IBM Developer
SKILLS NETWORK

SpaceX Falcon 9
a data-driven journey to space

9th December 2024

TAMARA ROMERO HERNANDEZ

OUTLINE

EXECUTIVE SUMMARY

This project aims to analyze and predict the outcome of **SpaceX Falcon 9**'s first-stage landings based on historical launch data. The methodology includes data collection, data wrangling, exploratory data analysis (EDA), interactive visual analytics, and predictive analysis using classification models.

DATA COLLECTION

- The launch data was collected through the **SpaceX REST API**, offering detailed insights into launches, rockets, payloads, and outcomes, supplemented by **web scraping** from Wikipedia.

DATA WRANGLING

- The dataset was cleaned by filtering out irrelevant observations, handling missing values, and transforming categorical data into a suitable numerical format.

EXPLORATORY DATA ANALYSIS (EDA) USING VISUALIZATION AND SQL

- **Visualization**: Various plots, including pie charts and scatter plots, were created to explore trends in launch outcomes and the relationship between payload mass, orbit type, launch site, booster categories and mission success.
- **SQL**: SQL queries were employed to extract key insights, such as the total payload mass carried by boosters, the unique launch sites, and the first successful landing date, contributing to a deeper understanding of the data.

EXECUTIVE SUMMARY

INTERACTIVE VISUAL ANALYTICS USING FOLIUM AND PLOTLY DASH

- **Folium**: Various maps visualized the launch sites with markers indicating successful or failed outcomes. Distance calculations between launch sites and other key locations were also integrated.
- **Plotly Dash**: An interactive dashboard was built to allow real-time visual analytics, including pie charts for launch success rates and scatter plots for observing correlations between payload mass, launch sites, booster categories and mission outcomes.

PREDICTIVE ANALYSIS USING CLASSIFICATION MODELS

- Four classification models (**Logistic Regression**, **KNN**, **SVM**, and **Decision Tree**) were trained using the best parameters identified through GridSearchCV.
- Data normalization was applied for model training to ensure consistency in the input features.
- All models achieved an accuracy score of 0.83, with **KNN performing the best** in terms of Log Loss (0.366), which reflects the model's confidence in its predictions.
- The confusion matrix for KNN revealed a higher number of true positives (successful outcomes) compared to false positives, confirming its predictive reliability.

INTRODUCTION

This project examines **SpaceX's Falcon 9 rocket**, focusing on its first stage, which can be recovered and reused. This is a critical factor in reducing launch costs. By analyzing historical data, the goal is to predict whether the first stage will successfully land, which is essential for estimating the cost of a launch.

Understanding the likelihood of a successful landing not only aids cost estimation but also provides valuable insights for other companies competing against SpaceX in the space launch industry.

The analysis addresses the following questions:

- **What factors influence the successful landing of Falcon 9's first stage?**
- **Can we accurately predict the outcome of a first-stage landing based on launch data?**

This study aims to shed light on Falcon 9's operational efficiency and its implications for the space industry.

FALCON 9 OVERVIEW

HEIGHT
70m / 229.6ft

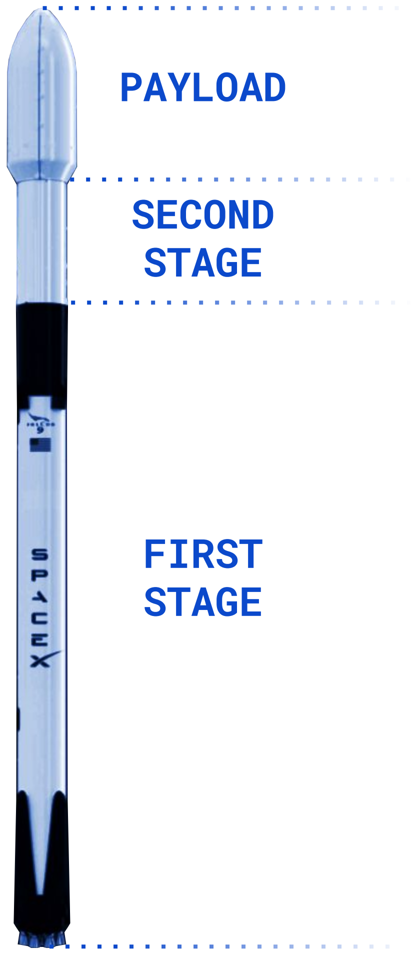
DIAMETER
3.7m / 12ft

PRICE
$69.75 M

LAUNCHES
405

LANDINGS
361

REFLIGHTS
335




METHODOLOGY

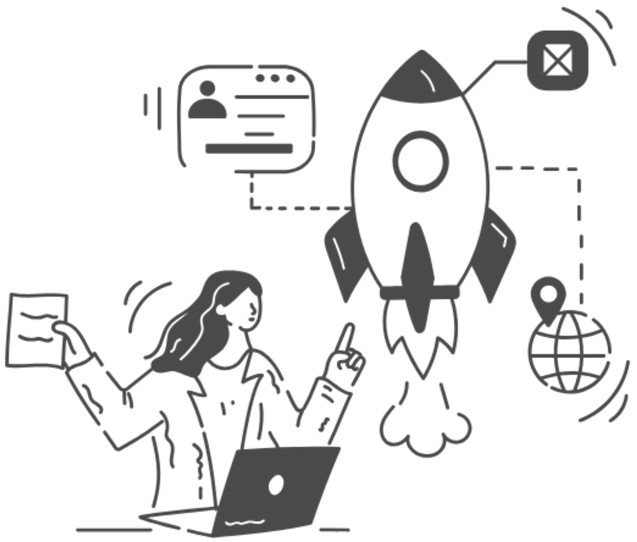
METHODOLOGY

DATA COLLECTION

The data was compiled using two methods: through the SpaceX REST API and employing web scraping techniques.

The **SpaceX REST API** provided detailed and structured information about launches, including the rocket used, payload details, launch specifications, and landing outcomes. Using the API ensured the data was reliable, consistent, and directly sourced from SpaceX's systems.

Web scraping was used to gather additional historical data. HTML tables from relevant Wikipedia pages were scraped to extract valuable Falcon 9 launch records.

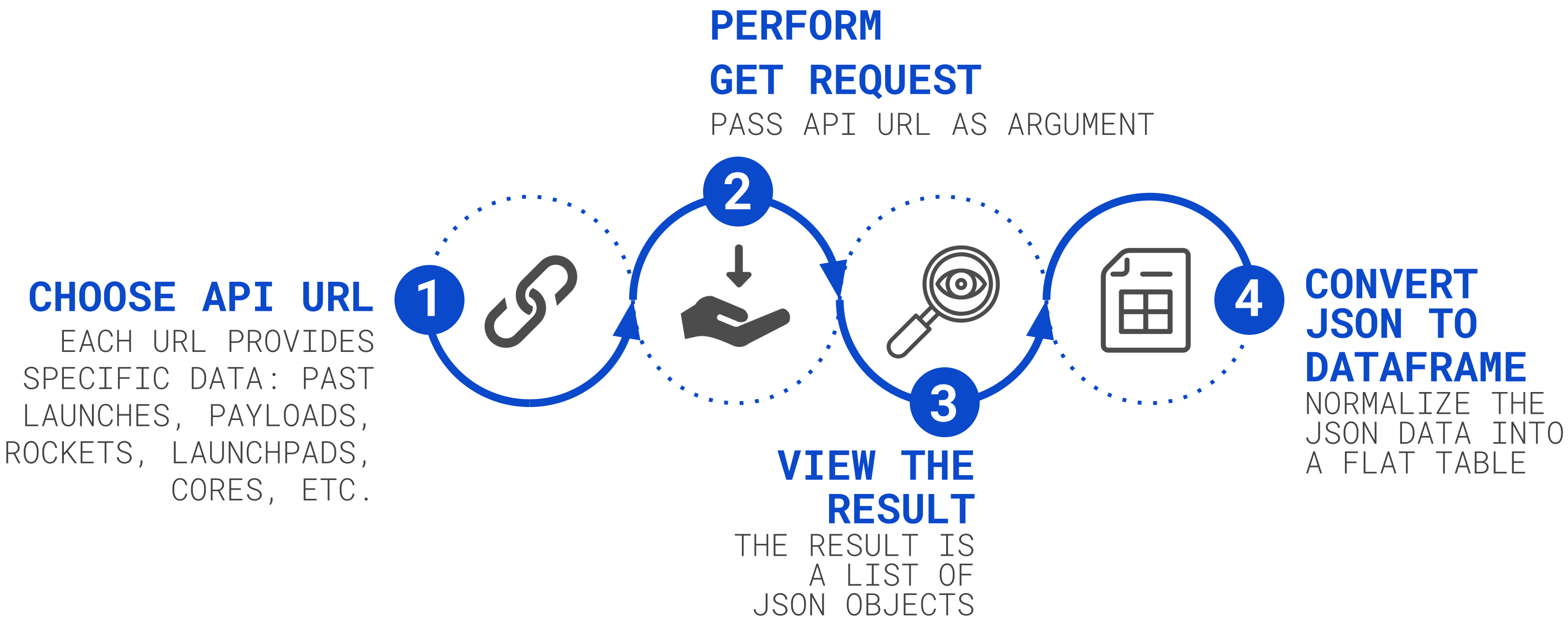


METHODOLOGY

DATA COLL. – SPACEX API

GitHub URL
Click Here

The flowchart represents the steps taken to gather data from the **SpaceX REST API**. This process ensured the retrieval of structured and reliable information about Falcon 9 launches.

