IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

CH V K N S N Moorthy
14 October 2022

Outline

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

Executive Summary

- Summary of methodologies
    - Collection of Data: With API and Web Scraping
    - Data Wrangling
    - Exploratory Data Analysis: using SQL, using Pandas and Matplotlib
    - Interactive Visual Analytics and Dashboards
    - Predictive Analysis - Classification
- Summary of all results
    - Data collected from public sources
    - EDA identified the features for the prediction of success launchings
    - Machine learning prediction – best model prediction

Introduction

- Project background and context
    - Evaluation of the feasibility of new organization Space Y which can compete with Space X
- Problems you want to find answers
    - The prediction of successful landings of the rockets' first stage by a comparative study on the estimation of the total cost for launches.
    - The prediction of the best place that can be used for launches

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Two methods have been used to get data from Space X
  - 1. with API – https://api.spacexdata.com/v4/rockets/
  - 2.. Using Web Scraping – https://en.Wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches
- Perform data wrangling
  - The features are summarized and analyzed to enhance the quality of the data using the creation of landing outcome

Methodology

Executive Summary

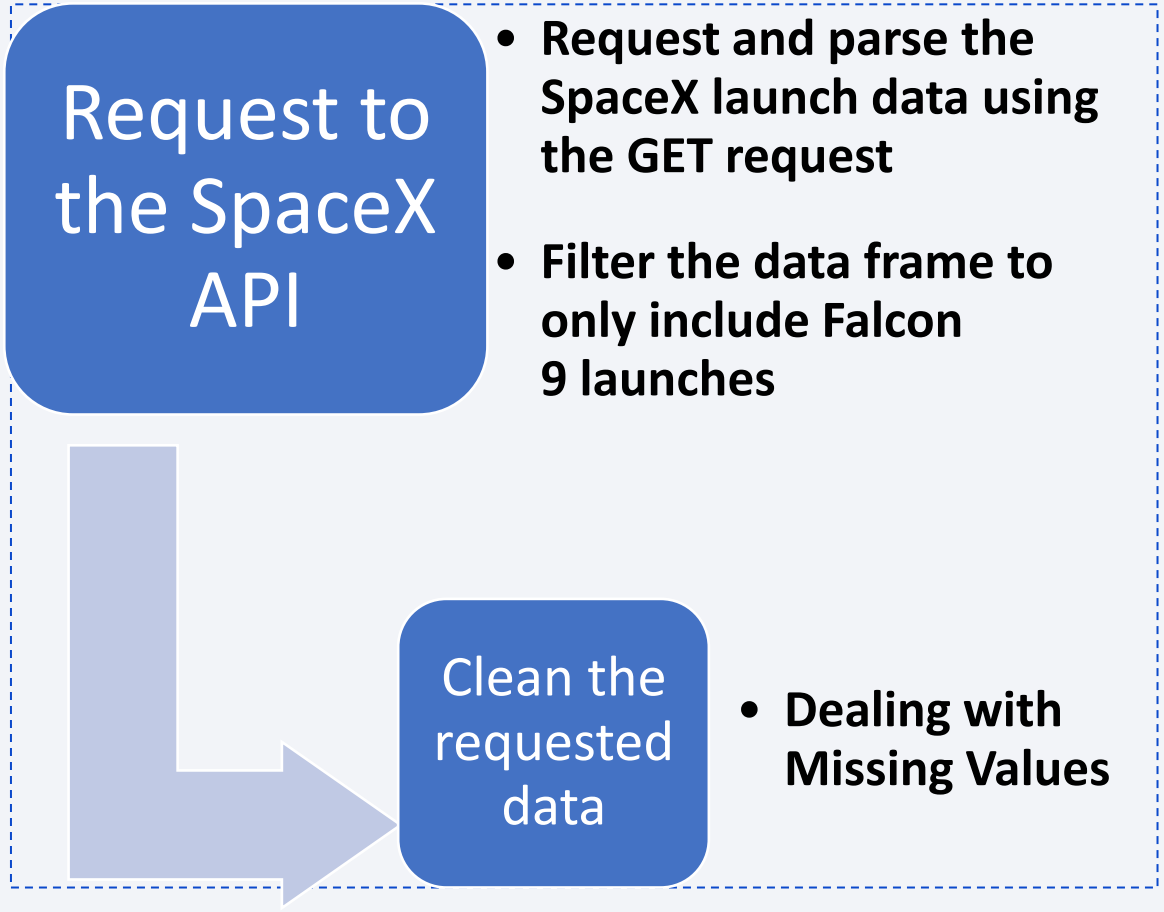
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - The data is first normalized and bifurcated as training and testing data sets. Then the data is evaluated using various classification techniques. The accuracy measure is used for the comparative analysis.

Data Collection

- Data Collection Methods
  - Two methods have been used to get data from Space X
    - 1. with API – https://api.spacexdata.com/v4/rockets/
    - 2.. Using Web Scraping – https://en.Wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches

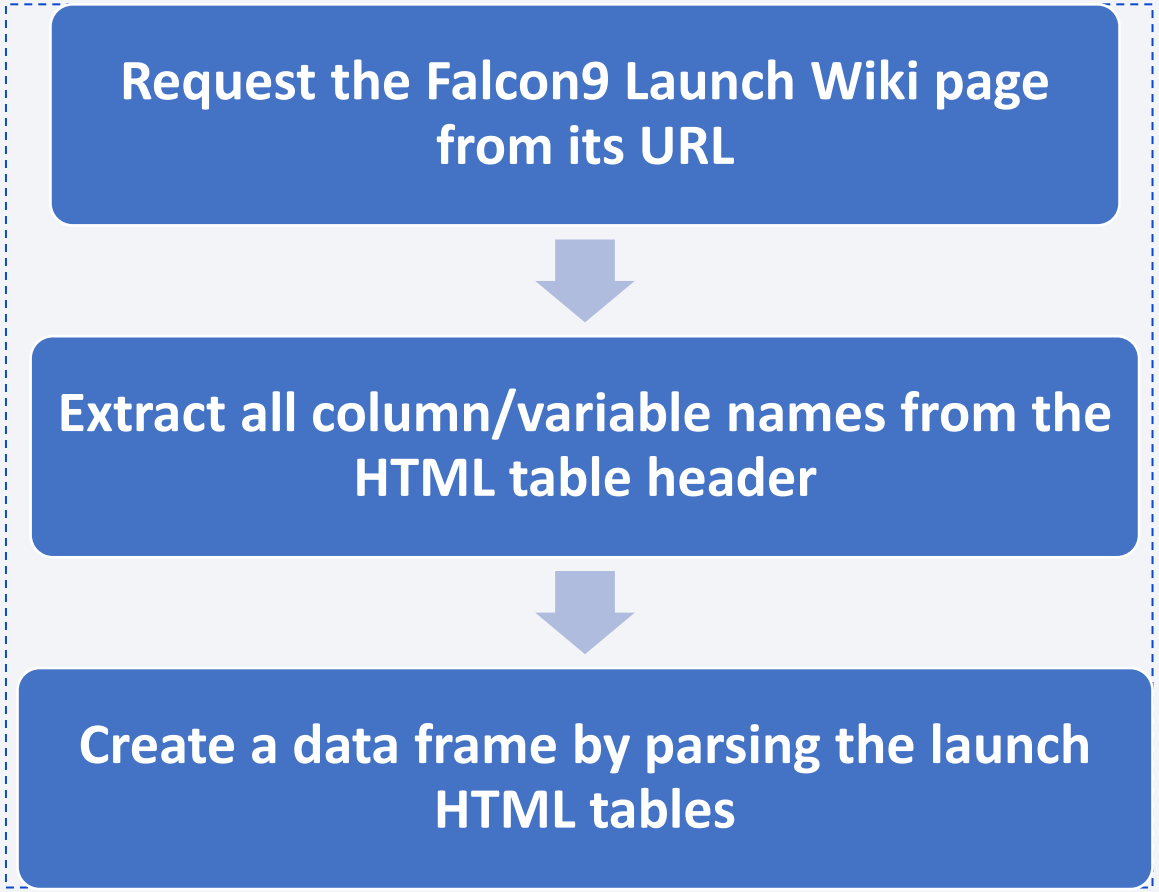
Data Collection – SpaceX API

- The main objective is to make a get request to the SpaceX API and perform some basic data wrangling and formatting.
- The two broad steps include
  - Request to the SpaceX API
  - Clean the requested data
- GitHub URL
- https://github.com/krishnaturbo/Capstone-IBM-Data-Scientist/blob/main/Data%20Collection%20API%20Capstone.ipynb



Data Collection - Scraping

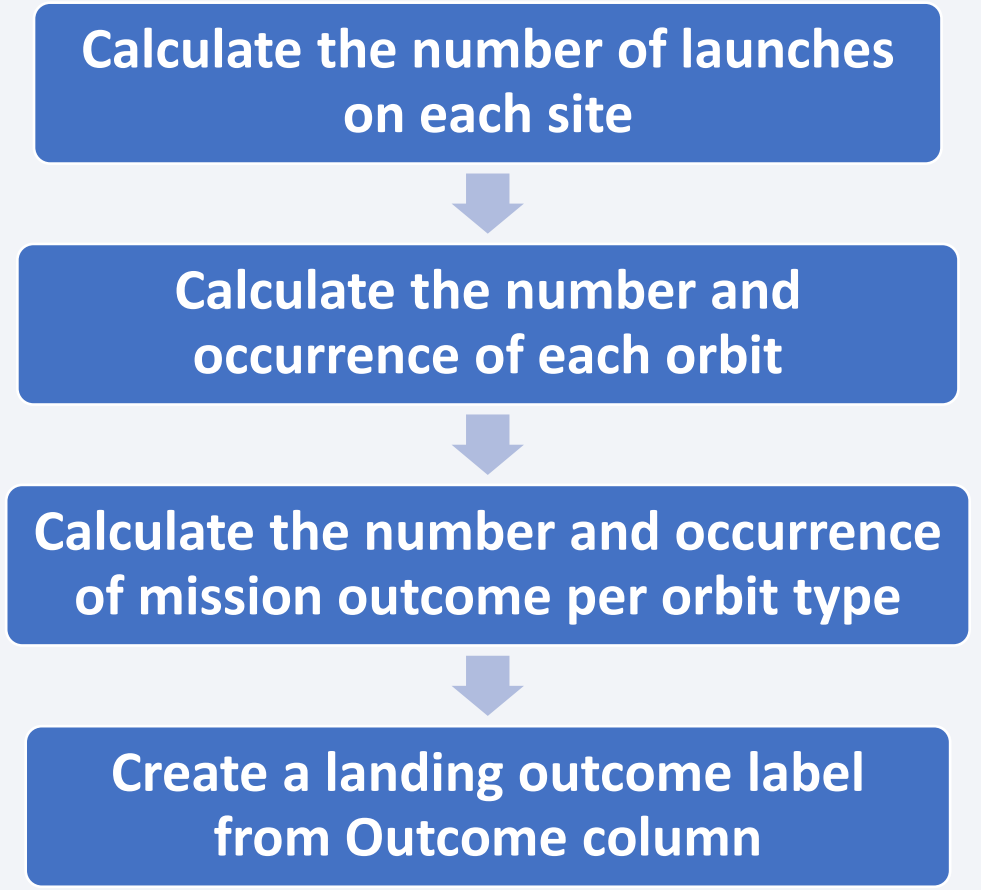
- Web scrap Falcon 9 launch records with BeautifulSoup:
  - Extract a Falcon 9 launch records HTML table from Wikipedia
  - Parse the table and convert it into a Pandas data frame
- GitHub URL
- https://github.com/krishnaturbo/Capstone-IBM-Data-Scientist/blob/main/Data%20Collection%20Web%20Scraping%20Capstone.ipynb



