IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Name : Rifatul Islam Majumder
Date : 31st March 2024

Outline

Executive Summary

Introduction

Methodology

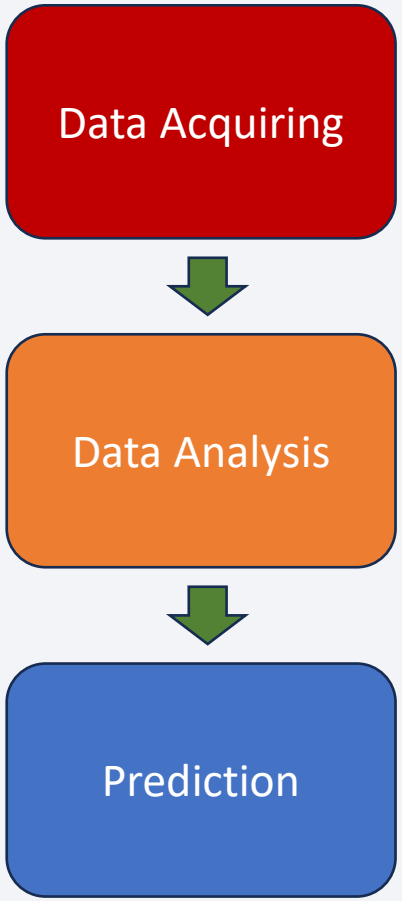
Results

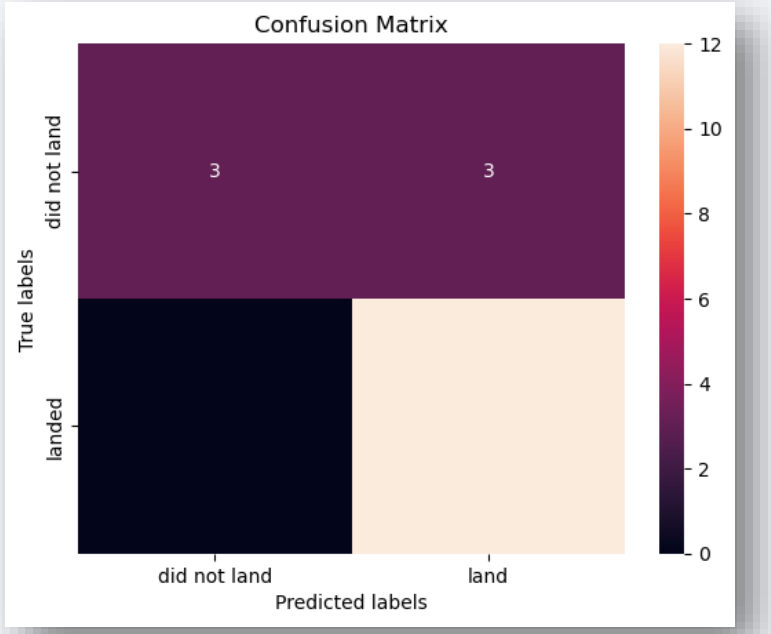
Conclusion

Appendix

Executive Summary

Project Description





Highly likely it will take more than this amount of effort and data to predict on this crucial project

Key Findings:

- SpaceX primary goal is to produce General transportable spacecraft
- Management improvement of simultaneously all facilities can improve success rate

Introduction

CONTEXT

- Finding out the motive of SpaceX
- Data Discovery

PROBLEM

- Prediction on a project before launching to minimize cost.
- Scope of improvement.

Why?

- Reductio of cost
- Reduction of energy and labor usage
- Consistent Improvement

