IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Mohammed Lubbad

14 June 2024

Outline

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

Executive Summary

Objective:

Predict if the Falcon 9 first stage will land successfully to determine the cost of a launch.

Methodologies:

Data collection via API and web scraping, data wrangling, exploratory data analysis (EDA), interactive visual analytics, and predictive modeling.

Results:

Insights on launch success rates, interactive dashboards, and a machine learning model to predict landing success.

Introduction

- **Project Background:**

SpaceX's cost-effective rocket launches due to reusable first stages.

- **Problem Statement:**

Determine if the first stage will land successfully to estimate launch costs.
Competing with SpaceX by predicting landing success using machine learning.

Section 1

Methodology

Methodology

Executive Summary

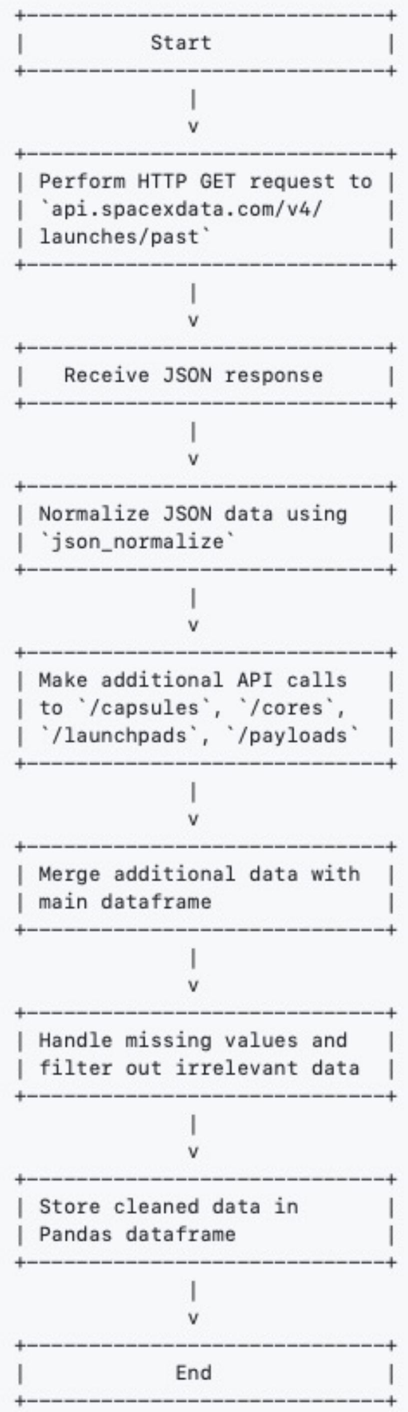
- Data collection methodology:
  - SpaceX REST API and Wikipedia web scraping were used.
- Perform data wrangling
  - Handling missing values, filtering for Falcon 9 launches, and converting categorical variables using one-hot encoding.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

- **Sources:** SpaceX REST API and Wikipedia web scraping.
- **API Endpoint:** api.spacexdata.com/v4/launches/past
- **Web Scraping:** Wikipedia page for Falcon 9 and Falcon Heavy launches.
- **Tools:** Python libraries requests, BeautifulSoup, and pandas.

Data Collection – SpaceX API

- The GitHub URL of the completed SpaceX API calls notebook [(https://github.com/mlubbad/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/jupyter-labs-spacex-data-collection-api-2.ipynb)](https://github.com/mlubbad/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/jupyter-labs-spacex-data-collection-api-2.ipynb)



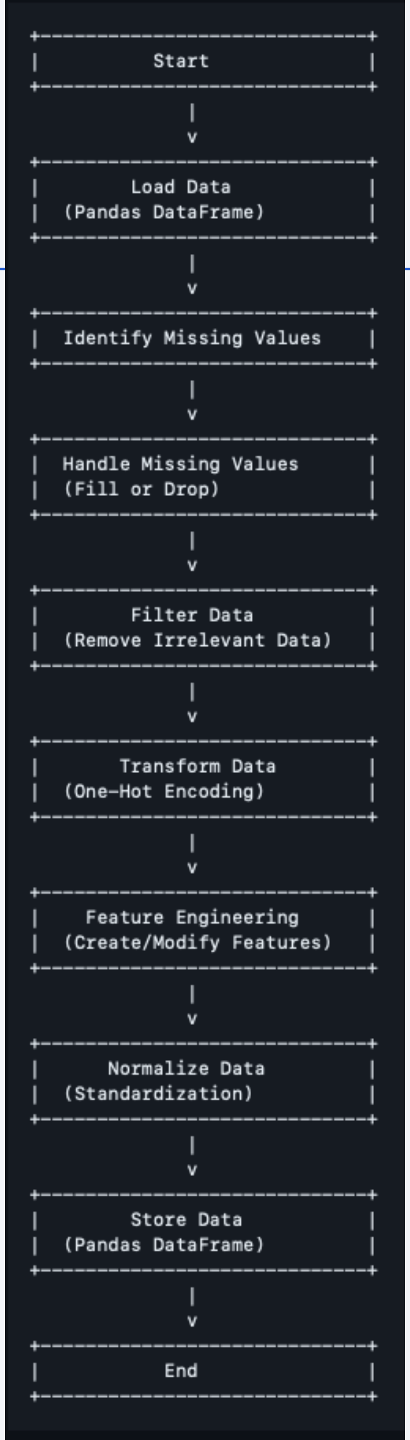
Data Collection - Scraping

- The GitHub URL of the completed web scraping notebook, (https://github.com/mlubbad/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/jupyter-labs-webscraping.ipynb)

```
+-----------------------------+
|            Start            |
+-----------------------------+
              |
              v
+-----------------------------+
| Send HTTP GET request to    |
| Wikipedia page URL          |
| `https://en.wikipedia.org/  |
| wiki/List_of_Falcon_9_and_  |
| Falcon_Heavy_launches`      |
+-----------------------------+
              |
              v
+-----------------------------+
| Receive HTML response       |
+-----------------------------+
              |
              v
+-----------------------------+
| Parse HTML content using    |
| BeautifulSoup               |
+-----------------------------+
              |
              v
+-----------------------------+
| Locate the HTML table       |
| containing launch records   |
+-----------------------------+
              |
              v
+-----------------------------+
| Extract table data and      |
| convert to Pandas dataframe |
+-----------------------------+
              |
              v
+-----------------------------+
| Clean and preprocess data   |
| (handle missing values,     |
| convert data types, etc.)   |
+-----------------------------+
              |
              v
+-----------------------------+
| Store cleaned data in       |
| Pandas dataframe            |
+-----------------------------+
              |
              v
+-----------------------------+
|             End             |
+-----------------------------+
```

Data Wrangling

- Data Loading: Load the raw data into a Pandas data frame.
- Handling Missing Values: Identify and handle missing values in the dataset.
- Data Filtering: Filter out irrelevant data, such as non-Falcon 9 launches.
- Data Transformation: Convert categorical variables into numerical values using techniques like one-hot encoding.
- Feature Engineering: Create new features or modify existing ones to improve the dataset's quality.
- Data Normalization: Standardize the data to ensure all features contribute equally to the model.
- Data Storage: Store the cleaned and processed data in a Pandas data frame for further analysis.
- The GitHub URL of your completed data wrangling-related notebooks [(https://github.com/mlubbad/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb)](https://github.com/mlubbad/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb).



EDA with Data Visualization

During the Exploratory Data Analysis (EDA) phase, various charts were plotted to uncover patterns, trends, and insights from the SpaceX launch data. Below is a summary of the charts that were plotted and the reasons for using them:

1. **Scatter Plot: Flight Number vs. Payload Mass**: A scatter plot was created with Flight Number on the x-axis and Payload Mass on the y-axis. To observe the relationship between the number of flight attempts and the payload mass. This helps in understanding if there is a trend in payload mass over successive launches.

2. **Bar Chart: Launch Site Success Rates**: A bar chart was plotted showing the success rates of launches for each launch site. To compare the success rates across different launch sites. This helps in identifying which launch sites have higher success rates and are more reliable.

3. **Line Chart: Success Rate Over Time**: A line chart was plotted with the year on the x-axis and the success rate on the y-axis. To visualize the trend in success rates over time. This helps in understanding if the success rate has improved, remained stable, or declined over the years.

4. **Pie Chart: Outcome Distribution**: A pie chart was created to show the distribution of different outcomes (e.g., successful landing, failed landing, etc.). To provide a quick visual summary of the proportion of different outcomes. This helps in understanding the overall success rate and the frequency of different types of outcomes.

EDA with Data Visualization

5. **Histogram: Payload Mass Distribution**: A histogram was plotted to show the distribution of payload mass. To understand the distribution and range of payload masses. This helps in identifying common payload mass ranges and any outliers.

6. **Box Plot: Payload Mass by Orbit Type**: A box plot was created to show the distribution of payload mass for different orbit types. To compare the payload mass distributions across different orbit types. This helps in understanding if certain orbits are associated with heavier or lighter payloads.

7. **Heatmap: Correlation Matrix:** A heatmap was plotted to show the correlation matrix of different numerical features. To identify the strength and direction of relationships between different numerical features. This helps in selecting features for predictive modeling.

