IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Alejandro

04/04/2023

Outline

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

Executive Summary

- Summary of methodologies
  - Data Gathering through web scraping and the SpaceX API;
  - Initial Data Exploration (EDA), encompassing data manipulation, data visualization, and interactive visual analytics
  - Machine Learning Predictions.
- Summary of all results
  - Valuable data was successfully obtained from public sources;
  - EDA enabled the identification of the most relevant features for predicting successful launches;
  - Machine Learning Predictions revealed the optimal model for determining crucial factors to guide this opportunity most effectively, using all gathered data.

Introduction

- Project background and context

  The goal is to assess the potential of new company Space Y in competing with Space X.

- Desired outcomes:

  The most effective method for estimating total launch costs by forecasting successful first-stage rocket landings;

  The ideal location for conducting launches.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data from Space X was sourced from two places:
    - 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
  - The collected data was enhanced by generating a landing outcome label based on outcome data following feature summarization and analysis
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - The collected data was enhanced by generating a landing outcome label based on outcome data following feature summarization and analysis

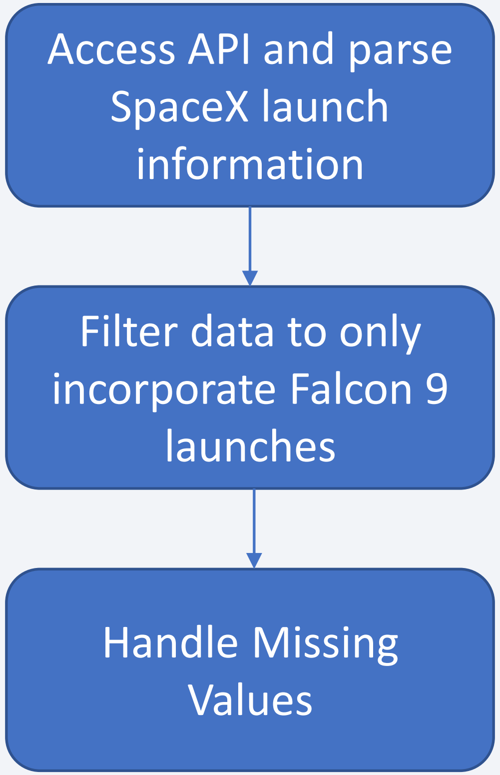
Data Collection

- Data sets were gathered from the Space X API (https://api.spacexdata.com/v4/rockets/)
- And Wikipedia (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches) using web scraping techniques.

Data Collection – SpaceX API

- This API was employed following the adjacent flowchart, and the data was then stored.

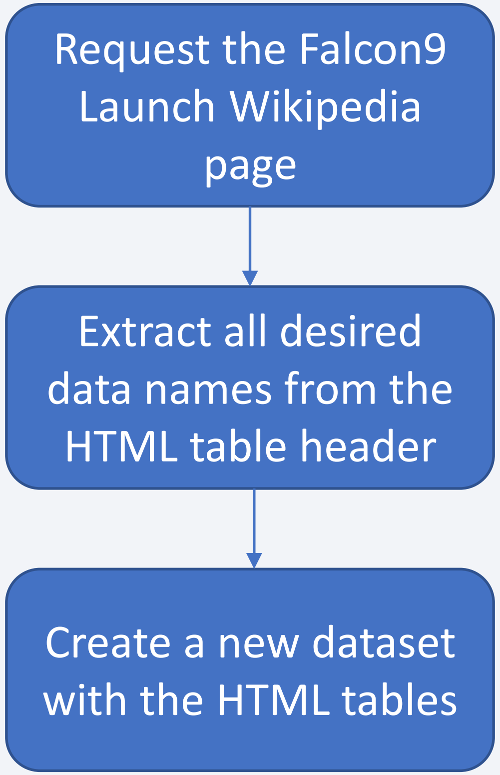
- https://github.com/DNAlex11/spacex-study/blob/f6011fc7319f021df95945751f607cd24ae4fac2/Data%20Collection%20API.ipynb



Data Collection – Scraping

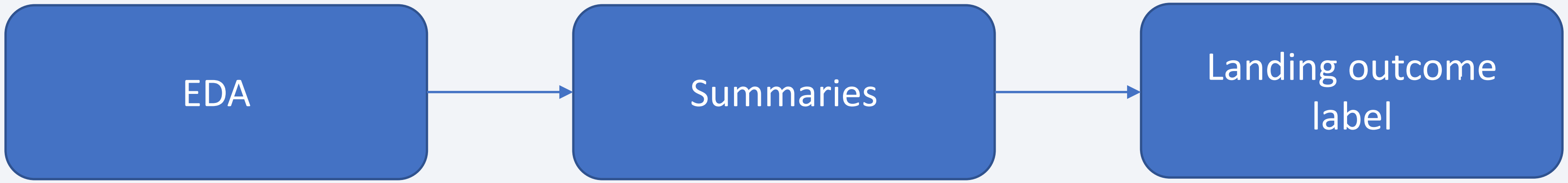
- Data from SpaceX launches can also be obtained from Wikipedia
- Data are downloaded from Wikipedia according to the flowchart and then put into our dataset.

- https://github.com/DNAlex11/spacex-study/blob/f6011fc7319f021df95945751f607cd24ae4fac2/Data%20Collection%20with%20Web%20Scraping.ipynb



Data Wrangling

- Initially, some Preliminary Data Analysis (EDA) was conducted on the dataset.
- Next, summaries of launches per site, occurrences of each orbit, and occurrences of mission outcomes per orbit type were determined.
- Lastly, the landing outcome label was generated from the Outcome column.



