IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Kulwinder Bhamra
31/05/2025

Outline

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

Executive Summary

- Methodologies
  - Data collection from the Space X rest API as well as web scraping Falcon 9 launch records
  - Data Wrangling
  - Exploratory data analysis using SQL and visualization techniques
  - Interactive visual analysis using Folium and Plotly Dash.
  - Predictive analysis using logistic regression, decision tree, support vector machine, and KNN
- Results
  - The Decision Tree classifier performed better than the other machine learning algorithms

Introduction

In this capstone project, I will take on the role of a data scientist working for a new rocket company called Space Y, which aims to compete with an existing company called Space X. I will need to investigate whether Space X can reuse its first stage, as this will help determine the cost of a launch. My role will involve gathering information about Space X and creating dashboards as well as training machine models to predict if Space X will reuse the first stage.

Section 1

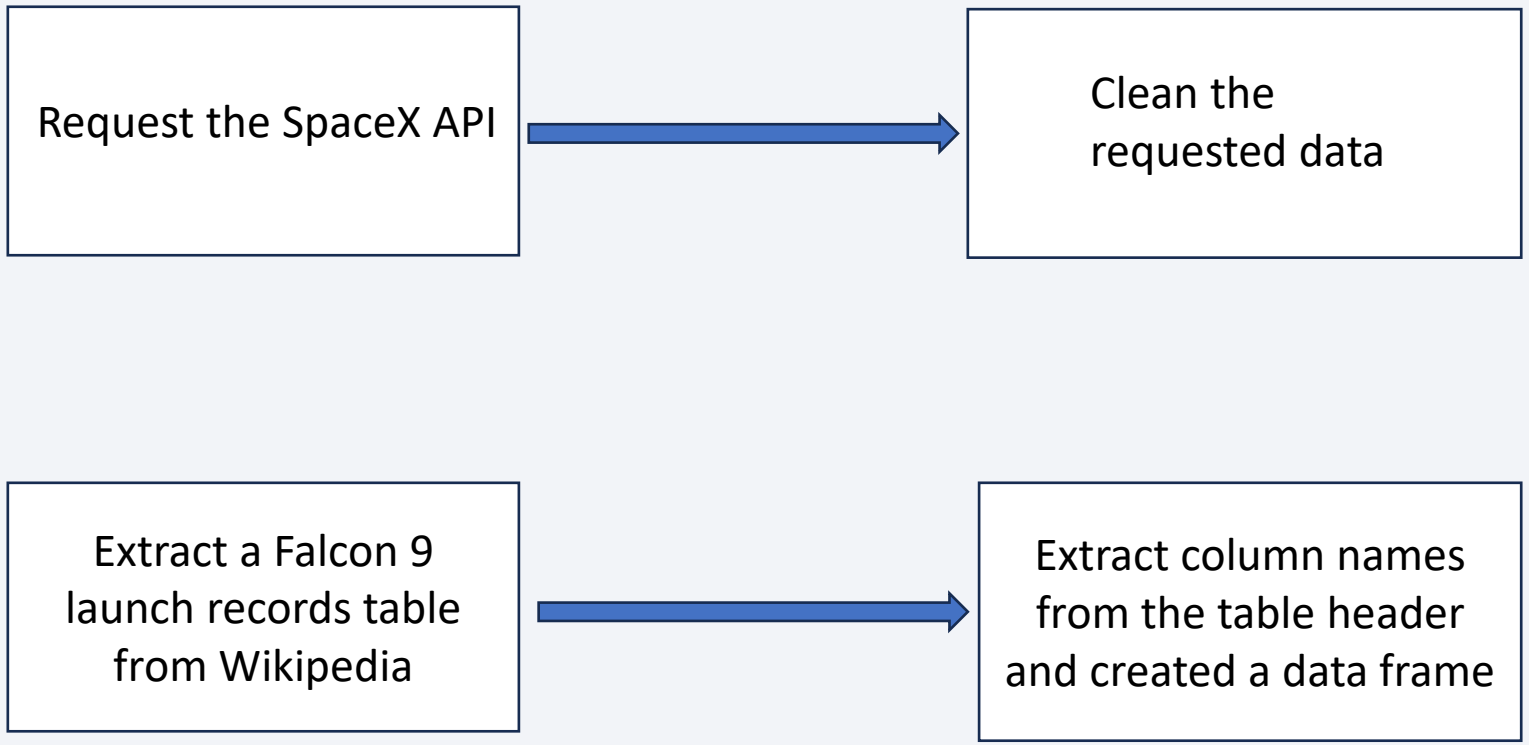
Methodology

Methodology

Executive Summary

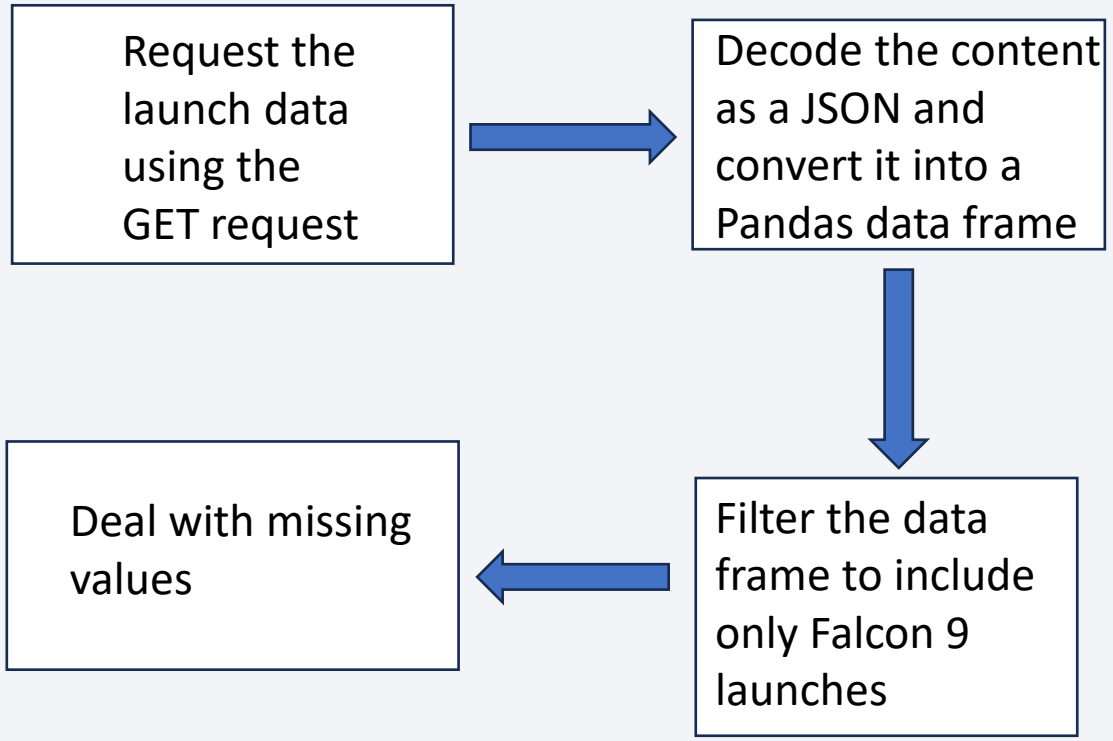
- Data collection methodology:
  - Data was collected from an API and scraped from a website.
- Perform data wrangling
  - The data collected from the API was in the form of a JSON object which was then converted to a data frame.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - The machine learning algorithms used were tuned to find the best hyperparameter and the corresponding best score.

Data Collection



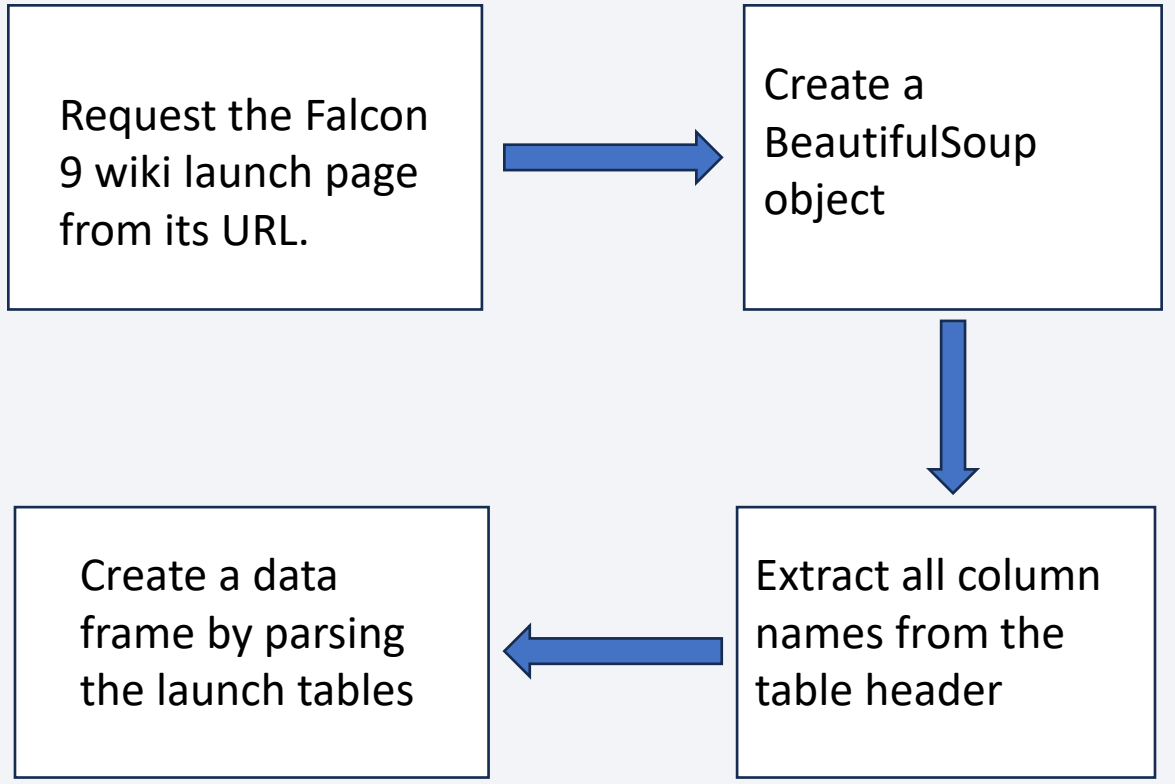
Data Collection – SpaceX API

- Applied-Data-Science-Capstone/jupyter-labs-spacex-data-collection-api.ipynb at main · kb1278/Applied-Data-Science-Capstone

