IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Jiayi Li
2023.6.14

Outline

- Executive Summary
- Introduction
- Methodology
- Data collection
- Data wrangling
- Results
- Data analysis
- Data visualization
- Conclusion
- Appendix

Executive Summary

- To review and analyze available data on SpaceX
- Using different methods for analysis
- EDA with SQL and Python
- Machine learning
- Others

https://github.com/LJY715/Capstone-project/tree/main

Introduction

- Objective - This project is carried out to predict and evaluate the cost and outcome of rocket launch based on website data

- Problems – the impact of rocket landing, under what conditions can SpaceX get the best result

Section 1

Methodology

Methodology

- Data collection methodology:
  - Data collection thru API
  - Web wrangling thru Pandas
- Perform data wrangling
  - Clean data by delete invalid data
  - Encode to make them reflect in correct format
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Building, tuning, evaluating classification models

Data Collection

Collect data with API	Create beautiful soup
Convert data and update row and column titles	Get column title and create launch dict
Filter data to focus on falcon9	Convert to data frame
Convert file to CSV	Convert file to CSV

Data Collection – SpaceX API

- Collect data with API calls
- Convert to data frame
- Update row and column
- Filter data to keep falcon 9 only
- Convert to CSV file

Data Collection - Scraping

- Create beautiful soup
- Get column titles
- Create launch dict
- Convert to data frame
- Convert to CSV file

Data Wrangling

Starting with exploratory data analysis, calculate the number of launches by different sites, then calculate the number and occurrence of mission outcome per orbit, then export the data to CSV file, finally to create landing outcomes labels.

EDA with Data Visualization

- Explore the data, by using scatterplots to visualize the relationship

EDA with SQL

- Names of the unique launch sites in the space mission
- Top 5 launch sites whose name begin with the string ‘CC
- Total payload mass carried by boosters launched by NASA (CRS)
- Average payload mass carried by booster version F9
- List the date when the first successful landing outcome was achieved
- List names of the boosters that have success in drone ship and have payload mass greater than 4000 but less than 6000
- List total number of successful and failure mission outcomes
- List booster versions names that have carried the maximum payload mass
- Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

Build an Interactive Map with Folium

- Utilized markers, circles, lines in the folium maps.
- Mark can indicate points, e.g. launch sites
- Circles are above marks on the map
- Lines can indicate the distance

Build a Dashboard with Plotly Dash

- Dashboard is a very important tool for visualization
- Use plotly to create graphs, charts and scatterplots.

Predictive Analysis (Classification)

- Compare 4 different classification models, including logistic regression, support vector machine, decision tree and K-nearest neighbors.

- To prepare and clean the data first, then test each models and compare the results.

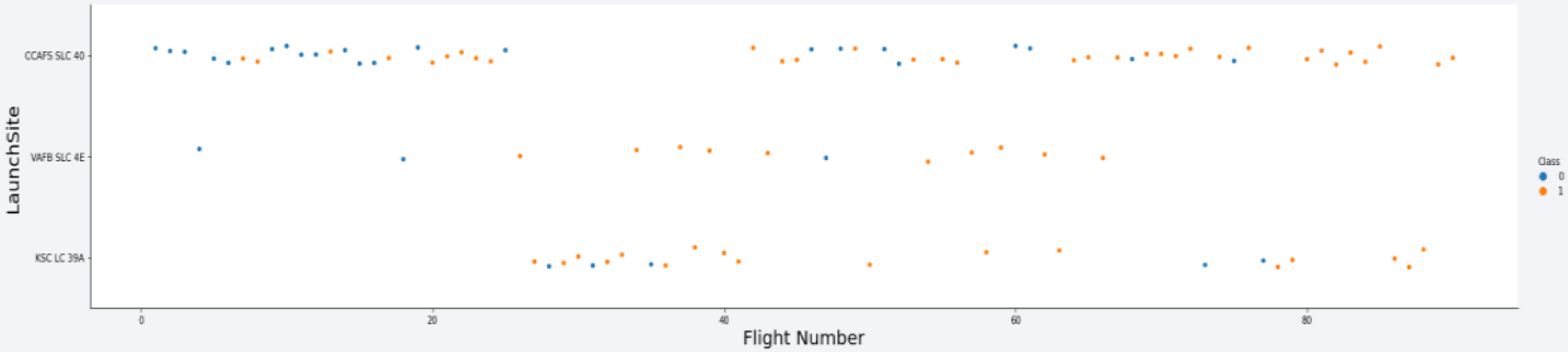
Results

- There are 4 different launch sites
- Basically all mission outcomes are successfully
- The landing successful rate gets better with time going by
- Most launches happened at east coast

