IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Jon Neubauer Rivera
06/24/2024

Outline

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

Executive Summary

- **Objective:** Create a data-driven model to predict SpaceX rocket landing success, leveraging the power of machine learning and data science techniques.
- **Methods:** Employed a systematic process encompassing diverse data sources, comprehensive data cleaning and transformation, rigorous exploratory data analysis, interactive visualization tools, and the development of multiple machine learning models.
- **Key Findings:** Discovered significant correlations between specific mission parameters and landing outcomes. The final predictive models achieved impressive accuracy, offering a reliable tool for assessing risk and optimizing mission planning for future SpaceX endeavors.

Introduction

- **Welcome to our project focused on predicting the success of SpaceX rocket landings.** In an era of rapid advancements in space exploration, understanding the factors that influence successful landings is paramount. SpaceX, an industry leader, has made significant strides with the Falcon 9 rocket. Our ability to predict the outcome of future Falcon 9 landings has far-reaching implications, not only for operational efficiency but also for improving safety and reducing costs.
- **Data Sources and Methods:** To achieve our goal, we've combined a diverse range of data sources and cutting-edge analytical methods. The foundation of our analysis is data obtained from the SpaceX REST API, which provides a wealth of information on past launches, including critical details on launch specifications and landing outcomes. We've further enriched our dataset by utilizing web scraping techniques to extract additional Falcon 9 launch records from relevant Wiki pages.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
    - Data from SpaceX was obtained from two sources:
        - SpaceX API(https://api.spacexdata.com/v4/rockets/)
        - Web Scraping (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches)
- Perform data wrangling
    - Collected data was enriched by creating a landing outcome label based on outcome data after summarizing and analyzing features
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

Methodology

Executive Summary

- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
    - Data that was collected until this step were normalized, divided in training and test data sets and evaluated by four different classification models, being the accuracy of each model evaluated using different combinations of parameters.

Data Collection

Source: SpaceX REST API (`api.spacexdata.com/v4/launches/past`)

Tools: `requests` (API calls), `pandas` (data handling)

Key Steps:

- Converted JSON response to DataFrame
- Filtered out Falcon 1 launches
- Imputed missing payload mass with mean

Data Collection – SpaceX API

API Request	Filter Data	Handle Missing Data
A programmatic request was sent to the SpaceX API to retrieve a comprehens ive dataset detailing their launch history. This initial dataset included information about all SpaceX launches, regardless of rocket type.	The retrieved dataset was refined to isolate only the launches conducted using the Falcon 9 rocket. This step was essential to focus the analysis on a specific vehicle and its performance characteristics.	A thorough assessment of the Falcon 9 launch data was performed to identify any missing or incomplete values. These gaps were then addressed through appropriate imputation techniques or, if necessary, by excluding specific data points, ensuring the reliability of the dataset for subsequent analysis and modeling.

https://github.com/jneubauer/ibm-data-science-cert/blob/main/Capstone/Data%20Collection%20API.ipynb

Data Collection – Scraping

Request Wiki Page	Extract from HTML	Parse to DataFrame
An HTTP GET request was sent to the Wikipedia page listing Falcon 9 launches, retrieving the page's HTML content for further processing.	The HTML content was parsed to identify the table containing launch data. The header row of this table was then extracted, providing the column names that would define the structure of the dataset.	The launch data within the HTML table was systematically extracted and organized into a structured format, resulting in a Pandas DataFrame. This DataFrame provided a convenient and efficient way to analyze and manipulate the launch data for subsequent analysis.

https://github.com/jneubauer/ibm-data-science-cert/blob/main/Capstone/Data%20Collection%20with%20Web%20Scraping.ipynb

Data Wrangling

Initial Exploratory Analysis → **Calculate SUmmary Statistics** → **Create Landing Outcome Label**

A preliminary exploration of the Falcon 9 launch dataset was conducted to gain insights into its structure, patterns, and potential relationships. This included examining descriptive statistics, distributions of key variables, and identifying potential outliers or anomalies.

Aggregate calculations were performed to summarize launches per site, occurrences of each orbit type, and mission outcomes by orbit type. This step provided a higher-level view of launch patterns and potential correlations.

A new categorical variable, "Landing Outcome," was derived from the existing "Outcome" column. This involved categorizing the various outcome descriptions into distinct landing success or failure labels, simplifying analysis and visualization.

https://github.com/jneubauer/ibm-data-science-cert/blob/main/Capstone/Data%20Wrangling.ipynb

EDA with SQL

- **Launch Site Diversity**: Identified all unique launch sites used for SpaceX missions.
- **Top CCA Launch Sites**: Determined the 5 most frequently used launch sites starting with "CCA".
- **NASA Payload Capacity**: Calculated the total payload mass transported by SpaceX for NASA's Commercial Resupply Services (CRS).
- **F9 v1.1 Performance**: Analyzed the average payload mass carried by the Falcon 9 v1.1 booster.
- **Ground Pad Landing Milestone**: Pinpointed the date of the first successful ground pad landing, a significant achievement in reusability.
- **Drone Ship Success & Payload Range**: Identified boosters successfully landing on drone ships with payloads between 4,000kg and 6,000kg.
- **Mission Outcomes Overview**: Summarized the total number of successful and failed missions, revealing the overall success rate.
- **Boosters with Highest Payload**: Highlighted the booster versions capable of carrying the maximum payload mass.
- **2015 Drone Ship Landing Challenges**: Examined failed drone ship landings in 2015, focusing on booster versions and launch sites.
- **Landing Outcome Trends (2010-2017)**: Ranked the frequency of different landing outcomes (e.g., drone ship failure, ground pad success) over a specific time period, revealing trends in landing technology.

https://github.com/jneubauer/ibm-data-science-cert/blob/main/Capstone/EDA%20with%20SQL.ipynb

Build an Interactive Map with Folium

- **Precise Location Pinpointing (Markers):** Markers precisely identify key locations like launch sites, providing immediate geographical context.
- **Highlighting Areas of Interest (Circles):** Circles emphasize specific regions, such as the NASA Johnson Space Center, visually connecting them to the broader launch data.
- **Clustering for Density Visualization (Marker Clusters):** Marker clusters reveal launch density at specific sites, simplifying the display of numerous launches in a single location.
- **Measuring Distances (Lines):** Lines illustrate distances between points of interest, potentially revealing correlations between launch locations and outcomes.

