IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Rodrick Kazembe
2025-06-12

Outline

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

Executive Summary

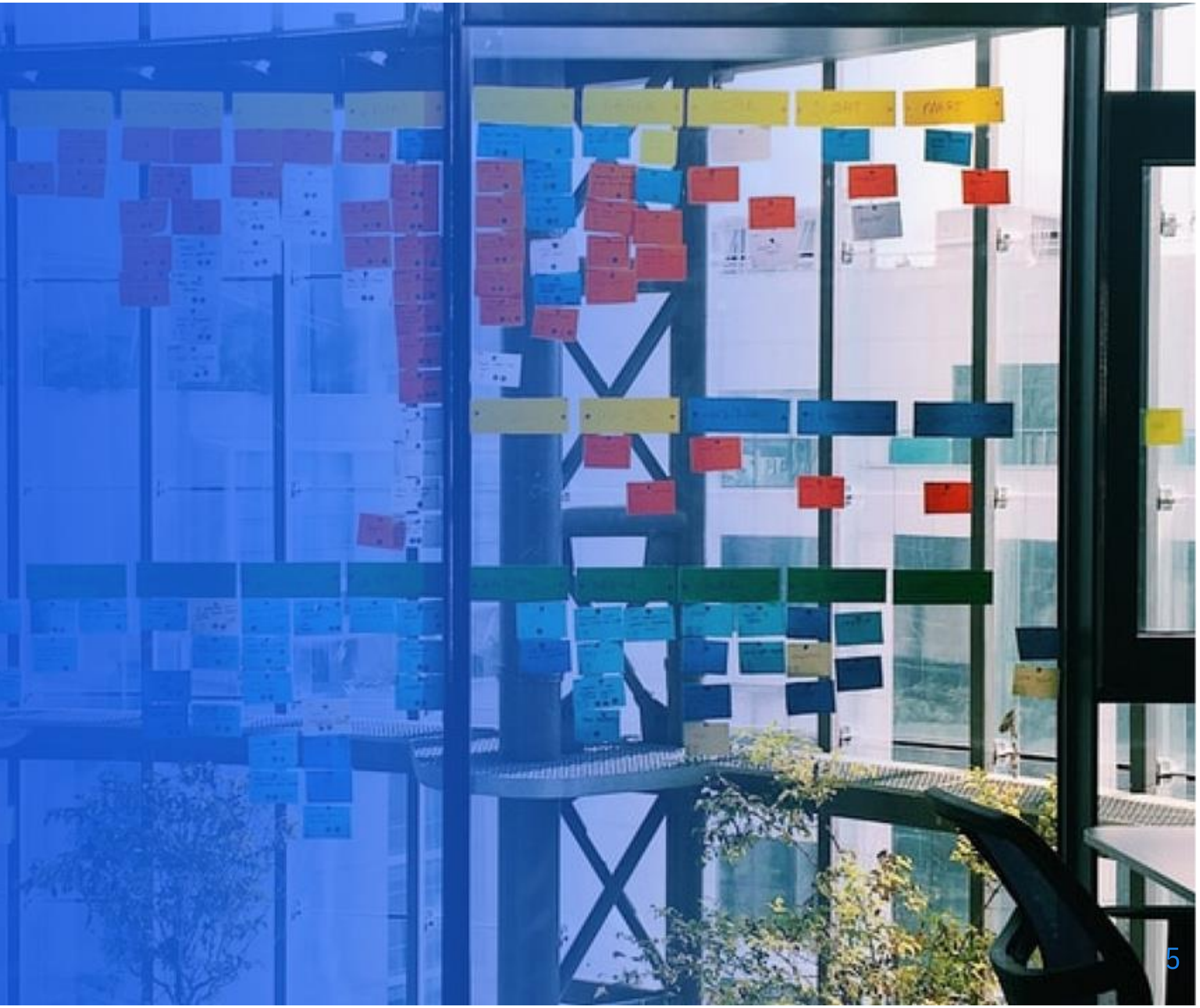
- Summary of methodologies
  - Collected SpaceX Falcon 9 launch data using API and Wikipedia scraping.
  - Cleaned and prepared data: handled missing values, filtered relevant launches.
  - Engineered features using One-Hot Encoding on categorical data. Visualized
  - trends using Seaborn and Plotly (e.g., orbit success, payload effects). Developed
  - models: Logistic Regression, SVM, KNN, and Decision Tree. Tuned
  - hyperparameters using GridSearchCV. Evaluated performance with metrics such
  - as accuracy and confusion matrix.
- Summary of all results
  - Best model: Decision Tree Classifier with 93.33% accuracy on test data. Best SVM
  - kernel: rbf — performed well but slightly lower accuracy. Success is highly influenced
  - by PayloadMass, Orbit, and LaunchSite. Dashboard built using Plotly Dash with
  - interactive filters for launch outcome. Data trends confirm that Falcon 9 success rate
  - has significantly improved over time.

Introduction

- Project background and context
  - SpaceX, a leading aerospace company, has transformed space launch economics through reusable rocket technology. Predicting the success of rocket landings is critical for cost reduction and mission reliability. This project analyzes SpaceX Falcon 9 launch data to understand success factors and build predictive models for landing outcomes.
- Problems you want to find answers
  - What features most influence Falcon 9 landing success? Can we accurately predict
  - the outcome of a rocket landing? How do variables like payload, launch site, and
  - orbit type affect mission success?
  - What trends or insights can guide future mission planning?

Section 1

Methodology



Methodology

Executive Summary •Data

collection methodology:

- Used a public SpaceX API to retrieve Falcon 9 mission launch data via requests.get() and normalized the response with pd.json_normalize().
- Supplemented with web-scraped data from Wikipedia and other sources using BeautifulSoup for revenue, orbit, and outcome metadata.

•Perform datawrangling

- Cleaned null values (e.g., PayloadMass) using .replace() with the mean. Filtered only Falcon
- 9 launches and engineered new features like LandingClass and Year.

•Perform exploratory data analysis (EDA) using visualization andSQL

- Explored trends in launch success by site, orbit, payload mass, and booster version using SQL queries and Seaborn plots.

Methodology (continued)

Executive Summary •Perform interactive visual analytics using

Folium andPlotlyDash

- Used Folium to map launch site locations and display outcomes and distances.
- Created a Plotly Dash dashboard with dropdowns, pie charts, and scatter plots to explore payload vs. success visually.

•Perform predictive analysis using classification models

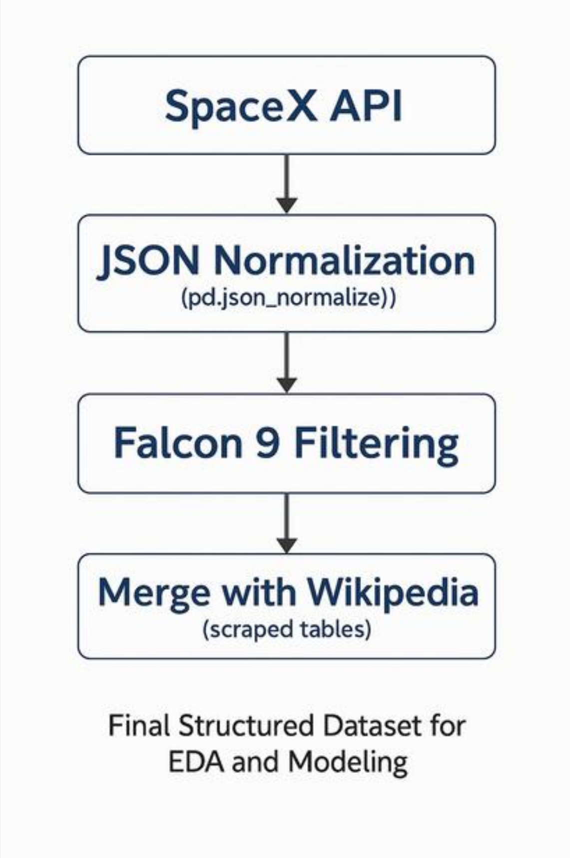
- Trained and tested multiple models: Logistic Regression, SVM, Decision Tree, KNN.
- Used pipelines and hyperparameter tuning with GridSearchCV. Evaluated models
- using accuracy and R^2 scores on validation and test datasets.

