IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Saw Eh Tar Moo
26th April, 2025

Outline

Executive Summary

Introduction

Methodology

Results

Conclusion

Appendix

Executive Summary

Summary of methodologies

Data Collection and Wrangling

Exploratory Data Analysis

Interactive Visual Analytics & Dashboard

Predictive Analytics

Presentation of Data Driven Insights

Summary of all results

The project identified critical factors affecting Falcon 9 first-stage landing success, success rates at various launch sites and for different orbit types, and trends in launch outcomes over time.

Among the tested models, Decision Trees achieved a peak accuracy of 93.33%, while SVM with rbf kernels surpassed linear SVM, and Logistic Regression established a baseline for comparison.

These insights offer a competing aerospace company valuable information for assessing potential market entry.

Furthermore, the project generated data-driven recommendations concerning launch strategies and risk management, ultimately improving the understanding of elements contributing to cost reduction via successful rocket landings.

Introduction

Introduction

Project background

SpaceX has revolutionized space transportation with its reusable rocket technology, reducing the cost of spaceflight. Analyzing SpaceX launch data is crucial for understanding mission success and improving future operations. This project addresses the need for data-driven insights in the aerospace industry.

Problems Statements

This project aims to answer the following questions:

- What factors influence the success of SpaceX Falcon 9 rocket launches and landings?
- Can we accurately predict launch outcomes based on historical data?

Introduction

Objectives

1. Collect and wrangle SpaceX launch data from various sources.
2. Explore the data to identify key relationships between different variables and launch outcomes.
3. Develop interactive visualizations to effectively present findings to stakeholders.
4. Build and evaluate classification models to predict launch outcomes (success or failure).

Methodology

Methodology

Executive Summary

Data collection

Data wrangling

Exploratory data analysis (EDA) using visualization and SQL

Interactive visual analytics using Folium and Plotly Dash

Predictive analysis using classification models

Data Collection – SpaceX API

Data on SpaceX launches, payloads, rockets, and launch sites was gathered through the SpaceX API and supplemented with historical launch outcome data obtained by web scraping relevant web pages.

Start → Connect to SpaceX API (requests) → Send GET request to endpoints:
- /launches
- /payloads
- /rockets
- /launchpads

→ Receive JSON responses → Parse and normalize JSON (pandas.json_normalize) → Save data into DataFrames → Initial analysis for missing launch outcome info → End

The Github URL of Complete API call for Data Collection

[here]

Data Collection - Scraping

Additional data, such as historical launch outcomes, was collected by scraping relevant web pages. This involved identifying the target website, inspecting the HTML structure, and using libraries like Beautiful Soup to extract the required information.

Start → Identify and target trusted websites(Wikipedia) → Send HTTP GET request (requests) → Fetch raw HTML content → Parse HTML (BeautifulSoup)

→ Locate and extract:
- Launch dates
- Rocket names
- Mission outcomes

→ Clean and organize extracted data → Match scraped data with SpaceX API data (e.g., by date, rocket name) → Merge into unified final dataset → End

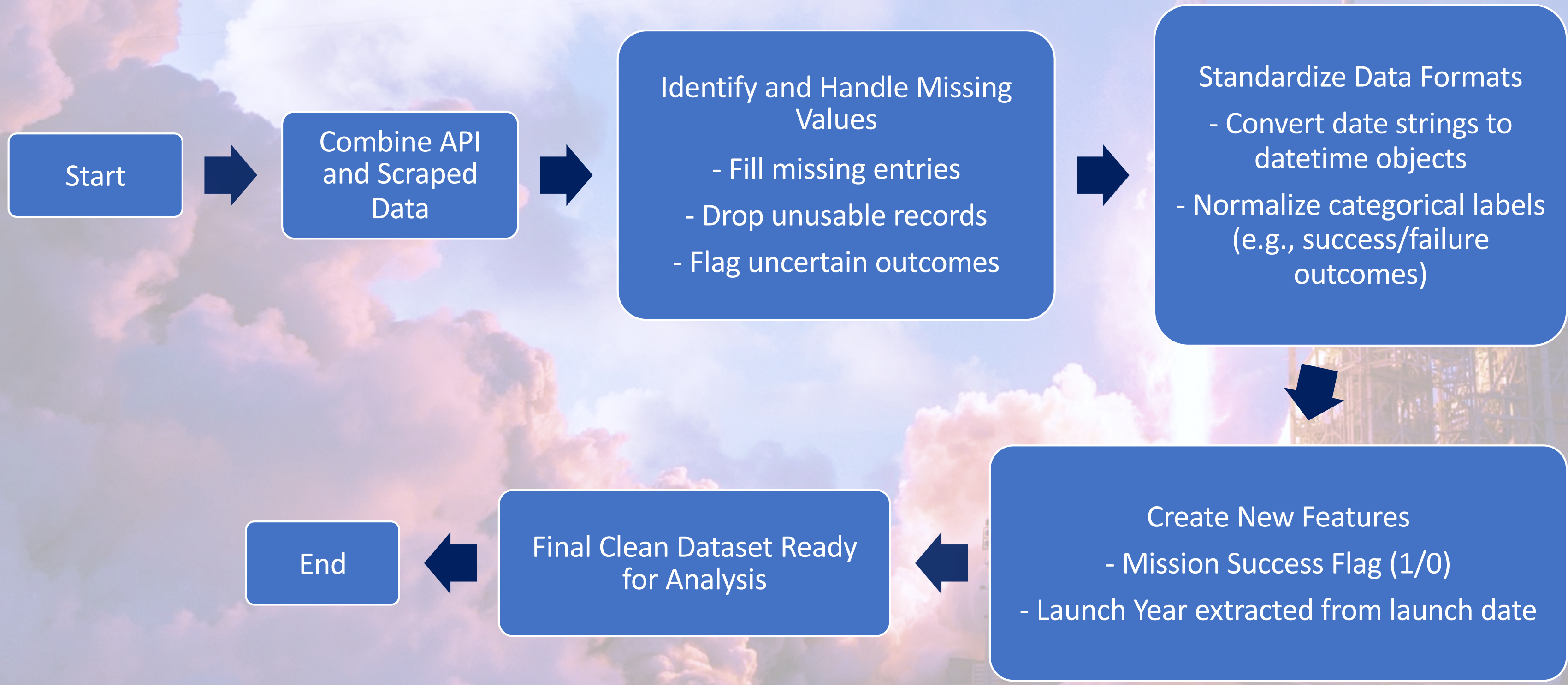
The Github URL of Complete API call for Data Collection

