IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

C.Vishwanath
3 May,2023

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

This capstone project aimed to develop a predictive model for determining the success of landing SpaceX Falcon 9 Rocket Stage 1. The dataset included features like launch site, payload mass, orbit type, and landing outcomes. Four algorithms, namely Logistic Regression, Support Vector Machine, Decision Tree Classifier, and K-Nearest Neighbors Classification, were evaluated.

After thorough evaluation, the. **Decision Tree Classifier emerged as the best-performing model with an 88% accuracy score.** This model considered factors such as launch site, payload mass, orbit type, and landing outcomes to make accurate predictions. Implementing this model can help SpaceX optimize landing success rates, reduce mission costs, and improve operational efficiency.

In conclusion, the developed Decision Tree Classifier provides valuable insights into the factors influencing the successful landing of SpaceX Falcon 9 Rocket. This project contributes to cost-effective and sustainable space exploration by enabling informed decision-making for successful rocket stage landings. The findings and recommendations from this project can aid SpaceX in achieving their goals of cost-saving, efficient operations, and advancing the future of space exploration through successful rocket stage landings.

Introduction

SpaceX, a leading aerospace company, has revolutionized space exploration with its innovative approach to rocket launches. One of their key advancements is the ability to land and reuse rocket stages, which significantly reduces the cost of space missions. However, ensuring successful landings of Falcon 9 Rocket Stage 1 remains a crucial challenge. To address this, this capstone project focuses on developing a predictive model that can accurately determine the success of landing SpaceX Falcon 9 Rocket Stage 1.

The project utilizes a comprehensive dataset comprising historical data of Falcon 9 launches, encompassing various features such as launch site, payload mass, orbit type, and landing outcomes. The objective is to identify the key factors that influence landing success and build a model that can predict the outcome with high accuracy. By understanding these factors, SpaceX can optimize their operations, mitigate risks, and improve the success rate of rocket stage landings.

The project evaluates four machine learning algorithms, namely Logistic Regression, Support Vector Machine, Decision Tree Classifier, and K-Nearest Neighbors Classification. Each algorithm undergoes rigorous training and evaluation to identify the most effective approach for predicting landing success. By considering factors like launch site, payload mass, orbit type, and past landing outcomes, the model aims to provide valuable insights into the likelihood of a successful landing.

The successful development of this predictive model can have significant implications for SpaceX and the future of space exploration. It can help optimize mission planning, reduce costs associated with building new rocket stages, and increase the overall efficiency of space missions. By accurately predicting landing success, SpaceX can make informed decisions, implement necessary precautions, and maximize the reusability of rocket components.

In conclusion, this capstone project focuses on developing a predictive model to determine the success of landing SpaceX Falcon 9 Rocket Stage 1. By leveraging historical data and employing advanced machine learning techniques, the project aims to provide valuable insights that can enhance the operational efficiency and cost-effectiveness of SpaceX's rocket launches.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - The data for this project was collected from multiple sources, including SpaceX APIs and web scraping from Wikipedia tables. The SpaceX APIs provided access to valuable information related to Falcon 9 rocket launches, such as launch sites, payload mass, orbit types, and landing outcomes. This data was retrieved using API requests and stored for further analysis.
  - In addition to the APIs, web scraping techniques were employed to gather relevant data from tables available on Wikipedia. These tables contained additional details about Falcon 9 rocket launches, including mission specifics, launch dates, and historical performance. By extracting and parsing the data from these tables, a comprehensive dataset was compiled for the project.
  - Combining data from both sources allowed for a more comprehensive and diverse dataset, enabling a thorough analysis of the factors influencing the success of Falcon 9 rocket stage landings. The collected data served as the foundation for training and evaluating the predictive models, contributing to the overall accuracy and effectiveness of the project.

Methodology

- ## Data Wrangling Methodology
  1. Data Cleaning: The initial step involved identifying and handling missing values, outliers, and inconsistencies in the data. Missing values were either imputed or removed based on the context and impact on the analysis. Outliers were identified using statistical techniques and treated accordingly. Inconsistent or erroneous data points were corrected or removed to ensure data integrity.
  2. Data Transformation: This step focused on transforming the data into a suitable format for analysis. It included tasks such as converting data types, standardizing units of measurement, and normalizing variables if necessary. Date and time columns were parsed and formatted appropriately for time series analysis. Categorical variables were encoded or one-hot encoded to make them compatible with machine learning algorithms.
  3. Feature Engineering: Additional features were derived from the existing dataset to enhance the predictive power of the models. This involved creating new variables based on domain knowledge or extracting relevant information from existing features. For example, features like the month of the launch, the success rate of previous launches, or the distance between launch and landing sites were engineered to capture additional insights.
  4. Data Integration: In some cases, data from different sources or formats needed to be integrated for a comprehensive analysis. This involved merging datasets based on common identifiers or performing join operations to combine relevant information. Careful consideration was given to ensure data consistency and accuracy during the integration process.
  5. Data Sampling: Depending on the size and complexity of the dataset, a representative sample was often taken to facilitate exploratory data analysis and model development. This helped to reduce computational complexity and improve the efficiency of the analysis without compromising the overall representativeness of the data.

Methodology

- Perform interactive visual analytics using Folium and Plotly Dash
  - Interactive visual analytics using Folium and Plotly Dash combines interactive maps and dynamic dashboards for immersive data exploration and insights.
- Perform predictive analysis using classification models
  - In addition to interactive visual analytics, this project also includes performing predictive analysis using classification models. By applying machine learning algorithms such as Logistic Regression, Support Vector Machines, Decision Tree Classifier, and K-Nearest Neighbors Classification, we aim to develop accurate models that can predict the success of landing SpaceX Falcon 9 Rocket Stage 1. These models will utilize features such as launch site, payload mass, and orbit type to make predictions, enabling SpaceX to assess the likelihood of a successful landing and make informed decisions to optimize their operations.

