IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Maria Veronica Alba Baroni
April 14th, 2024

Outline

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

Executive Summary

- The objective of this project was to predict if the Falcon 9 first stage will land successfully. If we can determine if the first stage will land, we can determine the cost of a launch.
- Summary of methodologies: Logistic Regression, Support Vector Machine, Decission Tree, K-nearest Neighbors
- Summary of all results: The method that performed best was the Decission Tree method.

Introduction

- SpaceX advertises Falcon9 rocket launches on its website with a cost of 62 million dollars.
- Why 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.

Section 1

Methodology

Methodology

Executive Summary

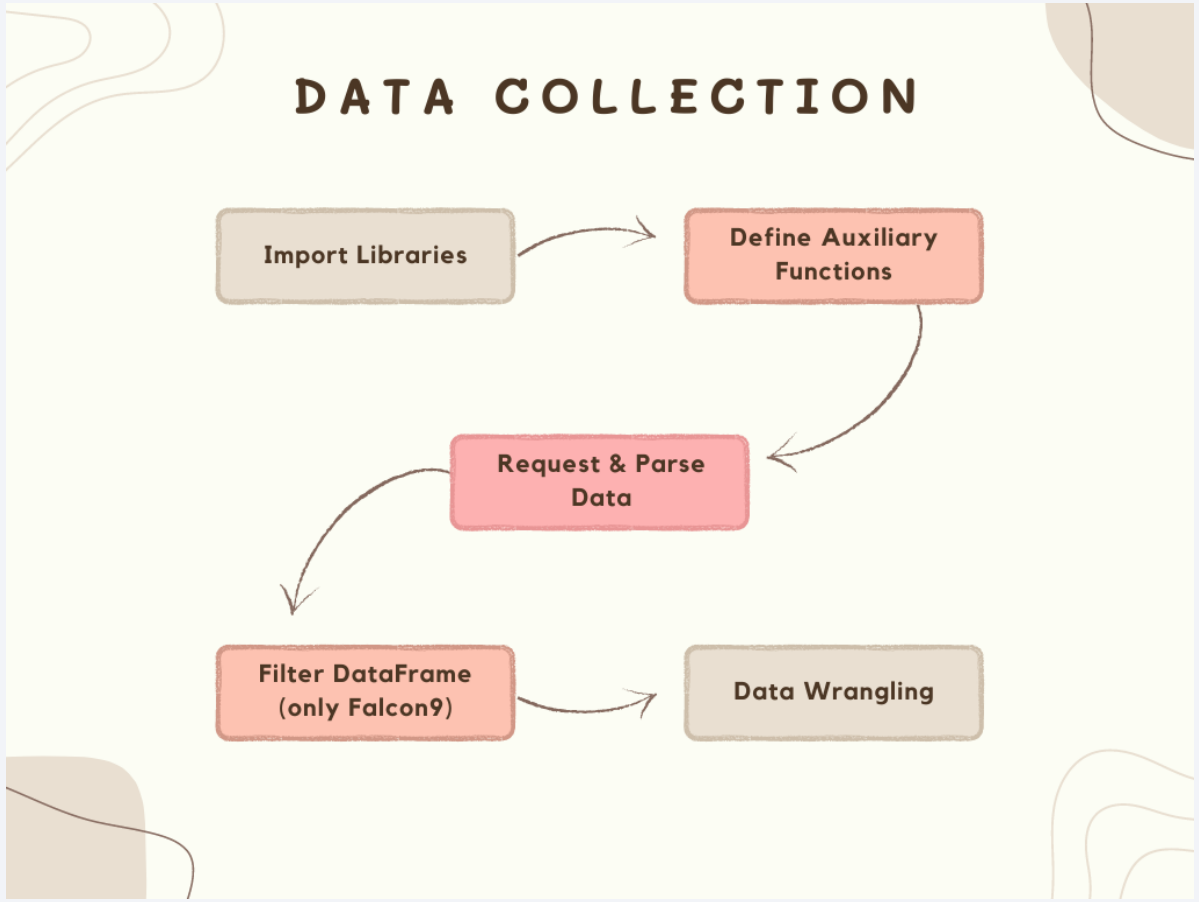
- Data collection methodology:
  - We collected data using a get request to the SpaceX API and we cleaned the requested data
- Perform data wrangling
  - We loaded the SpaceX dataset, calculated the number of launches on each site and the number of occurences of each orbit as well as the outcome of each mission.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - We standardized the data and split it into train and test sets. We created a Logistic Regression and tested its accuracy. We did the same process for a Support Vector Machine, Decision Tree Classifier and K-nearest Neighbors and compared results.

Data Collection

- Describe how data sets were collected.
- You need to present your data collection process use key phrases and flowcharts

