Winning Space Race with Data Science

Saman Aboutorab
2022/08/27

Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

- Project background and context

In this capstone, we will predict if the Falcon 9 first stage will land successfully. SpaceX advertises Falcon 9 rocket launches on its website, with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage. Therefore if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Web scarping
  - Space X API

  Perform data wrangling
  - Filtering the dataframe to only include Falcon 9 lunches
  - Dealing with Missing Values

- Perform exploratory data analysis (EDA) using visualization and SQL

- Performing some Exploratory Data Analysis (EDA) to find some patterns in the data and determine what would be the label for training supervised models.
- predicting if the Falcon 9 first stage will land successfully
- Visualizing the relationship between Flight Number, Payload, success rate, orbit type and Launch Site
- Visualize the launch success yearly trend

Methodology

Executive Summary

- Perform interactive visual analytics using Folium and Plotly Dash

- The launch success rate may depend on many factors such as payload mass, orbit type, and so on. It may also depend on the location and proximities of a launch site, i.e., the initial position of rocket trajectories. Finding an optimal location for building a launch site certainly involves many factors and hopefully we could discover some of the factors by analyzing the existing launch site locations.

- Perform predictive analysis using classification models

- Perform exploratory Data Analysis and determine Training Labels
- create a column for the class
- Standardize the data
- Split into training data and test data
- -Find best Hyperparameter for SVM, Classification Trees and Logistic Regression
- Find the method performs best using test data

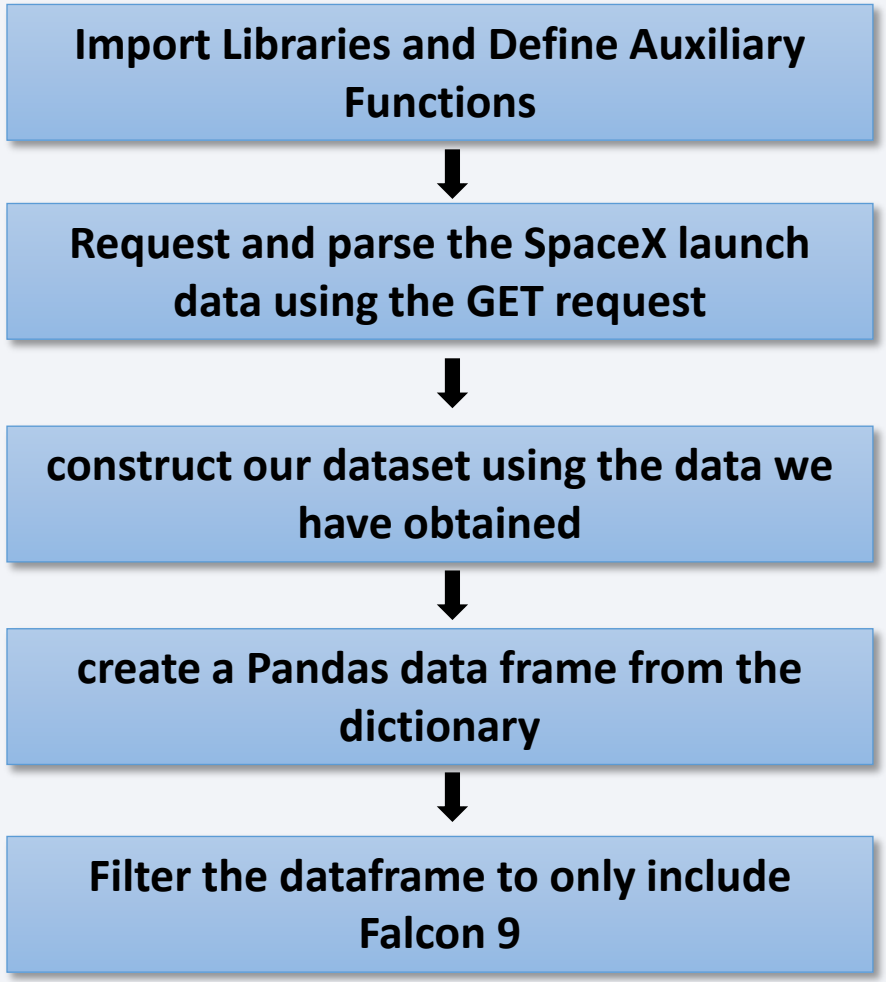
Data Collection

- Sources of Data:
  - Web scarping from: https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches
  - Space X API from: https://api.spacexdata.com/v4/rockets/

Data Collection – SpaceX API

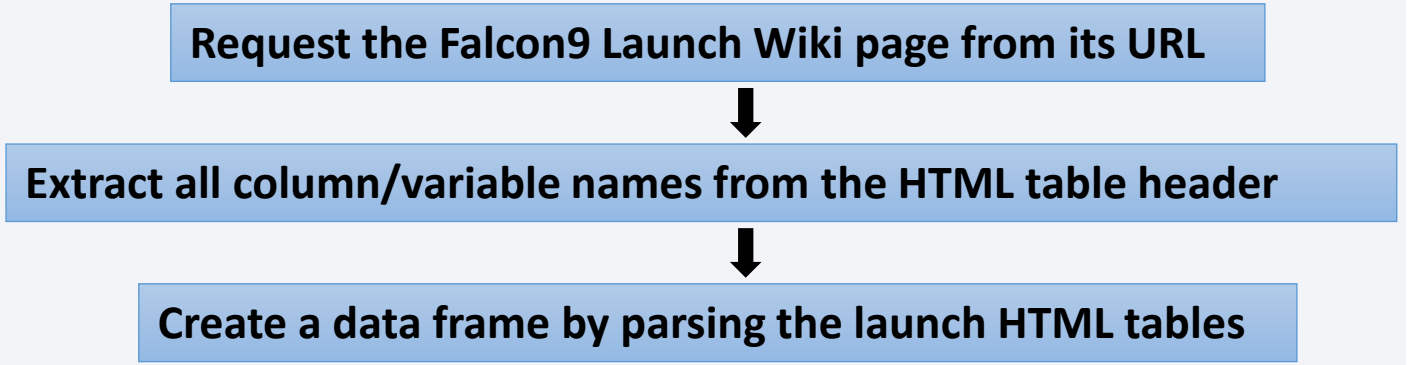
- Data collection with SpaceX REST calls using key phrases and flowcharts

- Source on Github: https://github.com/saman-aboutorab/IBM_capstone_project_space_Y/blob/master/01_Data%20collection%20from%20API.ipynb



Data Collection - Scraping

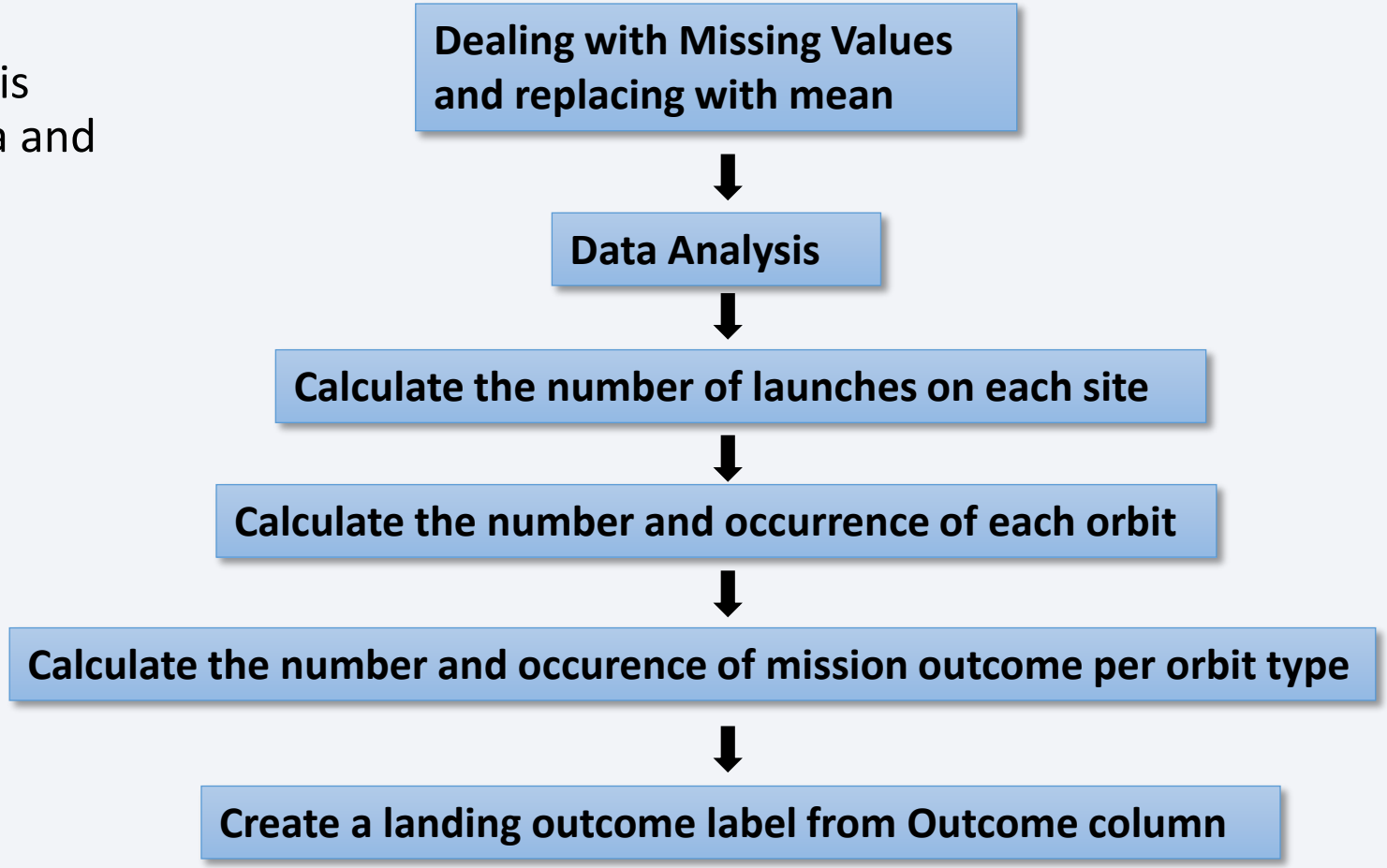
- The web scraping process using key phrases and flowcharts