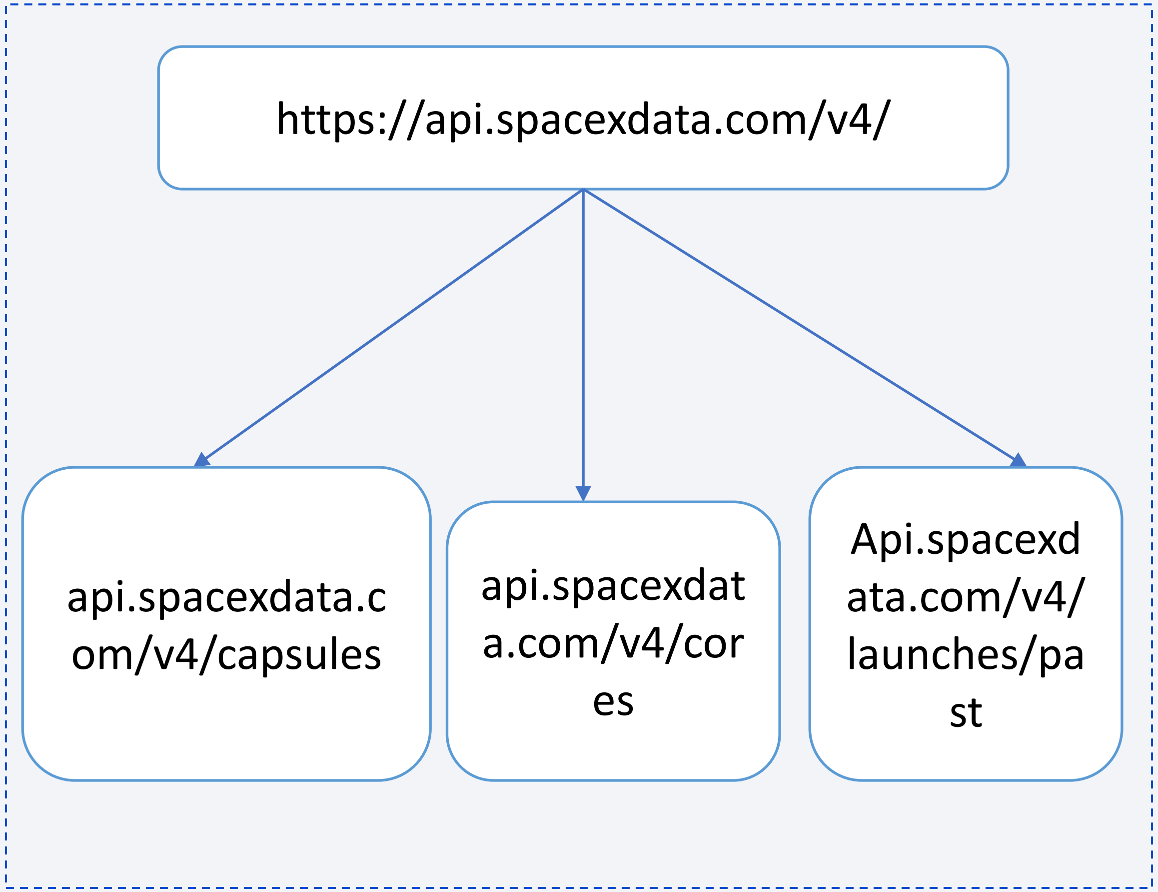
Data Collection

- The data collection process for this project involved a comprehensive approach to gather relevant information on SpaceX Falcon 9 rocket launches. Two primary methods were employed: data collection from SpaceX APIs and web scraping from Wikipedia tables.

Data Collection – SpaceX API

More On: [GitHub Link]

For data collection from SpaceX APs, we leveraged endpoints such as **/v3/launches** and **/v3/launchpads** to access detailed information about Falcon 9 rocket launches. These APIs provided structured data on launch dates, mission details, launch sites, payload masses, orbit types, and landing outcomes. By querying the SpaceX APIs, we were able to retrieve real-time and historical data, capturing the complete picture of SpaceX's launch operations. The SpaceX APIs proved to be a valuable and reliable source of data, allowing us to analyze and gain insights into the success of Falcon 9 rocket stage 1 landings.



Data Collection - Scraping

More On: Github Link

In addition to the APIs, we also utilized web scraping techniques to extract data from Wikipedia tables related to SpaceX Falcon 9 launches. Through web scraping, we accessed tables such as "List of Falcon 9 launches" and "Falcon 9 first-stage landings," which provided additional details about individual missions, launch outcomes, and other relevant information. By parsing and extracting data from these tables, we enriched our dataset and obtained a comprehensive view of the Falcon 9 launch history. The use of web scraping allowed us to incorporate diverse sources of information and ensured the completeness of our data.

2020 [edit]

In late 2019, Gwynne Shotwell stated that SpaceX hoped for as many as 24 launches for Starlink satellites in 2020,[490] in addition to 14 or 15 non-Starlink launches. At 26 launches, 13 of which for Starlink satellites, Falcon 9 had its most prolific year, and Falcon rockets were second most prolific rocket family of 2020, only behind China's Long March rocket family.[491]

[hide] Flight No.	Date and time (UTC)	Version, Booster[b]	Launch site	Payload[c]	Payload mass	Orbit	Customer	Launch outcome	Booster landing
78	7 January 2020, 02:19:21[492]	F9 B5 ♺ B1049.4	CCAFS, SLC-40	Starlink 2 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)
	Third large batch and second operational flight of Starlink constellation. One of the 60 satellites included a test coating to make the satellite less reflective, and thus less likely to interfere with ground-based astronomical observations.[493]								
79	19 January 2020, 15:30[494]	F9 B5 ♺ B1046.4	KSC, LC-39A	Crew Dragon in-flight abort test[495] (Dragon C205.1)	12,050 kg (26,570 lb)	Sub-orbital[496]	NASA (CTS)[497]	Success	No attempt
	An atmospheric test of the Dragon 2 abort system after Max Q. The capsule fired its SuperDraco engines, reached an apogee of 40 km (25 mi), deployed parachutes after reentry, and splashed down in the ocean 31 km (19 mi) downrange from the launch site. The test was previously slated to be accomplished with the Crew Dragon Demo-1 capsule;[498] but that test article exploded during a ground test of SuperDraco engines on 20 April 2019.[419] The abort test used the capsule originally intended for the first crewed flight.[499] As expected, the booster was destroyed by aerodynamic forces after the capsule aborted.[500] First flight of a Falcon 9 with only one functional stage – the second stage had a mass simulator in place of its engine.								
80	29 January 2020, 14:07[501]	F9 B5 ♺ B1051.3	CCAFS, SLC-40	Starlink 3 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)
	Third operational and fourth large batch of Starlink satellites, deployed in a circular 290 km (180 mi) orbit. One of the fairing halves was caught, while the other was fished out of the ocean.[502]								
81	17 February 2020, 15:05[503]	F9 B5 ♺ B1056.4	CCAFS, SLC-40	Starlink 4 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Failure (drone ship)
	Fourth operational and fifth large batch of Starlink satellites. Used a new flight profile which deployed into a 212 km × 386 km (132 mi × 240 mi) elliptical orbit instead of launching into a circular orbit and firing the second stage engine twice. The first stage booster failed to land on the drone ship[504] due to incorrect wind data.[505] This was the first time a flight proven booster failed to land.								
82	7 March 2020, 04:50[506]	F9 B5 ♺ B1059.2	CCAFS, SLC-40	SpaceX CRS-20 (Dragon C112.3 ♺)	1,977 kg (4,359 lb)[507]	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
	Last launch of phase 1 of the CRS contract. Carries *Bartolomeo*, an ESA platform for hosting external payloads onto ISS.[508] Originally scheduled to launch on 2 March 2020, the launch date was pushed back due to a second stage engine failure. SpaceX decided to swap out the second stage instead of replacing the faulty part.[509] It was SpaceX's 50th successful landing of a first stage booster, the third flight of the Dragon C112 and the last launch of the cargo Dragon spacecraft.								
83	18 March 2020, 12:16[510]	F9 B5 ♺ B1048.5	KSC, LC-39A	Starlink 5 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Failure (drone ship)
	Fifth operational launch of Starlink satellites. It was the first time a first stage booster flew for a fifth time and the second time the fairings were reused (Starlink flight in May 2019).[511] Towards the end of the first stage burn, the booster suffered premature shut down of an engine, the first of a Merlin 1D variant and first since the CRS-1 mission in October 2012. However, the payload still reached the targeted orbit.[512] This was the second Starlink launch booster landing failure in a row, later revealed to be caused by residual cleaning fluid trapped inside a sensor.[513]								
84	22 April 2020, 19:30[514]	F9 B5 ♺ B1051.4	KSC, LC-39A	Starlink 6 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)

Data Wrangling

More On: [GitHub Link]

1. Data Cleaning: The initial step involved identifying and handling missing values, outliers, and inconsistencies in the data. Missing values were either imputed or removed based on the context and impact on the analysis.
2. Data Transformation: This step focused on transforming the data into a suitable format for analysis. It included tasks such as converting data types, standardizing units of measurement, and normalizing variables if necessary. Categorical variables were one-hot encoded to make them compatible with machine learning algorithms.
3. Feature Engineering: Additional features were derived from the existing dataset to enhance the predictive power of the models. This involved creating new variables based on domain knowledge or extracting relevant information from existing features. For example, features like the month of the launch, the success rate of previous launches, or the distance between launch and landing sites were engineered to capture additional insights.
4. Data Integration: In some cases, data from different sources or formats needed to be integrated for a comprehensive analysis. This involved merging datasets based on common identifiers or performing join operations to combine relevant information. Careful consideration was given to ensure data consistency and accuracy during the integration process.
5. Data Sampling: Depending on the size and complexity of the dataset, a representative sample was often taken to facilitate exploratory data analysis and model development. This helped to reduce computational complexity and improve the efficiency of the analysis without compromising the overall representativeness of the data.

