Winning Space Race with Data Science

Outline

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

Executive Summary

- Summary of methodologies
  - Collection of Falcon 9 Launch data via SpaceX API, then stored as a csv file for the rest of the analysis
  - Data cleaning, replacing missing values, transforming categorical features into numerical form
  - Visualization of the data (descriptive)
  - Finally, performance of a predictive analysis to train and evaluate a best model and give prediction over the success of future landings
- Summary of all results
  - When Payload was greater than 7500 kg falcon rocket had a higher chance of successful landing
  - Among 11 orbit types , ES L1, GEO, HEO, SSO were 100% successful with less than 6000 kg payload
  - SpaceX has 4 launch sites, one is near California, the other three is near Florida and South Texas. All the sites are in near proximity to ocean and all the sites are bit far away from the city
  - Decision Tree classifier performed well in comparison with other models

Introduction

- Project background and context
- SpaceX is a company that aims to make commercial space travel more affordable for everyone.
- This company can launch rockets for a cost of around 60 million dollars. In contrast, other providers require 165 million dollars for one launch. This is due to the fact that SpaceX can reuse the first stage of the rocket Falcon9.
- In this project, SpaceY is going to compete with SpaceX
- The primary cost saving agent is the high success rate of stage 1 landing and thus its reusability in future launches
- The challenge here is to set a right costing forecast of the rocket launches through predicting its potential to land stage 1 successfully
- Problems you want to find answers
- What are features that contributes the most to predict whether the stage one of the rocket will land successfully?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected using The SpaceX REST API and Wikipedia Web scrapping using python BeautifulSoup
- Perform data wrangling
  - Data was processed using python pandas and numpy
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Classification models (Experiment usability and compatibility of SVM, Tree maps, KNN, Logistic Regression optimizing parameters) were built, evaluated and tuned using sklearn

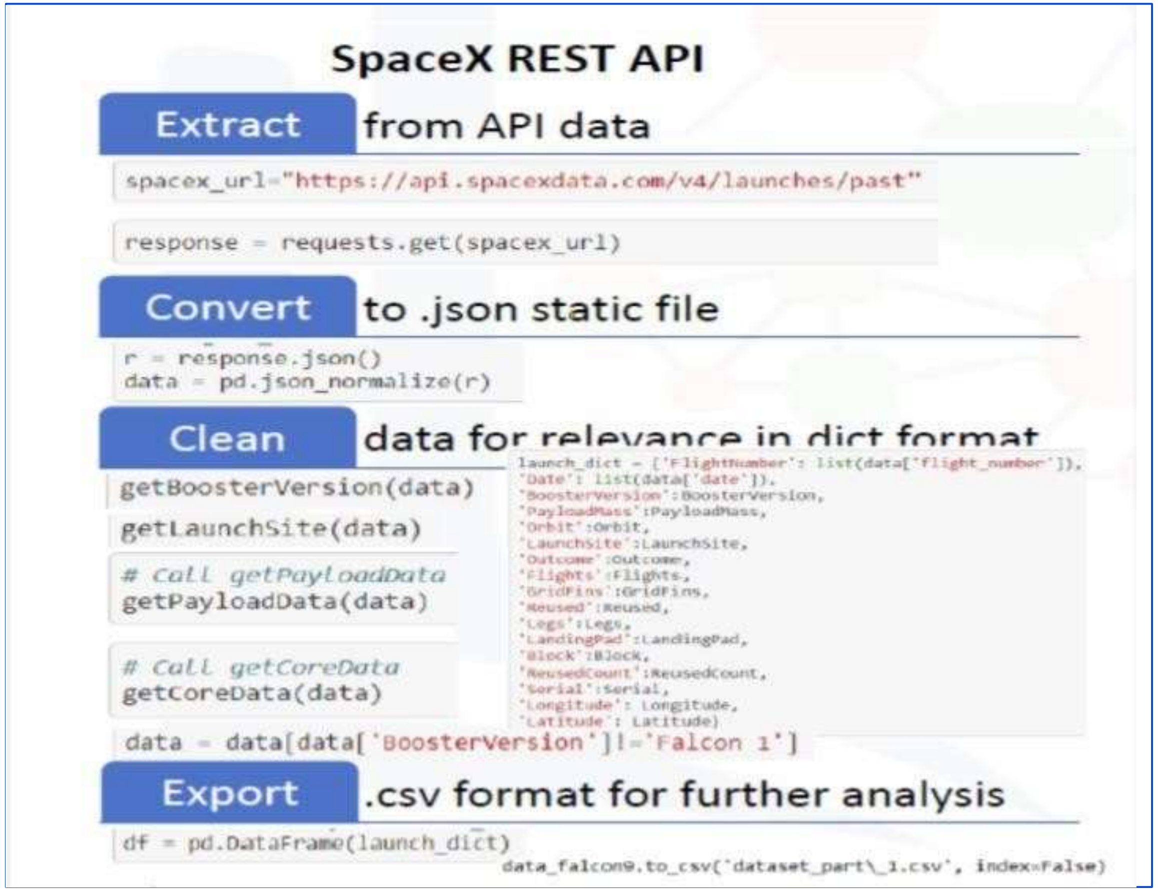
Data Collection

- Describe how data sets were collected.
  - 1. Using SpaceX API
  - 2. Using web scraping