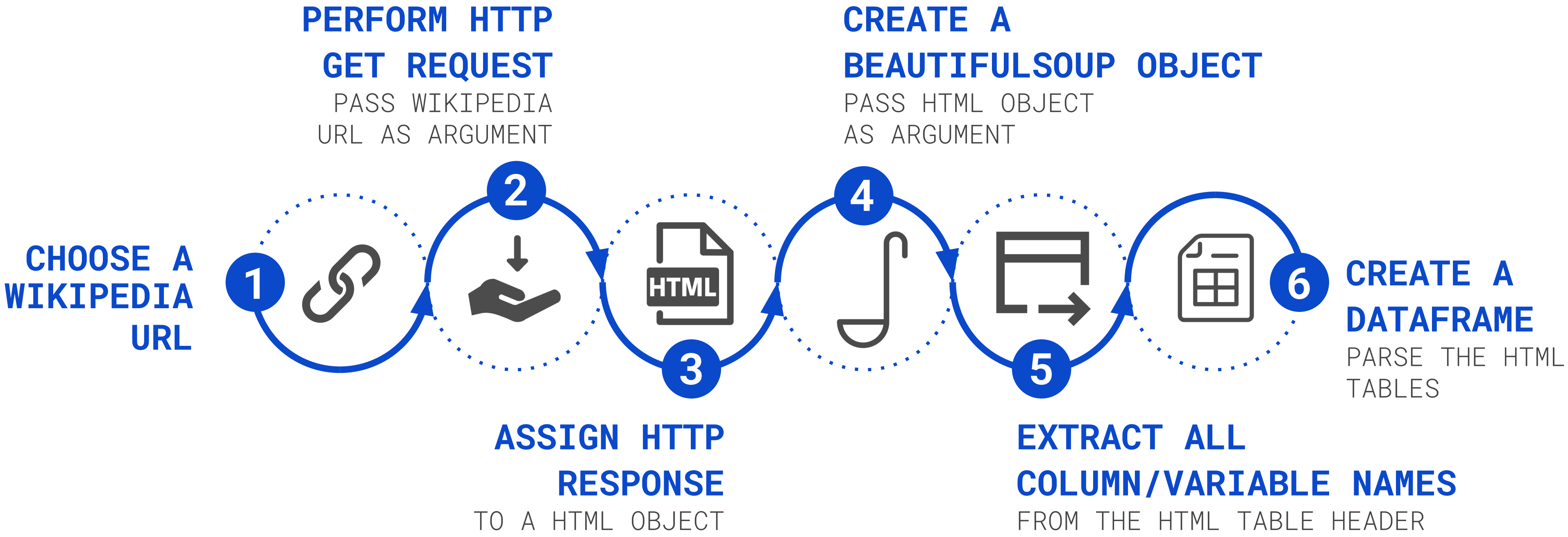


METHODOLOGY

DATA COLL. – WEB SCRAPING

GitHub URL
Click Here

The flowchart represents the steps taken to gather additional launch data through web scraping. This process involved using Python's BeautifulSoup library to extract valuable historical records from Wikipedia.

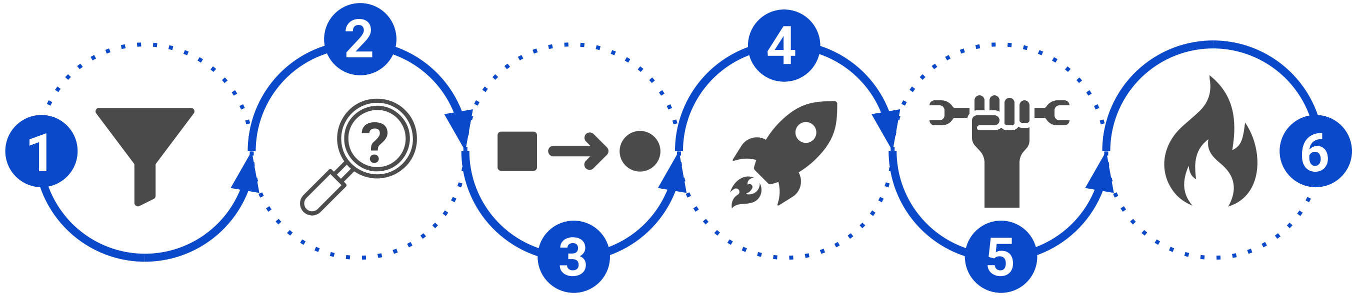


METHODOLOGY

DATA WRANGLING

GitHub URL
Click Here

The data was preprocessed to ensure quality and consistency, including handling missing values, feature selection, and transformation. Key features relevant to predicting first-stage landing success were identified and prepared for analysis.



DATA FILTERING

DEALING WITH MISSING VALUES
DATA IMPUTATION

DATA TYPE CONVERSION
DATES TO DATETIME OBJECTS, NUMERIC VALUES TO FLOATS

DEFINE DEPENDENT VARIABLE
'CLASS' COLUMN CREATED TO INDICATE LANDING OUTCOMES IN BINARY

FEATURE ENGINEERING
SELECTION OF INDEPENDENT VARIABLES

ONE-HOT ENCODING
CONVERT CATEGORICAL VARIABLES INTO NUMERICAL

METHODOLOGY

EDA – DATA VISUALIZATION

GitHub URL
Click Here

During Exploratory Data Analysis (EDA), various charts were plotted to uncover patterns in the data:

- **Flight Number vs. Launch Site**: A scatter plot to analyze launch attempts and outcomes at different launch sites.
- **Payload vs. Launch Site**: A scatter plot examining how payload mass influences launch outcomes across launch sites.
- **Success Rate vs. Orbit Type**: A bar chart exploring relationships between launch success rates and orbit types.
- **Flight Number vs. Orbit Type**: A scatter plot analyzing launch attempts, outcomes, and target orbit types.
- **Payload vs. Orbit Type**: A scatter plot investigating how payload mass and orbit types affect launch outcomes.
- **Launch Success Yearly Trend**: A line plot showing trends in launch success from 2010 to 2020.

METHODOLOGY

EDA – SQL

GitHub URL
Click Here

During the Exploratory Data Analysis (EDA), the following SQL queries were performed in the Jupyter Notebook:

- All Launch Site Names
- Launch Site Names Beginning with 'CCA'
- Total Payload Mass
- Average Payload Mass by F9 v1.1
- First Successful Ground Landing Date
- Successful Drone Ship Landing with Payload between 4000 and 6000
- Total Number of Successful and Failure Mission Outcomes
- Booster Versions that Carried Maximum Payload
- 2015 Launch Records
- Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

METHODOLOGY
INTERACTIVE MAP – FOLIUM

GitHub URL
Click Here

To analyze factors influencing the selection of optimal launch site locations, an interactive Folium map was created with various objects to represent key data:

- **Circles with Text Labels**: Each launch site was marked with a circle and labeled to provide clear identification of the site locations.
- **Marker Clusters:** Success and failure outcomes of launches at each site were visualized using marker clusters, allowing patterns to be observed while maintaining map clarity.
- **Icons**: Markers with icons were used to differentiate launch outcomes, with green icons for successful outcomes and red icons for unsuccessful ones.
- **Distance Calculations**: Distances from launch sites to nearby proximities were calculated and displayed with a line to assess geographic relationships and potential logistical considerations.

These visualizations provide an interactive way to explore existing launch site data and uncover factors relevant to optimal site selection.

METHODOLOGY

DASHBOARD – PLOTLY DASH

GitHub URL
Click Here

A Plotly Dash application was developed to enable real-time interactive visual analytics on SpaceX launch data.

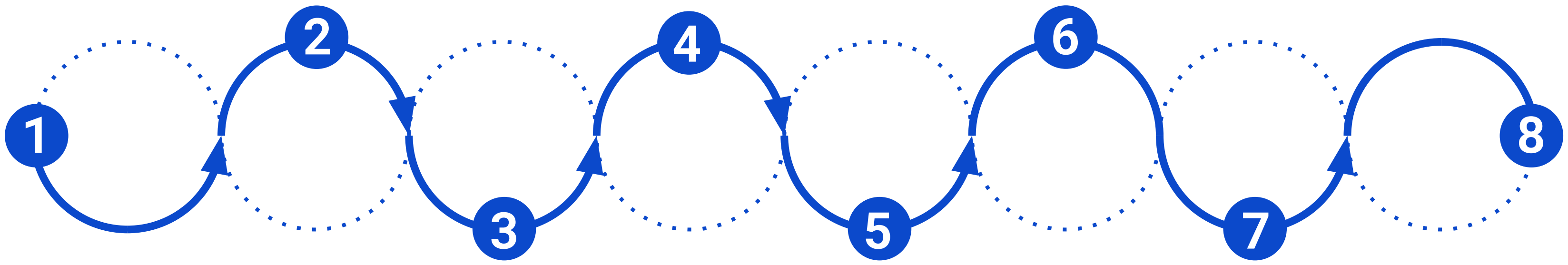
- **Pie Charts**: The dashboard includes pie charts that display launch success counts for individual launch sites as well as aggregated success counts across all sites. These charts help users quickly understand the performance of each site and overall success rates.
- **Scatter Plot**: A scatter plot was added to explore the relationship between payload mass and mission outcomes for selected sites. It also incorporates booster version categories to provide additional context. This plot helps identify potential correlations between payload and launch success.
- **Interactions**: The dashboard features a dropdown list for selecting specific launch sites and a range slider for filtering payload mass. These interactive components allow users to customize the visualizations and focus on specific subsets of the data, enhancing their ability to derive insights.

