IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Tarek Ali
October 16, 2022

Outline

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

Executive Summary

- **The following methodologies were used to analyze data:**
  - **Data Collection using web scraping and SpaceX API.**
  - **Exploratory Data Analysis (EDA), including data wrangling, data visualization and interactive**
  - **visual analytics.**
  - **Machine Learning Prediction.**
- **Summary of all results:**
  - **It was possible to collected valuable data from public sources.**
  - **EDA allowed to identify which features are the best to predict success of launchings.**
  - **Machine Learning Prediction showed the best model to predict which characteristics are important to drive this opportunity by the best way, using all collected data.**

Introduction

- Project background and context:
  - The objective is to evaluate the viability of the new company Space Y to compete with Space X.

- Problems you want to find answers:
  - The best way to estimate the total cost for launches, by predicting successful landings of the first stage of rockets.
  - Where is the best place to make launches.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Space X API (https://api.spacexdata.com/v4/rockets/)
  - Web Scraping (https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches
- Perform data wrangling
  - Collected data was enriched by creating a landing outcome label based on outcome data after summarizing and analyzing features
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models “evaluate classification models”

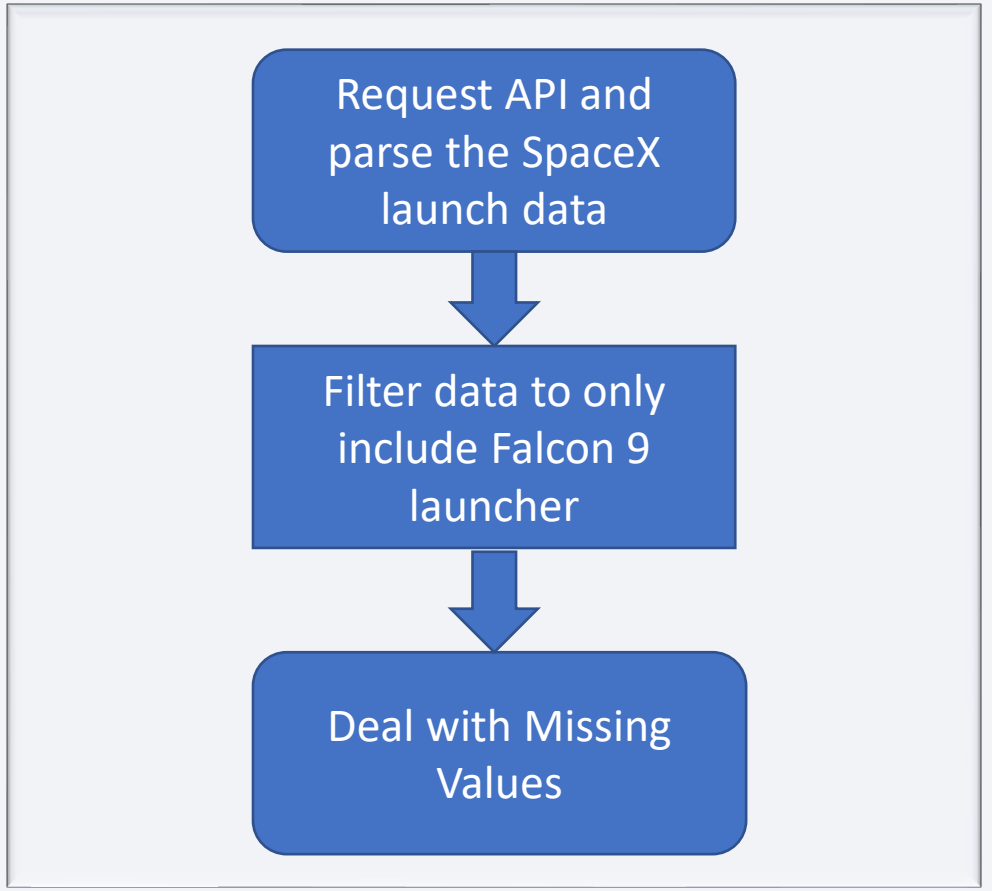
Data Collection

- Data sets were collected from:
  - Space X API (https://api.spacexdata.com/v4/rockets/)
  - Wikipedia (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches).

using web scraping technics.

Data Collection – SpaceX API

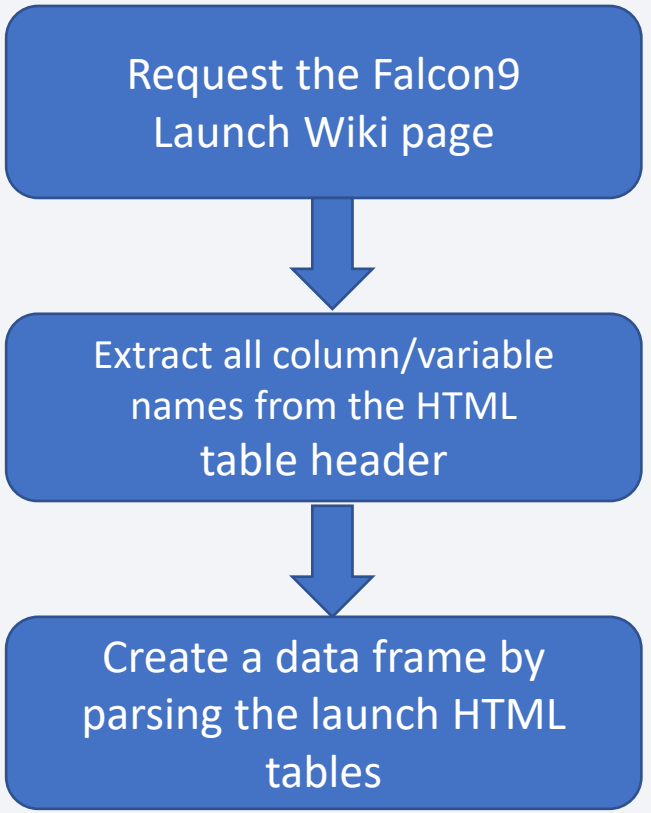
- SpaceX offers a public API from where data can be obtained and then used;
- This API was used according to the flowchart beside and then data is persisted
- Source code (https://github.com/tarek-ali20/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/jupyter-labs-spacex-data-collection-api.ipynb)



Data Collection - Scraping

- Data from SpaceX launches can also be obtained from Wikipedia

- Data are downloaded from Wikipedia according to the flowchart and then persisted.



