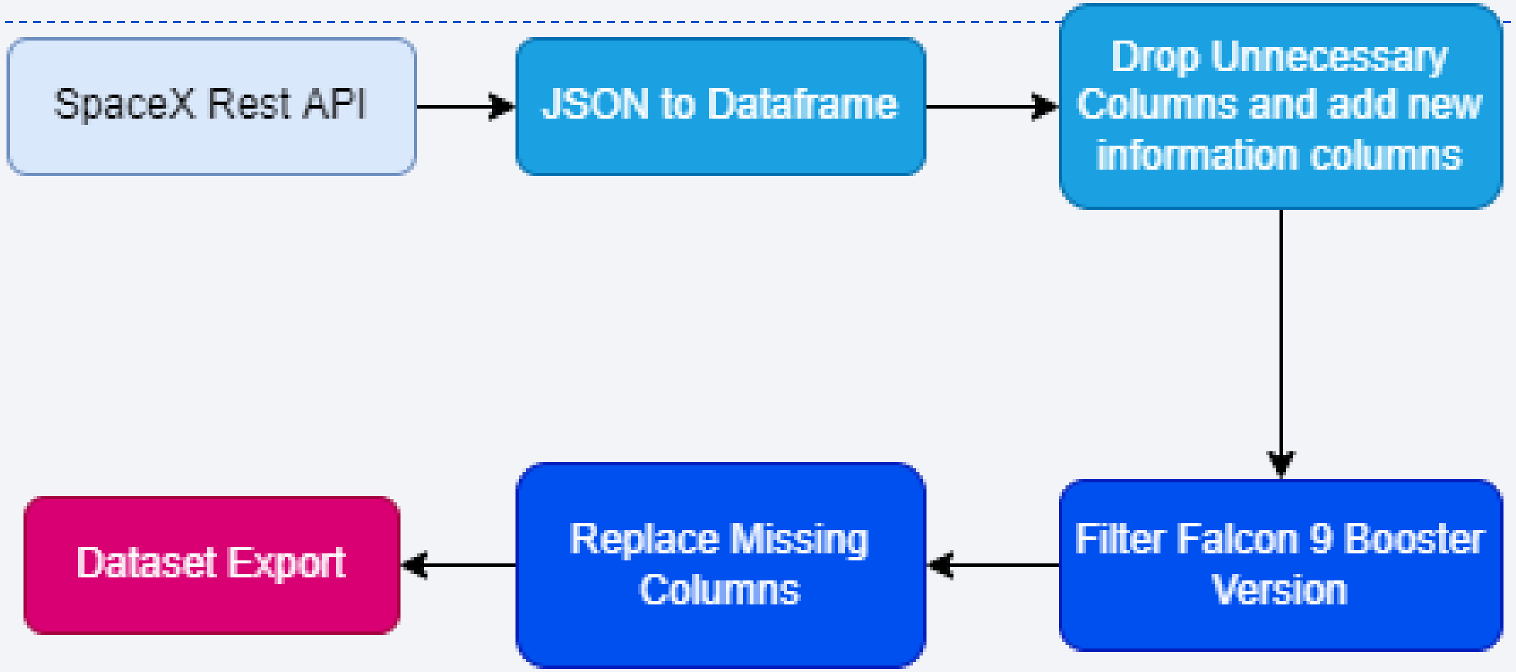
- Data collection methodology:
    - SpaceX API to DataFrame
    - SpaceX landing informations webscrapping on Wikipedia
- Perform data wrangling
    - Drop some columns
    - One hot encoding preprocessing
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
    - Tune machine learning models with different parameters and different classifiers

Data Collection – SpaceX API

- Data collected from spaceX https://api.spacexdata.com/v4/rockets/
- Converted to Json file from response String
- Apply Json Normalize operation to convert Pandas Dataframe
- Filter dataframe with necessary columns ('rocket', 'payloads', 'launchpad', 'cores', 'flight_number', 'date_utc')
- Append other columns clearly to Dataframe (flightNumber, BoosterVersion, PayloadMass, Orbit, LaunchSite, Outcome, Flights, GridFins, LandingPad, ReusedContent, Serial, Long-Lat)
- Filter only Falcon 9 Booster Version
- Replace missing values (payloadmass with mean, and landing pad with NaN)
- Dataset export and save.

Data Collection – SpaceX API

Flowchart of SpaceX API calls

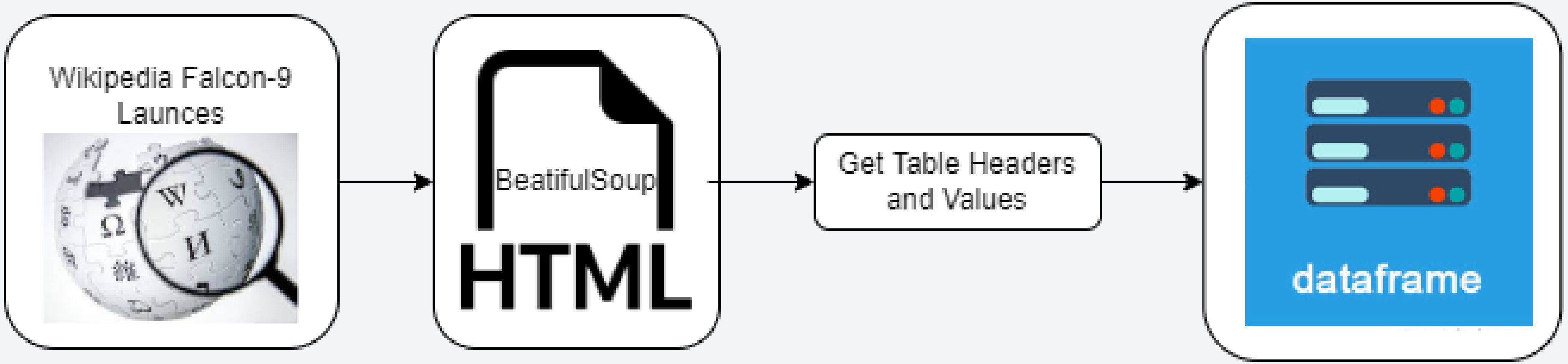


Data Collection – SpaceX API

- Data collected from Wikipedia https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches
- BeatifulSoup html parse from url
- Get column values with <th> object and extract to dataframe
- Dataframe dataset export and save.

Data Collection - Scraping

Flowchart of web scraping here



Data Wrangling

- Create a Column «outcome» when is bad assign 0 , when is not bad assign 1

- https://github.com/nuriozbey/CourseraTrainings/blob/main/Applied%20Data%20Science%20Capstone/Data%20wrangling.ipynb

EDA with Data Visualization

- Scatter plots
- Bar Charts
- Line Charts
- Feature engineering with one hot encoding of columns

- https://github.com/nuriozbey/CourseraTrainings/blob/main/Applied%20Data%20Science%20Capstone/exploratory-data-analysis-dataviz.ipynb

EDA with SQL

Performed SQL queries

- 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 acheived.
- 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 failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.
- Rank the count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order

https://github.com/nuriozbey/CourseraTrainings/blob/main/Applied%20Data%20Science%20Capstone/exploratory-data-analysis-SQL.ipynb

Build an Interactive Map with Folium

Folium map object is a map centered on NASA Johnson Space Center at Houson, Texas

- Red circle at NASA Johnson Space Center's coordinate with label showing its name
- Red circles at each launch site coordinates with label showing launch site name
- The grouping of points in a cluster to display multiple coordinates
- Markers to show successful and unsuccessful landings. Green for successful landing and Red for unsuccessful landing.
- Markers to show distance between launch site to key locations (railway, highway, coastway, city) and plot a line between them

https://github.com/nuriozbey/CourseraTrainings/blob/main/Applied%20Data%20Science%20Capstone/launch_site_location-follium.ipynb

Build a Dashboard with Plotly Dash

Manimulate Data in interactive filtering options real time

- Pie Chart for success rate
- Scatter chart for payload-mass vs outcome
- Drop-down menu for choosing lauch sites

https://github.com/nuriozbey/CourseraTrainings/blob/main/Applied%20Data%20Science%20Capstone/spacex_dash_app.py

Predictive Analysis (Classification)

Data preparation

- Load dataset
- Normalize data
- Split data into training and test sets.