METHODOLOGY

PREDICTIVE ANALYSIS

GitHub URL
Click Here

The dataset was split into training and test sets, with normalization applied to the training data. **Classification models** (Logistic Regression, KNN, SVM, Decision Tree) were trained using optimal hyperparameters identified through GridSearchCV and evaluated on the test set to predict first-stage landing success.



1. **IDENTIFY INDEPENDENT & DEPENDENT VARIABLES**
   X AND Y RESPECTIVELY
2. **STANDARDIZE DATA IN X**
   USING .STANDARDSCALER()
3. **SPLIT DATA INTO TEST AND TRAINING SET**
4. **CREATE THE MODEL OBJECT**
   LOGISTIC REGRESSION / KNN / SVM / DECISION TREE
5. **CREATE THE GRIDSEARCHCV OBJECT**
6. **FIT THE GRIDSEARCHCV OBJECT TO THE TRAINING DATA**
   PASS THE MODEL OBJECT AS AN ESTIMATOR
7. **TRAIN A NEW MODEL**
   WITH THE BEST PARAMETERS IDENTIFIED THROUGH GRIDSEARCHCV
8. **PREDICT ON THE TEST SET DATA**

RESULTS

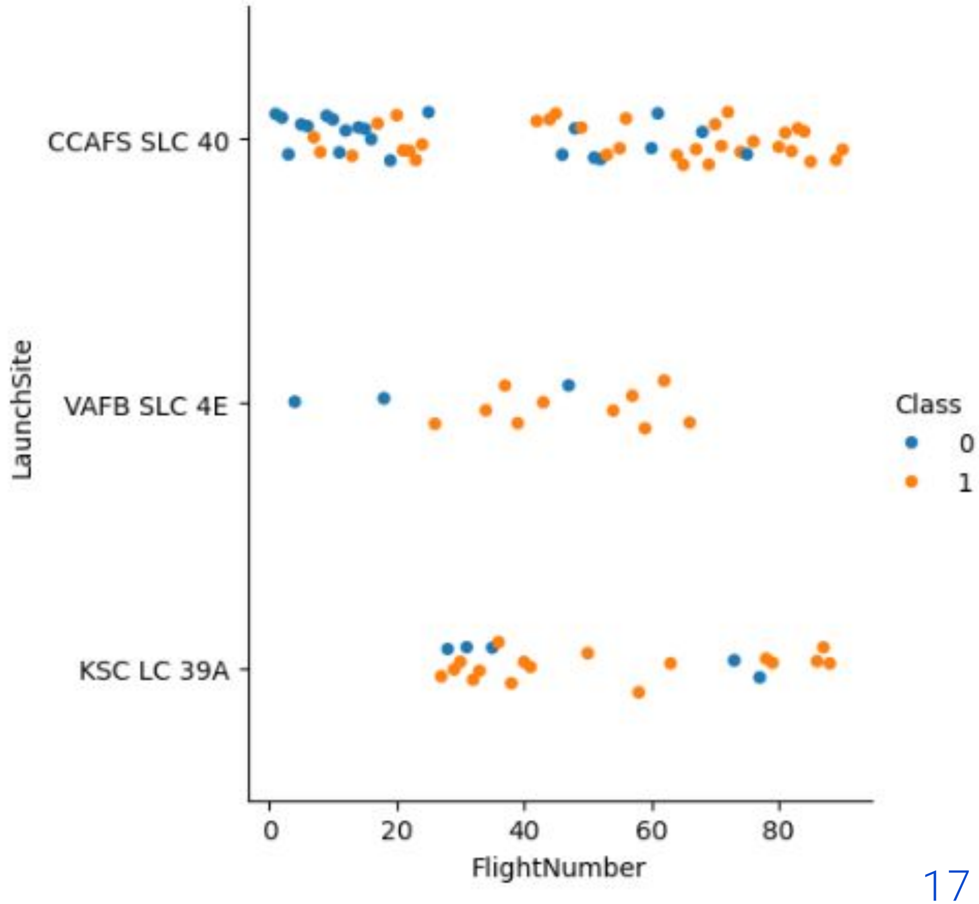
RESULTS

EDA – DATA VISUALIZATION

FLIGHT NUMBER vs. LAUNCH SITE

Most launch attempts and most unsuccessful outcomes (class 0) occurred at **CCAFS-SLC-40** launch site, while **VAFB-SLC-4E** launch site had the **fewest attempts** and least unsuccessful outcomes.

A potential explanation lies in the frequency of launches and operational challenges at each site. A higher number of launches at CCAFS-SLC-40 likely increased the probability of having more unsuccessful outcomes.



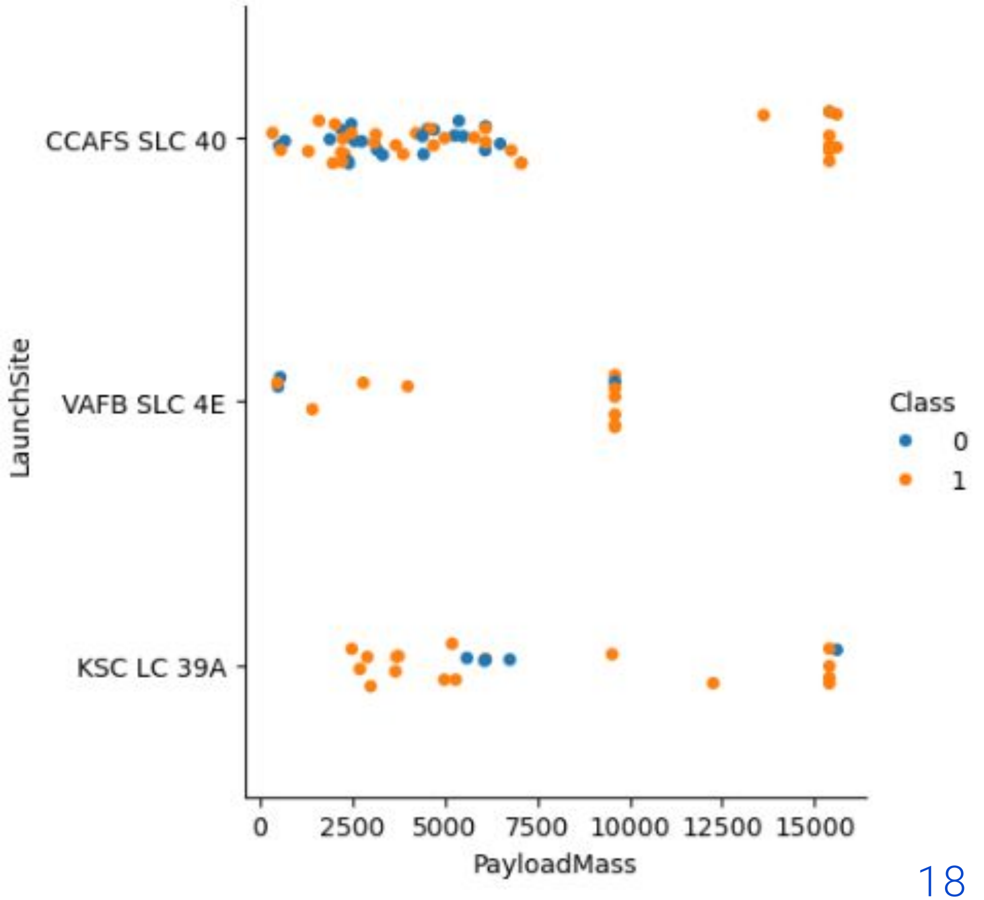
RESULTS

EDA – DATA VISUALIZATION

PAYLOAD vs. LAUNCH SITE

Most launch attempts across **all sites** involved rockets with payloads **below 10,000 kg**, where the majority of unsuccessful outcomes (class 0) also occurred. Additionally, **VAFB-SLC-4E** had **no launches** with heavy payloads **exceeding 10,000 kg**.

This observation suggests a potential focus on smaller payload missions, as rocket launches with payloads above 10,000 kg appear to be rare.



