IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Hrisita Mohapatra
19-01-2026

Outline

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

Executive Summary

- Collected SpaceX launch data from APIs and Wikipedia, followed by thorough data cleaning and preparation.
- Conducted exploratory data analysis (EDA) using SQL queries and visualizations to uncover trends in launch outcomes, payload mass, and launch sites.
- Built interactive geographic maps with Folium and analytical dashboards using Plotly Dash to visualize launch locations and success rates.
- Engineered features and trained machine learning classification models (Logistic Regression, SVM, KNN, and Decision Tree).
- Decision Tree model performed best in predicting landing success.
- Key insights show that launch site location, payload mass, and booster version play a significant role in determining first-stage landing success.

Introduction

Project Background and Context

SpaceX has significantly reduced the cost of space launches by reusing the Falcon 9 first stage, offering launches at approximately $62 million, compared to over $165 million by other providers.

The ability to accurately predict first-stage landing success is critical, as successful landings directly impact launch cost, reusability, and competitiveness.

Using publicly available launch data and machine learning techniques, this project aims to predict whether the Falcon 9 first stage will successfully land.

Problems We Aim to Answer

What factors (payload mass, launch site, number of flights, orbit type) influence first-stage landing success?

How does the success rate of landings change over time?

Which machine learning model performs best for predicting landing success?

Section 1

Methodology

Methodology

Executive Summary

- Data Collection Methodology - Launch data was collected from the SpaceX REST API and enhanced with web-scraped Wikipedia data to ensure completeness and historical coverage of Falcon 9 launches.
- Data Wrangling - Data cleaning involved handling missing values, correcting inconsistencies in landing outcomes, encoding categorical variables, and selecting relevant features for analysis and modeling.
- Data Processing - The processed data was organized into structured dataframes and SQL tables, enabling efficient querying, visualization, and machine learning workflows.

Project Link- https://github.com/hrisitamohapatra/Applied-Data-Science-Capstone

Methodology

- Exploratory Data Analysis (EDA) - EDA was conducted using Python visualizations and SQL queries to explore relationships between payload mass, launch site, orbit type, booster version, and landing success.
- Interactive Visual Analytics - Interactive visualizations were developed using Folium maps and Plotly Dash dashboards to analyze geographic patterns, launch site performance, and success distributions.
- Predictive Analysis - Several classification models, Logistic Regression, SVM, KNN, and Decision Tree, were built to predict Falcon 9 first-stage landing success.
- Model Building and Evaluation - Models were tuned using GridSearchCV and evaluated with accuracy metrics and confusion matrices. The Decision Tree model achieved the highest accuracy, demonstrating strong predictive capability.

Data Collection

Key Phrases:

- Publicly available data sources
- SpaceX REST API
- Wikipedia web scraping
- Automated data retrieval
- Structured data extraction
- Historical Falcon 9 launch records

Data Sources:

- SpaceX REST API
  - Launch dates and locations
  - Rocket and payload information
  - First-stage landing outcomes
- Wikipedia
  - Historical Falcon 9 launch tables
  - Mission status (success/failure)
  - Orbit and payload details

Start
↓
Access SpaceX REST API
↓
Extract launch, rocket, payload data
↓
Web scrape Wikipedia Falcon 9 launch tables
↓
Merge API and Wikipedia datasets
↓
Store data in Pandas DataFrames
↓
Ready for Data Wrangling & Analysis

Data Collection

Step 1: SpaceX API Request
- Initiate API requests to retrieve Falcon 9 launch data
- Fetch detailed launch, rocket, payload, and landing outcome information
- Store raw API responses locally for further processing

Step 2: Wikipedia Web Scraping
- Extract HTML launch tables from Wikipedia pages
- Parse and clean the data using BeautifulSoup
- Convert extracted information into structured Pandas DataFrames

Step 3: Data Integration
- Combine SpaceX API data with Wikipedia launch records
- Merge datasets using common launch identifiers and dates
- Create a final integrated dataset ready for data wrangling, analysis, and modeling

Data Collection – SpaceX API

Step 1: Initiate API Request

- Use Python’s requests library to connect to the SpaceX REST API
- Endpoint used: https://api.spacexdata.com/v4/launches
- Send GET requests to retrieve Falcon 9 launch data

Step 2: Parse API Response

- Parse JSON responses into Python objects using response.json()
- Extract relevant fields:
  - Launch date and launch site
  - Rocket and booster version
  - Payload mass and orbit type
  - First-stage landing outcome
-

  Step 3: Store Data Locally
- Convert extracted data into Pandas DataFrames
- Store structured data locally (CSV / database)
- Prepare data for wrangling, analysis, and modeling

Github:-https://github.com/hrisitamohapatra/Applied-Data-Science-Capstone

Initiate API Request (requests.get)

↓

Fetch Launch Data (dates, sites, payloads, outcomes)

↓

Parse JSON Response (response.json())

↓

Extract Relevant Fields

↓

Store Data in Pandas DataFrame

↓

Save Data Locally (CSV / Database)

Data Collection - Scraping

Step 1: Initiate Web Scraping

- Use Python’s requests library to fetch the HTML content of the Wikipedia page
- Target URL:https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launchesRetrieve raw HTML for Falcon 9 launch history

Step 2: Parse HTML Content

- Use BeautifulSoup to parse the HTML structure
- Identify and extract the HTML tables containing Falcon 9 launch records
- Select relevant rows and columns for analysis

Step 3: Convert to DataFrame

- Convert extracted HTML tables into Pandas DataFrames
- Clean and format data:Remove missing valuesStandardize column names
- Ensure data consistency
- Prepare structured data for integration with API data

Github:-https://github.com/hrisitamohapatra/Applied-Data-Science-Capstone

Request Web Page (requests.get)

↓

Parse HTML Content (BeautifulSoup)

↓

Extract Launch Records (HTML tables)

↓

Convert Table to Pandas DataFrame

↓

Clean and Format Data

Data Wrangling

Step 1: Handle Missing Values

- Identify missing values in payload mass, landing outcome, and orbit fields
- Remove or replace missing entries where appropriate

Step 2: Data Cleaning

- Standardize landing outcome labels (Success / Failure)
- Normalize categorical values such as launch sites and orbit types
- Remove unnecessary or duplicate columns

Raw API & Scraped Data

↓

Identify Missing & Invalid Values

↓

Clean & Standardize Data

↓

Feature Transformation & Encoding

↓

Final Clean Dataset

Data Wrangling

Step 3: Feature Preparation

- Convert categorical variables into numerical formats
- Create new features required for analysis and modeling
- Ensure consistent data types across all columns

Step 4: Store Processed Data

- Save cleaned and processed data into structured Pandas DataFrames
- Store datasets for EDA, SQL analysis, and machine learning

Github:-https://github.com/hrisitamohapatra/Applied-Data-Science-Capstone

EDA with Data Visualization

Charts Used and Purpose

- Scatter Plots (Payload Mass vs Landing Outcome)-Used to explore how payload mass influences first-stage landing success and to identify payload ranges associated with failures or successes.
- Bar Charts (Launch Site vs Success Rate)-Used to compare landing success rates across different launch sites and evaluate site-specific performance.
- Pie Charts (Success vs Failure)-Used to show the overall proportion of successful and failed landings, providing a quick summary of model outcomes.
- Time-Series / Trend Plots-Used to analyze how landing success rates evolve over time, highlighting improvements due to technology and operational experience.
- Categorical Plots (Orbit Type and Booster Version)-Used to assess how mission orbit types and booster versions affect landing success.

EDA with Data Visualization

Key Insights from Visual Analysis

- Landing success improves significantly over time, reflecting technological advancements.
- Certain launch sites consistently demonstrate higher success rates.
- Newer booster versions show greater reliability, even with heavier payloads.
- Payload mass and orbit type are important factors influencing landing outcomes.

Github:-https://github.com/hrisitamohapatra/Applied-Data-Science-Capstone

EDA with SQL

SQL Queries Performed

- Retrieved and filtered launch records based on year and date ranges to analyze historical trends.
- Counted successful and failed landing outcomes across different missions.
- Grouped launches by launch site to calculate and compare success rates.
- Analyzed landing outcomes by booster version to assess performance improvements over time.
- Queried payload mass statistics (min, max, average) for different mission categories.
- Identified the earliest successful landings and key milestone launches.
- Ranked landing outcome frequencies within specified time periods.