Data Wrangling

- Exploratory Data Analysis was performed, and Training Labels were determined

- GitHub URL
- https://github.com/krishnaturbo/Capstone-IBM-Data-Scientist/blob/main/Data%20Wrangling%20Capstone.ipynb



EDA with Data Visualization

- Exploratory Data Analysis was performed along with Feature Engineering using Pandas and Matplotlib
- Charts developed and their corresponding inferences:
  1. **Flight Number vs. Launch Site scatter point plot** : Site wise flight launch visual
  2. **Payload Vs. Launch Site scatter point chart:** for the VAFB-SLC launch site there are no rockets launched for heavy payload mass(greater than 10000).
  3. **bar chart of success rate:** to find which orbits have high success rate
  4. **Flight Number Vs Orbit type scatter point plot:** LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.
  5. **Payload Vs orbit type scatter point plot:** With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.
  6. **Launch success Yearly trend chart**: to observe that the success rate since 2013 kept increasing till 2020
- GitHub URL: https://github.com/krishnaturbo/Capstone-IBM-Data-Scientist/blob/main/EDA%20with%20Data%20Visualization%20Capstone.ipynb

EDA with SQL

The SQL queries for the following tasks were executed

1. Display the names of the unique launch sites in the space mission
2. Display 5 records where launch sites begin with the string 'CCA'
3. Display the total payload mass carried by boosters launched by NASA (CRS)
4. Display average payload mass carried by booster version F9 v1.1
5. List the date when the first successful landing outcome in ground pad was acheived.
6. List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
7. List the total number of successful and failure mission outcomes
8. List the names of the booster versions which have carried the maximum payload mass. Use a subquery
9. List the failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015
10. 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 URL: https://github.com/krishnaturbo/Capstone-IBM-Data-Scientist/blob/main/EDA%20with%20SQL%20Capstone.ipynb

Build an Interactive Map with Folium

The following tasks were performed

- TASK 1: Mark all launch sites on a map
- TASK 2: Mark the success/failed launches for each site on the map
- TASK 3: Calculate the distances between a launch site to its proximities
- After completed the above tasks, some geographical patterns about launch sites were easily identified.
- GitHub URL
- https://github.com/krishnaturbo/Capstone-IBM-Data-Scientist/blob/main/Interactive%20Visual%20Analytics%20with%20Folium%20Capstone.ipynb

Build a Dashboard with Plotly Dash

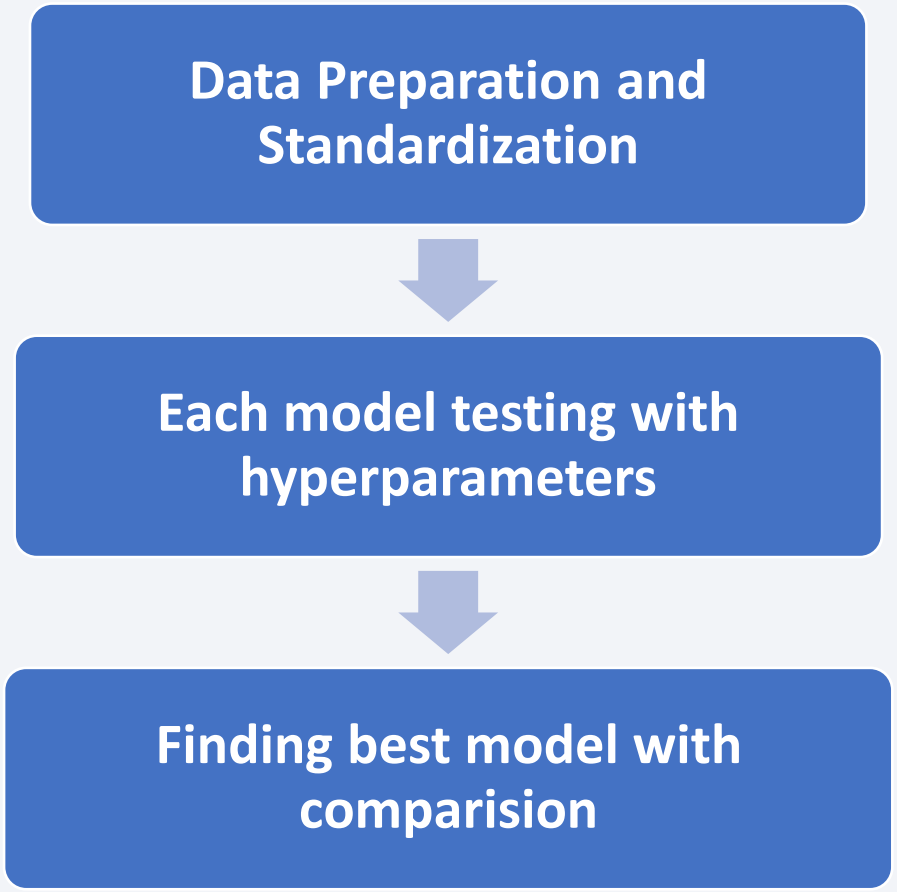
- The following were the objectives in building this dashboard with Plotly Dash
  - 1. to analyze payloads vs launch sites
  - 2. to identify the best place for launching

- To achieve the above the following plots were developed
  - Percentage of launches by site
  - Payload range

- GitHub URL: https://github.com/krishnaturbo/Capstone-IBM-Data-Scientist/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

- Perform exploratory Data Analysis and determine Training Labels
    - create a column for the class
    - Standardize the data
    - Split into training data and test data
- Find best Hyperparameter for SVM, Classification Trees and Logistic Regression
    - Find the method performs best using test data
- GitHub URL:
- https://github.com/krishnaturbo/Capstone-IBM-Data-Scientist/blob/main/Machine%20Learning%20Prediction%20Capstone.ipynb

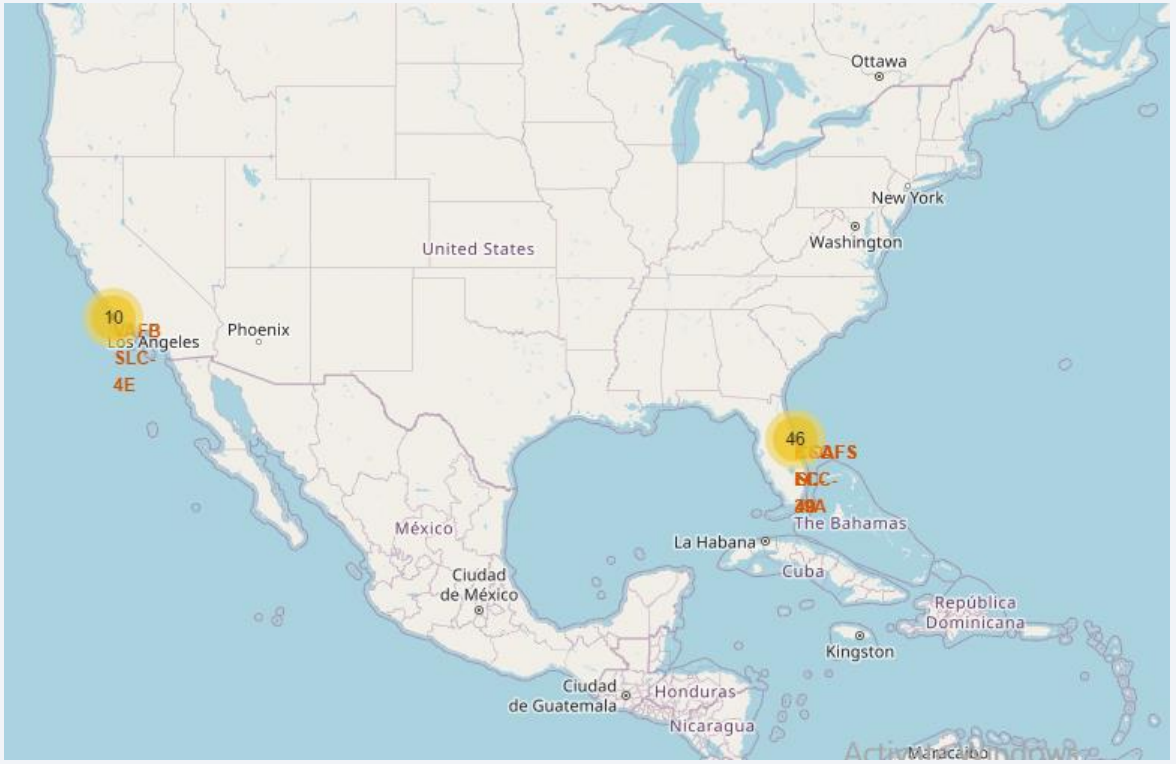


Results

- Exploratory data analysis results
  - as the flight number increases, the first stage is more likely to land successfully.
  - The payload mass is also important; it seems the more massive the payload, the less likely the first stage will return
  - different launch sites have different success rates. CCAFS LC-40, has a success rate of 60 %, while KSC LC-39A and VAFB SLC 4E has a success rate of 77%.
  - for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).
  - in the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.
  - 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
  - the sucess rate since 2013 kept increasing till 2020

