IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Lorby Bi>
<Jul, 2024>

Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

- Project background and context
  - As commercial space industry is rapidly evolving, SpaceX has gained significant attention for its reusable rockets and cost-effective space missions, which include sending payloads to the International Space Station and deploying the Starlink satellite constellation.
- Problems that needs to be answered
  - Estimating the cost of each rocket launch.
  - Predicting the likelihood of successful recovery and reuse of the first stage of SpaceX’s Falcon 9 rocket.

Section 1

Methodology

Methodology

Executive Summary

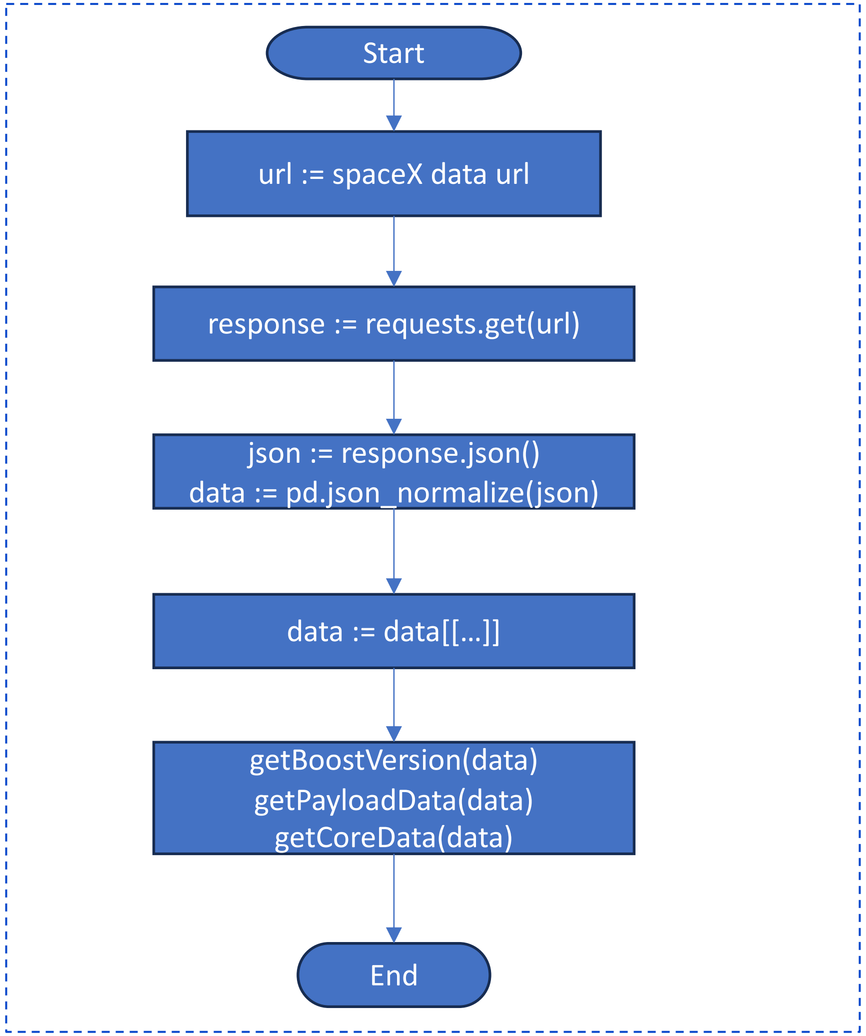
- Data collection methodology:
- Perform data wrangling:
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

Data Collection

- We collected data using the SpaceX REST API, targeting the endpoint for past launches to obtain detailed information about each launch.
- Additional data was gathered by web scraping relevant Wikipedia pages using the BeautifulSoup package.

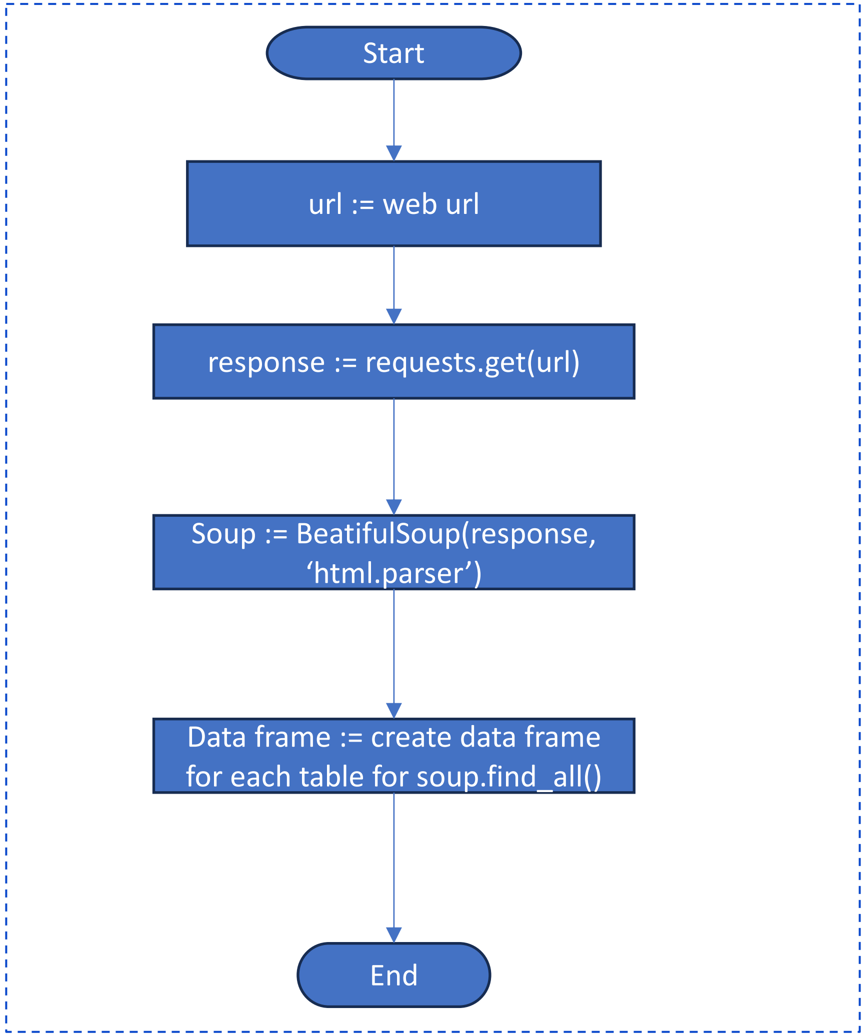
Data Collection – SpaceX API

- To collect the data with SpaceX API, we performed with the following steps:
  1. Determine the URL of data IDs
  2. Using HTTP GET to get the data IDs from the determined URL with requests.get()
  3. Decode the response content which is in JSON format with .json()
  4. Convert the decoded json file to pandas dataframe
  5. Extract needed fields (IDs of the real data) with relevant pandas APIs
  6. Get the real data from the data ID
     1) getBoosterVersion:
        i. For each ID, using HTTP GET with REST API for getting rocket
     2) getPayloadData
        i. For each ID, using HTTP GET with REST API for getting payload data
     3) getCoreData
        i. For each ID, using HTTP GET with REST API for getting each core data respectively
- Reference: https://github.com/Lorby04/IBMCapstone/blob/master/module_1_L2/jupyter-labs-spacex-data-collection-api.ipynb