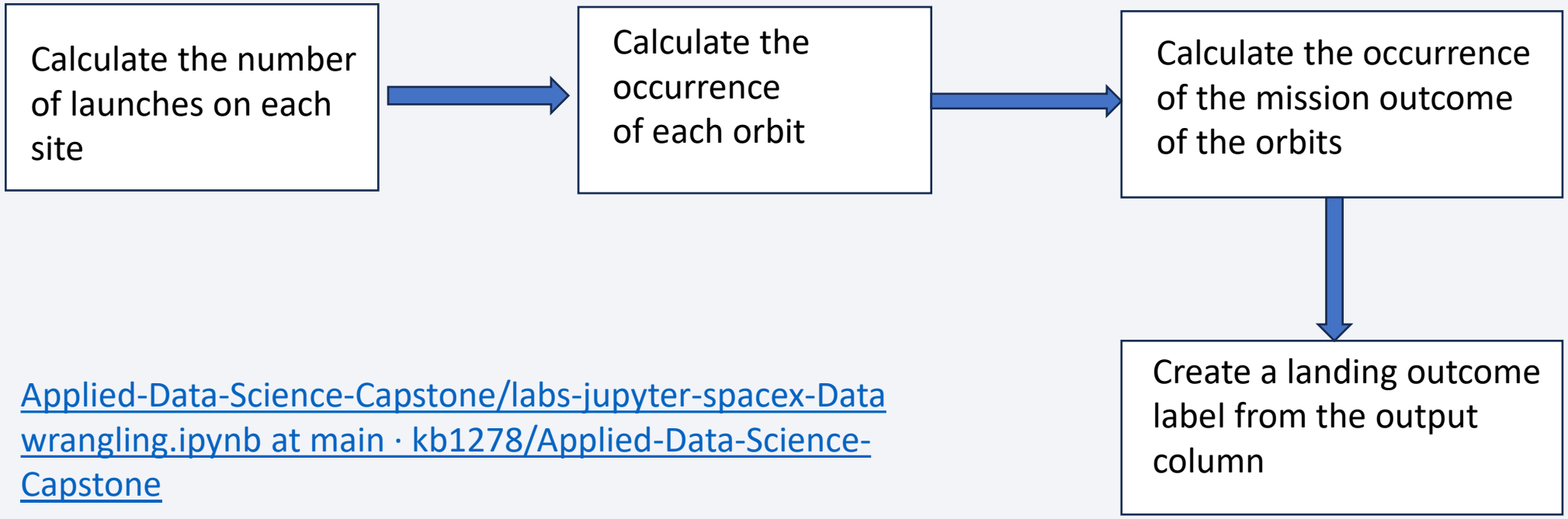


Data Collection - Scraping

[Applied-Data-Science-Capstone/jupyter-labs-webscraping.ipynb at main · kb1278/Applied-Data-Science-Capstone]



Data Wrangling



[Applied-Data-Science-Capstone/labs-jupyter-spacex-Data wrangling.ipynb at main · kb1278/Applied-Data-Science-Capstone]

EDA with Data Visualization

Here is a summary of the charts that I plotted:

- Catplot to visualize the relationship between Flight Number and Payload Mass
- Catplot to visualize the relationship between Flight Number and Launch Site
- Catplot to visualize the relationship between Payload Mass and Launch Site
- Bar chart to visualize the relationship between success rate of each orbit type
- Catplot to visualize the relationship between Flight Number and Orbit type
- Catplot to visualize the relationship between Payload Mass and Orbit type
- Line chart to visualize the launch success yearly trend

Applied-Data-Science-Capstone/jupyter-labs-eda-dataviz-v2.ipynb at main · kb1278/Applied-Data-Science-Capstone

EDA with SQL

The SQL queries I performed included:

- Displaying the names of the launch sites
- Displaying 5 records of launch sites that begin with a specific string sequence
- Displaying the total payload mass
- Displaying the average payload mass
- Listing the date of the first successful landing outcome
- Listing the names of boosters that meet a certain category
- Listing the total number of successful and failed missions
- Listing booster versions that have the maximum payload mass
- Displaying records that meet a specific criterion
- Ranking the count of landing outcomes that meet a specific criterion

Applied-Data-Science-Capstone/jupyter-labs-eda-sql-coursera_sqllite.ipynb at main · kb1278/Applied-Data-Science-Capstone

Build an Interactive Map with Folium

Markers were created to label the launch sites

Circles were created around launch sites

Polylines were drawn between the launch sites and other locations

Applied-Data-Science-Capstone/lab-jupyter-launch-site-location-v2.ipynb at main · kb1278/Applied-Data-Science-Capstone

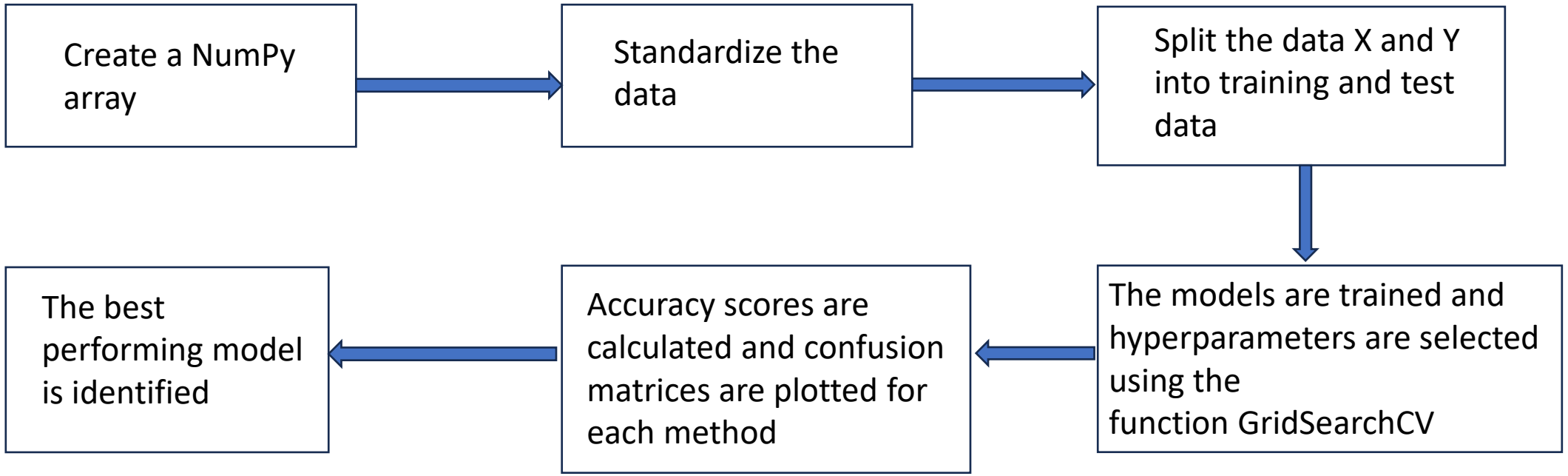
Build a Dashboard with Plotly Dash

- Pie charts were created to show the percentage of successful and failed missions for all launch sites, as well as for each site
- Scatter plots were created to show the correlation between payload and success for all sites, as well as individual sites
- A dropdown menu was added to select the required launch site
- A range slider was added to select different payload ranges

Applied-Data-Science-Capstone/spacex-dash-app (1).py at main · kb1278/Applied-Data-Science-Capstone

models are trained and hyperparameters are selected using the function GridSearchCV

