IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Marianela Guerra
September 19th, 2022

Outline

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

Executive Summary

- Data collection methodology:

SpaceX Launch data is gathered from the SpaceX REST API. The API provides launch data such as rocket used, payload delivered, launch specifications, landing specifications and landing outcome. JSON objects represent each launch. Python BeautifulSoup package was also used to scrape HTML tables with Falcon 9 launch records.

- Perform data wrangling

Missing values in the dataset existed and needed to be fixed before proceeding. In order to do this, the mean for the column with missing values (PayloadMass) was calculated and used to replace N/A values using the .replace() function. Falcon 1 launches were also removed as only Falcon 9 launches were of interest in this project.

- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

Introduction

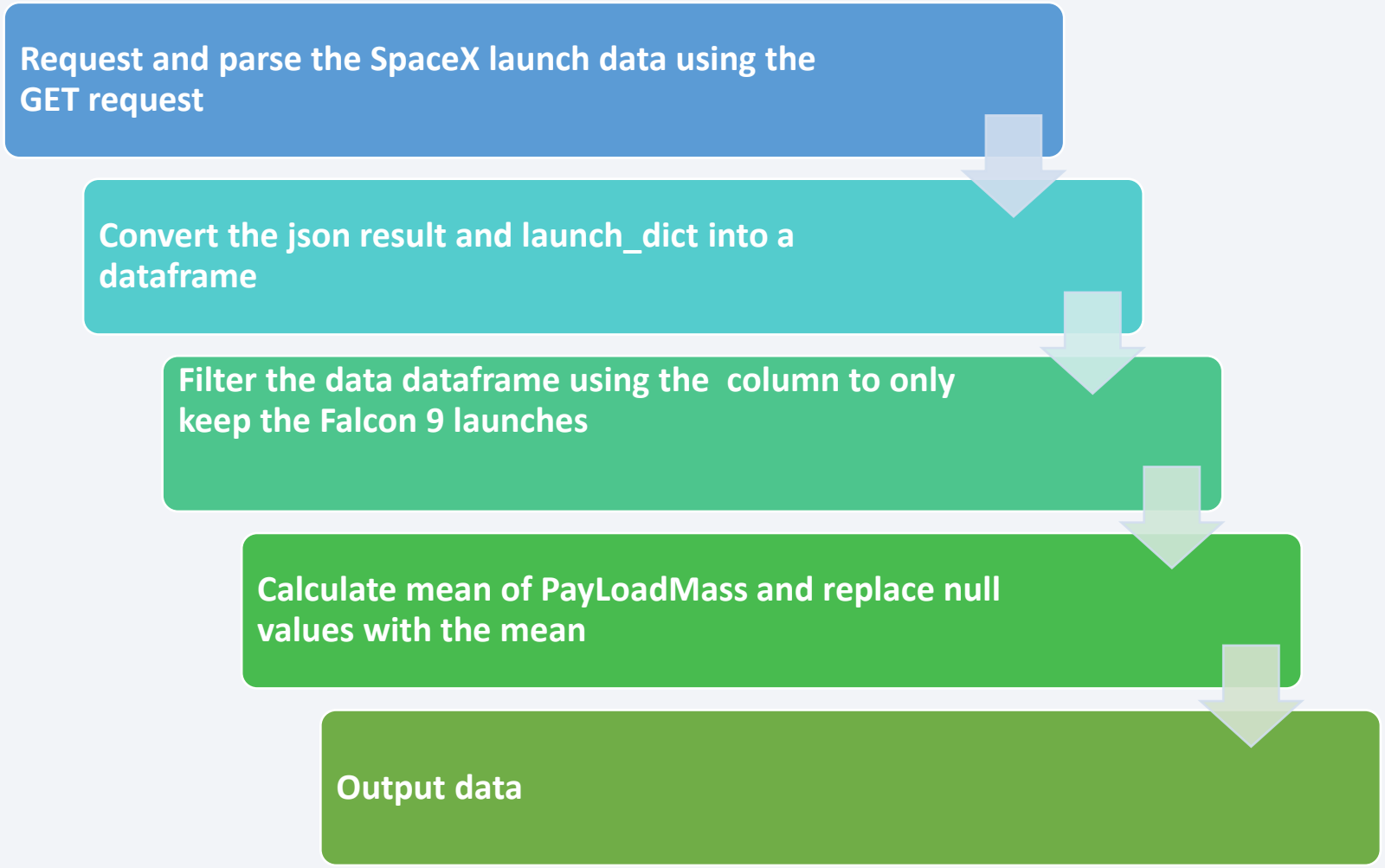
SpaceX advertises Falcon 9 rocket launches on its website for a much lower cost than other providers. The company can reuse the first stage if landing is successful.

As the data scientist on this project, my aim is to predict if the first stage will land successfully in order to determine the cost of a launch.

