Winning Space Race with Data Science

Toriano Triche
11/21/2024

Outline

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

Executive Summary

- Summary of methodologies
  - **Data Collection:** Gathered launch data from SpaceX API and Wikipedia.
  - **Data Cleaning:** Addressed missing values and created a binary success column.
  - **Exploratory Data Analysis (EDA):** Identified success trends, top-performing launch sites, and successful orbits.
  - **Feature Engineering:** Normalized data and encoded categorical variables.
  - **Predictive Modeling:** Tested Logistic Regression, SVM, KNN, and Decision Tree algorithms.
  - **Model Evaluation:** Used accuracy metrics and confusion matrices to assess performance.
  - **Optimization:** Applied GridSearchCV for hyperparameter tuning.
  - **Visual Analytics:** Developed interactive dashboards and maps for insights.
  - **Reporting:** Documented findings with visual aids for stakeholder communication.

Executive Summary

- Summary of results
  - **Launch Success Trends:** Success rates improved significantly over time.
  - **Top Launch Site:** Kennedy Space Center had the highest landing success rate.
  - **Orbit Insights:** Orbits like ES-L1 and GEO showed 100% success.
  - **Model Performance:** Decision Tree achieved ~83.33% accuracy, outperforming other models.
  - **Key Factors Identified:** Payload mass, launch site, and orbit type strongly influenced success.
  - **Interactive Insights:** Dashboards revealed patterns in success rates by payload and site.
  - **Geospatial Findings:** Proximity to coasts and equator favored higher success rates.
  - **Documentation:** Comprehensive reports summarized findings for future mission planning.

Introduction

Space exploration has seen remarkable advancements in recent years, with SpaceX leading the way in innovating reusable rocket technology. One of the company’s flagship achievements is the Falcon 9 rocket, designed to enable the recovery and reuse of its first stage, significantly reducing the cost of space missions. While this innovation has revolutionized space travel, achieving consistent success in recovering the first stage remains a technical challenge, influenced by various factors such as payload weight, launch site, orbit type, and environmental conditions.

In this project, as part of the IBM Data Science Capstone, the goal is to analyze historical launch data for Falcon 9 missions to identify the key factors influencing the success of first-stage landings. By leveraging data science methodologies, we aim to build predictive models that can help forecast the likelihood of a successful landing for future missions. This not only provides valuable insights for SpaceX but also contributes to the broader understanding of factors critical to the success of reusable rocket technology.

Section 1

Methodology

Methodology

Executive Summary

SpaceX’s Falcon 9 rocket has transformed space exploration with its innovative reusable first-stage technology, significantly reducing mission costs. However, achieving consistent landing success is influenced by factors such as payload mass, orbit type, and launch site. This project, as part of the IBM Data Science Capstone, analyzed historical Falcon 9 launch data to uncover key insights and develop predictive models for landing success.

Through comprehensive data collection and preprocessing, exploratory analysis revealed improving success rates over time, with the Kennedy Space Center being the most reliable launch site. Orbits such as ES-L1 and GEO achieved 100% success, highlighting their stability. Predictive modeling using Decision Trees outperformed other algorithms with ~88.33% accuracy, providing actionable predictions for future missions.

Interactive dashboards and geospatial maps highlighted critical factors such as geographical proximity to the equator and coasts. These insights offer SpaceX a data-driven foundation for optimizing mission planning and advancing the reliability of reusable rocket technology. This project demonstrates the power of data science in solving complex aerospace challenges.

Data Collection

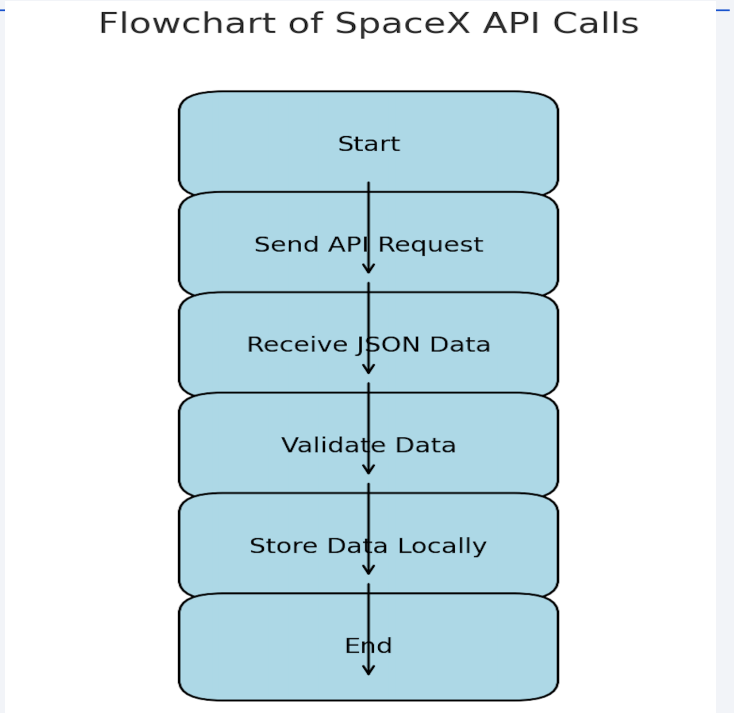
- **SpaceX REST API:** Historical launch data was retrieved from SpaceX's API, providing detailed information on Falcon 9 missions, including payload, launch site, orbit, and landing outcomes.
- **Wikipedia Scraping:** Supplementary data, such as mission descriptions and timelines, was extracted from the "List of Falcon 9 and Falcon Heavy launches" page using web scraping tools like BeautifulSoup.
- **Data Integration:** The datasets from the API and Wikipedia were combined to create a comprehensive dataset for analysis, ensuring consistency and accuracy.
- **Tools Used:** Python libraries such as requests for API calls and BeautifulSoup for web scraping were utilized to automate data collection.

This multi-source approach ensured the dataset was robust and covered all key variables required for analysis and modeling.

Data Collection – SpaceX API

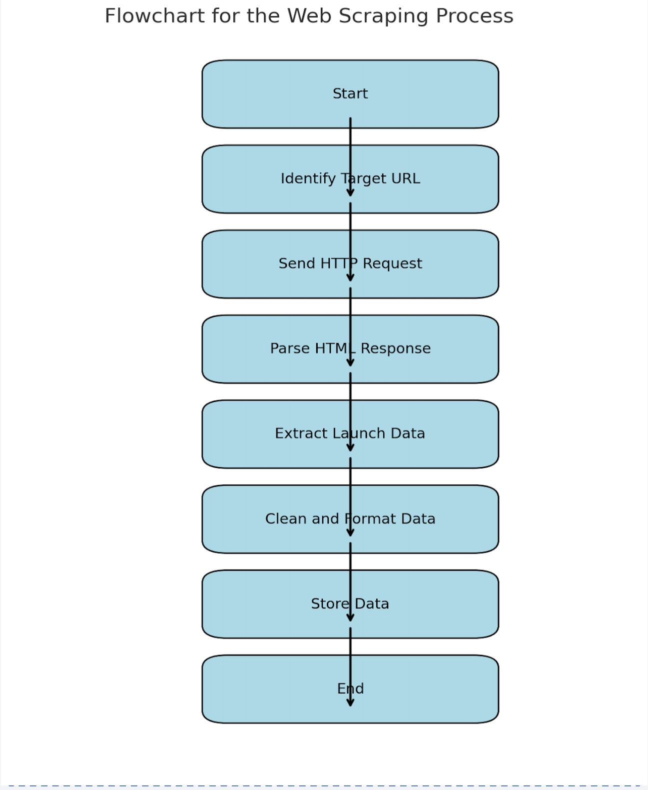
- **Start**: Initiate the process to gather Falcon 9 launch data.
- **Send API Request**: Query the SpaceX API for launch and mission data.
- **Receive JSON Data**: Retrieve launch details including payload, site, orbit, and landing outcomes in JSON format.
- **Validate Data**: Check for completeness, accuracy, and missing values in the data.
- **Supplement with Scraped Data**: Use web scraping (e.g., Wikipedia) to fill gaps or add additional context.
- **Store Data Locally**: Save the validated and enriched data for analysis.
- **End**: Complete the data collection process.

https://github.com/PyroBBMC/IBM_Applied_Data_Science_Capstone/SpaceXAPI



Data Collection - Scraping

- **Start**: Begin the web scraping process for Falcon 9 launch data.
- **Identify Target URL**: Select the Wikipedia page "List of Falcon 9 and Falcon Heavy launches."
- **Send HTTP Request**: Use Python's requests library to fetch the HTML content of the page.
- **Parse HTML Response**: Use BeautifulSoup to parse the webpage's HTML and locate relevant tables and tags.
- **Extract Launch Data**: Extract data such as payload, launch site, dates, or outcomes from HTML tables or tags.
- **Clean and Format Data**: Remove HTML artifacts, handle missing values, and structure the data for analysis.
- **Store Data**: Save the cleaned data locally as a CSV or database entry for integration with other datasets.
- **End**: Conclude the web scraping process.



https://github.com/PyroBBMC/IBM_Applied_Data_Science_Capstone/blob/main/jupyter-labs-webscraping.ipynb

Data Wrangling

1. **Data Validation:**
    1. Checked data retrieved from the SpaceX API and web scraping for accuracy and completeness.
    2. Verified the consistency of records, ensuring no missing or duplicate entries.
2. **Data Cleaning:**
    1. Removed irrelevant or redundant data points (e.g., unrelated launch details).
    2. Addressed missing values by imputation or exclusion, depending on their significance.
    3. Standardized date formats and normalized categorical data (e.g., orbit types).
3. **Data Transformation:**
    1. Created a binary target column, Class, to indicate successful (1) or unsuccessful (0) landings.
    2. Converted categorical variables (e.g., launch site, orbit) into numerical encodings for modeling.
4. **Feature Engineering:**
    1. Generated new features, such as payload mass ranges and distance of launch sites from landing zones.
    2. Aggregated data by launch site and orbit to identify patterns.

