IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Hariyanto
19 Feb 2025

Outline

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

Executive Summary

- We want to predict the outcome of first stage launch mission of SpaceX rockets, since this will help in saving cost.
- Data was collected from SpaceX and Wikipedia. They were combined and explored to get a better sense of the data. The data was cleaned and prepared to be used in developing the predictive models.
- The result shows that we can predict the outcome of launch mission by using Decision Tree Classifier

Introduction

- In this project, we will predict if the Falcon 9 first stage will land successfully.
- SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars
  - Other providers cost upward of 165 million dollars each
  - Much of the savings is because SpaceX can reuse the first stage
- Problems we want to find answers:
  - If we can predict the ffirst stage launch outcome, we can determine the cost of the launch

Section 1

Methodology

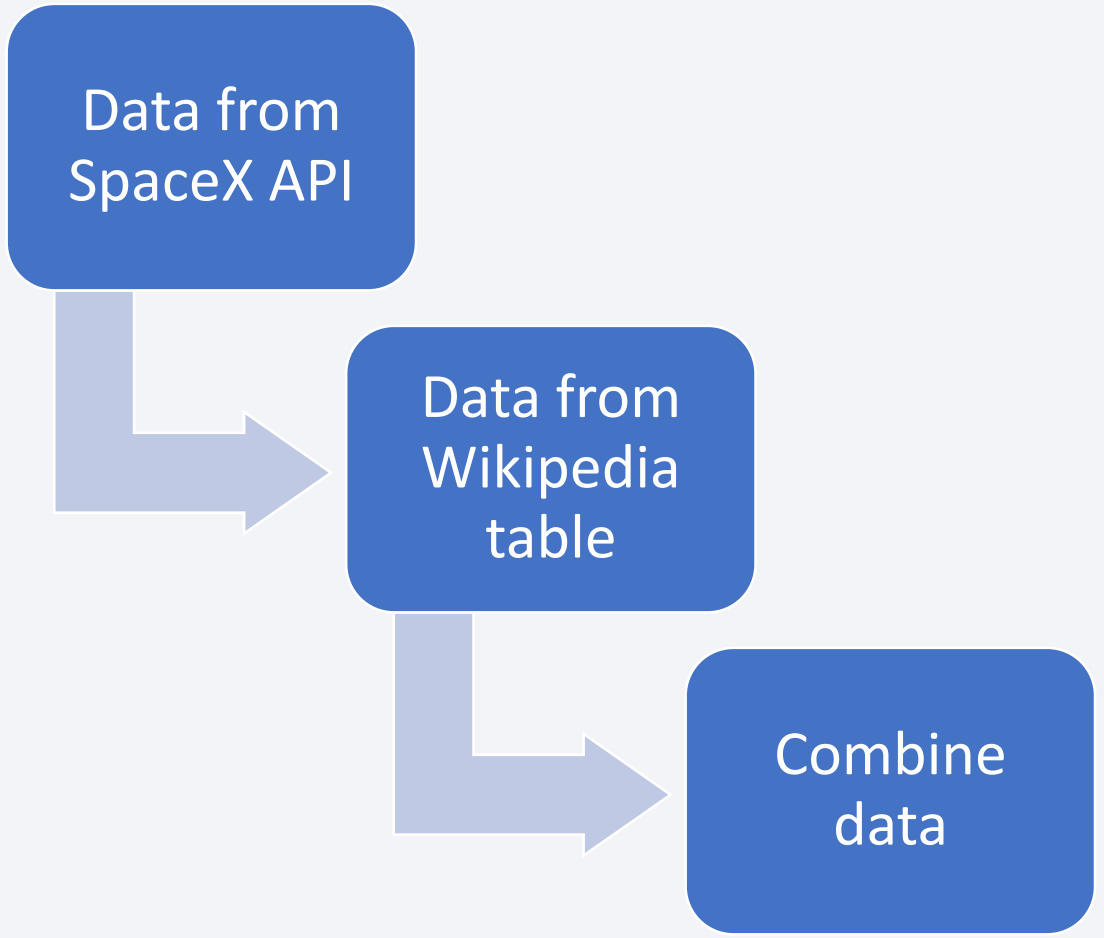
Methodology

Executive Summary

- Data collection methodology:
  - SpaceX API
  - List of Falcon 9 and Falcon Heavy launches (Wikipedia)
- Perform data wrangling
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
  - Create interactive map to visualize the launch site and their proximities
  - Create interactive dashboard to visualize success rate
- Perform predictive analysis using classification models
  - Using 4 model and compare the score to get the best model

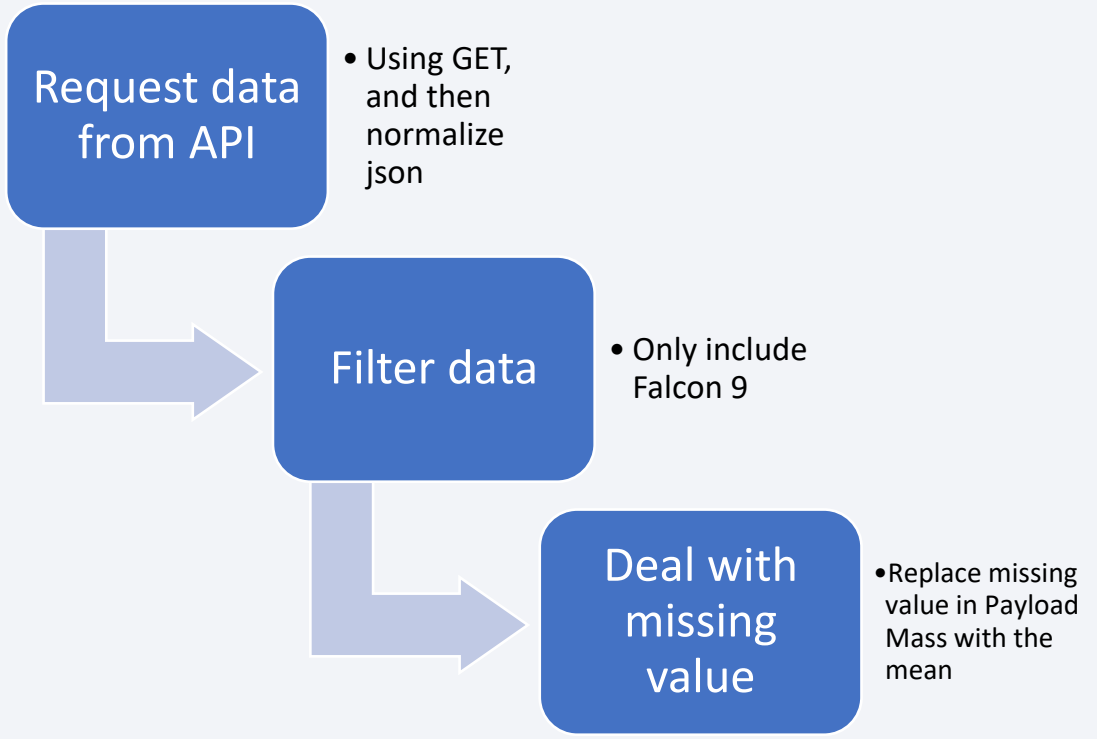
Data Collection

- Data is collected mainly from SpaceX API and Wikipedia.
- API calls were made and the data is combined with the data scraped from a Wikipedia table.