- Source on Github:
- https://github.com/saman-aboutorab/IBM_capstone_project_space_Y/blob/master/01_webscraping.ipynb

Data Wrangling

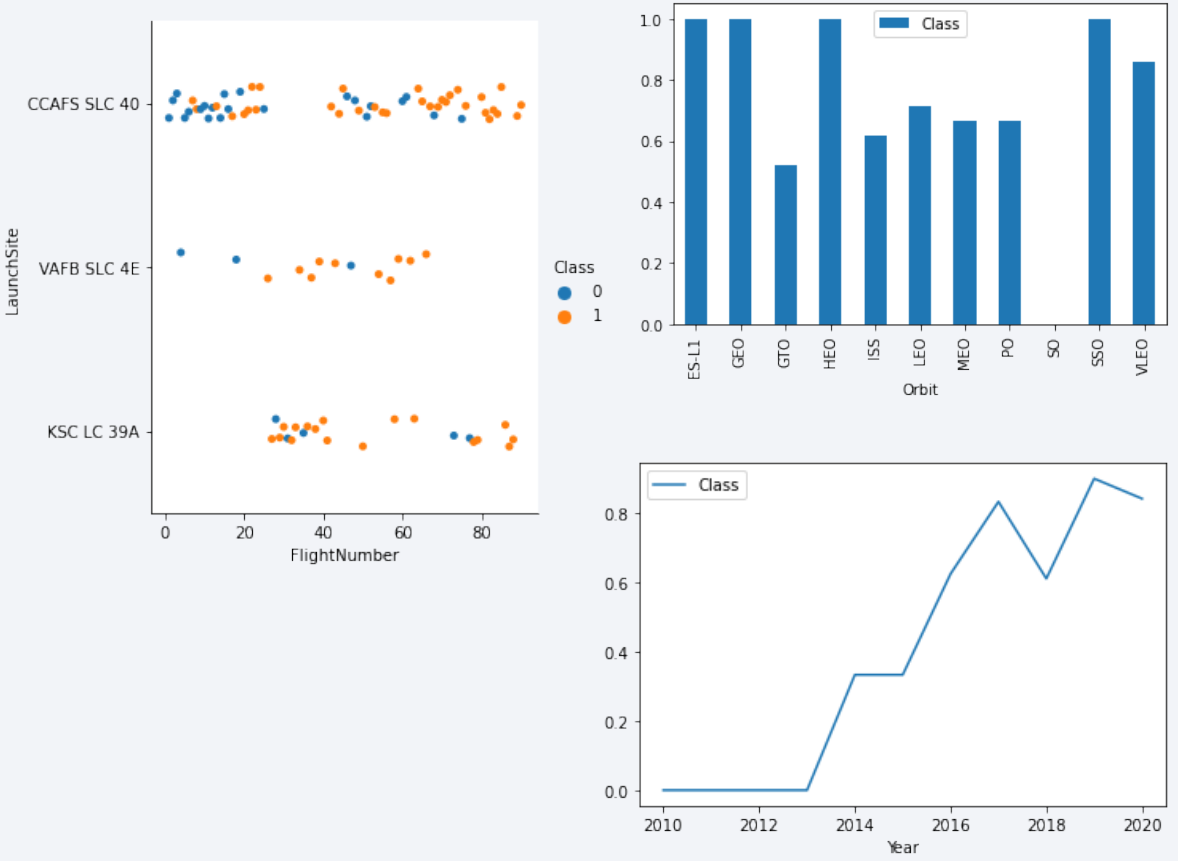
- perform some Exploratory Data Analysis (EDA) to find some patterns in the data and determine what would be the label for training supervised models.



- Source:
- https://github.com/saman-aboutorab/IBM_capstone_project_space_Y/blob/master/02_Data%20Wrangle.ipynb

EDA with Data Visualization

- Scatter plot, line chart and bar chart are used to Visualizing the relationship between Flight Number, Payload, success rate, orbit type and Launch Site



- Source: https://github.com/saman-aboutorab/IBM_capstone_project_space_Y/blob/master/04_EDA_dataviz.ipynb

EDA with SQL

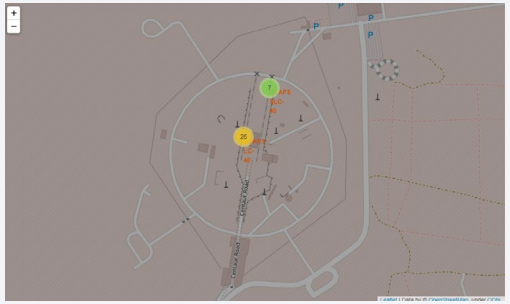
The SQL queries that was performed:

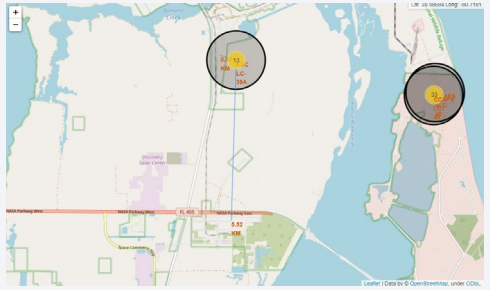
- **Display the names of the unique launch sites in the space mission**
- **Display 5 records where launch sites begin with the string 'CCA'**
- **Display the total payload mass carried by boosters launched by NASA (CRS)**
- **Display average payload mass carried by booster version F9 v1.1**
- **List the date when the first successful landing outcome in ground pad was 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**
- **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.**

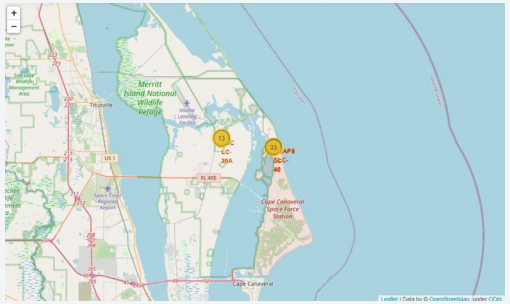
- **https://github.com/saman-aboutorab/IBM_capstone_project_space_Y/blob/master/03_EDA_SQL.ipynb**

Build an Interactive Map with Folium

Markers, circles, lines and marker clusters were used with Folium Maps







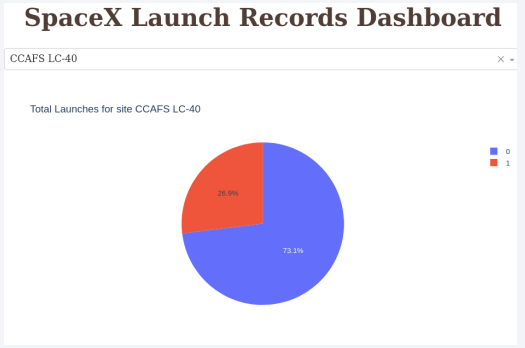
- Markers indicate points like launch sites;
- Circles indicate highlighted areas around specific coordinates,
- Marker clusters indicates groups of events in each coordinate
- Lines are used to indicate distances between two coordinates.