Section 1

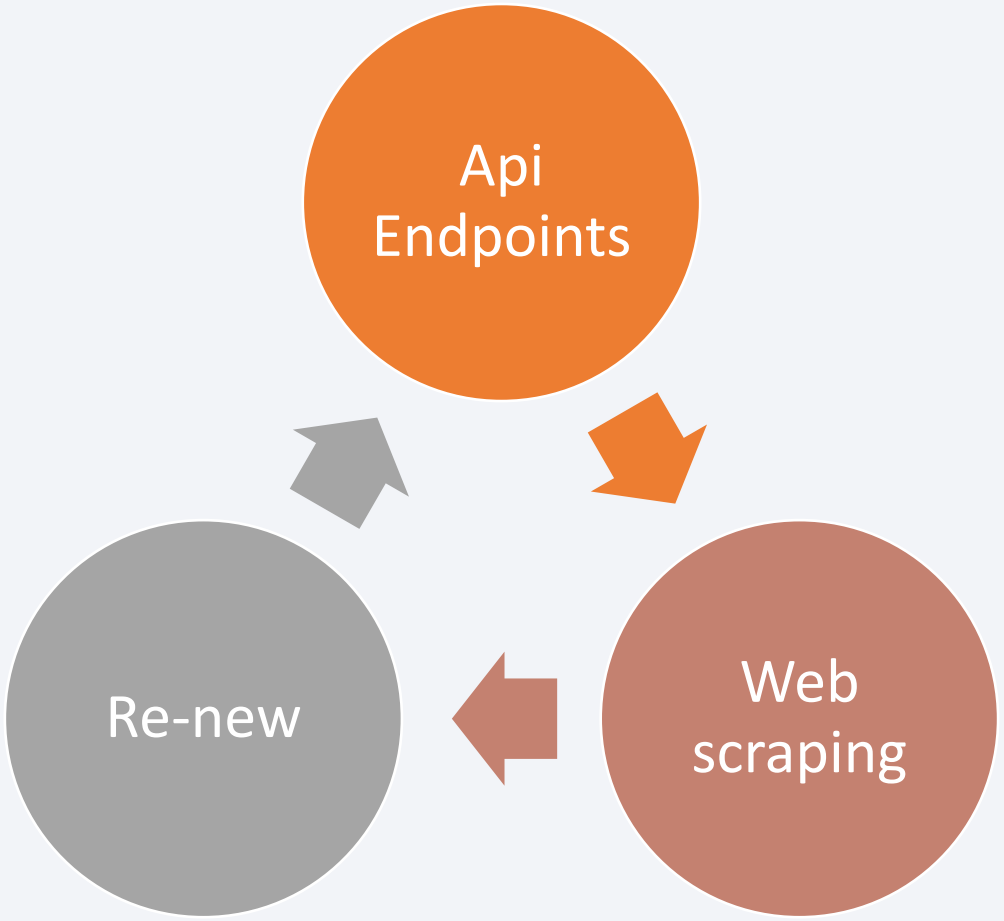
Methodology

Methodology

Executive Summary

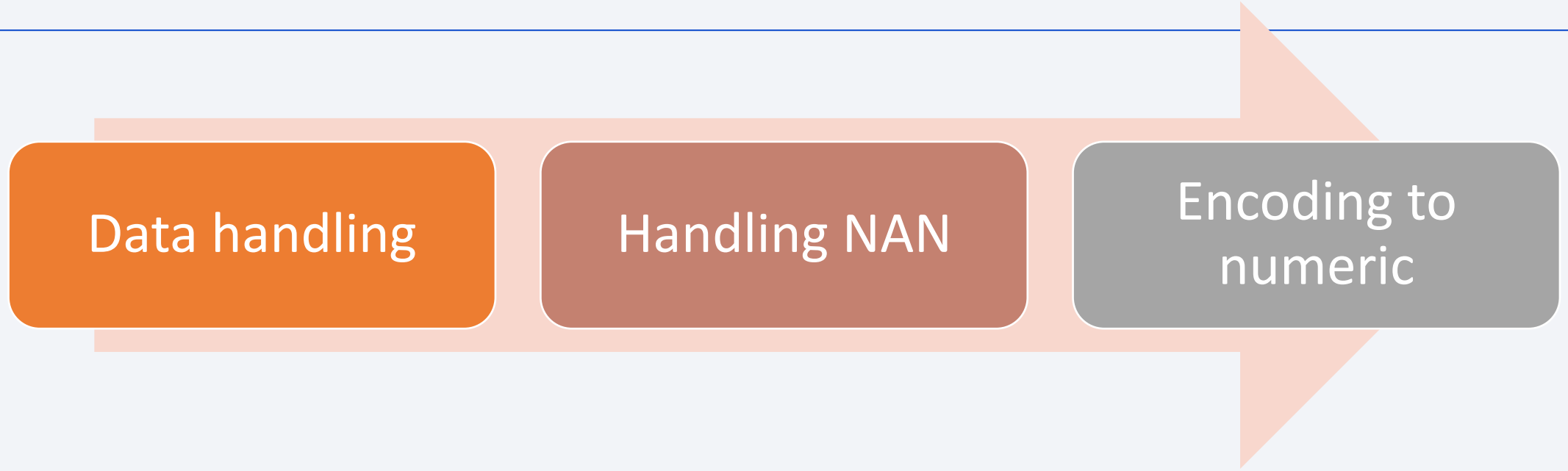
- Data collection methodology:
  - Data was gathered through multiple API endpoint & web scraping
- Perform data wrangling
  - Processing includes dealing with NAN value(not a number),converting categorical features to numeric features.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Model is build on Cross-Validation and 80%,20% training testing split
  - Logistic regression , Support Vector Machine ,Decision Tree , K Nearest Neighbors algorithms are tested .