- https://github.com/DNAlex11/spacex-study/blob/f6011fc7319f021df95945751f607cd24ae4fac2/Data%20Wrangling.ipynb

EDA with Data Visualization

- Scatterplots and barplots were utilized to examine data, displaying the relationship between pairs of features:
  - Payload Mass X Flight Number, Launch Site X Flight Number, Launch Site X Payload Mass, Orbit and Flight Number, Payload and Orbit

- Add the GitHub URL of your completed EDA with data visualization notebook, as an external reference and peer-review purpose

EDA with SQL

The following SQL queries were executed:

- Unique launch site names in the space mission;
- Top 5 launch sites with names starting with 'CCA';
- Total payload mass carried by NASA-launched boosters (CRS);
- Average payload mass carried by the F9 v1.1 booster version;
- Date of the first successful ground pad landing outcome;
- Names of boosters with drone ship successes and payload masses between 4000 and 6000 kg;
- Total number of successful and failed mission outcomes;
- Names of booster versions carrying the maximum payload mass;
- Failed drone ship landing outcomes, their booster versions, and launch site names in 2015; and
- Ranking of landing outcome counts (e.g., Failure (drone ship) or Success (ground pad)) between 2010-06-04 and 2017-03-20.

- https://github.com/DNAlex11/spacex-study/blob/f6011fc7319f021df95945751f607cd24ae4fac2/EDA.ipynb

Build an Interactive Map with Folium

- Folium Maps employed markers, circles, lines, and marker clusters
  - Markers represent points like launch sites
  - Circles highlight areas around specific coordinates, such as NASA Johnson Space Center
  - Marker clusters indicate groups of events at each coordinate, like launches at a launch site
  - Lines denote distances between two coordinates.

- https://github.com/DNAlex11/spacex-study/blob/f6011fc7319f021df95945751f607cd24ae4fac2/Interactive%20Visual%20Analytics%20with%20Folium%20lab.ipynb

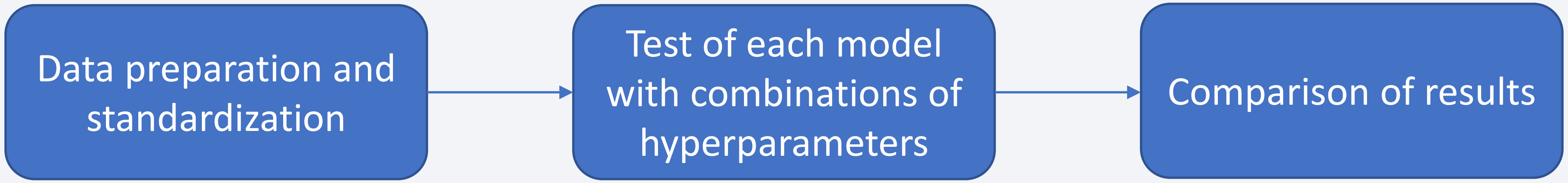
Build a Dashboard with Plotly Dash

- The following charts and plots were employed to visualize data
    - Launch site percentages
    - Payload range
- This combination facilitated a quick analysis of the relationship between payloads and launch sites, aiding in identifying the optimal launch location based on payloads.

- https://github.com/DNAlex11/spacex-study/blob/f6011fc7319f021df95945751f607cd24ae4fac2/spacex_dash_app.py

Predictive Analysis (Classification)

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



- https://github.com/DNAlex11/spacex-study/blob/f6011fc7319f021df5945751f607cd24ae4fac2/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

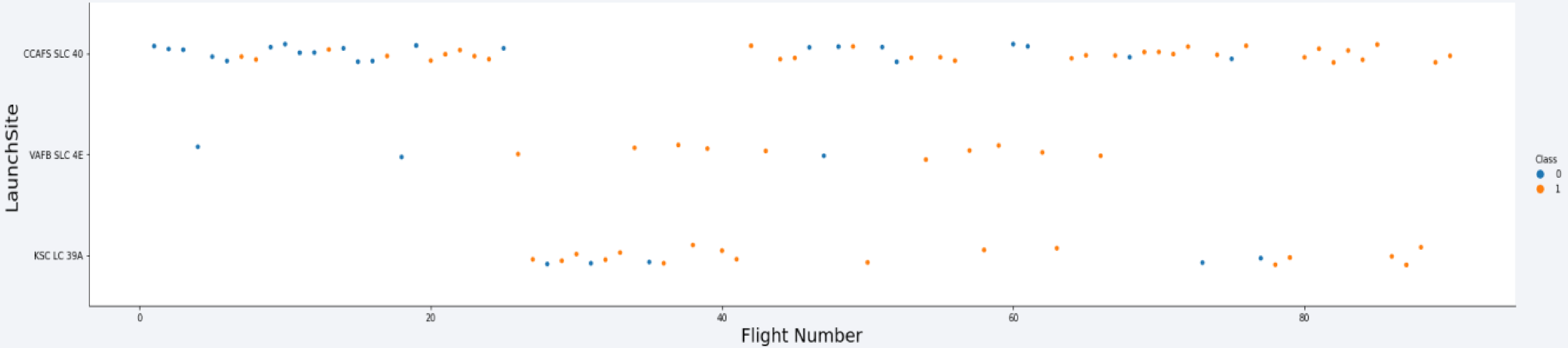
Results

- Preliminary data analysis findings:
  - Space X utilizes 4 distinct launch sites;
  - Initial launches were conducted for Space X itself and NASA;
  - The average payload of the F9 v1.1 booster is 2,928 kg;
  - The first successful landing outcome occurred in 2015, five years after the first launch;
  - Several Falcon 9 booster versions successfully landed on drone ships with payloads above the average;
  - Nearly 100% of mission outcomes were successful;
  - Two booster versions failed to land on drone ships in 2015: F9 v1.1 B1012 and F9 v1.1 B1015;
  - The number of successful landing outcomes increased over the years.

