IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Satria F. Bagaskara
11 April 2023

Outline

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

Executive Summary

- Summary of methodologies
  A. Collecting Data using Webscraping from API
  B. Data Wrangling
  C. Exploratory Data Analysis (EDA) including data wrangling & data visualization
  D. Interactive Visual Analytics using Dashboard with Folium map
  E. Predictive Analysis (Classification)

- Summary of all results
  A. EDA can be used to features selection to predict success of launching and landing
  B. There's possibility to drive the opportunity for Space Y with better success prediction of launching and landing

Introduction

- The Goals of Project

Evaluating feasibility of the new Space Y to compete with Space X

- Solutions:

a. The best way to minimize the cost per launch is by predicting the success rate of rocket landings, which can lead to the re-use of rockets for future launches

b. Higher success rate of rocket launches by selecting best place for launching

Section 1

Methodology

Methodology

Executive Summary

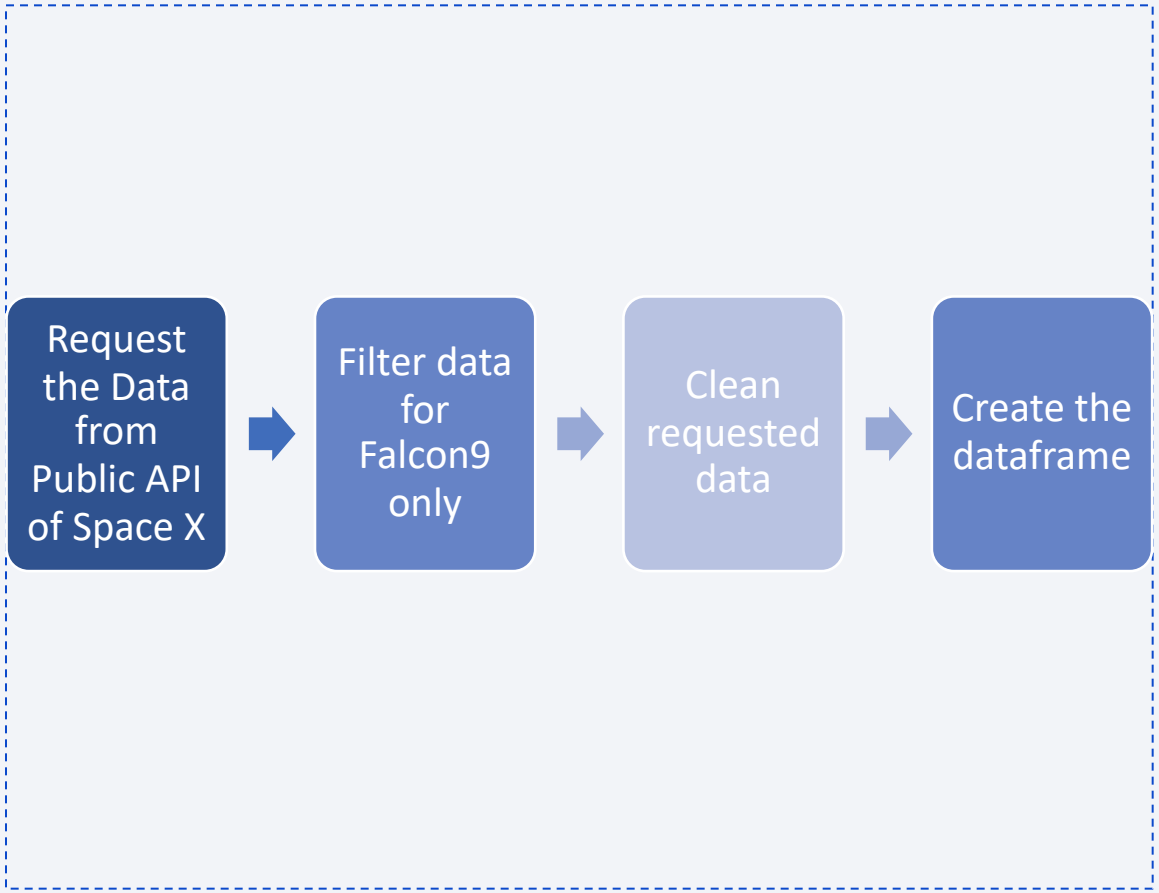
- Data collection methodology:
  - Data from Space X was collected from:
    - Space X API (https://api.spacexdata.com/v4/rockets/) & WebScraping (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches)
- Perform data wrangling
  - Checking missing values, cleaning and creating target feature based on outcome data
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Normalizing, split test data, modelling with 4 classification models using GridSearchCV and evaluating the model using accuracy on train/test dataset and confusion matrix evaluation

Data Collection

- Data from Space X was collected using WebScraping Technique from:
  - Space X API (https://api.spacexdata.com/v4/rockets/)
  - WebScraping (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches)