Section 1

Methodology

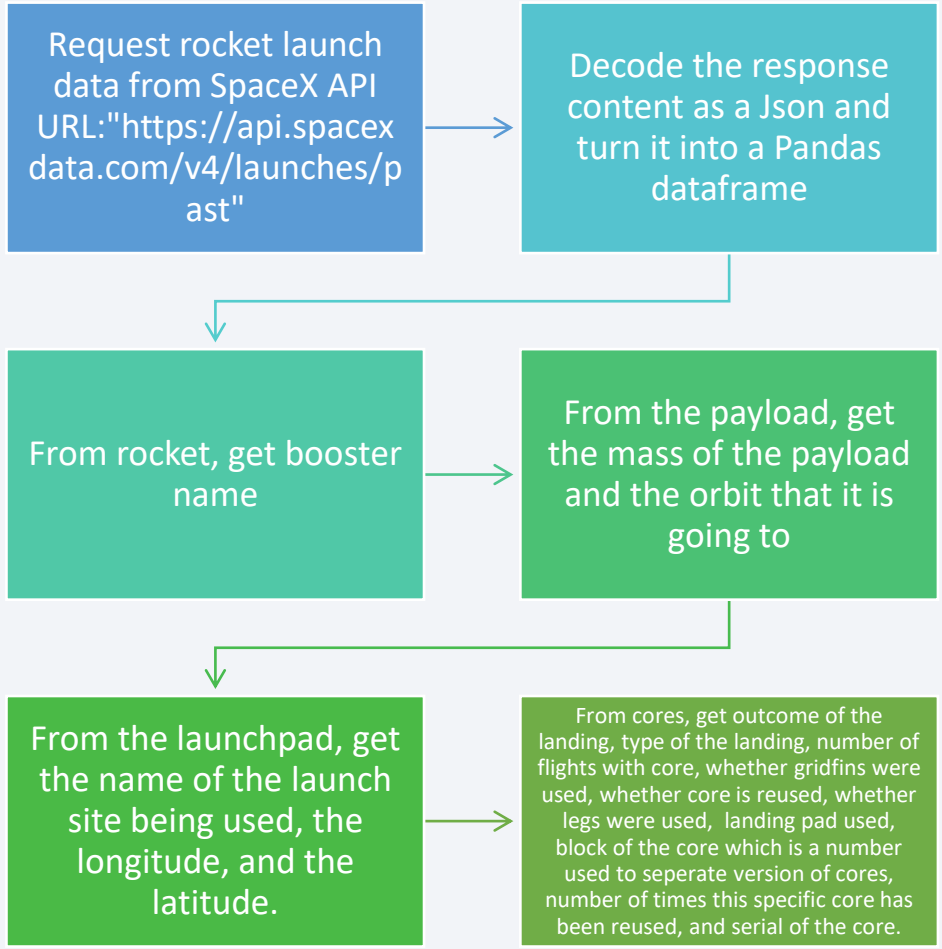
Data Collection



Data Collection – SpaceX API

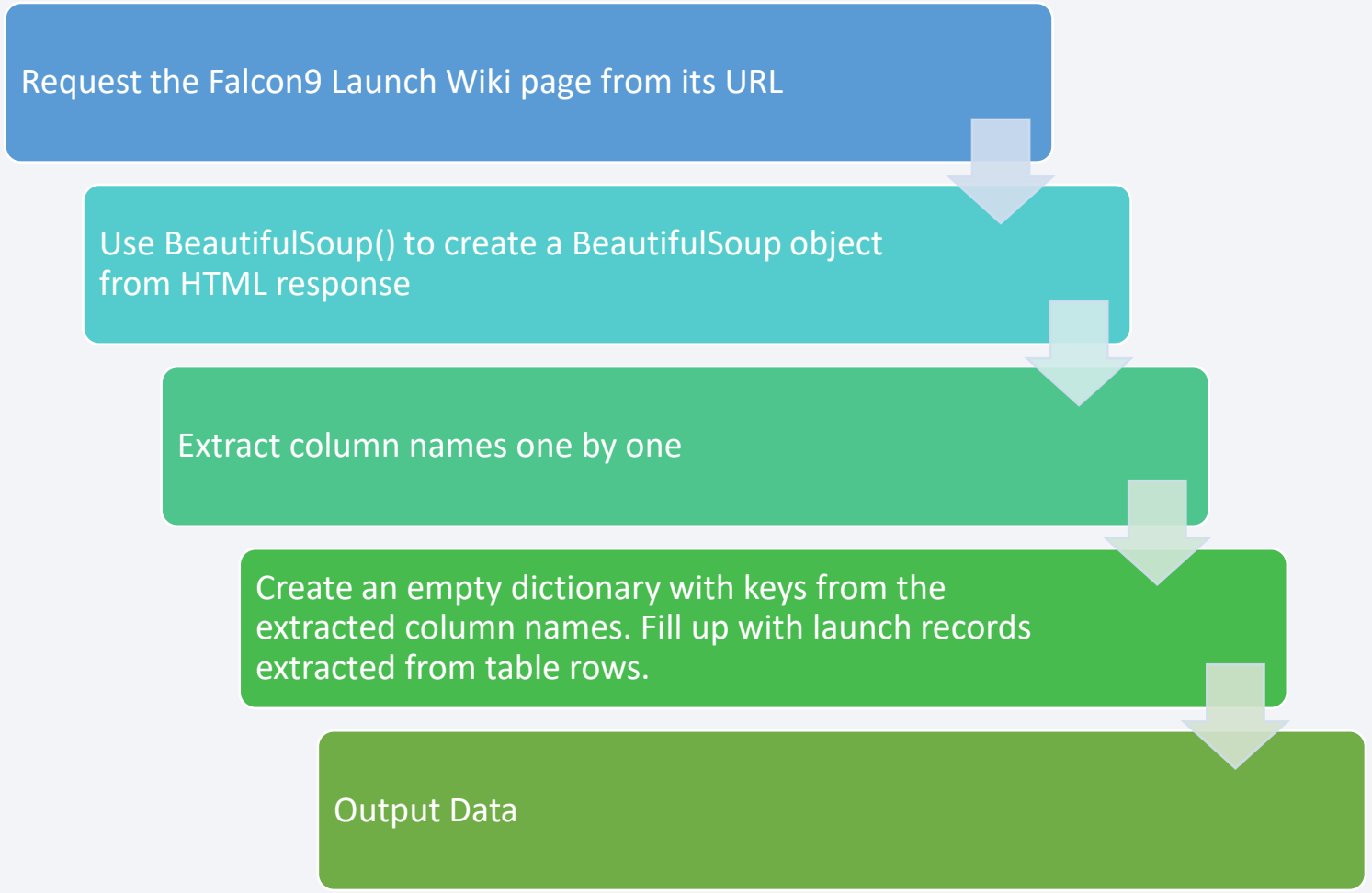
- The flowchart on the process of collecting data from SpaceX API can be seen on the right.

GITHUB LINK



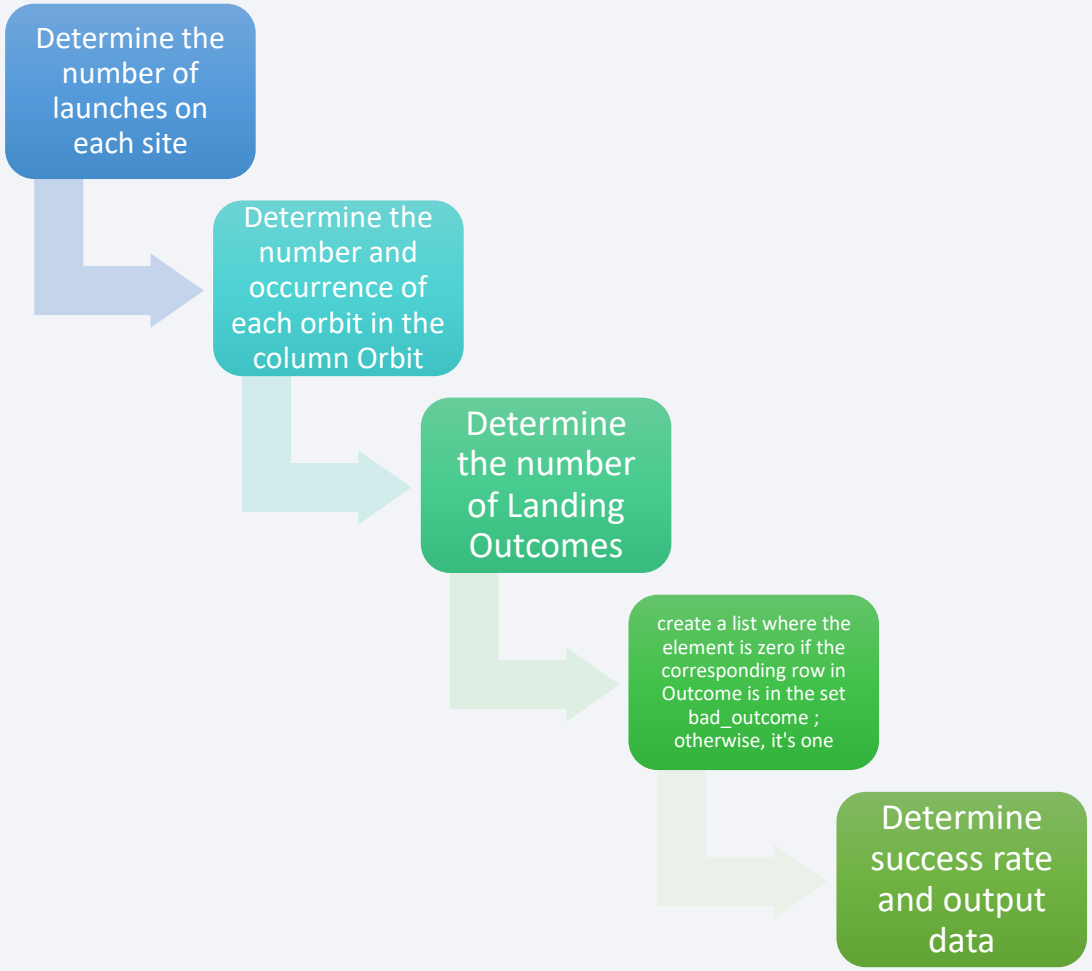
Data Collection - Scraping

- Web scraping process can be seen on the right.
- Web scraping GITHUB Link



Data Wrangling

- See Data Wrangling flowchart on the right:
- Data Wrangling GitHub



EDA with Data Visualization

- Charts used include:

- Scatter Plot: Flight Number vs Payload Mass, Flight Number vs Launch Sites, Payload Mass vs Launch Sites, Flight Number vs Orbit Type, Payload Mass vs Orbit Type

- Bar Chart: Success Rate per Orbit

- Line graph: Success Rate per year

- EDA with Visualization Github

EDA with SQL

Summary of 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 successful 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 successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

- EDA with SQL Github

Build an Interactive Map with Folium

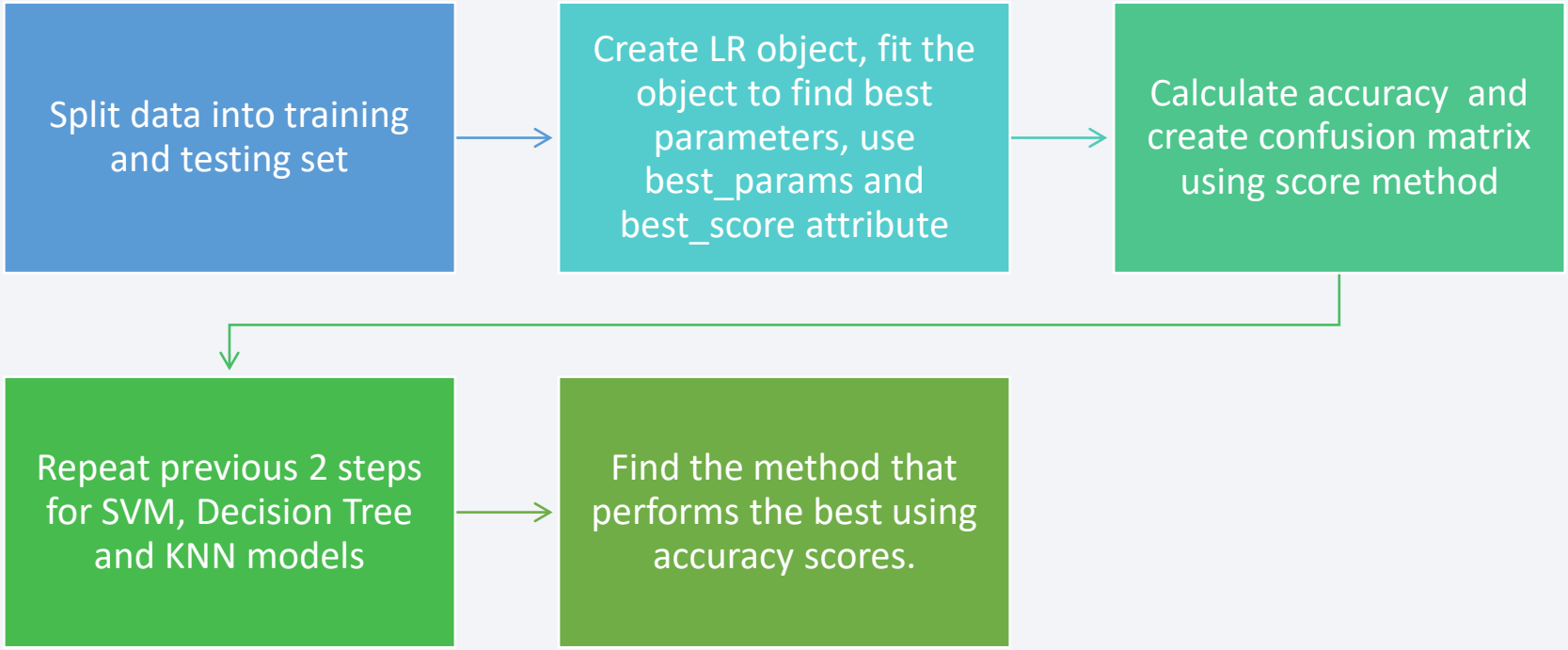
- I used folium.circle and folium.marker to add highlighted circle areas with a text label on the specific coordinated for the launch sites.
- Note that all launch sites are in proximity to the equator line as well as in proximity to the coast.
- Interactive Map with Folium Github

