IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Kenrick Saputra Tedja
27-06-2023

Outline

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

Executive Summary

- Methodologies

The data is collected using web scraping and SpaceX API utilizing Python to perform Exploratory Data Analysis including data wrangling, data visualization, making an interactive visual dashboard, and performing a machine learning process to predict the landing outcome

- Result

Public sources could be utilized to collect data for analysis and with EDA, the insight of the best feature correlated to the landing outcome is determined and the machine learning prediction shows the best model to use for predicting the landing outcome

Introduction

- The objective of this project is to determine the cost per launch from SpaceX for Space Y to find insight to compete with SpaceX
- The outcome of this project is to estimate the total cost of each launch based on the successful landing of the first stage

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was obtained from SpaceX API and List of Falcon 9 and Falcon Heavy launches
- Perform data wrangling
  - Landing Outcome feature created based on the Outcome feature after summarizing data
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - The dataset was normalized and divided into training and test datasets and evaluated using four classification model

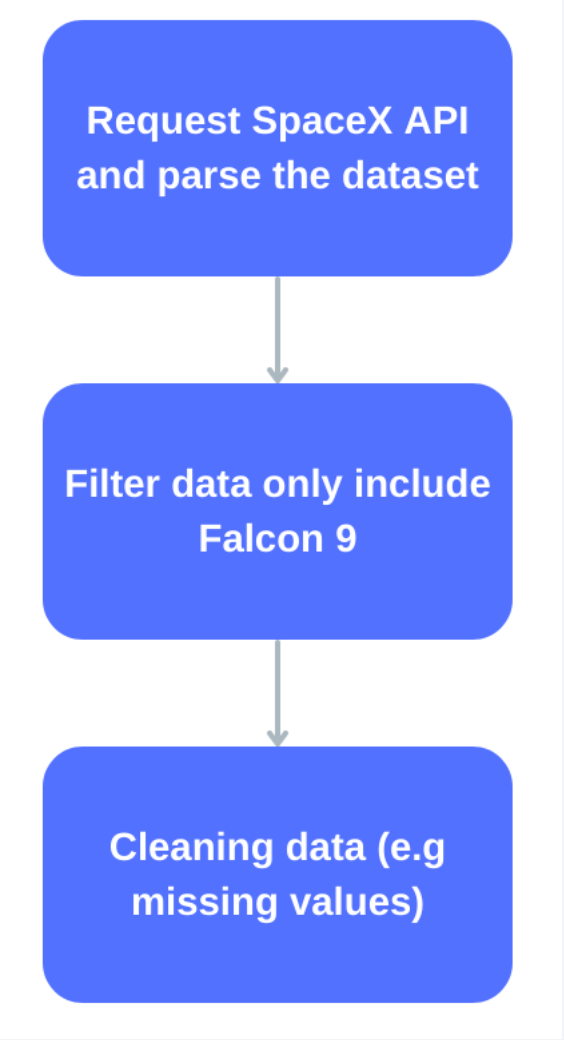
Data Collection

- First data comes from SpaceX API
- Second data web scraped from List of Falcon 9 and Falcon Heavy launches

Data Collection – SpaceX API

- SpaceX API offers public data and requested using “request” package from Python. The data was used according to the flowchart

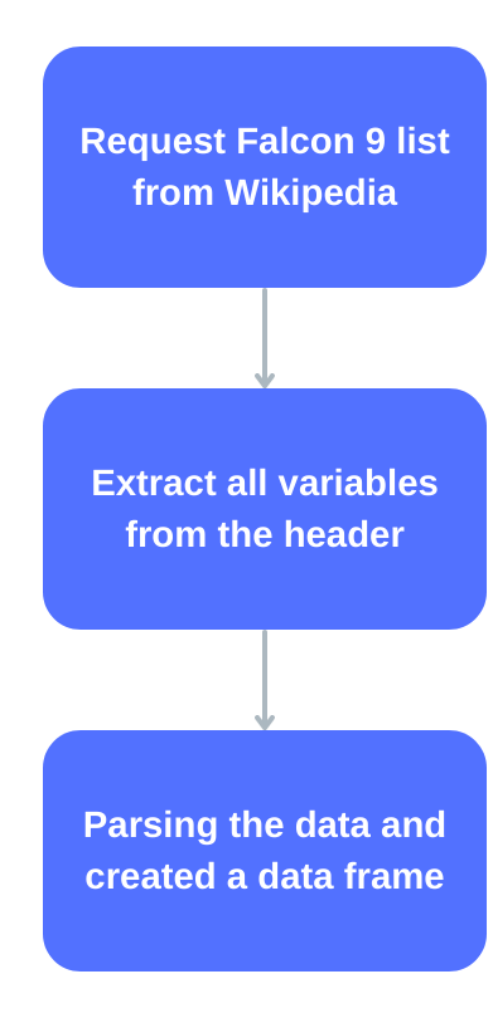
- The code for the following process is presented through this link (Data Collection)



Data Collection - Scraping

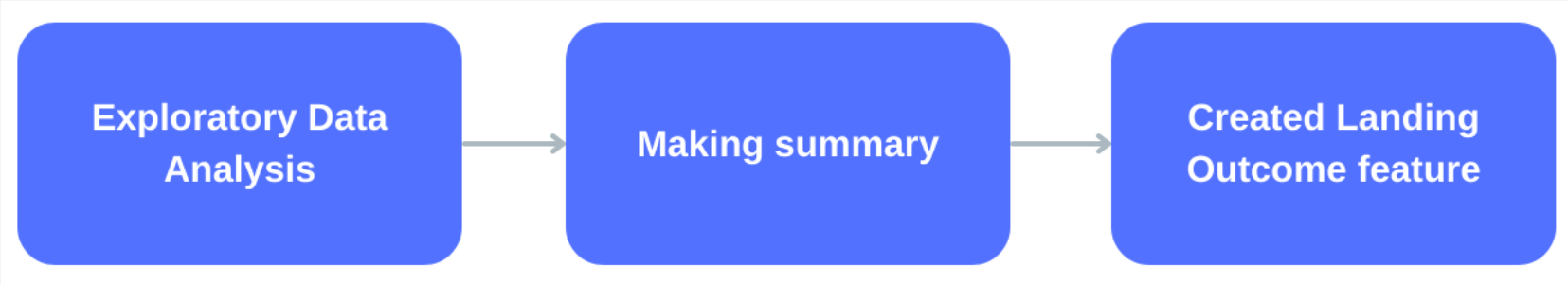
- The list of Falcon 9 launches was requested using JSON from Wikipedia

