IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Naim Pilon
17-08-2024

Outline

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

Executive Summary

Summary of Methodologies:

1. Data Collection:
    - Utilized SpaceX API to gather comprehensive launch data
    - Performed web scraping to collect additional information
    - Conducted data wrangling to prepare the dataset for analysis
2. Exploratory Data Analysis (EDA):
    - Used various visualization techniques including scatter plots, bar charts, and line charts
    - Employed SQL queries for data exploration and insights
3. Interactive Visualization:
    - Created interactive maps using Folium to display launch sites and outcomes
    - Developed dashboards with Plotly Dash for dynamic data exploration
4. Predictive Analysis:
    - Implemented classification models to predict launch outcomes
    - Evaluated and compared different models including Decision Trees, Logistic Regression, SVM, and KNN

Executive Summary

Summary of Results

1. Launch Success Trends: Observed an increasing success rate with higher flight numbers across all launch sites
2. Noted a clear upward trend in success rate from 2010 to 2020
3. Payload and Orbit Analysis: Found no clear correlation between payload mass and launch success
4. Identified varying success rates across different orbit types
5. Launch Site Performance: KSC LC-39A showed the highest proportion of successful launches (41.7% of total successes)
6. KSC LC-39A also demonstrated a high success rate of 76.9% for its launches
7. Predictive Model Performance: Decision Tree model achieved the highest cross-validated score among tested models
8. The best model showed high accuracy in predicting successful landings but struggled with predicting landing failures
9. Geographical Insights: Launch sites are strategically located near transportation routes and coastlines, while maintaining safe distances from urban areas

Introduction

Project background and context

- This project focuses on analyzing SpaceX's Falcon 9 rocket launches. SpaceX is a pioneering private space company that has revolutionized the space industry by developing reusable rockets, significantly reducing the cost of space missions. The Falcon 9 rocket is a key component of SpaceX's operations, used for various missions including satellite deployments, cargo resupply to the International Space Station, and crewed missions.
- The analysis covers a period from 2010 to 2020, encompassing the early developmental stages of the Falcon 9 through to its more mature operational phase. This timeframe allows for the examination of SpaceX's progress, learning curve, and improvements in launch success rates over time.
- The project utilizes data from multiple sources, including the SpaceX API and web scraping, to create a comprehensive dataset covering various aspects of each launch, such as payload details, launch sites, booster versions, and landing outcomes.

Introduction

Problems and Questions

- Launch Success Factors:
  - How does the flight number (i.e., launch experience) correlate with launch success?
  - Is there a relationship between payload mass and launch success?
  - Do certain orbit types have higher success rates than others?
- Launch Site Performance:
  - Which launch sites have the highest success rates?
  - How do different launch sites compare in terms of the number of successful launches?
- Temporal Trends:
  - How has the launch success rate changed over time from 2010 to 2020?
  - Are there any notable milestones or turning points in SpaceX's launch history?
- Payload and Mission Characteristics:
  - What types of payloads and missions has SpaceX undertaken?
  - How do payload masses vary across different orbit types?
- Geographical and Infrastructure Considerations:
  - How are launch sites positioned in relation to key infrastructure and population centers?
  - What are the strategic considerations for launch site locations?
- Predictive Capabilities:
  - Can machine learning models accurately predict the success of future launches?
  - Which factors are most influential in determining launch success?
- Reusability and Cost Efficiency:
  - How has the introduction of reusable boosters affected launch success rates?
  - Are there any patterns in the performance of reused boosters?
- Technical Evolution:
  - How have different booster versions performed over time?
  - Is there evidence of technological improvements leading to higher success rates?

Section 1

Methodology

Executive Summary

- Data Collection: Leveraged the SpaceX API to gather comprehensive launch data, including rocket IDs, launchpad IDs, payload IDs, and core IDs
- Performed targeted API calls to enrich the dataset with additional details such as booster versions, launch site coordinates, payload masses, and core reuse information
- Conducted web scraping using BeautifulSoup to extract relevant information from HTML tables
- Executed thorough data wrangling processes, including data exploration, feature engineering, and label creation to prepare the dataset for in-depth analysis
- Exploratory Data Analysis (EDA): Utilized a variety of visualization techniques:
  - Scatter plots to analyze relationships between launch success, payload mass, flight number, and orbit type
  - Bar charts to compare success rates across different orbit types and visualize launch site distribution
  - Line charts to illustrate success rate trends over time
- Employed SQL queries for advanced data exploration, including:
  - Identifying unique launch sites
  - Calculating total payload mass for specific mission types
  - Determining average payload mass for specific booster versions
  - Finding key dates such as the first successful ground pad landing

Executive Summary

- Interactive Visualization: Developed interactive maps using Folium to display:
    - Launch site locations with precise markers
    - Color-coded indicators for launch outcomes at each site
    - Proximity analysis of launch sites to key infrastructure
- Created dynamic dashboards with Plotly Dash, enabling:
    - Real-time exploration of launch success rates by site
    - Interactive analysis of payload mass versus launch outcome across different booster versions
- Predictive Analysis: Implemented and compared multiple classification models:
    - Decision Trees
    - Logistic Regression
    - Support Vector Machines (SVM)
    - K-Nearest Neighbors (KNN)
- Conducted model evaluation using cross-validation techniques
- Analyzed confusion matrices to assess model performance in detail

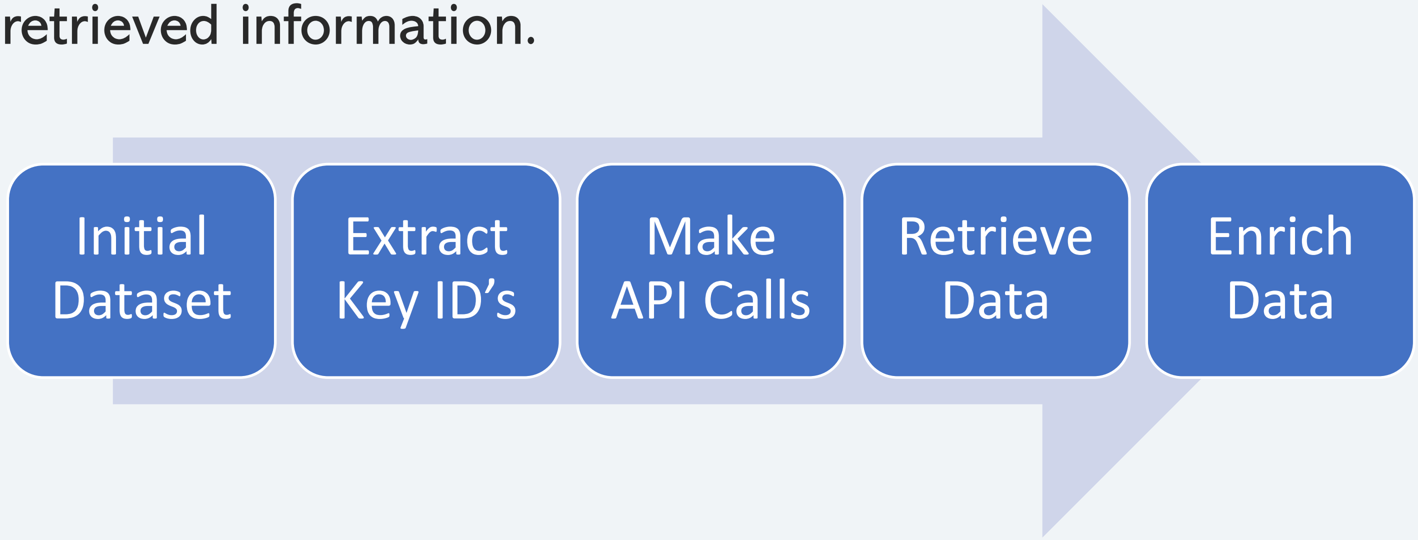
Data Collection

The dataset was collected by querying the SpaceX API, which provides detailed information about past SpaceX launches. The process begins with the extraction of key identifiers, such as rocket IDs, launchpad IDs, payload IDs, and core IDs, from an initial dataset. Using these identifiers, the notebook makes targeted API calls to retrieve additional data related to each launch. For instance, it fetches rocket details like booster versions, launch site coordinates, payload masses, and core reuse information. This enriched data is then stored in lists and integrated back into the original dataset, resulting in a comprehensive and detailed record of SpaceX launches.

Data Collection – SpaceX API

- Initial dataset was processed to extract key identifiers.
- API calls were made using these identifiers to retrieve additional data.
- The dataset was then enriched with the newly retrieved information.

