IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Gbenga Amode
12th April, 2024

Outline

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

Executive Summary

Summary of methodologies

•Data was collected from the SpaceX public API and publically available data on Wikipedia. Data wrangling included extracting launch outcome information to serve as the dependent variable in the Machine Learning models.

• SQL queries and data visualizations (static plots, interactive maps, and an interactive dashboard) were created to discover insights about the data set and answer questions.

• Predictive analysis was pursued using Logistic Regression, SVM (Support Vector Machine), Decision Tree, and KNN (k-Nearest Neighbors) Machine Learning models.

Summary of all results

• Launch data include info about flight number, date of launch, payload mass, orbit type, launch site, mission outcome and other variables.

• Logistic Regression, SVM (Support Vector Machine), and KNN (k-Nearest Neighbors) all perform equally well for Machine Learning models on this dataset.

Introduction

- Project background and context
- SpaceX advertises Falcon 9 rocket launches on its website, with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage.
- If we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.
- Therefore, this project task is to predict if the first stage of the SpaceX Falcon 9 rocket will land successfully.

Section 1

Methodology

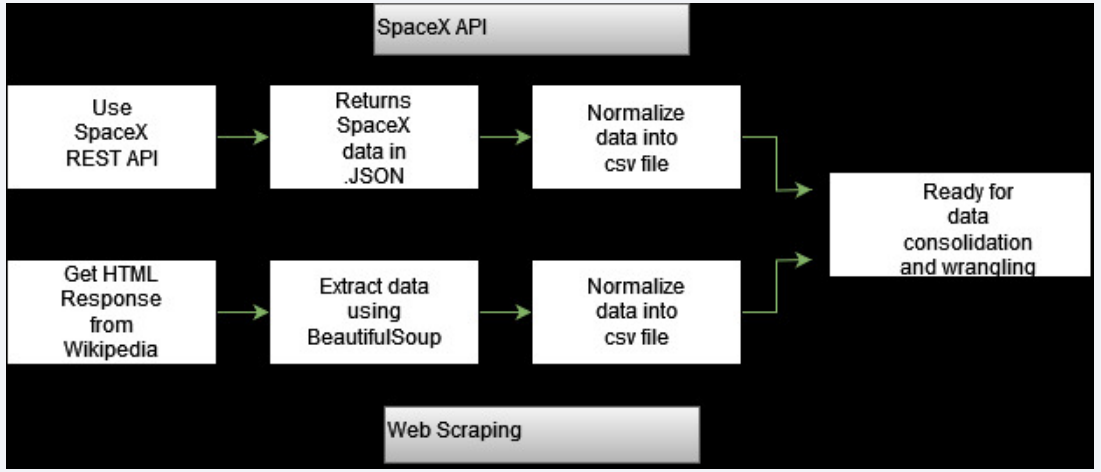
Methodology

Executive Summary

- SpaceX API and Wikipedia launch table data was collected.
- Data was cleaned in preparation for visualization, queries and machine learning model creation.
- Exploratory data analysis (EDA) was done using visualization and SQL.
- Interactive visual analytics were created using Folium and Plotly Dash.
- Predictive analysis using classification models was done.

Data Collection

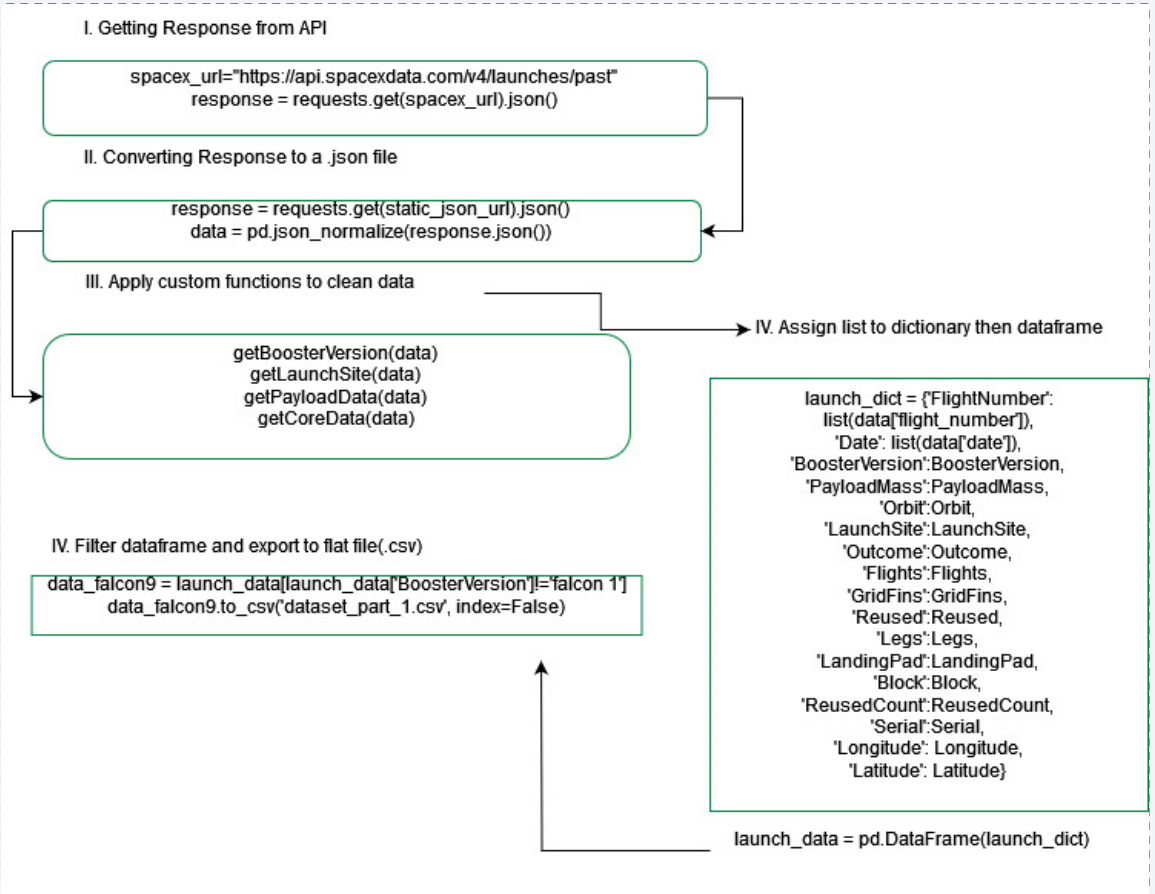
- Data was collected using the following:
  - SpaceX launch data was obtained from the SpaceX REST API.
  - The API contains data rocket launch, including information about the rocket used, payload delivered, launch specification, landing specification, and landing outcome.
  - Another source of data is web scraping Wikipedia using BeautifulSoup.



Data Collection – SpaceX API

Data collection with SpaceX REST calls.

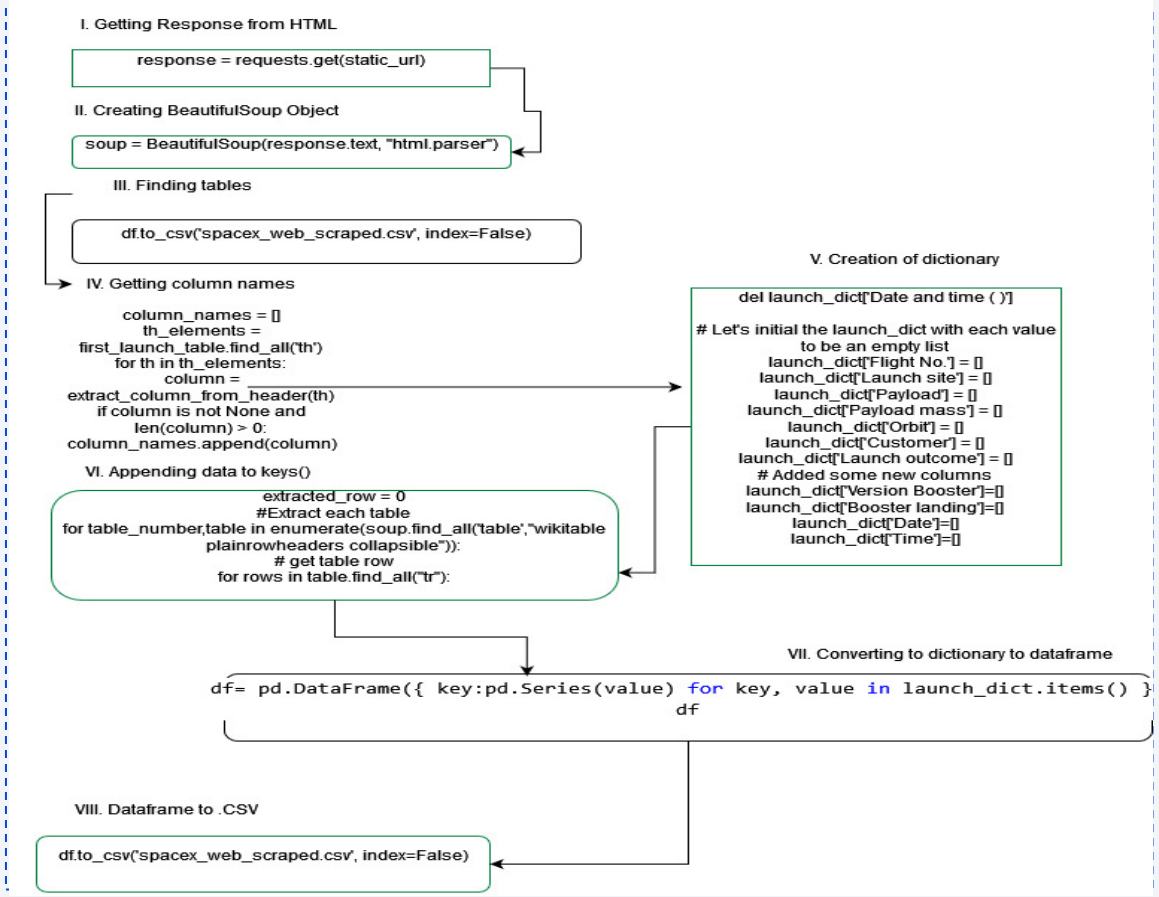
https://github.com/uriameAmode/uriame-IBM-Coursera_Data_Science_Capstone/blob/main/spacex-data-collection-api.ipynb