- The code for the following process is presented through this link (Web Scraping)



Data Wrangling

- Exploratory Data Analysis conducted toward the dataset using data visualization and SQL
- The summaries of the dataset including the launches and outcomes with correlation with each feature made. Then, the Landing Outcome feature is created
- The code for the following process is presented through this link (Data Wrangling)



EDA with Data Visualization

- Scatterplot and Barplot were utilized to see the correlation between each features
- The code for the following process is presented through this link (Data Visualization)

EDA with SQL

- Using SQL, these queries were perfomed:
  - 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 achieved
  - 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
  - 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 successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order
- The code for the following process is presented through this link (EDA with SQL)

Build an Interactive Map with Folium

- Markers, circles, lines, and markers cluster were made using Folium
- These items were used to highlight points and cluster similar points also measure the distance between two coordinates
- The code for the following process is presented through this link (Folium Analysis)

Build a Dashboard with Plotly Dash

- Using Plotly Dash, pie chart and scatter plot were made
- The chart and plot were utilized to see the percentage of launches and correlation between payload and launch sites
- The code for the following process is presented through this link (Plotly Dashboard)

Predictive Analysis (Classification)

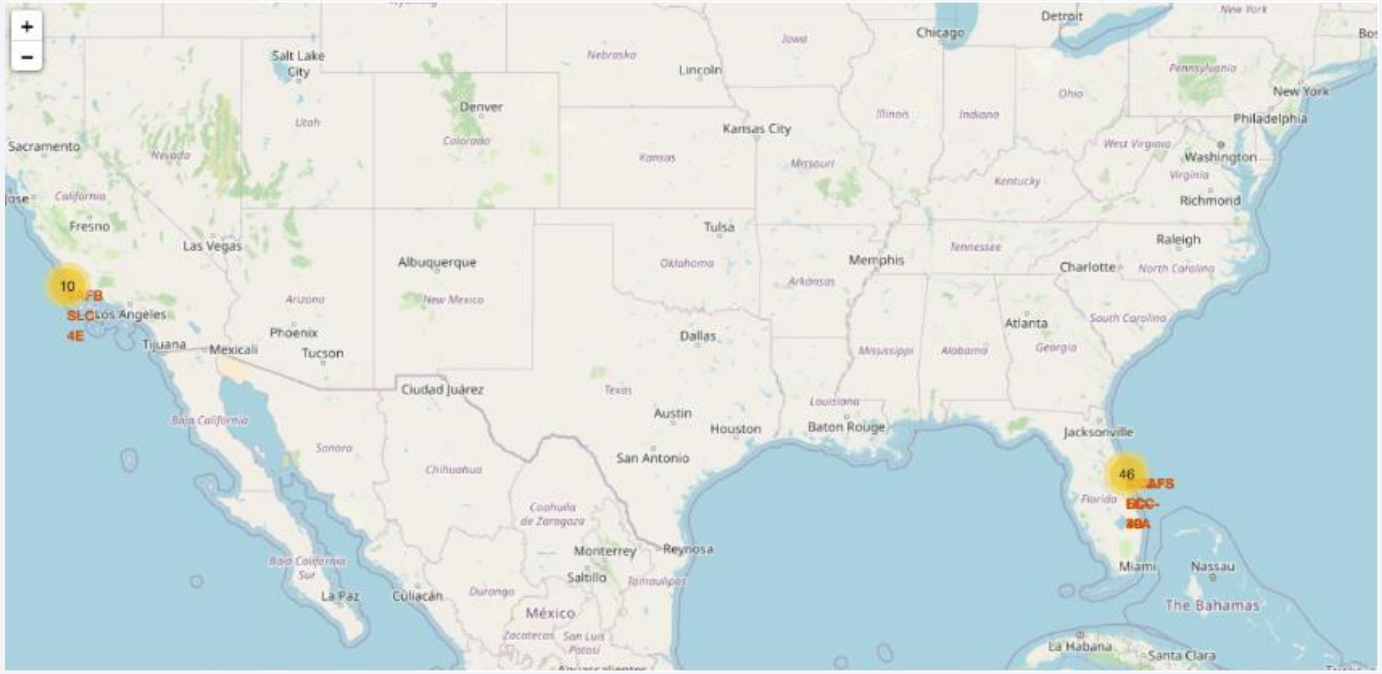
- The prediction was conducted using four classification models including Logistic Regression, SVM, Decision Tree, and KNN
- In this process, the data is prepared and normalized then tested using the classification model to see which model perform the best
- The code for the following process is presented through this link (Predictive Analysis)

Results

- Exploratory data analysis results
  - SpaceX uses four different launch sites: CCAFS LC-40, CCAFS SLC-40, KSC LC-39A. VAFB SLC-4E
  - The average payload for Falcon 9 v1.1 is 2928.4
  - First landing outcome comes at 2018
  - Almost every launch was successful

