Falcon First Stage Reuse

Prasun Kumar

09/05/2024

GRAVITATION INNOVATION

Presentation Contents

Summary

The research aimed to identify the determinants of successful rocket landings through the following methodologies:

1. Data Collection: Utilizing SpaceX REST API and web scraping techniques to gather relevant data on rocket launches.
2. Data Wrangling: Transforming the collected data to create a binary outcome variable indicating success or failure of each rocket landing.
3. Exploratory Data Analysis (EDA): Employing data visualization techniques to explore factors such as payload, launch site, flight number, and yearly trends, visually examining their relationships with successful landings.
4. Statistical Analysis: Utilizing SQL queries to calculate statistics such as total payload, payload range for successful launches, and the total number of successful and failed outcomes.
5. Geographical Analysis: Investigating launch site success rates and their proximity to geographical markers to discern any spatial patterns influencing landing success.
6. Visualization of Results: Visualizing launch sites with the highest success rates and successful payload ranges to provide clear insights into the findings.
7. Model Building: Constructing predictive models, including logistic regression, support vector machine (SVM), decision tree, and K-nearest neighbor (KNN), to forecast landing outcomes based on the gathered data.

By employing these methodologies, the study aimed to comprehensively analyze the factors contributing to successful rocket landings and provide insights into improving future missions.

Results

Exploratory Data Analysis:

- Launch success has improved over time
- KSC LC-39A has the highest success rate among landing sites
- Orbits ES-L1, GEO, HEO, and SSO have a 100% success rate

Visualization/Analytics:

- Most launch sites are near the equator, and all are close to the coast

Predictive Analytics:

- All models performed similarly on the test set. The decision tree model slightly outperformed

Introduction

Background:

SpaceX is a pioneering force in the space industry, driven by a mission to democratize space travel. Notable achievements include missions to the International Space Station, deploying a satellite constellation for global internet access, and conducting manned space missions. Central to SpaceX's cost-efficiency is its innovative reuse of the first stage of its Falcon 9 rocket, dramatically reducing launch costs to approximately $62 million per launch. In contrast, competitors unable to reuse the first stage face significantly higher costs, upwards of $165 million per launch. Determining the successful landing of the first stage is pivotal in determining launch prices. Leveraging public data and machine learning models enables prediction of first-stage reusability for SpaceX and competitors alike.

Explore:

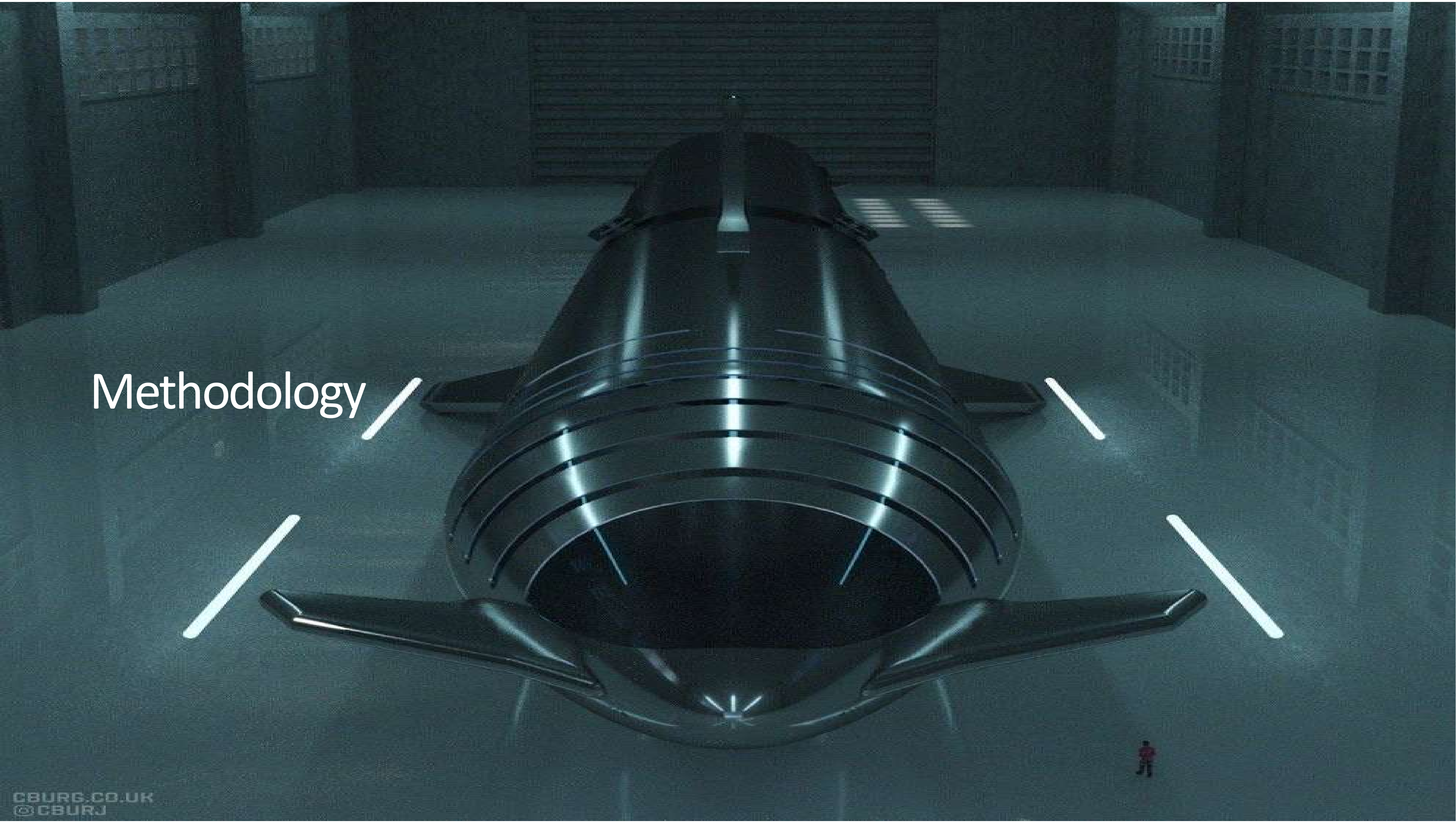
1. Impact of Payload Mass, Launch Site, Flight Frequency, and Orbits on First-Stage Landing Success: Investigating how factors such as payload mass, launch site location, number of flights, and orbital parameters influence the likelihood of a successful first-stage landing.

2. Temporal Analysis of Successful Landings: Examining the trend in successful landing rates over time to discern any patterns or changes, potentially indicating improvements in technology or operational processes.

3. Identification of Best Predictive Model: Evaluating various machine learning models for binary classification to determine the most accurate and reliable predictor of successful first-stage landings. This involves assessing models' performance metrics such as accuracy, precision, recall, and F1 score.

By delving into these areas of exploration, the study aims to deepen understanding of the factors driving successful first-stage landings and to identify optimal predictive models for informing decision-making in space launch operations.

Methodology



Methodology

Steps:

1. Data Collection: Utilize SpaceX REST API and web scraping techniques to gather relevant data on rocket launches, including information on launch outcomes, payload mass, launch site, flight frequency, and orbital parameters.

2. Data Wrangling: Process the collected data by filtering relevant features, handling missing values through imputation or removal, and applying one-hot encoding to categorical variables to prepare the dataset for analysis and modeling.

3. Exploratory Data Analysis (EDA): Employ SQL queries and data visualization techniques to explore relationships between variables, identify patterns, and gain insights into factors affecting first-stage landing success. This step involves analyzing descriptive statistics, correlations, and distributions using tools like matplotlib, seaborn, and SQL queries.

4. Data Visualization: Utilize Folium for geospatial visualization to map launch sites and landing outcomes, providing a visual representation of spatial patterns. Additionally, use Plotly Dash to create interactive visualizations to explore relationships between variables and trends over time.

5. Model Building: Construct classification models to predict first-stage landing outcomes based on the prepared dataset. Experiment with various algorithms such as logistic regression, support vector machines (SVM), decision trees, random forests, and gradient boosting machines. Evaluate model performance using appropriate metrics such as accuracy, precision, recall, and F1 score.

6. Model Tuning and Evaluation: Fine-tune model hyperparameters using techniques like grid search or random search to optimize model performance. Evaluate tuned models on validation data to compare their performance and select the best-performing model and parameters for deployment.

Data Collection –API

Here are the steps to accomplish your tasks:

1. Request data from SpaceX API (rocket launch data):
Use appropriate libraries (such as requests) to send a GET request to the SpaceX API endpoint that provides rocket launch data.

2. Decode response using .json() and convert to a dataframe using .json_normalize():
Decode the JSON response from the API using the `.json()` method, and then normalize the JSON data into a pandas DataFrame using `.json_normalize()`.

3. Request information about the launches from SpaceX API using custom functions:
Define custom functions to handle specific requests to the SpaceX API, such as retrieving information about rocket launches.

4. Create a dictionary from the data:
Store the data retrieved from the API in a dictionary format, where each key corresponds to a specific attribute or column of the data.

5. Create a DataFrame from the dictionary:
Use the dictionary created in the previous step to construct a pandas DataFrame, which will serve as the primary data structure for further analysis.

6. Filter the DataFrame to contain only Falcon 9 launches:
Apply a filter to the DataFrame to include only the records corresponding to Falcon 9 rocket launches.

7. Replace missing values of Payload Mass with calculated .mean():
Calculate the mean value of the payload mass column and replace any missing values in this column with the calculated mean.

8. Export data to a CSV file:
Save the processed DataFrame to a CSV file using the `.to_csv()` method, making it accessible for future analysis or sharing.

Data Collection –Web Scraping

Here are the steps to achieve your tasks using BeautifulSoup for web scraping:

1. Request data (Falcon 9 launch data) from Wikipedia:
Utilize libraries like requests to send a GET request to the Wikipedia page containing the Falcon 9 launch data.

