IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Amir Zubair

February 2025

Outline

- Executive Summary
- Introduction
- Methodology
    - Data Collection (API & Web Scraping)
    - Data Wrangling & EDA
    - Interactive Visual Analysis (Folium & Plotly Dash)
    - Predictive Modeling
- Results
    - Exploratory Data Analysis
    - Interactive Analytics
    - Predictive Model Performance
- Conclusion
- Appendix

Executive Summary

- Summary of methodologies
  - Data Collection
    - Used SpaceX API
    - Web Scraping from Wikipedia to collect launch data.
  - Data Processing: Data Wrangling and feature engineering.
  - EDA & Visual Analytics:
    - Applied SQL, Matplotlib/Seaborn visualizations, interactive Folium maps, and Plotly Dash dashboards.
  - Predictive Analytics:
    - Built classification models (Logistic Regression, SVM, Decision Trees, KNN) with hyperparameter tuning.
  - Key Insights:
    - Identified relationship between launch sites, payloads, booster versions, and success outcomes.

Introduction

- Background & Context:
  - The Commercial Space Age is here
  - SpaceX’s cost savings come from reusing the Falcon 9 first stage.
  - Space Y wants to compete with SpaceX
  - Predicting landing success will mean determining the cost of each launch
- Problem Statement:
  - Which factors (launch site, payload, booster version) influences landing success?
- Project Objectives:
  - Build an interactive dashboard for real-time data exploration.
  - Develop and evaluate predictive models to forecast landing outcomes.

Section 1

Methodology

Methodology

Executive Summary

- **Data Collection & Integration:**
  - Gathered SpaceX launch data using REST API calls and web scraping (Wikipedia)
  - Combined with pre-built CSV datasets (e.g., *spacex_launch_dash.csv*)

- **Data Wrangling & Feature Engineering:**
  - Cleaned and standardized data; handled missing values
  - Derived key features (e.g., “Class” for landing success) and applied one-hot encoding

- **Exploratory Analysis:**
  - Visualized relationships among Flight Number, Payload, Launch Site, Orbit, etc.
  - Executed SQL queries to summarize launch statistics and trends

- **Interactive Visual Analytics:**
  - Built interactive maps with Folium (markers, circles, distance lines)
  - Developed a dynamic dashboard with Plotly Dash (dropdowns, sliders, scatter & pie charts)

- **Predictive Modelling:**
  - Built and tuned classification models (Logistic Regression, SVM, Decision Tree, KNN)
  - Evaluated models with GridSearchCV and confusion matrices

Data Collection

- Process Description:
  - Collected launch records from SpaceX's API
  - Enhanced the dataset with additional details (rocket booster info, payload, launch site, etc.).

Data Collection - SpaceX API

- **Process:**
  - Utilized SpaceX REST API calls to retrieve historical launch records
  - Automated data collection using Python (requests module) and stored results as CSV
- **Key Steps:**
  - Define API endpoints, send GET requests, and parse JSON responses
  - Validate and store data for further processing
- **Reference:**
  - Amir-Zubair-Khalid/Data_Science_Capstone

Data Collection - Scraping

•**Process Description:**
•Scraped launch records, the “List of Falcon 9 and Falcon Heavy launches” Wikipedia page, using BeautifulSoup.
•Extracted HTML table data, parsed HTML tables and cleaned the content into a Pandas DataFrame, extracting launch details.

•**Key Steps:** Identify and extract relevant HTML table elements
•Clean and process text (e.g., remove references, normalize formatting)

•**Reference:**
- Amir-Zubair-Khalid/Data_Science_Capstone

Data Wrangling

- Data Processing Steps:
  - Combined API and scraped data with pre-existing CSV files
  - Cleaned data: addressed missing values, corrected data types, and standardized formats

- Engineered features:
  - Created target variable “Class” (0 = failure, 1 = success)
  - Applied one-hot encoding to categorical variables (e.g., Booster Version Category)

- Reference:
  - Amir-Zubair-Khalid/Data_Science_Capstone

EDA with Data Visualization

- **Charts & Plots:**
  - Scatter plots:
    - Flight Number vs. Launch Site
    - Payload vs. Launch Site
    - Flight Number vs. Orbit Type
    - Payload vs. Orbit Type
  - Bar chart: Success Rate vs. Orbit Type.
  - Line chart: Yearly Launch Success Trend.

- **Purpose:**
  - To identify patterns and relationship between launch variables.

- **Insights:**
  - Clear trends between flight progression and success rates
  - Relationships between payload mass and landing outcome

- Reference:
  - Amir-Zubair-Khalid/Data_Science_Capstone

EDA with SQL

- **SQL Queries Performed:**
  - Retrieved unique launch sites and filtered records (e.g., sites beginning with "CCA")
  - Calculated total payload mass for NASA launches
  - Calculated average payload for F9 v1.1
  - Identified first successful ground landing date
  - Listed boosters with payload between 4000 and 6000 for successful landing
  - Calculated total number of successful and failed mission outcomes
  - Listed boosters with maximum payload
  - Listed the failed outcomes, booster version, launch site for 2015
  - Ranked the launch outcomes by count between a date range
- Key Insights: Identified launch site distribution and performance differences.
- Reference:
  - Amir-Zubair-Khalid/Data_Science_Capstone

Build an Interactive Map with Folium

- Map Components: Added markers, circles, and lines to visualize launch site locations.
    - (e.g., CCAFS LC-40, VAFB SLC-4E, KSC LC-39A)
- Color-coded markers indicating launch success (green) and failure (red).
- Calculated distances with distance lines between launch sites, showing proximities (coastline, highway, railway).
- Purpose: Visualize geographic patterns and assess how proximity factors might impact launch success

- Reference:
    - [Amir-Zubair-Khalid/Data_Science_Capstone]

Build a Dashboard with Plotly Dash

- Dashboard Components:
  - Dropdown: Allows selection of “All Sites” or a specific launch site.
  - Pie Chart: Displays overall or site-specific success vs. failure counts.
  - Range Slider: Filters payload mass range.
  - Scatter Plot: Shows correlation between payload mass and launch outcome, color-coded by Booster Version Category.
- Interactions: Dynamic filtering that updates visualizations in real time
- Purpose: Provide an interactive platform for stakeholders to explore launch performance metrics

- Reference:
  - Amir-Zubair-Khalid/Data_Science_Capstone

Predictive Analysis (Classification)

