IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Emmanuel Obeng Agyen
07/10/2022

Outline

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

Executive Summary

- The methodologies used to analyze data:
  - Data Collection was completed using the SpaceX API and web scraping.
  - Data wrangling, data visualization and interactive visual analytics.
  - Machine learning prediction
- Summary of all results:
  - The exploratory data analysis activity revealed features to consider which were best suitable to predict the success of a landing and launch.
  - Machine learning classification models build helps to predict the best characteristics to drive the space company needs to be able to be competitive.

Introduction

- The objective of the project is to assist company Space Y with successful landings of their rockets to enable them compete with Space X
- This project predicts successful landings of the first stage of the rockets which will help Space Y.
- Finally knowing the best locations to make launches helps to ensure successful landings and thus the project tackles that also.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data from Space X was obtained from:
    - Space X API
    - Web Scraping
- Perform data wrangling
  - A landing outcome label was created, and it contained the outcome of the lunch landing.
- Perform exploratory data analysis (EDA) using visualization and SQL

Methodology

Executive Summary

- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
    - The data was split into test and training data sets. Four different classification models were used and for each the accuracy and also confusion matrix for the model was determined to measure the performance of the model.

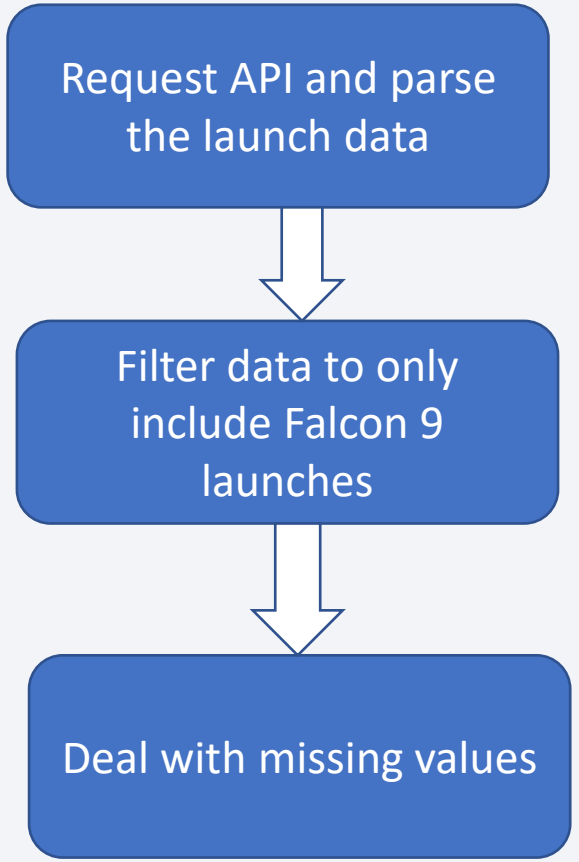
Data Collection

- Data sets were acquired from:
  - Space X API
  - Wikipedia using web scraping

Data Collection – SpaceX API

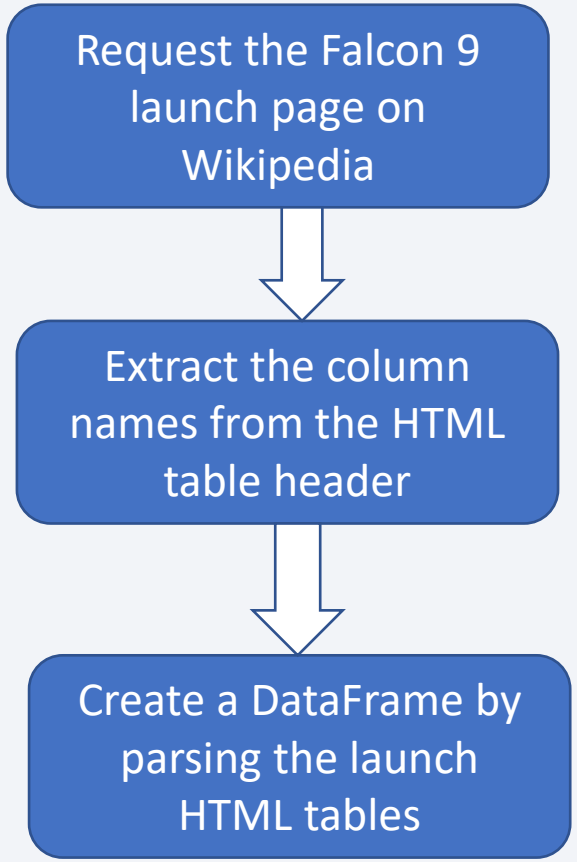
- SpaceX API offers the data about SpaceX launches.

- GitHub URL: https://github.com/eoagyen/ibm-machine-learning/blob/master/Data%20Collection%20API.ipynb



Data Collection - Scraping

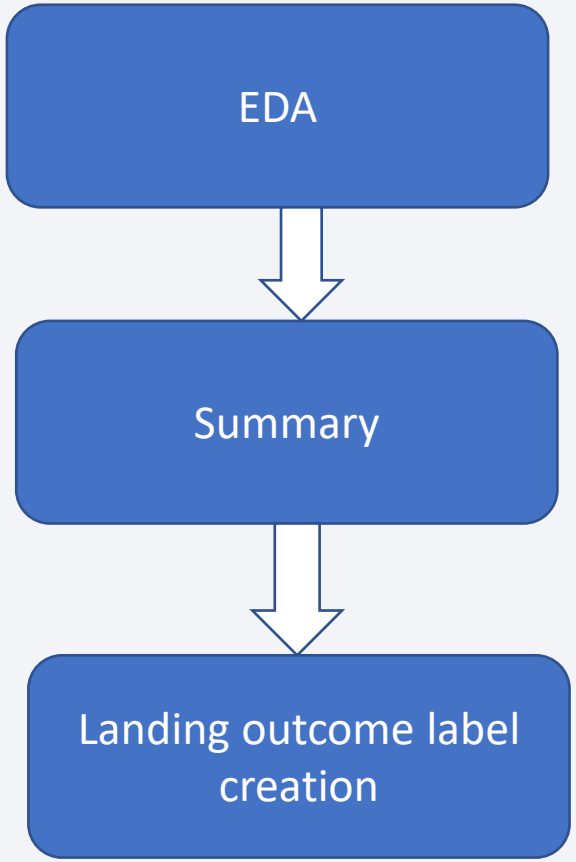
- SpaceX launch data is also obtained from Wikipedia.
- The data are downloaded as shown in the chart.



- GitHub URL: https://github.com/eoagyen/ibm-machine-learning/blob/master/Data%20Collection%20with%20Web%20Scraping.ipynb

Data Wrangling

- Exploratory Data Analysis (EDA) was done on the dataset.
- Occurrence of the mission outcome and summary of launches done per site was evaluated.
- The landing outcome label was also created from the column for outcome.
- GitHub URL: https://github.com/eoagyen/ibm-machine-learning/blob/master/EDA.ipynb



EDA with Data Visualization

- Scatter plots and bar charts were employed to explore the data using visualization and determine relationships between features
- GitHub URL: https://github.com/eoagyen/ibm-machine-learning/blob/master/EDA%20with%20Data%20Visualization.ipynb

EDA with SQL

- The SQL queries performed were:
  - 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
  - Date when the first successful landing outcome in ground pad was achieved
  - Names of the boosters which have success in drone ship
  - Total number of successful and failure mission outcomes and the booster versions which have carried the maximum payload mass
  - Failed landing outcomes in drone ship, their booster versions, and launch site names
  - Rank of the count of landing outcomes between the date 2010-06-04 and 2017-03-20.
- GitHub URL: https://github.com/eoagyen/ibm-machine-learning/blob/master/EDA%20with%20SQL.ipynb