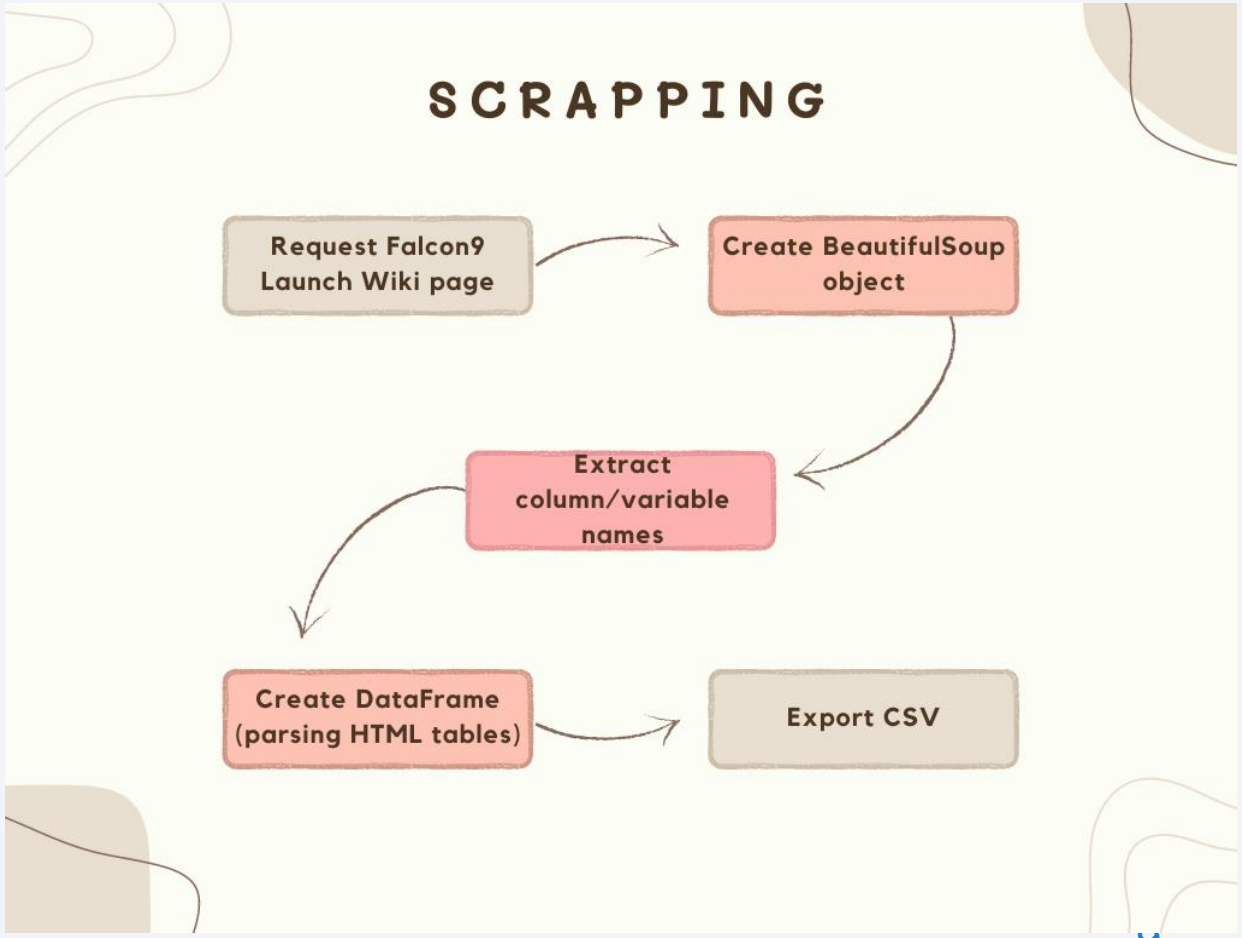
Data Collection – SpaceX API

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



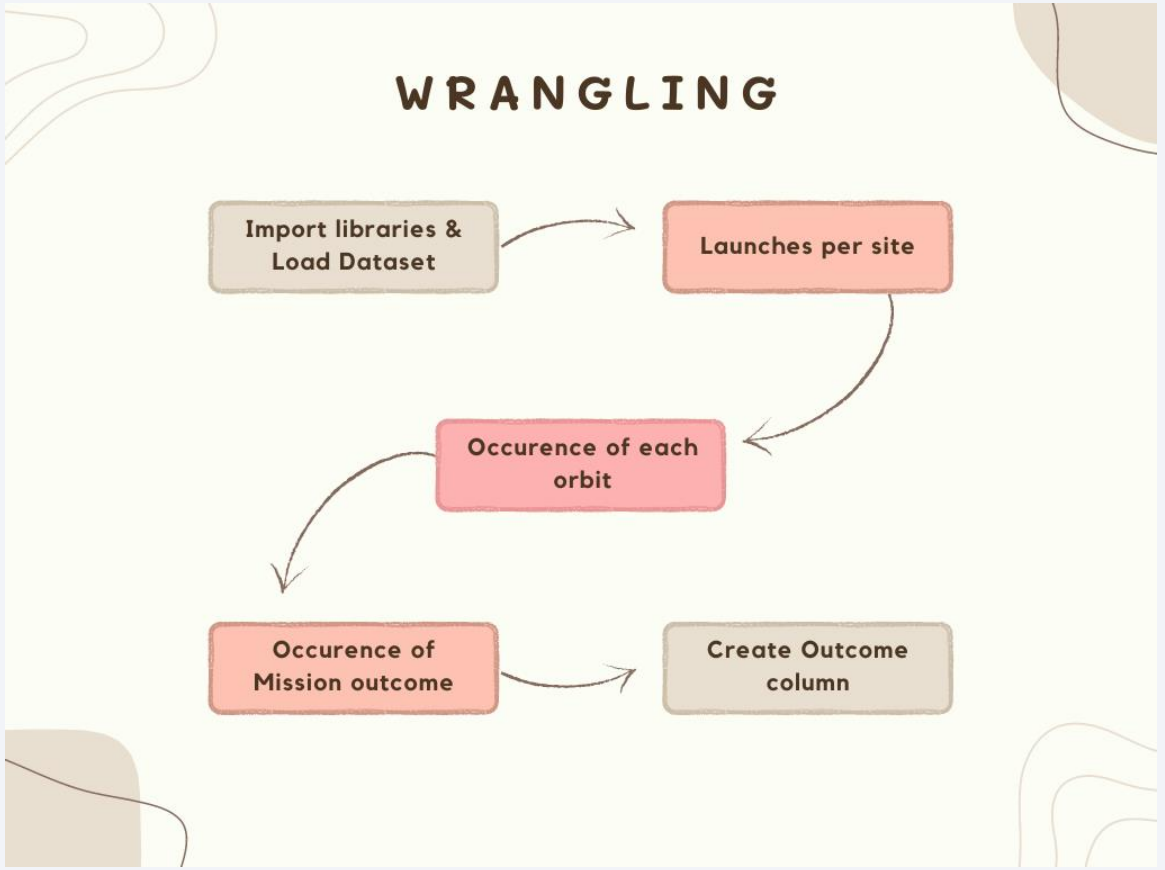
Data Collection - Scraping

- https://github.com/veritoalba/Applied-Data-Science-Capstone/blob/main/jupyter-labs-webscraping%20(1).ipynb



Data Wrangling

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



EDA with Data Visualization

- Scatterplot: comparing different variables and how their relationships affect the launch outcome.
- Bar chart to visualize the relationships between success rate of each orbit type.
- Scatterplot comparing different variables with orbit type and how this relationship affects launch outcome.
- Line chart depicting the success rate in a yearly basis.

- https://github.com/veritoalba/Applied-Data-Science-Capstone/blob/main/edadataviz%20(1).ipynb

EDA with SQL

- Display the unique launch sites' names
- Display 5 records where launch sites begin with "CCA"
- Display total Payload Mass
- Display average Payload Mass
- List date of first successful landing
- Total number of success/failure missions
- List Booster versions which carried maximum payload
- List month, failure landing, booster versions and launch sites in 2015
- Rank landing outcomes between 2010-06-04 and 2017-03-20 (descending order)
- https://github.com/veritoalba/Applied-Data-Science-Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

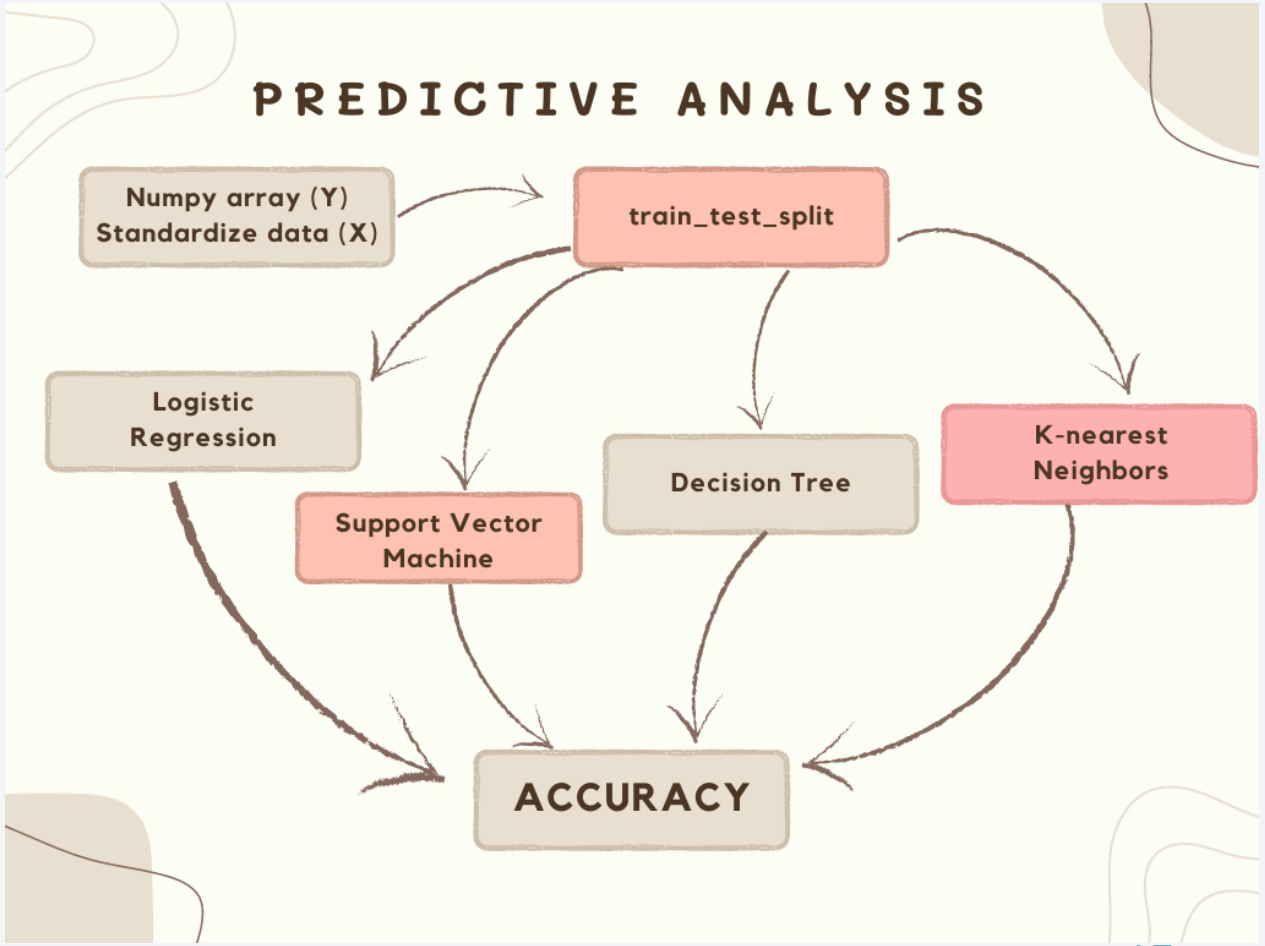
- We created and added circles and markers for each launch site. We created and added lines to measure the distance between each launch site and the closest railway, highway, coastlines and cities.
- We marked those objects to see if launch sites are in close proximity to those points of interest.
- https://github.com/veritoalba/Applied-Data-Science-Capstone/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- We added a Dropdown to enable launch site selection. We included a pie chart showing the total successful launches for all sites or the one selected in the Dropdown. We added a Slider to select the payload range and a scatter chart to show the correlation between payload and launch success. Summarize what plots/graphs and interactions you have added to a dashboard
- https://github.com/veritoalba/Applied-Data-Science-Capstone/blob/main/spacex_dash_app%20(1).py

Predictive Analysis (Classification)

- https://github.com/veritoalba/Applied-Data-Science-Capstone/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5%20(1).ipynb

