Winning Space Race with Data Science

Taha Kerim Sezer

10.10.2022

GITHUB REPOSITORY LINK

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection by using SQL, Web Scraping and SpaceX API
  - Data Wrangling
  - Data Analysis
  - Interactive Visual Analytics by using Folium
  - Predictive Analysis
- Summary of all results
  - Data Analysis using Visual Analytics
  - Finding best model for Predictive Analysis

Introduction

- Project background and context

  In this project we predicted if the first stage of Falcon 9 will land succesfully. If this happens Space X will save much cost. So, if we found out if the first stage land succesfully, this info wil be very help for many companies.

- Problems you want to find answers

  If the first stage of Falcon 9 land succesfully and factors

  Best conditions for succesfull landing

Section 1

Methodology

Methodology

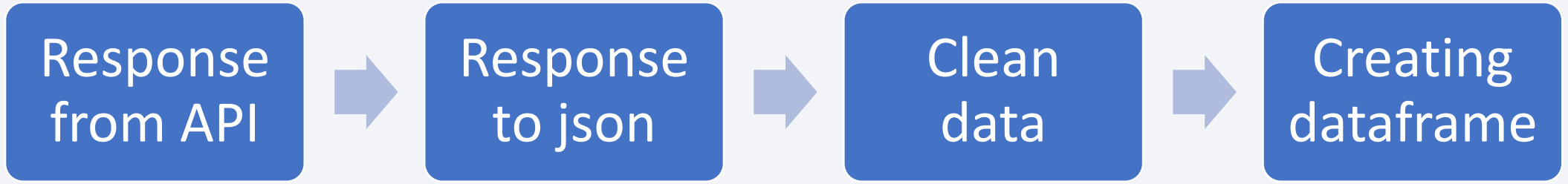
Executive Summary

- Data collection methodology:
  - Using Space X rest API
- Perform data wrangling
  - Dropping irrelevant columns etc.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Building classification models

Data Collection

- How datasets collected?
    1. Acquire data from rest API
    2. Forming dataframes
    3. Filtering the dataframe
    4. Export to a csv

Data Collection – SpaceX API



GITHUB link

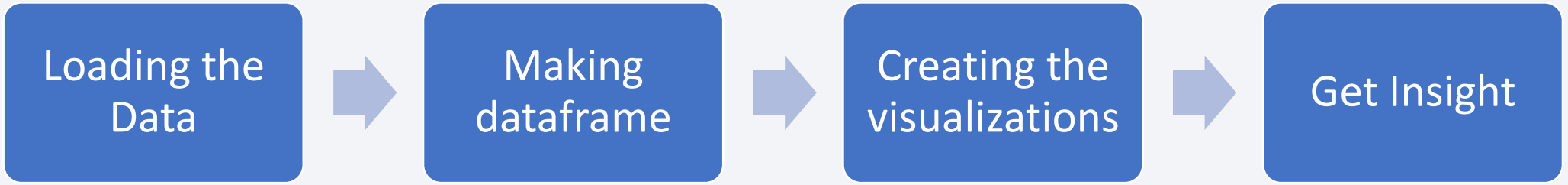
Data Wrangling

- How were data processed?

  Cleaning messy and complex data for easy analysis. Data wrangling done by converting messy data to simple form.

GITHUB link

EDA with Data Visualization

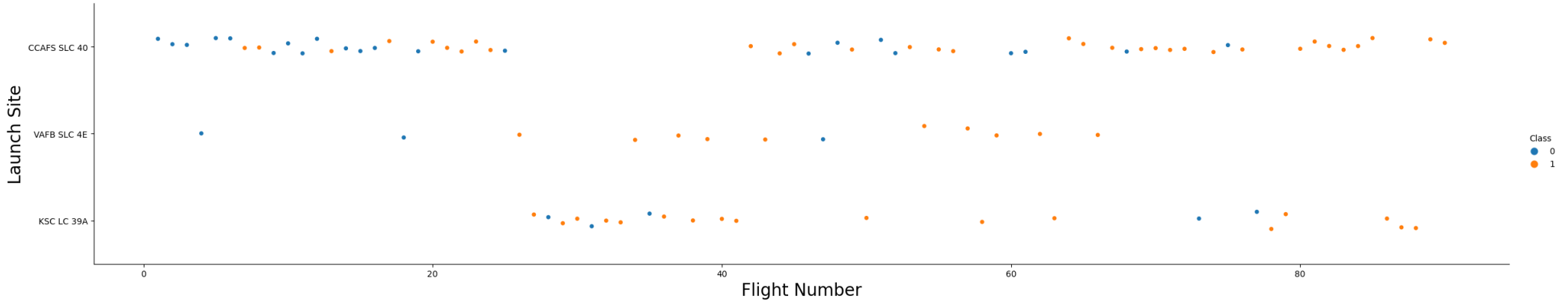


GITHUB link

EDA with Data Visualization

- Scatter Graphs
  - Payload Mass vs Flight Number
  - Launch Site vs Flight Number
  - Launch Site vs Payload Mass
  - Orbit vs Flight Number
  - Outcome vs Payload Mass

