IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Dustin Cole
8/16/2023

Outline

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

Executive Summary

- Summary of methodologies:
  - Collected data we used the SpaceX API via Jupyter labs.
  - Then used Jupyter Labs to wrangle and analyze.
  - Developed visualizations and interactive dashboards via Jupyter and Plotly
  - Developed machine learning algorithms for prediction via Jupyter.
- Summary of all results
  - Found features that best predict the odds of a successful mission.

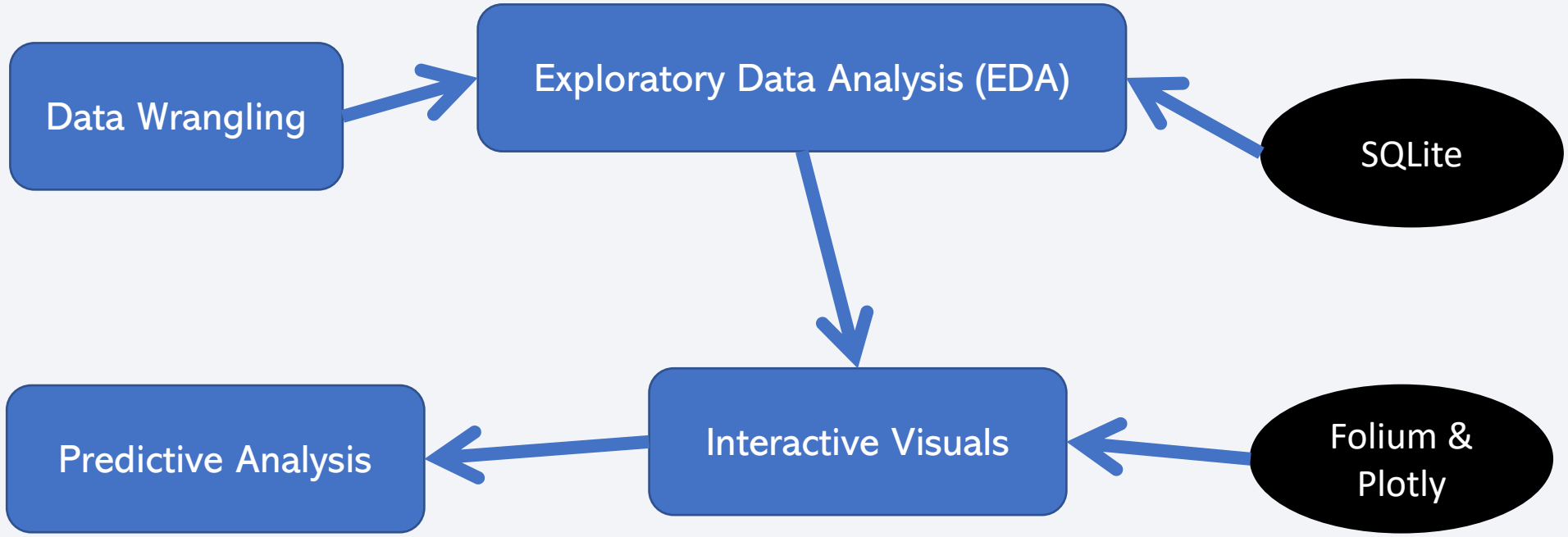
Introduction

- We wanted to see if a new company ‘SpaceY’ could be as successful as ‘SpaceX’, with spacecraft landings. We used data science to solve the problem.
- We want to identify best places to launch spacecraft from and see how much it will cost, in addition to predicting success of the missions.

Section 1

Methodology

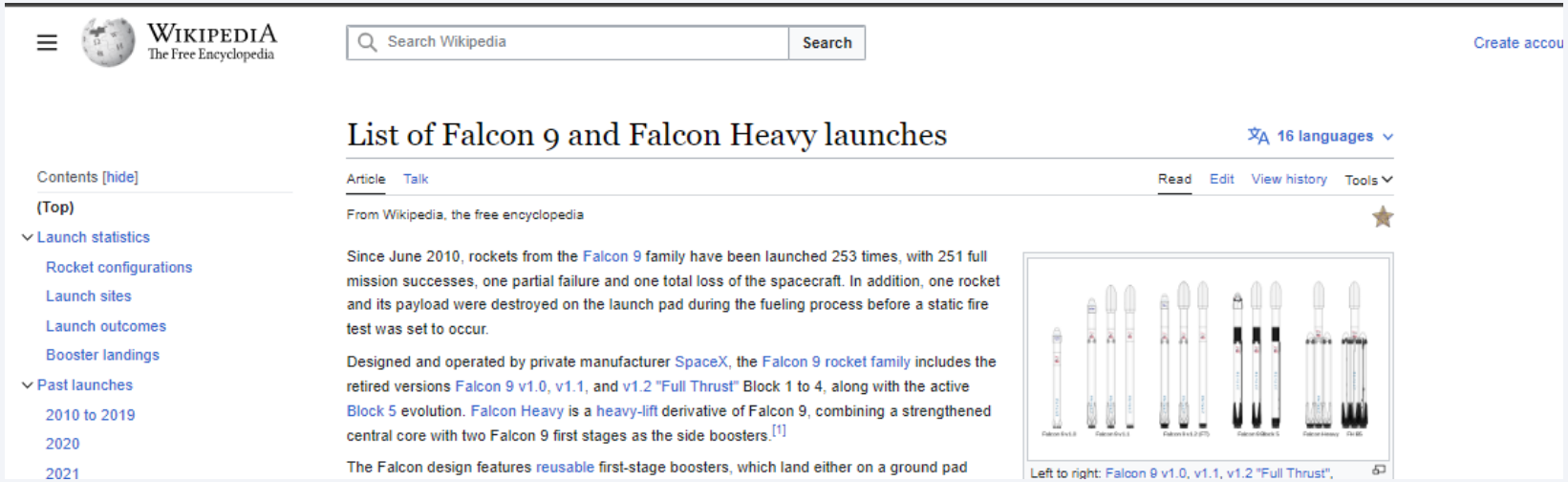
Methodology



Data Collection

Sources:
-Retrieved data from https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-

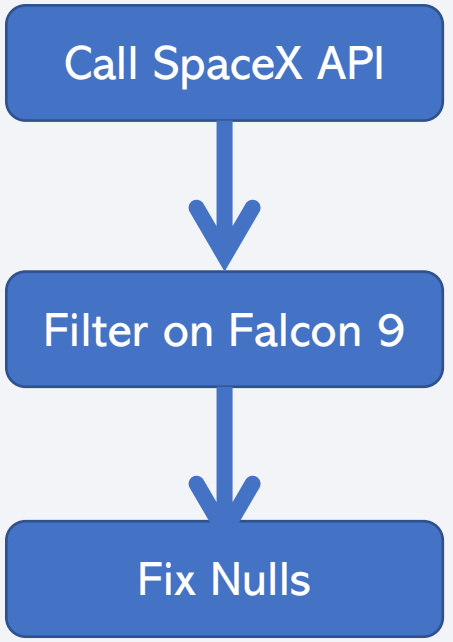
-SkillsNetwork/datasets/dataset_part_1.csv
Retrieved data from List of Falcon 9 and Falcon Heavy launches – Wikipedia



Data Collection – SpaceX API

- Utilize SpaceX Public API

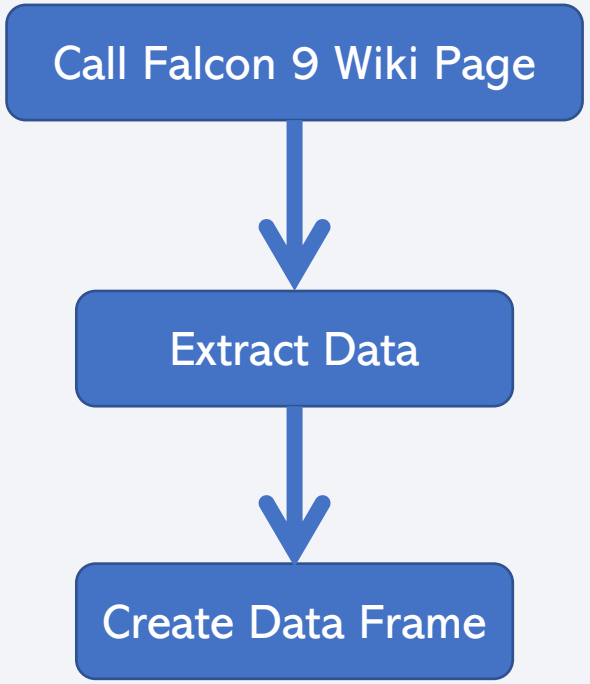
- https://github.com/dcole2200/Applied-Data-Science-Capstone/blob/main/spacex-data-collection-api.ipynb



