IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Nicolás González Jaramillo
15/08/2022

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection using API's and Webscraping
  - EDA using different python libraries
  - Model Development based on Machine Learning techniques.
- Summary of all results
  - Successful data collection from public sources
  - The best features to describe the model that predict success of launchings were identified by using EDA techniques.
  - Using the data collected, Machine Learning models were developed to find the characteristics that are important to get the best results in the launch of a rocket.

Introduction

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

- Problems that need a solution:
  - Cost for launches depend on successful landings of the first stage of rockets.
  - Place from where rocket is launched may affect if it is successful or not.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Collection by API (https://api.spacexdata.com/v4/rockets/)
  - Collection by Webscraping (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_)
- Perform data wrangling
  - New variables were created depending on landing outcome
- 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
    - Selected data that will compose the model is normalized and separated in training and testing sets. Then, different models were evaluated to determine which was more accurate.

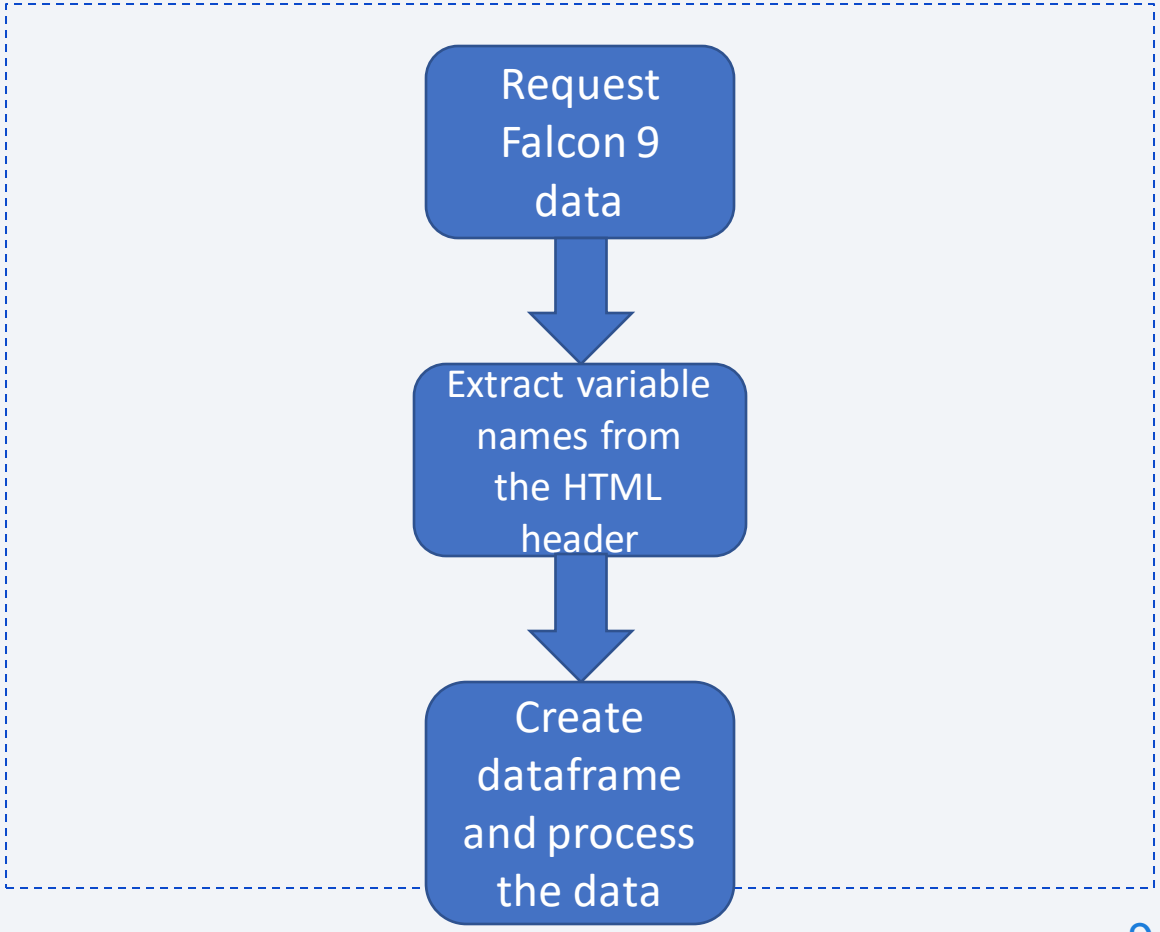
Data Collection – SpaceX API

- Space X Public API allow everyone to access some data to be obtained and then used.
- GitHub: https://github.com/nigoja97/Coursera/blob/main/IBM%20Data%20Science/Applied%20Data%20Science%20Capstone/Data%20Collection%20API.ipynb

Data Collection - Scraping

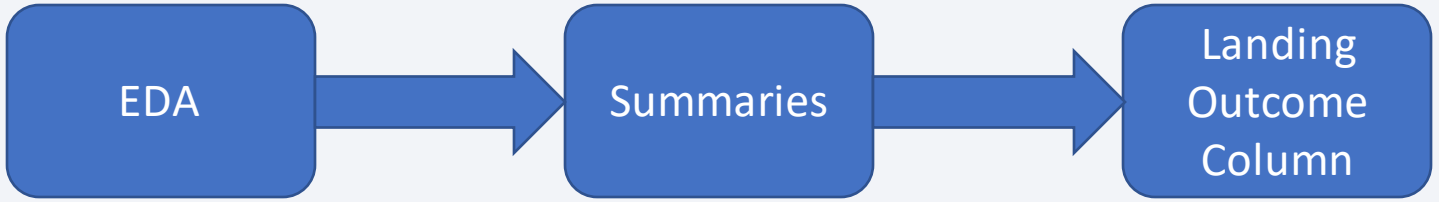
- Wikipedia had some data that was extracted by webscraping techniques.