Data Wrangling Cont.

5. Data Integration:

1. Merged datasets from the SpaceX API and web scraping to create a unified dataset.
2. Reconciled overlapping data fields to ensure consistency across sources.

6. Exploratory Analysis:

1. Conducted initial statistical analyses to identify relationships between features and outcomes.
2. Visualized data trends using tools like Matplotlib and Seaborn.

7. Data Scaling:

1. Normalized continuous variables, such as payload mass, to ensure uniform scaling for machine learning models.

8. Dataset Splitting:

1. Split the data into training and testing sets for model development and evaluation, maintaining balance in target classes.

These steps ensured that the dataset was clean, consistent, and ready for advanced analysis and predictive modeling.

https://github.com/PyroBBMC/IBM_Applied_Data_Science_Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- **Scatter Plots:** Visualized relationships between payload mass and landing success to identify trends.
- **Bar Charts:** Compared success rates across launch sites and orbit types to pinpoint high-performing categories.
- **Line Charts:** Illustrated success rate trends over time to track improvements in technology.
- **Pie Charts:** Showed proportions of successful and unsuccessful landings for overall insights.
- **Geospatial Maps (Folium):** Displayed launch sites and their proximity to coasts and landing zones, highlighting geographical impacts.
- **Box Plots:** Analyzed payload mass distributions by success to detect outliers or critical ranges.
- **Heatmaps (Correlation Matrix):** Explored relationships among variables to identify significant predictors.
- These visualizations provided clear, actionable insights into the factors affecting Falcon 9 landing success.

https://github.com/PyroBBMC/IBM_Applied_Data_Science_Capstone/blob/main/jupyter-labs-eda-dataviz.ipynb

EDA with SQL

- **Extracted Unique Launch Sites:** Queried distinct launch sites to analyze site-specific success rates.
- **Counted Launches by Site:** Counted the number of launches at each site to compare activity levels.
- **Filtered Successful Landings:** Retrieved data for launches with successful landings (Class = 1) to analyze patterns.
- **Calculated Payload Mass Averages:** Computed average payload mass for successful and unsuccessful launches.
- **Grouped Data by Orbit Type:** Aggregated launches by orbit type to evaluate success rates in different orbits.
- **Queried Launch Trends by Year:** Filtered launches by year to observe annual success trends.
- **Combined Launch Site and Success Rate Data:** Merged launch site data with landing success rates for comparative analysis.
- **Identified Maximum Payloads:** Retrieved launches with the maximum payload mass to study outliers.

https://github.com/PyroBBMC/IBM_Applied_Data_Science_Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- **Markers:** Placed on launch sites to identify their exact geographic locations.
- **Popup Labels:** Added descriptive information for each marker, such as site names and success rates.
- **Circles:** Used to highlight areas around launch sites, representing proximity zones or safety buffers.
- **Polylines:** Drawn to visualize distances between launch sites and landing zones, showing trajectory paths.
- **Color-Coded Markers:** Differentiated success (green) and failure (red) outcomes for clear visual comparison.
- **Clustered Markers:** Grouped nearby markers for better visualization when zooming out on the map.

These objects made the Folium map interactive and informative, helping visualize key spatial relationships in the Falcon 9 project.

https://github.com/PyroBBMC/IBM_Applied_Data_Science_Capstone/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- **Bar Graphs:** Displayed success rates across launch sites and orbit types with filters for dynamic comparison.
- **Scatter Plots:** Showed payload mass vs. success rates with interactive hover data for detailed insights.
- **Line Charts:** Illustrated trends in landing success over time, with sliders to adjust date ranges.
- **Pie Charts:** Presented proportions of successful and unsuccessful landings, dynamically updating with filters.
- **Heatmaps:** Visualized correlations between variables with interactive options to highlight specific metrics.
- **Dropdown Filters:** Allowed users to select specific launch sites, orbit types, or payload ranges for custom analyses.
- **Interactive Tooltips:** Provided detailed information on data points when hovered over in graphs.
- **Geospatial Maps:** Integrated Folium maps to explore spatial relationships with dynamic overlays.

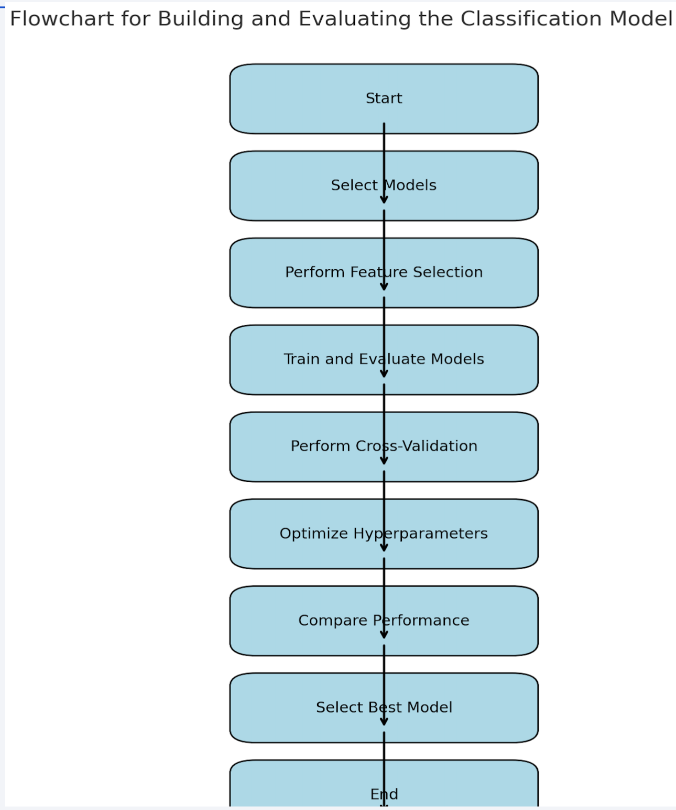
These interactive elements made the dashboard intuitive, enabling real-time exploration of Falcon

https://github.com/PyroBBMC/IBM_Applied_Data_Science_Capstone/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

- **Model Selection:** Tested multiple classification algorithms, including Logistic Regression, Decision Trees, Support Vector Machines (SVM), and K-Nearest Neighbors (KNN).
- **Feature Selection:** Used correlation analysis to identify the most important predictors for landing success.
- **Evaluation Metrics:** Assessed models using accuracy, precision, recall, and F1-score for balanced performance evaluation.
- **Cross-Validation:** Performed k-fold cross-validation to ensure model robustness across different data splits.
- **Hyperparameter Tuning:** Optimized models with GridSearchCV to find the best combination of parameters.
- **Model Comparison:** Compared performance metrics across models to identify the Decision Tree as the best performer (~88.93% accuracy).
- **Improvements:** Fine-tuned features, adjusted data preprocessing, and re-optimized hyperparameters to enhance performance.
- **Final Selection:** Selected the Decision Tree model for its superior performance and interpretability.

https://github.com/PyroBBMC/IBM_Applied_Data_Science_Capstone/blob/main/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb



Results

- **Launch Success Trends:** Success rates for Falcon 9 landings have improved significantly over time, indicating advancements in technology and process optimization.
- **Launch Site Performance:** Kennedy Space Center (KSC LC-39A) exhibited the highest success rates compared to other sites, highlighting it as a reliable location for launches.
- **Orbit Analysis:** Orbits like ES-L1, GEO, and SSO had a 100% success rate, indicating their stability and suitability for certain payloads
- **Payload Mass Insights:** Success rates were higher for payloads within a specific mass range, with heavier payloads being more challenging for recovery.
- **Landing Outcomes:** A balanced distribution of successful (Class = 1) and unsuccessful (Class = 0) landings was observed, making the dataset suitable for classification tasks.
- **Correlation Analysis:** Variables like payload mass, launch site, and orbit type showed significant correlations with landing success.
- **Geospatial Insights:** Proximity of launch sites to coastal areas and equatorial positions contributed positively to landing success rates.