Data Collection - Scraping

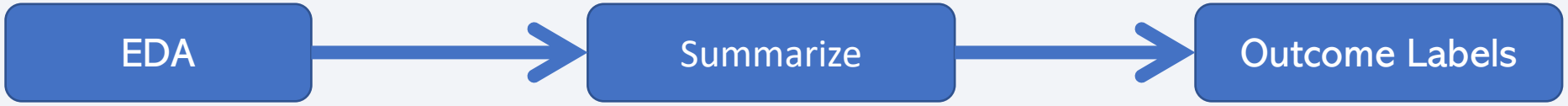
- Public data available via Wikipedia.

- https://github.com/dcole2200/Applied-Data-Science-Capstone/blob/main/webscraping.ipynb



Data Wrangling

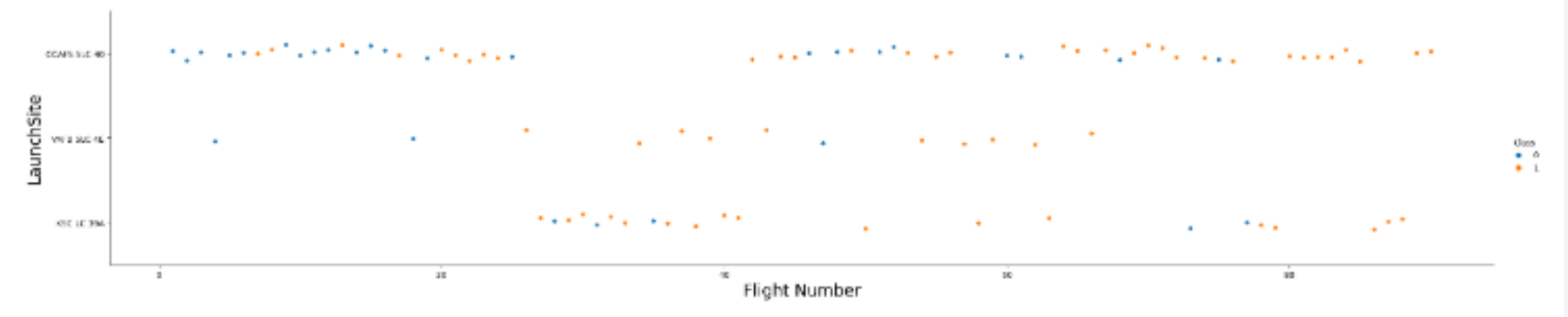
- Perform EDA, then Summarize the Data, then Create labels from Outcomes.



https://github.com/dcole2200/Applied-Data-Science-Capstone/blob/main/data_wrangling.ipynb

EDA with Data Visualization

- Visuals Included: y="PayloadMass", x="FlightNumber“, y="LaunchSite", x="FlightNumber“, y="LaunchSite", x="PayloadMass"
- Example Below:



Applied-Data-Science-Capstone/dataviz.ipynb at main · dcole2200/Applied-Data-Science-Capstone (github.com)

EDA with SQL

- Display the names of the unique launch sites in the space mission
- Display 5 records where launch sites begin with the string 'CCA'
- Display average payload mass carried by booster version F9 v1.1
- Display the total payload mass carried by boosters launched by NASA (CRS)
- 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 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.

Applied-Data-Science-Capstone/sqllite.ipynb at main · dcole2200/Applied-Data-Science-Capstone (github.com)

