IBM Developer
SKILLS NETWORK

Revolutionizing Space Launch Economics: Predicting Falcon 9 First Stage Reusability with Data Science

Elísio Henrique
Borges dos Santos
Souto
26/11/2023

Outline

Executive Summary

Introduction

Methodology

Results

Conclusion

Appendix

Executive Summary

Summary of Methodologies

- Data Collection
- Data Wrangling
- Exploratory Data Analysis (EDA) with Data Visualization
- EDA with SQL
- Building an Interactive Map with Folium
- Building a Dashboard with Plotly Dash:
- Predictive Analysis (Classification)

Summary of All Results

- **EDA Results**: Findings from the exploration of data patterns, correlations, and outliers.
- **Interactive Analytics Demo in Screenshots**: Visual representations capturing the interactive analytics experience.
- **Predictive Analysis Results**: Outcomes and insights derived from predictive modeling, specifically in the context of classification.

Introduction

Project background and context

In the realm of space travel economics, SpaceX has emerged as a trailblazer. Their Falcon 9 rocket launches are prominently featured on their website at a cost of $62 million, a stark contrast to competitors' prices exceeding $165 million. A key factor in their cost efficiency is the reuse of the first stage. This project delves into predicting the success of Falcon 9 first stage landings and estimating launch costs. By leveraging public information and advanced machine learning models, our goal is to provide practical insights for companies navigating the competitive landscape of space launch contracts.

Questions to be answered:

- How do variables such as payload mass, launch site, number of flights, and orbits affect the success of the first stage landing?

- Does the rate of successful landings increase over the years?

- What is the best algorithm that can be used for binary classification in this case?

Section 1

Methodology

Methodology

Data Collection Methodology

- SpaceX Rest API
- Webscraping from Wikipedia

Data Wrangling

- Filtering the data
- Dealing with missing values
- Using One Hot Encoding to prepare the data for binary classification

Performed exploratory data analysis (EDA) using visualization and SQL

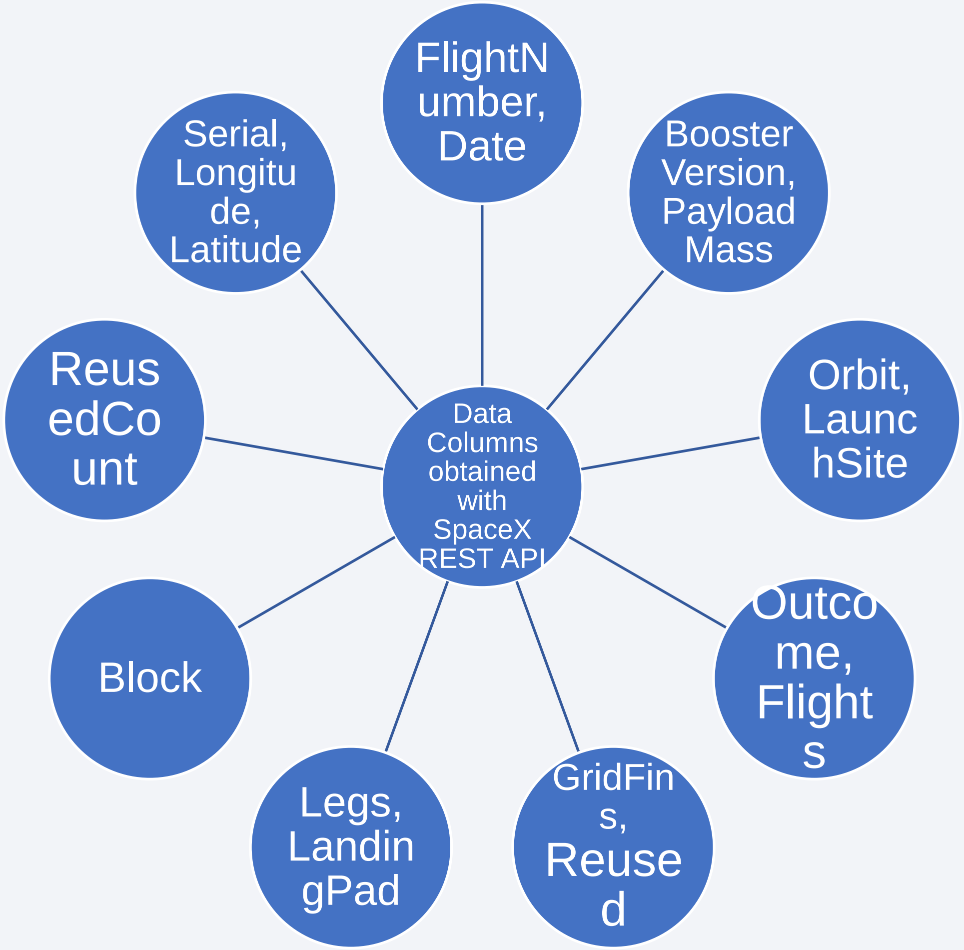
Performed interactive visual analytics using Folium and Plotly Dash

Performed predictive analysis using classification models

Building, tuning and evaluation of classification models to ensure the best results

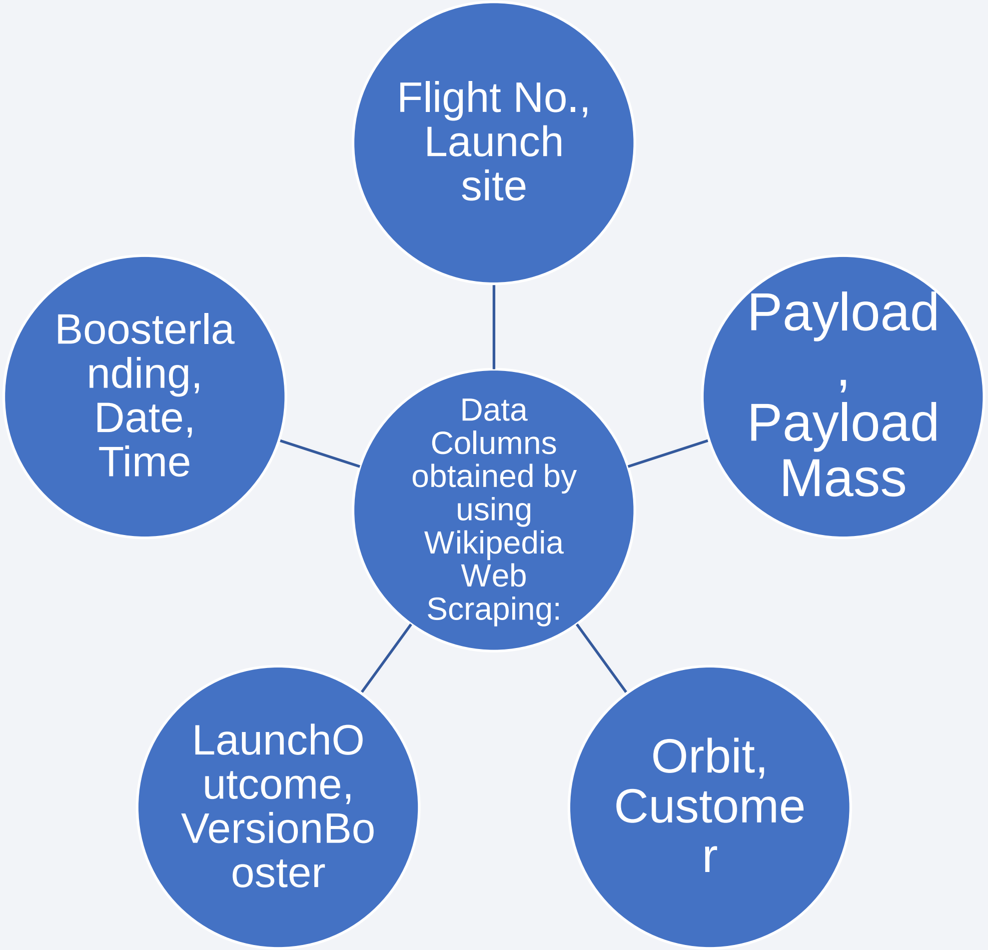
Data Collection

Data collection involved a dual approach, combining API requests from SpaceX REST API with web scraping from a table in SpaceX's Wikipedia entry. This ensured comprehensive information retrieval for a detailed analysis of the launches.



