IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Dylanon Wichiramala
18/12/2023

Outline

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

Executive Summary

Summary of methodologies:

- Data Collection: Gathering data from SpaceX api, including information on various parameters that related and the results parameter of success of failure of landing attempts.
- Data Cleansing: Handling missing and irregular data.
- Exploratory Data Analysis: Analyzing distributions, patterns, and correlations between data.
- Feature Engineering: Select feature and enhancing model performance through variable transformations.
- Model Selection and Training: Employed machine learning algorithms, compare, optimized through cross-validation.

Summary of results:

- Promising model performance metrics (accuracy, precision, recall, F1 score), and visualize using confusion matrix.

Introduction

- Project background and context
- Problems you want to find answers

Project Background and Context:

- Focus on creating an ML model for predicting Falcon 9 landing success.
- Utilizing data from SpaceX, including telemetry, weather, and mission details.
- Aiming to enhance decision-making processes and improve operational efficiency.

Problems to Address:

Prediction of Landing Success:

- Develop an ML model to forecast the probability of Falcon 9 landing success.
- Utilize historical data to train and validate the model for accurate predictions.

Identification of Significant Factors:

- Determine the key factors influencing the success or failure of Falcon 9 landings.
- Explore in each parameter details to uncover critical predictors

Section 1

Methodology

Methodology

Executive Summary

- Collected data from SpaceX API
- Perform data wrangling, Addressed missing values, outliers, splitting data, normalize, and perform an one-hot encoding for categorical features.
- Create visualizations and interactive visual analytics using Dashboard Plotly Dash.
- Perform predictive analysis using classification models evaluate and compare between each model. Employed cross-validation techniques to ensure robustness. Adjust parameter for optimal predictive capabilities.

Data Collection

The project acquiring and preparing data from various table sources in SpaceX API . We got different source of data during the extraction phase. After Transforming data, we combined the data into single table. The final step involved loading the processed data into a designated SQL database

Extract

- Utilized the Python Request library to getting data from the SpaceX API.
- Extracted data on rockets, launchpads, payloads, and other relevant information into one table.

Transform

- Gathered data from diverse sources, including rockets, launchpads, payloads, etc.
- Merged and transformed data into a unified table for comprehensive analysis.
- Aligned data structures, ensuring consistency and compatibility.
- Transformed data types and shapes to meet the desired format for analysis.

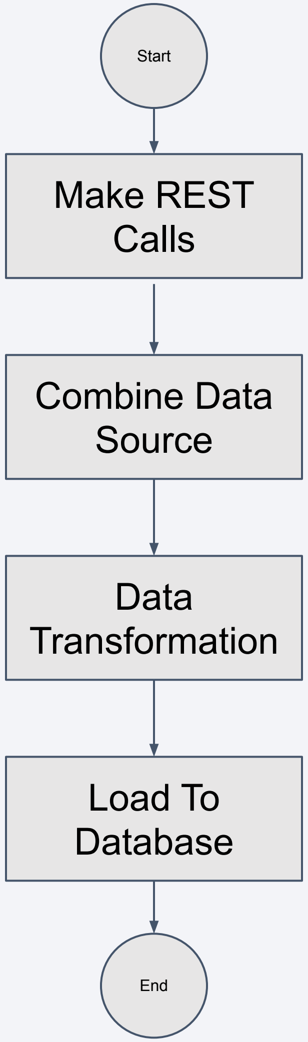
Load

- Prepared a destination database using SQL for efficient data storage.
- Executed SQL queries to load transformed data into the one destination database.

Data Collection – SpaceX API Calls

- Utilized the Python Request library to getting data from the SpaceX API.
- Extracted data on /rockets, /launchpads, /payloads, /cores, and /pasts and combine into single table
- Transform data into a unified format
- Prepare destination database using SQL and execute SQL queries to load data.

Project file: [SpaceY-project/Data collection api.ipynb at main · DylanonWicchiramala/SpaceY-project (github.com)]



Data Wrangling

- Handling Missing Data - Identify missing data using the `pandas.isnull()` function and drop the columns that have null values.
- Data Type and Formatting Conversion
  - Identify Categorical data by look the datatype, Handling Categorical data using one-hot encoding
  - Convert landing outcomes data into 2 categorical value (success, and fail)
  - Casing all numeric data into float64 for ML modeling.
- Normalization the data using standardization.

Project file:[SpaceY-project/Data wrangling.ipynb at main · DylanonWicchiramala/SpaceY-project (github.com)]

EDA with Data Visualization

To understand the relationship between each feature and the landing success rate, we conduct Exploratory Data Analysis (EDA) and represent the findings through various visualizations. This involves comparing each feature with the landing success using charts, such as bar plots illustrating the success rate for each orbit and categorical plots depicting the relationship between launch site and payload mass. Additionally, we explore other visualization techniques to provide a comprehensive understanding of the factors influencing the landing success.

Project file: SpaceY-project/EDA dataviz.ipynb at main · DylanonWicchiramala/SpaceY-project (github.com)

EDA with SQL

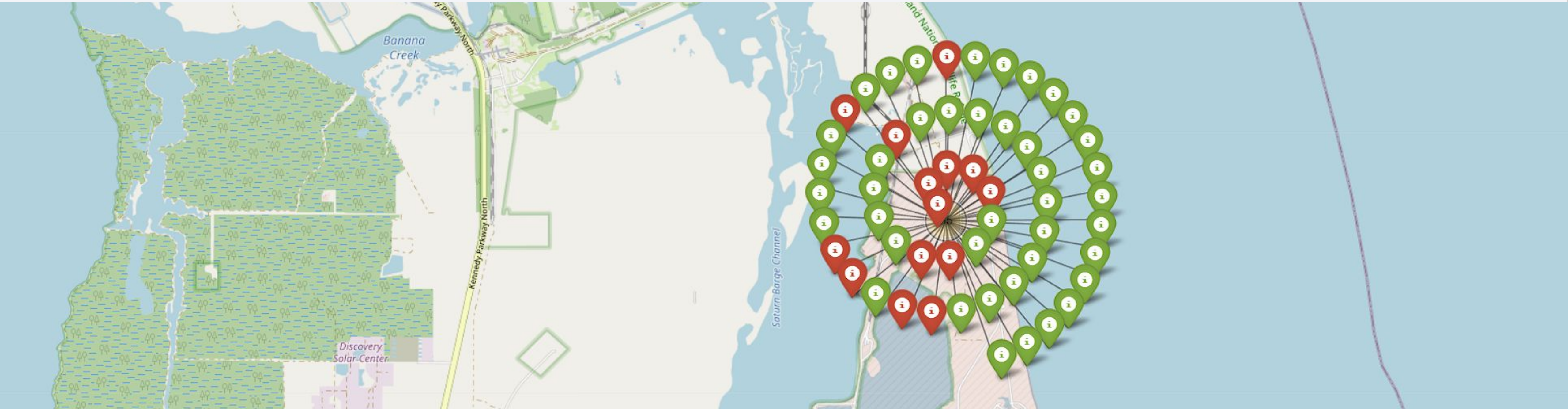
- Total Mission Outcomes:
  - List the total number of successful and failure mission outcomes.
- Successful Drone Ship Landings:
  - List the names of the boosters with success in a drone ship and payload mass between 4000 and 6000.
- Failure Outcomes in 2015:
  - List the records displaying month names, failure landing outcomes on a drone ship, booster versions, and launch sites for the months in the year 2015.
- Rank Landing Outcomes:
  - Rank the count of landing outcomes between the date 2010-06-04 and 2017-03-20 in descending order.
- ETC.

Project file: [SpaceY-project/EDA sql.ipynb at main · DylanonWicchiramala/SpaceY-project (github.com)]

Build an Interactive Map with Folium

In the Folium map, I created and added markers with pop-up information for each spaceflight record from the dataset. Marker colors were based on the 'Class' column, differentiating between categories. A MarkerCluster was used for cleaner visualization, and the map provided an interactive display of spaceflight data, highlighting geographical locations, outcomes, and additional details.

