IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Shayne Voris
April 2, 2024

Outline

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

Executive Summary

- SpaceX data was examined, using multiple variables to determine factors that impact a successful launch; and to examine whether or not launch performance has increased.
- In Summary, SpaceX has improved the success rate of launches as both time, and the number of launches has increased. Improvements in booster design, have also resulted in a higher rate of successful launches. It's also clear that some launch sites have had a much higher success rate than others.
- A predictive model created from the data, can predict, with a relatively high rate of confidence whether or not a future SpaceX launch will be successful.

Introduction

- This project is an examination of historical data from past SpaceX Launches.
- Has SpaceX improved launch success rates over time?
- Are some SpaceX Launch Sites more effective at launching successfully?
- Does Payload Mass have a significant impact on the Success Rate?
- Are revisions to booster design leading to more successful launches?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Launch Data was Collected from SpaceX’s API, as well as Web Scraping from the SpaceX Wikipedia page.
- Perform data wrangling
  - Launch Data was processed to remove any Null values. Launches were grouped by Launch Site, and Orbit type.
- 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

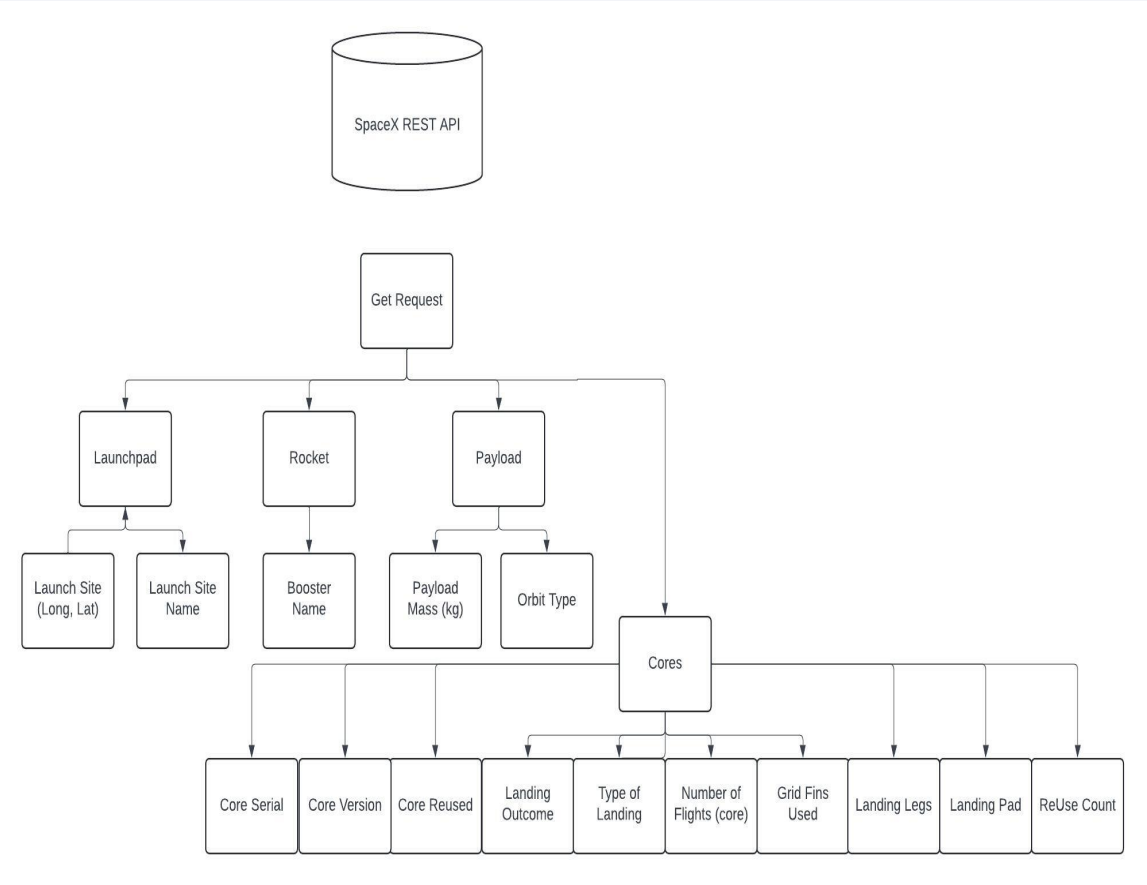
SpaceX Launch Data was collected from SpaceX's own API.

Additional data was Web Scraped from SpaceX's Wikipedia page.

Data Collection – SpaceX API

- Present your data collection with SpaceX REST calls using key phrases and flowcharts

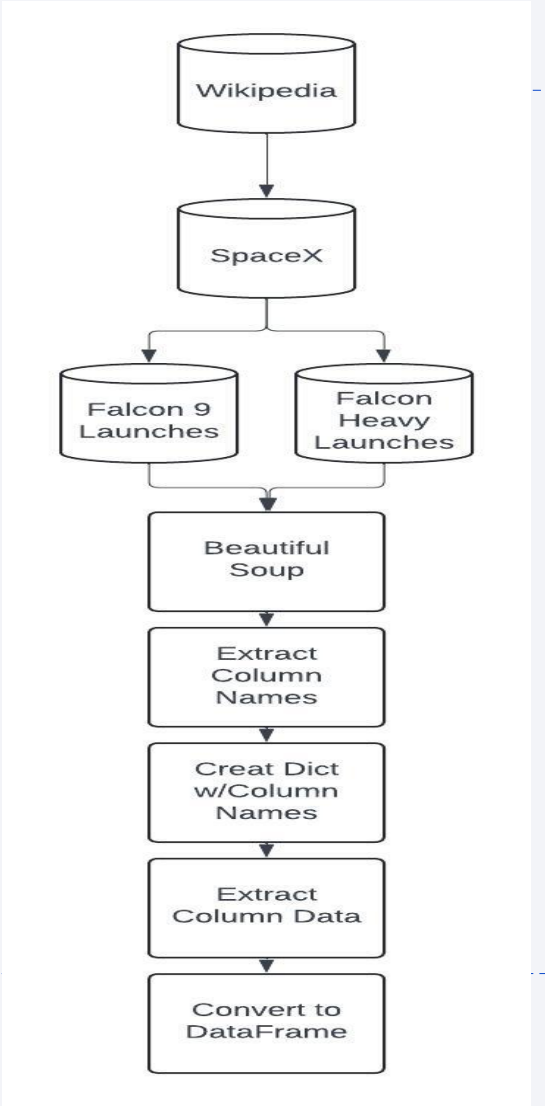
- https://github.com/SmVoris/Data-Science-Capstone/blob/main/jupyter-labs-spacex-data-collection-api_COMPLETED.ipynb



Data Collection - Scraping

- Web Scraping Process

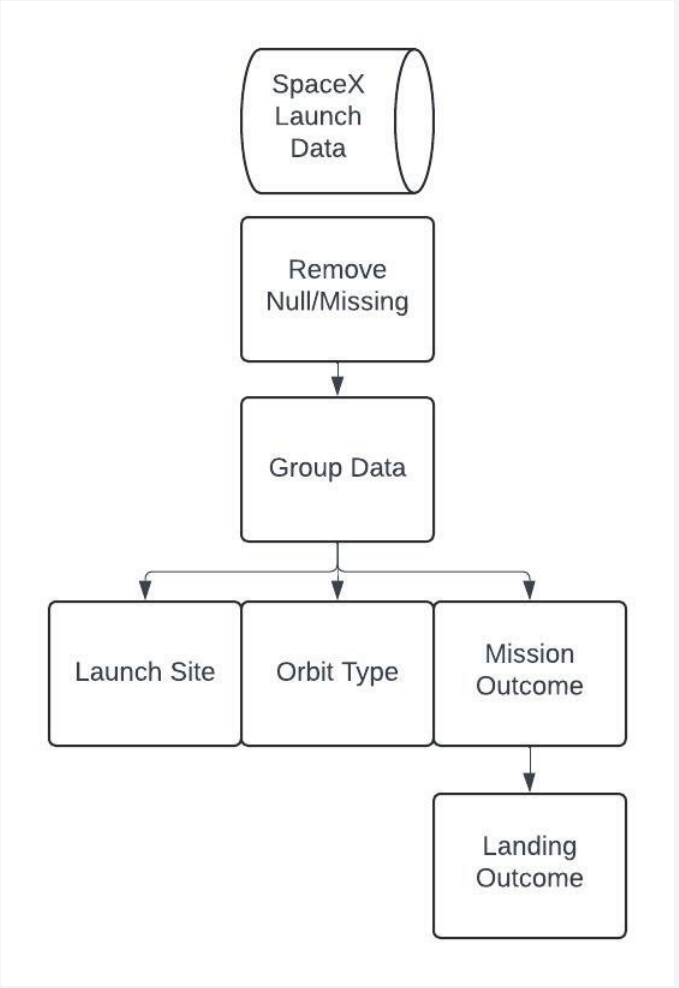
- https://github.com/SmVoris/Data-Science-Capstone/blob/main/jupyter-labs-webscraping_COMPLETED.ipynb



Data Wrangling

Data was processed to remove null values, then grouped to examine the number of launches per site, Orbit Type, and Landing Outcome.

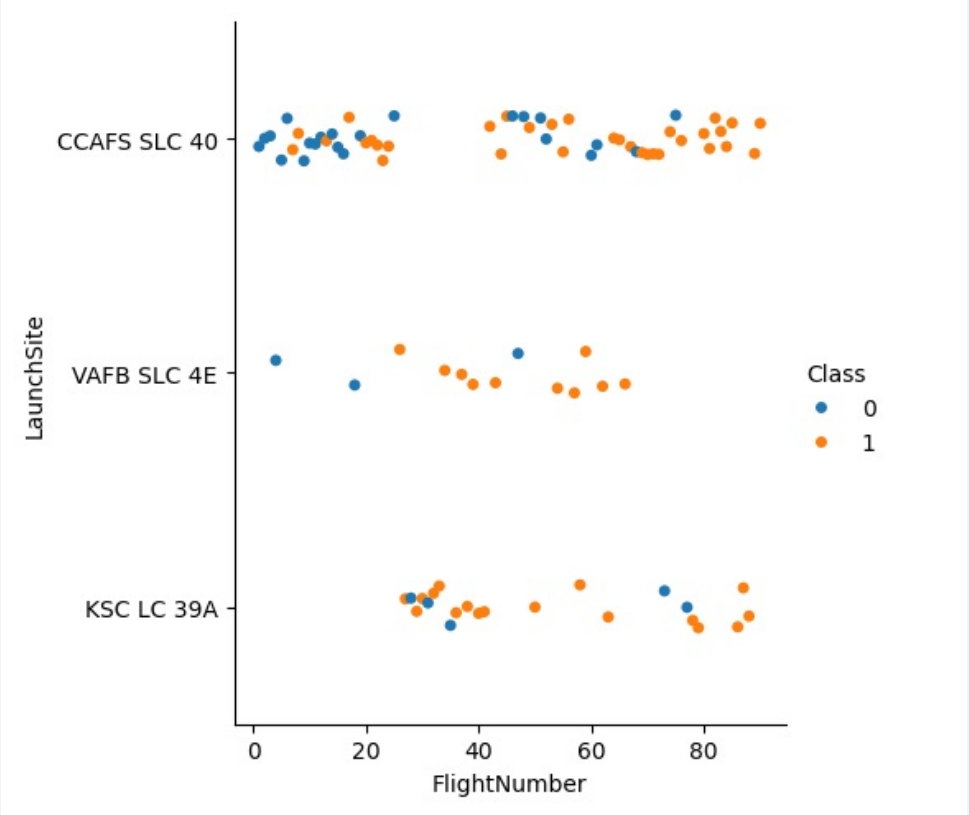
https://github.com/SmVoris/Data-Science-Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb



EDA with Data Visualization

Scatter Plots and Bar Plots were used to examine the relationships between: Flight Number, Payload, Launch Site, Orbit Type, and Success Rate.