Data Collection



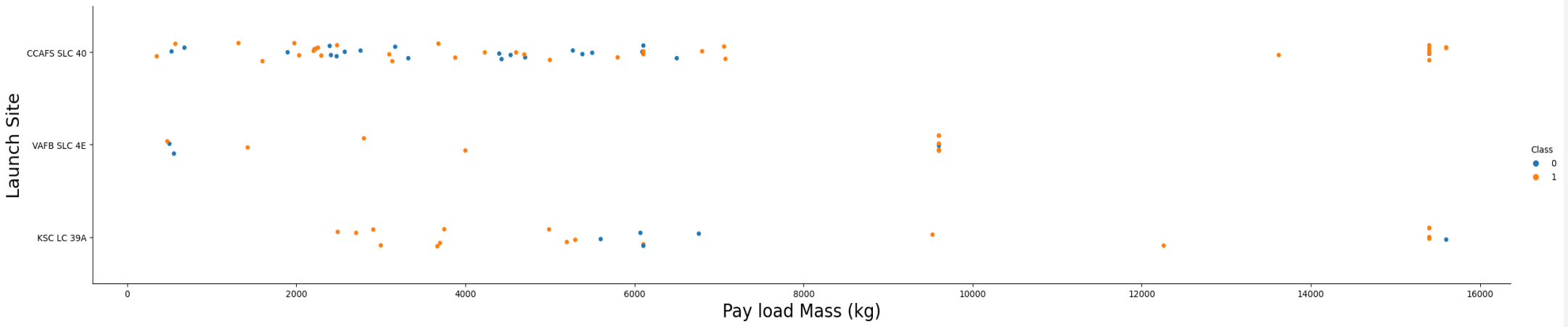
GET NOOTBOOK

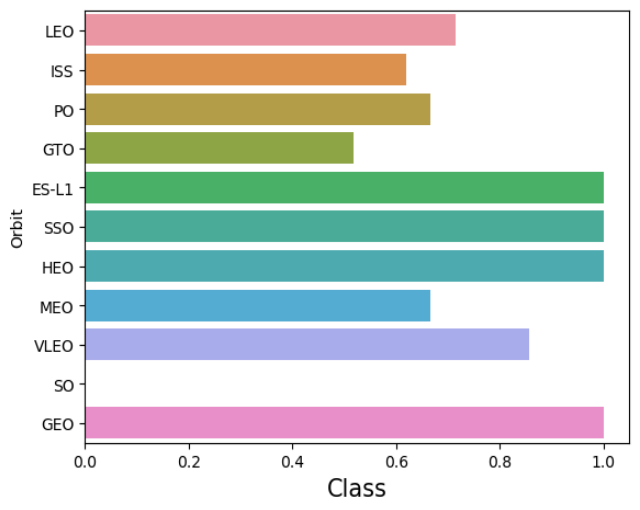
Data Wrangling

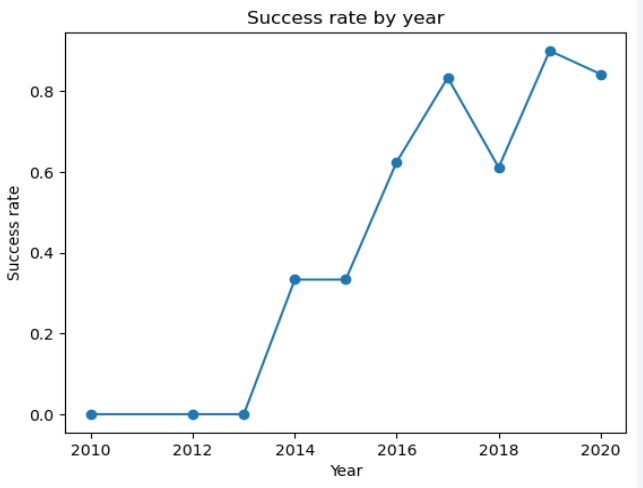


GET NOOTBOOK

EDA with Data Visualization







GET NOOTBOOK

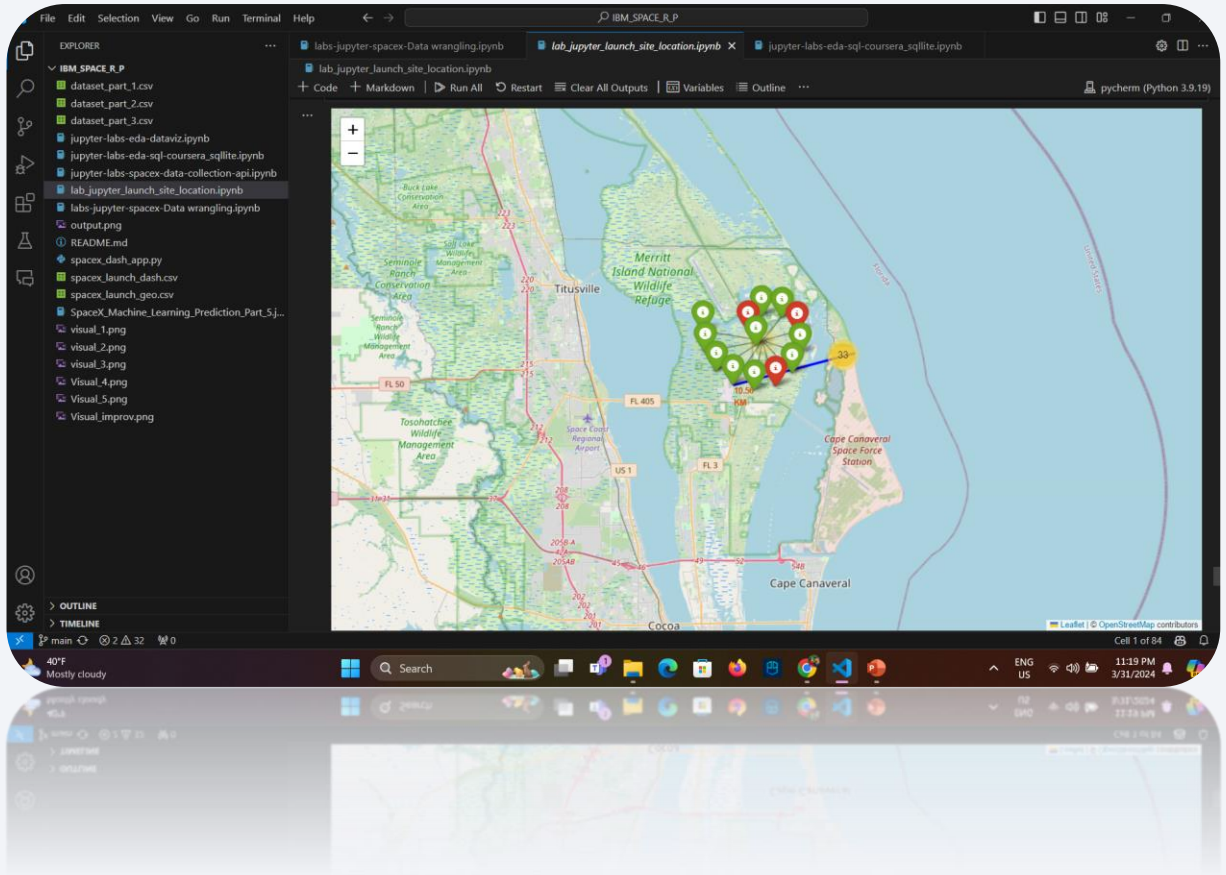
EDA with SQL

