IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Patric William F. da Silva
8 Jul 2024

Outline

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

Executive Summary

- Summary of methodologies
  - Collection of data from API's and web scraping
  - Cleaning and formatting the data
  - Finding relations between variables or features with exploratory data analysis (EDA) using visualization and SQL
  - Building interactive visual analytics using Folium (for Maps) and Plotly Dash (for Dashboards)
  - Predicting outcomes using classification models
- Summary of all results
  - Exploratory data analysis results
  - Demo of interactive analytics
  - Outcome of predictive analysis

Introduction

- Now a Days there is high competition between Space Agencies to launch rockets In low costs. One of them is SpaceX who aims to make reusable First Stage of Falcon 9 rockets which can reduce the cost of each launch.
- With the help of the past data we can predict that the First Stage of Falcon 9 will land Successfully or not ?
- This prediction will help us to determine the overall cost of each launch.
- Further we can use this information if an alternate company wants to bid against SpaceX for a rocket launch.

Section 1

Methodology

Methodology

Executive Summary

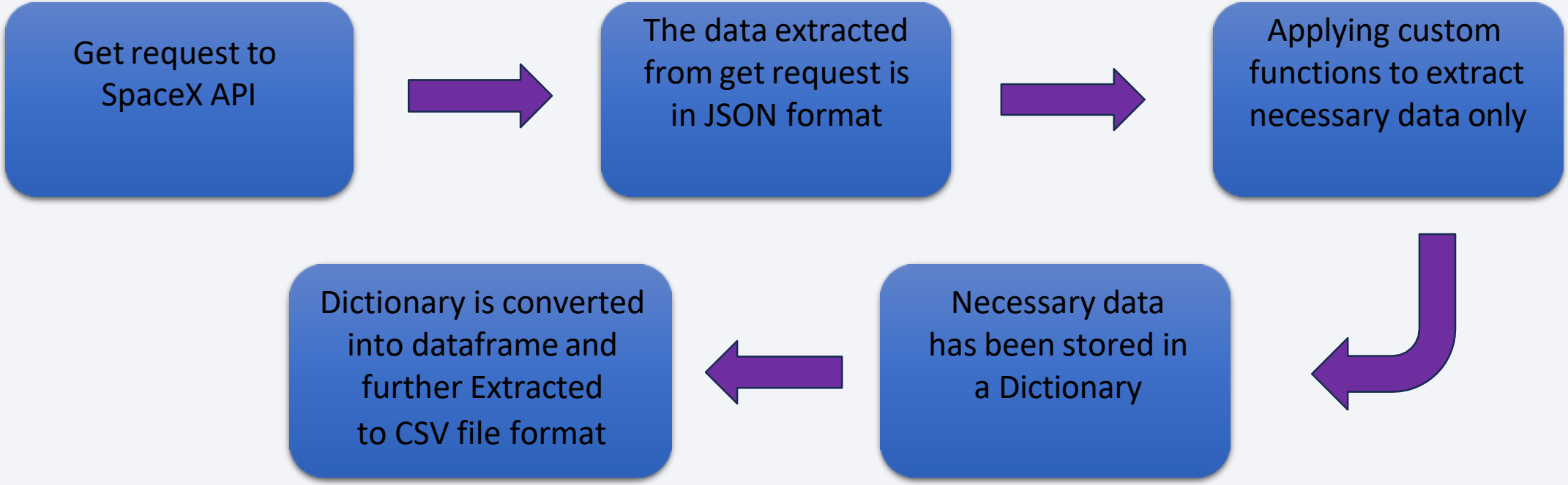
- Data collection methodology:
  - Data has been collected using following methods -
    - SpaceX Rest API
    - Web scraping from Wikipedia
- Perform data wrangling
  - Data has been processed by dealing with unavailable records/features, formatting the features with proper type, cleaning the noise from data.
- Exploratory data analysis (EDA) using visualization and SQL
- Interactive visual analytics using Folium and Plotly Dash
- Predictive analysis using classification models

  Building the classifier model and tuning it with proper algorithms to increase accuracy. Also using part of data for training and remaining for testing to evaluate model.

Data Collection

- For the purpose of this analysis we required data of Falcon 9 rockets. This data can be collected through web pages and API's.
- We are going to use following methods to extract data :-
    1. Collecting data from SpaceX Rest API
    2. Collecting data from Wikipedia with web scraping

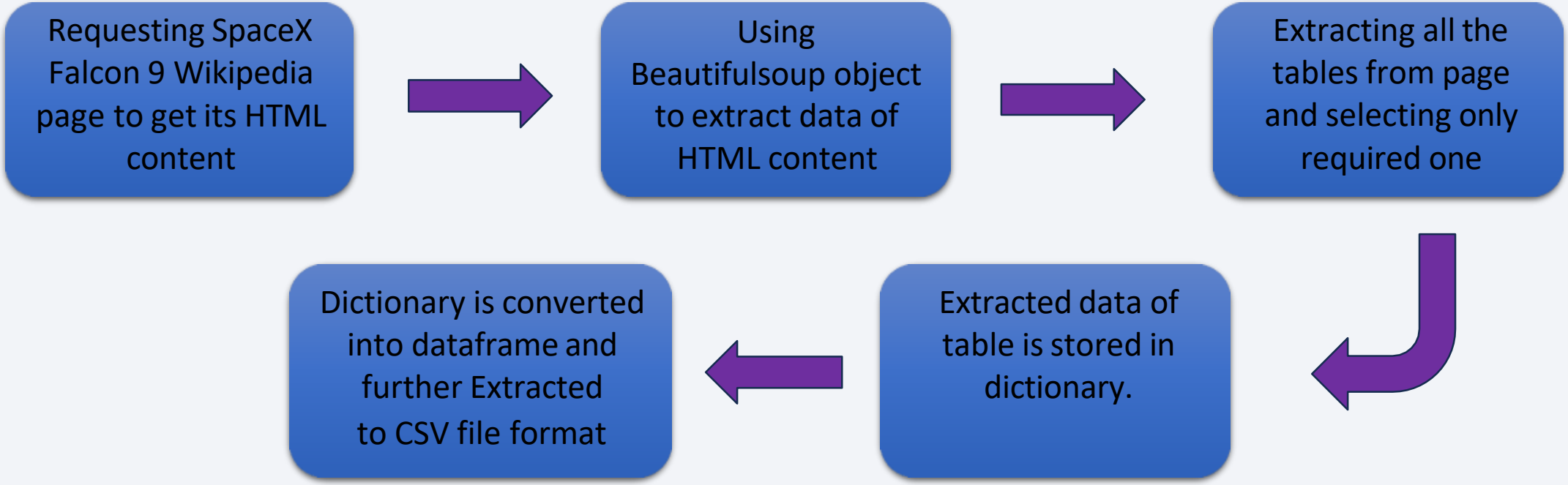
Data Collection - SpaceX API



The data has been extracted by sending the Get request to SpaceX API, Which contains the data in JSON format.

