IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Ulises Ramsés Prado Valderrábano
November 3, 2023

Outline

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

Executive Summary

Summary of methodologies

1. Data Collection using API and Web Scraping
2. Data Wrangling
3. EDA with SQL
4. EDA with Visualization
5. Interactive Visual Analytics and Dashboard
6. Predictive Analysis

Summary of all results

- EDA showed important features to develop predictive analysis
- Machine Learning methods were evaluated to find the best to predict successful Space X first stage landings

Introduction

- Project Background

Space X can reuse the first stage of their Falcon 9 rocket launches with a total cost of 62 million dollars, other providers cost upward of 165 millions dollars.

- Desirable answers

Determine if the first stage of the Falcon 9 rocket launch will land successfully to estimate the cost of a launch.

Section 1

Methodology

Methodology

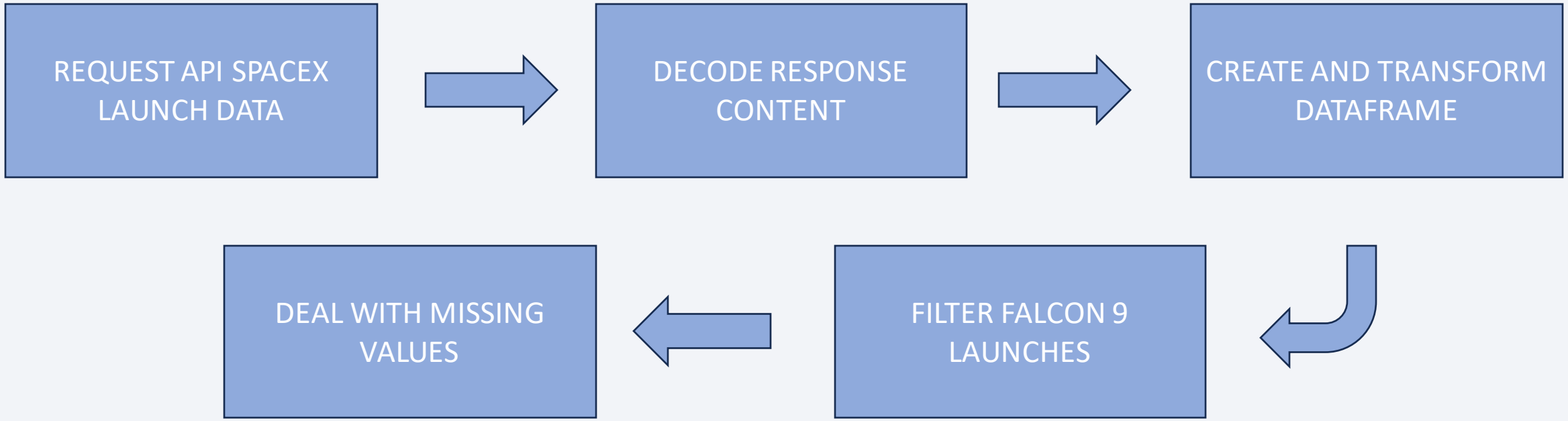
Executive Summary

- Data collection methodology
  - Data were collected from the Space X API (api.spacexdata.com/v4/) and performing Web Scraping from Wikipedia (https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches)
- Perform data wrangling
  - After analyzing features, a landing outcome label was created from the outcome column
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Data was standardized, trained and evaluated with four classification methods.

Data Collection

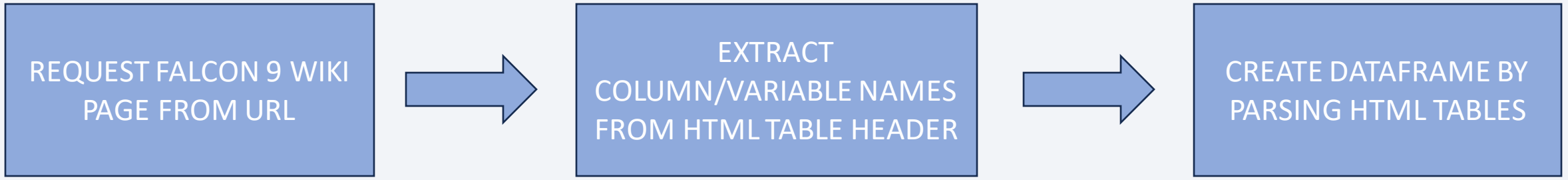
- Data related to rocket specifications and launches information were collected using the Space X API
- Historical Falcon 9 launch records were collected performing Web Scraping

Data Collection – SpaceX API



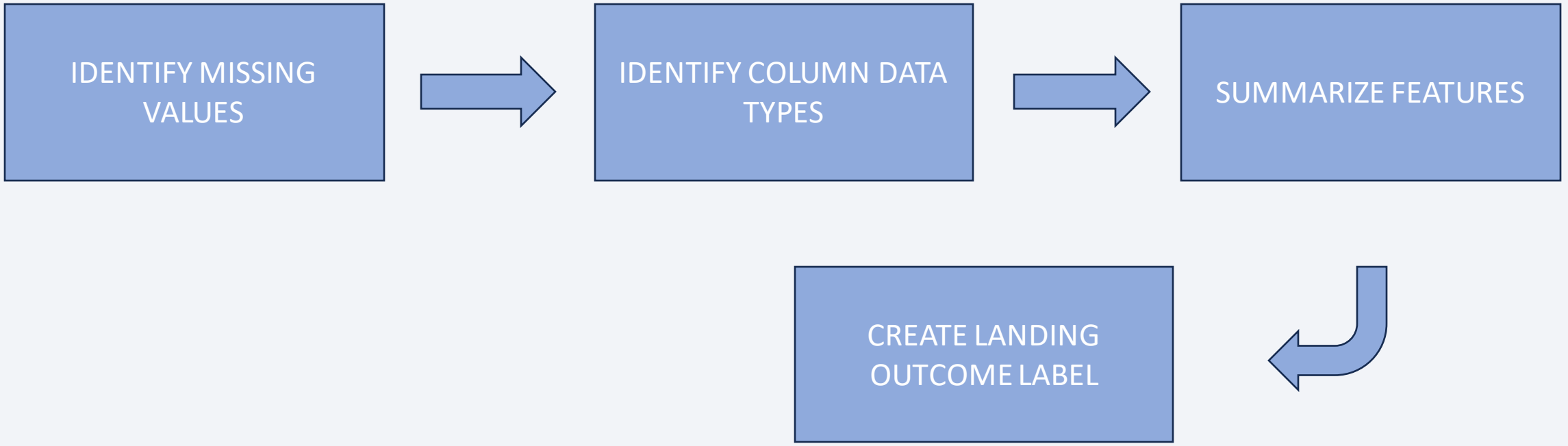
Github URL: https://github.com/UlisesPrado09/Applied-Data-Science-Capstone/blob/main/Data-Collection-API.ipynb

Data Collection - Scraping



GitHub URL: https://github.com/UlisesPrado09/Applied-Data-Science-Capstone/blob/main/Data-Collection-WebScraping.ipynb

Data Wrangling



GitHub URL: https://github.com/UlisesPrado09/Applied-Data-Science-Capstone/blob/main/Data_Wrangling.ipynb

EDA with Data Visualization

The following chart types were used:

- Scatter Plot: Flight Number vs Payload Mass, Flight Number vs Launch Site, Payload Mass vs Launch Site, Flight Number vs Orbit Type, Payload Mass vs Orbit Type.
- Bar Plot: Orbit Success Rate.
- Line Plot: Success Rate Over Time.

Scatter Plots were useful to identify relationships between variables, Bar Charts helped to find the orbit with the highest success rate, and with a Line Chart it was easy to identify an increase in success rate over time.

GitHub URL: https://github.com/UlisesPrado09/Applied-Data-Science-Capstone/blob/main/EDA-with-DataVisualization.ipynb

EDA with SQL

Performed SQL queries:

- Display the names of the unique launch sites in the space mission
- Display 5 records where launch sites begin with the string 'CCA'
- Display the total payload mass carried by boosters launched by NASA (CRS)
- Display average payload mass carried by booster version F9 v1.1
- List the date when the first successful landing outcome in ground pad was achieved.
- 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.
- 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 between the date 2010-06-04 and 2017-03-20, in descending order.

GitHub URL: https://github.com/UlisesPrado09/Applied-Data-Science-Capstone/blob/main/EDA-with-SQL.ipynb

Build an Interactive Map with Folium

- Markers for Launch Sites

Circle Markers with Popup Label and Text Label were added for all launch sites using their latitude and longitude.

- Colored Markers for launch outcomes

Colored Markers were plotted for each launch outcome, green for successful launches and red for failed launches.