Build a Dashboard with Plotly Dash

- With Plotly Dash, I created Pie charts and scatter charts to visualize success rates of launches by Space X.
- These charts were able to visually display launch success per launch site. The data also gave us an insight into the variables that could have contributed to the launch's success such as booster versions and payload mass.
- Plotly Dash Github

Predictive Analysis (Classification)

- Best Hyperparameter for SVM, KNN, Classification Trees and Logistic Regression was found using the steps on the right:



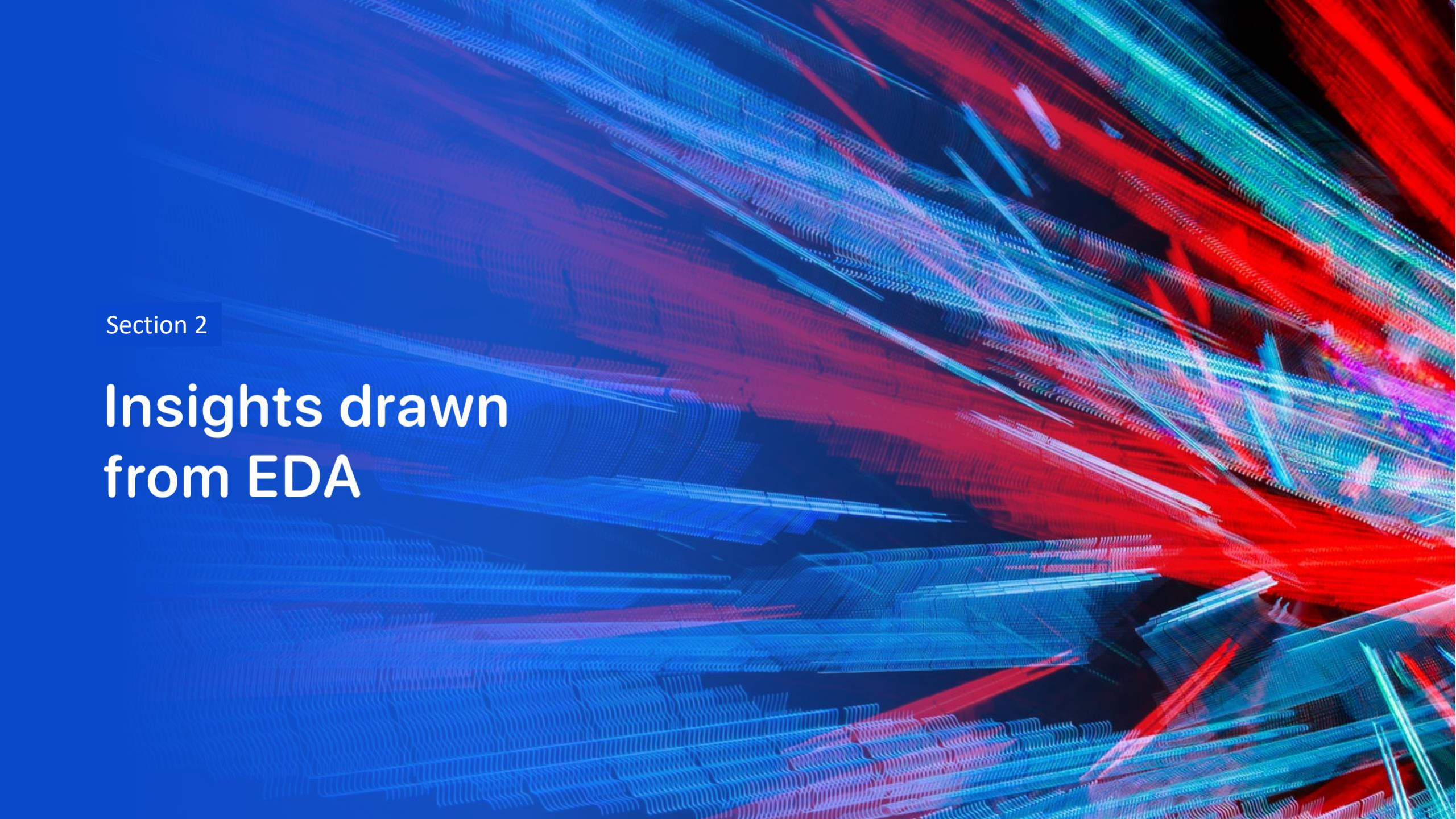
Results

Successful landing outcomes have increased since 2013.

Launch sites are conveniently located near the coast, highways and railways to facilitate transportation and landings.

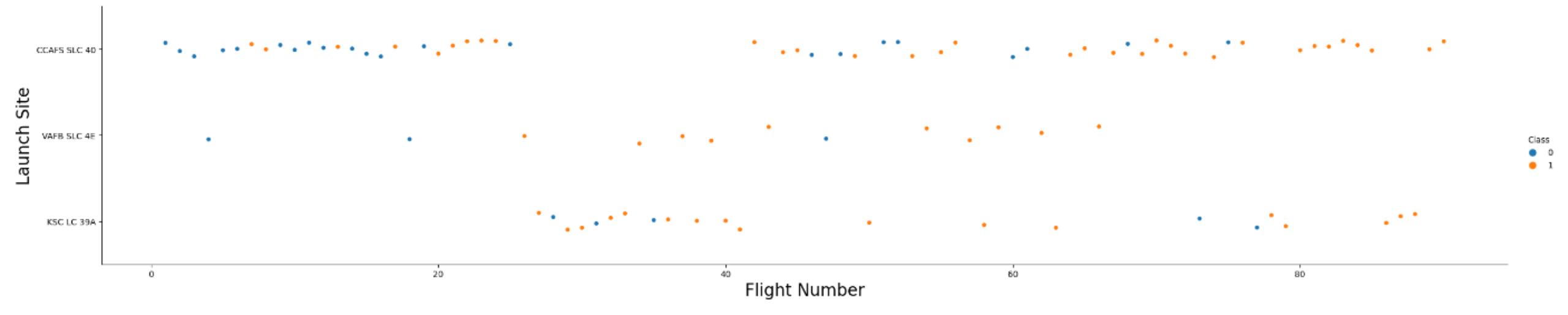
Launch sites are located away from cities to prevent interference with public activity.

All models had an accuracy score of 83.33333%


Section 2
Insights drawn from EDA

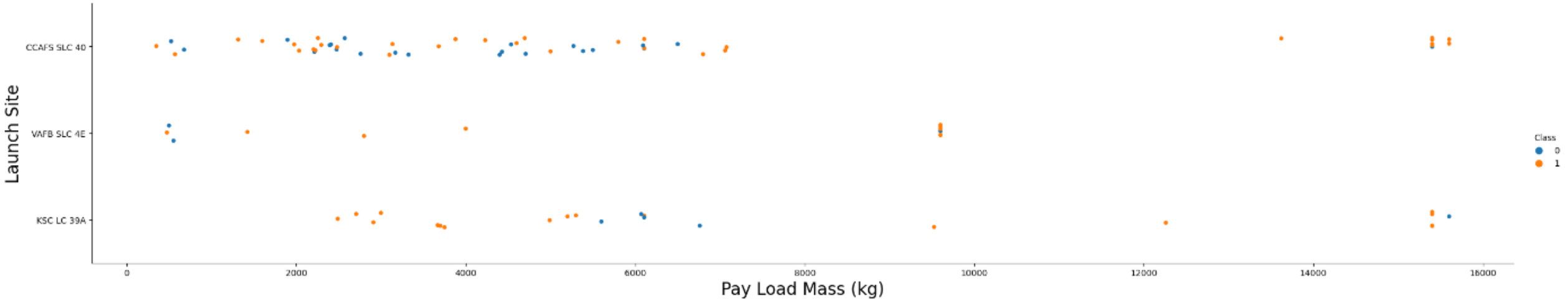
Flight Number vs. Launch Site

- CCAFS LC-40 had the most launches. The blue dots represent failed landing outcomes while the orange represents successful landing outcomes