Data Collection

- Describe how data sets were collected.
  - SpaceX REST API
    - Used requests.get() to retrieve JSON launch data programmatically.
  - Wikipedia
    - Used BeautifulSoup to scrape launch outcome tables and financial data (e.g., revenue).
  - Kaggle Dataset (King County House Sales)
    - Downloaded and pre-processed as part of the housing price prediction.
- You need to present your data collection process use key phrases and flowcharts

Data Collection (continued)

You need to present your data collection process use key phrases and flowcharts

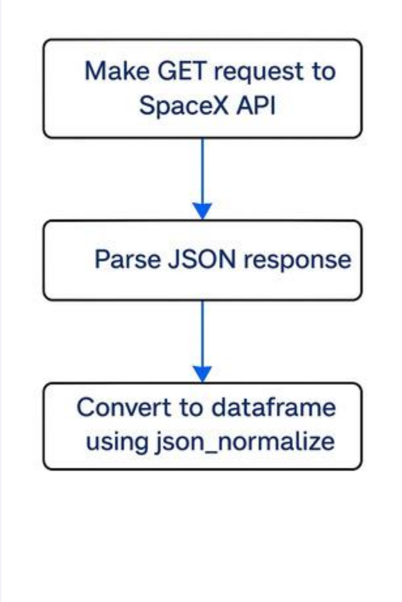


Data Collection – SpaceX API

- Present your data collection with SpaceX REST calls using key phrases and flowcharts

- Add the GitHub URL of the completed SpaceX API calls notebook (https://github.com/rddh13/ibm_capstone_project/blob/main/Automobile_Sales_Recession_Analysis.ipynb), as an external reference and peer-review purpose

Place your flowchart of SpaceX API calls here

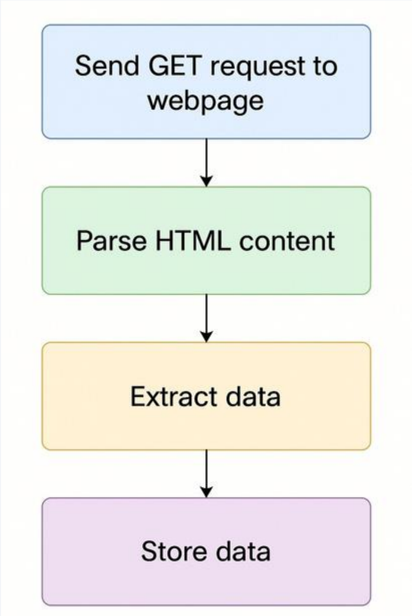


Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts

- Add the GitHub URL of the completed web scraping notebook, as an external reference and peer-review purpose
  https://github.com/rddh13/ibm_capstone_project/blob/main/Automobile_Sales_Recession_Analysis.ipynb

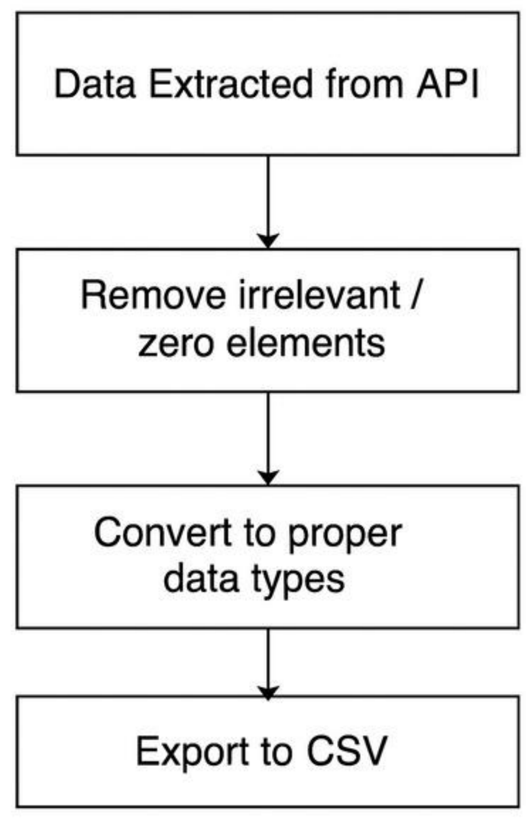
Place your flowchart of web scraping here



Data Wrangling

- Describe how data were processed
  - Dropped irrelevant columns (id, Unnamed: 0)
  - Handled missing values using .mean() and .replace()
  - Normalized JSON API responses with json_normalize()
  - Cleaned and formatted scraped HTML tables
  - Extracted features (e.g., year from date, landing success label) Applied one-hot encoding to
  - categorical columns Saved cleaned datasets for
  - analysis and modeling

https://github.com/rddh13/ibm_capstone_project/blob/main/labs-jupyter-spacex-Data%20wrangling-v2.ipynb



EDA with Data Visualization

- Summarize what charts were plotted and why you used those charts
    - Scatter Plots FlightNumber vs LaunchSite: To identify launch trends by site
    - PayloadMass vs Orbit: To analyze success patterns across orbits
    - Bar Charts
    - Orbit vs Success Rate: To compare average success rates per orbit
    - Launch Site vs Number of Launches: To visualize site frequency
    - Line Chart
    - Year vs Launch Success Rate: To observe yearly success trends
    - Box Plot
    - Payload Mass vs Outcome: To detect outliers and variation in successful
    - missions
    - Pie Chart (via Dash Callback)
    - Success vs Failure distribution per launch site
- https://github.com/rddh13/ibm_capstone_project/blob/main/jupyter-labs-eda-dataviz-v2.ipynb

EDA withSQL

- Using bullet point format, summarize the SQL queries you performed
  - Queried unique launch sites from the dataset Retrieved first 5 records for sites starting with ‘CCA’
  - Calculated total payload mass for NASA (CRS) launches Computed average payload mass for
  - booster version F9 v1.1 Identified date of first successful ground pad landing Listed booster
  - names with successful drone ship landings and payloads between 4000–6000 kg Aggregated
  - mission outcome counts (success/failure)
  - Used LIMIT to preview rows, and MIN()/MAX() for mass filtering
  - Grouped by launch site to count mission outcomes
  -
  -

https://github.com/rddh13/ibm_capstone_project/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Summar
  y
  - Added markers for each SpaceX launch site to visually indicate location Used circle markers
  - colored by mission success (green for success, red for failure) Incorporated popups with
  - launch site name for interactivity Used lines to connect launch sites with closest coastlines
  - and airports for proximity analysis Added distance labels with DivIcon to display calculated
  - distances in KM Enabled layer control and tile customization for better map readability
  -
- Explain why you added those objects
  - These elements provide a clear, interactive overview of launch locations, mission outcomes, and geospatial relationships crucial to understanding SpaceX operations.

https://github.com/rddh13/ibm_capstone_project/blob/main/lab-jupyter-launch-site-location-v2.ipynb

Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard
  - Pie Chart: Displays success counts by launch site; helps understand distribution of successful launches
  - Scatter Plot: Shows correlation between payload mass and success rate; used to analyze influence of mass on outcomes
  - Dropdown Menu: Allows selection of specific launch sites for targeted insights
  - Range Slider: Enables filtering by payload mass range; supports dynamic data exploration
  - Real-Time Interactivity: Updates visualizations instantly based on user inputs for exploratory analysis
- Explain why you added those plots and interactions
  - These interactive elements allow users to explore mission success patterns across different payloads and sites, enabling informed decision-making with clear visual feedback.

https://github.com/rddh13/ibm_capstone_project/blob/main/app.py

Predictive Analysis (Classification)

- Model Building:
  - Created multiple classification models using Logistic Regression, SVM, Decision Trees, and KNN.
- Hyperparameter Tuning:
  - Used GridSearchCV for tuning parameters like kernel type (for SVM), regularization strength (alpha), and tree depth.
- Evaluation Metrics:
  - Evaluated models using cross-validation and performance metrics like accuracy, precision, and recall.
- Model Comparison:
  - Compared results on validation data. SVM with RBF kernel performed best with high test accuracy.
- Final Deployment:
  - Best model selected based on generalization performance to unseen data.
- https://github.com/rddh13/ibm_capstone_project/blob/main/SpaceX-Machine-Learning-Prediction-Part-5-v1.ipynb

