IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Ali Fadlalla Ali
February-2022

Outline

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

Executive Summary

- The data analysis process involved the utilization of several methodologies:
  - Data Collection via web scraping and SpaceX API.
  - Exploratory Data Analysis (EDA) incorporating data wrangling, visualization and interactive visual. analytics.
  - Machine Learning Prediction.
- Summary of results:
  - Valuable data obtained from public sources.
  - EDA identified key features affecting launch success.
  - Machine Learning Prediction demonstrated optimal model for predicting important characteristics to drive opportunity using collected data.

Introduction

- The goal is to see if a new company called Space Y can compete with Space X. To do this, we want to:
  - Figure out the cost of launches by predicting if the first part of the rocket will land safely.
  - Find the best place to do launches.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Gathering data on Space X from two sources: the Space X API "https://api.spacexdata.com/v4/rockets".
  - Web scraping https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches.
  - Perform data wrangling:
  - The gathered information was improved by generating a landing result classification using the result information, once the features have been distilled and examined.

Methodology

Executive Summary

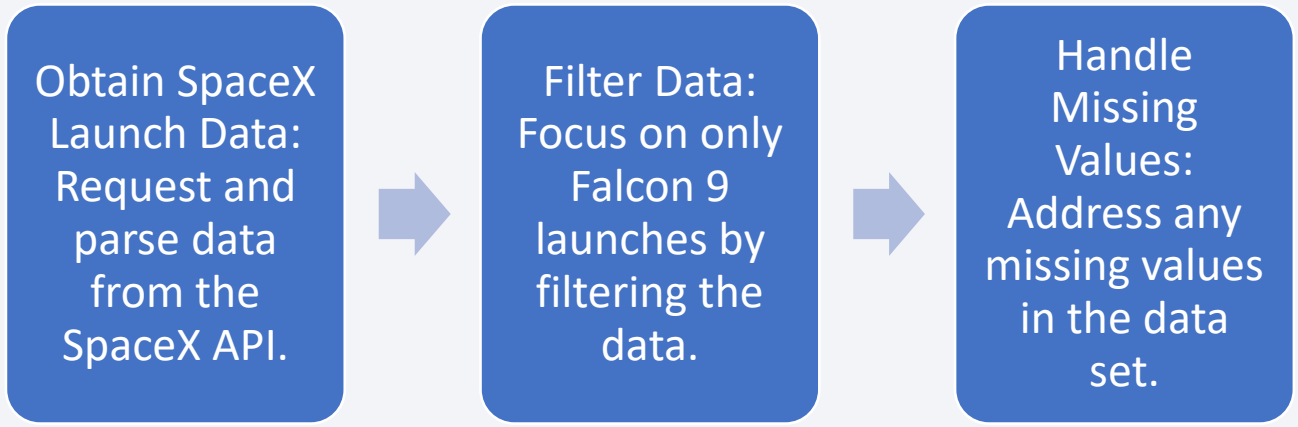
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models:
    - The data that had been collected underwent normalization, was divided into training and test datasets, and was evaluated using four different classification models. The accuracy of each model was then assessed using various parameter combinations.

Data Collection

- The data sets were obtained from the 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 techniques. The data collection process involved the following steps:
  - Identification of relevant sources (Space X API and Wikipedia)
  - Utilization of web scraping techniques to extract the data from these sources
  - Storage of the collected data for further analysis and processing.

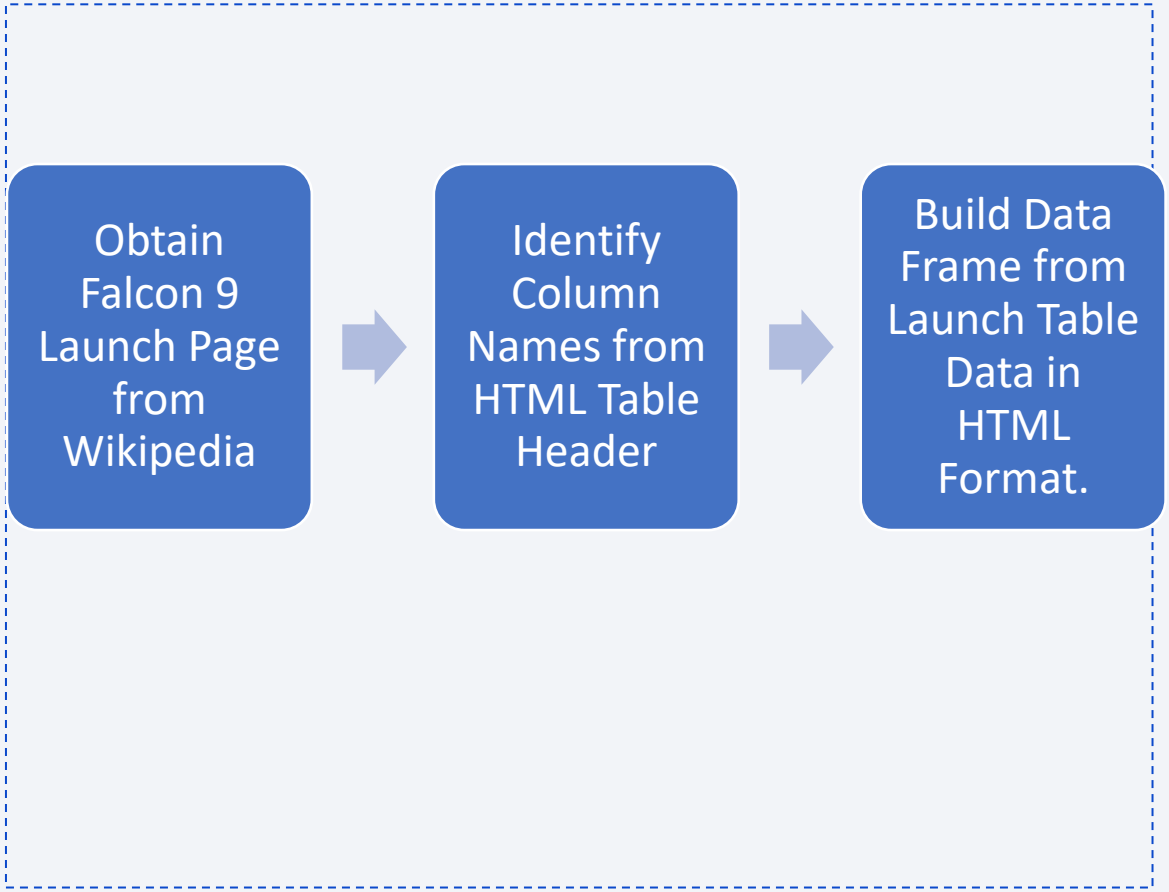
Data Collection – SpaceX API

- SpaceX provides a public API as a source of data that can be obtained and utilized.
- The API was accessed following the flowchart described, and the obtained data was saved. The source code for this process can be found at the following GitHub URL:
  - https://github.com/alifadl009/Applied-Data-Science/blob/main/Data%20Collection%20API.ipynb



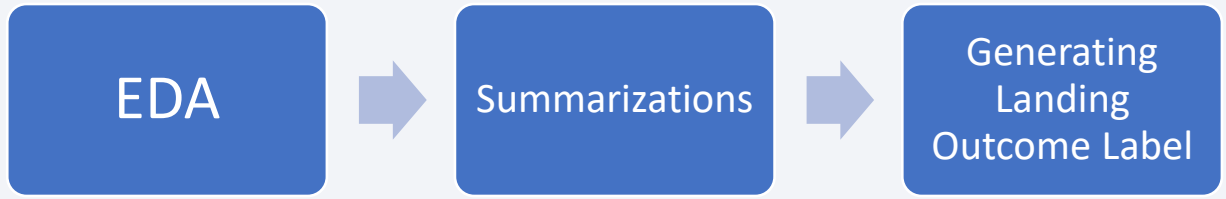
Data Collection - Scraping

- Acquire SpaceX Launch Information: Gather data on SpaceX launches from Wikipedia.
- Store Data: Store the obtained data for future use.
- Source code:
  - https://github.com/alifadl009/Applied-Data-Science/blob/main/Data%20Collection%20with%20Web%20Scraping.ipynb



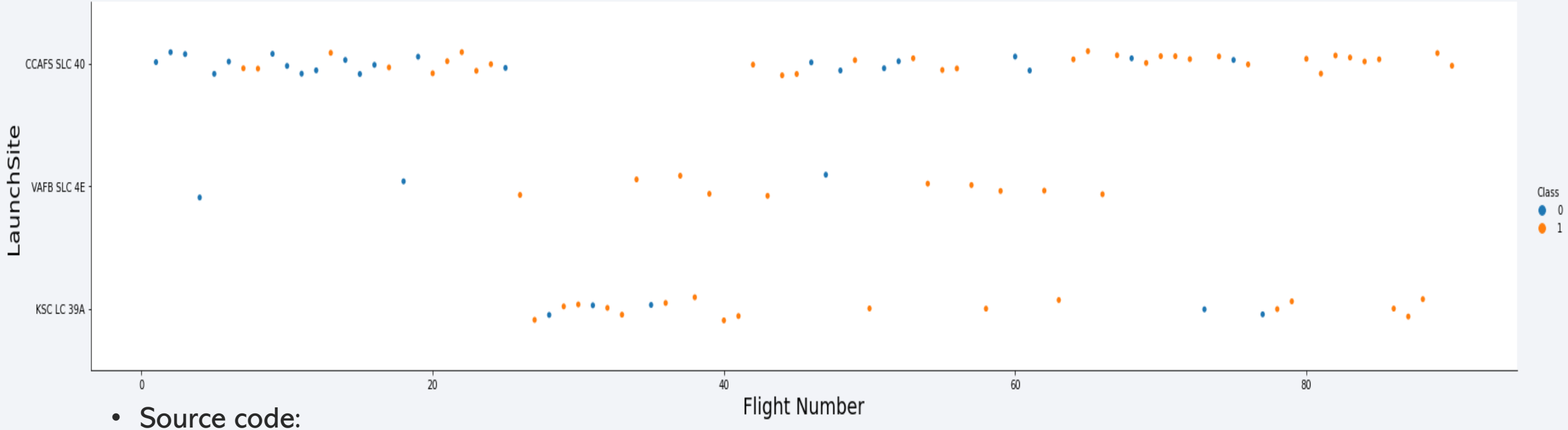
Data Wrangling

- The first step in the analysis of the data was conducting an Exploratory Data Analysis (EDA).
- Then calculating the number of launches per site, the frequency of each orbit, and the frequency of mission outcomes for each orbit type.
- The final step was creating a landing outcome label from the Outcome column data.
- Source code:
  - https://github.com/alifadl009/Applied-Data-Science/blob/main/Data%20Wrangling.ipynb



EDA with Data Visualization

- Scatterplots and barplots were utilized to examine relationships between various feature pairs. This included: Payload Mass and Flight Number, Launch Site and Flight Number, Launch Site and Payload Mass, Orbit and Flight Number, and Payload and Orbit.



