IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Pablo Ramos
07/04/2023

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection: Web scraping and SpaceX API
  - Exploratory Data Analysis (EDA): Data wrangling, data visualization, interactive visual analytics
  - Machine Learning Prediction
- Results Summary:
  - Collected valuable data from public sources
  - EDA identified best features for predicting launch success
  - Machine Learning Prediction determined the best model for leveraging collected data

Introduction

- Goal:
  - Assess the potential of Space Y to rival SpaceX
- Key Insights Needed:
  - Forecasting total launch expenses through successful first stage rocket landings predictions
  - Determining the optimal site for conducting launches

Section 1

Methodology

Methodology

Executive Summary

- Method for gathering data:
  - Acquired Space X data from two distinct resources:
  - Space X API (https://api.spacexdata.com/v4/rockets/)
  - Web Scraping (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches)
- Conduct data manipulation
  - Enhanced the collected data by generating a landing result label from outcome information after examining and summarizing attributes
- Carry out exploratory data analysis (EDA) employing visualizations and SQL

Methodology

Executive Summary

- Carry out interactive visual analysis with Folium and Plotly Dash
- Execute predictive assessments employing classification algorithms
  - Up to this stage, gathered data was standardized, separated into training and testing datasets, and assessed with four distinct classification techniques. Each model's accuracy was measured using various parameter combinations

Data Collection

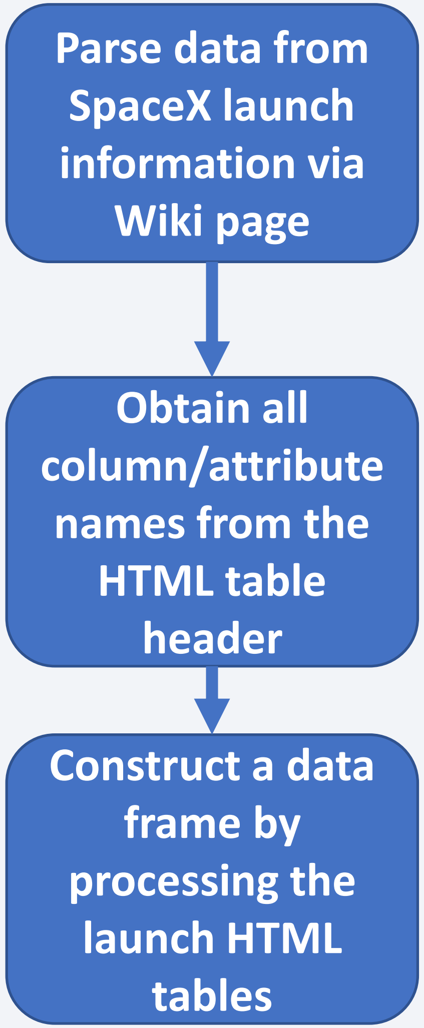
- Information was obtained from a pair of sources: Space X API (https://api.spacexdata.com/v4/rockets/rockets/) and Wikipedia (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches), employing web scraping methods.

Data Collection – SpaceX API

- SpaceX provides an accessible API for data acquisition and utilization;
- Following the adjacent flowchart, the API was employed, and the data was stored.
- Source code: https://github.com/bloramos/IBM-Applied-Data-Science-Capstone/blob/main/Data%20Collection%20API.ipynb

Data Collection - Scraping

- Information on SpaceX launches can be sourced from Wikipedia
- Data is retrieved following the flowchart and subsequently stored.
- Source code:https://github.com/bloramos/IBM-Applied-Data-Science-Capstone/blob/main/Data%20Collection%20with%20Web%20Scraping.ipynb

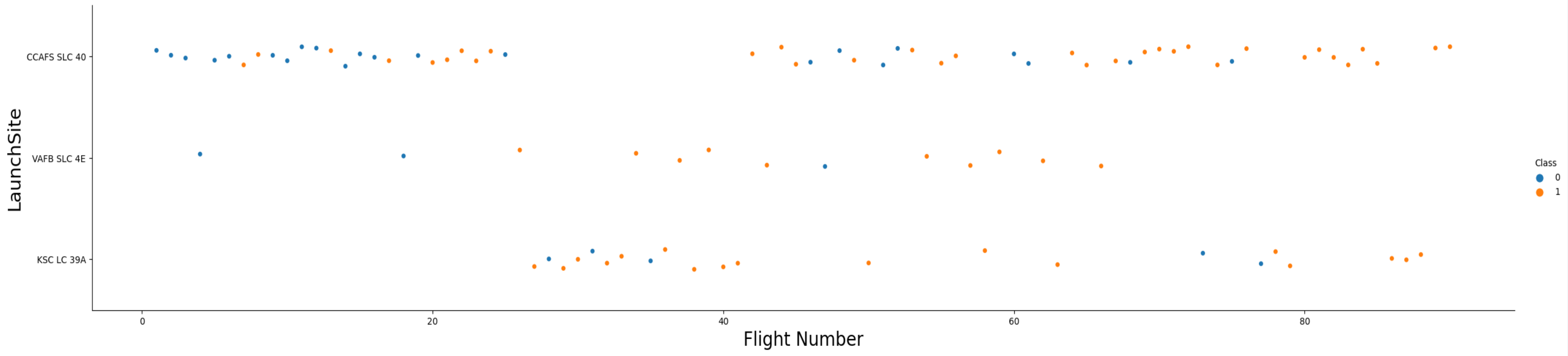


Data Wrangling

- At first, Exploratory Data Analysis (EDA) was conducted on the dataset.
- Next, calculations were made for the number of launches per site, frequency of each orbit, and mission outcomes for each orbit type.
- Ultimately, the landing outcome label was derived from the Outcome column.
- Source code: https://github.com/bloramos/IBM-Applied-Data-Science-Capstone/blob/main/Data%20Wrangling.ipynb

EDA with Data Visualization

- For data exploration, scatterplots and barplots were employed to display the connections between pairs of attributes:
- Payload Mass vs. Flight Number, Launch Site vs. Flight Number, Launch Site vs. Payload Mass, Orbit and Flight Number, Payload and Orbit.
- Source code: https://github.com/bloramos/IBM-Applied-Data-Science-Capstone/blob/main/EDA%20with%20Data%20Visualization.ipynb



EDA with SQL

- SQL queries 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: https://github.com/bloramos/IBM-Applied-Data-Science-Capstone/blob/main/EDA.ipynb

Build an Interactive Map with Folium

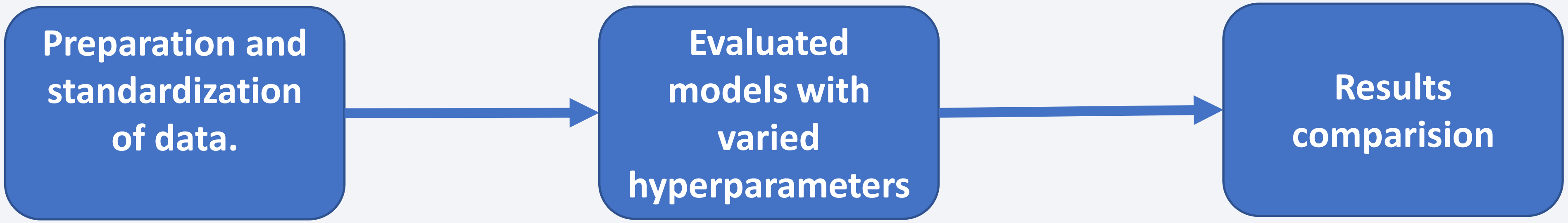
- Using Folium Maps, the following elements were employed:
  - Markers, which pinpoint locations such as launch sites;
  - Circles, highlighting areas surrounding specific coordinates, like NASA Johnson Space Center
  - Marker clusters, representing groups of events at each coordinate, like launches at a launch site; and
  - Lines, utilized for illustrating distances between two coordinates.
  - Source code: https://github.com/bloramos/IBM-Applied-Data-Science-Capstone/blob/main/Interactive%20Visual%20Analytics%20with%20Folium%20lab.ipynb