Data Collection – SpaceX API

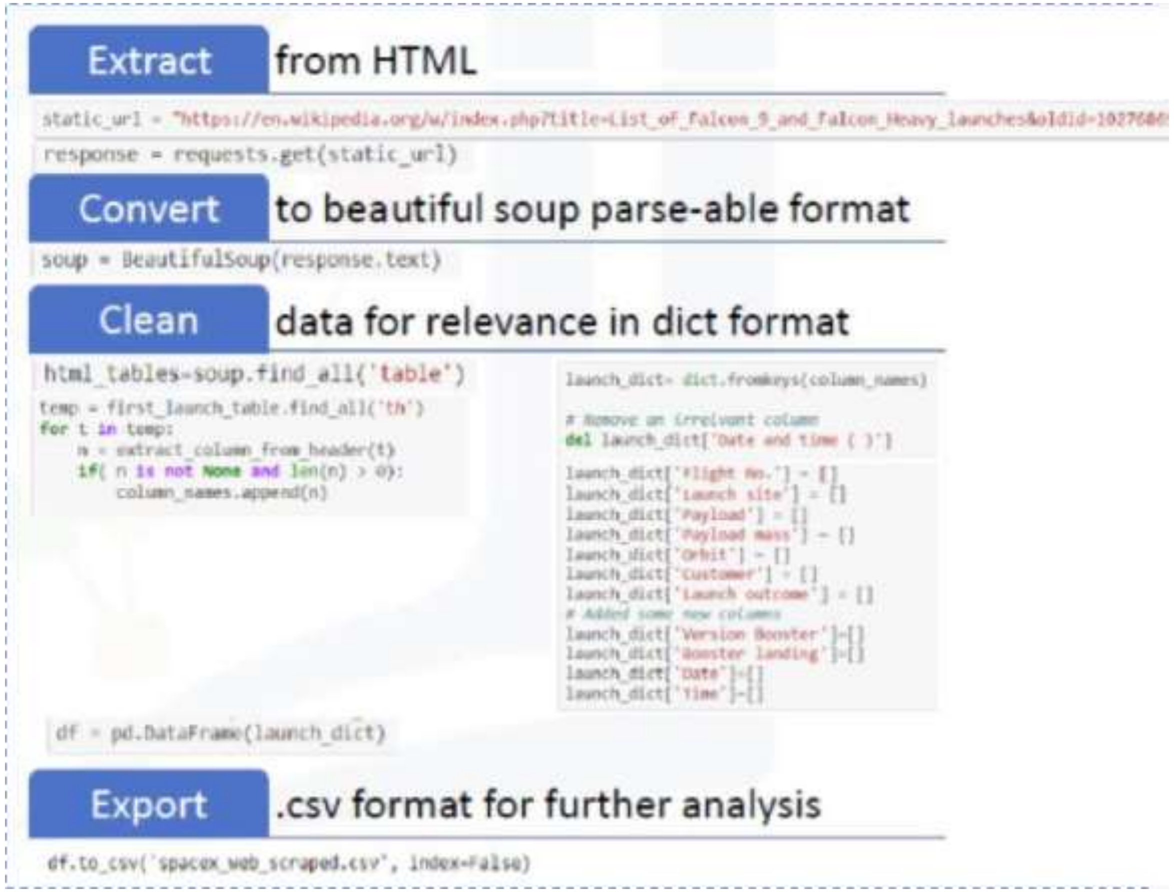
- Request launch data to SpaceX URL
- Parse the data from html text
- Request More data using Launch IDs
- Preprocess and Construct the Data
- Store the Data in CSV File

- GitHub link: https://github.com/rayhanozzy/IBM-Applied-Data-Science-Capstone/blob/main/Collecting%20the%20data.ipynb



Data Collection - Scraping

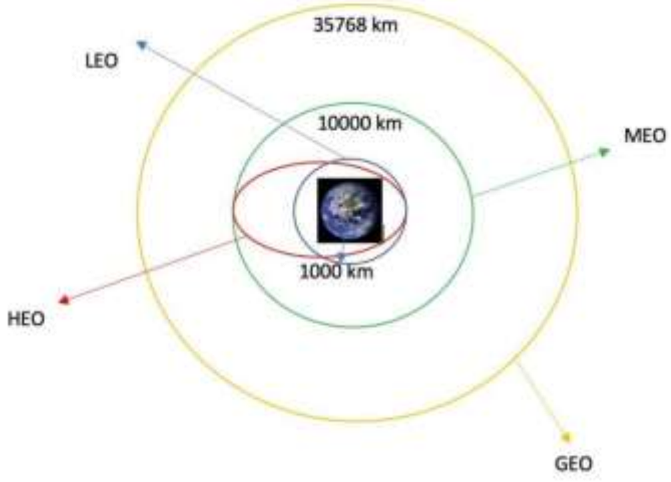
- Request rocket launch data
- Parse the table data from html text using bs4
- Create pandas dataframe from table data
- Construct the Data and
- Store it in CSV
- GitHub link: https://github.com/rayhanozzy/IBM-Applied-Data-Science-Capstone/blob/main/Collecting%20the%20data.ipynb



Data Wrangling

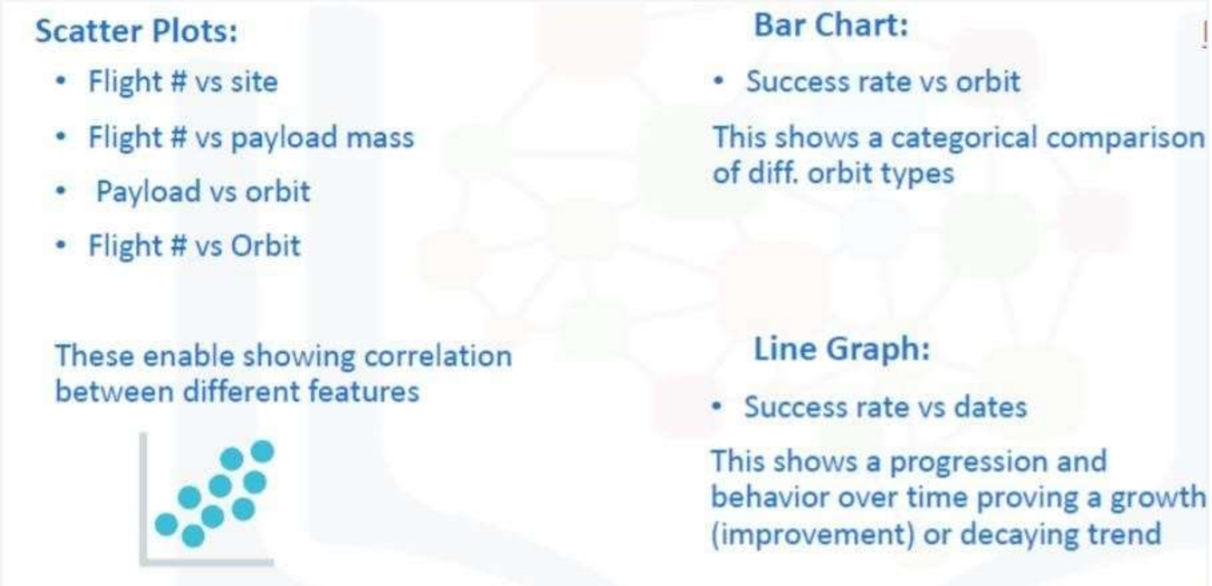
- Describe how data were processed
    1. Calculate the number of launches per site
    2. Number of occurrence of each orbit
    3. Number of occurrences of outcome per orbit
    4. Create landing outcome label
- GitHub: https://github.com/rayhanozzy/IBM-Applied-Data-Science-Capstone/blob/main/Data%20Wrangling.ipynb