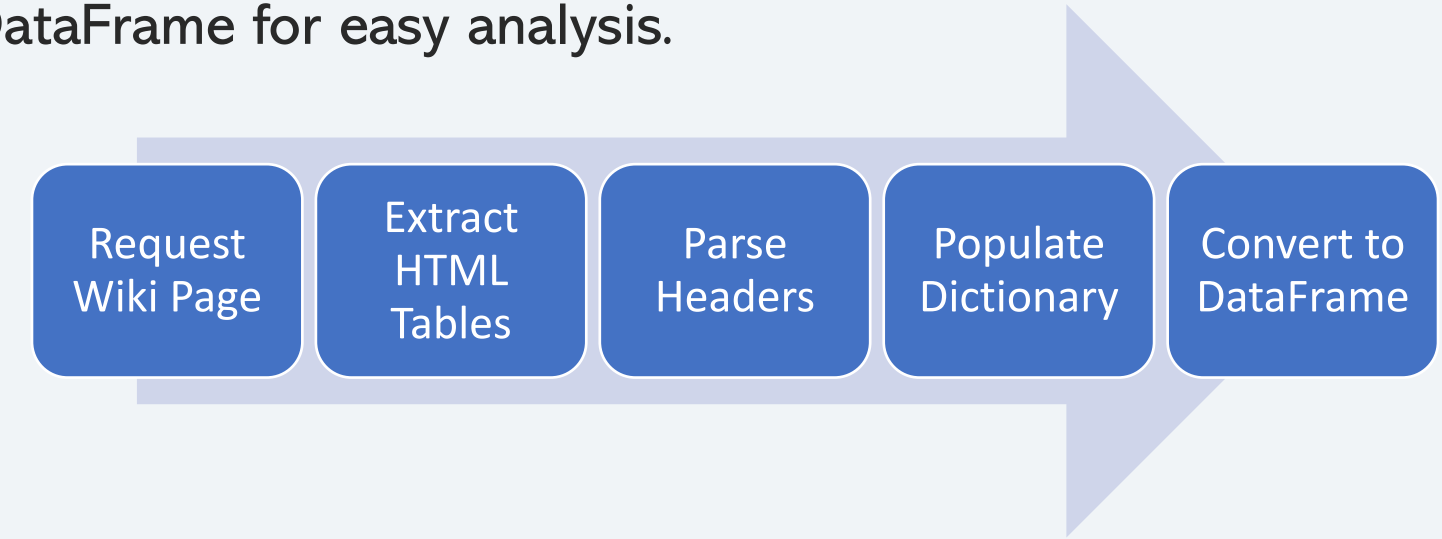
[GitHub Link of the completed SpaceX API calls notebook]



Data Collection – Web Scraping

- Web scraping retrieves and parses HTML content to extract specific data.
- BeautifulSoup navigates the HTML structure to extract tables or elements.
- Extracted data is organized into a Pandas DataFrame for easy analysis.

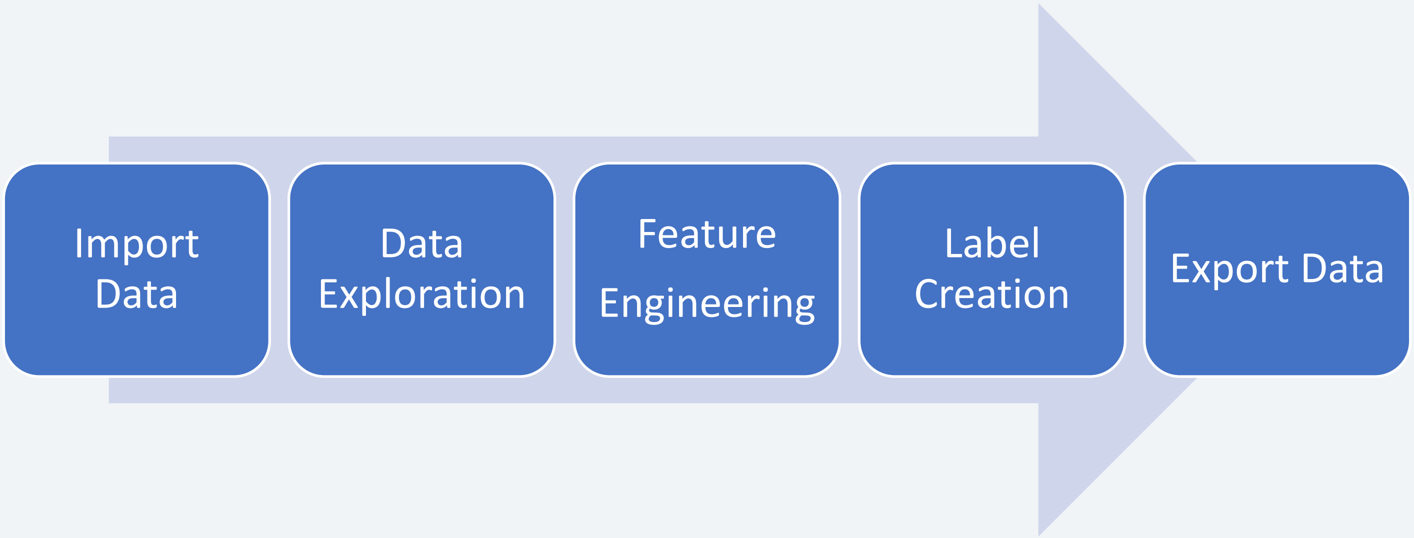
[GitHub Link of the completed web scraping notebook]



Data Collection – Data Wrangling

- Import libraries and load the dataset for analysis.
- Perform EDA to identify patterns and prepare data for modeling.
- Convert outcomes to labels and export the processed data.

[GitHub Link of the completed Data Wrangling notebook]



EDA with Data Visualization

Chart Types and Their Purposes

[GitHub Link of the completed EDA Visualization notebook]

Scatter Plots:

- Used to visualize the relationship between two numerical variables.
- Effective for identifying trends, correlations, and outliers.
- In this case, used to analyze the relationship between launch success, payload mass, flight number, and orbit type.

Bar Chart:

- Used to compare categorical data.
- Effective for visualizing differences in quantities across groups.
- In this case, used to compare success rates across different orbit types and to visualize the distribution of launch sites.

Line Chart:

- Used to visualize trends over time.
- Effective for identifying patterns and changes in data over a period.
- In this case, used to show the change in success rate over the years.

EDA with SQL

Here's a summary of the SQL queries used in the notebook:

[GitHub Link of the completed EDA SQL notebook]

- 1. Display unique launch sites: SELECT DISTINCT "Launch_Site" FROM SPACEXTBL;
- 2. Show 5 records where launch sites begin with "CCA": SELECT * FROM SPACEXTBL WHERE "Launch_Site" LIKE 'CCA%' LIMIT 5;
- 3. Calculate total payload mass for NASA (CRS) missions: SELECT SUM("PAYLOAD_MASS__KG_") AS Total_Payload_Mass FROM SPACEXTBL WHERE "Customer" = 'NASA (CRS)';
- 4. Find average payload mass for F9 v1.1 booster: SELECT AVG("PAYLOAD_MASS__KG_") AS Average_Payload_Mass FROM SPACEXTBL WHERE "Booster_Version" = 'F9 v1.1';
- 5. Get date of first successful ground pad landing: SELECT MIN("Date") AS First_Successful_Landing FROM SPACEXTBL WHERE "Landing_Outcome" = 'Success (ground pad)';
- 6. List boosters with drone ship success and payload 4000-6000 kg: SELECT "Booster_Version" FROM SPACEXTBL WHERE "Landing_Outcome" = 'Success (drone ship)' AND "PAYLOAD_MASS__KG_" > 4000 AND "PAYLOAD_MASS__KG_" < 6000;
- 7. Count successful and failed mission outcomes: SELECT SUM(CASE WHEN "Landing_Outcome" LIKE 'Success%' THEN 1 ELSE 0 END) AS Successful_Outcomes SUM(CASE WHEN "Landing_Outcome" LIKE 'Failure%' THEN 1 ELSE 0 END) AS Failed_Outcomes FROM SPACEXTBL;
- 8. Find boosters carrying maximum payload mass: SELECT "Booster_Version" FROM SPACEXTBL WHERE "PAYLOAD_MASS__KG_" = (SELECT MAX("PAYLOAD_MASS__KG_") FROM SPACEXTBL);
- 9. List 2015 drone ship failures with month, booster, and launch site: SELECT substr("Date", 6, 2) AS Month, "Landing_Outcome", "Booster_Version", "Launch_Site" FROM SPACEXTBL WHERE "Landing_Outcome" = 'Failure (drone ship)' AND substr("Date", 0, 5) = '2015';

Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
- Explain why you added those objects
- Add the GitHub URL of your completed interactive map with Folium map, as an external reference and peer-review purpose

Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard
- Explain why you added those plots and interactions
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model
- You need present your model development process using key phrases and flowchart
- Add the GitHub URL of your completed predictive analysis lab, as an external reference and peer-review purpose