- GitHub: https://github.com/nigoja97/Coursera/blob/main/IBM%20Data%20Science/Applied%20Data%20Science%20Capstone/Data%20Collection%20with%20Web%20Scraping.ipynb



Data Wrangling

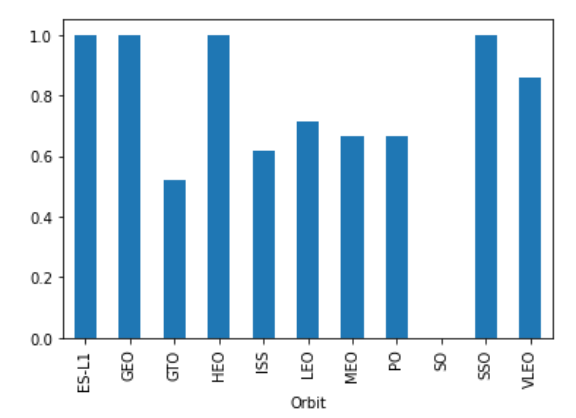
- Exploratory Data Analysis (EDA) to understand the dataset.
- Summaries of launches per site, occurrences of each orbit and mission outcome per orbit type were calculated.
- 'Landing Outcome' label was created from Outcome column.



GitHub: https://github.com/nigoja97/Coursera/blob/main/IBM%20Data%20Science/Applied%20Data%20Science%20Capstone/Data%20Wrangling.ipynb

EDA with Data Visualization

- Barplots and Scatterplots were used to understand and visualize the relationship existent between the success of a launch and other variables like Payload, Flight Number and Orbit type.



- GitHub: https://github.com/nigoja97/Coursera/blob/main/IBM%20Data%20Science/Applied%20Data%20Science%20Capstone/Exploratory%20Analysis%20with%20Visualitzation.ipynb

EDA with SQL

- The following SQL queries were performed:
  - Names of the unique launch sites in the space mission.
  - Top 5 launch sites where launch sites 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 (groundpad)) between the date 2010 06 04 and 2017 03 20.
- GitHub: https://github.com/nigoja97/Coursera/blob/main/IBM%20Data%20Science/Applied%20Data%20Science%20Capstone/Exploratory%20Analysis%20Using%20SQL.ipynb

Build an Interactive Map with Folium

- Markers, circles and lines were created with Folium to indicate:
  - Points like launch sites (Markers)
  - Specific coordinates (Circles)
  - Events in each coordinate (Marker cluster)
  - Distance between two coordinates (Lines)

- GitHub: https://github.com/nigoja97/Coursera/blob/main/IBM%20Data%20Science/Applied%20Data%20Science%20Capstone/Interactive%20Visual%20Analytics%20with%20Folium.ipynb

Build a Dashboard with Plotly Dash

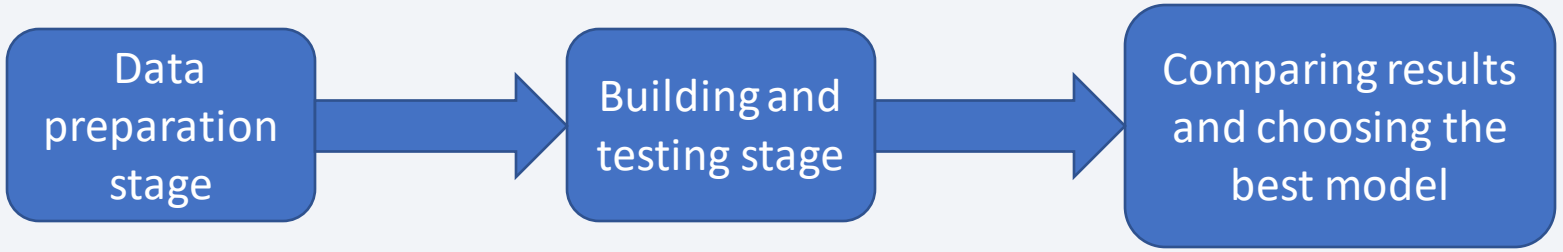
- The graphs and plots were used to visualize data were Percentage of launches by site and Payload range.

- The visualization methods mentioned above helped to identify the best place to launch depending on payloads.

- GitHub: https://github.com/nigoja97/Coursera/tree/main/IBM%20Data%20Science/Applied%20Data%20Science%20Capstone

Predictive Analysis (Classification)

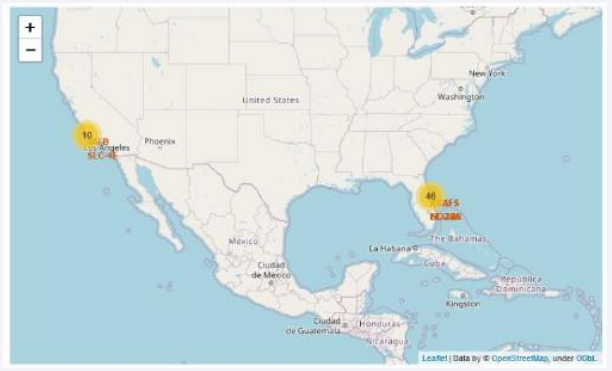
- Four Machine Learning algorithms were built and compared to identify which fits the best. (Logit Model, SVM, KKN and Decision Tree)



- GitHub: https://github.com/nigoja97/Coursera/blob/main/IBM%20Data%20Science/Applied%20Data%20Science%20Capstone/Machine%20Learning%20Prediction.ipynb

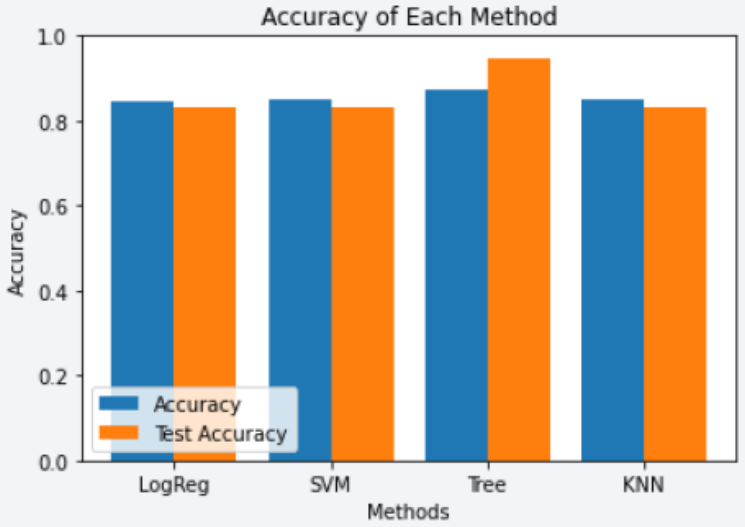
Results

- Exploratory data analysis results:
  - There are 4 different launch sites used by Space X.
  - The average payload of Falcon 9 booster is 2,928 kg;
  - The first success landing outcome happened in 2015
  - 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.