Data Collection – SpaceX API

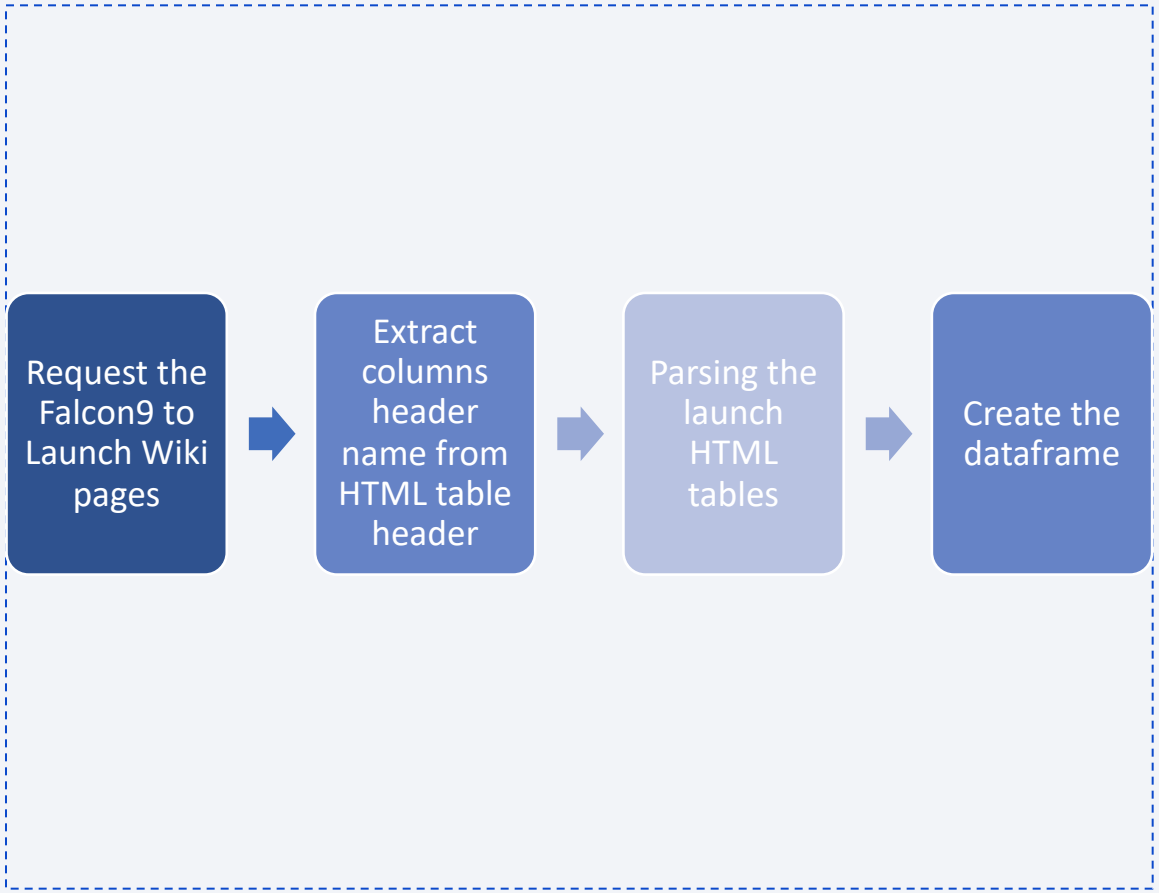
- Present your data collection with SpaceX REST calls using key phrases and flowcharts
- Requesting the data from Public API of Space X which step-by-step as following from the flowchart
- Add the GitHub URL of the completed SpaceX API calls notebook (must include completed code cell and outcome cell), as an external reference and peer-review purpose



Data Collection - Scraping

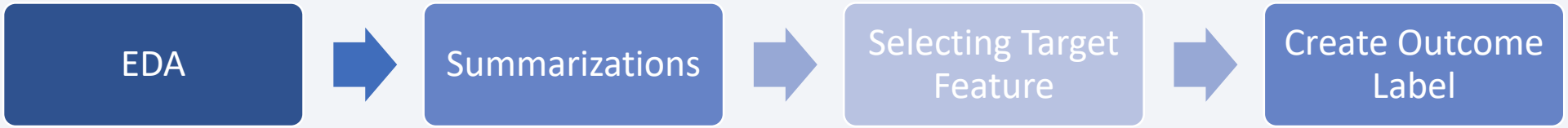
- Scraping the data from Wikipedia which step-by-step as following from the flowchart

- Add the GitHub URL of the completed web scraping notebook, as an external reference and peer-review purpose



Data Wrangling

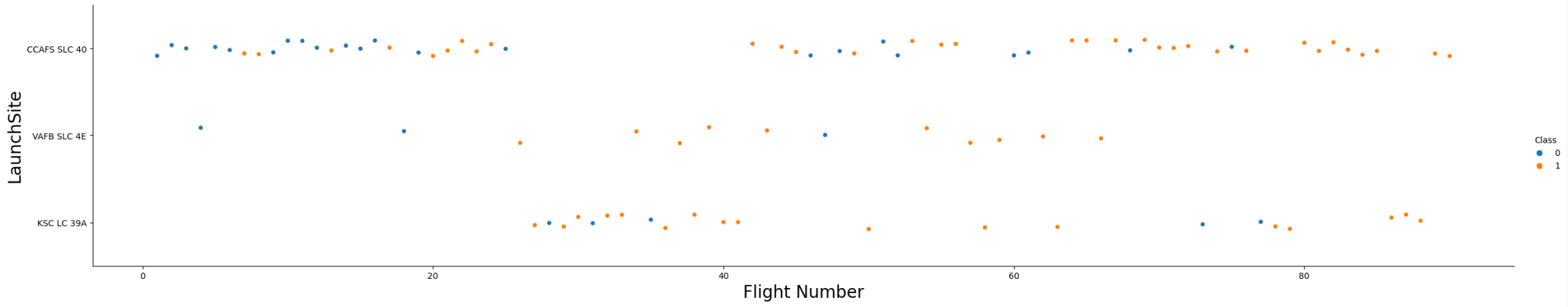
- First of all, perform Exploratory Data Analysis (EDA)
- Summary the launch site, occurrences of mission, and other features that have possibility to affect success rate of landing
- Selecting outcome column that can be used for classification modelling, and create outcome label (success rate of landing)



- Add the GitHub URL of your completed data wrangling related notebooks, as an external reference and peer-review purpose

EDA with Data Visualization

- In exploratory data analysis, the data visualized using scatterplot and barplot to show relationship between features like Payload Mass and Flight Number, Launch Site and Flight Number, Launch Site and Payload Mass, Orbit and Flight Number, Payload and Orbit
- Add the GitHub URL of your completed EDA with data visualization notebook, as an external reference and peer-review purpose



EDA with SQL

SQL queries were performed :

1. Names of the distinct launch sites used in the space mission
2. The top 5 launch sites that start with 'CCA' in terms of usage
3. Total payload mass transported by NASA's (CRS) boosters
4. Mean payload mass carried by the F9 v1.1 booster variant
5. Date of the first successful ground pad landing
6. Boosters that achieved successful drone ship landings and transported payloads weighing between 4000 and 6000 kg
7. Total number of successful and unsuccessful mission outcomes
8. Booster models that transported the heaviest payloads
9. Failed landing outcomes on drone ships, along with their booster versions and launch site names, for the year 2015
10. Rank of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between June 4, 2010 and March 20, 2017.

- Add the GitHub URL of your completed EDA with SQL notebook, as an external reference and peer-review purpose

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, like NASA Johnson Space Center;
  - Marker clusters indicates groups of events in each coordinate, like launches in a launch site; and
  - Lines are used to indicate distances between two coordinates.
- Add the GitHub URL of your completed interactive map with Folium map, as an external reference and peer-review purpose

Build a Dashboard with Plotly Dash

- The following graphs and plots were used to visualize data
  - Percentage of launches by site
  - Payloadrange
- This combination allowed to quickly analyze the relation between payloads and launch sites, helping to identify where is best place to launch according to payloads.
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

Predictive Analysis (Classification)

- Four classification models were compared: logistic regression, support vector machine, decision tree and k nearest neighbors.