- Add the GitHub URL of your completed EDA with data visualization notebook, (https://github.com/mlubbad/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/edadataviz.ipynb)

EDA with SQL

- **Select All Data**: SELECT * FROM launches;
- **Count Total Launches**: SELECT COUNT(*) FROM launches;
- **Filter by Rocket Type**: SELECT * FROM launches WHERE rocket = 'Falcon 9';
- **Group by Launch Site:** SELECT launch_site, COUNT(*) FROM launches GROUP BY launch_site;
- **Calculate Success Rate**: SELECT launch_site, AVG(success) FROM launches GROUP BY launch_site;
- **Find Maximum Payload**: SELECT MAX(payload_mass) FROM launches;
- **Join with Payload Table**: SELECT launches.*, payloads.mass FROM launches JOIN payloads ON launches.payload_id = payloads.id;
- **Order by Date:** SELECT * FROM launches ORDER BY date DESC;
- Add the GitHub URL of your completed EDA with SQL notebook, (https://github.com/mlubbad/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/jupyter-labs-eda-sql-coursera_sqllite-2.ipynb)

Build an Interactive Map with Folium

Folium Map Objects Summary:

- **Markers:** Pinpoint exact launch site locations. To identify and differentiate between launch sites.
- **Circle Markers:** Highlight areas around launch sites. To provide a sense of scale and proximity.
- **Lines**: Connect launch sites to nearby points of interest. To visualize distances and relationships.
- **Popups:** Display additional information about launch sites. To enhance interactivity and informativeness.
- **Marker Clusters:** Group nearby markers to avoid clutter. To improve map readability.
- The GitHub URL of your completed interactive map with Folium map, (https://github.com/mlubbad/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/lab_jupyter_launch_site_location.ipynb).

Build a Dashboard with Plotly Dash

Dashboard Plots/Graphs and Interactions Summary Plots/Graphs Added:

- Success-Pie-Chart: To show the success rate of launches for the selected site. Provides a quick visual summary of the success rate for the selected launch site, helping users to easily understand the performance of different sites.
- Success-Payload-Scatter-Chart: To display the relationship between payload mass and launch success for the selected payload range. Helps in visualizing how payload mass affects the success rate, allowing users to identify trends and patterns.

Interactions Added:

- Launch Site Dropdown: Allows users to select a specific launch site to filter the data. Enhances interactivity by allowing users to focus on data from a specific launch site, making the analysis more targeted.
- Payload Range Slider: Enables users to filter the data based on the payload range. Allows users to explore the impact of different payload ranges on launch success, providing deeper insights into the data.
- The **GitHub URL** of your completed **Plotly Dash lab**, (https://github.com/mlubbad/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/spacex_dash_app.py).

Predictive Analysis (Classification)

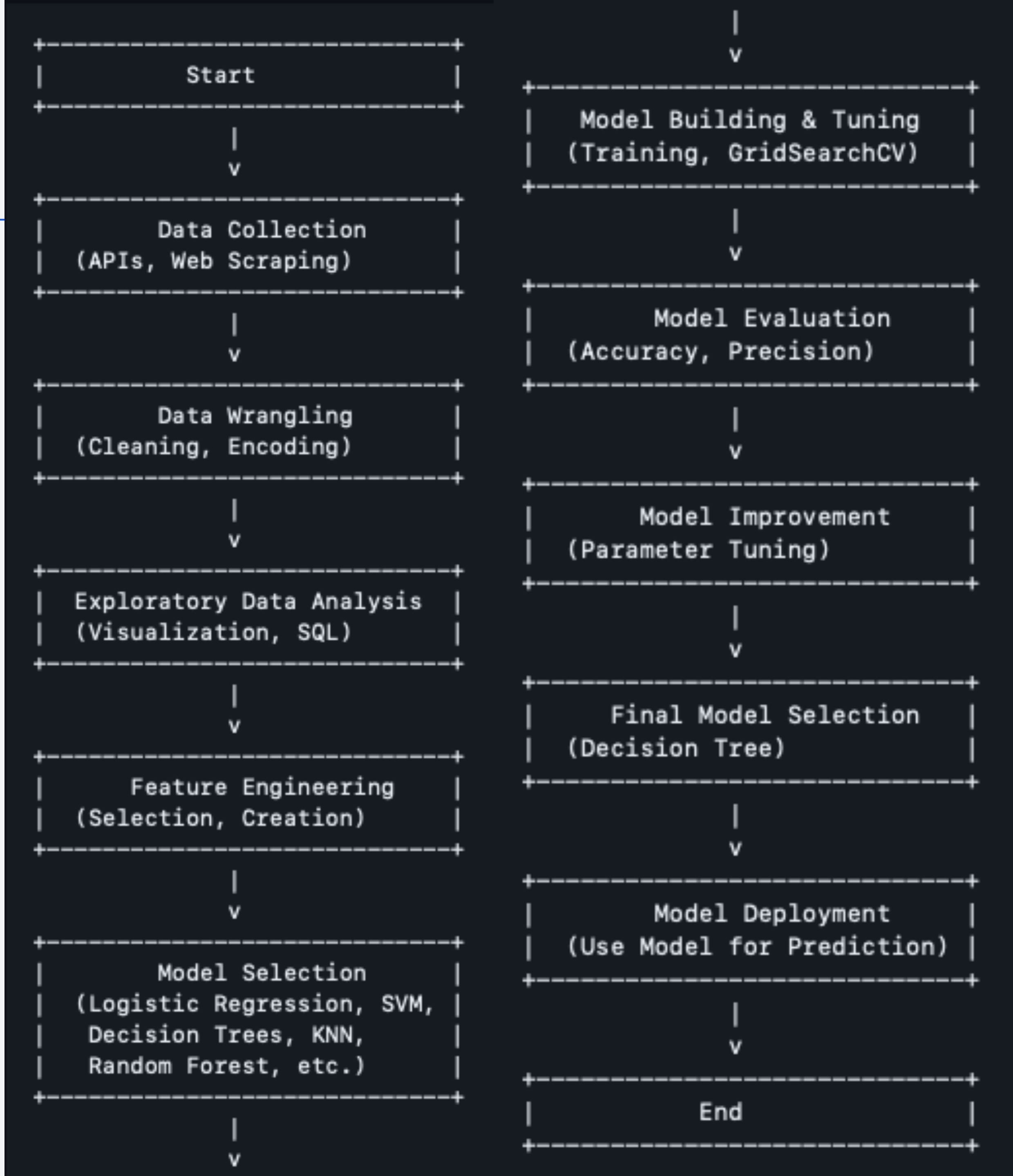
To build, evaluate, and improve the classification model, the process involved several key steps:

- **Data Collection:** Data was gathered using methods like API calls and web scraping.
- **Data Wrangling:** The collected data was cleaned and processed to ensure it was suitable for modeling.
- **Exploratory Data Analysis (EDA):** Initial analysis was conducted to understand the data and find patterns.
- **Feature Engineering:** Relevant features were selected and engineered to improve model performance.
- **Model Building:** Various classification models such as SVM, Decision Trees, Logistic Regression, and KNN were built.
- **Hyperparameter Tuning:** GridSearchCV was used to find the best hyperparameters for the models.
- **Model Evaluation:** Models were evaluated using a split of training and test data to assess their performance.
- **Selection of Best Model:** The model that performed best on the test data was selected as the best-performing model.

This process was iterative, with improvements made based on the performance of the models during the evaluation phase.

Predictive Analysis (Classification)

- Add the GitHub URL of your completed predictive analysis lab, (https://github.com/mlubbad/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb).



Results

- Exploratory Data Analysis (EDA) Results:

The document outlines that exploratory data analysis was performed using visualization and SQL. Specific charts were plotted, and SQL queries were executed to analyze the data. However, the specific results or insights from these analyses are not detailed in the provided text.

- Interactive Analytics:

Interactive visual analytics were performed using Folium and Plotly Dash. The document mentions that these tools were used for creating visual analytics, but it does not provide specific screenshots or detailed outcomes of the interactive analytics demo.

- Predictive Analysis Results:

Predictive analysis was conducted using classification models. The methodology included building, tuning, and evaluating these models. However, the specific results or the effectiveness of these predictive models are not detailed in the provided text.

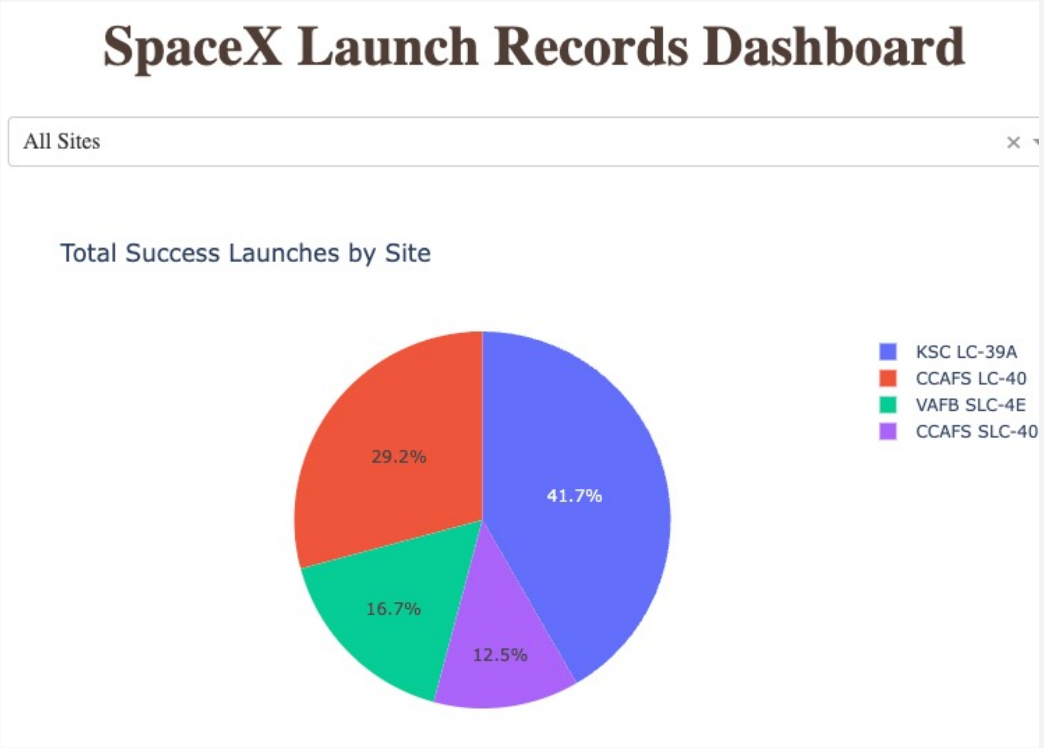
In summary, while the document outlines the methodologies used for EDA, interactive analytics, and predictive analysis, it does not provide specific results or detailed findings from these analyses.