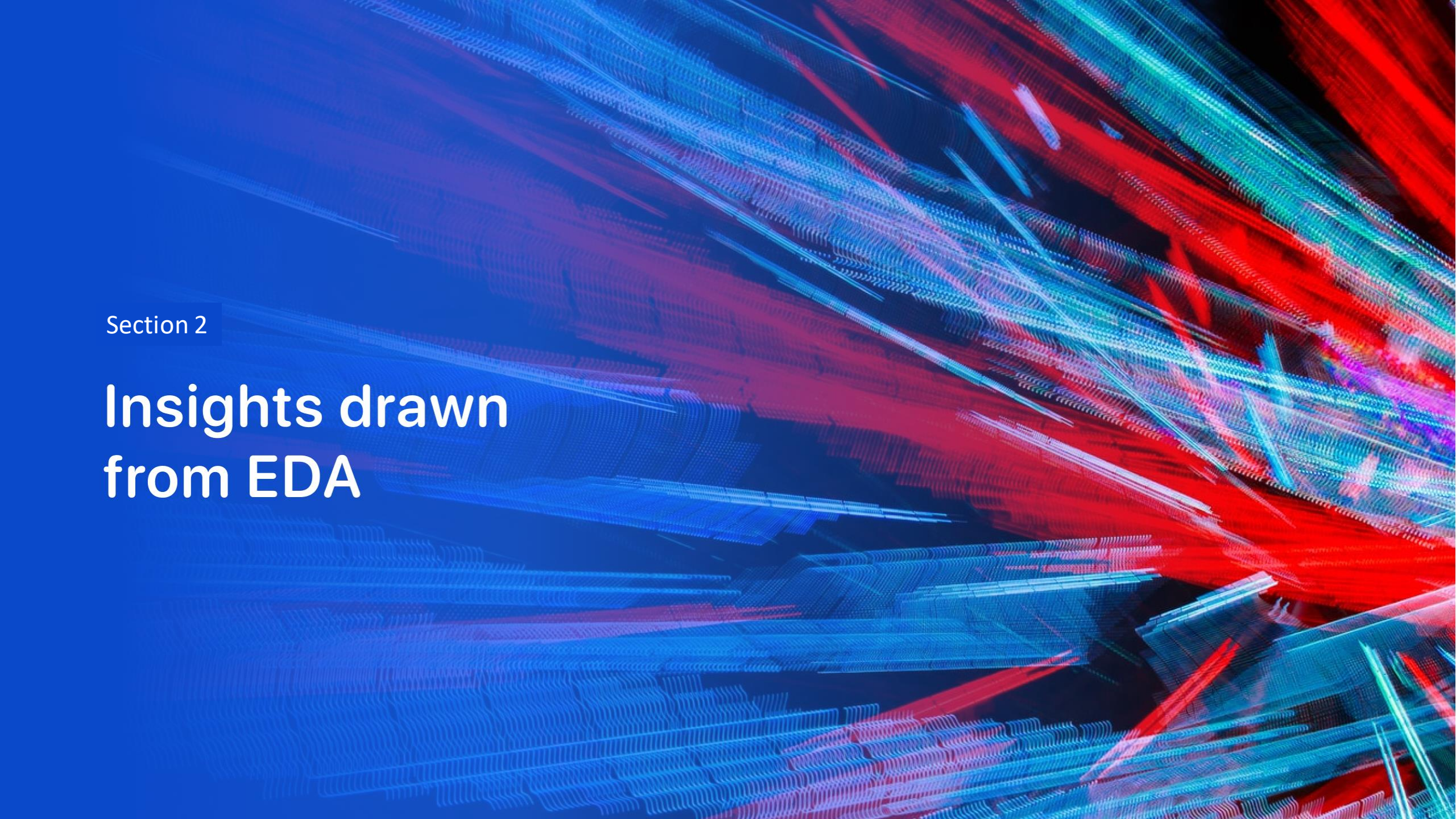


Results

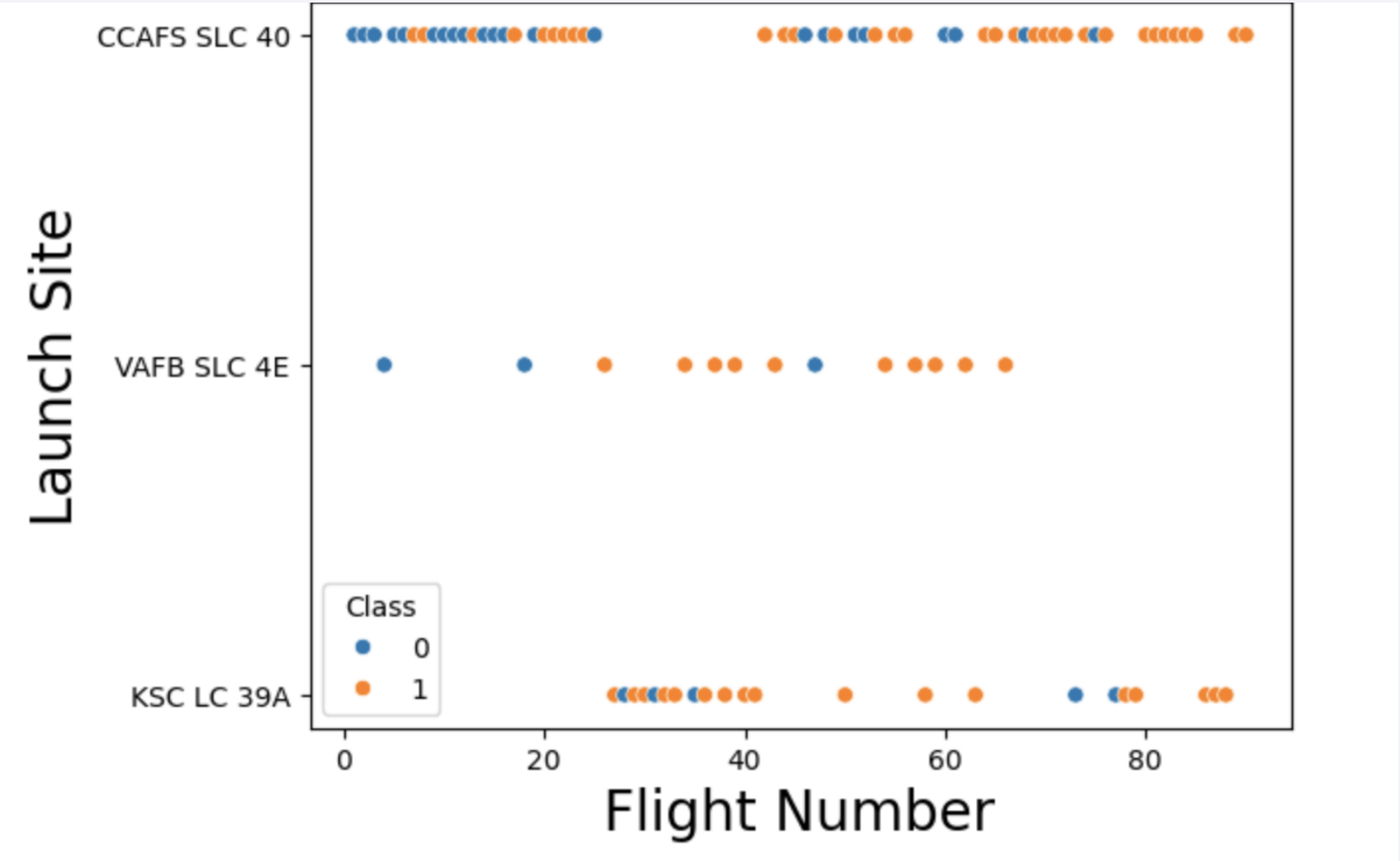
- The plots created during the Exploratory data analysis show that as the flight number increases, the first stage is more likely to land successfully and the more massive the payload mass, the less likely the first stage will return.
- We can also analyze which orbits have success rate and visualize the relationship between the launch site and the orbits and also the payload mass and the orbit type in the success of the missions.
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

