IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Furkan Eşref Yazıcı
01/09/2023

Outline

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

Executive Summary

Methodology Summaries

The research attempts to identify the factors for a successful rocket landing. To make this determination, the following methodologies where used:

- **Collect** data using SpaceX REST API and web scraping techniques
- **Wrangle** data to create success/fail outcome variable
- **Explore** data with data visualization techniques, considering the following factors: payload, launch site, flight number and yearly trend
- **Analyze** the data with SQL, calculating the following statistics: total payload, payload range for successful launches, and total # of successful and failed outcomes
- **Explore** launch site success rates and proximity to geographical markers
- **Visualize** the launch sites with the most success and successful payload ranges
- **Build Models** to predict landing outcomes using logistic regression, support vector machine (SVM), decision tree and K-nearest neighbor (KNN)

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, a pioneering force in the aerospace industry, is dedicated to democratizing space travel by striving to make it affordable for all. The company's impressive track record includes achievements such as delivering spacecraft to the International Space Station, establishing a satellite constellation for global internet access, and successfully conducting crewed missions. A key driver behind SpaceX's cost-effectiveness lies in its groundbreaking practice of reusing the first stage of its Falcon 9 rocket, substantially reducing launch costs to a mere $62 million per mission. In stark contrast, competitors unable to implement such reusability models incur launch expenses upwards of $165 million per launch. The ability to predict first-stage landing success plays a pivotal role in determining launch costs. Leveraging publicly available data and machine learning models, this analysis aims to forecast whether SpaceX or competing firms can successfully reuse the first stage of their rockets.

Exploration Objectives

– **Investigate the Impact of Variables:** Assess how variables such as payload mass, launch site, number of prior flights, and specific orbits influence the probability of a successful first-stage landing.
– **Temporal Analysis:** Analyze the historical trend in the rate of successful first-stage landings over time, shedding light on the evolution of SpaceX's reusability efforts.
– **Model Selection:** Identify the most effective predictive model for successful first-stage landing, employing binary classification techniques. In our pursuit of these objectives, we aim to provide insights that contribute to the optimization of cost-effective space travel and highlight the significance of SpaceX's innovative approach to rocket reusability.

Section 1

Methodology

Methodology

Executive Summary

Data Collection:
We acquired the necessary data through a two-fold process:
- Utilizing the SpaceX REST API, we retrieved relevant information.
- We complemented this dataset with data obtained through web scraping techniques.

Data Wrangling:
Data was subjected to a meticulous wrangling process to ensure its readiness for analysis and modeling. Key steps included:
- Data filtering to retain pertinent information.
- Handling missing values through appropriate techniques.
- Application of one-hot encoding for categorical variables.

Exploratory Data Analysis (EDA):
Our analysis began with EDA, a critical step that involved:
- Utilizing SQL for data exploration and retrieval.
- Leveraging data visualization techniques to gain insights into the dataset.

Interactive Visual Analytics:
To enhance data understanding and interactivity, we employed two powerful tools:
- Folium for geospatial visualization.
- Plotly Dash for dynamic and interactive visualizations.

Predictive Analysis Using Classification Models:
In our quest to predict first-stage landing outcomes, we adopted a structured approach:
- We developed various classification models.
- These models were meticulously built, tuned, and evaluated to identify the best-performing model and optimal parameters.

Data Collection

Process

Request Data from SpaceX API:
- Initiate data collection by requesting information from the SpaceX API, focusing on rocket launch data.

Decode Response and Convert to DataFrame:
- Decode the received response using **.json()** to ensure readability.
- Convert the decoded data into a structured DataFrame using **.json_normalize()**.

Request Launch Information from SpaceX API:
- Utilize custom functions to specifically request detailed information about the launches from the SpaceX API.

Create a Dictionary from the Data:
- Organize the obtained data into a dictionary for structured storage and manipulation.

Create a DataFrame from the Dictionary:
- Transform the dictionary into a DataFrame format for comprehensive data analysis.

Filter DataFrame for Falcon 9 Launches:
- Isolate and retain data related to Falcon 9 rocket launches, focusing our analysis.

Replace Missing Payload Mass Values:
- Address missing values in the Payload Mass column by replacing them with the calculated mean value.

Export Data to CSV File:
- Conclude the data collection phase by exporting the cleaned and structured data to a CSV file for future use and reference.

Github Repo: https://github.com/NonsensicalInsane/IBMDataScienceCourse/blob/main/Capstone/IBM_DataScience_Spacex_DataCollectionAPI.ipynb

Data Collection - Scraping

Web Scrapping Process

Request Data from Wikipedia:
- Commence data acquisition by requesting Falcon 9 launch data from Wikipedia.

Create BeautifulSoup Object:
- Construct a BeautifulSoup object to facilitate the parsing of the HTML response.

Extract Column Names from HTML Header:
- Identify and extract column names from the HTML table header, ensuring data structure understanding.

Collect Data from Parsed HTML Tables:
- Parse the HTML tables to extract the relevant data for analysis.

Create a Dictionary from the Data:
- Organize the collected data into a structured dictionary format, enhancing data management.

Create a DataFrame from the Dictionary:
- Transform the dictionary into a DataFrame for structured data analysis.

Export Data to CSV File:
- Conclude the data collection process by exporting the cleaned and structured data to a CSV file, facilitating future reference and analysis.

Github Repo: https://github.com/NonsensicalInsane/IBMDataScienceCourse/blob/main/Capstone/IBM_DataScience_Spacex_Webscrapping.ipynb

Data Wrangling

Perform Exploratory Data Analysis (EDA):

- Initiate data wrangling by conducting exploratory data analysis (EDA) to understand the data's structure and labels.

Calculate Key Metrics:

- Determine essential metrics, including:
- Number of launches for each launch site.
- Number and occurrence of different orbits.
- Number and occurrence of mission outcomes per orbit type.

Create Binary Landing Outcome Column:

- Create a binary landing outcome column, serving as the dependent variable for our analysis.

Export Data to CSV File:

- Conclude the data wrangling phase by exporting the cleaned and processed data to a CSV file for future reference and analysis.

Binary Landing Outcome Categories:

1) **False Ocean:** Represents an unsuccessful landing in a specific ocean region.
2) **True RTLS:** Denotes a successful landing on a ground pad.
3) **False RTLS:** Signifies an unsuccessful landing on a ground pad.
4) **True ASDS:** Indicates a successful landing on a drone ship.
5) **False ASDS:** Represents an unsuccessful landing on a drone ship.

Outcomes Conversion:

- Successful landings are represented by '1'.
- Unsuccessful landings are represented by '0'.

Github Repo: https://github.com/NonsensicalInsane/IBMDataScienceCourse/blob/main/Capstone/IBM_DataScience_Spacex_DataWrangling.ipynb

EDA with Data Visualization

Summarized Charts and Their Purpose:

Flight Number vs. Payload:
- This scatter plot illustrates the relationship between flight numbers and payload mass.
- *Purpose*: To discern any patterns or trends in payload mass across different launch missions.

