IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Rolf Chung
2025. 06.02

Outline

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

Executive Summary - Summary of methodologies

- **Data Collection:**
  - Used the SpaceX REST API and web scraping techniques to gather historical data on rocket launches.
  - Processed JSON data into structured formats for analysis, filtering out irrelevant data like Falcon 1 launches.
- **Data Preprocessing:**
  - Addressed missing values by imputing the mean for numerical fields (e.g., PayloadMass) and applied one-hot encoding for categorical variables.
  - Created a target variable class
- **Visualization and Dashboarding:**
  - Leveraged Folium for geospatial analysis and interactive maps to study launch site proximities.
  - Developed a dashboard using Plotly Dash to provide dynamic insights through interactive charts and controls.
- **Predictive Modeling:**
  - Built a machine learning pipeline to predict first-stage landing success.
  - Tested multiple classifiers (Logistic Regression, SVM, Decision Tree, KNN) with hyperparameter tuning via Grid Search.
- **Performance Evaluation:**
  - Assessed model accuracy, precision, recall, and confusion matrix to identify the most effective predictor.

Executive Summary - Summary of results

- **Model Performance:**
    - The Decision Tree classifier achieved the best accuracy (83.3%) among all tested models, including KNN, SVM, and Logistic Regression.
    - It excelled in identifying successful launches (precision: 92.3%, recall: 85.7%), making it highly reliable for predicting successful outcomes.

- **Predictive Insights:**

    - The model provides valuable predictions to improve operational planning, risk management, and cost estimation for SpaceX launches.
    - While not revealing direct causes of success, it highlights factors associated with landing outcomes, offering a starting point for further investigation.

- **Monitoring Tools:**
    - Interactive dashboards and maps developed with Plotly Dash and Folium allow stakeholders to explore launch data dynamically.
    - These tools facilitate monitoring and operational decision-making by visualizing launch patterns and proximities of landing sites.

Overall, the combination of predictive modeling and interactive analytics tools offers a comprehensive solution for evaluating and optimizing SpaceX operations.

Introduction

- The commercial space industry is transforming space travel into a more affordable and accessible venture, with SpaceX leading the charge through innovative cost-saving strategies.
- SpaceX's Falcon 9 rocket is a key contributor to its success due to its reusable first stage, significantly reducing launch costs compared to competitors.
- A Falcon 9 launch costs approximately $62 million, whereas other providers charge upwards of $165 million.
- Reusability of the first stage not only makes launches more economical but also sets SpaceX apart in a competitive market.
- The project aims to analyze public data to predict the likelihood of Falcon 9’s first stage successfully landing, which directly impacts launch cost estimation.
- Developing maps and dashboards for monitoring the SpaceX operation.
- Prediction of Landing Success: Can machine learning models accurately predict whether the Falcon 9 first stage will land successfully?

Section 1

Methodology

Methodology

- Data collection methodology:
  - The SpaceX launch data was gathered from an API, specifically the SpaceX REST API.
  - The other source for obtaining Falcon 9 Launch data was web scraping related Wiki pages.
- Perform data wrangling
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

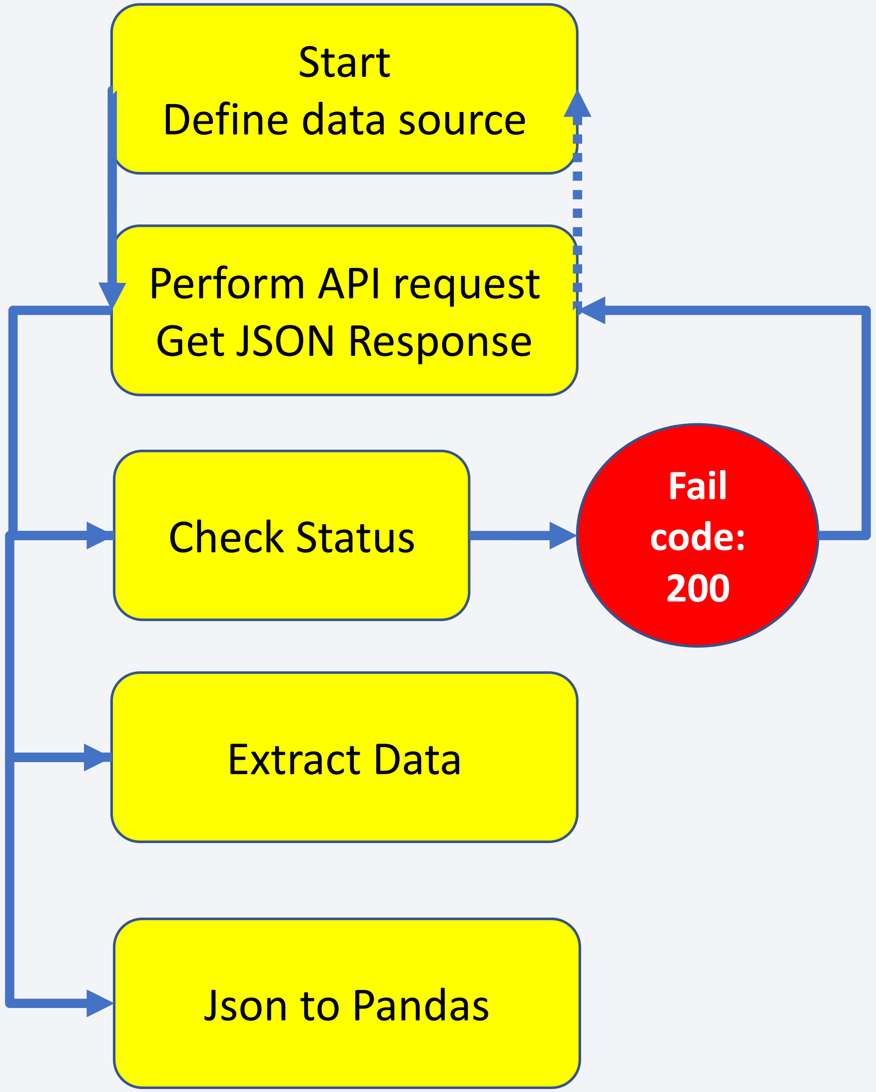
Data Collection

The SpaceX launch data was gathered from an API, specifically the SpaceX REST API. The response were a list of JSON objects. These were converted into a data frame using the Python Pandas library.

The other source for obtaining Falcon 9 Launch data was web scraping related Wiki pages. The Python BeautifulSoup package was used to web scrape some HTML tables that contain valuable Falcon 9 launch records.

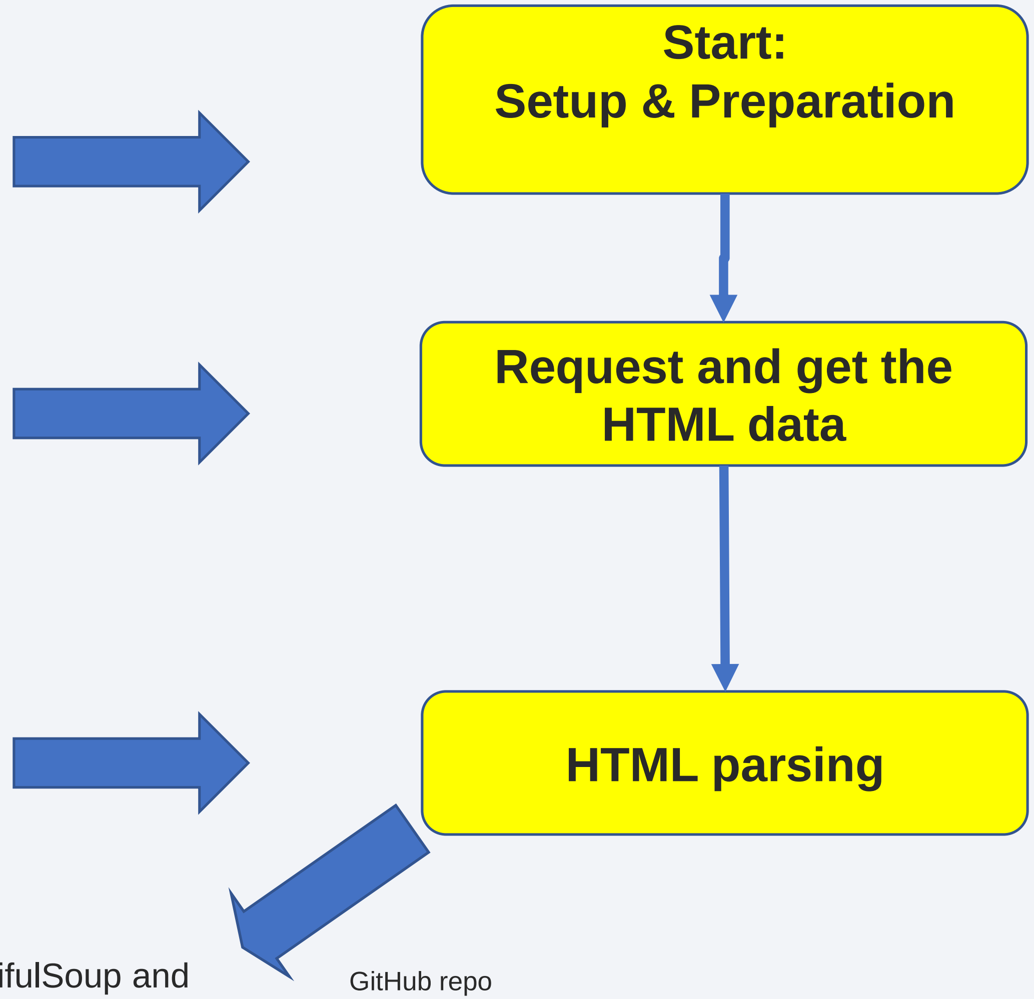
Data Collection – SpaceX API

- **Data collection with SpaceX REST API calls:**
  - Identify the API and specific endpoint (e.g., `api.spacexdata.com/v4/launches/past`)
  - Perform API Request
  - Use the `requests` library to send a GET request to the API endpoint.
  - Receive JSON Response in JSON format.
  - Check Response Status: Verify status code success = 200
  - If unsuccessful**, log the error and stop. If successful**, proceed.Extract Data**
  - Call the `.json()`and Normalize JSON method to parse the JSON response into Python objects and a tabular structure.
  - Convert the normalized JSON into a Pandas DataFrame.
  - Inspect & Store Data
- GitHub repo
  - https://github.com/RolfChung/Interactive_Vis_SpaceX/blob/main/jupyter-labs-spacex-data-collection-api.ipynbäö##ä



Data Collection - Scraping

- **Setup & Preparation:**
  - Installing and importing Packages: BeautifulSoup , requests
  - Define Helper Functions
- **Request and get the HTML data:**
  - Conducting the HTTP Request and set target URL: Wikipedia Falcon 9 launches page
  - Send GET Request and verify Response: 200 status code
- **HTML parsing:**
  - Get raw HTML content from response
  - Extract HTML Text and BeautifulSoup Object
  - Verify and locate tables Page
- **Further actions**:
  - Transform the HTML through data wrangling using Python, BeautifulSoup and the helper functions



GitHub repo

https://github.com/RolfChung/Interactive_Vis_SpaceX/blob/main/jupyter-labs-webscraping.ipynb

Data Wrangling

Goals

The primary goals for the data wrangling process were:

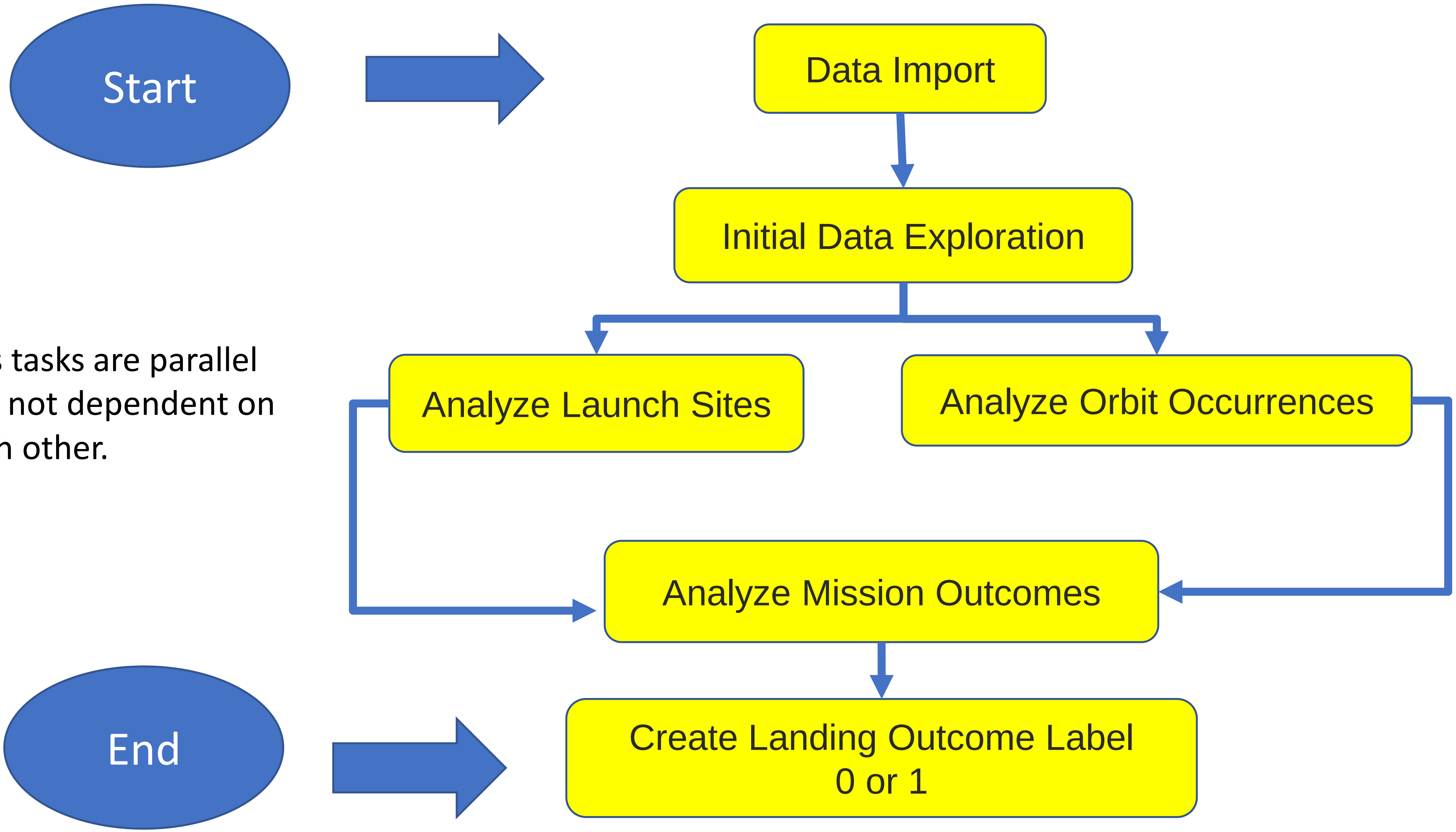
Perform exploratory data analysis to understand the data. Without this understanding, the next task would not have been possible

Determine the appropriate labels (1 for successful landing, 0 for unsuccessful) for training supervised machine learning models.

Data Wrangling tasks

- Data Import from CSV to Pandas DataFrame.
- Initial Data Exploration:
  - missing values, data types
- Analyze Launch Sites:
  - “value_counts()” to count the number of launches from each unique site
- Analyze Orbit Occurrences:
  - “value_counts()” to determine the frequency of each orbit type
- Analyze Mission Outcomes:
  - “value_counts()” to tally the occurrences of different landing outc-omes.
  - Defining “bad_outcomes” include all outcomes that represent an unsuccessful first stage landing
- Create Landing Outcome Label:
  - A new column ‘class’ is created.
  - For bad_outcomes \`0\` is set, \`bad_outcomes\` Otherwise is set \`1\`

Data Wrangling – flow chart



This tasks are parallel and not dependent on each other.

EDA with Data Visualization

What variables are plotted?	Plot type	What is the plot saying?
Relationship of Flight Number and PayloadMass by Class'	sns.catplot as scatterplot	As the flight number increases, the first stage is more likely to land successfully. With more massive payloads, the first stage often returns successfully.
Launch Site of Flight Number and PayloadMass by Class	sns.catplot as scatterplot	As the flight number increases the successful launches are increasing, too. CCFA has the most launches. VAFB has the smallest number of launches. The number of successful launches are similar distributed over the launch sites.
Relationship of launch site and PayloadMass by Class	sns.catplot as scatterplot	Now if you observe Payload Mass Vs. Launch Site scatter point chart you will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).
Mean success rate by orbit	sns.barplot	The ES-L1, GEO and HEO are the three highest success rates. The ISS and the GTO have low success rates. SO has got a success rate of 0.

EDA with Data Visualization

What variables are plotted?	Plot type	What is the plot saying?
Relationship of FlightNumber and Orbit type by Class	sns.catplot as scatterplot	The LEO orbit, success seems to be related to the number of flights. Conversely, in the GTO orbit, there appears to be no relationship between flight number and success. MEO, VLEO were only used with high FlightNumbers.It seems with increasing FlighNumbers additional orbits were tested. The VLEO orbit has a high success rate an. The SSO has only successes but has not many FlightNumbers. LEO and GTO were used with low FlightNumbers.
Success rate over year	sns.lineplot	From 2010 to 2013 there were none successful launches.The sucess rate since 2013 kept increasing till 2020. In 2018 there was a setback. The peak was 2019 and to 2020 there is small drop.

GitHub URL:
https://github.com/RolfChung/Interactive_Vis_SpaceX/blob/main/edadataviz.ipynb

Additionally feature engineering was conducted.

EDA with SQL

Performed SQL queries:

Display the names of the unique launch sites in the space mission:
pd.read_sql("SELECT DISTINCT Launch_Site FROM SPACEXTABLE", con)

Display 5 records where launch sites begin with the string 'CCA‘
query_2 = "SELECT Launch_Site FROM SPACEXTABLE WHERE Launch_Site LIKE 'CCA%' LIMIT 5“

Display the total payload mass carried by boosters launched by NASA (CRS)
query_3 = "SELECT SUM(PAYLOAD_MASS__KG_) AS TotalPayloadMass_Nasa FROM SPACEXTABLE WHERE Customer LIKE '%NASA (CRS)%' “

GitHub URL:
https://github.com/RolfChung/Interactive_Vis_SpaceX/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

EDA with SQL

- Display average payload mass carried by booster version F9 v1.1
  query_4 = "SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTABLE WHERE Booster_Version LIKE 'F9 v1.1'“

- Display average payload mass carried by booster version F9 v1.1
  query_4 = "SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTABLE WHERE Booster_Version LIKE 'F9 v1.1'"

- List the date when the first succesful landing outcome in ground pad was acheived
  query_5 = """SELECT MIN(Date) FROM SPACEXTABLE WHERE Landing_Outcome LIKE 'Success (ground pad)'""“

- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
  query_6 = """SELECT Booster_Version FROM SPACEXTABLE WHERE Landing_Outcome LIKE 'Success (drone ship)' AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000""“

- List the total number of successful and failure mission outcomes
  query_7a = """SELECT CASE WHEN Mission_Outcome LIKE '%Success%' THEN 'Success' WHEN Mission_Outcome LIKE '%Failure%' THEN 'Failure‘ ELSE 'Error‘ END AS success_or_failure, COUNT(*) AS Total_Count FROM SPACEXTABLE GROUP BY success_or_failure;"""

EDA with SQL

- List all the booster_versions that have carried the maximum payload mass. Use a subquery.
  query_8 = """SELECT Booster_Version FROM SPACEXTABLE WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTABLE);""“

- List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.
  query_9 = """SELECT Booster_Version, substr(Date, 6,2) as month, substr(Date, 1, 4) as year, Launch_Site FROM SPACEXTABLE WHERE Landing_Outcome LIKE 'Failure (drone ship)’ AND substr(Date, 1, 4) = '2015';

- 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.
  query_9 = """SELECT CASE WHEN Landing_Outcome LIKE "%Success%" THEN "success“ WHEN Landing_Outcome LIKE "%Failure%" THEN "failure“ ELSE "Other“ END AS result, COUNT(Landing_Outcome) Total_Outcome FROM SPACEXTABLE WHERE Date BETWEEN "2010-06-04" AND "2017-03-20“ GROUP BY result ORDER BY Total_Outcome DESC;""""

Build an Interactive Map with Folium

The base site map was build centered around the launch sites

site_map = folium.Map(location=nasa_coordinate, zoom_start=10)

All launch sites were marked on the map

Added a highlighted circle area with a text label on specific launch sites
folium.Circle(nasa_coordinate, radius=1000, color='#d35400', fill=True)

Add Folium markers. The markers represent launch sites.
marker = folium.map.Marker(nasa_coordinate, icon=DivIcon() ...)

Success/failed launches were marked for each site cluster

Added a marker cluster. Marker clusters are a good way to simplify a map containing many markers having the same coordinate.

For each launch result in spacex_df data frame a folium.Marker was added to marker_cluster. The markers within the cluster makes it easy to distinguish between successful and failed launches easily.

marker_cluster = MarkerCluster()

Build an Interactive Map with Folium

- **Creating distance markers and line between a launch site and locations in proximity**
  - The locations selected for visualization were a coastline location and the neighboring Florida town of Titusville.
  - Distances markers were created and placed on the map.
    distance_marker = folium.Marker(location=coastline_point, icon=DivIcon(...))
  - PolyLines between a launch site to the selected locations were drawn.
    PolyLine(locations=[coastline_coord, launch_site_coord], ...).add_to(site_map)
  - Distance markers and PolyLines are valuable for numerous applications like traveling, operations or logistics.
  - The launch sites are close to coastline. They have access to nearby roads like the Centaur row and railway proximity to the Titan railway. The neighboring town of Titusvile is in some distance. This makes the Florida launch site a good choice for operations and security of SpaceX.

GitHub URL:
https://github.com/RolfChung/Interactive_Vis_SpaceX/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

A SpaceX Launch Records Dashboard od Plotly graphs was created with Dash app.layout and equipped with charts.

Pie chart

A pie chart was added depicting Success vs Failure for every launch site. The specific launch site is selected by a drop-down selection.

The Success vs Failure for every launch site is valuable information. Launch site with a higher success rate could be better launch sites for future launches.

Scatterplot

A 'success-payload-scatter-chart‘ was added depicting the relationship between payload and successes. The specific payload for the launch is selected with a slider graph. The graph is created for a specific launch site selected with the drop-down above. The data points are colored aligned with the booster version.

If given, a correlation between success and payload is valuable information and can guide adjustment to the boosters to increase chances of success.

GitHub URL:
https://github.com/RolfChung/Interactive_Vis_SpaceX/blob/main/new_dash_app.py

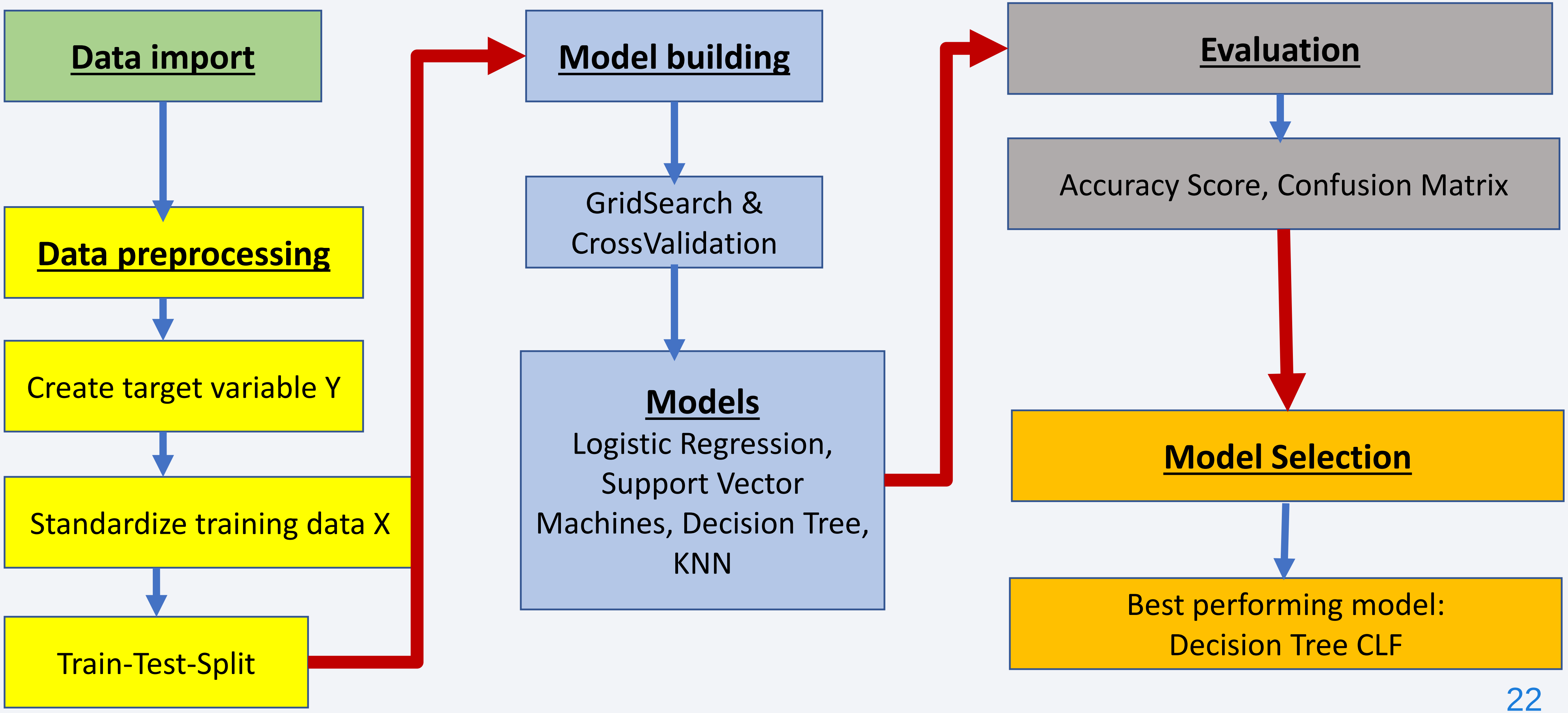
Predictive Analysis (Classification)

The machine learning pipeline for SpaceX landing prediction involved:

- **Data Preparation**: Separating features and labels, standardizing the feature data, and splitting it into training and test sets.
- **Model Building & Optimization:** Multiple classification models (Logistic Regression, SVM, Decision Tree, KNN) were built. Each model's hyperparameters were tuned using `GridSearchCV` with 10-fold cross-validation on the training data.
- **Model Evaluation:** Performance was rigorously assessed using the test data, calculating accuracy scores, visualizing results with confusion matrices, and providing detailed metrics via classification reports.
- **Best Model Selection:** The final best performing model was identified by comparing the test accuracy scores across all optimized algorithms.