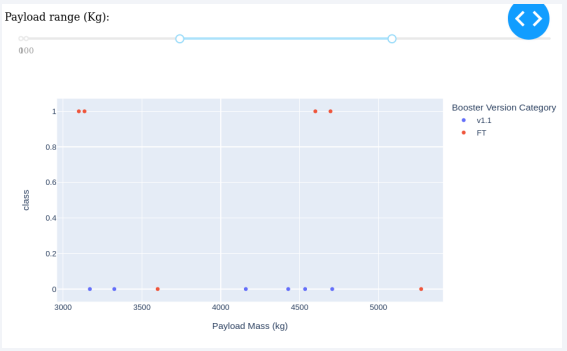
- https://github.com/saman-aboutorab/IBM_capstone_project_space_Y/blob/master/05_launch_site_location_Folium.ipynb

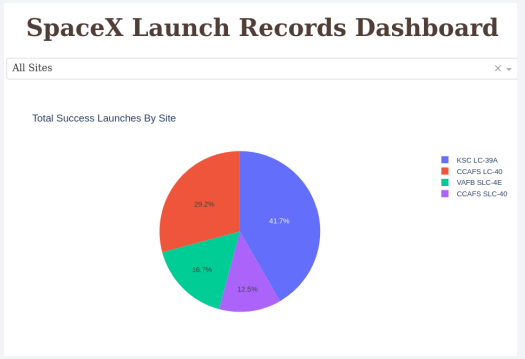
Build a Dashboard with Plotly Dash

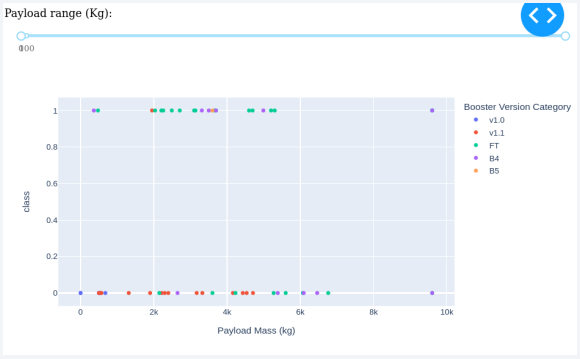
The following graphs and plots were used to visualize data

- Percentage of launches by site
- Payload range





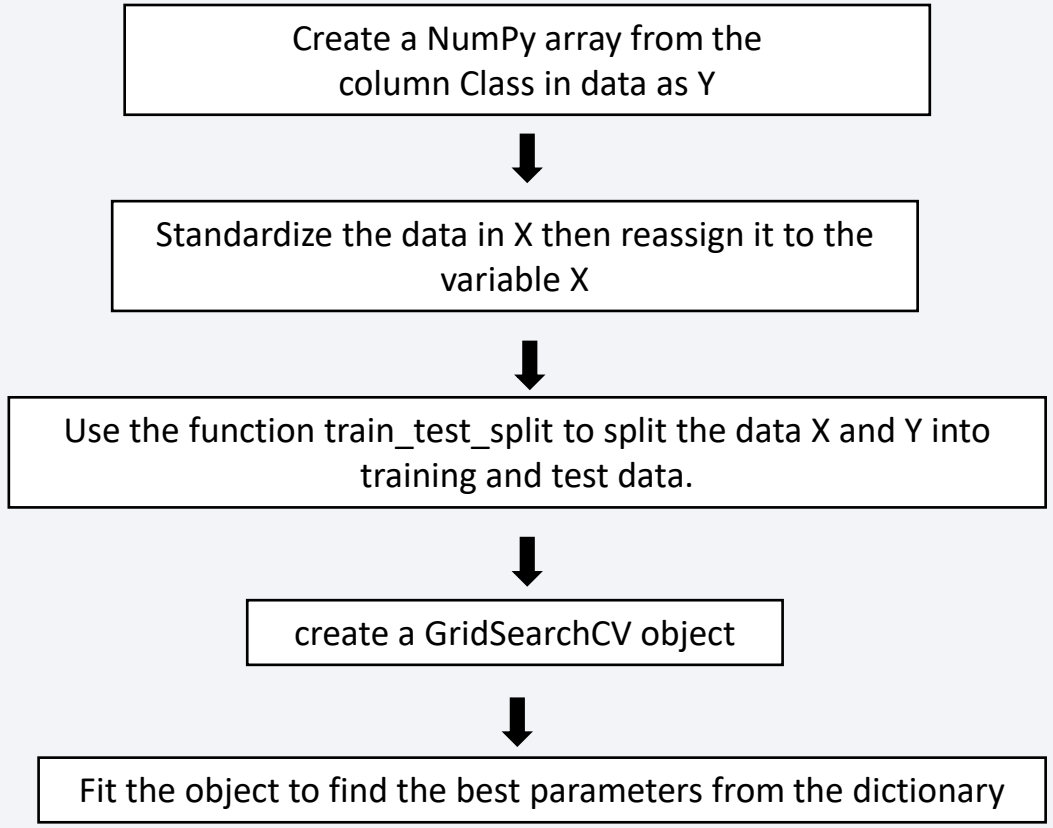




Predictive Analysis (Classification)

- In order to find the most accurate answer in each Machine learning model, a range of relevant parameters for each model are inserted into GridSearchCV.

- Source: https://github.com/saman-aboutorab/IBM_capstone_project_space_Y/blob/master/06_SpaceX_Machine%20Learning%20Prediction.ipynb



Results

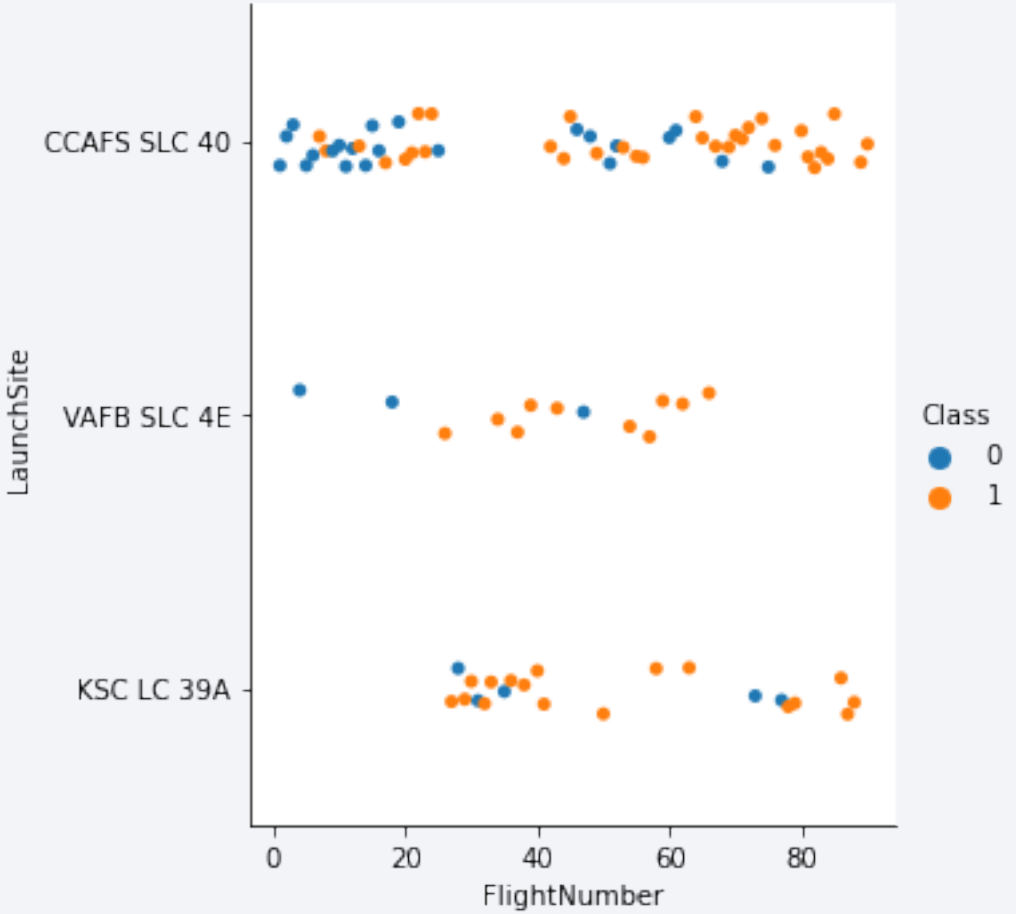
- Exploratory data analysis results:
- Space X uses 4 different launch sites;
- The first launches were done to Space X itself and NASA;
- The first success landing outcome happened in 2015 fiver year after the first launch;
- Many Falcon 9 booster versions were successful at landing in drone ships having payload above the average;
- The number of landing outcomes became as better as years passed.