Section 2

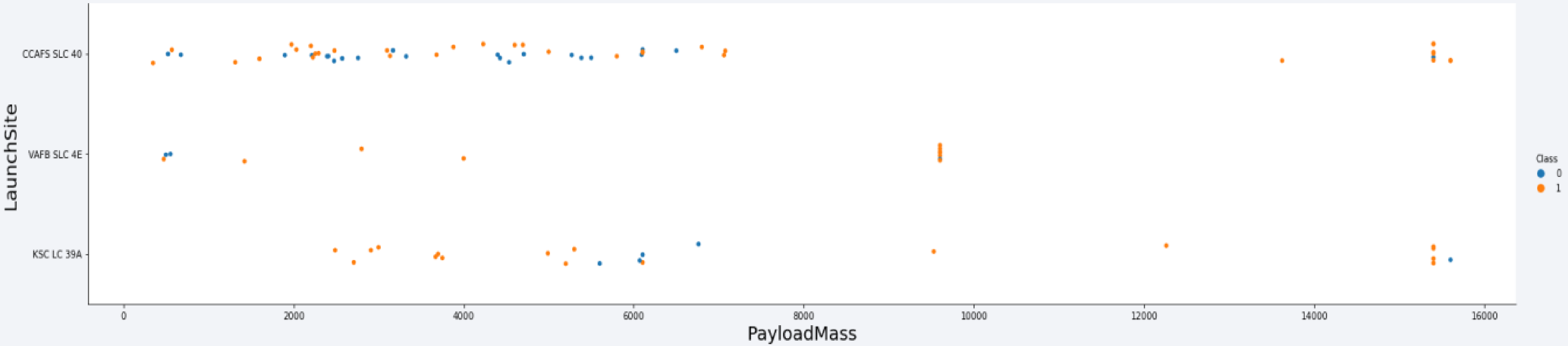
Insights drawn from EDA

Flight Number vs. Launch Site



- Based on the plot above, it can be observed that the current best launch site is CCAFS SLC 40, where most recent launches have been successful;
- VAFB SLC 4E ranks second, followed by KSC LC 39A in third place;
- The overall success rate has improved over time.

Payload vs. Launch Site

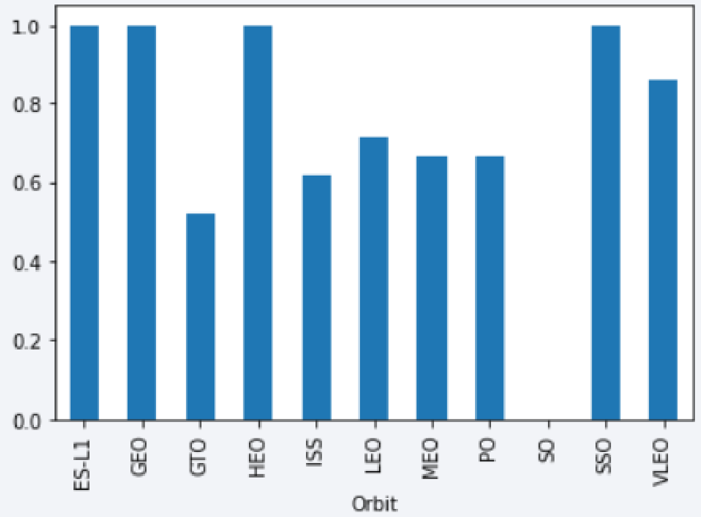


- Payloads above 9,000 kg (approximately the weight of a school bus) exhibit a high success rate;

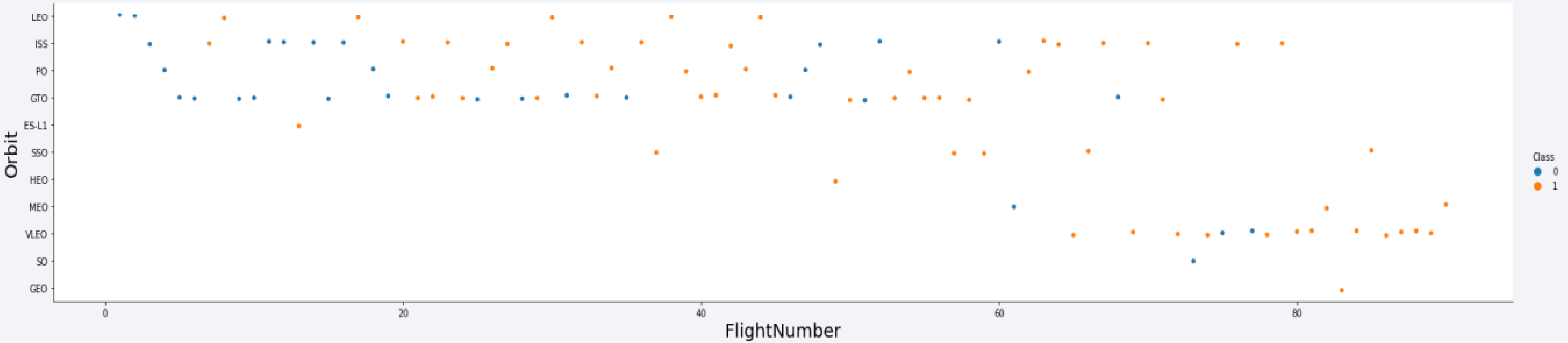
- Payloads exceeding 12,000 kg seem feasible only at CCAFS SLC 40 and KSC LC 39A launch sites.