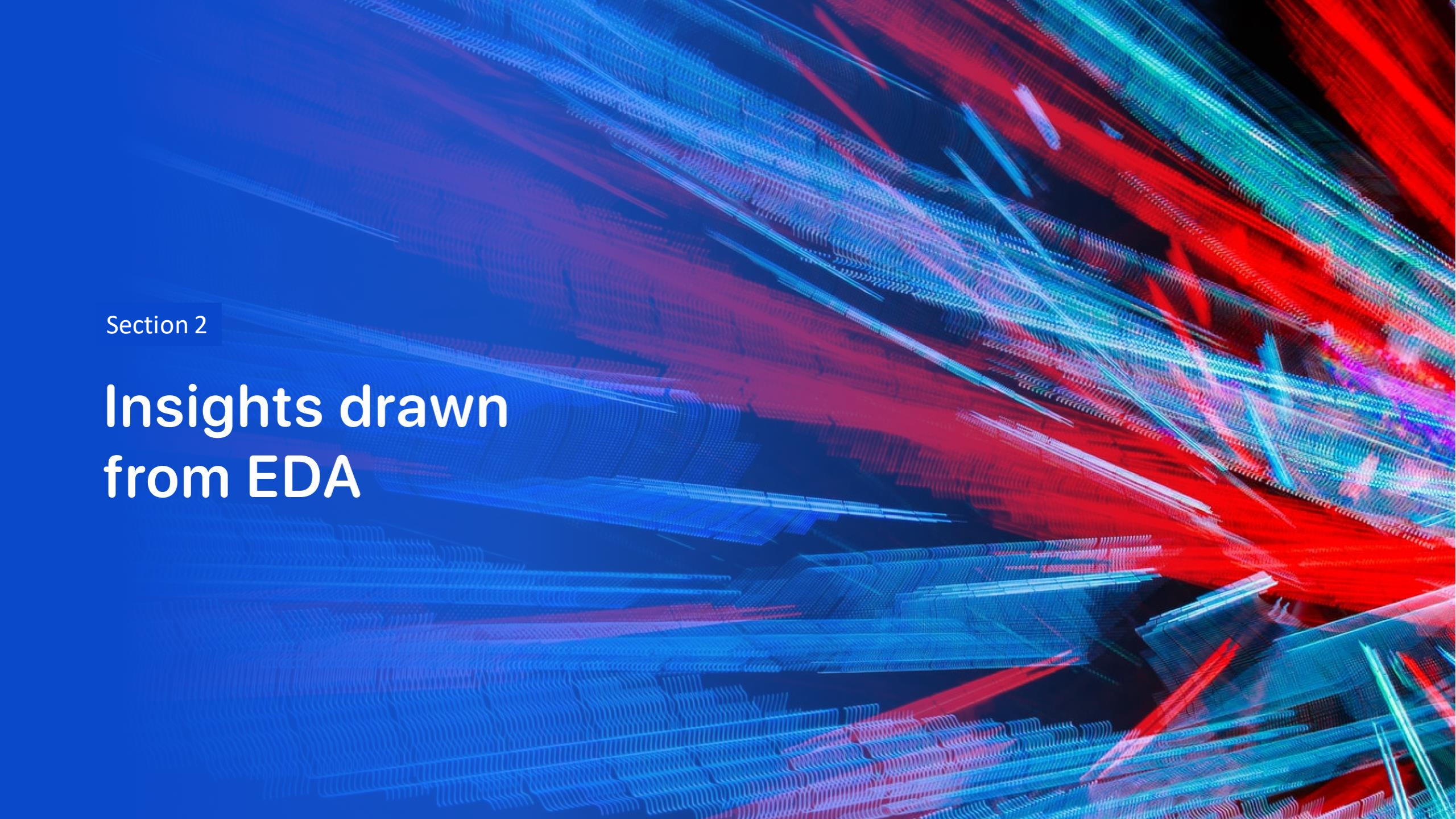
Results

- Predictive analysis results
  - After building, testing and comparing the four models (Logit Model, SVM, KKN and Decision Tree) the analysis showed that the best behaving model was the Decision Tree Classifier having accuracy over 87% and accuracy on testing data of 94%