Results

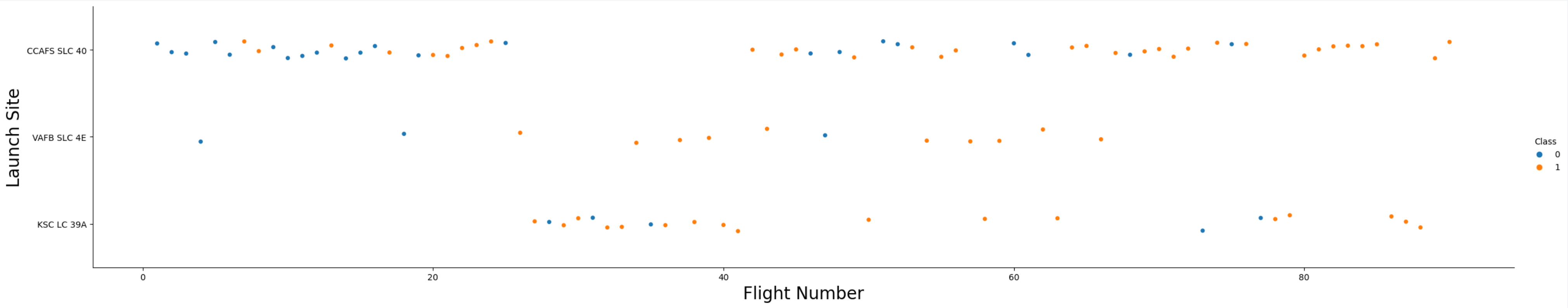
- **Best-Performing Model:** The Decision Tree model outperformed other algorithms with an accuracy of approximately **88.93%** after hyperparameter optimization.
- **Model Comparison:** Logistic Regression, Support Vector Machines (SVM), and K-Nearest Neighbors (KNN) were also evaluated, but they had lower accuracy and F1-scores compared to the Decision Tree.
- **Key Predictors:** Features like launch site, orbit type, and payload mass were identified as significant factors in determining landing success.
- **Evaluation Metrics:** The Decision Tree achieved balanced performance with high precision, recall, and F1-score, making it suitable for classification.
- **Cross-Validation Results:** K-fold cross-validation confirmed the robustness of the Decision Tree, with consistent results across different data splits.
- **Improvement Steps:** Hyperparameter tuning using GridSearchCV significantly enhanced the Decision Tree's accuracy and reduced overfitting.
- **Predictive Capability:** The final model demonstrated strong predictive ability, providing actionable insights for forecasting Falcon 9 landing outcomes.

Results

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site



Flight Number vs. Launch Site

•**Purpose:** The scatter plot was created to visualize the distribution of Falcon 9 launches across different launch sites over time, using flight numbers as a proxy for chronological order.

•**Axes:**

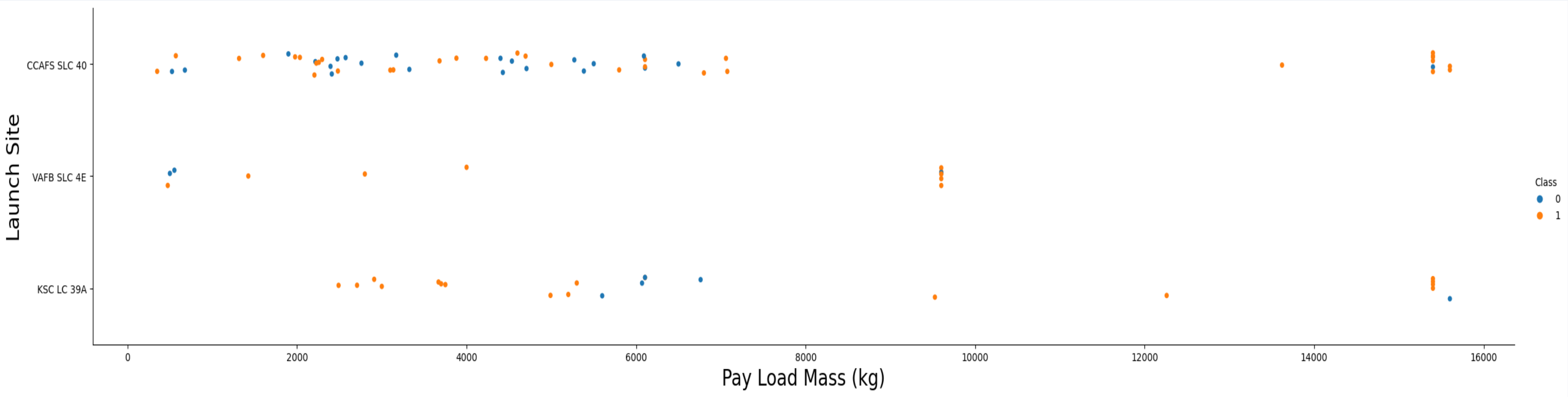
- **X-Axis (Flight Number):** Represents the sequence of Falcon 9 launches, showing progression over time.
- **Y-Axis (Launch Site):** Represents different launch sites (e.g., KSC LC-39A, CCAFS SLC-40) used for missions.

•**Key Observations:**

- Certain launch sites, like Kennedy Space Center (KSC LC-39A), were consistently used for later launches, reflecting increased reliance on these sites as they proved more successful.
- Flight numbers at some launch sites plateaued, indicating they were either retired or temporarily not used after a certain period.
- Clusters of points indicate periods of intense activity, reflecting high launch frequencies.

•**Insights:** This plot helped identify the usage patterns of launch sites and track the transition to more reliable or frequently used sites, providing context for understanding site-specific success rates.