Build a Dashboard with Plotly Dash

- Various graphs and plots were employed for data visualization:
  - Proportion of launches per site
  - Payload spectrum
- This combination facilitated swift analysis of the relationship between payloads and launch sites, aiding in determining the optimal location for launches based on payload requirements.
- Source code: https://github.com/bloramos/IBM-Applied-Data-Science-Capstone/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

- Four classification methods were evaluated: logistic regression, support vector machine, decision tree, and k-nearest neighbors.
- Source code: https://github.com/bloramos/IBM-Applied-Data-Science-Capstone/blob/main/Machine%20Learning%20Prediction.ipynb

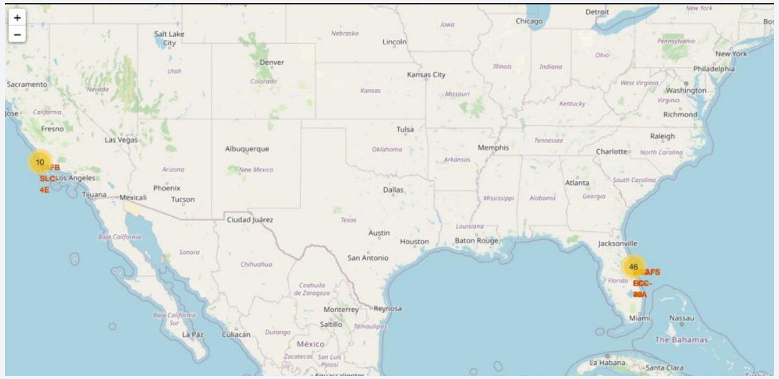


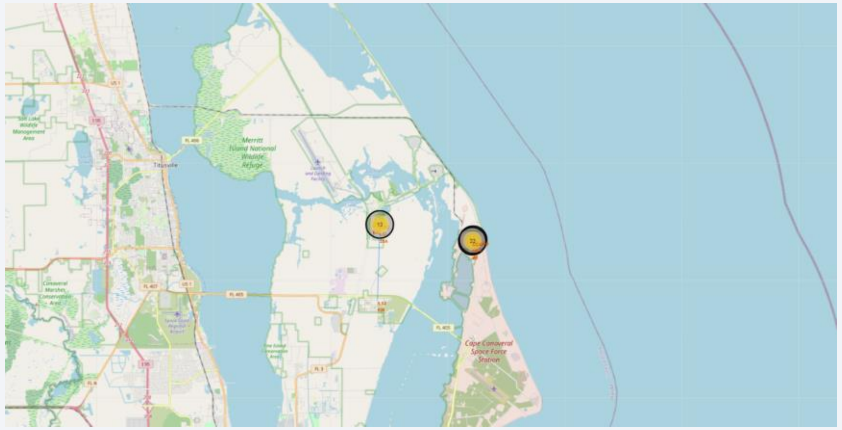
Results

- Exploratory data analysis results:
  - Space X operates from 4 distinct launch sites;
  - Early launches were conducted for Space X and NASA;
  - F9 v1.1 booster has an average payload of 2,928 kg;
  - Successful landing outcomes began 5 years after the first launch in 2015;
  - Mission outcomes were successful in nearly 100% of cases;
  - Two booster versions, F9 v1.1 B1012 and F9 v1.1 B1015, failed to land on drone ships in 2015;
  - The rate of successful landing outcomes improved over time.

Results

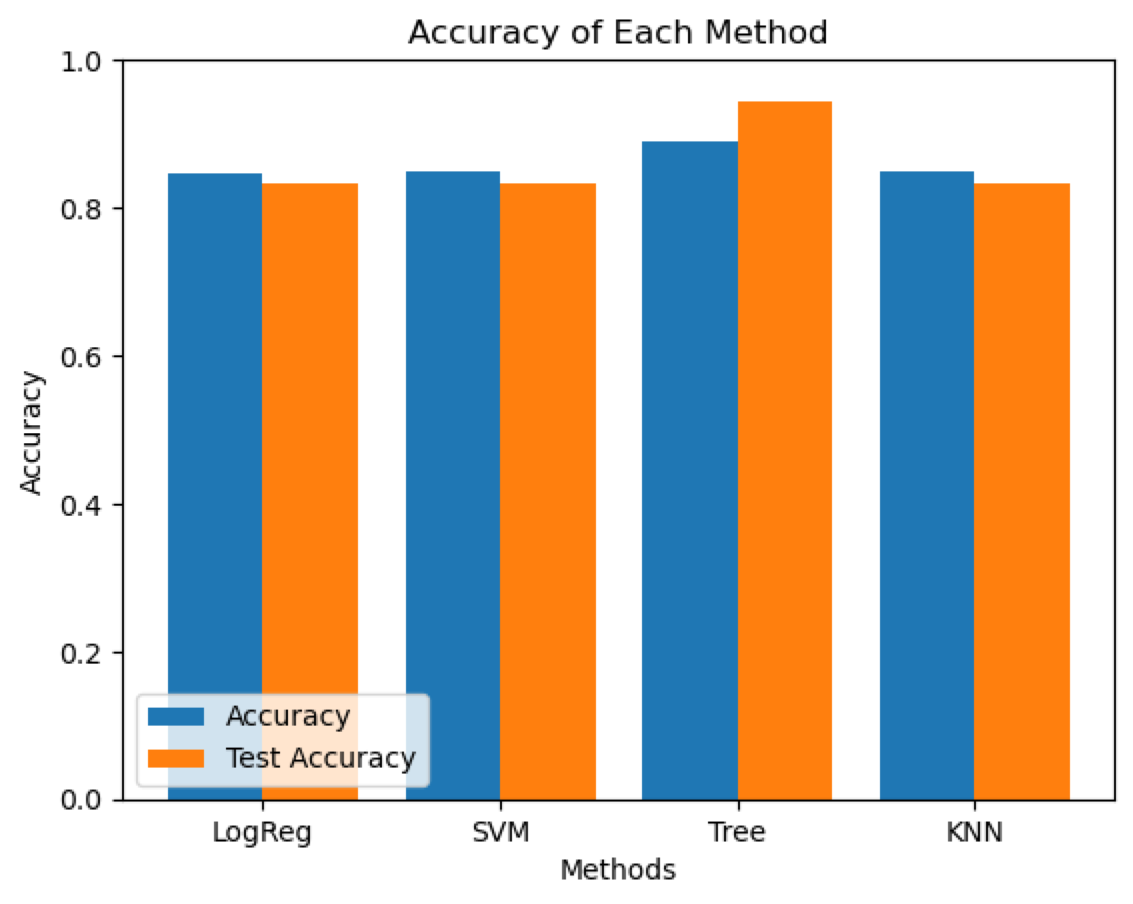
- Through interactive analytics, it was feasible to determine that launch sites are typically located in secure areas near bodies of water, with well-established logistical infrastructure nearby.
- The majority of launches occur at launch sites located on the east coast.





Results

- According to the results of predictive analysis, the Decision Tree Classifier proved to be the most effective model in predicting successful landings, with an accuracy exceeding 87% and a test data accuracy of over 94%.The majority of launches occur at launch sites located on the east coast.