- Compensative possible SQL integration can be possible
- Creates possibility of we-application

GET NOOTBOOK

Build an Interactive Map with Folium

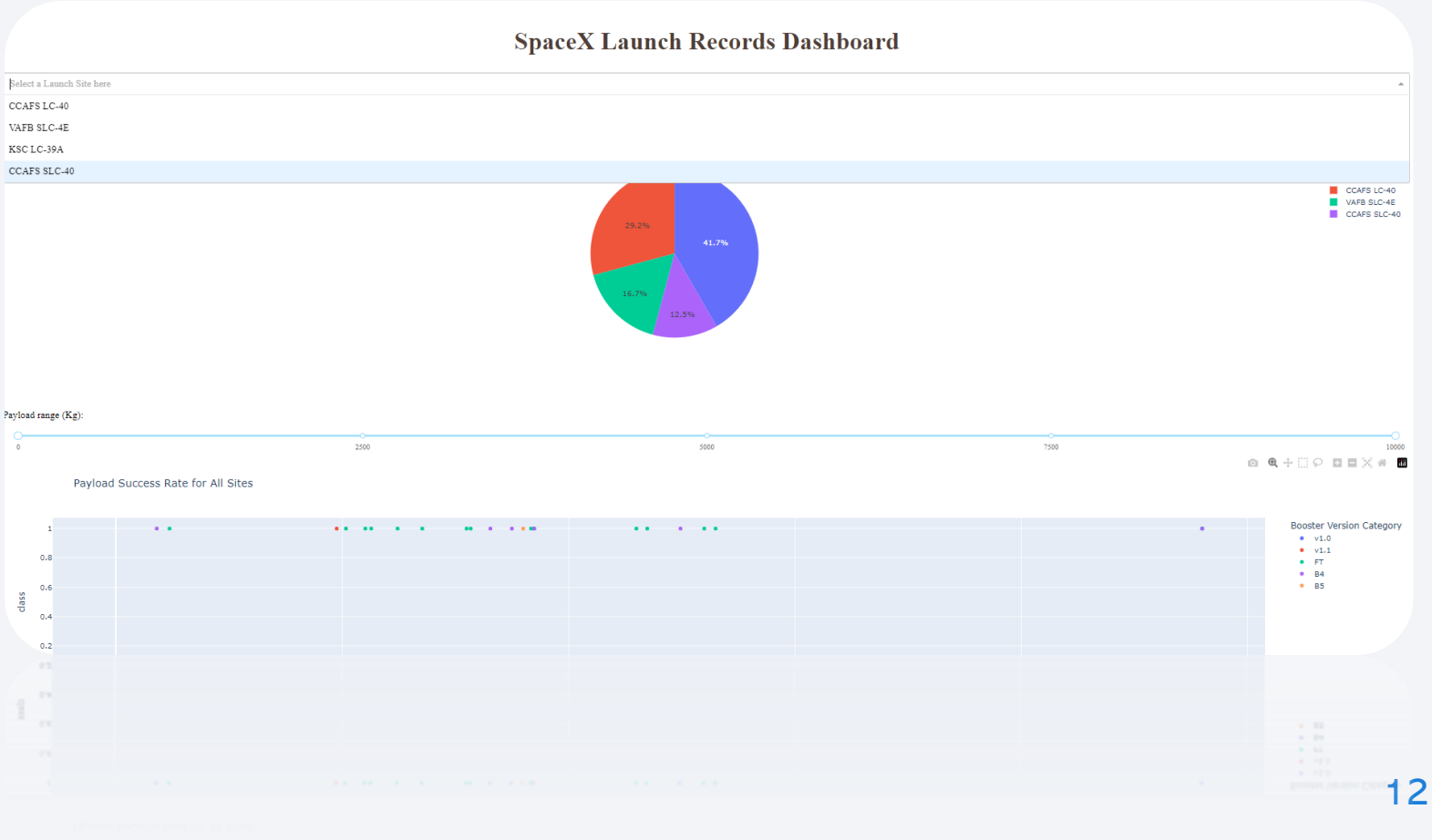
- Integration of folium and Jupiter-lab based application is used for better visualization of geo-data.