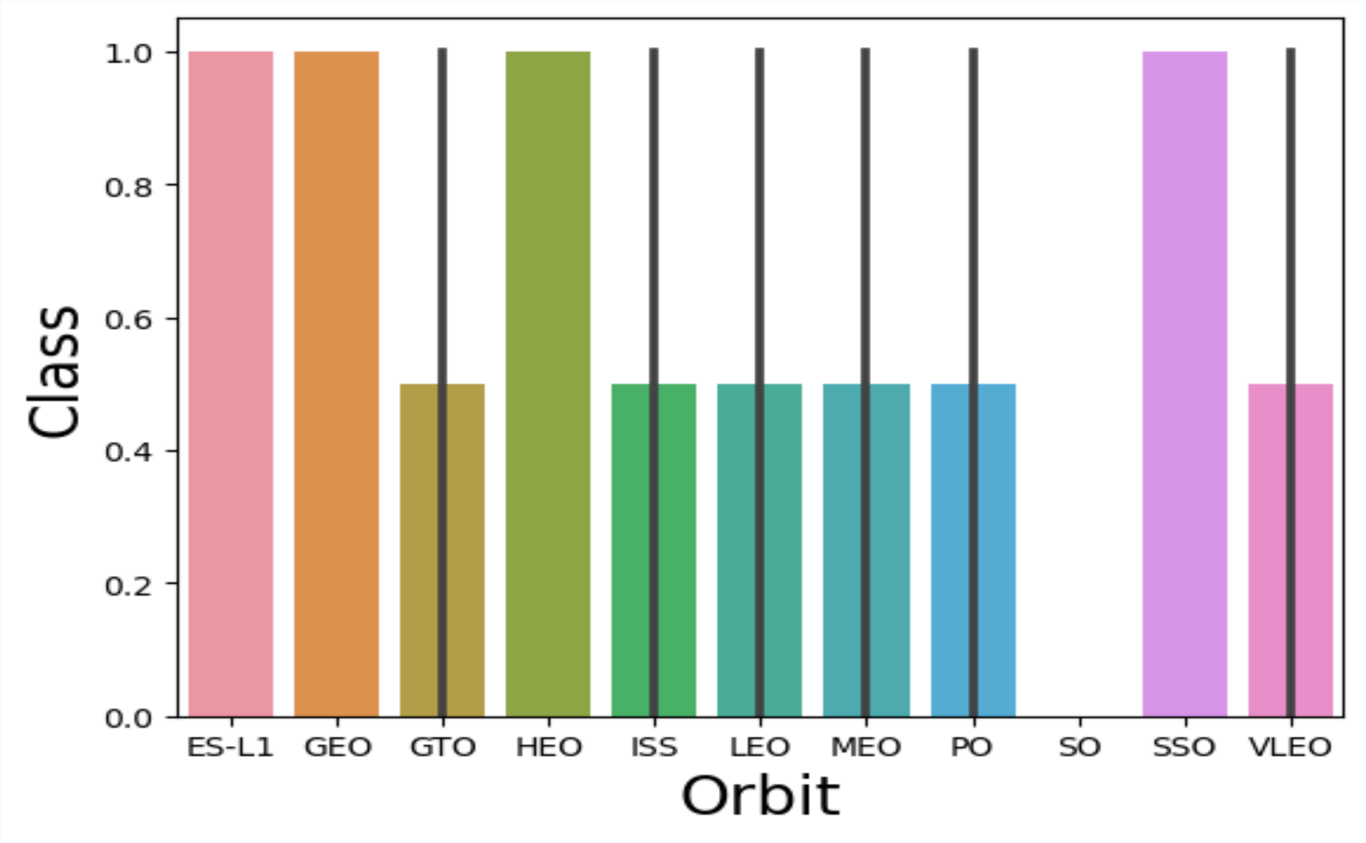
Payload vs. Launch Site



Payload vs. Launch Site

- **Purpose:** This scatter plot visualizes the relationship between payload mass and launch site, highlighting trends in payload capacity and performance across different sites.
- **Axes:**
  - **X-Axis (Launch Site):** Represents the distinct launch sites (e.g., KSC LC-39A, CCAFS SLC-40) used for Falcon 9 missions.
  - **Y-Axis (Payload Mass):** Represents the payload mass (in kilograms) carried by each launch.
- **Key Observations:**
  - Different launch sites handled varying payload capacities, with some sites (e.g., Kennedy Space Center) accommodating heavier payloads more frequently.
  - Clusters of points indicate the typical payload range for specific sites, showing specialization in certain payload sizes.
  - Outliers, such as unusually heavy payloads, were identified, providing insights into unique missions (e.g., interplanetary launches or military payloads).
- **Insights:**
  - Heavier payloads tended to correspond with certain sites, likely due to their superior facilities or proximity to equatorial orbits.
  - Sites with a narrow range of payload masses indicate focused operational use, while sites with a wide range suggest versatility.
- This scatter plot helped analyze how payload capacities influenced launch site selection and performance, offering insights into the operational characteristics of each site.

Success Rate vs. Orbit Type



Success Rate vs. Orbit Type

•**Purpose:** Analyzes the relationship between orbit type and Falcon 9 landing success rates to identify patterns and performance trends.

•**Key Observations:**

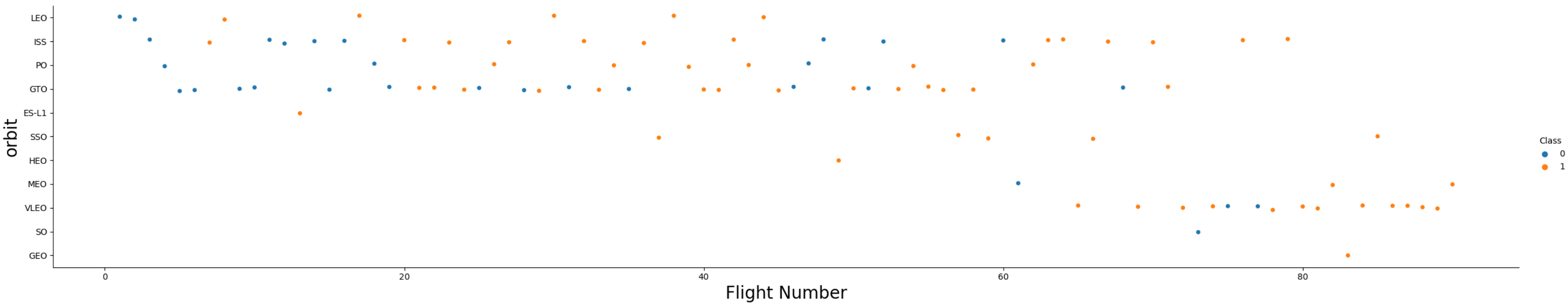
- **SSO (Sun-Synchronous Orbit)** and **GEO (Geostationary Orbit)** achieved 100% success, reflecting optimized procedures and manageable mission conditions.
- **LEO (Low Earth Orbit)** showed moderate success rates, highlighting variability due to diverse payloads and mission profiles.
- Orbits with lower success rates were linked to challenging trajectories, heavier payloads, or higher-energy missions.

•**Insights:**

- Success rates vary by orbit type, with GEO and SSO being most reliable.
- Orbits like LEO require improved strategies for consistent outcomes.
- Understanding orbit-specific challenges can guide better mission planning and resource allocation.

This plot emphasizes the impact of orbit type on landing success, offering actionable insights for optimizing Falcon 9 mission performance.

Flight Number vs. Orbit Type



Flight Number vs. Orbit Type

•**Purpose:** Shows the distribution of orbit types across Falcon 9 launches, using flight numbers as a timeline.
•**Key Observations:**

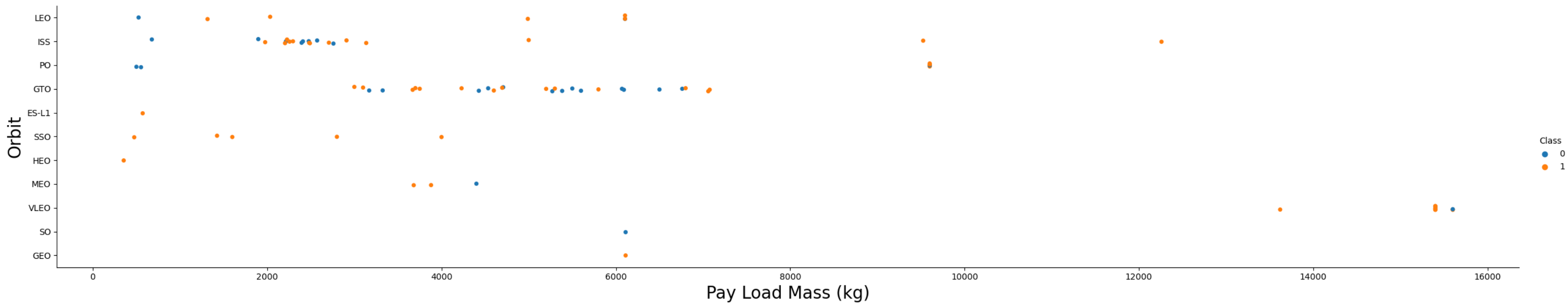
- **LEO (Low Earth Orbit):** Consistently used across all flight numbers, indicating frequent, diverse missions.
- **GEO (Geostationary Orbit) and SSO (Sun-Synchronous Orbit):** Became more common in later launches, reflecting SpaceX’s growing technical capabilities.
- Clusters of specific orbit types suggest periods focused on particular missions, such as satellite deployments or geostationary transfers.

•**Insights:**

- Increased variety in orbit types over time reflects Falcon 9’s growing versatility.
- Later missions targeting complex orbits highlight SpaceX’s technological advancements.
- This distribution reveals trends in SpaceX’s evolving mission objectives and capabilities.

This plot captures how Falcon 9 missions progressed and diversified over time, showcasing SpaceX’s operational growth.

Payload vs. Orbit Type



Payload vs. Orbit Type

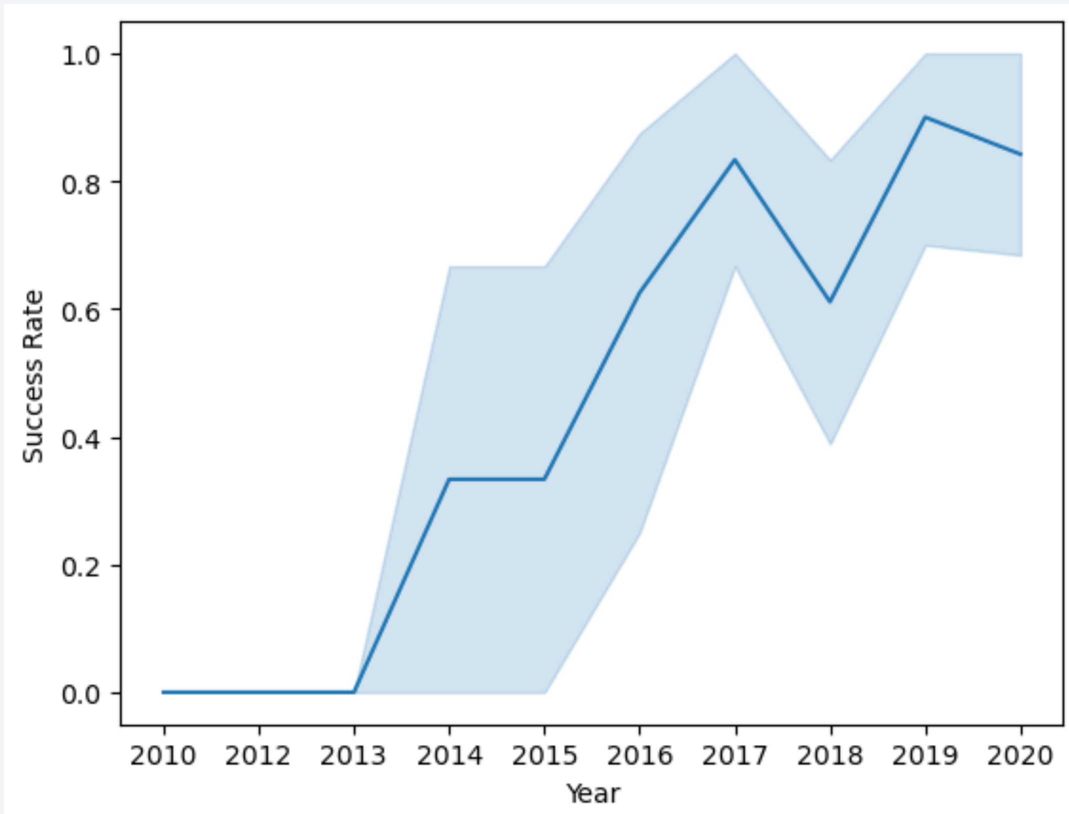
•**Purpose:** Examines the relationship between payload mass and orbit type to identify trends in mission requirements.
•**Axes:**
- **X-Axis (Orbit Type):** Represents different orbit types (e.g., LEO, GEO, SSO).
- **Y-Axis (Payload Mass):** Shows payload mass (in kilograms) delivered to each orbit.

