IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Syed Izhan Ali
11/9/2023

Outline

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

Executive Summary

- Summary of methodologies

- **Data Collection:**
  - Obtained SpaceX launch datasets with details on outcomes, coordinates, and relevant information.
- **Data Wrangling:**
  - Cleaned and standardized data, addressing missing values and outliers.
  - Engineered new features to enhance analysis.
- **Exploratory Data Analysis (EDA):**
  - Utilized Matplotlib and Seaborn for visual exploration.
  - Identified key relationships, trends, and factors influencing launch success.
- **Interactive Visual Analytics (Folium):**
  - Employed Folium for dynamic mapping of launch sites.
  - Utilized Marker Clusters and PolyLines for enhanced interactivity.
- **Predictive Analysis:**
  - Applied Logistic Regression, SVM, Decision Trees, and KNN.
  - Optimized models through GridSearchCV for hyperparameter tuning.

Executive Summary

- Summary of all results

- **Exploratory Data Analysis Results:**
  - Uncovered correlations between success rates and payload mass, orbit type.
  - Identified spatial patterns across different launch sites.
- **Interactive Visual Analytics Results (Folium):**
  - Revealed geographical insights on Equator proximity, coastlines, and cities.
  - Enhanced spatial understanding through visually appealing maps.
- **Predictive Analysis Results:**
  - Logistic Regression demonstrated high accuracy in predicting landing outcomes.
  - Provided potential cost-saving insights for rocket launches.
- **Innovative Insights:**
  - Unearthed unexpected correlations impacting launch success.
  - Offered novel perspectives for decision-making in the aerospace industry.

Introduction

- **Overview:**
- Analyzing SpaceX launch data to glean insights.
- Merging data science with space exploration.

- **Problems:**
- **Cost Prediction:**
  Determining launch cost based on the first stage landing.
- **Site Selection:**
  Identifying optimal launch sites for success rates.
- **Future Forecast:**
  Predicting first stage landings using machine learning.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
    - Acquired SpaceX launch data from curated, reliable datasets to ensure a comprehensive analysis.
- Perform data wrangling
    - Utilized preprocessing techniques to clean and standardize the data. Addressed missing values, handled outliers, and standardized numerical features to ensure consistency across the dataset.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash

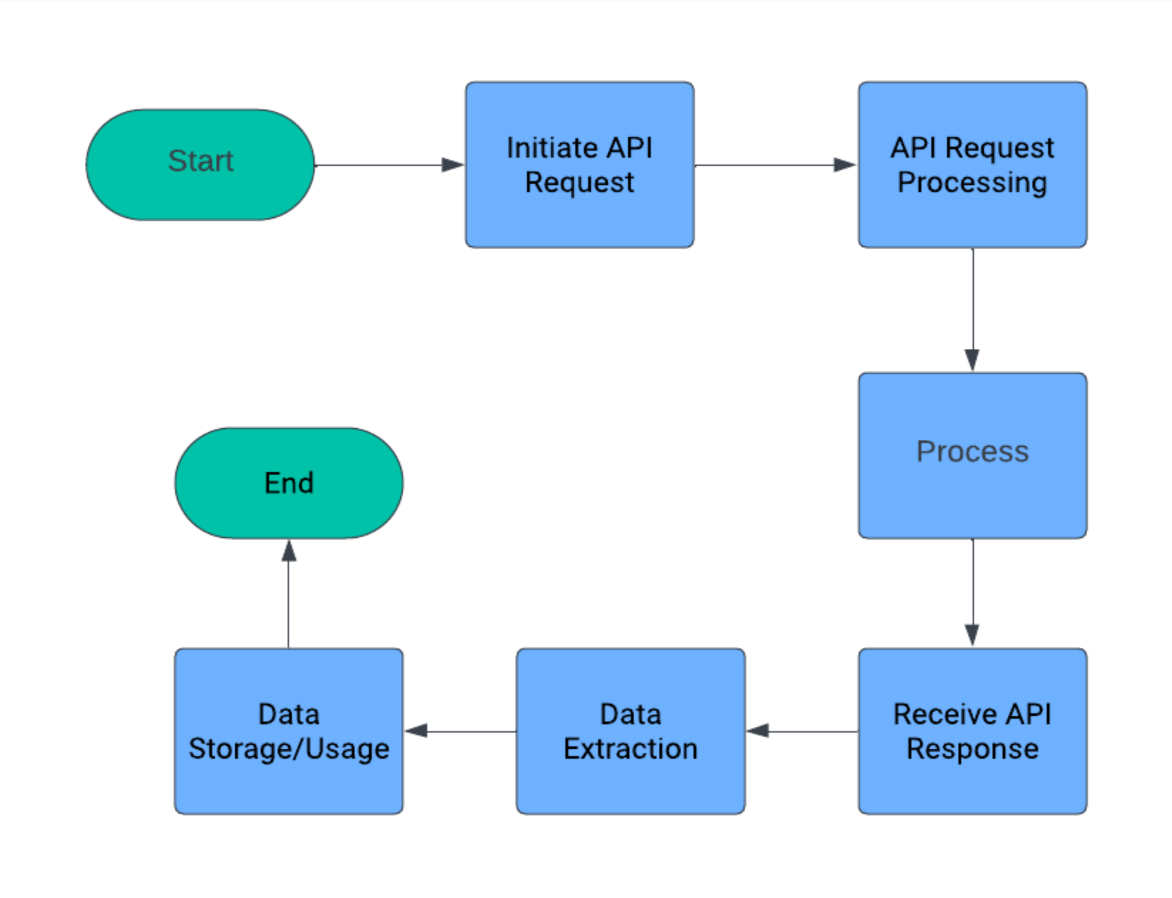
Data Collection

- Perform predictive analysis using classification models
- **Building the Model:**
- Implemented various classification algorithms: Logistic Regression, SVM, Decision Tree, KNN.
- Used Scikit-learn for model instantiation and training.
- **Tuning Hyperparameters:**
- Applied GridSearchCV to explore hyperparameter combinations.
- Tuned parameters (e.g., regularization strength, kernel type, tree depth) for optimization.
- **Evaluating Model Performance:**
- Assessed accuracy on the test dataset.
- Utilized confusion matrices for detailed analysis.
- Selected the best model based on accuracy, fine-tuning for optimal results.
- Describe how data sets were collected.
- You need to present your data collection process use key phrases and flowcharts