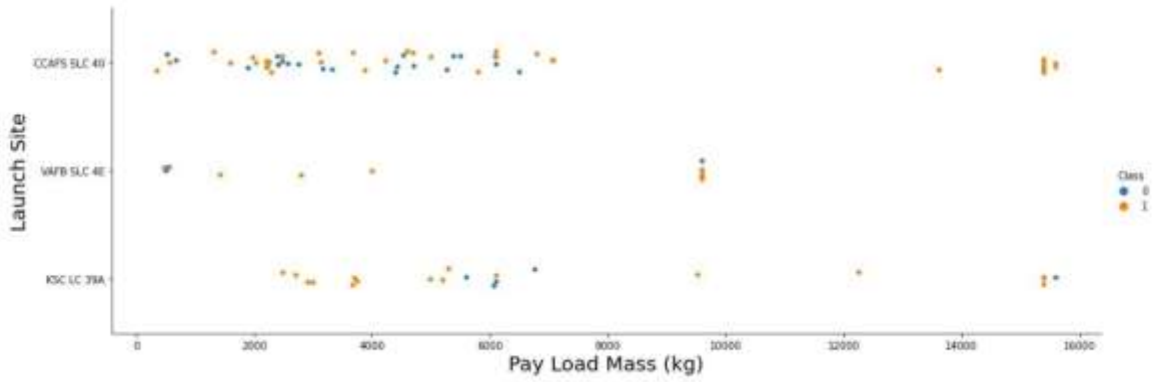
```
In [5]: # Apply value_counts() on column LaunchSite
        df["LaunchSite"].value_counts()

Out[5]: CCAFS SLC 40    55
        KSC LC 39A      22
        VAFB SLC 4E     13
        Name: LaunchSite, dtype: int64
```

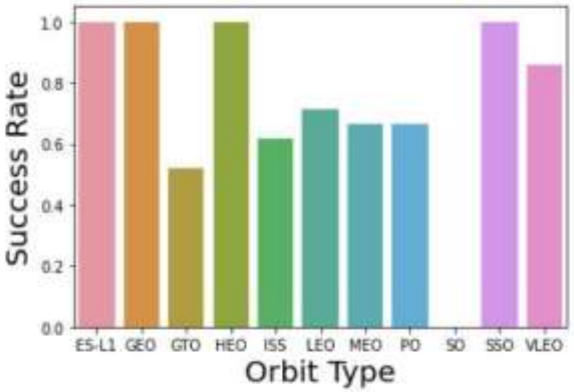


EDA with Data Visualization

- GitHub: https://github.com/rayhanozzy/IBM-Applied-Data-Science-Capstone/blob/main/Data%20visualization%20with%20Folium.ipynb







EDA with SQL

- **SQL queries performed**
- Display the names of the unique launch sites in the space mission
- Display 5 records where launch sites begin with the string 'CCA‘ Link github
- Display the total payload mass carried by boosters launched by NASA (CRS)
- Display average payload mass carried by booster version F9 v1.1
- List the date when the first successful landing outcome in ground pad was achieved
- 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
- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
- 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: https://github.com/rayhanozzy/IBM-Applied-Data-Science-Capstone/blob/main/EDA%20with%20SQL.ipynb**

Build an Interactive Map with Folium

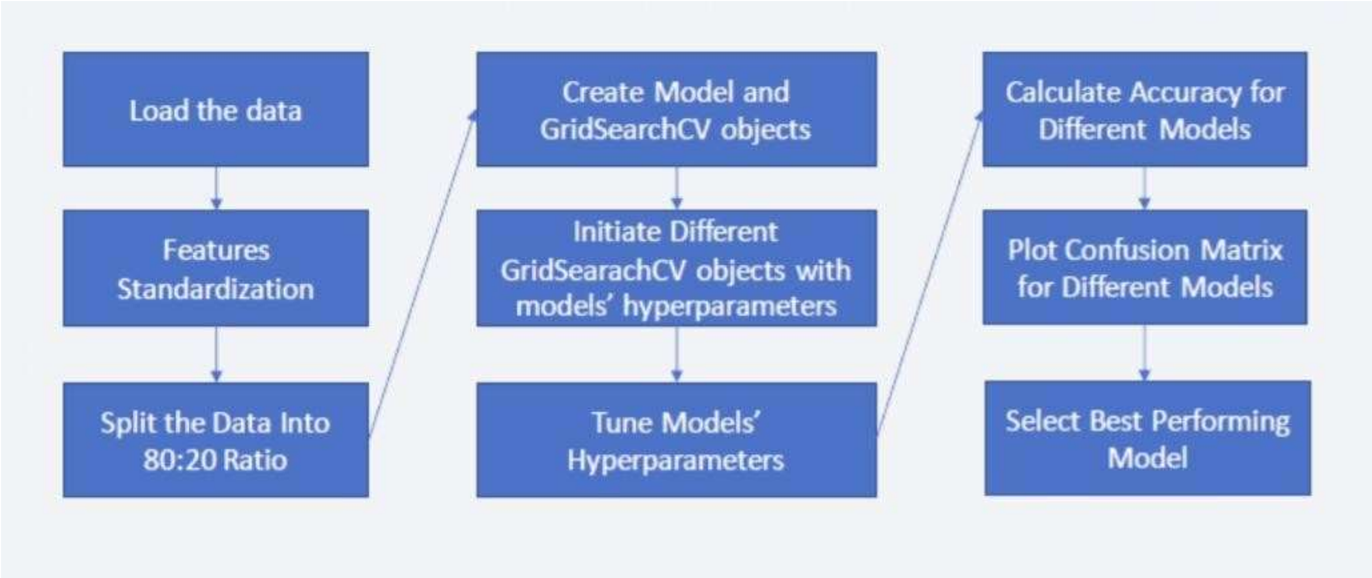
- Map objects which are created and added to the folium map are given below
  - Markers : Added to mark a specific area with a text label on a specific coordinate
  - Circles : Added to highlight circle areas with a text label on a specific coordinate
  - MarkerCluster : Marker clusters were used to simplify the containing many markers having the same coordinates.
  - MousePosition : Used to get coordinate for a mouse over a point on the map (proximities). It helps to find the coordinates easily of any points of interests while exploring the map
  - PolyLine : It draws polyline overlays on a map. It was used to denote the distance between a launch site and its proximities. (such as Railway station, city etc.)
- Github: https://github.com/rayhanozzy/IBM-Applied-Data-Science-Capstone/blob/main/Data%20visualization%20with%20Folium.ipynb