Success Rate vs. Orbit Type

- The highest success rates are associated with orbits:
    - ES-L1
    - GEO
    - HEO
    - SSO.
- Next in line are:
    - VLEO (over 80%)
    - LFO (over 70%)

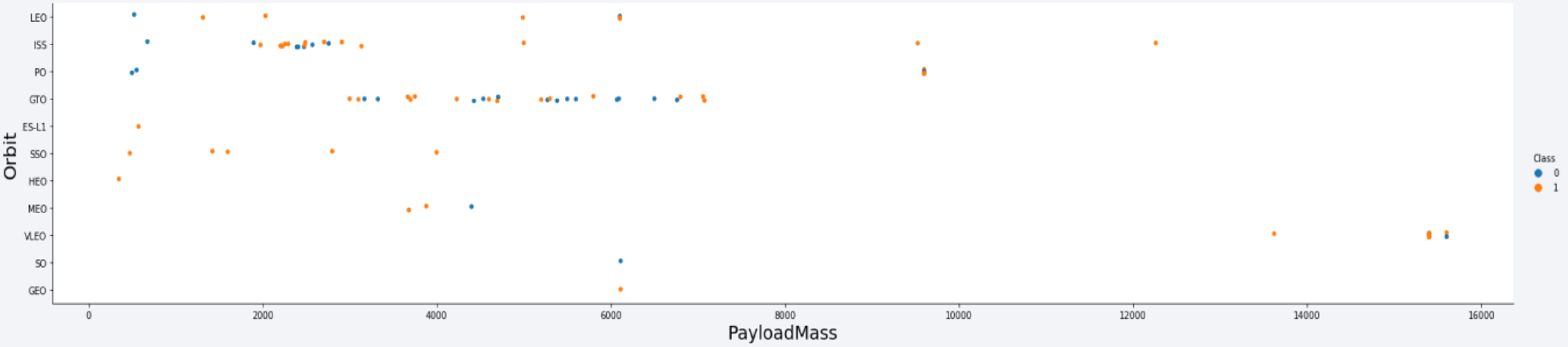


Flight Number vs. Orbit Type



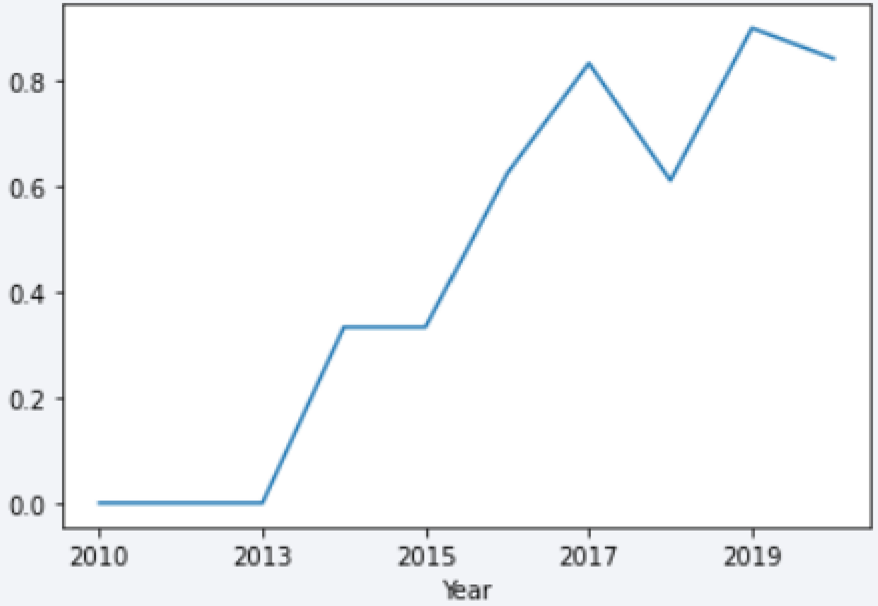
- Success rate seems to have improved over time for all orbits;
- The VLEO orbit appears to be a new business opportunity due to its recent increase in frequency

Payload vs. Orbit Type



- There doesn't seem to be a correlation between payload and success rate for GTO orbit;
- ISS orbit has the broadest payload range and a solid success rate;
- Only a few launches target SO and GEO orbits.

Launch Success Yearly Trend



- The success rate began to rise in 2013 and continued through 2020;
- The first three years seem to have been a period of adjustments and technological advancements.

All Launch Site Names

- The data reveals four launch sites:
  - CCAFS LC-40
  - CCAFS SLC-40
  - KSC LC-39A
  - VAFB SLC-4E
- These are acquired by selecting unique instances of "launch_site" values from the 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

- In this query, we can see five examples of Cape Canaveral launches.

Total Payload Mass

- Calculate the total payload carried by boosters from NASA
  - Total Payload (kg)
  - 111.268
- The total payload calculated is obtained by adding all payloads with codes containing 'CRS,' which corresponds to NASA.

Average Payload Mass by F9 v1.1

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

  Avg Payload (kg)

  2.928

• By filtering data for the booster version F9 v1.1 and calculating the average payload mass, we get the value of 2,928 kg.