GitHub URL:
https://github.com/RolfChung/Interactive_Vis_SpaceX/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Predictive Analysis (Classification)

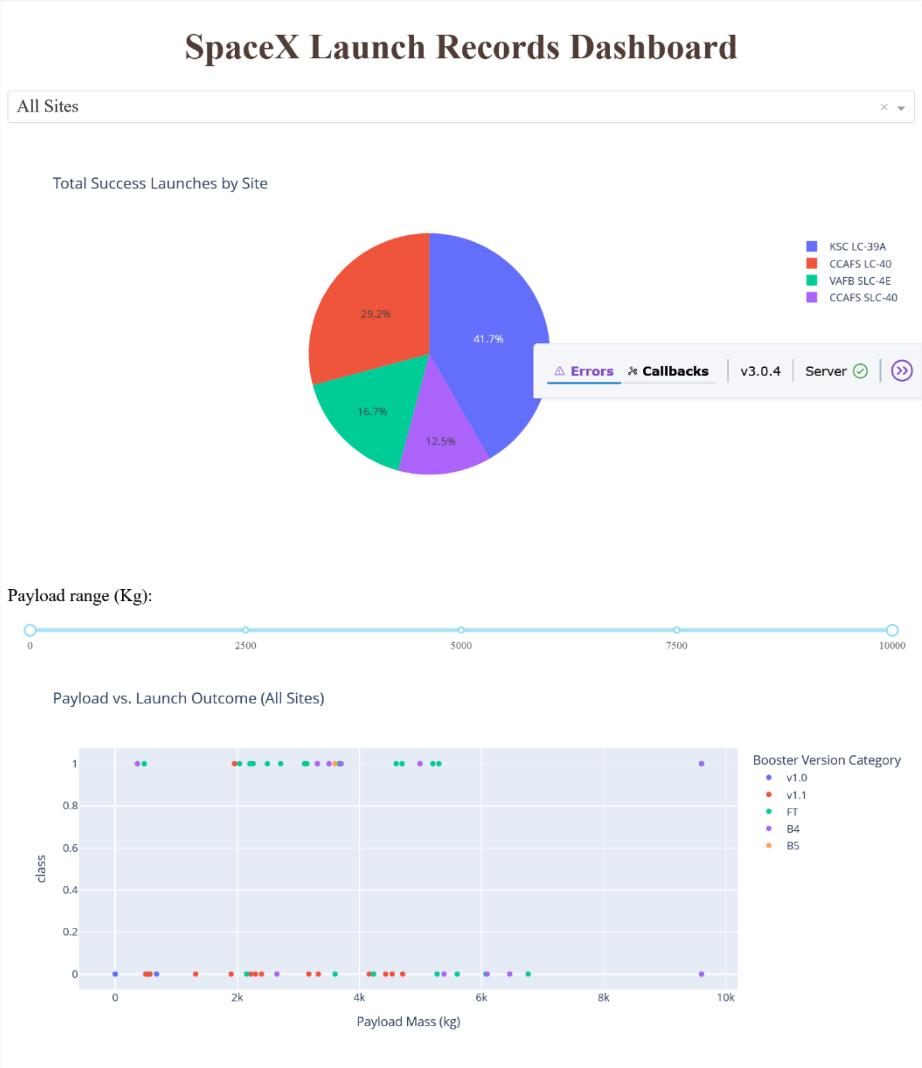


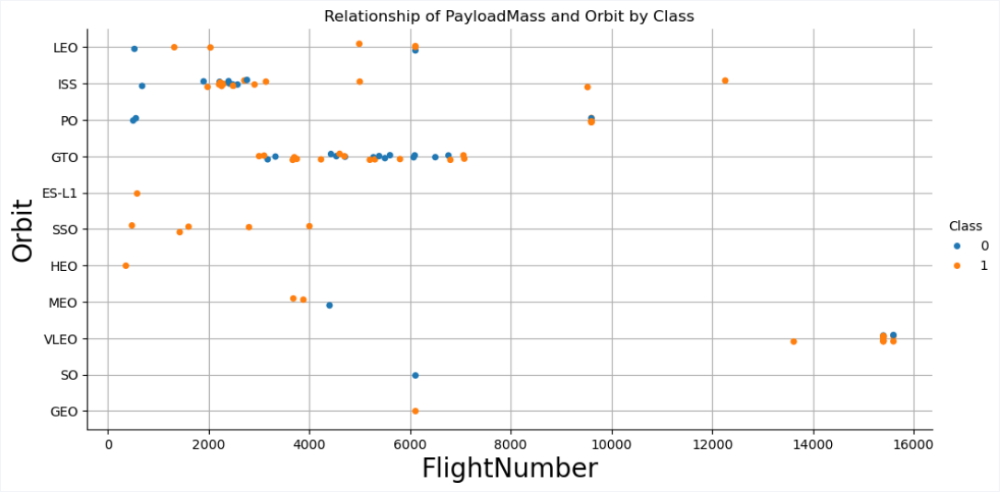
Results - Exploratory data analysis

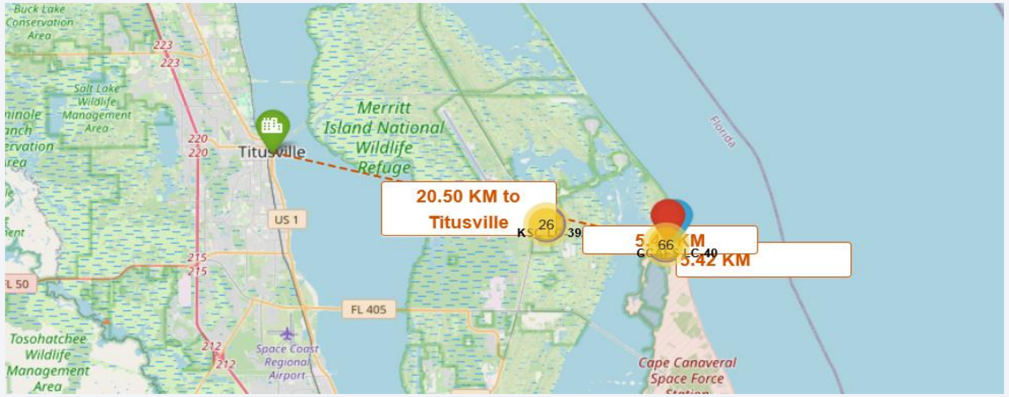
- Higher flight numbers improves success.
- Payload mass affects landing success.
- Launch sites vary in activity and payload capacity.
- Orbit type significantly impacts success rates.
- Historical trend shows improving success over time.

y data analysis results

Results - Interactive analytics







y data analysis results

Results - Predictive analysis

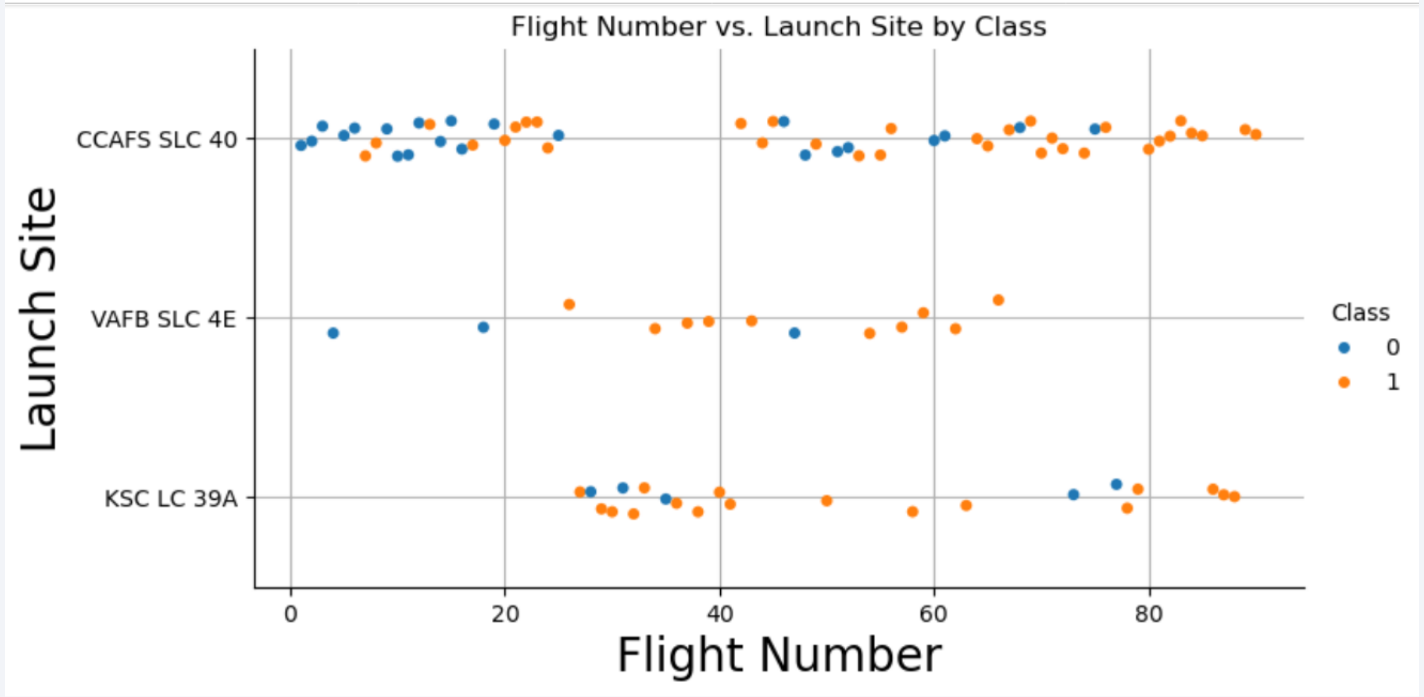
- The Decision Tree model outperforms other classifiers such as KNN, SVM, and Logistic Regression on this dataset, achieving the highest accuracy of approximately 83.3%. In general the model adapts well to small datasets.
- The classification report reveals that the model performs especially well in identifying successful launches (class 1), with a high precision (92.3%) and recall (85.7%). This means the model is effective at correctly predicting successful launches while minimizing false positives and false negatives.
- The result show the decision tree model can be a practical tool for forecasting SpaceX launch outcomes,
- Accurate prediction of launch success enables SpaceX and its clients to better manage mission risks, allocate resources efficiently, and reduce costly failures.
- The model provides actionable clues that engineers and analysts can explore further to identify and validate the true causes behind successful launches.