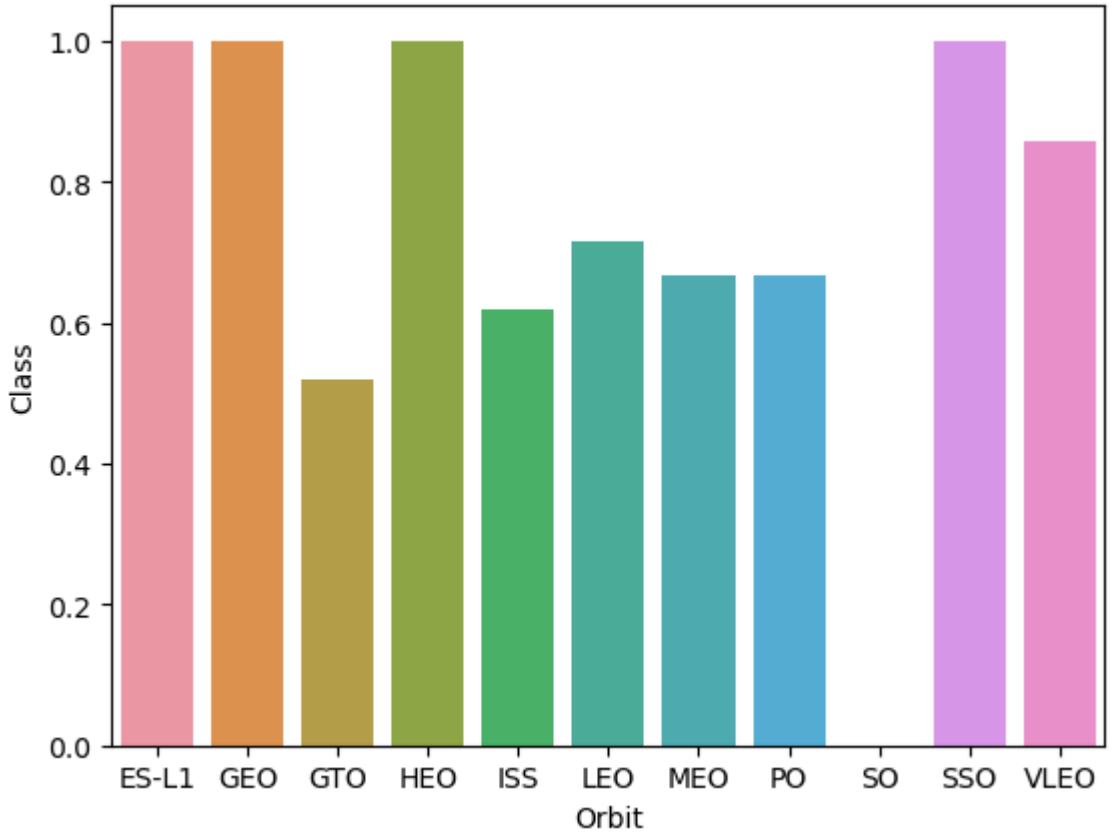
  Scatter graphs showing that the dependency of each attribute to each other. And gives us insight about their relation. They help us to predict which factors will give maximum success probability.



EDA with Data Visualization

- Bar Graph
  - Succes rate vs Orbit

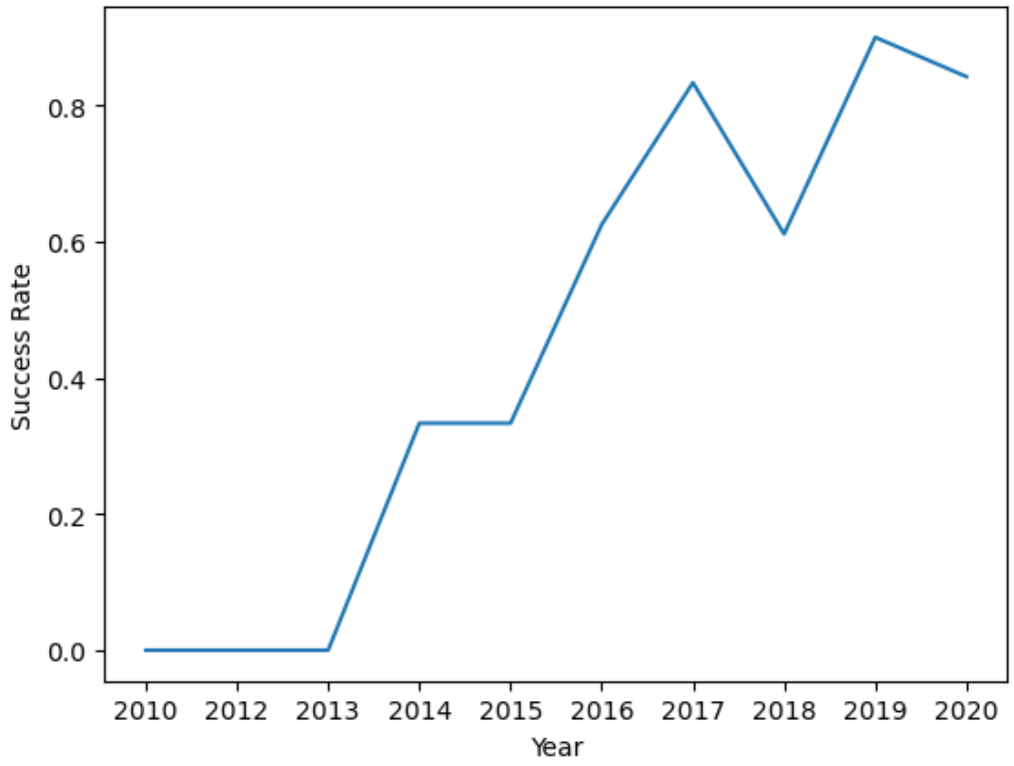
  To see which orbit have highest succes rate and compare the orbits.



EDA with Data Visualization

- Line Graph
  - Success rate vs year

  In line graphs we can see the trends clearly.