Github URL

Data Collection - Scraping

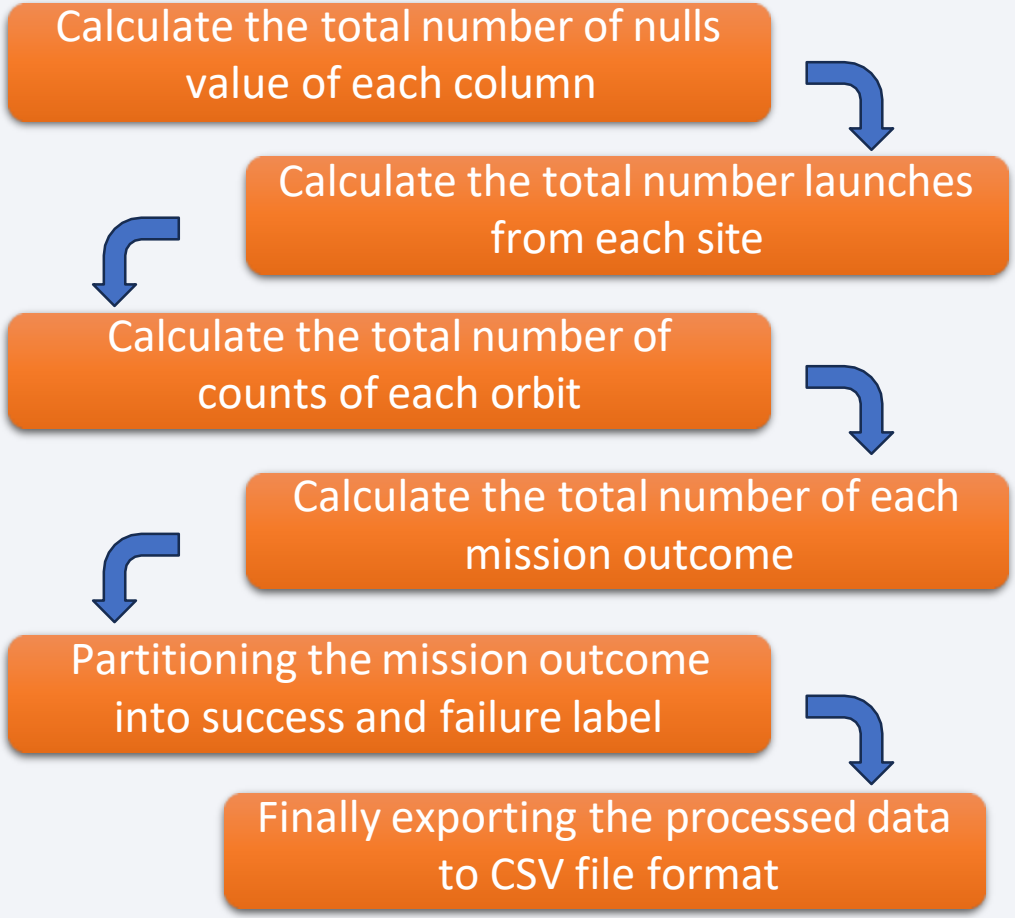


BeautifulSoup is one of the popular model used for web scraping. With this module we can extract required data from HTML page content.

Github URL

Data Wrangling

- The data collected from API's and web scraping contains several impurities and noise which can affects our model accuracy and precision.
- There may also missing values and improper data formats in dataset
- The dataset contains the column named 'Landing Outcome' which contains different categorical values in the form of outcome from landing of first stage. E.g. 'True Ocean', 'False Ocean', 'None None'.
- We will mainly convert those outcomes into Training Labels with '1' means the booster successfully landed '0' means it was unsuccessful.



- Github URL

EDA with Data Visualization

- To find trends from different variables of SpaceX dataset we use categorical plots, time series analysis.
- By using Categorical plots we plotted following plots -
  - PayloadMass Vs FlightNumber
  - FlightNumber Vs LaunchSite
  - PayloadMass Vs LaunchSite
  - FlightNumber Vs Orbit
  - PayloadMass Vs Orbit
- For time series analysis we plotted Year Vs SuccessRate.
- Using this Visuals we can answer most of the questions based on the patterns found in dataset.
- Github URL

EDA with SQL

- SQL queries performed on dataset -
  - Displaying the names of unique launch sites in the space mission
  - Displaying 5 records which have names of launch sites starting with CCA
  - 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 when the first succesful landing outcome in ground pad was acheived
  - Listing the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
  - Listing the total number of successful and failure mission outcomes
  - Listing the names of the booster_versions which have carried the maximum payload mass
  - Listing the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.
  - Ranking 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.

Build an Interactive Map with Folium

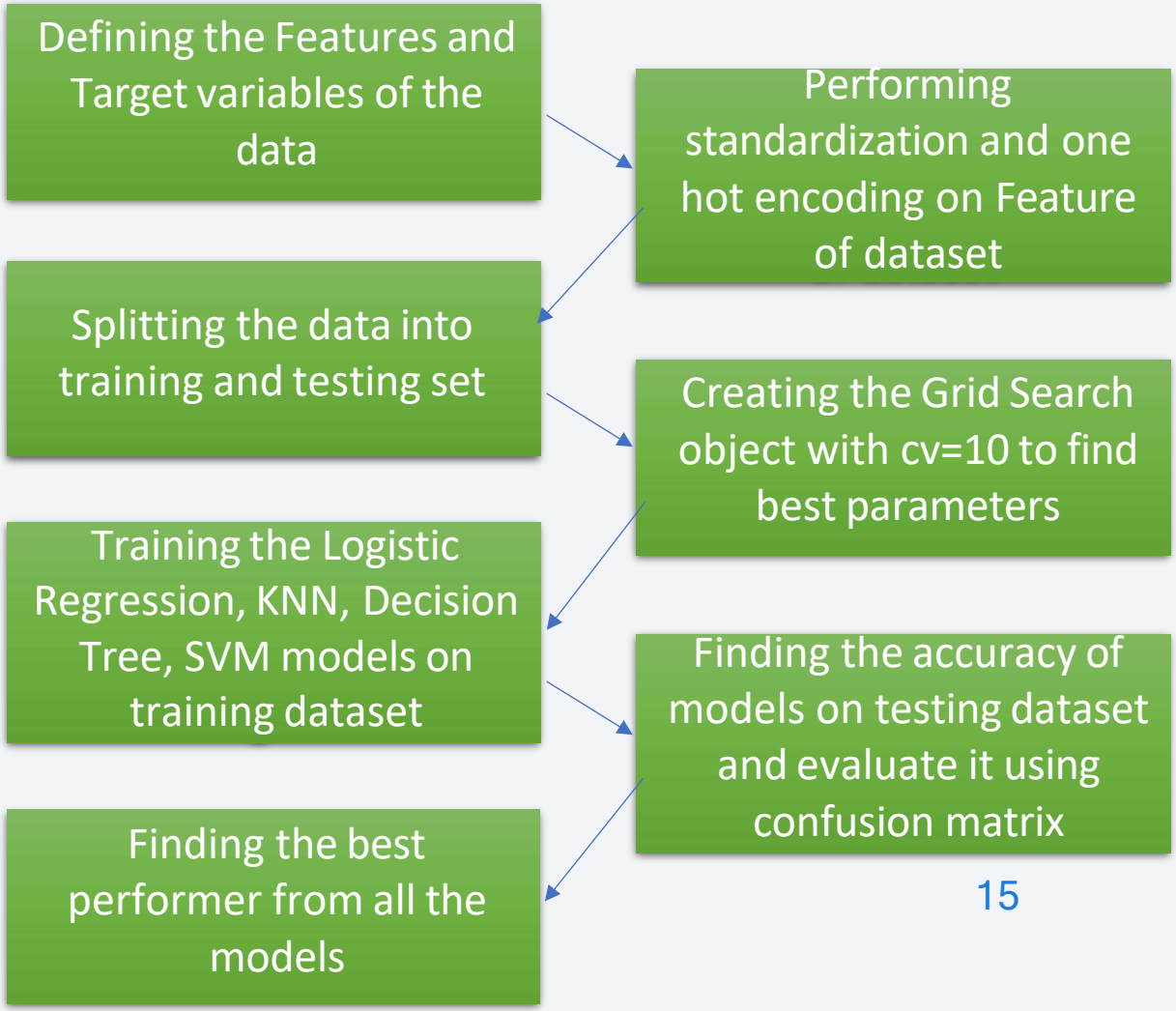
- Building the interactive map of launch sites to perform analysis of proximities
- Objects used in map -
  - Circle - To highlight the launch site
  - Marker- To denote the location
  - Icon- To Customize the marker
  - Cluster - To plot many markers on a same location
  - Mouse Position- To display co-ordinates of mouse pointer in map
  - Polyline - To plot the line between launch site and its proximities
- Displaying the distance between launch site and its proximities
- Displaying the Successful landings in Green marker and Failed landings in Red marker for each landing site
- Github URL

Build a Dashboard with Plotly Dash

- Plots/Graphs used in Dashboard -
  - Pie Chart - To display the success rate for each site by providing dropdown to select landing site
  - Scatter Plot – To display the relations between mass of payload and class that determines outcome of landing by providing slider which decides the range of payload mass
- By using this interactive plots/graphs we can analyze data for each site effectively which will helpful for further predictions.
- Github URL

Predictive Analysis (Classification)

