IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

- By RAVI TEJ

Date: 02/09/2023

Outline

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

Executive Summary

- **Methodologies:**
  - **Data Collection:** Reliable sources for Falcon 9 landing data.
  - **Data Wrangling:** Cleaning and preprocessing for quality data.
  - **EDA:** Visualizations and SQL for insights.
  - **Interactive Analytics:** Folium and Plotly Dash for accessibility.
  - **Predictive Analysis:** Classification models for landing prediction.
- **Results**:
  - **Successful Landing Prediction:** Significant cost savings.
  - **Cost Implications:** Influence on launch decisions.
  - **Data Insights:** Valuable findings from EDA.
  - **Model Performance:** Achieved high-performance metrics (e.g., accuracy, precision, recall, F1-score) validating the model's effectiveness.

Introduction

- **Project Background and Context:**
  - SpaceX Falcon 9 Rocket Launches and Cost-Efficiency.
  - **Reusability of the First Stage:** Cost-saving advantage.
  - **Cost Discrepancy:** SpaceX vs. competitors.
  - Significance of Predicting Landings for Cost Estimation.

- **Problems to Solve:**
  - **Challenge:** Predicting landing success for cost estimation.
  - **Competitive Advantage:** Affecting rocket launch market competition.
  - **Data-Driven Approach:** Leveraging data science and predictive modeling.
  - **Real-World Impact:** Aerospace industry implications.

Section 1

Methodology

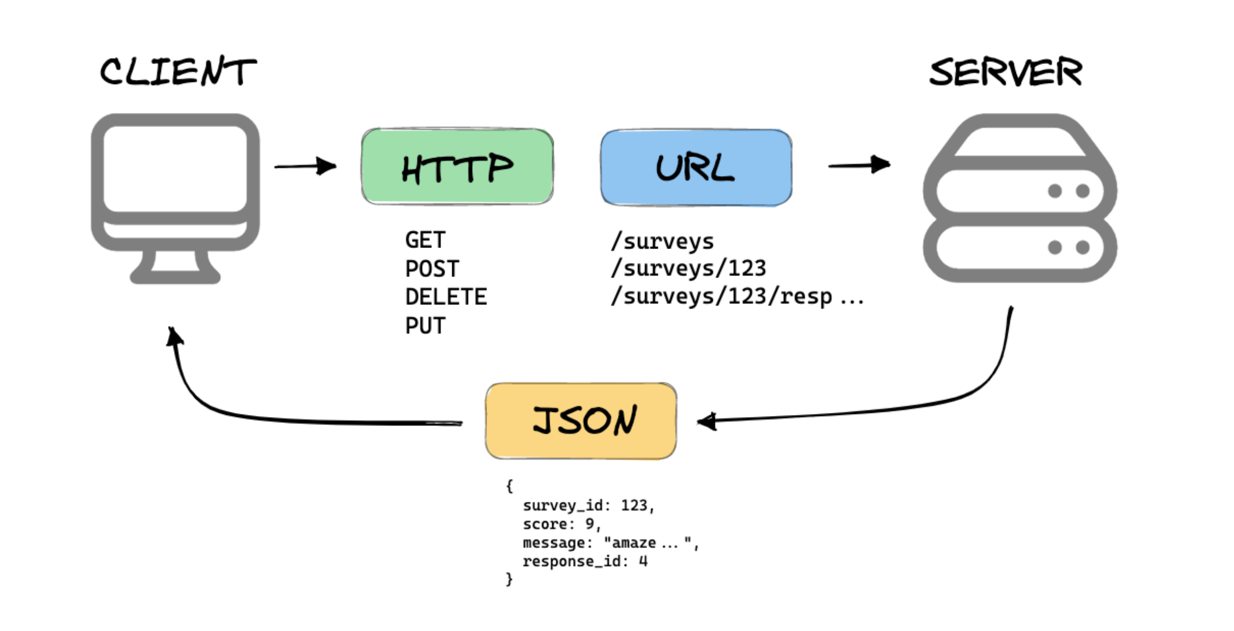
Methodology

Executive Summary

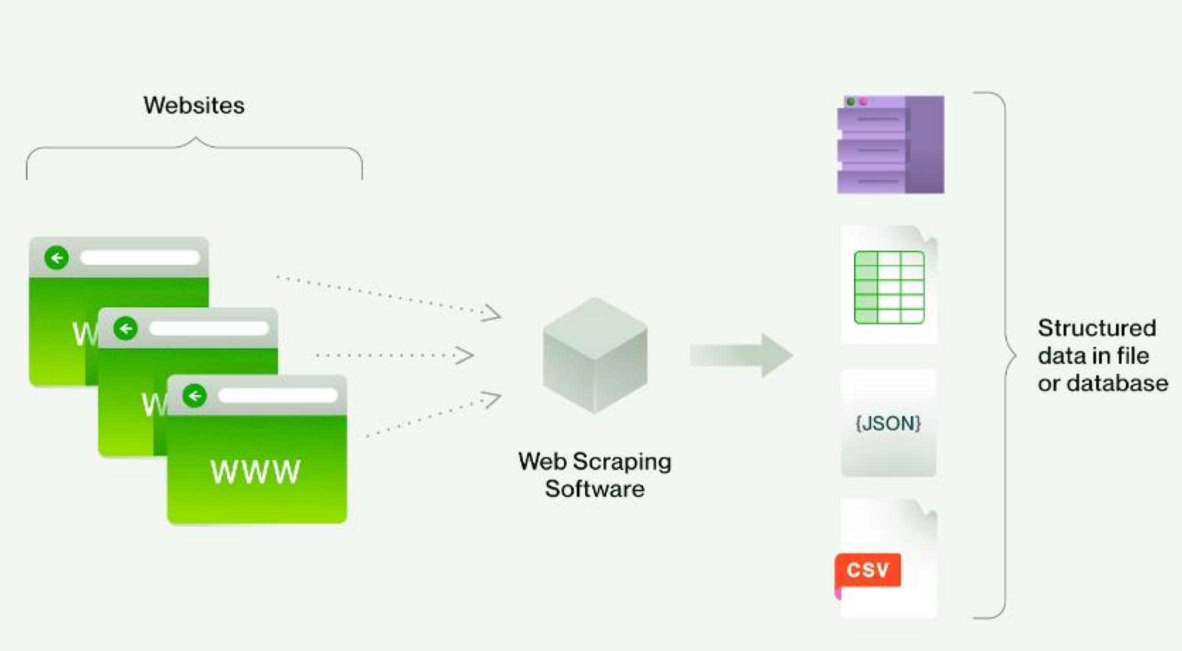
- **Data Collection Methodology:** Collected Falcon 9 first-stage landing data using a **REST API** and **web scraping.**
- **Data Wrangling:** Transformed collected data into a Pandas data frame and performed initial data wrangling by **cleaning** & **processing** the data.
- **Exploratory Data Analysis (EDA):**
  - Conducted exploratory data analysis by manipulating data, created scatter plots and bar charts for data analysis in a **Pandas** data frame.
  - Executed **SQL** queries to select and sort data for analysis.
  - Utilized **data visualization** skills to extract meaningful patterns.
- **Interactive Visual Analytics:**
  - Built an interactive dashboard with **Plotly Dash** containing pie charts and scatter plots.
  - Calculated distances, generated interactive maps, plotted coordinates, and marked clusters with **Folium**.
- **Predictive Analysis using Machine Learning Classification Models:**
  - Trained different classification models, including **Logistic Regression, SVM**, **Decision Trees**, and **KNN**.
  - Optimized hyperparameters using **grid search**.
  - Utilized machine learning skills to build a **predictive model** for determining first-stage Falcon 9 landing success.

Data Collection

- **Data Sets Collection Process:**
  - **REST API Usage:** Data retrieval through a REST API.
  - **Web Scraping:** Data gathering via web scraping.