Build a Dashboard with Plotly Dash

- **Site-Specific Launch Distribution (Pie Charts):** Interactive pie charts reveal the proportion of launches conducted at each site, offering a quick comparison of site usage.

- **Outcome vs. Payload Mass Relationship (Scatter Plots):** Scatter plots, segmented by booster version, dynamically illustrate how payload mass may influence mission outcome, allowing for in-depth analysis and pattern identification.

Predictive Analysis (Classification)

- **Data Preparation and Transformation:** Utilizing NumPy and Pandas, we loaded, cleaned, and transformed the SpaceX launch dataset to prepare it for modeling. The data was then split into training and testing sets.

- **Model Selection and Hyperparameter Tuning:** A range of machine learning models were evaluated using GridSearchCV to identify optimal hyperparameters, ensuring a thorough exploration of potential model configurations.

- **Model Refinement and Optimization:** Model performance was iteratively improved through feature engineering and algorithm tuning, focusing on maximizing accuracy as the primary metric.

- **Best Model Selection:** The best-performing classification model was identified based on its accuracy on the testing set, demonstrating its ability to predict launch outcomes effectively.

Results

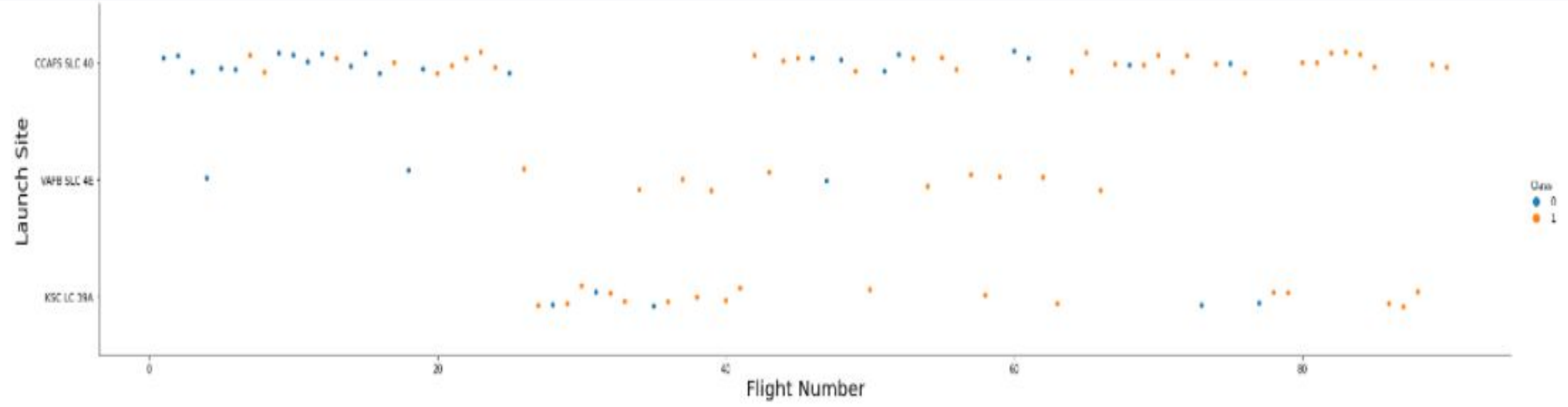
- **Operational Diversity:** SpaceX utilizes four distinct launch sites for its missions.
- **Early Partners:** Initial launches primarily served SpaceX's own needs and NASA missions.
- **Payload Efficiency:** The Falcon 9 v1.1 booster boasts an average payload capacity of 2,928 kg.
- **Landmark Reusability:** A major milestone was achieved in 2015 with the first successful landing, five years after the inaugural launch.
- **Payload & Landing Success Correlation:** A significant number of Falcon 9 boosters successfully landed on drone ships when carrying above-average payloads.
- **Mission Success Rate:** SpaceX has demonstrated an impressive mission success rate, approaching 100%.
- **Early Landing Challenges:** Two specific Falcon 9 v1.1 booster versions faced challenges with drone ship landings in 2015.
- **Continuous Improvement in Landing:** Landing outcomes have shown a positive trend over time, indicating advancements in SpaceX's landing technologies.