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
    - Min Date 2015-12-22
- By filtering data for successful ground landing outcomes and finding the earliest date, we identify the first occurrence on December 22, 2015.

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the 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

- Selecting distinct booster versions based on the filters, the following four results are obtained.

Total Number of Successful and Failure Mission Outcomes

- Calculate the total 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 produces the summary below.

Boosters Carried Maximum Payload

- List 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

- The following boosters have carried the maximum payload mass recorded in the dataset.

2015 Launch Records

- List 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

- 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

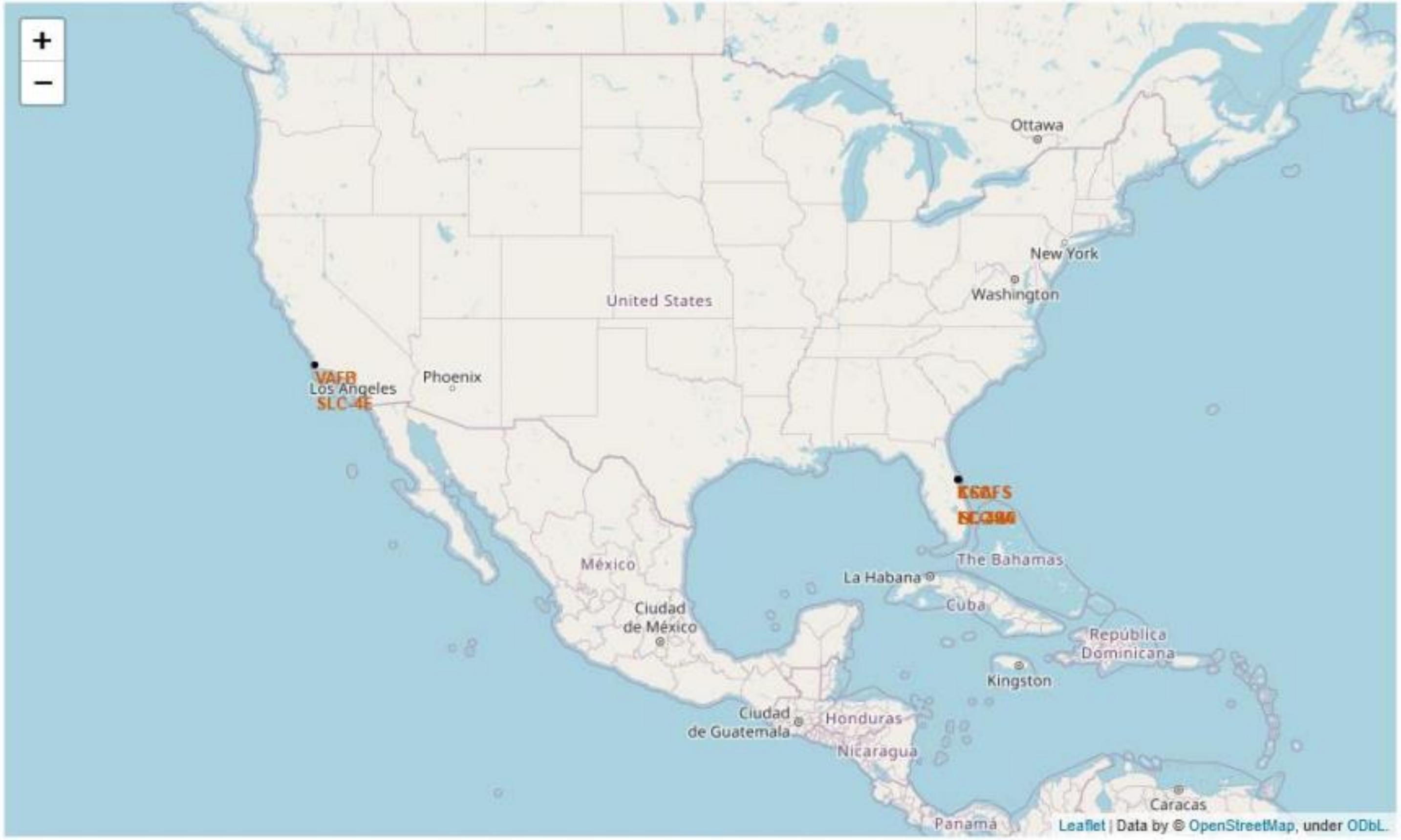
Landing Outcome	Occurrences
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