Build an Interactive Map with Folium

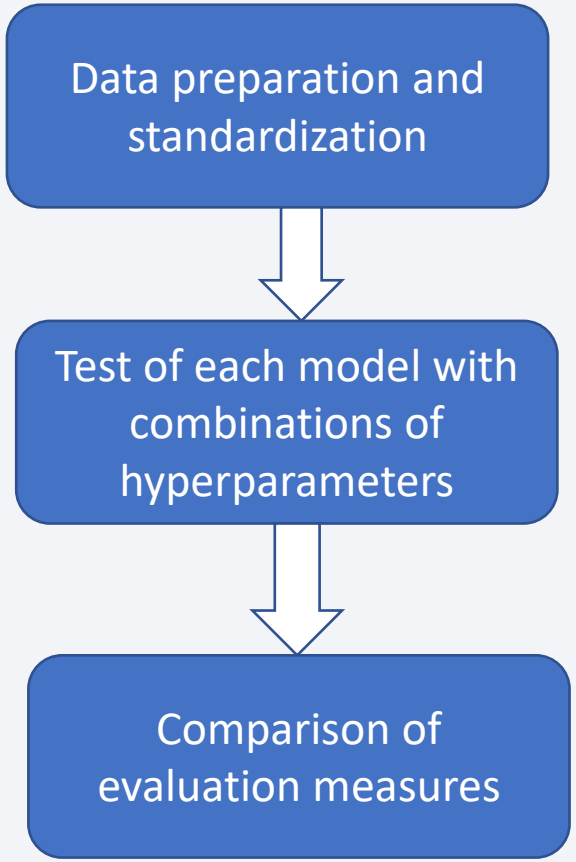
- The map objects such as markers, circles, lines and marker clusters were added to the folium maps created.
- The markers:
  - Were used to indicate points of the launch sites
  - Highlight areas around some specific coordinates
  - To indicate events at a particular coordinate
  - To indicate the distances between two specific coordinates
- GitHub URL: https://github.com/eoagyen/ibm-machine-learning/blob/master/Interactive%20Visual%20Analytics%20with%20Folium%20lab.ipynb

Build a Dashboard with Plotly Dash

- Graphs and plots were used to visualize data to show the percentage of launches by site and payload range
- These plots helped to identify the best place to launch according to payloads.

Predictive Analysis (Classification)

- Four classification models were trained and compared.
- The models are:
  - Logistic regression
  - Support vector machine
  - Decision tree
  - K nearest neighbors
- GitHub URL: https://github.com/eoagyen/ibm-machine-learning/blob/master/Machine%20Learning%20Prediction.ipynb

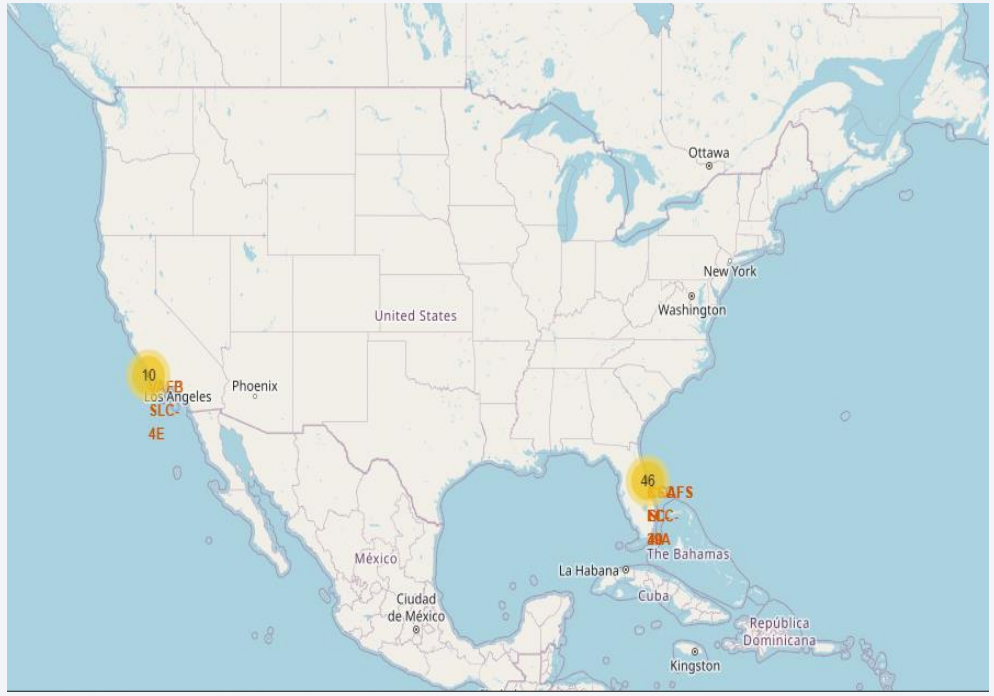


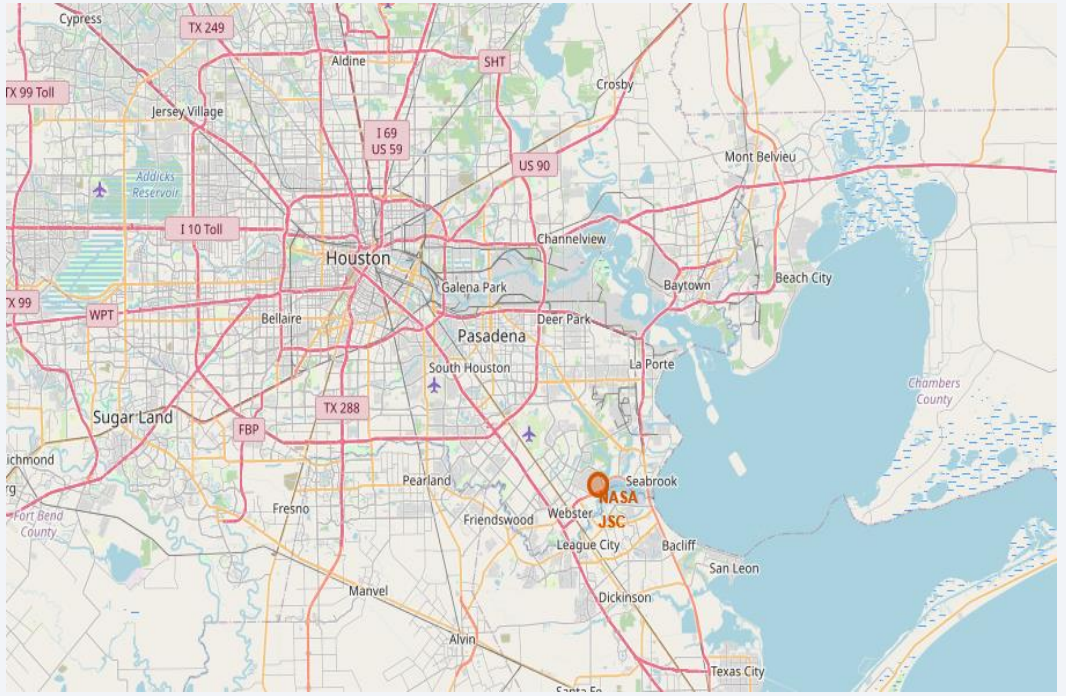
Results

- Exploratory data analysis results
  - 4 different launch sites were used by SpaceX
  - The mean payload of F9 v1.1 booster is 2,928 kg
  - The first successful landing outcome occurred in 2015, five years after the first launch
  - Most success by the Falcon 9 booster versions occurred by using drone ships payload above the average
  - The number of successful landing outcomes increased as the years went by

Results

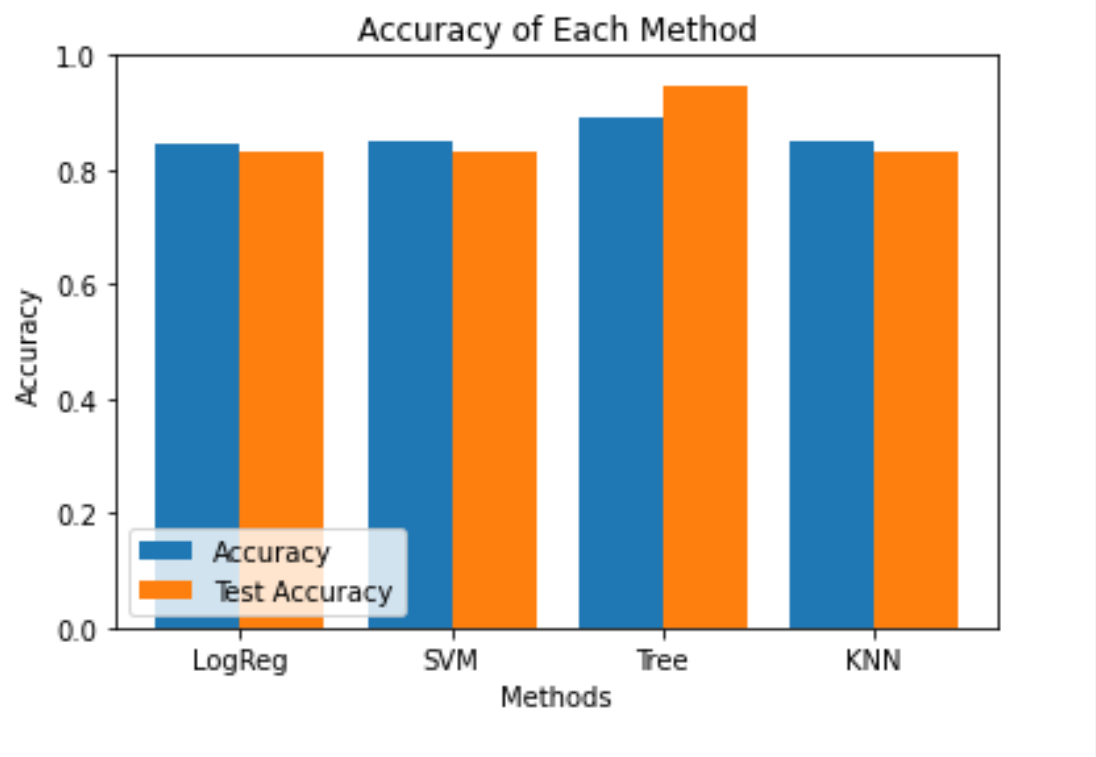
- Interactive analytics demo:

