IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Trey Murray
04/04/2024

Outline

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

Executive Summary

- Summary of methodologies

  In this project the following methodologies were used:

  Data Collection.

  Data Cleaning

  Data Exploration

  Data Visualisation

  Predictive Analyisis

- Summary of all results

  The results of this project is a predictive model to determine if the first stage was reused.

Introduction

- Project background and context
- Problems you want to find answers

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - The data was collected through the Space X data API
- Perform data wrangling
  - The data was processed by looking for and filling in missing values.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

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

GITHUB URL:
https://github.com/TreyMurray01/IBM-DS-Specialisation-Capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping

- https://github.com/TreyMurray01/IBM-DS-Specialisation-Capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Place your flowchart of web scraping here

Data Wrangling

- Describe how data were processed
- You need to present your data wrangling process using key phrases and flowcharts
- https://github.com/TreyMurray01/IBM-DS-Specialisation-Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- This part of the project used scatter plots, category plots and bar charts.
  - The scatter plots were chosen to see if there was any relationship between two variables.
  - The category plots were chosen to
- https://github.com/TreyMurray01/IBM-DS-Specialisation-Capstone/blob/main/jupyter-labs-eda-dataviz.ipynb

EDA with SQL

- Using bullet point format, summarize the SQL queries you performed
- https://github.com/TreyMurray01/IBM-DS-Specialisation-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
- Map Objects Used:
  - Markers
  - Circles
  - Lines
  - Marker Clusters
- Markers were added to indicate the location of the lunch sites
- Circles were used to highlight the area of the launch sites
- Marker Clusters were used to group the markers of the same location
- A line was used to show the distance to the closest coastline.
- GITHUB URL:

Build a Dashboard with Plotly Dash

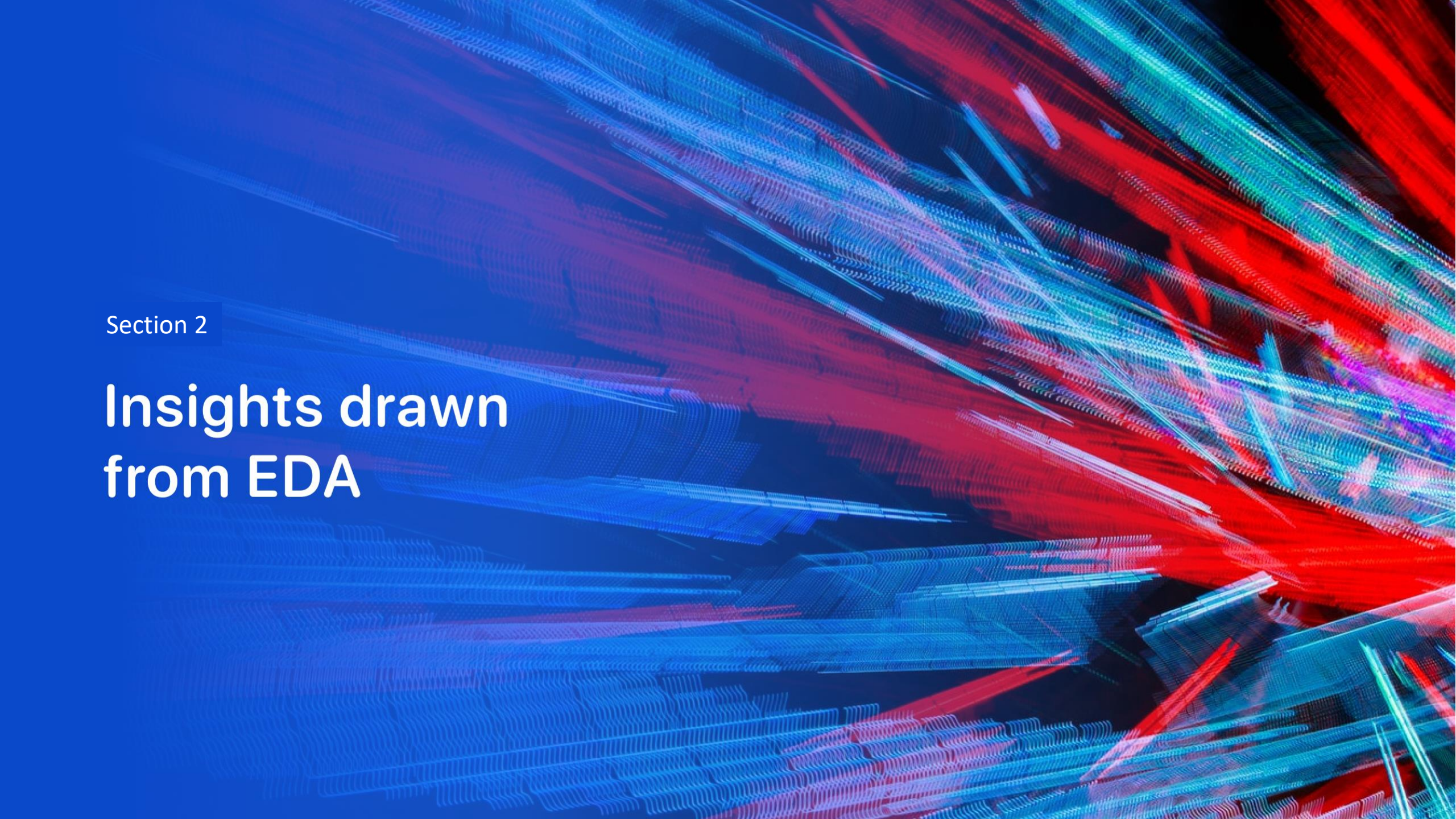
- Plots Used:
  - Pie Chart
  - Scatter Plot
  - Interactions:
    - Slider
    - Drop down menus
- The pie chart was chosen to show the distribution between successful and failed launches. The scatter plot was chosen to display the correlation between successful launches and the payload. The range slider was used to filter the payload and the dropdown menus were used to filter by launch sites.
- GITHUB URL:

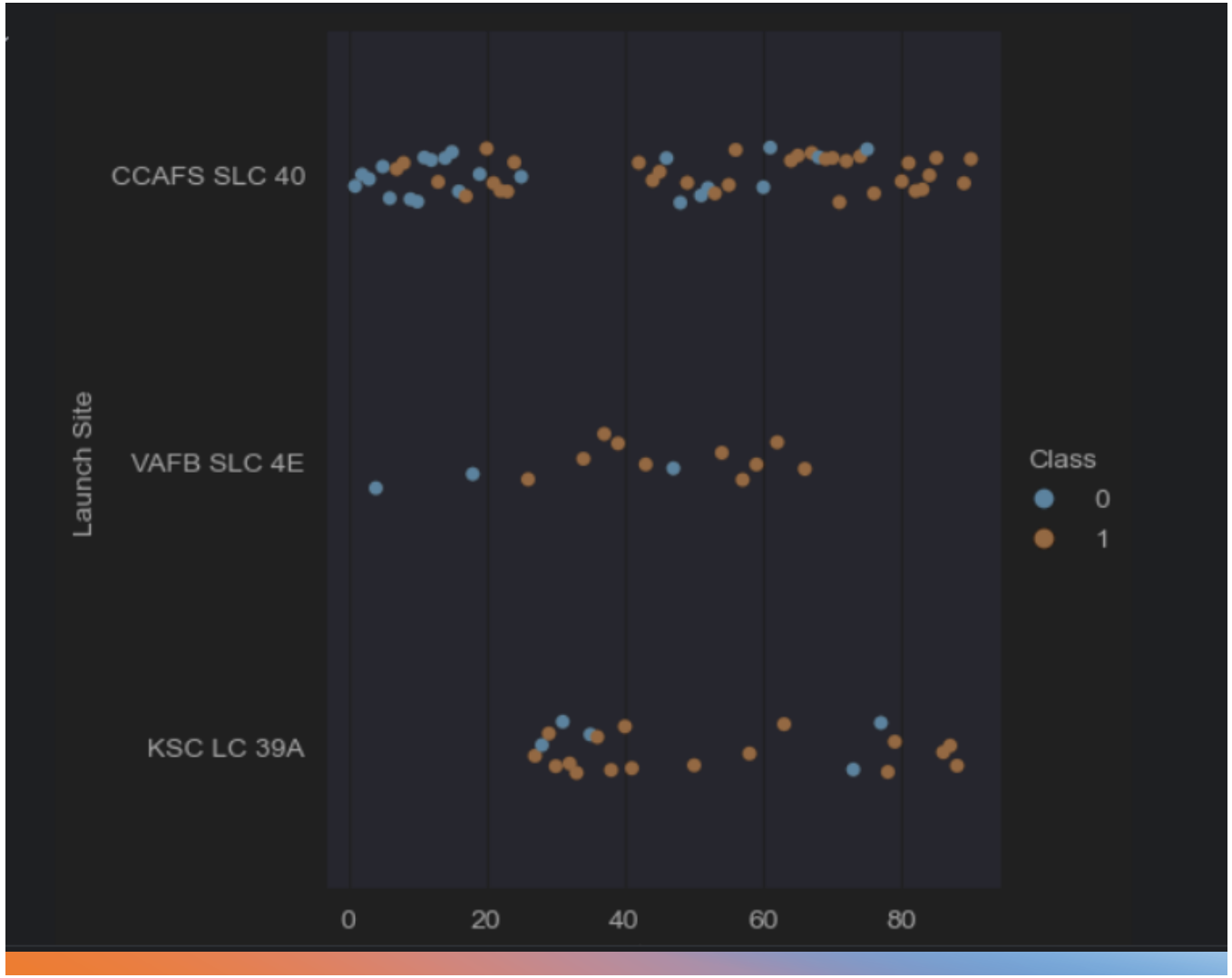
Predictive Analysis (Classification)

- To build the best performing model we first took the features and split them into the train and test set. Then using these sets we built various models: regression, decision tree, SVM, KNN. And used grid search to train and find the best parameters. Lastly comparing the results to get the best performing model.
- FLOW CHART
- GITHUB URL:

Results

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


Section 2
Insights drawn from EDA

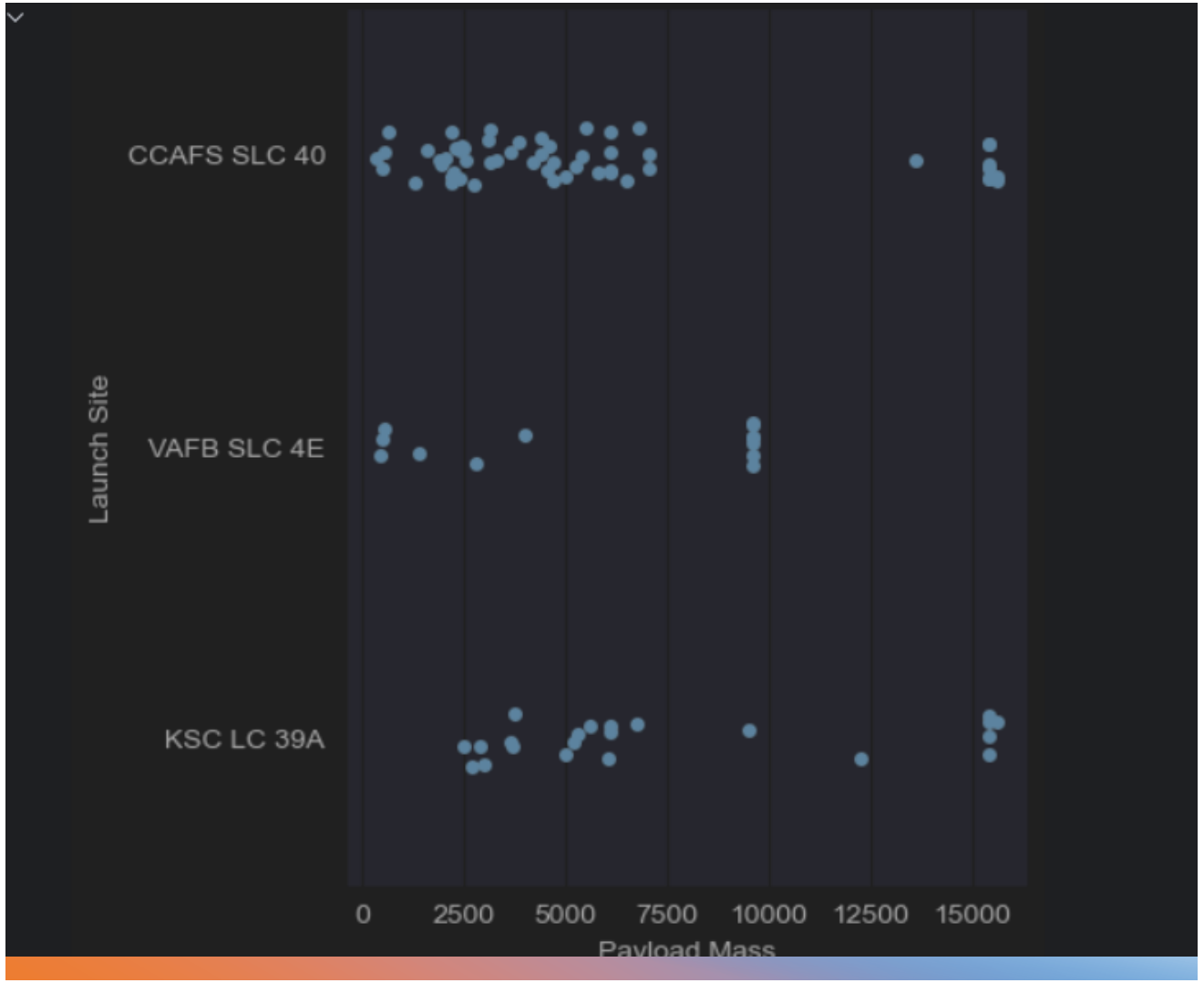


Flight Number vs. Launch Site

- As the flight number increases so does the number of successful landings.

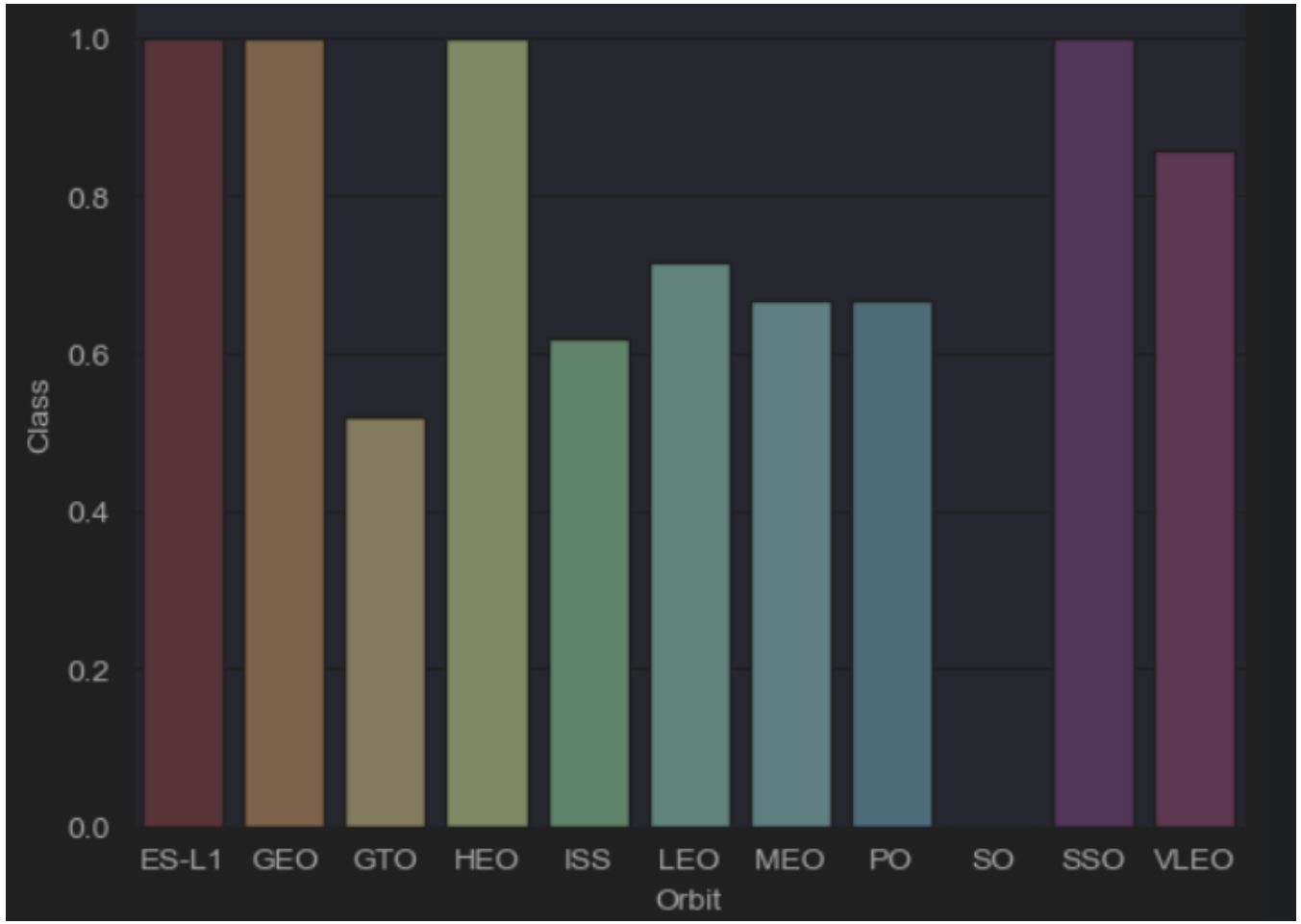
Payload vs. Launch Site

- Site VAFB SLC 4E is the sole site hasn’t handled payloads masses in excess of 10000 kg.