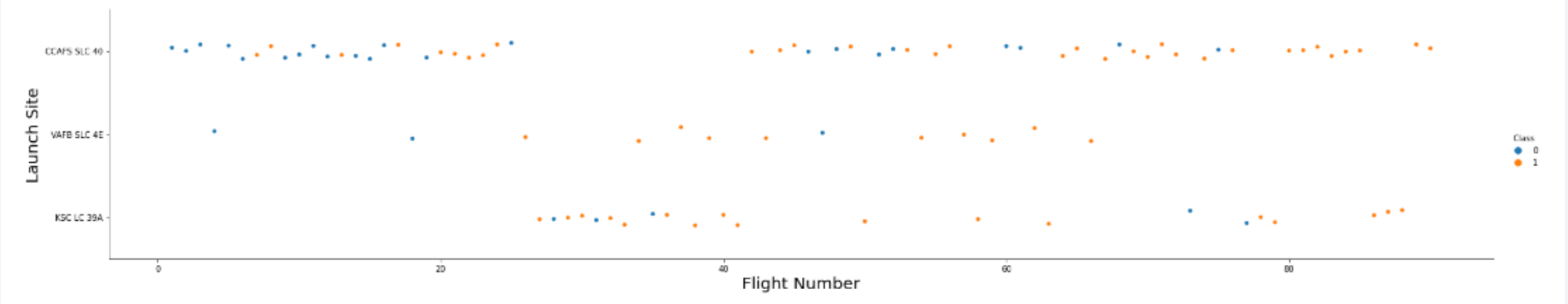


Section 2

Insights drawn from EDA

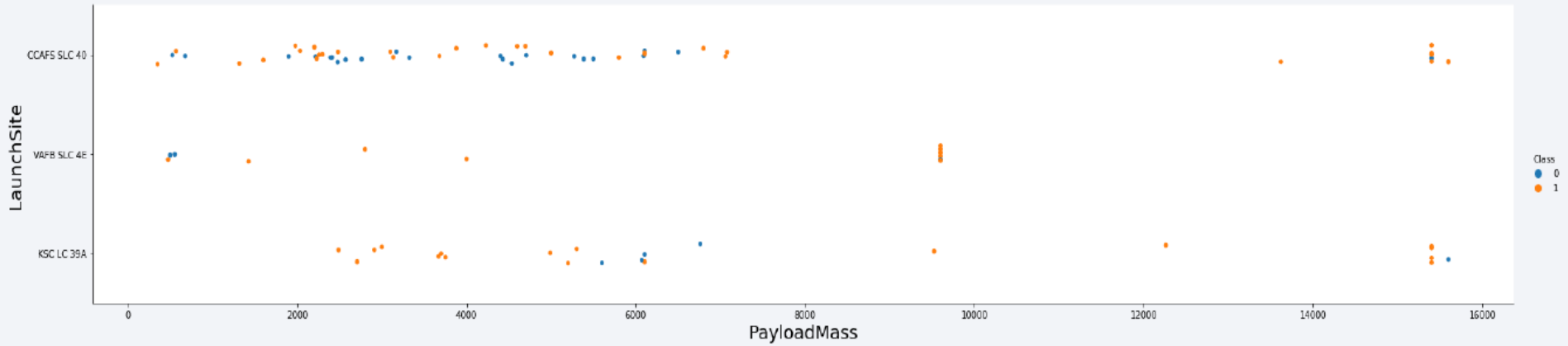


Flight Number vs. Launch Site



- It's possible to get to the conclusion that the success rates become better over time.
- Also, it's clear that CCAPS had the best behaving scenarios for launchings.

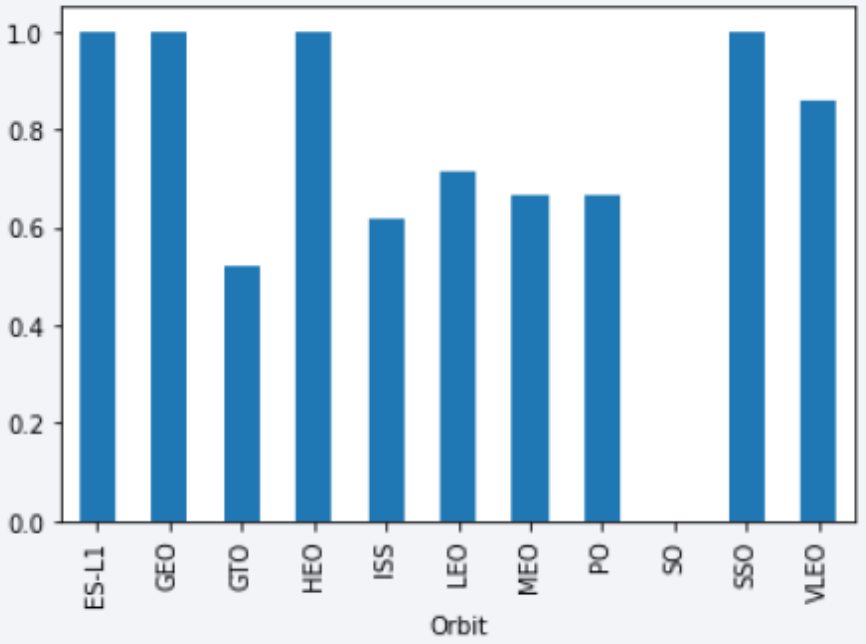
Payload vs. Launch Site



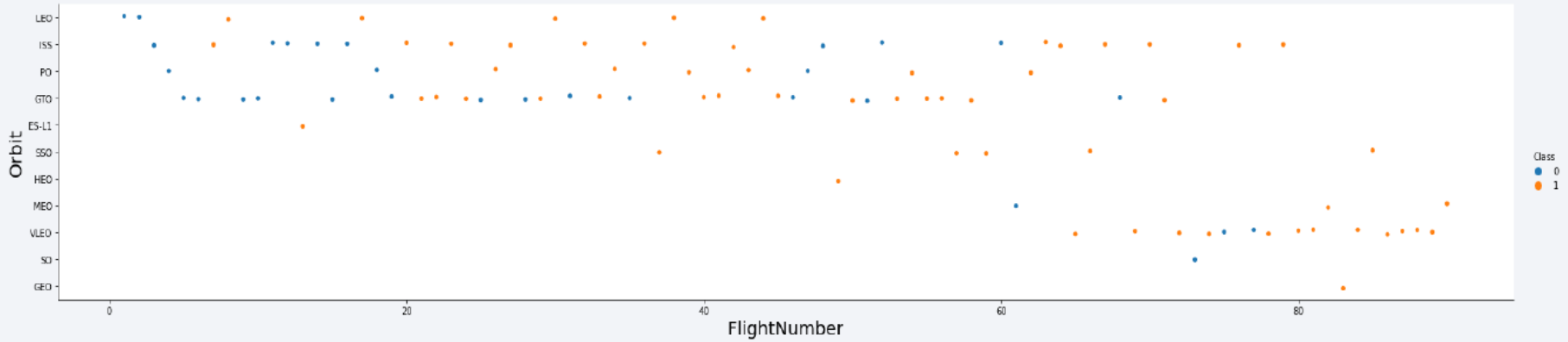
- After 9000Kg the successful rate increases exponentially
- Launches over 9000Kg are not possible on VAFB SLC 4E launch site.