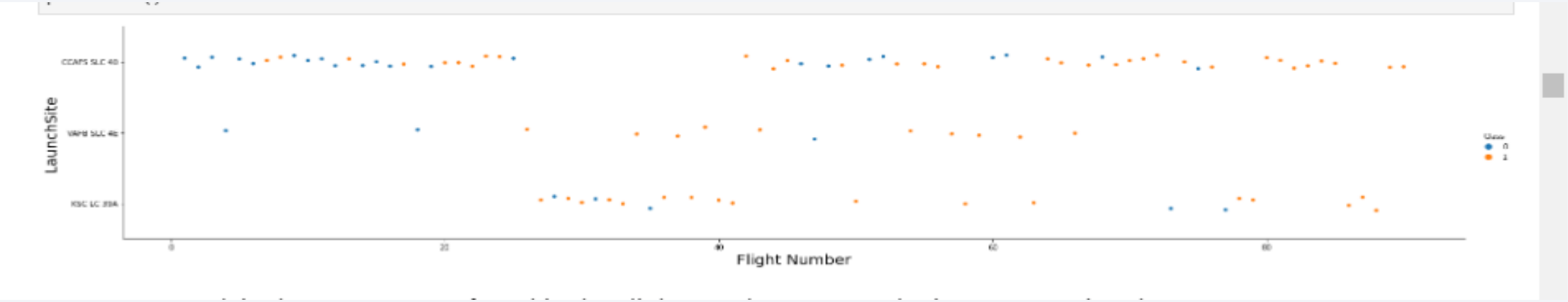
Data Wrangling

- Initially some Exploratory Data Analysis (EDA) was performed on the dataset.
- Then the summaries launches per site, occurrences of each orbit and occurrences of mission outcome per orbit type were calculated.
- Finally, the landing outcome label was created from Outcome column
- Source code:

https://github.com/tarek-ali20/Space-X-Falcon-9-First-Stage-Land-Data-wrangling-ing-Prediction/blob/main/spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- To explore data, scatterplots and barplots were used to visualize the relationship between pair of features:
  - Payload Mass X Flight Number, Launch Site X Flight Number, Launch Site X Payload Mass, Orbit and Flight Number, Payload and Orbit



Source Code : https://github.com/tarek-ali20/SpaceX-Falcon-9-First-Stage-LandinExploring-and-Preparing-Datag-Prediction-/blob/main/eda-dataviz.ipynb

EDA with SQL

The following SQL queries were performed:

- Names of the unique launch sites in the space mission;
- Top 5 launch sites whose name begin with the string 'CCA’;
- Total payload mass carried by boosters launched by NASA (CRS);
- Average payload mass carried by booster version F9 v1.1;
- Date when the first successful landing outcome in ground pad was achieved;
- Total number of successful and failure mission outcomes;
- Names of the booster versions which have carried the maximum payload mass;

Source code :

https://github.com/tarek-ali20/EDA-with-SQL/blob/main/eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Markers, circles, lines and marker clusters were used with Folium Maps
  - Markers indicate points like launch sites.
  - Circles indicate highlighted areas around specific coordinates, like NASA Johnson Space Center.
  - Marker clusters indicates groups of events in each coordinate, like launches in a launch site.
  - Lines are used to indicate distances between two coordinates.
- Code source : https://github.com/tarek-ali20/Launch-Sites-Locations-Analysis-with-Folium

Build a Dashboard with Plotly Dash

- The following graphs and plots were used to visualize data
  - Percentage of launches by site
  - Payload range
- This combination allowed to quickly analyze the relation between payloads and launch sites, helping to identify where is best place to launch according to payloads.

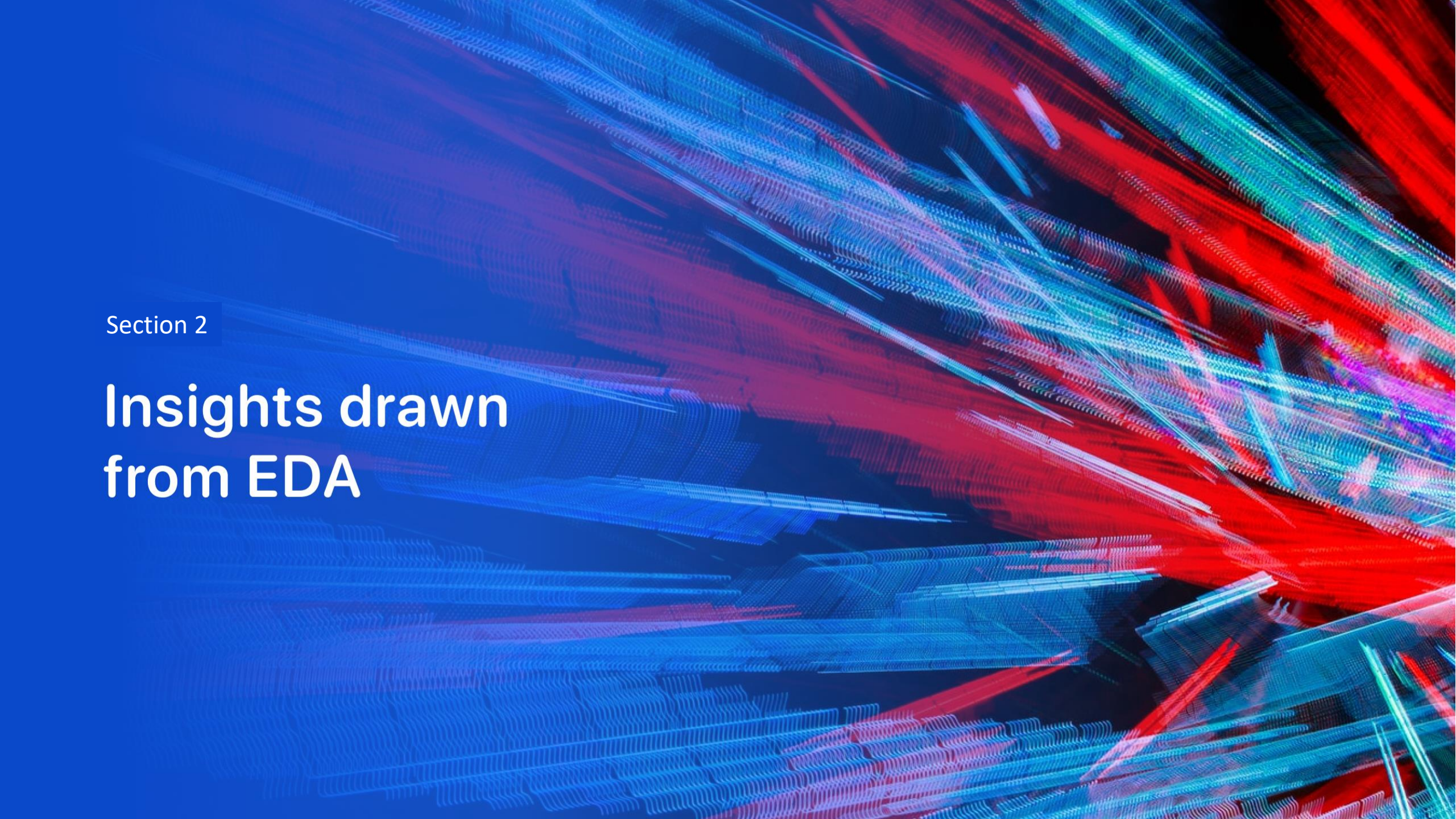
Predictive Analysis (Classification)

