IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Shubhankar Nandakumar
2023-03-07

Outline

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

Executive Summary

- The following methodologies were used to analyze data:
    1. Data Collection using API and Web Scraping
    2. Exploratory Data Analysis (EDA) including Data Visualization, Wrangling and interactive visual analysis
    3. Machine Learning Prediction

- Summary
    1. EDA allowed us to identify features for prediction of successful launching
    2. Machine Learning Prediction helped us identify the best model to predict which characteristics are important

Introduction

- The objective is to evaluate the viability of the new company Space Y to compete with Space X

- Desired Outcomes:
  1. Estimate cost for launches
  2. Locations for successful launching

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Space X API (https://api.spacexdata.com/v4/launches/past)
  - Web scraping
    (https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches)
- Perform data wrangling
  - We created landing outcome label based on outcome data after analyzing features
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash

Methodology

- Perform predictive analysis using classification models
    - The data was normalized, split into training and testing data, and evaluated by 4 classification models – Logistic Regression, Support Vector Machines, Decision Trees and K-Nearest Neighbours. Finally, they were evaluated based on accuracy

Data Collection

- Data Sets were collected from
  - Space X API (https://api.spacexdata.com/v4/launches/past)
  - Web scraping from Wikipedia
    (https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches)

Data Collection – SpaceX API

SPACE X OFFERS PUBLIC API FROM WHERE DATA CAN BE OBTAINED AND THEN USED

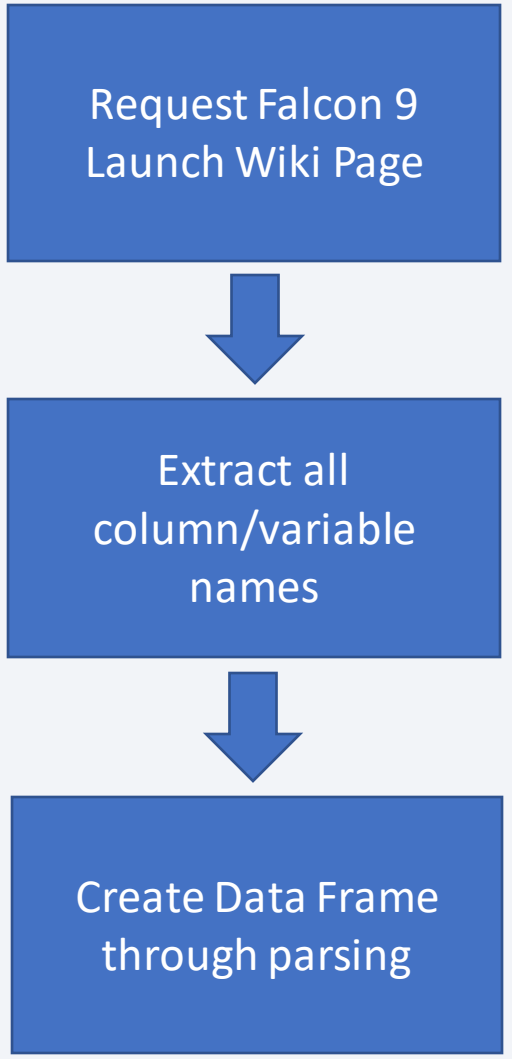
THE API WAS USED AS THE SHOWN FLOWCHART

Source Code : https://github.com/Shubhankar07/IBM-Capstone-Project/blob/main/Data%20Collection.ipynb

Data Collection - Scraping

- Data was also scrapped from Wikipedia

- The data was downloaded as shown in the flowchart



Source Code : https://github.com/Shubhankar07/IBM-Capstone-Project/blob/main/Data%20Collection%20with%20Web%20Scraping.ipynb

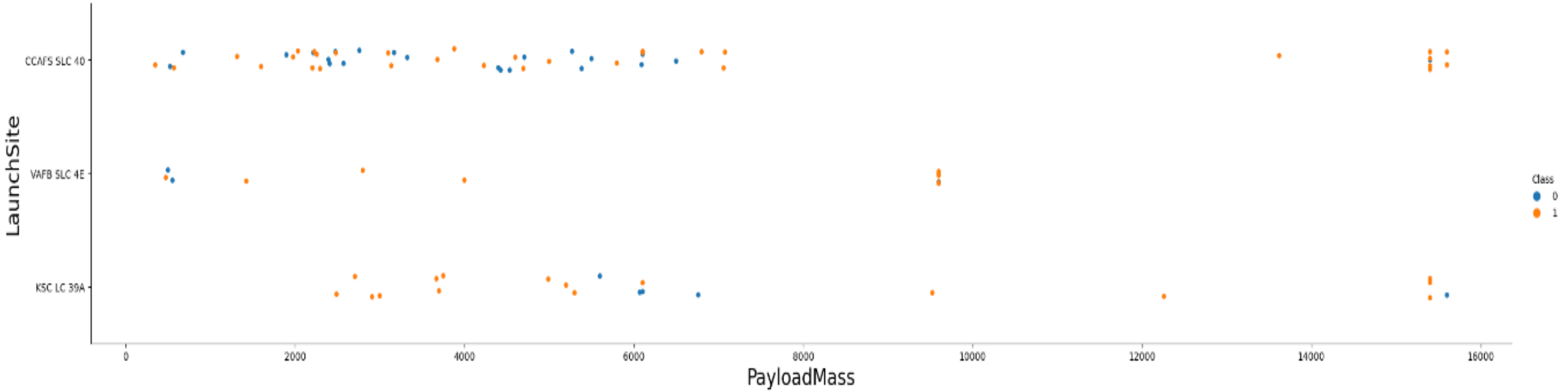
Data Wrangling

- Initially some EDA was performed on the dataset
- Then the summaries launches per site, occurrences of each orbit and occurrences of mission outcome per orbit type was calculated.
- Finally, the landing outcome labels were created

Source Code : https://github.com/Shubhankar07/IBM-Capstone-Project/blob/main/Data%20Wrangling.ipynb

EDA with Data Visualization

- To explore data, scatterplots and bar plots were used to visualize the relationship between features such as payload mass, flight number, launch site



Source Code : https://github.com/Shubhankar07/IBM-Capstone-Project/blob/main/EDA%20with%20Visualization.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
  - Names of the boosters which have success in drone ship and have payload mass between 4000 and 6000 kg
  - Total number of successful and failure mission outcomes
  - Names of the booster versions which have carried the maximum payload mass
  - Failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015
  - Rank of the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20.

Source Code : https://github.com/Shubhankar07/IBM-Capstone-Project/blob/main/EDA%20with%20SQL.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; and
- Lines are used to indicate distances between two coordinates.

Source Code : https://github.com/Shubhankar07/IBM-Capstone-Project/blob/main/Interactive%20Visual%20Analytics%20with%20Folium.ipynb

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.

