IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Saksham
Chauhan

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection
  - Data Wrangling
  - EDA with Data Visualization
  - EDA with SQL
  - Building an interactive map with Folium
  - Building a dashboard with Plotly Dash
  - Predictive Analytics - Classification
- Summary of all results
  - EDA Results
  - Interactive Analysis/Visualization
  - Predictive Analytics

Introduction

- Project background and context
  - Spacex advertises it's Falcon 9 rocket launches, with a cost of $62mn with other providers costing upwards of $165mn for each launch. Much of these savings are due to SpaceX's capability of reusing first stage launches
- Problems you want to find answers
  - Predicting success of first stage landing of the SpaceX Falcon 9 rocket

Section 1

Methodology

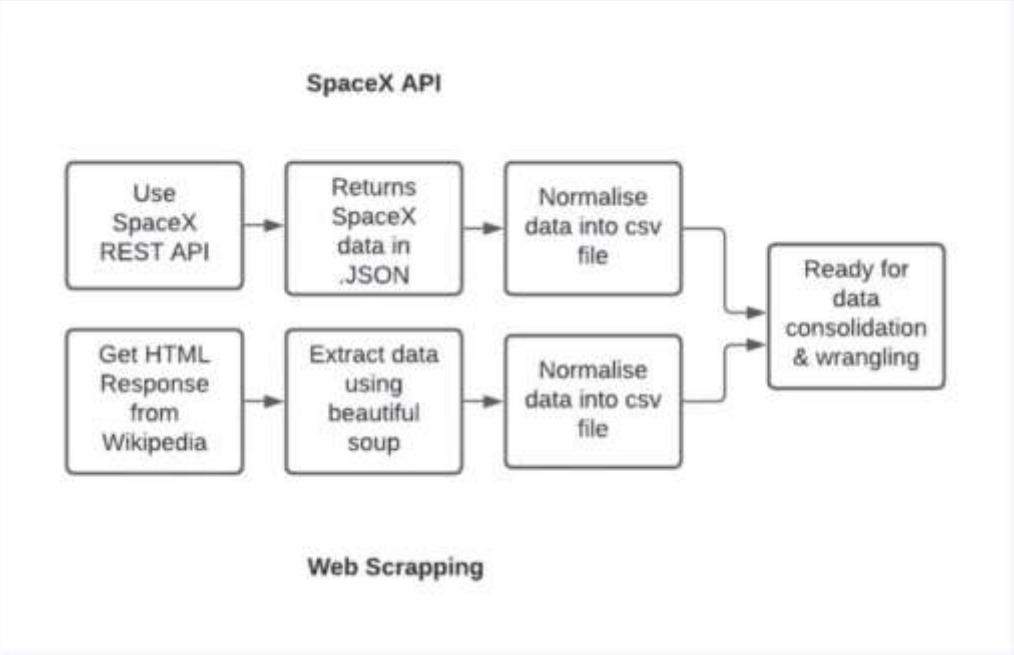
Methodology

Executive Summary

- Data collection methodology:
  - SpaceX Rest API
  - Web Scrapping from Wikipedia
- Perform data wrangling
  - Data cleaning and One Hot Encoding of data fields to prepare data for analysis and predictive modeling
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Logistic Regression, KNN, SVM and Decision Tree models used for predictions

Data Collection

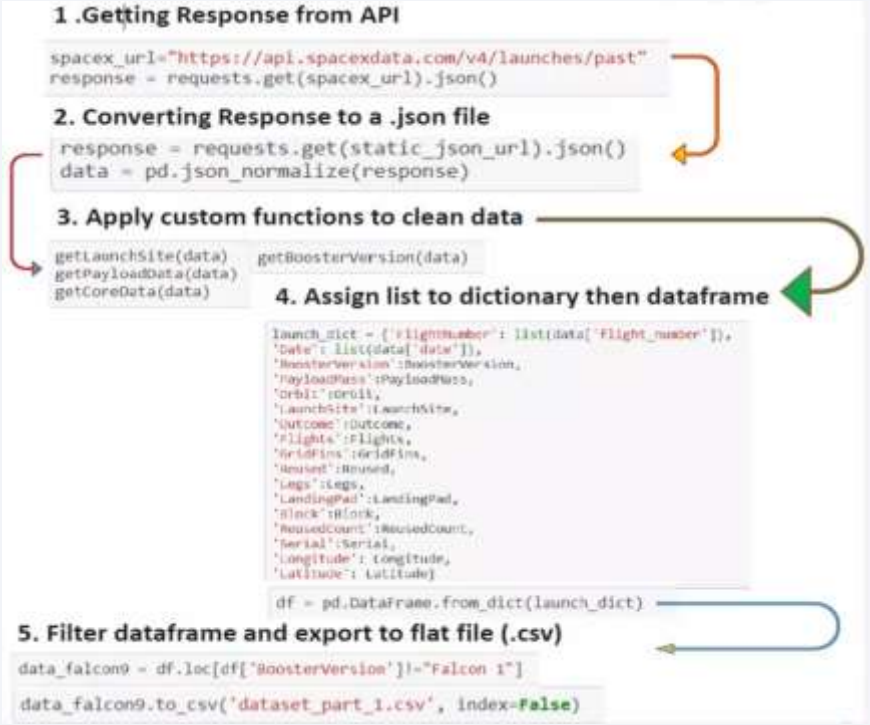
- The following datasets were collected
  - SpaceX launch data collected from the SpaceX REST API
  - Data provides info on rockets used, payload delivered, launch specifications, landing specifications and landing outcomes
  - Data was also gathered from Wikipedia for the Falcon 9 launches through webscrapping using BeautifulSoup



Data Collection - SpaceX API

- Data collection with SpaceX Rest API

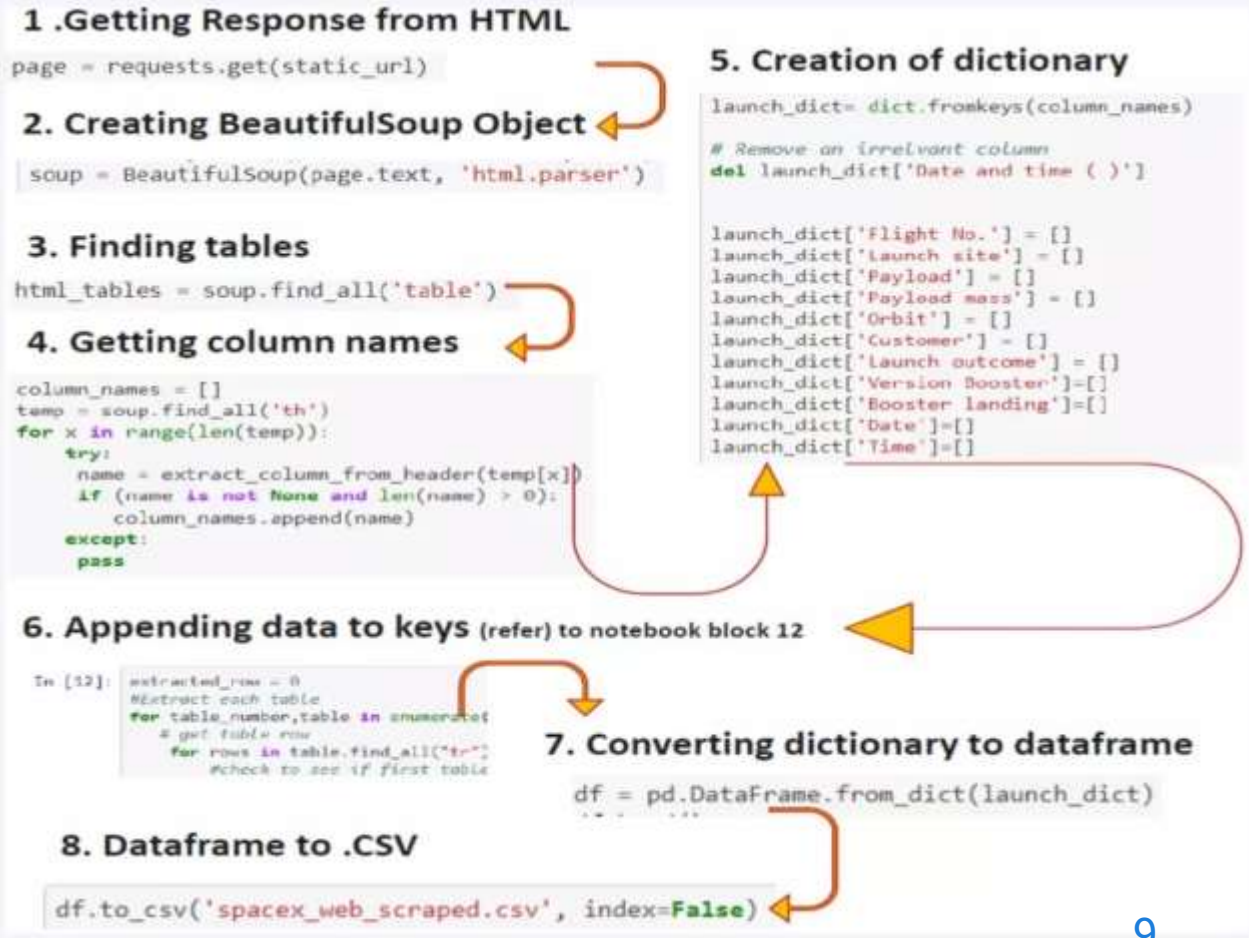
- Github link : https://github.com/rkohli0010/SpaceX-Capstone-Project/blob/803bd7ecc229f17454bb3174342534ac92197e05/SpaceX%20Part%201%20-%20Data%20Collection%20with%20REST%20API.ipynb