here

Data Wrangling

The collected data was cleaned and transformed to ensure consistency and suitability for analysis. This involved handling missing values, standardizing data formats, and creating new features.

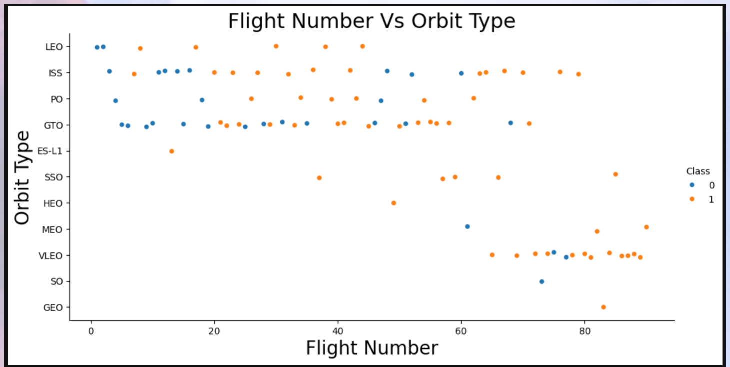
Github URL of Complete Data Wrangling notebook [here]

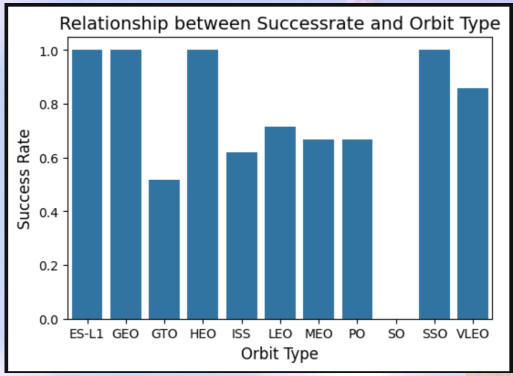


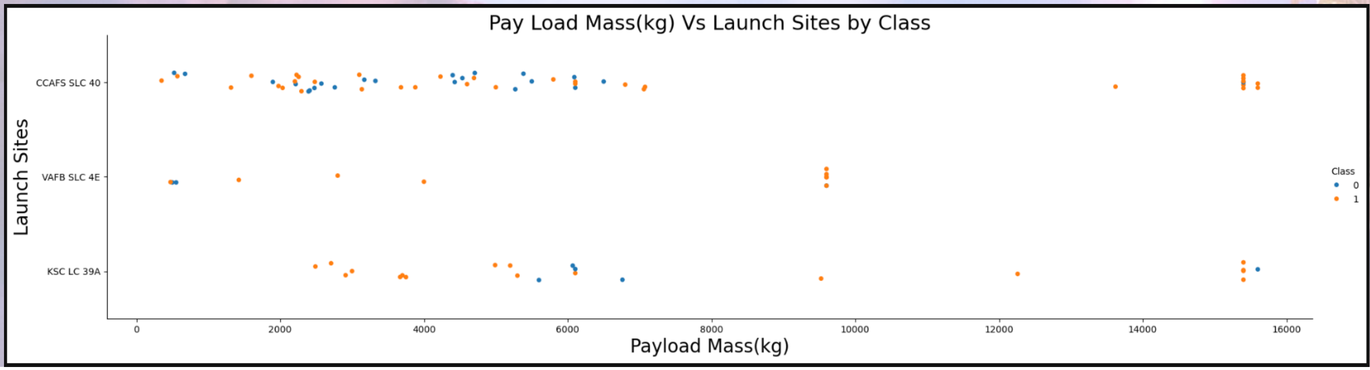
EDA with Data Visualization

EDA was performed to understand the characteristics of the data and identify potential relationships between variables.