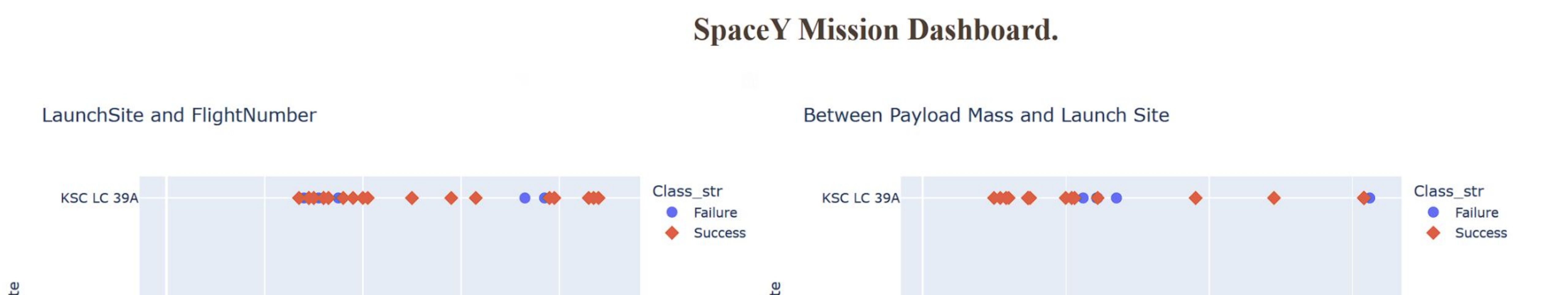
Project file: [SpaceY-project/Dataviz map.ipynb at main · DylanonWicchiramala/SpaceY-project (github.com)]



Build a Dashboard with Plotly Dash

To understand the relationship between each feature and the landing success rate, we conduct interactive dashboard and represent the findings through various visualizations. This involves comparing each feature with the landing success using charts, such as bar plots illustrating the success rate for each orbit and succession rate trend for each year. Additionally, we explore other visualization techniques to provide a comprehensive understanding of the factors influencing the landing success.

Project file:[SpaceY-project/Dashbord.py at main · DylanonWicchiramala/SpaceY-project (github.com)]



Predictive Analysis (Classification)

- **Model Selection:** Explored logistic regression, support vector classifier (SVC), k-nearest neighbors (KNN), and decision tree models for classification.
- **Grid Search Hyperparameter Tuning:** Applied grid search for each model to optimize hyperparameters. Tuned hyperparameters to enhance model performance.
- **Model Training:** Trained logistic regression, SVC, KNN, and decision tree models using the training dataset.
- **Model Evaluation:** Evaluated model performance using accuracy and confusion matrix. Determined that logistic regression, SVC, and KNN achieved high accuracy (1.0).
- **Model Selection:** Chose KNN as the best-performing model due to its fastest training time among the top-performing models.

Project Source:[SpaceY-project/Machine_Learning_Prediction.ipynb at main · DylanonWicchiramala/SpaceY-project (github.com)]

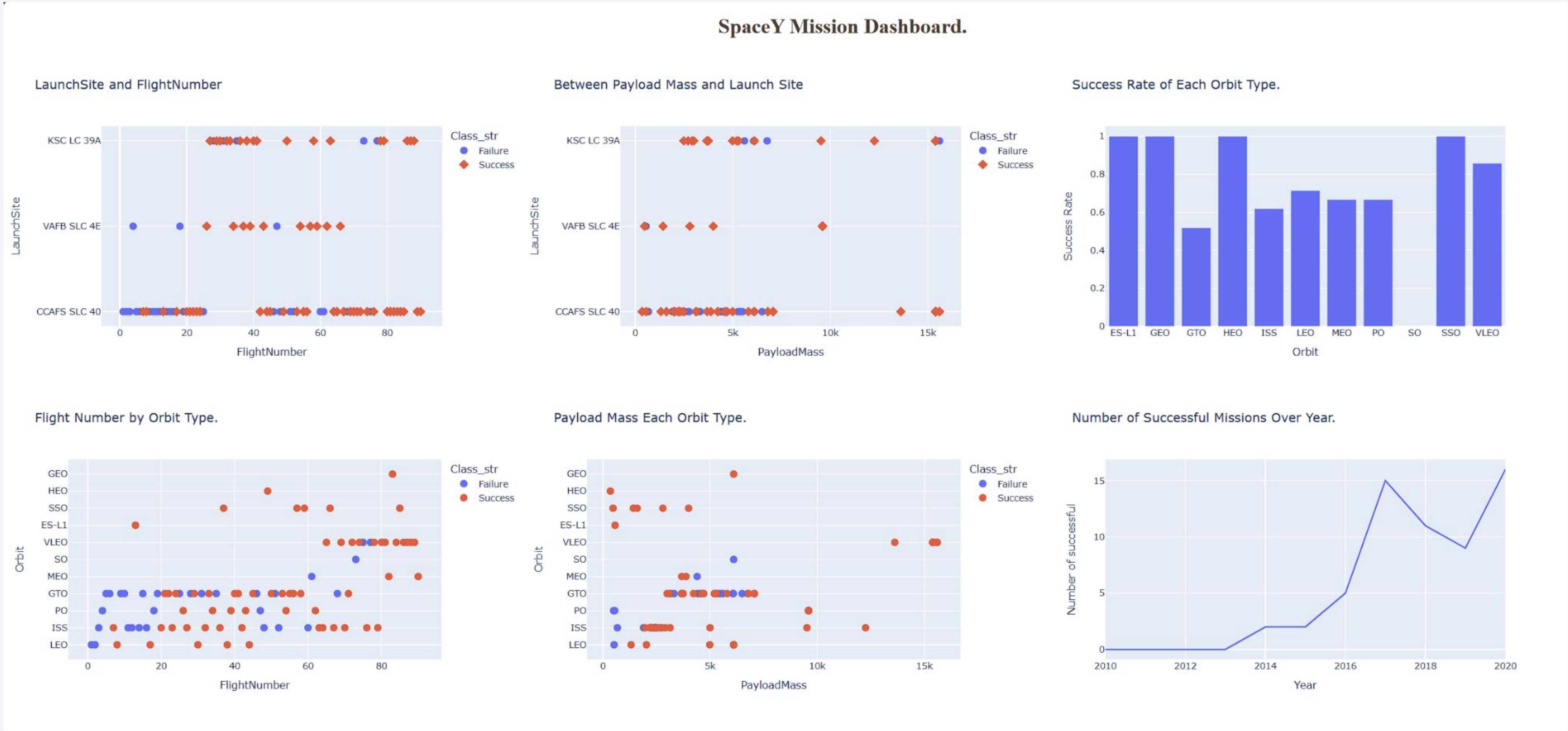
Results (1/2)

- Exploratory data analysis results
  - Explored dataset features, identifying patterns and outliers.
  - Analyzed variable distributions and relationships.
  - Utilized visualizations (e.g., histograms, scatter plots) for insights.

- Predictive analysis results
  - Applied KNN model for prediction.
  - Achieved a model accuracy of 100%.

Results (2/2)