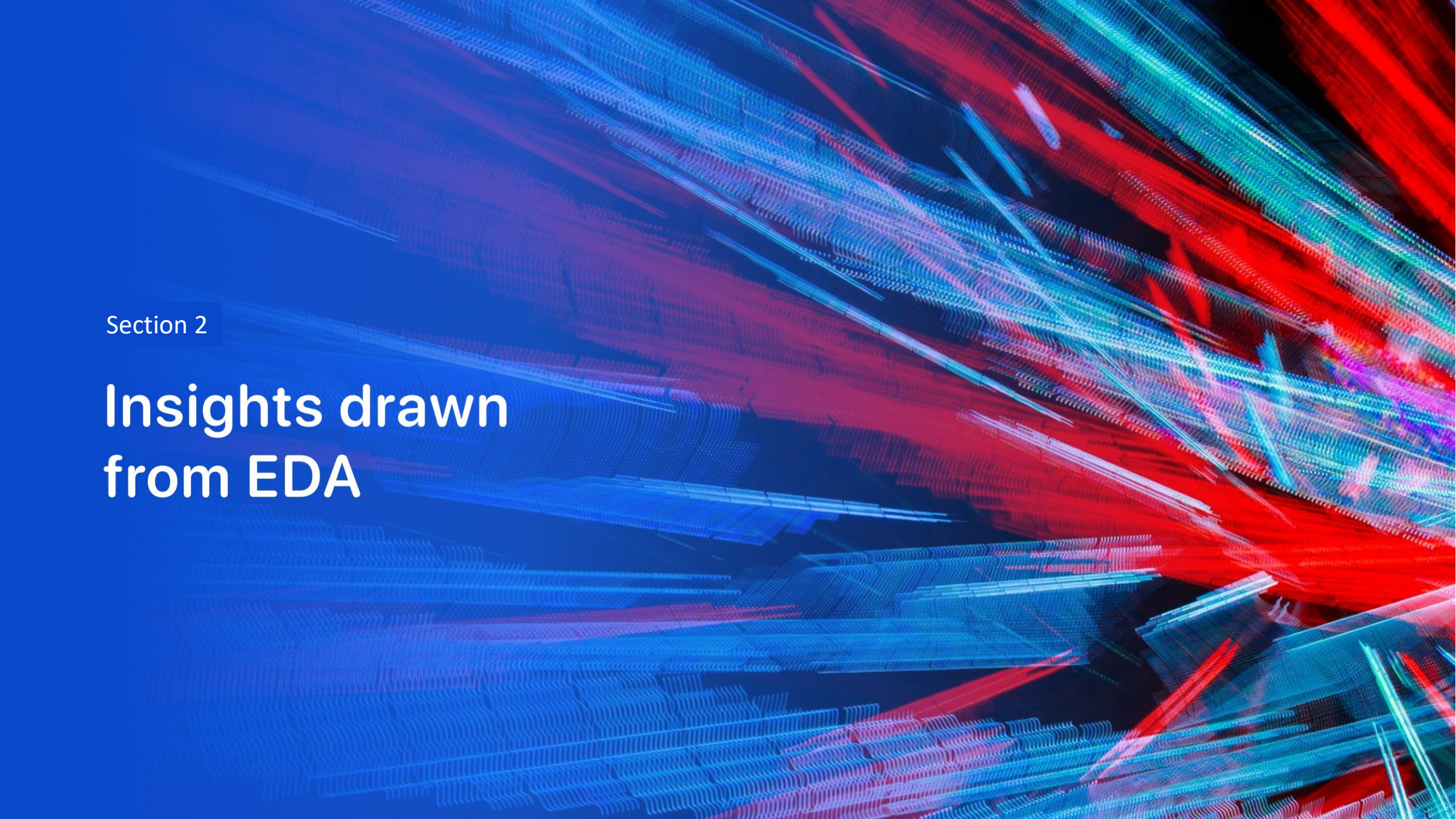
Results

- Interactive maps make it easier to identify launch sites and its surrounding



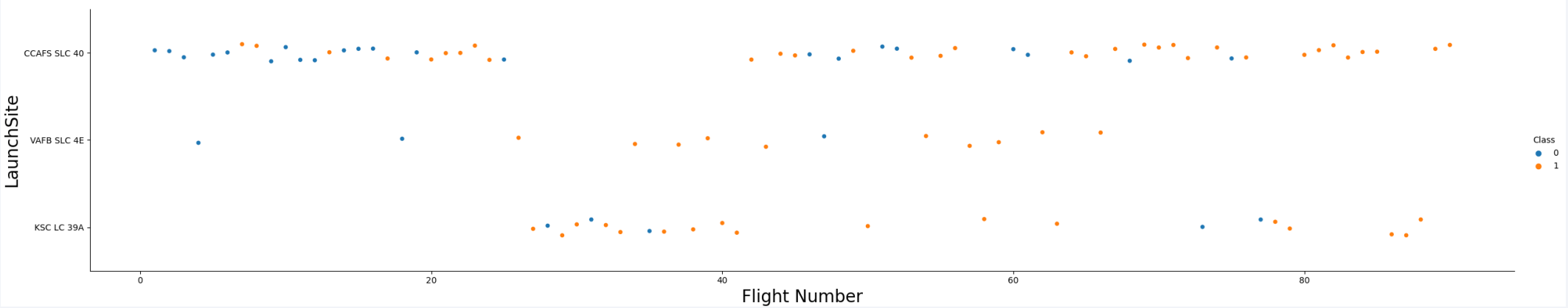
Results

- The predictive analysis shows that Decision Tree is the best model with an accuracy of 89.11%


Section 2
Insights drawn from EDA

Flight Number vs. Launch Site

- According to the plot, CCAFS SLC 40 has most recent successful launch
- The success rate increases over time

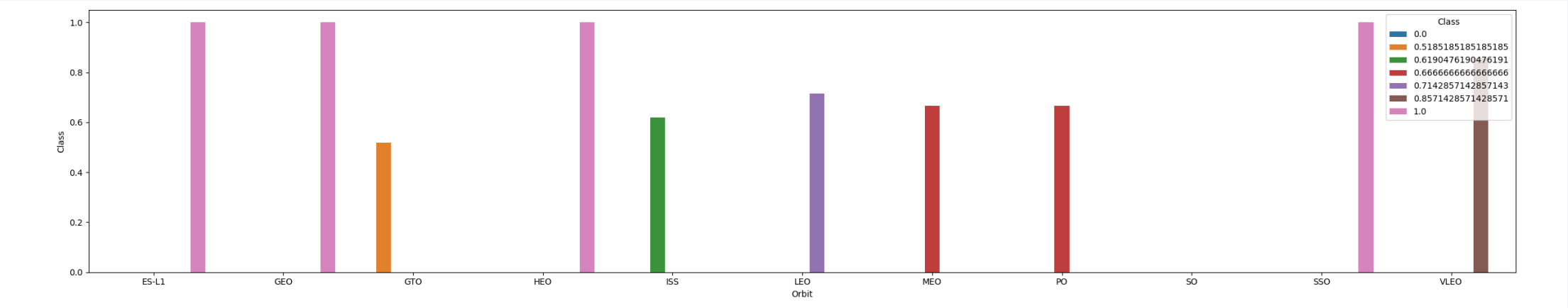


Payload vs. Launch Site

- Higher payloads have higher success rates and according to the plot, only two launch sites are capable of payload over 12000 kg