EDA with SQL

GITHUB link

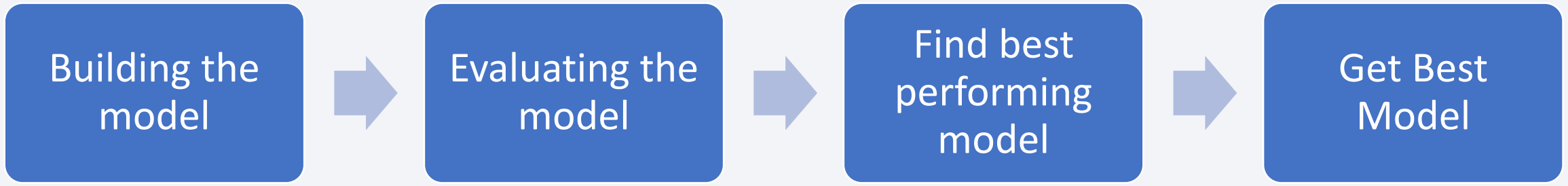
- SQL queries
  - Selecting unique launch sites
  - Selecting 5 records which their launch site is CCA
  - Selecting total payload mass boosters launched by NASA
  - Display average payload mass carried by booster version F9 v1.1
  - List the date when the first succesful landing outcome in ground pad 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 successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

Build an Interactive Map with Folium

GITHUB link

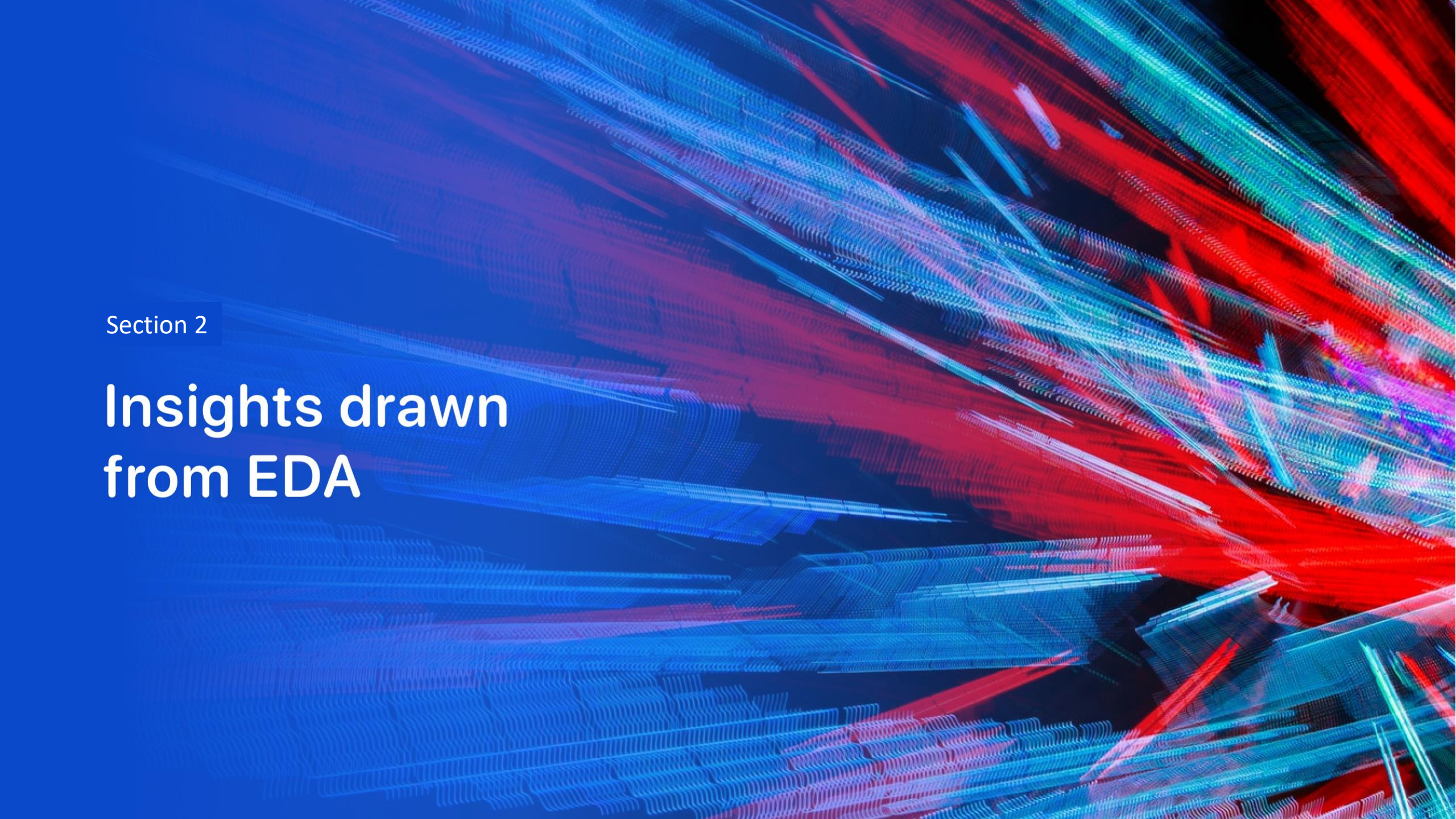
- Map objects created are map marker, icon marker,circle marker,polyline,marker cluster object and anthpath
- Map marker is for to make a mark on map, icon marker is for the creating an icon on the map, circle marker create a circle where marker is placed,polyline create line between points, marker cluster object simply markers having same location and antpath is for animated line between points.