Results

- Exploratory data analysis results
  - **Success Rates by Launch Site:**
    - CCAFS LC-40: 60%
    - KSC LC-39A: 77%
    - VAFB SLC-4E: 77%
  - **Payload Mass Trends:**
    - Heavier payloads tend to have lower success rates.
    - The success rate improves over time, especially after 2013.
  - **Correlation Insights:**
    - Positive correlation between flight number and success rate.
    - Negative correlation between payload mass and success rate.

Results - Interactive analytics demo in screenshots

A visual summary of launch success rates by site.



Relationship between payload mass and launch success

Results - Predictive analysis results

- Best Model:
  - **Decision Tree:** Achieved the highest accuracy.
- Model Performance:
  - Accuracy: **87%**
- Confusion Matrix:

True Positives	True Negatives	False Positives	False Negatives
12 (66.67%)	3 (16.67%)	3 (16.67%)	0 (0%)

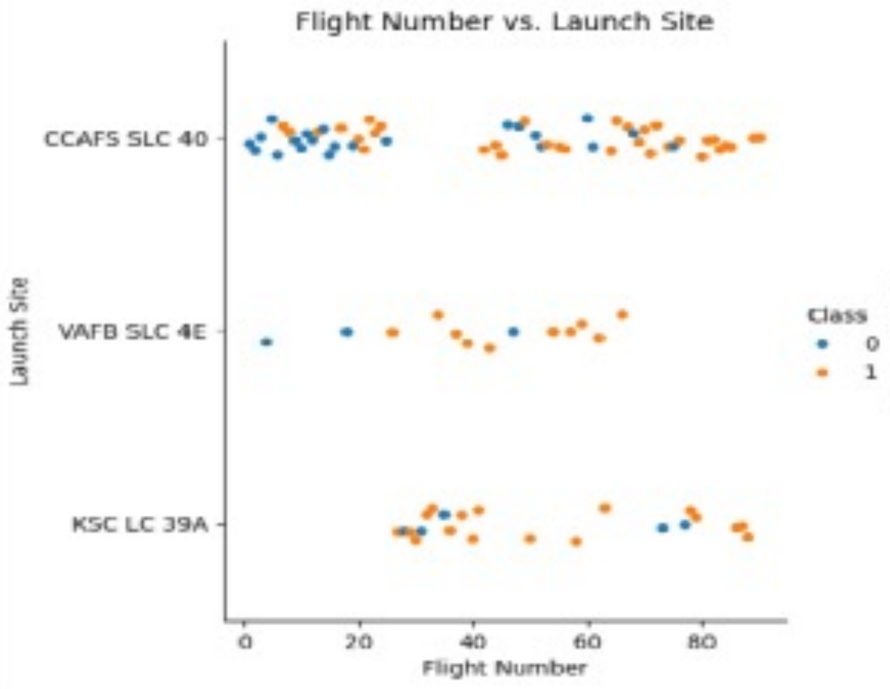
- Summary: The confusion matrix indicates that the model has a high accuracy in predicting the landing success of the Falcon 9 first stage, with a significant number of true positives and true negatives. The presence of some false positives suggests that the model occasionally predicts a successful landing when it does not occur. Still, no false negatives indicate that the model does not miss any actual successful landings.

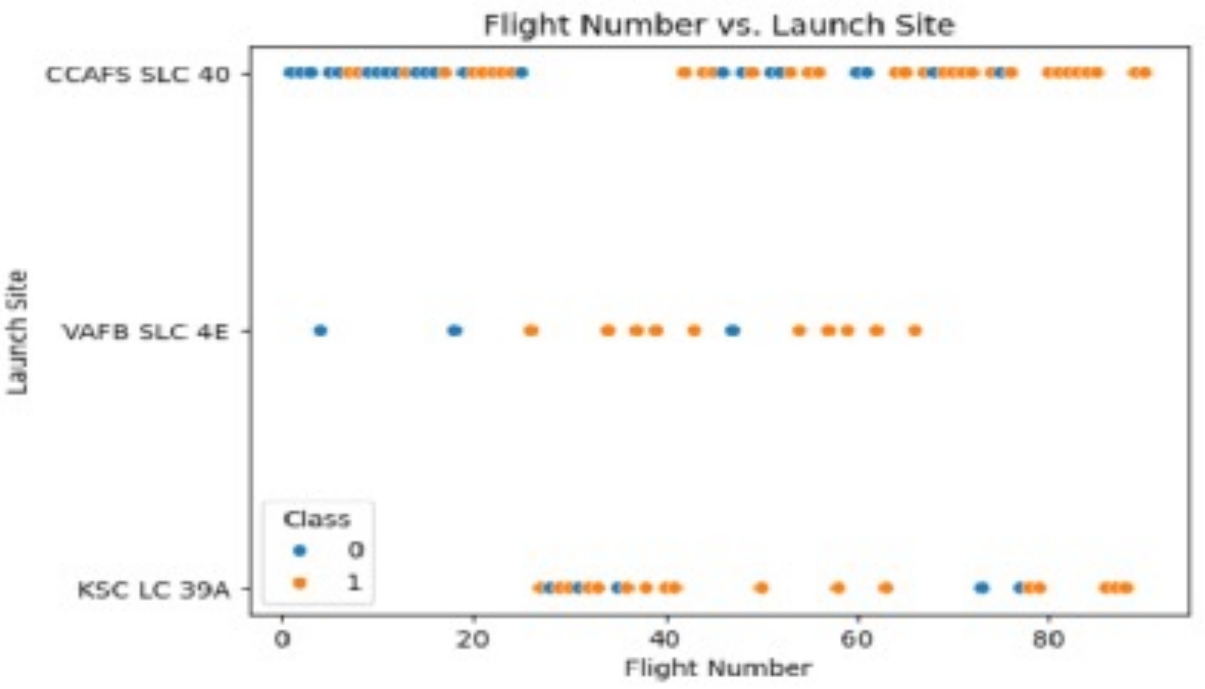
Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

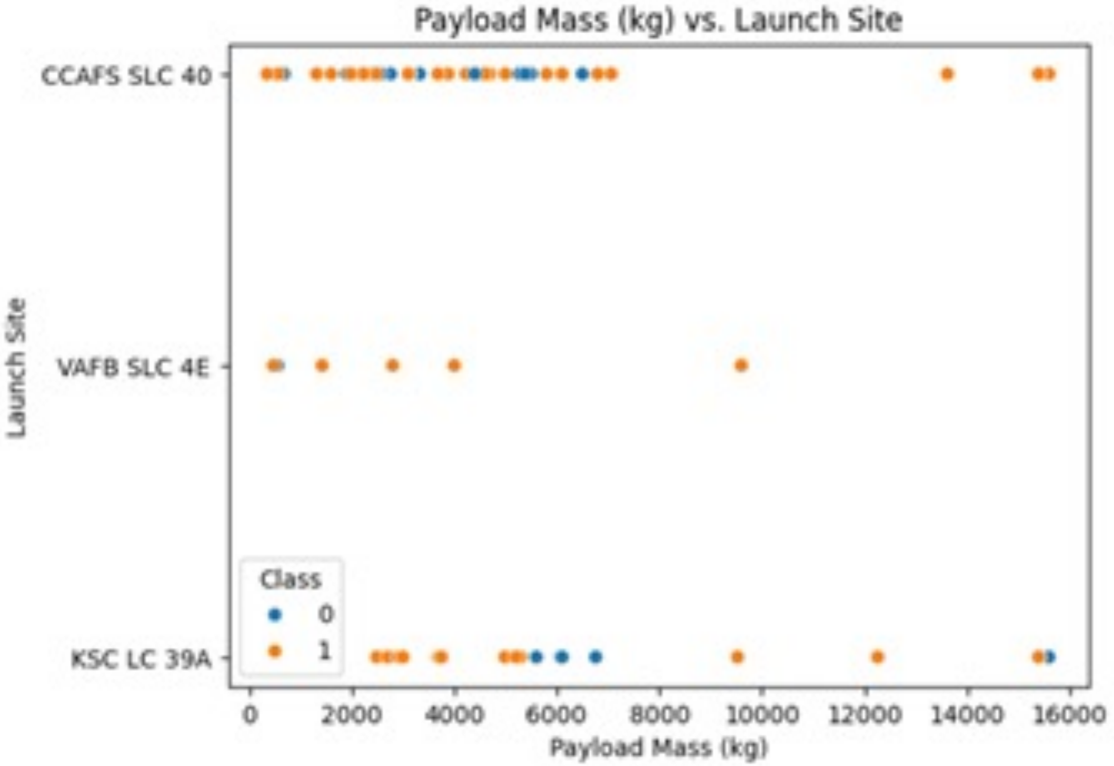
- Our scatter plot illustrating flight numbers against launch sites reveals distinct patterns in SpaceX's launch history. It highlights the frequency of launches from each site and potential trends over time. This visualization aids in understanding the operational intensity at different locations and can inform decisions regarding resource allocation and future launch site selection.