Data Preparation and Normalization → Test of each model with GridSearchCV → Gain best result for each model → Comparing for each model

- Add the GitHub URL of your completed predictive analysis lab, as an external reference and peer-review purpose

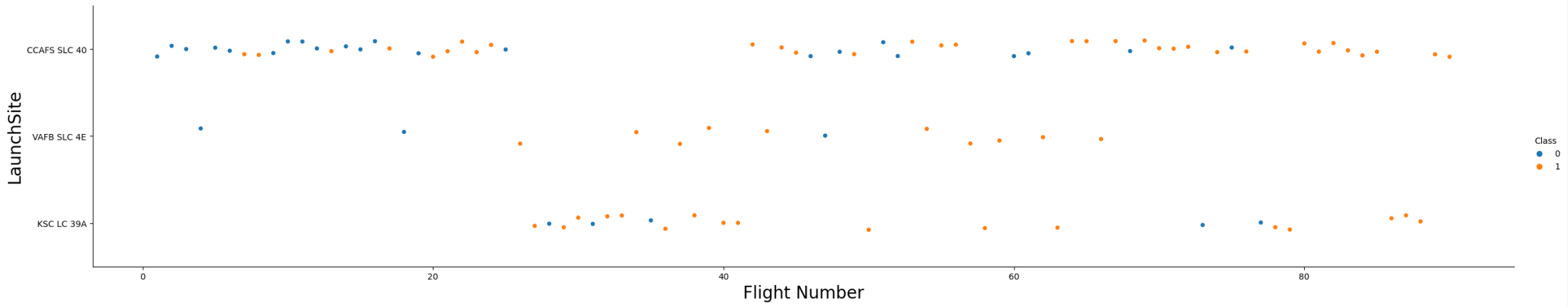
Results

- Exploratory data analysis results
  - Space X have 4 different launch sites
  - The first launches by Space X itself and NASA
  - CCAFS SLC 40 was pioneering with many fail occurrences, but it seems contribute data for others
  - The average payload of F9 v1.1 booster is 2.928 kg
  - After the first launch in 2010, the first success landing is in 2015
  - Two booster versions (F9 v1.1 B1012 and F9 v1.1 B101) failed at landing in drone ships in 2015
  - Many Falcon 9 booster versions were successful at landing in drone ships
  - Almost 100% of mission outcomes were successful
  - The number of landing outcomes became as better as years passed, especially when its 20th flight above

Section 2

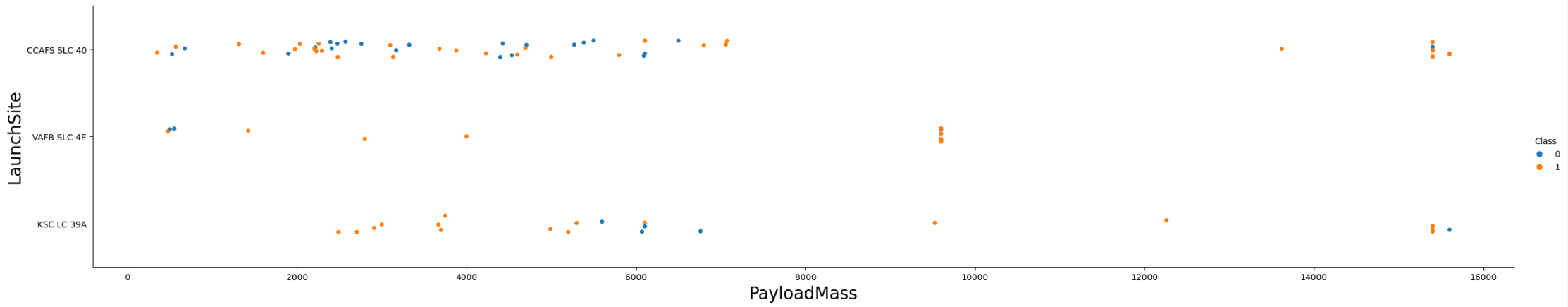
Insights drawn from EDA

Flight Number vs. Launch Site



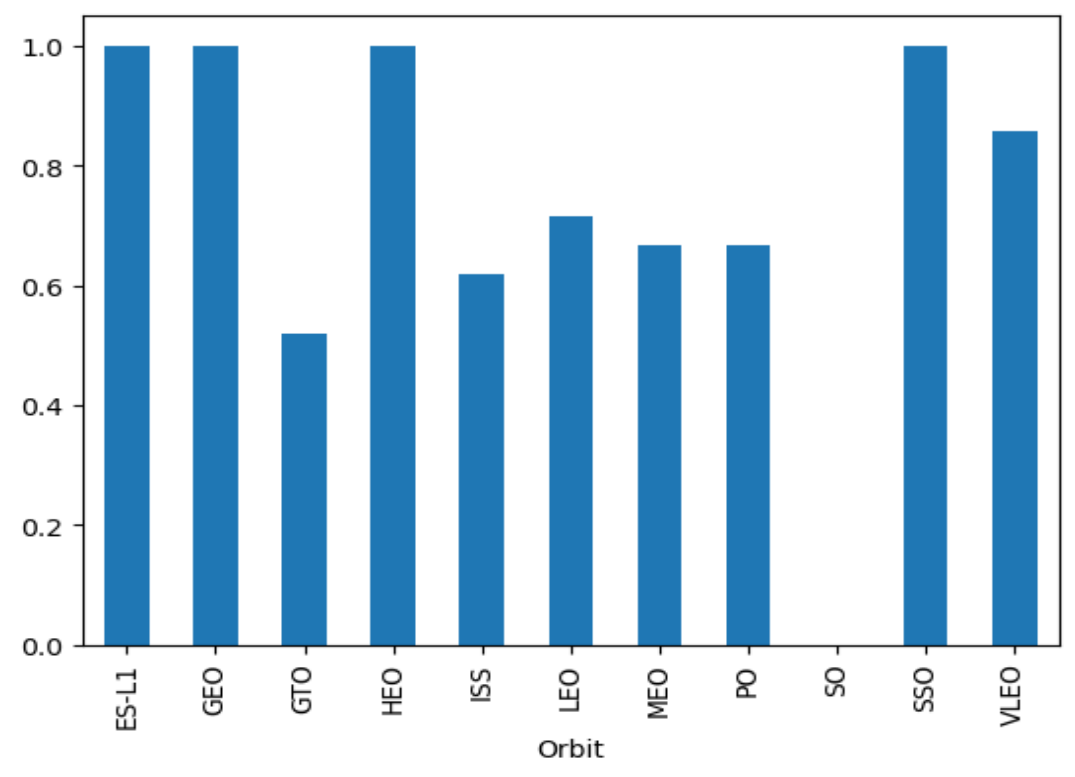
- According to the plot above, CCAF5 SLC 40 had been as pioneer launch site, where from 17th flight is significantly gaining improved success rate over time
- In second place VAFB SLC 4E and third place KSC LC 39A continue the flights
- In general success rate improved over time.