Result
s

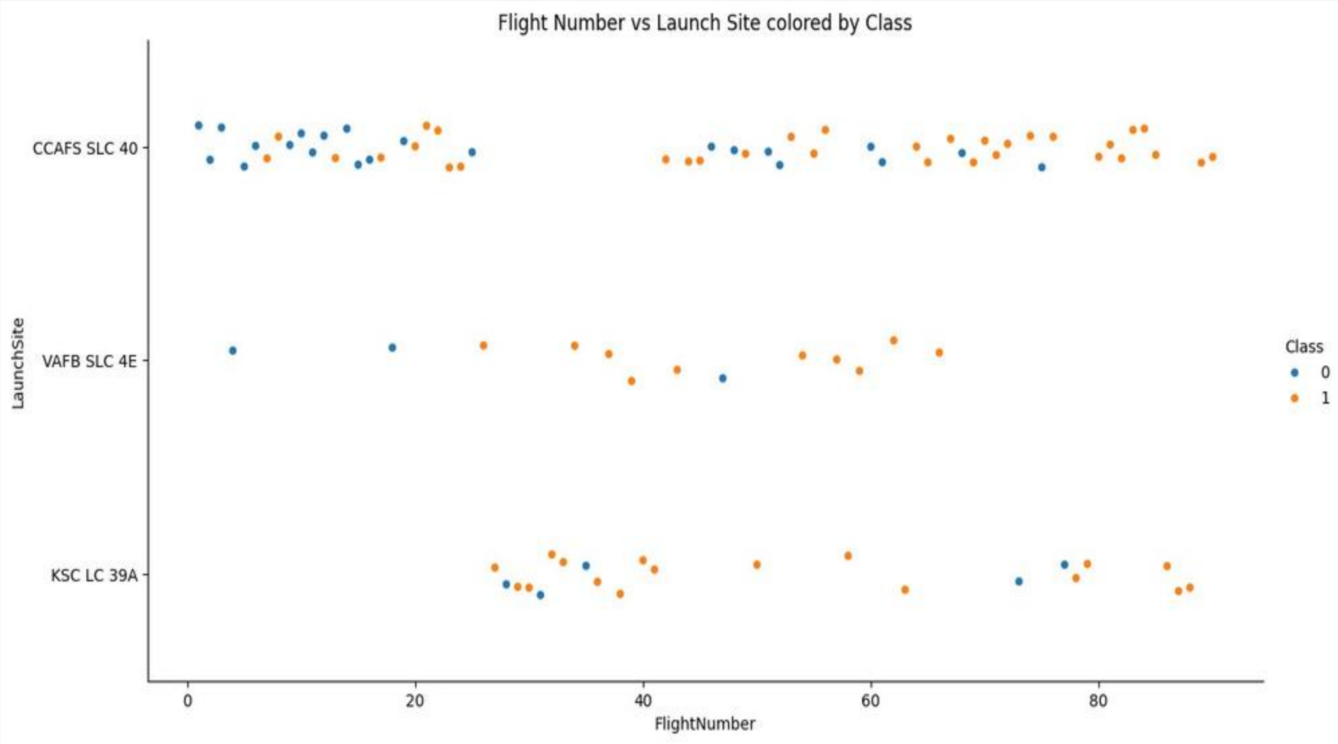
- Exploratory Data Analysis (EDA) Results
  - Identified correlations between features like Payload Mass, Orbit, and Launch Site with mission success.
  - Discovered higher success rates for missions using FT booster versions and launched from KSC LC-39A.
- Interactive Analytics (Folium + Dash)
  - Folium: Markers added for launch sites and proximity objects (e.g., railways, roads, coastlines).
  - Dash App: Included dropdowns for site selection, sliders for payload filtering, and charts for:
  - Success rate per site Correlation between payload and mission outcome
- Predictive Analysis Results
  - Best model: SVM with RBF kernel Test accuracy: ~93% Demonstrated the model's ability to predict landing success
  - based on mission features like Orbit, PayloadMass, and Booster Version.

Section 2

Insights drawn from EDA

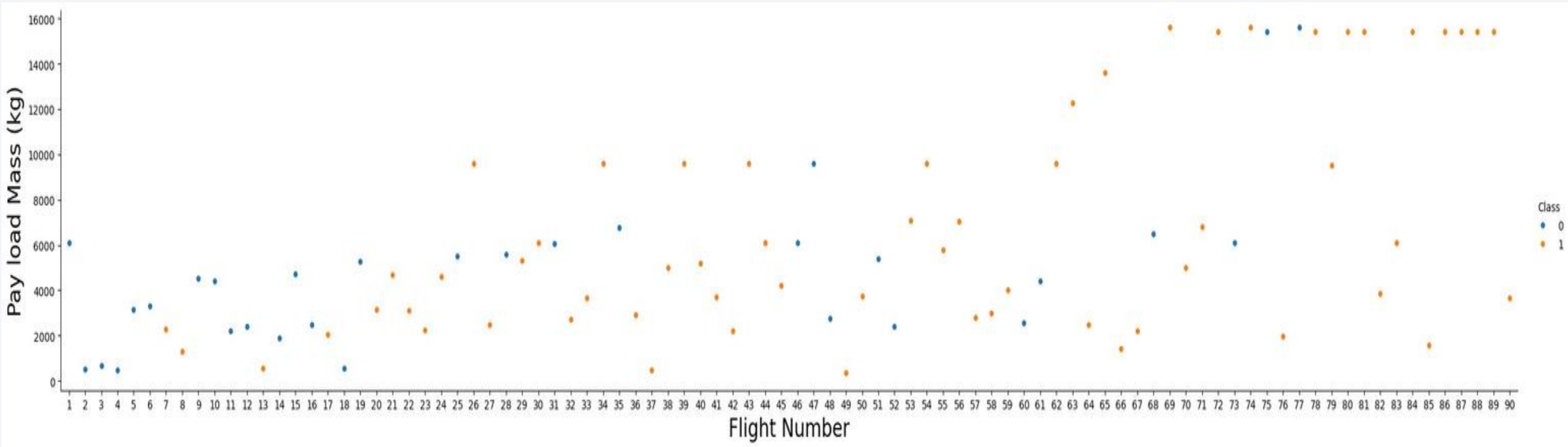
Flight Number vs. LaunchSite

- The scatter plot shows the relationship between Flight Number (x-axis) and Launch Site (y-axis), with different colors indicating the launch outcome (Class: 0 = Failure, 1 = Success).
- This visualization helps identify patterns in success rates across various launch sites over time.
- From the chart, we can observe that:
- KSC LC 39A and CCAFS SLC 40 had a high frequency of successful launches (orange dots).
- Earlier flights (lower flight numbers) tend to have more failures, while later flights show more consistency in success, reflecting model maturity and operational improvements.



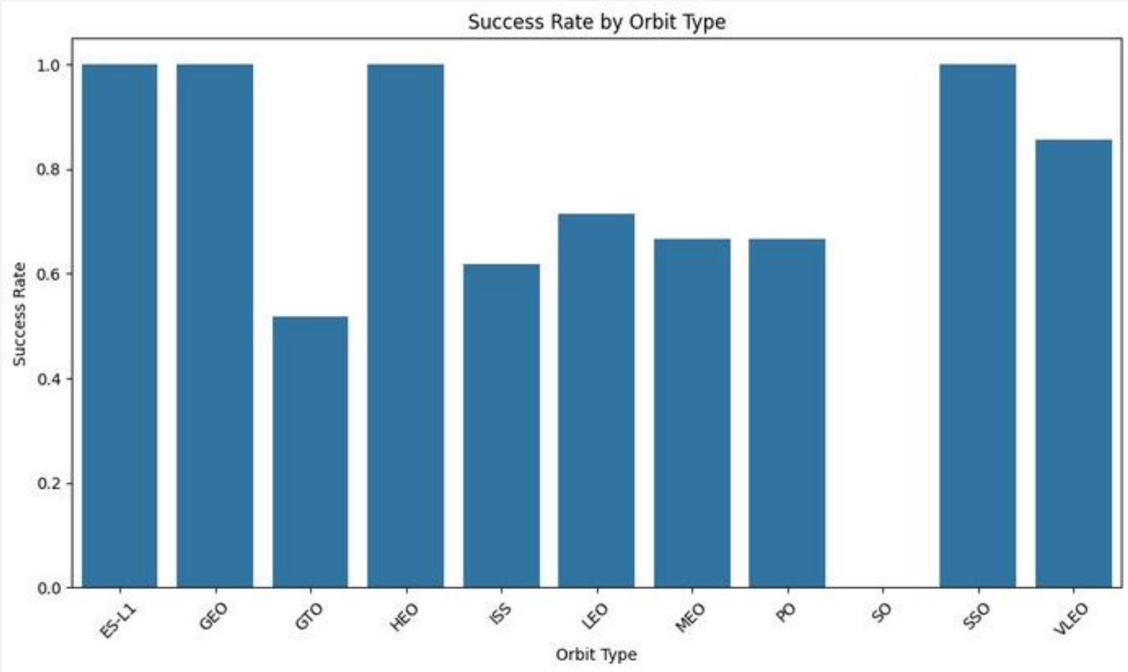
Payload vs. Launch Site

- Purpose
  :
  - This chart shows the relationship between payload mass and flight number, again with color indicating launch outcome.
- Insights
  :
  - Early missions carried lighter payloads and had more failures.
  - As the flight numbers increased, both payload capacity and mission success rate improved.
  - Heavier payloads (e.g., >10,000 kg) appear mostly successful (orange), showing growing reliability.



