IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Muhammad Ismail Sadaqat
26/03/2024

Outline

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

Executive Summary

- Summary of methodologies
  - API Data Collection
  - Data Collection with Web Scaping
  - Exploratory Data Wrangling
  - Exploratory Data Analysis with SQL
  - Data Analysis with Data Visualization
  - Interactive Visual Analytics with Folium lab
  - Machine Learning Prediction
- Summary of all results
  - Data collection and data analysis results
  - Visualize the data and extract meaningful patterns using data visualization
  - Machine learning model predictions

Introduction

- SpaceX advertises Falcon 9 rocket launches on its website, with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage. If we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

- In a role of a data scientist for a new rocket company, Space Y, aiming to compete with SpaceX. The task involves determining launch prices, gathering SpaceX information, and predicting the reuse of the first stage using machine learning and public data,

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - 2 sources used for Data collection:
  - SpaceX REST API (https://api.spacexdata.com/v4/rockets/)
  - Web scraping from a wikipedia page (https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches)
- Perform data wrangling
  - convert outcomes into training labels with 1 means the booster successfully landed 0 means it was unsuccessful.

Methodology

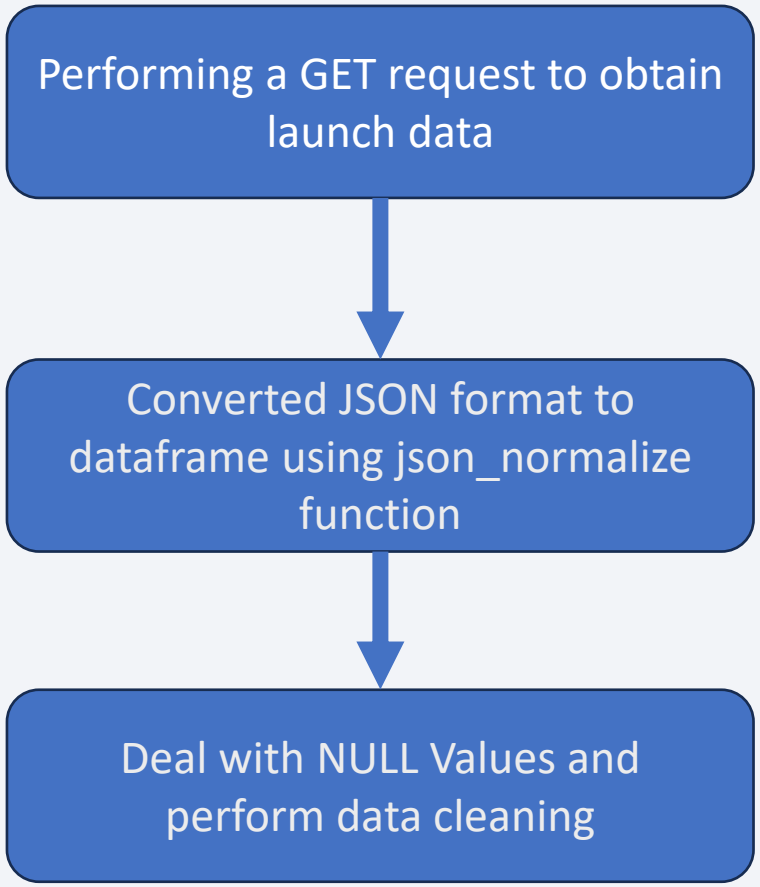
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Collected data was standardized, Split into training data and test data. Then found the best Hyperparameter for SVM, Classification Trees and Logistic Regression and found which method performs the best

Data Collection

- Data was collected using 2 sources: SpaceX REST API (https://api.spacexdata.com/v4/rockets/) and wikipedia page (https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches)

Data Collection – SpaceX API

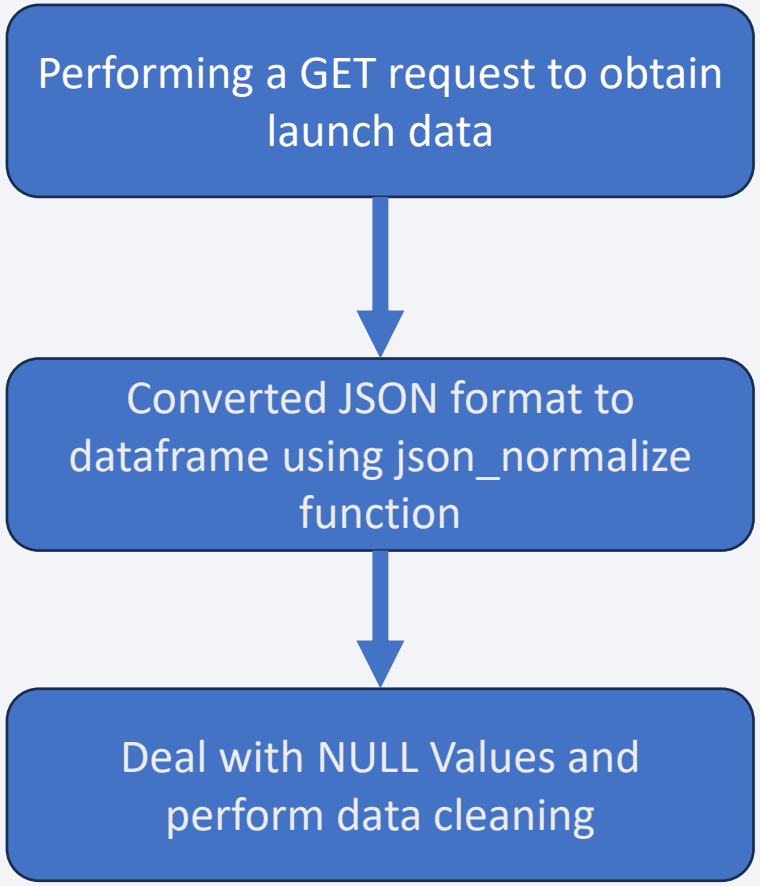
- Used SpaceX REST API to gather launch data, including rocket details, payload information, launch and landing specifications, and outcomes.



https://github.com/Ismail-Sadaqat/Applied-Data-Science-Capstone/blob/2bd2e2d33fc8616250bcb4548c1fb72a170428d7/Data%20Collection%20API%20Lab.ipynb

Data Collection - Scraping

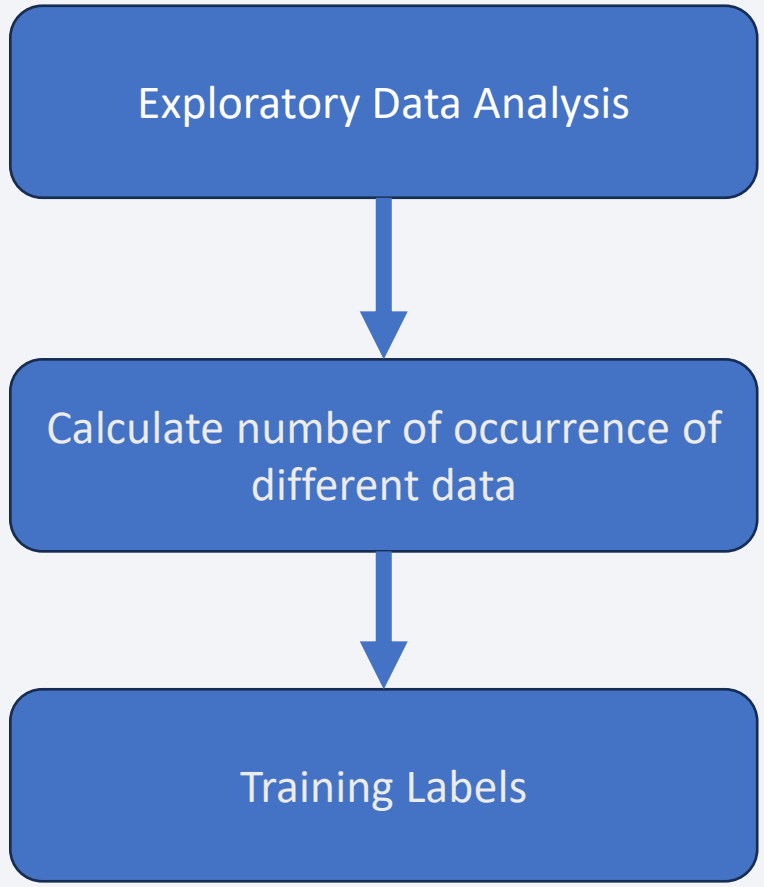
- Used Wikipedia pages for Falcon 9 launch records.
- Flowcharts present web scraping process



https://github.com/Ismail-Sadaqat/Applied-Data-Science-Capstone/blob/2bd2e2d33fc8616250bcb4548c1fb72a170428d7/Data%20Collection%20with%20Web%20Scraping.ipynb