Data Collection

Data collection involved a dual approach, combining API requests from SpaceX REST API with web scraping from a table in SpaceX's Wikipedia entry. This ensured comprehensive information retrieval for a detailed analysis of the launches.



Data Collection – SpaceX API

- Retrieved rocket launch data from SpaceX API
- Decoded the response content using .json() and transformed it into a dataframe using .json_normalize()
- Extracted necessary launch information from SpaceX API through custom functions
- Organized obtained data into a dictionary
- Created a dataframe from the dictionary
- Filtered the dataframe to exclusively include Falcon 9 launches
- Addressed missing values in the Payload Mass column by replacing them with the calculated mean for this column
- Exported the data to a CSV file

[GitHub URL: Data Collection - API]()

Data Collection – Web Scraping

1. Retrieved rocket launch data from SpaceX API
2. Requested Falcon 9 launch data from Wikipedia
3. Generated a BeautifulSoup object from the HTML response
4. Extracted all column names from the HTML table header
5. Collected the data by parsing HTML tables
6. Organized obtained data into a dictionary
7. Created a dataframe from the dictionary
8. Exported the data to a CSV file

GitHub URL: Data Collection - Web Scraping

Data Wrangling

In this lab, Exploratory Data Analysis (EDA) is conducted to identify patterns in the data and establish labels for training supervised models. The dataset includes various cases where booster landings were unsuccessful, differentiated by outcomes such as True Ocean, False Ocean, True RTLS, False RTLS, True ASDS, and False ASDS, each signifying different landing scenarios. The primary objective is to convert these outcomes into Training Labels, where 1 indicates a successful booster landing, and 0 denotes an unsuccessful landing.

Data Wrangling

Import the dataset containing information about booster landing outcomes

Explore and identify different landing scenarios, including True Ocean, False Ocean, True RTLS, False RTLS, True ASDS, and False ASDS

Convert the varied outcomes into Training Labels, assigning 1 for successful landings and 0 for unsuccessful ones

Conduct EDA to analyze patterns and gain insights into the dataset

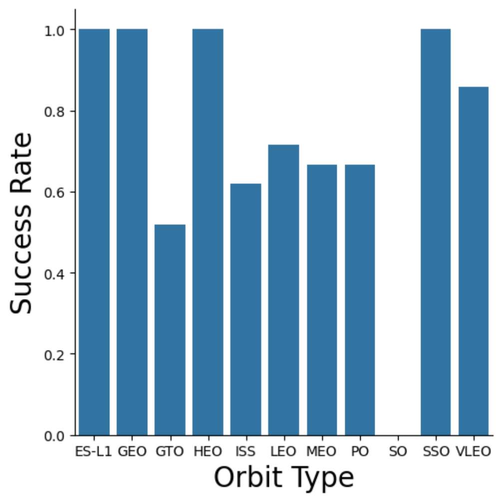
Finalize the Training Labels based on the EDA findings

EDA with Data Visualization

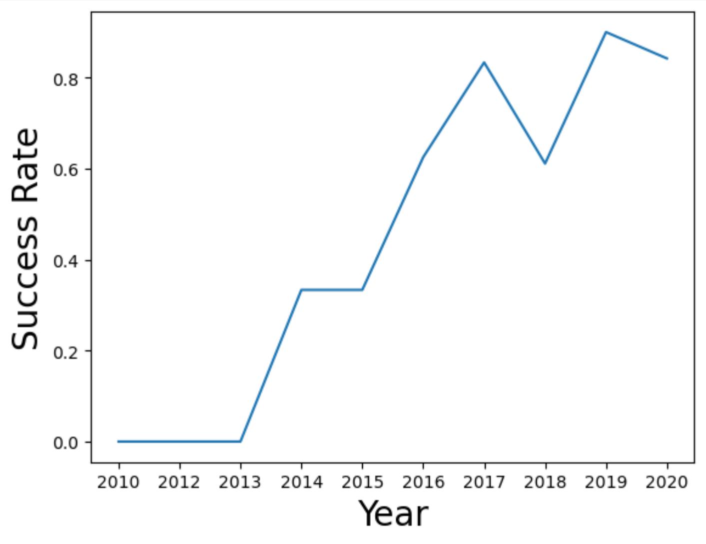
Various charts were plotted in the analysis, including Flight Number vs. Payload Mass, Flight Number vs. Launch Site, Payload Mass vs. Launch Site, Orbit Type vs. Success Rate, Flight Number vs. Orbit Type, Payload Mass vs. Orbit Type, and Success Rate Yearly Trend. These charts were chosen to visually explore relationships and patterns within the data. Scatter plots were employed to examine correlations between variables, bar charts to compare categorical data, line charts to observe trends over time, and collectively, they provided a comprehensive visual representation of key insights and trends in the dataset.

[GitHub URL: Exploratory Data Analysis – Data Visualization]

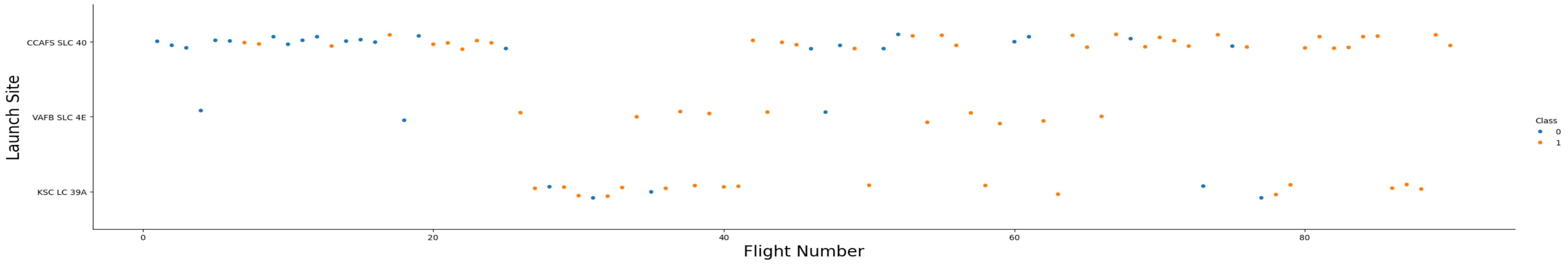
EDA with Data Visualization



1.1. Bar chart - Orbit Type vs Success Rate



1.2. Line chart – Year vs Success Rate



1.3. Scatter plot - Flight Number vs Launch Size

EDA with SQL

- Displayed unique launch site names in the space mission.
- Displayed 5 records where launch sites begin with the string ‘CCA.'
- Displayed the total payload mass carried by NASA (CRS) boosters.
- Displayed the average payload mass carried by booster version F9 v1.1.
- Listed the date of the first successful ground pad landing outcome.
- Listed names of boosters with success on a drone ship and payload mass between 4000 and 6000.
- Listed the total number of successful and failed mission outcomes.
- Listed the names of booster versions carrying the maximum payload mass.
- Listed failed landing outcomes on a drone ship, along with booster versions and launch site names for 2015.
- Ranked the count of landing outcomes between June 4, 2010, and March 20, 2017, in descending order.

GitHub URL: Exploratory Data Analysis with SQL

Build an Interactive Map with Folium

Markers

- Placed on the map to denote the locations of selected launch sites
- Different colors were used to represent the success and failure of launches at each site

Circles

- Drawn around the launch sites to signify different proximities (such as railways, highways, coastline, cities)
- Serve as visual boundaries, highlighting the spatial relationships of launch sites to key infrastructure

Lines

- Utilized to connect launch sites with specific points on the map, representing calculated distances
- Provide a clear visual representation of the measured distances from launch sites to critical infrastructures

Build a Dashboard with Plotly Dash

Launch Sites Dropdown List:

- Facilitates user-friendly launch site selection.
- Enables dynamic data filtering based on launch site preferences, enhancing user exploration.

Pie Chart for Success Launches:

- Illustrates total successful launches across all sites.
- Offers a quick overview of success distribution; dynamic adjustment aids in site-specific analysis.