Section 2

Insights drawn from EDA

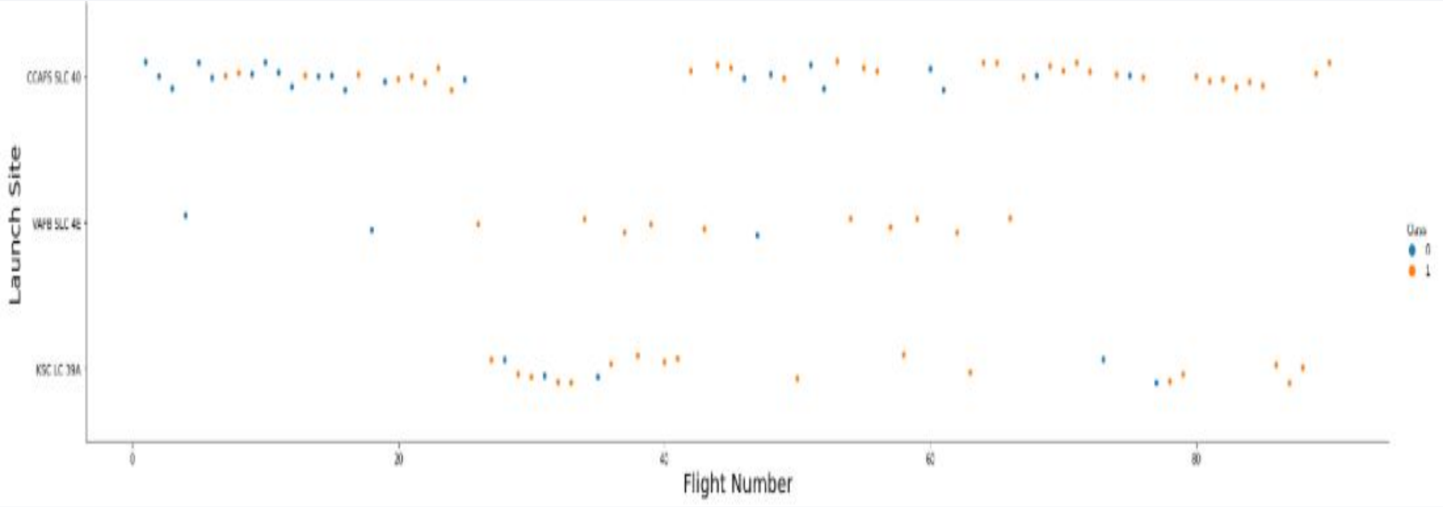
Flight Number vs. Launch Site

Our analysis reveals a strong positive correlation between the number of flights launched from a site and the success rate at that site. This suggests that experience and operational familiarity play a significant role in achieving successful launches.



Payload vs. Launch Site

At CCAFS SLC 40, launching heavier payloads appears to enhance mission success rates. This insight could be leveraged for strategic planning and resource allocation at this specific launch site.



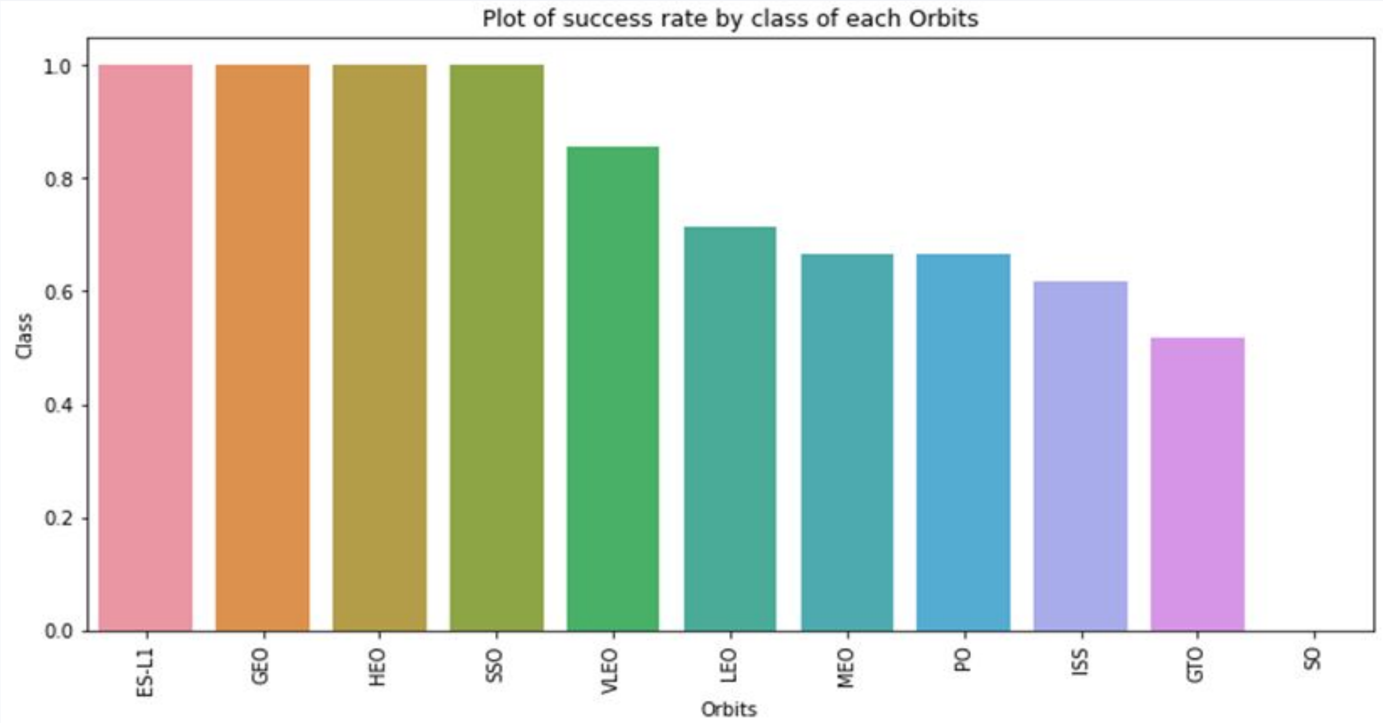
Success Rate vs. Orbit Type

- Show a bar chart for the success rate of each orbit type

- Show the screenshot of the scatter plot with explanations

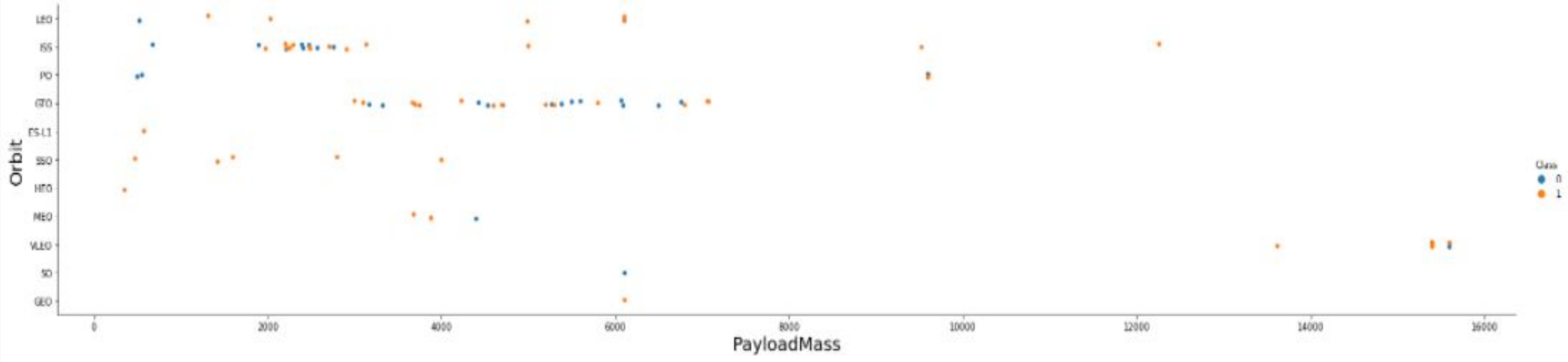
Flight Number vs. Orbit Type

Our analysis reveals that certain orbit types, specifically ES-L1, GEO, HEO, SSO, and VLEO, consistently demonstrate the highest success rates for SpaceX missions. This suggests that these orbits may present more favorable conditions or less inherent risk compared to others.