Data Wrangling

- Perform exploratory data analysis
  - Calculate the number of launches on each site
  - Calculate the number and occurrence of each orbit
  - Calculate the number and occurrence of mission outcome per orbit type
- Convert Landing outcomes into training labels with 1 means the booster successfully landed 0 means it was unsuccessful.



https://github.com/Ismail-Sadaqat/Applied-Data-Science-Capstone/blob/2bd2e2d33fc8616250bcb4548c1fb72a170428d7/Data%20Wrangling.ipynb

EDA with Data Visualization

- Used scatter graph to Visualize and find relationship between:
  - Flight Number and Payload Mass
  - Flight Number and Launch Site
  - Payload and Launch Site
  - Flight Number and Orbit type
  - Payload and Orbit type
- Created a bar chart to visualize the relationship between success rate of each orbit type and a line chart to visualize the launch success yearly trend
- Used features engineering for success prediction in the future module.

https://github.com/Ismail-Sadaqat/Applied-Data-Science-Capstone/blob/2bd2e2d33fc8616250bcb4548c1fb72a170428d7/EDA%20with%20Visualization.ipynb

EDA with SQL

- Performed the following 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. Use a subquery
  - 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 (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order.

https://github.com/Ismail-Sadaqat/Applied-Data-Science-Capstone/blob/2bd2e2d33fc861625Obcb4548c1fb72a170428d7/EDA%20with%20SQL.ipynb

Build an Interactive Map with Folium

- Markers, circles, lines created and added to a folium map:
  - Mark all launch sites on a map
  - Mark the success/failed launches for each site on the map
  - Calculate the distances between a launch site to its proximities

https://github.com/Ismail-Sadaqat/Applied-Data-Science-Capstone/blob/2bd2e2d33fc8616250bcb4548c1fb72a170428d7/Visual%20Analytics%20with%20Folium.ipynb

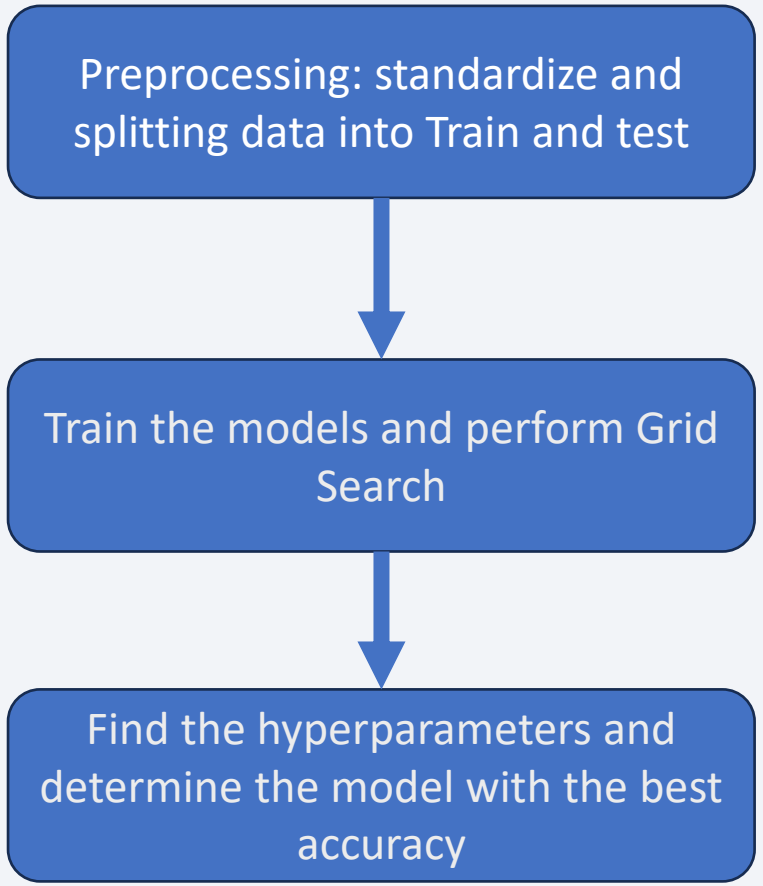
Build a Dashboard with Plotly Dash

- Plots, graphs and interactions added to a dashboard:
  - Pie charts - Percentage of launches by certain launch site
  - Scatter graphs – relationship between Outcome and Payload Mass

https://github.com/Ismail-Sadaqat/Applied-Data-Science-Capstone/blob/2bd2e2d33fc8616250bcb4548c1fb72a170428d7/spacex_dash_app.py

Predictive Analysis (Classification)

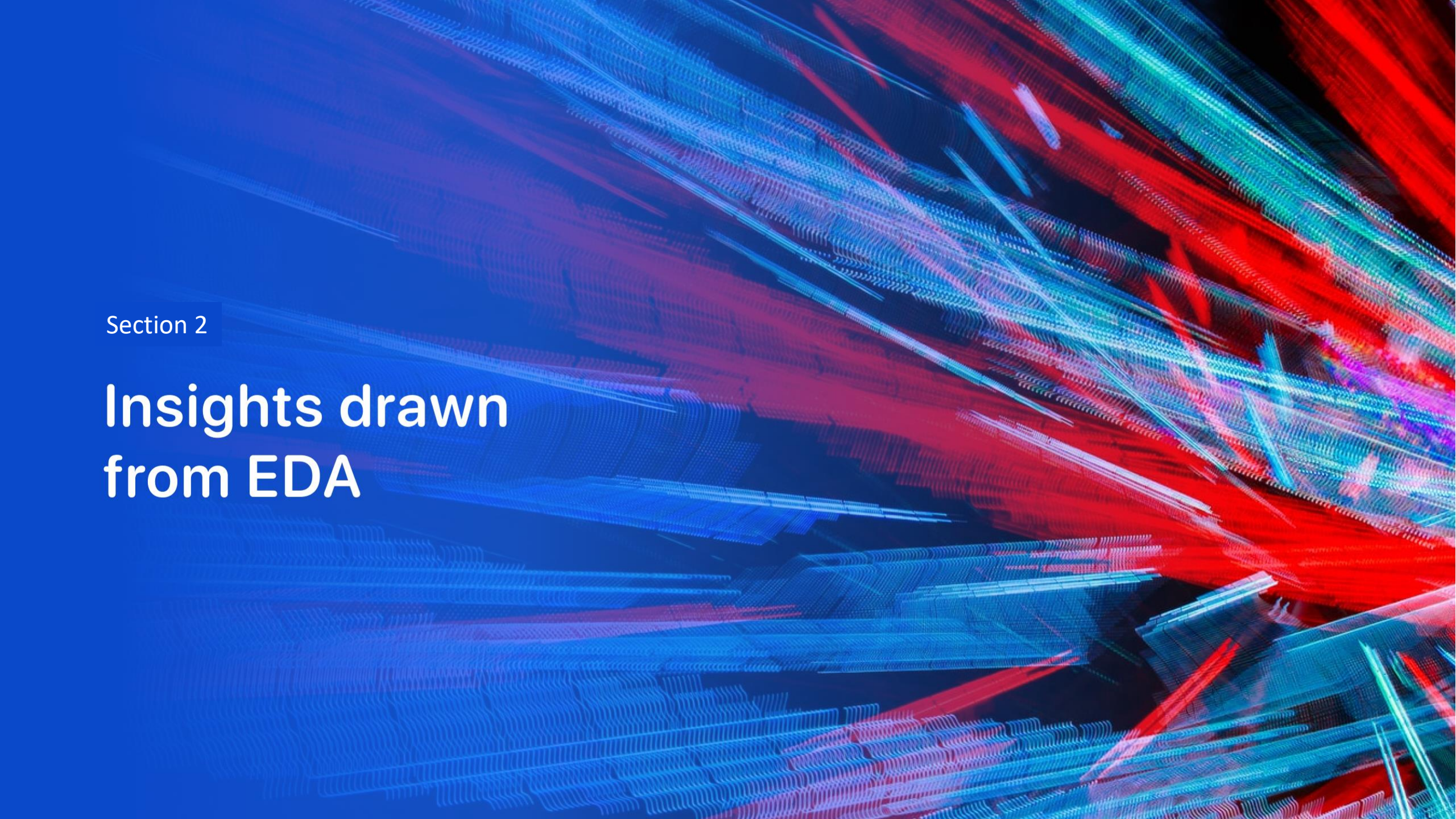
- Four classification models prepared and compared:
  - Logistic Regression
  - Support Vector machines
  - Decision Tree Classifier
  - K-nearest neighbors
- The flowchart presents the model development process.