Section 2

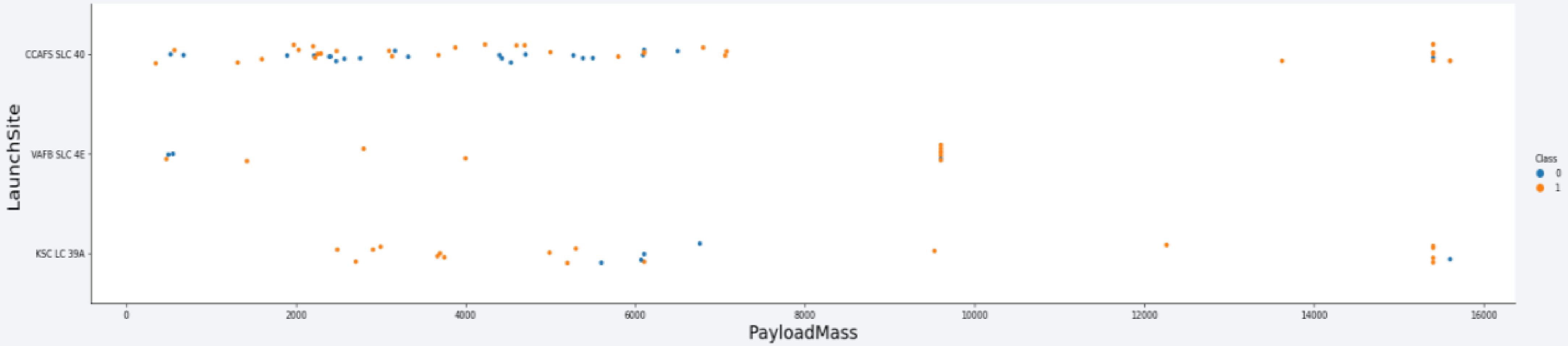
Insights drawn from EDA

Flight Number vs. Launch Site



There's positive relationship between the number of flights and success rate

Payload vs. Launch Site



with payload mass increase, the success rate increase as well

Success Rate vs. Orbit Type

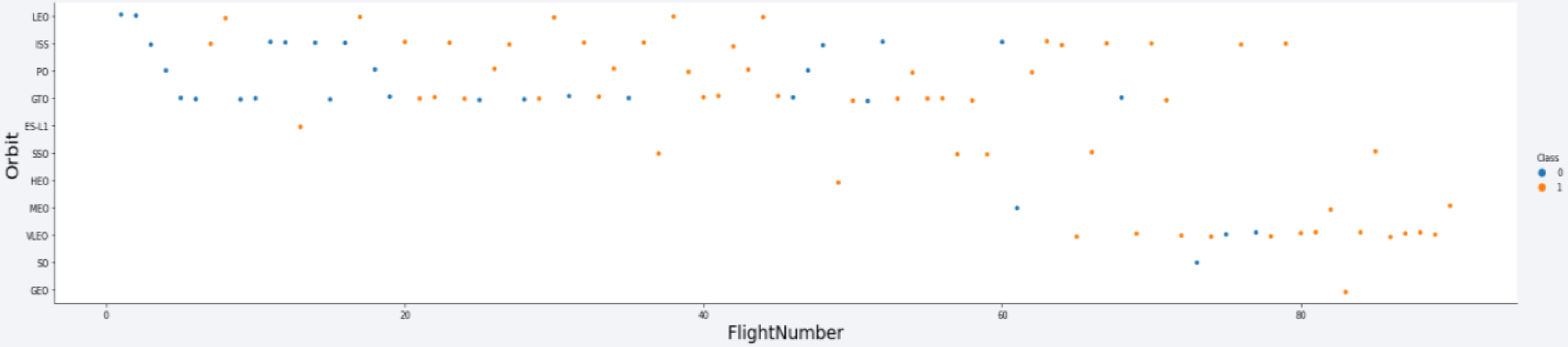
- The biggest success rates happens to orbits:

ES-L1; GEO; HEO; SSO.

