Outline

- Executive Summary

- Introduction

- Methodology

- Results

- Conclusion

- Appendix

Executive Summary

- Methodologies used for data analysis:

  - Collection through SpaceX API

  - EDA, data wrangling, data visualization, interactive dashboard analytics

  - Machine Learning predictive analytics (supervised learning)

- Result Summary:

  - Machine learning prediction model had an accuracy of above 83%

  - EDA showed the right features for prediction of successful launches

Introduction

- The objective of this data science applied capstone is to see how good another company named Space Y is in terms of launch successes vs SpaceX

- Answers I looked for:

  - A reason why SpaceX does so well at a lower budget is because of successful landings, I want to estimate cost for launches through the success of rocket landings as an independent variable.

  - What geo-location ensures highest rate of launch landing success

Section 1

Methodology

Methodology

- Data collection methodology:

  - Obtained through SpaceX API

  - WebScraping Wikipedia

- Perform data wrangling

  - Cleaned a few columns and finalized some datatype changes for certain columns.

- Perform exploratory data analysis (EDA) using visualization and SQL

- Perform interactive visual analytics using Folium and Plotly Dash

- Perform predictive analysis using classification models

  - Normalized Data divided into training + test sets, evaluating with 4 different classification models, and finding the best one with > 87% accurate prediction.

6

Data Collection

- Data was collected through SpaceX API (https://api.spacexdata.com/v4/rockets/)

- Data was collected through Wikipedia WebScraping
  (https://en.Wikipedia.org/wiki/List of falcon/ 9/ and Falcon Heavy Launches).
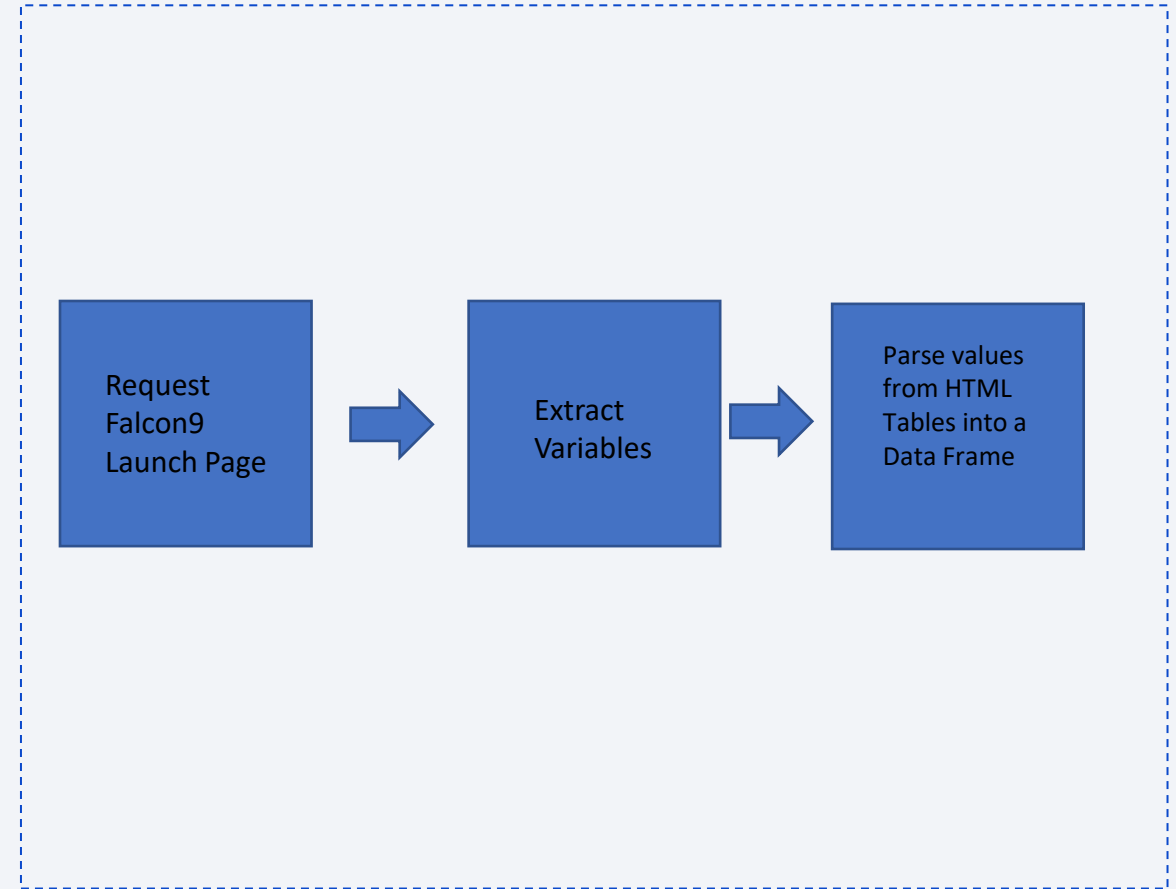
Data Collection – SpaceX API

- Please Refer to flowchart on right for data Collection process:

- Source:
  https://github.com/mgmike123/finalcapstone/blob/main/dataccollectionAPI.ipynb
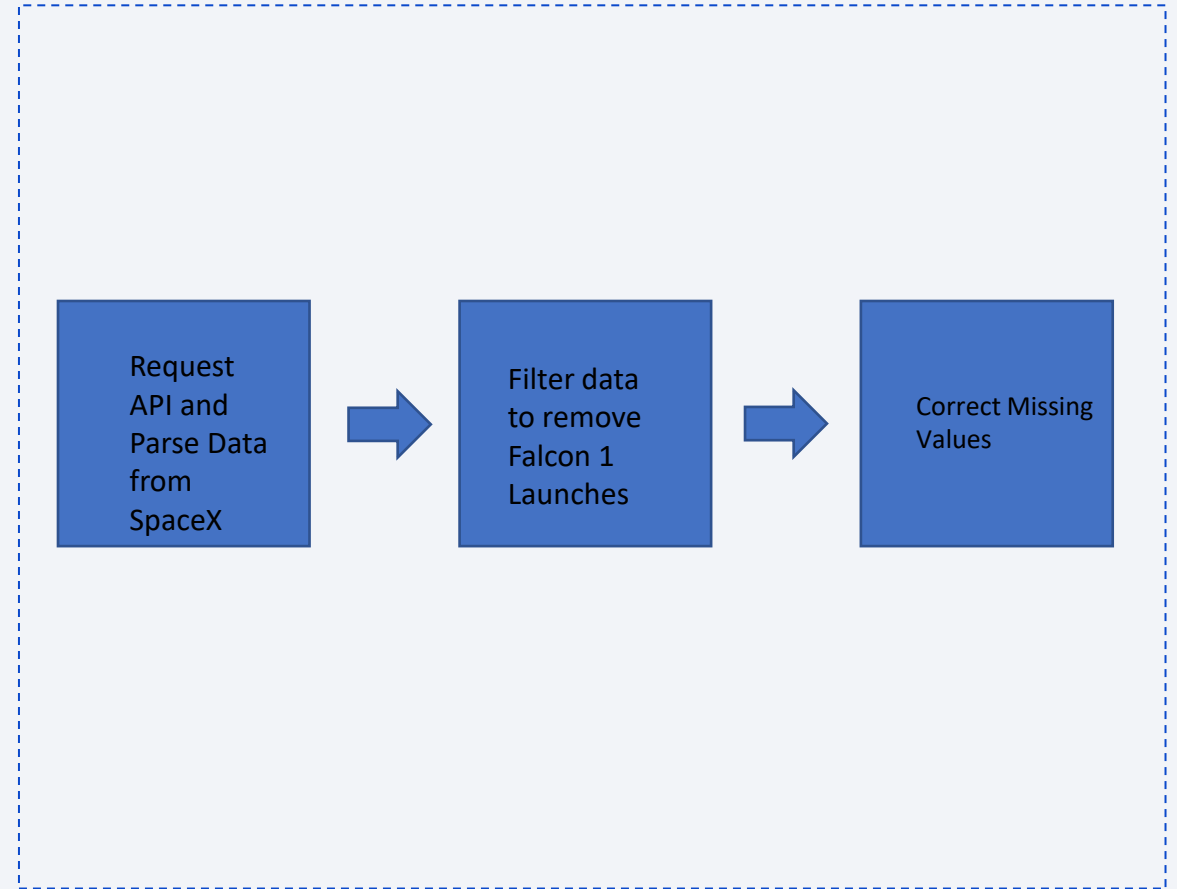
Data Collection - Scraping

- Please Refer to flowchart on right for data Collection process:

- Source: https://github.com/mgmike1 23/finalcapstone/blob/main/ WebScraping.ipynb

```
┌─────────────────────────────────────────────────────────┐
│                                                          │
│   ┌──────────┐       ┌──────────┐       ┌──────────┐     │
│   │ Request  │       │          │       │ Parse    │     │
│   │ Falcon9  │  ──▶  │ Extract  │  ──▶  │ values   │     │
│   │ Launch   │       │ Variables│       │ from HTML│     │
│   │ Page     │       │          │       │ Tables   │     │
│   │          │       │          │       │ into a   │     │
│   │          │       │          │       │ Data Frame│    │
│   └──────────┘       └──────────┘       └──────────┘     │
│                                                          │
└─────────────────────────────────────────────────────────┘
```
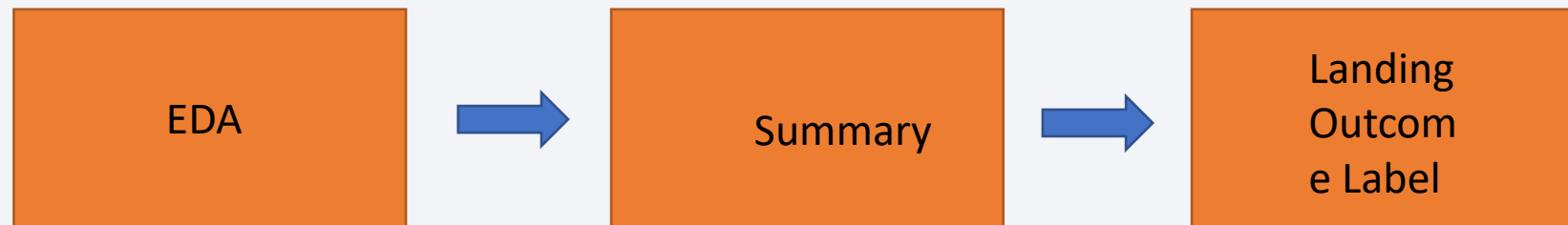
Data Wrangling

- EDA(Exploratory Data Analysis) was performed

- Please refer to the following flowchart

Source:
https://github.com/mgmike123/finalcapstone/blob/main/dataWrangling.ipynb

```
EDA  →  Summary  →  Landing Outcome Label
```

EDA with Data Visualization

- Scatterplots and Bar charts as well as Bar plots were the most used visuals for this lab as I had decided that they did their best job respectively to display the necessary insights on what data was being assessed.

- Source: https://github.com/mgmike123/finalcapstone/blob/main/EDAwithvisualization.ipynb

EDA with SQL

- What I did SQL Query Wise:

  - Names of unique launch sites

  - Top 5 launch sites beginning with 'CCA'

  - Total payload mass on boosters launched by NASA CRS

  - Average payload mass carried by booster F9 1.1

  - Date of first successful landing

  - Names of boosters with success in droning AND payload mass = Between 4000-6000kg

  - Total # of successful missions

  - Total # of failure missions

  - Booster names with highest payload mass

  - Failed landing outcomes for 2015

  - Rank of landing outcomes between 2010-2017

Build an Interactive Map with Folium

- Markers/Circles/Lines/Marker Clusters used in Folium Maps:

    - Used because:

        - Markers: Good for launch site indication

        - Circles: Good for displaying NASA Johnson Space Center

        - Lines: Calculating distance between coordinates on a map

        - Marker Clusters: Groups of events occurring in each coordinate found.