These charts show how the above variables affect the success of SpaceX Launches.



EDA with SQL

- Unique Landing Outcomes
- Launch Sites beginning with “CCA”
- Total Payload Mass by NASA (CRS)
- Average Payload Mass for F9 v1.1
- First Successful Ground Pad Landing
- Successful Drone Ship Landing for Payload > 4000kg & <6000kg
- Total # of Successful and Failure Mission Outcomes
- Boosters which have carried Max Payload
- Drone Ship Landing Failures for 2015
- Ranking Landing Outcomes 2010-06-04 to 2017-03-20

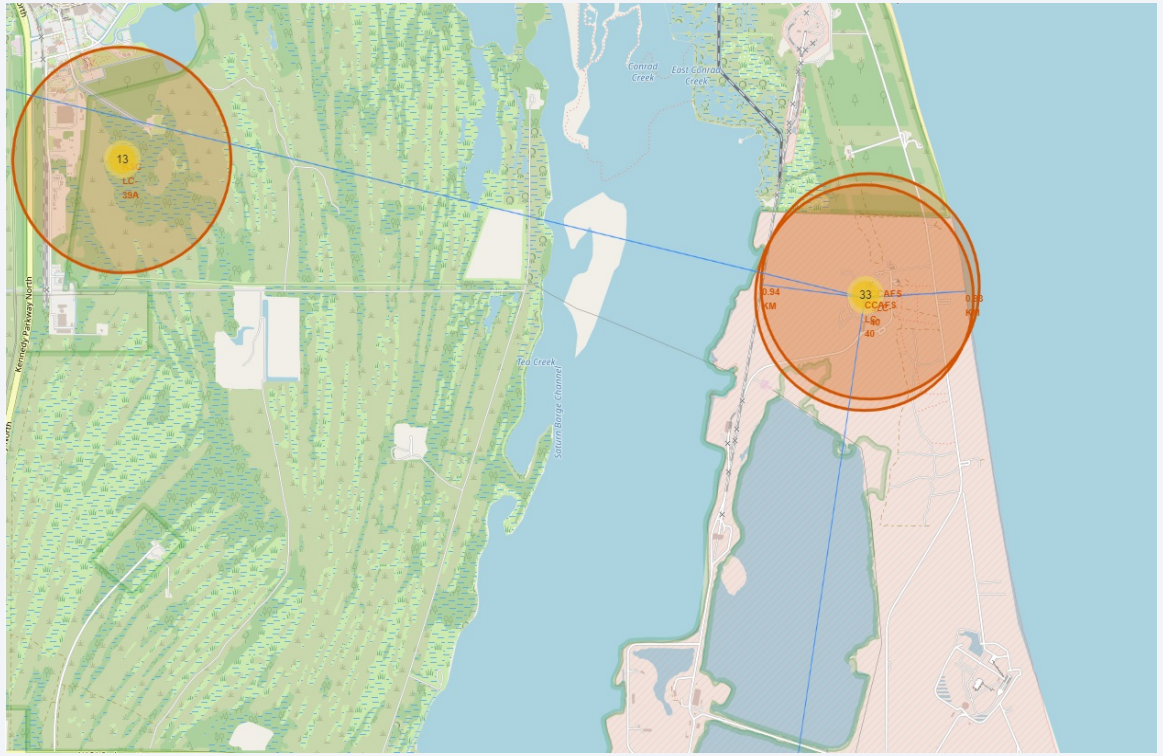
https://github.com/SmVoris/Data-Science-Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

Folium Map was created with Markers to denote each SpaceX Launch Site.

Circles were added to show successful and failed launches at each site.

Lines were added to show the distance of Launch Site from nearby highways, railways, and cities.



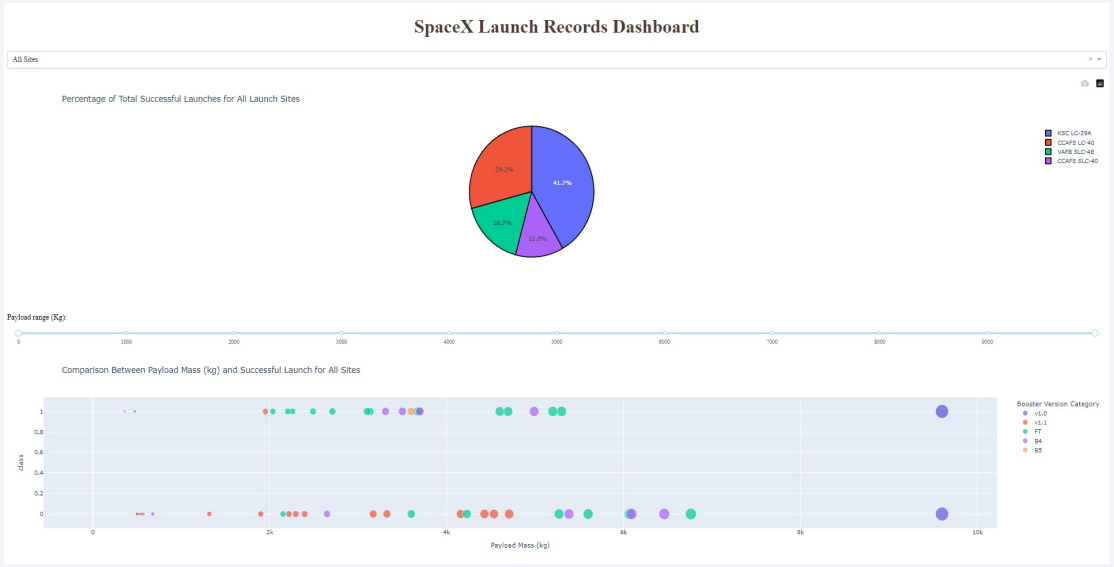
https://github.com/SmVoris/Data-Science-Capstone/blob/main/lab_jupyter_launch_site_location.jupyterlite_COMPLETED.ipynb

Build a Dashboard with Plotly Dash

Pie charts were used to show which Launch Sites had the highest percentage of Successful Launches.

As well as showing Success vs Failure for each individual site.

A Scatter chart was used to show the relationship between Payload Mass and Success vs Failure for all sites, and for each individually selected site.



https://github.com/SmVoris/Data-Science-Capstone/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

Models were constructed using the Cleaned SpaceX Launch Data, to predict the likelihood of a successful launch.

Features were converted to Feature Arrays to train the models, using train_test_split to separate the data into Training and Testing sets.

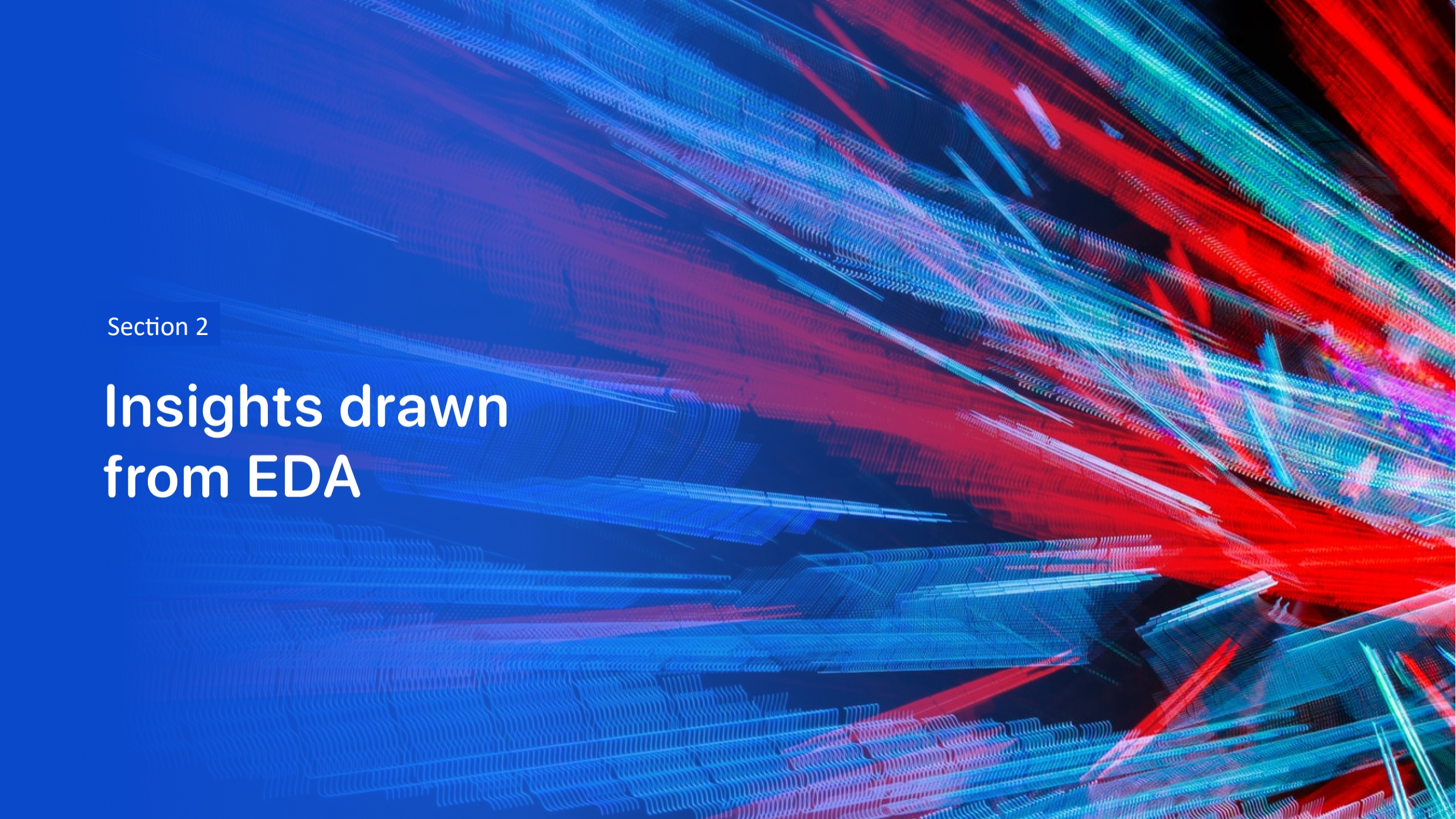
Logistic Regression, Support Vector Machine, Decision Tree, and KNN models were evaluated and compared.