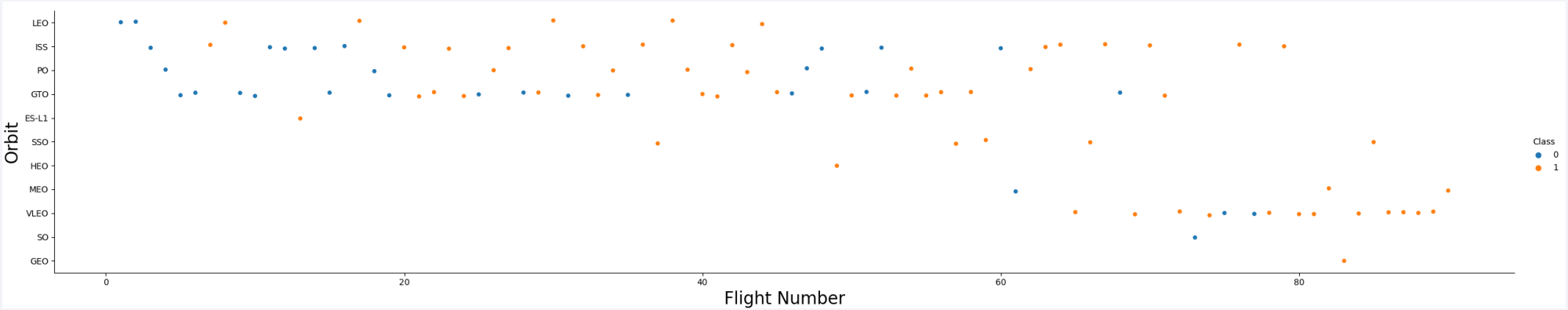
Success Rate vs. Orbit Type

- ES-L1, GEO, HEO, and SSO shows the highest success rate



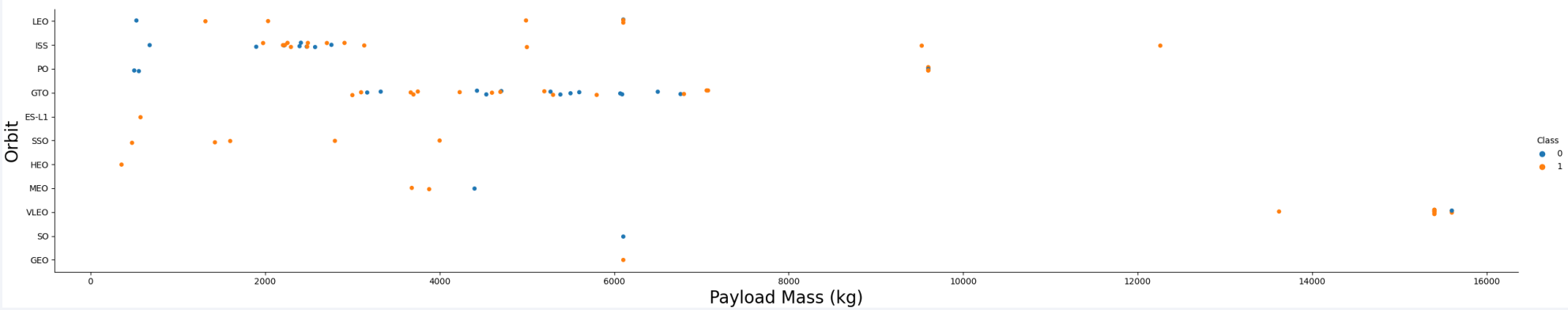
Flight Number vs. Orbit Type

- Success rate to all orbit increases over time



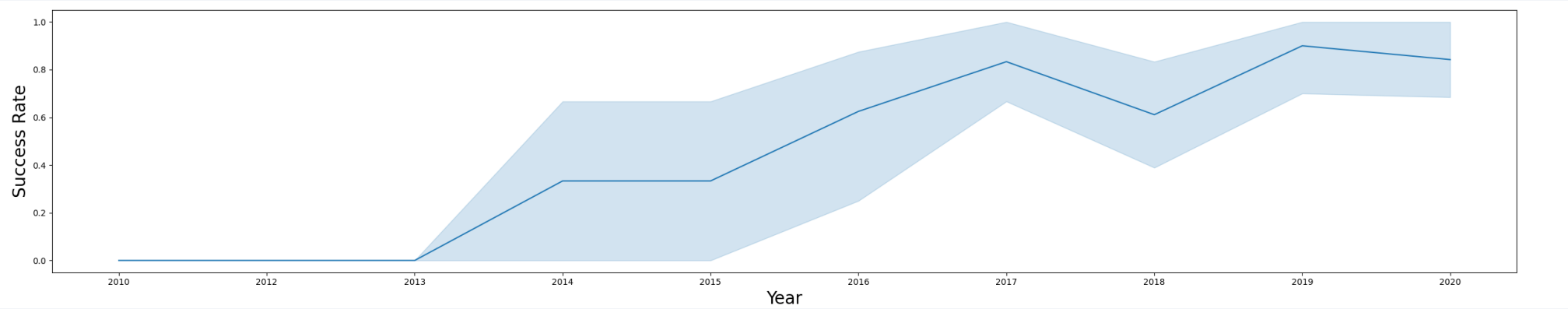
Payload vs. Orbit Type

- Payload and success rate not correlated for GTO
- ISS has better success rate with a wider range
- Few launches for SO and GEO



Launch Success Yearly Trend

- Generally, the launch success is increasing over the year with flops during 2017-2018 and 2019-2020



All Launch Site Names

- There are four launch sites obtained by unique function in launch_site

Launch_Site
None
CCAFS LC-40
CCAFS SLC-40
KSC LC-39A
VAFB SLC-4E

Launch Site Names Begin with 'CCA'

- The record shows launch sites begin with ‘CCA’

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
06/04/2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0.0	LEO	SpaceX	Success	Failure (parachute)
12/08/2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0.0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
22/05/2012	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525.0	LEO (ISS)	NASA (COTS)	Success	No attempt
10/08/2012	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500.0	LEO (ISS)	NASA (CRS)	Success	No attempt
03/01/2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677.0	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- The total payload was calculated using sum function from the payload_mass_kg_

TOTAL_PAYLOAD
111268.0

Average Payload Mass by F9 v1.1

- The average payload mass by F9 v1.1 was calculated using avg function on payload_mass_kg_

AVERAGE_PAYLOAD
2928.4

First Successful Ground Landing Date

- The first successful landing date is determined using min function on date

FIRST_SUCCESS_GP
01/08/2018

Successful Drone Ship Landing with Payload between 4000 and 6000

- List of the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000. These obtained by using distinct function on booster_version

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

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes calculated count function after grouping the mission_outcomes

Mission_Outcome	QTY
None	898
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- List of the names of the booster which have carried the maximum payload mass. These obtained by using distinct function after filtering the maximum payload

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

2015 Launch Records

- List of the failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015. There are only two failure.

month	Date	Booster_Version	Launch_Site	Landing_Outcome
10	01/10/2015	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	14/04/2015	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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

- Rank of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20. The data included no attempt