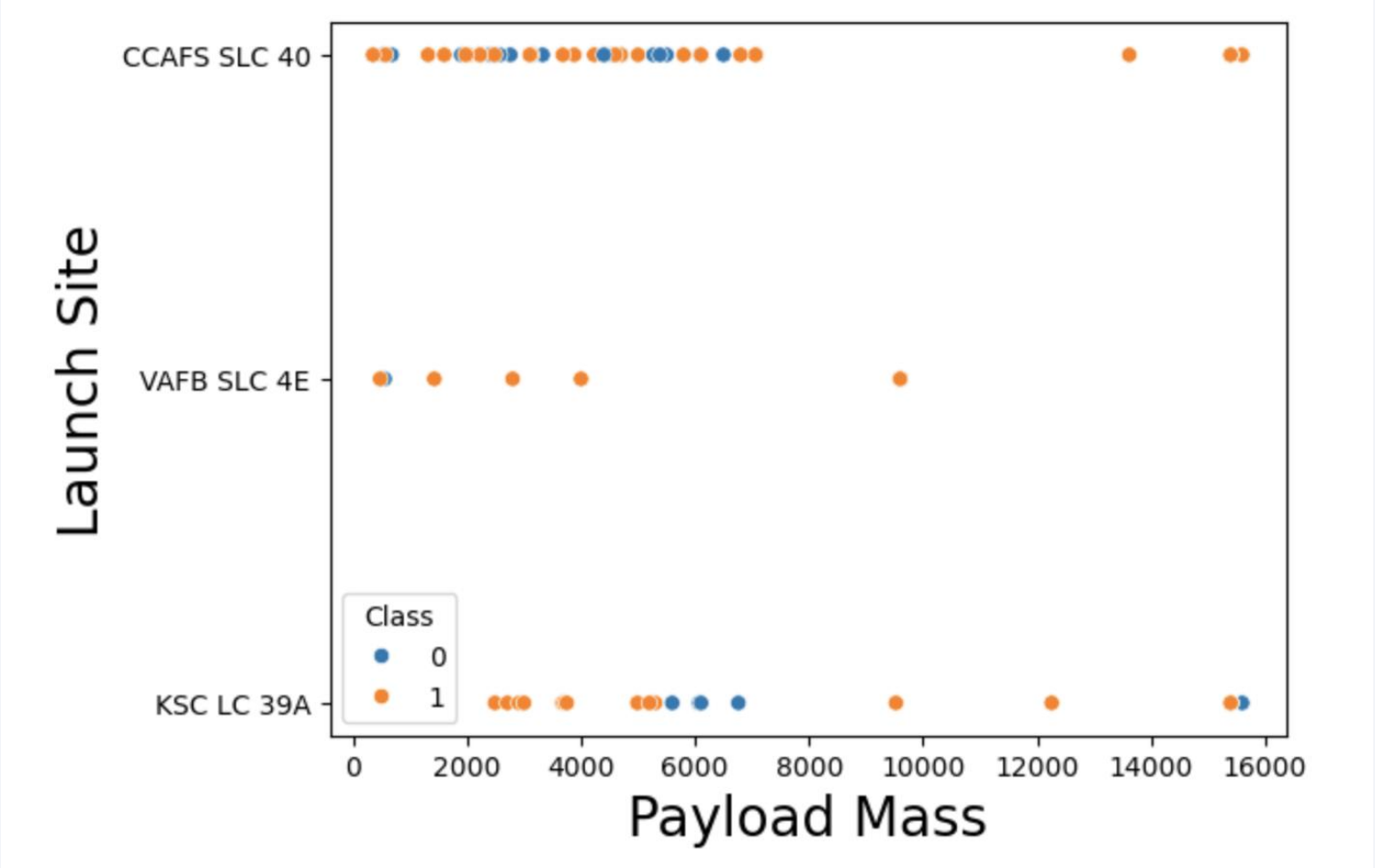
Insights drawn from EDA



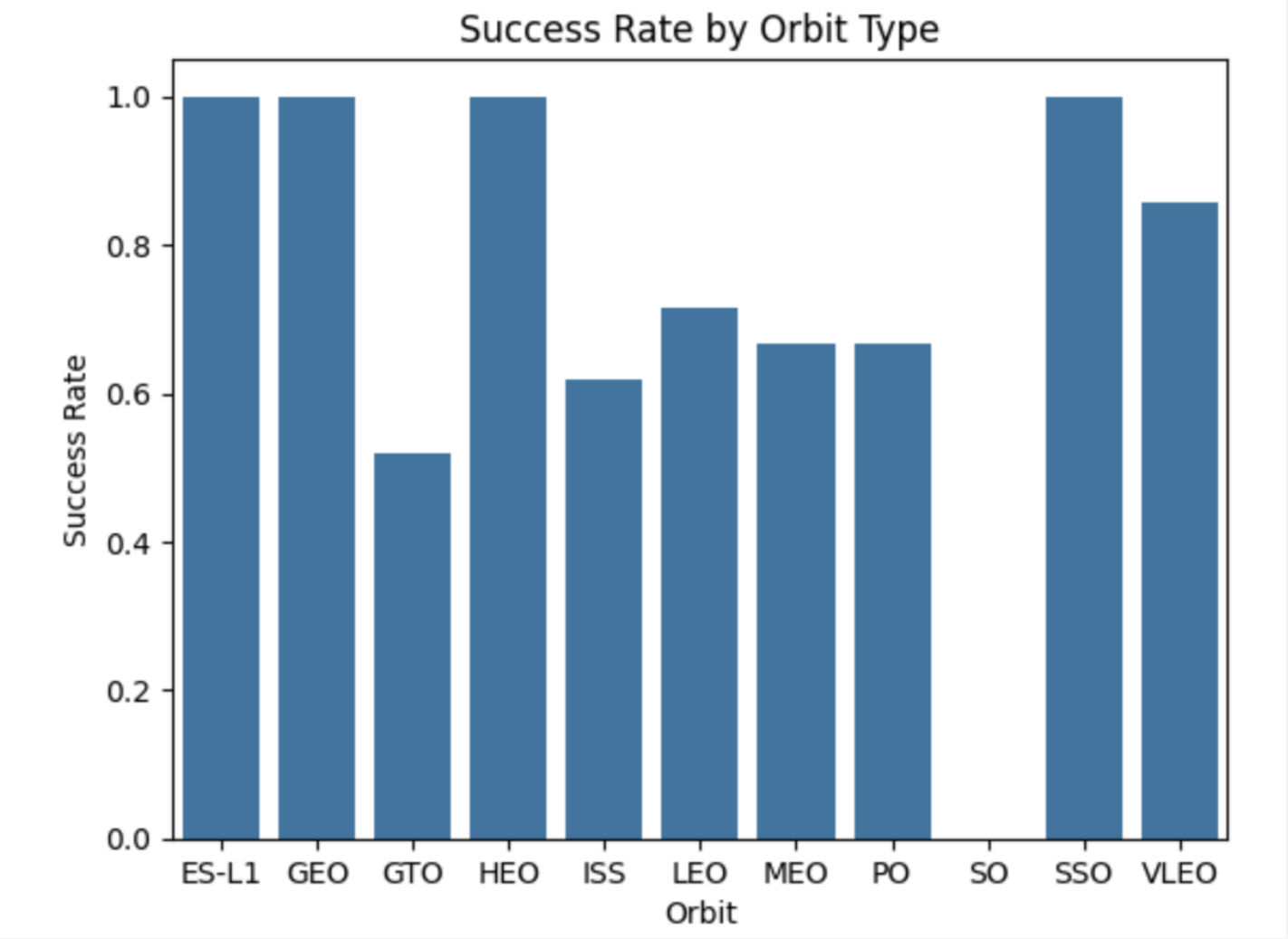
Flight Number vs. Launch Site



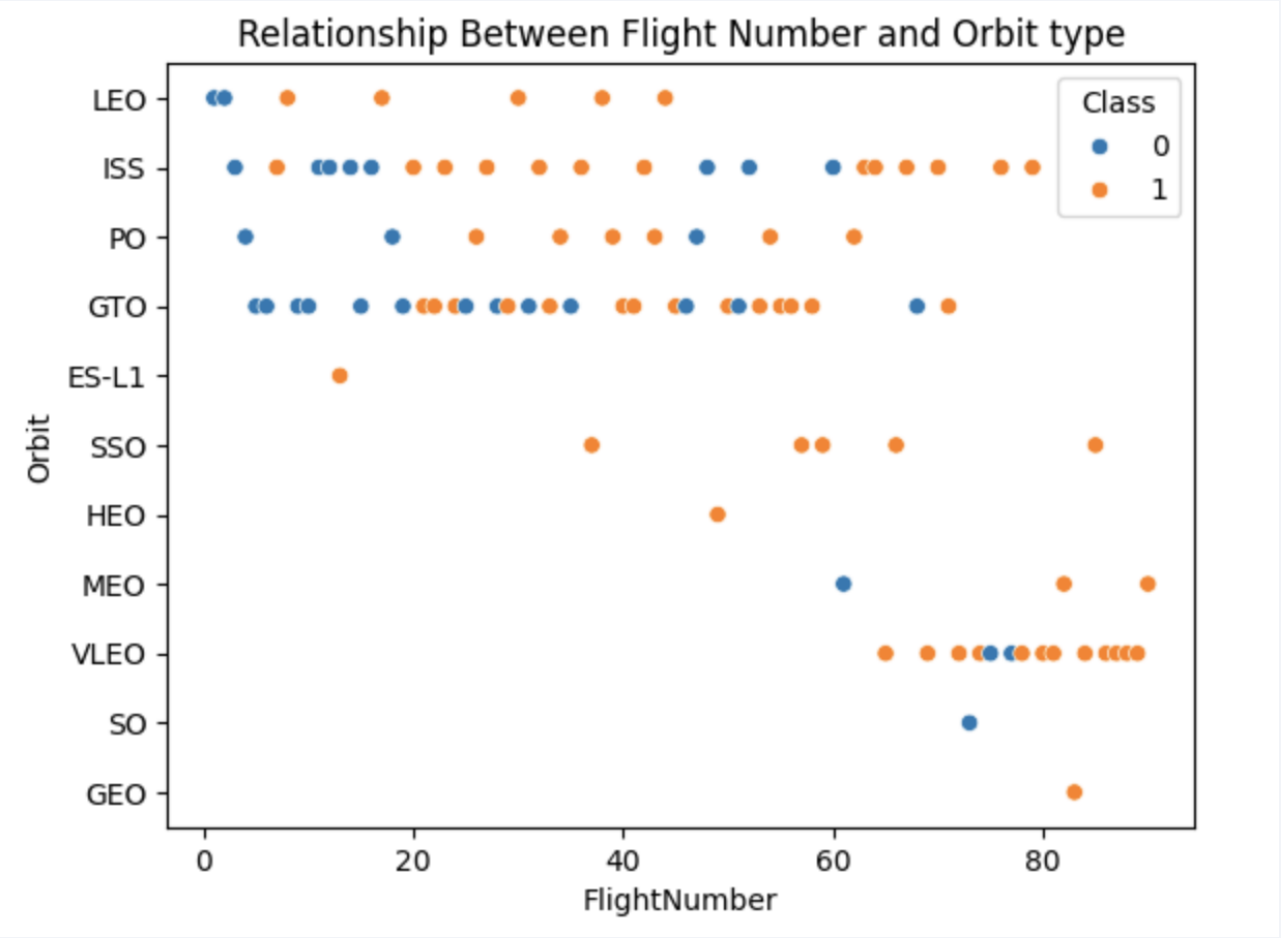
Payload vs. Launch Site



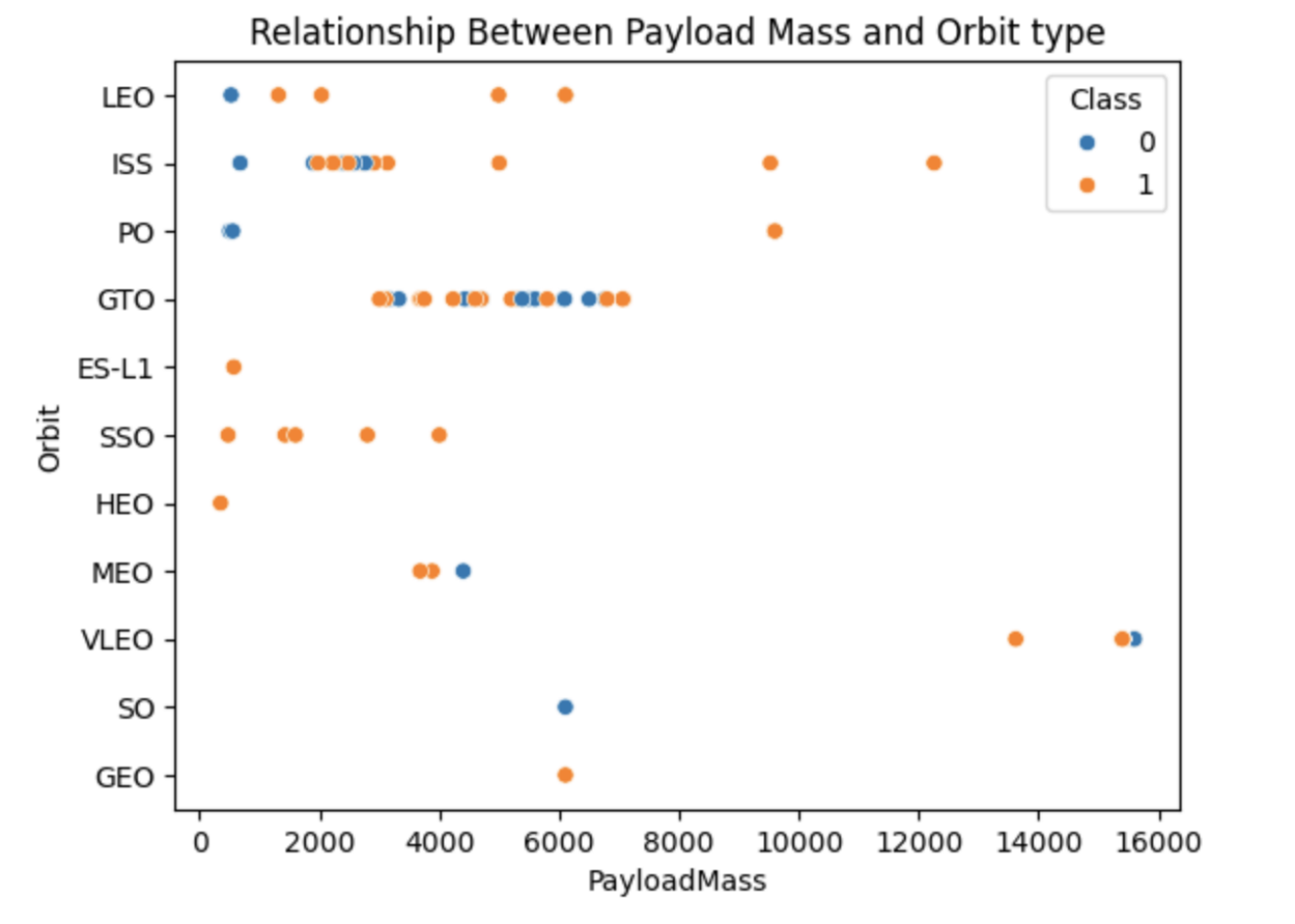
Success Rate vs. Orbit Type



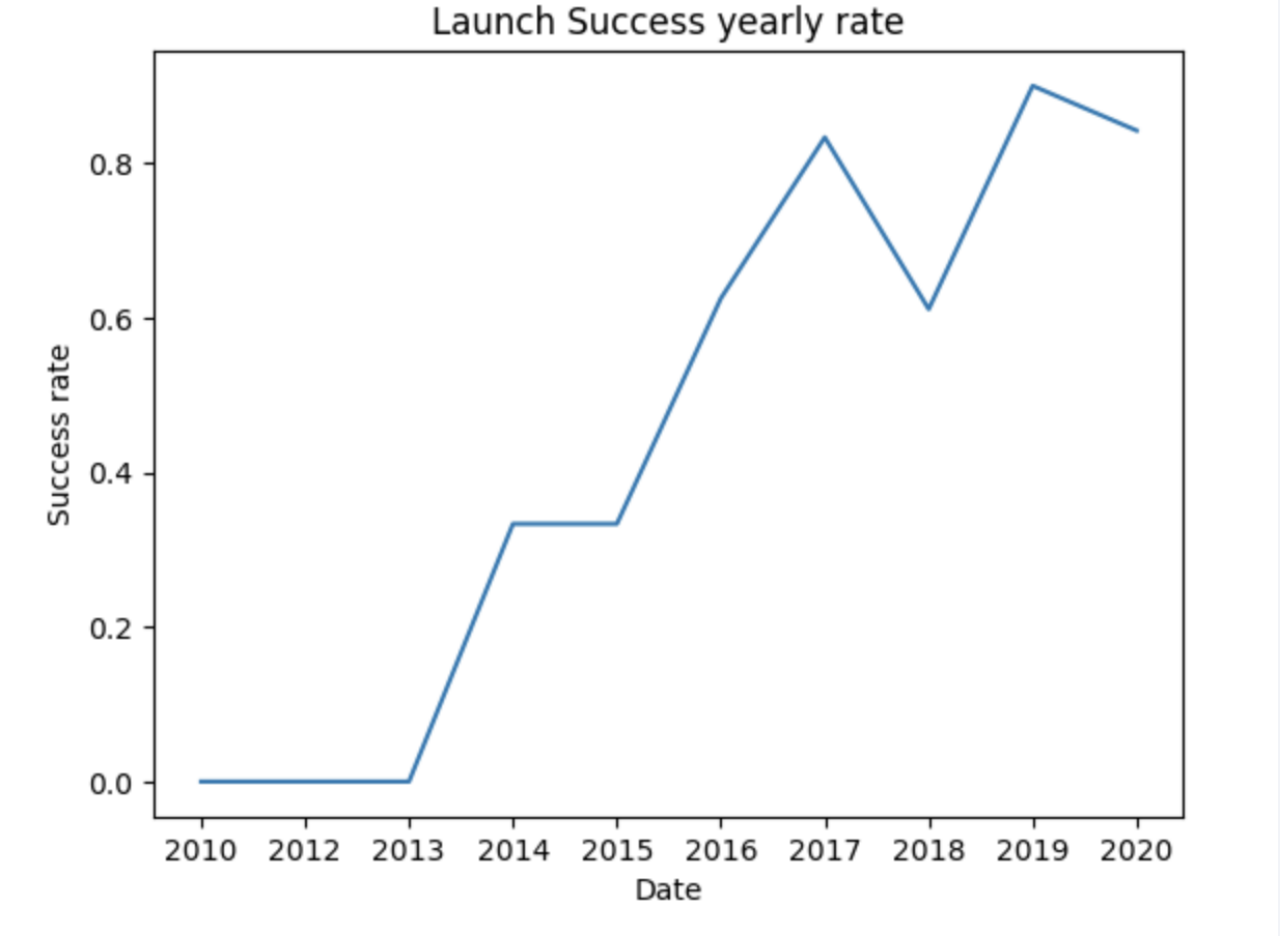
Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

- We use the SELECT DISTINCT query to group Launch Sites and avoid repetition.

```
%sql SELECT DISTINCT "Launch_Site" FROM SPACEXTABLE;
```

```
 * sqlite:///my_data1.db
Done.
```

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

Launch Site Names Begin with 'CCA'

- We include the condition "Launch_Site" LIKE (CCA%) to list the launch sites beginning in CCA and we LIMIT it to 5

```
%sql SELECT * FROM SPACEXTABLE WHERE "Launch_Site" LIKE ('CCA%') LIMIT 5;
```

```
 * sqlite:///my_data1.db
Done.
```

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcom
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Succes
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	Succes
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Succes
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Succes
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Succes

Total Payload Mass