After evaluation of the Accuracy Score, and Confusion Matrix for each model type, it was determined that the Decision Tree Model had the highest accuracy at 87.6%.

https://github.com/SmVoris/Data-Science-Capstone/blob/main/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite%20(1).ipynb

Results

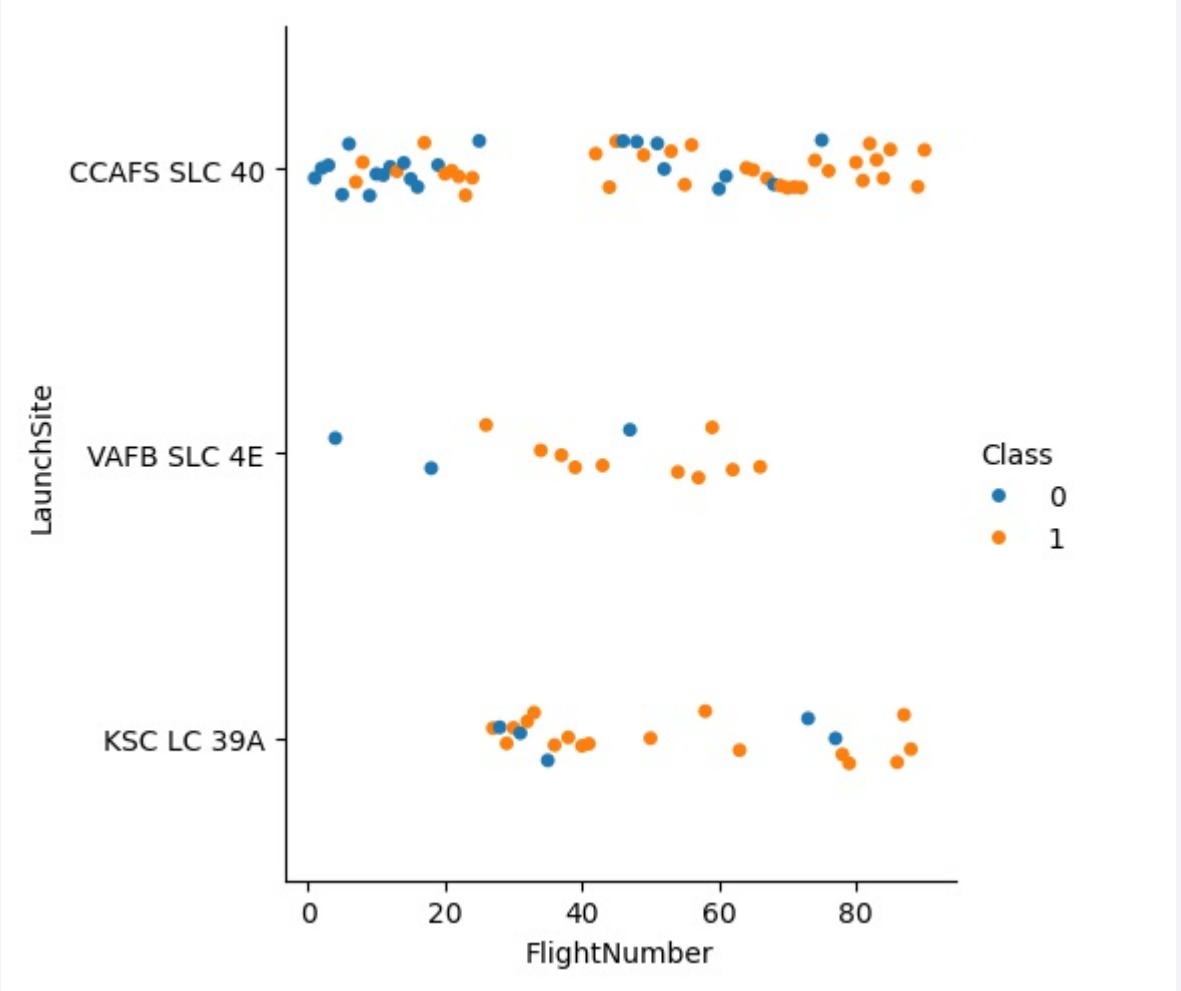
- Exploratory Data Analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results


Section 2
Insights drawn from EDA

Flight Number vs. Launch Site

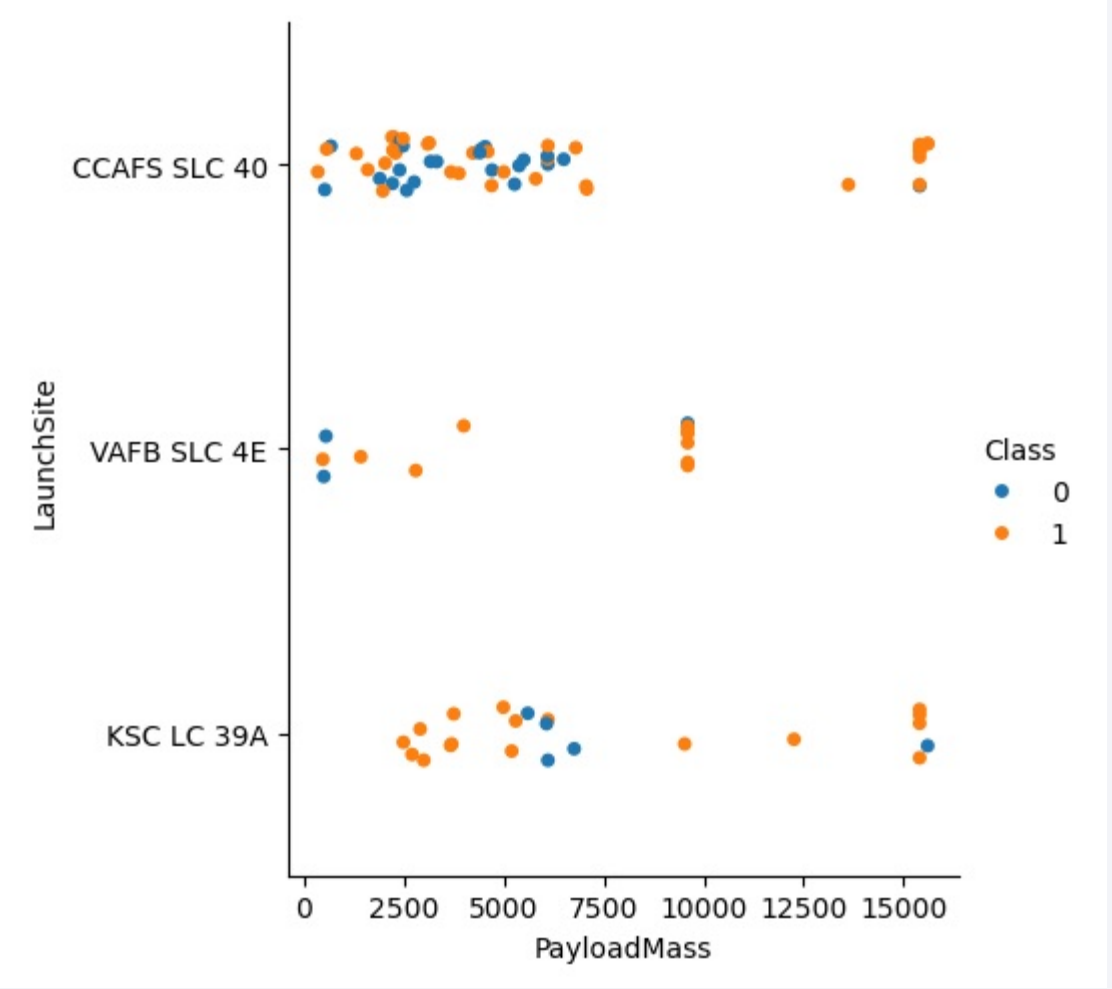
- Show a scatter plot of Flight Number vs. Launch Site

- Class “1” = Success
- Class “0” = Failure



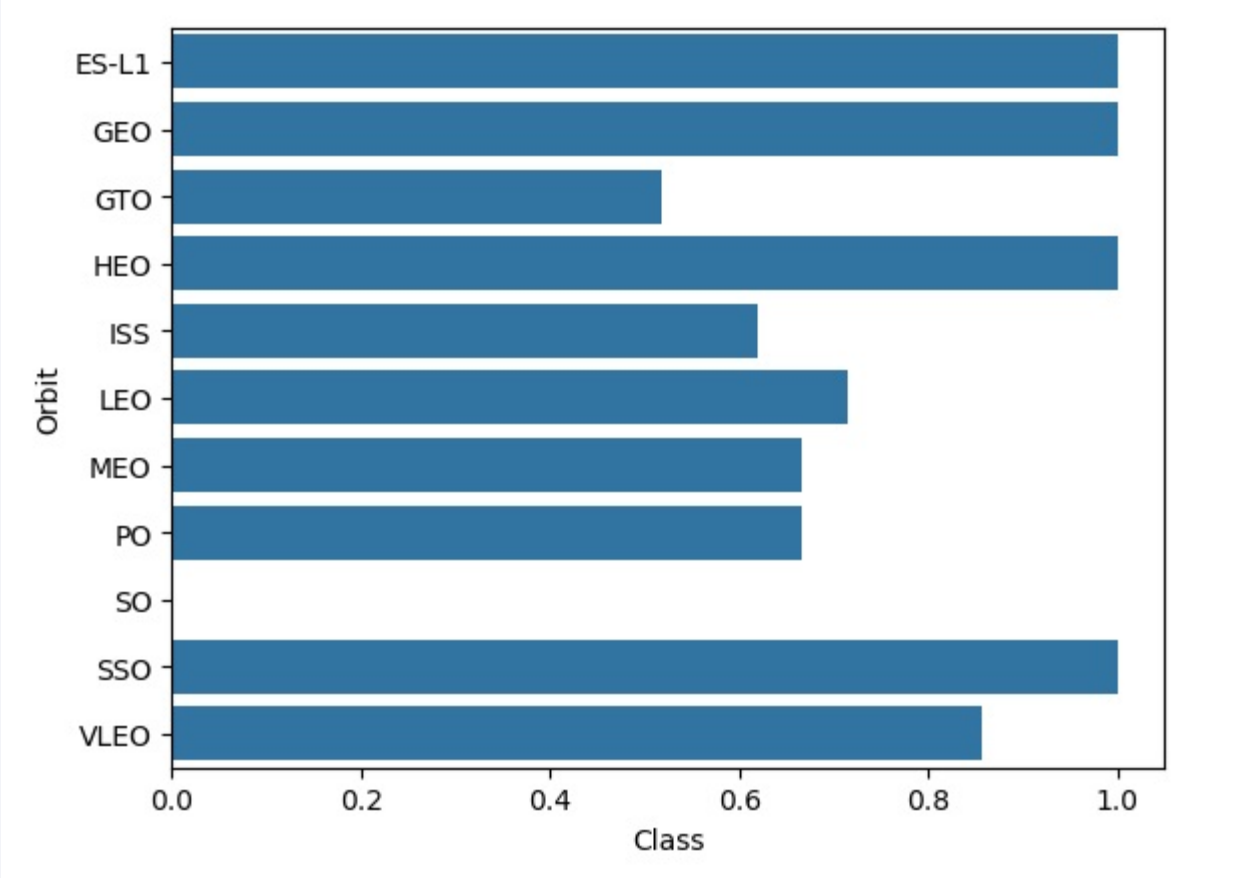
Payload vs. Launch Site

- Scatter plot of Payload vs. Launch Site
- Launches with 7,500kg + Payload Mass have a higher Success Rate at all Sites.