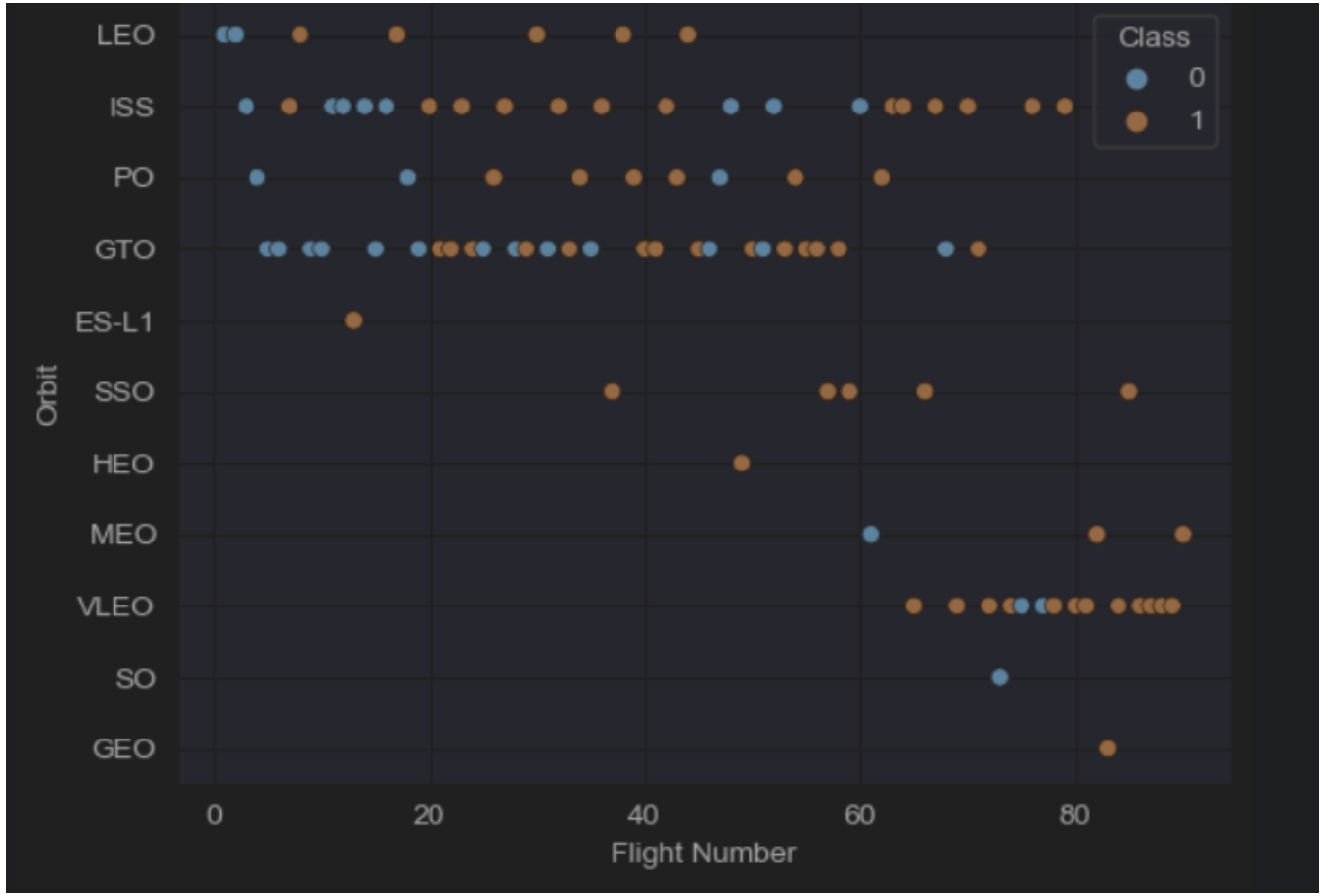
Success Rate vs. Orbit Type

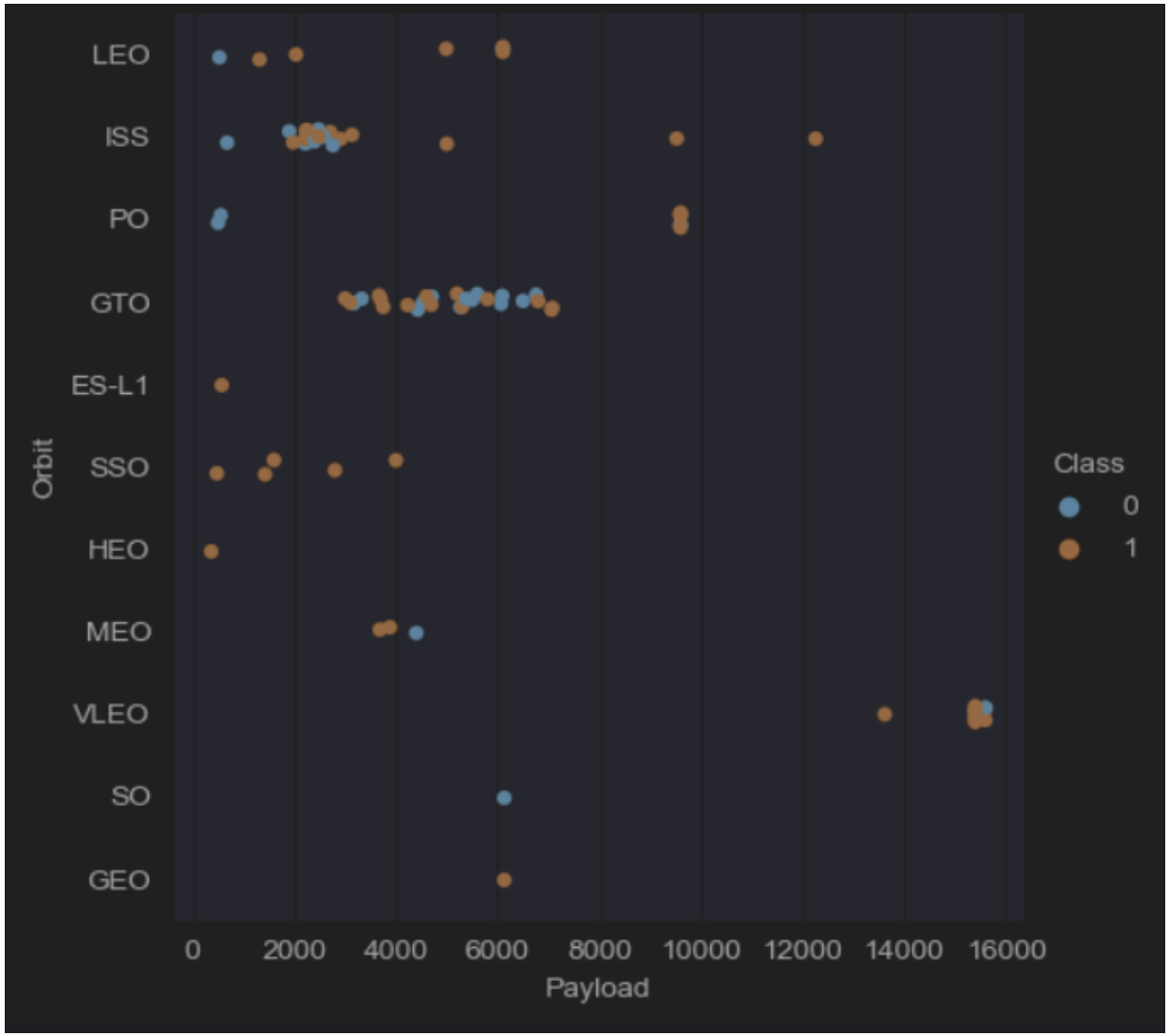
- There are four types of orbit with a 100% success rate (ES-L1, GEO, SSO, VLO) followed by VLEO.