Using REST API



Using Web Scraping

Data Collection – SpaceX API

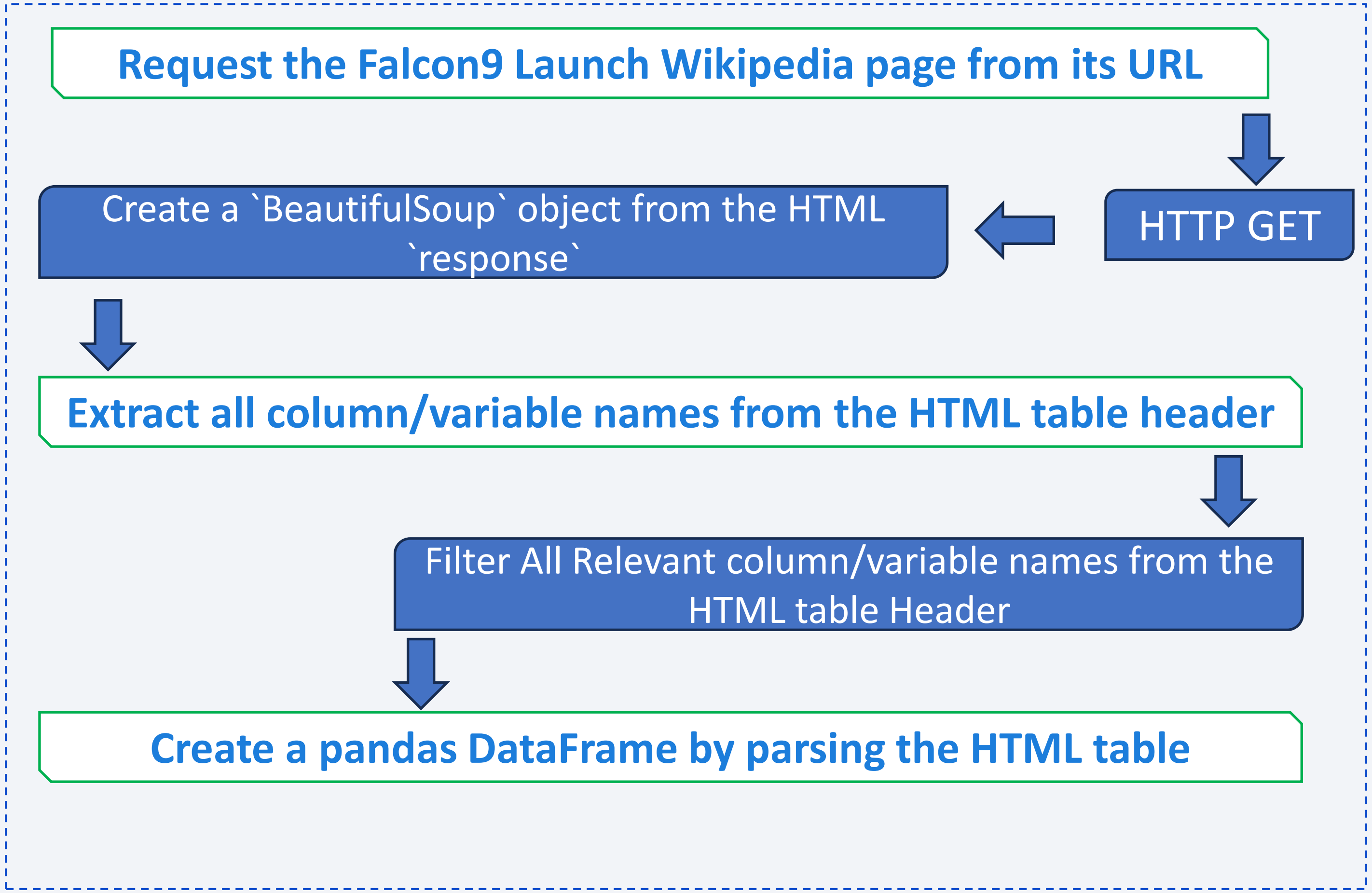
- **SpaceX REST GET :**Requested and parse the SpaceX launch data using the GET request to get the JSON Object.
- Converted JSON into the pandas DataFrame.
- Filtered the DataFrame to only include `Falcon 9` launches

```
spacex_url = "https://api.spacexdata.com/v4/launches/past"

response = requests.get(spacex_url)

data_json = response.json()

data = pd.json_normalize(data_json)

data_falcon = data1[data1['BoosterVersion']='Falcon 9']
```


GitHub

Data Collection - Scraping

- **Web scrap Falcon 9 launch records with `BeautifulSoup`:**
  - Extracted a Falcon 9 launch records HTML table from Wikipedia.
  - Parsed the table and converted it into a Pandas dataFrame.



Data Wrangling

- **Data Processing Description:**

  - Dealt with **missing values** for the columns 'PayloadMass' & 'LandingPad' by replacing them with the averages.

  - Observed the bar chart Visualizations - number of launches on each site, number and occurrence of each orbit, number and occurrence of mission outcome per orbit type.

  - Created a New column named "**Class**" having data type as bool for landing outcome label from Outcome column.

GitHub

EDA with Data Visualization

- Charts Summary:
  - **Charts Plotted:**
    - **Scatter Plots:**
      - Flight Number vs. Payload Mass, overlaying the Launch Success(Class) :(To see the launch success with payload mass)
      - Flight Number vs Launch Site, overlaying the Launch Success(Class) :(To see the launch success at Launch sites)
      - Payload Mass vs Launch Site, overlaying the Launch Success(Class) :(To see the launch success with Both PayloadMass & Launch site)
      - Flight Number vs Orbit type, overlaying the Launch Success(Class) :(To see the launch success with with orbit types)
      - Payload Mass vs Orbit type, overlaying the Launch Success(Class) :(To see the launch success with both Payload Mass & Orbit type)
    - **Bar Chart:**
      - Avg Launch Success rate for each Orbit type :(To see the launch success rate with orbit type)
      - Avg Launch Success rate for Each year :(To see the launch success rate per year)
    - **Line plot:**
      - Launch success yearly trend :(To see the continuous trend for launch success with year)
  - **Reasons:**

GitHub

EDA with SQL

- **SQL Queries Summary:**
  - Displayed unique launch site names in the space mission data.
  - Displayed 5 records where launch sites begin with 'CCA'.
  - Displayed the total payload mass carried by boosters launched by NASA (CRS).
  - Displayed the average payload mass carried by booster version F9 v1.1.
  - Listed the date of the first successful landing outcome on the ground pad using the min function.
  - Listed the names of boosters with success in drone ship landings and payload mass between 4000 and 6000.
  - Listed the total number of successful and failure mission outcomes.
  - Listed the names of booster versions that carried the maximum payload mass using a subquery.
  - Listed records displaying month names, failure landing outcomes on the drone ship, booster versions, and launch sites for the year 2015.
  - Ranked the count of landing outcomes (e.g., Failure (drone ship) or Success (ground pad)) between June 4, 2010, and March 20, 2017, in descending order.

GitHub

Build an Interactive Map with Folium

- **Utilized Map Objects Summary:**
  - **Markers:** Added markers to represent launch sites on the map.
  - **Circles:** Utilized circles to visualize launch site proximity areas.
  - **Marker Clusters:** Implemented marker clusters to showcase successes at each launch site.
  - **Polylines (Lines):** Incorporated polylines to showcase distances between proximities and launch sites.
- **Object Justification:**
  - **Markers:** Added for clear launch site identification and location reference.
  - **Circles:** Included to display launch site proximity for analytical purposes.
  - **Marker Clusters:** Utilized for efficient grouping and display of success outcomes.
  - **Polylines (Lines):** Used to showcase distances and spatial relationships for better data interpretation.

GitHub

Build a Dashboard with Plotly Dash

- **Dashboard Components Summary:**
  - **Pie Charts:** Added pie charts to visualize launch success distributions among all sites. And launch success & failure distributions for each Individual sites also(By changing the dropdown) .
  - **Scatter Plots:** Included scatter plots to showcase Correlation between Payload Mass and Launch Success for All Sites and individual sites as well.
  - **Dropdown Menus:** Implemented dropdown menus for interactive site selection.
  - **Slider Controls:** Utilized slider controls to filter data based on payload mass.
- **Component Justification:**
  - **Pie Charts:** Added for a clear overview of launch success and failure rates.
  - **Scatter Plots:** Included to allow users to explore relationships and patterns in the data.
  - **Dropdown Menus:** Implemented for interactive site selection, enhancing user experience.
  - **Slider Controls:** Utilized to filter data by payload mass, enabling data exploration.

GitHub

Predictive Analysis (ML Classification)

- **Model Development Summary:**
  - **Built Multiple Models:** Developed various ML classification models to predict Falcon 9 first-stage landing success.
  - **Evaluation and Improvement:** Evaluated model performance using metrics and iteratively improved them.
  - **Best Performing Model:** Identified the best-performing classification model for accurate predictions.
- **Model Development Process:**

Data Splitting (Train, Test): → Model Selection (e.g. Logistic Regression, SVM, Decision Trees, KNN)

↓

Hyperparameter Tuning (Optimized model hyperparameters)

↓

Evaluation Metrics: key metrics (e.g., accuracy, F1-score) for model performance.

