IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Davio Hong
5 Oct 2023

Outline

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

Executive Summary

- This capstone project revolves around predicting the successful landing of Falcon 9 first stages. Data acquisition was accomplished through the utilization of the SpaceXdata API and web scraping from Wikipedia. Subsequent to data collection, an extensive data wrangling process was carried out, which included column renaming, addressing missing values, column classification, and basic exploratory data analysis (EDA).
- Following the initial data cleaning, a deeper EDA was conducted employing SQL, and data analysis was facilitated through various visualization techniques, employing an array of plots. Further preparation of the data was performed to facilitate modeling. Visual analytics was also applied, utilizing the Folium library to gain insights into physical and environmental factors.
- A dashboard was created to present key findings. In the final phase, predictive modeling was undertaken using the preprocessed and cleaned data. A selection of classification models was employed, with hyperparameter tuning for optimization.
- The analysis revealed a correlation between an increasing flight number and a higher success rate. Additionally, it was discovered that launch sites tend to be located in close proximity to highways and coastlines. Ultimately, the Decision Tree Classifier emerged as the best-performing model, exhibiting the highest accuracy on test data.

Introduction

- **Project Overview:**
- SpaceX offers Falcon 9 rocket launches at a significantly lower cost than competitors, mainly due to their ability to reuse the first stage. This project aims to develop a machine learning model to predict the success of the first stage landing, allowing for cost estimation of the entire mission. The project seeks to answer key questions:

- **Key Questions:**
- What factors determine successful rocket stage landings?
- How do these factors interact to affect success rates?
- What operational conditions are crucial for successful landings?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - One dataset was extracted by using SpaceXData API
  - Other dataset was webscraped using BeautifulSoup from wikipedia
- Perform data wrangling
  - Performed basic EDA
  - Determined training labels
- Perform exploratory data analysis (EDA) using visualization and SQL

• Visualized and analysed data using scatterplots

- Perform interactive visual analytics using Folium and Plotly Dash
  - Built a dashboard to view piechart and scatter plots according to each site
- Perform predictive analysis using classification models

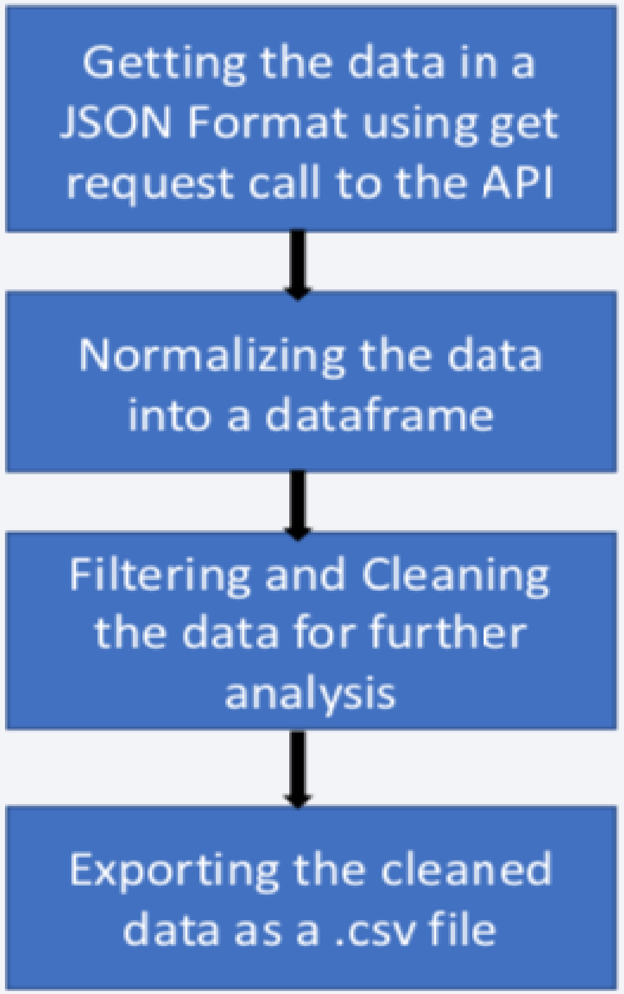
• Trained classification models and determined the best model, and also plotted confusion matrix for each model.

Data Collection

- **Project Objective:**
- The primary objective was to acquire SpaceX launch data with the aim of conducting data cleansing, analysis, visualization, and modeling to ascertain the likelihood of a successful launch.
- **Data Collection and Processing:**
- Two distinct datasets were obtained and processed as follows:
- One dataset was sourced through the SpaceX API, and subsequent normalization of the JSON responses was performed to construct a structured dataframe. Subsequent phases included additional preprocessing and in-depth analysis.
- Another dataset, encompassing launch dates and pertinent information, was gathered through web scraping from the Falcon9 Wikipedia page, employing the BeautifulSoup package. This process involved the systematic extraction of relevant data from tables and its transformation into a dataframe, thus facilitating further comprehensive analysis.

Data Collection – SpaceX API

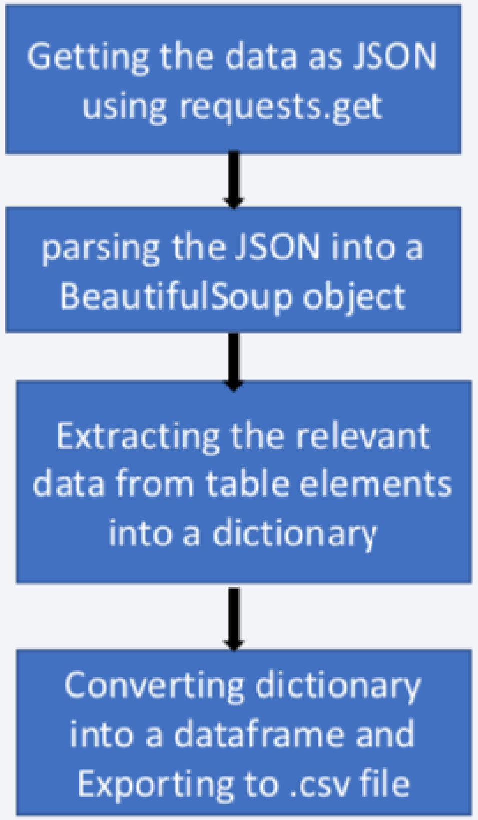
- **Data Acquisition from SpaceX API:**
- The initial dataset is meticulously gathered through Python utilizing the SpaceX API. The process involves executing HTTP GET requests to retrieve the requisite data. Subsequently, we undertake the normalization of the JSON content, transforming it into a structured dataframe. Employing a combination of specialized functions and the Pandas library, we systematically extract pertinent information, conduct data refinement procedures, and ultimately export the refined dataset.
- GitHub URL



Data Collection - Scraping

- An essential step in our data acquisition process involv
scraping to obtain historical launch records of the Falcc
dedicated Wikipedia page titled 'List of Falcon 9 and Fa
launches.' This was accomplished by employing the
response.get() method to request the page and subseq
parsing the retrieved data as a BeautifulSoup object. Fr
we systematically extracted all column and variable nar
the HTML table header and expertly transformed this i
into a structured dataframe.

- GitHub URL



Data Wrangling

- **Prior to Data Wrangling, Comprehensive Exploratory Data Analysis (EDA) Was Conducted:**
- The following analytical tasks were undertaken:
- Computation of the total number of launches at each launch site.
- Determination of the frequency and distribution of each targeted orbit for every launch.
- Examination of the frequency and occurrence of mission outcomes for each orbit type.
- Additionally, a pivotal step was taken to generate a landing outcome label based on the 'Outcome' column, where a classification of 0 indicated a failure, while a classification of 1 signified a successful landing.
- Upon the completion of these critical analyses, the meticulously cleaned data was exported to a .csv file for further utilization.