Results

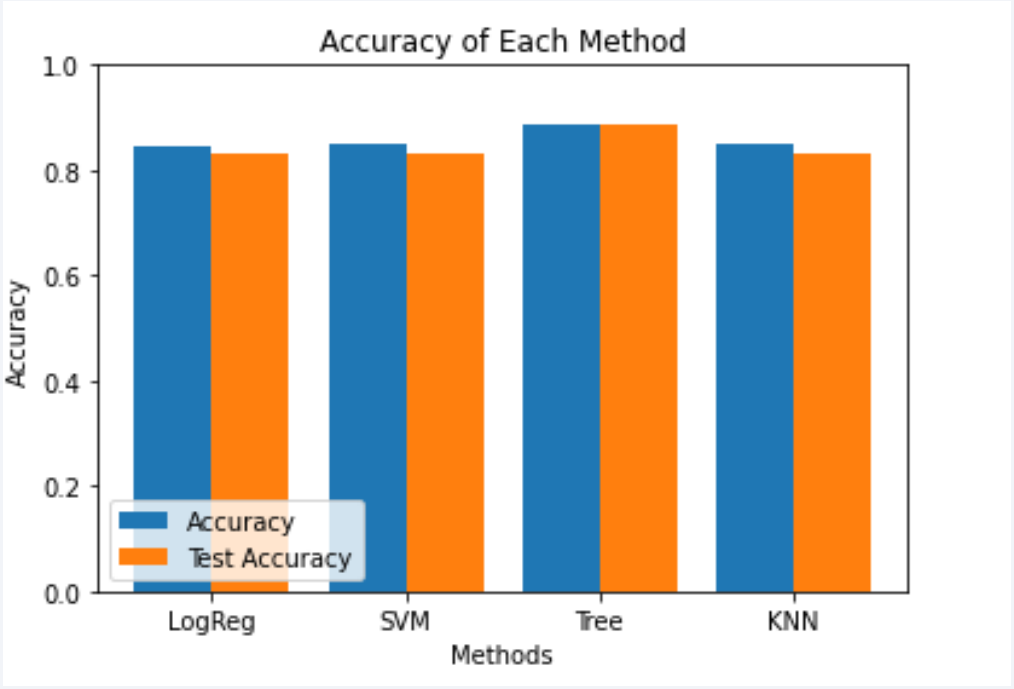
- Interactive analytics demo in screenshots
  - launch sites were marked on a map
  - the success/failed launches for each site were marked on the map
  - the distances between a launch site to its proximities were calculated

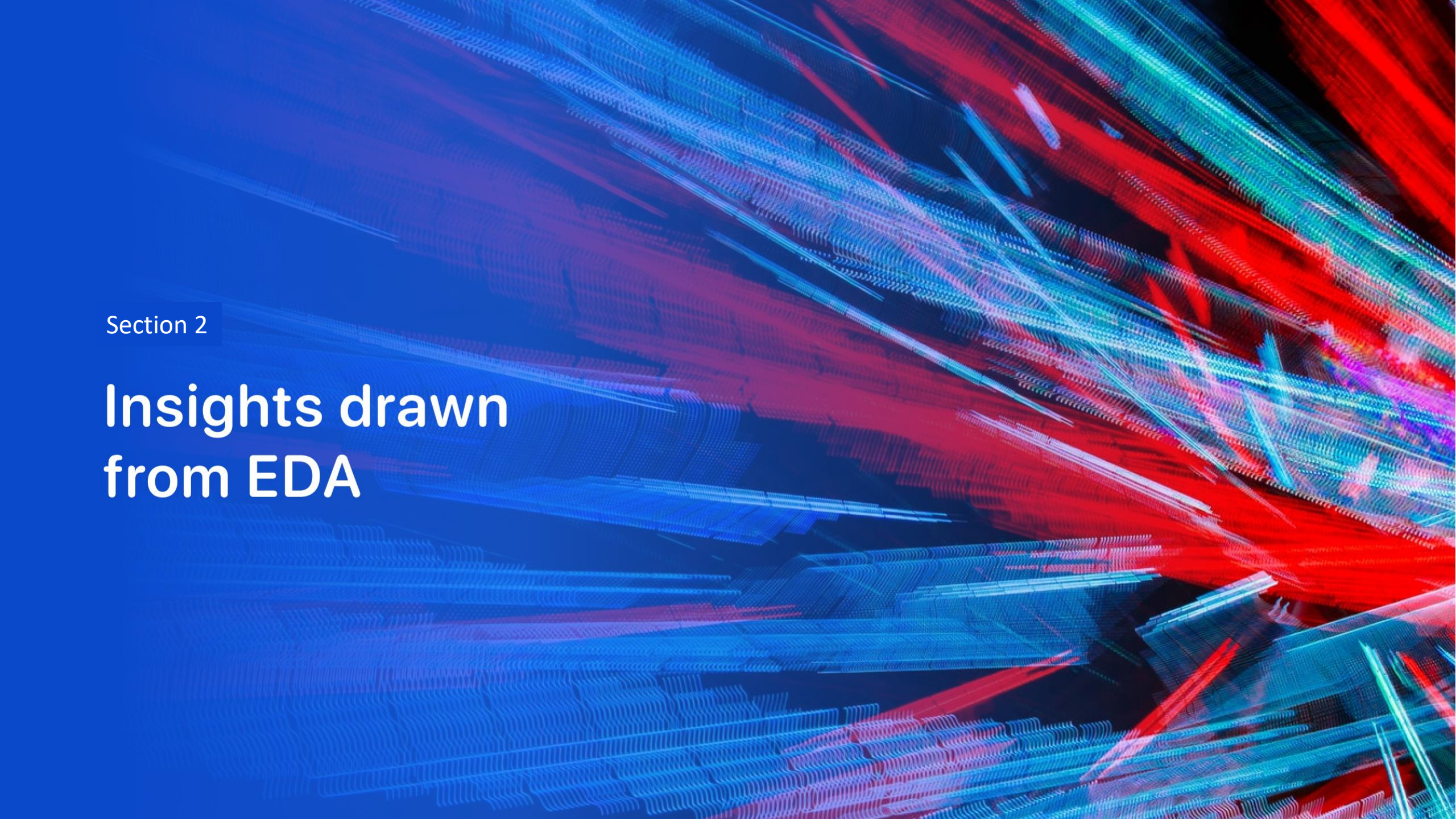


Results

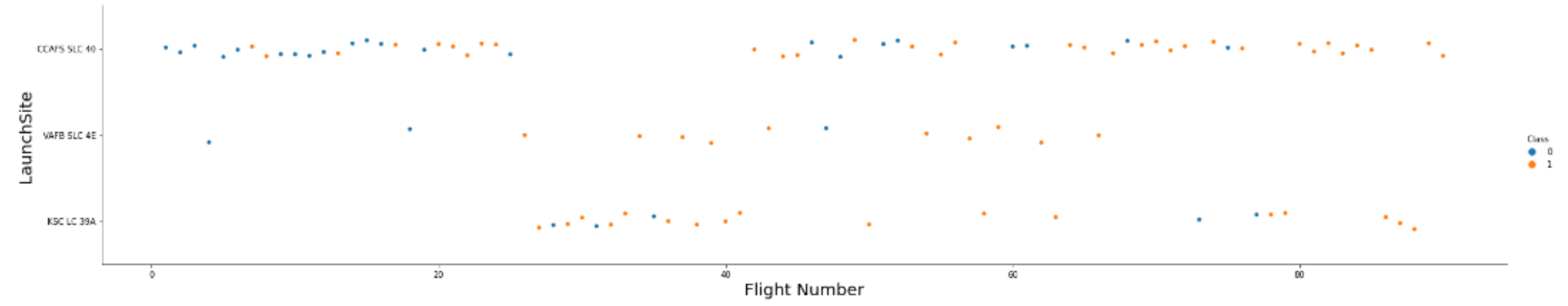
- Predictive analysis results

Decision Tree Model was found to be best model with 88.75% Accuracy and 88.89% Test accuracy



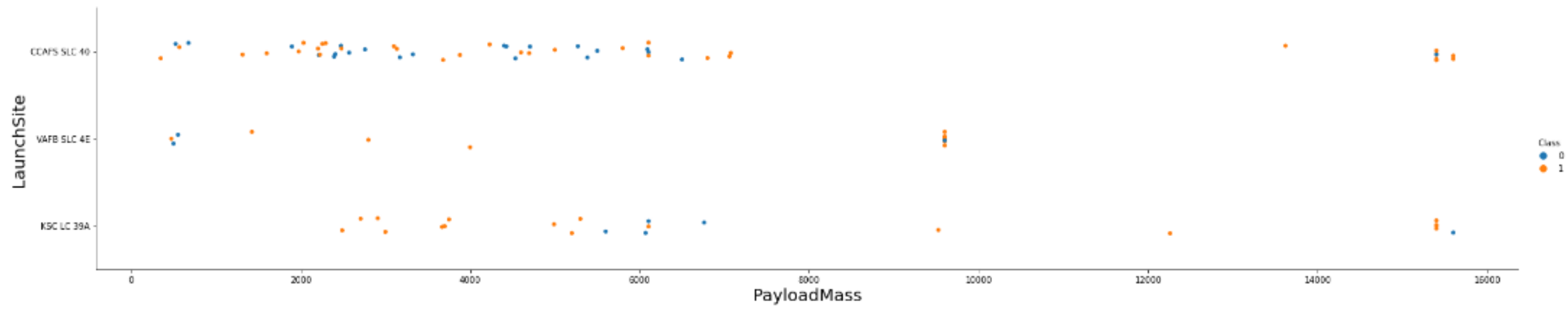

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



- Launch sites in the order of **their best performance** are as follows
  - 1. CCAF5 SLC 40
  - 2. VAFB SLC 4E
  - 3. KSC LC 39A