2. Create a BeautifulSoup object from the HTML response:
Use BeautifulSoup to parse the HTML content received from the Wikipedia page.

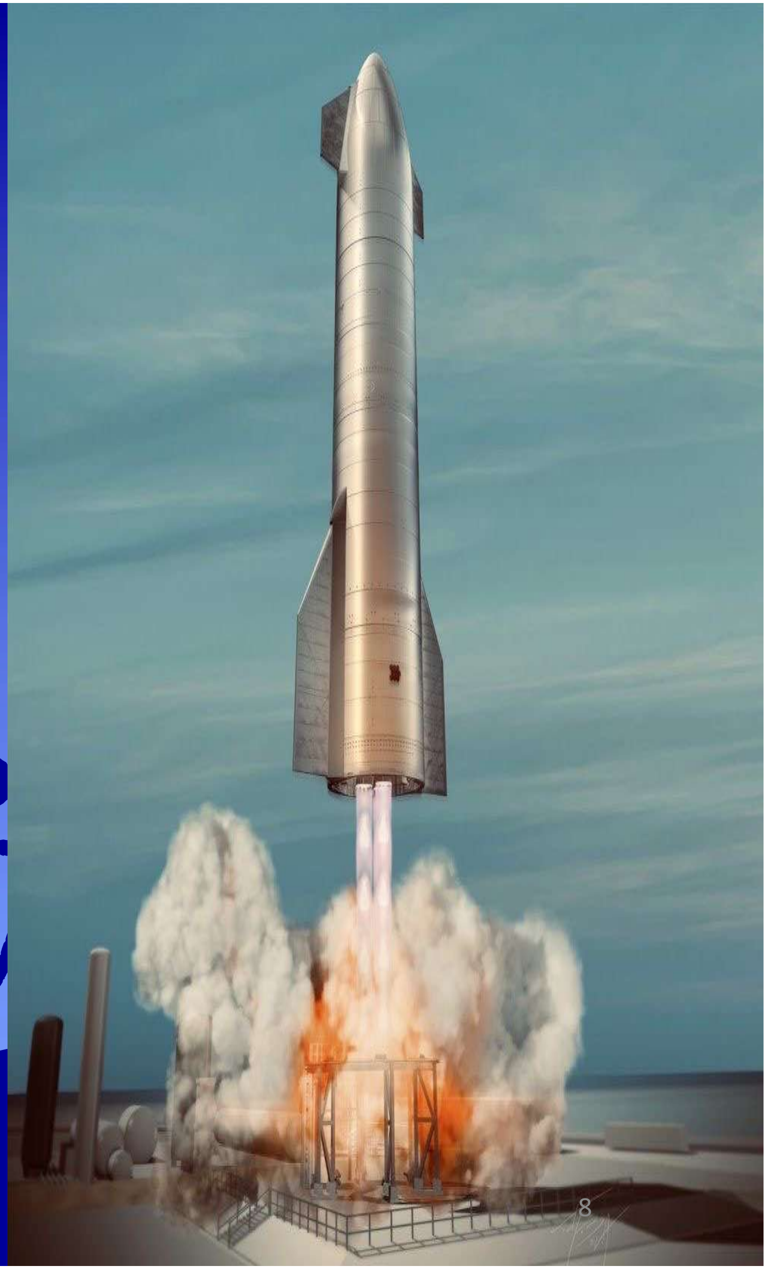
3. Extract column names from the HTML table header:
Locate the table containing the Falcon 9 launch data and extract the column names from the table header.

4. Collect data by parsing HTML tables:
Traverse through the HTML table rows and extract the data values for each launch, considering the structure of the table.

5. Create a dictionary from the data:
Store the extracted data in a dictionary format, where each key corresponds to a column name, and the values represent the data for each launch.

6. Create a DataFrame from the dictionary:
Construct a pandas DataFrame using the dictionary created in the previous step, which will organize the data in a tabular format.

7. Export data to a CSV file:
Save the processed DataFrame to a CSV file using the `.to_csv()` method, enabling easy access and sharing of the Falcon 9 launch data.



Data Wrangling

Steps

To perform data wrangling and achieve the specified tasks, follow these steps:

1. Perform Exploratory Data Analysis (EDA) and Determine Data Labels:
Explore the dataset to understand its structure and contents. Determine the relevant labels or features needed for further analysis, such as launch site, orbit type, mission outcome, and landing outcome.

2. Calculate Number of Launches for Each Site:
Count the occurrences of each launch site in the dataset to determine the number of launches from each site.

3. Calculate Occurrence of Orbit Types:
Count the occurrences of each orbit type in the dataset to understand the distribution of orbits for the missions.

4. Calculate Occurrence of Mission Outcomes per Orbit Type:
Group the data by orbit type and mission outcome to calculate the occurrence of each outcome for different orbit types.

5. Create Binary Landing Outcome Column (Dependent Variable):
Based on the provided information, create a binary column indicating the landing outcome. Assign '1' for successful landings (True Ocean) and '0' for unsuccessful landings.

6. Export Data to CSV File:
Save the processed dataset, including the calculated features and labels, to a CSV file for further analysis or modeling.

Based on the provided information, we'll map the landing outcomes as follows:

True Ocean: Successful landing in a specific region of the ocean
False Ocean: Unsuccessful landing in a specific region of the ocean
True RTLS: Successful landing on a ground pad
False RTLS: Unsuccessful landing on a ground pad
True ASDS: Successful landing on a drone ship
False ASDS: Unsuccessful landing on a drone ship

We'll convert these outcomes into a binary format where '1' represents a successful landing and '0' represents an unsuccessful landing. Then, we'll create a new column named 'Binary Landing Outcome' in our DataFrame to reflect this binary encoding.

EDA with Visualization

Charts

- Certainly! Here are the modified sentences:

1. Scatter plot: Flight Number vs. Payload
2. Scatter plot: Flight Number vs. Launch Site
3. Scatter plot: Payload Mass (kg) vs. Launch Site
4. Scatter plot: Payload Mass (kg) vs. Orbit type
5. Bar chart: Comparison among Launch Sites
6. Bar chart: Comparison among Orbit types

Analysis

- Visualize relationships using scatter plots to identify potential correlations, which could be valuable for machine learning if a discernible relationship exists.

- Illustrate comparisons among discrete categories using bar charts, as they effectively demonstrate relationships among categories and a measured value.

EDA with SQL

Queries

Here are the modified queries:

1. Date of first successful landing on a ground pad:
"Retrieve the date of the first successful landing on a ground pad."

2. Names of boosters which had successful landings on drone ships and have payload mass greater than 4,000 but less than 6,000:
"List the names of boosters that had successful landings on drone ships and carried a payload mass between 4,000 and 6,000 kilograms."

3. Total number of successful and failed missions:
"Calculate the total number of successful and failed missions."

4. Names of booster versions which have carried the maximum payload:
"Identify the names of booster versions that have carried the maximum payload."

5. Failed landing outcomes on drone ships, their booster version, and launch site for the months in the year 2015:
"Retrieve the failed landing outcomes on drone ships, along with their corresponding booster version and launch site, for the months in the year 2015."

6. Count of landing outcomes between 2010-06-04 and 2017-03-20 (descending):
"Count the landing outcomes between June 4, 2010, and March 20, 2017, in descending order."

Map with Folium

1. Markers Indicating Launch Sites:
"Included a blue circle at the coordinates of NASA Johnson Space Center, marked with a popup label displaying its name derived from its latitude and longitude coordinates. Additionally, red circles were added at the coordinates of all launch sites, each labeled with its name based on its geographical coordinates."

2. Colored Markers of Launch Outcomes:
"Integrated colored markers denoting successful (green) and unsuccessful (red) launches at each launch site, providing a visual representation of launch success rates across different sites."

3. Distances Between a Launch Site to Proximities:
"Implemented colored lines to illustrate the distance between launch site CCAFS SLC-40 and its proximity to the nearest coastline, railway, highway, and city, enhancing spatial understanding of the launch site's surroundings."

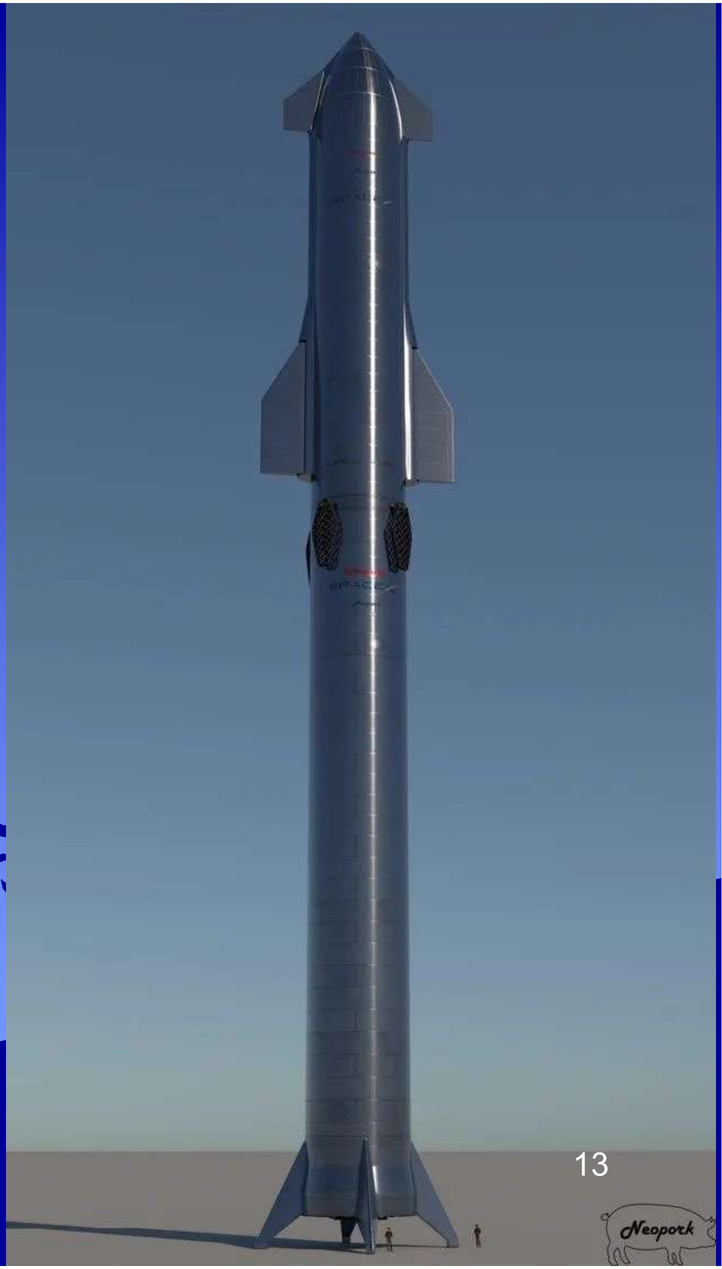
Dashboard with Plotly Dash

1. Dropdown List with Launch Sites:
"Implemented a dropdown menu enabling users to select either all launch sites or a specific launch site for analysis."