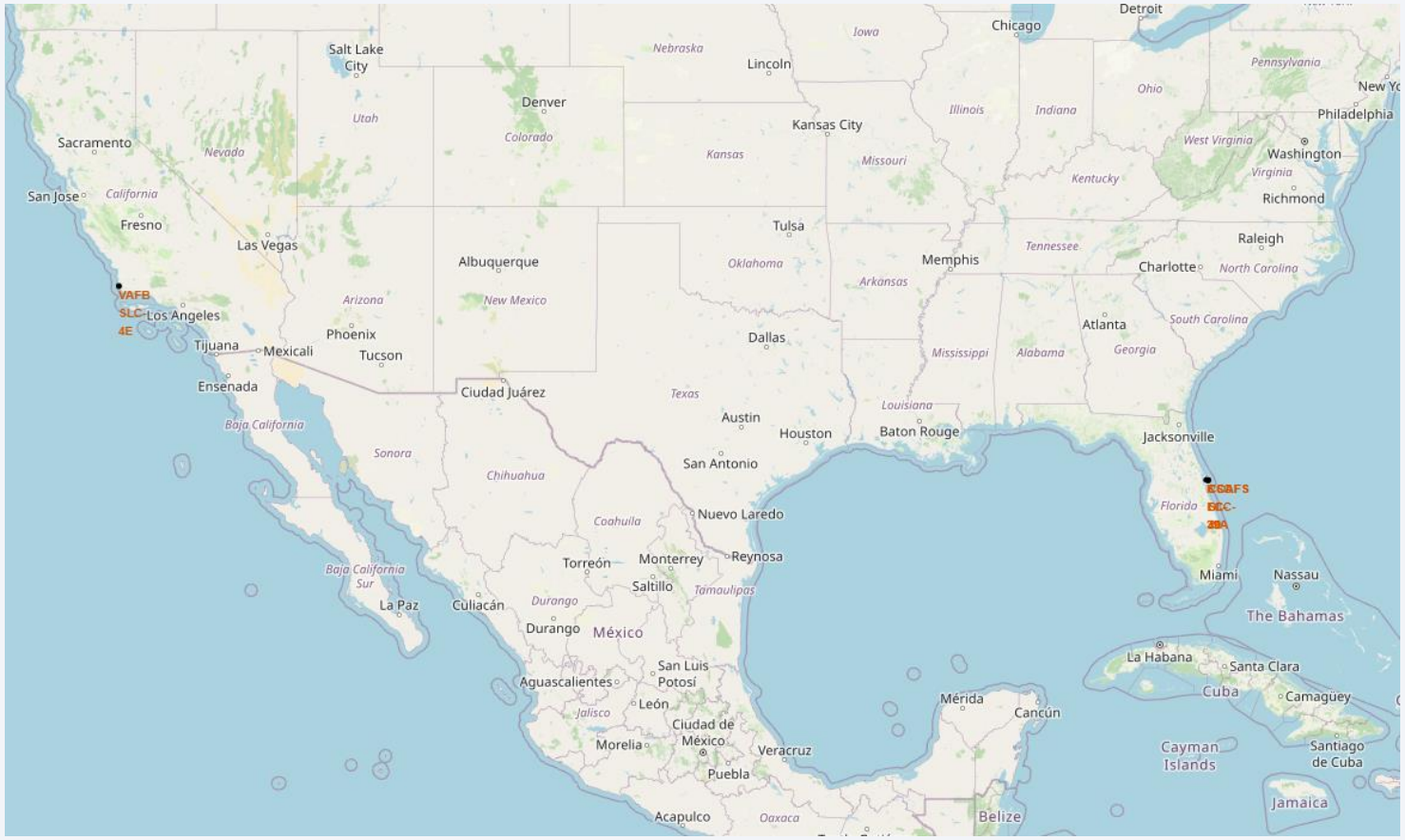
Landing_Outcome	count_outcomes
Success	20
No attempt	10
Success (drone ship)	8
Success (ground pad)	7
Failure (drone ship)	3
Failure	3
Failure (parachute)	2
Controlled (ocean)	2
No attempt	1