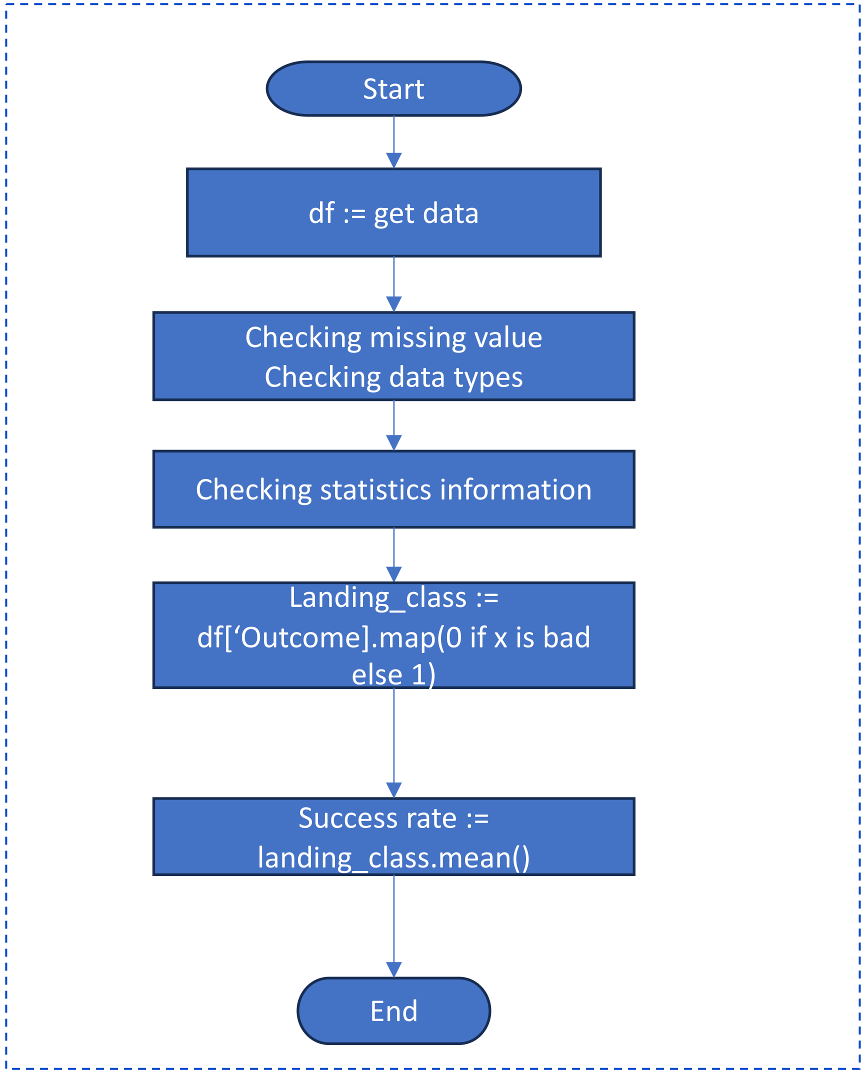
Data Collection - Scraping

- To collect the data with SpaceX API, we performed with the following steps:
    1. Determine the URL of the web page
    2. Using HTTP GET to get the page with requests.get()
    3. Buidling BeatfulSoup instance from the response with html_parser
    4. Creating a dataframe from the soup object with soup find APIs
- Reference: https://github.com/Lorby04/IBMCapstone/blob/master/module_1_L2/jupyter-labs-webscraping.ipynb



Data Wrangling

- To collect the data with SpaceX API, we performed with the following steps:
  1. Get data from data source
  2. Identifying missing value and data type of each column
  3. Checking statistics information, such as count of launch site, mission outcome per orbits
  4. Creating a landing_class column from outcome column
  5. Calculating the mean of landing_class column to get the success rate
- Reference: https://github.com/Lorby04/IBMCapstone/blob/master/module_1_L3/labs-jupyter-spacex-Data%20wrangling.ipynb



EDA with Data Visualization

- **The charts that are used for EDA:**
  - **Scatter Plots: scatter plots can be used to inspect the correlations or patterns between two numerical variables, it is used to view the following relationships**
    - the impact to success launching of relationship between flight number payload
    - the impact to success launching of relationship between flight number and launch site
    - the impact to success launching of relationship between launch site and payload
    - the impact to success launching of relationship between flight number and orbit type
    - the impact to success launching of relationship between payload and orbit type
  - **Bar Charts: bar charts can be used to show the frequency distribution of categorical variables, it is used to demonstrate:**
    - The success rate distributed against the orbit type
  - **Line Plots: line plots can be used to display the trend of a continuous data over time, it is used to visualize:**
    - The success rate climbing trend over years
- **Reference:** https://github.com/Lorby04/IBMCapstone/blob/master/module_2/Visualization/edadataviz.ipynb

EDA with SQL

- To perform EDA with SQL, the following operations are used to show relevant results:
  - SELECT DISTINCT "Launch_Site" FROM SPACEXTBL: to display the names of unique launch sites
  - SELECT * FROM SPACEXTBL WHERE "Launch_Site" LIKE 'CCA%' LIMIT 5: to display 5 records where launch sites begin with "CCA"
  - SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE "Customer" = 'NASA (CRS)' : to calculate total payload mass carried by boosters launched by NASA (CRS)
  - SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE "Booster_Version" LIKE 'F9 v1.1' : to calculate average payload mass carried by booster version F9 v1.1
  - SELECT MIN(Date) FROM SPACEXTBL WHERE "Landing_Outcome" LIKE 'Success (ground pad)' : to find the date of first successful landing outcome on ground pad
  - SELECT "Booster_Version" FROM SPACEXTBL WHERE "Landing_Outcome" LIKE 'Success (drone ship)' AND "PAYLOAD_MASS__KG_" BETWEEN 4000 AND 6000 : to list boosters with successful drone ship landings and payload mass between 4000-6000 kg
  - SELECT COUNT(CASE WHEN "Mission_Outcome" LIKE 'Success%' THEN 1 END) AS Successful, COUNT(CASE WHEN "Mission_Outcome" LIKE 'Failure%' THEN 1 END) AS Failure FROM SPACEXTBL : to count total number of successful and failed mission outcomes
  - SELECT "Booster_Version" FROM SPACEXTBL WHERE "PAYLOAD_MASS__KG_" = (SELECT MAX("PAYLOAD_MASS__KG_") FROM SPACEXTBL) : to list booster versions that carried maximum payload mass
  - SELECT CASE *(month to name mapping)* ... END AS month, Landing_Outcome, Booster_Version, Launch_Site FROM SPACEXTBL WHERE Landing_Outcome LIKE 'Failure (drone ship)' AND substr(Date, 0, 5) = '2015' : to list month names, failure landing outcomes on drone ships, booster versions, and launch sites for 2015
  - SELECT "Landing_Outcome", COUNT(*) as outcomes FROM SPACEXTBL WHERE "Date" BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY "Landing_Outcome" ORDER BY outcomes DESC : to rank count of landing outcomes between 2010-06-04 and 2017-03-20 in descending order
- **Reference:** https://github.com/Lorby04/IBMCapstone/blob/master/module_2/SQLLIte/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- **Objects added in the map:**
  - A highlighted circle for each launch site.
  - A marker with the name of site for each launch site
  - A marker cluster with green marker for success and red marker for failure on each launch site
  - A distance marker at coastline near to the site CCAFS SLC-40
  - A distance line from the site CCAFS SLC-40 to the coastline
- **Reasons to add the objects in the map:**
  - The highlighted circle plus with the marker with site name provides intuitive insights for people to be aware of where all the sites are.
  - To group markers with colored indication of success or failure of launch in a marker cluster for each site provides a good map readability
  - The distance marker on the closest coastline and distance line between the site and coastline gives visually connected site and the coastline for people to understand the geographic position of the site better .
- **Reference**:

  https://github.com/Lorby04/IBMCapstone/blob/master/module_3/interactive/lab_jupyter_launch_site_location.ipynb