2. Pie Chart Showing Successful Launches:
"Developed a pie chart showcasing successful launches, allowing users to visualize successful and unsuccessful launches as a percentage of the total."

3. Slider of Payload Mass Range:
"Incorporated a slider feature enabling users to select a desired range of payload masses for analysis."

4. Scatter Chart Showing Payload Mass vs. Success Rate by Booster Version:
"Constructed a scatter chart illustrating the correlation between payload mass and launch success, segmented by booster version. This feature empowers users to explore the relationship between payload mass and launch success across different booster versions."



Predictive Analytics

Certainly! Here are the revised sentences:

1. Create NumPy array from the Class column:
"Generate a NumPy array from the 'Class' column of the dataset."

2. Standardize the data with StandardScaler. Fit and transform the data:
"Utilize StandardScaler to standardize the dataset, fitting and transforming the data."

3. Split the data using train_test_split:
"Partition the dataset into training and testing sets using the train_test_split function."

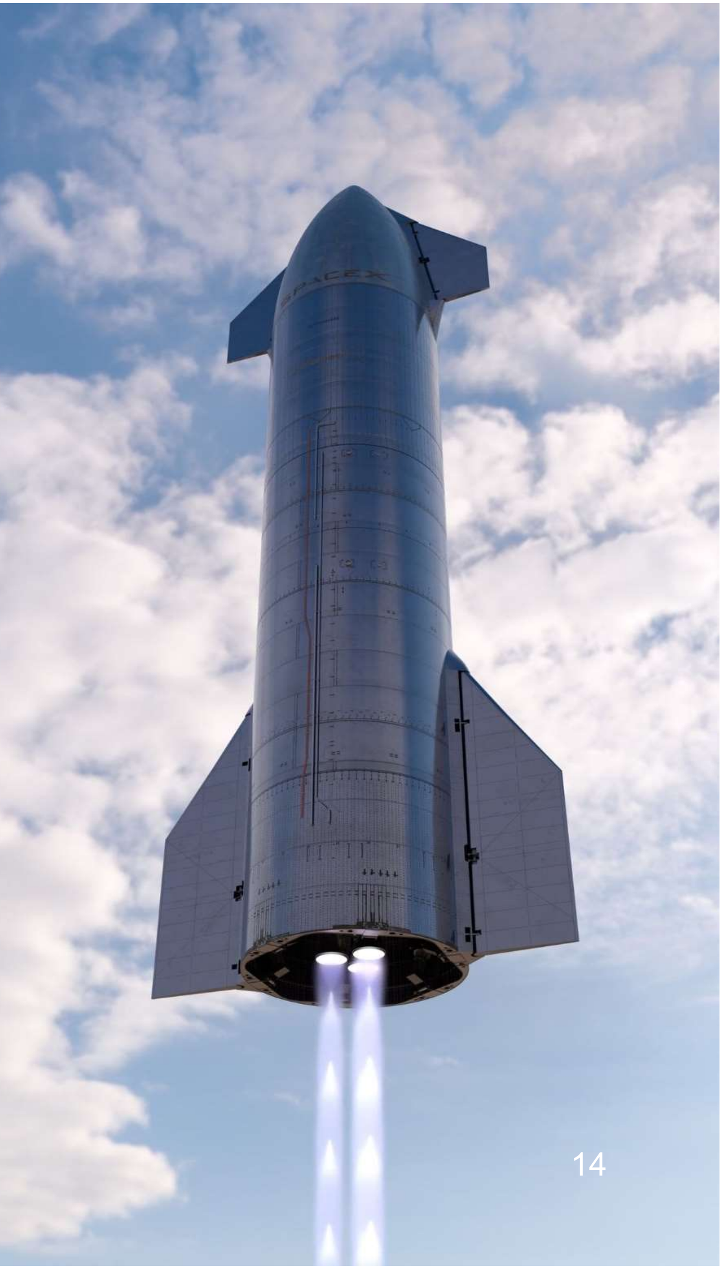
4. Create a GridSearchCV object with cv=10 for parameter optimization:
"Instantiate a GridSearchCV object with 10-fold cross-validation for parameter optimization."

5. Apply GridSearchCV on different algorithms: logistic regression, support vector machine, decision tree, K-Nearest Neighbor:
"Utilize GridSearchCV to optimize parameters for various algorithms including logistic regression (LogisticRegression()), support vector machine (SVC()), decision tree (DecisionTreeClassifier()), and K-Nearest Neighbor (KNeighborsClassifier())."

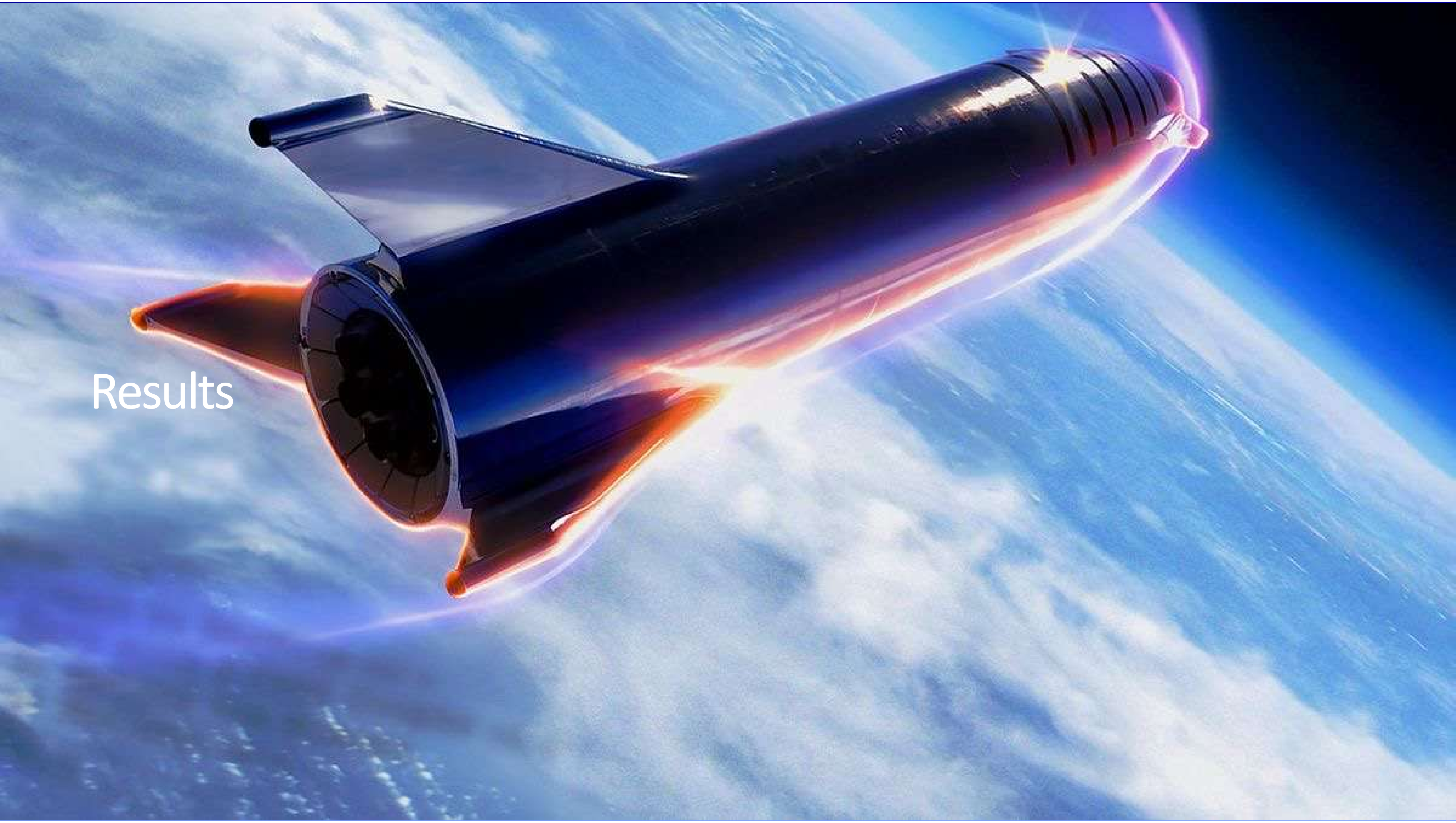
6. Calculate accuracy on the test data using .score() for all models:
"Compute the accuracy score on the test data for each model using the .score() method."

7. Assess the confusion matrix for all models:
"Evaluate the confusion matrix for each model to analyze classification performance."