- Interactive analytics demo

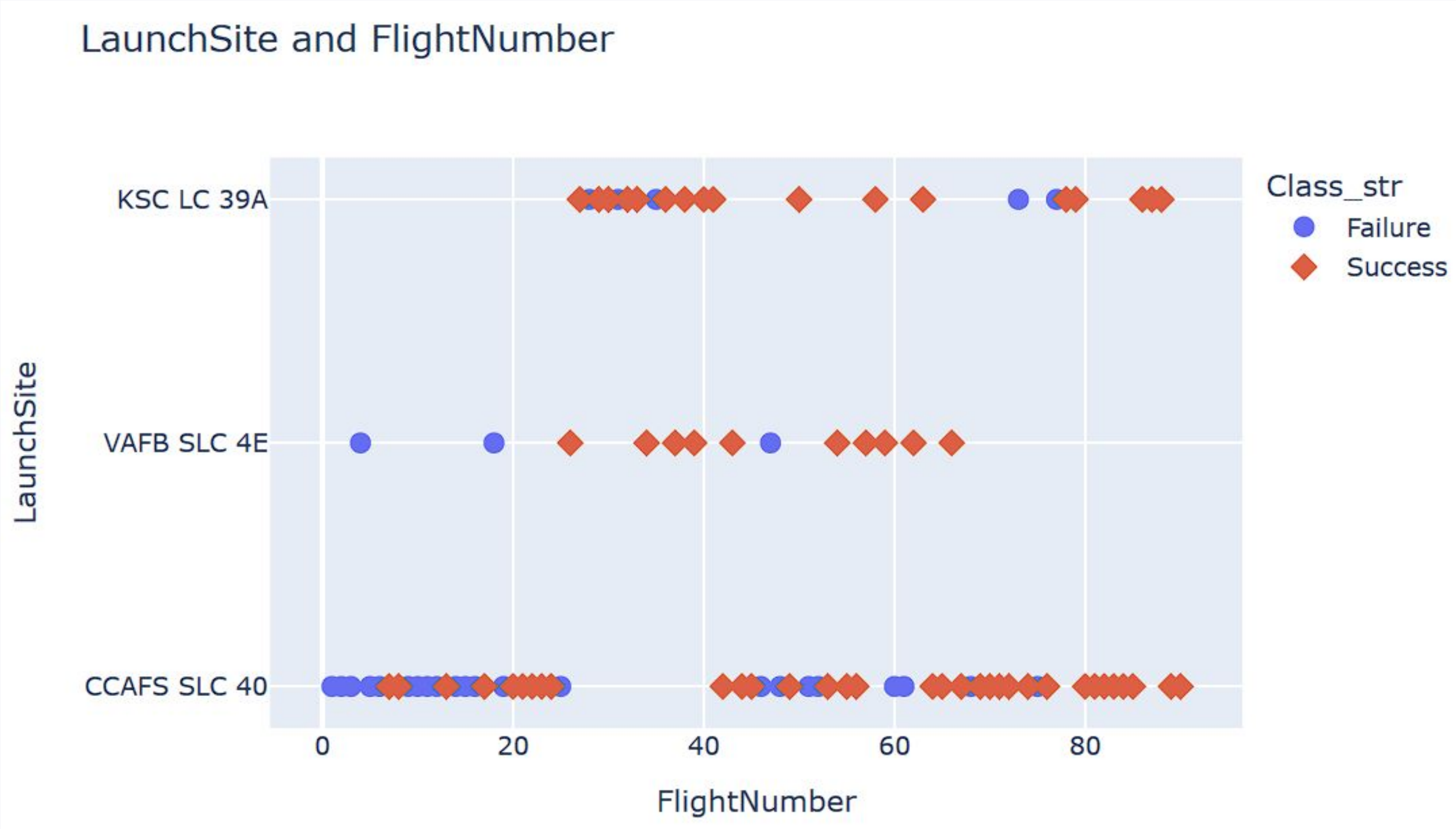


Section 2

Insights drawn from EDA

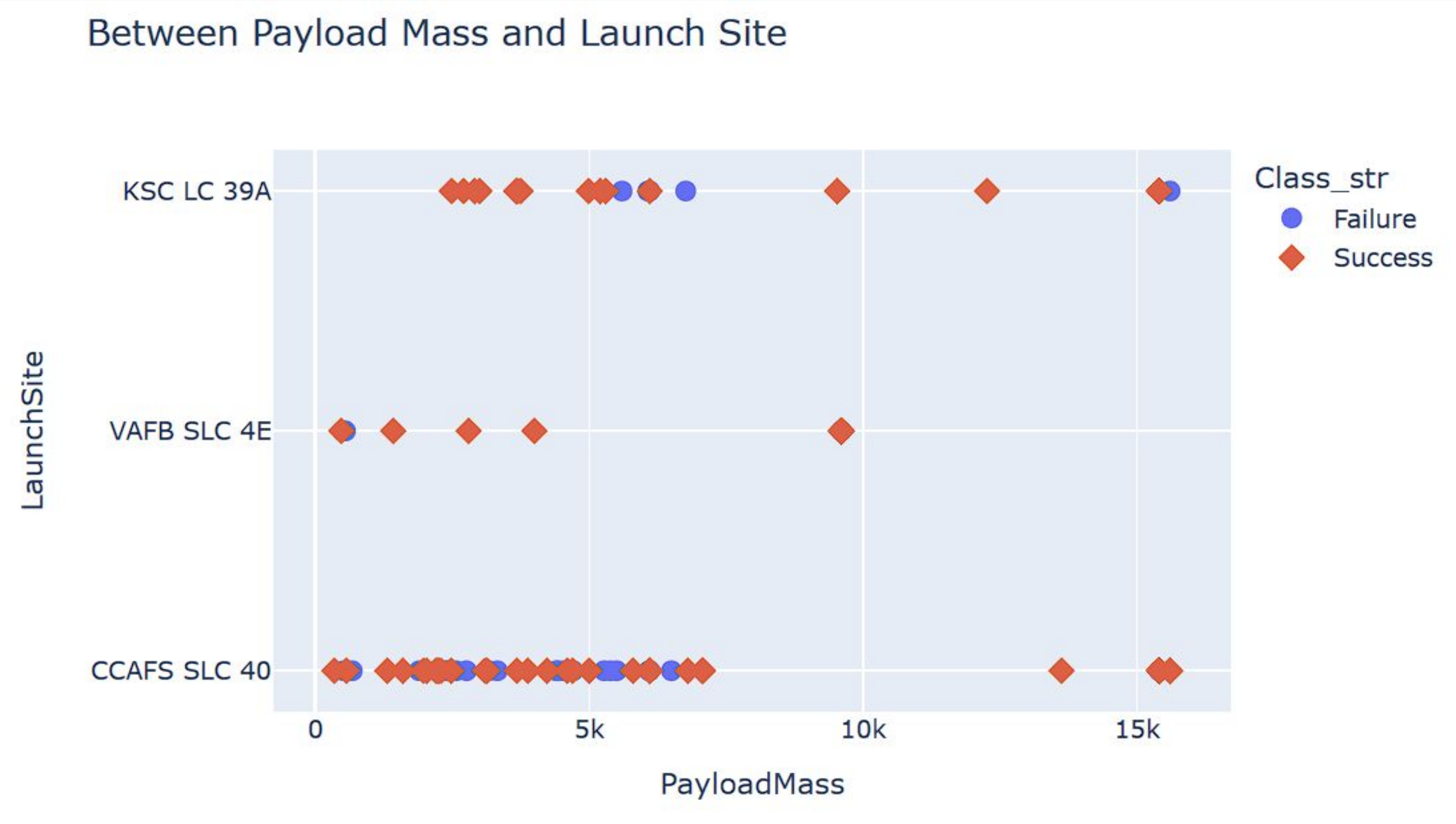
Flight Number vs. Launch Site

This chart indicate the relation between the launch site and flight number, with the color of the plot indicating the mission's success.



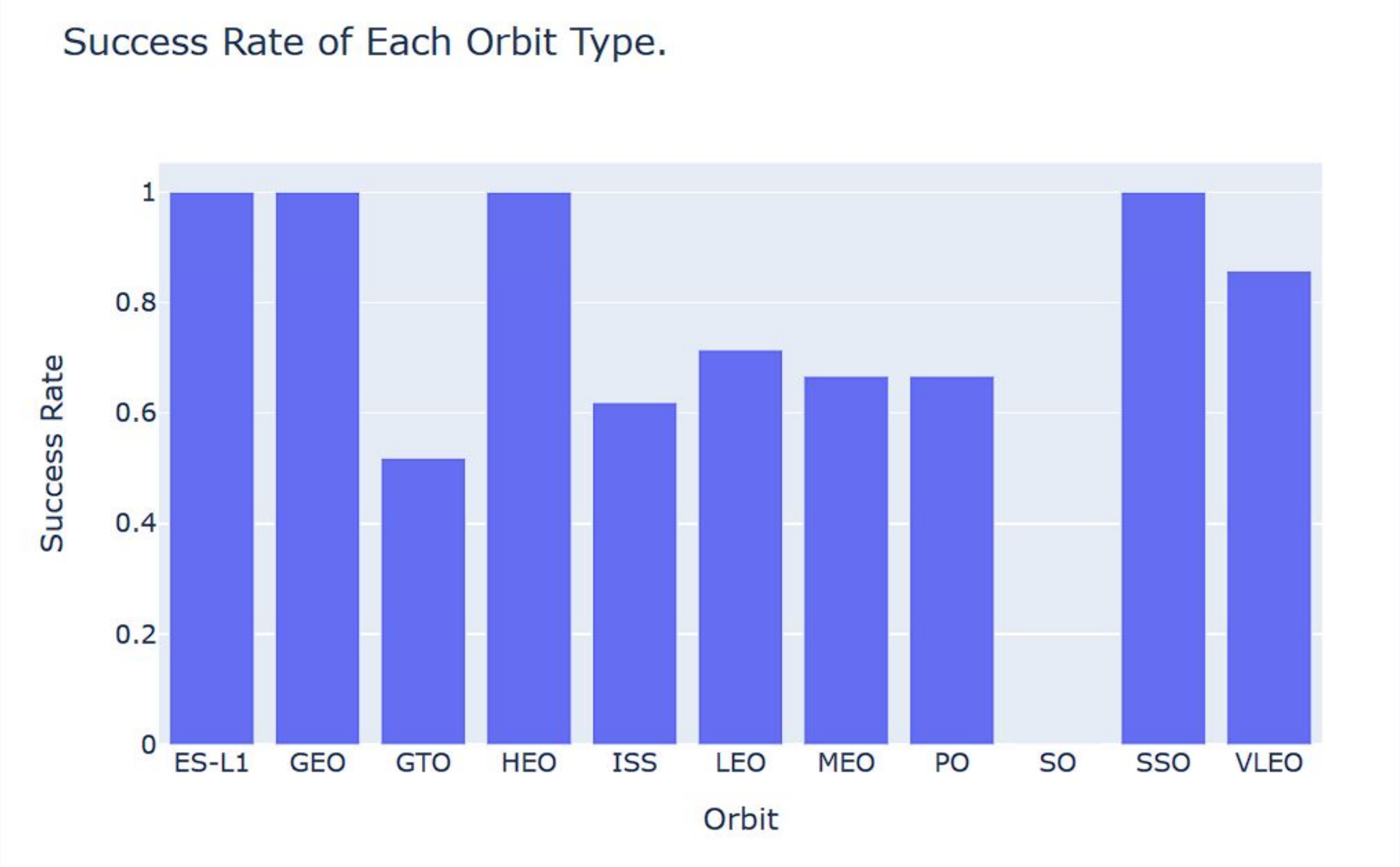
Payload vs. Launch Site

This chart indicate the launch site and payload mass of rocket.