Results

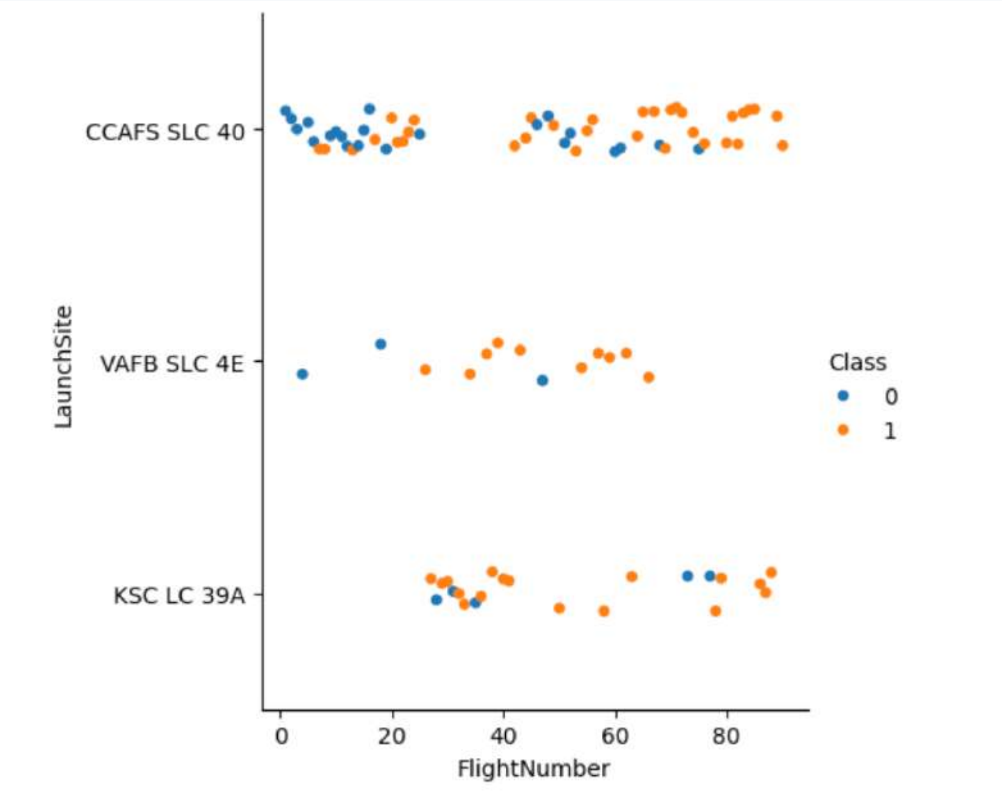
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

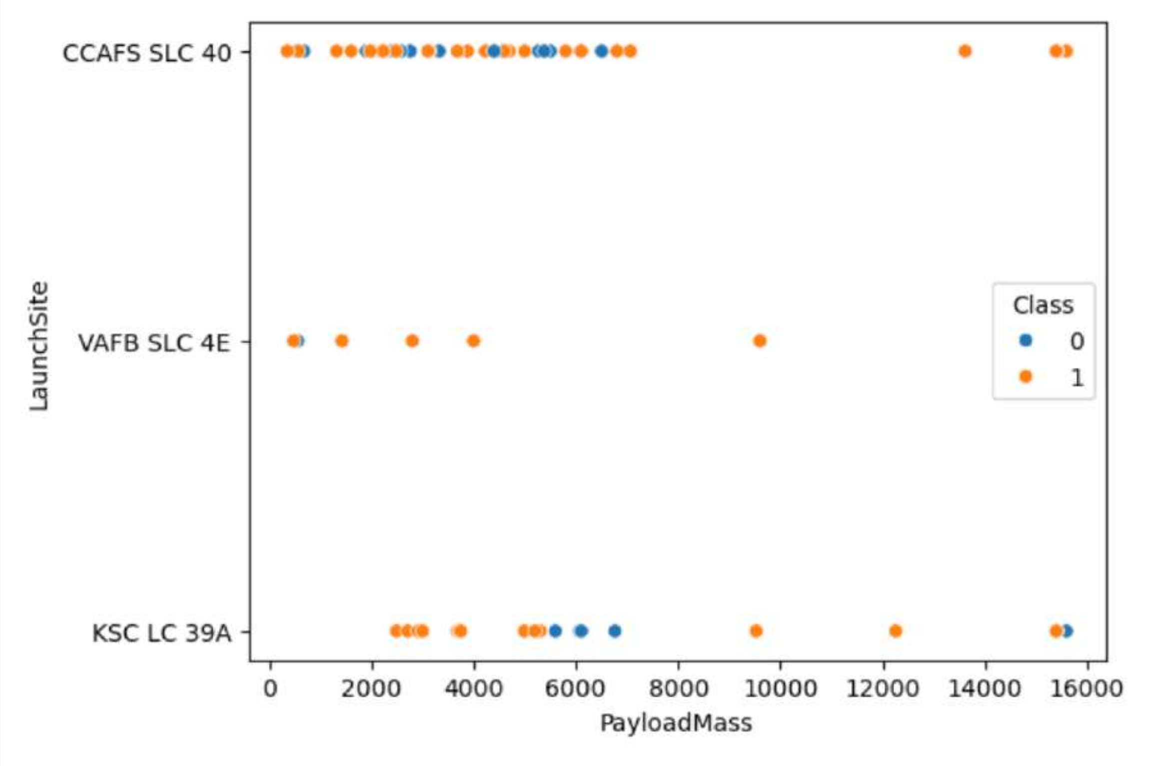
Flight Number vs. Launch Site

- The plot shows the relationship between SpaceX Falcon 9 launch success (Class 1 for success, 0 for failure), and flight number, categorized by launch site.
- There is a clear trend of increasing success rate with higher flight numbers, regardless of launch site.
- While all launch sites show similar success rates overall, VAFB SLC 4E has a slightly higher proportion of successful launches (Class 1) compared to the other two sites.



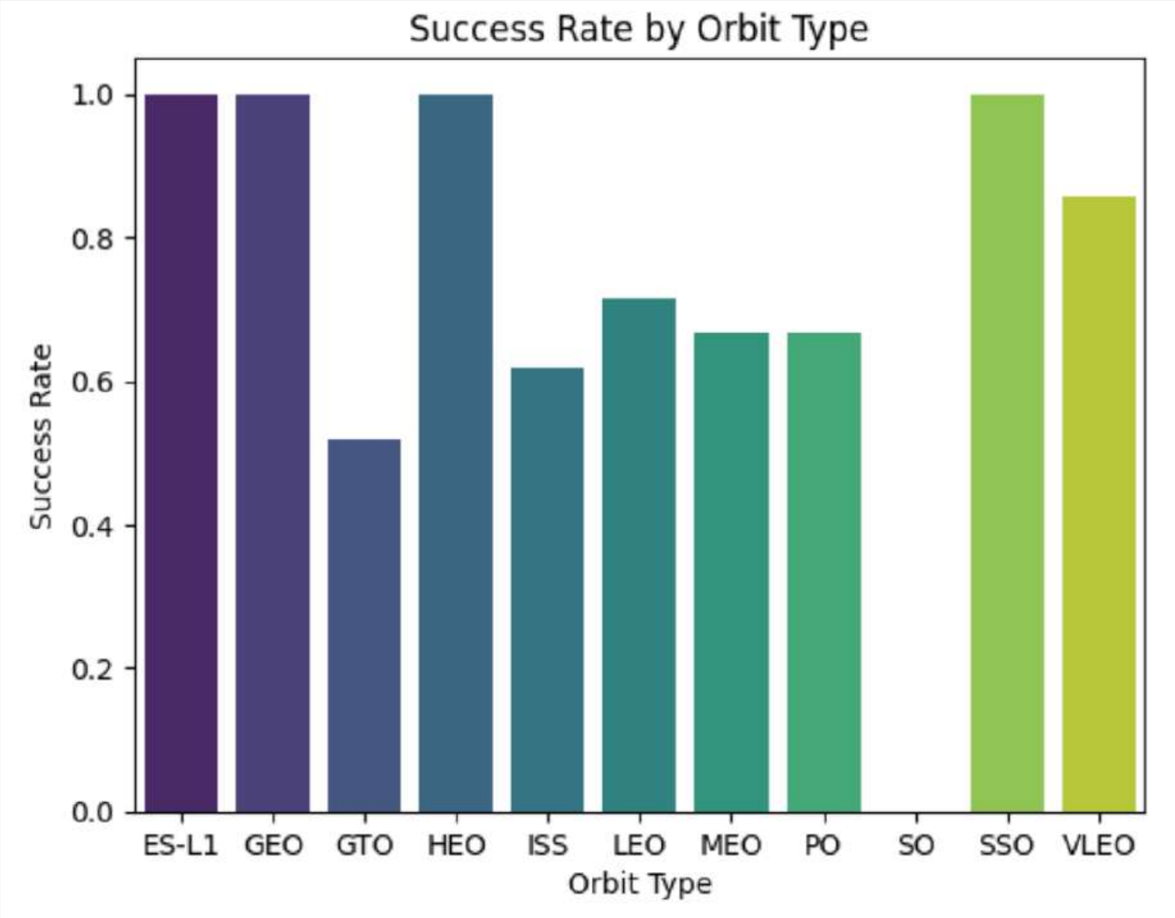
Payload vs. Launch Site

- The scatter plot reveals a relationship between payload mass and launch success (Class 1 for success, 0 for failure), categorized by launch site.
- There seems to be no clear correlation between payload mass and launch success across all launch sites.