Predictive Analysis (Classification)



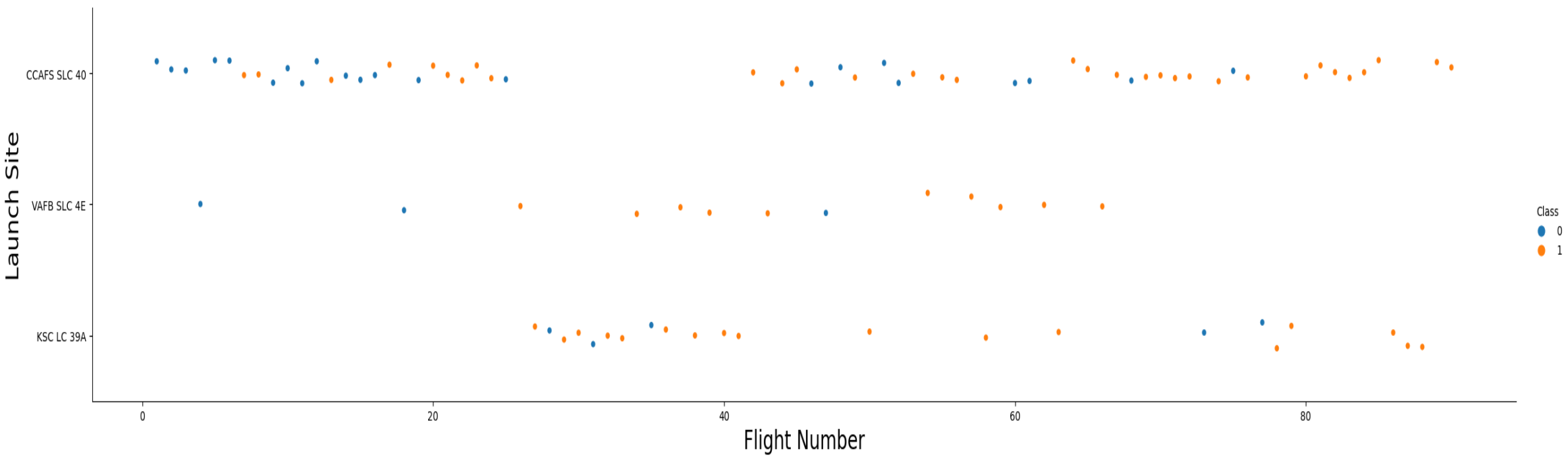
Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results


Section 2
Insights drawn from EDA

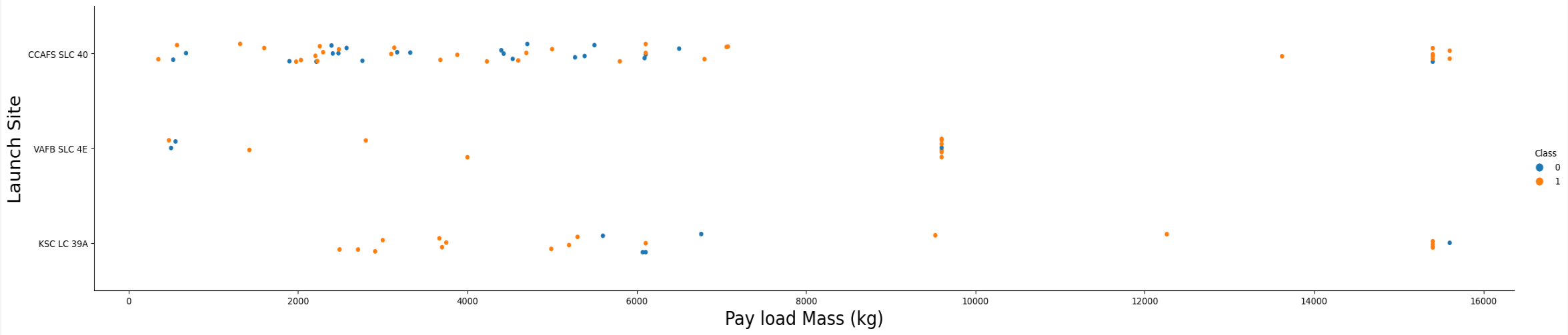
Flight Number vs. Launch Site

- Succes rate increases after the flight number passes the thirty (30)