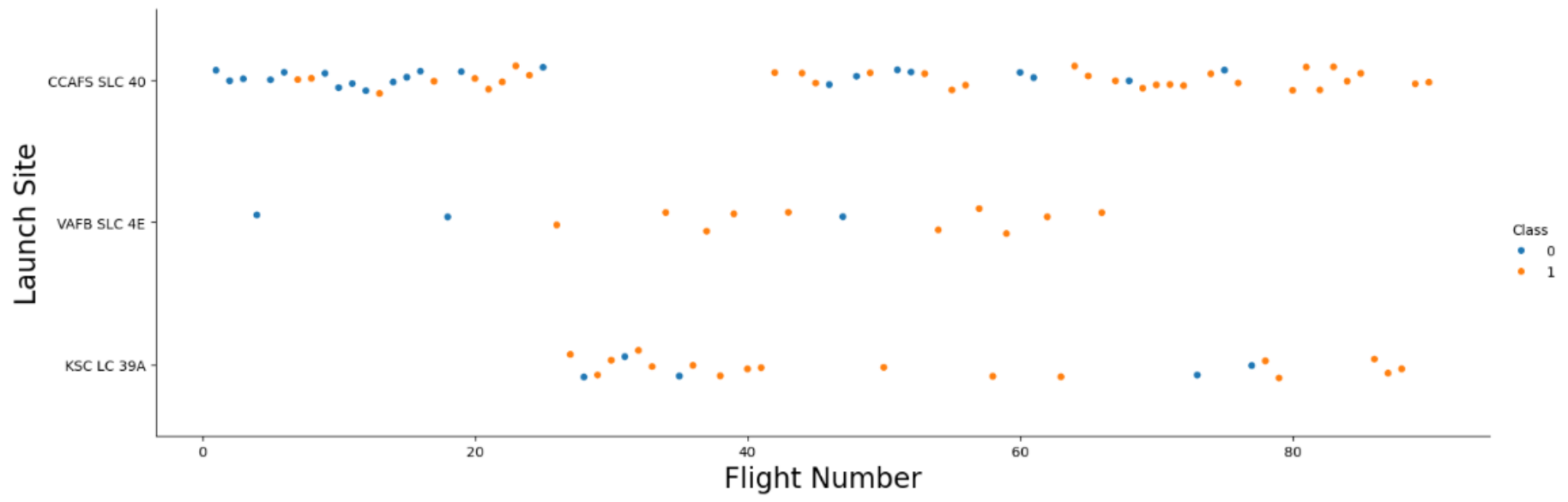
https://github.com/Ismail-Sadaqat/Applied-Data-Science-Capstone/blob/2bd2e2d33fc8616250bcb4548c1fb72a170428d7/Machine%20Learning%20Prediction.ipynb

Results

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

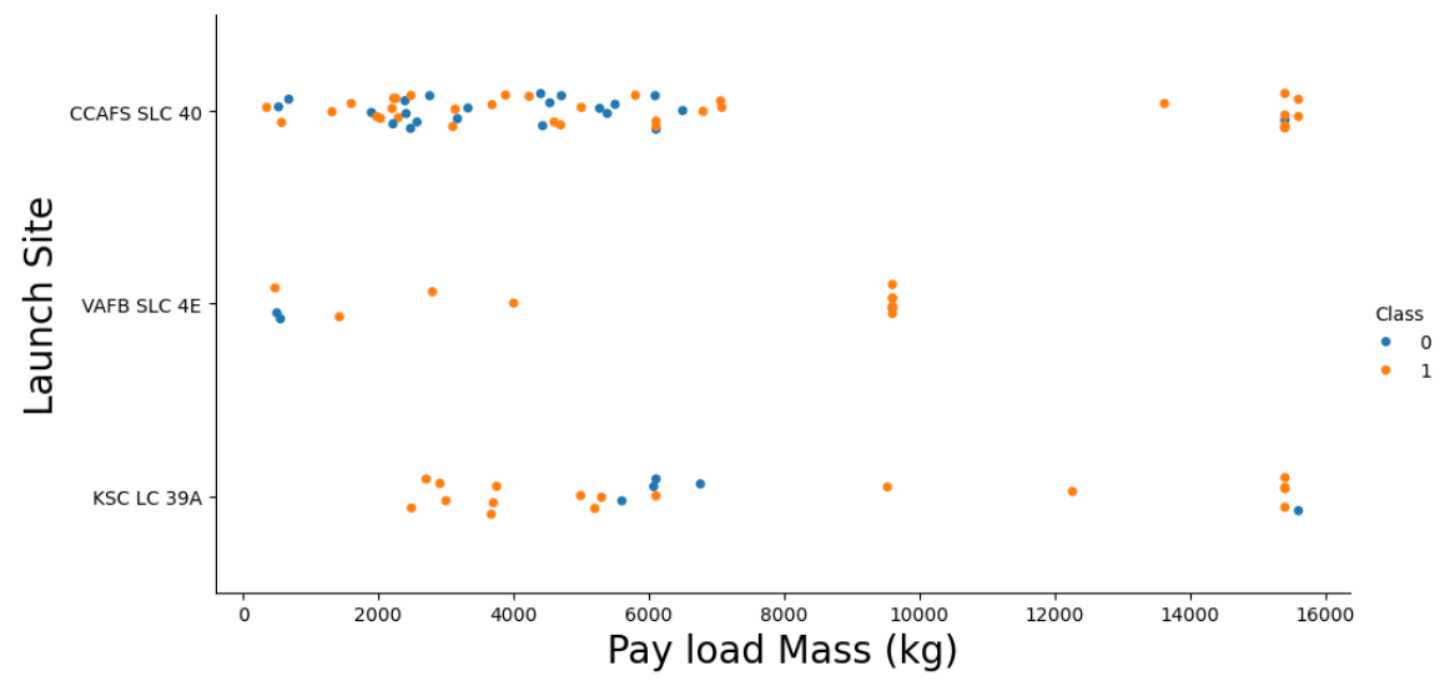

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



- The Scatter graph shows that the most successful Launch site is the CCAFS SLC 40
- The Scatter graph also shows that the larger the flight number the greater the success rate.