- **Model Development:**
  - Built four models: Logistic Regression, SVM, Decision Tree, and KNN
  - Used GridSearchCV for hyperparameter tuning with cross-validation (cv=10)
- **Evaluation:**
  - Compared model accuracy on test data (Logistic Regression, SVM, Decision Tree and KNN achieved ~83% test accuracy. Analyzed confusion matrices to identify prediction errors (notably false positives)
- **Outcome:**
  - Best performing models (Logistic Regression, SVM, Decision Tree, KNN) achieved similar accuracy, demonstrating robustness in predicting first stage landing success
- **Reference:**
  - Amir-Zubair-Khalid/Data_Science_Capstone

Results

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

Section 2

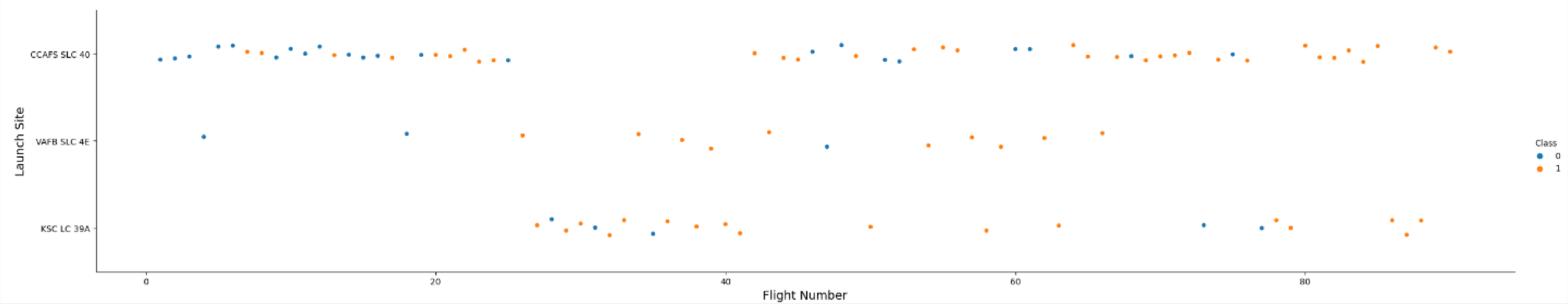
Insights drawn from EDA

Flight Number vs. Launch Site

- **Chart:** Scatter plot of Flight Number (x-axis) vs. Launch Site (y-axis) with markers colored by landing outcome (class).

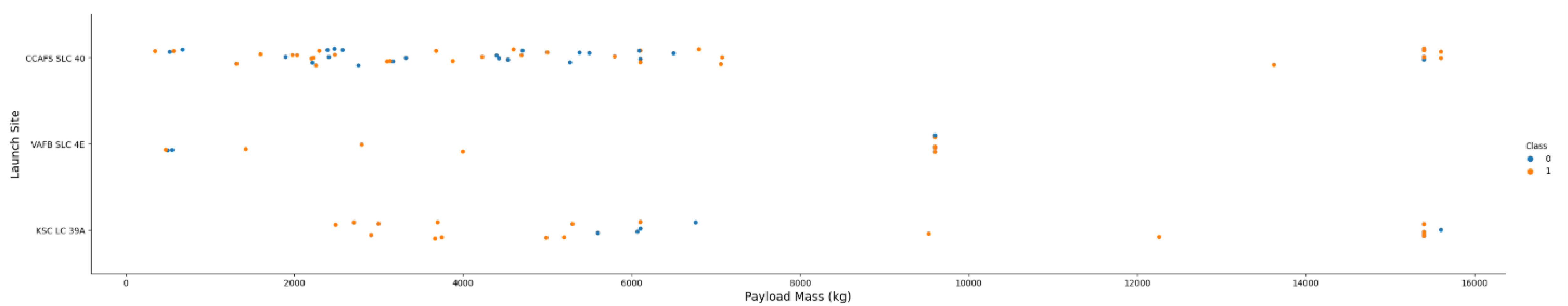
- **Key Insights:**
  - Higher flight numbers generally correspond with improved success at specific sites.
  - Trends suggest that as SpaceX gained experience, certain launch sites (e.g., CCAFS LC-40) demonstrated better outcomes.

- **Explanation:** This plot shows temporal progress across sites and highlights operational improvements.



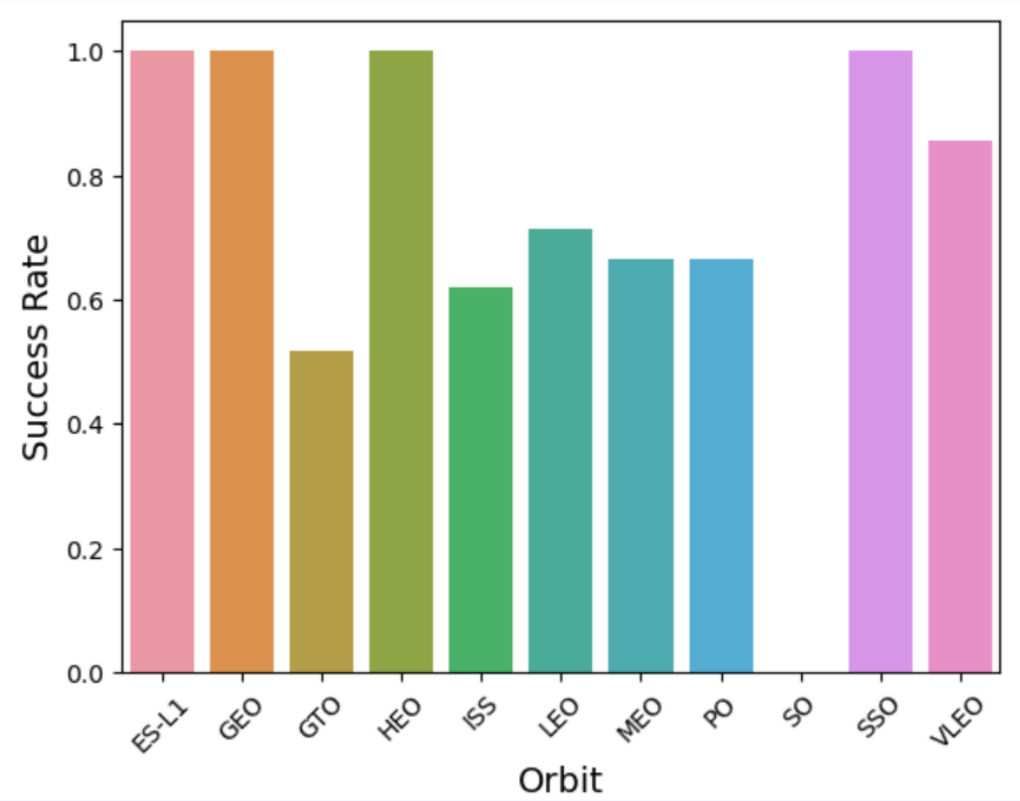
Payload vs. Launch Site

- **Chart:** Scatter plot of Payload Mass (kg) (x-axis) vs. Launch Site (y-axis) with marker color representing landing outcome.
- **Key Insights:**
  - Variations in payload mass across launch sites can affect landing performance.
  - Certain site, such as CCAFS SLC 40, handle heavier payloads with higher success rates compared to others.
- **Explanation:** This analysis helps identify which sites are optimized for high-mass launches.