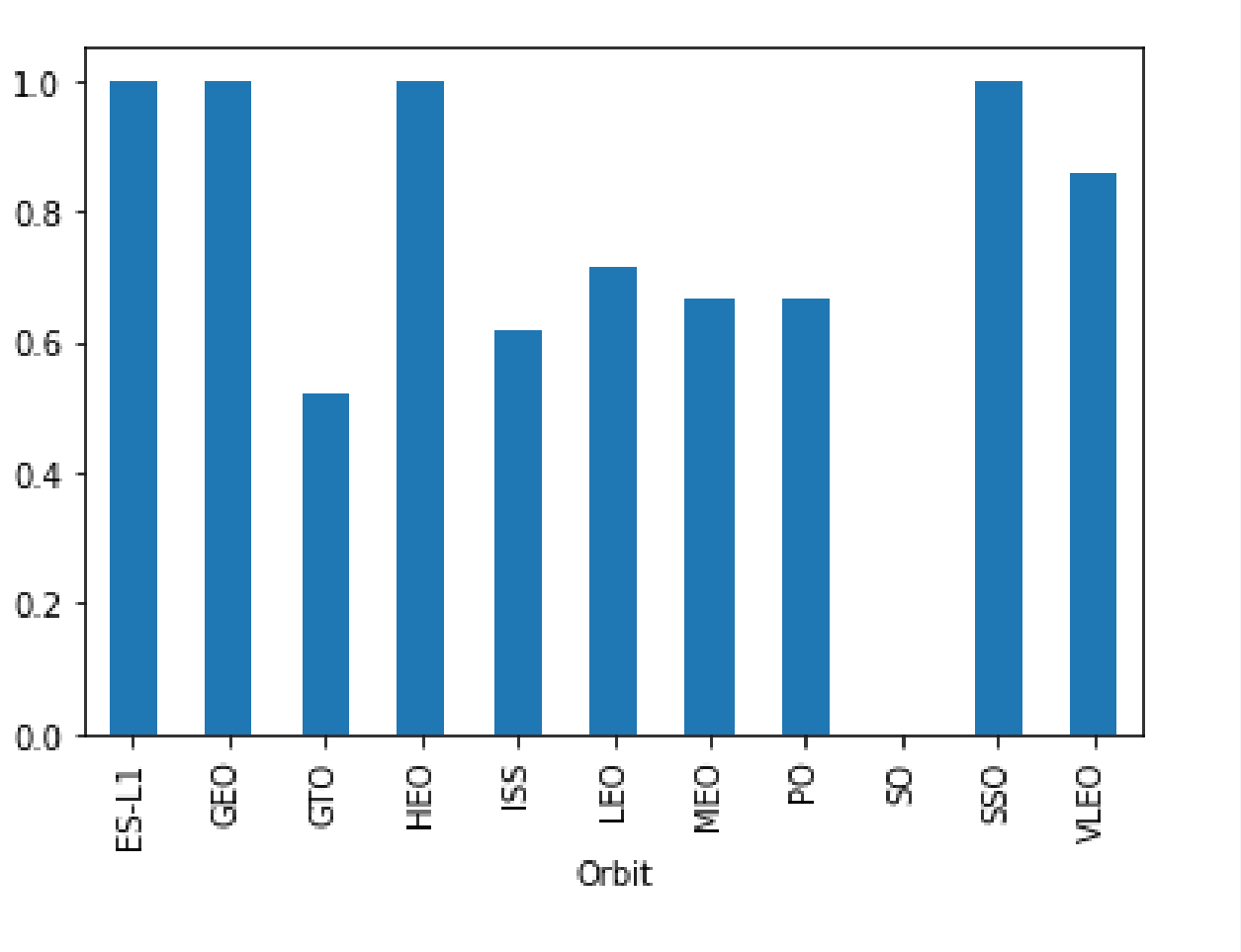
Payload vs. Launch Site



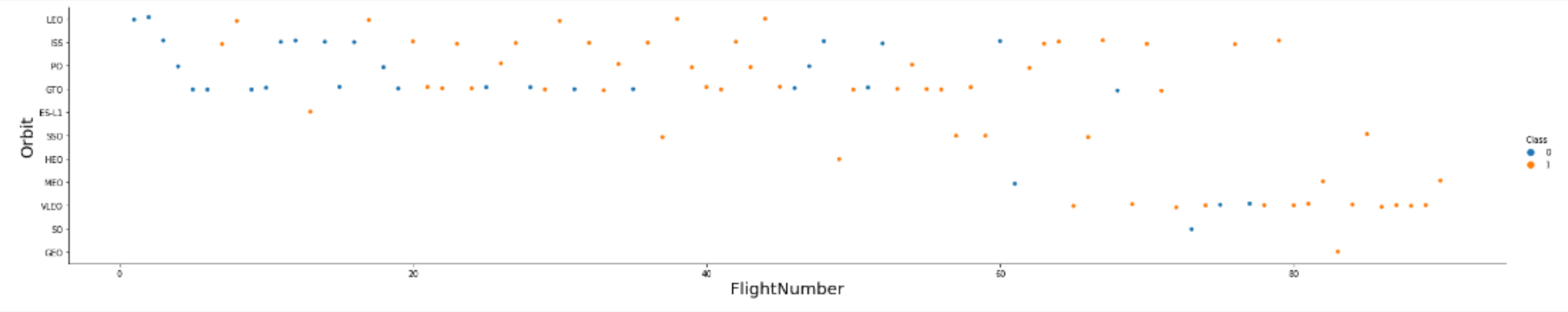
- for the VAFB-SLC launch site there are no rockets launched for heavy payload mass(greater than 10000).

Success Rate vs. Orbit Type

- Nearly 100% success rates happens to orbits:
- ES-L1;
- GEO;
- HEO; and
- SSO.
- Next success rates above 70%
- VLEO (above 80%); and
- LFO (above 70%).



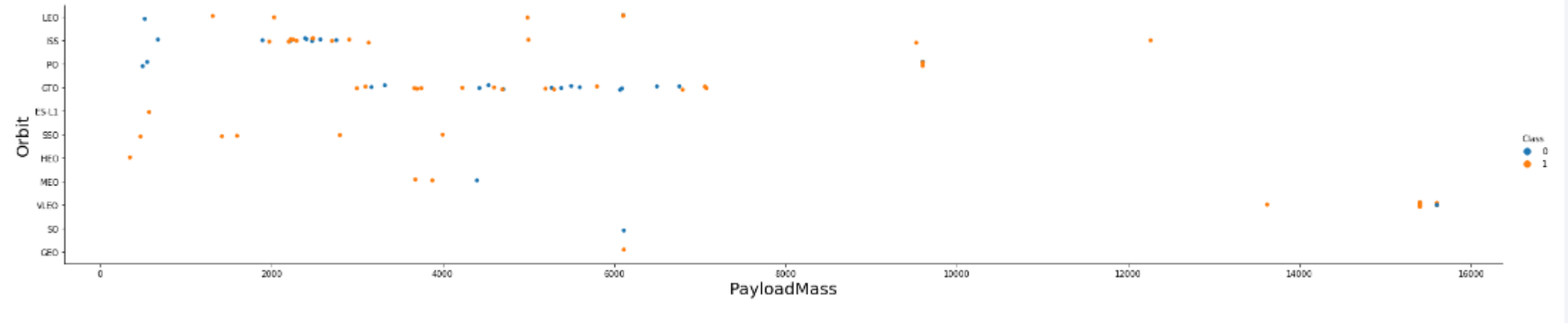
Flight Number vs. Orbit Type



- In the LEO orbit the Success appears related to the number of flights;

- on the other hand, there seems to be no relationship between flight number when in GTO orbit.

Payload vs. Orbit Type

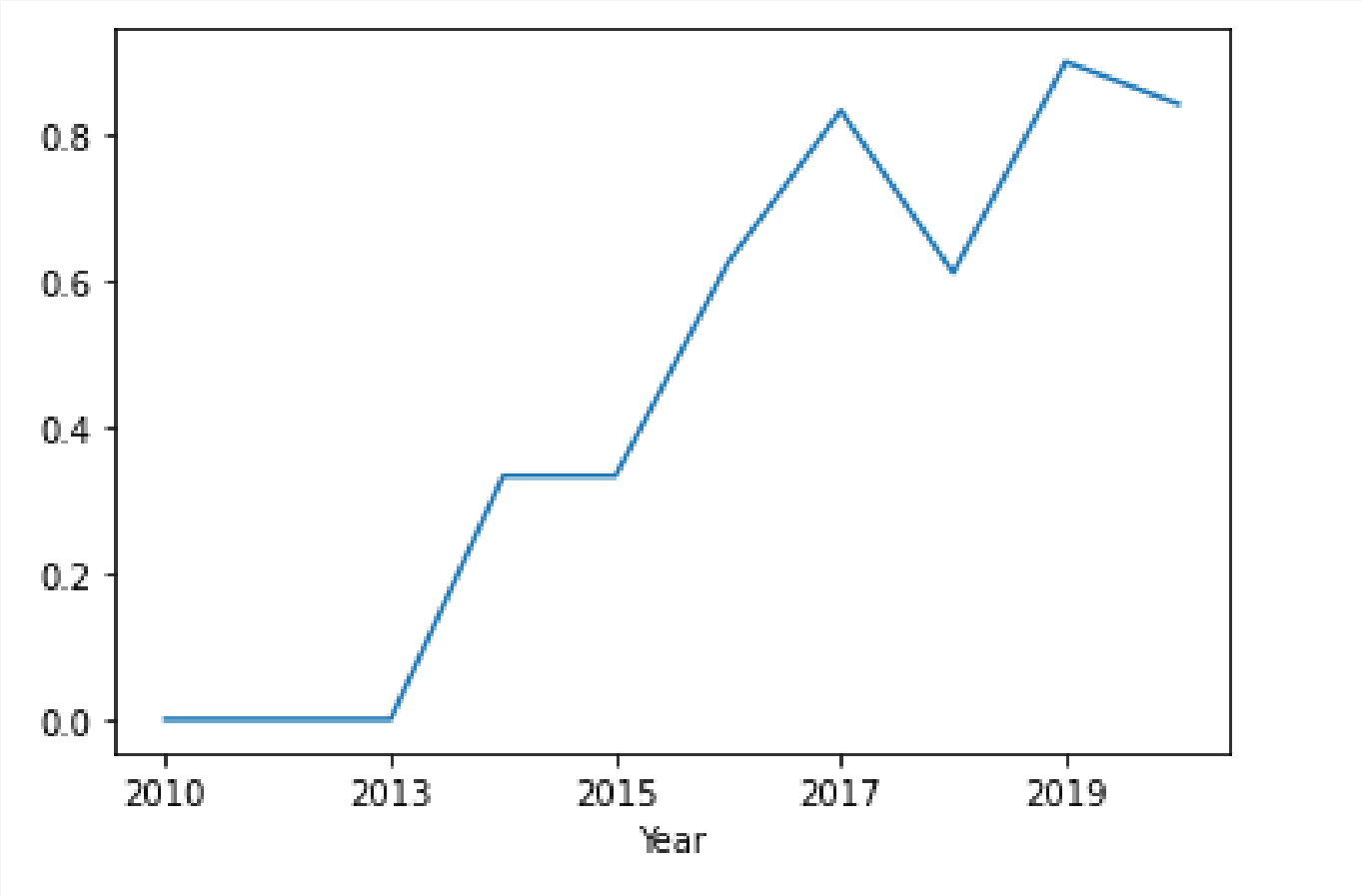


- With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.