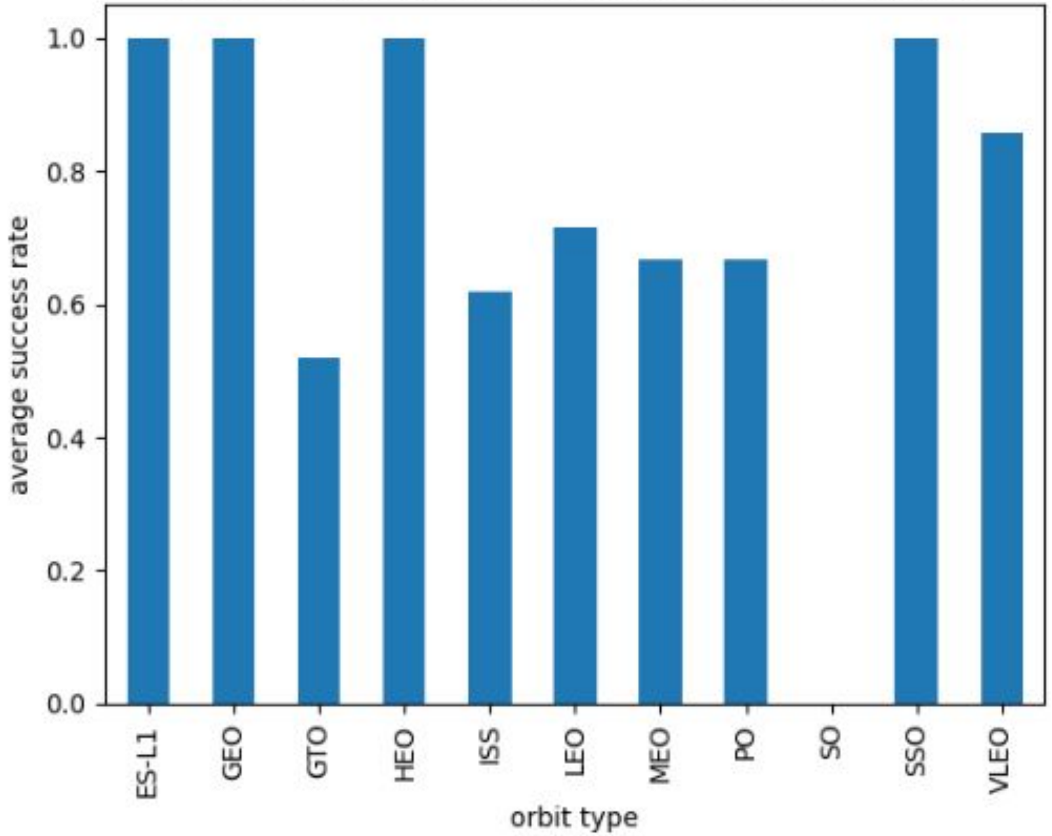
RESULTS

EDA – DATA VISUALIZATION

SUCCESS RATE vs. ORBIT TYPE

Orbits such as **ES-L1**, **GEO**, **HEO**, and **SSO** have **higher success rates**, while the **GTO** orbit stands out with the **lowest success rate**.

This suggests that certain orbit types may be more conducive to successful launches than others, with GTO experiencing more challenges and complexities.

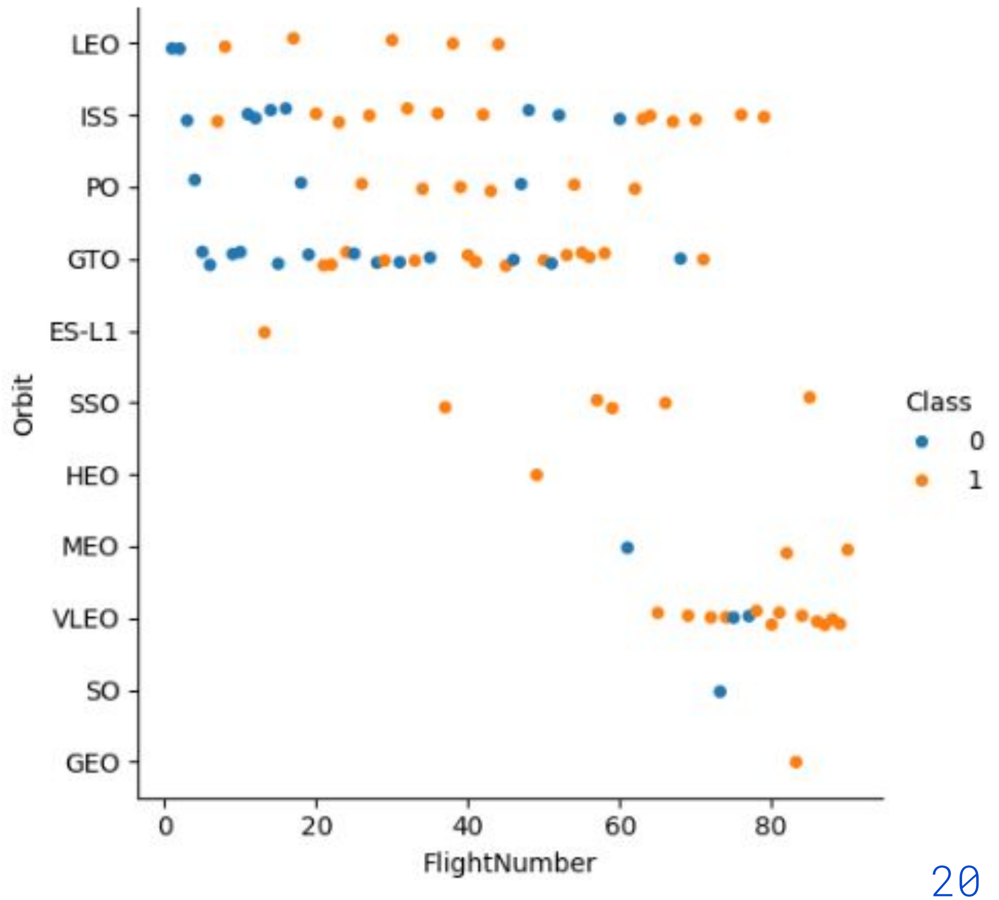


RESULTS

EDA – DATA VISUALIZATION

FLIGHT NUMBER vs. ORBIT TYPE

In the **LEO** orbit, the more flights that are conducted, the higher the success rate appears to be. In contrast, for the **GTO** orbit, there doesn't seem to be any clear connection between the number of flights and the likelihood of success. **Launch attempts** are **more frequent** in **GTO** and **ISS** orbits, with GTO having the most unsuccessful outcomes (class 0).

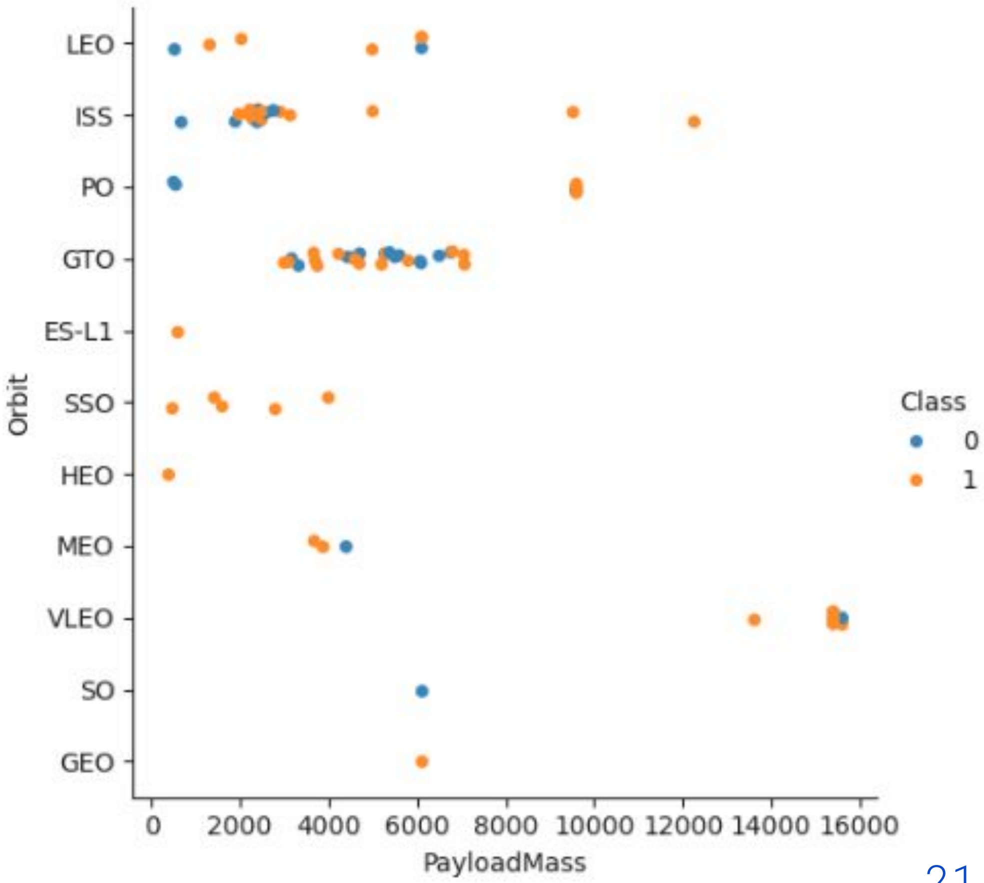


RESULTS

EDA – DATA VISUALIZATION

PAYLOAD vs. ORBIT TYPE

Heavy payloads (greater than 10,000 kg) tend to have a **higher success rate** for launches in **VLEO** and **ISS** orbits. However, in the **GTO** orbit, there appears to be no clear relationship between payload mass and successful launch outcomes (class 1).