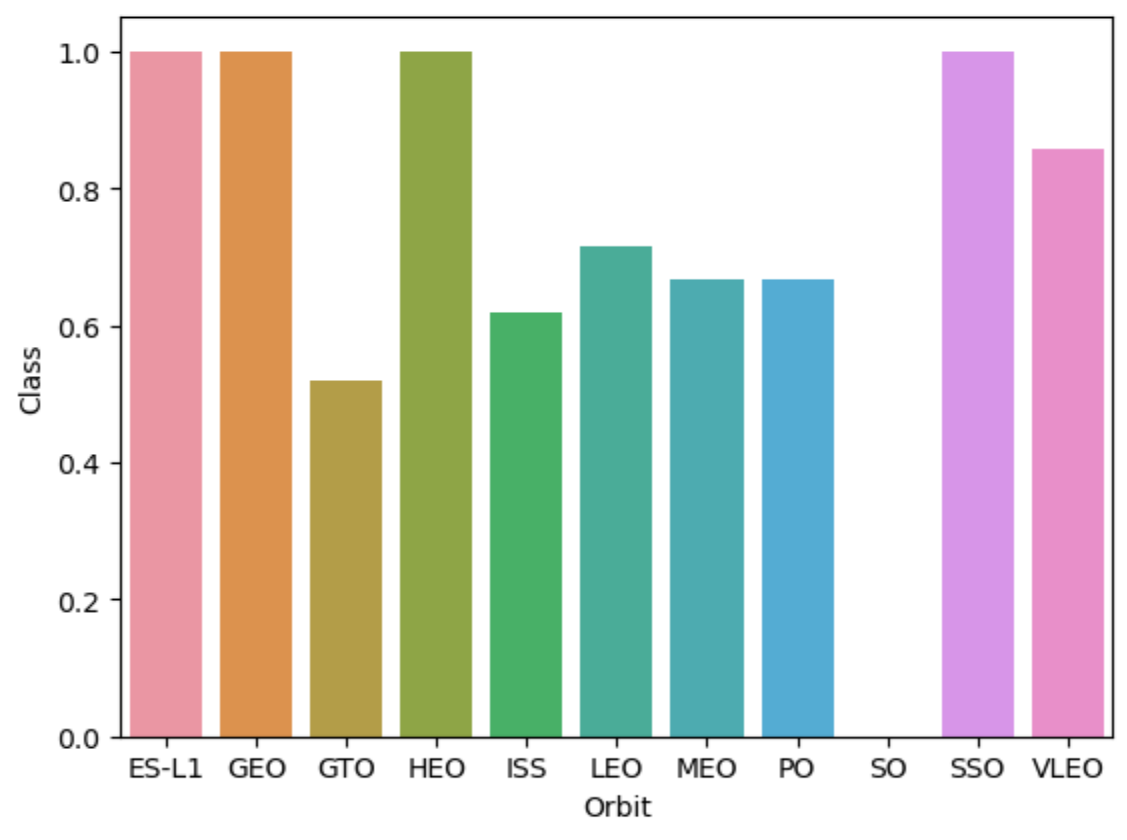
Payload vs. Launch Site

- After 7k kg payload mass clearly seen that succes rate is rising.



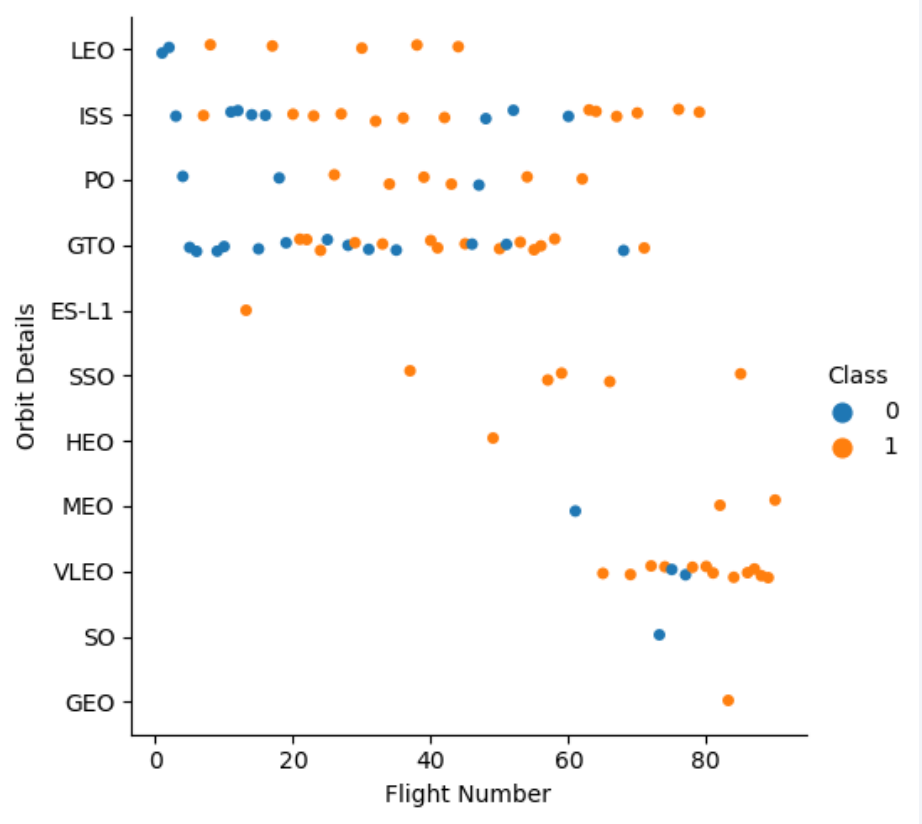
Success Rate vs. Orbit Type

- From the graph ES-L1, GEO, HEO and SSO have the highest succes rates.