Model preparation

- Selection of machine learning algorithms
- Set parametersfor each algorithm to GridSearchCV
- Training GridSearchModel models with training dataset

Model evaluation

- Get best hyperparameters for each type of model
- Compute accuracy for each model with test dataset
- Plot Confusion Matrix

Model comparison

- Comparison of models according to their accuracy
- The model with the best accuracy will be chosen (see Notebook for result)

https://github.com/nuriozbey/CourseraTrainings/blob/main/Applied%20Data%20Science%20Capstone/Machine-Learning-Prediction.ipynb

Results

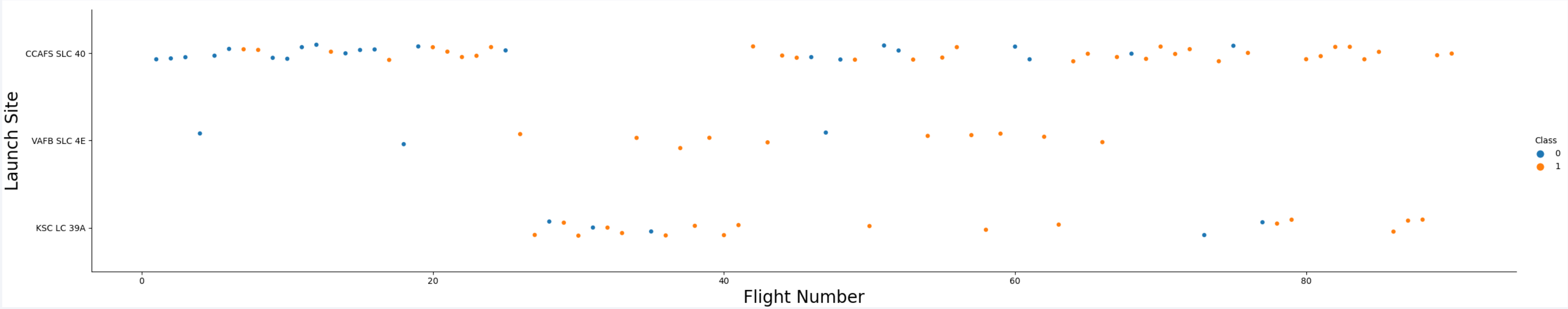
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