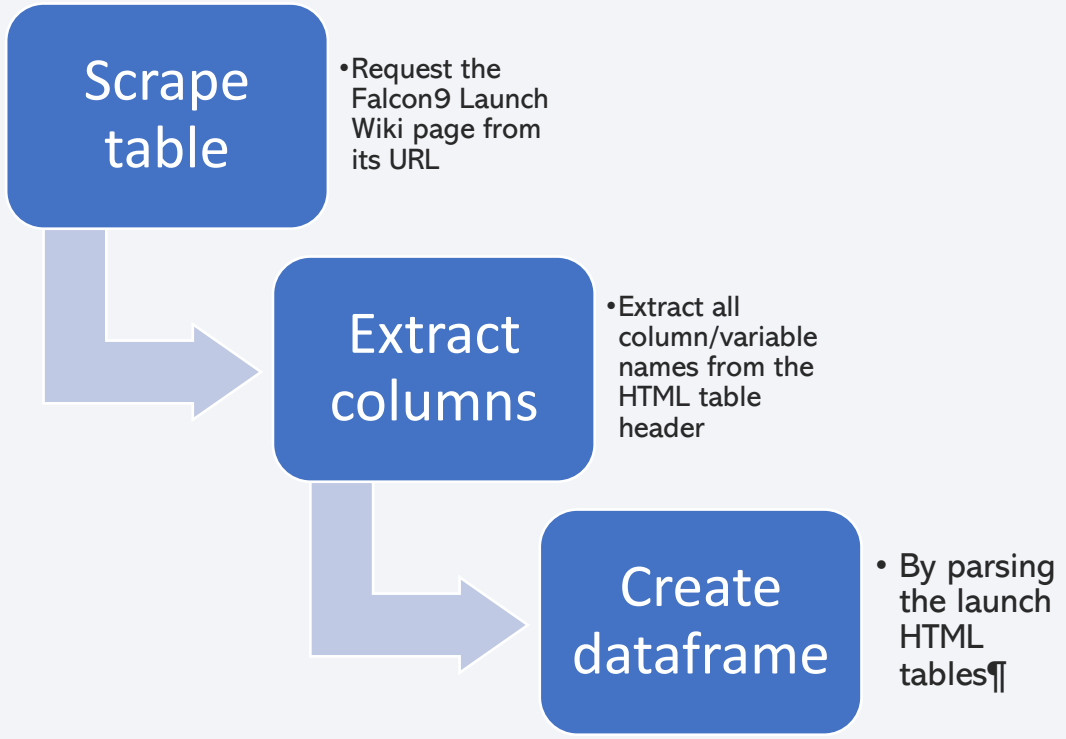
Data Collection – SpaceX API

- Data collection from SpaceX API:
  - Request and parse the SpaceX launch data using the GET request
  - Filter the dataframe to only include Falcon 9 launches
  - Dealing with Missing Values
- Github URL:
  - https://github.com/harichang/ibmdatasciencecapstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

- Data collection from Wikipedia:
  - Request the Falcon9 Launch Wiki page from its URL
  - Extract all column/variable names from the HTML table header
  - Create a data frame by parsing the launch HTML tables
- Github URL:
  - https://github.com/harichang/ibmdatascicencecapstone/blob/main/jupyter-labs-webscraping.ipynb



Data Wrangling

- Some Exploratory Data Analysis (EDA) is performed to find some patterns in the data and determine what would be the label for training supervised models.
- Process:
  - Calculate the number of launches on each site
  - Calculate the number and occurrence of each orbit
  - Calculate the number and occurence of mission outcome of the orbits
  - Create a landing outcome label from Outcome column
- Github URL:
  - https://github.com/harichang/ibmdatasciencecapstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- Scatter/catplot and bar chart was used to visualize the trend and relationships among variables:
  - Visualize the relationship between Flight Number and Launch Site
  - Visualize the relationship between Payload Mass and Launch Site
  - Visualize the relationship between success rate of each orbit type
  - Visualize the relationship between FlightNumber and Orbit type
  - Visualize the relationship between Payload Mass and Orbit type
  - Visualize the launch success yearly trend
- Github:
  - https://github.com/harichang/ibmdatasciencecapstone/blob/main/edadataviz.ipynb

EDA with SQL

- These SQL queries were performed:
  - Display the names of the unique launch sites in the space mission
  - Display 5 records where launch sites begin with the string 'CCA'
  - Display the total payload mass carried by boosters launched by NASA (CRS)
  - Display average payload mass carried by booster version F9 v1.1
  - List the date when the first succesful landing outcome in ground pad was acheived.
  - List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
  - List the total number of successful and failure mission outcomes
  - List the names of the booster_versions which have carried the maximum payload mass. Use a subquery
  - 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.
  - 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.
- Github:
  - https://github.com/harichang/ibmdatasciencecapstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

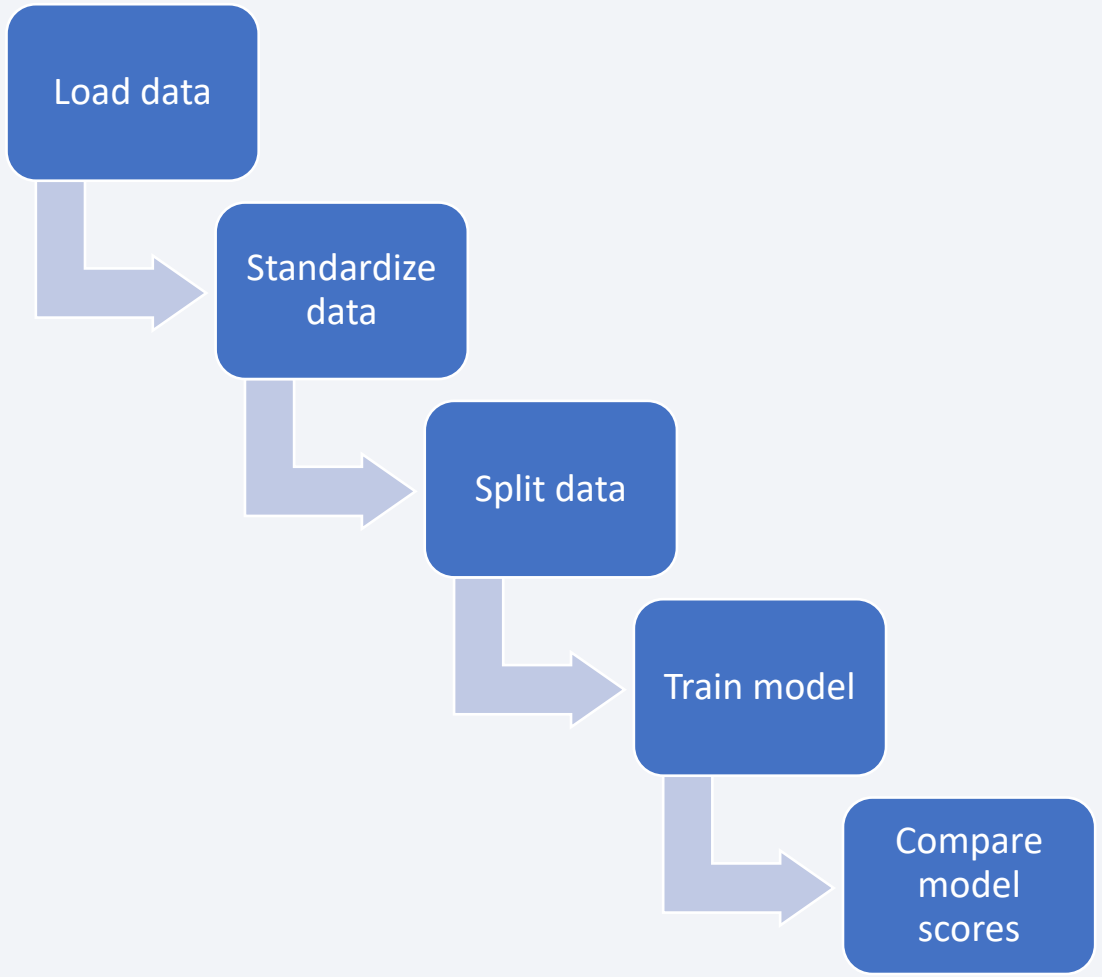
- Folium map built to visualize the location and proximities:
  - Mark all launch sites on a map
  - Mark the success/failed launches for each site on the map
  - Calculate the distances between a launch site to its proximities:
    - Coastline
    - Railway
    - Highway
    - City
- Some proximities were added to get a better sense of the surrounding of the launch site.
- Github:
  - https://github.com/harichang/ibmdatasciencecapstone/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- The Plotly-Dash dashboard were created:
  - Pie chart to show the total successful launches count for all sites or for each site selected
  - Scatter chart to show the correlation between payload and launch success for all sites or for each site
  - Dropdown list to select Launch Site
  - Range Slider to select Payload range
- The dropdown and range slider is necessary so the user can choose/filter the the data for the charts.
- Github:
  - https://github.com/harichang/ibmdatasciencecapstone/blob/main/spacex_dash_app.py