- Followed by:

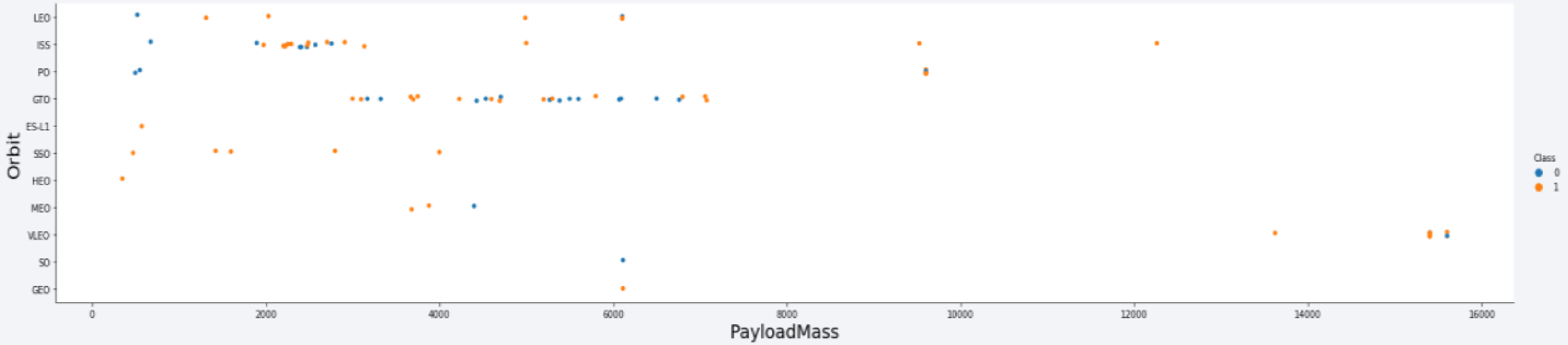
VLEO (above 80%); and LFO (above 70%)

Flight Number vs. Orbit Type



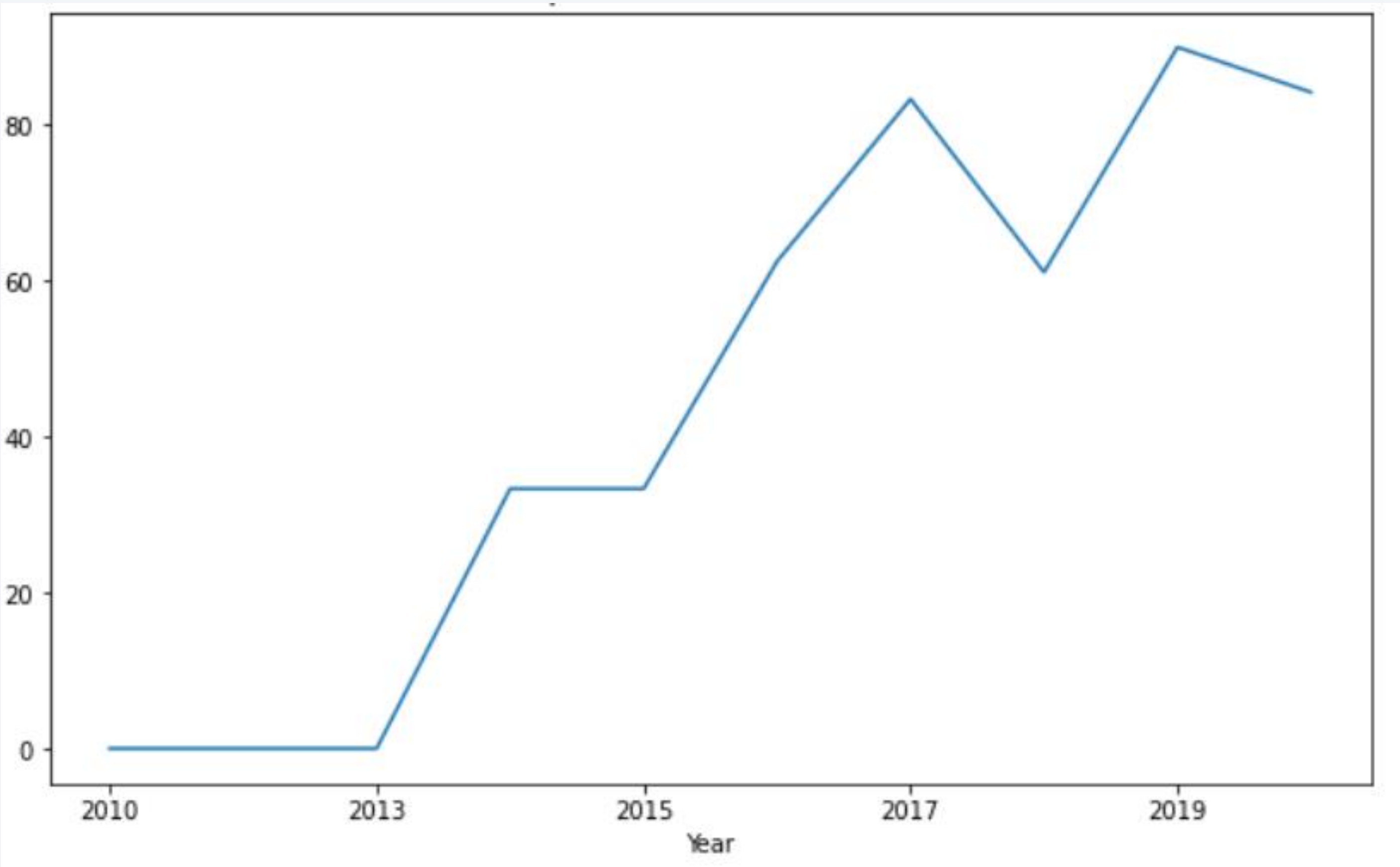
From the chart, there is no clear relationship between filght number and orbit