- Four classification models were compared:
  - logistic regression
  - support vector machine
  - decision tree
  - k nearest neighbors

- Code Source : https://labs.cognitiveclass.ai/v2/tools/jupyterlab?ulid=ulid-2563c3f548aca5a9774a3c5b1b4ed8e4809b5c0b

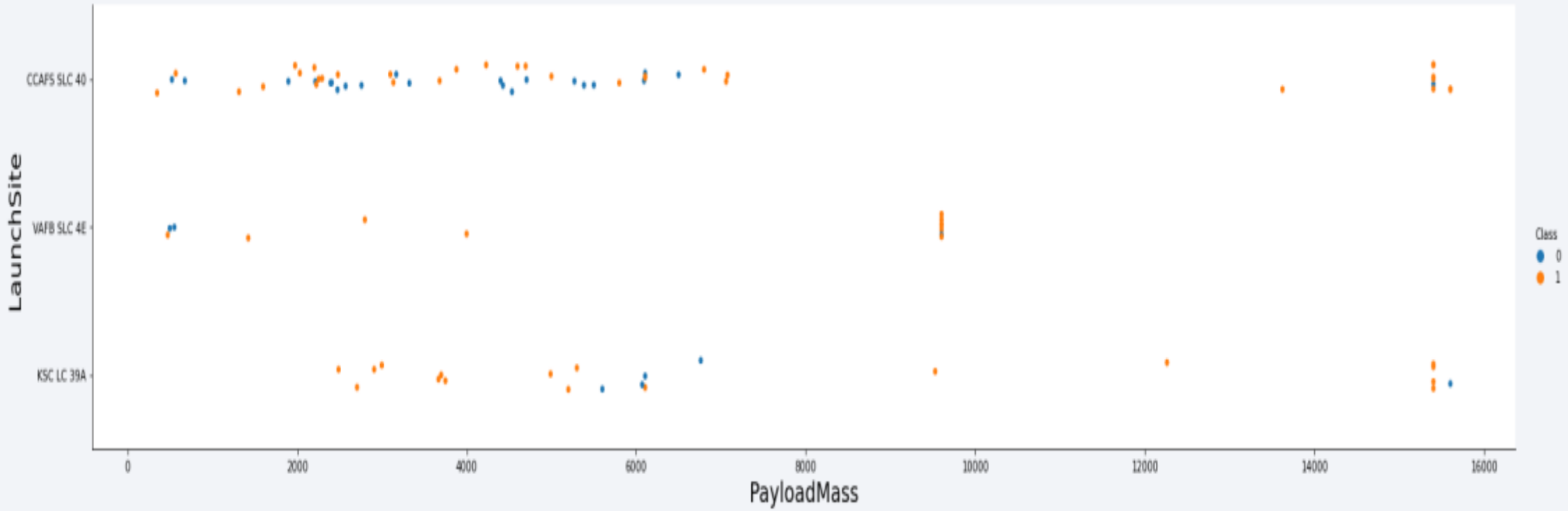
Results

- **Exploratory data analysis results:**
  - Space X uses 4 different launch sites;.
  - The first launches were done to Space X itself and NASA.
  - The average payload of F9 v1.1 booster is 2,928 kg.
  - The first success landing outcome happened in 2015 fiver year after the first launch.
  - Many Falcon 9 booster versions were successful at landing in drone ships having payload above the average.
  - Almost 100% of mission outcomes were successful.
  - Two booster versions failed at landing in drone ships in 2015: F9 v1.1 B1012 and F9 v1.1 B1015.
  - The number of landing outcomes became as better as years passed.