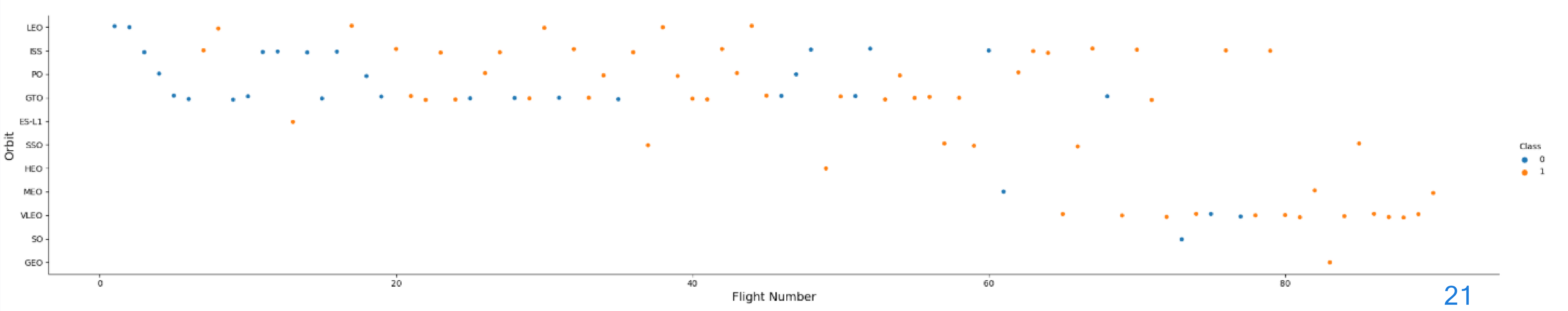
Success Rate vs. Orbit Type

- **Chart:** Bar chart showing the average success rate (mean of “class”) for each orbit type.
- **Key Insights:**
  - Orbits such as LEO and ISS show higher success rates.
  - Variations in performance across orbit types may reflect differing mission complexities.
- **Explanation:** This visualization reveals which orbital missions are most reliably executed.



Flight Number vs. Orbit Type

- **Chart:** Scatter plot with Flight Number (x-axis) and Orbit type (y-axis) with markers colored by landing outcome.
- **Key Insights:**
  - A clear trend emerges for LEO missions, where success increases with flight number.
  - For other orbits (e.g., GTO), the relationship is less pronounced.
- **Explanation:** The plot illustrates how flight experience impacts mission success differently depending on the orbit.



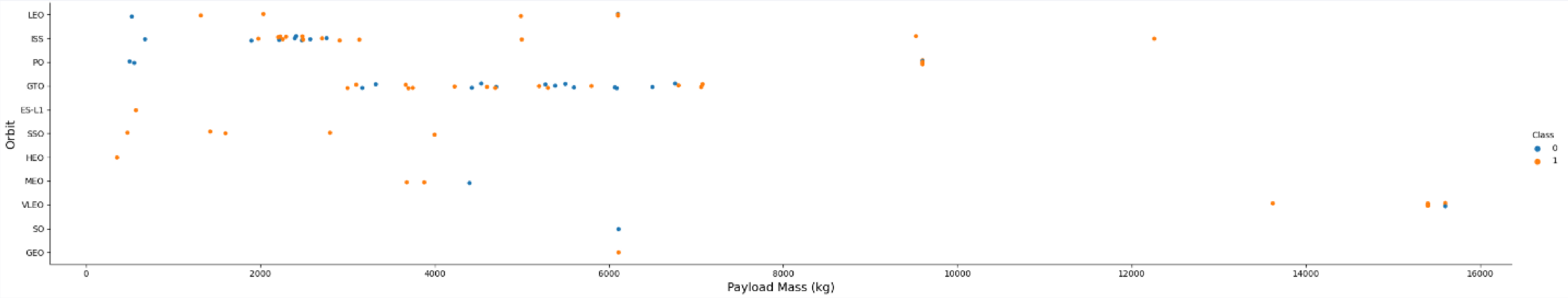
Payload vs. Orbit Type

•**Chart:** Scatter plot of Payload Mass (kg) (x-axis) vs. Orbit type (y-axis) with outcome indicated by color.

