IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Rhys Powell
21/09/2022

Outline

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

Executive Summary

Summary of methodologies

- Data collection through web scraping and calling the SpaceX API.
- Data analysis, including data wrangling, data visualization and interactive visual analytics.
- Predictive analysis via machine learning techniques

Summary of all results

- Valuable and useful information successfully collected from the public data sources outlined above.
- Data analysis helped identify which features were best to predict the success of launches.
- Machine learning predictions showed the best model to predict which characterizes are import to drive this investigation using the data collected.

Introduction

Background & Context

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

Problems to address

- What is 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 launch rockets from?

Section 1

Methodology

Methodology

Executive Summary

Data collection methodology:

- Space X specific data obtained from 2 sources:
  - Space X API – http://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 enhanced by creating a landing outcome label based on outcome data after summarizing and analyzing various features.

Perform exploratory data analysis (EDA) using visualization and SQL

Methodology

Perform interactive visual analytics using Folium and Plotly Dash

Perform predictive analysis using classification models

- Data collected and manipulated was first normalized.
- Normalized data divided into training and test data sets.
- Datasets were evaluated by four different classification models.
- Accuracy of each method evaluated using different combinations of parameters.

Data Collection

Space X specific data obtained from 2 sources:

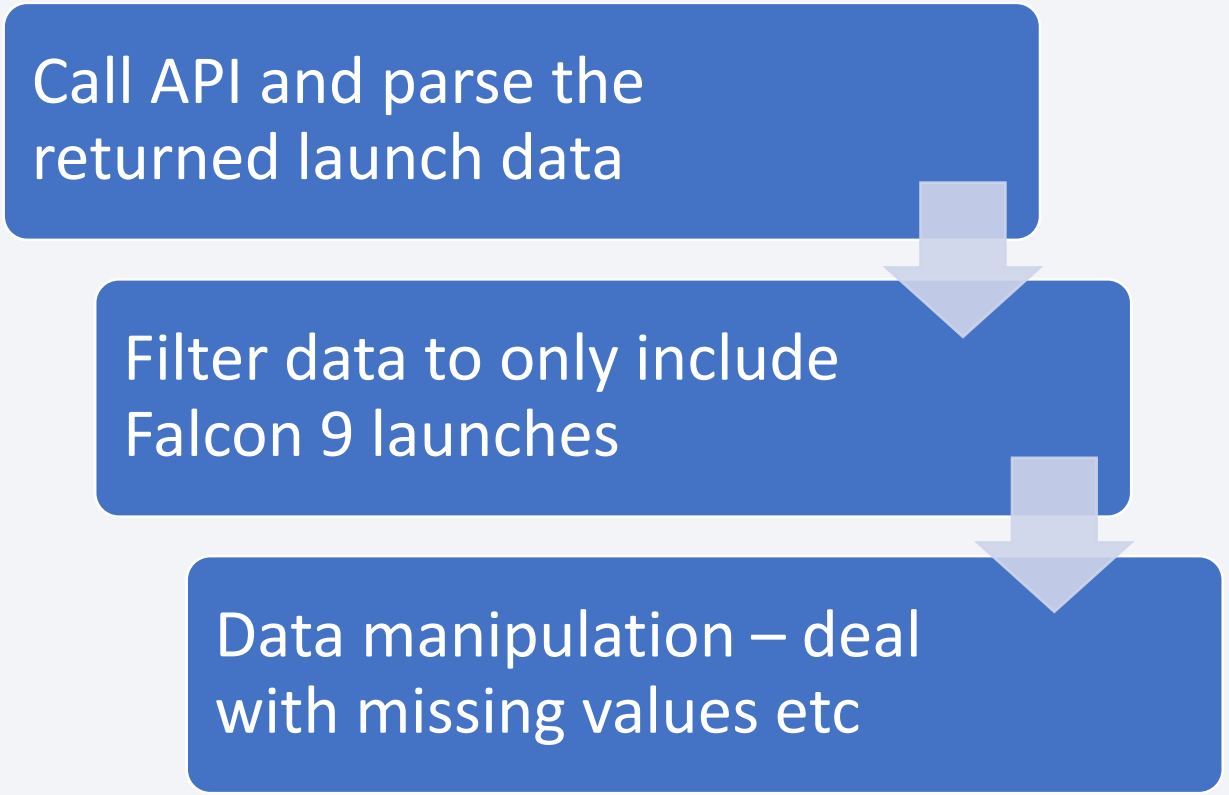
- Space X API – http://api.spacexdata.com/v4/rockets/
- Web Scraping - https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

How was the data collected?

- Utilising python requests and beautiful soup libraires to scrape the data.
- Python's Pandas library used to transform the data into a useful format.

Data Collection – SpaceX API

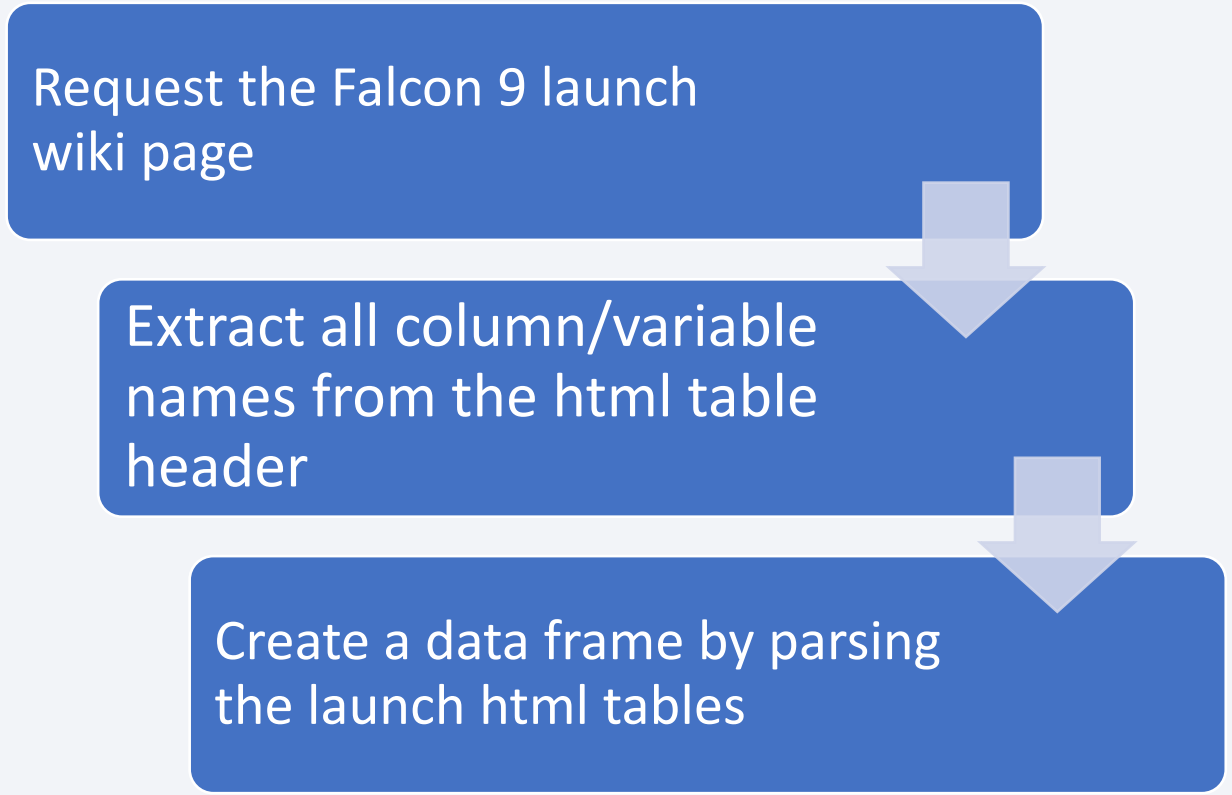
- Space X offers a public API from where data relating to launches can be obtained and used.
- The API was used per the workflow outlined in the following flowchart.