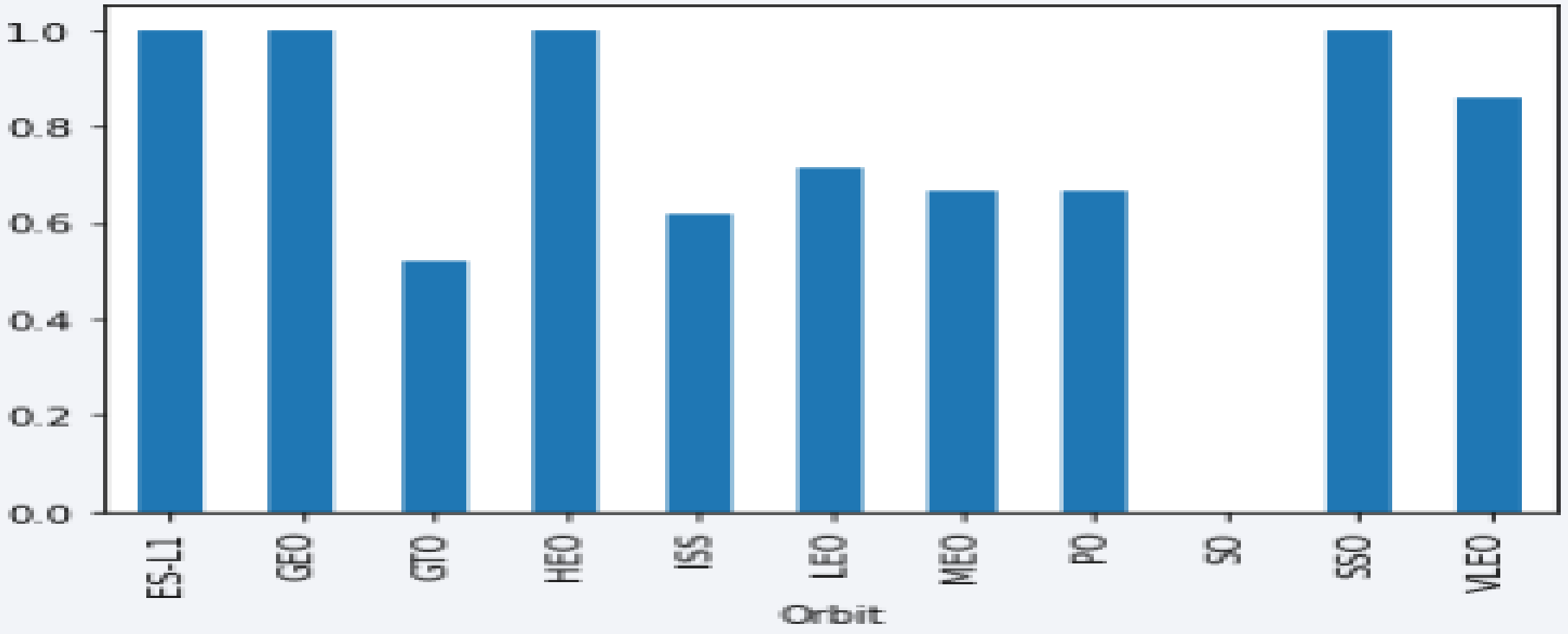

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site

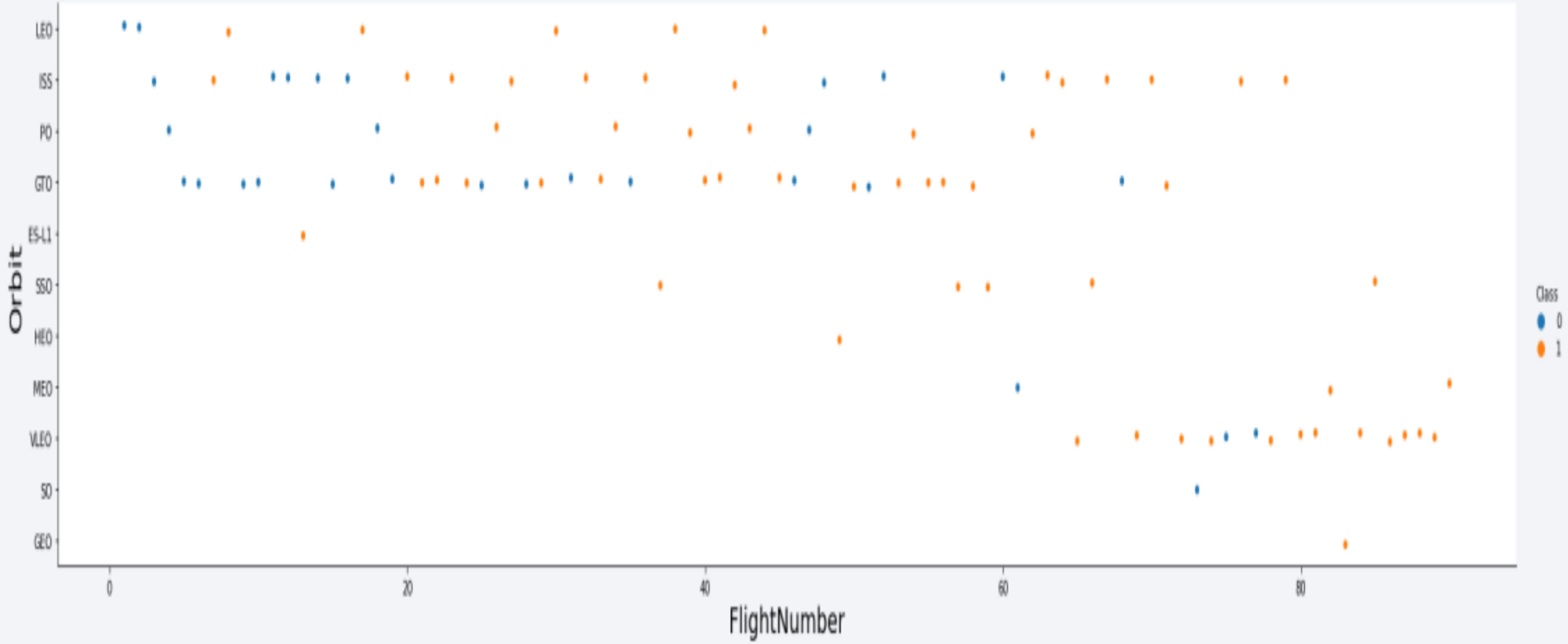
Payload vs. Launch Site



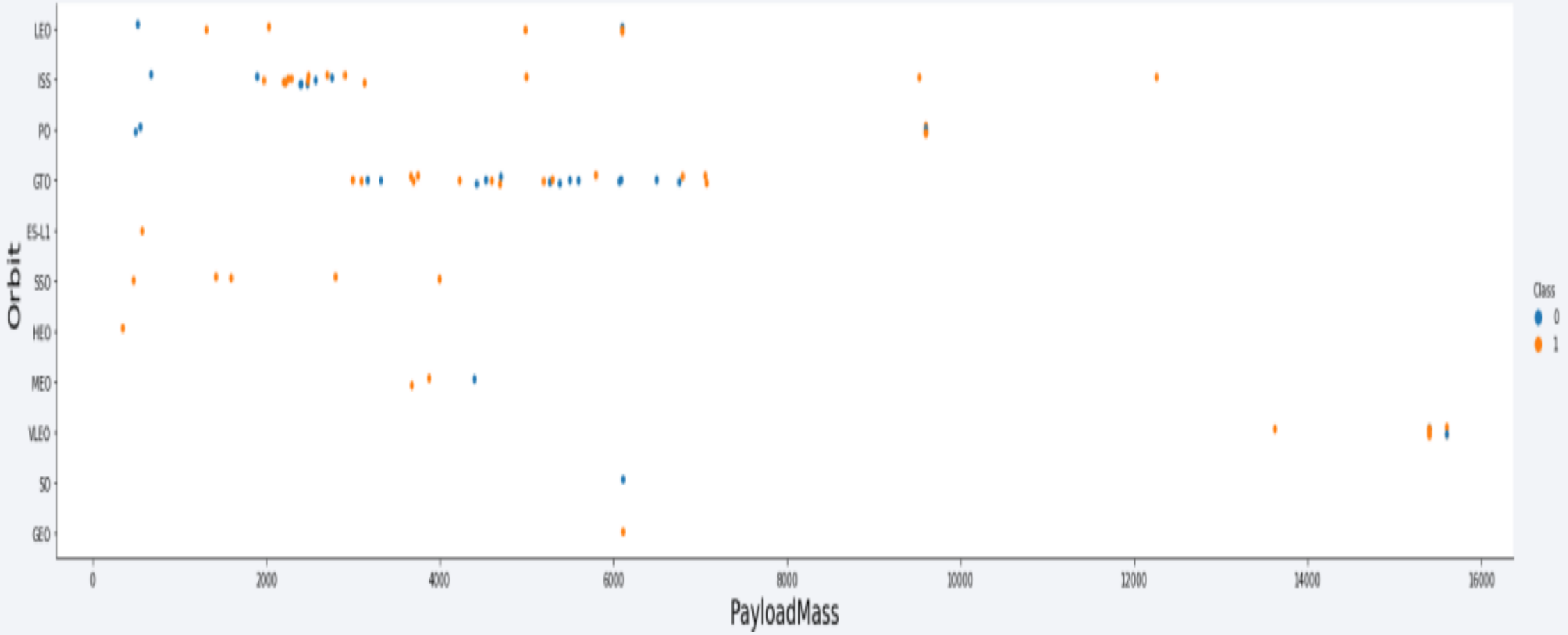
Success Rate vs. Orbit Type



Flight Number vs. Orbit Type

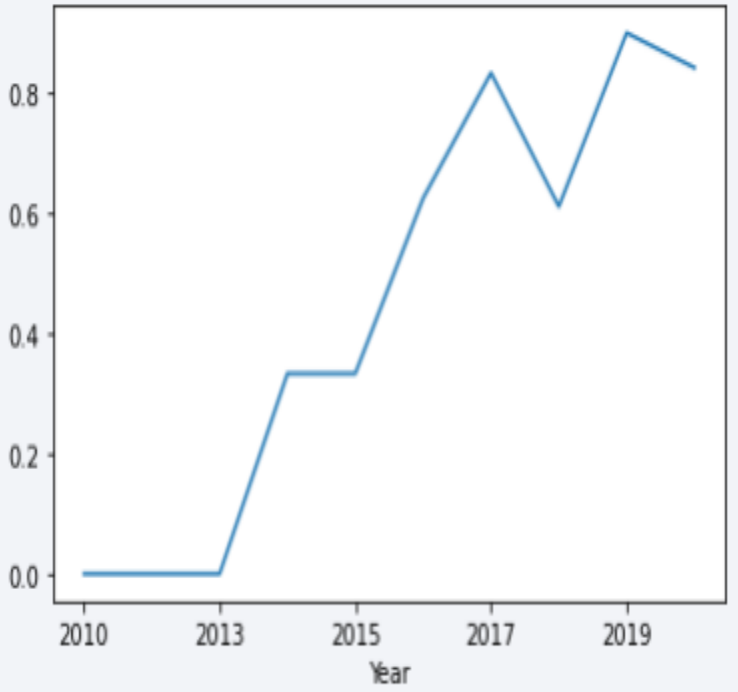


Payload vs. Orbit Type



Launch Success Yearly Trend

- Success rate started increasing in 2013 and kept until 2020;
- It seems that the first three years were a period of adjusts and improvement of technology.



All Launch Site Names

- Find the names of the unique launch sites

Launch Site
CCAFS LC-40
CCAFS SLC-40
KSC LC-39A
VAFB SLC-4E

- Present your query result with a short explanation here

They are obtained by selecting unique occurrences of “launch_site” values from the dataset.

Launch Site Names Begin with 'CCA'

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

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	**CCA**FS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	15:43:00	F9 v1.0 B0004	**CCA**FS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2012-05-22	07:44:00	F9 v1.0 B0005	**CCA**FS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	F9 v1.0 B0006	**CCA**FS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	**CCA**FS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attemp

Total Payload Mass

- Calculate the total payload carried by boosters from NASA

Total Payload (kg)
111.268

- **Total payload calculated above, by summing all payloads whose codes contain ‘CRS’, which corresponds to NASA.**

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1

Avg Payload (kg)
2.928

- **Filtering data by the booster version above and calculating the average payload mass we obtained the value of 2,928 kg.**

First Successful Ground Landing Date

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

Min Date
2015-12-22

- **By filtering data by successful landing outcome on ground pad and getting the minimum value for date it's possible to identify the first occurrence, that happened on 12/22/2015.**

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

Booster Version
F9 FT B1021.2
F9 FT B1031.2
F9 FT B1022
F9 FT B1026

- **Selecting distinct booster versions according to the filters above, these 4 are the result.**

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

Mission Outcome	Occurrences
Success	99
Success (payload status unclear)	1
Failure (in flight)	1

- Grouping mission outcomes and counting records for each group led us to the summary above.

Boosters Carried Maximum Payload

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

Booster Version (...)
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
F9 B5 B1051.3

Booster Version
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

- **These are the boosters which have carried the maximum payload mass registered in the dataset.**