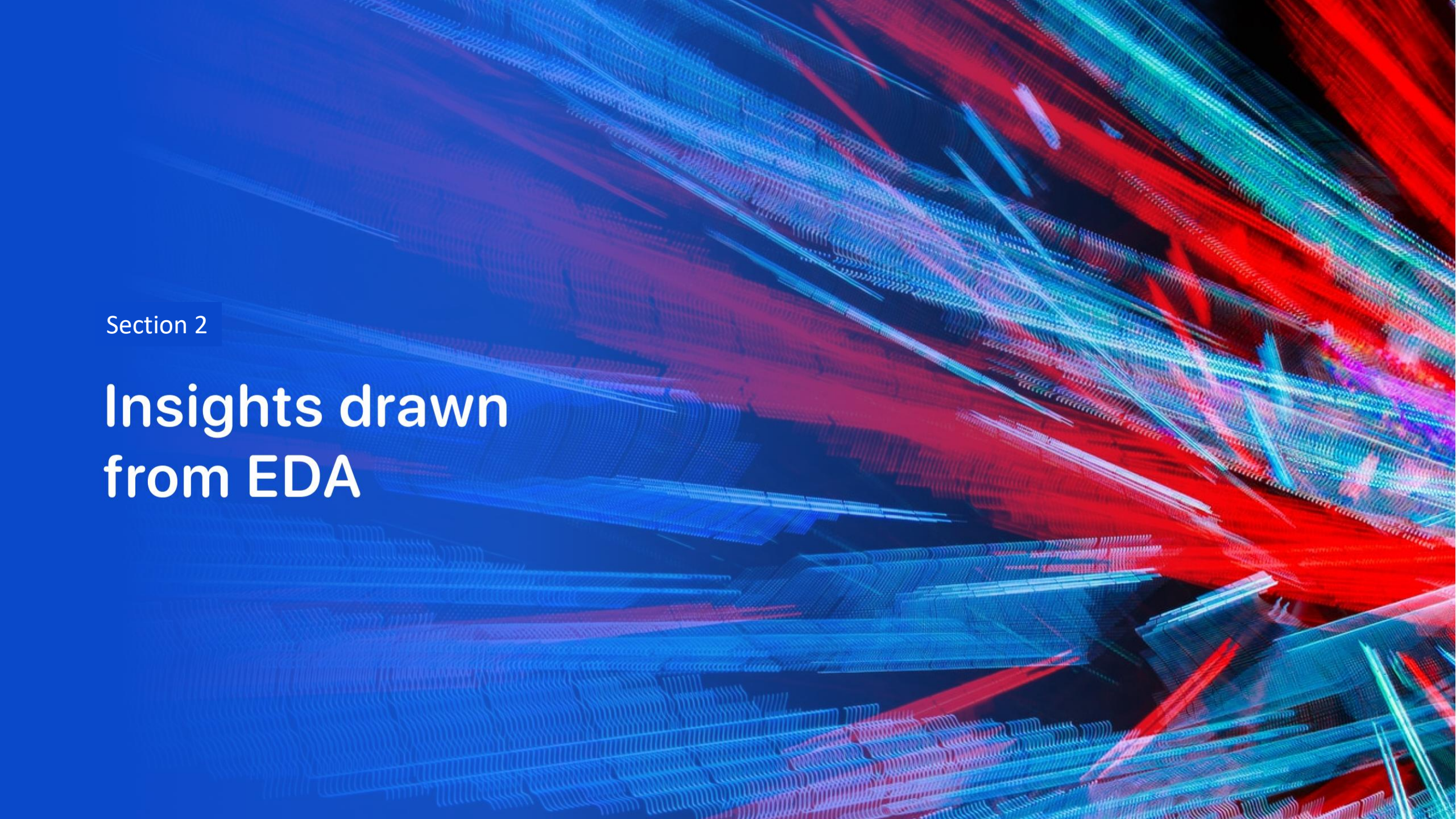
Predictive Analysis (Classification)



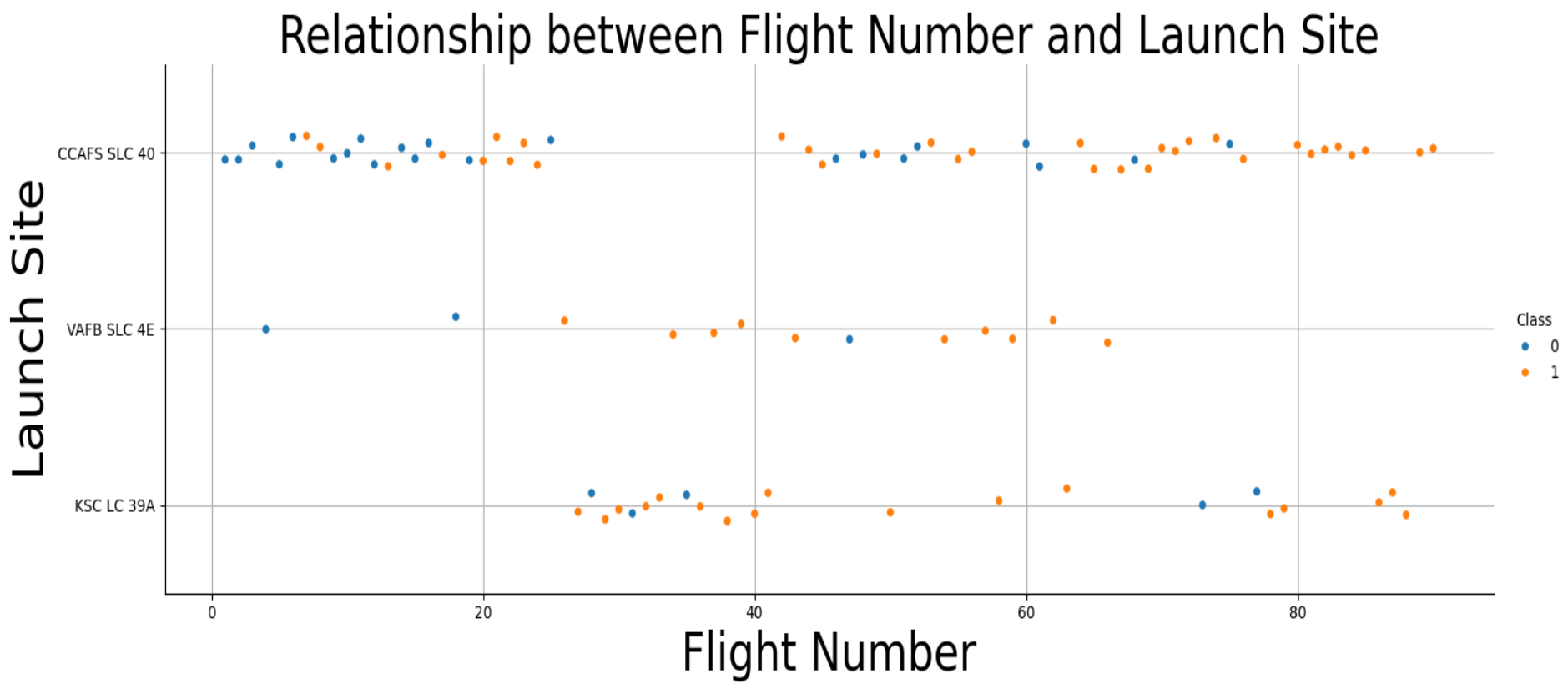
[Applied-Data-Science-Capstone/SpaceX-Machine-Learning-Prediction-Part-5-v1 (2).ipynb at main · kb1278/Applied-Data-Science-Capstone]

Results

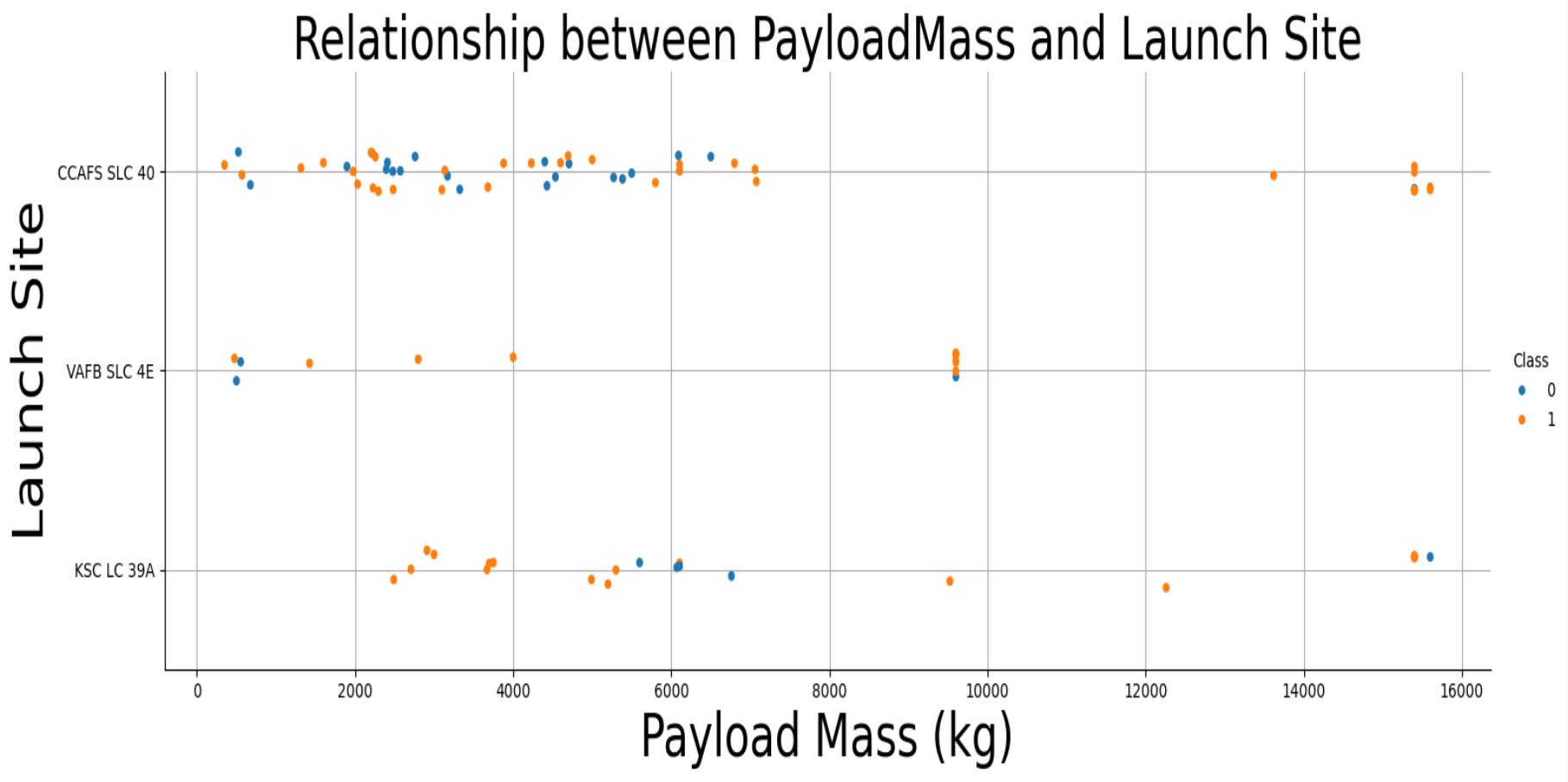
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results