•**Key Observations:**
- **LEO:** Displays the widest range of payloads, reflecting its versatility for diverse missions.
- **GEO and SSO:** Tend to handle moderate to heavy payloads, often for specialized tasks.
- Outliers with extremely high payloads correspond to unique missions, such as interplanetary launches.

•**Insights:**
- Heavier payloads align with high-energy orbits like GEO, requiring careful planning.
- The range of payloads highlights SpaceX's ability to adapt to varied mission needs.
- This plot helps optimize mission design and efficiency for specific orbit types.

It reveals how payload capacities vary by orbit, offering insights into mission complexity and SpaceX's operational flexibility.

Launch Success Yearly Trend



Launch Success Yearly Trend

•**Purpose:** Shows Falcon 9 landing success rates over time to assess reliability improvements.

•**Axes:**

- **X-Axis (Time):** Sequence of launches (e.g., by year or flight number).
- **Y-Axis (Success Rate):** Percentage of successful first-stage landings.

•**Key Observations:**

- A clear upward trend reflects consistent improvements in success rates.
- Early launches had lower success rates, showing initial development challenges.
- Later launches stabilized near 100%, demonstrating operational maturity.

•**Insights:**

- The trend highlights SpaceX’s learning curve and iterative improvements.
- Stabilization at high success rates underscores Falcon 9’s reliability.
- The chart showcases how innovation and refinement lead to operational excellence.

This chart effectively captures SpaceX’s progress in achieving reliable landings over time.

All Launch Site Names

%sql select DISTINCT LAUNCH_SITE from SPACEXTBL;

- This query retrieves all distinct entries in the Launch_Site column from the **SPACEXTBL** table, ensuring each launch site is listed only once.

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

Launch Site Names Begin with 'CCA'

%sql select * from SPACEXTBL WHERE LAUNCH_SITE LIKE 'CCA%' limit 5

- This query retrieves all distinct entries in the Launch_Site column from the **SPACEXTBL** table, ensuring each launch site is listed only once.

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Lar
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Fa
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	Fa
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

- Calculate the total payload carried by boosters from NASA

SUM(PAYLOAD_MASS__KG_)
45596

%sql Select SUM(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE CUSTOMER == 'NASA (CRS)’

This query will calculate the total payload carried by the boosters from nasa. It will filter only NASA boosters then total the payload for each.

Average Payload Mass by F9 v1.1

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

AVG(PAYLOAD_MASS__KG_)
2534.6666666666665

%sql select AVG(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE Booster_Version Like 'F9 v1.1%'

- **AVG(Payload_Mass):** Computes the average payload mass for the filtered records.
- **WHERE Booster_Version = 'F9 v1.1':** Filters the data to include only launches with the booster version F9 v1.1.

First Successful Ground Landing Date

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

min(DATE)
2015-12-22

%sql select min(DATE) from SPACEXTBL WHERE Landing_Outcome == 'Success (ground pad)’;

- **MIN(Date):** Retrieves the earliest date from the filtered results.
- **WHERE Landing_Outcome = 'Success (ground pad)':** Filters for records where the landing outcome was a success on a ground pad.

Successful Drone Ship Landing with Payload between 4000 and 6000

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

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

%sql select Booster_Version, PAYLOAD_MASS__KG_ From SPACEXTBL where Landing_Outcome = 'Success (drone ship)’

- **SELECT DISTINCT Booster_Version**: Retrieves unique booster versions that meet the specified criteria.
- **WHERE Landing_Outcome = 'Success (drone ship)'**: Filters for missions with successful drone ship landings.
- **AND Payload_Mass BETWEEN 4000 AND 6000**: Further filters for payload masses within the specified range.

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

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

%sql select Mission_Outcome, count(*) From SPACEXTBL group by Mission_Outcome

- **COUNT(*):** Counts the total number of rows for each landing outcome.
- **GROUP BY Mission_Outcome:** Groups the data by each unique value in the Mission_Outcome column.

Boosters Carried Maximum Payload

%sql Select Booster_Version, PAYLOAD_MASS__KG_ From SPACEXTBL WHERE PAYLOAD_MASS__KG_ == (Select max(PAYLOAD_MASS__KG_) From SPACEXTBL)

- **SELECT MAX(Payload_Mass):** Identifies the maximum payload mass from the dataset.
- **WHERE Payload_Mass = (SELECT MAX(Payload_Mass)):** Filters records where the payload mass matches the maximum value.
- **Booster_Version:** Retrieves the names of the boosters associated with the maximum payload.

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

Landing_Outcome	Booster_Version	Launch_Site	Month	Year
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40	01	2015
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40	04	2015

%sql select Landing_Outcome, Booster_Version, Launch_Site, substr(Date,6,2) AS Month, substr(Date,0,5) AS Year From SPACEXTBL WHERE Year Like '2015' and Landing_Outcome='Failure (drone ship)'

- **Landing_Outcome = 'Failure (drone ship)':** Filters for missions with failed drone ship landings.
- **YEAR(Date) = 2015:** Filters records to include only those from the year 2015.
- **SELECT Booster_Version, Launch_Site, Landing_Outcome, Date:** Retrieves the relevant details for these missions.

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

%sql SELECT Landing_Outcome, COUNT(*) AS qty FROM SPACEXTBL WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY Landing_Outcome ORDER BY qty DESC; WHERE Date BETWEEN '2010-06-04' AND '2017-03-20': Filters records within the specified date range

- **GROUP BY Landing_Outcome:** Groups records by each unique landing outcome.
- **COUNT(*):** Counts the occurrences of each landing outcome.
- **ORDER BY Outcome_Count DESC:** Sorts the results in descending order based on the count.

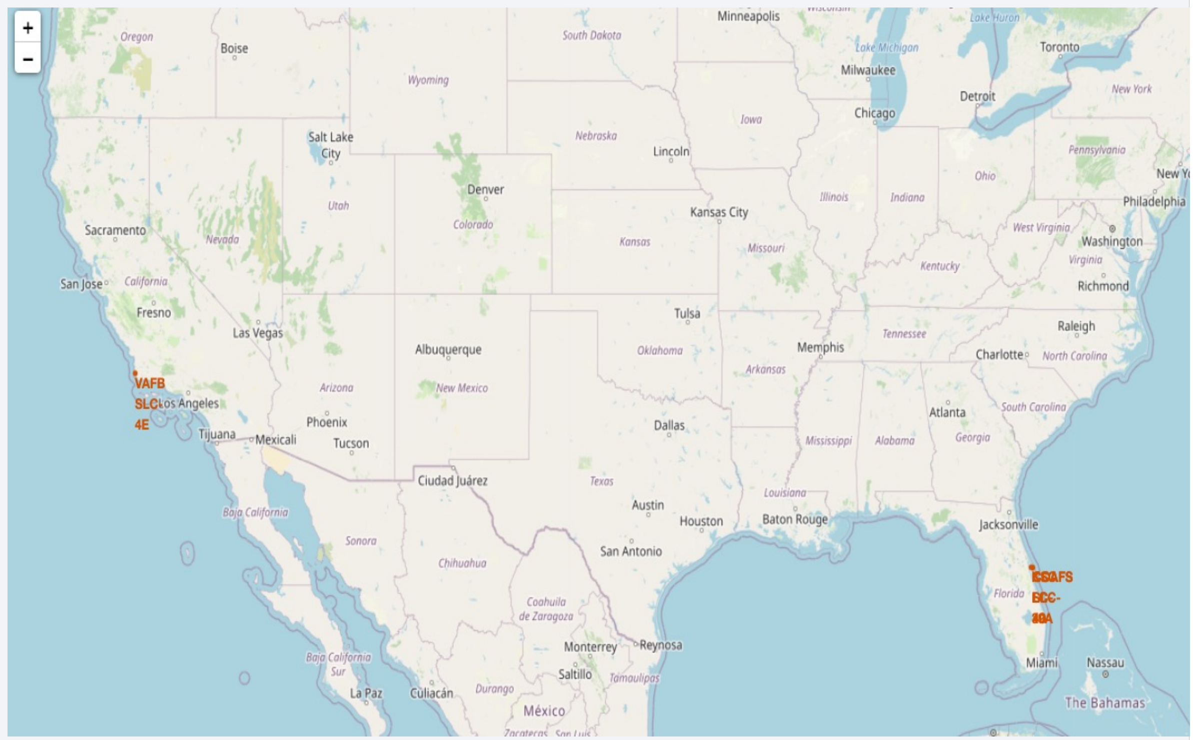
Landing_Outcome	qty
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

