IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Thato Mogotsi
29 April 2024

Outline

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

Executive Summary

- **Summary of methodologies**

  - Data Collection and Data wrangling (REST APIs & Web Scraping)
  - Exploratory Data Analysis, EDA (SQL, Python Pandas, Numpy, Matplotlib)
  - Data Visualizations (Folium and Python Dashboard, Plotly Dash)
  - Predictive Model Building

- **Summary of all results**

  - Exploratory Data Analysis Results
  - Interactive Visualization Dashboard
  - Predictive Analysis by Classification

Introduction

SpaceX is a private aerospace manufacturer and space transportation company founded by Elon Musk in 2002. The company is known for its ambitious goals of reducing space transportation costs and enabling the colonization of Mars. SpaceX has achieved several milestones in the aerospace industry, including the development of the Falcon 1, Falcon 9, and Falcon Heavy launch vehicles, as well as the Dragon spacecraft.

Our goal is to forecast the successful landing of the Falcon 9 first stage. SpaceX prominently features Falcon 9 rocket launches on its website, pricing each launch at 62 million dollars. In contrast, other providers charge upwards of 165 million dollars per launch. A significant portion of SpaceX's cost savings stems from their ability to reuse the first stage. Thus, accurately predicting the success of the first stage landing enables us to estimate the overall launch cost. This insight can prove invaluable for competing companies considering bidding against SpaceX for a rocket launch contract.

Objectives:

- Which factor has the most significant impact on a successful landing?
- How does the success rate of landings change over time?
- Which predictive model provides the most accurate forecast for a successful landing?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
    - Gather data utilizing REST API and web scraping methodologies.
- Perform data wrangling
    - Refining the data, addressing null values, and implementing one-hot encoding to ready the dataset for analysis and modeling.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
    - Utilize classification models to make predictions regarding landing outcomes. Refine and evaluate the models through tuning to identify the most suitable model and parameter configurations.

Data Collection

• Initially, data acquisition commenced via the SpaceX API, a RESTful interface, facilitated by executing a GET request to the designated SpaceX API endpoint. This involved the development of auxiliary functions to streamline API utilization, specifically for extracting launch-related details utilizing unique identification numbers within the launch dataset. Subsequently, the rocket launch data was retrieved from the SpaceX API URL.

• Subsequent to retrieving the requested JSON outcomes, efforts were made to ensure uniformity in the SpaceX launch data. This entailed executing a GET request to fetch the SpaceX launch data, followed by decoding the content of the response as JSON. The resultant JSON was then transformed into a structured Pandas data frame for further processing.

• Additionally, web scraping techniques were employed to gather historical Falcon 9 launch records from a Wikipedia page titled "***List of Falcon 9 and Falcon Heavy launches*.**" The launch records were embedded within HTML format on the Wikipedia page. Leveraging the Beautiful Soup and requests libraries, the Falcon 9 launch HTML table records were extracted, parsed, and subsequently converted into a Pandas data frame for integration into the analysis pipeline.

Data Collection – SpaceX API

1. Initiate a request to retrieve data from the SpaceX API.
2. Decode the response using the `.json()` method and convert it into a dataframe using `json_normalize`.
3. Utilize custom functions to request information about launches from the SpaceX API.
4. Construct a dictionary from the retrieved data.
5. Generate a dataframe from the constructed dictionary.
6. Filter the dataframe to include only Falcon 9 launches.
7. Substitute missing values in the Payload Mass column with the calculated mean.
8. Export the processed data to a CSV file.

- Detailed code and insights pulled from above process flow can be found at: ***https://github.com/SirTido/IBM-Data-Science-Capstone-Project-SpaceX.git***

Data Collection - Scraping

- Implemented sophisticated web scraping techniques to procure comprehensive historical data pertaining to Falcon 9 rocket launches from a dedicated Wikipedia repository. Leveraging widely-recognized tools including BeautifulSoup and the requests library in Python, the intricate HTML representation of the Wikipedia page's tabular content was meticulously parsed and extracted. This meticulous process facilitated the creation of a meticulously structured dataframe, poised to undergo rigorous analysis and scrutiny. The resulting dataset offers invaluable insights into the trajectory and evolution of Falcon 9 launch missions, serving as a foundational resource for further research and investigation within the aerospace domain.

- URL: ***https://github.com/SirTido/IBM-Data-Science-Capstone-Project-SpaceX.git***

Data Wrangling

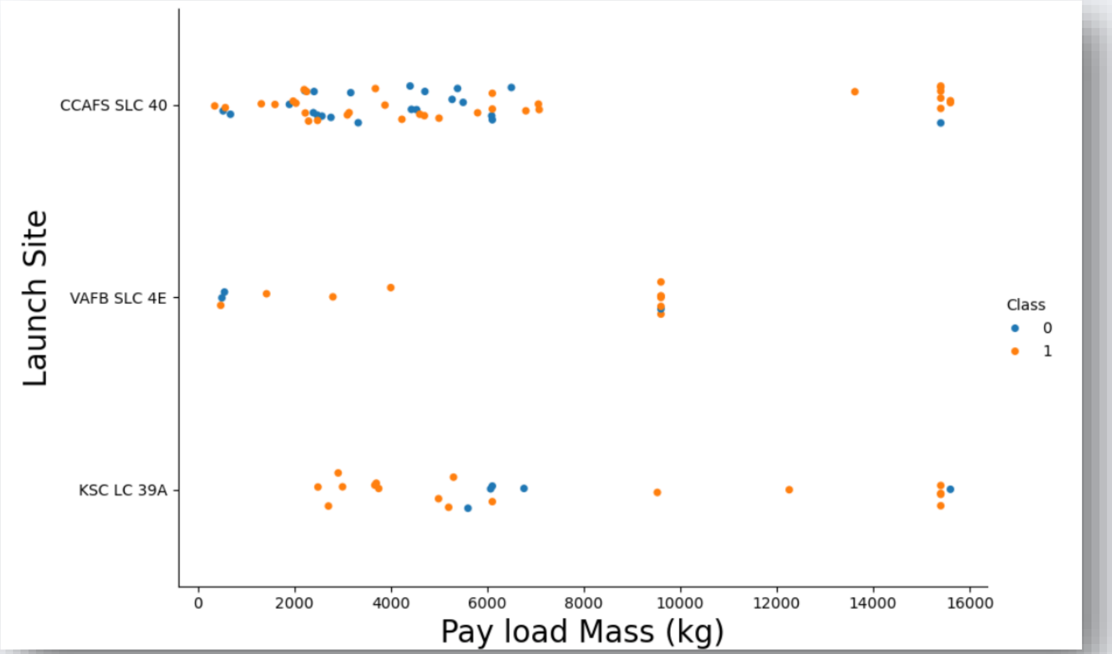
- Following the acquisition and creation of a Pandas DataFrame from the gathered dataset, filtering procedures were applied utilizing the ***BoosterVersion*** column to exclusively retain data pertinent to Falcon 9 launches. Subsequently, handling of missing values within the ***LandingPad*** and ***PayloadMass*** columns ensued. In the case of the ***PayloadMass*** column, absent data entries were imputed utilizing the mean value derived from the respective column.
- Furthermore, Exploratory Data Analysis (EDA) was conducted to unveil underlying patterns within the dataset. This comprehensive examination aimed to identify potential features suitable for training supervised models and to discern appropriate labels for predictive modeling endeavors.