    Source: https://github.com/mgmike123/finalcapstone/blob/main/folium.ipynb

Build a Dashboard with Plotly Dash

- Graphs/Plots used in my dashboard:

  - Payload Range

  - Launches by site

- These were the best pair of variables to use to find best places for launches to occur

- Source: https://github.com/mgmike123/finalcapstone/blob/main/spacexdashapp.py

Predictive Analysis (Classification)

- Classification is a supervised machine learning method:

  - I used the four methods of classifications taught: Regression, SVM, Decision Tree, KNN

    - Please refer to the flow chart below:

Data Prep + Normalization	→	Find best parameters by testing models	→	Compare accuracy for each model

Source:

https://github.com/mgmike123/finalcapstone/blob/main/Machine%25Learning.ipynb
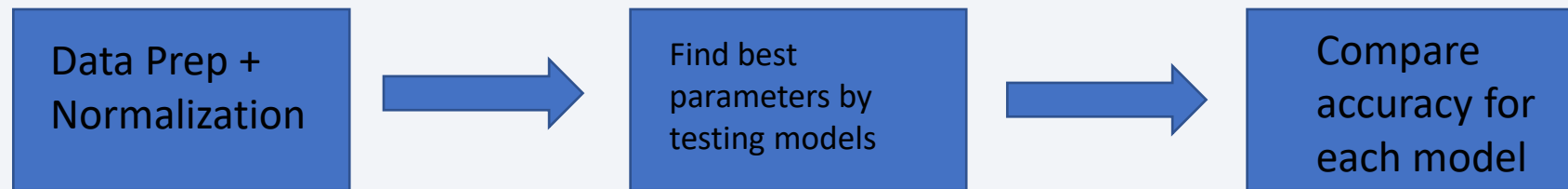
Results

- Data analysis results:

  - 4 different launch sites located

  - First launch was done to spaceX and NASA

  - First success landing occurred in 2015, 5 years after first launch

  - Falcon 9 booster versions still had good landing when above average payload having

  - Near 100% missions' outcomes were classified as successful

  - Two booster versions failed landing n 2015, f9 B1012 and f9 B1015

  - Linear increase in successful landing outcomes overtime

Results

- According to interactive analytics, most launches occurred on east coast launch sites as shown below:
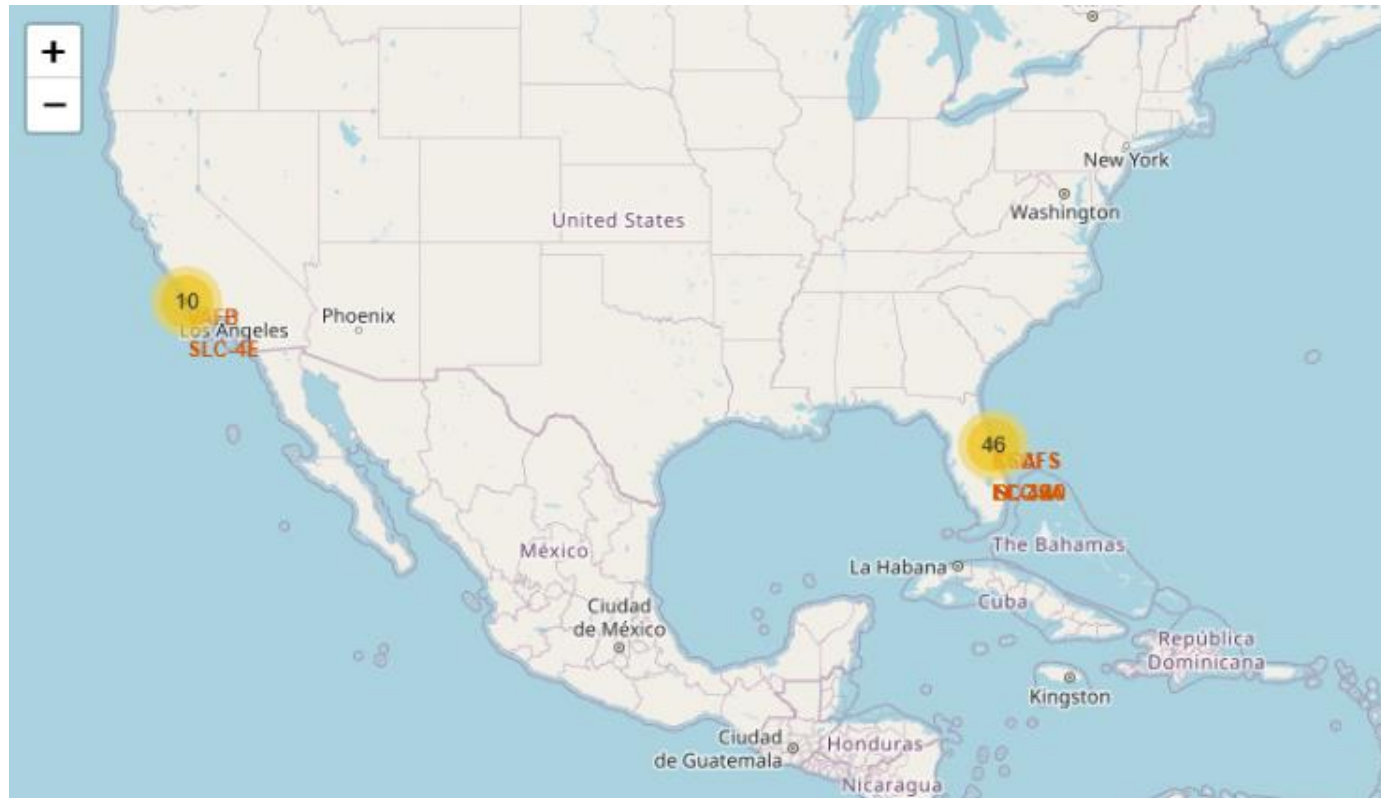
Results