Source Code : https://github.com/Shubhankar07/IBM-Capstone-Project/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

- The data was normalized, split into training and testing data, and evaluated by 4 classification models – Logistic Regression, Support Vector Machines, Decision Trees and K-Nearest Neighbours. Finally, they were evaluated based on accuracy

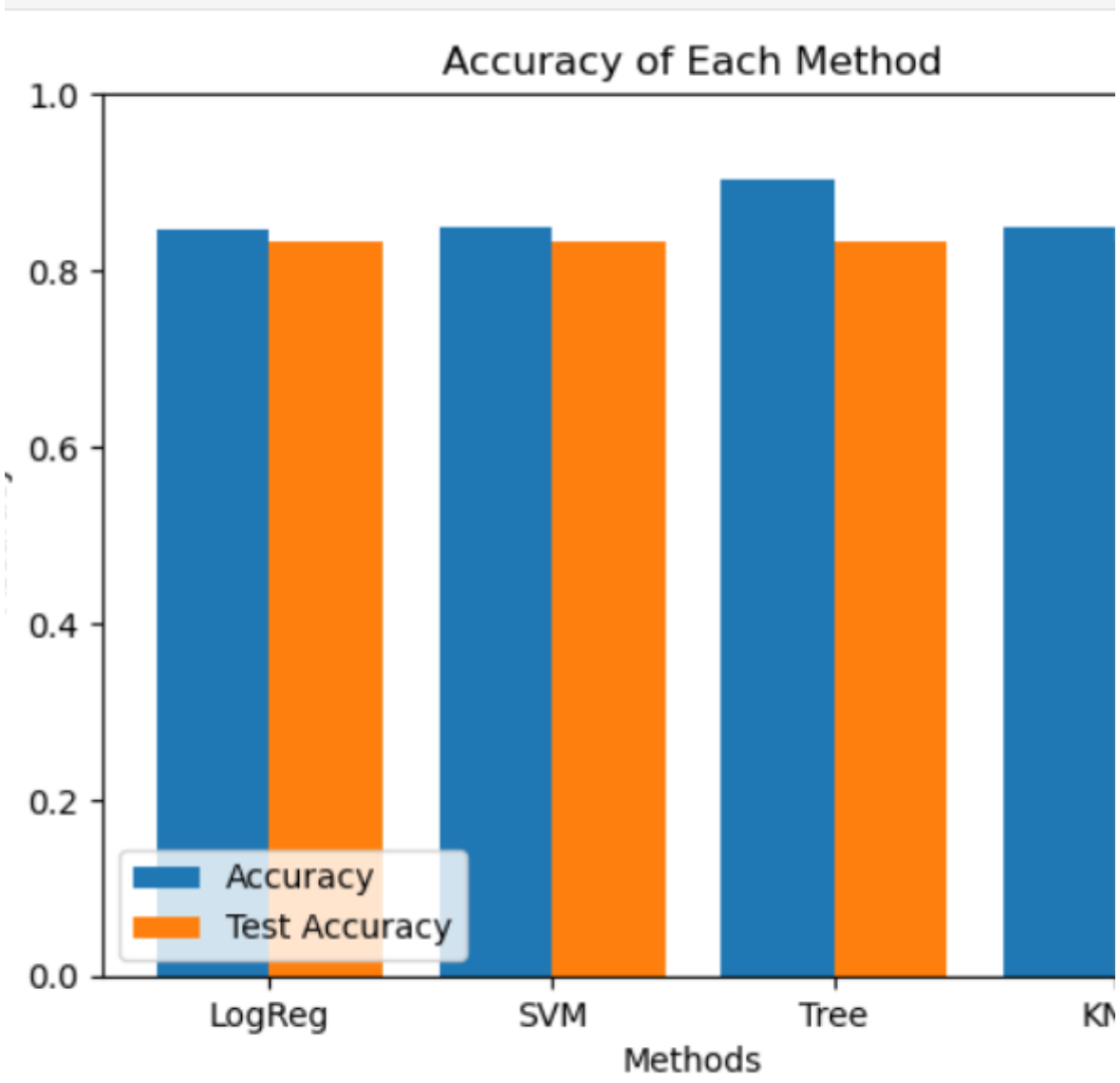
Source Code : https://github.com/Shubhankar07/IBM-Capstone-Project/blob/main/Machine%20Learning%20Prediction.ipynb

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.