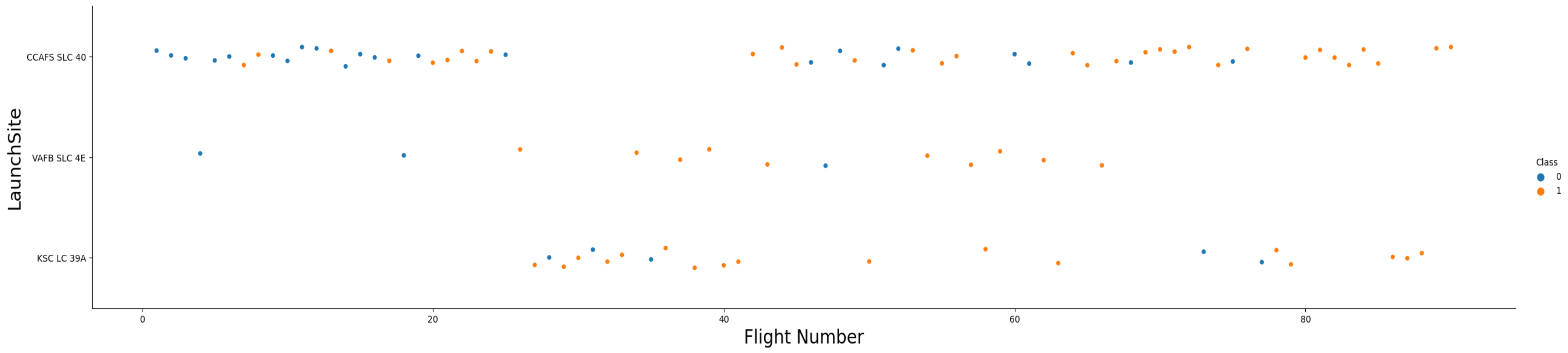


Section 2

Insights drawn from EDA

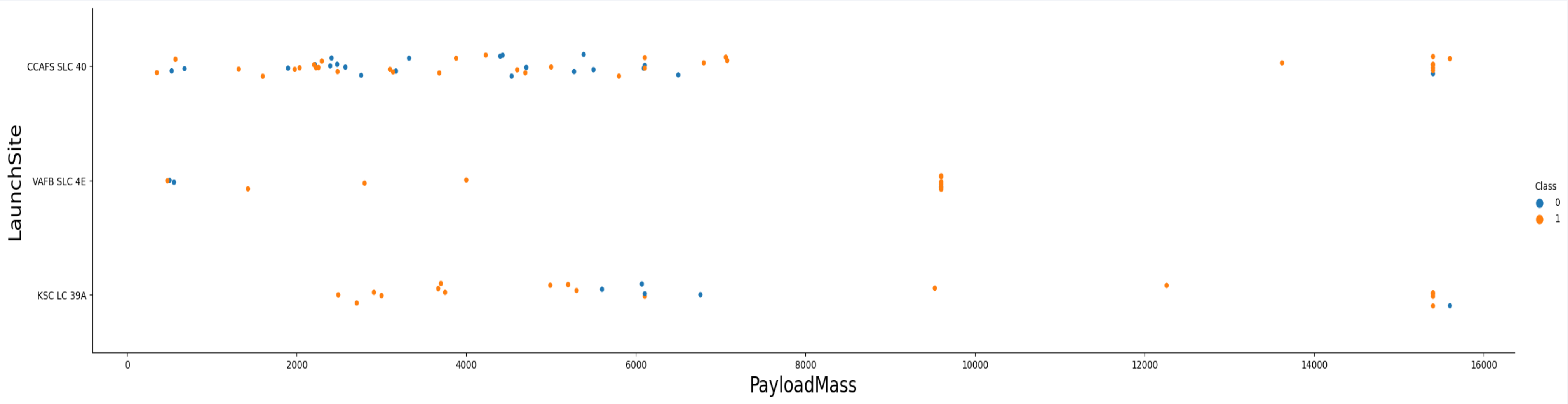
Flight Number vs. Launch Site

- Based on the above graph, it is evident that the most successful launch site presently is CCAF5 SLC 40, with the majority of recent launches achieving success.
- Second place is occupied by VAFB SLC 4E, with KSC LC 39A coming in third.
- The success rate, in general, has improved over time.



Payload vs. Launch Site

- Payloads weighing more than 9,000kg, equivalent to the weight of a school bus, display an excellent success rate.
- Only the CCAFS SLC 40 and KSC LC 39A launch sites appear capable of accommodating payloads exceeding 12,000kg.

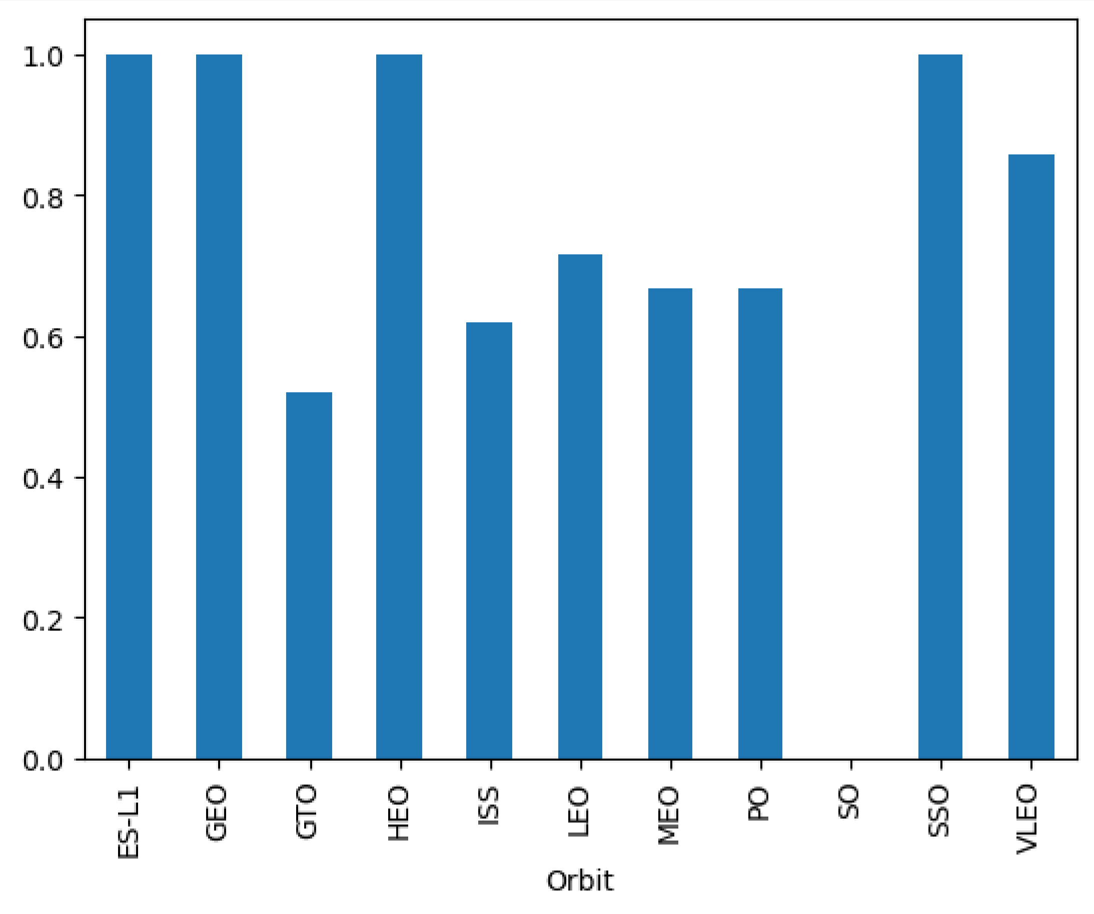


Success Rate vs. Orbit Type

- The highest success rates are observed for the following orbits:
    - ES L1
    - GEO
    - HEO
    - SSO

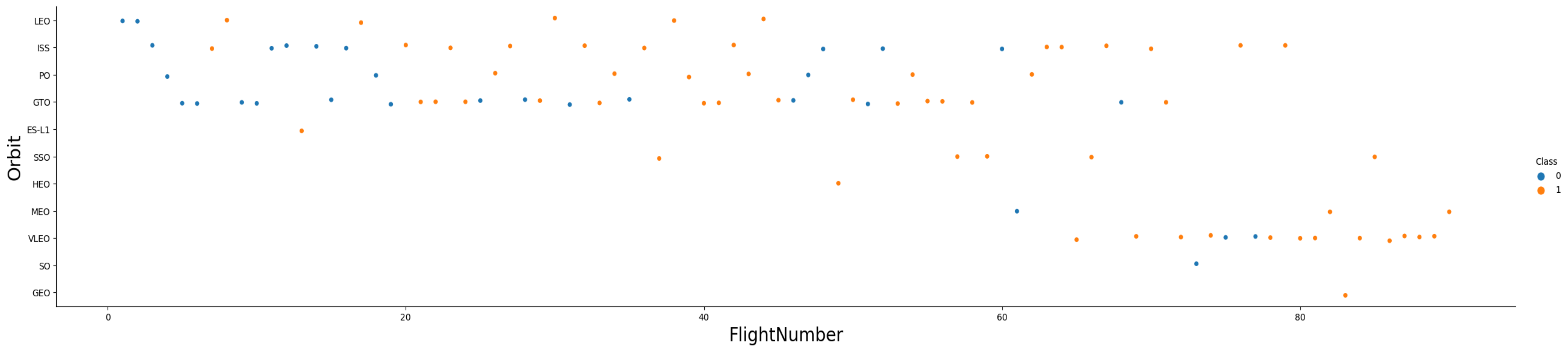
The next most successful orbits are:

- VLEO (above 80%); and
- LFO (above 70%).