- Lines for distances between points of interest

Lines were added showing distance between launch sites and proximities like railways, highways, coastlines and cities.

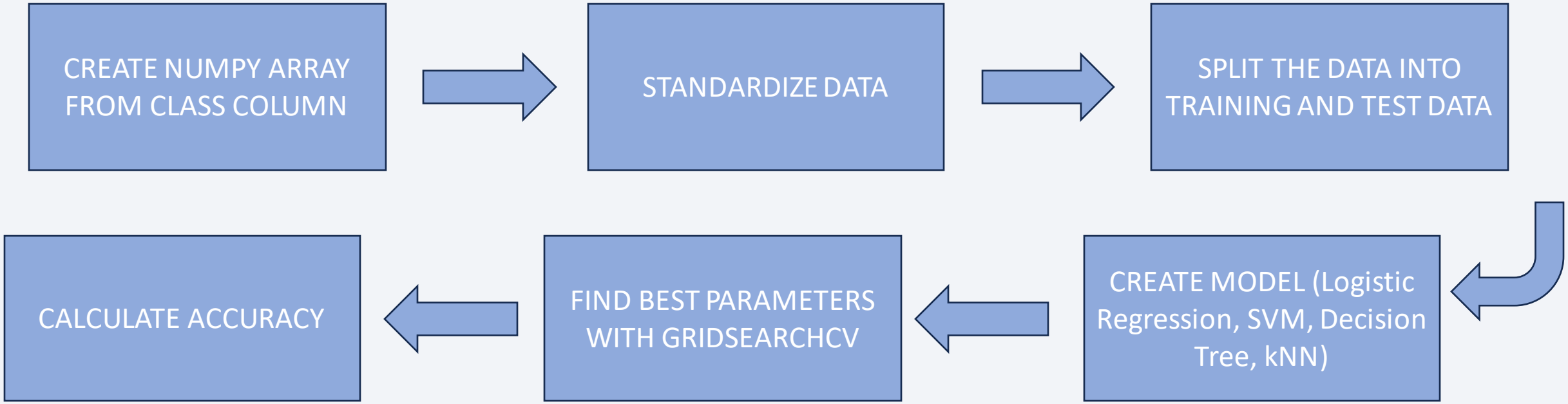
GitHub URL: https://github.com/UlisesPrado09/Applied-Data-Science-Capstone/blob/main/Visualization-with-Folium.ipynb

Build a Dashboard with Plotly Dash

- Added a Launch Site drop-down menu for launch site selection.
- Added a pie chart to visualize total launches for all sites or successful vs failed launches for a specific launch site.
- A scatter plot is displayed to visualize the relationship between payload mass and successful launches.
- A slider is included to select payload mass range.

GitHub URL: https://github.com/UlisesPrado09/Applied-Data-Science-Capstone/blob/main/SpaceX-Dashboard.py

Predictive Analysis (Classification)



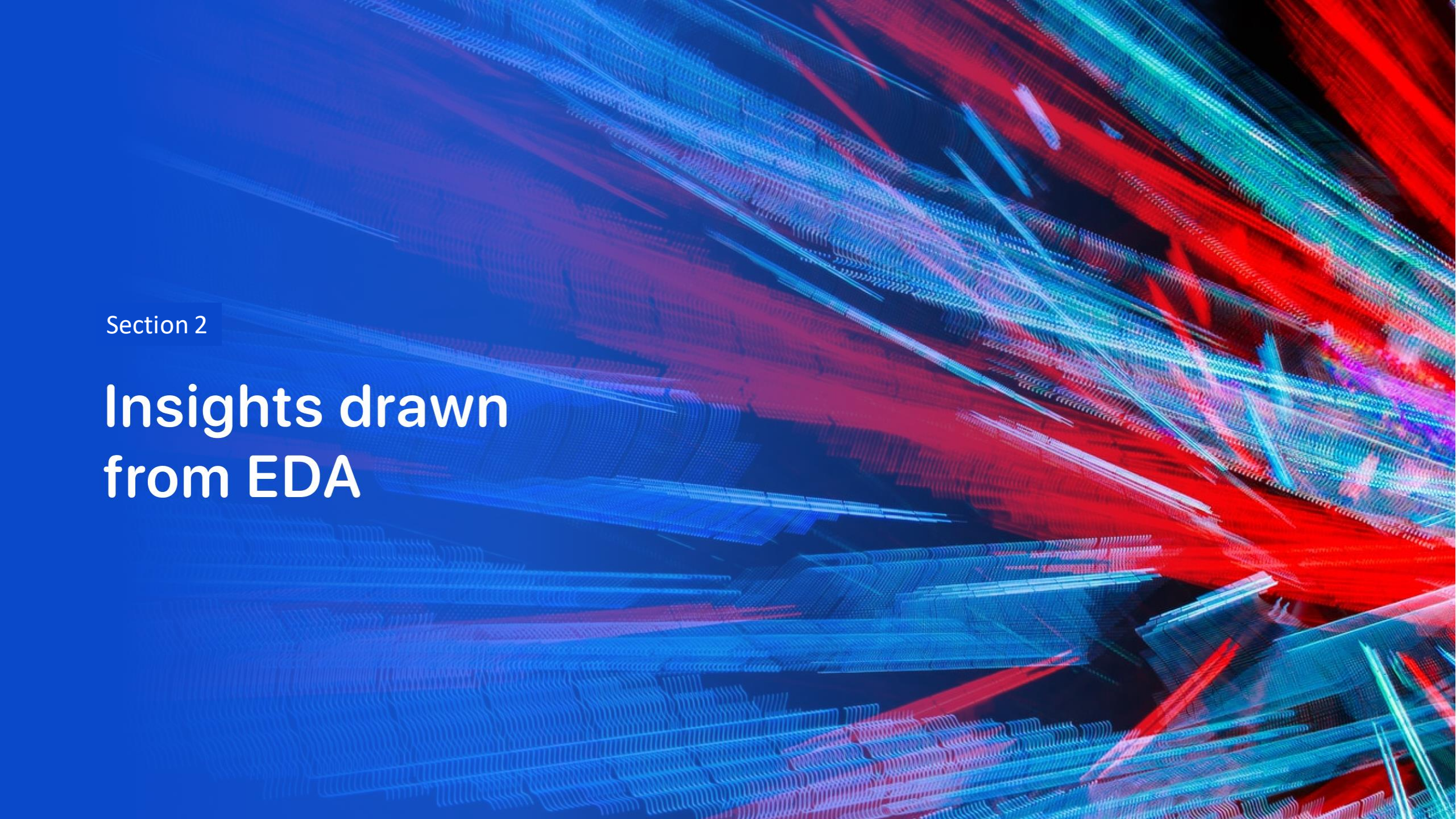
GitHub URL: https://github.com/UlisesPrado09/Applied-Data-Science-Capstone/blob/main/MachineLearning-Prediction.ipynb