Flight Number vs. Launch Site:
- Another scatter plot showcasing the correlation between flight numbers and launch sites.
- *Purpose*: To investigate whether there are any patterns or preferences in launch site selection over time.

Payload Mass (kg) vs. Launch Site:
- A scatter plot highlighting the association between payload mass and launch site.
- *Purpose*: To explore the influence of launch site on payload mass, potentially revealing site-specific characteristics.

Payload Mass (kg) vs. Orbit Type:
- This visualization involves a scatter plot that examines the connection between payload mass and orbit type.
- *Purpose*: To uncover any patterns or dependencies between payload mass and the type of orbit used.

EDA with Visualization Analysis:
- **Scatter Plots:**
- Scatter plots have been employed to visualize relationships between variables.
- *Rationale*: These plots help us identify potential correlations that could prove valuable for machine learning if significant relationships exist within the data.

Bar Charts:
- Bar charts have been utilized for comparing discrete categories and measured values.
- *Rationale*: Bar charts effectively showcase relationships among categorical variables and provide insights into the distribution of data across different categories.

Github Repo: https://github.com/NonsensicalInsane/IBMDataScienceCourse/blob/main/Capstone/IBM_DataScience_Spacex_Visualization.ipynb

EDA with SQL

Summary of SQL Queries

Names of Unique Launch Sites:
– Retrieve and display the names of launch sites that are distinct within the dataset.

Launch Sites Beginning with 'CCA':
– Showcase five records where the launch site name commences with 'CCA'.

Total Payload Mass by NASA (CRS):
– Calculate and present the total payload mass carried by boosters launched under NASA's Commercial Resupply Services (CRS) program.

Average Payload Mass for Booster Version F9 v1.1:
– Compute and list the average payload mass carried by boosters of version F9 v1.1.

Date of First Successful Ground Pad Landing:
– Identify and display the date of the first successful landing on a ground pad.

Boosters with Successful Drone Ship Landing (Payload 4,000 - 6,000 kg):
– List the names of boosters that achieved successful landings on drone ships and carried payloads greater than 4,000 kg but less than 6,000 kg.

Total Successful and Failed Missions:
– Calculate the total number of successful and failed missions.

Booster Versions with Maximum Payloads:
– Identify the names of booster versions that have carried the maximum payload mass.

Failed Drone Ship Landings in 2015:
– Present the outcomes of failed drone ship landings, including booster version and launch site, for the year 2015.

Count of Landing Outcomes (2010-06-04 to 2017-03-20):
Calculate and display the count of landing outcomes between June 4, 2010, and March 20, 2017, in descending order.

Github Repo: https://github.com/NonsensicalInsane/IBMDataScienceCourse/blob/main/Capstone/IBM_DataScience_Spacex_SQLITE.ipynb

Build an Interactive Map with Folium

Map Objects

Markers Indicating Launch Sites:
- Blue circles were added at the coordinates of NASA Johnson Space Center, accompanied by popup labels displaying its name, utilizing latitude and longitude coordinates.
- Red circles were included at the coordinates of all launch sites, with popup labels showcasing their respective names, utilizing latitude and longitude coordinates.

Map with Folium Colored Markers of Launch Outcomes:
- Colored markers were introduced to represent successful (green) and unsuccessful (red) launches at each launch site, visually conveying launch success rates.

Distances Between a Launch Site and Proximities:
- Colored lines were integrated to illustrate the distance between launch site CCAFS SLC40 and its proximity to the nearest coastline, railway, highway, and city, providing valuable geographical context.

Rationale Map Objects

Markers Indicating Launch Sites:
- To pinpoint the location of significant launch sites, enhancing the viewer's understanding of their geographical distribution.

Map with Folium Colored Markers of Launch Outcomes:
- To visually represent launch outcomes and highlight which launch sites have high success rates, aiding in data interpretation.

Distances Between a Launch Site and Proximities:
To provide spatial context and visualize the proximity of launch sites to key geographic features, assisting in assessing launch site suitability.

Github Repo: https://github.com/NonsensicalInsane/IBMDataScienceCourse/blob/main/Capstone/IBM_DataScience_Spacex_FoliumLocationVisulatization.ipynb

Build a Dashboard with Plotly Dash

Map Elements

Dropdown List with Launch Sites:
- Enables users to choose either all launch sites or a specific one for visualization, enhancing user interactivity and customization.

Dashboard with Plotly Dash:
- Incorporates a slider for selecting payload mass ranges, allowing users to narrow down data based on their preferences.

Pie Chart Showing Successful Launches:
- Presents data in a pie chart format to illustrate the proportion of successful and unsuccessful launches as a percentage of the total, providing a clear overview of launch outcomes.

Scatter Chart Showing Payload Mass vs. Success Rate by Booster Version:
- Depicts the correlation between payload mass and launch success rate by displaying data in a scatter chart, offering insights into the relationship between these variables.

Rationale for Map Elements

Dropdown List with Launch Sites:
- Facilitates user-driven exploration by offering the choice to focus on specific launch sites or view data collectively.

Dashboard with Plotly Dash:
- Provides users with the ability to filter data by payload mass range, enabling customization of the displayed information.

Pie Chart Showing Successful Launches:
- Offers a concise visual representation of launch success proportions, aiding in quick comprehension.

Scatter Chart Showing Payload Mass vs. Success Rate by Booster Version:
Allows users to explore the correlation between payload mass and launch success, helping identify potential trends or dependencies.

Github Repo: https://github.com/NonsensicalInsane/IBMDataScienceCourse/blob/main/Capstone/IBM_DataScience_Spacex_DashPlotly.py

Predictive Analysis (Classification)

Create NumPy Array from Class Column:
- Convert the target variable (Class) into a NumPy array for modeling purposes.

Standardize the Data with StandardScaler:
- Utilize StandardScaler to standardize the data, ensuring consistent scales.
- Fit and transform the data to make it suitable for modeling.

Split the Data using train_test_split:
- Divide the dataset into training and testing sets to enable model evaluation.

Create a GridSearchCV Object for Parameter Optimization:
- Implement GridSearchCV with cv=10 to perform parameter optimization, enhancing model performance.

Apply GridSearchCV on Different Algorithms:
- Employ GridSearchCV on multiple algorithms, including logistic regression, support vector machine, decision tree, and K-Nearest Neighbor, to identify the best-performing model.

Calculate Accuracy on the Test Data:
- Assess the accuracy of all models on the test data using the .score() method, gauging their predictive capabilities.

Assess the Confusion Matrix for All Models:
- Examine confusion matrices for all models to evaluate their ability to classify data correctly.

Identify the Best Model using Jaccard Score, F1 Score, and Accuracy:
- Determine the optimal model by considering performance metrics such as Jaccard Score, F1 Score, and Accuracy, enabling a comprehensive assessment of model efficacy.

Github Repo: https://github.com/NonsensicalInsane/IBMDataScienceCourse/blob/main/Capstone/IBM_DataScience_Spacex_Machine_Learning_Prediction_Part_5.ipynb

Results Summary

Exploratory Data Analysis (EDA) Results:

– Notable improvement in launch success rates over time.
– Kennedy Space Center's Launch Complex 39A (KSC LC-39A) stands out with the highest success rate among landing sites.
– Orbits ES-L1, GEO, HEO, and SSO exhibit a remarkable 100% success rate, indicating precision in mission execution.