Build a Dashboard with Plotly Dash

- Pie chart (total launches for a selected site or the total sites collection)
  - Shows relative proportions of different sites successful landing distribution
  - Shows % of success vs. failure for a given site
- Scatter Plot
  - Showing the correlation between Outcome and Payload Mass (Kg) for different Booster Versions with freedom of selection of the range of payload mass of interest
- Github: https://github.com/rayhanozzy/IBM-Applied-Data-Science-Capstone/blob/main/Interactive%20Dashboard%20with%20Ploty%20Dash.ipynb

Predictive Analysis (Classification)

- Classification model development
- Github: https://github.com/rayhanozzy/IBM-Applied-Data-Science-Capstone/blob/main/Machine%20Learning%20Prediction.ipynb



Results

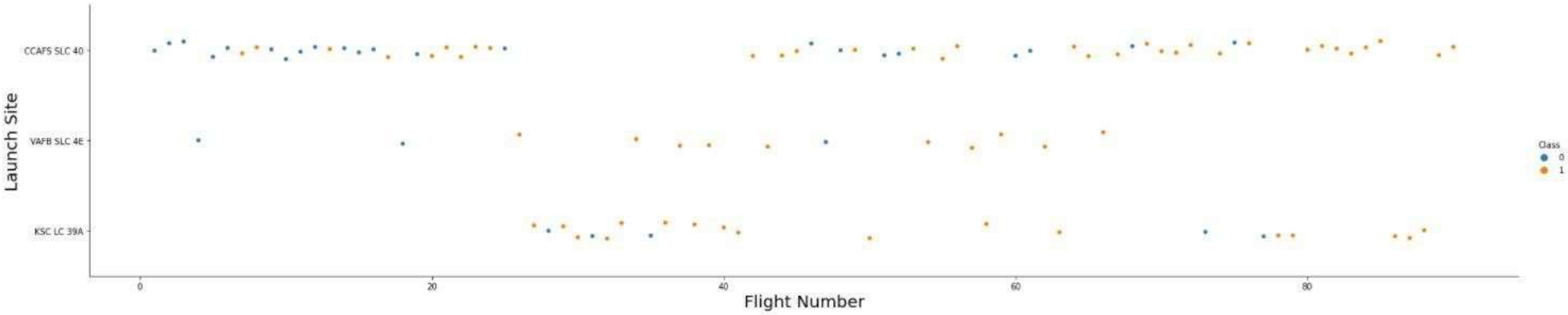
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

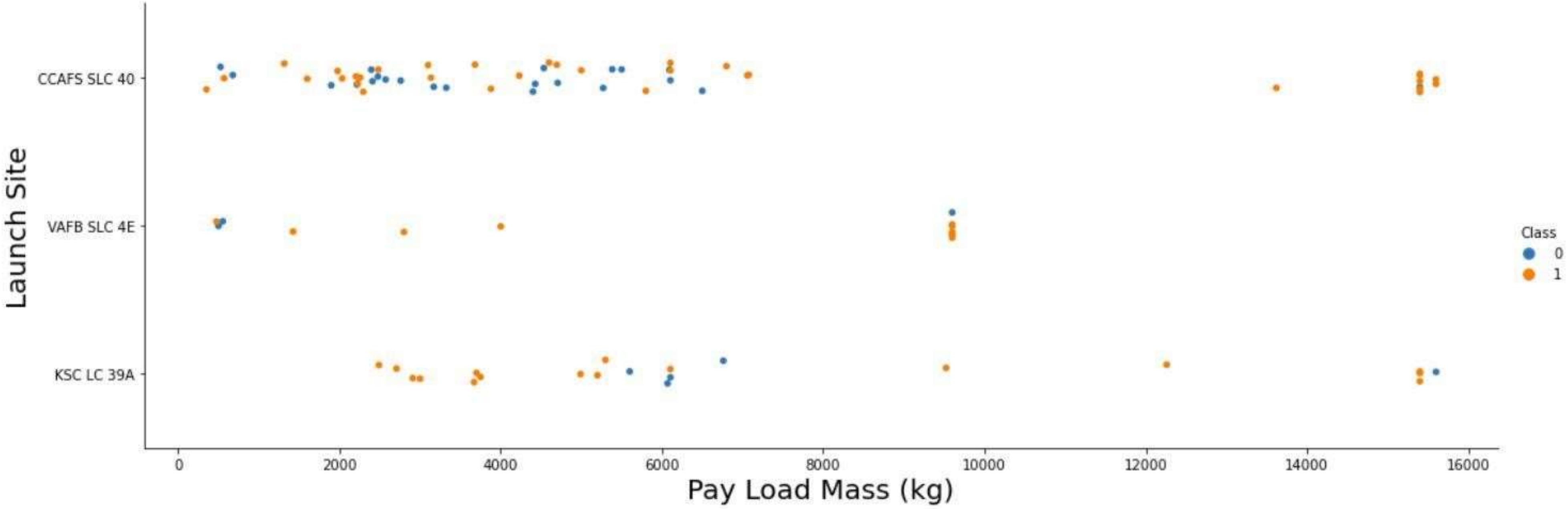
Flight Number vs. Launch Site

- The more orange dots in each horizontal distribution the higher the success rate at a launch site