[Github URL of Data Visualization with EDA]



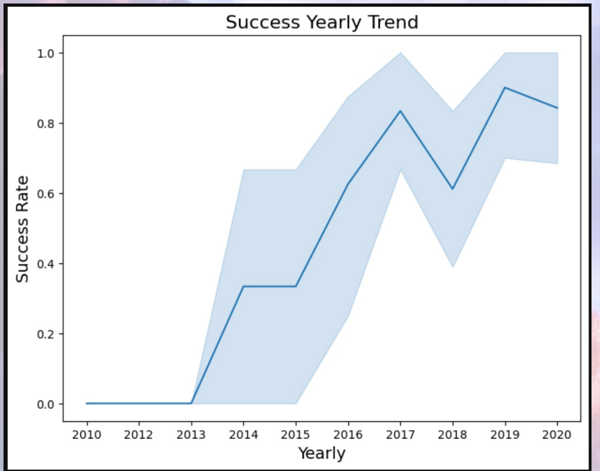


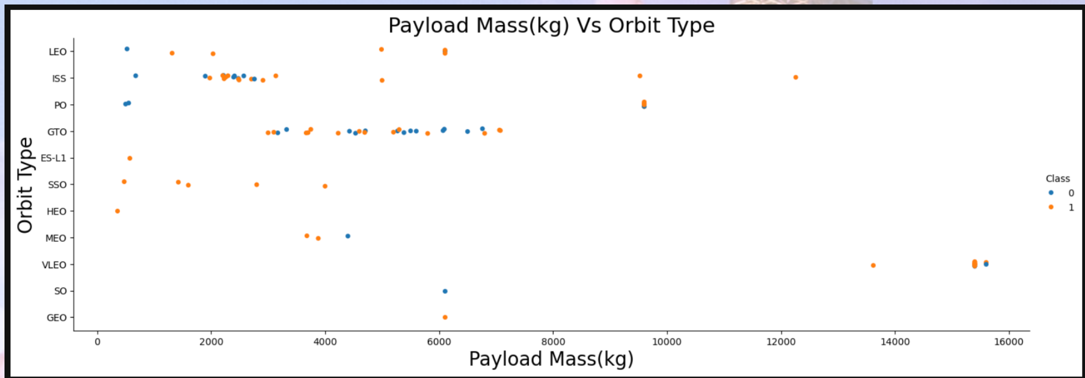


EDA with Data Visualization

EDA was performed to understand the characteristics of the data and identify potential relationships between variables.

[Github URL of Data Visualization with EDA]





EDA with SQL

SQL queries were used to extract and aggregate data from a relational database. These queries provided insights into specific aspects of the data, such as launch site statistics and payload information.

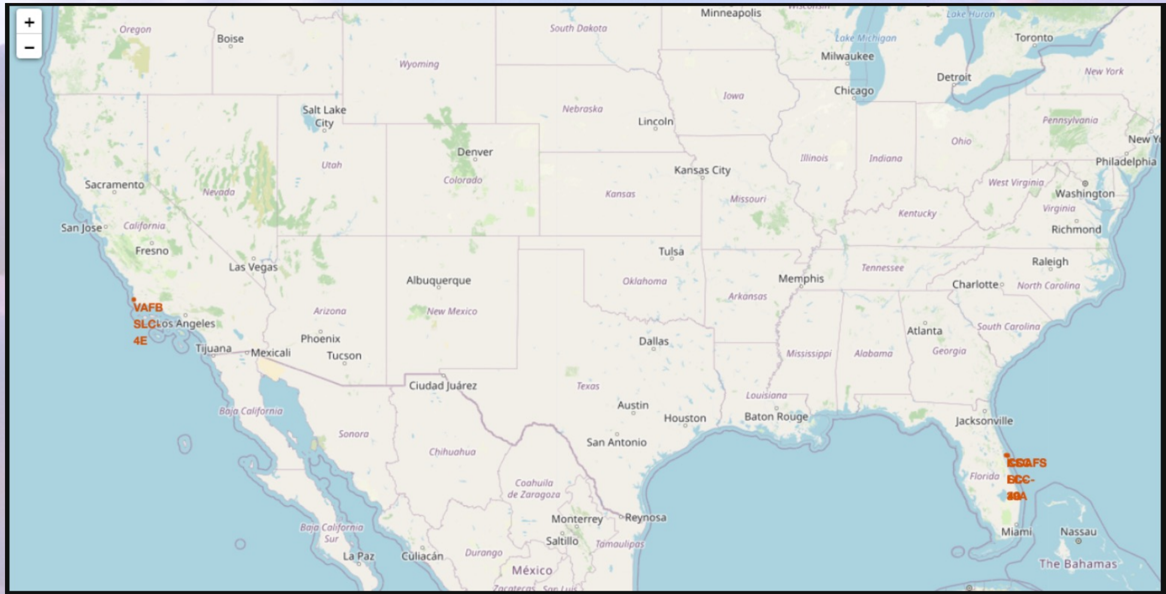
[Github URL of SQL quiries]

- Display the names of the unique launch sites in the space mission
- Display 5 records where launch sites begin with the string 'CCA'
- Display the total payload mass carried by boosters launched by NASA (CRS)
- List the date when the first succesful landing outcome in ground pad was acheived.
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- List the total number of successful and failure mission outcomes
- List all the booster_versions that have carried the maximum payload mass. Use a subquery.
- List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order.

Build an Interactive Map with Folium

A Folium map was created to visualize launch site locations and launch outcomes. Markers, circles, and lines were used to represent different aspects of the data, such as launch site coordinates, launch success/failure, and rocket trajectories.