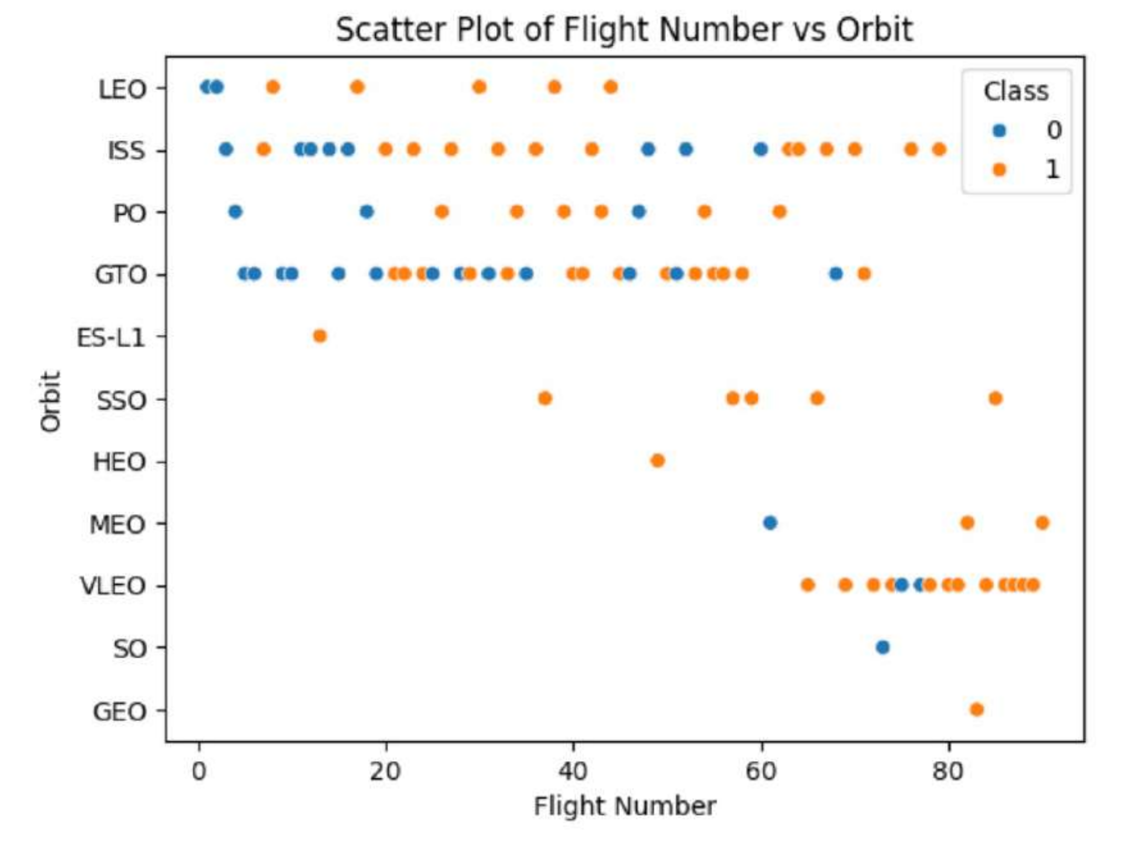
Success Rate vs. Orbit Type

- The chart visualizes the success rate of SpaceX launches across different orbit types.
- The success rate varies significantly between orbit types, with some orbits demonstrating consistently high success rates (e.g., ES-L1, GEO, GTO) and others showing lower success rates (e.g., HEO, ISS).
- The chart provides a clear overview of the performance of SpaceX launches in achieving different orbit types, highlighting areas of strength and potential areas for improvement.



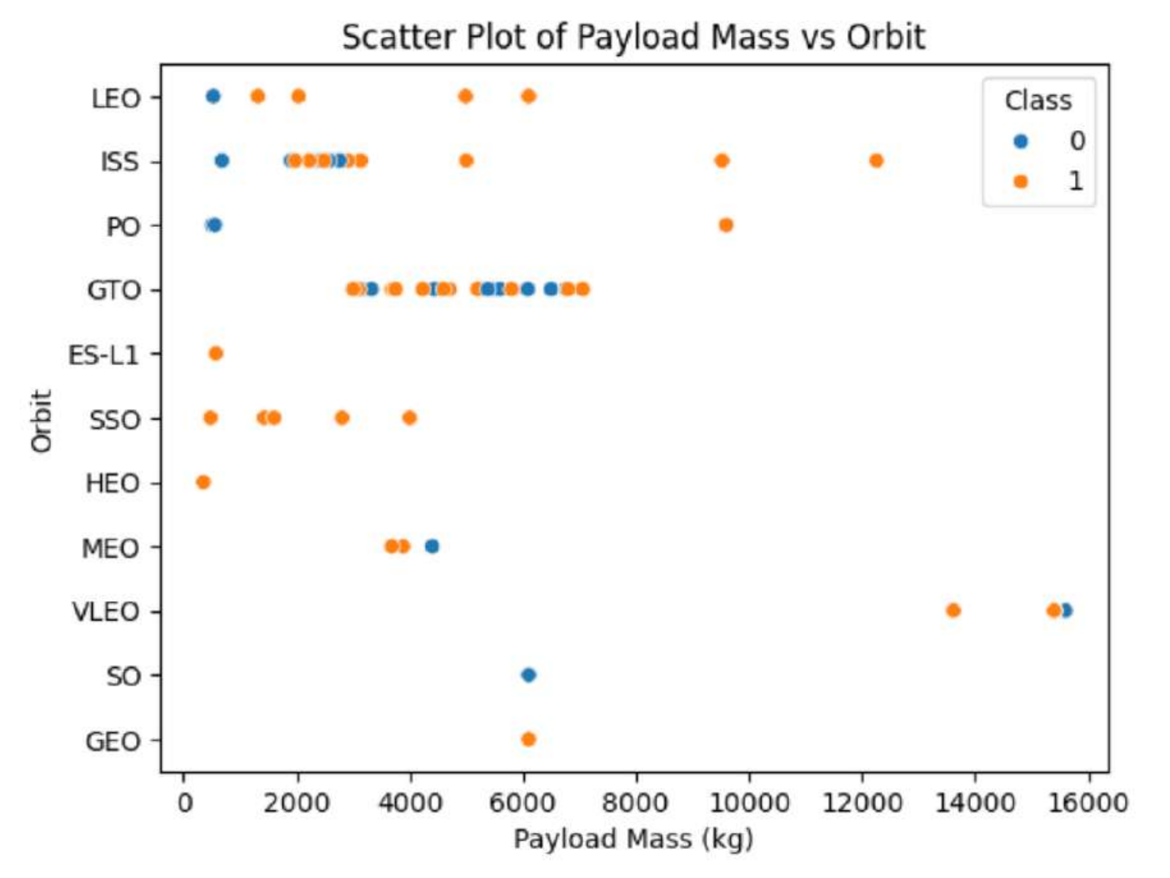
Flight Number vs. Orbit Type

- The scatter plot illustrates the relationship between SpaceX flight number and the orbit type of the mission, with success (Class 1) and failure (Class 0) differentiated by color.
- There is a clear preference for certain orbit types, particularly LEO, ISS, and GTO, with a higher number of launches across flight numbers.
- A potential correlation exists between flight number and success in LEO orbits, with a trend towards higher success rates for later flights. Conversely, no clear relationship between flight number and success is evident in GTO orbits.