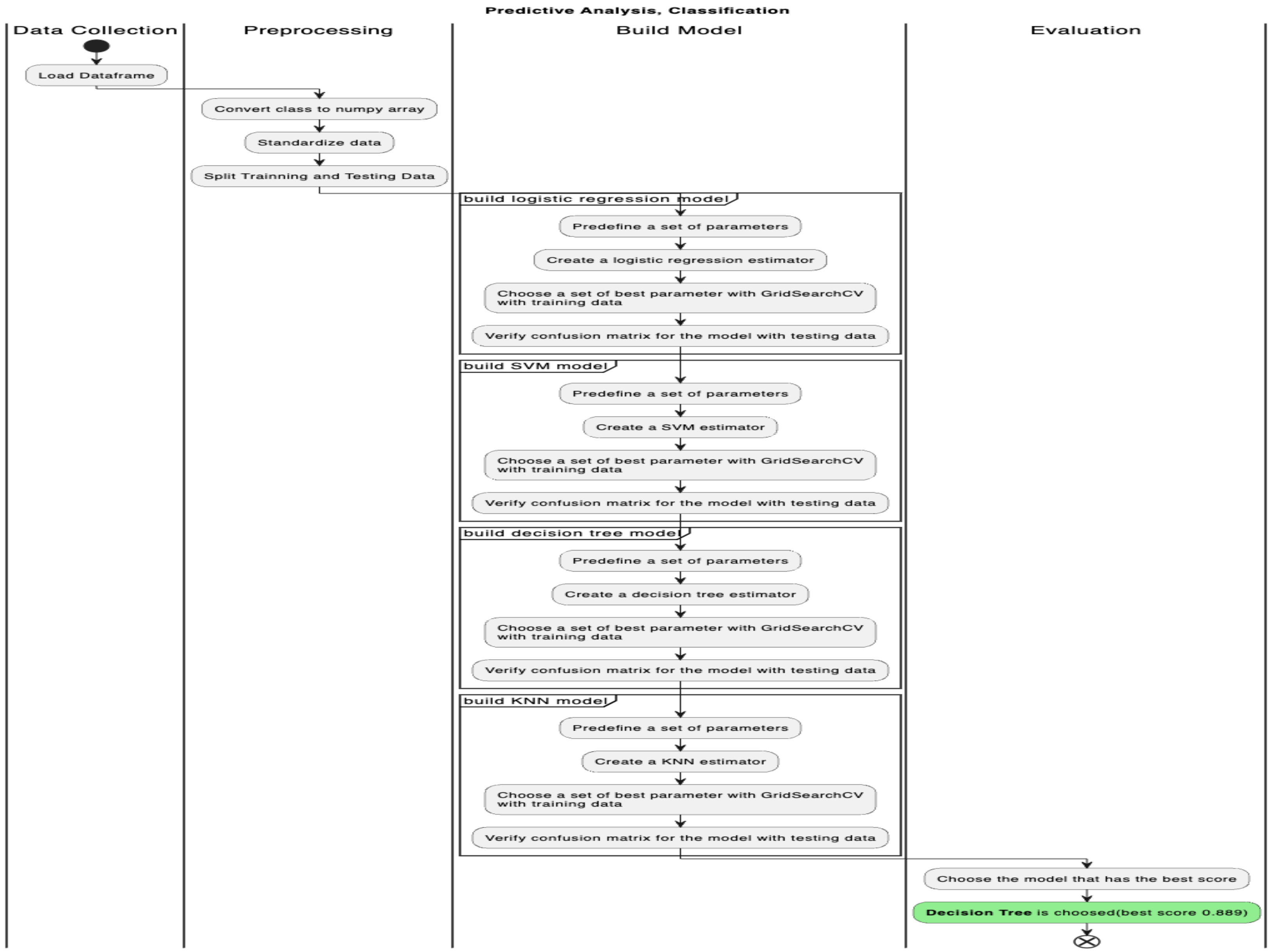
Build a Dashboard with Plotly Dash

- A dashboard is built with Plotly Dash including the following elements:
  - Dropdown menu for launch site selection
  - Pie chart for success rate visualization
  - Range slider for payload mass selection
  - Scatter plot for payload mass vs. launch outcome
- Reason to add the elements:
  - The dashboard needs to be clean to show major information only, so it cannot include too many elements.
  - The dropdown menu (site-dropdown allows users to select a specific launch site or view data for all sites, which is crucial for comparing performance across different launch sites quickly.
  - The pie chart (success-pie-chart) visualizes the success rate of launches of "All sites" or success vs failure ratio for selected sites. It gives the user a quick intuitive understanding of the reliability of each site and overall situation.
  - The range slider (payload-slider) and the corresponding scatter plot (success-payload-scatter-chart) allows user to analyze how payload mass might affect the outcome of launch which enables users to focus on specific payload ranges of interest.
- Reference: https://github.com/Lorby04/IBMCapstone/blob/master/module_3/dashboard/spacex_dash_app.py

Predictive Analysis (Classification)

- To build the classification, here is the steps being adopted:
  - **Data Collection**
  - **Data Preprocessing**
    - Standardize the data
    - Split the data into training and test sets
  - **Model Building and Selection**
    - Logistic Regression
    - SVM
    - Decision Tree
    - KNN
    - Running GridSearchCV to select the best parameters for each model
  - **Evaluation**
    - Evaluate the models, select the one producing the best score
    - In the above models, SVM is the winner with a high score of 0.944
- Flowchart:
  - See the next slide
- Reference: https://github.com/Lorby04/IBMCapstone/blob/master/module_4/prediction/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Predictive Analysis (Classification, Flowchart)



Results

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

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

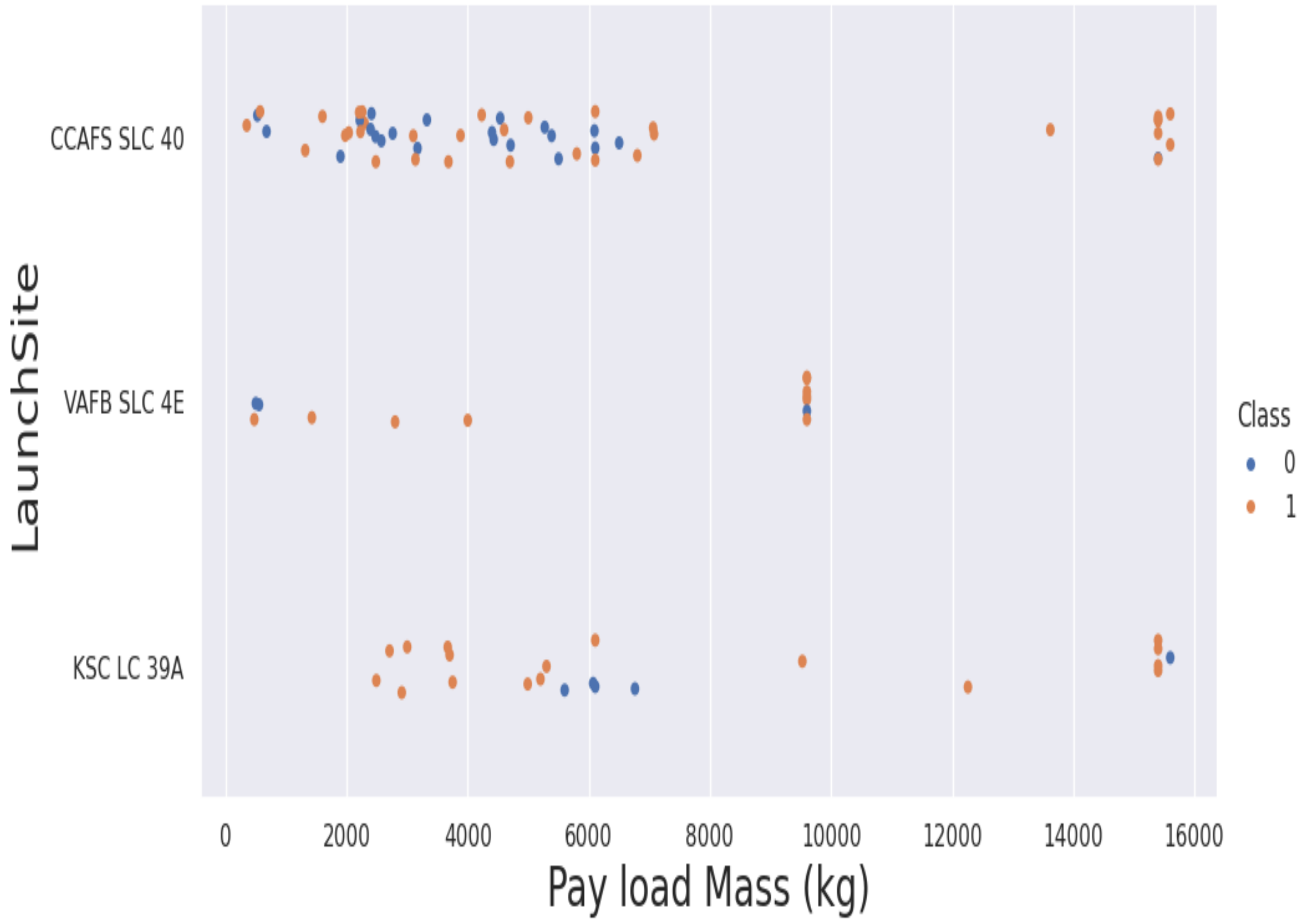
Key observations from the plot:

- **Learning Curve**:
  - More frequent of failures in all sites for earlier flights
  - More successful landing in all sites for later flights
- **Launch Site Performance**:
  - **CCAFS SLC 40**: The site takes more flights than the other 2 sites, it is the primary site. Successful rate is also the highest in all the 3 sites for later flights although the total success rate of this site is lower than others due to high failure rate in the beginning.
  - **VAFB SLC 4E**: The least flights, a few failure.
  - **KSC LC 39A**: The success rate is also high, and it is improved along with more flights being tried.
- **Consistency**:
  - CCAFS SLC 40 and KSC LC 39A show increasing reliability.
  - No enough data to give conclusion on VAFB SLC .