GET NOOTBOOK

Build a Dashboard with Plotly Dash

- For better visualization for Business and stake holders interactive studio uas created.



GET PYTHON

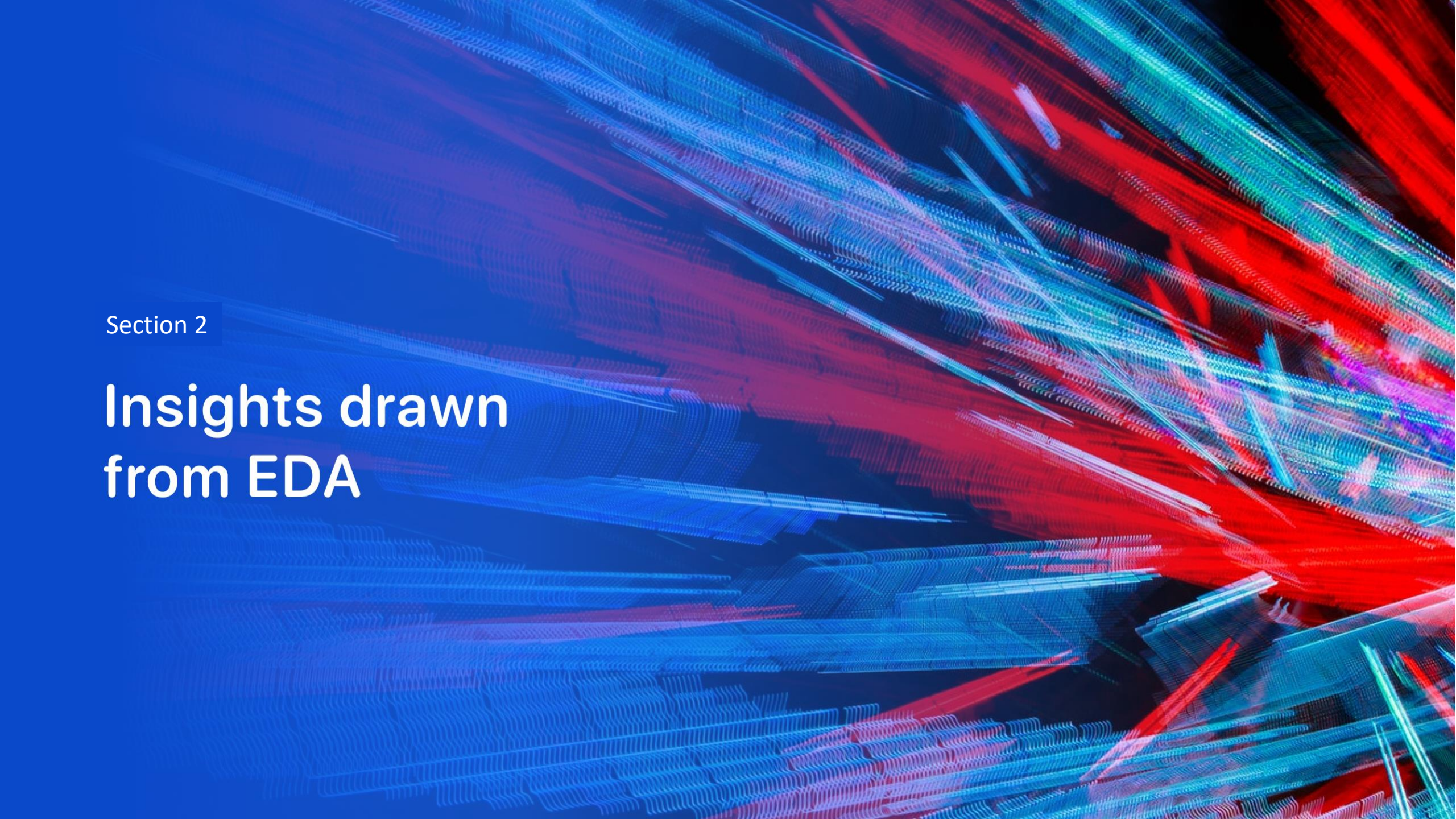
Predictive Analysis (Classification)

- Multiple Algorithms used to predict on Success or failure
  - Model is build on Cross-Validation and 80%,20% training testing split
  - Logistic regression , Support Vector Machine ,Decision Tree , K Nearest Neighbors algorithms are tested .