Source Code: IBM-Data-Science-Professional-Certificate/01_Data_Collection_API.ipynb at main · powellrhys/IBM-Data-Science-Professional-Certificate (github.com)

Data Collection - Scraping

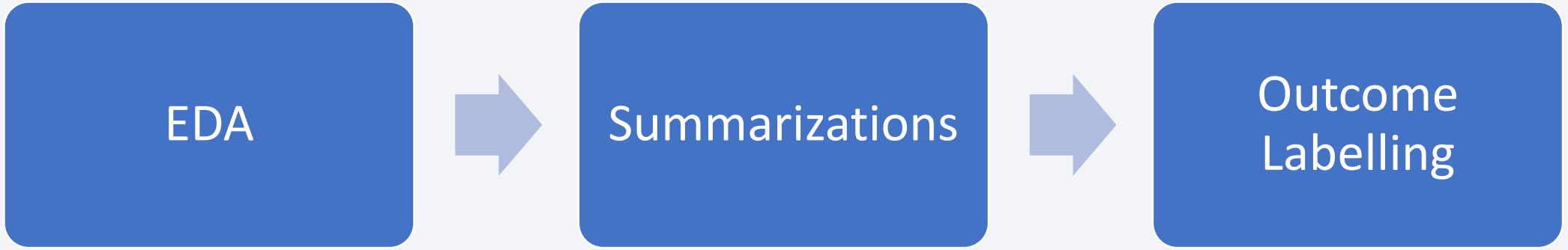
- Additional data regarding the Space X launches can be obtained from Wikipedia
- Data retrieved from Wikipedia utilising the following workflow outlined in the attached flowchart



Source Code: IBM-Data-Science-Professional-Certificate/10 Applied Data Science Capstone at main · powellrhys/IBM-Data-Science-Professional-Certificate (github.com)

Data Wrangling

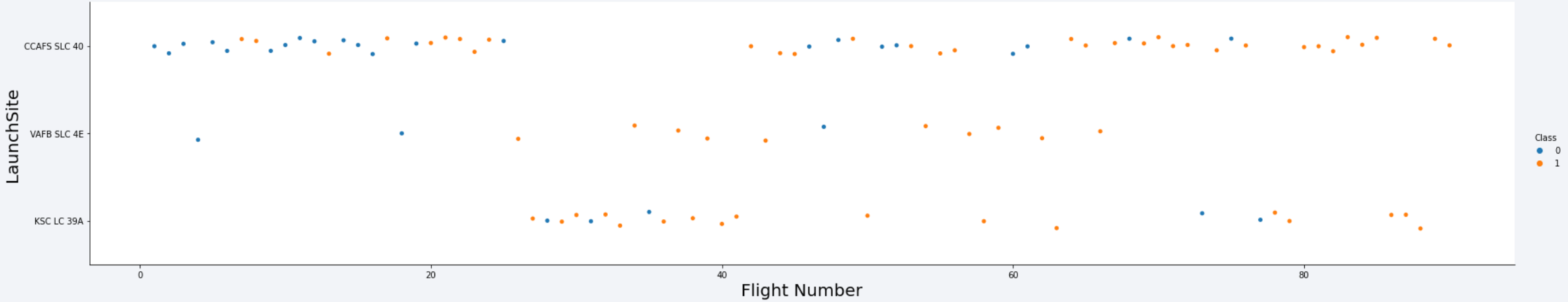
- Before going any further some initial exploratory data analysis (EDA) was performed on the dataset.
- Following this, summaries of the launches per site, occurrences of each orbit and occurrences of mission outcome per orbit type were calculated.
- Finally, a landing outcome label was established through the outcome column.



Source Code: IBM-Data-Science-Professional-Certificate/10 Applied Data Science Capstone at main · powellrhys/IBM-Data-Science-Professional-Certificate (github.com)

EDA with Data Visualization

- As part of the exploration process, scatterplots and bar plots 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: [IBM-Data-Science-Professional-Certificate/05_EDA_with_Data_Visualization.ipynb at main · powellrhys/IBM-Data-Science-Professional-Certificate (github.com)]

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 being 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 6000kg.
  - 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 in 2015.
  - Rank of the count of landing outcomes between the date 2010-06-04 and 2017-03-20.

Source Code: IBM-Data-Science-Professional-Certificate/04_EDA.ipynb at main · powellrhys/IBM-Data-Science-Professional-Certificate (github.com)

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 JSC.
  - Marker clusters indicates groups of events in each coordinate, like launches in a launch site.
  - Lines are used to indicate distances between two coordinates.

Source Code: IBM-Data-Science-Professional-Certificate/06_Interactive_Visual_Folium.ipynb at main · powellrhys/IBM-Data-Science-Professional-Certificate (github.com)

Build a Dashboard with Plotly Dash

The following plots and graphs were used to visualize the data:

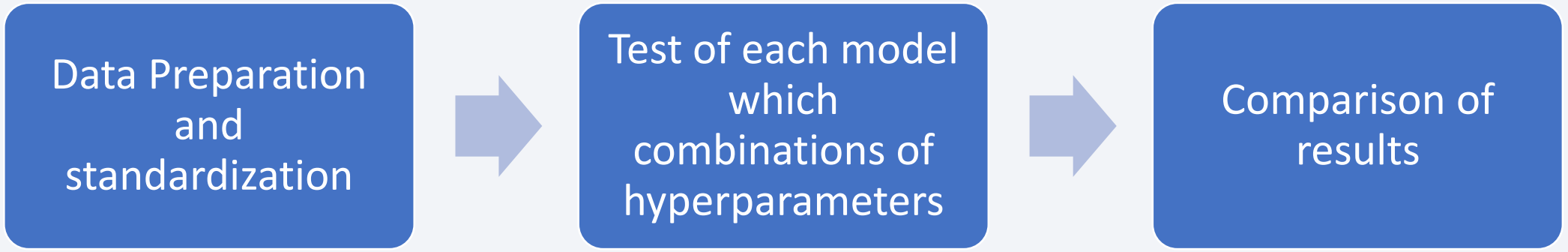
- Percentage of launches by site
- Payload range

This combination allowed quick analysis of relationships between payloads and launch sites, helping to identify where is best to launch according to payloads

Source Code: IBM-Data-Science-Professional-Certificate/07_spacex_dash_app.py at main · powellrhys/IBM-Data-Science-Professional-Certificate (github.com)

Predictive Analysis (Classification)

- Four classification models were compared: logistic regression, support vector machine, decision tree and k nearest neighbour.



Source Code: IBM-Data-Science-Professional-Certificate/08_ML_Prediction.ipynb at main · powellrhys/IBM-Data-Science-Professional-Certificate (github.com)

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 2,929kg.
  - The first successful landing outcome happened in 2015, five years after the first launch.
  - Almost 100% of mission outcomes were successful.
  - Tow booster versions failed at landing in drone ships: F9 v1.1 B1012 & F9 v1.1 B1015.
  - The number of landing outcomes improved over time.

