IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

RUI.XU
2023.5.10

Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

- Project background and context
- Problems you want to find answers

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - To know where are the data stored
  - Use the python`s tool to scratch information
- Perform data wrangling
  - To deal with the confusion data(like missing values)
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection – SpaceX API

- **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**

Reason

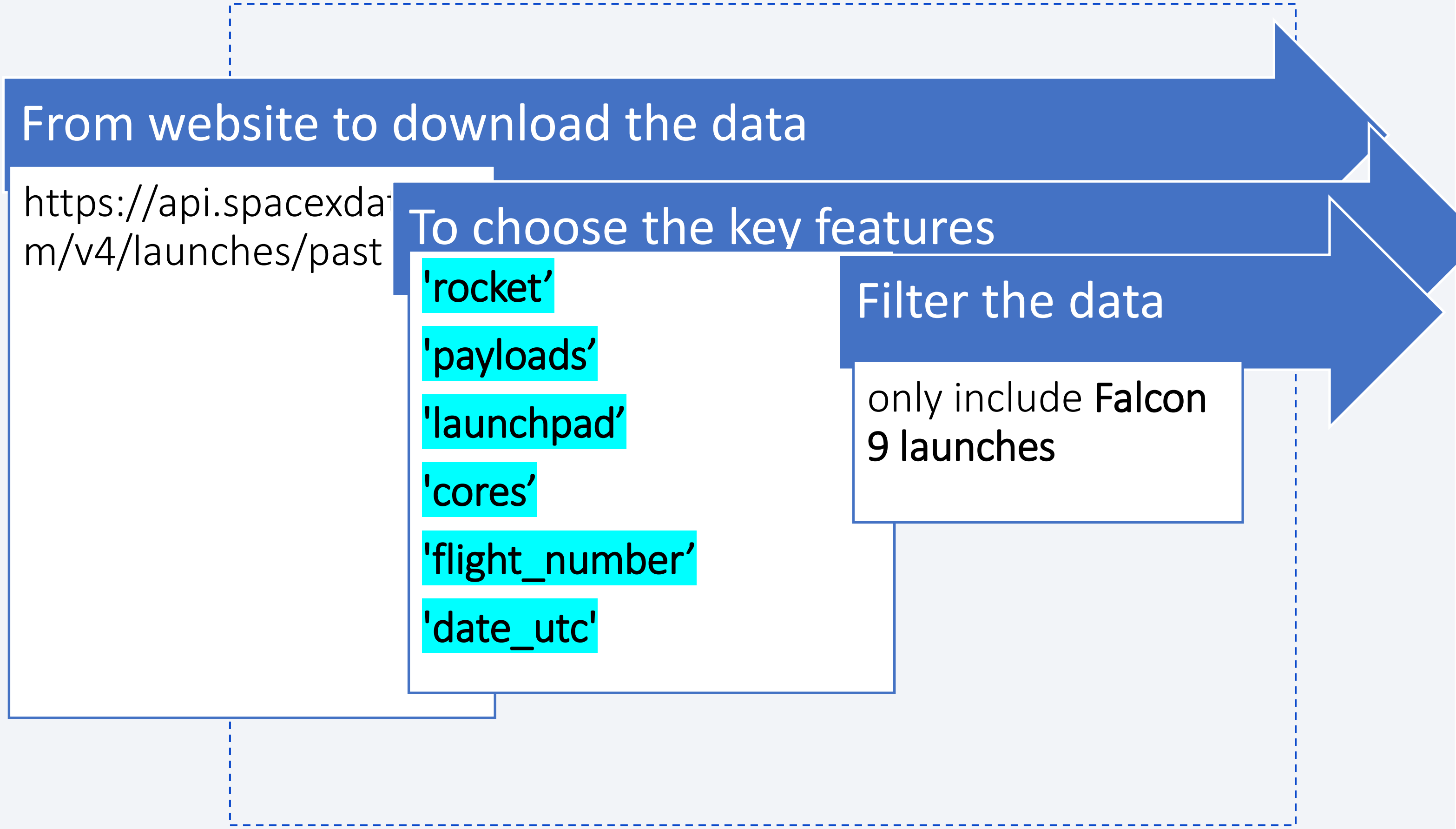
- Reuse the first stage

Study

- The key factors
  - Frist stage
  - successfully
- Object
  - The Falcon 9

Data Collection - Scraping

- Use the structure of the website , analysis the content
- Discover relevant information
- Get 200 records from the website
- Normalize the data form as json
- Get the main columns
- Transform form like dataframe



https://github.com/yvonne1989-cloud/finaltest/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Data Wrangling

- Deal with the null data
  - As use the code—(isnull)—to find how much null were there
  - We find 5 of PlayloadMass and 26 of LandingPad

- Deal with the missing data
  - Use mean of those data to fill
  - As result we can get a table with no missing values

https://github.com/yvonne1989-cloud/finaltest/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

We use about three plots to show the relationship between those factors:

1.**Bar**:can help us better understand and communicate data, compare the size and trend of different categories or metrics, and identify outliers and trends from them

2.**Catplot**: used to display the relationship between one continuous variable and one or more categorical variables

3.**Line**: displays the relationship between two variables as a series of connected points

https://github.com/yvonne1989-cloud/finaltest/blob/main/jupyter-labs-eda-dataviz.ipynb

EDA with SQL

- 1. SELECT DISTINCT [col] from[table]
- 2. SELECT * FROM [table] WHERE [conditions]
- 3. SELECT [formal] FROM [table] WHERE [conditions]
- 4. SELECT [col1],[col2] FROM [table] GROUP BY [col3]

https://github.com/yvonne1989-cloud/finaltest/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Markers: Used to mark a point on the map and associate it with a specific location.
- Circles: Created with a specific location as the center, typically used to represent an area or a radius.
- Lines: Used to connect two or more points on a map, continuous lines can be used to represent routes or paths.
- Popups: Used to display a text box when a user clicks on a location on the map. Popup boxes can be used to add explanations or descriptions for locations or issues that appear on the map.
- Tiles: Used to set the base style of the map, such as setting a map theme, satellite imagery, or a black and white map.

https://github.com/yvonne1989-cloud/finaltest/blob/main/lab_jupyter_launch_site_location.ipynb