- GitHub URL

EDA with Data Visualization

- A systematic approach to data visualization was employed to discern pertinent relationships between key variables:
- Scatterplots were generated to visualize the correlation between Flight Number and Launch Site, Payload and Launch Site, Flight Number and Orbit Type, as well as Payload and Orbit Type. These scatterplots served as valuable tools for trend observation and the elucidation of overall variable relationships.
- To facilitate side-by-side comparisons, bar charts were utilized to visualize the success rates associated with each orbit type. This graphical representation allowed for a clear assessment of success rates within different orbit categories.
- A line plot was deployed to vividly depict the annual trend in launch success. Line plots excel at illustrating trends, revealing whether success rates have been on an upward or downward trajectory over time.
- In preparation for data modeling, a critical step involved the application of one-hot encoding to the features, ensuring that the data was suitably structured for subsequent modeling and analysis.

- GitHub URL

EDA with SQL

- Our initial step involved the seamless loading of the dataset into an IBM Db2 database, facilitated through the utilization of the IBM Db2 API. This strategic database connection provided us with a robust platform for conducting an in-depth Exploratory Data Analysis (EDA) through SQL-based queries.
- Within this analytical framework, the following key insights were extracted:
- Identification of unique launch sites.
- Querying and retrieval of launch sites commencing with the acronym 'CCA.'
- Calculation of the total payload mass carried by boosters launched under NASA's CRS program.
- Computation of the average payload mass carried by booster version F9 v1.1.
- Determination of the date marking the achievement of the first successful landing outcome on a ground pad.
- Aggregation of the total count of successful and failed outcomes.
- Compilation of the names of booster versions responsible for carrying the maximum payload mass.
- Ranking the count of landing outcomes within a specified date range, offering valuable insights into the success rate dynamics.

- GitHub URL

Build an Interactive Map with Folium

- To enhance the geographical representation within our project, several key elements were seamlessly integrated using the Folium library:
- **Circles:** These were employed to delineate highlighted circular areas on specific sites, accompanied by informative text labels.
- **Markers:** Utilized to pinpoint and denote specific sites of interest on the map.
- **Marker Clusters:** A practical inclusion to streamline the map display, particularly when numerous markers shared identical coordinates.
- **Mouse Position:** Implemented to provide precise coordinates corresponding to the position where the mouse cursor is placed on the map, enhancing interactivity and geospatial awareness.
- **Polyline:** Employed for the purpose of drawing connecting lines from sites to their nearest geographical features, such as coastlines, cities, and highways. This feature served to visually enhance the contextual relationships between sites and their surroundings on the map.

- GitHub URL

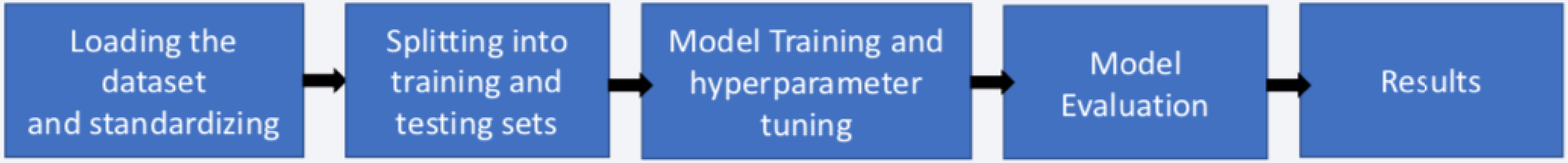
Build a Dashboard with Plotly Dash

Meticulously crafted an interactive dashboard using the Plotly Dash framework, featuring the following functionalities:

- **Dropdown Selection:** Integration of a dropdown menu to enable users to select a specific launch site. This selection triggered the display of a pie chart, offering a concise representation of the relative success and failure rates for launches at the chosen site.
- **Range Slider:** Incorporation of a range slider, specifically designed for payload specifications. This feature allowed users to define a payload range of interest and subsequently visualize a scatterplot illustrating the correlation between payload and class, all while filtering the data based on the selected launch site and the provided payload range.
- These dynamic visualizations and interactive components served as powerful tools for expeditiously examining the intricate relationship between payloads and launch sites. This, in turn, facilitated the identification of the optimal launch site for Falcon 9 missions.

- GitHub URL

Predictive Analysis (Classification)



- Initiated the process by organizing predictor and target variables into separate dataframes. The predictor data underwent standardization using the StandardScaler.
- Subsequently, we divided the data into training and test sets, with the "class" column serving as our target variable and all other columns as features.
- For our classification task, we explored various models including logistic regression, SVM, Decision Trees, and KNN. Model optimization was achieved through GridSearchCV with 10-fold cross-validation to identify the most suitable hyperparameters.
- To assess model performance, we created confusion matrices and model scores using the test data. The results will be consolidated to determine the best-performing machine learning model.

- GitHub URL

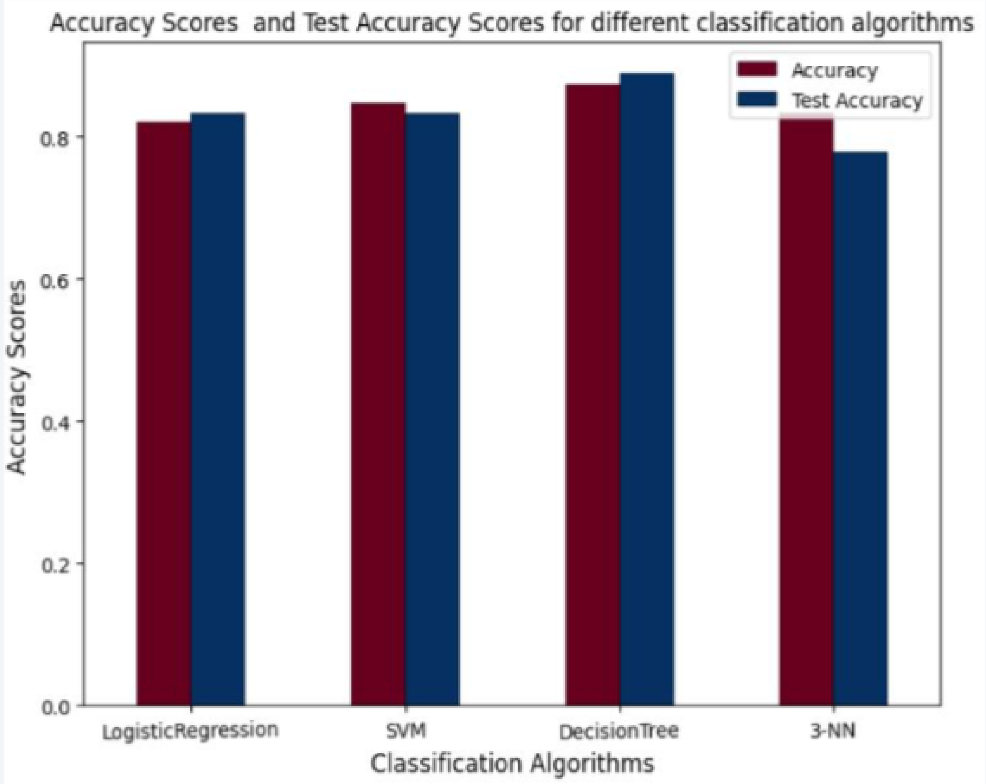
Results

- **Key Findings from Exploratory Data Analysis:**
- Our exploration of the data unveiled several pivotal insights:
- **Payload Capacity:** Newer rockets demonstrated an increased capacity for carrying payloads, highlighting technological advancements.
- **Payload Success:** Payloads exceeding 8000kg exhibited a notably high success rate, indicating the efficiency of SpaceX's capabilities in handling heavier payloads.
- **Success Rate Trend:** From 2013 to 2020, there was a consistent upward trajectory in the success rate of launches, reflecting continual improvement.
- **Early Challenges:** The period between 2010 and 2013 exhibited no recorded successful launches, underscoring initial challenges faced by SpaceX.
- **Launch Sites:** SpaceX operates from four distinct launch sites, emphasizing their geographical diversity.
- **VLEO Orbit:** Launches to Very Low Earth Orbit (VLEO) demonstrated a 14-launch history with an impressive 85% success rate, suggesting efficiency in executing these missions.
- **Successful Landings:** The first successful ground pad landing occurred in 2015, marking a significant milestone five years after the inaugural launch.
- **Booster F9:** Notably, missions utilizing booster version F9 consistently achieved successful outcomes.
- **Landing Outcomes:** Around 70 landing attempts were successful, while approximately 22 missions had no landing attempts, and about 10 resulted in failures.
- **Technological Advancements:** Over time, the success rate exhibited a discernible upward trend, attributed primarily to advancements in technology and operational expertise.