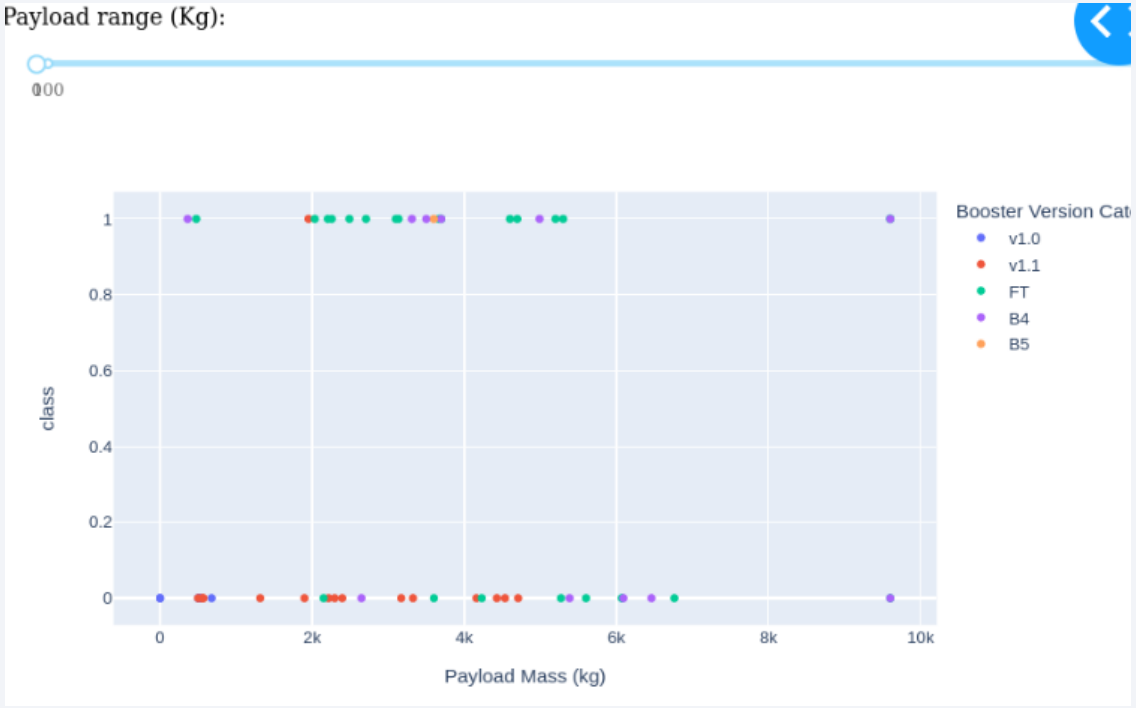
Build an Interactive Map with Folium

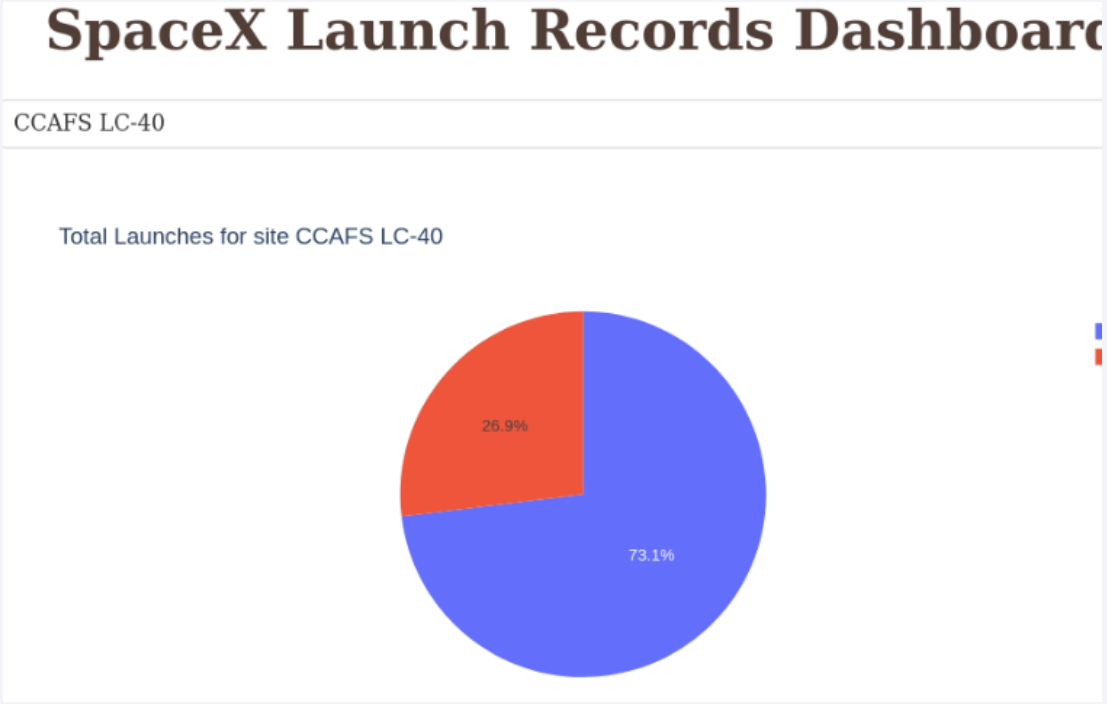
- Used markers, marker clusters, circles, and lines in Folium maps
- Used these objects to show coordinates, groups of coordinates, launch sites, and distances between.

Applied-Data-Science-Capstone/launch_site_location.ipynb at main · dcole2200/Applied-Data-Science-Capstone (github.com)

Build a Dashboard with Plotly Dash

- Created dashboard to analyze site launches and payloads





https://github.com/dcole2200/Applied-Data-Science-Capstone/blob/main/dashapp.py

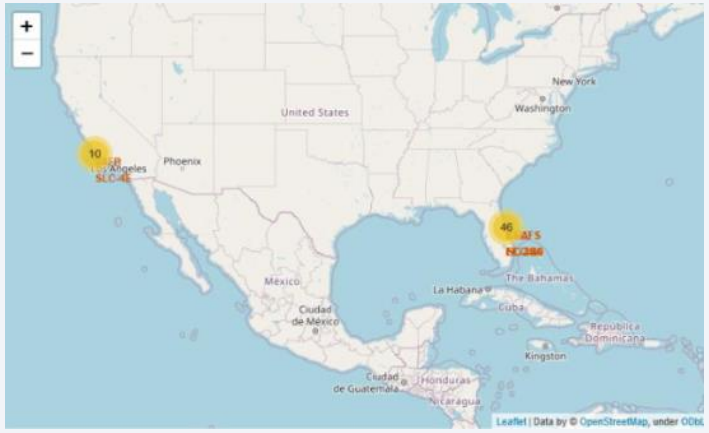
Predictive Analysis (Classification)

- Used logistic regression, decision trees, k nearest neighbors, and support vector machine for predictive analytics.

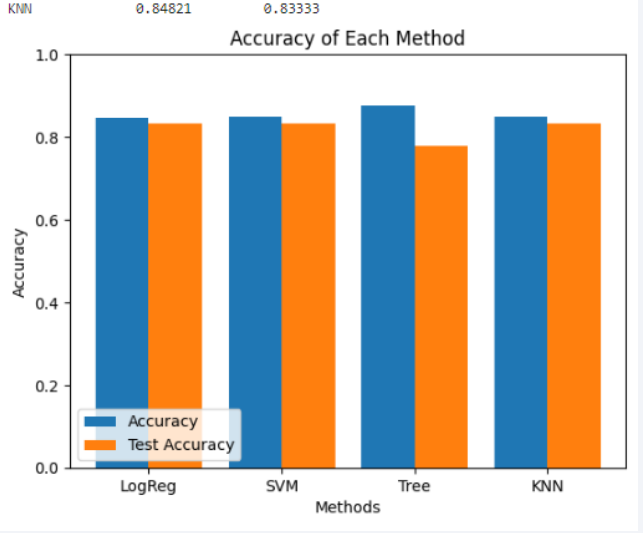
dcole2200/Applied-Data-Science-Capstone (github.com)

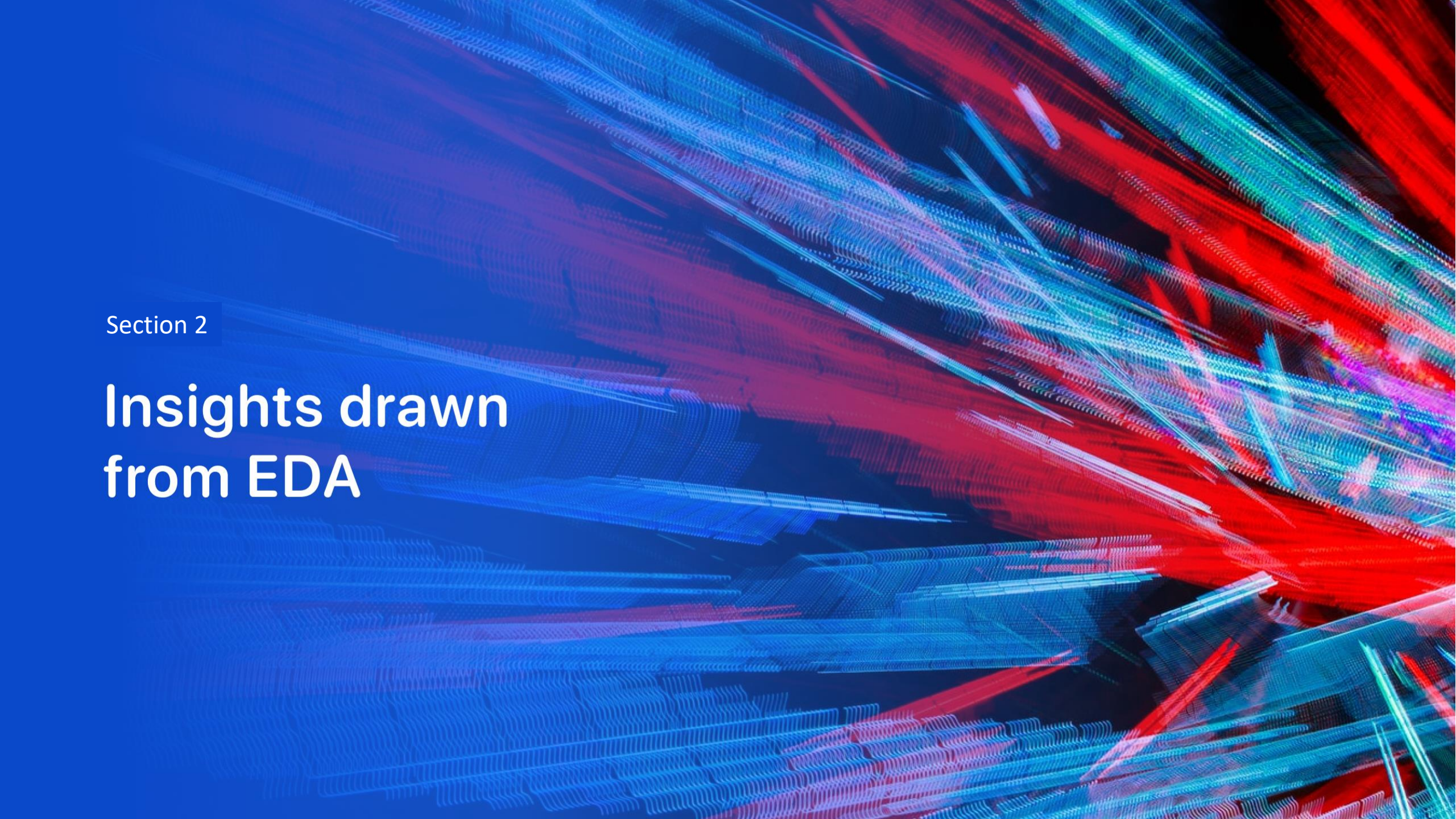
Results

- Exploratory data analysis results: Very high mission success rate, 4 launch sites used, avg F9 payload is around 2.9K kg, landing successes have improved.
- Interactive analytics: Most launches are on the coasts. (See images below)
- Predictive analysis results: Decision Tree is the best method and had an 87% + accuracy.