Key Insights from SQL Analysis

- Landing success rates increase significantly in later years. Certain launch sites show consistently higher success rates.
- Newer booster versions are associated with improved landing performance.
- Github:-https://github.com/hrisitamohapatra/Applied-Data-Science-Capstone

Build an Interactive Map with Folium

Map Objects Created

- Markers to represent SpaceX launch sites
- Color-coded markers to indicate successful and failed launches
- Circles to highlight launch site regions and safety zones
- Lines (Polylines) to show distances between launch sites and nearby features (cities, coastlines, highways)
- Popups and labels to display launch site names and distance information

Github:-
https://github.com/hrisitamohapatra/Applied-Data-Science-Capstone

Why These Objects Were Added
- Markers locate launch sites on the map.
- Color coding highlights success vs failure outcomes.
- Circles show proximity and safety zones.
- Lines display distances to nearby cities and infrastructure.
- Popups enhance interactivity and readability.

Key Insights from the Map
- Launch sites are positioned near coastlines to reduce public risk.
- Some sites show higher concentrations of successful launches.
- Distance analysis emphasizes safety considerations in site selection.

Build a Dashboard with Plotly Dash

Build a Dashboard with Plotly Dash

- Plots and Interactions Added
- Pie Chart showing launch success vs failure distribution
- Scatter Plot of payload mass versus launch outcome
- Dropdown filter to select different launch sites
- Range slider to filter launches by payload mass

Why These Plots and Interactions Were Added

- Pie charts provide a quick overview of success rates.
- Scatter plots help analyze the relationship between payload mass and launch success.
- Dropdowns allow interactive comparison across launch sites.
- Sliders enable users to explore performance across payload ranges.

Key Insights from the Dashboard

- Launch success varies significantly by launch site.
- Higher success rates are associated with specific payload ranges.
- Interactive controls make trend exploration more intuitive and flexible.

Github:-https://github.com/hrisitamohapatra/Applied-Data-Science-Capstone

Predictive Analysis (Classification)

Model Development Summary

- Defined the target variable as first-stage landing success (Success / Failure).

- Selected key features such as payload mass, launch site, orbit type, booster version, and number of flights.

- Split the dataset into training and testing sets.

- Trained multiple classification models: Logistic Regression, Support Vector Machine (SVM),K-Nearest Neighbors (KNN,)Decision Tree.

Model Evaluation and Improvement

-Tuned hyperparameters using GridSearchCV.

-Evaluated models using accuracy scores and confusion matrices.

-Compared model performance to identify the best-performing classifier.

-Decision Tree achieved the highest accuracy, making it the most effective model.

Github:-https://github.com/hrisitamohapatra/Applied-Data-Science-Capstone

Cleaned Dataset

↓

Feature Selection & Encoding

↓

Train-Test Split

↓

Train Classification Models

↓

Hyperparameter Tuning (GridSearchCV)

↓

Model Evaluation

↓

Best Model Selection

Results

Exploratory Data Analysis

Landing success rates increased over time, indicating technological improvement.

Launch site, payload mass, and booster version significantly influence landing outcomes.

Interactive Analytics

Folium maps highlighted geographic patterns and proximity to coastlines.

Plotly Dash dashboards enabled interactive exploration of success rates by launch site and payload mass.

Predictive Analysis Results

Decision Tree achieved the highest accuracy among all models on Training.

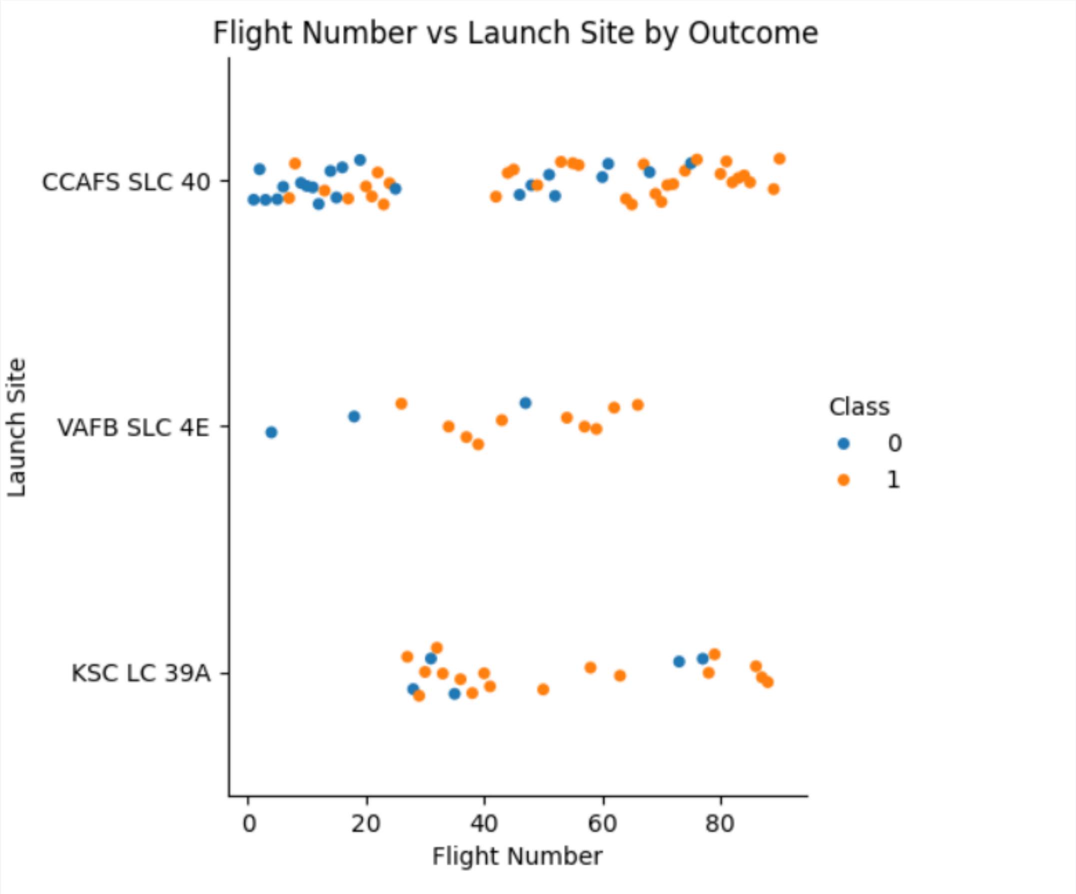
Payload mass and launch site were the most influential features.