Results

- Our interactive analytics have yielded valuable observations:
- **Geographical Proximity:** Launch sites exhibit a strategic placement near the equator, optimizing the efficiency of space launches.
- **East Coast Dominance:** A significant majority of launches occur at launch sites situated along the east coast, emphasizing their preferred geographical locations.
- **Transportation Accessibility:** Launch sites are strategically positioned in close proximity to railways and highways, ensuring efficient logistical support for missions.
- **Coastline Proximity:** Many launch sites are strategically located near coastlines, likely for ease of access and safety considerations.
- **Urban Buffer:** Launch sites maintain a distinct distance from urban areas, underscoring the importance of safety and minimizing potential risks associated with rocket launches.
- These insights collectively reflect a thoughtful and strategic approach in the selection and positioning of launch sites.

Results

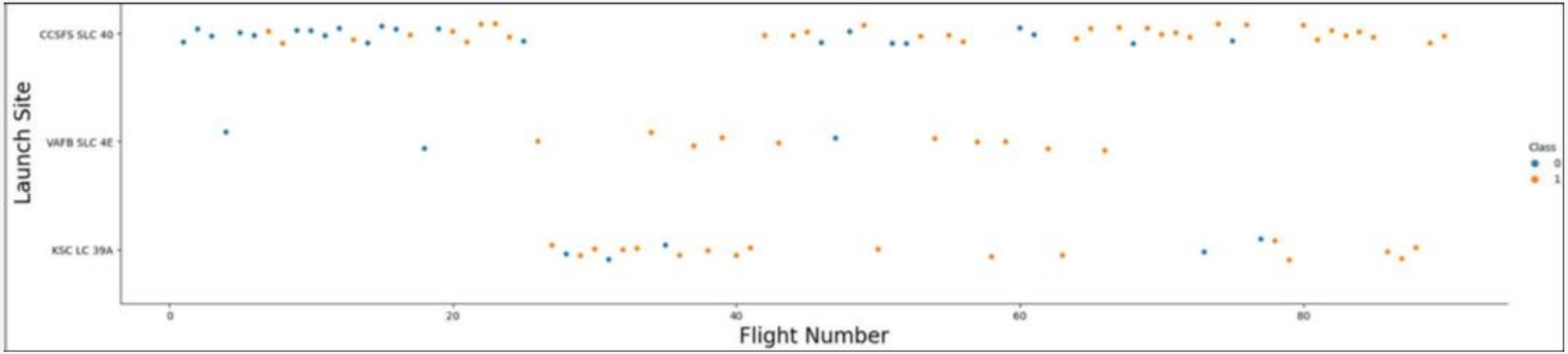


- The results of our predictive analysis reveal essential insights:
- **Model Performance Consistency:** With the exception of KNN, all models displayed comparable levels of accuracy on the test data, indicating relative consistency in predictive capability.
- **Top-Performing Model:** The Decision Tree Classifier emerged as the most effective model, boasting an impressive accuracy rate of approximately 87.5% on the training data and a commendable 88.9% accuracy on the test data.
- **Least Performing Model:** KNN demonstrated the lowest performance among the models, with an approximate mean accuracy rate of 80%, albeit still a respectable level of accuracy.
- These findings underscore the suitability of the Decision Tree Classifier as the preferred choice for accurate predictive modeling in this context.

Section 2

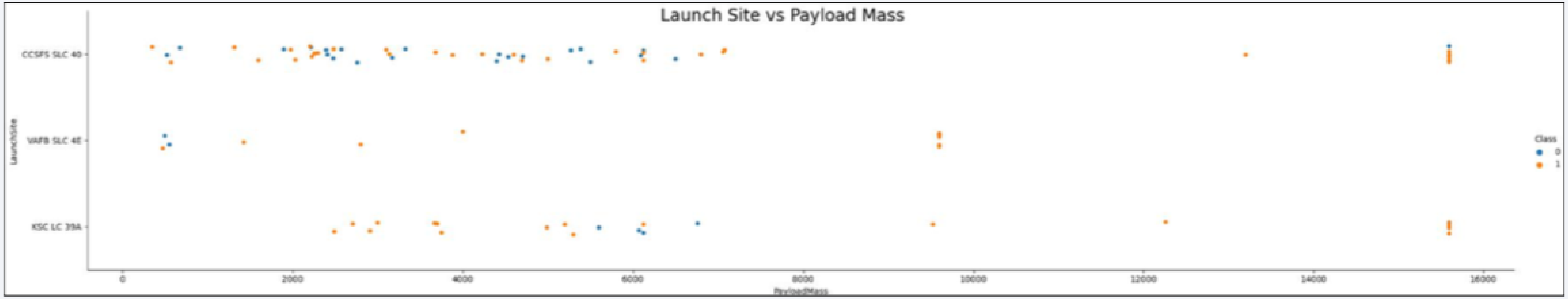
Insights drawn from EDA

Flight Number vs. Launch Site



- Analysis has yielded significant observations:
- **CCAFS LC-40 Performance:** CCAFS LC-40 exhibits an overall lower success rate compared to the other two launch sites, primarily due to a significant number of failures during its initial flights.
- **KSC LC-39A and VAFB SLC 4E:** These two launch sites, KSC LC-39A and VAFB SLC 4E, showcase nearly identical success rates. Their relatively higher flight numbers contribute to a lower failure rate, highlighting their consistent performance.
- **Optimal Site:** CCAFS SLC-40 emerges as the preferred launch site, with a notably high success rate in recent times. This suggests its reliability and effectiveness for space launches, positioning it as the top choice among the options.

