



IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Aditya Challa



Outline

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

Executive Summary

- The following methodologies were used to analyze data:
 - Data Collection using web scraping and SpaceX API;
 - Exploratory Data Analysis (EDA), including data wrangling, data visualization and interactive visual analytics;
 - Machine Learning Prediction.
- Summary of all results
 - It was possible to collect valuable data from public sources;
 - EDA allowed to identify which features are the best to predict success of launchings;
 - Machine Learning Prediction showed the best model to predict which characteristics are important to drive this opportunity by the best way, using all collected data.

Introduction

- The objective is to evaluate the viability of the new company Space Y to compete with Space X.
- Outcomes:
 - The best way to estimate the total cost for launches
 - The best place to make launches.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
 - Data from Space X was scraped from 2 sources:
 - Space X API (<https://api.spacexdata.com/v4/rockets/>)
 - WebScraping
(https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches)
- Perform data wrangling
 - Collected data was enriched by creating a landing outcome label based on outcome data after summarizing and analyzing features.
- 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
 - Data was normalized, divided in training and test data sets and evaluated by four different classification models, and the accuracy of each model evaluated.

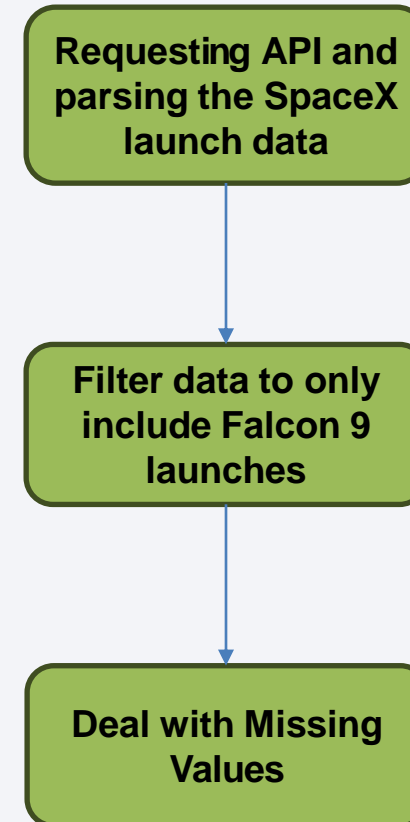
Data Collection

- Data sets were collected from Space X API (<https://api.spacexdata.com/v4/rockets/>) and from Wikipedia (https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches), using web scraping technics.

Data Collection – SpaceX API

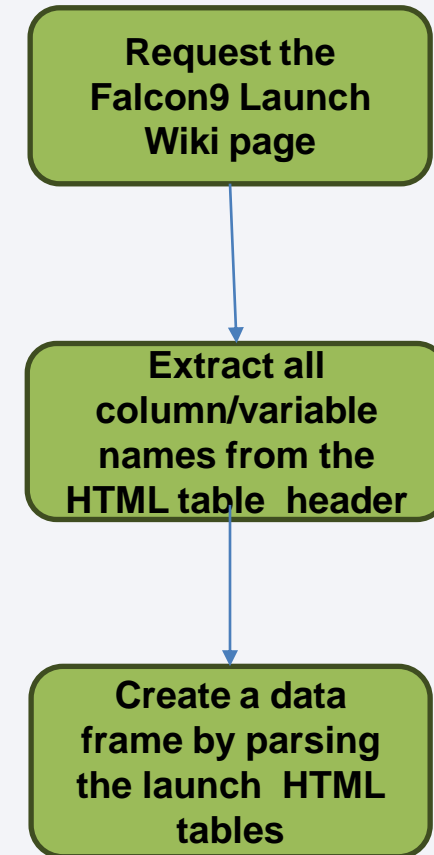
- SpaceX offers a public API from where data can be obtained and then used;
- This API was used according to the flowchart beside.

- Source code: [capstone_coursera/Spacex data collection api.ipynb at main · challa-srmap/capstone_coursera \(github.com\)](#)



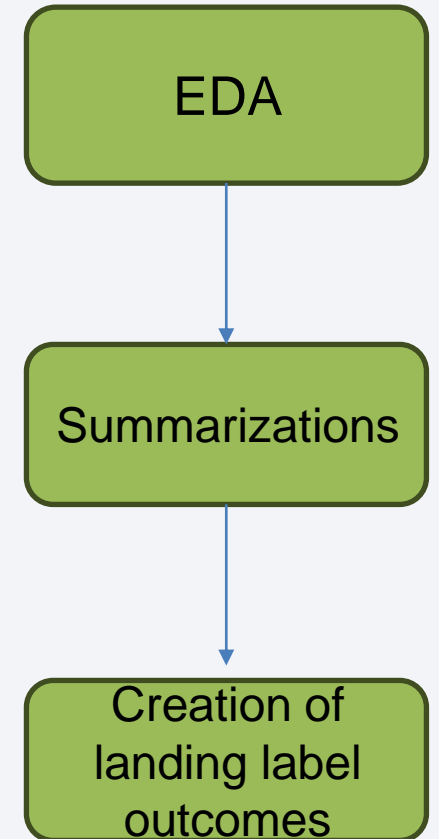
Data Collection - Scraping

- Data from SpaceX launches can also be obtained from Wikipedia;
- Data are downloaded from Wikipedia according to the flowchart.