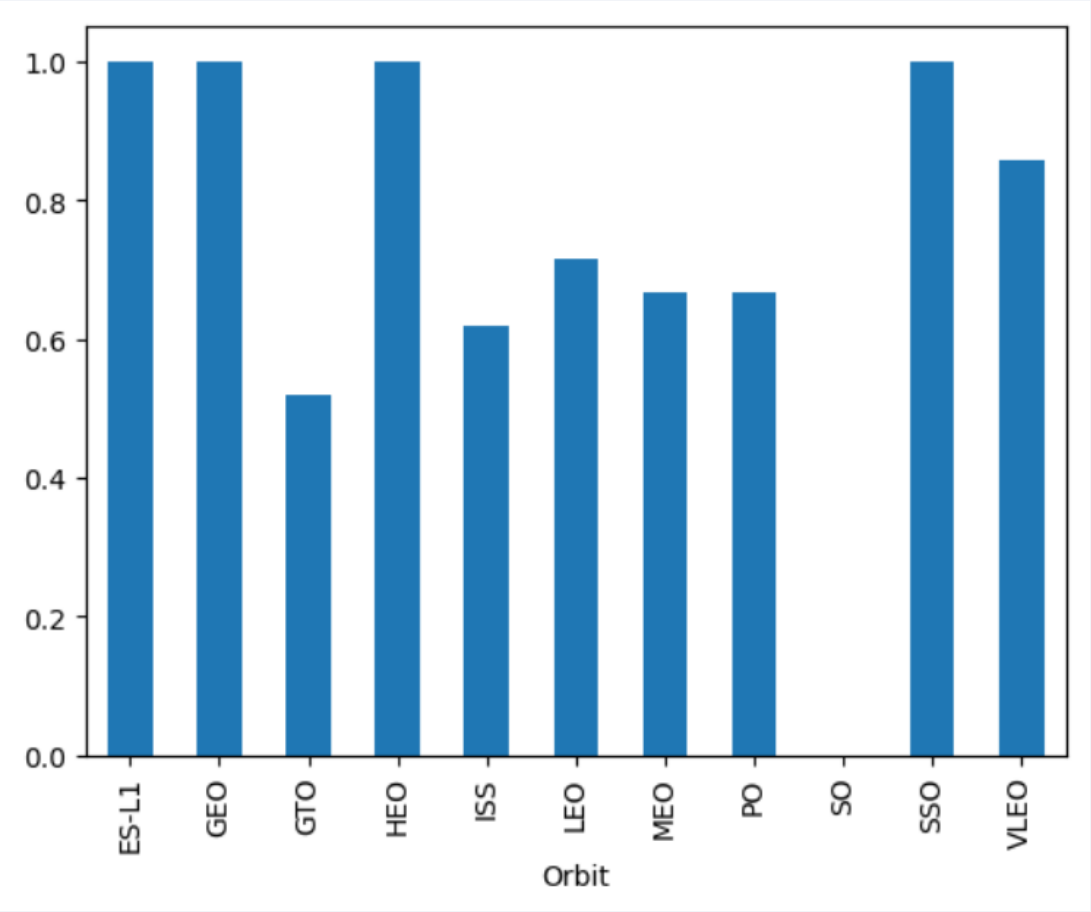
Payload vs. Launch Site



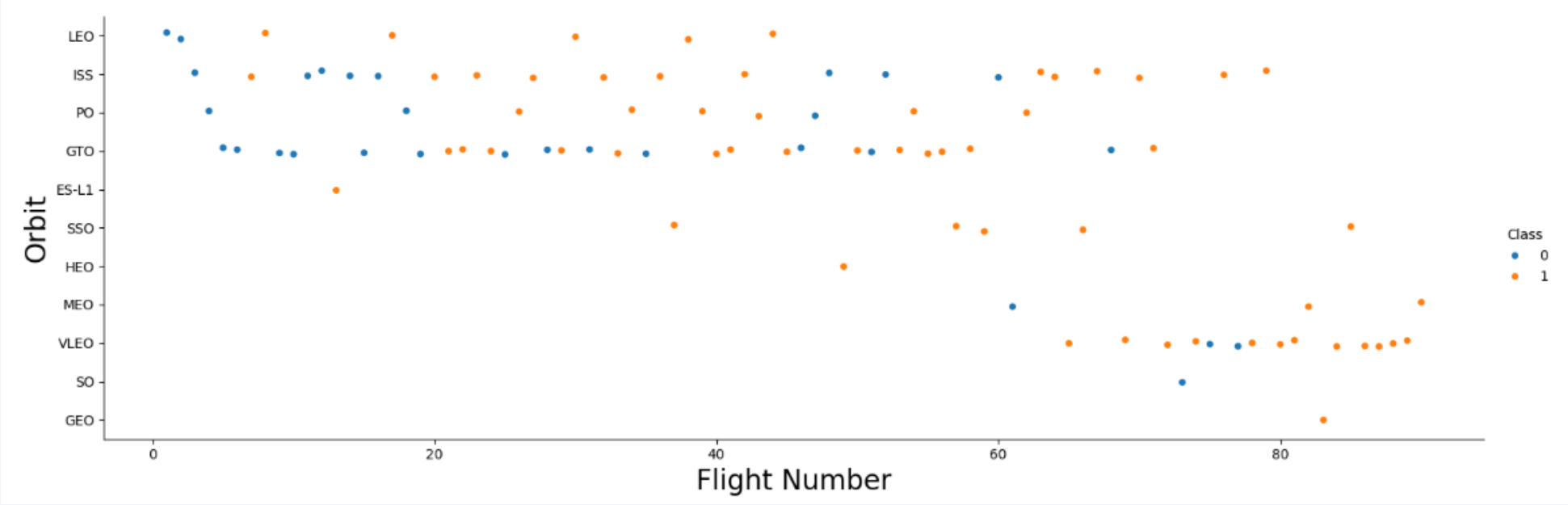
- Payload above 8000 kg the probability of success rate increases
- The scatter graph suggests that payload above 10000kgs is only possible in CCAFS SLC 40 and KSC LC 39A

Success Rate vs. Orbit Type

- The graph shows the probability of success full landing outcome depending on the orbit type
- Orbits with 100% success rate:
  - ES-L1
  - GWO
  - HEO
  - SSO

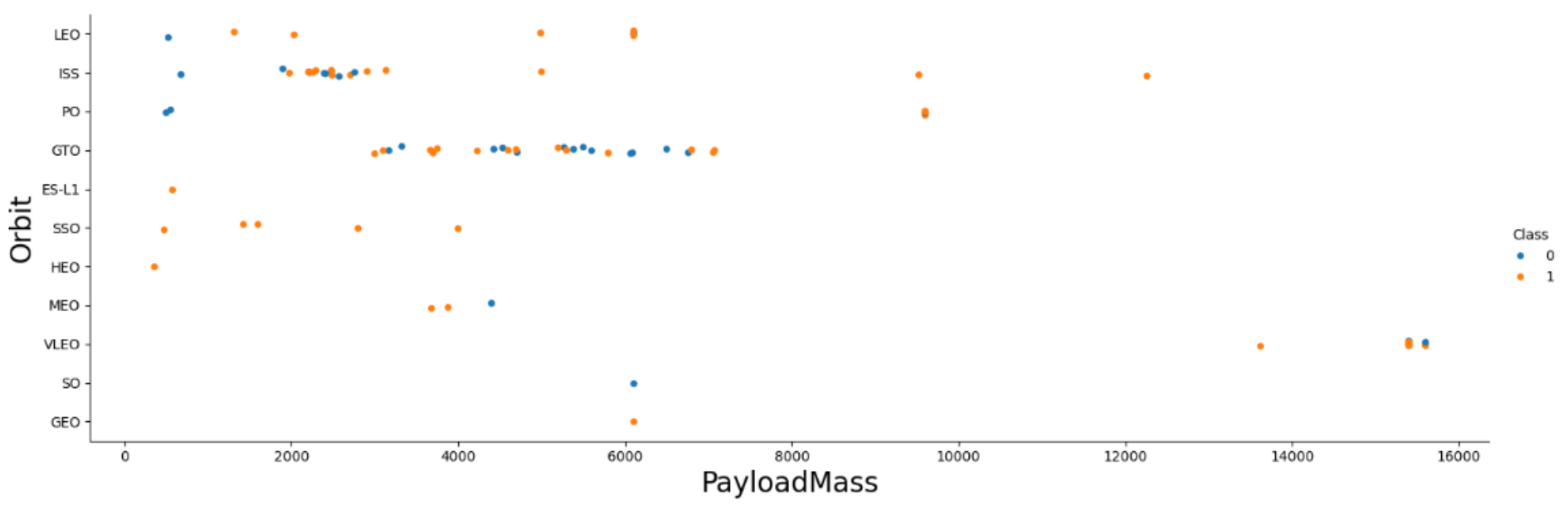


Flight Number vs. Orbit Type



- The scatter graph suggest that most of the Flights were done in LEO, ISS, PO and GTO
- Success rate is also increased with the increase in Flight Number