GET NOOTBOOK

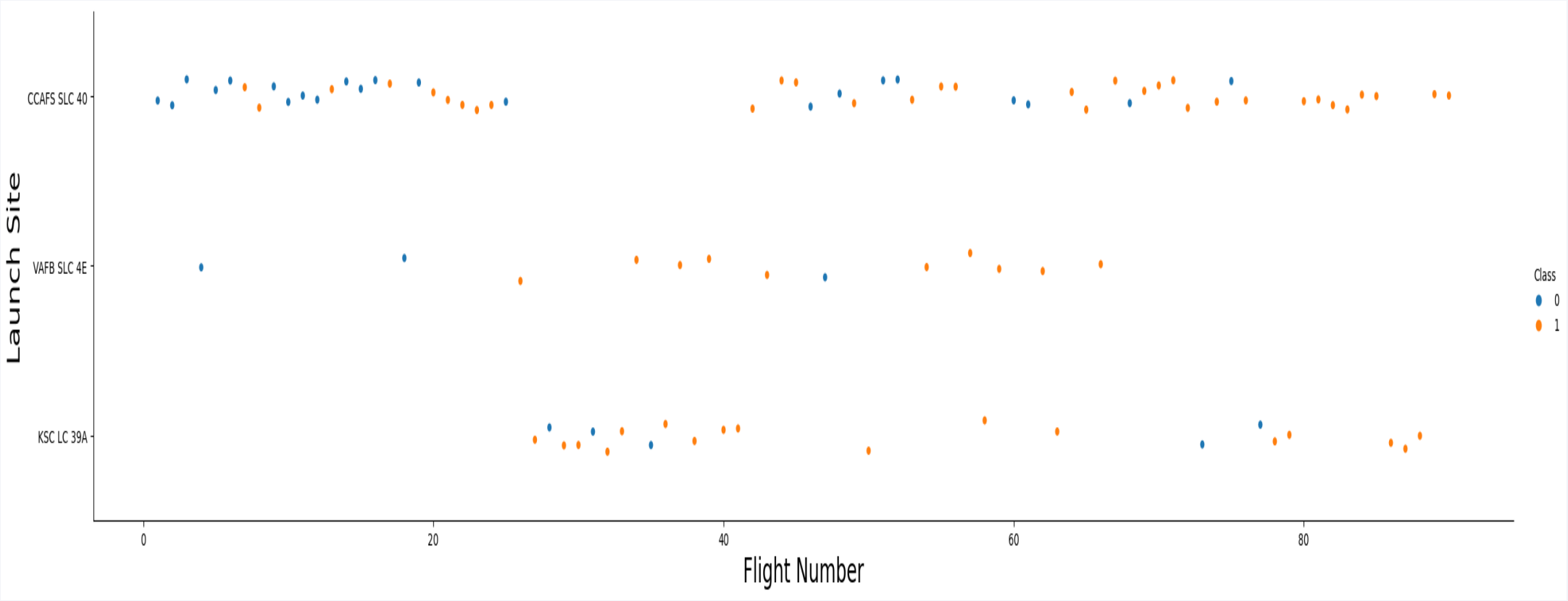
Results

- Though The predictive answer scope is narrow but descriptive answer can be found.