Results

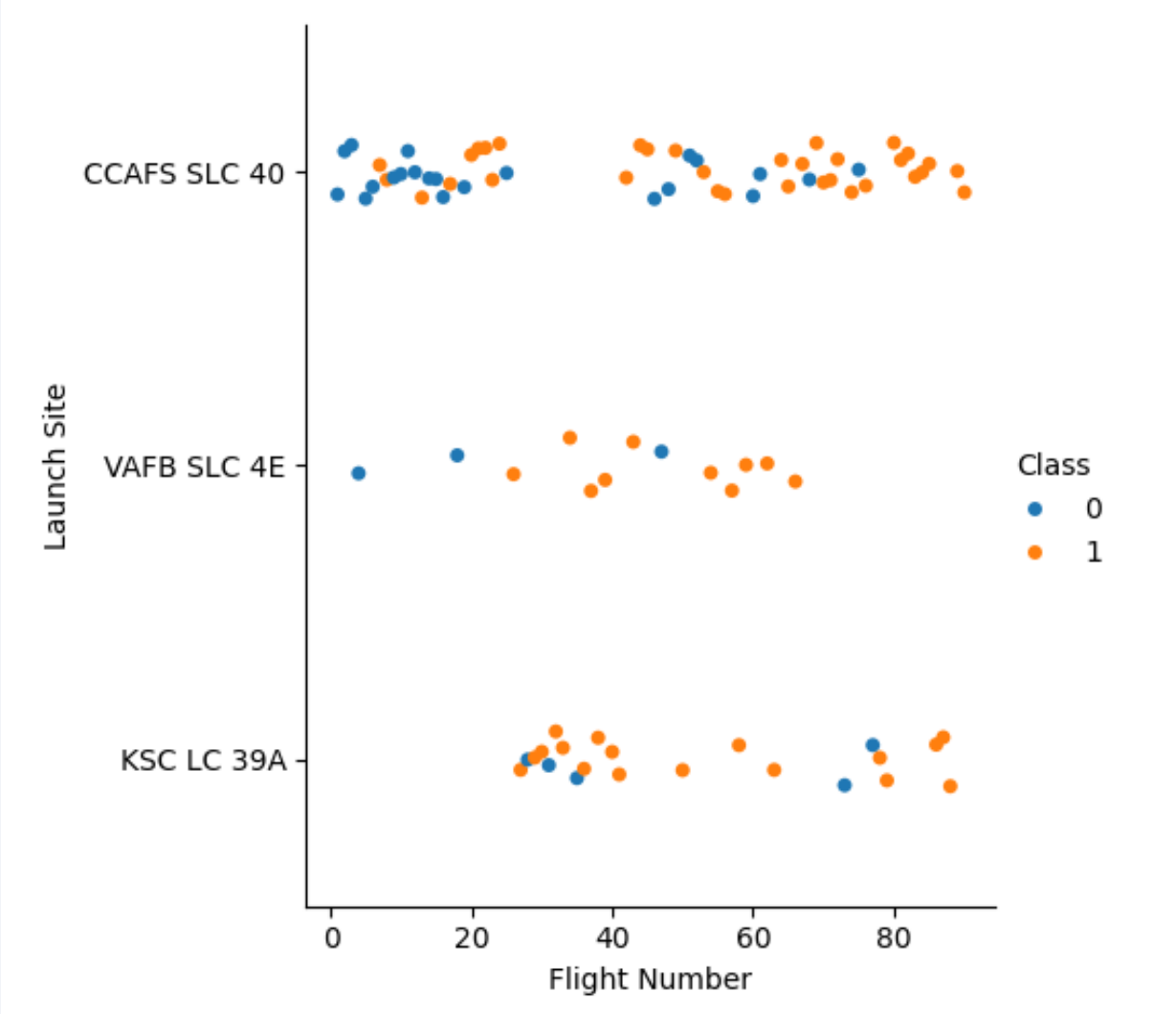
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA



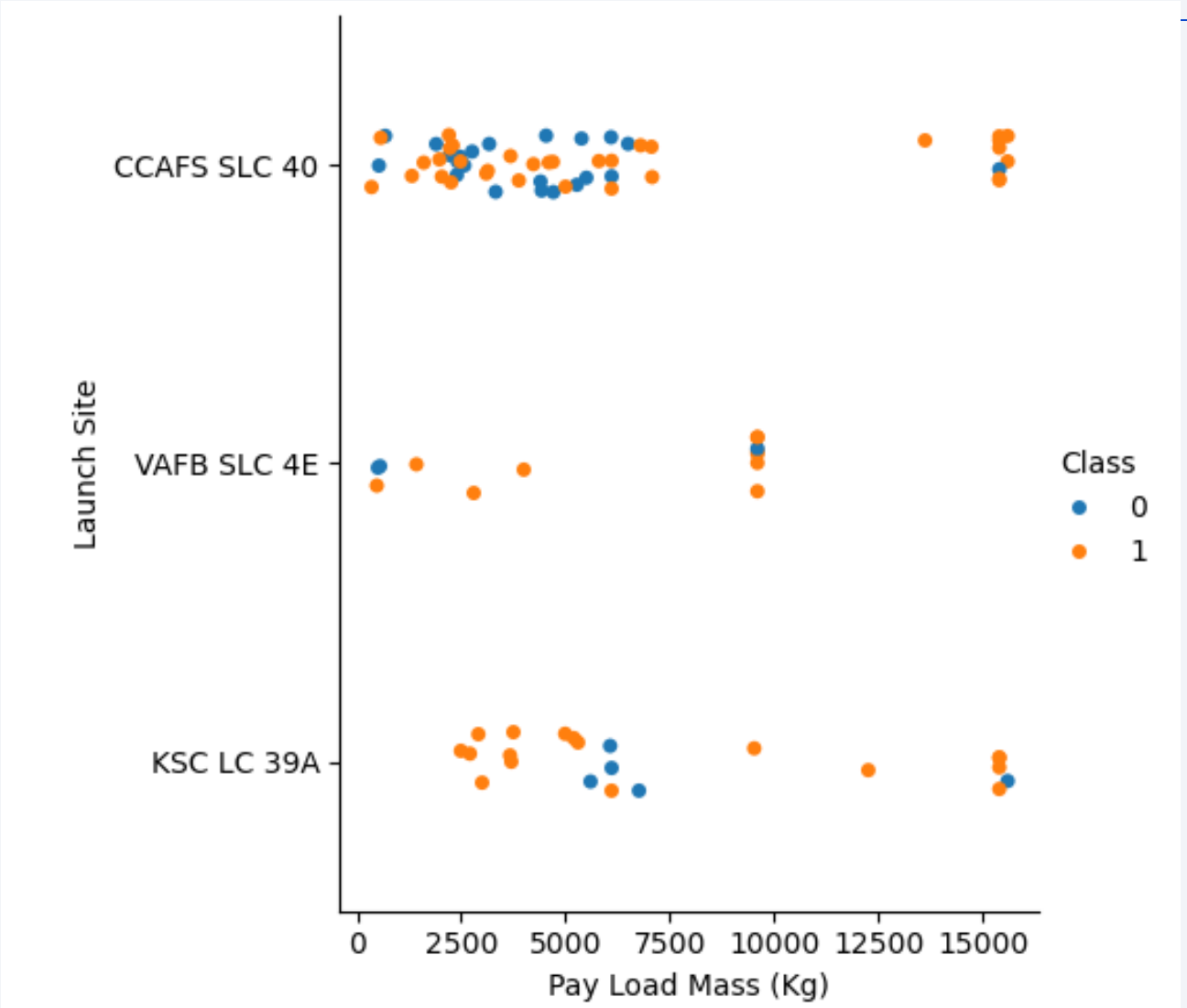
Flight Number vs. Launch Site



In launch site CCAFS SLC 40 occurred more unsuccessful landings for flight numbers between 0 and 20.

There is no visible strong correlation between flight number and other launch sites.