Payload vs. Launch Site

Key observations from the plot:

- **CCAFS SLC 40:**
    - Most frequently used site, same observation as previous one
    - Handles a wide range of payload masses for (0-8000 kg) and (12000-16000 kg)
    - Low success rate on light payload (0-8000 kg)
    - High success rate(no failure) on heavy payload (12000-16000 kg), the reason could be the heavy payload was launched later
- **VAFB SLC 4E:**
    - The least launches
    - Mostly light to middle masses payloads (0-4000 kg) and (8000-10000 kg)
    - Success rate is high on middle masses payload (8000-10000 kg)
- **KSC LC 39A:**
    - Handles a wide rage of payload masses (2000-16000 kg)
    - Good performance on all payload masses, esp. with the light payload masses.
    - Few failures with mid-range payloads
- **Summary**:
    - Success rates seem higher for heavier payloads
    - Each site specializes in different payload mass ranges
    - CCAFS SLC 40 appears most successful with heavy payloads



Success Rate vs. Orbit Type

Key observations from the plot:

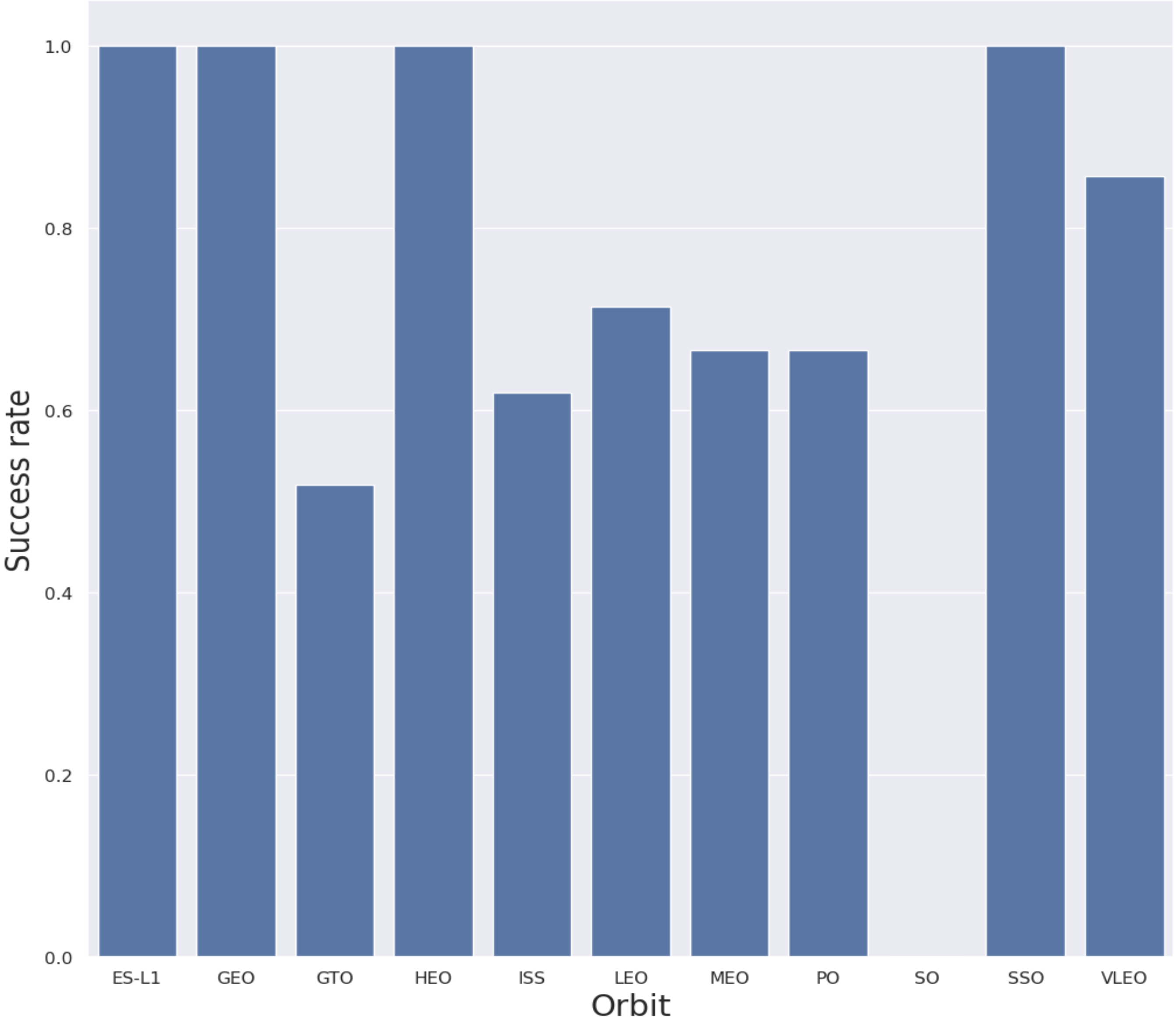
- **High Success Rates (100%)**:
  - ES-L1
  - GEO
  - HEO

  These orbits have a perfect success rate, indicating reliable missions to these destinations.
- **Moderate Success Rates (> 50%)**:
  - ISS (~62%)
  - LEO (~70%)
  - MEO (~65%)
  - PO (~65%)
  - SSO (~99%)
  - VLEO (~85%)

  These orbits have high but not perfect success rates, showing good reliability.
- **Low Success Rate (<= 50%)**:
  - GTO (~50%)
  - SO(0%)

  This orbit has the lowest success rate, indicating significant challenges or risks associated with missions to this destination.
- **Summary** :

  The data suggests that orbits at higher altitudes or with specific purposes (ES-L1, GEO, HEO) have perfect success rates. On the other hand, GTO and SO, involving complex maneuvers, has lower success rates. LEO and related orbits (ISS, VLEO, SSO, etc.) demonstrate good but not perfect success rates.