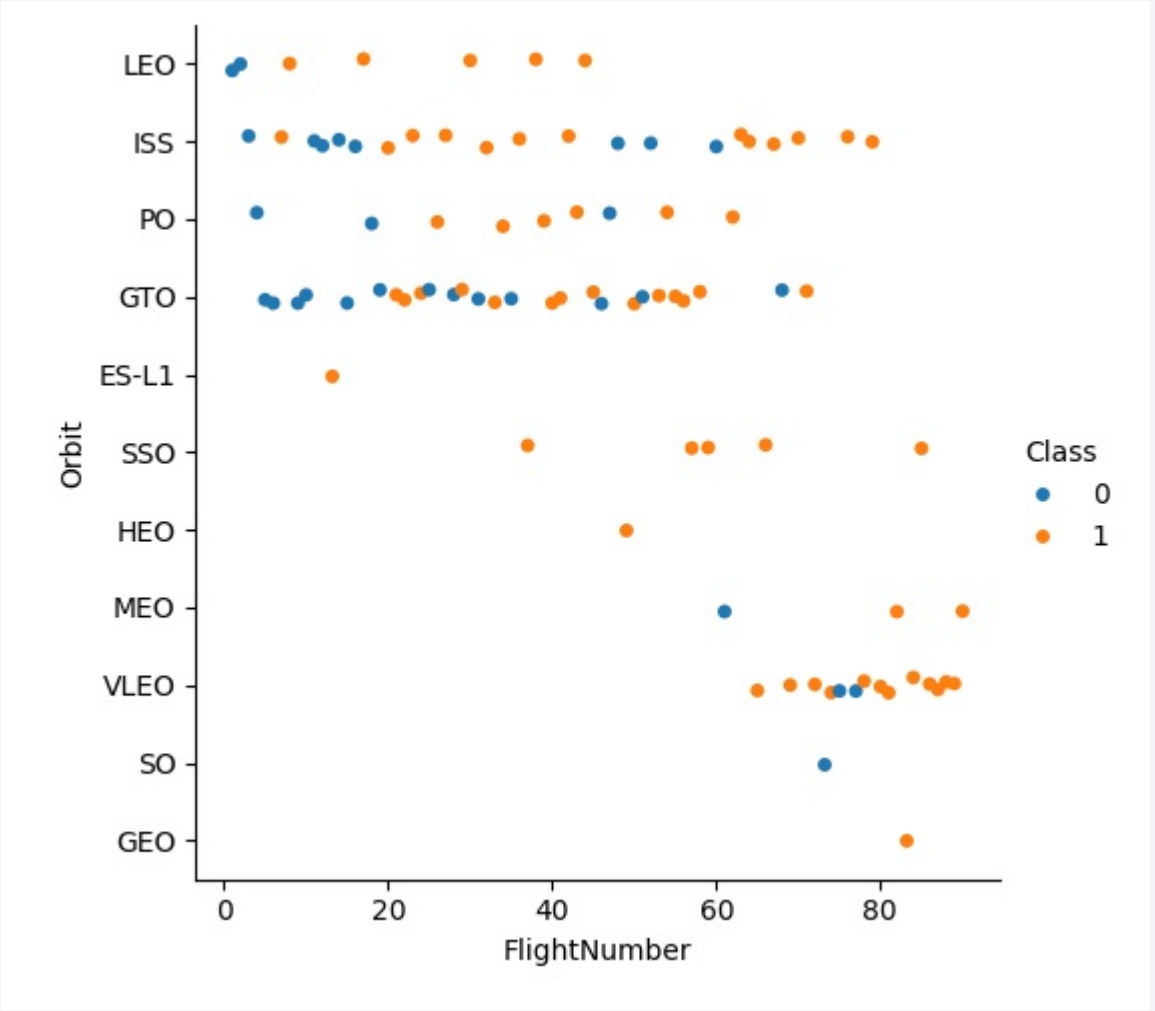
Success Rate vs. Orbit Type

- Bar chart for the success rate of each orbit type
- ES-L1, GEO, HEO, and SSO Orbits have a 100% Success Rate



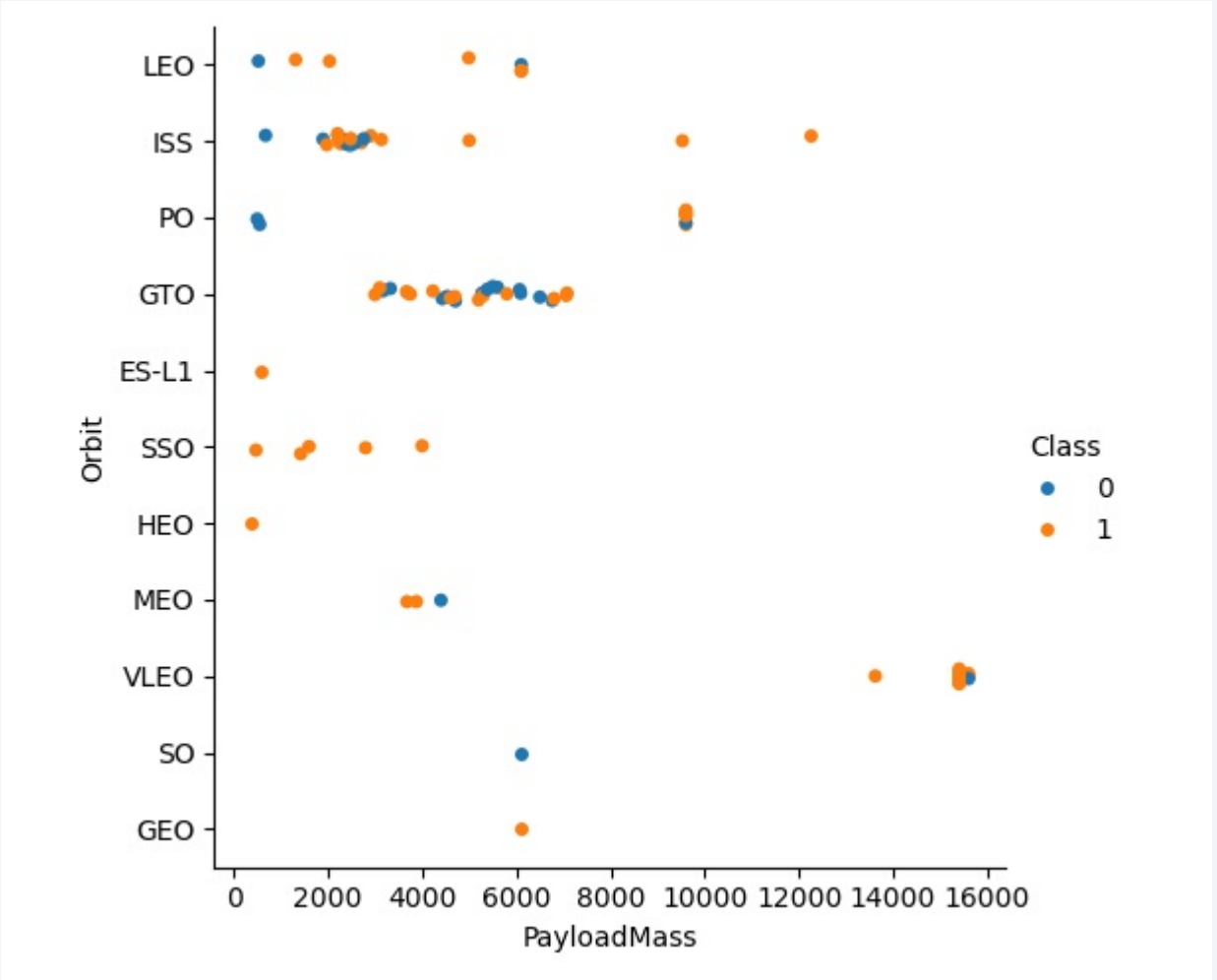
Flight Number vs. Orbit Type

- Scatter point of Flight number vs. Orbit type
- Success Rate for most Orbits has increased with more Flights.



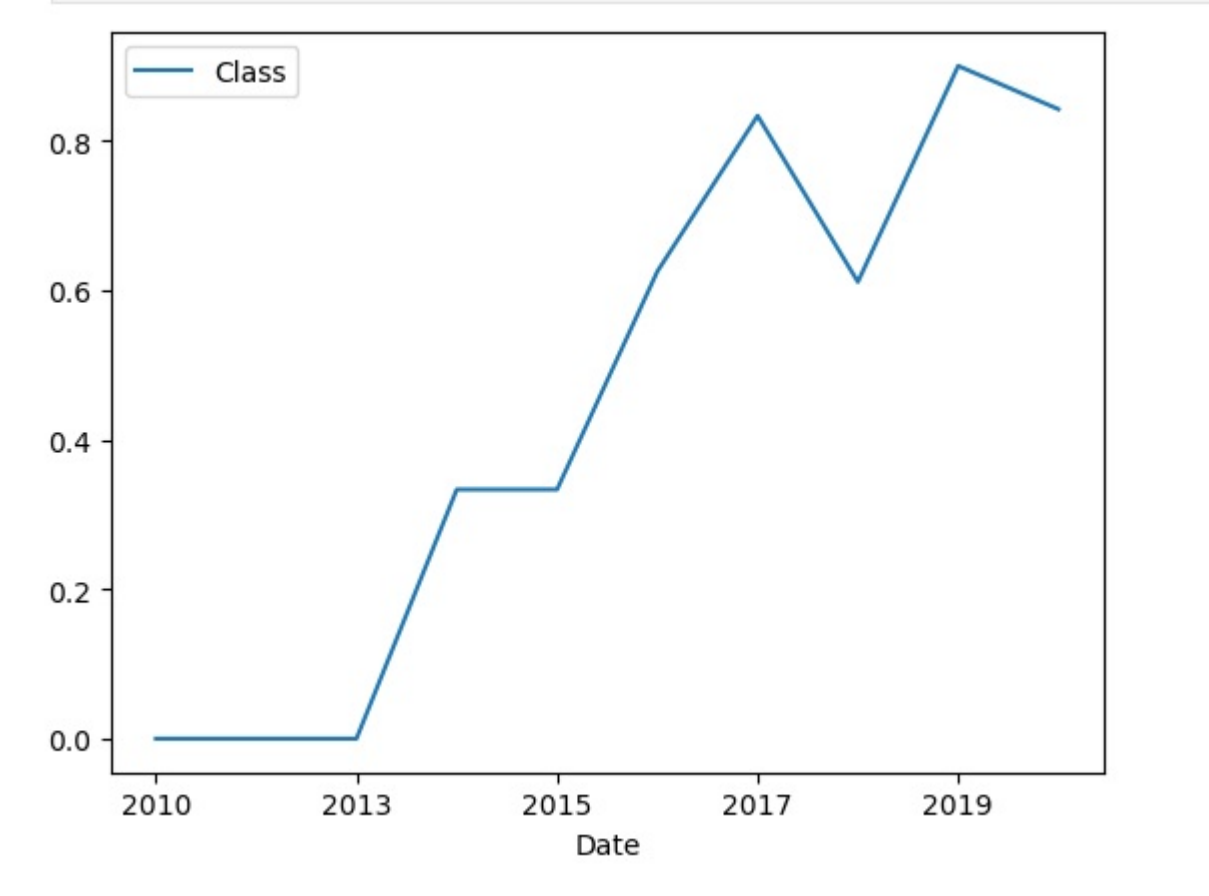
Payload vs. Orbit Type

- Show a scatter point of payload vs. orbit type
- Payloads greater than 8,000kg have only been attempted for ISS, PO, and VLEO Orbits.



Launch Success Yearly Trend

- SpaceX Launch Success Rate by Year



All Launch Site Names

List of 4 Unique SpaceX Launch Sites

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

Launch Site Names Begin with 'CCA'

5 Records from Launch Sites beginning with “CCA”.

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

Total Payload Mass

Total Payload Mass (kg) for Customer: NASA (CRS)

Total Mass	Customer
45596.0	NASA (CRS)

Average Payload Mass by F9 v1.1

Average Payload Mass (kg) carried by Booster Version F9 v1.1

Average Payload Mass	Booster Version
2534.6666666666665	F9 v1.1 B1003

First Successful Ground Landing Date

First Successful Ground Pad Landing Date

First Successful Ground Pad Landing Date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

Falcon 9 Boosters that have achieved a Successful Drone Ship Landing, with Payload Mass greater than 4000kg, and less than 6000kg.

Booster Version	Payload Mass	Landing Outcome
F9 FT B1022	4696	Success (drone ship)
F9 FT B1026	4600	Success (drone ship)
F9 FT B1021.2	5300	Success (drone ship)
F9 FT B1031.2	5200	Success (drone ship)

Total Number of Successful and Failure Mission Outcomes

Total Number of Successful and Failure Mission Outcomes.

Mission Outcome	Count
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

Booster which have carried the Maximum Payload Mass of 15,600kg.

Booster Version	Payload Mass (kg)
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

2015 Launch Records

Launches in 2015 that Failed to land on Drone Ship.

Month	Year	Booster Version	Launch Site	Landing Outcome
01	2015	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	2015	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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

Rank of the Number of Landing Outcomes by Landing Outcome Type, between 2010-06-04 and 2017-03-20.