- Source code: https://github.com/alifadl009/AppliedDataScience/blob/main/EDA%20with%20Data%20Visualization.ipynb

EDA with SQL

- The following SQL queries were performed on the SpaceX launch data:
  - Names of unique launch sites.
  - Top 5 launch sites starting with "CCA“.
  - Payload mass carried by NASA (CRS).
  - Average payload mass carried by F9 v1.1 boosters.
  - Date of first successful landing on ground pad.
  - Successful drone ship boosters with payload between 4000 and 6000 kg.
  - Count of successful and failure missions.
  - Boosters with highest payload mass.
  - Failed landing outcomes on drone ship in 2015, with booster version and launch site names.
  - Rank of landing outcome counts from 2010-06-04 to 2017-03-20.

- Source code: “https://github.com/alifadl009/Applied-Data-Science/blob/main/EDA.ipynb”

Build an Interactive Map with Folium

- Folium Maps were utilized with markers, circles, lines, and marker clusters.
- Markers were used to represent launch sites.
- Circles were used to highlight areas around specific coordinates, like the NASA Johnson Space Center.
- Marker clusters were used to group events in each coordinate, such as launches at a launch site.
- Lines were used to show the distances between two coordinates.

- Source code: "https://github.com/alifadl009/Applied-Data-Science/blob/main/Interactive%20Visual%20Analytics%20with%20Folium%20lab.ipynb"

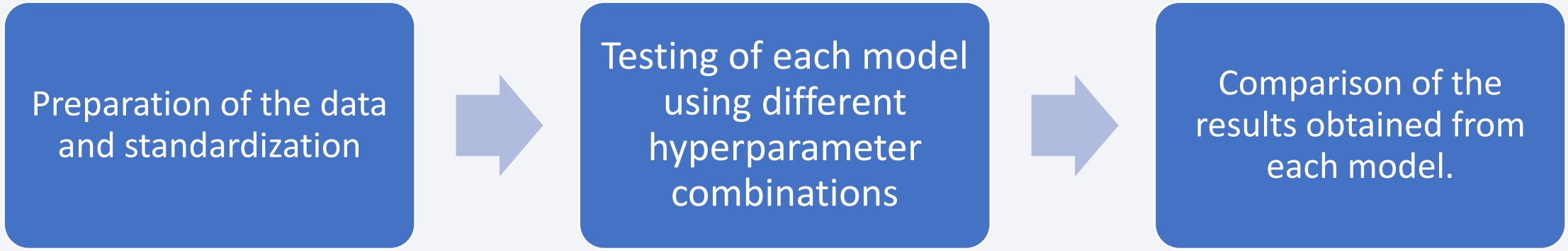
Build a Dashboard with Plotly Dash

- The data was visualized through graphs and plots, including:
  - A representation of the proportion of launches at each site.
  - A display of the range of payloads.
  - These visualizations enabled a fast examination of the relationship between payloads and launch sites, aiding in the discovery of the optimal launch location based on payloads.

- Source code: “https://github.com/alifadl009/Applied-Data-Science/blob/main/spacex_dash_app.py”

Predictive Analysis (Classification)

- Four classification models were compared and evaluated: Logistic Regression, Support Vector Machine, Decision Tree, and k-Nearest Neighbors.



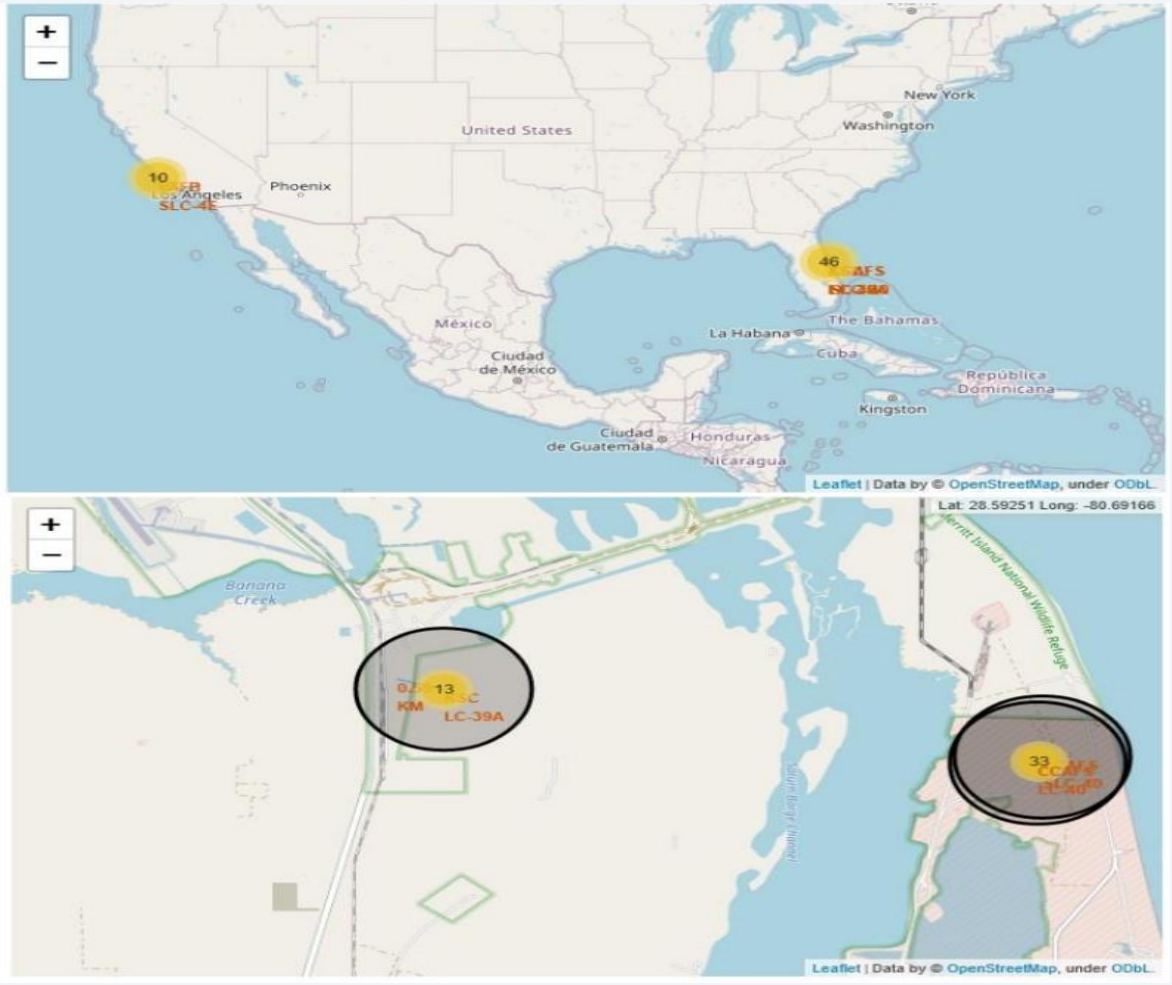
Source code: "https://github.com/alifadl009/Applied-Data-Science/blob/main/Machine%20Learning%20Prediction.ipynb"

Results

- SpaceX has utilized four distinct launch facilities throughout its operations.
- The early launches performed by the company were primarily for internal and NASA purposes.
- The average payload capacity of the F9 v1.1 booster is recorded to be 2,928 kilograms.
- The first recorded instance of a successful landing outcome took place in 2015, which was five years after the initiation of launches.
- A considerable number of Falcon 9 booster versions have demonstrated successful landing outcomes on drone ships, with payloads exceeding the average.
- A near-complete rate of successful mission outcomes has been achieved by SpaceX.
- In 2015, two specific booster versions, namely F9 v1.1 B1012 and F9 v1.1 B1015, were recorded as having failed landing outcomes on drone ships.
- A trend of improvement in the number of successful landing outcomes has been observed over time.