Visual Analytics:

– Most launch sites are strategically located near the equator, optimizing launch trajectories.
– Launch sites are strategically positioned, ensuring adequate distance from densely populated areas (cities), major transportation routes (highways and railways), while maintaining proximity for logistical support.

Predictive Analytics:

– The Decision Tree model emerges as the best predictive model for the dataset, demonstrating its efficacy in making informed predictions based on historical data.

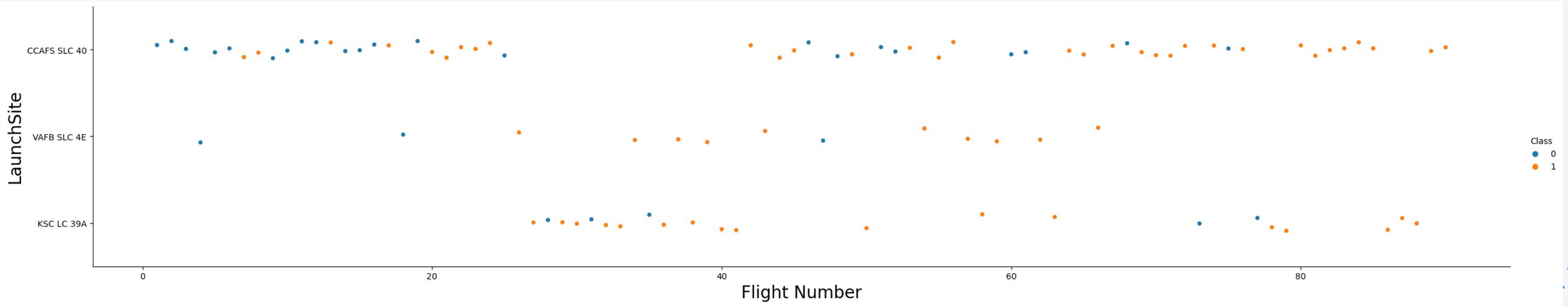
Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

Exploratory Data Analysis

- Earlier flights had a lower success rate (blue = fail)
- Later flights had a higher success rate (orange = success)
- Around half of launches were from CCAFS SLC 40 launch site
- VAFB SLC 4E and KSC LC 39A have higher success rates
- We can infer that new launches have a higher success rate



Payload vs. Launch Site

Payload Mass and Success Rate Relationship:

– Generally, there is a positive correlation between payload mass (kg) and launch success rate. As the payload mass increases, the success rate tends to rise.

Success with Heavy Payloads:

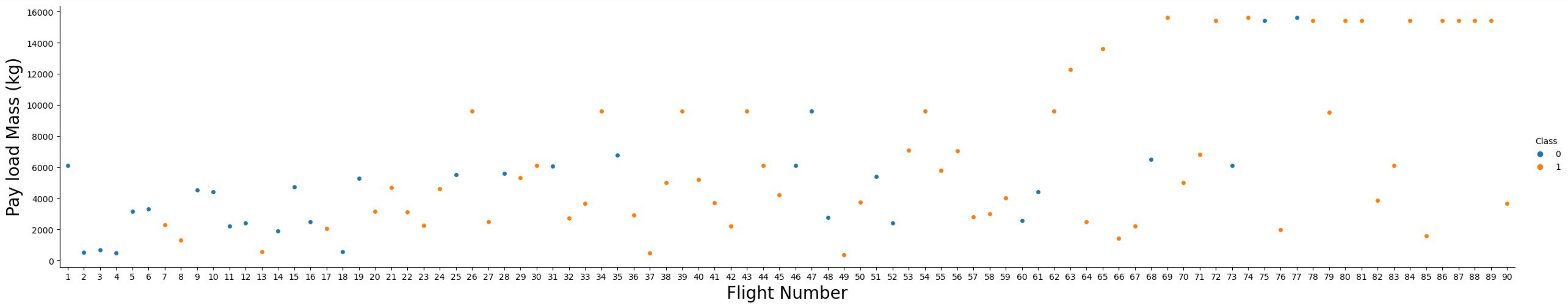
– Notably, a significant portion of launches with a payload greater than 7,000 kg achieved success, underlining the industry's capability to handle heavy payloads successfully.

KSC LC 39A Success Rate for Lighter Payloads:

– Kennedy Space Center's Launch Complex 39A (KSC LC 39A) boasts a remarkable 100% success rate for launches with payloads less than 5,500 kg, reflecting its reliability for lighter payloads.

Limitation at VAFB SKC 4E:

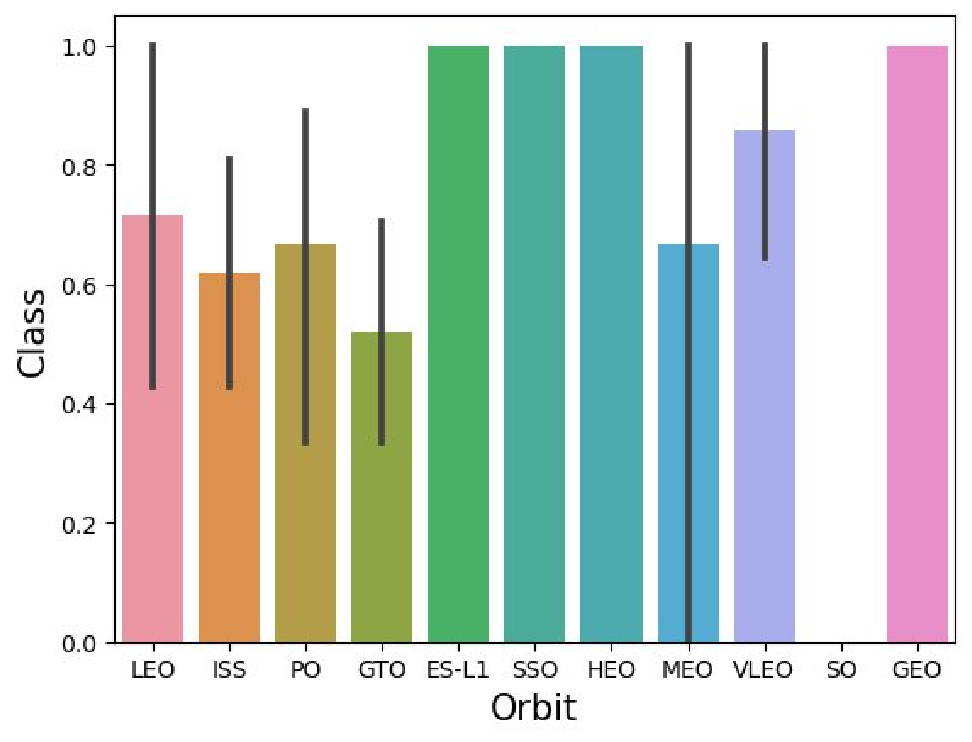
– Vandenberg Air Force Base Space Launch Complex 4E (VAFB SKC 4E) has not conducted any launches with payloads exceeding approximately 10,000 kg, indicating a limitation in handling heavier payloads.