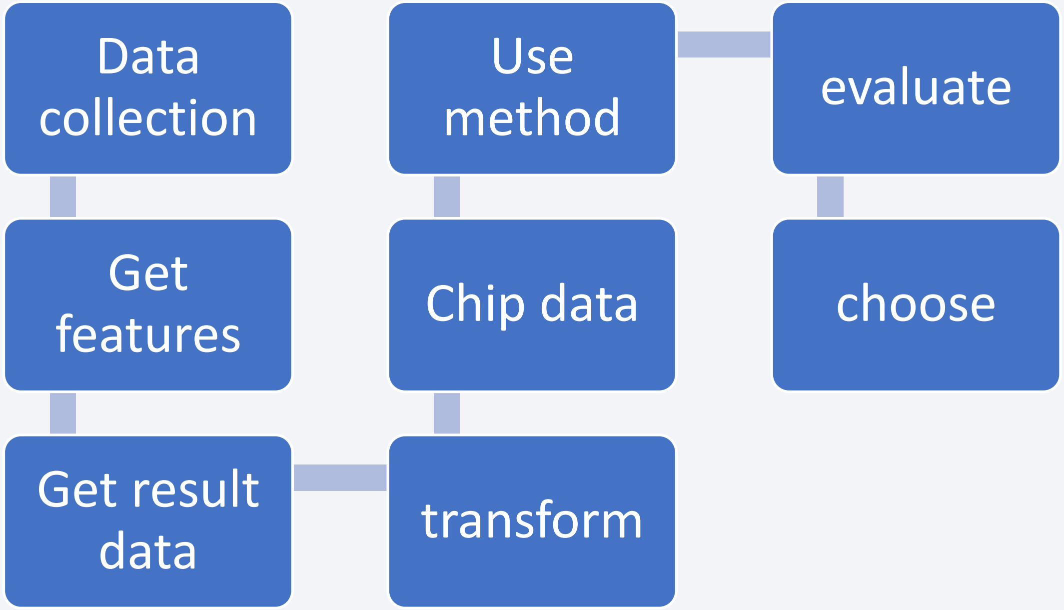
Build a Dashboard with Plotly Dash

- We got two graphs in this dash
  - 1.Pie:when we choose the different locations, they will show the percentage of success
  - 2.Scatter:according to the range of payload , they will show the distributions
- It`s a explore of looking for the factors how to effect the results

https://github.com/yvonne1989-cloud/finaltest/blob/main/Dashboard%20Application%20with%20Plotly%20Dash.py

Predictive Analysis (Classification)

- We have these steps to get the result:
  - 1.get class of success
  - 2.transform these factors
  - 3.chip the data between test and train
  - 4.use LR\SVC\CSV\Decision Tree\KNN\ method
  - 5.use confusion matrix to evaluate
  - 6.choose the best way to predict



https://github.com/yvonne1989-cloud/finaltest/blob/main/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