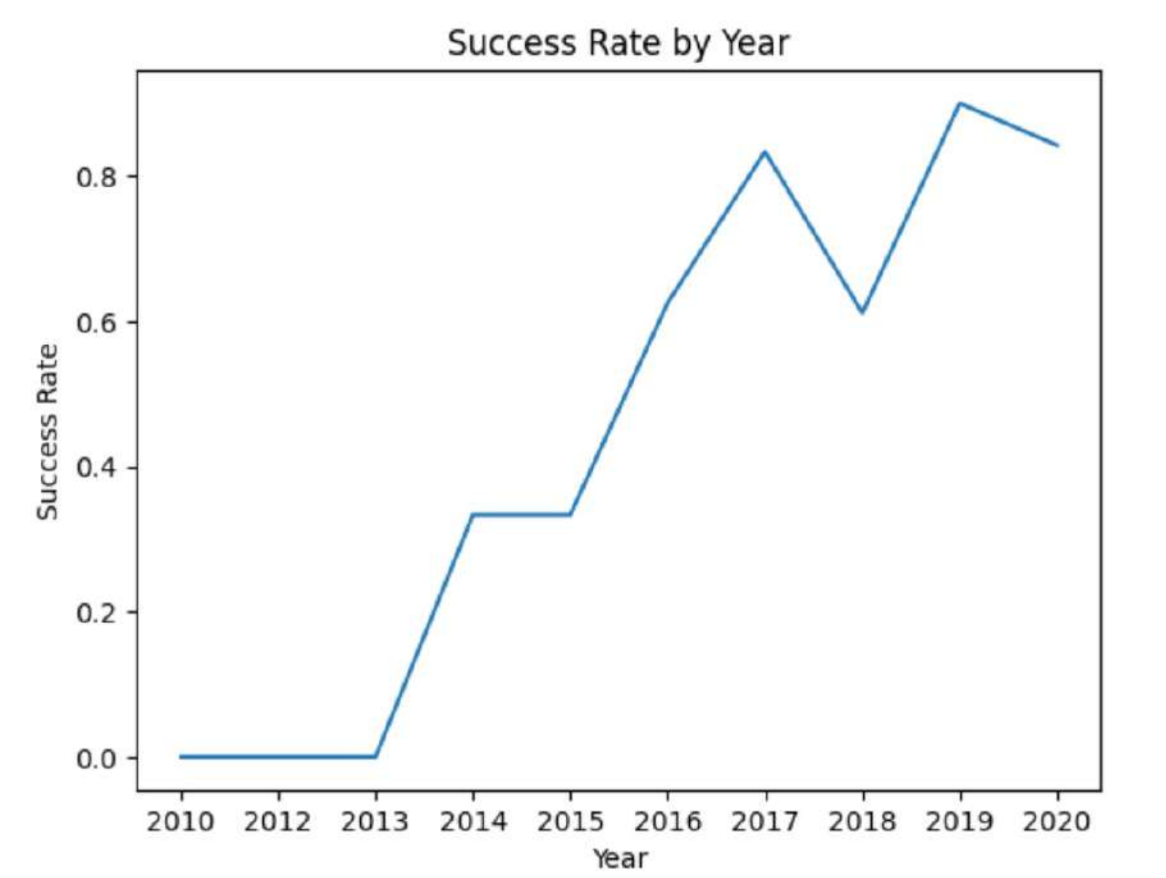
Payload vs. Orbit Type

- The scatter plot illustrates the relationship between payload mass and orbit type, with successful launches (Class 1) and failures (Class 0) differentiated by color.
- There is a clear variation in payload mass across different orbit types, with some orbits (like LEO and ISS) handling a wider range of payload masses compared to others (like ES-L1 and GEO).
- While some orbits (like LEO and ISS) show a mix of successful and failed launches across different payload masses, others (like GTO and SSO) predominantly consist of successful launches with varying payload weights.



Launch Success Yearly Trend

- The graph depicts a clear upward trend in success rate over the years from 2010 to 2020. This is consistent with the general trend observed in some of the previous graphs, particularly those showing an increase in success rate with flight number or time.
- There are notable jumps in success rate between certain years, indicating significant improvements or advancements in technology or processes. For example, the increase between 2013 and 2014 is particularly pronounced.
- The graph paints a positive picture of the performance over time, suggesting continuous progress and improvements in reliability of rockets being launched.



All Launch Site Names

- SQL Query: SELECT DISTINCT "Launch_Site" FROM SPACEXTBL;
- The unique names of the launch site:
  - CCAFS LC-40
  - VAFB SLC-4E
  - KSC LC-39AC
  - CAFS SLC-40

Launch Site Names Begin with 'CCA'

- SQL Query:
  SELECT * FROM SPACEXTBL WHERE "Launch_Site" LIKE 'CCA%' LIMIT 5;

- Insights:
  - **Early Focus:** The table covers early Falcon 9 launches, suggesting a focus on the development and testing phase of the rocket.
  - **Payload Variety:** The payloads include Dragon spacecraft, CubeSats, and other experimental cargo, indicating a diverse range of missions in the early stages.
  - **NASA Collaboration:** NASA appears to be a significant customer for these early launches, with missions like COTS (Commercial Orbital Transportation Services) and CRS (Commercial Resupply Services).
  - **Mission Success:** All listed missions have been marked as "Success," suggesting a strong track record for early Falcon 9 launches.

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

Total Payload Mass

- Total Payload Mass is 45596
- SQL Query:
  SELECT SUM("PAYLOAD_MASS__KG_") AS Total_Payload_Mass FROM SPACEXTBL WHERE "Customer" = 'NASA (CRS)’;
- This SQL query calculates the total payload mass (in kg) for SpaceX launches with the customer "NASA (CRS).”
- It sums up all the values in the "PAYLOAD_MASS__KG_" column.
- The WHERE clause filters for rows where the customer is 'NASA (CRS).’
- The final result is 45,596 kg, representing the total payload mass for these missions.

Total_Payload_Mass
45596

Average Payload Mass by F9 v1.1

- Total Average Payload Mass: 2928.4
- SQL Query:
  SELECT AVG("PAYLOAD_MASS__KG_") AS Average_Payload_Mass
  FROM SPACEXTBL WHERE "Booster_Version" = 'F9 v1.1';

- This SQL query calculates the average payload mass (in kg) for SpaceX launches with the "F9 v1.1" booster version.
- The AVG function is used to find the mean value of the "PAYLOAD_MASS__KG_" column.
- The WHERE clause filters for rows where the booster version is specified as 'F9 v1.1.'
- The final result is an average payload mass of 2,928.4 kg for these launches.

Average_Payload_Mass
2928.4

First Successful Ground Landing Date

- First Successful Landing: 2015-12-22
- SQL Query: SELECT MIN("Date") AS First_Successful_Landing FROM SPACEXTBL WHERE "Landing_Outcome" = 'Success (ground pad)';
- The query identifies the earliest date of a successful ground pad landing by SpaceX.
- It uses the MIN function to find the oldest date in the "Date" column.
- The WHERE clause filters for records where the "Landing_Outcome" is 'Success (ground pad).'
- The result is December 22, 2015, marking SpaceX's first successful ground pad landing.

First_Successful_Landing
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- SQL Query: SELECT "Booster_Version" FROM SPACEXTBL WHERE "Landing_Outcome" = 'Success (drone ship)' AND "PAYLOAD_MASS__KG_" > 4000
  AND "PAYLOAD_MASS__KG_" < 6000;

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

- The SQL query retrieves a list of booster versions that successfully landed on a drone ship.
- It filters for records where the "Landing_Outcome" is 'Success (drone ship).'
- The query also includes a condition where the payload mass is between 4,000 and 6,000 kg.
- The result includes the following booster versions: F9 FT B1022, F9 FT B1026, F9 FT B1021.2, and F9 FT B1031.2.

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes
- SQL Query:
  SELECT SUM(CASE WHEN "Landing_Outcome" LIKE 'Success%' THEN 1 ELSE 0 END) AS Successful_Outcomes, SUM(CASE WHEN "Landing_Outcome" LIKE 'Failure%' THEN 1 ELSE 0 END) AS Failed_Outcomes FROM SPACEXTBL;

- The SQL query counts the number of successful and failed landing outcomes from the SPACEXTBL table.
- It uses SUM with a CASE statement to count rows where "Landing_Outcome" starts with 'Success' as Successful_Outcomes.
- Another SUM with a CASE statement counts rows where "Landing_Outcome" starts with 'Failure' as Failed_Outcomes.
- The results show 61 successful outcomes and 10 failed outcomes.

Successful_Outcomes	Failed_Outcomes
61	10

Boosters Carried Maximum Payload

- SQL Query:
  SELECT "Booster_Version" FROM SPACEXTBL WHERE "PAYLOAD_MASS__KG_" = (SELECT MAX("PAYLOAD_MASS__KG_") FROM SPACEXTBL);
- The SQL query selects the "Booster_Version" from the SPACEXTBL table where the "PAYLOAD_MASS__KG_" equals the maximum payload mass found in the table.
- It uses a subquery to determine the maximum value of "PAYLOAD_MASS__KG_" across all rows in the SPACEXTBL table.
- The outer query then retrieves all "Booster_Version" entries corresponding to this maximum payload mass.
- The result lists multiple booster versions, each associated with the maximum payload mass, indicating these versions all carried the highest payload recorded in the table.