- We SUM all the Payload Mass to obtain the Total Payload Mass where the customer was NASA.

```
%%sql SELECT SUM("PAYLOAD_MASS__KG_") AS "TOTAL PAYLOAD MASS" FROM SPACEXTABLE
WHERE "Customer" LIKE '%NASA (CRS)%';
```

```
 * sqlite:///my_data1.db
Done.
```

TOTAL PAYLOAD MASS
48213

Average Payload Mass by F9 v1.1

- We calculate the Average Payload Mass where the Booster version was F9 v1.1.

```
%%sql SELECT AVG("PAYLOAD_MASS__KG_") AS "AVERAGE PAYLOAD MASS" FROM SPACEXTABLE
WHERE "Booster_Version" LIKE "%F9 v1.1%";
```

```
 * sqlite:///my_data1.db
Done.
```

AVERAGE PAYLOAD MASS
2534.6666666666665

First Successful Ground Landing Date

- We obtained the date of the first successful landing outcome.

```
%%sql SELECT MIN("Date") AS "FIRST SUCCESSFULL OUTCOME" FROM SPACEXTABLE
WHERE "Landing_Outcome" LIKE "%Success%";
```

```
 * sqlite:///my_data1.db
Done.
```

FIRST SUCCESSFULL OUTCOME
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- According to this query, there are no successful Drone ship landing with Payload between 4000 and 6000

```
%%sql SELECT "Booster_Version" FROM SPACEXTABLE
WHERE "Mission_Outcome" LIKE "%Success%"
AND "PAYLOAD_MASS__KG_" > 4000 AND "PAYLOAD_MASS__KG" < 6000;
```

```
 * sqlite:///my_data1.db
Done.
```