•**Key Insights:**

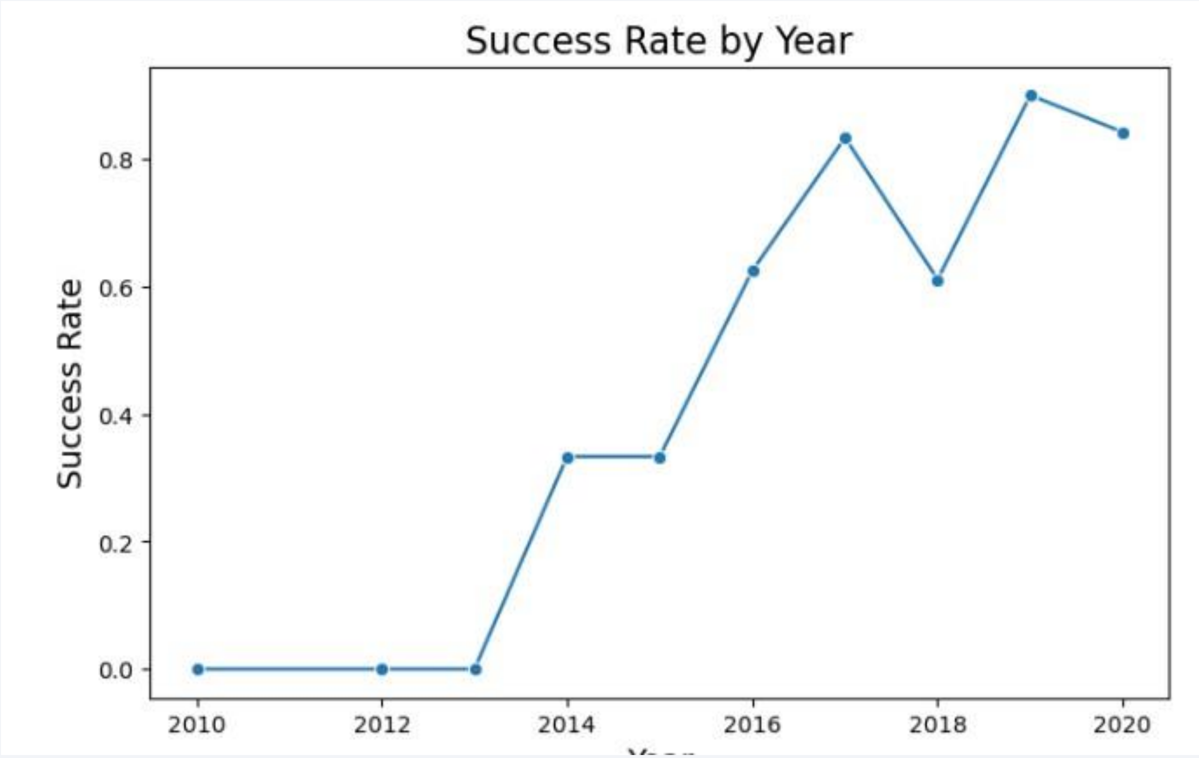
- •Heavy payloads affect launch success variably across orbit types.
- •Specific orbits may have operational thresholds for payloads that optimize success.

Explanation: This chart supports decision-making regarding payload allocations per orbit.



Launch Success Yearly Trend

- **Chart:** Line chart showing the yearly average success rate (grouped by Year from the Date field).
- **Key Insights:**
  - A notable upward trend in landing success is observed from 2013 to 2017, stabilizing post-2015.
  - This improvement aligns with technological and procedural refinements over time.
- **Explanation:** The trend demonstrates SpaceX's increasing reliability in launch recovery.



All Launch Site Names

Query

SELECT DISTINCT LAUNCH_SITE FROM SPACEXTBL;

- **Result:**
  - Identified unique sites: “CCAFS LC-40”, “VAFB SLC-4E”, “KSC LC-39A”, “CCAFS SLC-40”.

- **Explanation:**
  - This query shows the primary locations used by SpaceX, which will be used for further analysis.

```
%sql SELECT DISTINCT LAUNCH_SITE FROM SPACEXTBL;
```

```
 * sqlite:///my_data1.db
Done.
```

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

Launch Site Names Begin with 'CCA'

- **Query:**
- SELECT * FROM SPACEXTBL WHERE "Launch_Site" LIKE 'CCA%' LIMIT 5;

- **Result:**
  - Displays records for sites starting with “CCA”, indicating Cape Canaveral’s involvement.

- **Explanation:**
  - This filter highlights the subset of launch sites associated with the Cape Canaveral region.

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

```
 * sqlite:///my_data1.db
Done.
```

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Cust
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	S
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)	(C
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	(C
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	

Total Payload Mass

- **Query:**
- SELECT SUM(PAYLOAD_MASS__KG_), Customer FROM SPACEXTBL WHERE Customer = 'NASA (CRS)';
- **Result:**
  - Total payload mass (e.g., 45,596 kg) carried by boosters for NASA CRS missions.
- **Explanation:**
  - This metric illustrates the overall contribution of NASA’s CRS payloads in the dataset.

```
%%sql
SELECT SUM(PAYLOAD_MASS__KG_), Customer
FROM SPACEXTBL
WHERE Customer = 'NASA (CRS)';
```

```
 * sqlite:///my_data1.db
Done.
```

SUM(PAYLOAD_MASS__KG_)	Customer
45596	NASA (CRS)

Average Payload Mass by F9 v1.1

- **Query:**
- SELECT AVG(PAYLOAD_MASS__KG_) AS "Avg_Payload_Mass", Customer, Booster_Version FROM SPACEXTBL WHERE Booster_Version LIKE 'F9 v1.1%’;

```
%%sql
SELECT AVG("PAYLOAD_MASS__KG_")AS "Avg_Payload_Mass", Customer,
Booster_Version
FROM SPACEXTBL
WHERE Booster_Version LIKE 'F9 v1.1%';
```

```
 * sqlite:///my_data1.db
Done.
```

Avg_Payload_Mass	Customer	Booster_Version
2534.6666666666665	MDA	F9 v1.1 B1003

- **Result:**
  - The average payload mass for F9 v1.1 (e.g., approximately 2534.67 kg).

- **Explanation:**
  - This average helps in comparing performance between different booster versions.

First Successful Ground Landing Date

- **Query:**
- SELECT MIN(DATE) FROM SPACEXTBL WHERE "Landing_Outcome" = "Success (ground pad)";

- **Result:**
  - Earliest ground landing date identified as 2015-12-22.

- **Explanation:**
  - Marks a milestone in recovery technology and operational success.

```
%%sql SELECT MIN(DATE)
FROM SPACEXTBL
WHERE "Landing_Outcome" = "Success (ground pad)";
```

```
 * sqlite:///my_data1.db
Done.
```

MIN(DATE)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- **Query:**
- SELECT DISTINCT Booster_Version, Payload, PAYLOAD_MASS__KG_ FROM SPACEXTBL WHERE "Landing_Outcome" = "Success (drone ship)" AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000;

- **Result:**
  - Shows a dataframe with list of booster versions and corresponding payloads meeting the criteria.

- **Explanation:**
  - This result isolates successful drone ship recoveries under specific payload conditions, providing insight into operational limits.

```
%%sql
SELECT DISTINCT Booster_Version, Payload, "PAYLOAD_MASS__KG_"
FROM SPACEXTBL
WHERE "Landing_Outcome" = "Success (drone ship)"
    AND PAYLOAD_MASS__KG_ > 4000
    AND PAYLOAD_MASS__KG_ < 6000;
```

```
 * sqlite:///my_data1.db
Done.
```

Booster_Version	Payload	PAYLOAD_MASS__KG_
F9 FT B1022	JCSAT-14	4696
F9 FT B1026	JCSAT-16	4600
F9 FT B1021.2	SES-10	5300
F9 FT B1031.2	SES-11 / EchoStar 105	5200

Total Number of Successful and Failure Mission Outcomes

- **Query:**
- SELECT "Mission_Outcome", COUNT(*) AS Total FROM SPACEXTBL GROUP BY "Mission_Outcome";

- **Result:**
  - Breakdown of outcomes (e.g., “Success”: 98, “Failure (in flight)”: 1, etc.)