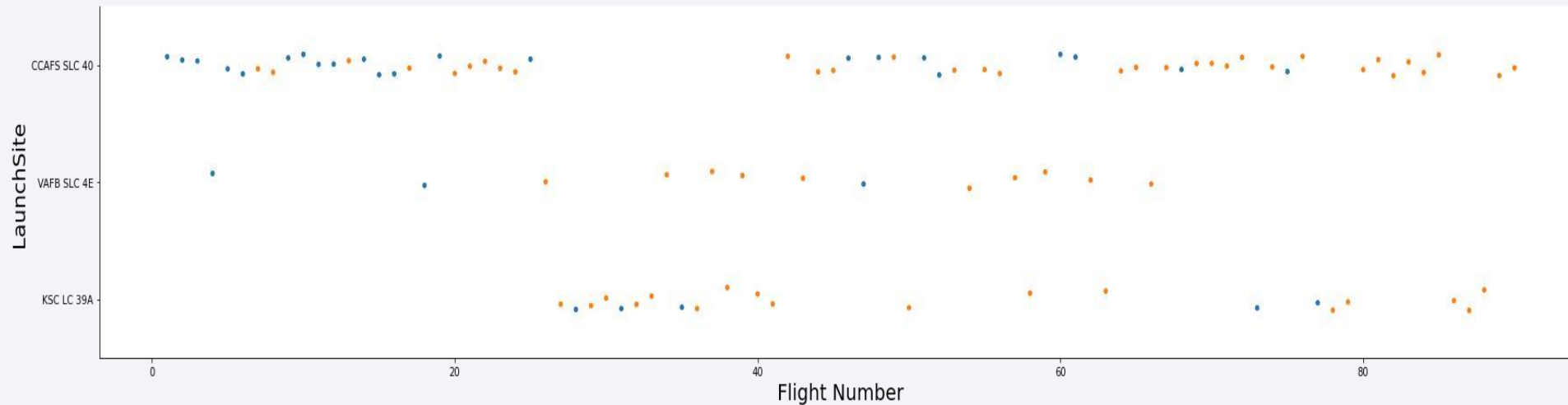
Data Wrangling

- Exploratory Data Analysis (EDA) was performed on the dataset.
- Launches per site, occurrences of each orbit and occurrences of mission outcome per orbit type were calculated.
- Finally, the landing outcome label was created from Outcome column.
- Source code: [capstone_coursera/Data wrangling.ipynb at main · challa-srmap/capstone_coursera \(github.com\)](#)



EDA with Data Visualization

- To explore data, scatterplots and barplots 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: [capstone_coursera/EDA with python.ipynb](https://github.com/challa-srmap/capstone_coursera) at main · challa-srmap/capstone_coursera (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 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; and
 - Rank of the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20.
- Source code: [capstone_coursera/EDA with sql.ipynb at main · challa-srmap/capstone_coursera \(github.com\)](#)

Build an Interactive Map with Folium

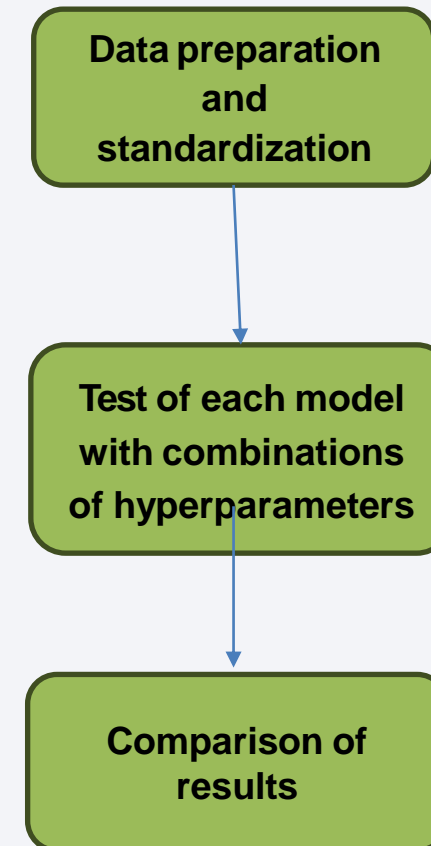
- Markers, circles, lines and marker clusters were used with Folium Maps
 - Markers indicate launch sites;
 - Circles indicate highlighted areas around specific coordinates, in this case like the NASA Johnson Space Center;
 - Marker clusters indicates groups of events in each coordinate, in this case like the launches in a launch site;
 - Lines are used to indicate distances between two coordinates.
- Source code: [capstone_coursera/Launch site location with folium.ipynb at main · challa-srmap/capstone_coursera \(github.com\)](#)

Build a Dashboard with Plotly Dash

- The following graphs and plots were used to visualize data
 - Percentage of launches by site
 - Payload range
- This combination allowed to quickly analyze the relation between payloads and launch sites, helping to identify where is best place to launch according to payloads.
- Source code: [capstone_coursera/spacex_dash_app.py at main · chall-srmap/capstone_coursera \(github.com\)](https://github.com/chall-srmap/capstone_coursera/blob/main/dash_app.py)

Predictive Analysis (Classification)

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



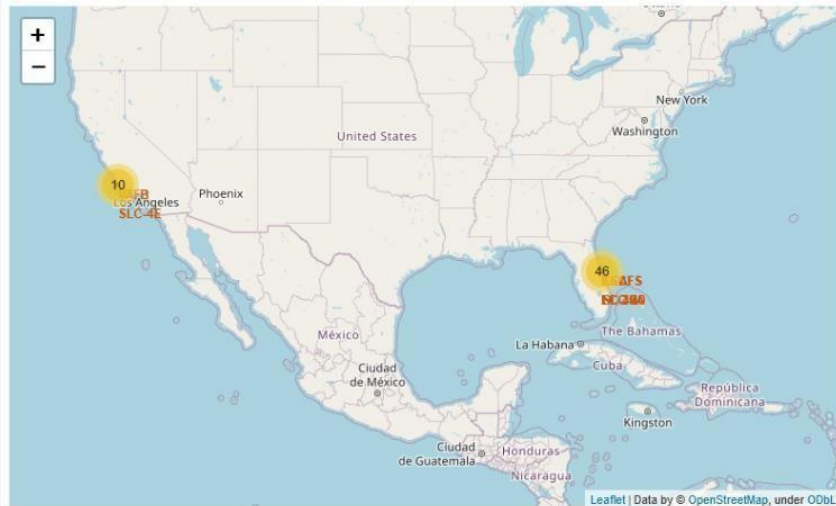
- Source code: [capstone_coursera/Machine Learning Prediction.ipynb](https://github.com/challa-srmap/capstone_coursera) at main · challa-srmap/capstone_coursera (github.com)