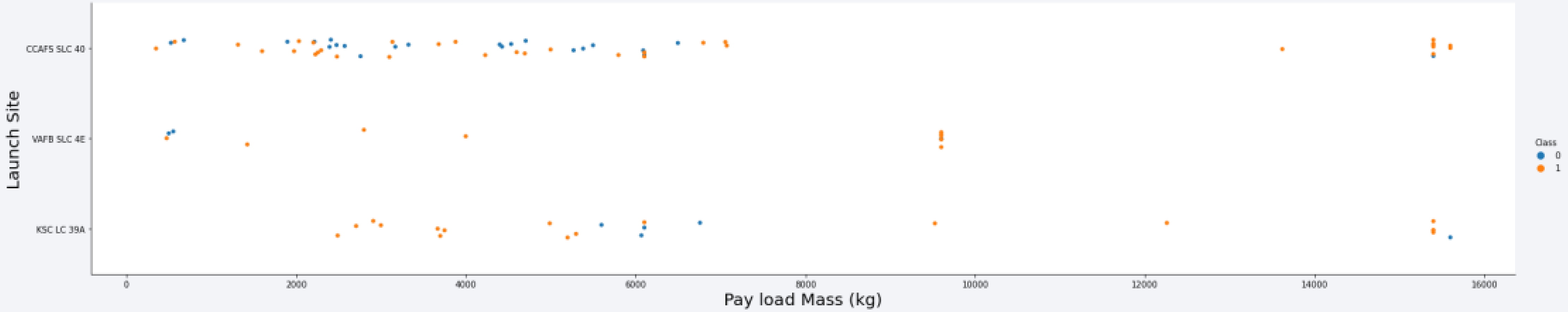
Flight Number vs. Launch Site

- Scatter plot of Flight Number vs. Launch Site



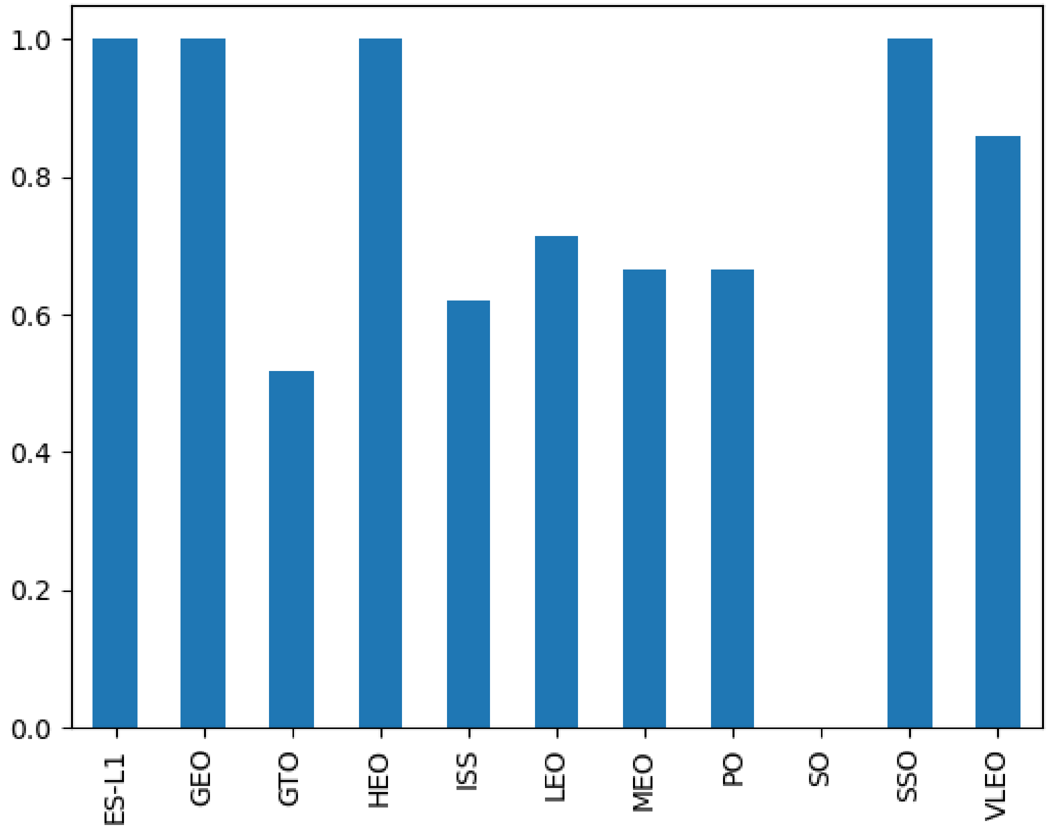
Payload vs. Launch Site

- Scatter plot of Payload vs. Launch Site



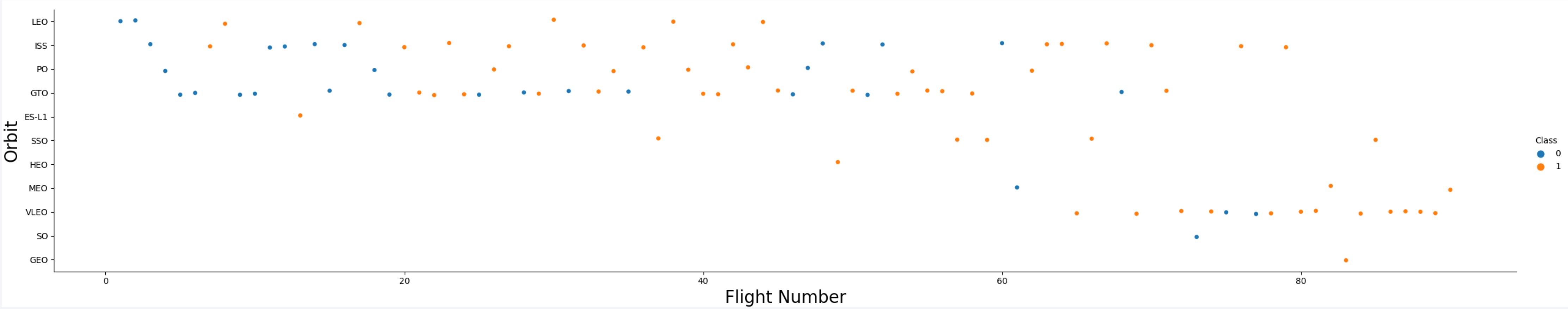
Success Rate vs. Orbit Type

- Bar chart for the success rate of each orbit type



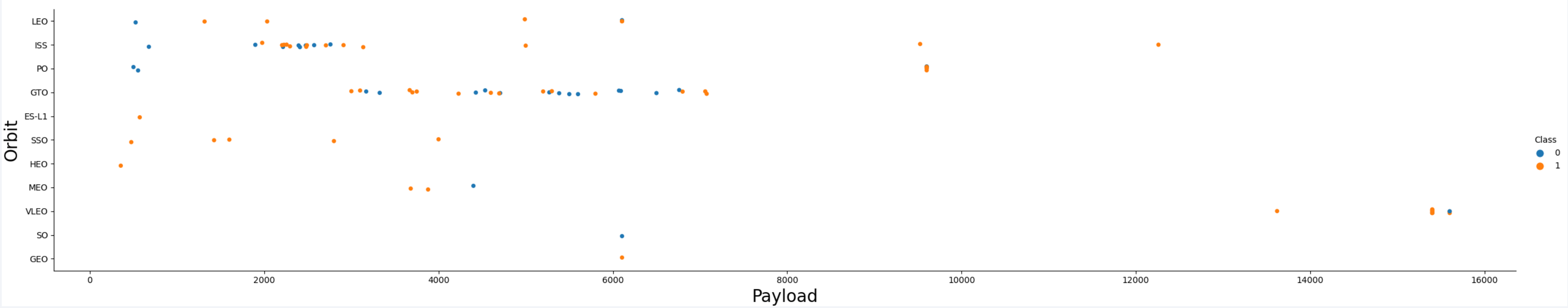
Flight Number vs. Orbit Type

- Scatter point of Flight number vs. Orbit type



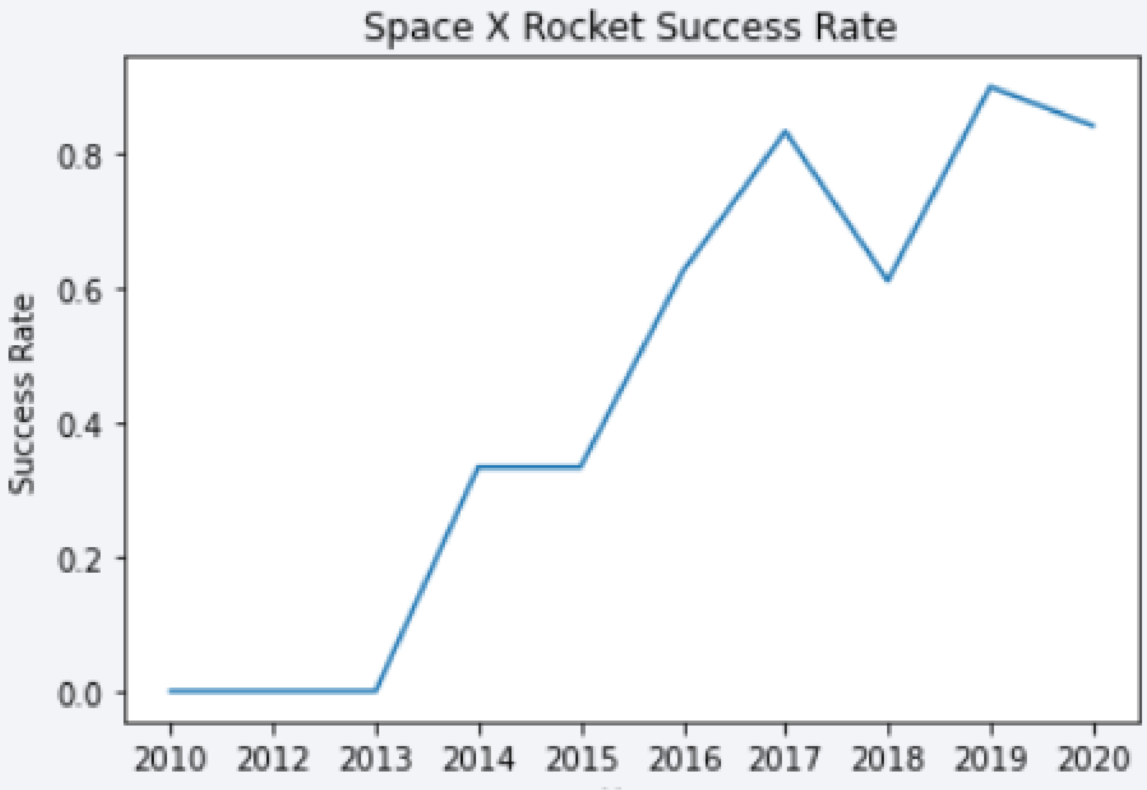
Payload vs. Orbit Type

- Scatter point of payload vs. orbit type



Launch Success Yearly Trend

- Line chart of yearly average success rate



All Launch Site Names

- SQL Query:

SELECT DISTINCT "LAUNCH_SITE" FROM SPACEXTB

- Result:

Out[12]:

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

Launch Site Names Begin with 'CCA'

- SQL Query:

SELECT * FROM SPACEXTBL WHERE "LAUNCH_SITE" LIKE '%CCA%' LIMIT 5

- Result:

Out[13]:

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

Total Payload Mass

- SQL Query:

SELECT SUM("PAYLOAD_MASS__KG_") FROM SPACEXTBL WHERE "CUSTOMER" = 'NASA (CRS)’

- Result:

Done.

Out[14]:

SUM("PAYLOAD_MASS__KG_")
45596

Average Payload Mass by F9 v1.1

- SQL Query:

SELECT AVG("PAYLOAD_MASS__KG_") FROM SPACEXTBL WHERE "BOOSTER_VERSION" LIKE '%F9 v1.1%'

- Result:

Out[15]:

AVG("PAYLOAD_MASS__KG_")
2534.6666666666665

First Successful Ground Landing Date

- SQL Query:

SELECT MIN("DATE") FROM SPACEXTBL WHERE "Landing _Outcome" LIKE '%Success%’

Result:

Out[16]:

MIN("DATE")
01-05-2017

Successful Drone Ship Landing with Payload between 4000 and 6000

- SQL Query:

```
SELECT "BOOSTER_VERSION" FROM SPACEXTBL WHERE "LANDING
_OUTCOME" = 'Success (drone ship)' AND "PAYLOAD_MASS__KG_" >= 4000
AND "PAYLOAD_MASS__KG_" < 6000;
```

Result:

Out[18]:

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:

SELECT (SELECT COUNT("MISSION_OUTCOME") FROM SPACEXTBL WHERE "MISSION_OUTCOME" LIKE '%Success%') AS SUCCESS, (SELECT COUNT("MISSION_OUTCOME") FROM SPACEXTBL WHERE "MISSION_OUTCOME" LIKE '%Failure%') AS FAILURE

Result:

Out[19]:

SUCCESS	FAILURE
100	1

Boosters Carried Maximum Payload

- SQL Query:

SELECT DISTINCT "BOOSTER_VERSION" FROM SPACEXTBL WHERE "PAYLOAD_MASS__KG_" = (SELECT max("PAYLOAD_MASS__KG_") FROM SPACEXTBL)

Result:

Out[20]:

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 Query:

SELECT substr("DATE", 4, 2) AS MONTH, "BOOSTER_VERSION", "LAUNCH_SITE" FROM SPACEXTBL WHERE "LANDING _OUTCOME" = 'Failure (drone ship)' and substr("DATE",7,4) = '2015'

Result:

Out[21]:

MONTH	Booster_Version	Launch_Site
01	F9 v1.1 B1012	CCAFS LC-40
04	F9 v1.1 B1015	CCAFS LC-40

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

- SQL Query:

SELECT "LANDING _OUTCOME", COUNT("LANDING _OUTCOME") FROM SPACEXTBL WHERE "DATE" >= '04-06-2010' and "DATE" <= '20-03-2017' and "LANDING _OUTCOME" LIKE '%Success%’ GROUP BY "LANDING _OUTCOME" ORDER BY COUNT("LANDING _OUTCOME") DESC ;

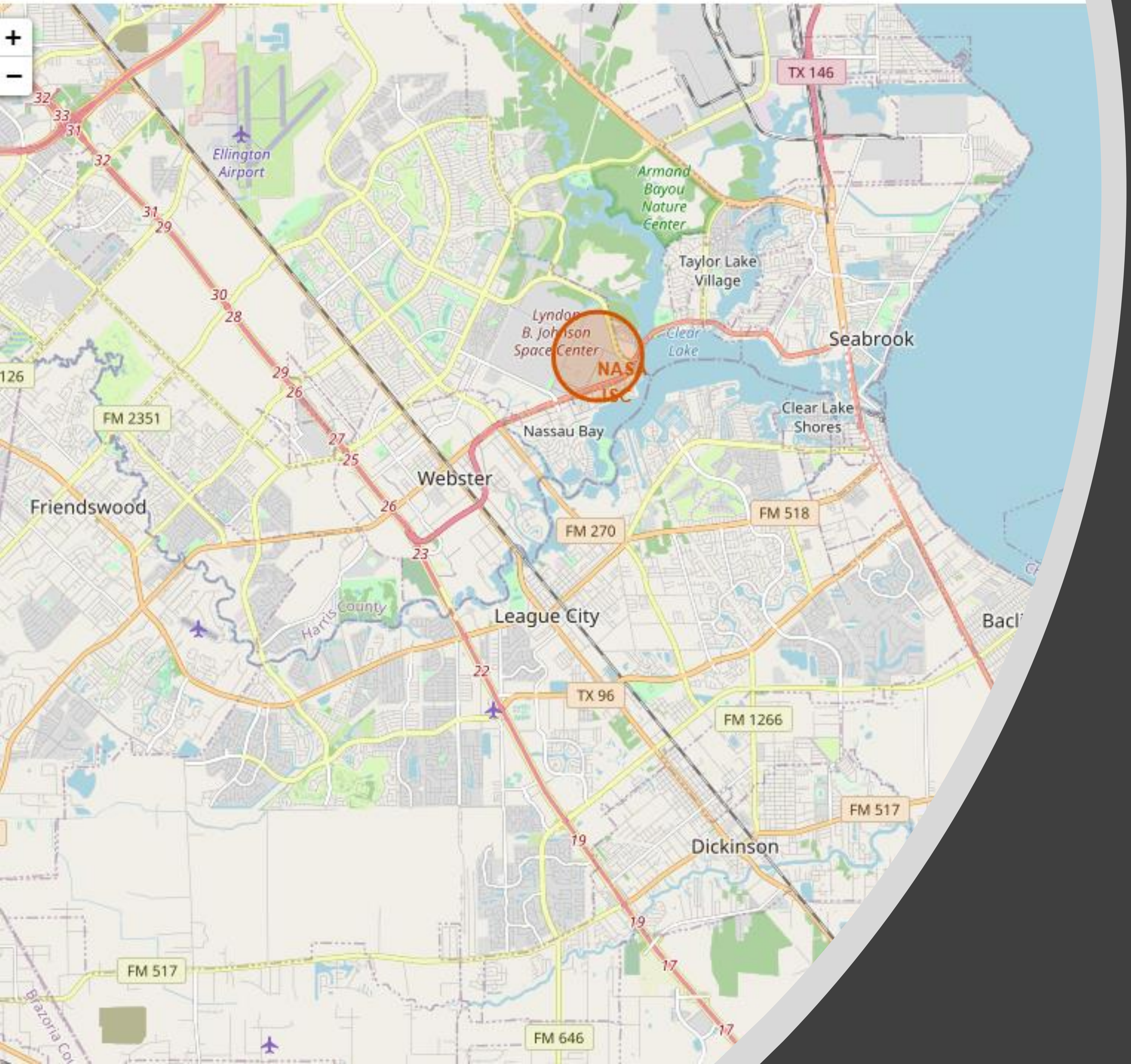
Result:

Out[22]:

Landing _Outcome	COUNT("LANDING _OUTCOME")
Success	20
Success (drone ship)	8
Success (ground pad)	6