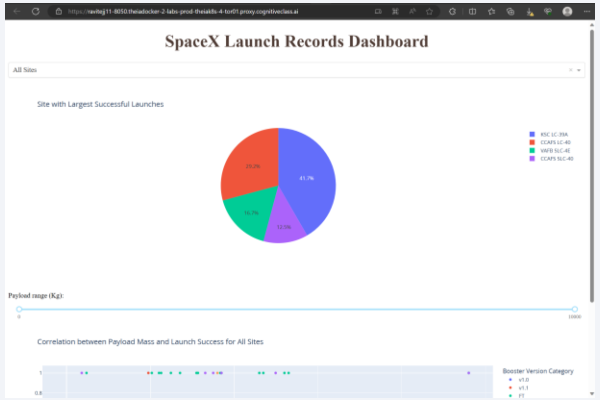
↓

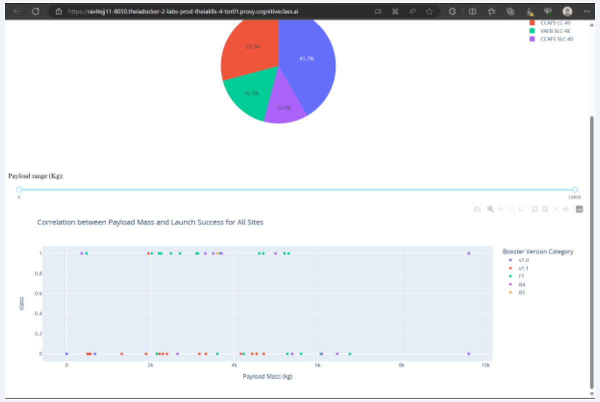
Select Model having highest performance (Evaluation Metrics)

GitHub

Results

- **Exploratory Data Analysis Results:**
  - **Key Findings:**
    - **Success Rates:** Uptrend with progressing year, except some of the downs in 2018 & 2020.
    - **Payload Mass distribution:** Success is more within (1952,5300).
    - **Launch Site analysis:** Highest Success Launches at KSC LC – 39A.
- **Interactive Analytics Demo Screenshots:**





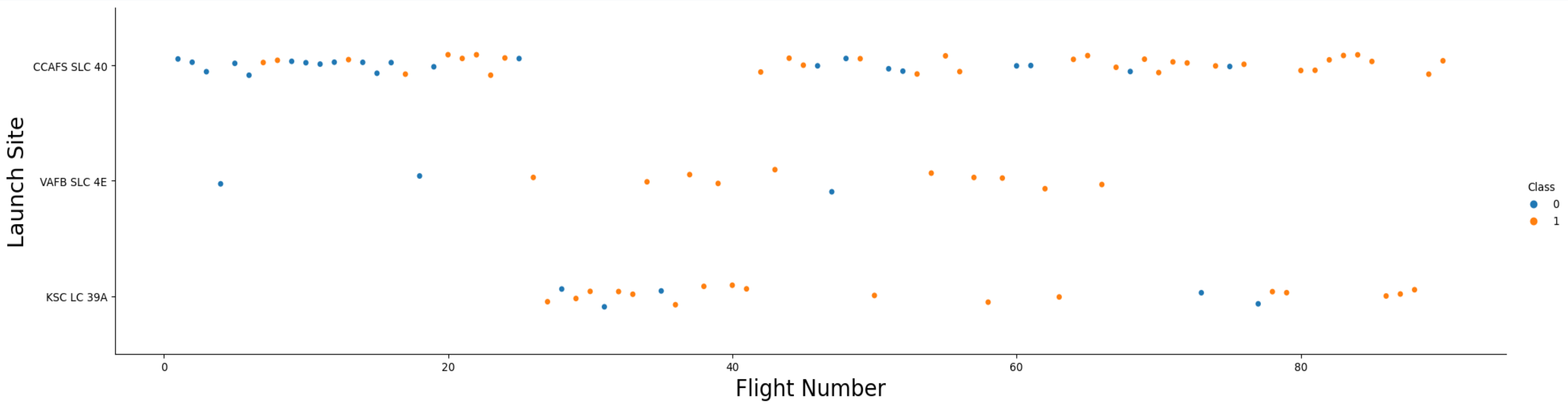
- **Predictive Analysis Results:**
  - **Model Performance: (** Best Performance Model - DECISION TREE CLASSIFIER)
    - **Accuracy:** '0.8888888888888888' (For - DECISION TREE CLASSIFIER)
    - **False Positives in the CONFUSION MATRIX:** '1' (For - DECISION TREE CLASSIFIER)
    - **Key Findings:** With the given Features about the Falcon 9, we can Predict the First stage launch success close to 89%.

GitHub

Section 2

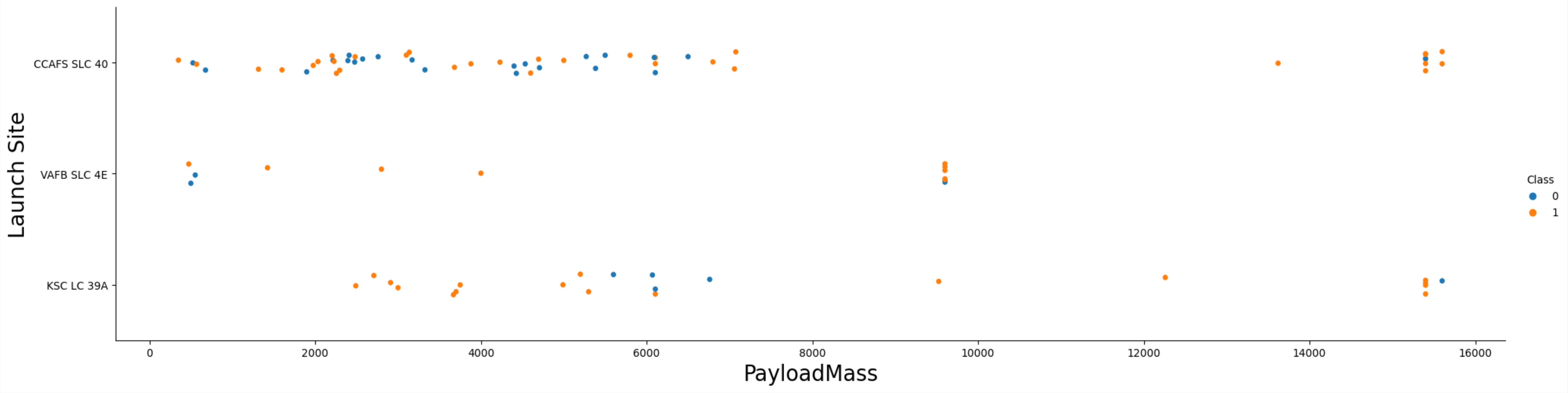
Insights drawn from EDA

Flight Number vs. Launch Site



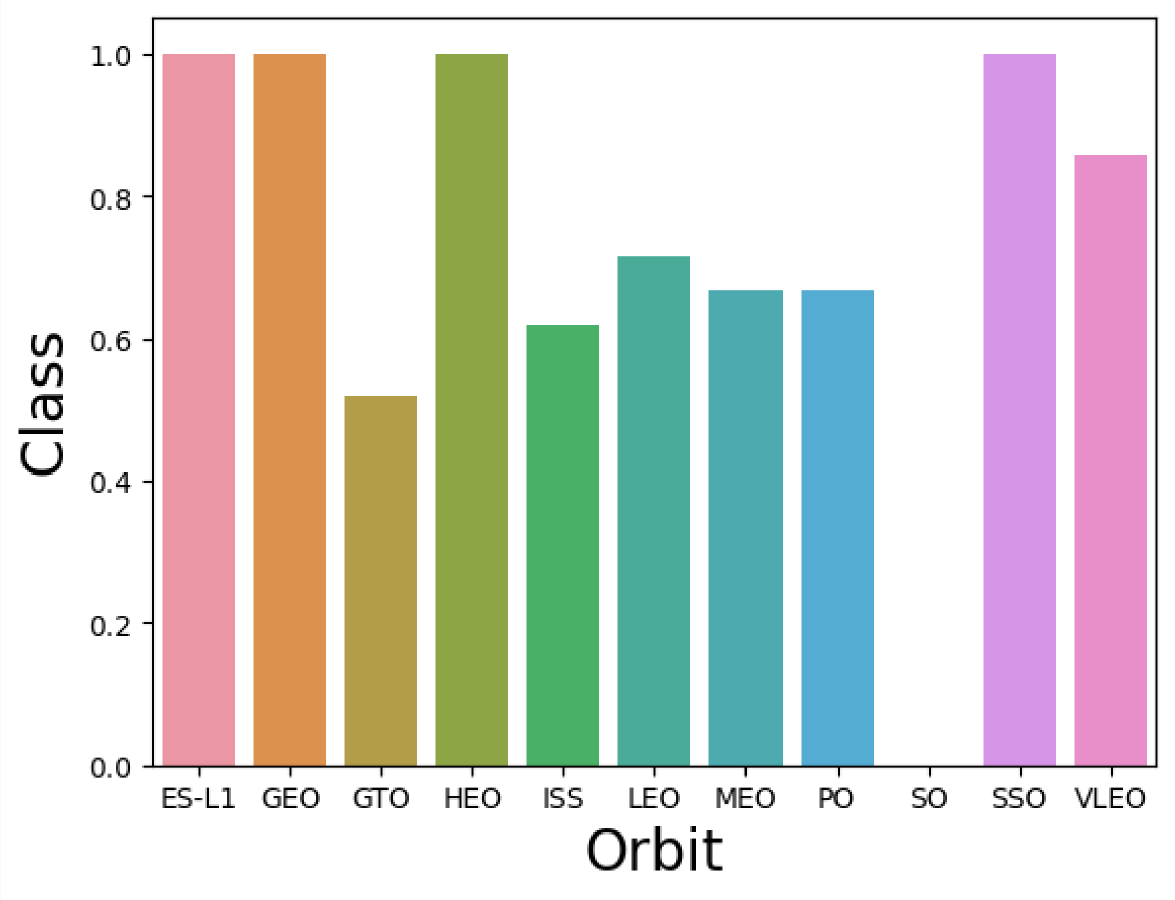
- Explanations:
  - For Launch site 'CCAFS 40' Success has increased with higher Flight Number.
  - For Launch Site 'KSC LC 39A' very few (=5) of Failures have been there.