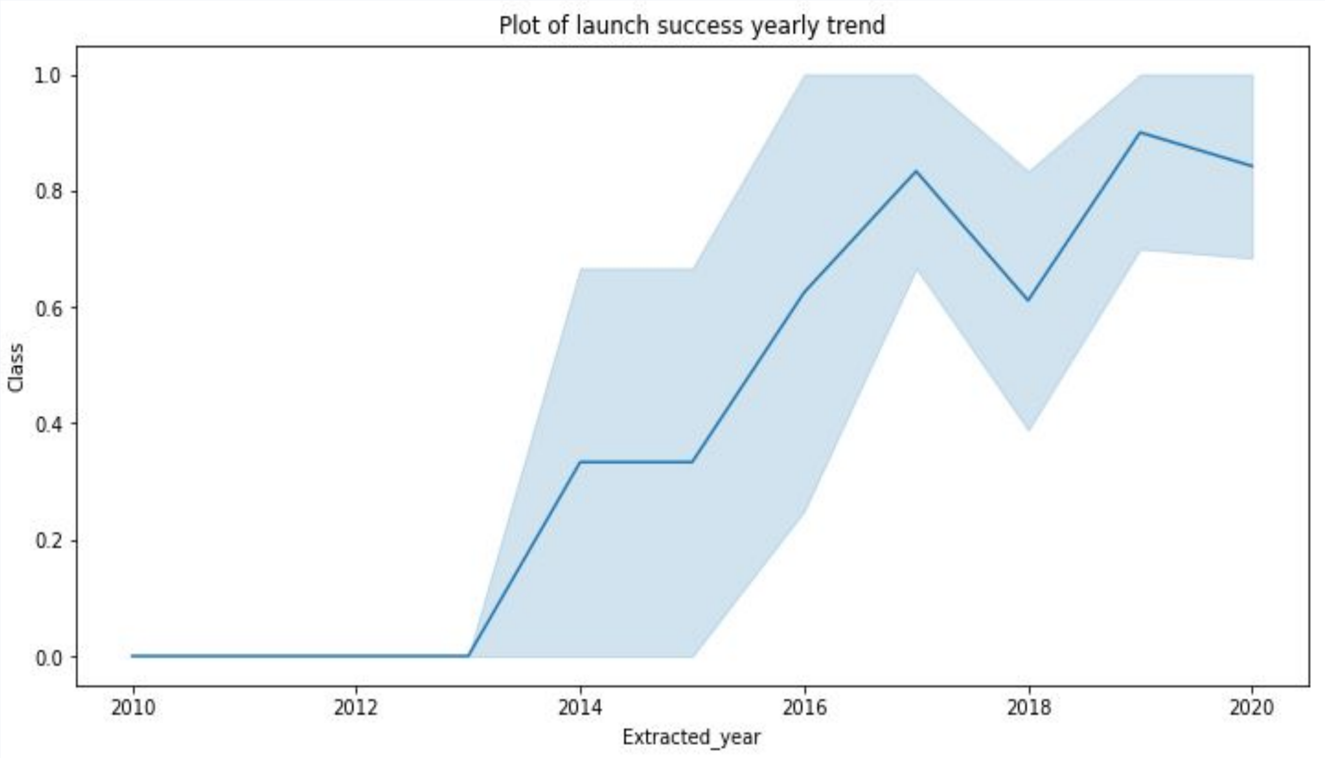
Payload vs. Orbit Type

Our analysis reveals that heavier payloads are more frequently associated with successful landings in PO, LEO, and ISS orbits. This suggests that these orbits may offer advantageous conditions or mission profiles for managing the complexities of heavier payloads.



Launch Success Yearly Trend

Our analysis reveals a remarkable upward trend in launch success rates from 2013 to 2020, demonstrating SpaceX's continuous improvement and commitment to reliability.



All Launch Site Names

By leveraging the DISTINCT keyword in our SQL query, we extracted the complete list of unique locations from which SpaceX has launched missions, showcasing the breadth of their global operations.

Display the names of the unique launch sites in the space mission

In [10]:
```
task_1 = '''
        SELECT DISTINCT LaunchSite
        FROM SpaceX
'''
create_pandas_df(task_1, database=conn)
```

Out[10]:

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

Launch Site Names Begin with 'CCA'

We identified five unique SpaceX launch sites with names beginning with "CCA", all located at Cape Canaveral Air Force Station in Florida. This highlights the importance of this location for SpaceX's launch operations.

Display 5 records where launch sites begin with the string 'CCA'

In [11]:

```
task_2 = '''
        SELECT *
        FROM SpaceX
        WHERE LaunchSite LIKE 'CCA%'
        LIMIT 5
        '''
create_pandas_df(task_2, database=conn)
```

Out[11]:

	date	time	boosterversion	launchsite	payload	payloadmasskg	orbit	customer	missionoutcome	landingoutcome
0	2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
1	2010-08-12	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of...	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2	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
3	2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
4	2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

Our analysis reveals that SpaceX boosters have transported a total of 45,596 kilograms of payload for NASA missions, demonstrating a significant contribution to the agency's scientific and logistical endeavors.

Display the total payload mass carried by boosters launched by NASA (CRS)

In [12]:
```
task_3 = '''
        SELECT SUM(PayloadMassKG) AS Total_PayloadMass
        FROM SpaceX
        WHERE Customer LIKE 'NASA (CRS)'
        '''
create_pandas_df(task_3, database=conn)
```

Out[12]:

	total_payloadmass
0	45596

Average Payload Mass by F9 v1.1

Our analysis reveals that the average payload mass carried by the Falcon 9 v1.1 booster is 2,928.4 kilograms, offering valuable insights into this specific version's capabilities and performance.