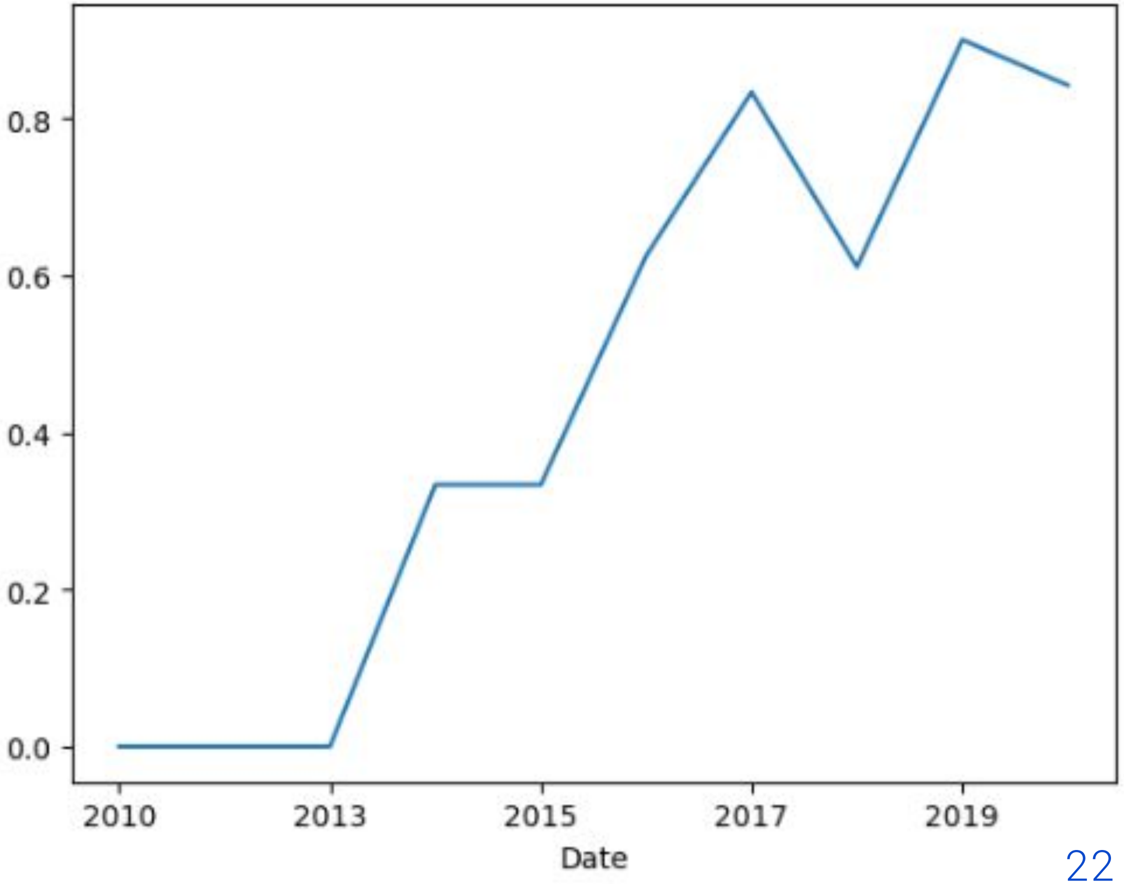


RESULTS

EDA – DATA VISUALIZATION

LAUNCH SUCCESS YEARLY TREND

The success rate of SpaceX launches has seen a significant upward trend since 2013, with a noticeable dip in 2018. The **first successful launch** occurred in **2014**, marking a turning point in SpaceX's development and operational capabilities.



RESULTS

EDA – SQL

LAUNCH SITE NAMES

```
%sql SELECT DISTINCT "Launch_Site" FROM SPACEXTABLE
```

```
 * sqlite:///my_data1.db
Done.
```

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

The SQL query identifies all unique launch sites from the dataset. By using the DISTINCT keyword, we ensure that only unique launch site names are selected.

RESULTS

EDA – SQL

LAUNCH SITE NAMES BEGINNING WITH ‘CCA’

```
%sql SELECT * FROM SPACEXTABLE WHERE "Launch_Site" like 'CCA%' LIMIT 5;
```

```
 * sqlite:///my_data1.db
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

The SQL query retrieves launch sites starting with 'CCA', showing 5 records. The observations returned are for launch site **CCAFS LC-40**.

RESULTS

EDA – SQL

TOTAL PAYLOAD MASS

```
%sql SELECT SUM("PAYLOAD_MASS__KG_") FROM SPACEXTABLE WHERE "Customer"='NASA (CRS)'
```

```
 * sqlite:///my_data1.db
Done.
```

SUM("PAYLOAD_MASS__KG_")
45596

The SQL query calculates the total payload mass carried by NASA boosters, which is **45,596 kg**.

RESULTS

EDA – SQL

AVERAGE PAYLOAD MASS BY F9 V1.1

```
%sql SELECT AVG("PAYLOAD_MASS__KG_") FROM SPACEXTABLE WHERE "Booster_Version"='F9 v1.1'

 * sqlite:///my_data1.db
Done.
```

AVG("PAYLOAD_MASS__KG_")
2928.4

The SQL query calculates the average payload mass carried by the F9 v1.1 booster version, which is **2,928.4 kg**.

RESULTS

EDA – SQL

FIRST SUCCESSFUL GROUND LANDING DATE

```
%sql SELECT MIN("Date") FROM SPACEXTABLE WHERE "Landing_Outcome"='Success (ground pad)'
```

```
 * sqlite:///my_data1.db
Done.
```

MIN("Date")
2015-12-22

The SQL query retrieves the date of the first successful landing outcome on ground pad, which occurred on **2015-12-22**.

RESULTS

EDA – SQL

SUCCESSFUL DRONE SHIP LANDING WITH PAYLOAD BETWEEN 4,000 AND 6,000 KG

```
%sql SELECT "Booster_Version" FROM SPACEXTABLE WHERE "PAYLOAD_MASS__KG_">=4000 AND "PAYLOAD_MASS__KG_"<=6000
 * sqlite:///my_data1.db
Done.
```

Booster_Version
F9 v1.1
F9 v1.1 B1011
F9 v1.1 B1014
F9 v1.1 B1016
F9 FT B1020
F9 FT B1022
F9 FT B1026
F9 FT B1030
F9 FT B1021.2
F9 FT B1032.1
F9 B4 B1040.1
F9 FT B1031.2
F9 B4 B1043.1
F9 FT B1032.2
F9 B4 B1040.2
F9 B5 B1046.2
F9 B5 B1047.2
F9 B5 B1046.3
F9 B5B1054
F9 B5 B1048.3
F9 B5 B1051.2
F9 B5B1060.1
F9 B5 B1058.2
F9 B5B1062.1

The SQL query retrieves a list of boosters that successfully landed on a drone ship with a payload mass between 4,000 kg and 6,000 kg.

RESULTS

EDA – SQL

TOTAL NUMBER OF SUCCESSFUL AND FAILURE MISSION OUTCOMES

```
%sql SELECT SUM(CASE WHEN "Landing_Outcome" LIKE 'Success%' THEN 1 ELSE 0 END) AS Success_Count, SUM(CASE WHEN "Landing_Outcome" LIKE 'Failure%' THEN 1 ELSE 0 END) AS Failure_Count FROM SPACEXTAB
```

```
 * sqlite:///my_data1.db
Done.
```

Success_Count	Failure_Count
61	10

The data was filtered to focus only on 'Success' and 'Failure' outcomes, allowing us to clearly assess the total number of successful and failed mission outcomes. The full list of landing outcomes in the dataset includes: Controlled (ocean), Failure, Failure (drone ship), Failure (parachute), No attempt, Preclude (drone ship), Success, Success (drone ship), Success (ground pad), and Uncontrolled (ocean).

RESULTS

EDA – SQL

BOOSTERS WHICH CARRIED MAXIMUM PAYLOAD

```
%sql SELECT "Booster_Version" FROM SPACEXTABLE WHERE "PAYLOAD_MASS__KG_" = (SELECT MAX("PAYLOAD_MASS__KG_") FROM SPACEXTABLE)
```

* sqlite:///my_data1.db
Done.

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

The SQL query retrieves a list of boosters that carried the maximum payload mass.

RESULTS

EDA – SQL

2015 LAUNCH RECORDS FOR FAILED DRONE SHIP OUTCOME

```
%sql SELECT "Date", "Landing_Outcome", "Booster_Version", "Launch_Site" FROM SPACEXTABLE WHERE "Landing_Outcome"='Failure (drone ship)' AND "Date" like '2015%'
```

```
 * sqlite:///my_data1.db
Done.
```

Date	Landing_Outcome	Booster_Version	Launch_Site
2015-01-10	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
2015-04-14	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

The SQL query displays the failed drone ship landing outcomes in 2015, along with the corresponding booster versions and launch site names.

RESULTS

EDA – SQL

RANK LANDING OUTCOMES BETWEEN 2010-06-04 AND 2017-03-20

```
%sql SELECT "Landing_Outcome", COUNT("Landing_Outcome") AS COUNT, DENSE_RANK() OVER (ORDER BY COUNT("Landing_Outcome") DESC) AS rank FROM SPACEXTABLE WHERE "Date" BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY "Landing_Outcome" ORDER
```