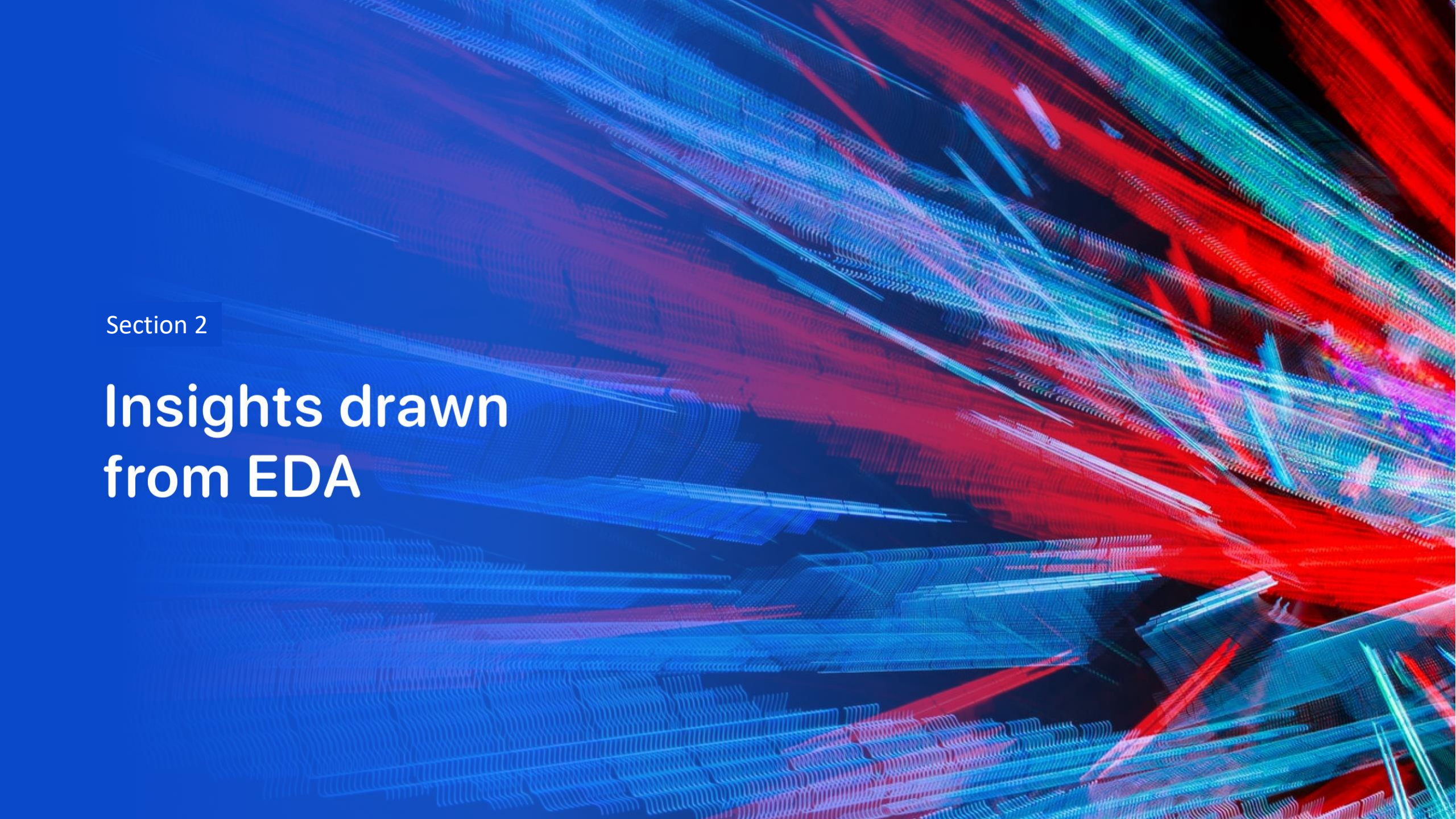
Predictive Analysis (Classification)

- Steps in the predictive analysis:
  - Load data into X (features) and Y (target)
  - Standarize the data
  - Split the data with 20% test data
  - Train the model using Grid Search on four methods:
    - Logistic Regression
    - SVM
    - Decision Tree Classifier
    - KNN
  - Compare the accuracy score on training and test data to determine the best model
- Github:
  - https://github.com/harichang/ibmdatascience capstone/blob/main/SpaceX_Machine%20Learnin g%20Prediction_Part_5.ipynb

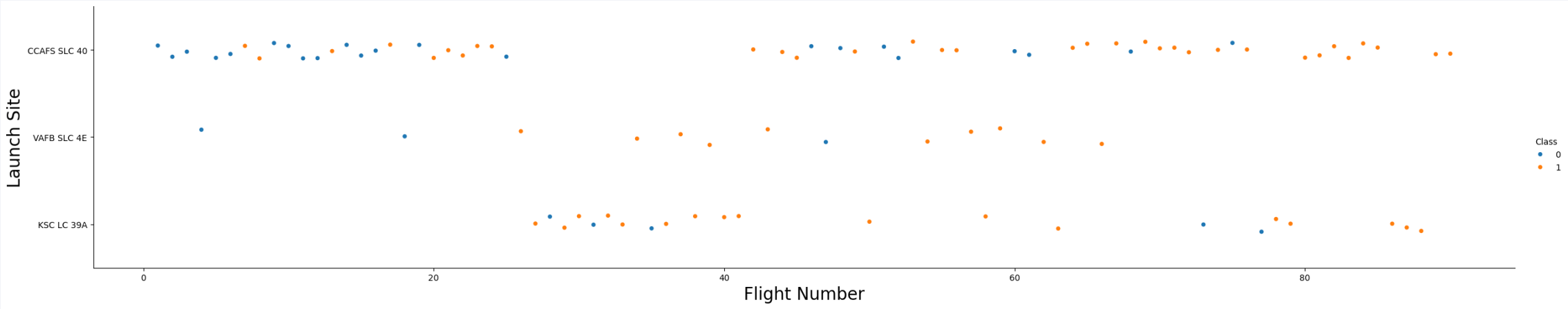


Results

- Exploratory data analysis results
  - From the exploratory analysis, we can see the relationship among variables.
  - But we cannot conclude the main factor affecting the mission outcomes.
  - Predictive model were needed to predict mission outcomes.
- Predictive analysis results:
  - Four models were trained: Logistic Regression, SVM, Decision Tree Classifier, and KNN
  - The best predictive model is Decision Tree Classifier.

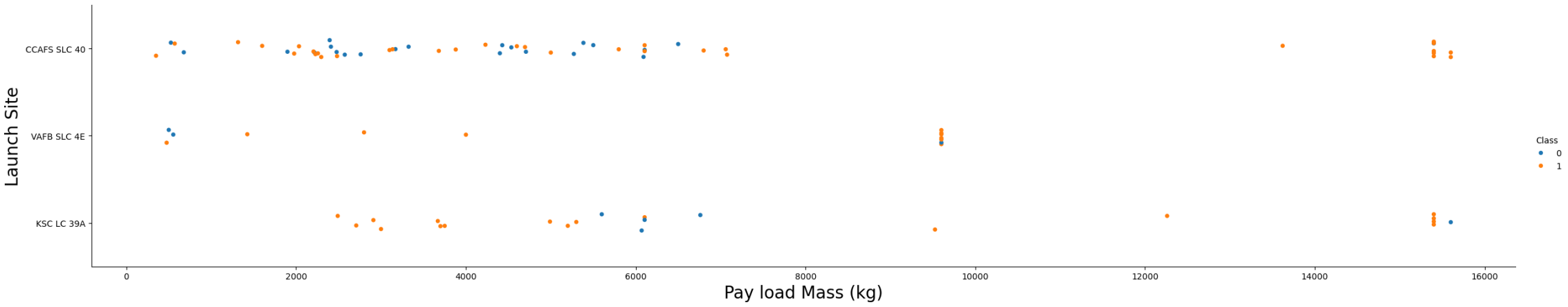

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



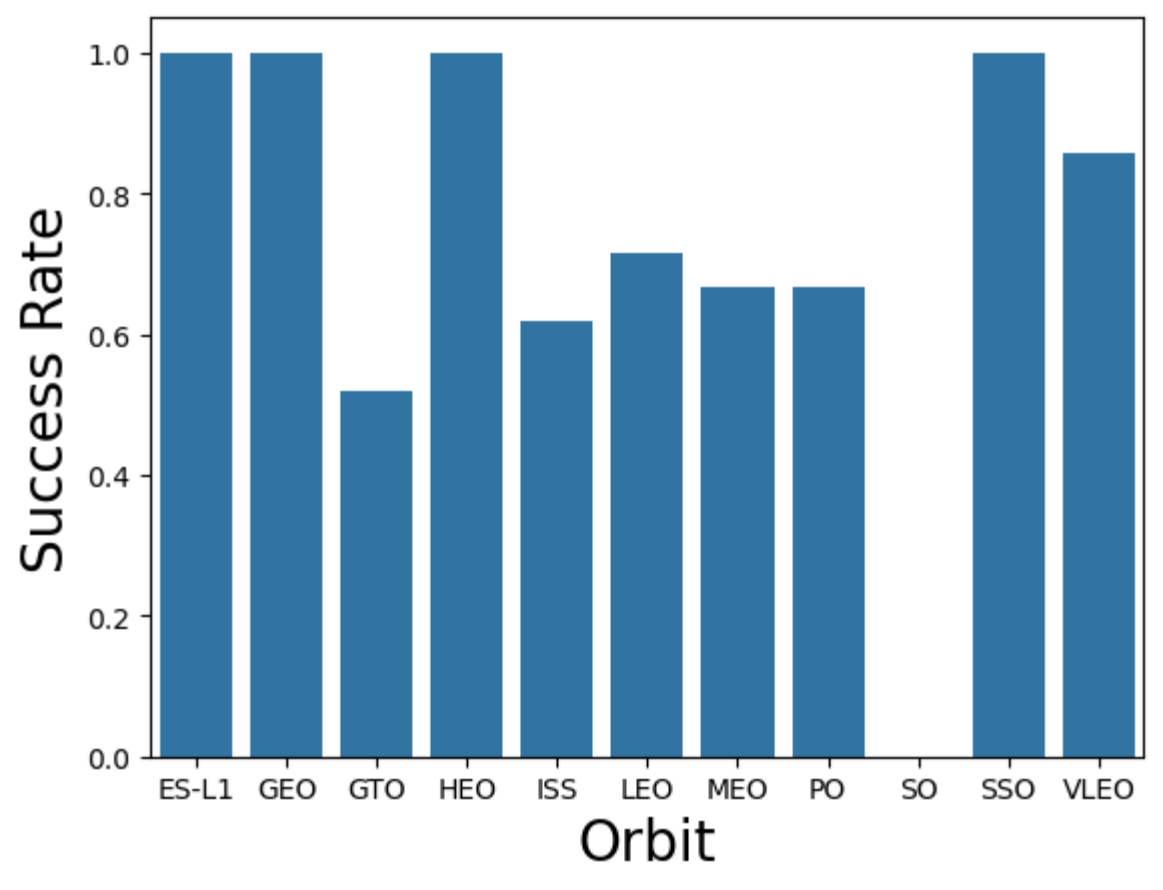
The scatter plot shows that there is no correlation between launch site and flight number.
It shows that higher flight numbers have more successful landing.