Payload vs. Launch Site



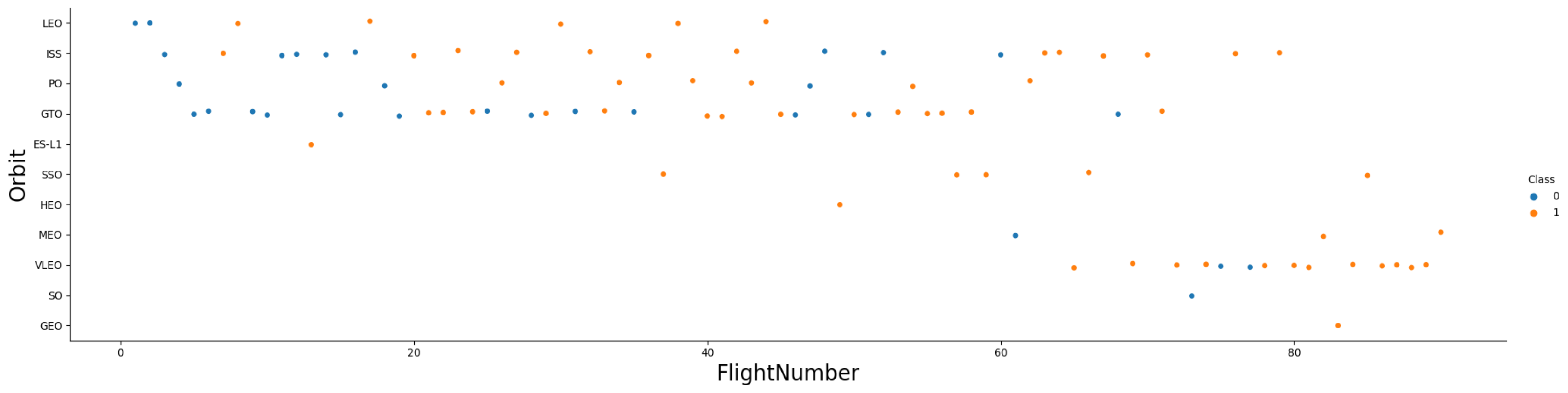
- Explanations:
  - For Launch site ‘CCAFS 40’ & ‘KSC LC 39A’ Success has also been for heavy payload mass(>10000kg).
  - For Launch Site ‘VAFB-SLC’, there are no rockets launched for heavy payload mass(>10000kg).

Success Rate vs. Orbit Type



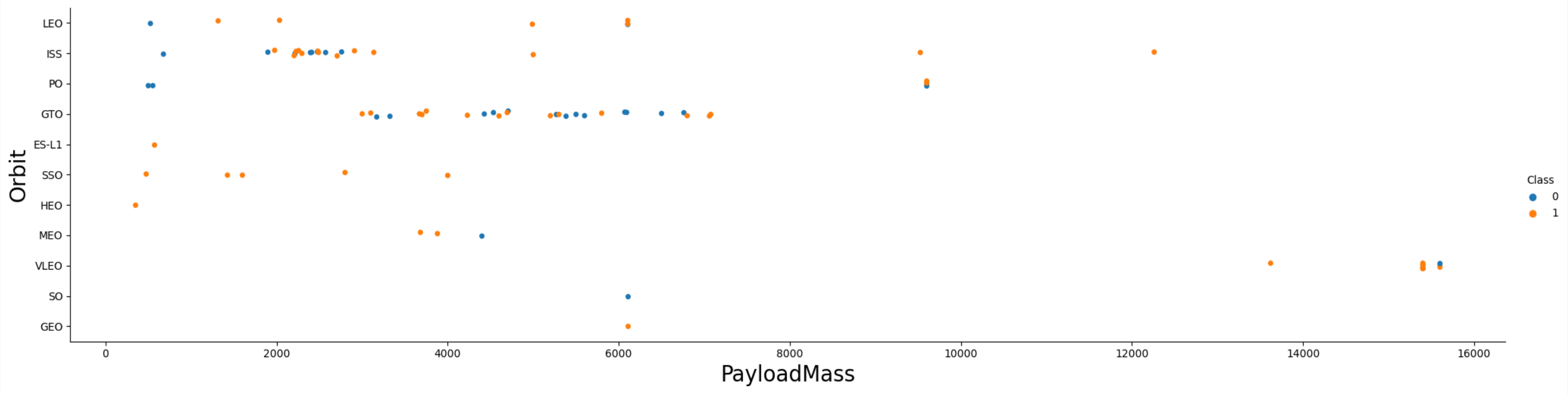
- Explanations:
  - For Orbit types 'ES-L1(Lagrange point)', 'GEO(Circular Geosynchronous)', 'HEO(highly Elliptical Orbit)', and 'SSO(Sun-synchronous orbit)' Success rate has been **100%.**
  - For Orbit type 'GTO(Geosynchronous)' Success rate has been **least** (around **50%).**

Flight Number vs. Orbit Type



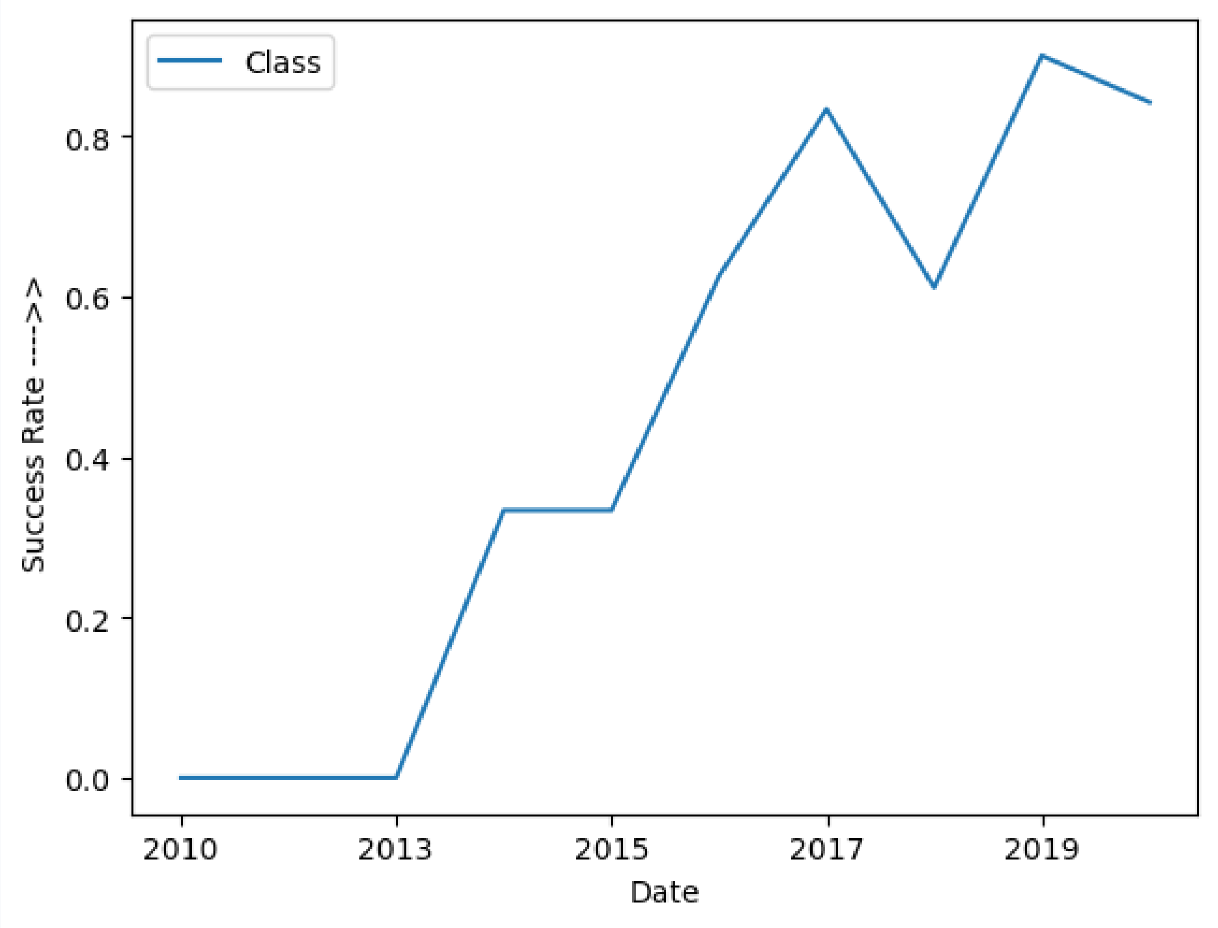
- Explanations:
  - At Larger Flight Number(around 80 and greater than 80), almost all the orbit types are giving Success
  - 'LEO(Low Earth orbit)' orbit the Success appears related to the number of flights.
  - There seems to be no relationship between flight number when in 'GTO(Geosynchronous)' orbit.

