IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Griffin Ulsh
November, 2024

Outline

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

Executive Summary

Methodology Overview:

- Collect, clean, and prepare Falcon9 launch data.
- Exploratory analysis of the Falcon9 launch data.
- Visualize the data in an interactive dashboard.
- Predictive model development, accuracy evaluation, and results analysis.

Summary of Results:

- Key trends and variables identified in Exploratory Analysis
- Visual insights on first-stage success factors
- Predictive Model accuracy in forecasting launch outcomes

Introduction

Project Background and Context:

SpaceX leads in the commercial space race with competitive launch costs, advertising $62M per launch compared to other providers' $165M+. This affordability stems from SpaceX's reusable first-stage rockets. Predicting the success of these first stages can further refine cost assessments for launches."

Key Questions:

- What factors determine a successful SpaceX rocket landing?
- Which variables most influence first-stage success rates?
- How accurately can we estimate launch costs by predicting first-stage success?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
    - Data collected using Python's Request library, making HTTP requests/SpaceX's REST API
    - Web scraping from Wikipedia using BeautifulSoup
- Perform data wrangling
    - Evaluated all orbits and outcomes, used one-hot-encoding to create a new column in our dataframe labeling each launch as either successful or failed.
- 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 standardization, splitting data into training and testing sets, tuning our hyperparameters (SKL GridSearchCV), model deployment and evaluation (logistic regression, support vector machine, decision tree, KNN).

Data Collection

We collected data from 2 sources:

- SpaceX API
- Wikipedia article containing past launch data

We needed to filter specifically for Falcon9 launch data and load our 2 dataframes into one CSV we could work with in later steps.

SpaceX REST API

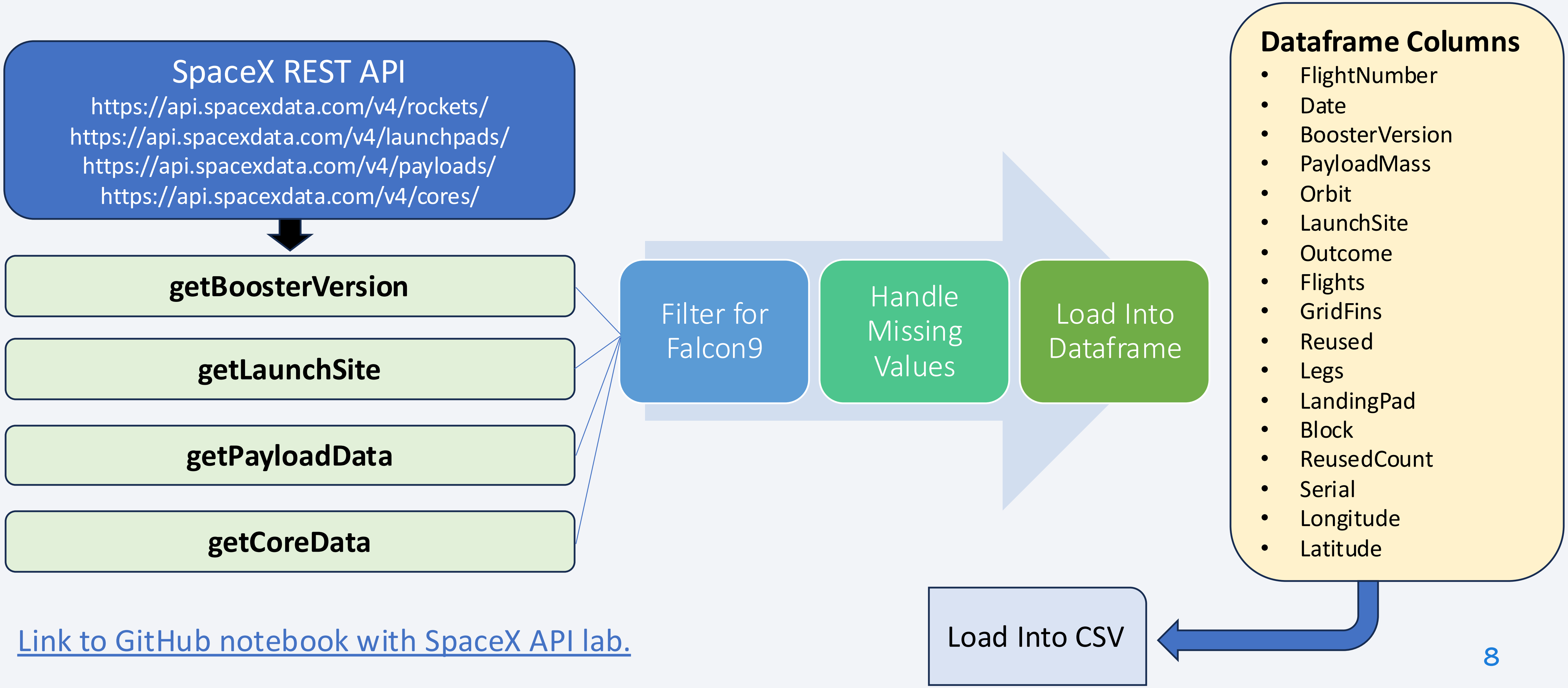
1. **Setup**: Configured SpaceX API and retrieved data.
2. **Extraction**: Defined functions to pull key launch details.
3. **Dataset Build**: Structured data into a DataFrame.
4. **Filter**: Kept only Falcon 9 launches.
5. **Clean**: Accounted for missing values.

SpaceX Wikipedia Article

1. **Setup**: Loaded required packages and helper functions.
2. **Request**: Accessed the static Wikipedia page with Falcon 9 launch data.
3. **Extract Headers**: Parsed table headers for column names.
4. **Parse Data**: Scraped and cleaned each launch record.
5. **Build DataFrame**: Structured parsed data into a DataFrame.

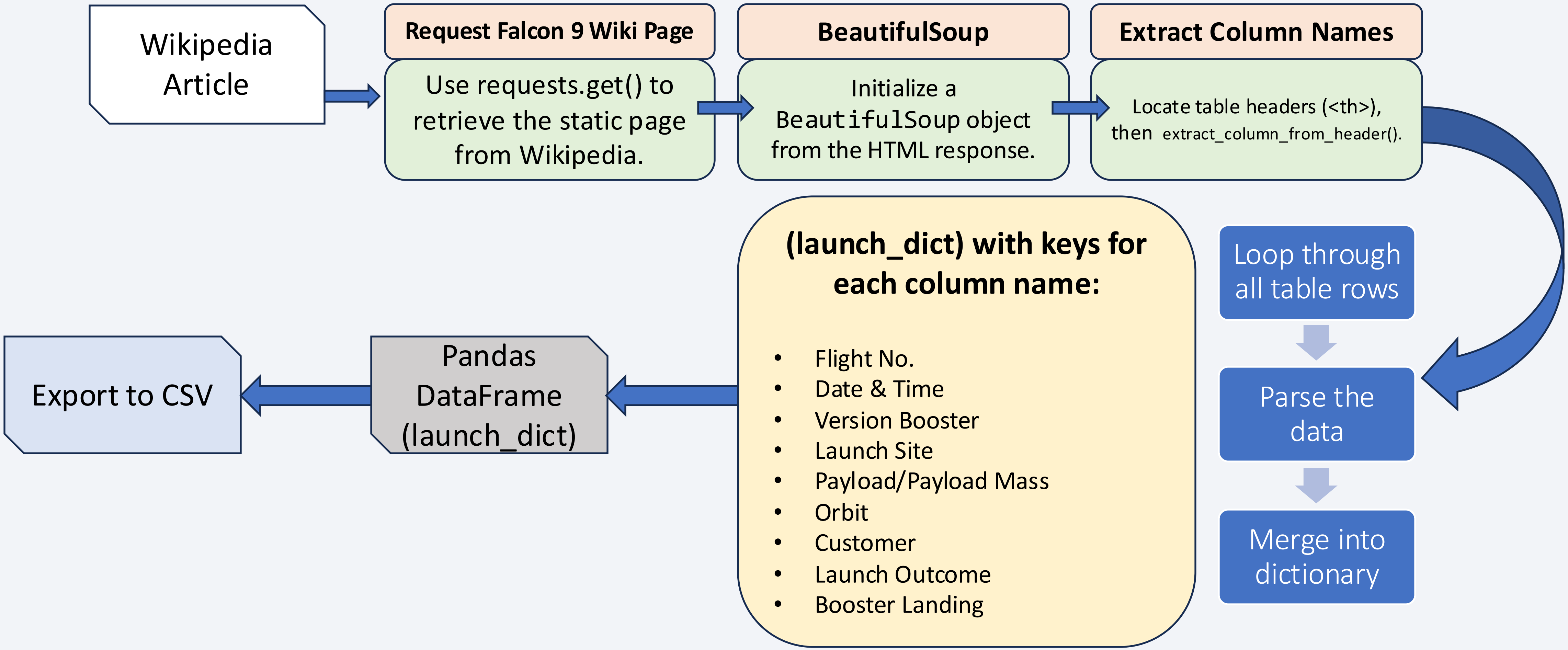
Export to CSV for future steps.