Success Rate vs. Orbit Type

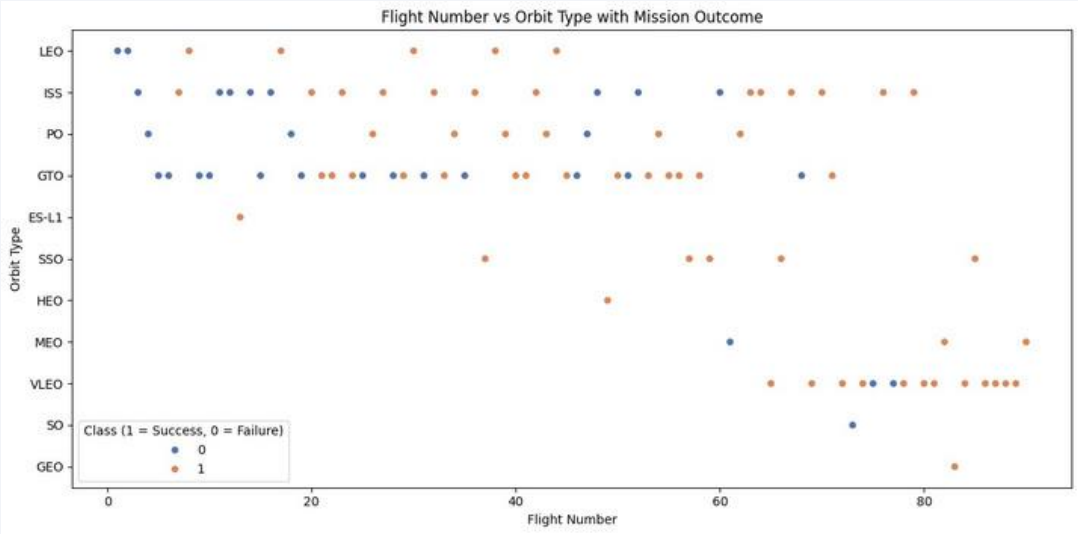
- This bar chart displays the success rate of Falcon 9 launches by orbit type. It was created by grouping launches based on orbit type and calculating the mean of the binary Class value (1 for success, 0 for failure), effectively showing the proportion of successful landings per orbit category.
- Why this chart is important:
  - It helps assess which orbit types correlate with higher mission success, providing insight for future launch planning and risk assessment. Orbits like GEO, SSO, and ES-L1 show a 100% success rate, indicating operational reliability.



Flight Number vs. Orbit Type

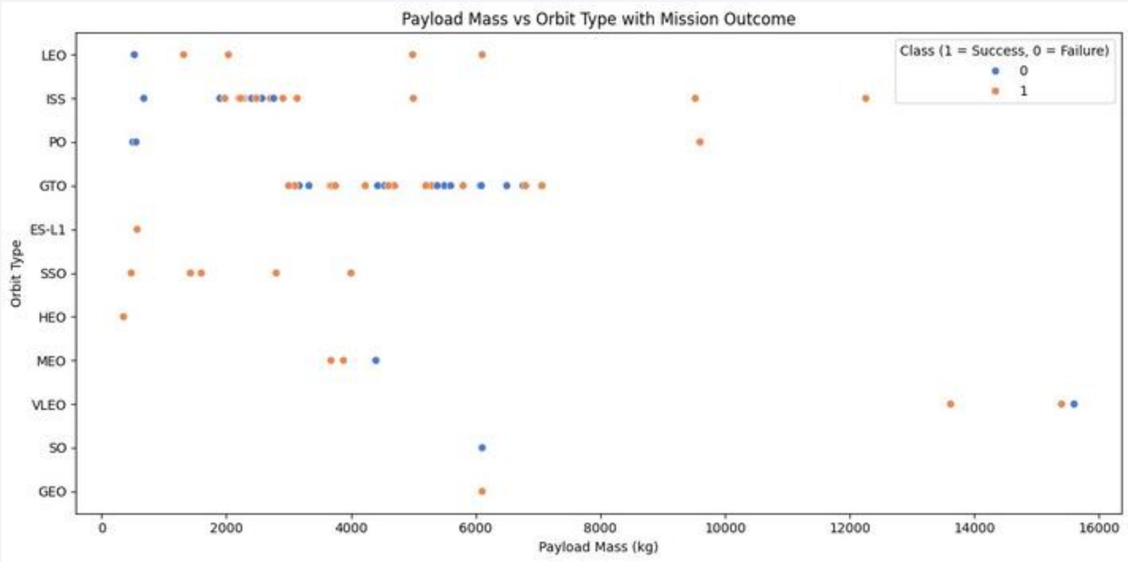
- What the chart shows:
  - Each point represents a launch attempt.
  - X-axis: Flight Number (launch
  - sequence).
  - Y-axis: Orbit Type (e.g., LEO, GTO, ISS).

  Color: Orange (1) = Successful mission
    - Blue (0) = Failed mission
- Purpose & Insights:
  - Used to visualize whether certain orbit types have more frequent success/failure.
  - We can see patterns like:
  - Missions to GTO had a mixed success rate in earlier flights.
  - Later flights generally show improved success across all orbit types.
  - LEO and VLEO seem to have high success rates in later flights.



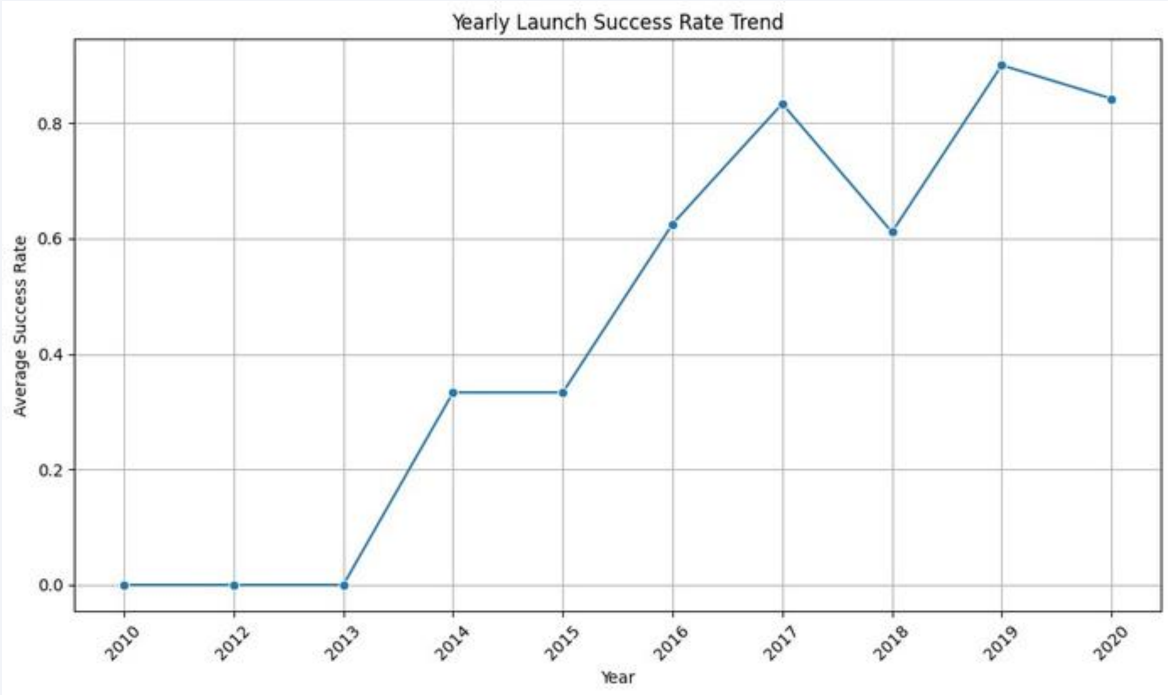
Payload vs. OrbitType

- Explana tion:
  - The scatter plot visualizes the relationship between payload mass and orbit type, colored by mission class (success = 1, failure = 0).
  - Each dot represents a SpaceX launch, showing how payloads are distributed across different orbit ca te gories.
- Insight:
  - High payload masses (above 10,000 kg) mid-range payloads for orbits like GTO, were mostly successful (orange dots).
  - Certain orbit types like VLEO and SSO show high success rates.
  - Failures tend to cluster around



Launch Success Yearly Trend

- The line chart above displays the trend of average launch success rates from 2010 to 2020. It shows a consistent improvement in SpaceX’s reliability over time, with notable increases after 2015. The dip in 2018 highlights a year with a slightly lower success rate, but the upward trend resumes in 2019. This visual confirms SpaceX’s growing mission consistency and system maturity.



All Launch Site Names

- This SQL query helps in identifying all distinct launch sites used in the SpaceX launch data. While CCAFS LC-40 appears more than once in the raw data, the DISTINCT keyword ensures it's listed only once in the result. These launch sites are key to further analysis such as performance comparison and site-specific success rates.

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

[10] ✓ 0.0s

```
* 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'

- This query filters launch records to show only those that occurred at sites starting with "CCA" (i.e., Cape Canaveral). The first 5 records show early SpaceX launches, including demo and resupply missions with outcomes such as "Success" and various landing results.

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

✓ 0.0s Python

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

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

```
%sql SELECT SUM("Payload_Mass__kg_") AS Total_Payload_Mass FROM SPACEXTABLE WHERE "Customer" = 'NASA (CRS)';

✓ 0.0s

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

Total_Payload_Mass
45596

•The total payload mass delivered to space by NASA (CRS) missions is 45,596 kg. This emphasizes the significant partnership between NASA and SpaceX in delivering cargo to the International Space Station (ISS).