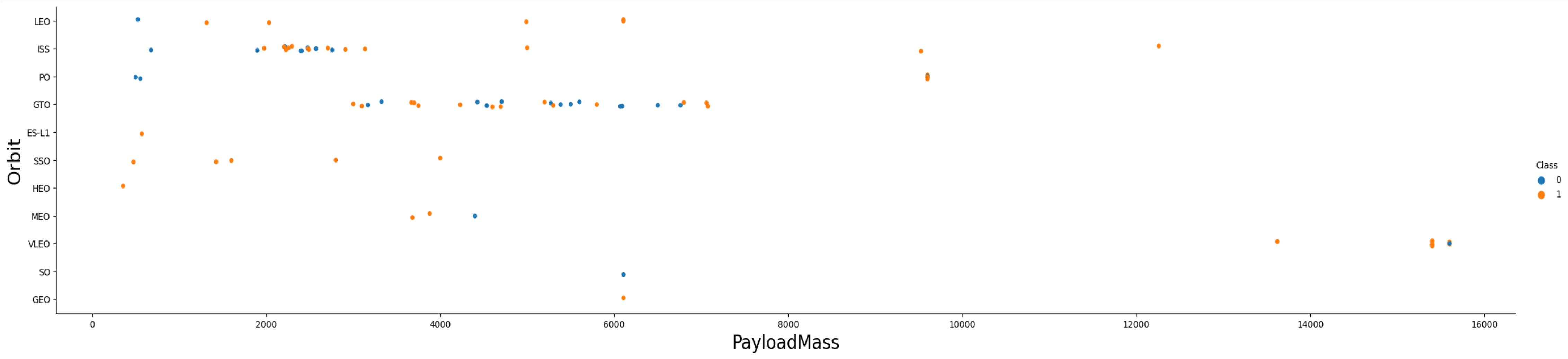
Flight Number vs. Orbit Type

- Success rates have seemingly improved over time for all orbits.
- The frequency of VLEO orbit launches has recently increased, indicating a potential new business opportunity.



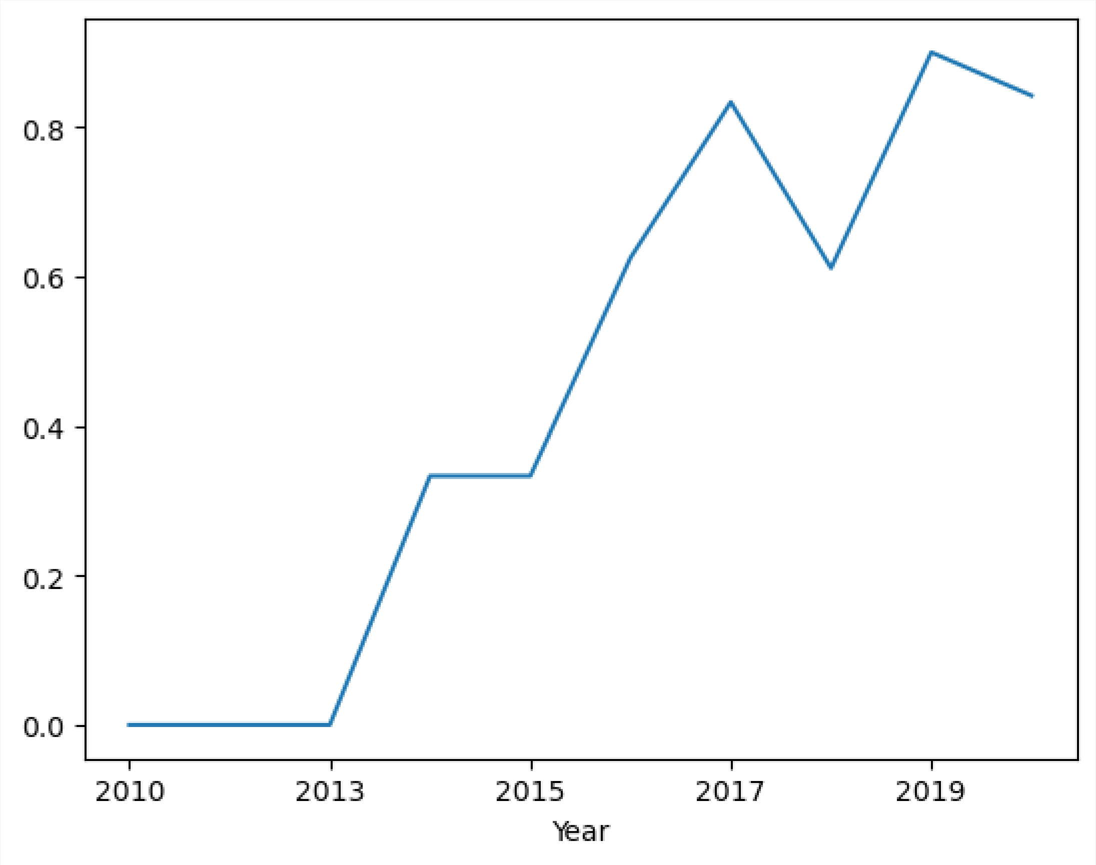
Payload vs. Orbit Type

- There appears to be no correlation between payload and success rate for the GTO orbit.
- The ISS orbit displays the broadest range of payloads and a commendable success rate.
- Launches to the SO and GEO orbits are relatively infrequent.



Launch Success Yearly Trend

- The success rate has been on an upward trend since 2013 and persisted until 2020.
- The initial three years appear to have been a period of adjustments and technological improvements.



All Launch Site Names

- There are four launch sites, which are determined by selecting the unique instances of "launch_site" values from the dataset.

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

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`
- The following displays five examples of Cape Canaveral launches.

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

Total Payload Mass

- The total payload transported by boosters from NASA:

Total Payload (kg)
111.268

- The overall payload was calculated by summing the payloads associated with codes containing 'CRS', which correspond to NASA.

Average Payload Mass by F9 v1.1

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

Avg Payload (kg)
2928

- After filtering the data to include only the specified booster version and calculating the average payload mass, it was determined to be 2,928 kg.

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad

Min Date
22/12/2015

- By filtering the data to include only successful landing outcomes on ground pads and identifying the minimum date value, the initial occurrence can be determined, which transpired on 22/12/2015

Successful Drone Ship Landing with Payload between 4000 and 6000

- The boosters that have successfully landed on drone ships and had payload masses greater than 4000 but less than 6000.

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

- After applying the aforementioned filters to select unique booster versions, the resulting count was 4.

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

- By grouping mission outcomes and tallying the corresponding number of records for each group, we arrived at the above summary.

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
F9 B5 B1051.4

Booster Version
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

- The following boosters transported the highest recorded payload masses in the dataset.

2015 Launch Records

- The booster versions and launch site names associated with failed landing outcomes on drone ships during the year 2015.