Booster_Version

Total Number of Successful and Failure Mission Outcomes

- We grouped the total number of Success and Failure outcomes

```
%sql SELECT "Mission_Outcome", COUNT(*) AS "Total"FROM SPACEXTABLE GROUP BY "Mission_Outcome";
```

```
 * sqlite:///my_data1.db
Done.
```

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

Boosters Carried Maximum Payload

- We present a list of the booster versions that carried maximum Payload.

```
%%sql SELECT "Booster_Version" FROM SPACEXTABLE
WHERE "PAYLOAD_MASS__KG_" = (SELECT MAX("PAYLOAD_MASS__KG_")FROM SPACEXTABLE);
```

```
 * sqlite:///my_data1.db
Done.
```

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 Failure Records

- We filtered the months in 2015 which had a Failure in the Landing outcome.

```
%%sql

SELECT
    CASE SUBSTR("Date", 6, 2)
        WHEN '01' THEN 'January'
        WHEN '02' THEN 'February'
        WHEN '03' THEN 'March'
        WHEN '04' THEN 'April'
        WHEN '05' THEN 'May'
        WHEN '06' THEN 'June'
        WHEN '07' THEN 'July'
        WHEN '08' THEN 'August'
        WHEN '09' THEN 'September'
        WHEN '10' THEN 'October'
        WHEN '11' THEN 'November'
        WHEN '12' THEN 'December'
    END AS "Month",
    "Landing_Outcome",
    "Booster_Version",
    "Launch_Site"
FROM
    SPACEXTABLE
WHERE
    SUBSTR("Date", 0, 5) = '2015'
    AND "Landing_Outcome" LIKE '%Failure (drone ship)%';
```

```
 * sqlite:///my_data1.db
Done.
```

Month	Landing_Outcome	Booster_Version	Launch_Site
January	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
April	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

- We ranked in decreasing order, the total landing outcomes between 2010-06-04 and 2017-03-20

```
%%sql

SELECT
    "Landing_Outcome",
    COUNT(*) AS "Count"
FROM
    SPACEXTABLE
WHERE
    "Date" BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY
    "Landing_Outcome"
ORDER BY
    "Count" DESC;
```

```
 * sqlite:///my_data1.db
Done.
```

Landing_Outcome	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

Launch Sites

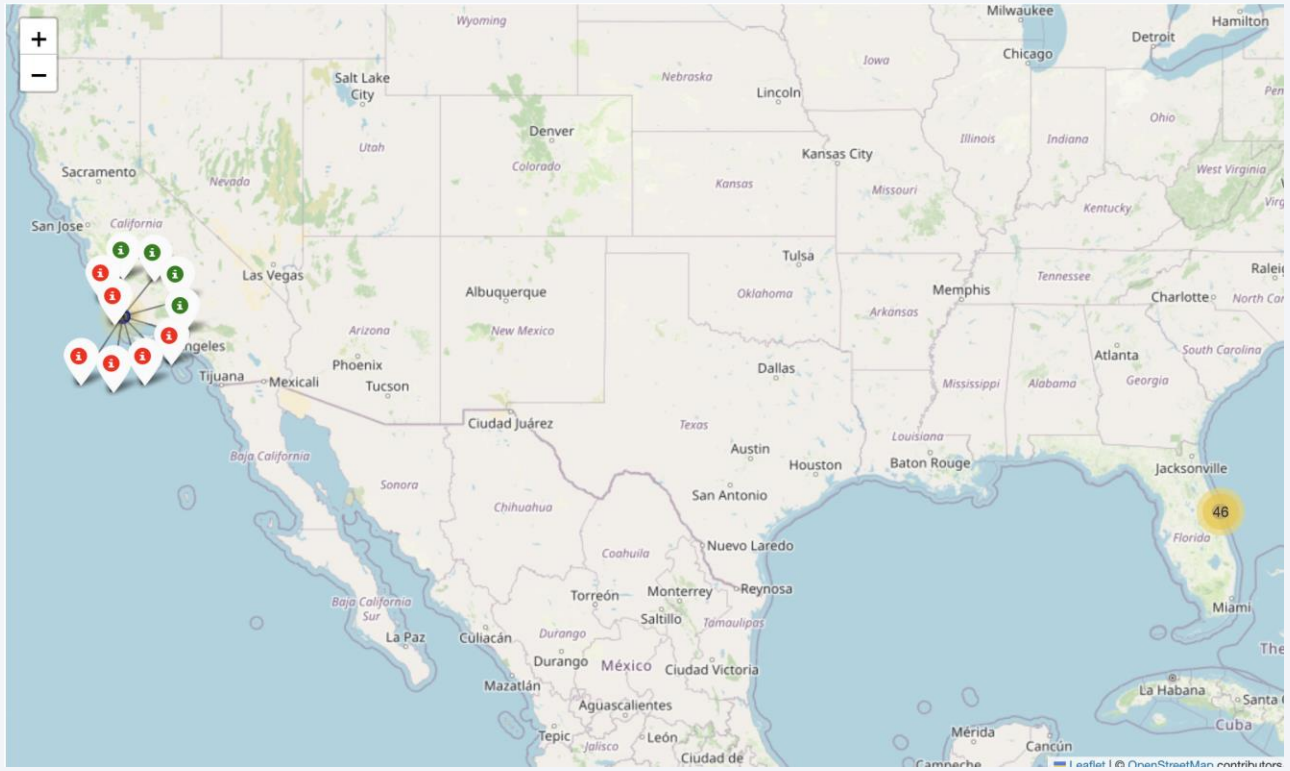
- We Marked the for launch sites in a map

Launch outcomes

- In this Launch Site we have a total of 13 launches, 3 of which were a failure (red) and the rest were successful (green).

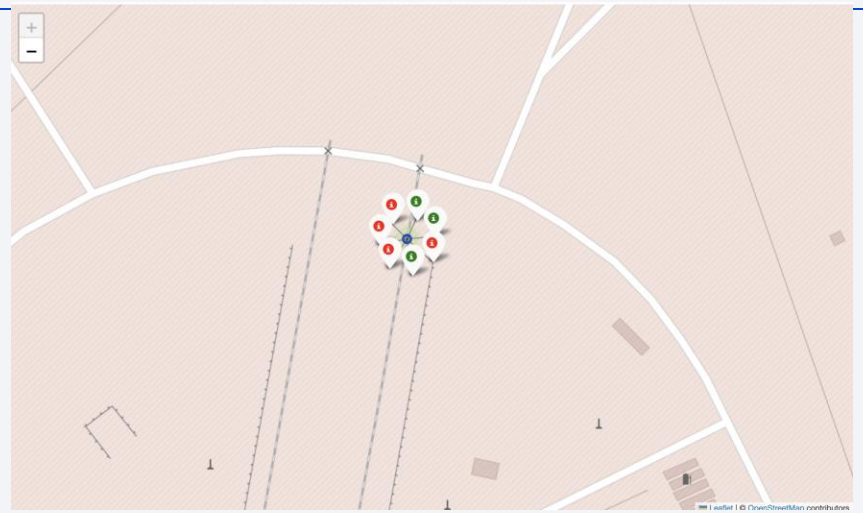
Launch outcomes

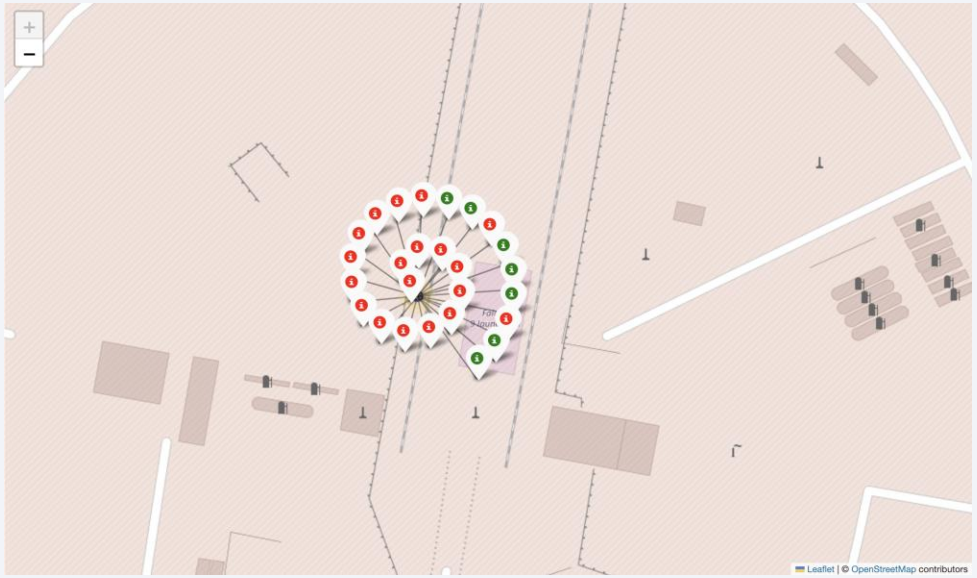
- In this Launch Site we have a total of 10 launches, 6 of which were a failure (red) and the rest were successful (green).