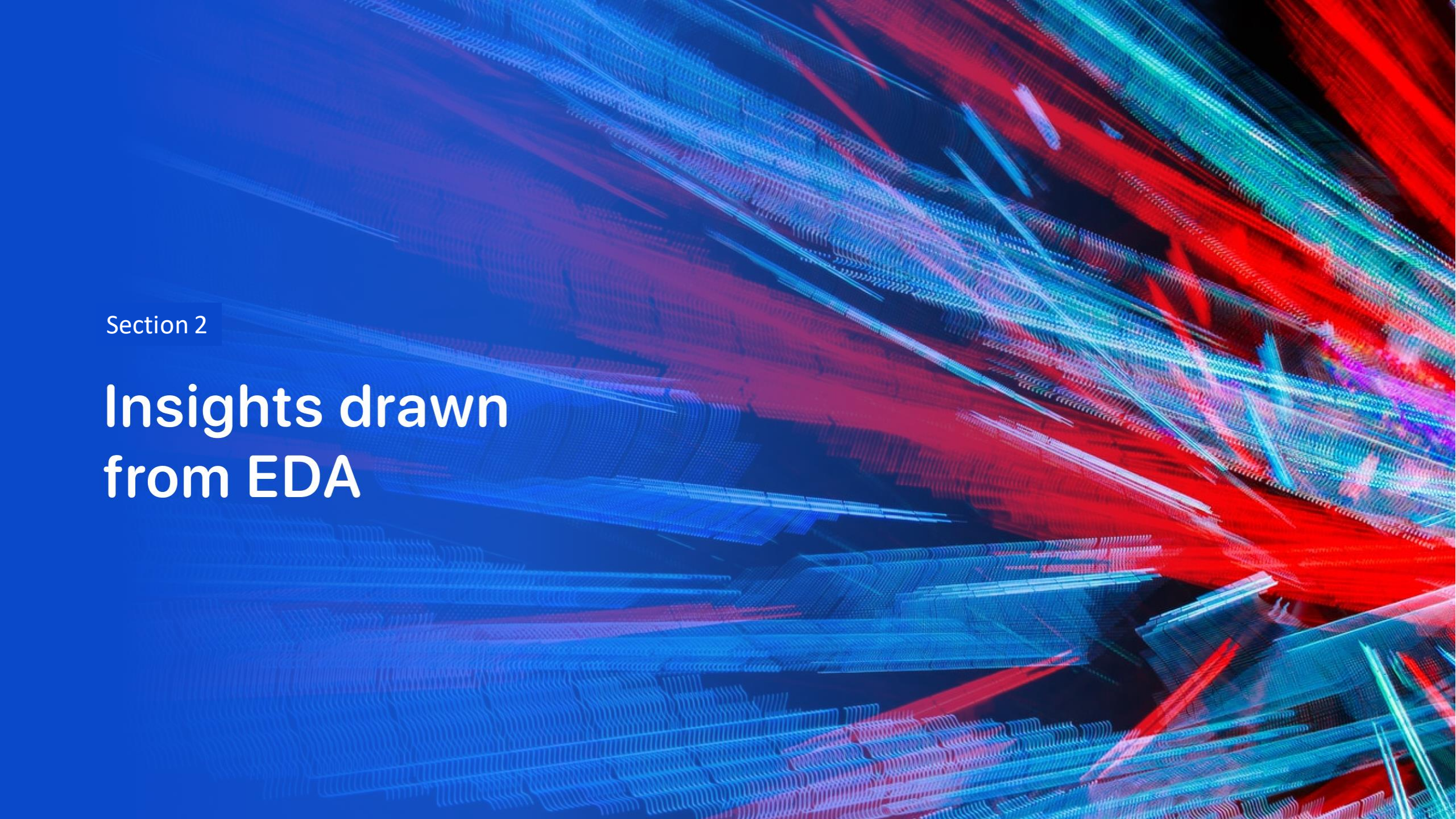
Results

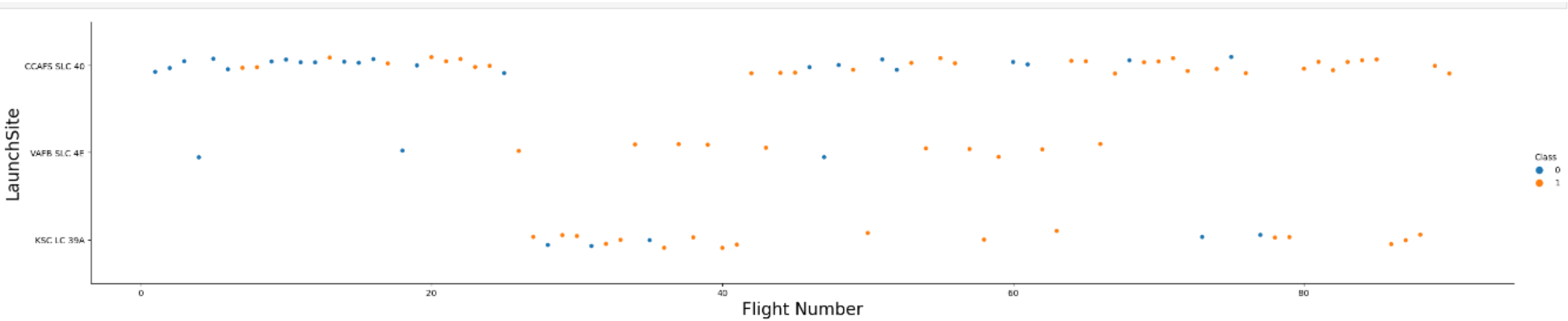
- Predictive Analysis showed that Decision Tree performed the best having an accuracy of 90.17% and Test accuracy of 83.33%



Section 2

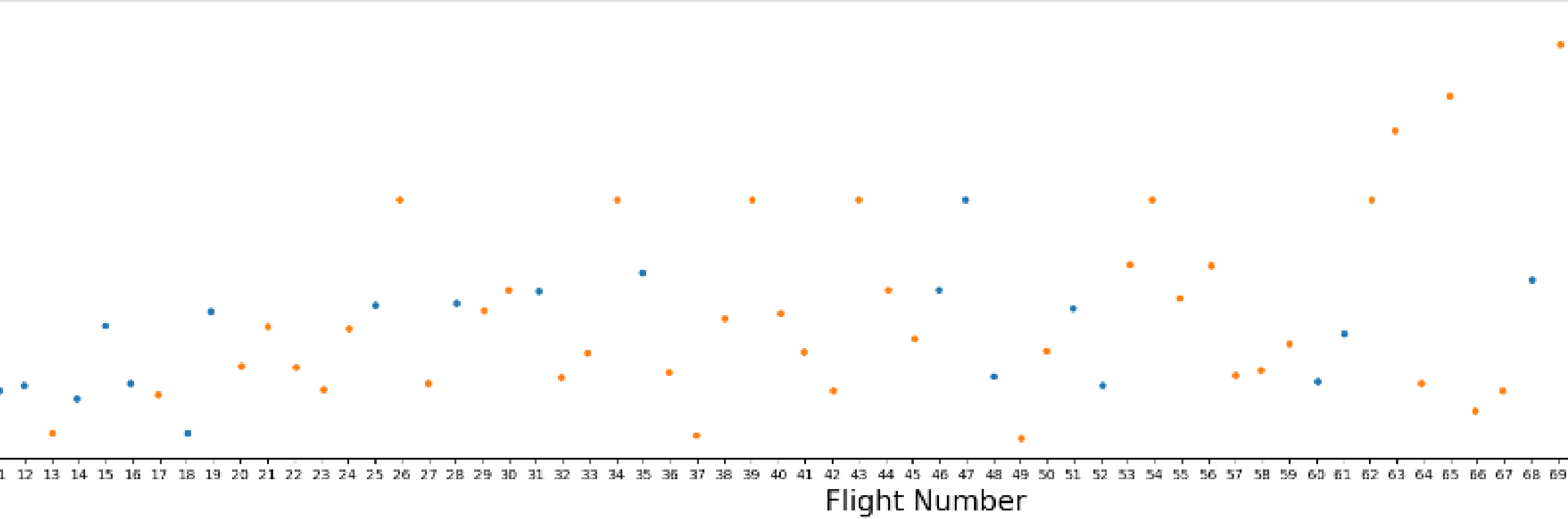
Insights drawn from EDA





Flight Number vs. Launch Site

- According to the graph, the best launch site is CCAF5 SLC 40
- Also, the general success rate has improved over time.