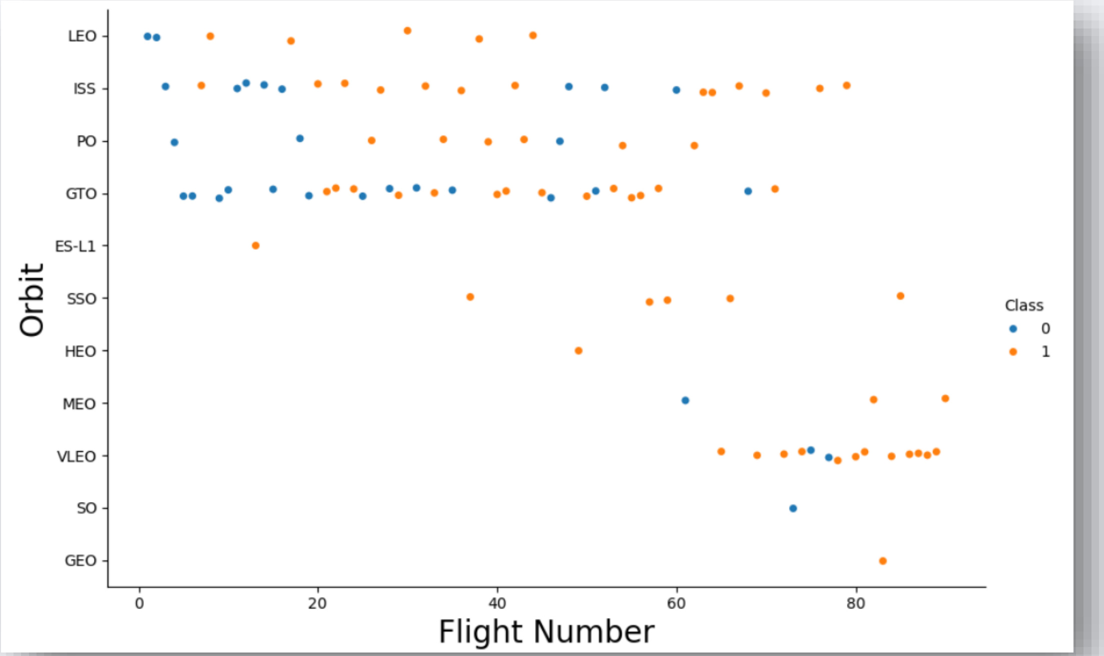
```
# landing_class = 0 if bad_outcome
# landing_class = 1 otherwise
df['Class'] = df['Outcome'].apply(lambda x: 0 if x in bad_outcomes else 1)
df['Class'].value_counts()
```

- URL: [***https://github.com/SirTido/IBM-Data-Science-Capstone-Project-SpaceX.git***](https://github.com/SirTido/IBM-Data-Science-Capstone-Project-SpaceX.git)

EDA with Data Visualization

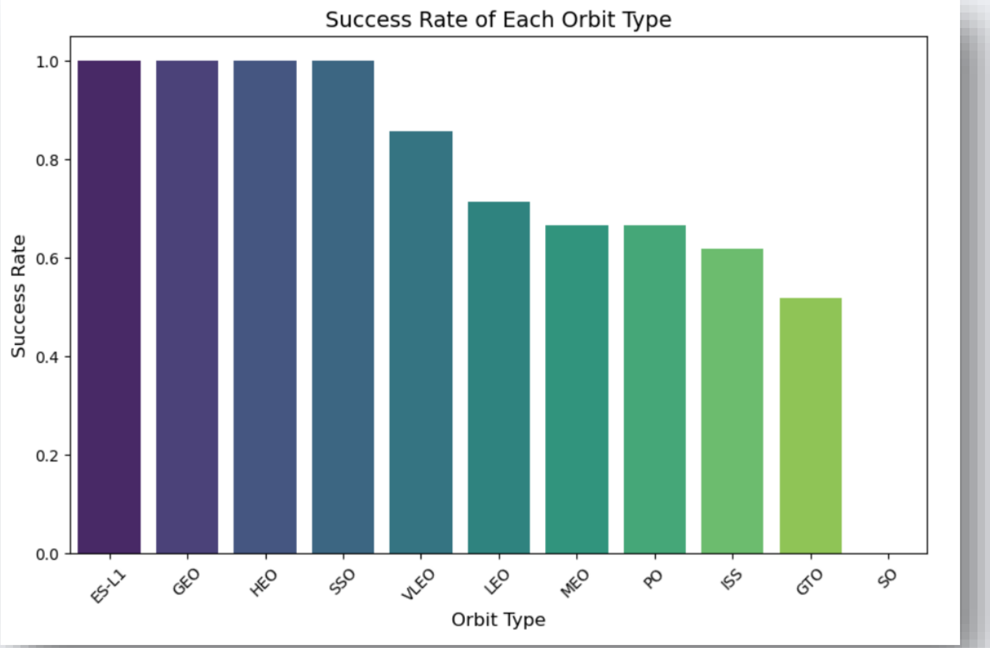
- The data visualization phase of the Exploratory Data Analysis (EDA) encompassed the generation of several essential plots, including the catplot, scatter chart, bar chart, and line chart.
- The catplot and scatter chart served to elucidate the relationships between launch sites and flight numbers, offering insights into launch frequency and distribution across different sites. Additionally, the scatter chart depicted correlations between launch sites and payload mass (in kilograms), as well as the associations between orbits and flight numbers, thereby providing a multifaceted view of mission characteristics.

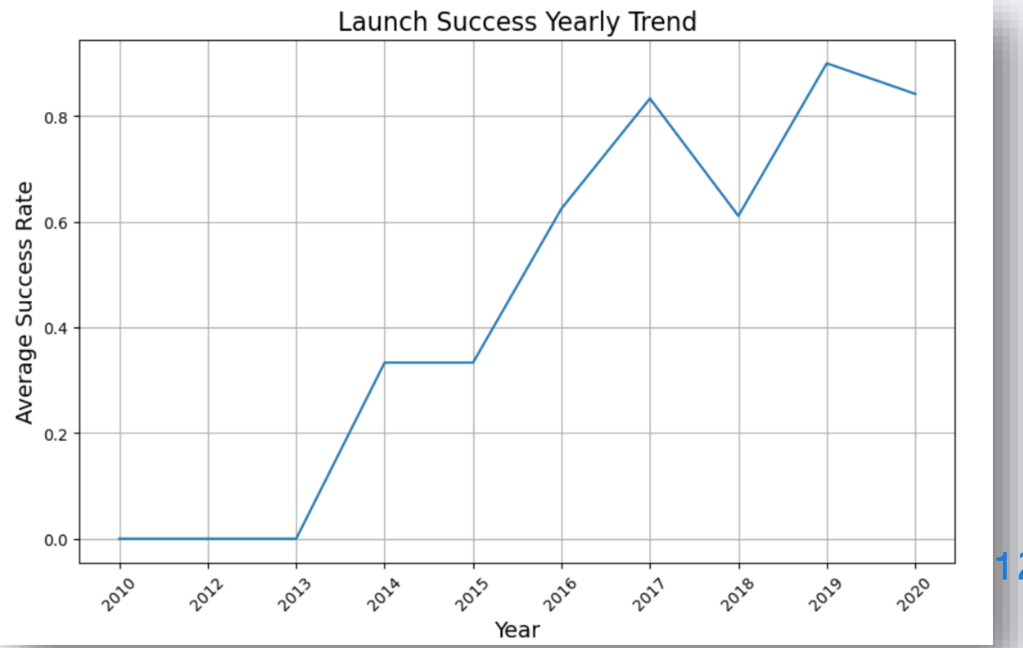




EDA with Data Visualization (cont…)

- The bar chart played a pivotal role in comparing and contrasting the success rates among different orbits. By visually representing the success rates of various orbits, this chart facilitated the identification of disparities and patterns across different mission trajectories.
- Lastly, the line chart provided a longitudinal perspective by illustrating the trend of successful launch rates from 2010 to 2020. This graphical representation enabled the observation of temporal patterns and fluctuations in launch success over the specified timeframe, thereby offering valuable insights into the historical performance of launch missions.





EDA with SQL

- Display the names of the unique launch sites in the space mission
- Display 5 records where launch sites begin with the string 'KSC'
- Display the total payload mass carried by boosters launched by NASA (CRS)
- Display average payload mass carried by booster version F9 v1.1
- List the date where the successful landing outcome in drone ship was achieved.
- List the names of the boosters which have success in ground pad and have payload mass greater than 4000 but less than 6000
- List the total number of successful and failure mission outcomes
- List the names of the booster versions which have carried the maximum payload mass. Use a subquery
- List the records which will display the month names, successful landing outcomes in ground pad, booster versions, launch site for the months in year 2017
- 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

- URL: ***https://github.com/SirTido/IBM-Data-Science-Capstone-Project-SpaceX.git***

Build an Interactive Map with Folium

- Placed a blue circle at the geographic coordinates of NASA Johnson Space Center, accompanied by a popup label displaying its name, derived from its latitude and longitude coordinates.
- Positioned red circles at the coordinates of all launch sites, each accompanied by a popup label exhibiting its name based on its latitude and longitude coordinates.
- Integrated colored markers denoting successful launches (green) and unsuccessful launches (red) at each launch site, providing visual insight into sites with high success rates.
- Incorporated colored lines to depict the distance between launch site CCAFS SLC 40 and its proximity to the nearest coastline, railway, highway, and city, facilitating spatial comprehension.