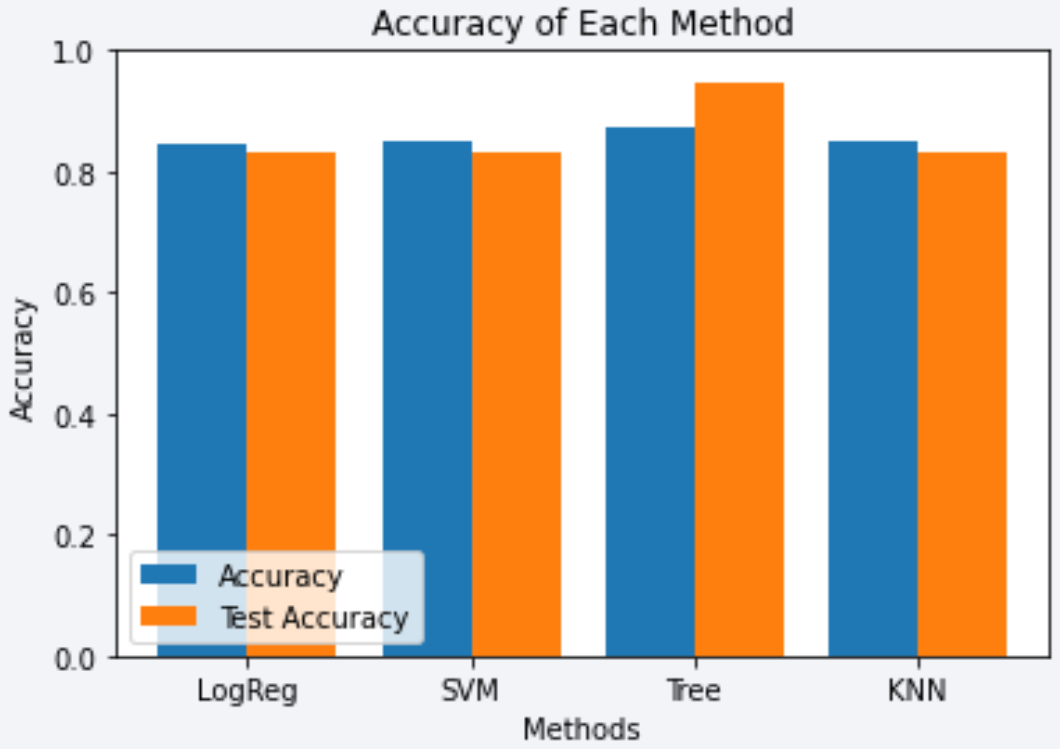
Results

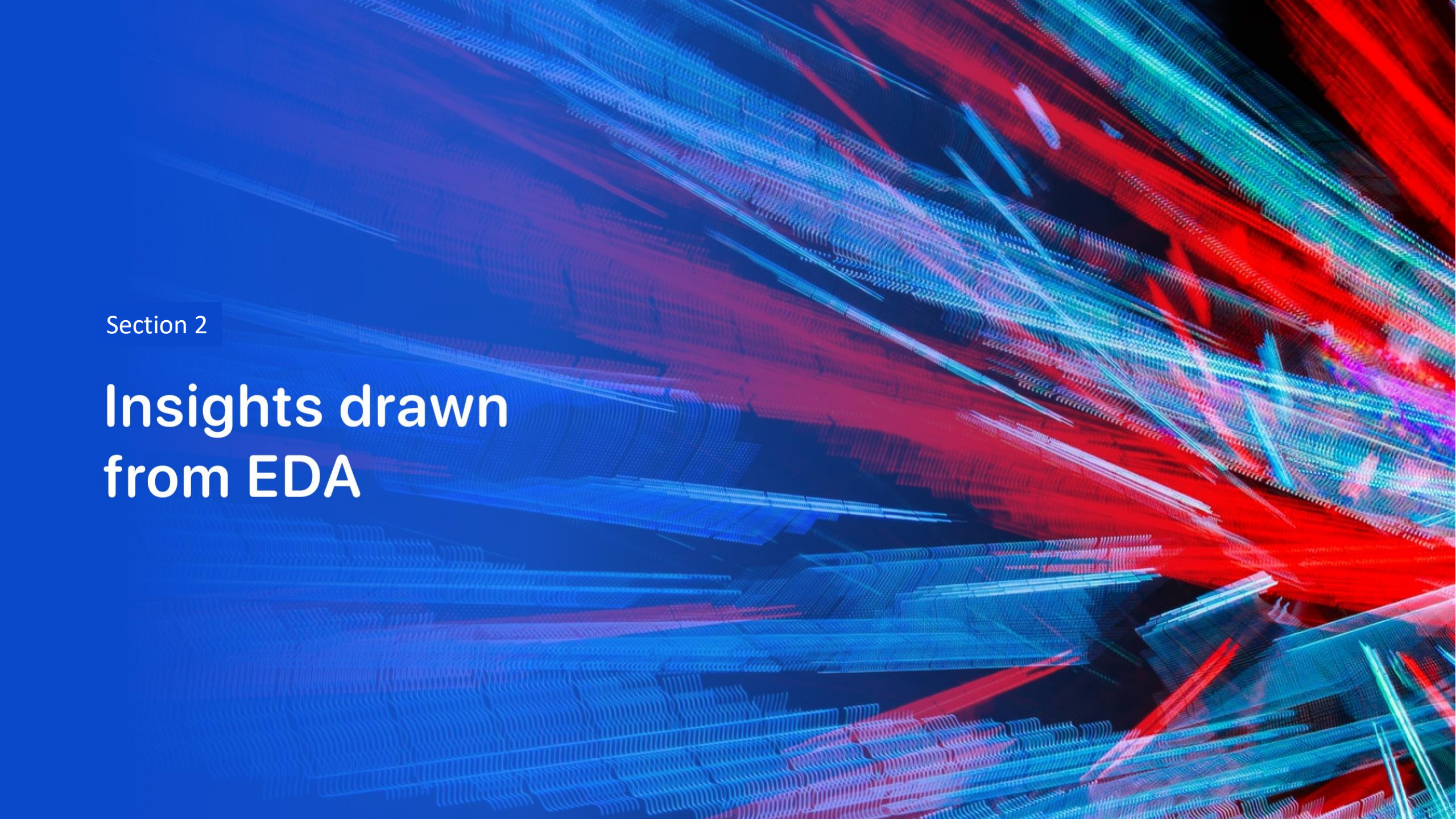
- Using interactive analytics, it was identified that launch sites were positioned near the coast due to safety concerns and due to having good logistical infrastructure in place.

- Most launches happened at the east cost launch sites.