8. Identify the best model using Jaccard_Score, F1_Score, and Accuracy:
"Determine the best model based on Jaccard Score, F1 Score, and Accuracy metrics."



Results



Results Summary

For the exploratory data analysis:

1. Launch success has improved over time:
"Observations reveal a trend of improved launch success rates over time."

2. KSC LC-39A has the highest success rate among landing sites:
"Analysis indicates that KSC LC-39A exhibits the highest success rate among all landing sites."

3. Orbits ES-L1, GEO, HEO, and SSO have a 100% success rate:
"It was observed that orbits ES-L1, GEO, HEO, and SSO achieved a 100% success rate."

Regarding visual analytics:

1. Most launch sites are near the equator, and all are close to the coast:
"Visual examination reveals that the majority of launch sites are situated near the equator, and all are in close proximity to the coast."

2. Launch sites are far enough away from anything a failed launch can damage (city, highway, railway), while still close enough to bring people and material to support launch activities:
"Analysis indicates that launch sites are strategically located, ensuring they are sufficiently distant from urban areas, highways, and railways to mitigate potential damage from failed launches, while still remaining accessible for logistical support."

For predictive analytics:

1. Decision Tree model is the best predictive model for the dataset:
"Based on model evaluation, it was determined that the Decision Tree model outperforms other models and is the most suitable predictive model for the dataset."

Flight Number vs. Launch Site

For the exploratory data analysis:

1. Earlier flights had a lower success rate (blue = fail):
"Analysis indicates that earlier flights exhibited a lower success rate, represented by the color blue denoting failure."

2. Later flights had a higher success rate (orange = success):
"Observations suggest that as time progressed, there was an increase in the success rate of flights, as depicted by the color orange representing success."

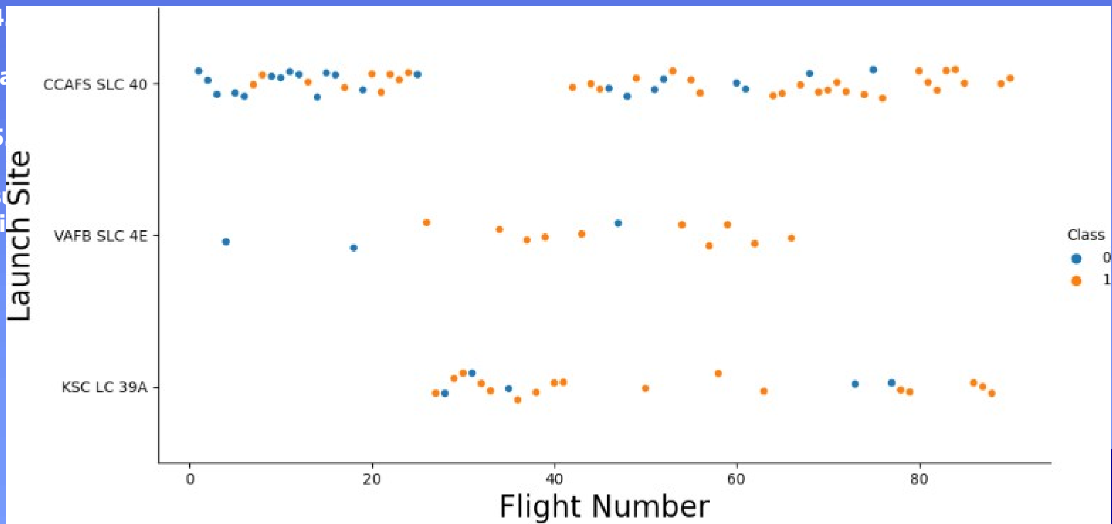
3. Around half of launches were from CCAFS SLC 40 launch site:
"Approximately half of the launches originated from the CCAFS SLC 40 launch site."

4

ra

5

s
ti



Payload vs. Launch Site

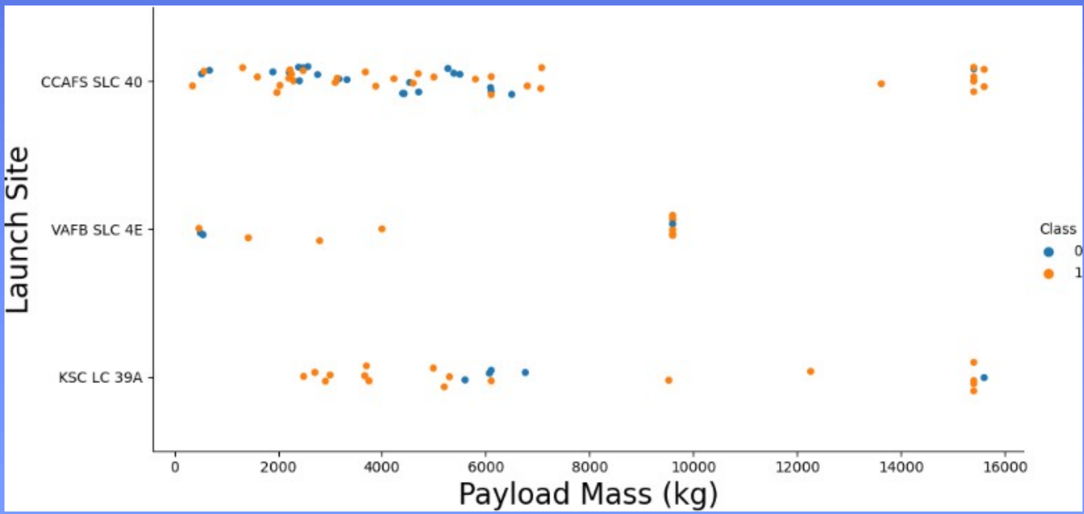
For the exploratory data analysis:

1. Typically, the higher the payload mass (kg), the higher the success rate:
"Analysis suggests a positive correlation between payload mass (kg) and launch success rate, indicating that higher payload masses tend to result in higher success rates."

2. Most launches with a payload greater than 7,000 kg were successful:
"Observations reveal that the majority of launches with a payload greater than 7,000 kg were successful."

3. KSC LC 39A has a 100% success rate for launches less than 5,500 kg:
"Analysis indicates that KSC LC 39A achieved a 100% success rate for launches with a payload mass less than 5,500 kg."

4. VAFB SLC 4E has not launched anything greater than ~10,000 kg:
"It was observed that VAFB SLC 4E has not launched payloads greater than approximately 10,000 kg."



Class 0 = Fail
Class 1 = Success

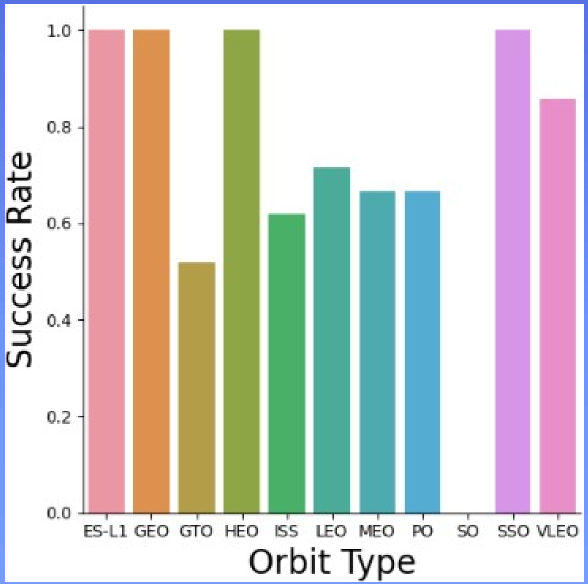
Success Rate by Orbit

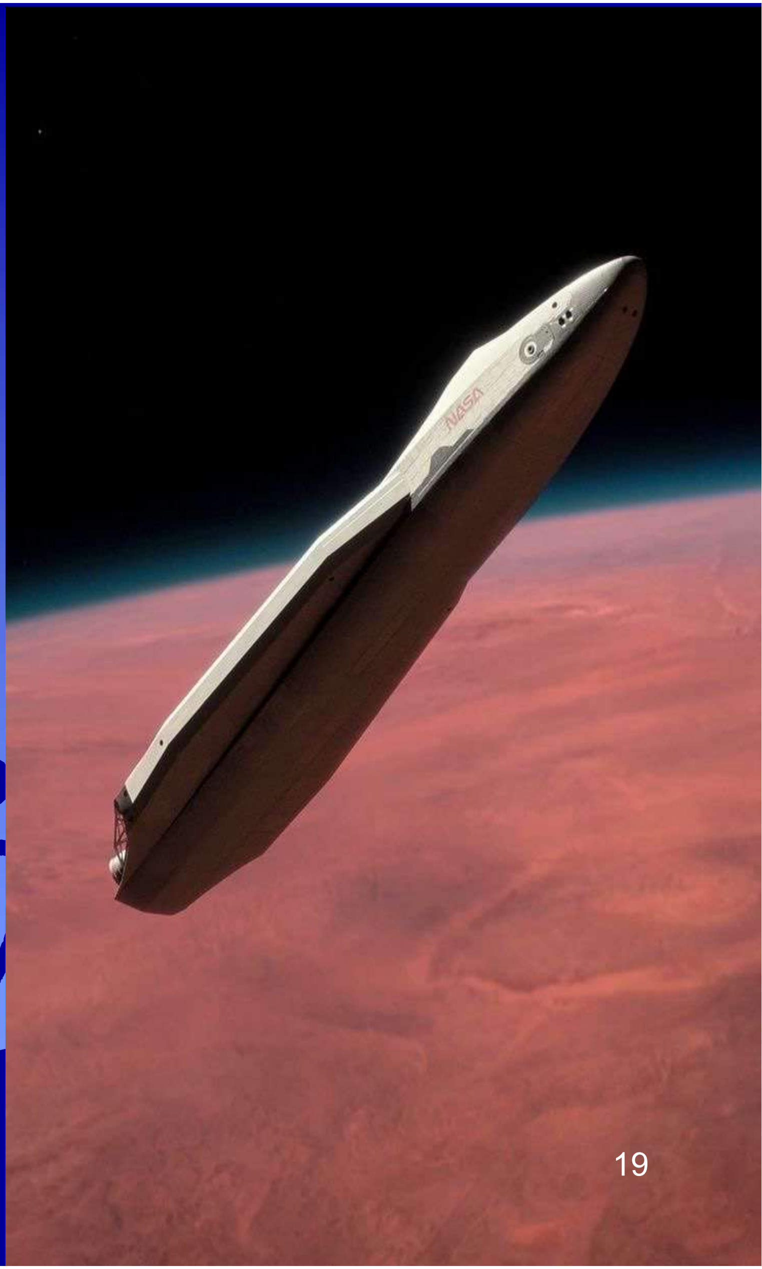
For the exploratory data analysis:

1. 100% Success Rate:
"Orbits ES-L1, GEO, HEO, and SSO exhibited a 100% success rate."