Data Collection - Scraping

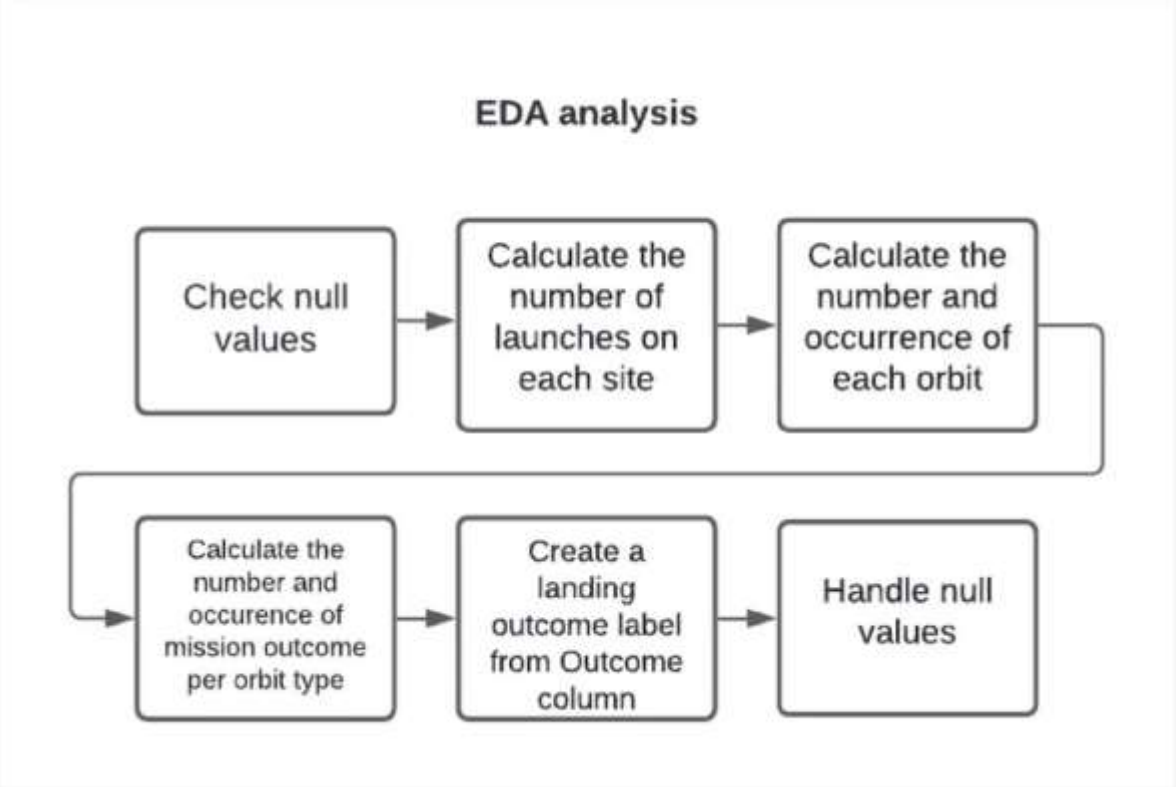
- Webscrapping from Wikipedia

- Github Link : https://github.com/rkohli0010/SpaceX-Capstone-Project/blob/f671f15dcfe5cf85ba193acb42ec570bd3b2c99a/SpaceX%20Part%202%20-%20Webscrapping%20with%20Beautifulsoup

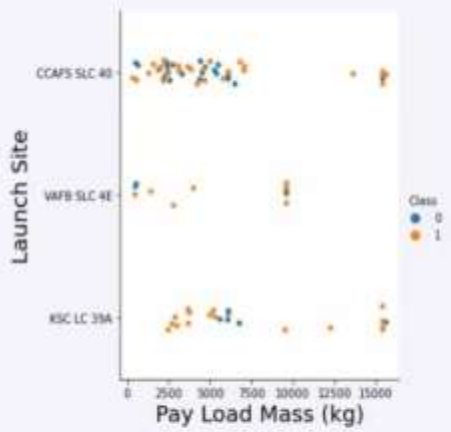


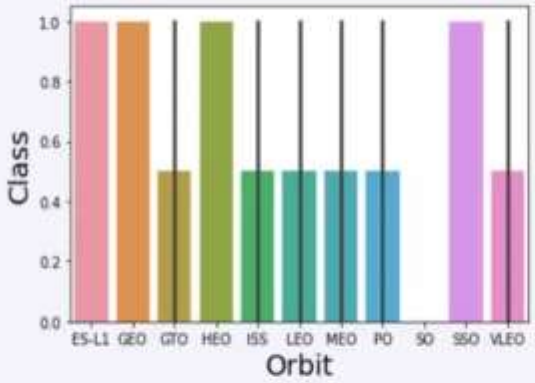
Data Wrangling

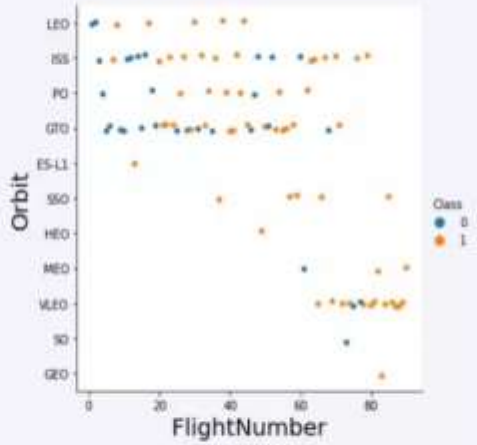
- Github Link - https://github.com/rkohli0010/SpaceX-Capstone-Project/blob/1541d26c94dee91da839fd801ec533a0b8077415/Spacex%20Part%203%20:%20Data%20Wrangling.ipynb

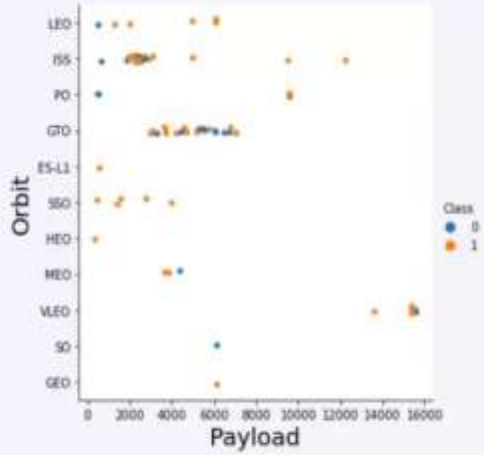


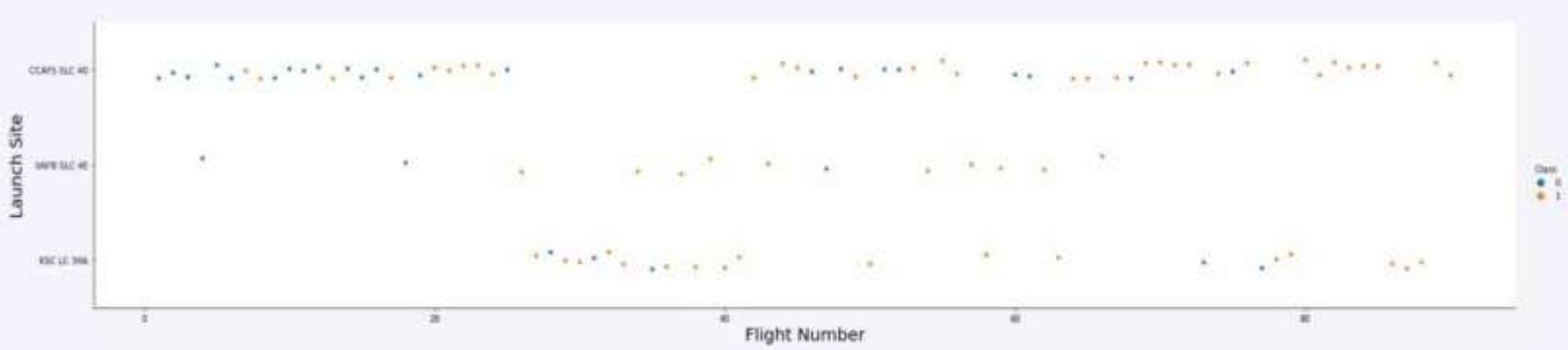
EDA with Data Visualization

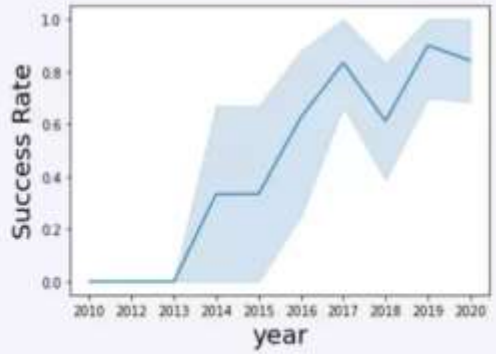












- Github Link - https://github.com/rkohli0010/SpaceX-Capstone-Project/blob/b446aa440e39ffd358c18a657bc9a233e6c10529/Spacex%20Part%204%20-%20EDA%20and%20Data%20Visualization.ipynb