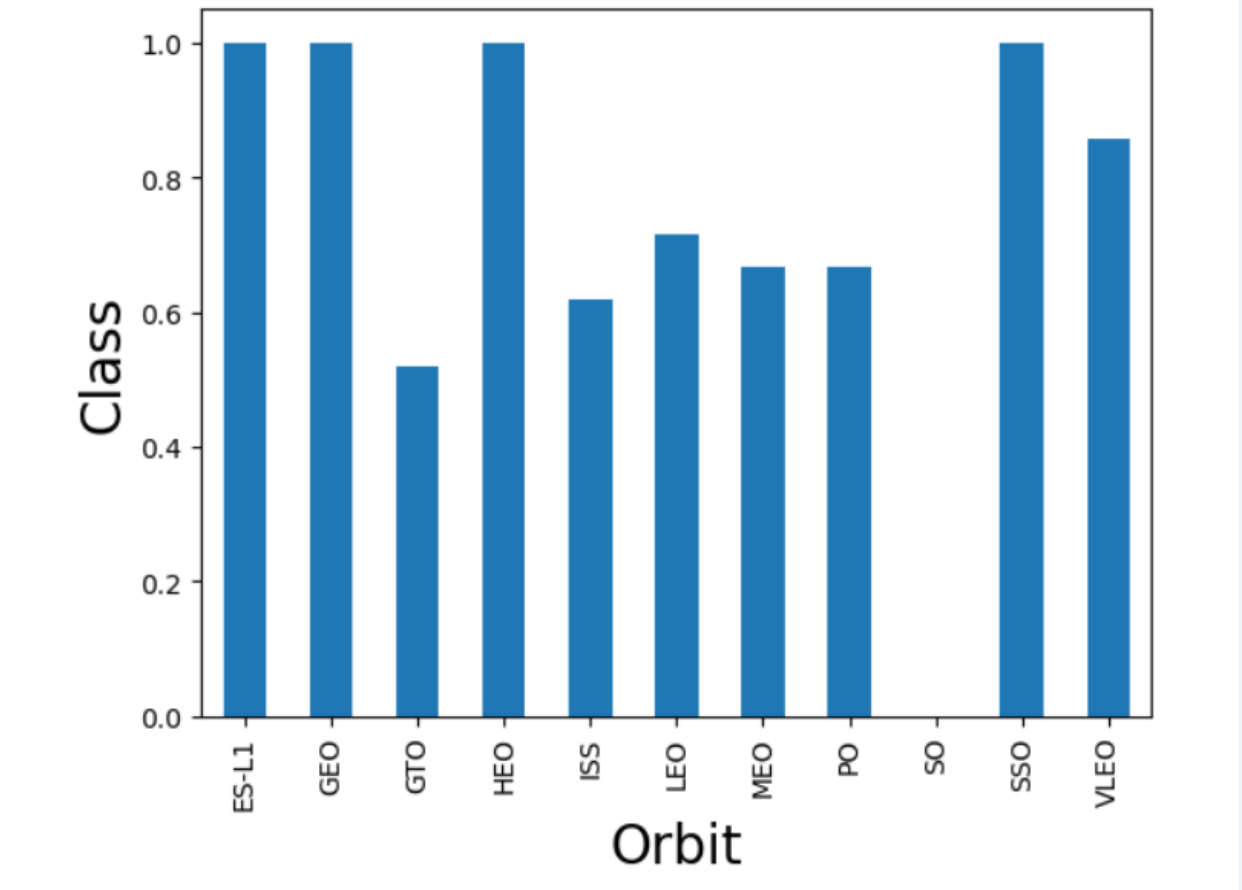
Payload vs. Launch Site

- For the VAFB-SLC launchsite, there are no rockets launched for heavypayload mass(greater than 10000); whereas, the other launchsites do.



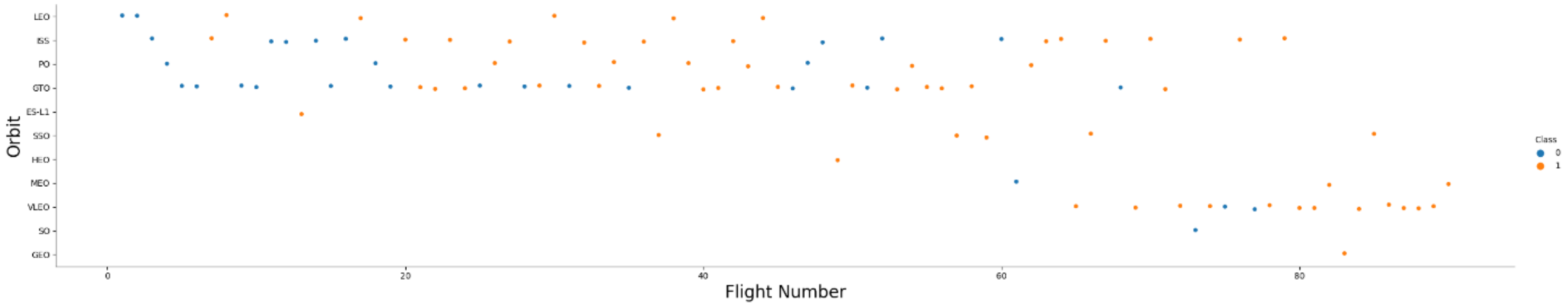
Success Rate vs. Orbit Type

- Orbits ES-L1, GEO, HEO and SSO have the highest success rates, with no failures.
- GTO has the lowest success rate at ~0.55



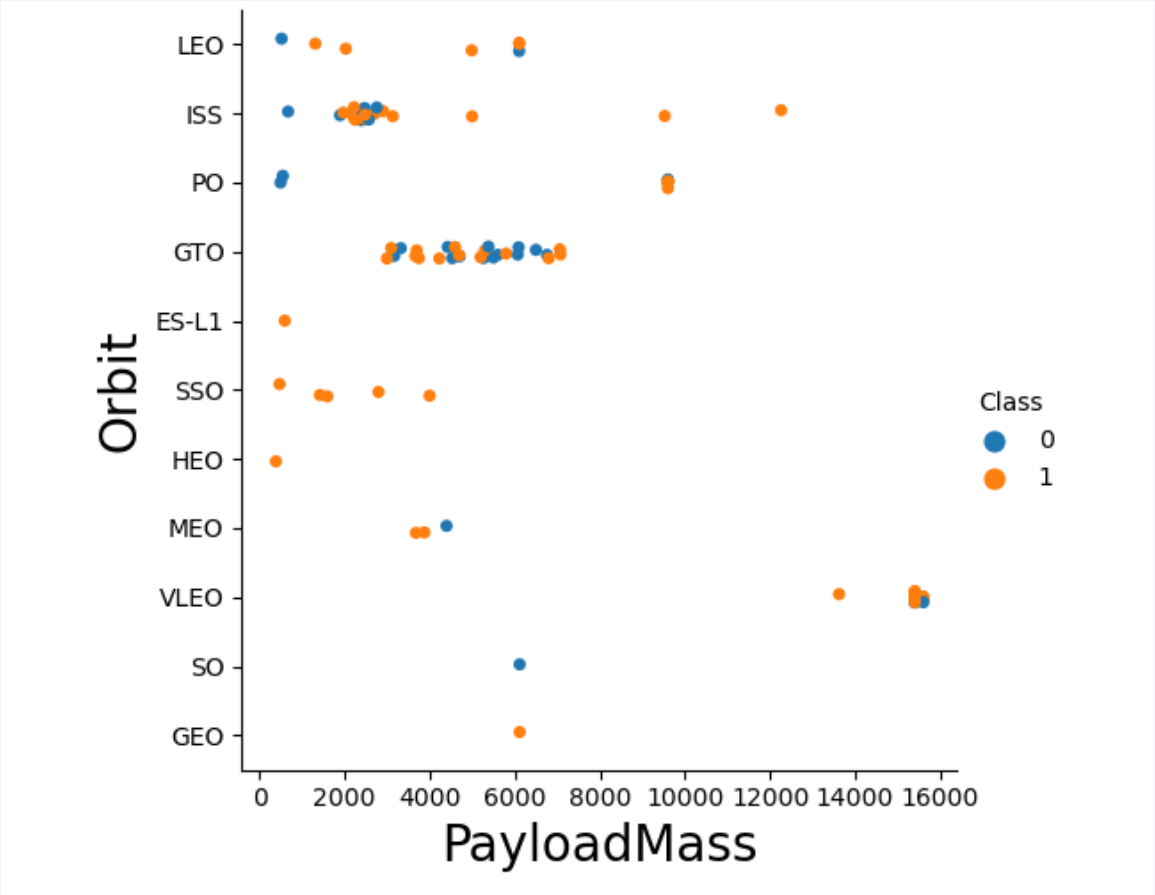
Flight Number vs. Orbit Type

In the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.