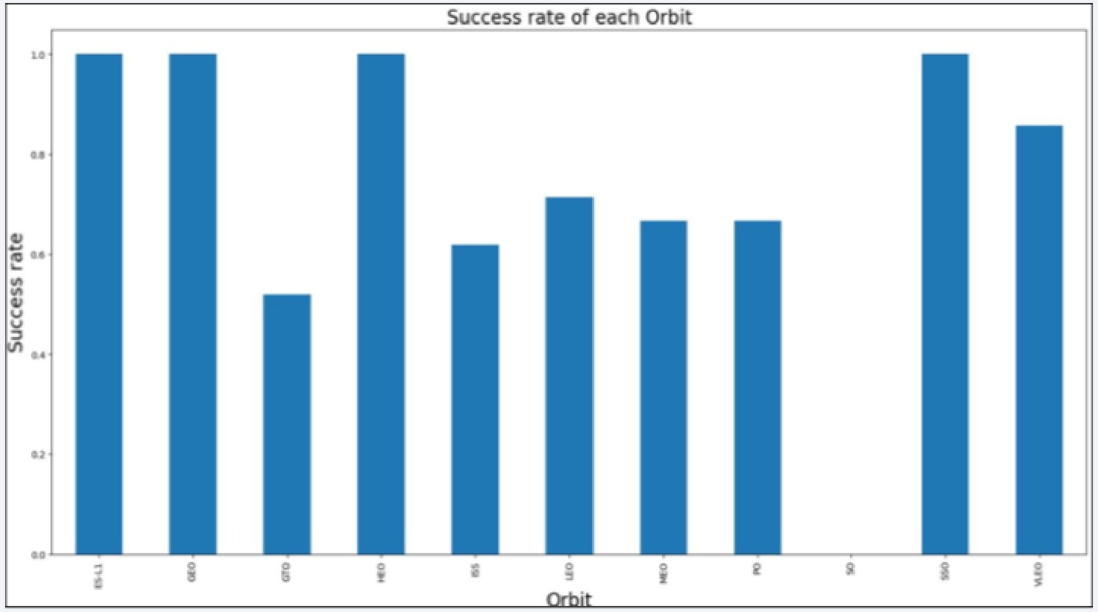
Payload vs. Launch Site



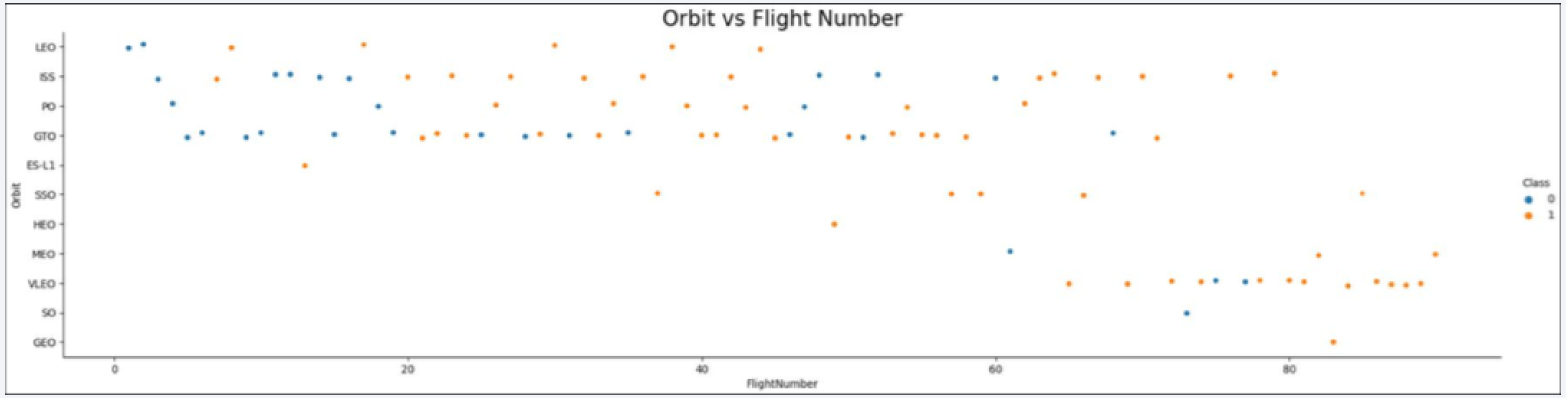
- Several noteworthy observations have emerged from the analysis:
- **VAFB-SLC LaunchsiteA:** Notably, no rockets carrying heavy payloads (greater than 10,000 kg) were launched from the VAFB-SLC LaunchsiteA. This suggests a potential limitation or strategic focus for this launch site.
- **Payload Success:** Payloads exceeding 8,000 kg consistently exhibit a high success rate, underscoring the robust capabilities of the launch system in handling heavier payloads.
- **CCAFS SLC-40:** Conversely, payloads with mass less than 6,000 kg at the CCAFS SLC-40 launch site are associated with a higher failure rate. This observation indicates potential challenges or factors specific to this launch site that warrant further investigation and mitigation measures.

Success Rate vs. Orbit Type

- Analysis, supported by the visual representation in the bar chart, has yielded insightful observations:
- **GEO, HEO, ES-L1, SSO:** These specific orbit types exhibit a 100% success rate, reflecting the effectiveness and reliability of launches targeting these orbits.
- **SO Orbit:** In contrast, the SO orbit has a single launch with a 0% success rate, suggesting challenges or unique complexities associated with this particular orbital destination.
- **ISS Orbit:** The ISS orbit features 21 launches with a success rate of approximately 61%. This relatively moderate success rate may indicate specific challenges associated with missions to the International Space Station.
- **VLEO Orbit:** Launches to Very Low Earth Orbit (VLEO) are marked by a notably high success rate of approximately 85%, emphasizing the efficiency and proficiency in executing missions to this orbit.
- **PO Orbit:** For Polar Orbit (PO), there are nine launches with an approximate success rate of 65%, denoting consistent performance in achieving mission objectives.



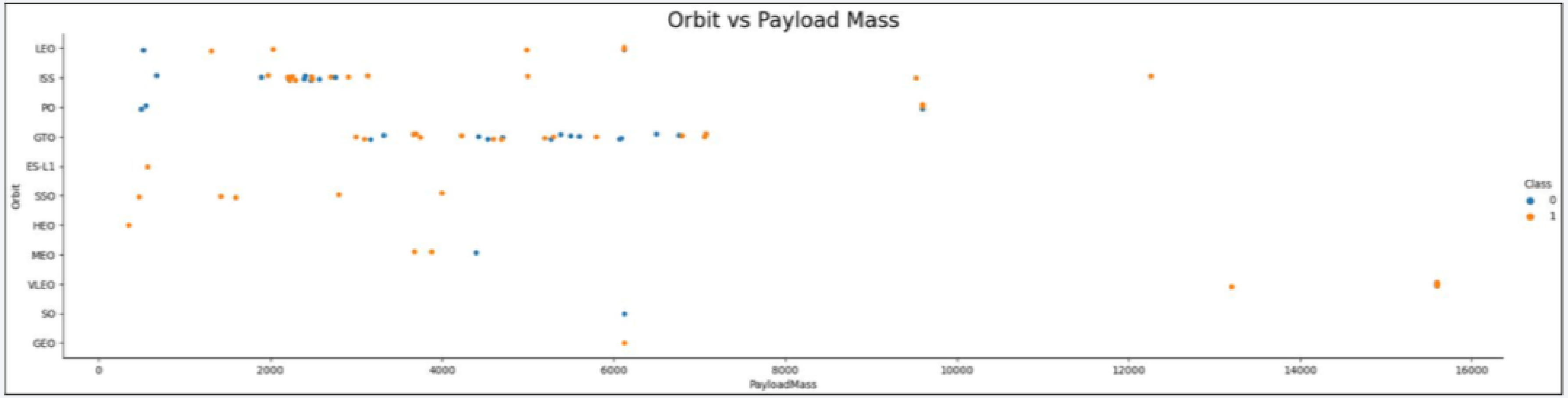
Flight Number vs. Orbit Type



Careful examination of the data reveals the following noteworthy patterns:

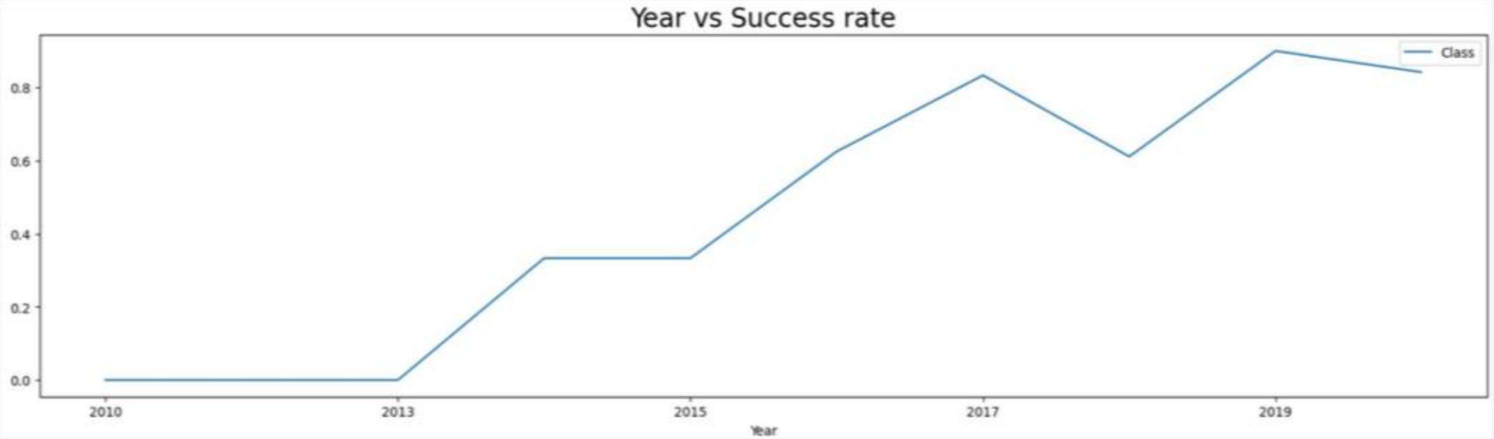
- **LEO Orbit:** In the case of Low Earth Orbit (LEO), it appears that the success rate is influenced by the frequency of flights. Higher numbers of flights correspond to higher success rates. However, in contrast, this relationship does not hold for flights to Geostationary Transfer Orbit (GTO).
- **Initial Flight Success:** Interestingly, there is a noticeable trend of lower success rates for the first flight, followed by a consistent increase in success rates as the flight number progresses. This trend suggests a learning curve or an improvement in operational procedures over time.

Payload vs. Orbit Type



- Analysis has yielded pertinent insights:
- **Heavy Payloads:** When dealing with heavy payloads, there is a noticeable trend of higher rates of successful or positive landings for Polar Orbit (PO), Low Earth Orbit (LEO), and International Space Station (ISS) orbits. This suggests a robust capability for handling heavier payloads in these scenarios.
- **GTO Orbit Complexity:** In the case of Geostationary Transfer Orbit (GTO), distinguishing trends related to heavy payloads is less straightforward. GTO exhibits a mix of positive landing rates and negative landing outcomes (unsuccessful missions), making it more complex to discern specific patterns associated with heavy payloads in this orbital context.

Launch Success Yearly Trend



- Analysis has revealed significant trends related to success rates:
- **Upward Success Rate Trend:** Notably, there has been a consistent upward trend in success rates from 2013 to 2020. This extended period of improvement underscores the advancement in operational efficiency and mission success over time.
- **Challenges in the Early Years:** In contrast, the period spanning from 2010 to 2013 exhibited no recorded success rates, indicating initial challenges and difficulties faced by the program during its inception.

All Launch Site Names

- The names of the sites were obtained from the database utilizing the "Distinct" keyword in the query.

```
%sql select distinct launch_site from spacex
✓ 0.8s

 * ibm_db_sa://hnr90643:***@21fecfd8-47b7-4937-840d-d79
Done.
```

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

Launch Site Names Begin with 'CCA'

```
%sql select * from spacex where launch_site like 'CCA%' limit 5
✓ 0.4s                                                                    Python

 * ibm_db_sa://hnr90643:***@21fecfd8-47b7-4937-840d-d791d0218660.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31864/bludb
Done.
```

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

- Employed the "LIKE" operator in conjunction with the "LIMIT" keyword to present a restricted selection of records. Specifically, we showcased the names of sites commencing with "CCA," providing a concise display of the top 5 matching records.

Total Payload Mass

```
%sql select sum(payload_mass_kg_) as SUM from spacex where customer like 'NASA (CRS)'
✓ 0.5s

* ibm_db_sa://hnr90643:***@21fecfd8-47b7-4937-840d-d791d0218660.bs2io90l08kqb1od8lcg.databa
Done.

SUM
45596
```

- Employed the "LIKE" operator, as well as the "=" operator, to obtain the cumulative payload mass carried by boosters involved in NASA's Commercial Resupply Services (CRS) missions.

Average Payload Mass by F9 v1.1

```
%sql select avg(payload_mass_kg_) as AVG from spacex where booster_version = 'F9 v1.1'
✓ 0.6s

 * ibm_db_sa://hnr90643:***@21fecfd8-47b7-4937-840d-d791d0218660.bs2io90l08kqb1od8lcg.databases.a
Done.

AVG
2928
```

- Computed the average payload mass carried by booster version F9 v1.1 using the "=" operator and the "avg()" aggregate function, as depicted in the aforementioned process.

First Successful Ground Landing Date

```
%sql select min(date) as DATE from spacex where landing_outcome like '%ground pad%'
✓ 0.5s

* ibm_db_sa://hnr90643:***@21fecfd8-47b7-4937-840d-d791d0218660.bs2io90l08kqb1od8lcg.databases.a
Done.
```

DATE
2015-12-22

- Determined the date of the first successful landing outcome on a ground pad by employing the "min()" aggregate function in conjunction with the "LIKE" operator.

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%%sql
select booster_version as name from spacex
where landing_outcome like '%drone%' and payload_mass_kg_ > 4000 and payload_mass_kg_ < 6000
```

✓ 0.8s

```
 * ibm_db_sa://hnr90643:***@21fecfd8-47b7-4937-840d-d791d0218660.bs2io90l08kqb1od8lcg.databases.app
Done.
```

name
F9 FT B1020
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

- To isolate boosters that have achieved successful landings on drone ships, we utilized the "WHERE" clause to filter the relevant criteria and applied the "AND" condition to refine the query results.

Total Number of Successful and Failure Mission Outcomes

- To ascertain the total count of both successful and failed outcomes, we leveraged the "count()" aggregate function in conjunction with the "WHERE" clause to specify the relevant criteria and obtain the desired results.

```
%%sql
select count(*) as successful_launches from spacex where mission_outcome like '%Success%';
✓ 0.5s

* ibm_db_sa://hnr90643:***@21fecfd8-47b7-4937-840d-d791d0218660.bs2io90l08kqb1od8lcg.databases.a
Done.
```

successful_launches
100

```
%sql select count(*) as failed_launches from spacex where mission_outcome like '%Failure%'
✓ 0.5s

* ibm_db_sa://hnr90643:***@21fecfd8-47b7-4937-840d-d791d0218660.bs2io90l08kqb1od8lcg.databases.a
Done.
```

failed_launches
1

Boosters Carried Maximum Payload

- Executed a subquery to calculate the maximum payload and subsequently utilized the "=" operator to filter and refine the query results accordingly.

```
%%sql
select booster_version as MAX_PAYLOAD_BOOSTERS from
    spacex where payload_mass_kg_ =
    (select max(payload_mass_kg_) from spacex)
```

✓ 0.6s

```
* ibm_db_sa://hnr90643:***@21fecfd8-47b7-4937-840d-d791d0218
Done.
```

max_payload_boosters
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

```
%%sql
select booster_version, launch_site, landing_outcome from spacex where Year(date) = 2015 and landing_outcome like '%Failure%drone%'
```

✓ 0.5s

```
 * ibm_db_sa://hnr90643:***@21fecfd8-47b7-4937-840d-d791d0218660.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31864/bludb
Done.
```

booster_version	launch_site	landing_outcome
F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

- Identified the failed landing outcomes on drone ships in the year 2015 by applying the 'Year()' function in conjunction with the 'AND' condition and the 'LIKE' operator. This allowed us to retrieve relevant data, including the associated booster versions and launch site names.

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

- We executed a query to rank the count of landing outcomes, including Failure (drone ship) and Success (ground pad), within the date range from 2010-06-04 to 2017-03-20. The results were sorted in descending order to provide a comprehensive view of the outcomes.