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 2,928 kg;
 - The first success landing outcome happened in 2015 five year after the first launch;
 - Many Falcon 9 booster versions were successful at landing in drone ships having payload above the average;
 - 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.

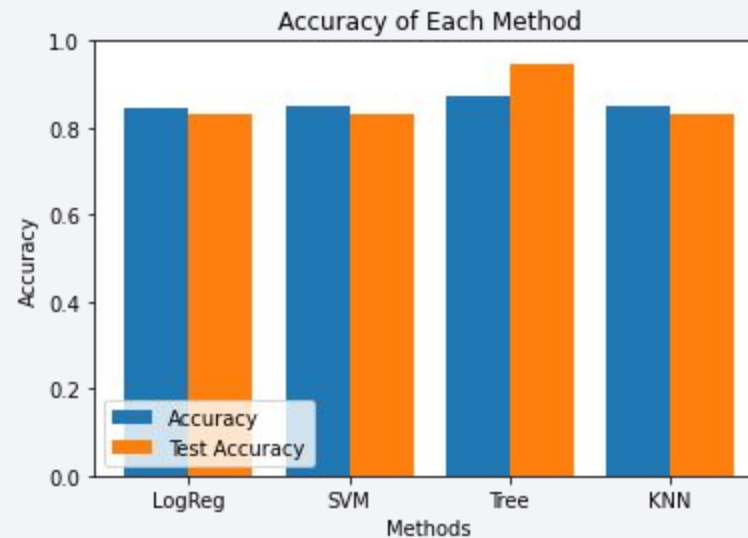
Results

- Using interactive analytics was possible to identify that launch sites use to be in safety places, near sea, for example and have a good logistic infrastructure around.
- Most launches happens at east cost launch sites.



Results

- Predictive Analysis showed that Decision Tree Classifier is the best model to predict successful landings, having accuracy over 87% and accuracy for test data over 94%.

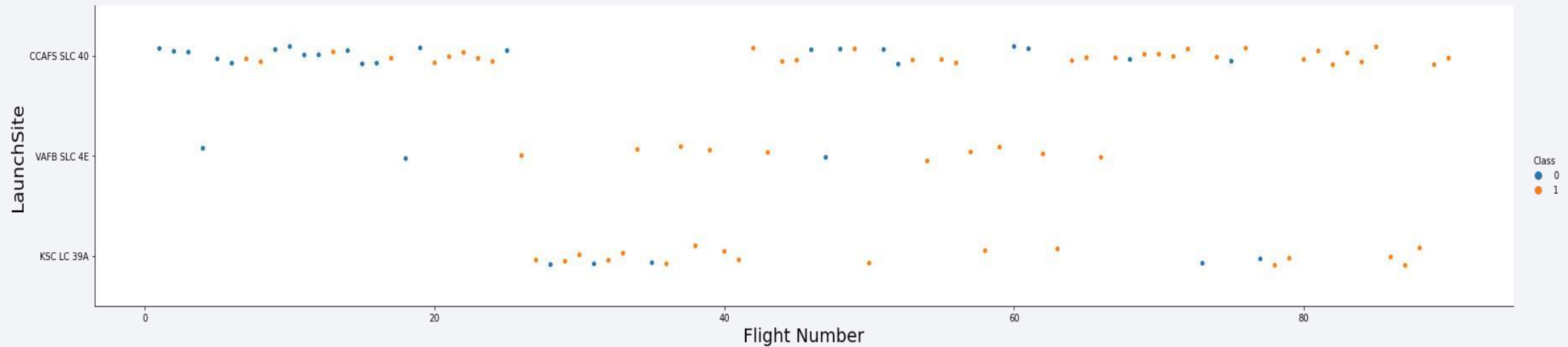


The background of the slide is an abstract composition. It features a solid blue area on the left side, which transitions into a dynamic pattern of diagonal streaks in shades of blue, red, and teal on the right. These streaks are layered over a faint, grid-like pattern, creating a sense of depth and movement.