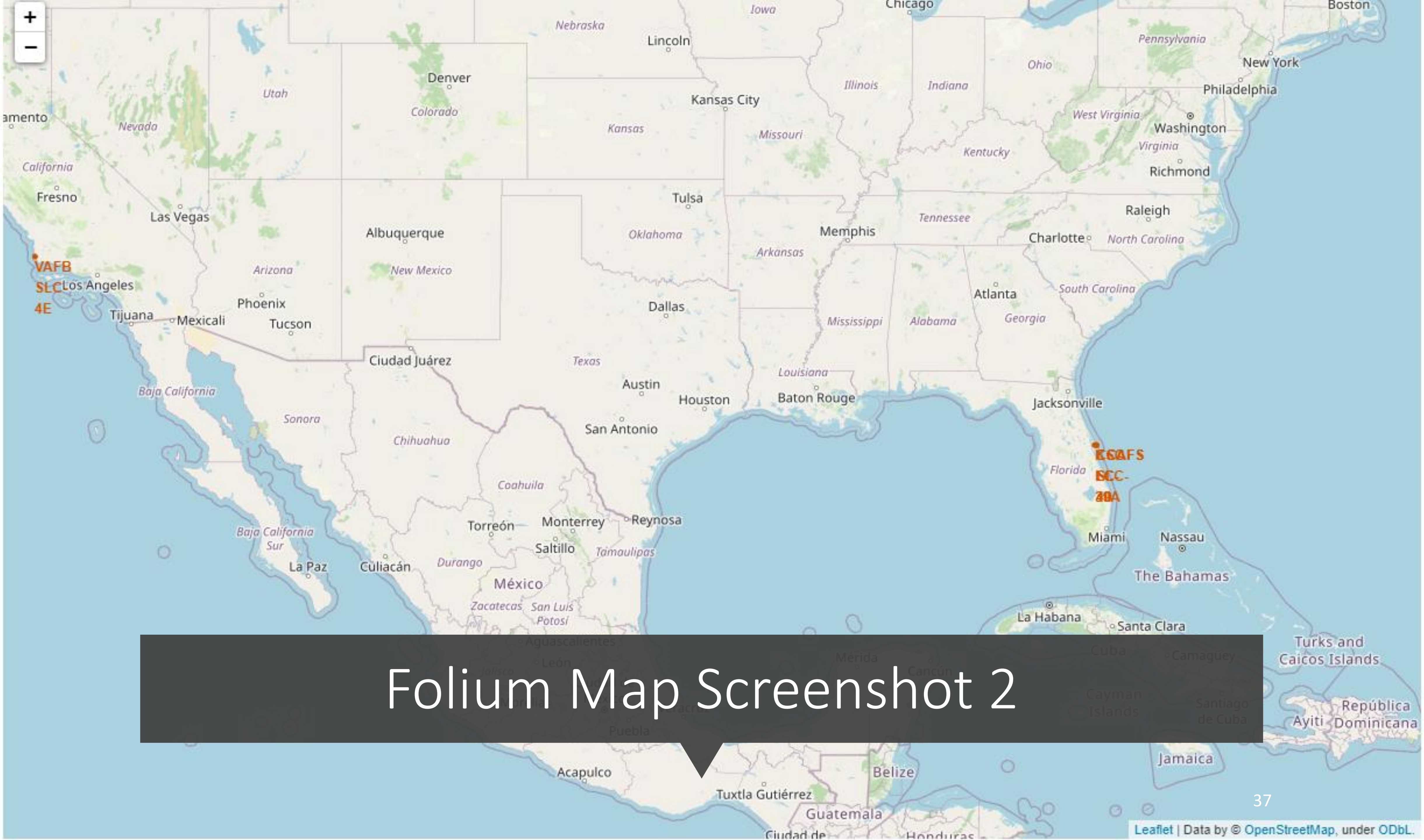
Section 3

Launch Sites Proximities Analysis



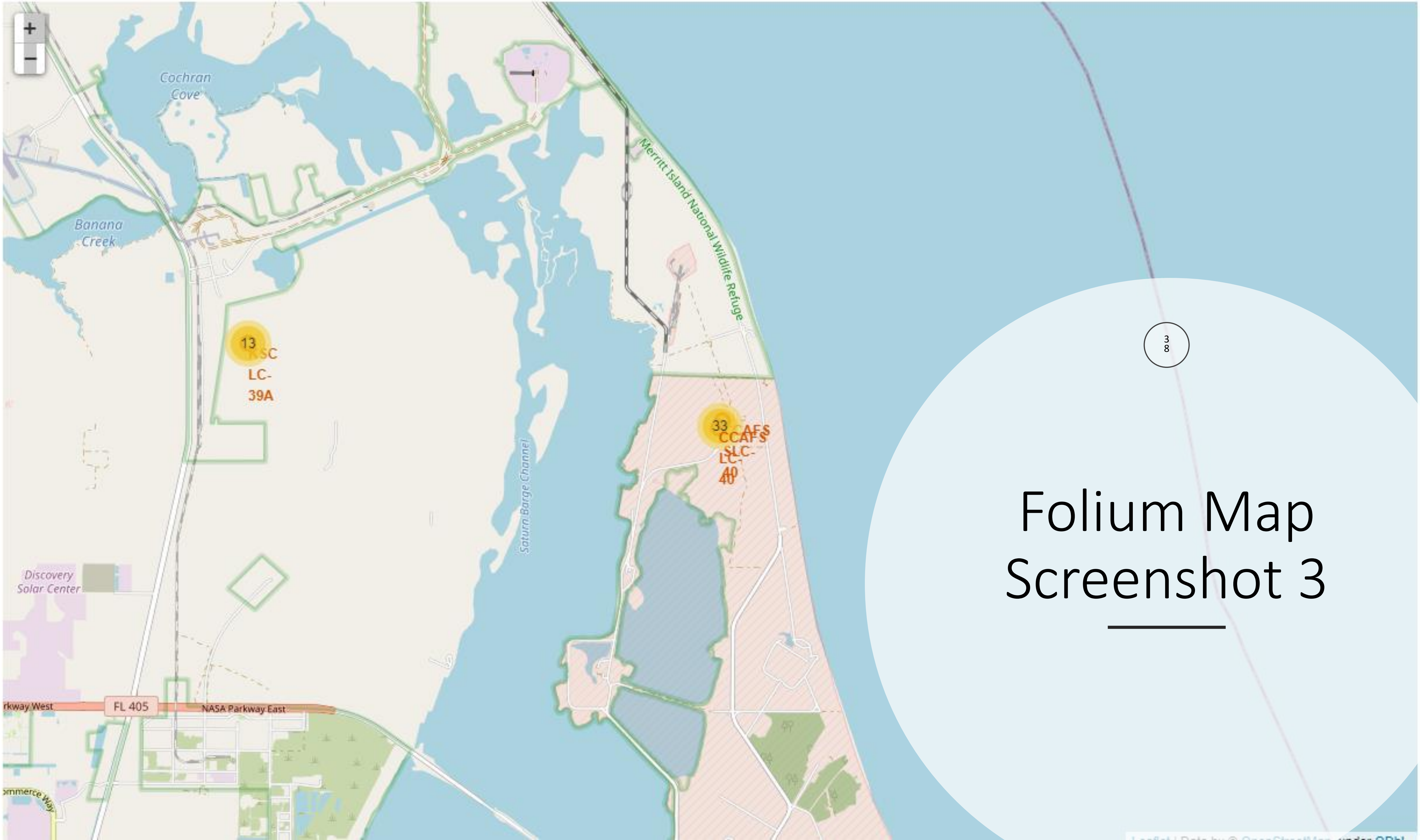
Folium Map Screenshot 1

[9]:



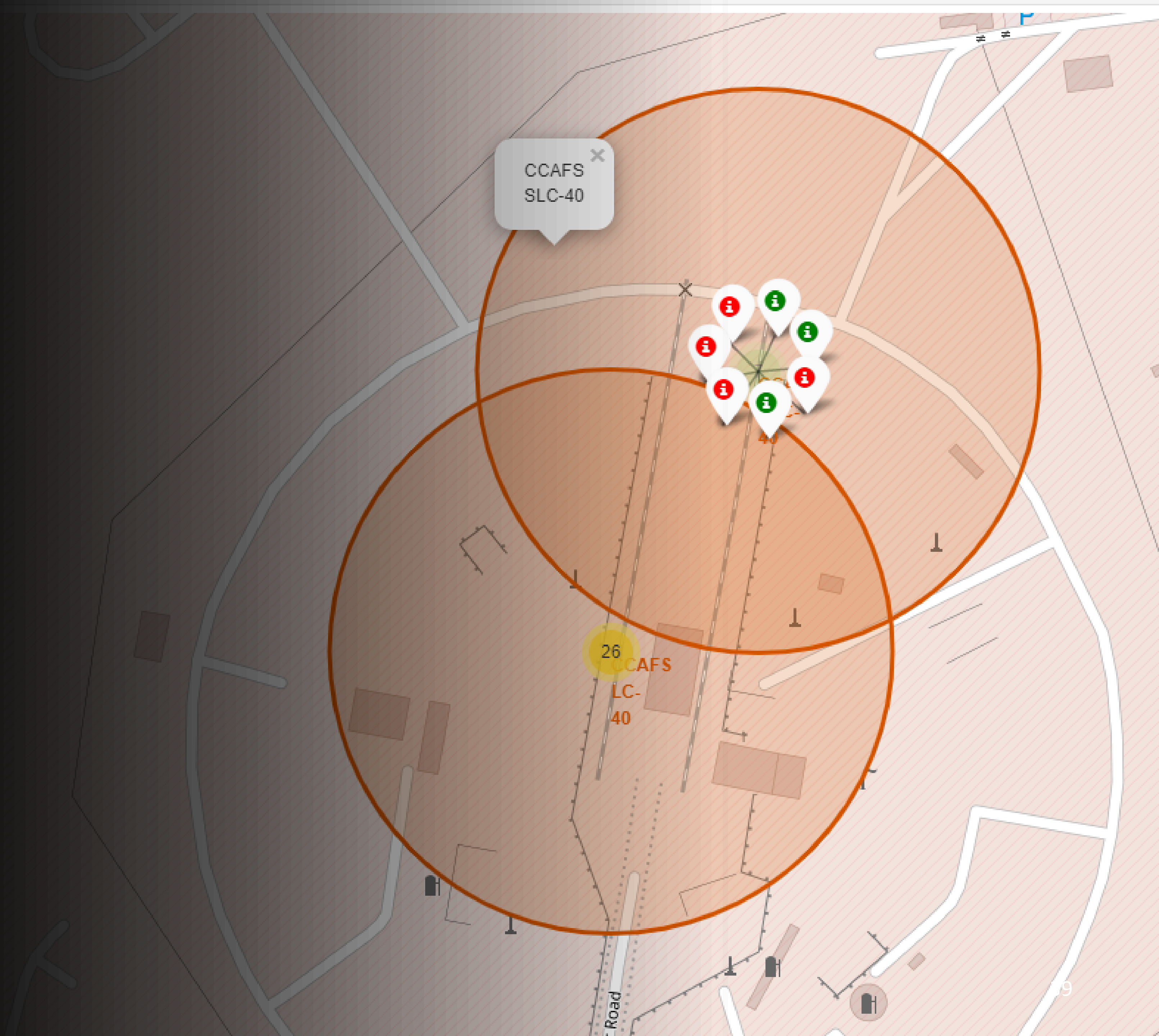
Folium Map Screenshot 2

[14]:

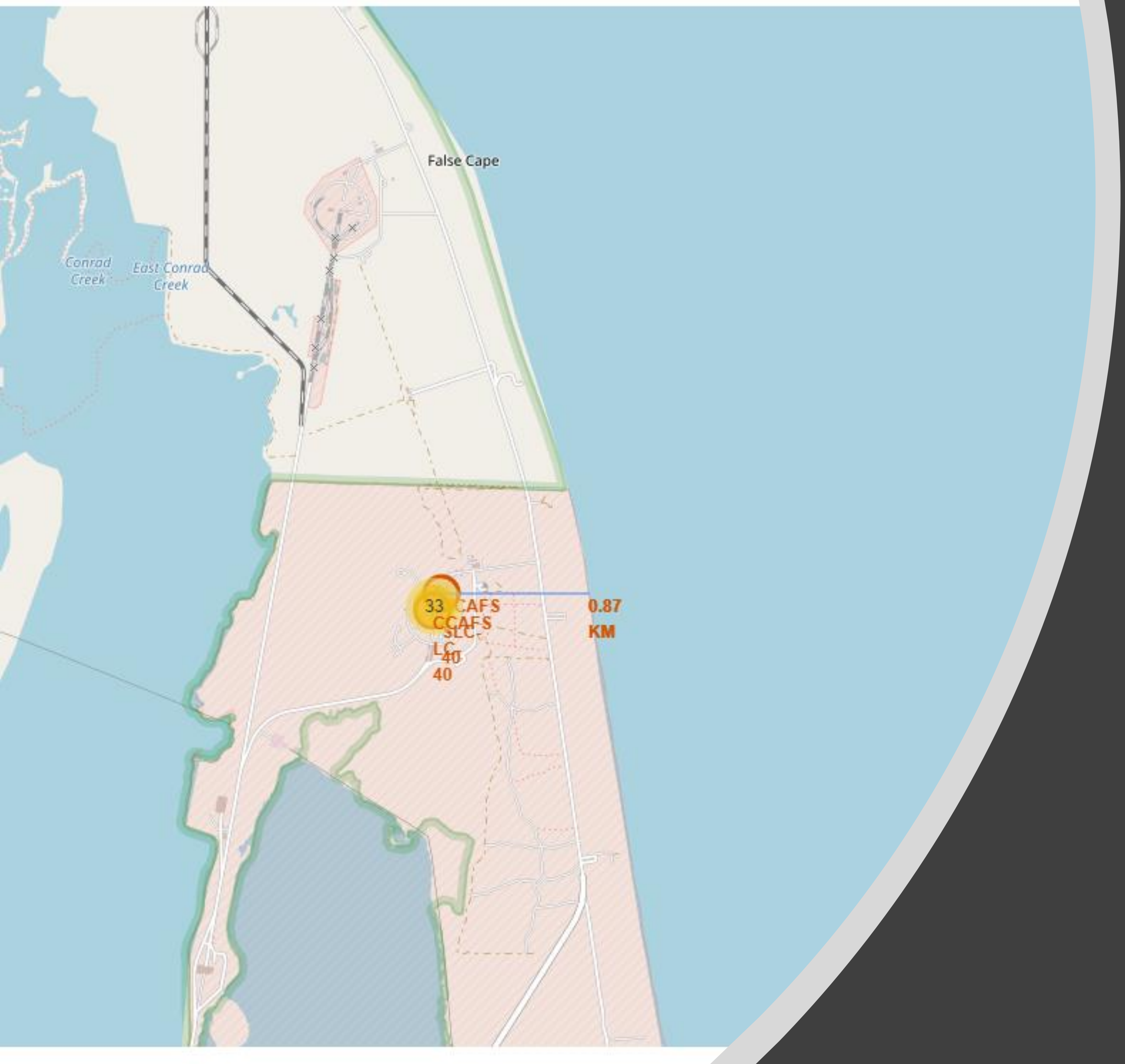


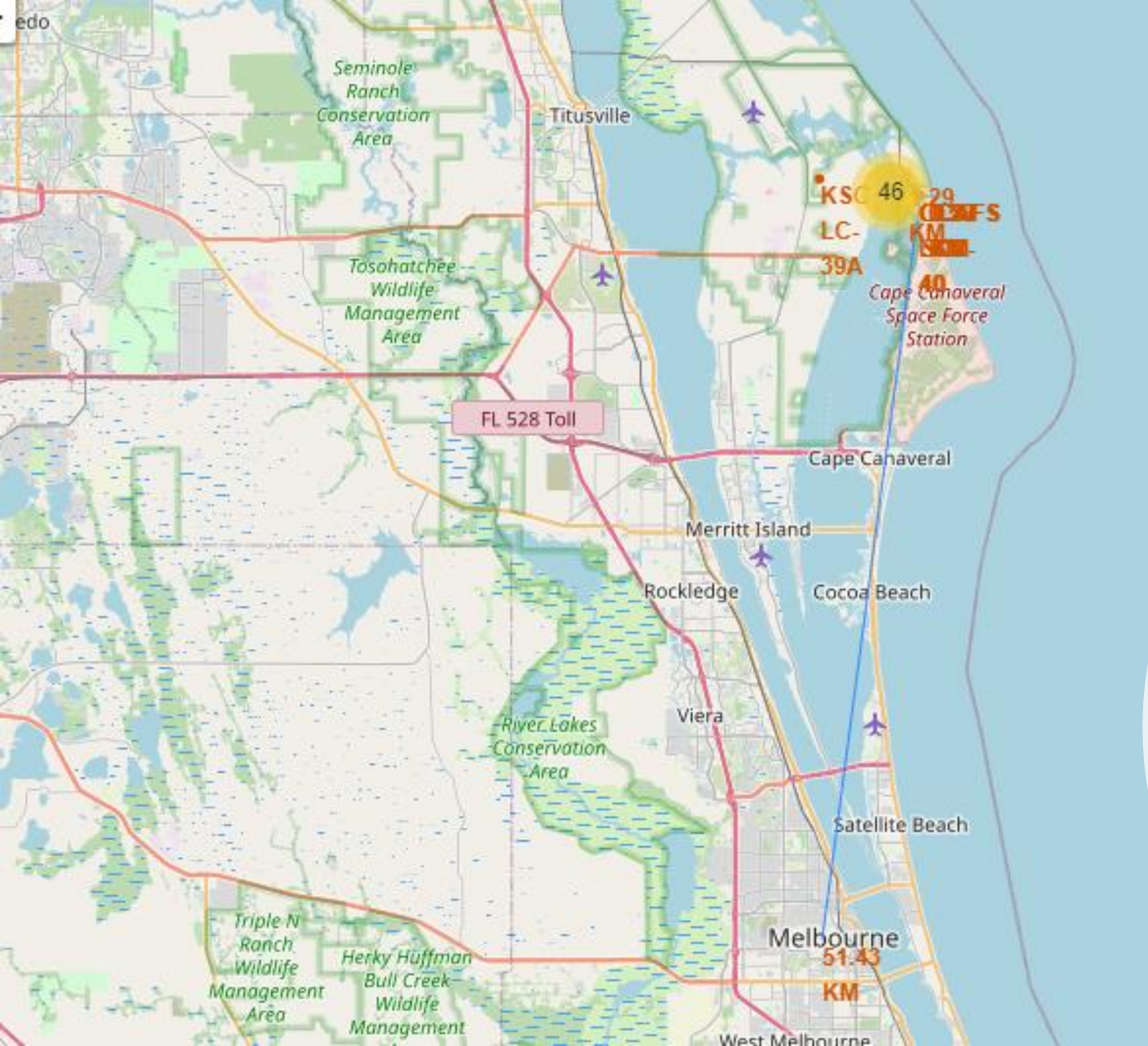
Folium Map Screenshot 3

Folium Map Screenshot 4



Folium Map Screenshot 5



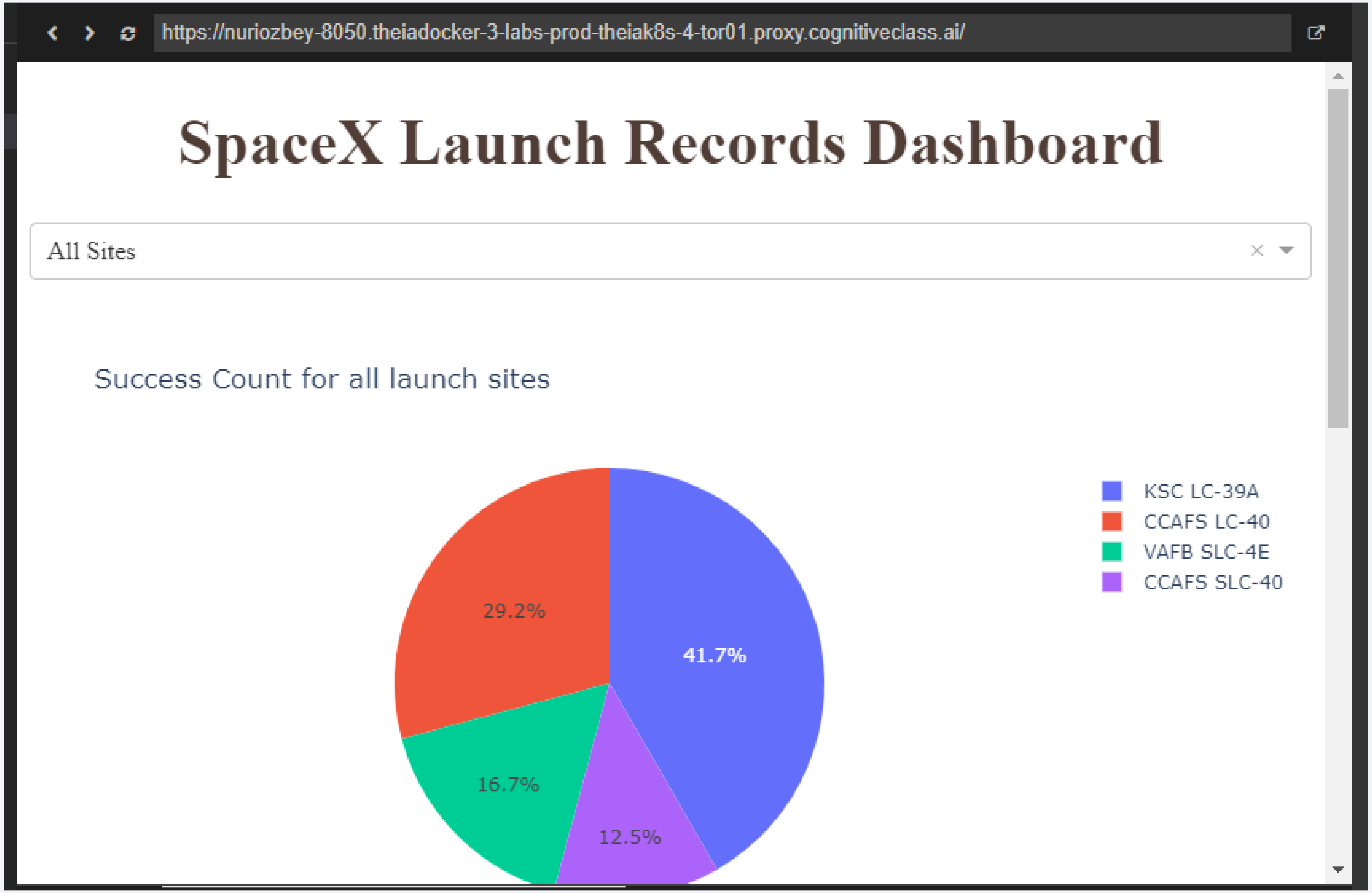


Folium Map Screenshot 6

Section 4

Build a Dashboard with Plotly Dash

Interactive Dashboard Results



Interactive Dashboard Results

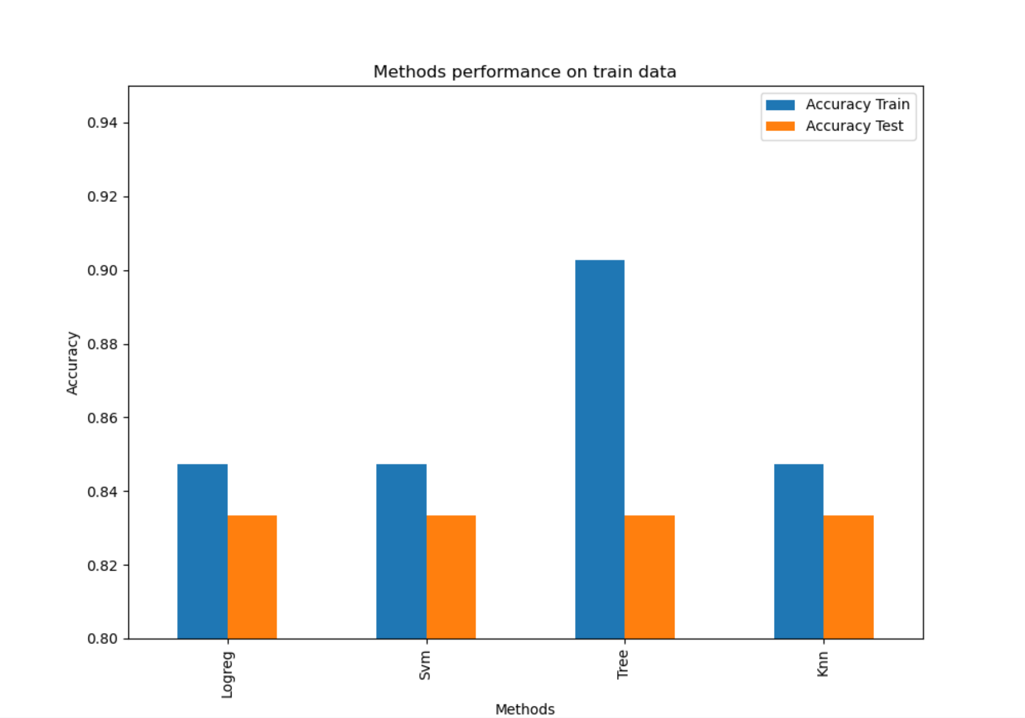
Section 5

Predictive Analysis (Classification)