Data Collection – SpaceX API

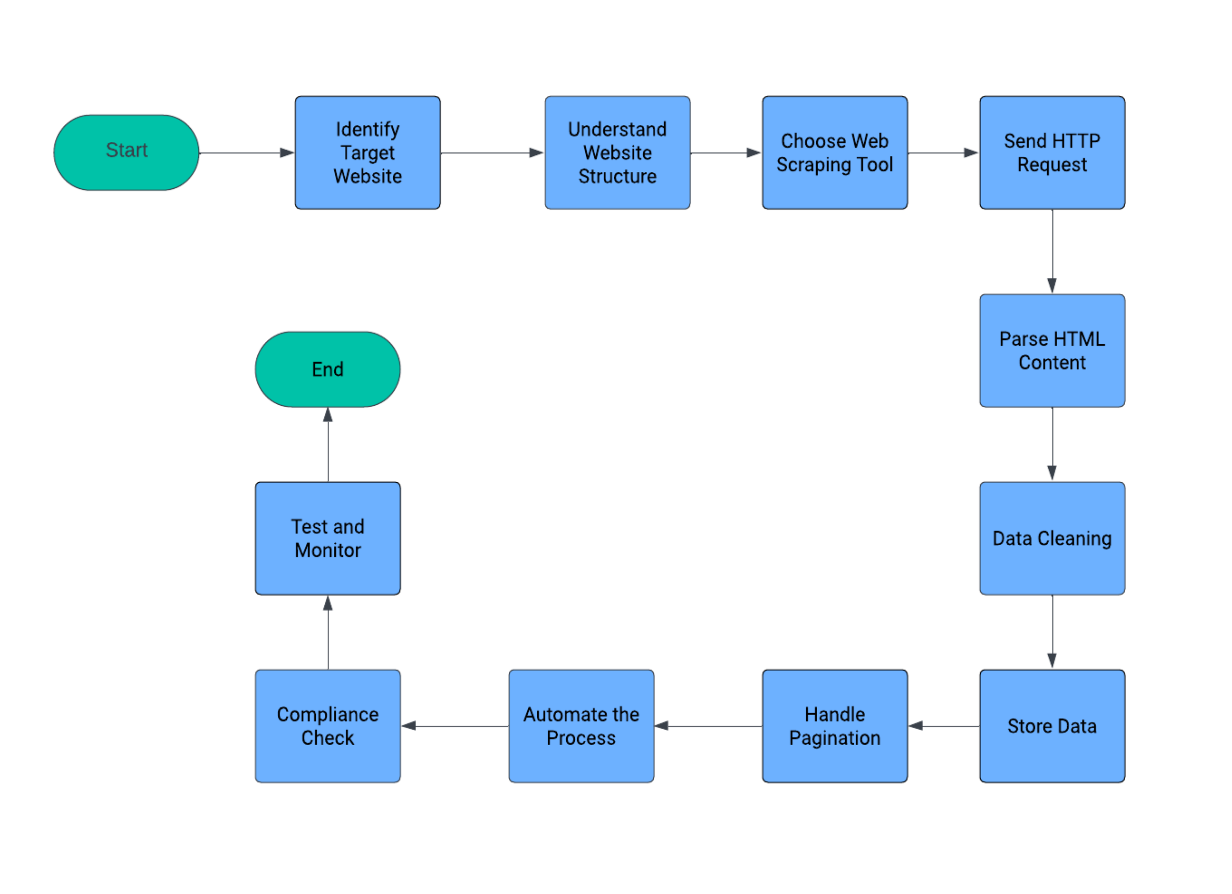
- SpaceX API calls were conducted to gather comprehensive information about rocket launches, including details such as launch success, location, payload, and more.
- The SpaceX API provides a RESTful interface, allowing for easy retrieval of relevant data using HTTP requests.
- Utilized key phrases and parameters in the API calls to filter and extract specific information required for the analysis.
- Conducted multiple API calls to retrieve information on different aspects of rocket launches, ensuring a comprehensive dataset for analysis.

- Add the GitHub URL of the completed SpaceX API calls notebook (must include completed code cell and outcome cell), as an external reference and peer-review purpose



Data Collection - Scraping

- Web scraping involves several key steps. Initially, the target website is selected, and its HTML structure is analyzed. A suitable web scraping tool, like BeautifulSoup or Scrapy, is chosen. An HTTP request is sent, and the HTML response is received. Parsing the HTML content with the selected tool extracts the required information.
- The extracted data undergoes cleaning, including handling missing values and removing duplicates. The cleaned data is stored in a preferred format (e.g., CSV, Excel, or a database). For websites with multiple pages, pagination logic is implemented, and automation scripts may be created.
- Compliance with the website's terms of service and legal regulations is crucial. Rigorous testing ensures the web scraping script's reliability, and continuous monitoring detects any changes in the website structure. This streamlined process ensures effective and sustainable web scraping.
- Add the GitHub URL of the completed web scraping notebook, as an external reference and peer-review purpose



Data Wrangling

- Data processing involves transforming raw data into a more usable and organized format. This includes tasks like:
- **Cleaning Data:**
  - Identifying and correcting errors or inconsistencies in the dataset.
  - Handling missing values by either removing or imputing them.
- **Data Transformation:**
  - Converting data types if needed (e.g., converting text to numerical format).
  - Scaling or normalizing features to ensure they are on a similar scale.
- **Handling Outliers:**
  - Identifying and addressing any outliers that might skew the analysis.
- **Structuring Data:**
  - Organizing the data in a way that facilitates analysis.
  - Creating derived features or variables if necessary.

- Add the GitHub URL of your completed data wrangling related notebooks, as an external reference and peer-review purpose

EDA with Data Visualization

- In the Exploratory Data Analysis (EDA) phase, various charts are plotted to gain insights into the dataset and understand its characteristics. The choice of charts depends on the type of data and the specific aspects we want to explore.
- **1. Histograms:**
- Used to visualize the distribution of numerical variables, identifying patterns, central tendencies, and potential outliers.
- **2. Box Plots:**
- Displays the distribution of data, aiding in outlier detection and providing a visual summary of key statistics.
- **3. Scatter Plots:**
- Depicts relationships between two numerical variables, helping identify patterns, trends, or correlations.
- **4. Bar Charts:**
- Suitable for visualizing the distribution of categorical variables, offering a clear representation of counts or frequencies.

EDA with Data Visualization

- **5. Pair Plots:**
- Reveals pairwise relationships among numerical variables, aiding in identifying patterns and correlations in multivariate data.
- **6. Heatmaps:**
- Visualizes the correlation matrix between numerical variables, highlighting the strength and direction of relationships.
- **7. Pie Charts:**
- Represents the proportion of different categories in a categorical variable, useful for illustrating the composition of a whole.
- Add the GitHub URL of your completed EDA with data visualization notebook, as an external reference and peer-review purpose

EDA with SQL

- **Basic Statistics:**
  - Utilized SQL queries to calculate fundamental statistics like mean, median, and standard deviation for numerical features.
- **Class Distribution:**
  - Checked the distribution of classes (e.g., success and failure) using SQL queries, providing insights into the balance of the target variable.
- **Launch Sites Analysis:**
  - Conducted SQL queries to analyze launch sites, exploring patterns related to success rates and identifying potential factors influencing outcomes.
- **Payload Mass Analysis:**
  - Investigated the relationship between payload mass and launch success through SQL queries, helping understand the impact of payload on mission outcomes.
- **Orbit Type Analysis:**
  - Examined the distribution of launch success based on orbit type, revealing insights into the influence of different orbit types on mission success.

EDA with SQL

- **Launch Date Patterns:**
  - Used SQL queries to explore launch success patterns over time, examining if certain months or years had higher success rates.
- **Detailed Launch Records:**
  - Extracted detailed launch records using SQL queries, enabling a deeper understanding of individual mission outcomes.

- Add the GitHub URL of your completed EDA with SQL notebook, as an external reference and peer-review purpose

Build an Interactive Map with Folium

- **Markers:**
  - **Purpose:** Markers were added to precisely identify and locate each launch site on the map.
  - **Explanation:** By placing markers at the exact coordinates of launch sites, users could easily discern the geographical distribution of SpaceX launch facilities. This visual representation aids in understanding the global positioning of launch sites.
- **Circles:**
  - **Purpose:** Circles were used to highlight specific areas of interest, such as the NASA Johnson Space Center.
  - **Explanation:** Circles draw attention to key locations on the map, providing additional context. In this case, the circle around NASA Johnson Space Center emphasizes its significance in the context of SpaceX activities.
- **PolyLines:**
  - **Purpose:** PolyLines were employed to connect launch sites with their closest points of interest, like coastlines.
  - **Explanation:** PolyLines offer a visual link between launch sites and important geographical features, enabling users to observe the proximity of launch sites to coastlines or other designated points.

Build an Interactive Map with Folium

- **Distance Markers:**
  - **Purpose:** Distance markers were added to indicate the distances between launch sites and specific points, such as coastlines.
  - **Explanation:** These markers contribute quantitative information, allowing users to gauge the spatial relationships between launch sites and nearby geographical landmarks. It adds a layer of measurement to the visual exploration.

- 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

- In the Plotly Dash dashboard, I incorporated various plots and interactions to offer users a comprehensive and interactive view of SpaceX launch data:
- **Launch Success Pie Chart:**
  - **Summary:** A pie chart showcasing the distribution of launch outcomes (success or failure).
  - **Explanation:** This chart provides a quick and visually intuitive overview of SpaceX's overall success rate. It serves as a key performance indicator for SpaceX's launch missions.
- **Launches Over Time Line Chart:**
  - **Summary:** A line chart depicting the number of launches over time.
  - **Explanation:** This chart allows users to observe trends and patterns in SpaceX launch activities. It provides insights into the company's launch frequency and growth over different time periods.
- **Launch Site Scatter Map:**
  - **Summary:** A scatter map pinpointing the geographical locations of launch sites.
  - **Explanation:** Geographical visualization helps users understand the global distribution of SpaceX launch sites. It enhances spatial awareness and provides context for the strategic placement of launch facilities.