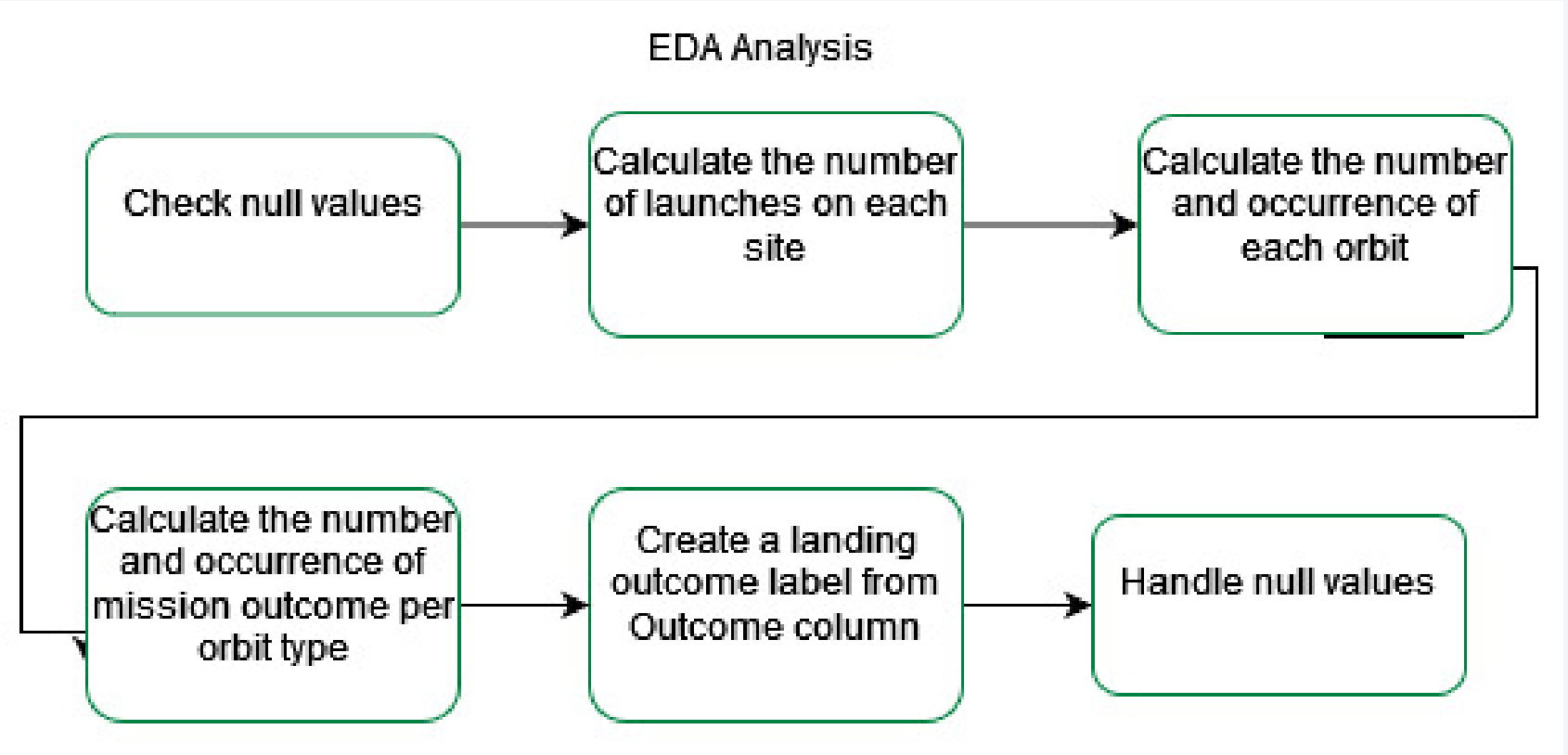
Data Collection - Scraping

- Web scraping from Wikipedia

https://github.com/uriameAmode/uriame-IBM-Coursera_Data_Science_Capstone/blob/main/jupyter-labs-webscraping.ipynb

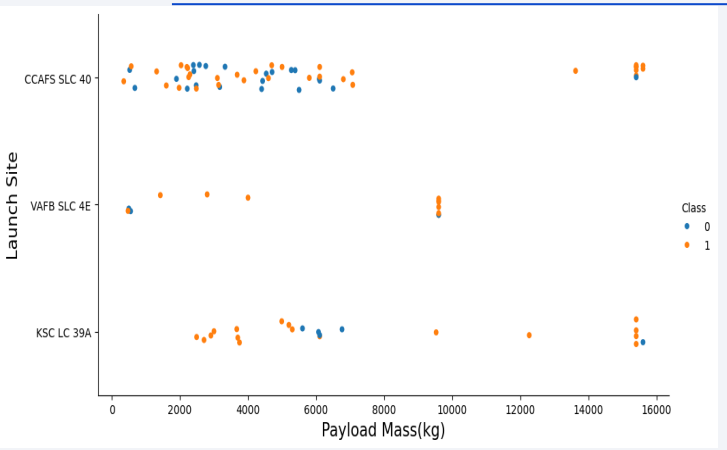


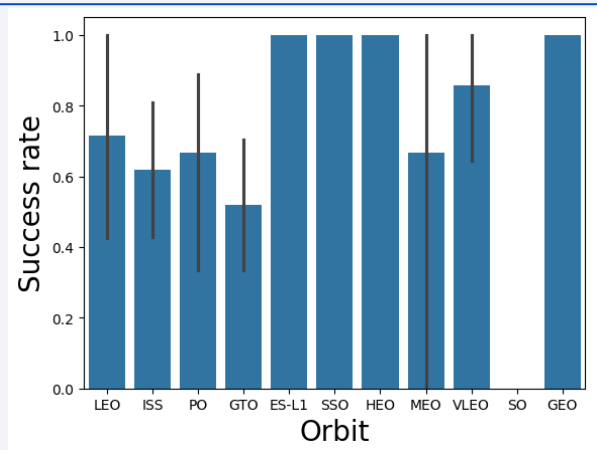
Data Wrangling

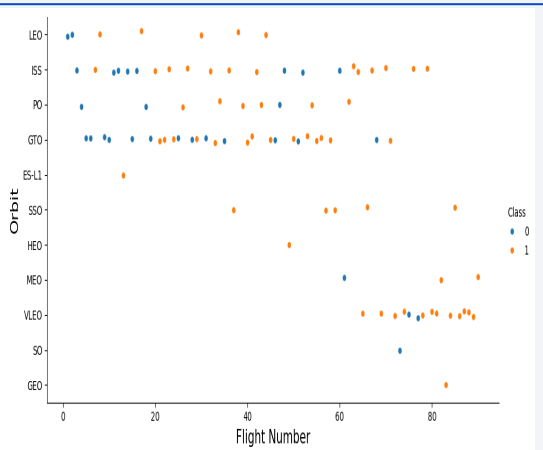


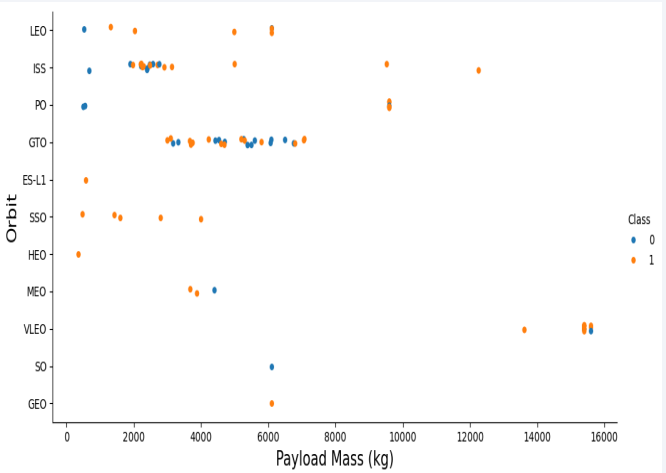
https://github.com/uriameAmode/uriame-IBM-Coursera_Data_Science_Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

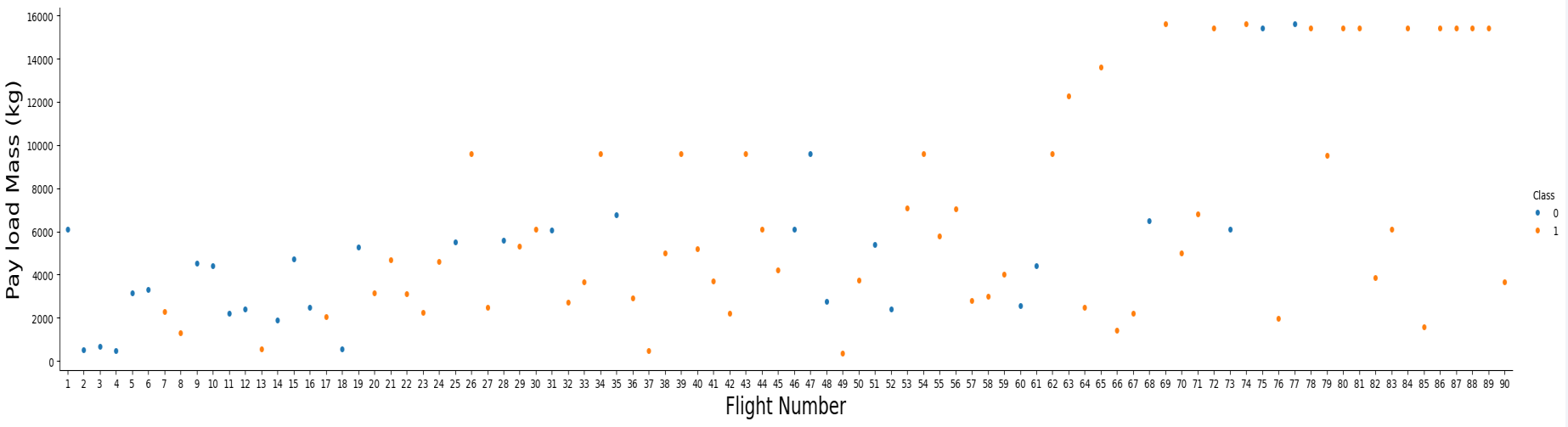
EDA with Data Visualization

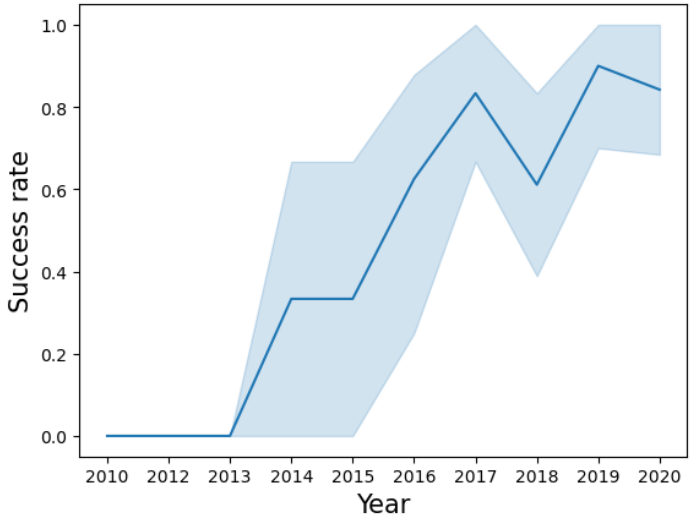












https://github.com/uriameAmode/uriame-IBM-Coursera_Data_Science_Capstone/blob/main/edadataviz.ipynb

EDA with SQL

- Queries were written to extract information about:
  - Launch sites
  - Payload masses
  - Dates
  - Booster types
  - Mission outcomes
- GitHub URL (EDA with SQL):

https://github.com/uriameAmode/uriame-IBM-Coursera_Data_Science_Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
  - Markers were added for launch sites and the NASA Johnson Space Center.
  - Circles were added for the launch sites.
  - Lines were added to show the distance to the nearby features:
    - Distance from CCAFS LC-40 to the coastline
    - Distance from CCAFS LC-40 to the rail line
    - Distance from CCAFS LC-40 to the perimeter road

- Github url(Folium Maps):

https://github.com/uriameAmode/uriame-IBM-Coursera_Data_Science_Capstone/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- The input dropdown is used to select one or all launch sites for the pie chart and scatter plot.
- The pie chart displays one of two things:
  - For All Sites – the distribution of successful Falcon 9 first stage landings between the sites.
  - For One Site – the distribution of successful and failed Falcon 9 first stage landings for that site.
- The input slider is used to filter the payload masses for the scatterplot.
- The scatterplot displays the distribution of Falcon 9 first stage landings split by payload mass, mission outcome and by booster version category.