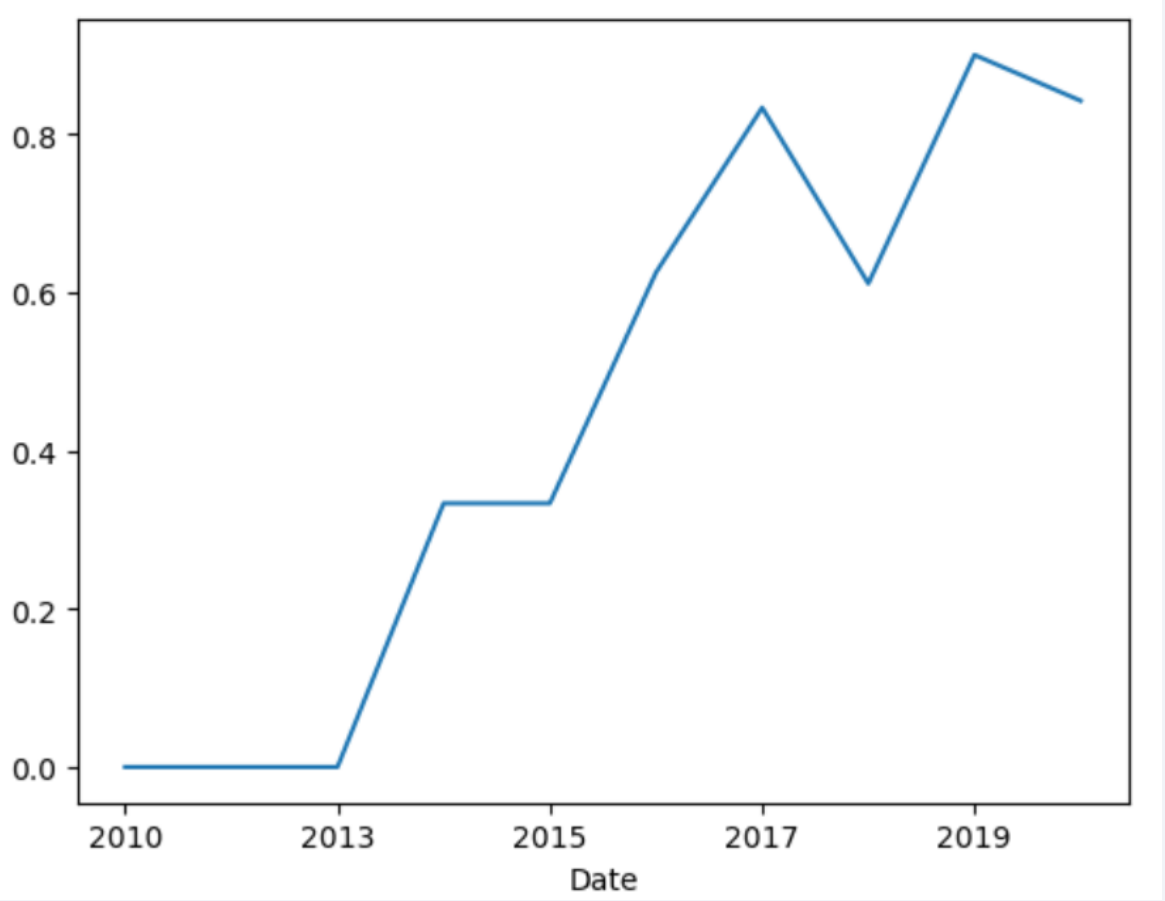
Payload vs. Orbit Type



- There seem to be no relationship between payload and orbit Type

Launch Success Yearly Trend

- The line chart shows a clear increase in yearly average success rate starting from 2013 till 2017.
- After 2017 there has been few decrease in the success rate, but overall, as time passes success rate increases.



All Launch Site Names

- There are four unique launch sites:

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

- They were obtained by using the DISTINCT keyword

Launch Site Names Begin with 'CCA'

- Five 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	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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
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

Total Payload Mass

- Total payload carried by boosters from NASA:

Total_Payload
111268

- This was obtained by summing all payload mass whose codes contain ‘CRS’

Average Payload Mass by F9 v1.1

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

Average_Payload_Mass
2928.4

- This was obtained by averaging all payload mass where booster version is F9 v1.1

First Successful Ground Landing Date

- Date of the first successful landing outcome on ground pad:

min(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

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

Total Number of Successful and Failure Mission Outcomes

- Total number of successful and failure mission outcomes:

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

Boosters Carried Maximum Payload

- Names of the booster which have carried the maximum payload mass:

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

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

Months	Landing_Outcome	Booster_Version	Launch_Site
01	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

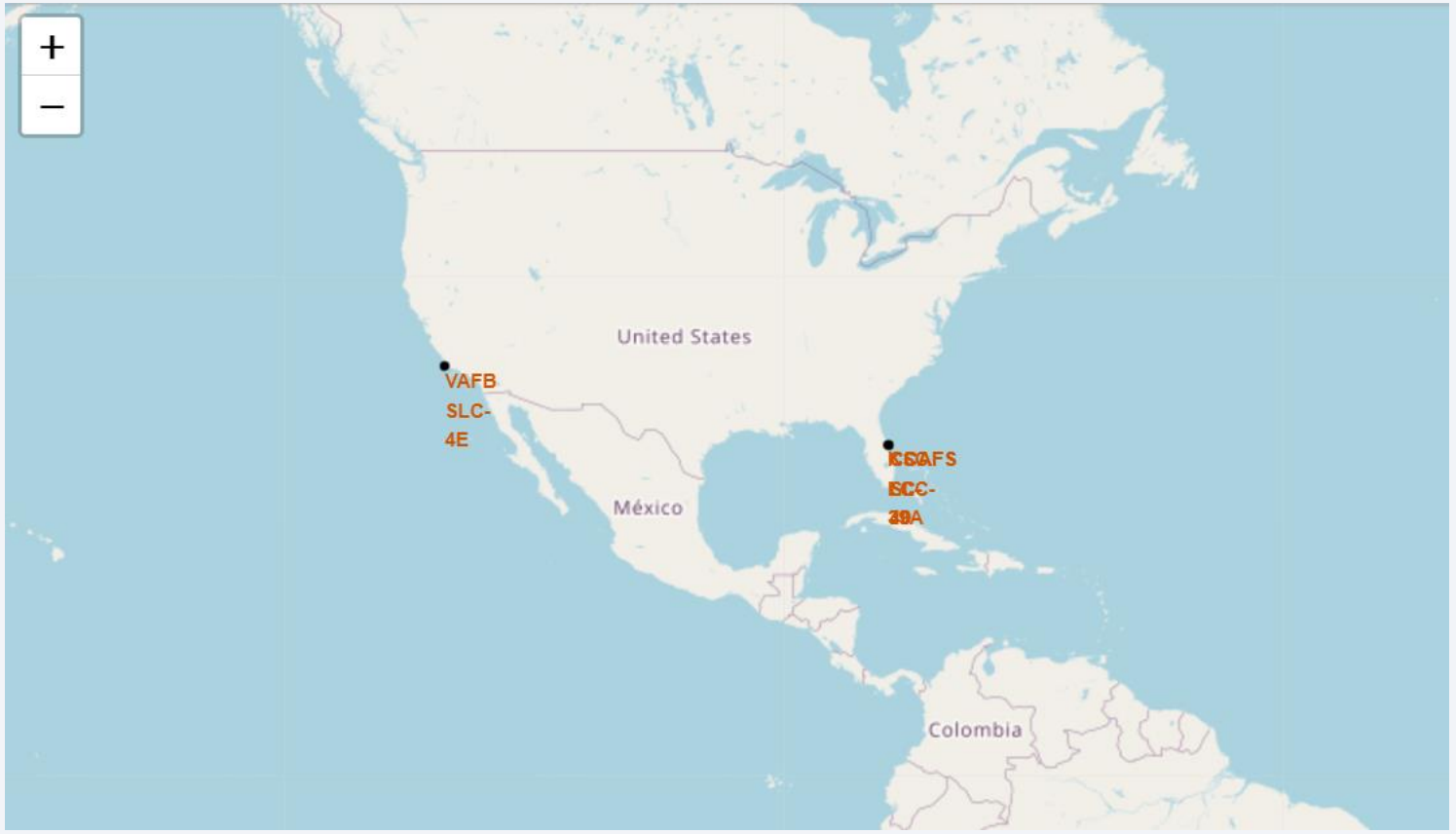
- Count of landing outcomes between the date 2010-06-04 and 2017-03-20, in ranked in descending order