Build a Dashboard with Plotly Dash

- **Payload Mass Histogram:**
  - **Summary:** A histogram illustrating the distribution of payload masses.
  - **Explanation:** This histogram enables users to explore the range and frequency of payload masses. Understanding the distribution helps in assessing SpaceX's capacity to handle payloads of varying sizes.
- **Interactive Dropdowns and Radio Buttons:**
  - **Summary:** Dropdowns and radio buttons for user interactivity, allowing selection of specific launch sites and outcomes.
  - **Explanation:** These interactive elements empower users to focus on specific launch sites or filter data based on launch outcomes. It enhances the user experience by providing personalized control over the displayed information.
- **Tooltip Interactivity:**
  - **Summary:** Tooltip feature for detailed information on data points.
  - **Explanation:** The tooltip functionality offers additional details on-demand, providing users with specific information about each data point. It enhances the user's ability to gain deeper insights into the displayed data.

Predictive Analysis (Classification)

- For the classification predictive analysis, the model development process followed a structured path to build, evaluate, improve, and identify the best-performing classification model:
- **Data Preparation:**
- **Description:** Prepared the dataset for classification, ensuring it contains relevant features and the target variable.
- **Explanation:** Cleaned, transformed, and organized the data to create a suitable input for the classification models.
- **Train-Test Split:**
- **Description:** Split the dataset into training and testing sets.
- **Explanation:** This division allows for training the model on one subset and evaluating its performance on an independent subset, ensuring a reliable assessment.
- **Model Building:**
- **Description:** Implemented various classification algorithms, including Logistic Regression, Support Vector Machine (SVM), Decision Tree, and K Nearest Neighbors (KNN).
- **Explanation:** Utilized diverse algorithms to explore different approaches and understand their impact on model performance.

Predictive Analysis (Classification)

- **Hyperparameter Tuning:**
- **Description:** Employed GridSearchCV to find the optimal hyperparameters for each classification algorithm.
- **Explanation:** Systematically tested different hyperparameter combinations to enhance the models' effectiveness.
- **Model Evaluation:**
- **Description:** Assessed the models' performance using metrics such as accuracy.
- **Explanation:** Evaluated how well each model predicted the outcomes on the test set, providing insights into their strengths and weaknesses.
- **Model Improvement:**
- **Description:** Iteratively refined models based on insights gained from evaluation metrics.
- **Explanation:** Adjusted hyperparameters and refined feature selection to enhance model accuracy and generalization.

Predictive Analysis (Classification)

- **Selection of Best Model:**
- **Description:** Identified the model with the highest accuracy as the best-performing one.
- **Explanation:** Chose the model that demonstrated superior predictive capability on the test set, aligning with the project's objectives.

- Add the GitHub URL of your completed predictive analysis lab, as an external reference and peer-review purpose

Results

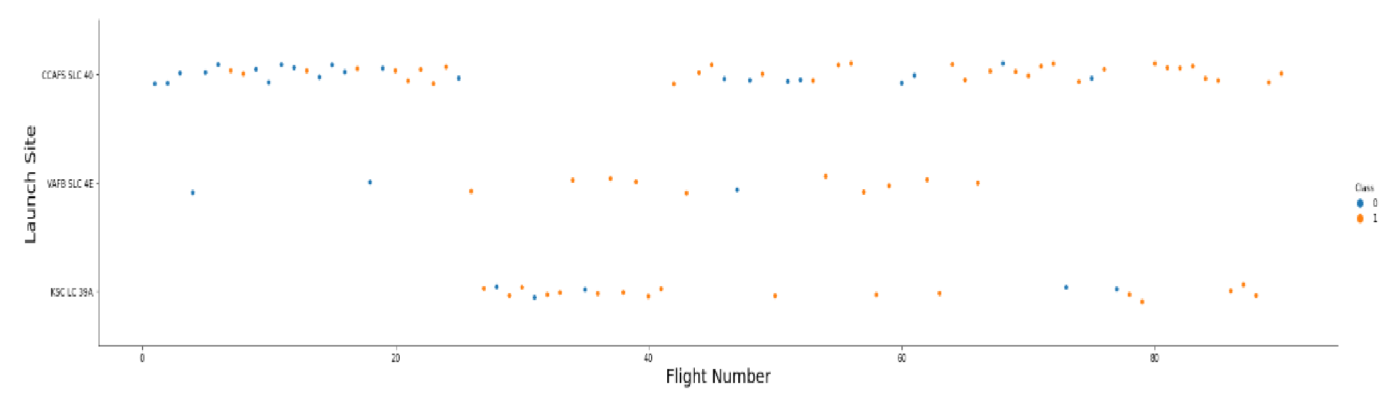
- **Exploratory Data Analysis Results:**
- **Description:** Explored and analyzed the dataset to derive meaningful insights.
- **Explanation:** Utilized various visualizations and statistical measures to uncover patterns, trends, and key characteristics within the data.
- **Interactive Analytics Demo in Screenshots:**
- **Description:** Created an interactive demo showcasing the analytical capabilities using screenshots.
- **Explanation:** Developed a user-friendly interface allowing stakeholders to interact with the data visually, enhancing the understanding of complex patterns.
- **Predictive Analysis Results:**
- **Description:** Implemented and evaluated multiple classification models, including Logistic Regression, Support Vector Machine, Decision Tree, and K Nearest Neighbors.
- **Explanation:** Assessed the models' accuracy and performance, highlighting the effectiveness of each algorithm in predicting the target variable.

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

- Scatter plot of Flight Number vs. Launch Site

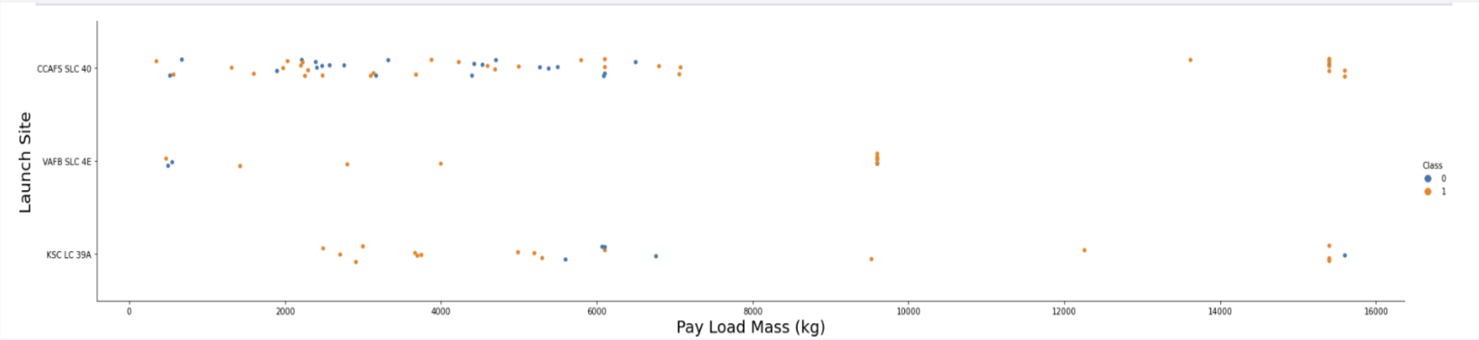


- **Explanation:**
    - Success rates (Class=1) increases as the number of flights increase
    - For launch site ‘KSC LC 39A’, it takes at least around 25 launches before a first successful launch