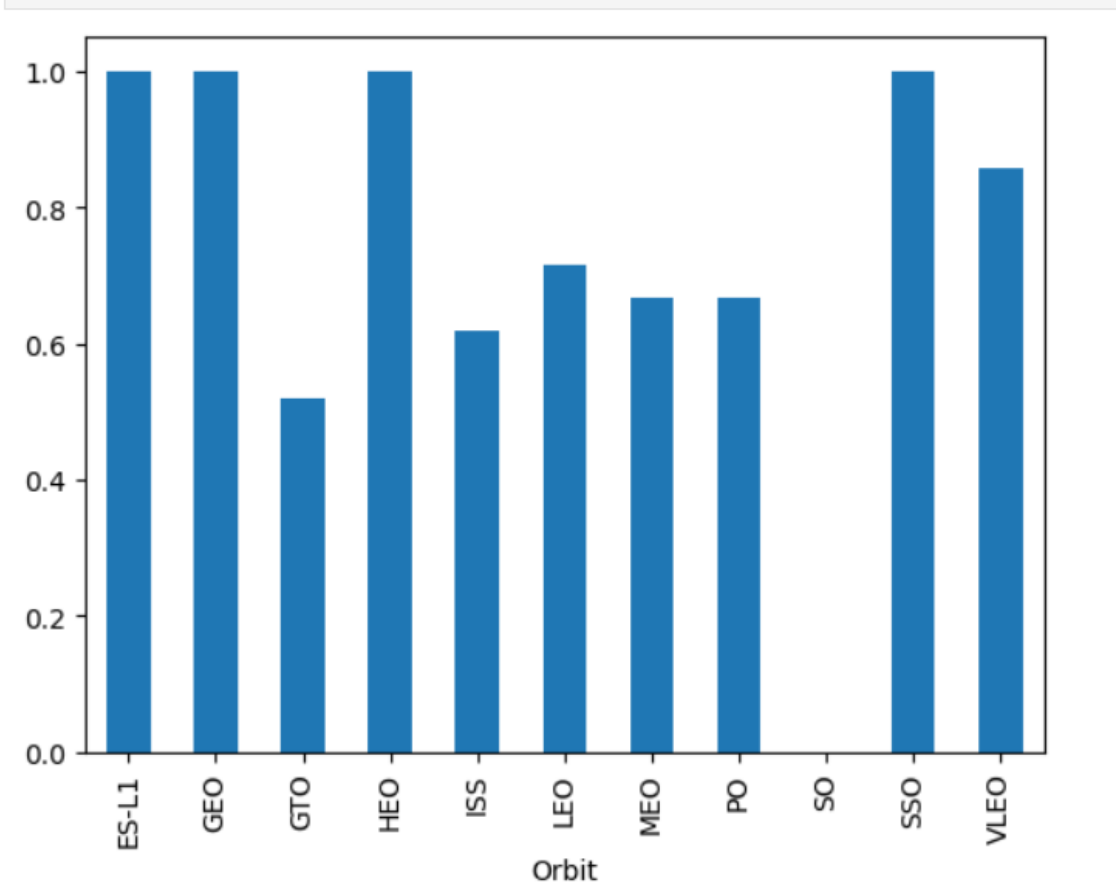


Payload vs. Launch Site

- Payloads over 9000kg have brilliant success rate
- Payloads over 12000kg are only possible at CCAFS SLC 40 and KSC LC 39A sites

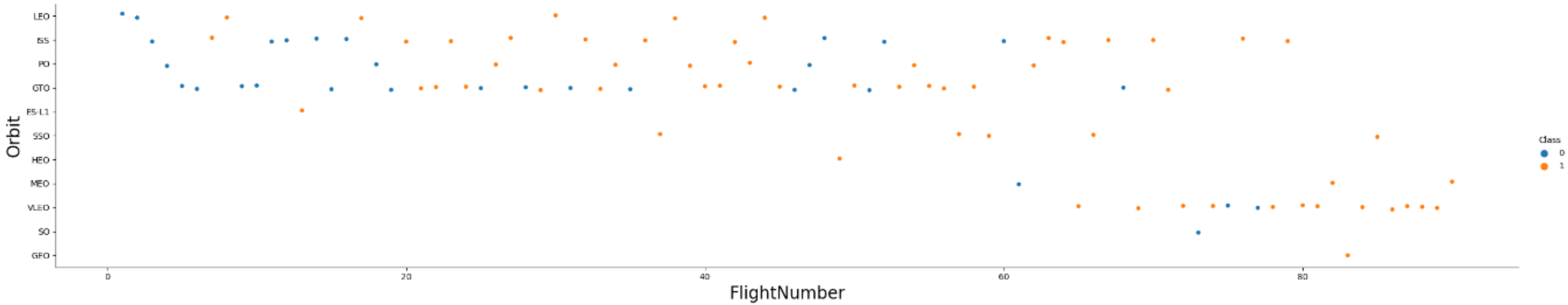
Success Rate vs. Orbit Type

- The biggest success rates happens to orbits:
  - ES-L1
  - GEO
  - HEO
  - SSO



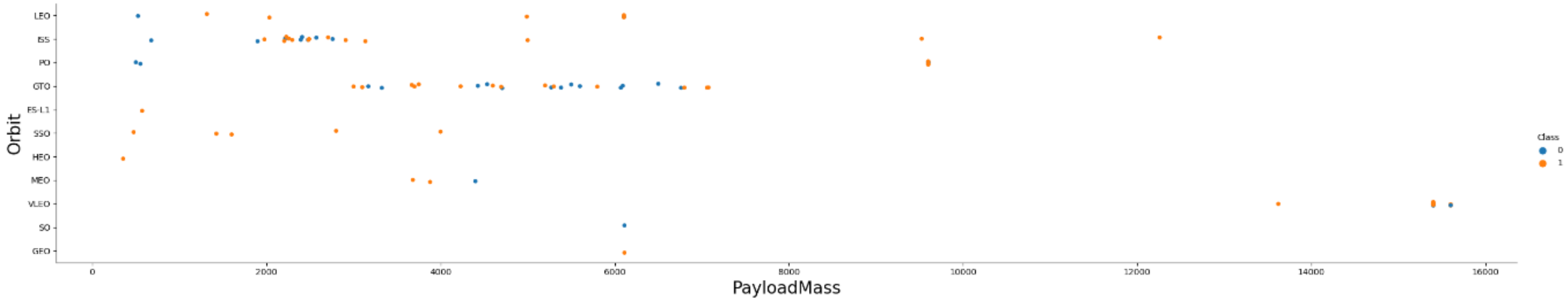
Flight Number vs. Orbit Type

- The success rate improved over time for all orbits
- VELO orbit seems like a recent like of customers

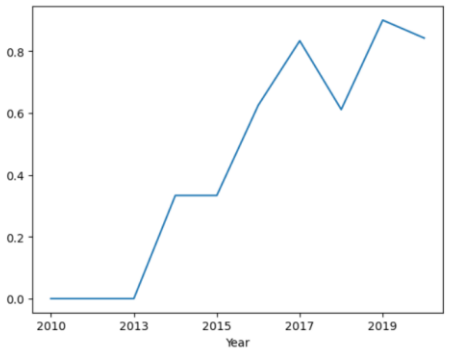


Payload vs. Orbit Type

- There is no significant relation between payload and success rate to orbit GTO
- ISS orbit has the widest range of payload and rate of success
- There are few launches to the orbits SO and GEO



- Success rate started increasing in 2013 till 2020



Launch Success Yearly Trend

All Launch Site Names

- According to the dataset, there are 4 launch sites

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

Launch Site Names Begin with 'CCA'

- Here are five samples of launches from Cape Canaveral

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	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)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- Total Payload of NASA launches

TOTAL_PAYLOAD
111268

Average Payload Mass by F9 v1.1

- Average Payload Mass by F9 v1.1
- Filtering the required booster and calculating the average payload mass

AVG_PAYLOAD
2928.4

First Successful Ground Landing Date

- This was the first successful ground landing date found out through filtering

Successful Drone Ship Landing with Payload between 4000 and 6000

- There are the boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 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

- There are the total number of successful and failure mission outcomes
- These were found by grouping records based on outcome and counting them

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

Boosters Carried Maximum Payload

- These are the names of the boosters 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
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