Payload vs. Launch Site



The scatter plot shows that there is no correlation between launch site and payload mass
It shows that higher payload mass might have more successful landing, but need further analysis.

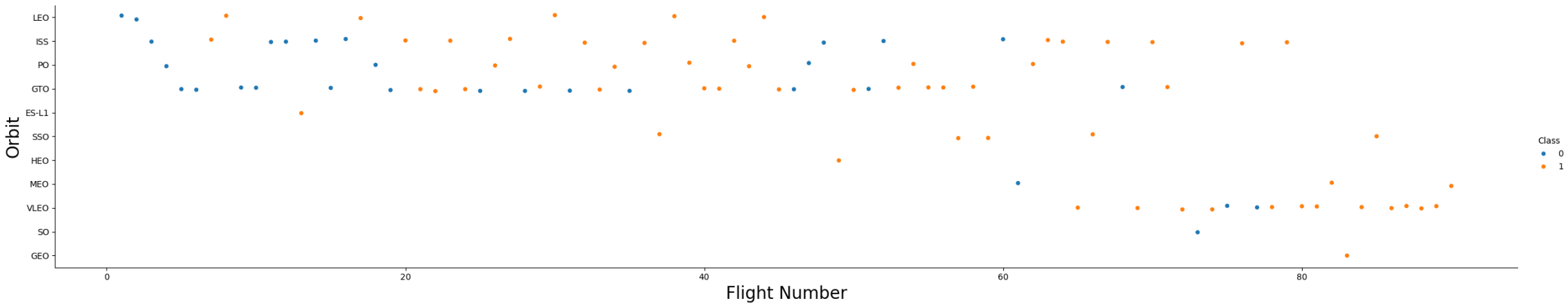
Success Rate vs. Orbit Type



Each orbit type have different success rate:

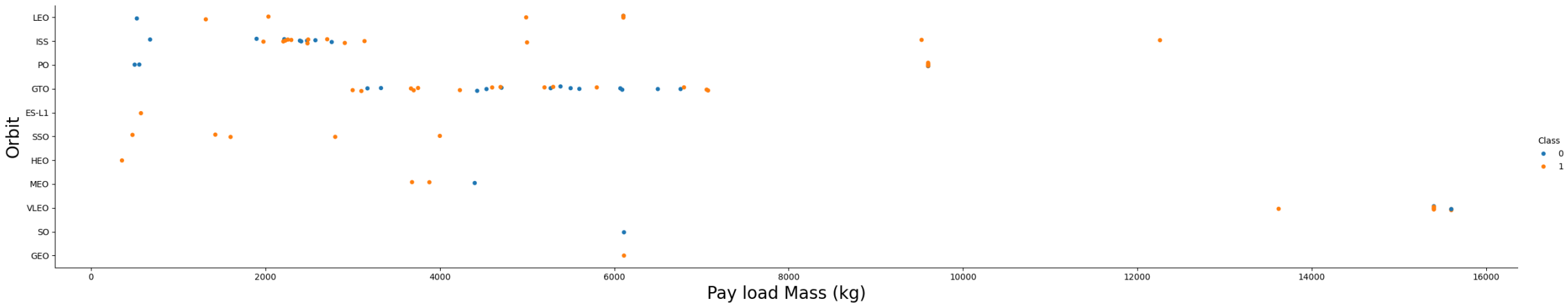
- Orbit type ES-L1, GEO, HEO, and SSO have 100% success rate.
- Other orbit types have around 60% success rate.
- GTO orbit type has 50% success rate
- SO orbit type has 0% success rate

Flight Number vs. Orbit Type



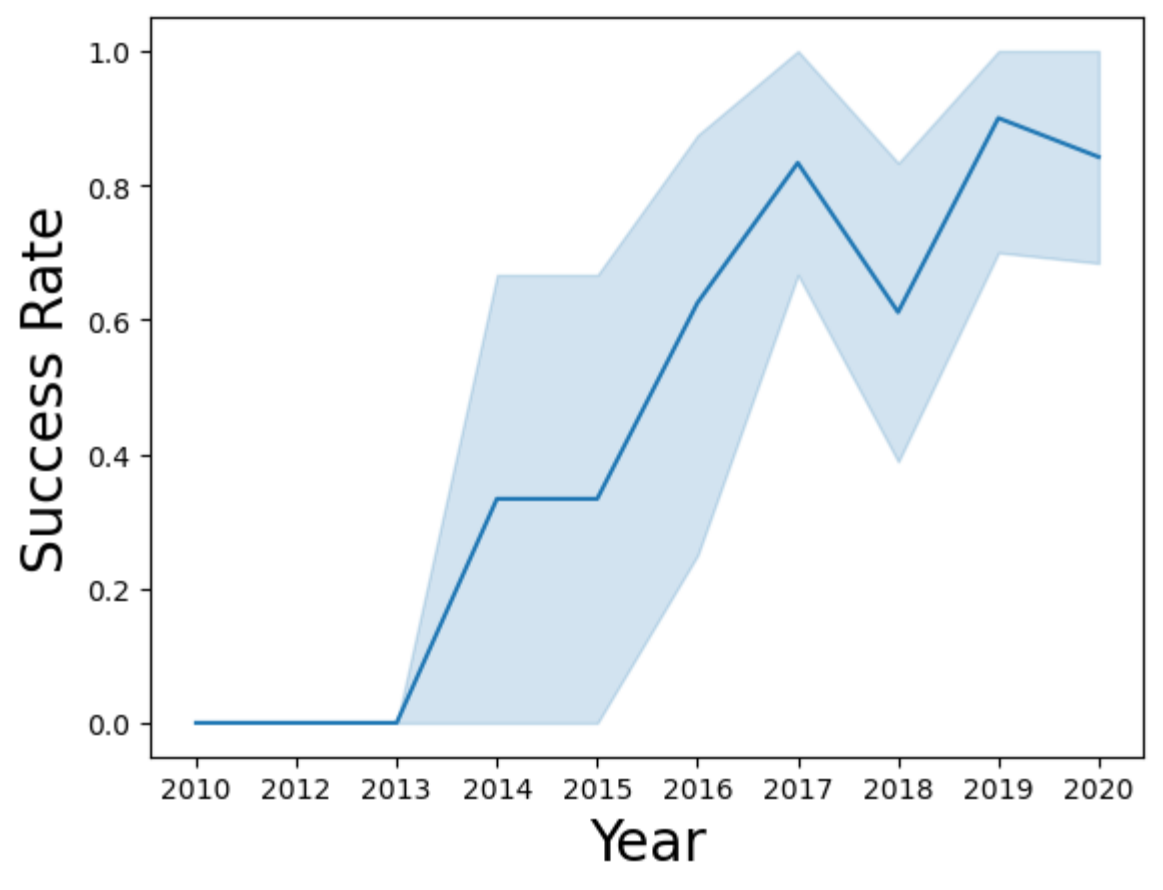
The scatter plot shows that certain orbit types have higher flight number.
It also shows that higher flight number might have more successful landing.

Payload vs. Orbit Type



The scatter plot shows that certain orbit types handles higher payload mass.
It shows that most payload mass are around 2000 an 7000 kg.

Launch Success Yearly Trend



From 2013, success rate kept increasing.

Since then, the lowest success rate was in 2018.
And the highest success rate was in 2019.