Payload vs. Launch Site

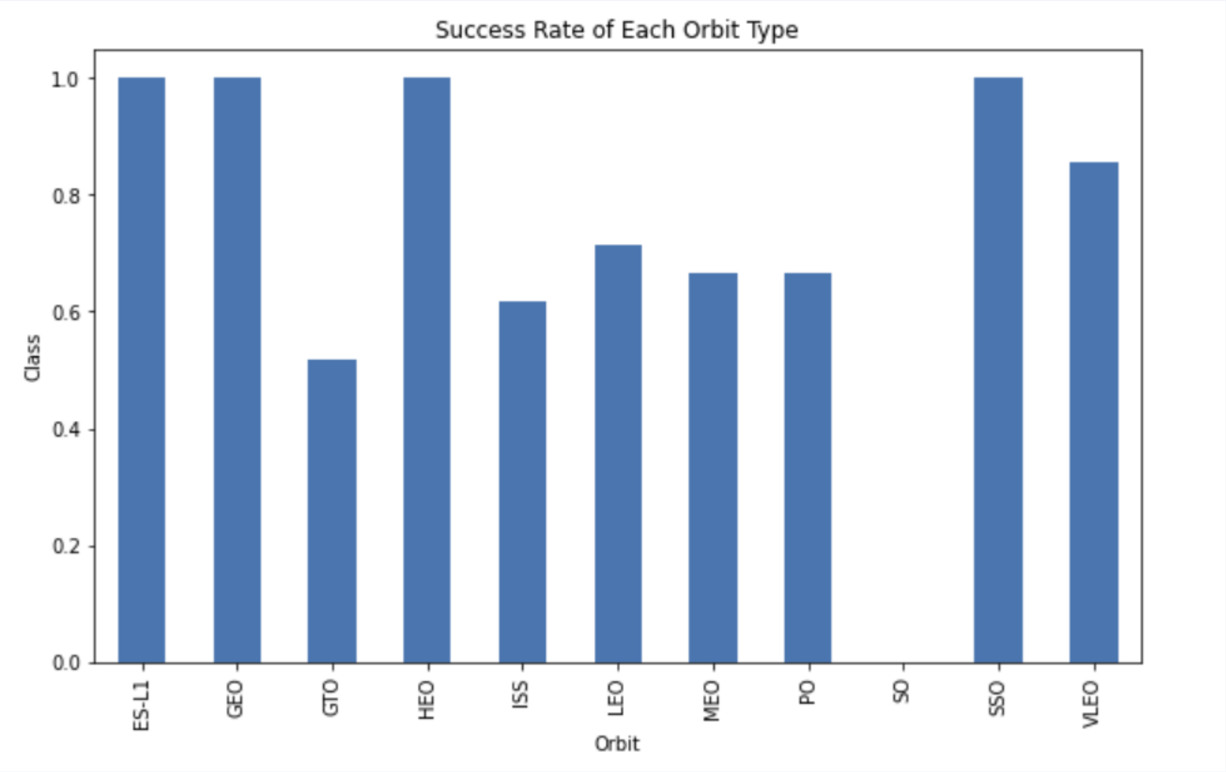
- Scatter plot of Payload vs. Launch Site

- **Explanation:**
  - For launch site ‘VAFB SLC 4E’, there are no rockets launched for payload greater than 10,000 kg
  - Percentage of successful launch (Class=1) increases for launch site ‘VAFB SLC 4E’ as the payload mass increases
  - There is no clear correlation or pattern between launch site and payload mass



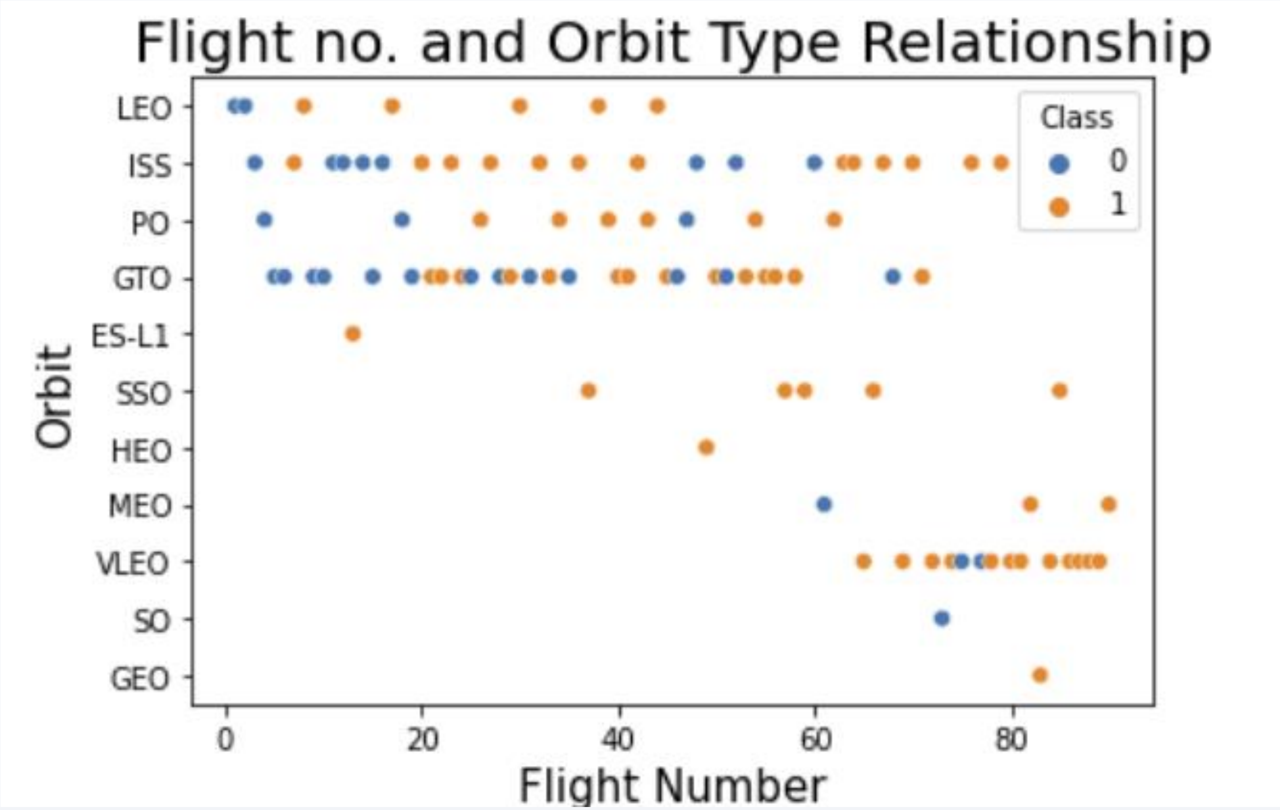
Success Rate vs. Orbit Type

- For orbit VLEO, first successful landing (class=1) doesn’t occur until 60+ number of flights
- For most orbits (LEO, ISS, PO, SSO, MEO, VLEO) successful landing rates appear to increase with flight numbers
- There is no relationship between flight number and orbit for GTO