Section 2

Insights drawn from EDA

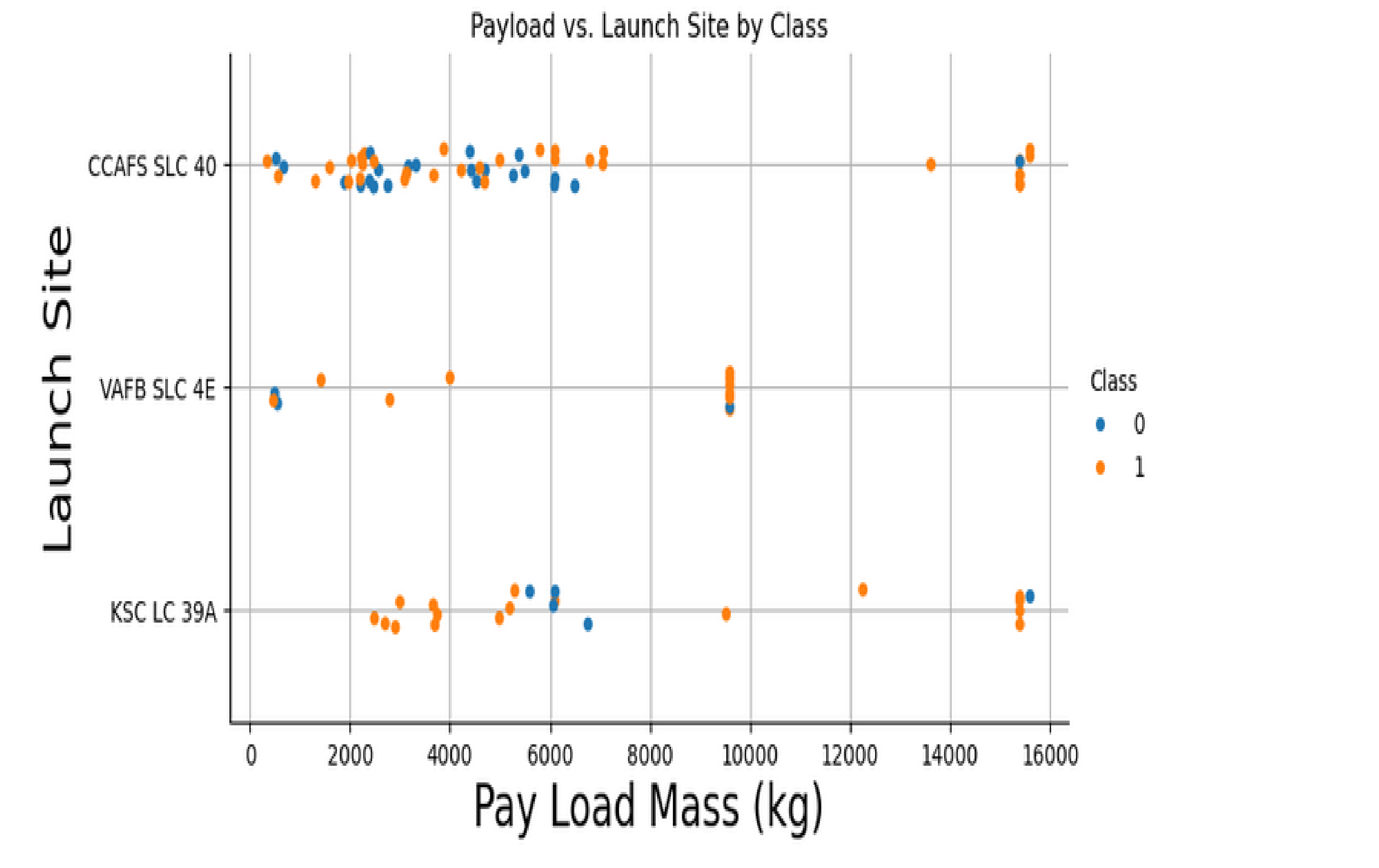
Flight Number vs. Launch Site

- As the flight number increases the successful launches are increasing, too.
- CCFA has the most launches. VAFB has the smallest number of launches. The number of successful launches are similar distributed over the launch sites. Even VAFB has more successful launches than failures.



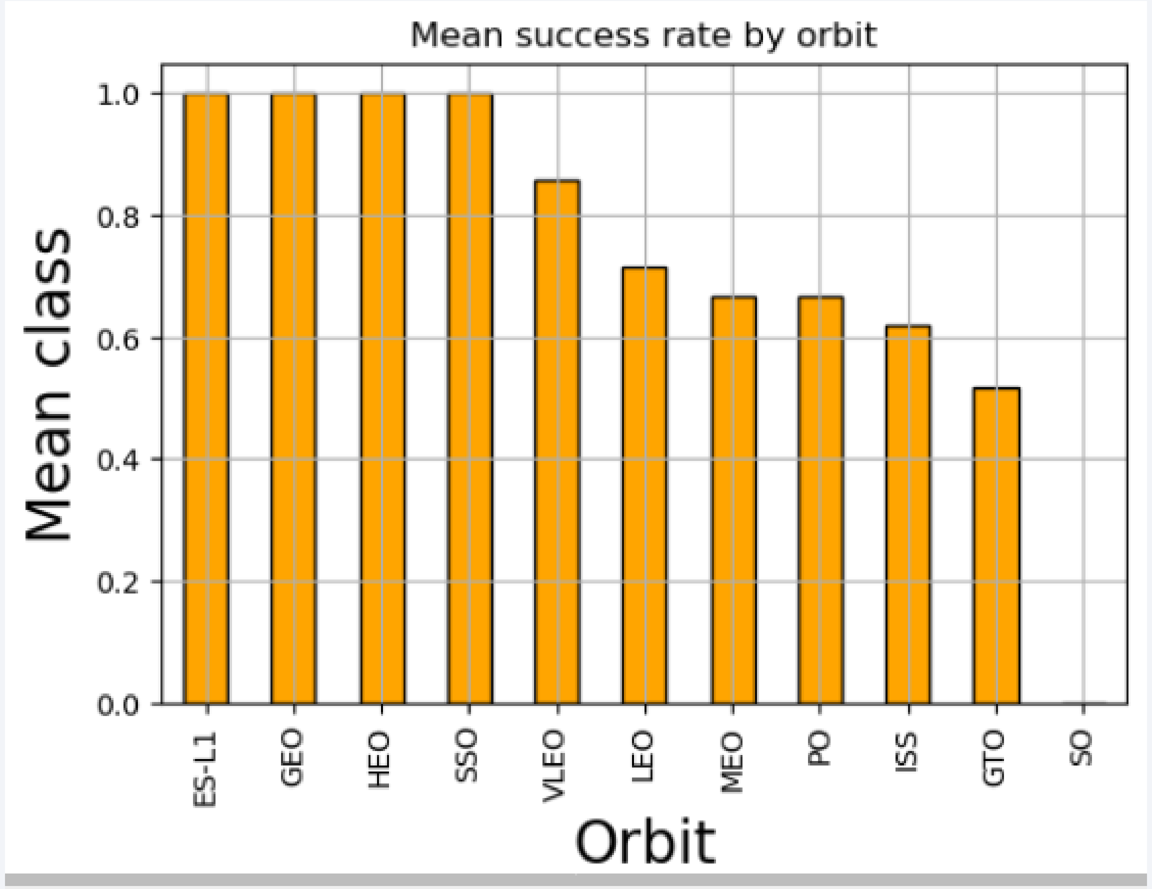
Payload vs. Launch Site

- Now if you observe Payload Mass Vs. Launch Site scatter point chart you will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).
- The other launch site have successfully launched payloads of nearly 16000 kg, while have each one failure there.
- Most launches are up to 6000 kg.



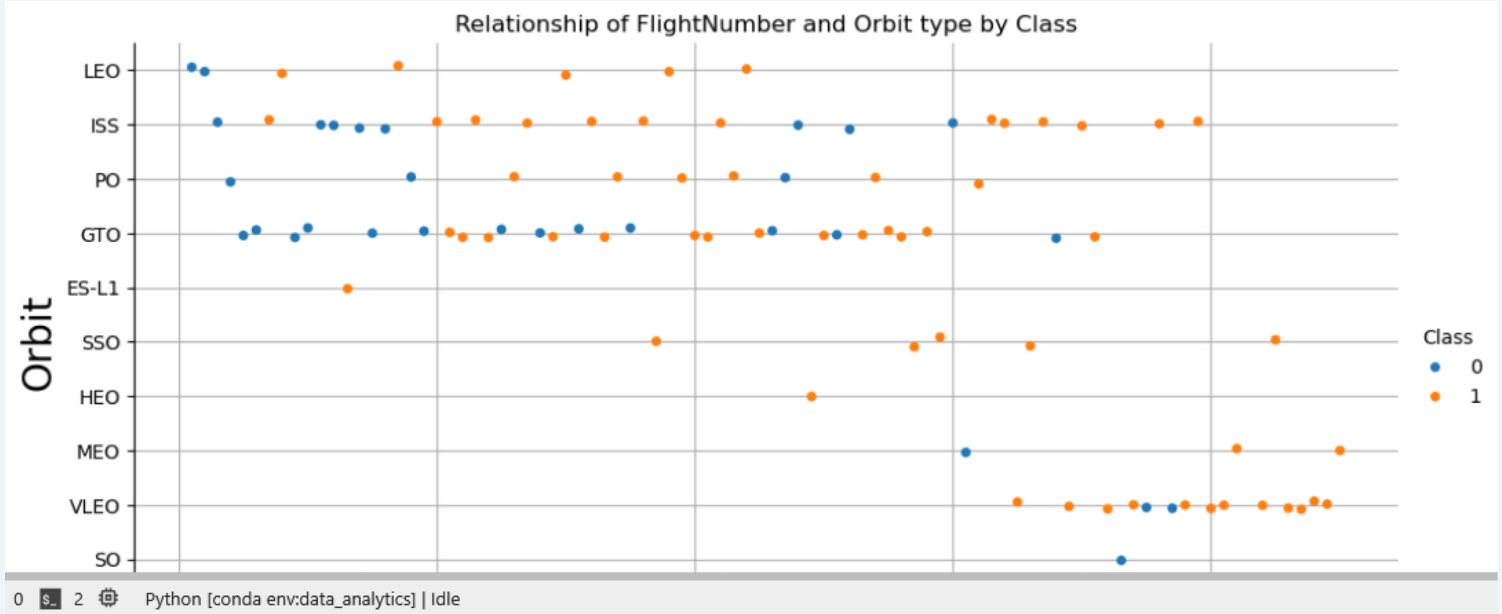
Success Rate vs. Orbit Type

- The ES-L1, GEO and HEO are the three highest success rates.
- The ISS and the GTO have low success rates.
- SO has got a success rate of 0.
- Space X should go with the sites having the highest success rates.