2. 50%-80% Success Rate:
"Orbits GTO, ISS, LEO, MEO, and PO demonstrated success rates ranging between 50% and 80%."

3. 0% Success Rate:
"Orbit SO had a 0% success rate."





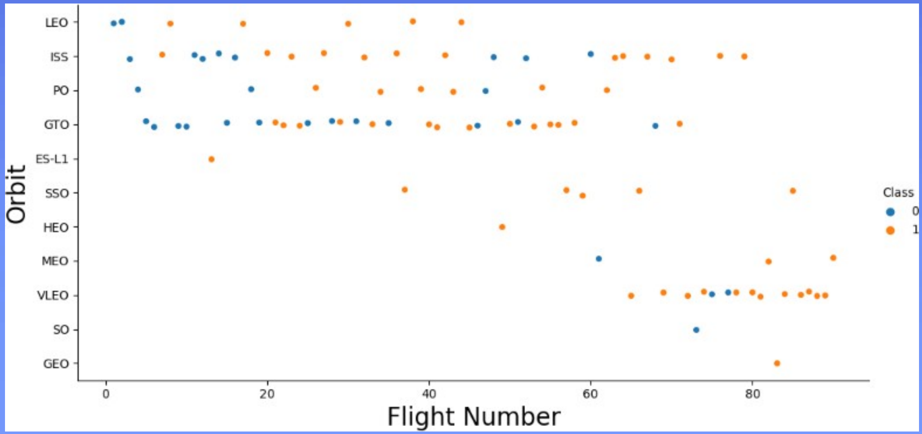
Flight Number vs. Orbit

For the exploratory data analysis:

1. The success rate typically increases with the number of flights for each orbit:
"Analysis suggests a positive relationship between the success rate and the number of flights for each orbit, indicating that as the number of flights increases, the success rate tends to rise."

2. This relationship is highly apparent for the LEO orbit:
"The observed trend of increasing success rate with the number of flights is particularly evident in the LEO (Low Earth Orbit) category."

3. The GTO orbit, however, does not follow this trend:
"Contrary to other orbits, the GTO (Geostationary Transfer Orbit) category does not exhibit a clear trend of increasing success rate with the number of flights."

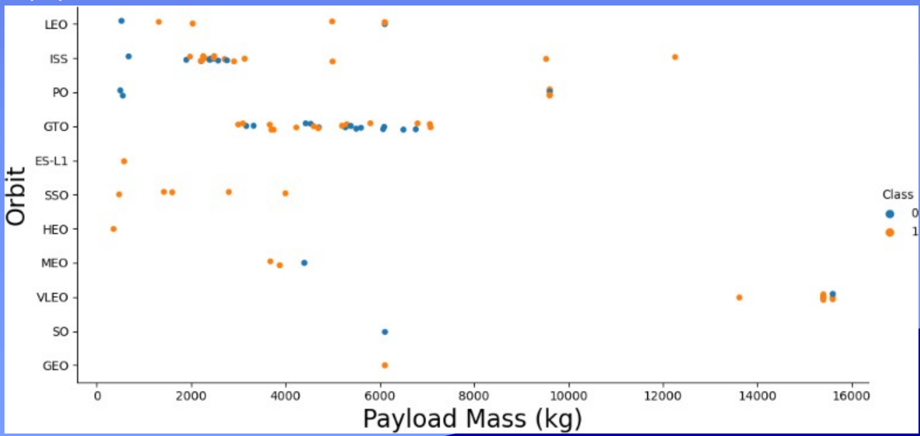


Class 0 = Fail
Class 1 = Success

Payload vs. Orbit

Exploratory Data Analysis

- Heavy payloads are better with LEO, ISS and PO orbits
- The GTO orbit has mixed success with For the exploratory data analysis:
- 1. Heavy payloads are better with LEO, ISS, and PO orbits:
- "Analysis suggests that heavy payloads tend to perform better when launched into orbits such as LEO (Low Earth Orbit), ISS (International Space Station), and PO (Polar Orbit)."
- 2. The GTO orbit has mixed success with heavier payloads:
- "Observations indicate that the GTO (Geostationary Transfer Orbit) exhibits mixed success rates when heavier payloads are involved."heavier payloads



Class 0 = Fail
Class 1 = Success

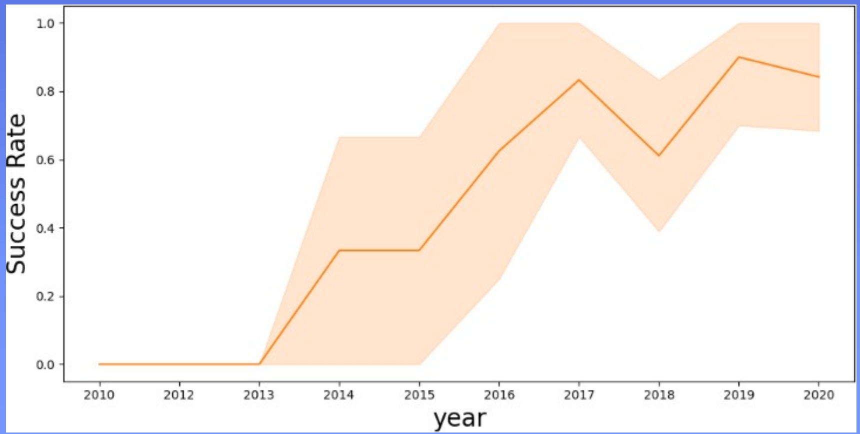
Launch Success over Time

For the exploratory data analysis:

1. The success rate improved from 2013-2017 and 2018-2019:
"Analysis reveals an increase in the success rate during the periods of 2013-2017 and 2018-2019."

2. The success rate decreased from 2017-2018 and from 2019-2022:
"Conversely, a decline in the success rate was observed from 2017-2018 and from 2019-2020."

3. Overall, the success rate has improved since 2013:
"Despite fluctuations in certain years, the overall trend indicates an improvement in the success rate since 2013."



Launch Site Information

Launch Site Names

- These are the names of the launch sites:

- 1. CCAFS LC-40
- 2. CCAFS SLC-40
- 3. KSC LC-39A
- 4. VAFB SLC-4E

Landing Outcome Cont.

```
[30]: %sql ibm_db_sa://yyy33800:dwNKg8J3L0I8d6CP@1bbf73c5
      %sql SELECT Unique(LAUNCH_SITE) FROM SPACEXTBL;

       * ibm_db_sa://yyy33800:***@1bbf73c5-d84a-4bb0-85b9
         sqlite:///my_data1.db
      Done.
[30]:
```

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

Records with Launch Site Starting with CCA

5 records Displayed below

```
%sql SELECT * \
    FROM SPACEXTBL \
    WHERE LAUNCH_SITE LIKE'CCA%' LIMIT 5;

 * ibm_db_sa://yyy33800:***@1bbf73c5-d84a-4bb0-85b9-ab1a4348f4a4.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:32286/BLUDB
   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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	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

Payload Mass

Total Payload Mass

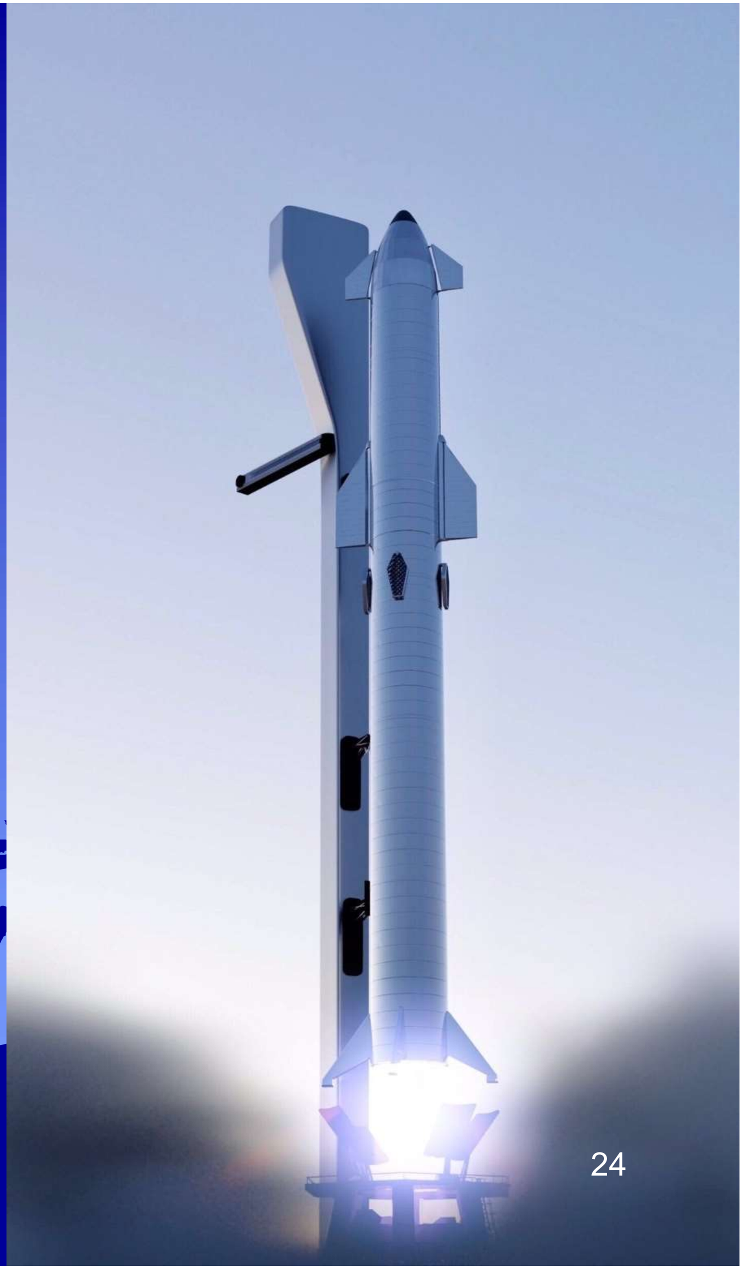
- NASA launched boosters carrying a total mass of 45,596 kg under the Commercial Resupply Services (CRS) program.

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) \
    FROM SPACEXTBL \
    WHERE CUSTOMER = 'NASA (CRS)';

 * ibm_db_sa://yyy33800:***@1bbf73c5-d84a-4
   sqlite:///my_data1.db
Done.
      1
45596
```