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 substr("Date", 6, 2) AS Month, "Landing_Outcome", "Booster_Version", "Launch_Site" FROM SPACEXTBL WHERE "Landing_Outcome" = 'Failure (drone ship)' AND substr("Date", 0, 5) = '2015';
- The SQL query extracts the month from the "Date" column, along with the "Landing_Outcome", "Booster_Version", and "Launch_Site" from the SPACEXTBL table.
- It filters the results to include only rows where the "Landing_Outcome" is 'Failure (drone ship)'.
- Additionally, it restricts the results to entries where the year portion of the "Date" column is '2015'.
- The result shows the month of failure, the booster version involved, and the launch site for failures on drone ships in 2015, listing specific failures in January and April at CCAFS LC-40.

Month	Landing_Outcome	Booster_Version	Launch_Site
01	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

- SQL Query: SELECT "Landing_Outcome", COUNT(*) AS Outcome_Count FROM SPACEXTABLE WHERE "Date" BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY "Landing_Outcome" ORDER BY Outcome_Count DESC;
- The SQL query counts the number of occurrences for each "Landing_Outcome" within a specified date range from '2010-06-04' to '2017-03-20'.
- It uses the COUNT(*) function to aggregate the number of entries for each type of landing outcome.
- The results are grouped by "Landing_Outcome" and sorted in descending order based on the count of each outcome.
- The result shows the frequency of each landing outcome within the specified date range, with 'No attempt' being the most frequent and 'Precluded (drone ship)' being the least frequent.

Landing_Outcome	Outcome_Count
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

Section 3

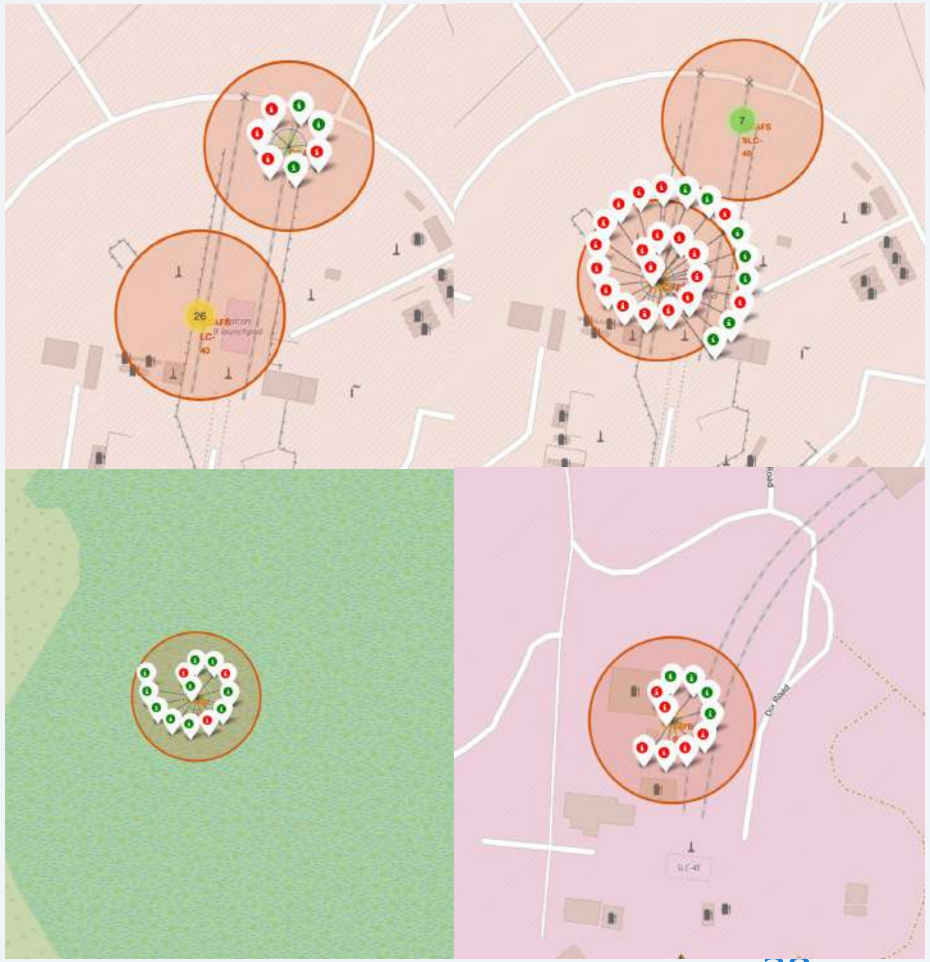
Launch Sites Proximities Analysis

Folium Map: Location Markers

- The markers indicate specific launch sites used by SpaceX for rocket launches, including CCAFS LC-40, CCAFS SLC-40, KSC LC-39A, and VAFB SLC-4E, each identified by their latitude and longitude coordinates.
- These launch sites are strategically located at Cape Canaveral Air Force Station (CCAFS), Kennedy Space Center (KSC), and Vandenberg Air Force Base (VAFB), highlighting key facilities utilized by SpaceX for various missions.

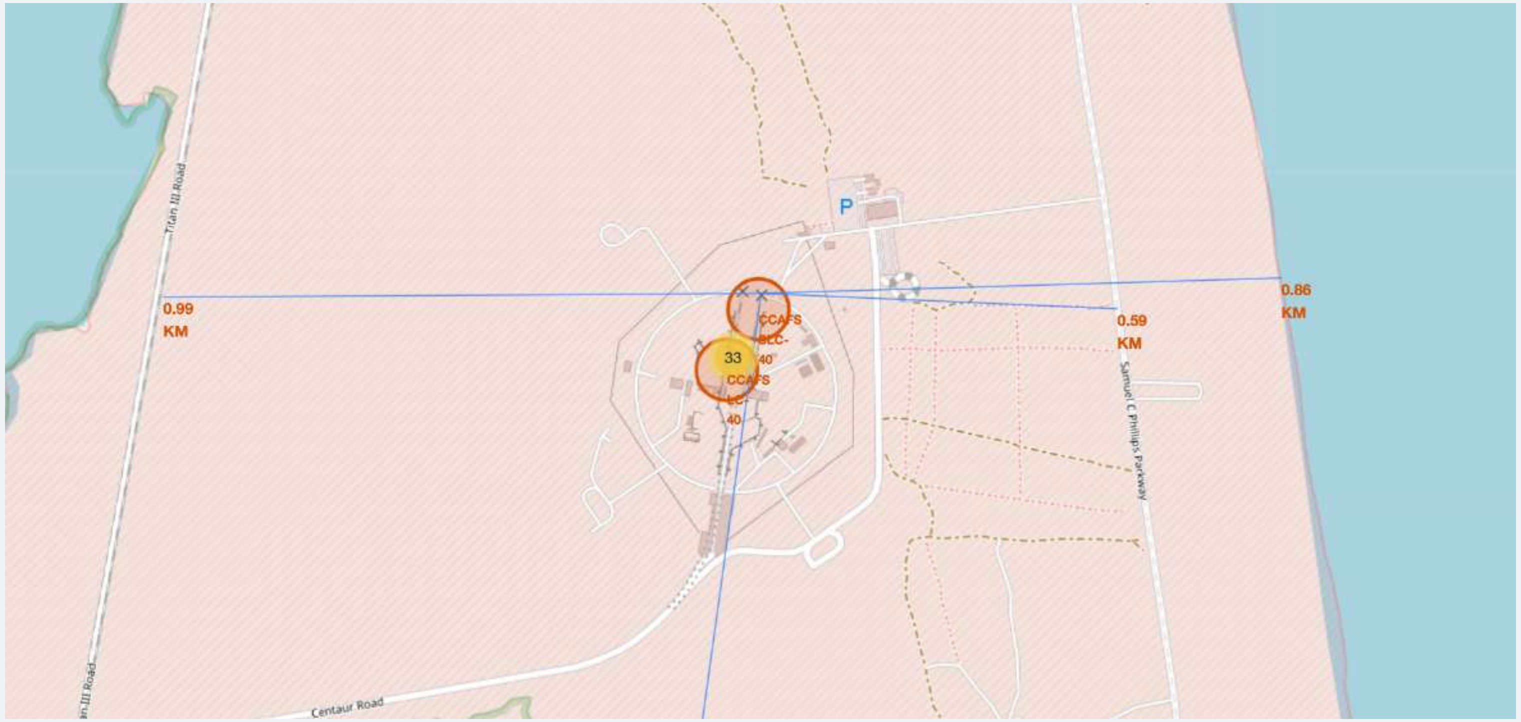
Folium Map: Color-labeled Launch Outcomes

- The image shows four separate launch sites (CCAFS LC-40, CCAFS SLC-40, KSC LC-39A, and VAFB SLC-4E), each marked with red and green indicators that represent the outcomes of SpaceX launches, where red markers indicate failed launches and green markers indicate successful launches.
- The distribution of markers at each site visually highlights the frequency and success rate of launches at these locations, providing an immediate comparison of performance across different launch sites.