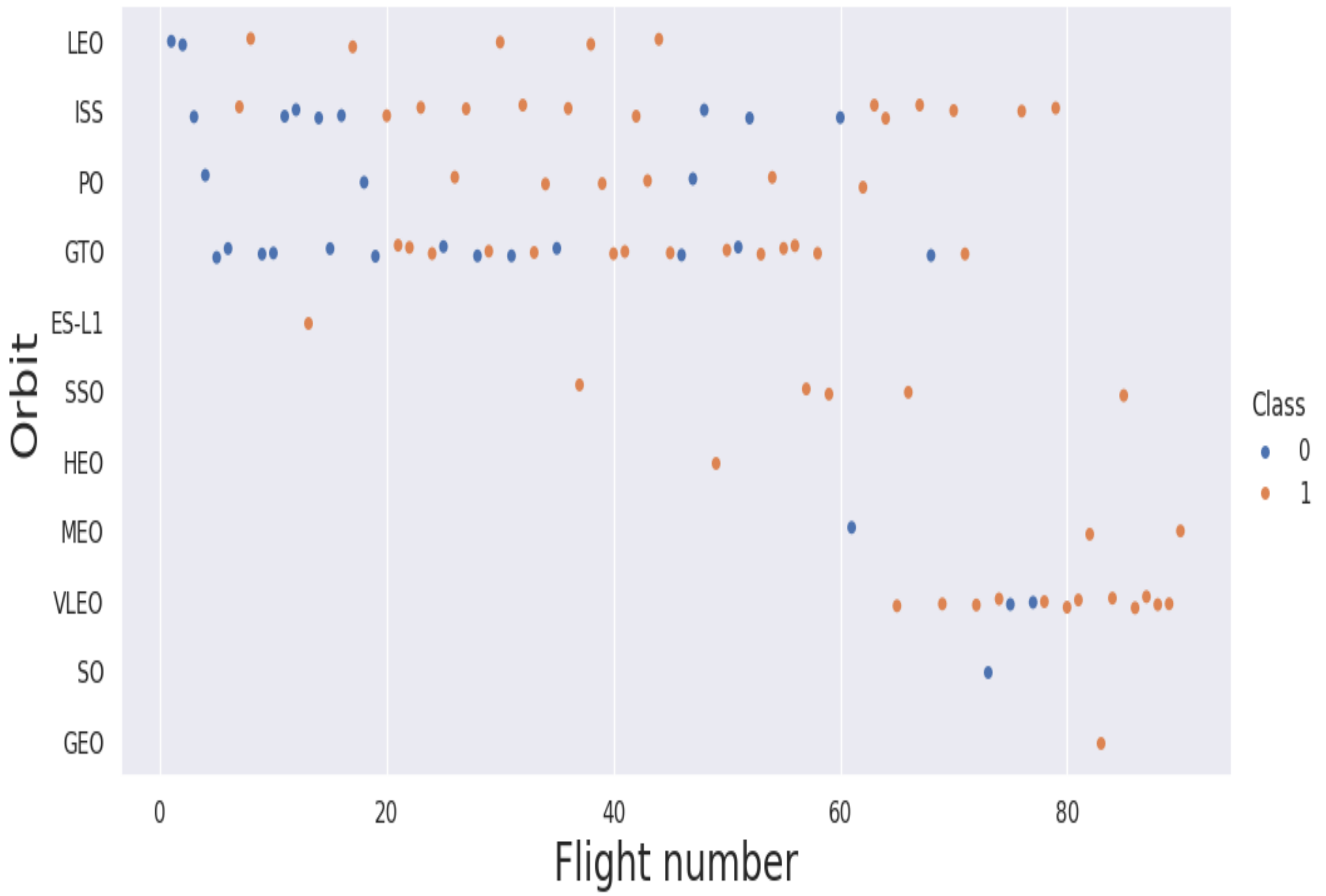


Flight Number vs. Orbit Type

Key observations from the plot:

- **Facts:**
  - **Diverse Orbit Types in Later Flights:** More orbit types are added in later flight.
  - **Improving Success Rates:** Success rates generally improve over time with increasing of flight number for most orbit types, indicating technical and planning improvement.
  - **Higher Failure Rate On GTO:** GTO stands out with a consistently higher failure rate, suggesting it is a more challenging target.
  - **Only one try on SO:** Failed
- **Conclusion:**

  This data suggests a clear evolution in launch capabilities, with an increase in the diversity of mission types and targeted orbits. There is a general trend of improving success rates, reflecting technological advancements and gained experience. However, certain orbits, particularly GTO and SO, remain challenging with a higher incidence of failures. Which indicates potential technical challenges

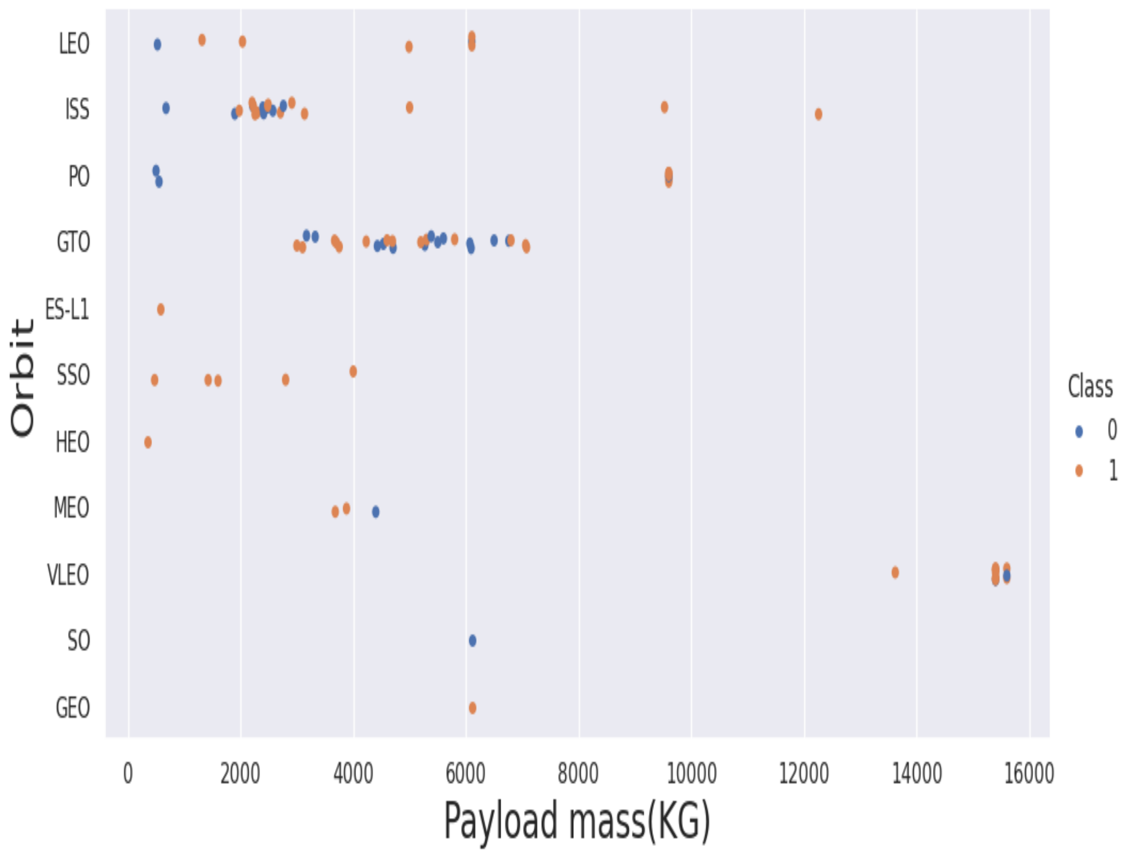


Payload vs. Orbit Type

Key observations from the plot:

- **Facts:**
  - LEO: Mainly for light payload masses, mostly successful
  - ISS: Wide range of payload masses from less than 2000kg to more than 12000kg, success rate is high, among which heavier load performs even better.
  - PO: Light and moderate payload masses, all light payload masses failed, and all moderate payload masses succeeded
  - GTO: Moderate masses of payload, range (2000-8000 kg), about half success rate, lower than other sites.
  - ES-L1: Single light payload, successful
  - SSO: Several launches with light mass payload (less than 4000 kg), all successful
  - HEO: Single light payload, successful
  - MEO: Few launches with moderate mass of payload (~4000kg), mixed success
  - VLEO: Heaviest payloads (13000-16000 kg), mostly successful
  - SO: Single moderate payload (~6000 kg), failed
  - GEO: Single moderate payload (~6000 kg), successful

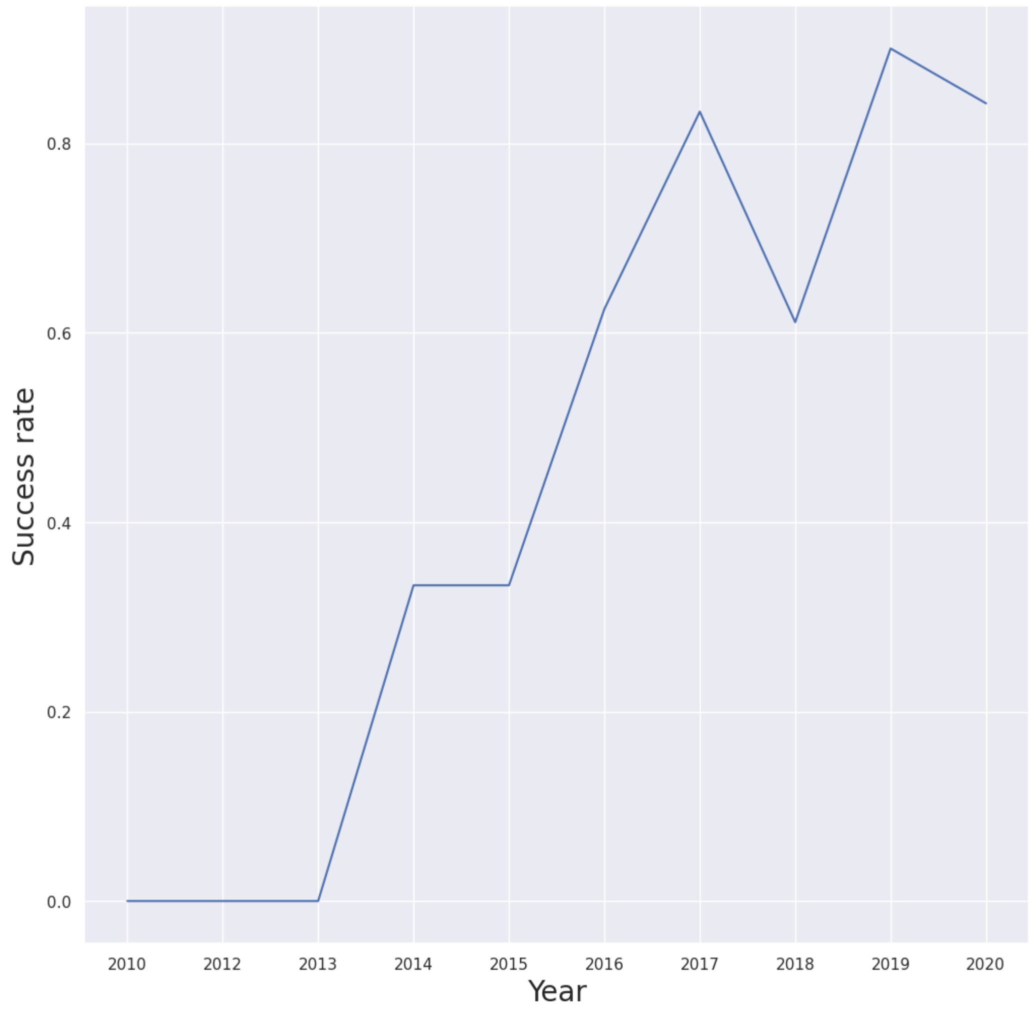
- **Summary:**
  - Heavier payloads tend towards higher success rates
  - Different orbits specialize in different payload mass ranges
  - More failure rate on GTO
  - The heaviest payload are launched on VLEO solely