All Launch Site Names

- The names of the unique launch sites:

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

- There are only 4 unique launch site in the data

Launch Site Names Begin with 'CCA'

- 5 records where launch sites begin with `CCA`:

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

- The first lauch from site ‘CCAFS LS-40’ is from year 2010

Total Payload Mass

- The total payload carried by boosters from NASA:

SUM(PAYLOAD_MASS__KG_)
45596

- Total Payload mass for customer NASA is 45,596 kg.

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1:

AVG(PAYLOAD_MASS__KG_)
2928.4

- Average payload mass carried by booster version F9 v1.1 is 2928.4 kg

First Successful Ground Landing Date

- The date of the first successful landing outcome on ground pad:

MIN(Date)
2015-12-22

- The first successful landing on ground pad is on 22 Dec 2015

Successful Drone Ship Landing with Payload between 4000 and 6000

- Names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000:

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

- There are only 4 booster version with payload between 4000 and 600 that had successfully landed on drone ship.

Total Number of Successful and Failure Mission Outcomes

- Total number of successful and failure mission outcomes:

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

- Most mission are successful with only one mission had failure as outcome.

Boosters Carried Maximum Payload

- Names of the booster which have carried the maximum payload mass:

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

- There are 12 booster version that carried the maximum payload mass

2015 Launch Records

- Failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015:

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

- There are two failed landing outcome in drone ship in year 2015

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

- Rank of count of landing outcomes between the date 2010-06-04 and 2017-03-20, in descending order:

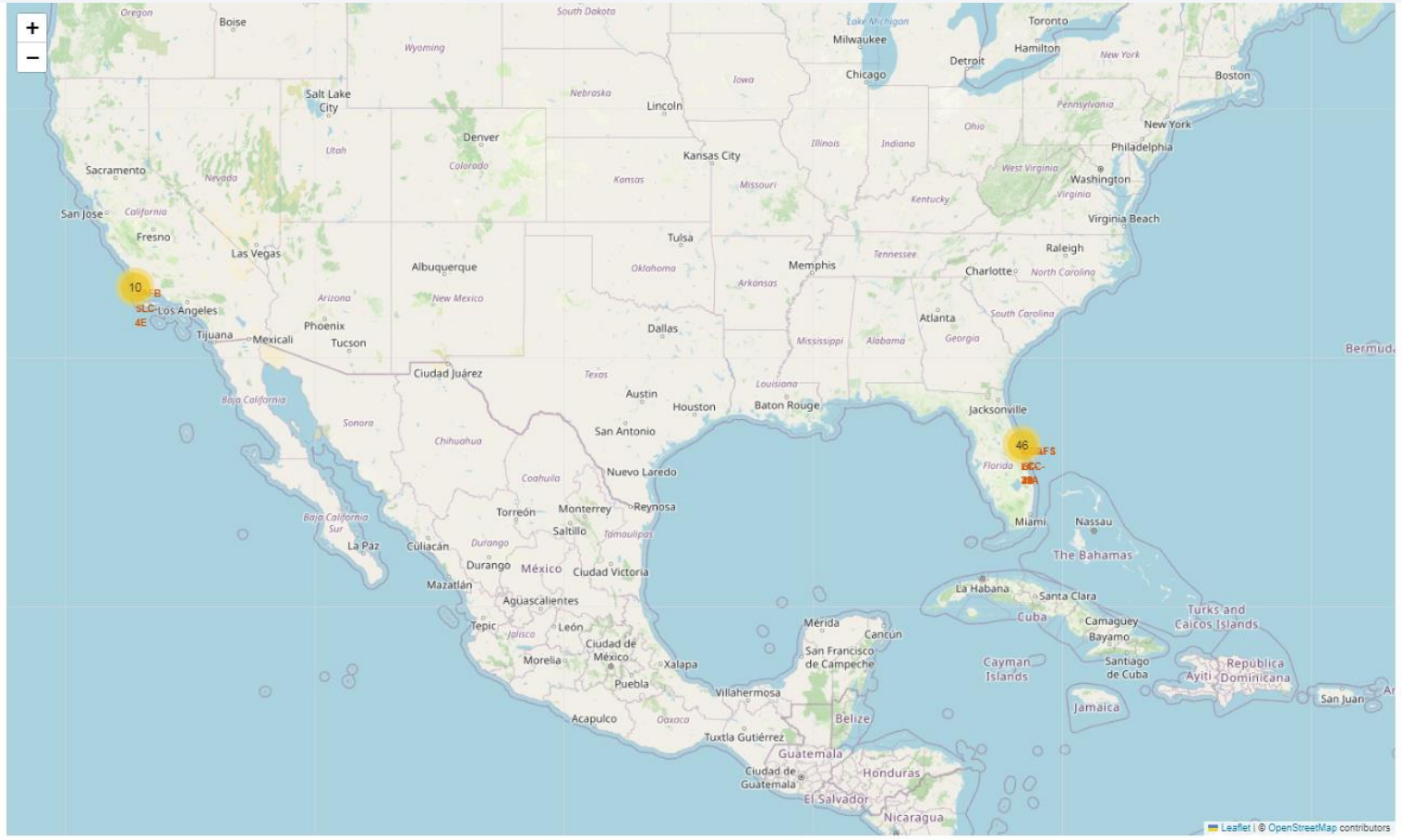
Landing_Outcome	COUNT(*)
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

- Ten missions between 2010-06-04 and 2017-03-20 had no landing attempt.
- Ten mission had drone ship landing attempt, with only half were successful