Data Collection – SpaceX API



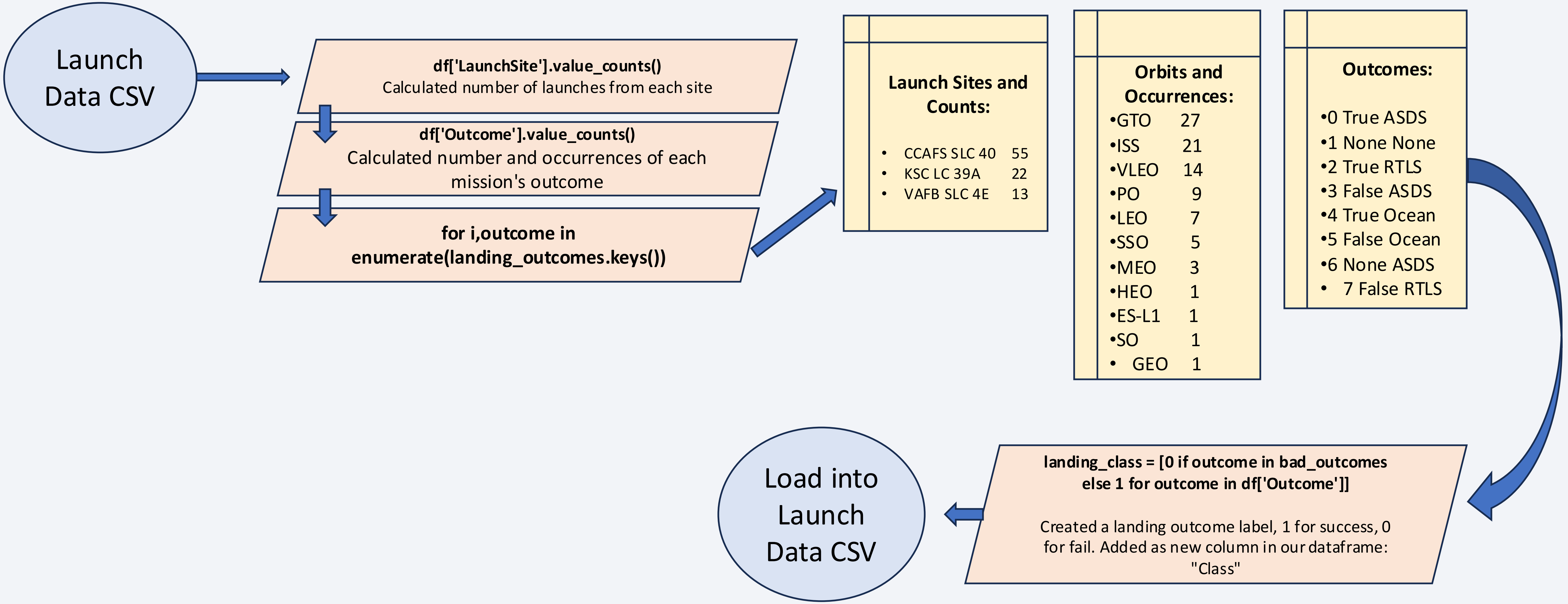
[Link to GitHub notebook with SpaceX API lab.]()

Data Collection - Scraping



[Link to GitHub notebook with Webscraping lab.]()

Data Wrangling



[Link to GitHub notebook with Data Wrangling Lab]

EDA with Data Visualization

Types of Data Visualizations Used and Why

•**Scatter Plots**

•Used to explore relationships between key variables in the dataset.

•Included hue="Class" to differentiate between successful and failed launches.

•Visualized the following comparisons:

- **Payload Mass vs. Flight Number**: To track trends in payload mass over time and its impact on success.
- **Launch Site vs. Flight Number**: To analyze site usage and success rates over time.
- **Payload Mass vs. Launch Site**: To assess the correlation between payload mass and launch site performance.
- **Orbit vs. Flight Number**: To observe orbital destinations over time and their success rates.
- **Orbit vs. Payload Mass**: To study how payload mass varies by orbital destination and its success outcomes.

Bar Chart

- **Orbit vs. Success Rate:** Used a bar chart to show the average success rate for each orbit. This provided a clear summary of which orbital types were most reliably associated with successful launches.

Line Chart

- **Year vs. Success Rate:** Used a line chart to track the improvement in success rates over time, providing insights into SpaceX's performance trends year by year.

[Link to GitHub notebook with EDA Data Viz Lab]

EDA with SQL

SQL Queries Used

- **Unique Launch Sites**:
  - Used SELECT DISTINCT to list all unique launch sites in the dataset.
- **Filter Launch Sites**:
  - Queried 5 records where the Launch_Site starts with "CCA" using LIKE.
- **Total Payload Mass for NASA (CRS)**:
  - Calculated the total payload mass for NASA (CRS) missions using SUM.
- **Average Payload Mass for F9 v1.1**:
  - Computed the average payload mass for booster version F9 v1.1 using AVG.
- **First Successful Ground Pad Landing**:
  - Found the earliest date for a successful ground pad landing using MIN.
- **Drone Ship Success with Specific Payload Range**:
  - Listed booster versions with successful drone ship landings and payload mass between 4000 and 6000 using conditional filters.
- **Count of Successful and Failed Landings**:
  - Counted total successful and failed landing outcomes using COUNT and UNION ALL.
- **Booster Versions with Maximum Payload**:
  - Identified booster versions carrying the maximum payload mass using a subquery.
- **Failed Drone Ship Landings in 2015**:
  - Extracted records for 2015 with failed drone ship landings, including month and launch site, using substr to extract year and month.
- **Landing Outcome Counts by Date**:
  - Ranked landing outcomes (e.g., success or failure) by count between 2010-06-04 and 2017-03-20 using filtering and ORDER BY.

Link to GitHub notebook with SQL EDA Lab

Build an Interactive Map with Folium

Map Objects Used and Their Purpose

- **Markers**:
  - **Launch Sites**: Labeled markers show site names and distances from the equator for easy identification.
  - **Launch Outcomes**: Green for successes, red for failures—visualizing site-specific success rates.
  - **Proximities**: Markers for railways, highways, coastlines, and cities with distance labels to assess logistical factors.
- **Circles**:
  - Highlighted launch site locations for better visibility.
- **Lines**:
  - Connected sites to proximities (railways, highways, coastline, cities) to visualize distances and relationships.
- **Equator Line**:
  - Added for reference to evaluate proximity of sites to the equator.

Link to GitHub notebook (via NBViewer) with Interactive Folium Maps

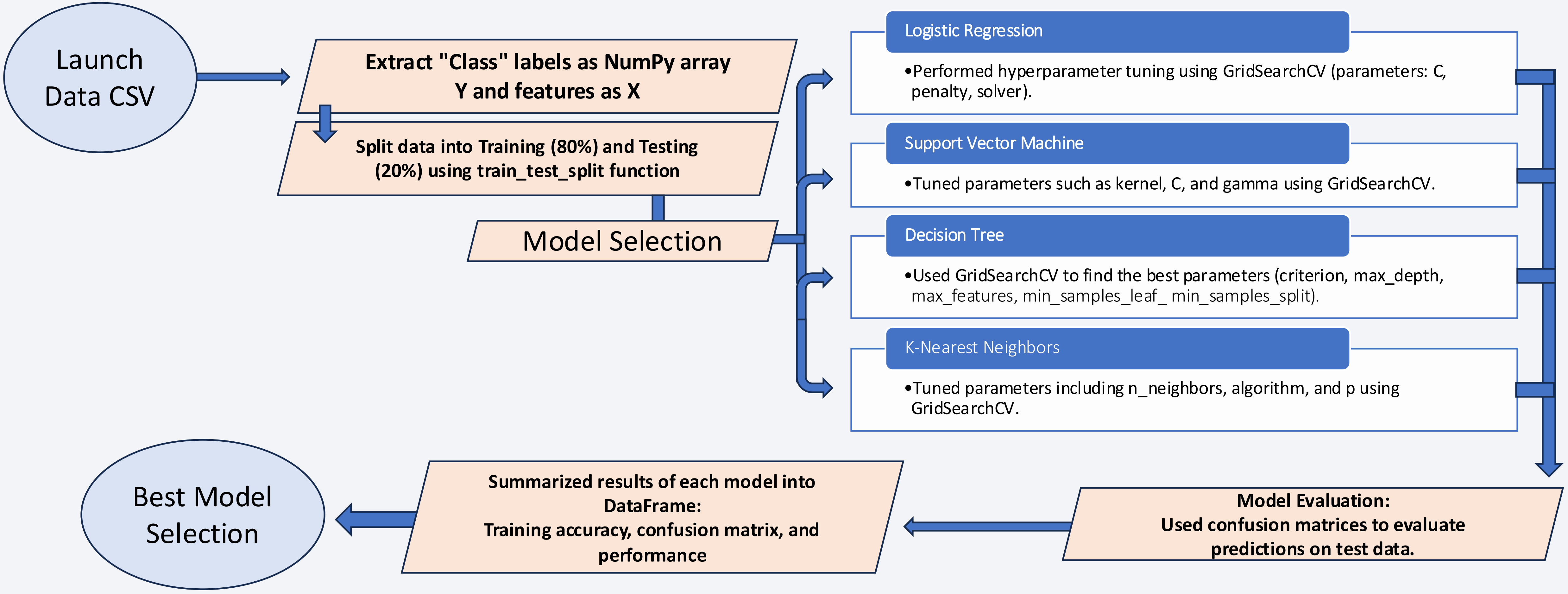
Build a Dashboard with Plotly Dash

Plots Used, Interactions, and Their Purpose

- **Pie Chart**:
  - Shows total successes across all sites or success vs. failure for a selected site.
  - Added to provide an overview of launch performance by site.
- **Scatter Plot**:
  - Visualizes the relationship between payload mass and success, with booster versions color-coded for context.
  - Helps uncover trends or anomalies in payload and success rates.
- **Dropdown Menu**:
  - Enables selection of specific sites or all sites.
  - Provides flexibility to focus on site-specific or overall data.
- **Payload Range Slider**:
  - Filters launches by payload mass.
  - Highlights patterns within specific payload ranges.
  - These elements were added to create an interactive dashboard, making it easy to explore SpaceX data and identify key insights visually.