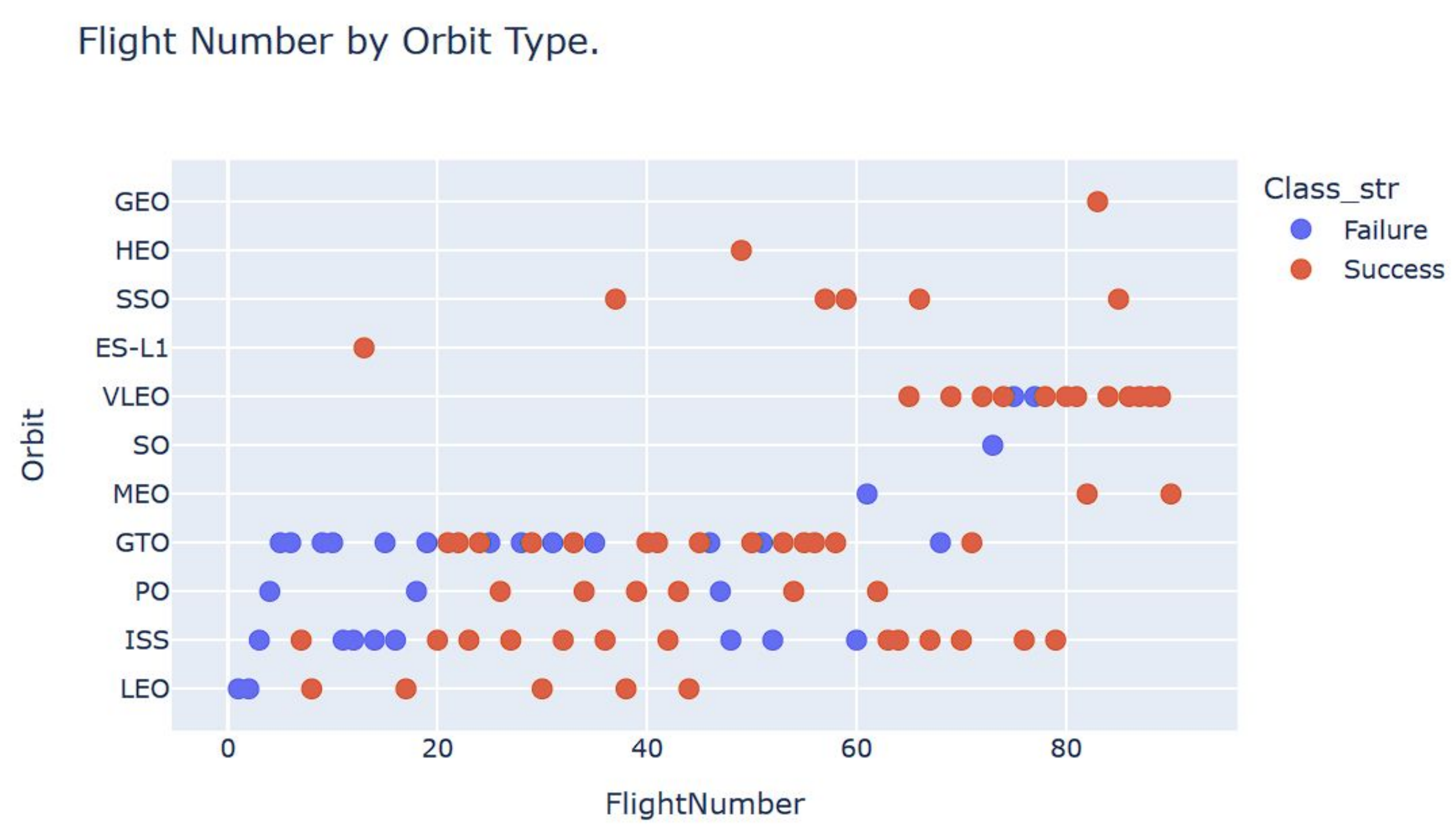
Success Rate vs. Orbit Type

This chart displays the success rates for each orbit type. It is evident that several orbit types exhibit a 100% success rate, while the "SO" orbit type has a success rate of 0%.