Success Rate vs. Orbit Type

Success Rates by Orbit Type:

- Notably, certain orbit types have consistently achieved a 100% success rate, including ES-L1, GEO, HEO, and SSO, underscoring their reliability in space missions.
- Another group of orbit types, including GTO, ISS, LEO, MEO, and PO, has experienced success rates ranging from 50% to 80%, indicating their viability for a variety of missions.
- In contrast, the SO orbit type has recorded a 0% success rate, suggesting challenges or limitations associated with this specific orbit type.



Flight Number vs. Orbit Type

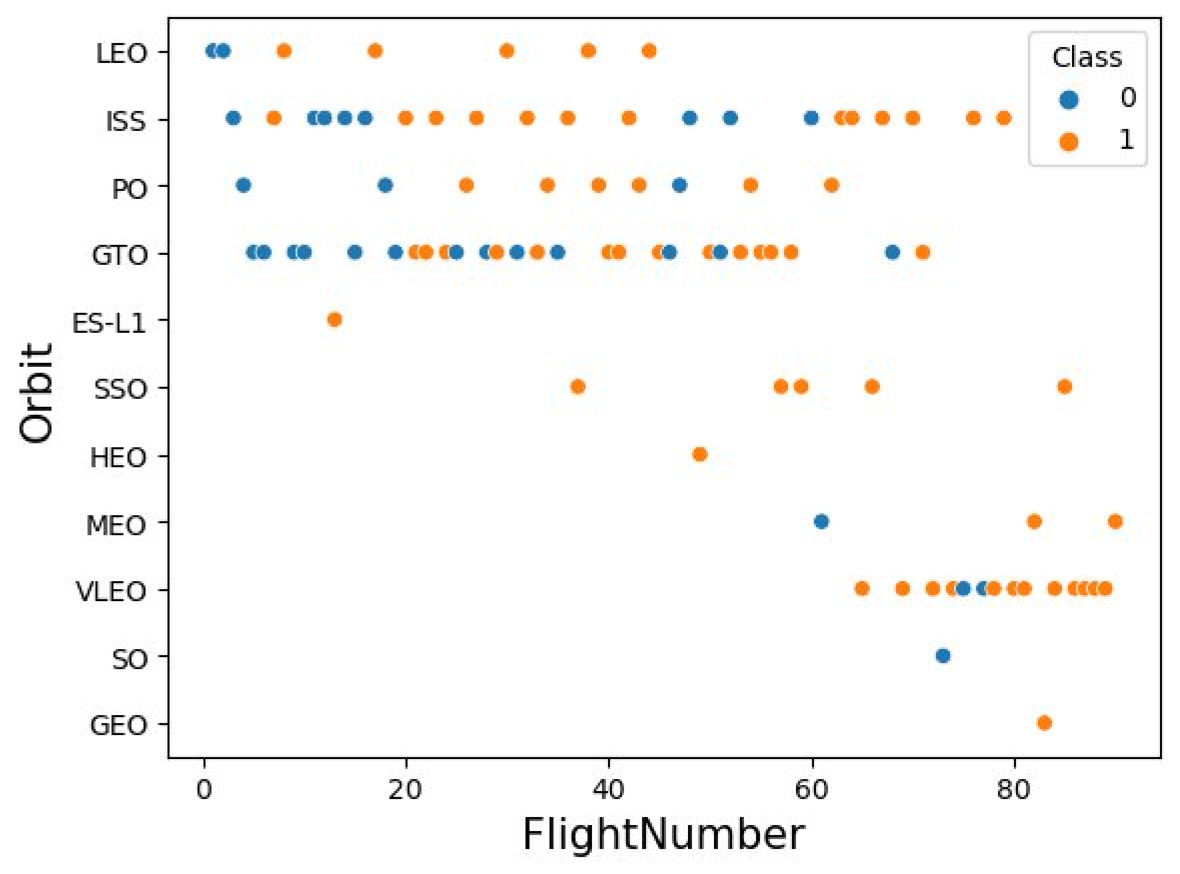
Relationship between Flight Number and Success Rate:

– In general, there appears to be a positive correlation between the number of flights and the success rate for each orbit type. As the number of flights increases, the success rate tends to rise.

Distinct Patterns for LEO and GTO Orbits:

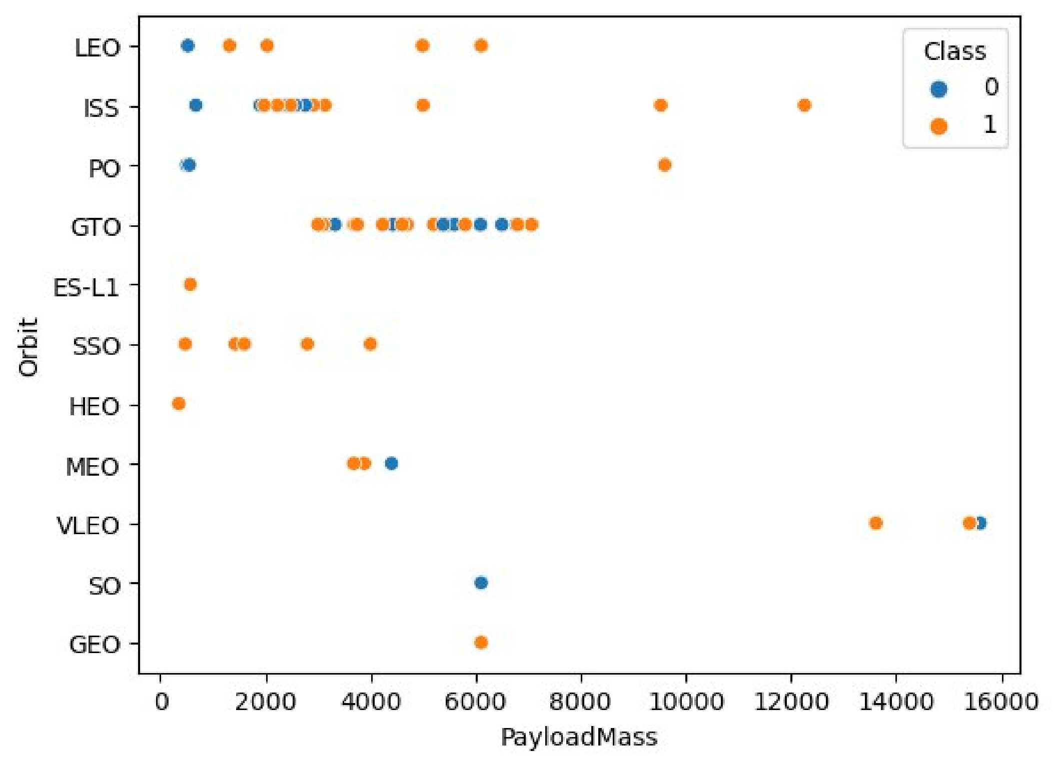
– This relationship is particularly noticeable when analyzing the Low Earth Orbit (LEO). LEO missions exhibit a clear and consistent increase in success rate as the number of flights accumulates. This suggests that experience and repetition contribute to improved success rates in this orbit.