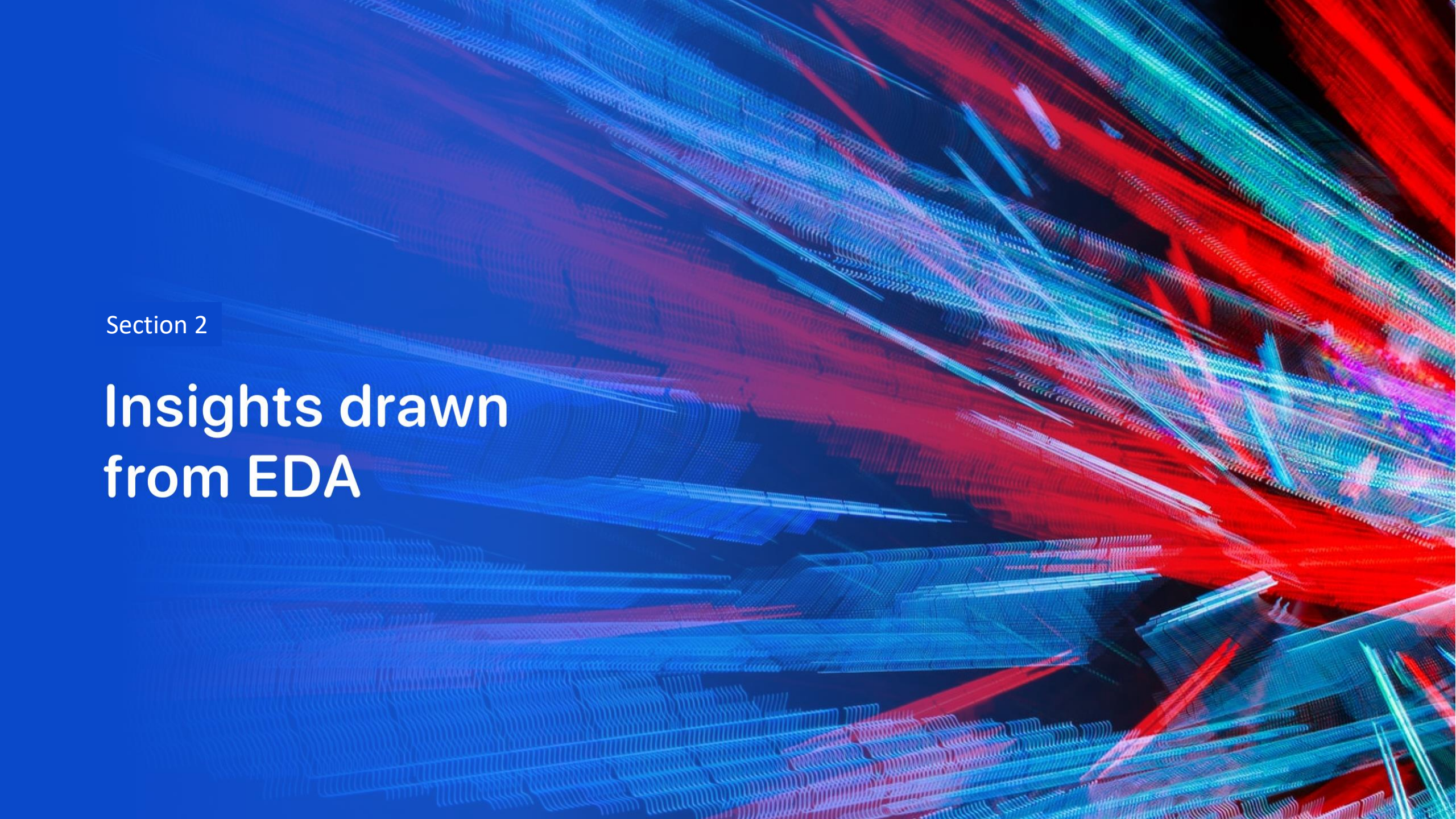




Results

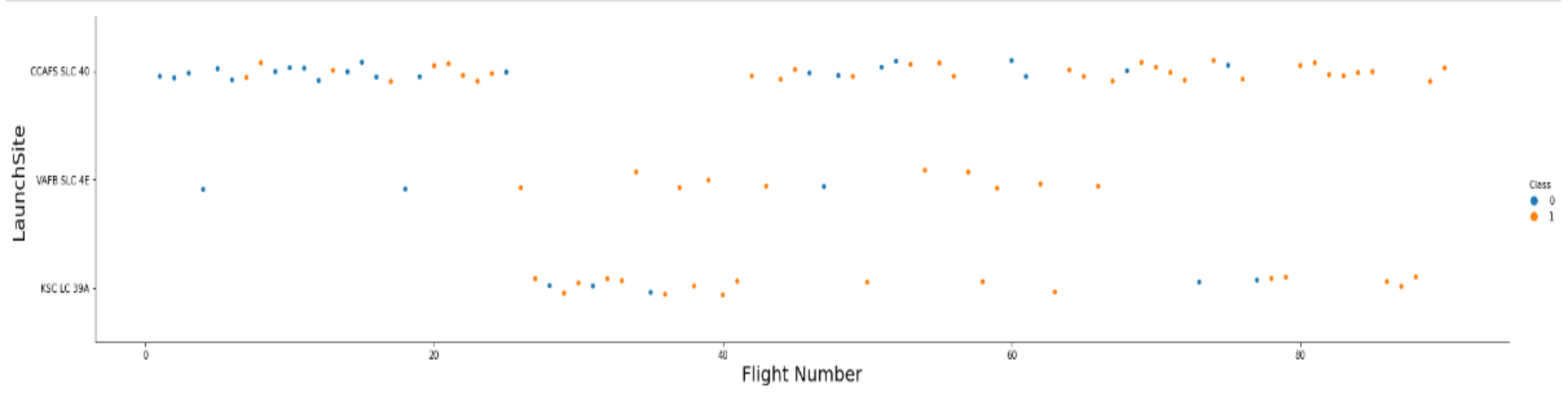
- The Decision Tree Classifier is the model with the highest accuracy as shown from the bar chart.




Section 2
Insights drawn from EDA

Flight Number vs. Launch Site

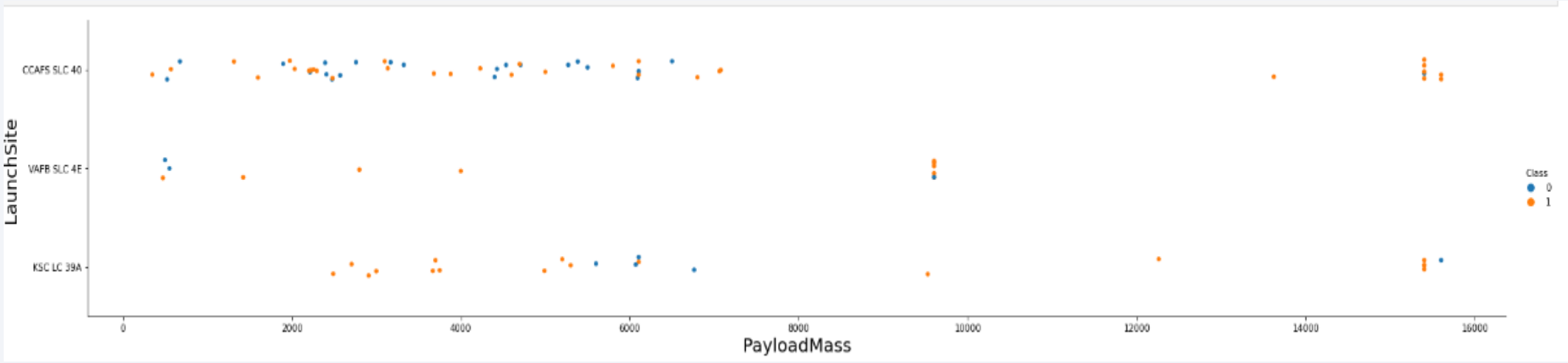
- From the plot below, a chunk of the recent successful launches took place at the CCAF5 SLC 40.