- For predictive analysis we used different classification algorithms provided by scikit-learn library.
- For the process of evaluation data has been split into two parts i.e. training data and testing data.
- Finally to improve the performance of model the Grid Search algorithm is used which fits the model with best parameters.
- Github URL

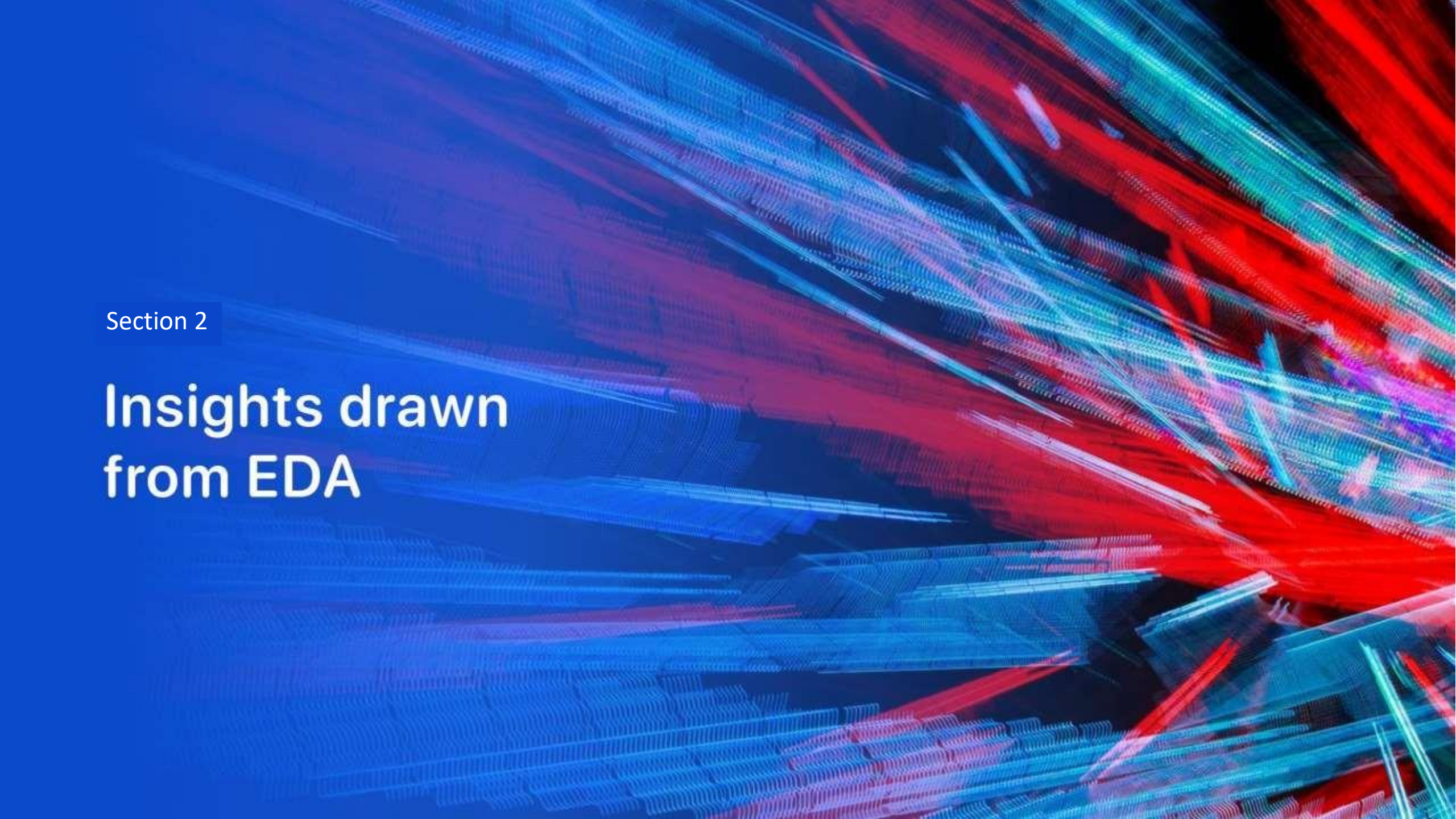


Results

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

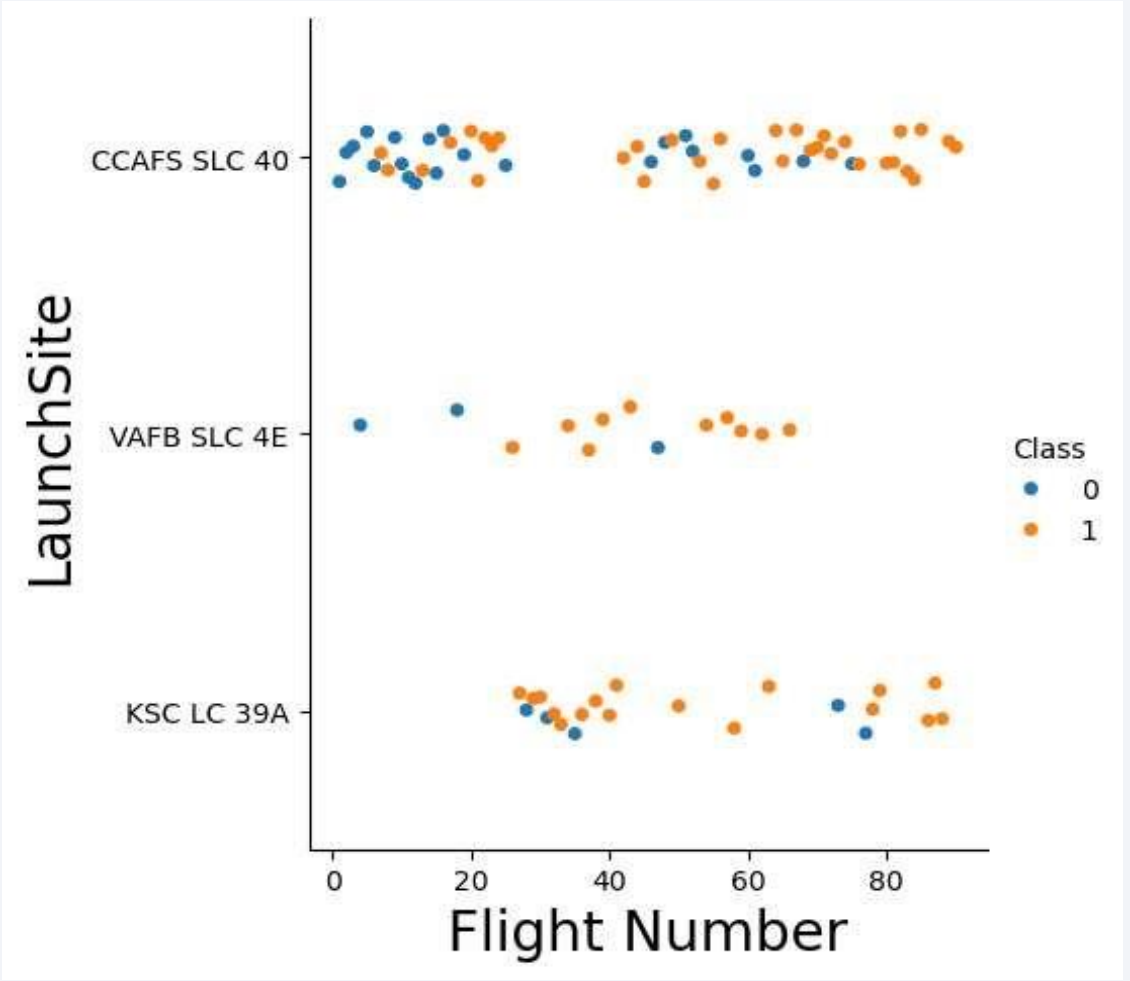
Section 2

Insights drawn from EDA



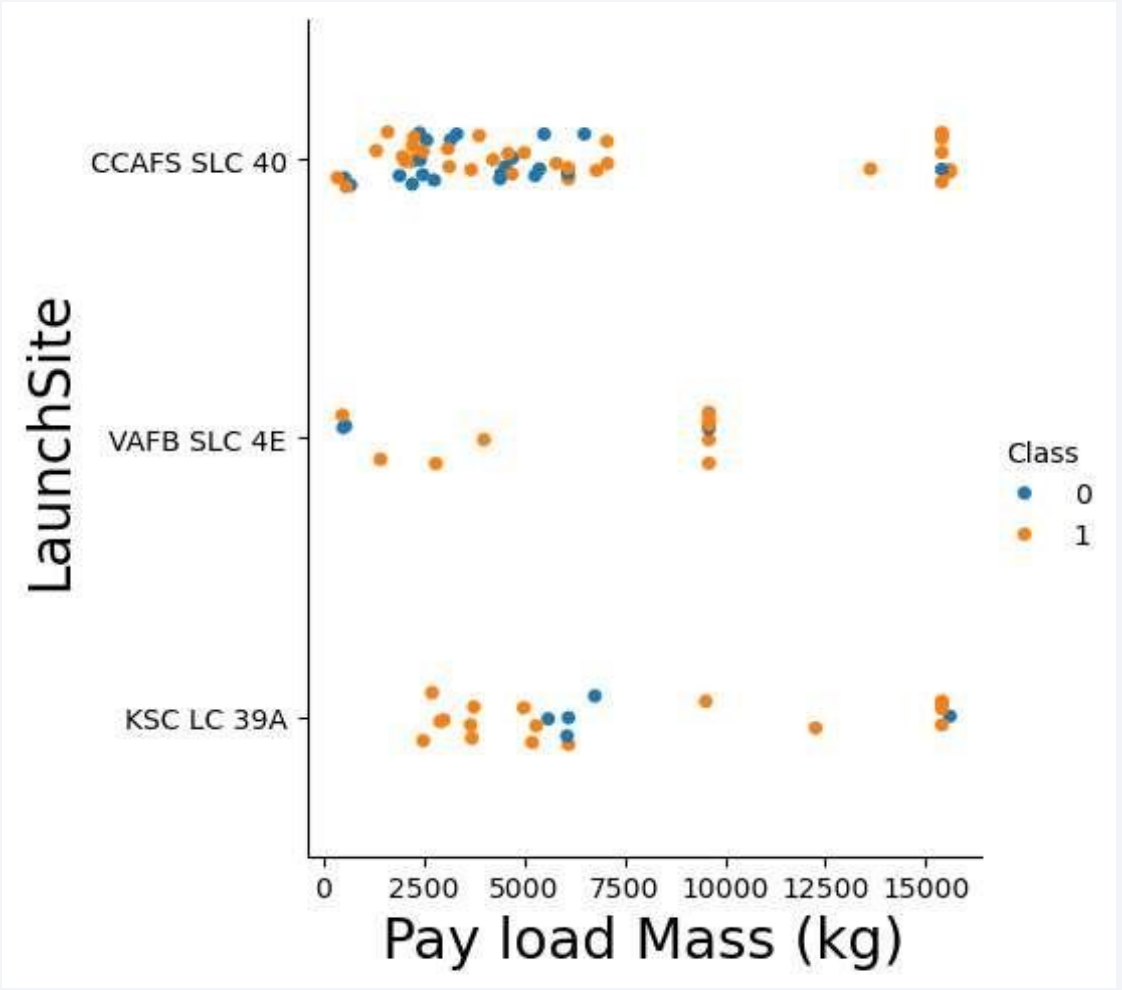
Flight Number vs. Launch Site

- This plot is a representation of relation between flight number and launch site
- Insights drawn from this plot -