Link to GitHub notebook with Plotly Dash Lab

Predictive Analysis (Classification)



[Link to GitHub notebook with Predictive Analysis Lab]()

Results

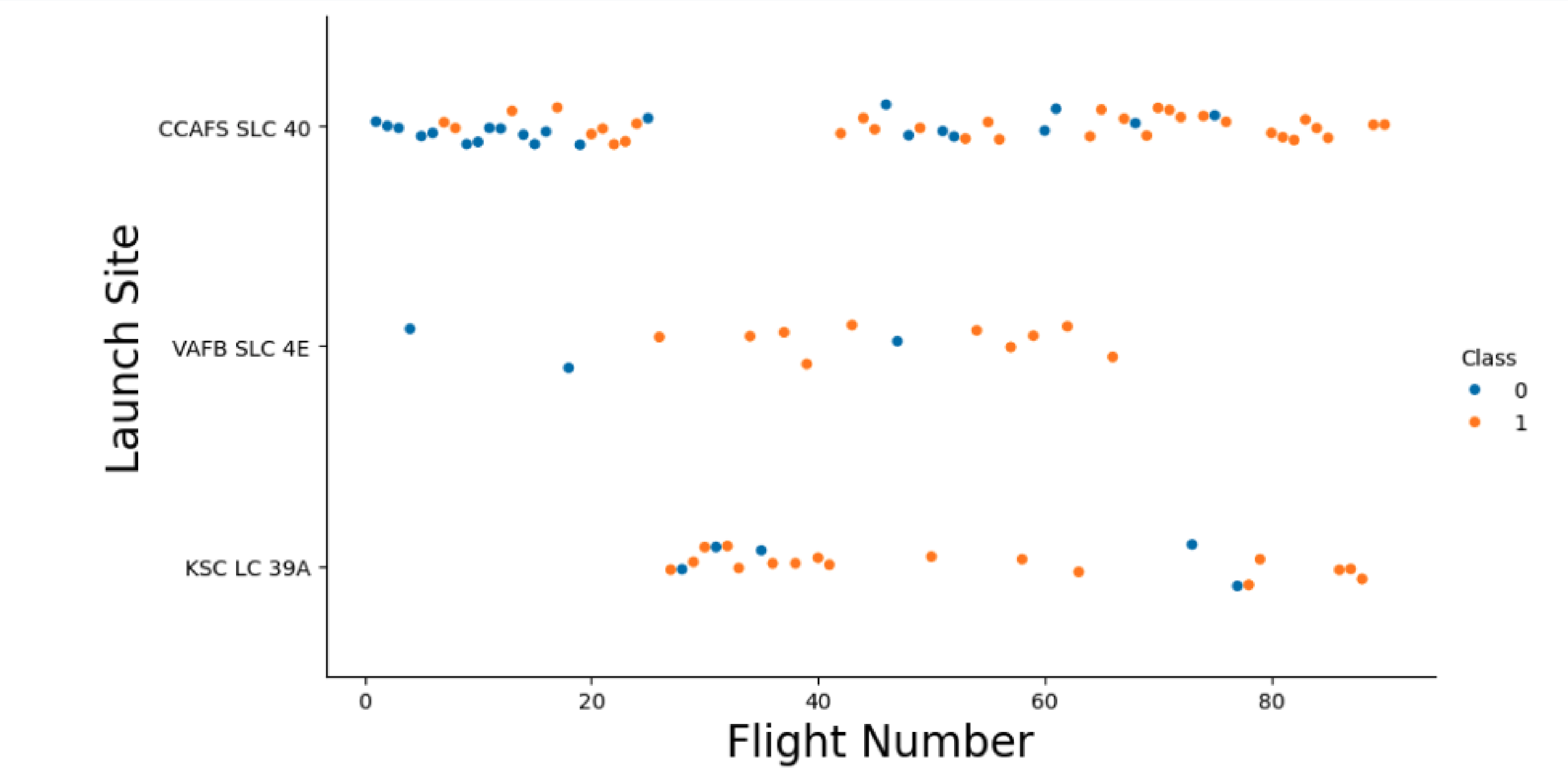
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

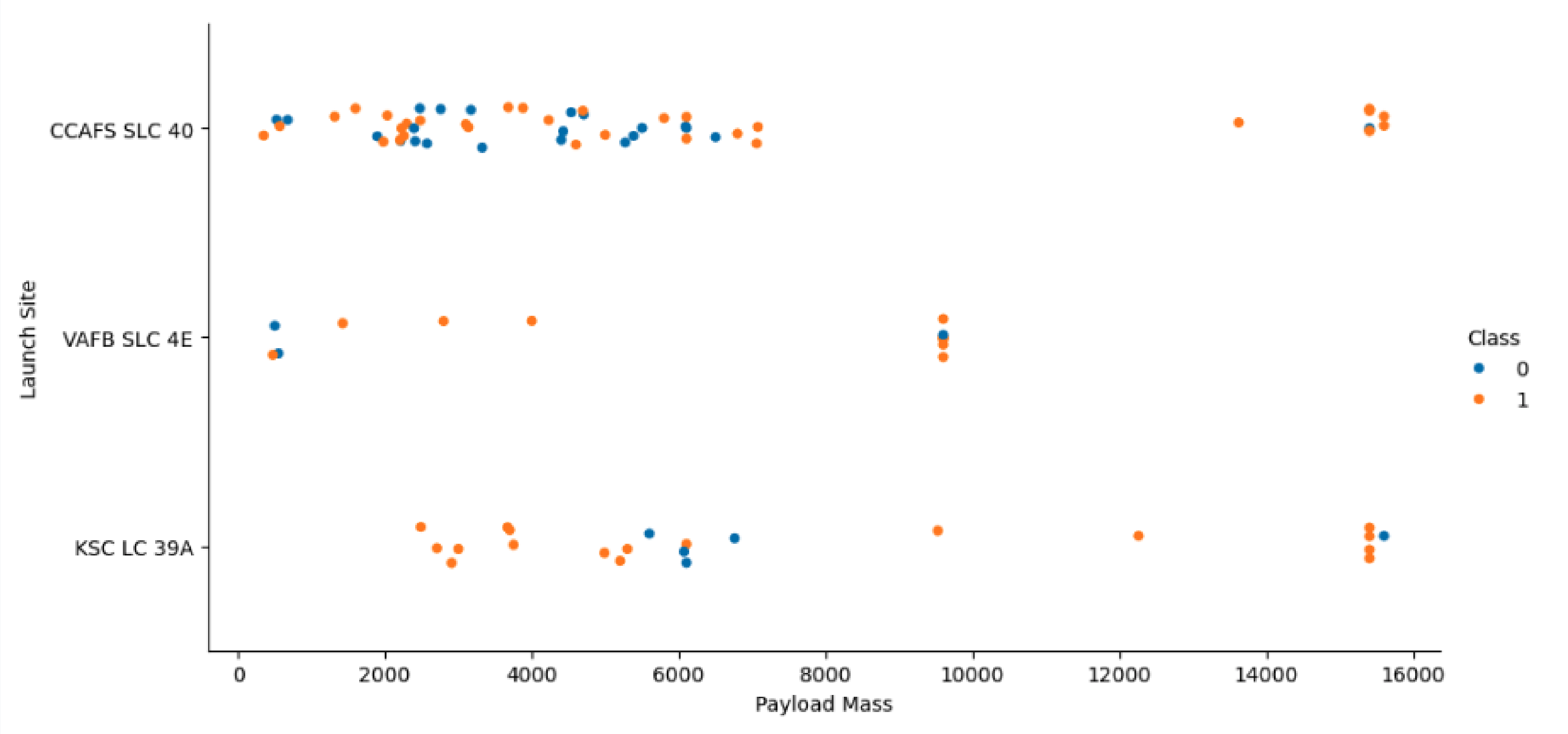
Flight Number vs. Launch Site

- Through this scatter plot we can tell that success rate for rocket launches improved over time.
- CCAFS saw the most launches of any launch site.



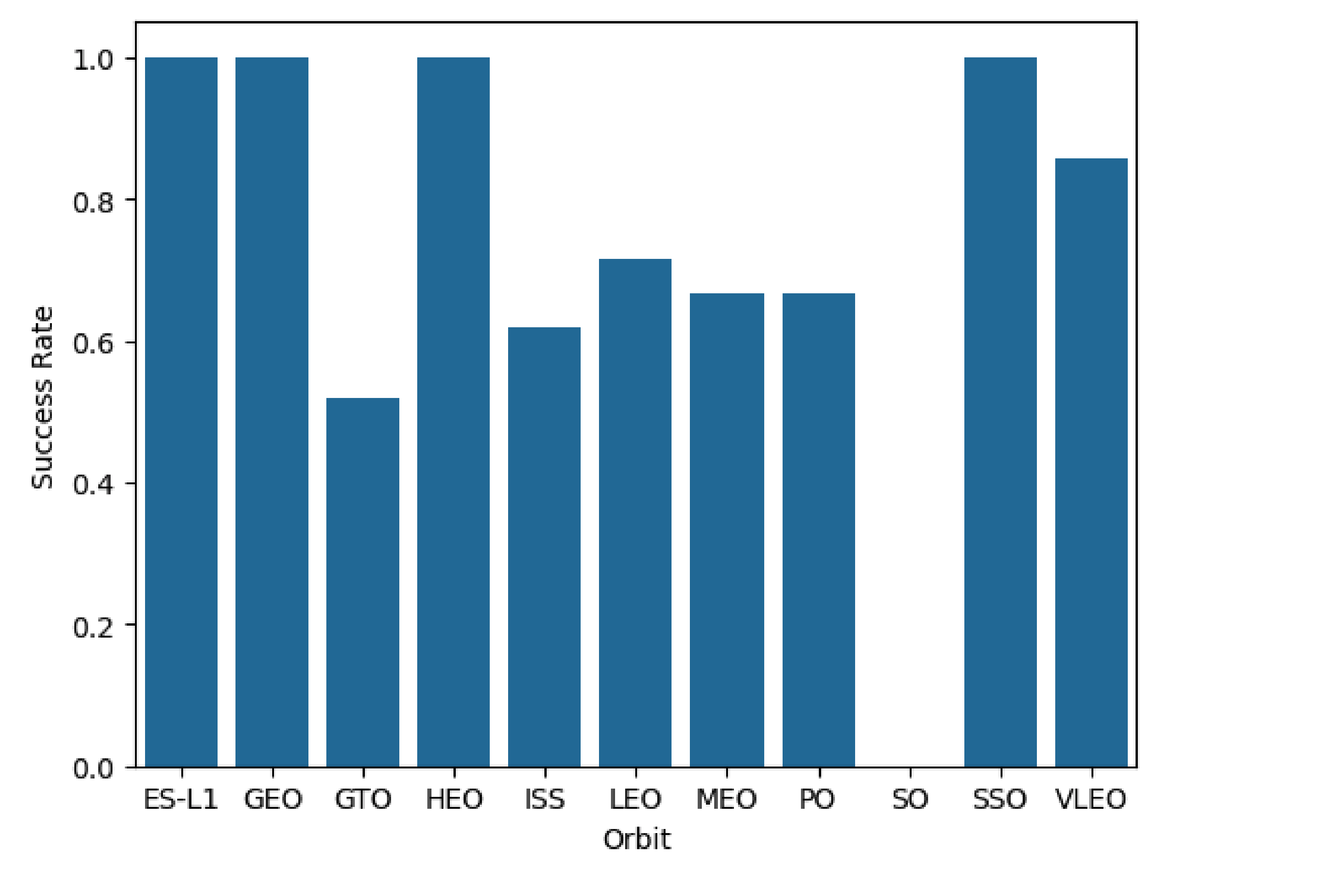
Payload vs. Launch Site

- The larger the payload mass, the higher success rate.
- When payload mass exceeded 8,000kg, almost every rocket launch had a successful outcome.