Payload vs. Launch Site

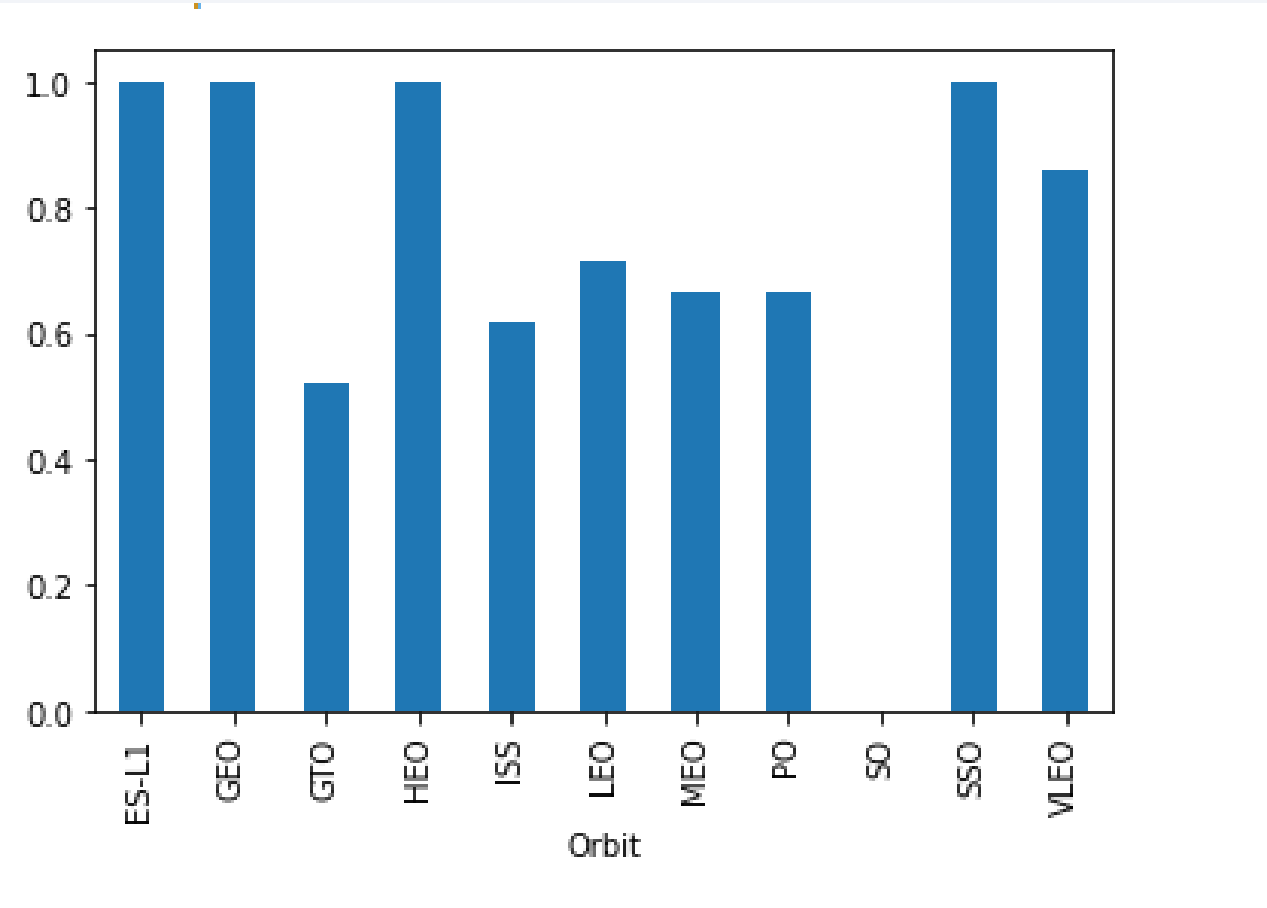
- From the plot, payloads over 9,000 kg have good success.

- Most of the successful launches with payload over 12,000 kg occurred at CCAF5 SLC 40



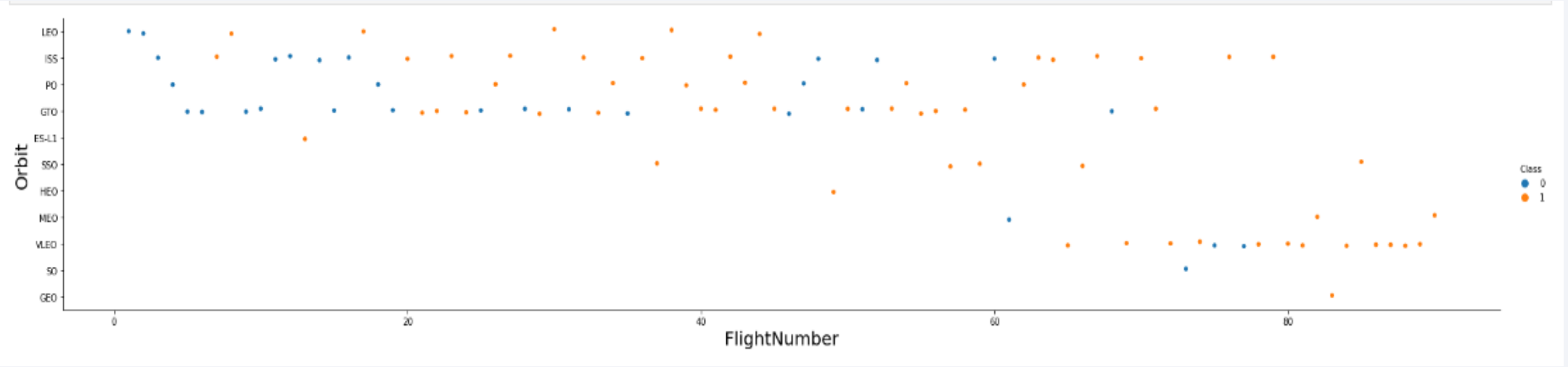
Success Rate vs. Orbit Type

- The highest success rates occurred at:
  - ES-L1
  - GEO
  - HEO
  - SSO



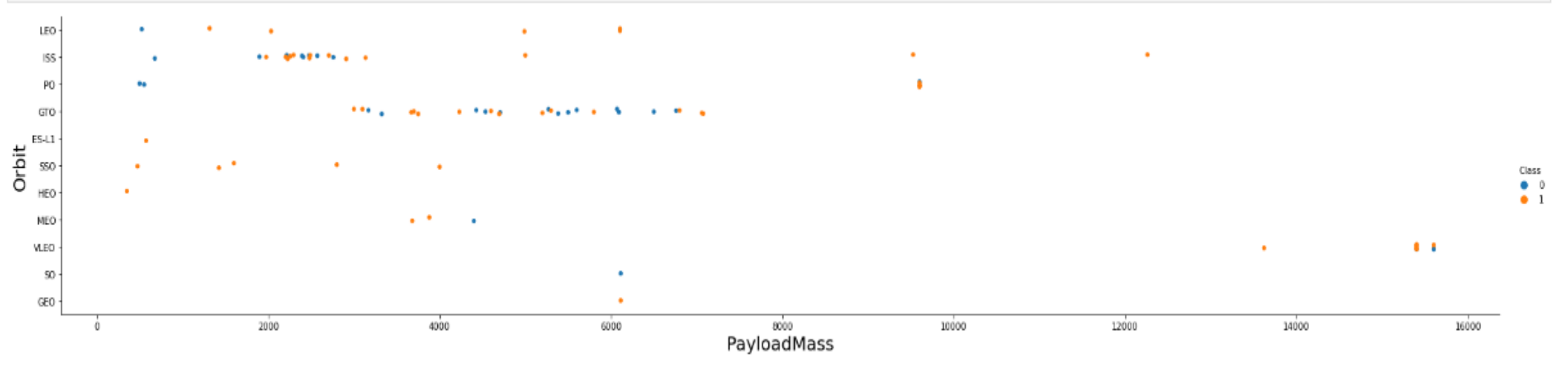
Flight Number vs. Orbit Type

- VLEO orbit had an increase of its recent success and thus looks more to be explored further.