Section 2

Insights drawn from EDA

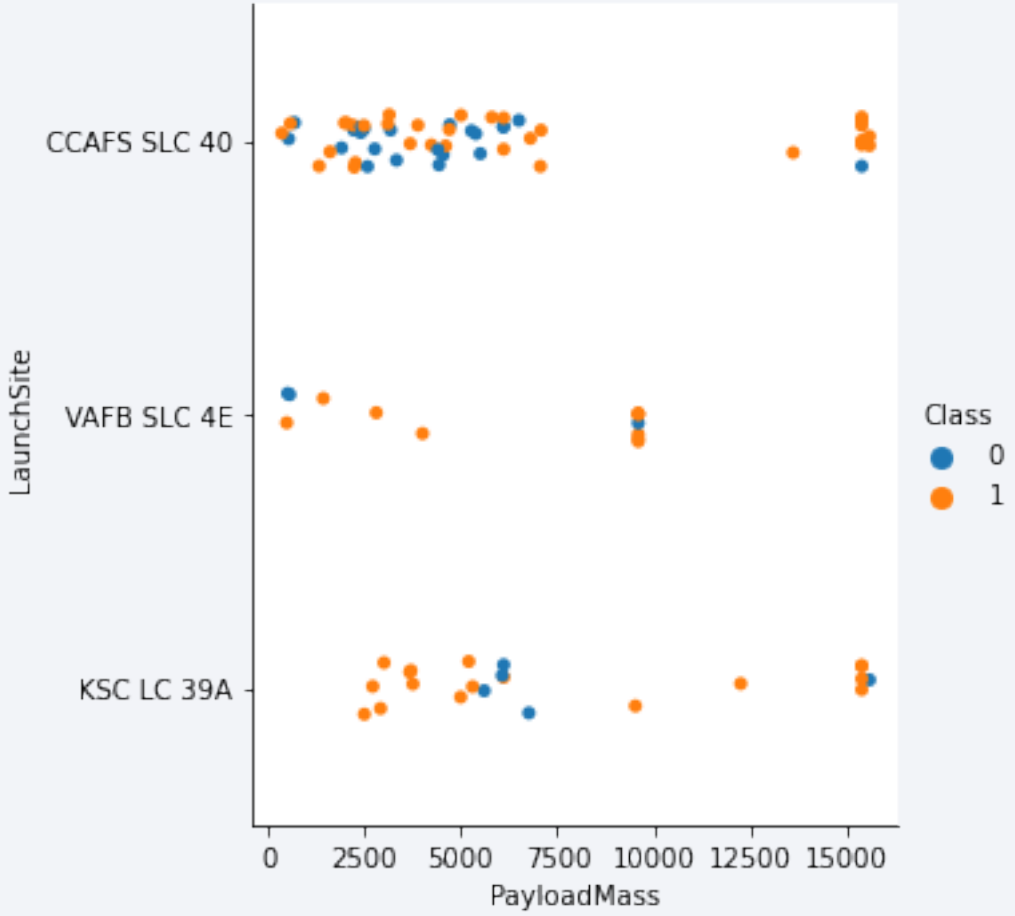
Flight Number vs. Launch Site

- Generally the higher Flight number the higher rate of success
- For flight numbers more than 70 in CCAFS lunch site, the success is guaranteed



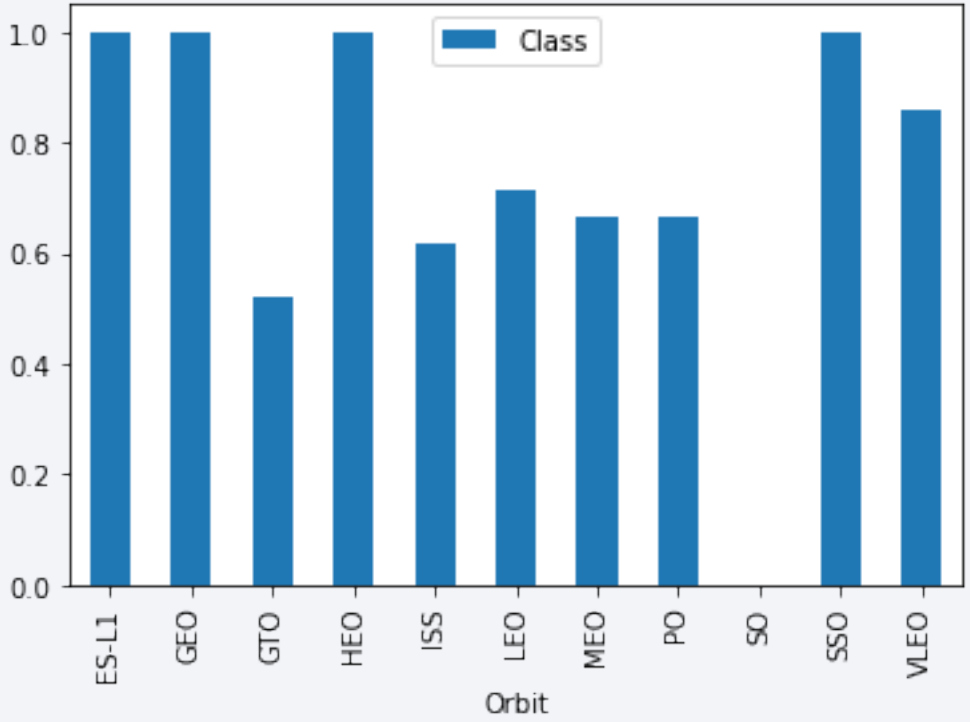
Payload vs. Launch Site

- In CCAFF lunch site, payloads less than 7500 does not show a relation between Payload mass and success. While higher than 15000 would work.
- In KSC LC lunch site payload between 5000 to 7500 would result failure while higher or lower than that is mainly succeeded



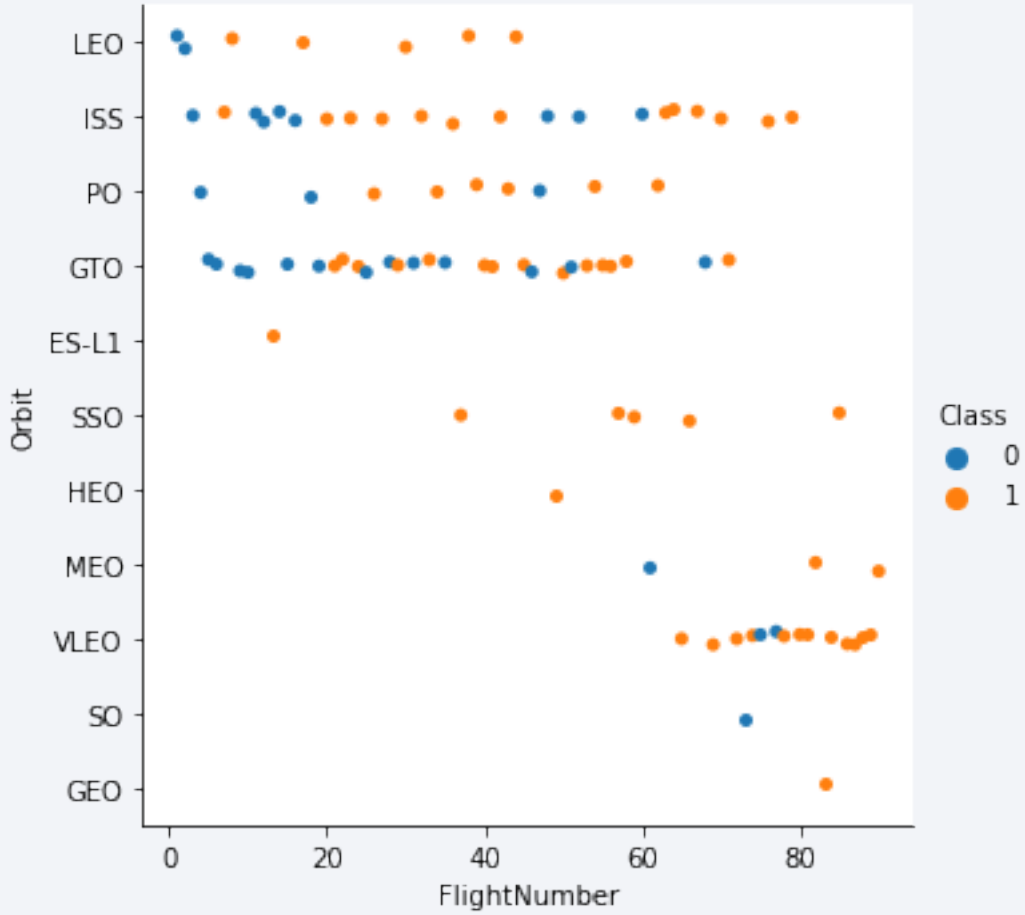
Success Rate vs. Orbit Type

- The bar chart shows the mean of success rate which is the highest for ES-L1, GEO, HEO and SSO, and the lowest for GTO