Launch outcomes

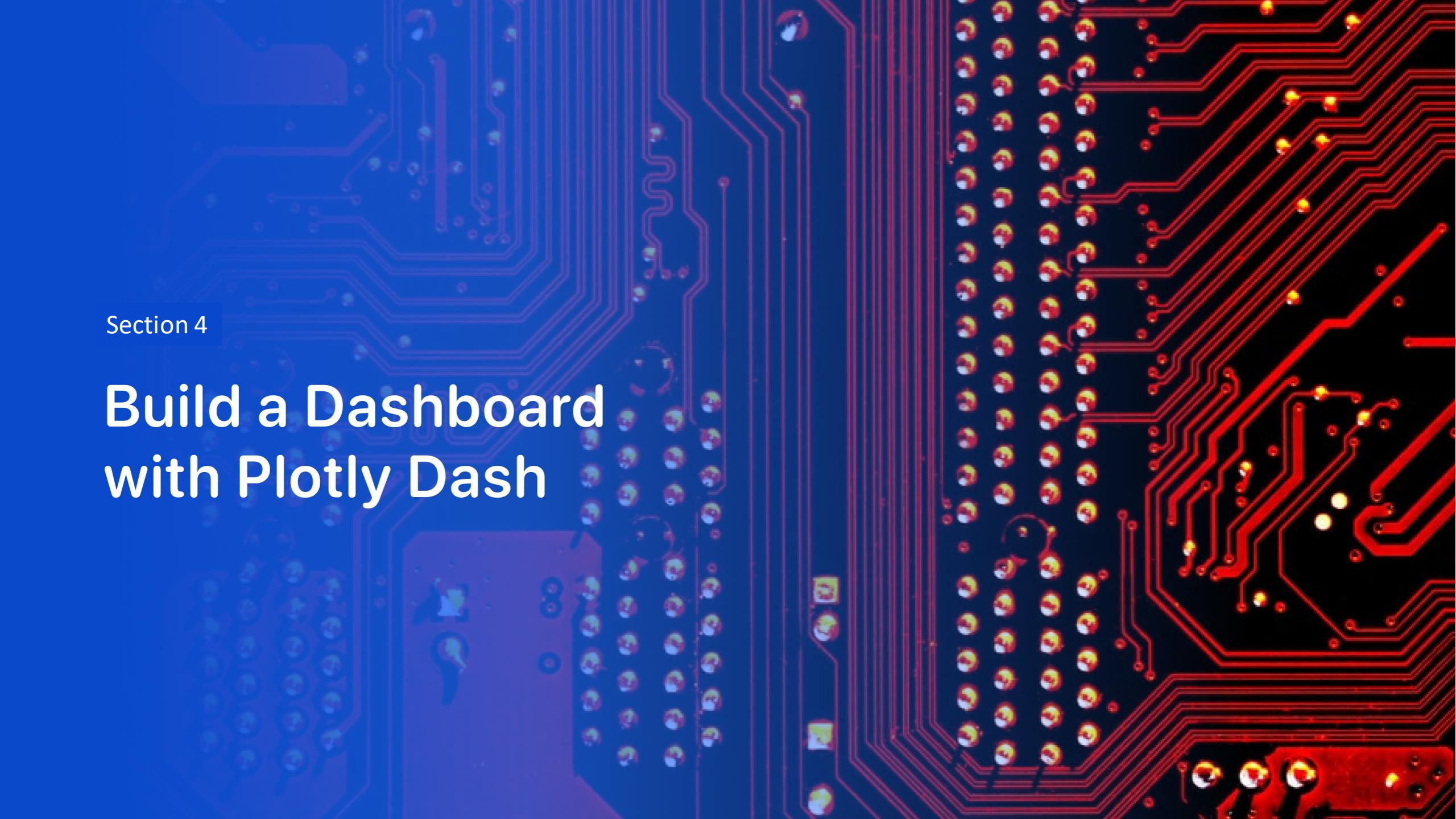
- In the first Launch Site we have a total of 7 launches. 4 were a failure (red) and the rest were successful (green).
- In the second Launch Site we have a total of 26 launches. 19 were a failure and the rest were successful.





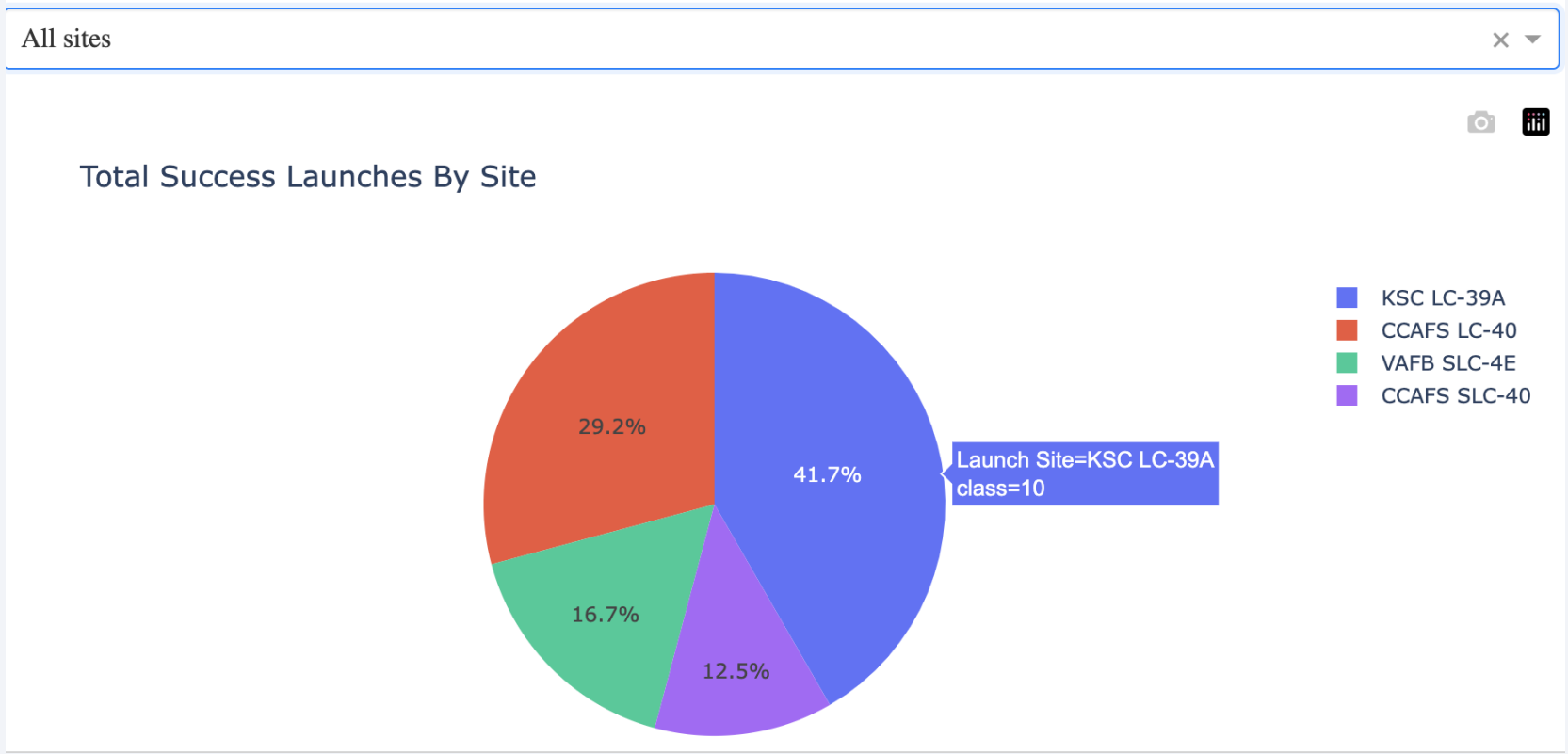
Closest points of interest

- Here we see two launch sites and their proximities to railway, highway, coastline marked with blue lines