Section 3

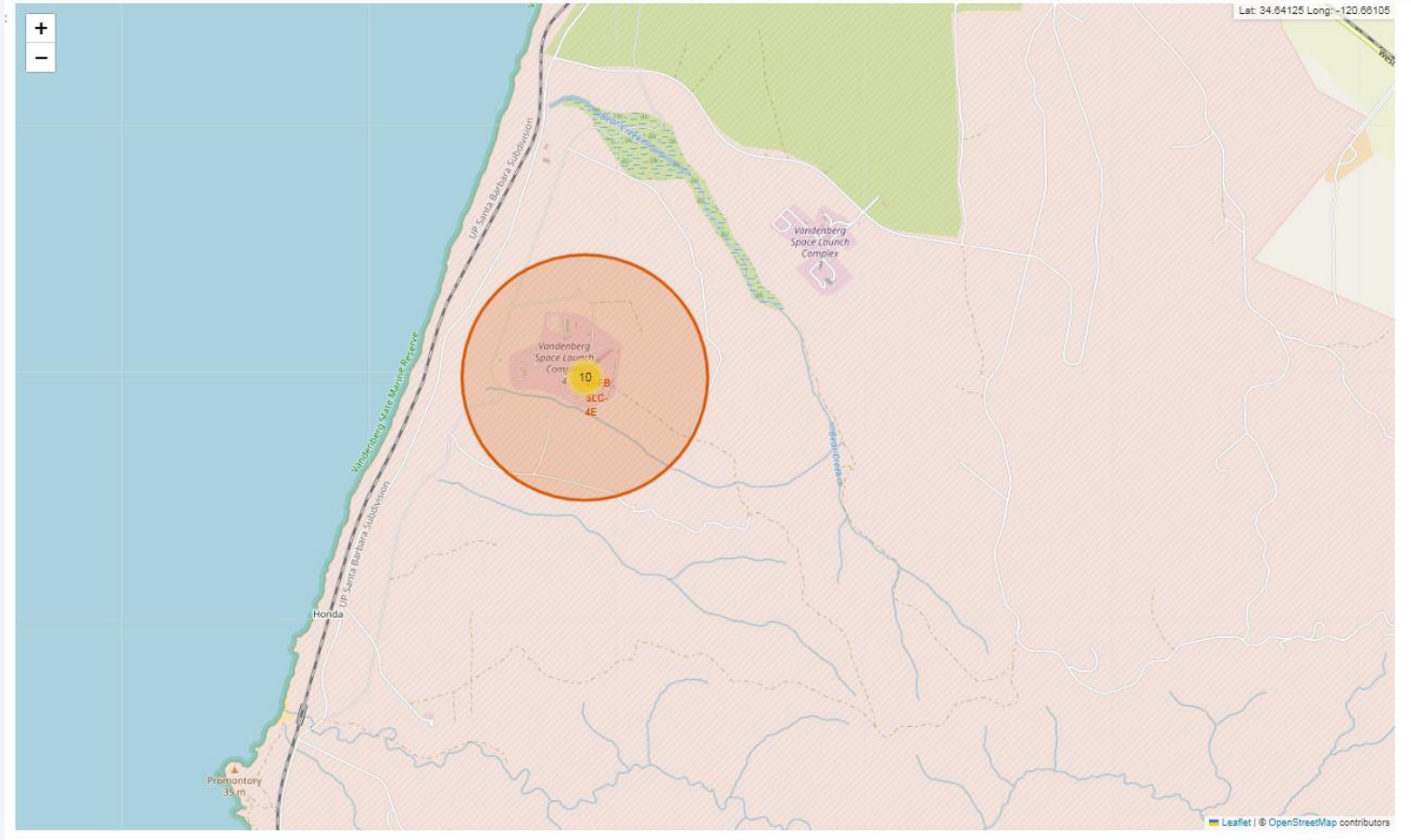
Launch Sites Proximities Analysis

Map of Lauch Sites



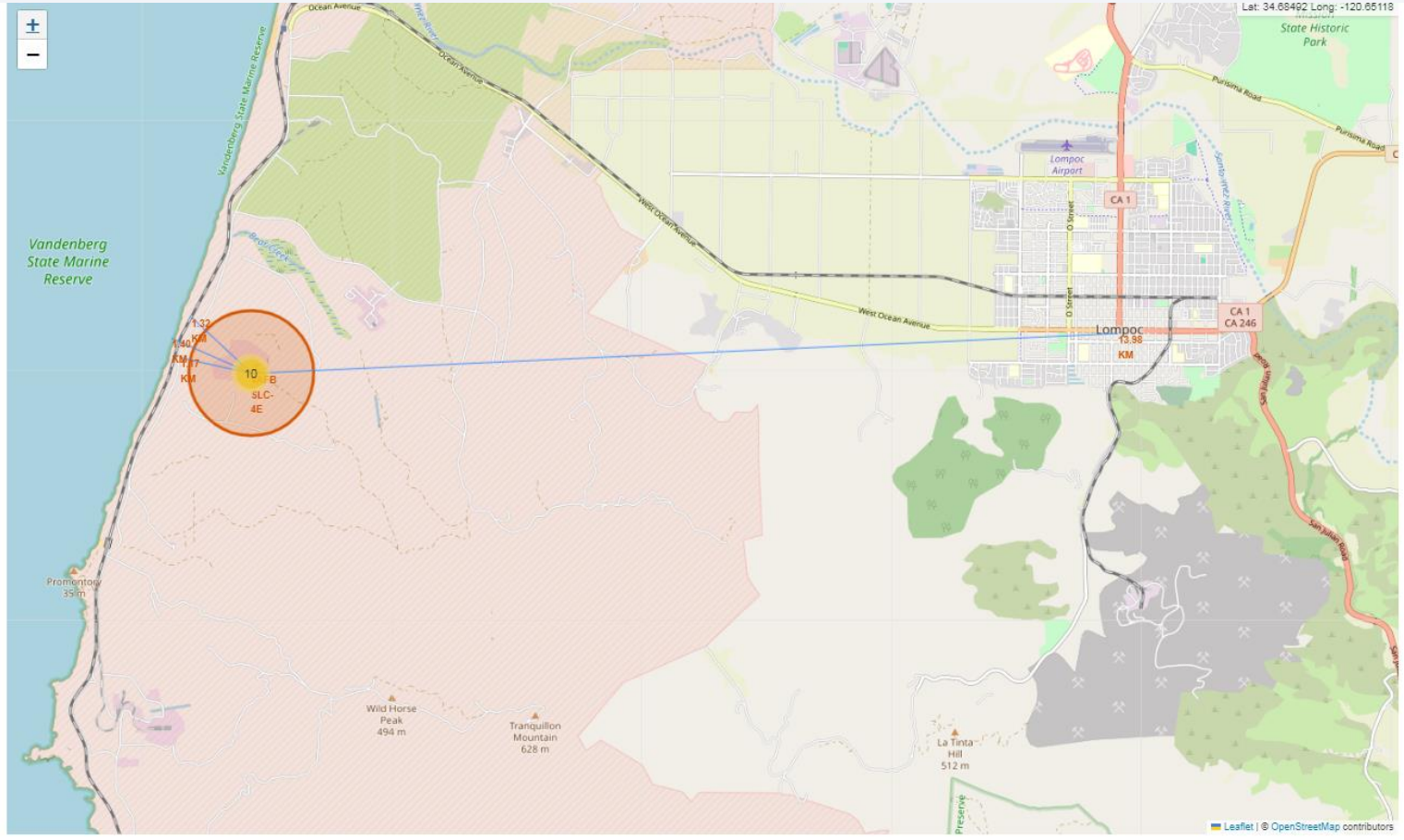
- The map shows all launch site locations.
- Most launches were from the east cost.

Map of Launch Outcomes

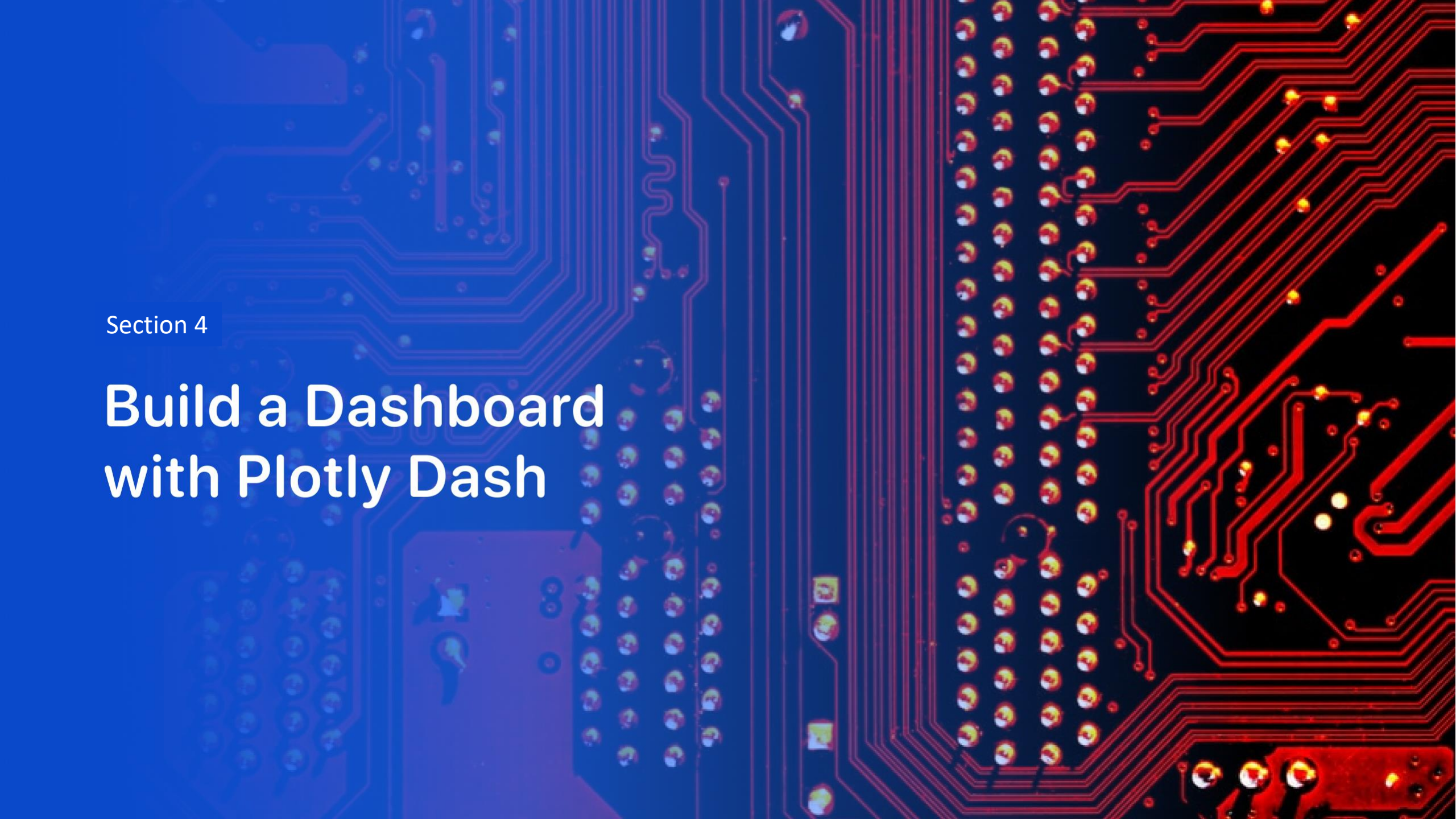


- The map shows launch site locations on west coast.