Section 3

Launch Sites Proximities Analysis

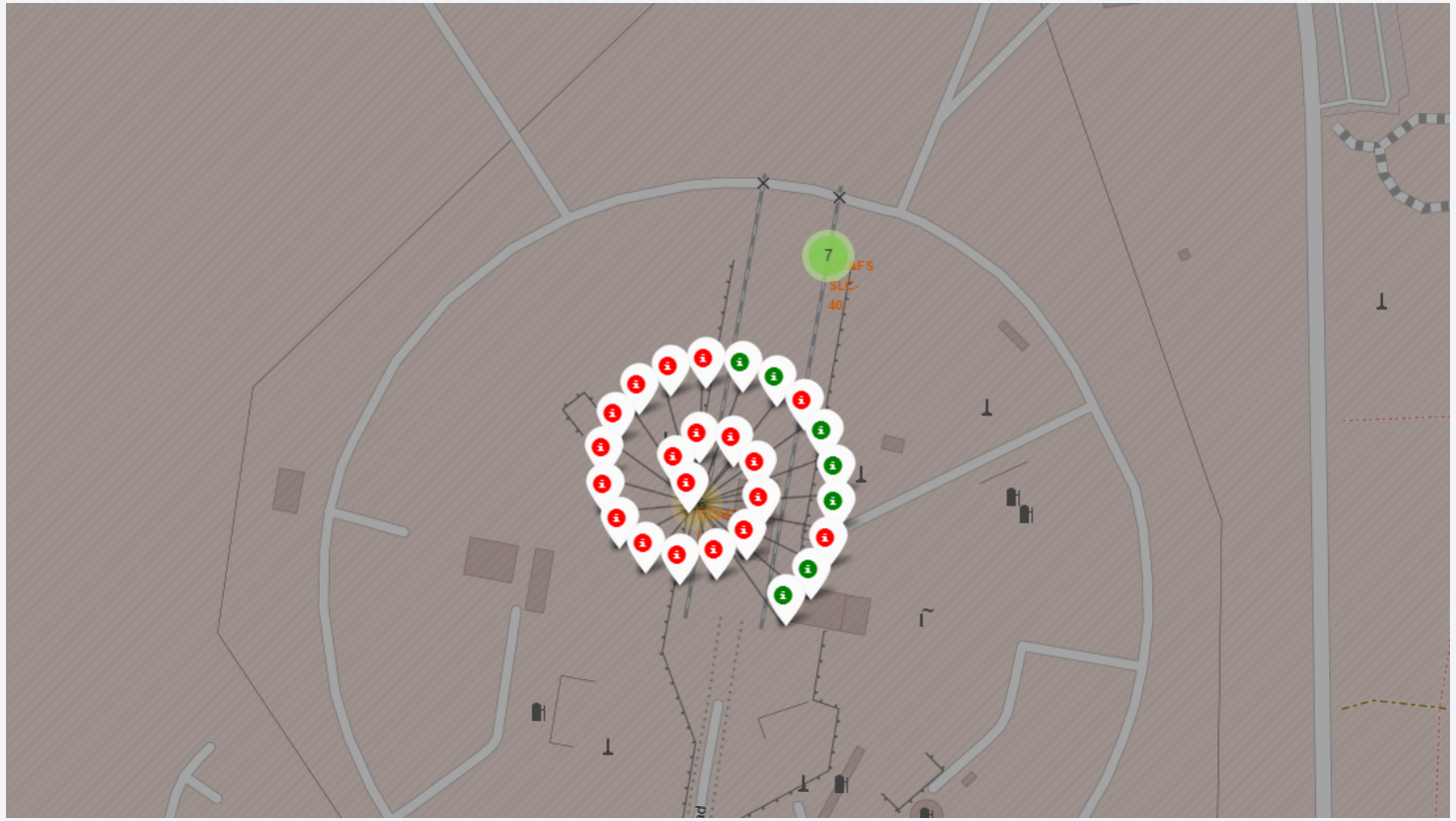
Launch Sites Location

- Every launch site build near railroad as access point and near sea.



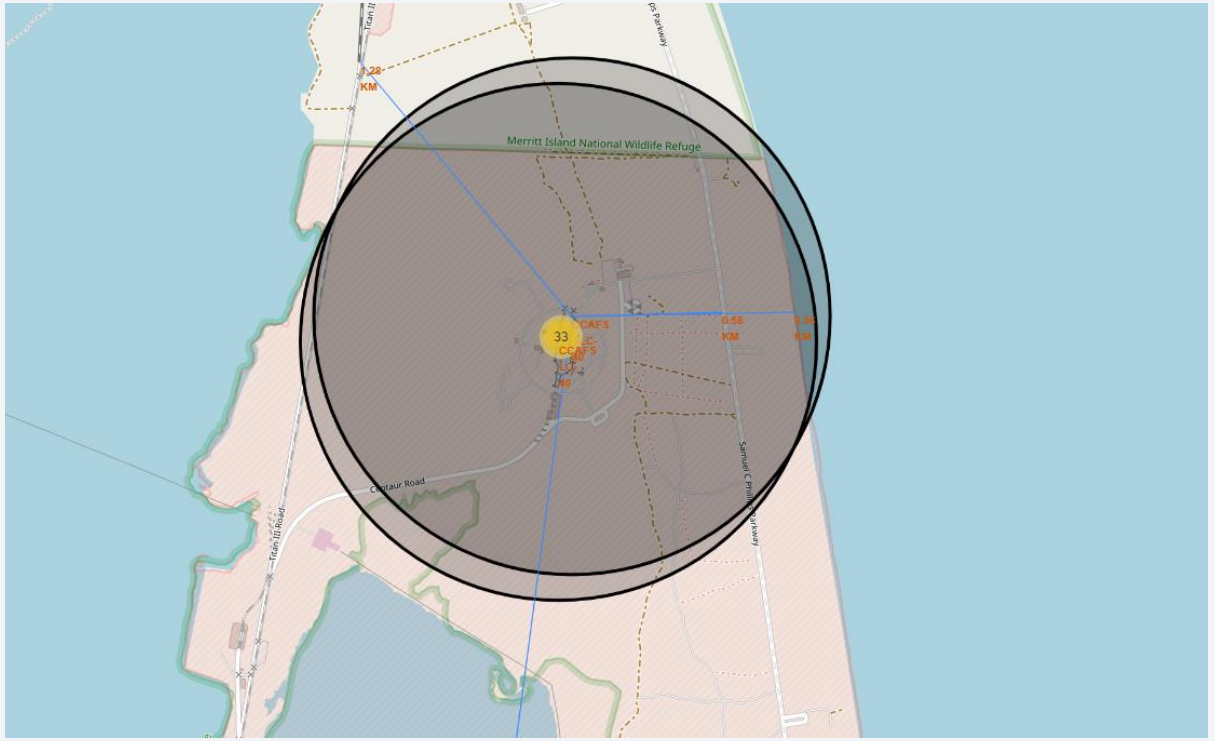
Launch Outcomes

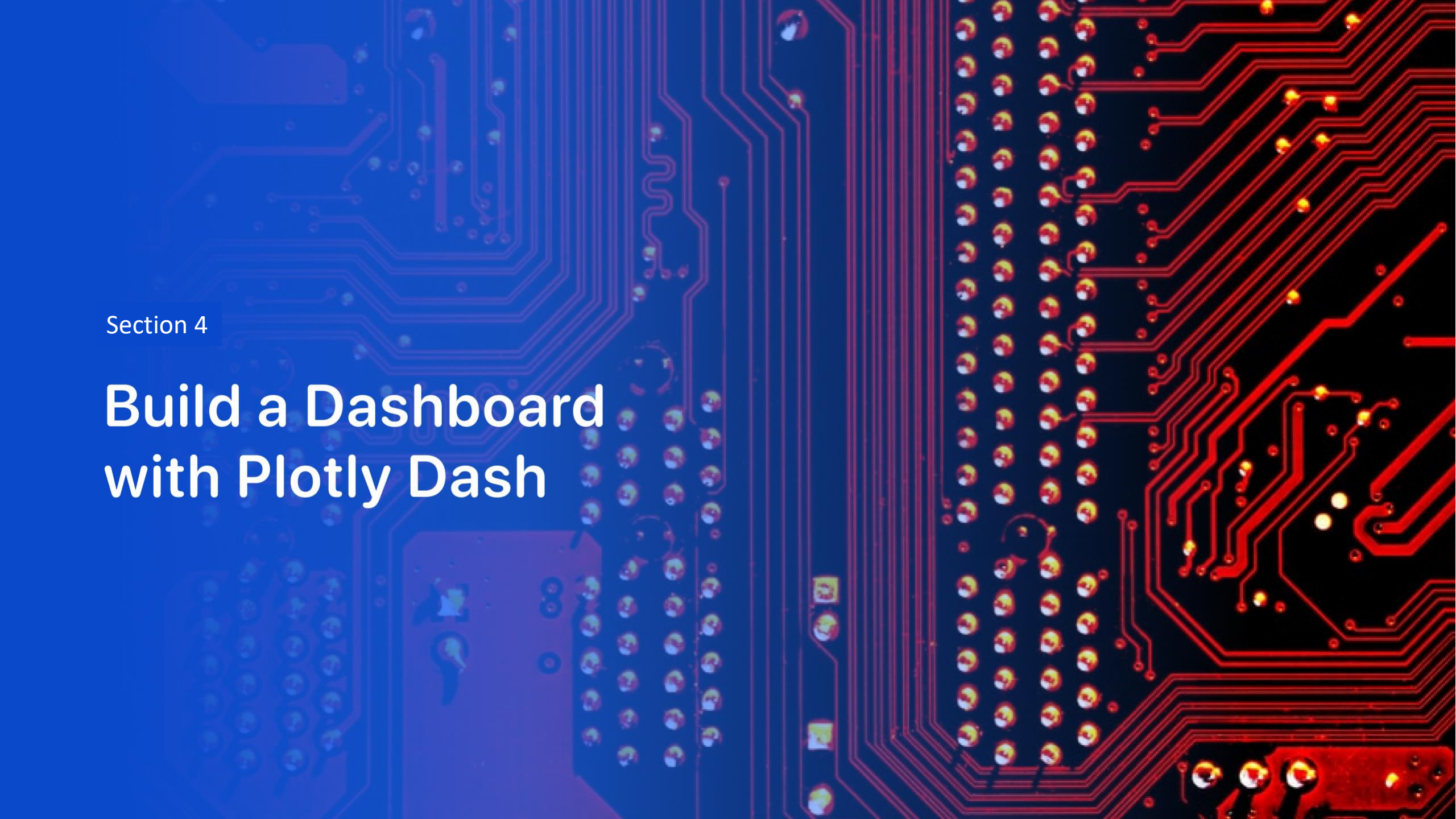
- Launch outcomes are displayed with green as success and red as failure.