[Github URL for Visualization with Folium]()



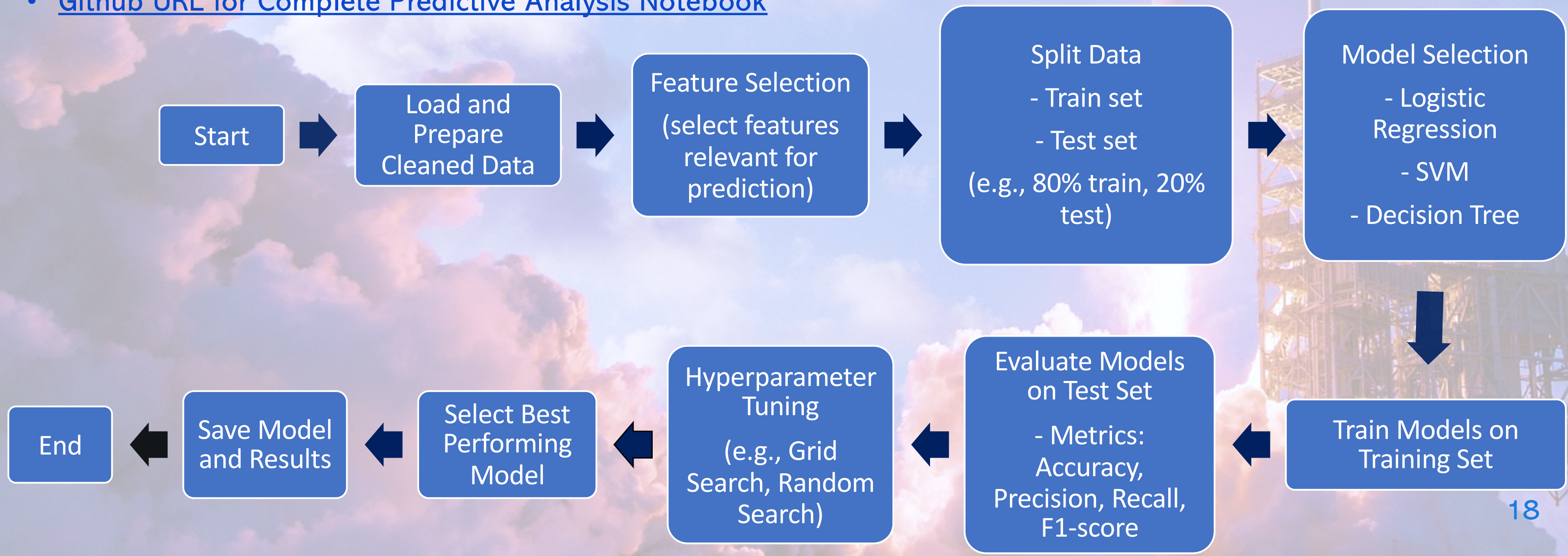
Build a Dashboard with Plotly Dash

A Plotly Dash dashboard was developed to provide an interactive overview of the launch data. The dashboard includes various plots and graphs, such as pie charts and scatter plots, along with interactive elements like dropdown menus and sliders.

Github URL for Dash Application dashboard

Predictive Analysis (Classification)

- Classification models were built to predict launch outcomes (success or failure) based on historical data. This involved selecting relevant features, training different models, evaluating their performance, and tuning their parameters.
- The process included: data splitting, model selection (e.g., Logistic Regression, SVM, Decision Tree), training, evaluation (using metrics like accuracy, precision, recall, F1-score), and hyperparameter tuning to find the best-performing model.
- [Github URL for Complete Predictive Analysis Notebook]()



Results

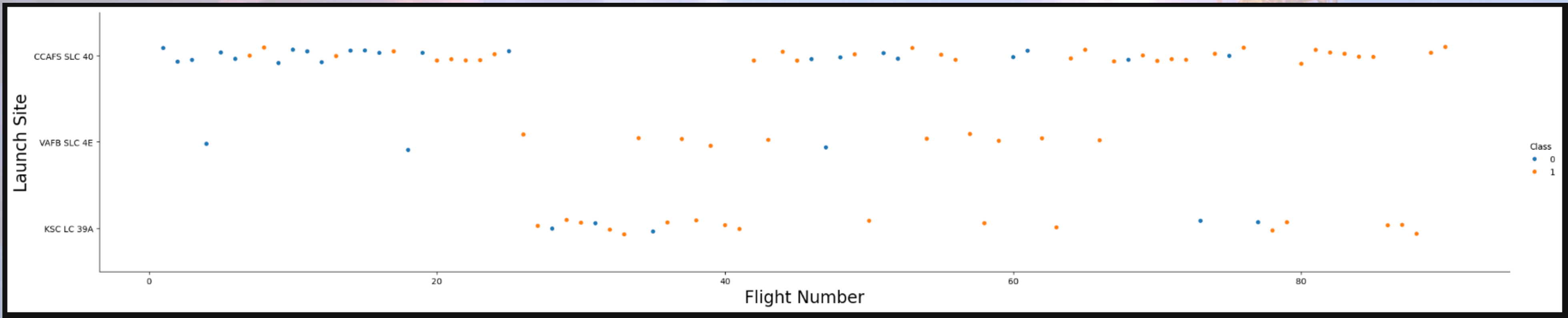
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results


Section 2

Flight Number vs. Launch Site

This visualization shows if there is any correlation between the flight number and the launch site.