Section 2
Insights drawn from EDA

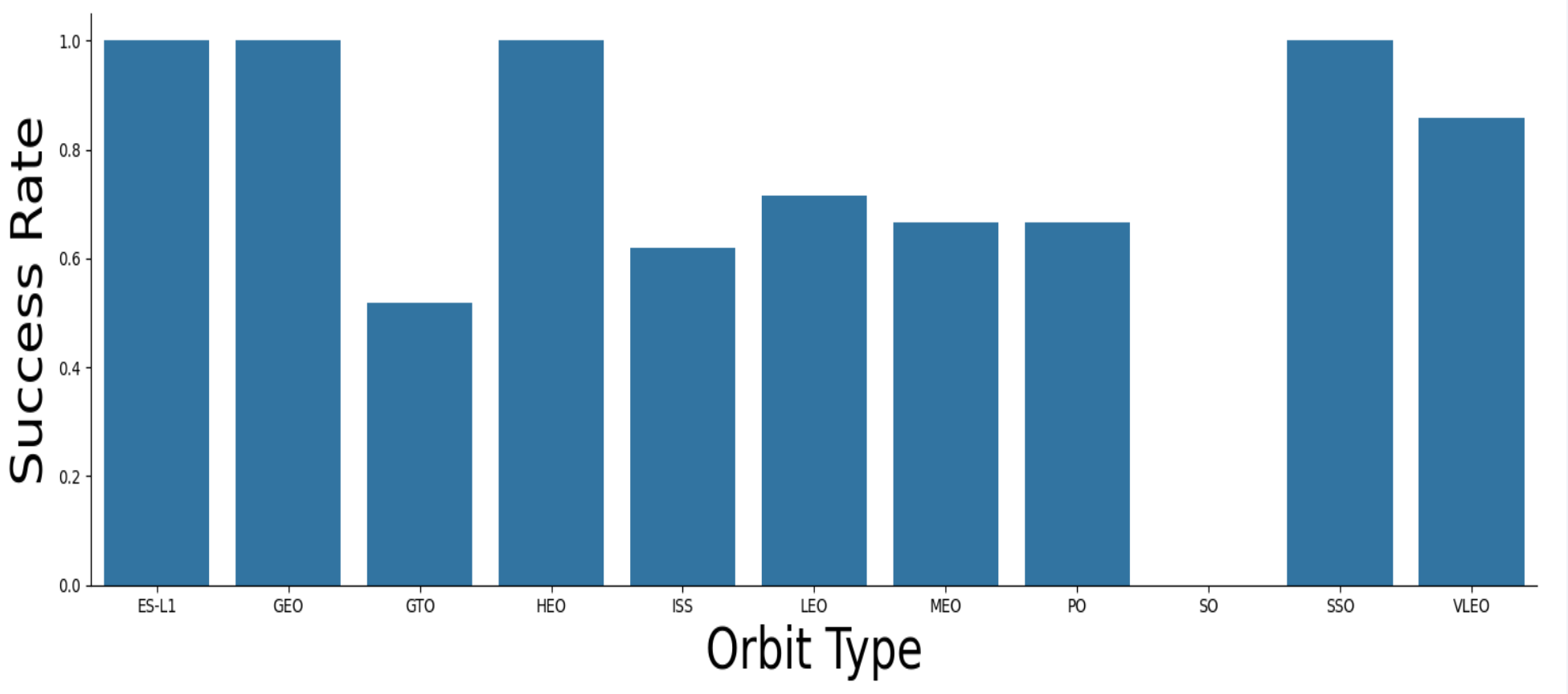
Flight Number vs. Launch Site



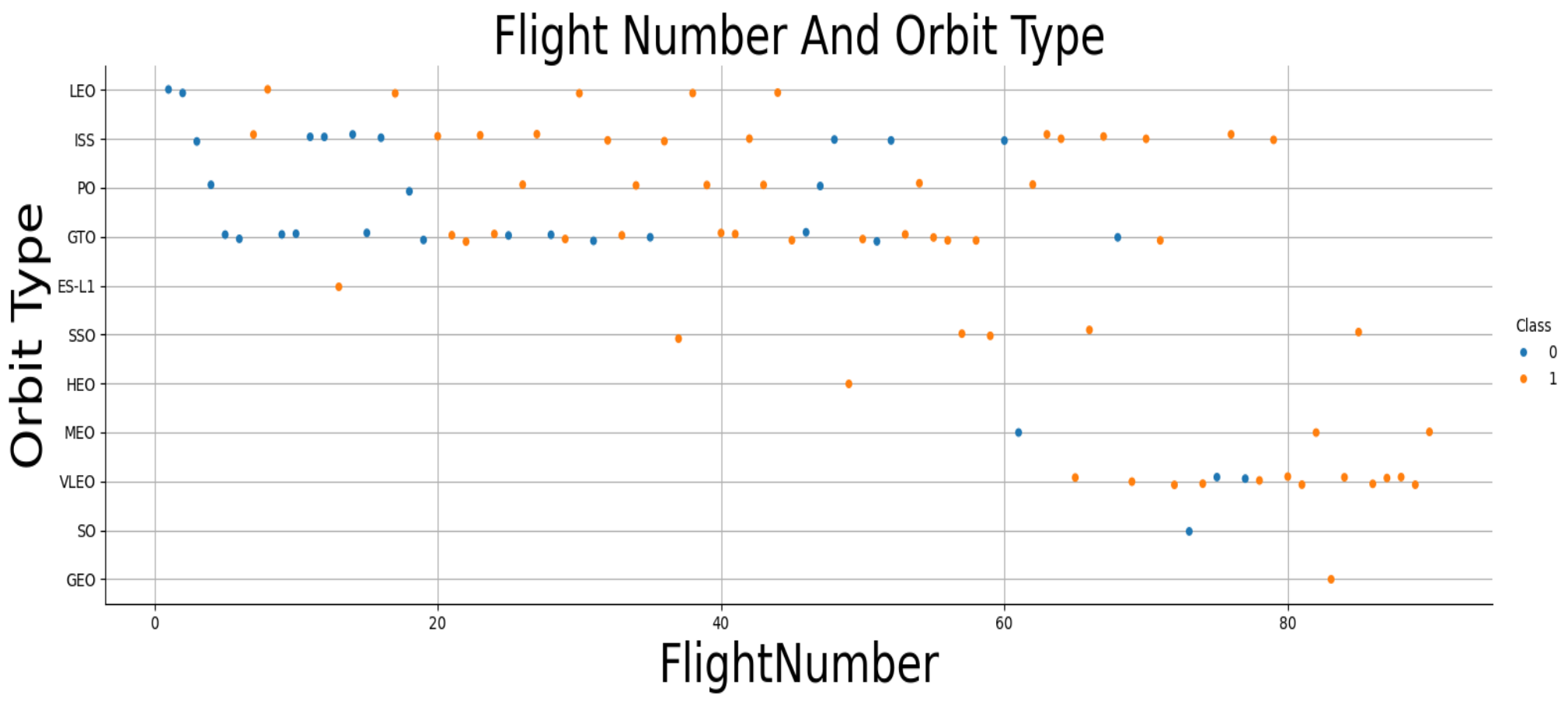
Payload vs. Launch Site



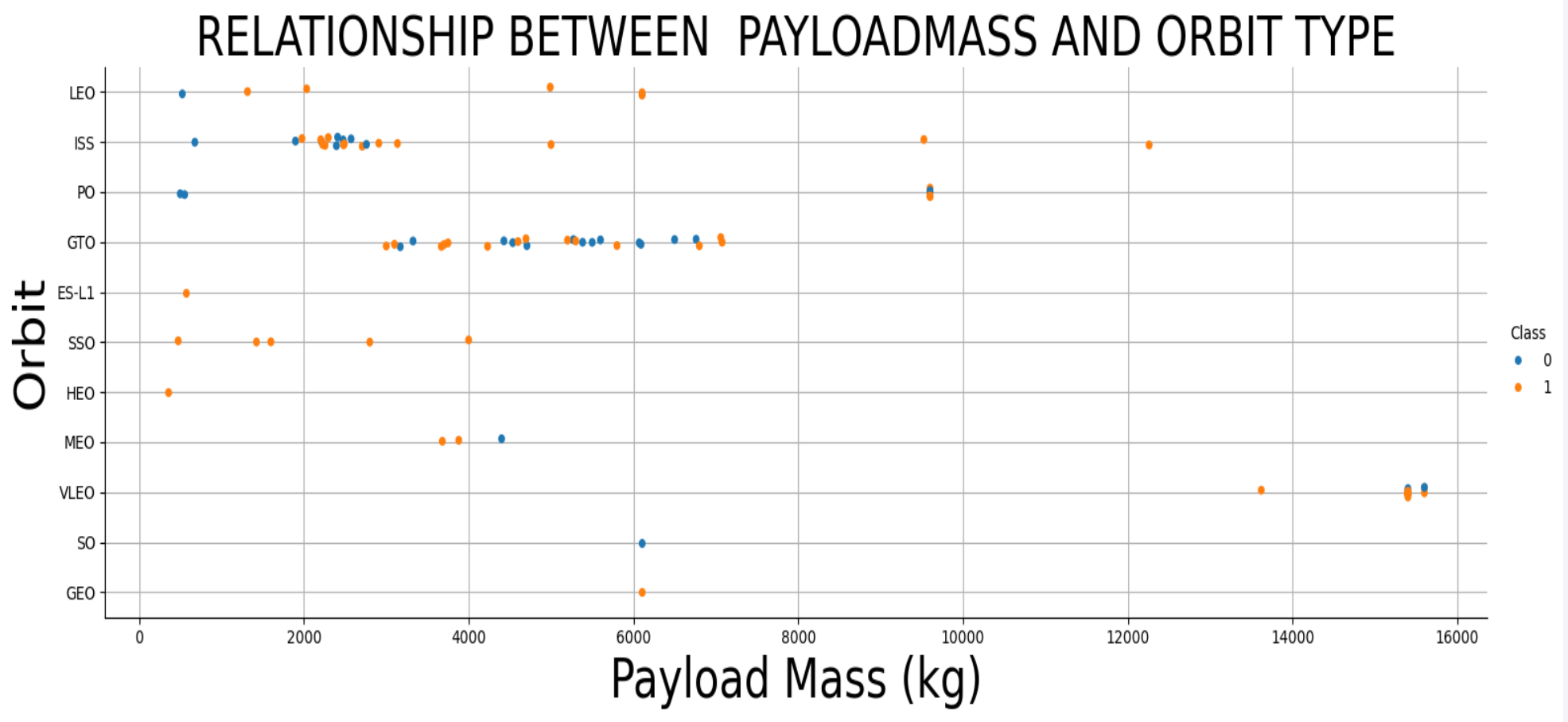
Success Rate vs. Orbit Type



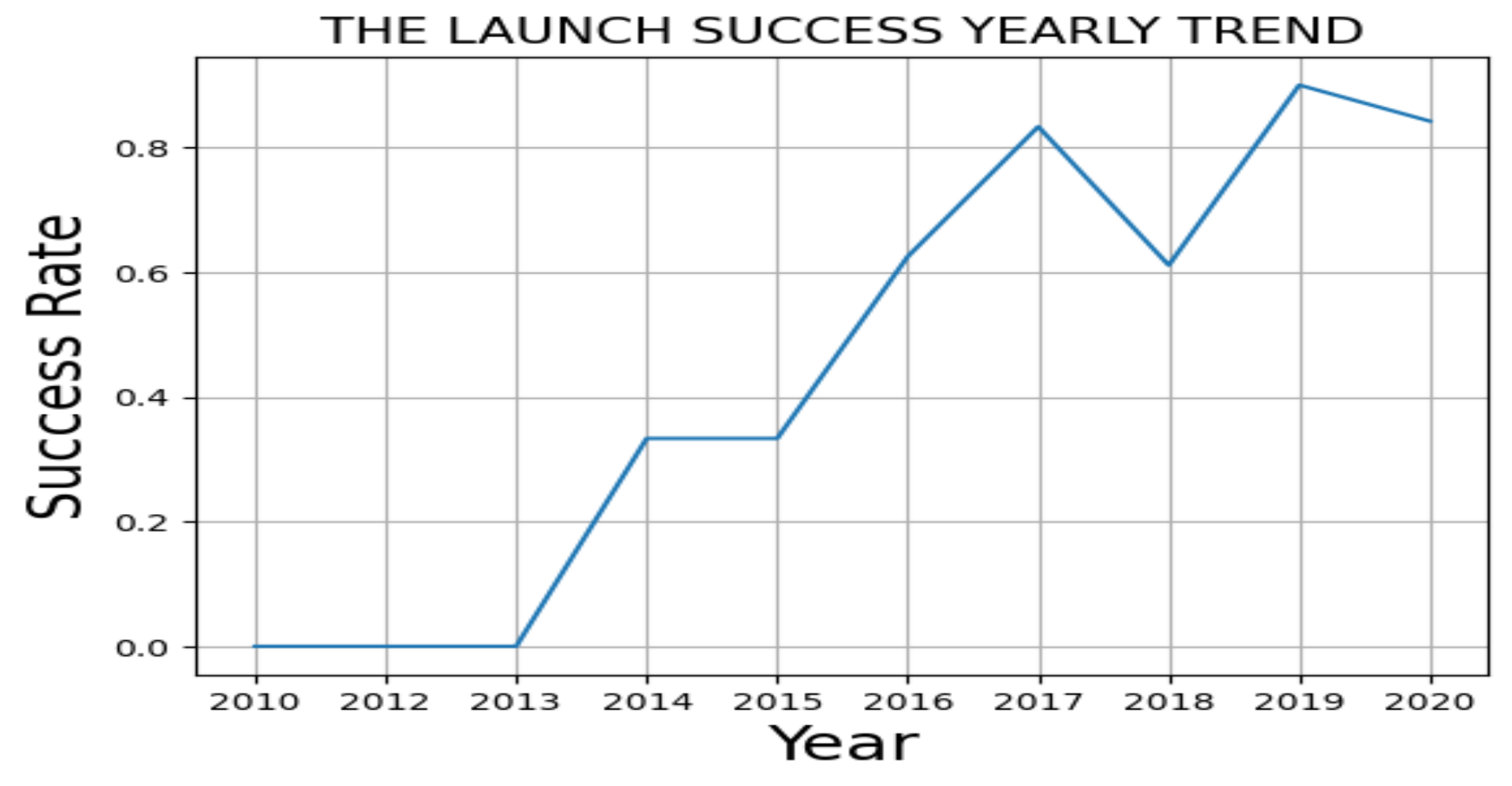
Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

%sql select distinct launch_site from SPACEXTABLE;

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

Launch Site Names Begin with 'CCA'

- %sql select * from SPACEXTABLE 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	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

- Calculate the total payload carried by boosters from NASA
- %sql select sum(payload_mass__kg_) as total_payload_mass from SPACEXTABLE where customer = 'NASA (CRS)';

total_payload_mass
45596

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
- %sql select avg(payload_mass__kg_) as average_payload_mass from SPACEXTABLE where booster_version like '%F9 v1.1%';

average_payload_mass
2534.6666666666665

First Successful Ground Landing Date

- Find the date of the first successful landing outcome on a ground pad
- %sql select min(date) as first_successful_landing from SPACEXTABLE where landing_outcome = 'Success (ground pad)';

first_successful_landing
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000
- %sql select booster_version from SPACEXTABLE where landing_outcome = 'Success (drone ship)' and payload_mass__kg_ between 4000 and 6000;

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

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes
- %sql select mission_outcome, count(*) as total_number from SPACEXTABLE group by mission_outcome;

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

Boosters Carried Maximum Payload

%sql select booster_version from SPACEXTABLE where payload_mass__kg_ = (select max(payload_mass__kg_) from SPACEXTABLE);

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

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
- %sql select date, booster_version, launch_site, landing_outcome from SPACEXTABLE where landing_outcome = 'Failure (drone ship)' and date <='2015-12-31';

Date	Booster_Version	Launch_Site	Landing_Outcome
2015-01-10	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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