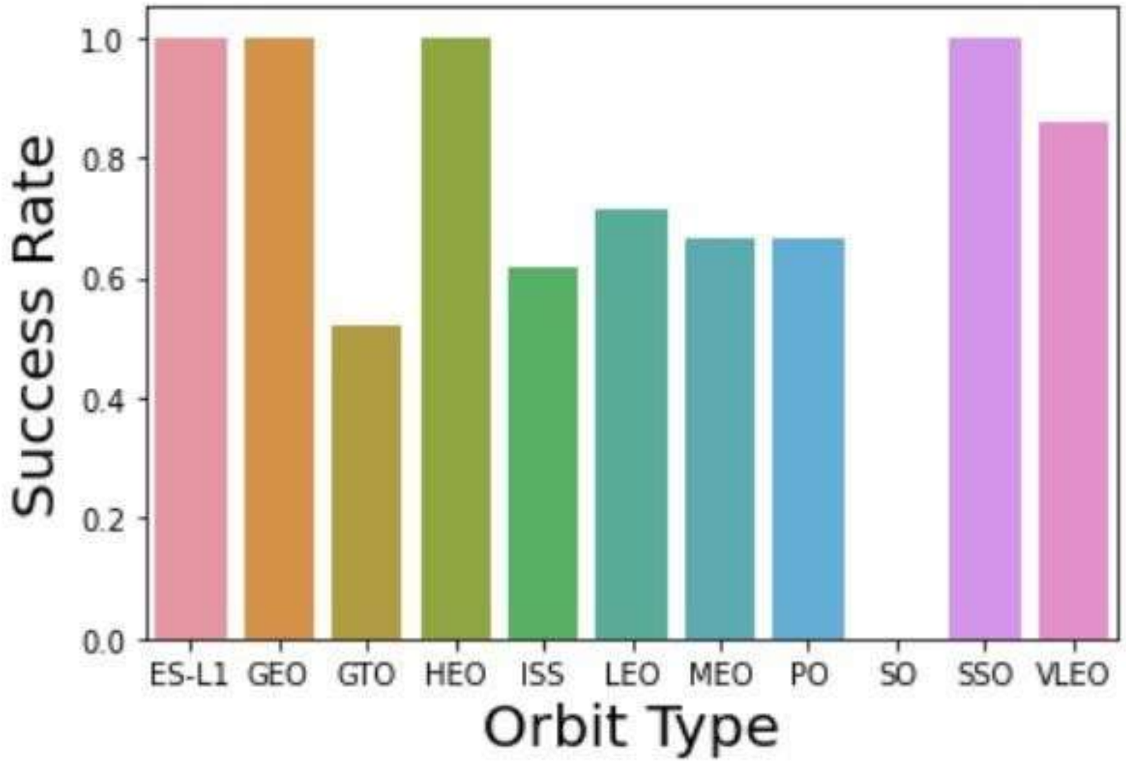
Payload vs. Launch Site

•Looks like the higher the payload mass the greater the chance for a successful landing (orange dots)



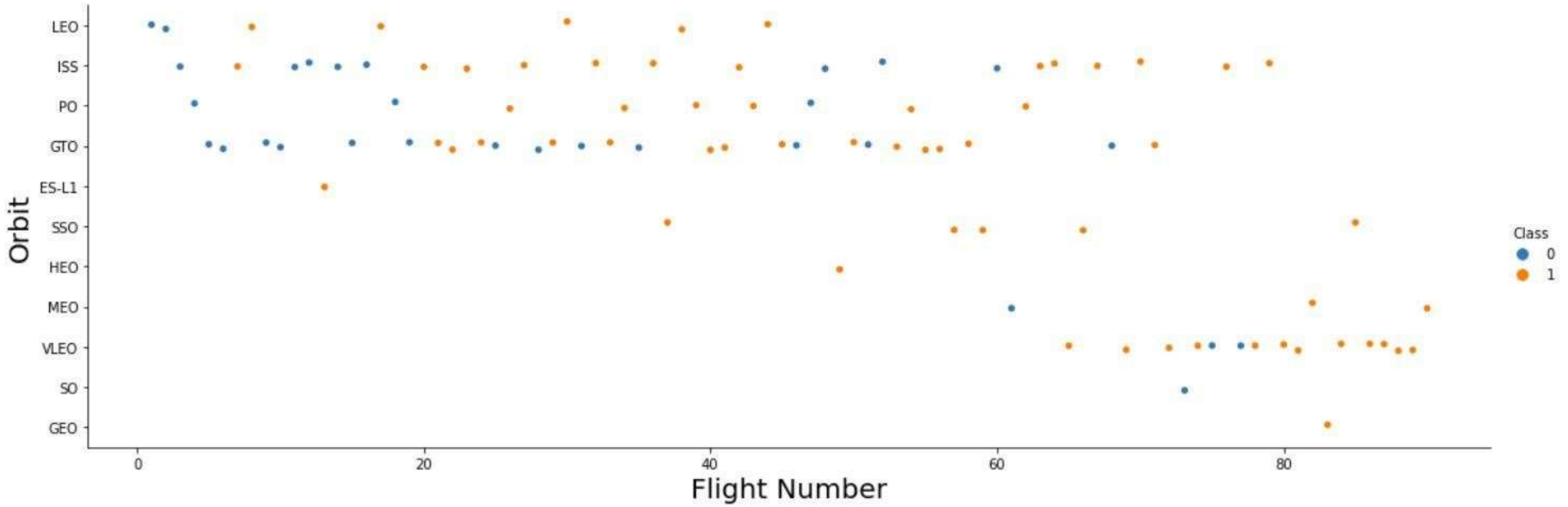
Success Rate vs. Orbit Type

- “SO” orbit had no success
- Orbit ES-L1, GEO,HEO,SSO, have the highest success Rate (100% successful)
- “GTO” had approximately 50% success rate whereas the rest of