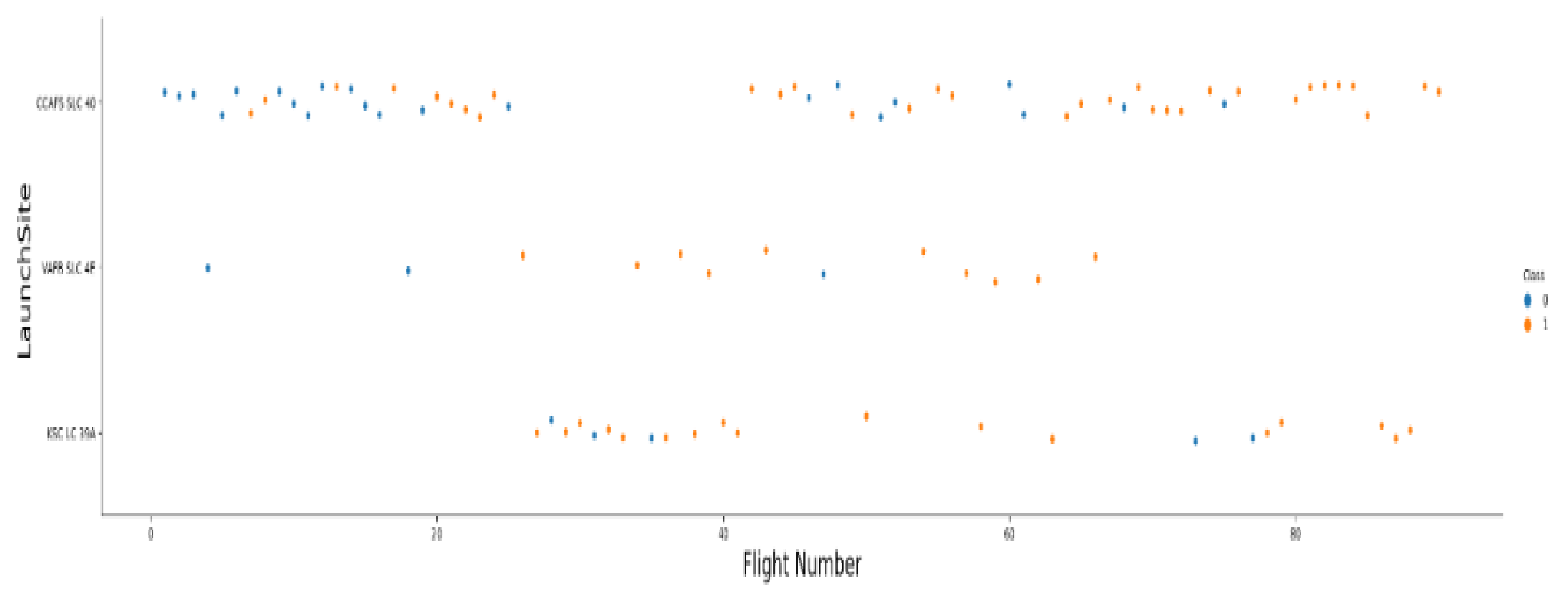
Results

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

Section 2

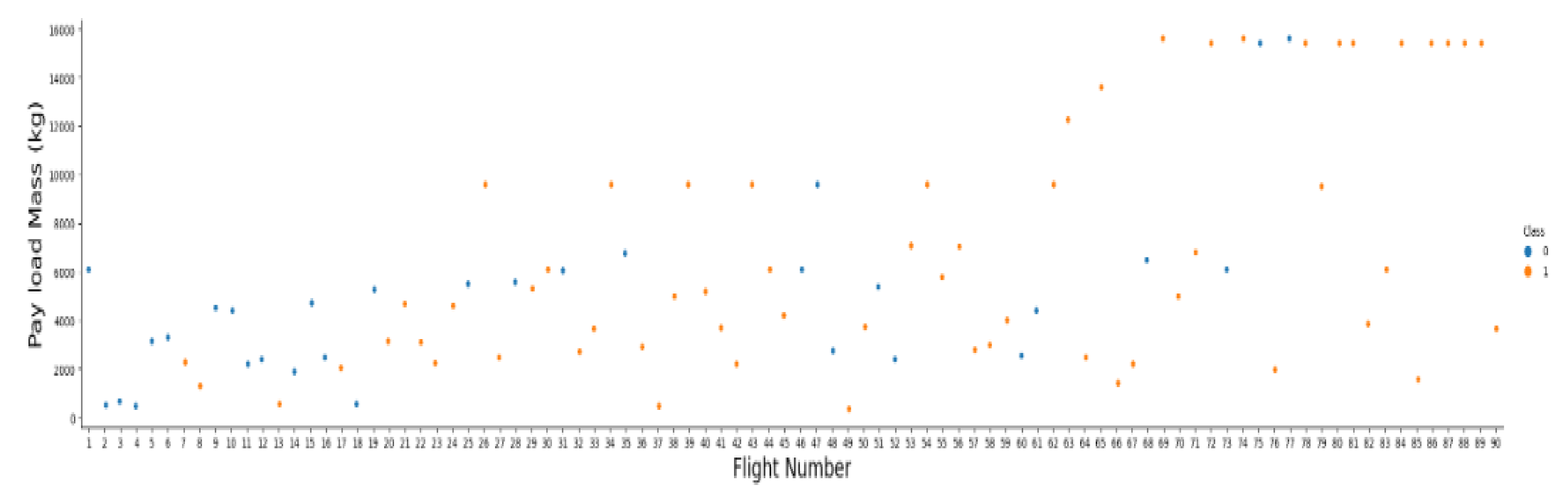
Insights drawn from EDA

Flight Number vs. Launch Site



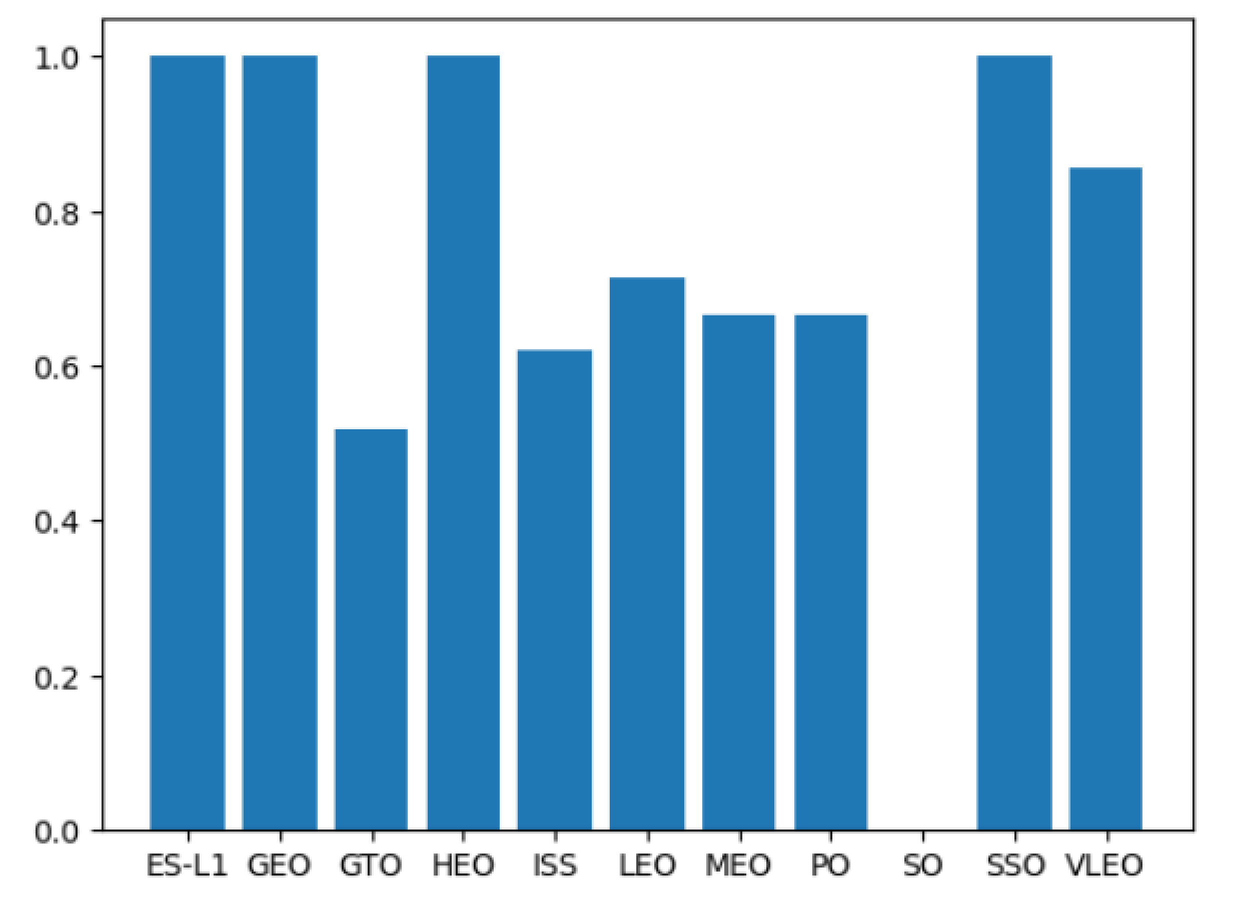
To the end of the tests, CCAFS LC-40 have got a higher successful percentage.