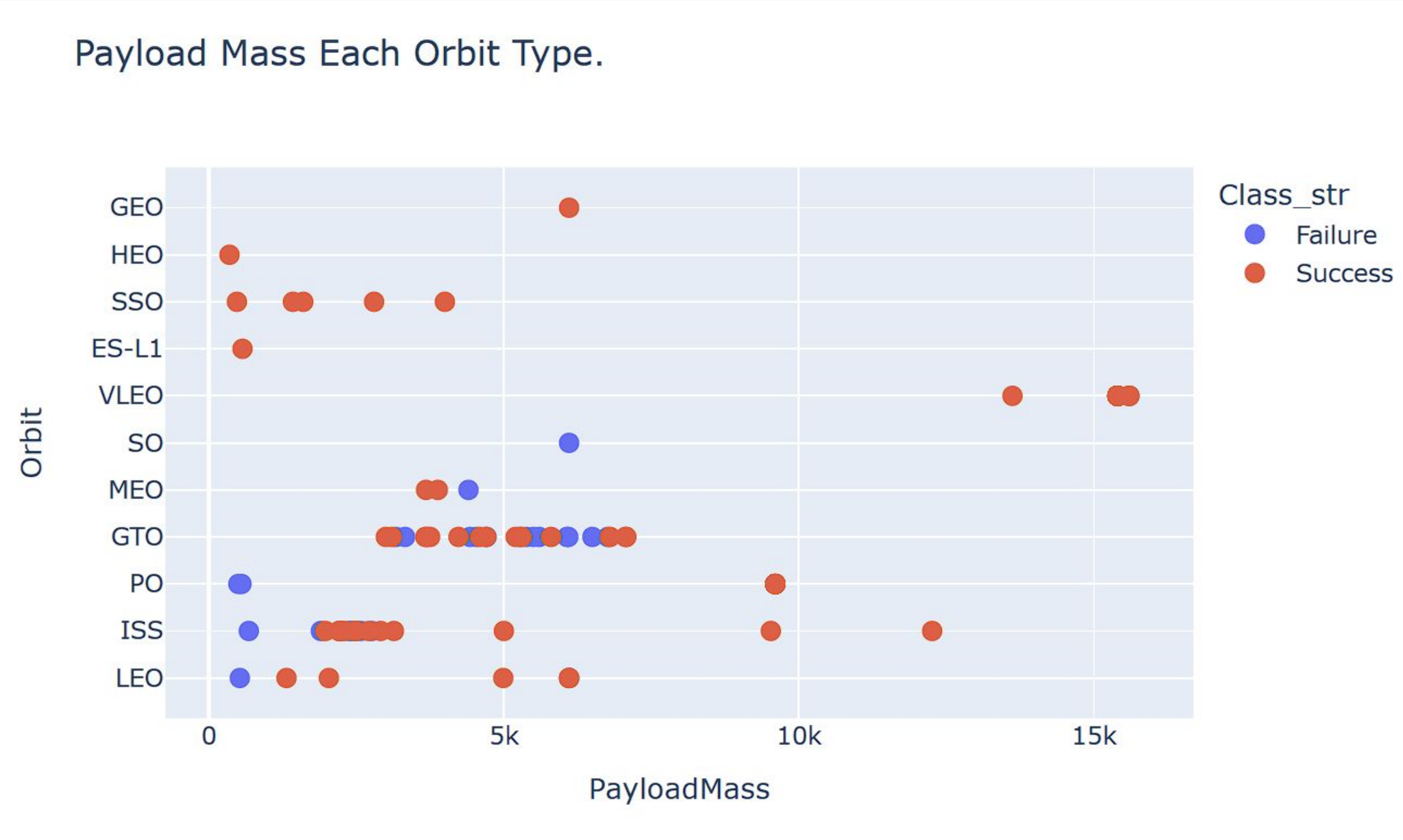
Flight Number vs. Orbit Type

This chart show relation between orbit type and fight number



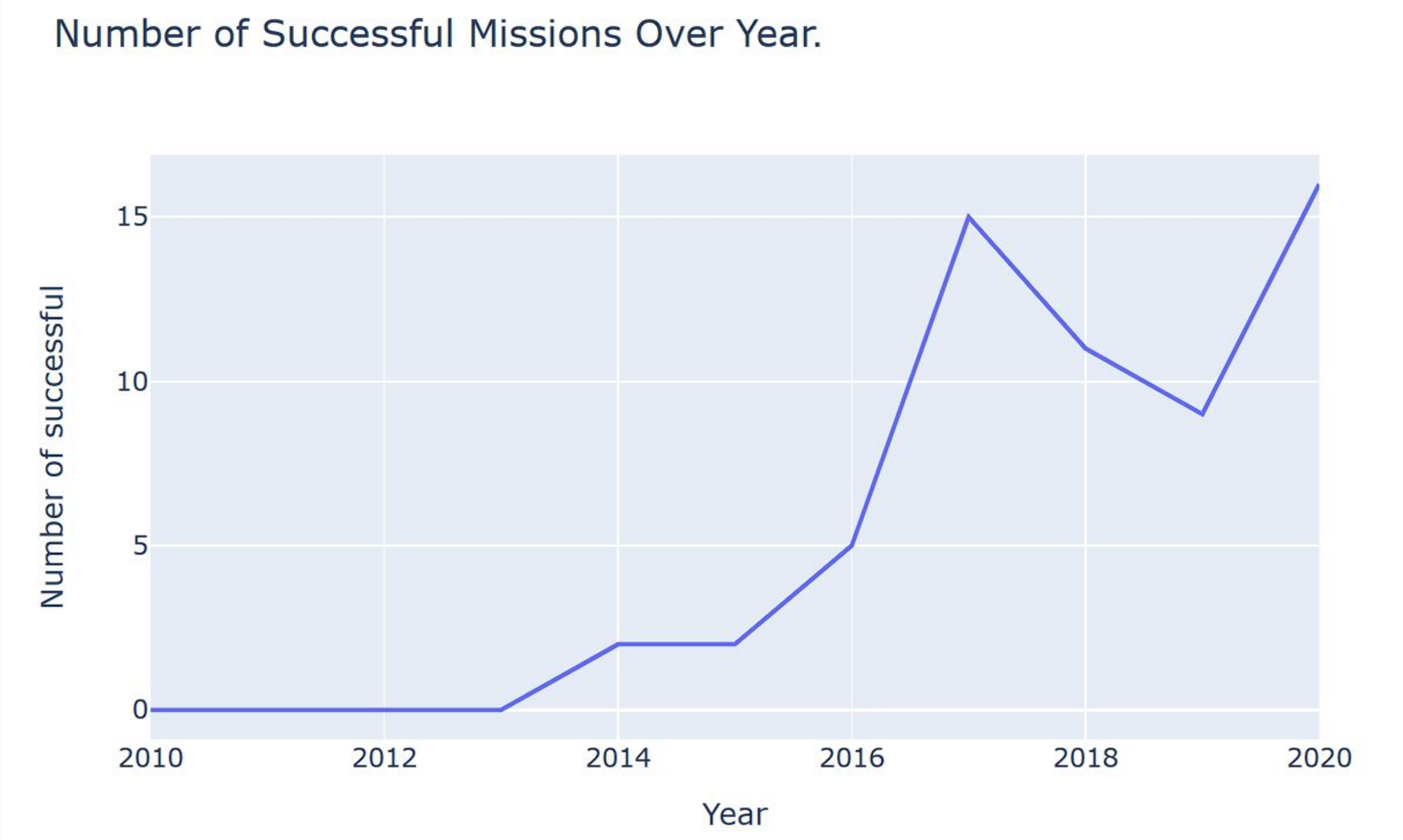
Payload vs. Orbit Type

This chart show payload mass of each orbit



Launch Success Yearly Trend

The chart show the trend of number of success over year is increased.



All Launch Site Names

These are the names of unique launch sites the location from which the launch took place, and there are four distinct launch sites.

	Launch_Site
0	CCAFS LC-40
1	VAFB SLC-4E
2	KSC LC-39A
3	CCAFS SLC-40

Launch Site Names Begin with 'CCA'

The provided data represents information about SpaceX launches. Each row corresponds to a specific launch event, and the columns provide details about various aspects of the launch: Date, Time (UTC): The date and the universal time launch. Booster Version: The version of the Falcon 9 booster used for the launch. Launch Site, Payload: The name or description of the payload carried by the rocket. PAYLOAD_MASS__KG_: The mass of the payload in kilograms. Orbit: The type of orbit achieved by the payload (e.g., Low Earth Orbit - LEO). Customer: The customer or organization for whom the launch was conducted. Mission Outcome: The overall success or failure of the mission. Landing Outcome: The outcome of the rocket's landing attempt, if applicable.

	Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
0	2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
1	2010-12-08	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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
3	2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
4	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

The data represents the total payload mass (in kilograms) for different booster versions used in SpaceX launches.

	Booster_Version	SUM(PAYLOAD_MASS__KG_)
0	F9 B4 B1039.2	2647
1	F9 B4 B1039.1	3310
2	F9 B4 B1045.2	2697
3	F9 B5 B1056.2	2268
4	F9 B5 B1058.4	2972

Average Payload Mass by F9 v1.1

The data represents the average payload mass (in kilograms) for different booster versions used in SpaceX launches.

	Booster_Version	AVG(PAYLOAD_MASS__KG_)
0	F9 B4 B1039.2	2647.0
1	F9 B4 B1040.2	5384.0
2	F9 B4 B1041.2	9600.0
3	F9 B4 B1043.2	6460.0
4	F9 B4 B1039.1	3310.0
5	F9 B4 B1040.1	4990.0
6	F9 B4 B1041.1	9600.0
7	F9 B4 B1042.1	3500.0
8	F9 B4 B1043.1	5000.0
9	F9 B4 B1044	6092.0

First Successful Ground Landing Date

The first successful landing outcome on ground pad is 22 December 2015.

	Date	Landing_Outcome
0	2015-12-22	Success (ground pad)

Successful Drone Ship Landing with Payload between 4000 and 6000

This is the names of boosters which have successfully landed on drone ship and had payload mass between 4000 and 6000.

	Booster_Version	Landing_Outcome
0	F9 FT B1022	Success (drone ship)
1	F9 FT B1026	Success (drone ship)
2	F9 FT B1021.2	Success (drone ship)
3	F9 FT B1031.2	Success (drone ship)

Total Number of Successful and Failure Mission Outcomes

The data summarizes the mission outcomes for SpaceX launches, categorizing them based on different outcomes.

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

Boosters Carried Maximum Payload

The names of the booster which have carried the maximum payload mass

	Booster_Version	Payload_Mass
0	F9 B5 B1048.4	15600
1	F9 B5 B1049.4	15600
2	F9 B5 B1051.3	15600
3	F9 B5 B1056.4	15600
4	F9 B5 B1048.5	15600
...	...	...
96	F9 v1.1 B1003	500
97	F9 FT B1038.1	475
98	F9 B4 B1045.1	362
99	F9 v1.0 B0003	0
100	F9 v1.0 B0004	0

2015 Launch Records

This is the failed landing outcomes in drone ship, their booster versions, and launch site names in year 2015

	month	year	Landing_Outcome	Booster_Version	Launch_Site
0	01	2015	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
1	04	2015	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

This is the rank of the count of landing outcomes between the date 2010-06-04 and 2017-03-20, in descended order