BoosterVersion	LaunchSite
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 from June 4, 2010, to March 20, 2017. This perspective on the data highlights the significance of taking "No attempt" into account

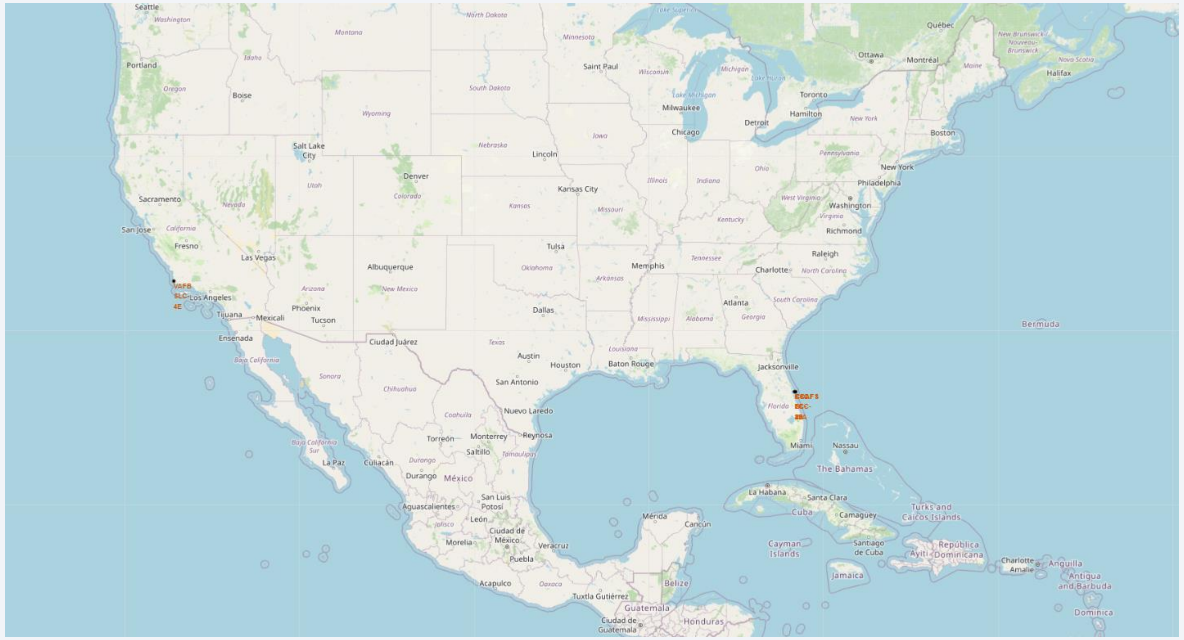
Landing Outcome	Occurrences
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled(ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

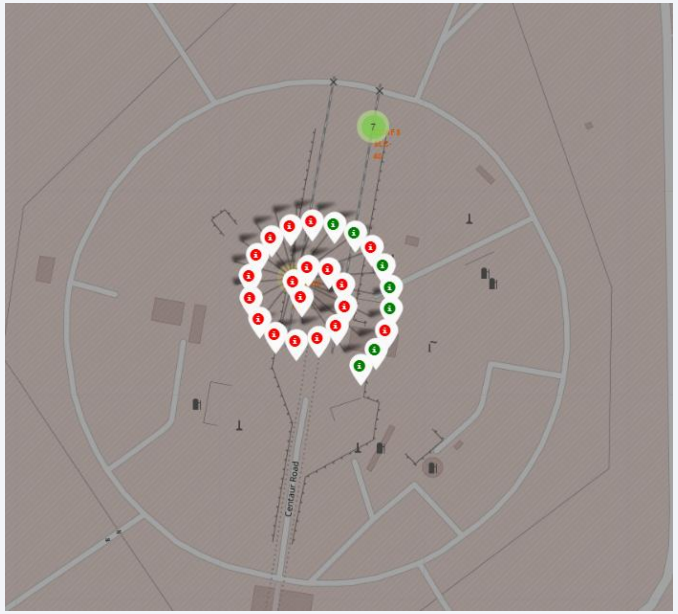
All launch sites

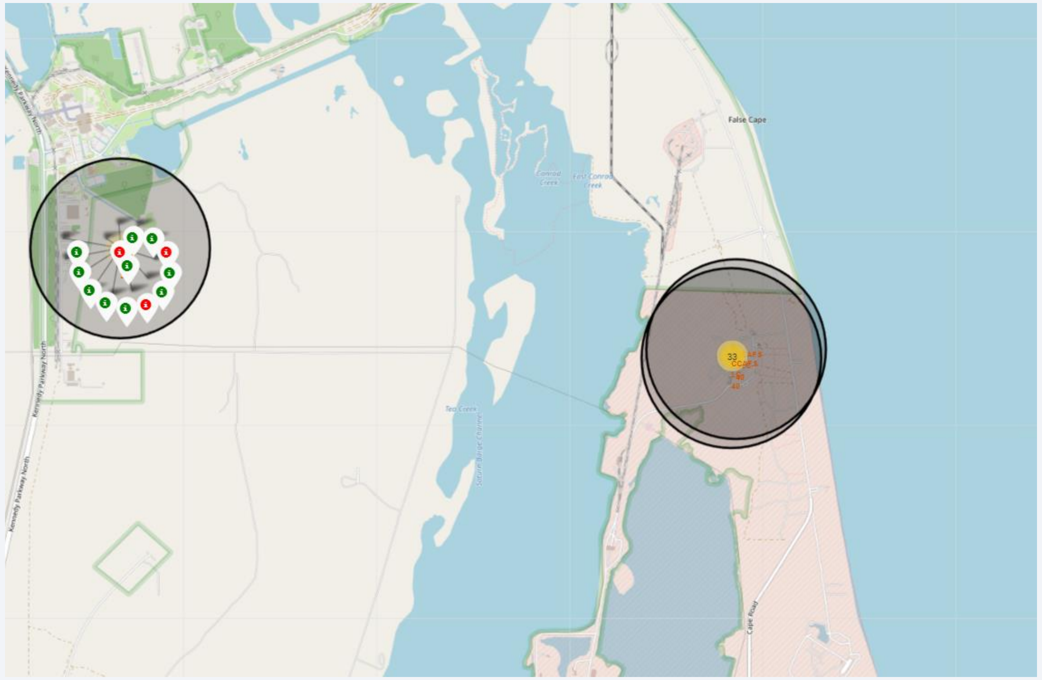
- Launch sites are situated in close proximity to bodies of water, most likely for safety reasons, while remaining accessible via roads and railroads.



Launch Outcomes by Site

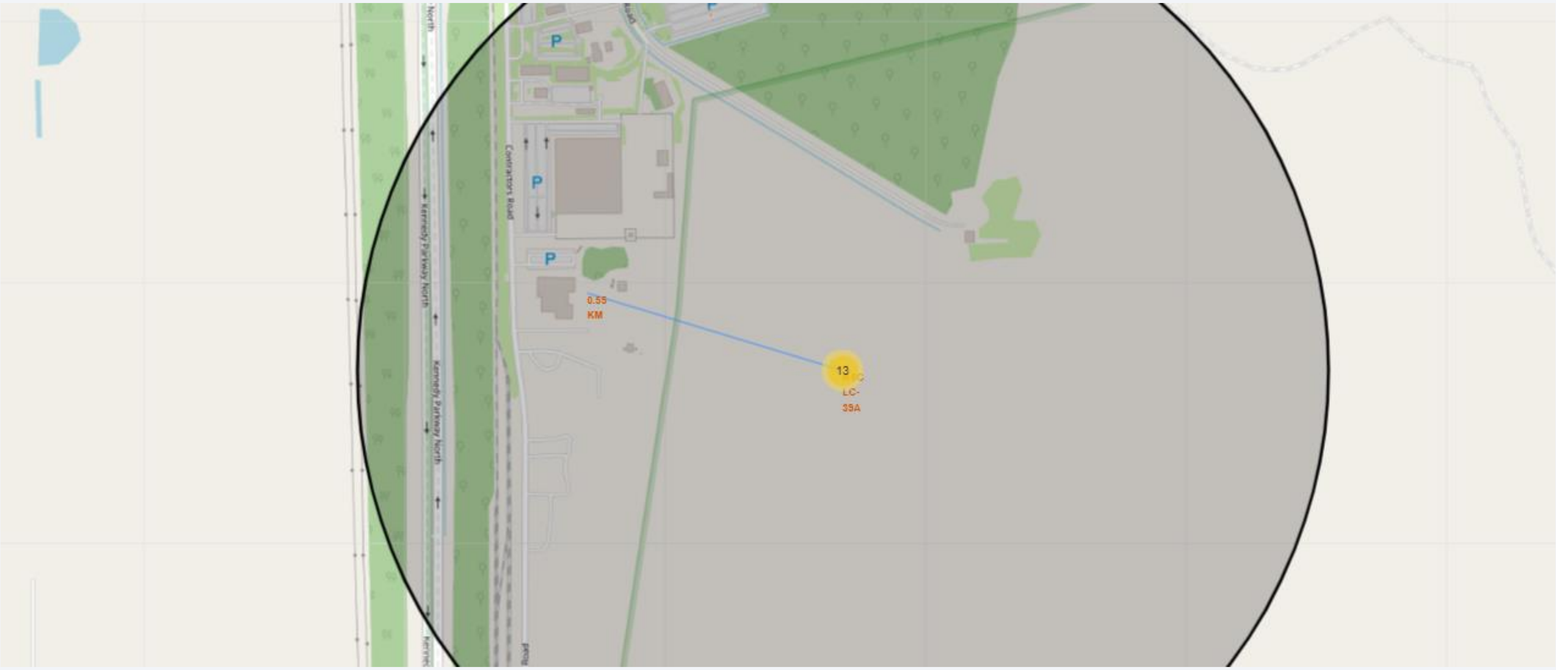
- Example of launch site launch outcomes