Payload vs. Launch Site



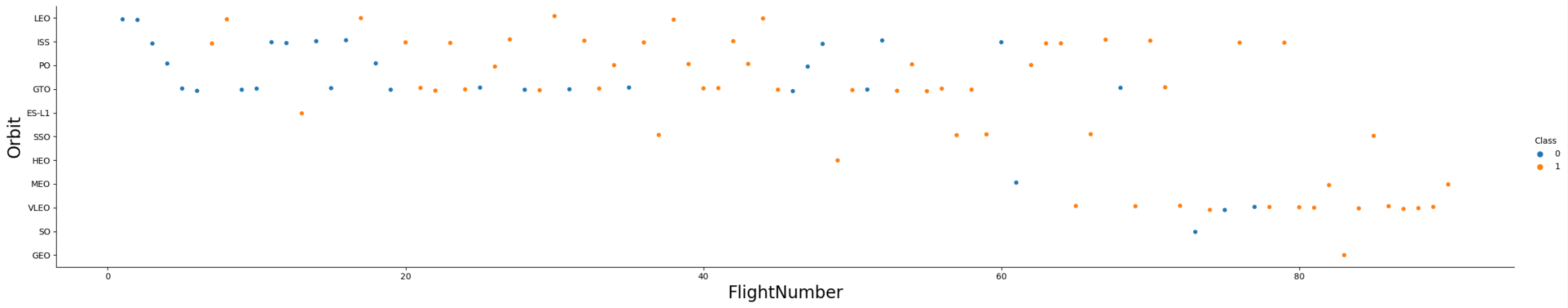
- Payloads over 8.000 kg have significantly high success rate
- VAFB SLC 4E of launch site seems have payloads under 10.000 kg

Success Rate vs. Orbit Type



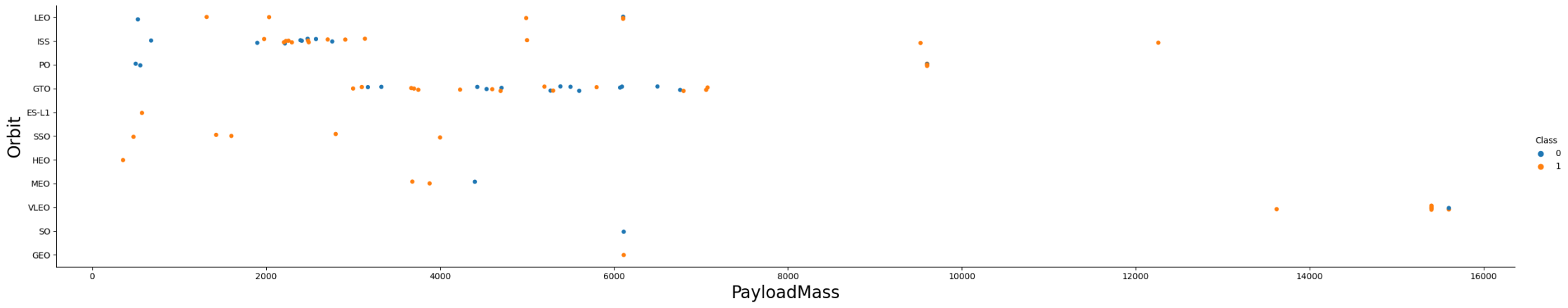
- The biggest success rates happens to orbits:
  - ES-L1
  - GEO
  - HEO
  - SSO

Flight Number vs. Orbit Type



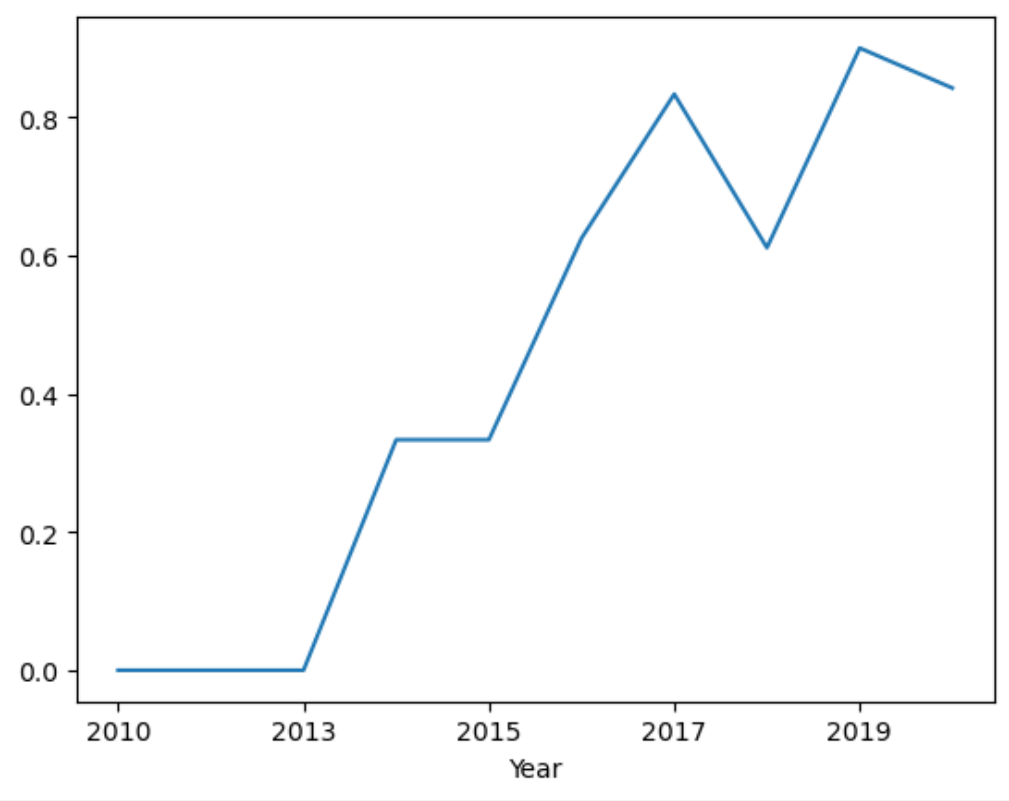
- In general success rate improved over time to all orbits

Payload vs. Orbit Type



- In general there is no relation between payload and success rate to orbit GTO
- ISS orbit and GTO orbit have wide range of payload
- There are few launches to the orbits SO and GEO.