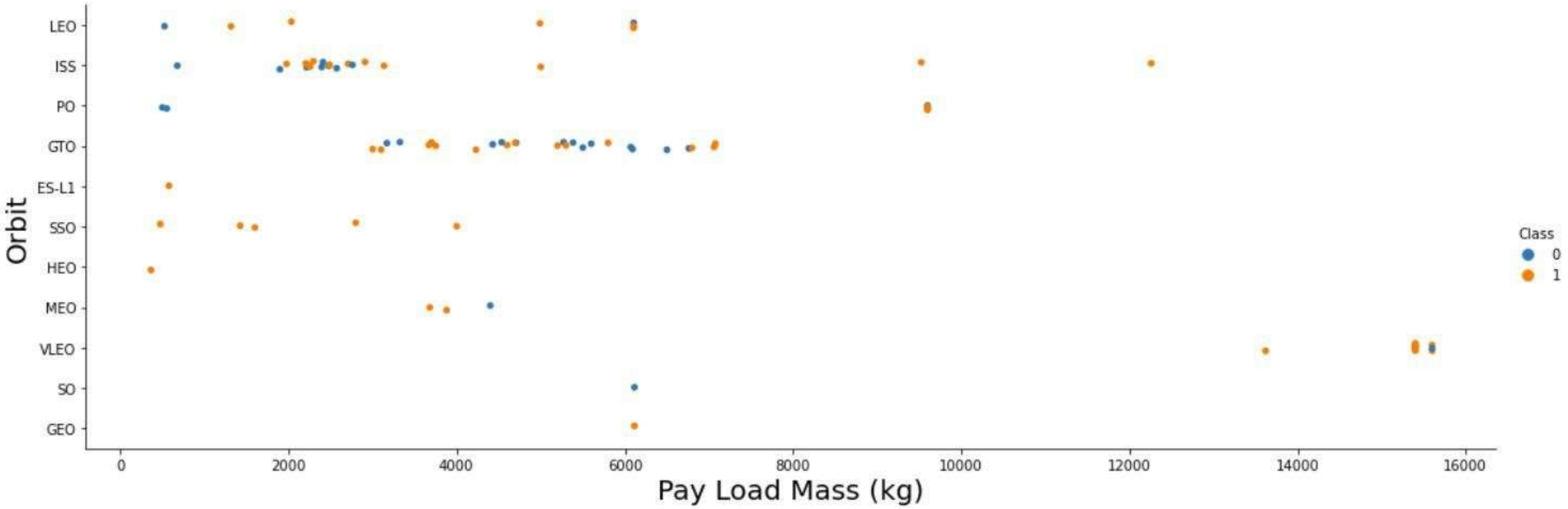
Flight Number vs. Orbit Type

- LEO , ISS, PO ,GTO had more flight counts than other orbit types
- SO,GEO,HEO orbit type had only 1 flight which was a failure, success and success respectively
- ES L1, SSO, GEO, HEO had 100% success landing in terms of flights



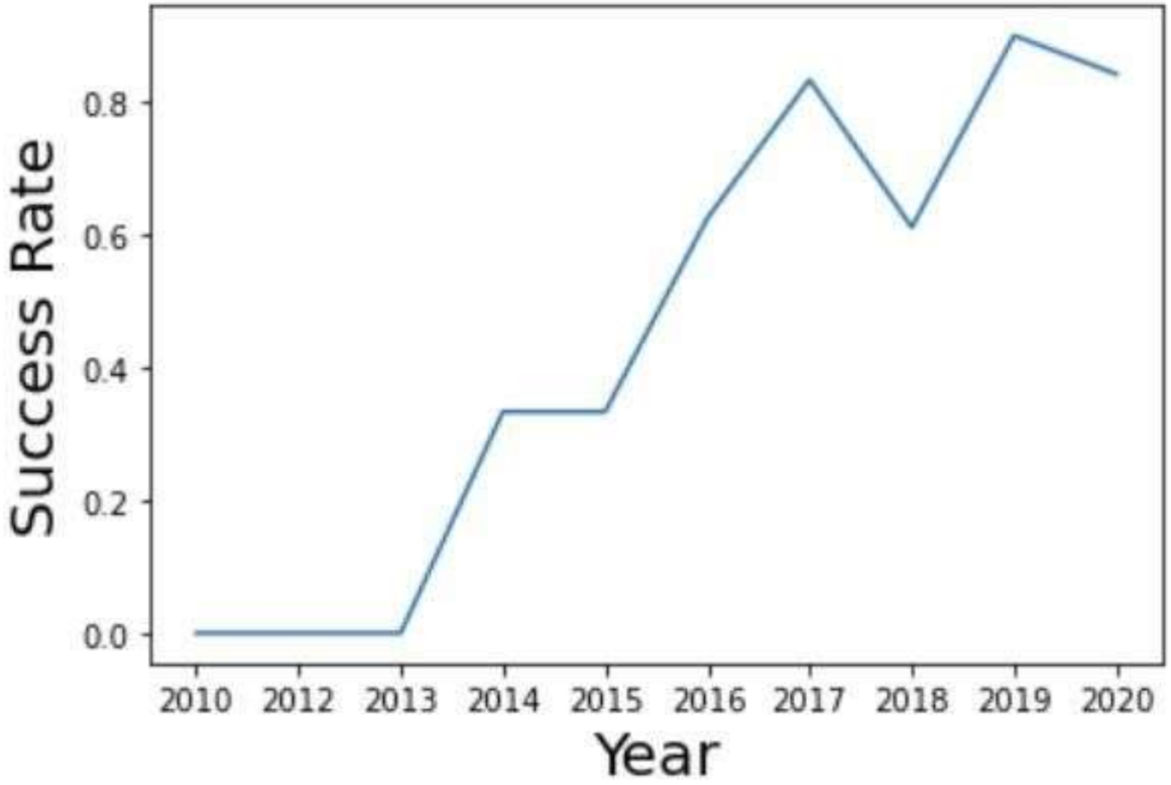
Payload vs. Orbit Type

- • Only ISS, PO and VLEO orbits carried more than 8000 kg payload
- ES L1, SSO, HEO,GEO had less than 6000 kg payload in their orbit with100% success



Launch Success Yearly Trend

- Success rate improves over time specifically since 2013



All Launch Site Names

- Find the names of the unique launch sites

	Launch Site	Lat	Long
0	CCAFS LC-40	28.562302	-80.577356
1	CCAFS SLC-40	28.563197	-80.576820
2	KSC LC-39A	28.573255	-80.646895
3	VAFB SLC-4E	34.632834	-120.610746

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`

DATE	time__utc_	booster_version	launch_site
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40
2014-04-18	19:25:00	F9 v1.1	CCAFS LC-40
2014-07-14	15:15:00	F9 v1.1	CCAFS LC-40
2014-09-21	5:52:00	F9 v1.1 B1010	CCAFS LC-40
2015-04-14	20:10:00	F9 v1.1 B1015	CCAFS LC-40

Total Payload Mass

- Calculate the total payload carried by boosters from NASA

total_payload
48213

Average Payload Mass by F9 v1.1

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

average_payload_carried_booster_f9_v1
2534

First Successful Ground Landing Date

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

date_first_successful_landing
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

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

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

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

Boosters Carried Maximum Payload

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

booster_version	payload_mass__kg_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

2015 Launch Records

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

landing__outcome	booster_version	launch_site	DATE
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40	2015-01-10
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40	2015-04-14

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

- 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

landing__outcome	RANK
Uncontrolled (ocean)	30
Uncontrolled (ocean)	30
Success (ground pad)	27
Success (ground pad)	27
Success (ground pad)	27
Success (drone ship)	22
Success (drone ship)	22
Success (drone ship)	22
Success (drone ship)	22
Success (drone ship)	22
Precluded (drone ship)	21
Failure (parachute)	9
Failure (parachute)	9
Failure (drone ship)	4
Failure (drone ship)	4
Failure (drone ship)	4
Failure (drone ship)	4
Failure (drone ship)	4
Controlled (ocean)	1
Controlled (ocean)	1
Controlled (ocean)	1