Section 2

Insights drawn from EDA

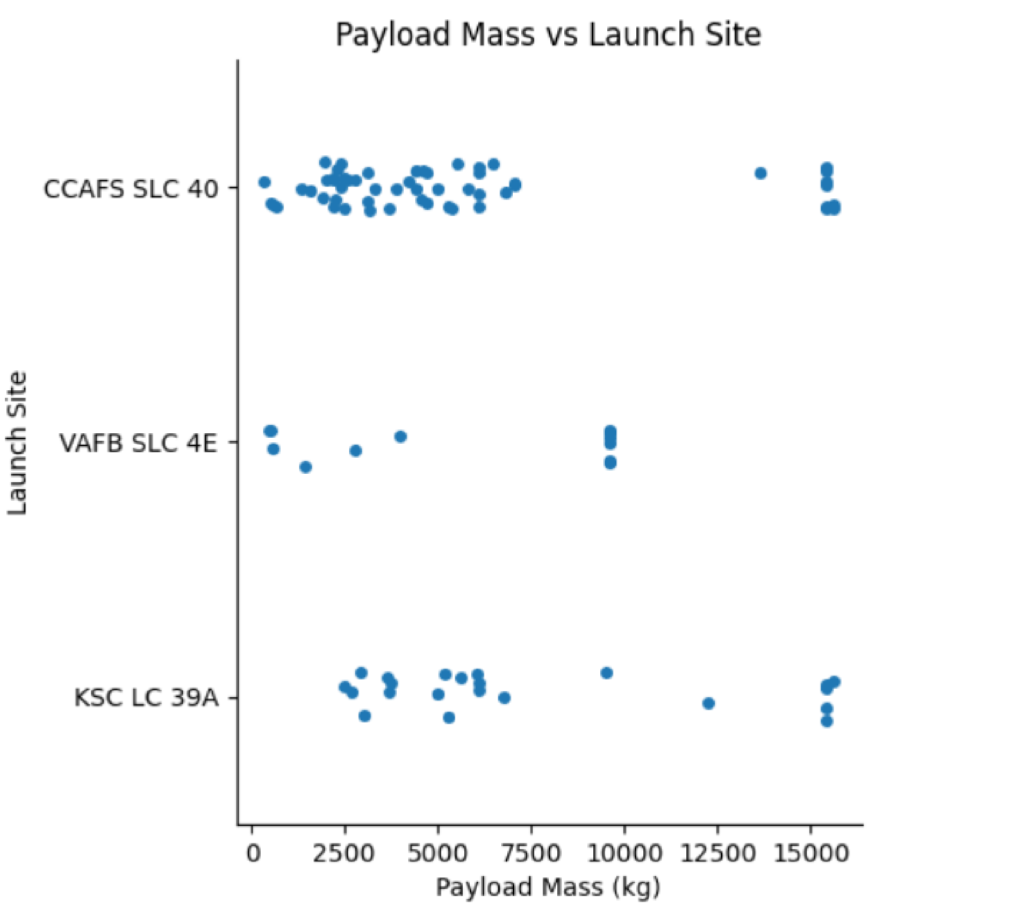
Flight Number vs. Launch Site

- Earlier flights across all launch sites show higher failure rates, highlighting initial operational challenges.
- Launch success improves with increasing flight number, suggesting learning effects and technological improvements over time.
- Major launch sites (CCAFS SLC 40 and KSC LC 39A) show stronger performance in later missions, indicating that landing success depends on multiple factors beyond launch site alone



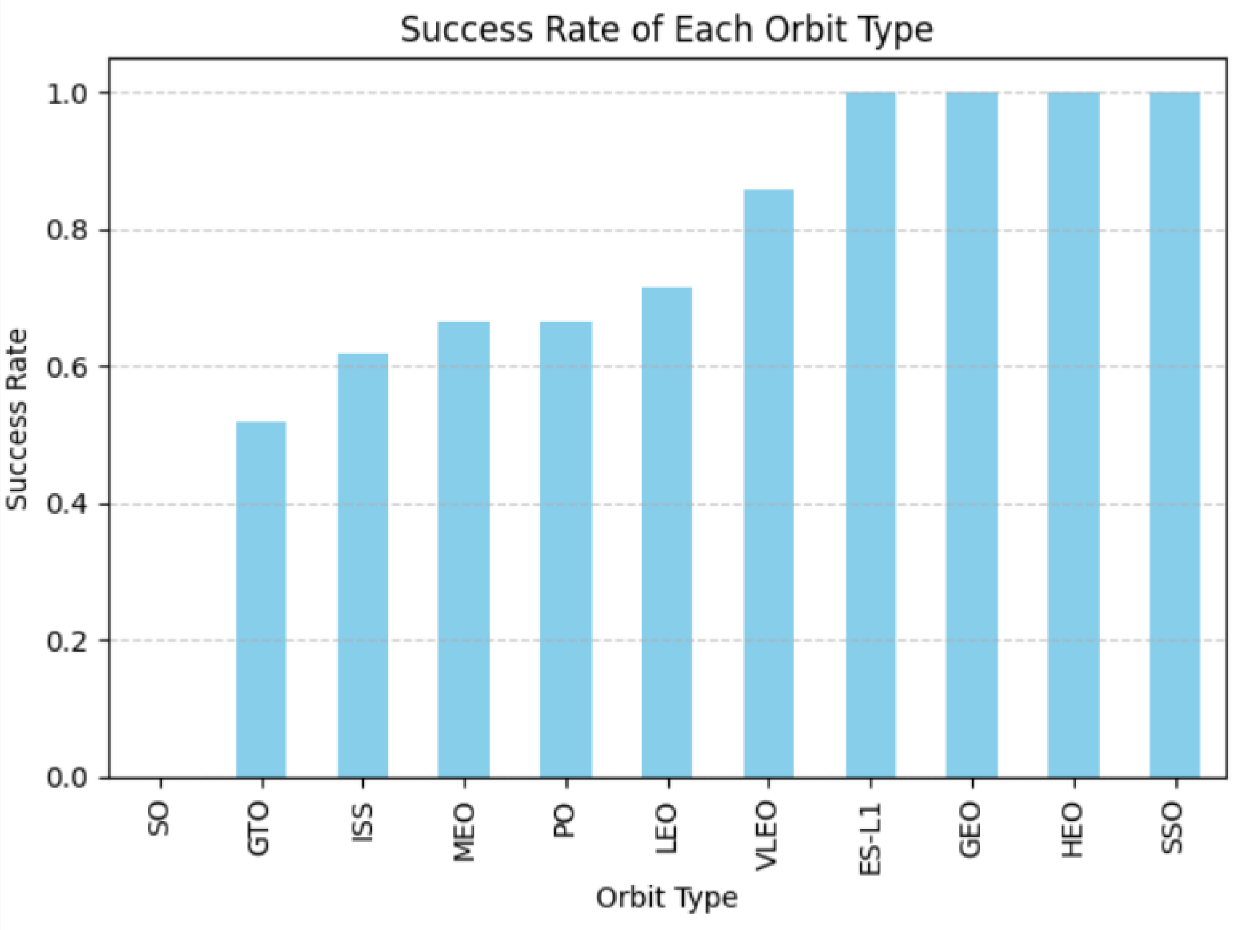
Payload vs. Launch Site

- Payload masses vary across launch sites, with CCAFS SLC 40 supporting a wide range of payloads, including many mid-to-high mass launches.
- KSC LC 39A handles some of the heaviest payloads, suggesting it is used for more demanding missions.
- VAFB SLC 4E shows fewer launches and generally lighter payloads, indicating more specialized mission profiles