Launch Success Yearly Trend



- The first three years were a period of adjusts and improvement of technology
- Success rate has significant increasing in 2013 until 2020
- Temporarily success rate decrease in 2017 until 2018, but increase 2018 above

All Launch Site Names

- There are 4 launch sites which are CCAFS LC-40, CCAFS SLC-40, KSC LC-39A, and VAFB SLC-4E

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

Launch Site Names Begin with 'CCA'

- 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
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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	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

Total Payload Mass

- The total payload carried by boosters from NASA whose codes contain “CRS” is 111.268 kg

total_payload
111268

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1 by filtering data by booster version above, the result we obtained the value of 2.928 kg.

avg_payload
2928

First Successful Ground Landing Date

- The first successful landing outcome on ground pad with minimum value for date to identify the first occurrence is happened on 12/22/2015

first_success_gp
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000, below is the result

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

Total Number of Successful and Failure Mission Outcomes

- Total number of successful and failure mission outcomes, below mission outcome and occurrence for each group:

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

Boosters Carried Maximum Payload

- The booster which have carried the maximum payload mass is below

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

2015 Launch Records

- Failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015 only have 2 occurences:

booster_version	launch_site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

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

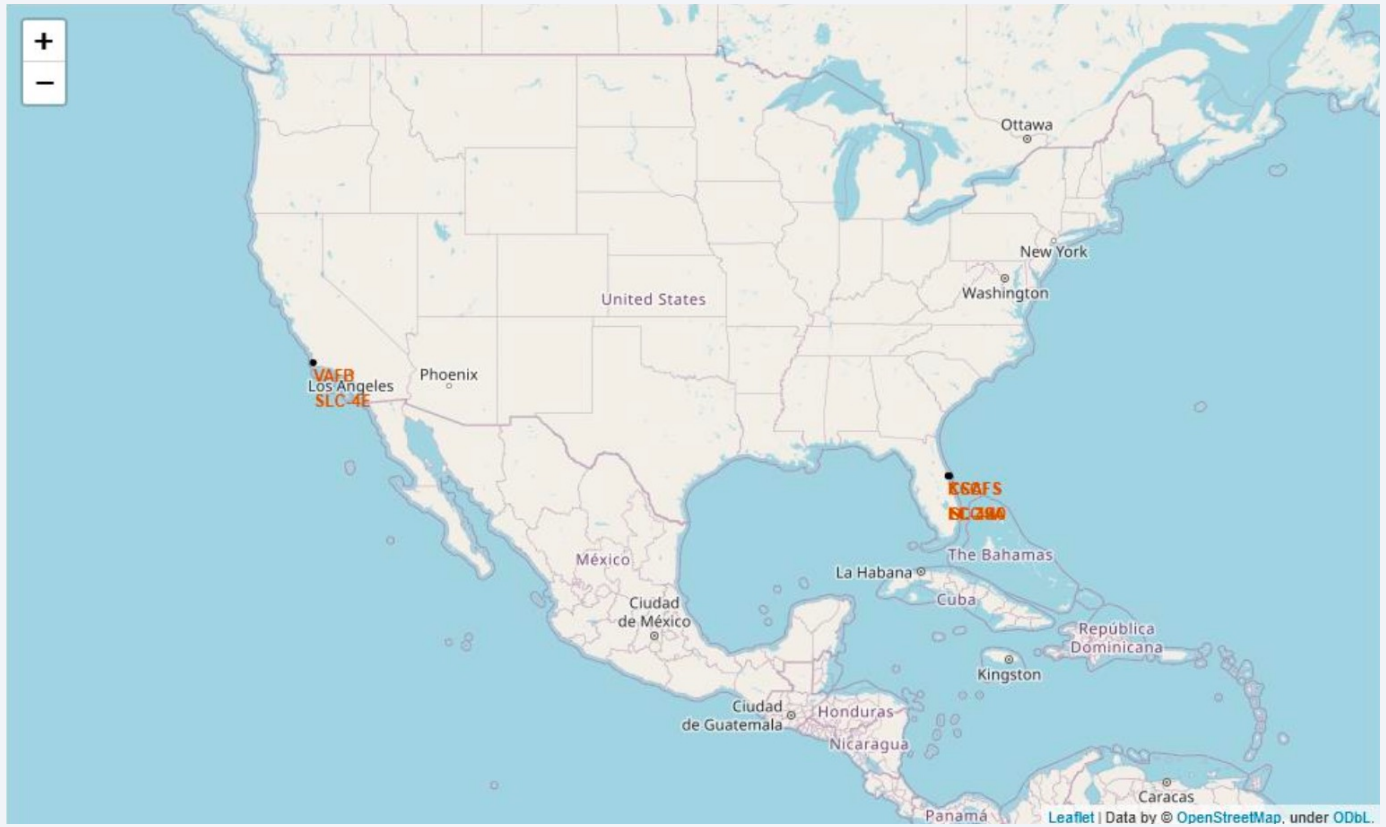
- Ranking of landing outcomes Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