Payload vs. Launch Site



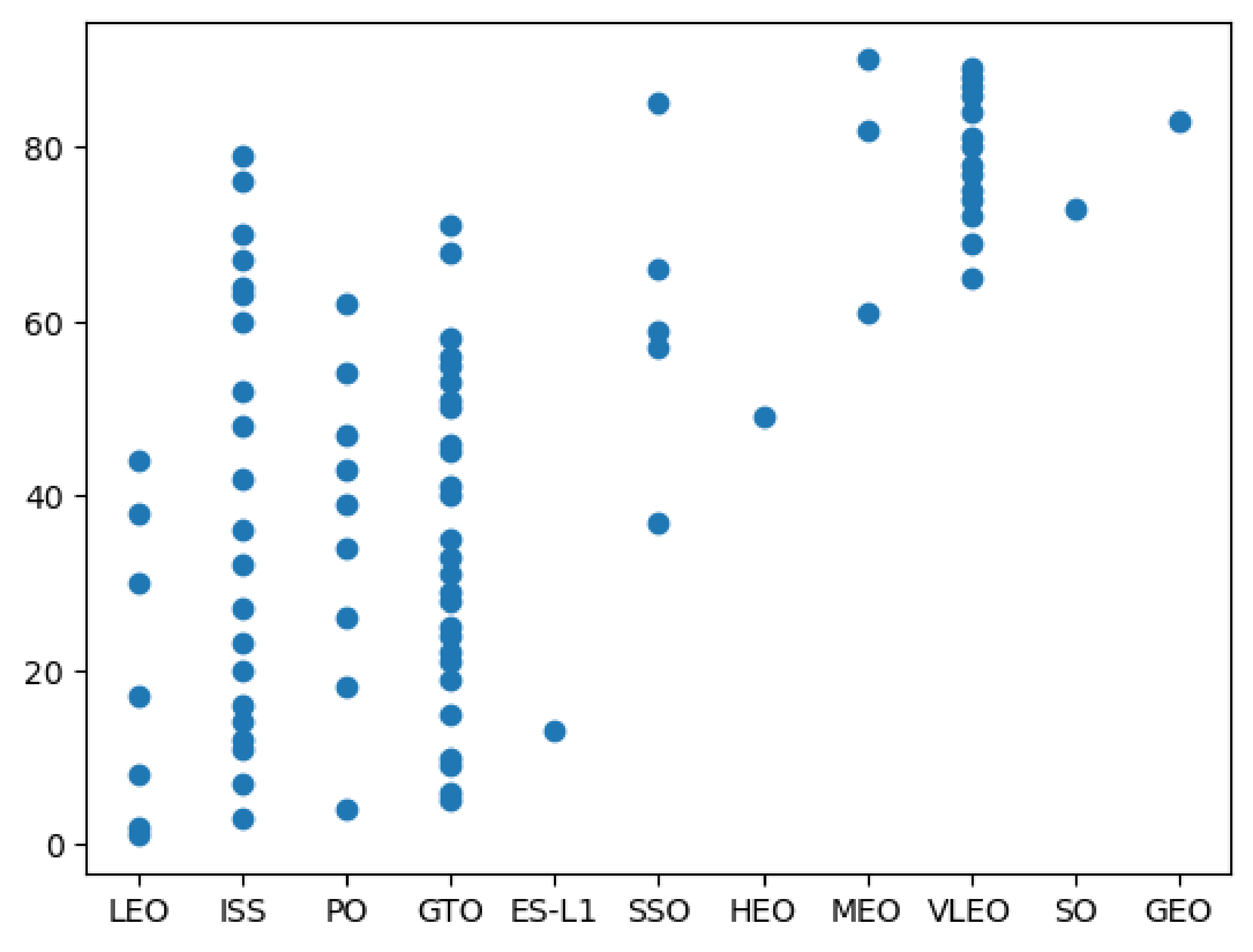
At the beginning of the test, they focus on the lighter load mass . The further we go, the more they choose heavy lift rockets.

Success Rate vs. Orbit Type



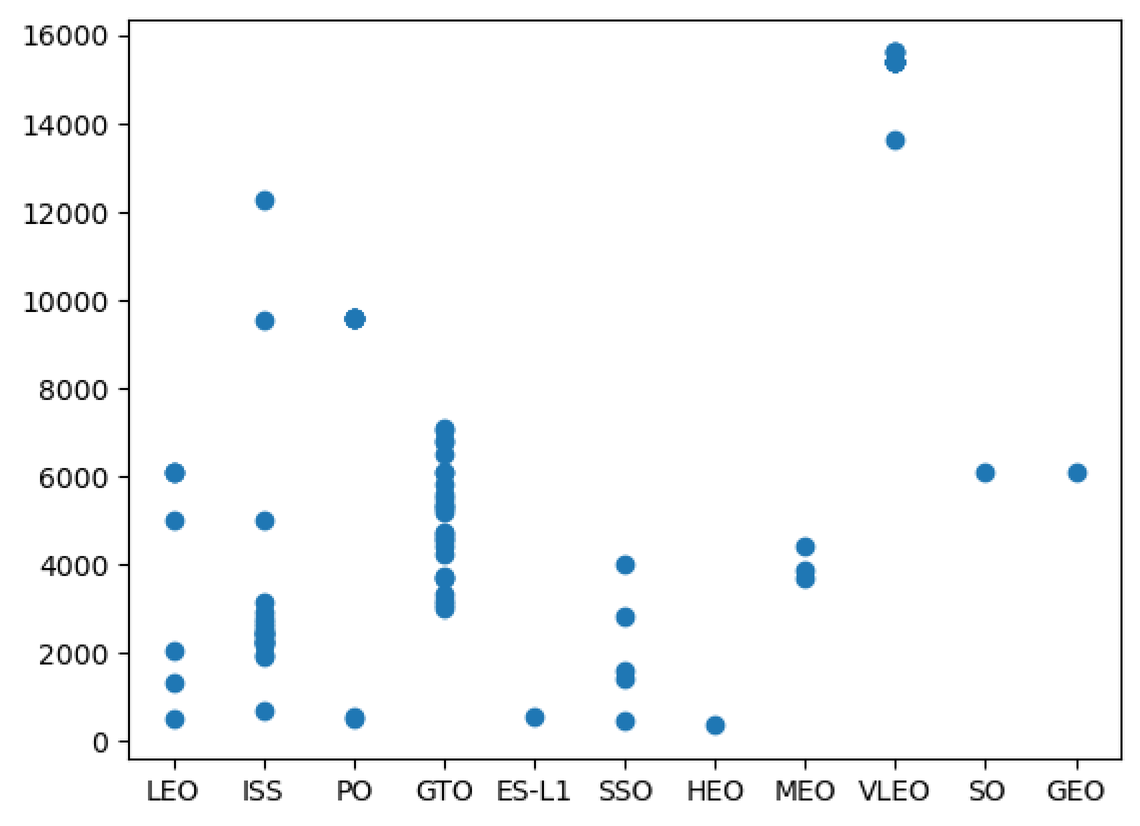
Often we think the maintain safe orbit.,
have a better condition to submit.