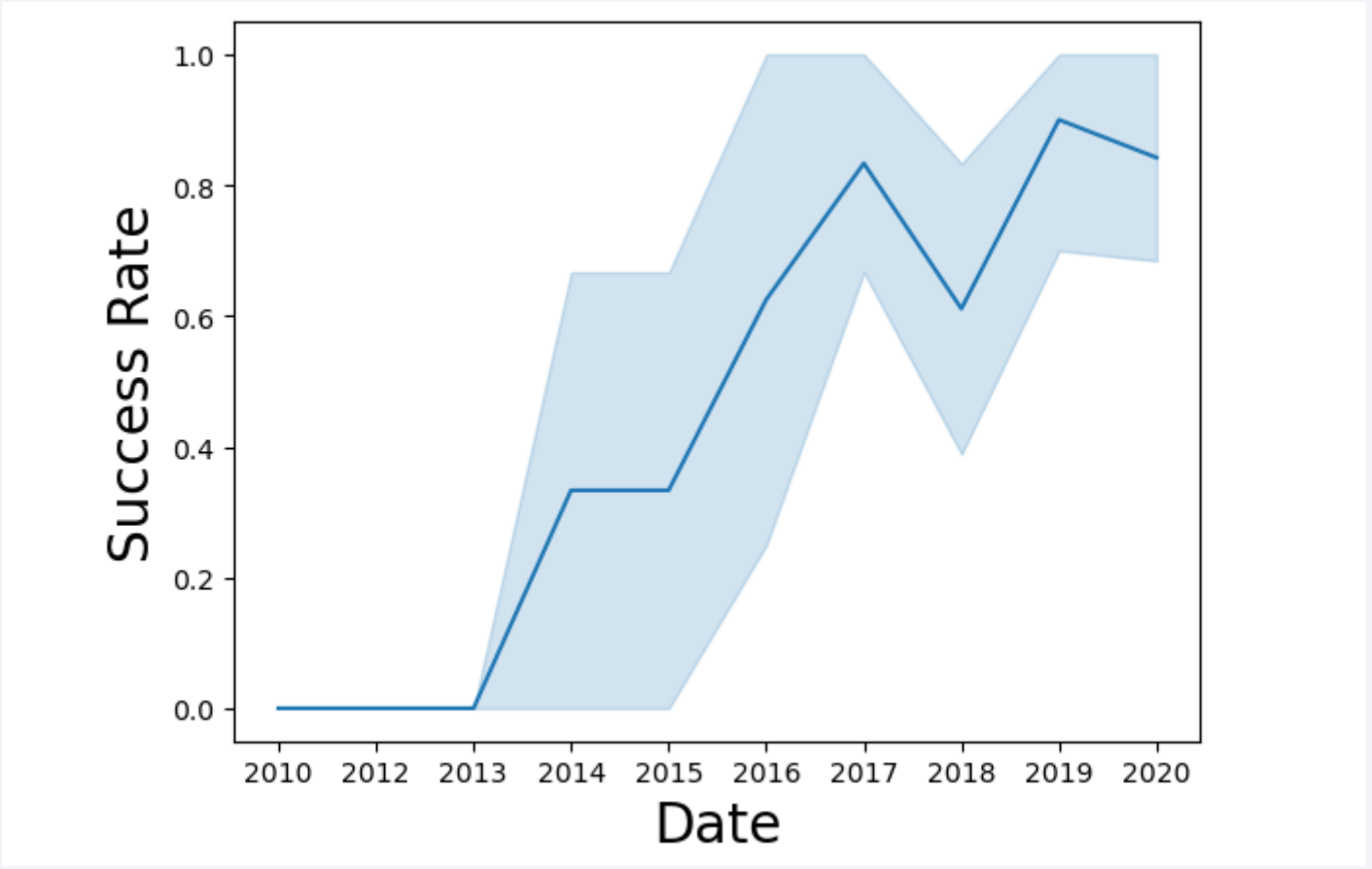
Payload Mass vs. Orbit Type

- With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.
- However for GTO we cannot distinguish this accurately as both positive landing rate and negative landing are present.



Launch Success Yearly Trend

- The success rate has increased since 2013. There was a dip in 2018.



All Launch Site Names

- There are 4 unique launch sites in the data. The result was acquired using the DISTINCT function. See results below:

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

Launch Site Names Begin with 'CCA'

The LIKE and LIMIT functions were used to get the results below:

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

- The SUM function was used to get the result below for boosters launched by NASA (CRS).

SUM (PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

- The AVG function was used to get the results below for booster version F9 v1.1.

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

- The MIN function was used to get the first successful landing date for groundpad.

FIRST_SUCCESS_LANDING
01-05-2017

Successful Drone Ship Landing with Payload between 4000 and 6000

- Boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000 can be seen on the bottom right.

- The MIN and BETWEEN functions were used to get the results

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

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes can be seen below.
- COUNT and GROUP BY functions were used.

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

Boosters Carried Maximum Payload

- The list on the right contains the boosters which have carried the maximum payload mass.
- A subquery was used as well as the MAX function.

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

- The failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015 can be seen below.

- There were 2 results for 2015.

month	year	Mission_Outcome	Landing _Outcome	Booster_Version	Launch_Site
01	2015	Success	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	2015	Success	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

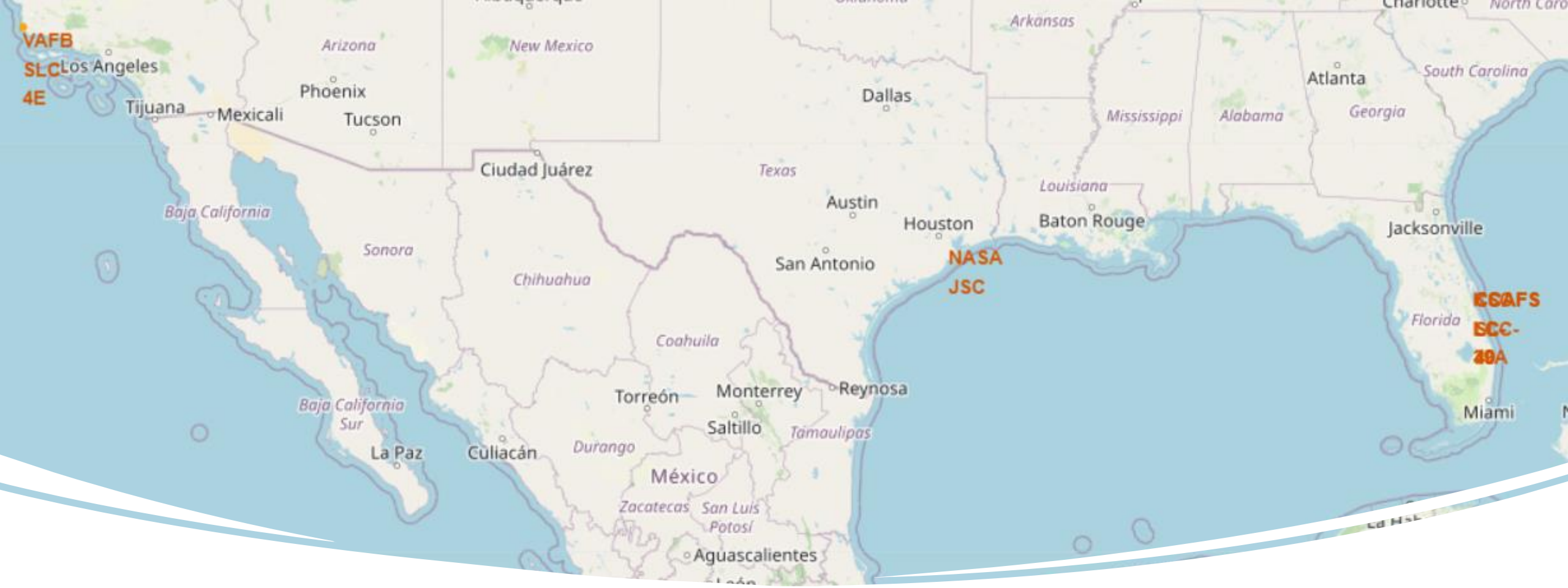
- Successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order can be seen below.

- COUNT, GROUP BY and ORDER BY functions were used.