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

Section 3

Launch Sites Proximities Analysis

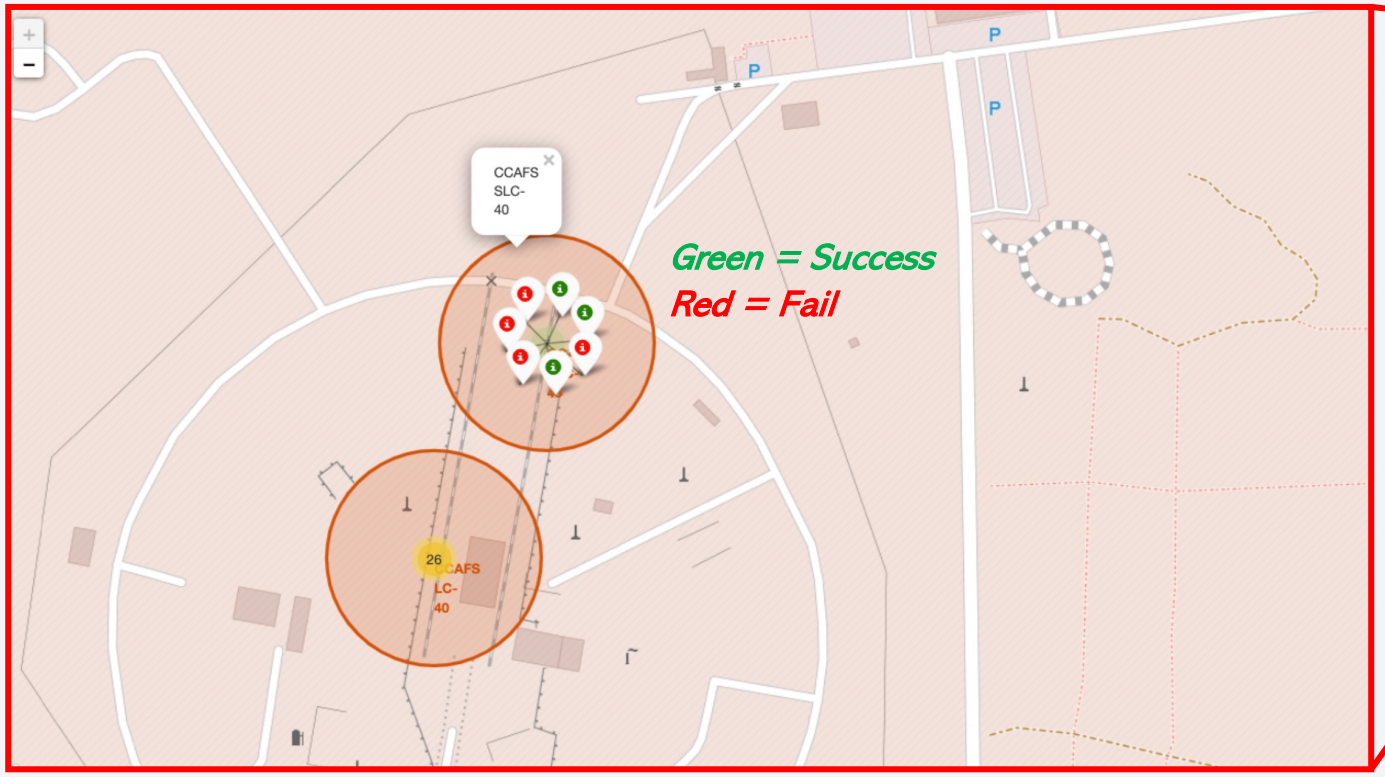
All Launch Sites

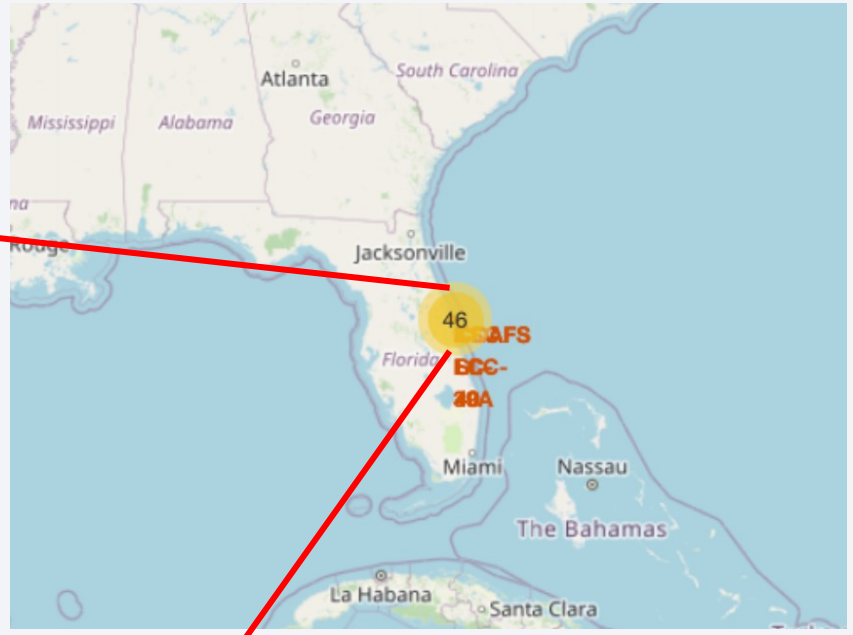


- Launch sites are still near the existing infrastructure, hence it can be still access with road or railroad