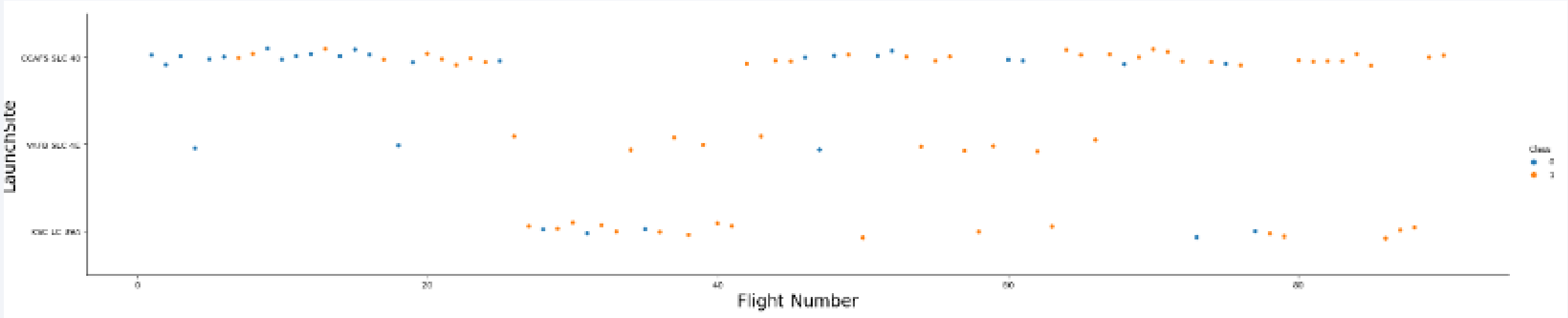
Model	Accuracy	TestAccuracy
LogReg	0.84643	0.83333
SVM	0.84821	0.83333
Tree	0.875	0.77778
KNN	0.84821	0.83333




Section 2
Insights drawn from EDA

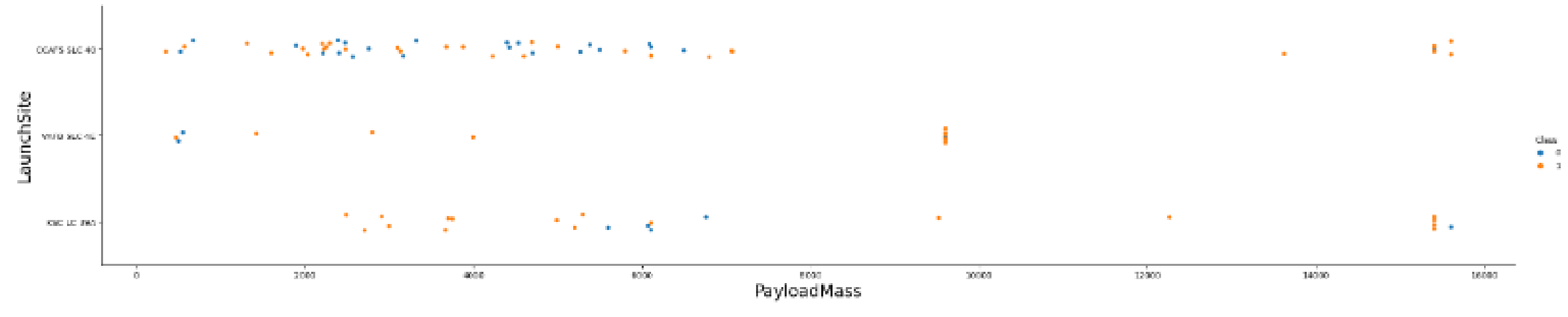
Flight Number vs. Launch Site

- It appears success rates may have improved.



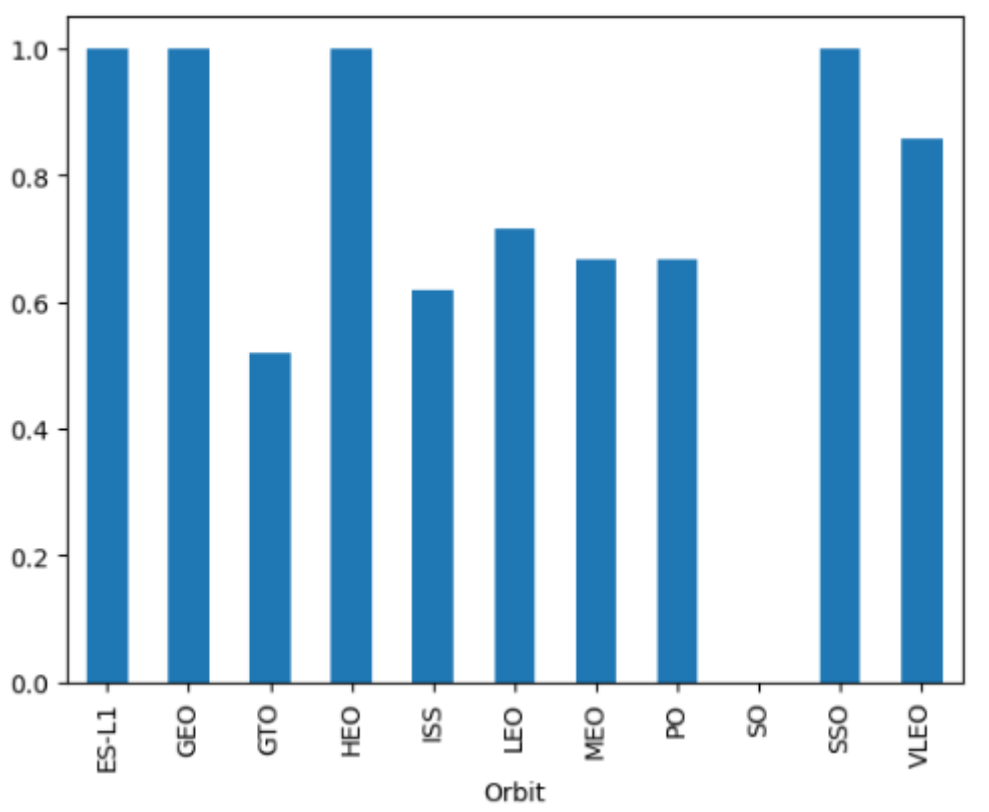
Payload vs. Launch Site

- For the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000)