2015 Launch Records

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

Booster Version	Launch Site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

- **The list above has the only two occurrences.**

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

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

- **This view of data alerts us that “No attempt” must be taken in account.**

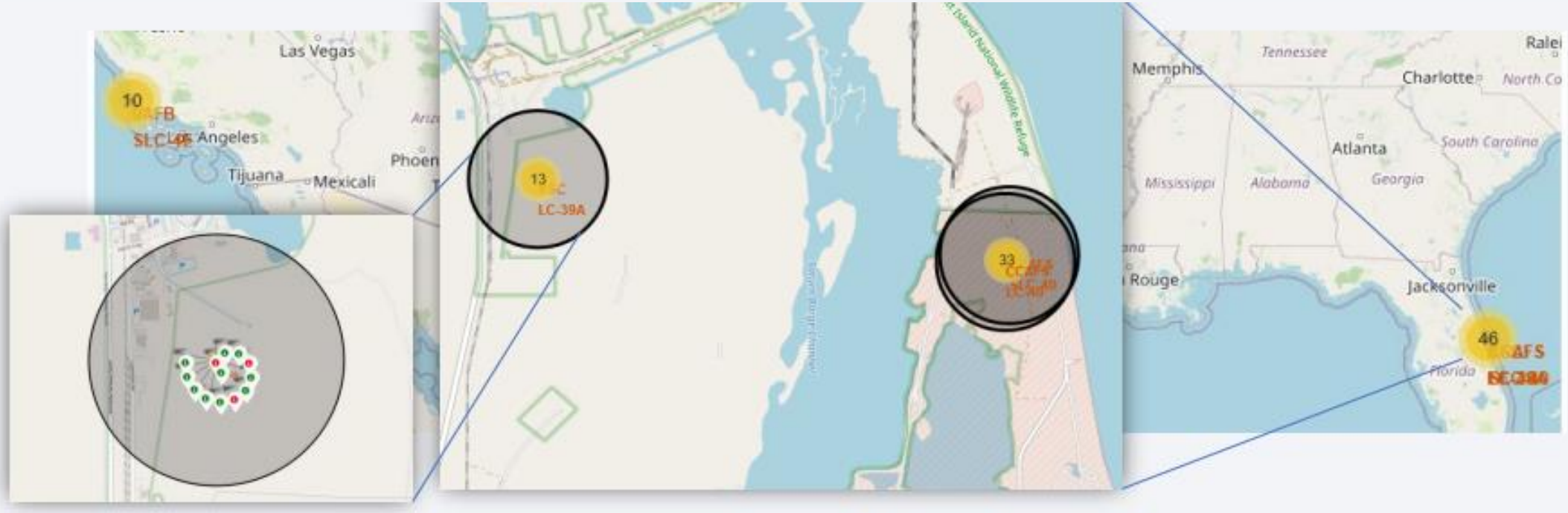
Section 3

Launch Sites Proximities Analysis

All launch sites

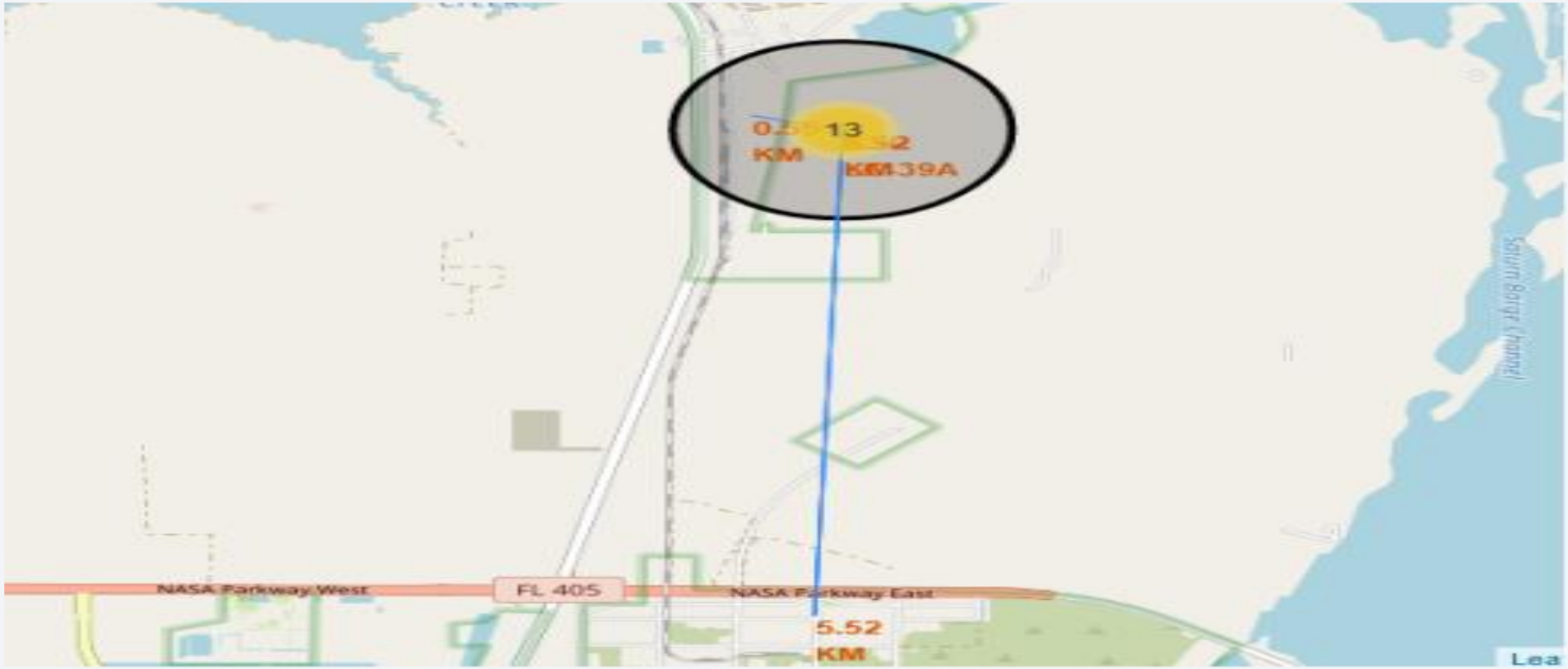
Launch sites are near sea, probably by safety, but not too far from roads and railroads.

Launch Outcomes by Site

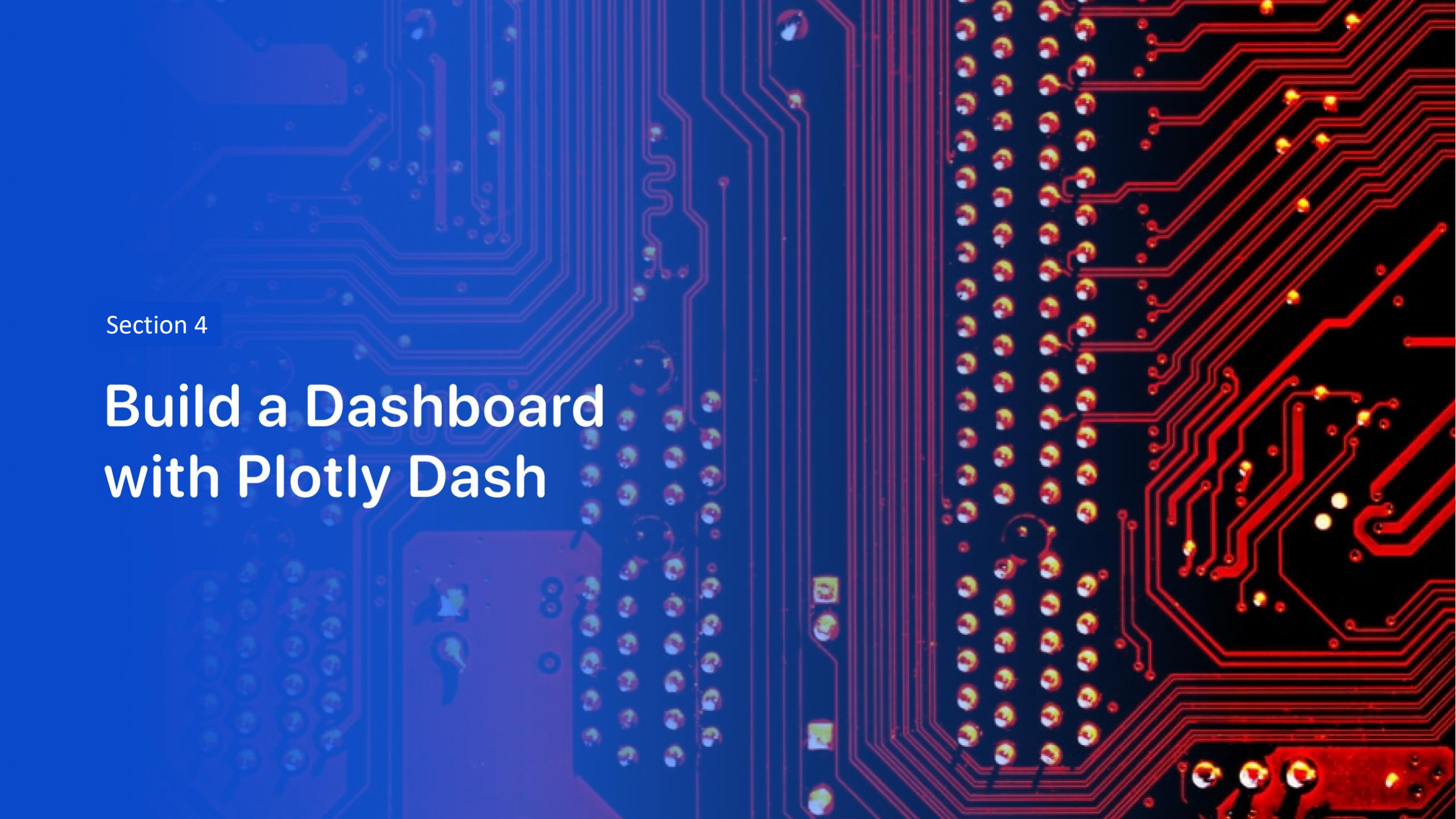


Green markers indicate successful and red ones indicate failure.