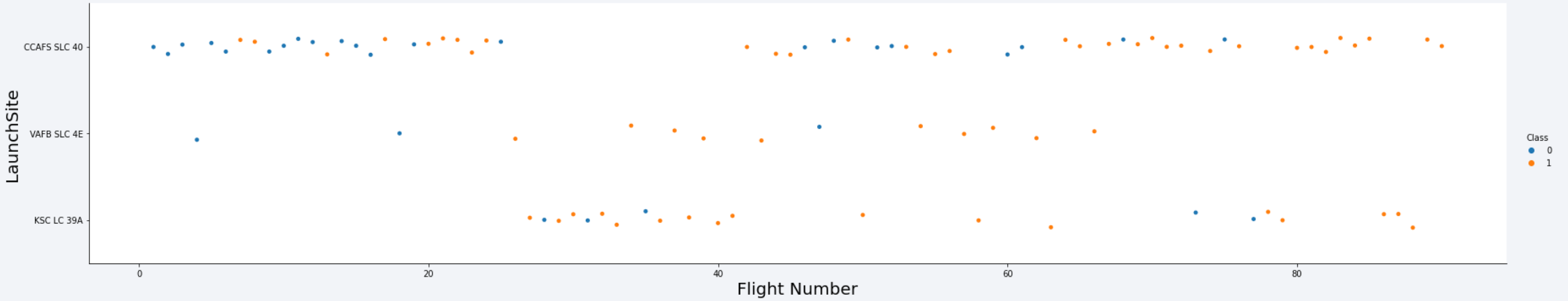
Results

- Predictive analysis showed that decision tree classifier is the best model to predict successful landing, having an accuracy over ##% and a test accuracy over ##%



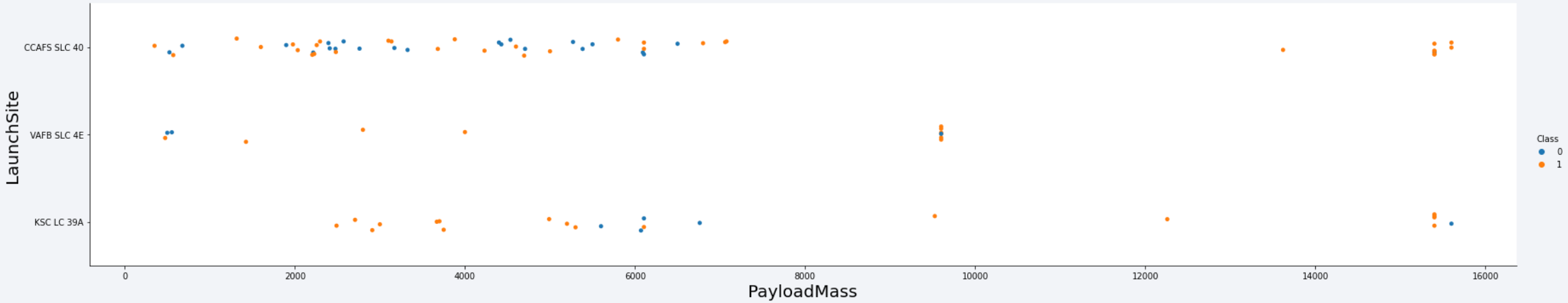

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



- CCAF5 SLC 40 has been the most popular launch destination.
- Most successful launch point is recent years is CCAF5 SLC 40.
- Over time, launches have become more successful.

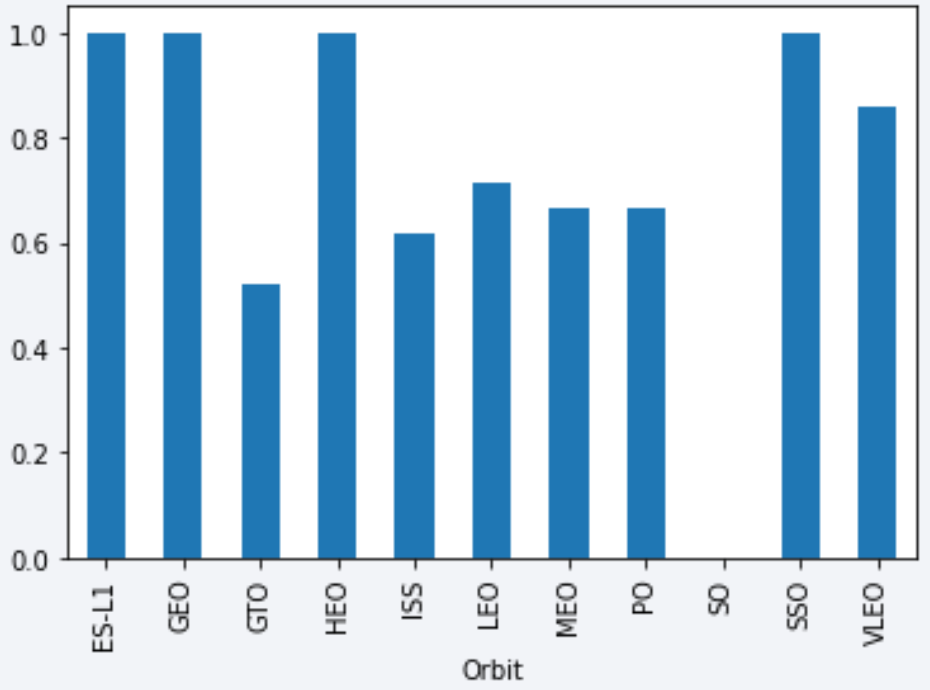
Payload vs. Launch Site



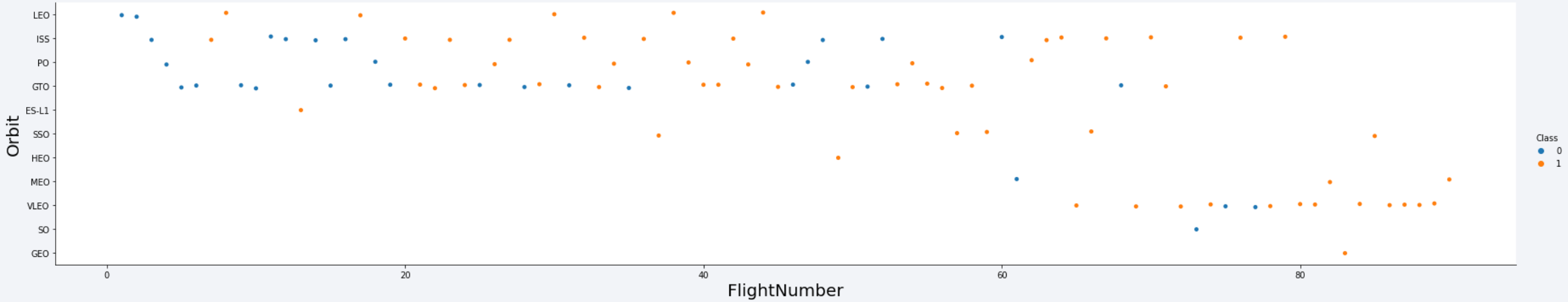
- Payloads over 9000kg have an excellent success rate.
- Launches over 12000kg appear to only be possible from CCAFS SLC 40 KSC LC 39A.

Success Rate vs. Orbit Type

- The highest success rate are for the following orbits:
  - ES-L1
  - GEO
  - HEO
  - SSO
- Followed by:
  - VLEO (above 80%)
  - LFO (above 70%)



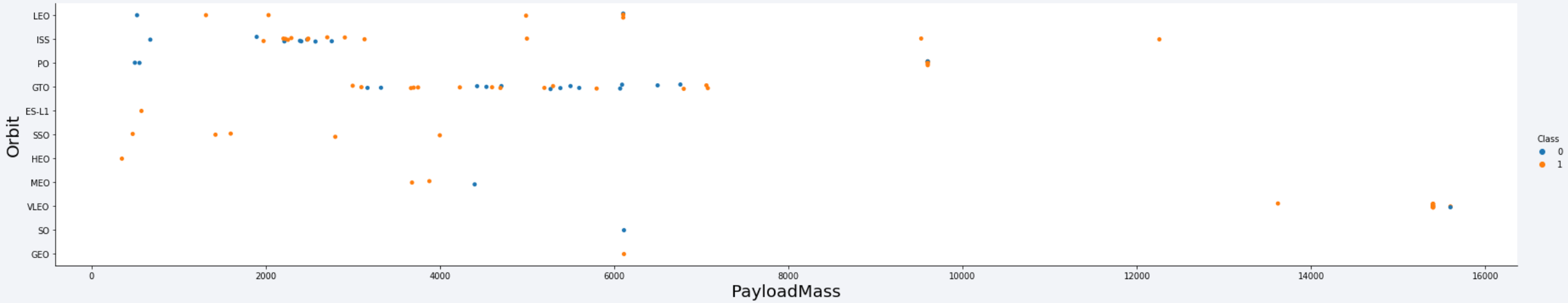
Flight Number vs. Orbit Type



- For most if not all orbits, success rate improved over time.