Launch Success Yearly Trend

Key observations from the plot:

- Before the year of 2013, the success rate is 0
- Since 2013 till 2020, the success rate kept improving, except a slight dip in 2018.



All Launch Site Names

- Find the unique launch site with the SQL query :

```
SELECT DISTINCT "Launch_Site" FROM SPACEXTBL
```

- The result coming from the query is [('CCAFS LC-40',), ('VAFB SLC-4E',), ('KSC LC-39A',), ('CCAFS SLC-40',)], which means the unique launch sites are:
  - CCAFS LC-40
  - VAFB SLC-4E
  - KSC LC-39A
  - CCAFS SLC-40

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA` with the SQL command:

```
SELECT * FROM SPACEXTBL
WHERE "Launch_Site" LIKE 'CCA%'
LIMIT 5
```

- The 5 records begin with `CCA` retrieved from database is as the following:
  - ('2010-06-04', '18:45:00', 'F9 v1.0 B0003', 'CCAFS LC-40', 'Dragon Spacecraft Qualification Unit', 0, 'LEO', 'SpaceX', 'Success', 'Failure (parachute)'),
  - ('2010-12-08', '15:43:00', 'F9 v1.0 B0004', 'CCAFS LC-40', 'Dragon demo flight C1, two CubeSats, barrel of Brouere cheese', 0, 'LEO (ISS)', 'NASA (COTS) NRO', 'Success', 'Failure (parachute)'),
  - ('2012-05-22', '7:44:00', 'F9 v1.0 B0005', 'CCAFS LC-40', 'Dragon demo flight C2', 525, 'LEO (ISS)', 'NASA (COTS)', 'Success', 'No attempt'),
  - ('2012-10-08', '0:35:00', 'F9 v1.0 B0006', 'CCAFS LC-40', 'SpaceX CRS-1', 500, 'LEO (ISS)', 'NASA (CRS)', 'Success', 'No attempt'),
  - ('2013-03-01', '15:10:00', 'F9 v1.0 B0007', 'CCAFS LC-40', 'SpaceX CRS-2', 677, 'LEO (ISS)', 'NASA (CRS)', 'Success', 'No attempt')

Total Payload Mass

- Calculate the total payload carried by boosters from NASA with the command:

```
SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTBL
WHERE "Customer" = 'NASA (CRS)'
```

- The result coming from the query is [(45596,)] which means the total payload mass carried by boosters from NASA is 45596 kg.

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1with the following SQL query:

```
SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTBL
WHERE "Booster_Version" LIKE 'F9 v1.1’
```

- The result from the query is [(2928.4,)] which means the average payload mass carried by F9 v1.1 is 2928.4kg

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad with the SQL query:

```
SELECT MIN(Date) FROM SPACEXTBL
WHERE "Landing_Outcome" LIKE 'Success (ground
pad)
```

- The result coming from the query is [('2015-12-22',)] which means the date of the first successful landing outcome on ground pad is 2015-12-22

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 with the SQL query:

```
SELECT "Booster_Version" FROM SPACEXTBL
WHERE "Landing_Outcome" LIKE 'Success (drone ship)'
AND "PAYLOAD_MASS__KG_" BETWEEN 4000 AND 6000
```

- The requested result from the query is:

  [('F9 FT B1022',), ('F9 FT B1026',), ('F9 FT B1021.2',), ('F9 FT B1031.2',)]

  It indicates the name of the boosters are:

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

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes with the following SQL query:

```
SELECT
COUNT(
    CASE
        WHEN "Mission_Outcome" LIKE 'Success%' THEN 1
        END) AS Successful,
COUNT(
    CASE
        WHEN "Mission_Outcome" LIKE 'Failure%' THEN 1
        END) AS Failure
FROM SPACEXTBL
```

- The result from the query is [(100, 1)] which indicates 100 successful and 1 failure mission outcomes

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass with the SQL query:

```
SELECT "Booster_Version" FROM SPACEXTBL
WHERE "PAYLOAD_MASS__KG_" = (
     SELECT MAX("PAYLOAD_MASS__KG_") FROM SPACEXTBL
)
```

- The result is:

[('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 ‘,)]

- According the result, the names of booster which carried the maximum payload mass are:

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

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015 with the SQL query:

```
SELECT
CASE
      WHEN substr(Date, 6, 2) = '01' THEN 'January'
      WHEN substr(Date, 6, 2) = '02' THEN 'February'
      WHEN substr(Date, 6, 2) = '03' THEN 'March'
      WHEN substr(Date, 6, 2) = '04' THEN 'April'
      WHEN substr(Date, 6, 2) = '05' THEN 'May'
      WHEN substr(Date, 6, 2) = '06' THEN 'June'
      WHEN substr(Date, 6, 2) = '07' THEN 'July'
      WHEN substr(Date, 6, 2) = '08' THEN 'August'
      WHEN substr(Date, 6, 2) = '09' THEN 'September'
      WHEN substr(Date, 6, 2) = '10' THEN 'October'
      WHEN substr(Date, 6, 2) = '11' THEN 'November'
      WHEN substr(Date, 6, 2) = '12' THEN 'December'
END AS month,
Landing_Outcome,
Booster_Version,
Launch_Site
FROM SPACEXTBL
WHERE Landing_Outcome LIKE 'Failure (drone ship)'
AND substr(Date, 0, 5) = '2015’
```

- The outcome from the query is: [('January', 'Failure (drone ship)', 'F9 v1.1 B1012', 'CCAFS LC-40'), ('April', 'Failure (drone ship)', 'F9 v1.1 B1015', 'CCAFS LC-40’)]. It means there are 2 failures in drone ship happened in year 2015, both with the booster 'F9 v1.1 B1015’ in the site 'CCAFS LC-40’, one failure happened in January and the other happened in April.