– However, when assessing the Geostationary Transfer Orbit (GTO), a different pattern emerges. Unlike LEO, GTO missions do not exhibit the same trend of increasing success rates with the number of flights. This deviation from the overall trend implies that factors specific to GTO missions may influence their success rates differently.



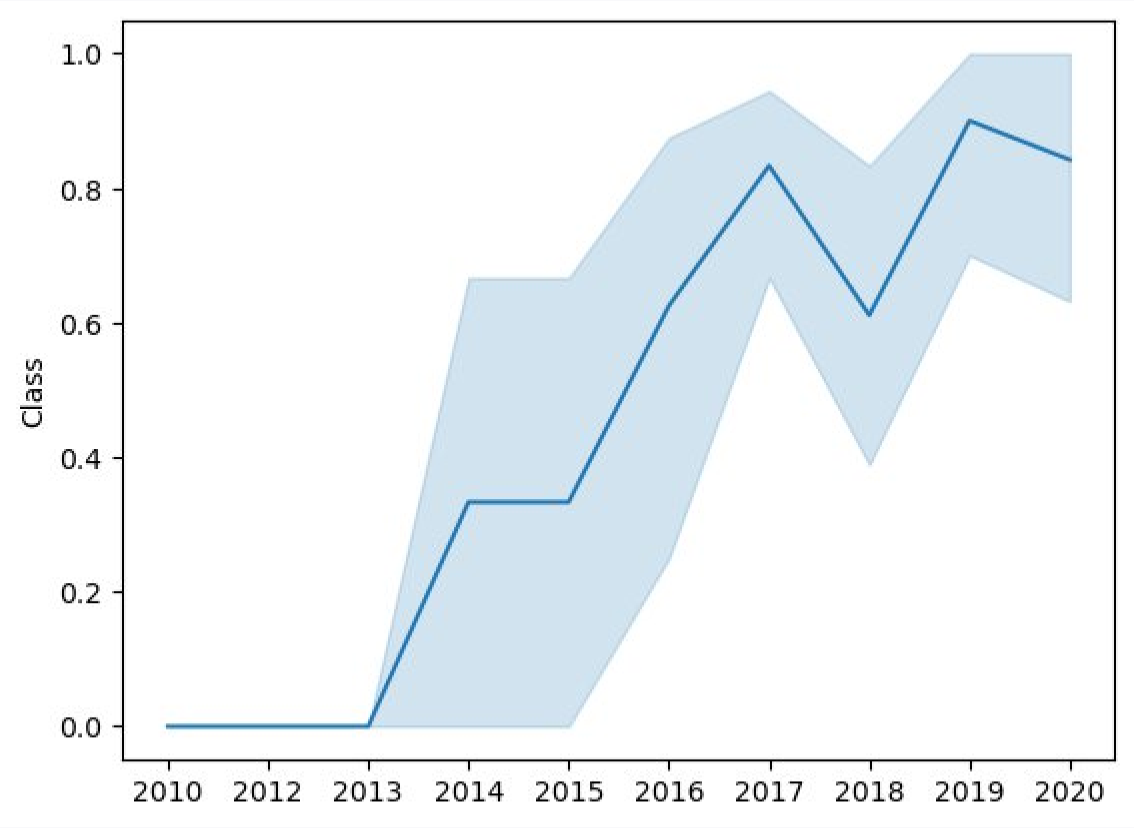
Payload vs. Orbit Type

- Heavy payloads tend to exhibit higher success rates when launched into Low Earth Orbit (LEO), deployed to the International Space Station (ISS), or directed towards Polar Orbit (PO). These orbits appear to be well-suited for handling substantial payloads effectively.
- In contrast, the Geostationary Transfer Orbit (GTO) presents mixed success rates when dealing with heavier payloads, suggesting that this orbit type may pose unique challenges for such missions.



Launch Success Yearly Trend

- Success rates showed improvement during two periods: from 2013 to 2017 and from 2018 to 2019.
- Conversely, success rates declined between 2017 and 2018, as well as from 2019 to 2020.
- In the broader context, the overall success rate has exhibited improvement since 2013.



All Launch Site Names

Launch Site Names
- CCAFS LC-40
- CCAFS SLC-40
- KSC LC-39A
- VAFB SLC-4E

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

```
%%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

Launch Site Names Begin with 'CCA'

Records with Launch Site Starting with CCA

• Displaying 5 records below

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

```
%%sql
SELECT *
FROM SPACEXTABLE
WHERE Launch_Site LIKE'CCA%'
LIMIT 5
```

MagicPython

* sqlite:///my_data1.db
Done.

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success
2010-08-12	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
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success

Total and Average Payload Mass

Total Payload Mass

• 45,596 kg (total) carried by boosters launched by NASA (CRS)

Average Payload Mass

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

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

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

MagicPython

* sqlite:///my_data1.db
Done.

Customer	Total_Payload
NASA (CRS)	45596

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

```
%%sql
SELECT Booster_Version, AVG(PAYLOAD_MASS__KG_) as "Average of Payload"
FROM SPACEXTABLE
WHERE Booster_Version ="F9 v1.1"
```

MagicPython

* sqlite:///my_data1.db
Done.

Booster_Version	Average of Payload
F9 v1.1	2928.4

Landing & Mission Info

1st Successful Landing in Ground Pad

- 12/22/2015

```
%%sql

SELECT Min(Date)
FROM SPACEXTABLE
WHERE Landing_Outcome = "Success (ground pad)"
ORDER BY Date
```

* sqlite:///my_data1.db
Done.

Min(Date)
2015-12-22

Booster Drone Ship Landing

- Booster mass greater than 4,000 but less than 6,000
- JSCAT-14, JSCAT-16, SES-10, SES-11 / EchoStar 105

```
%%sql
SELECT Booster_Version, PAYLOAD_MASS__KG_
FROM SPACEXTABLE
WHERE Landing_Outcome="Success (drone ship)" AND (PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000)
```

0.0s MagicPython

* sqlite:///my_data1.db
Done.

Booster_Version	PAYLOAD_MASS__KG_
F9 FT B1022	4696
F9 FT B1026	4600
F9 FT B1021.2	5300
F9 FT B1031.2	5200

List the total number of successful and failure mission outcomes

```
%%sql
SELECT MISSION_OUTCOME, COUNT(*) AS total
FROM SPACEXTABLE
GROUP BY MISSION_OUTCOME;
```

* sqlite:///my_data1.db
Done.

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

Total Number of Successful and Failed Mission Outcomes

- 1 Failure in Flight
- 99 Success
- 1 Success (payload status unclear)

Boosters Carried Maximum Payload

Carrying Max Payload

- 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

Task 8

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

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

MagicPython

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

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

2015 Launch Records

In 2015

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

```
%%sql
SELECT substr(Date, 6, 2) AS Month, Landing_Outcome, Booster_Version, Launch_Site
FROM SPACEXTABLE
WHERE substr(Date, 0, 5) = '2015'
AND Landing_Outcome = 'Failure (drone ship)';
```

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

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

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

Ranked Descending

• Count of landing outcomes between 2010-06-04 and 2017-03-20 in descending order