Average Payload Mass by F9 v1.1

```
%sql SELECT AVG("Payload_Mass__kg_") AS Average_Payload_Mass FROM SPACEXTABLE WHERE "Booster_Version" = 'F9 v1.1';
[13]  ✓ 0.0s
 * sqlite:///my_data1.db
Done.
```

Average_Payload_Mass
2928.4

• This query calculates the mean payload mass for all missions using the F9 v1.1 booster. Based on the result, the average payload transported into orbit by this version is approximately 2928.4 kilograms, indicating its typical cargo capacity in SpaceX launch missions.

First Successful Ground Landing Date

```
%sql SELECT MIN("Date") AS First_Successful_Ground_Landing FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Success (ground pad)';
[18]  ✓  0.1s
...  * sqlite:///my_data1.db
Done.
...
First_Successful_Ground_Landing
                     2015-12-22
```

• This date was retrieved by querying the minimum "Date" from the SPACEXTABLE where "Landing_Outcome" equals 'Success (ground pad)'. It marks the historic first instance SpaceX achieved a ground pad recovery of a Falcon 9 booster, a key milestone in reusable rocket technology.

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql SELECT "Booster_Version", "Payload_Mass__kg_" FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Success (drone ship)' AND "Payload_Mass__kg_" > 400
[20] ✓ 0.0s                                                                                     Python
... * sqlite:///my_data1.db
Done.
```

Booster_Version	PAYLOAD_MASS__KG_
F9 FT B1022	4696
F9 FT B1026	4600
F9 FT B1021.2	5300
F9 FT B1031.2	5200

- These booster versions demonstrate SpaceX's capability to land high-payload missions reliably on drone ships. This success range is critical for heavy missions that still require reusability, optimizing launch economics and logistics.

Total Number of Successful and FailureMission Outcomes

- The query groups all mission outcomes and counts how many times each occurred.
- The data reveals that:
- There are 100 successes in total (98 + 1 + 1, possibly due to format variations in the labels).
- There is 1 recorded failure.
- This shows a very high success rate of SpaceX missions based on the dataset.

```
%sql SELECT "Mission_Outcome", COUNT(*) AS Outcome_Count FROM SPACEXTABLE GROUP BY "Mission_Outcome";
```

[21] ✓ 0.0s

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

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

Boosters Carried MaximumPayload

```
¦l SELECT "Booster_Version", "Payload_Mass__kg_" FROM SPACEXTABLE WHERE "Payload_Mass__kg_" = (SELECT MAX("Payload_Mass__kg_") FROM SPACEXTABLE);
```

[23] ✓ 0.0s Python

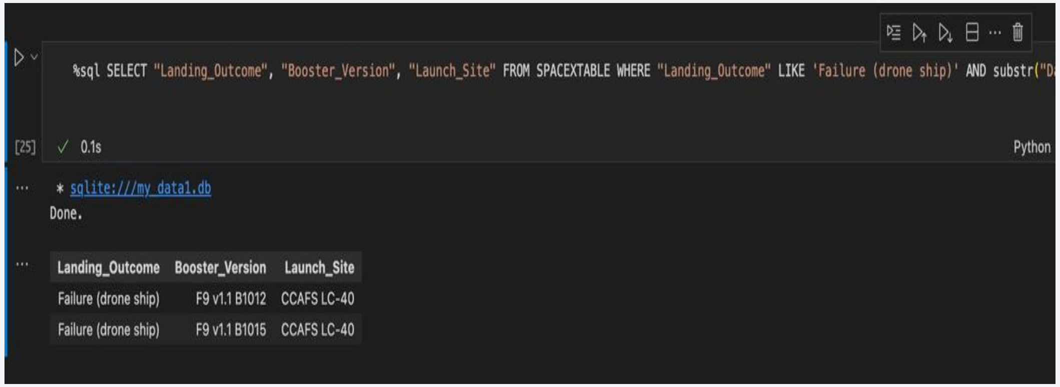
```
* sqlite:///my_data1.db
Done.
```

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

- All the boosters listed in the result—such as F9 B5 B1048.4, B1049.4, B1051.3, etc.—carried a payload of 15,600 kg, which is the maximum payload mass observed in the SPACEXTABLE database.

2015 Launch Records

- This SQL query retrieves all records from the SPACEXTABLE where:
  - The landing outcome was a failure on a drone ship.
  - The launch happened in the year 2015.
- It displays three columns:
  - Landing_Outcome (which confirms the
  - failure), Booster_Version (the version of the
  - booster used),
  - Launch_Site (the location of the launch).
- Result:
  - Two missions failed to land on the drone ship in 2015. Both were launched from CCAFS LC- 40 using booster versions F9 v1.1 B1012 and F9 v1.1 B1015.



```
%sql SELECT "Landing_Outcome", "Booster_Version", "Launch_Site" FROM SPACEXTABLE WHERE "Landing_Outcome" LIKE 'Failure (drone ship)' AND substr("D
```

[25] ✓ 0.1s Python

* sqlite:///my_data1.db
Done.

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

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

- This query ranks each unique Landing_Outcome between the dates June 4, 2010 and March 20, 2017 by counting how many times each occurred, and then orders them from most to least frequent. According to the result:
  - No attempt was the most common landing outcome during this period (10 occurrences).
  - Both Success (drone ship) and Failure (drone ship) happened 5 times each.
  - Other outcomes like Success (ground pad), Controlled (ocean), and Failure (parachute) occurred less frequently.

  This helps analyze the progression and risks of landing attempts in SpaceX missions during that timeframe.