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

Section 3

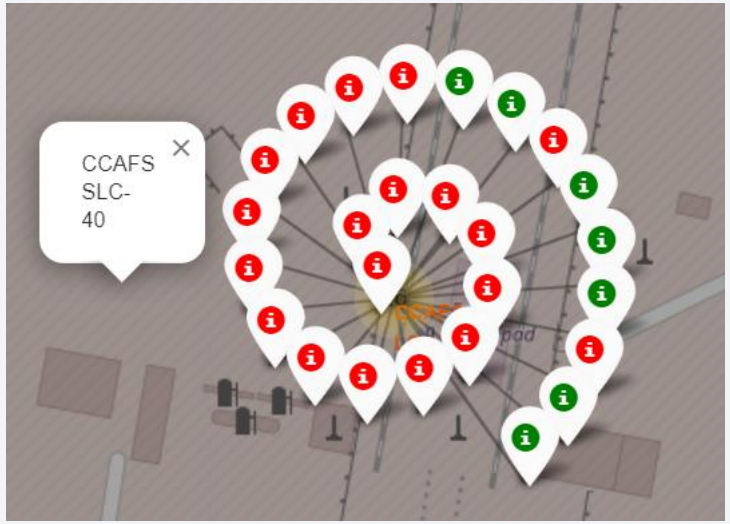
Launch Sites Proximities Analysis

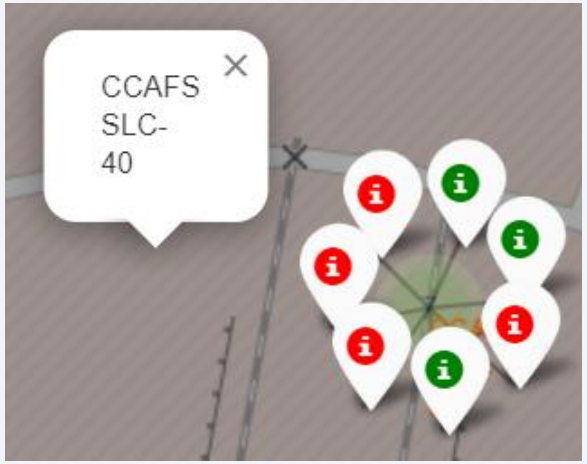
Launch Sites



- From the map we can see that all launch sites are in proximity to the Equator line
- We can also see all launch sites are in very close proximity to the coast

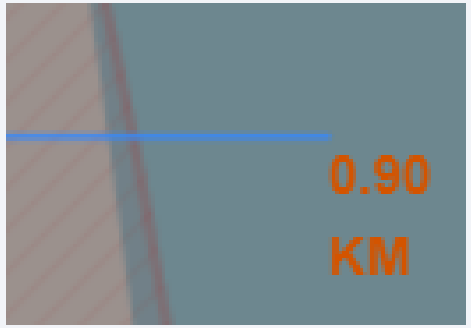
launch outcomes



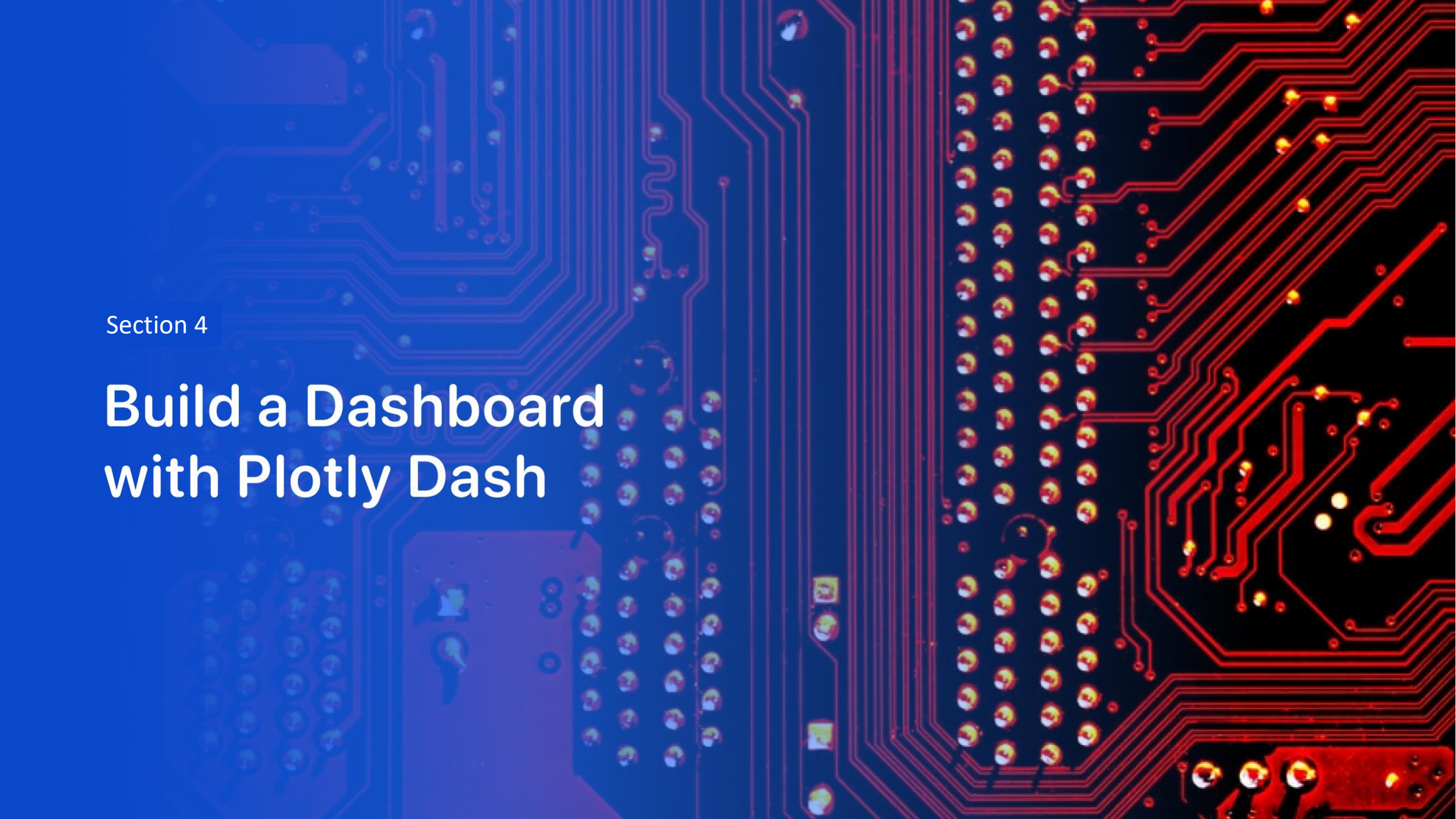


- The color-labeled markers identify which launch sites have relatively high success rates.
- Green color-labeled markers indicates successful and red color-labeled markers indicates falure