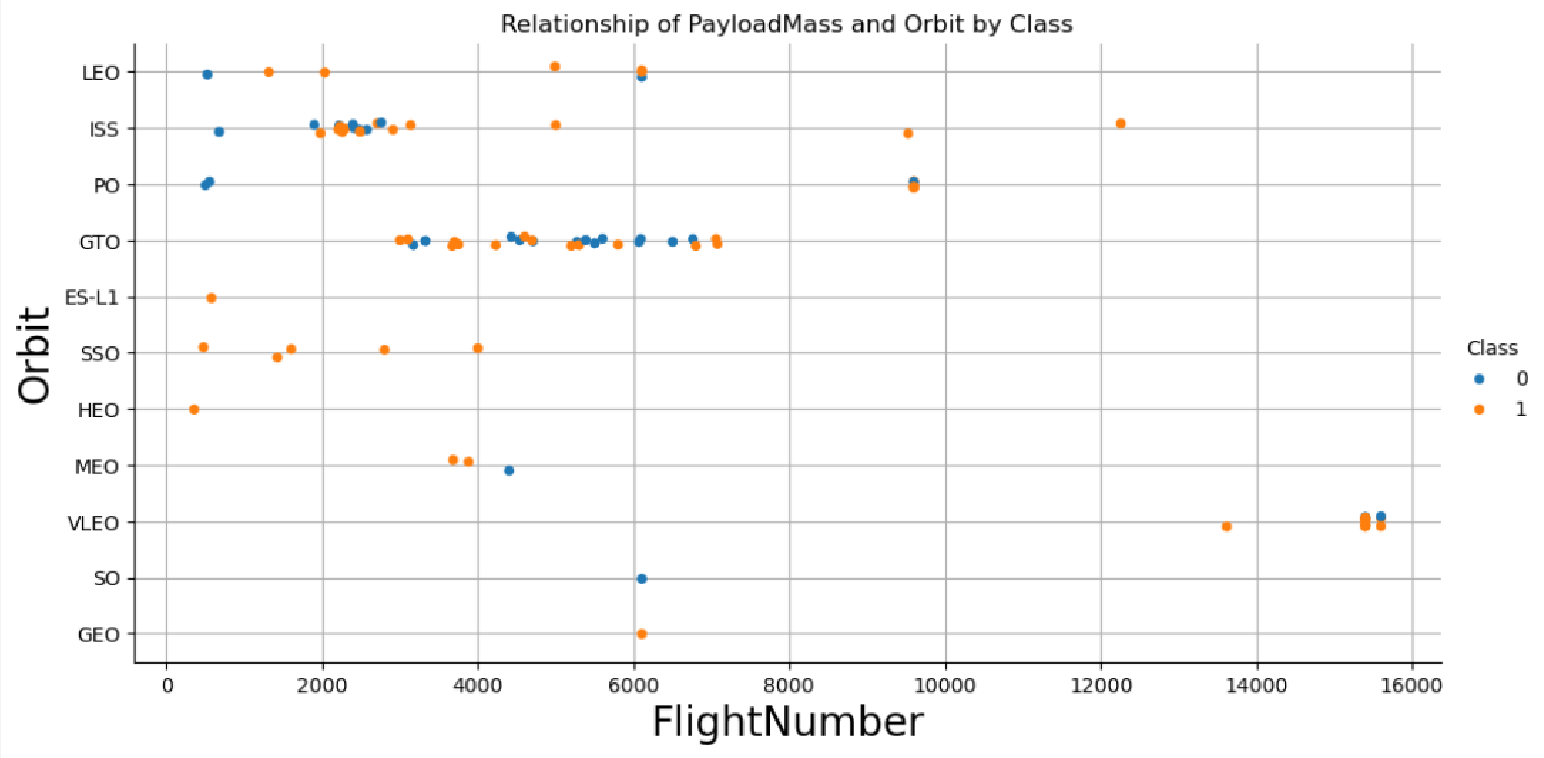
Flight Number vs. Orbit Type

- The LEO orbit has a high success rate but a low number of flights. Thesis: Is the number of successes increasing with the number of flights.
- Conversely, in the GTO orbit, there appears to be no relationship between flight number and success. MEO, VLEO were only used with high FlightNumbers.
- It seems with increasing FlighNumbers additional orbits were tested.
- The VLEO orbit has a high success rate an and was used at later flight numbers.
- The SSO has only successes but has not many FlightNumbers.
- LEO and GTO were used with low FlightNumbers.



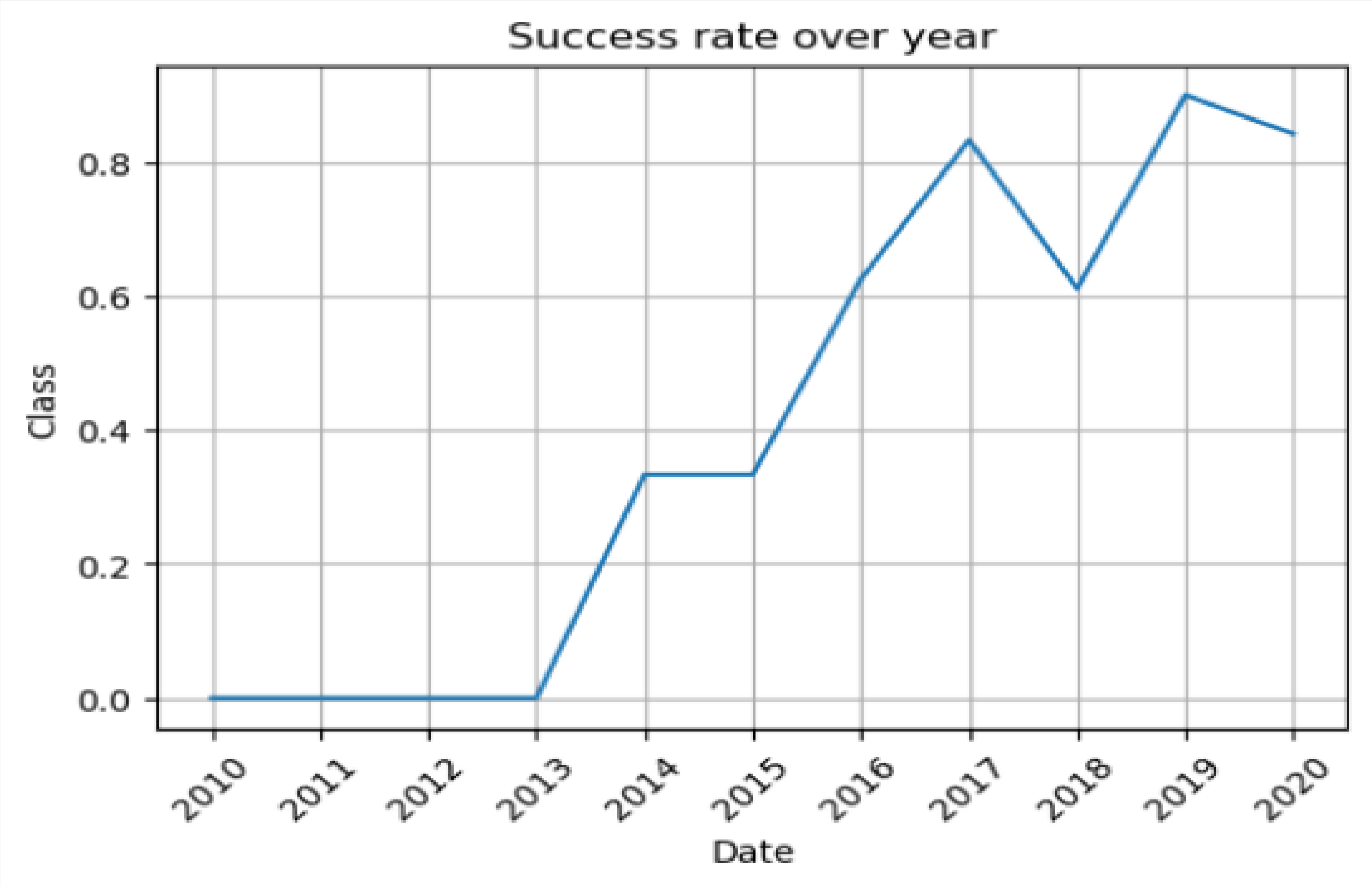
Payload vs. Orbit Type

- With heavy payloads the successful landing or positive landing rate is higher for Polar,LEO and ISS.
- However, for GTO, it's difficult to distinguish between successful and unsuccessful landings as both outcomes are present.
- Further analysis is necessary there.
- The MEO, VLEO and other orbits are less used at this point.



Launch Success Yearly Trend

- From 2010 to 2013 there were none successful launches.
- The sucess rate since 2013 kept increasing till 2020.
- In 2018 there was a setback.
- The peak was 2019 and to 2020 there is small drop.



All Launch Site Names

- pd.read_sql("SELECT DISTINCT Launch_Site FROM SPACEXTABLE", con)
- The unique sites ['CCAFS LC-40', 'VAFB SLC-4E', 'KSC LC-39A', 'CCAFS SLC-40']
- Knowing the sites allows to make comparisons with regard to success or failures. Below are the number of launches per site derived.

Launch_Site	
CCAFS SLC-40	34
CCAFS LC-40	26
KSC LC-39A	25
VAFB SLC-4E	16

	Launch_Site
0	CCAFS LC-40
1	VAFB SLC-4E
2	KSC LC-39A
3	CCAFS SLC-40

Launch Site Names Begin with 'CCA'

query_2 = \

"SELECT Launch_Site FROM SPACEXTABLE WHERE Launch_Site LIKE 'CCA%' LIMIT 5"

There are two sites:
CCAFS LC-40, CCAFS SLC-40

What does the launch sites have in common?

	Launch_Site
0	CCAFS LC-40
1	CCAFS LC-40
2	CCAFS LC-40
3	CCAFS LC-40
4	CCAFS LC-40

Total Payload Mass

The Total Payload Mass for the customer NASA is 48213 kg.

Several organizations of NASA are customers.
It is necessary to summarize all NASA departments to calculate the total payload.

```
query_3 = \
"SELECT SUM(PAYLOAD_MASS__KG_) AS TotalPayloadMass_Nasa FROM SPACEXTABLE WHERE Customer LIKE '%NASA (CRS)%' "
result_3 = pd.read_sql(query_3, con)

print(result_3)
```

```
   TotalPayloadMass_Nasa
0                  48213
```

Average Payload Mass by F9 v1.1

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

The mass is 2928.4 kg.

```
query_4 = "SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTABLE WHERE Booster_Version LIKE 'F9 v1.1'"
result_4 = pd.read_sql(query_4, con)
print(result_4)
```

```
   AVG(PAYLOAD_MASS__KG_)
0                  2928.4
```

First Successful Ground Landing Date

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

The the first successful landing outcome on ground pad was: 2015-12-22.

It highlights SpaceX's progress in achieving reusable rocket technology, reducing costs, and advancing space exploration.

```
query_5 = """SELECT MIN(Date) FROM SPACEXTABLE
WHERE Landing_Outcome LIKE 'Success (ground pad)'"""
result_5 = pd.read_sql(query_5, con)
print(result_5)
```

```
    MIN(Date)
0  2015-12-22
```

Successful Drone Ship Landing with Payload between 4000 and 6000

It demonstrates SpaceX's success in landing rockets on drone ships under specific payload conditions, which are challenging because of the additional precision required at sea.

This list is a testament to SpaceX's engineering prowess and its achievements in reusability and precision landing technology.

```
query_6 = """
SELECT Booster_Version FROM SPACEXTABLE
WHERE Landing_Outcome LIKE 'Success (drone ship)'
AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000"""
result_6 = pd.read_sql(query_6, con)
print(result_6)
```

```
  Booster_Version
0     F9 FT B1022
1     F9 FT B1026
2  F9 FT  B1021.2
3  F9 FT  B1031.2
```

Total Number of Successful and Failure Mission Outcomes

Provides a high-level view of SpaceX's mission success rate, crucial for assessing performance and reliability.

Highlights failure counts, which may point to areas for improvement or investigation.

```
query_7a = """
SELECT
    CASE
        WHEN Mission_Outcome LIKE '%Success%' THEN 'Success'
        WHEN Mission_Outcome LIKE '%Failure%' THEN 'Failure'
        ELSE 'Error'
    END AS success_or_failure,
    COUNT(*) AS Total_Count
FROM SPACEXTABLE
GROUP BY success_or_failure;
"""
result_7a = pd.read_sql(query_7a, con)
print(result_7a)
```

```
  success_or_failure  Total_Count
0            Failure            1
1            Success          100
```

Boosters Carried Maximum Payload

Payload capacity is a key metric for any rocket. This result shows which boosters performed the most challenging missions, carrying the heaviest payloads SpaceX can handle.

The result highlights SpaceX's leading position in payload capacity, cost reduction through reusability, and the operational reliability of Falcon 9 Block 5 boosters. This strengthens their market competitiveness and supports their mission of making space exploration more accessible and affordable.

```
query_8 = """
SELECT
Booster_Version
FROM SPACEXTABLE
WHERE PAYLOAD_MASS__KG_ =
(SELECT MAX(PAYLOAD_MASS__KG_ )
FROM SPACEXTABLE)
;
"""
result_8 = pd.read_sql(query_8, con)
print(result_8)
```

```
   Booster_Version
0    F9 B5 B1048.4
1    F9 B5 B1049.4
2    F9 B5 B1051.3
3    F9 B5 B1056.4
4    F9 B5 B1048.5
5    F9 B5 B1051.4
6    F9 B5 B1049.5
7   F9 B5 B1060.2
8   F9 B5 B1058.3
9    F9 B5 B1051.6
10   F9 B5 B1060.3
11  F9 B5 B1049.7
```

2015 Launch Records

Knowing failures in drone ship landings during a specific year (2015) is crucial for identifying patterns or factors contributing to landing failures.

Early missions in 2015 might have faced technological or operational challenges as SpaceX was perfecting its drone ship landing technology.

```
query_9 = """
SELECT Booster_Version,  substr(Date, 6,2) as month,
substr(Date, 1, 4) as year, Launch_Site
FROM SPACEXTABLE
WHERE Landing_Outcome LIKE 'Failure (drone ship)'
AND
substr(Date, 1, 4) = '2015'
;
"""
result_9 = pd.read_sql(query_9, con)
print(result_9)
```

```
  Booster_Version month  year  Launch_Site
0   F9 v1.1 B1012    01  2015  CCAFS LC-40
1   F9 v1.1 B1015    04  2015  CCAFS LC-40
```

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

The relatively low success count (8) and failure count (7) reflect that this was the early era of SpaceX's landing attempts (before 2017).