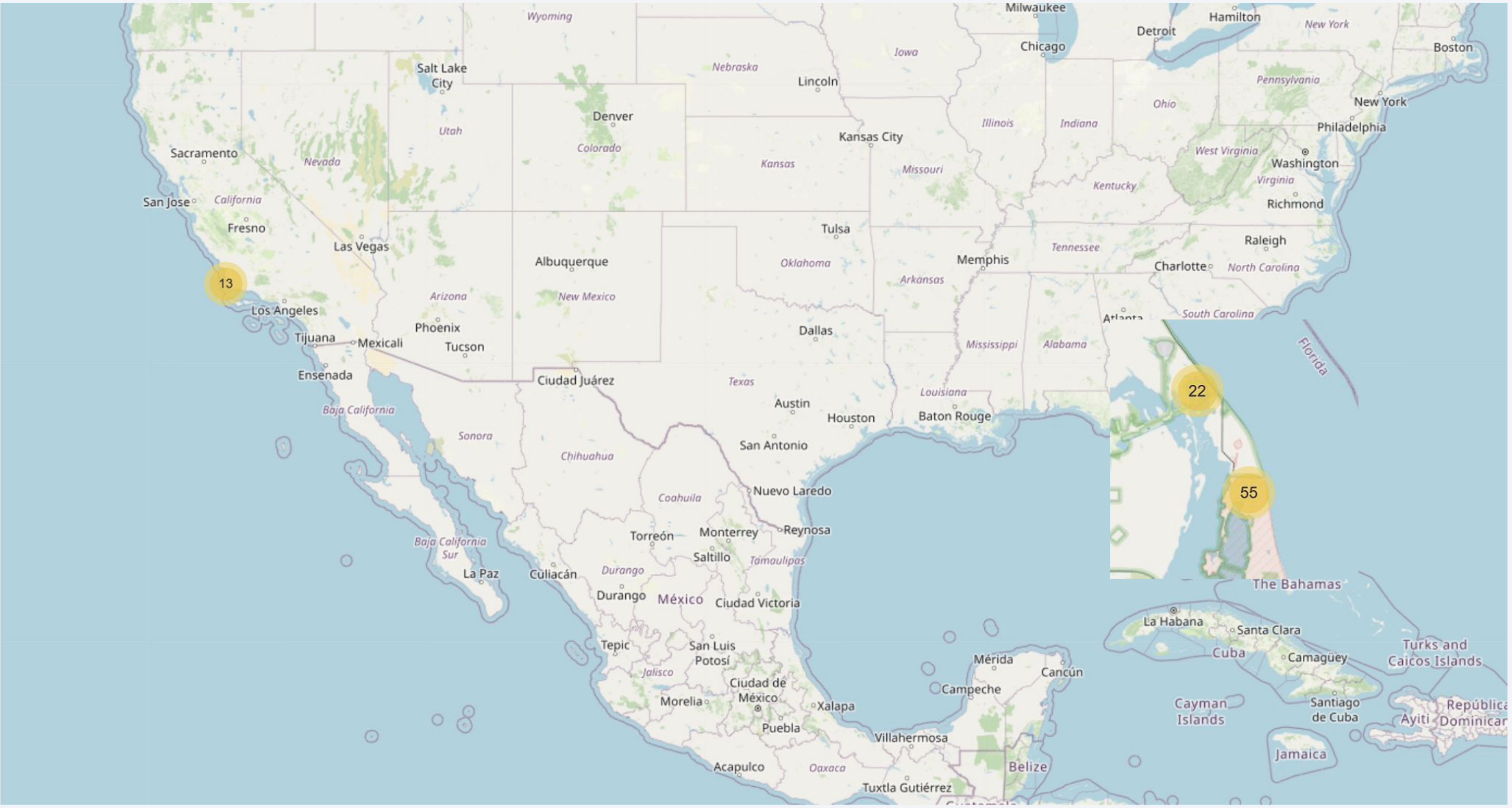
	Landing_Outcome	Outcome_Count
0	No attempt	10
1	Success (drone ship)	5
2	Failure (drone ship)	5
3	Success (ground pad)	3
4	Controlled (ocean)	3
5	Uncontrolled (ocean)	2
6	Failure (parachute)	2
7	Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

Launch site locations

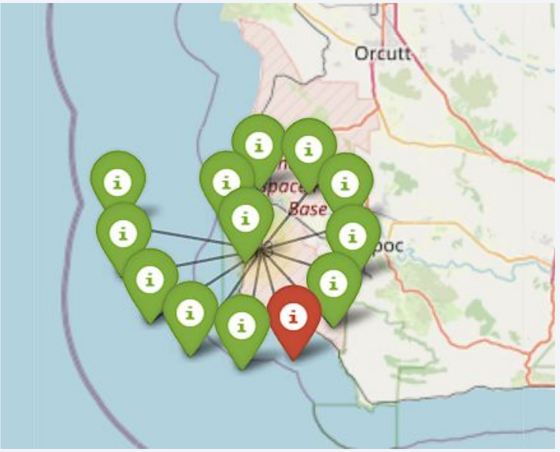
This the locations of the launch site, indicate three launch site. there are one launch at Los angeles and two at the Florida



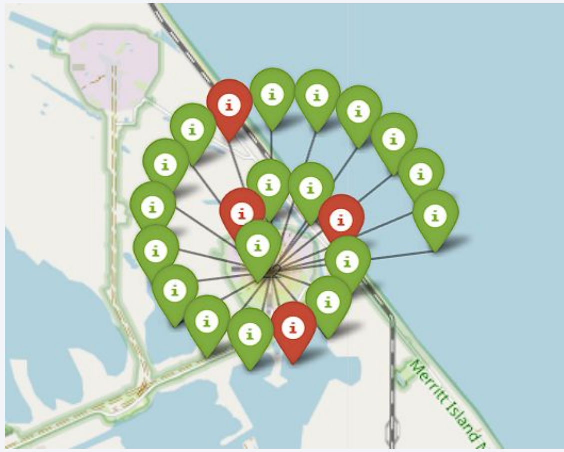
Launch sites success

This map highlights three distinct launch sites, with the color red indicating failed landings and green indicating successful landing.

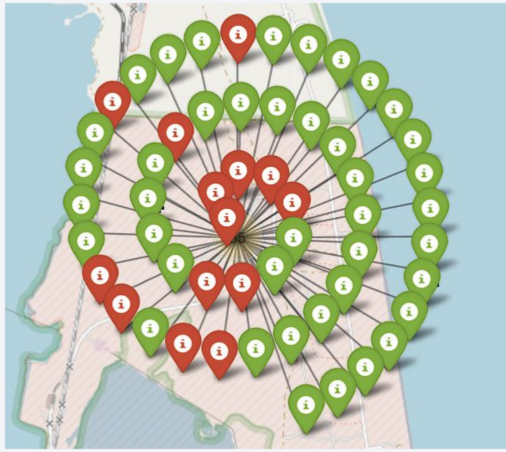
LC-40



LC-39A



SLC-40



<Folium Map Screenshot 3>

- Replace <Folium map screenshot 3> title with an appropriate title

- Explore the generated folium map and show the screenshot of a selected launch site to its proximities such as railway, highway, coastline, with distance calculated and displayed

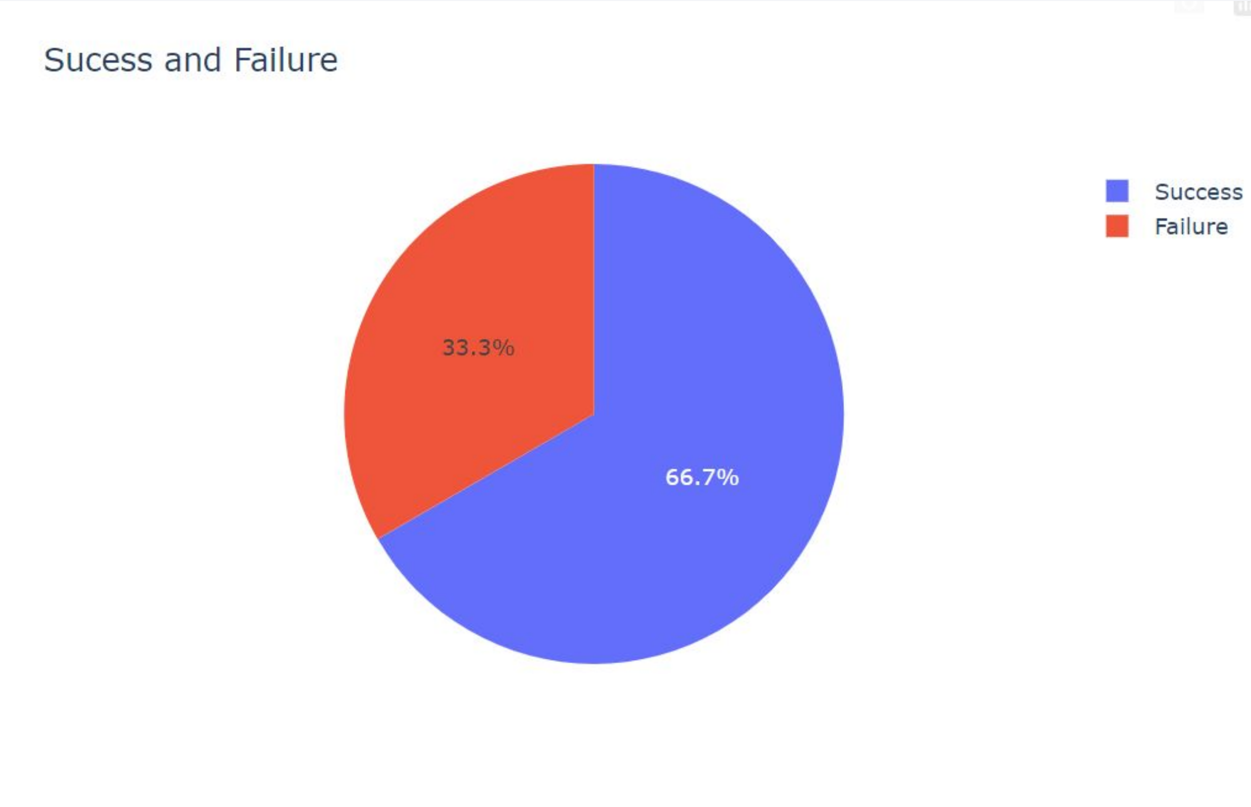
- Explain the important elements and findings on the screenshot