Payload vs. Orbit Type

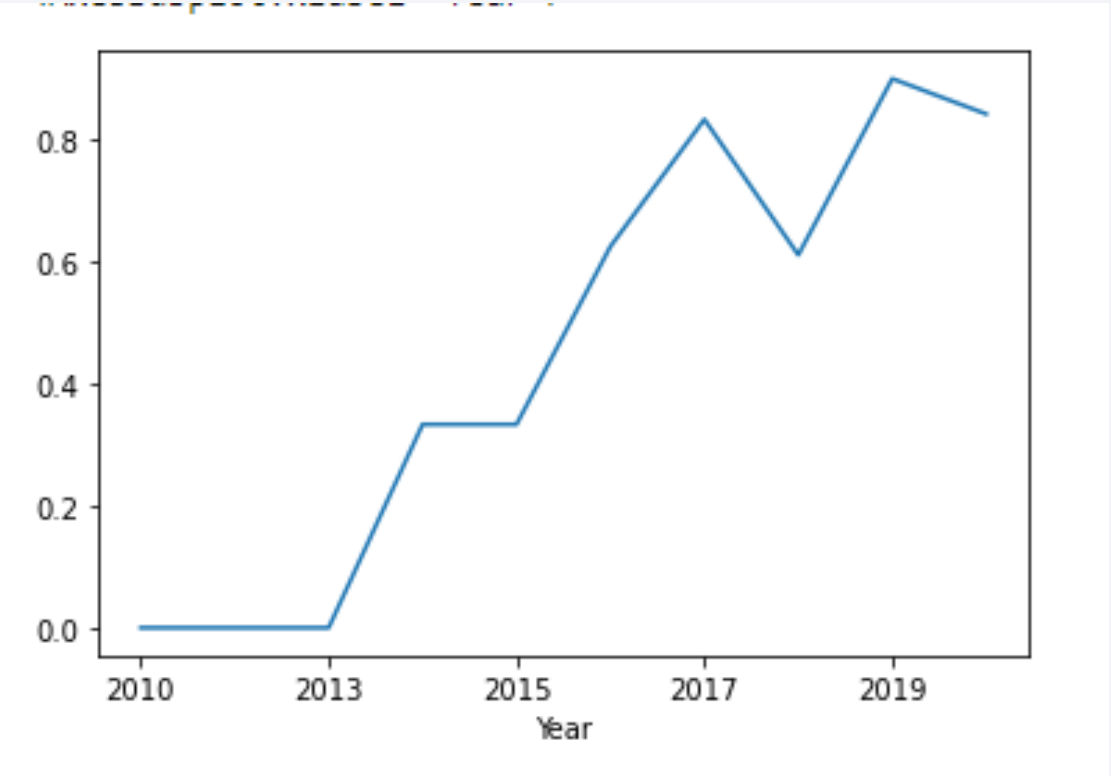
- GEO had a few launches as compared to the other orbits



Launch Success Yearly Trend

- The success rate increases from 2013 and kept raising till 2020

- It seems there were no success from 2010 to 2013.



All Launch Site Names

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

Launch Site Names Begin with 'CCA'

- 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	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 attemp

Total Payload Mass

- Total payload carried by boosters from NASA

Total Payload (kg)
111.268

Average Payload Mass by F9 v1.1

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

Avg Payload (kg)
2.928

First Successful Ground Landing Date

- First successful landing outcome on ground pad occurred in:

Min Date
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

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes

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

Boosters Carried Maximum Payload

- 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

2015 Launch Records

- 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

Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

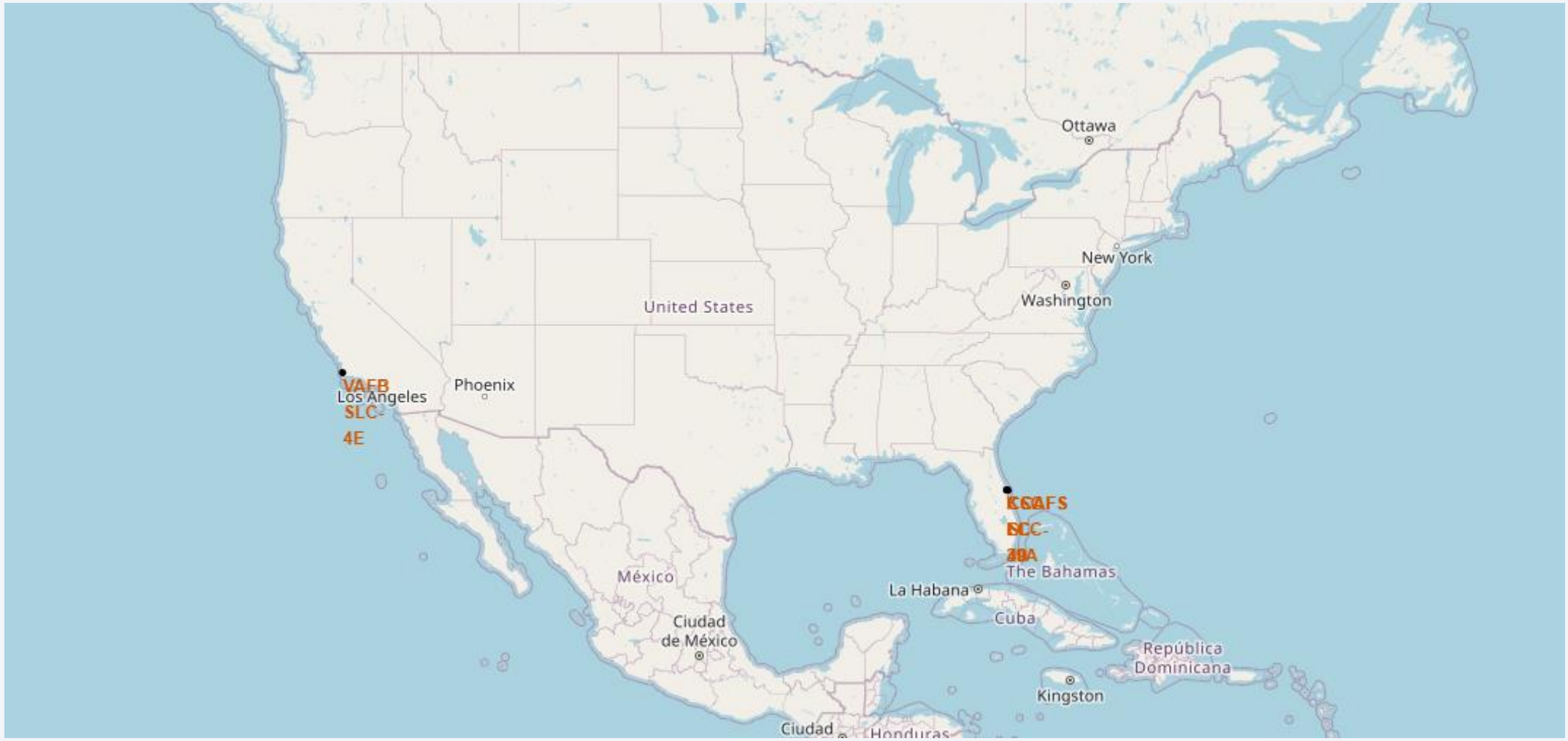
- Ranking of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