2015 Launch Records

Booster_Version	Launch_Site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

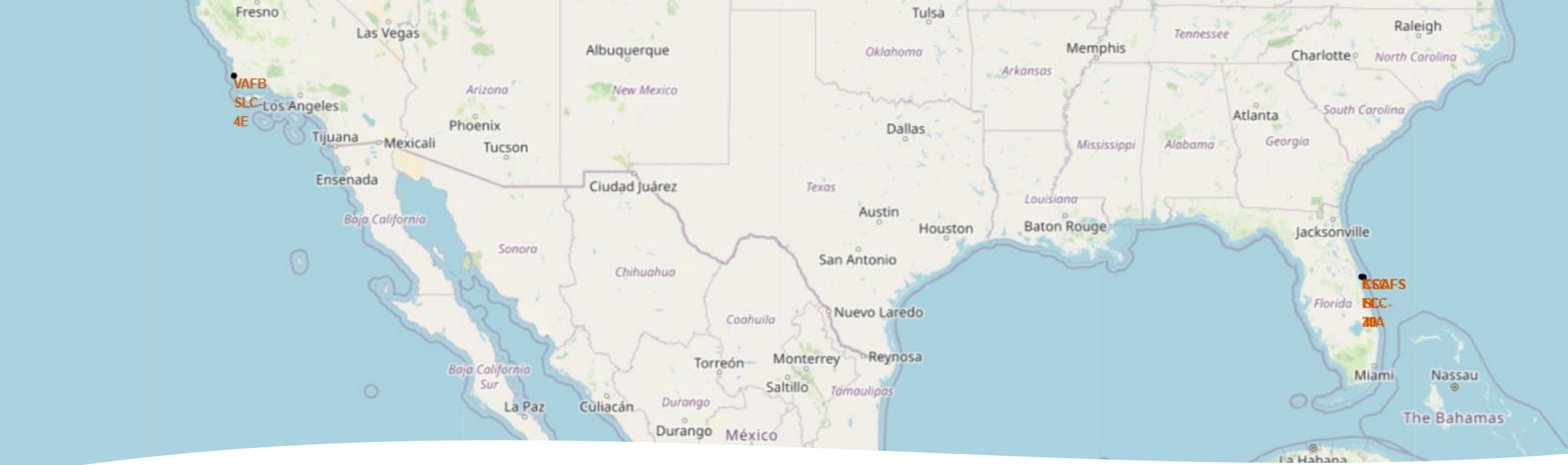
- There are the list of failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015

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

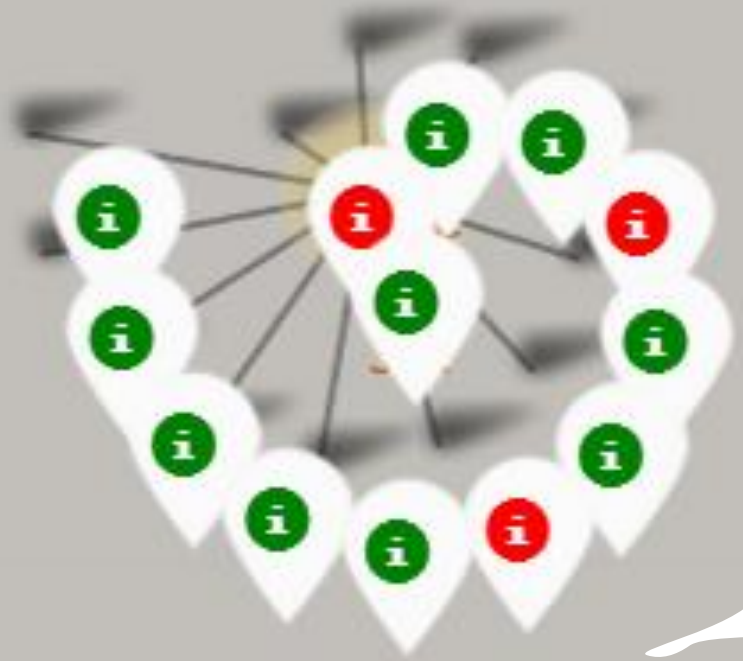
Landing _Outcome	QTY
Success	20
No attempt	10
Success (drone ship)	8
Success (ground pad)	6
Failure (drone ship)	4
Failure	3
Controlled (ocean)	3
Failure (parachute)	2
No attempt	1


Section 3
Launch Sites
Proximities Analysis



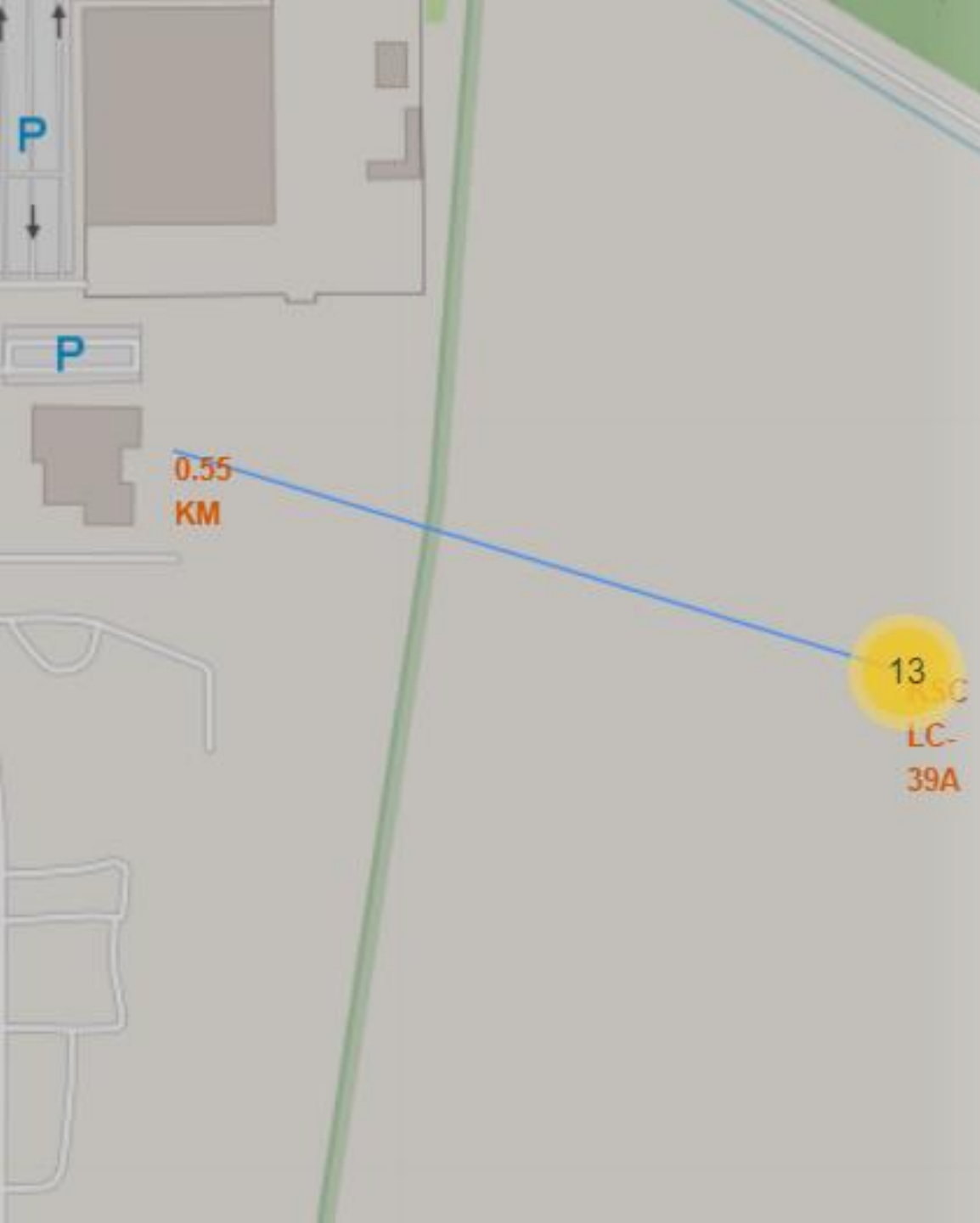
All Launch Sites

- All Launch Sites are near the sea and close to road and railways for safety as well as quick transportation of goods and resources