Payload Mass Range Slider:

- Allows selection of payload mass range.
- Empowers users to filter launches based on payload criteria, providing focused analytical insights.

Scatter Chart of Payload Mass vs. Success Rate:

- Visualizes correlation between payload mass and launch success.
- Provides an insightful representation of how payload mass influences success across booster versions.

GitHub URL: SpaceX Dash App

Predictive Analysis (Classification)

Extracted the "Class" column from the dataset and created a NumPy array

Applied StandardScaler to standardize the data, enhancing model performance

Utilized the train_test_split function to divide the dataset into training and testing sets

Created a GridSearchCV object with 10-fold cross-validation to identify optimal model parameters

Implemented GridSearchCV on Logistic Regression, SVM, Decision Tree, and KNN models

Calculated accuracy on the test data using the .score() method for each model

Examined the confusion matrix for a comprehensive understanding of model performance

Evaluated model performance using Jaccard score and F1 score metrics

Determined the best-performing model based on comprehensive metric analysis

GitHub URL: Machine Learning Prediction

Results

Unique Launch Sites:

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

Date of First Successful Landing on Ground Pad:

- First Successful Landing: 2015-12-22

Total Payload Mass Launched by NASA (CRS):

- Total Payload Mass: 45596 kg

Average Payload Mass of Booster Version F9 v1.1:

- Average Payload Mass: 2534.67 kg

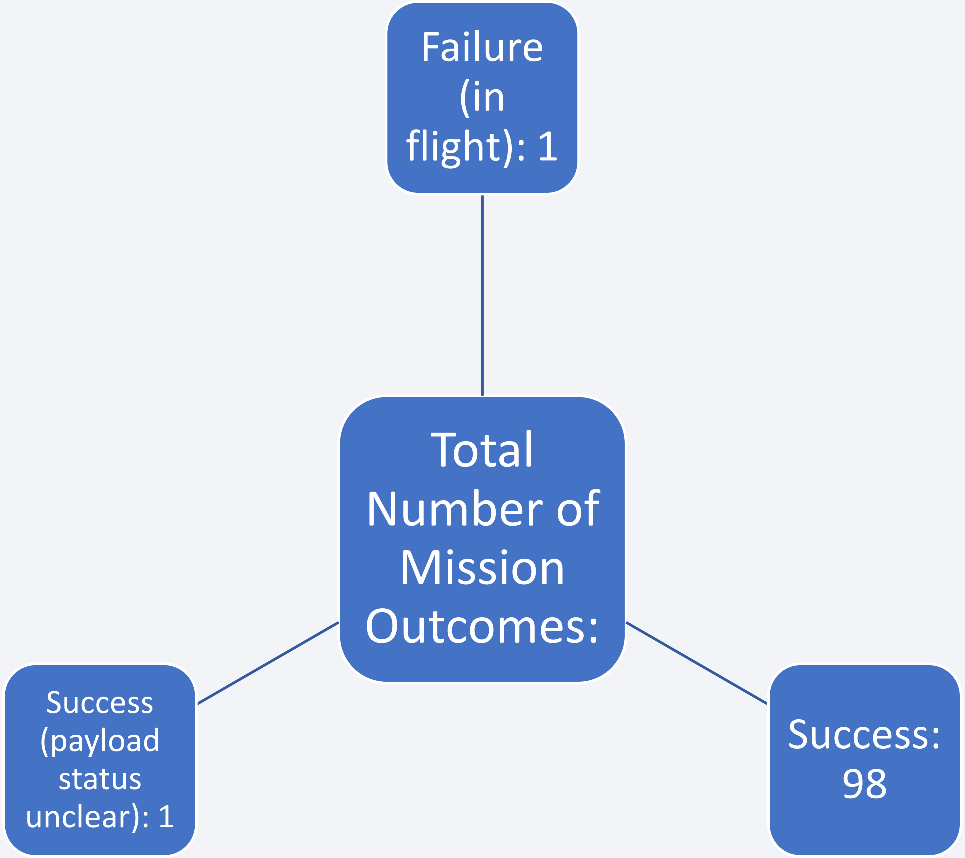
Boosters with Maximum Payload Mass:

- F9 B5 B1048.4
- F9 B5 B1049.4

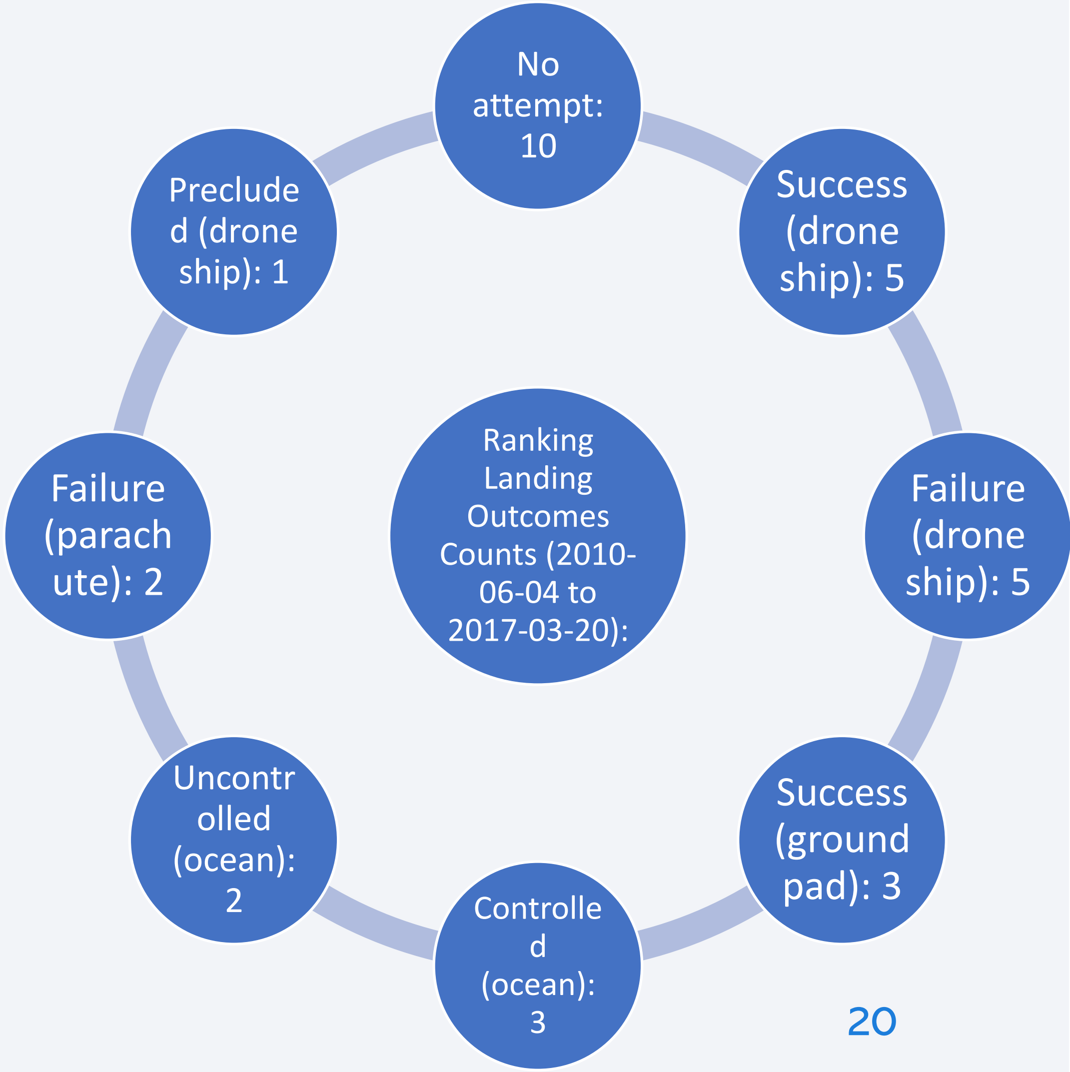
Boosters with Success on Drone Ship and Payload Mass 4000-6000 kg:

- F9 FT B1022
- F9 FT B1031.2
- F9 FT B1026
- F9 FT B1021.2

Results


Failure (in flight): 1
Total Number of Mission Outcomes:
Success (payload status unclear): 1
Success: 98


January 2015: F9 v1.1 B1012, CCAFS LC-40, Failure (drone ship)
Records for Failure Landing Outcomes on Drone Ship in 2015:
April 2015: F9 v1.1 B1015, CCAFS LC-40, Failure (drone ship)


Ranking Landing Outcomes Counts (2010-06-04 to 2017-03-20):
No attempt: 10
Success (drone ship): 5
Failure (drone ship): 5
Success (ground pad): 3
Controlled (ocean): 3
Uncontrolled (ocean): 2
Failure (parachute): 2
Precluded (drone ship): 1

Results

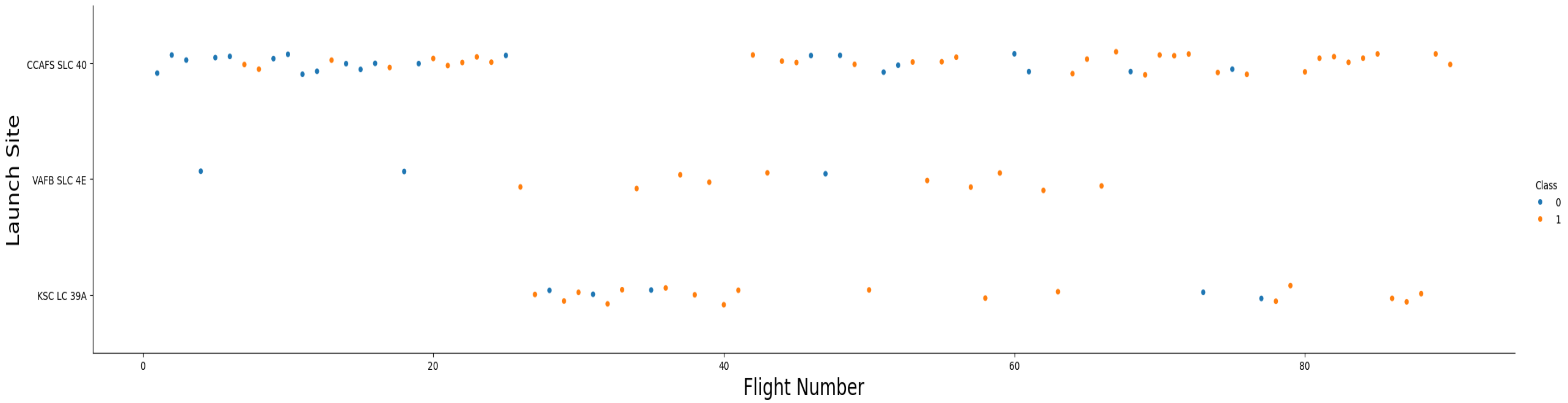
Upon evaluating the Test Set scores, a definitive determination regarding the superior-performing method cannot be confidently made. The consistency in scores across different methods within the Test Set may be attributed to its relatively small sample size (18 samples); to address this limitation, a comprehensive assessment was conducted, considering the entire dataset.

Upon evaluating the complete dataset, the results conclusively identify the Decision Tree Model as the most effective: this model not only boasts higher scores but also achieves the highest accuracy among all methods considered.

Section 2

Insights drawn from EDA

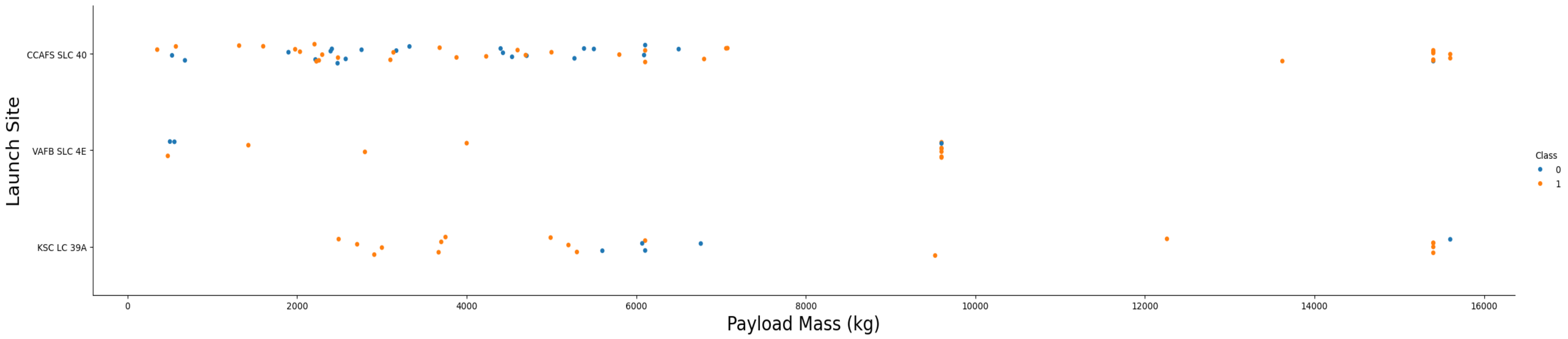
Flight Number vs. Launch Site



Pattern explanation:

- The earliest flights all failed while the latest flights all succeeded.
- The CCAFS SLC 40 launch site has about a half of all launches.
- VAFB SLC 4E and KSC LC 39A have higher success rates.
- It can be assumed that each new launch has a higher rate of success.

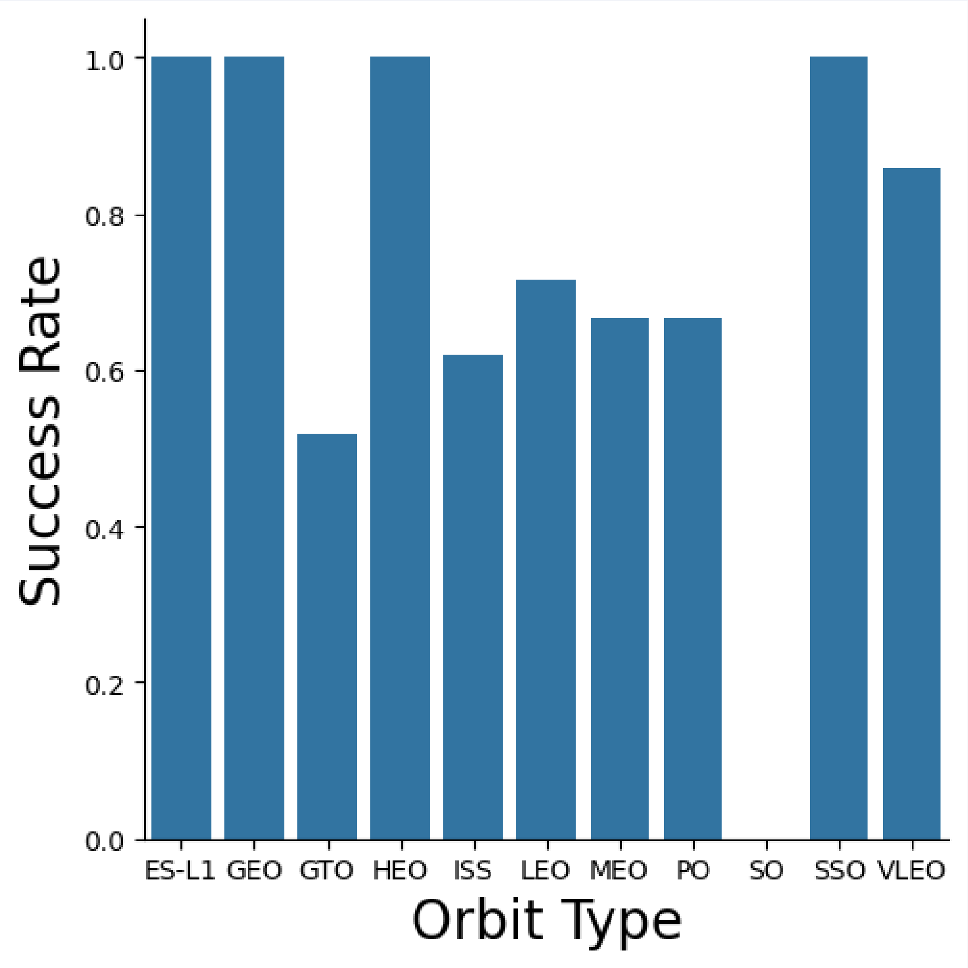
Payload Mass(kg) vs. Launch Site



Pattern explanation:

- There seems to be a positive correlation between payload mass and success rate across all launch sites: generally, higher payload masses coincide with higher success rates.
- Notably, a majority of launches with payload masses exceeding 7000 kg resulted in successful missions.
- KSC LC 39A stands out with a remarkable 100% success rate, even for payload masses under 5500 kg.
- In regards to the VAFB-SLC launch site, it appears that there are no recorded rocket launches with payload masses surpassing 10000 kg, which suggests a specific operational characteristic or limitation for this launch site in handling heavy payloads.