- Decision tree classification was best for prediction of successful landings:
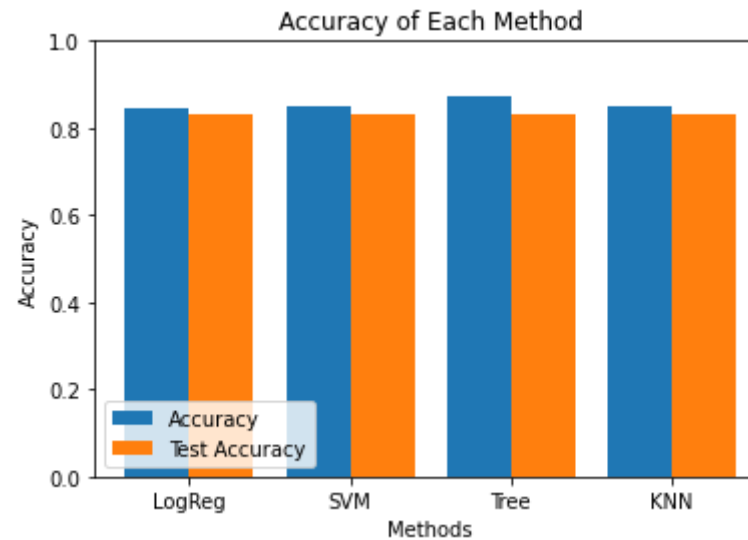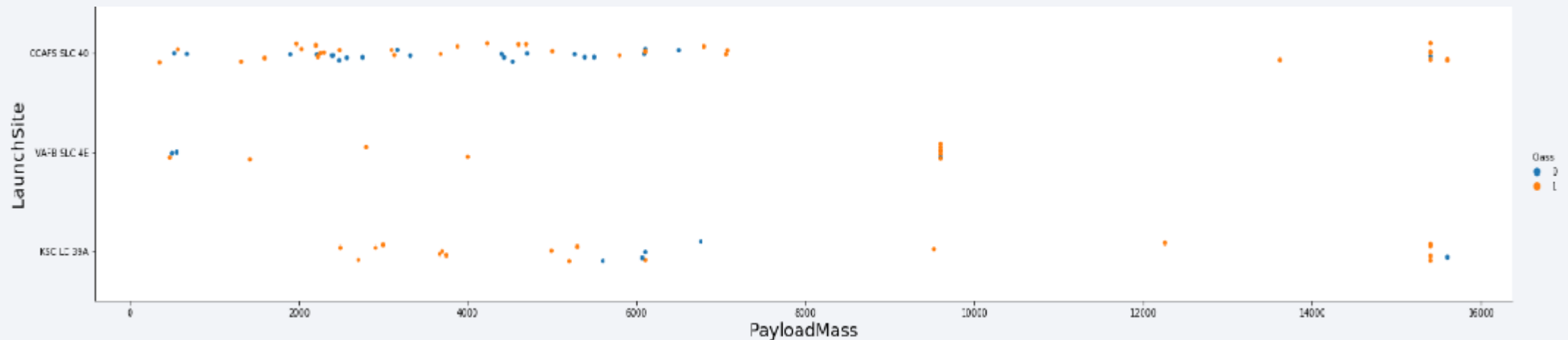
Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

- As shown below on my scatter plot: safe conclusion is that best launch site is CCAF5 SLC40

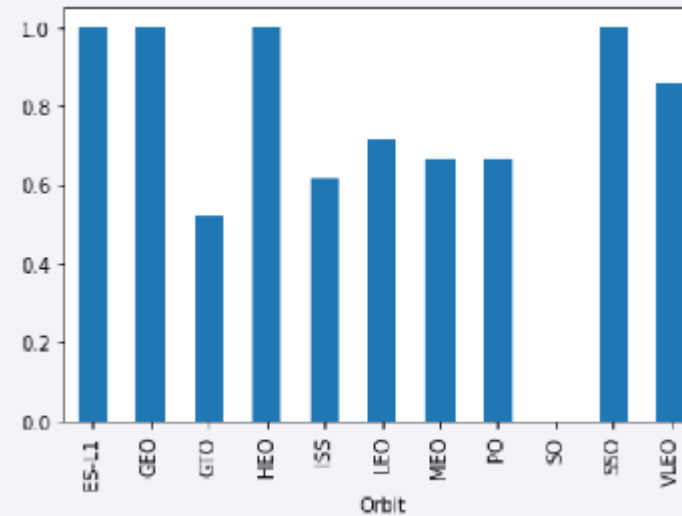- Overtime improvement in success rate is also displayed
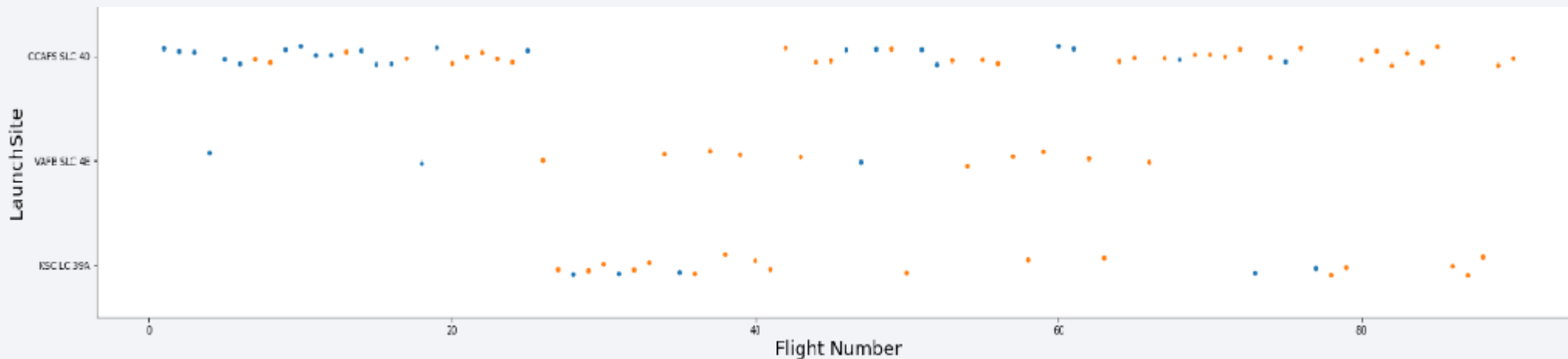
Payload vs. Launch Site