EDA with SQL

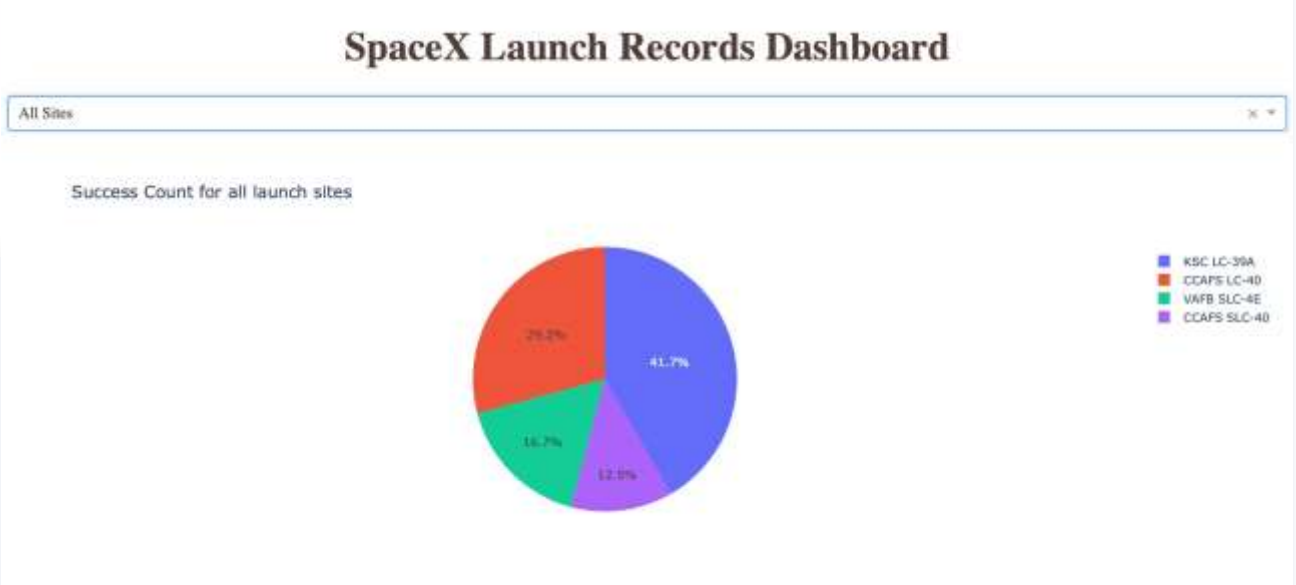
- SQL Queries performed :
  - Displaying names of unique launch sites
  - Displaying 5 records where launch site begins with ‘KSC’
  - Displaying the total payload mass carried by boosters launched by NASA(CRS)
  - Displaying average payload mass carried by booster version F9 v1.1
  - Listing the date where the successful landing outcome in drone ship was achieved
  - Listing names of boosters having success on ground pad and payload mass between 4000 and 6000
  - Listing total number of successful and failure mission outcomes
  - Listing booster versions with maximum payload mass
  - Listing records displaying month, successful landing outcomes in ground pad and booster
  - Versions and launch sites for months in year 2017
  - Ranking count of successful landing outcomes between 206/04/2010 and 03/20/2017 in descending order

  - Github Link - https://github.com/rkohli0010/SpaceX-Capstone-Project/blob/fe5fa90d3ea5471052a3fa977e7bc350031374af/SpaceX%20Part%205%20-%20EDA%20with%20SQL.ipynb

Build an Interactive Map with Folium

- Github link - https://github.com/rkohli0010/SpaceX-Capstone-Project/blob/340f49f692b8b78cc1bcb1d4914e1beb9681b0a9/SpaceX%20Part%206%20-%20Visualization%20with%20maps-2.ipynb

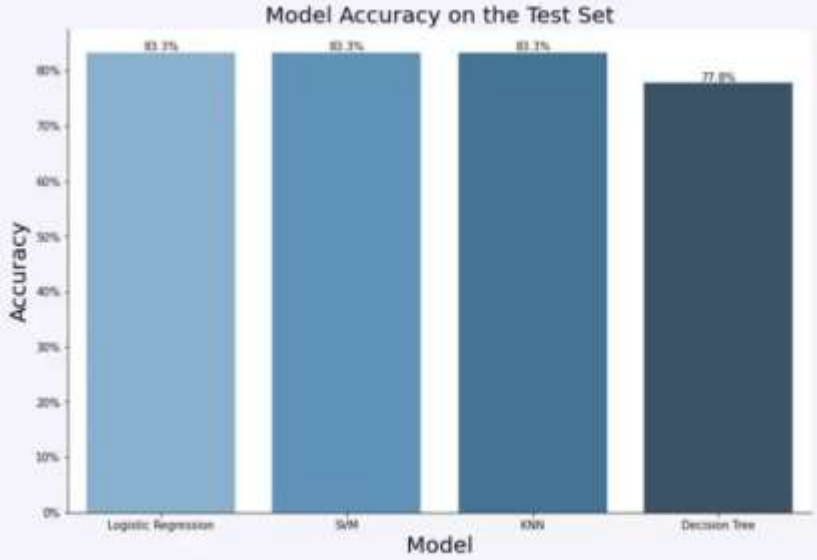
Build a Dashboard with Plotly Dash

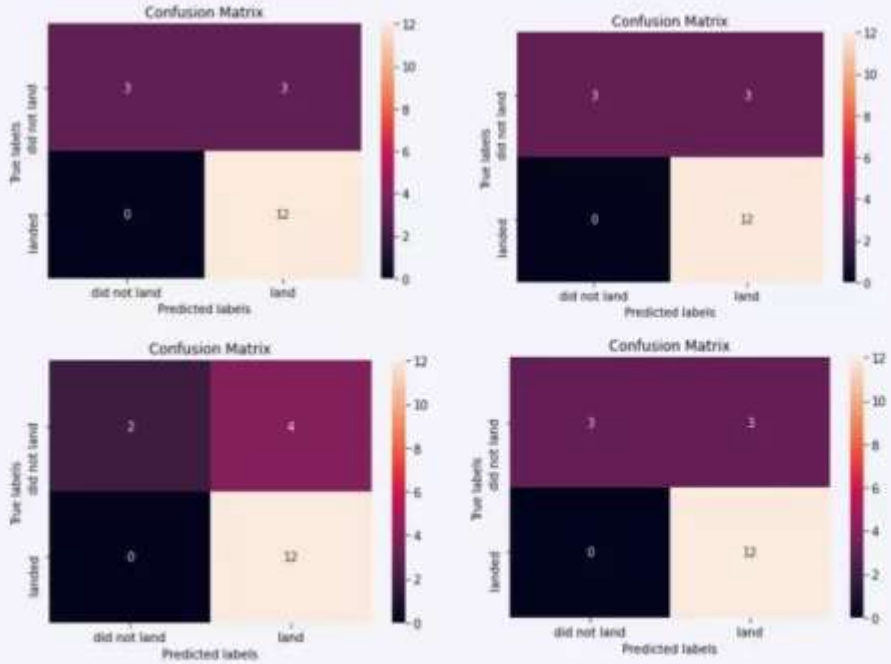


- Github link - https://github.com/rkohli0010/SpaceX-Capstone-Project/blob/23a7dd01ac3ba42bdab6db1283a5002cbf411d14/SpaceX%20Part%207%20-%20Dash%20App

Predictive Analysis (Classification)

- The SVM, KNN, and Logistic Regression model achieved the highest accuracy at 83.3%, while the SVM performs the best in terms of Area Under the Curve at 0.958.