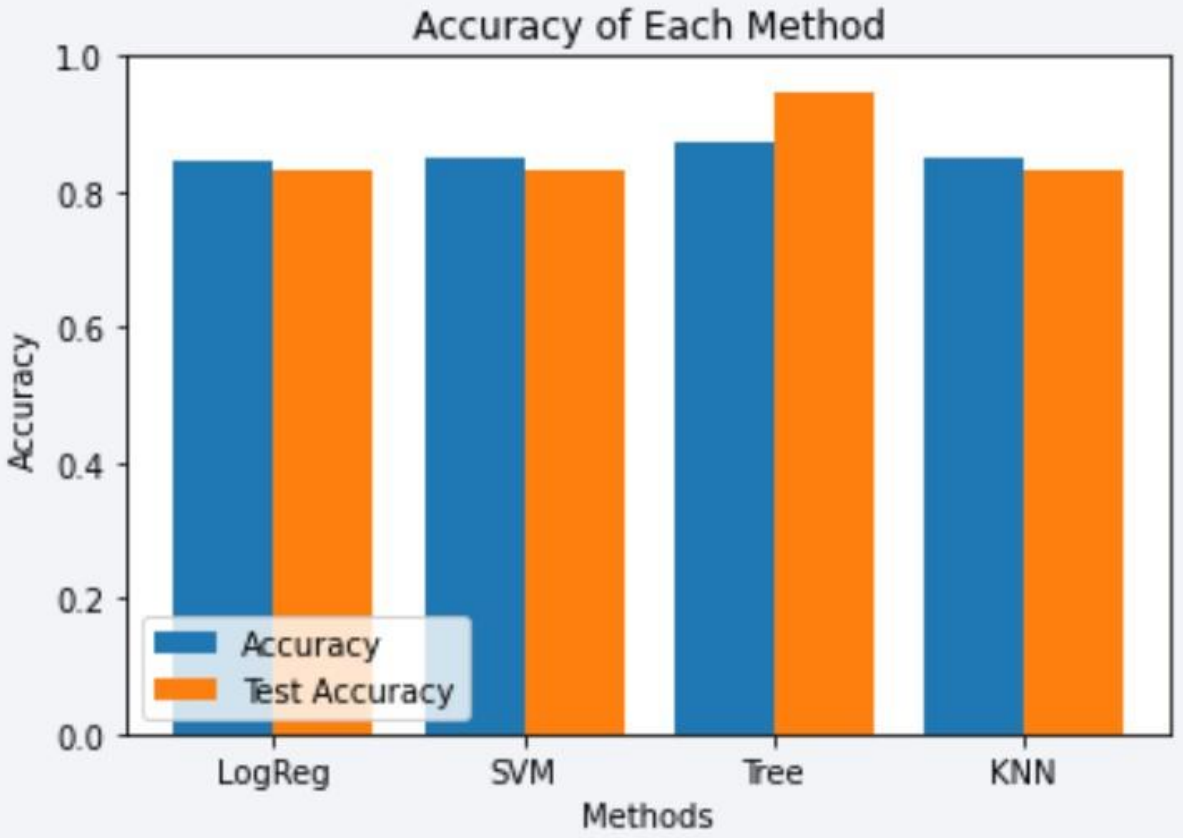
Results

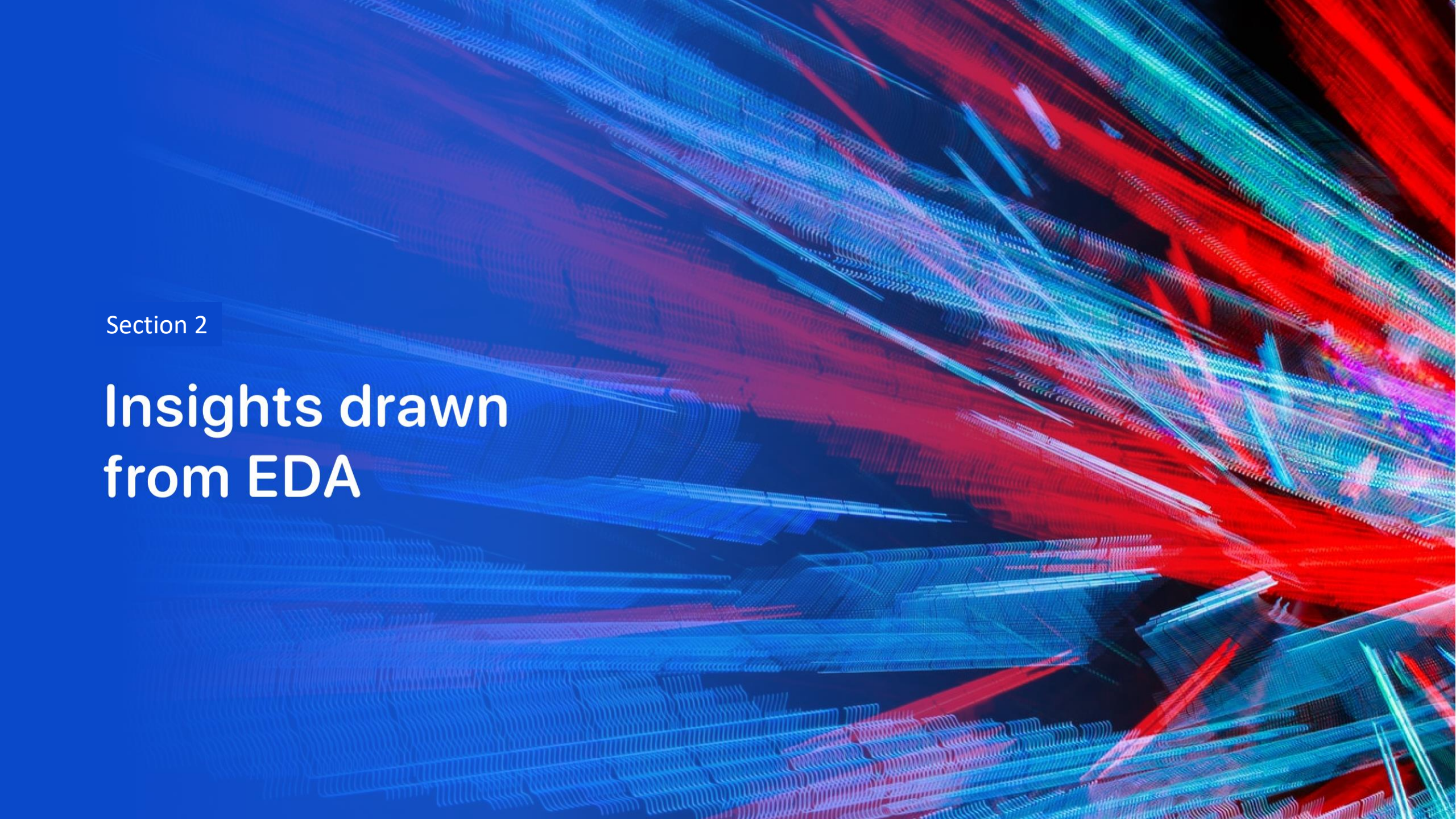
- The utilization of interactive analytics has highlighted the strategic placement of SpaceX's launch sites in safe areas near bodies of water, with substantial logistical support.
- The east coast launch facilities of the company have been the primary site of launch operations.



Results

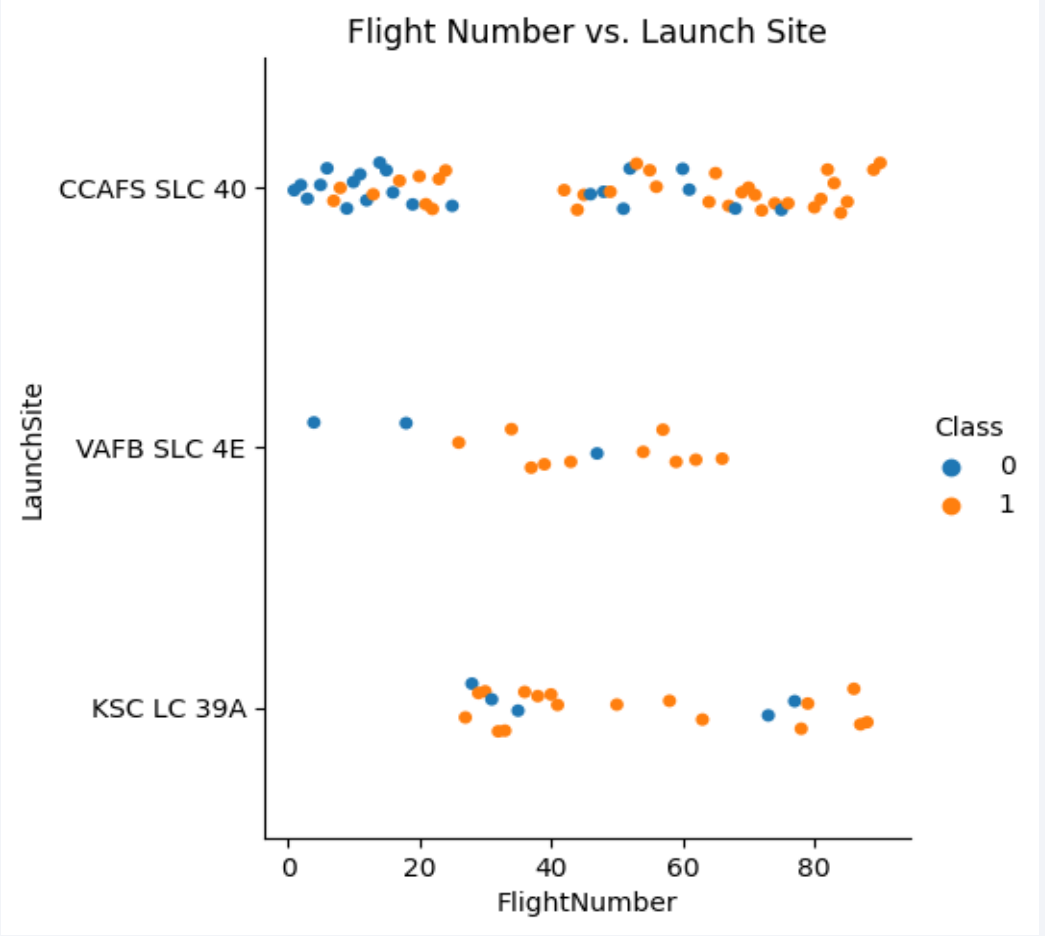
- The results of the Predictive Analysis revealed that the Decision Tree Classifier is the most suitable model for predicting successful landings, exhibiting an accuracy rate of over 87%.