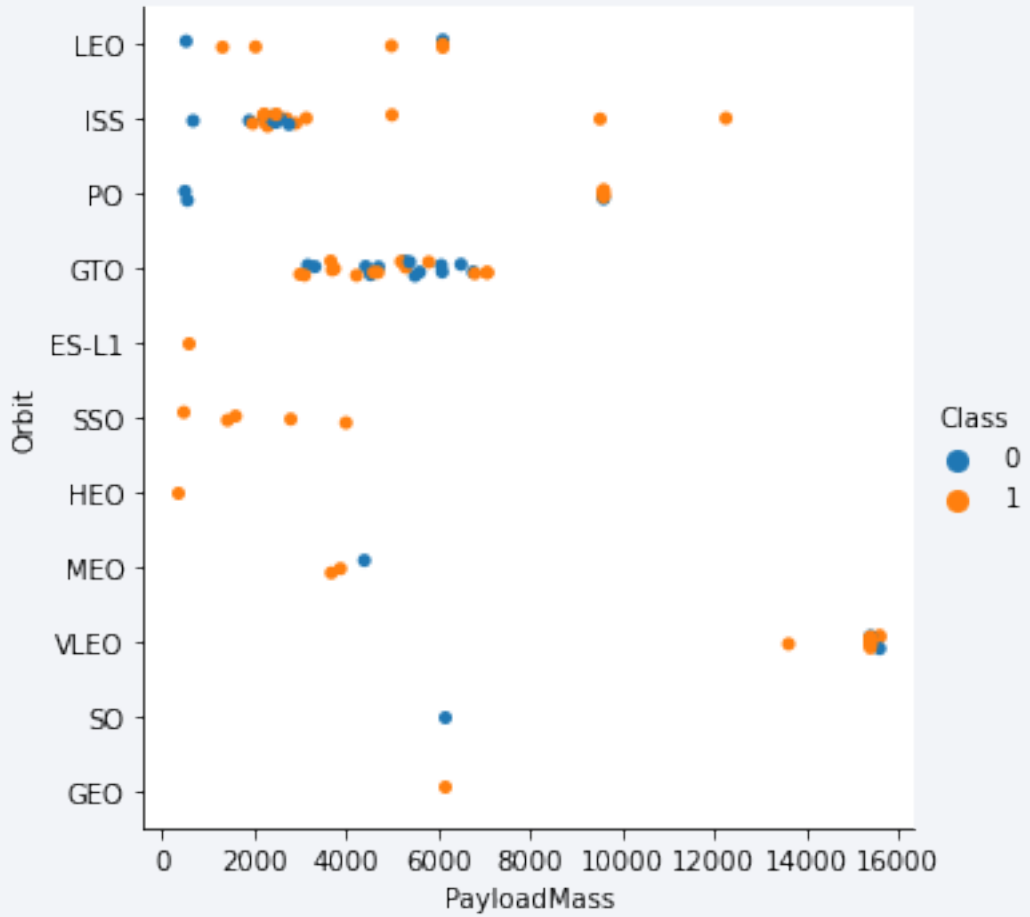
Flight Number vs. Orbit Type

- The higher the Flight number, the more success we have generally for Orbit types
- Orbits like LEO clearly shows higher flight numbers than 10 would result in success
- Orbits like GTO don't show a clear correlation between the flight numbers and success
- Orbit SSO shows success over all range of flight numbers



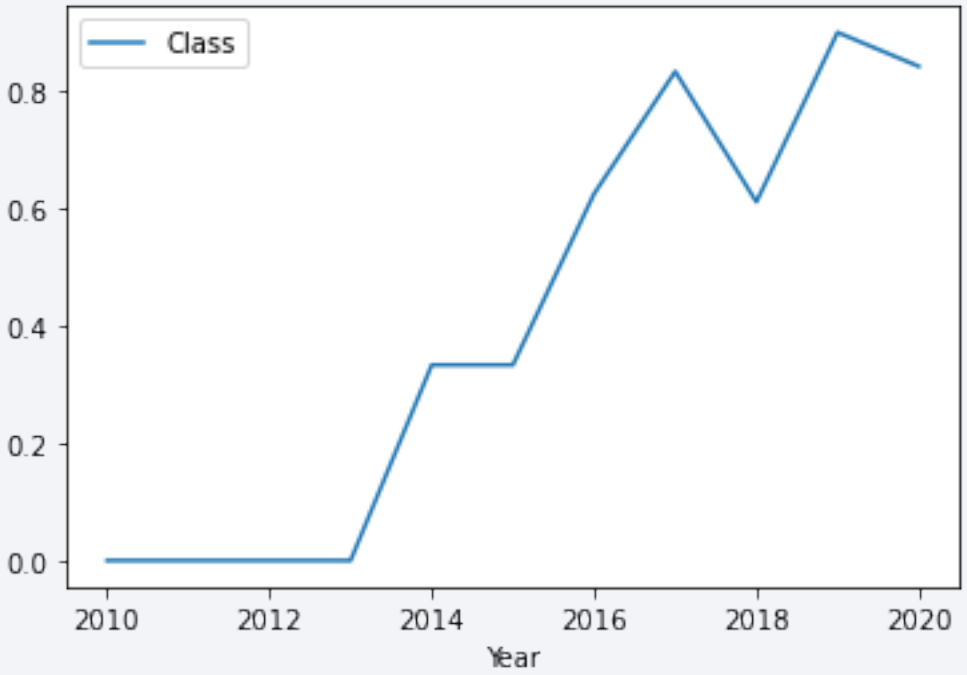
Payload vs. Orbit Type

- Generally, the higher payload mass, the higher success class
- In GTO, ISS Orbit, rate of success is not proportioned to the payload mass
- SSO orbit shows success for payload between 1000 to 6000



Launch Success Yearly Trend

- From 2013 the rate of success has been increased on a general base, except for 2018 with a drop



All Launch Site Names

- The unique name of lunch sites from SQL data base:

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

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites 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

- The total payload carried by boosters from NASA

TOTAL_PAYLOAD
111268

Average Payload Mass by F9 v1.1

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

AVG(PAYLOAD_MASS__KG_)
2928.4

Successful Drone Ship Landing with Payload between 4000 and 6000

- List of the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

Booster_Version
F9 v1.0 B0003
F9 v1.0 B0004
F9 v1.0 B0005
F9 v1.0 B0006
F9 v1.0 B0007
F9 v1.1 B1003
F9 v1.1
F9 v1.1
F9 v1.1
F9 v1.1
F9 v1.1
F9 v1.1 B1011
F9 v1.1 B1010
F9 v1.1 B1012
F9 v1.1 B1013
F9 v1.1 B1014

Total Number of Successful and Failure Mission Outcomes

The total number of successful and failure mission outcome

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

Boosters Carried Maximum Payload

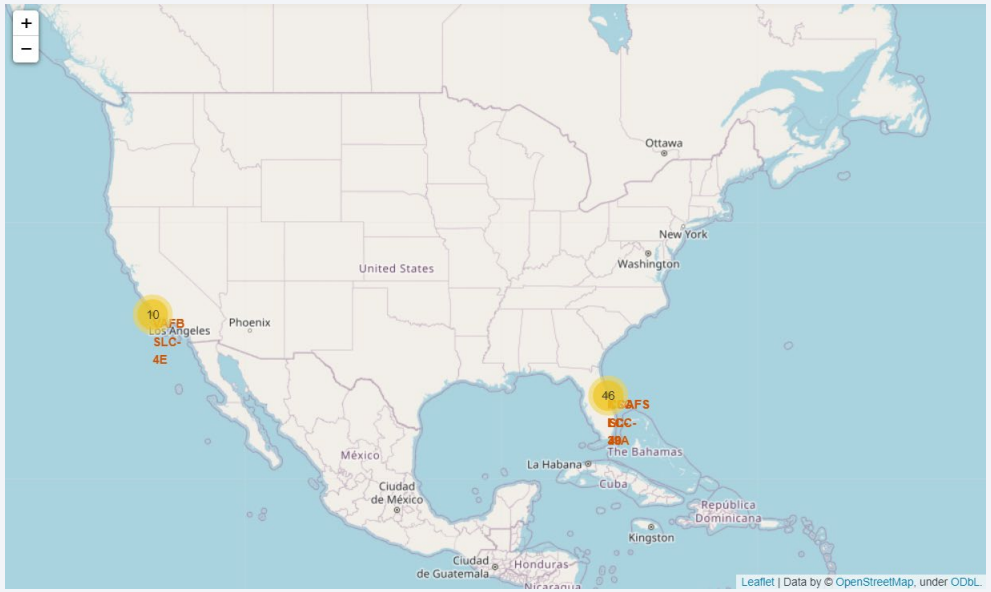
- List the names of the booster which have carried the maximum payload mass