- **Explanation:**
  - Summarizes overall mission performance and reliability.

```
%%sql
SELECT "Mission_Outcome", COUNT(*) as Total
FROM SPACEXTBL
GROUP BY "Mission_Outcome";
```

```
 * sqlite:///my_data1.db
Done.
```

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

Boosters Carried Maximum Payload

- **Query:**
- SELECT Booster_Version, Payload, PAYLOAD_MASS__KG_ FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL);

- **Result:**
  - Identifies booster versions that carried the maximum payload (e.g., 15,600 kg).

- **Explanation:**
  - Highlights the upper operational limit achieved in the dataset.

```
%%sql
SELECT "Booster_Version", "Payload", "PAYLOAD_MASS__KG_"
FROM SPACEXTBL
WHERE "PAYLOAD_MASS__KG_" = (
    SELECT MAX("PAYLOAD_MASS__KG_")
    FROM SPACEXTBL);
```

```
 * sqlite:///my_data1.db
Done.
```

Booster_Version	Payload	PAYLOAD_MASS__KG_
F9 B5 B1048.4	Starlink 1 v1.0, SpaceX CRS-19	15600
F9 B5 B1049.4	Starlink 2 v1.0, Crew Dragon in-flight abort test	15600
F9 B5 B1051.3	Starlink 3 v1.0, Starlink 4 v1.0	15600
F9 B5 B1056.4	Starlink 4 v1.0, SpaceX CRS-20	15600
F9 B5 B1048.5	Starlink 5 v1.0, Starlink 6 v1.0	15600
F9 B5 B1051.4	Starlink 6 v1.0, Crew Dragon Demo-2	15600
F9 B5 B1049.5	Starlink 7 v1.0, Starlink 8 v1.0	15600
F9 B5 B1060.2	Starlink 11 v1.0, Starlink 12 v1.0	15600
F9 B5 B1058.3	Starlink 12 v1.0, Starlink 13 v1.0	15600
F9 B5 B1051.6	Starlink 13 v1.0, Starlink 14 v1.0	15600
F9 B5 B1060.3	Starlink 14 v1.0, GPS III-04	15600
F9 B5 B1049.7	Starlink 15 v1.0, SpaceX CRS-21	15600

2015 Launch Records

- **Query:**
- SELECT substr("Date", 6, 2) AS "Month", substr("Date", 1, 4) AS "Year", Booster_Version, Launch_Site, Payload, PAYLOAD_MASS__KG_, Mission_Outcome, Landing_Outcome FROM SPACEXTBL WHERE substr(Date,1,4)='2015' AND "Landing_Outcome" = 'Failure (drone ship)';

- **Result:**
  - Displays 2015 records where drone ship landing attempts failed.

- **Explanation:**
  - Provides insight into specific challenges during 2015 operations.

```
%%sql
SELECT substr("Date", 6, 2) AS "Month", substr("Date", 1, 4) AS "Year",
"Booster_Version", "Launch_Site", "Payload", "PAYLOAD_MASS__KG_",
"Mission_Outcome", "Landing_Outcome"
FROM SPACEXTBL
WHERE substr(Date,1,4)='2015'
    AND "Landing_Outcome" = 'Failure (drone ship)';
```

```
 * sqlite:///my_data1.db
Done.
```

Month	Year	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Mission_Outcom
01	2015	F9 v1.1 B1012	CCAFS LC-40	SpaceX CRS-5	2395	Succes
04	2015	F9 v1.1 B1015	CCAFS LC-40	SpaceX CRS-6	1898	Succes

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

- **Query:**
- SELECT "Landing_Outcome", COUNT(*) AS "Count" FROM SPACEXTBL WHERE "Date" BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY "Landing_Outcome" ORDER BY "Count" DESC;

- **Result:**
  - Ranks different landing outcomes (e.g., “No attempt” being the most frequent, followed by “Success (drone ship)” and others).

- **Explanation:**
  - This ranking helps identify the most common outcomes and operational trends over the specified period.

```
%%sql
SELECT "Landing_Outcome",
       COUNT(*) AS "Count"
FROM SPACEXTBL
WHERE "Date" BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY "Landing_Outcome"
ORDER BY "Count" DESC;
```

```
 * sqlite:///my_data1.db
Done.
```

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

Section 3

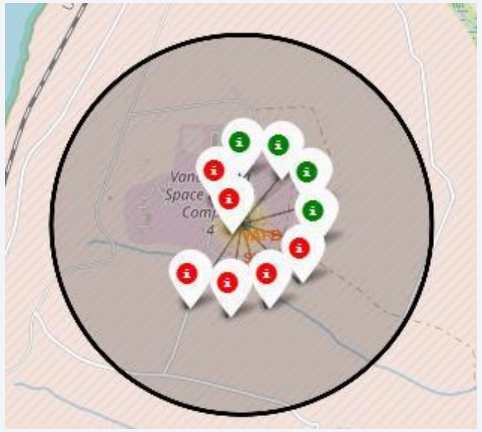
Launch Sites Proximities Analysis

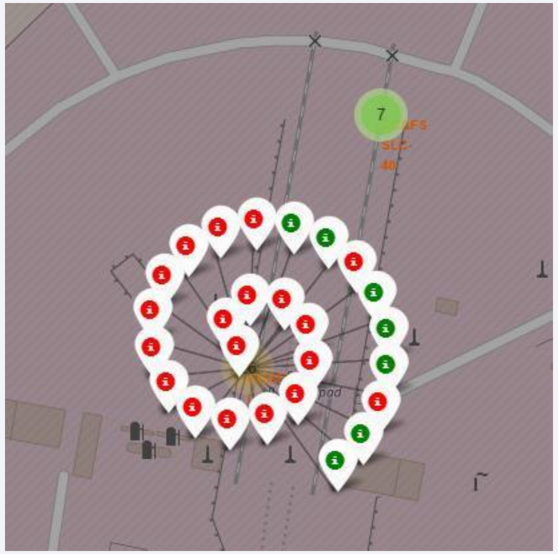
<Global Launch Sites – Location Markers>