Rank	Count	Landing Outcome
1	10	No attempt
2	5	Success (drone ship)
2	5	Failure (drone ship)
4	3	Success (ground pad)
4	3	Controlled (ocean)
6	2	Uncontrolled (ocean)
6	2	Failure (parachute)
8	1	Precluded (drone ship)

Section 3

Launch Sites Proximities Analysis

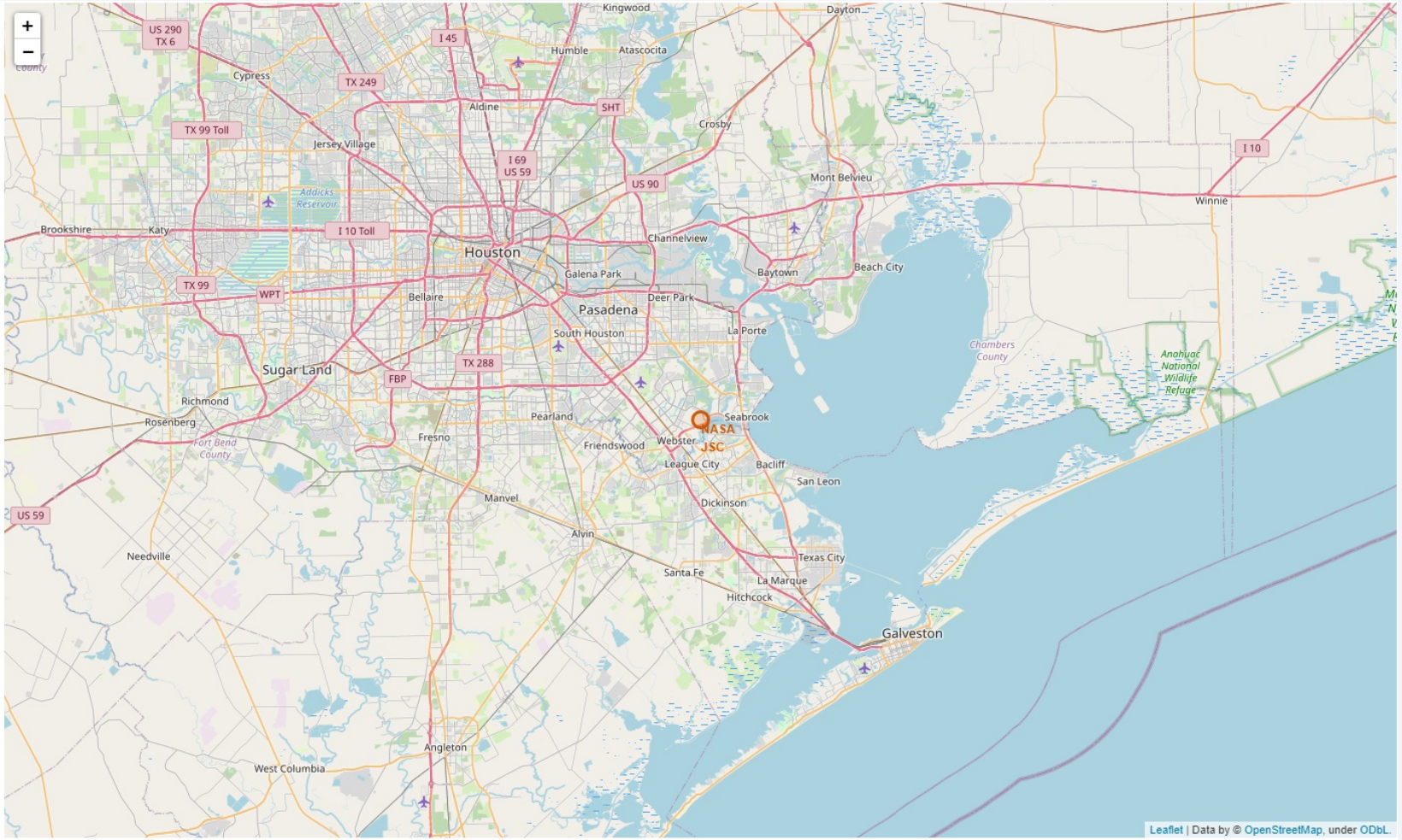
Folium Map – NASA Johnson Space Center

Johnson Space Center

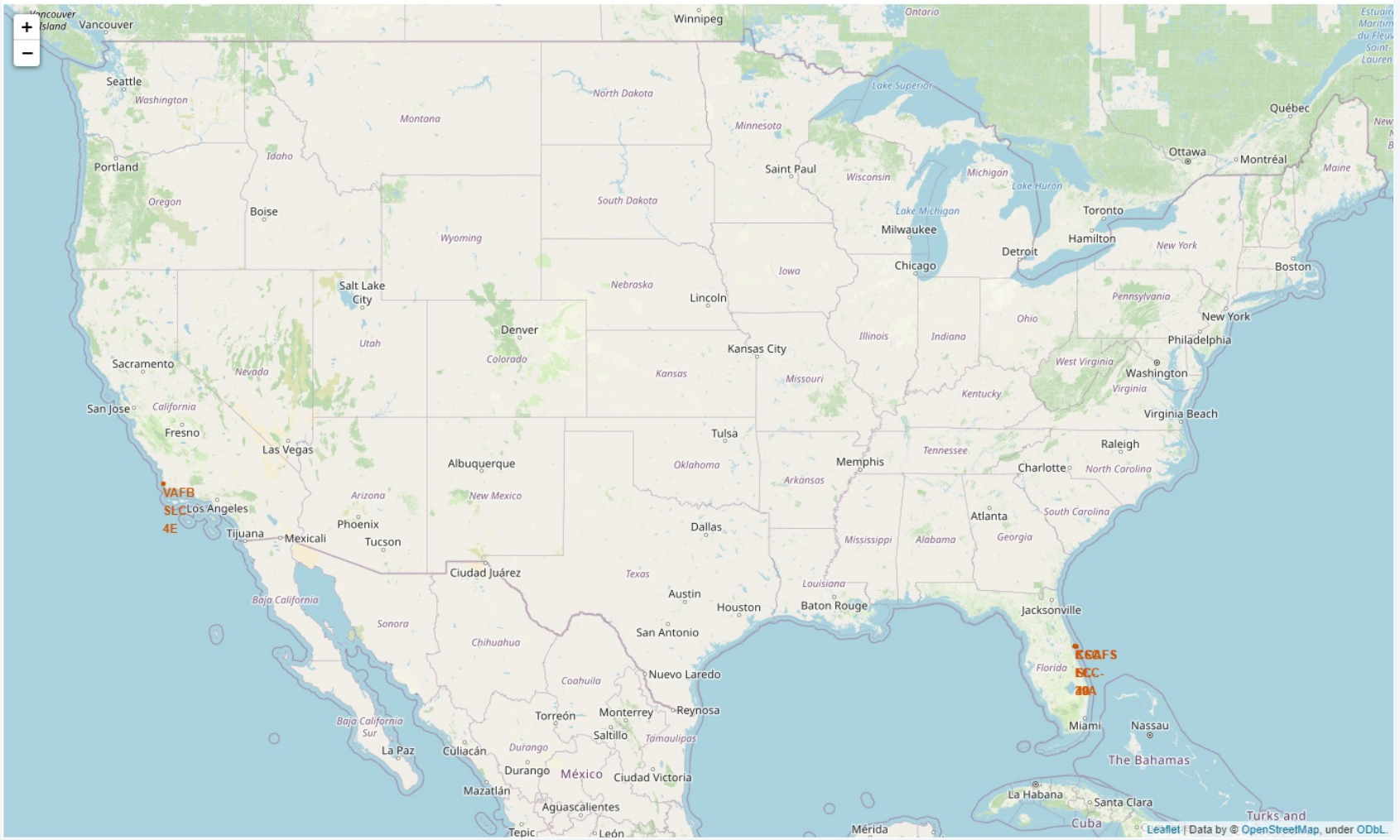
Houston, TX

Lat: 29.559684888503615