Success Rate vs. Orbit Type

- The best orbits to launch a rocket are:
  - ES-L1
  - GEO
  - HEO
  - SSE

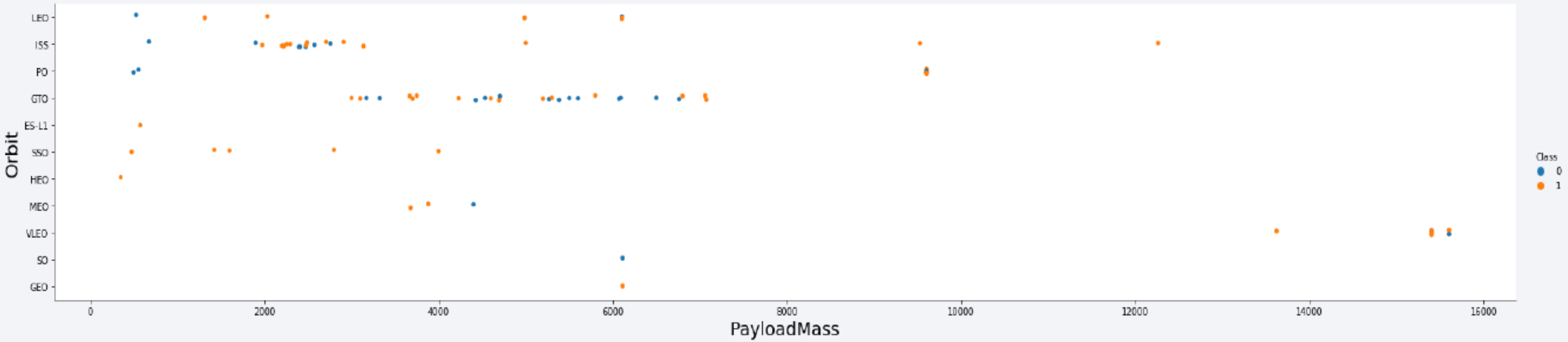


Flight Number vs. Orbit Type



- Apparently, the last orbits to be tested was GEO and SO

Payload vs. Orbit Type

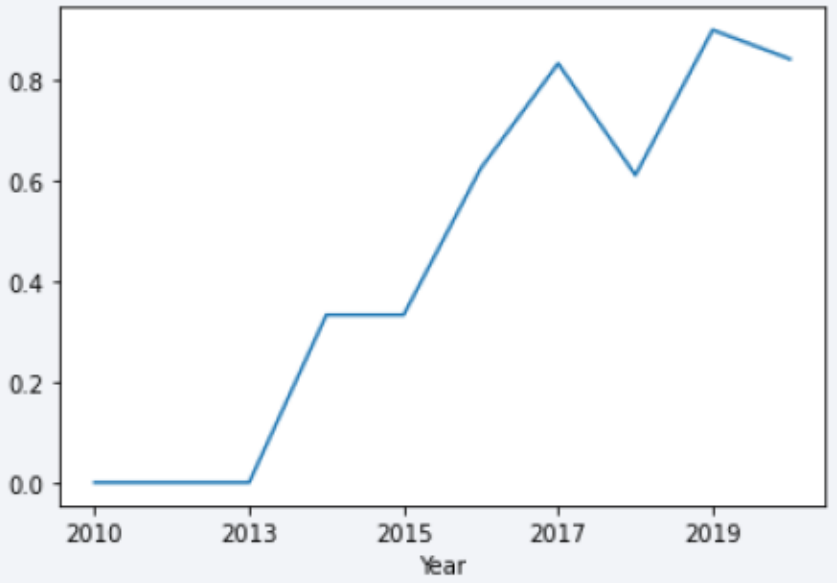


- Just 3 orbits (ISS, PO, VLEO) are used to do 'heavy' launches.

Launch Success Yearly Trend

- Success rate increased over time.

- There was a success rate of 0 between 2010 and 2013. That suggest that this period was of adjustments and preparation.



All Launch Site Names

- There are 4 launch sites

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

- Obtained by selecting unique values of 'launch_site' column in 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

- It shows only the launches that took place in Cape Canaveral (CCA)

Total Payload Mass

- Total payload carried by boosters from NASA

Total Payload (kg)
111.268

- It's kind of self-explanatory but this value represents the sum of all payloads from NASA

Average Payload Mass by F9 v1.1

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

Avg Payload (kg)
2.928

- In difference from the last slide, this value represents the average mass carried by Falcon 9 boosters and not the total sum of the mass.

First Successful Ground Landing Date

- Date of the first successful landing outcome on ground pad

Min Date
2015-12-22

- This is the first date to have a successful landing outcome recorded.

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

- There are four different booster names that successfully landed on drone ship and had mass between 4000Kg and 6000Kg.

Total Number of Successful and Failure Mission Outcomes

- Total number of successful and failure mission outcomes

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

- The summary is obtained by grouping outcomes and counting records.

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

- This are the names of the boosters that have carried the maximum payload mass registered in the dataset.