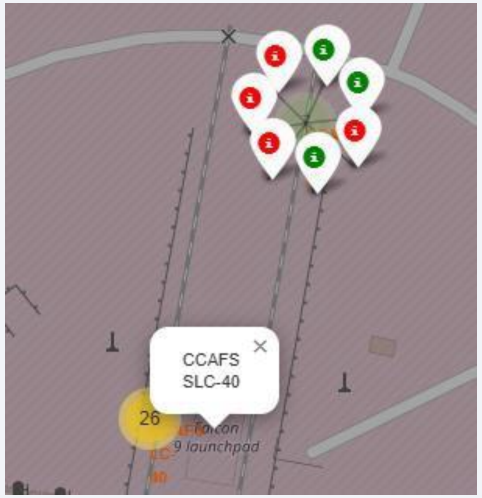
- **Description:**
  - A Folium map was created with NASA JSC as the initial center to ensure a global view.
  - All SpaceX launch sites were marked using circles and text labels.
  - Markers include key sites such as “CCAFS LC-40”, “VAFB SLC-4E”, and “KSC LC-39A”.
  - The map provides a clear visualization of the geographic distribution of launch facilities.
- **Explanation:**
  - The markers and accompanying circles illustrate the strategic placement of launch sites along the U.S. coasts and near the equator, emphasizing operational advantages like proximity to the ocean and favorable launch trajectories.

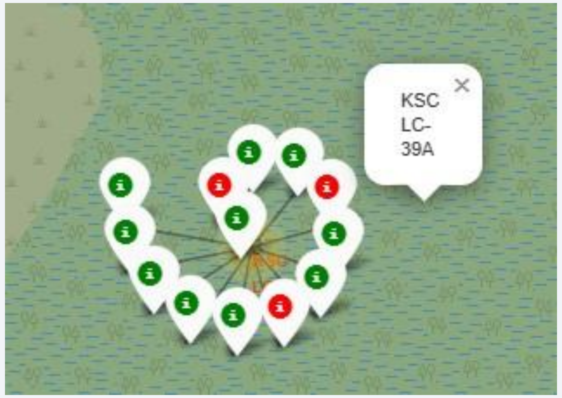
<Interactive Map – Launch Outcome Distribution>

- **Description:**
  - An enhanced Folium map uses a MarkerCluster to display individual launch records.
  - Each marker is color-coded: green indicates a successful landing (class=1) and red indicates a failure (class=0).
  - The clustering technique ensures that overlapping points are aggregated, offering a cleaner view.
  - Interactive popups display the launch site name and outcome, allowing detailed exploration.
- **Explanation:**
  - The color-coded markers reveal that the majority of launches at each site are successful.
  - This visual differentiation supports further analysis of performance trends by location.





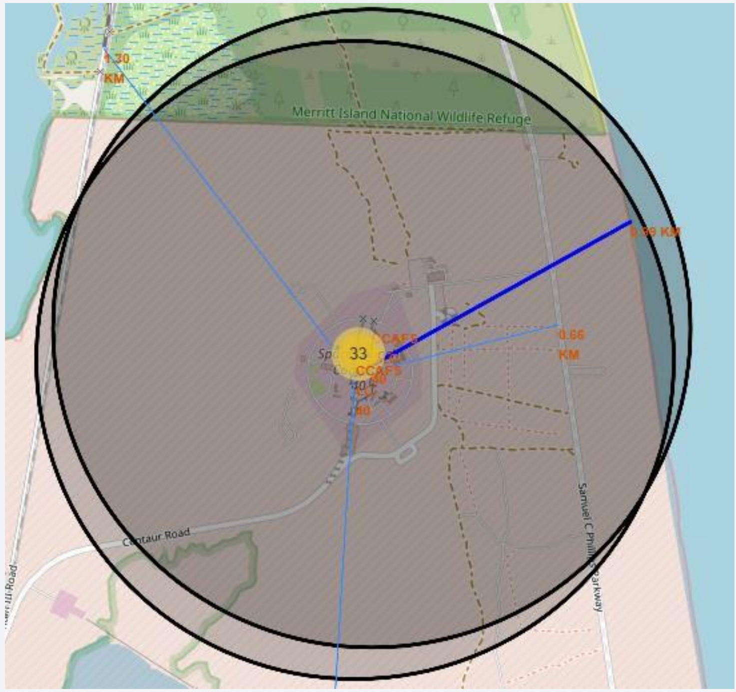




<Proximity Analysis for a Selected Launch Site>

- **Description:**
- Focused map view of a selected launch site (e.g., CCAFS LC-40) with additional markers.
- Markers for nearby features (coastline, highway, railroad, city) were added using calculated distances.
- Example distances: Coastline ~0.99 km; Highway ~0.66 km; Railroad ~1.30 km; Nearest city ~53.9 km.
- Lines (polyline objects) connect the launch site to each proximity marker to visually indicate distances.

- **Explanation:**
- The proximity analysis demonstrates that the launch site is optimally located near vital infrastructure (coastlines, highways, railways) while maintaining a safe distance from urban areas.
- Such placement supports operational safety and efficiency in launch and recovery operations.