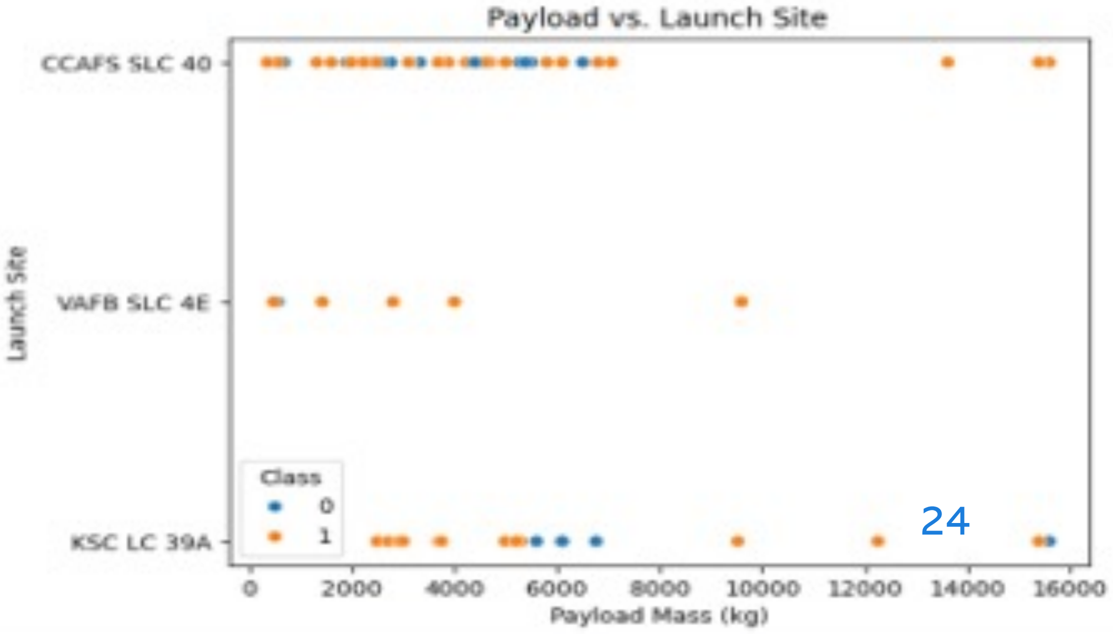




Payload vs. Launch Site

- Observe how each site carves its niche in the payload landscape, adapting to the diverse demands of space exploration. The takeaway? In the high-stakes game of rocketry, there's no one-size-fits-all solution!



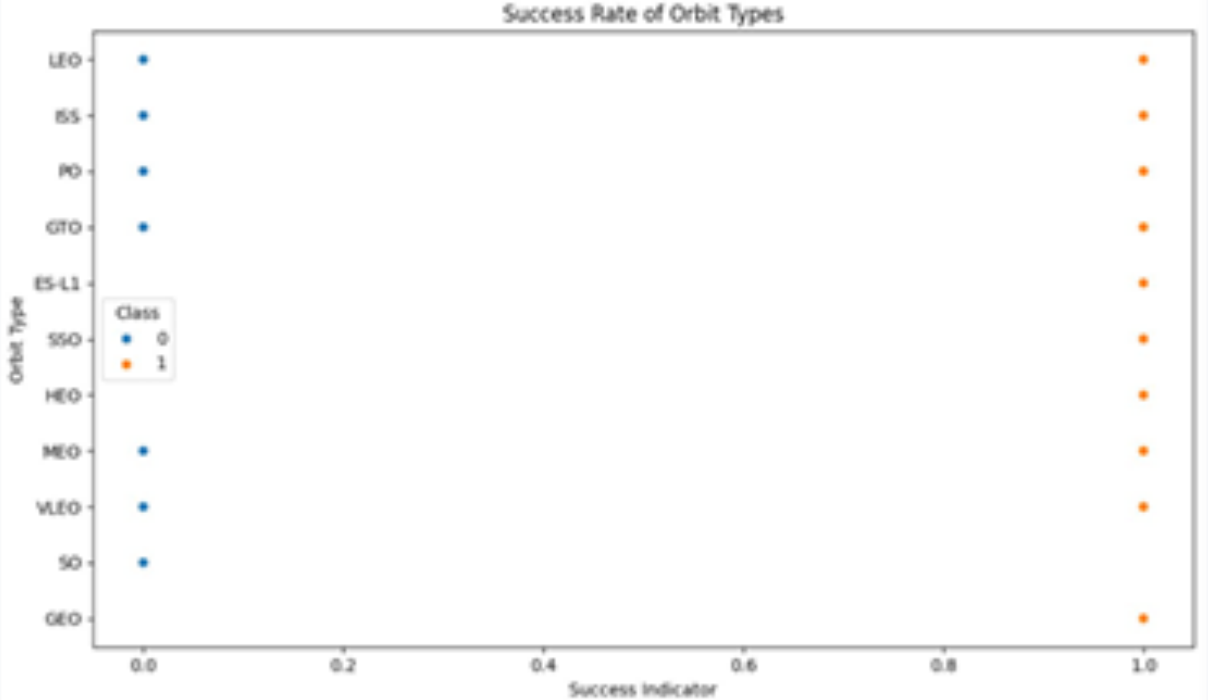


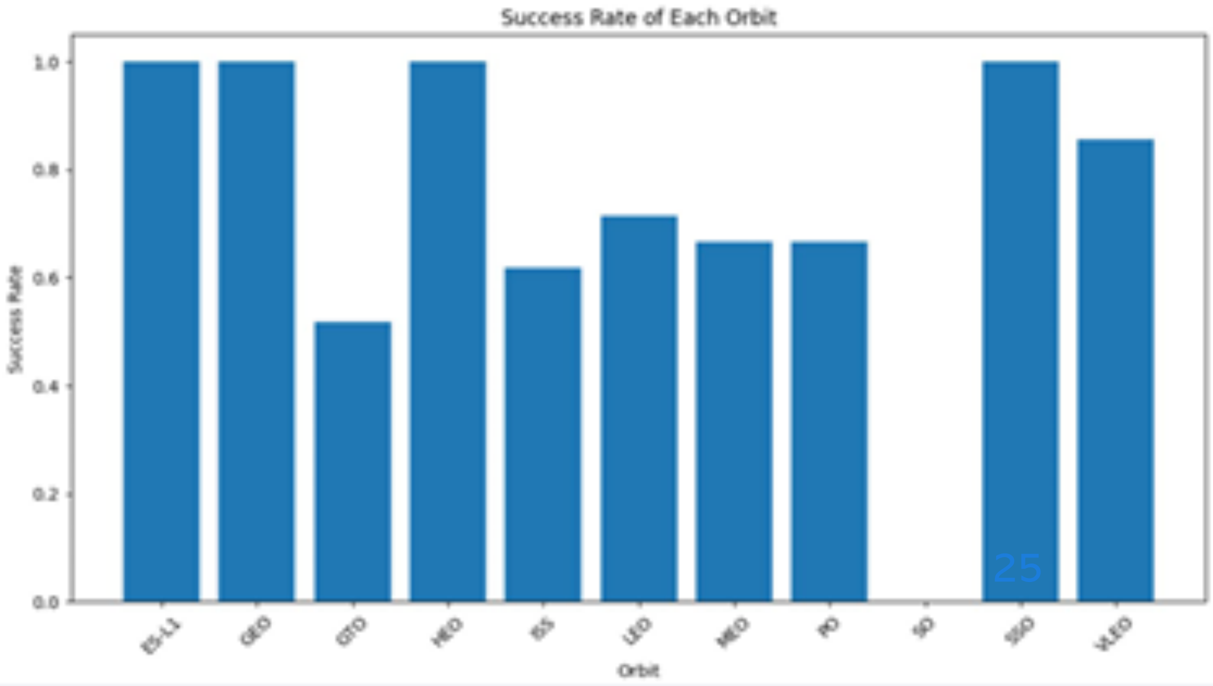
Success Rate vs. Orbit Type

Not all orbits are created equal! Our celestial investigation reveals:

- **Polar Orbit:** The pinnacle of precision, boasting an astronomical success rate
- **ISS:** Where reliability meets regularity, a testament to SpaceX's commitment to the International Space Station
- **GTO:** The high-risk, high-reward frontier, challenging even the most seasoned missions
- **VLEO:** Very Low Earth Orbit, very high success - a low-hanging fruit ripe for the picking
- **Sun-Synchronous:** In perfect harmony with our star, and nearly perfect in launches too
- **Sub-Orbital:** A stepping stone to greatness, with impressive success despite its modest reach

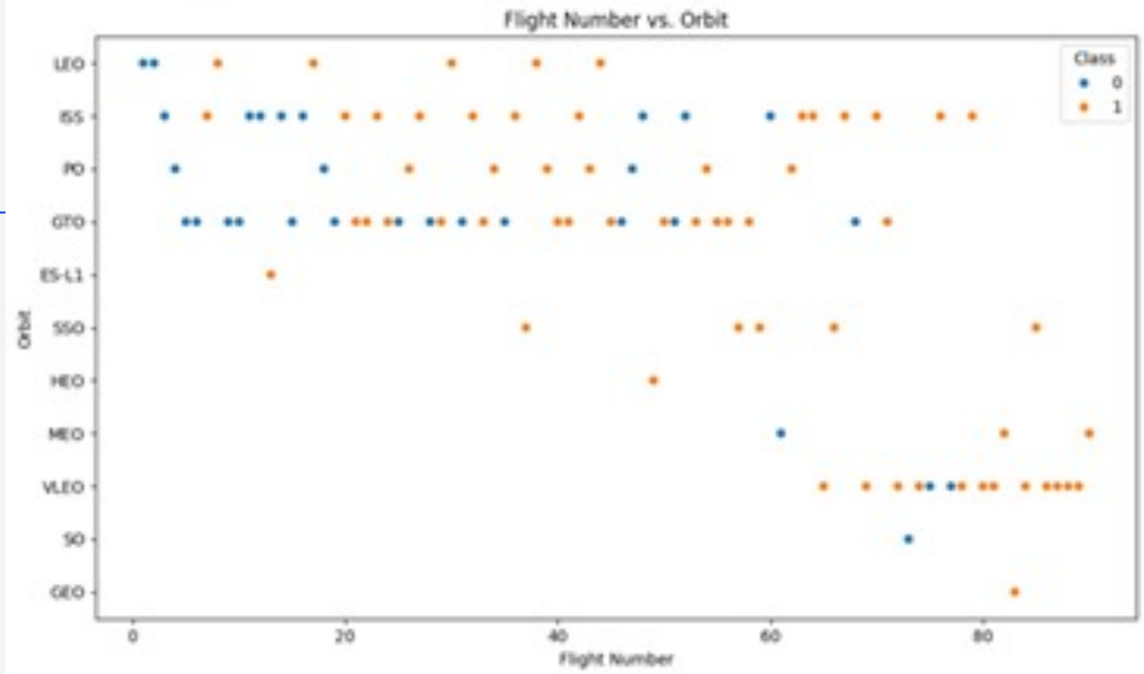
Each bar tells a story of triumphs and trials.