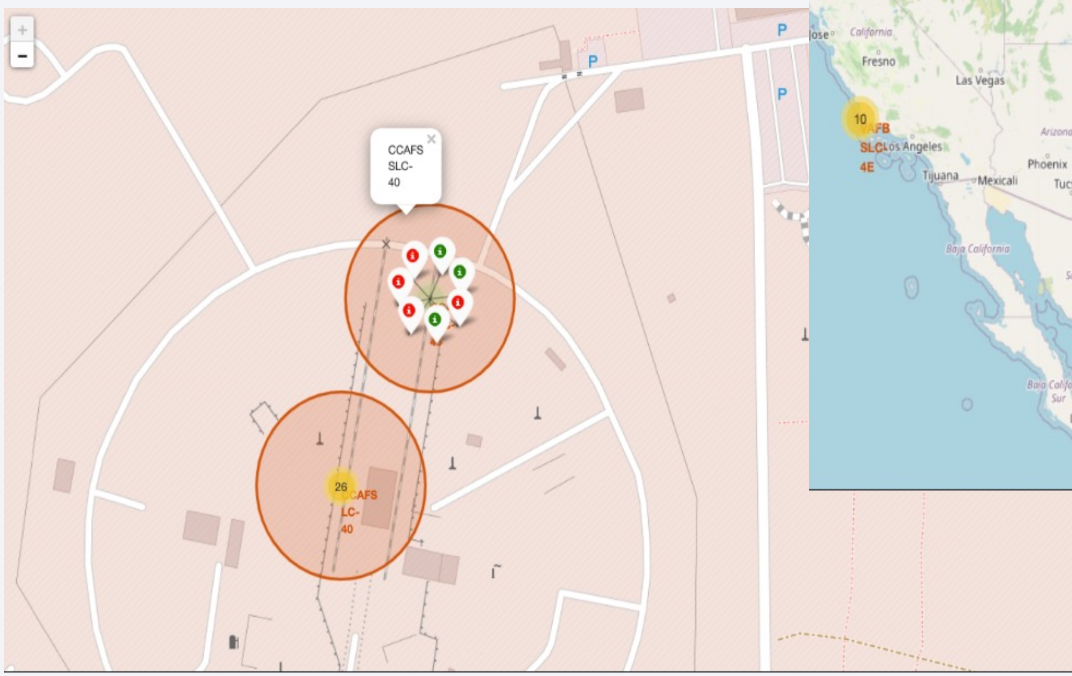
Launch Sites Proximities Analysis

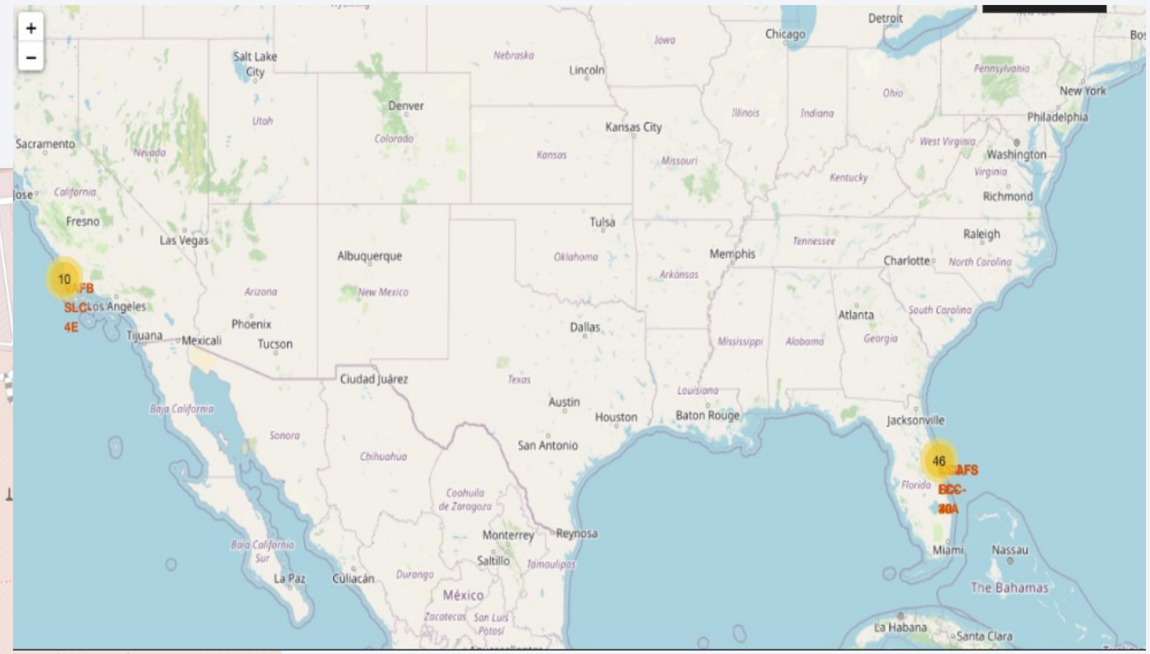
Launch Site Locations

- **Launch Site Locations**: The map highlights major SpaceX launch sites:
- **Cape Canaveral Space Force Station (CCSFS) Launch Complex 40 (LC-40)**: Located in Florida, this site has been pivotal for numerous Falcon 9 missions.
- **Kennedy Space Center (KSC) Launch Complex 39A (LC-39A)**: Also in Florida, LC-39A is notable for crewed missions and significant launches.
- **Vandenberg Space Force Base (VSFB) Space Launch Complex 4E (SLC-4E)**: Situated in California, this site supports polar orbit launches.
- **Starbase (Boca Chica)**: Located in Texas, Starbase is the primary site for Starship development and testing.
- **Proximity to the Equator**: Florida's launch sites are closer to the equator, providing a velocity advantage for eastward launches, which is beneficial for missions requiring higher orbital speeds.
- **Coastal Locations**: All sites are near coastlines, allowing rockets to ascend over open water, minimizing risks to populated areas in case of anomalies.
- **Launch Outcomes**:
- Markers on the map indicate the success or failure of launches from each site. Green markers denote successful missions, while red markers represent failures. This visual differentiation helps identify patterns in launch success rates across different sites.
- **Proximity Analysis**:
- The map allows users to assess the distances between launch sites and nearby infrastructure, such as highways, railways, and populated areas. This analysis is crucial for understanding logistical considerations and potential safety concerns.