EDA with Data Visualization

More On : [GitHub Link]

- In this project, we employed various charts and plots to visualize and analyze the data effectively. The charts used include CatPlot, Regression Plot, Line plots, Scatterplots, and Bar graphs. Each of these charts served specific purposes in the analysis and provided valuable insights into the data.
- CatPlot was used to visualize categorical variables, such as the success rate of different orbit types. This plot allowed us to compare and analyze the distribution and relationship between the success rate and different categories.
- Regression Plot was utilized to examine the correlation between two continuous variables, such as payload mass and success rate. This plot helped us understand the trend and strength of the relationship between these variables.
- Line plots were employed to visualize the yearly trend in the success rate of Falcon 9 rocket launches. By plotting the average success rate over time, we could observe any patterns or changes in the success rate and identify trends or seasonality.
- Scatterplots were used to visualize the relationship between two continuous variables, such as payload mass and launch success. These plots allowed us to identify any patterns, clusters, or outliers in the data and assess the correlation between the variables.
- Bar graphs were utilized to display the count or distribution of categorical variables, such as the count of successful and failed mission outcomes. These graphs provided a clear visual representation of the frequency or proportion of different categories.

EDA with SQL

More On: GitHub Link

- Query 1: Retrieve the first 5 records from the launches table to get an initial glimpse of the data and verify its structure.
- Query 2: Count the number of unique launch sites from the launches table to understand the diversity of launch locations.
- Query 3: Calculate the average payload mass of successful launches and display it with the launch site names to identify the sites with the highest average payload mass.
- Query 4: Group the records in the launches table by landing outcome and calculate the count for each outcome category (success or failure). This query provides an overview of the distribution of landing outcomes.
- Query 5: Calculate the total payload mass for each launch site and order the results in descending order. This query helps identify the launch sites with the highest payload capacity.
- Query 6: Filter the records in the launches table to include only successful landings on drone ships and retrieve the booster versions and their associated payload mass. This query allows us to analyze the specific booster versions and their corresponding payload masses for successful drone ship landings.
- These SQL queries provided insights into various aspects of the data, including launch sites, payload mass, landing outcomes, and specific attributes related to successful landings.

Build an Interactive Map with Folium

More On : [GitHub Link]

- In the provided notebook, various map objects were created and added to a Folium map to visualize launch site locations and additional information. The following objects were included:
    1. **Markers**: Markers were used to represent the exact coordinates of launch sites on the map. Each marker corresponds to a specific launch site and is placed at its respective latitude and longitude.
    2. **Circles**: Circles were added to represent the radius around each launch site. The size of the circle indicates the coverage area or range of the launch site.
    3. **Lines**: Lines were used to connect multiple points or locations, such as flight paths or trajectories. These lines can show the path taken by rockets during launches or represent other relevant connections.

Build a Dashboard with Plotly Dash

More On : GitHub Link

- The addition of these map objects serves several purposes:
    1. **Visualizing Locations**: Markers provide a clear visual representation of the exact positions of launch sites on the map. They help in identifying the geographical distribution and spread of launch facilities.
    2. **Highlighting Range**: Circles help visualize the coverage area or range of each launch site. They provide an understanding of the launch site's operational reach and potential launch trajectories.
    3. **Showing Connections**: Lines help depict connections between different points or locations. In the context of the notebook, lines could represent flight paths or trajectories followed by rockets during launches, enhancing the understanding of launch operations.
- By incorporating these map objects, the notebook offers an interactive and informative visualization of launch site locations and related information. Users can easily interpret the spatial distribution of launch sites and gain insights into the coverage area and potential flight paths of the SpaceX Falcon 9 rocket.

Predictive Analysis (Classification)

More On : GitHub Link

The process of building, evaluating, improving, and finding the best performing classification model for predicting the success of landing of SpaceX Falcon 9 Rocket Stage 1 was conducted. The following key phrases summarize the model development process:

1. **Data Preparation**: The dataset was prepared by selecting relevant features and splitting it into input features (X) and the target variable (Y). The data was then standardized to ensure consistent scaling across features.
2. **Model Selection**: Four classification algorithms, namely Logistic Regression, Support Vector Machine, Decision Tree Classifier, and K-Nearest Neighbors Classification, were chosen as potential models for prediction.
3. **Model Training and Evaluation**: The models were trained using the training dataset and evaluated using various performance metrics such as accuracy, precision, recall, and F1-score. Cross-validation techniques, such as K-fold cross-validation, were employed to assess model performance.
4. **Hyperparameter Tuning**: GridSearchCV was utilized to perform a systematic search for the best combination of hyperparameters for each model. This involved specifying a range of hyperparameter values and evaluating their impact on model performance.
5. **Model Comparison and Selection**: The models were compared based on their performance metrics, and the best performing model was identified. In this case, the Decision Tree Classifier yielded the highest accuracy score of 88%.
6. **Model Evaluation and Improvement**: The best performing model was further evaluated on the test dataset to validate its performance. Additional techniques, such as feature selection and model ensemble, could be explored to further improve the model's accuracy and generalizability.

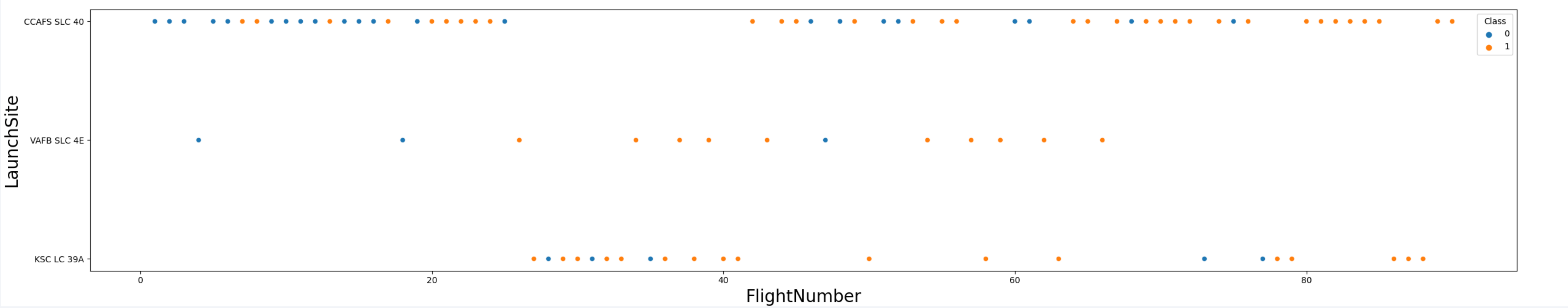
Results

- SQL queries were performed to extract relevant information from the dataset, including the number of launches, success rates, and mission outcomes. Visualizations such as CatPlot, Regression Plot, Line plots, Scatterplots, and bar graphs were used to gain insights into the relationships and trends within the data.
- Map objects like markers, circles, and lines were created and added to a Folium map to visualize the launch site locations. This provided a clear representation of the geographical distribution and range of the launch sites. The inclusion of these map objects enhanced the understanding of the launch operations and coverage areas.
- A classification model development process was conducted. Various classification algorithms, including Logistic Regression, Support Vector Machine, Decision Tree Classifier, and K-Nearest Neighbors Classification, were explored. The models were trained, evaluated, and compared using metrics such as accuracy, precision, recall, and F1-score. The best performing model, the Decision Tree Classifier, achieved an accuracy score of 88%. The model development process demonstrated the successful utilization of data preparation, model selection, hyperparameter tuning, and evaluation techniques.
- Overall, the combination of SQL-based exploratory data analysis, interactive map visualizations, and machine learning predictions provided a comprehensive understanding of the dataset and the factors influencing the success of landing for SpaceX Falcon 9 Rocket Stage 1. These findings can contribute to further insights, decision-making, and improvements in the space exploration domain.