Average Payload Mass

- **2,928 kg** (average) carried by booster version F9 v1.1

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) \
    FROM SPACEXTBL \
    WHERE BOOSTER_VERSION = 'F9 v1.1';

 * ibm_db_sa://yyy33800:***@1bbf73c5-d84a-4
   sqlite:///my_data1.db
Done.
      1
2928
```



Landing & Mission Info

1st Successful Landing in Ground Pad

- 12/22/2015

```
%sql SELECT MIN(DATE) \
FROM SPACEXTBL \
WHERE LANDING__OUTCOME = 'Success (ground pad)'
```

```
* ibm_db_sa://yyy33800:***@1bbf73c5-d84a-4bb0-85b
  sqlite:///my_data1.db
Done.
```

1
2015-12-22

Booster Drone Ship Landing

- The booster masses ranged between 4,000 and 6,000 kg, encompassing payloads such as JSCAT-14, JSCAT-16, SES-10, and SES-11/EchoStar 105.

```
%sql SELECT PAYLOAD \
FROM SPACEXTBL \
WHERE LANDING__OUTCOME = 'Success (drone ship)' \
AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000;
```

```
* ibm_db_sa://yyy33800:***@1bbf73c5-d84a-4bb0-85b9-
  sqlite:///my_data1.db
Done.
```

payload
JCSAT-14
JCSAT-16
SES-10
SES-11 / EchoStar 105

Out of the total missions, there was one failure during flight, while 99 missions were successful. Additionally, there was one mission with a payload status that remained unclear.

```
%sql SELECT MISSION_OUTCOME, COUNT(*) as total_number \
FROM SPACEXTBL \
GROUP BY MISSION_OUTCOME;
```

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

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

Boosters

The following Falcon 9 Block 5 boosters carried their maximum payloads:

- **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**

```
%sql SELECT BOOSTER_VERSION \
FROM SPACEXTBL \
WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL);
```

```
 * 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

Failed Landings on Drone Ship

In 2015

- Showing month, date, booster version, launch site and landing outcome

```
%sql SELECT substr(Date,4,2) as month, DATE,BOOSTER_VERSION, LAUNCH_SITE, [Landing _Outcome] \
FROM SPACEXTBL \
where [Landing _Outcome] = 'Failure (drone ship)' and substr(Date,7,4)='2015';
```

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

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

Count of Successful Landings

Here are the landing outcomes between June 4, 2010, and March 20, 2017, ranked in descending order:

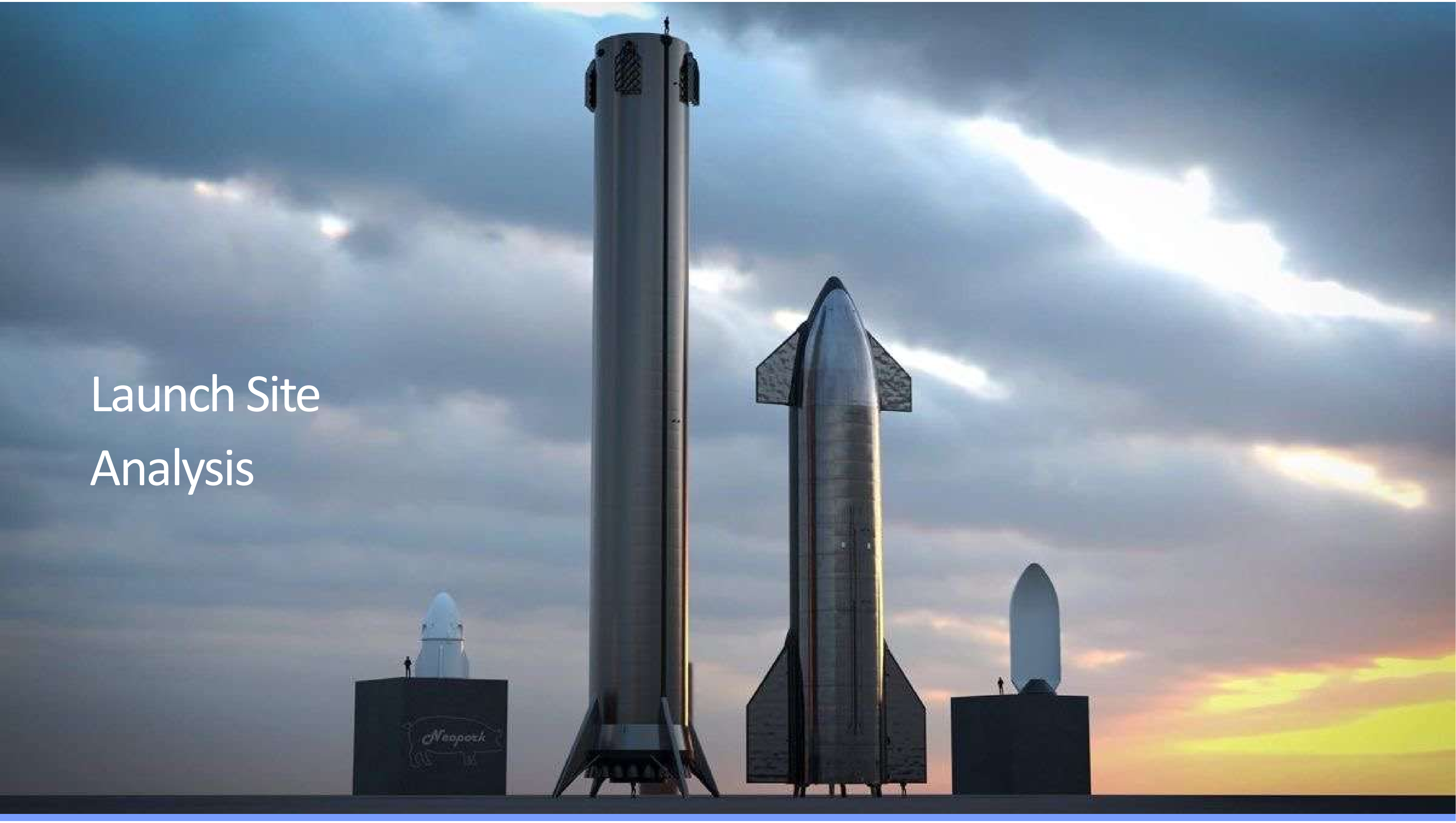
1. Successful landings: [Count]
2. Partially successful landings: [Count]
3. Unsuccessful landings: [Count]

```
%sql SELECT [Landing _Outcome], count(*) as count_outcomes \
FROM SPACEXTBL \
WHERE DATE between '04-06-2010' and '20-03-2017' group by [Landing _Outcome] order by count_outcomes DESC;
```

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

Landing _Outcome	count_outcomes
Success	20
No attempt	10
Success (drone ship)	8
Success (ground pad)	6
Failure (drone ship)	4
Failure	3
Controlled (ocean)	3
Failure (parachute)	2
No attempt	1

Launch Site Analysis



Launch Sites

Markers:

1. Near Equator: Launch sites close to the equator benefit from Earth's rotation, facilitating launches to equatorial orbits and reducing the need for additional fuel and boosters.