This data highlights the challenges SpaceX faced and the milestones achieved in refining their technology.

```
query_9 = """
SELECT

CASE
WHEN Landing_Outcome LIKE "%Success%" THEN "success"
WHEN Landing_Outcome LIKE "%Failure%" THEN "failure"
ELSE
"Other"
END AS result,
COUNT(Landing_Outcome) Total_Outcome

FROM SPACEXTABLE
WHERE Date BETWEEN "2010-06-04" AND "2017-03-20"
GROUP BY result
ORDER BY Total_Outcome DESC


;
"""
result_9 = pd.read_sql(query_9, con)
print(result_9)
```

```
    result  Total_Outcome
0    Other             16
1  success              8
2  failure              7
```

Section 3

Launch Sites Proximities Analysis

Map include all launch sites location markers on a global map

The map highlights the geographical distribution of SpaceX's launch sites.

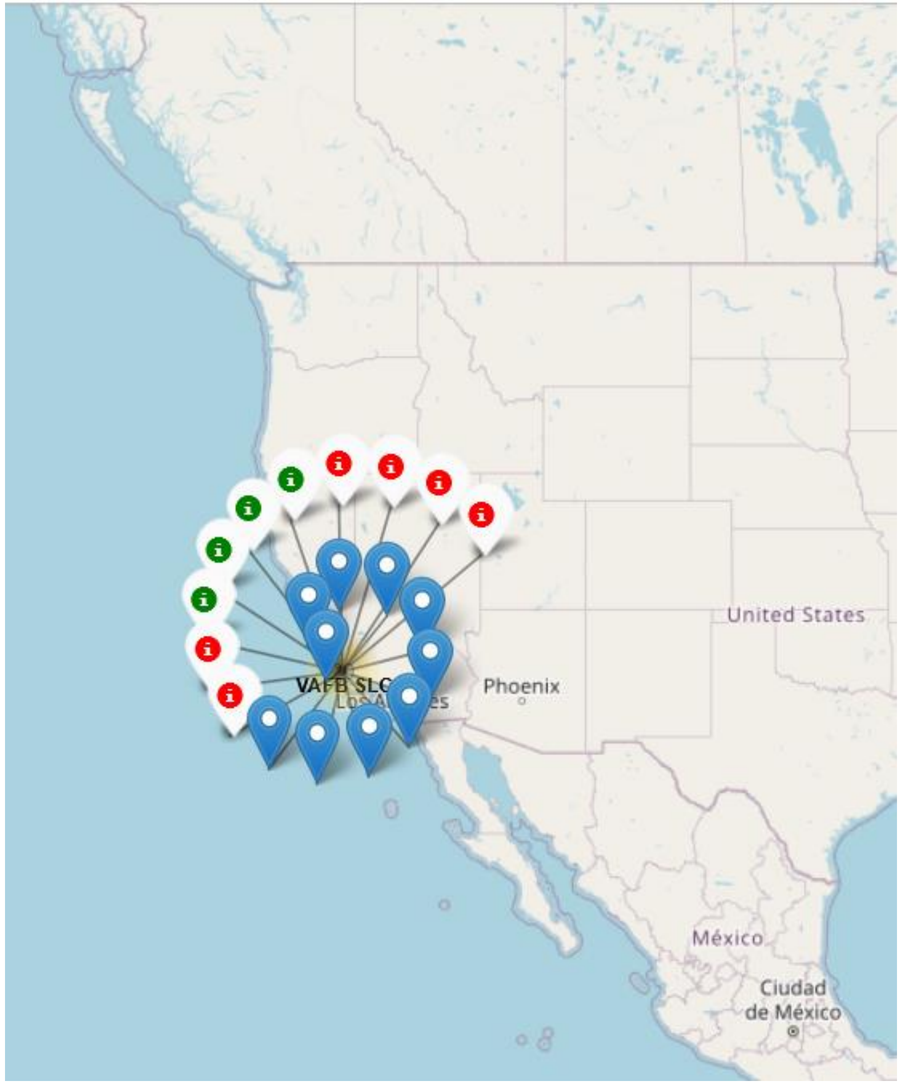
Key launch sites visible include CCAFS LC-40, CCAFS SLC-40 (Space Launch Complex 40 at Cape Canaveral), KSC LC-39A (Kennedy Space Center, Launch Complex 39A)

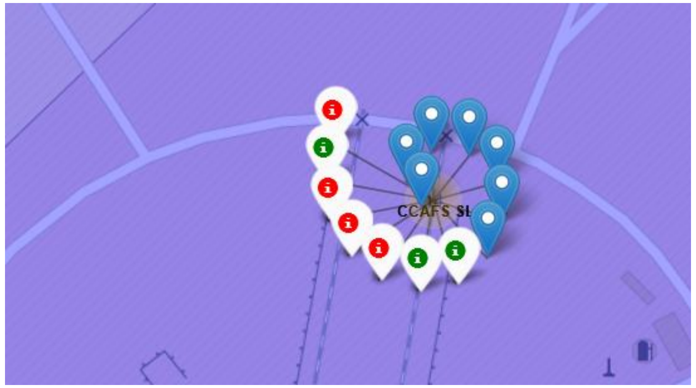
Most launch sites are clustered in Florida, indicating the region's strategic importance for space launches due to its proximity to the equator and large open waters for safety.

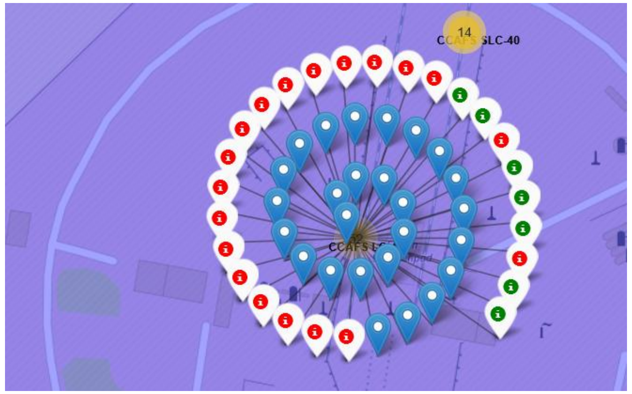
Map to show the color-labeled launch outcomes on the map

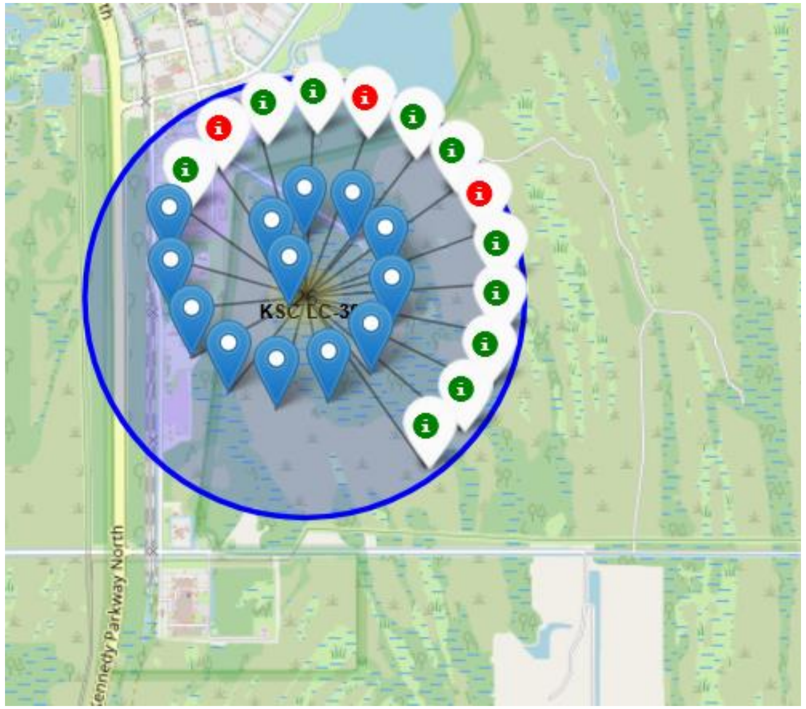
- The marker_color field dynamically assigns colors to markers based on launch outcomes in the SpaceX DataFrame.
- Success: green successful launch outcomes. / Failure: red failed launch outcomes.
- Clusters of success markers highlight consistently successful launch sites mostly in Florida.
- It helps SpaceX pinpoint areas for technical improvements or policy adjustments.
- Patterns in failure-prone areas might guide further investigations into environmental, technical, or logistical factors.

Map include all launch sites location markers on a global map





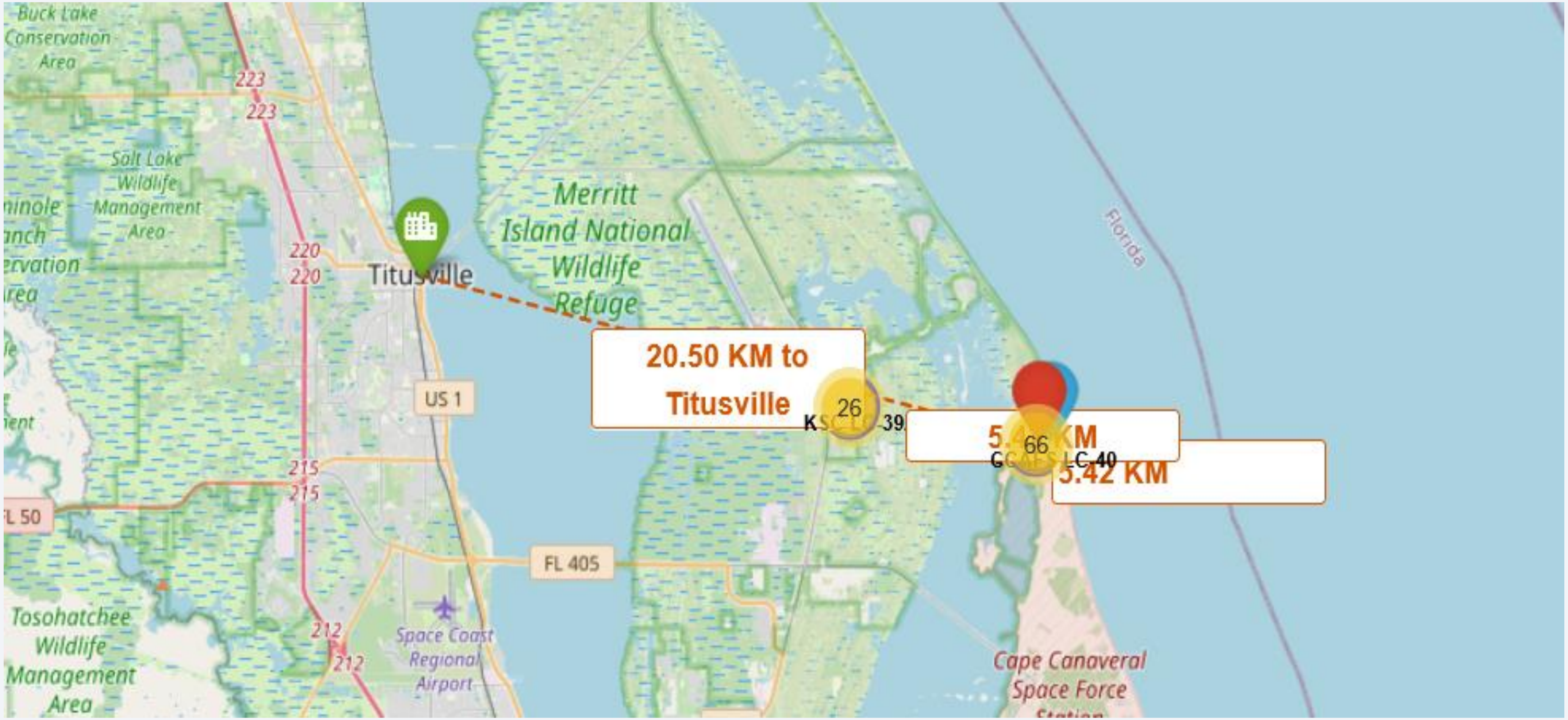




Map showing proximities such as railway, highway, coastline, with distance calculated and displayed

- The interactive map is created using Folium and allows users to explore distances and relationships between SpaceX launch sites and nearby points of interest, in this case Titusville Florida
- A dashed Polyline connects Titusville and the SpaceX launch site.
- A custom label is placed at the midpoint of the Polyline, showing the calculated distance between Titusville and the launch site (20.5 km in this example).
- Titusville Marker: Green marker with a "city" icon, representing Titusville, FL.
- Launch Site Marker: Red marker with a "rocket" icon, denoting the SpaceX launch site.
- The map emphasizes the close distance between Titusville and the SpaceX launch site (approximately 20.5 km). This proximity underlines the strategic location of Titusville as a support hub for SpaceX operations..