Section 2

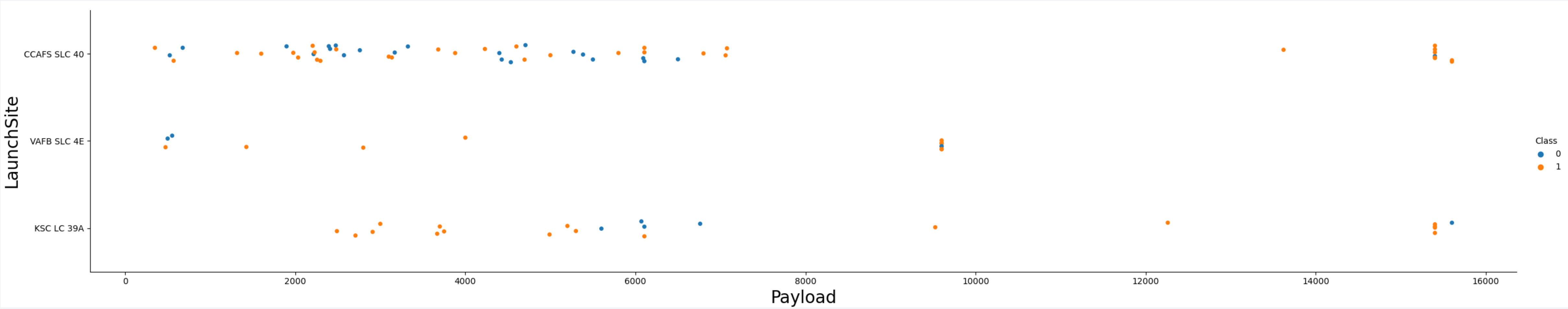
Insights drawn from EDA

Flight Number vs. Launch Site



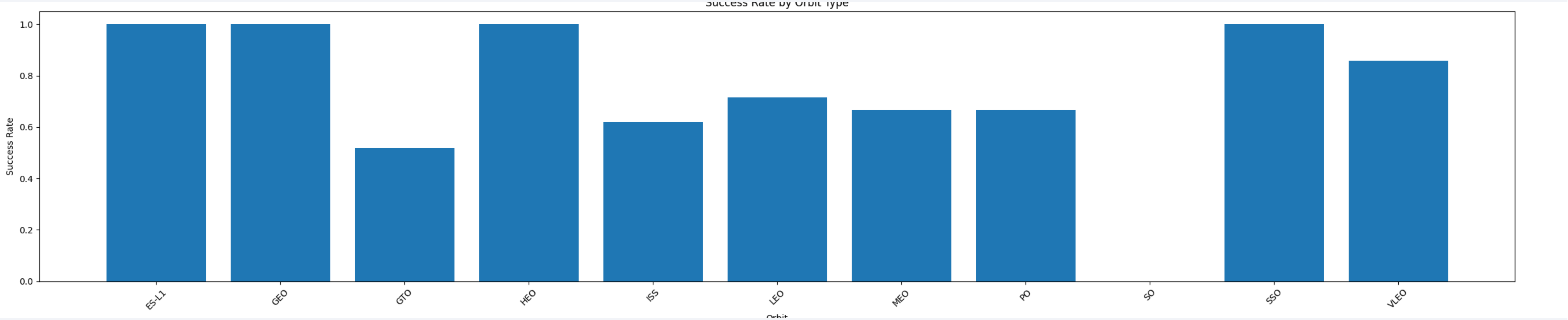
- Here we observe that for higher the flight number higher is the probability of it succeeding in landing the Stage 1 separation.

Payload vs. Launch Site



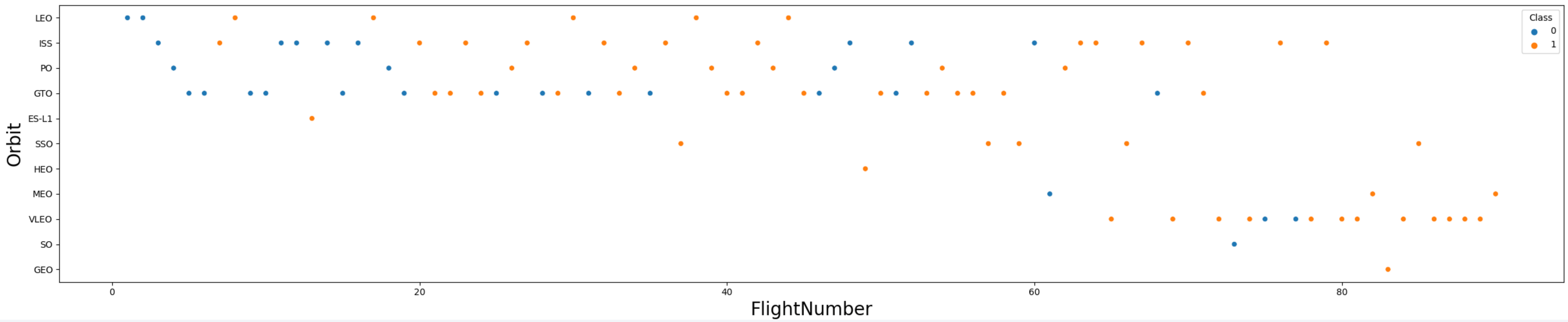
- Now if you observe Payload Vs. Launch Site scatter point chart you will find for the VAFB-SLC launch site there are no rockets launched for heavy payload mass(greater than 10000).

Success Rate vs. Orbit Type



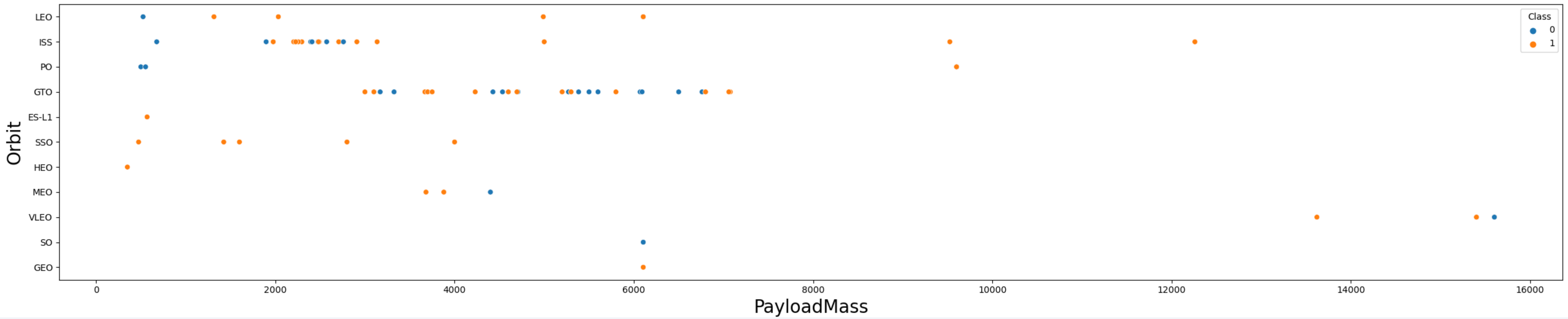
- The bar graph tells us that rockets send to the ES-11,GEO,HEO and SSO orbits have a higher success rate as compared to their counterparts