```
%%sql
SELECT Landing_Outcome, COUNT(*) AS Total
FROM SPACEXTABLE
WHERE Date BETWEEN "2010-06-04" AND "2017-03-20"
GROUP BY Landing_Outcome
ORDER BY Date;
```

[16] ✓ 0.0s

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

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

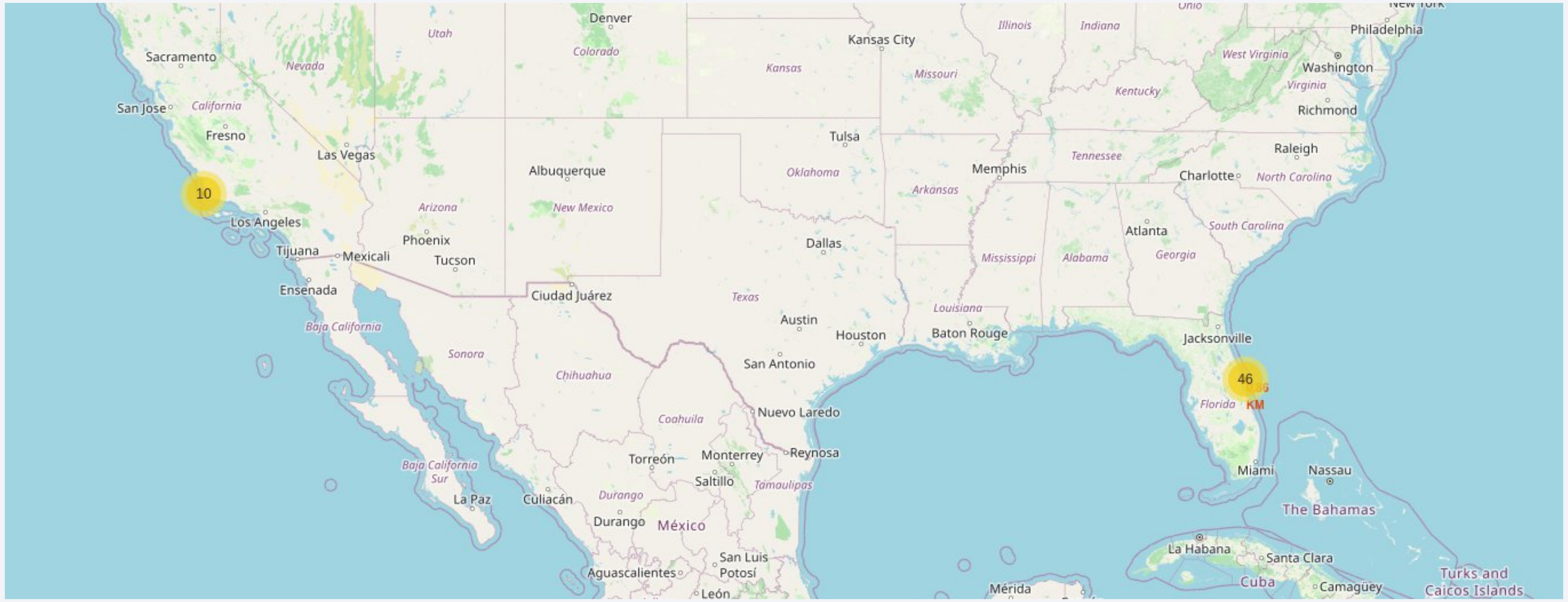
Section 3

Launch Sites Proximities Analysis

Launch Sites

With Markers

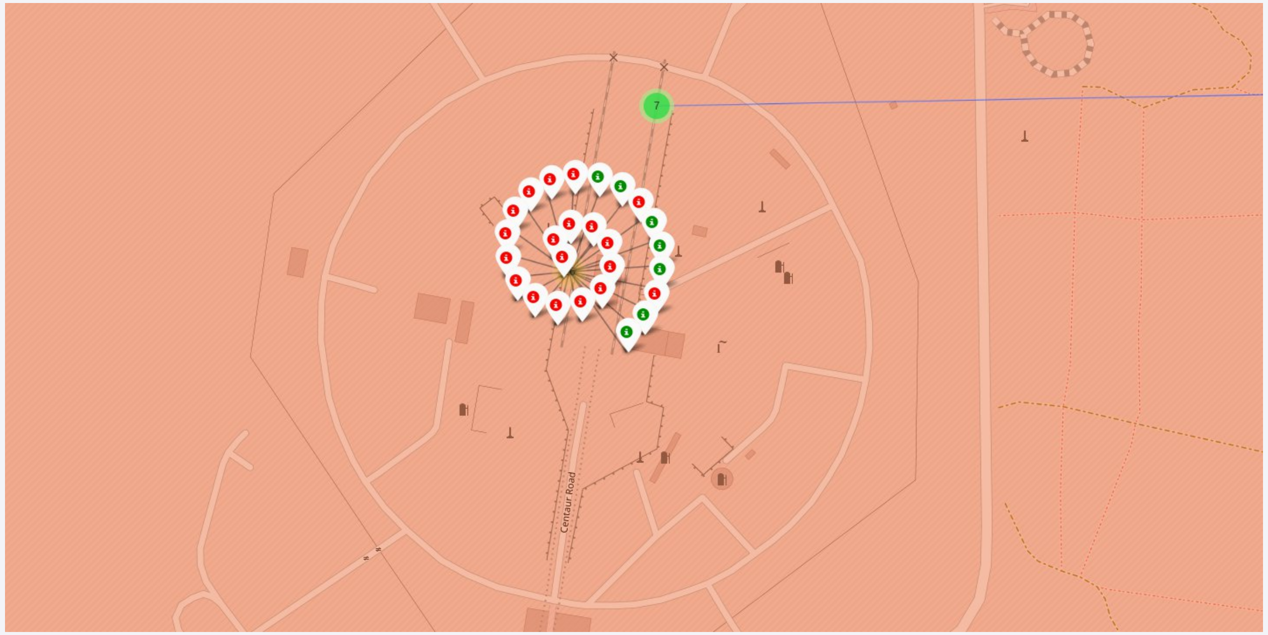
- Near Equator: the closer the launch site to the equator, the easier it is to launch to equatorial orbit, and the more help you get from Earth's rotation for a pro grade orbit. Rockets launched from sites near the equator get an additional natural boost-due to the rotational speed of earth-that helps save the cost of putting in extra fuel and boosters.



Launch Outcomes

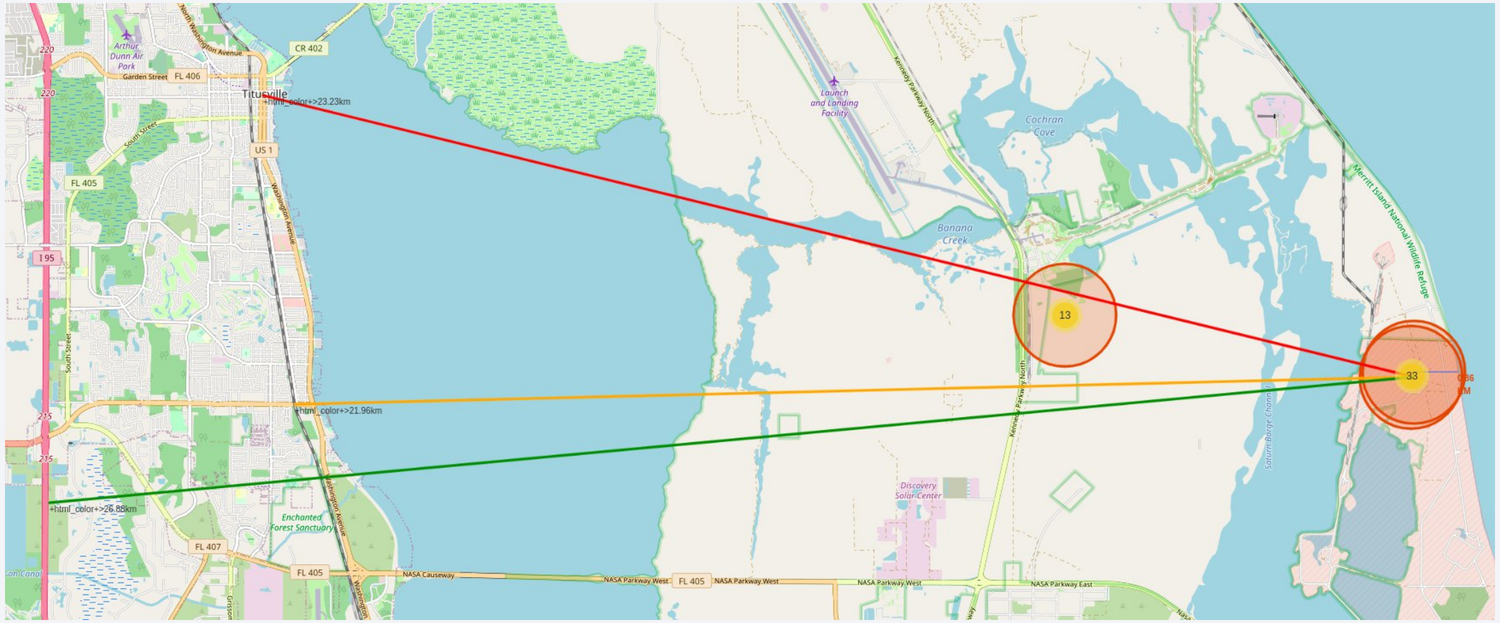
At Each Launch Site

- Outcomes:
- Greenmarkers for successful launches
- Redmarkers for unsuccessful launches
- Launch site CCAFS SLC-40 has a 3/7 success rate (42.9%)



Distance to Proximities

CCAFS SLC-40
• .86 km from nearest coastline
• 21.96 km from nearest railway
• 23.23 km from nearest city •26.88 km from nearest highway



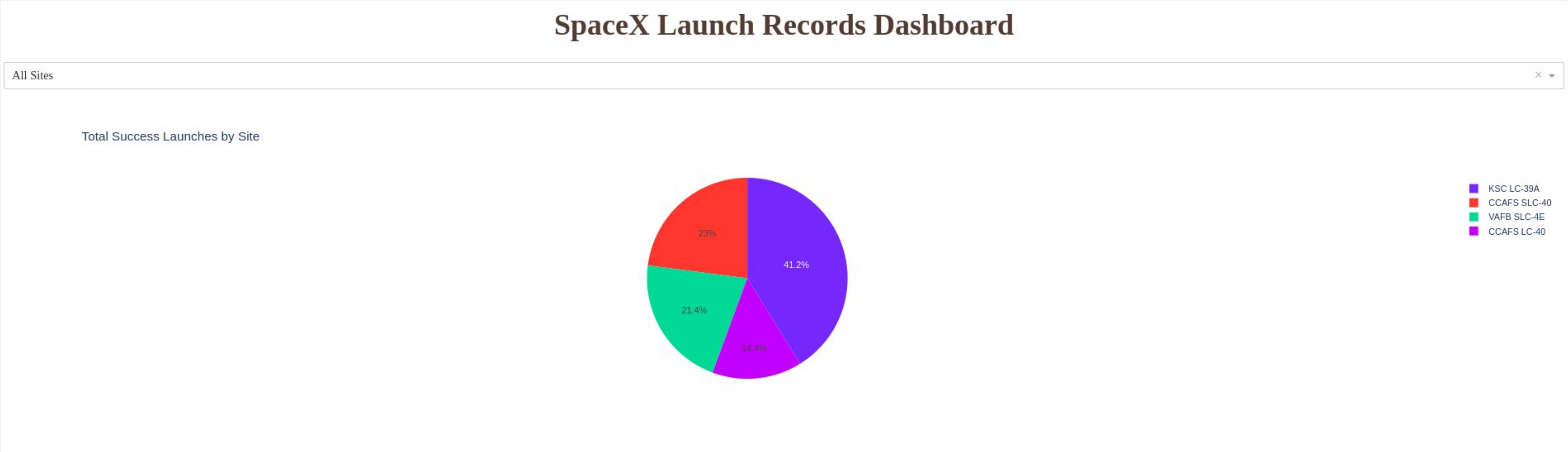