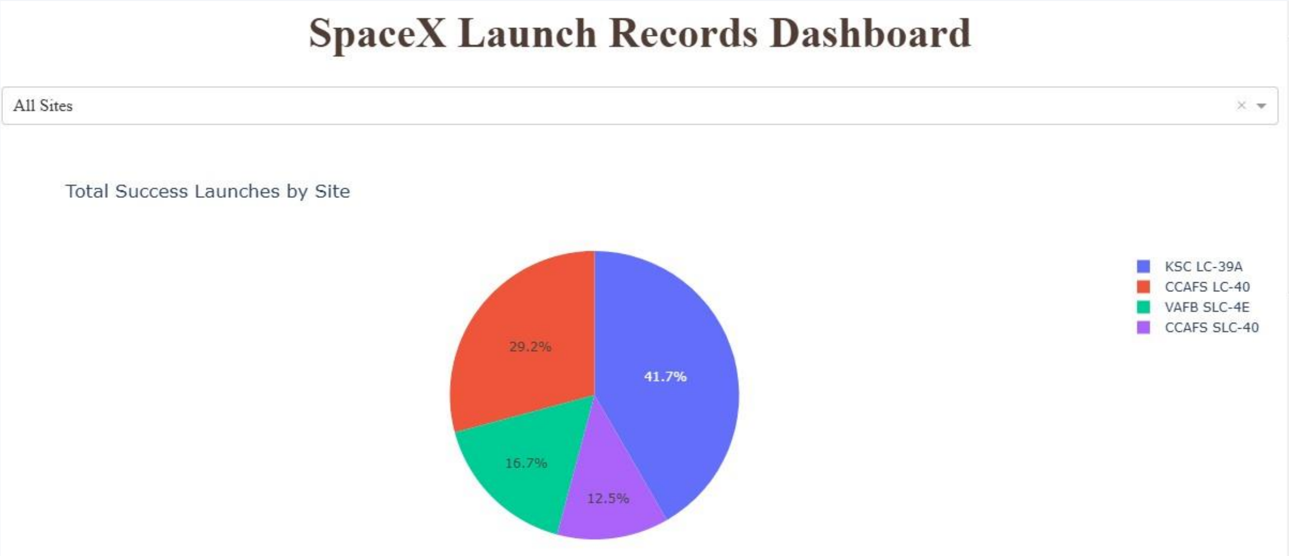


Section 4

Build a Dashboard with Plotly Dash

<Launch Success Count - All Sites>

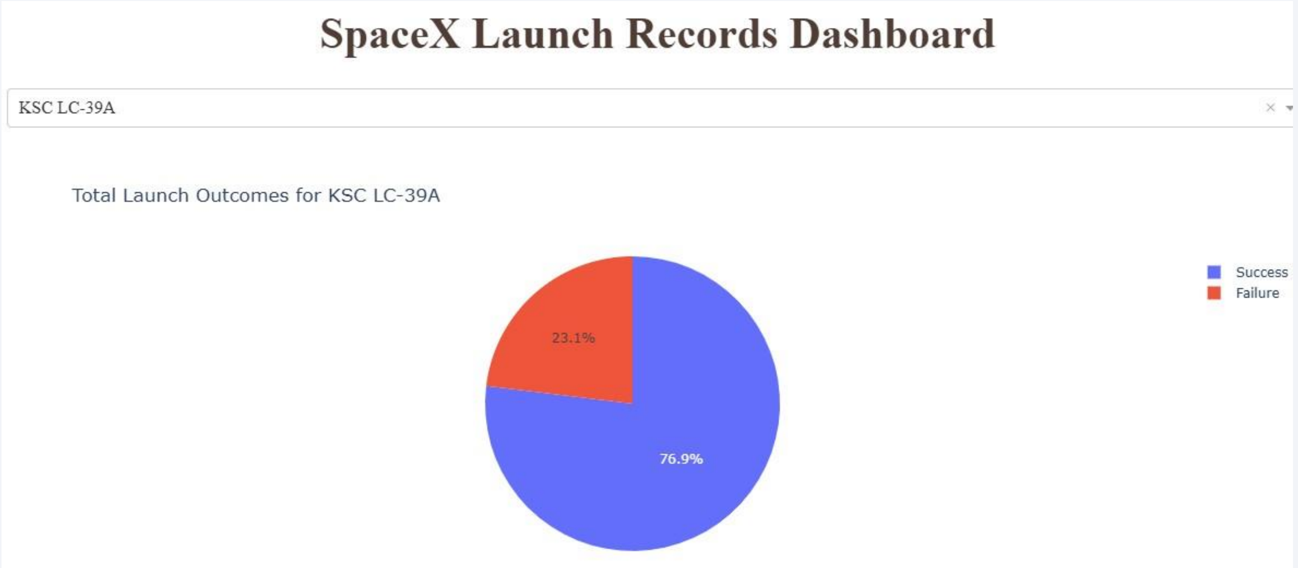
- **Description:**
  - The dashboard opens with a dropdown (default “ALL”) that aggregates data across all SpaceX launch sites.
  - The visualization supports a high-level assessment of performance across the network.



- **Explanation:**
  - KSC LC-39A had the highest number of successful launches across all sites.
  - The pie chart shows the relative contribution of each launch site to SpaceX’s overall success.
  - It serves as a summary of site performance before drilling down into individual site metrics.

<Detailed Launch Outcomes for Selected Site>

- **Description:**
- When a specific launch site is chosen from the dropdown, the pie chart updates to show the breakdown of successes versus failures.
- For example, selecting a site with a high success ratio displays a dominant “Success” segment with a smaller “Failure” portion.
- This targeted view allows for deeper insights into the performance metrics of each site.
- **Explanation:**
- KSC LC-39A has the highest success rate of 76.9%, among all sites



<Payload vs. Launch Outcome – Interactive Analysis>

- **Description:**
- An interactive scatter plot correlates payload mass (x-axis) with launch outcome (y-axis; 0 = Failure, 1 = Success).
- Data points are color-coded by Booster Version Category, revealing trends across different hardware iterations.
- A payload range slider allows dynamic filtering, enabling analysis of specific mass ranges.
- This scatter plot shows relationships between payload size, across booster version, and landing success.

Explanation:

There are more failures than success. Launches that are successful tend to be on the light range, before 6000kg. The FT category also has the most success, followed by B4, and v1.1