Display average payload mass carried by booster version F9 v1.1

In [13]:

```
task_4 = '''
        SELECT AVG(PayloadMassKG) AS Avg_PayloadMass
        FROM SpaceX
        WHERE BoosterVersion = 'F9 v1.1'
        '''
create_pandas_df(task_4, database=conn)
```

Out[13]:

	avg_payloadmass
0	2928.4

First Successful Ground Landing Date

On December 22, 2015, SpaceX achieved a groundbreaking milestone with the first successful landing of a Falcon 9 rocket on a ground pad, marking a major step towards reusable spaceflight technology.

```
In [14]:
task_5 = '''
        SELECT MIN(Date) AS FirstSuccessfull_landing_date
        FROM SpaceX
        WHERE LandingOutcome LIKE 'Success (ground pad)'
        '''
create_pandas_df(task_5, database=conn)
```

Out[14]:

	firstsuccessfull_landing_date
0	2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

Several SpaceX boosters have successfully landed on drone ships while carrying payloads between 4,000kg and 6,000kg, demonstrating the versatility and precision of their landing technology for a range of mission requirements.

In [15]:

```
task_6 = '''
        SELECT BoosterVersion
        FROM SpaceX
        WHERE LandingOutcome = 'Success (drone ship)'
            AND PayloadMassKG > 4000
            AND PayloadMassKG < 6000
        '''
create_pandas_df(task_6, database=conn)
```

Out[15]:

	boosterversion
0	F9 FT B1022
1	F9 FT B1026
2	F9 FT B1021.2
3	F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

Our analysis of SpaceX missions reveals a remarkable trend: the vast majority of missions have resulted in success, highlighting SpaceX's impressive track record in spaceflight operations.

List the total number of successful and failure mission outcomes

In [16]:

```
task_7a = '''
        SELECT COUNT(MissionOutcome) AS SuccessOutcome
        FROM SpaceX
        WHERE MissionOutcome LIKE 'Success%'
        '''

task_7b = '''
        SELECT COUNT(MissionOutcome) AS FailureOutcome
        FROM SpaceX
        WHERE MissionOutcome LIKE 'Failure%'
        '''
print('The total number of successful mission outcome is:')
display(create_pandas_df(task_7a, database=conn))
print()
print('The total number of failed mission outcome is:')
create_pandas_df(task_7b, database=conn)
```

```
The total number of successful mission outcome is:
```

	successoutcome
0	100

```
The total number of failed mission outcome is:
```

Out[16]:

	failureoutcome
0	1

Boosters Carried Maximum Payload

Our analysis reveals that the Falcon Heavy booster holds the record for the highest payload mass carried by any SpaceX launch vehicle, showcasing its exceptional capacity for ambitious missions.

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

In [17]:

```
task_8 = '''
        SELECT BoosterVersion, PayloadMassKG
        FROM SpaceX
        WHERE PayloadMassKG = (
                                SELECT MAX(PayloadMassKG)
                                FROM SpaceX
                                )
        ORDER BY BoosterVersion
        '''
create_pandas_df(task_8, database=conn)
```

Out[17]:

	boosterversion	payloadmasskg
0	F9 B5 B1048.4	15600
1	F9 B5 B1048.5	15600
2	F9 B5 B1049.4	15600
3	F9 B5 B1049.5	15600
4	F9 B5 B1049.7	15600
5	F9 B5 B1051.3	15600
6	F9 B5 B1051.4	15600
7	F9 B5 B1051.6	15600
8	F9 B5 B1056.4	15600
9	F9 B5 B1058.3	15600
10	F9 B5 B1060.2	15600
11	F9 B5 B1060.3	15600

2015 Launch Records

In 2015, SpaceX encountered challenges with specific Falcon 9 booster versions attempting drone ship landings. These challenges, primarily occurring at Cape Canaveral, highlight a key phase in the development of SpaceX's reusable launch technology.

List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

In [18]:

```
task_9 = '''
        SELECT BoosterVersion, LaunchSite, LandingOutcome
        FROM SpaceX
        WHERE LandingOutcome LIKE 'Failure (drone ship)'
            AND Date BETWEEN '2015-01-01' AND '2015-12-31'
        '''
create_pandas_df(task_9, database=conn)
```

Out[18]:

	boosterversion	launchsite	landingoutcome
0	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
1	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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

Analysis of SpaceX landing outcomes between June 2010 and March 2017 reveals a clear trend: While both successful and failed landings occurred, failures were slightly more frequent during this early period of landing technology development.

Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad))

In [19]:

```
task_10 = '''
        SELECT LandingOutcome, COUNT(LandingOutcome)
        FROM SpaceX
        WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20'
        GROUP BY LandingOutcome
        ORDER BY COUNT(LandingOutcome) DESC
        '''
create_pandas_df(task_10, database=conn)
```

Out[19]:

	landingoutcome	count
0	No attempt	10
1	Success (drone ship)	6
2	Failure (drone ship)	5
3	Success (ground pad)	5
4	Controlled (ocean)	3
5	Uncontrolled (ocean)	2
6	Precluded (drone ship)	1
7	Failure (parachute)	1