Section 2
Insights drawn from EDA

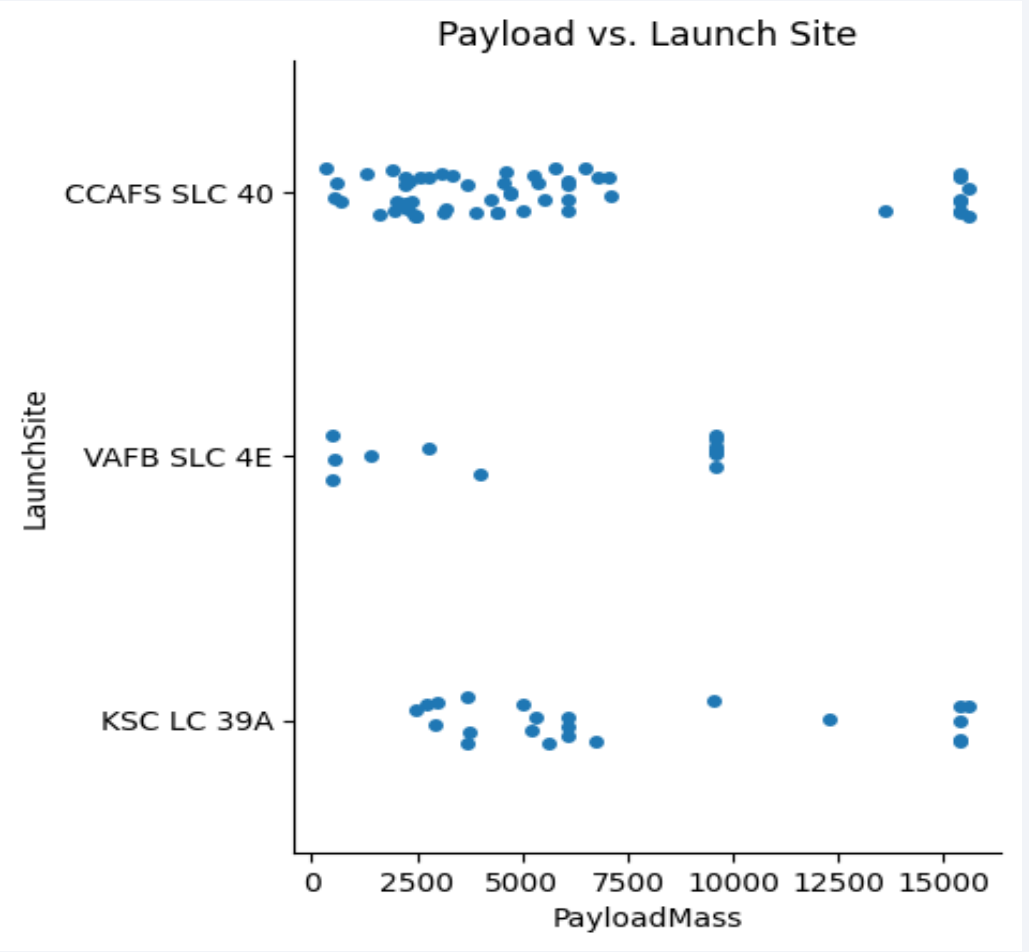
Flight Number vs. Launch Site

- The plot indicates that the most effective launch site at present is the CCAF5 SLC 40, where the majority of recent launches have been successful.
- The VAFB SLC 4E is the second-best site, followed by KSC LC 39A in third place.
- Additionally, the overall success rate of launches has shown an upward trend over time.



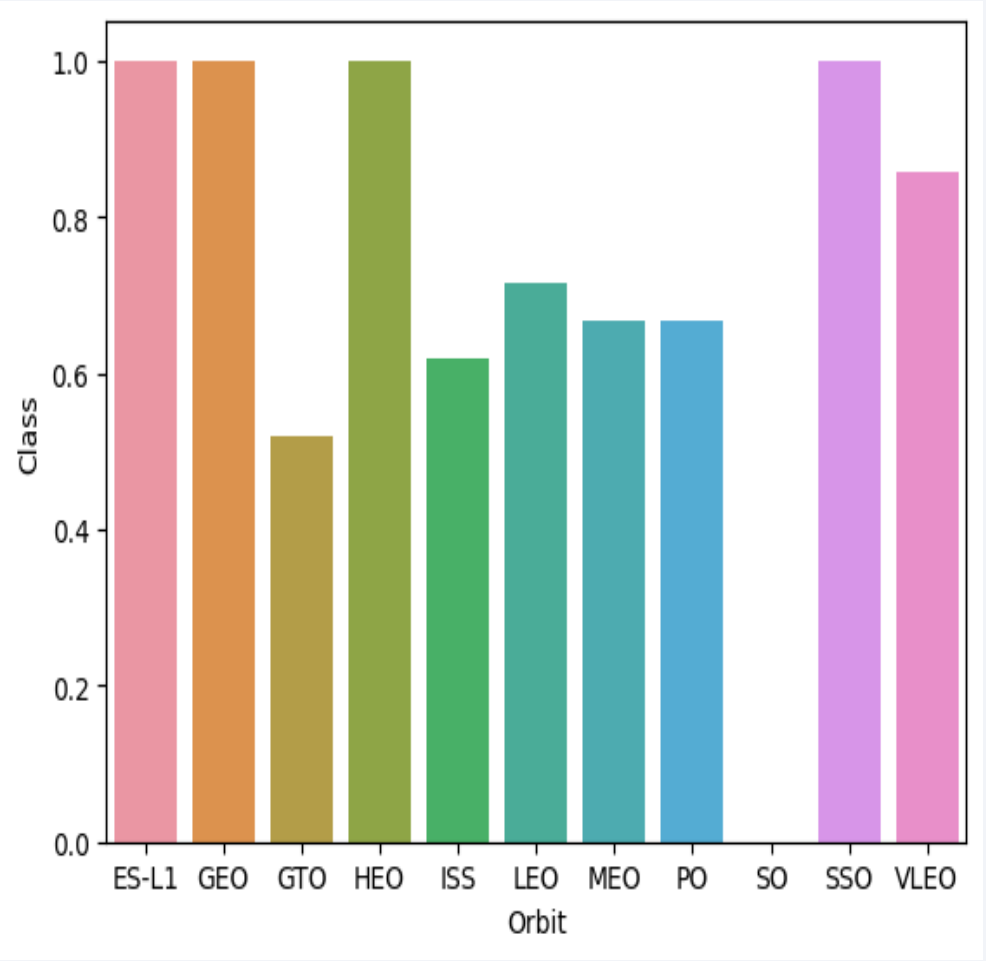
Payload vs. Launch Site

- The analysis shows that payloads exceeding 9,000 kilograms (equivalent to the weight of a school bus) demonstrate a high rate of successful outcomes.
- It appears that payloads over 12,000 kilograms can only be launched from the CCAFS SLC 40 and KSC LC 39A launch sites.



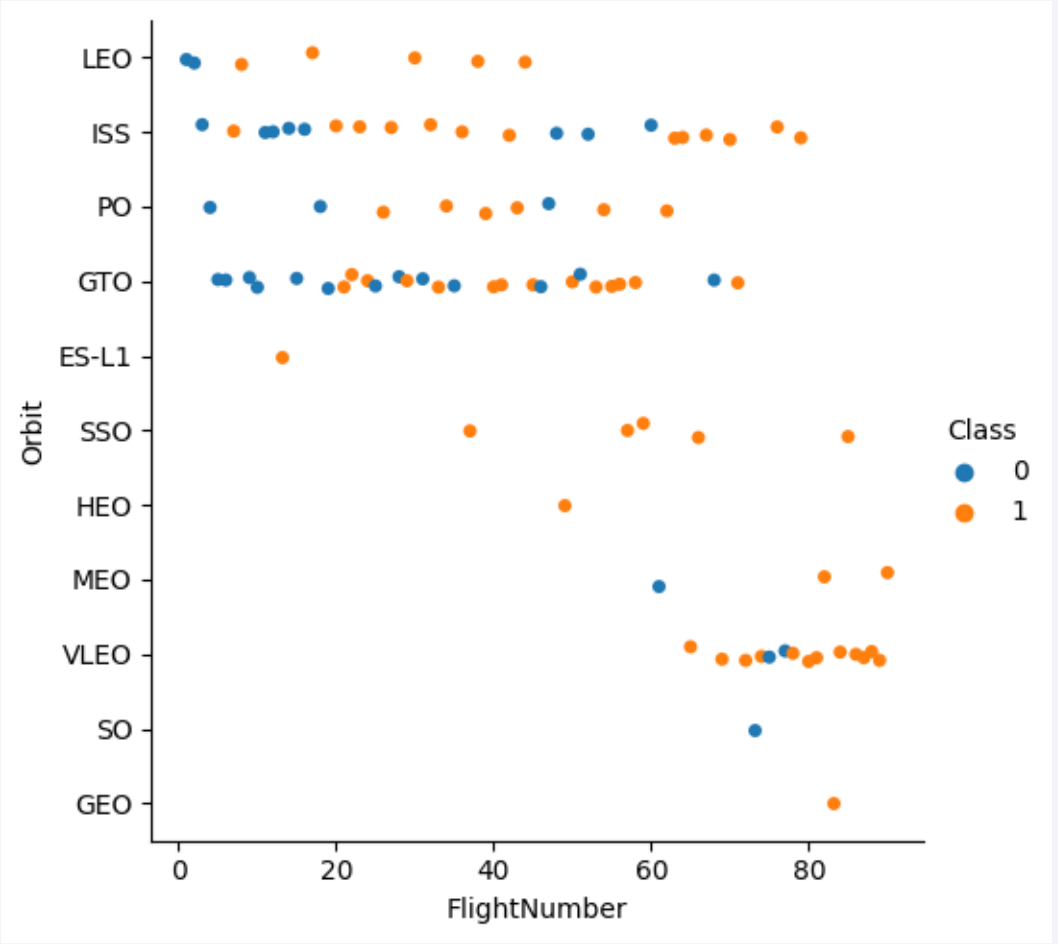
Success Rate vs. Orbit Type

- The highest success rates have been recorded for the following orbits: ES-L1, GEO, HEO, and SSO. Subsequently, the VLEO orbit has exhibited a success rate of above 80%, and the LFO orbit has shown a success rate of above 70%.