Section 2

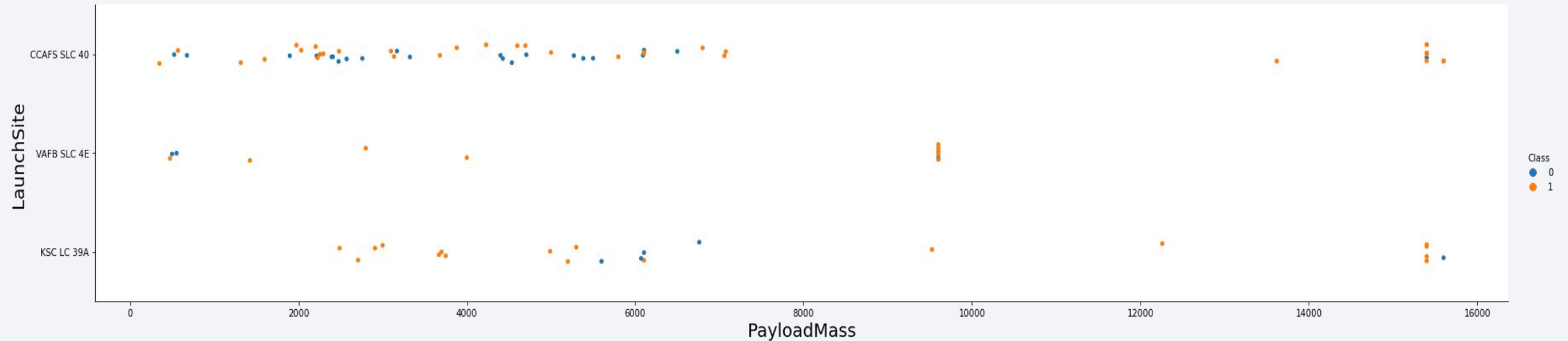
Insights drawn from EDA

Flight Number vs. Launch Site



- According to the plot above, it's possible to verify that the best launch site nowadays is CCAFS SLC 40, where most of recent launches were successful;
- In second place VAFB SLC 4E and third place KSC LC 39A;
- It's also possible to see that the general success rate improved over time.

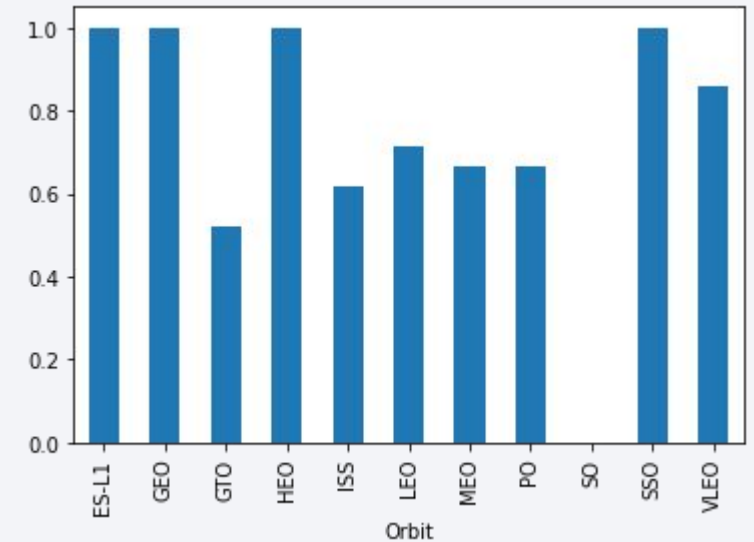
Payload vs. Launch Site



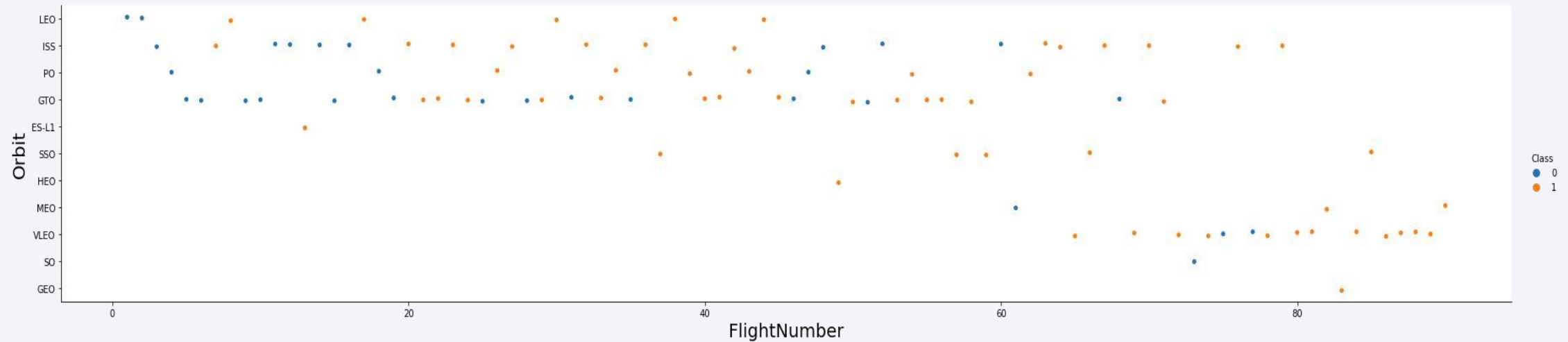
- Payloads over 9,000kg (about the weight of a school bus) have excellent success rate;
- Payloads over 12,000kg seems to be possible only on CCAFS SLC 40 and KSC LC 39A launch sites.

Success Rate vs. Orbit Type

- The biggest success rates happens to orbits:
 - ES-L1;
 - GEO;
 - HEO; and
 - SSO.
- Followed by:
 - VLEO (above 80%); and
 - LFO (above 70%).