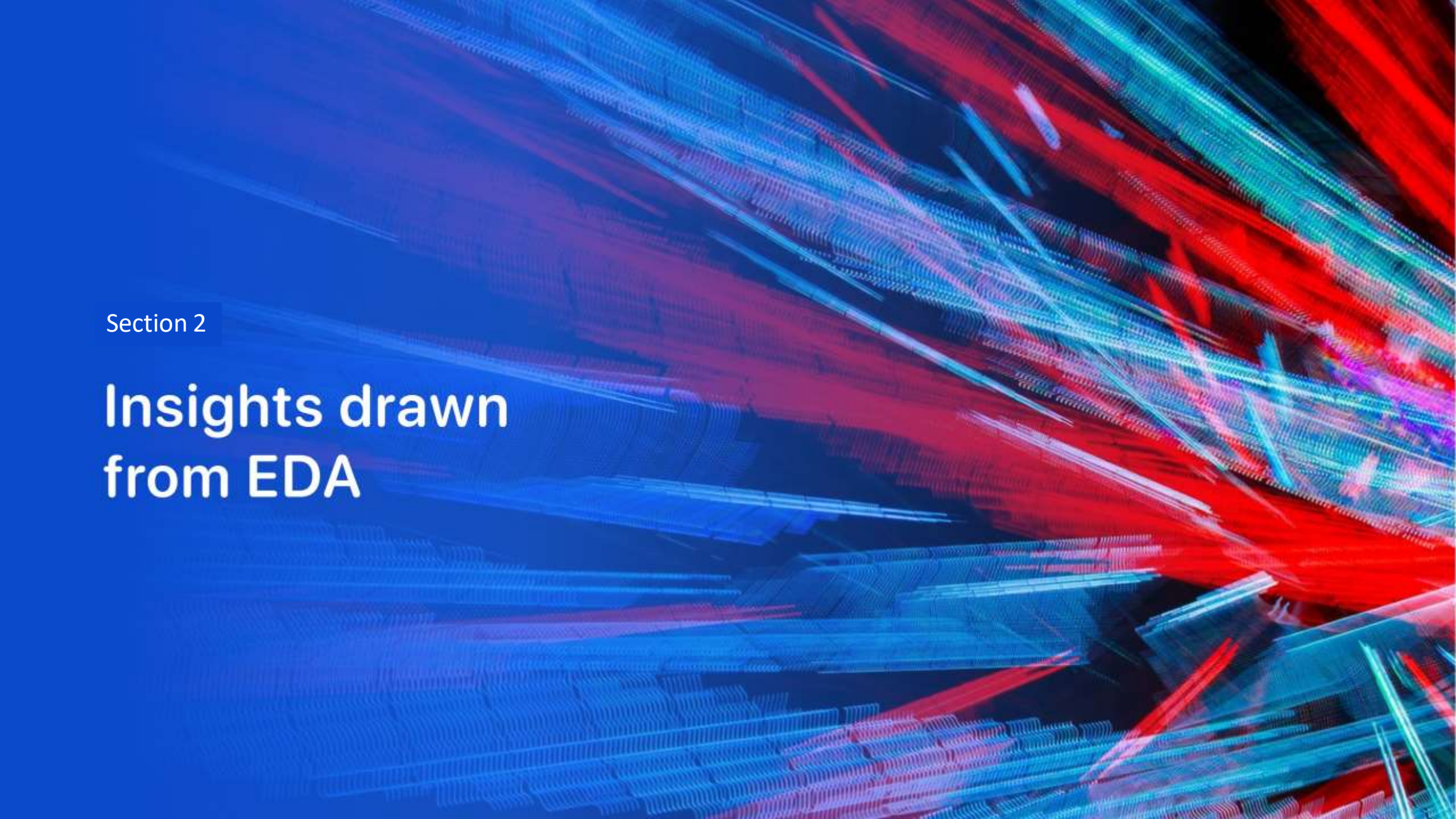




Github Link : https://github.com/rkohli0010/SpaceX-Capstone-Project/blob/a2300ec7cbe5a166d2074541b1831fcd3c272d55/SpaceX%20Part%208%20-%20Machine%20Learning-2.ipynb

Results

- Low weighted payloads perform better than heavier payloads
- Success rates for SpaceX launches are directly proportional to time in years
- KSC LC 39A had the most successful launches from all sites
- Orbit GEO, HEO, SSO, and ES L1 had the best success rates

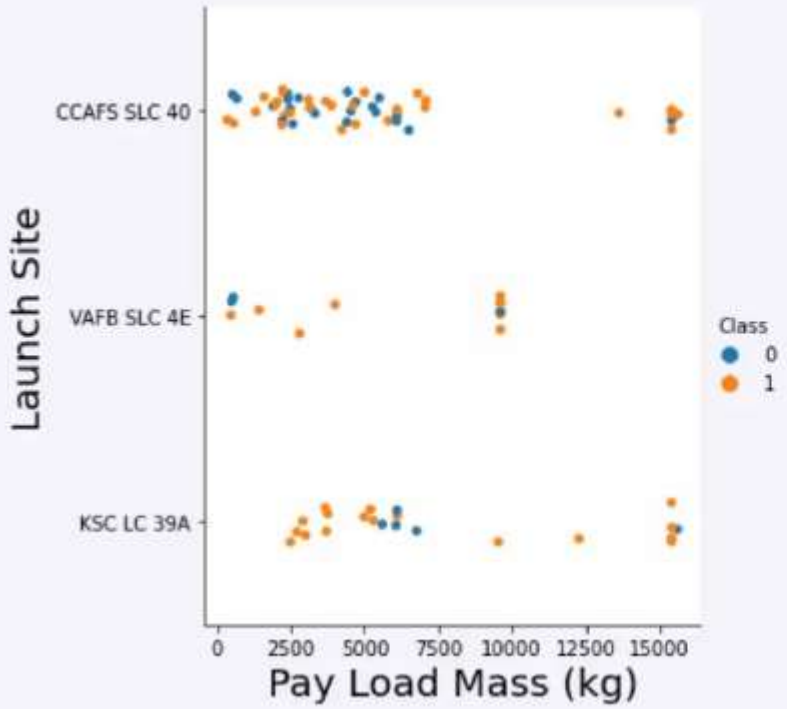

Section 2
Insights drawn
from EDA

Flight Number vs. Launch Site

- Launches from the site of CCAFS SLC 40 are significantly higher than launches form other sites.

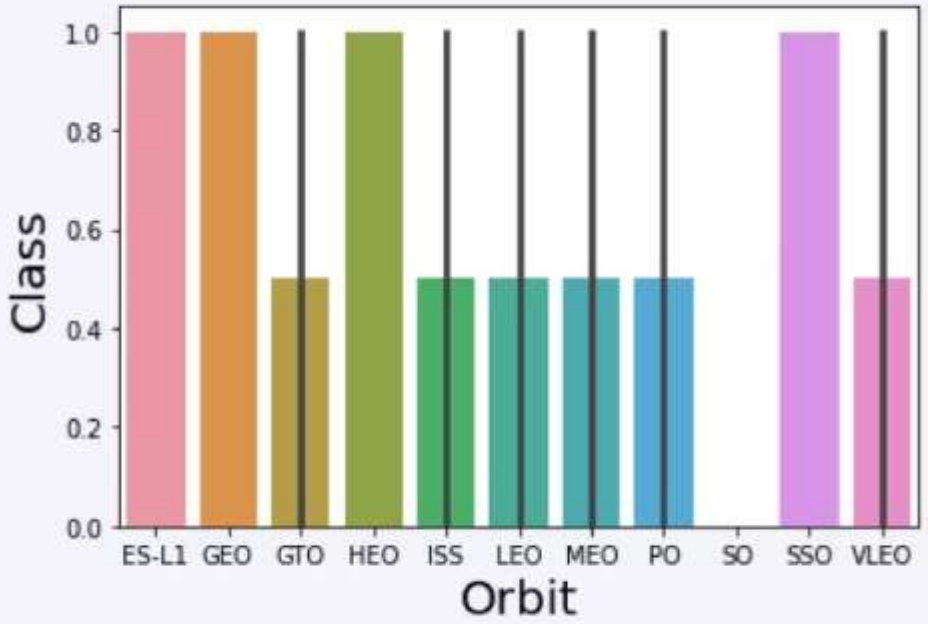
18

Payload vs. Launch Site



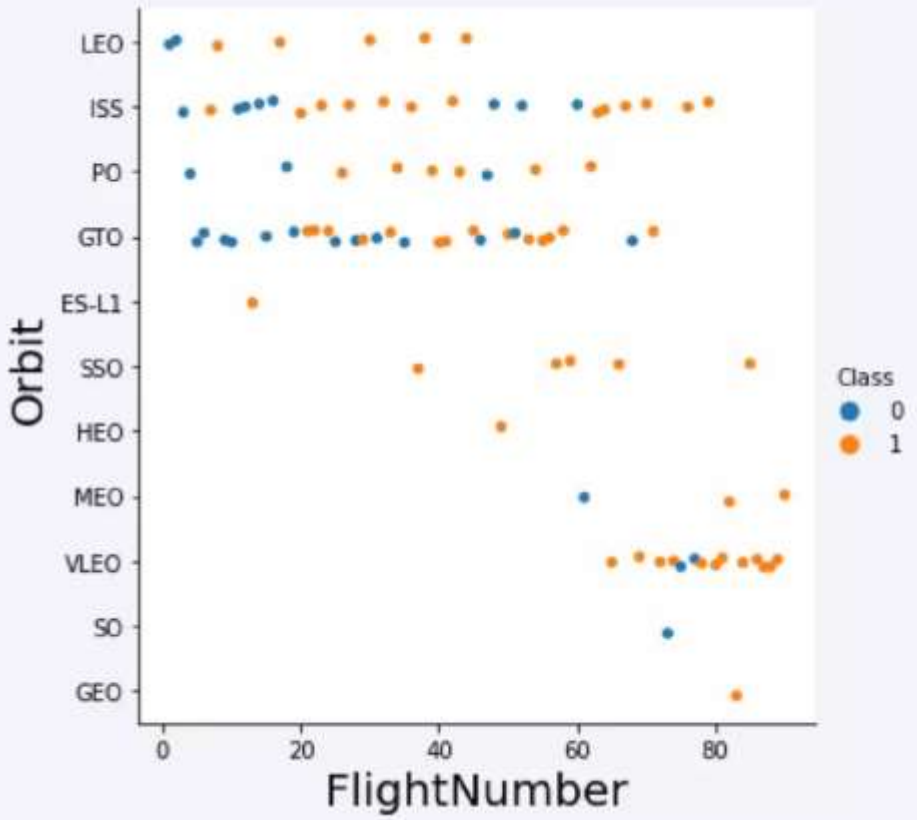
- The majority of lPay Loads with lower Mass have been launched from CCAFS SLC 40.