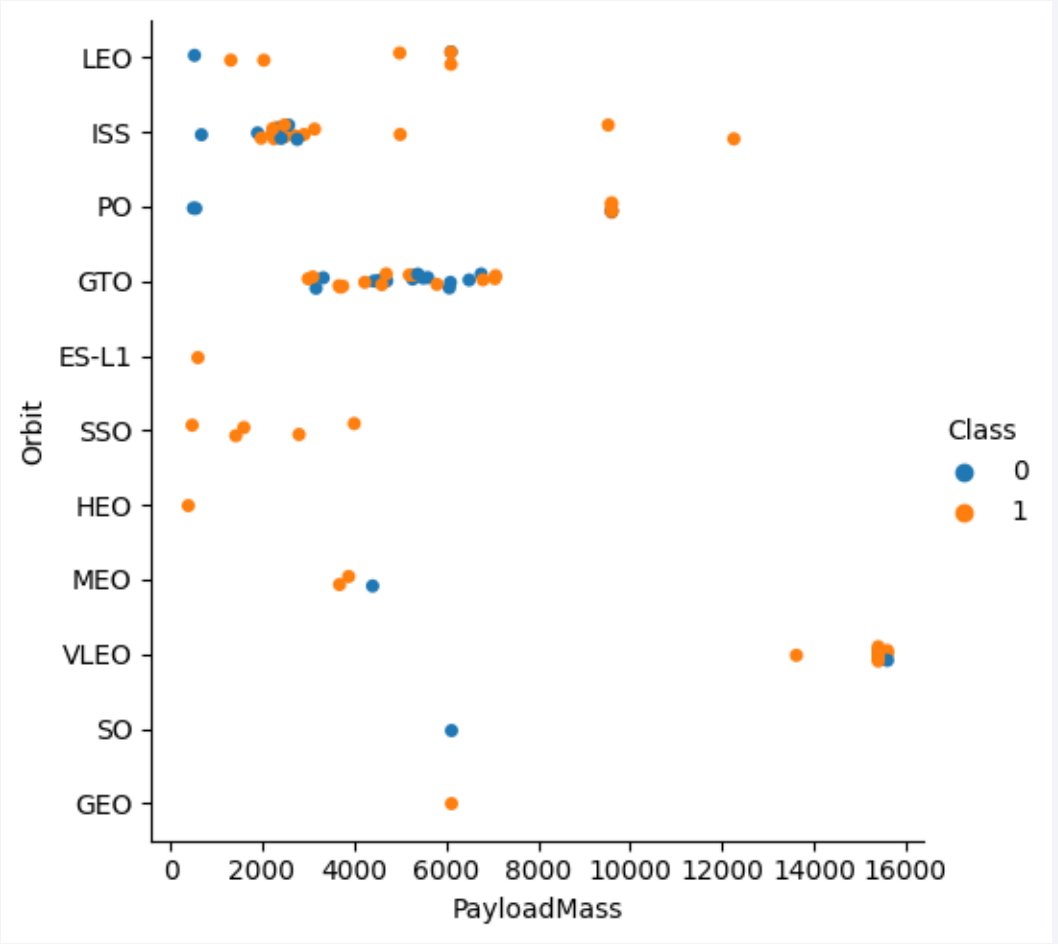
Flight Number vs. Orbit Type

- The analysis suggests that the success rate has improved over time for all orbital categories.
- Additionally, the recent increase in frequency of launches in the VLEO orbit presents a potential new business opportunity for the company.



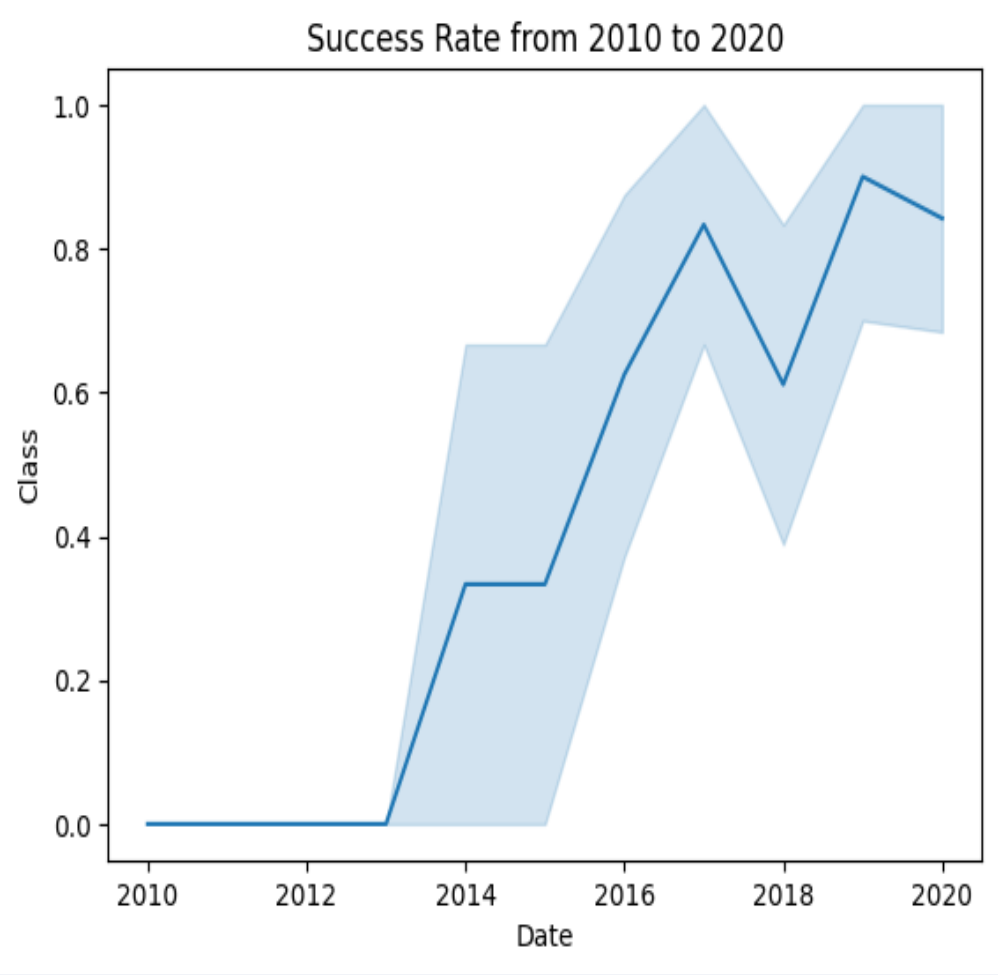
Payload vs. Orbit Type

- The analysis has indicated a lack of correlation between payload and success rate for launches to the Geostationary Transfer Orbit (GTO).
- The International Space Station (ISS) orbit has a substantial range of payload capacity and a high rate of successful missions.
- Relatively few launches have been conducted to the Sun-Synchronous Orbit (SO) and Geostationary Orbit (GEO).



Launch Success Yearly Trend

- The success rate of landing outcomes has shown an upward trend beginning in 2013 and sustained until 2020. This suggests that the initial three years were a phase of adjustment and improvement of technology.



All Launch Site Names

- The data analysis has identified four distinct launch facilities, as determined through the process of extracting unique values of "launch_site" from the dataset.

Launch_Site	COUNT(Launch_Site)
CCAFS LC-40	26
CCAFS SLC-40	34
KSC LC-39A	25
VAFB SLC-4E	16

Launch Site Names Begin with 'CCA'

- The analysis showcases five samples of launches originating from Cape Canaveral.

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS 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_mass
45596

- Total payload mass carried by boosters launched by NASA (CRS).

Average Payload Mass by F9 v1.1

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

average_payload_mass
2534.6666666666665

- The analysis involved filtering the data based on the booster version and calculating the average payload mass.

First Successful Ground Landing Date

- The data analysis process involved filtering the data by the criterion of successful landing outcomes on the ground pad and determining the minimum value of the date field. This resulted in the identification of the first instance of a successful landing outcome, which occurred on December 22, 2015.

Successful Drone Ship Landing with Payload between 4000 and 6000

- The data analysis process involved filtering boosters based on the criteria of successful landing outcomes on drone ships and payload masses that were greater than 4,000 kilograms but less than 6,000 kilograms. The result of this analysis was the identification of four distinct booster versions that met these criteria.

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

Total Number of Successful and Failure Mission Outcomes

- The data analysis involved grouping mission outcomes and counting the number of records for each group, resulting in a summary of the number of successful and failed mission outcomes.

Mission_Outcome	Number_of_Missions
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- Names of the booster which have carried the maximum payload mass

- The data analysis aimed at identifying the boosters that carried the maximum payload mass, resulting in the identification of the boosters that had the largest payload capacity.

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

2015 Launch Records

- The following is a list of unsuccessful drone ship landing outcomes, corresponding booster versions, and launch site names in 2015:

month	Booster_Version	Launch_Site
01	F9 v1.1 B1012	CCAFS LC-40
04	F9 v1.1 B1015	CCAFS LC-40

- Present your query result with a short explanation here

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

- Here is the ranking of the count of landing outcomes (such as failure on drone ship or success on ground pad) between the date range of 2010-06-04 to 2017-03-20.