Flight Number vs. Orbit Type

- Orbits LEO, ISS,GTO followed by VLEO has the most flights
- Orbits with near 100% success rates as well as fail rates with the have to few flights.





Payload vs. Orbit Type

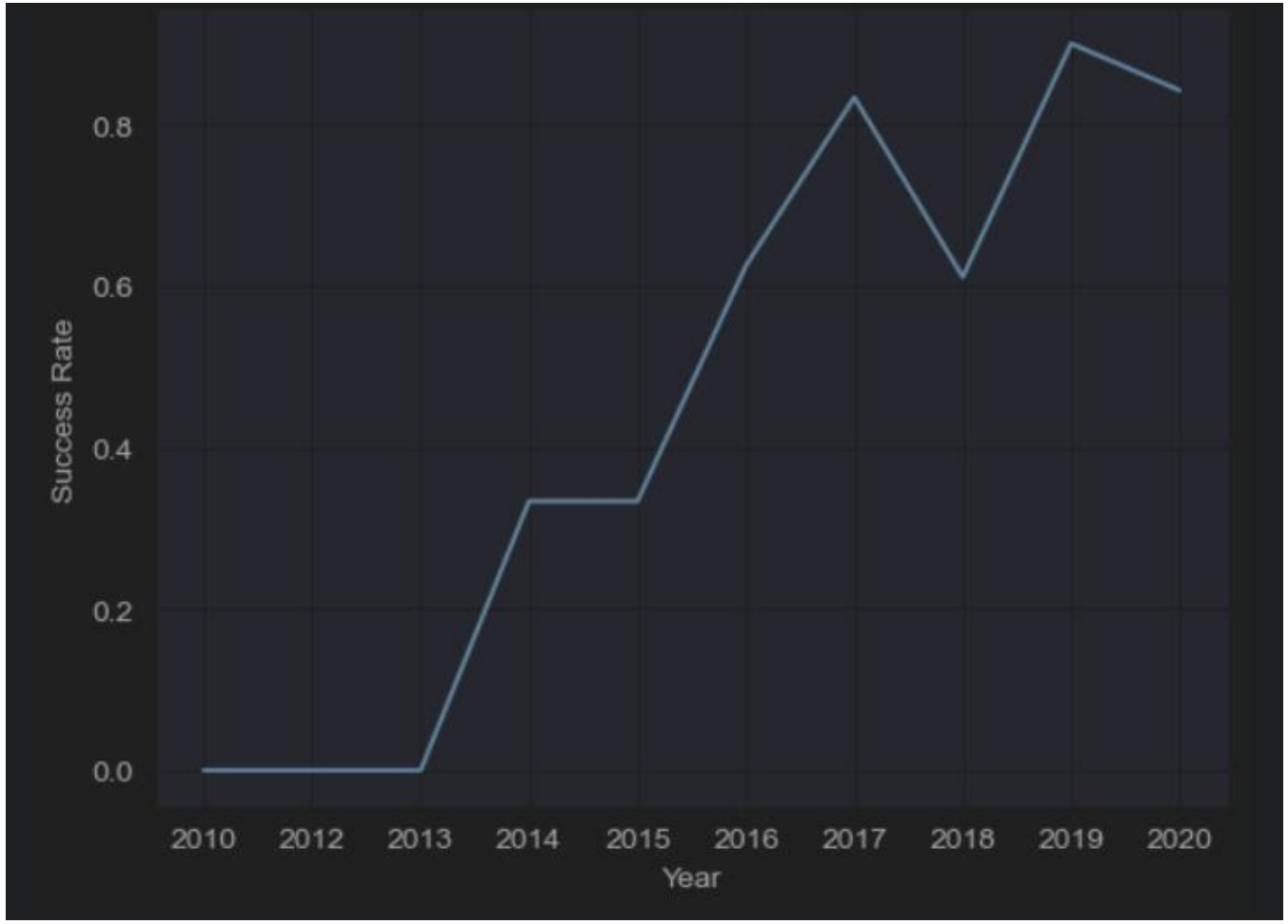
- Show a scatter point of payload vs. orbit type

- Show the screenshot of the scatter plot with explanations

Launch Success Yearly Trend

- Show a line chart of yearly average success rate

- Show the screenshot of the scatter plot with explanations



All Launch Site Names

- Find the names of the unique launch sites
- Present your query result with a short explanation here

Launch Site Names Begin with 'CCA'

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

```
[('2010-06-04', '18:45:00', 'F9 v1.0  B0003', 'CCAFS LC-40', 'Dragon Spacecraft Qualification Unit', 0, 'LEO', 'SpaceX', 'Success', '
 ('2010-12-08', '15:43:00', 'F9 v1.0  B0004', 'CCAFS LC-40', 'Dragon demo flight C1, two CubeSats, barrel of Brouere cheese', 0, 'LE(
 ('2012-05-22', '7:44:00', 'F9 v1.0  B0005', 'CCAFS LC-40', 'Dragon demo flight C2', 525, 'LEO (ISS)', 'NASA (COTS)', 'Success', 'No
 ('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')]
```