Landing _Outcome	Total
Success	20
Success (drone ship)	8
Success (ground pad)	6

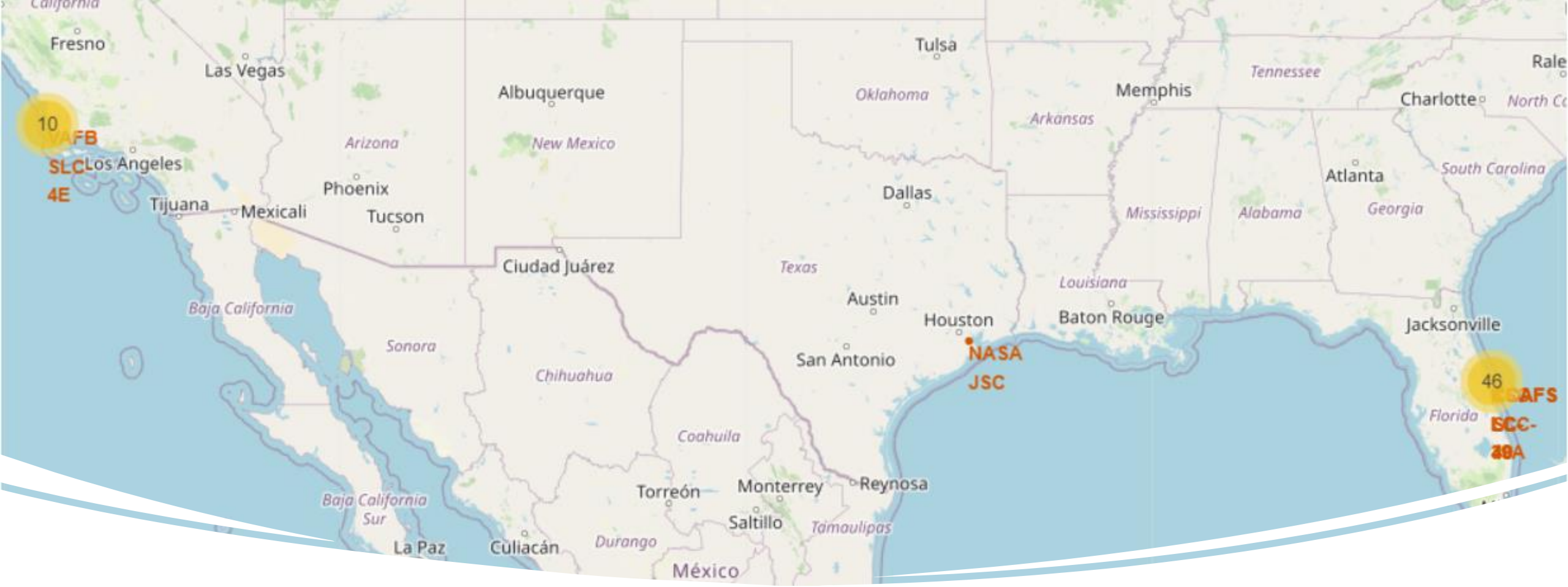
Section 3

Launch Sites Proximities Analysis



Launch Sites Folium Map

- The screenshot shows the highlighted circles and labels of each launch site in the data set consisting of circles and markers



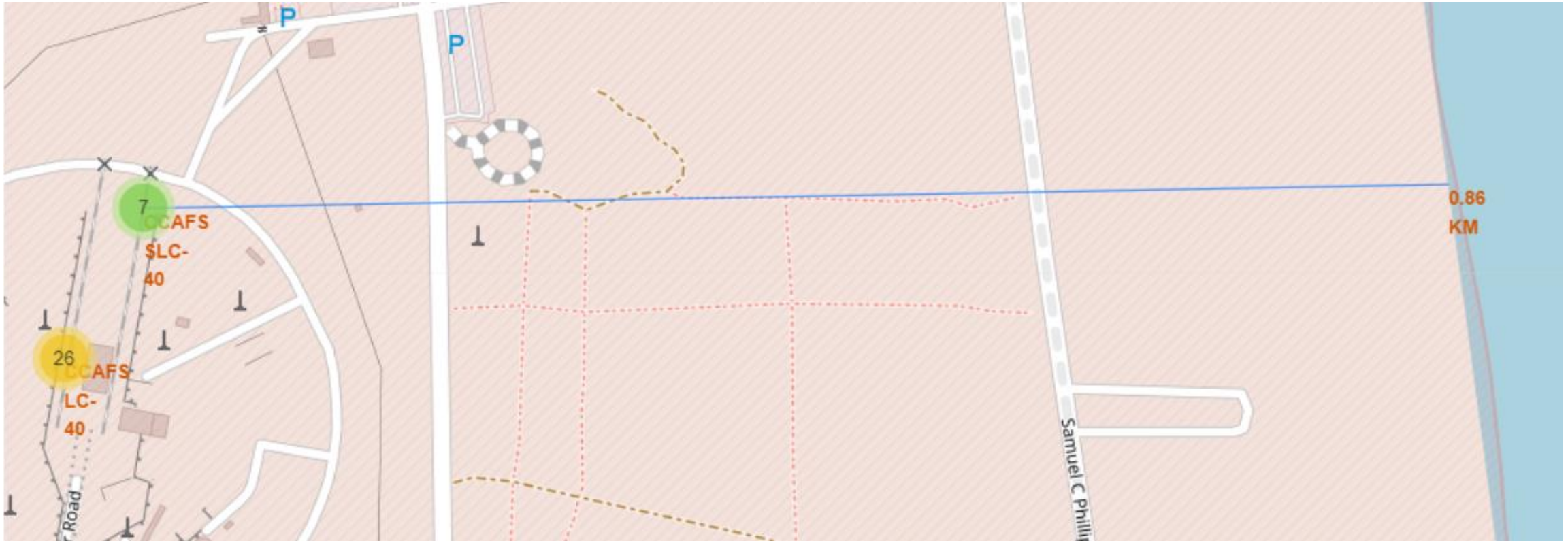
Launch Sites Success & Failures (Folium Map)

- Note this map is different from the previous slide as numbers are now included to indicate the launch site's respective success and failures.

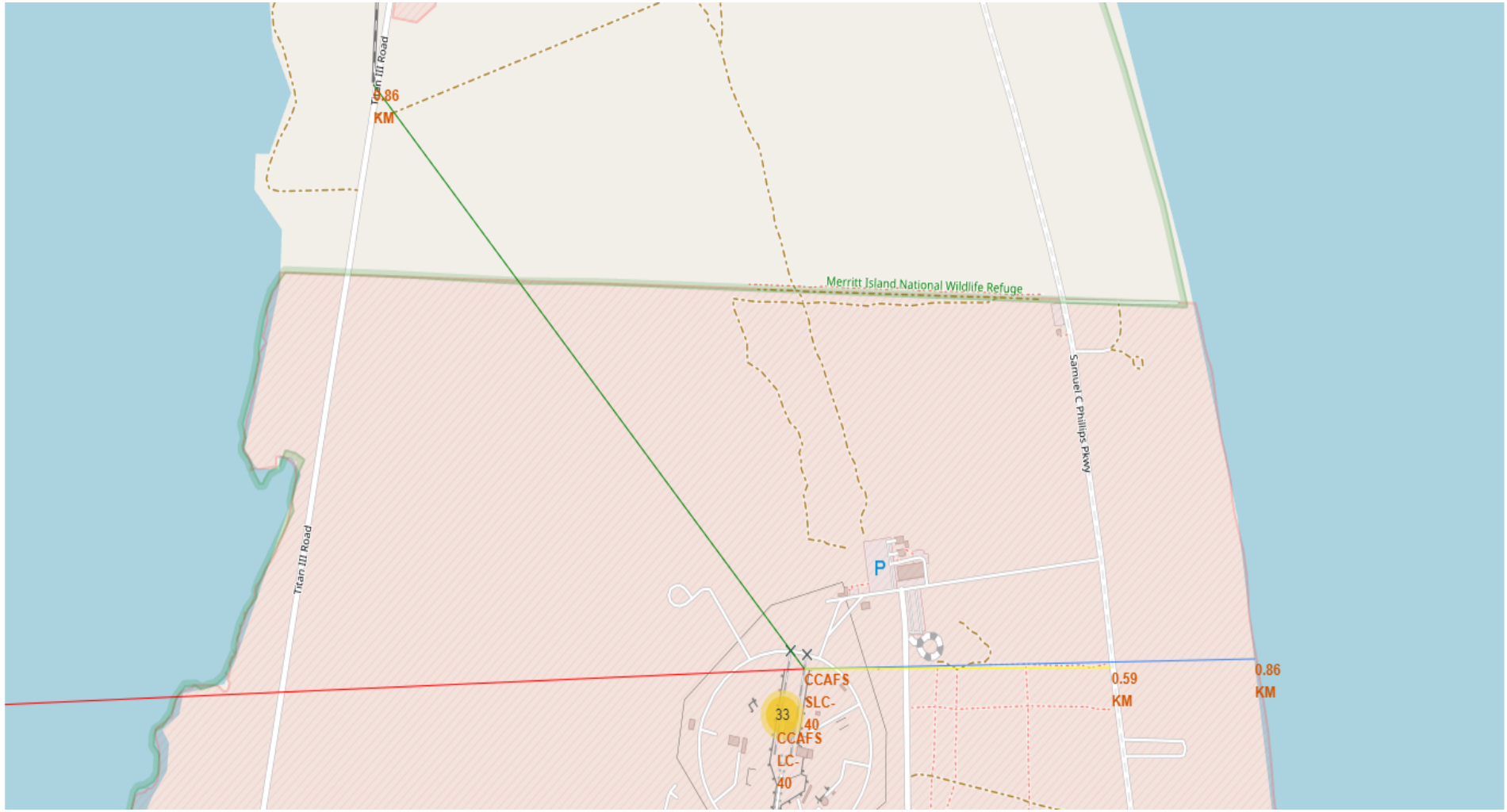
Zoomed in screenshot of previous map

PolyLine between Launch Site

- the coastline coordinates and launch site coordinates were used to create the line between the site and the coasat. Note the distance between those in the image below is 0.86 KM.