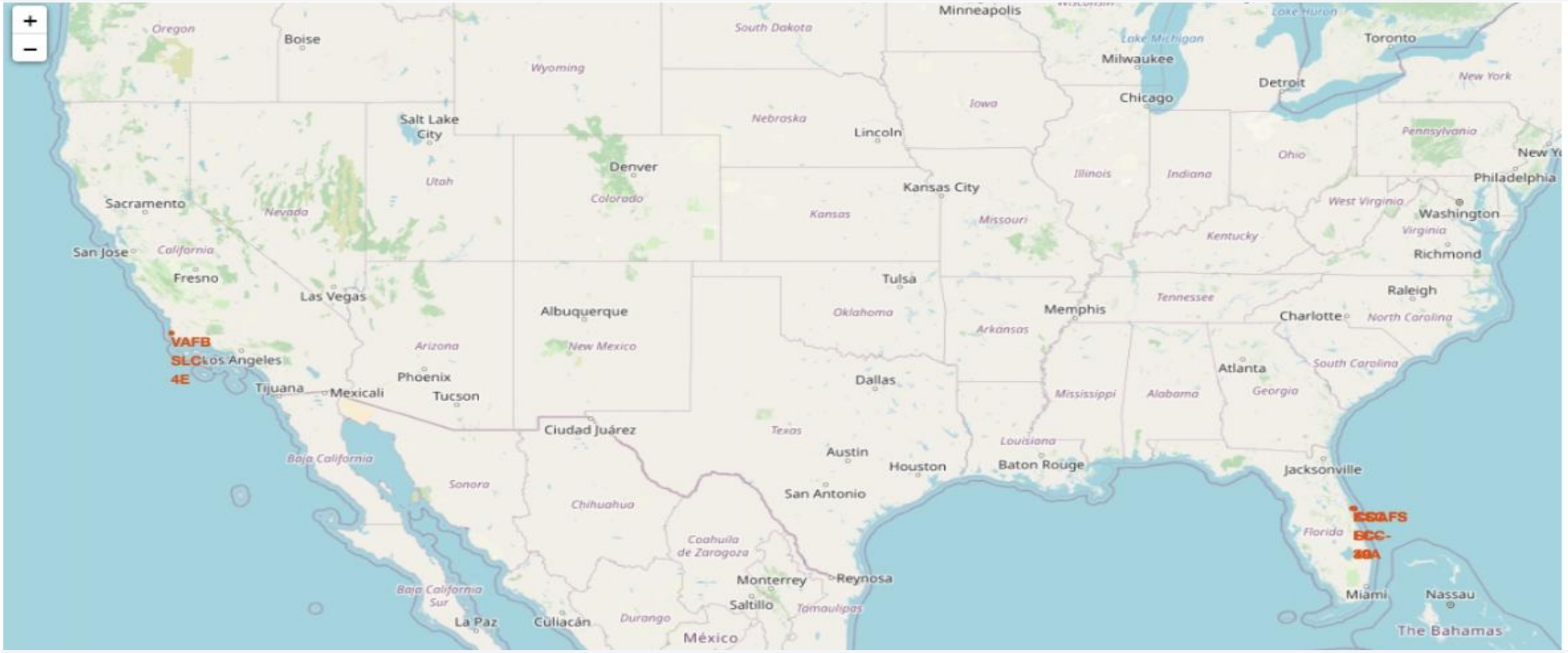
- %sql select landing_outcome, count(*) as count_outcomes from SPACEXTABLE
  where date between '2010-06-04' and '2017-03-20'
  group by landing_outcome
  order by count_outcomes desc;