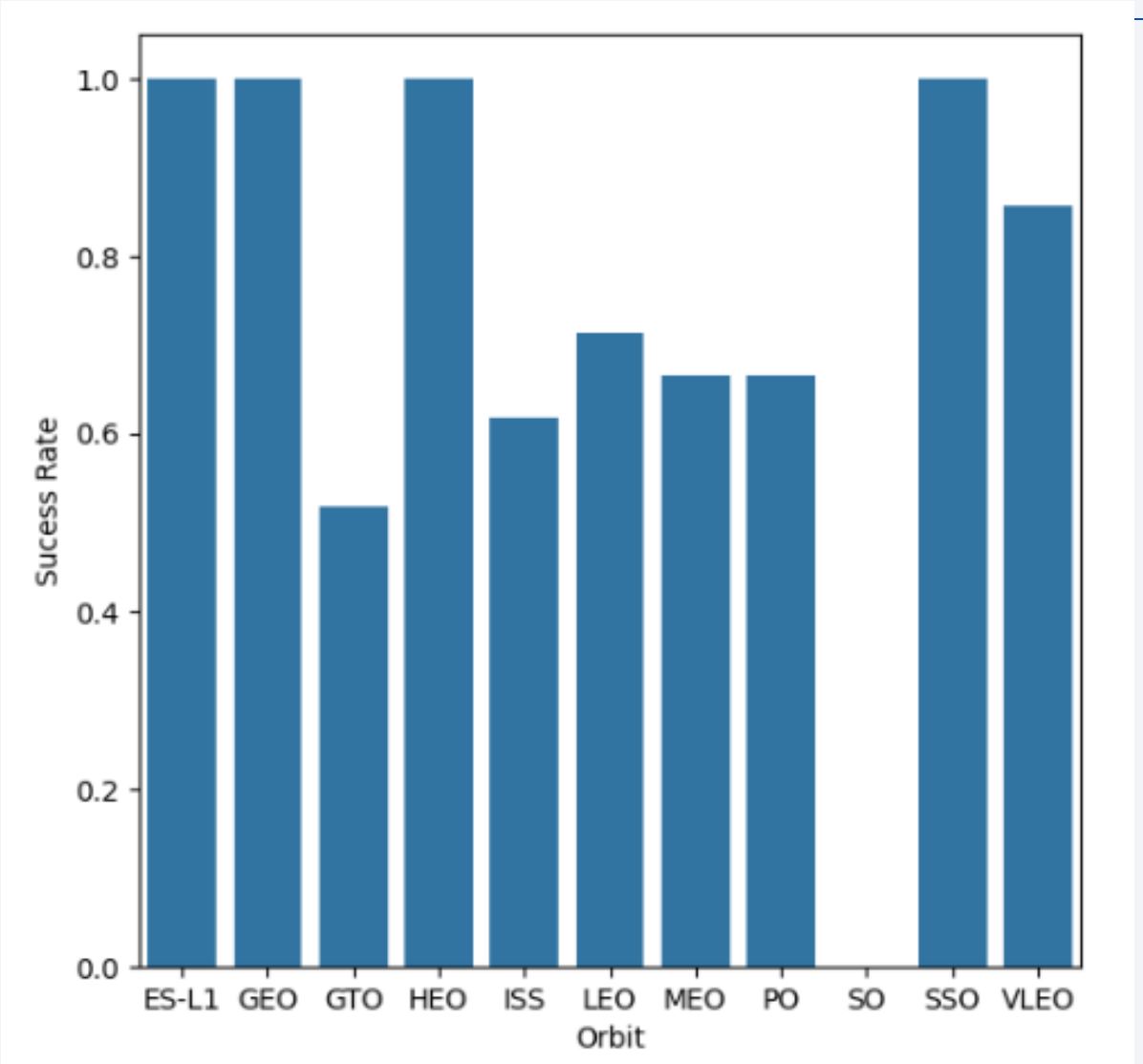
Payload vs. Launch Site



CCAFS SLC 40 launch site showed more failed landings for payload mass between 0 and 7500 kg.

For VAFB SLC 43 launch site there are no rocket launches with payload mass greater than 10000 kg.

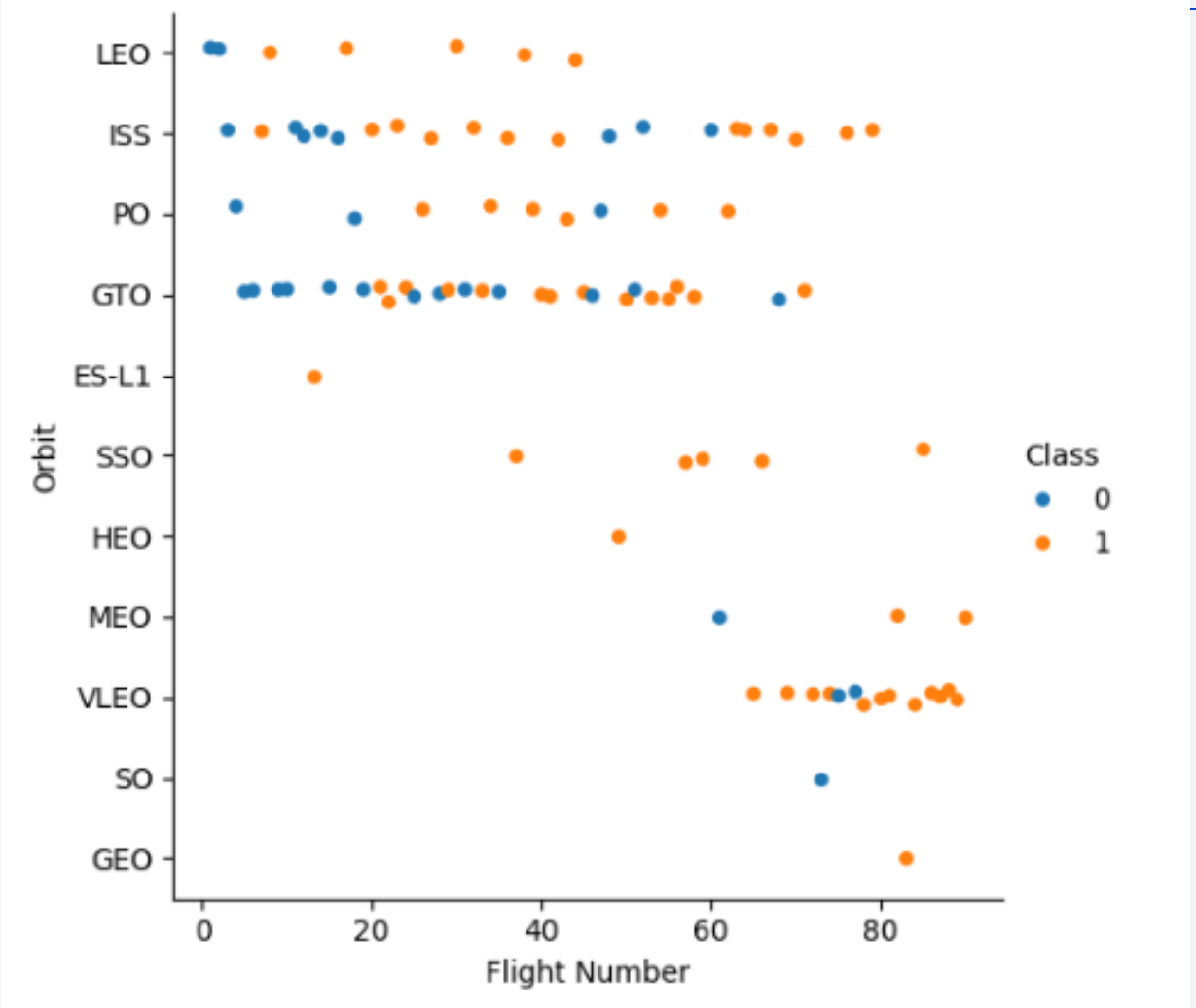
Success Rate vs. Orbit Type



ES-L1, GEO, HEO and SSO orbits showed the highest success rate.

For orbit SO the success rate is 0.

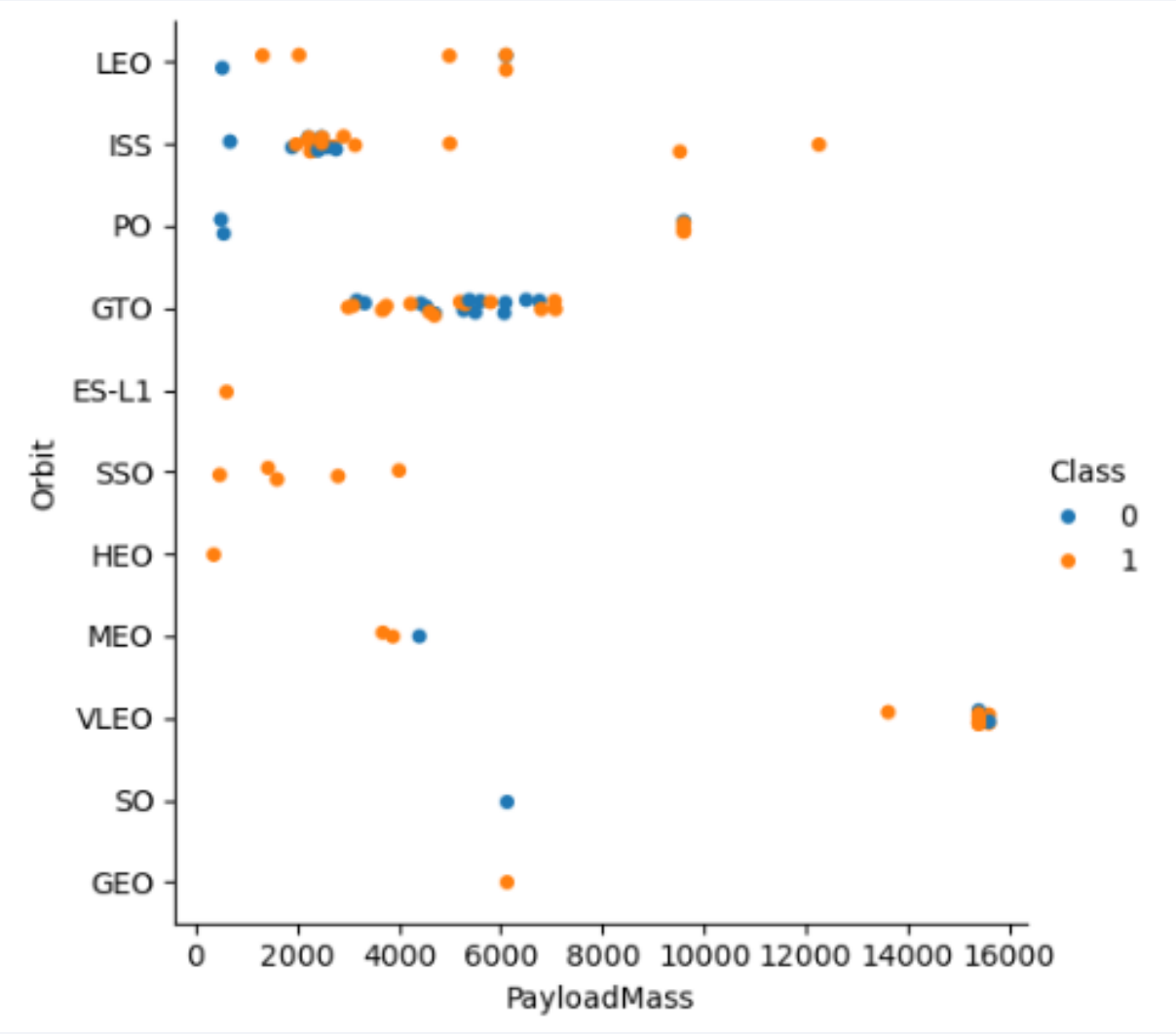
Flight Number vs. Orbit Type



In LEO orbit there is only successful launches for flight numbers greater than 10.