2015 Launch Records

- 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

- There are only 2 failures registered

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

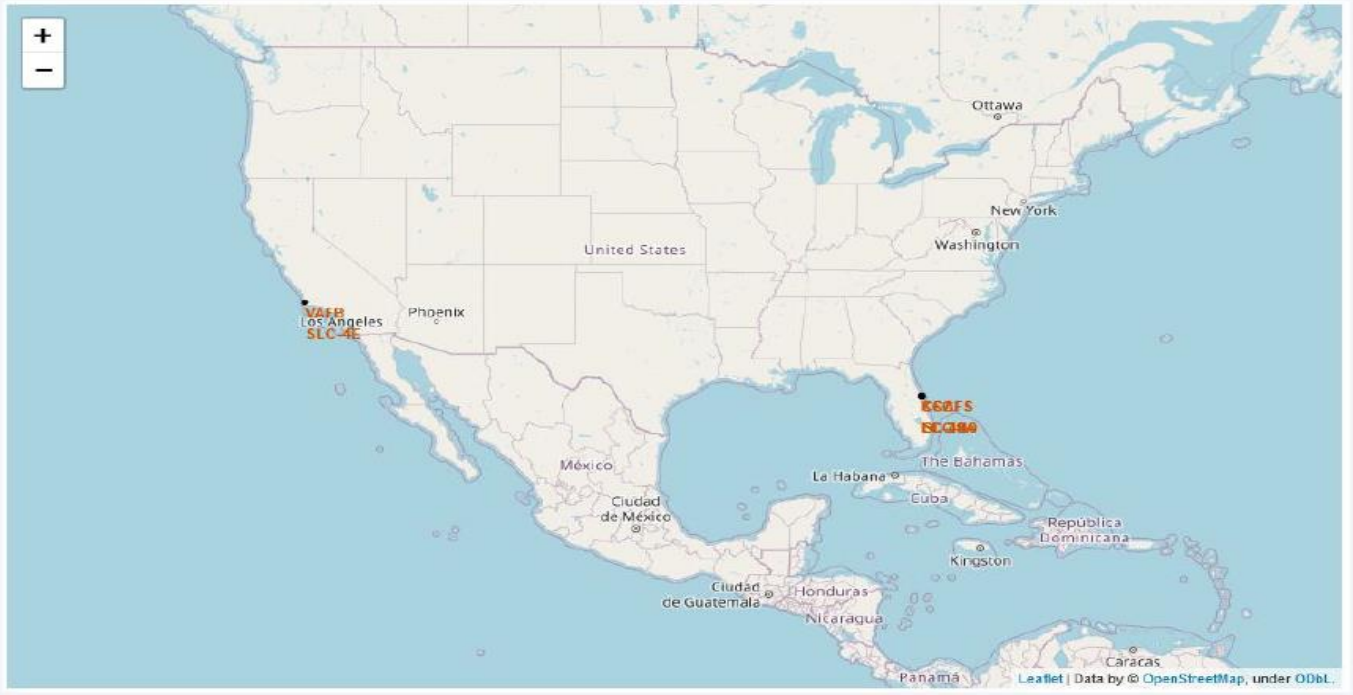
- The count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, ranked in descending order

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

- No attempt has a lot of occurrences that must be considered.