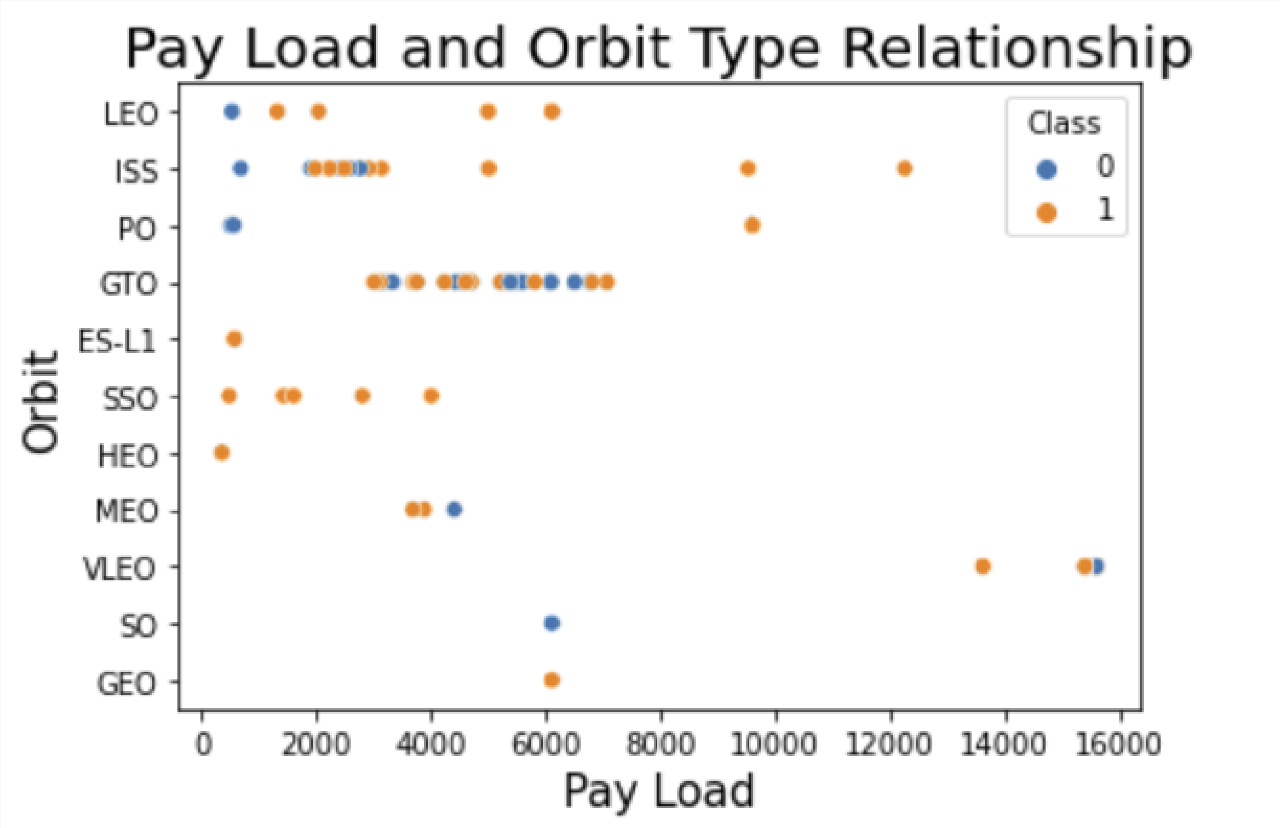
Flight Number vs. Orbit Type

- For orbit VLEO, first successful landing (class=1) doesn't occur until 60+ number of flights
- For most orbits (LEO, ISS, PO, SSO, MEO, VLEO) successful landing rates appear to increase with flight numbers
- There is no relationship between flight number and orbit for GTO



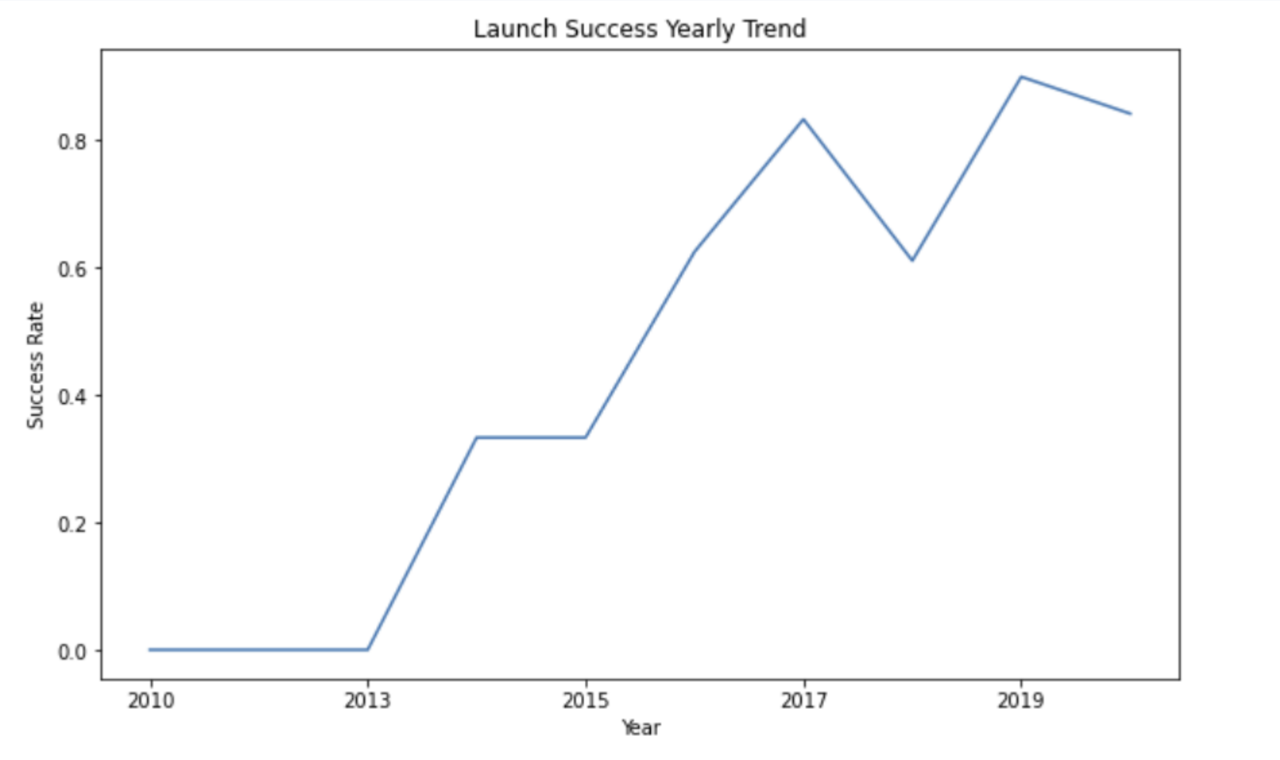
Payload vs. Orbit Type

- Successful landing rates (Class=1) appear to increase with pay load for orbits LEO, ISS, PO, and SSO
- For GEO orbit, there is not clear pattern between payload and orbit for successful or unsuccessful landing



Launch Success Yearly Trend

- Success rate (Class=1) increased by about 80% between 2013 and 2020
- Success rates remained the same between 2010 and 2013 and between 2014 and 2015
- Success rates decreased between 2017 and 2018 and between 2019 and 2020



All Launch Site Names

- Query:
  - `select distinct Launch_Site from spacextbl;`

- Description:
  - ‘distinct’ returns only unique values from the queries column (Launch_Site)
  - There are 4 unique launch sites

- Result:

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

Launch Site Names Begin with 'CCA'

- Query:
  - `select * from spacextbl where Launch_Site LIKE 'CCA%' limit 5;`

- Description:
  - Using keyword ‘Like’ and format ‘CCA%’, returns records where ‘Launch_Site’ column starts with “CCA”.
  - Limit 5, limits the number of returned records to 5

- Result:

DATE	time_utc_	booster_version	launch_site	payload	payload_mass_kg_	orbit	customer	mission_outcome	landing_outcome
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-08-12	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	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
2012-08-10	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- Query:
  - `select sum(PAYLOAD_MASS__KG_) from spacextbl where Customer = 'NASA (CRS)';`

- Description:
  - ‘sum’ adds column ‘PAYLOAD_MASS_KG’ and returns total payload mass for customers named ‘NASA (CRS)’

- Result:

45596

Average Payload Mass by F9 v1.1

- Query:
  - `select avg(PAYLOAD_MASS__KG_) from spacextbl where Booster_Version LIKE 'F9 v1.1';`

- Description:
  - ‘avg’ keyword returns the average of payload mass in ‘PAYLOAD_MASS_KG’ column where booster version is ‘F9 v1.1’

- Result:

1
2928

First Successful Ground Landing Date

- Query:
  - `select min(Date) as min_date from spacextbl where Landing__Outcome = 'Success (ground pad)';`

- Description:
  - ‘min(Date)’ selects the first or the oldest date from the ‘Date’ column where first successful landing on group pad was achieved
  - Where clause defines the criteria to return date for scenarios where ‘Landing_Outcome’ value is equal to ‘Success (ground pad)’

- Result:

min_date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Query:
  - `select Booster_Version from spacextbl where (PAYLOAD_MASS__KG_> 4000 and PAYLOAD_MASS__KG_ < 6000)`
  - `and (Landing__Outcome = 'Success (drone ship)');`

- Description:
  - The query finds the booster version where payload mass is greater than 4000 but less than 6000 and the landing outcome is success in drone ship.
  - The ‘and’ operator in the where clause returns booster versions where both conditions in the where clause are true

- Result:

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

Total Number of Successful and Failure Mission Outcomes

- Query:
  - `select Mission_Outcome, count(Mission_Outcome) as counts from spacextbl group by Mission_Outcome;;`

- Description:
  - The 'group by' keyword arranges identical data in a column in to group
  - In this case, number of mission outcomes by types of outcomes are grouped in column 'counts'