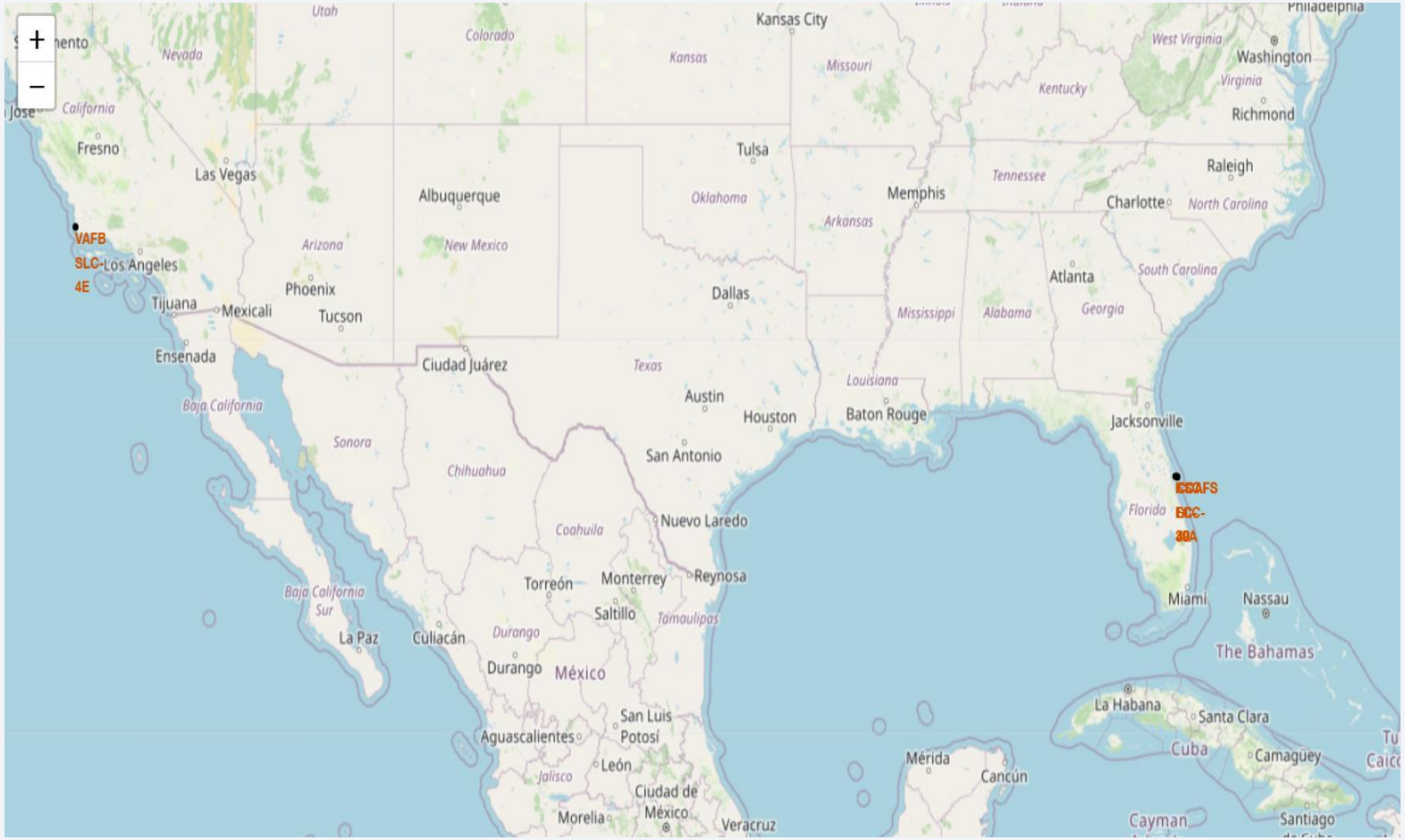
Landing _Outcome	successful_landing_outcomes_count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

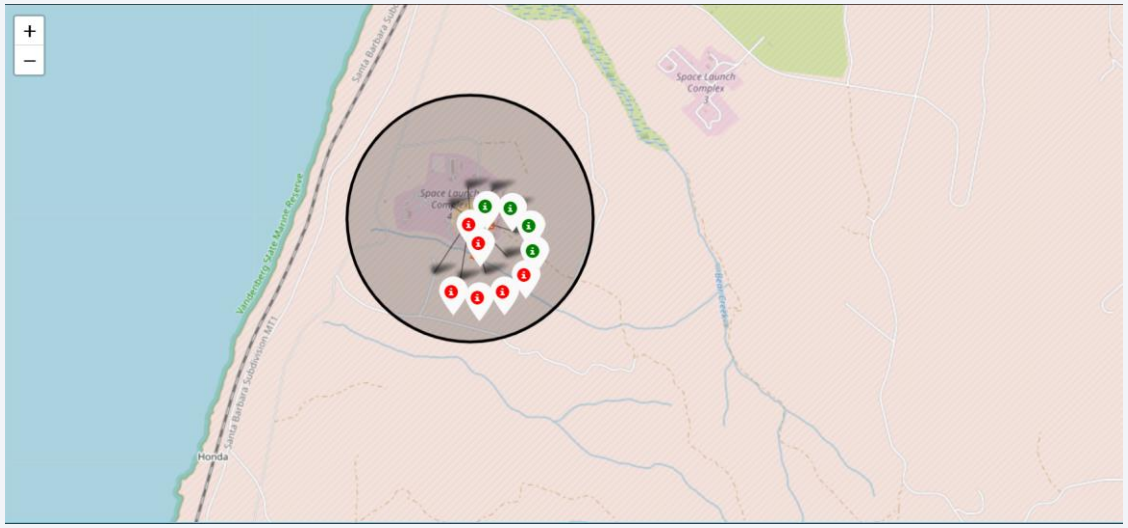
All launch sites

- The proximity of launch sites to the sea is primarily for safety reasons, however, they are strategically located in close proximity to roads and railroads to ensure convenient access and transportation.