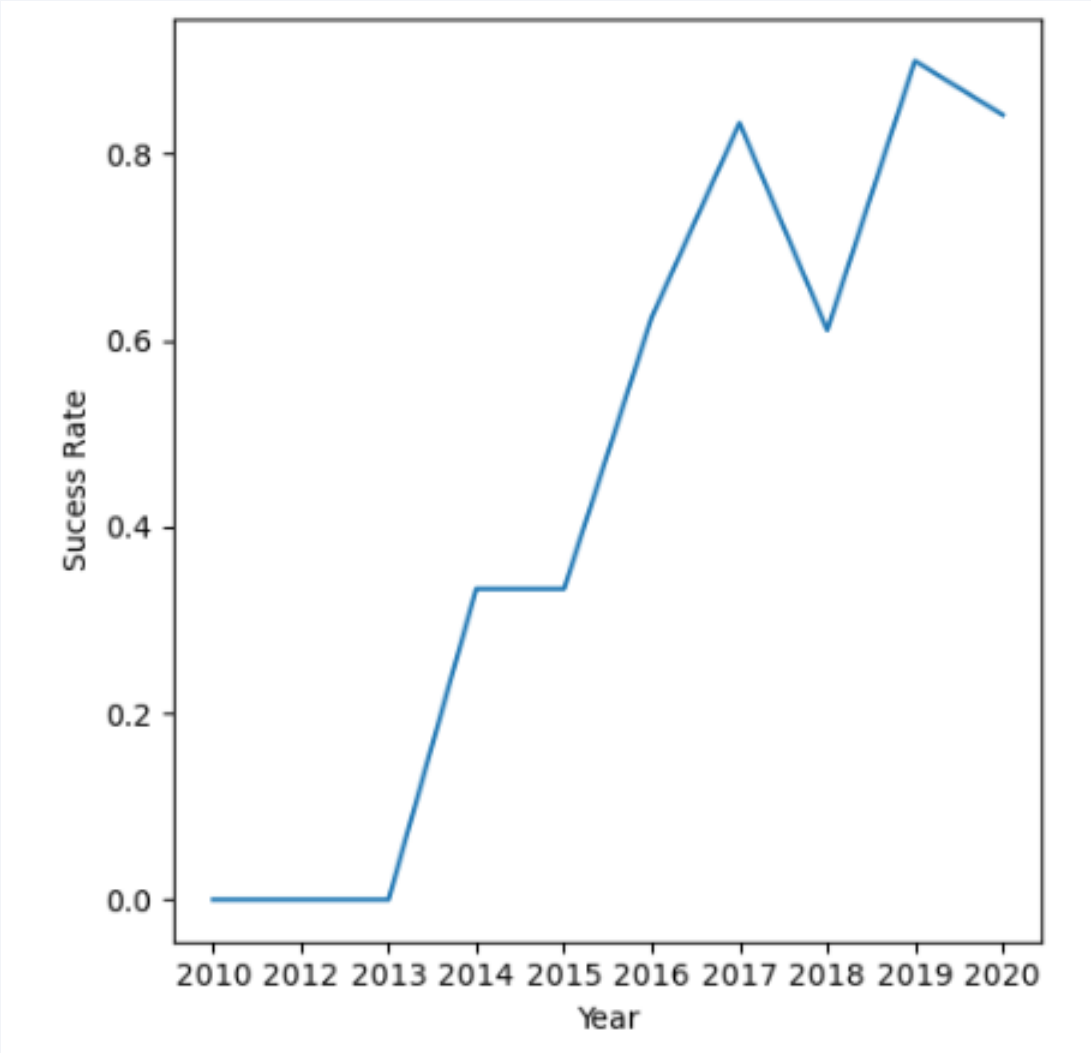
There is no relationship between GTO orbit and successful flight numbers.

Payload vs. Orbit Type



With heavy payloads the successful landing rate are more for Polar, LEO and ISS.

Launch Success Yearly Trend



This line chart showed that success rate since 2013 kept increasing till 2020.

All Launch Site Names

Display the names of the unique launch sites in the space mission

```
%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'

Display 5 records where launch sites begin with the string 'CCA'

```
%sql SELECT "Launch_Site" FROM SPACEXTABLE WHERE "Launch_Site" LIKE "CCA%" LIMIT 5
```

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

Launch_Site
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40

Total Payload Mass

Display the total payload mass carried by boosters launched by NASA (CRS)

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

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

SUM("PAYLOAD_MASS__KG_")
45596

Average Payload Mass by F9 v1.1

Display average payload mass carried by booster version F9 v1.1

```
%sql SELECT AVG("PAYLOAD_MASS__KG_") AS AVERAGE_MASS FROM SPACEXTABLE WHERE "Booster_Version" == "F9 v1.1"
```

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

AVERAGE_MASS
2928.4

First Successful Ground Landing Date

List the date when the first succesful landing outcome in ground pad was acheived.

Hint:Use min function

```
%sql SELECT MIN(DATE) FROM SPACEXTABLE WHERE "Landing_Outcome" == "Success (ground pad)"
```

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

MIN(DATE)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

List the names of the boosters which have success in drone ship and have 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_" > 4000 AND "PAYLOAD_MASS__KG_" < 6000)
```

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

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

Total Number of Successful and Failure Mission Outcomes

List the total number of successful and failure mission outcomes

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

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

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

Boosters Carried Maximum Payload

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

```
%%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 Records

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.

Note: SQLLite does not support monthnames. So you need to use substr(Date, 6,2) as month to get the months and substr(Date,0,5)='2015' for year.

```
%%sql SELECT substr(Date,6,2) AS Month, "Landing_Outcome", "Booster_Version", "Launch_Site" FROM SPACEXTABLE
WHERE "Landing_Outcome" = "Failure (drone ship)" AND substr(Date,0,5) = "2015"
```

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

Month	Landing_Outcome	Booster_Version	Launch_Site
10	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	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(*) AS TOTAL FROM SPACEXTABLE
WHERE "Date" BETWEEN "2010-06-04" AND "2017-03-20" GROUP BY "Landing_Outcome"
ORDER BY TOTAL DESC
```