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


Section 3
Launch Sites
Proximities Analysis

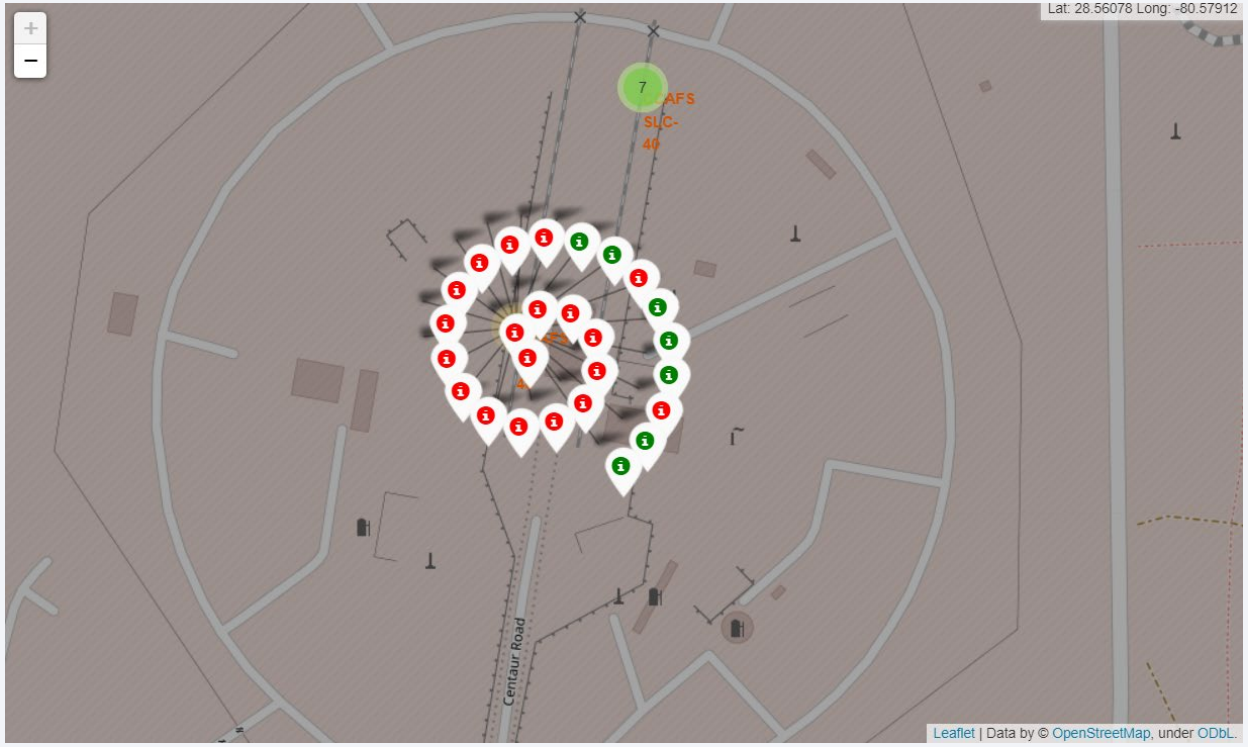
Launch sites' location markers on a global map

- Exploring the generated folium map and making a proper screenshot to include all launch sites' location markers on a global map



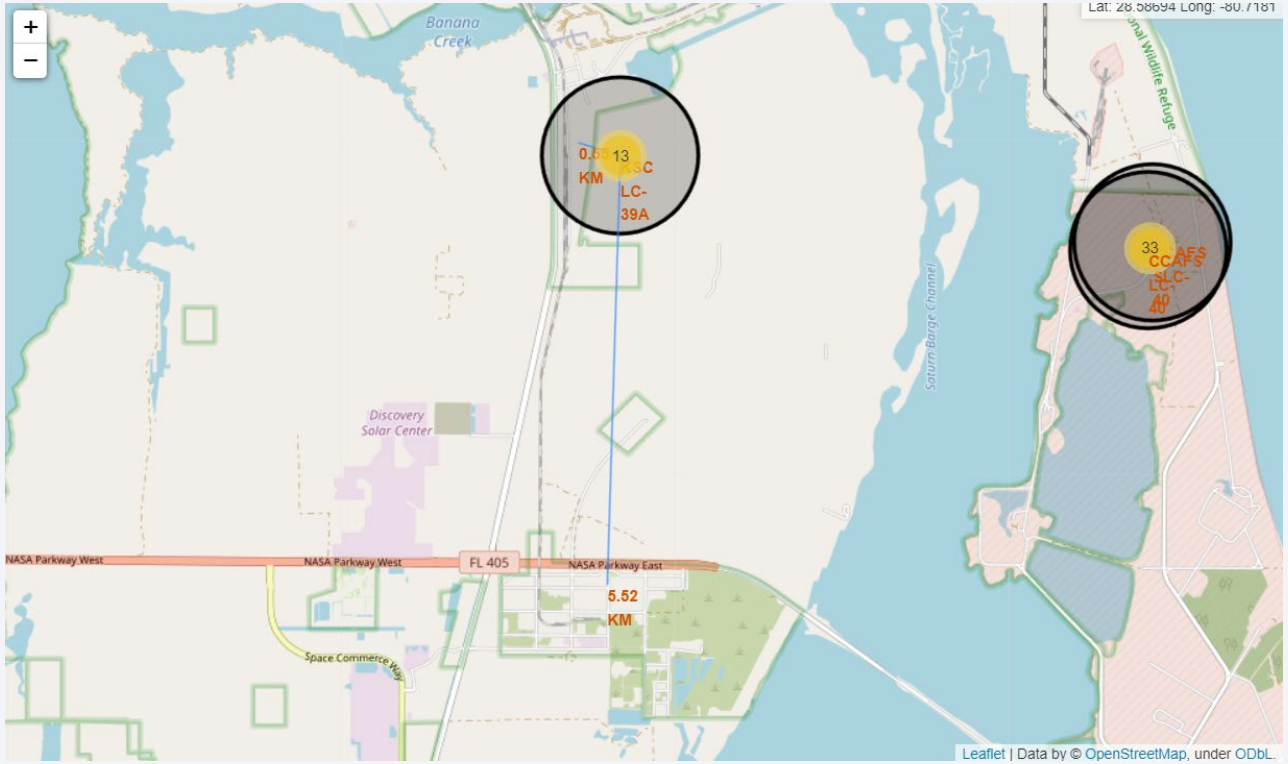
Color-labeled launch outcomes

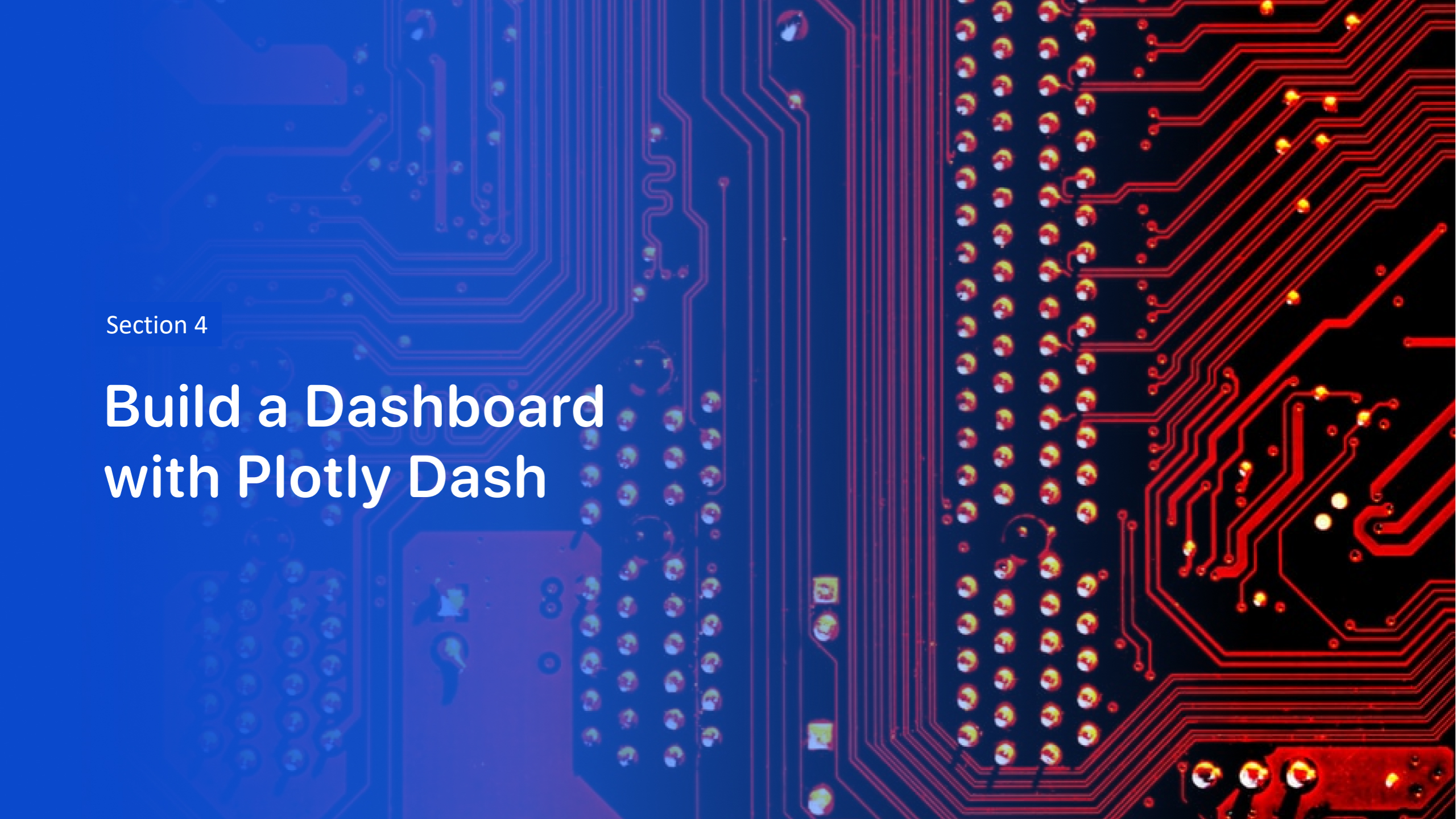
- Exploring the folium map and making a proper screenshot to show the color-labeled launch outcomes on the map