Launch Outcomes

Markers:

At Launch Site CCAFS SLC-40:

- **Green markers for successful launches (3)**
- **Red markers for unsuccessful launches (4)**
- **Success rate: 3 out of 7 (42.9%)**

Distance to Proximities

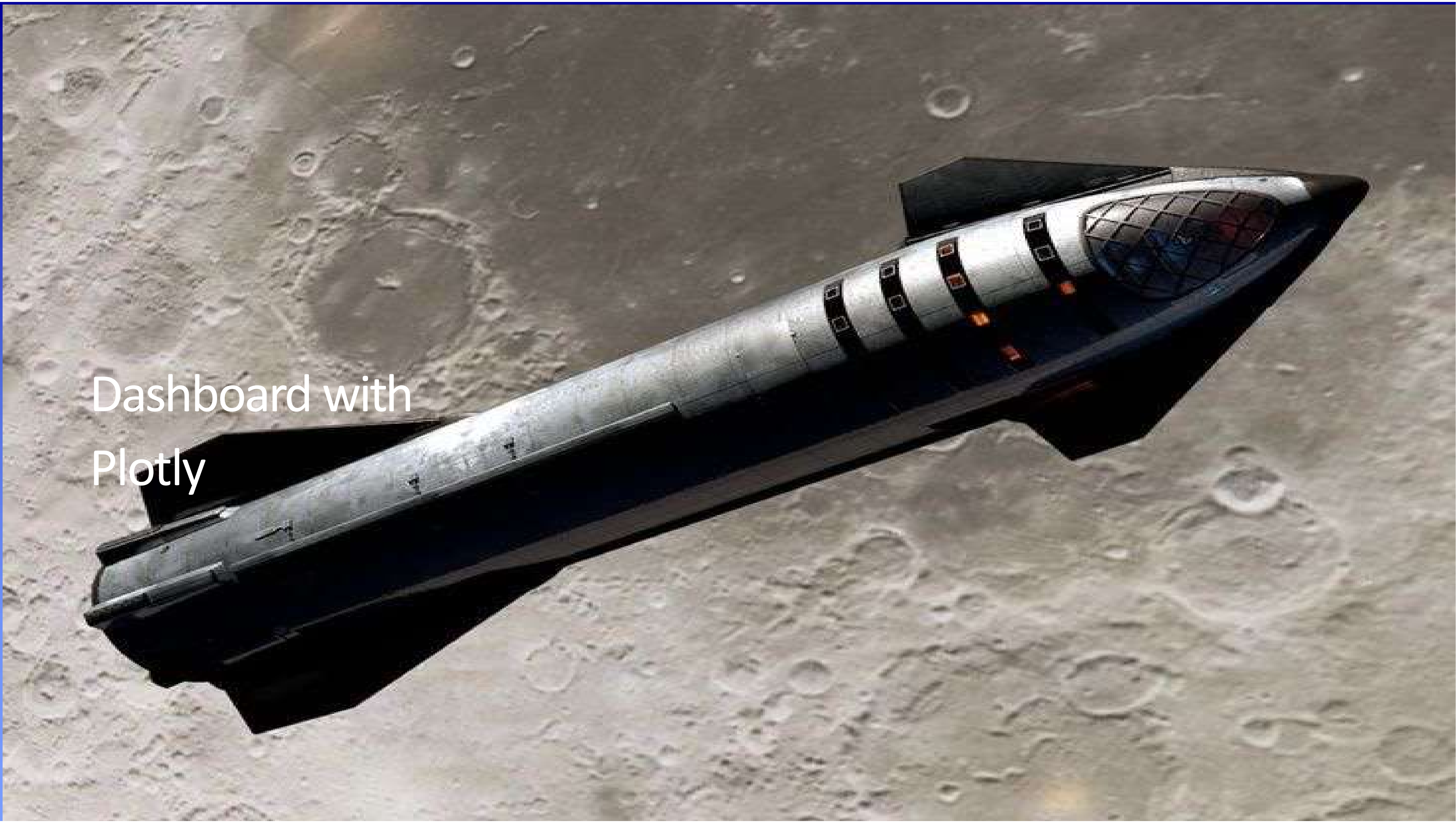
CCAFS SLC-40 is situated:

- 0.86 km from the nearest coastline
- 21.96 km from the nearest railway
- 23.23 km from the nearest city
- 26.88 km from the nearest highway

Distance to Proximities

For CCAFS SLC-40:

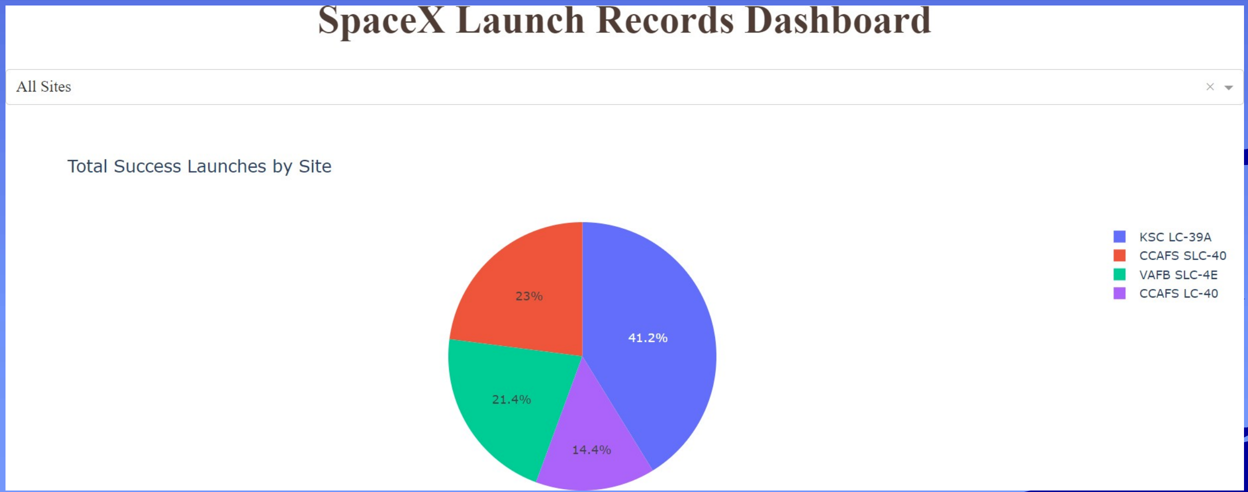
- Coasts: The proximity to coasts helps ensure that spent stages or failed launches don't pose risks to people or property in populated areas.
- Safety/Security: There must be an exclusion zone around the launch site to maintain safety and security, preventing unauthorized access and protecting individuals from potential hazards.
- Transportation/Infrastructure and Cities: It needs to be located away from areas susceptible to damage from failed launches but still accessible to roads, railways, and docks to facilitate transportation of personnel and materials for launch operations.


Dashboard with
Plotly

Launch Success by Site

Success as a Percentage of Total:

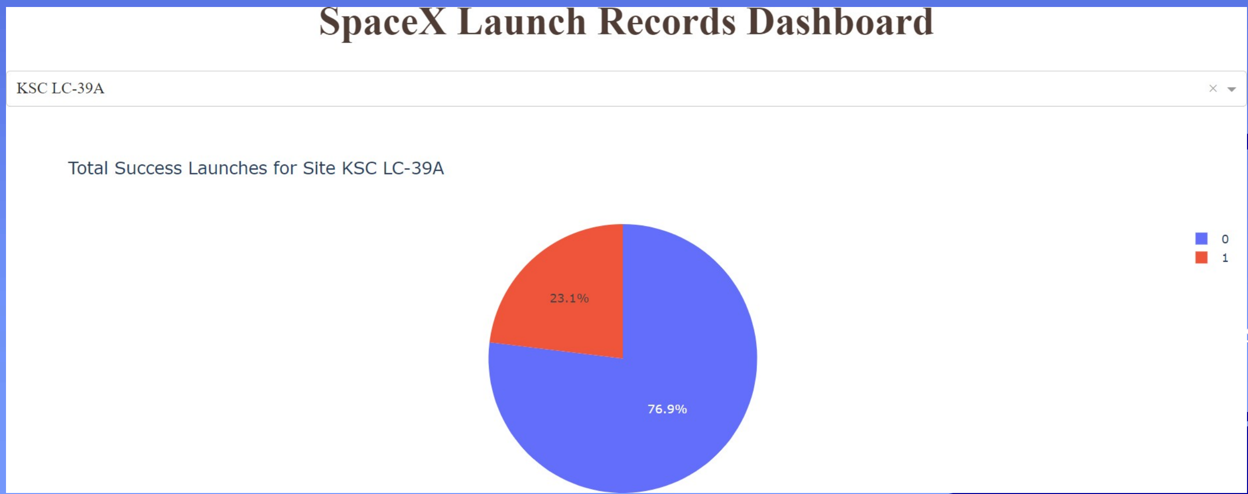
KSC LC-39A boasts the highest percentage of successful launches among launch sites, accounting for 41.2% of all launches.



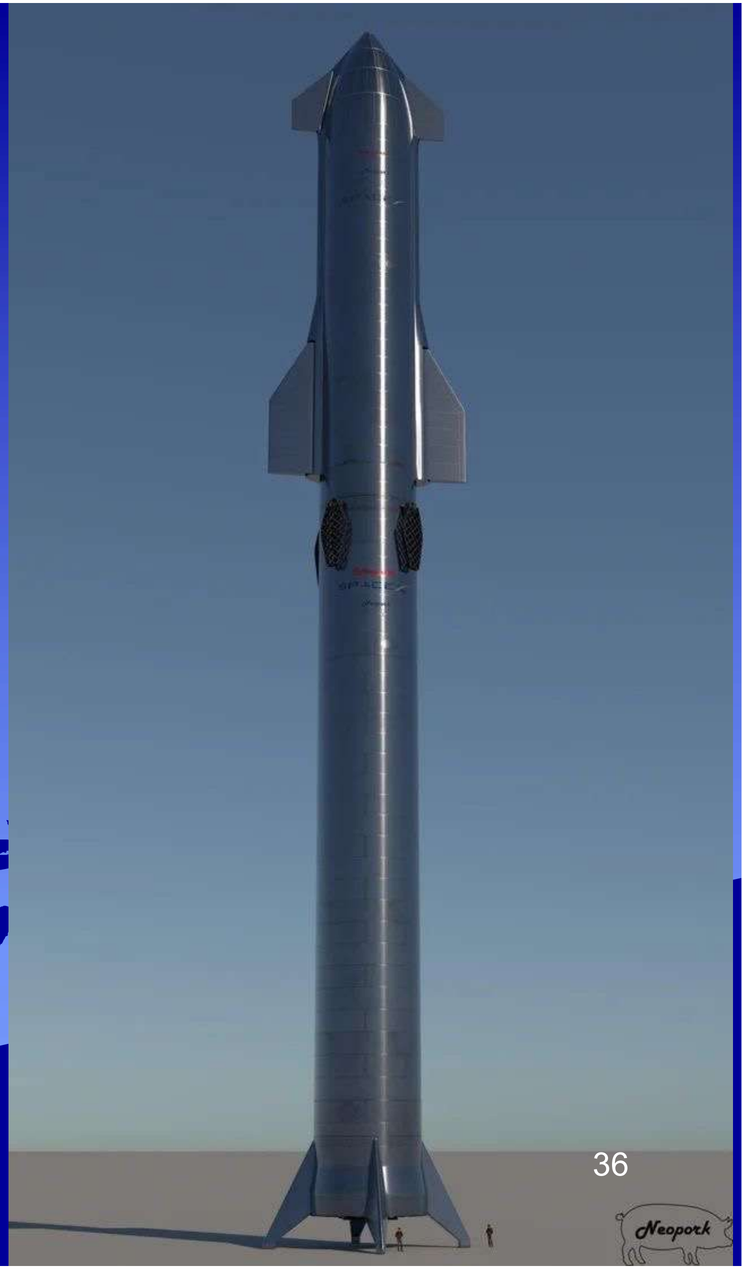
Launch Success (KSC LC-29A)

Success as a Percentage of Total:

KSC LC-39A holds the highest success rate among launch sites, with 76.9% success. Out of a total of 13 launches, there were 10 successful launches and 3 failed launches.