Payload vs. Orbit Type



- Explanations:
  - There have been very Few Launches with heavy Payload Mass(>8000kg), and most of them have success.
  - With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS.
  - However for GTO we cannot distinguish this well as both positive landing rate and negative landing(Unsuccessful Mission) are both there here.

Launch Success Yearly Trend



- **Explanations:**
  - Success rate has been increased monotonically over time from 2013 onwards, except some of the small falls in 2017 and 2020.

All Launch Site Names

- **'CCAFS LC-40’**
- **'CCAFS SLC-40’**
- **'KSC LC-39A’**
- **'VAFB SLC-4E’**

	Launch Site	Lat	Long
0	CCAFS LC-40	28.562302	-80.577356
1	CCAFS SLC-40	28.563197	-80.576820
2	KSC LC-39A	28.573255	-80.646895
3	VAFB SLC-4E	34.632834	-120.610745

- Maximum Lunches from site **'CCAFS LC-40’ .**

Launch Site Names Begin with 'CCA'

	Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
0	2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
1	2010-08-12	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of...	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2	2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
3	2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
4	2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

- 5 Records where launch sites begin with `CCA`

- There are only two site names begin with `CCA`:
  - **'CCAFS LC-40’**
  - **'CCAFS SLC-40’**

Total Payload Mass

- **Query Result:** Total Payload Mass Carried by NASA (CRS) Boosters: 45596 kg

- **Explanation:**
  - Understanding the total payload mass carried by NASA is crucial for mission planning, cost estimation, and evaluating NASA's contribution to space missions.
  - It provides valuable insights into the significance of NASA's involvement in space exploration.

Average Payload Mass by F9 v1.1

- **Query Result:** Average Payload Mass Carried by Booster Version 'F9 v1.1': **2928.4 kg**

- **Explanation:**
  - Analyzing the average payload mass for "F9 v1.1" booster versions provides insights into their performance and capabilities.
  - Aiding in mission planning and cost-efficiency assessments.

First Successful Ground Landing Date

- **Query Result:** Date of First Successful Landing on Ground Pad: **2015-12-22**

- **Explanation:**
  - The first successful ground landing outcome marks a historic achievement for SpaceX.
  - It significantly reduced launch costs by enabling booster reusability.
  - This milestone is a testament to SpaceX's innovation and commitment to revolutionizing space travel.

Successful Drone Ship Landing with Payload between 4000 and 6000

- **Query Result:**
  - Names of Desired Boosters (Dron ship Landing Criteria)
    - **F9 FT B1022**
    - **F9 FT B1026**
    - **F9 FT B1021.2**
    - **F9 FT B1031.2**

- **Explanation:**
  - Identifying boosters with successful drone ship landings in this specific payload range is essential for evaluating their performance and reliability in handling medium-sized payloads for cost-effective space missions.

Boosters Carried Maximum Payload

- **Query Result:**
  - Booster Versions with Maximum Payload Mass:
    - **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**

- **Explanation:**
  - Identifying the booster(s) that achieved the maximum payload mass showcases their exceptional capabilities in handling heavy payloads.
  - It is pivotal for cost-effective space missions and demonstrates SpaceX's engineering achievements.

2015 Launch Records

- **Query Result:** List of failed landing outcomes in drone ship, along with their booster versions and launch site names for the year 2015.

“Month”	“ Landing Outcome”	“Booster Version”	“Launch Site”
October	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
April	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40

- **Explanation:**
  - Analyzing the launch records from 2015, specifically focusing on failed landing outcomes in drone ships, provides valuable insights for SpaceX's continuous improvement in space missions.
  - Understanding the specific incidents and their details can lead to enhanced performance and safety.

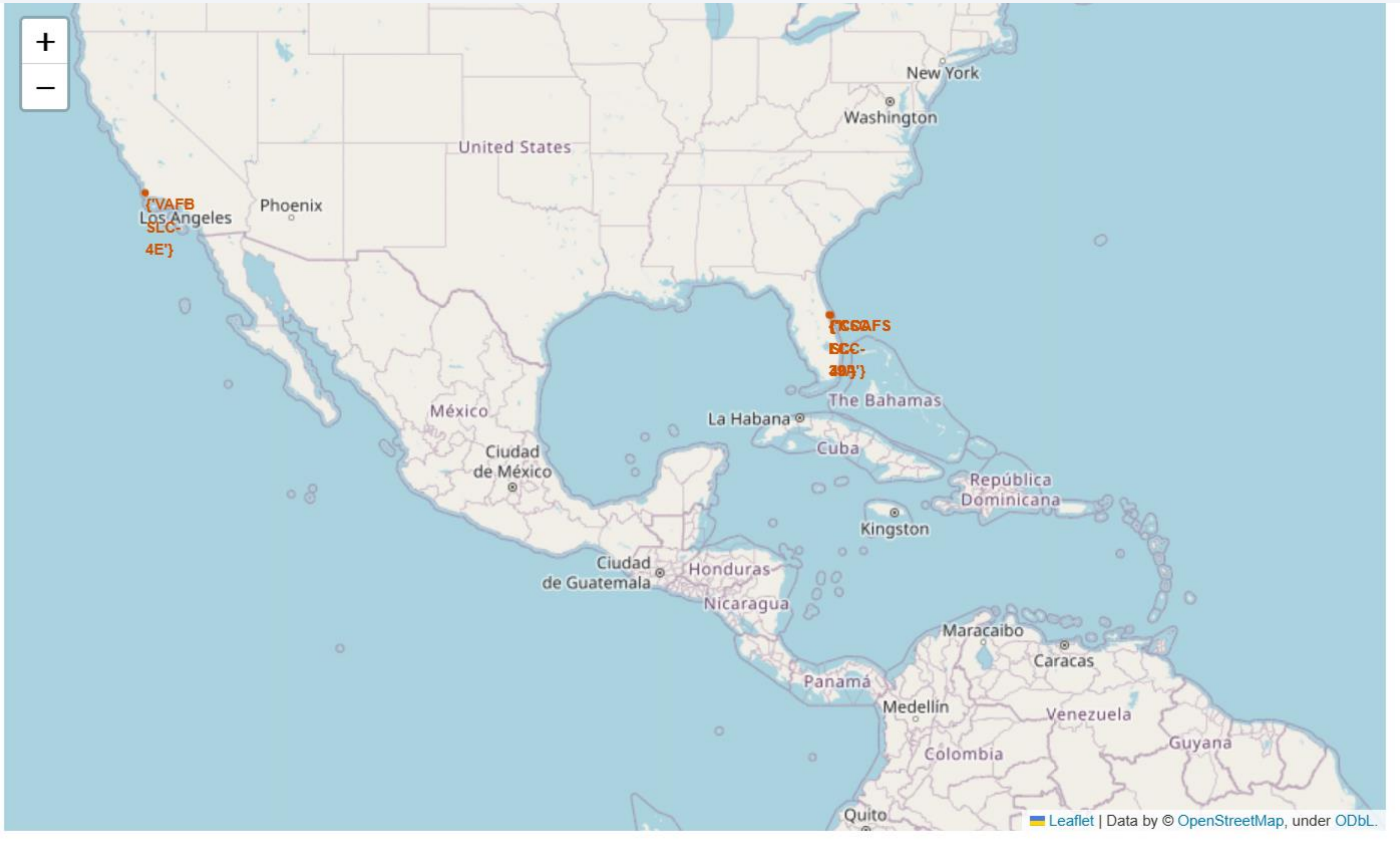
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- **Query Result:**
  - Ranked Landing Outcomes between 2010-06-04 and 2017-03-20 in Descending Order:
    - No attempt :- 10
    - Success (drone ship) :- 5
    - Success (ground pad) :- 5
    - Failure (drone ship) :- 5
    - Controlled (ocean) :- 3
    - Uncontrolled (ocean) :- 2
    - Failure (parachute) :- 1
    - Precluded (drone ship) :- 1
- **Explanation:**
  - Ranking the landing outcomes during this specific time frame offers valuable insights into the overall performance of SpaceX missions.
  - It helps identify trends and patterns in mission outcomes, aiding in continuous improvement and decision-making.