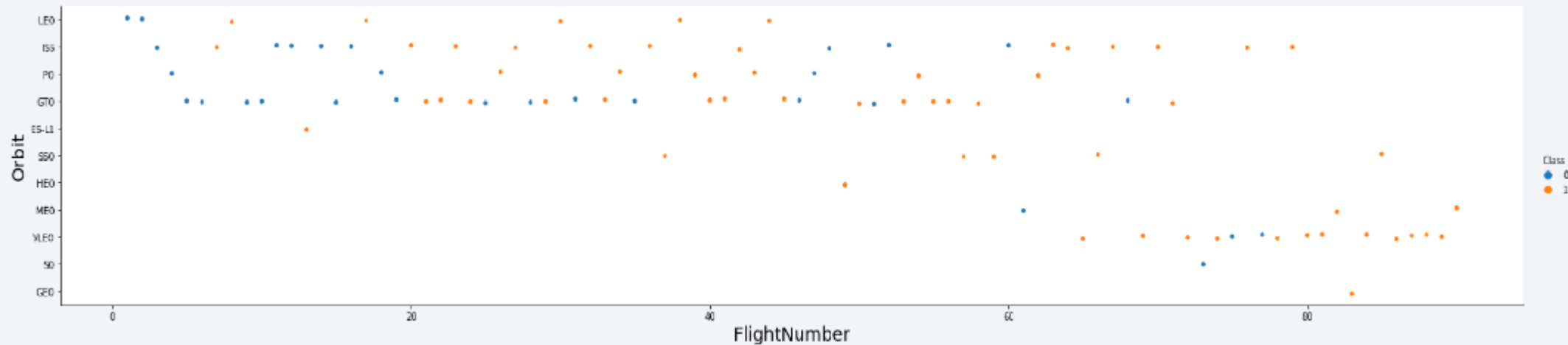- Good success rate = payloads > 9000 kg

Success Rate vs. Orbit Type

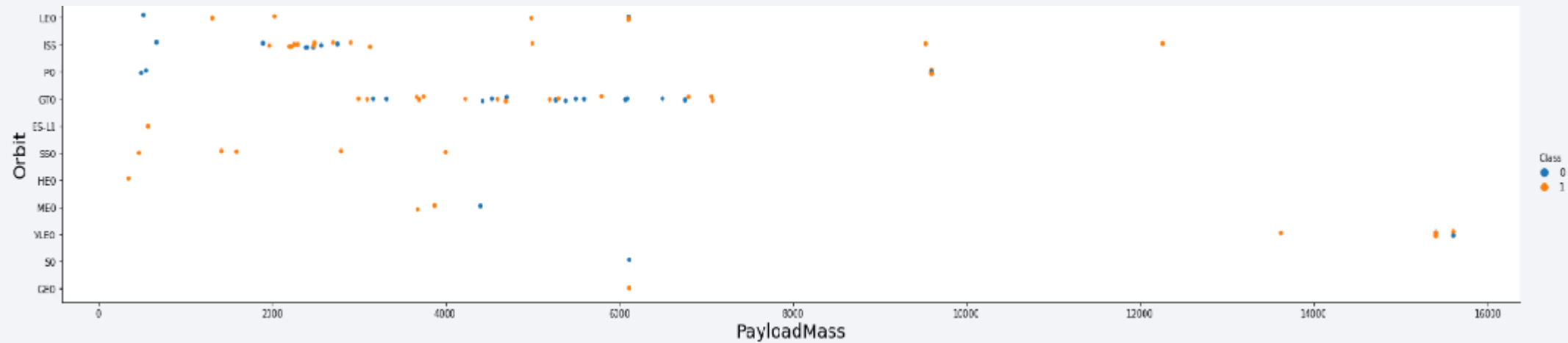- Best success rate: ES-L1

- Worst success rate: GTO

Flight Number vs. Orbit Type

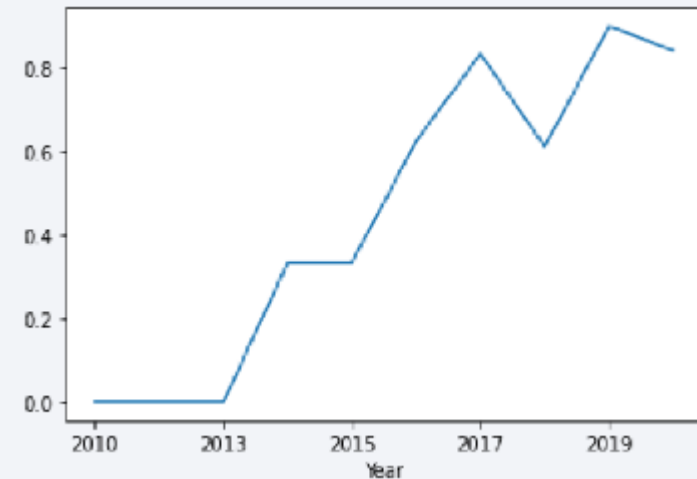- Linear increase overtime in success rates across all orbit variables

Payload vs. Orbit Type

- Hard to find connection between two axis, however ISS seems promising

Launch Success Yearly Trend

- Huge increase in success rate from 2013 to 2017.

All Launch Site Names

- All launch site names:

  - CCAFS LC-40

  - CCAFS SLC-40

  - KSC LC-39A

  - VAFB SLC-4E

  - Obtained through querying unique rows with launch site values from dataset

Launch Site Names Begin with 'CCA'

- As shown below, we see the CCA tag in bold in the dataset.

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

Total Payload Mass

- Total payload carried by boosters from NASA:

  - 111,268kg

  - Calculated using simple mathematics summing all payloads with CRS in their code respectively.