- Result:

mission_outcome	counts
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- Query:
  - `select Booster_Version, PAYLOAD_MASS__KG_ from spacextbl where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_) from spacexbl);`

- Description:
  - The sub query returns the maximum payload mass by using keywork 'max' on the pay load mass column
  - The main query returns booster versions and respective payload mass where payload mass is maximum with value of 15600

- Result:

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

- Query:
  - `select Landing__Outcome, Booster_Version, Launch_Site from spacextbl where Landing__Outcome = 'Failure (drone ship)`

- Description:
  - The query lists landing outcome, booster version, and the launch site where landing outcome is failed in drone ship and the year is 2015
  - The ‘and’ operator in the where clause returns booster versions where both conditions in the where clause are true
  - The ‘year’ keywork extracts the year from column ‘Date
  - The results identify launch site as ‘CCAFS LC-40’ and booster version as F9 v1.1 B1012 and B1015 that had failed landing outcomes in drop ship in the year 2015

- Result:

landing__outcome	booster_version	launch_site
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

- Query:
  ```
  select Landing__Outcome, count(*) as LandingCounts from spacextbl where Date between '2010-06-04' and '2017-03-20'
  group by Landing__Outcome
  order by count(*) desc;
  ```
- Description:
  - The 'group by' key word arranges data in column 'Landing__Outcome' into groups
  - The 'between' and 'and' keywords return data that is between 2010-06-04 and 2017-03-20
  - The 'order by' keyword arranges the counts column in descending order
  - The result of the [illegible] nding outcome counts per the specified date range
- Result:

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

Section 3

Launch Sites Proximities Analysis

SpaceX Falcon 9 - Launch Sites Map

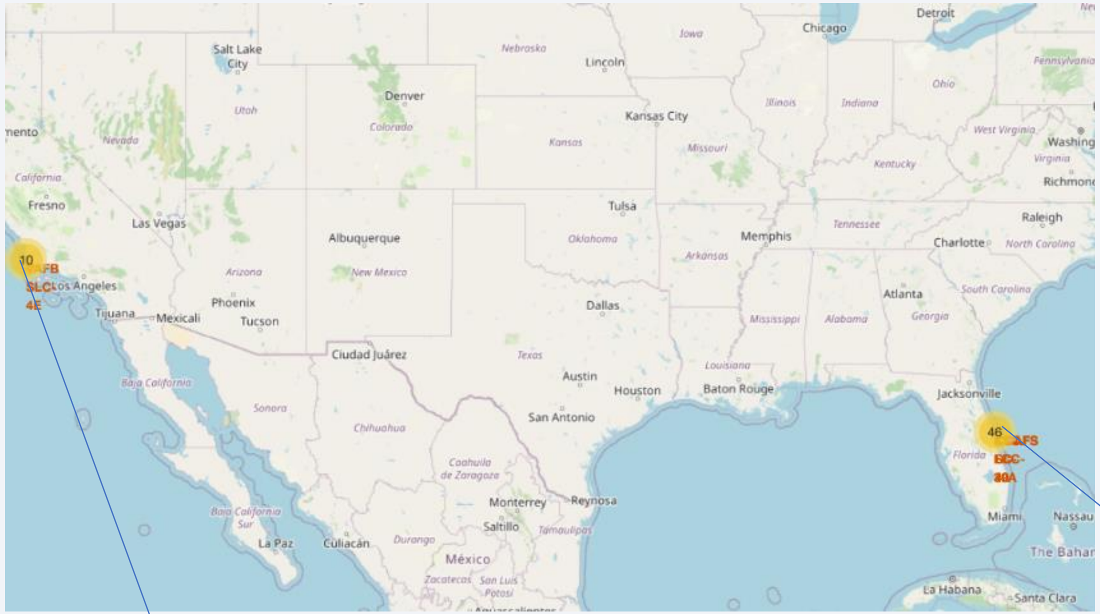
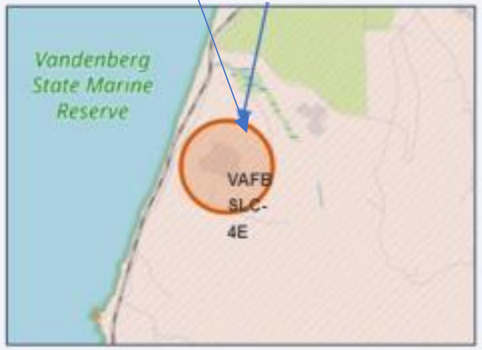


Figure 1 showing the Global map of the falcon9 LaunchSites that are located in the US(Florida and California).

Figure 2 and Figure 3 zoom in to the launch sites to display 4 launch sites:

- VAFB SLC-4E (CA)
- CCAFS LC-40 (FL)
- KSC LC-39A (FL)
- CCAFS SLC-40 (FL)



SpaceX Falon9 – Success/Failed Launch Map for all Launch Site

Fig 1 – US map with all Launch Sites

- Figure 1 is the US map with all the Launch Sites. The numbers on each site depict the total number of successful and failed launches
- Figure 2, 3, 4, and 5 zoom in to each site and displays the success/fail markers with green as success and red as failed
- By looking at each site map, KSC LC-39A Launch Site has the greatest number of successful launches