Flight Number vs. Orbit Type



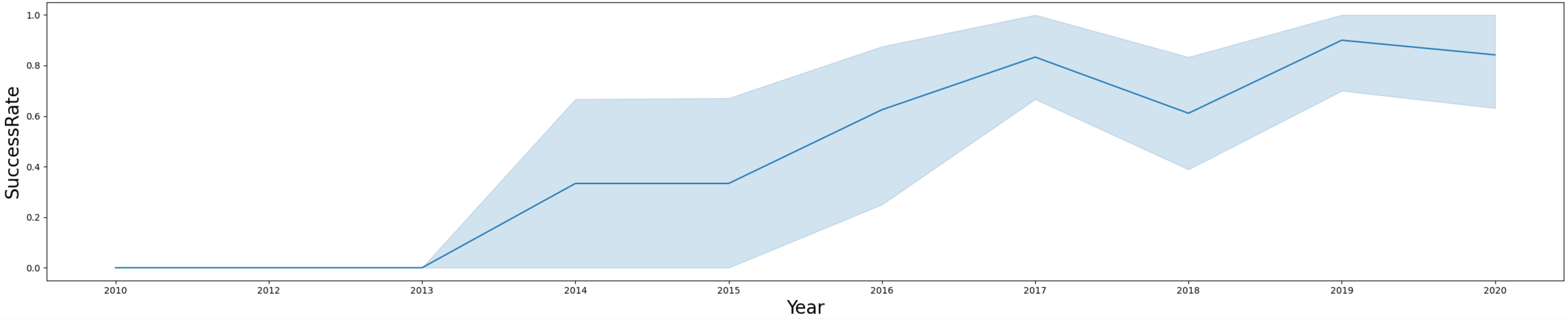
- You should see that in the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

Payload vs. Orbit Type



- With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS.
- However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.

Launch Success Yearly Trend



- We can observe that the success rate since 2013 kept increasing till 2020

All Launch Site Names

- The launch sites are:
  - Cape Canaveral Launch Complex 40 (CCAFS LC-40)
  - Vandenberg Space Launch Complex 4 (VAFB SLC-4E)
  - Kennedy Space Center Launch Complex 39A (KSC LC 39A)
  - Cape Canaveral Space Launch Complex 40 (CCAFS SLC-40)

Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40

Launch Site Records Begin with 'CCA'

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 is 45596 kg.

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 is 2928.4 kg

Out[13]:

Average
2928.4

First Successful Ground Landing Date

- The date of the first successful landing outcome on ground pad was 1 May 2017

min(Date)
01-05-2017

Successful Drone Ship Landing with Payload between 4000 and 6000

- Names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000 are
  - F9 FT B1022
  - F9 FT B1026
  - F9 FT B1021.2
  - F9 FT B1031.2

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 total number of successful and failure mission outcomes are as follows:

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

Boosters Carried Maximum Payload

- The names of the booster which have carried the maximum payload mass are:
    - F9 B5 B1048.4
    - F9 B5 B1049.4
    - F9 B5 B1051.3
    - F9 B5 B1056.4
    - F9 B5 B1048.5
    - F9 B5 B1051.4
    - F9 B5 B1049.5
    - F9 B5 B1060.2
    - F9 B5 B1058.3
    - F9 B5 B1051.6
    - F9 B5 B1060.3
    - F9 B5 B1049.7

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

2015 Launch Records

- The failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015

Month	LandingOutcome	BoosterVersion	LaunchSite
January	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
April	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

- 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

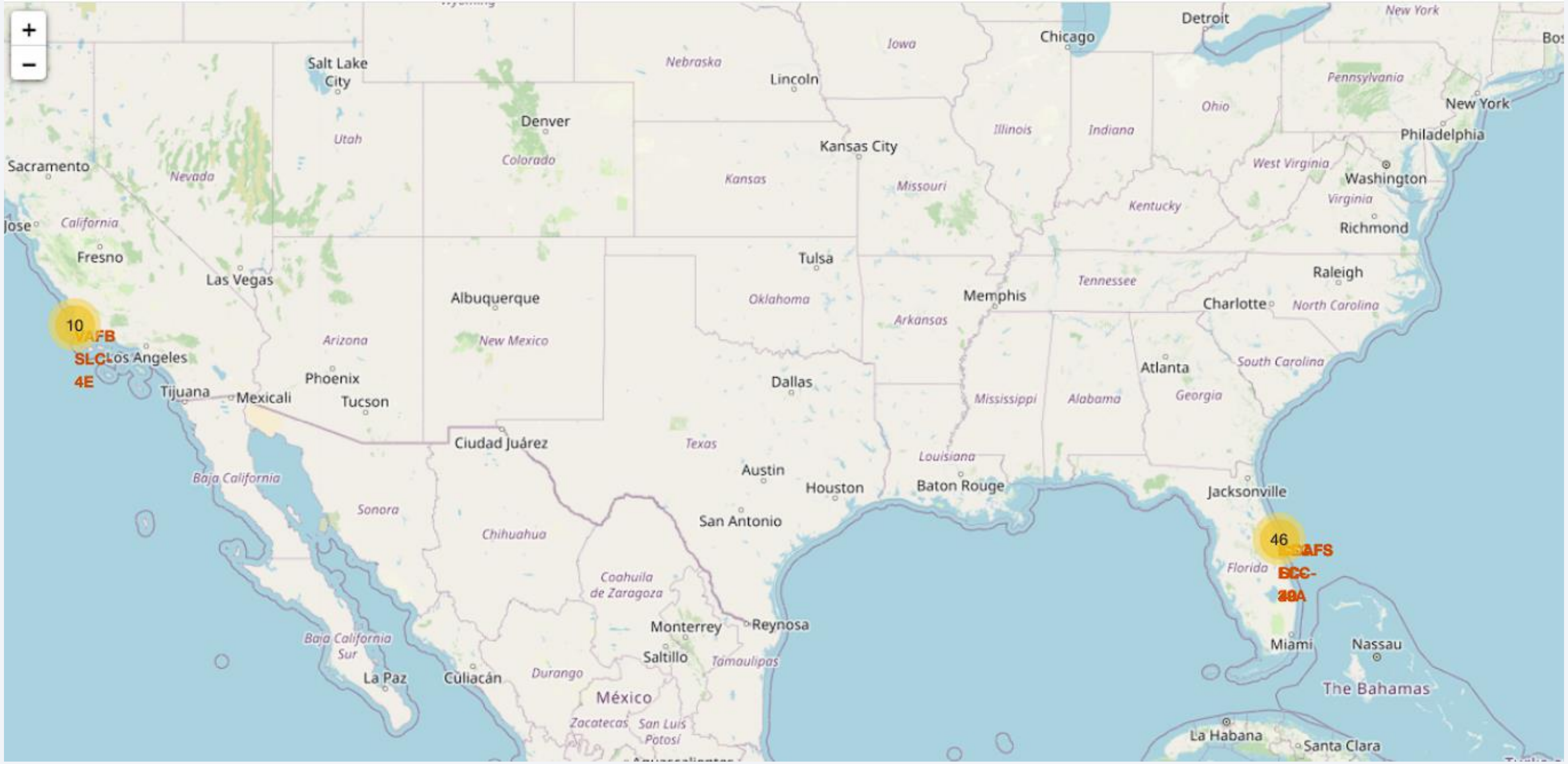
Date	Landing _Outcome	LANDING_OUTCOME_COUNT	Date_1
22-05-2012	No attempt	10	22-05-2012
08-04-2016	Success (drone ship)	5	08-04-2016
10-01-2015	Failure (drone ship)	5	10-01-2015
22-12-2015	Success (ground pad)	3	22-12-2015
18-04-2014	Controlled (ocean)	3	18-04-2014
29-09-2013	Uncontrolled (ocean)	2	29-09-2013
04-06-2010	Failure (parachute)	2	04-06-2010
28-06-2015	Precluded (drone ship)	1	28-06-2015