Section 3
Launch Sites
Proximities Analysis

All Launch Sites

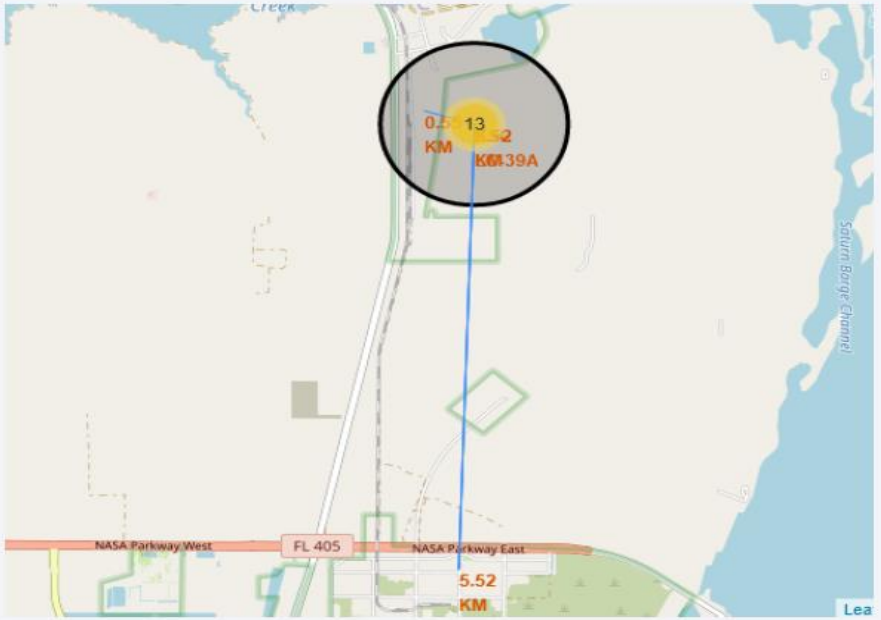


- Launch sites are near coast due to safety issues

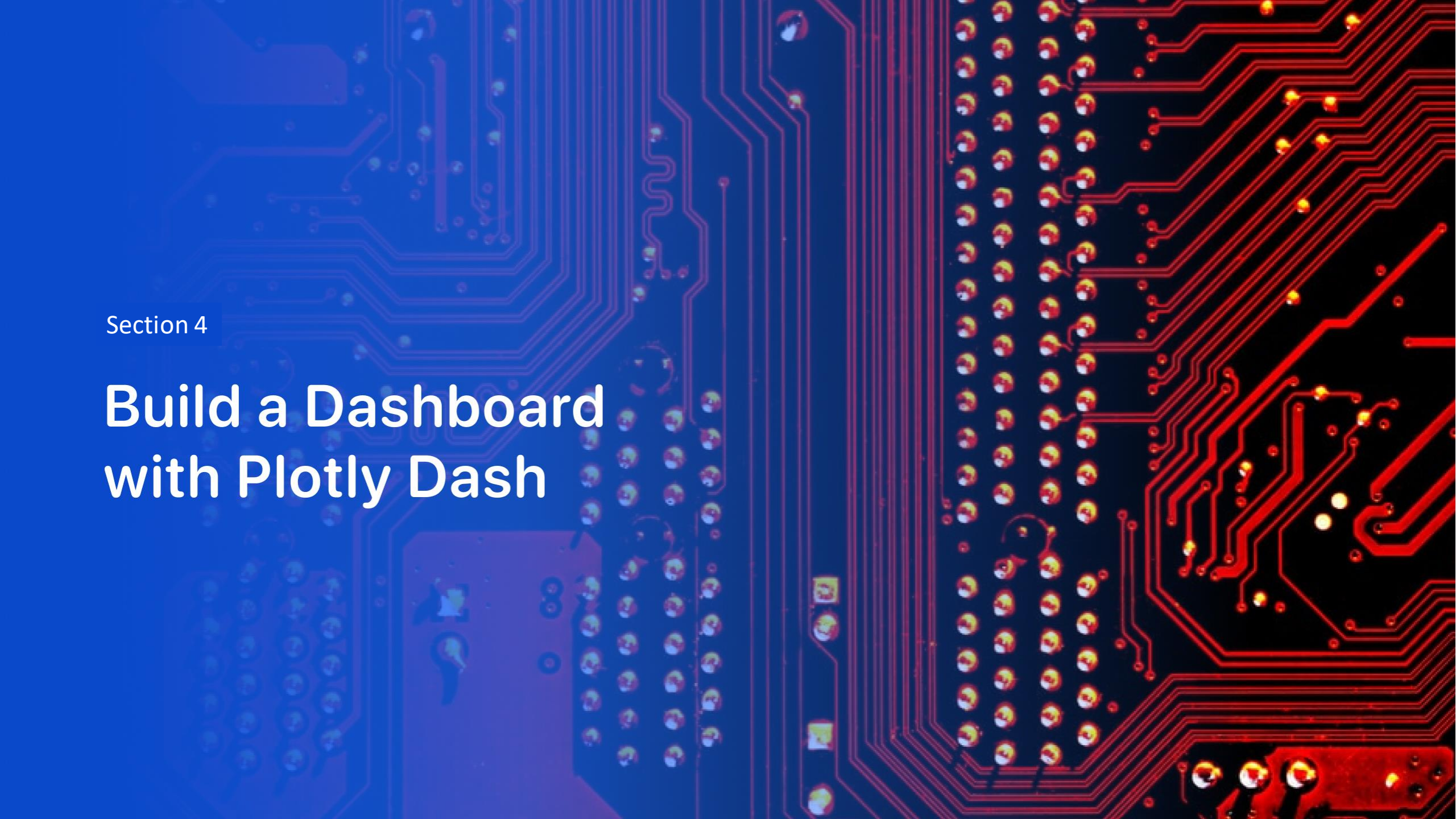
Launch Outcomes by Launch Site

- Green markers indicate success and red ones indicate failure.

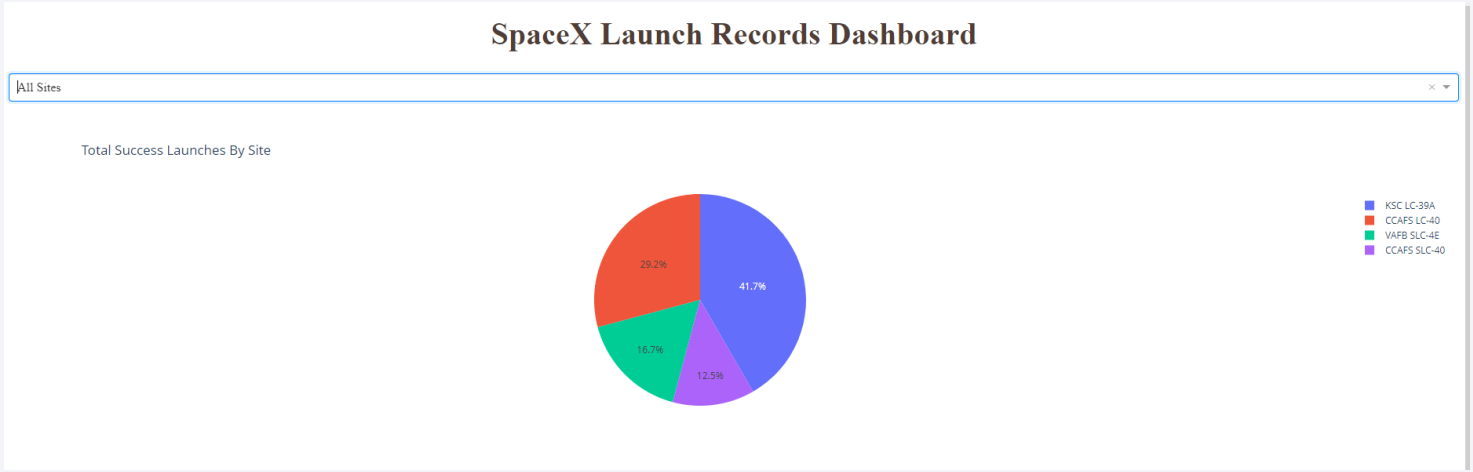
Logistics and Safety



- KSC LC-39A is very accessible but far enough to guarantee safety.