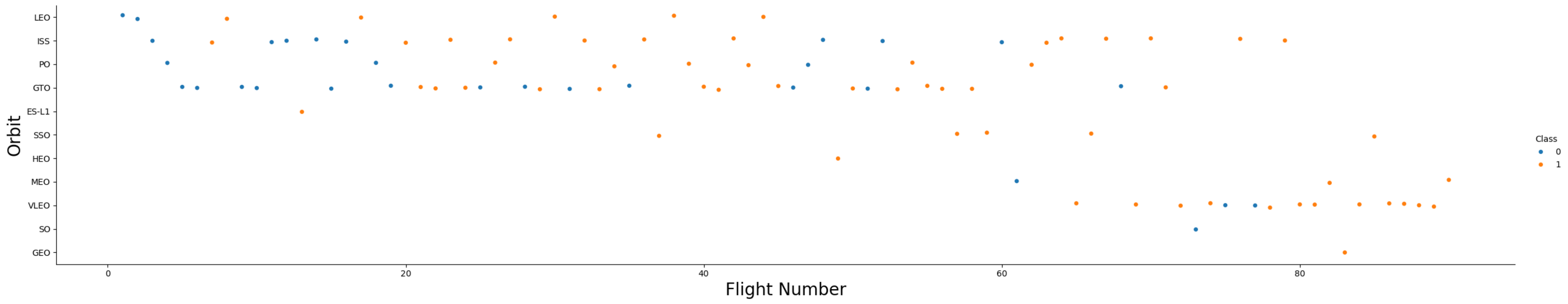
Orbit Type vs. Success Rate



Pattern explanation:

- Orbits with 100% success rate:
  - ES-L1
  - GEO
  - HEO
  - SSO
- Orbits with success rate between 50% and 85%:
  - GTO
  - ISS
  - LEO
  - MEO
  - PO
- Orbits with 0% success rate:
  - SO

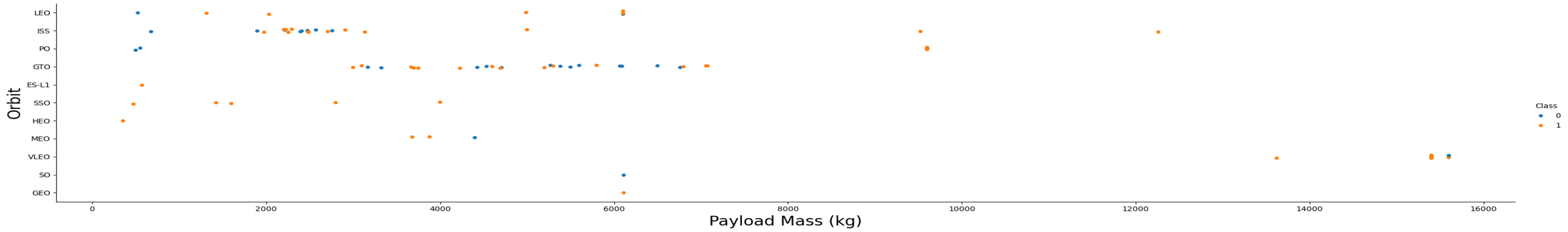
Flight Number vs. Orbit Type



Pattern explanation:

- Success rate appears to be influenced by the number of flights conducted in this orbit. In contrast, for missions in Geostationary Transfer Orbit (GTO), there seems to be no apparent correlation between the flight number and the mission's success.
    - This, in turn, suggests that factors affecting success may vary between LEO and GTO orbits, highlighting the need for further investigation into the specific dynamics of each orbital scenario.

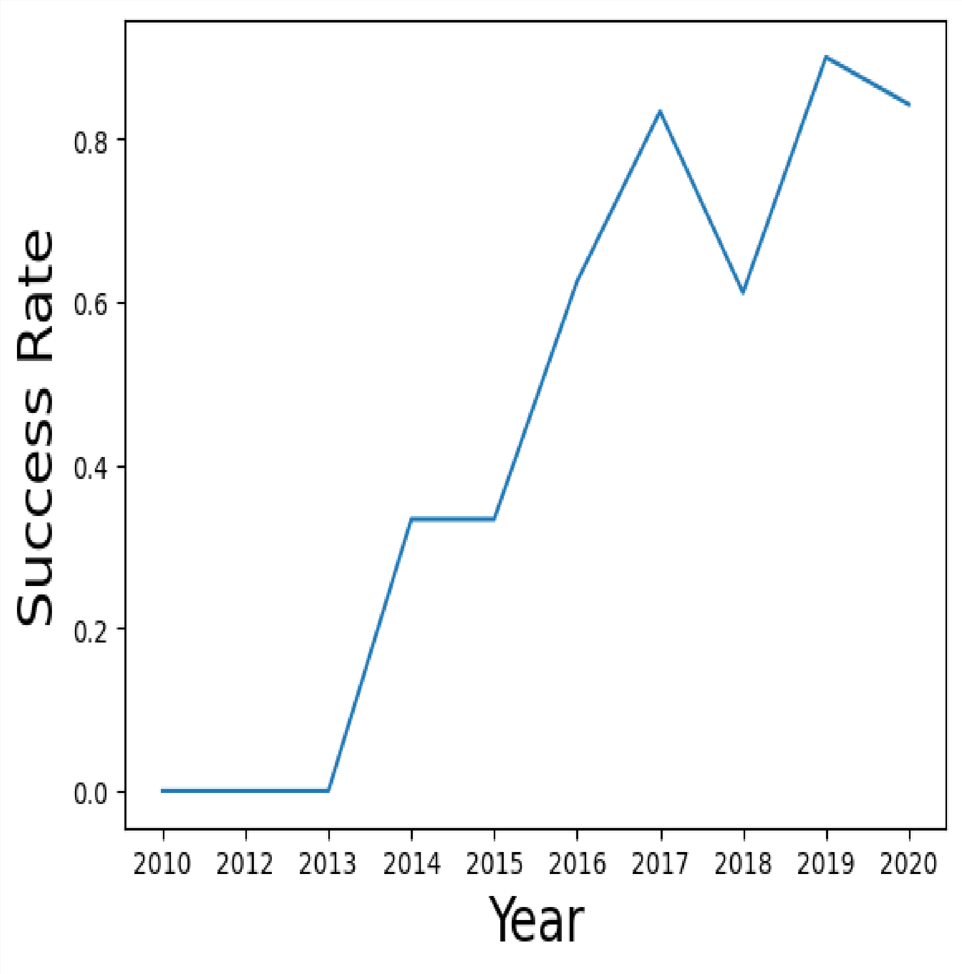
Payload Mass(kg) vs. Orbit Type



Pattern explanation:

- On heavy payloads, we observe a higher rate of successful landings - positive landing outcomes - particularly in Polar, Low Earth Orbit (LEO), and International Space Station (ISS) missions. However, distinguishing such patterns becomes challenging in Geostationary Transfer Orbit (GTO) scenarios.
  - In GTO, both positive landing rates and negative landing outcomes (unsuccessful missions) coexist, making it less clear-cut to discern specific trends associated with heavy payload missions in this orbital category.

