Data Science Journey of SpaceX

Zixuan Yi
04.02.2024

Executive Summary

- Methodologies
  - Data Collection with web scraping and API
  - Data wrangling
  - Data Analysis
  - Data visualization
  - Machine Learning prediction
- Summary
  - We collect the dataset from the open source dataset and analyze it insightful
  - We applied several machine learning methods on dataset to see the relationship among features

Introduction

We aim to learn how SpaceX be successful on both launch successfully and budget control

- If we can determine if the first stage will land, we can determine the cost of a launch
- We can use this information if an alternate company wants to big against SpaceX for a rocket launch

Data Collection

- We collect data from Web scraping and SpaceX API
    - Web scraping (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches)
    - SpaceX API ((https://api.spacexdata.com/v4/rockets/))

Data Wrangling

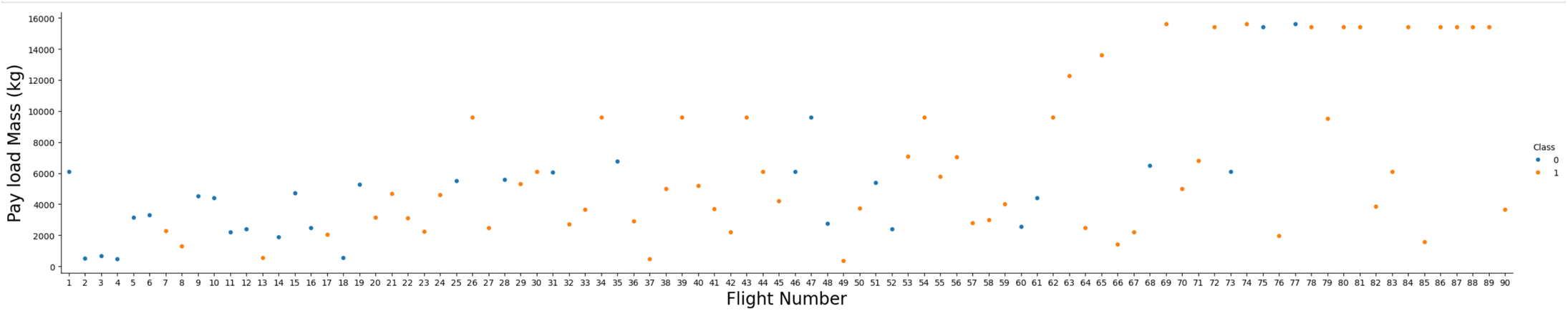
- After analyzing features, we create labels for original data

EDA with SQL

- Limit, order condition to limit the output of data and Rank the dataset in specific order
- The function as count(), sum() etc to calculate columns value
- The window operator as groupby to analyze the data between categories or groups

EDA with Data visualization

- To deep in the data, we use scatter plots, barplot and line plot to display the relationship between features
- Example below is the pay load Mass and flight number relationship for different classes.