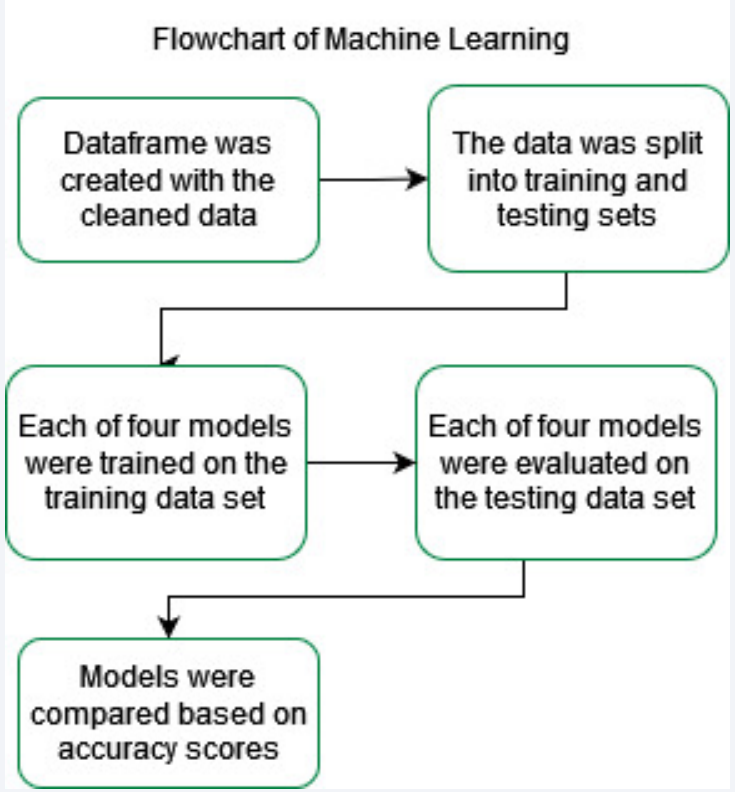
- Github url(Dashboard)

https://github.com/uriameAmode/uriame-IBM-Coursera_Data_Science_Capstone/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

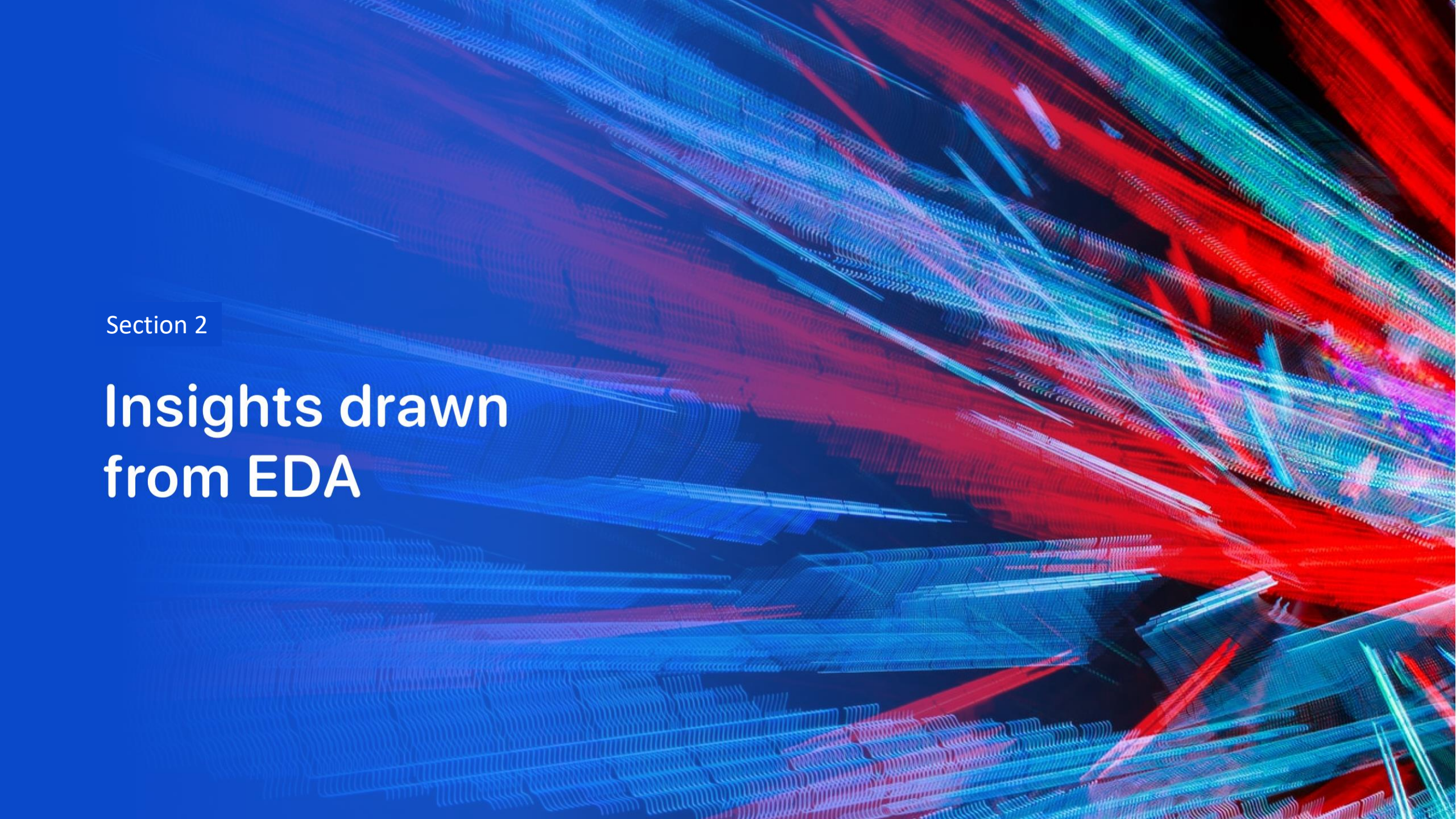
- The dataset was split into training and testing sets.
- Logistic Regression, SVM (Support Vector Machine), Decision Tree, and KNN (k-Nearest Neighbors) machine learning models were trained on the training data set.
- Hyper-parameters were evaluated using GridSearchCV() and the best was selected using '.best_params_'.
- Using the best hyper-parameters, each of the four models were scored on accuracy by using the testing data set.
- GitHub URL (Machine Learning):

https://github.com/uriameAmode/uriame-IBM-Coursera_Data_Science_Capstone/blob/main/SpaceX_Machine%20Learning%20Prediction.ipynb



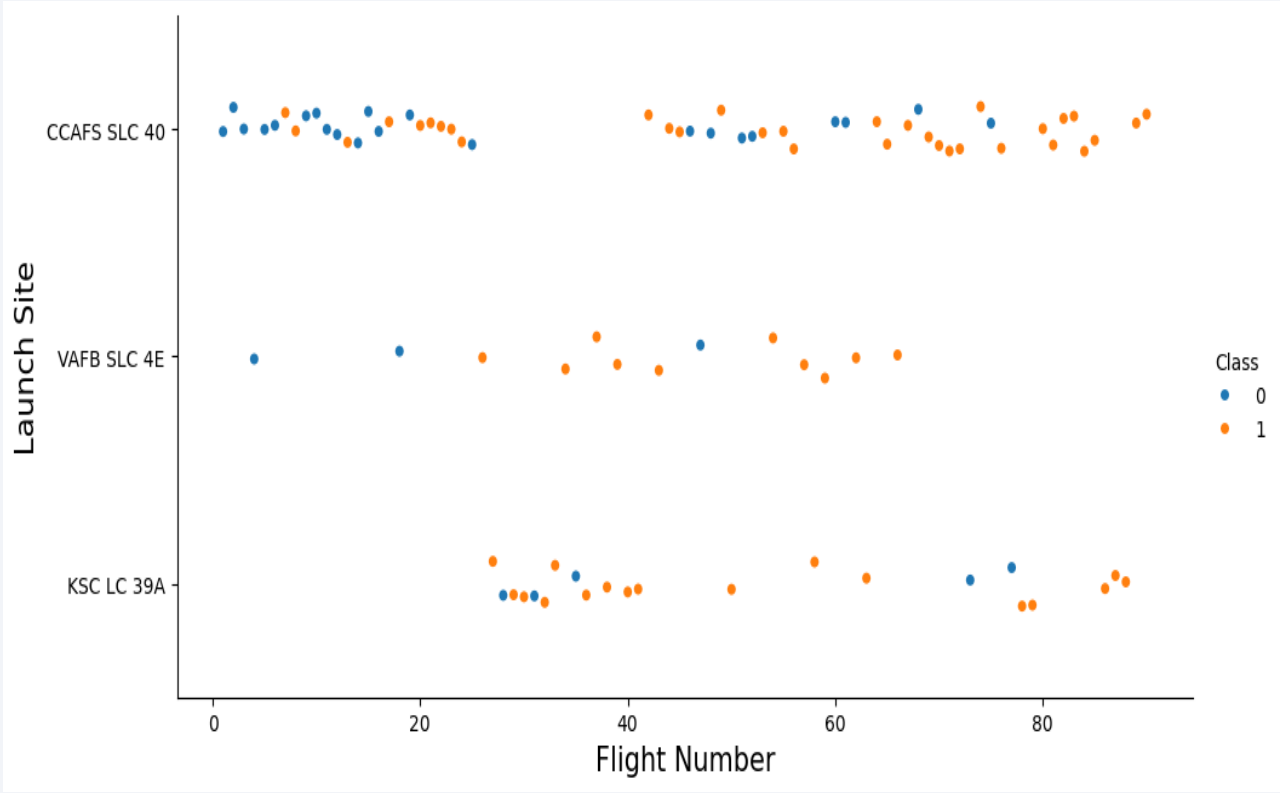
Results

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


Section 2
Insights drawn from EDA

Flight Number vs. Launch Site

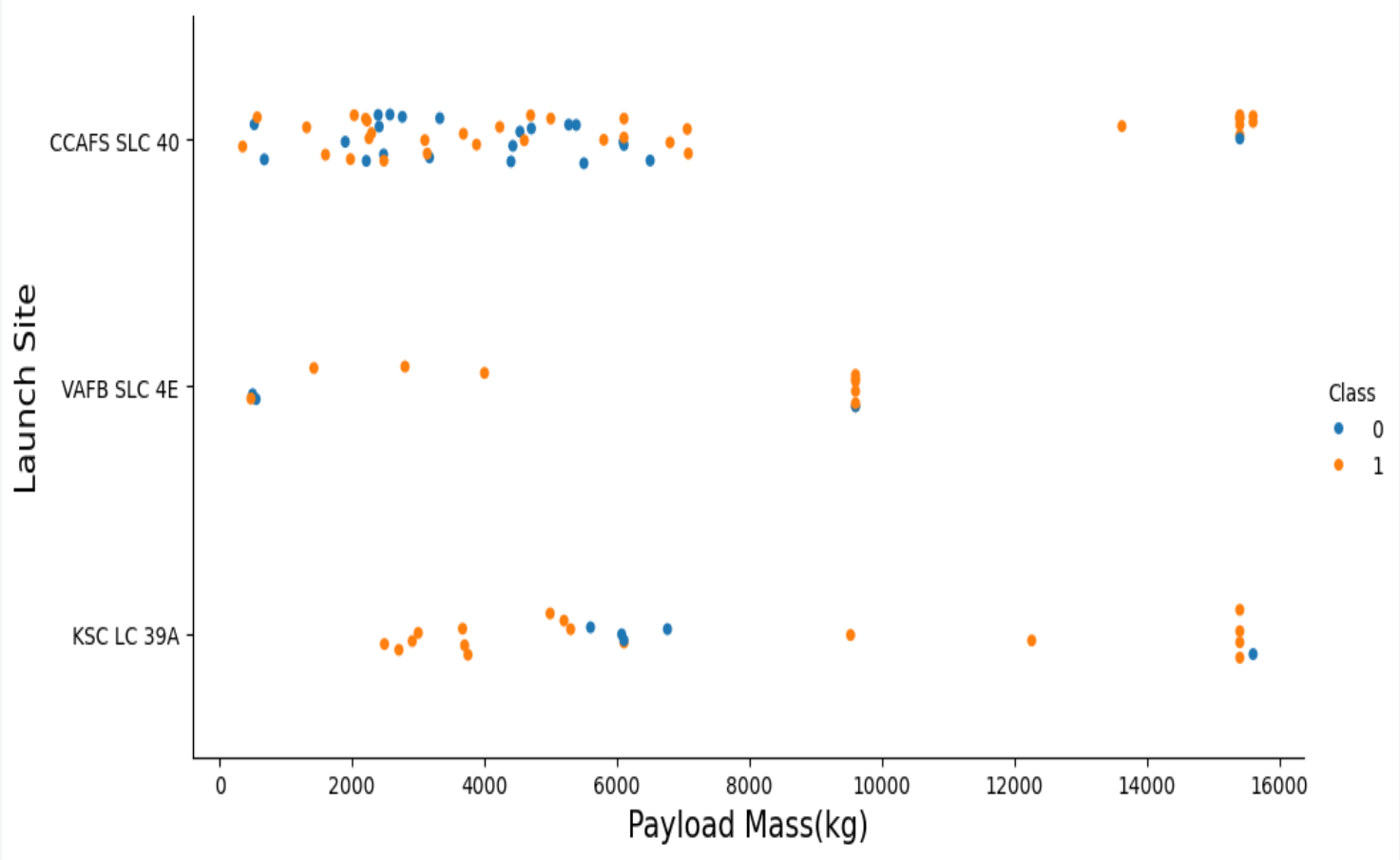
- Success rate varies noticeably with launch site.
- Successful Falcon 9 first stage landings appear to become more prevalent as the flight number increases.