```
 * sqlite:///my_data1.db
Done.
```

Landing_Outcome	COUNT	rank
No attempt	10	1
Success (drone ship)	5	2
Failure (drone ship)	5	2
Success (ground pad)	3	3
Controlled (ocean)	3	3
Uncontrolled (ocean)	2	4
Failure (parachute)	2	4
Precluded (drone ship)	1	5

The SQL query presents the count of landing outcomes between 2010-06-04 and 2017-03-20, ranked in descending order.

RESULTS

INTERACTIVE MAP – FOLIUM

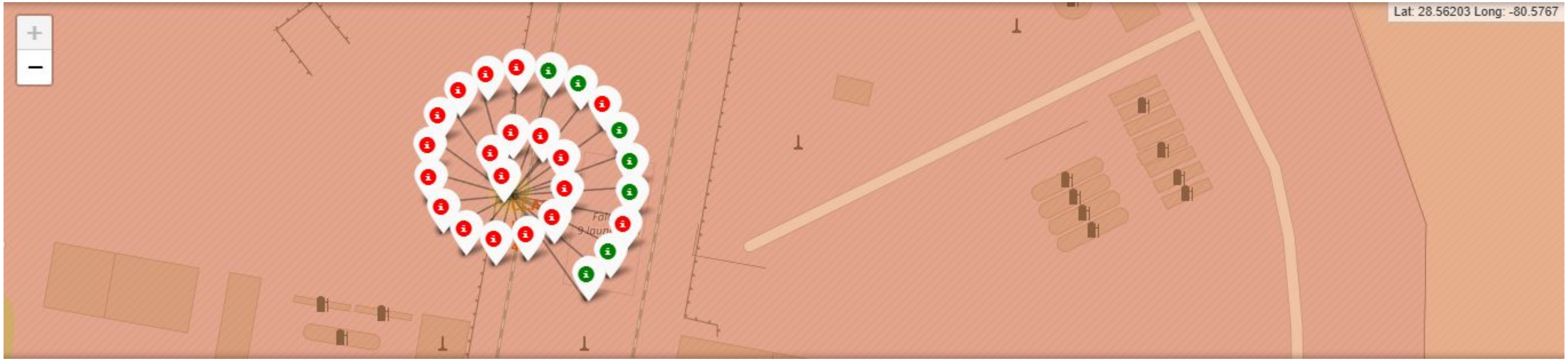
LAUNCH SITES LOCATION MARKERS ON GLOBAL MAP

The Folium map above displays **SpaceX Falcon 9 launch sites**, represented by yellow cluster markers. These markers help group the sites for better visualization and clarity when zoomed out. We can observe that all SpaceX Falcon 9 launch sites are located in the states of Florida and California, in the United States. This geographic distribution likely reflects strategic considerations such as proximity to the ocean for safety, access to equatorial launch trajectories, and existing aerospace infrastructure in these regions.

RESULTS

INTERACTIVE MAP – FOLIUM

COLOR-LABELED LAUNCH OUTCOMES

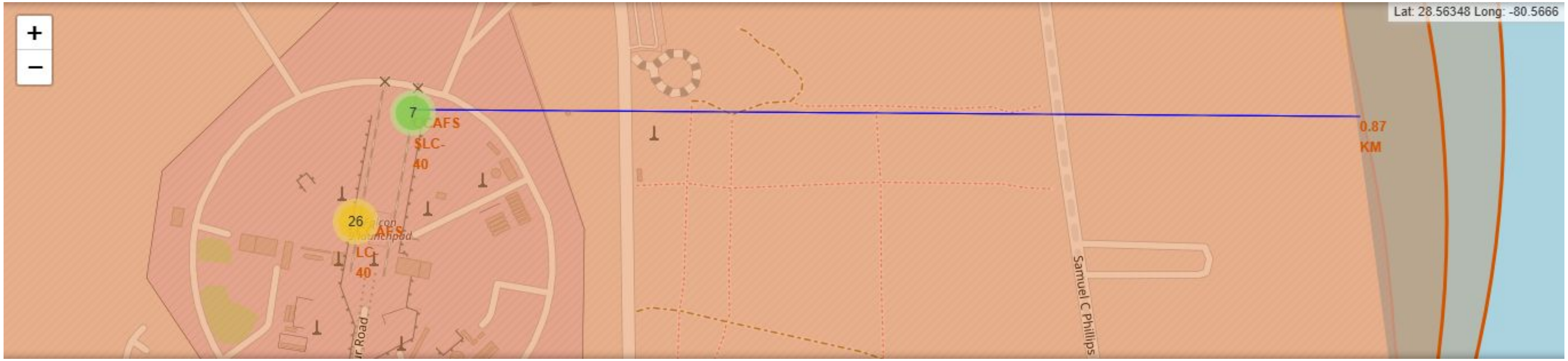


The Folium map above shows a zoomed in view of a launch site. We can observe colored icon markers used to differentiate launch outcomes: green for successful outcomes and red for unsuccessful ones. This visual distinction allows for a quick assessment of launch success patterns at the site, offering insights into its historical performance and reliability. Such information is valuable for understanding factors contributing to successful outcomes and for making data-driven decisions about future launches.

RESULTS

INTERACTIVE MAP – FOLIUM

DISTANCE BETWEEN LAUNCH SITE AND COASTLINE



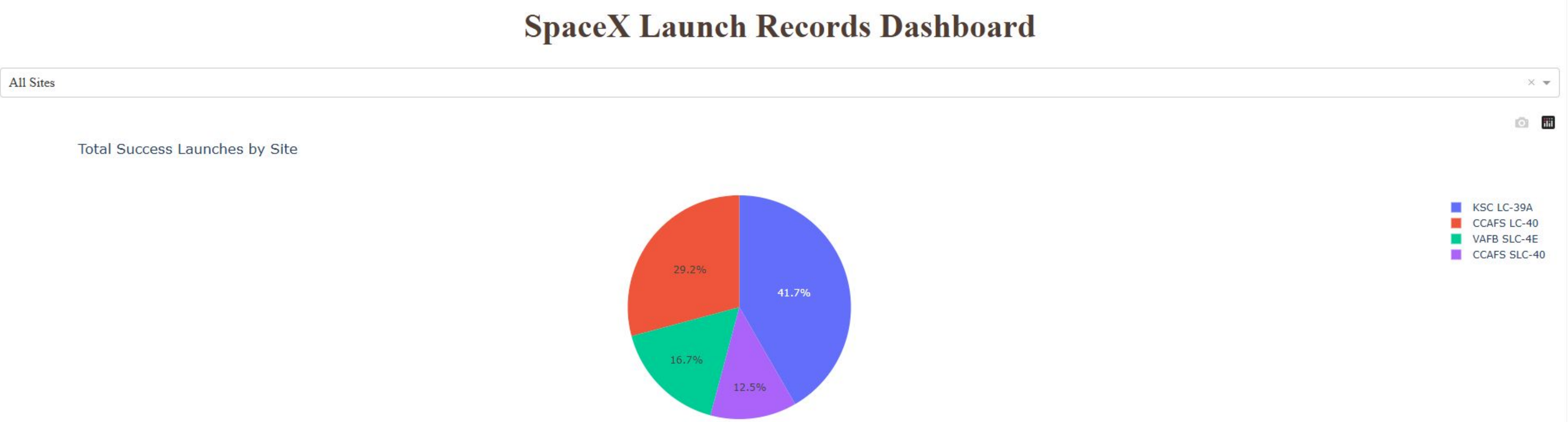
The Folium map above shows a zoomed in view of a launch site. A line marker indicates the distance between the launch site and the Florida coastline. Such distance has been calculated and it's equal to 0.87 Km.
This measurement is essential for assessing the feasibility of the launch site, as proximity to the coastline could impact logistics, safety considerations, and potential recovery operations. Such insights help evaluate the suitability of the launch site for efficient rocket operations.

RESULTS

DASHBOARD – PLOTLY DASH

SCAN ME

LAUNCH SUCCESS COUNT FOR ALL SITES



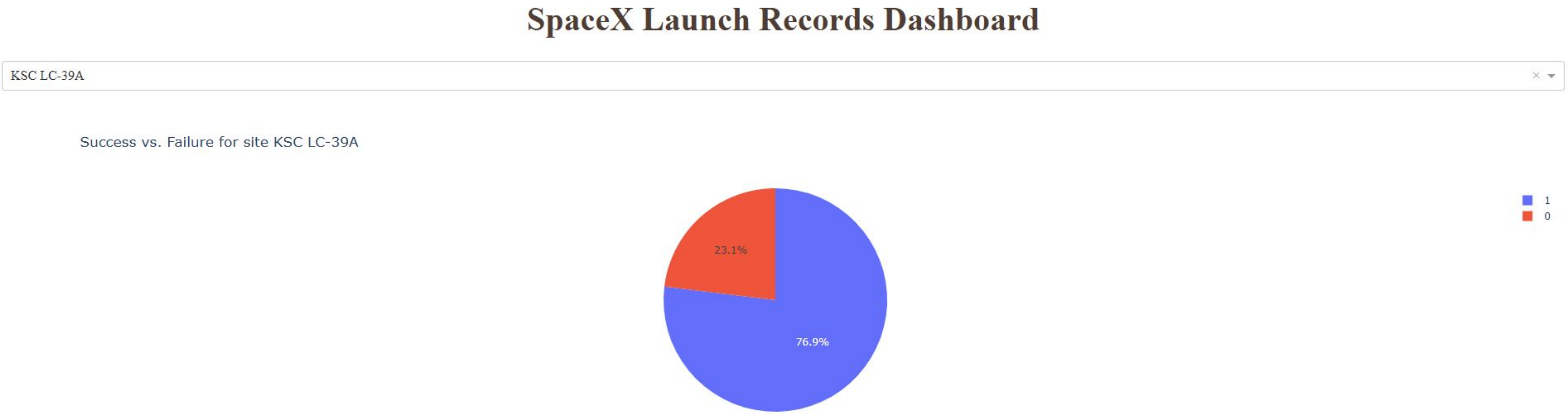
The pie chart illustrates the percentage distribution of launch successes across al SpaceX launch sites. We can observe that the site with **highest success rate** is **KSC LC-39A** and the site with **least success rate** is **CCAFS SLC-40**. These insights help assess the performance of each launch site and inform decisions about future operations.

RESULTS

DASHBOARD – PLOTLY DASH

SCAN ME

LAUNCH SITE WITH HIGHEST LAUNCH SUCCESS RATIO



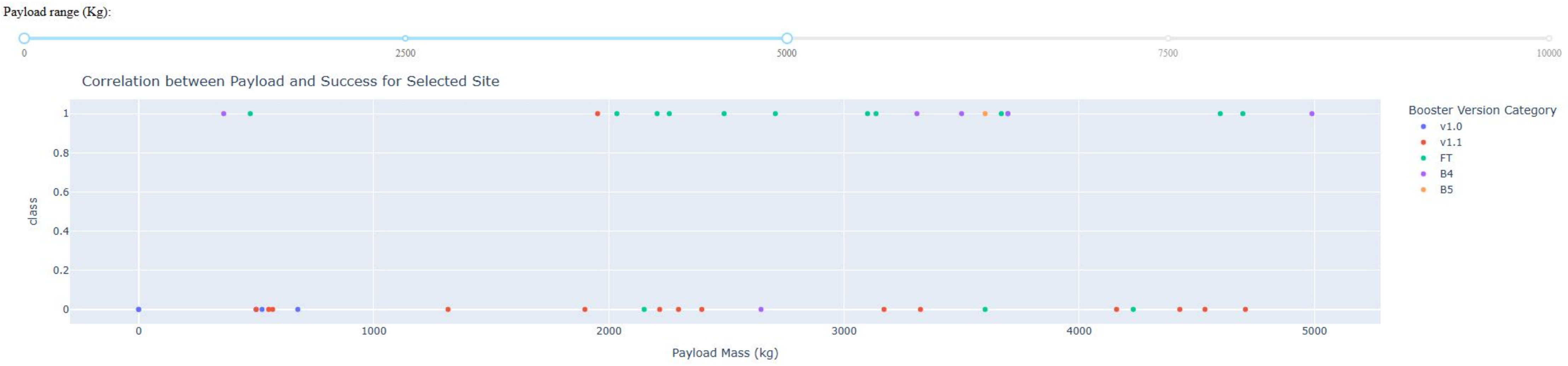
The pie chart above displays the proportion of successful and unsuccessful launches at **KSC LC-39A**, the site with the highest success rate. The chart reveals that **76.9% of launches** at this site have been **successful**, while **23.1% have been unsuccessful**. Understanding these proportions is crucial for evaluating site performance and identifying areas for improvement to minimize unsuccessful outcomes.