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

- In the above query, when executed it returns the first 5 records that begins with ‘CCA’

Total Payload Mass

```
%sql SELECT "Customer", SUM("PAYLOAD_MASS__KG_") FROM SPACEXTABLE  WHERE "Customer" = 'NASA (CRS)';
```

- From the above query the total payload mass from NASA is 45596 KG

```
[('NASA (CRS)', 45596)]
```

Average Payload Mass by F9 v1.1

```
%sql SELECT "BOOSTER_VERSION", AVG("PAYLOAD_MASS__KG_") AS AVERAGEPAYLOAD FROM "SPACEXTABLE" WHERE "BOOSTER_VERSION" LIKE 'F9 v1.1%'
```

- From the query above the average payload was found to be 2534.67
- The Query calculates the average of the Payload Mass KG column after filtering by Booster Version where the booster version begins with “F9 v1.1”.

First Successful Ground Landing Date

- The first successful landing was dated 2018-07-22.

```
%sql SELECT MIN("Date") FROM "SPACEXTABLE" WHERE "LANDING_OUTCOME" = 'Success';
```

- In the above query it selects the oldest date where the landing outcome is “Success”.

```
%sql  SELECT DISTINCT "Booster_Version" FROM SPACEXTABLE WHERE "PAYLOAD_MASS__KG_" > 4000 AND "PAYLOAD_MASS__KG_" < 6000
```
Executed at 2024.04.11 14:10:38 in 48ms

Successful Drone Ship Landing with Payload between 4000 and 6000

- F9 v1.1', 'F9 v1.1 B1011', 'F9 v1.1 B1014', 'F9 v1.1 B1016', 'F9 FT B1020', 'F9 FT B1022', 'F9 FT B1026', 'F9 FT B1030', 'F9 FT B1021.2’, 'F9 FT B1032.1', 'F9 B4 B1040.1’, 'F9 FT B1031.2’, 'F9 B4 B1043.1’, 'F9 FT B1032.2’ , 'F9 B4 B1040.2’, 'F9 B5 B1046.2', 'F9 B5 B1047.2’ ,'F9 B5B1054’, 'F9 B5 B1048.3’, 'F9 B5 B1051.2 ‘, 'F9 B5B1060.1','F9 B5 B1058.2 ','F9 B5B1062.1’
- In the above query, the Booster version is selected where the payload mass is between the set range.

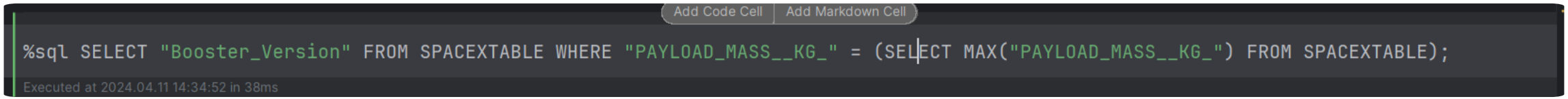
```
[('Failure (in flight)', 1),
 ('Success', 98),
 ('Success ', 1),
 ('Success (payload status unclear)', 1)]
```

```
%sql SELECT "MISSION_OUTCOME", COUNT("MISSION_OUTCOME") FROM SPACEXTABLE GROUP BY "MISSION_OUTCOME"

Executed at 2024.04.11 11:44:41 in 162ms
```

Total Number of Successful and Failure Mission Outcomes

- There is 100 flights that is classified as successful and one as failure.
- The above query gets the count of each of the mission outcome.



Boosters Carried Maximum Payload

- '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 ‘
- In the above query a subquery is run to get the maximum payload then using the result all booster that carried that payload is selected.

2015 Launch Records

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

- Present your query result with a short explanation here

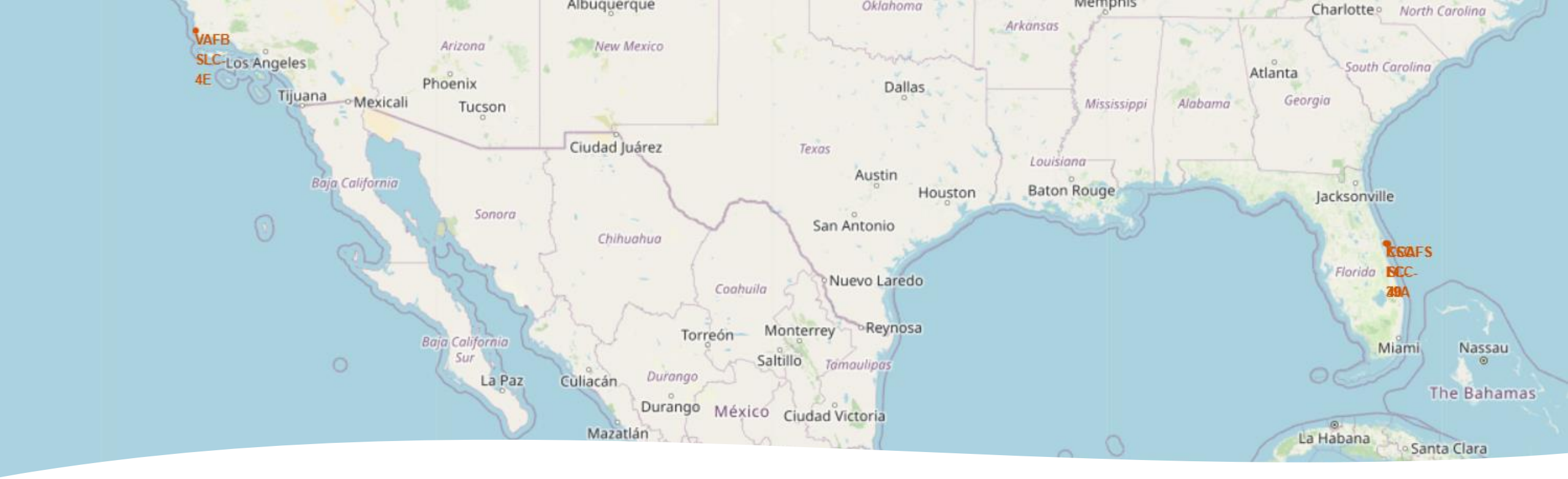
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- 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