Build a Dashboard with Plotly Dash

- The dashboard provides interactive visualization and analysis tools for exploring space mission data:
- **Dropdown List with Launch Sites:** Users can choose to view data for all launch sites or a specific one. This feature enables targeted analysis based on launch site locations.
- **Slider of Payload Mass Range:** Users can define a range of payload masses to analyze. This functionality allows for filtering data based on the payload's weight, facilitating focused investigations.
- **Pie Chart Showing Successful Launches:** This chart illustrates the proportion of successful and unsuccessful launches relative to the total number of launches. Users can quickly grasp the success rate visually.
- **Scatter Chart Showing Payload Mass vs. Success Rate by Booster Version:** Users can explore the relationship between payload mass and launch success rates across different booster versions. This scatter chart facilitates the identification of any correlations between payload mass and launch success, aiding in understanding mission performance.

Predictive Analysis (Classification)

- Convert the 'Class' column into a NumPy array for further processing.
- Standardize the data using StandardScaler, fitting and transforming it to ensure uniformity in scale.
- Employ train_test_split to partition the data into training and testing sets, facilitating model evaluation.
- Instantiate a GridSearchCV object with cv=10 for parameter tuning and optimization.
- Utilize GridSearchCV to explore various algorithms, including logistic regression (LogisticRegression()), support vector machine (SVC()), decision tree (DecisionTreeClassifier()), and K Nearest Neighbor (KNeighborsClassifier).
- Evaluate the accuracy of each model on the test data using the .score() method.
- Analyze the confusion matrix for each model to assess its performance in classifying data points.
- Determine the best model based on Jaccard Score, F1 Score, and overall Accuracy metrics.

Results

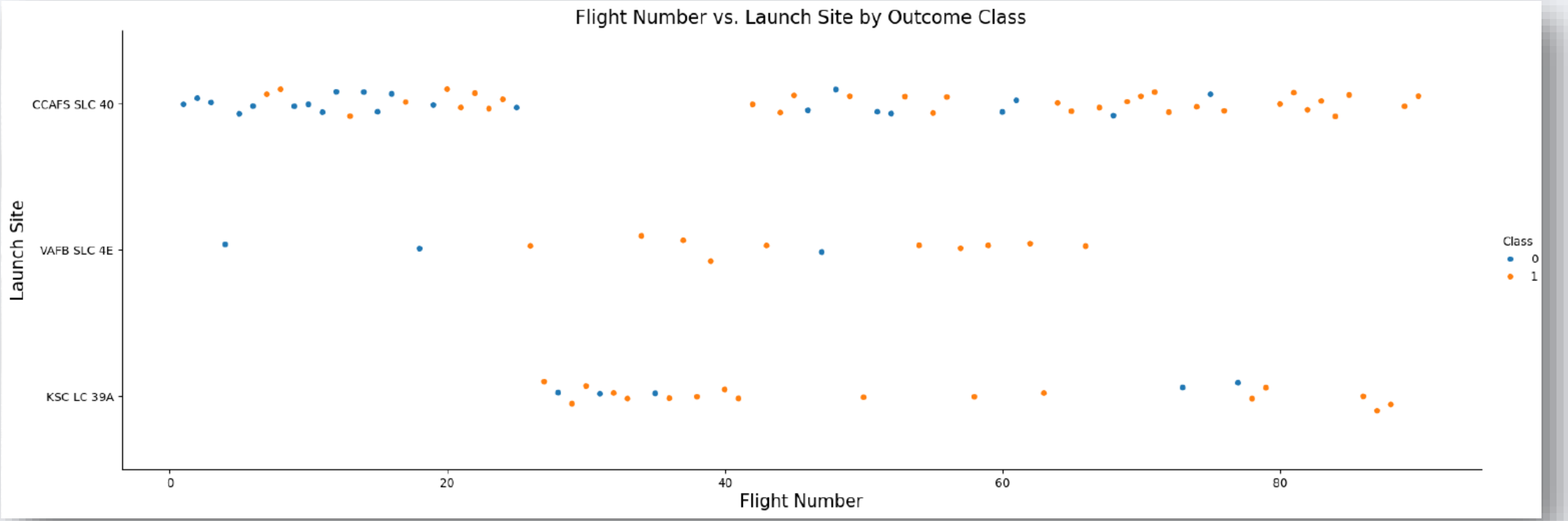
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

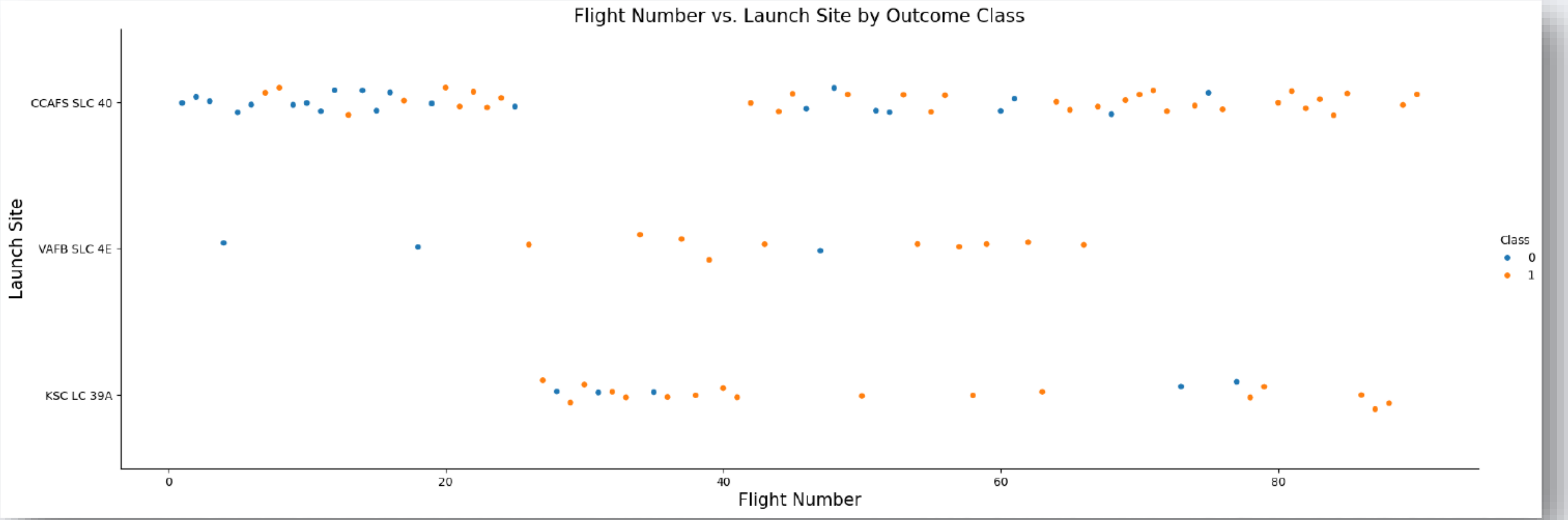
Flight Number vs. Launch Site

- Historical data indicates that earlier flights exhibited a lower success rate, depicted by the blue trend line.
- Conversely, later flights demonstrate a notable improvement in success rates, as denoted by the orange trend line.
- Approximately half of all launches originated from the CCAFS SLC 40 launch site, indicating its significance in space missions.
- Notably, the VAFB SLC 4E and KSC LC 39A launch sites exhibit comparatively higher success rates, suggesting their operational efficacy.
- From the observed trends, it can be inferred that newer launches tend to boast higher success rates, showcasing advancements and improvements in spaceflight technology and operations.