Payload vs. Orbit Type



Based on the chart, there is no clear relationship between payload and successful rate

Launch Success Yearly Trend



The successful rate start to increase from 2013, and keep the trend till 2020

All Launch Site Names

launch site
KSC LC-39A
CCAFS LC-40
CCAFS SLC-40
VAFB SLC-4E

Select unique launch site values from the data

Launch Site Names Begin with 'CCA'

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcom
4	1	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None
5	2	2012-05-22	Falcon 9	525.0	LEO	CCSFS SLC 40	None None
6	3	2013-03-01	Falcon 9	677.0	ISS	CCSFS SLC 40	None None
7	4	2013-09-29	Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean
8	5	2013-12-03	Falcon 9	3170.0	GTO	CCSFS SLC 40	None None

Average Payload Mass by F9 v1.1

Average payload mass

45596

Filter booster version and calculate the average payload mass

First Successful Ground Landing Date

first successful date
2015/12/22

Filter successful data and the minimum landing outcome is 2015/12/22

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

Boosters that successfully landing on the drone ship and have payload above 4000 but less than 6000.

After filters, there’re 4.

Total Number of Successful and Failure Mission Outcomes

Successful mission outcome

Success outcome	
0	100

Failure mission outcome

Failure outcome	
0	1

Boosters Carried Maximum Payload

Boosters that have the highest payload mass

Booster version	Payload mass
F9 B5 B1048.4	15600
F9 B5 B1048.5	15600
F9 B5 B1049.4	15600
F9 B5 B1049.5	15600
F9 B5 B1049.7	15600
F9 B5 B1051.3	15600
F9 B5 B1051.4	15600
F9 B5 B1051.6	15600
F9 B5 B1056.4	15600
F9 B5 B1058.3	15600
F9 B5 B1060.2	15600
F9 B5 B1060.3	15600

2015 Launch Records

There are only two 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

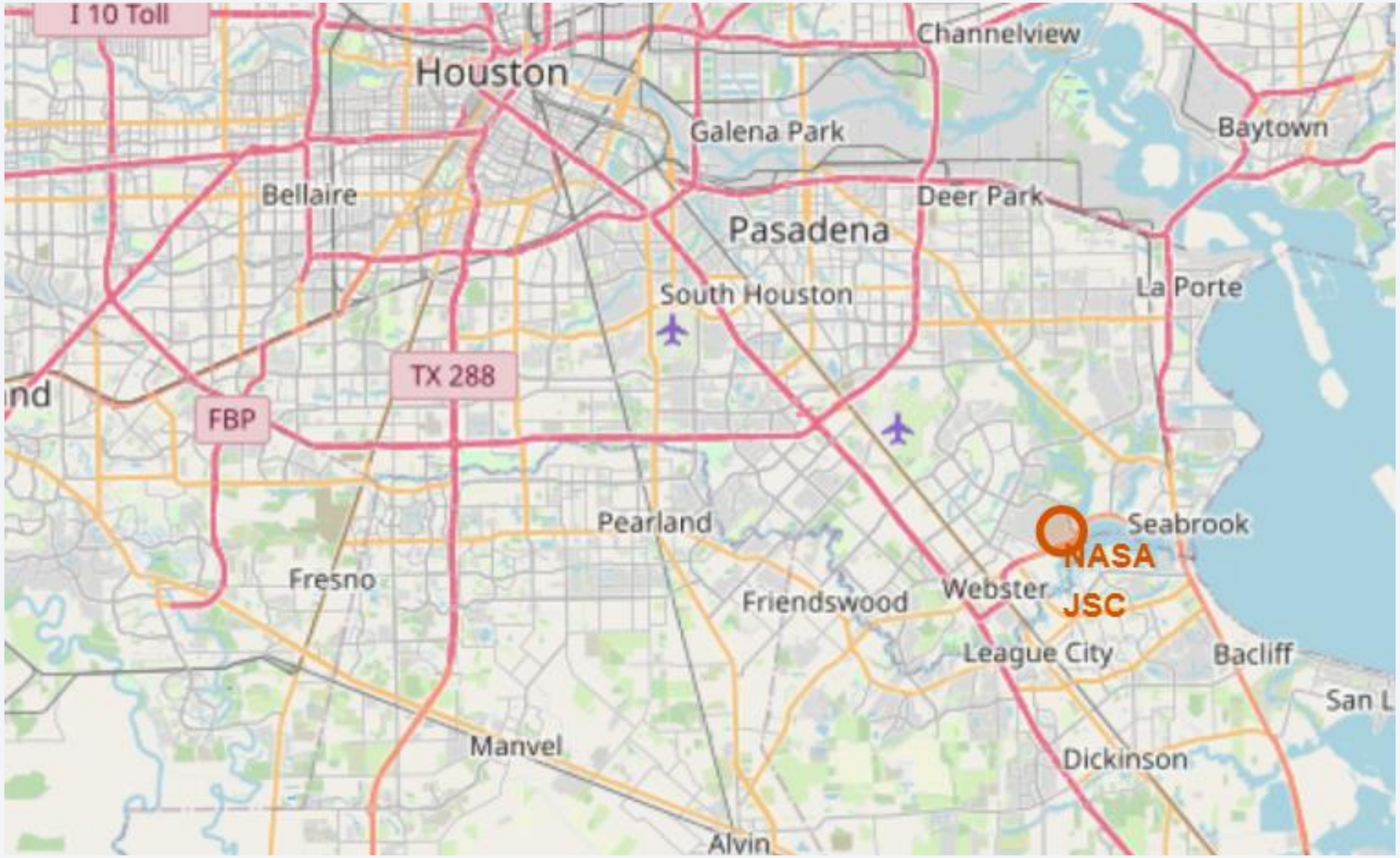
landing outcome	count
No attempt	10
Success (drone ship)	6
Failure (drone ship)	5
Success (ground pad)	5
Controlled (ocean)	3
Uncontrolled (ocean)	2
Precluded (drone ship)	1
Failure (parachute)	1