  - Most of the launches are Carried out at 'CCAFS SLC 40' launch site
  - As the number of flight increases at each site, the success rate also increases
  - At the initial number of flights the failure rate is high



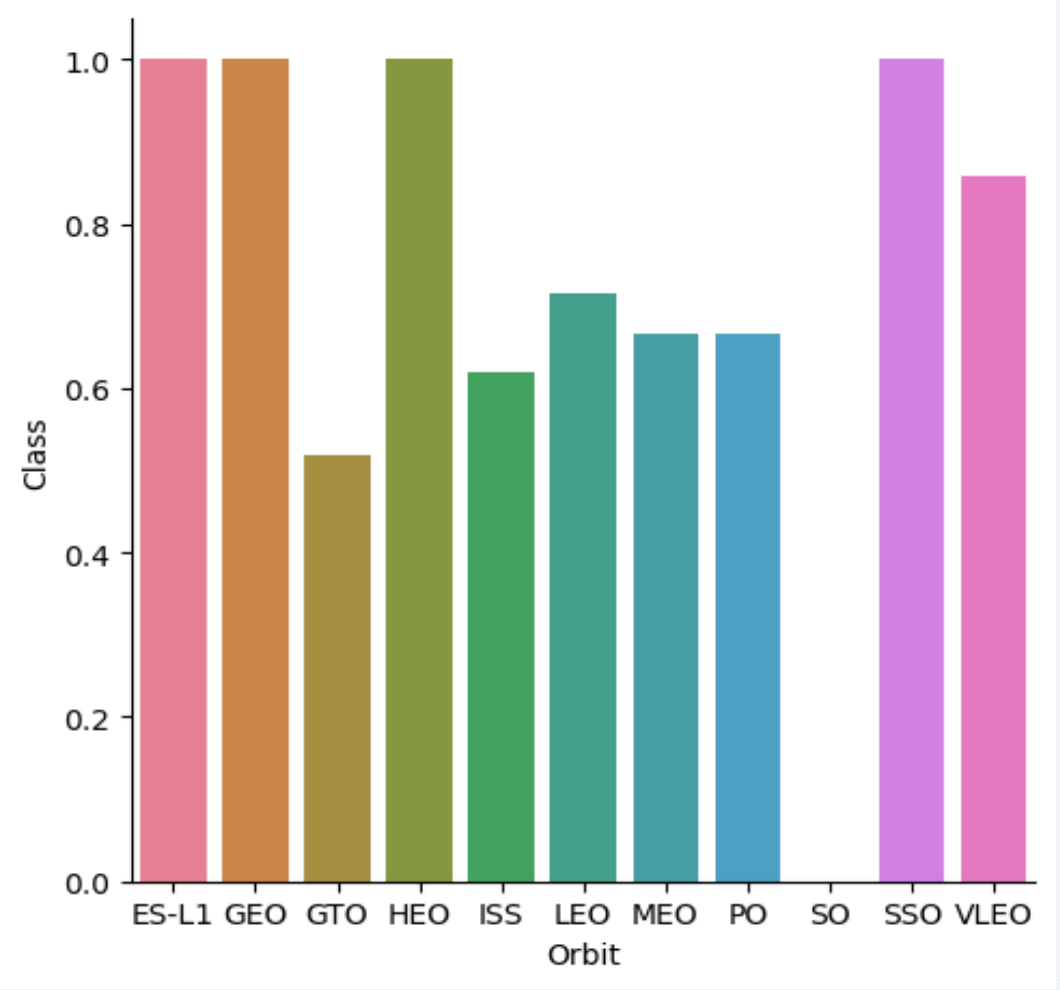
Payload vs. Launch Site

- This plot is a representation of relation between Payload Mass and launch site
- Insights drawn from this plot -
  - Most of the launches are Carried out for payloads having mass in range 2500 to 7500 kg
  - The launches with payload mass above 10000 kg, the success rate is high
  - For 'KSC LC 39A' launch site the success rate is 100% for payload mass below 5000 kg
  - At launch site 'VAFB SLC 4E', there is no any launches for payload mass greater than 10000 kg