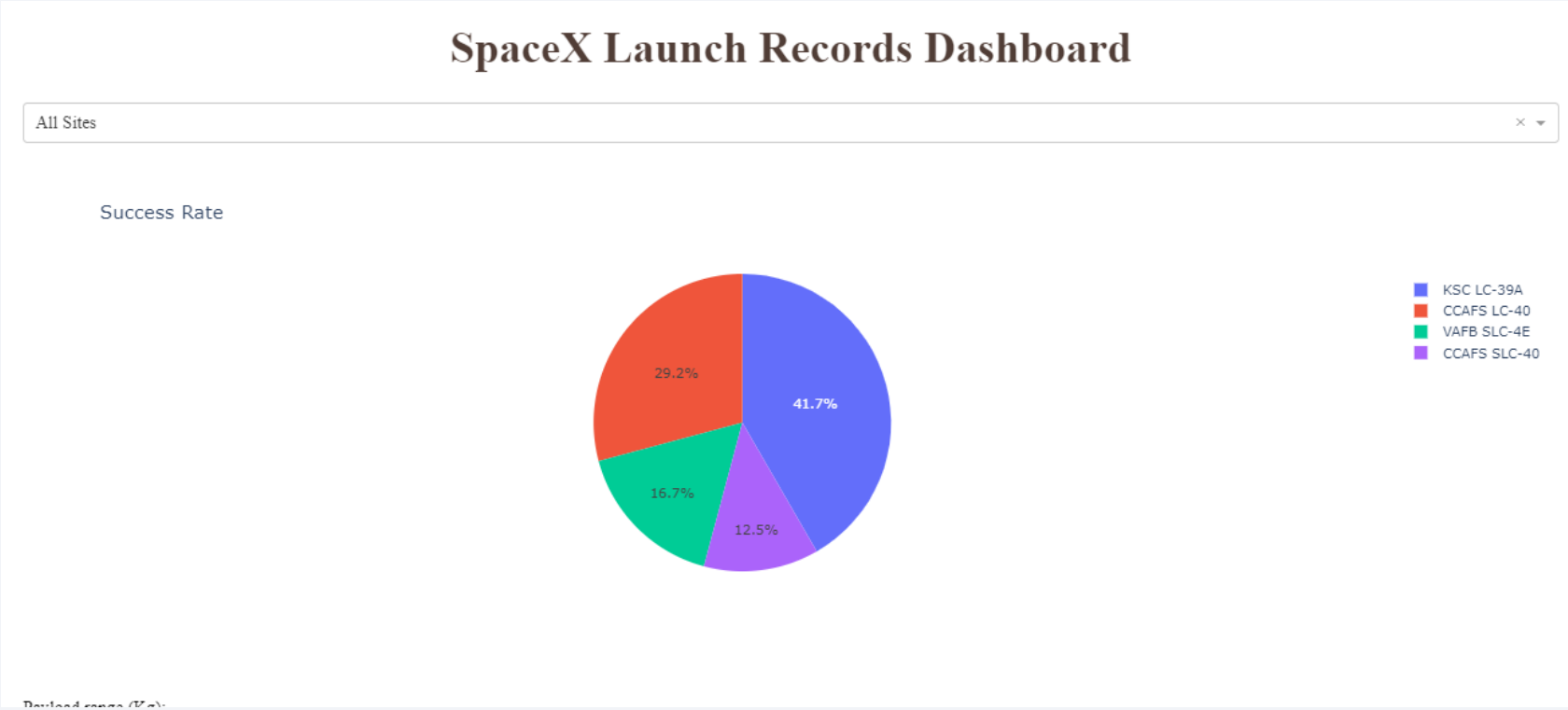
Map of proximity distances from a Launch Site



- The map shows the point of proximities from one launch site and their distances:
  - Coastline
  - Railway
  - Highway
  - City

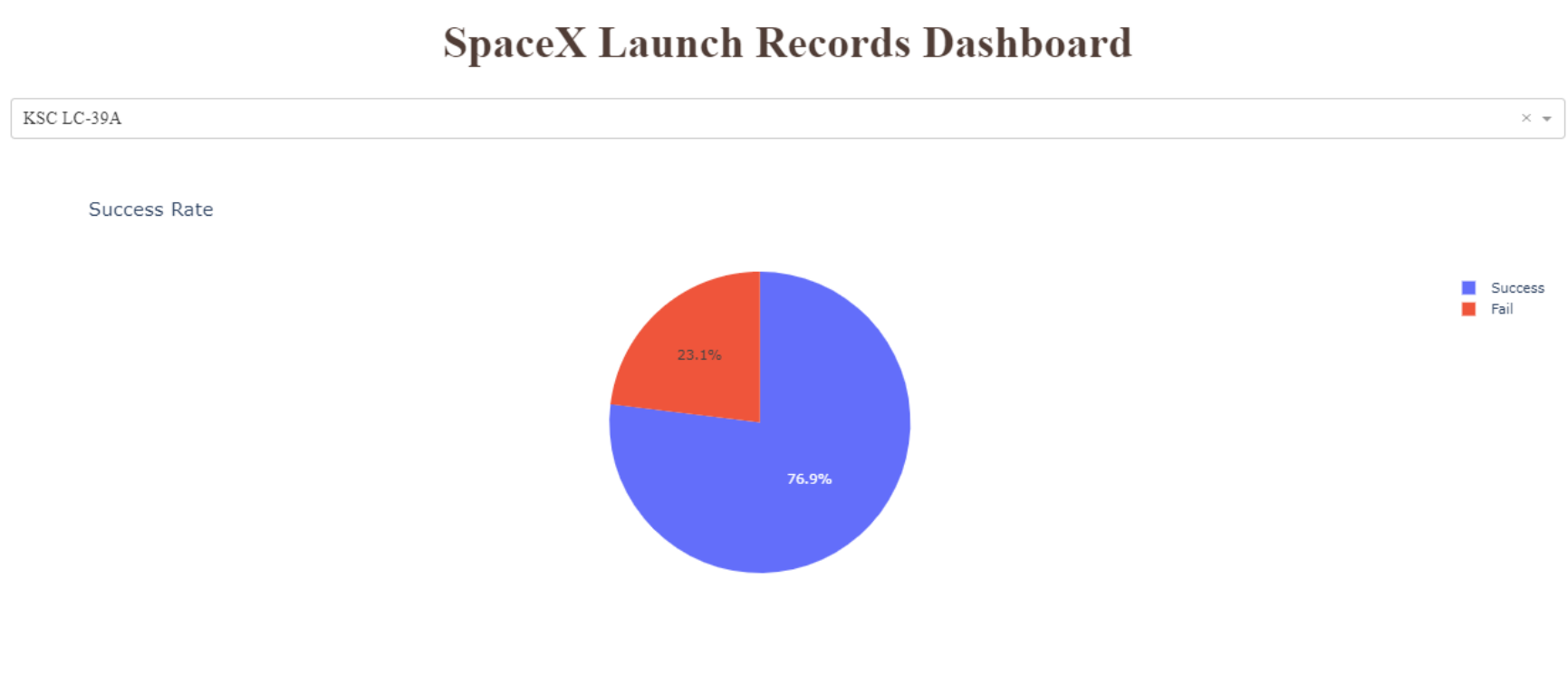

Section 4
Build a Dashboard with Plotly Dash

Launch Success Count for All Sites



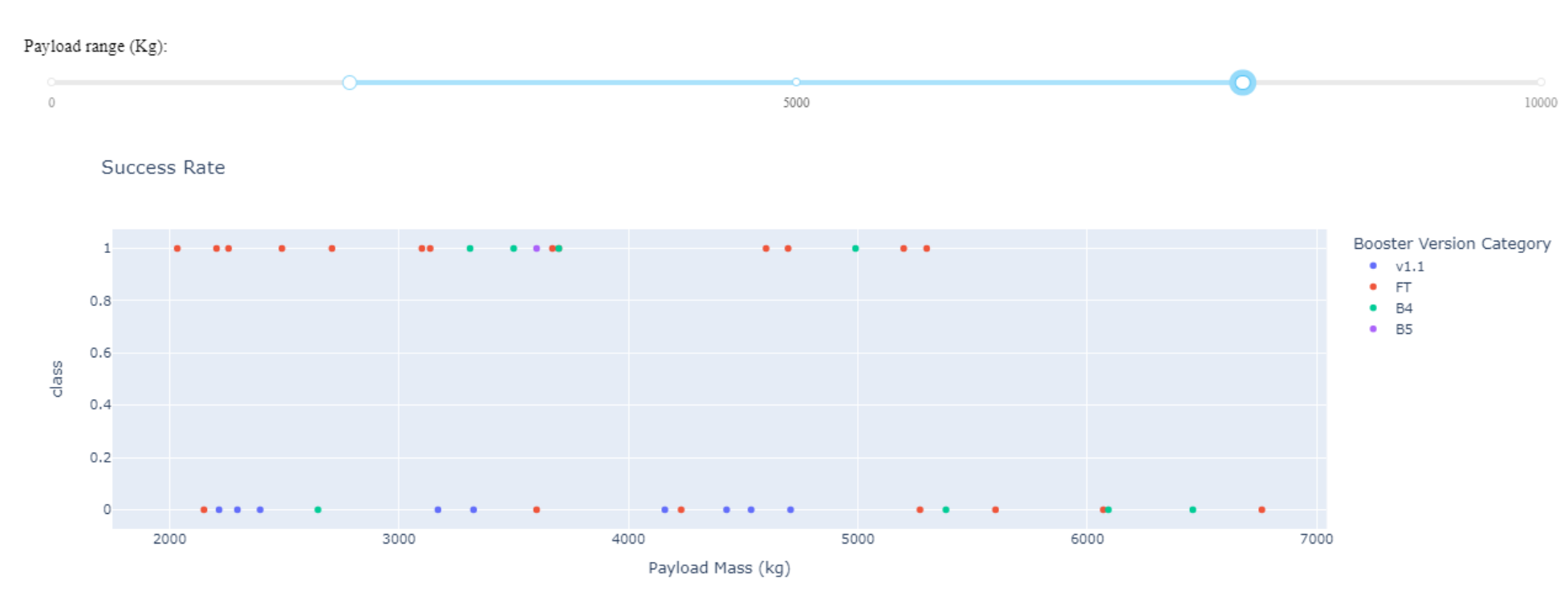
- KSC LC-39A is the launch site with the most success count.
- CCAFS SLC-40 is the launch site with the least success count.