Launch Site Proximities: Analyzing Distances

- KSC LC 39A benefits from excellent logistics, given its strategic location near both rail and road networks, while also being situated at a safe distance from residential areas.

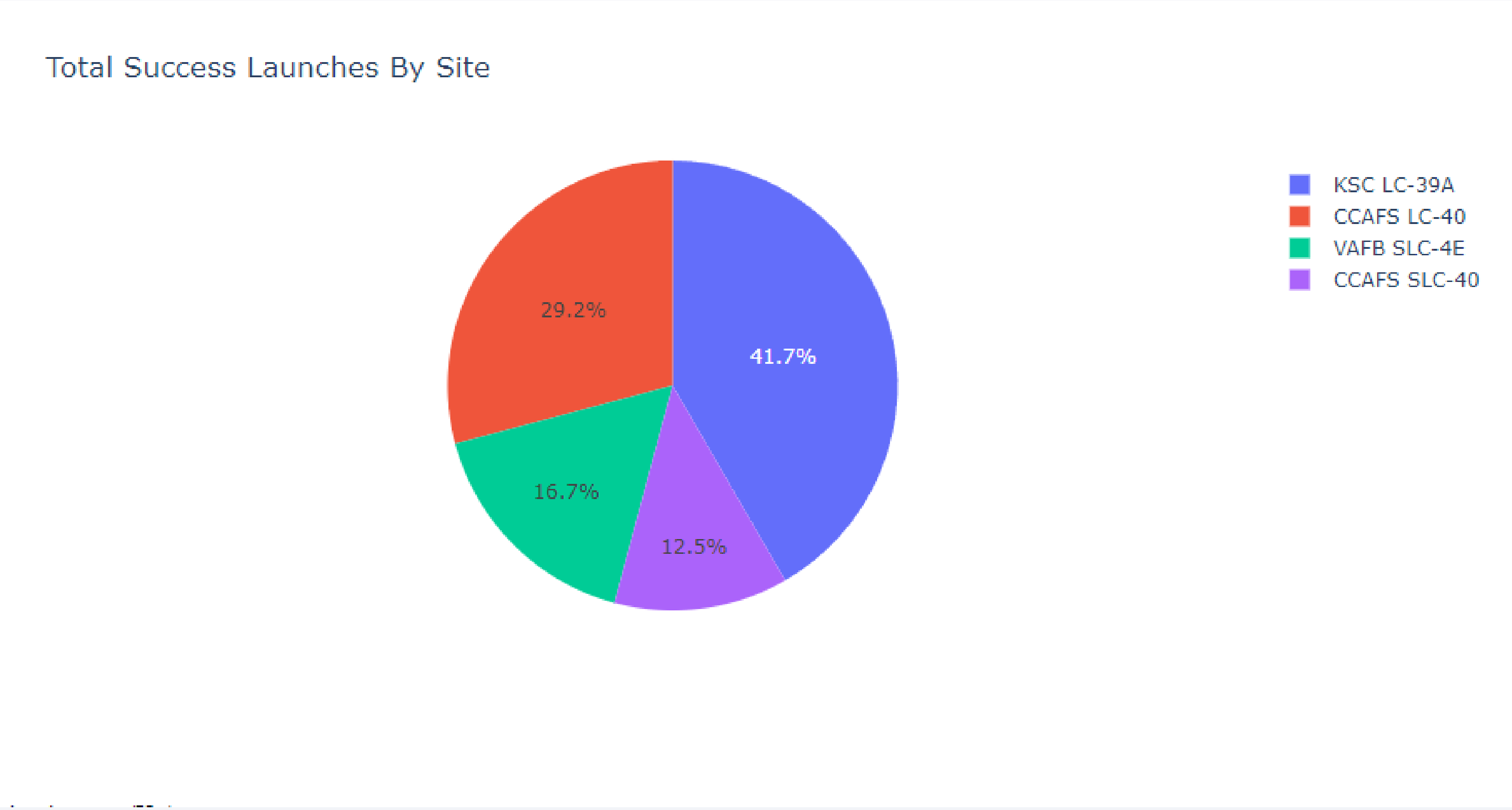


Section 4

Build a Dashboard with Plotly Dash

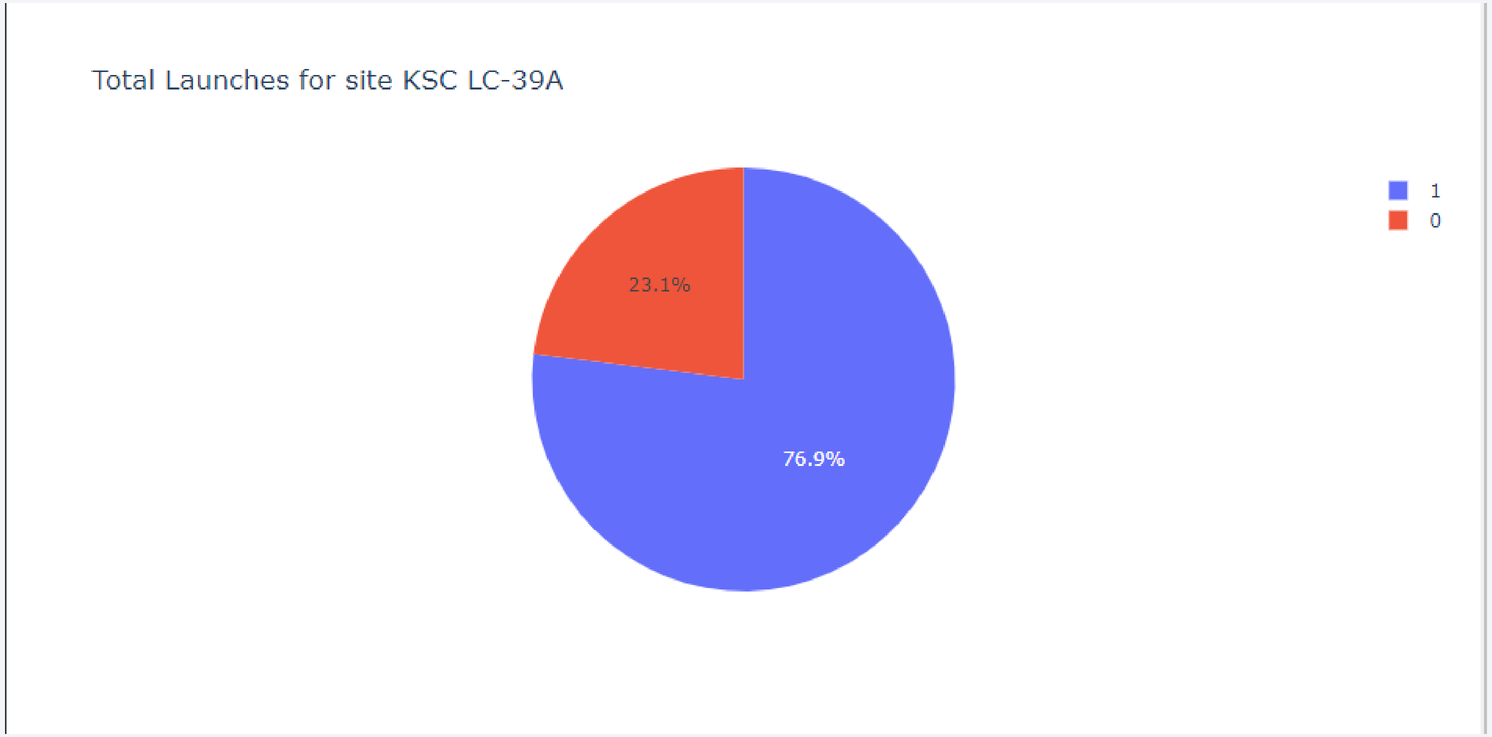
Achievement of Launch Sites

- The launch site appears to be a crucial factor in the success of missions.