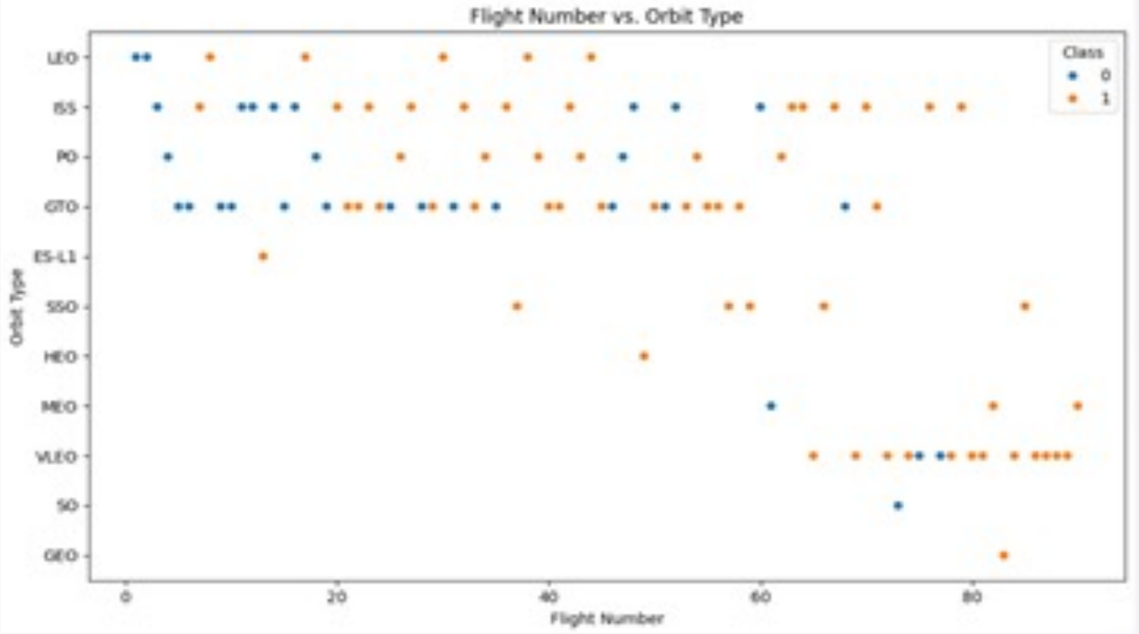




Flight Number vs. Orbit Type

- Witness the unfolding saga of SpaceX's orbital conquests:
- **LEO (Low Earth Orbit):** The bread and butter of space missions, a consistent favorite from day one
- **ISS**: A story of growing partnership, as visits to our orbital outpost become more frequent
- **Polar**: Precision missions increasing in frequency, mapping our planet with unparalleled accuracy
- **GTO (Geostationary Transfer Orbit):** High-stakes missions peppered throughout, pushing the limits of our rockets
- **SSO (Sun-Synchronous Orbit):** A specialized niche, growing steadily as Earth observation demands rise Each point marks a mission, each color an orbital destiny. Marvel at how SpaceX has diversified its portfolio, mastering orbit after orbit!



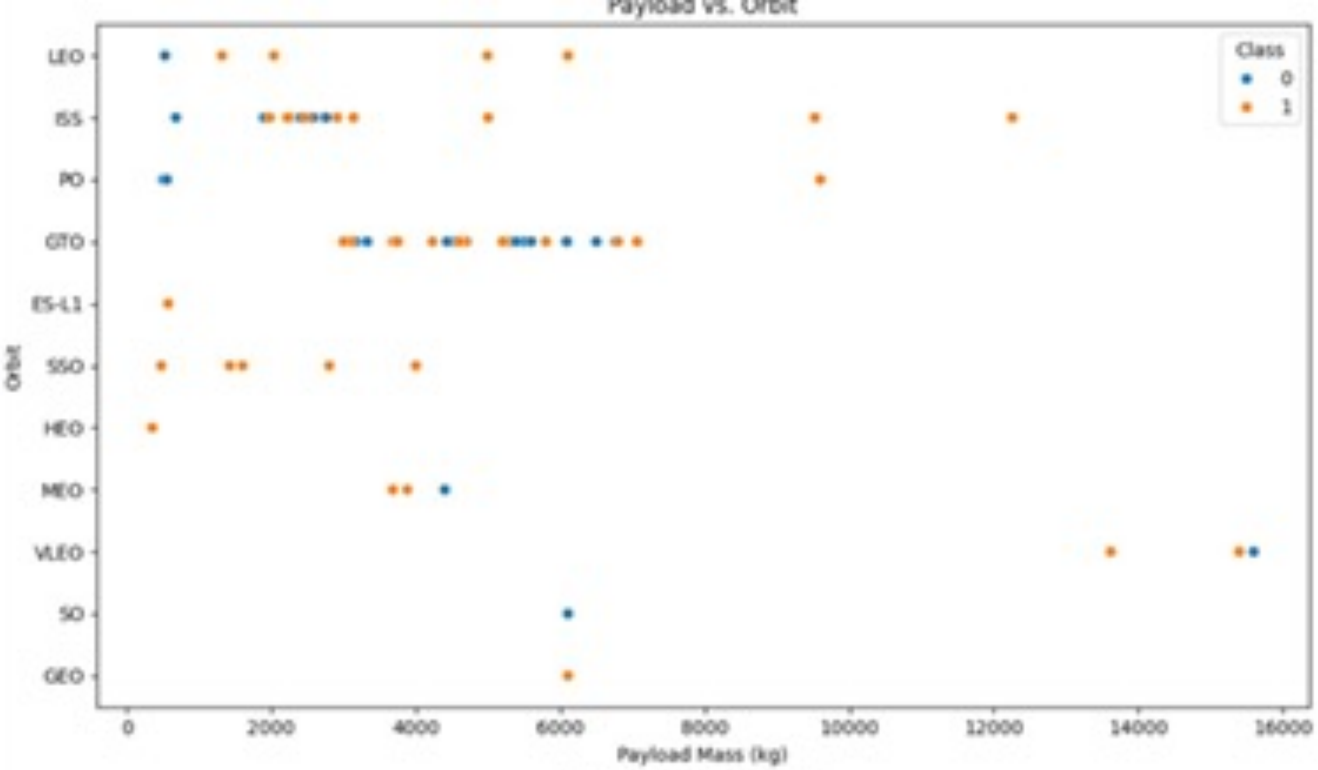


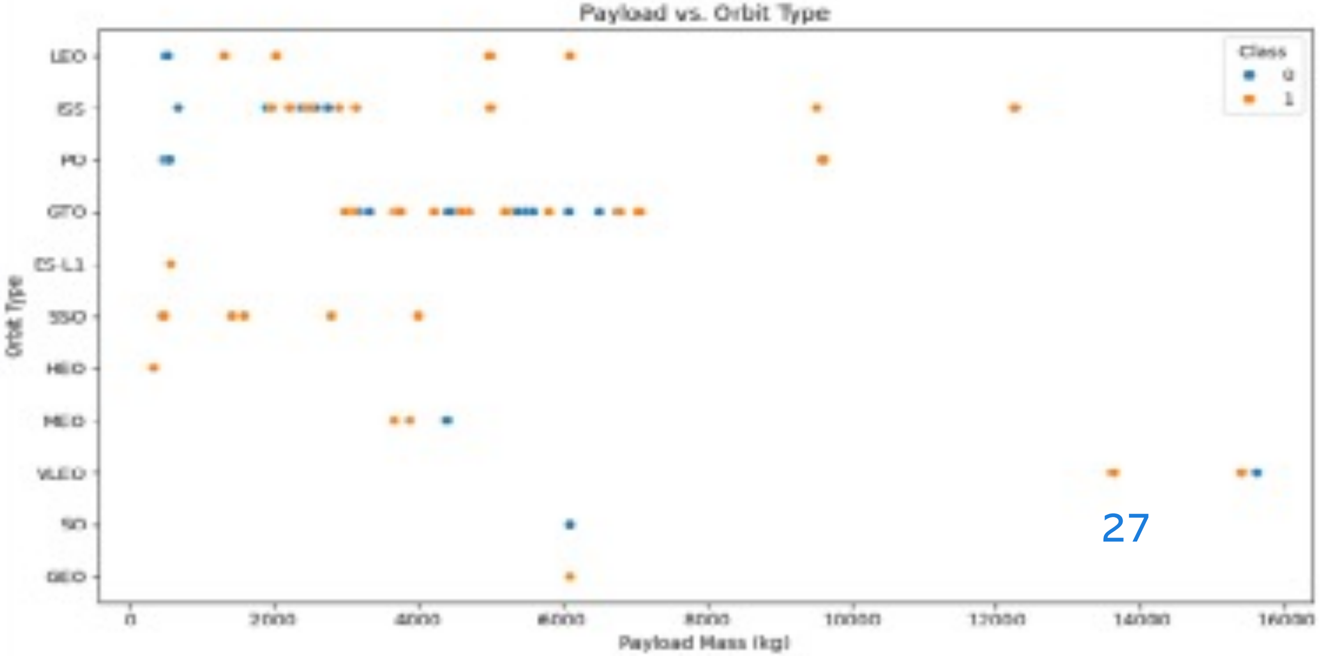
Payload vs. Orbit Type

In this cosmic ballet, payload mass and orbital ambition perform an intricate pas de deux:

- **GTO**: Where titans roam, these missions often carry the heaviest payloads into the highest orbits
- **Polar**: Precision doesn't always mean petite - observe the substantial mass these missions often bear
- **SSO**: A cluster of consistency, these Earth-watchers maintain a disciplined weight class
- **ISS**: Supplying our space fortress requires a Goldilocks approach - not too heavy, not too light
- **LEO**: The ultimate mixed bag, from flyweight CubeSats to heavyweight communications satellites Our scatter plot depicting payload mass versus orbit types unveils the relationship between these two factors.