Launch site distance to its proximities

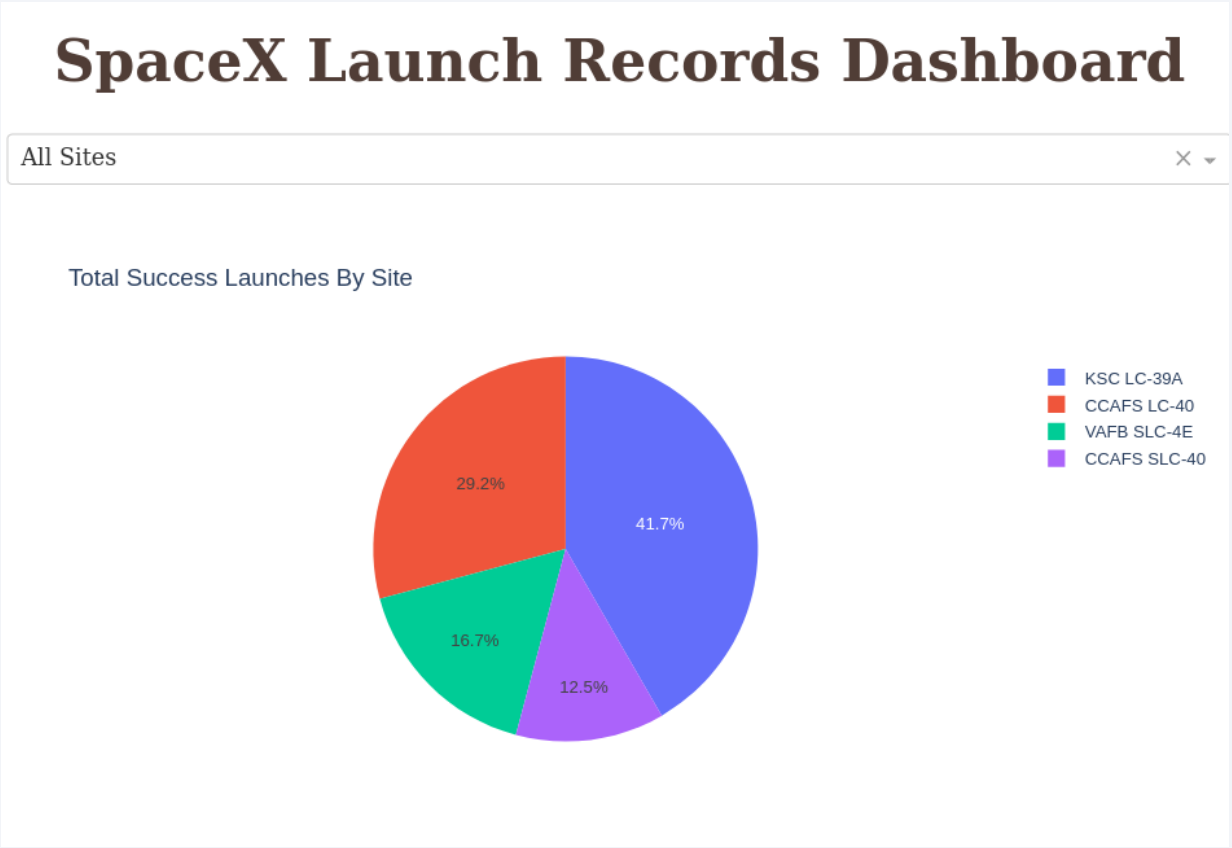
- Exploring the generated folium map and showing the screenshot of a selected launch site to its proximities such as railway, highway, coastline, with distance calculated and displayed