- VLEO orbit appears to be lucrative in recent years due to its increase in frequency over time.

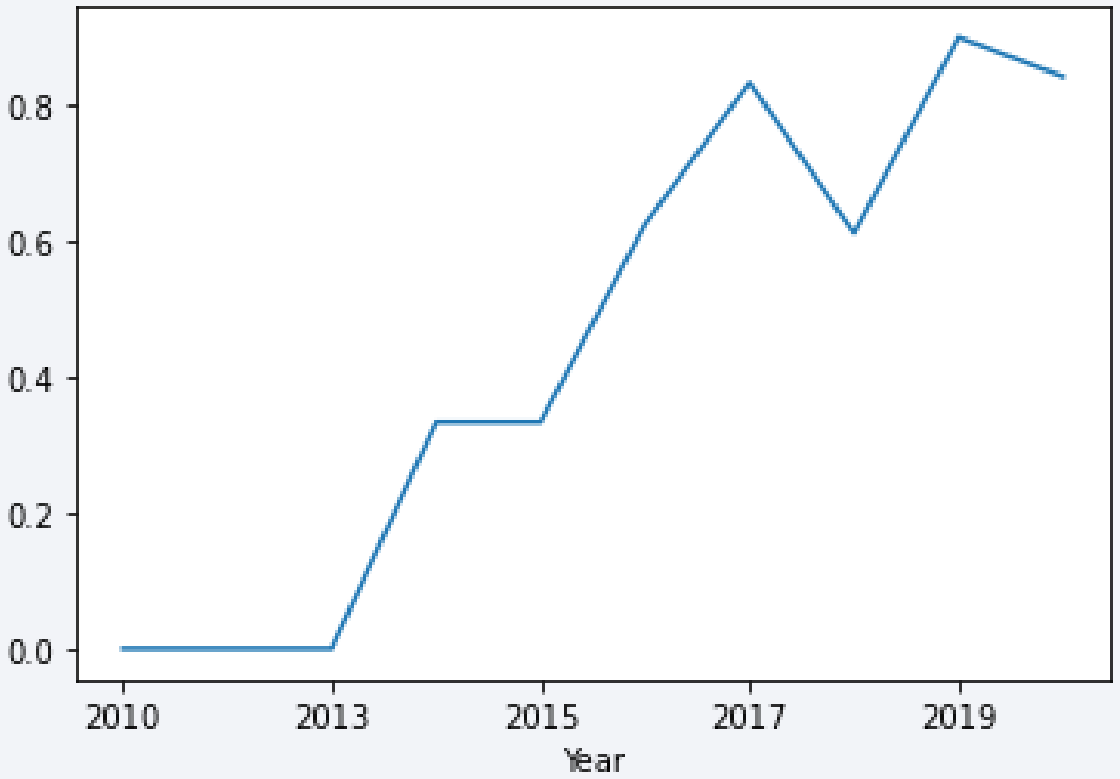
Payload vs. Orbit Type



- There does not appear to be a clear relationship between orbit and payload mass.
- ISS orbit has the greatest range in payload and a good success rate.
- There are a few launches to the orbits of SO and GEO

Launch Success Yearly Trend

- Success rate started increasing in 2013 until 2017.
- First 3 years appeared to be a process of improvement until the technology was refined.
- 2019 saw the highest success rates for launches.



All Launch Site Names

- According to the data, there are four launch sites:
  - CCAFS LC-40
  - CCAFS SLC-40
  - KSC LC-39A
  - VAFB SLC-4E

- This data was obtained by selecting the unique occurrences of “launch site” values from the dataset.

Launch Site Names Begin with 'CCA'

- 5 samples from launch sites including 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	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
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	Success	Failure (parachute)
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	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

Total Payload Mass

- Total payload carried by boosters from NASA:

Total payload (kg)
111268

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

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1

Avg payload (kg)
2928

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

First Successful Ground Landing Date

- First successful landing outcome on ground pad:

First success Ground Pad
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 a successful ladning happened on the ground pad: 2015/12/22.

Successful Drone Ship Landing with Payload between 4000 and 6000

- 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 results.

Total Number of Successful and Failure Mission Outcomes

- Number of successful and failure mission outcomes:

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

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

Boosters Carried Maximum Payload

- Boosters which have carried the maximum payload mass

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

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

- Failed landing outcomes in drone ship, their booster versions, and launch site names for the 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

- Ranking of all landing outcomes between the date 2010-06-04 and 2017-03-20:

- This view of data alerts us that “no attempt” must be considered.

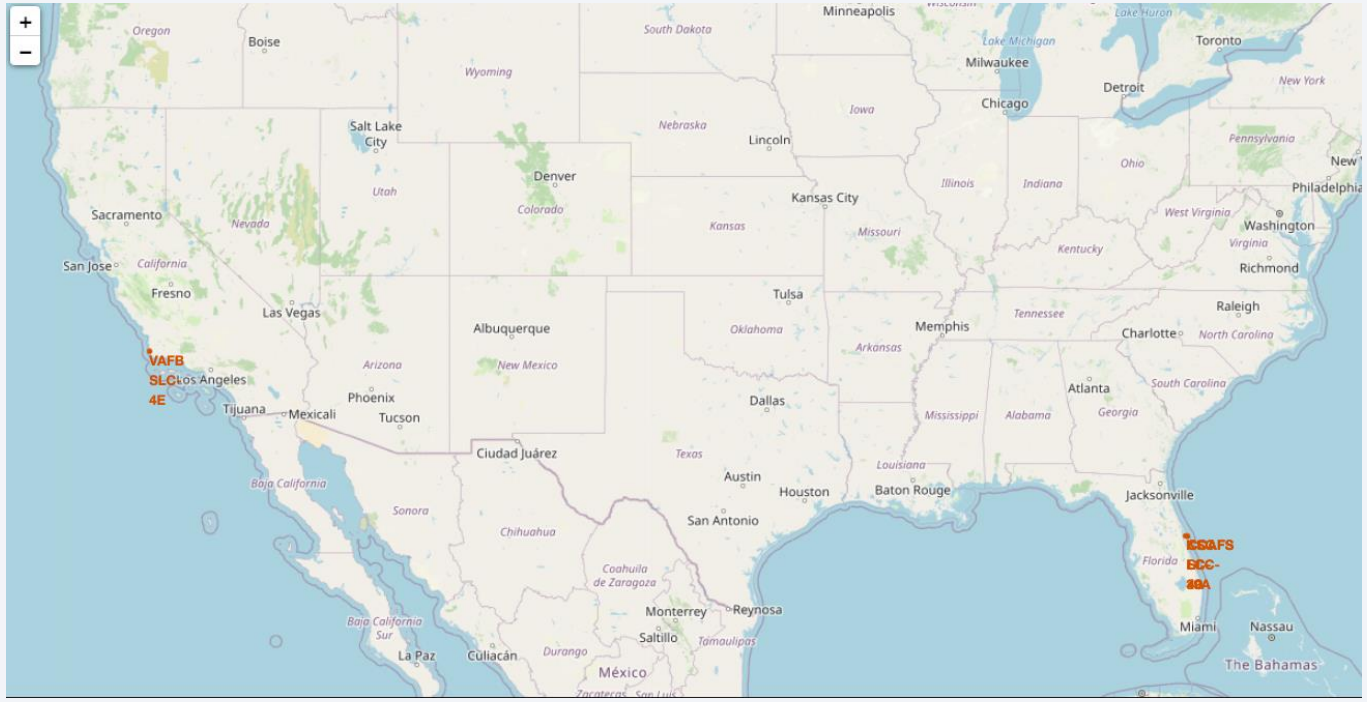
Landing outcome	qty
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