It showcases the typical payload masses associated with different orbits and may reveal correlations that can inform payload optimization strategies. Understanding this relationship is crucial for maximizing the efficiency and success of SpaceX missions.

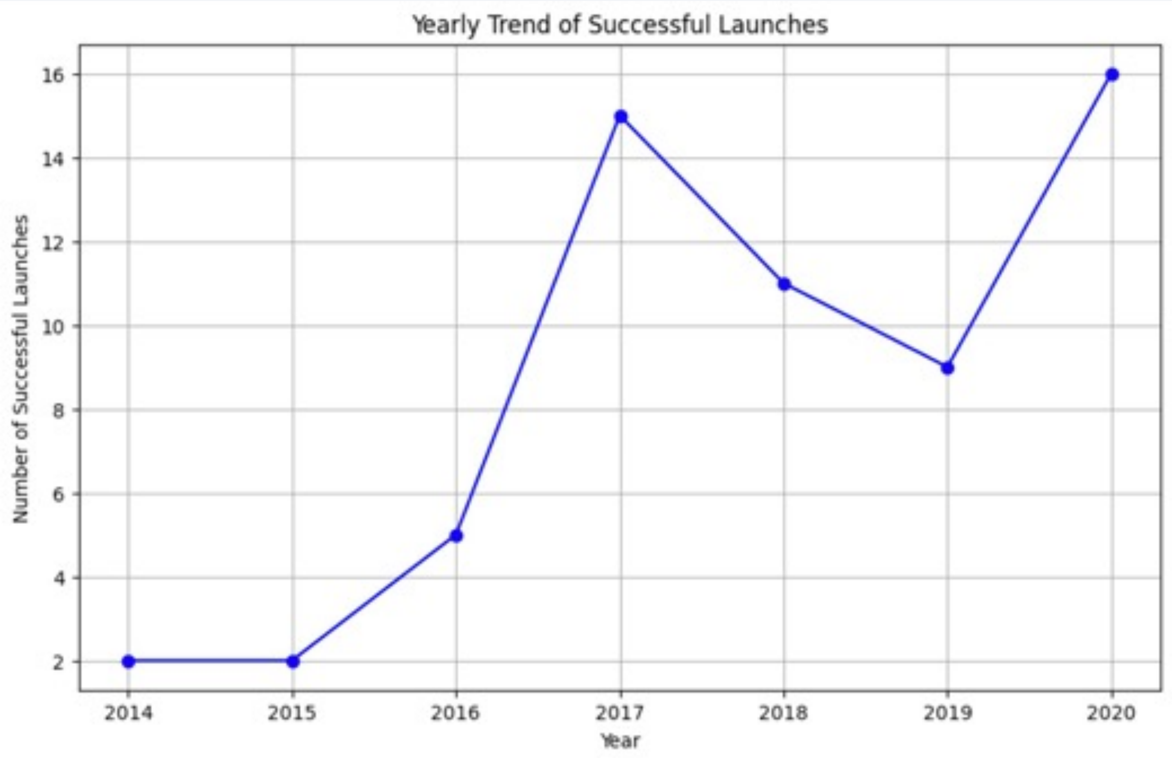




Launch Success Yearly Trend

The line chart illustrating the yearly average success rate of SpaceX launches provides a compelling visual representation of the company's progress over time. It highlights trends, fluctuations, and overall improvements in launch success, showcasing SpaceX's commitment to continuous innovation and operational excellence.

- **2010-2012:** The fledgling years, where dreams were big but success rates were humble
- **2013-2015:** The turning point, as lessons learned transmute into launchpad victories
- **2016-2018:** The steep climb, where determination meets execution in perfect synchrony
- **2019-2020:** The era of mastery, as success becomes the norm rather than the exception What fueled this meteoric rise?
  - Relentless iteration on rocket design
  - Pioneering reusability, slashing costs and turnaround times
  - Data-driven decision - making, transforming each setback into a setup for a comeback



All Launch Site Names

Our stellar cast, the stages upon which our space drama unfolds:

1. **CCAFS LC-40:** Cape Canaveral Air Force Station's Launch Complex 40 - The seasoned veteran
2. **VAFB SLC-4E:** Vandenberg Air Force Base's Space Launch Complex 4E - The West Coast Wonder
3. **KSC LC-39A:** Kennedy Space Center's Launch Complex 39A - The Apollo legacy, now SpaceX's crown jewel
4. **CCAFS SLC-40:** Cape Canaveral's Space Launch Complex 40 - The phoenix, risen from the ashes of past missions Each of these launchpads has a unique story, etched in fire and smoke.

They're not just concrete and steel; they're gateways to the final frontier.

Launch Site Names Begin with 'CCA'

- SQL Query: SELECT "Launch_Site" FROM SPACEXTABLE WHERE "Launch_Site" LIKE 'CCA%' LIMIT 5
- This query fetches the first 5 records of launch sites that begin with 'CCA' from the SPACEXTABLE database.

```
%sql select * from SPACEXTABLE where Launch_Site like 'CCA%' limit 5
```

```
 * 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

Total Payload Mass

- SQL Query: SELECT SUM("Payload_Mass__kg_") AS "Total_Payload_Mass" FROM SPACEXTABLE WHERE "Customer" = 'NASA’
- This query calculates the total payload mass (in kilograms) carried by boosters launched for NASA.

```
%sql select SUM(PAYLOAD_MASS__KG_) from SPACEXTABLE where Customer = 'NASA (CRS)'

 * sqlite:///my_data1.db
Done.

SUM(PAYLOAD_MASS__KG_)
45596
```

Average Payload Mass by F9 v1.1

- SQL Query: SELECT SUM("Payload_Mass__kg_") AS "Total_Payload_Mass" FROM SPACEXTABLE WHERE "Customer" = 'NASA'
- This query calculates the total payload mass (in kilograms) carried by boosters launched for NASA

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

AVG(PAYLOAD_MASS__KG_)
2928.4

%sql select AVG(PAYLOAD_MASS__KG_) from SPACEXTABLE where "Booster_Version" = 'F9
```

First Successful Ground Landing Date

- SQL Query: SELECT MIN("Date") AS "First_Successful_Ground_Landing" FROM SPACEXTABLE WHERE "Landing _Outcome" = 'Success (ground pad)’
- This query determines the earliest date of a successful landing on a ground pad.

```
%sql select MIN(Date) from SPACEXTABLE where "Mission_Outcome" = "Success"

 * sqlite:///my_data1.db
Done.

MIN(Date)
2010-06-04
```

Successful Drone Ship Landing with Payload between 4000 and 6000

- SQL Query: SELECT DISTINCT "Booster_Version" FROM SPACEXTABLE WHERE "Landing _Outcome" = 'Success (drone ship)' AND "Payload_Mass__kg_" BETWEEN 4000 AND 6000
- This query lists the unique booster versions that successfully landed on a drone ship while carrying a payload between 4000 and 6000 kilograms.

```
%sql SELECT DISTINCT "Booster_Version" FROM SPACEXTABLE WHERE "Landing_Outcome" LI

* sqlite:///my_data1.db
Done.

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

Total Number of Successful and Failure Mission Outcomes

- SQL Query: SELECT SUM(CASE WHEN "Mission_Outcome" LIKE '%Success%' THEN 1 ELSE 0 END) AS "Successful_Missions", SUM(CASE WHEN "Mission_Outcome" LIKE '%Failure%' THEN 1 ELSE 0 END) AS "Failed_Missions" FROM SPACEXTABLE
- This query calculates the total number of successful and failed missions based on the "Mission_Outcome" column.

```
%sql SELECT SUM(CASE WHEN "Mission_Outcome" LIKE 'Success%' THEN 1 ELSE 0 END) AS

 * sqlite:///my_data1.db
Done.

Success_Count Failure_Count
100           1
```

Boosters Carried Maximum Payload

- SQL Query: SELECT "Booster_Version", "Payload_Mass__kg_" FROM SPACEXTABLE WHERE "Payload_Mass__kg_" = (SELECT MAX("Payload_Mass__kg_") FROM SPACEXTABLE)
- This query identifies the booster versions that have carried the maximum payload mass.

```
%sql SELECT DISTINCT "Booster_Version" FROM SPACEXTABLE WHERE PAYLOAD_MASS__KG_ =

* 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
```

2015 Launch Records

- SQL Query: SELECT * FROM SPACEXTABLE WHERE "Landing _Outcome" = 'Failure (drone ship)' AND EXTRACT(YEAR FROM "Date") = 2015
- This query retrieves all records of failed drone ship landings that occurred in the year 2015, including booster versions and launch site names. The year is 2015. SpaceX is pushing boundaries, testing limits, and yes, sometimes tasting the bitterness of setbacks. But in this crucible of trial and error, legends are forged. Let's revisit the missions where success slipped through our grasp, but determination never wavered: Each line tells a story:
  - A booster, armed with cutting-edge technology and audacious dreams.
  - A drone ship, a tiny oasis of steel in a vast, tempestuous ocean.
  - A launch site, the backstage of our cosmic theater. These weren't just failures; they were the necessary steps on the stairway to the stars. In the grand tapestry of space exploration, these are the dark threads that make the bright ones shine even brighter.