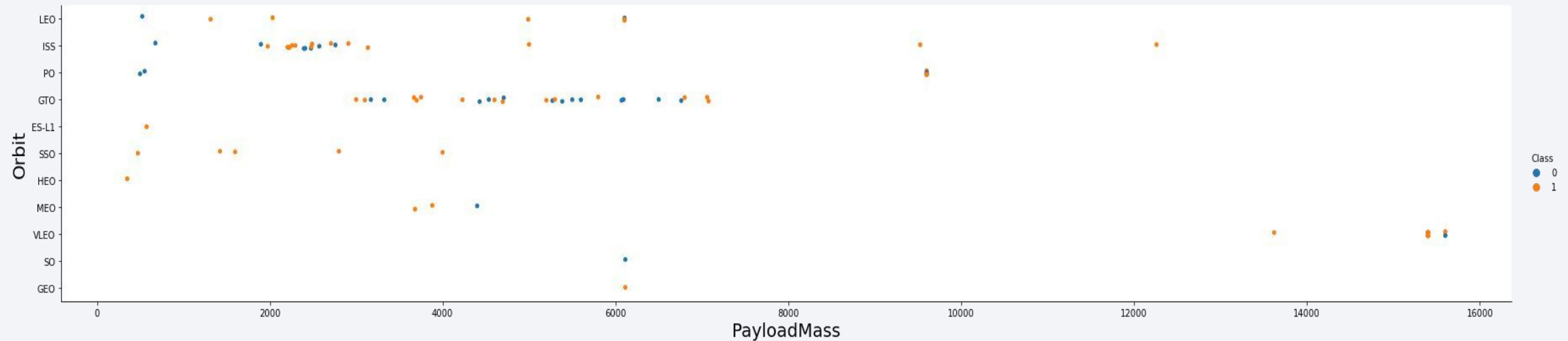


Flight Number vs. Orbit Type



- Apparently, success rate improved over time to all orbits;
- VLEO orbit seems a new business opportunity, due to recent increase of its frequency.

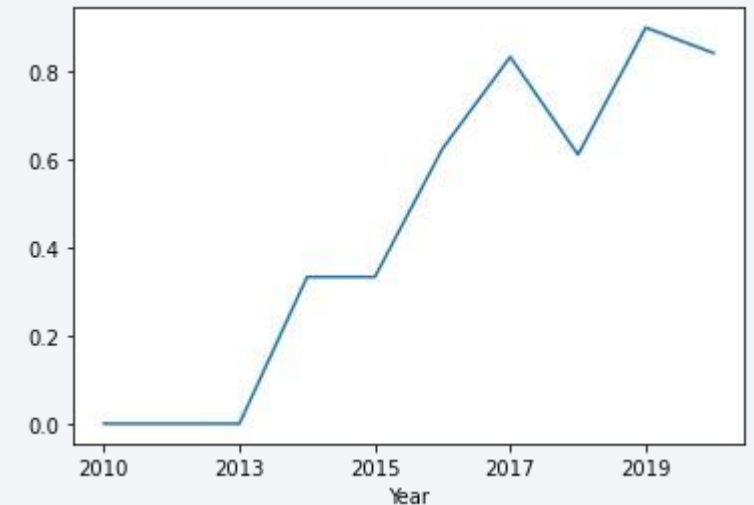
Payload vs. Orbit Type



- From the graph there is no relation between payload and success rate to orbit GTO;
- ISS orbit has the widest range of payload and a good rate of success;
- There are few launches to the orbits SO and GEO.

Launch Success Yearly Trend

- Success rate started increasing in 2013 and kept until 2020;
- It seems that the first three years were a period of adjusts and improvement of technology.



All Launch Site Names

- According to data, there are four launch sites:

| 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 | 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 attempt |

Total Payload Mass

- Total payload carried by boosters from NASA:

| Total Payload (kg) |
|--------------------|
| 111.268 |

Average Payload Mass by F9 v1.1

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

| Avg Payload (kg) |
|------------------|
| 2.928 |

- Filtering data by the booster version above and calculating the average payload mass.

First Successful Ground Landing Date

- First successful landing outcome on ground pad:

| Min Date |
|------------|
| 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 happened on 12/22/2015.

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

- Number of successful and failure mission outcomes:

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

| 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

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

| 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 |

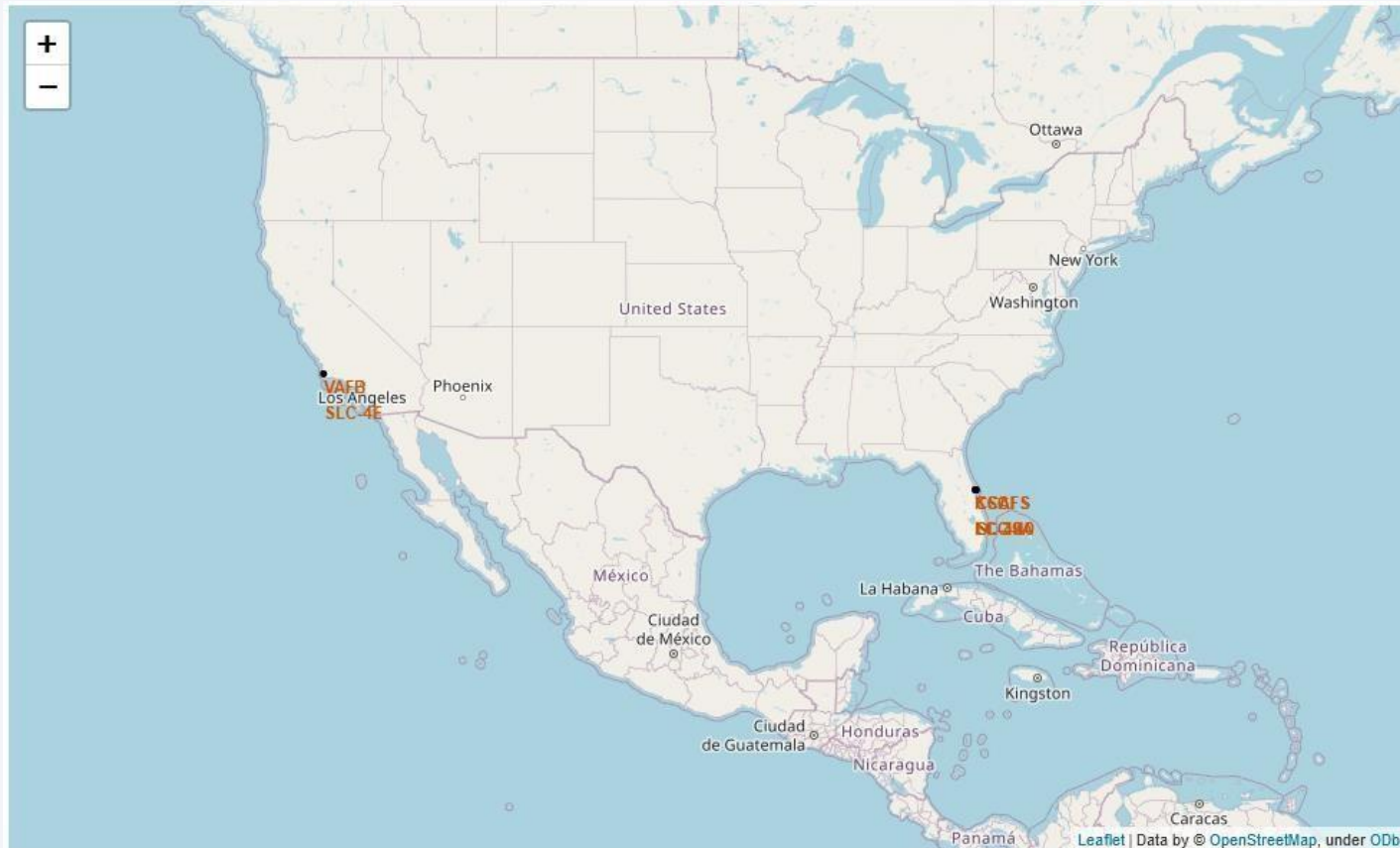
- The high number of “No attempt” cannot be ignored.