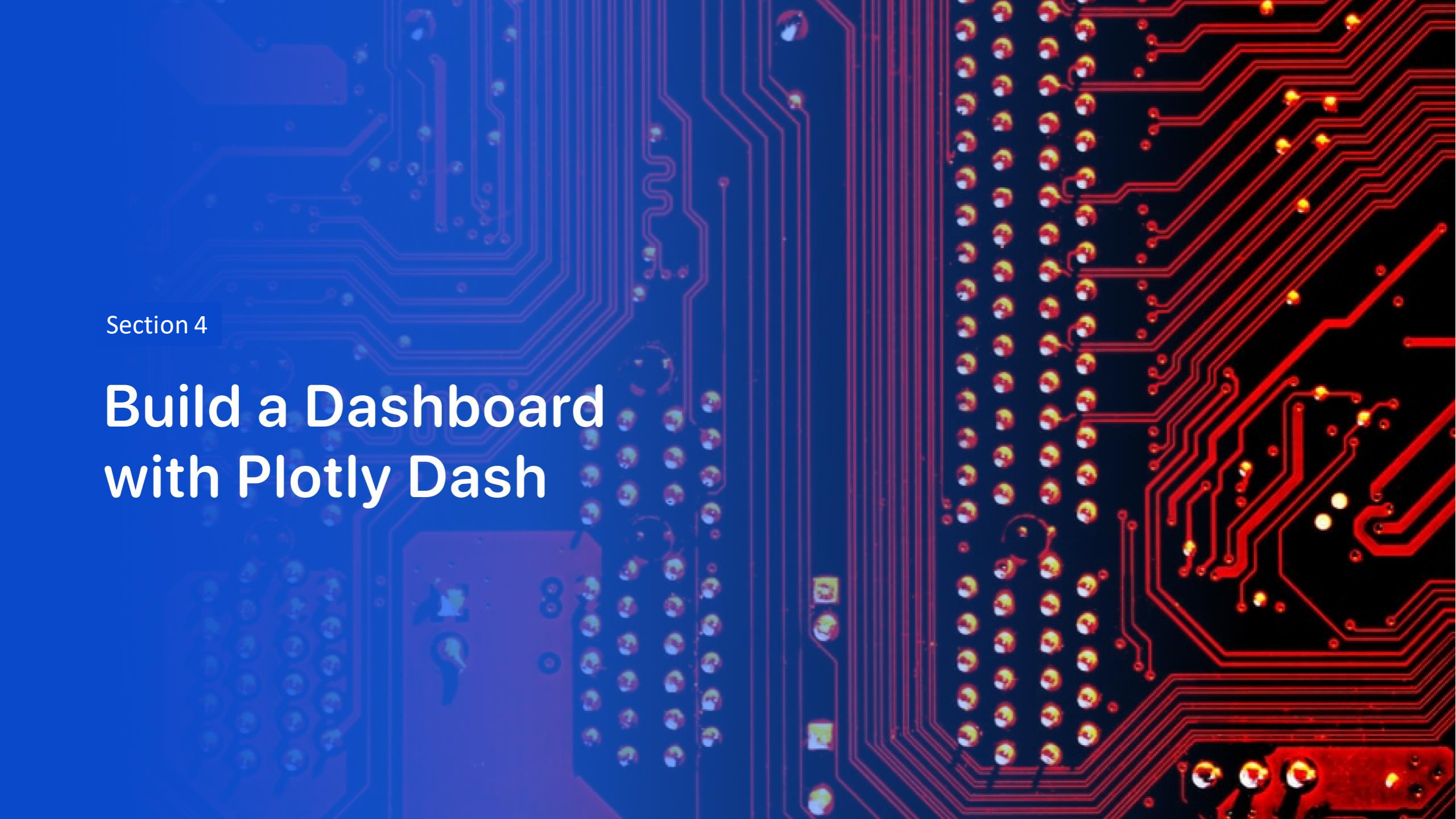
List of Launches

- The green dots represent successful launches and the red dots represent unsuccessful launches