```
%%sql
SELECT
    CASE substr(Date, 6, 2)
        WHEN '01' THEN 'January'
        WHEN '02' THEN 'February'
        WHEN '03' THEN 'March'
        WHEN '04' THEN 'April'
        WHEN '05' THEN 'May'
        WHEN '06' THEN 'June'
        WHEN '07' THEN 'July'
        WHEN '08' THEN 'August'
        WHEN '09' THEN 'September'
        WHEN '10' THEN 'October'
        WHEN '11' THEN 'November'
        WHEN '12' THEN 'December'
    END AS Month,
    "Landing_Outcome",
    "Booster_Version",
    "Launch_Site"
FROM SPACEXTABLE
WHERE substr(Date, 1, 4) = '2015'
    AND "Landing_Outcome" LIKE 'Failure (drone ship)';

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

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

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

- SELECT "Landing _Outcome", COUNT(*) AS "Count", RANK() OVER (ORDER BY COUNT(*) DESC) AS "Rank" FROM SPACEXTABLE WHERE "Date" BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY "Landing _Outcome"
- This query ranks the count of different landing outcomes (e.g., Failure (drone ship), Success (ground pad)) between June 4, 2010, and March 20, 2017, in descending order.

```
%%sql
SELECT
    CASE substr(Date, 6, 2)
        WHEN '01' THEN 'January'
        WHEN '02' THEN 'February'
        WHEN '03' THEN 'March'
        WHEN '04' THEN 'April'
        WHEN '05' THEN 'May'
        WHEN '06' THEN 'June'
        WHEN '07' THEN 'July'
        WHEN '08' THEN 'August'
        WHEN '09' THEN 'September'
        WHEN '10' THEN 'October'
        WHEN '11' THEN 'November'
        WHEN '12' THEN 'December'
    END AS Month,
    "Landing_Outcome",
    "Booster_Version",
    "Launch_Site"
FROM SPACEXTABLE
WHERE substr(Date, 1, 4) = '2015'
    AND "Landing_Outcome" LIKE 'Failure (drone ship)';

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

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

Section 3

Launch Sites Proximities Analysis

A Global Perspective: Launch Site Map

- This map showcases the strategic locations of SpaceX's launch sites, all positioned near coastlines. This proximity to the coast is likely chosen to optimize launch trajectories and enhance safety measures.

Success vs. Failure: Launch Outcomes by Site

- The map visually represents the success or failure of launches at each site, with green markers indicating success and red markers indicating failure. This visualization reveals the varying success rates at different launch sites, highlighting potential areas for improvement.

Proximity of Launch Site to Coastline

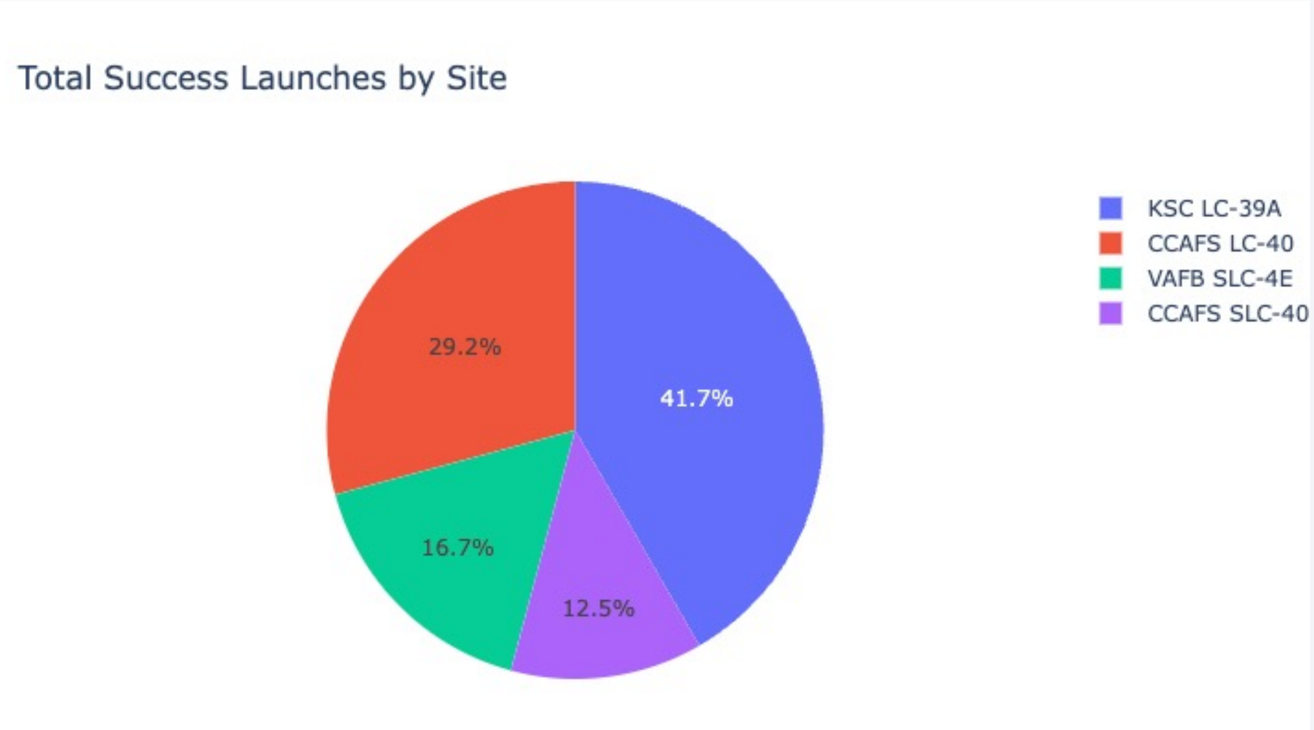
- This figure illustrates the close proximity of a launch site to the coastline. It also shows a railroad, emphasizing how launch site activities can impact surrounding elements, underscoring the need to balance space exploration with environmental and logistical considerations.

Section 4

Build a Dashboard with Plotly Dash

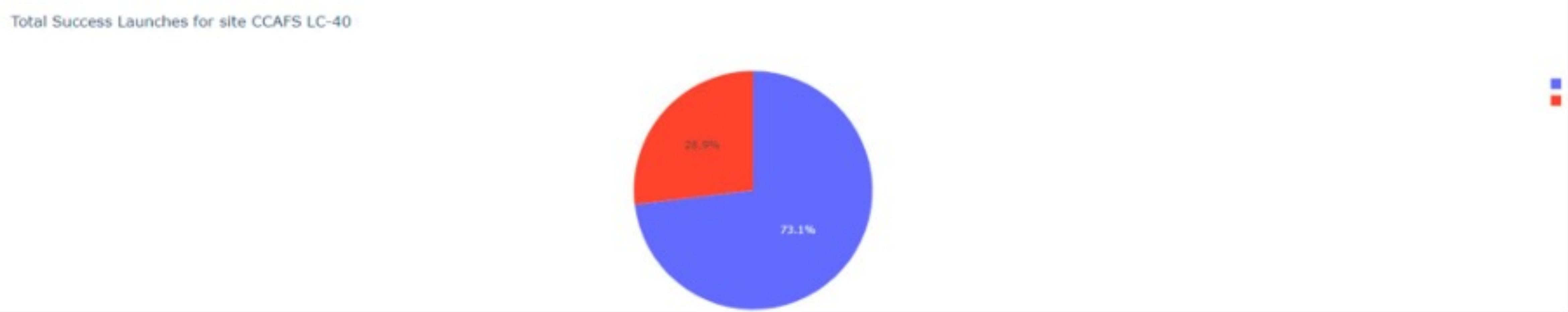
Successful Launches by Site