- Present your query result with a short explanation here

Section 3

Launch Sites Proximities Analysis



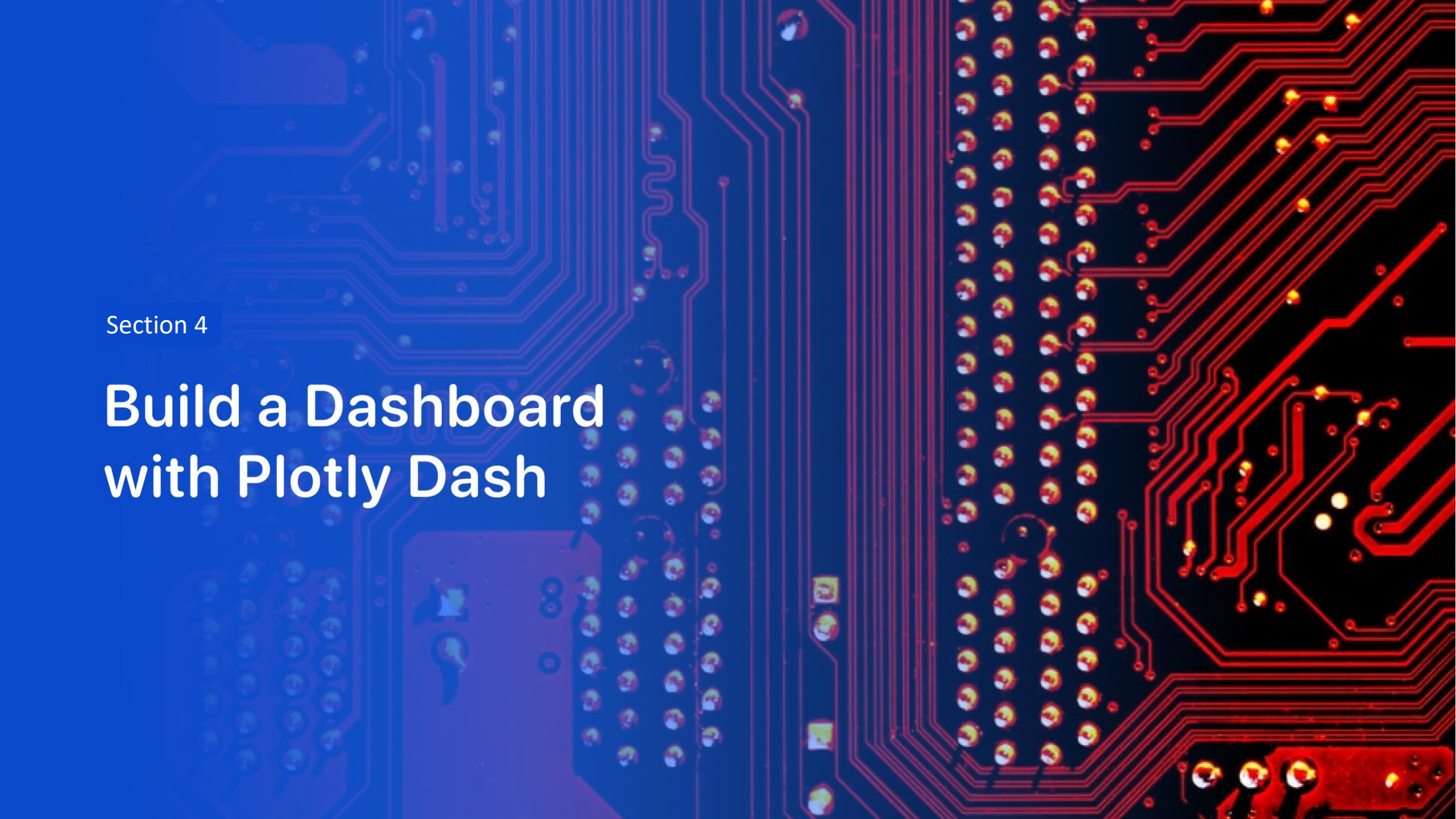
Map Indication Launch Site Locations

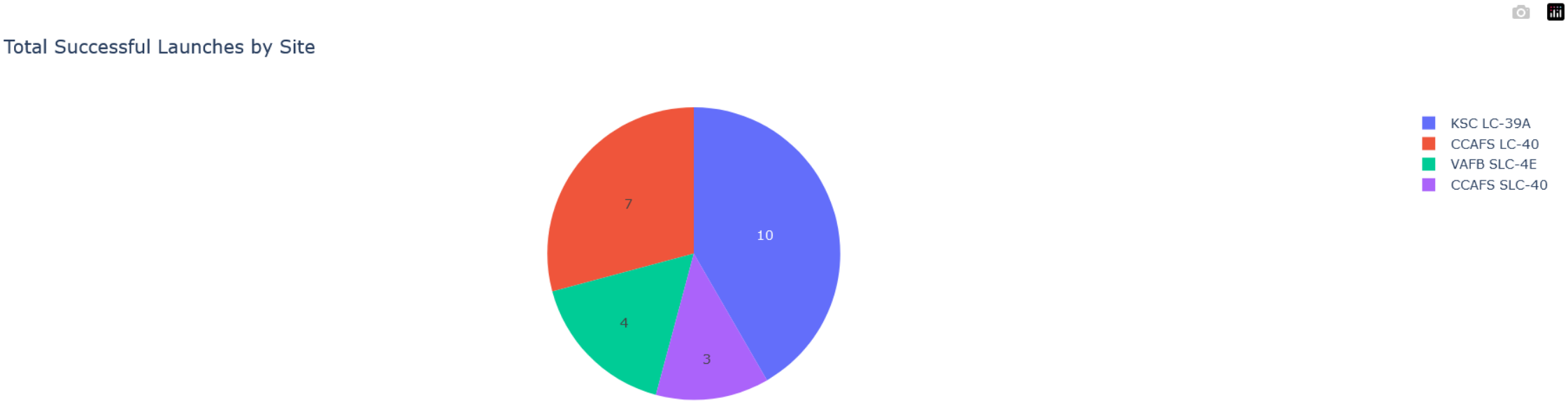
Launch Site With Markers

- The markers with red text indicates a failed attempt and green indicates a sucess