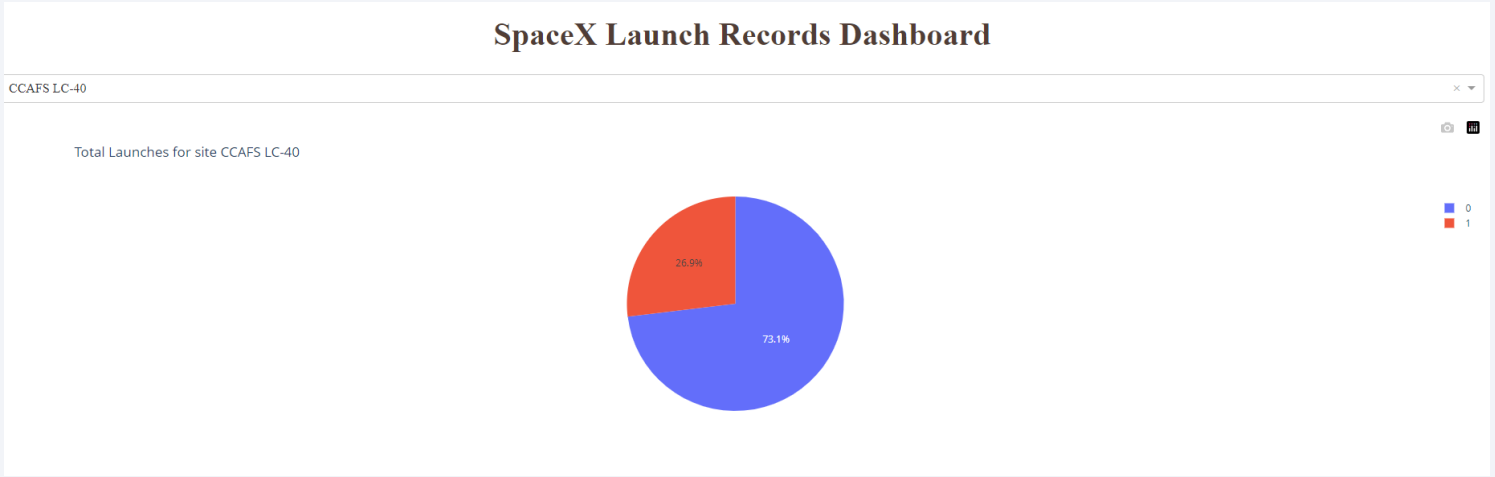

Section 4
Build a Dashboard with Plotly Dash

Successful Launches by Site



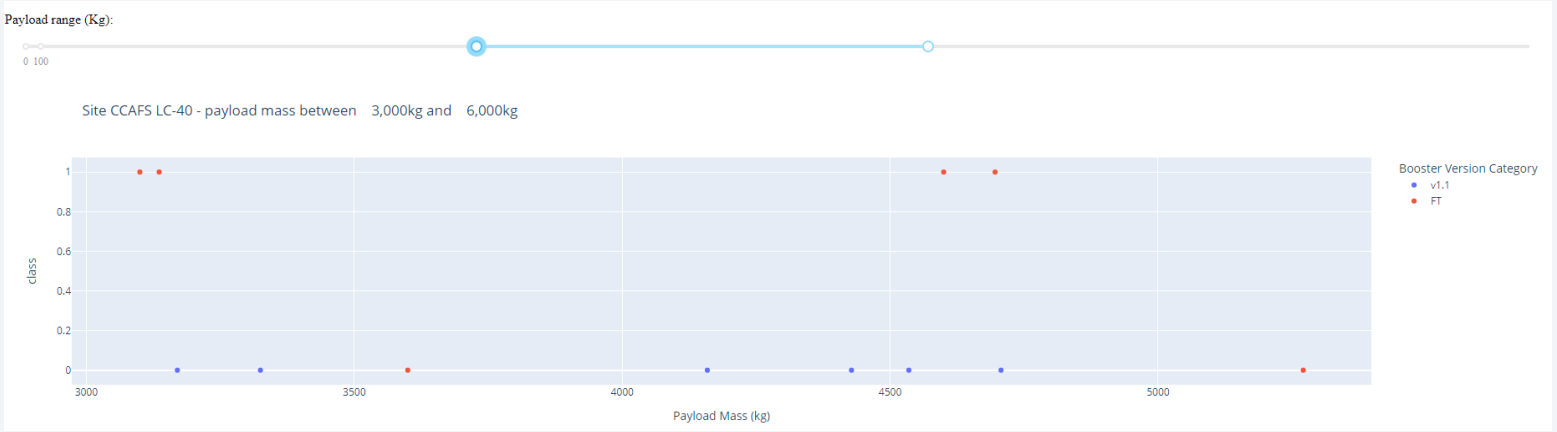
- The launch site has great influence on the outcome of the launch.

Launch Success Rate of KSC LC-39A



- KSC LC-39A has a success rate of 76.9%

Payload Mass vs Outcome



- It's difficult to determine whether the masses between 3000Kg and 6000Kg are 'good' or not because of the high variance.

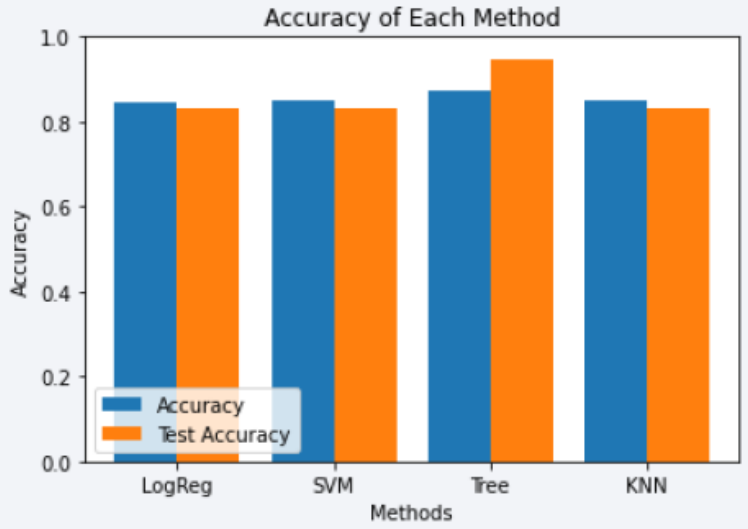
Section 5

Predictive Analysis (Classification)

Classification Accuracy

- Four Machine Learning models were built: Logit Model, SVM, KKN and Decision Tree

- The Decision Tree Classifier model has the accuracy over 87% and the accuracy on testing data of 94%



Confusion Matrix

- The confusion matrix of the Decision Tree model it's a prove of it's good behavior and 'fitness' to the problem.

Conclusions

- The best launch site is KSC LC-39A
- The heavier the mass the safer the launch.
- Even though successful rates are high they got even better over time.
- The Decision Tree Classifier Model was the best behaving model and the one that fits the best the data. It could be used to determine future data and improve profits.

Thank you!