```
%%sql
select landing_outcome, count(landing_outcome) as COUNT from spacex
    group by landing_outcome order by count(landing_outcome) desc
```

0.5s

```
* ibm_db_sa://hnr90643:***@21fecfd8-47b7-4937-840d-d791d0218660.bs2io90l08
Done.
```

landing_outcome	COUNT
Success	38
No attempt	22
Success (drone ship)	14
Success (ground pad)	9
Controlled (ocean)	5
Failure (drone ship)	5
Failure	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

Launch Sites

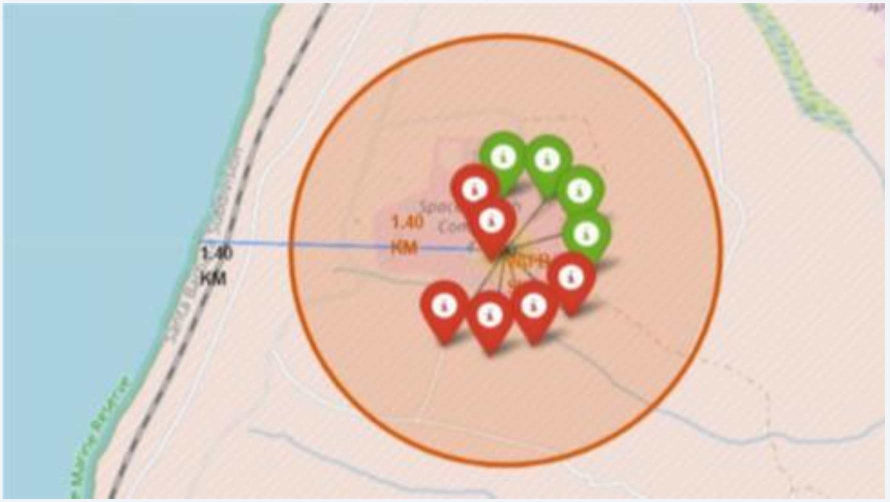
Analysis of the Falcon 9 launch sites map has yielded significant insights:

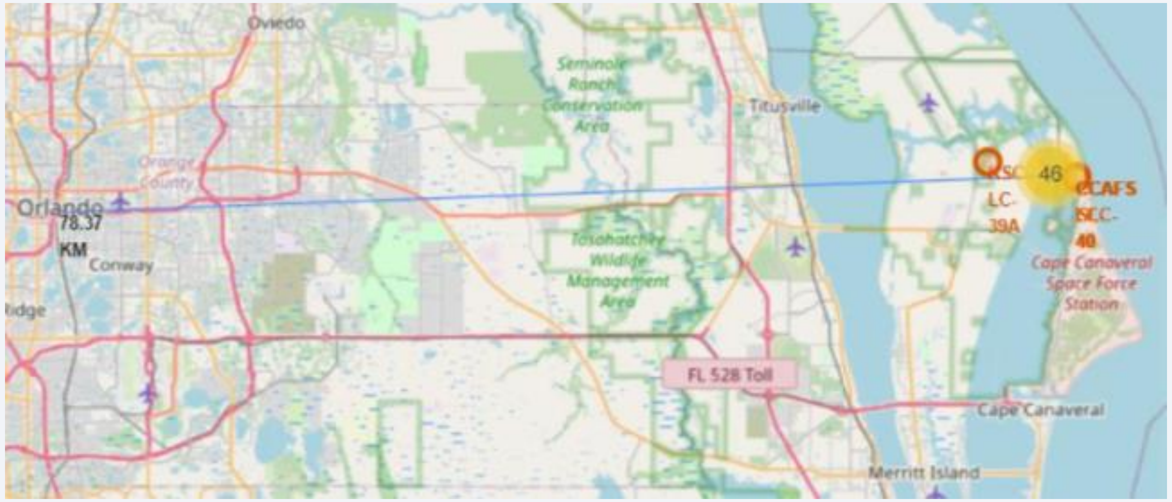
- **Proximity to Coastline:** Notably, the launch sites are strategically positioned in close proximity to coastlines. This is a deliberate choice to minimize risk, especially given the relatively higher failure rates (approximately 5-10%) associated with rocket launches. Placing launch sites near coastlines helps mitigate potential dangers to civilian areas.
- **East Coast Dominance:** Of the four launch sites, three are located on the east coast of the United States, while one is situated on the west coast. This distribution reflects the geographic diversity of launch facilities, catering to a range of mission requirements.
- **Equatorial Proximity:** All launch sites exhibit a common feature – their proximity to the equator. This strategic positioning is based on the advantage of requiring less fuel for rocket launches from the equator. It enhances fuel efficiency and minimizes operational costs, aligning with the considerations of cost-effective space exploration.

Number of Launches for Each Site

- **Green Markers:** These green markers symbolize successful launches, indicating launch sites with commendable success rates.
- **Red Markers:** Conversely, the red markers represent unsuccessful launches, signifying those sites with less favorable outcomes.
- Notably, the preponderance of green markers, denoting successful launches, underscores the higher success rates achieved by certain launch sites. Additionally, a predominant concentration of launches is observed on the east coast, emphasizing the significance of launch facilities in that region.

Distance of Sites from Nearby Proximities





Analysis involved measuring the distance of each launch site from the west coast and calculating the distance from Orlando. Several key observations emerged:

- **Proximity to Infrastructure:** Launch sites were strategically positioned in close proximity to railways, highways, and coastlines. This deliberate placement ensures efficient logistical support and transportation access for mission requirements.
- **Urban Buffer:** Notably, launch sites maintain a certain distance from populated urban areas. This separation underscores safety and risk mitigation considerations associated with rocket launches.
- These observations collectively reflect a thoughtful and strategic approach in the selection and positioning of launch sites, balancing the need for accessibility with safety and operational considerations.