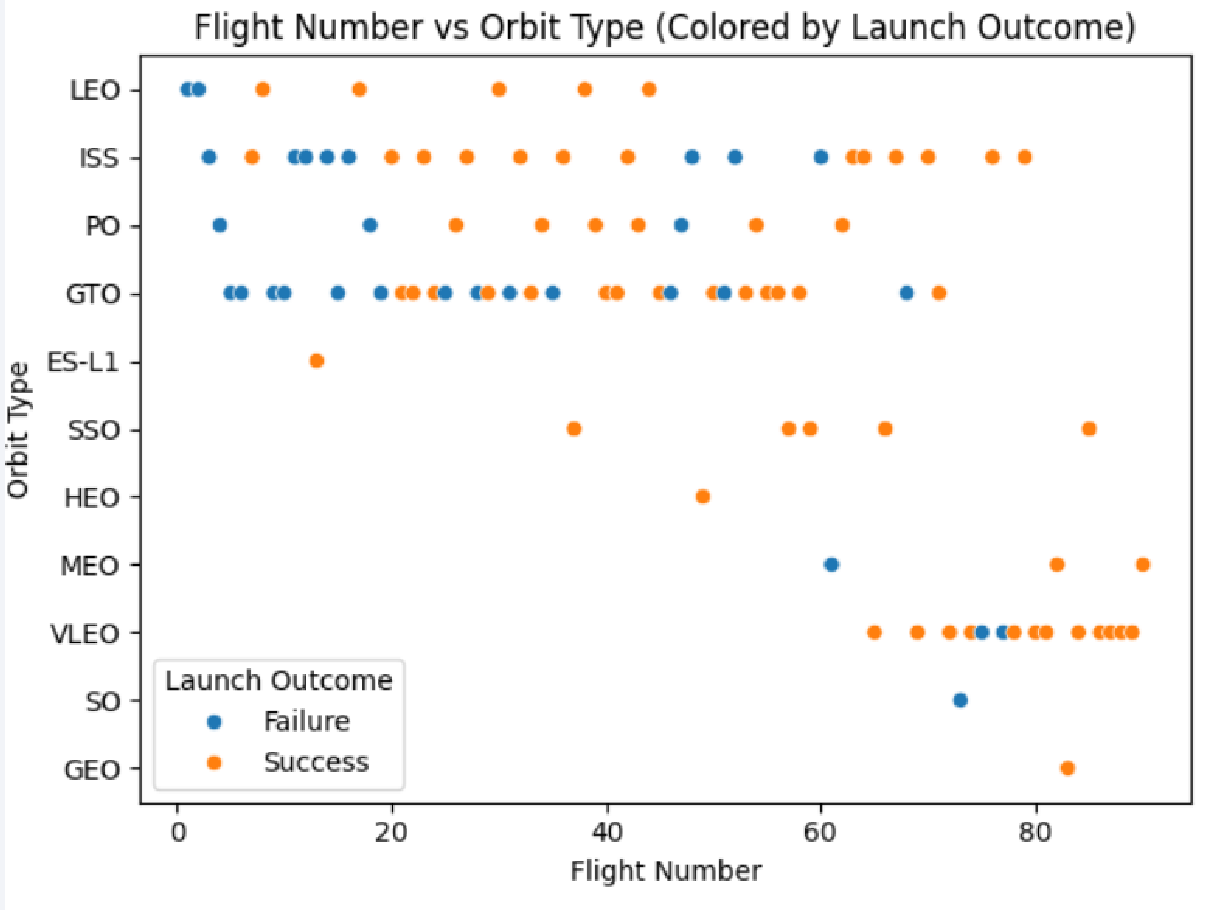
Success Rate vs. Orbit Type

- Success rates vary by orbit type, with GTO, ISS, MEO, PO, and LEO showing moderate to high success levels.
- Higher success rates are observed for VLEO, ES-L1, GEO, HEO, and SSO, indicating more consistent mission outcomes.
- Lower success rates in earlier or more complex orbit types suggest that mission profile and orbital requirements influence landing success



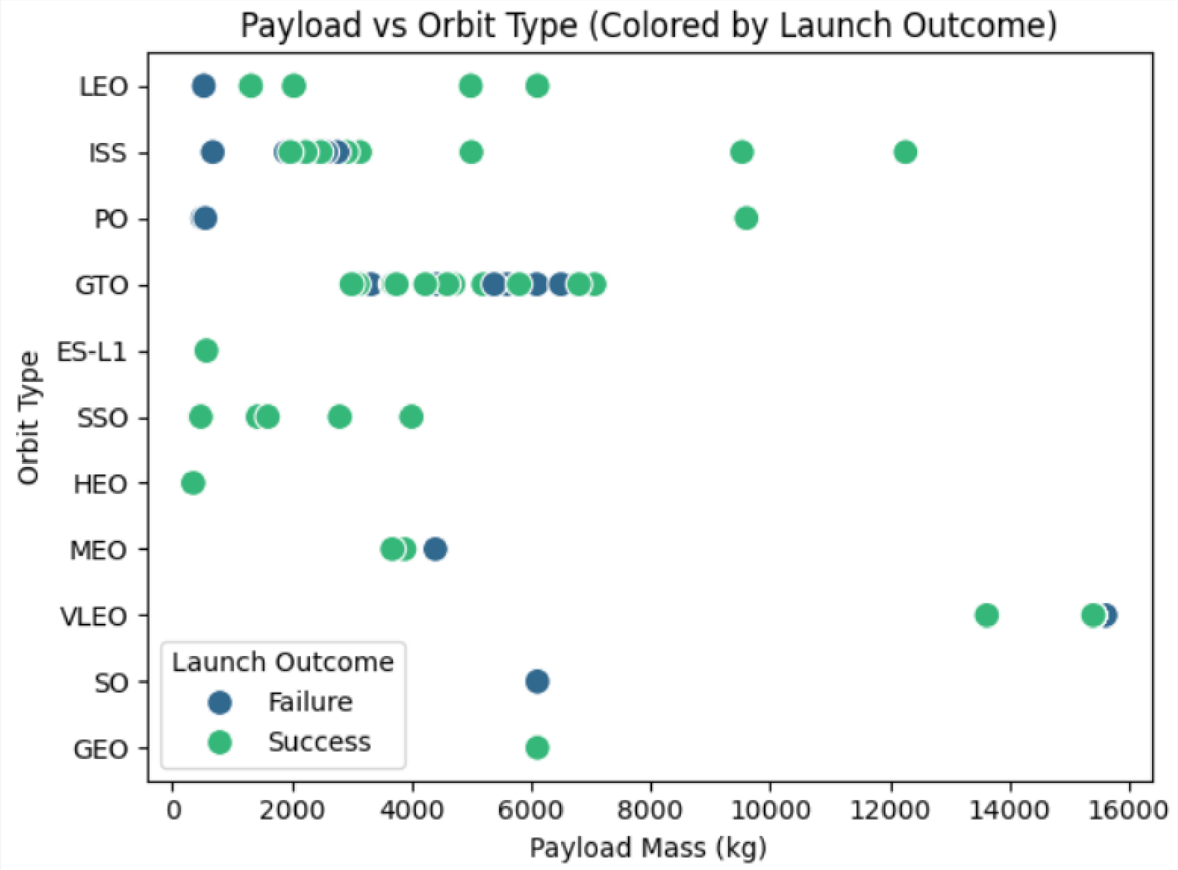
Flight Number vs. Orbit Type

- Earlier flights across multiple orbit types show a higher number of failures, particularly for GTO, LEO, and ISS missions.
- Launch success increases in later flight numbers across most orbit types, indicating improved reliability over time.
- Certain orbit types (such as SSO, VLEO, GEO, and ES-L1) show predominantly successful outcomes, suggesting mission profile influences landing success



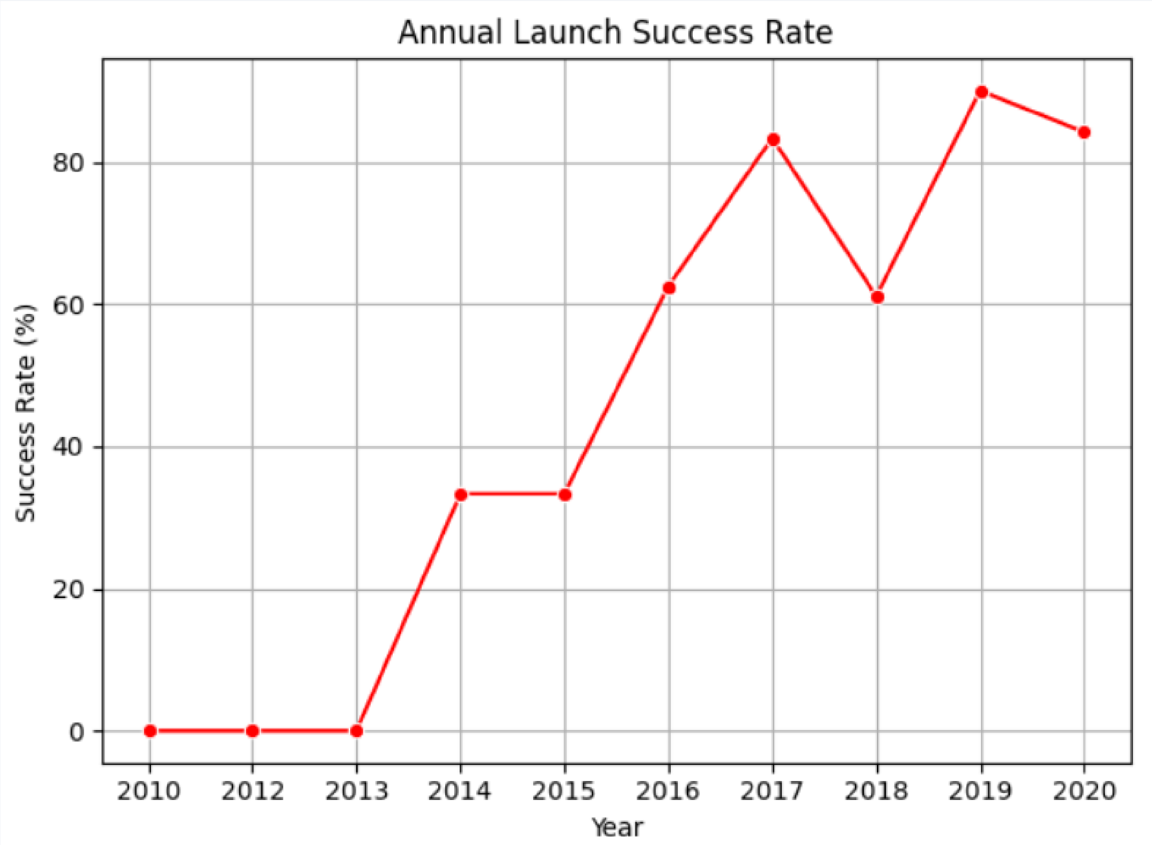
Payload vs. Orbit Type

- Successful launches occur across a wide range of payload masses, indicating that payload mass alone does not determine launch success.
- Certain orbit types (such as GTO, ISS, and LEO) support a broad range of payload masses with mostly successful outcomes.
- Higher-mass payloads are more commonly associated with VLEO and GEO missions, which also show high success rates



Launch Success Yearly Trend

- Early years show low or zero success rates, reflecting initial development and operational challenges.
- Launch success improves significantly over time, with a sharp increase beginning around 2014.
- Despite minor fluctuations, success rates remain consistently high in later years, indicating improved reliability and maturity of the launch system.
- Experience and iterative improvements over time play a critical role in increasing launch success.