Section 4

Launch Sites Proximities Analysis



All launch sites



- Launch sites are near sea, probably by safety, but not too far from roads and railroads.

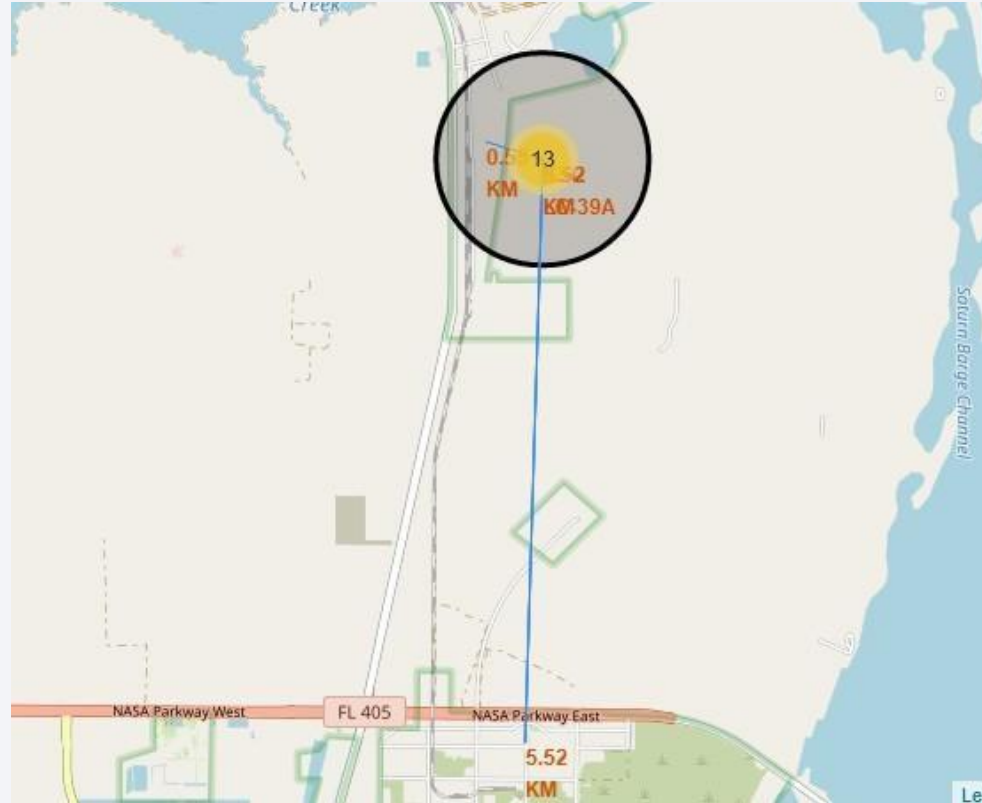
Launch Outcomes by Site

- Example of KSC LC-39A launch site launch outcomes



- Green markers indicate successful and red ones indicate failure.

Logistics and Safety



- Launch site KSC LC-39A has good logistics aspects, being near railroad and road and relatively far from inhabited areas.