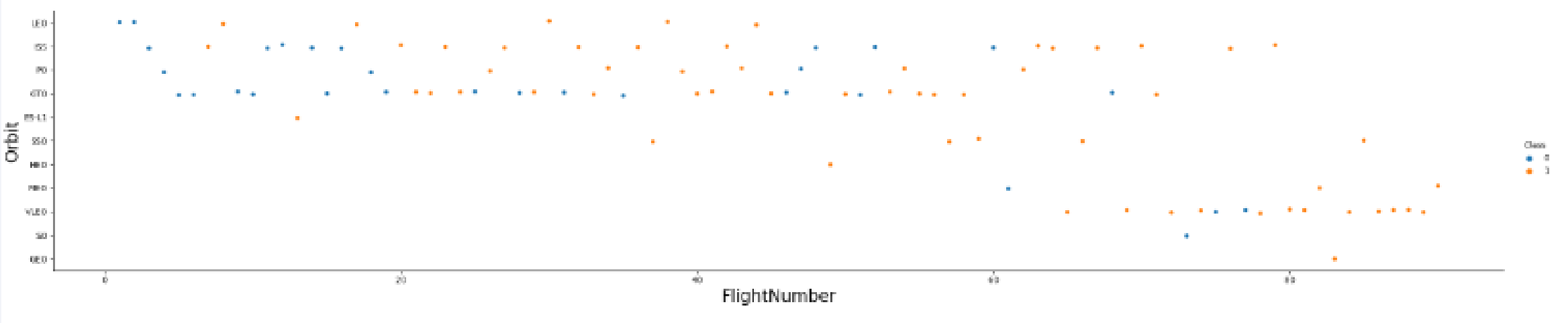
Success Rate vs. Orbit Type

- Most Successful:
  - ES-L1
  - GEO
  - HEO
  - SSO



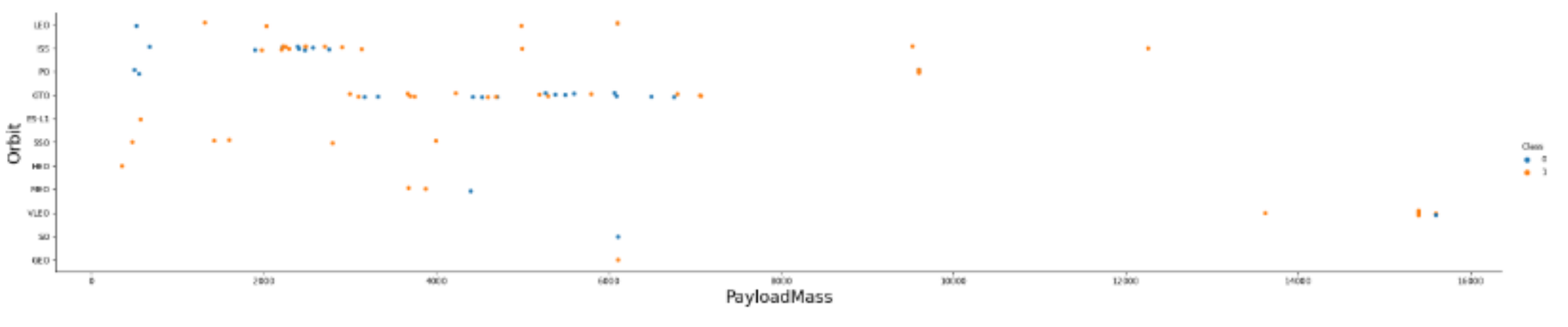
Flight Number vs. Orbit Type

- Success has improved in every orbit



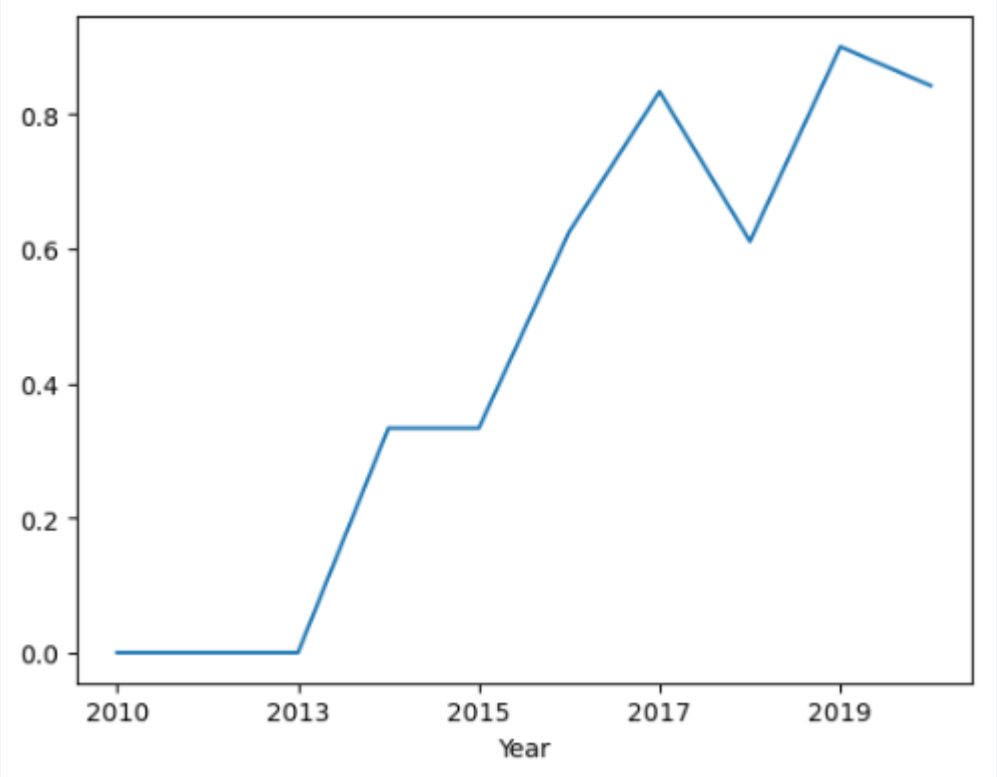
Payload vs. Orbit Type

- Very little relationship between payload & success.



Launch Success Yearly Trend

- Vast improvement with a dip in 2018



All Launch Site Names

- There are 4 launch sites with details below:

Out[20]:

	Launch Site	Lat	Long
0	CCAFS LC-40	28.562302	-80.577356
1	CCAFS SLC-40	28.563197	-80.576820
2	KSC LC-39A	28.573255	-80.646895
3	VAFB SLC-4E	34.632834	-120.610745

Launch Site Names Begin with 'CCA'

- 5 ‘CCA%’ Records

Out[9]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcom
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachut
2010-08-12	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 (parachut
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attemp
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attemp
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attemp

Total Payload Mass

- Total Payload in kg for NASA booster. Sum of PAYLOAD LIKE '%CRS%'.

Out[10]:

TOTAL_PAYLOAD
111268

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1
- Average booster WHERE BOOSTER_VERSION = 'F9 v1.1'

Out[11]:

AVG_PAYLOAD
2928.4

Total Number of Successful and Failure Mission Outcomes

- Total number of successful and failure mission outcomes
- SQL Query counting by total Mission_Outcome

Out[14]:

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

Boosters Carried Maximum Payload

- Names of the booster which have carried the maximum payload mass
- Obtained using “MAX”

Out[15]:

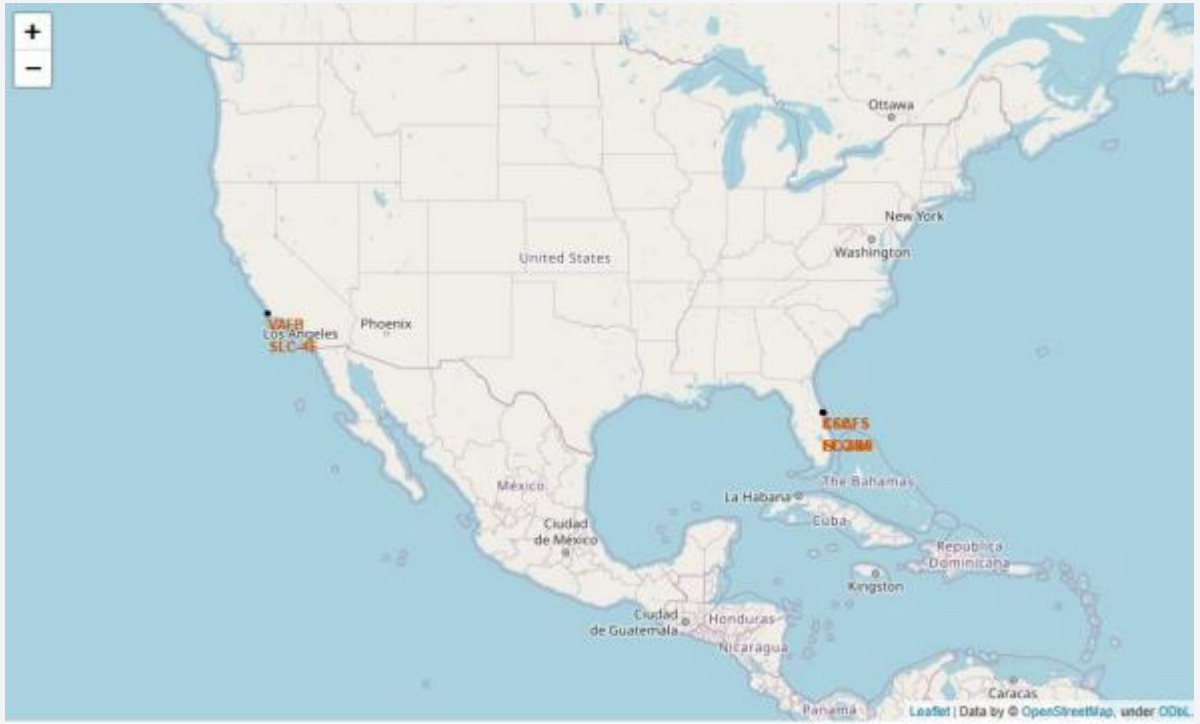
Booster_Version
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
F9 B5 B1051.3
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