- We see that ‘No attempt’ has 10 occurrences

Section 3

Launch Sites Proximities Analysis

Launch sites

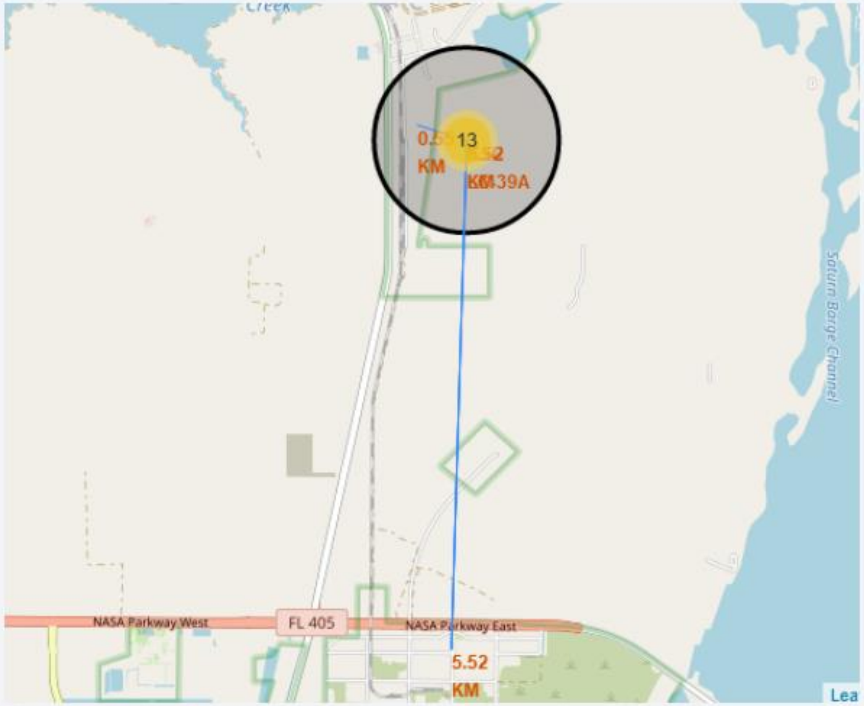


- Launch sites are typically situated near the sea, likely for safety reasons, but also remain in proximity to roads and railroads.

Launch Outcomes

- An example of launch outcomes at the KSC LC-39A site
- Green markers represent successful launches, while red ones indicate failures.

Launch site characteristics

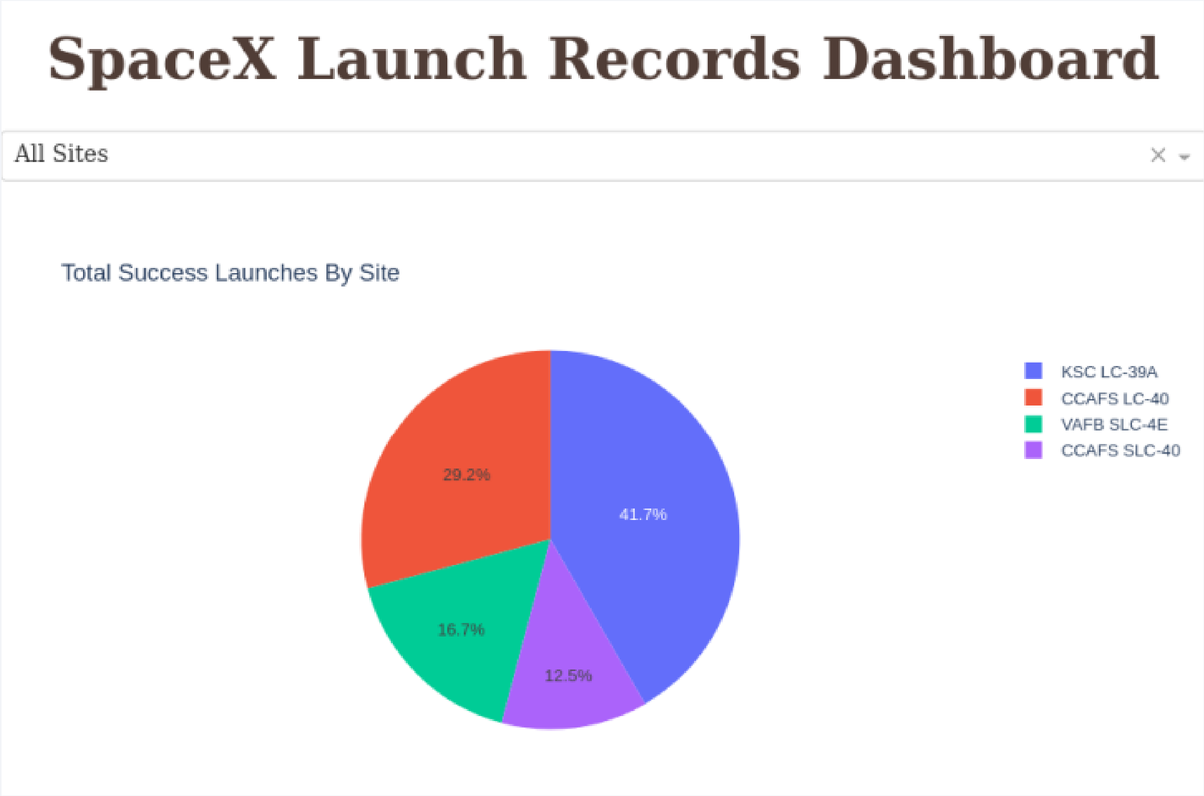


- The KSC LC-39A launch site exhibits good logistical aspects, being close to railroads and roads while remaining relatively distant from populated areas.