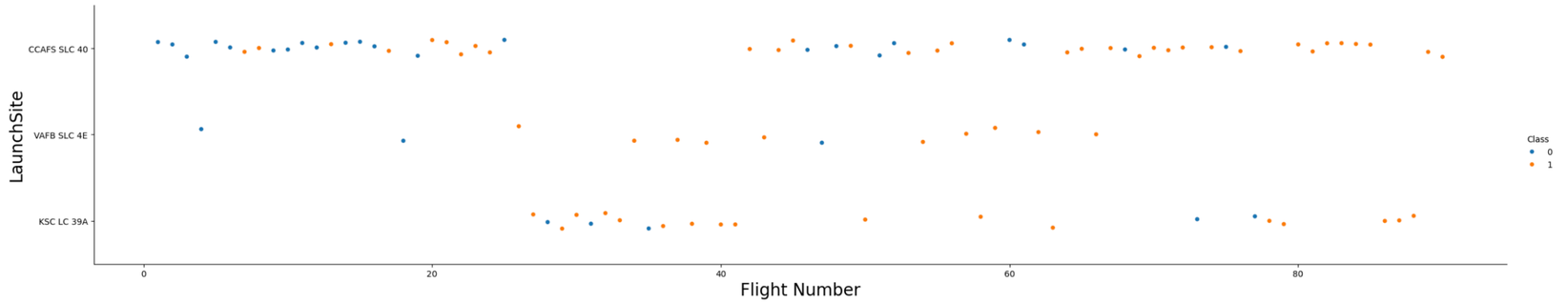
Interactive visual analytics

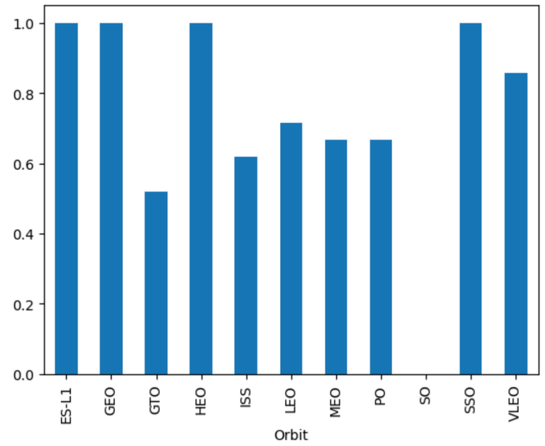
- Create Interactive visual with folium package
- Use Markers to mark launch sites
- Marker cluster can mark multi marks on the map to show the different event
- Lines represent the distance between two destinations
- Using Plotly Dash to build a Dashboard
- Callback decorator to automatically aligned with changes

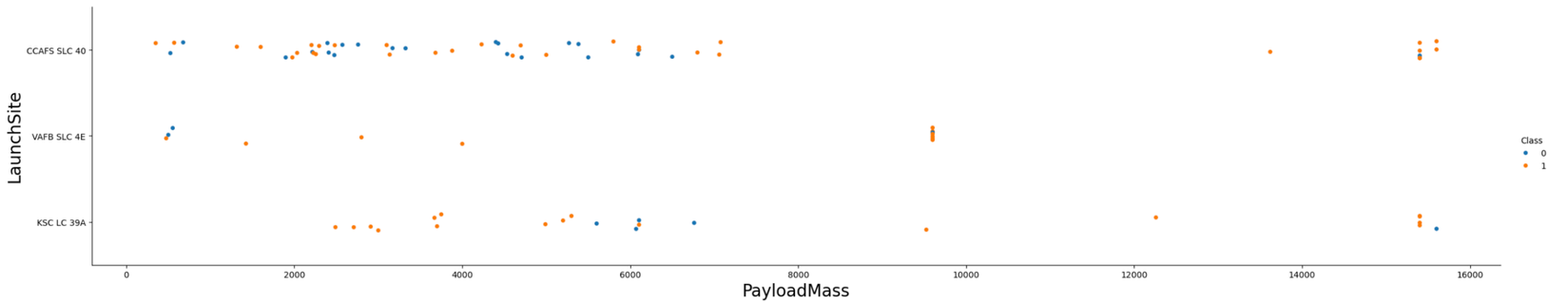
Predictive Analysis

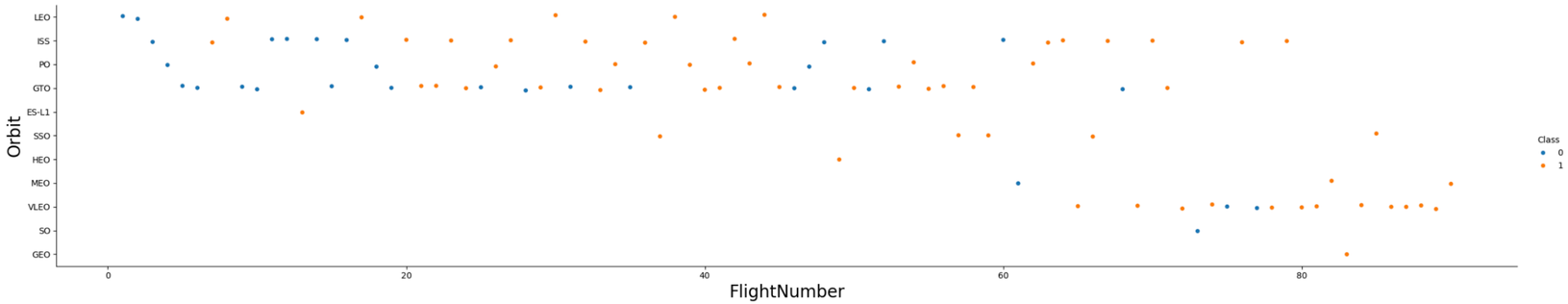
- Using logistic regression and svm to do classification
- Search the parameters show the best scores
- Analyze the result with confusion matrix