Launch Site with Highest Launch Success Ratio



- KSC LC-39A is the launch site with the most success count.
- Ten launches were successful
- Only 3 launches failed

<Dashboard Screenshot 3>

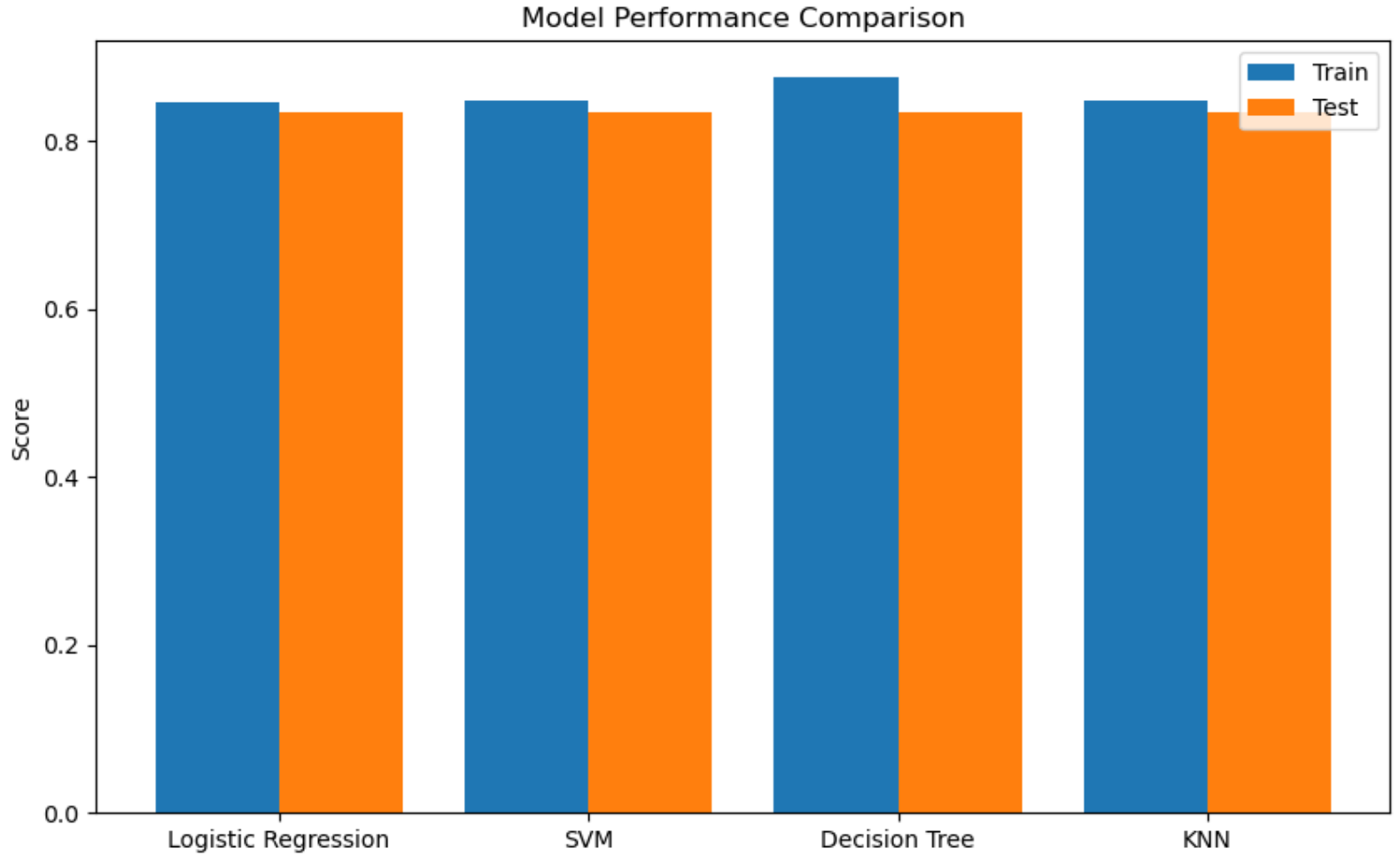


- The most successful launches were with payload between 2000 and 6000
- Booster version category v1.1 had the most failures.
- Booster version category FT had the most successes.

Section 5

Predictive Analysis (Classification)

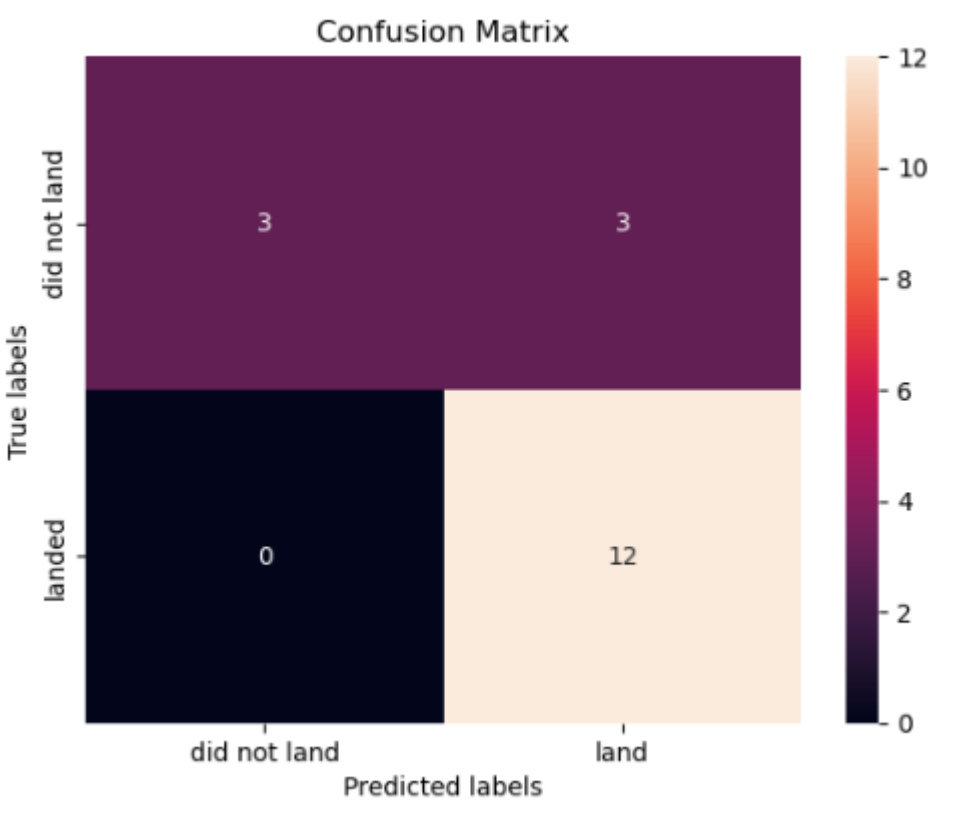
Classification Accuracy



- Decision Tree Classifier has the highest accuracy score:

Model	Train	Test
Logistic Regression	0.846429	0.833333
SVM	0.848214	0.833333
Decision Tree	0.875000	0.833333
KNN	0.848214	0.833333

Confusion Matrix



- The confusion matrix of the Decision Tree Classifier shows that it correctly predicted 12 of all 12 successful landing.
  - Accuracy: 15/18 = 83.33%
  - Precision: 12/15 = 80%
  - Recall: 12/12 = 100%

Conclusions

- The best model to predict successful landing is Decision Tree Classifier, with accuracy score of 83.33% (87.5% on training data)
- The best parameter for this model are:
  - 'criterion': 'gini'
  - 'max_depth': 2
  - 'max_features': 'sqrt'
  - 'min_samples_leaf': 4
  - 'min_samples_split': 2
  - 'splitter': 'random'

Appendix

- List of Falcon 9 and Falcon Heavy launche:
  - https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches
- Project Github URL:
  - https://github.com/harichang/ibmdatasciencecapstone

Thank you!