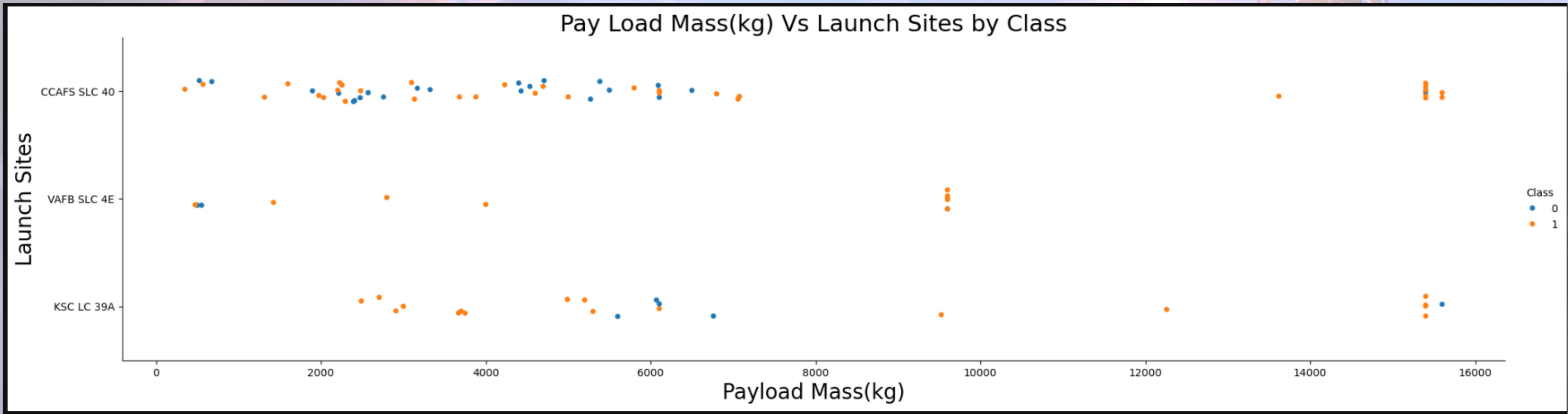
- CCSFS SLC-40 hosted the most launches with a seemingly mixed record of mission outcomes throughout its flight history in this dataset.
- KSC LC-39A also experienced a significant number of launches with a visible distribution of both mission outcome categories.
- VAFB SLC-4E had the fewest launches among the three sites and appears to have a potentially skewed distribution towards one of the mission outcome categories.



Payload vs. Launch Site

This visualization helps understand if the payload mass varies significantly between different launch sites.

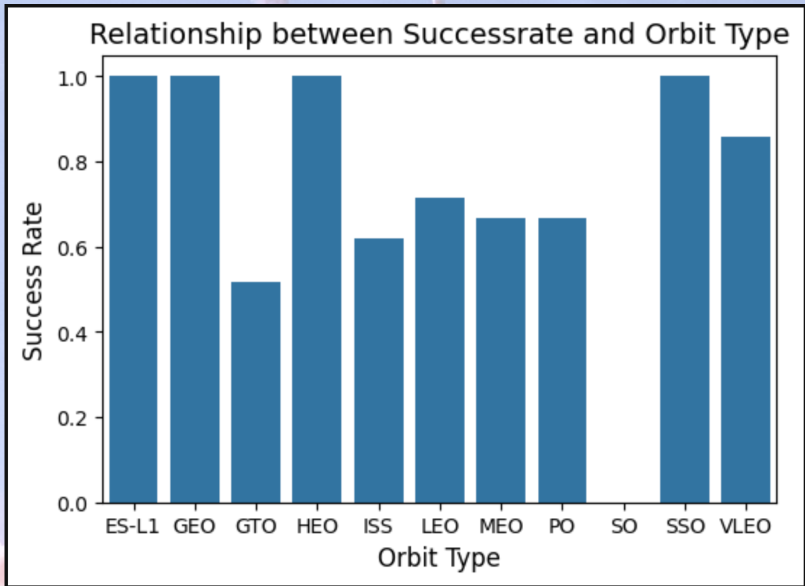
- CCSFS SLC-40 accommodated launches across a wide range of payload masses, with both mission outcome categories (0 and 1) observed throughout this range.
- VAFB SLC-4E saw launches with relatively lower payload masses compared to the other sites, and the outcomes appear to be predominantly of one class within the observed payload range.
- KSC LC-39A supported launches with a more varied range of payload masses, showing a mix of both mission outcome categories across those payloads.



Success Rate vs. Orbit Type

This chart illustrates the success rate for each orbit type, highlighting which orbits have historically been more successful.

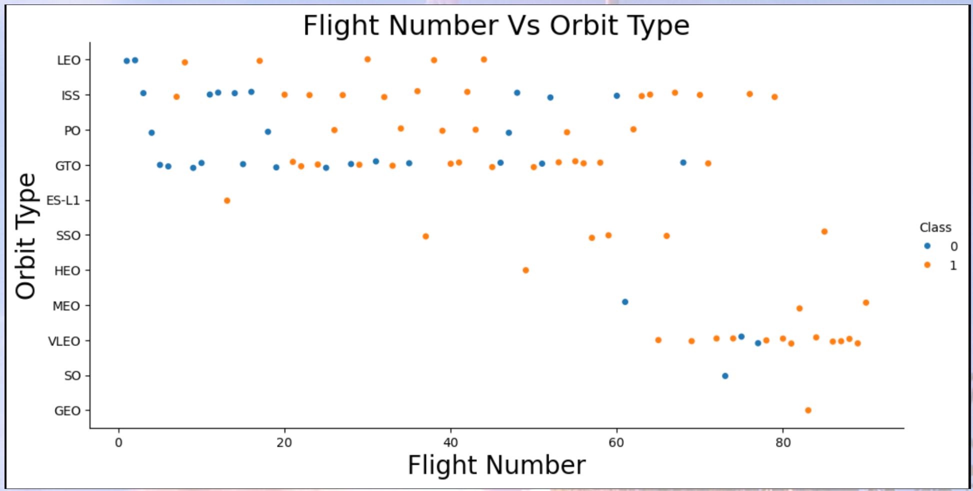
- ES-L1, GEO, HEO, and SSO orbits saw almost perfect success, meaning nearly all missions to these orbits were successful.
- GTO orbit had a lower success rate, with only about half of the missions reaching this orbit successfully.
- ISS orbit also had a moderate success rate, with roughly six out of ten missions succeeding.
- LEO, MEO, and PO orbits showed a similar moderate success rate, with about two-thirds of the missions being successful.
- SO orbit had a perfect success rate, indicating all missions to this orbit were successful.
- VLEO orbit had a good success rate, with over eighty percent of its missions being successful.