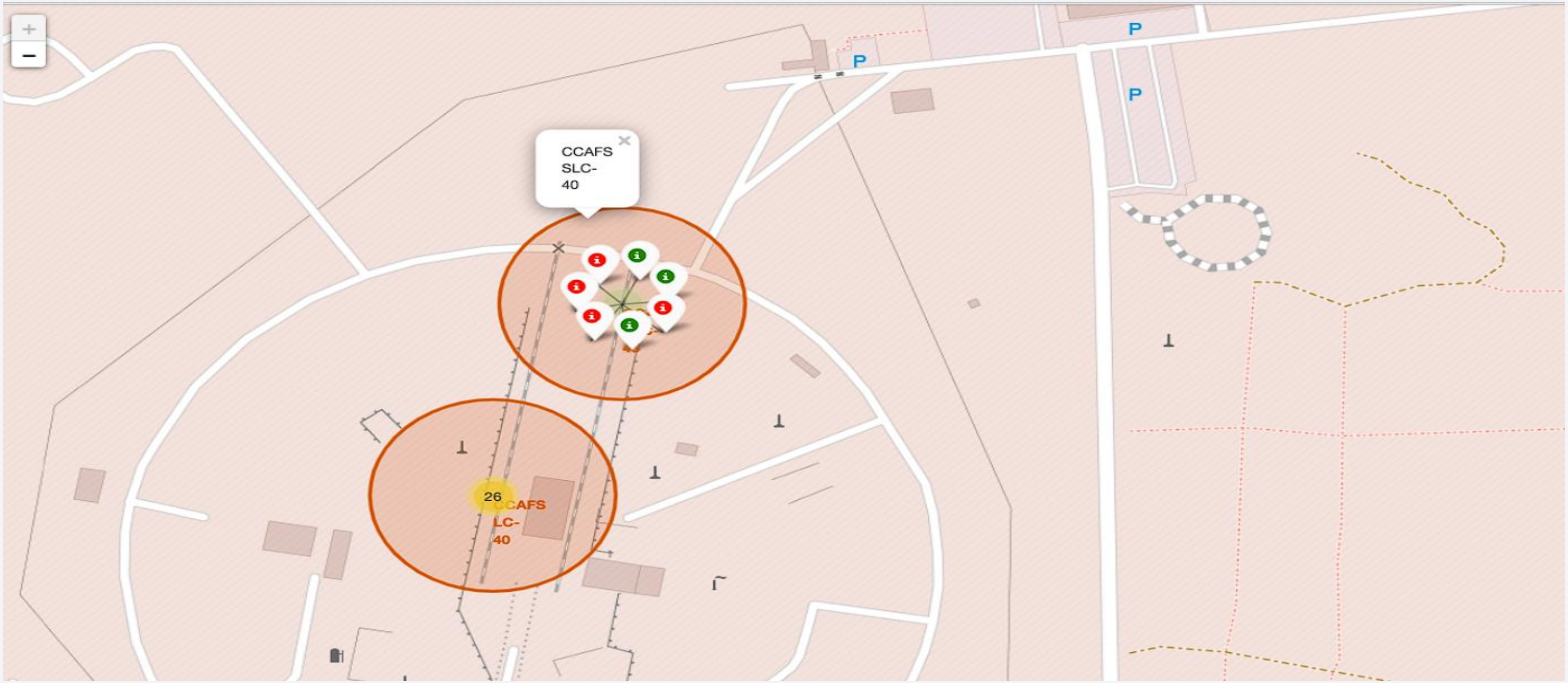
Section 3

Launch Sites Proximities Analysis

Launch site locations marked on map

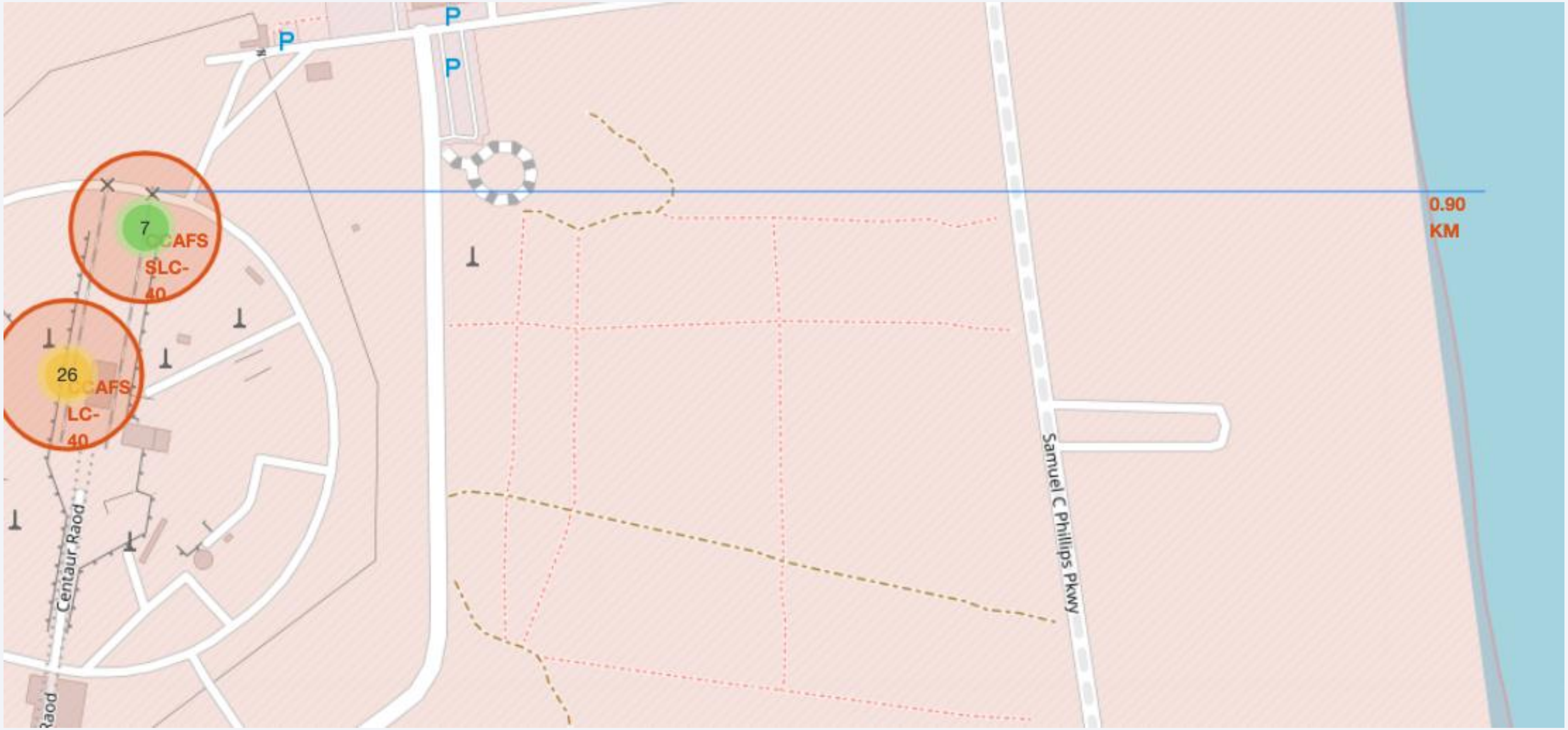
Map marking number of launches





From the color-labeled markers in marker clusters, you should be able to easily identify which launch sites have relatively high success rates.