Map showing proximities SpaceX and Titusville

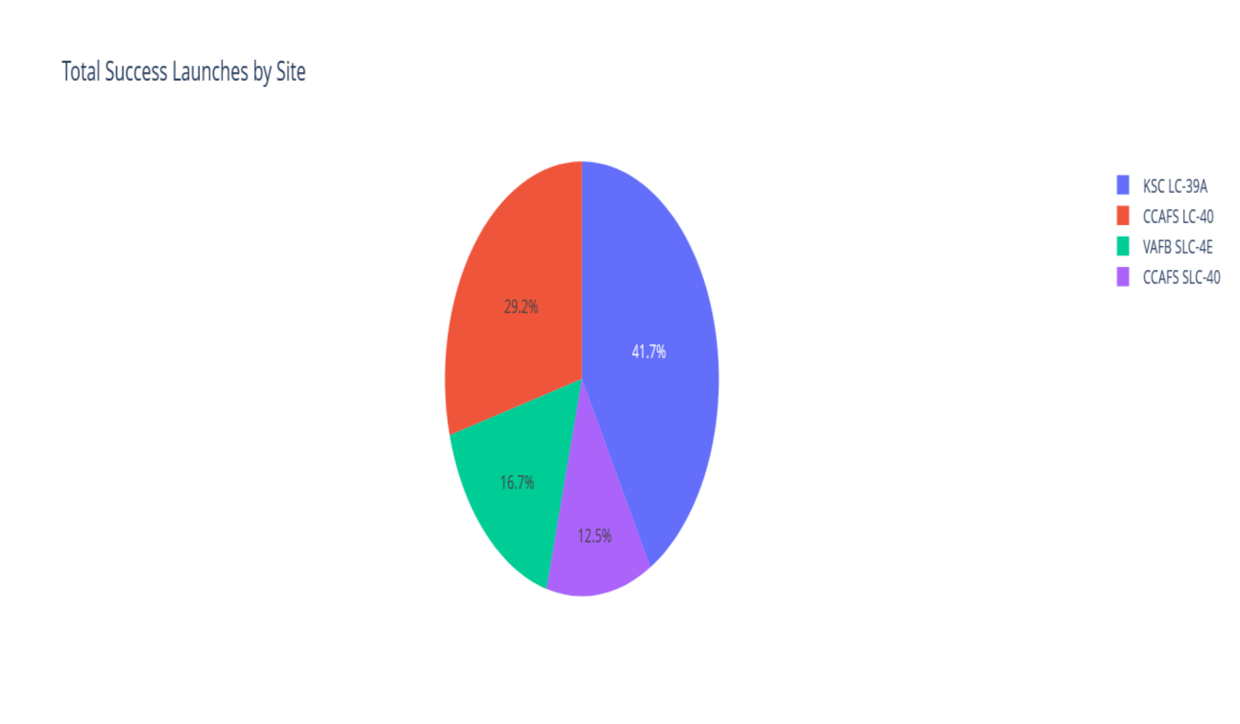


Section 4

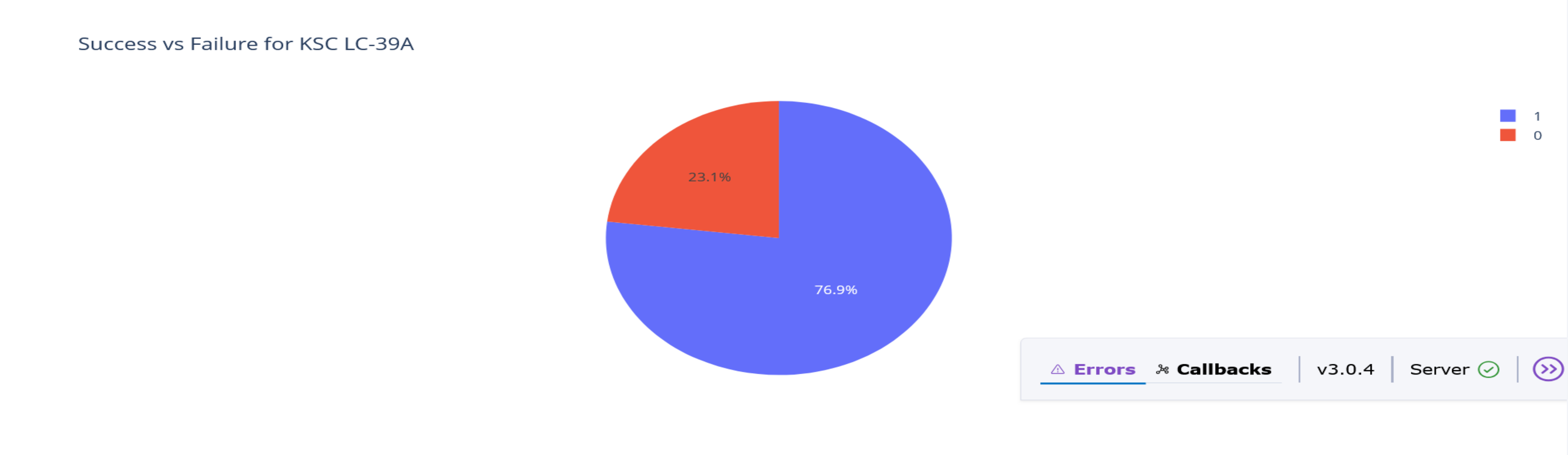
Build a Dashboard with Plotly Dash

Graph of launch success count for all sites in a piechart

- KSC LC-39A's Dominance Handling 41.7% of successful launches signifies its importance as SpaceX's flagship site, capable of supporting diverse and high-profile missions.
- If KSC LC-39A has fewer total launches compared to other sites, the success rate percentage might overstate the importance.
- Cape Canaveral Air Force Station's LC-40 (29.2%) is the second-most significant contributor.
- Supports a substantial portion of SpaceX's launch operations, potentially focusing on commercial satellite launches or medium-scale payloads.
- VAFB share indicates a more limited role, possibly supporting specific mission types or a smaller number of launches.
- The success rate of the CCAF sites could be combined. Since both sites are geographically close and managed under the same larger facility, treating them as one entity could better represent the region's overall contribution to SpaceX operations.

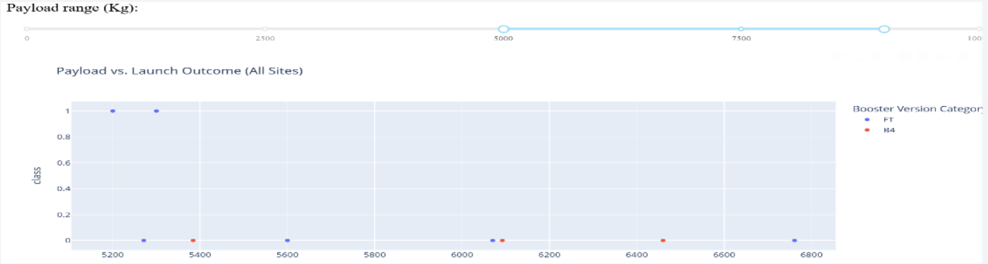


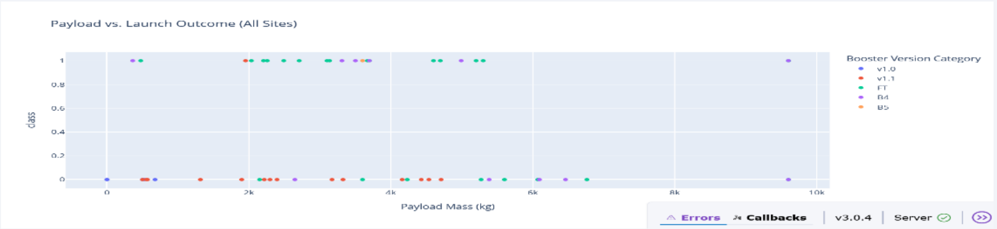
Graph of site with the highest success rate

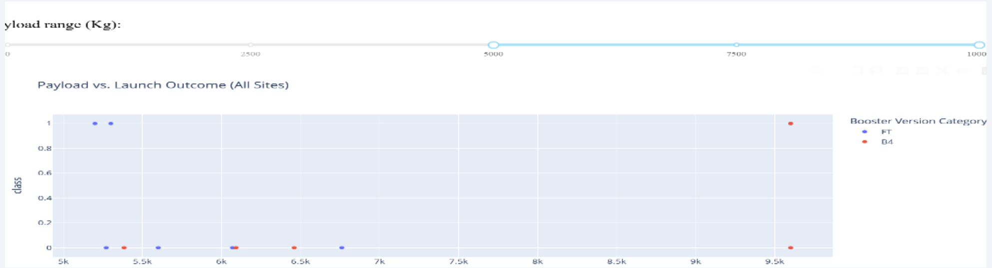


- The graph indicates that "KSC LC-39A" has a success rate of "76.9%". However, the percentage alone doesn't tell the full story. If KSC LC-39A has fewer total launches compared to other sites, the success rate percentage might overstate its operational reliability. Also the CCAF launch sites. This impairs the percentages too.
- Given this caveat KSC LC-39A may have optimized infrastructure and processes, allowing it to achieve higher success rates.
- This site might have been used for higher-priority or carefully planned missions, which could explain the higher success rate.
- A high success rate suggests that KSC LC-39A is a reliable site for SpaceX’s operations, likely benefiting from extensive testing and lessons learned from past launches.

Graph- Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider







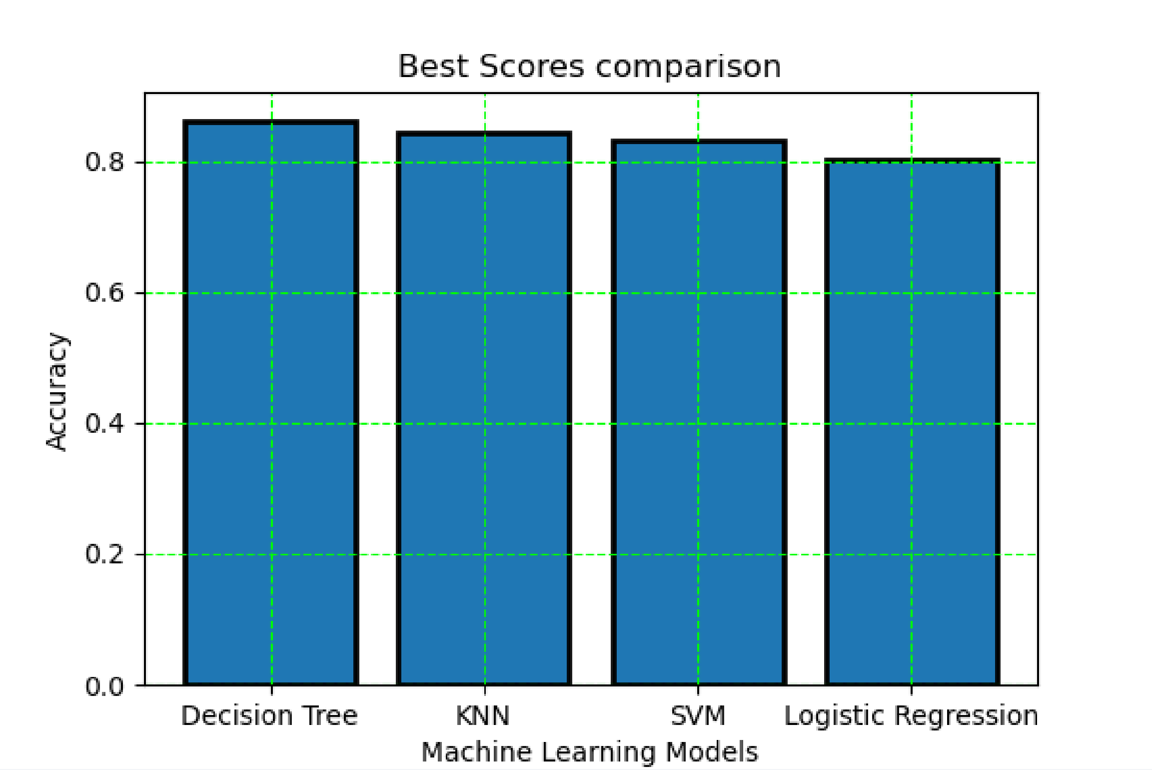
- The payload scatter chart is a visualization that explores the relationship between payload mass and launch success outcomes, categorized by booster versions
- Lower Payload Mass might show higher success rates as they are less demanding on the rocket's capacity. The challenge for SpaceX is to be successful at higher payloads.
- Each booster version might exhibit specific payload mass ranges and associated success rates:
- For instance, "F9 v1.1" shows an average payload mass of \~2928 kg, suggesting its design is optimized for medium payloads.
- Booster versions like "F9 B5 (various iterations)" are associated with the highest payload masses, showcasing SpaceX's technological progress in launching and reusing rockets with significant capacity.
- By filtering payload ranges with the slider users can focus on specific operational windows, such as low-payload missions (<2000 kg) or high-payload launches (>6000 kg), and study trends in success outcomes.
- The transition from older booster versions to newer ones (e.g., F9 v1.1 → F9 FT → F9 B5) reflects SpaceX's increasing ability to handle heavier payloads with higher success rates.