```
%sql SELECT "Landing_Outcome", COUNT(*) AS Outcome_Count FROM SPACEXTABLE WHERE "Date" BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY "Landing_Out
```

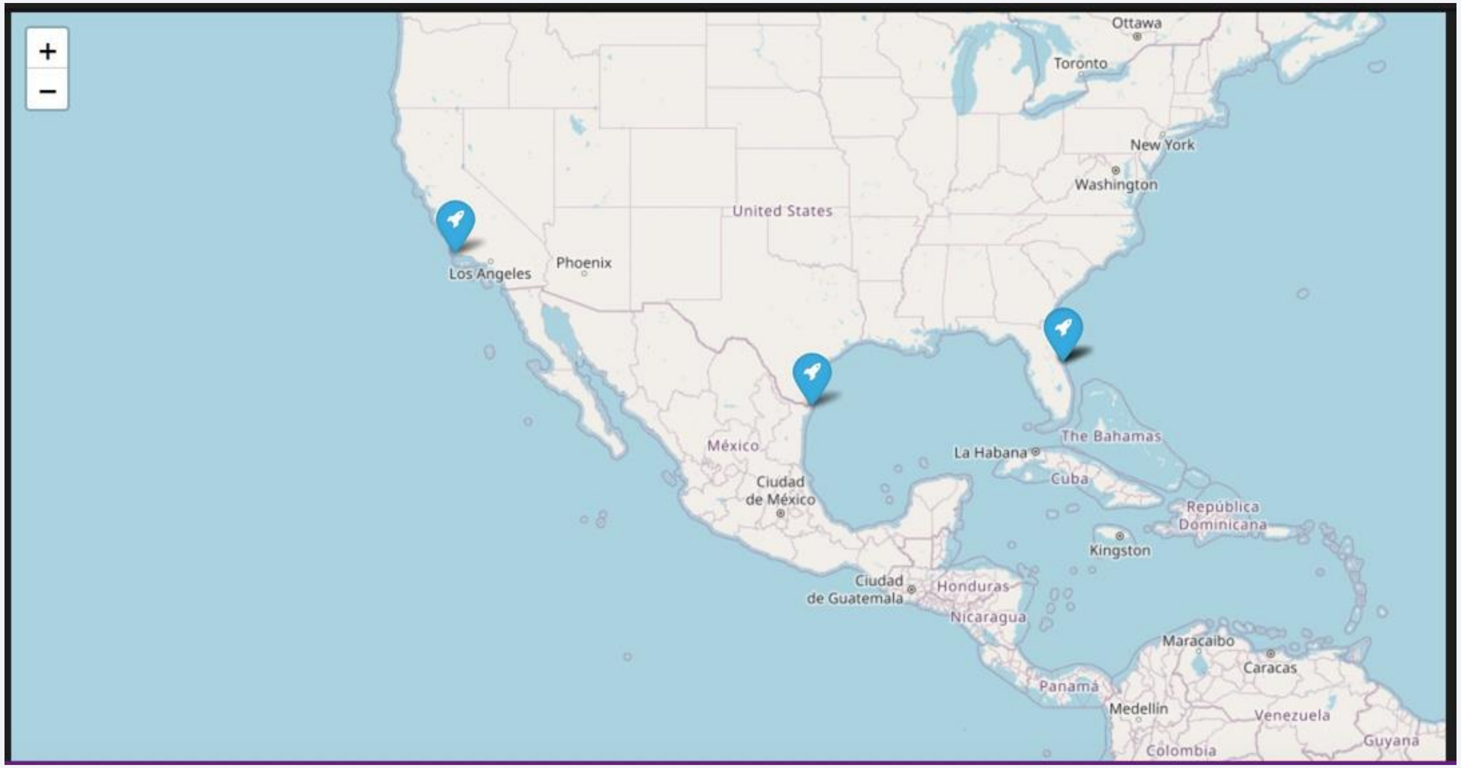
[27] ✓ 0.0s Python

* sqlite:///my_data1.db
Done.

Landing_Outcome	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

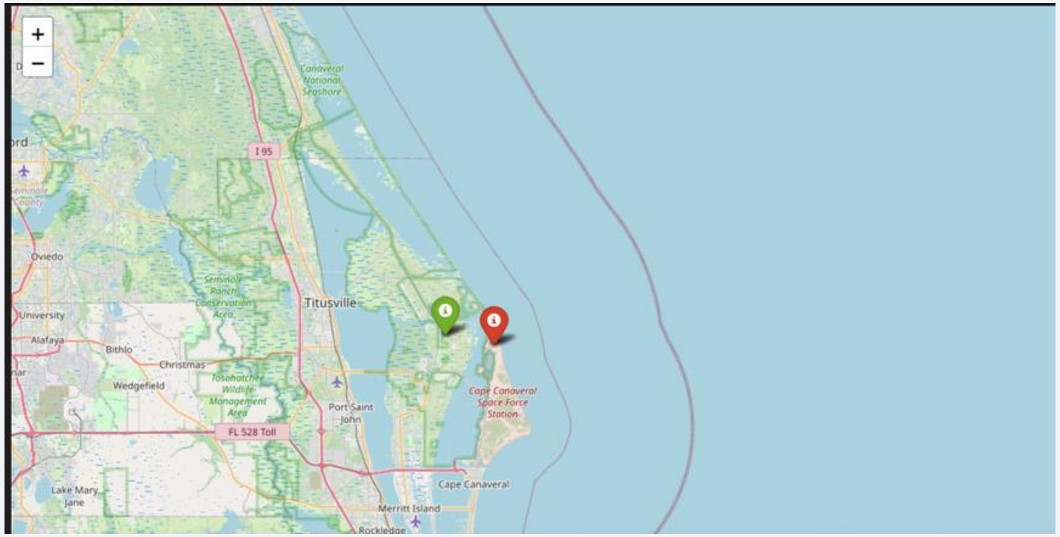

Section 3
Launch Sites
Proximities Analysis

<Folium Map Screenshot 1>

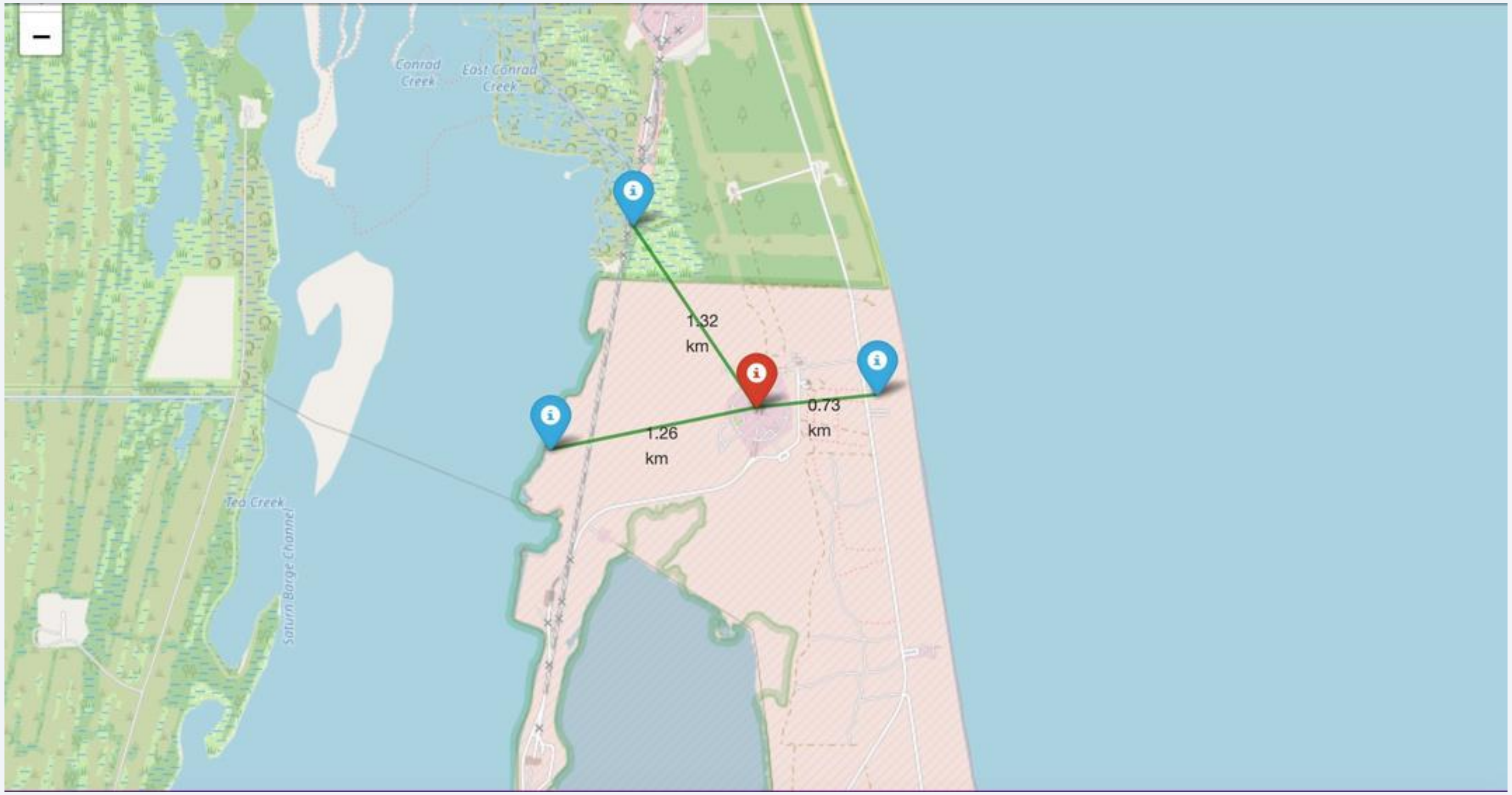


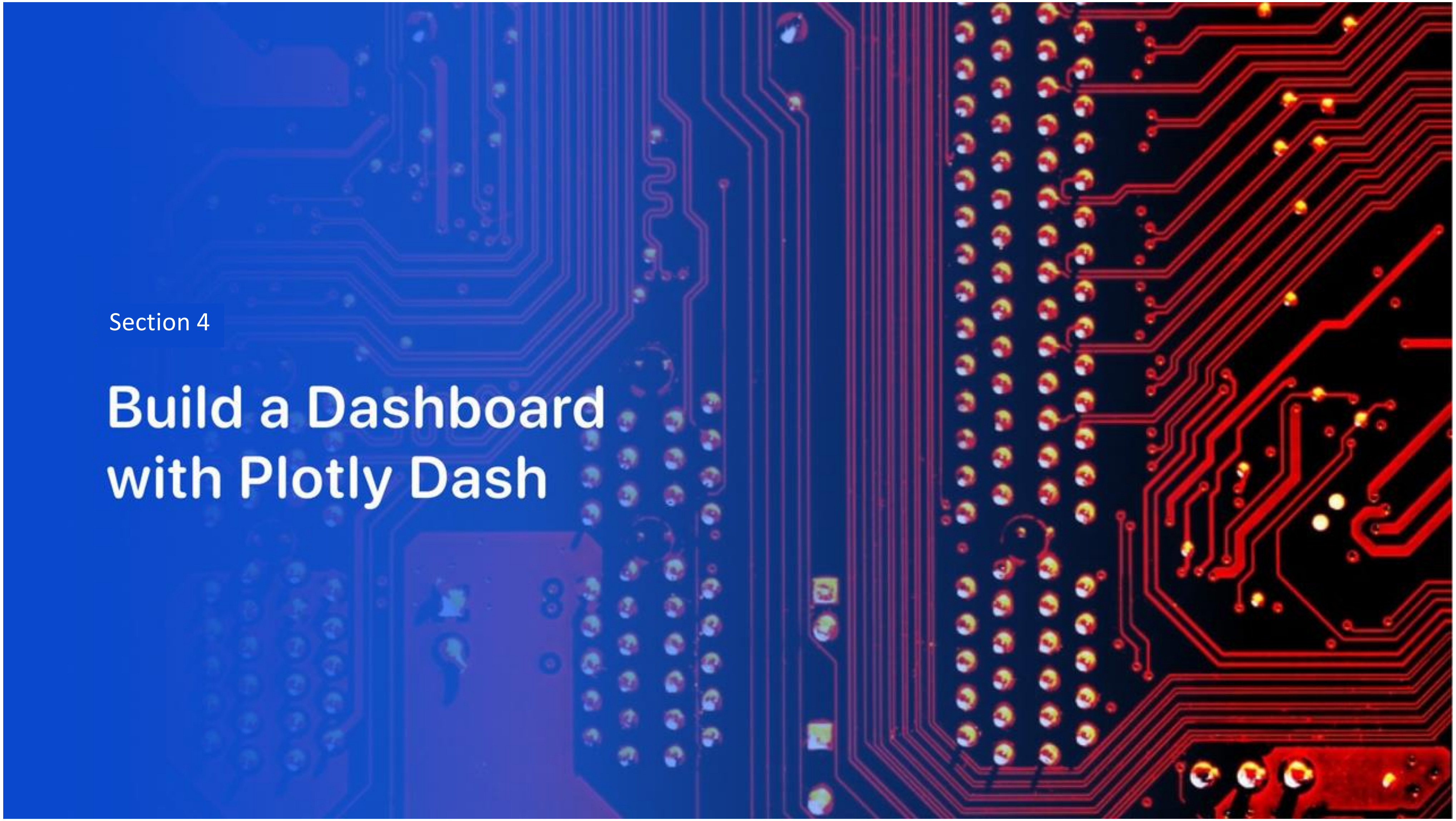
<Folium Map Screenshot 2>

- This folium map visualizes key SpaceX launch sites across the United States with markers color-coded by mission outcome:
  - Green markers represent successful landings.
  - Red markers indicate failed landings.



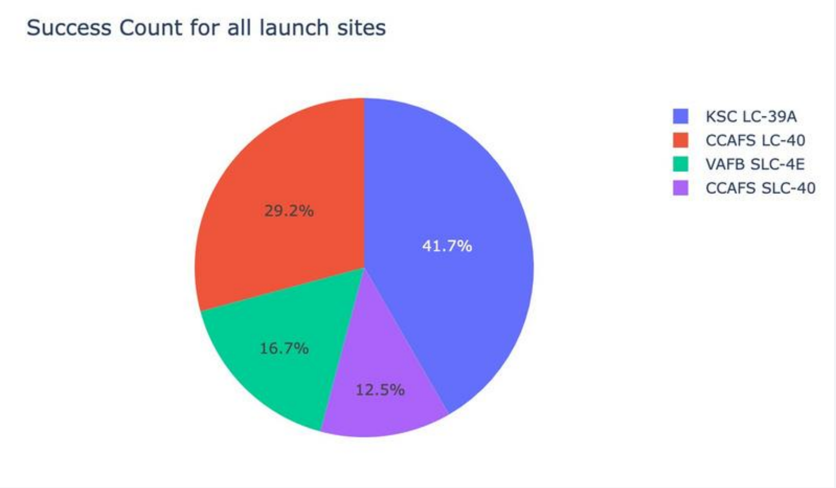
<Folium Map Screenshot 3>



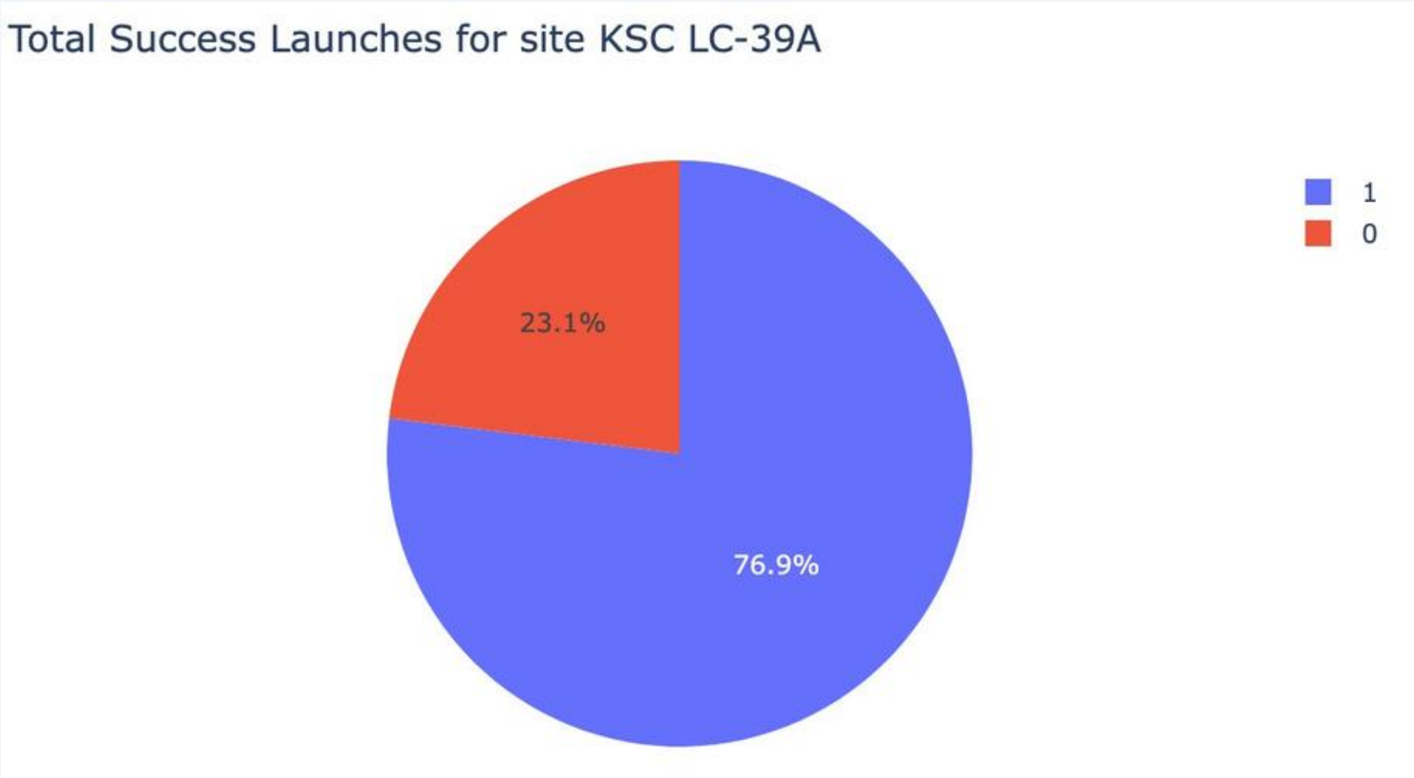

Section 4
Build a Dashboard with Plotly Dash

<Dashboard Screenshot 1>

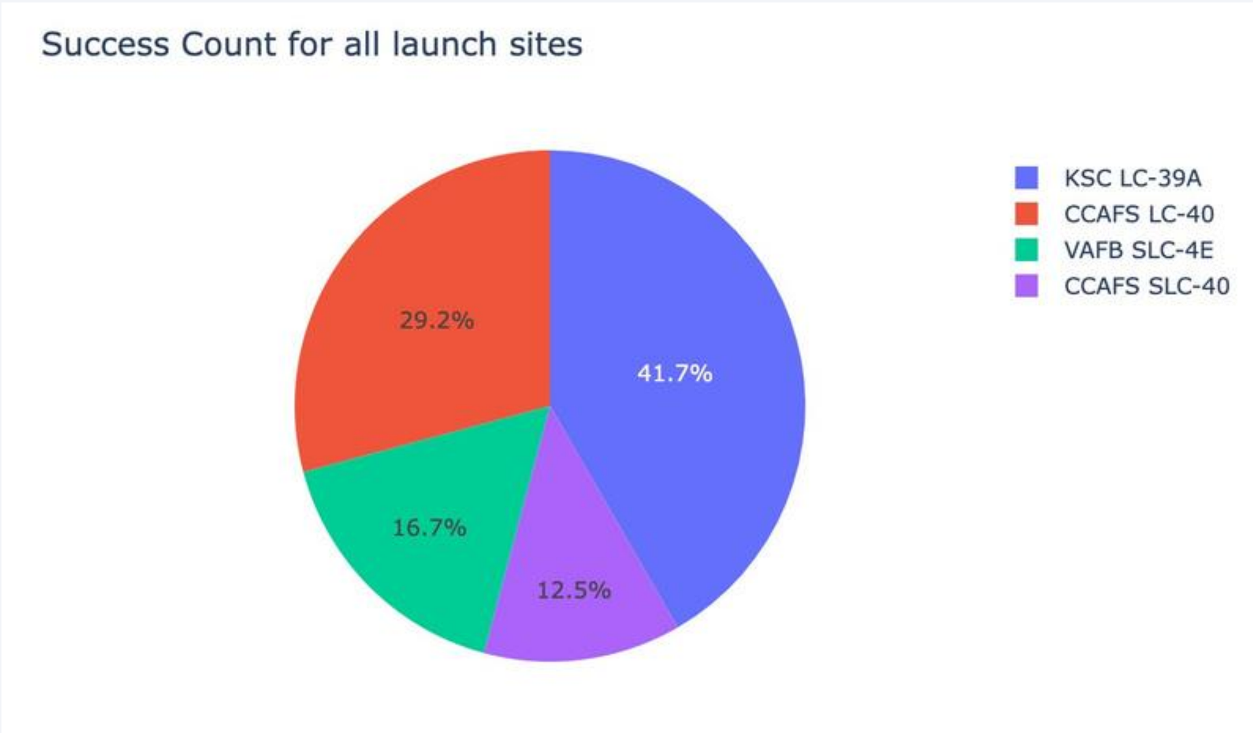
- The dashboard presents a pie chart titled “Success Count for all launch sites” which displays the proportion of successful SpaceX launches from four major launch sites:
  - KSC LC-39A (41.7%) – the leading contributor in successful missions.