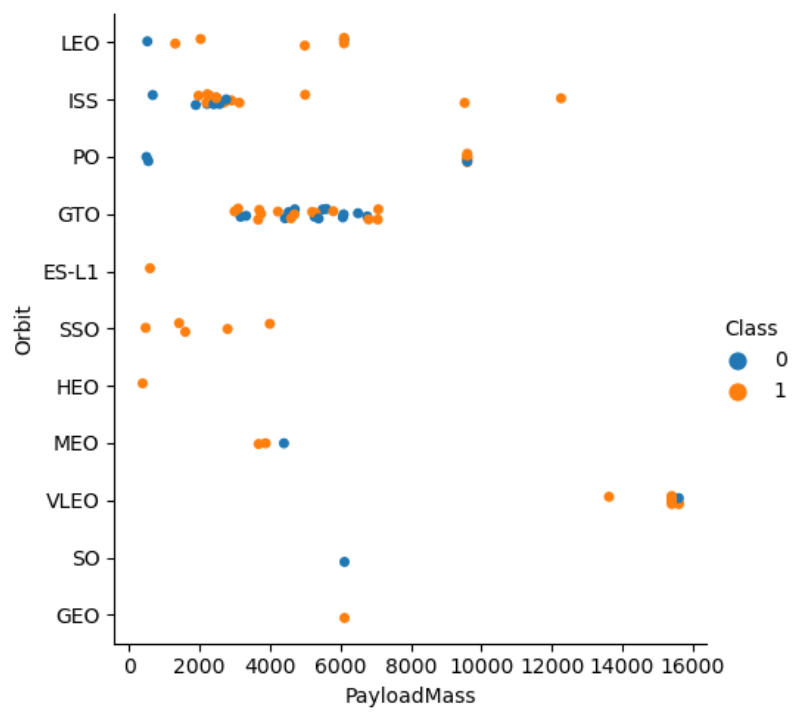
Flight Number vs. Orbit Type

- In LEO succes rate increases with the increasing flight number



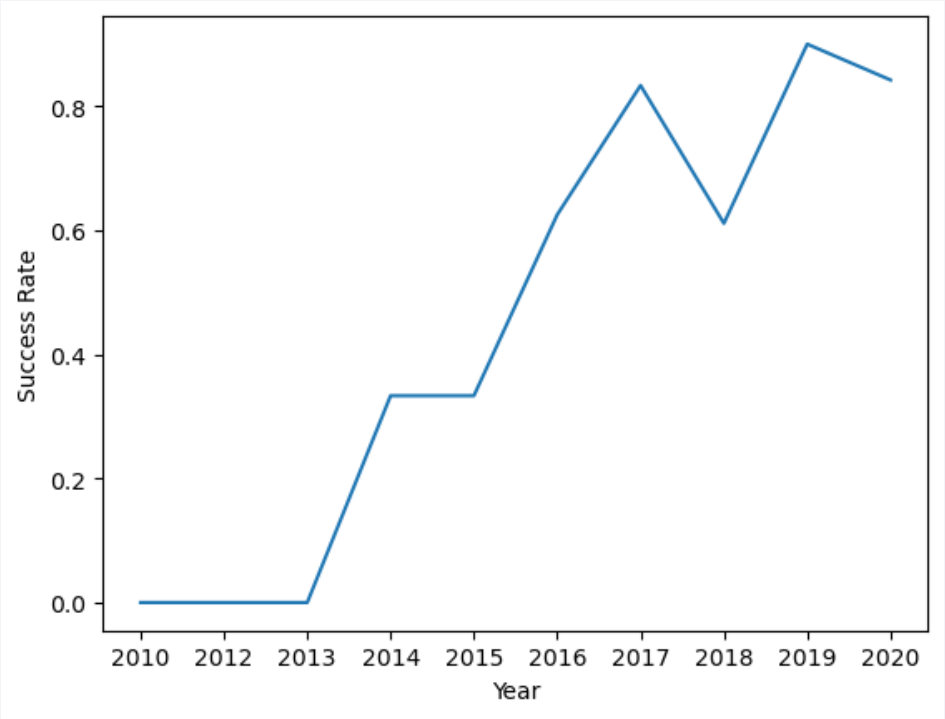
Payload vs. Orbit Type

- Heavy loads have high succes rate in LEO



Launch Success Yearly Trend

- Succes rate kept increasing after 2013



All Launch Site Names

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

Launch Site Names Begin with 'CCA'

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	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)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

Total Payload Mass by NASA (CRS)
45596

Average Payload Mass by F9 v1.1

Average Payload Mass by Booster Version F9 v1.1
2928.4

First Successful Ground Landing Date

First Succesful Landing Outcome in Ground Pad
01-05-2017

Successful Drone Ship Landing with Payload between 4000 and 6000

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

Total Number of Successful and Failure Mission Outcomes

Successful Mission	Failure Mission
100	1

Boosters Carried Maximum Payload

Booster Versions which carried the Maximum Payload Mass
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

2015 Launch Records

Month	Booster_Version	Launch_Site
01	F9 v1.1 B1012	CCAFS LC-40
04	F9 v1.1 B1015	CCAFS LC-40

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

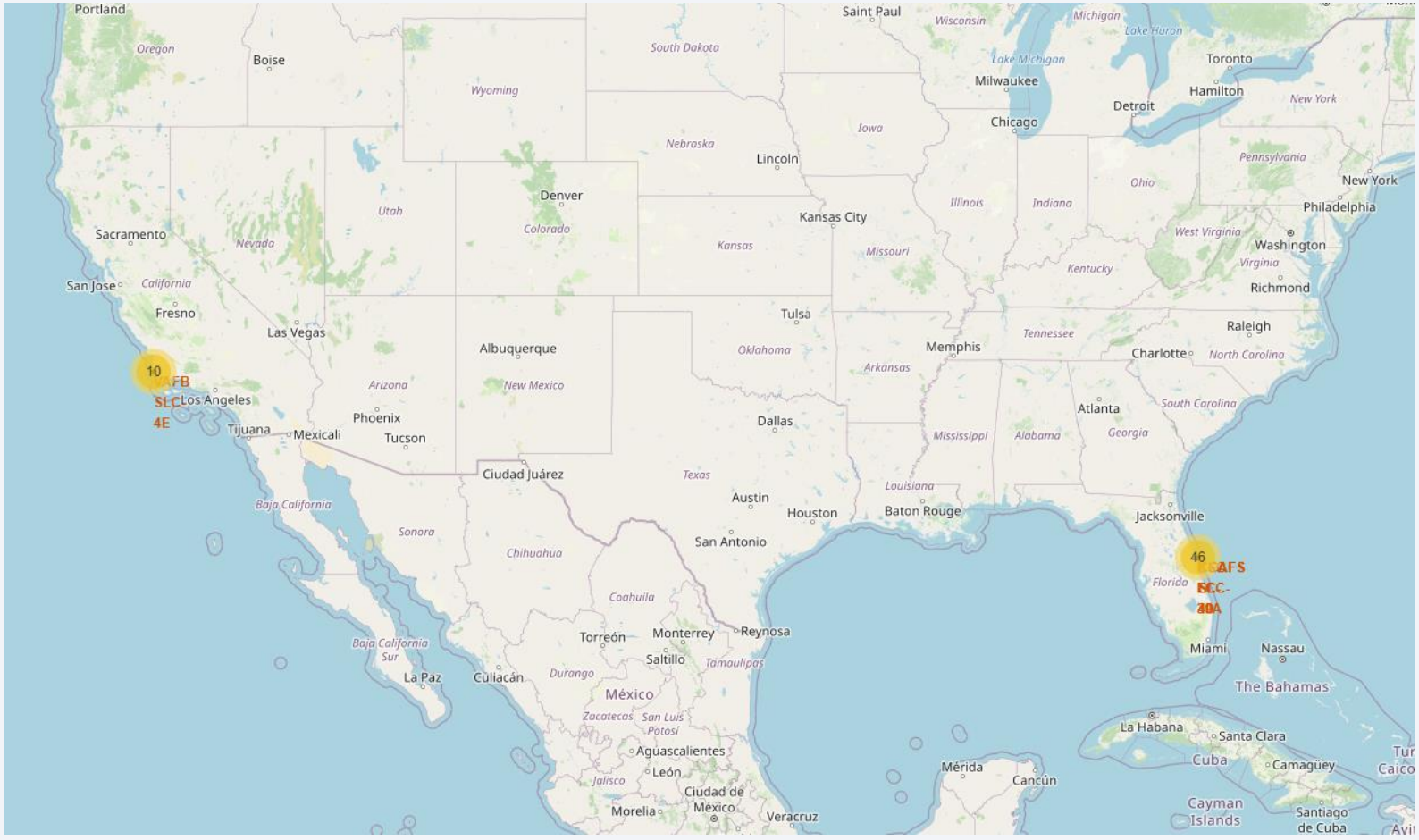
Rank success count between 2010-06-04 and 2017-03-20
40