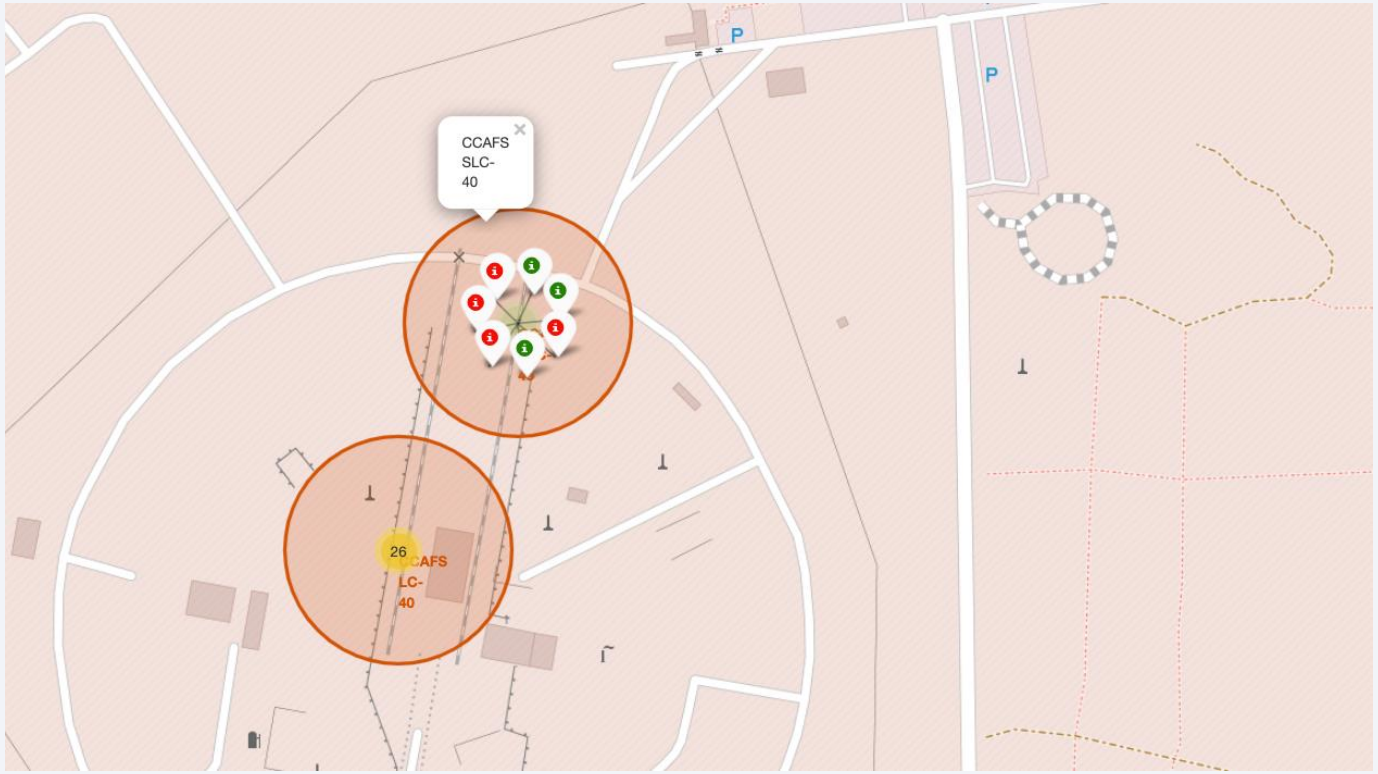
All Launch Sites

- Launch sites are near the sea, probably due to safety concerns, ensuring any failed launches were away from civilians.
- Launch sites nearer the equator to increase initial launch velocity.