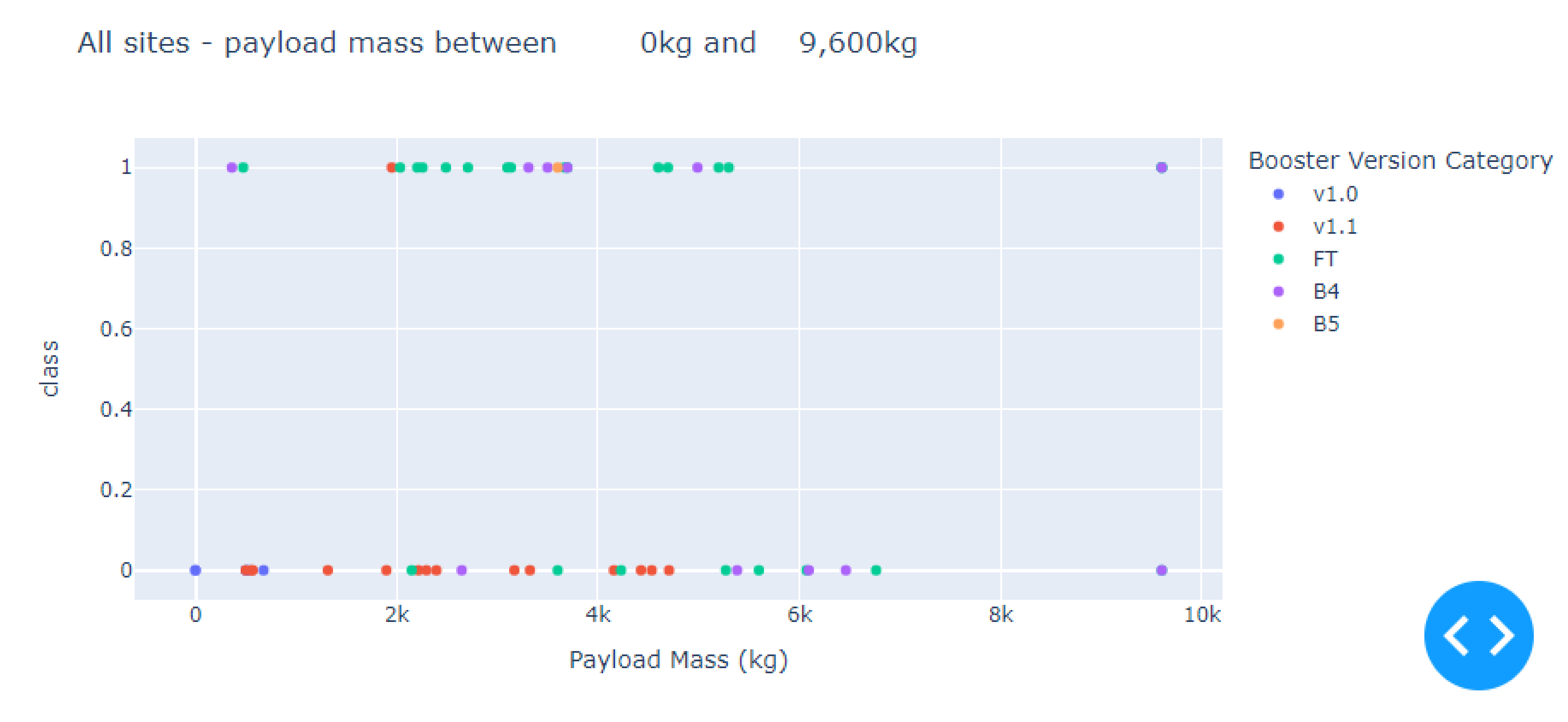
KSC LC 39A Launch Success

KSC LC 39A boasts a 76.9% success rate for launches.



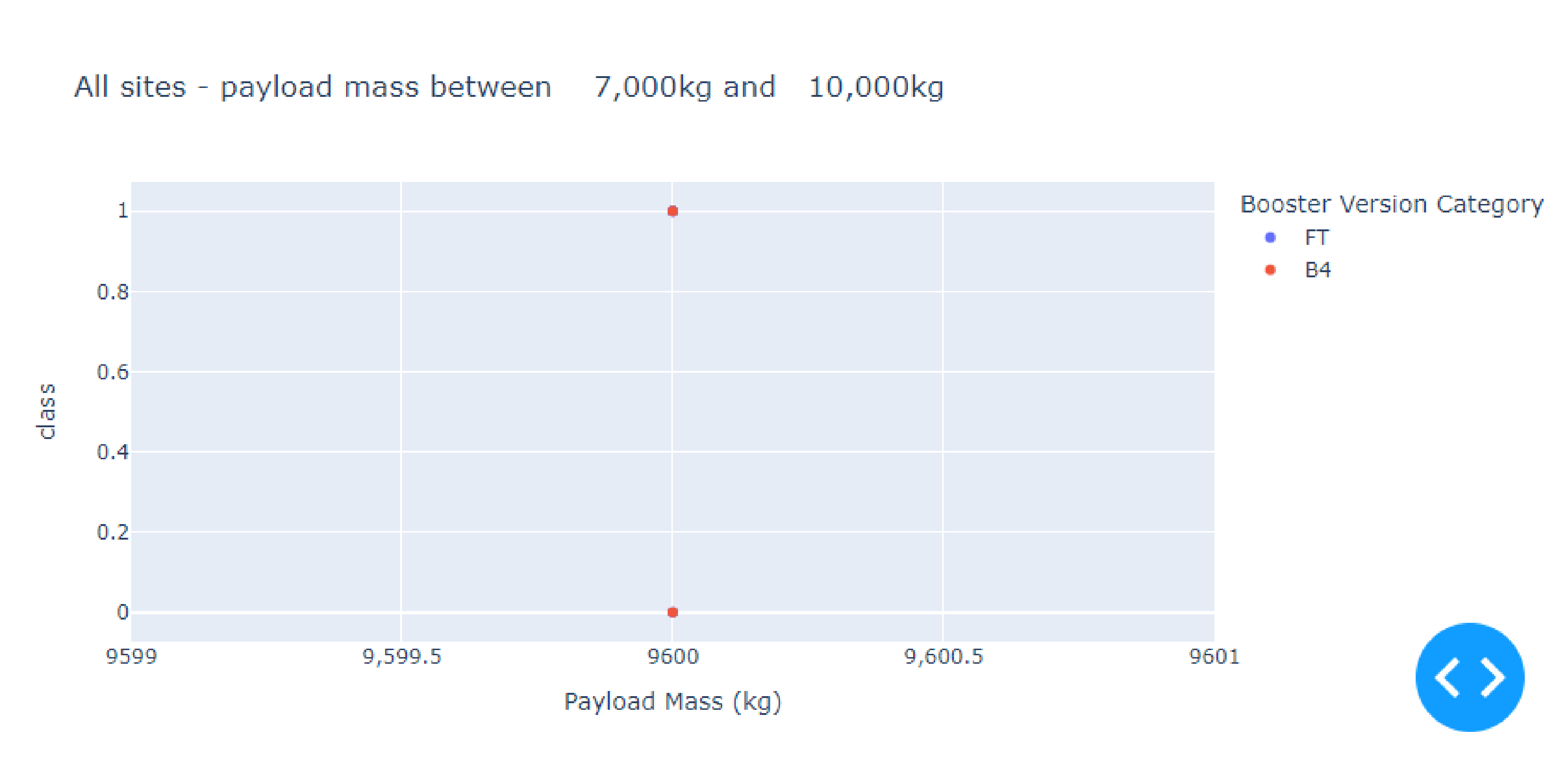
Relationship between Payload and Launch Result

- The most successful combination appears to be payloads under 6,000kg paired with FT boosters.