Section 3

Launch Sites Proximities Analysis

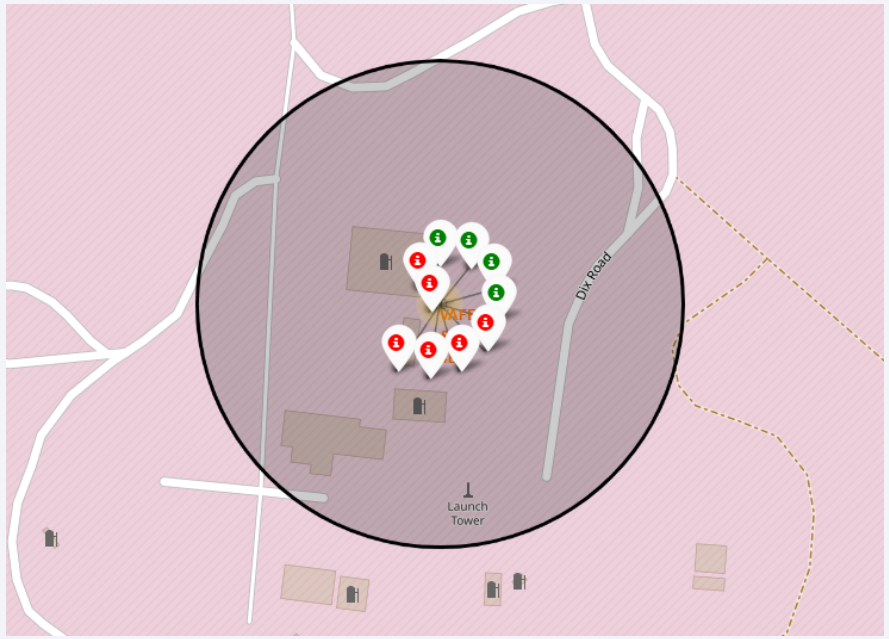
All Launch Sites

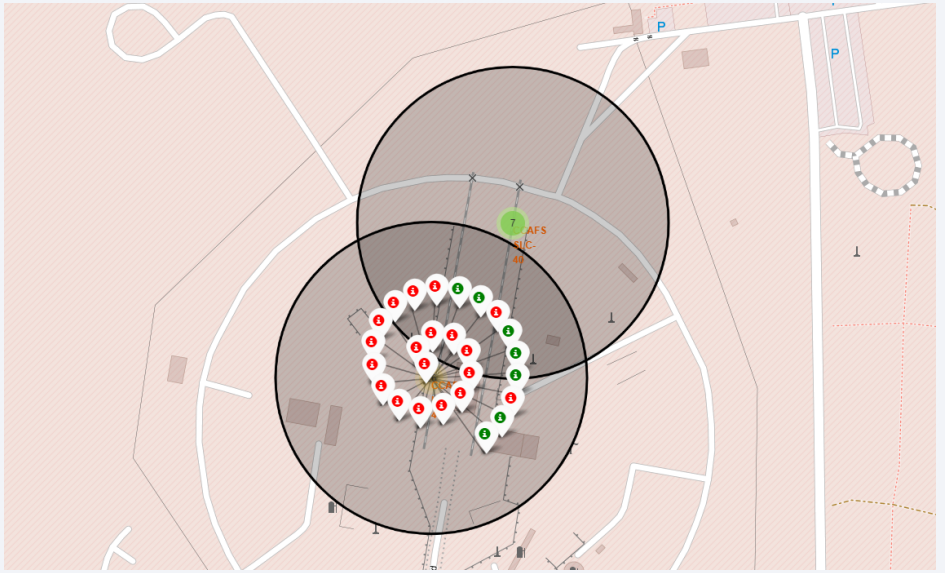
- Wee can see that all launch sites in USA.