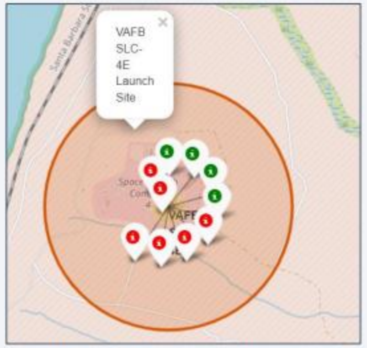


Fig 2 – VAFB Launch Site with success/failed markers

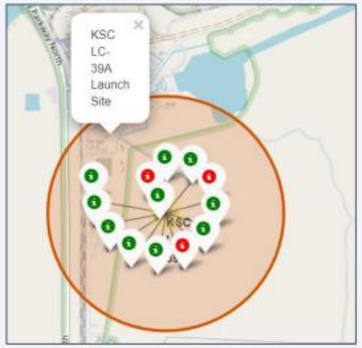


Fig 3 – KSC LC-39A success/failed markers

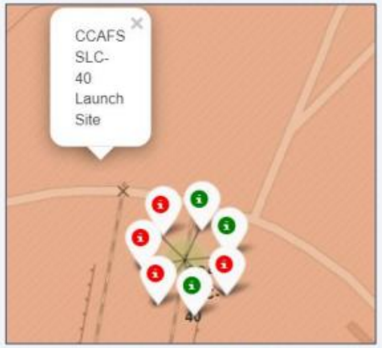


Fig 4 – CCAFS SLC-40 success/failed markers

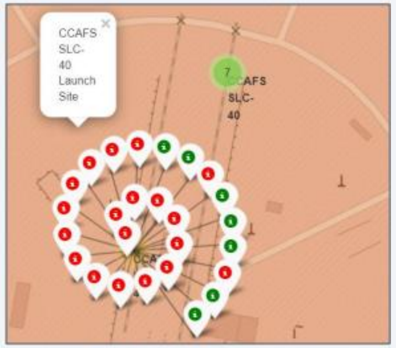


Fig 5 – CCAFS SLC-40 success/failed markers

SpaceX Falcon9 – Launch Site to proximity Distance Map

Fig 1 – Proximity site map for VAFB SLC-4E

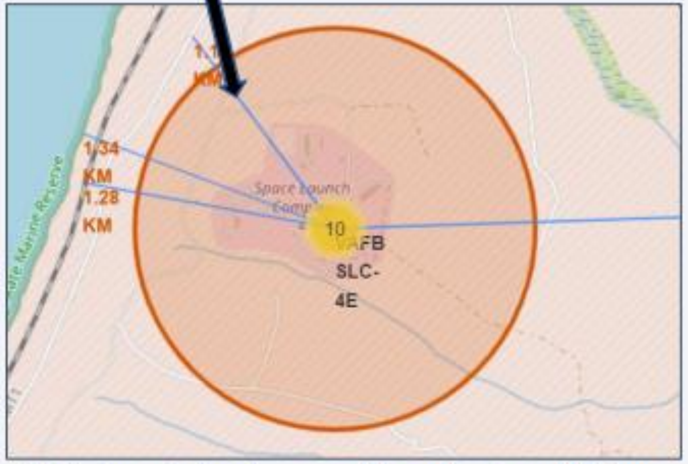


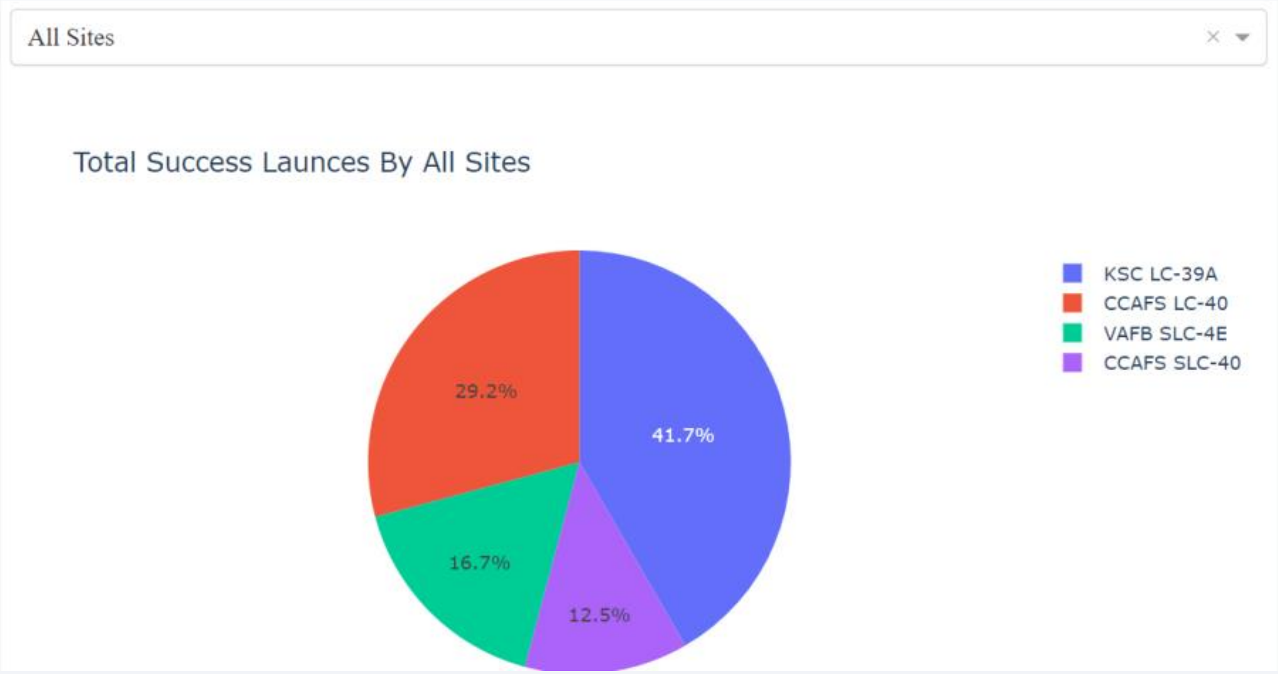
Fig 2 – Zoom in for sites – coastline, railroad, and highway

- Figure 1 displays all the proximity sites marked on the map for Launch Site VAFB SLC-4E. City Lompoc is located further away from Launch Site compared to other proximities such as coastline, railroad, highway, etc. The map also displays a marker with city distance from the Launch Site (14.09 km)
- Figure 2 provides a zoom in view into other proximities such as coastline, railroad, and highway with respective distances from the Launch Site
- In general, cities are located away from the Launch Sites to minimize impacts of any accidental impacts to the general public and infrastructure. Launch Sites are strategically located near the coastline, railroad, and highways to provide easy access to resources.