Section 4

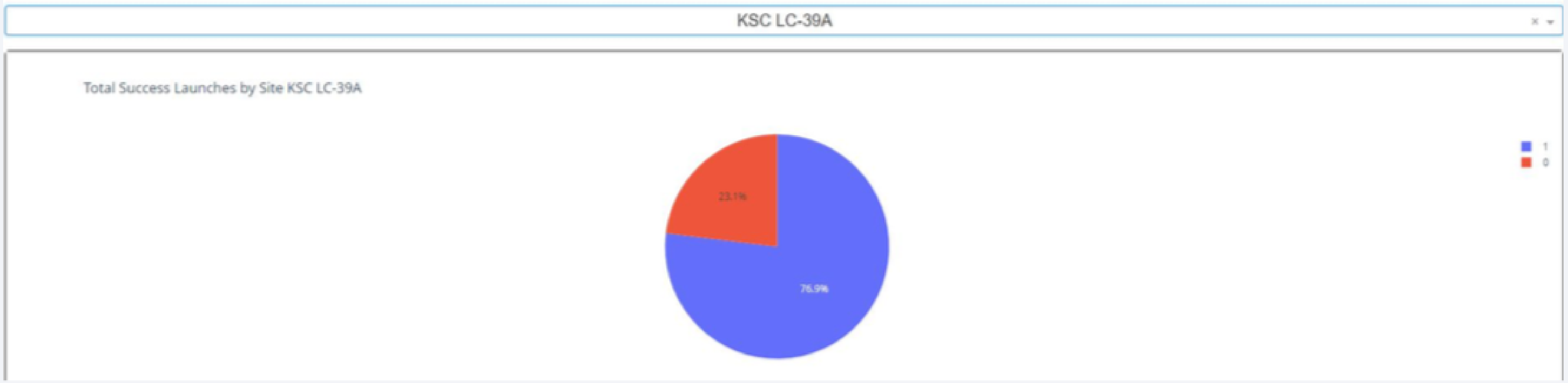
Build a Dashboard with Plotly Dash

Successful Launches From All Sites

Examination of the pie chart reveals the distribution of successful launches across all sites. Key observations include:

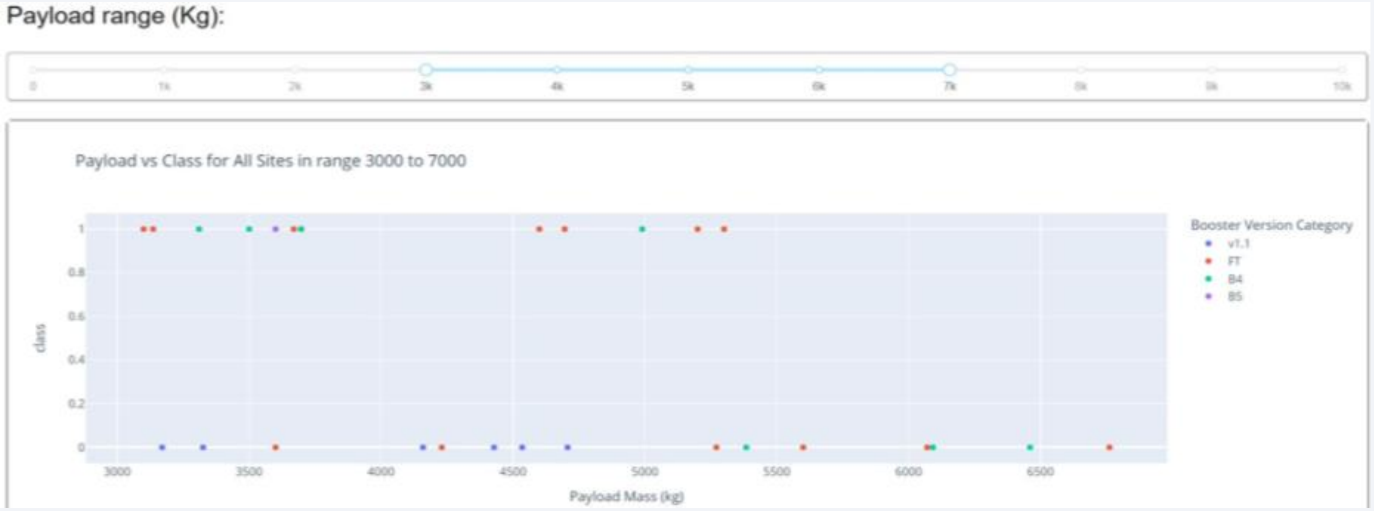
- **KSC LC-39A Dominance:** KSC LC-39A emerges as the launch site with the highest number of successful launches among all sites, indicating its consistent track record of success.
- **CCAFS SLC-40 Runner-Up:** CCAFS SLC-40 follows as the second-highest site in terms of successful launches, further emphasizing its effectiveness in achieving mission objectives.
- These findings underscore the notable performance of specific launch sites and their contribution to the overall success of Falcon 9 missions.

Launch Site with Highest Launch Success



- Analysis of the pie chart for KSC LC-39A launch site highlights the distribution of successful and failed launches. Key insights include:
- **Success Dominance:** KSC LC-39A boasts the highest number of successful launches among all sites, contributing to a significant success rate.
- **Success Percentage:** The pie chart underscores KSC LC-39A's impressive 76.9% success rate, reaffirming its role as a prominent and reliable launch site.
- These observations underscore the exceptional track record of KSC LC-39A in achieving successful mission outcomes.

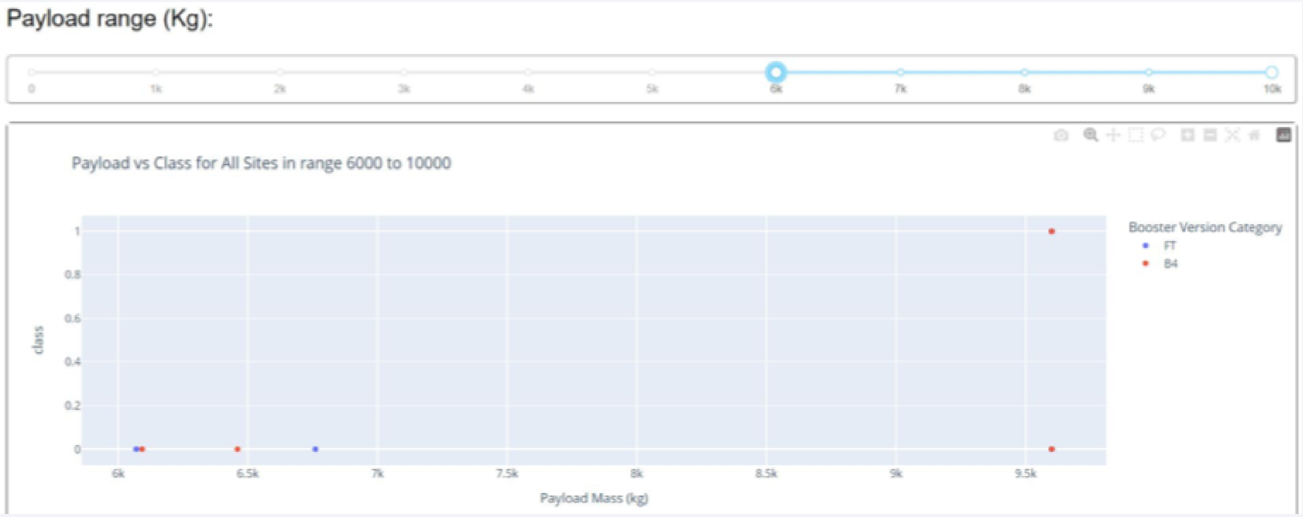
Payload vs Launch Outcome for All Sites



In the selected payload range of 3000 to 7000 kg, several crucial elements and findings are evident:

- **B4 Booster Success:** The B4 booster stands out with a notably high success rate within this payload range. This indicates its reliability in handling payloads falling within this specified weight category.
- **FT Booster Performance:** While the FT booster exhibits an average success rate, it is noteworthy that it experiences a relatively higher failure rate for high payloads within the specified range. This observation highlights a potential area for improvement or optimization in mission planning and execution.
- These insights provide valuable information about the performance of specific booster versions and their effectiveness in managing payloads within the defined range.

Payload vs Launch Outcome for All Sites



In the selected payload range of 6000 to 10000 kg, a notable observation is the presence of a high failure rate. This finding suggests that payloads falling within this specified weight category tend to pose increased challenges or complexities, leading to a comparatively lower success rate in achieving mission objectives. Further investigation and operational improvements may be warranted to address these challenges and enhance mission success within this payload range.

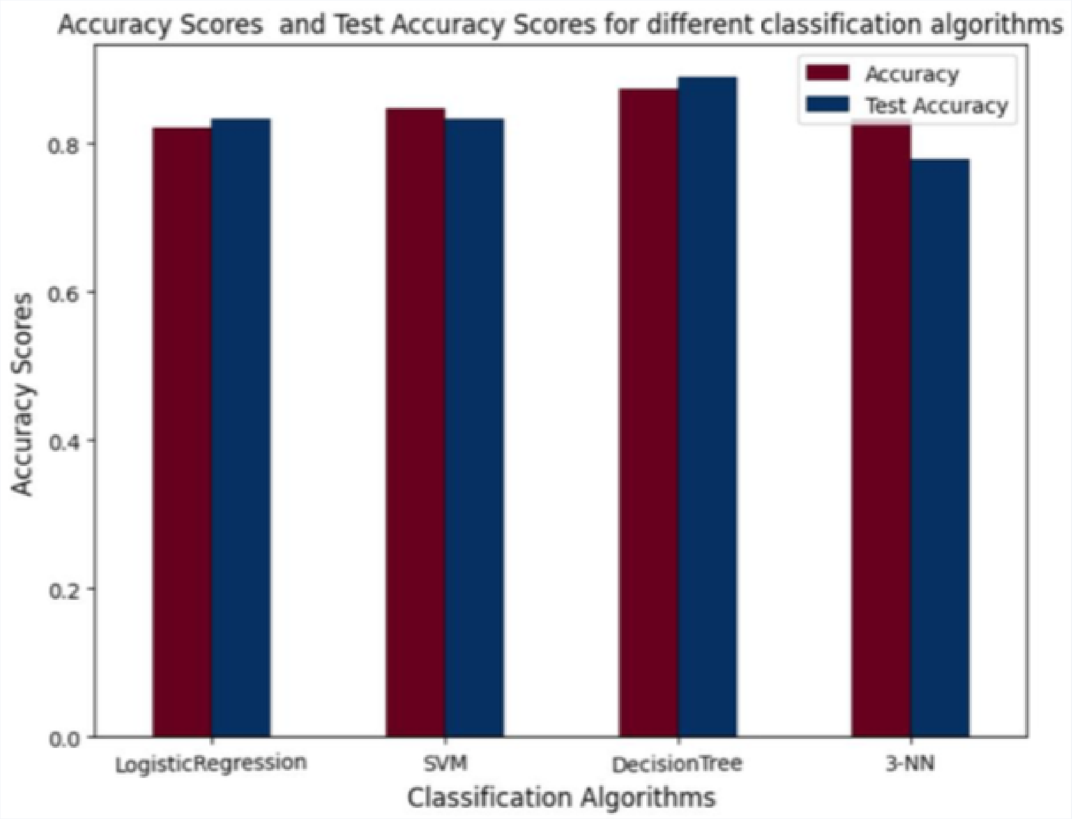
Section 5

Predictive Analysis (Classification)

Classification Accuracy

The bar chart depicting the accuracy of all classification models used in the project provides valuable insights:

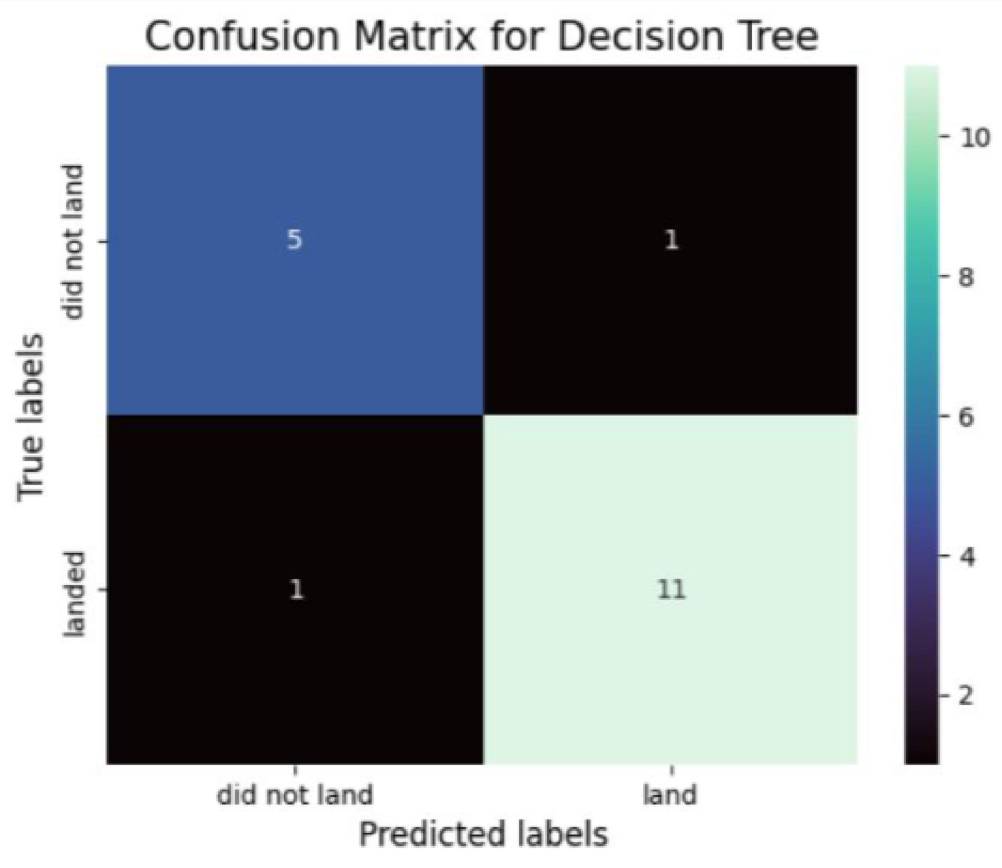
- **Decision Tree Classifier:** Notably, the Decision Tree Classifier emerges as the top-performing model with the highest accuracy score among all models. This signifies its effectiveness in making accurate predictions.
- **KNN with K=3:** Conversely, the bar chart indicates that the K-Nearest Neighbors (KNN) model with K=3 exhibits the lowest accuracy score among the models, suggesting room for improvement or alternative model selection.
- These visualized accuracy scores serve as a valuable reference for assessing the performance of each classification model, facilitating informed decision-making in model selection and refinement.



Confusion Matrix

An examination of the confusion matrix reveals a favorable distribution of true positive and true negative classifications compared to false classifications. Specifically, for the Decision Tree model:

- **True Positives and True Negatives:** The model correctly classified a substantial number of test points as true positives and true negatives, indicating its ability to accurately identify successful and unsuccessful outcomes.
- **False Classifications:** The matrix also shows a relatively low number of false classifications, reflecting the model's proficiency in minimizing misclassifications.
- These results highlight the Decision Tree model's effectiveness in correctly classifying the majority of test data points, with only a limited number of misclassifications.



Conclusions

- The project has yielded several key conclusions and insights:
- **Optimal Launch Site:** KSC LC-39A has emerged as the preferred launch site, demonstrating consistent success in achieving mission objectives.
- **Payload Success:** Payloads exceeding 8000 kg consistently exhibit a high success rate, underlining the robust capabilities of the launch system for handling heavier payloads.
- **VLEO Orbit Suitability:** Very Low Earth Orbit (VLEO) proves to be an excellent choice for launches, showcasing a high success rate across a significant number of missions.
- **New Launch Success:** New launches exhibit a relatively low failure rate, indicating efficient planning and execution of initial missions.
- **Heavy Payload Success:** Payloads with heavy masses experience higher success rates in Polar Orbit (PO), Low Earth Orbit (LEO), and International Space Station (ISS) missions.
- **Strategic Site Placement:** Launch sites are strategically situated near the equator and in proximity to coastlines, railways, and highways, optimizing operational efficiency and logistical support.
- **Preferred Model:** The Decision Tree Classifier emerges as the best-performing model for predicting the success or failure of upcoming launches, making it a valuable tool for mission planning and risk assessment.
- These conclusions provide valuable insights for decision-making in future space missions and underscore the importance of data-driven analysis in space exploration endeavors.

Thank you!