Launch Records

- Green ones are succesfull and red ones are failures





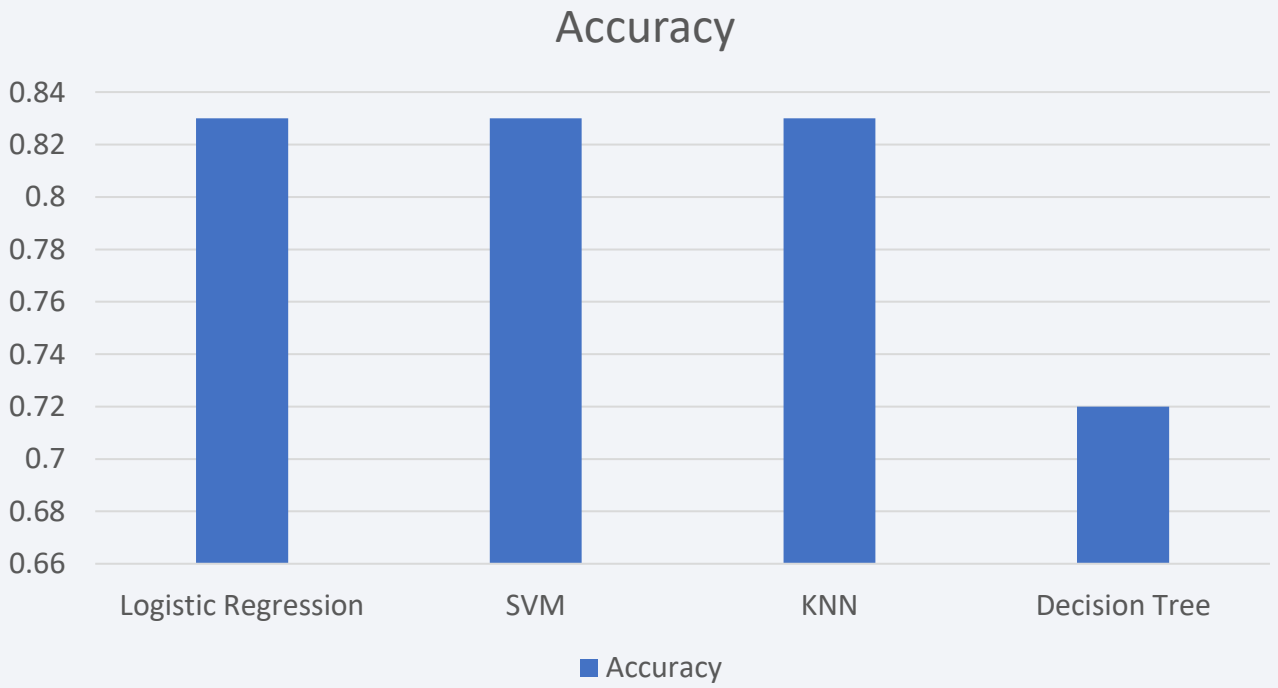
Launch site distances to important locations

Section 4

Predictive Analysis (Classification)

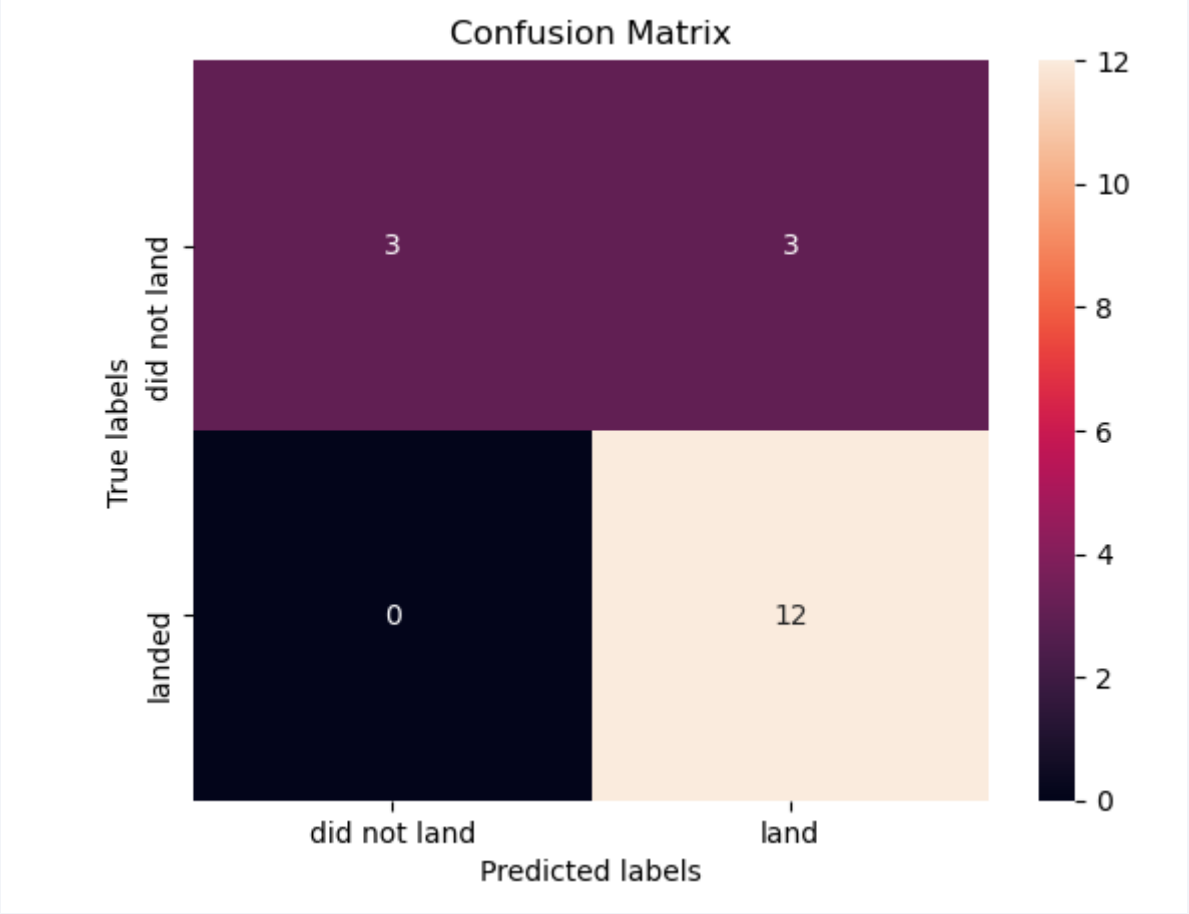
Classification Accuracy

- Accuracy of Decision Tree less than the other trained models.



Confusion Matrix

- Confusion matrix is same for all models



Conclusions

- Succes rate of launches increasing with passing years
- Other than decision tree other models quite satisfying
- Orbits ES-L1, GEO, HEO and SSO have high success rate

Thank you!