Launch Outcomes by Site

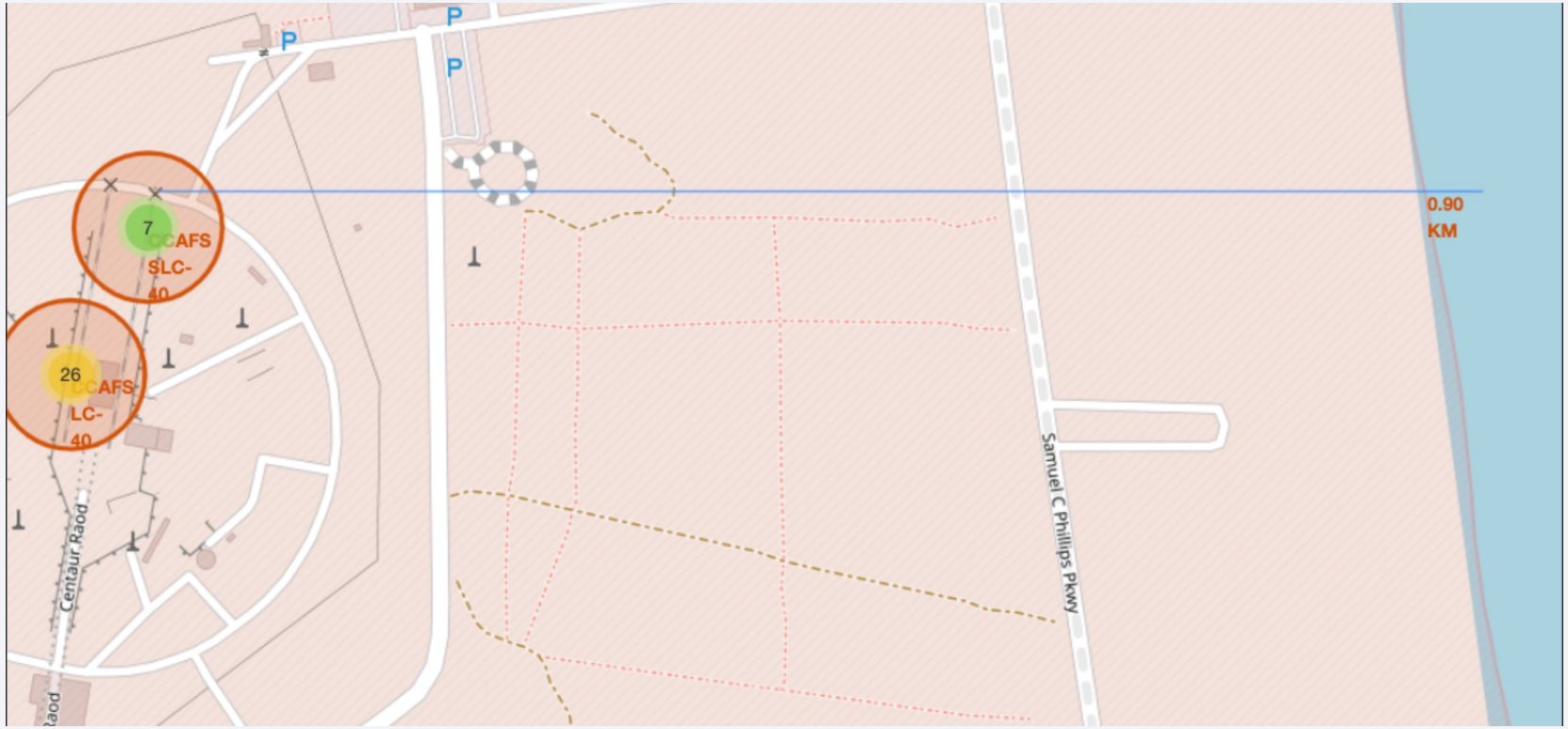
- Example of CCAFS SLC-40 launch site





Logistic and Safety

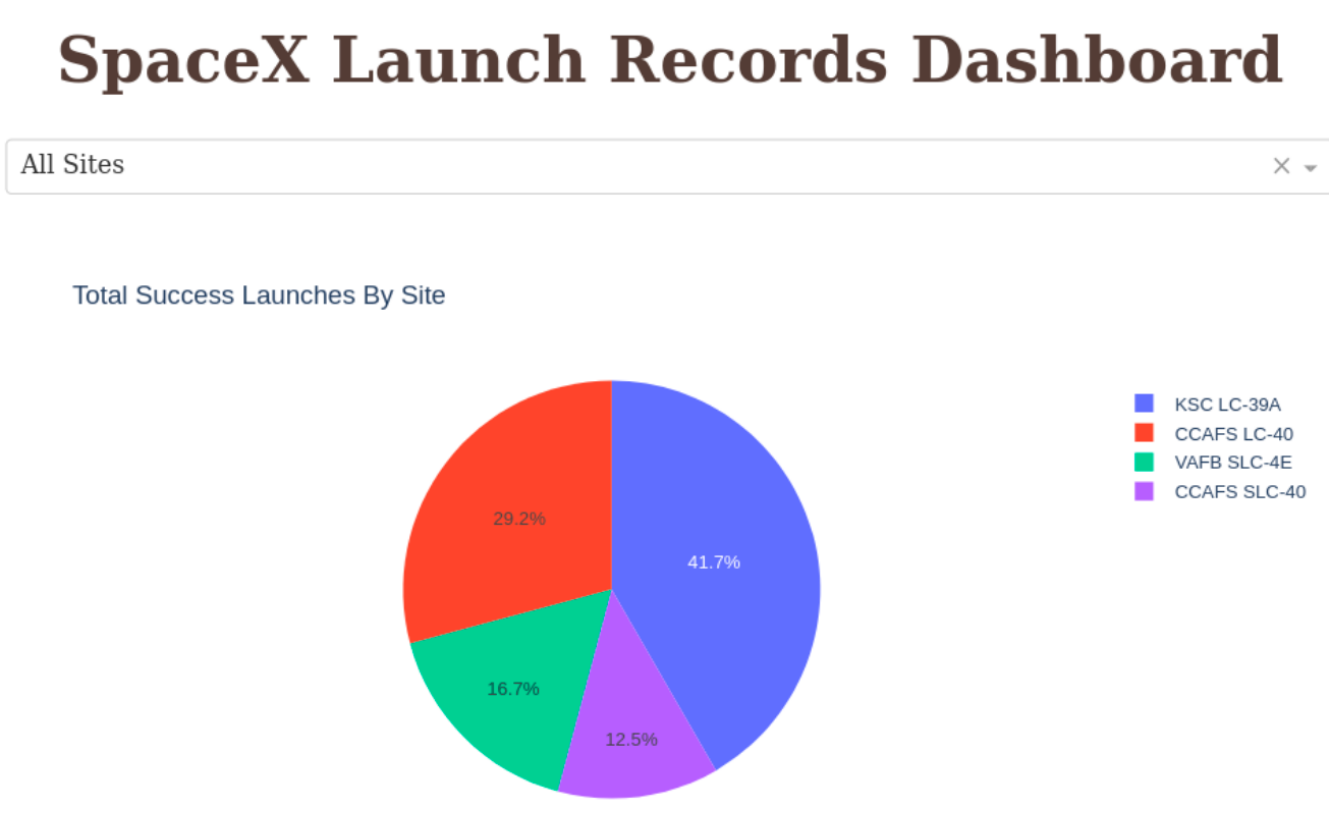
- Launch site CCAFS SLC-40 has good access of transportation which relatively near with coastline