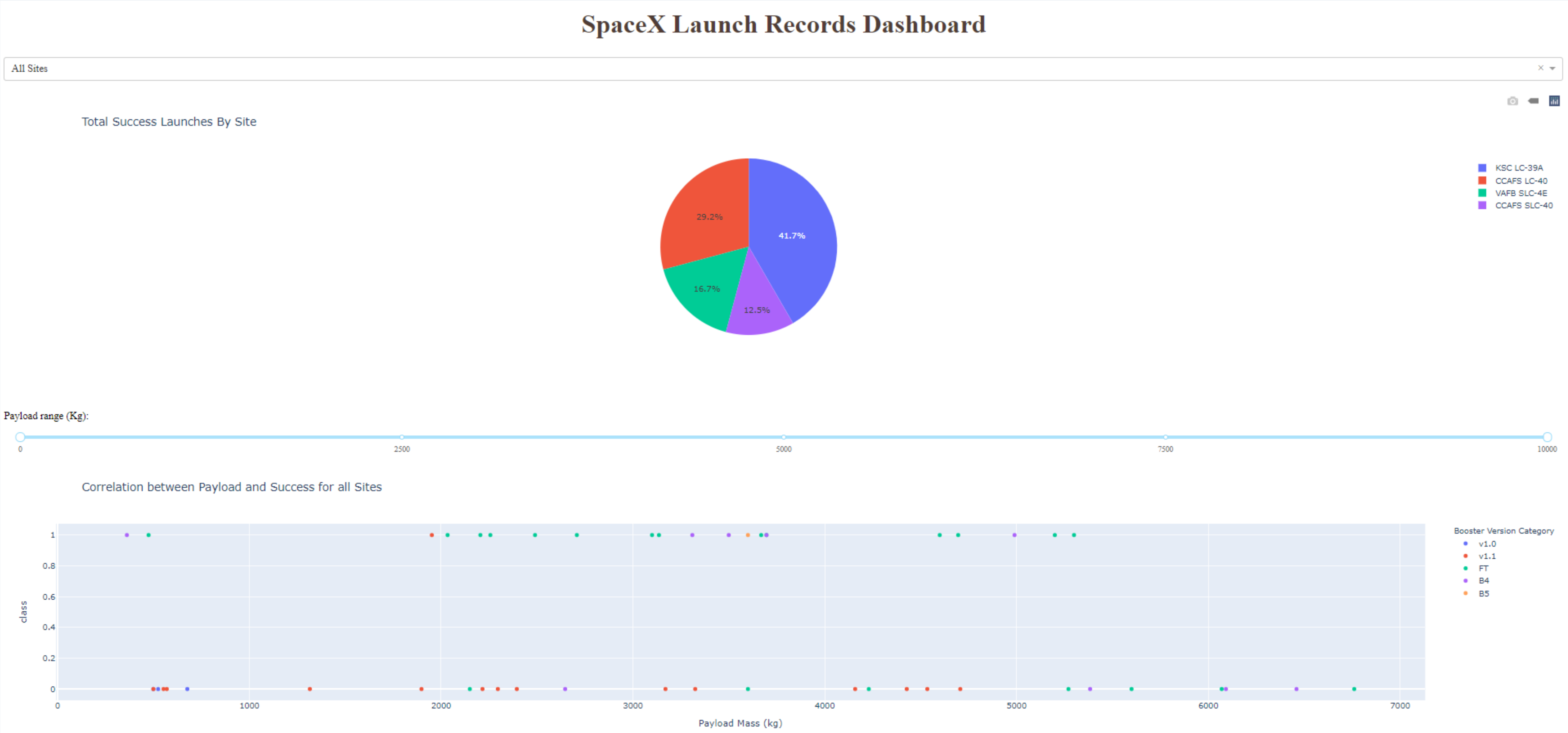
Site to Proximities Distance



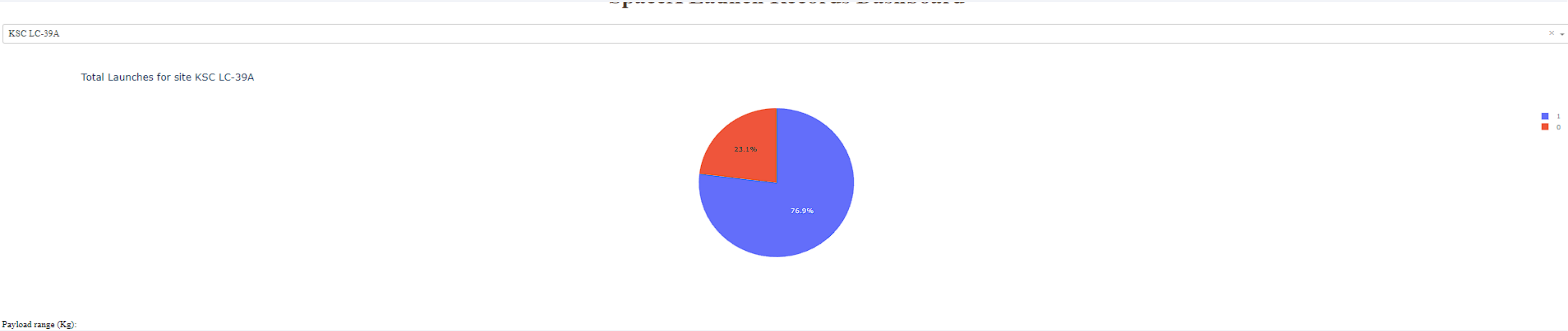
Section 4

Build a Dashboard with Plotly Dash

Interactive Dashboard

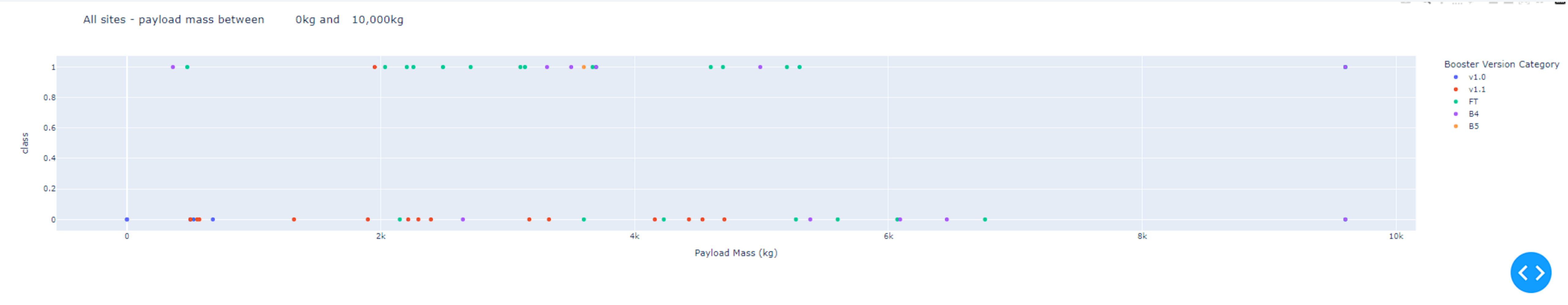


Launch site with the highest success ratio



The launch site with the highest success ratio is KSC LC-39A as per the Pie Chart