Section 3

Launch Sites Proximities Analysis

Launch Sites

- Launch sites are by the ocean

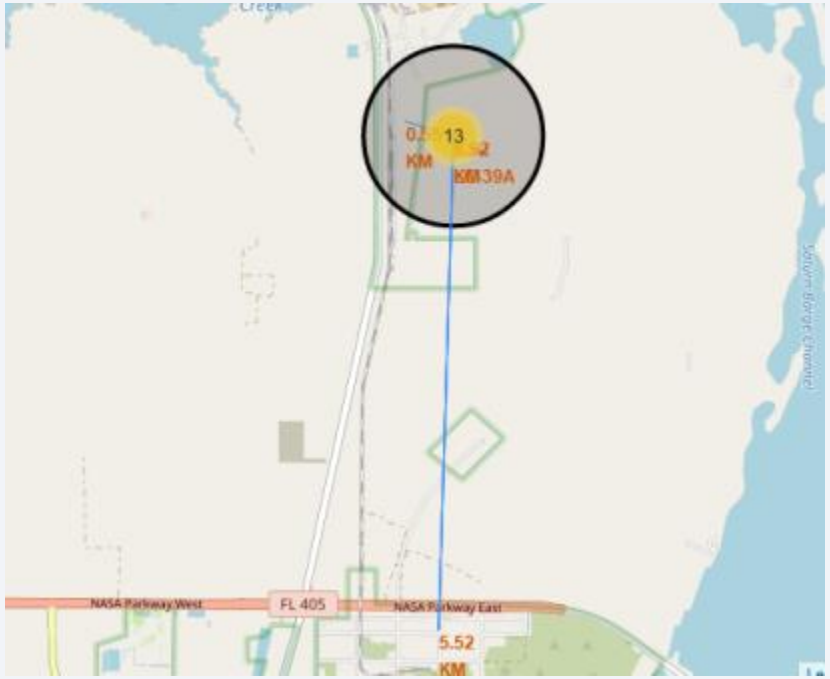


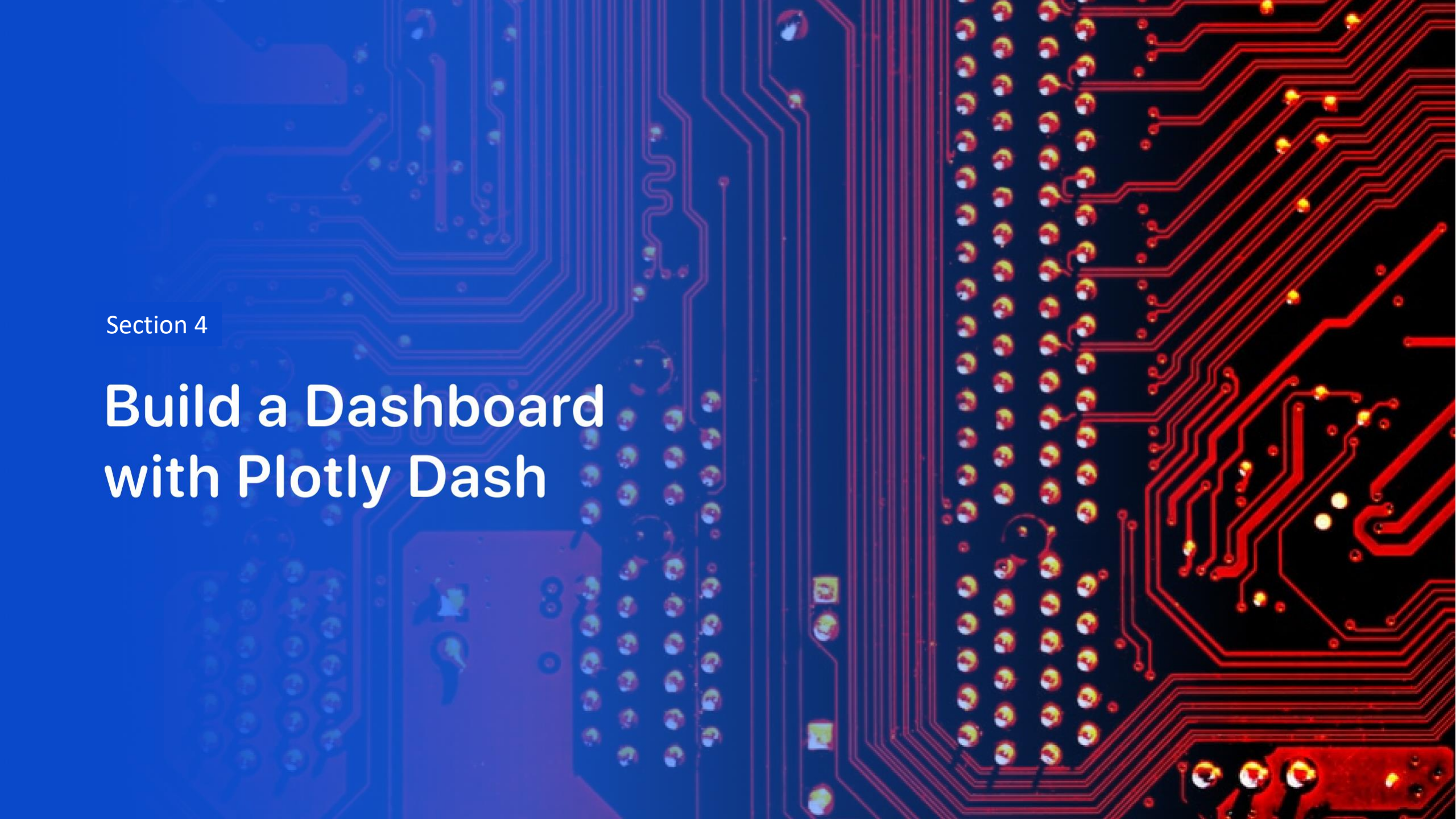
Success vs. Failure

- Successes and failures marked Green or Red

Logistics