Section 2
Insights drawn from EDA

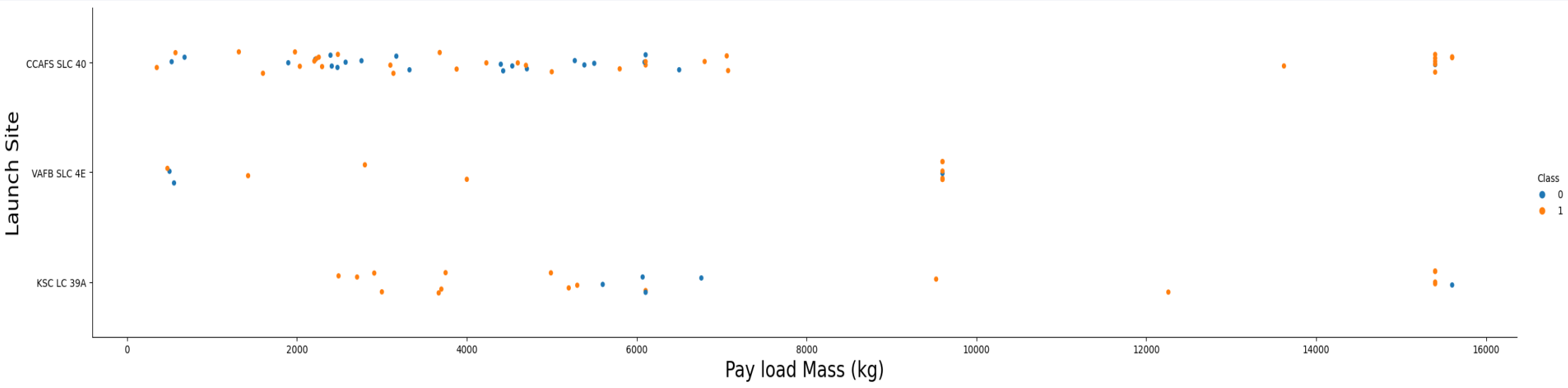
Flight Number vs. Launch Site

- SpaceX uses their one station too much frequently for every purpose



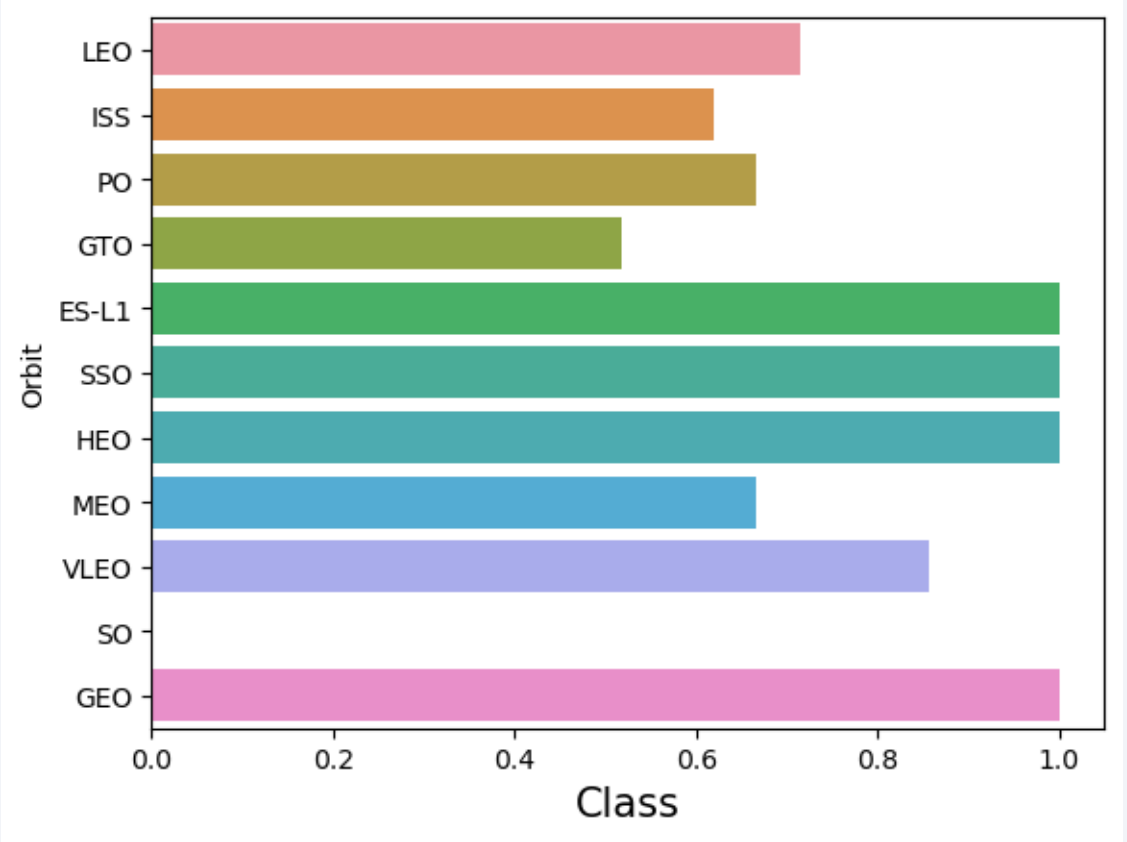
Payload vs. Launch Site

- WAFB SLC 4E is not used for higher payload



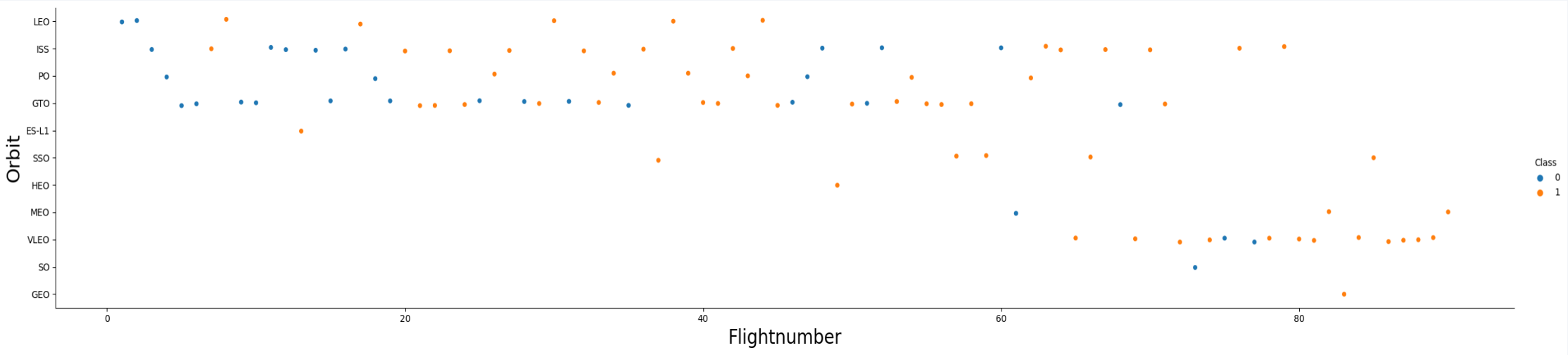
Success Rate vs. Orbit Type