Classification Accuracy

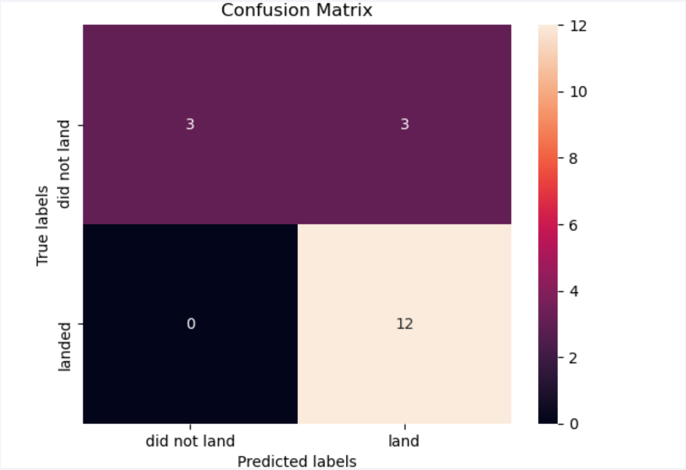
	Accuracy Train	Accuracy Test
Tree	0.902778	0.833333
Logreg	0.847222	0.833333
Svm	0.847222	0.833333
Knn	0.847222	0.833333

- The best train accuracy in Decision Tree Classifier but test result generally gives the same results % 83.33

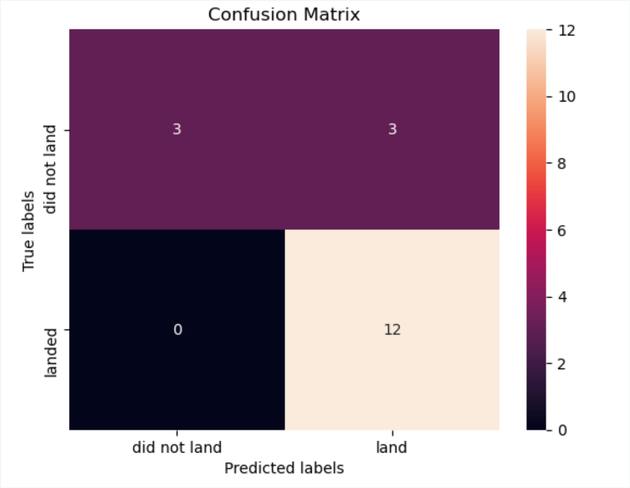


Confusion Matrix

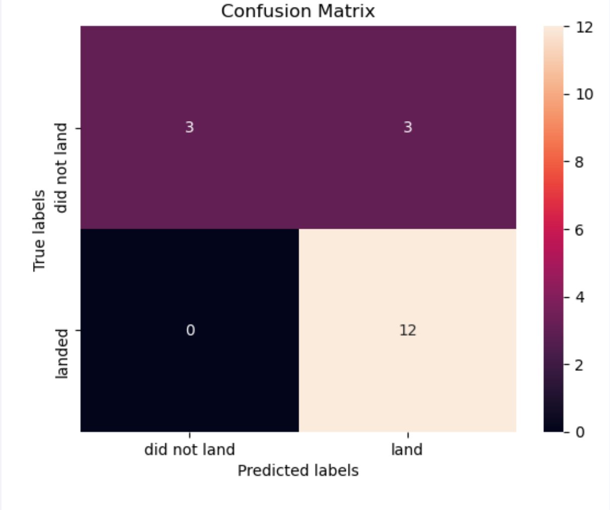
KNN Classifier



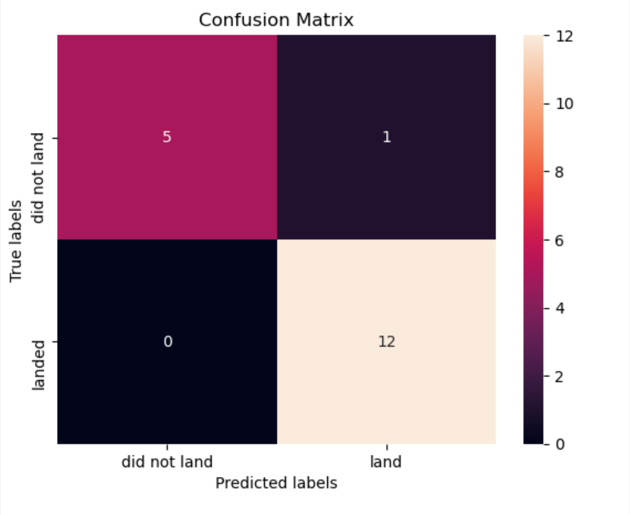
Logistic Regression Classifier



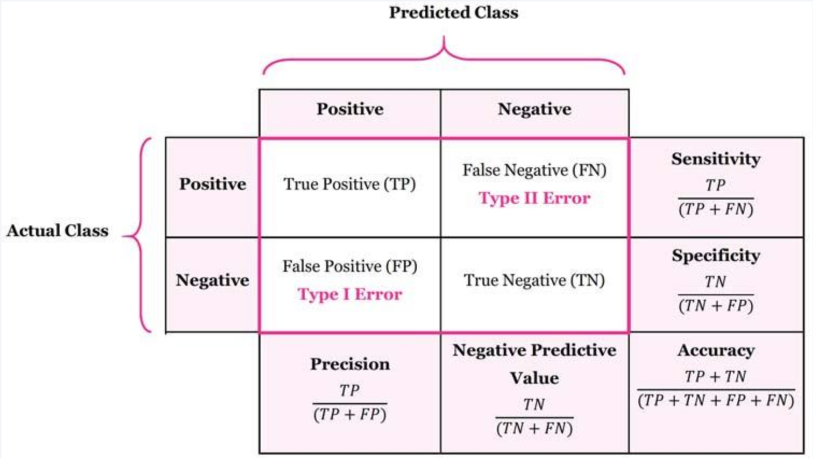
SVM Classifier



Decision Tree Classifier



- Accuracy Confusion Matrices



https://manisha-sirsat.blogspot.com/2019/04/confusion-matrix.html

Conclusions

- The four of models from this report can predict with an 83,3% of accuracy when SpaceX will successfully land the first stage booster.
- The orbits with the highest success rates are GEO, HEO,SSO, ES-L1.
- Heaviy payload mass reports positive success rates.
- As a result, with this model SpaceX can save money with predicted machine learning model.

Appendix

- https://github.com/nuriozbey/CourseraTrainings/blob/main/Applied%20Data%20Science%20Capstone/Machine-Learning-Prediction.ipynb
- https://github.com/nuriozbey/CourseraTrainings/blob/main/Applied%20Data%20Science%20Capstone/Data%20wrangling.ipynb
- https://github.com/nuriozbey/CourseraTrainings/blob/main/Applied%20Data%20Science%20Capstone/best_model_performance.png
- https://github.com/nuriozbey/CourseraTrainings/blob/main/Applied%20Data%20Science%20Capstone/data-collection-api.ipynb
- https://github.com/nuriozbey/CourseraTrainings/blob/main/Applied%20Data%20Science%20Capstone/data-collection-webscraping.ipynb
- https://github.com/nuriozbey/CourseraTrainings/blob/main/Applied%20Data%20Science%20Capstone/exploratory-data-analysis-SQL.ipynb
- https://github.com/nuriozbey/CourseraTrainings/blob/main/Applied%20Data%20Science%20Capstone/exploratory-data-analysis-dataviz.ipynb
- https://github.com/nuriozbey/CourseraTrainings/blob/main/Applied%20Data%20Science%20Capstone/launch_site_location-follium.ipynb
- https://github.com/nuriozbey/CourseraTrainings/blob/main/Applied%20Data%20Science%20Capstone/spacex_dash_app.py

Thank you!