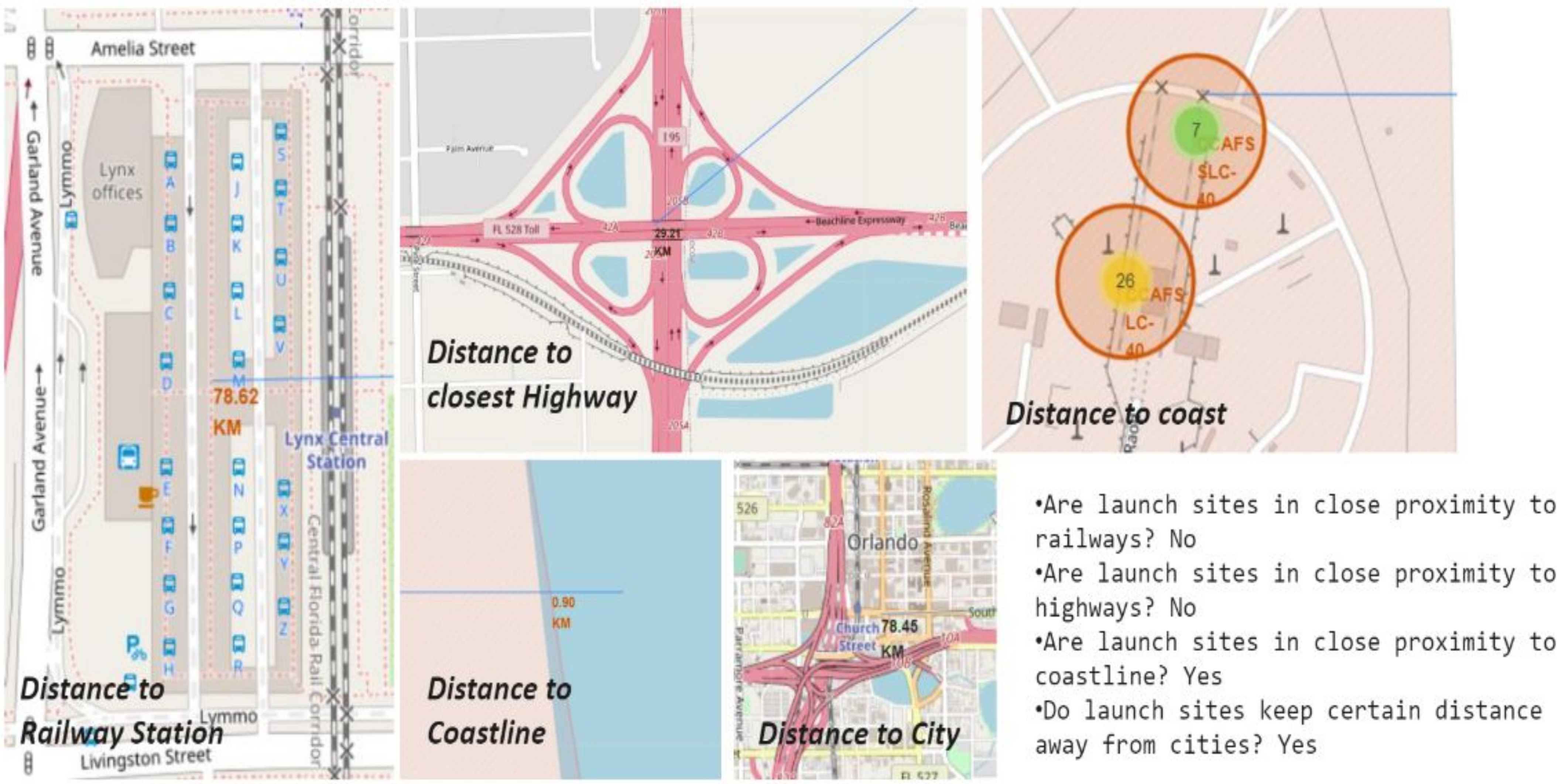
Section 3

Launch Sites Proximities Analysis

Visualizing SpaceX Launch Sites Around the World

Global Launch Locations: A Color-Coded Guide to SpaceX Activity

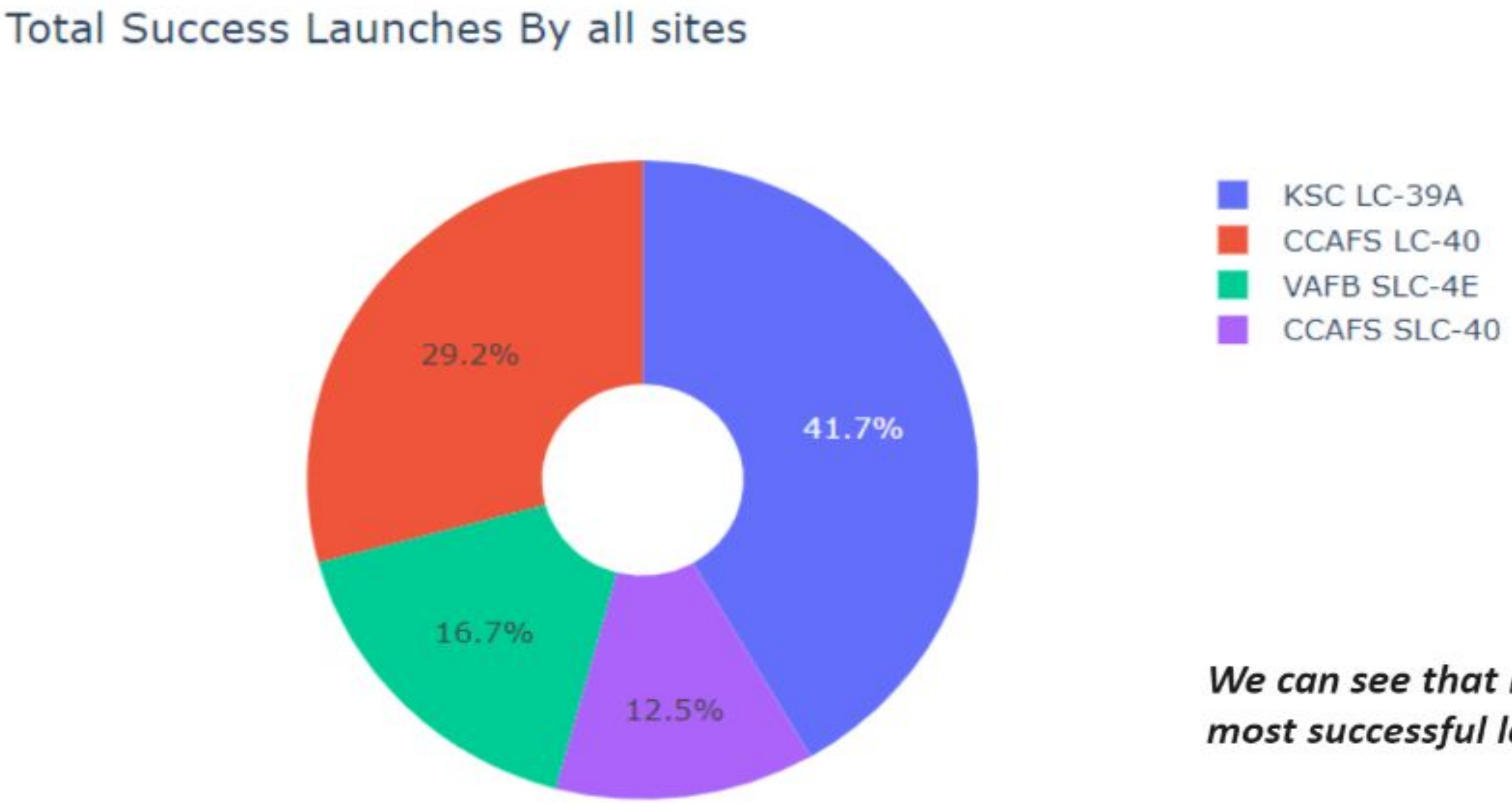
Visualizing Launch Site Geography: Proximity to Landmarks