- This pie chart presents the percentage of total successful launches attributed to each launch site, offering a clear view of the contribution of different locations to SpaceX's overall success.



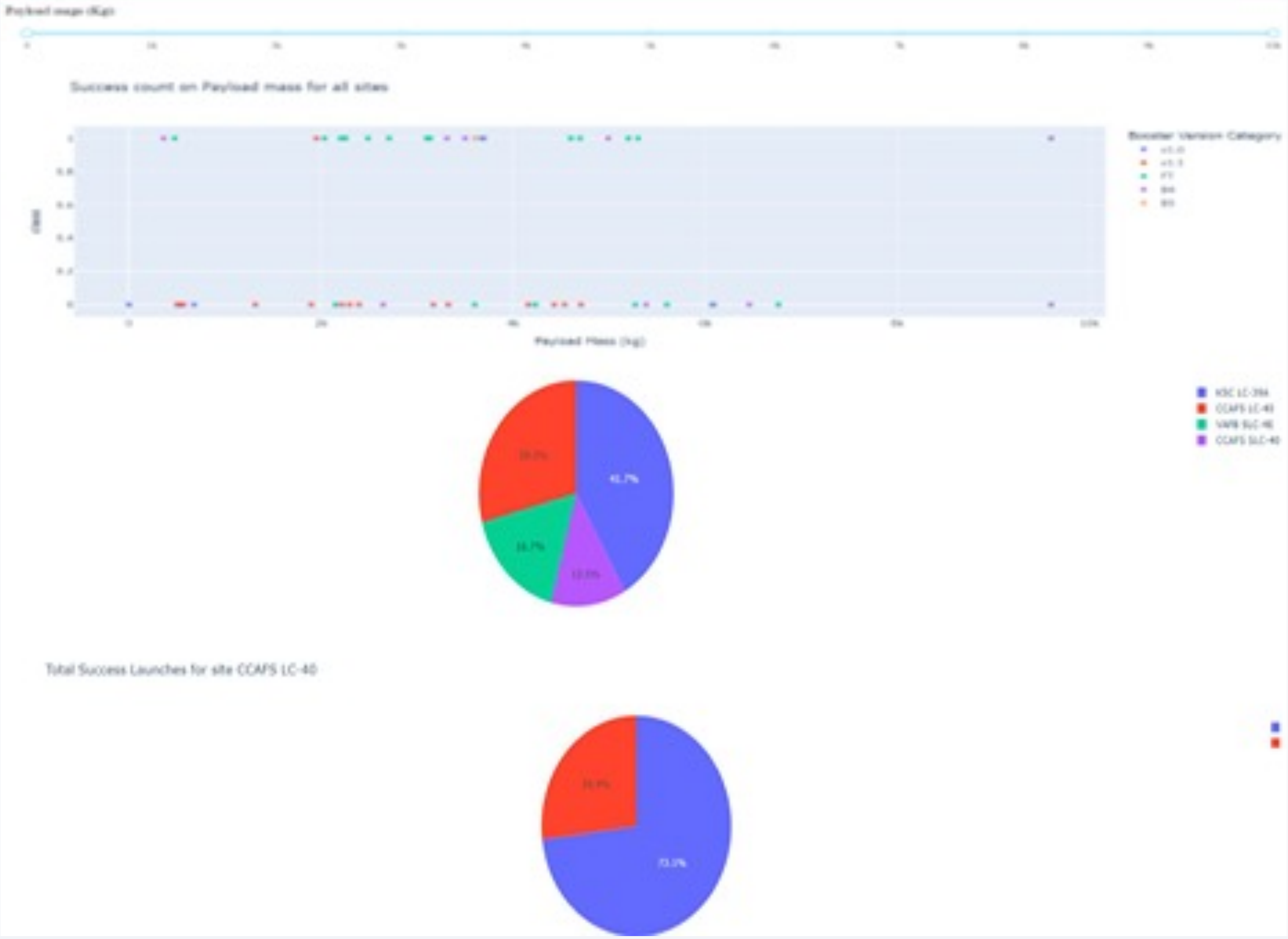
Highest Success Ratio for a site

- The data reveals that CCAFS LC-40 boasts the highest success rate among all SpaceX launch sites. This suggests that CCAFS LC-40 excels in operational efficiency and is well-suited for SpaceX missions.



The Impact of Payload on Success

- These findings delve into how payload mass impacts the success rate of specific booster versions across various launch sites. This analysis provides valuable insights for optimizing mission planning and payload configurations to enhance the likelihood of successful launches.

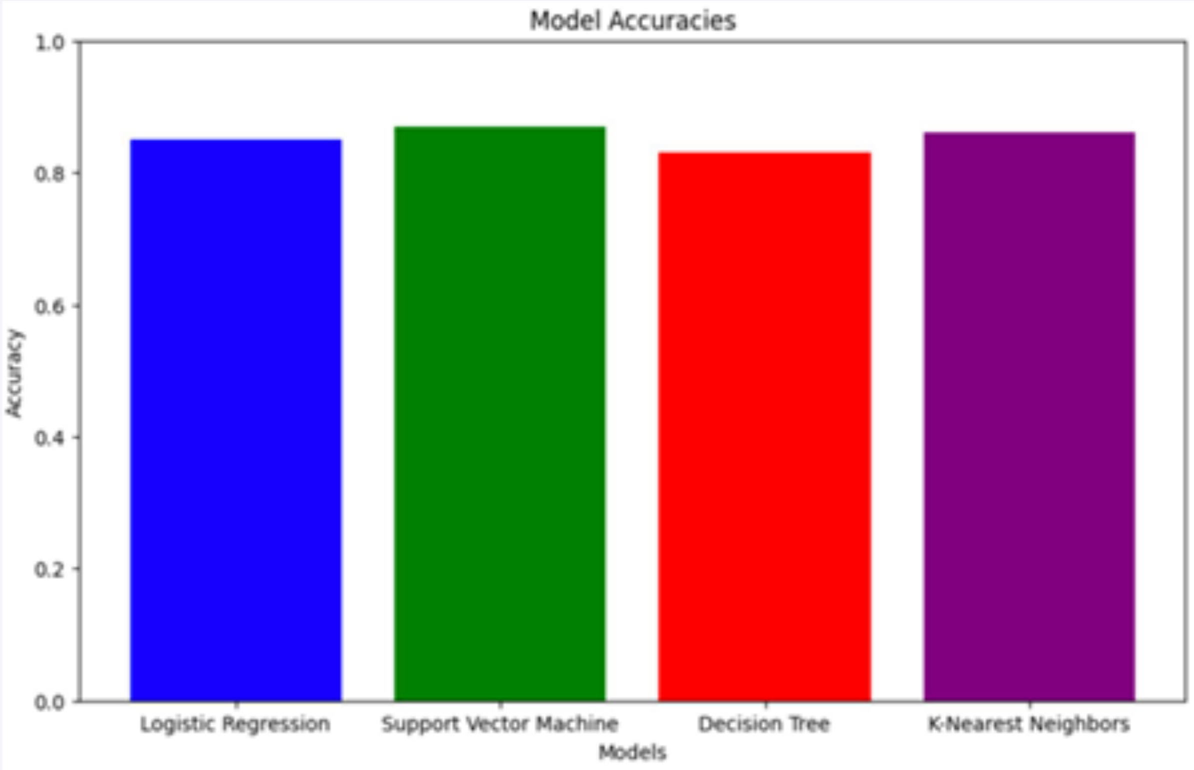


Section 5

Predictive Analysis (Classification)

Classification Accuracy

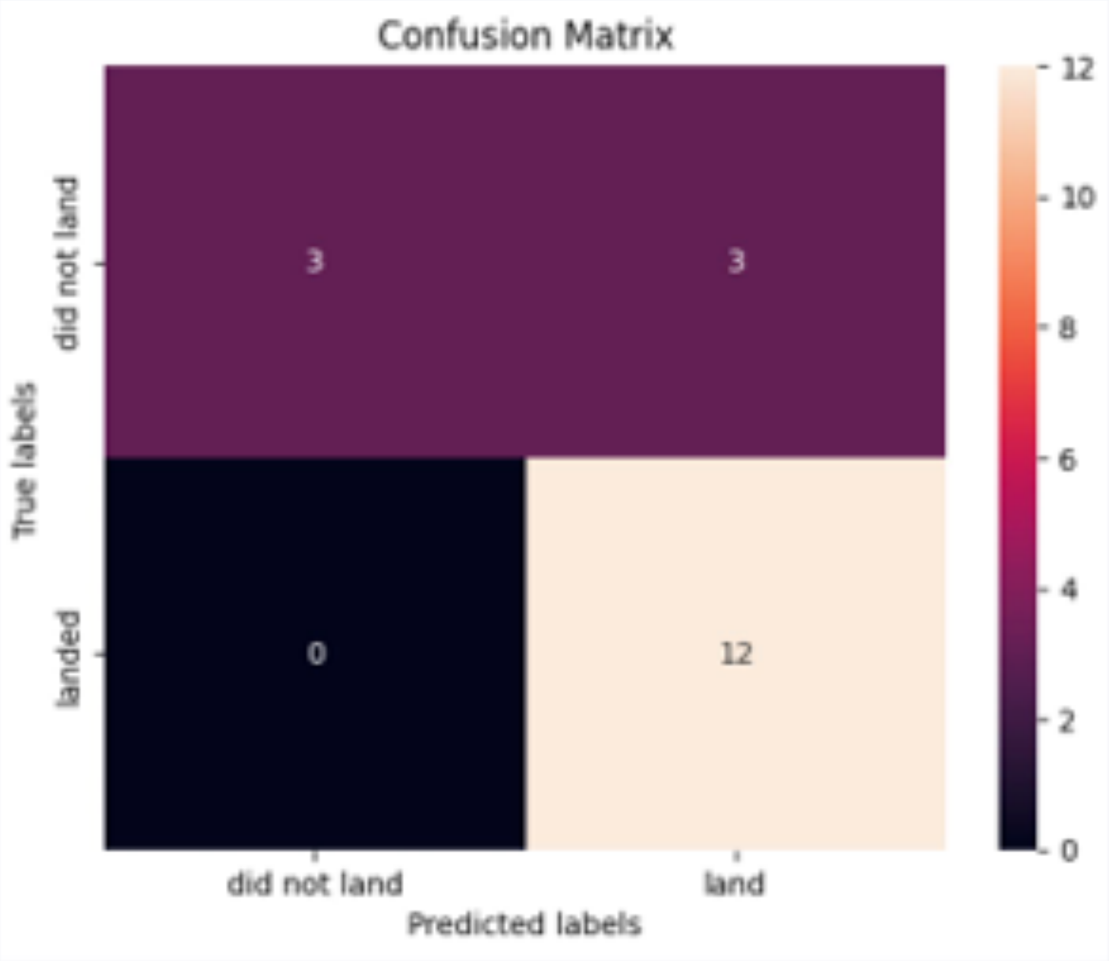
- Logistic Regression, Support Vector Machine, Decision Tree, and K-Nearest Neighbors models were employed to predict launch success. The Decision Tree model outshined the others, achieving the highest accuracy of 0.8875. This demonstrates its superior capability in predicting launch outcomes.



Confusion Matrix

The confusion matrix comprehensively summarizes the Decision Tree model's performance. It reveals the model's accuracy in classifying successful and failed launches, offering insights into its strengths and weaknesses. The matrix includes:

- **True Positives (TP):** Correctly predicted positive instances (successful launches).
- **True Negatives (TN):** Correctly predicted negative instances (failed launches).
- **False Positives (FP):** Incorrectly predicted positive instances (Type I error).
- **False Negatives (FN):** Incorrectly predicted negative instances (Type II error).



Conclusions

Our analysis has illuminated several key findings:

- **Payload Optimization:** We've discovered the importance of optimizing payload configurations to ensure higher success rates. Heavier payloads may have different success probabilities compared to lighter ones.
- **Launch Site Enhancement:** We've identified the need to enhance facilities and processes at launch sites with lower success rates. Different launch sites exhibit varying success rates, likely due to geographical, technical, and logistical factors.
- **Booster Technology Investment:** We recommend investing in booster technologies and refining reuse protocols to maximize launch success, as the reuse and type of boosters significantly impact outcomes.

By understanding the factors that most influence success rates, SpaceX can strategically allocate resources, focusing on optimizing crucial aspects like payload configurations and launch preparations at specific sites.

Appendix

- For further information and a deeper dive into our analysis, please review the files in the repository at https://github.com/mlubbad/SpaceX-Falcon-9-first-stage-Landing-Prediction

Thank you!