Must note there're "No attempt"

Section 3

Launch Sites Proximities Analysis

All launch site

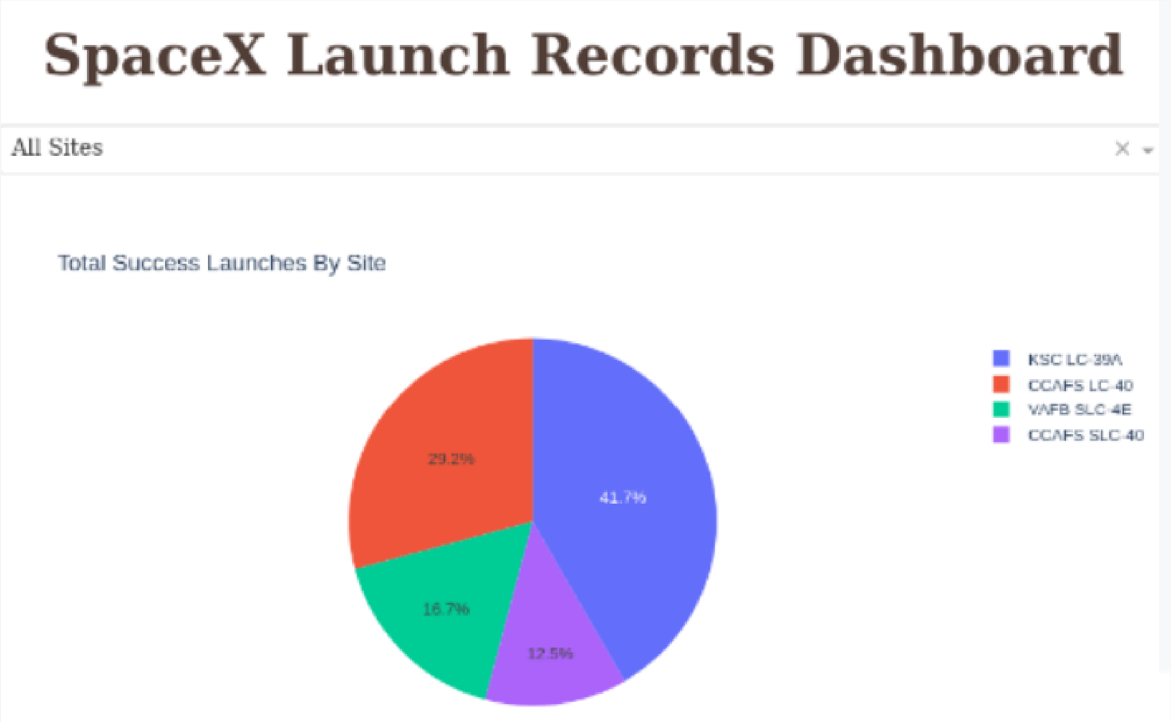


Launch outcome

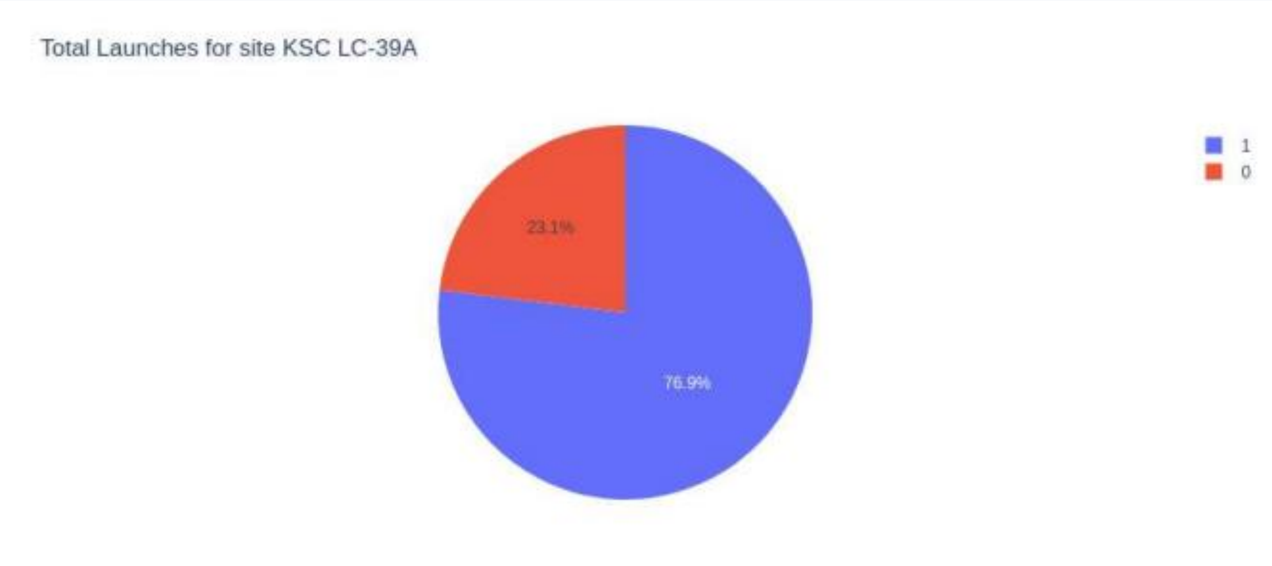
Section 4