Launch Success Yearly Trend



Pattern explanation:

- Examining the success rates from 2013 to 2020, we reveal a consistent and upward trajectory. Over this period, the success rate steadily increased, indicating a positive trend in mission outcomes.
  - This implies a significant enhancement in mission success during the specified timeframe, possibly attributable to technological advancements, improved operational efficiency, or other influential factors in the field of space exploration.

All Launch Site Names

```
%sql select distinct launch_site from SPACEXTBL;
```

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

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

2.1. This SQL query retrieves the distinct launch sites recorded in the "SPACEXTBL" dataset. The result provides a list of unique launch sites where space missions have been conducted. In this specific case, the launch sites are CCAFS LC-40, VAFB SLC-4E, CCAFS SLC-40 and KSC LC-39A.

Launch Site Names Begin with 'CCA'

```
%sql select * from SPACEXTBL where launch_site like 'CCA%'
```

```
 * 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
2013-			CCAFS LC-						

2.2. This SQL query calculates the total payload mass for missions commissioned by NASA under the CRS program, represented by the result named "total_payload_mass,". The result set contains 60 rows in total, but the screenshot provided only displays the first five results.

Total Payload Mass

```
%sql select sum(payload_mass__kg_) as total_payload_mass from SPACEXTBL where customer = 'NASA (CRS)';
```

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

total_payload_mass
45596

2.3. This SQL query calculates the total payload mass for missions commissioned by NASA under the Commercial Resupply Services (CRS) program, in kilograms (kg). The result, named "total_payload_mass," represents the cumulative mass of all payloads carried in CRS missions - 45596 kg, in this case.

Average Payload Mass by F9 v1.1

```
%sql select avg(payload_mass__kg_) as average_payload_mass from SPACEXTBL where booster_version like '%F9 v1.1%';
```

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

average_payload_mass
2534.6666666666665

2.4. This SQL query computes the average payload mass for missions involving the Falcon 9 version 1.1 booster. The result, named "average_payload_mass," represents the mean payload mass across all instances of this specific booster version - in this case, approximately 2534.67 kg.

First Successful Ground Landing Date

```
%sql select min(date) as first_successful_landing from SPACEXTBL where landing_outcome = 'Success (ground pad)';
```

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

first_successful_landing
2015-12-22

2.5. This SQL query identifies the earliest date on which a successful landing occurred on a ground pad. The result, labeled "first_successful_landing," represents the minimum (earliest) date for such successful ground pad landings - in this instance, December 22, 2015.

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql select booster_version from SPACEXTBL where landing_outcome = 'Success (drone ship)' and payload_mass__kg_ between 400
```

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

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

2.6. This SQL query retrieves the booster versions for space missions with a successful landing on a drone ship and a payload mass between 4000 and 6000 kg; the screenshot provided display a snippet of the full query.

Total Number of Successful and Failure Mission Outcomes

```
%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

2.7. This SQL query categorizes space missions based on their outcomes and counts the occurrences for each outcome; the result, labeled "total_number," provides a summary of the total count for each mission outcome category. The displayed result set lists distinct mission outcomes, such as "Success," "Failure (in flight)," and "Success (payload status unclear)," along with their respective counts.

Boosters Carried Maximum Payload

```
%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

2.8. This SQL query identifies the booster versions associated with space missions that carried the maximum payload mass recorded in the dataset, through use of a subquery; the result set, labeled "booster_version," displays the booster versions corresponding to missions with the highest payload mass. The displayed result includes multiple booster versions, as there may be more than one mission with the maximum payload mass.

2015 Launch Records

```
%%sql
select substr(date, 6, 2) as month, substr(date, 0, 5) as year, date, booster_version, launch_site, landing_outcome
from SPACEXTBL
where landing_outcome = 'Failure (drone ship)' and substr(date, 0, 5) = '2015';
```

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

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

2.9. This SQL query extracts specific details from the dataset, including the month, year, date, booster version, launch site, and landing outcome for missions with a landing outcome of 'Failure (drone ship)' in the year 2015; it provides insights into the failures on drone ship landings during that specific year.

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

```
%%sql select landing_outcome, count(*) as count_outcomes from SPACEXTBL
    where date between '2010-06-04' and '2017-03-20'
    group by landing_outcome
    order by count_outcomes desc;
```

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

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

2.10. This SQL query summarizes landing outcomes for space missions that occurred between June 4, 2010, and March 20, 2017; the result set, labeled "count_outcomes," provides a count of occurrences for each landing outcome category, ordered in descending order. Common landing outcomes include 'No attempt,' 'Success (drone ship),' 'Failure (drone ship),' 'Success (ground pad),' and others, each associated with their respective counts.

Section 3

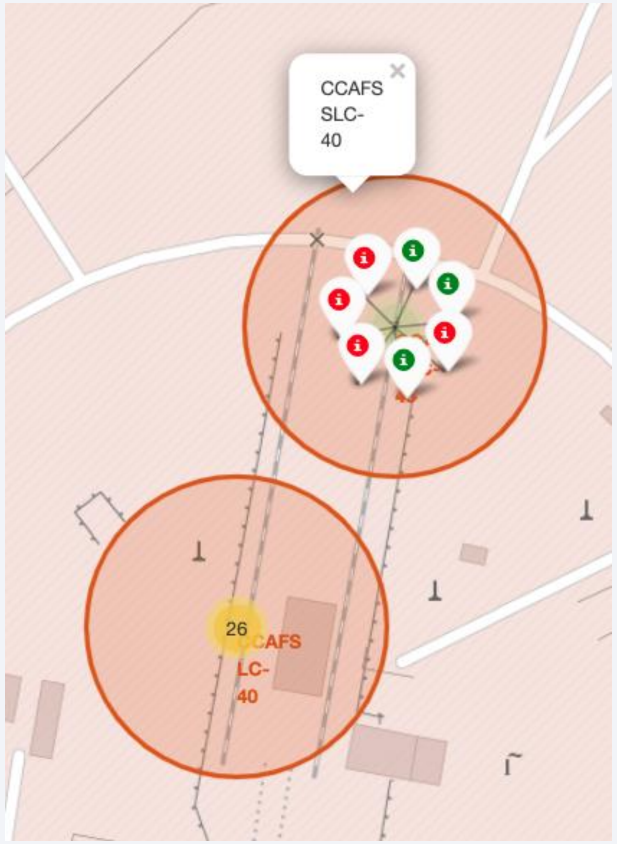
Launch Sites Proximities Analysis

Optimal Launch Sites: Equatorial Advantage & Coastal Safety

3.1. Latitude and longitude coordinates for each site’s location (in orange)

- The selected launch sites are strategically located near the Equator line: this positioning takes advantage of the Earth's maximum rotational speed at the equator (1670 km/h). Launching from the equator provides a substantial velocity boost, nearly 500 km/h more compared to launching from a point halfway to the North Pole.
- All chosen launch sites for this project are in immediate proximity to coastlines - launching rockets towards the ocean serves a critical safety purpose by minimizing the risk of debris falling or exploding in areas inhabited by people - this measure enhances the overall safety and risk mitigation strategies associated with space launches.