Flight Number vs. Orbit Type



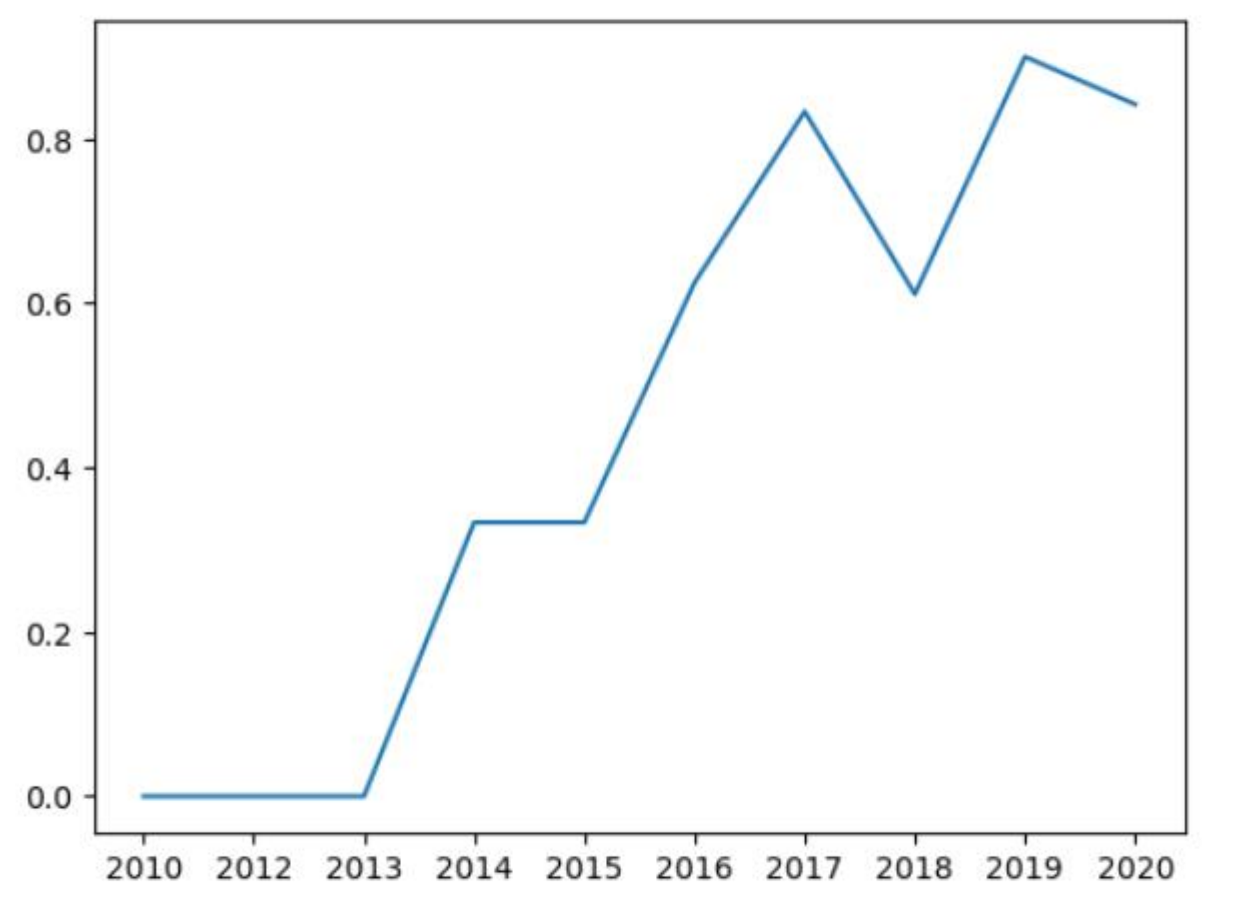
Taking into account the safety of the orbit, a rocket launch would choose this type of orbit.

Payload vs. Orbit Type



With heavy payloads the successful landing or positive landing rate are more for PO,LEO and ISS

Launch Success Yearly Trend



Overall, from 2010 to 2020, the success rate of space launches has significantly improved, growing from 0% to around 80%.

All Launch Site Names

Launch_Site
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

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

AVG(PAYLOAD_MASS__KG_)
2534.6666666666665

First Successful Ground Landing Date

MIN(Date)
01-05-2017

Successful Drone Ship Landing with Payload between 4000 and 6000

Booster_Version
F9 FT B1032.1
F9 B4 B1040.1
F9 B4 B1043.1

Total Number of Successful and Failure Mission Outcomes

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

Boosters Carried Maximum Payload

Booster_Version	MAX(PAYLOAD_MASS__KG_)
F9 B4 B1039.2	2647
F9 B4 B1040.2	5384
F9 B4 B1041.2	9600
F9 B4 B1043.2	6460
F9 B4 B1039.1	3310
F9 B4 B1040.1	4990
F9 B4 B1041.1	9600
F9 B4 B1042.1	3500
F9 B4 B1043.1	5000
F9 B4 B1044	6092

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

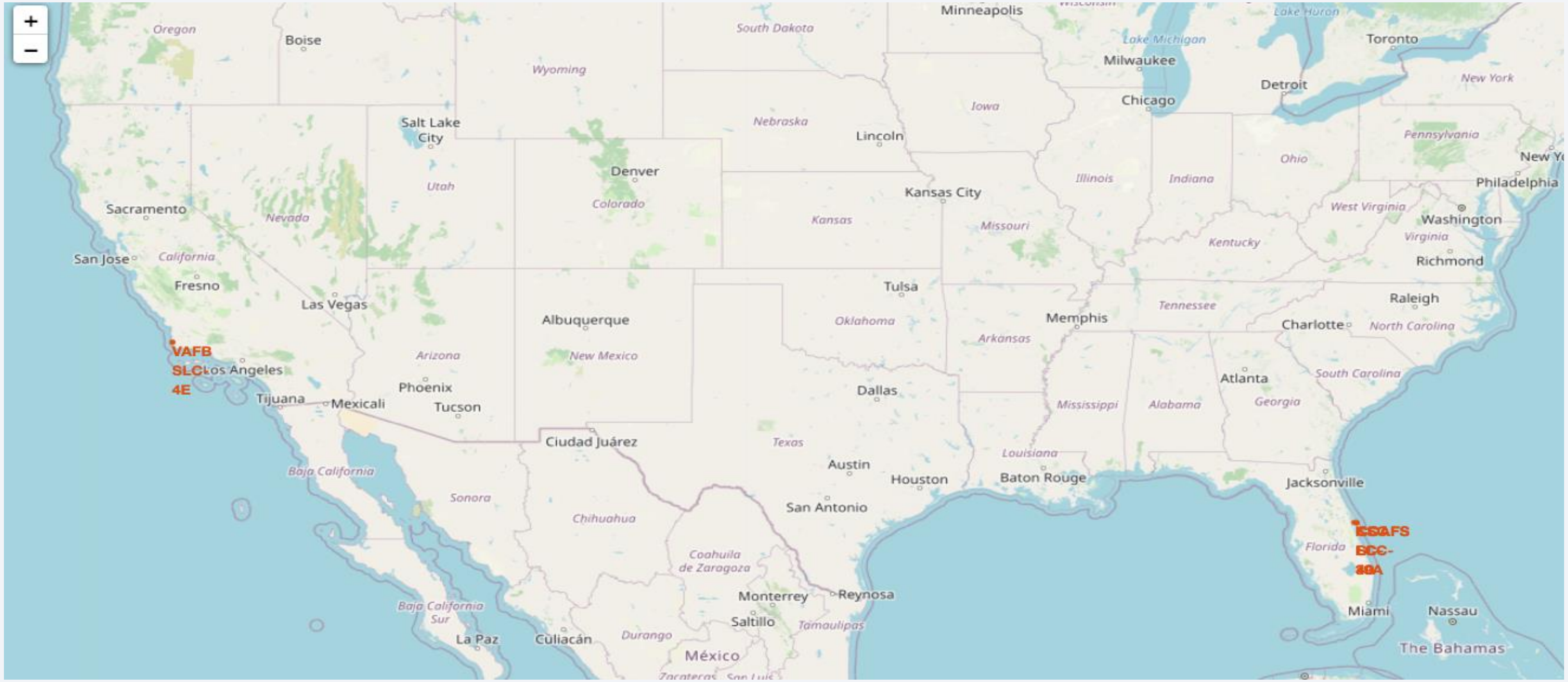
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