Section 4

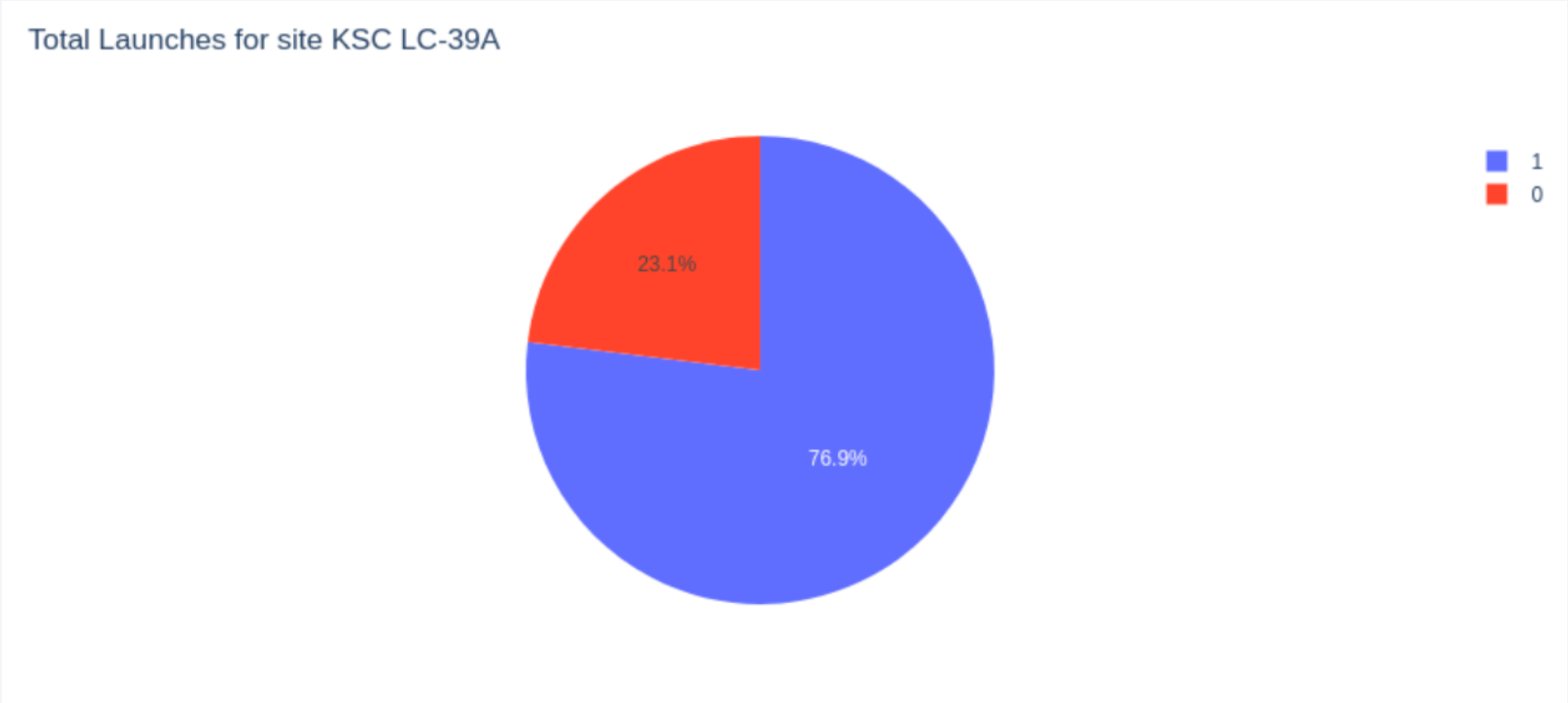
Build a Dashboard with Plotly Dash

Succesful Launches by Site



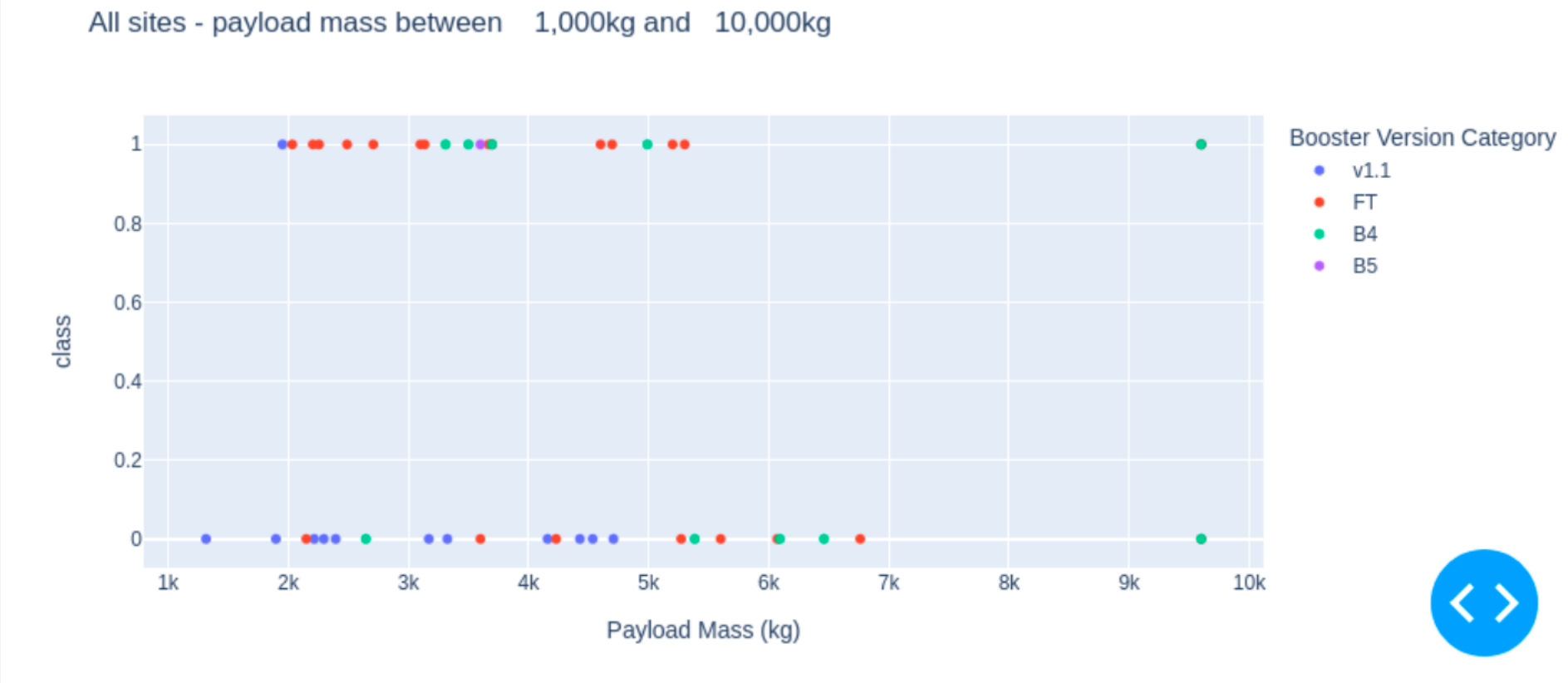
Mostly launch site from KSC LC-39A. The place for launch site seems could be factor of launch success

Launch Success Ratio for KSC LC-39A



- 76.9% of launches are successful on KSC LC-39A

Payload vs Launch Outcome



- Payload between 1.5k and 5.5k with booster version FT are the great combination for highly chance successful launch

Section 5

Predictive Analysis (Classification)

Classification Accuracy

- After modelling using Logistic Regression, Support Vector Machine, Decision Tree Classifier and KNeighbors Classiffier, the accuracy of those models are plotted fig above
- The model with highest accuracy is Decision Tree Classifier with value of 89%. Other models like Logistic Regression and Support Vector Machine are seems overfitting with accuracy of test < accuracy of train

Confusion Matrix

- Confusion matrix of the Decision Trees Classifier shows the high value of True Positive and True Negative compared to low value of False Positive and False Negative.

Conclusions

- The best launch sites is KC LC-39A
- Orbit types that tend high chance are ES-L1, GEO, HEO, SSO
- Even though launch outcomes are tend to highly successful, landing outcomes could improving along the time based on improvement of the technology
- In order to predict successful landings, machine learning model that can be fit is Decision Tree Classifier, hence the company can tight the cost and increase the profits

Appendix

- Folium didn’t show maps on Github, therefore screenshot is uploaded

Thank you!