- 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

- the success rate since 2013 kept increasing till 2020



All Launch Site Names

- Four unique launch sites are available in the given data

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

- The above result was achieved by using DISTINCT query for finding unique occurrences of “launch_site”

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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	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

Total Payload Mass

- the total payload carried by boosters from NASA

total_payload
111268

- Sum of all payloads mass carried by boosters launched by NASA (CRS)

Average Payload Mass by F9 v1.1

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

avg_payload
2928

- Data is filtered with the corresponding booster version and average payload mass was obtained

First Successful Ground Landing Date

- the dates of the first successful landing outcome on ground pad

first_success_gp
2015-12-22

- The data is filtered by successful landing outcome on ground pad using Min function on date

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

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

Total Number of Successful and Failure Mission Outcomes

- the total number of successful and failure mission outcomes

mission_outcome	qty
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

- Mission outcomes are grouped and each group's records are counted

Boosters Carried 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

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

2015 Launch Records

- List of the failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015

booster_version	launch_site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

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

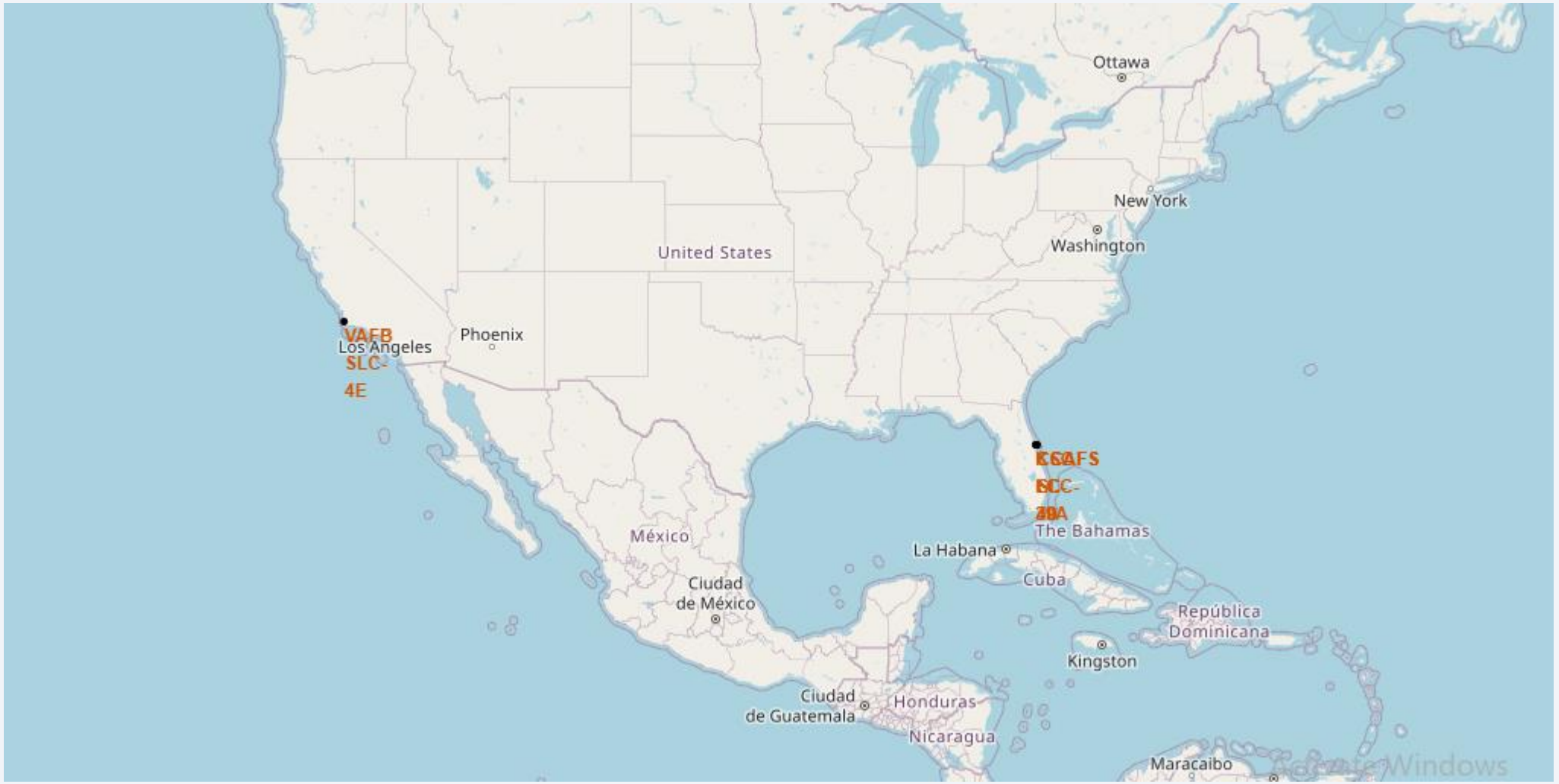
- Ranks of the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