Success Rate vs. Orbit Type



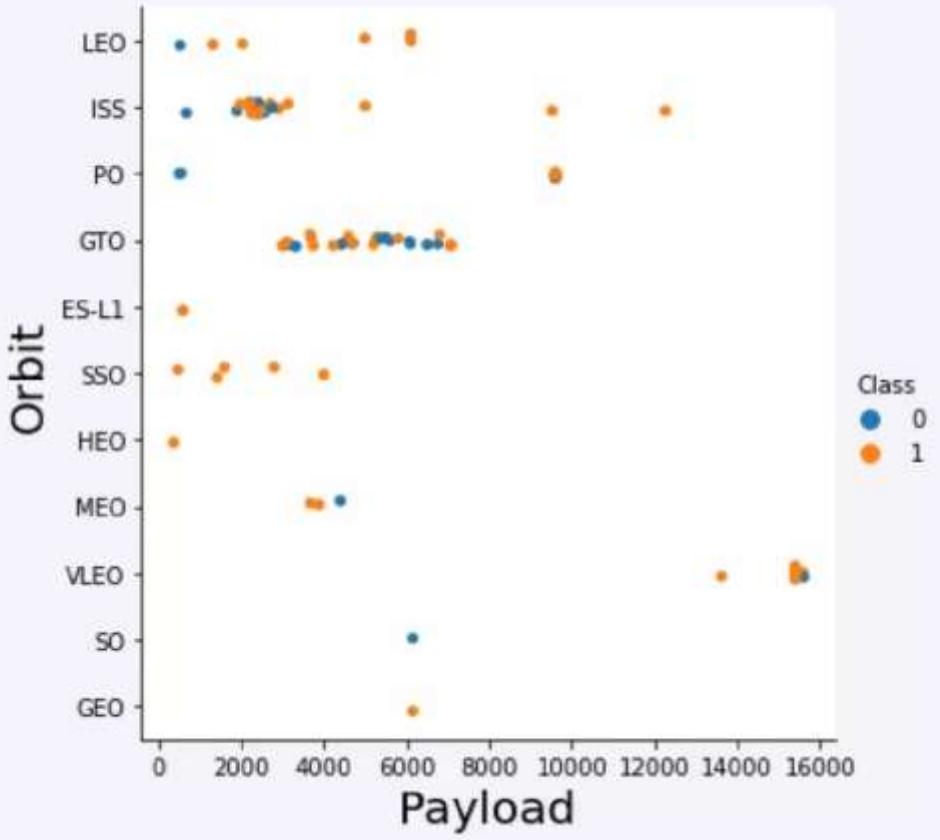
- The orbit types of ES-L1, GEO, HEO, SSO are among the highest success rate.

Flight Number vs. Orbit Type



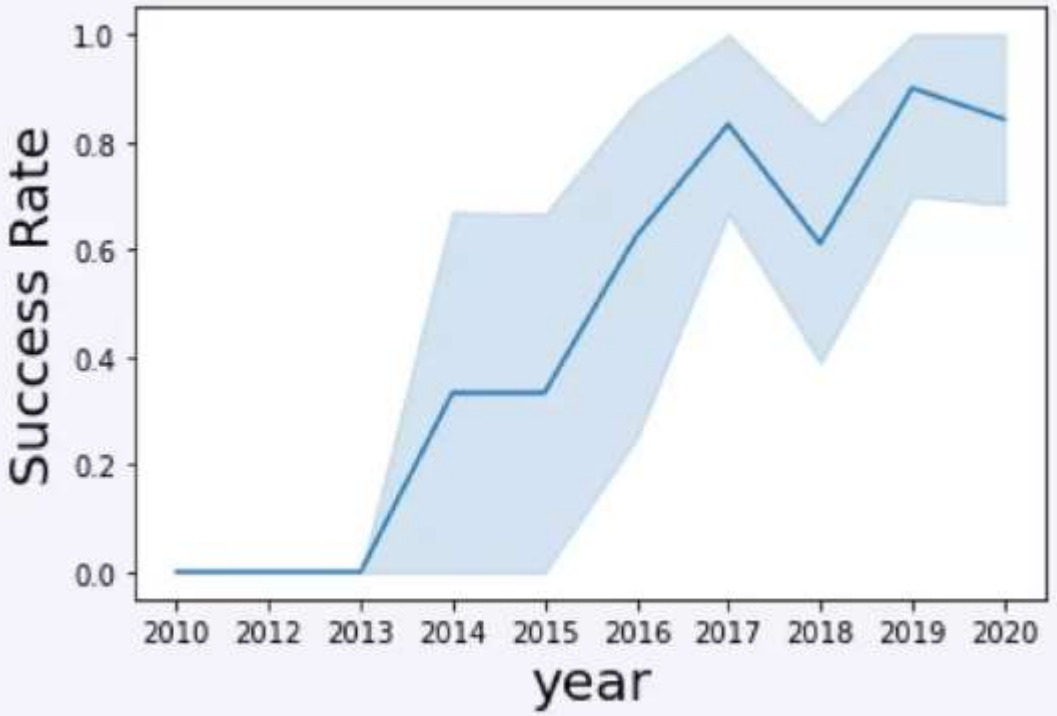
- A trend can be observed of shifting to VLEO launches in recent years.

Payload vs. Orbit Type



- There are strong correlation between ISS and Payload at the range around 2000, as well as between GTO and the range of 4000-8000.