Section 3

Launch Sites Proximities Analysis

SpaceX Launch Sites

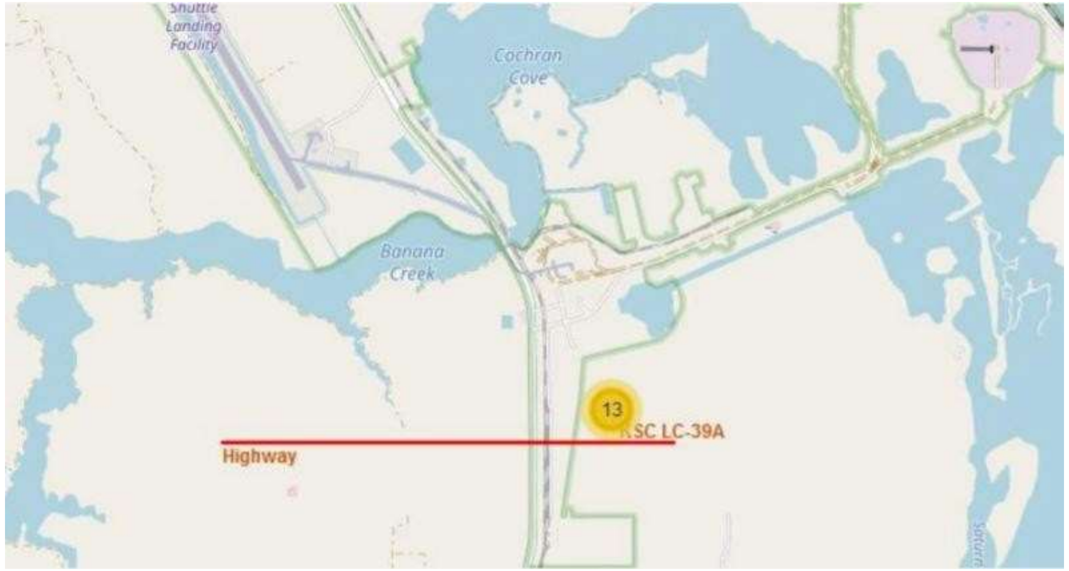
Explore the generated folium map and make a proper screenshot to include all launch sites' location markers on a global map

Succeeded/Failed Launches

- Marker clusters is used to simplify the map containing many markers having the same coordinate (check 3 locations near Florida).
- Successful launches are marked using a green marker and failed launches are marked using a red marker

Distance from launch site to nearest highway

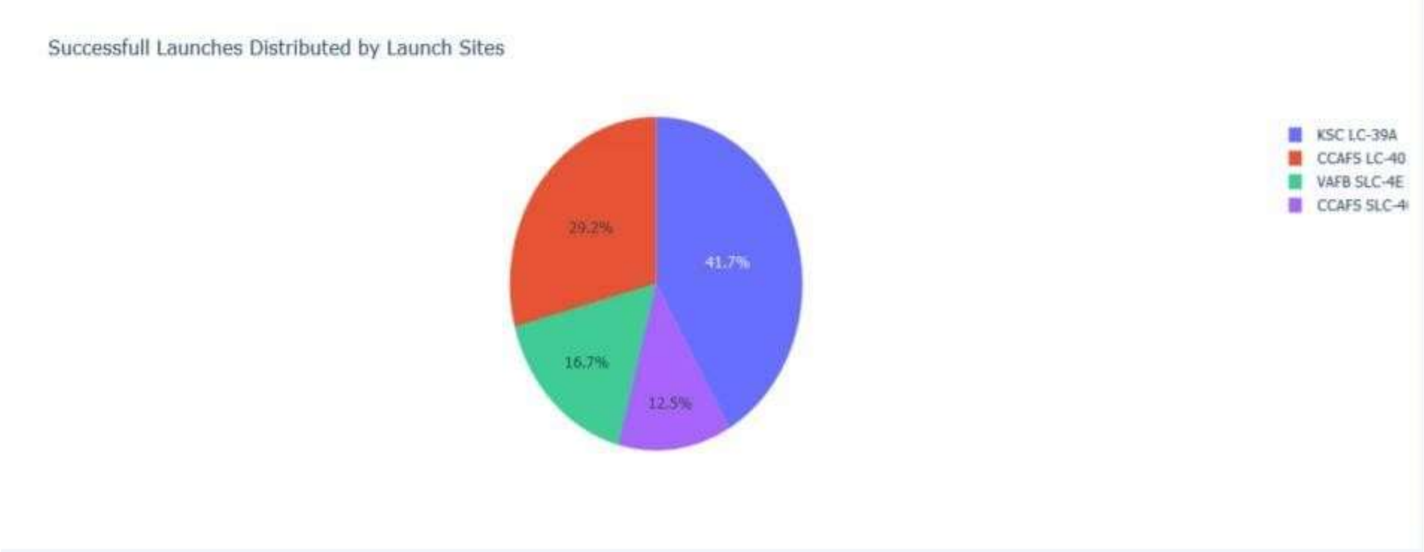
- the screenshot shows a red distance line denotes the distance between a launch site and its nearest highway