All Launch Site Names

Display the names of the unique launch sites in the space mission

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

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

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

This SQL query retrieves the unique launch sites used in the SpaceX missions dataset.

The results show that launches were conducted from CCAFS LC-40, VAFB SLC-4E, and KSC LC-39A, indicating three distinct launch locations in the dataset.

Launch Site Names Begin with 'CCA'

Display 5 records where launch sites begin with the string 'CCA'

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

```
 * 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

- This SQL query filters the dataset to display five records where the launch site name begins with “CCA”, corresponding to launches from CCAFS LC-40.
- The output shows early SpaceX missions launched from this site, including mission details such as payload, orbit type, customer, and launch and landing outcomes.

Total Payload Mass

Display the total payload mass carried by boosters launched by NASA (CRS)

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

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

SUM("PAYLOAD_MASS__KG_")
45596

- This SQL query calculates the total payload mass carried by SpaceX launches for the NASA (CRS) program.
- The result shows that NASA (CRS) missions delivered a combined payload mass of 45,596 kg.

Average Payload Mass by F9 v1.1

Display average payload mass carried by booster version F9 v1.1

```
%sql SELECT AVG("PAYLOAD_MASS__KG_") FROM SPACEXTABLE WHERE "Booster_Version" = 'F9 v1.1';
```

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

AVG("PAYLOAD_MASS__KG_")
2928.4

- This SQL query computes the average payload mass carried by launches using the Falcon 9 v1.1 booster.
- The result shows that F9 v1.1 missions carried an average payload of 2,928.4 kg.

First Successful Ground Landing Date

List the date when the first succesful landing outcome in ground pad was acheived.

Hint:Use min function

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

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

MIN("Date")
2015-12-22

- This SQL query identifies the earliest date on which a successful ground pad landing occurred.
- The result shows that the first successful ground pad landing was achieved on December 22, 2015.

Successful Drone Ship Landing with Payload between 4000 and 6000

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

```
%sql SELECT DISTINCT "Booster_Version" FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Success (drone ship)' AND "PAYLOAD_MASS__
```

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

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

- This SQL query selects distinct booster versions from the SpaceX dataset where the landing outcome was a successful drone ship landing and the payload mass was between 4,000 kg and 6,000 kg:
  - SELECT DISTINCT "Booster_Version" FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Success (drone ship)' AND "PAYLOAD_MASS__KG_" > 4000 AND "PAYLOAD_MASS__KG_" < 6000;
- The results show that the boosters F9 FT B1022, F9 FT B1026, F9 FT B1021.2, and F9 FT B1031.2 met these conditions.

Total Number of Successful and Failure Mission Outcomes

List the total number of successful and failure mission outcomes

```
%sql SELECT "Mission_Outcome", COUNT(*) AS "Total" FROM SPACEXTABLE WHERE "Mission_Outcome" IN ('Success', 'Failure') GROUP
```

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

Mission_Outcome	Total
Success	98

- This SQL query counts the total number of successful and failed missions by grouping records based on the mission outcome:
  - SELECT "Mission_Outcome", COUNT(*) AS "Total" FROM SPACEXTABLE WHERE "Mission_Outcome" IN ('Success', 'Failure') GROUP BY "Mission_Outcome";
- The result shows that there are 98 successful missions recorded in the dataset

Boosters Carried Maximum Payload

List all the booster_versions that have carried the maximum payload mass, using a subquery with a suitable aggregate function.

```
%sql SELECT DISTINCT "Booster_Version" FROM SPACEXTABLE WHERE "PAYLOAD_MASS__KG_" = (SELECT MAX("PAYLOAD_MASS__KG_") FROM SP
```

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

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

- This SQL query uses a subquery with the MAX() aggregate function to identify the highest payload mass carried in the dataset and then selects all booster versions that carried that maximum payload:
  - SELECT DISTINCT "Booster_Version" FROM SPACEXTABLE WHERE "PAYLOAD_MASS__KG_" = (SELECT MAX("PAYLOAD_MASS__KG_") FROM SPACEXTABLE)
- ;The results list all Falcon 9 Block 5 booster versions that have successfully carried the maximum payload mass recorded in the dataset.

2015 Launch Records

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.

Note: SQLLite does not support monthnames. So you need to use substr(Date, 6,2) as month to get the months and substr(Date,0,5)='2015' for year.