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

Section 3

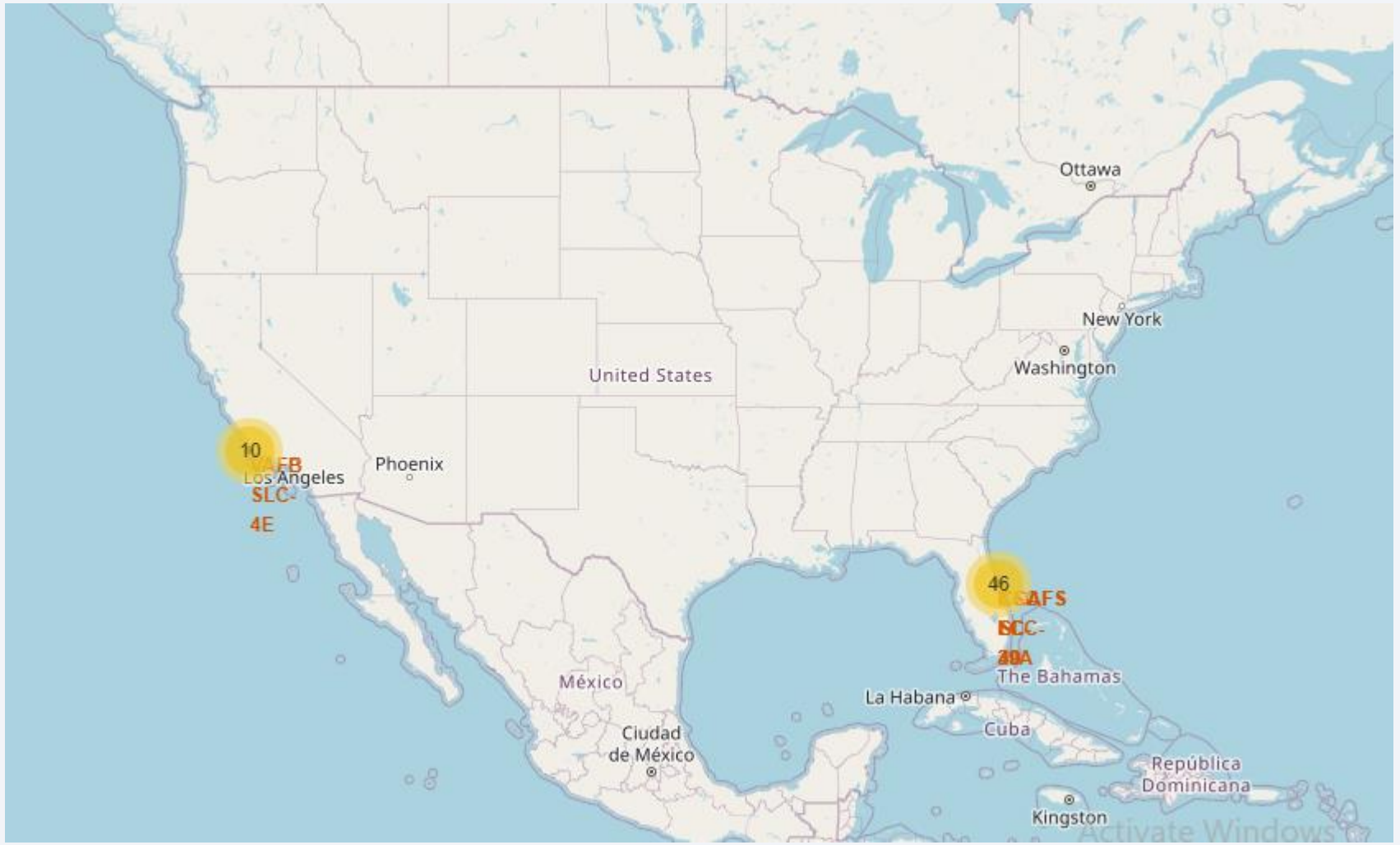
Launch Sites Proximities Analysis

Launch Sites



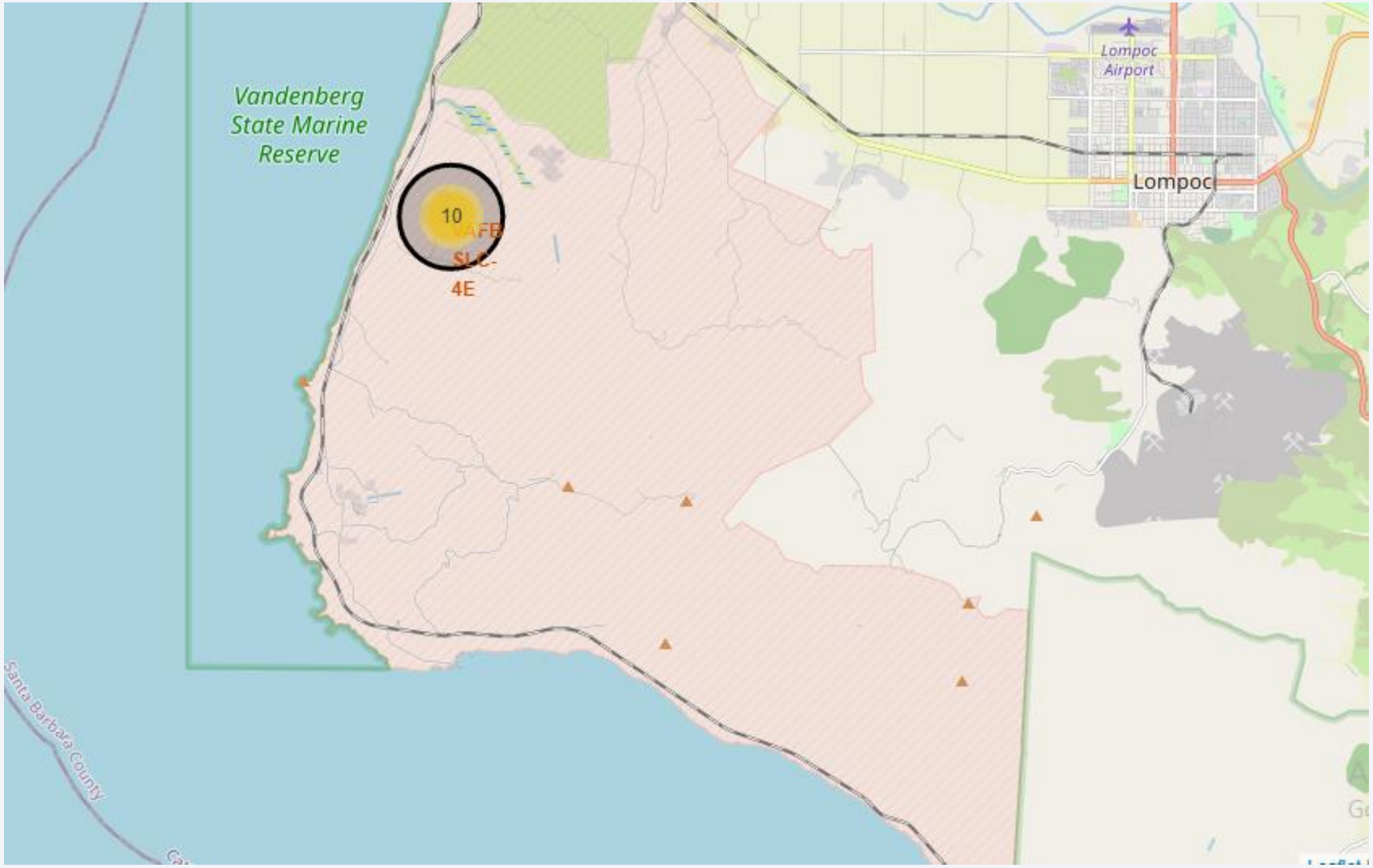
launch_site
CCAFS LC-40
CCAFS SLC-40
KSC LC-39A
VAFB SLC-4E