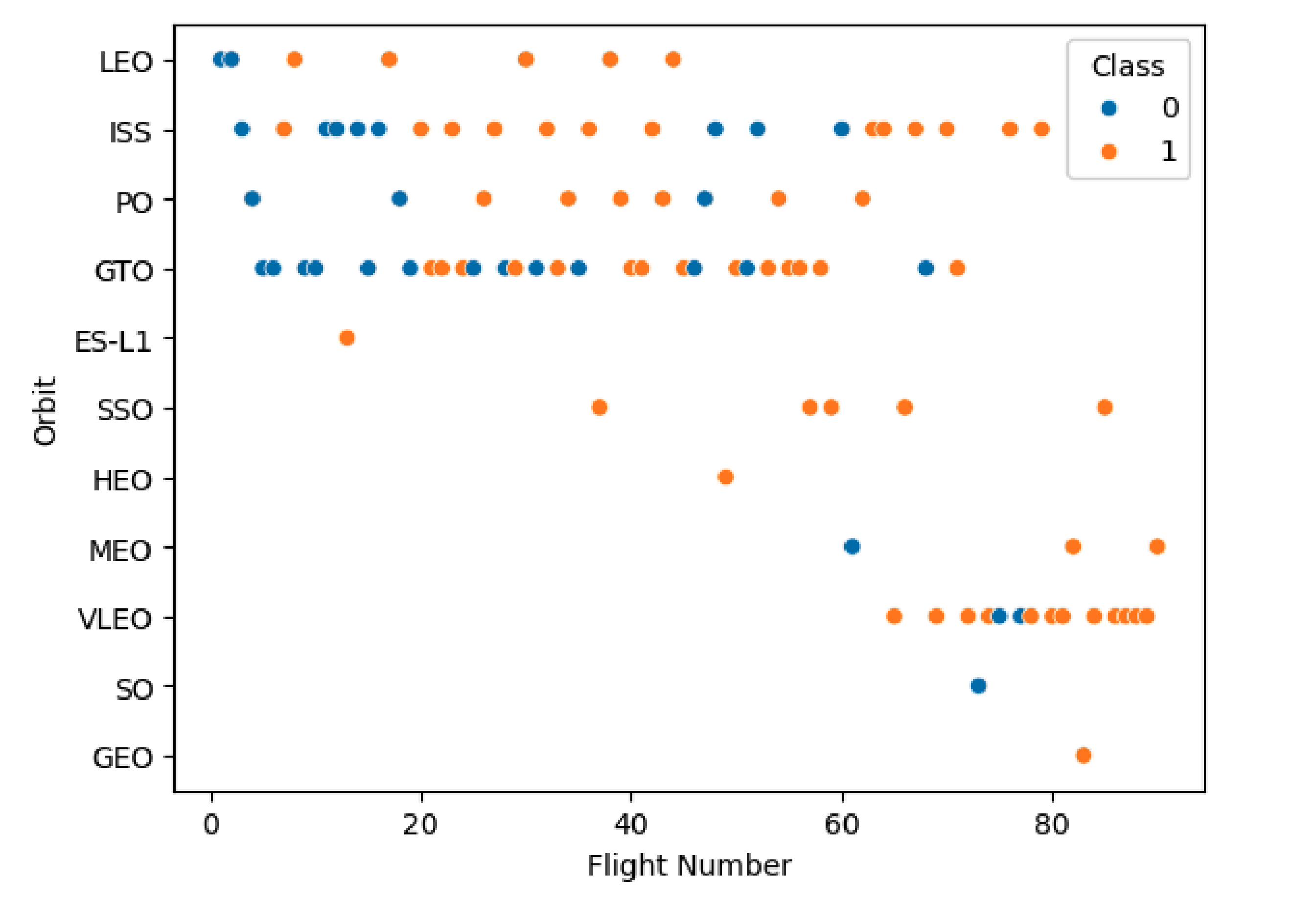
Success Rate vs. Orbit Type

- The ES-L1, GEO, HEO, and SSO orbits had the highest success rates.
- The rest of the orbits averaged around a 70% success rate.



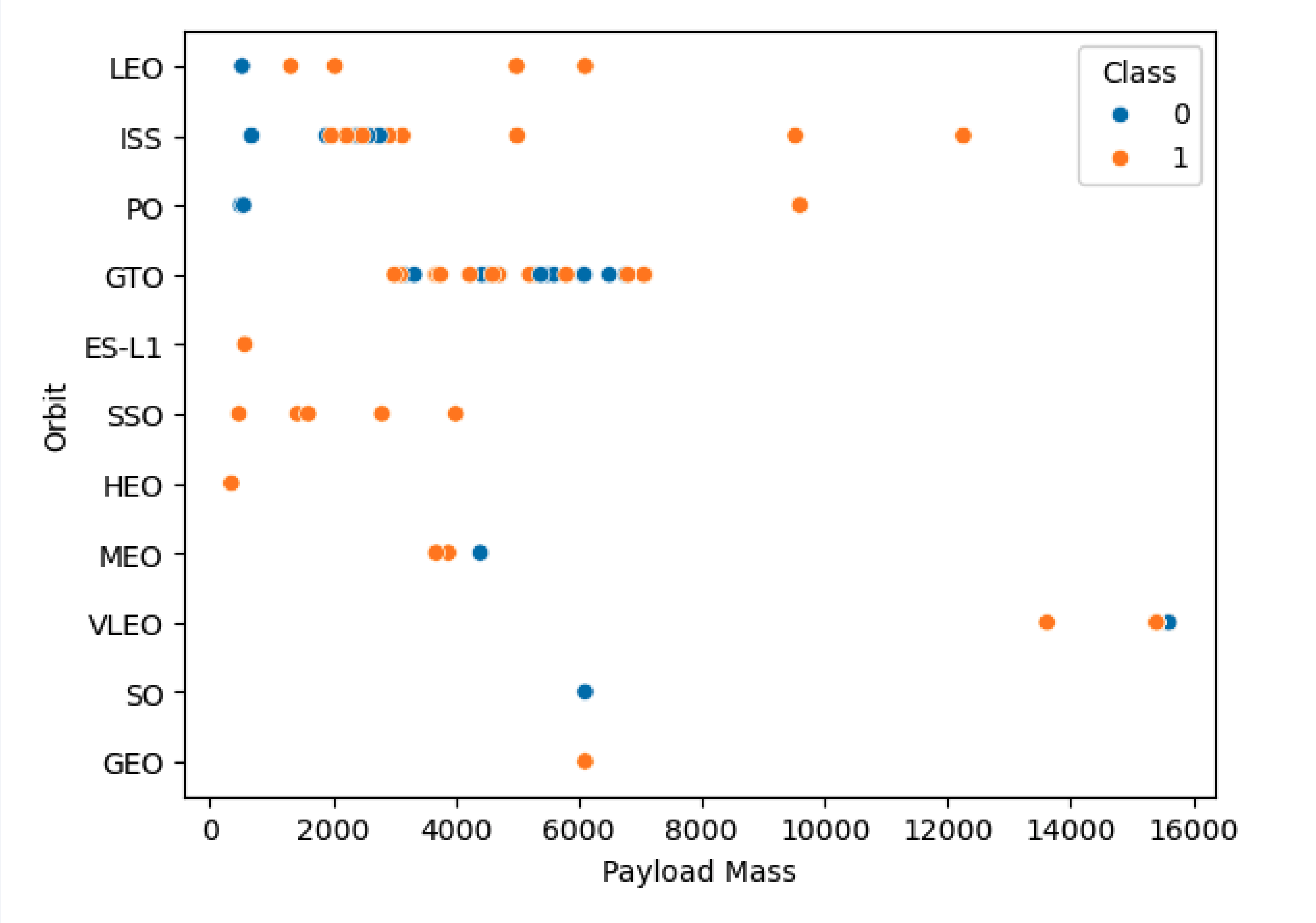
Flight Number vs. Orbit Type

- Again we can see that over time, the success rate increased, even when comparing across various orbits.