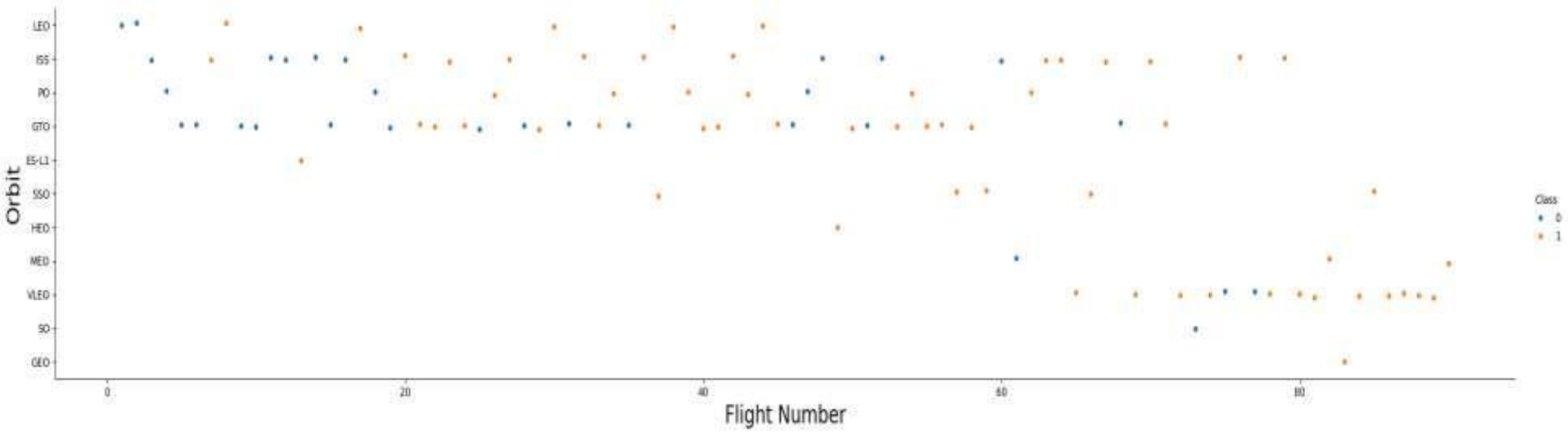


Success Rate vs. Orbit Type

- This plot is a representation of relation between success rate and orbit
- Insights drawn from this plot -
  - For the orbit 'ES', 'GEO', 'HEO' and 'SSO' the success rate is 100%
  - For orbit 'SO' the success rate is totally 0%
  - 'VLEO' orbit have also high success rate greater than 80%



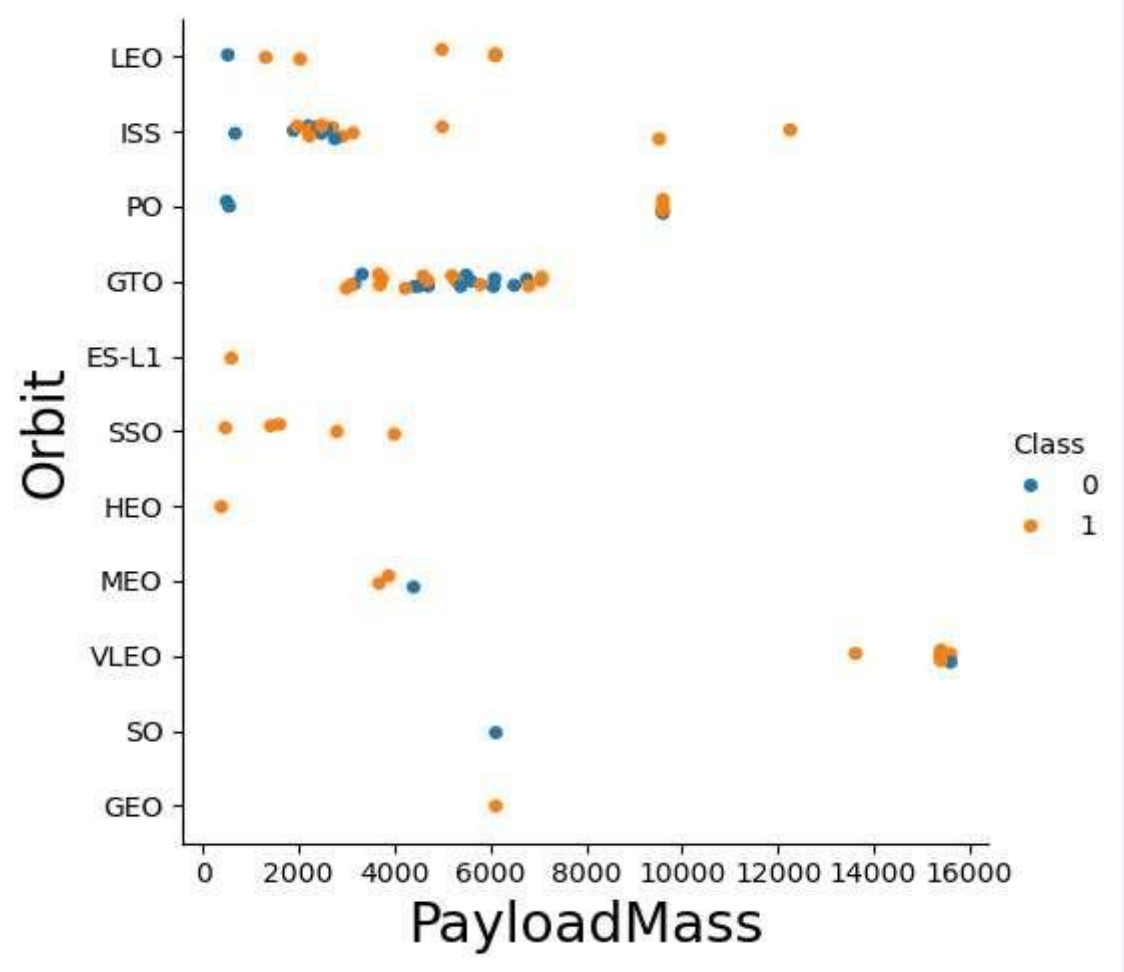
Flight Number vs. Orbit Type



- This plot is a representation of flight number and orbit
- Insights drawn from this plot -
  - For first 20 flights the success rate is lowest
  - For orbit 'VLEO' and 'LEO' the success rate is high
  - After first 60 flights the success rate is in increasing order
  - For the orbit 'ISS' and 'GTO' the number of flights are highest

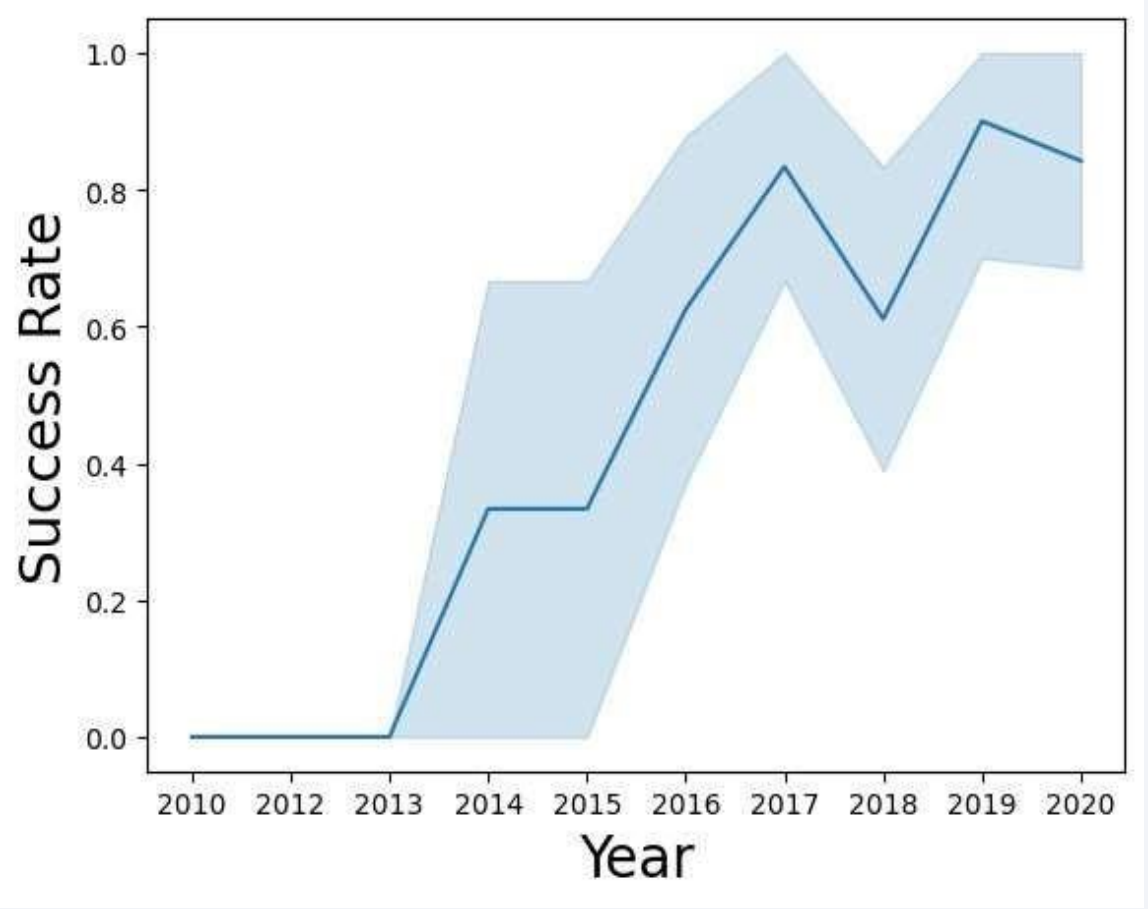
Payload vs. Orbit Type

- This plot is a representation of relation between payload mass and orbit
- Insights drawn from this plot -
  - 'VLEO' orbit has the flights which have payload mass greater than 13000 kg
  - Most of the orbits has flights having payload mass less than 8000 kg
  - As the payload mass increases in 'ISS' orbit the success rate also increases