Section 4
Build a Dashboard with Plotly Dash

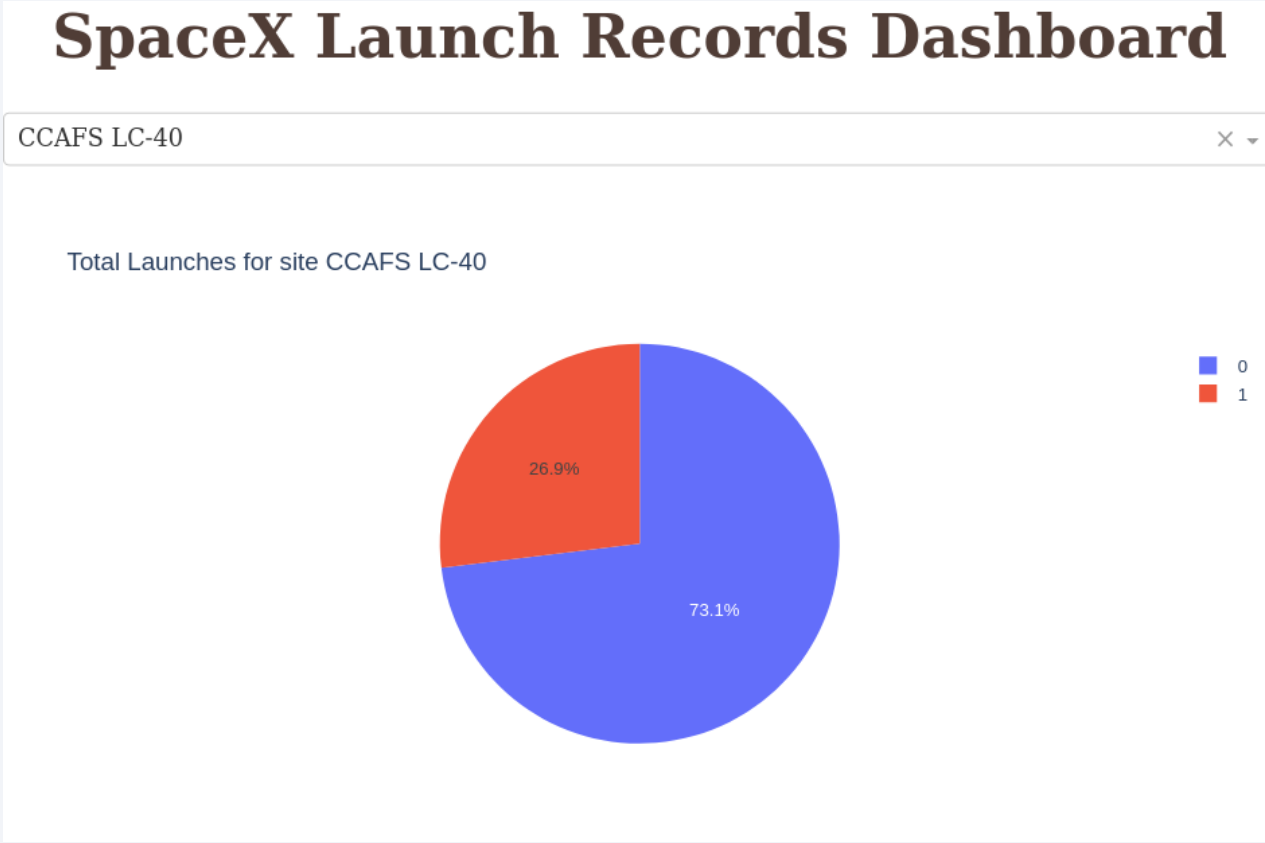
Launch success count for all sites

- Showing the screenshot of launch success count for all sites, in a piechart



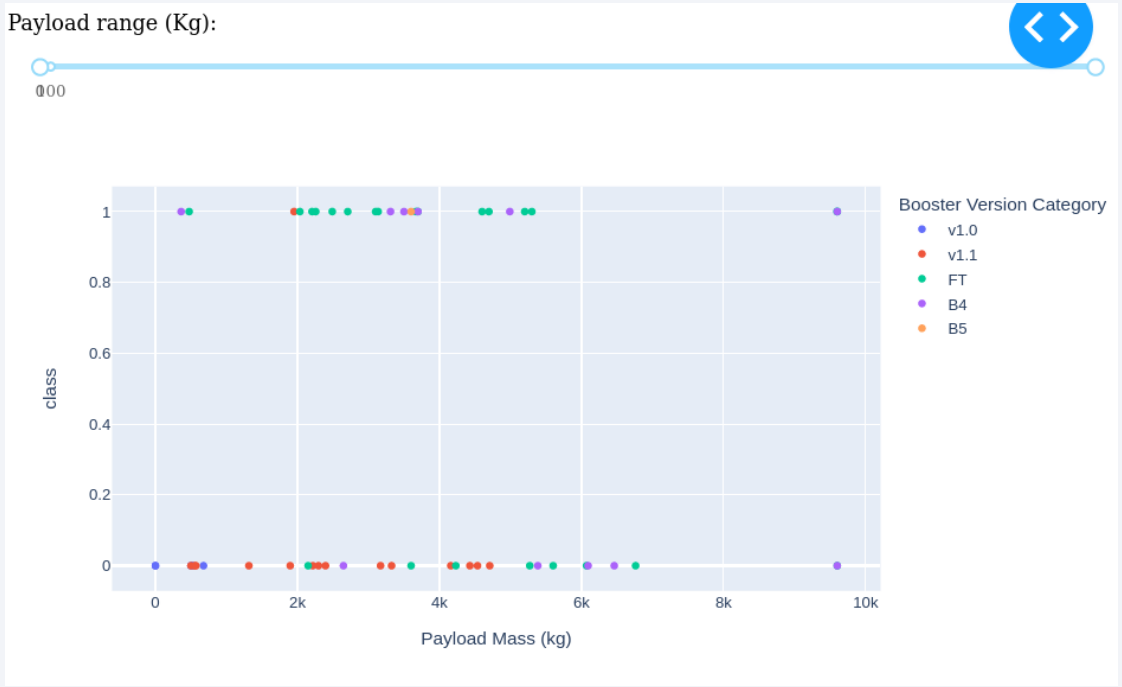
Launch site with highest launch success ratio

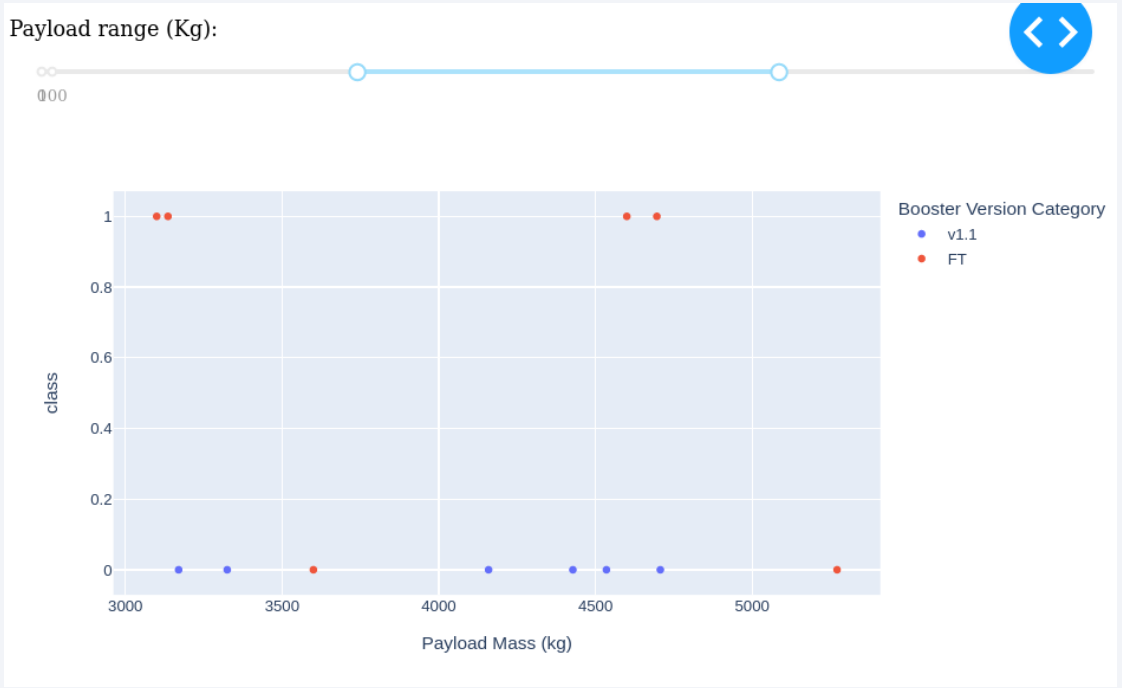
- Showing the screenshot of the piechart for the launch site with highest launch success ratio



Payload vs. Launch Outcome

- Showing screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider



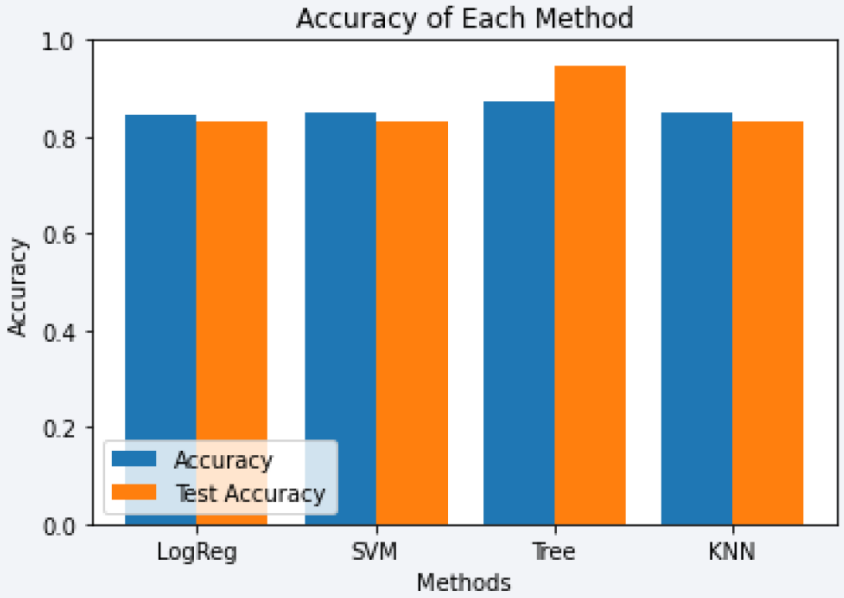


Section 5

Predictive Analysis (Classification)

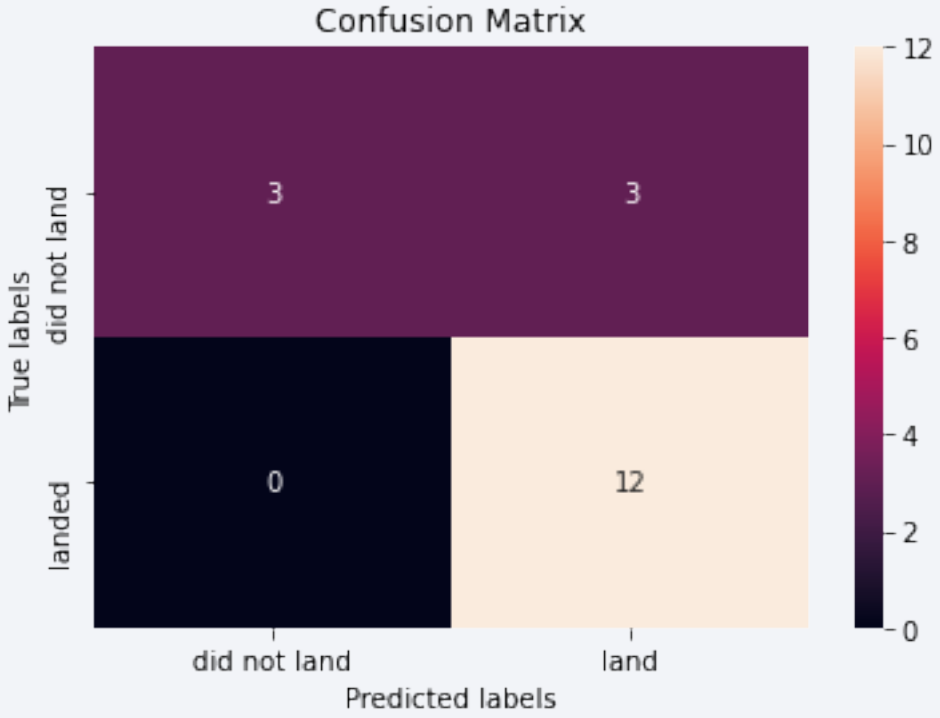
Classification Accuracy

- The Tree model has the highest Accuracy on training model and highest accuracy on Test model with more than 87%



Confusion Matrix

Confusion matrix of Decision Tree Classifier proves its accuracy by showing the big numbers of true positive and true negative compared to the false ones.



Conclusions

- The Rate of success has been increased as time has passed since 2013
- The best launch site is KSC LC-39A;
- Launches above 7,000kg are less risky;
- Although most of mission outcomes are successful, successful landing outcomes seem to improve over time, according the evolution of processes and rockets;
- Decision Tree Classifier can be used to predict successful landings and increase profits.
- The more the Flighter number each lunch site had, the higher the chance of success

Appendix

- Please refer to the project's Github repository for python codes: https://github.com/saman-aboutorab/IBM_capstone_project_space_Y/tree/master

Thank you!