```
%%sql
SELECT
CASE
 WHEN substr("Date",6,2) = '01' THEN 'January'
 WHEN substr("Date",6,2) = '02' THEN 'February'
 WHEN substr("Date",6,2) = '03' THEN 'March'
 WHEN substr("Date",6,2) = '04' THEN 'April'
 WHEN substr("Date",6,2) = '05' THEN 'May'
 WHEN substr("Date",6,2) = '06' THEN 'June'
 WHEN substr("Date",6,2) = '07' THEN 'July'
 WHEN substr("Date",6,2) = '08' THEN 'August'
 WHEN substr("Date",6,2) = '09' THEN 'September'
 WHEN substr("Date",6,2) = '10' THEN 'October'
 WHEN substr("Date",6,2) = '11' THEN 'November'
 WHEN substr("Date",6,2) = '12' THEN 'December'
 ELSE 'Unknown'
END AS "Month_Name",
"Landing_Outcome",
"Booster_Version",
"Launch_Site"
FROM SPACEXTABLE
WHERE substr("Date",0,5) = '2015'
AND "Landing_Outcome" LIKE '%Failure%'
AND "Landing_Outcome" LIKE '%Drone Ship%';
```

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

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

- This SQL query lists SpaceX launch records from 2015 where a drone ship landing attempt failed.
- The query filters results to include only failure landing outcomes on a drone ship and displays the month name, landing outcome, booster version, and launch site, highlighting when and where these failures occurred during 2015.

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

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

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

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

Landing_Outcome	Total_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

- This SQL query ranks landing outcomes by their total number of occurrences between June 4, 2010 and March 20, 2017.
- It groups launch records by landing outcome, counts how many times each outcome occurred, and orders the results in descending order of frequency, showing which landing outcomes were most common during this period.

Section 3

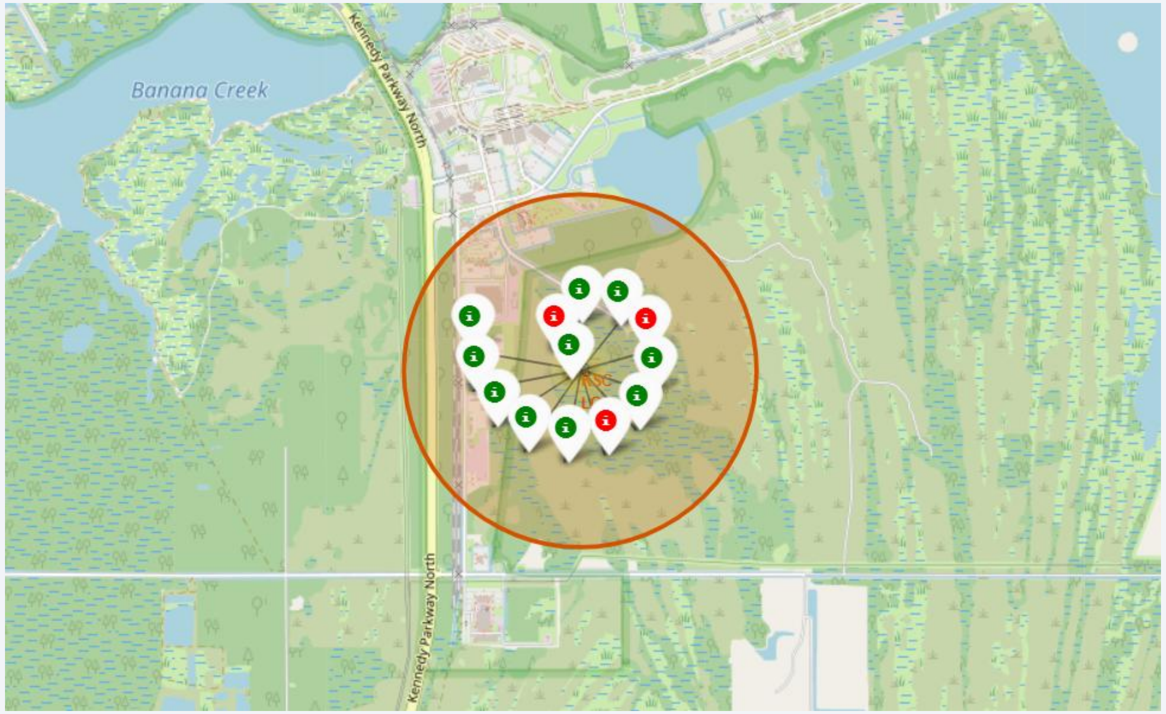
Launch Sites Proximities Analysis

launch sites location markers on a global map

- The map displays the geographic distribution of all SpaceX launch sites using location markers on a global (continental) map
- .Most launch sites are located at lower latitudes, closer to the equator, which allows rockets to take advantage of the Earth’s higher rotational speed to gain additional velocity during launch.
- This natural boost helps reduce fuel consumption and improves the ability of spacecraft to reach and maintain orbit.
- All launch sites are positioned close to coastlines, enabling rockets to launch over open ocean.
- Launching toward the ocean significantly reduces safety risks, as any debris from failed launches is less likely to fall over populated areas.
- Overall, the visualization highlights how orbital efficiency and public safety strongly influence the selection of launch site locations.

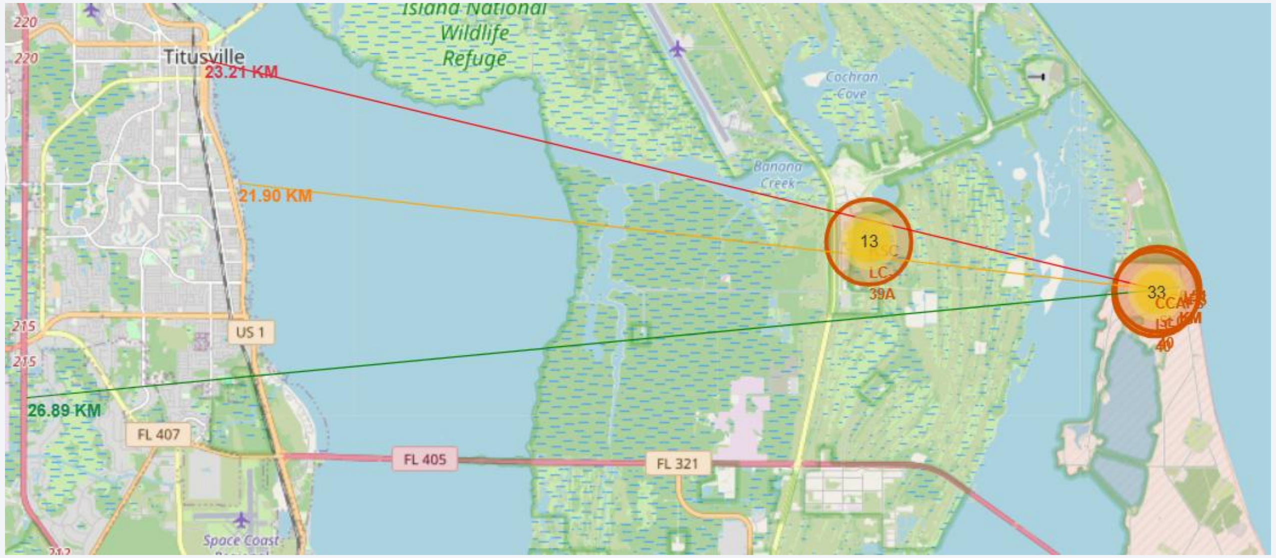
Launch Success Distribution at KSC LC-39A