Landing_Outcome	COUNT(*)
Success	20
Success (drone ship)	8
Success (ground pad)	7

Section 3

Launch Sites Proximities Analysis

<Folium Map Screenshot 1>



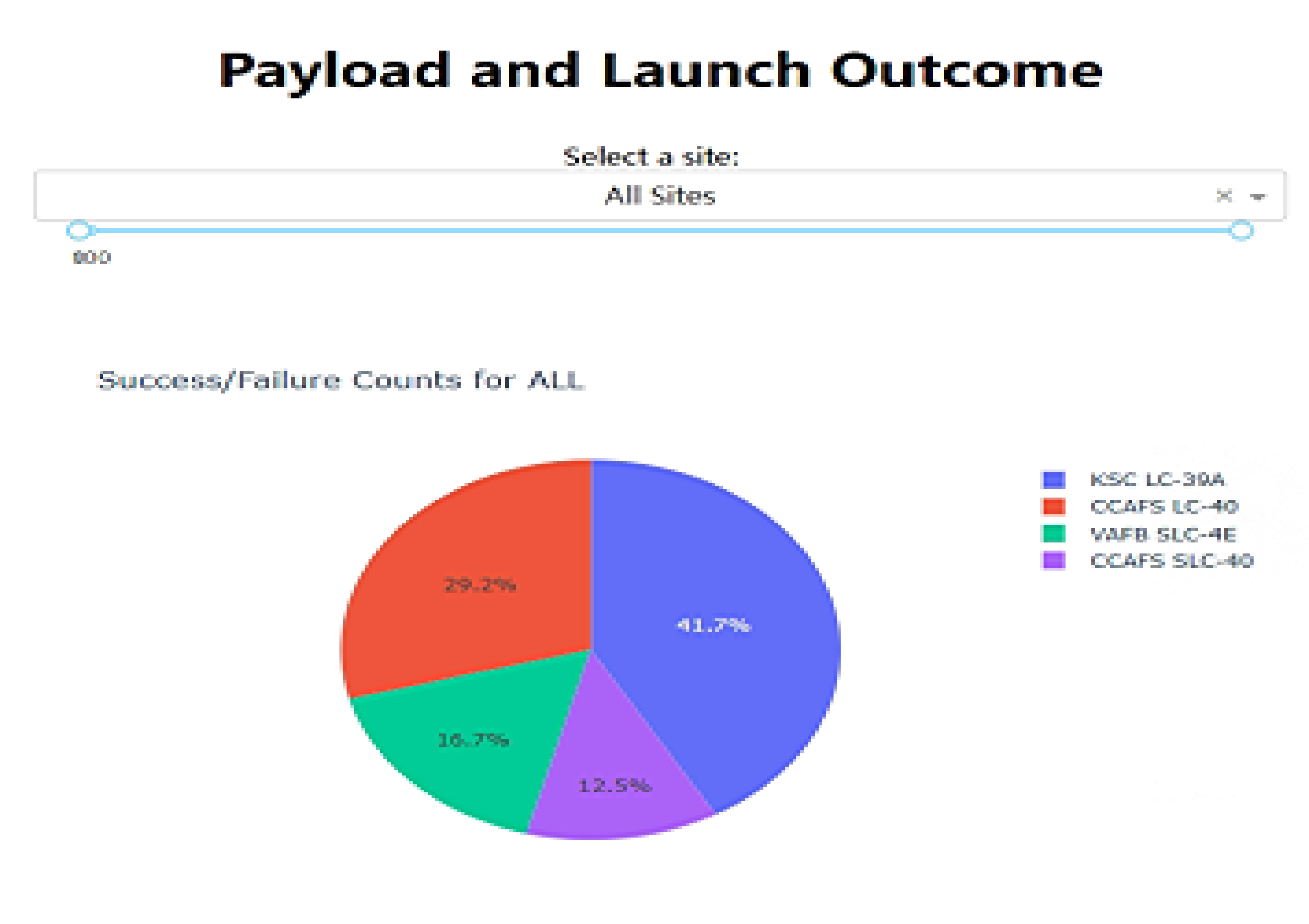
<Folium Map Screenshot 2>

<Folium Map Screenshot 3>

Section 4

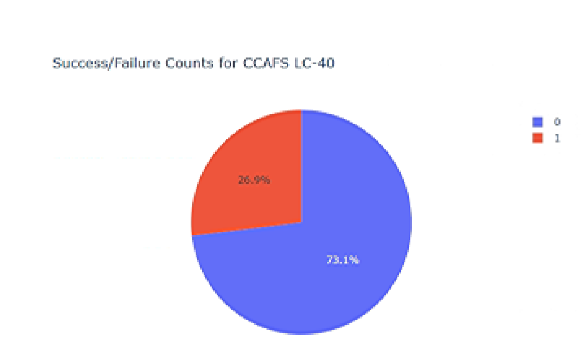
Build a Dashboard with Plotly Dash