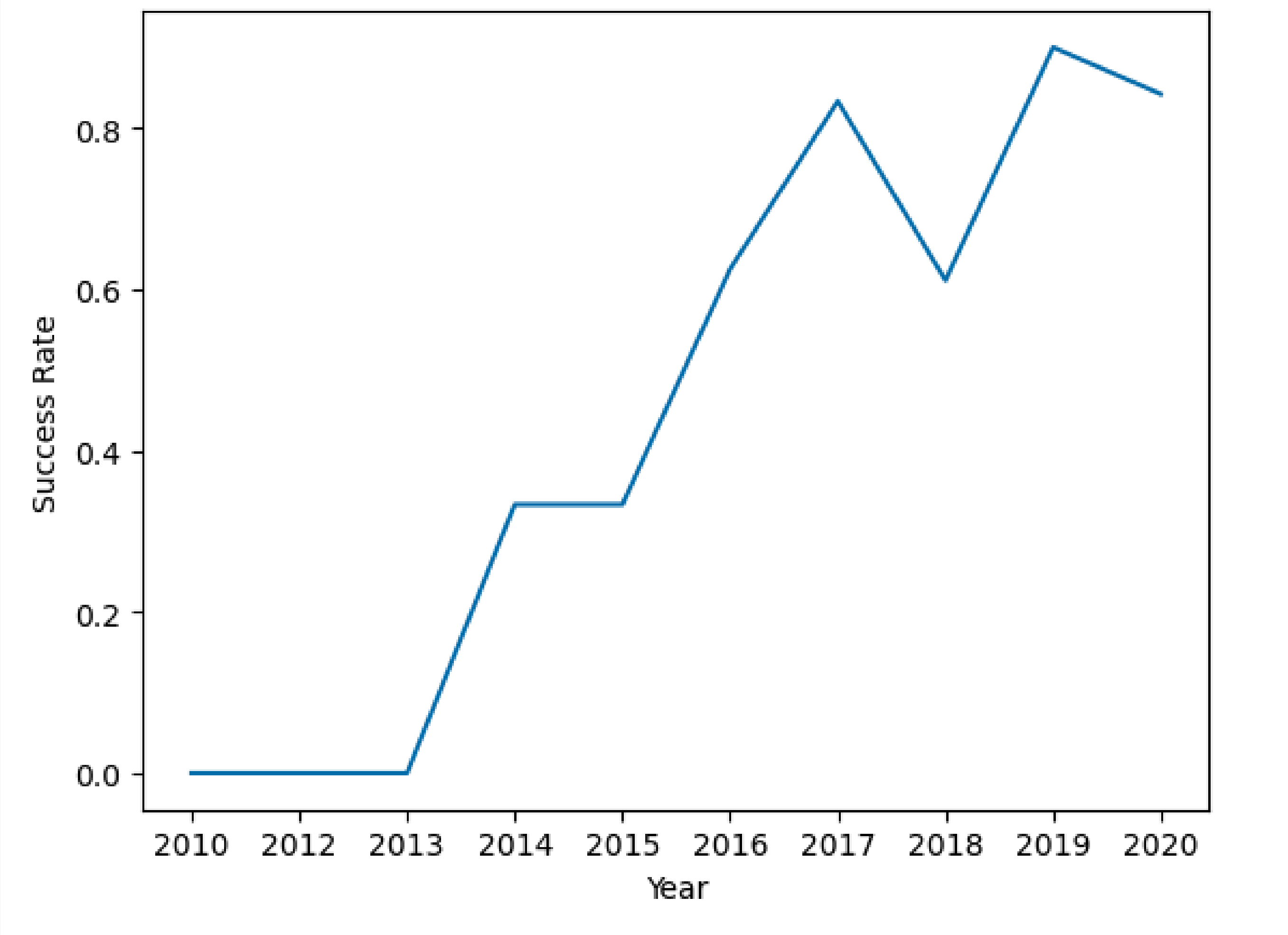
Payload vs. Orbit Type

- The ISS, GTO, and VLEO orbits had the highest success rates.
- When payload mass exceeded 8,000kg, almost every rocket launch had a successful outcome with these orbits.



Launch Success Yearly Trend

- This is the best visual to see how success rate improved over time.
- With the exception of a small dip in 2018, success rate only went up year over year.



All Launch Site Names

- To retrieve each launch site, we use the DISTINCT SQL function on the Launch_Site column.

```
SELECT DISTINCT "Launch_Site" FROM SPACEXTABLE;
```

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

Launch Site Names Begin with 'CCA'

- To pull 5 instances of launch sites starting with CCA, we had to use LIKE. When putting the % after CCA we can allow for all instances where the launch site starts with CCA but may end with different characters.

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
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

- Calculated the total PAYLOAD_MASS__KG for NASA CRS missions by summing the values, casting as FLOAT, and assigning the result to a new column, total_payload_mass_nasacrs.

```
SELECT SUM(CAST("PAYLOAD_MASS__KG_" AS FLOAT)) AS
  total_payload_mass_nasacrs FROM SPACEXTABLE WHERE
  "Customer" = 'NASA (CRS)';
```

total_payload_mass_nasacrs
91192.0

Average Payload Mass by F9 v1.1

- This query calculates the average payload mass for launches using the F9 v1.1 booster version by filtering relevant records and computing the mean. The result is stored as avg_payload_mass_f9_1_1.

```
SELECT AVG(CAST("PAYLOAD_MASS__KG_" AS FLOAT)) AS avg_payload_mass_f9_1_1
    FROM SPACEXTABLE WHERE "Booster_Version" = 'F9 v1.1';
```

avg_payload_mass_f9_1_1
2928.4

First Successful Ground Landing Date

- This query retrieves the earliest date on which a successful landing on a ground pad occurred, providing insight into the timeline of SpaceX's landing achievements.

```
SELECT MIN("Date") FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Success (ground
    pad)';
```

MIN("Date")
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- This query lists all unique booster versions that achieved a successful landing on a drone ship while carrying a payload mass between 4000 and 6000 kg, helping identify boosters used for specific mission profiles.

```
SELECT DISTINCT "Booster_Version" FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Success
(drone ship)' AND "PAYLOAD_MASS__KG_" > 4000 AND "PAYLOAD_MASS__KG_" < 6000;
```

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

Total Number of Successful and Failure Mission Outcomes

- This query counts the total number of successful and failed landings, categorizing them as Success or Failure. It combines the results into a single output using UNION ALL for a clear comparison of landing outcomes.

```
SELECT COUNT(*) AS total_landings, 'Success' AS landing_type FROM SPACEXTABLE WHERE
  "Landing_Outcome" LIKE 'Success%' UNION ALL SELECT COUNT(*) AS total_landings, 'Failure' AS
  landing_type FROM SPACEXTABLE WHERE "Landing_Outcome" LIKE 'Failure%';
```

total_landings	landing_type
122	Success
20	Failure

Boosters Carried Maximum Payload

- This query identifies all unique booster versions that carried the maximum payload mass recorded in the dataset by using a subquery to find the highest payload mass and filtering for matching records.

```
SELECT DISTINCT "Booster_Version" FROM SPACEXTABLE WHERE "PAYLOAD_MASS__KG_"
    = (SELECT MAX("PAYLOAD_MASS__KG_") FROM SPACEXTABLE);
```

Booster_Version
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1051.3
F9 B5 B1056.4
F9 B5 B1048.5
F9 B5 B1051.4
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3
F9 B5 B1051.6
F9 B5 B1060.3
F9 B5 B1049.7

2015 Launch Records

- This query extracts the month and other details (landing outcome, booster version, and launch site) for failed drone ship landings that occurred in 2015. It uses the substr function to isolate the month and year from the Date column.

```
SELECT substr("Date", 6, 2) AS MONTH, "Landing_Outcome", "Booster_Version",
    "Launch_Site" FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Failure (drone ship)'
    AND substr("Date", 0, 5) = '2015';
```

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

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

- This query retrieves all records from the dataset where the date falls between June 4, 2010, and March 20, 2017, inclusive. The results are sorted in descending order by date to display the most recent records first.

SELECT * FROM SPACEXTABLE WHERE DATE >= '2010-06-04' AND DATE <= '2017-03-20' ORDER BY DATE DESC;

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2017-03-16	6:00:00	F9 FT B1030	KSC LC-39A	EchoStar 23	5600	GTO	EchoStar	Success	No attempt
2017-03-16	6:00:00	F9 FT B1030	KSC LC-39A	EchoStar 23	5600	GTO	EchoStar	Success	No attempt
2017-02-19	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
2017-02-19	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
2017-01-14	17:54:00	F9 FT B1029.1	VAFB SLC-4E	Iridium NEXT 1	9600	Polar LEO	Iridium Communications	Success	Success (drone ship)
2017-01-14	17:54:00	F9 FT B1029.1	VAFB SLC-4E	Iridium NEXT 1	9600	Polar LEO	Iridium Communications	Success	Success (drone ship)
2016-08-14	5:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2016-08-14	5:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2016-07-18	4:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
2016-07-18	4:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
2016-06-15	14:29:00	F9 FT B1024	CCAFS LC-40	ABS-2A Eutelsat 117 West B	3600	GTO	ABS Eutelsat	Success	Failure (drone ship)
2016-06-15	14:29:00	F9 FT B1024	CCAFS LC-40	ABS-2A Eutelsat 117 West B	3600	GTO	ABS Eutelsat	Success	Failure (drone ship)
2016-05-27	21:39:00	F9 FT B1023.1	CCAFS LC-40	Thaicom 8	3100	GTO	Thaicom	Success	Success (drone ship)
2016-05-27	21:39:00	F9 FT B1023.1	CCAFS LC-40	Thaicom 8	3100	GTO	Thaicom	Success	Success (drone ship)
2016-05-06	5:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2016-05-06	5:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2016-04-08	20:43:00	F9 FT B1021.1	CCAFS LC-40	SpaceX CRS-8	3136	LEO (ISS)	NASA (CRS)	Success	Success (drone ship)
2016-04-08	20:43:00	F9 FT B1021.1	CCAFS LC-40	SpaceX CRS-8	3136	LEO (ISS)	NASA (CRS)	Success	Success (drone ship)
2016-03-04	23:35:00	F9 FT B1020	CCAFS LC-40	SES-9	5271	GTO	SES	Success	Failure (drone ship)
2016-03-04	23:35:00	F9 FT B1020	CCAFS LC-40	SES-9	5271	GTO	SES	Success	Failure (drone ship)
2016-01-17	18:42:00	F9 v1.1 B1017	VAFB SLC-4E	Jason-3	553	LEO	NASA (LSP) NOAA CNES	Success	Failure (drone ship)
2016-01-17	18:42:00	F9 v1.1 B1017	VAFB SLC-4E	Jason-3	553	LEO	NASA (LSP) NOAA CNES	Success	Failure (drone ship)
2015-12-22	1:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)
2015-12-22	1:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)
2015-06-28	14:21:00	F9 v1.1 B1018	CCAFS LC-40	SpaceX CRS-7	1952	LEO (ISS)	NASA (CRS)	Failure (in flight)	Precluded (drone ship)
2015-06-28	14:21:00	F9 v1.1 B1018	CCAFS LC-40	SpaceX CRS-7	1952	LEO (ISS)	NASA (CRS)	Failure (in flight)	Precluded (drone ship)