Launch Site Success Analysis

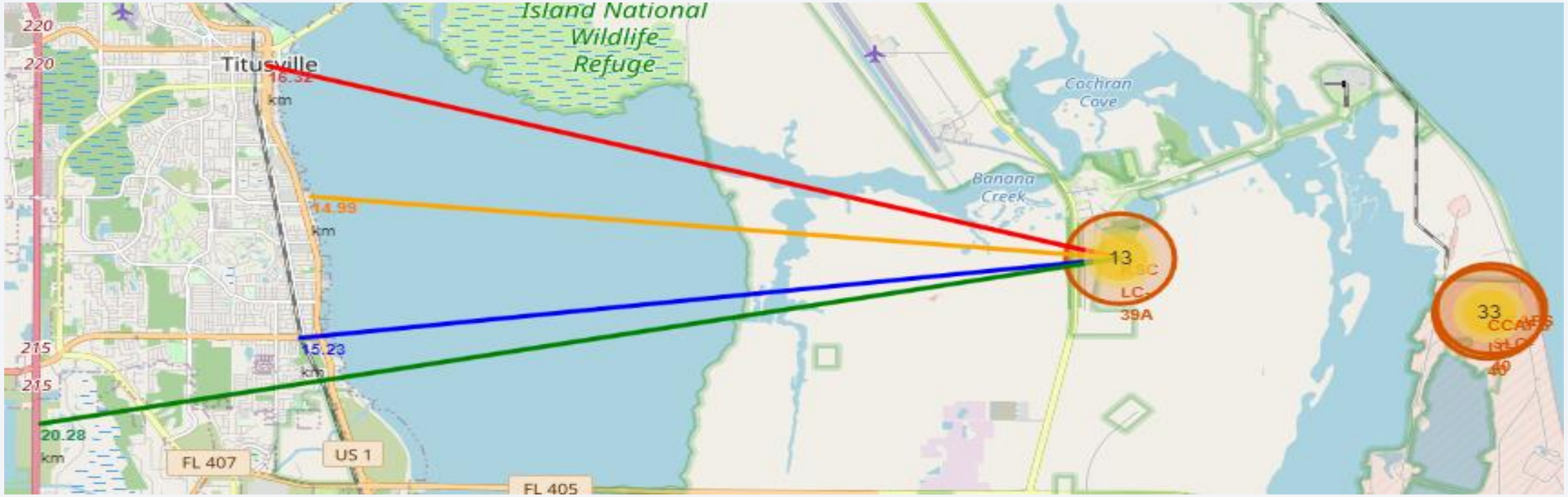


3.2. Zoom-in of one of the launch sites

- By observing markers with distinct colors, it becomes straightforward to distinguish launch sites with comparatively high success rates.

Green Marker = Successful Launch
Red Marker = Failed Launch

Launch Site KSC LC-39A Proximity Analysis: Assessing Risks and Considerations



3.3. Markers indicating launch sites and their distances to a nearby city (red), railway (blue), highway (green), and coastline (yellow), outlined by drawn lines of the respective colors.

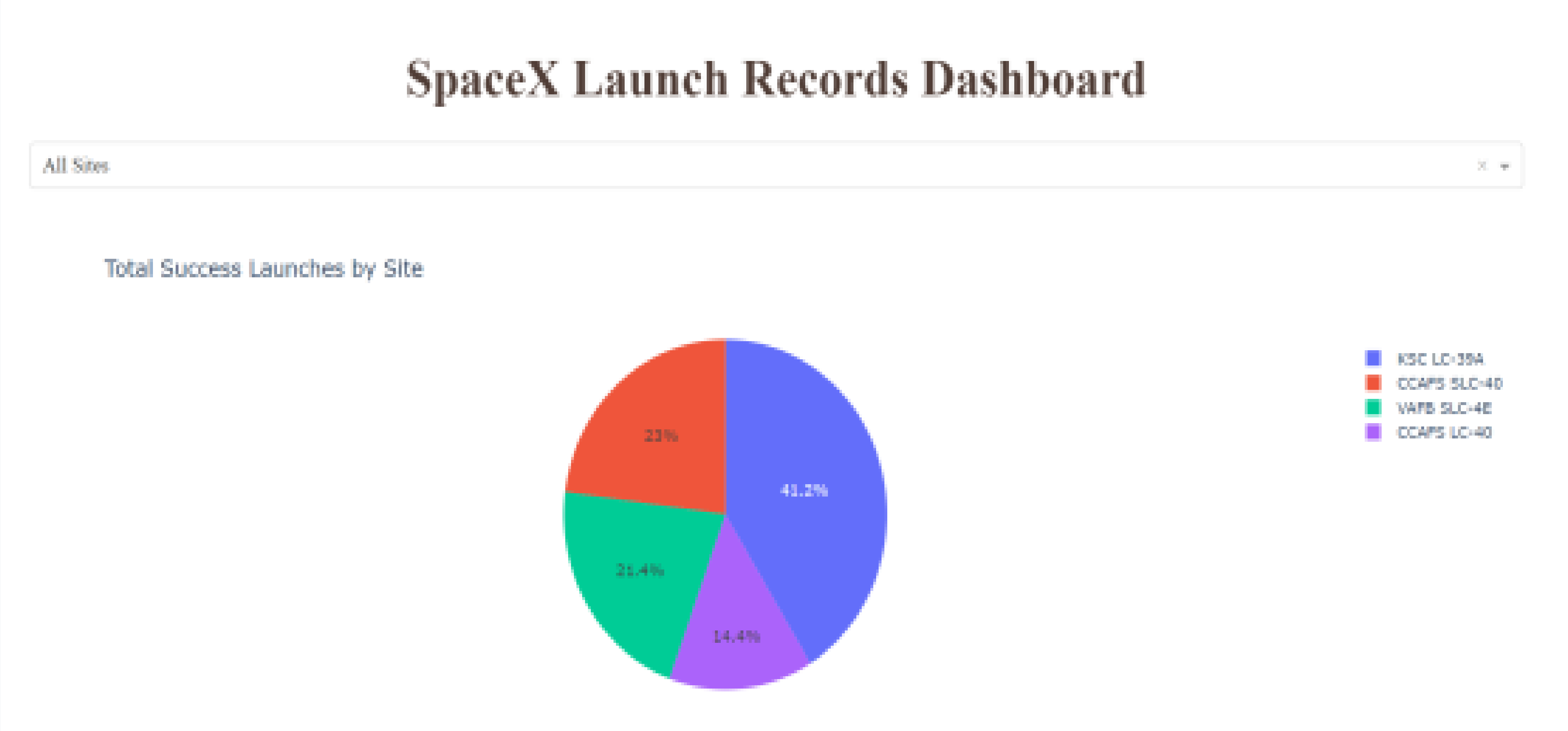
- Railway Proximity: the launch site is relatively close to a railway, approximately 15.23 km away.
- Highway Proximity: similarly, it is relatively close to a highway, situated approximately 20.28 km from the launch site.
- Coastline Proximity: the launch site also demonstrates a close proximity to the coastline, at a distance of approximately 14.99 km.
- City Proximity: furthermore, the launch site maintains relative closeness to its nearest city, Titusville, positioned approximately 16.32 km away.

A failed rocket, propelled at high speed, can cover distances of 15-20 km in mere seconds. The close proximity of the launch site to these infrastructures raises considerations for potential hazards to populated areas, emphasizing the need for careful assessment and risk management in launch site selection and operational planning.

Section 4

Build a Dashboard with Plotly Dash

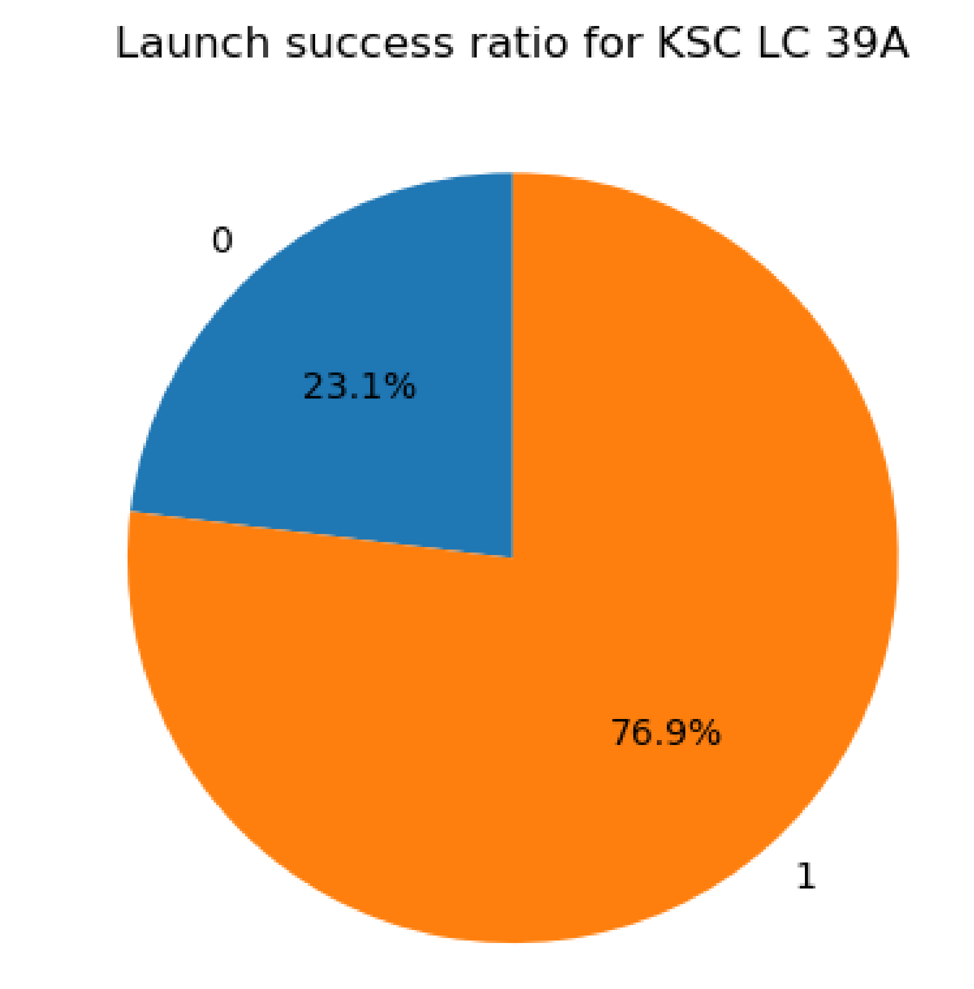
Visualization of Successful Launch Distribution by Launch Site