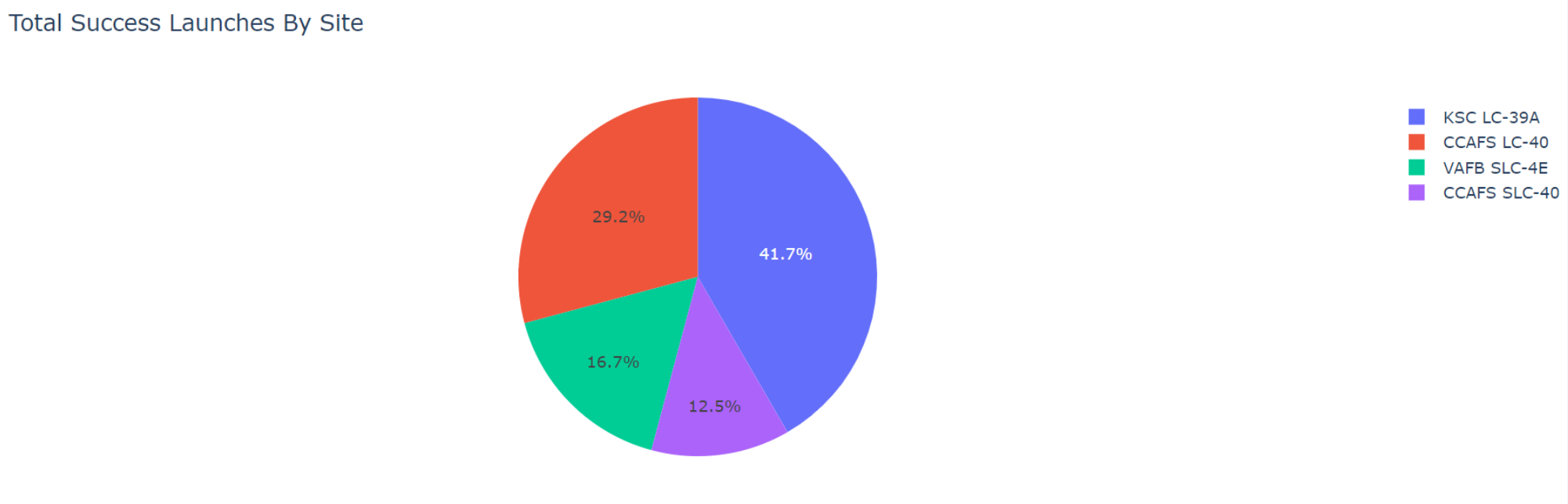
launch site distance calculated from landmarks



- The launch sites are in close proximity to railways, highways and coastline?
- The launch sites keeps certain distance away from cities

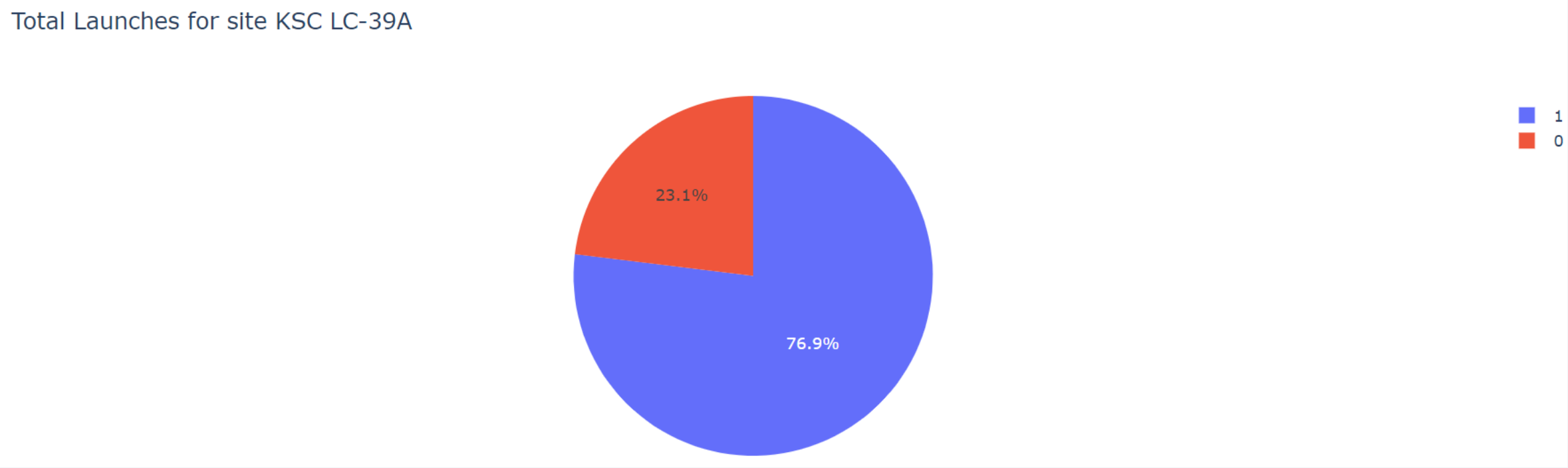

Section 4
Build a Dashboard with Plotly Dash

launch success for all sites



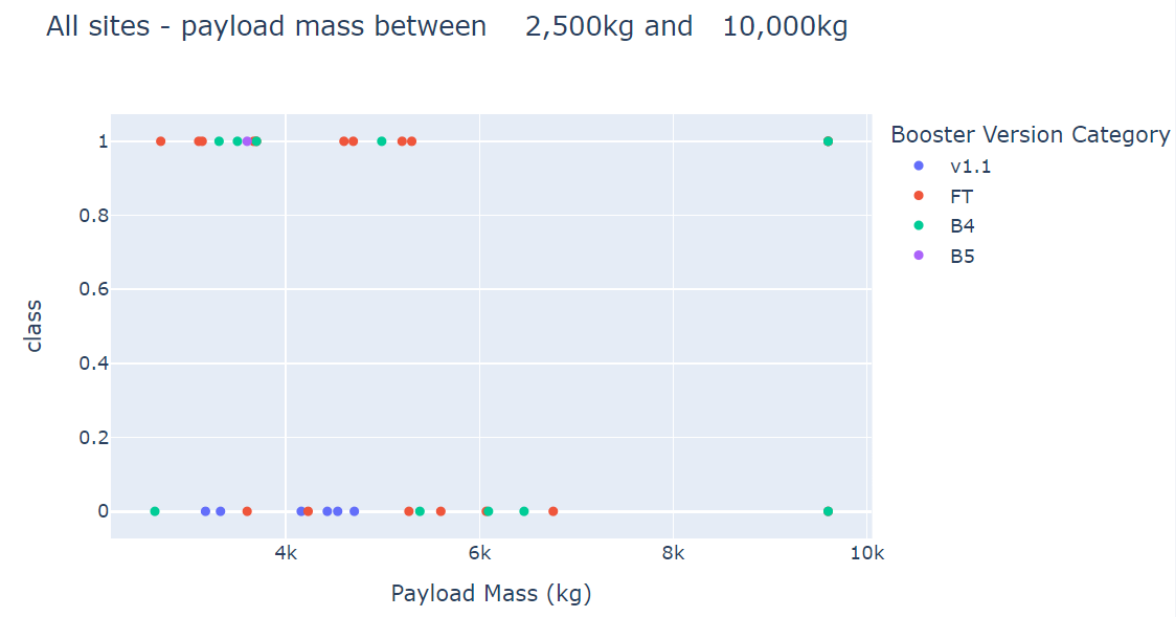
- KSC LC-39A launch site has a high launch success rate