Falcon 9 first stage failed landings are indicated by the ‘0’ Class (blue markers) and successful landings by the ‘1’ Class (orange markers).

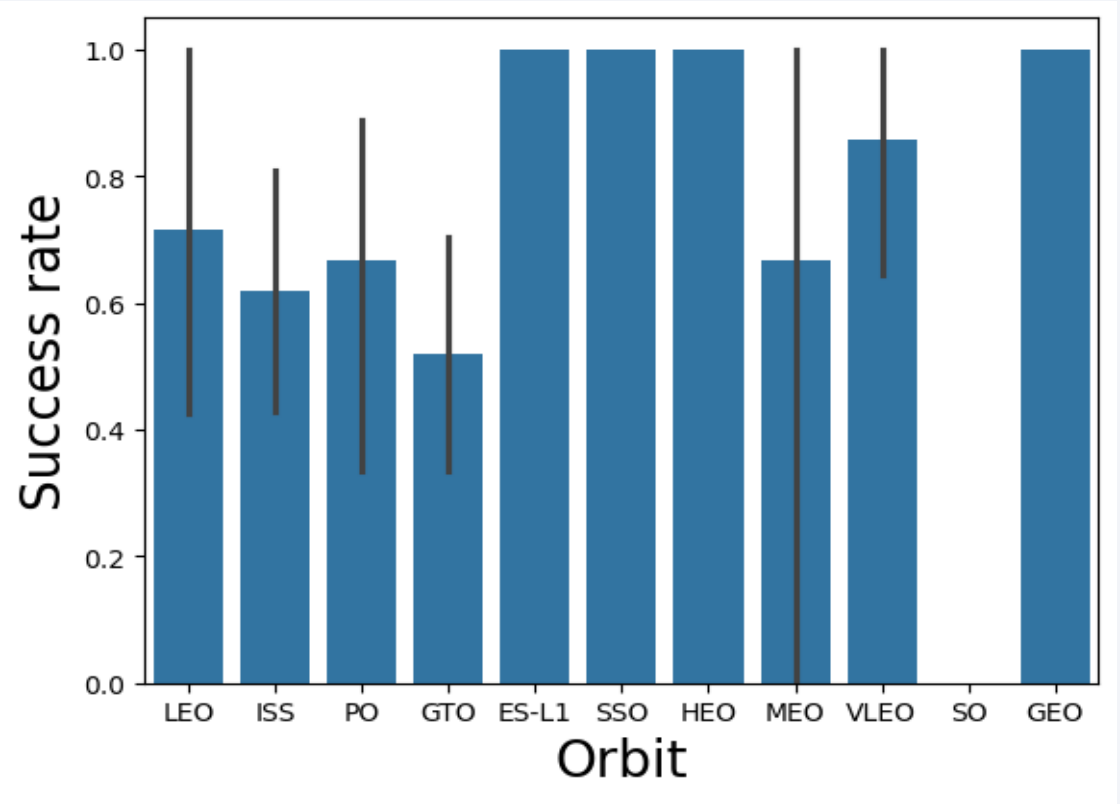
Payload vs. Launch Site

- For the CCAFS SLC 40 launch site, the payload mass and the landing outcome appear to not be strongly correlated.
- • The failed landings at the KSC LC 39A launch site are all grouped around a narrow band of payload masses.



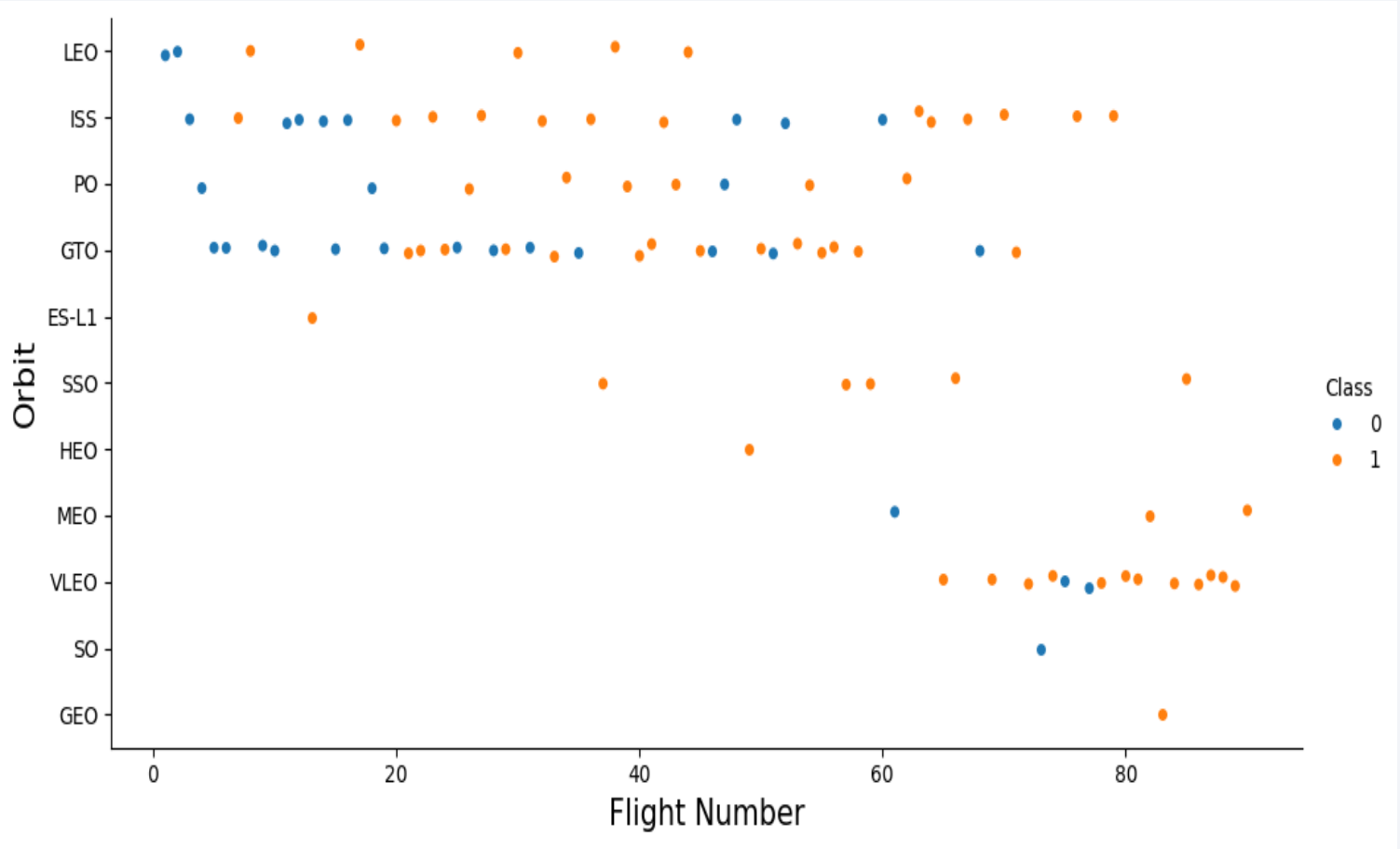
Success Rate vs. Orbit Type

- ES-L1, SSO, HEO and GEO orbits have no failed first stage landings.
- SO orbits have no successful first stage landings.



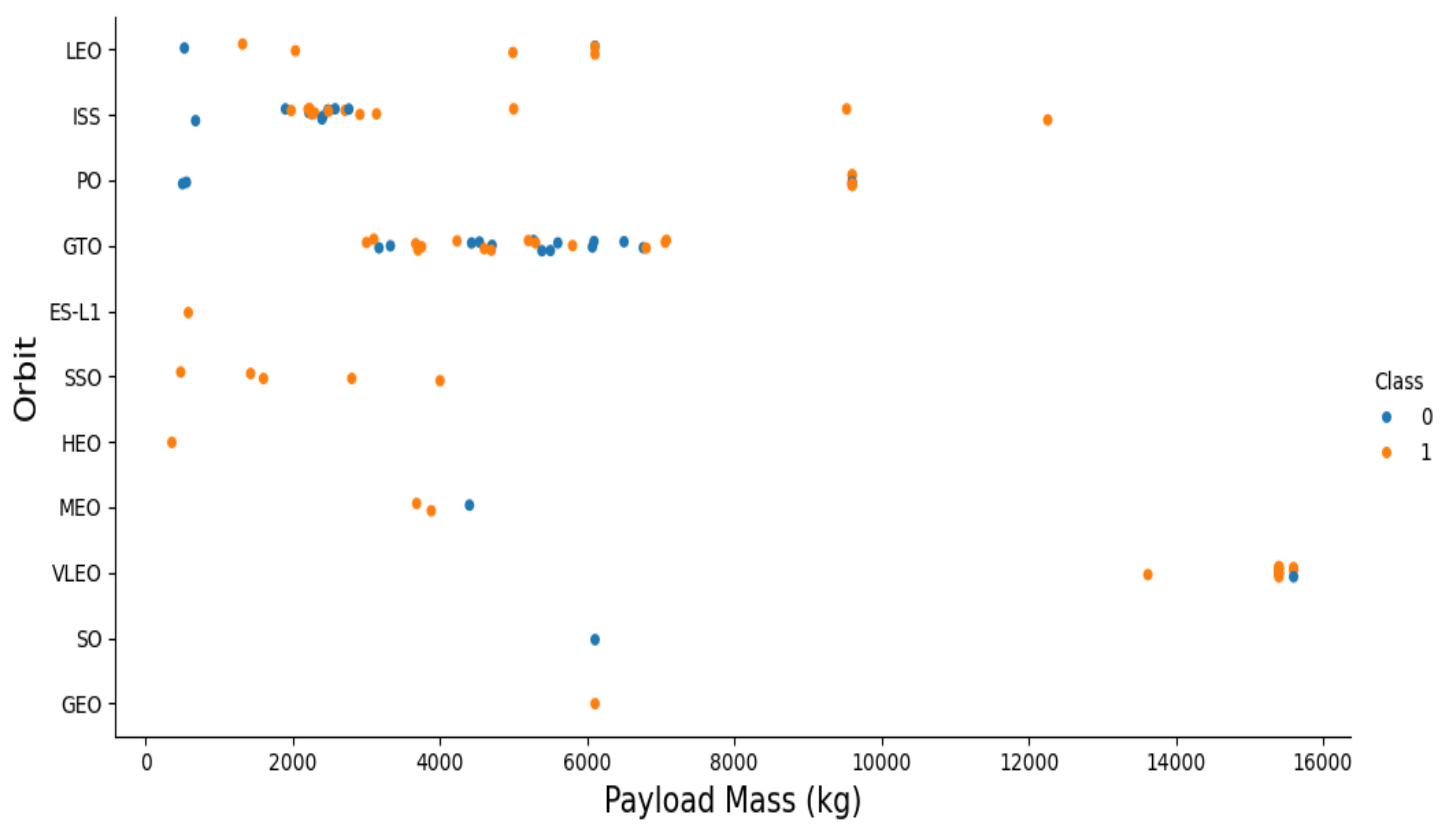
Flight Number vs. Orbit Type

- There is a correlation between flight number and success rate with larger flight numbers being associated with higher success rates.