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

Section 3

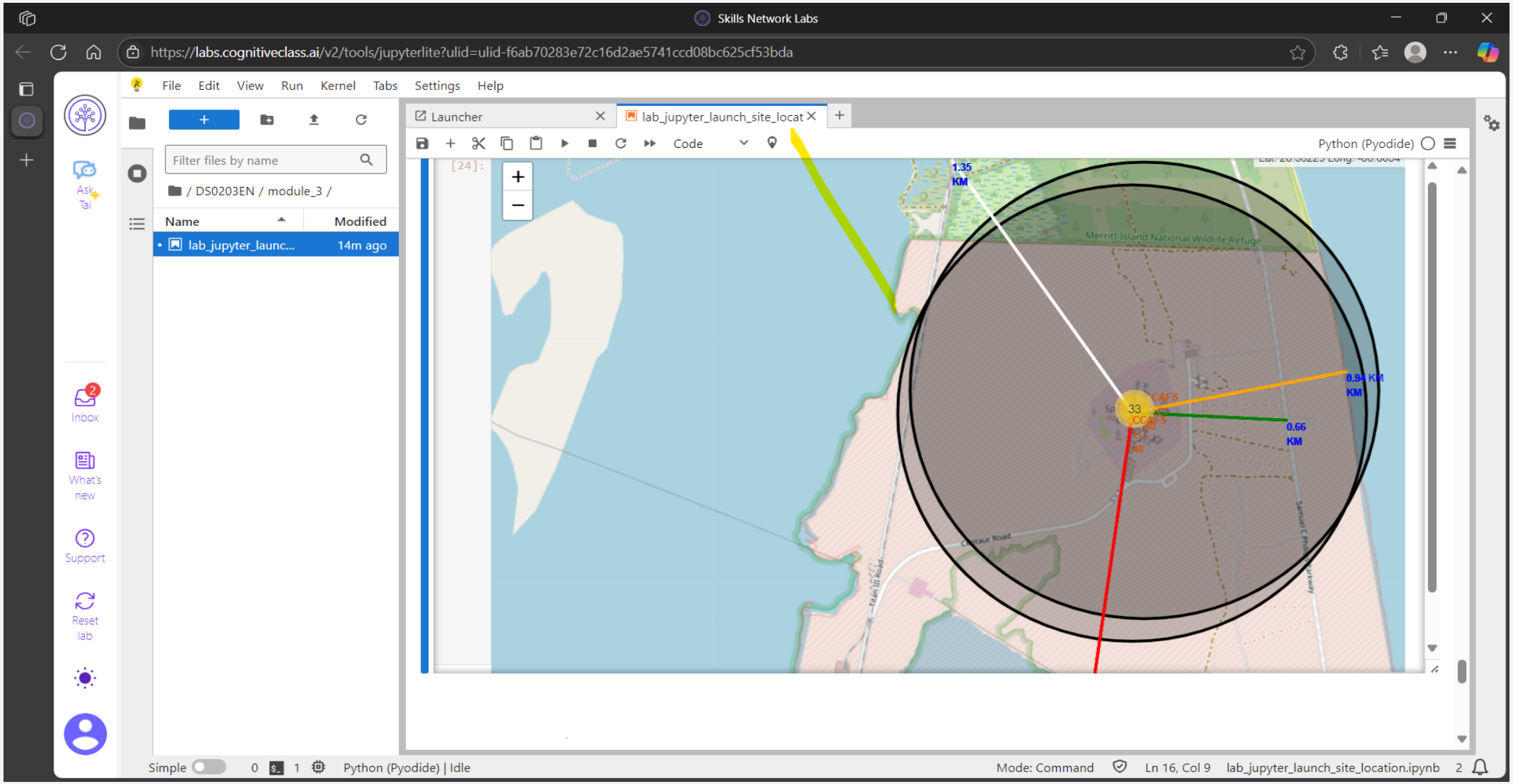
Launch Sites Proximities Analysis

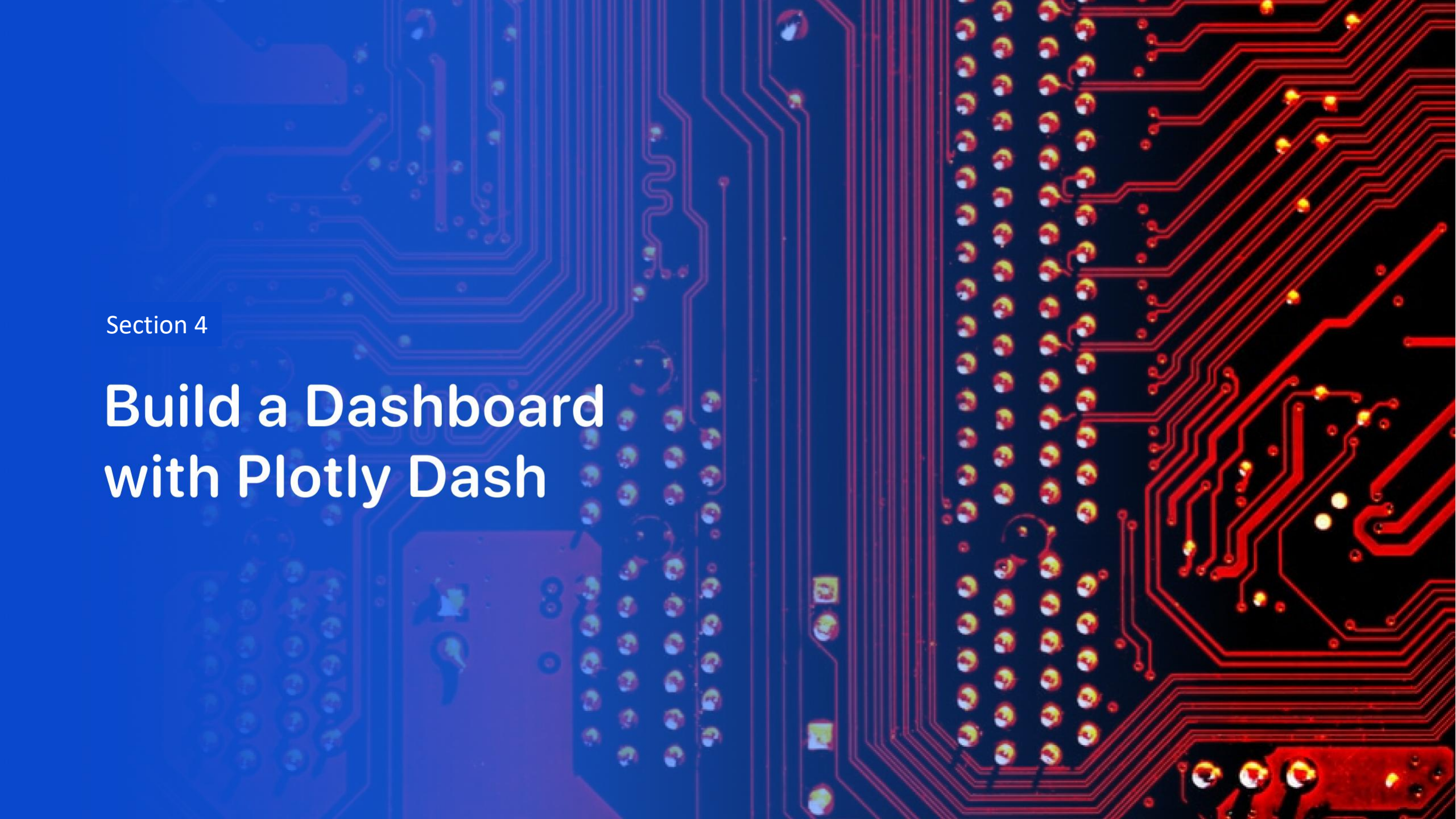
All Launch Sites



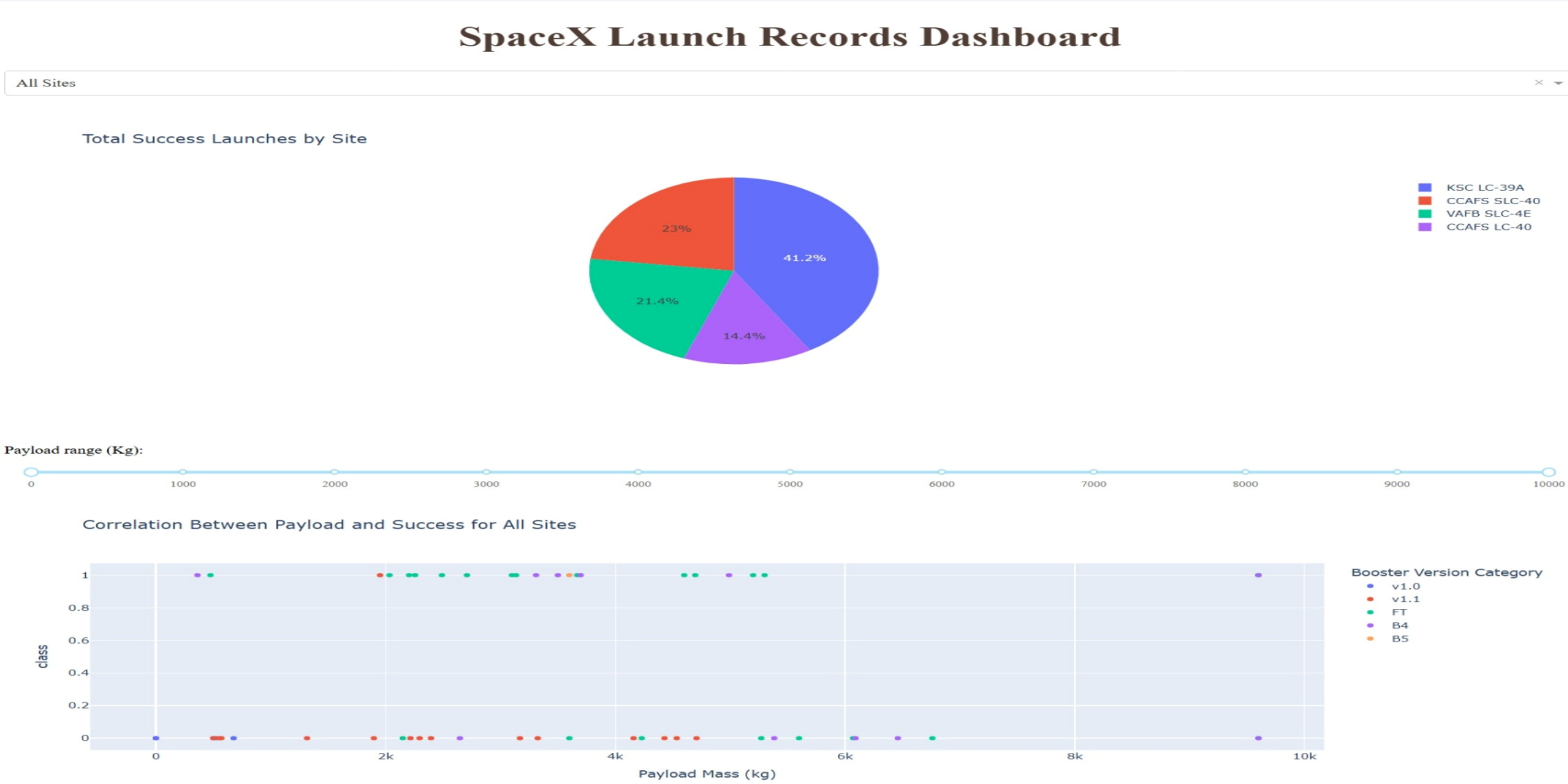
Site Showing Successful/Failed Launches

Launch Site And Its Proximities

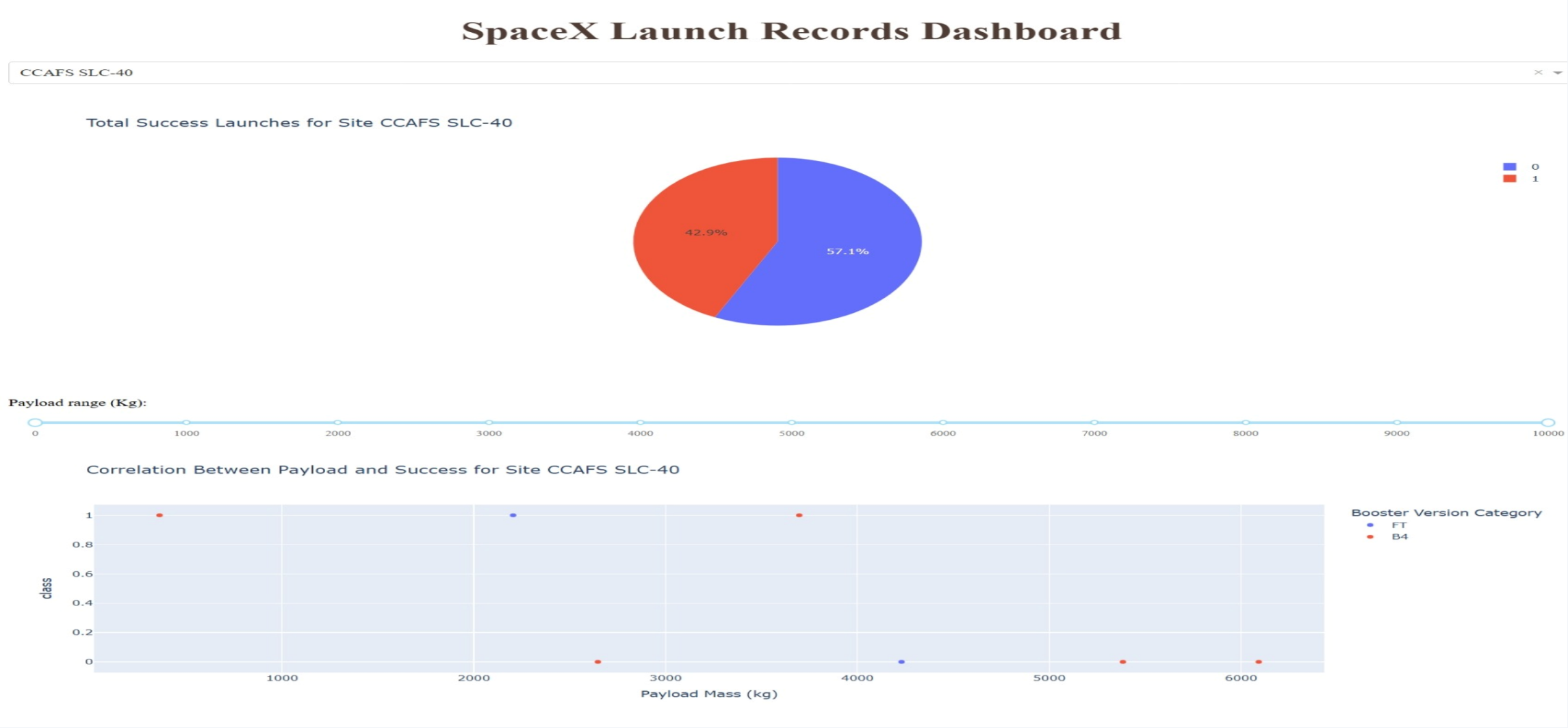