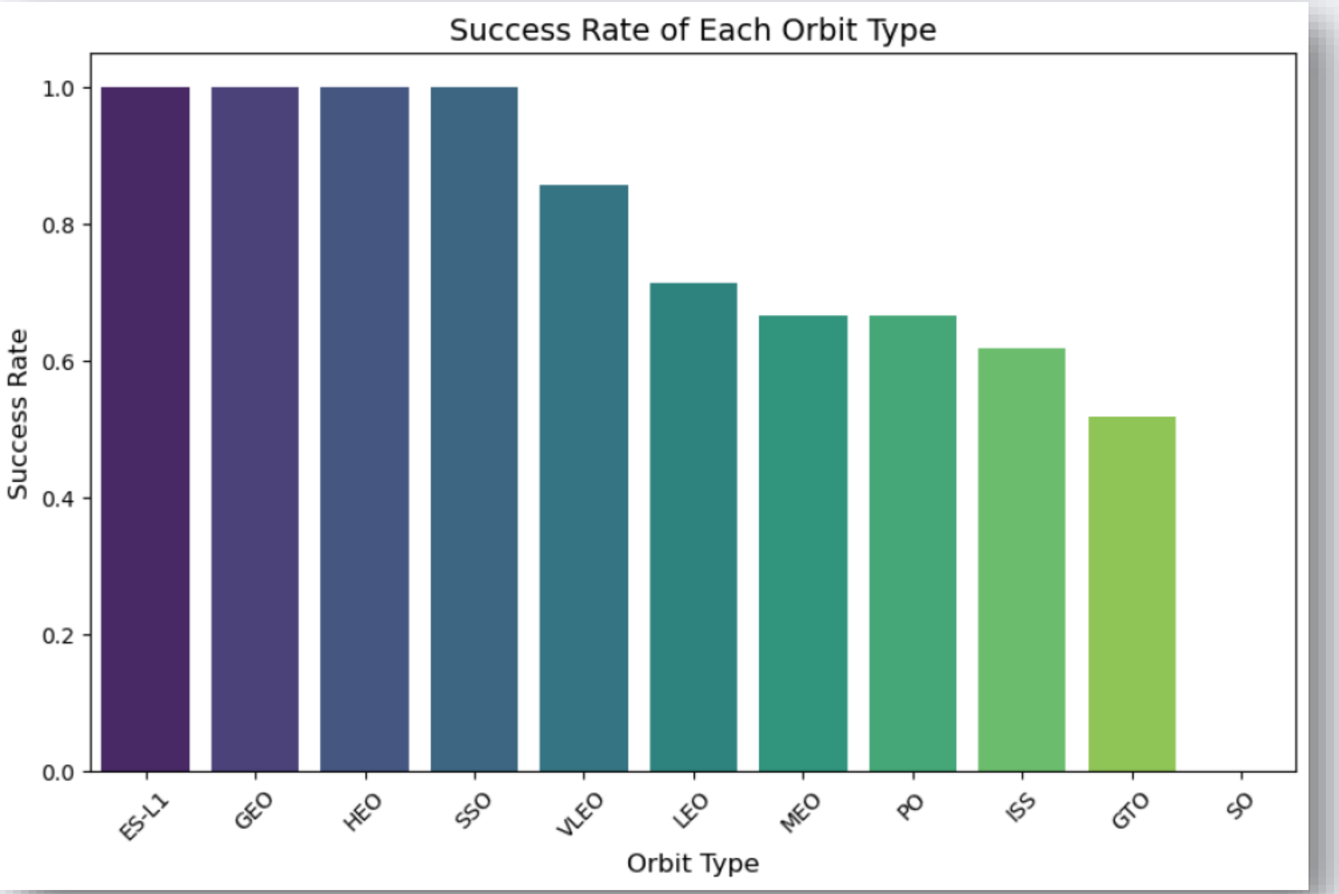
Payload vs. Launch Site

- Generally, there's a positive correlation between payload mass (kg) and success rate, indicating that higher payload masses tend to result in higher success rates.
- Specifically, the majority of launches with a payload exceeding 7,000 kg were successful, underscoring the robustness of missions with heavier payloads.
- Notably, launches from the KSC LC 39A site with payloads less than 5,500 kg achieved a 100% success rate, highlighting the reliability of this site for smaller payloads.
- Conversely, the VAFB SLC 4E launch site has not conducted any launches with payloads greater than approximately 10,000 kg, suggesting limitations or operational constraints specific to this site for heavier payloads.



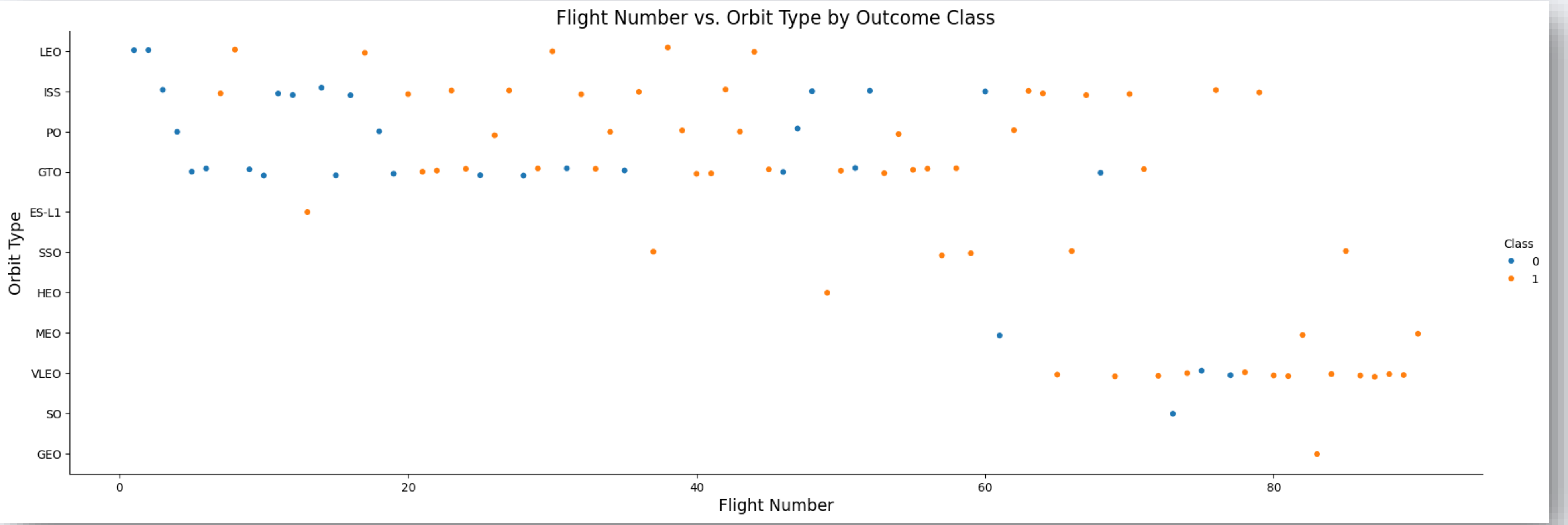
Success Rate vs. Orbit Type

- Launch missions to ES L1, GEO, HEO, and SSO orbits have achieved a 100% success rate, reflecting their reliability and effectiveness.
- Missions to GTO, ISS, LEO, MEO, and PO orbits have demonstrated success rates ranging from 50% to 80%, indicating moderate to high levels of success in these mission categories.
- However, missions to the SO orbit have yet to achieve success, as indicated by a 0% success rate, highlighting potential challenges or difficulties specific to this orbital trajectory.



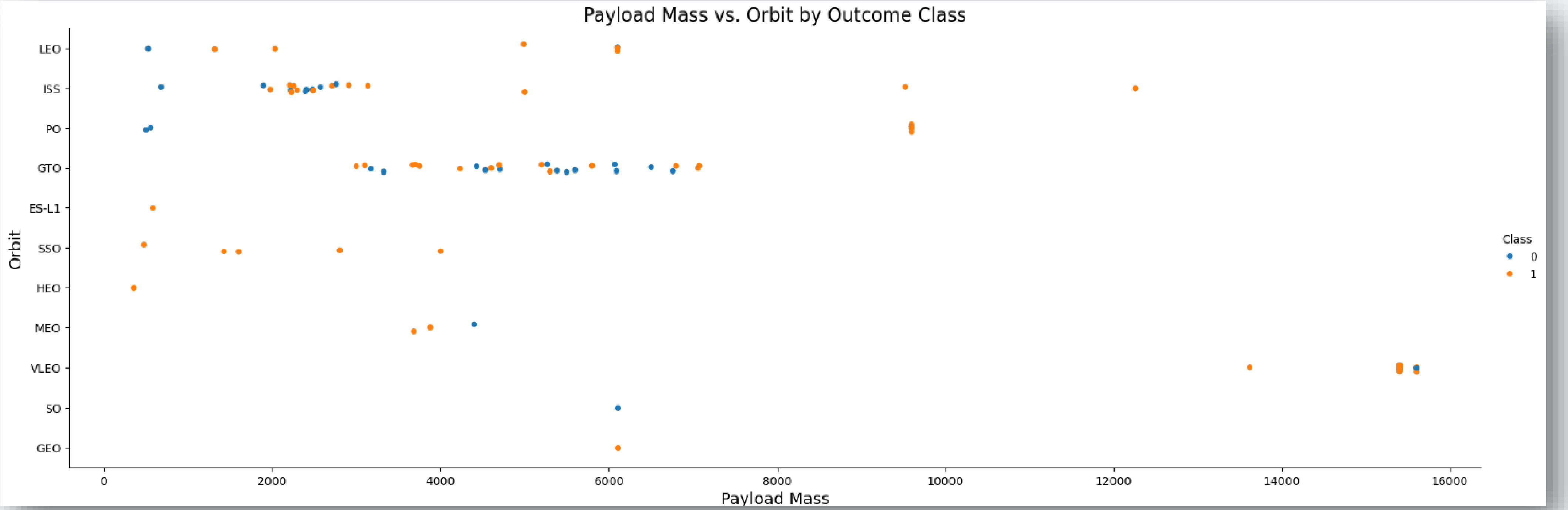
Flight Number vs. Orbit Type

- Success rates in Low Earth Orbit (LEO) demonstrate a positive correlation with flight numbers, indicating an increasing reliability and success rate with each subsequent flight.
- The Sun-Synchronous Orbit (SSO) consistently achieves a 100% success rate despite having fewer flights, indicating a high level of reliability and effectiveness in missions to this orbit.
- In contrast, there seems to be no discernible correlation between flight numbers in the Geostationary Transfer Orbit (GTO), suggesting that the success rate does not exhibit a consistent trend with increasing flight numbers.
- Notably, success rates tend to be higher for flight numbers exceeding 40, indicating an improvement in mission reliability and operational efficiency over time.