Section 3

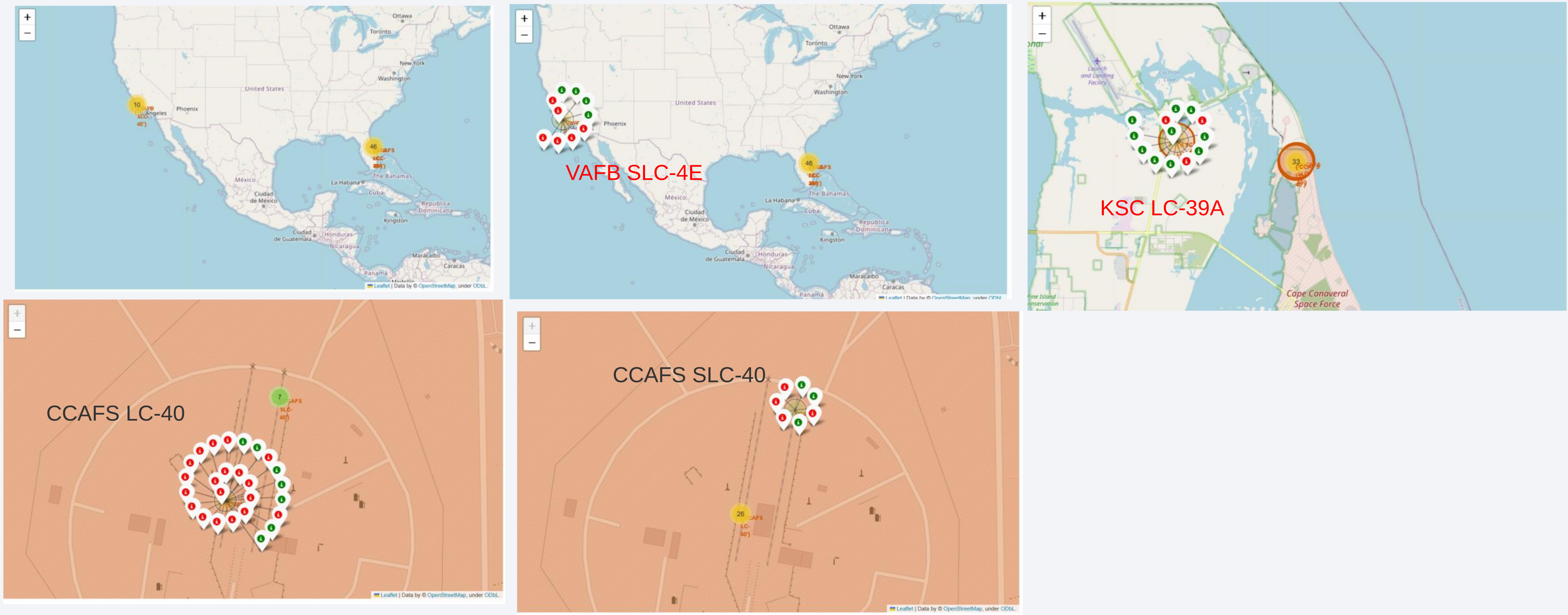
Launch Sites Proximities Analysis

Folium Map: All Launch sites Location Markers on a Global Map



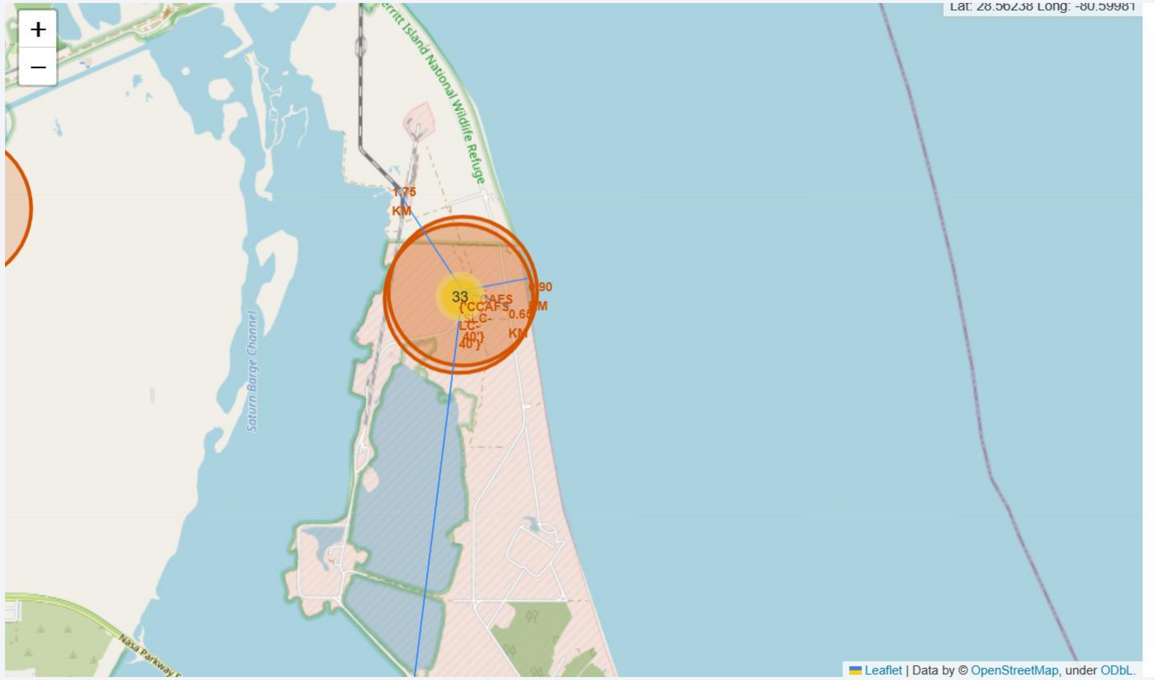
- **Explanation:** Almost all the sites are in very close proximity to the coast.

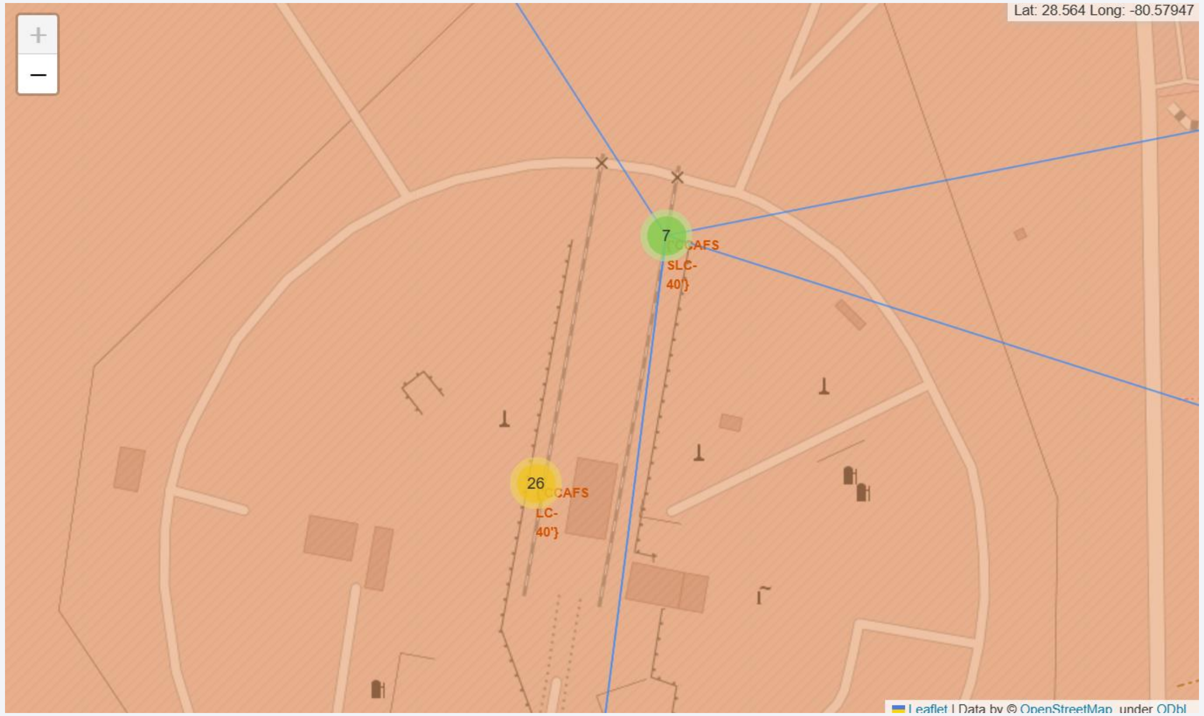
Folium Map: Color Labeled launch Outcome for each site on the map



- Explanation:
  - Site “KSC LC-39A” have Relatively high Success Rate

Folium Map: Distances between a launch site (CCAFS SLC-40) to its proximities.