Launches: Success Failure

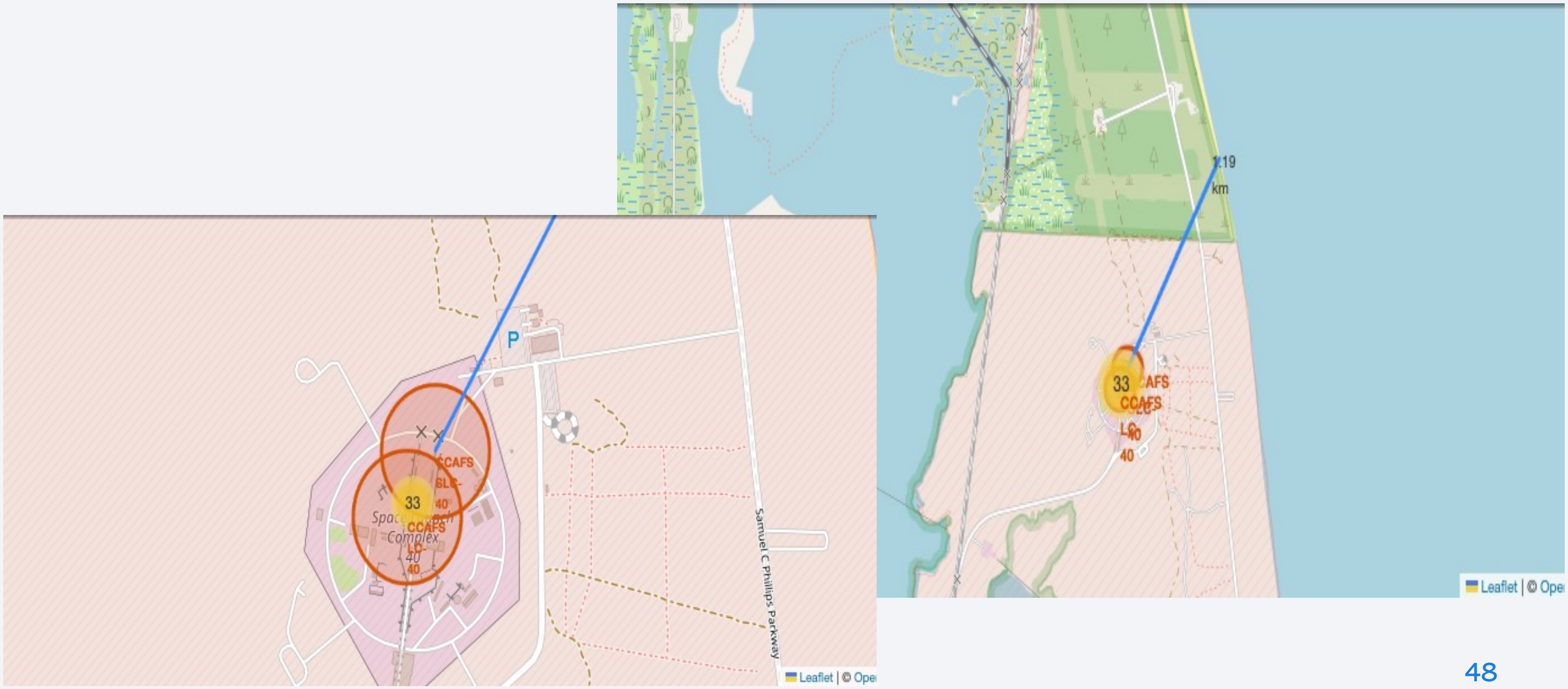




Launches: Success Failure Cont.

- **Outcome Markers**:
  - **Success vs. Failure**: Launch outcomes are represented with distinct markers—typically green for successful launches and red for failures. This clear distinction helps users quickly identify the performance record of each site.
  - **Hover Information**: Users can hover over the markers to view detailed information, such as the mission name, date, payload, and orbit type, providing deeper insights into each launch.
- **Geographical Distribution**:
  - Launch sites with concentrated success (e.g., Kennedy Space Center and Cape Canaveral) demonstrate operational reliability, while sites with mixed results (e.g., testing locations like Starbase) indicate developmental progress.
  - Coastal locations are emphasized, ensuring launches occur over open water, which is safer and allows for easier recovery of failed stages.
- **Success Patterns**:
  - Sites like LC-40 and LC-39A show a high density of successful missions, reflecting their use for operational and high-priority launches.
  - Testing sites, such as Starbase, may show a mix of outcomes due to the experimental nature of the missions conducted there.
- **Interactive Analysis**:
  - Users can zoom in and out to analyze trends across different regions and sites. Clusters of markers can indicate high activity or areas where SpaceX is focusing its efforts.
  - The ability to filter by year, mission type, or outcome adds further depth to the analysis, enabling specific trends to be studied over time.
- **Logistical Insights**:
  - The proximity of launch sites to recovery zones, highways, and supply chains is visually evident, showcasing how SpaceX integrates logistics with operational planning.
  - Sites near SpaceX's production facilities, like Starbase in Texas, highlight the efficiency of minimizing transport distances for experimental hardware.

Launch Site Coastline Proximity



Launch Site Coastline Proximity

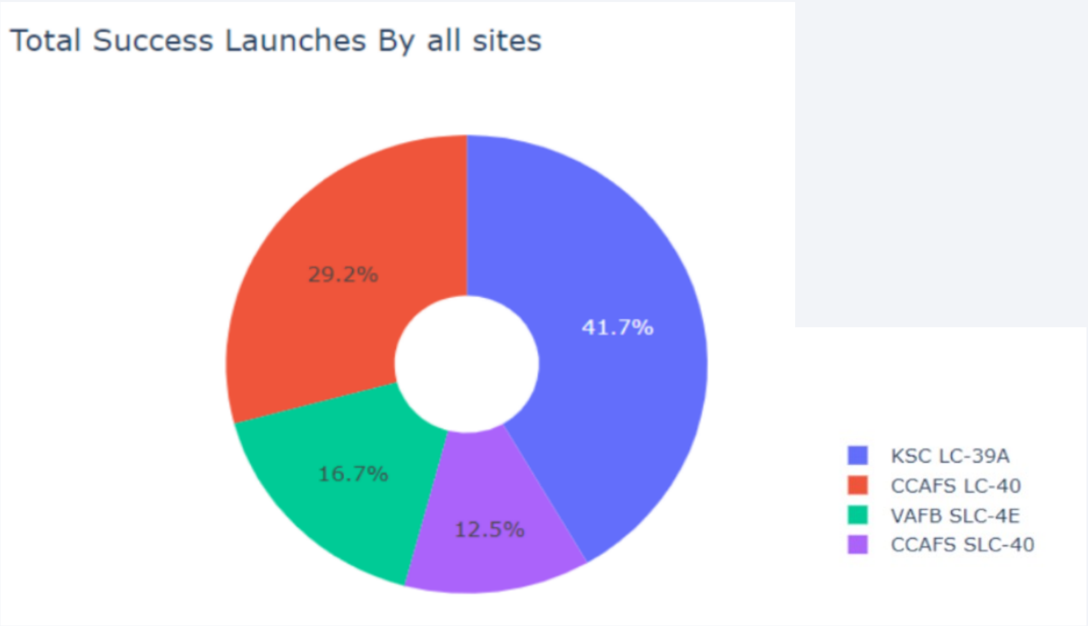
- **Proximity to the Coastline**:
  - **Safety Considerations**: The map highlights how launch sites are strategically located near coastlines, allowing rockets to ascend over open water. This minimizes risks to populated areas in the event of an anomaly during launch.
  - **Environmental Impact**: Coastal proximity ensures that debris from launches or failed missions falls into the ocean, reducing the risk to land-based infrastructure and communities.
- **Geographical Advantages**:
  - **Orbital Trajectories**: Launch sites like Kennedy Space Center and Cape Canaveral benefit from their location along the east coast, enabling efficient eastward launches into geostationary orbits or interplanetary trajectories.
  - **Testing and Recovery**: Coastal locations facilitate stage recovery operations, as SpaceX can deploy recovery ships nearby to retrieve boosters and fairings.
- **Safety Buffers**:
  - The map visually demonstrates the buffer zones around launch sites, showing how their positioning reduces the risk of rocket debris impacting populated areas or infrastructure.
  - Proximity markers may indicate the distances between the launch pad and the coastline, emphasizing strategic placement for safety.
- **Launch Site Specifics**:
  - **Kennedy Space Center (LC-39A)**: Close to the Atlantic Ocean, ideal for a wide range of orbital missions and reusability testing.
  - **Cape Canaveral (LC-40)**: Similarly positioned, offering a comparable safety and operational profile.
  - **Vandenberg Space Force Base (SLC-4E)**: Located along California's coast, it specializes in polar orbit launches over the Pacific Ocean.
  - **Starbase (Boca Chica)**: Positioned near the Gulf of Mexico, this site facilitates testing of experimental Starship systems while minimizing risk to populated areas.
- **Interactive Features**:
  - Markers may show exact distances from the launch sites to the nearest coastline.
  - Users can zoom in to analyze the surrounding environment, including access roads, recovery zones, and nearby water bodies.
- **Insights for Operational Planning**:
  - The proximity to coastlines underscores SpaceX's strategic planning to maximize safety and operational efficiency.
  - The map visually demonstrates how site locations align with SpaceX's goals of safe launches, cost-effective recoveries, and streamlined operations.