Section 5

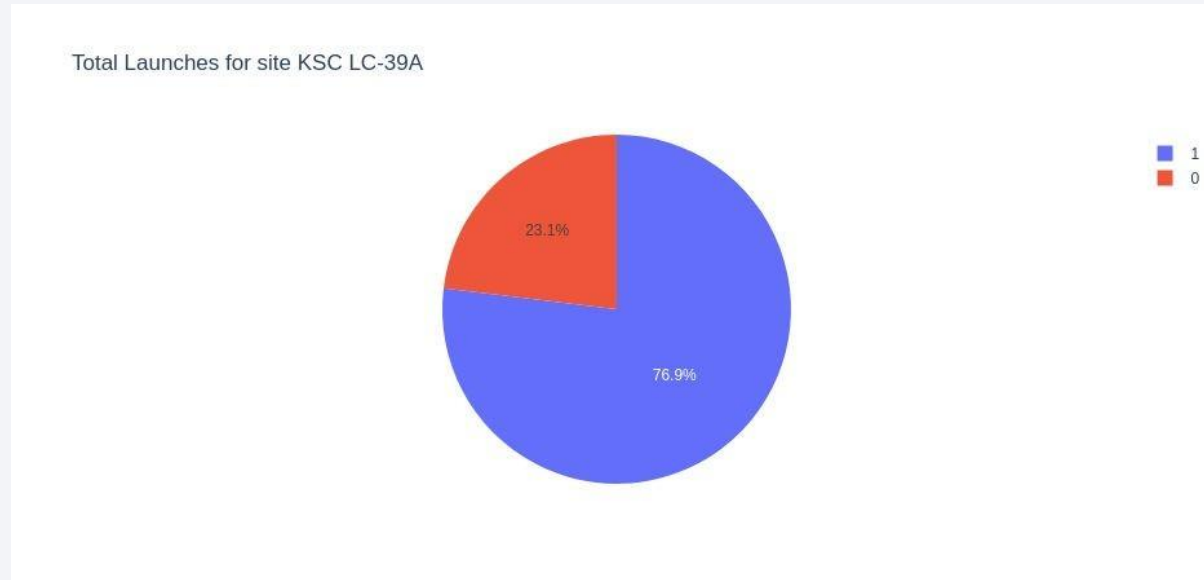
Build a Dashboard with Plotly Dash

Successful Launches by Site



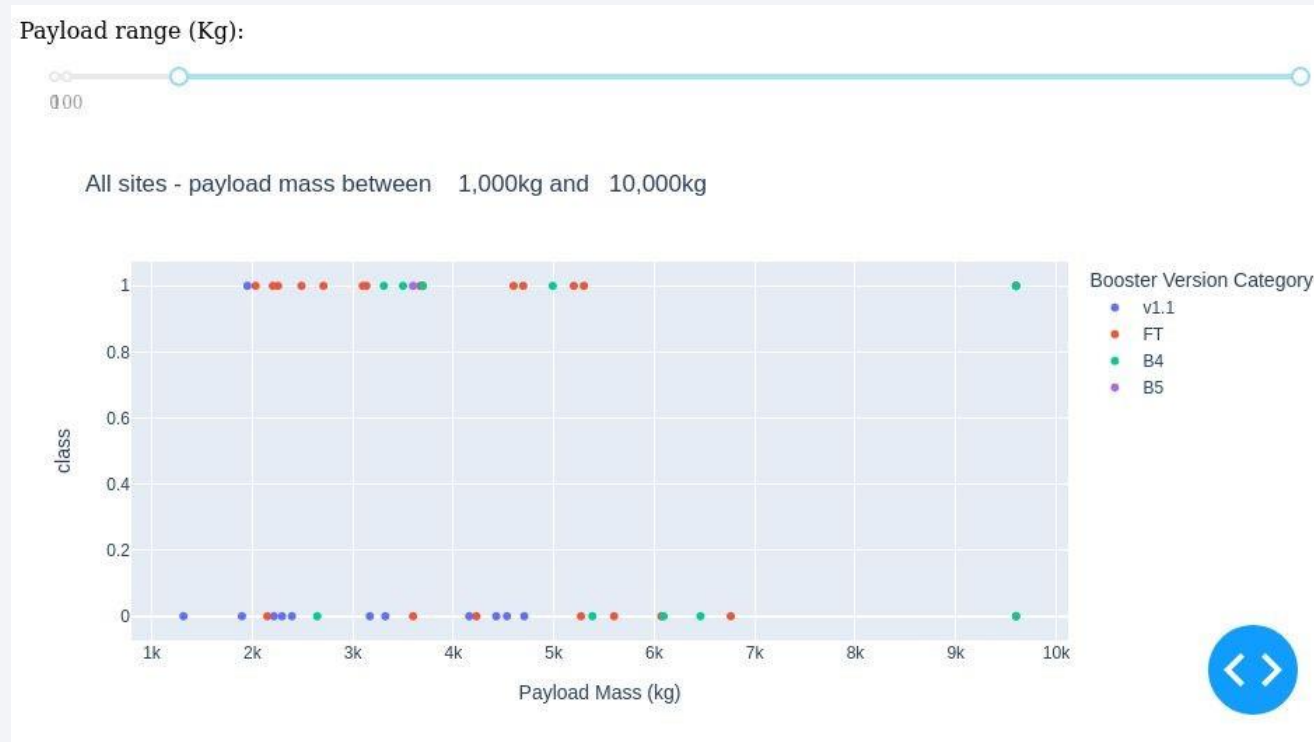
- The place from where launches are done seems to be a very important factor of success of missions.