Section 3

Launch Sites Proximities Analysis

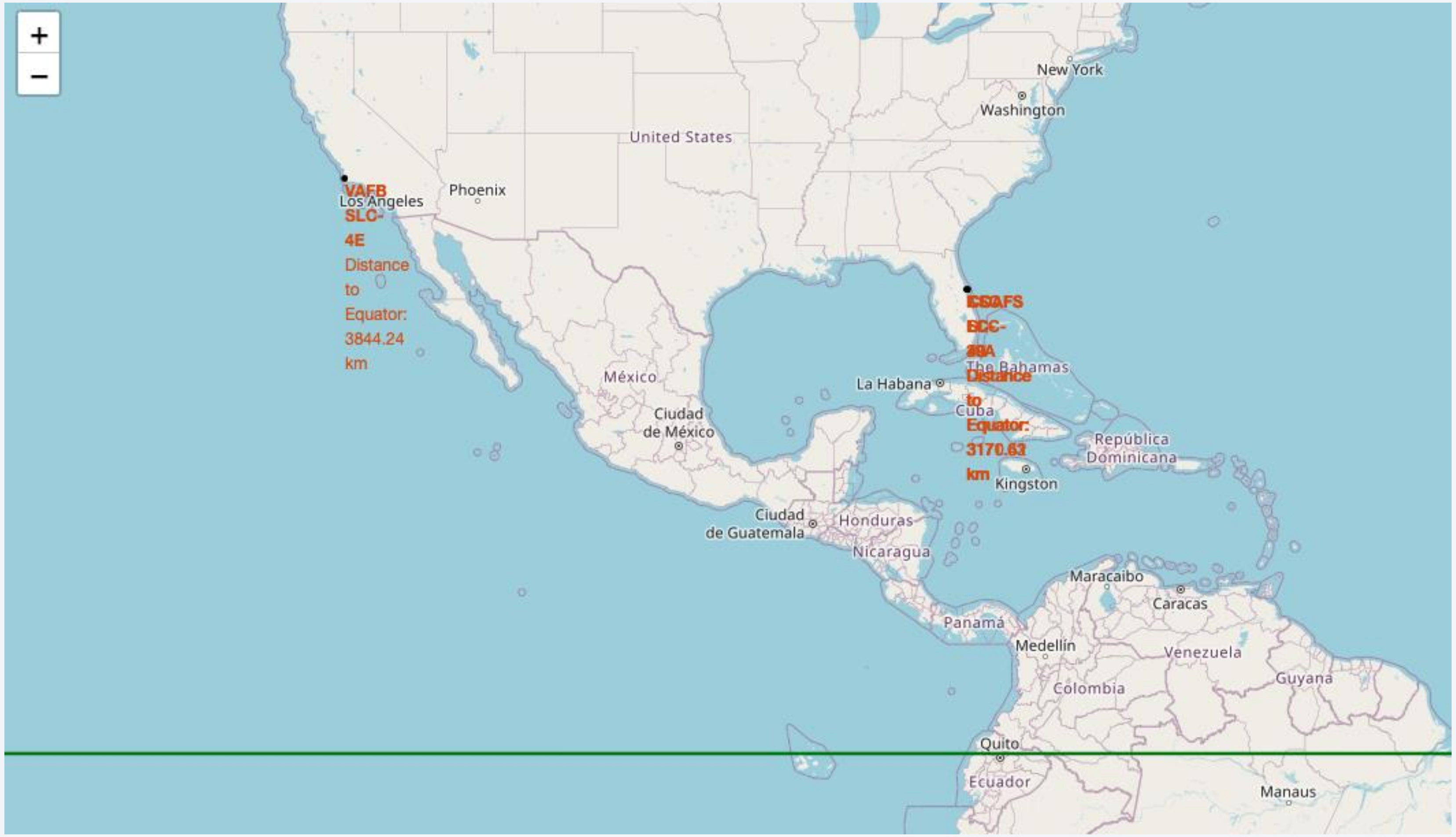
Global Map of SpaceX Launch Sites

Map Elements:

- Markers: Each launch site is marked with:
- A label showing the site name.
- Distance from the equator displayed below the name.
- Circles: Highlight each launch site location for better visibility.
- Equator Line: A green line across latitude 0 for reference.

Key Findings:

- Proximity to Equator: Launch sites are relatively far from the equator globally but are closer when viewed in the U.S. context (~3000–4000 km).
- Coastal Locations: All sites are near coastlines, likely for safety and logistical reasons during launches.



Launch Outcomes by Location: Successes and Failures

Map Elements

Markers:

- Green: Represent successful launches (class = 1).
- Red: Represent failed launches (class = 0).
- Popups on markers display the corresponding launch site names.

Marker Clusters:

- Group overlapping markers to simplify the map and make it more interactive.
- Allow users to zoom in and explore individual launch outcomes.

Key Findings

Launch Outcome Patterns:

- Successes and failures are distributed across all launch sites.
- Certain sites, like VAFB SLC-4E (shown right), show a mix of successes and failures, indicating varying performance outcomes.

Geographical Insights:

- Marker clustering indicates high activity at specific sites.
- Visualizing outcomes by site helps identify locations with consistently high success rates.

Proximity Analysis of Launch Site to Key Landmarks

Map Elements

Markers:

- Labeled with calculated distances from the launch site to:
- Coastline: Key for sea-based safety during launches.
- City: Demonstrates sufficient distance for safety and noise reduction.
- Highway: Indicates accessibility for logistical operations.

Lines:

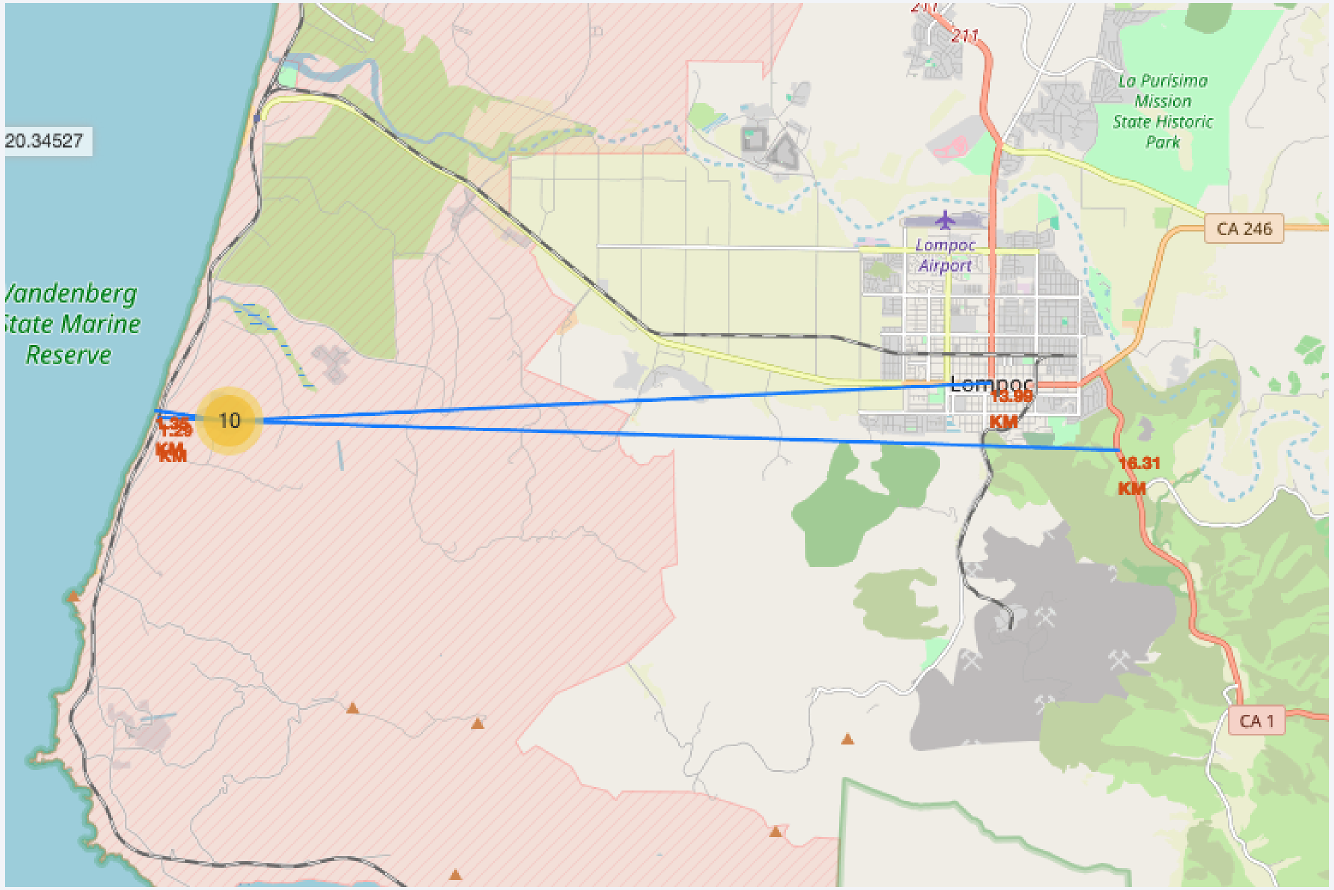
- Connect the launch site to its nearest city, coastline, and highway for clear visualization of proximities.

Launch Site:

- Highlighted as the central location connecting to other landmarks.

Key Findings

- Proximity to Coastline: Launch sites are positioned very close to coastlines, likely to mitigate risks in case of launch failures.
- Distance from Cities: The site maintains a safe distance (~14 km) from cities to ensure safety and minimize noise disturbances.
- Accessibility: Proximity to highways (~18 km) shows logistical considerations for transport and operations.