Folium Map: Launch Site Proximities

- The image highlights the proximity of a SpaceX launch site to key infrastructure, showing the launch site's closeness to a railway (0.99 km), a highway (0.59 km), and the coastline (0.86 km), while being significantly farther from the nearest city (19.62 km).
- These distances suggest that launch sites are typically situated near critical transportation routes and coastlines for logistical reasons, yet maintain a safe distance from urban areas to minimize risks to populated regions.

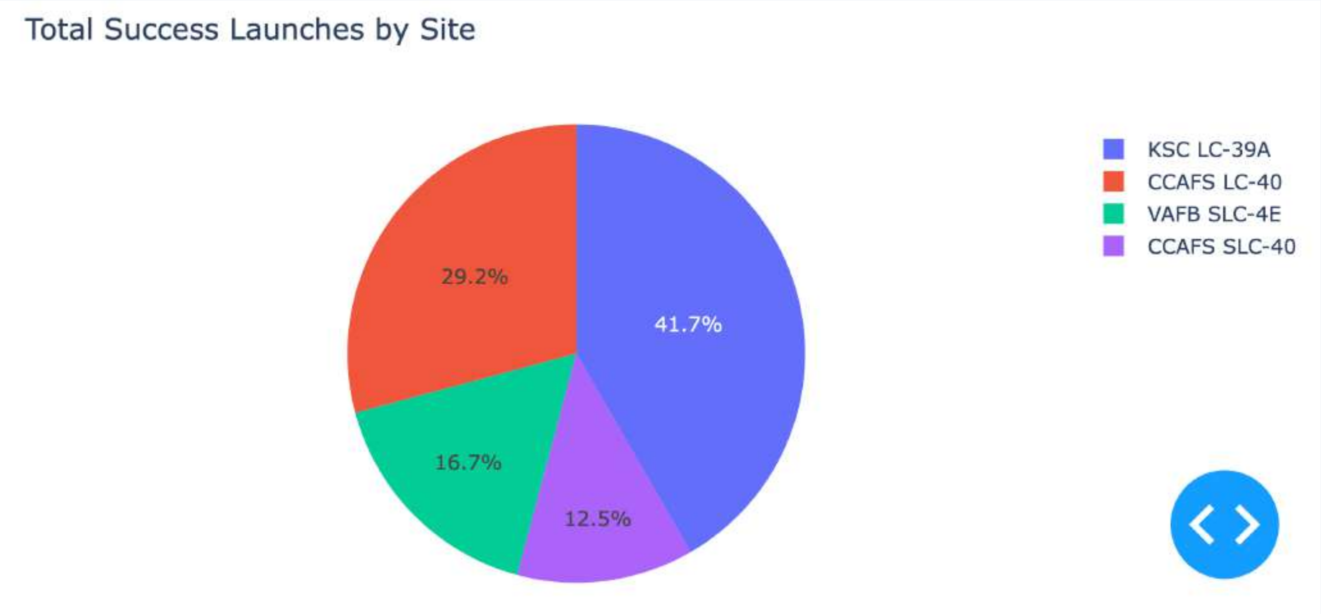


Section 4

Build a Dashboard with Plotly Dash

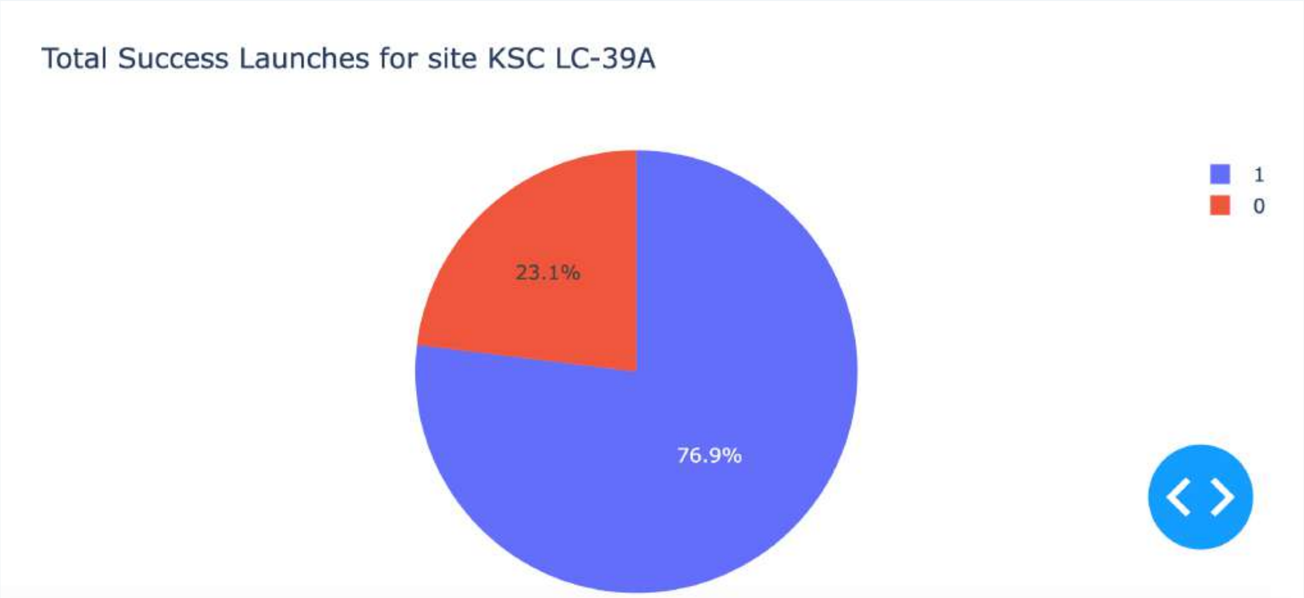
<Dashboard Screenshot 1>

- KSC LC-39A has the highest proportion of successful launches, accounting for 41.7% of the total successful launches, indicating it is SpaceX's most frequently used and likely most reliable launch site.
- CCAFS LC-40 follows with 29.2% of successful launches, reflecting its significance as a major launch site for SpaceX operations.
- VAFB SLC-4E and CCAFS SLC-40 contribute smaller shares, with 16.7% and 12.5% of the successful launches, respectively, highlighting their roles as secondary launch sites.



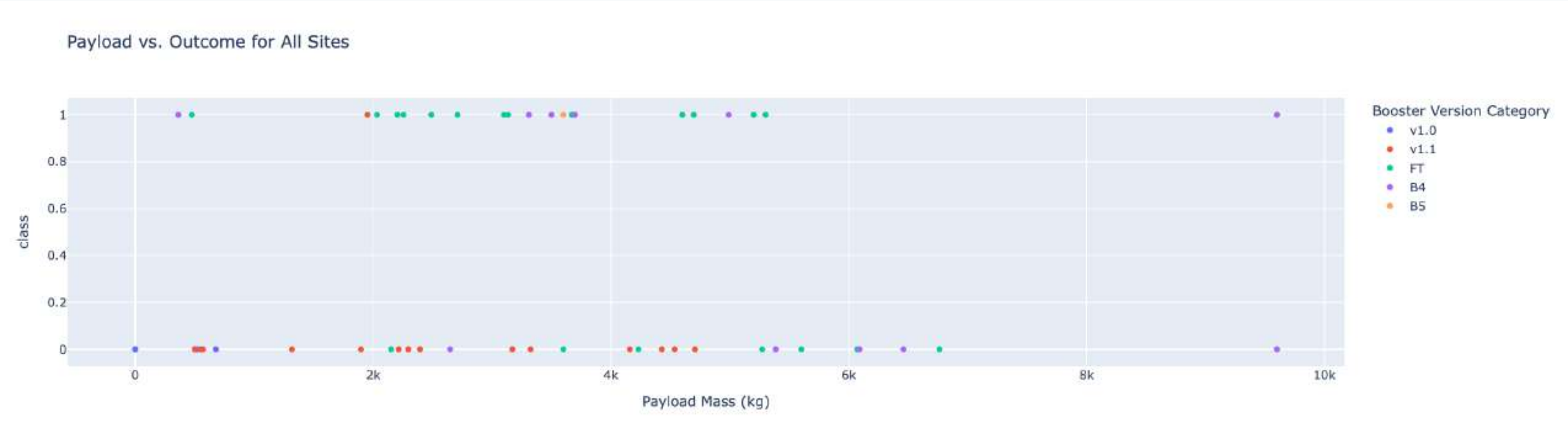
<Dashboard Screenshot 2>

- Site KSC LC-39A has a high rate of successful launches. The pie chart indicates that 76.9% of launches from this site have been successful. This suggests that it is a reliable and efficient launch pad.
- There is still room for improvement. While the success rate is high, 23.1% of launches have failed. Investigating the reasons for these failures could help improve the overall performance of the site.