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

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

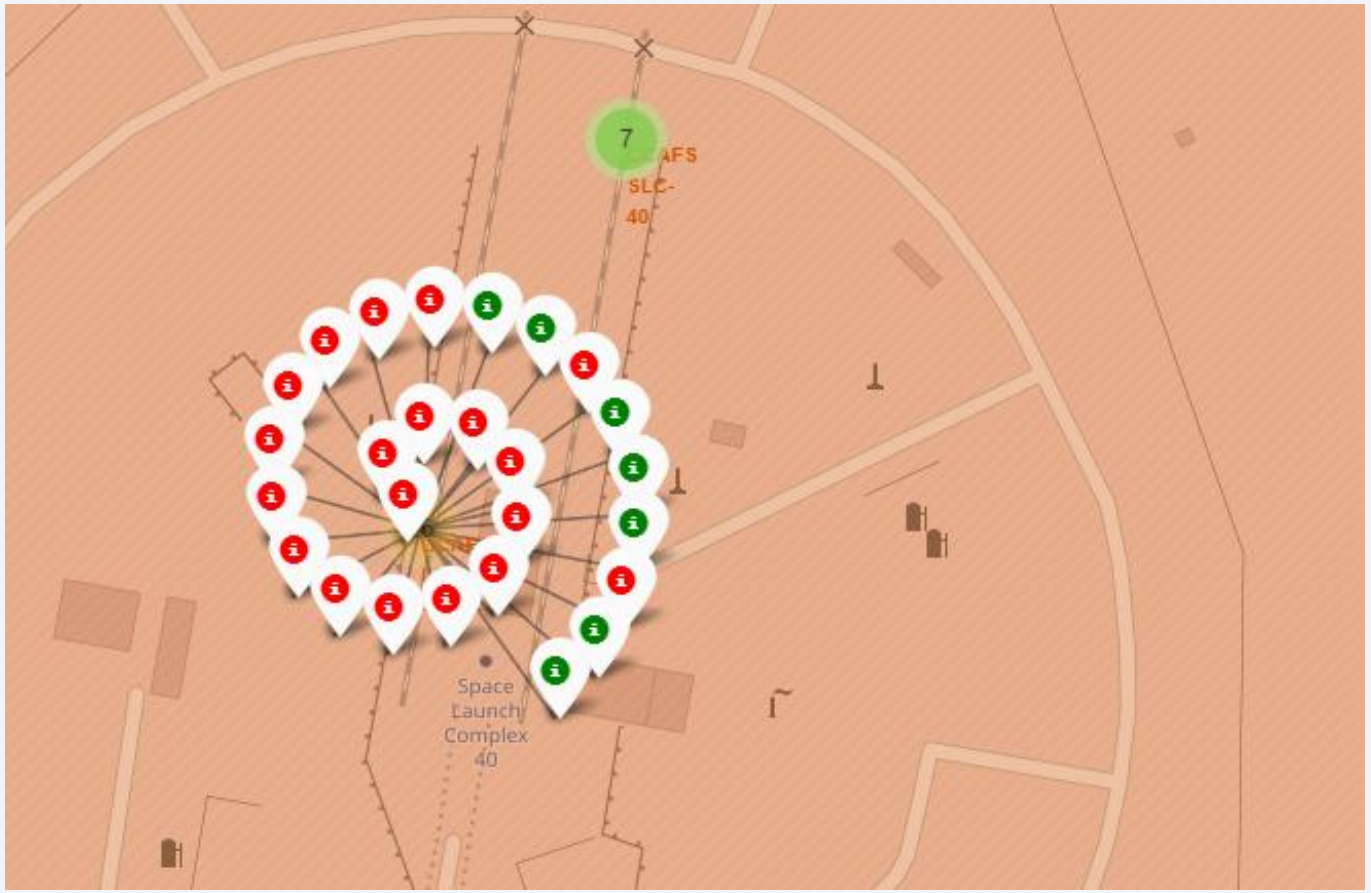

Section 3
Launch Sites
Proximities Analysis

All Launch Sites Map

- All launch sites were plotted with markers in a folium map.

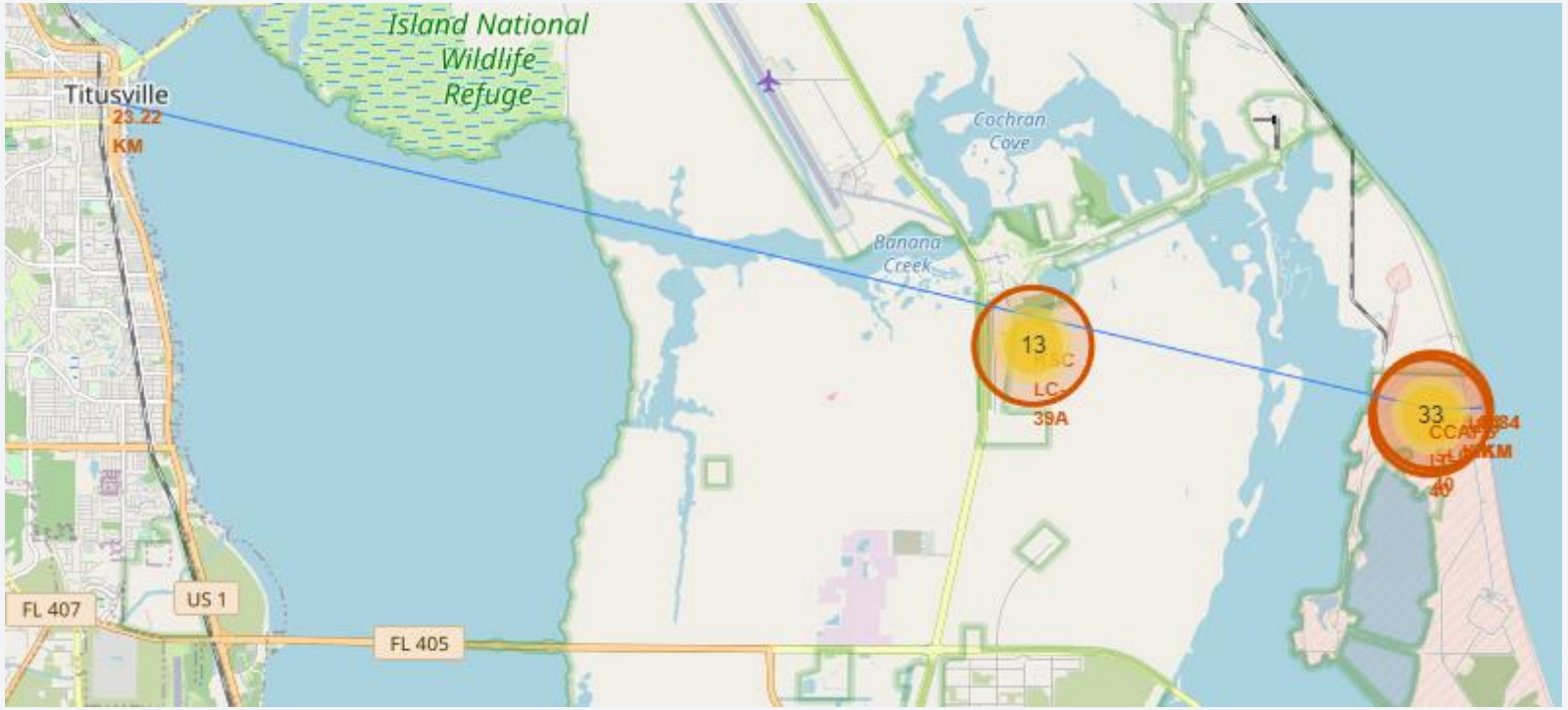
Color-labeled launch outcomes

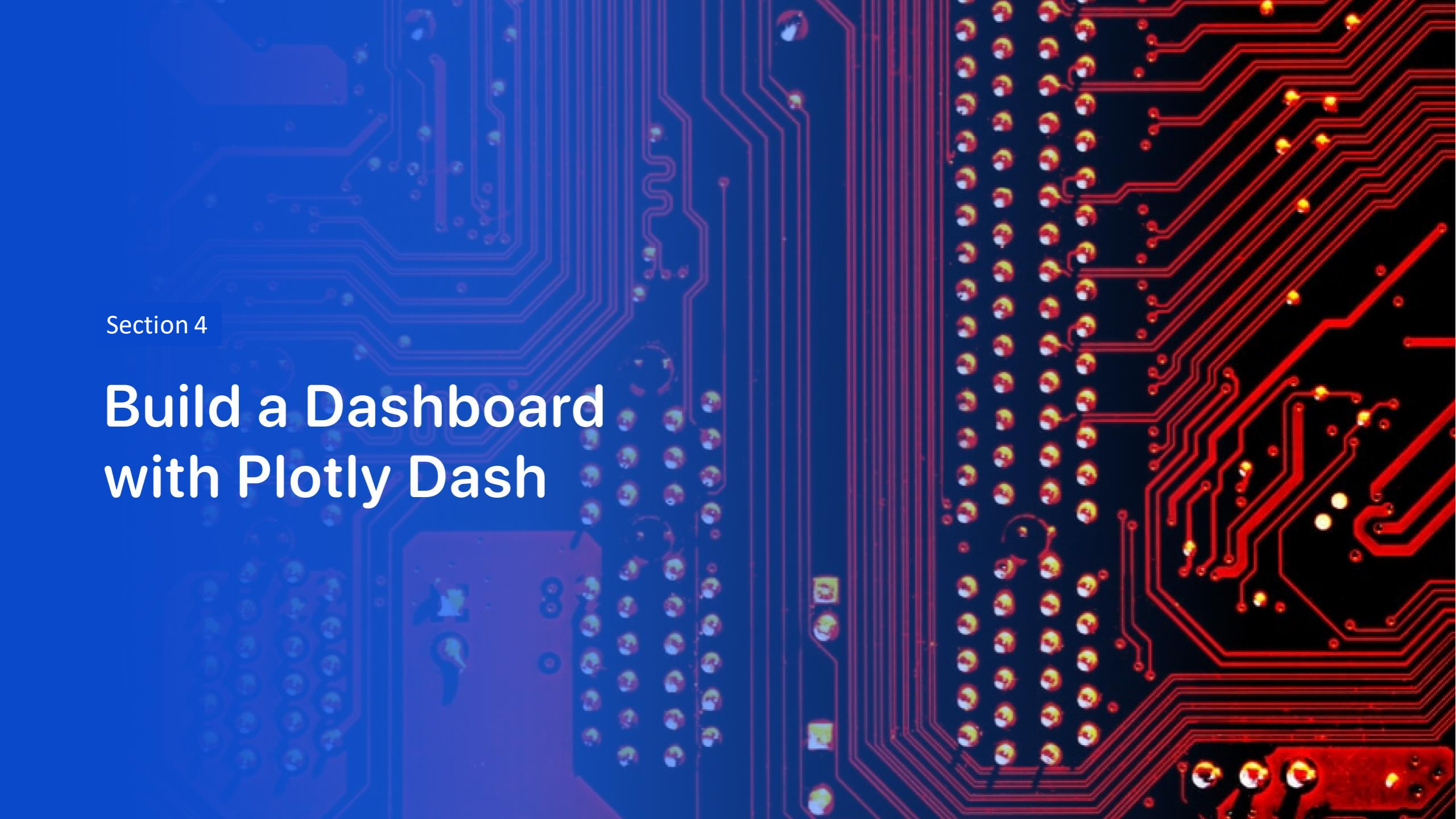
- Markers for all launch records were plotted and labeled with green color for successful launches and red for failed launches