Flight Number vs. Orbit Type

This visualization shows if there is any correlation between the flight number and the orbit type.

- LEO and ISS orbits were common targets throughout the flights. GTO orbit was also frequently used.
- Some orbits like MEO and VLEO became more common later.
- Other orbits had fewer missions overall.
- Mission outcomes (success/failure) seem mixed across all orbits and flight numbers.



Payload vs. Orbit Type

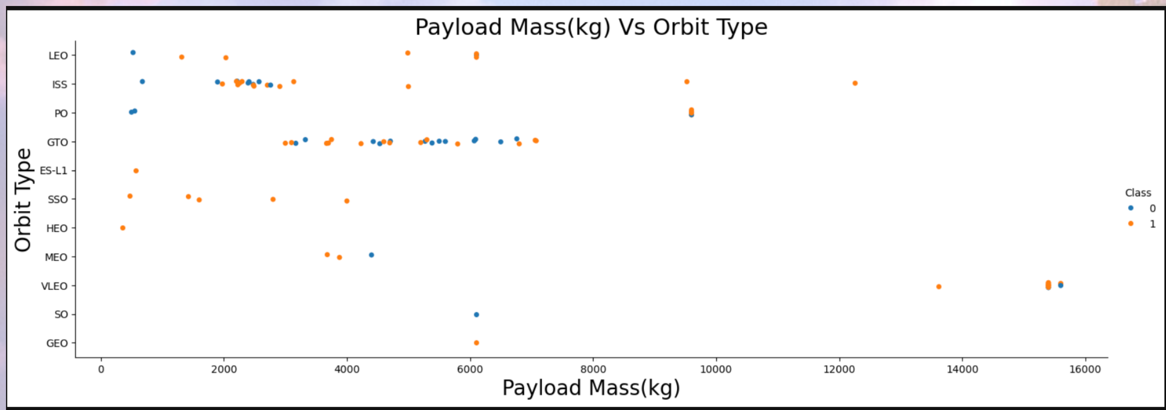
This visualization helps in understanding the relationship between payload mass and the type of orbit.

LEO and ISS orbits accommodated a wide range of payload masses, from lighter to heavier. GTO orbit also saw launches with varying payload weights, generally on the lighter to medium side.

HEO and SSO orbits appear to have handled mostly lighter payloads.

MEO and VLEO orbits were used for some of the heaviest payloads in this dataset.

The mission outcomes (indicated by color) are distributed across different payload masses within each orbit type.



Launch Success Yearly Trend

This chart shows how the success rate has changed over the years, indicating any improvements or trends.

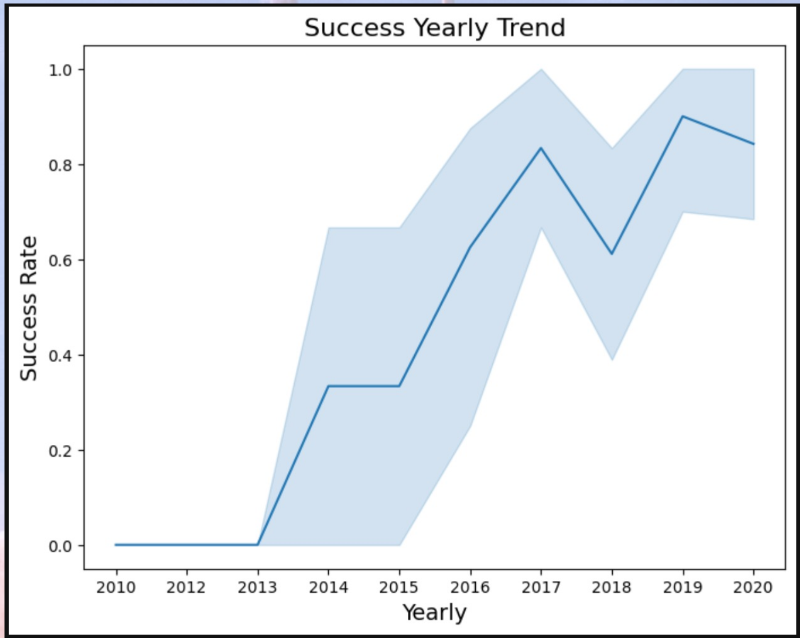
The success rate was zero in the early years (2010-2013).

There was a noticeable increase in success rate around 2014.

The success rate fluctuated in the mid-2010s, showing both increases and decreases.