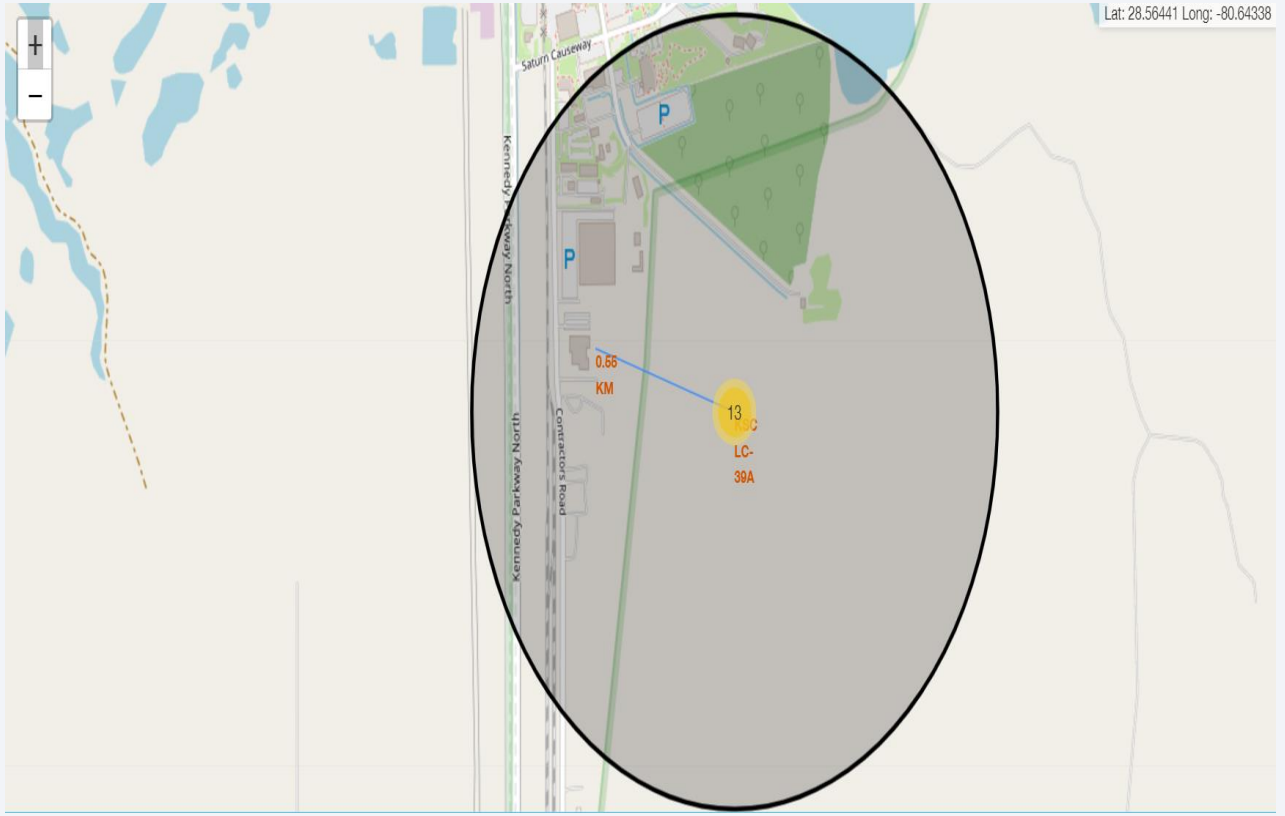
Launch Outcomes by Site

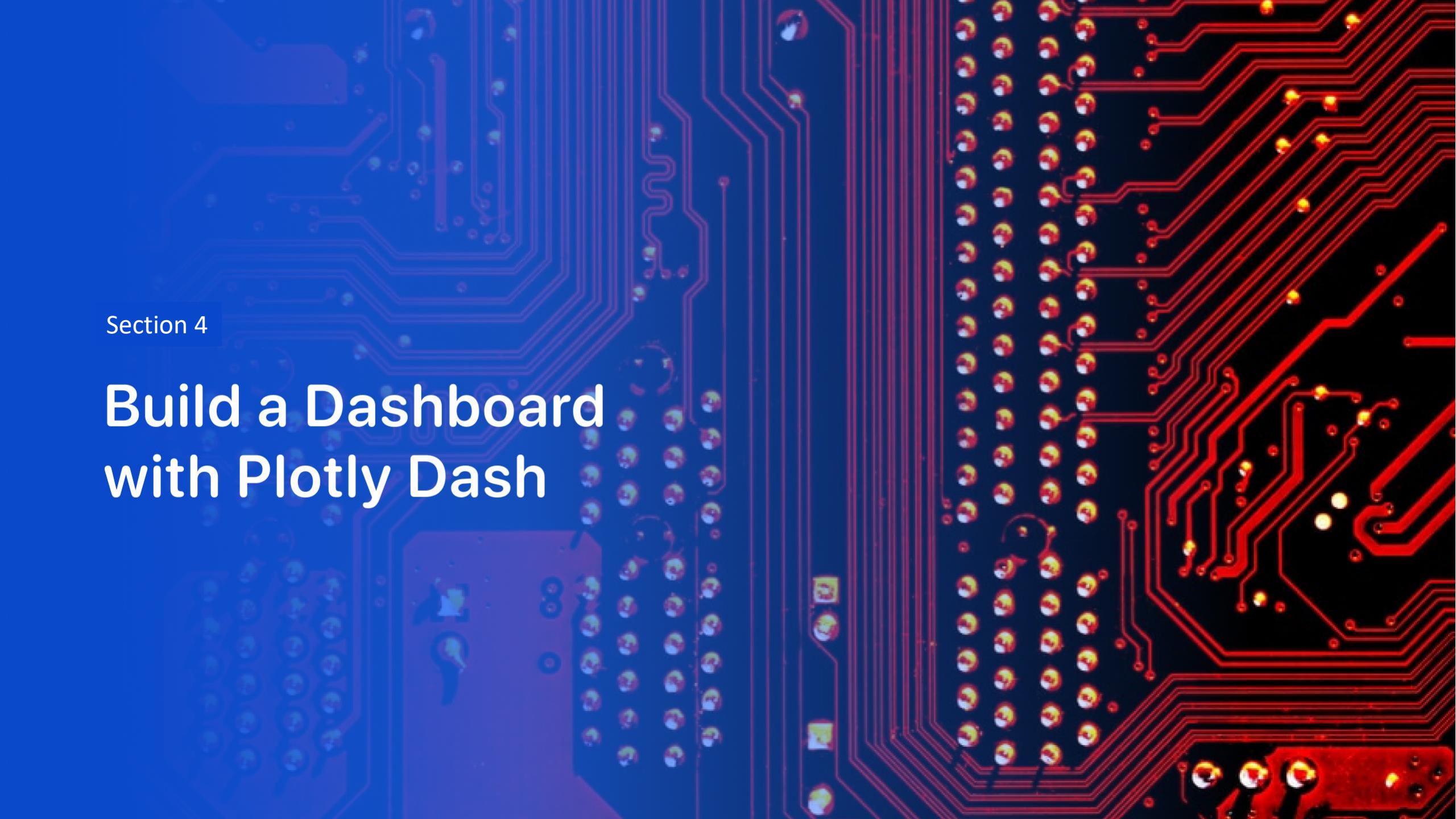
- An illustration of the outcomes of launches from Kennedy Space Center's LC-39A launch site, with green markers signifying successful launches and red markers denoting failures.



<Folium Map Screenshot 3>

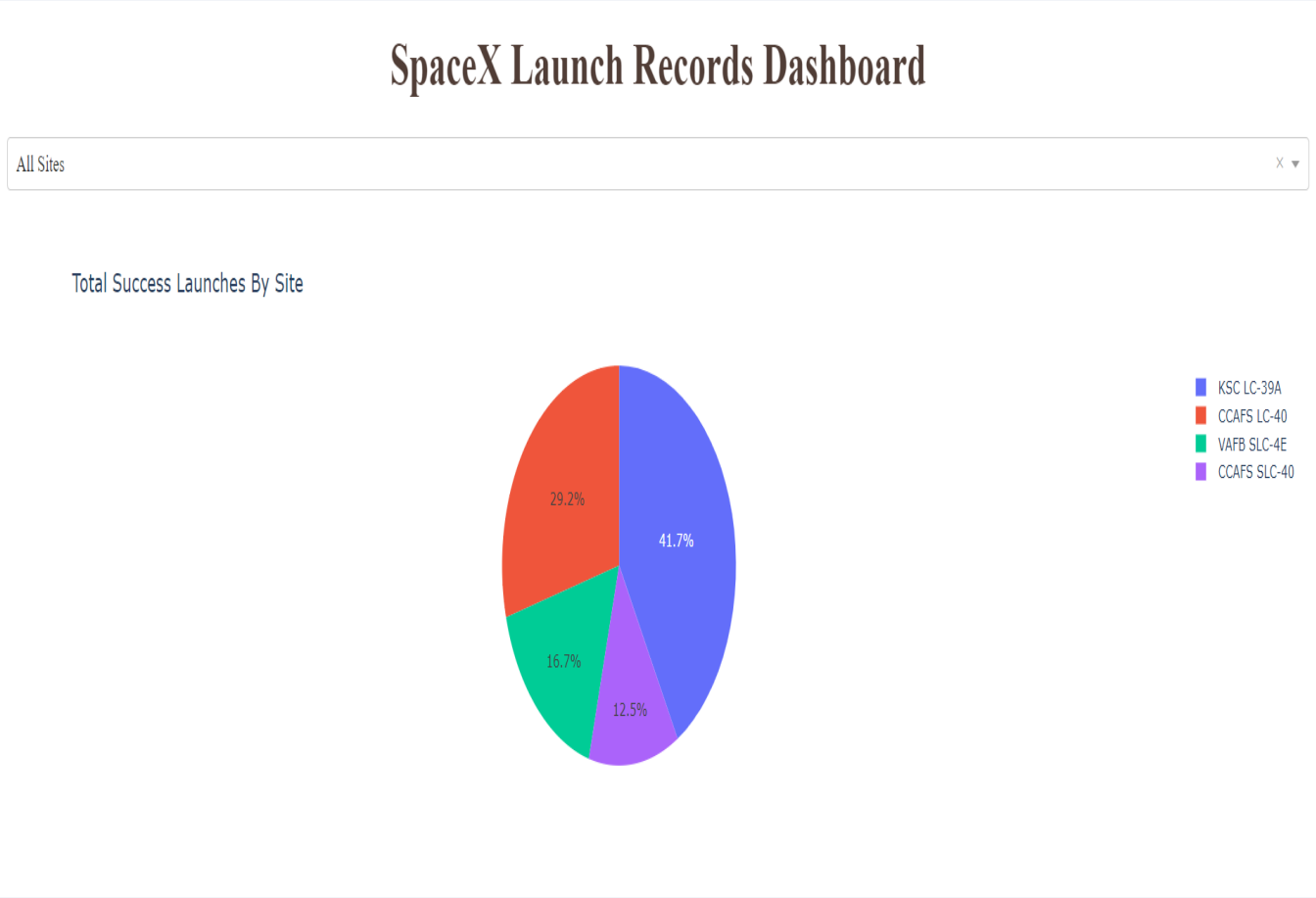
- The Kennedy Space Center's LC-39A launch site boasts advantageous logistical features, situated near both a railway and road and at a safe distance from populated areas.




Section 4
Build a Dashboard with Plotly Dash

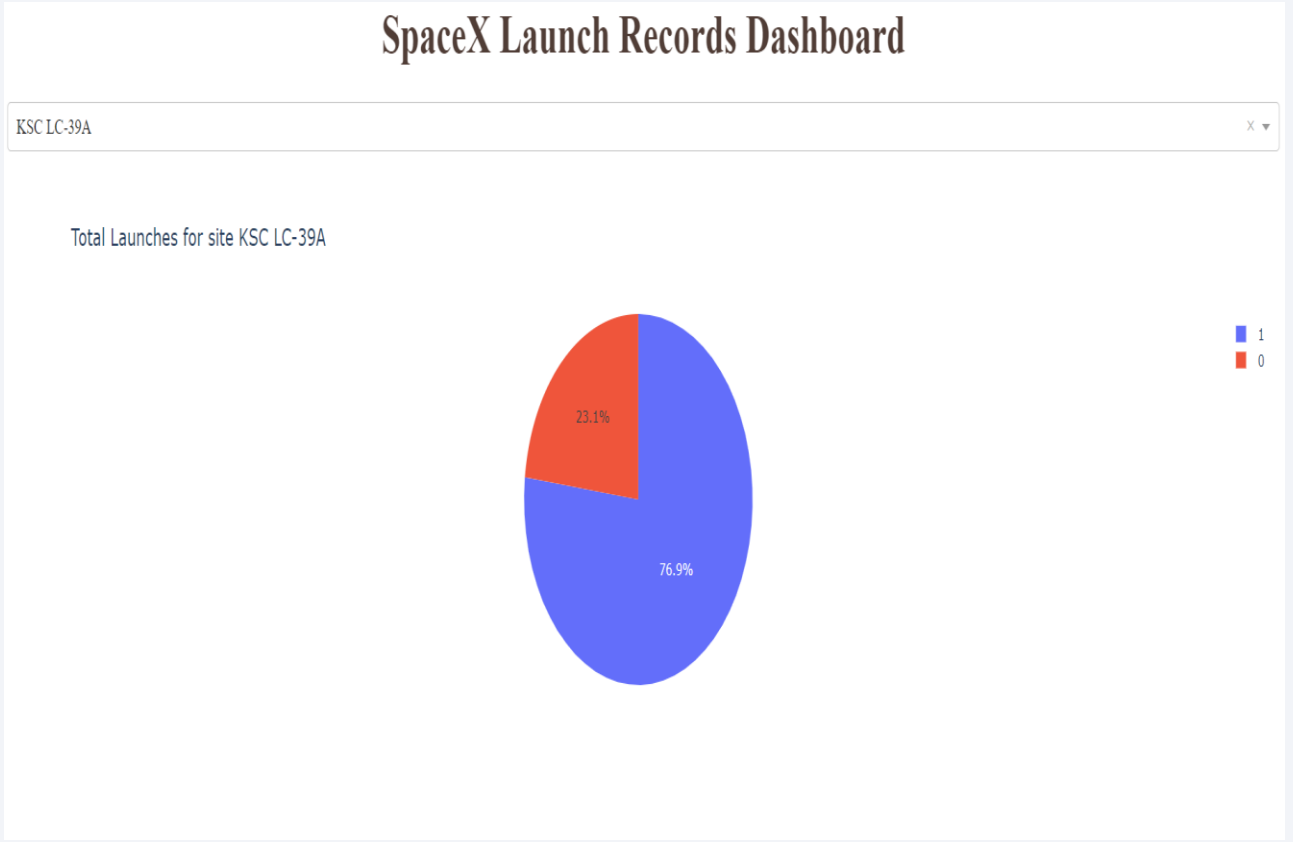
Successful Launches by Site

- The location from which launches are conducted appears to play a crucial role in determining the success of missions.