Section 4

Build a Dashboard with Plotly Dash

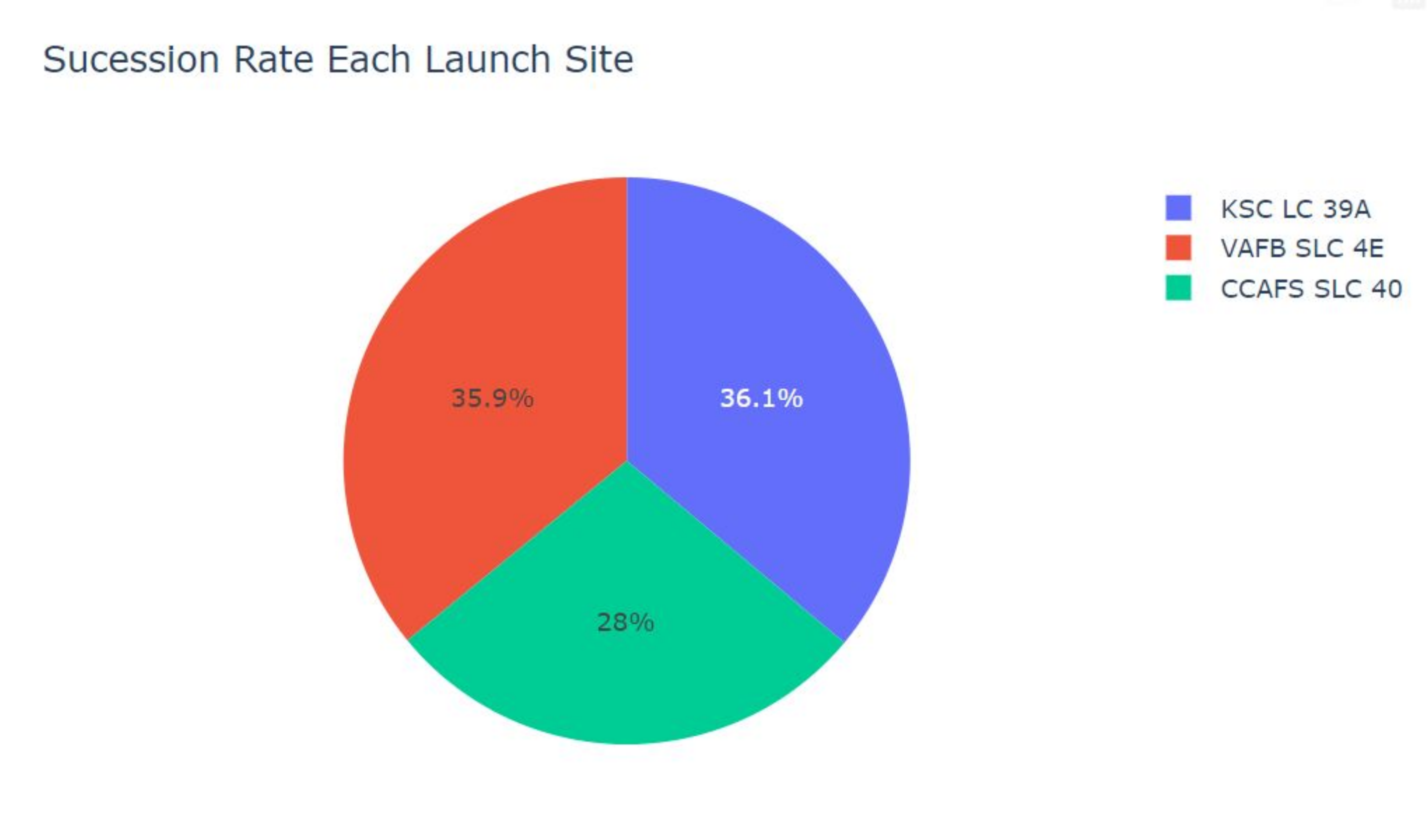
Percent of Succession and Failure

This pie chart indicate percent of mission succession and failure. The probability of success is more than failure.



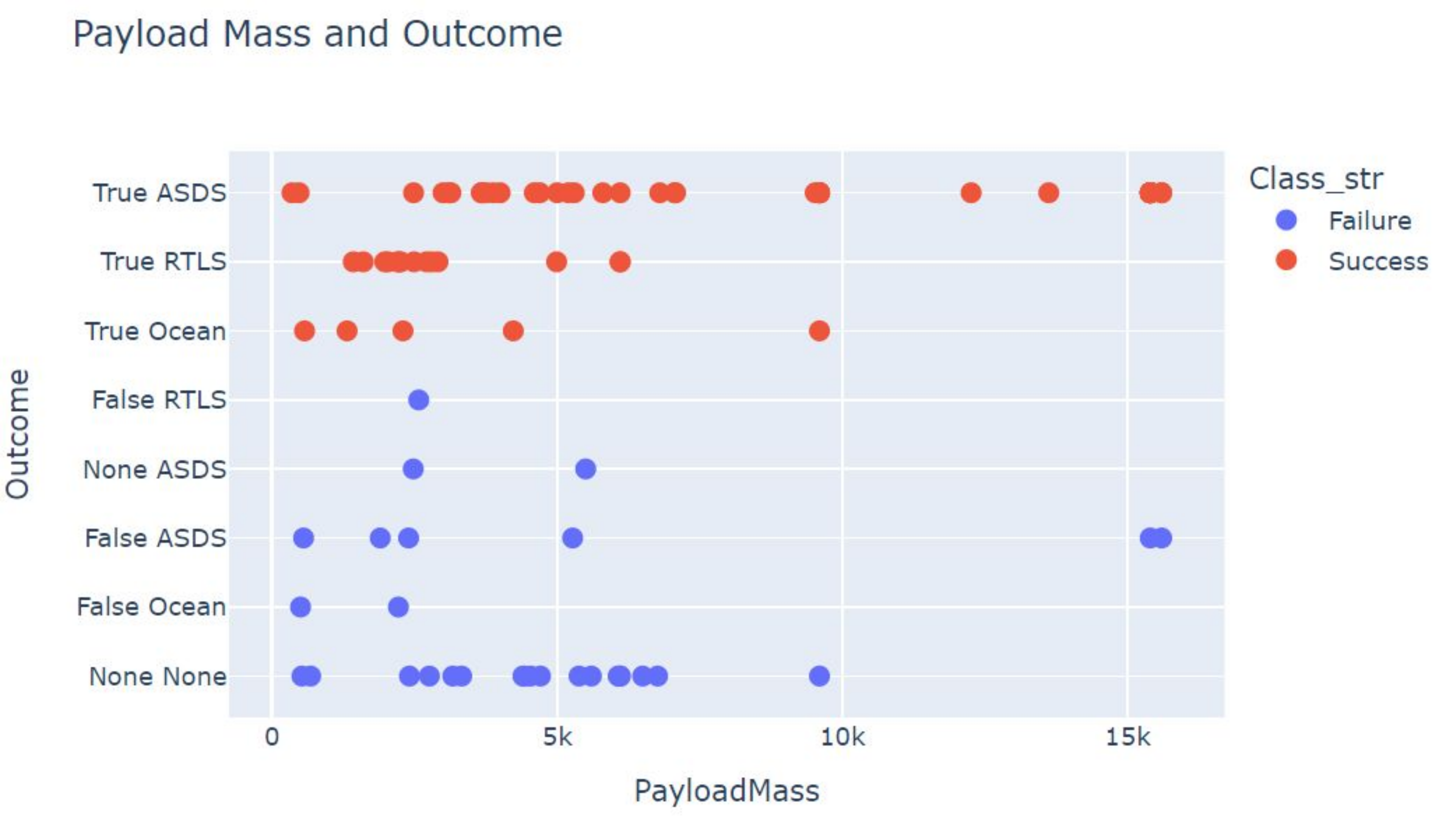
Percent of Succession Each Site

The pie chart show what launch site have more succession rate, The moss success launch site is KSC.



Payload Mass Missions Outcome

This chart show relation between payload mass and mission outcome. red is Fail, and blue is success.