Launch Success Ratio for KSC LC-39A



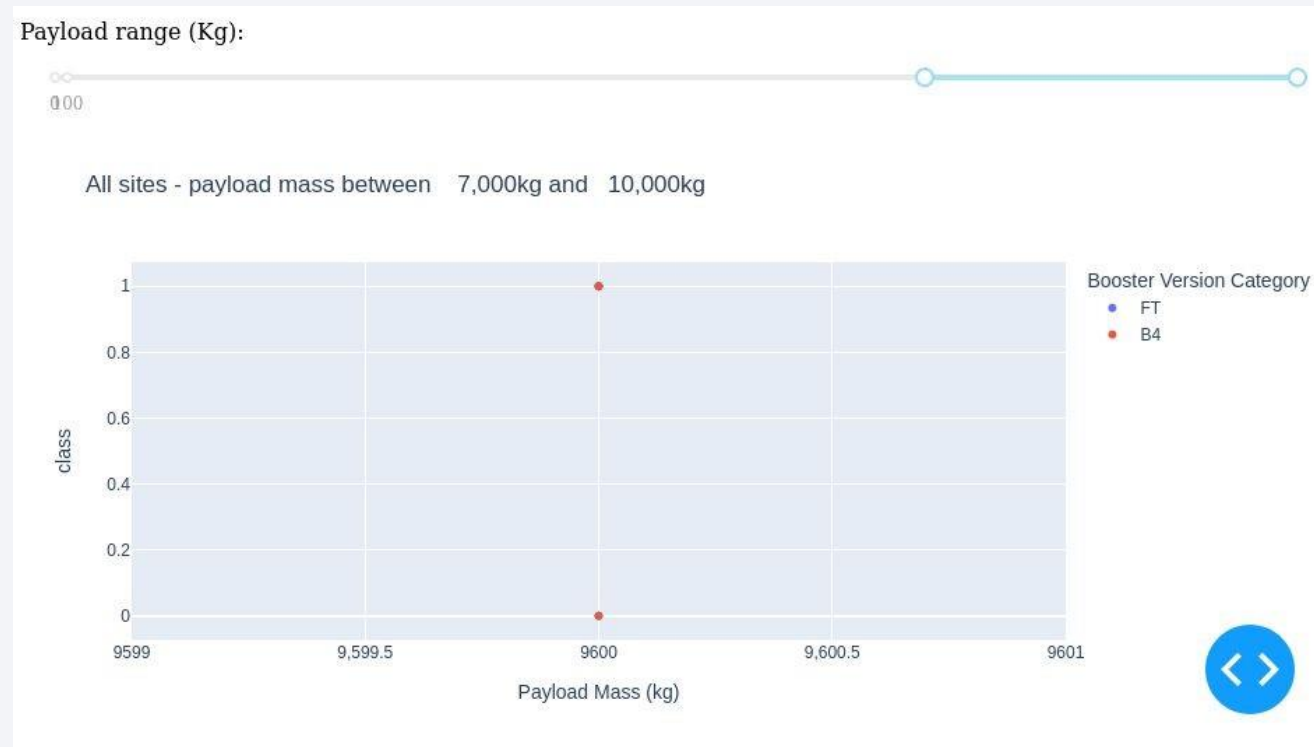
- 76.9% of launches are successful in this site.

Payload vs. Launch Outcome



- Payloads under 6,000kg and FT boosters are the most successful combination.

Payload vs. Launch Outcome



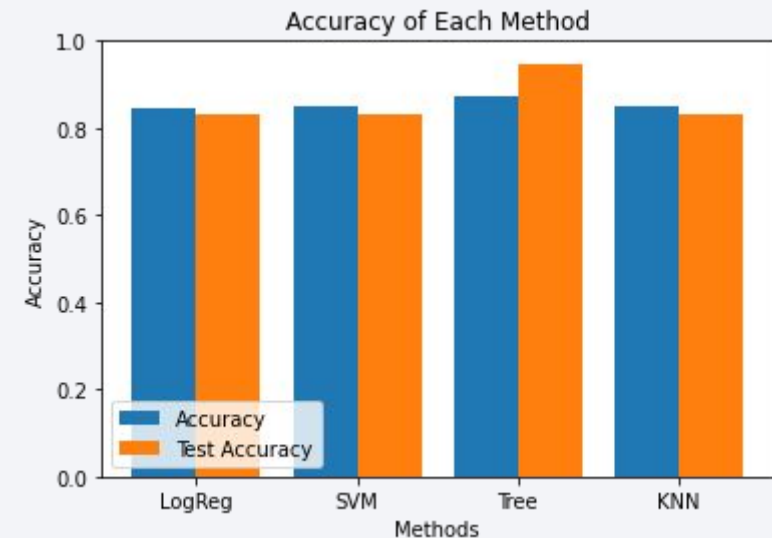
- There's not enough data to estimate risk of launches over 7,000kg

Section 6

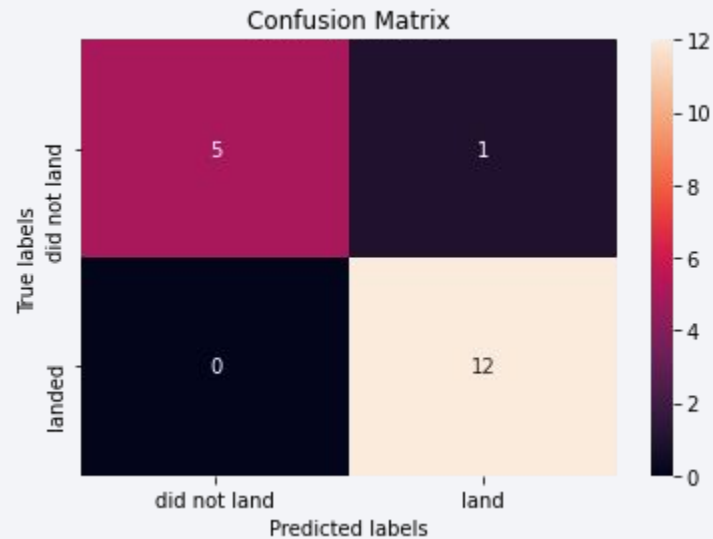
Predictive Analysis (Classification)

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 than 87%.



Confusion Matrix of Decision Tree Classifier



- Confusion matrix of Decision Tree Classifier proves its accuracy as is shown by true positive and true negative values which are much higher than the False Positive and False Negative values.

Conclusions

- Different data sources were analyzed, refining conclusions along the process;
- The best launch site is KSC LC-39A;
- Launches above 7,000kg are less risky;
- Although most of mission outcomes are successful, successful landing outcomes seem to improve over time, according the evolution of processes and rockets;
- Decision Tree Classifier can be used to predict successful landings and increase profits.

Thank you!