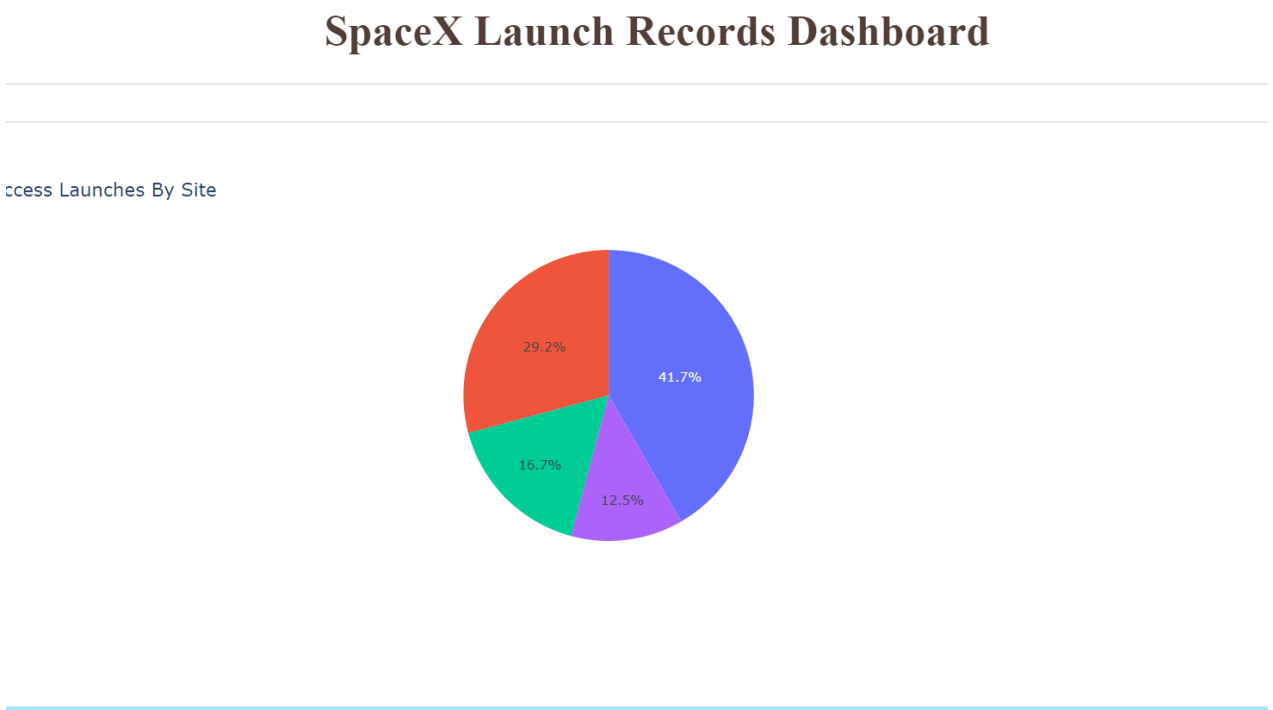


Launch Site to its Proximities

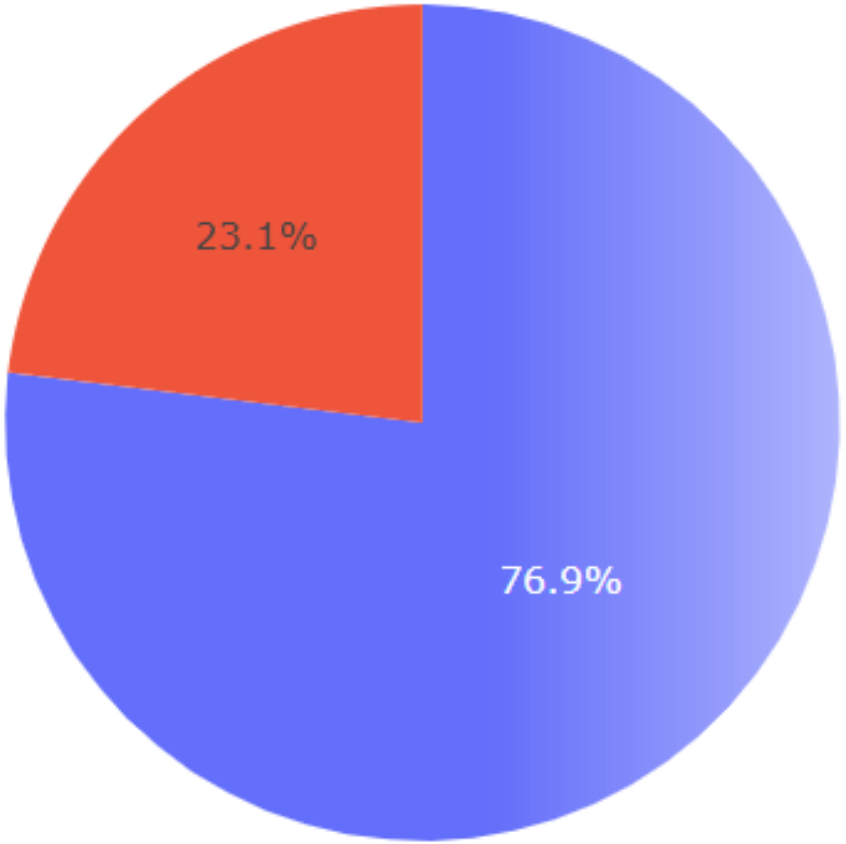
- It shows the distance between the launch site and events such as roads, railways