Section 4
Build a Dashboard with Plotly Dash

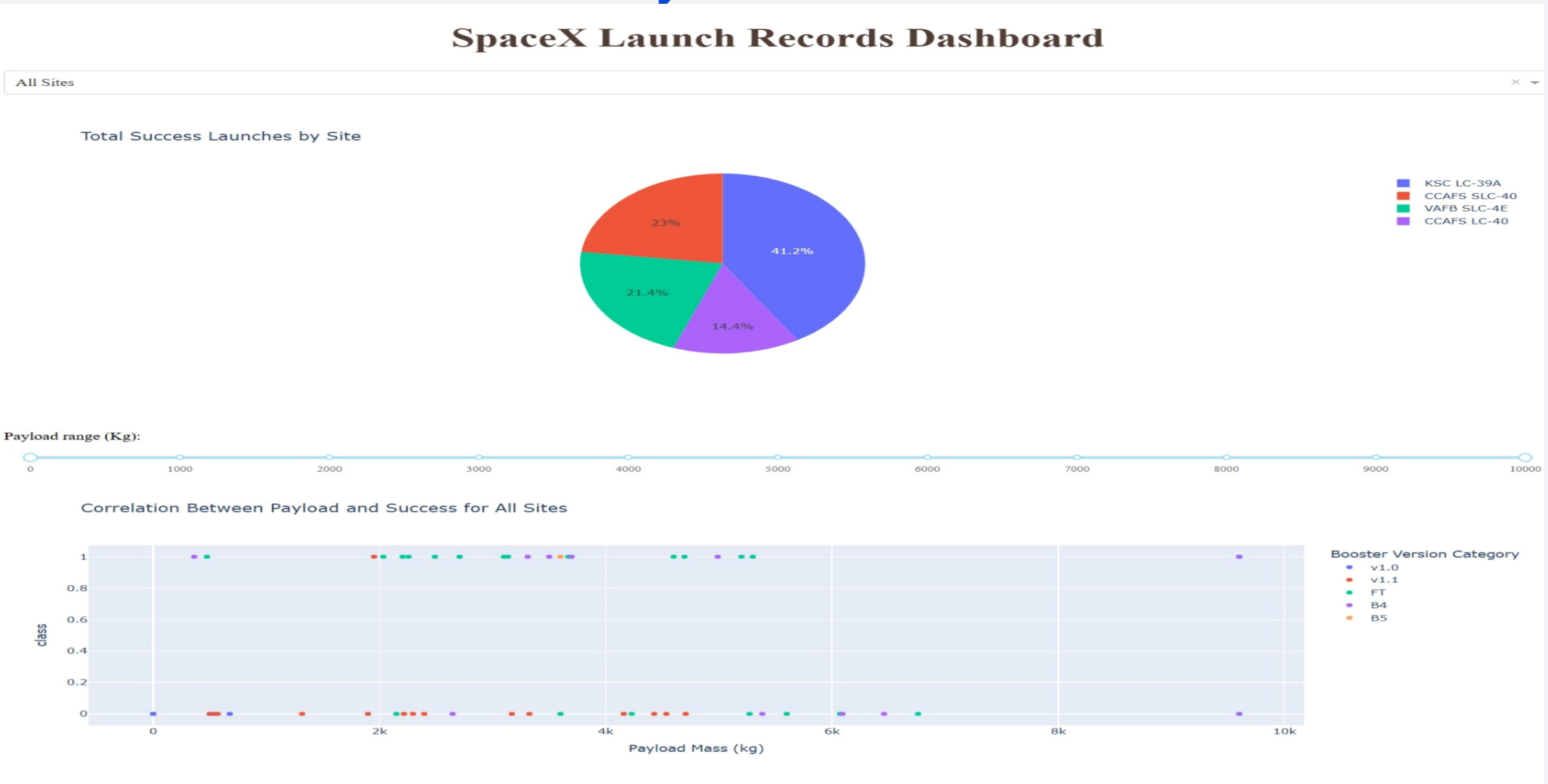
Total Success Launches By Site



Total Success Launches For Site CCAFS SLC-40



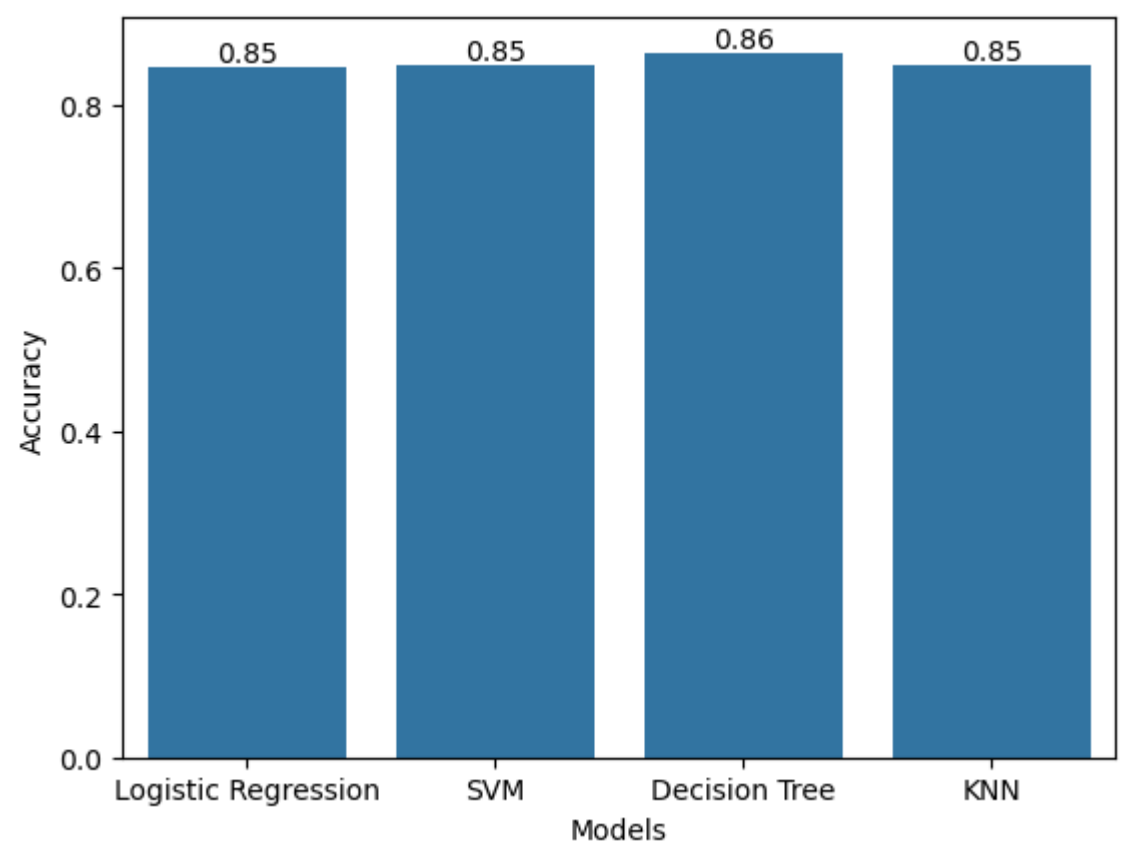
Correlation Between Payload And Success For All Sites



Section 5

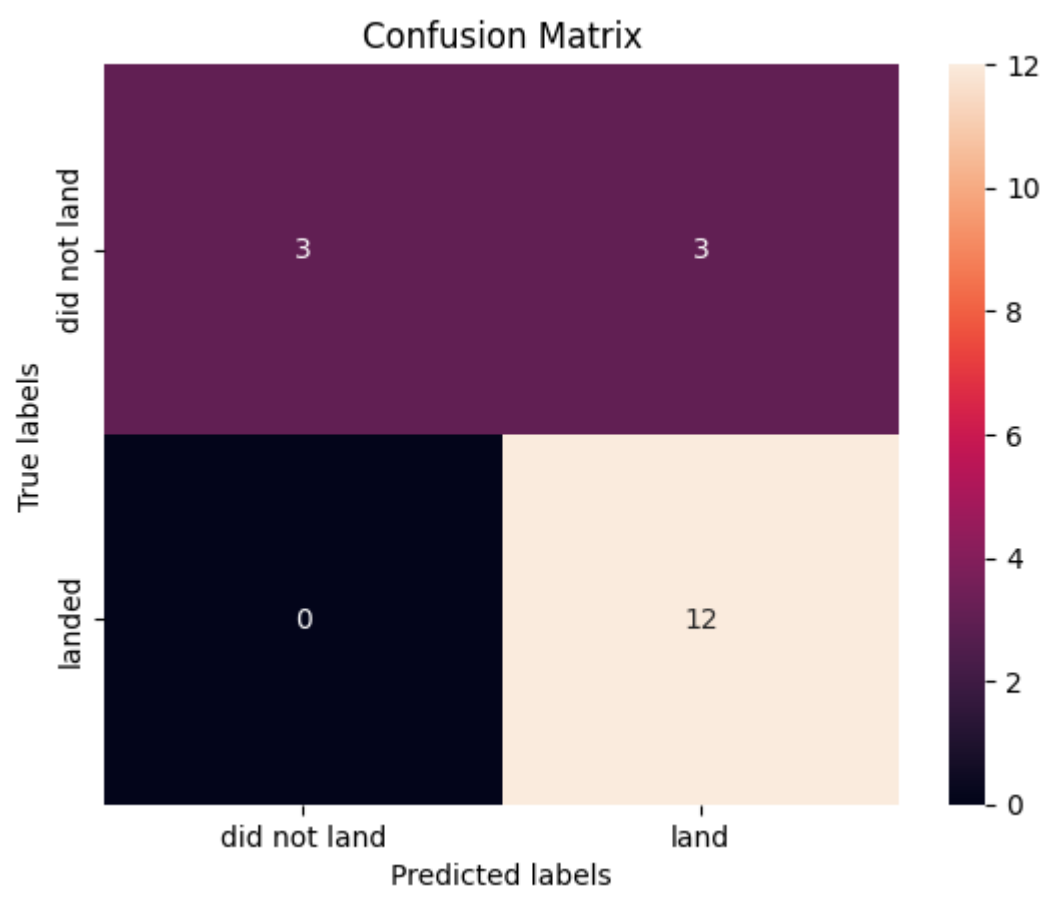
Predictive Analysis (Classification)

Classification Accuracy



	Models	Accuracy
0	Logistic Regression	0.846429
1	SVM	0.848214
2	Decision Tree	0.875000
3	KNN	0.848214

Confusion Matrix



True positive = 12
True negative = 3
False positive = 3
False negative = 0

Conclusions

- The decision tree classifier object gave the most accurate results
- The KSC LC 39A site had the most successful launches

Thank you!