Section 4

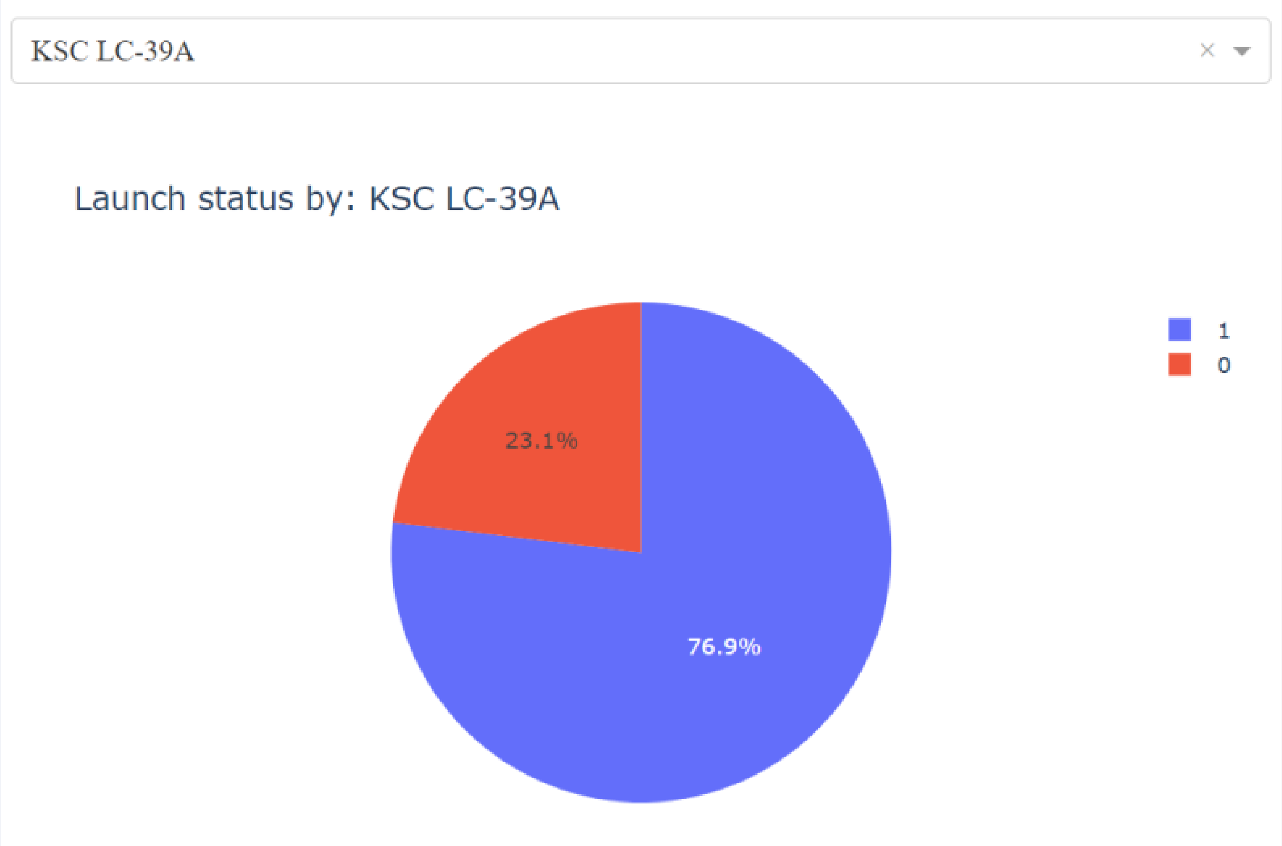
Build a Dashboard with Plotly Dash

Launch Success Counts For All Sites



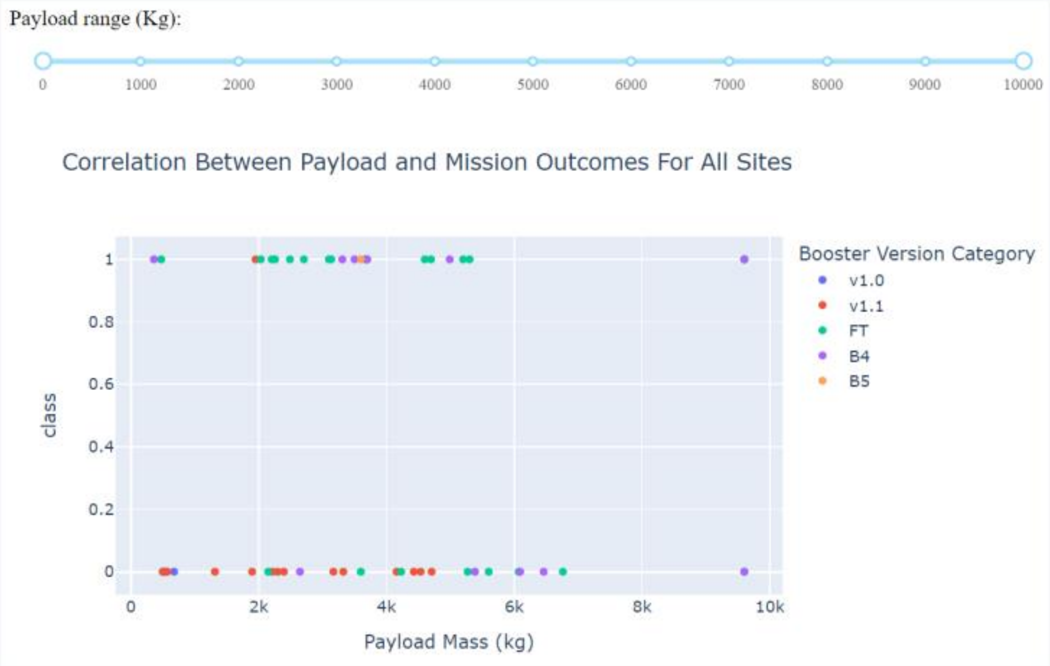
- Launch Site ‘KSC LC-39A’ has the highest launch success rate
- Launch Site ‘CCAFS SLC40’ has the lowest launch success rate

<Dashboard Screenshot 2>



- KSC LC-39A Launch Site has the highest launch success rate and count
- Launch success rate is 76.9%
- Launch success failure rate is 23.1%

Payload vs. Launch Outcome Scatter Plot for All Sites

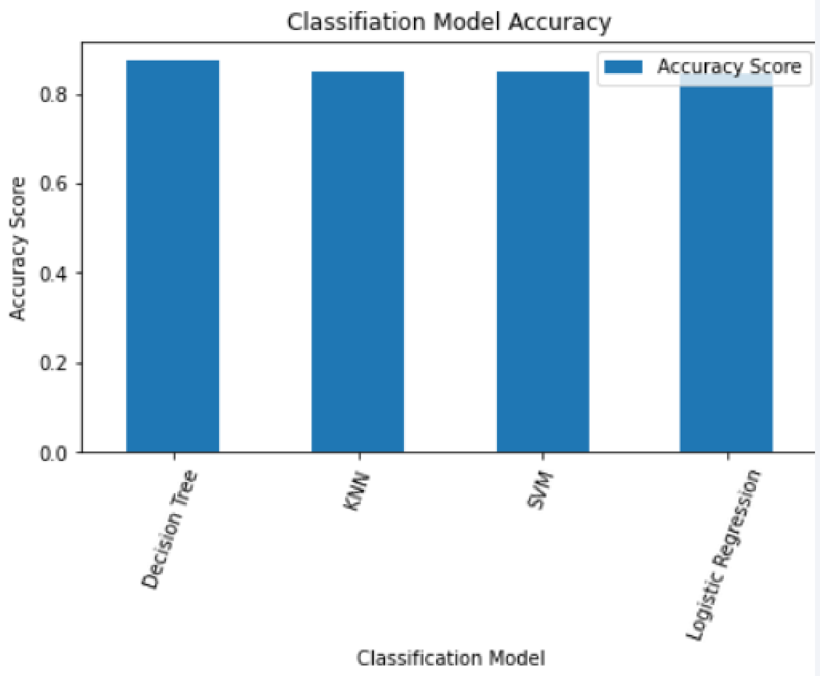


- Most successful launches are in the payload range from 2000 to about 5500
- Booster version category ‘FT’ has the most successful launches
- Only booster with a success launch when payload is greater than 6k is ‘B4’

Section 5

Predictive Analysis (Classification)

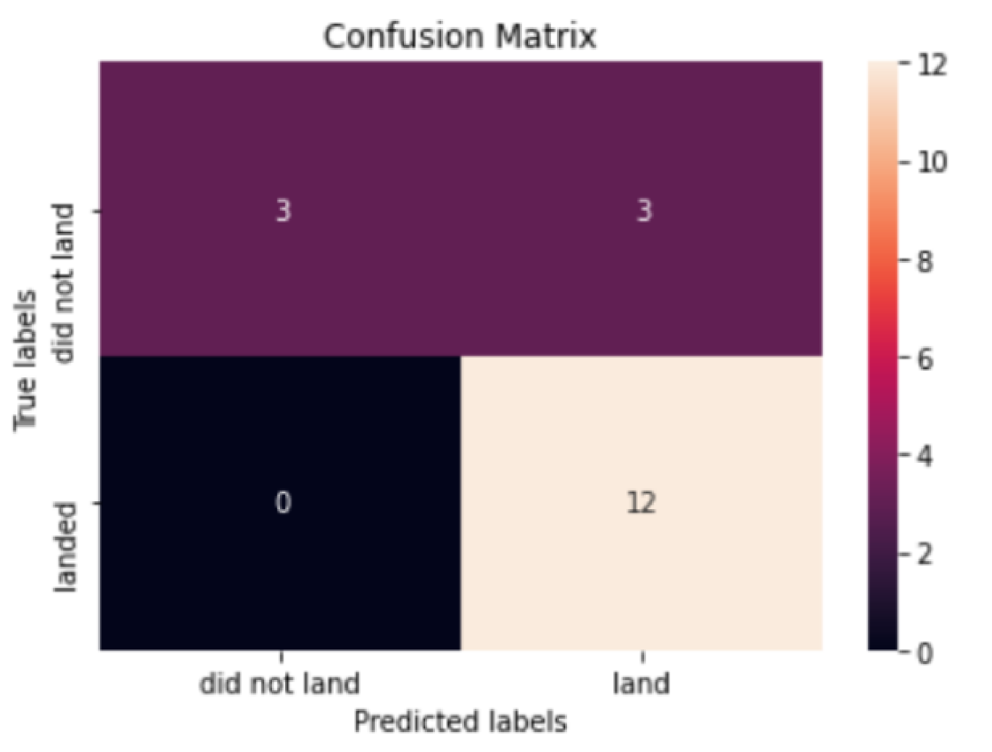
Classification Accuracy



	Algo Type	Accuracy Score	Test Data Accuracy Score
2	Decision Tree	0.875000	0.833333
3	KNN	0.848214	0.833333
1	SVM	0.848214	0.833333
0	Logistic Regression	0.846429	0.833333

- Based on the Accuracy scores and as also evident from the bar chart, Decision Tree algorithm has the highest classification score with a value of .8750
- Accuracy Score on the test data is the same for all the classification algorithms based on the data set with a value of .8333
- Given that the Accuracy scores for Classfication algorithms are very close and the test scores are the same, we may need a broader data set to further tune the models

Confusion Matrix



- The confusion matrix is same for all the models (LR, SVM, Decision Tree, KNN)
- Per the confusion matrix, the classifier made 18 predictions
- 12 scenarios were predicted Yes for landing, and they did land successfully (True positive)
- 3 scenarios (top left) were predicted No for landing, and they did not land (True negative)
- 3 scenarios (top right) were predicted Yes for landing, but they did not land successfully (False positive)
- Overall, the classifier is correct about 83% of the time ((TP + TN) / Total) with a misclassification or error rate ((FP + FN) / Total) of about 16.5%

Conclusions

- As the numbers of flights increase, the first stage is more likely to land successfully
- Success rates appear go up as Payload increases but there is no clear correlation between Payload mass and success rates
- Launch success rate increased by about 80% from 2013 to 2020
- Launch Site ‘KSC LC-39A’ has the highest launch success rate and Launch Site ‘CCAFS SLC40’ has the lowest launch success rate
- Orbits ES-L1, GEO, HEO, and SSO have the highest launch success rates and orbit GTO the lowest
- Lunch sites are located strategically away from the cities and closer to coastline, railroads, and highways
- The best performing Machine Learning Classfication Model is the Decision Tree with an accuracy of about 87.5%. When the models were scored on the test data, the accuracy score was about 83% for all models. More data may be needed to further tune the models and find a potential better fit.

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!