Section 5

Predictive Analysis (Classification)

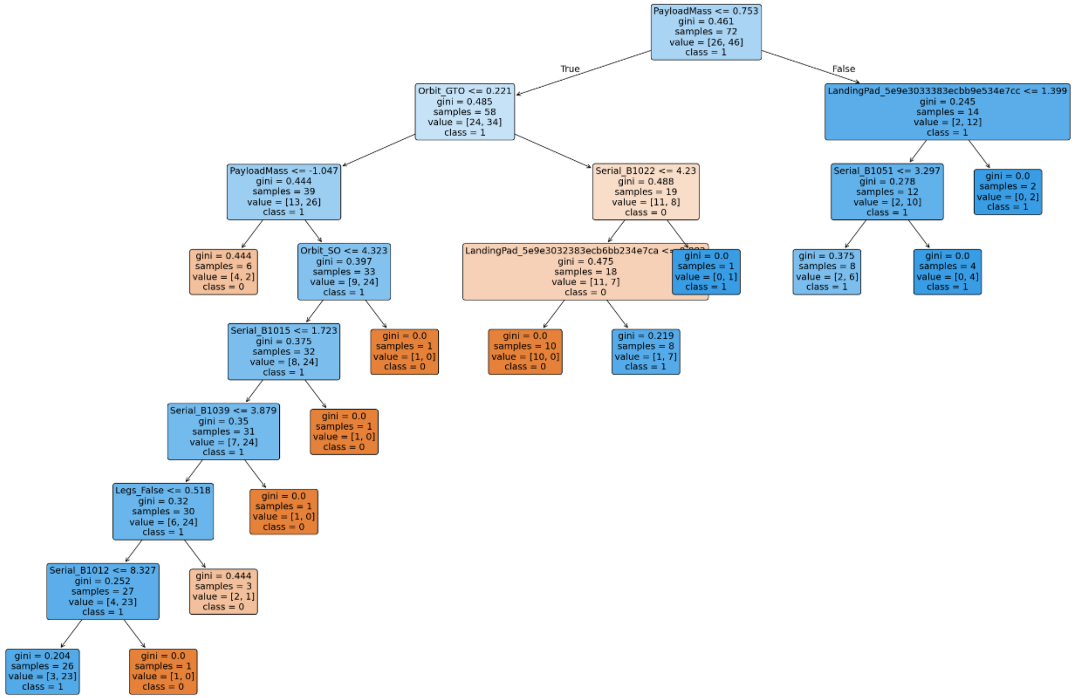
Classification Accuracy

- The decision tree model has the best accuracy score.
- Given that the dataset is small, the decision tree's ability to adapt to specific non-linear patterns likely contributes to its high accuracy. Its simplicity ensures that the model doesn't overfit while still capturing essential relationships in the data.

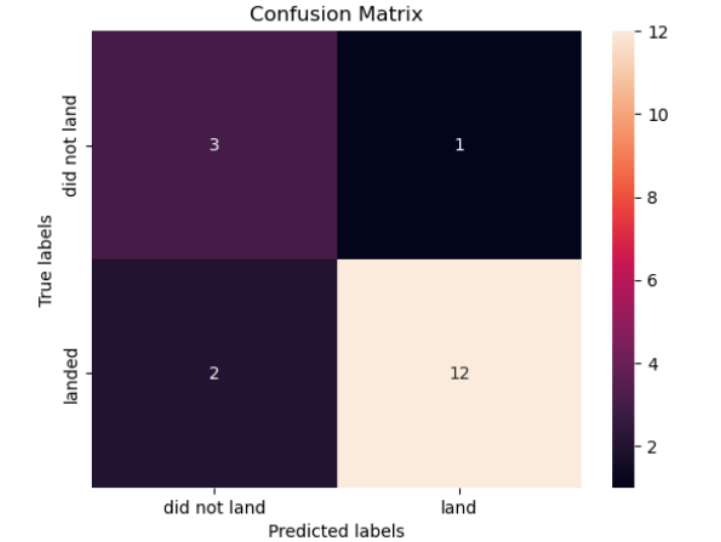


Classification Accuracy

- This image shows the plot of the decision tree model.
- The root node is PayloadMass.
- The first split is between Orbit and LaunchingPad.
- The Orbit branch is developed further.



Confusion Matrix



Classification report of decision tree model

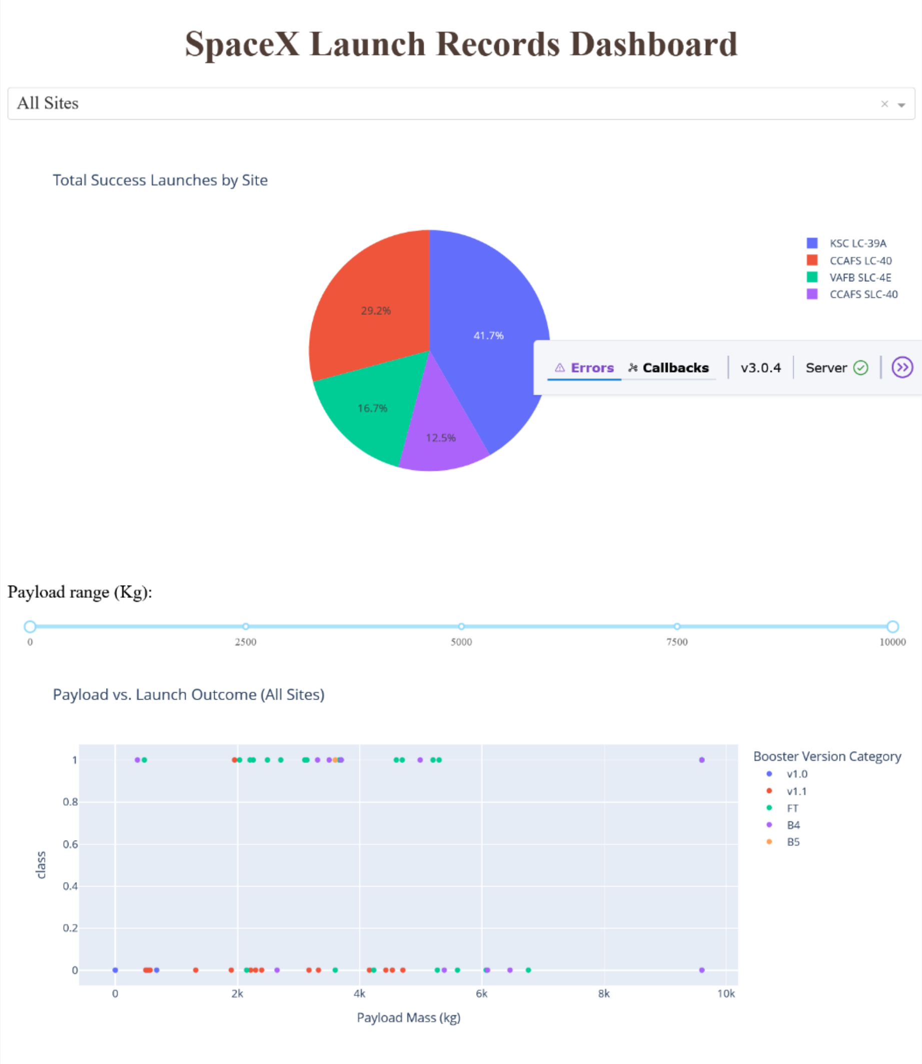
	precision	recall	f1-score	support
0	0.500000	0.750000	0.600000	4.000000
1	0.916667	0.785714	0.846154	14.000000
accuracy	0.777778	0.777778	0.777778	0.777778
macro avg	0.708333	0.767857	0.723077	18.000000
weighted avg	0.824074	0.777778	0.791453	18.000000

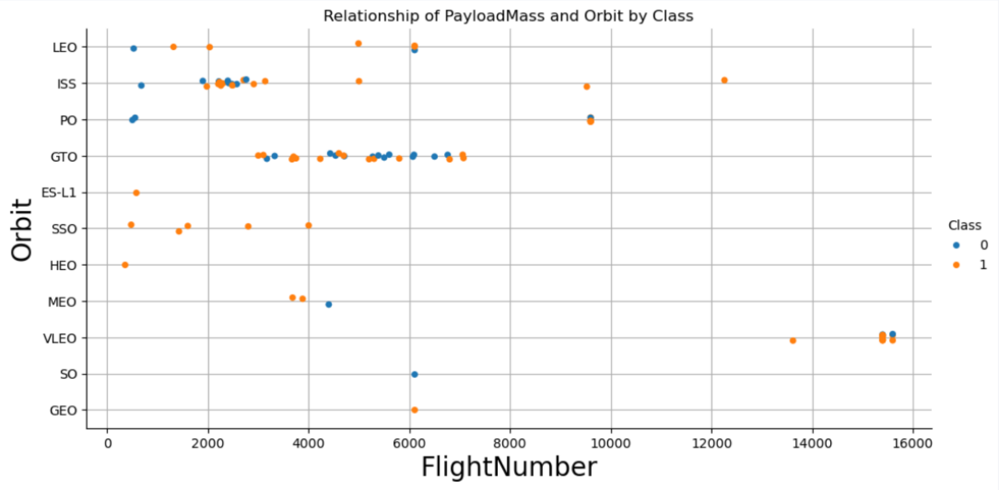
- The model does well predicting class 1 (12 correct vs 2 missed).
- It performs moderately on class 0 (3 correct vs 1 incorrect).
- Overall, it has more correct predictions than mistakes, but some confusion exists especially for class 1 being predicted as 0 (false negatives).

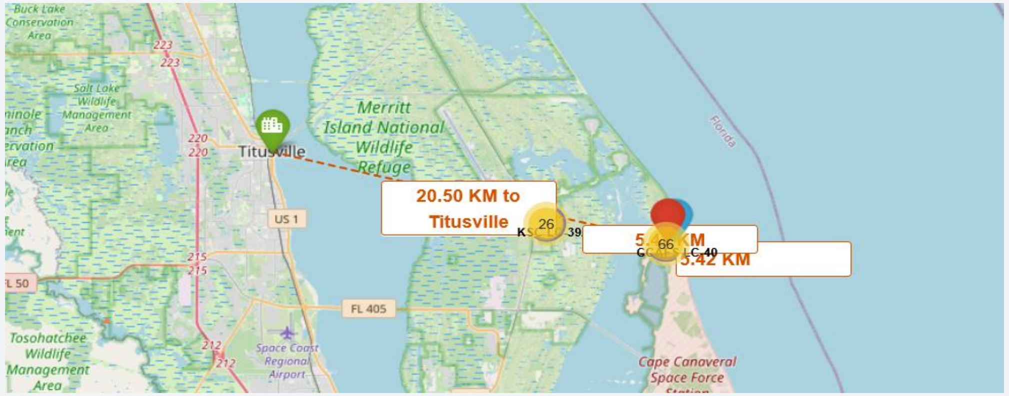
Conclusions

- The Decision Tree model outperforms other classifiers such as KNN, SVM, and Logistic Regression on this dataset, achieving the highest accuracy of approximately 83.3%. In general the model adapts well to small datasets.
- The classification report reveals that the model performs especially well in identifying successful launches (class 1), with a high precision (92.3%) and recall (85.7%). This means the model is effective at correctly predicting successful launches while minimizing false positives and false negatives.
- The result show the decision tree model can be a practical tool for forecasting SpaceX launch outcomes,
- Accurate prediction of launch success enables SpaceX and its clients to better manage mission risks, allocate resources efficiently, and reduce costly failures.
- The model provides actionable clues that engineers and analysts can explore further to identify and validate the true causes behind successful launches.

Appendix


SpaceX Launch Records Dashboard
All Sites
Total Success Launches by Site
KSC LC-39A
CCAFS LC-40
VAFB SLC-4E
CCAFS SLC-40
41.7%
29.2%
16.7%
12.5%
Errors
Callbacks
v3.0.4
Server
Payload range (Kg):
0
2500
5000
7500
10000
Payload vs. Launch Outcome (All Sites)
Booster Version Category
v1.0
v1.1
FT
B4
B5
class
Payload Mass (kg)


Relationship of PayloadMass and Orbit by Class
Orbit
LEO
ISS
PO
GTO
ES-L1
SSO
HEO
MEO
VLEO
SO
GEO
FlightNumber
Class
0
1


Merritt Island National Wildlife Refuge
Titusville
20.50 KM to Titusville
5.42 KM
Cape Canaveral Space Force Station

Thank you!