Section 4

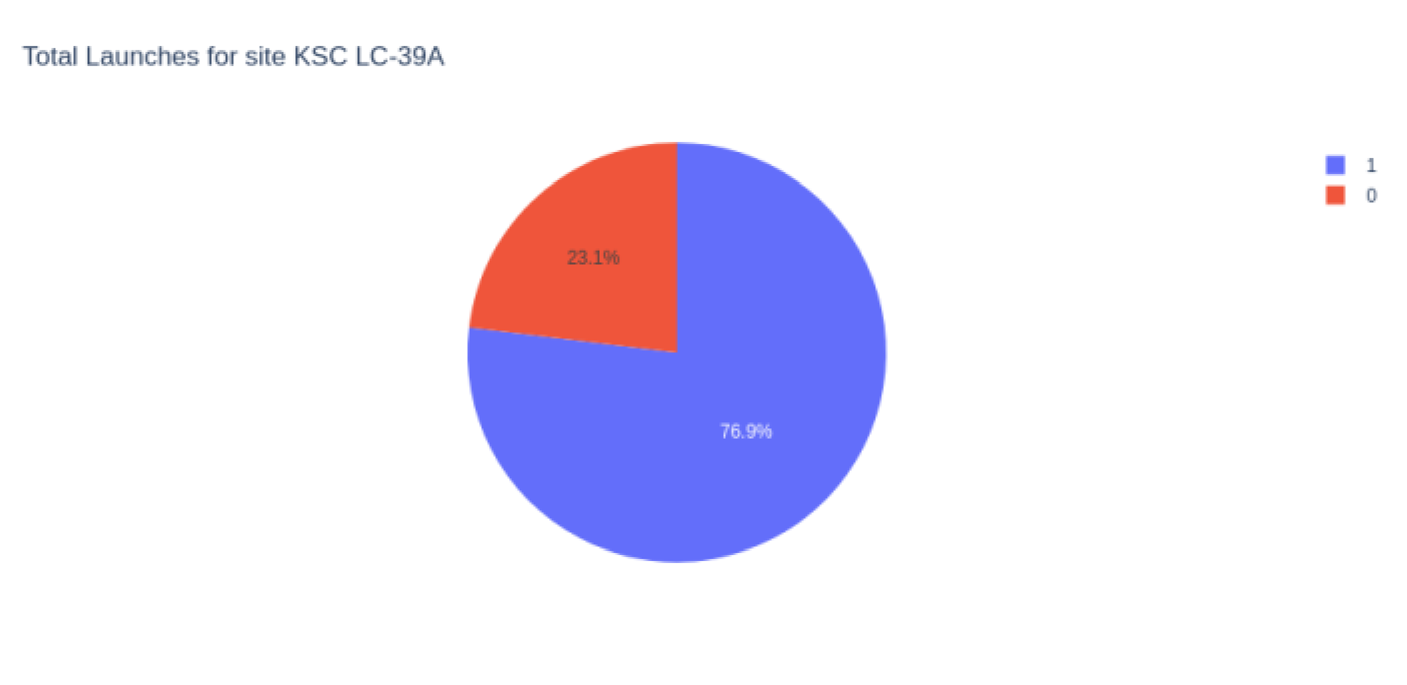
Build a Dashboard with Plotly Dash

Launch Success



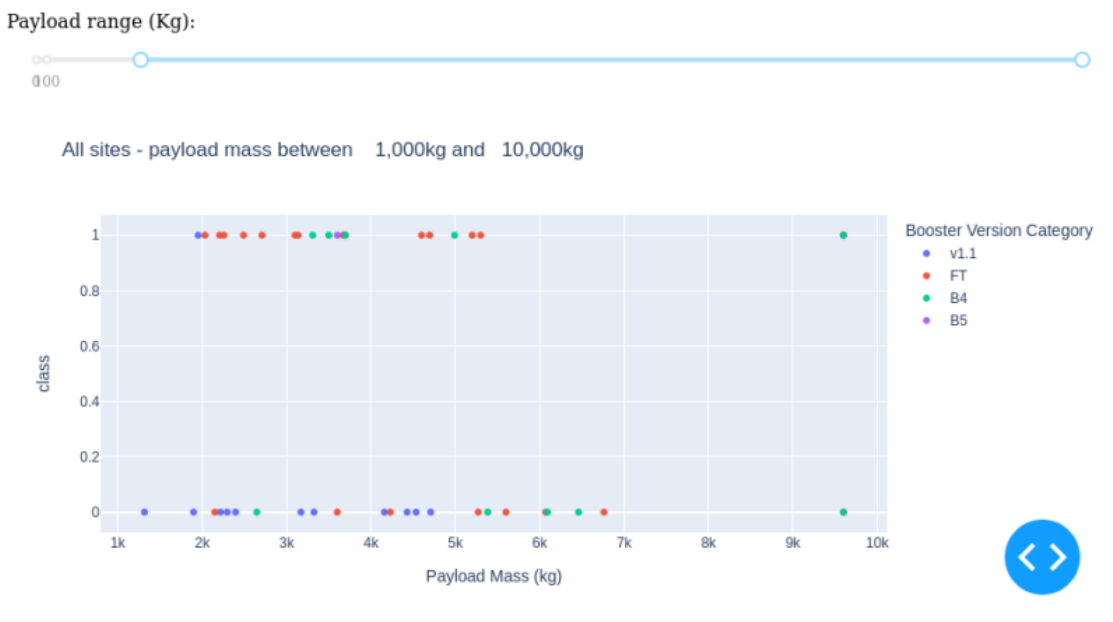
- The location from which launches are conducted appears to be a crucial factor in mission success.

KSC LC-39A Success Ratio



- Almost 77% of launches are successful

Payload vs Launch Outcome



- Explain the important elements and findings on the screenshot, such as which payload range or booster version have the largest success rate, etc.