Launch Success Yearly Trend

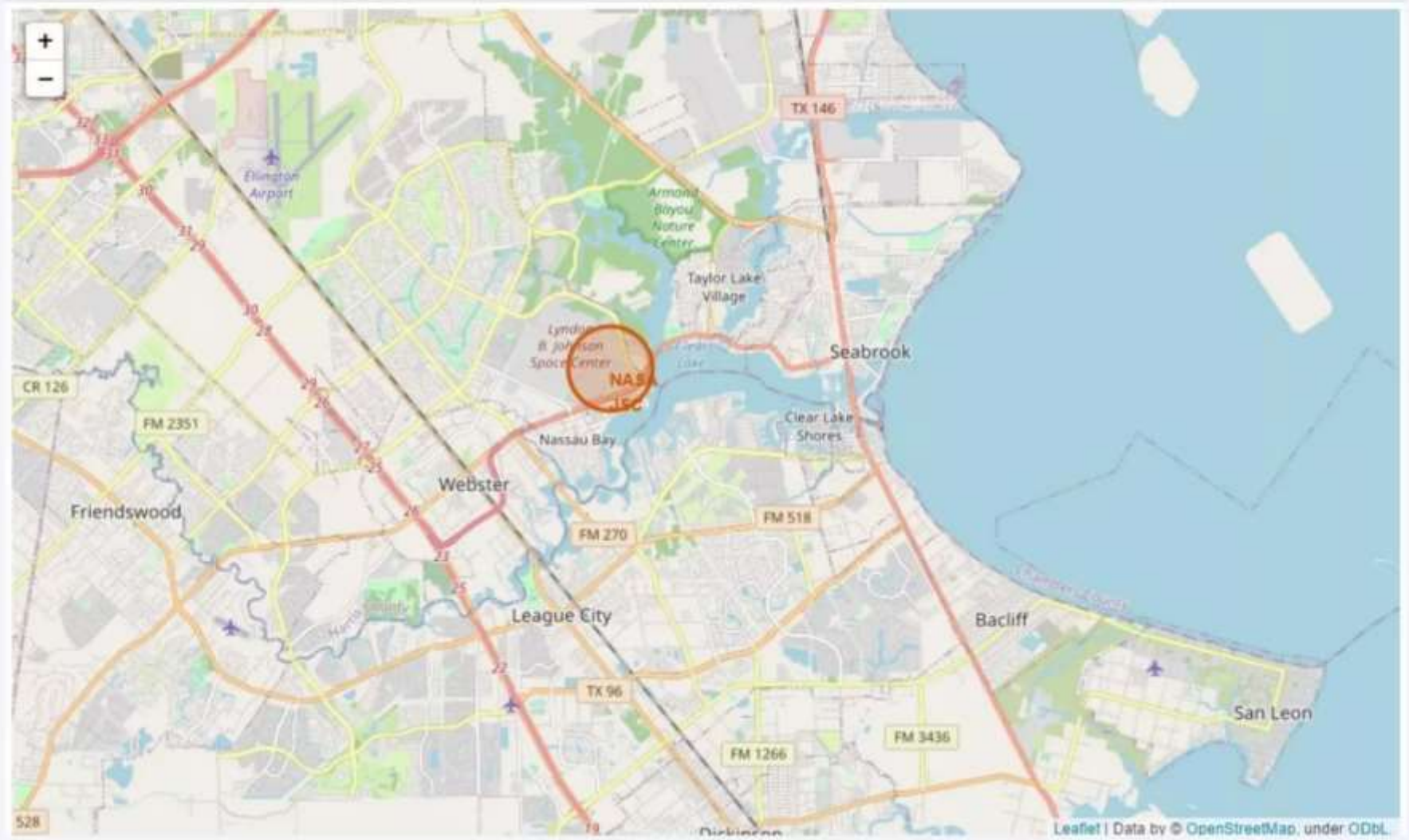


- Launch success rate has increased significantly since 2013 and has stablised since 2019, potentially due to advance in technology and lessons learned.