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

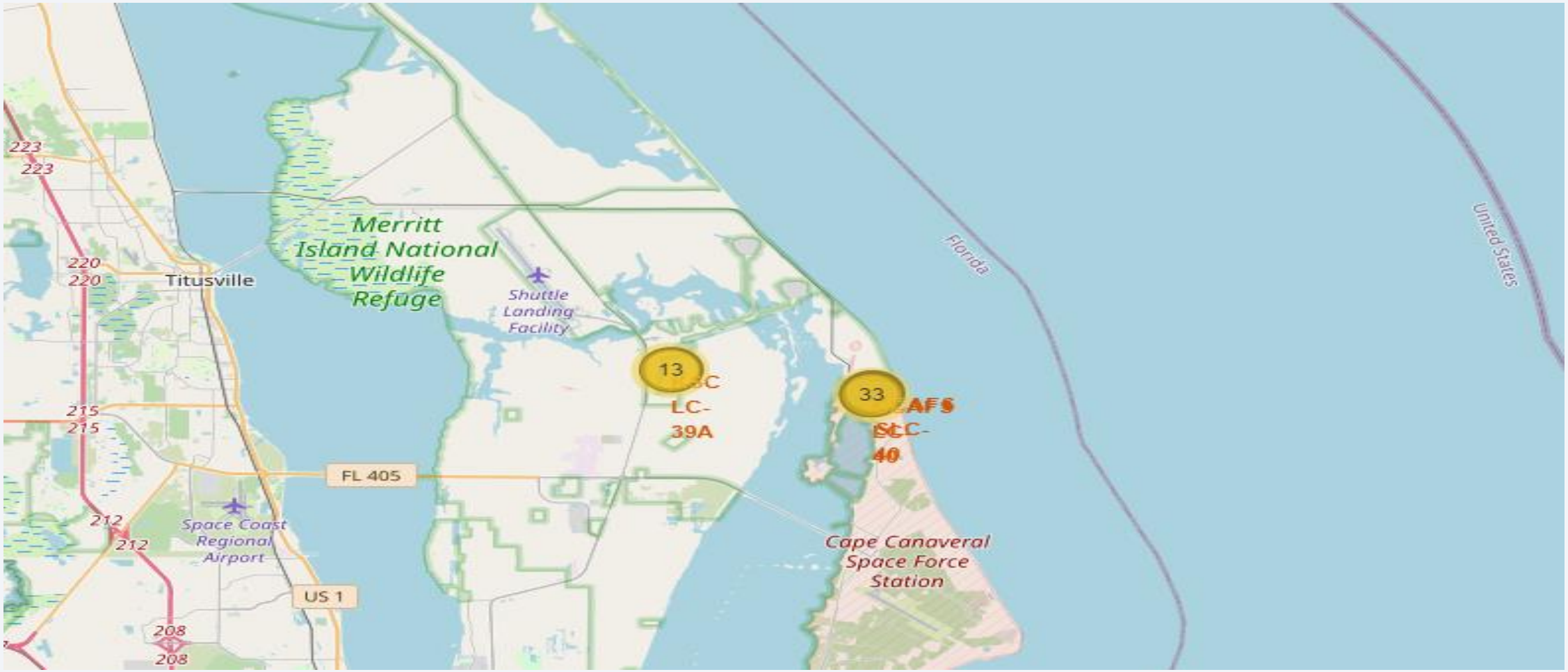
Section 3

Launch Sites Proximities Analysis

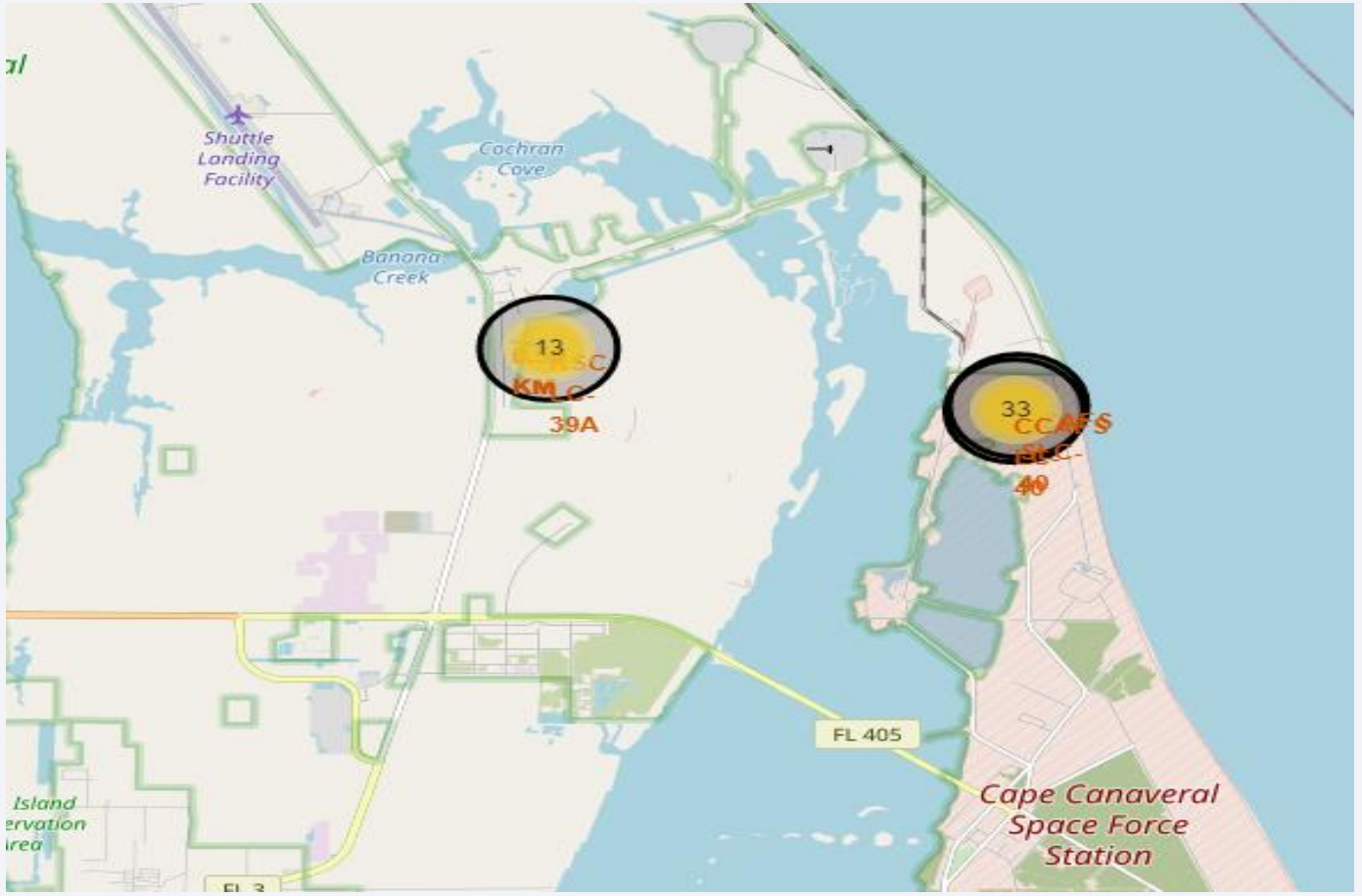
All Launch Sites

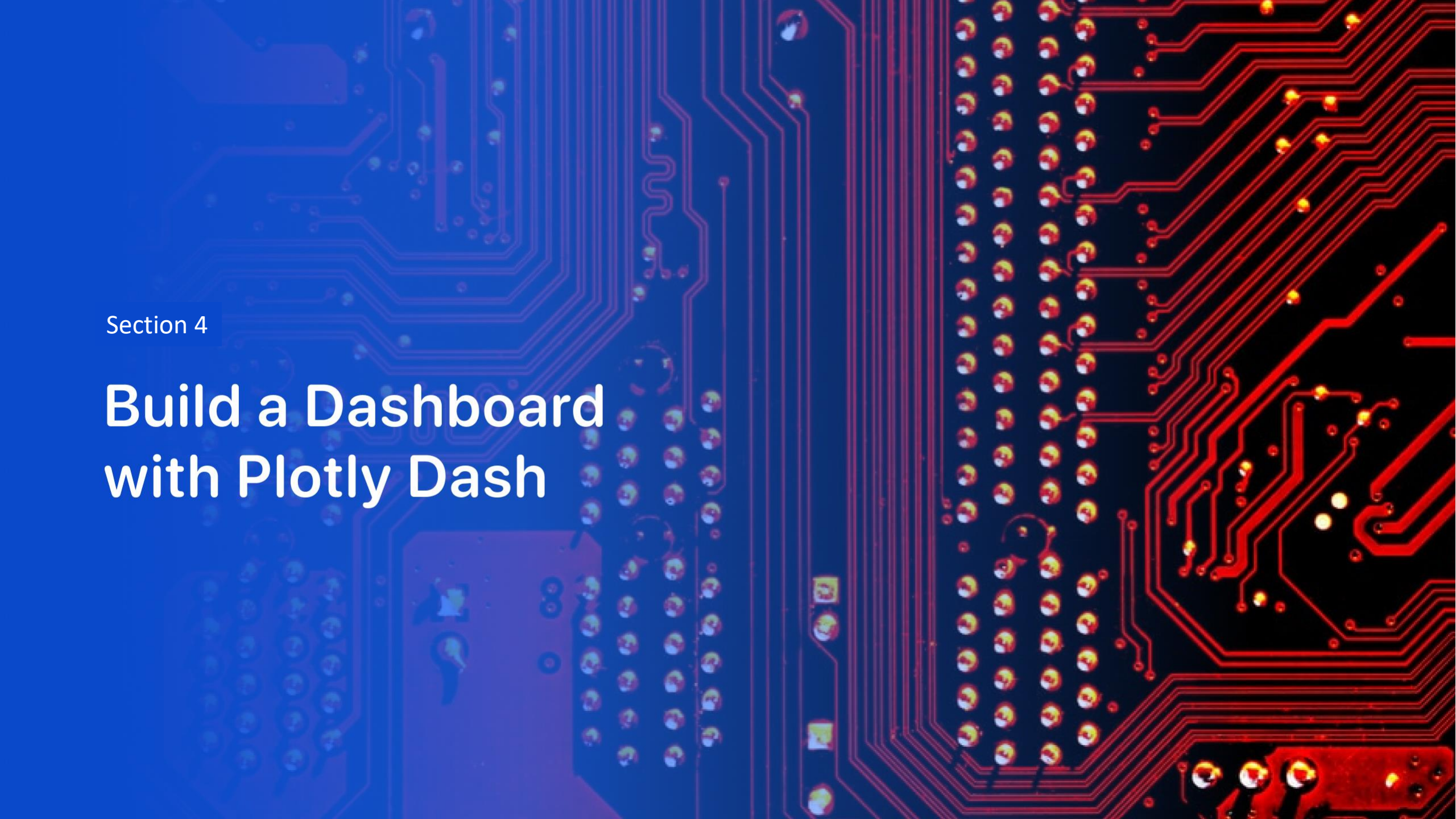


Launch Outcomes by Site



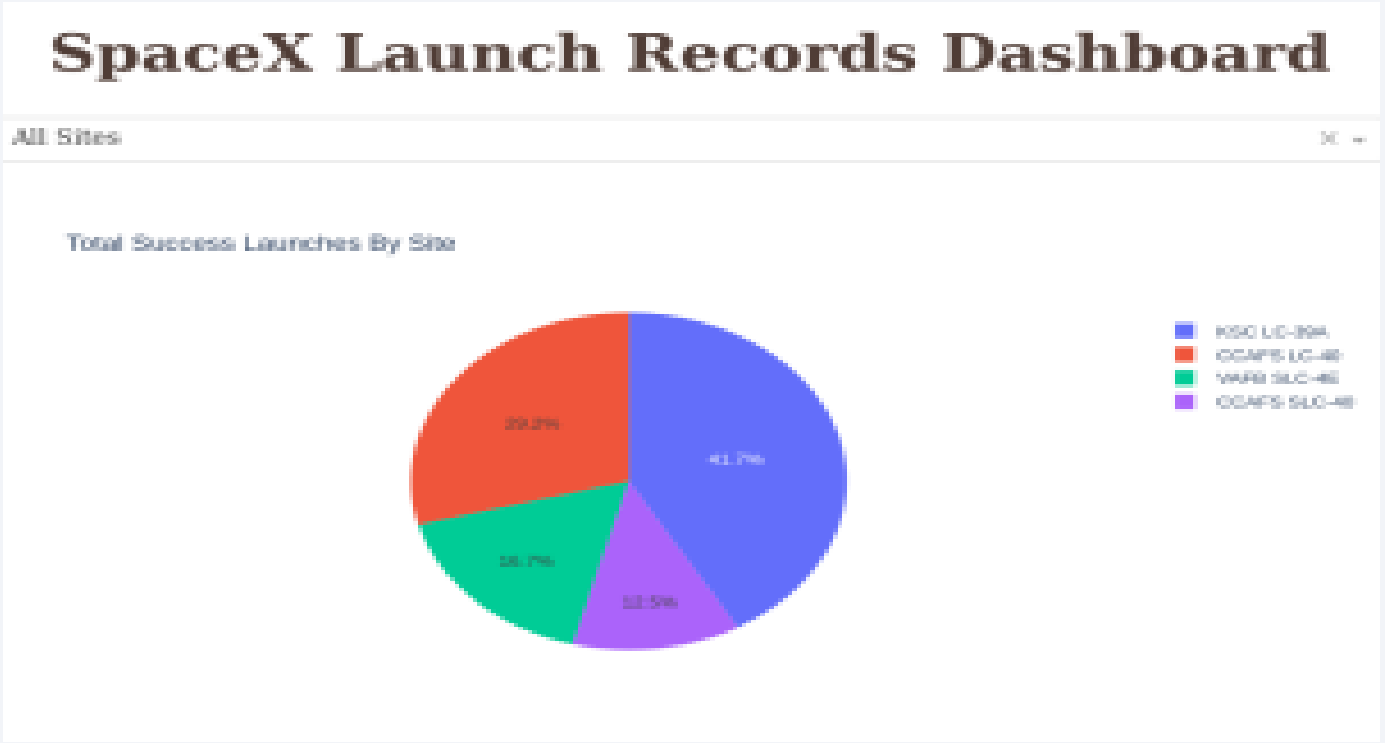
Logistics and Safety




Section 4
Build a Dashboard with Plotly Dash

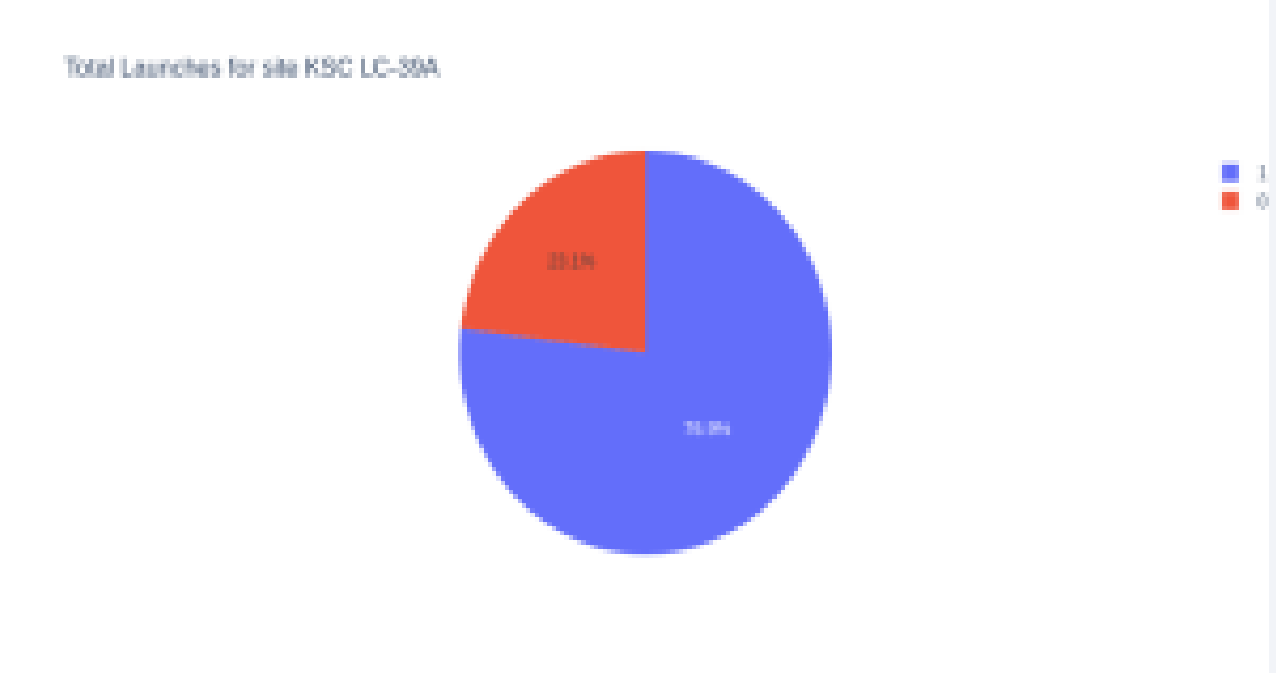
Successful Launches by Site

- The location where the launches takes place affects the success of the launch



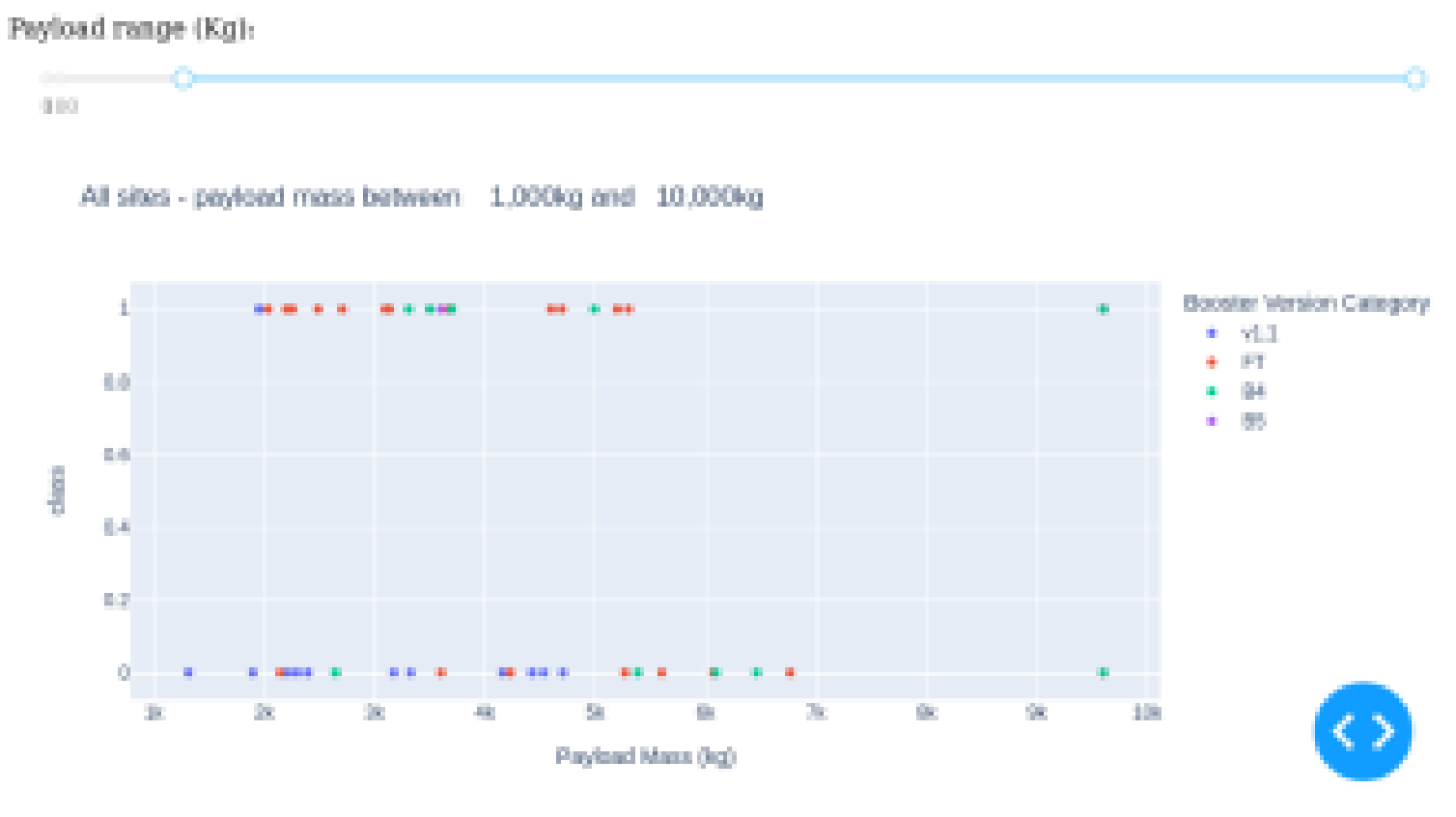