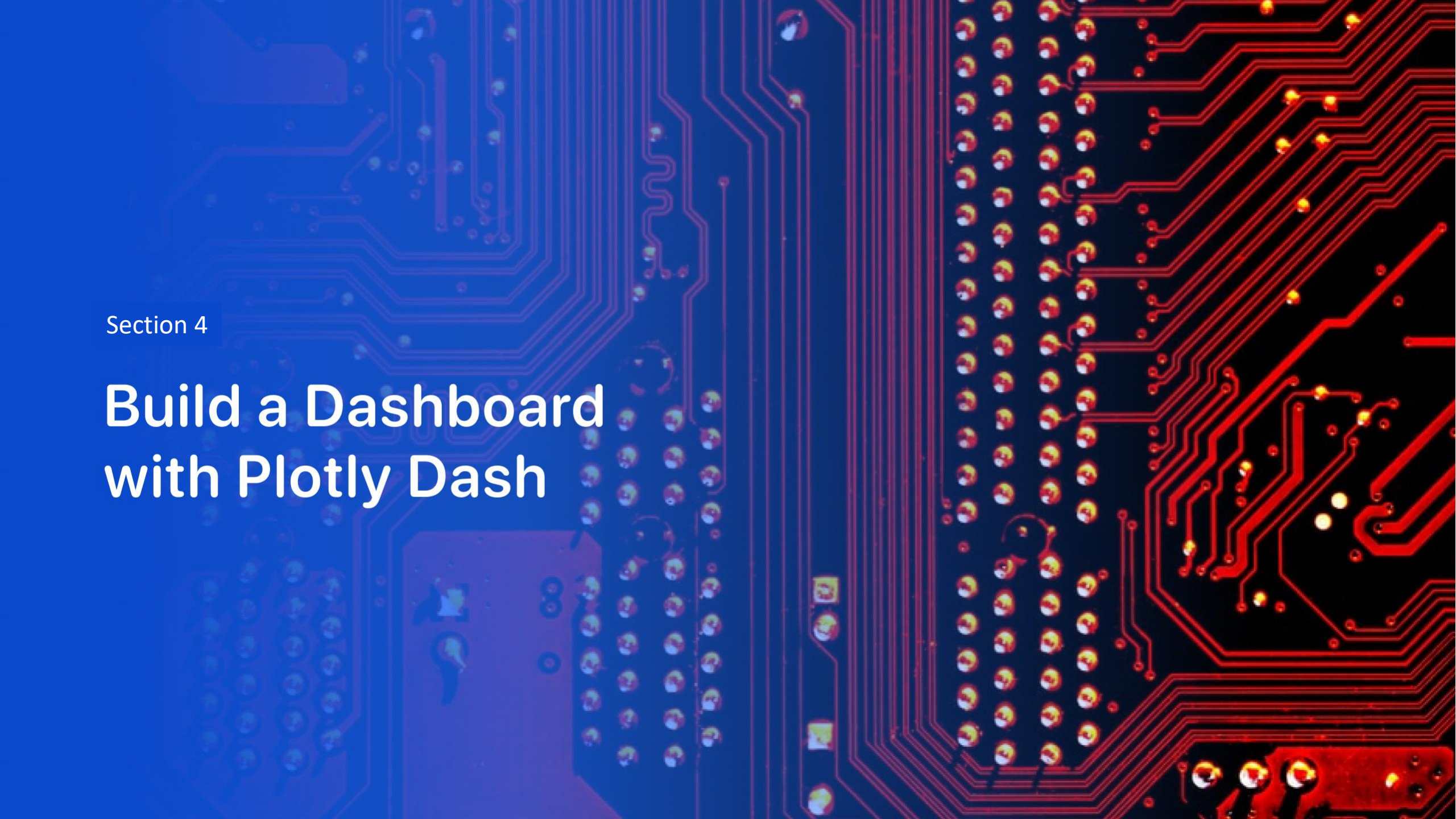


Launch site to closest city, railway, highway



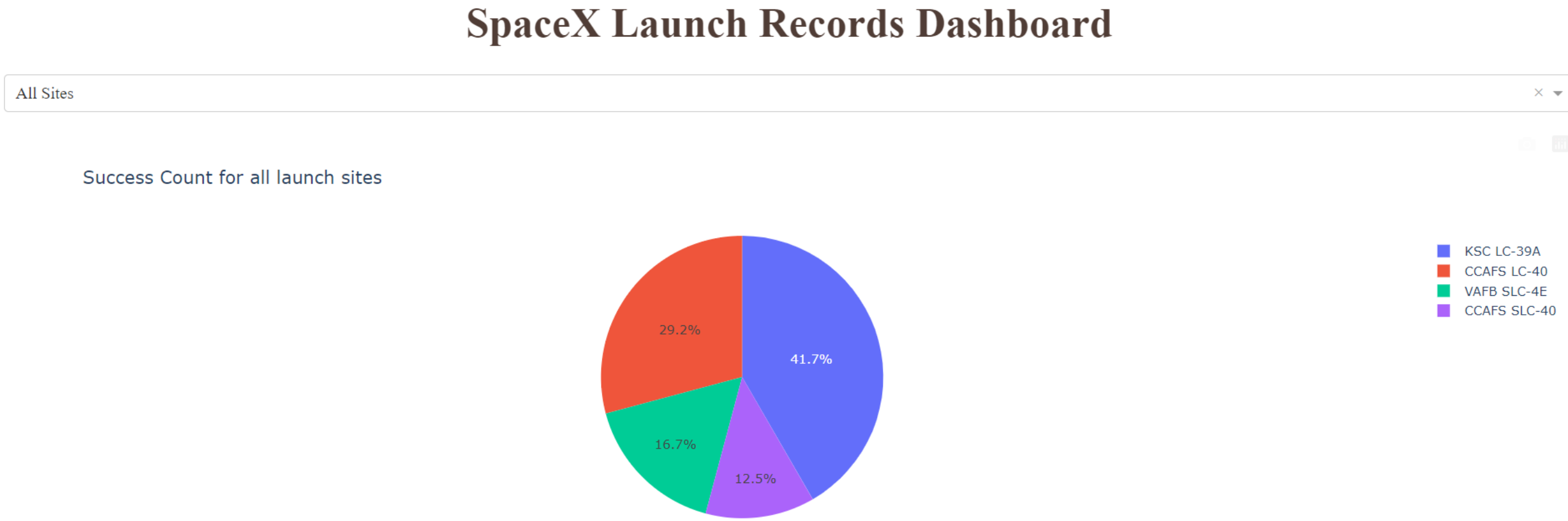
Different color lines point to different symbols near the launch site.

Note that launch sites are in close proximity to railways, highways and coastlines, but are far from cities.


Section 4
Build a Dashboard with Plotly Dash

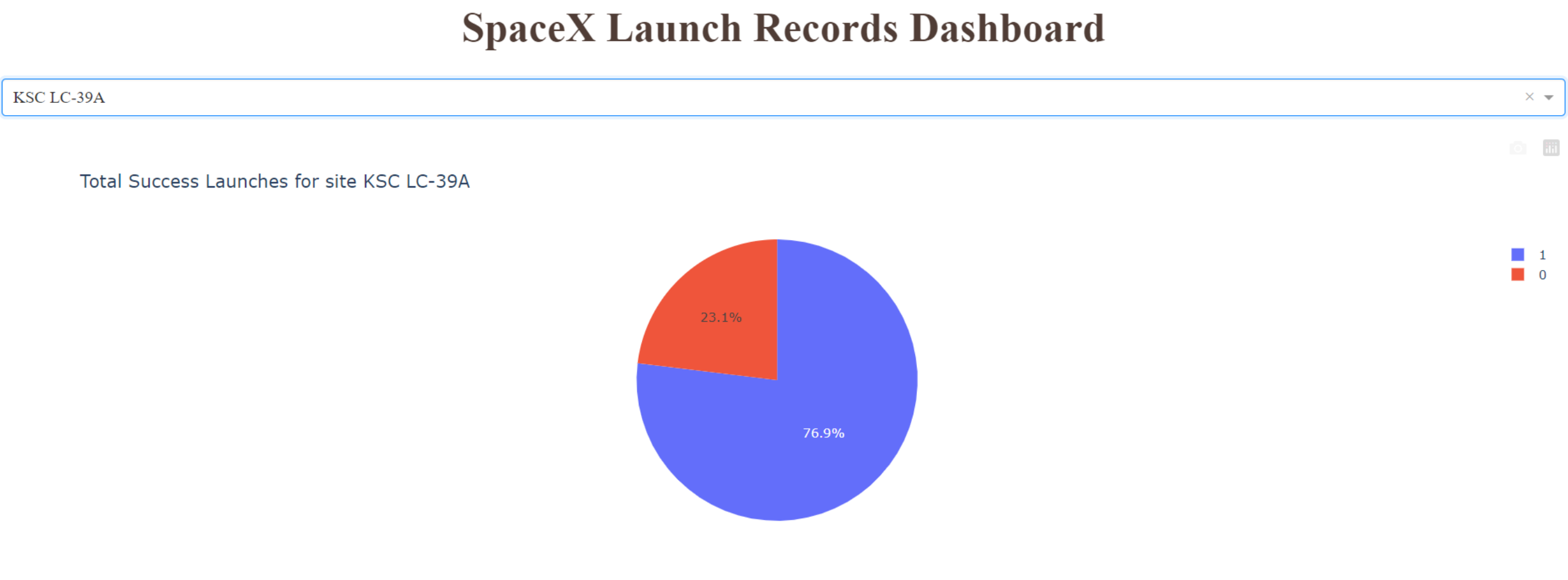
Pie Chart

- Each division in the pie chart represents a launch site. The success rates can be seen within its respective division.