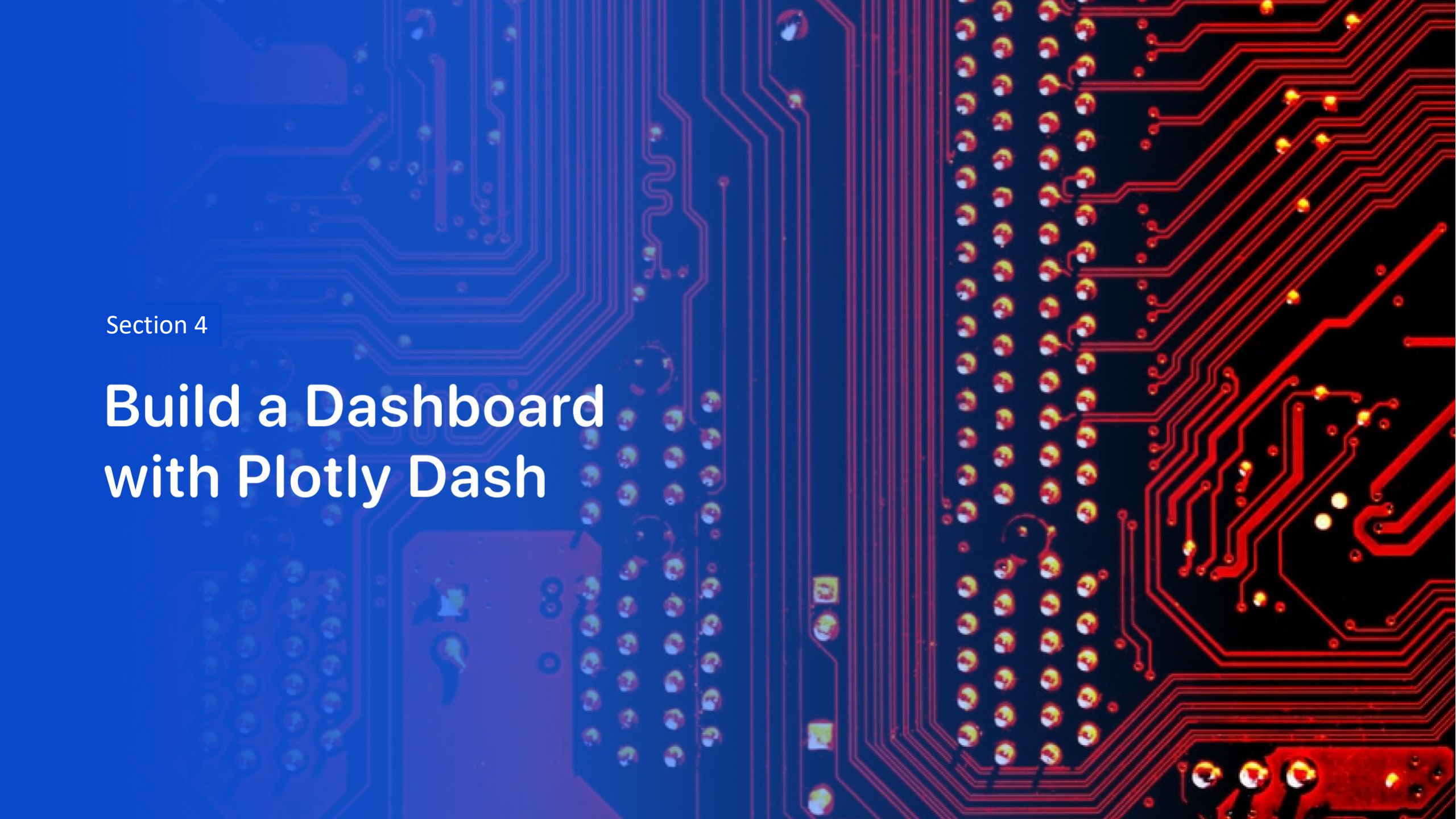
Launch Outcomes



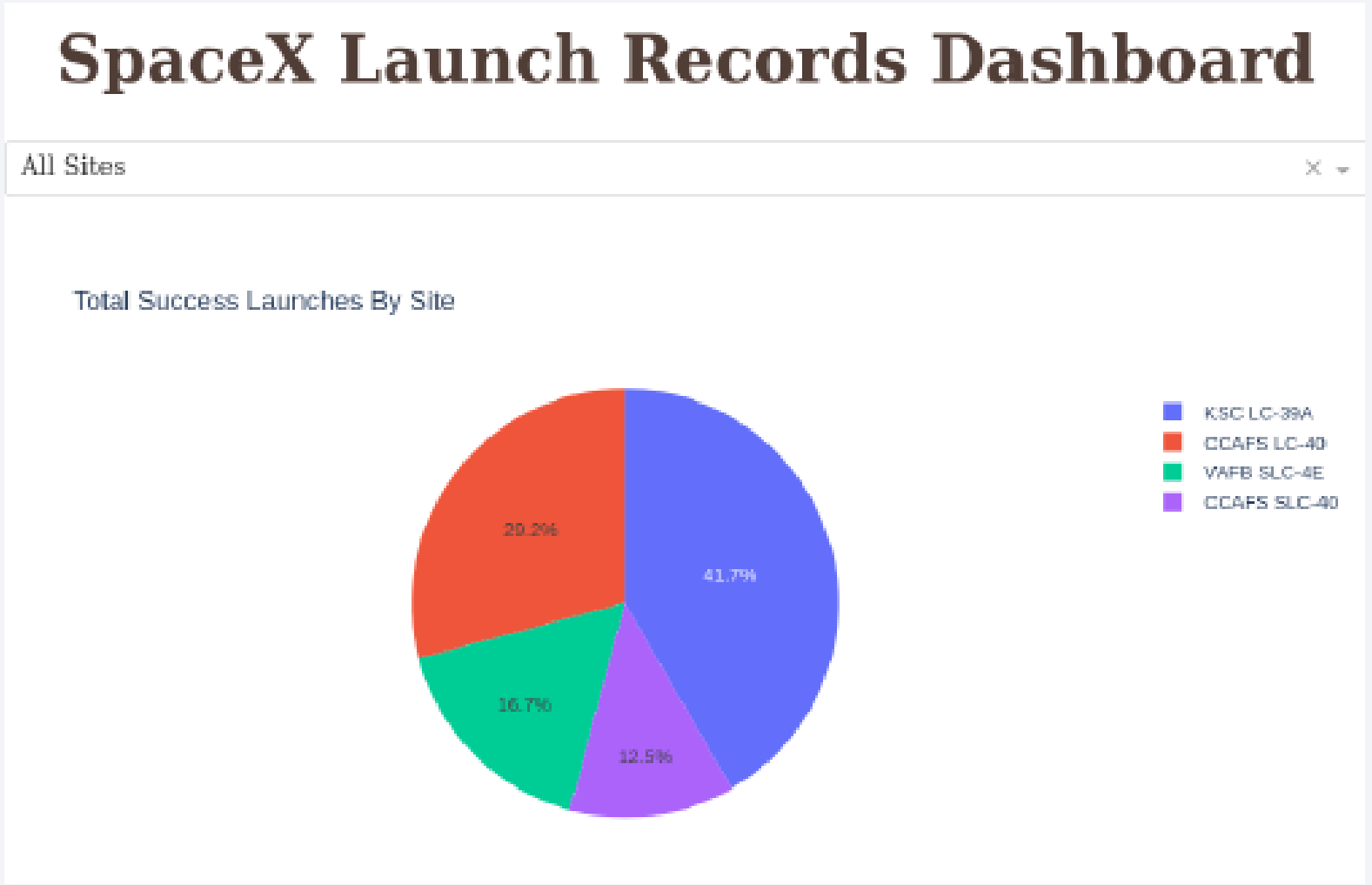
Logistics and Safety

- 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

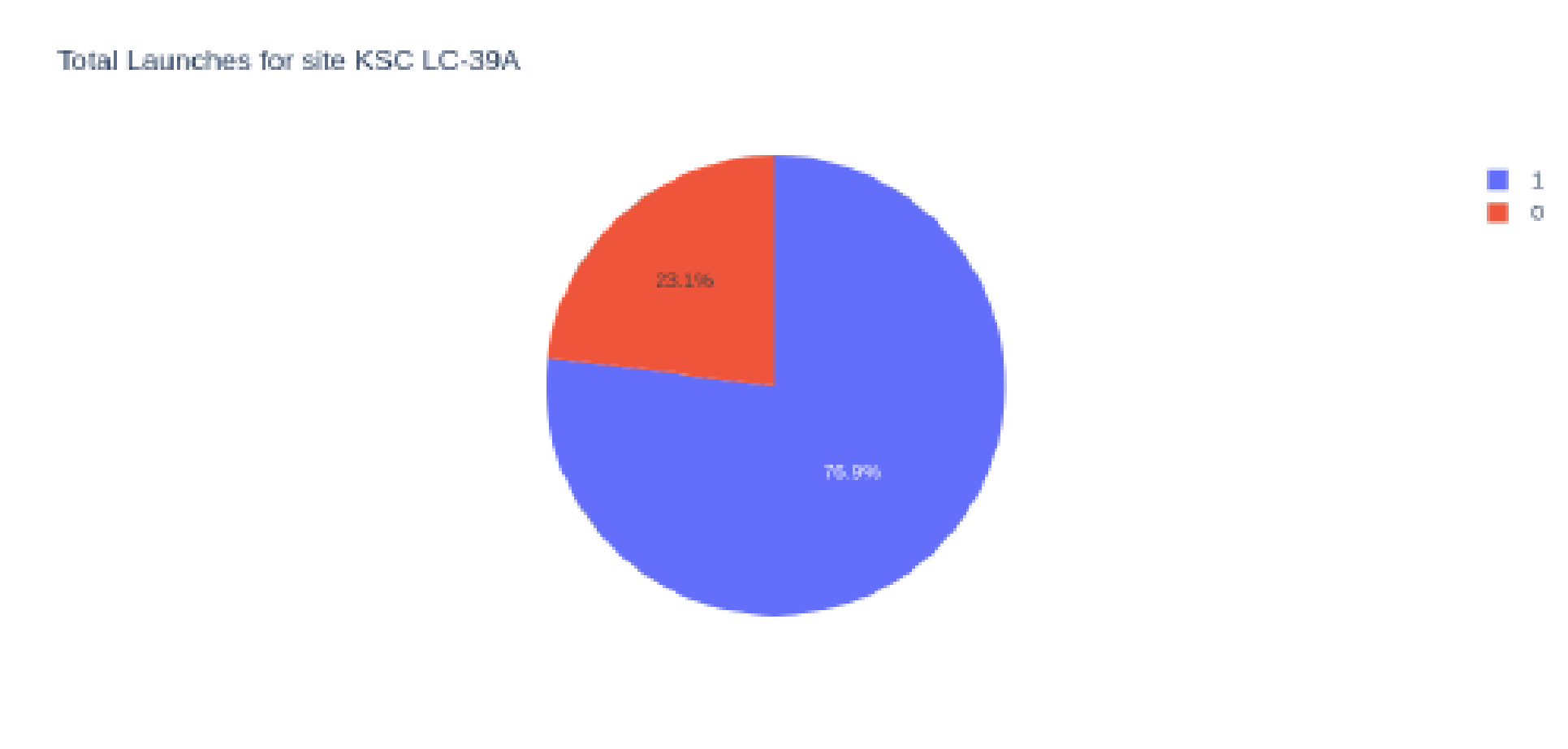



Section 4
Build a Dashboard with Plotly Dash

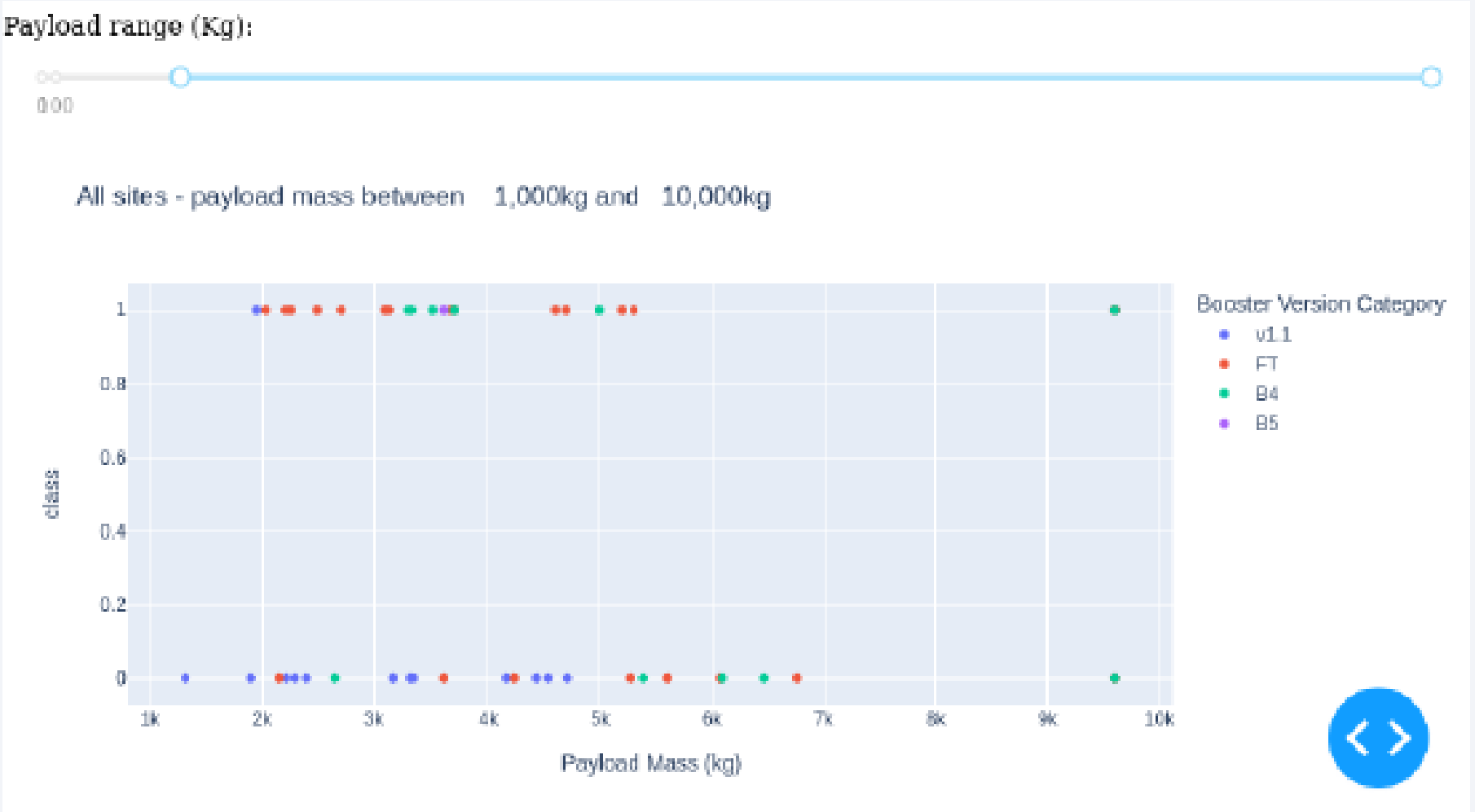
Successful Launches by Site



Launch site with highest success ratio



Payload vs. Launch Outcome scatter plot for all sites



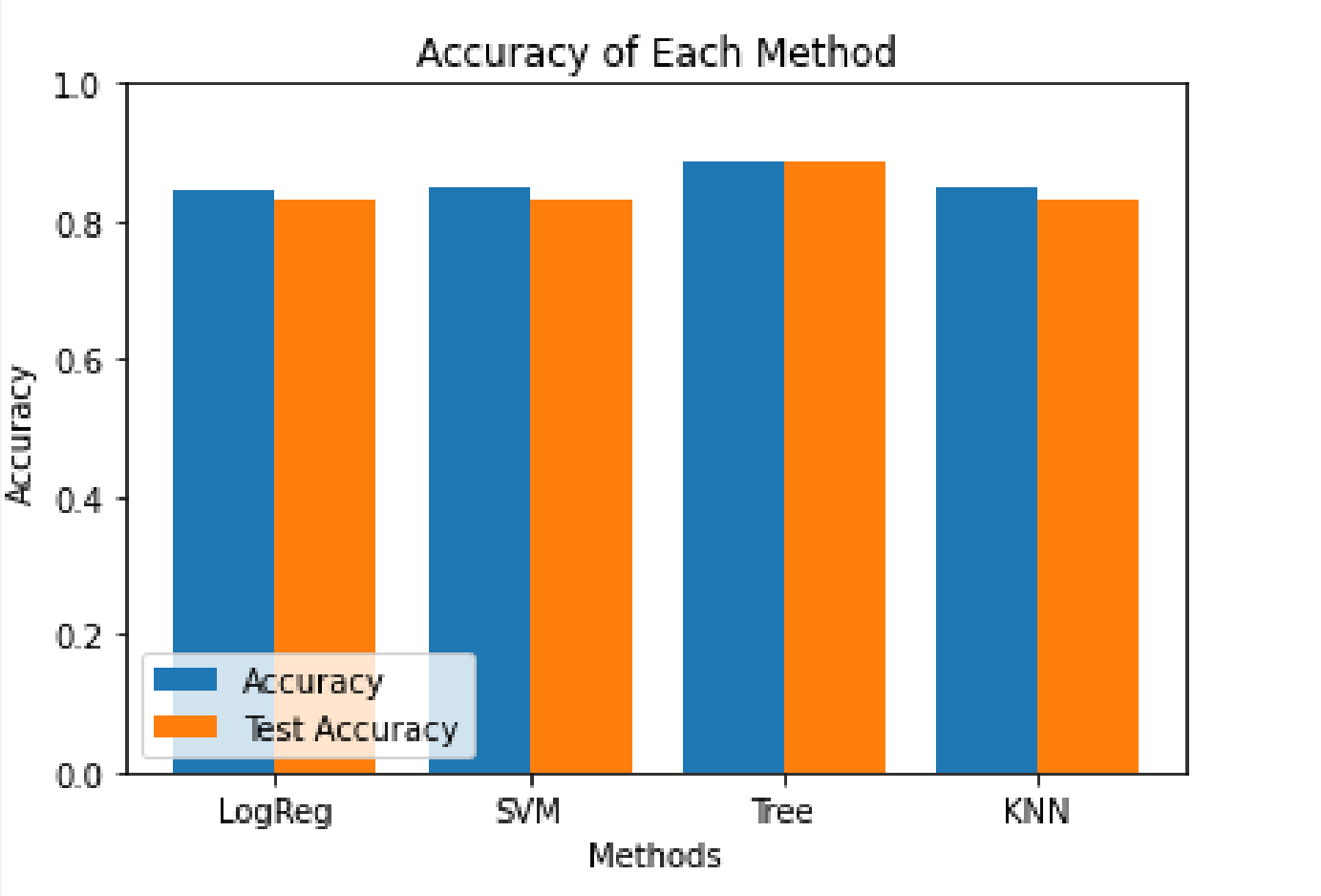
Section 5

Predictive Analysis (Classification)

Classification Accuracy

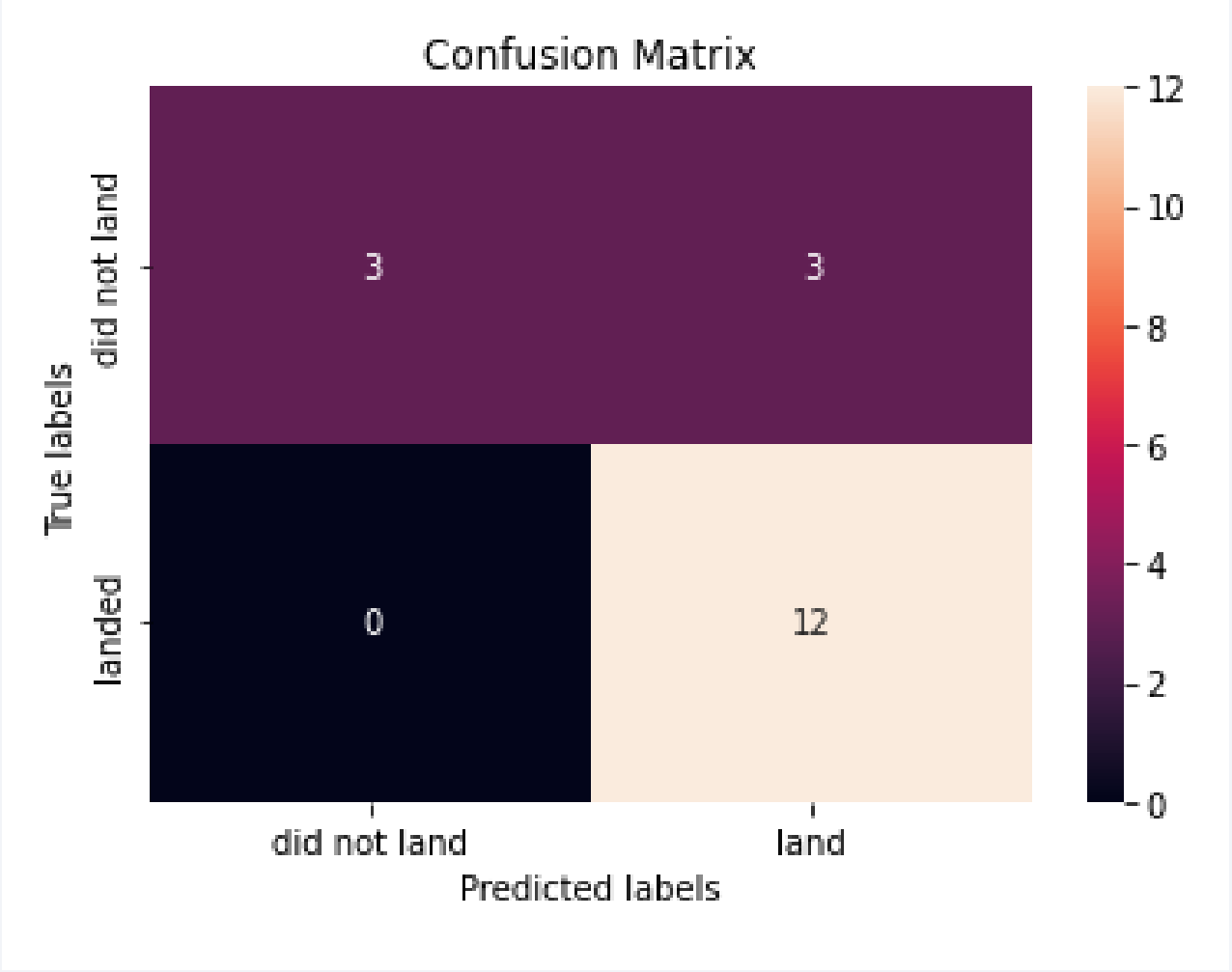
- Predictive analysis results

Decision Tree Model was found to be best model with 88.75% Accuracy and 88.89% Test accuracy



Confusion Matrix: Best performing model

Decision Tree Model was found to be best model with 88.75% Accuracy and 88.89% Test accuracy

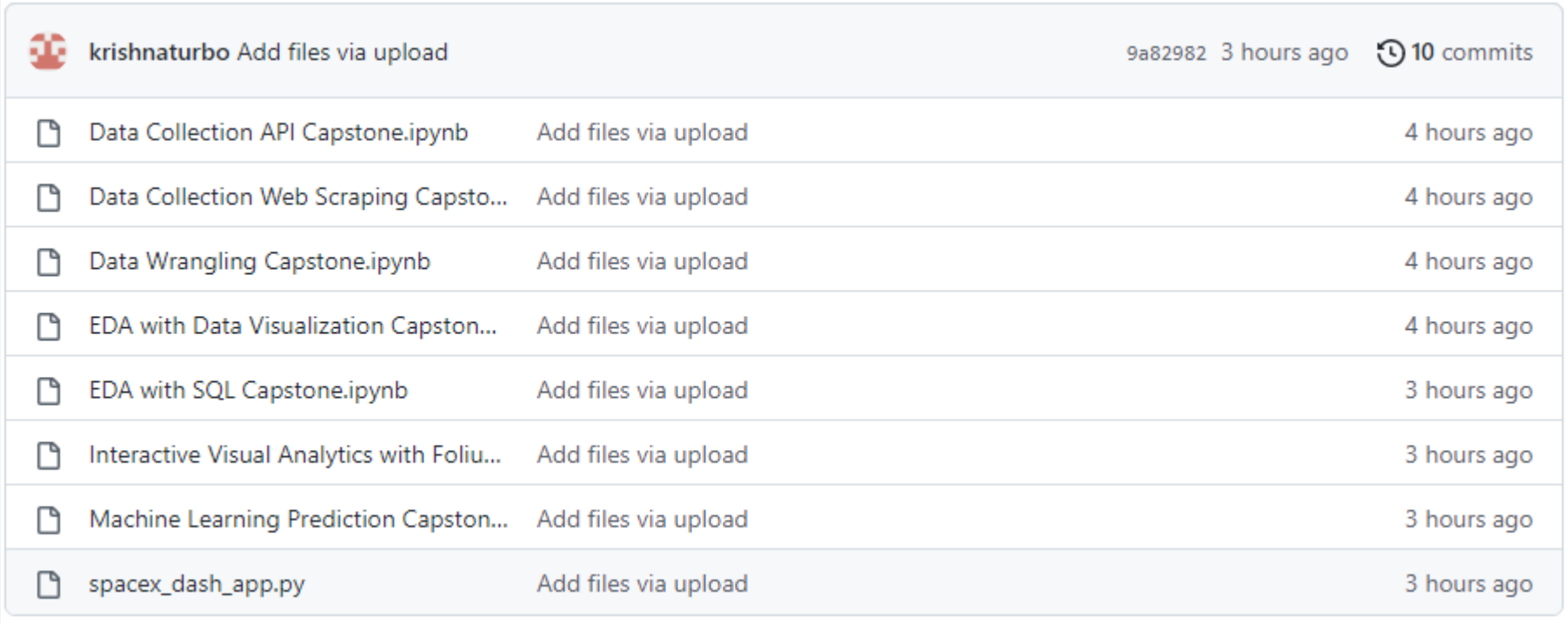


Conclusions

- Analysis of Various data sources were performed
- KSC LC – 39A was found to be best launch site
- Maximum payload – critical value is found to be 7000Kg
- Successful outcomes are observed
- Success has been improved over time
- Decision Tree classifier is found to be best machine learning model to predict the successful landings and consequently enhance the profits

Appendix

- The executed files along with code cells are attached at
- https://github.com/krishnaturbo/Capstone-IBM-Data-Scientist



krishnaturbo Add files via upload — 9a82982 3 hours ago — 10 commits

Data Collection API Capstone.ipynb	Add files via upload	4 hours ago
Data Collection Web Scraping Capsto...	Add files via upload	4 hours ago
Data Wrangling Capstone.ipynb	Add files via upload	4 hours ago
EDA with Data Visualization Capston...	Add files via upload	4 hours ago
EDA with SQL Capstone.ipynb	Add files via upload	3 hours ago
Interactive Visual Analytics with Foliu...	Add files via upload	3 hours ago
Machine Learning Prediction Capston...	Add files via upload	3 hours ago
spacex_dash_app.py	Add files via upload	3 hours ago

Thank you!