Section 4

Build a Dashboard with Plotly Dash

Launch Success by Site

Success as Percent of Total

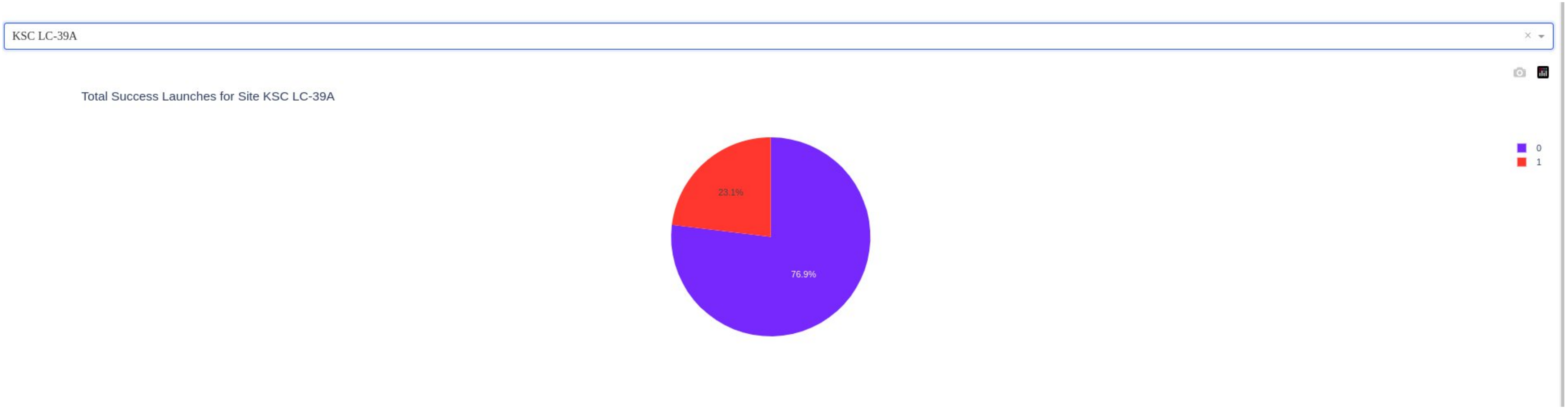
• KSC LC-39A has the most successful launches amongst launch sites (41.2%)



Launch Success (KSC LC-29A)

Success as Percent of Total

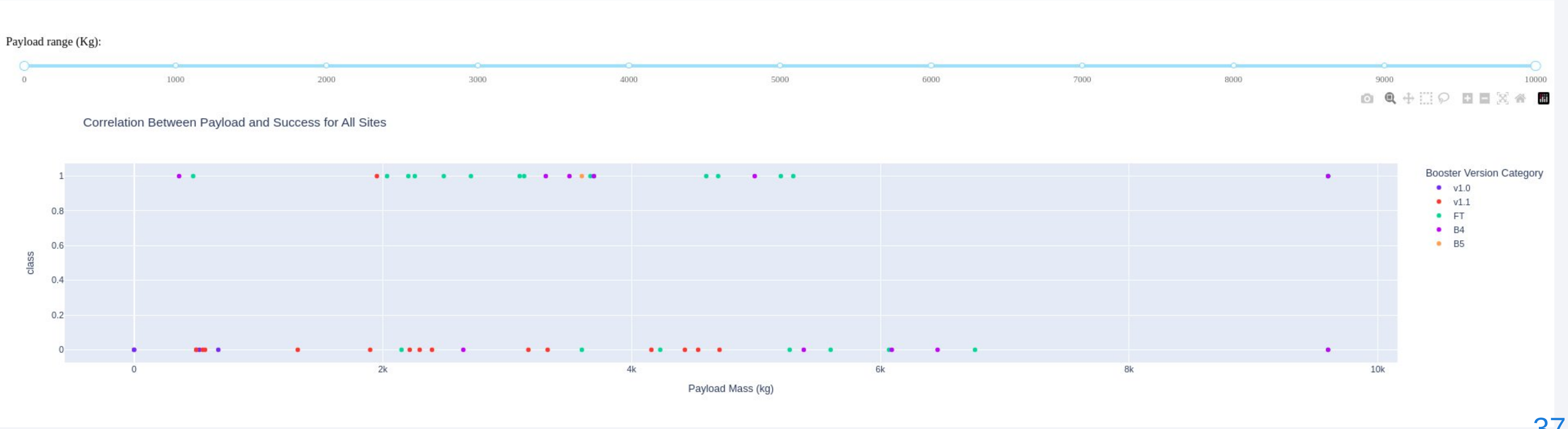
- KSC LC-39A has the highest success rate amongst launch sites (76.9%)
- 10 successful launches and 3 failed launches



Payload Mass and Success

By Booster Version

- Payloads between2,000 kg and 5,000 kg have the highest success rate
- 1 indicating successful outcome and 0 indicating an unsuccessful outcome



Section 5

Predictive Analysis (Classification)

Classification Accuracy

Accuracy

•All the modelsperformed at about the same level and had the same scores and accuracy. This is likely due to the small dataset. The Decision Tree modelslightly outperformed the restwhenlookingat.best_score_

• .best_score_ is the average of all cv folds for a single combination of the parameters

	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

Confusion Matrix

Performance Summary

•A confusion matrix summarizes the performance of a classification algorithm