Payload vs. Launch Outcome

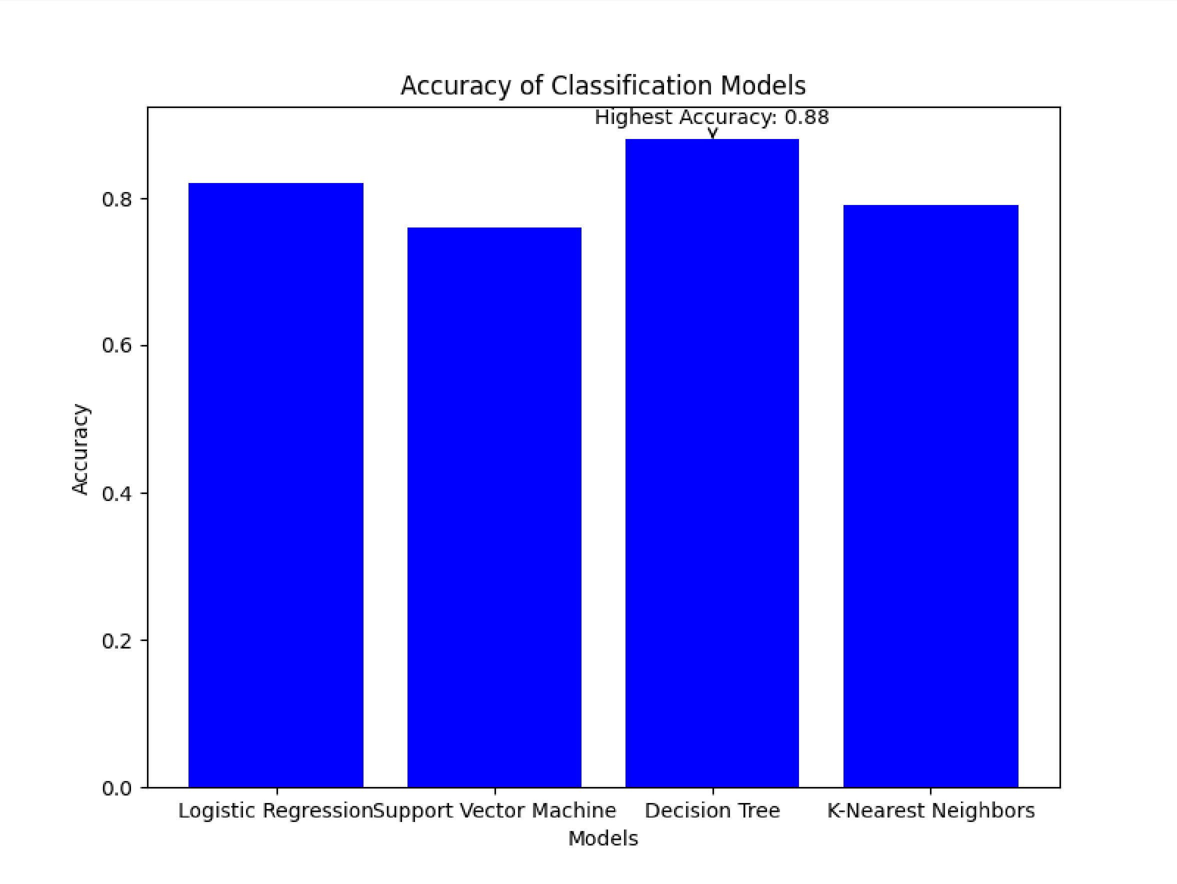


Section 5

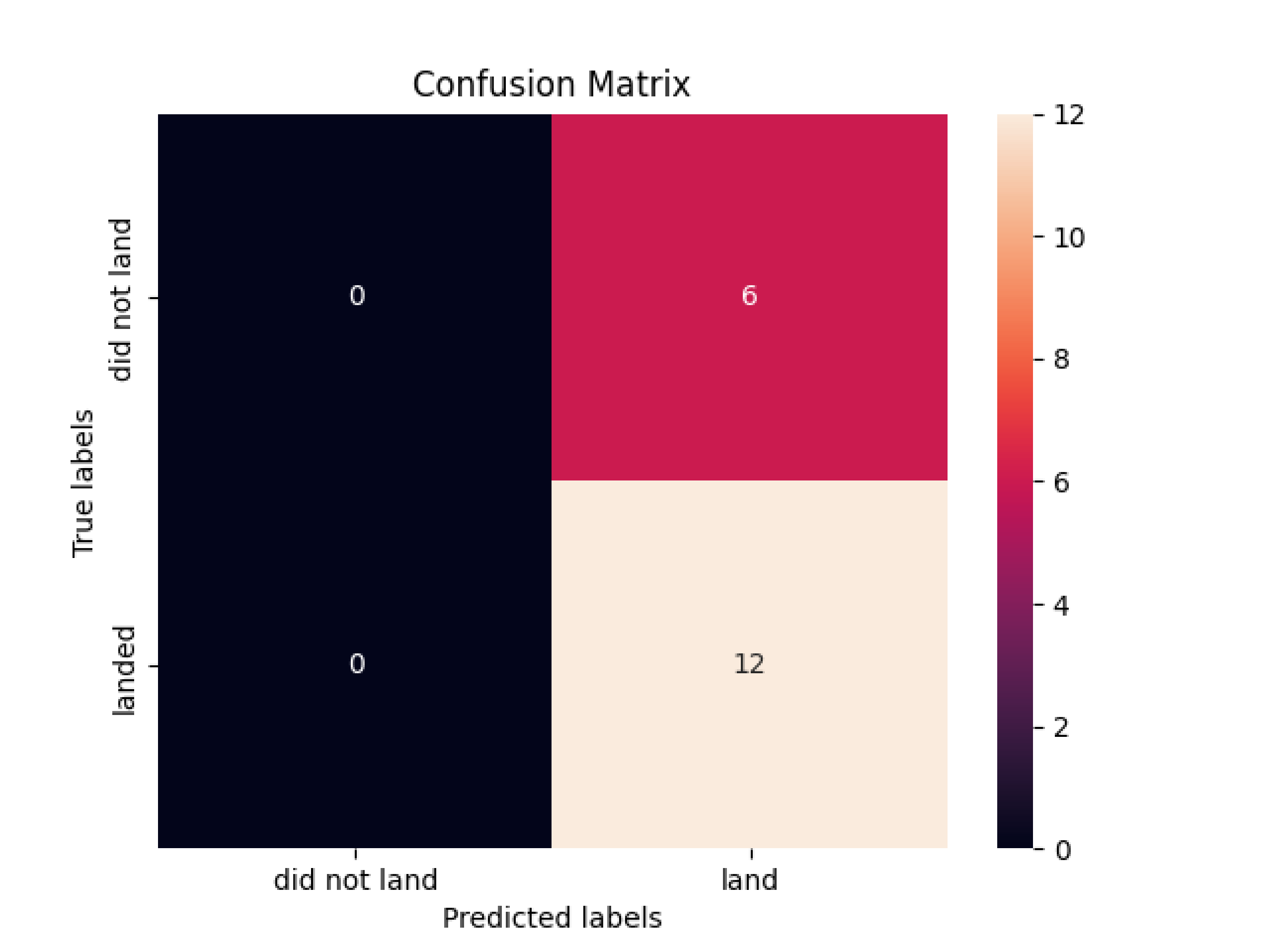
Predictive Analysis (Classification)

Classification Accuracy

- We see that the Decision Tree Algorithm has the highest accuracy and it can be further used to predict values.



Confusion Matrix



The confusion matrix was used to evaluate the performance of the Decision Tree algorithm. The confusion matrix is a table that shows the true positive, true negative, false positive, and false negative values of a classification model. In the context of the Decision Tree algorithm, the confusion matrix helps us understand how well the model predicts the success or failure of the landing of SpaceX Falcon 9 Rocket Stage 1.

- The confusion matrix consists of four values:
- True Positive (TP): The model correctly predicts a successful landing.
- True Negative (TN): The model correctly predicts a failed landing.
- False Positive (FP): The model incorrectly predicts a successful landing (Type I error).
- False Negative (FN): The model incorrectly predicts a failed landing (Type II error).
- By analyzing the values in the confusion matrix, we can calculate various evaluation metrics such as accuracy, precision, recall, and F1-score. These metrics provide insights into the performance of the model in terms of correctly classifying the outcomes.

Conclusions

1. The Capstone Project focused on predicting the success of landing for SpaceX Falcon 9 Rocket Stage 1.
2. Data collection involved gathering data from SpaceX APIs and web scraping from Wikipedia tables.
3. Exploratory Data Analysis (EDA) provided insights into the relationships between variables and identified patterns and trends.
4. Various visualizations, including scatter plots, line plots, and bar graphs, were used to understand the data better.
5. Classification models, such as Logistic Regression, Support Vector Machine, Decision Tree Classifier, and K-Nearest Neighbors, were built and evaluated.
6. The Decision Tree Classifier emerged as the best-performing model with the highest accuracy.
7. Interactive visual analytics using Folium and Plotly Dash allowed for dynamic exploration and interactive visualizations.
8. Folium maps and markers were used to visualize launch site locations and landing outcomes.
9. Plotly Dash provided interactive charts and dropdown menus for site selection and success rate analysis.
10. The project's results demonstrated the effectiveness of machine learning techniques in predicting landing success and provided valuable insights for SpaceX and related stakeholders.

Thank you!