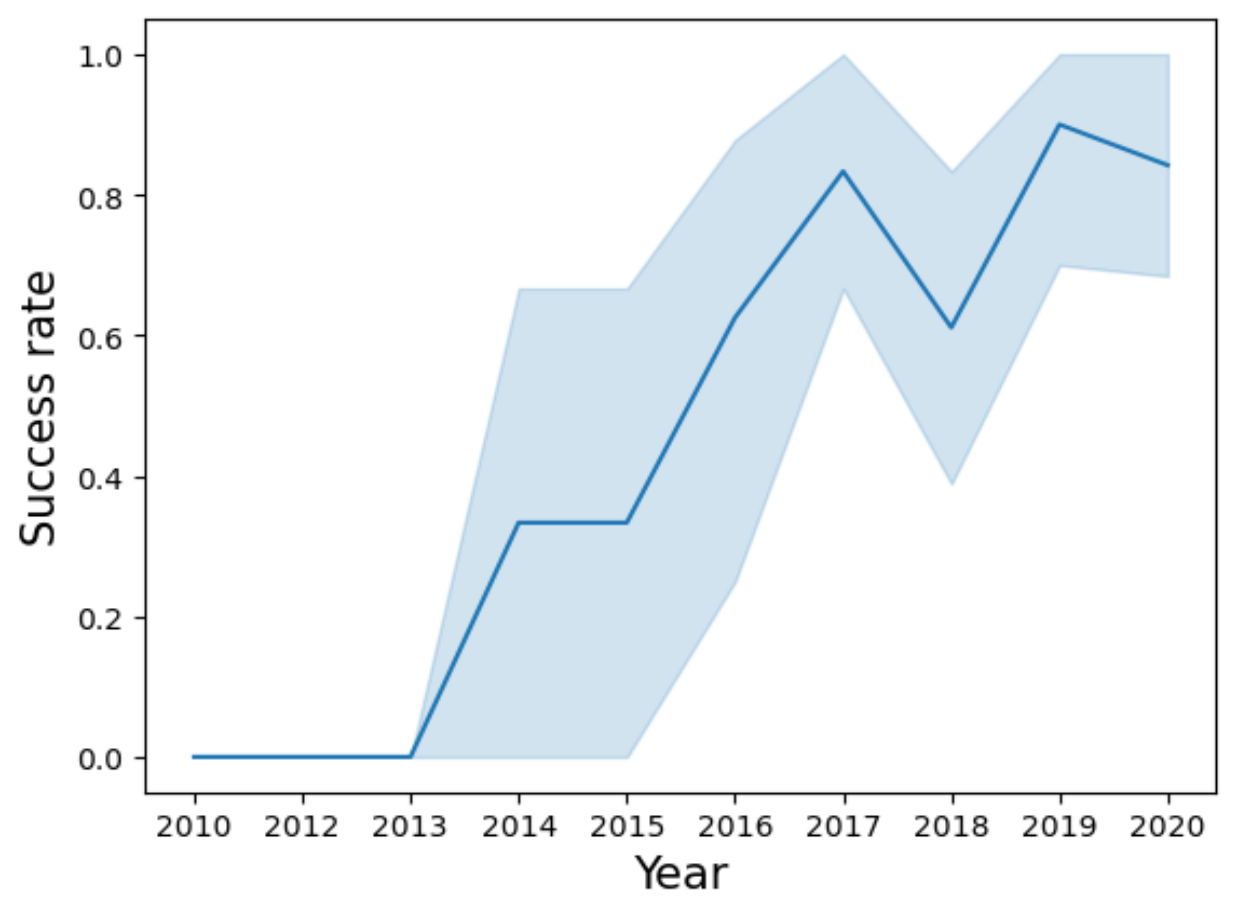
Payload vs. Orbit Type

- Some orbit types have better success rates than others.
- Success rate appears to have no obvious correlation with payload mass.



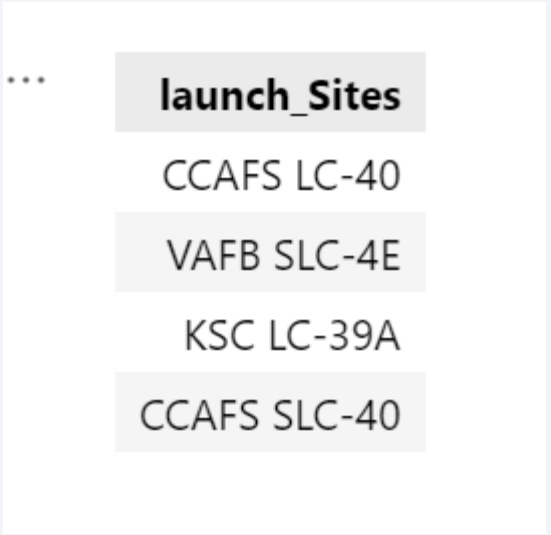
Launch Success Yearly Trend

- The success rate has increased significantly over the years.



All Launch Site Names

- `%sql SELECT DISTINCT LAUNCH_SITE as "launch_Sites" FROM SPACEXTBL`



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

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`

```
%sql SELECT * FROM SPACEXTBL 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 in Kg of Boosters launched
by NASA" FROM SPACEXTBL WHERE CUSTOMER = 'NASA (CRS)'
```

Total Payload mass in Kg of Boosters launched by NASA
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 in Kg of Booster version
F9 v1.1" FROM SPACEXTBL WHERE BOOSTER_VERSION = 'F9 v1.1'
```

Average Payload Mass in Kg of Booster version F9 v1.1
2928.4

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad

```
%sql SELECT MIN(DATE) as "Date of first successful landing Outcome in ground pad" FROM
SPACEXTBL WHERE LANDING_OUTCOME = 'Success (ground pad)'
```

Date of first successful landing Outcome in ground pad
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 SPACEXTBL 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 number_of_success_outcomes, number_of_failure_outcomes FROM (SELECT
COUNT(*) AS number_of_success_outcomes FROM SPACEXTBL WHERE MISSION_OUTCOME like
'Success%') success_table, (SELECT COUNT(*) number_of_failure_outcomes FROM SPACEXTBL
WHERE MISSION_OUTCOME like 'Failure%') failure_table
```

...

number_of_success_outcomes	number_of_failure_outcomes
100	1

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass

```
%sql SELECT DISTINCT BOOSTER_VERSION FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT
MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL)
```

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 FROM SPACEXTBL WHERE substr(DATE,0,5) =
'2015' AND LANDING_OUTCOME = 'Failure (drone ship)'
```

Date	Booster_Version	Launch_Site
2015-01-10	F9 v1.1 B1012	CCAFS LC-40
2015-04-14	F9 v1.1 B1015	CCAFS LC-40

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

- 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