EDA visualization result

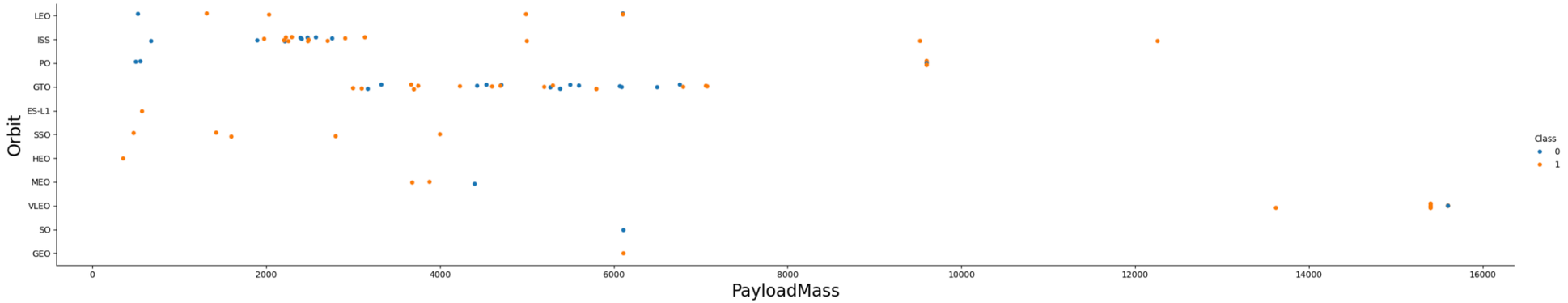




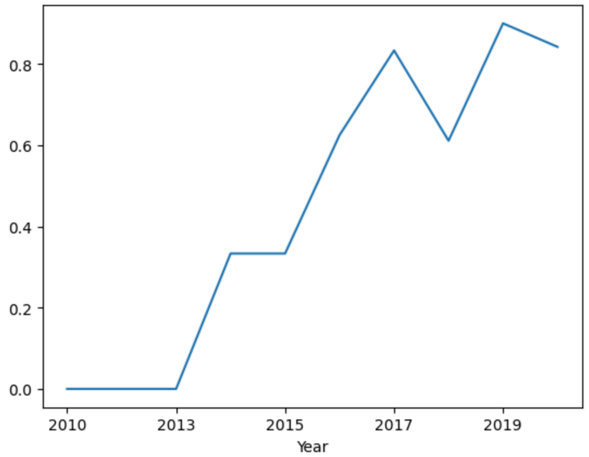




EDA visualization result



[13]: <AxesSubplot:xlabel='Year'>



EDA SQL Result

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

[10]:

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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	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
111268