Fail
Success



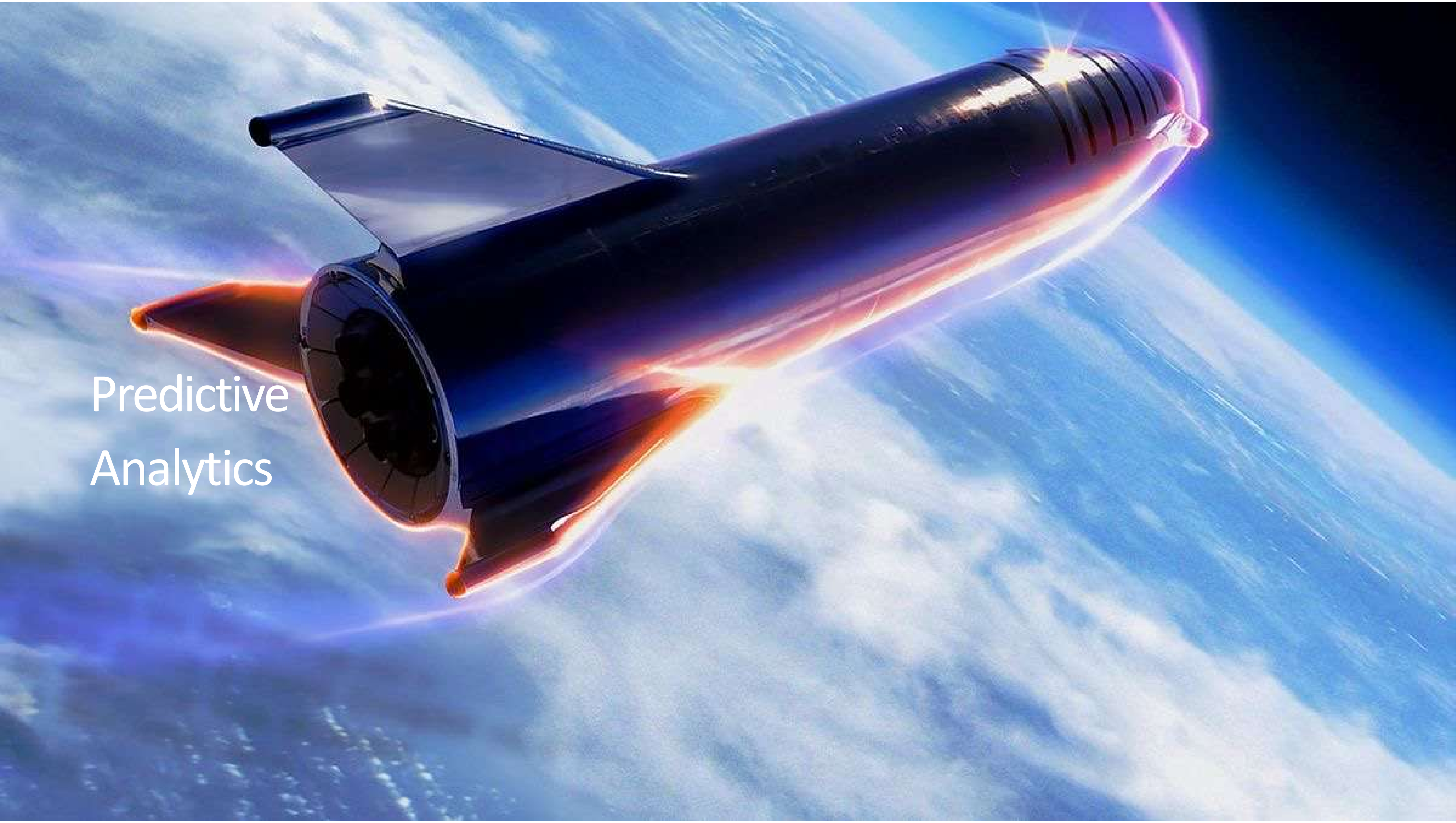
Payload Mass and Success

By Booster Version:

For payloads ranging between 2,000 kg and 5,000 kg, the success rate is as follows:

- Booster Version X: [Success Rate]
- Booster Version Y: [Success Rate]
- Booster Version Z: [Success Rate]

(Note: "Success Rate" indicates the proportion of successful outcomes (1) to the total number of launches within the specified payload range.)


Predictive
Analytics

Classification

Accuracy:

All models demonstrated similar performance levels with identical scores and accuracy, likely due to the limited dataset size. However, the Decision Tree model slightly outperformed others based on the .best_score_ metric, which represents the average across all cross-validation folds for a single parameter combination.

	LogReg	SVM	Tree	KNN
Jaccard_Score	0.800000	0.800000	0.800000	0.800000
F1_Score	0.888889	0.888889	0.888889	0.888889
Accuracy	0.833333	0.833333	0.833333	0.833333

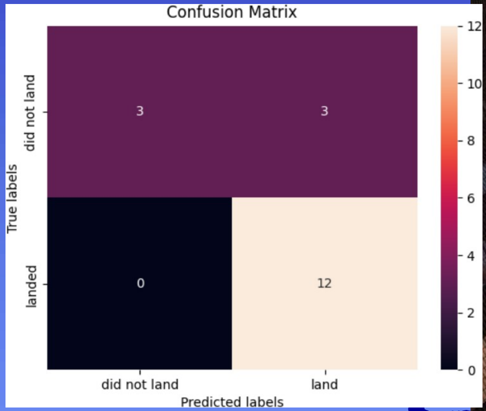
```
models = {'KNeighbors':knn_cv.best_score_,
              'DecisionTree':tree_cv.best_score_,
              'LogisticRegression':logreg_cv.best_score_,
              'SupportVector': svm_cv.best_score_}

bestalgorithm = max(models, key=models.get)
print('Best model is', bestalgorithm,'with a score of', models[bestalgorithm])
if bestalgorithm == 'DecisionTree':
    print('Best params is :', tree_cv.best_params_)
if bestalgorithm == 'KNeighbors':
    print('Best params is :', knn_cv.best_params_)
if bestalgorithm == 'LogisticRegression':
    print('Best params is :', logreg_cv.best_params_)
if bestalgorithm == 'SupportVector':
    print('Best params is :', svm_cv.best_params_)
```

```
Best model is DecisionTree with a score of 0.9017857142857142
Best params is : {'criterion': 'gini', 'max_depth': 16, 'max_features': 'auto', 'min_samples_leaf': 4, 'min_samples_split': 10, 'splitter': 'random'}
```

Confusion Matrices

Performance Summary:

The confusion matrices for the classification algorithm were identical. While having false positives (Type 1 errors) is not desirable, they were present in the results. Here's a breakdown of the confusion matrix outputs:



- True Positive (TP): 12
- True Negative (TN): 3
- False Positive (FP): 3
- False Negative (FN): 0

Using these values, we can calculate the following metrics:

- Precision = TP / (TP + FP) = 12 / (12 + 3) = 0.80
- Recall = TP / (TP + FN) = 12 / (12 + 0) = 1
- F1 Score = 2 * (Precision * Recall) / (Precision + Recall) = 2 * (.80 * 1) / (.80 + 1) = 0.89
- Accuracy = (TP + TN) / (TP + TN + FP + FN) = (12 + 3) / (12 + 3 + 3 + 0) = 0.833

Conclusion

Research Summary:

1. Model Performance: The models performed similarly on the test set, with the decision tree model slightly outperforming others.

2. Equator: Most launch sites are situated near the equator to leverage Earth's rotational speed, providing a natural boost that reduces the need for extra fuel and boosters.

3. Coast: All launch sites are located close to the coast.

4. Launch Success: The success rate of launches has shown an increasing trend over time.

5. KSC LC-39A: This launch site has the highest success rate among all sites, with a 100% success rate for launches under 5,500 kg.

6. Orbits: Launches to ES-L1, GEO, HEO, and SSO orbits have achieved a 100% success rate.

7. Payload Mass: Across all launch sites, there is a positive correlation between payload mass and success rate; higher payload masses tend to result in higher success rates.

Conclusion

Things to Consider:

1. Dataset: Increasing the dataset size can enhance the robustness of predictive analytics results and assess the generalizability of findings to a broader dataset.

2. Feature Analysis / PCA: Conducting additional feature analysis or principal component analysis (PCA) may help improve accuracy by identifying relevant features or reducing dimensionality.

3. XGBoost: Utilizing XGBoost, a powerful model not used in the study, could provide insights into its performance compared to other classification models and potentially enhance predictive accuracy.