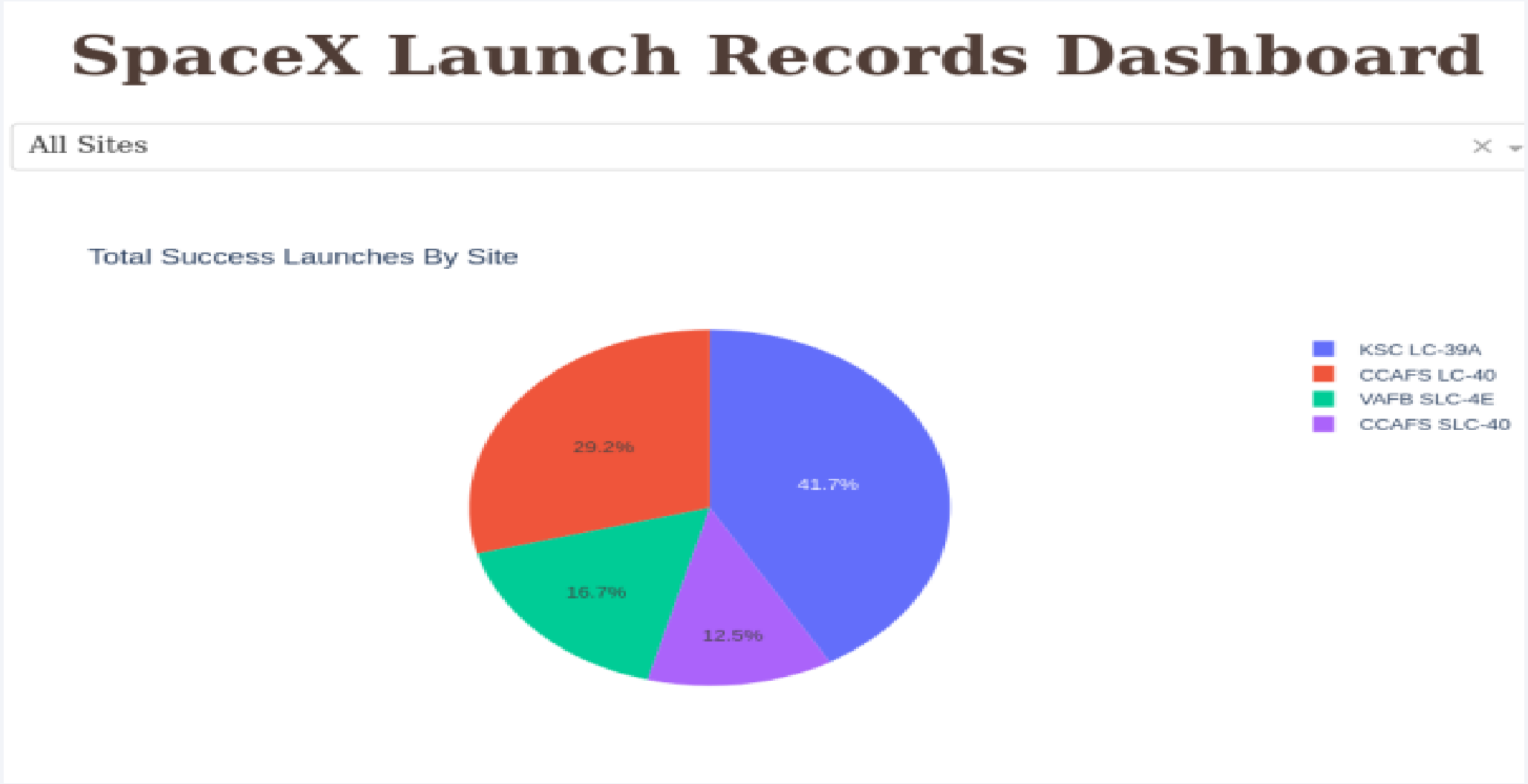
Logistics and Safety



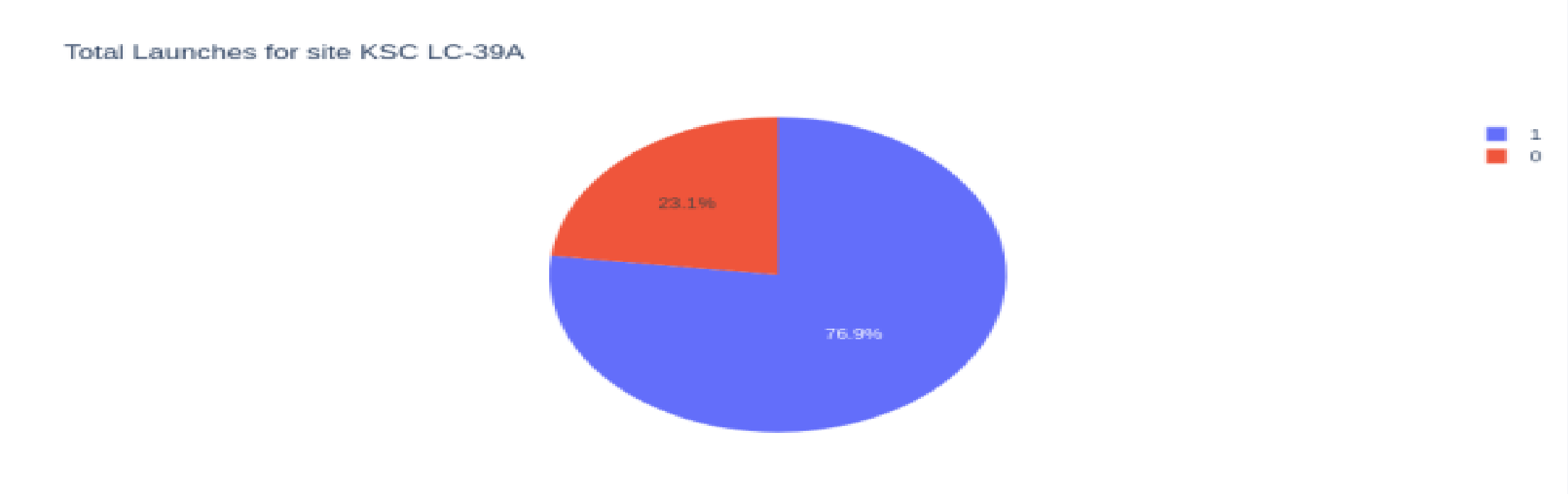
Launch site KSC LC-39A has good logistics aspects, being near railroad and road and relatively far from inhabited areas.