Build a Dashboard with Plotly Dash

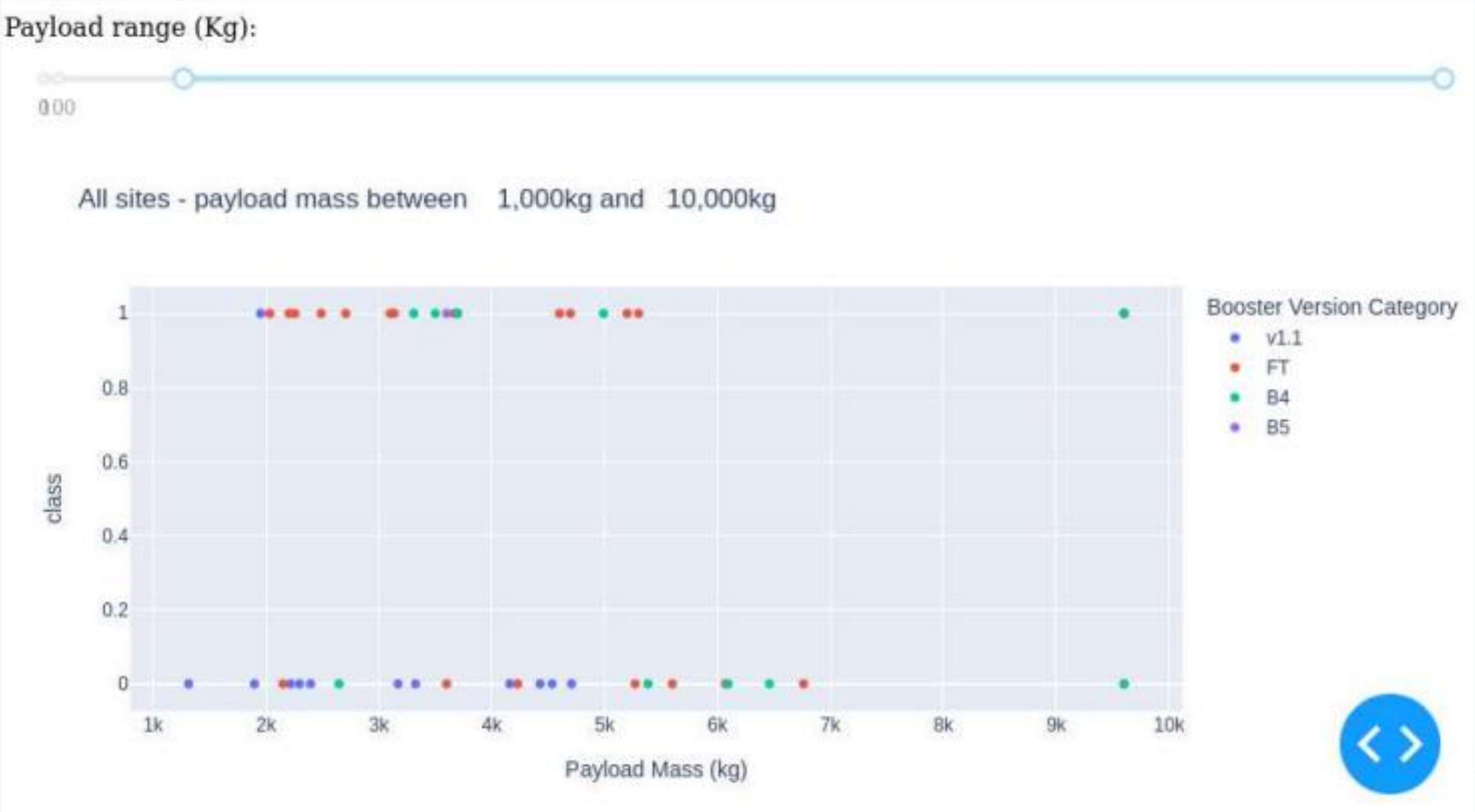
Launch success – dashboard



Launch success rate – dashboard 2



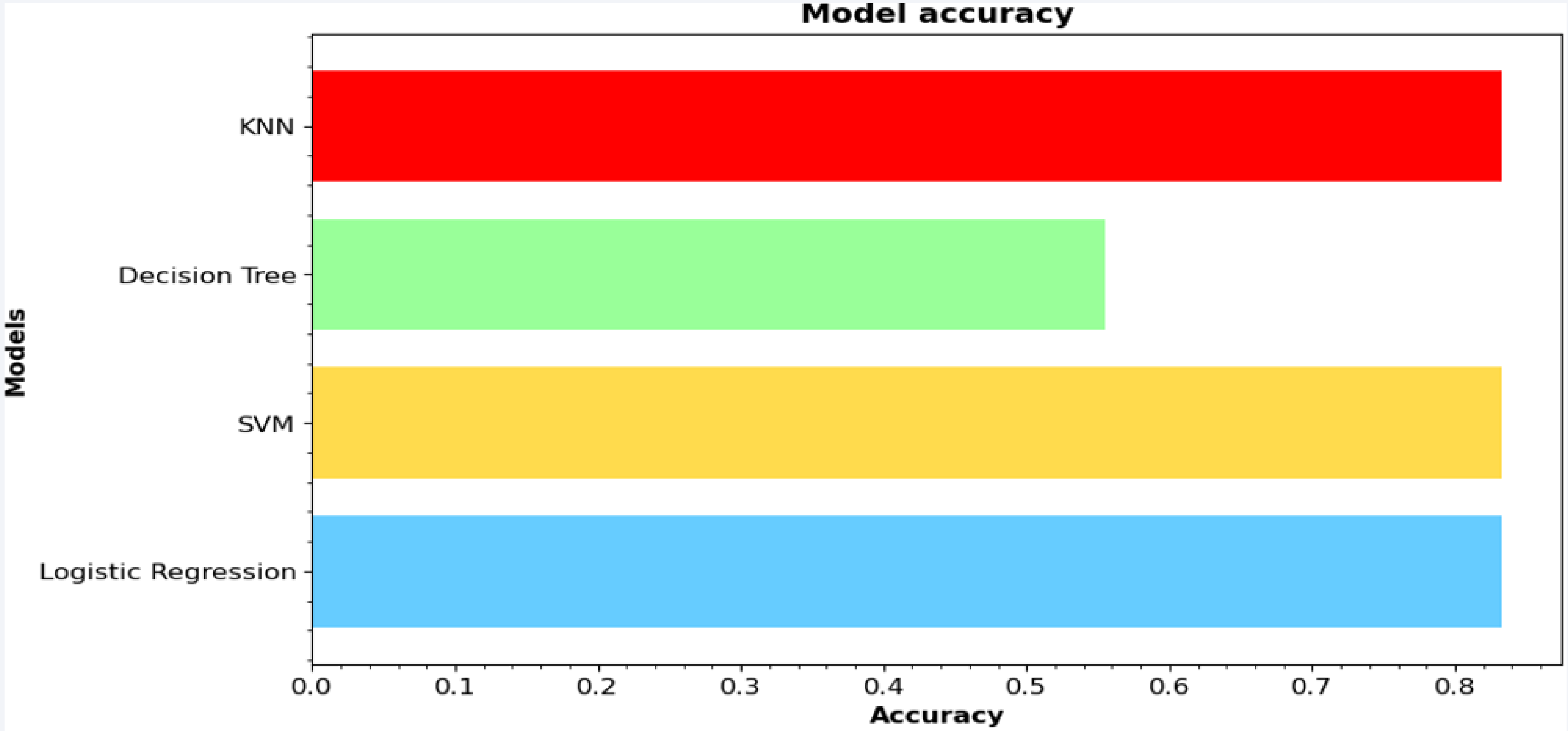
Success rate vs payload