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

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

```
SELECT "Landing_Outcome",
COUNT(*) as outcomes
FROM SPACEXTBL
WHERE "Date" BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY "Landing_Outcome"
ORDER BY outcomes DESC
```

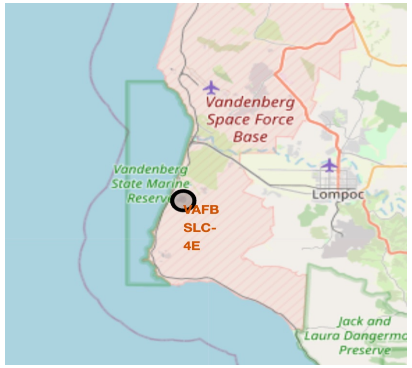
- The result is: [('No attempt', 10), ('Success (drone ship)', 5), ('Failure (drone ship)', 5), ('Success (ground pad)', 3), ('Controlled (ocean)', 3), ('Uncontrolled (ocean)', 2), ('Failure (parachute)', 2), ('Precluded (drone ship)', 1)]

- According to the query, between 2010-06-04 and 2017-03-20, there are:
  - 10 times 'No attempt',
  - 11 times landing on drone ship with 5 times failure , 5 times success and 1 precluded
  - 3 times landing on ground pad , all succeeded
  - 5 times landing on ocean, in which 3 times are controlled and the other 2 times are uncontrolled
  - 2 times landing with parachute, both failed

Section 3

Launch Sites Proximities Analysis

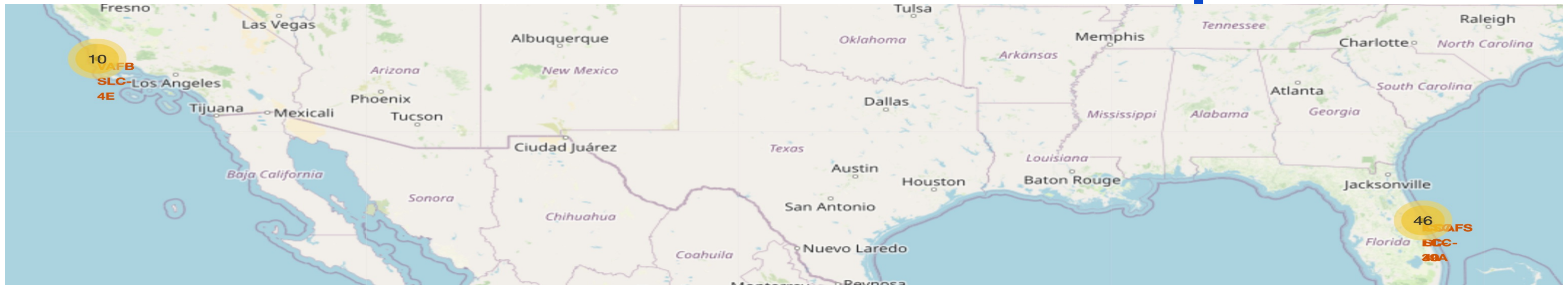
Launch Sites On Map

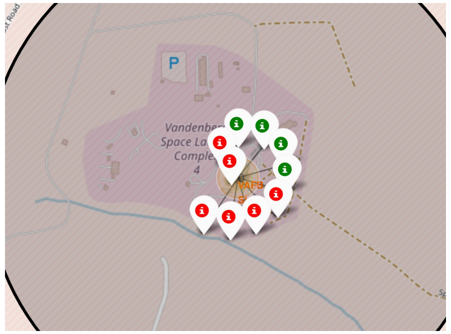


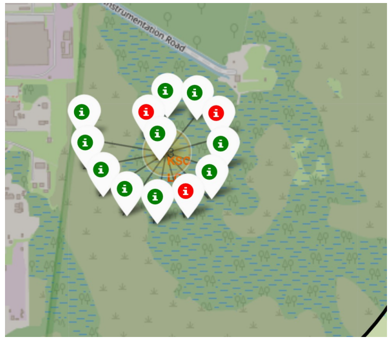
Key Observations:

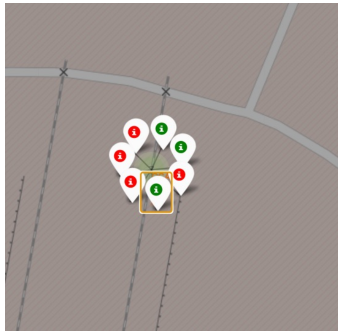
- All the four sites are located close to ocean. 1 (VAFB SLC-4E) is located at west coast in California and all the other 3 sites (KSC LC-39A, CCAFS SLC-40, CCAFS LC-40) are at southeast coast in Florida.
- The latitude of all the 4 sites are relatively low, the 3 sites in Florida are closer to the south.

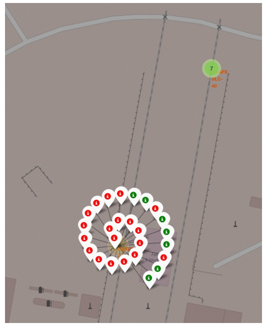
Launch Outcomes On The Map







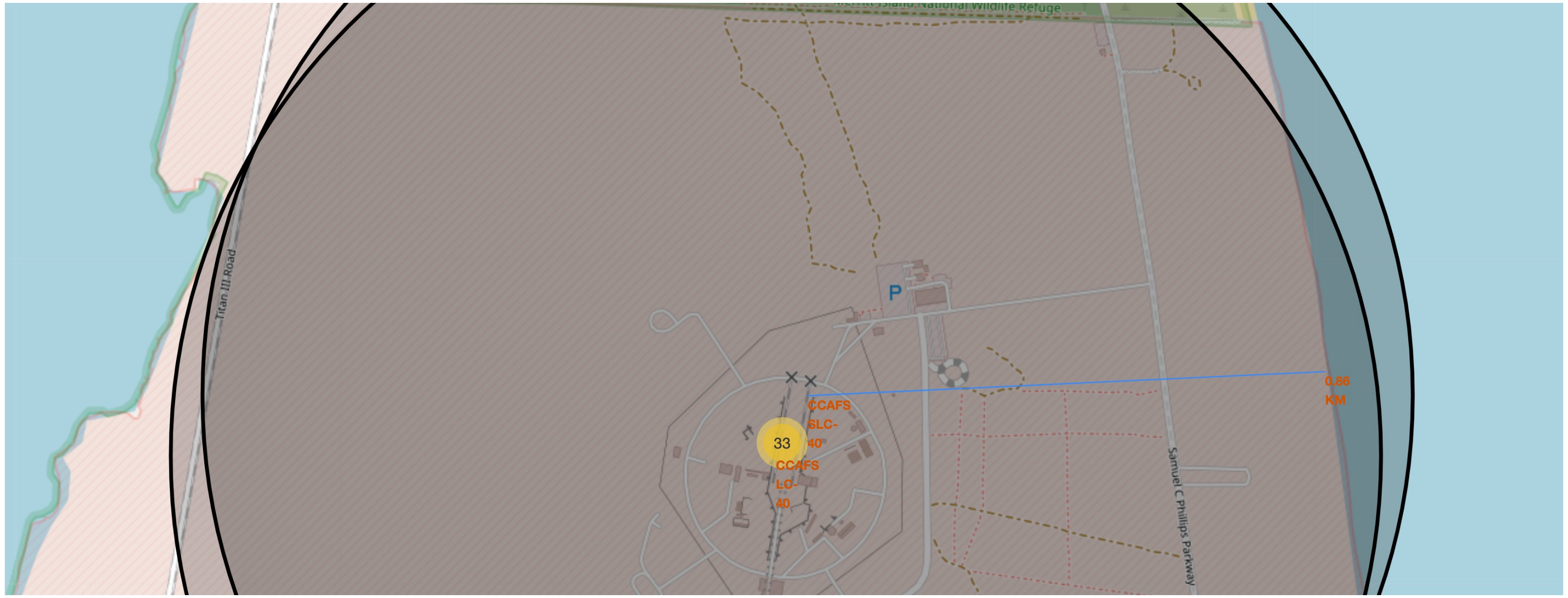




Key Observations:

- The map with outcomes provided an intuitive for people to understand the launch outcomes per site.
- From the insight view of each site, the failure rate on site VAFB SLC-4E and CCAFS LC-40 are higher than the other 2 sites.
- The success rate on site KSC LC39-A is much higher than other sites.
- CCAFS LC-40 had much more launches than other sites and CCAFS SLC-40 had the least launches.

Proximities Exploring



Key Observations:

- The launch site is close to railway and highway networks, which facilitates the efficient transportation of materials and cargoes necessary for launch operations.
- Specifically, the site is less than 1 km away from the coastline. This proximity is crucial for safety reasons and enables ocean-based landings, which are often used for recovering rocket stages after launch.