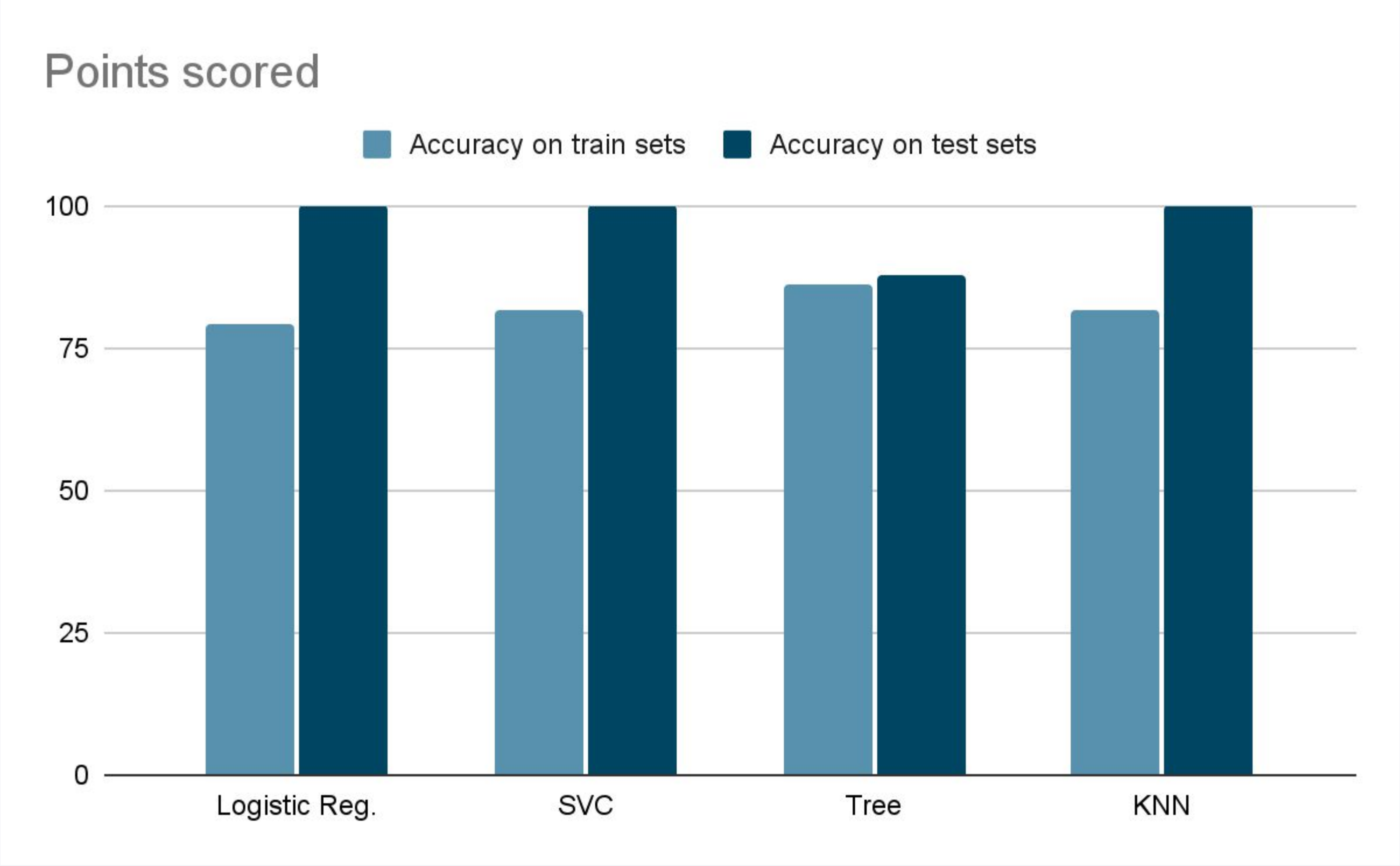


Section 5

Predictive Analysis (Classification)

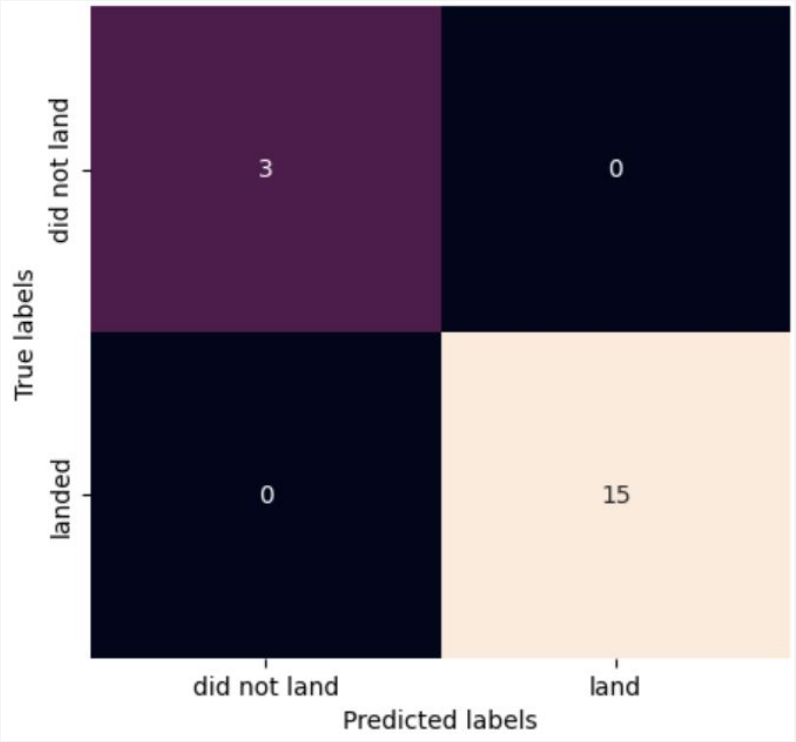
Classification Accuracy

The bar chart show the training and testing accuracy in each predictive models. The most perform model is KNN, and SVC that have most testing accuracy and training accuracy.

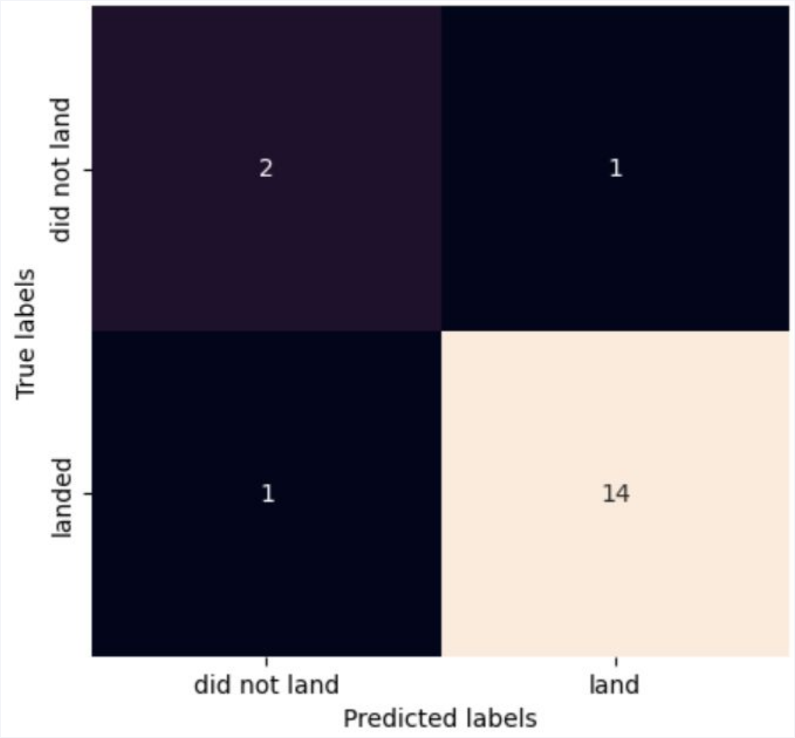


Confusion Matrix

This confusion matrix reveals that the decision tree model made false predictions, unlike the other three models, suggesting a potential issue with imbalanced classes in the dataset.



KNN, SVM, Log Reg.



Decision Tree

Conclusions

- The KNN, and SVC model demonstrated high accuracy in predicting the outcomes
- The decision tree model exhibited false predictions, indicating potential challenges in handling certain patterns within the data.
- The confusion matrix suggests that the dataset might have imbalanced classes, mya impact the performance of the models when to use in real life, especially in negative class.
- Among the models tested, KNN was chosen due to its superior training and evaluation speed. However, further exploration and optimization may be needed to enhance its predictive capabilities

Appendix

- Repository of All Project Files And Dashbord: https://github.com/DylanonWicchiramala/SpaceY-project/

Thank you!