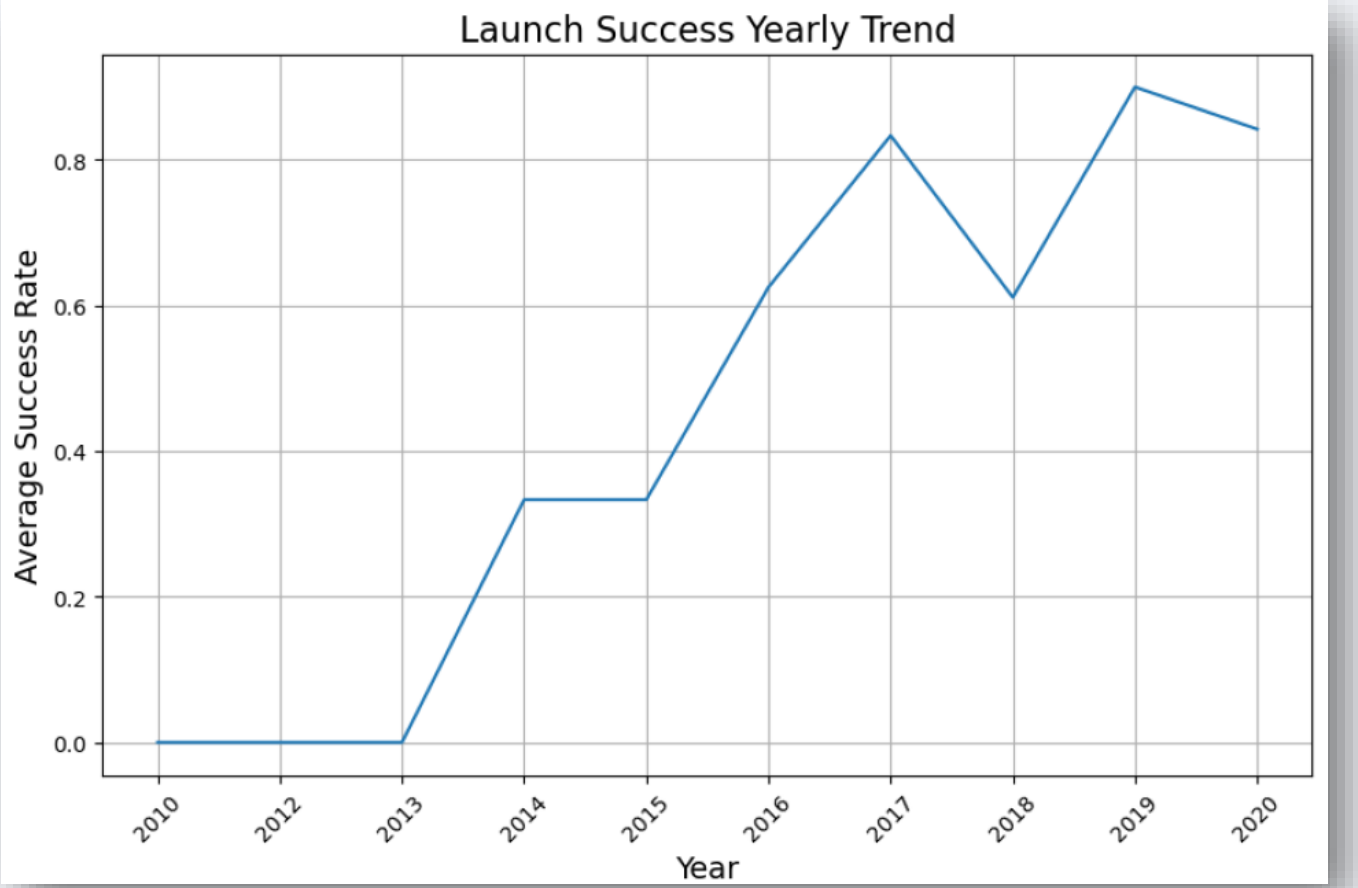
Payload vs. Orbit Type

- Very Low Earth Orbit (VLEO) typically accommodates the heaviest payloads among all orbits, indicating its capability to support missions with substantial payload masses.
- On the other hand, the Geostationary Transfer Orbit (GTO) exhibits mixed success rates when handling heavier payloads, suggesting variability in mission outcomes and operational challenges associated with heavier payloads in this orbit.



Launch Success Yearly Trend

- The success rate experienced improvement between 2013 and 2017, with further enhancements noted between 2018 and 2019.
- Conversely, a decrease in the success rate occurred from 2017 to 2018, as well as from 2019 to 2020.
- Despite fluctuations in specific years, the overall trend indicates an improvement in the success rate since 2013.



All Launch Site Names

- Utilized the DISTINCT clause to extract unique launch sites from the 'Launch_Site' column in the SPACEXTABLE.
- Identified distinct launch sites as CCAFS LC-40, VAFB SLC-4E, KSC 39A, and 40.

```
%%sql
SELECT DISTINCT launch_site
FROM SPACEXTABLE;
```

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

Launch Site Names Begin with 'CCA'

- Employed the LIMIT clause to restrict the output to the first 5 results.
- Utilized the LIKE clause to filter launch sites with names beginning 'CCA'.
- Displayed only the launch sites meeting the specified criteria.

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	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	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)
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

```
%%sql
SELECT * FROM SPACEXTABLE
WHERE launch_site LIKE 'CCA%'
LIMIT 5;
```

Total Payload Mass

- Applied the SUM() function to compute the total payload mass.
- Utilized the WHERE clause to filter boosters launched by NASA.
- Calculated the total payload mass carried by NASA boosters as 45596.

```
%%sql
SELECT SUM(payload_mass__kg_) AS Total_Payload_Mass, customer
FROM SPACEXTABLE
WHERE customer = 'NASA (CRS)';
```

```
 * sqlite:///data/SpacexDB.db
Done.
```

Total_Payload_Mass	Customer
45596	NASA (CRS)

Average Payload Mass by F9 v1.1

- The command WHERE Customer == 'NASA (CRS)' filters the boosters associated with NASA's Commercial Resupply Services (CRS).
- Subsequently, the SUM() command calculates the total payload mass of the filtered boosters.

```
%%sql
SELECT AVG(payload_mass__kg_) AS Average_Payload_Mass, booster_version
FROM SPACEXTABLE
WHERE booster_version = 'F9 v1.1';
```

```
 * sqlite:///data/SpacexDB.db
Done.
```

Average_Payload_Mass	Booster_Version
2928.4	F9 v1.1

First Successful Ground Landing Date

- The command WHERE Booster_Version == 'F9 v1.1' filters the launches featuring the booster version F9 V1.1.
- Following this, the command AVG() calculates the average payload mass of the filtered launches.

```
%%sql
SELECT MIN(date) AS Date_of_First_Successful_Landing_in_Ground_Pad
FROM SPACEXTABLE
WHERE landing_outcome = 'Success (ground pad)';
```

```
 * sqlite:///data/SpacexDB.db
Done.
```

Date_of_First_Successful_Landing_in_Ground_Pad
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- The command WHERE Landing_Outcome == 'Success (ground pad)' filters the launches with successful outcomes on the ground pad.
- Subsequently, the command MIN() retrieves the earliest date among the filtered launches.

```
%%sql
SELECT booster_version
FROM SPACEXTABLE
WHERE landing_outcome = 'Success (drone ship)'
AND payload_mass__kg_ > 4000
AND payload_mass__kg_ < 6000;
```

```
 * sqlite:///data/SpacexDB.db
Done.
```