Section 4

Build a Dashboard with Plotly Dash

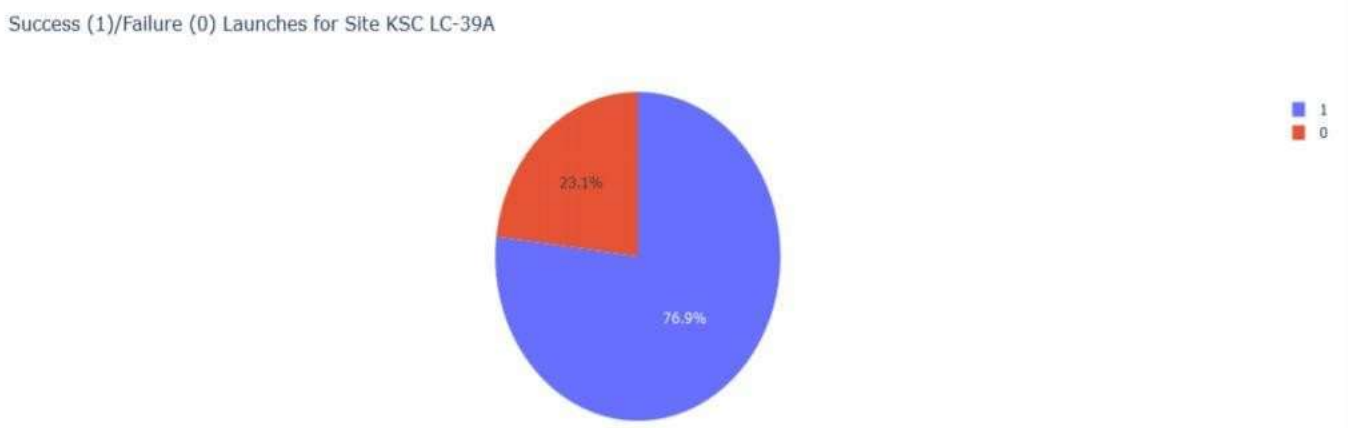
Successful launches by site



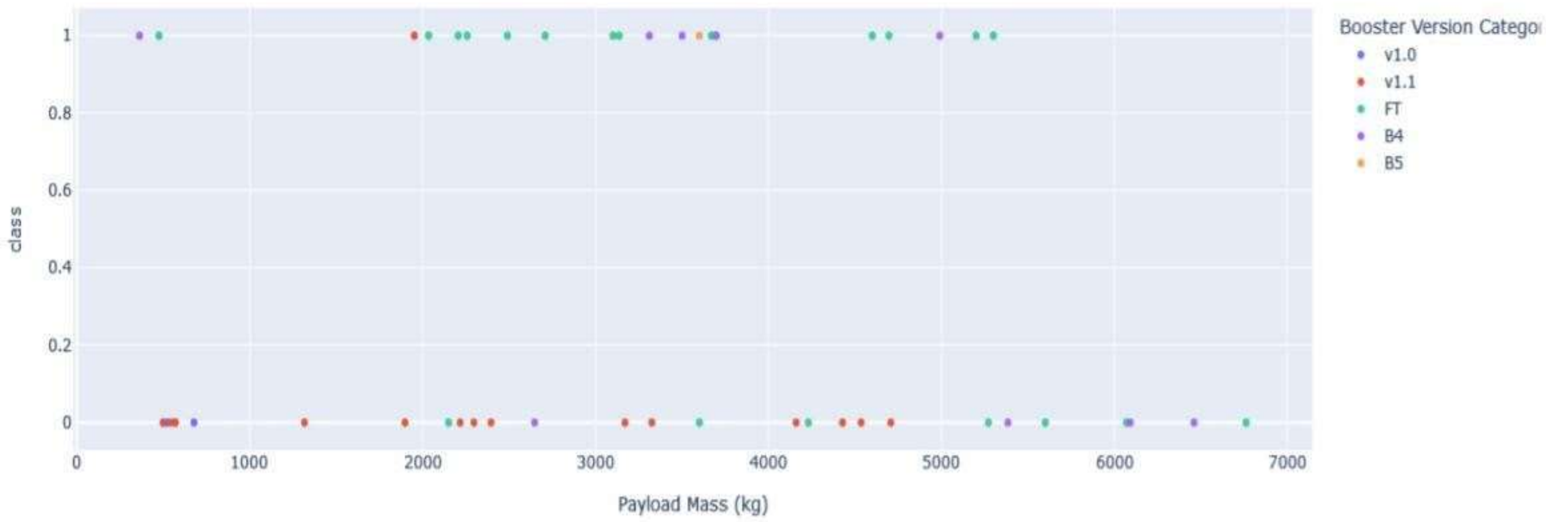
•"KSC LC 39A" has the highest success ratio whereas "VAFB SLC 4E" had the lowest success ratio among all four launch sites

Site With Highest Launch Success Ratio

- The KSC LC-39A has almost a 77% of success ratio and a 23 % failure ratio



Payload vs. Launch Outcome

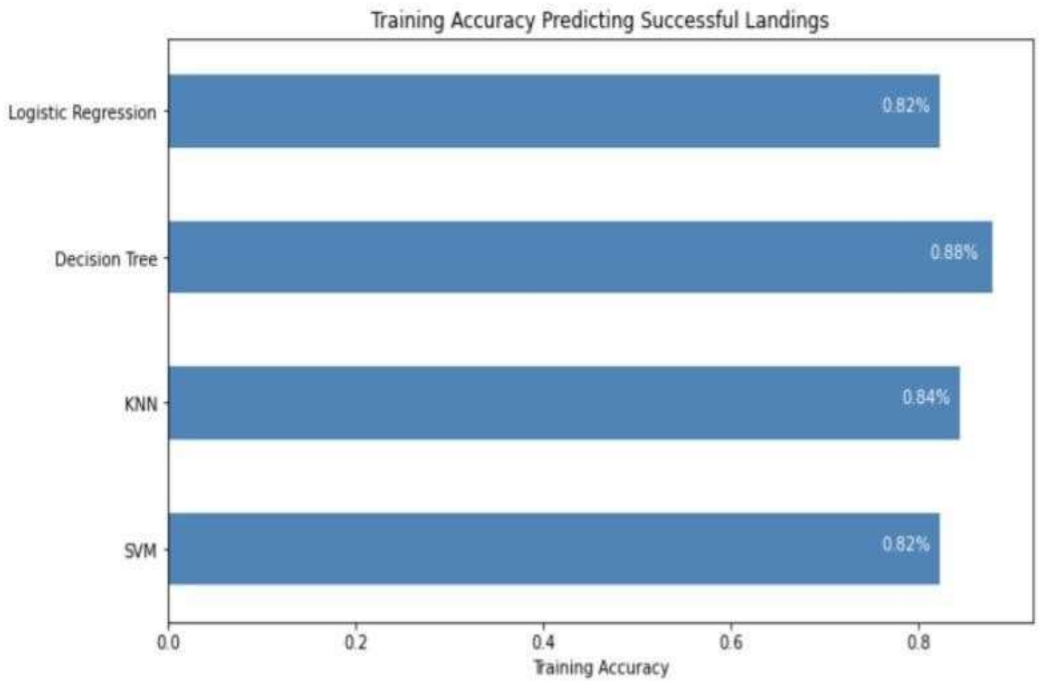


Section 5

Predictive Analysis (Classification)

Classification Accuracy

- Model accuracy for all built classification models
- Decision Tree has the highest classification accuracy



Thank you