RESULTS

DASHBOARD – PLOTLY DASH

SCAN ME

PAYLOAD vs. LAUNCH OUTCOME FOR ALL SITES (PAYLOAD RANGE 0 KG - 5,000 KG)



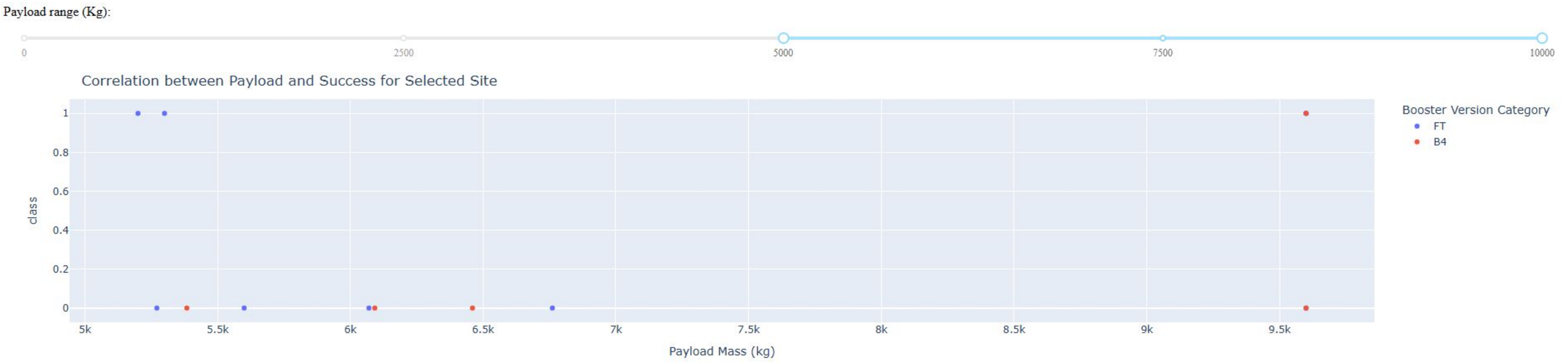
The scatter plot above visualizes the relationship between payload mass (ranging from 0 kg to 5,000 kg), launch outcomes, and booster version categories. This payload range has the highest number of successful outcomes. Additionally, the plot reveals that the **FT booster has the highest success** rate across all booster types, while the **v1.1 booster has the highest failure rate**. These findings are essential for understanding how payload capacity and booster technology contribute to launch success, providing valuable insights for optimizing future missions.

RESULTS

DASHBOARD – PLOTLY DASH

SCAN ME

PAYLOAD vs. LAUNCH OUTCOME FOR ALL SITES (PAYLOAD RANGE 5,000 KG - 10,000 KG)



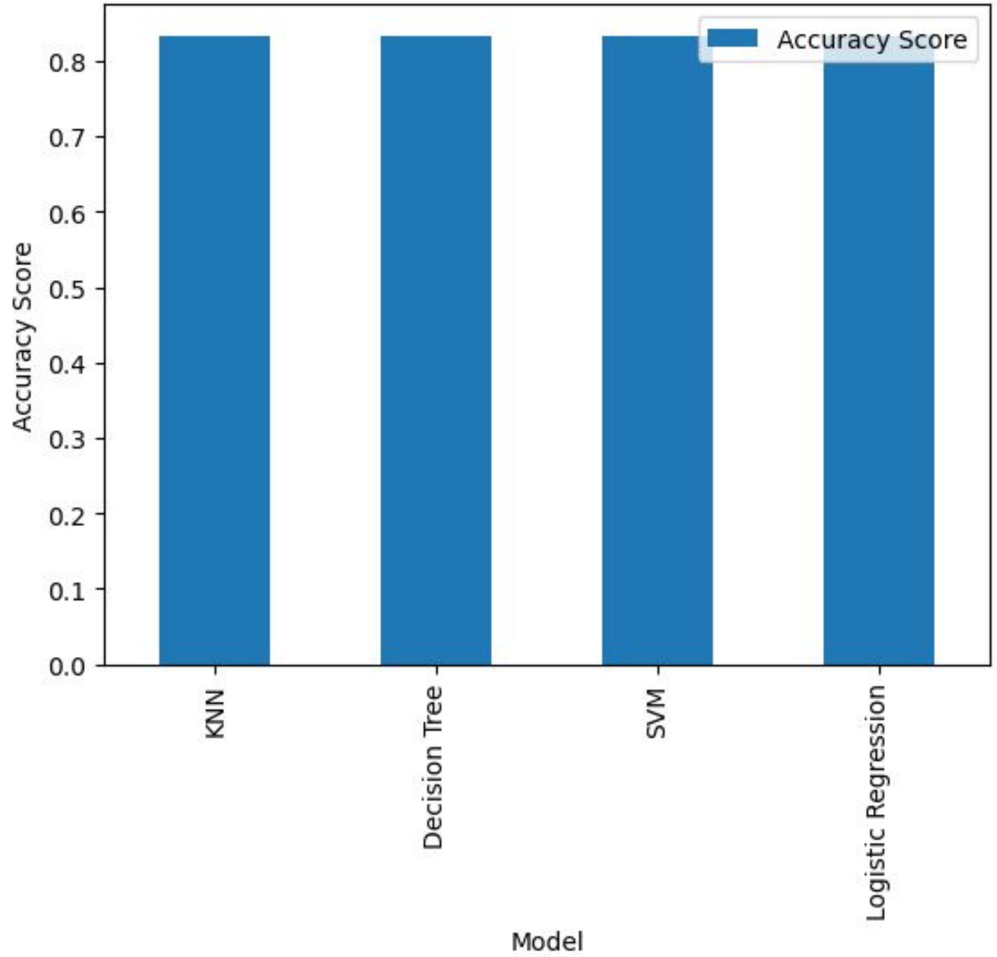
The scatter plot above visualizes the relationship between payload mass (ranging from 5,000 kg to 10,000 kg), launch outcomes, and booster version categories. This payload range has the least number of successful outcomes. Additionally, the plot highlights that the **FT booster version has the largest success** rate in comparison to B4 booster. These findings are essential for understanding how payload capacity and booster technology contribute to launch success, providing valuable insights for optimizing future missions.

RESULTS

PREDICTIVE ANALYSIS

CLASSIFICATION ACCURACY

All models – **KNN**, **Decision Tree**, **SVM**, and **Logistic Regression** – were trained using the optimal hyperparameters identified through **GridSearchCV**. After training, the accuracy scores for each model were calculated. Interestingly, all models achieved the **same accuracy score** of **0.83**. This indicates that, within the context of this project, these models performed similarly in predicting the launch outcomes, providing a robust foundation for further analysis or deployment.



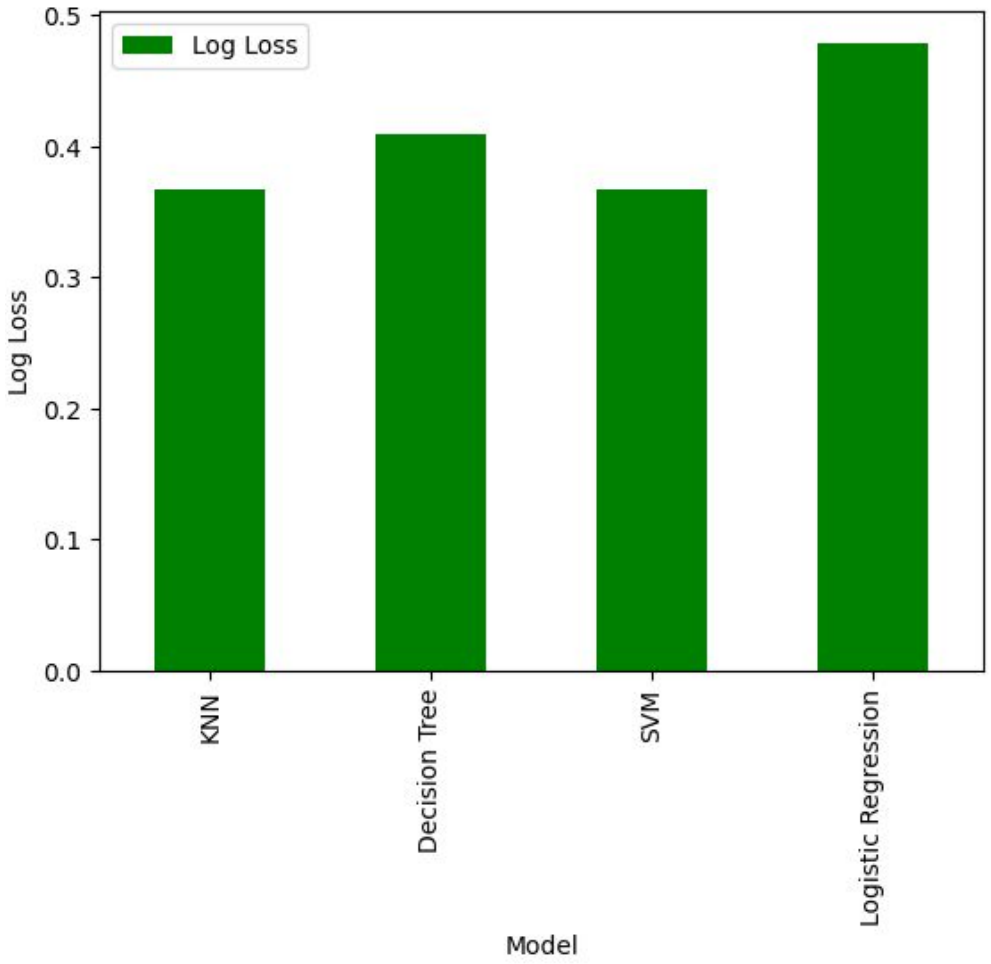
RESULTS

PREDICTIVE ANALYSIS

CLASSIFICATION LOG LOSS

Since the accuracy score across all models was identical, an alternative evaluation metric, **Log Loss**, was used to assess the confidence of the predictions. Log Loss measures how well the predicted probabilities align with the actual class labels. Lower Log Loss values indicate more accurate probability predictions.

After evaluating the models, the **KNN** model emerged as the **best performer**, with a **Log Loss score** of **0.366**. This suggests that KNN provided the most accurate probability predictions among the models, making it the most reliable model for this task.

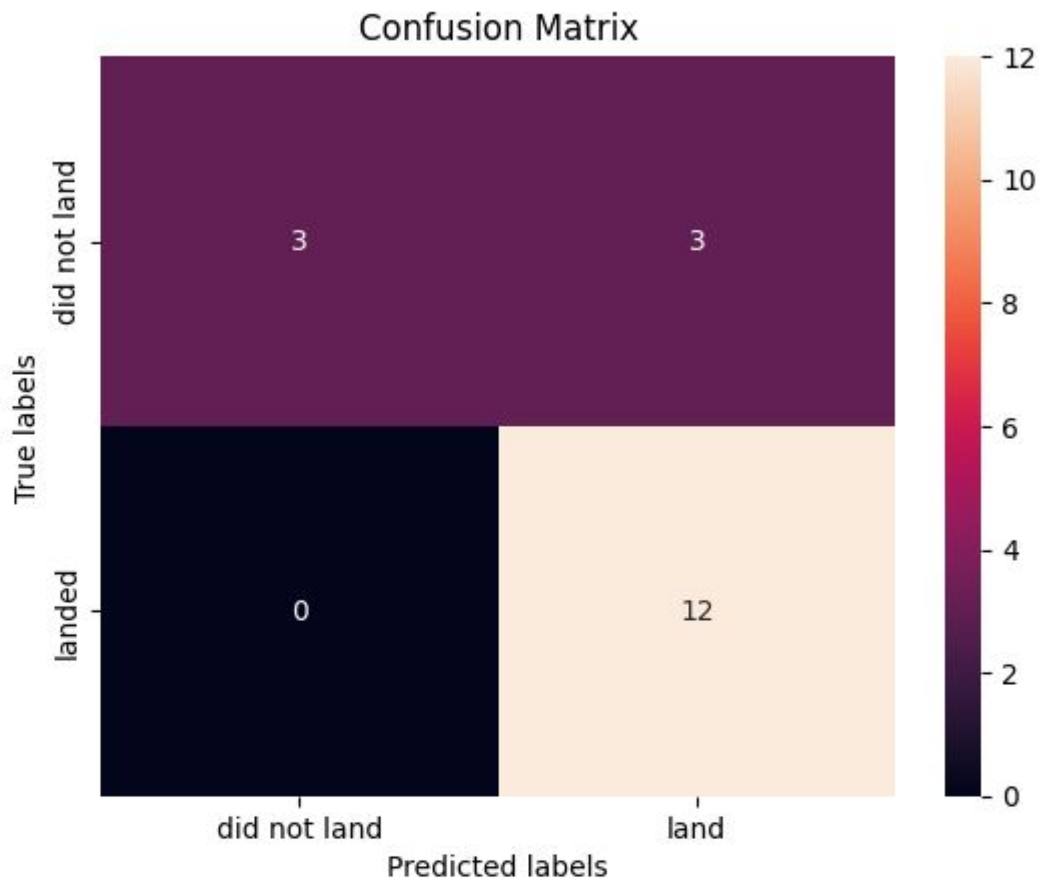


RESULTS

PREDICTIVE ANALYSIS

KNN MODEL CONFUSION MATRIX

The confusion matrix for the best performing model, **KNN**, compares the predicted labels with the true labels. From the matrix, we observe that there are **more true positives than false positives**, indicating that the model is effectively identifying successful launch outcomes. This suggests that the KNN model is reliably predicting successful launches, while minimizing the number of incorrect predictions, which is a favorable outcome for this task.