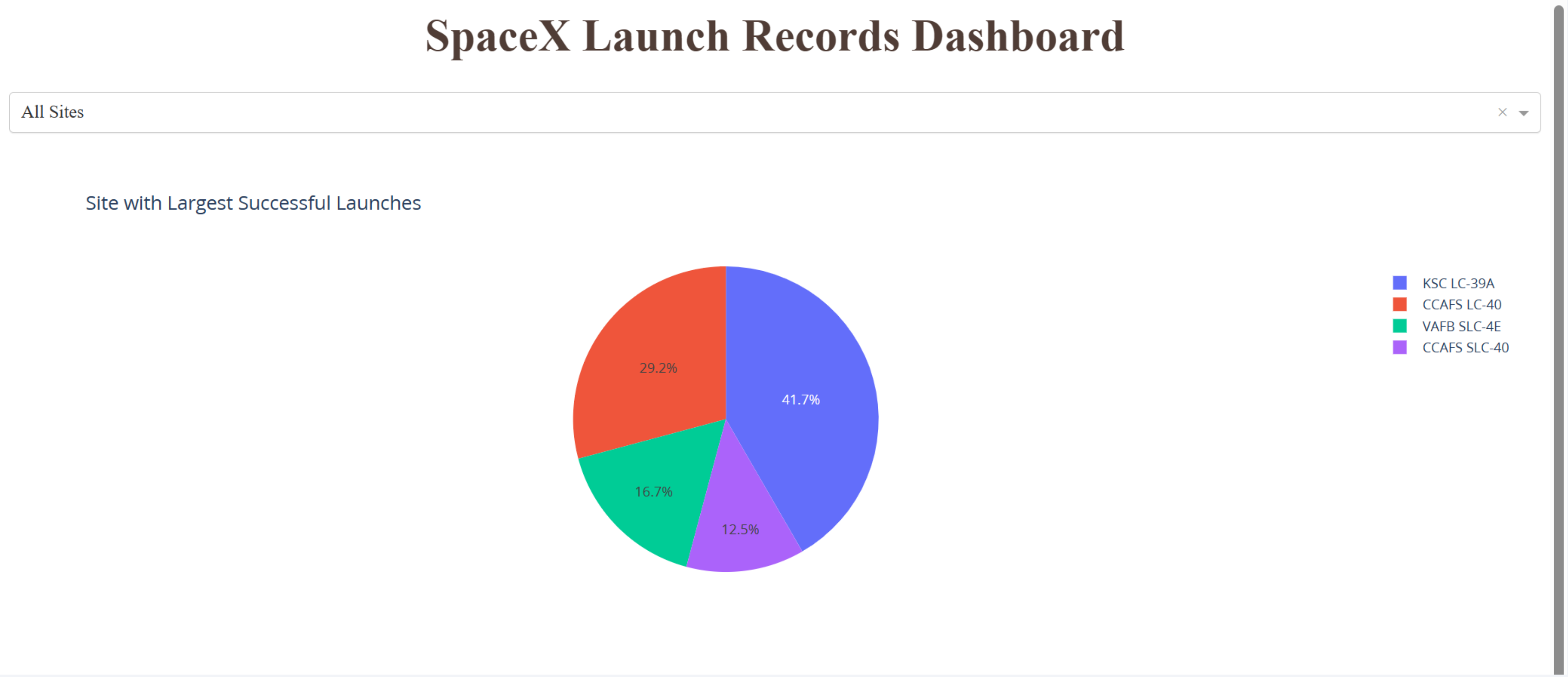


- Explanation:
  - We found that Launch site "CCAFS SLC-40" is in close proximity to Railway, Highway,& Coastline (1.7491725095424817 Km, 0.6532857532514599 Km, and 0.90 km respectively)
  - But not to the city (51.499840490511794 Km).

Section 4

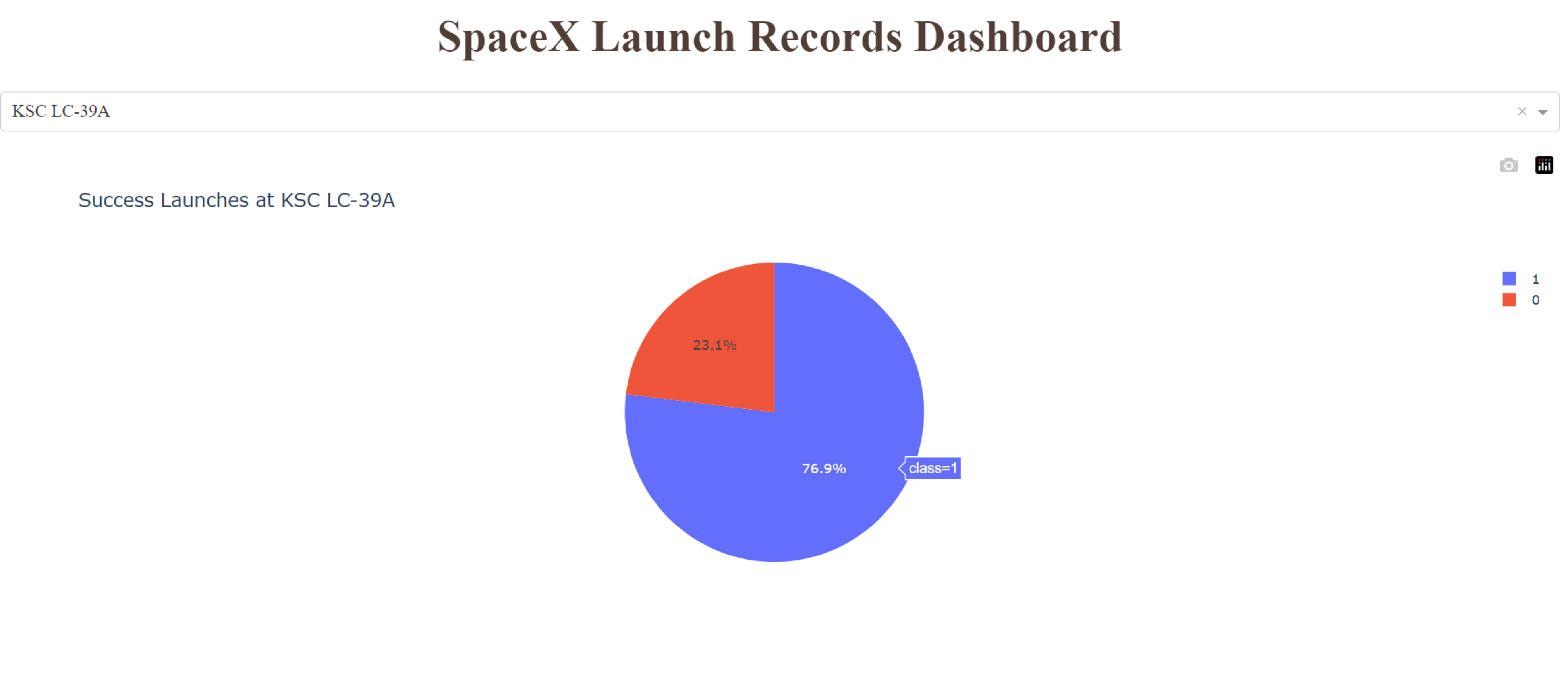
Build a Dashboard with Plotly Dash

Pie Chart: Launch Success counts for All Sites



- Explanation:
  - Site with Largest Successful Launches is "KSC LC-39A".