Launch Success Yearly Trend

- This plot is a representation of yearly trend in launch success
- Insights drawn from this plot -
  - From year 2013 the success rate is rapidly increasing
  - In year 2014 to 2015 the success rate is constant
  - During year 2017 to 2018 there is decrease in success rate
  - Up to year 2013 the success rate is totally 0%



All Launch Site Names

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

- There are total 4 unique launch sites for space mission.

Launch Site Names 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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	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

- 5 records where launch sites begin with `CCA`
- Landing outcome of this 5 records are failed and not attempted

Total Payload Mass

total_payload_mass
45596

- The total payload mass carried by NASA boosters is 45596 Kg with the help of different boosters in rockets

Average Payload Mass by F9 v1.1

AVG(PAYLOAD_MASS__KG_)
2928.4

- There is average payload mass of 2928.4 kg carried by booster version F9 v1.1

First Successful Ground Landing Date

first_successful_landing
2015-12-22

- The first successful ground landing is done on 22 Dec 2015
  - This landing is done on ground pad in the ocean

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

- This boosters had successfully landed on the Drone Ship and payload mass carried by the boosters are in range 4000 to 6000 kg

Total Number of Successful and Failure Mission Outcomes

Mission_Outcome	count
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

- There are total 101 mission are conducted
- And out of this 100 are successful and 1 is failed

Boosters Carried Maximum Payload

- This are listed booster versions that carried maximum payload mass for missions

Booster_Version
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	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

- In 2015 there are 2 failed landings on Drone Ship
- The launches of this missions are done from CCSFS LC-40

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

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

- Ranking the count of landing outcomes between the date 2010-06-04 and 2017-03-20

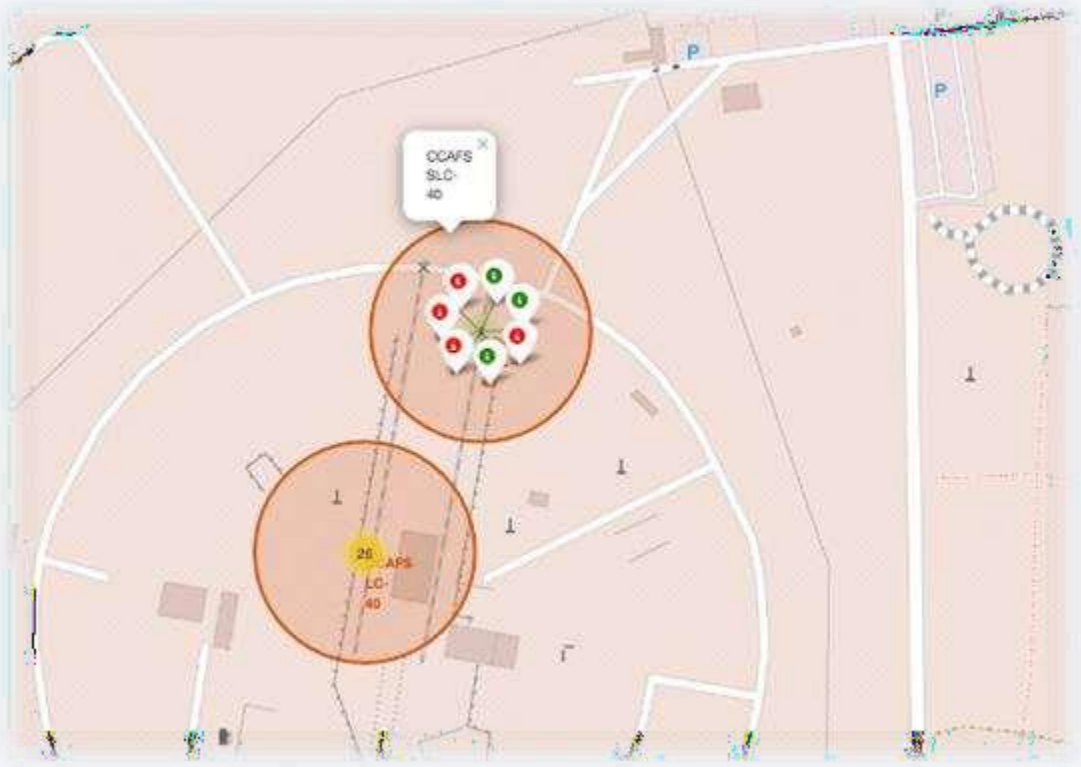
Section 3

Launch Sites Proximities Analysis

Launch sites on global map

- This global map represents the 4 launch sites location
- All the 4 launch sites are located near the coast line area
- This all locations are isolated from populated area which prevent the harm from any disaster

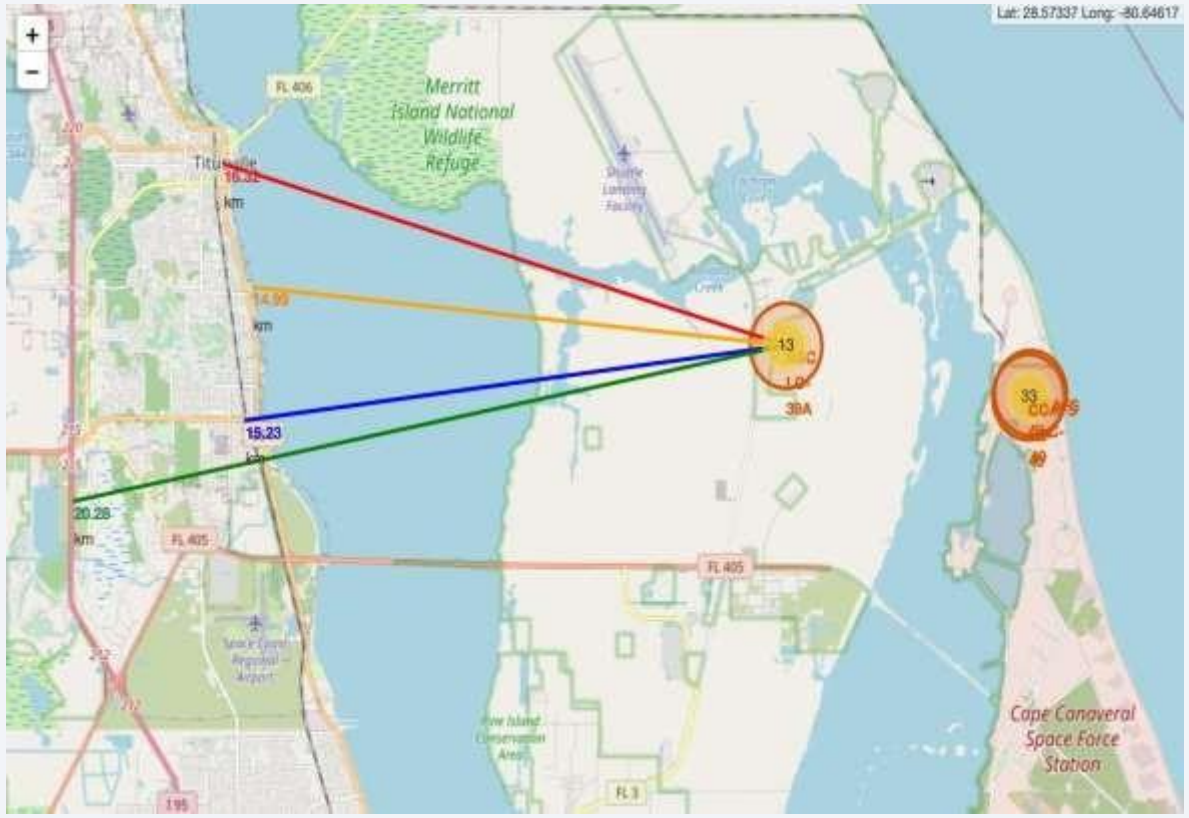
Color-labeled launch outcomes on map



- By using color-labeled markers we have plotted the outcomes on each launch site
- The Green marker represent the success in landing outcome, and Red marker shows the failure on each site
- We can say that launch site KSC LC-39A has highest success rate