Relationship between Payload and Launch Result

- Not enough data of huger than 7000 kg.

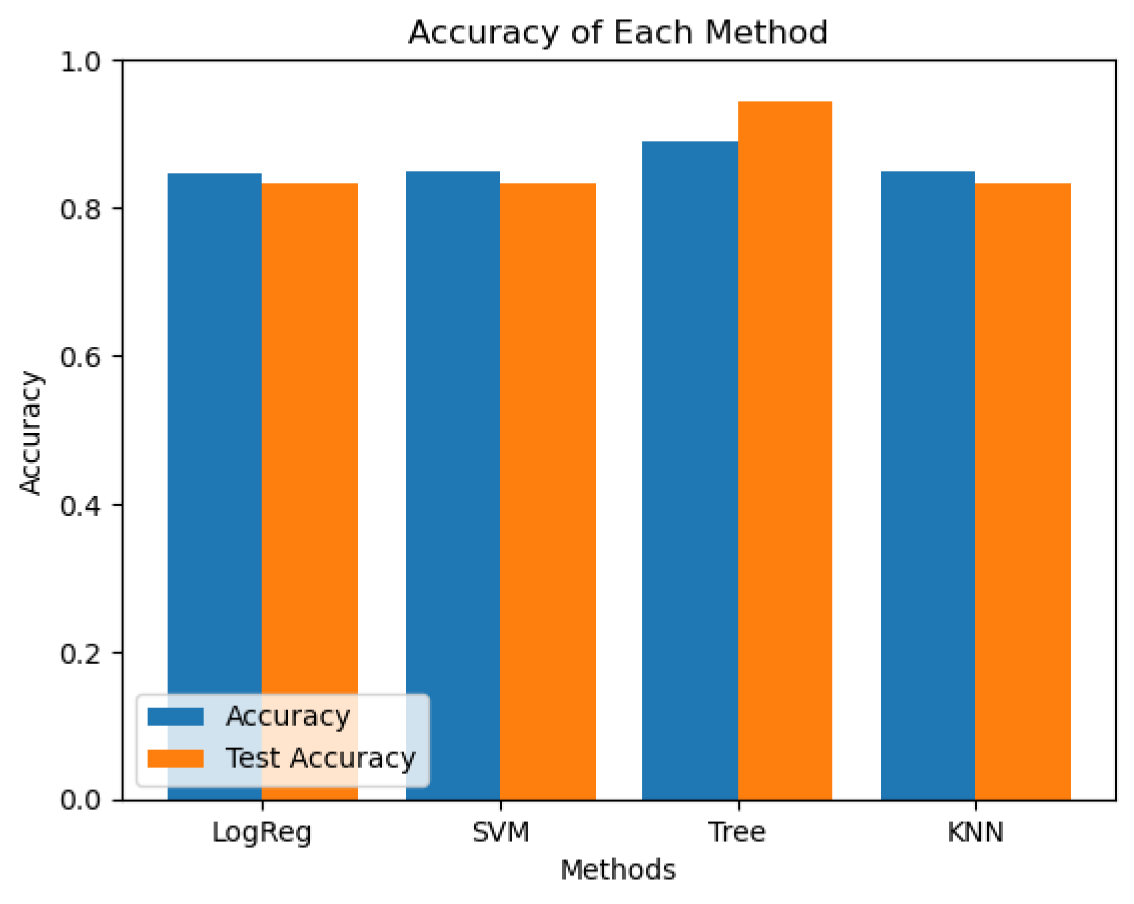


Section 5

Predictive Analysis (Classification)

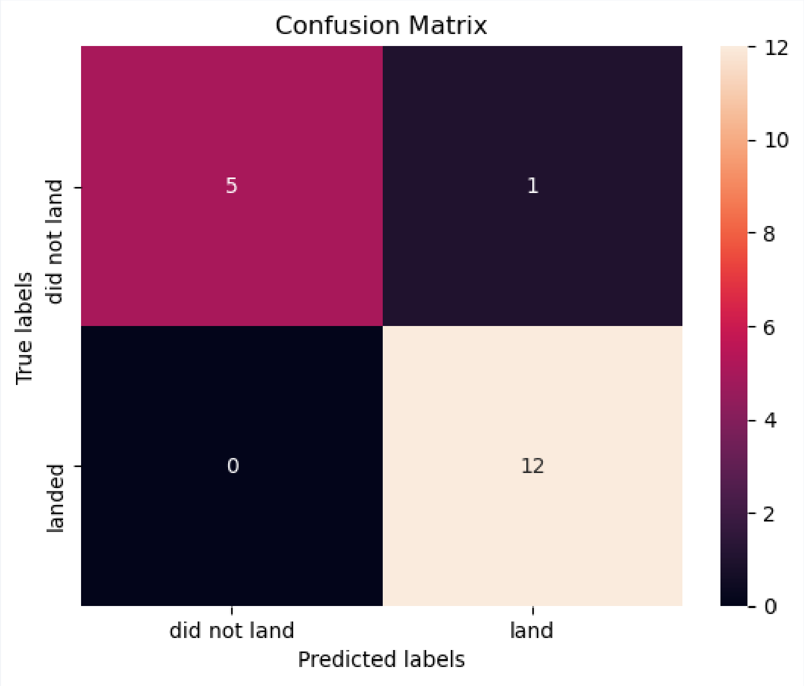
Classification Accuracy

- The accuracies of four classification models were tested and plotted. The Decision Tree Classifier emerged as the top-performing model, boasting accuracy levels exceeding 87%.



Confusion Matrix

- The accuracy of the Decision Tree Classifier is supported by its confusion matrix, which demonstrates a high number of true positive and true negative values in comparison to false values.



Conclusions

- Diverse data sources examined for improved insights
- KSC LC 39A highlighted as best launch site
- Payloads beyond 7,000 kg regarded as lower risk
- Most mission outcomes found to be successful
- Incremental enhancement in landing success rates attributed to advancements in technology and methods
- Decision Tree Classifier deemed valuable for forecasting triumphant landings and boosting revenue

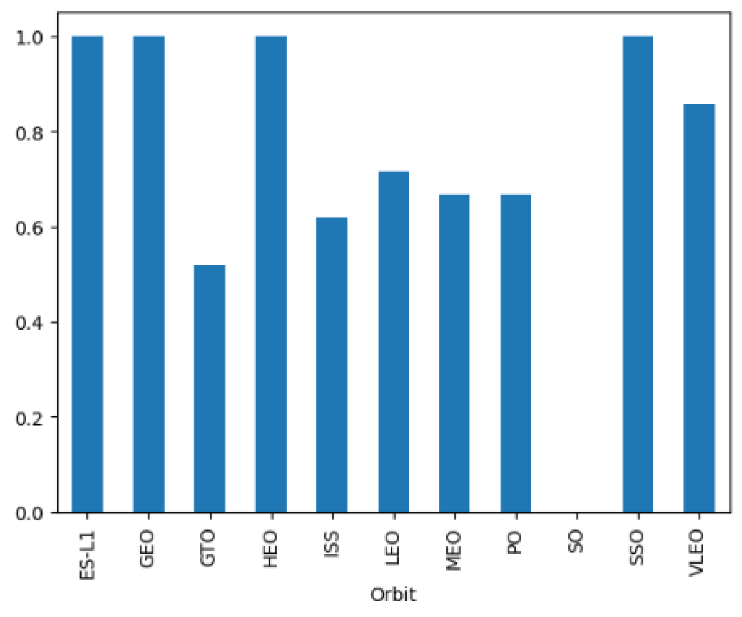
Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

```
# HINT use groupby method on Orbit column and get the mean of Class column
df.groupby('Orbit')['Class'].mean().plot.bar()
```

✓ 0.1s

<AxesSubplot: xlabel='Orbit'>



Thank you!