Access Point to Launch Sites

- The launch site are connected with railroads from many location which mean the site is very accessible



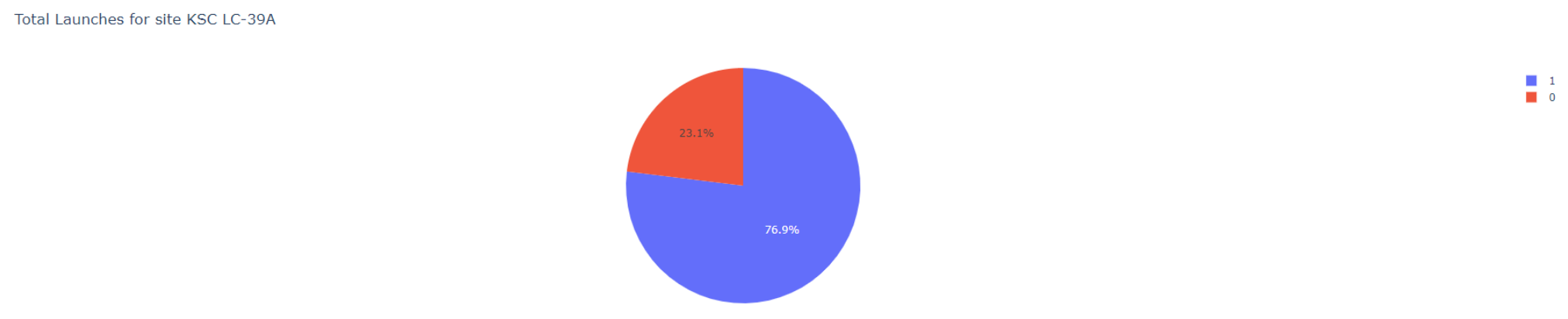

Section 4
Build a Dashboard with Plotly Dash

Successful Launch Sites

- Pie chart showing where the most successful launch site

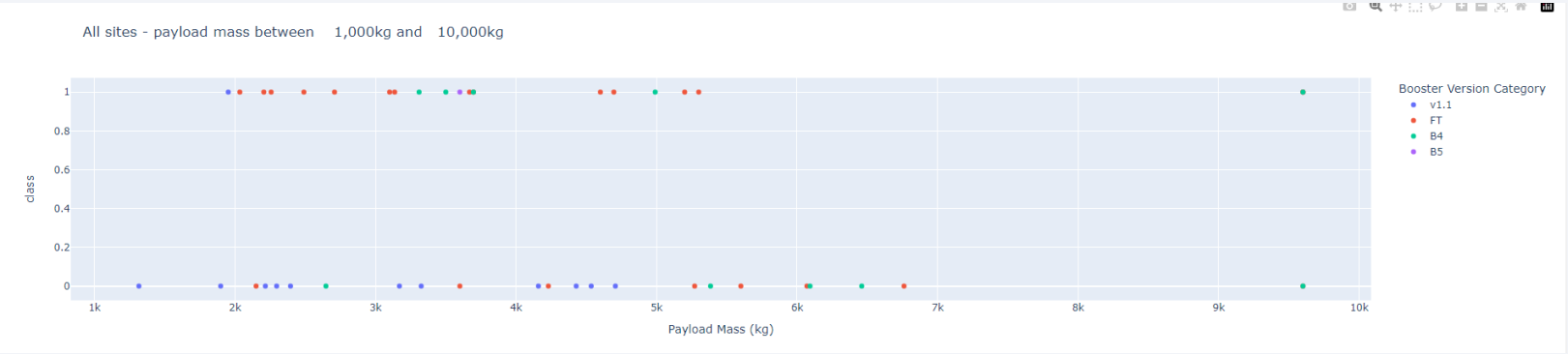
Highest Success Rate Launch Site

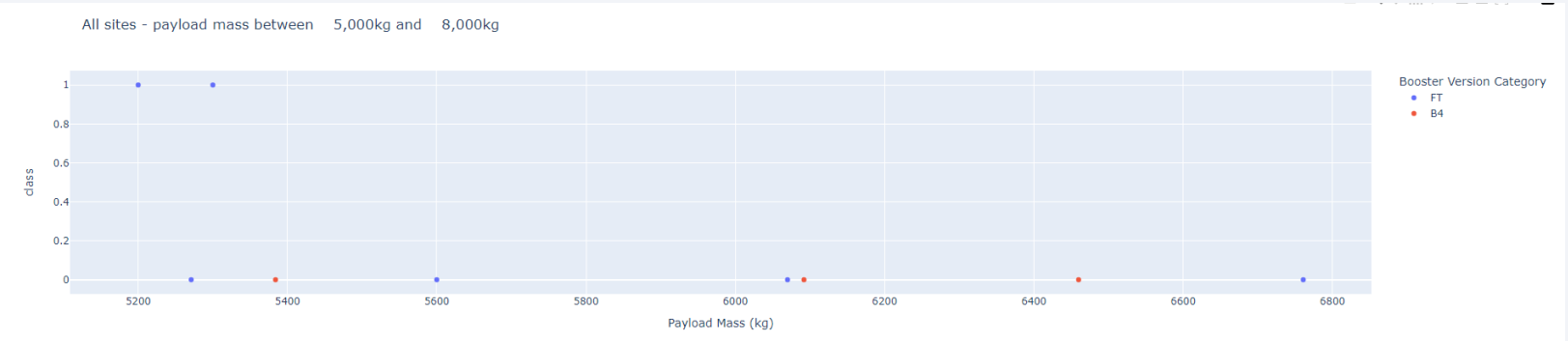
- KSC LC-39A is the most successful launch site with a success rate of 76.9%



<Dashboard Screenshot 3>

- Different payloads have different outcomes. Payload with in the range 2,000-6,000 has the most success rate while in the range of 5,000-8,000 have fewer data to calculate.