Launch Success Ratio for KSC LC-39A

- This site had a success rate of 76.9%



Payload vs. Launch Outcome

- The most successful combinations were payloads under 6,000 kg and FT boosters

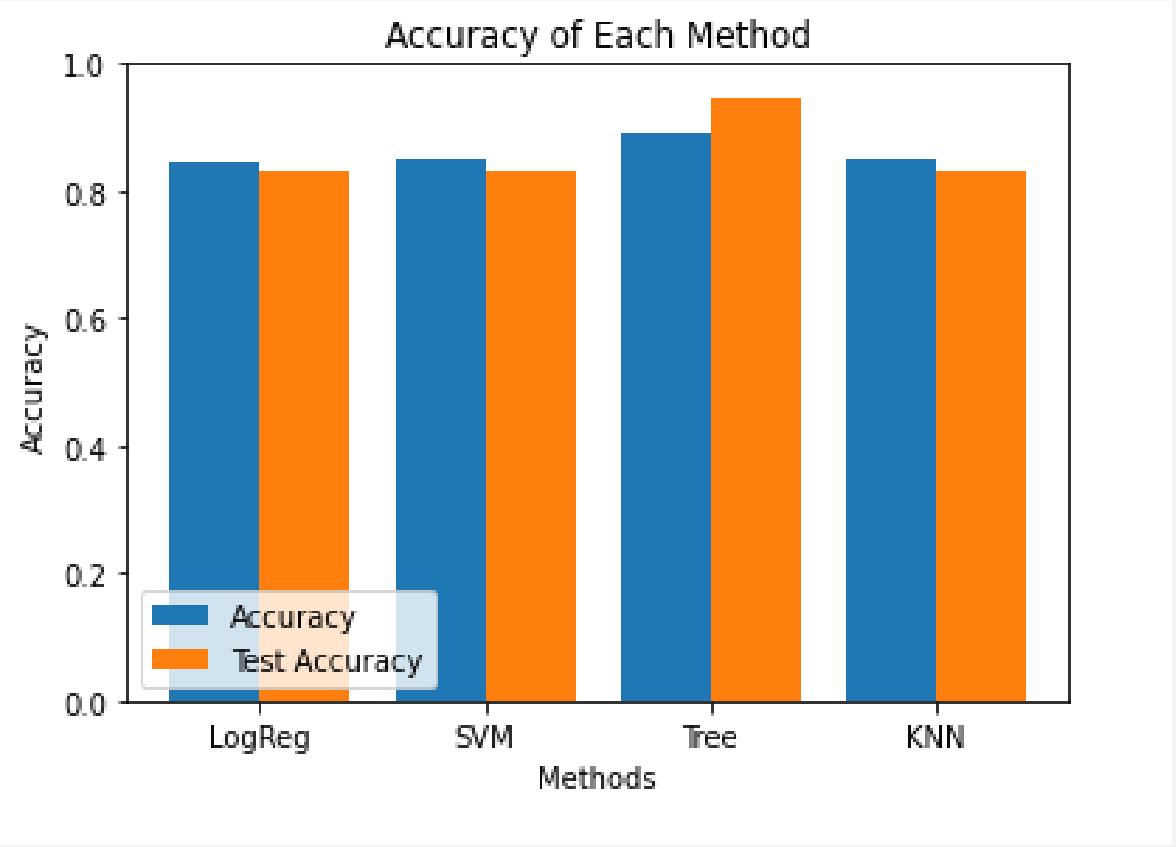


Section 5

Predictive Analysis (Classification)

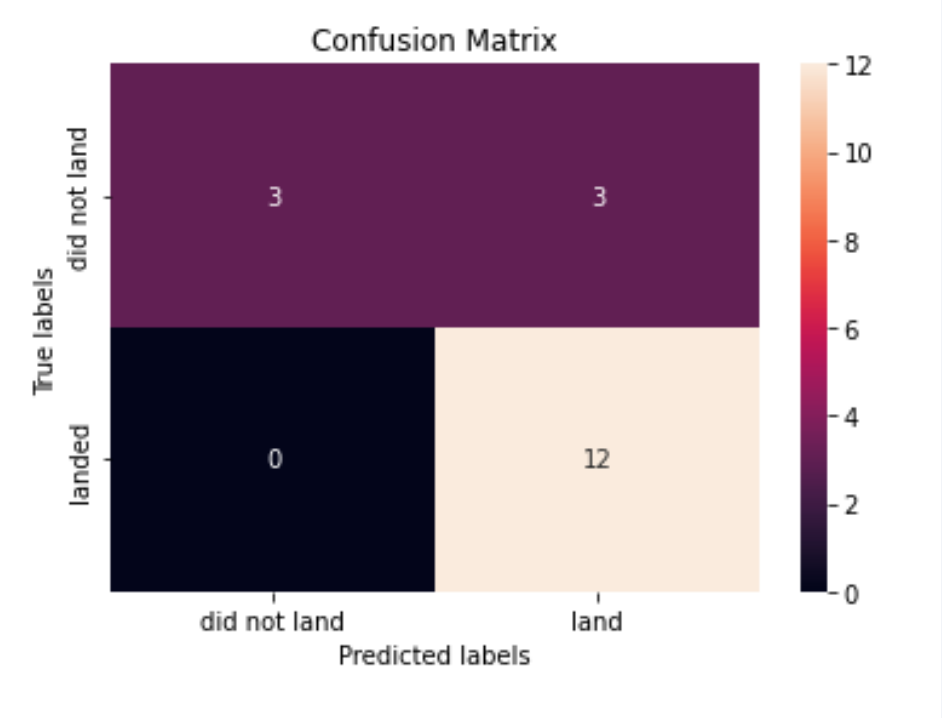
Classification Accuracy

- The accuracies of the four classification models are compared in the bar chart.
- The Decision Tree Classifier is the model with the highest accuracy as shown from the bar chart.



Confusion Matrix of Decision Tree Classifier

- The confusion matrix shows further proves the accuracy of the model as it shows a high number of true positives and true negatives.



Conclusions

- The mission outcomes of the launches improved over the years and more successful outcomes were recorded.
- The best site for a launch is KSC LC-39A
- Decision Tree Classifier is the best model that can be used to predict the success of the landing as and thus increase profits to make the Space company competitive.

Appendix

- The various plots and charts snippets were included to make it able to easily refer.

Thank you!