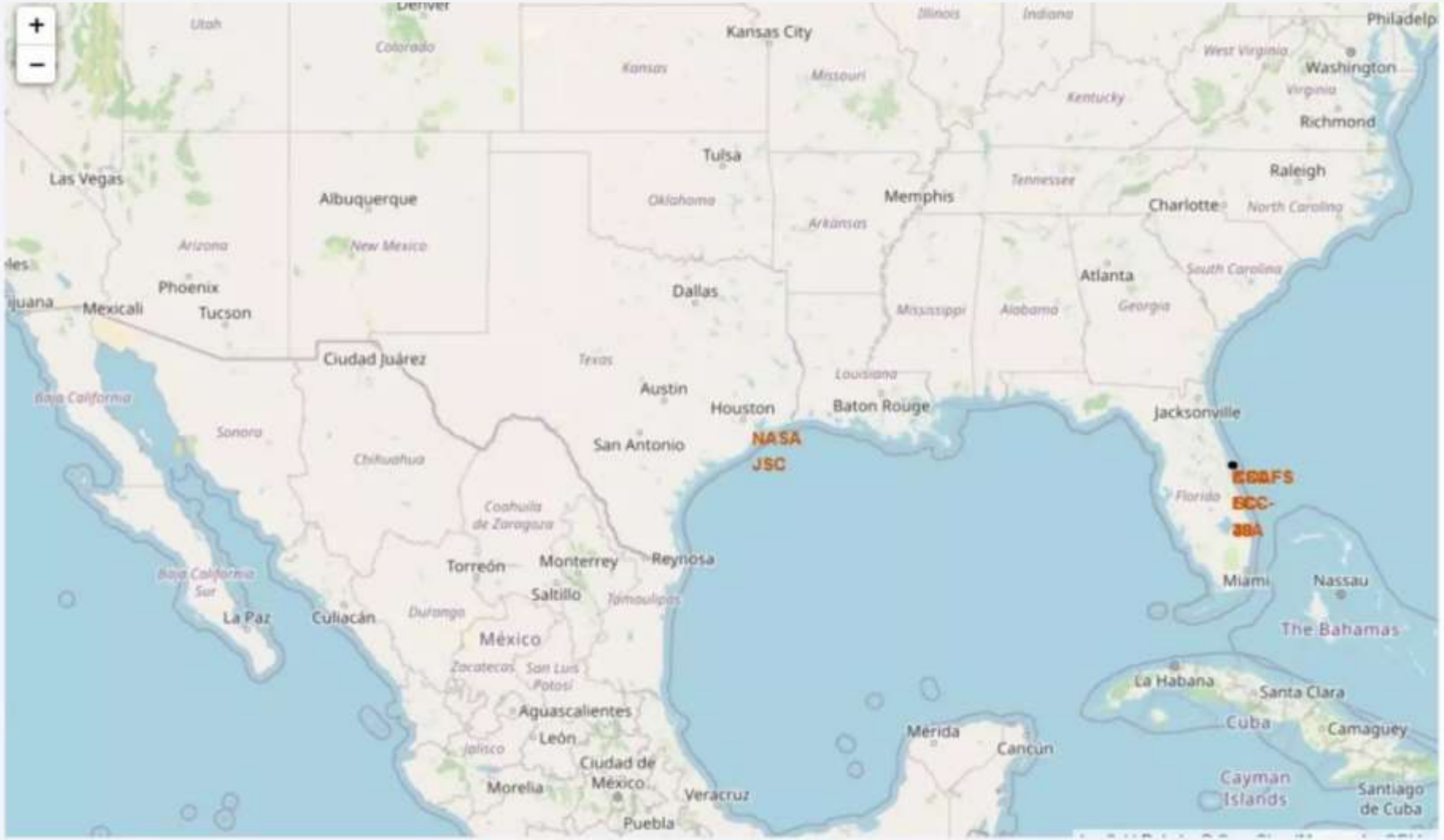
Section 3

Launch Sites Proximities Analysis

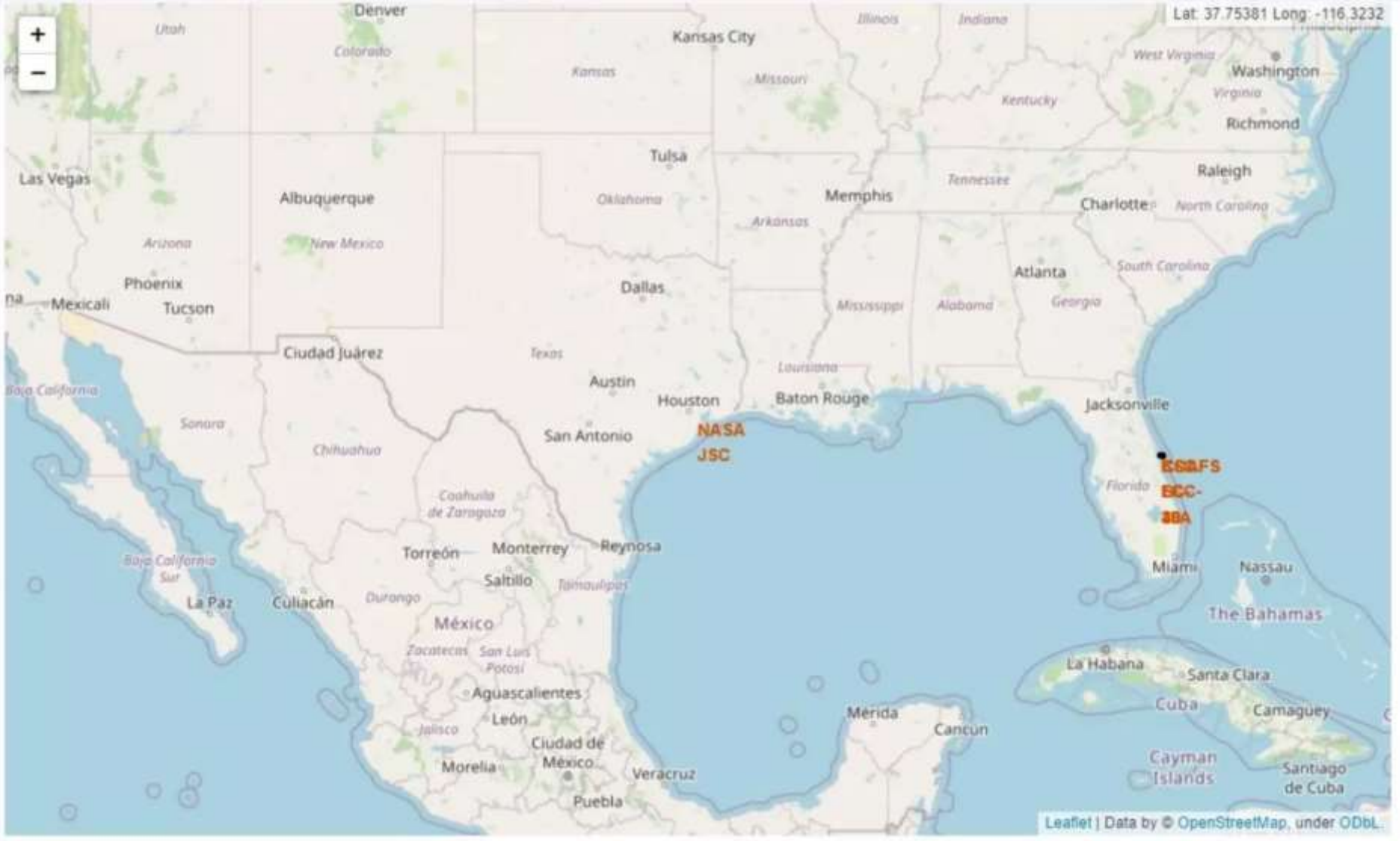
All launch sites marked on map



Success/failed launches marked

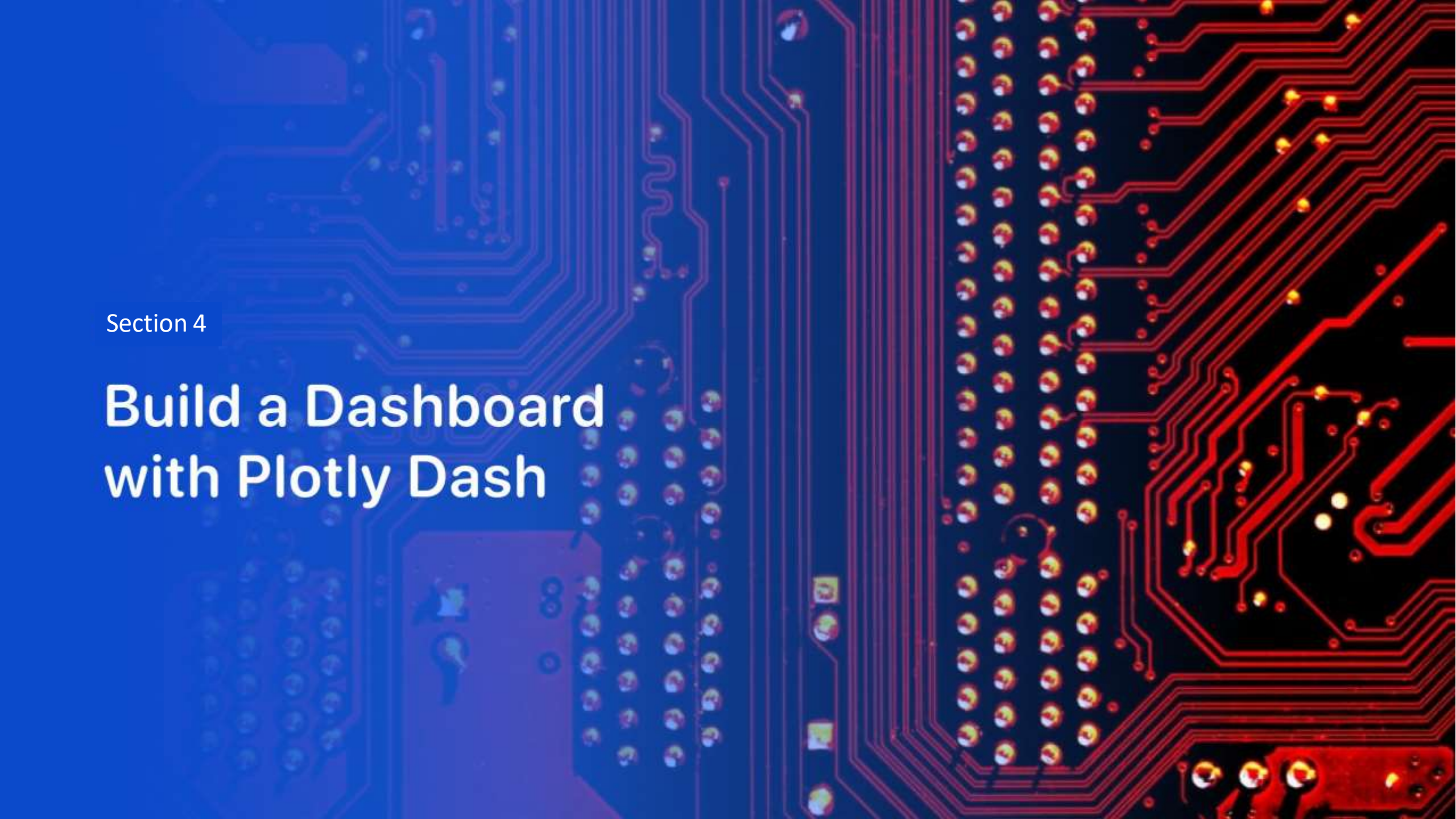


Distance b/w launch site and proximities

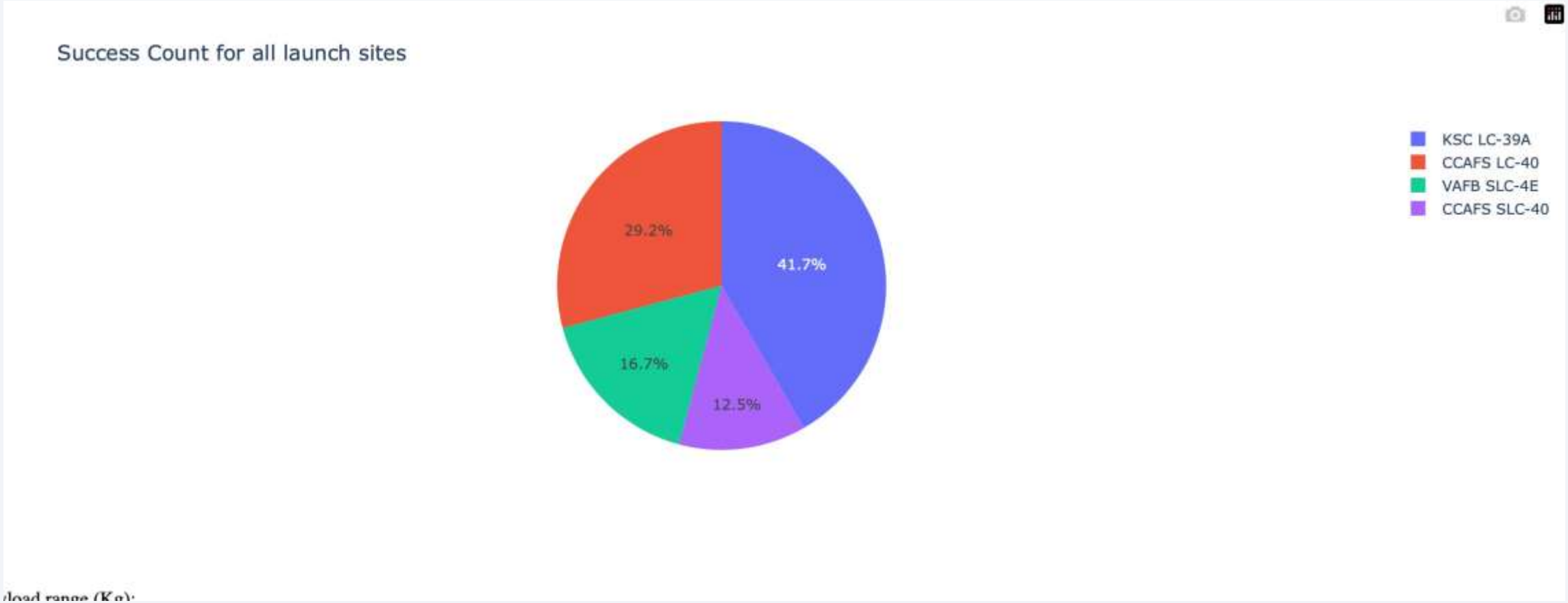