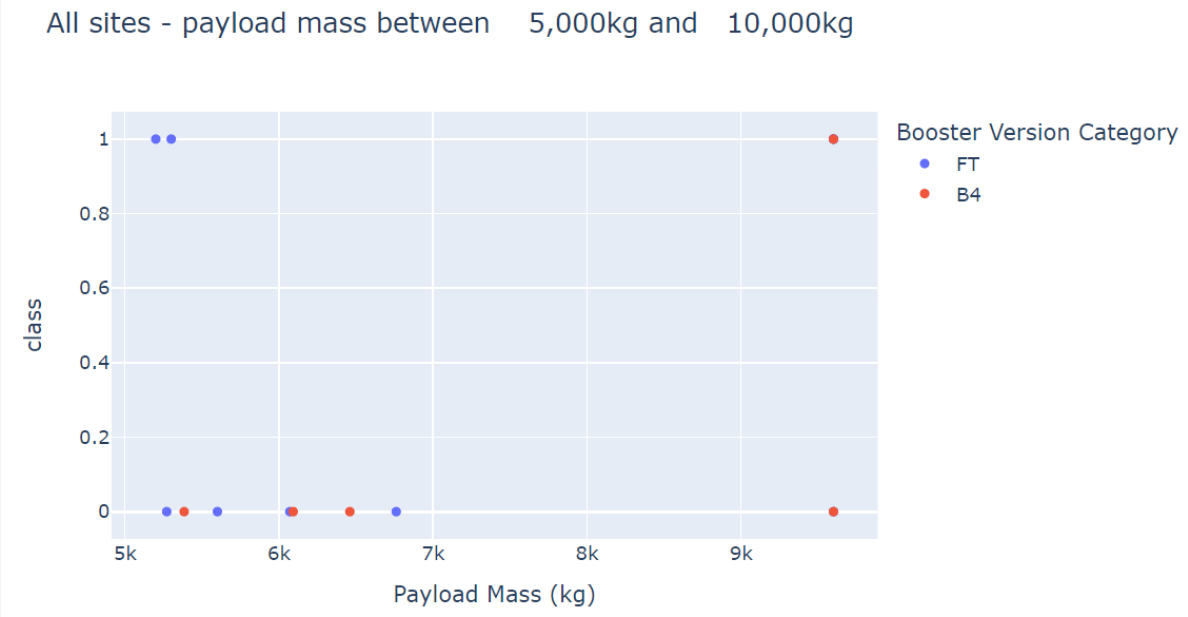
launch site with highest launch success



- 76.9% of launches done in KSC LC-39A are successful

Payload vs. Launch Outcome





- As the payload mass increase, successful launch decreases

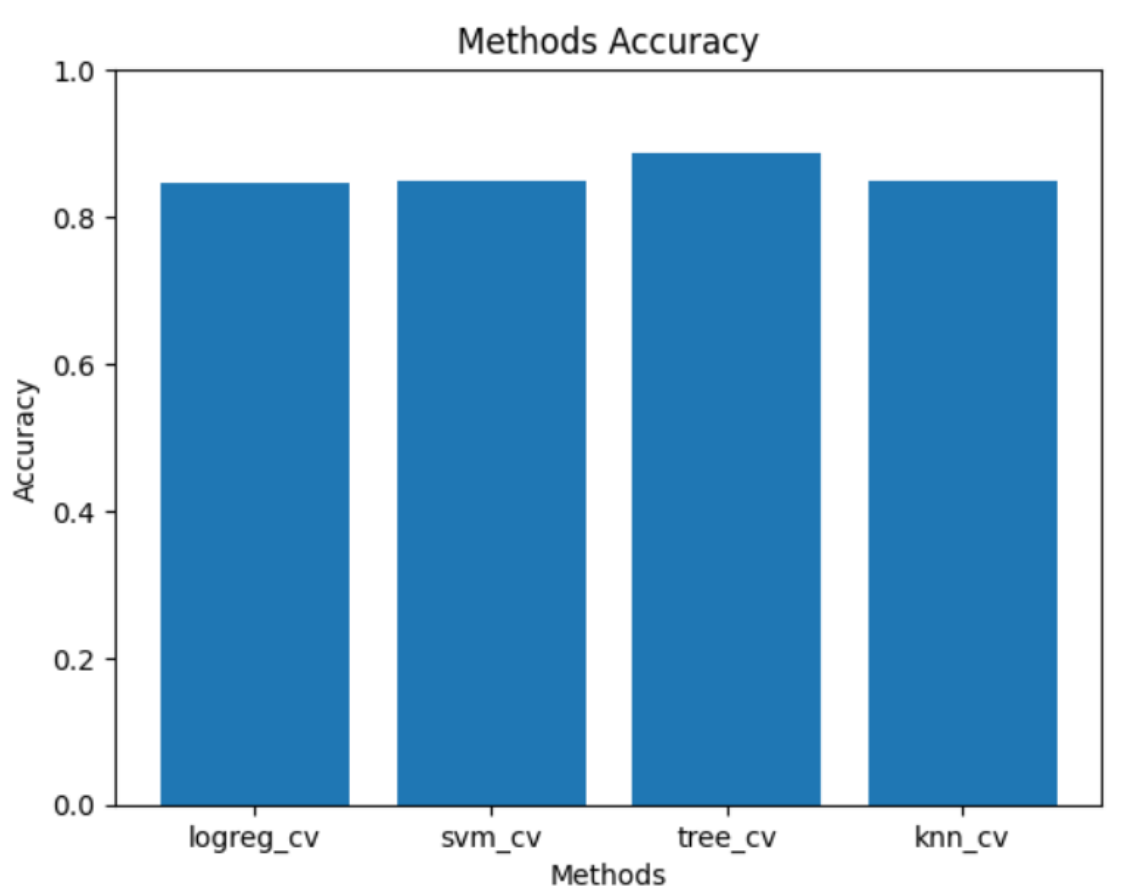
Section 5

Predictive Analysis (Classification)

Classification Accuracy

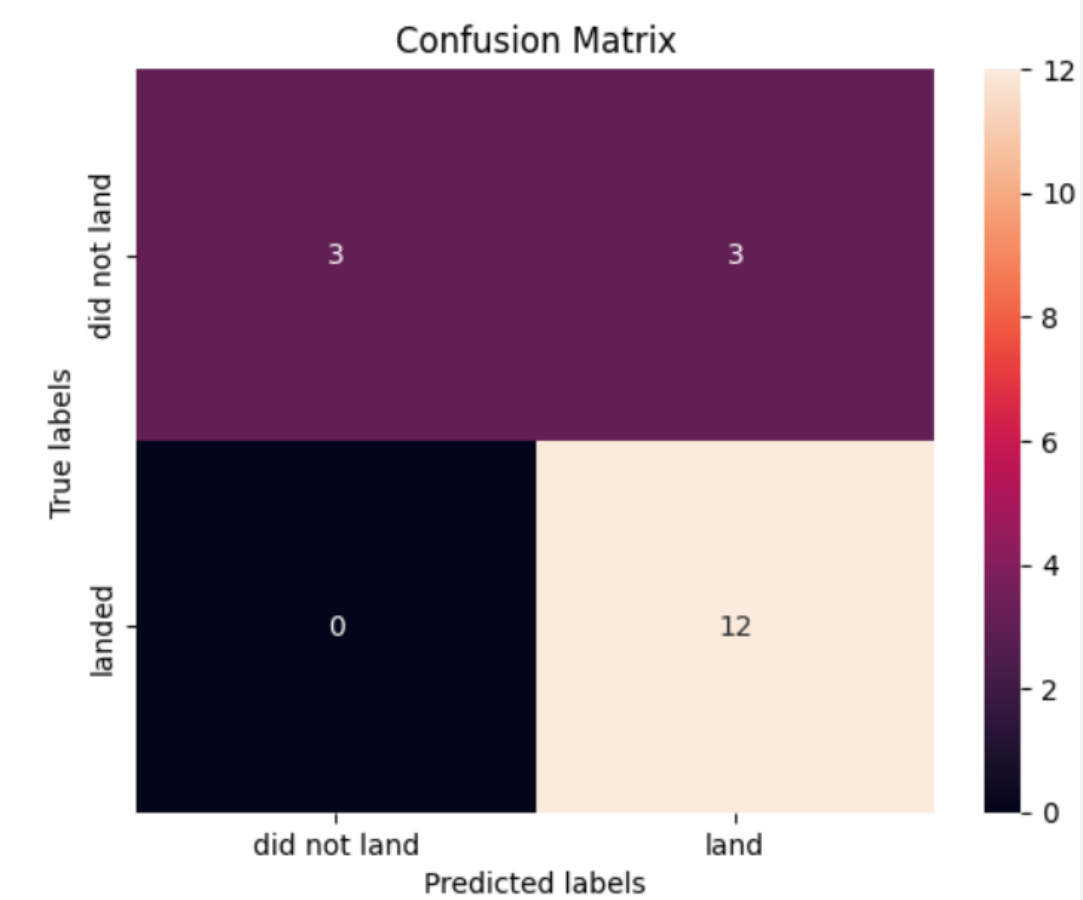
- The bar chart visualizes the built model accuracy for all built classification models

- Decision tree classifier was had the highest classification accuracy



Confusion Matrix

Confusion matrix of the best performing model shows that the model is accurate by showing big numbers in the true positive and true negative



Conclusions

- Different sources were used for the analysis
- KSC LC-39A launch site has a high launch success rate
- As the payload mass increase, successful launch decreases
- Decision tree classifier is the best model for this dataset

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!