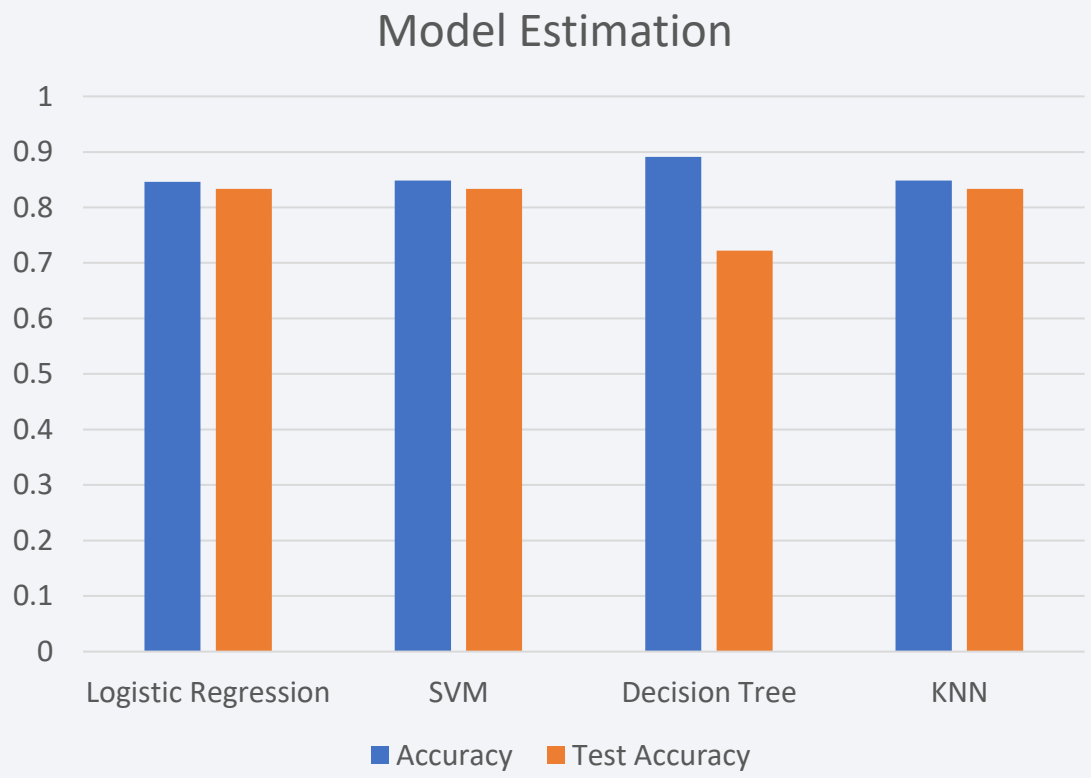


Section 5

Predictive Analysis (Classification)

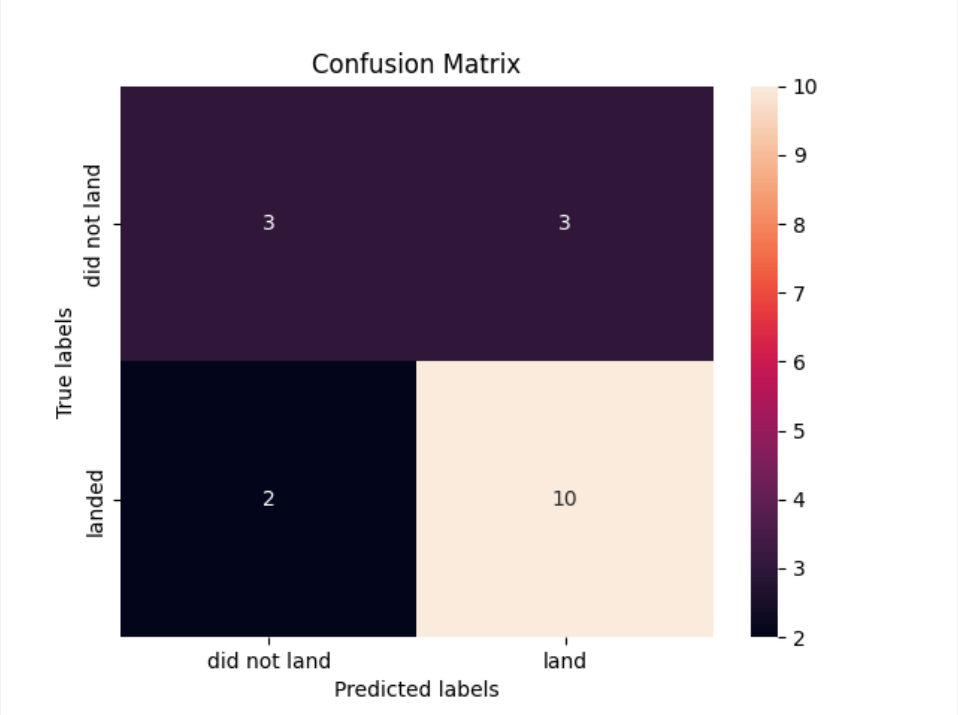
Classification Accuracy

- Decision Tree is the best model with the highest accuracy out of four different model
- Decision Tree has an accuracy of 89.11% and a test of 72.22%



Confusion Matrix

- The confusion matrix shows that Decision Tree has a high number of true positive and true negative



Conclusions

- Two datasets were analyzed and the result was interpreted
- KSC LC-39A is the launch site with the highest success rate
- Launch under 6,000 kg has a high success rate yet a high failure rate
- Success rate increased over time
- Decision Tree classifier is the best model to used made prediction of the landing outcome

Appendix

Thank you!