Proximities from launch sites

- From the visual analysis of the launch site KSC LC-39A we can clearly see that it is:
  - - relative close to railway (15.23 km)
  - - relative close to highway (20.28 km)
  - - relative close to coastline (14.99 km)
- Also the launch site KSC LC-39A is relative close to its closest city Titusville (16.32 km).
- Failed rocket with its high speed can cover distances like 15-20 km in few seconds. It could be potentially harmful

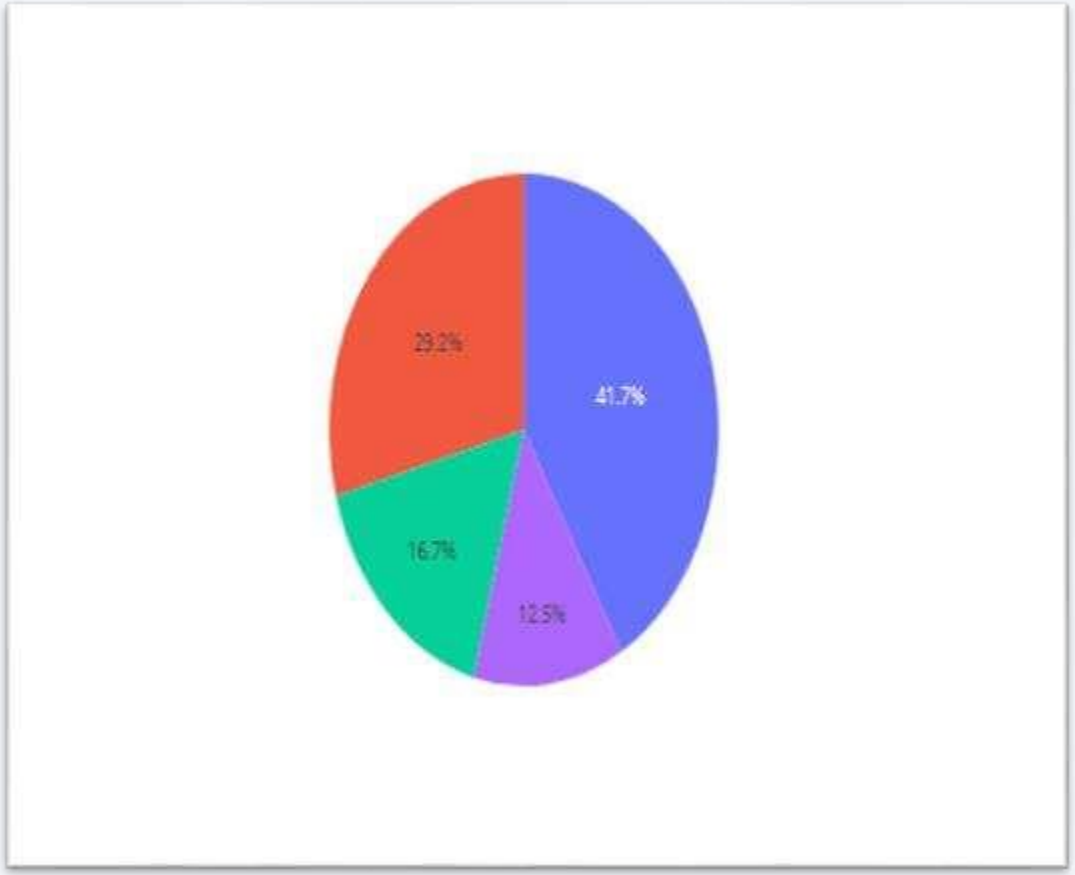


Section 4

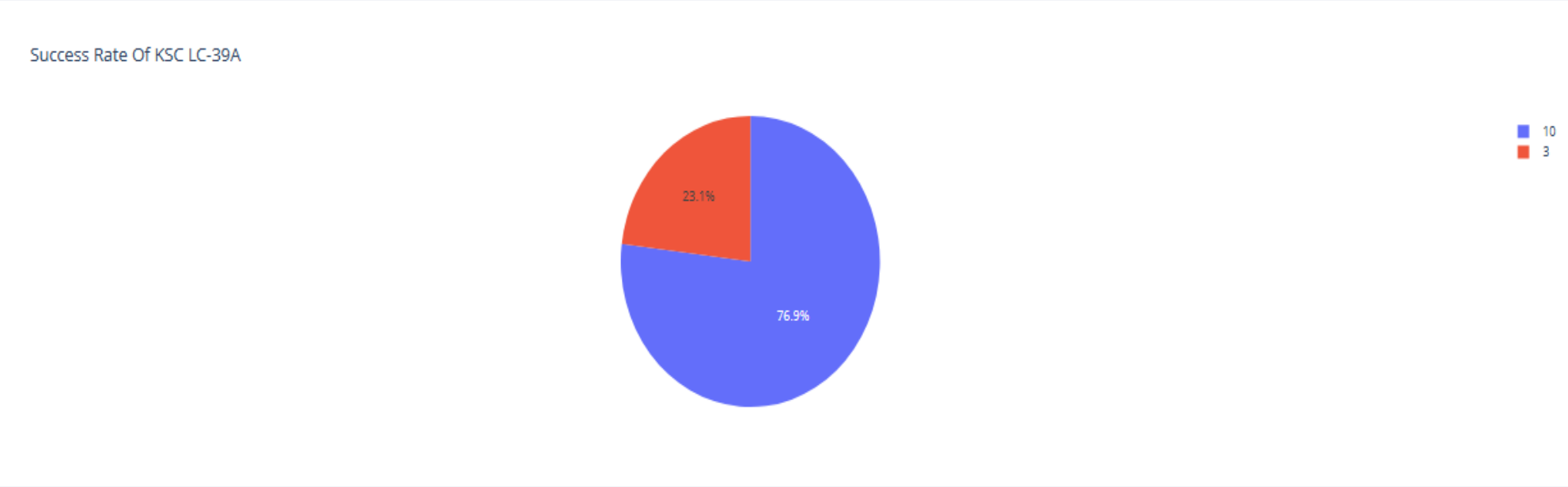
Build a Dashboard with Plotly Dash

Success rate of all launch sites

- This Pie chart represent the success rate of all launch sites
- The highest success rate is of KSC LC-39A launch site
- The lowest success rate is of CCAFS LC-40 launch site