Launch Sites and its proximities

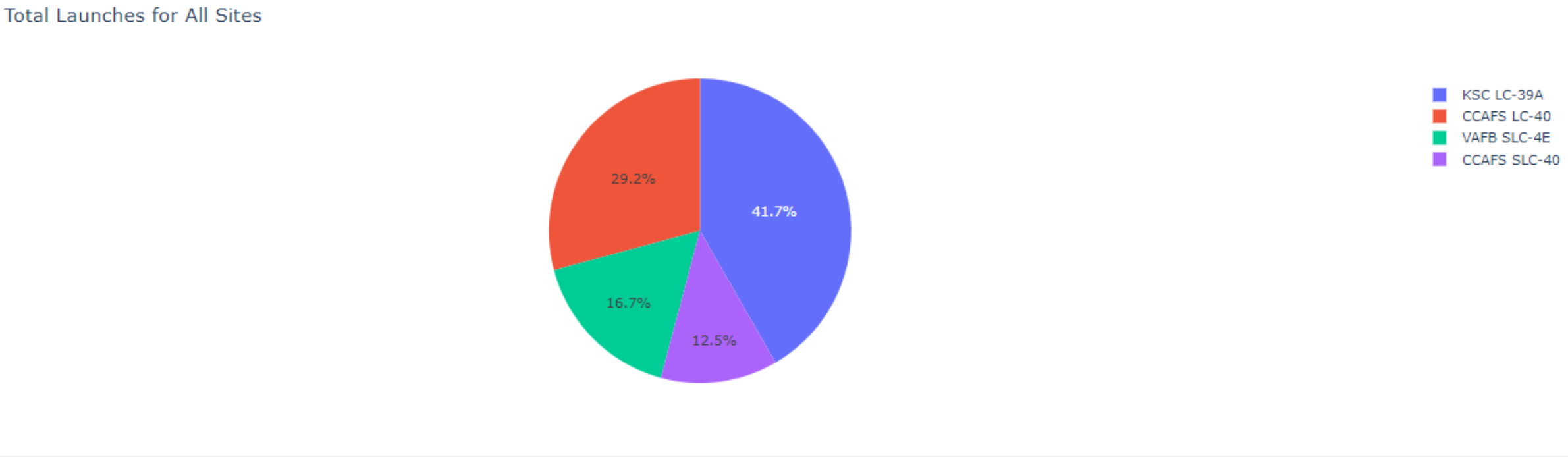
- A line can be displayed showing the distance between a launch site and its proximities like railways, highways, coastlines and cities.




Section 4
Build a Dashboard with Plotly Dash

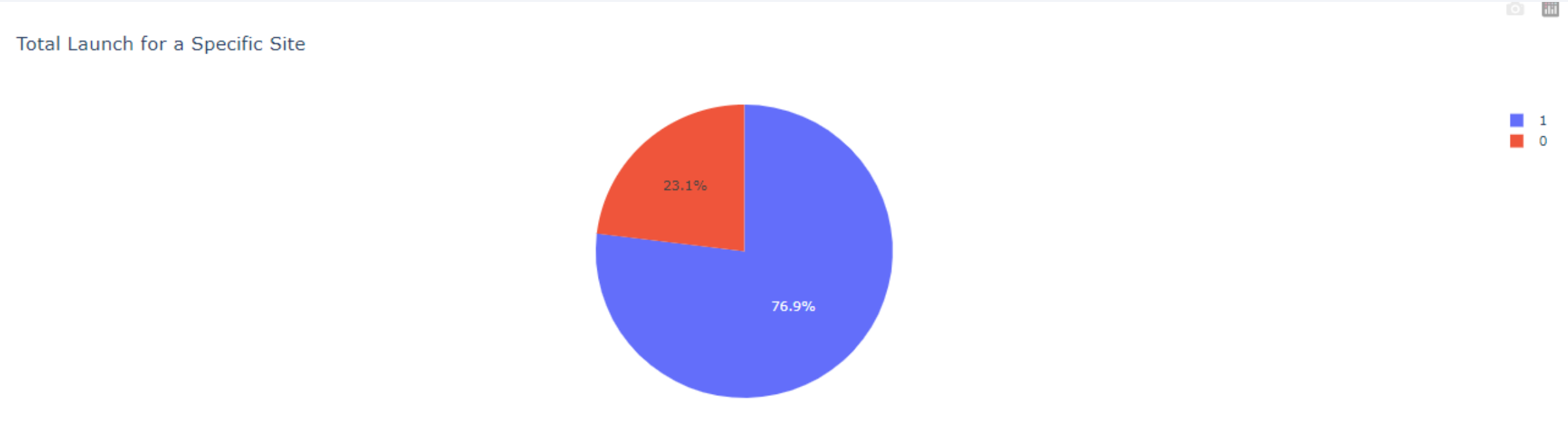
Pie Chart for Successful Launches

- The launch site with more successful launches was KSC LC 39A



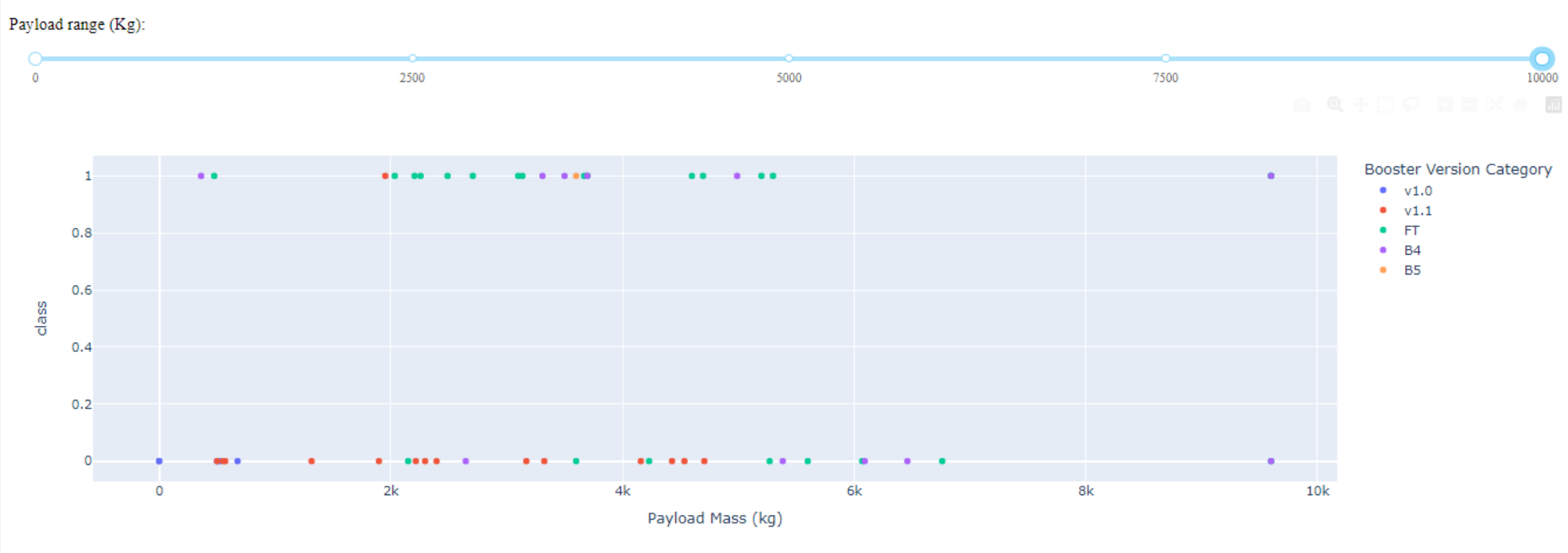
Pie Chart for KSC LC 39A launch site

- The KSC LC 39A launch site had 76.9% of successful launches.