Section 4

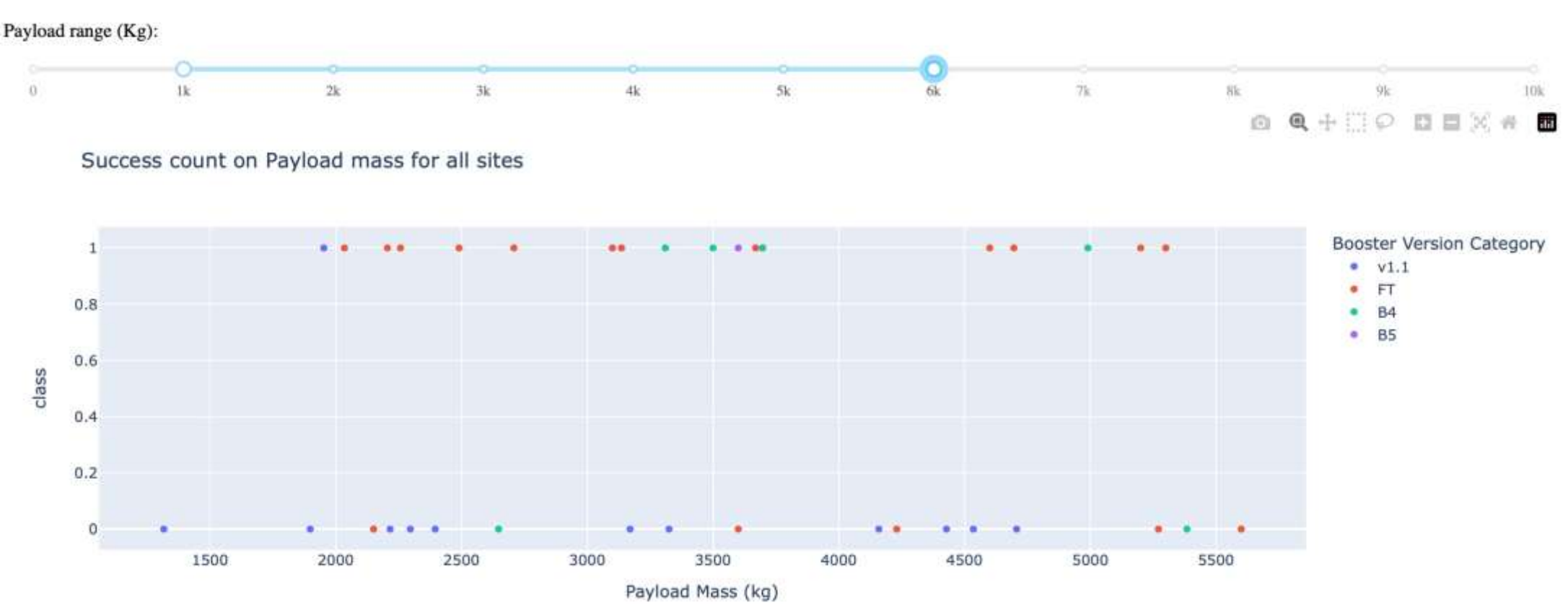
Build a Dashboard with Plotly Dash



Success rate for all launch sites



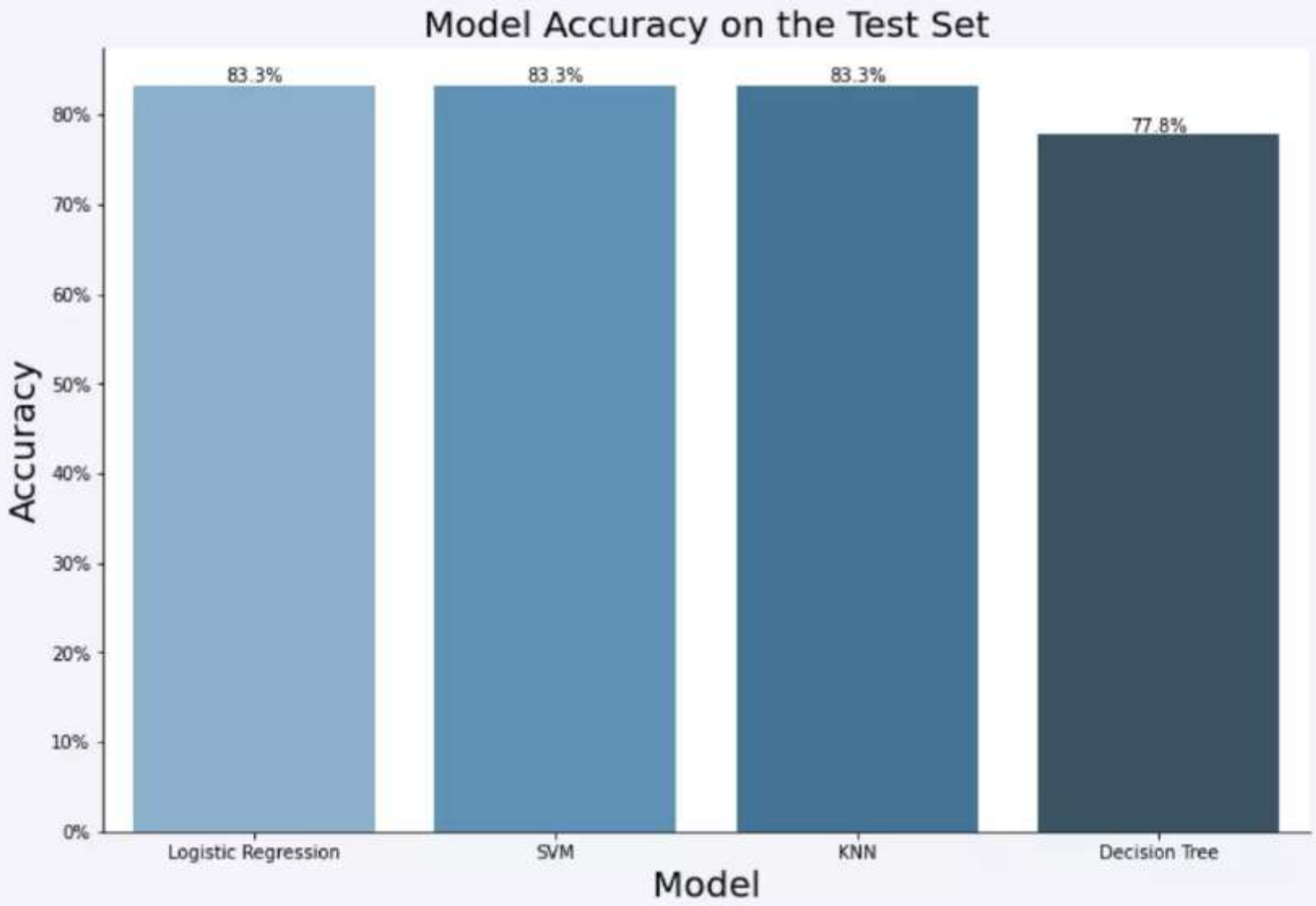
Payload Vs. Launch Outcome



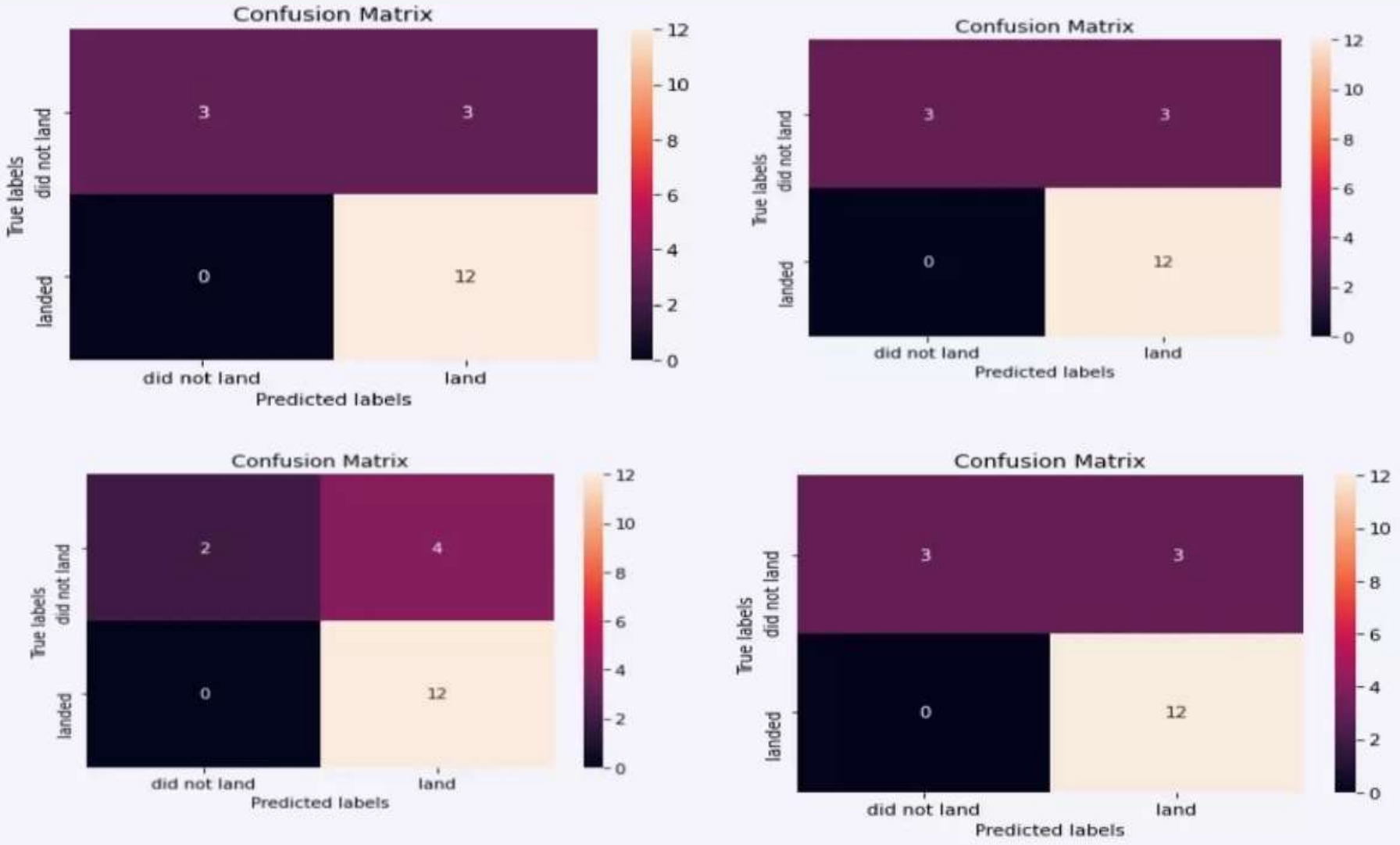
Section 5

Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix



Thank you!