•All the confusion matrices were identical

•The fact that there are false positives (Type 1 error) is not good

Confusion Matrix Outputs:

• 12 True positive

• 3 True negative

• 3 False positive

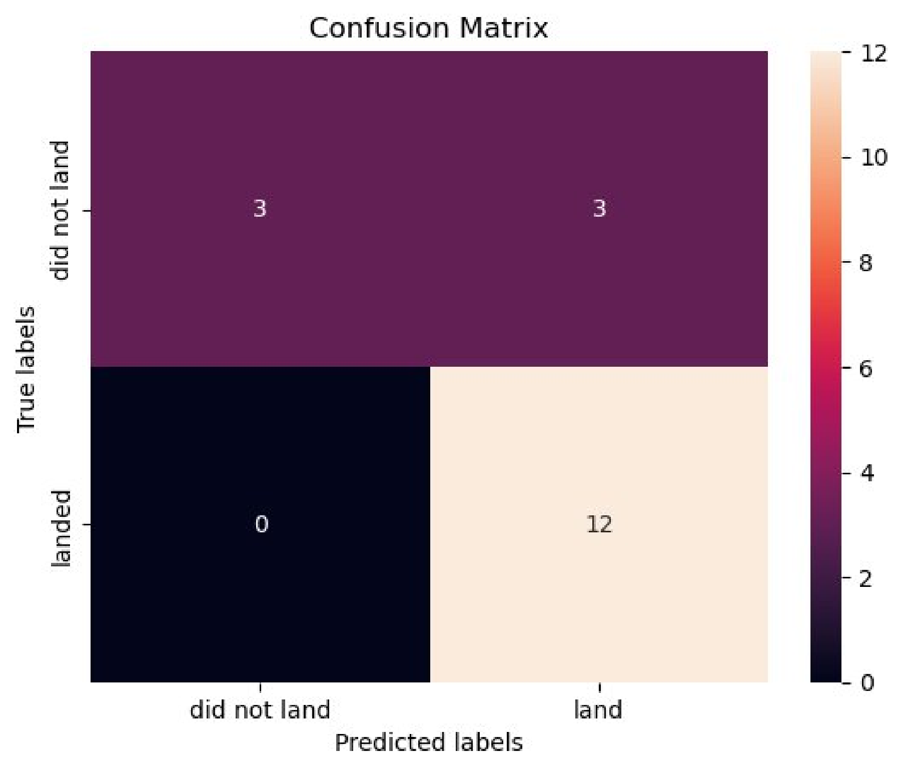
• 0 False Negative

• Precision= TP / (TP + FP) = 12 / 15 = .80

•Recall= TP / (TP + FN) = 12 / 12 = 1

• F1 Score= 2 * (Precision * Recall) / (Precision + Recall) = 2 * (.8 * 1) / (.8 + 1) = .89

• Accuracy= (TP + TN) / (TP + TN + FP + FN) = .833



Conclusions

- **Model Performance:** The models performed similarly on the test set with the decision tree model slightly outperforming
- **Equator:** Most of the launch sites are near the equator for an additional natural boost -due to the rotational speed of earth -which helps save the cost of putting in extra fuel and boosters
- **Coast:** All the launch sites are close to the coast
- **Launch Success:** Increases over time
- **KSC LC-39A:** Has the highest success rate among launch sites. Has a 100% success rate for launches less than 5,500 kg
- **Orbits:** ES-L1, GEO, HEO, and SSO have a 100% success rate
- **Payload Mass:** Across all launch sites, the higher the payload mass (kg), the higher the success rate

…

Thank you!