There's a trend of generally high success rates in the later years (around 2017-2020), although there's a dip in 2018.

The shaded area around the line likely represents the uncertainty or variability in the success rate for each year.



All Launch Site Names

- Finding the names of the unique launch sites

```
In [11]: %sql SELECT DISTINCT Launch_Site FROM SPACEXTABLE

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

Out[11]:

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

Launch Site Names Begin with 'CCA'

- Finding 5 records where launch sites begin with `CCA`

```
In [12]: %sql SELECT * FROM SPACEXTABLE WHERE Launch_Site LIKE "CCA%" LIMIT 5

 * sqlite:///my_data1.db
Done.
Out[12]:
```

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX
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
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)

Total Payload Mass

- Calculating the total payload carried by boosters from NASA

```
In [15]: %sql SELECT SUM(PAYLOAD_MASS__KG_) AS Total_payload_mass FROM SPACEXTABLE WHERE Customer LIKE "%NASA (CRS)%"

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

Out[15]:

Total_payload_mass
48213

Average Payload Mass by F9 v1.1

- Calculating the average payload mass carried by booster version F9 v1.1

```
In [17]: %sql SELECT AVG(PAYLOAD_MASS__KG_) AS average_payload_mass FROM SPACEXTABLE WHERE Booster_Version LIKE "%F9 v1.1%

 * sqlite:///my_data1.db
Done.

Out[17]:
```

average_payload_mass
2534.6666666666665

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad

```
In [19]: %sql SELECT MIN(Date) AS successful_landing_groundpad FROM SPACEXTABLE WHERE Landing_outcome LIKE "%Success (grou

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

Out[19]:

successful_landing_groundpad
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Listing the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

```
In [21]: %sql SELECT Booster_Version FROM SPACEXTABLE WHERE Landing_Outcome LIKE "%Success (drone ship)%" AND PAYLOAD_MASS_

 * sqlite:///my_data1.db
Done.
Out[21]:
```

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

Total Number of Successful and Failure Mission Outcomes

- Calculating the total number of successful and failure mission outcomes

```
In [23]: %sql SELECT Mission_Outcome, COUNT(*) AS total_successful_failure FROM SPACEXTABLE GROUP BY Mission_Outcome

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

Out[23]:

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

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass

```
In [25]: %sql SELECT Booster_Version, PAYLOAD_MASS__KG_ FROM SPACEXTABLE WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MAS

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

Out[25]:

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

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

- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

```
In [30]: %sql SELECT Landing_Outcome, COUNT(*) as Count_Landing_Outcome From SPACEXTABLE WHERE (Landing_Outcome LIKE '%Fai