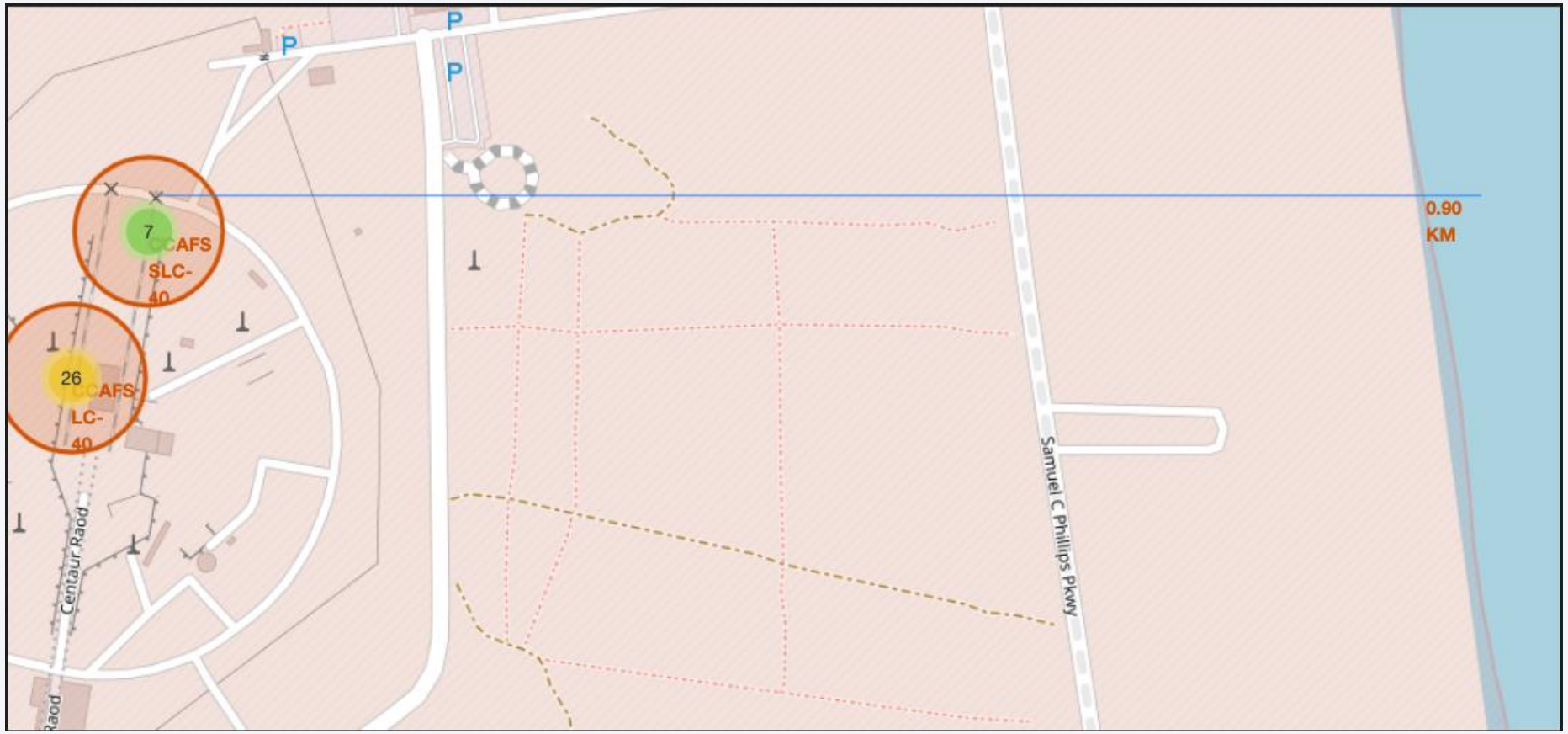
Launch Outcomes by Site

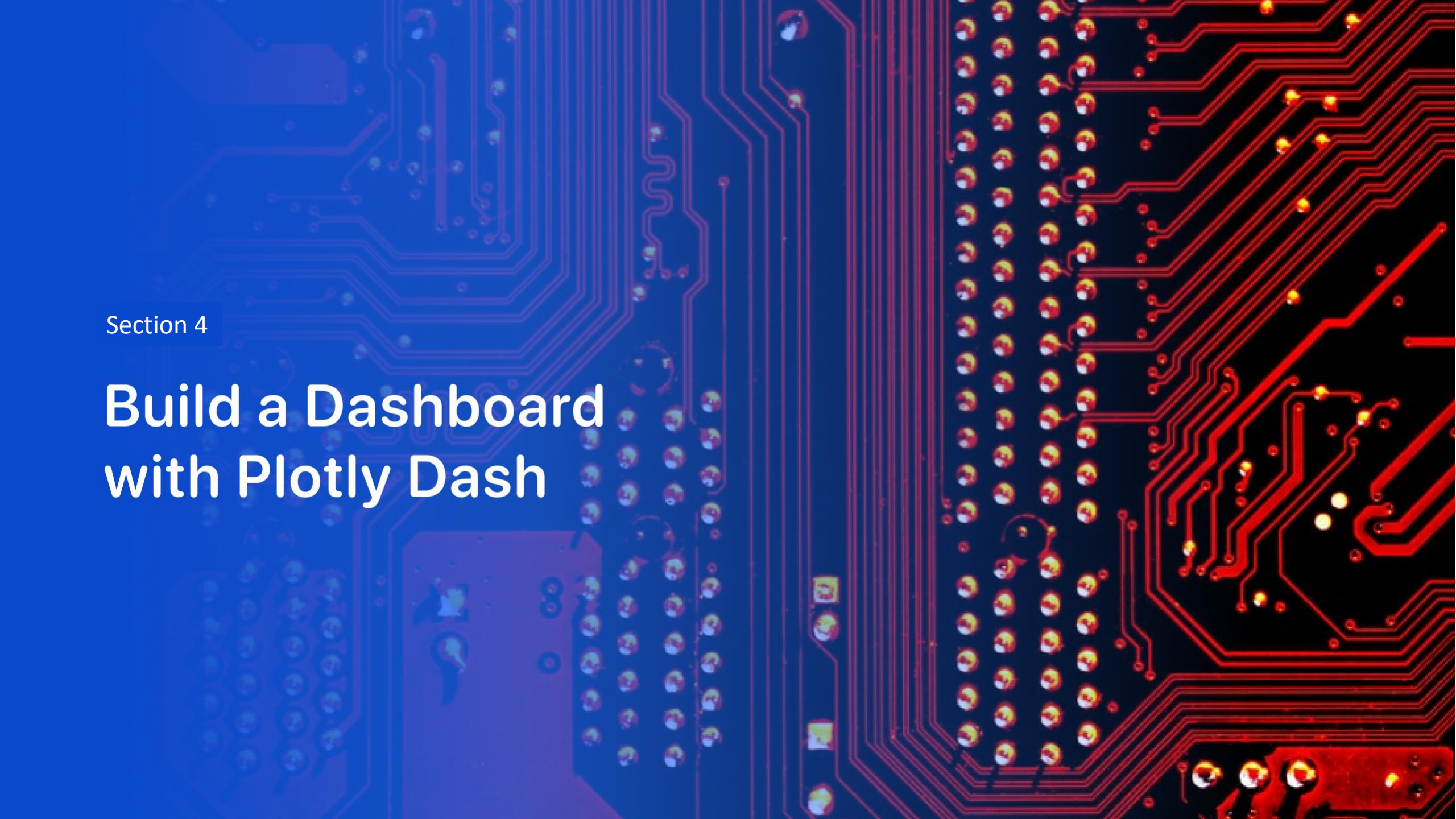
- Example of CCAFS SLC-40 launch site launch outcomes.
- Green markers indicate successful and red one indicate failure.



Logistics and Safety

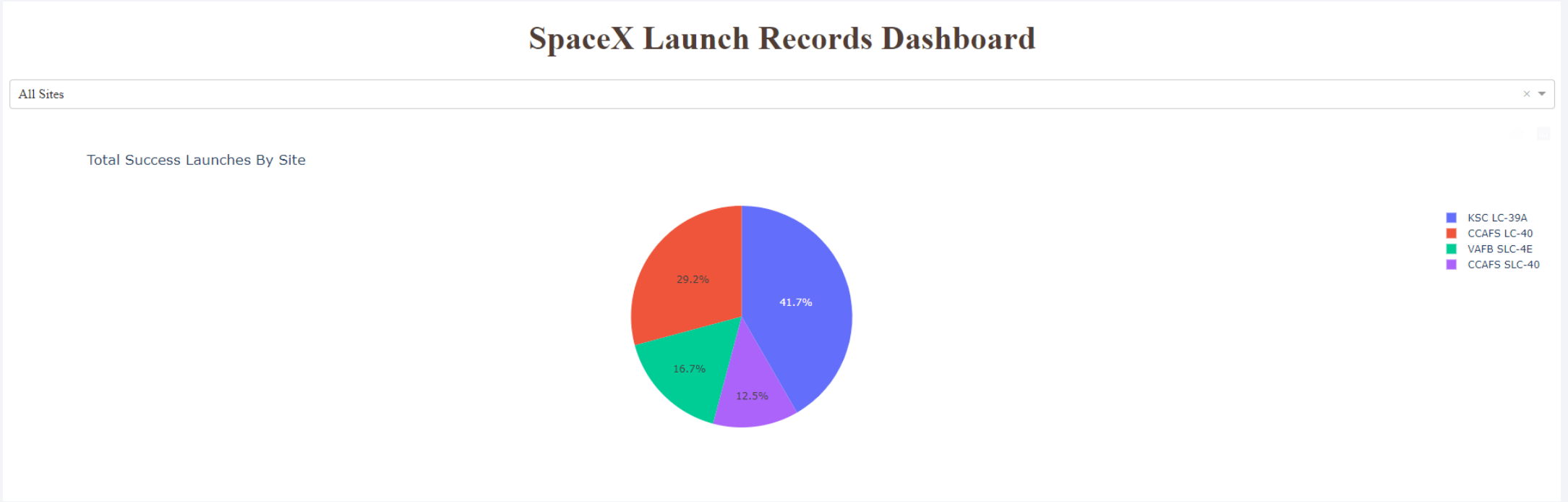
- Launch site CCAFS SLC-40 located 0.9km away from the coast for safety