Section 4
Build a Dashboard with Plotly Dash

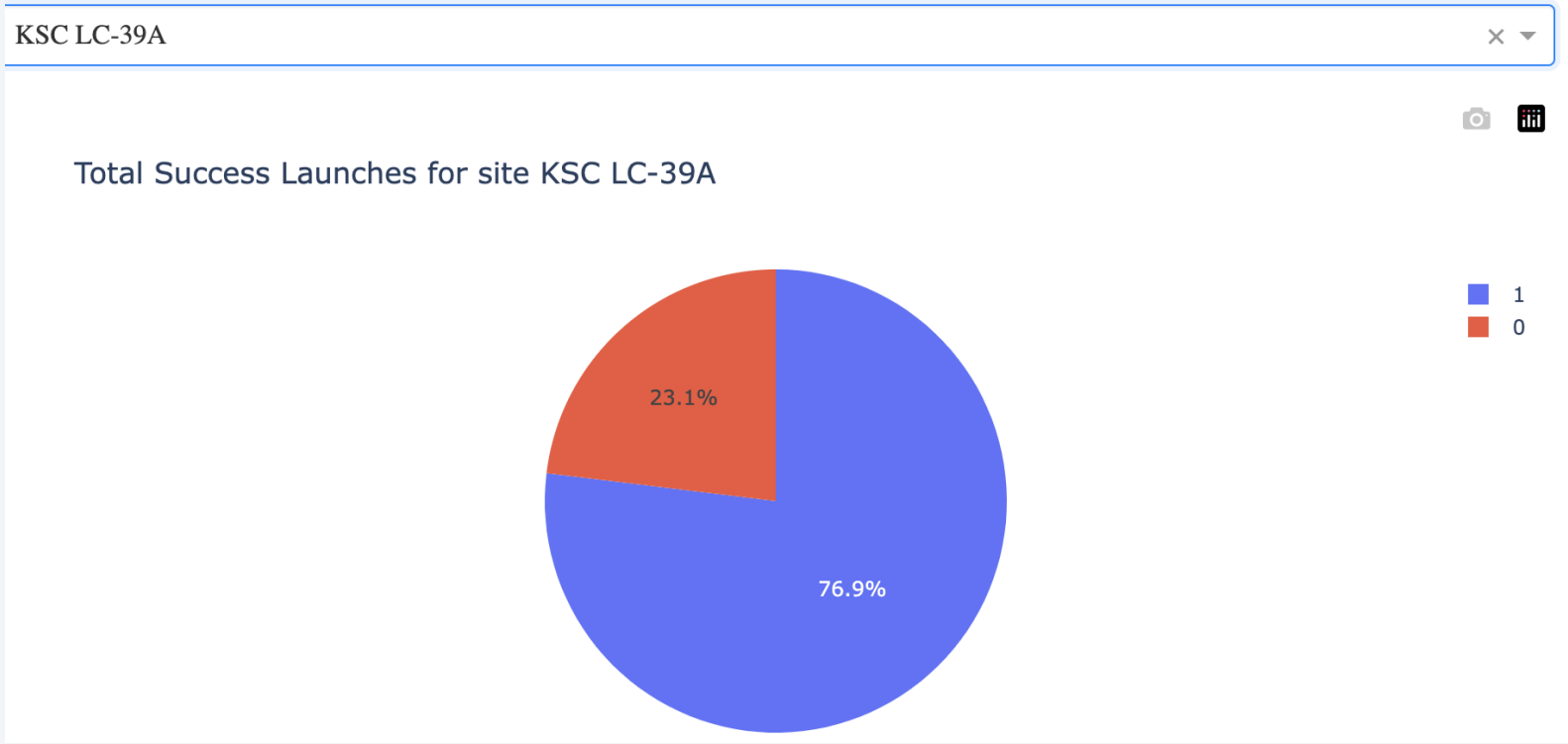
Launch Success count

- This pie chart shows the distribution of successful launches per launch site.

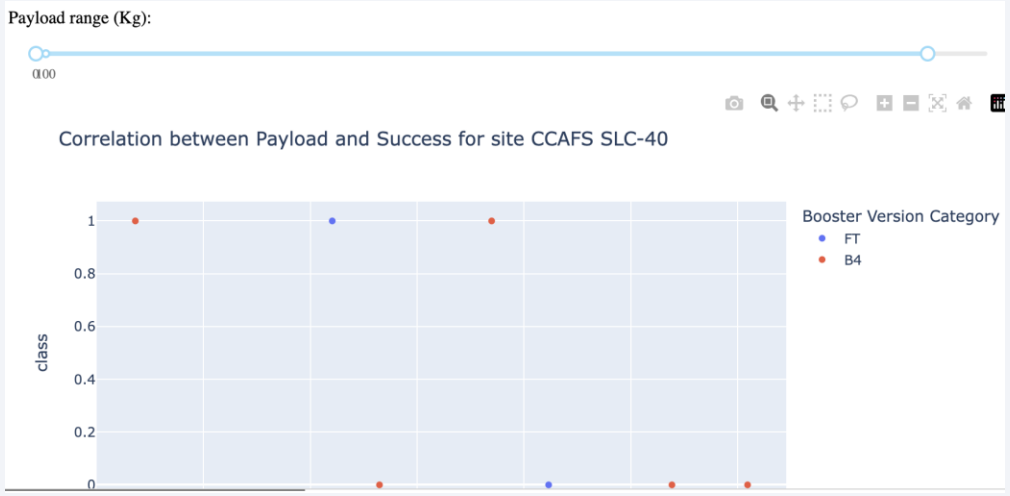


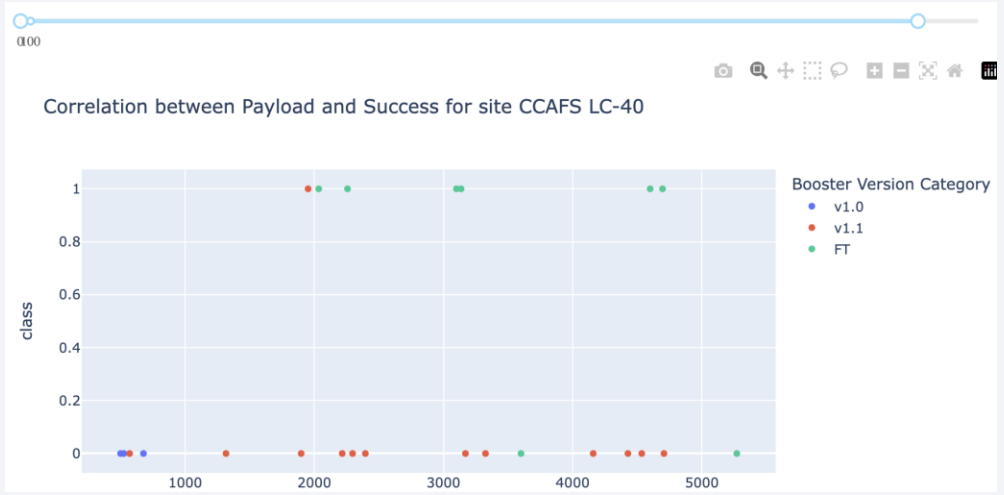
Highest Launch Success Ratio

The highest Launch Success Ratio was found in KSC LC-39A with 76.9% of success.



Payload vs. Launch Outcome per site





Section 5

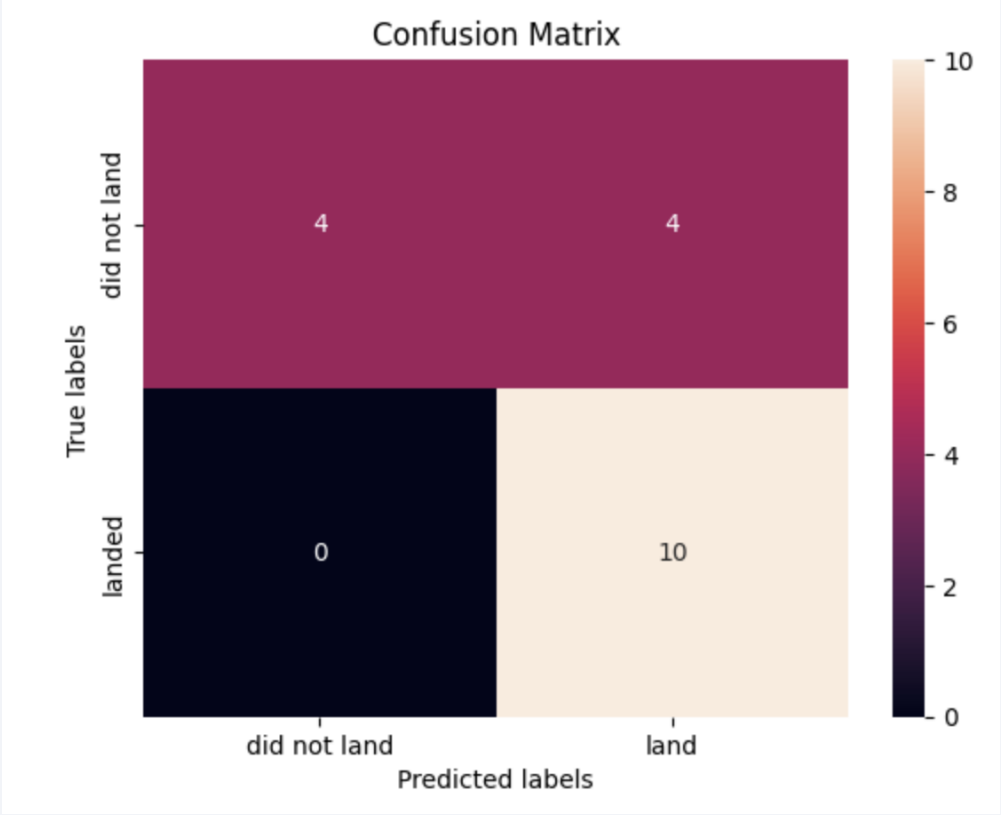
Predictive Analysis (Classification)

Classification Accuracy

```
print("Accuracy for Logistic Regression method is ", logreg_cv.score(X_test, Y_test))
print("Accuracy for Support Vector Machine method is ", svm_cv.score(X_test, Y_test))
print("Accuracy for Decission Tree method is ", tree_cv.score(X_test, Y_test))
print("Accuracy for K-nearest Neighbors is ", knn_cv.score(X_test, Y_test))
```

```
Accuracy for Logistic Regression method is  0.7777777777777778
Accuracy for Support Vector Machine method is  0.7777777777777778
Accuracy for Decission Tree method is  0.8333333333333334
Accuracy for K-nearest Neighbors is  0.7777777777777778
```

Confusion Matrix- Decision Tree



Conclusions

- The higher the flight number, the higher the success in launch site CCAFS SL40.
- The lower the payload mass, the higher number of success launches.
- The most successful orbit types are ES-L1, GEO, HEO and SSO.
- KSC-LC 39A has the highest count of successful launches with a ratio of 76.9%.
- The best Predictive analysis is made with the Decision Tree method.

Thank you!