Dashboard: Payload vs. Launch Outcome

- No clear correlation between payload mass and success. There appears to be no direct relationship between the amount of cargo a rocket carries and the likelihood of a successful launch. The data points are scattered across the graph without a clear pattern.
- Variation between booster versions. The different colors indicating the booster versions suggest that there might be performance differences between these versions. Some booster versions seem to have more successful launches with certain payload masses.

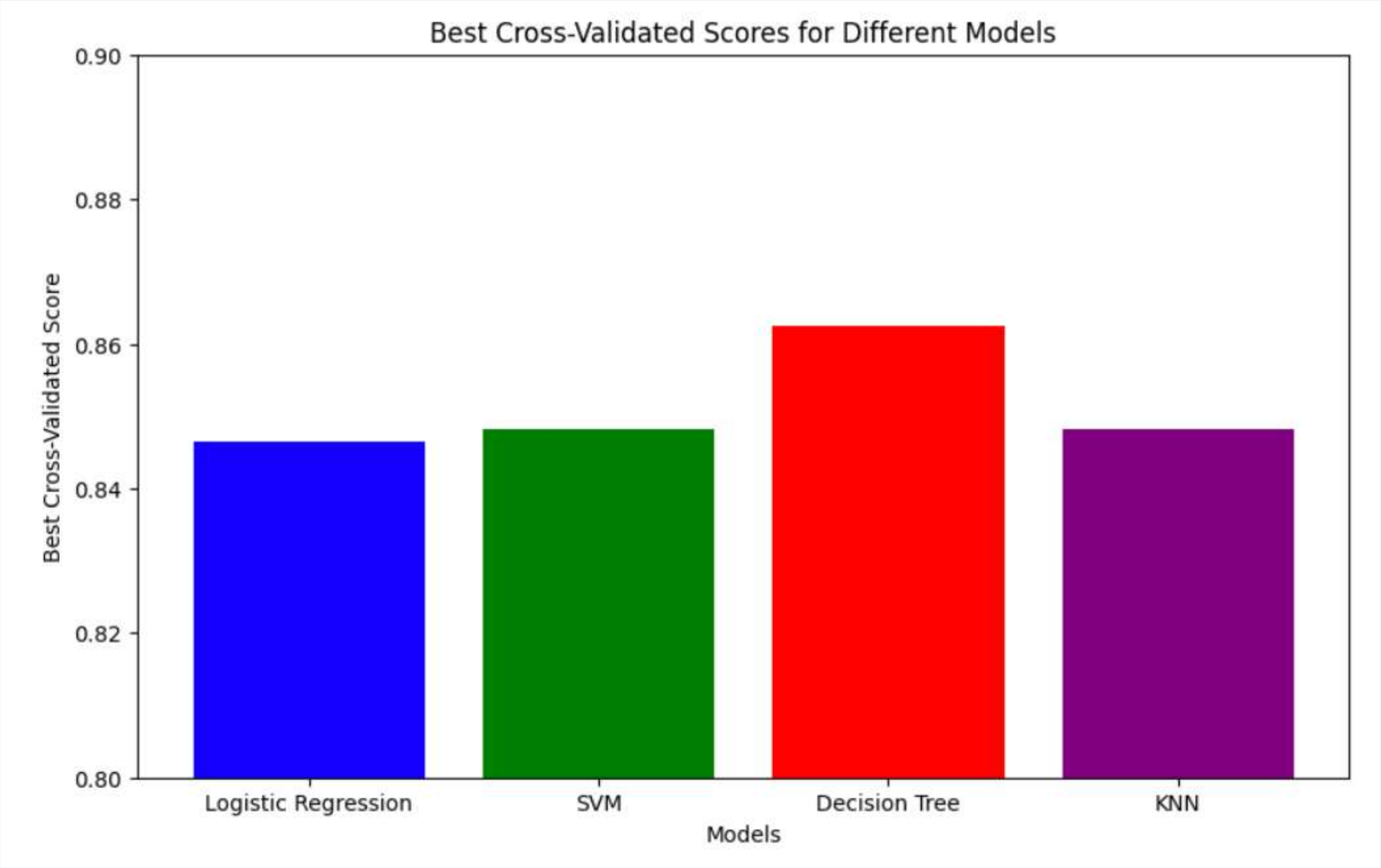


Section 5

Predictive Analysis (Classification)

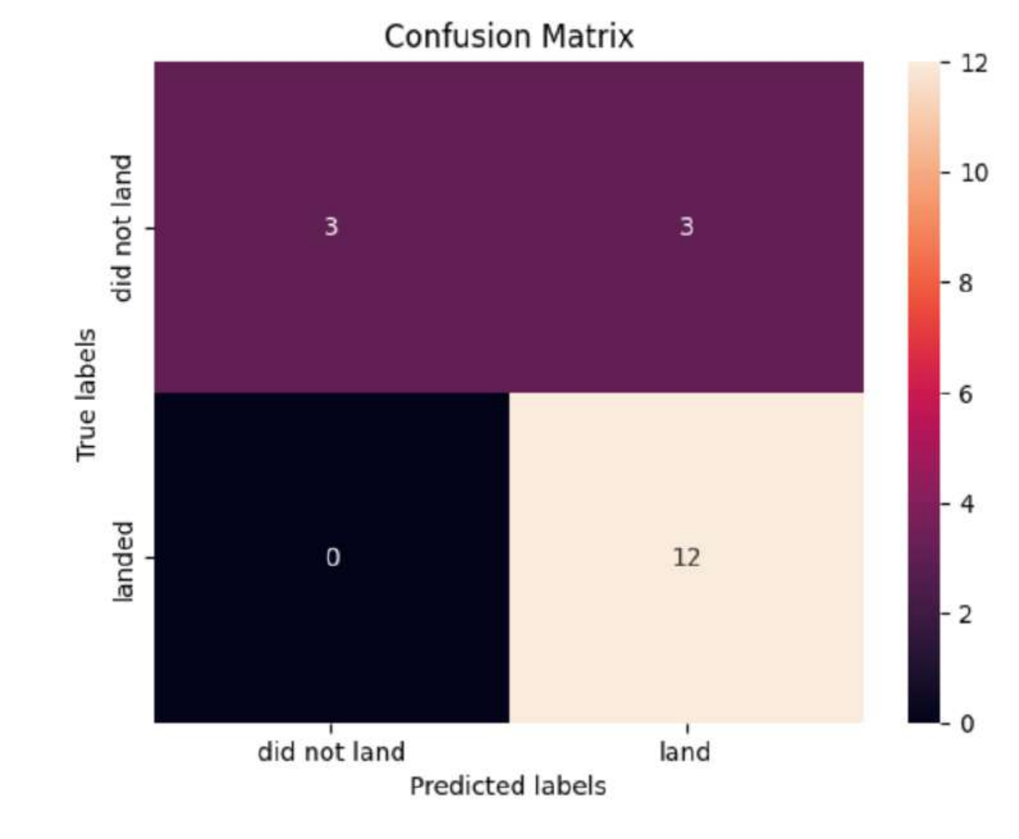
Classification Accuracy

- Decision Tree has the highest best cross-validated score.
- Logistic Regression has the lowest best cross-validated score.
- The scores for SVM, Decision Tree, and KNN are relatively close, with Decision Tree showing the highest performance.



Confusion Matrix: Decision Tree

- High accuracy for predicting 'landed' cases: The model correctly predicts 12 out of 12 instances where the actual outcome was a 'landed' spacecraft.
- Low accuracy for predicting 'did not land' cases: The model incorrectly predicts 3 out of 5 instances where the actual outcome was a 'did not land' spacecraft.
- Overall, the model demonstrates a strong performance in predicting 'landed' cases but struggles with accurately predicting 'did not land' cases.



Conclusions

- Launch Success Trends: Observed an increasing success rate with higher flight numbers across all launch sites
- Noted a clear upward trend in success rate from 2010 to 2020
- Payload and Orbit Analysis: Found no clear correlation between payload mass and launch success
- Identified varying success rates across different orbit types
- Launch Site Performance: KSC LC-39A showed the highest proportion of successful launches (41.7% of total successes)
- KSC LC-39A also demonstrated a high success rate of 76.9% for its launches
- Predictive Model Performance: Decision Tree model achieved the highest cross-validated score among tested models
- The best model showed high accuracy in predicting successful landings but struggled with predicting landing failures
- Geographical Insights: Launch sites are strategically located near transportation routes and coastlines, while maintaining safe distances from urban areas

Appendix

- Link to the notebooks and code

Thank you!