 * sqlite:///my_data1.db
Done.
Out[30]:
```

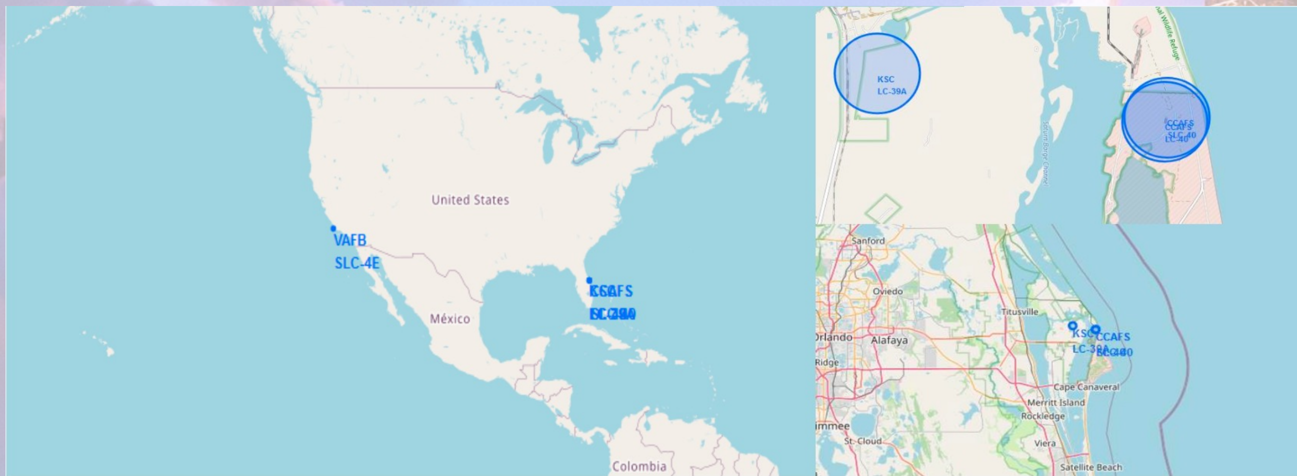
Landing_Outcome	Count_Landing_Outcome
Failure (drone ship)	5
Success (ground pad)	3

Section 3

Folium Map: Launch Sites.

All site locations are near the coast and Equator line, SpaceX focuses on locations that are close to water and the zeroth latitude for the purpose of avoiding any undesired accidents.

The launch sites are distributed in two states California and Florida



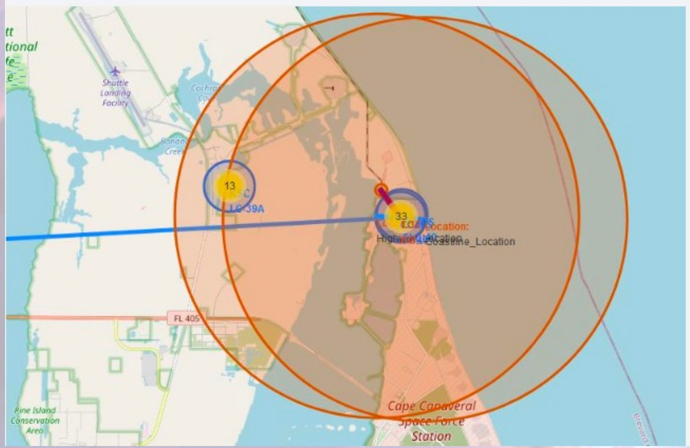
Folium Map: Success rate for each launch location

From the color-labeled markers in marker clusters, we can easily identify which launch sites have relatively high success rates.

Green Marker = Successful Return

Red Marker = Failed Return

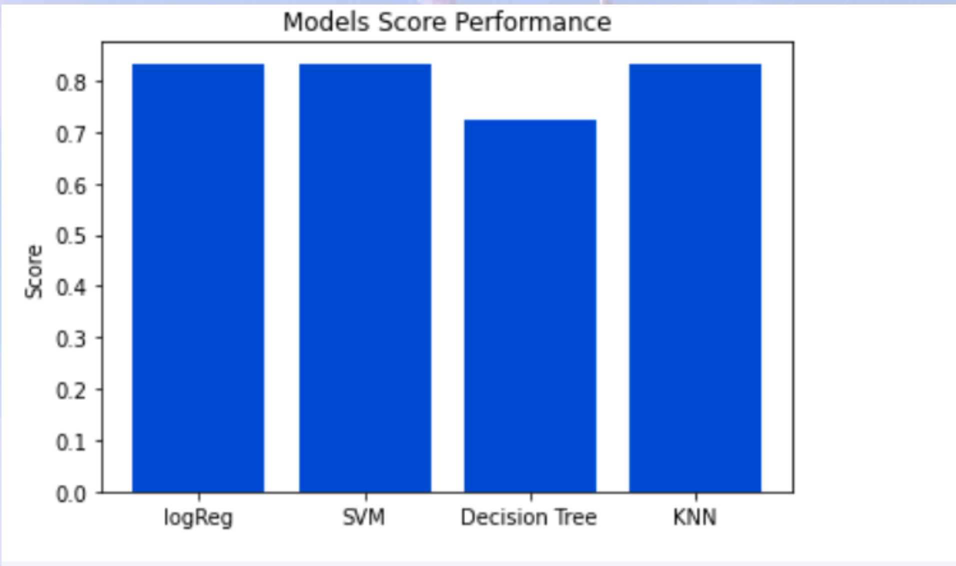
Folium Map: Closest Proximities to CCAFS LC-40




Section 5

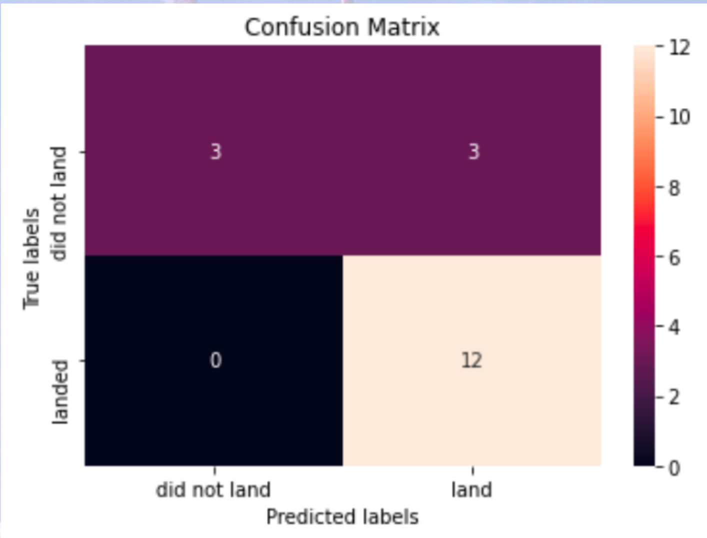
Classification Accuracy

- Logistic Regression , SVM and KNN has the same performance where the Jaccard Score is same : 0.8
- Where Decision Tree has the worst performance compared to other models.



Confusion Matrix

- Logistic Regression , SVM and KNN have the same confusion matrix and results:
- • True Positive = 12
- • False Positive = 0
- • True Negative = 3
- • False Negative = 3



Conclusions

In summary, this project analyzed SpaceX launch data to identify key factors influencing mission success and to develop a predictive model for launch outcomes.

Key findings:

- Certain launch sites may have a higher success rate than others.
- The type of orbit and payload mass can affect launch outcomes.
- The classification model can predict launch outcomes with a certain degree of accuracy.

Implications:

- These findings can help SpaceX to optimize its launch operations and improve mission success rates.
- The predictive model can be used to assess the risk associated with future launches.
- For a competing company: This analysis provides insights into SpaceX's successful strategies and areas where they excel. It can inform their own launch operations, strategy, and technology development to become more competitive.

Future directions:

- Further research could explore the impact of other factors, such as weather conditions and rocket reusability, on launch outcomes.
- The predictive model could be improved by incorporating more data and advanced machine learning techniques.

Appendix

SpaceX API URL "Click Here"

• SpaceX Static Wikipedia URL "Click Here"

• • SpaceX data used in ML training "Click Here"


Thank you