Payload Mass vs Launch Outcome

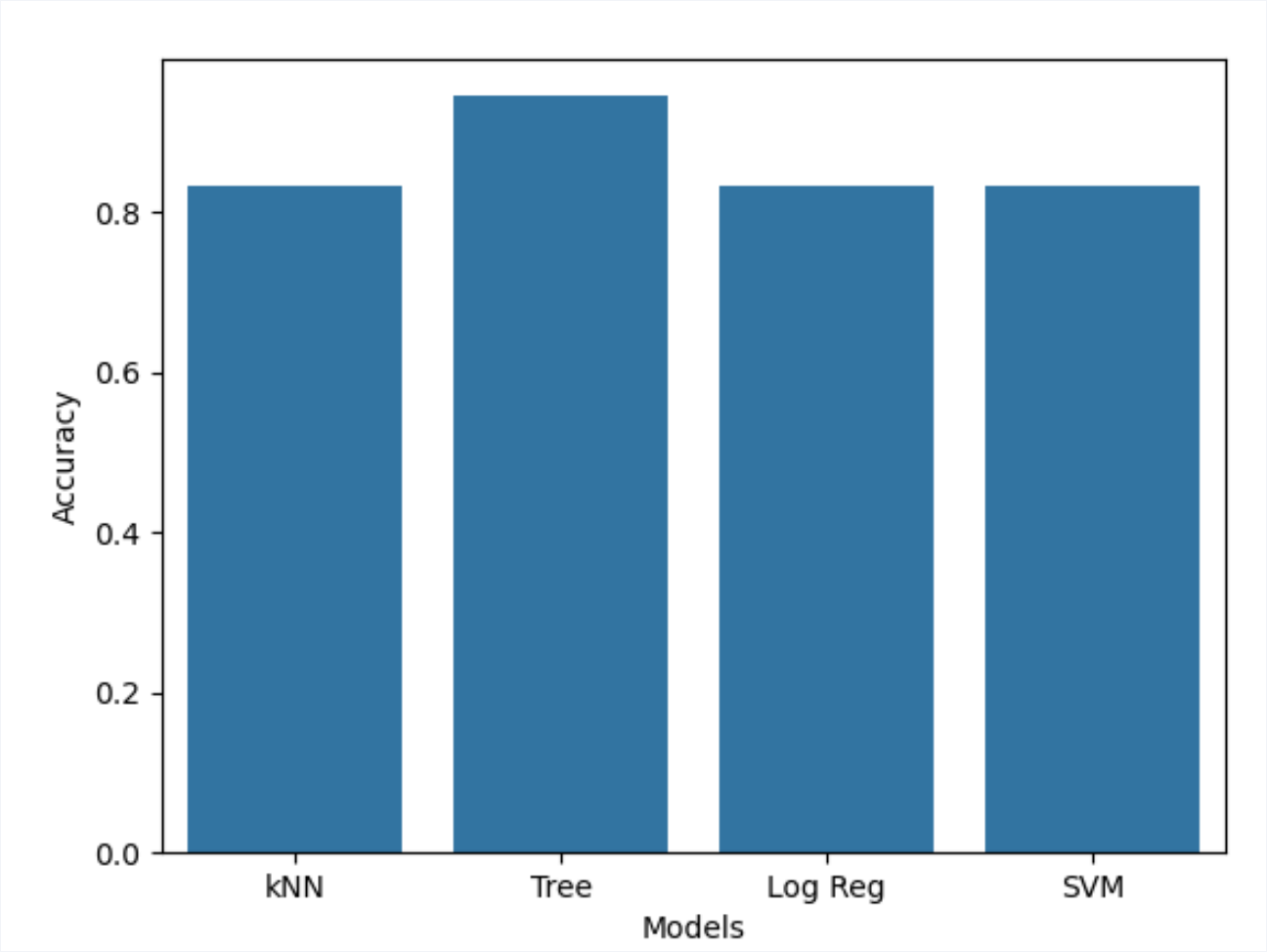
- Booster Version FT showed the highest number of successful launches, on the other hand, version v1.1 showed the highest number of failed launches.



Section 5

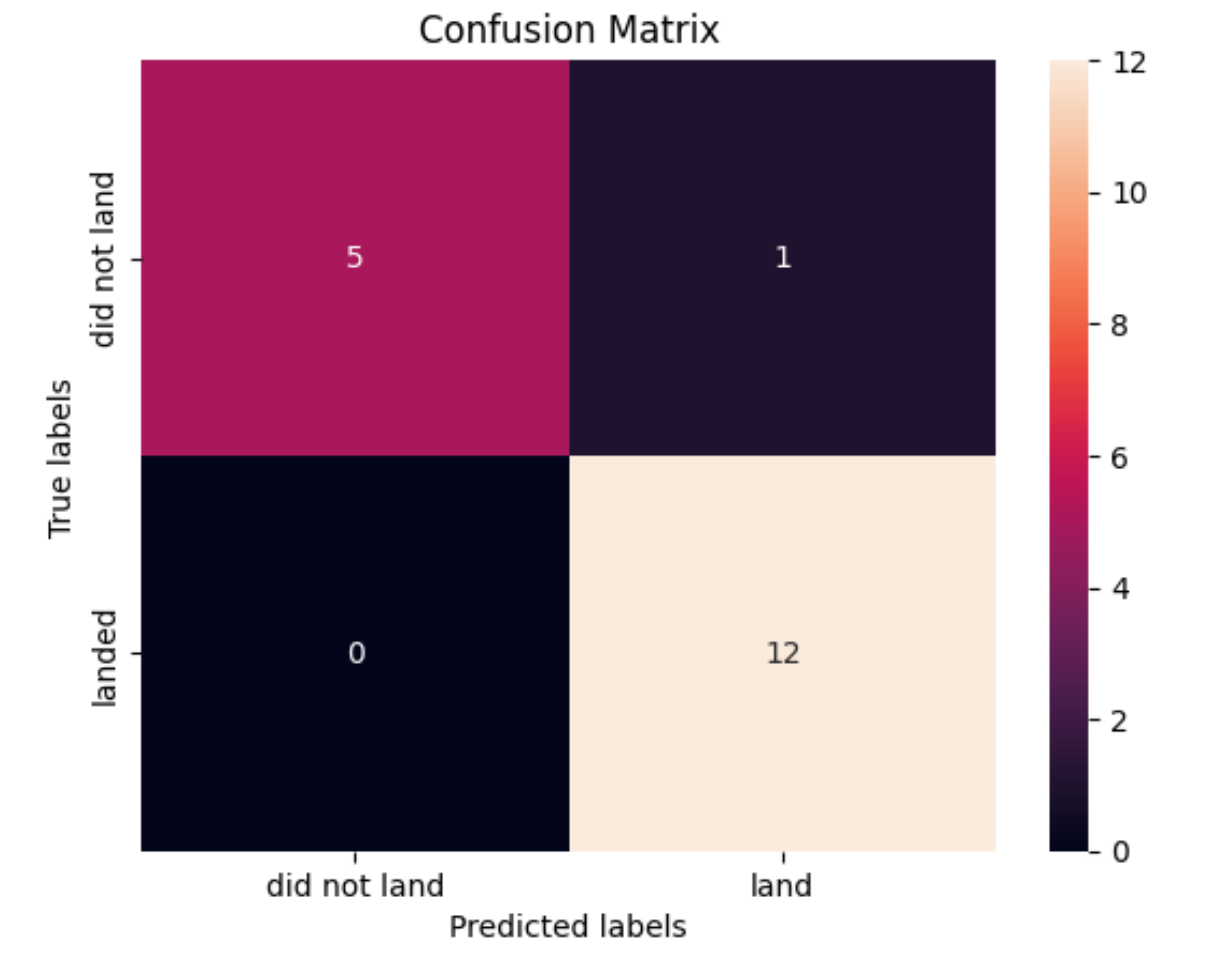
Predictive Analysis (Classification)

Classification Accuracy



The Decision Tree classification model obtained the best accuracy

Confusion Matrix



The Decision Tree classification model predicted correctly most of the launch cases, failing with just one failed launch.

Conclusions

- The Exploratory Data Analysis performed yielded significant information about SpaceX rocket launches
- There are some relationships between the extracted variables used for this study
- Four classification model were used to predict successful launches
- Decision Tree classification model obtained the best accuracy

Thank you!