Section 4
Build a Dashboard with Plotly Dash

Successful Launches by Site

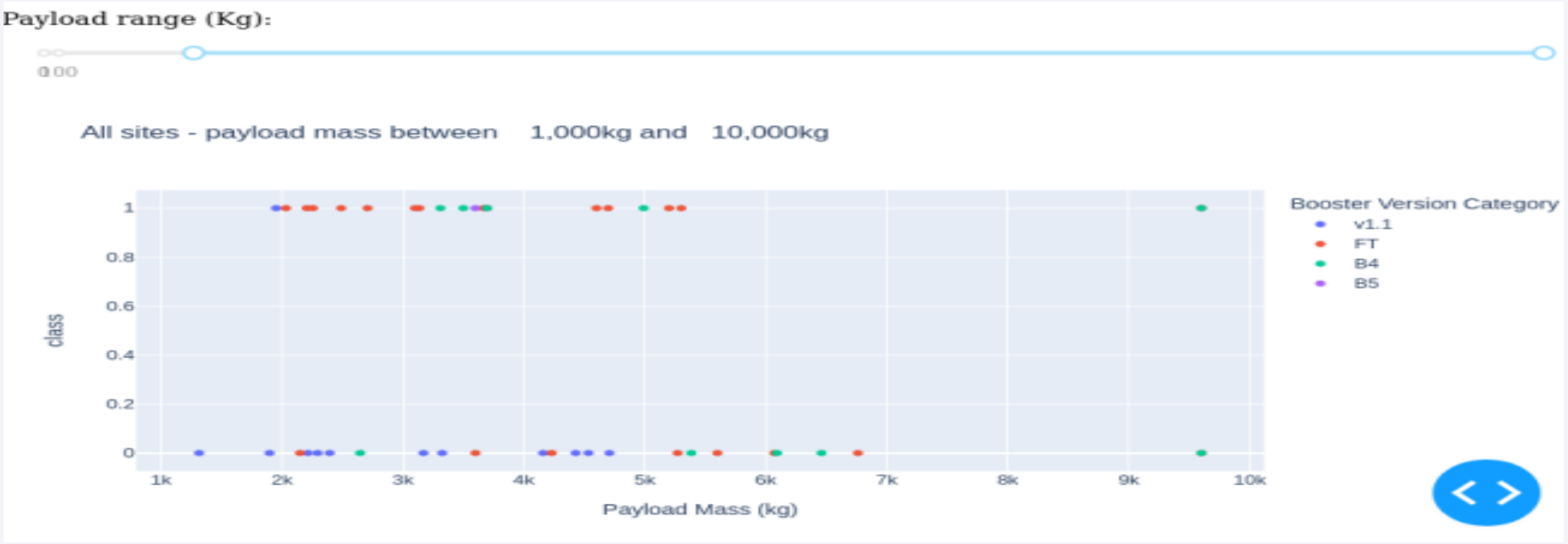


Launch Success Ratio for KSC LC-39A



- **76.9% of launches are successful in this site.**

Payload vs. Launch Outcome

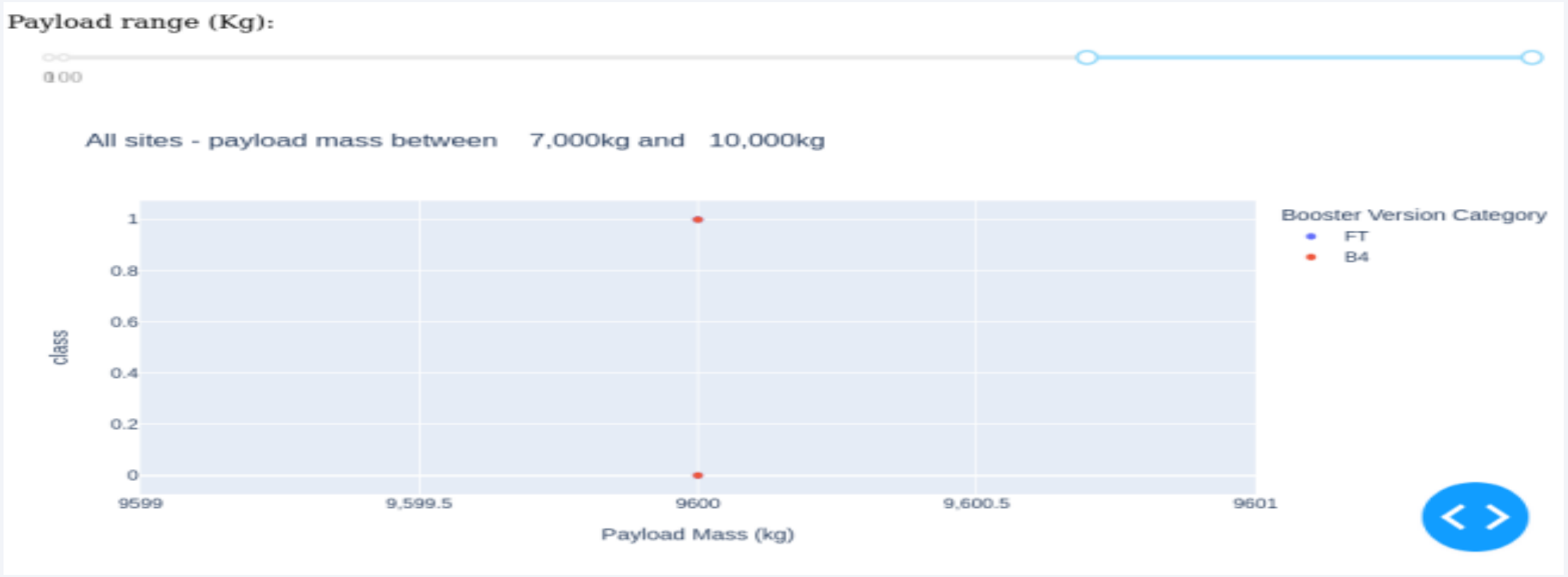


- **Payloads under 6,000kg and FT boosters are the most successful combination.**

Section 5

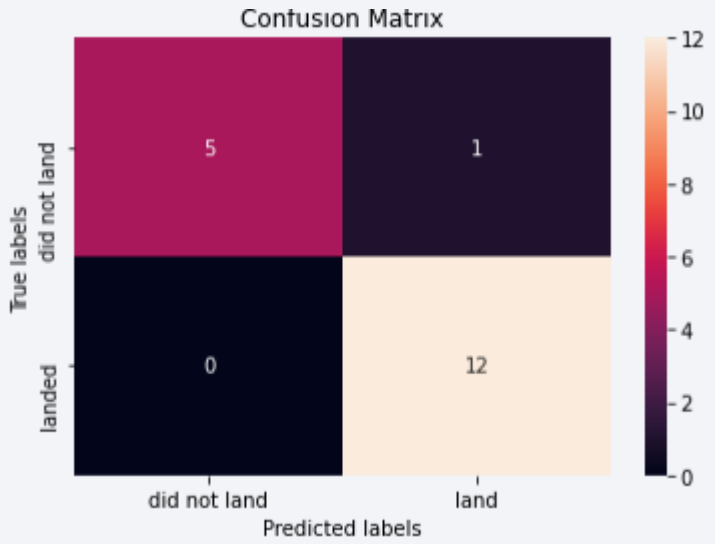
Predictive Analysis (Classification)

Classification Accuracy



- **There's not enough data to estimate risk of launches over 7,000kg**

Confusion Matrix



Confusion matrix of Decision Tree Classifier proves its accuracy by showing the big numbers of true positive and true negative compared to the false ones.

Conclusions

- **Different data sources were analyzed, refining conclusions along the process.**
- **The best launch site is KSC LC-39A.**
- **Launches above 7,000kg are less risky.**
- **Although most of mission outcomes are successful, successful landing outcomes seem to improve over time, according the evolution of processes and rockets.**
- **Decision Tree Classifier can be used to predict successful landings and increase profits**

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!