```
%sql SELECT LANDING_OUTCOME, COUNT(LANDING_OUTCOME) AS Landing_Count FROM SPACEXTBL
WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY LANDING_OUTCOME ORDER BY
COUNT(LANDING_OUTCOME) DESC
```

Landing_Outcome	Landing_Count
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

Falcon 9 Launch Site Locations

- VAFB SLC-4E (California, USA)

Vandenberg Air Force Base Space Launch Complex 4E

- KSC LC-39A (Florida, USA)

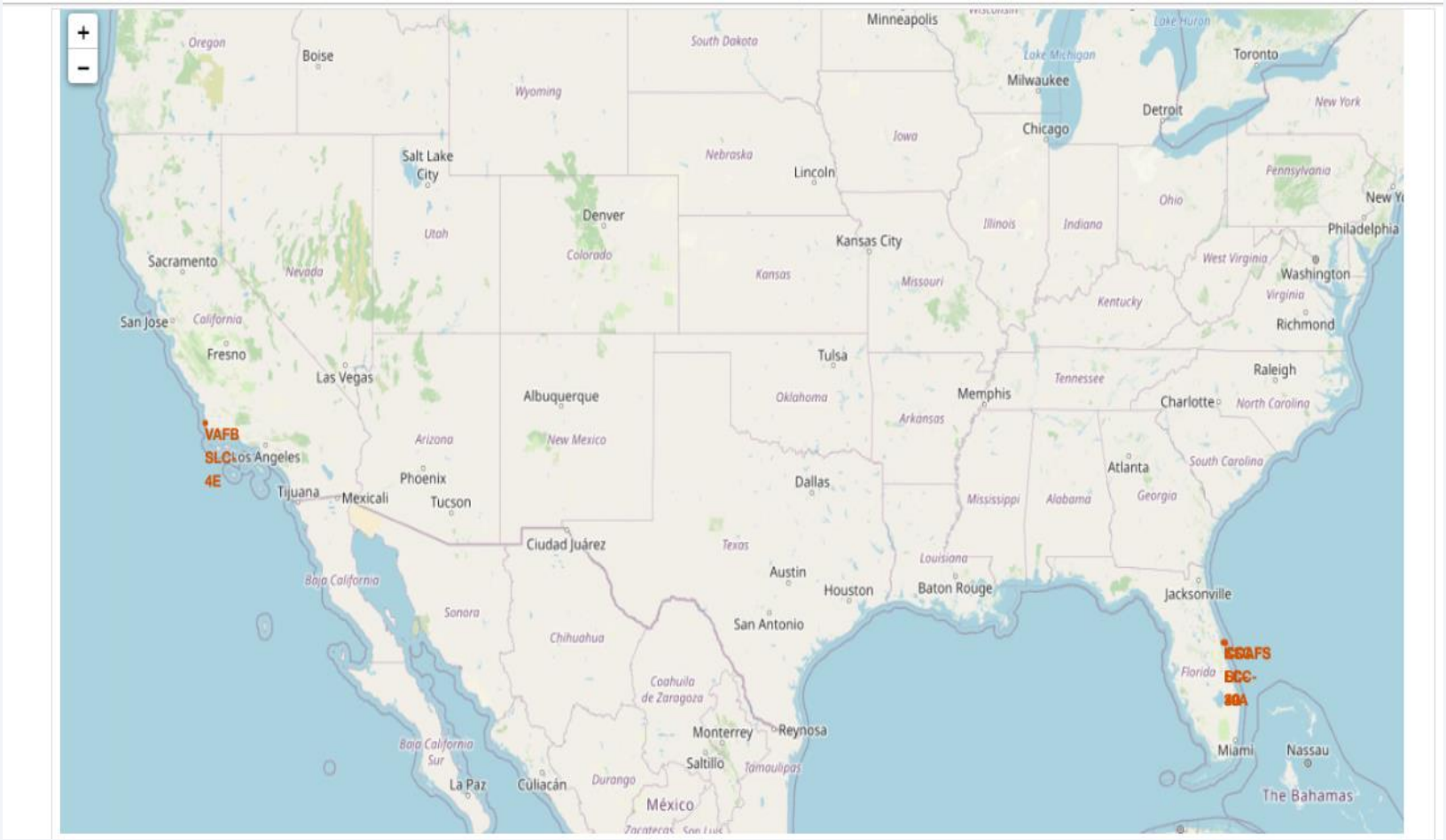
Kennedy Space Center Launch Complex 39A

- CCAFS LC-40 (Florida, USA)

Cape Canaveral Air Force Station Launch Complex 40

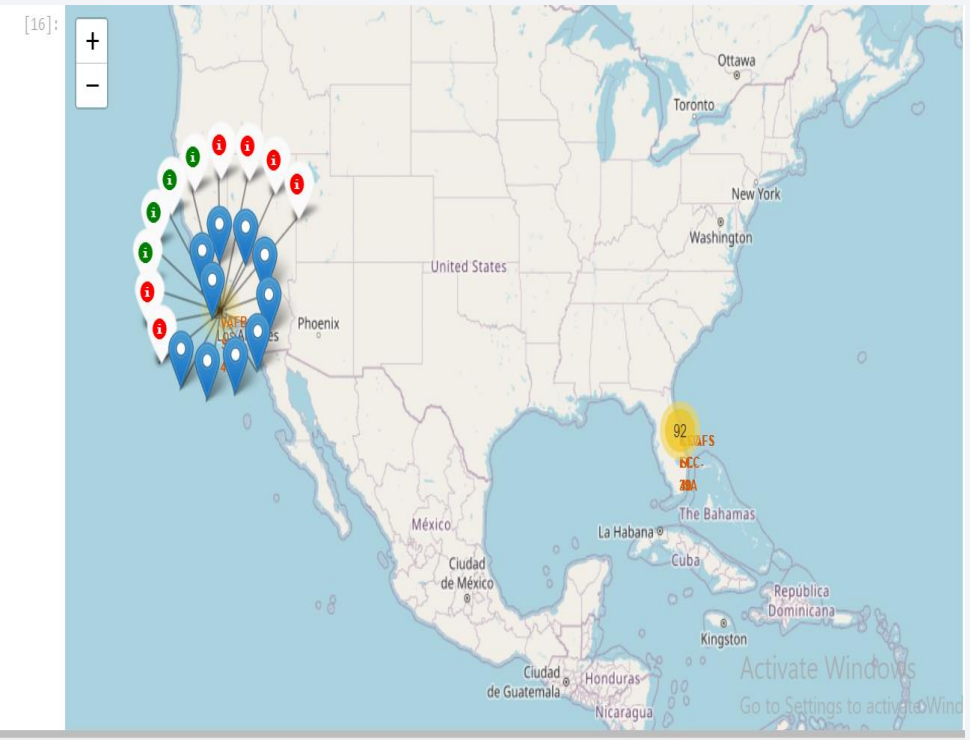
- CCAFS SLC-40 (Florida, USA)

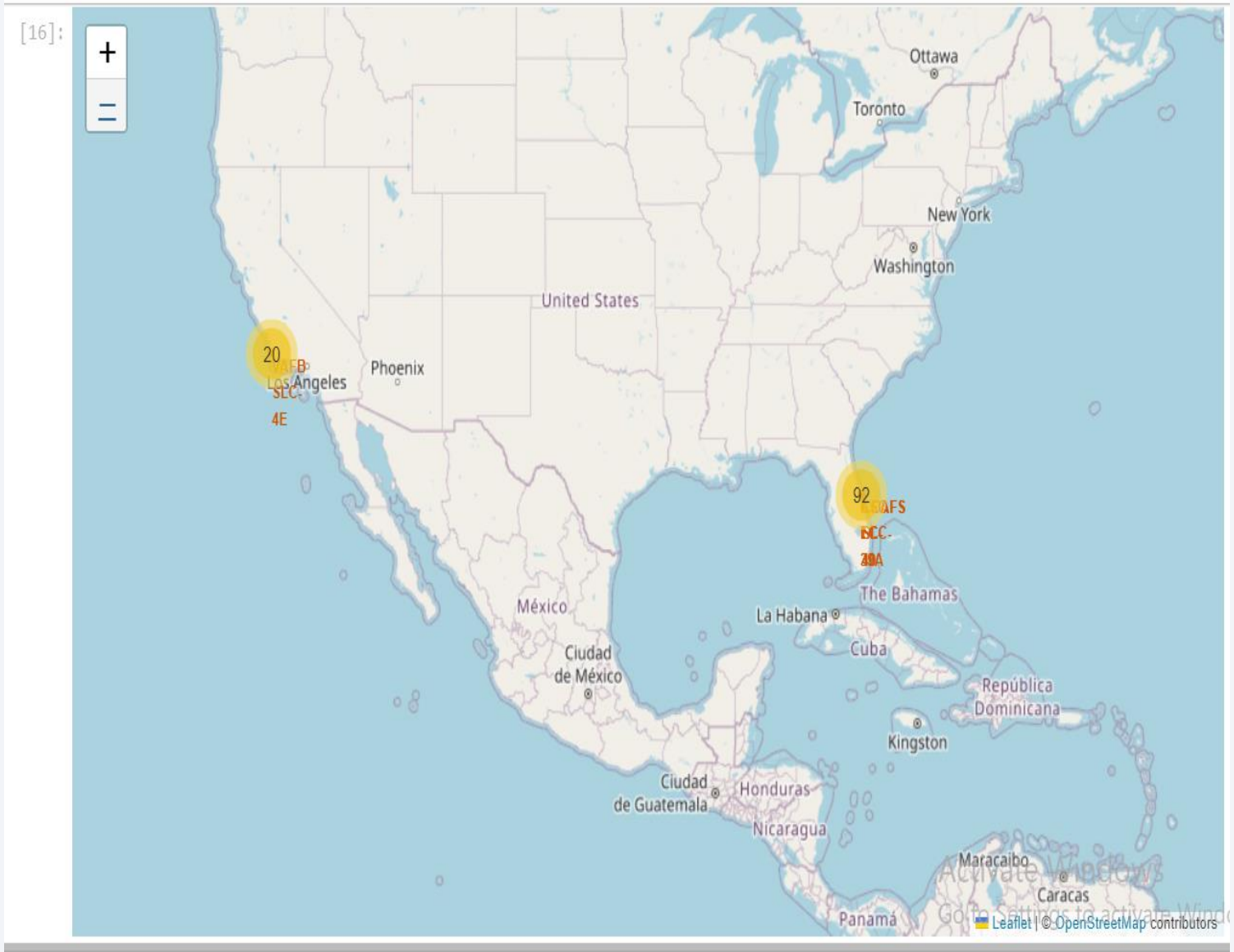
Cape Canaveral Air Force Station Space Launch Complex 40



Map Markers of Success/Failed Landings

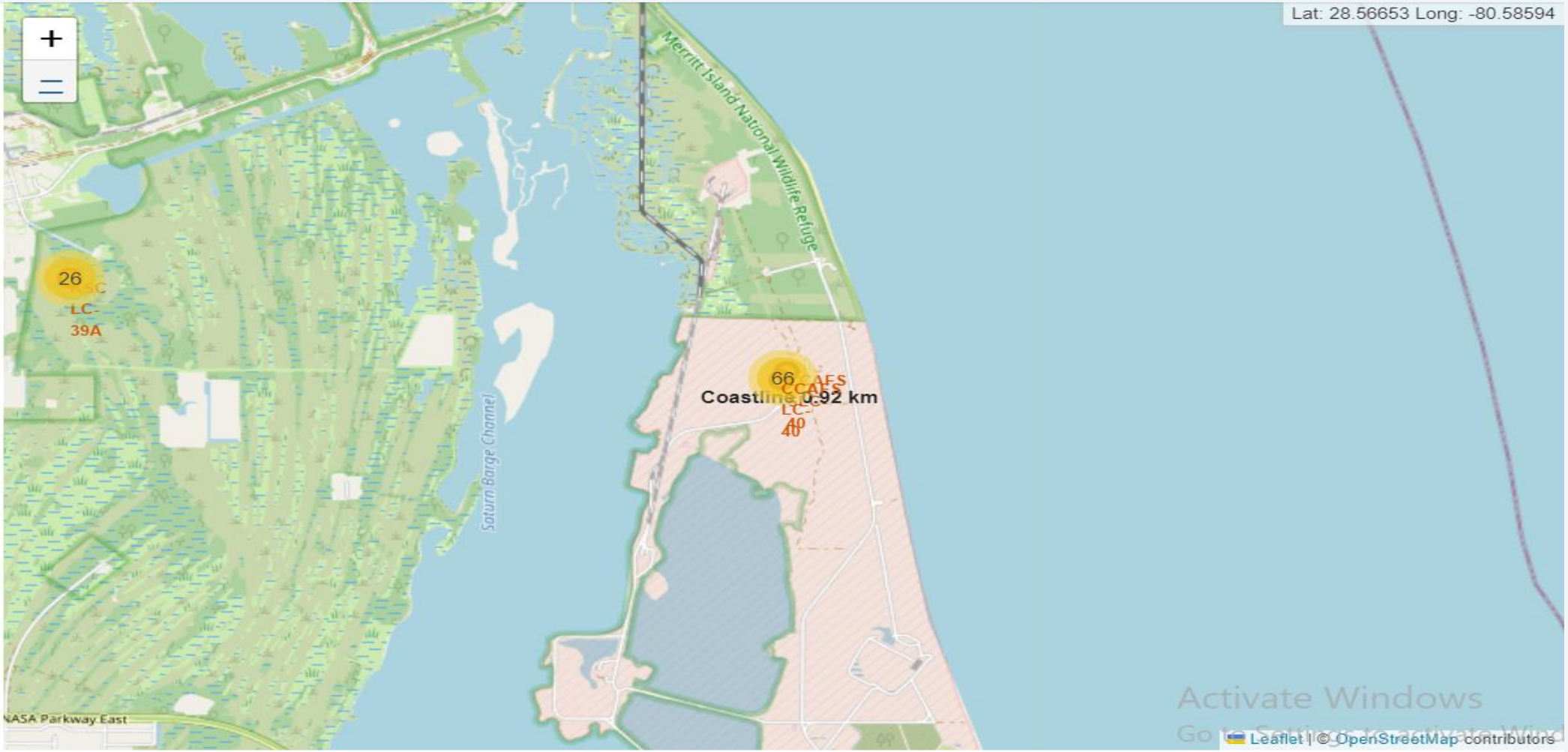
A sense of a launch site's success rate for Falcon 9 first stage landings can be gleaned from the relative number of green success markers to red failure markers.

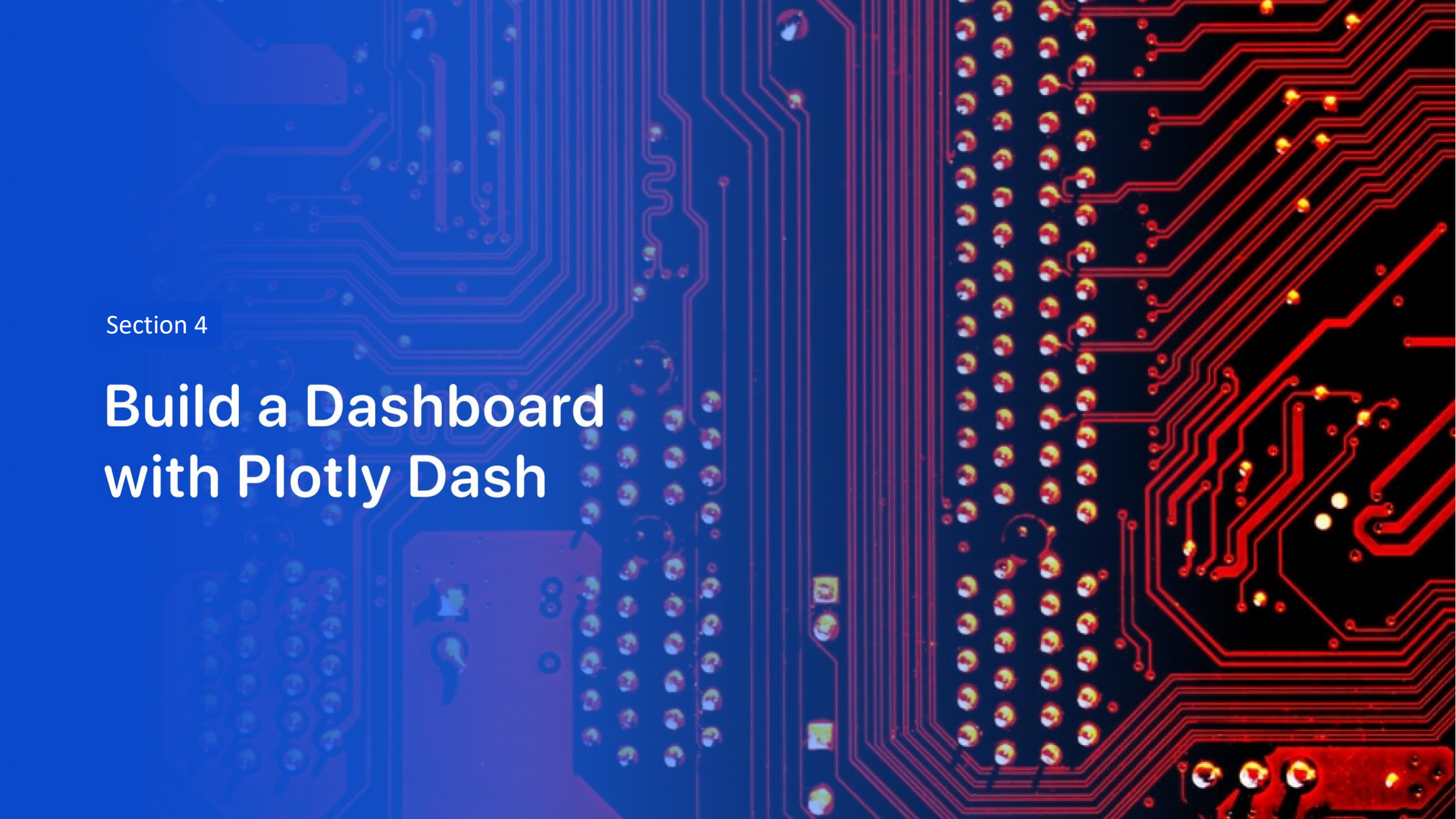




Distance from Launch Site to Proximities

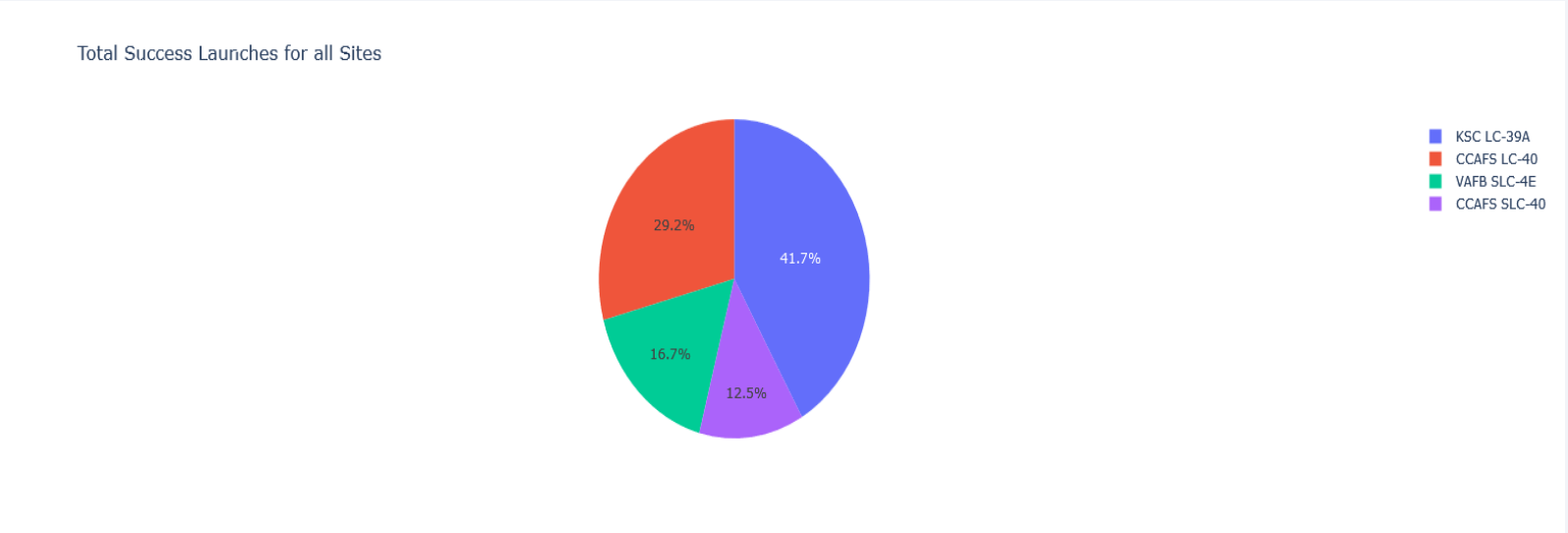
[34]:




Section 4
Build a Dashboard with Plotly Dash

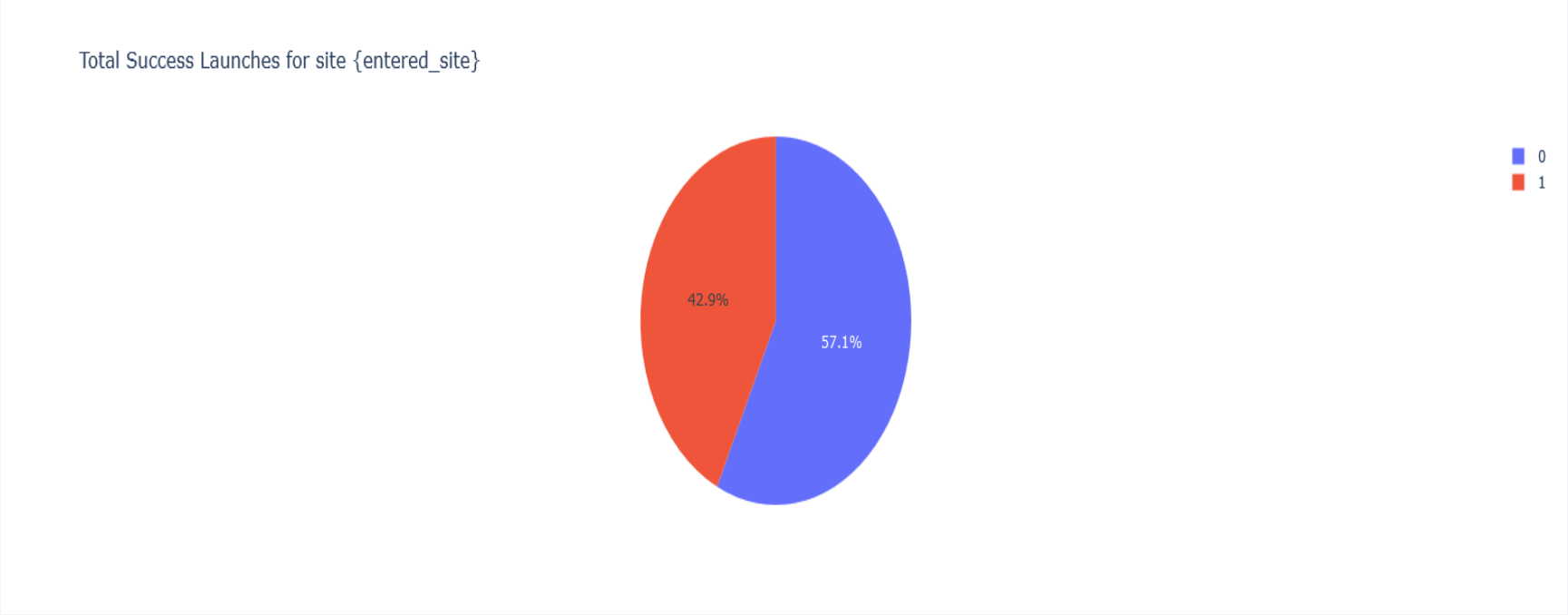
Launch Success Count for All Sites

- The dropdown menu allows the selection of one or all launch sites.