- Willingly trying to only focus on long term



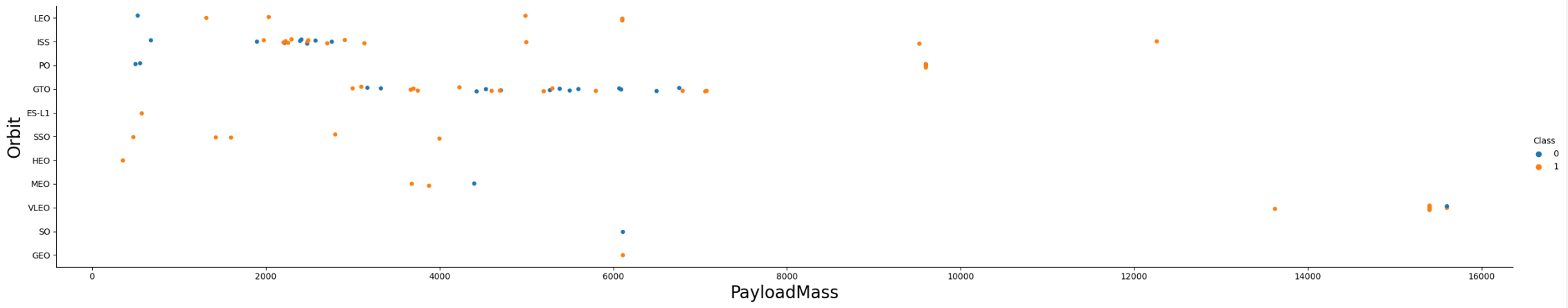
Flight Number vs. Orbit Type

- The grouping looks some relation comprehensive understanding required



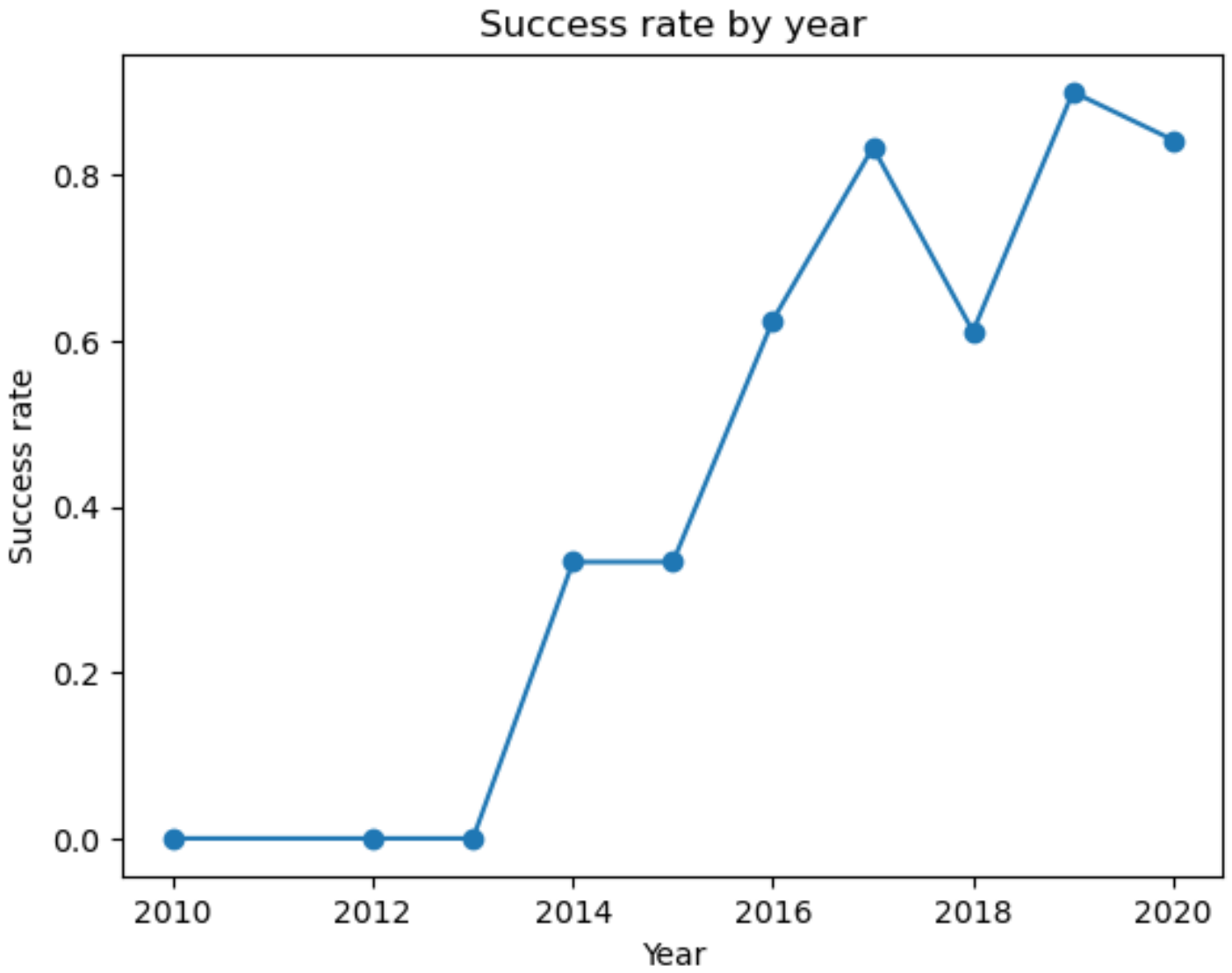
Payload vs. Orbit Type

- This graph also proofs SpaceX is more concerned about space traveling



Launch Success Yearly Trend

From 2013



Thank you!