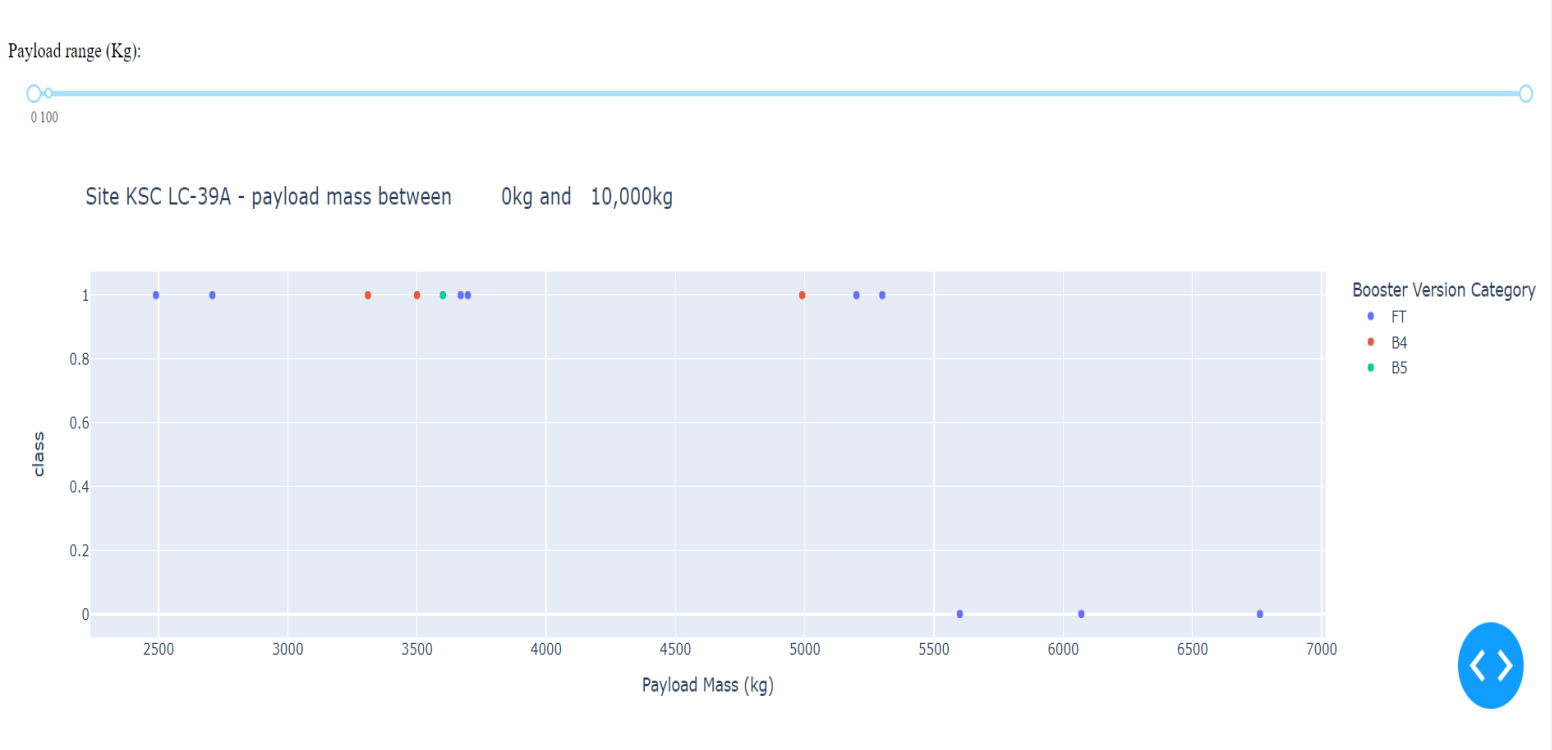
Launch Success Percentage by Site: Highlighting the Most Reliable Location

- At this site, out of all the launches conducted, 76.9% were successful. This represents a high success rate, indicating the efficiency and reliability of this location for launches.



Payload vs. Launch Outcome

- The combination of payloads weighing less than 6,000kg and FT boosters has proven to be the most successful.

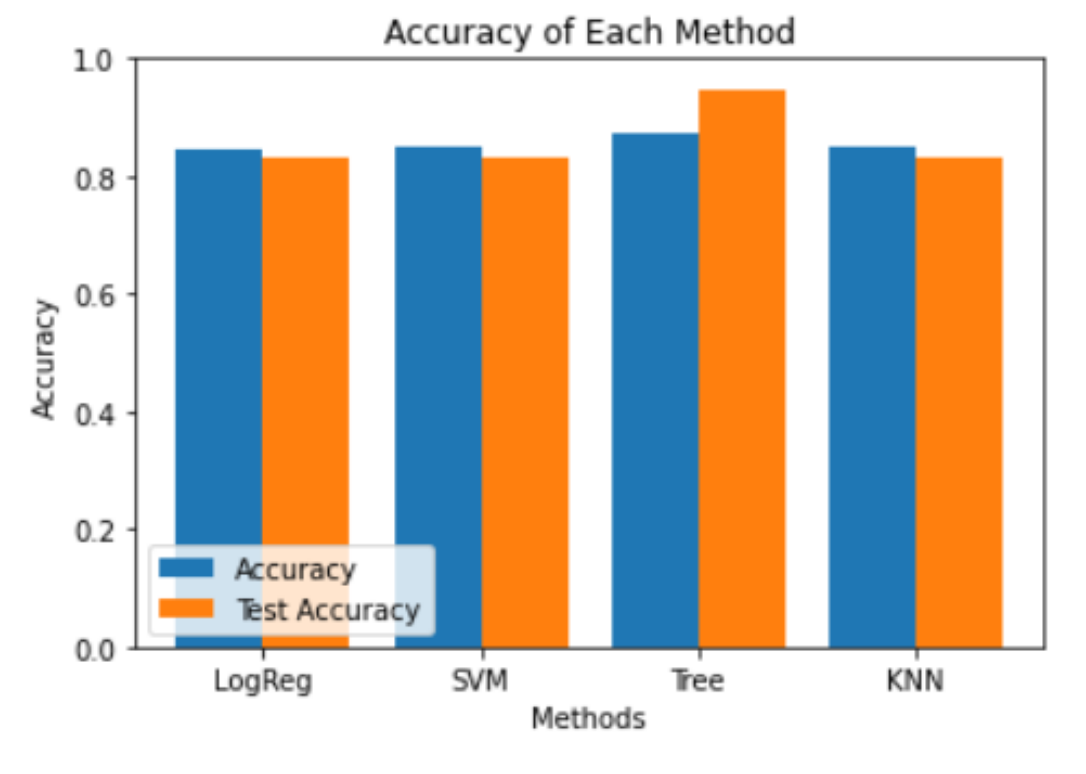


Section 5

Predictive Analysis (Classification)

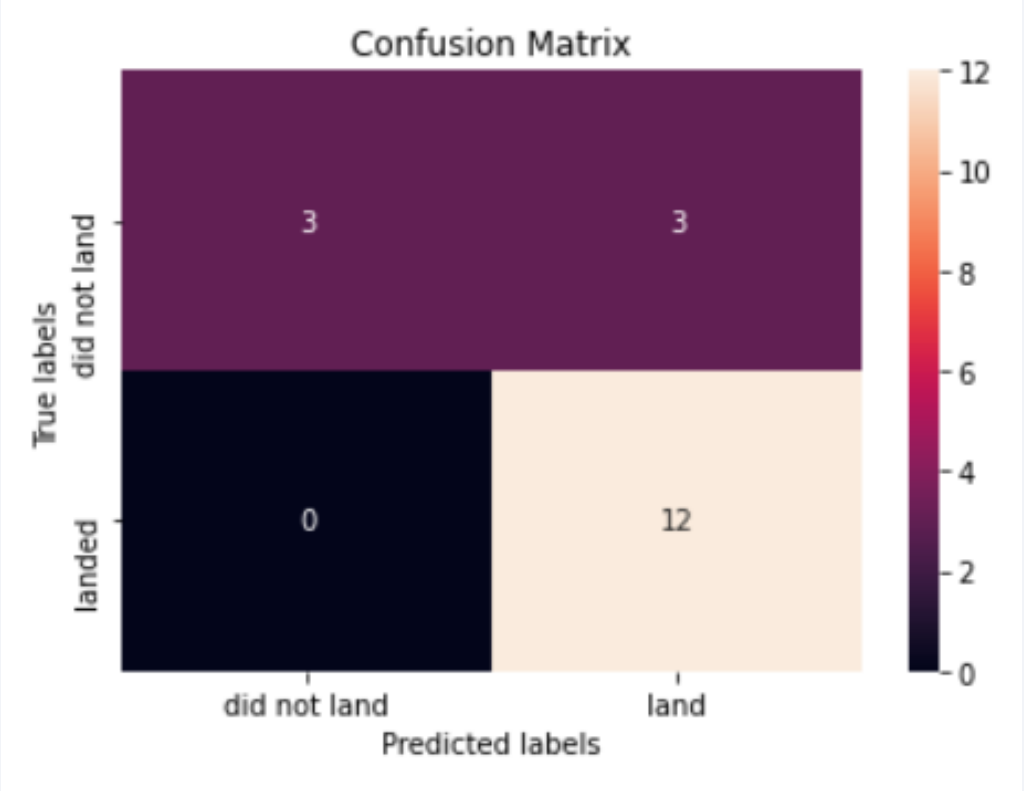
Classification Accuracy

- A comparison of four different classification models was conducted, and the accuracy of each was plotted. The Decision Tree Classifier was found to be the most accurate, with an accuracy of over 87%



Confusion Matrix

- The Confusion Matrix of the Decision Tree Classifier provides evidence of its accuracy by demonstrating a larger number of true positive and true negative results compared to false ones.



Conclusions

- Multiple data sources were analyzed in order to draw refined conclusions.
- The results indicated that the Kennedy Space Center LC-39A was the best launch site.
- Launches with payloads weighing over 7,000kg were found to be less risky.
- The success rate of missions has improved over time, and this trend is attributed to advancements in processes and rockets.
- The Decision Tree Classifier was found to be an effective tool for predicting successful landings and increasing profits.

Appendix

Thank you!