Average Payload Mass by F9 v1.1

- Avg payload mass carried by booster F9 v1.1:

  - 2,928kg

  - Data acquired through the filtering of booster versions and using math to calculate the average.

First Successful Ground Landing Date

- First successful landing outcome:

  - December 22nd, 2015

  - To get this data I filtered the data set by successful landing and queried the minimum possible value on the date column

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

    - F9 BT B1021.2

    - F9 FT B1031.2

    - F9 FT B1022

    - F9 FT B1026

    - Selecting distinct booster versions and using a between statement was used to query this information

Total Number of Successful and Failure Mission Outcomes

- Ratio of successful missions to failed missions:

  - 99 occurrences to 1 occurrence

  - Groupby statement used on mission outcomes lead to the obtainability of this dataset

Boosters Carried Maximum Payload

- Boosters that carried max payload mass:

    - F9 B5 B1048.4 AND .5

    - F9 B5 B1049.4 AND .5 AND .7

    - F9 B5 B1051.3 AND .4 AND .6

    - F9 B5 B1056.4

    - F9 B5 B1058.3

    - F9 B5 B1060.2 AND .3

2015 Launch Records

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

  - F9 v1.1 B1012 on CCAFS LC-40

  - F9 v1.1 b1015 on CCAFS LC-40

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

- Please examine the table on the right hand side for this information.

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 Site Locations
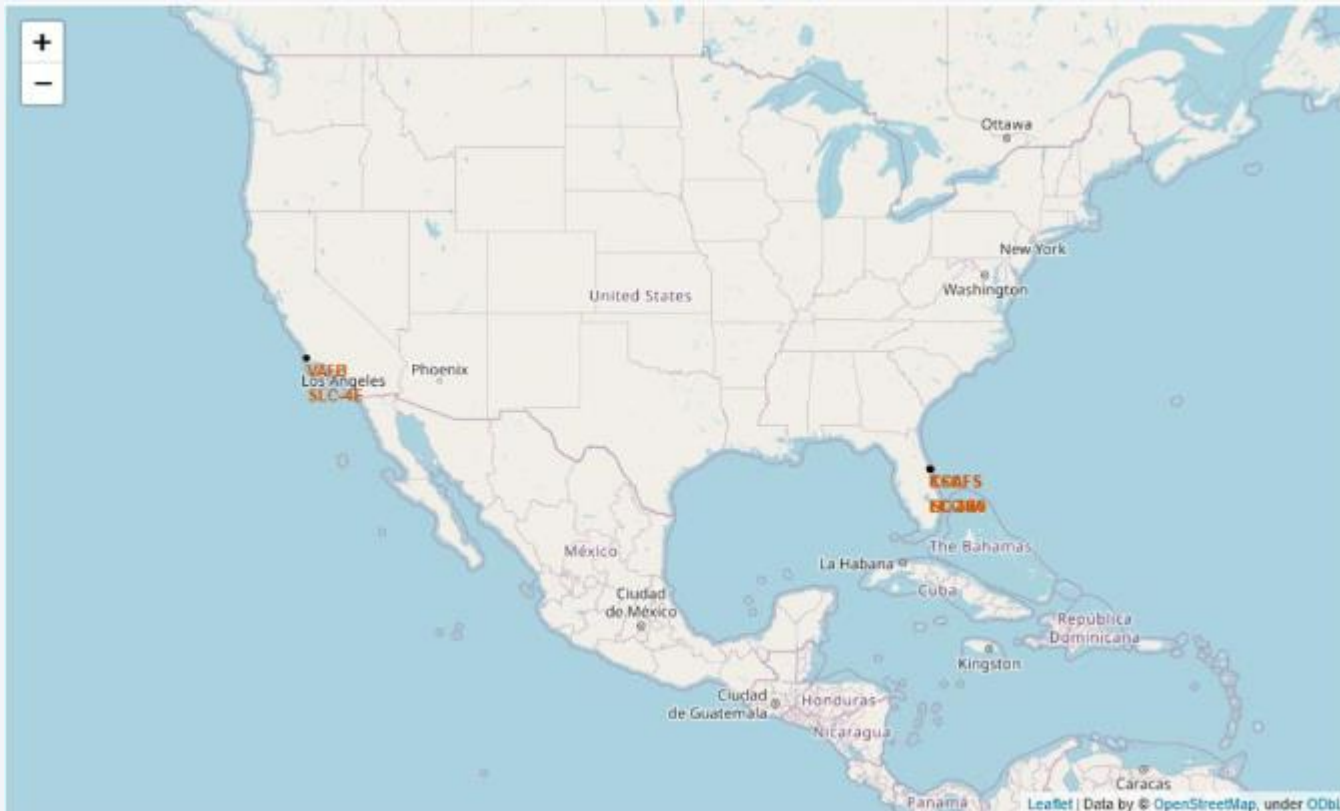
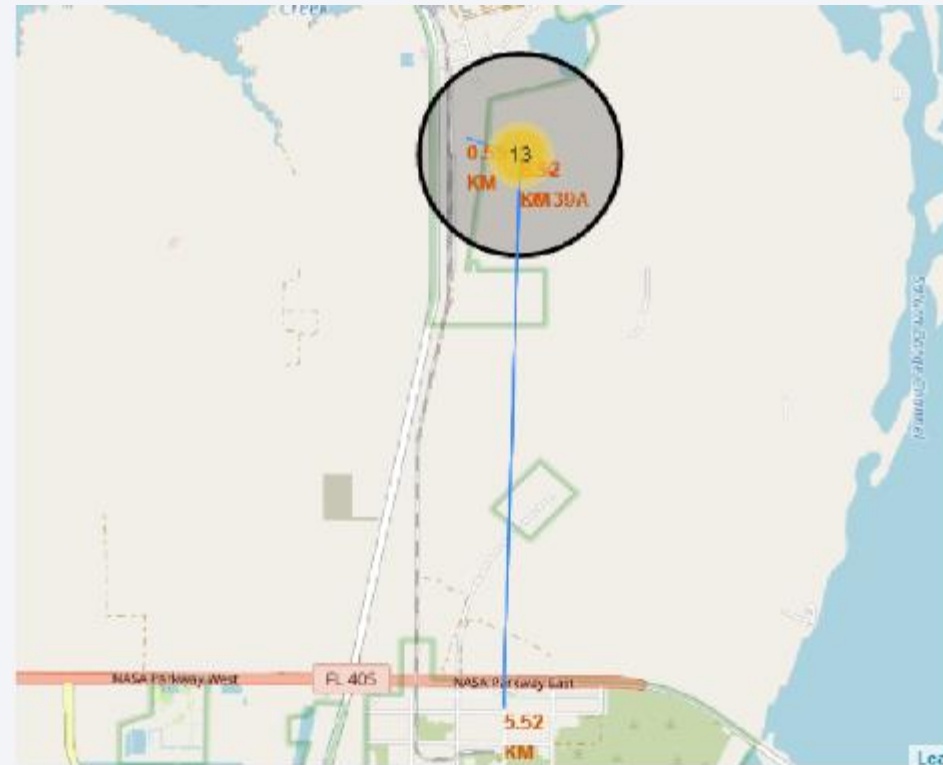- Most launch sites were located on east coast