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 command WHERE Landing_Outcome == 'Success (drone ship)' AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000 filters the launches based on specific criteria, including successful outcomes on the drone ship and payload masses between 4000 and 6000 kilograms.

```
%%sql
SELECT mission_outcome, COUNT(*) AS Total
FROM SPACEXTABLE
WHERE mission_outcome IN ('Success','Failure (in flight)','Success (payload status unclear)','Success')
GROUP BY mission_outcome;
```

```
 * sqlite:///data/SpacexDB.db
Done.
```

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

Boosters Carried Maximum Payload

- The subquery "SELECT MAX(PAYLOAD_MASS_KG_) FROM SPACEXTBL" retrieves the maximum payload mass.
- This subquery is included in the WHERE command to filter the boosters carrying the maximum payload mass.

```
%%sql
SELECT booster_version
FROM SPACEXTABLE
WHERE payload_mass__kg_ = (
    SELECT MAX(payload_mass__kg_)
    FROM SPACEXTABLE
);
```

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 WHERE command is utilized to filter the data according to specified criteria. Given that the date format appears as "20/06/2023," there is a need to use the substr() function to extract the month and the desired year (2015) for further analysis or processing.

```
%%sql
SELECT
CASE
    WHEN substr(date,6,2) = '01' THEN 'January'
    WHEN substr(date,6,2) = '02' THEN 'February'
    WHEN substr(date,6,2) = '03' THEN 'March'
    WHEN substr(date,6,2) = '04' THEN 'April'
    WHEN substr(date,6,2) = '05' THEN 'May'
    WHEN substr(date,6,2) = '06' THEN 'June'
    WHEN substr(date,6,2) = '07' THEN 'July'
    WHEN substr(date,6,2) = '08' THEN 'August'
    WHEN substr(date,6,2) = '09' THEN 'September'
    WHEN substr(date,6,2) = '10' THEN 'October'
    WHEN substr(date,6,2) = '11' THEN 'November'
    WHEN substr(date,6,2) = '12' THEN 'December'
END AS Month_Name, landing_outcome, booster_version, launch_site
FROM SPACEXTABLE
WHERE substr(date,0,5) = '2015'
    AND landing_outcome = 'Failure (drone ship)';
```

Month_Name	Landing_Outcome	Booster_Version	Launch_Site
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 substr() function is essential for selecting dates falling within the specified range of "2010-06-04" and "2017-03-20." This function allows for the extraction of the relevant date components (year, month, and day) from the date column, enabling accurate filtering based on the defined time frame.

```
%%sql
SELECT landing_outcome, COUNT(*) AS Count
FROM SPACEXTABLE
WHERE date BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY landing_outcome
ORDER BY Count DESC;
```

Landing_Outcome	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

SpaceX Launch Site Locations

All launch sites are situated in close proximity to the coast.

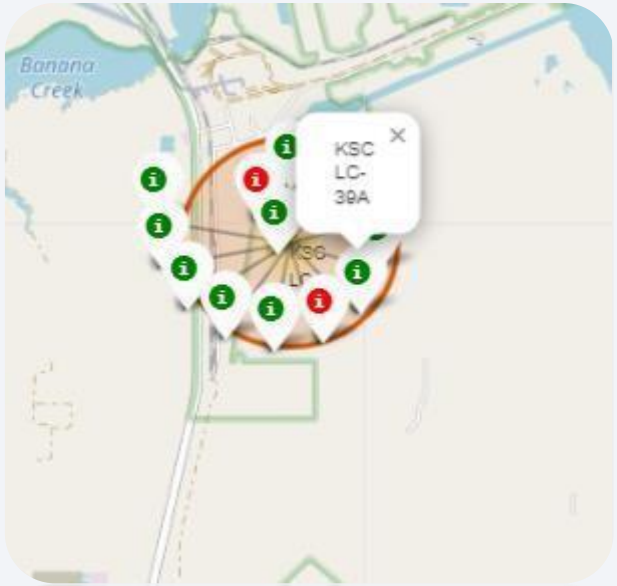
Due to the Earth's rotation, locations nearer to the equator benefit from a higher linear velocity, facilitating easier access to equatorial orbits with less fuel consumption. Rockets launched from sites close to the equator capitalize on this natural boost, leveraging Earth's rotational speed to reduce the need for additional fuel and boosters.

Consequently, launching from equatorial regions becomes more cost-effective and efficient compared to launches from higher latitudes.

Falcon 9 Success/Fail Launches by site

The successful launches are represented by a green marker, while the red marker represents failed rocket launches

Along the Eastern coast, Launch site KSC LC 39A exhibits comparatively higher success rates in contrast to CCAFS SLC 40 and CCAFS LC 40.



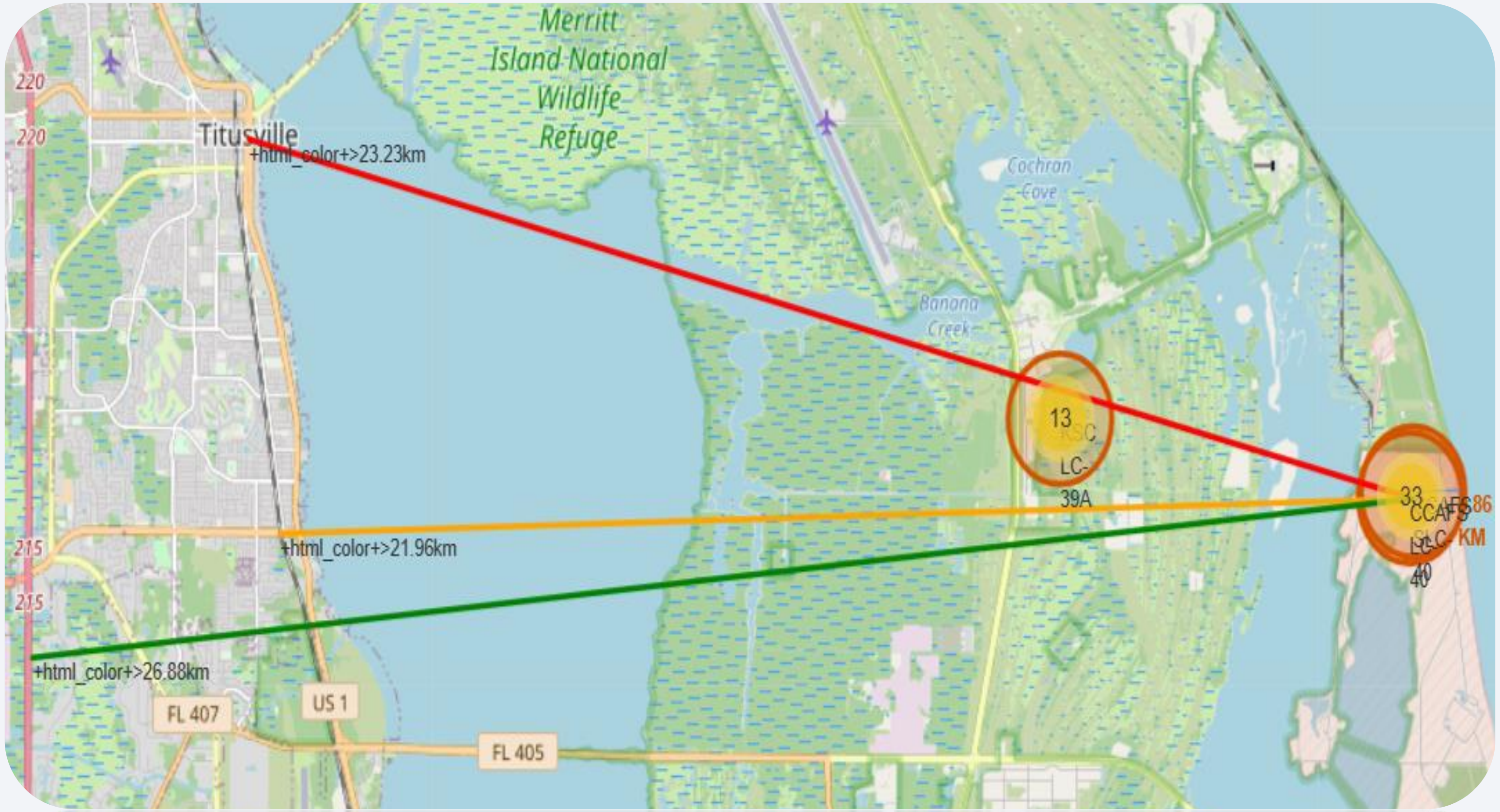
Falcon9 Launch Site Proximities

CCAFS SLC 40:

- Distance from nearest coastline: 86 km
- Distance from nearest railway: 21.96 km
- Distance from nearest city: 23.23 km
- Distance from nearest highway: 26.88 km

Considerations:

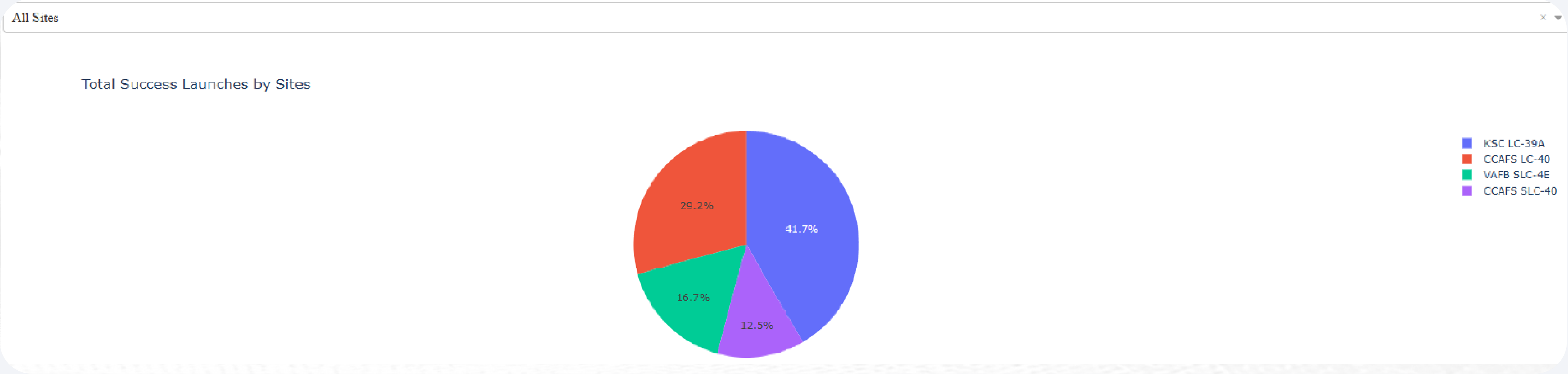
- Coasts: Ensure that spent stages or failed launches do not pose risks to people or property.
- Safety/Security: Implement an exclusion zone around the launch site to maintain safety and security by restricting unauthorized access.
- Transportation/Infrastructure and Cities: Balance the need to be far enough from potential damage caused by failed launches while still being close to transportation hubs such as roads, railways, and docks for logistical support during launch activities.



Section 4

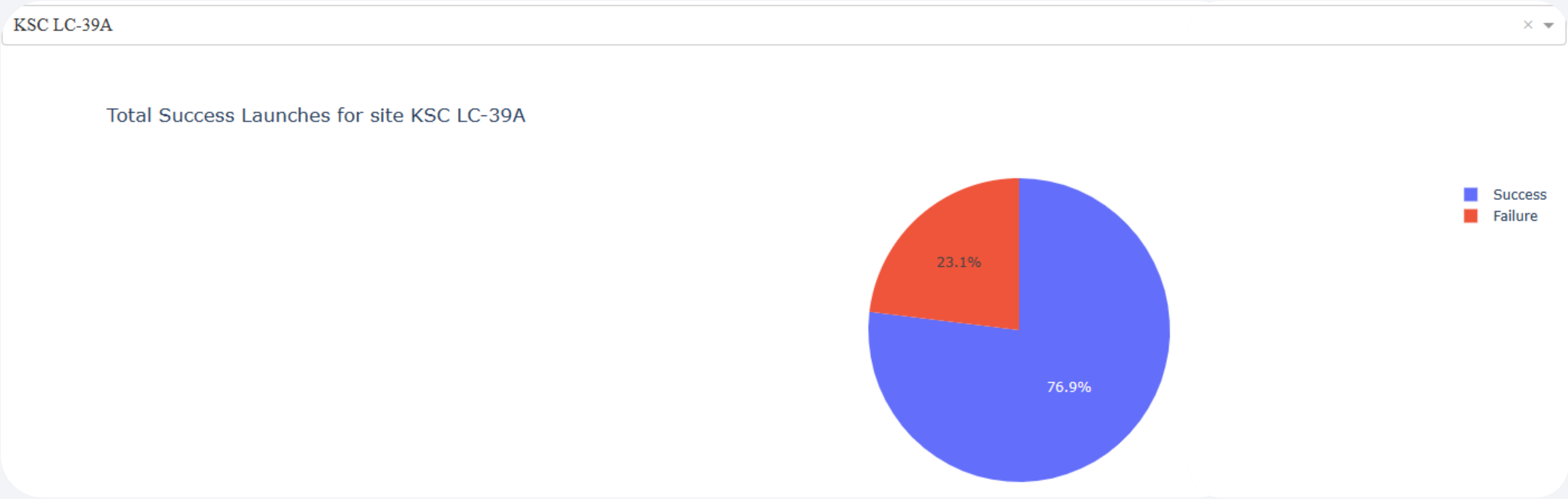
Build a Dashboard with Plotly Dash

Pie chart view of total Successful launces by Site



- Launch site KSC LC-39A has the highest launch success rate at 41.7%
- followed by CCAFS 40 with 29.2% success rate
- VAFB SLC-4E sis sitting at16.7%, and lastly
- 12.5% of launches at CCAFS SLC-40 were successful.

KSC LC-39A Launch Site Breakdown



Launch site CCAFS LC 40, which has the second largest success ratio, with 73% of its launches culminating in successful missions, while experiencing a 27% rate of unsuccessful launches.

This performance underscores the site's reliability and effectiveness in supporting successful rocket launches.
Such a high success rate signifies the site's proficiency in ensuring the smooth execution of missions, contributing significantly to its reputation as a dependable launch facility.

Payload vs Launch Sites for all Launch Sites

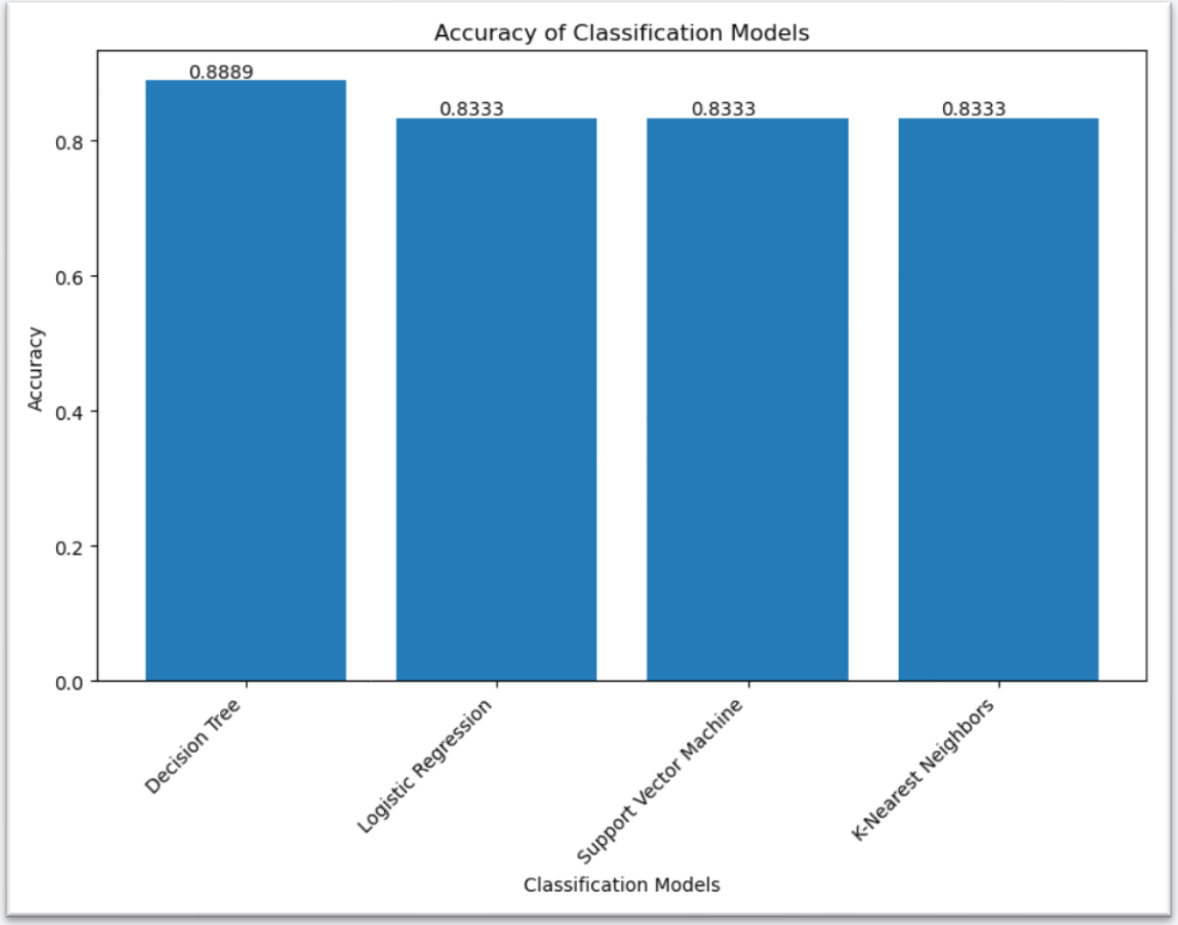
Analysis of the data reveals several notable trends:

- Payload masses ranging from 0 to 5,000 kilograms exhibit a higher success rate compared to payloads between 5,000 and 10,000 kilograms. This suggests that missions with lighter payloads tend to have a higher likelihood of success.
- Within the payload range of 2,000 to 4,000 kilograms, missions consistently achieve the highest success rate, indicating a favorable performance for payloads within this specific weight category.
- Booster version FT demonstrates the largest success rate among all booster versions, highlighting its reliability and effectiveness in facilitating successful launches.

Section 5

Predictive Analysis (Classification)

Classification Accuracy



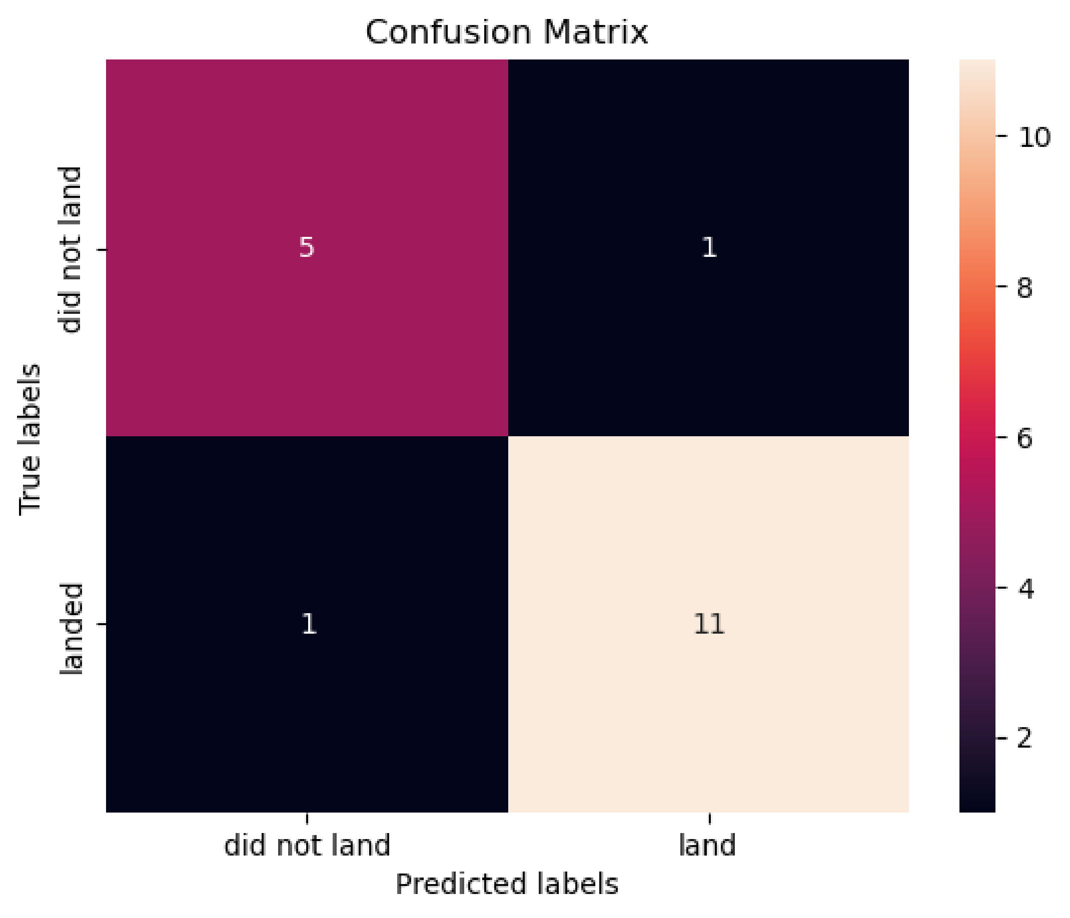
- Among the models evaluated, the Decision Tree Classifier stands out with its remarkable classification accuracy of 88.90%.

  This high level of accuracy suggests that the Decision Tree algorithm is adept at effectively discerning patterns and relationships within the dataset, enabling precise classification of data points into their respective categories.

  As a result, the Decision Tree Classifier emerges as a robust and reliable model for this particular classification task, showcasing its potential for real-world applications where accurate predictions are crucial.

Confusion Matrix

- The model only misclassified 1 instance each as false-negative and false-positive.
- This indicates that out of all the instances where the model made incorrect predictions, there were very few such instances.
- This suggests that the model has a high level of precision and accuracy, as it effectively minimizes the number of false classifications.
- Such a low error rate is indicative of the model's strong predictive capabilities and reliability in making accurate predictions.



Summary of Findings

- The project aimed to forecast the successful touchdown of SpaceX Falcon 9 first stages while discerning the key factors contributing to these successful landings.
- Data collection encompassed both API requests from SpaceX and web scraping from Wikipedia. This initial phase was succeeded by comprehensive data wrangling and exploratory data analysis (EDA).
- Exploratory data analysis (EDA) uncovered noteworthy findings, including an upward trend in success rates corresponding to flight numbers and divergent success rates across different orbit types.
- Interactive analytics, facilitated by Folium and Plotly Dash, generated visual representations depicting launch site locations, success rates, and payload outcomes for enhanced comprehension and exploration.
- Analysis revealed that launch sites are strategically positioned near coastlines and distant from urban areas, prioritizing safety. Additionally, proximity to railways and highways facilitates efficient transportation of personnel and equipment.
- Predictive analysis entailed training classification models to forecast landing outcomes. The decision tree classifier demonstrated superior performance, achieving an accuracy of 88.90%.
- The project furnished invaluable insights for stakeholders in space exploration, empowering informed decision-making in launch planning and operational endeavors.

Conclusions

- **Objective Questions:**
  - **Which factor has the most significant impact on a successful landing?**

    Our exploratory data analysis (EDA) uncovered numerous influential factors affecting the successful landing of the Falcon 9 first stage. These encompass flight number, orbit type, payload mass, launch site attributes, booster versions, and environmental conditions.

  - **How does the success rate of landings change over time?**

    Our analysis revealed discernible patterns suggesting heightened success rates with ascending flight numbers and certain orbit types such as ES-L1, GEO, HEO, and SSO. Moreover, payload mass demonstrated correlations with success rates across different orbit types. These findings serve as valuable indicators for assessing the probability of a successful landing.

  - **Which predictive model provides the most accurate forecast for a successful landing?**

    Using machine learning classification models trained on historical data, we can confidently predict the successful landing of a Falcon first stage. Among these models, the decision tree classifier emerged as the top performer in our project, achieving an impressive accuracy of 88.90%. This model proves to be highly effective in forecasting landing outcomes.

Appendix

- For a detailed technical and access to codes and data sets used, you can access my GitHub profile and Repositories here:
  [IBM Data Science Capstone Project SpaceX]

- 1 SpaceX Data Collection API.ipynb
- 2 SpaceX Data Collection Web Scrapping.ipynb
- 3 SpaceX Data Wrangling.ipynb
- 4 SpaceX Exploratory Data Analysis Using SQL....
- 5 SpaceX Exploratory Data Analysis with Visual...
- 6 SpaceX Interactive Visual Analytics with Foliu...
- 7 SpaceX Falcon9 launch Dashboard.py
- 8 SpaceX Machine Larning Prediction Analysis.i...

Thank you!