Pie Chart: Launch Site with Highest Launch Success Ratio



- **Explanation:**
  - At site “KSC LC-39A” Launch landing Success is 76.9%
  - At site “KSC LC-39A” Launch landing Failure is 76.9%

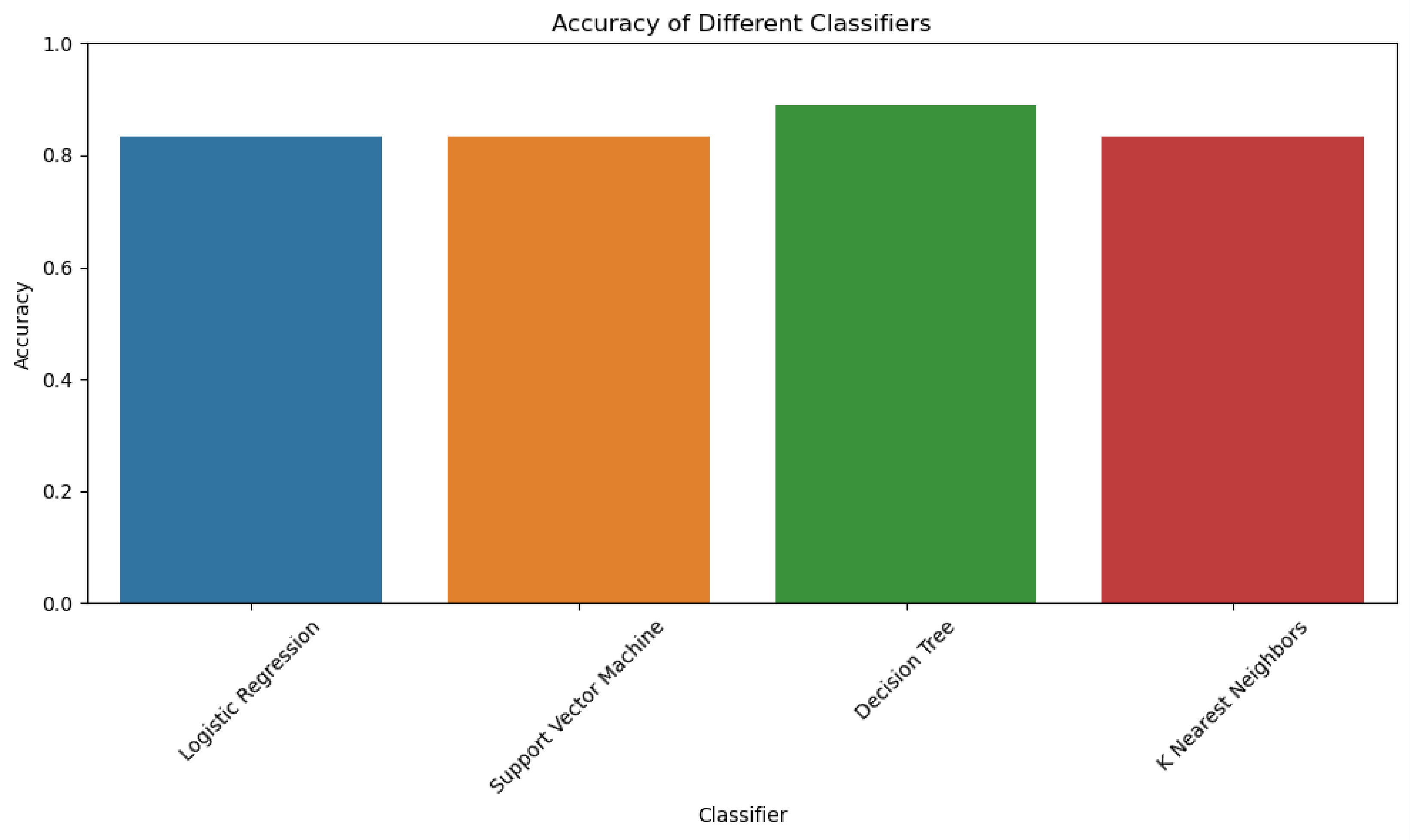
Scatter Plot: Payload vs Launch Outcome for All Sites

- **Explanation:**
  - 2000 kg to 4000 kg Payload mass have highest Success.
  - Booster Version “FT” given highest Success.

Section 5

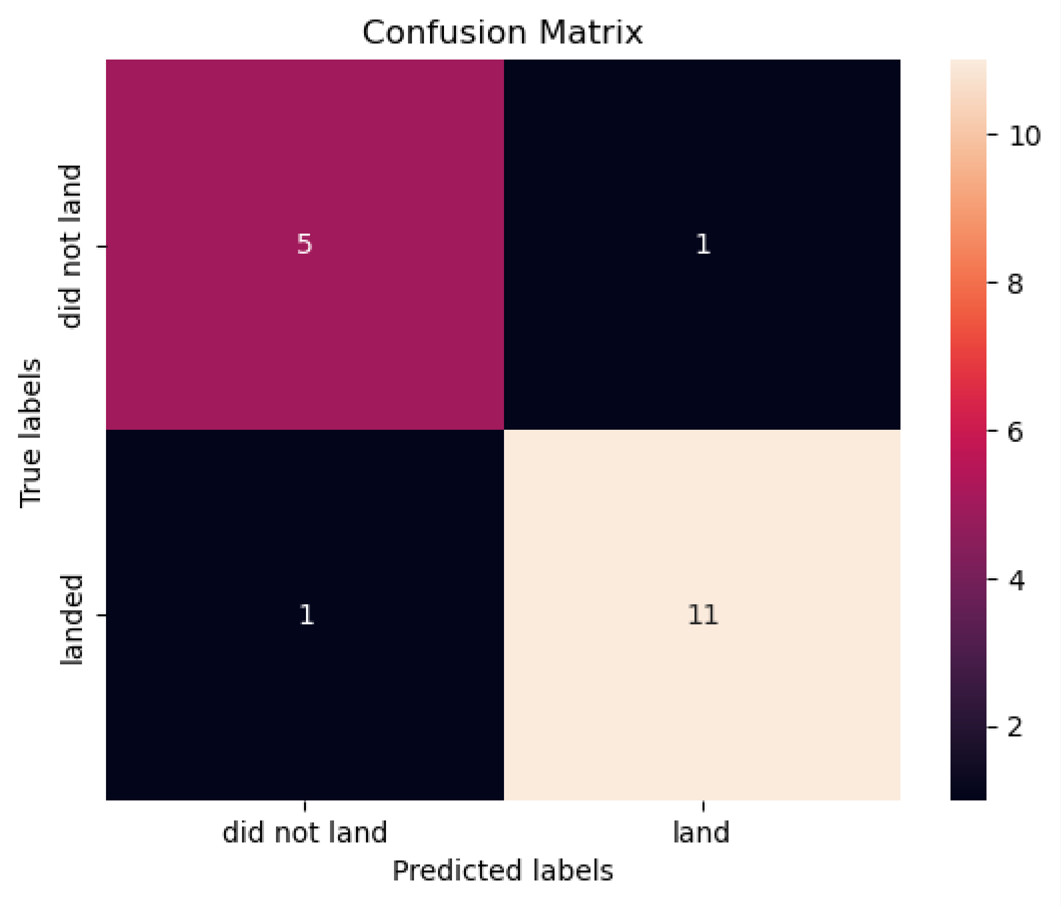
Predictive Analysis (Classification)

Classification Accuracy



- **Highest classification accuracy:**
  - DECISION TREE CLASSIFIER - 0.8888888888888888

Confusion Matrix



- **Decision Tree Classifier** (Best Performance Model)
  - Accuracy: 0.8888888888888888

- **Explanation:**

  Out of 18 test Samples

  - True Positives: 11
  - True Negatives: 5
  - False Negatives: 1
  - False Positives: 1

- In our Objective the False Positives(i.e. we expected to get success in first stage landing But Actually it didn't Land) are very crucial.
- Here we have **Least FALSE POSITIVES (= 1)**
- **Best Optimized Model**

Conclusions

- **Predictive Model Success:**
  - Successfully determines first stage landing outcomes with close to **89% Model accuracy**,
  - Enables accurate **cost estimations** for space missions.
- **Launch Site Analysis:**
  - Analysis of launch sites (Like “KSC LC-39A”) reveals their impact on mission success & **strategic site selection**.
- **Payload Mass Significance:**
  - Payload mass analysis (Like 2000 kg to 4000 kg &) provides insights into booster (Like “FT”) performance, affects mission success.
- **Historical Trends:**
  - historical trends(Like Uptrend from 2013) and patterns informs future mission planning and decision-making.
- **Global View:** A global perspective of launch site locations aids in strategic decision-making and mission planning.
- **Improvement Areas:** Such as enhancing booster capabilities, optimizing Payload Mass, and optimizing launch site selection
- **SpaceX's Competitive Edge:** The project findings contribute to SpaceX's competitive advantage in the space exploration industry.

Appendix

- Data Sets:
  - spacex_url = https://api.spacexdata.com/v4/launches/past
  - Wikipedia = https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922
  - Geo_URL = https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/spacex_launch_geo.csv
  - dataset_part_1.csv , dataset_part_2.csv, dataset_part_3.csv

- Notebooks: Python code snippets used for data analysis, visualization, and modeling.

- SQL Queries: Attach SQL queries used for data manipulation and analysis in the project.

Thank you!