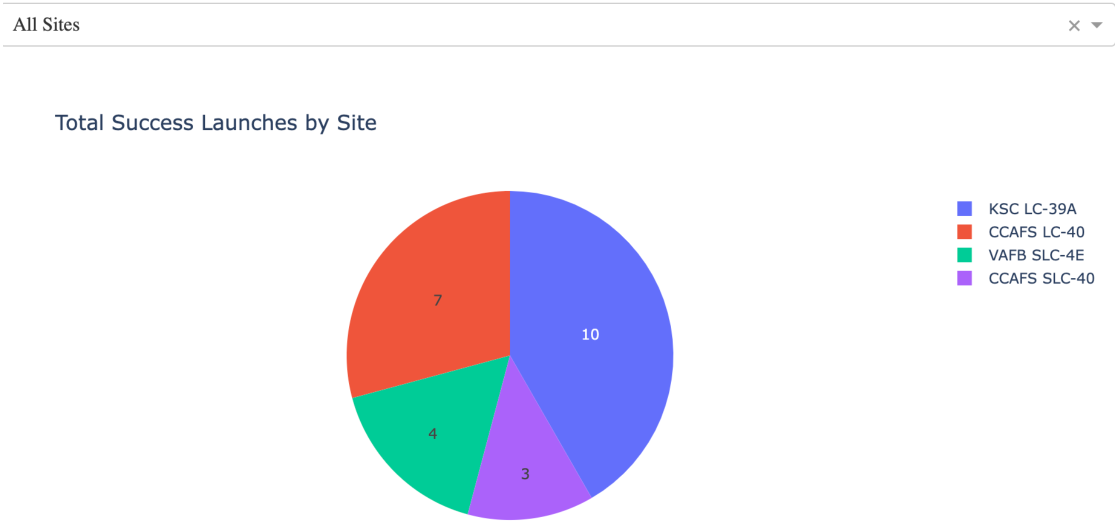
Section 4

Build a Dashboard with Plotly Dash

Successful Launches Overview

Key Observations:

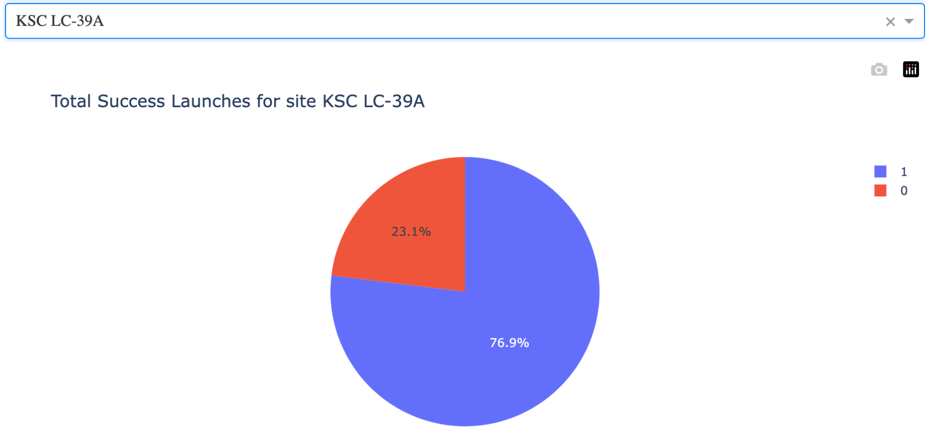
- The pie chart clearly shows the distribution of successful launches across four sites: KSC LC-39A (10 launches), CCAFS LC-40 (7 launches), VAFB SLC-4E (4 launches), and CCAFS SLC-40 (3 launches).
- The chart visually highlights KSC LC-39A as the most productive site for successful launches followed by CCAFS LC-40.



Most Productive Site Insights

Key Observations:

- The site with the highest success rate is KSC LC-39A
- The success rate of the site is 76.9%, close to 80%.



Launch Outcome With Different Payload

Key Observations:

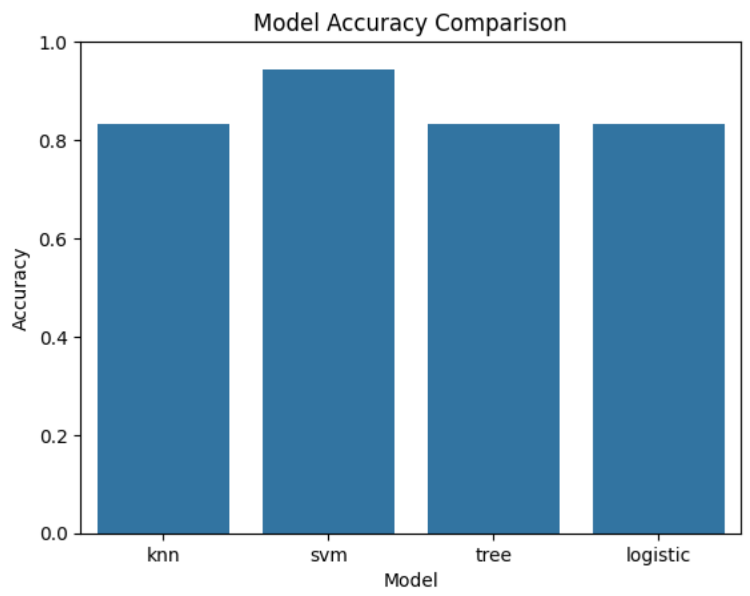
- Successes are distributed across different payloads lighter than 5500 kg and booster versions.
- All the payloads between 5500 and 7000 failed to launch.
- There is not significant evidence to show certain booster version that guarantees success, except that F9 v1.1 have more consistent successes.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

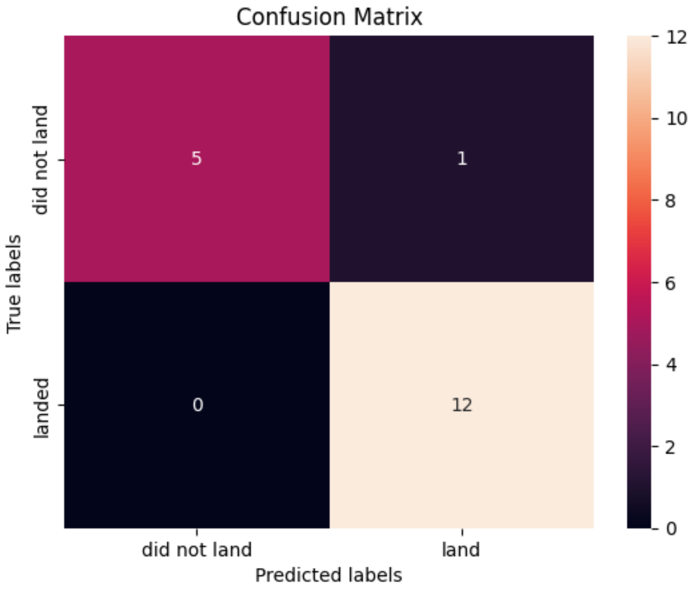
- The SVM model has the best accuracy among all the 4 models. The score is close to 0.95.



Confusion Matrix

What we got from the matrix:

- Base value:
  - TN: 5
  - FP: 1
  - TP: 12
  - FN: 0
  - Total: 18
- Performance Metrics:
  - Accuracy: (TP+TN)/Total = 17/18 = 0.944
  - Precision: TP/(TP+FP) = 12/13 = 0.92
  - Recall: TP/(TP+FN) = 12/12 = 1
  - F1 Score: 2*(Precision*Recall)/(Precision+Recall)=0.96



Conclusions

- **Project Undertakings**: This project entailed data collection from SpaceX's API and web scraping, followed by data wrangling, exploratory data analysis (EDA), and the development of predictive models using machine learning.
- **Data Preparing and EDA**:
  - **Data Processing**: Cleaning and processing the data allowed us to identify important features such as launch site, payload mass, and orbit type.
  - **Exploratory Analysis**: EDA revealed significant trends and relationships, showing that heavier payloads tend to have higher success rates and that launch success rates have generally improved over time. Interactive visualizations with Folium and Plotly Dash provided intuitive insights into launch site proximities and performance.
- **Predictive Modeling**:
  - **Model Evaluation**: Various classification models, including Logistic Regression, SVM, Decision Tree, and KNN, were evaluated.
  - **Optimal Model**: The SVM model emerged as the most accurate for predicting the success of the Falcon 9 first stage recovery.
- **Key Findings**:
  - The success rates of launches have increased over time since the year of 2013, with specific launch sites (KSC LC39-A), orbit types and payload ranges showing higher success.
  - Heavier payloads are generally more successful.
- **Predictions**:
  - The predictive models developed can estimate future launch costs and the likelihood of successful recovery and reuse of Falcon 9 rockets, supporting strategic planning and decision-making for upcoming missions.

Appendix

- All the code pieces and output including the report can be accessed through: https://github.com/Lorby04/IBMCapstone

Thank you!