- With all launch sites selected, the pie chart displays the distribution of successful Falcon 9 first stage landing outcomes between the different launch sites.
- The greatest share of successful Falcon 9 first stage landing outcomes (at 41.7% of the total) occurred at KSC LC-39A.



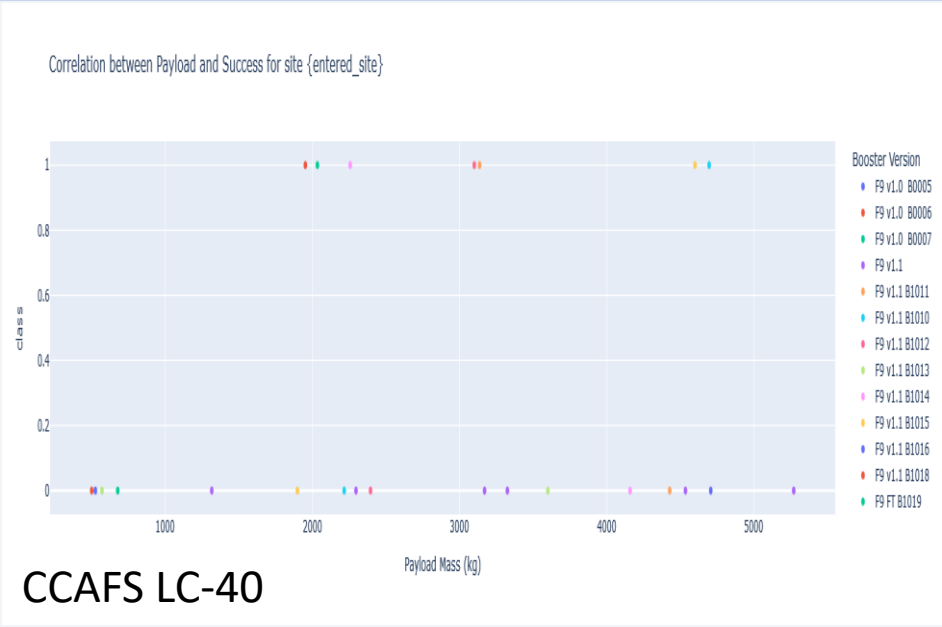
Launch Site with Highest Launch Success Ratio

- Falcon 9 first stage failed landings are indicated by the ‘0’ Class (■ blue wedge in the pie chart) and successful landings by the ‘1’ Class (■ red wedge in the pie chart).
- CCAFS SLC-40 was the launch site that had the highest Falcon 9 first stage landing success rate (42.9%).

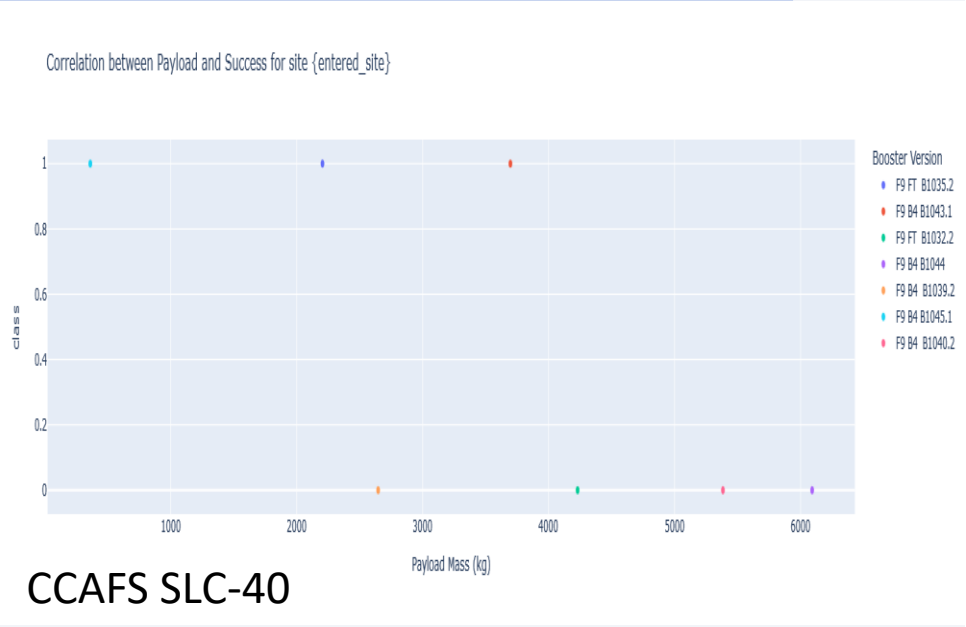


Payload vs. Launch Outcome

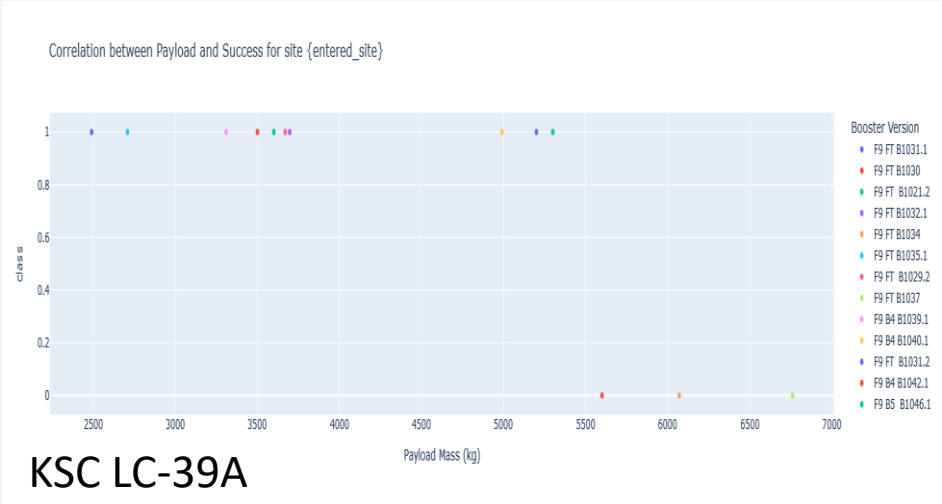
- These screenshots are of the Payload vs. Launch Outcome scatter plots for all sites, with different payload selected in the range slider.
- The payload range from about 2,000 kg to 5,000 kg has the largest success rate.
- The ‘FT’ booster version category has the largest success rate.



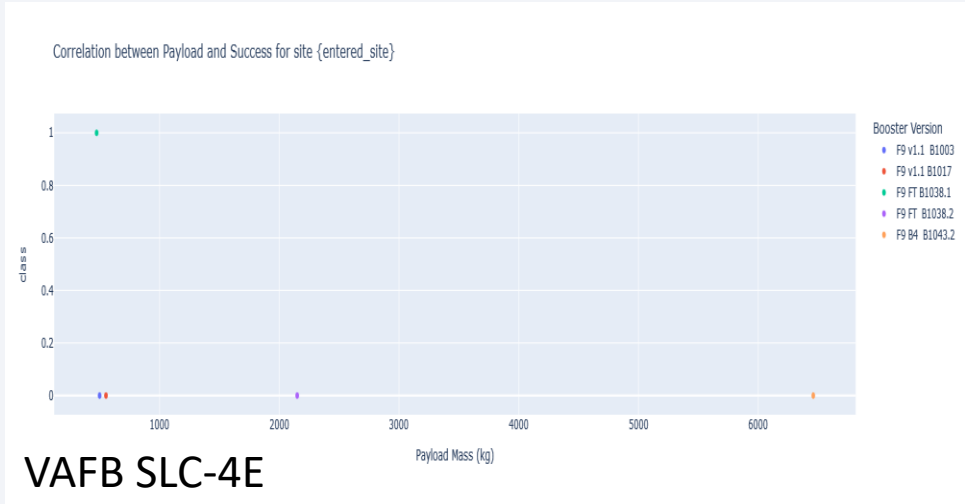
CCAFS LC-40



CCAFS SLC-40



KSC LC-39A



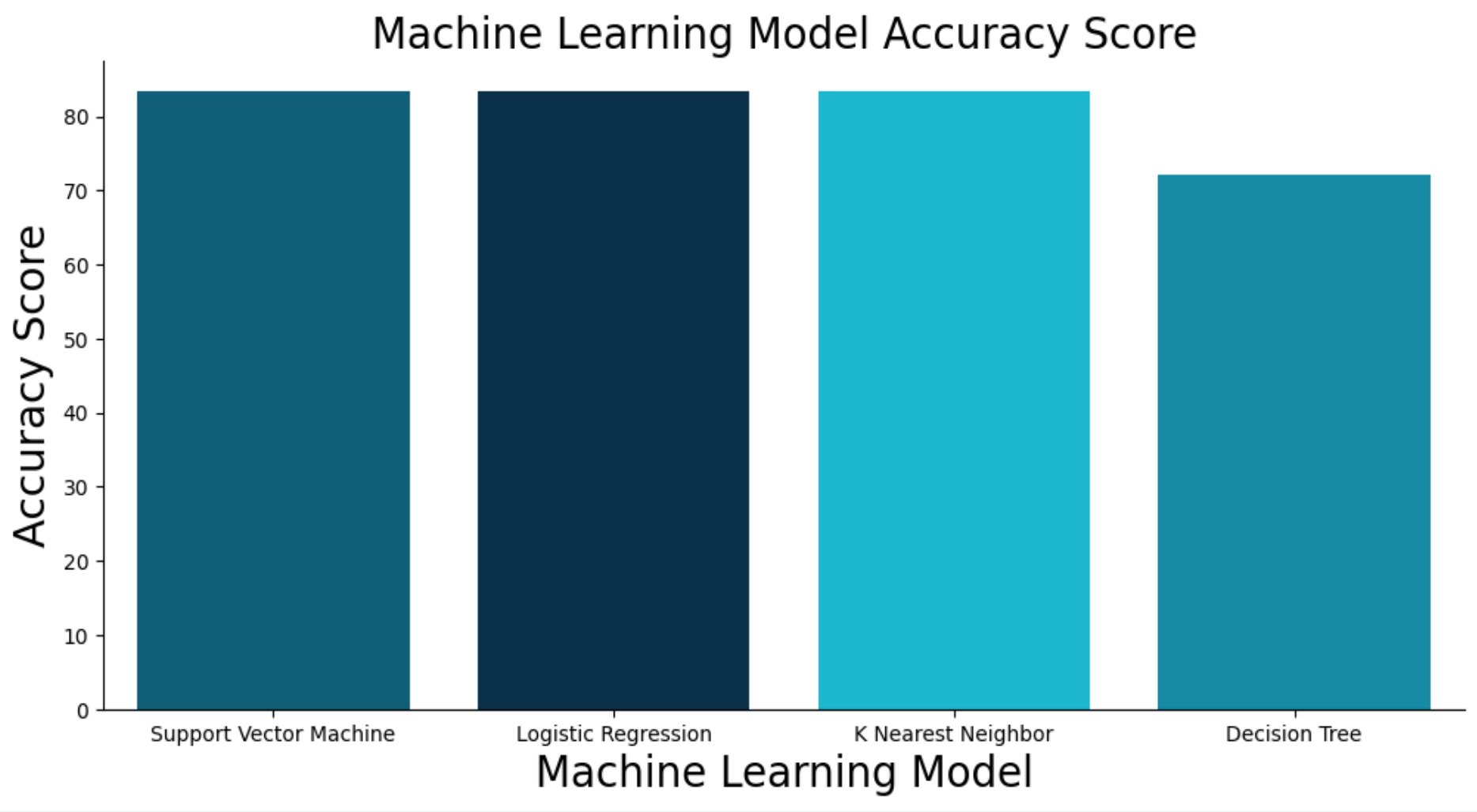
VAFB SLC-4E

Section 5

Predictive Analysis (Classification)

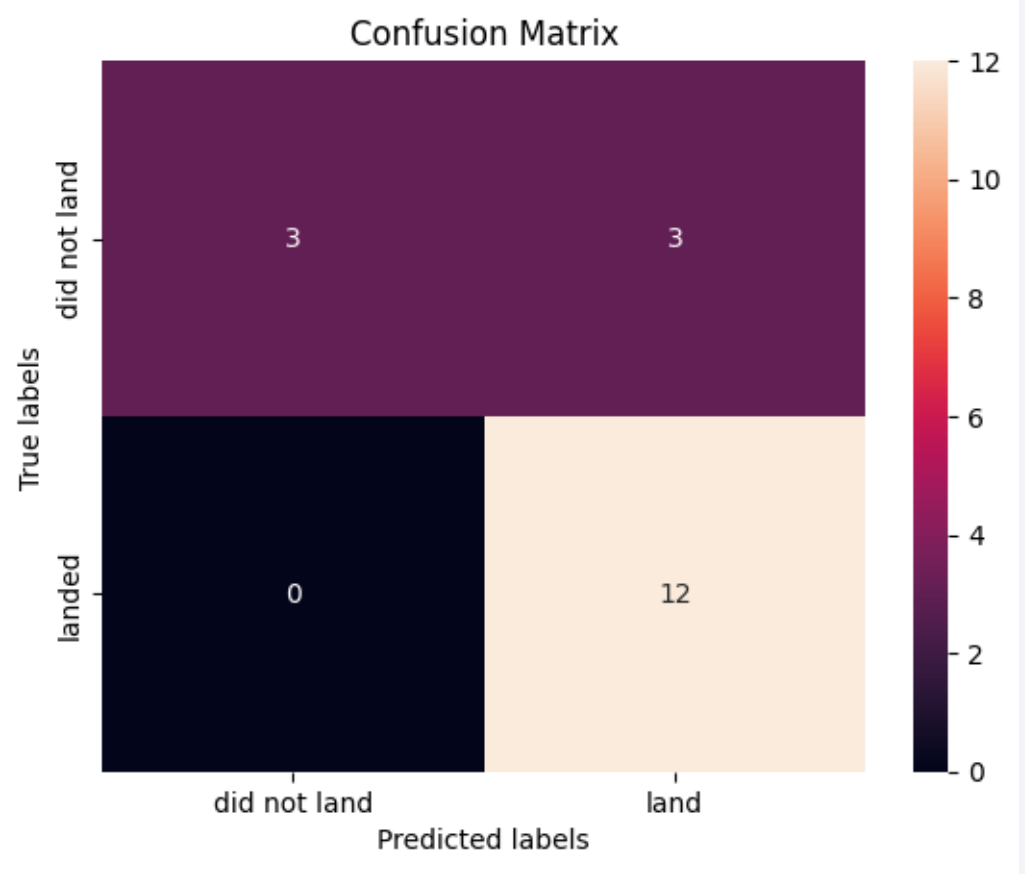
Classification Accuracy

- All models performed almost equally except for the Decision Tree model which performed poorly relative to the other models.



Confusion Matrix

- Show the confusion matrix of the best performing model with an explanation Shown here is the confusion matrix for
- Shown here is the confusion matrix of the best performing model, Logistic Regression.



Conclusions

- The SVM, KNN, and Logistic Regression models are the best in terms of predicting accuracy for this dataset.
- Low weighted payloads perform better than the heavier payloads.
- The success rates for SpaceX launches is directly proportional to the time in years they will eventually perfect the launches.
- KSC LC 39A had the most successful launches from all the sites.
- Oebit GEO, HEO, SSO has the best success rates.

Appendix

SpaceX API (JSON): https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/API_call_spacex_api.json

Wikipedia (Webpage): https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922

SpaceX (CSV): https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-

SkillsNetwork/labs/module_2/data/Spacex.csv?utm_medium=Exinfluencer&utm_source=Exinfluencer&utm_content=000026UJ&utm_term=10006555&utm_id=NA-SkillsNetwork-

Channel-SkillsNetworkCoursesIBMDS0321ENSkillsNetwork26802033-2022-01-01

Launch Geo (CSV): https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/spacex_launch_geo.csv

Launch Dash (CSV): https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/spacex_launch_dash.csv

Thank you!