- The map displays launch outcomes at KSC LC-39A using color-coded markers.
- Green markers represent successful launches, while red markers indicate failed launches.
- The circular boundary highlights the launch site area, allowing a focused spatial analysis.
- By visually comparing the number of green and red markers, we can quickly assess the overall launch success rate.
- The dominance of green markers shows that KSC LC-39A has a very high success rate, indicating strong launch reliability and operational maturity at this site.



Proximity of CCAFS SLC-40 Launch Site

- The map visualization uses distance markers and colored lines to show straight-line distances between CCAFS SLC-40 and surrounding geographic features, making proximity relationships easy to interpret.
- The results indicate that CCAFS SLC-40 is located within 20–30 km of populated and critical infrastructure areas, which highlights the importance of safety regulations and controlled launch zones.
- In the event of a launch failure, a rocket traveling at very high speed could cover 15–20 km within seconds, posing a potential risk to nearby cities, highways, and coastal regions if not properly managed.

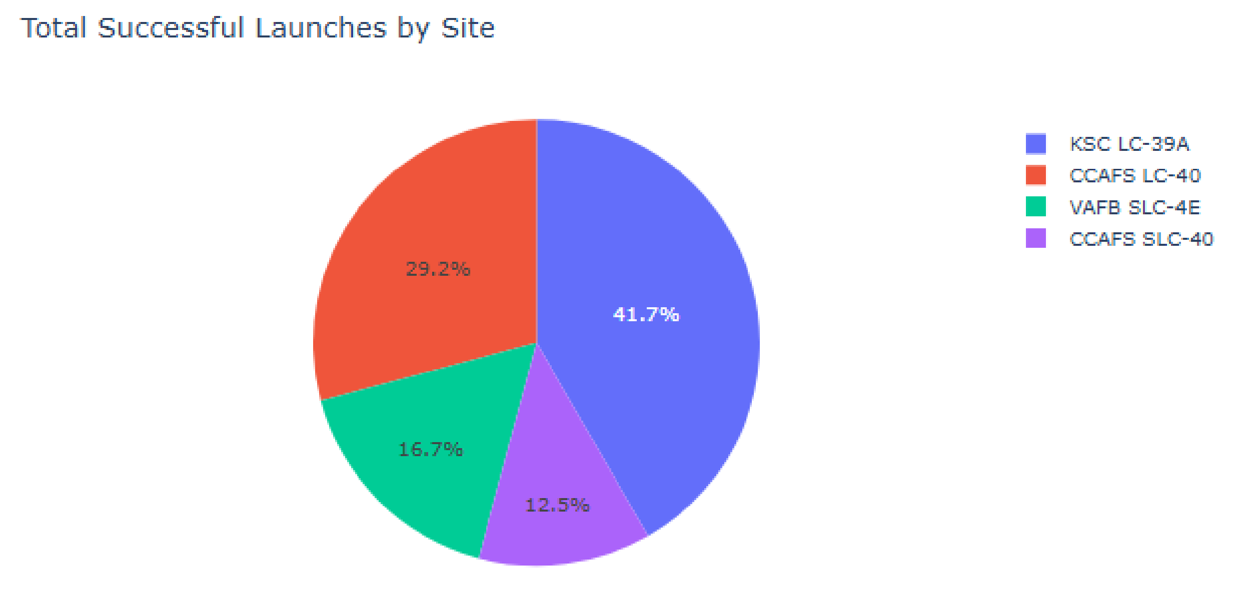


Section 4

Build a Dashboard with Plotly Dash

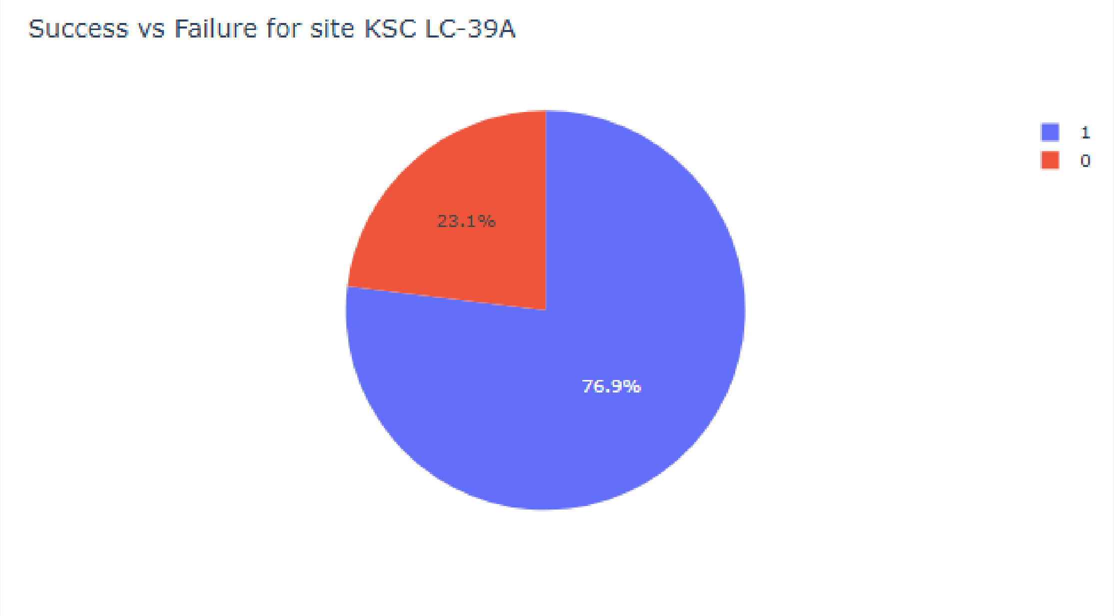
Launch success count for all sites

- The pie chart shows the distribution of total successful launches by launch site.
- KSC LC-39A contributes the largest share of successful launches at 41.7%, indicating the highest reliability.
- CCAFS LC-40 follows with 29.2% of successful launches.
- VAFB SLC-4E accounts for 16.7%, reflecting moderate launch activity.
- CCAFS SLC-40 (secondary pad) contributes the smallest share at 12.5%.
- Overall, the chart highlights that KSC LC-39A is the most successful and frequently used launch site.



Launch site with highest launch success ratio

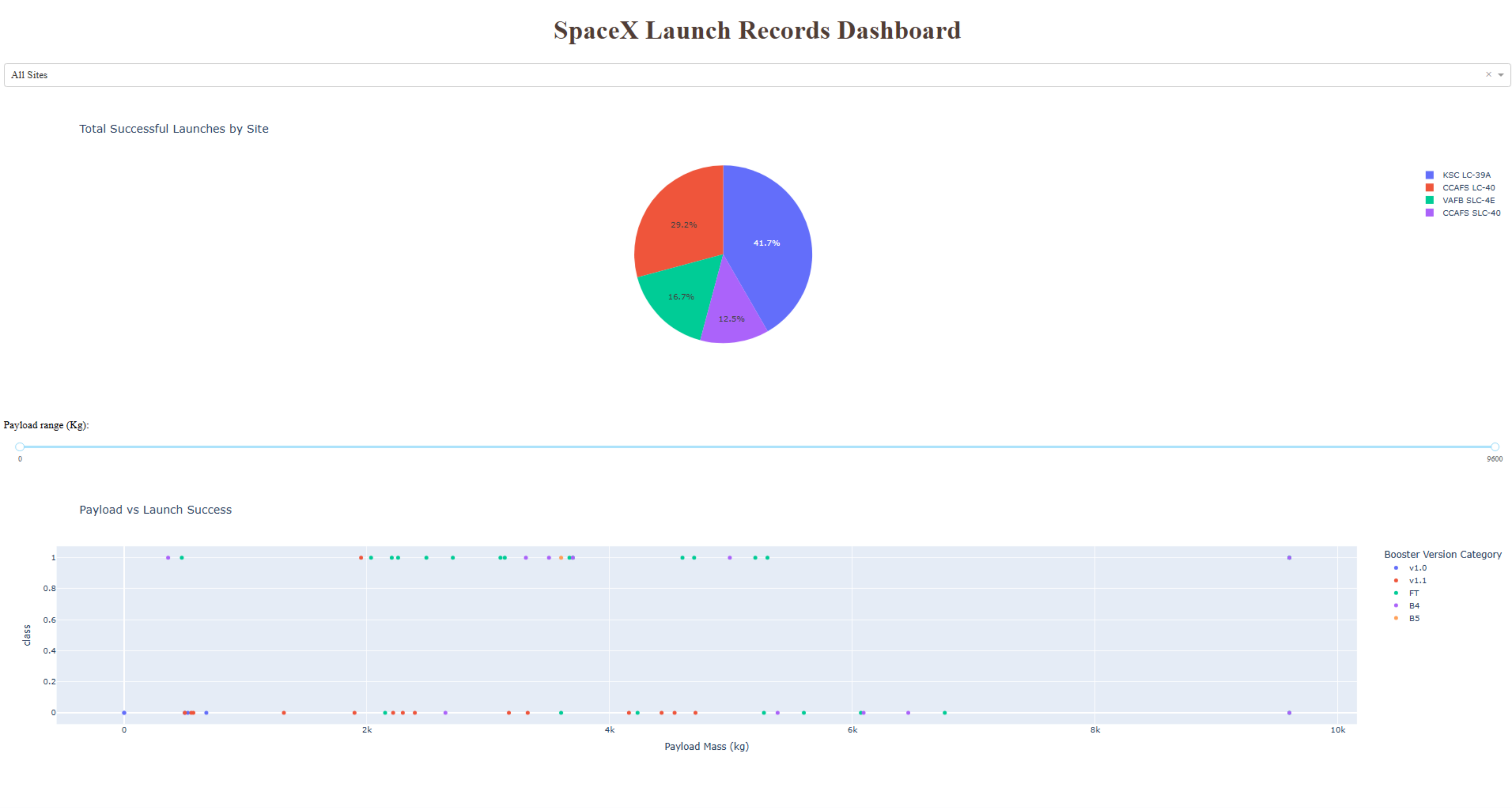
- The pie chart shows the launch outcome distribution for KSC LC-39A.
- Successful launches (1) account for 76.9%, indicating a strong success rate.
- Failed launches (0) make up 23.1% of the total.
- The large success portion highlights KSC LC-39A as a highly reliable launch site.
- Overall, the chart demonstrates consistent and effective launch performance at this site.



Payload vs. Launch Outcome scatter plot for all sites

- The scatter plot illustrates the relationship between payload mass (kg) and launch success (class).
- The y-axis represents launch outcome:
  - 1 = Successful launch
  - 0 = Failed launch
- Points are color-coded by booster version category (v1.0, v1.1, FT, B4, B5).
- Successful launches occur across a wide range of payload masses, indicating payload weight alone does not determine success.
- More recent booster versions (FT, B4, B5) show a higher concentration of successful launches, even at higher payload masses.