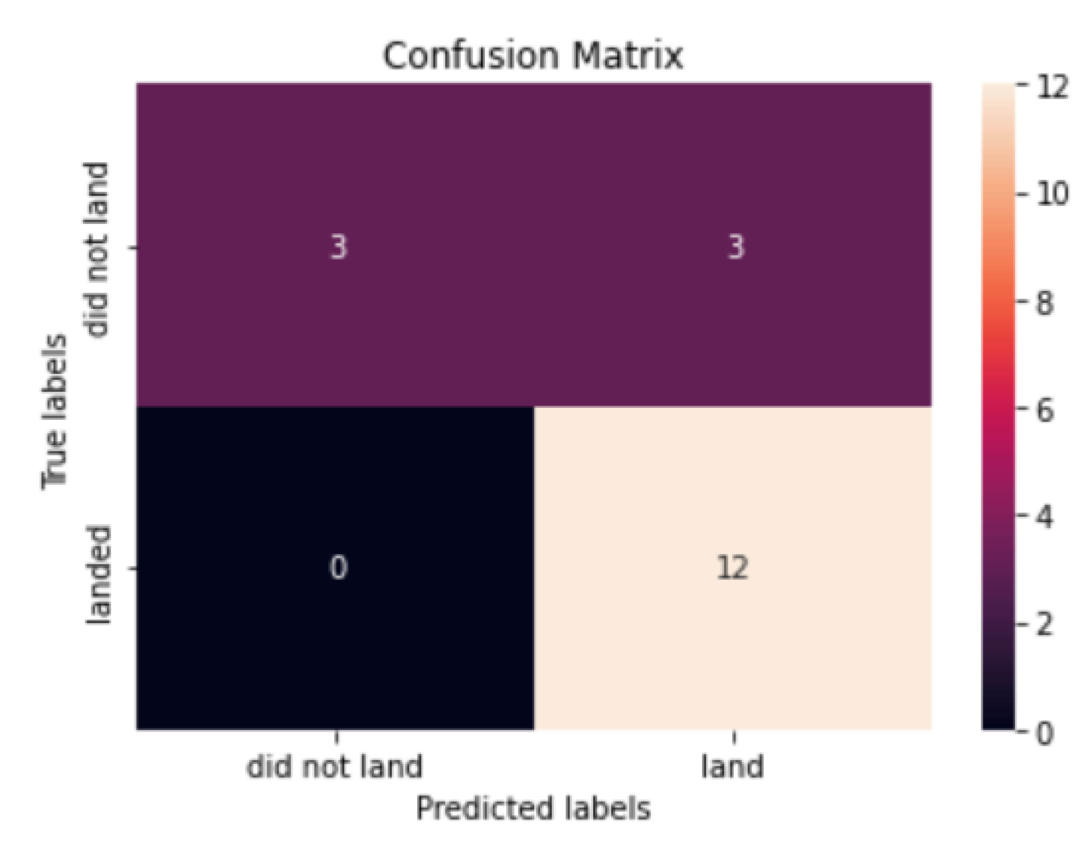
Section 5

Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix



Conclusions

- KSC LC-39A is the best launch
- Orbit GEO, HEO, SSO, ES-L1 have high success rate
- Low weight payloads perform better
- Success rate grow up with time going by
- Different prediction model drives to basically same precision

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!