Launch Outcomes by Site

- Green markers label successful attempts, and red markers label failed attempts.

Safety of launch sites by GeoLocation

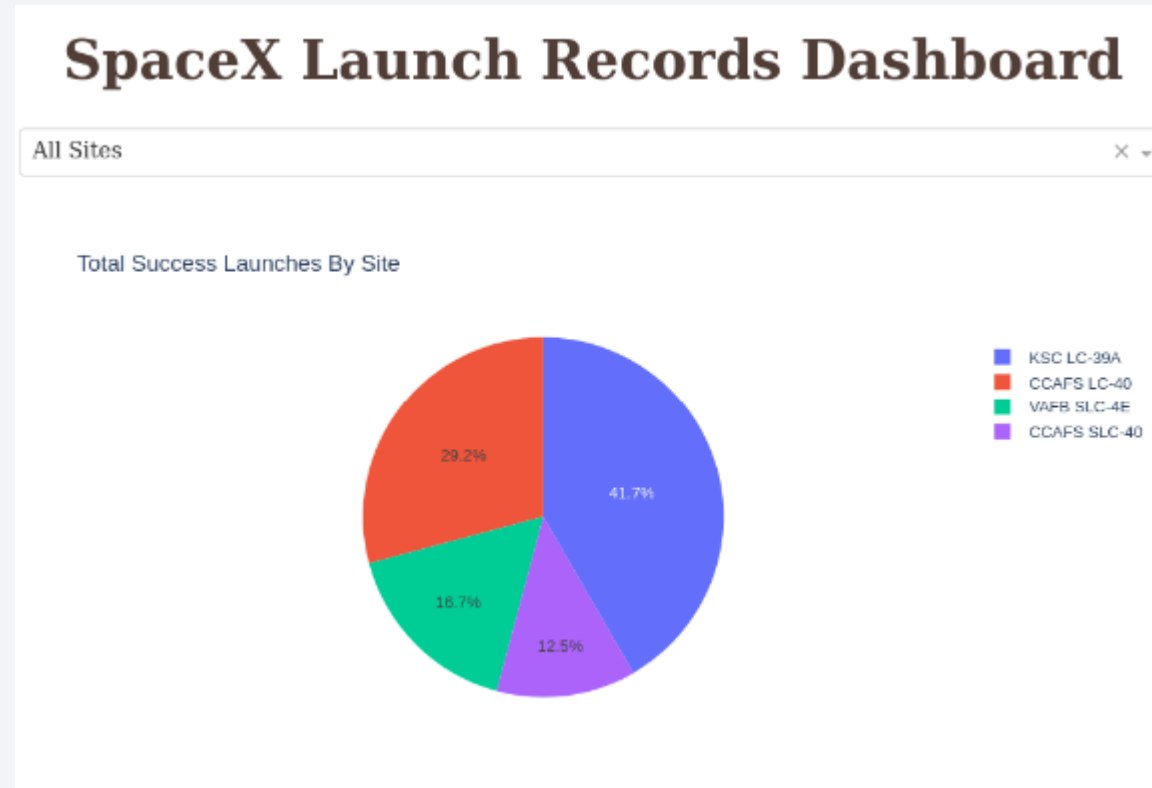- As shown on the right, launch site KSC LC-39A being 5.52 km from pathway.