Section 4

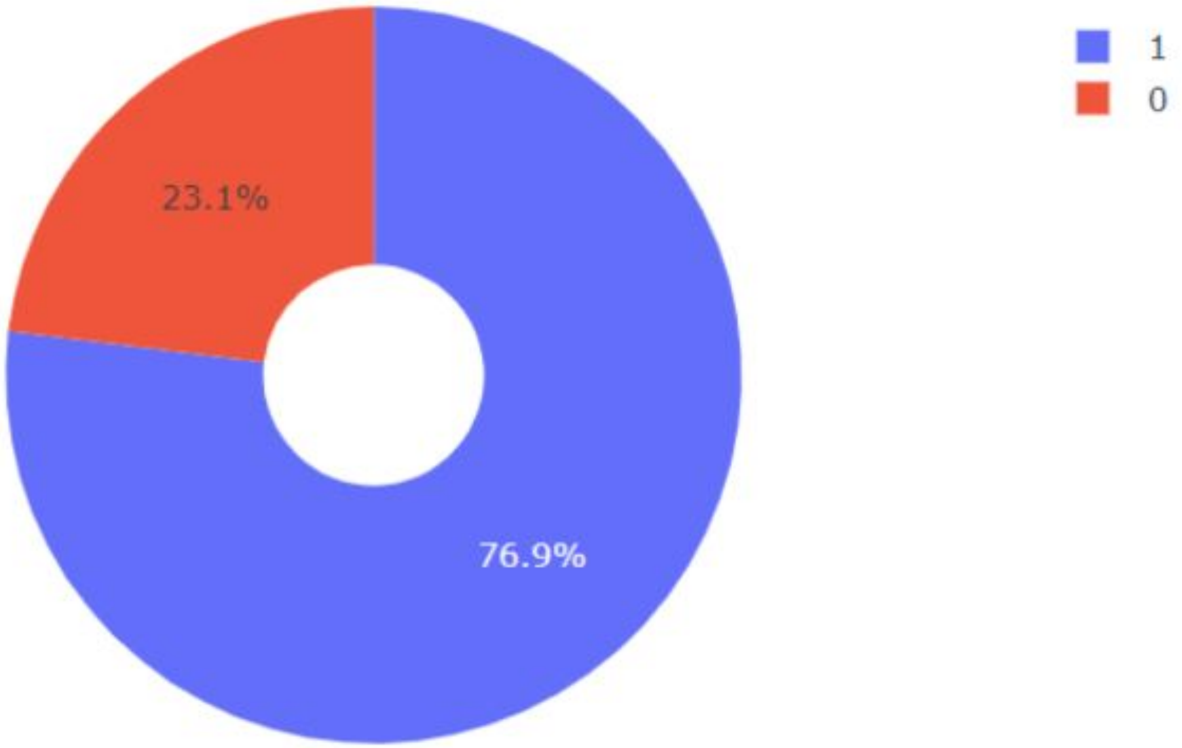
Build a Dashboard with Plotly Dash

SpaceX Launch Site Success Rates: A Visual Comparison



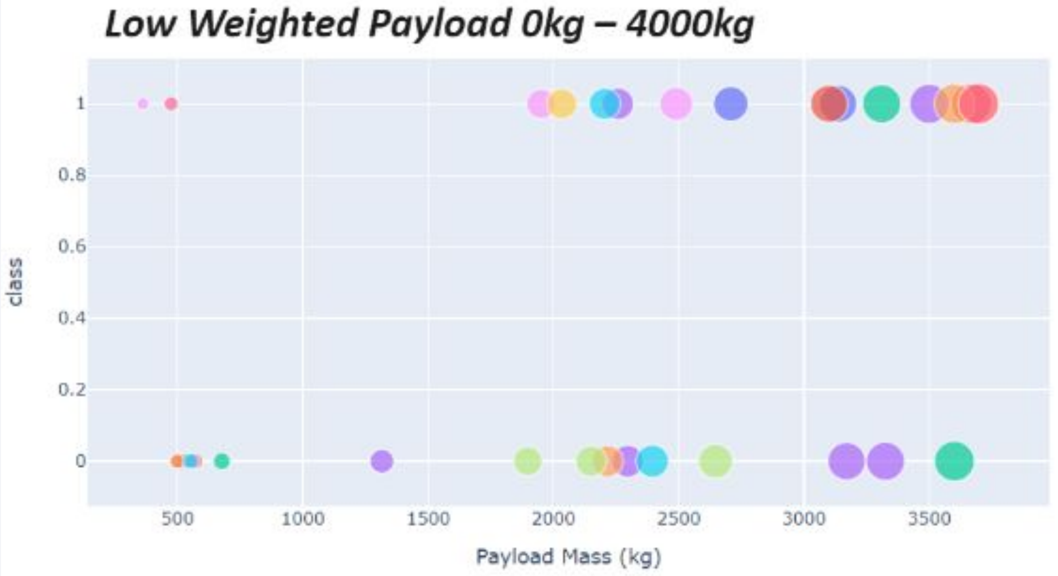
We can see that KSC LC-39A had the most successful launches from all the sites