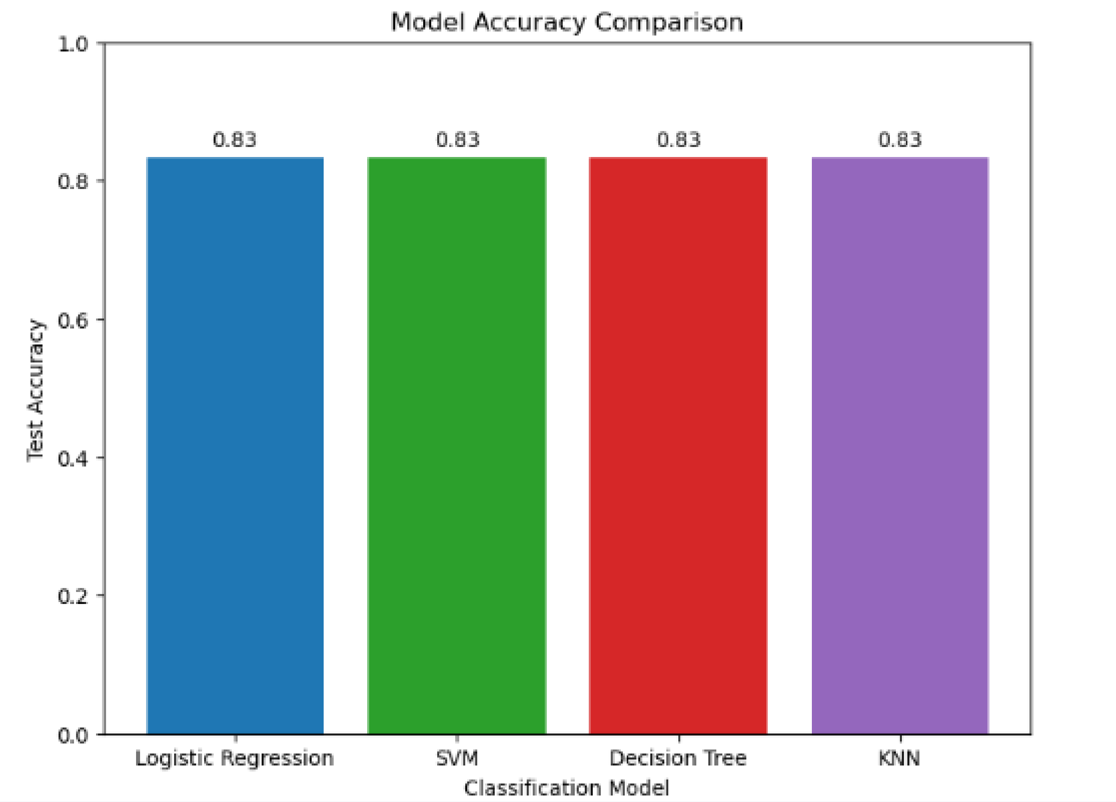
This supports strategic decisions by identifying optimal payload ranges and booster version that maximize successful outcomes.

Section 5

Predictive Analysis (Classification)

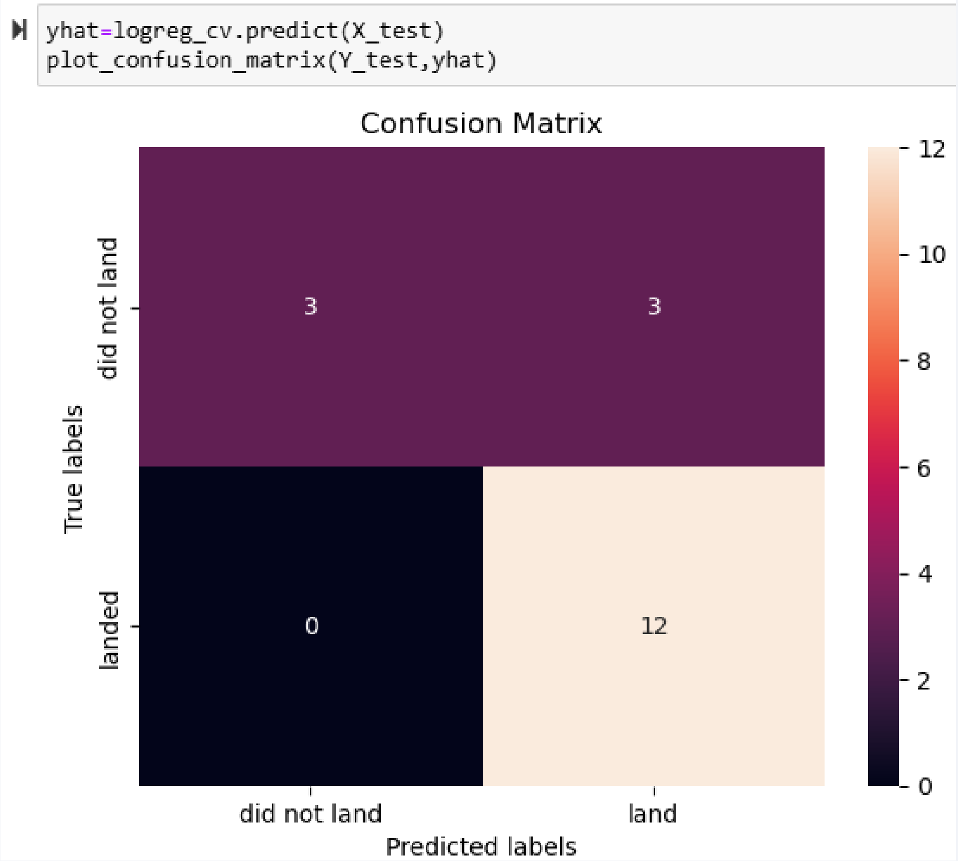
Classification Accuracy

- **Description:** Four classification models were built and tuned using GridSearchCV:
  - **Logistic Regression:** Test accuracy = 83.33%
  - **Support Vector Machine (SVM):** Test accuracy = 83.33%
  - **K-Nearest Neighbors (KNN):** Test accuracy = 83.33%
  - **Decision Tree:** Test accuracy = 83.33%
- Models were developed on standardized features with hyperparameters optimized via cross-validation.
- Logistic Regression, SVM, Decision Tree model and KNN all performed similarly.
- **Explanation:**
- The bar chart clearly shows that four models tie for the high accuracy, indicating a robust predictive signal in the features.



Confusion Matrix

- **Description:**
- All four models have the same confusion matrix. This includes:
  - **True Positives:** High count (12)
  - **False Positives:** Low count (3)
  - The model over predicted successful landing, counting 3 unsuccessful landing as landed.
  - **True negatives:**
    - For 3 launches, the models predicted unsuccessful landings when the true label was unsuccessful landing.
    - The models predicted 0 for did not land when there was successful landing.



Conclusions

- **Evaluation Process:**
  - Data were standardized and split into training/test sets.
  - Hyperparameters were optimized using GridSearchCV.
  - Multiple models were compared using test accuracy and confusion matrices.
- **Key Findings:**
  - Logistic Regression, SVM, Decision Tree model and KNN achieved 83.33% accuracy.
  - The confusion matrix for the best model shows a low false positive rate.
- **Future Improvements:**
  - Incorporate additional features and more historical data.
  - Explore ensemble methods to further boost accuracy.
  - Fine-tune feature engineering based on domain insights from launch metrics.

- **Explanation:** This summary demonstrates that predictive modeling can effectively forecast Falcon 9 landing outcomes. Ongoing enhancements in feature selection and model complexity may further improve performance.

Appendix

- Amir-Zubair-Khalid/Data_Science_Capstone

Thank you!