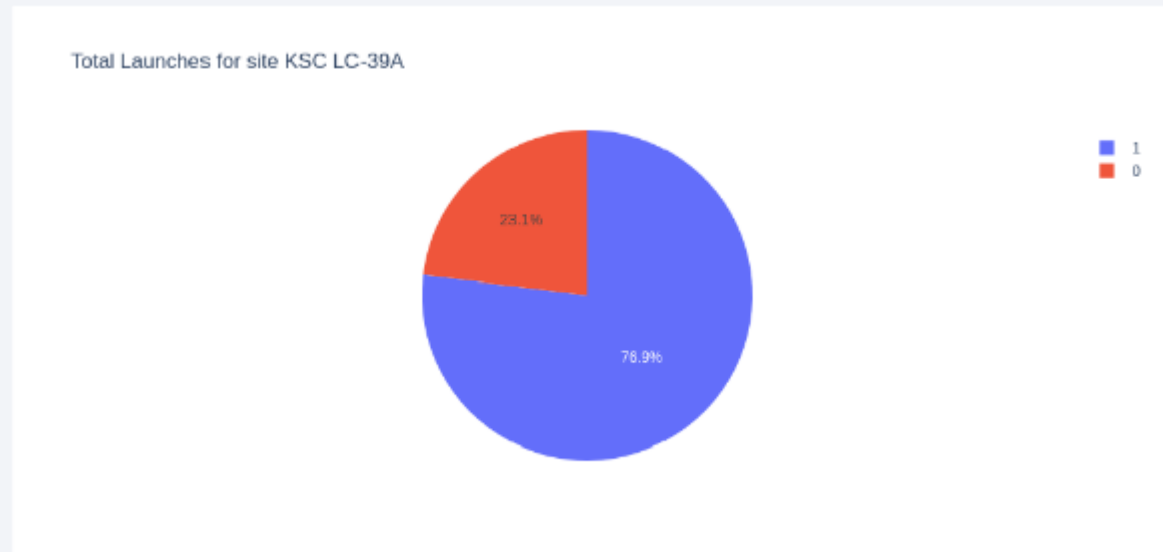Section 4

Build a Dashboard
with Plotly Dash

Successful launch count per site

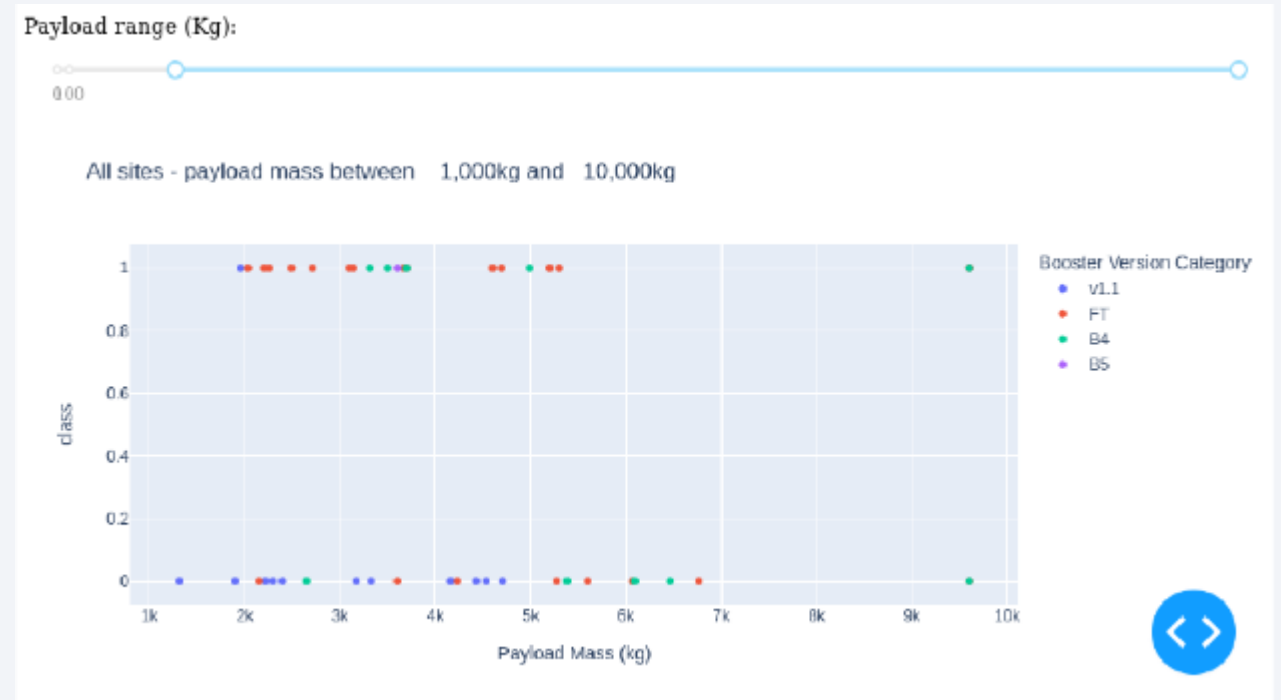- As shown on pie chart: 41.7% of successful launches occurred on KSC LC-39A

Diving deeper into KSC LC-39A's success

- KSC LC-39A represents a 76.9% success rate

Total Launches for site KSC LC-39A

Payload vs Launch Outcome for all sites

- Highest success combination:

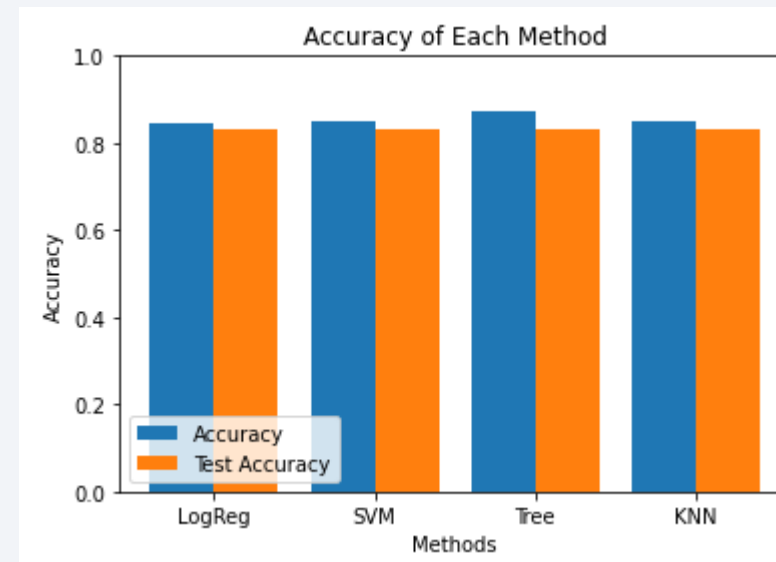 - Payloads < 6000kg

  - Booster version set to: FT

Section 5

Predictive Analysis (Classification)

Classification Accuracy

- As seen by the orange bars on the right, the highest reaching one was the decision tree model, reaching above 83%.

Confusion Matrix

- Confusion matrix performed for decision tree model showed high levels of true positives and true negatives.

Conclusions

- The best launch site was found to be KSC LC-39A

- Launches above 6000-7000kg are safer

- Decision Tree classifier is a safe classification method to use for this analysis

Appendix

- Had some issues on github for some reason, so everything that I was not able to upload there, I have uploaded here as screenshots ☺

Thank you!