Section 4

Build a Dashboard with Plotly Dash

Launch Success Count by Site: Pie Chart

Chart Elements

Pie Chart:

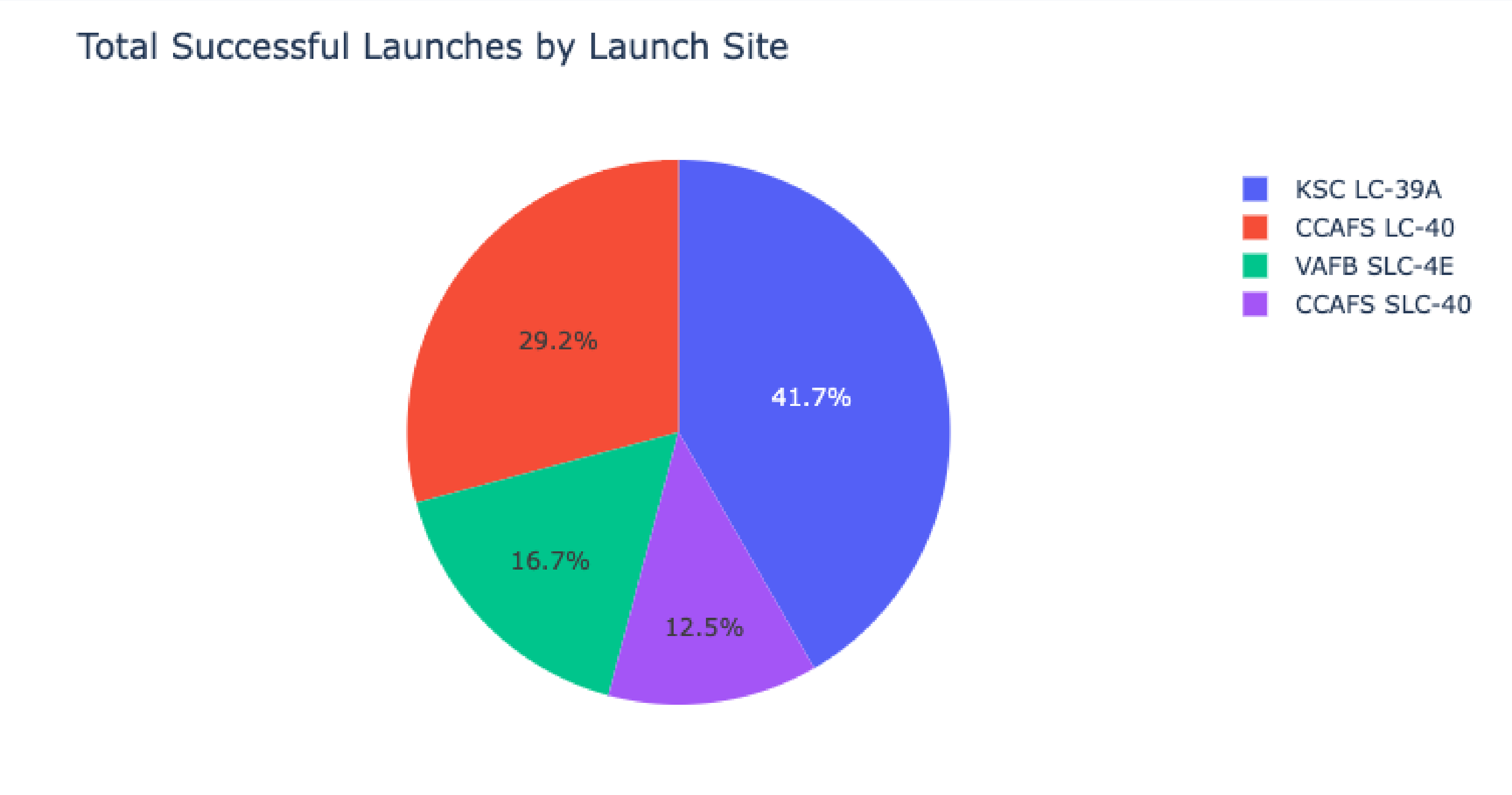
- Slices represent the proportion of successful launches at each site.
- Each slice is labeled with the site name and its corresponding percentage.

Title:

- Clearly indicates the data being visualized (Total Successful Launches by Launch Site).

Key Findings

- The distribution shows which sites contributed the most to successful launches.
- Insights such as whether one site is more consistently successful or whether launches are evenly distributed across all sites can be drawn from this visualization.



Launch Success Ratio: Best Performing Site

Chart Elements

Pie Chart:

Divided into two slices:

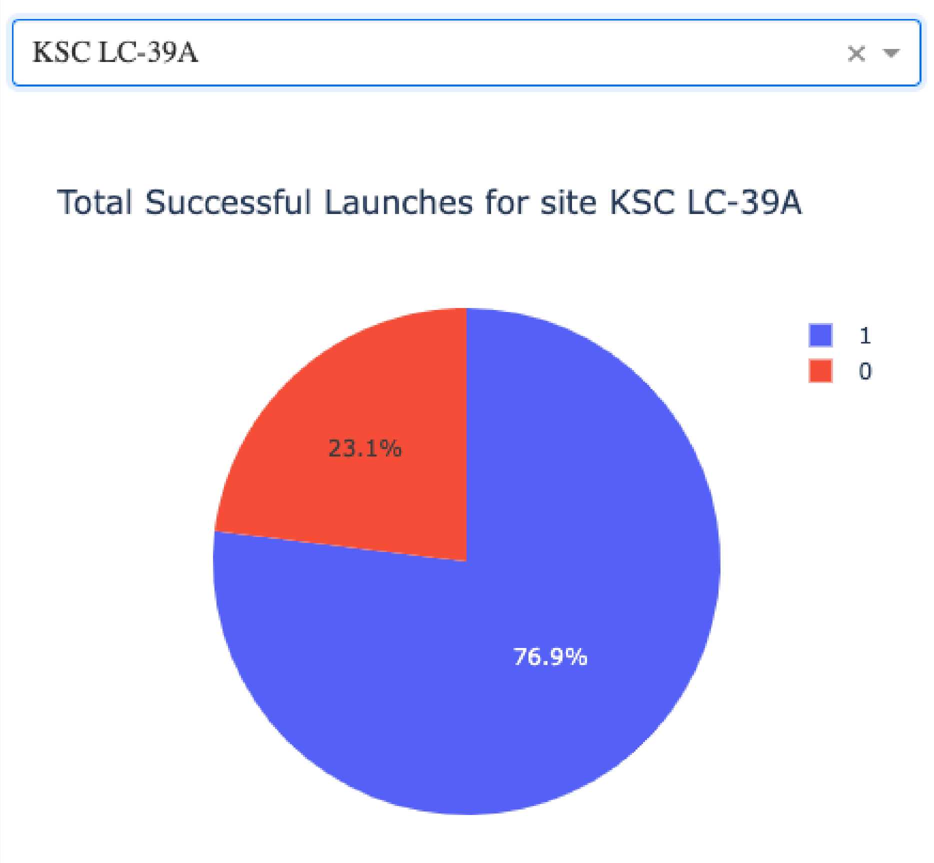
- Successes (1, blue): Represents the count of successful launches at the site.
- Failures (2, red): Represents the count of failed launches.
- Includes labels with the exact counts or percentages for each category.

Title:

- Clearly indicates the data being visualized (Total Successful Launches for KSC LC-39A).

Key Findings

- Kennedy Space Center had nearly 80% of all launches land successfully. This performance is better than any of the other launch sites we examined.



Payload vs. Launch Outcome for All Sites

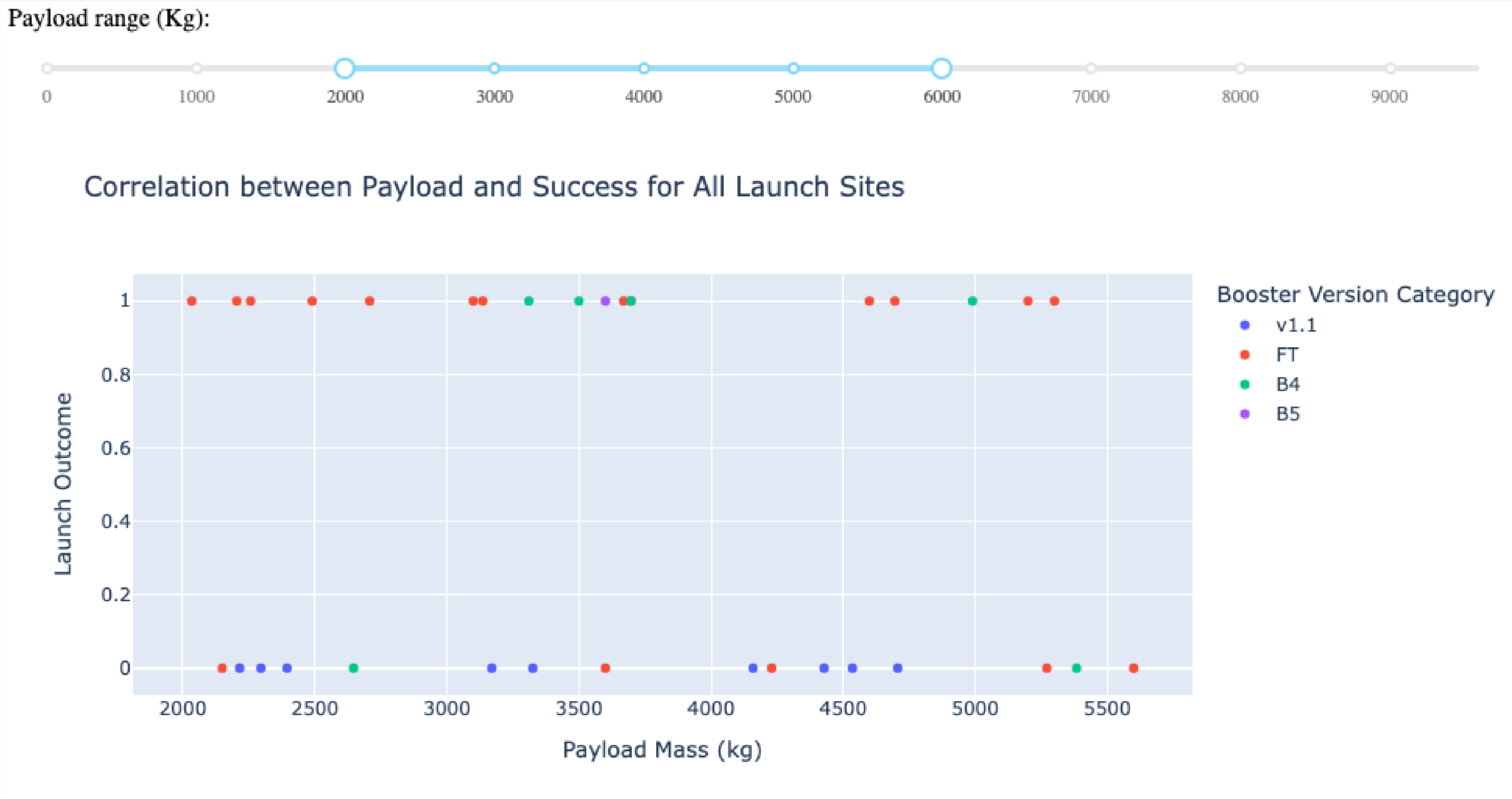
Chart Elements

Scatter Plot:

- X-axis: Payload Mass (kg).
- Y-axis: Launch Outcome (1 = Success, 0 = Failure).
- Color-coded points represent different Booster Version Categories.