Long: -95.0830971930759

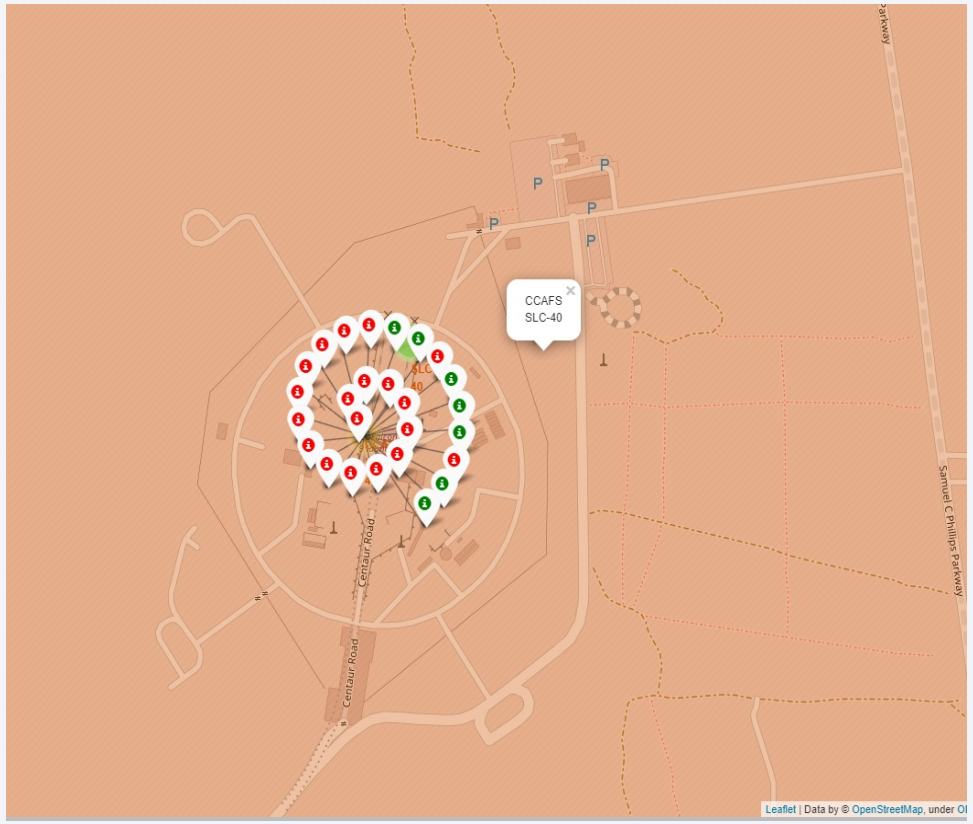


SpaceX Launch Sites

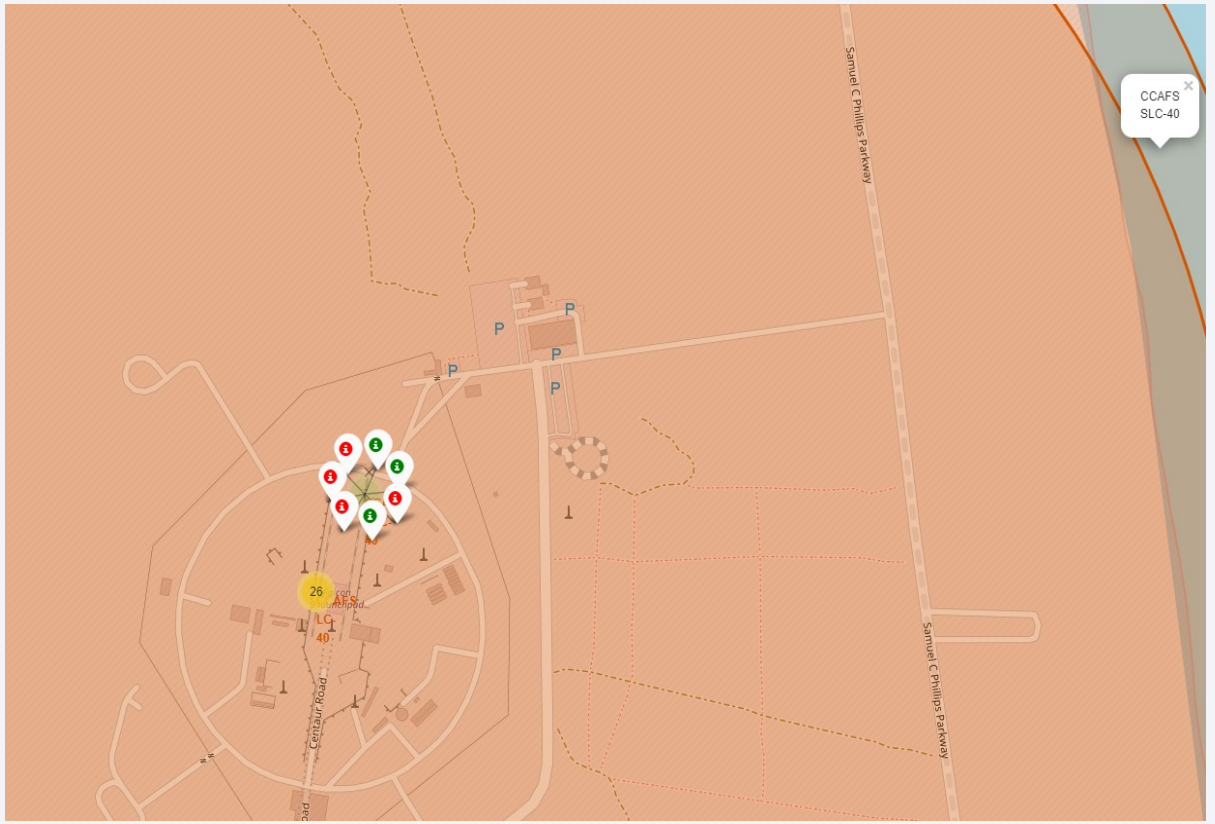


SpaceX Launch Sites

CCAFS-LC-40

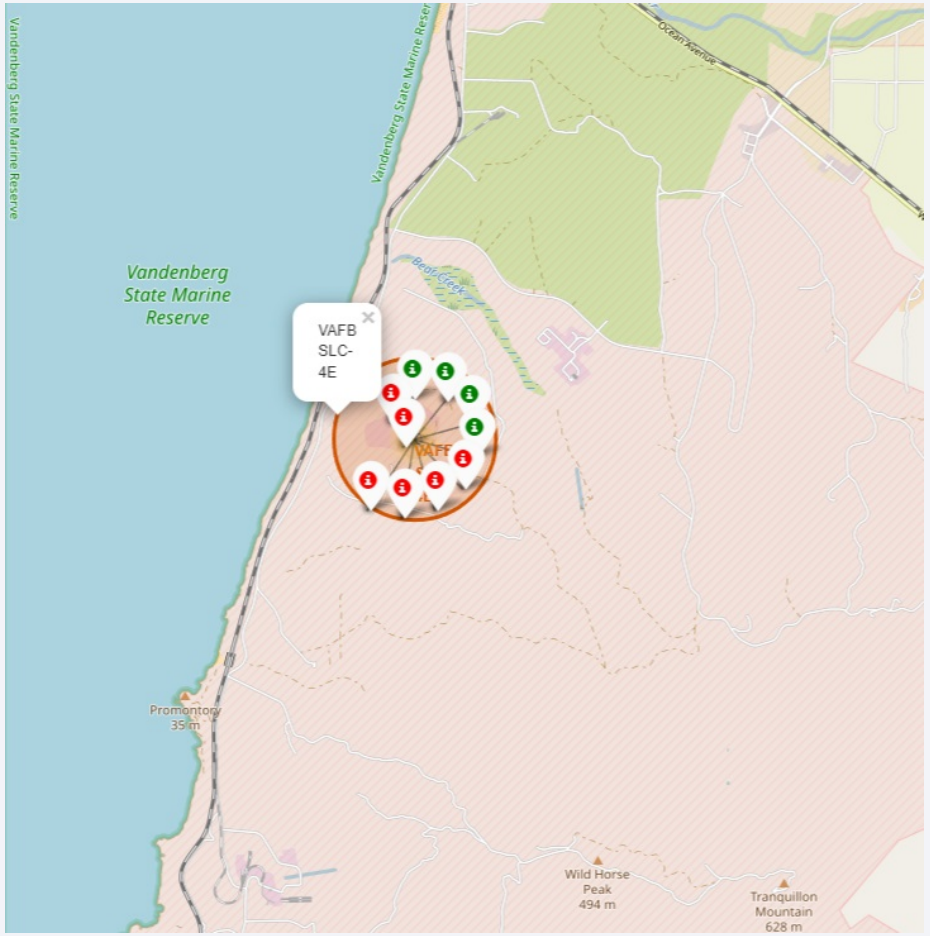


CCAFS-SLC-40

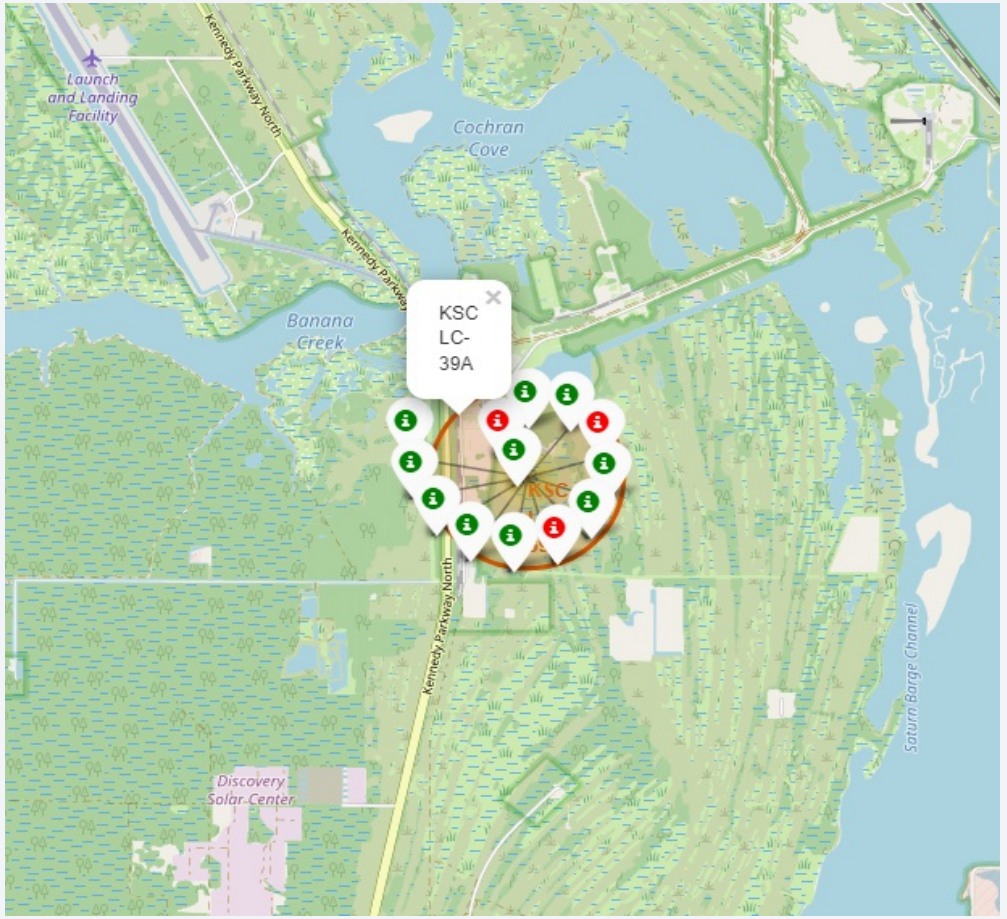


SpaceX Launch Sites

VAFB-SLC-4E



KSC-LC-39A



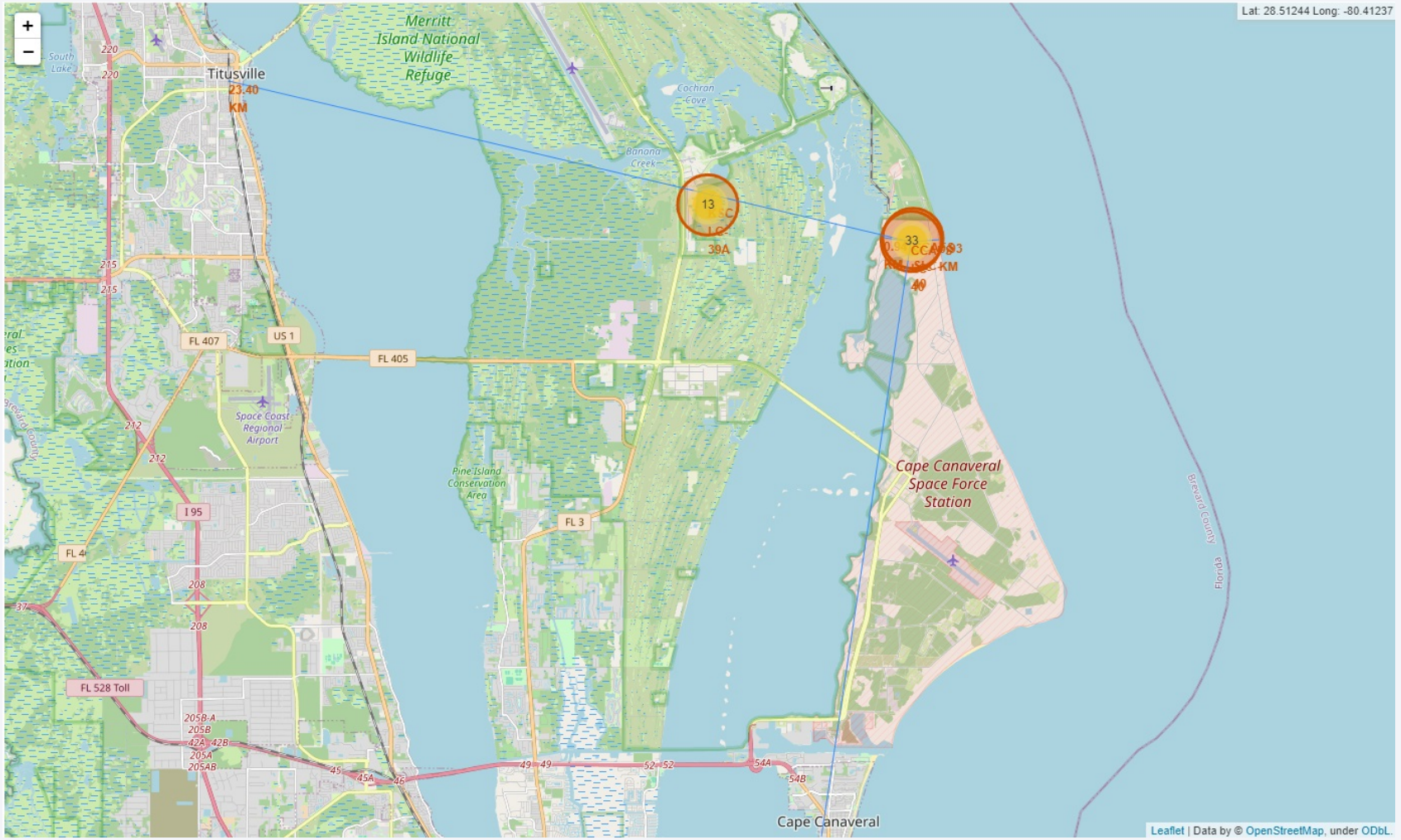
Launch Site Proximity to Cities

Titusville:

23.4km

Cape Canaveral:

20.11km



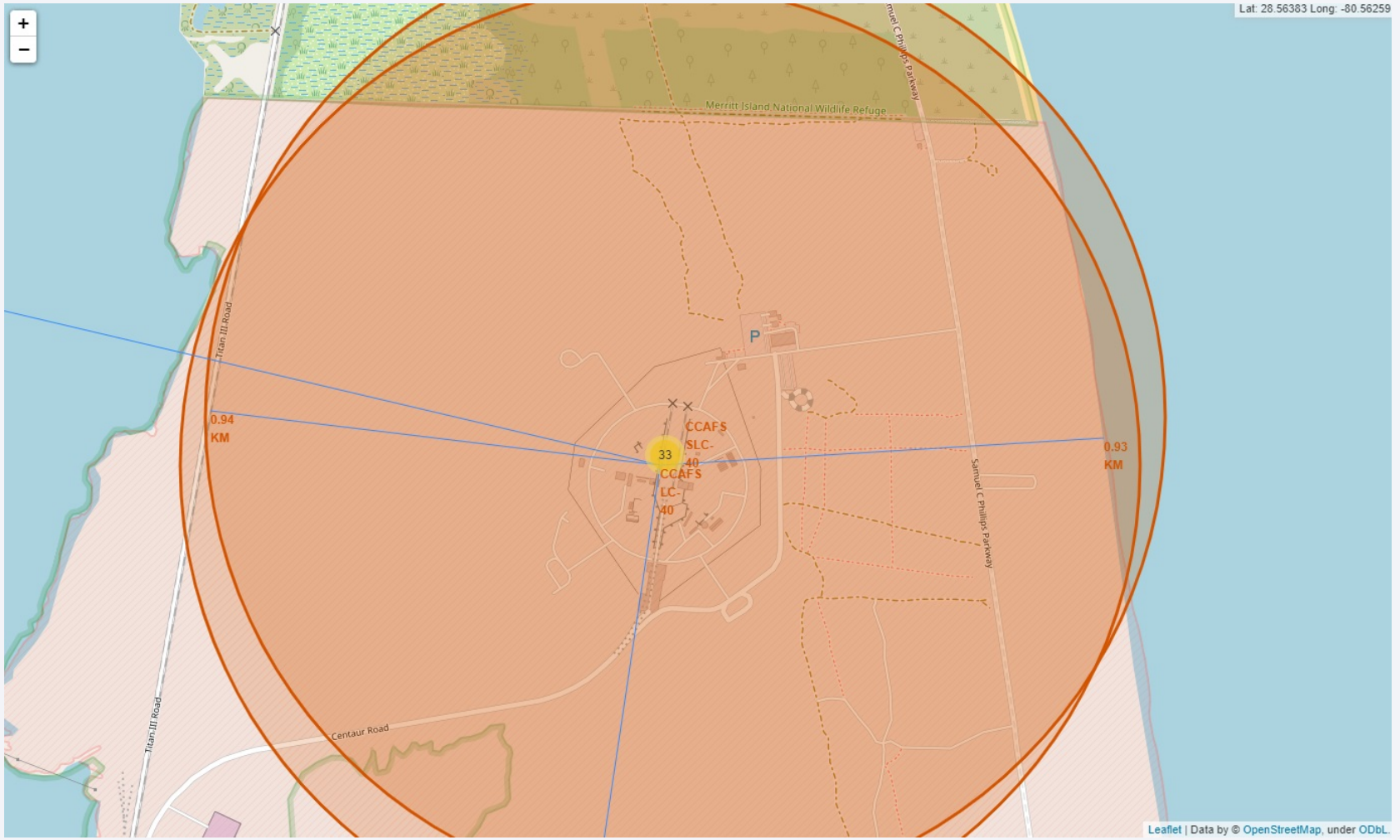
Launch Site Proximity to Railway and Coast

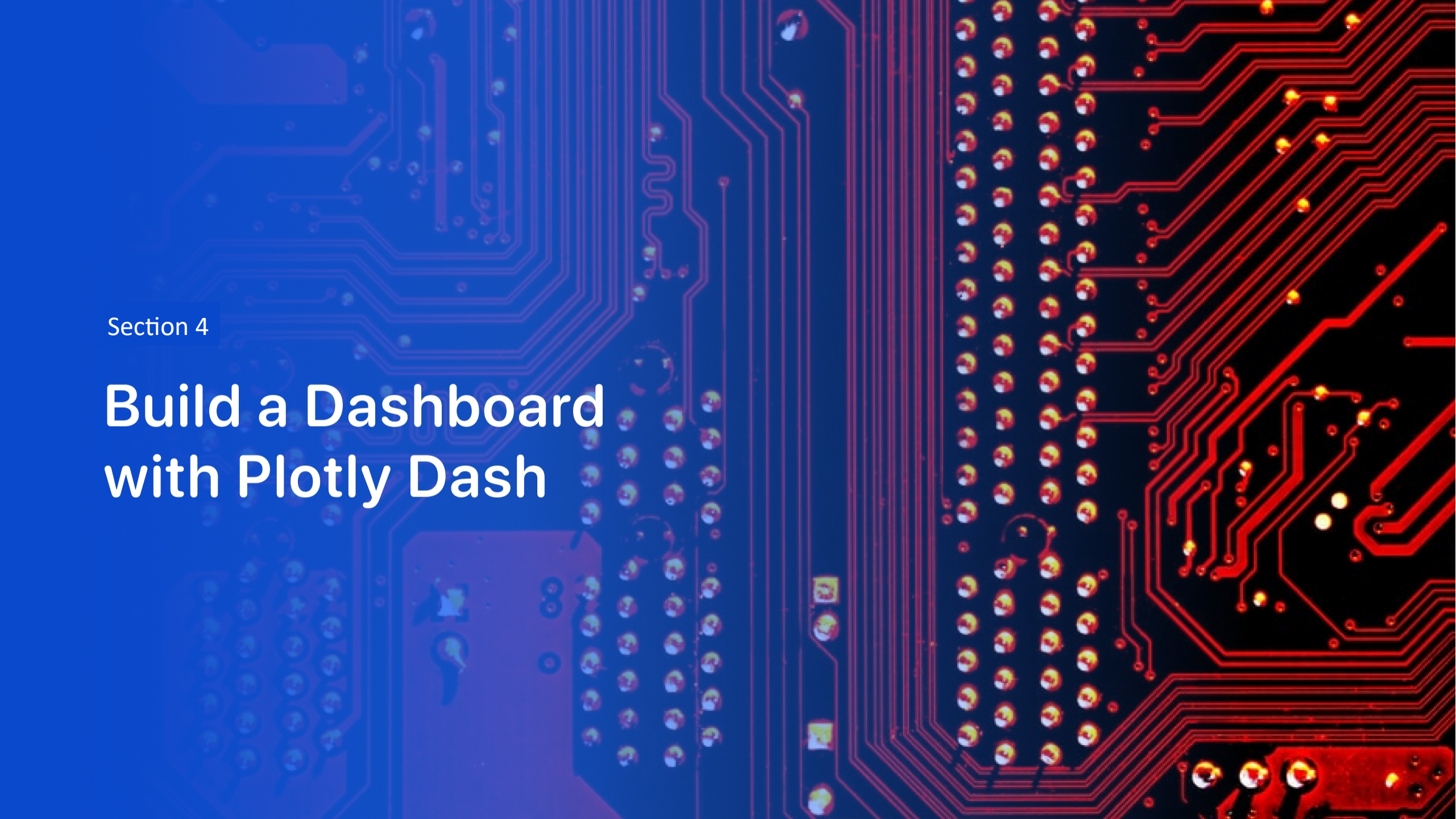
Rail:

0.94km

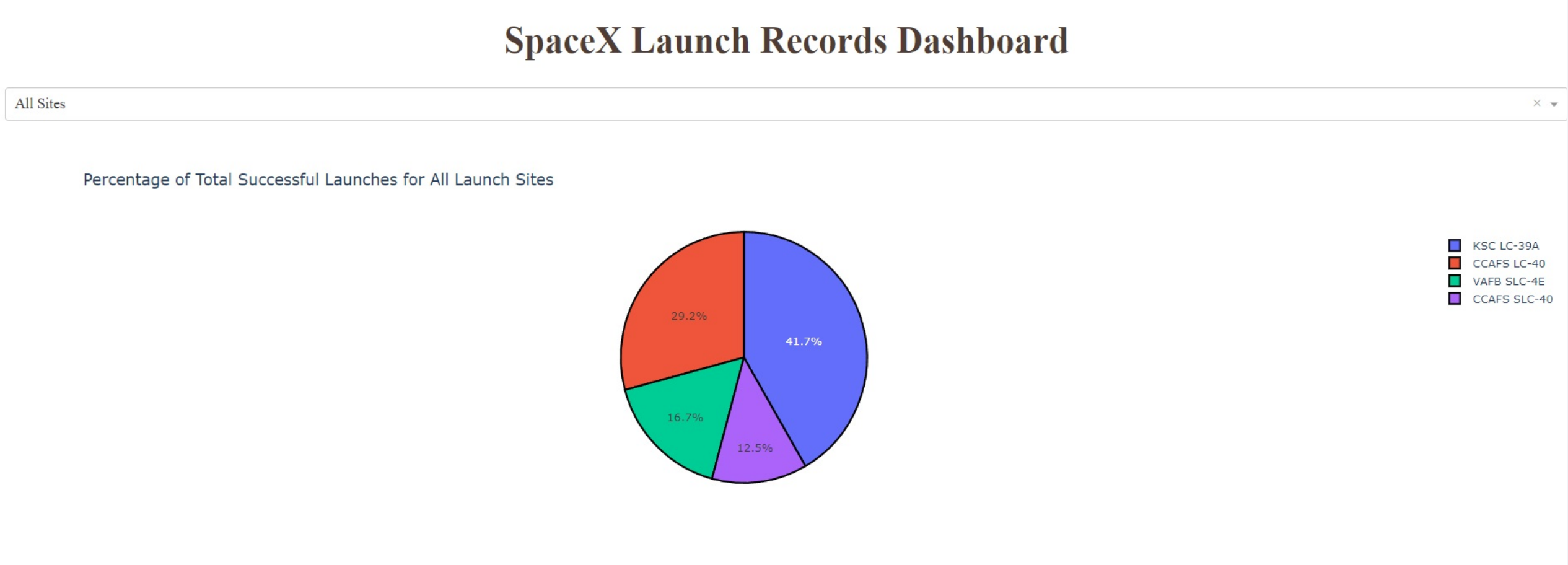
Coastline:

0.93km



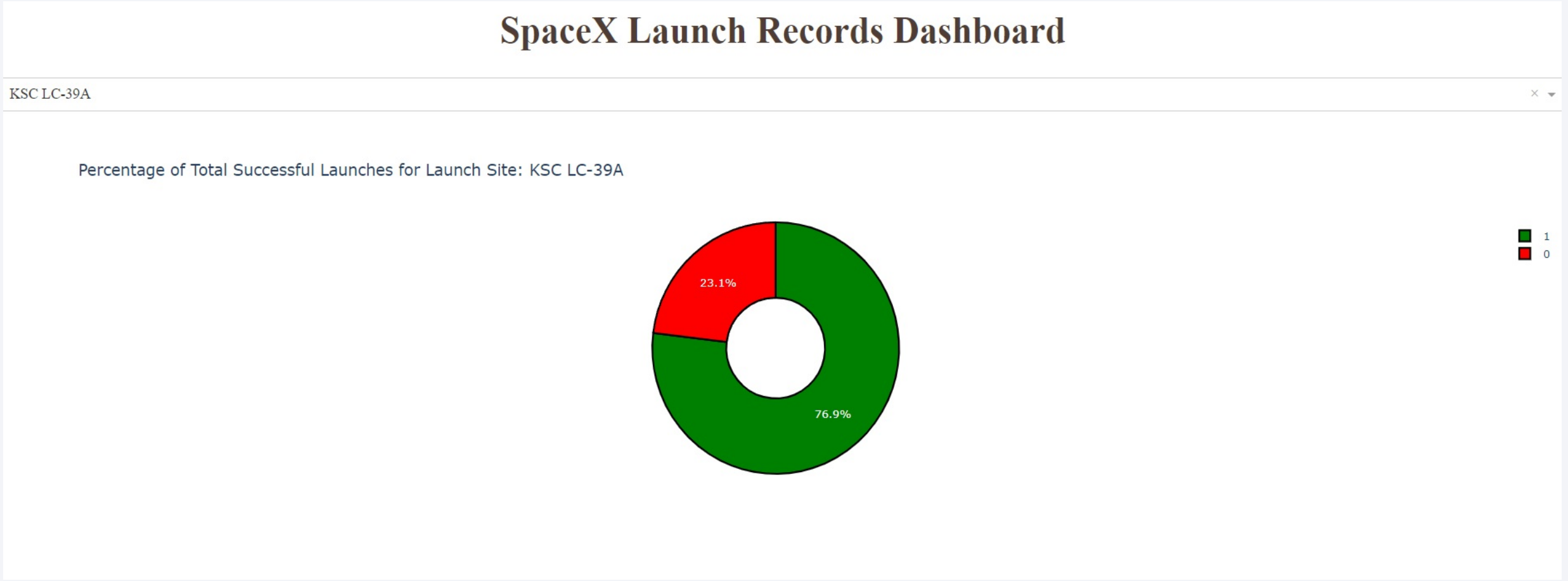

Section 4
Build a Dashboard with Plotly Dash

Dashboard – Success Count – All Sites



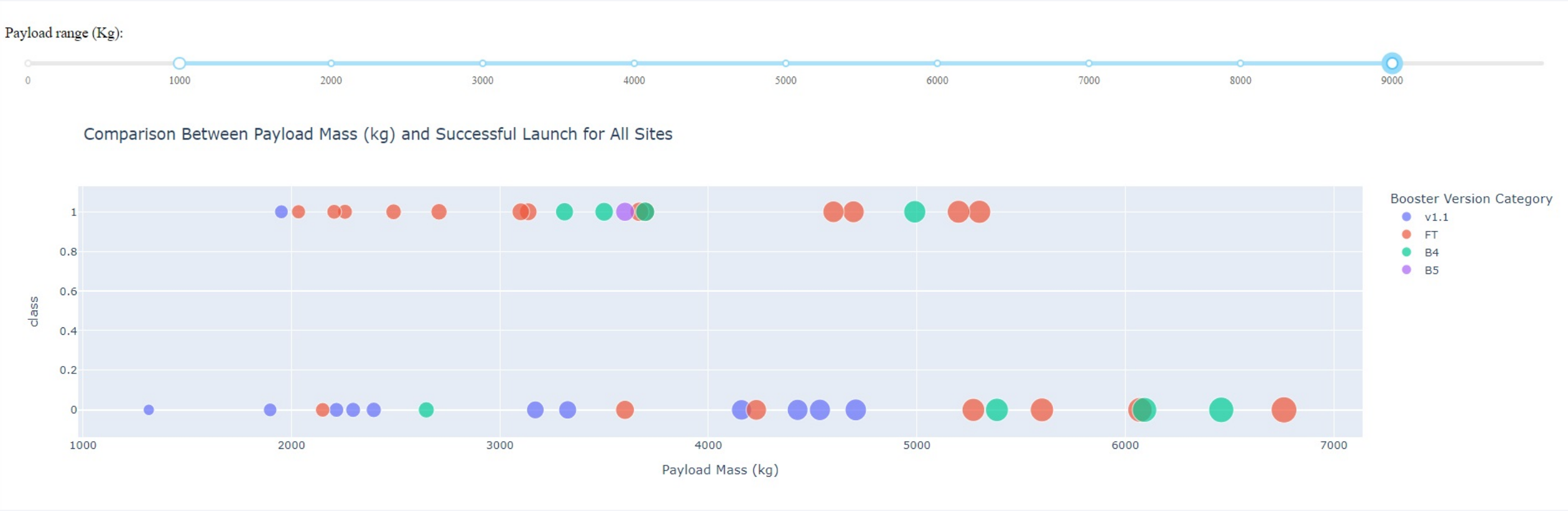
KSC LC-39A is the Most Successful Launch Site, With **41.7%** of All Successful Launches.

KSC LC-39A Success Rate



76.9% of Launches from KSC LC-39A Have been Successful.

Payload Range for All Sites



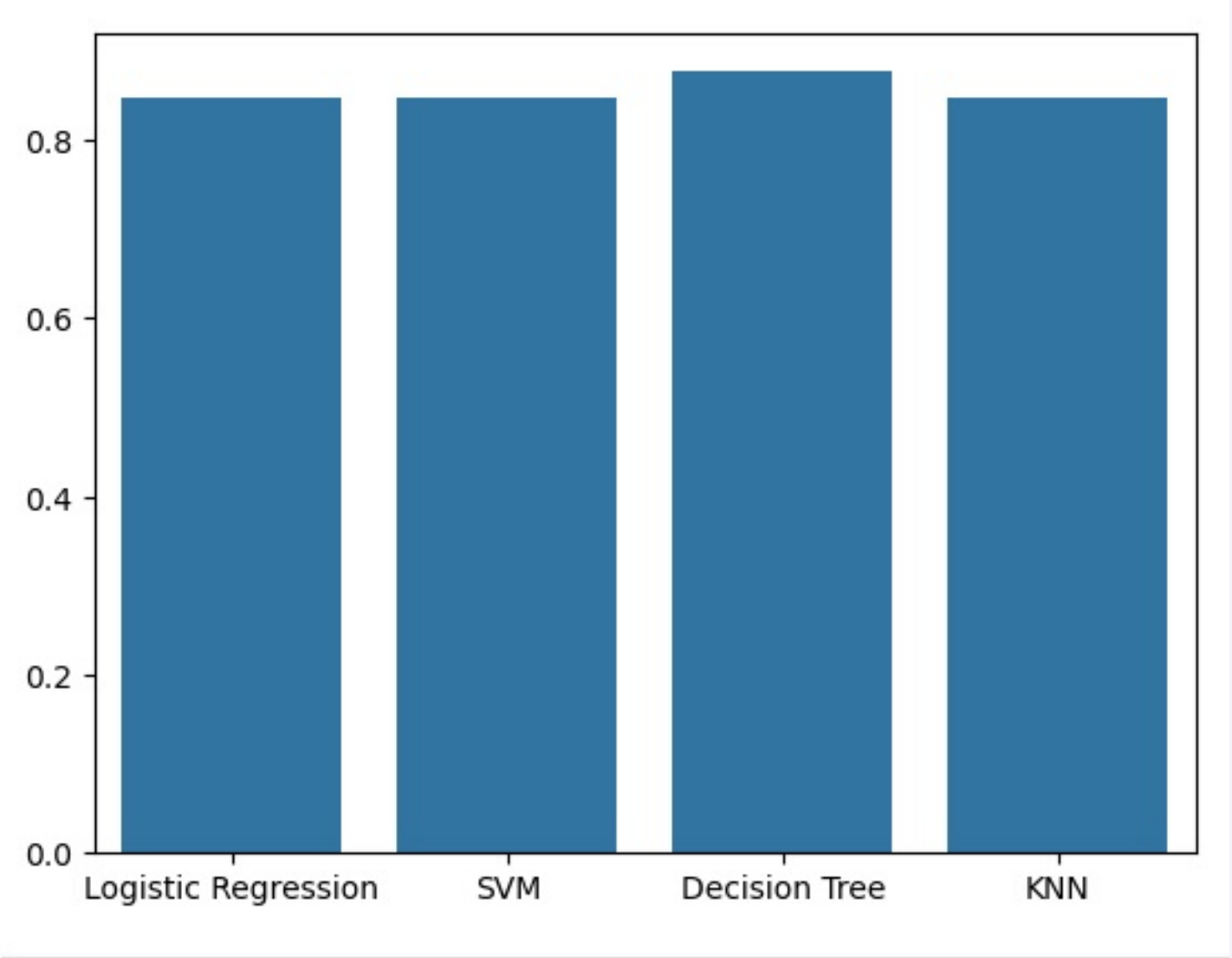
- Payload Range from **2000-5000kg** shows the greatest Success Rate
- Booster v1.1 has the greatest Failure Rate
- Booster B5 has 100% Success Rate, but only 1 recorded Launch
- Booster **FT** has the greatest Success Rate for Multiple Launches

Payload Range 2000 – 5000 kg

Section 5

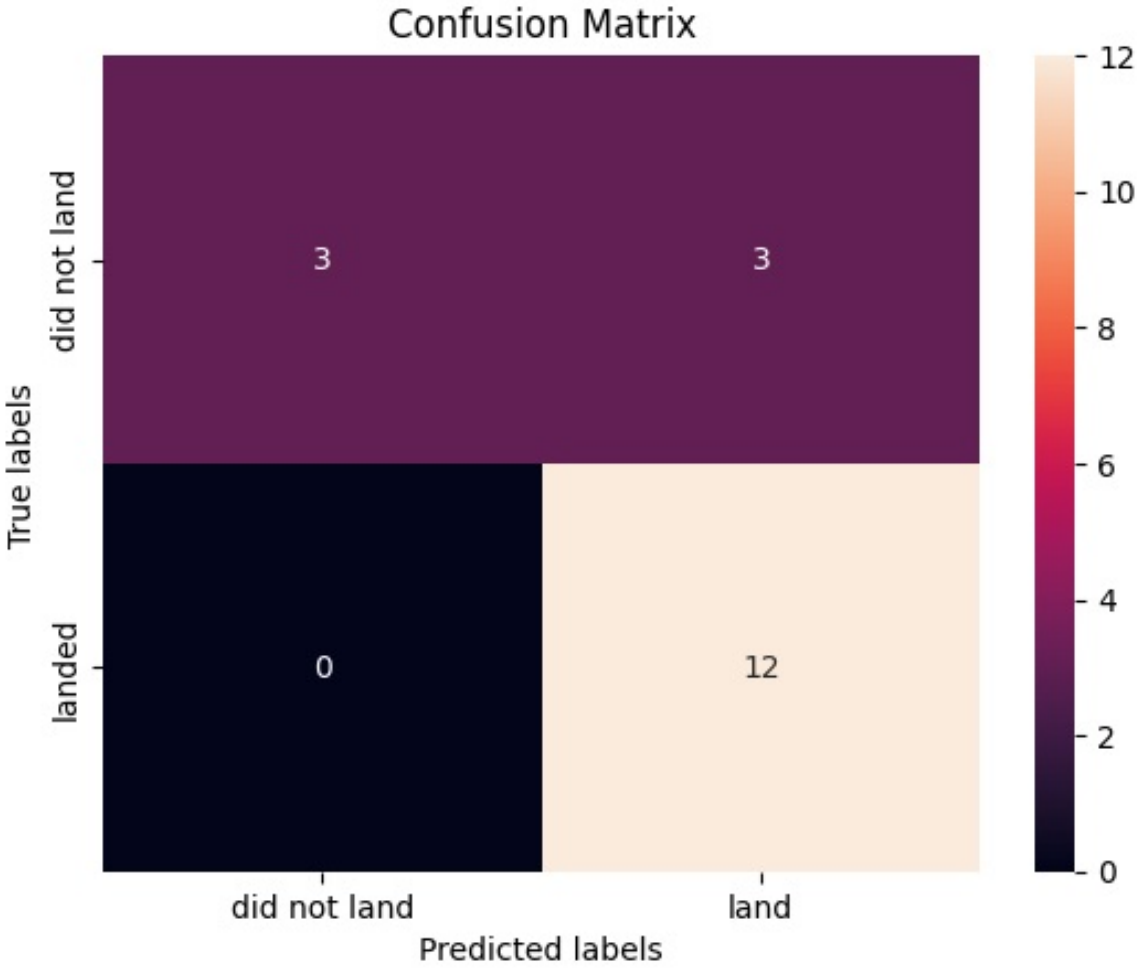
Predictive Analysis (Classification)

Classification Accuracy



Decision Tree Model has highest classification accuracy: 87.6%

Confusion Matrix



Decision Tree Model Confusion Matrix

Shows 80% Accuracy at predicting Landings. Predicting 12 out of 15 correctly.

Shows 100% accuracy at predicting Failure to Land. Predicting 3 out of 3 correctly.

Conclusions

- SpaceX has consistently improved performance, and Success Rates as the number of Flights has increased.
- KSC LC-39A: (Kennedy Space Center Launch Complex 39A) Has been the most Successful Launch Site with 76.9% Success Rate.
- Booster version FT has been the most successful booster version.
- Payloads between 2000 kg and 5000 kg, and above 7,500 kg have the greatest Success Rates.

Appendix

Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!