Section 4

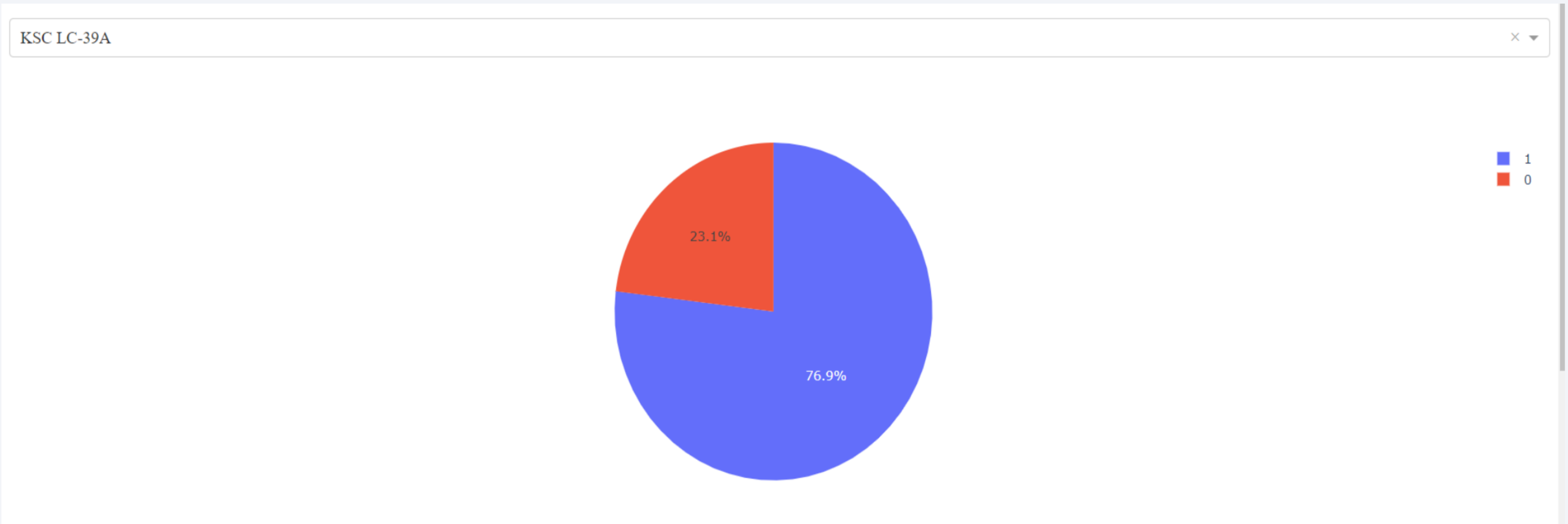
Build a Dashboard with Plotly Dash

Toral Success Launches By all Sites

- Pie Chart shows the percentage of successful launches per site.
- KSC LC-39A represented in blue color has the highest percentage.



Highest Launch Success Ratio



- KSC LC-39A Had the highest success rate at 76.9%

Payload Outcome vs. Launch Outcome

Payload Outcome vs. Launch Outcome

Key Elements:

1.Axes:

1. **X-axis**: Represents the payload mass (kg), showing the range of weights launched by SpaceX, from small payloads to heavy ones.
2. **Y-axis**: Represents the binary launch outcome, with distinct values for success (e.g., 1) and failure (e.g., 0).

2.Data Points:

1. Each point represents a single SpaceX launch.
2. **Marker Color**: Colors are often used to differentiate between success (e.g., green) and failure (e.g., red).
3. **Marker Size**: The size of the marker may vary to represent additional factors like the number of payloads carried or mission type.

3.Interactive Features:

1. Hovering over points provides detailed information, such as mission name, payload type, and orbit.
2. Filters allow users to focus on specific years, orbits, or launch sites, offering customized insights.

Insights and Findings:

1.Payload Weight vs. Success Rate:

1. **Lighter Payloads**: Typically, lighter payloads show a higher success rate, reflecting reduced stress on the rocket and systems.
2. **Heavier Payloads**: Some failures may be clustered at higher payload weights, potentially indicating challenges with launches at maximum capacity or experimental missions.

2.Clusters and Patterns:

1. Distinct clusters of points might reveal operational trends, such as frequent success for payloads within certain weight ranges (e.g., under 10,000 kg).
2. Experimental or developmental launches may appear as outliers with failures or unusually heavy payloads.

3.Mission Complexity:

1. Some heavy payloads coincide with successful outcomes, indicating advancements in SpaceX's capability to handle demanding missions.
2. Patterns may emerge for specific payload types or destinations (e.g., geostationary orbits vs. low Earth orbits).

4.Launch Outcome Distribution:

1. The scatter plot highlights the proportion of successes vs. failures across payload weights, offering insights into SpaceX's reliability for various mission profiles.

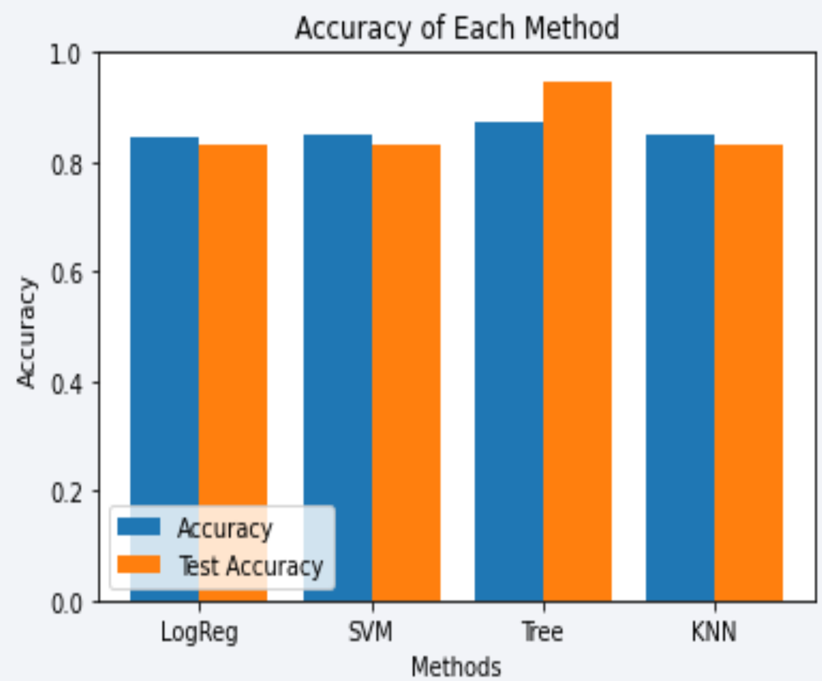
Usefulness

Section 5

Predictive Analysis (Classification)

Classification Accuracy

- Decision Tree had the best accuracy



Confusion Matrix

Decision Tree Confusion Matrix:

•**Axes:**

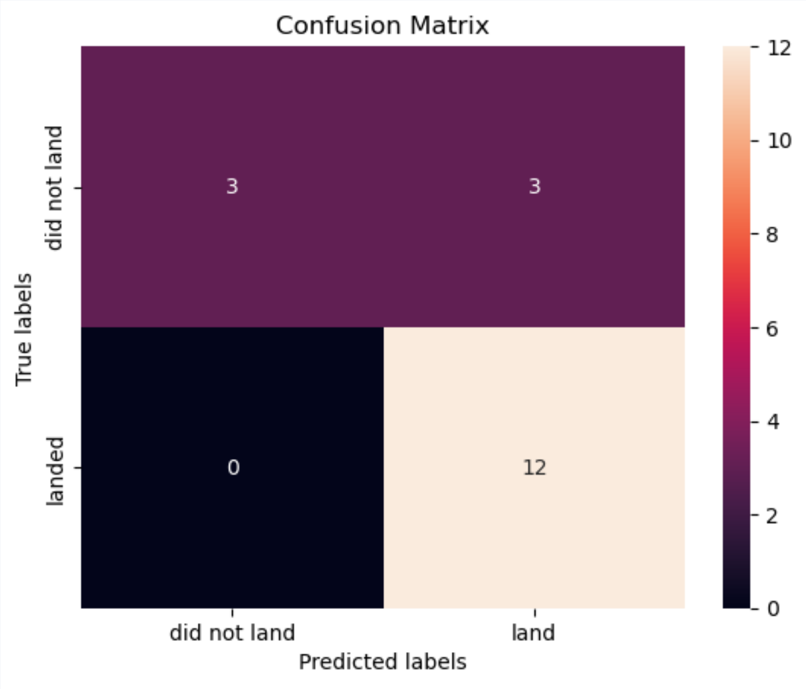
- **Rows (Actual Values):** Represent the true labels for success (1) and failure (0).
- **Columns (Predicted Values):** Represent the predicted labels by the model.

•**Key Metrics:**

- **True Positives (Top-left):** Correctly predicted successful outcomes.
- **True Negatives (Bottom-right):** Correctly predicted failures.
- **False Positives (Top-right):** Predicted success when the actual outcome was failure.
- **False Negatives (Bottom-left):** Predicted failure when the actual outcome was success.

•**Insights:**

- The confusion matrix highlights the model's accuracy in predicting successful and failed landings.
- High counts in the diagonal cells indicate a well-performing model with balanced predictions.



Conclusions

- Successfully analyzed Falcon 9 launch data to identify key factors influencing landing outcomes.
- Explored variables like launch sites, payload mass, orbit types, and booster versions to uncover significant trends.
- Highlighted the growing reliability of Falcon 9 landings and the impact of payload mass on outcomes.
- Identified Decision Tree as the best-performing model, achieving an accuracy of ~83.33%.
- Demonstrated strong predictive capabilities for forecasting landing success based on mission parameters.
- Used interactive visualizations (scatter plots, line charts, geospatial maps) for clear insights.
- Showcased SpaceX's operational progress and the role of data-driven decision-making in reusable rocket technology.
- Findings contribute to improved mission planning and emphasize continuous innovation in space exploration.

Appendix

https://github.com/PyroBBMC/IBM_Applied_Data_Science_Capstone


Thank you!