<Folium Map Screenshot 3>

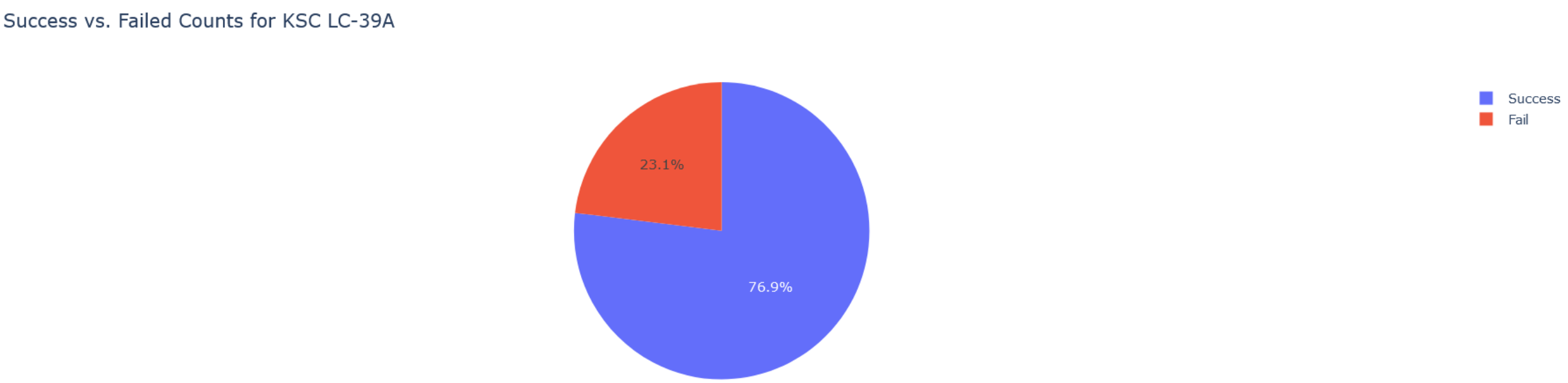
- Replace <Folium map screenshot 3> title with an appropriate title
- Explore the generated folium map and show the screenshot of a selected launch site to its proximities such as railway, highway, coastline, with distance calculated and displayed
- Explain the important elements and findings on the screenshot


Section 4
Build a Dashboard with Plotly Dash



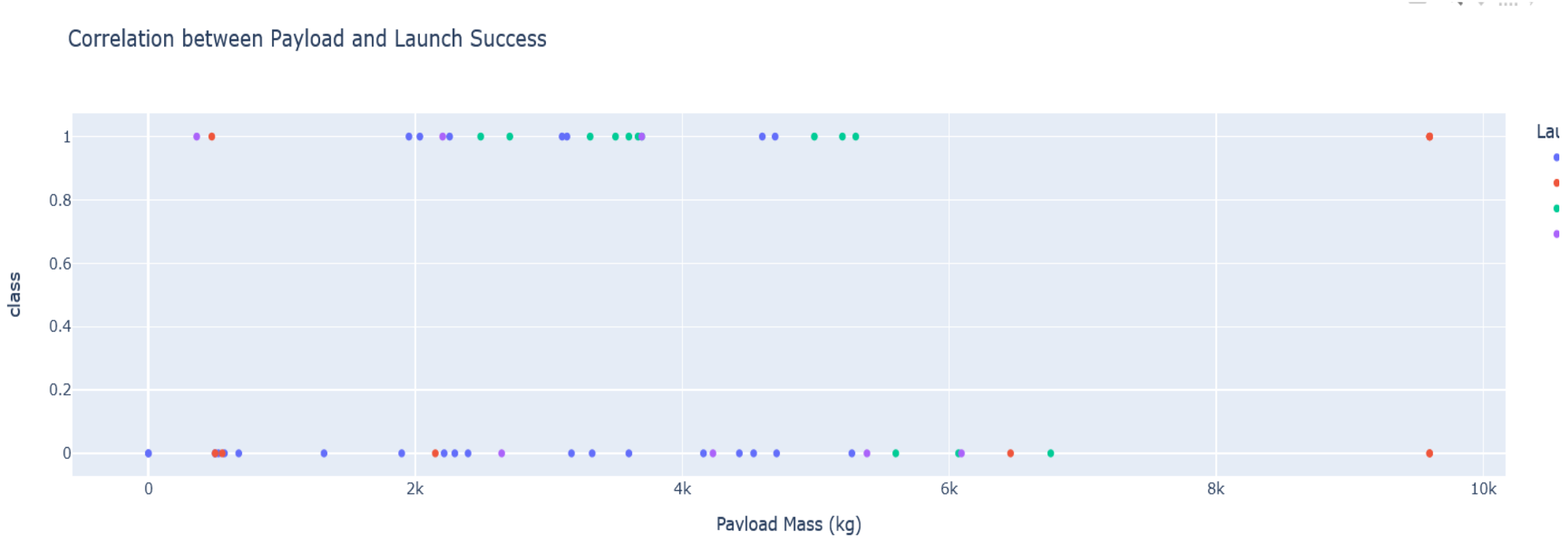
Successful Launch Count of Each Launch Site

- Launch site “KSC LC-39A” has the most successful launches.



Success VS Failure For KSC LC-39A

- Launches at this site has a 76.9% success rate.



Scatter Plot of All Launch VS Outcomes

- Explain the important elements and findigs on the screenshot, such as which payload range or booster version have the largest success rate, etc.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

- Visualize the built model accuracy for all built classification models, in a bar chart

- Find which model has the highest classification accuracy

Confusion Matrix

- Show the confusion matrix of the best performing model with an explanation

Conclusions

- Point 1
- Point 2
- Point 3
- Point 4
- …

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!