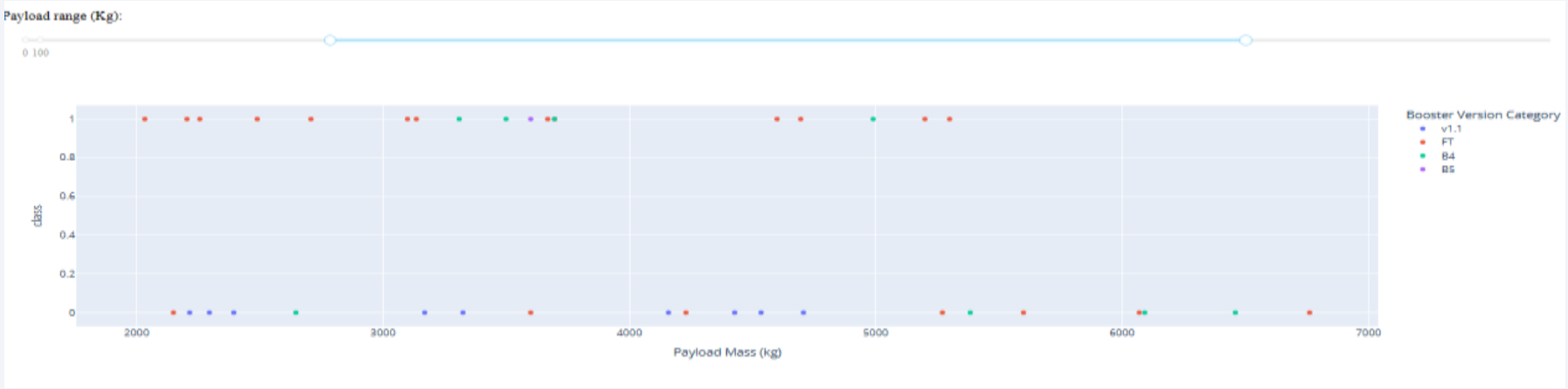
Success rate of KSC LC-39A



- This Pie chart represent the success and failure rate of the KSC LC-39A launch site
- From this we can state that around 76% of launches have success rate to land successfully

Payload Mass Vs Success rate relation

- In payload mass range 2000 to 5000 kg the success rate is highest

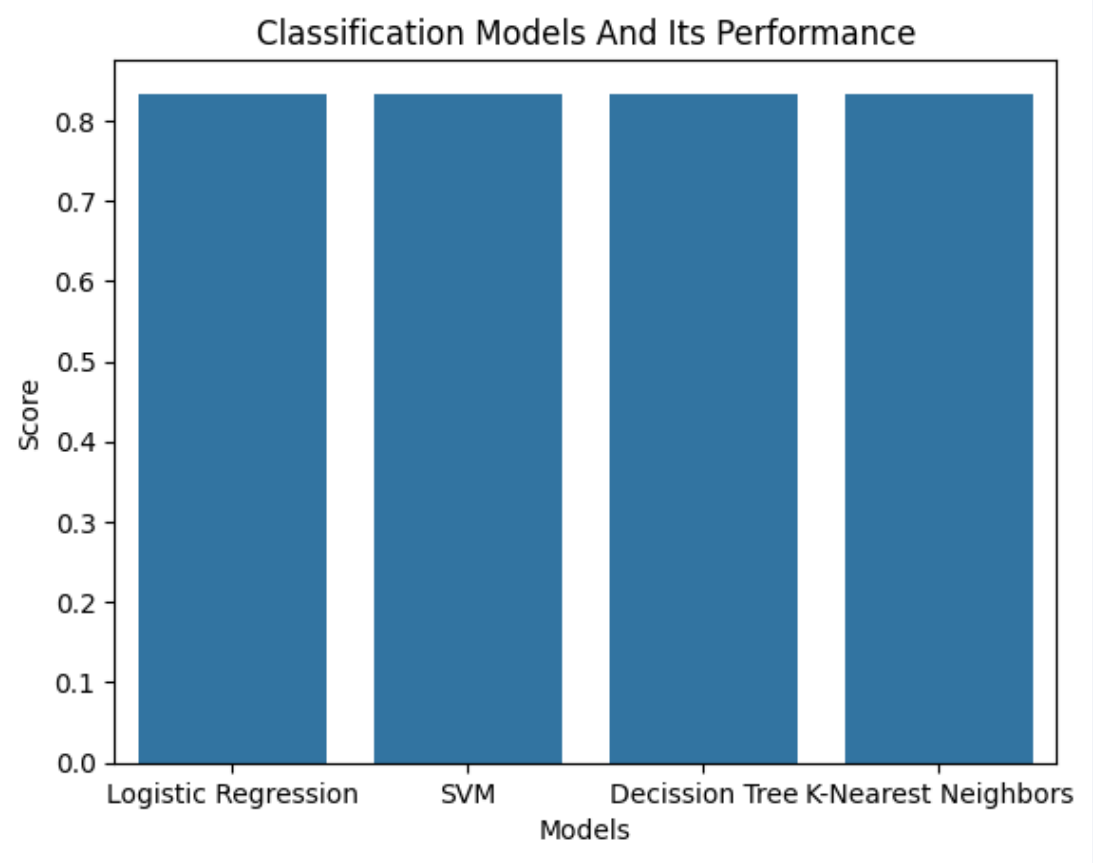


Section 5

Predictive Analysis (Classification)

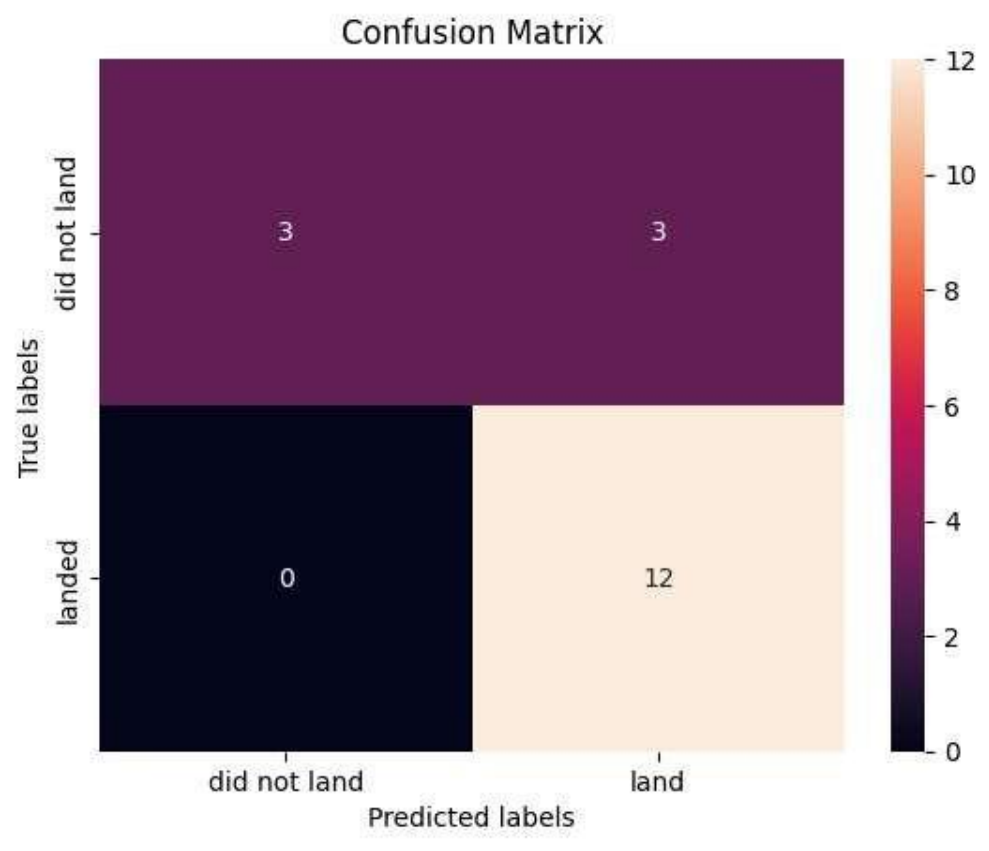
Classification Accuracy

- This bar plot represent the models and its accuracy on the test data set
- We have used Logistic Regression, KNN, Decision Tree and SVM models for classification
- All the models have same accuracy on the test dataset
- The accuracy we have got from this models is 83.33 %



Confusion Matrix

- This confusion matrix shows the distribution of predicted outcome of test records
- There are total 18 test records out of this 15 are correctly classified and 3 are incorrectly classified
- The incorrectly classified records are False Positive



Conclusions

- **From analyzing SpaceX dataset we can state following conclusions -**
- We can use any of the models for classification as we get same accuracy of 83.33%
- The launch site named KSC LC-39A have the highest success rate comparing to remaining sites
- The success rate of landing of first stage increases over the years
- The location of launch sites are near to coastal region and away from populated area
- Orbits ES-L1,SSO,HEO and GEO has success rate of 100%

Appendix

- **References -**
- Wikipedia
- Coursera
- IBM Skills Network
- **GitHub link to notebooks -** https://github.com/PvtricWilliam/SpaceX_landing_prediction.git

Thank you!