CONCLUSIONS

CONCLUSIONS

Returning to the questions we posed at the beginning of this analysis, we can now leverage the insights gathered throughout our exploratory data analysis, interactive analytics, and predictive modeling to provide answers.

What factors influence the successful landing of Falcon 9's first stage?

- **Payload Mass**: heavier payloads (greater than 10,000 kg) are more likely to result in successful landings, especially in VLEO and ISS orbits. This suggests that the payload mass plays a significant role in landing success, with heavier payloads correlating with higher success rates in these specific orbits.
- **Orbit Type**: orbits such as ES-L1, GEO, HEO, and SSO had higher success rates for first-stage landings. GTO orbit exhibited the lowest success rate, suggesting that potential challenges of this orbit may contribute to its lower success rate.
- **Booster Version**: the FT booster version stands out with the highest success rate, indicating that advancements in booster technology and newer versions likely play a key role in ensuring successful landings.

It remains unclear whether the **launch site** itself directly influences the success of Falcon 9's first-stage landings. Higher number of unsuccessful outcomes at busier launch sites, like CCAFS SLC 40, could be explained with the frequency of launches and increased probability of encountering operational challenges and unsuccessful outcomes, given the large volume of launch attempts.

CONCLUSIONS

Can we accurately predict the outcome of a first-stage landing based on launch data?

Yes, we can accurately predict the outcome of a first-stage landing based on launch data. The predictive models (KNN, Decision Tree, SVM, and Logistic Regression) achieved an **accuracy score** of **0.83**, demonstrating good precision in predicting landing outcomes. The **KNN** model **performed the best** in terms of predicting the probability of success, with a **Log Loss** of **0.366**. The confusion matrix confirmed that the model was effective at predicting successful landings, with more true positives than false positives. Overall, key features such as payload mass, orbit type, and booster version were found to significantly influence landing success, making it possible to predict outcomes accurately. Note that there is definitely potential for improvement, particularly by exploring additional data, optimizing the model further, and fine-tuning hyperparameters.

KEY TAKEAWAY

Building a successful company focused on reusing Falcon 9's first stage requires an emphasis on advanced booster technology, optimization of payload and orbit conditions, effective management of operational challenges at launch sites, and the use of data-driven insights to predict and improve landing outcomes.

APPENDIX

Please refer to the following datasets created during the project:

Data Collection - SpaceX REST API

- dataset_part_1.csv

Data Collection - Web Scraping

- spacex_web_scraped.csv

Data Wrangling

- dataset_part_2.csv
- dataset_part_3.csv

Dashboard - Plotly Dash

- spacex_launch_dash.csv


IBM Developer
SKILLS NETWORK
BON VOYAGE!