Section 4
Build a Dashboard with Plotly Dash

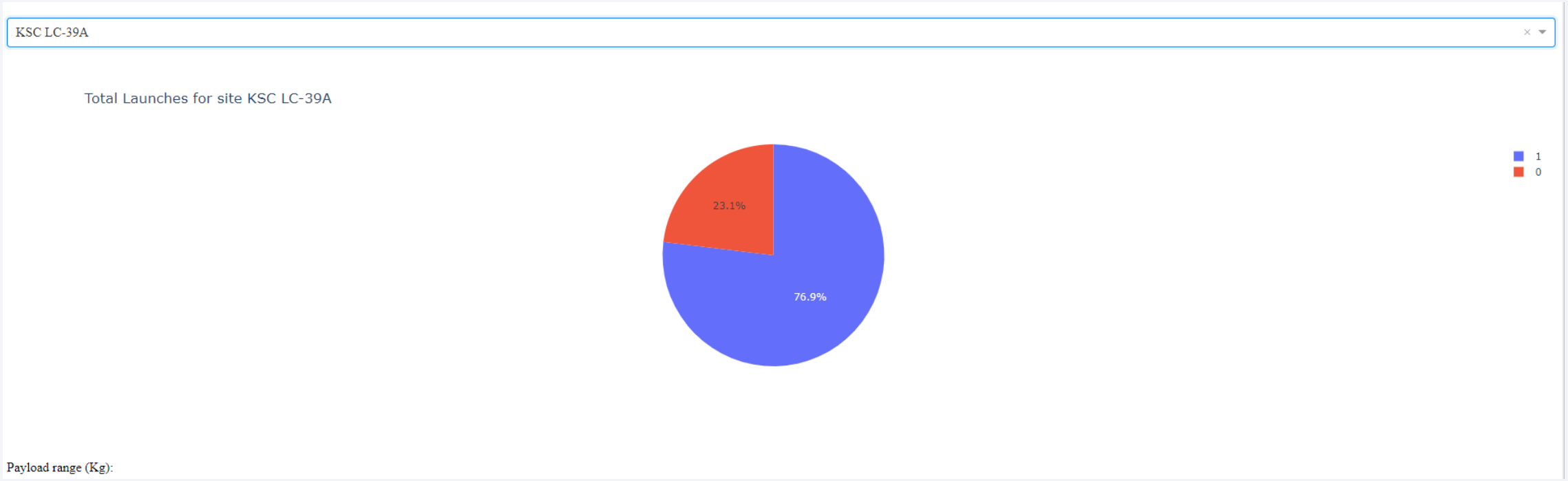
Successful Launches by Site

- The place from where launches are done seems to be a very important factor of success of mission



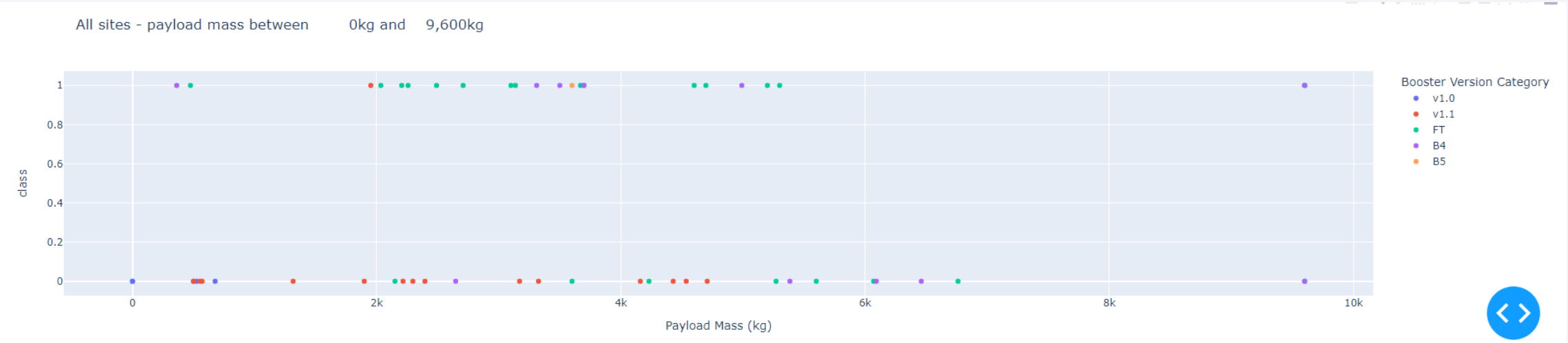
Launch Success Ratio for KSC LC-39A

- 76.9% of launches are successful from the KSC LC-39A site:



Payload vs Launch Outcome

- Payload under 6000kg and FT boosters are the most successful combination.



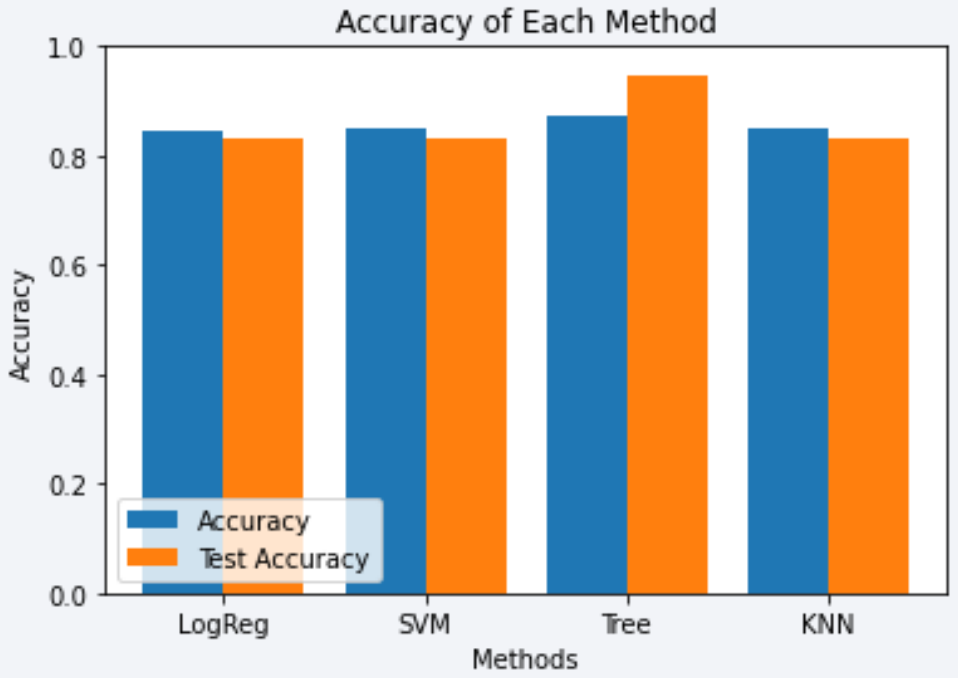
Section 5

Predictive Analysis (Classification)

Classification Accuracy

- Four classification models were tested, and their accuracies are plotted beside.

- The model with the highest classification accuracy is Decision Tree Classifier, which has accuracies over 87%.



Confusion Matrix of Decision Tree Classifier

- 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 analysed, refining conclusions along with process.
- The best launch site is KSC LC-39A.
- Launches above 7000kg are less risky.
- Although most of mission outcomes are successful, successful landing outcomes seem to improve over time.
- Decision tree classifier can be used to predict successful landing and increase profits.

Appendix

- Source code can be found here:
  - IBM-Data-Science-Professional-Certificate/10 Applied Data Science Capstone at main · powellrhys/IBM-Data-Science-Professional-Certificate (github.com)

- Folium plots don't render in GitHub and therefore the notebook will need to be pulled and view locally.

Thank you!