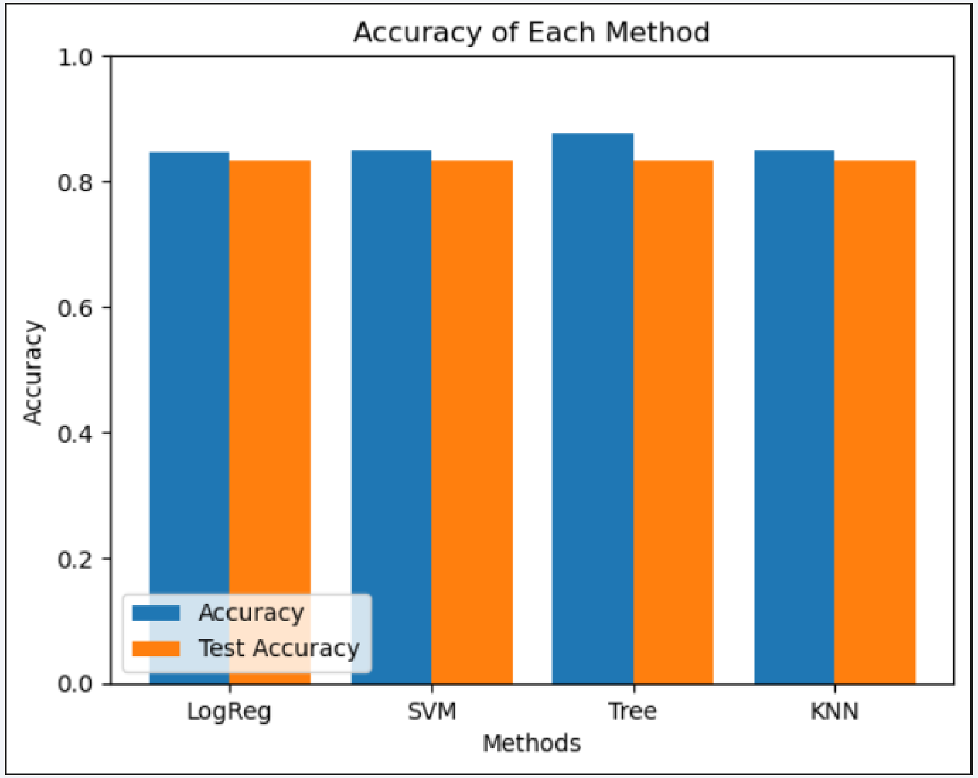
Section 5

Predictive Analysis (Classification)

Classification Accuracy

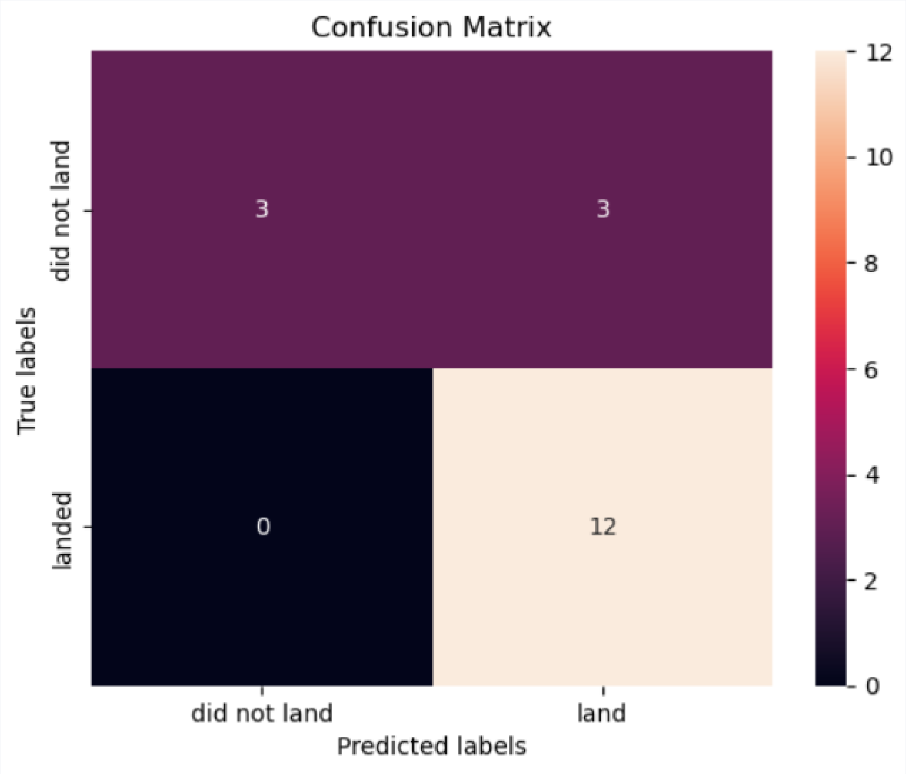
• Four classification models were tested, and their accuracies are plotted here

• All models yield the same test accuracy, roughly 84%



Confusion Matrix

- The model misclassifies half of the landings that didn't actually land



Conclusions

- From this study we can deduct:
  - Launches above 7,000 kg are less risky
  - The optimal launch site is KSC LC-39A
  - Decision Tree Classifier can be utilized to predict successful landings and enhance profits
  - While most mission outcomes are successful, successful landing outcomes appear to improve over time due to advancements in processes and rocket technology
  - Various data sources were examined, with conclusions being refined throughout the process.

Appendix

In addition to the findings discussed earlier, here are some creative insights that could be explored further:

- Investigate the impact of weather conditions and launch timing on the success rate of launches. This could help optimize launch schedules and potentially minimize risks associated with adverse weather.

- Analyze the role of partnerships and collaborations with other organizations (such as universities, private companies, or government agencies) in the success of space missions. Understanding how collaboration affects mission outcomes may guide future partnerships and resource allocation.

- Examine the influence of team composition, experience, and diversity on the success of launches. This could lead to more informed decisions when assembling teams and allocating resources.

Thank you!