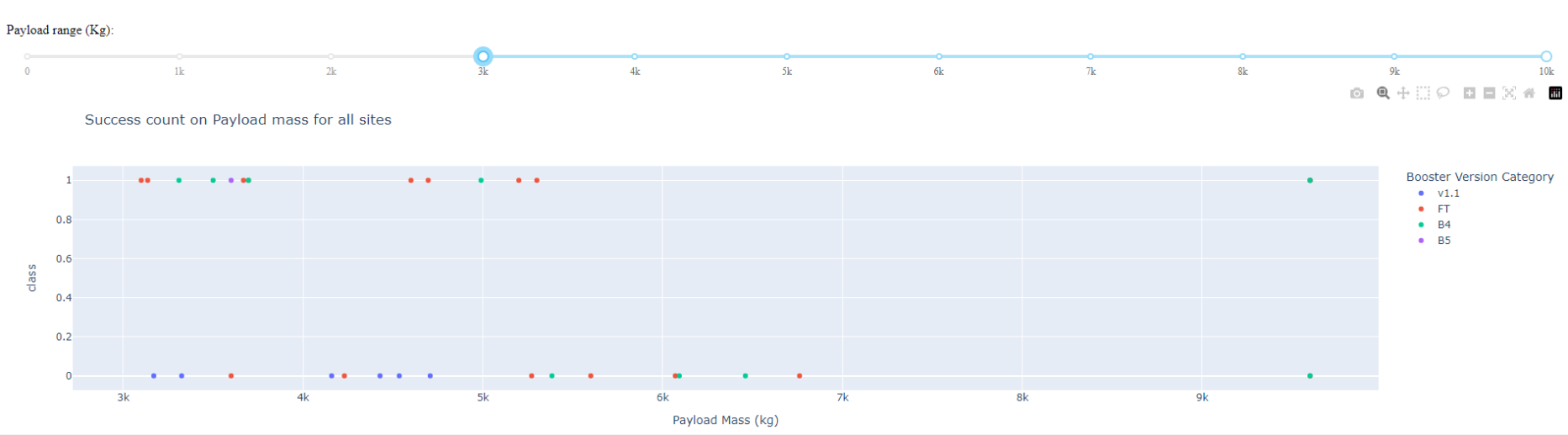


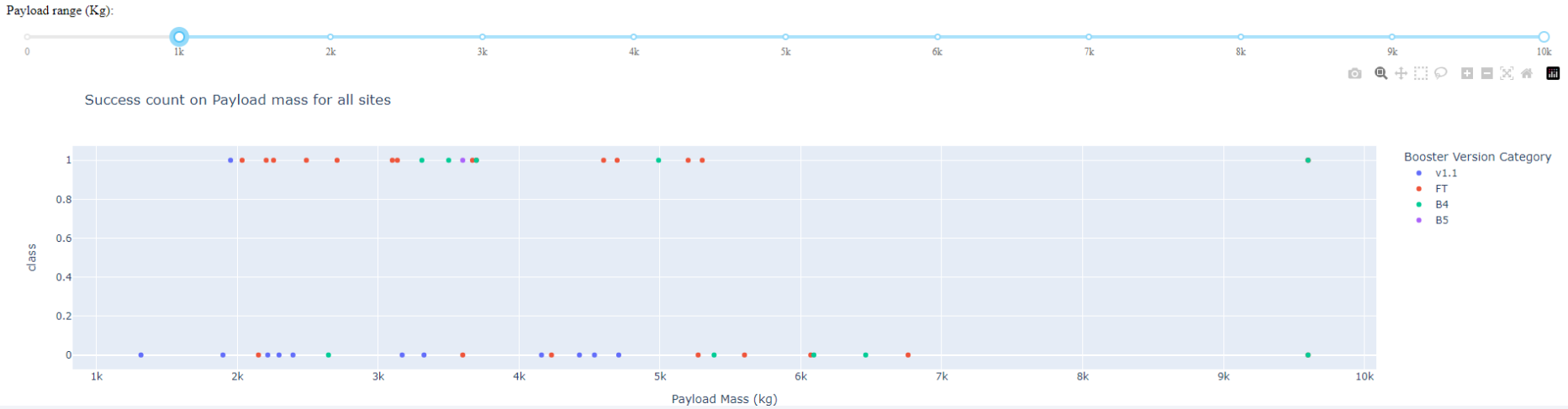
Launch Success Ratio

- KSC LC-39A has the most successful launch success ratio (successful launches/total launches)
- The blue division represents successful while red represents failed



Payload mass for all sites





On the left are payload masses with the payload range at different points on the slider.

There are less Booster Version Categories as the slider increases to the right.

Section 5

Predictive Analysis (Classification)

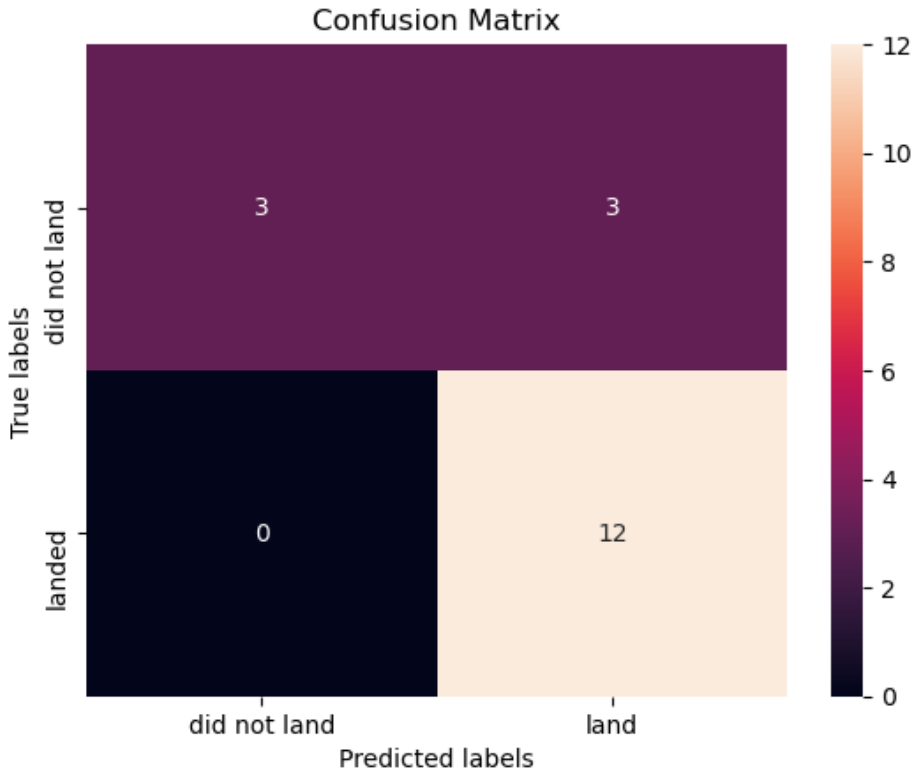
Classification Accuracy

- All models have the same level of accuracy.

	Model Method	Accuracy Score (%)
0	Support Vector Machine	83.333333
1	Logistic Regression	83.333333
2	K Nearest Neighbour	83.333333
3	Decision Tree	83.333333

Confusion Matrix

- Using the Logistics Regression, the following confusion matrix was produced
- The model failed to predict 3 labels. The major problem is false positives.



Conclusions

- Given the data, we are able to conclude that:
- All Launch Sites are located near the coast and away from cities.
- Launch Site KSC LC-39A has had the highest launch success rate with
- Success Rate among Landing Outcomes has increased since 2013 and is expected to continue.
- The Machine Learning Model is able to predict if the first stage will land successfully with 83.33333% accuracy.

Thank you!