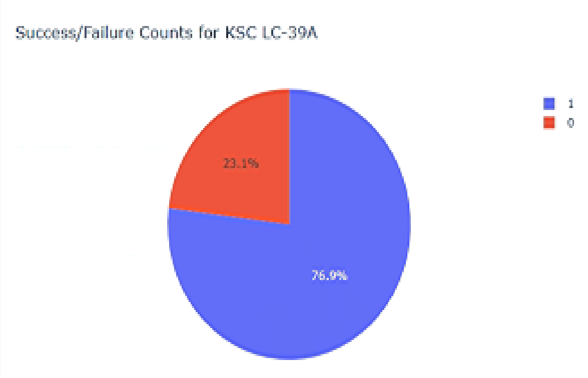
Payload and Launch Outcome

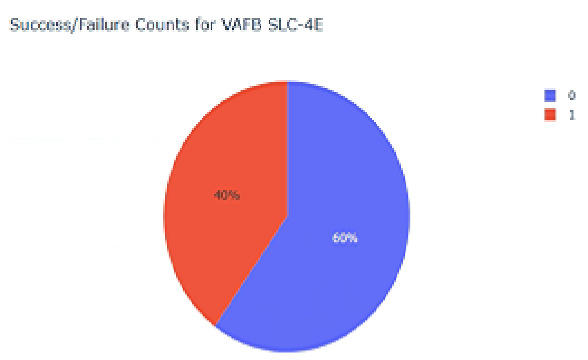


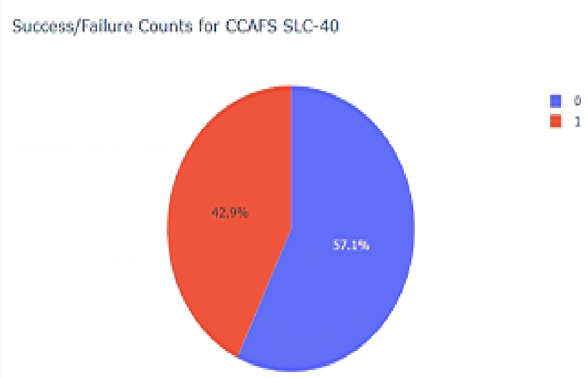
The biggest part is CCAFS LC-40

The part of KSC LC-39A Launch Outcome



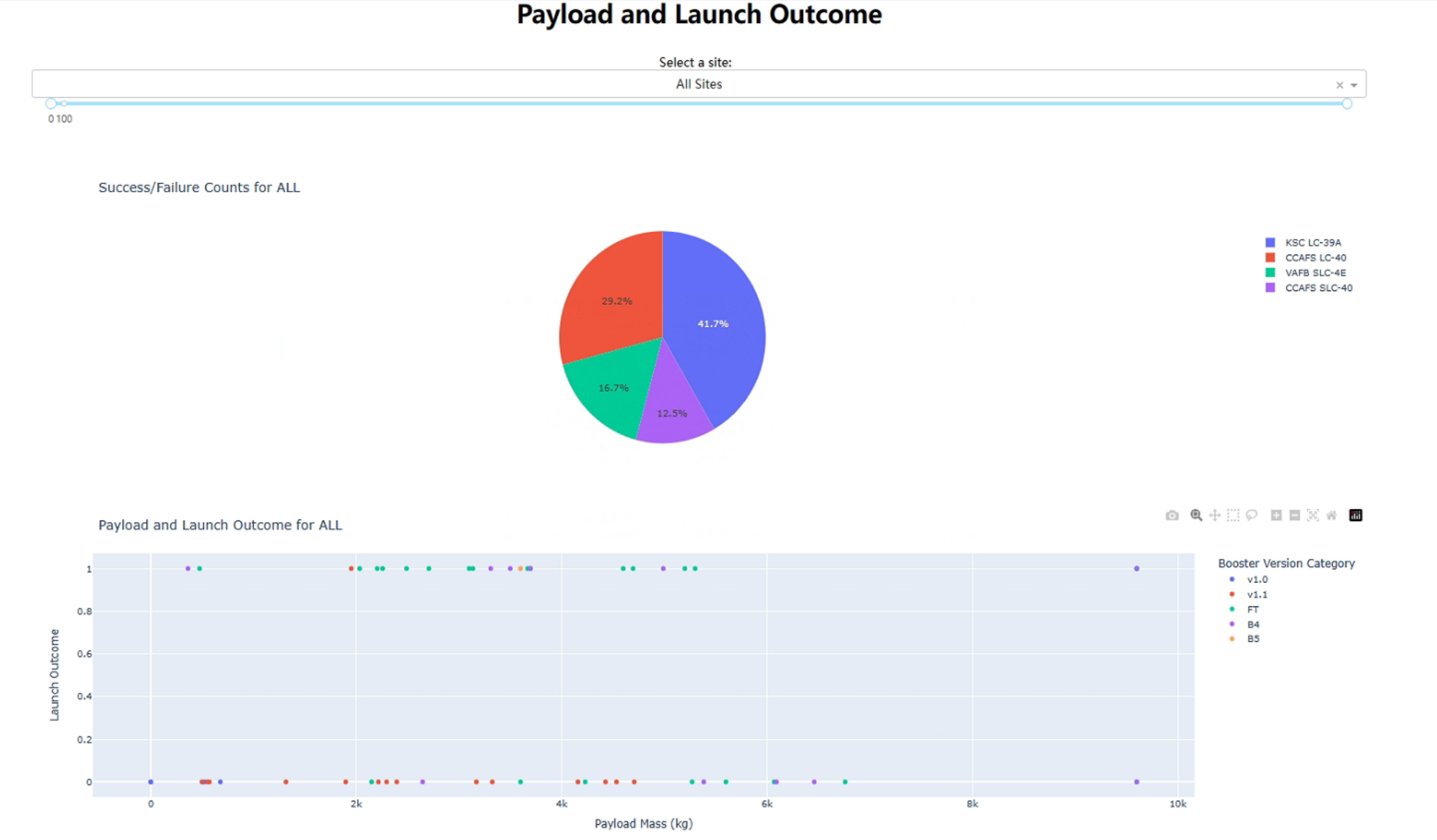






- As we see, the highest successful percentage is KSC LC-39A,is getting 76.9%.