Section 4
Build a Dashboard with Plotly Dash

Successful Launches by Site



- The launch site has important ramifications of the success of the flight

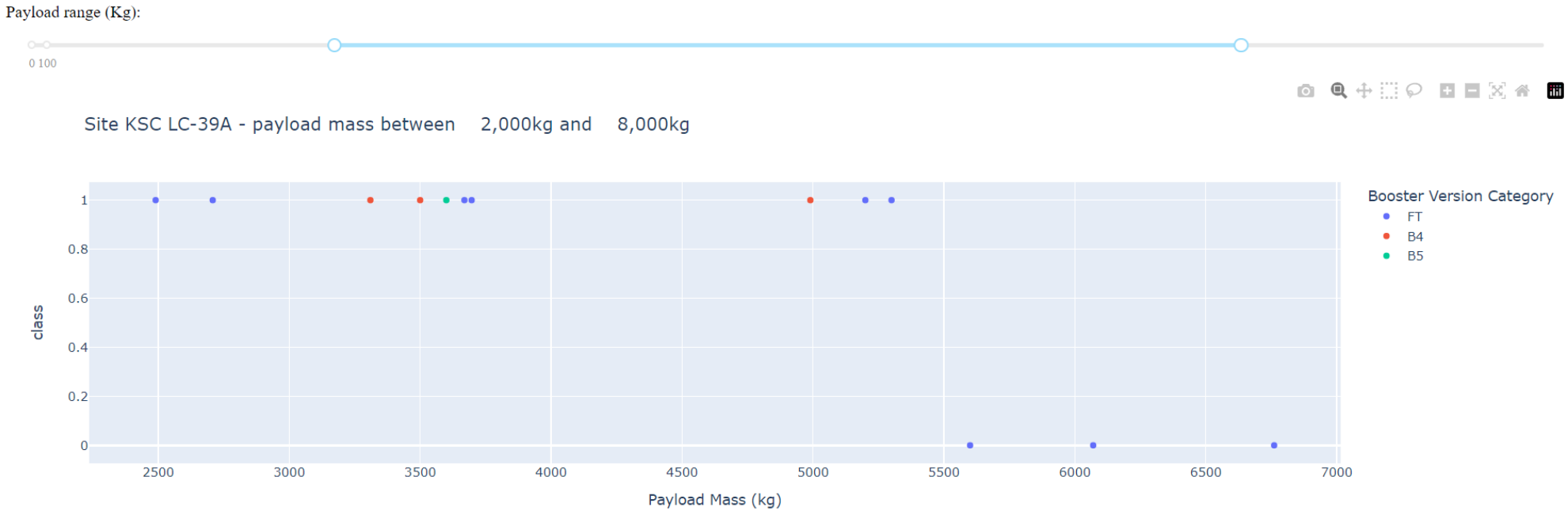


Highest Launch Success Ratio

- 76.9% success ratio

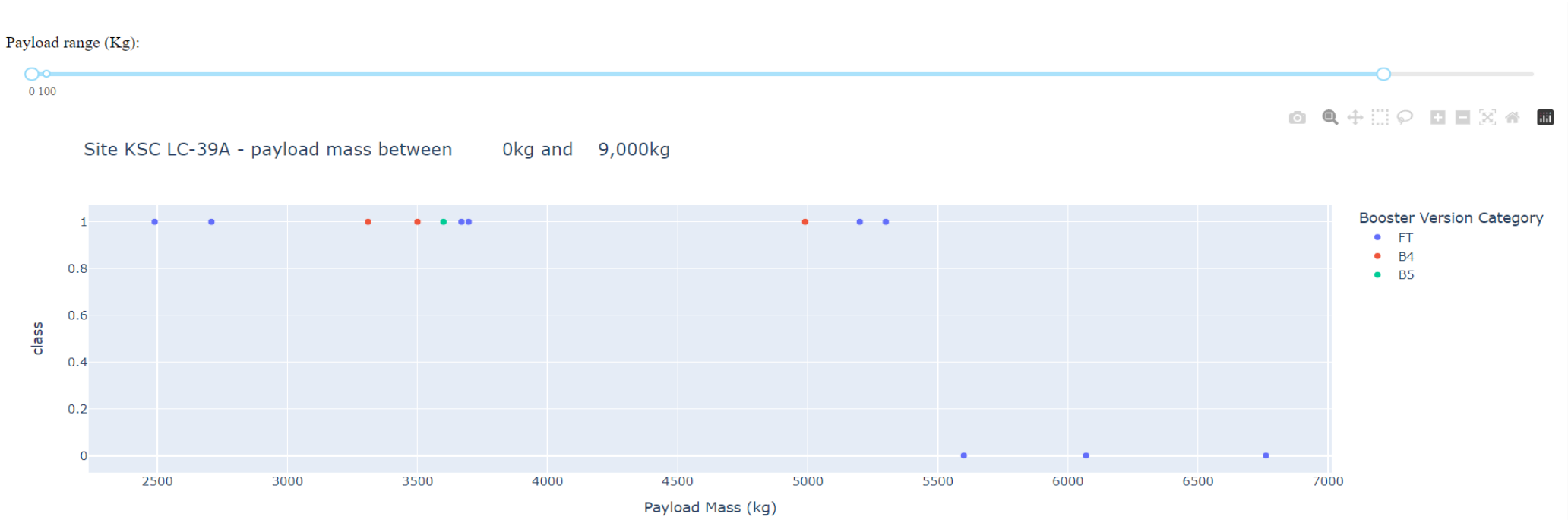
Payload vs Launch Outcome

- Shows launch of payloads between 2000kg and 8000kg



Payload vs Launch Outcome

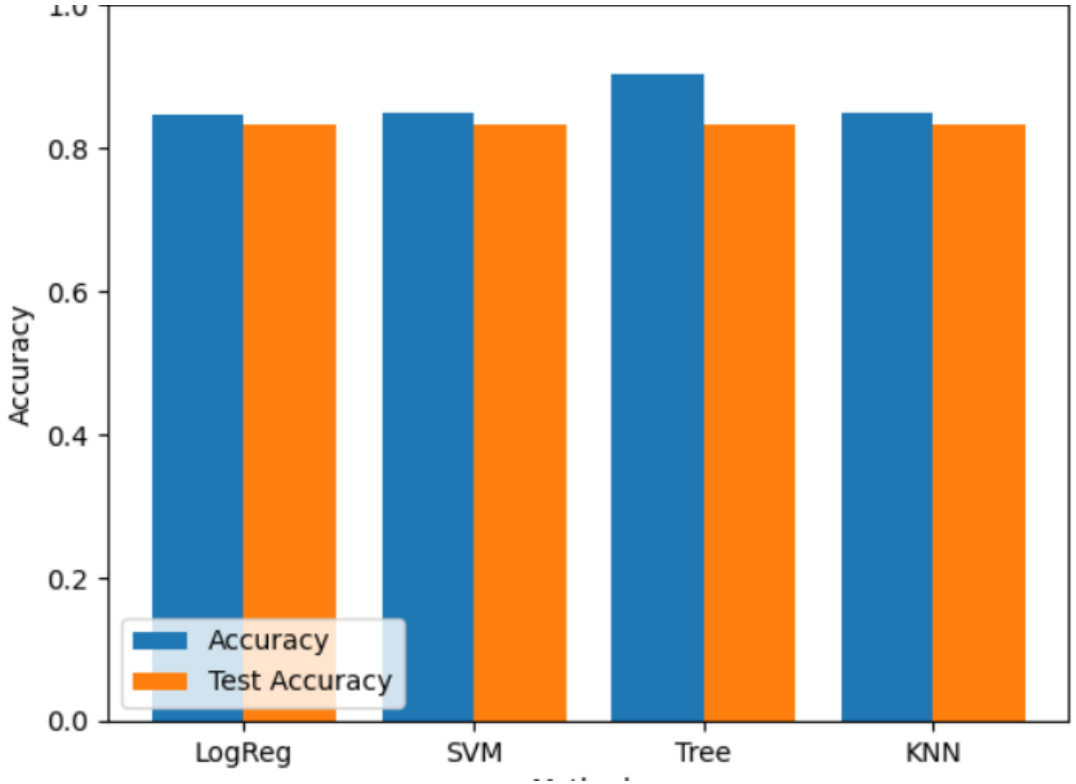
- FT and below 6000kg payload is the best combination



Section 5

Predictive Analysis (Classification)

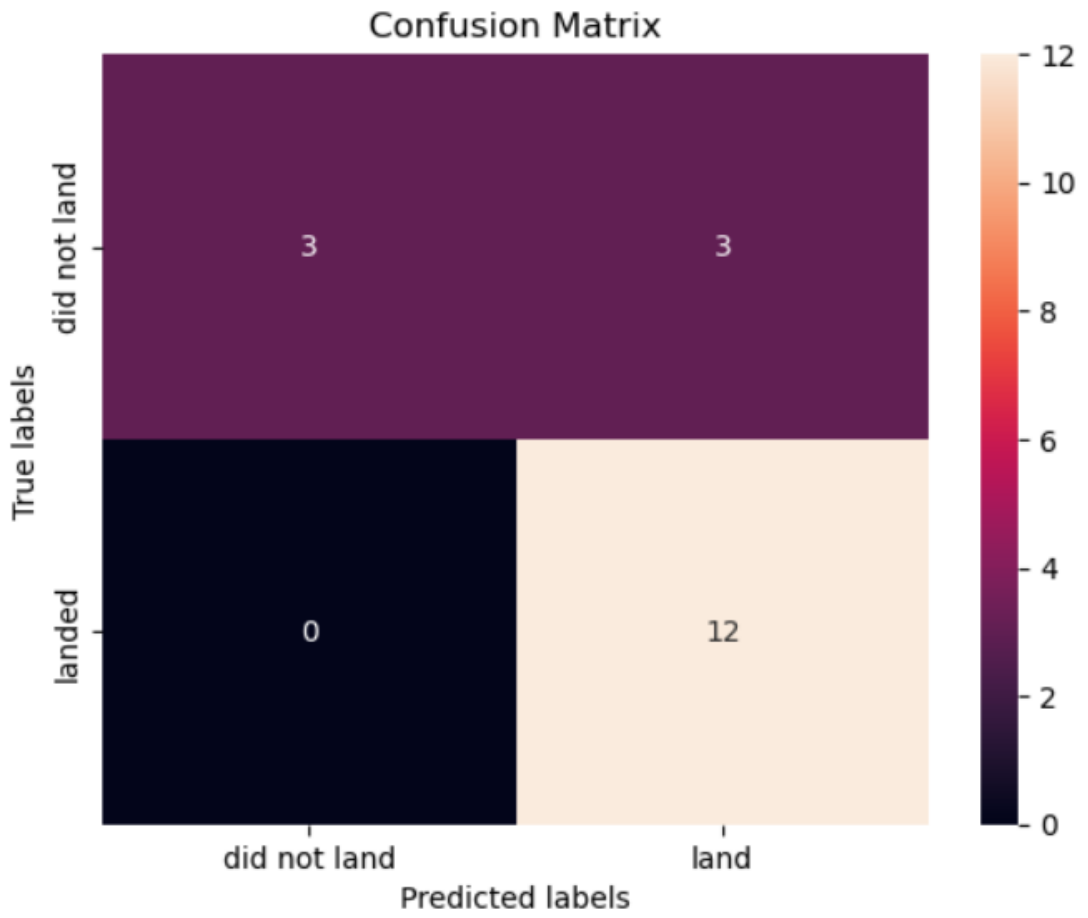
Classification Accuracy



- Predictive Analysis showed that Decision Tree performed the best having an accuracy of 90.17% and Test accuracy of 83.33%

Confusion Matrix

- The accuracy of Decision Tree is shown here as it has high True Positive and True Negative Reports



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

- Screenshots were taken wherever needed.

Thank you!