Range Slider:

- Filters payload values to explore trends within specific ranges.



Key Findings

Payload and Success Correlation:

- Higher success rates are often observed in specific payload ranges, such as mid-range payloads (e.g., 2000–5000 kg).

Booster Version Impact:

- Certain booster versions may have consistently higher success rates, visible through cluster patterns.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

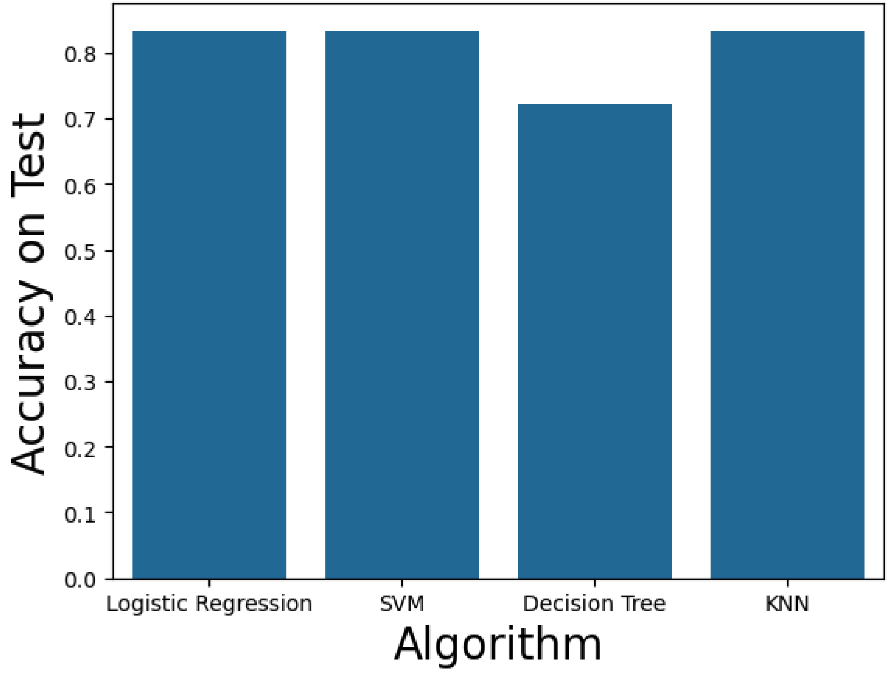
Model Selection

- **Logistic Regression**:
    - Performed hyperparameter tuning using GridSearchCV (parameters: C, penalty, solver).
    - Achieved training accuracy of **84.6%** and test accuracy of **83.3%**.
- **Support Vector Machine (SVM)**:
    - Tuned parameters such as kernel, C, and gamma using GridSearchCV.
    - Achieved training accuracy of **84.8%** and test accuracy of **83.3%**.
- **Decision Tree**:
    - Used GridSearchCV to find the best parameters (criterion, max_depth, etc.).
    - Achieved training accuracy of **87.5%** but test accuracy was lower at **72.2%**.
- **K-Nearest Neighbors (KNN)**:
    - Tuned parameters including n_neighbors, algorithm, and p using GridSearchCV.
    - Achieved training accuracy of **84.8%** and test accuracy of **83.3%**.

Best Model Selection

- **This was a 3-way tie between Logistic Regression, Support Vector Machine, and K-Nearest Neighbors. Each of these had an accuracy score of 83% on the testing data.**
- **With such a small dataset, we're finding that accuracy is going to be almost uniform across all models.**

	Algorithm	Best Score on Training	Accuracy on Test	Confusion Matrix	Performance
0	Logistic Regression	0.846429	0.833333	0-3-3-12	High
1	SVM	0.848214	0.833333	0-3-3-12	High
2	Decision Tree	0.875000	0.722222	0-5-1-12	Better
3	KNN	0.848214	0.833333	0-3-3-12	High



Confusion Matrix

Key Metrics

True Positives (TP = 12):

- Cases where the rocket actually landed, and the model correctly predicted it.

True Negatives (TN = 3):

- Cases where the rocket did not land, and the model correctly predicted it.

False Positives (FP = 3):

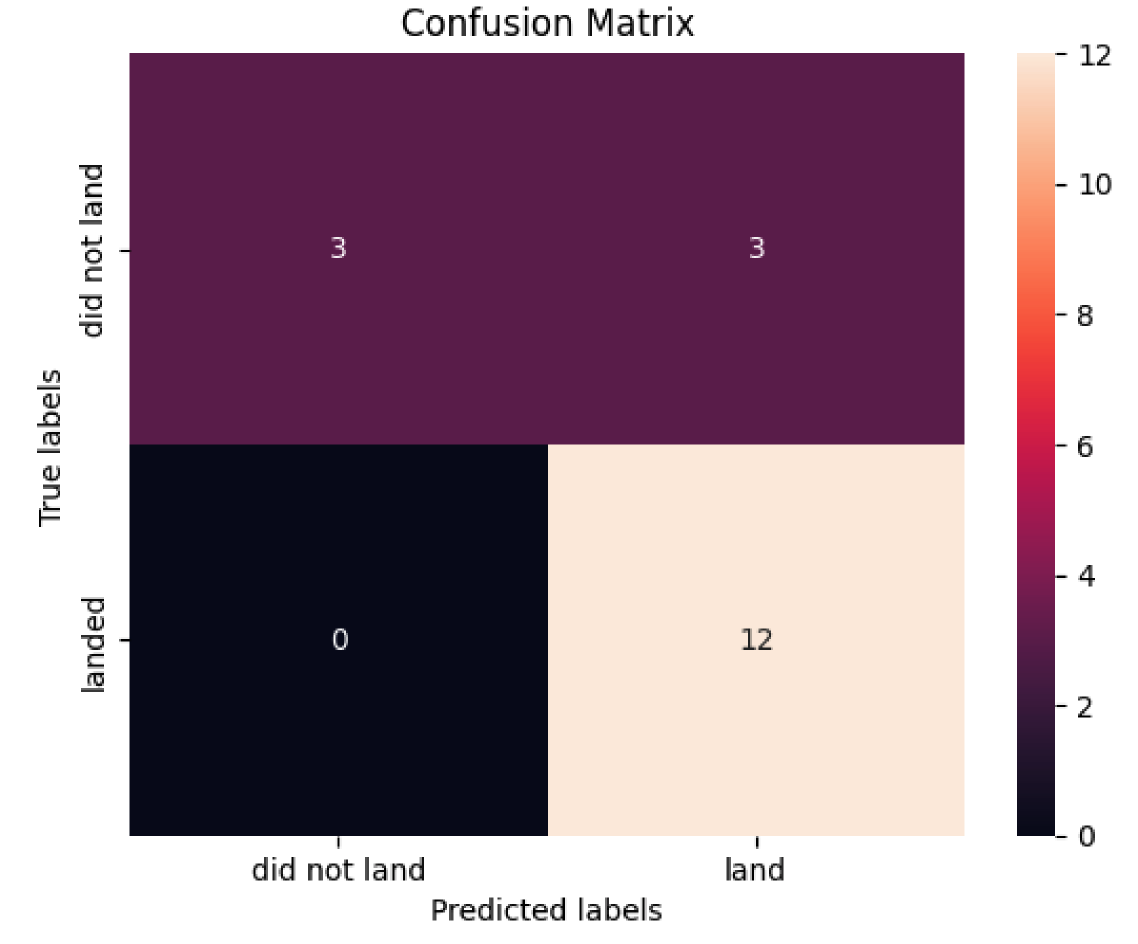
- Cases where the rocket did not land, but the model incorrectly predicted it landed.

False Negatives (FN = 0):

- Cases where the rocket landed, but the model incorrectly predicted it did not land.

Findings

- The model performs well in distinguishing between landed and not-landed classes.
- Strength: No false negatives, meaning the model successfully predicts all actual landings.
- Limitation: The model occasionally misclassifies unsuccessful landings as successful, reflected in the false positives.



Conclusions

What factors determine a successful SpaceX rocket landing?

- Geographical Factors: Launch sites located near coastlines and relatively closer to the equator optimize trajectory and safety, which are key to successful landings.
- Payload Mass: Moderate payloads generally lead to higher success rates, as heavier payloads can reduce the probability of a controlled landing.

Which variables most influence first-stage success rates?

- Orbit Type: Orbits like LEO (Low Earth Orbit) are associated with higher success rates compared to more demanding orbits like GTO (Geostationary Transfer Orbit).
- Payload Mass: A strong correlation was observed between payload mass and landing success, with optimal ranges yielding better outcomes.
- Booster Version: Advanced booster versions (e.g., Falcon 9 Block 5) demonstrated higher success rates, reflecting SpaceX's iterative improvements in technology.

How accurately can we estimate launch costs by predicting first-stage success?

- By predicting first-stage success rates with 83.3% accuracy (Logistic Regression), we can provide a reliable estimation of whether a mission can leverage reusable components.
- Successful landings directly reduce costs by enabling reusability, as reflected in the correlation between success predictions and estimated launch savings.

Overall Project Takeaways

- Data-Driven Optimization: This analysis underscores the importance of data in enhancing SpaceX's operations, from identifying key success factors to optimizing costs.
- Future-Focused Insights: Machine learning models can be further improved to predict not just success rates but also mission reliability, aiding in bid proposals and strategic planning.

Appendix

Thank you to Coursera and IBM for creating and hosting this data science certificate program.

Thank you!