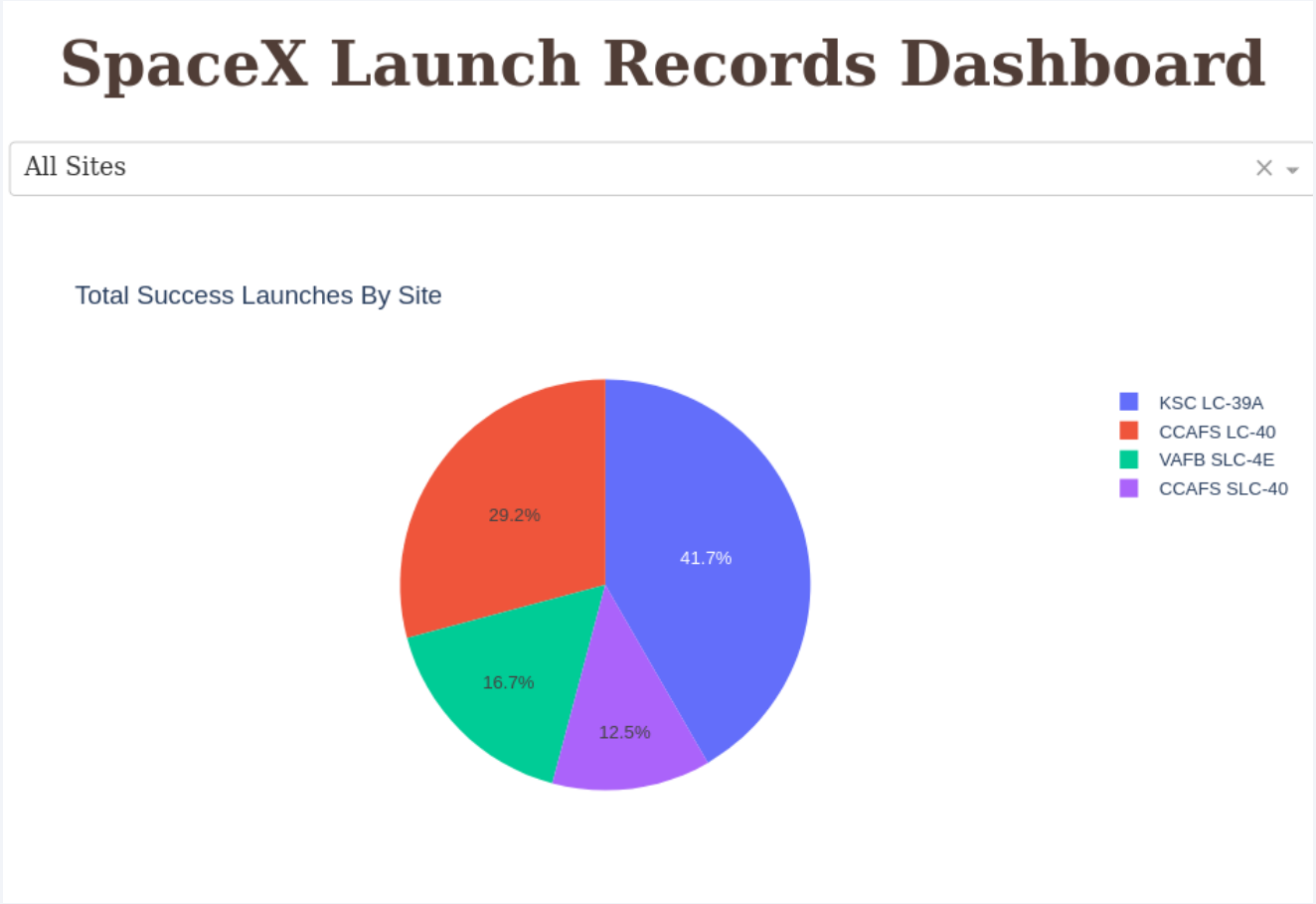
- This site is in a good location. Next to a railroad and a body of water.




Section 4
Build a Dashboard with Plotly Dash

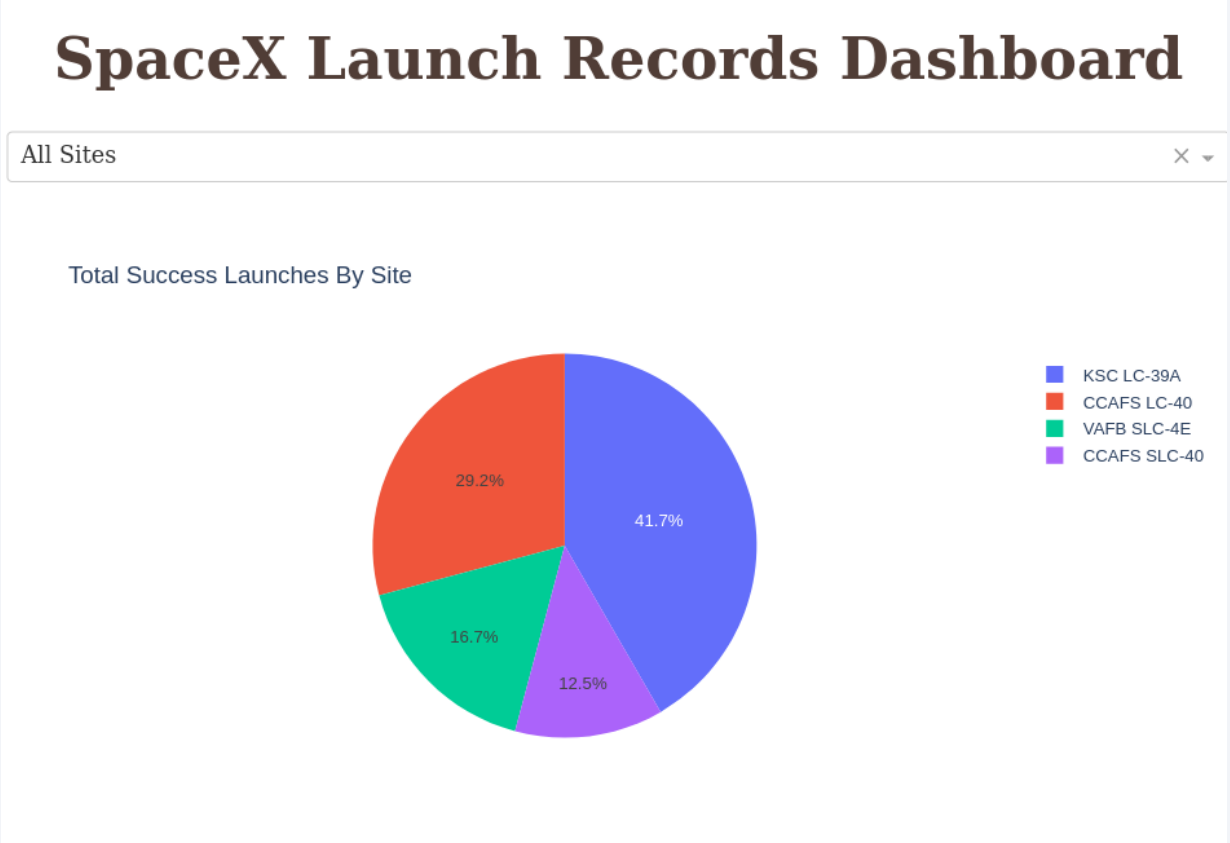
Launch Success

- Sites have a large impact on success.