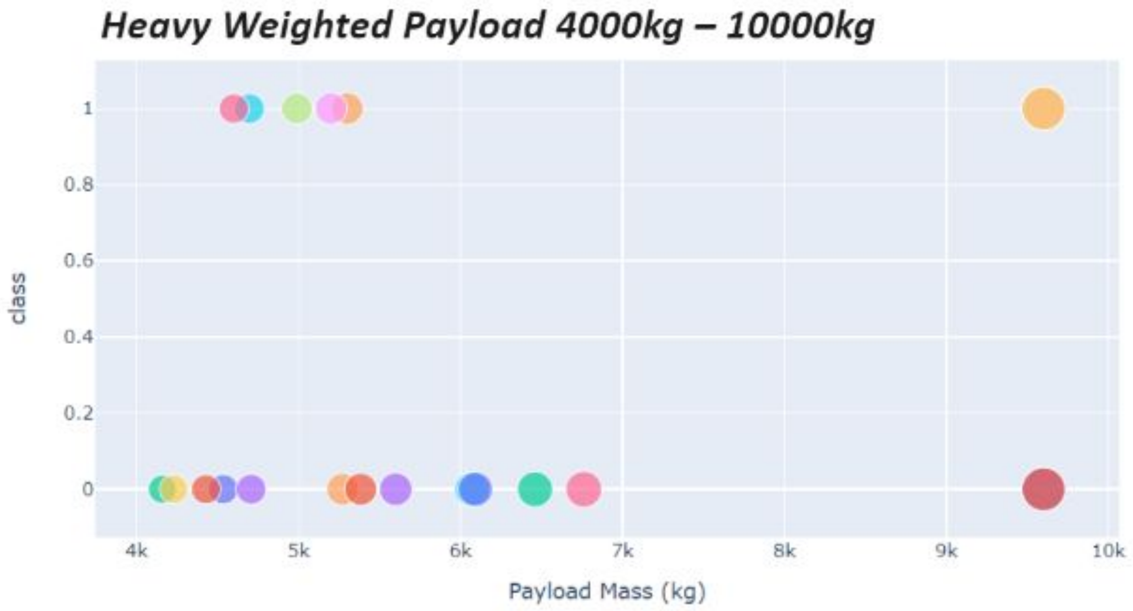
Launch Site Champion: Unveiling the Success Leader



KSC LC-39A achieved a 76.9% success rate while getting a 23.1% failure rate

Exploring Payload vs. Outcome: An Interactive Scatter Plot





We can see the success rates for low weighted payloads is higher than the heavy weighted payloads

Section 5

Predictive Analysis (Classification)

Classification Accuracy

Our analysis revealed the decision tree classifier as the top-performing model, achieving the highest classification accuracy among all tested algorithms.

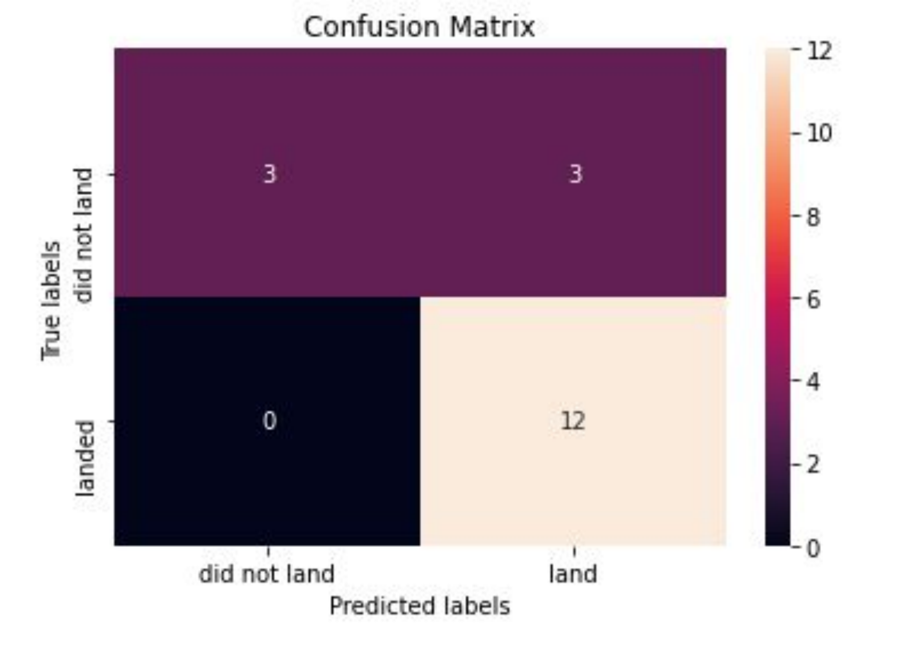
```
models = {'KNeighbors':knn_cv.best_score_,
              'DecisionTree':tree_cv.best_score_,
              'LogisticRegression':logreg_cv.best_score_,
              'SupportVector': svm_cv.best_score_}

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 0.8732142857142856
Best params is : {'criterion': 'gini', 'max_depth': 6, 'max_features': 'auto', 'min_samples_leaf': 2, 'min_samples_split': 5, 'splitter': 'random'}
```

Confusion Matrix

The decision tree classifier effectively distinguishes between landing outcomes,
but a key area for improvement is the rate of false positives, where unsuccessful landings are mistakenly predicted as successful.



Conclusions

- **Launch Site Experience Drives Success:** A higher number of launches at a site significantly correlates with improved mission success rates.
- **Success Trajectory:** SpaceX launch success rates have steadily increased from 2013 to 2020, highlighting a consistent upward trend in performance.
- **Optimal Orbits for Success:** ES-L1, GEO, HEO, SSO, and VLEO orbits have proven to be the most successful for SpaceX missions.
- **Launch Site Leader:** KSC LC-39A boasts the highest number of successful launches among all SpaceX launch sites.
- **Predicting Success:** The Decision Tree classifier emerged as the most effective machine learning model for accurately predicting SpaceX launch outcomes.

Thank you!