4.1. This pie chart illustrates the distribution of successful launches among different launch sites.

- Notably, KSC LC 39A emerges as the leader with the highest number of successful launches (41.2%) compared to other sites.

Launch Success (KSC LC 39A)



4.2. This piechart illustrates the success ratio of the launch site with the highest launch success ratio – 76.9% - which is KSC LC 39A.

Payload Mass(kg) and Success – By Booster Version

4.3. Graph demonstrating the correlation between Payload and Success for All Sites.

- Payloads within the weight range of 2,000 kg to 5,000 kg exhibit the highest success rate, with a success outcome represented by 1 and an unsuccessful outcome denoted by 0.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

	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

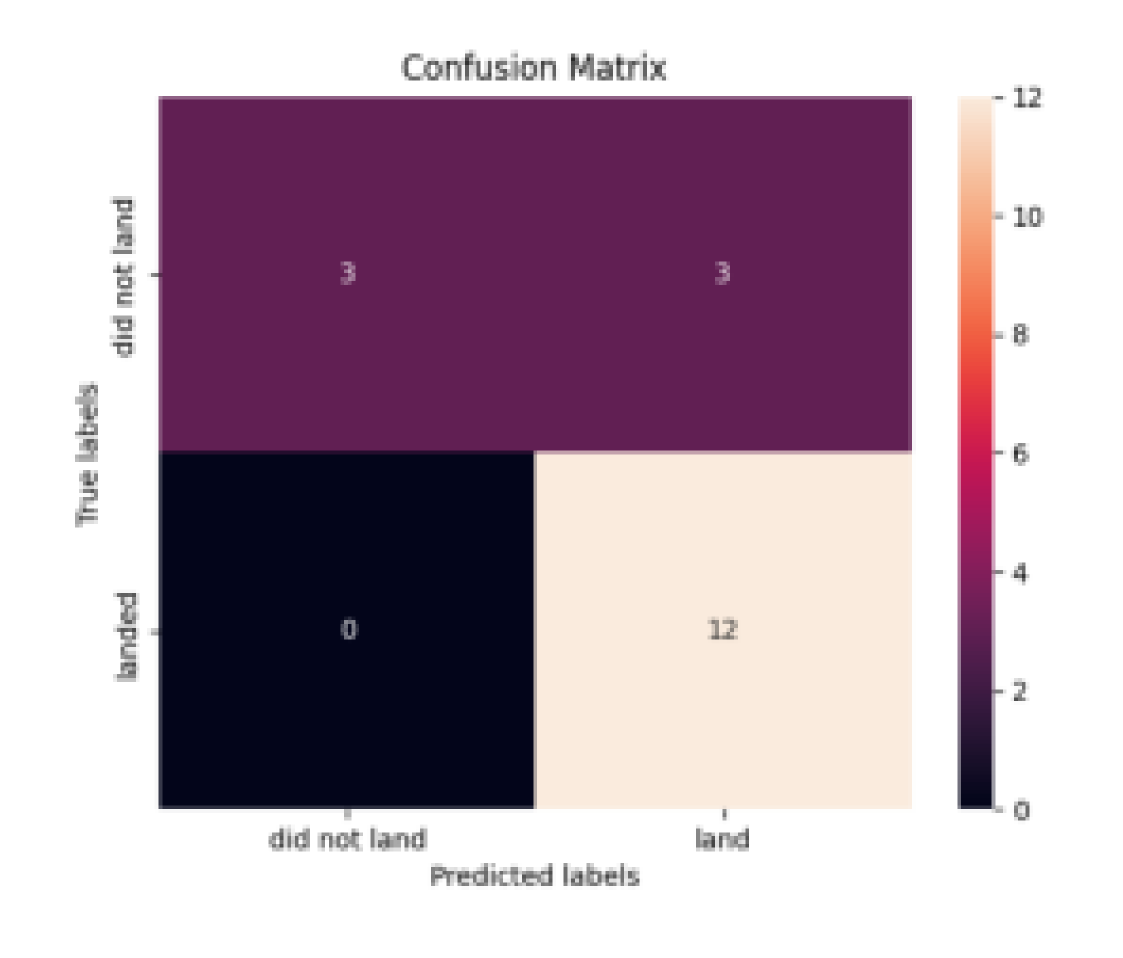
```
models = {'KNeighbors':knn_cv.best_score_,
              'DecisionTree':tree_cv.best_score_,
              'LogisticRegression':logreg_cv.best_score_,
              'SupportVector': svm_cv.best_score_}

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'}
```

5.1. Accuracy of all models

- All models demonstrated similar performance with consistent scores and accuracy, likely attributed to the limited dataset size. The Decision Tree model slightly outperformed the others, particularly evident in the evaluation of .best_score_. This metric represents the average across all cross-validation folds for a specific combination of parameters."

Confusion Matrix



5.2. Confusion Matrix of the Decision Tree Model

Performance Summary:

- The classification algorithm's performance is encapsulated in a confusion matrix.
- Notably, all confusion matrices exhibited identical patterns.
- The presence of false positives (Type 1 error) is a concern.
- Confusion Matrix Outputs:
  - 12 True positive
  - 3 True negative
  - 3 False positive
  - 0 False negative
- Precision (Precision = TP / (TP + FP)): 12 / 15 = 0.80
- Recall (Recall = TP / (TP + FN)): 12 / 12 = 1
- F1 Score (F1 Score = 2 * (Precision * Recall) / (Precision + Recall)): 2 * (0.8 * 1) / (0.8 + 1) = 0.89
- Accuracy (Accuracy = (TP + TN) / (TP + TN + FP + FN)): 0.833

Conclusions

Research Findings:

• Model Performance: The models demonstrated comparable performance on the test set, with the decision tree model exhibiting a slight edge.

• Equatorial Advantage: Most launch sites strategically located near the equator benefit from the Earth's rotational speed, offering a natural boost and cost savings in fuel and boosters.

• Coastal Proximity: All launch sites are strategically positioned close to coastlines.

• Temporal Trend: Launch success rates show a positive trend over time.

• KSC LC-39A: This launch site stands out with the highest success rate, achieving a 100% success rate for launches below 5,500 kg.

• Orbital Success: Specific orbits—ES-L1, GEO, HEO, and SSO—consistently achieve a 100% success rate.

• Payload Mass Influence: Across all launch sites, there is a positive correlation between higher payload mass (kg) and success rate.

Conclusions

Considerations for Future Research:

Dataset Expansion: Enlarging the dataset can enhance predictive analytics and determine the generalizability of findings to a broader context.

Feature Analysis/PCA: Conducting additional feature analysis or principal component analysis may contribute to improving accuracy.

XGBoost Exploration: The exploration of XGBoost, a powerful model not utilized in this study, could provide insights into its potential to outperform other classification models.

Thank you!