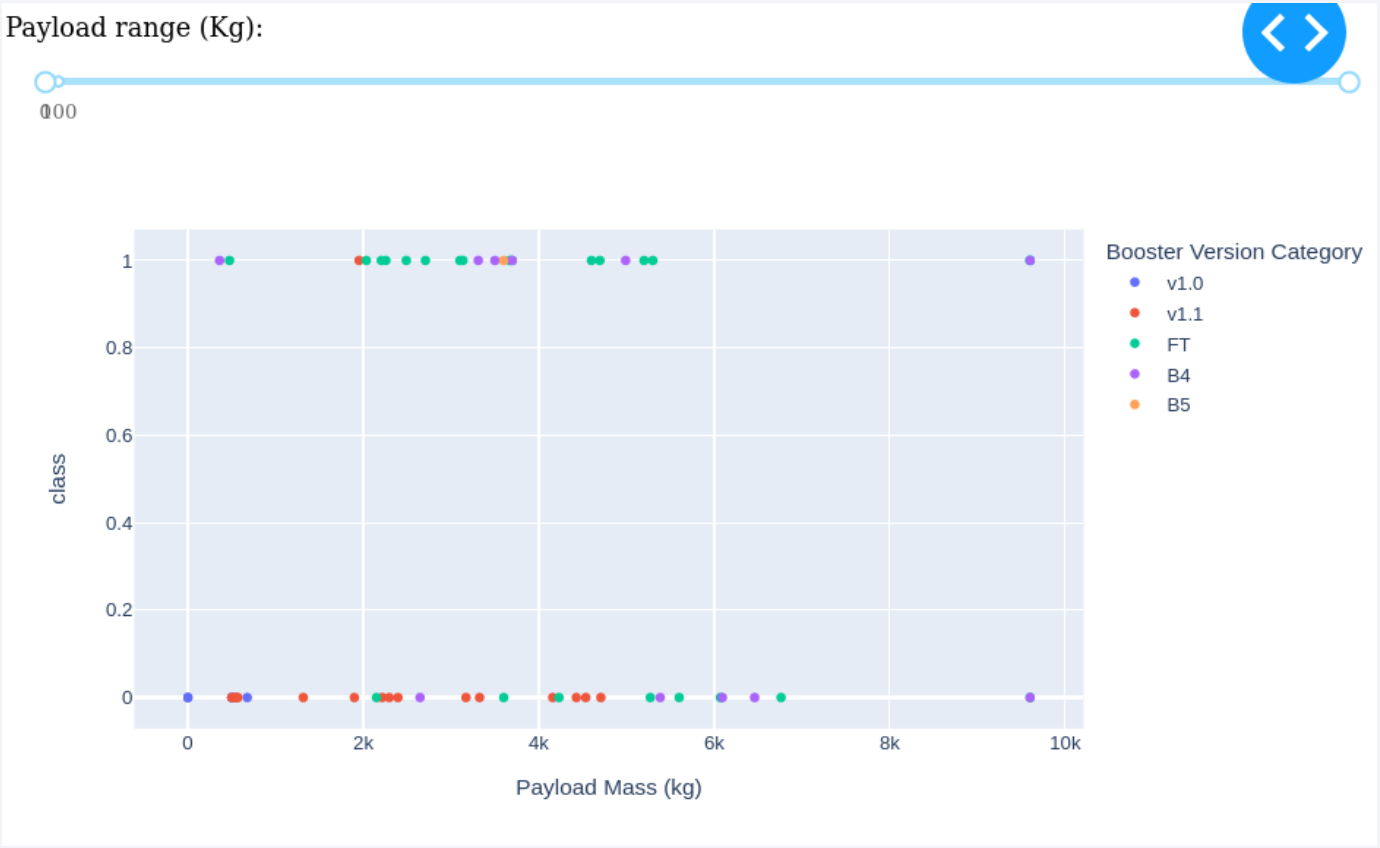
CCAFS LC-40

- This site had the highest success rate at 73.1%



Payload vs. Launch Outcome

- Payloads under 6k kg are typically the best performers.

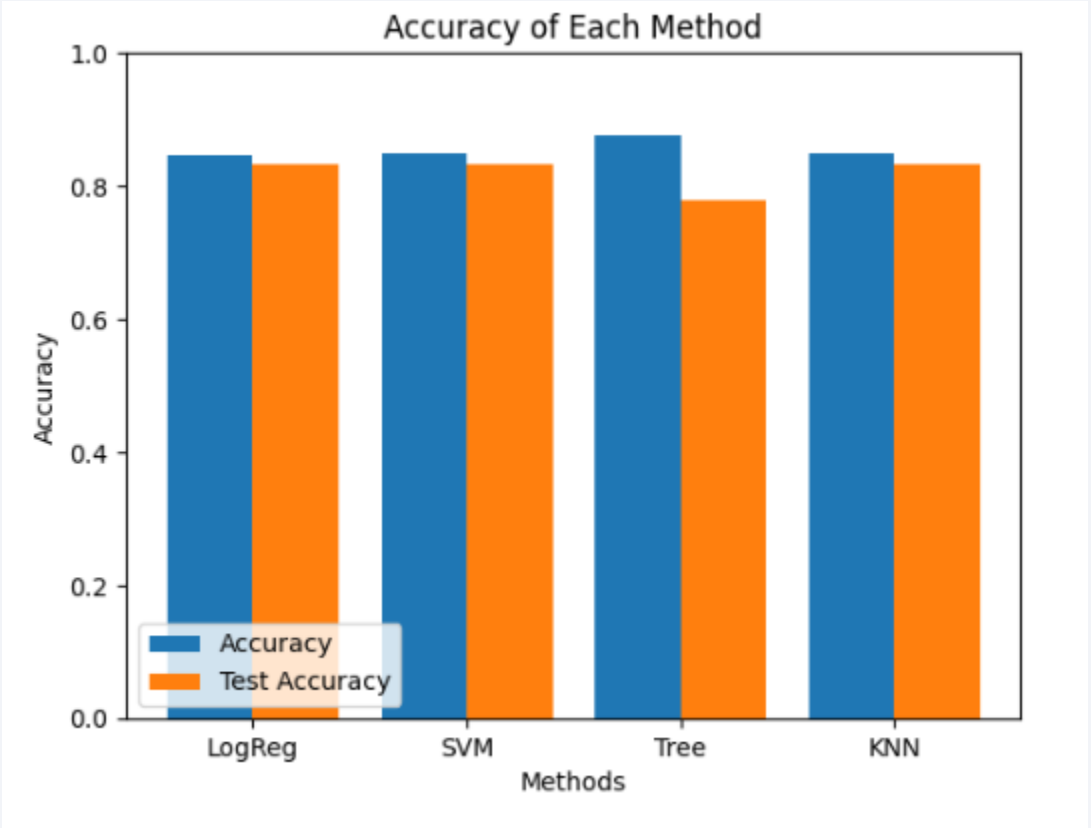


Section 5

Predictive Analysis (Classification)

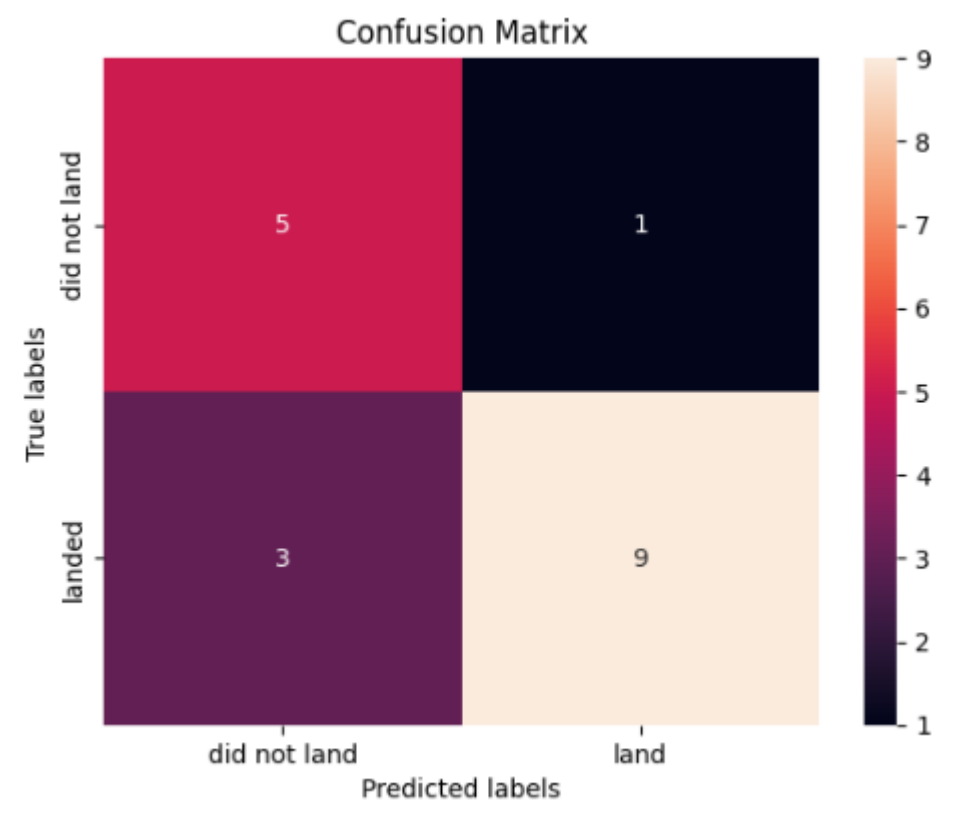
Classification Accuracy

- Decision Tree Classifiers were best for this exercise at 87% accuracy.



Confusion Matrix

- Confusion matrix of Decision Tree Classifier proves its accuracy by showing the big numbers of true positive and true negative.



Conclusions

- The best launch site is KSC LC-39A
- Decision Tree Classifier can be used to predict successful landings
- Launches above 6,000kg are least risky

Thank you!