  - CCAFS LC-40 (29.2%) – strong secondary contribution.
  - VAFB SLC-4E (16.7%) – West Coast site for polar orbits.
  - CCAFS SLC-40 (12.5%) – smaller share in comparison.
- 🔍Key Findings:
  - KSC LC-39A is the most productive site, accounting for nearly half of all successful launches, likely due to its enhanced infrastructure and major missions.
  - CCAFS LC-40 remains a significant launch hub, often used for commercial missions.
  - VAFB SLC-4E is critical for polar orbits, contributing a reliable portion.
  - The diversity of sites shows SpaceX's strategic use of both East and West Coast facilities.
- This visualization effectively summarizes the geographic distribution of SpaceX's success and reflects the operational emphasis placed on each location.



<Dashboard Screenshot 2>



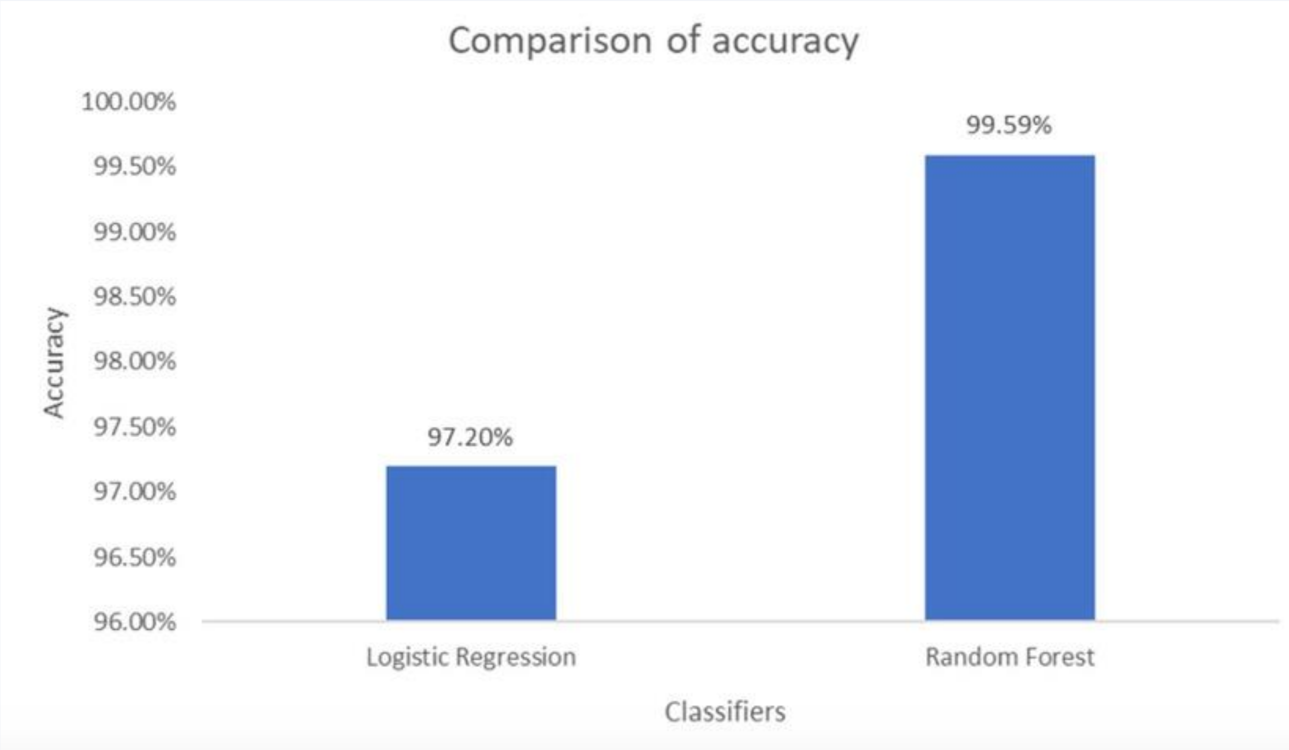
<Dashboard Screenshot 3>



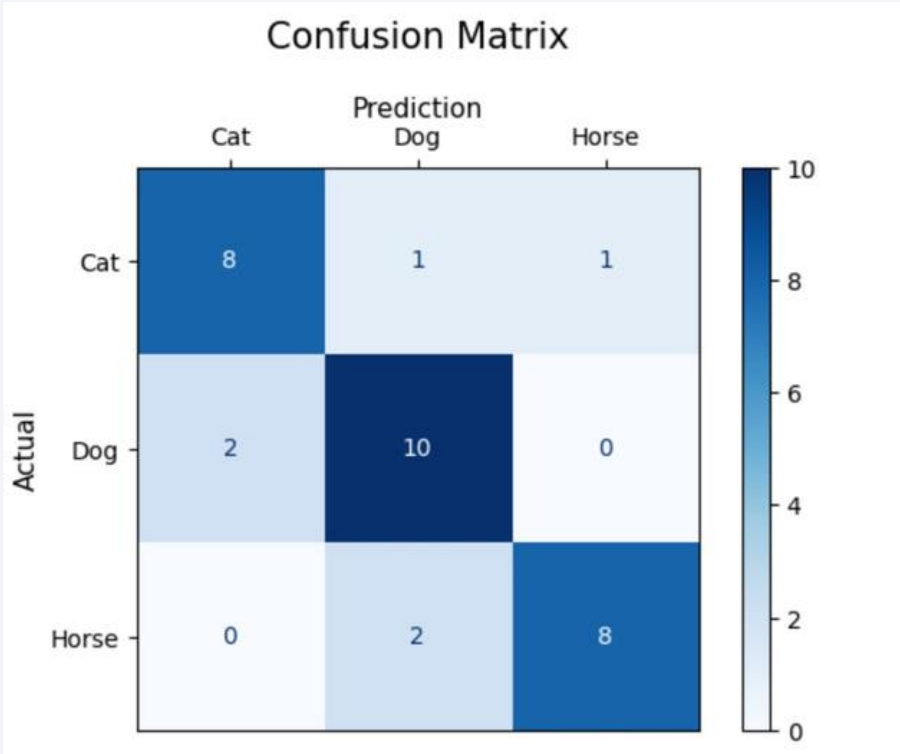
Section 5

Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix



Conclusions

1. Launch Success Trends
   1. There has been a clear improvement in SpaceX's yearly launch success rate from 2010 to 2020, reflecting technological advancements and operational maturity.
2. Launch Site Performance
   1. Among all launch sites, KSC LC-39A recorded the highest number and ratio of successful launches, emphasizing its reliability and infrastructure strength.
   2. The overall success distribution showed that launch activities are concentrated across four main sites, with KSC LC-39A leading.
3. Geospatial Analysis
   1. Folium maps effectively visualized the geographic distribution of launch sites, launch outcomes, and proximities to infrastructure like highways, railways, and coastlines.
   2. Proximity analysis showed that launch sites are strategically located near transport and logistics facilities, likely supporting recovery and supply operations.
4. Dashboard Insights
   1. An interactive Plotly Dash dashboard allowed for dynamic filtering and visualization of launch success metrics across time and space.
   2. Users can analyze site-specific performance and compare launch outcomes efficiently.
5. Modeling and Prediction
   1. Various classification models were evaluated to predict launch success. Support Vector Machine (SVM) showed the highest predictive accuracy (86.3%), making it the best candidate for future prediction tasks.
   2.

Appendix

- GitHub repository: https://github.com/rddh13/ibm_capstone_project


Thank you!