- Earlier booster versions (v1.0, v1.1) are associated with more failures, especially at lower payload ranges.
- Overall, the plot suggests that booster technology improvements significantly increase launch success, enabling heavier payloads to be launched reliably.

Complete Dashboard

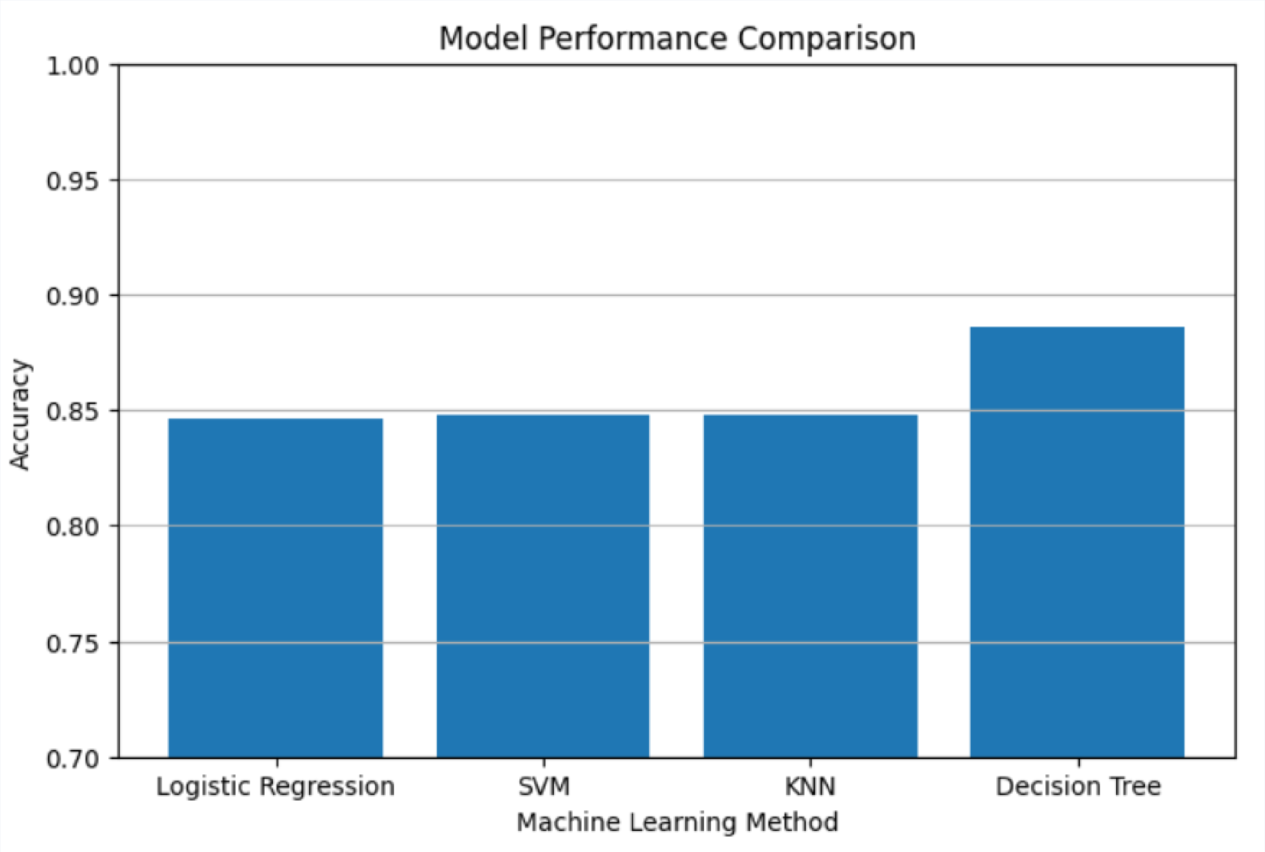


Section 5

Predictive Analysis (Classification)

Classification Accuracy

- The bar chart compares the training accuracy of four machine learning models used to predict launch success.
- The evaluated models are Logistic Regression, SVM, KNN, and Decision Tree.
- Decision Tree achieves the highest accuracy (~88.6%), indicating the best performance on the training data.
- SVM and KNN show very similar performance, both with accuracies of around 84.8%.
- Logistic Regression has a slightly lower accuracy but still performs competitively at around 84.6%.
- Overall, the results suggest that non-linear models, especially Decision Trees, capture the underlying patterns in the data more effectively than linear models.

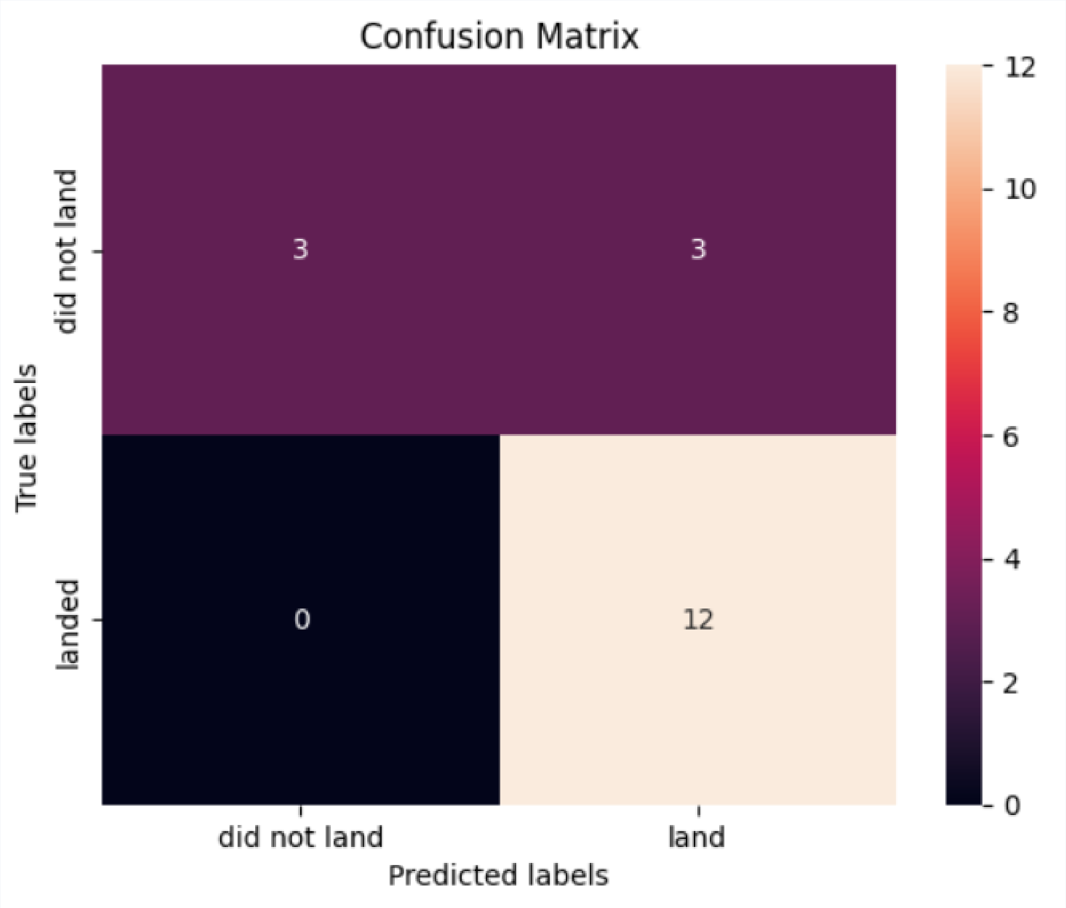


Confusion Matrix

Confusion Matrix for Decision Tree Model

- The confusion matrix summarizes how well the Decision Tree model predicts landing outcomes.
- True Positives (12):The model correctly predicted 12 successful landings as successful.
- True Negatives (3):The model correctly predicted 3 failed landings as failures.
- False Positives (3):The model predicted 3 launches as successful when they actually failed.
- False Negatives (0):The model did not miss any successful landings, which is a strong result.

Key Insight: The Decision Tree model shows excellent performance in identifying successful landings, with no false negatives, making it highly reliable for predicting successful launches.



Conclusions

- Publicly available SpaceX launch data can be successfully leveraged to analyze, visualize, and predict Falcon 9 first-stage landing outcomes.
- Data wrangling and exploratory analysis revealed that launch site location, payload mass, booster version, orbit type, and mission history play a critical role in landing success.
- Interactive visual analytics using Folium and Plotly Dash provided intuitive insights into geographic distribution, success rates, and operational patterns.
- Machine learning classification models demonstrated that landing success can be predicted with high accuracy, validating the effectiveness of data-driven approaches.
- Among all evaluated models, the Decision Tree classifier achieved the best performance, indicating strong capability in capturing non-linear relationships in the data.
- The findings highlight how predictive analytics can support cost estimation, mission planning, and reusability strategies in commercial space operations.

Appendix

Github:-https://github.com/hrisitamohapatra/Applied-Data-Science-Capstone

Thank you!