Payload and launch outcome for all

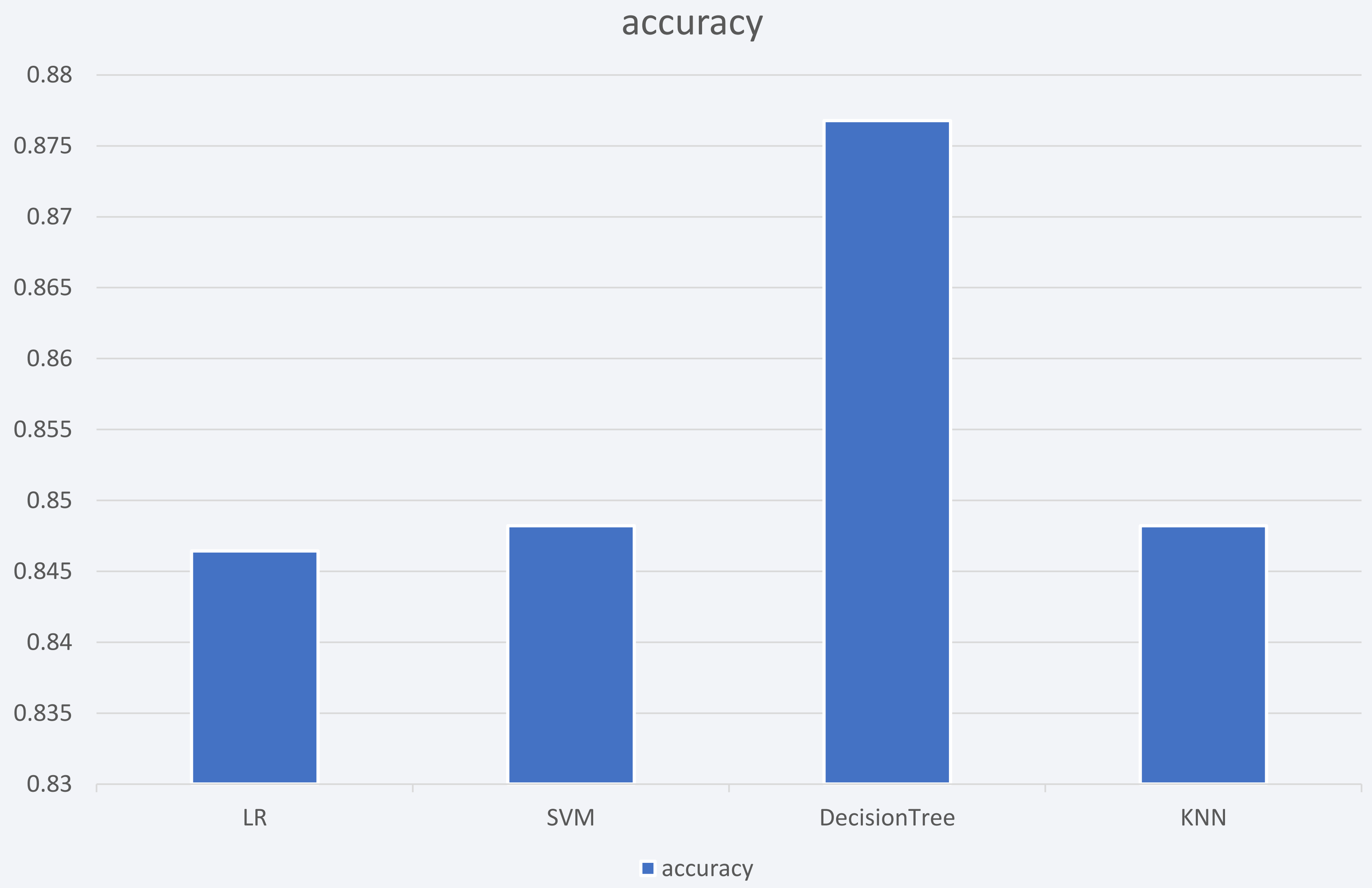


From 2k to 6k may be the best range of payload for submit

Section 5

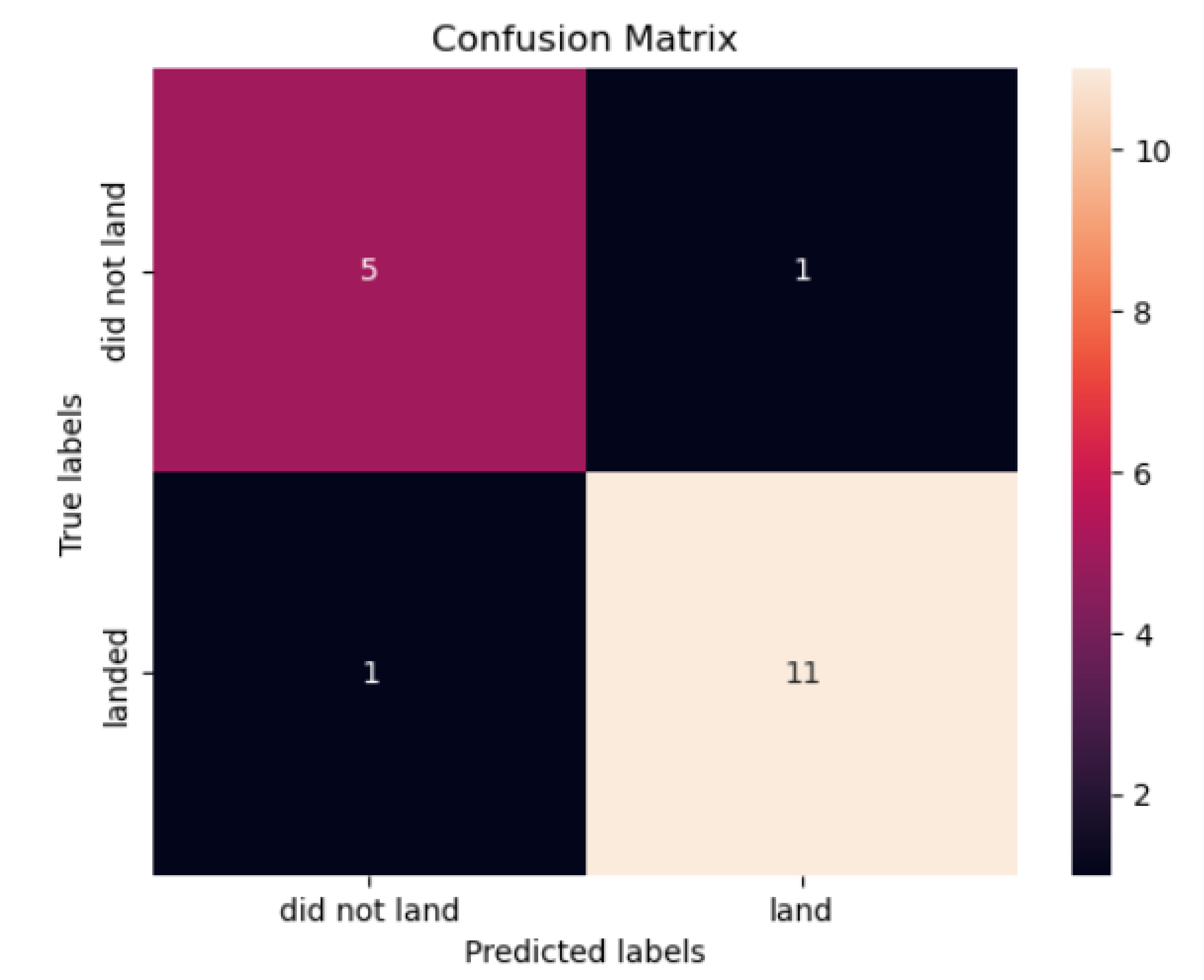
Predictive Analysis (Classification)

Classification Accuracy



DecisionTree is the biggest.

Confusion Matrix



Conclusions

- Location: CCAFS LC-40
- Loadmass: Avg 2534.66
- Orbit: ISS\LEO\PO
- Time: After 2017

Thank you!