AVG_PAYLOAD
2928.4

first_success_gp
2015-12-22

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

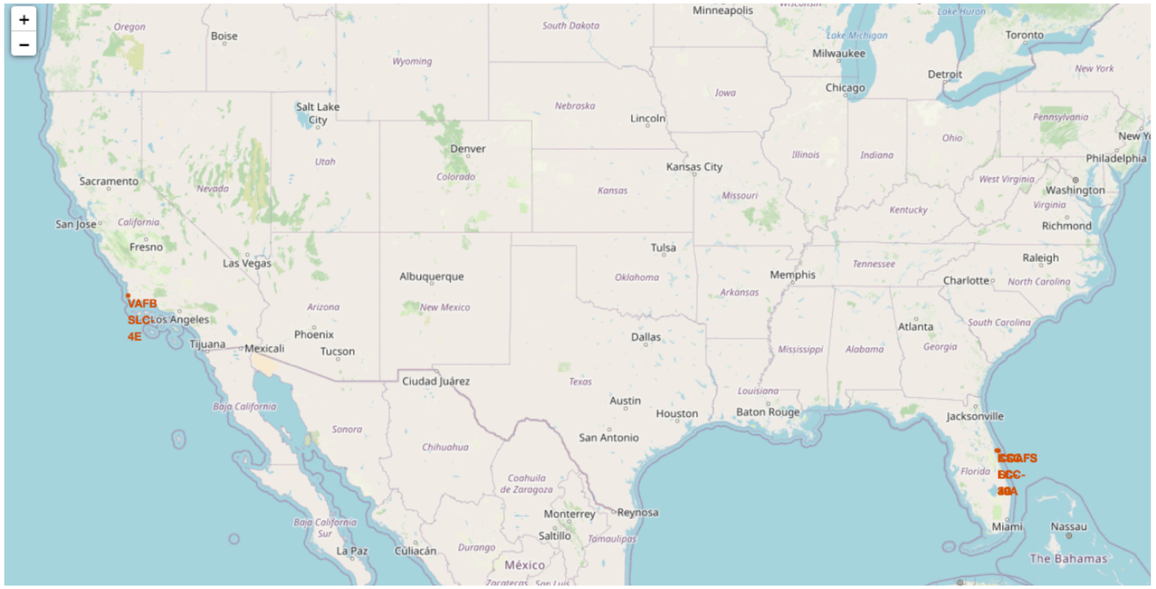
Mission_Outcome	QTY
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

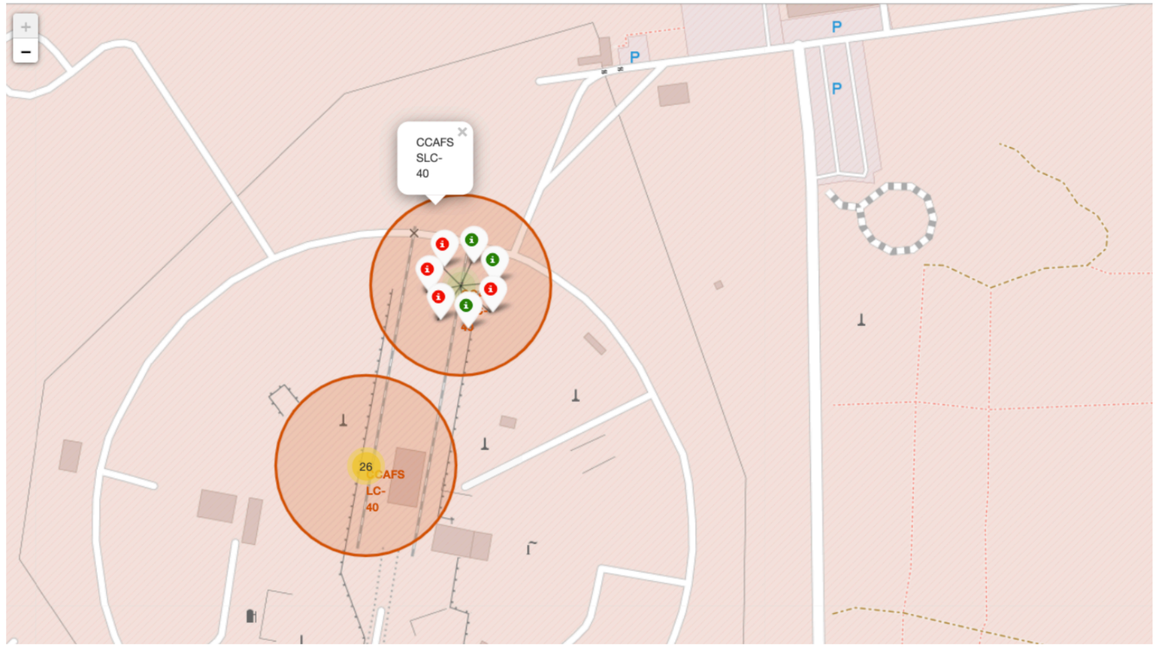
Booster_Version
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
F9 B5 B1051.3
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

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

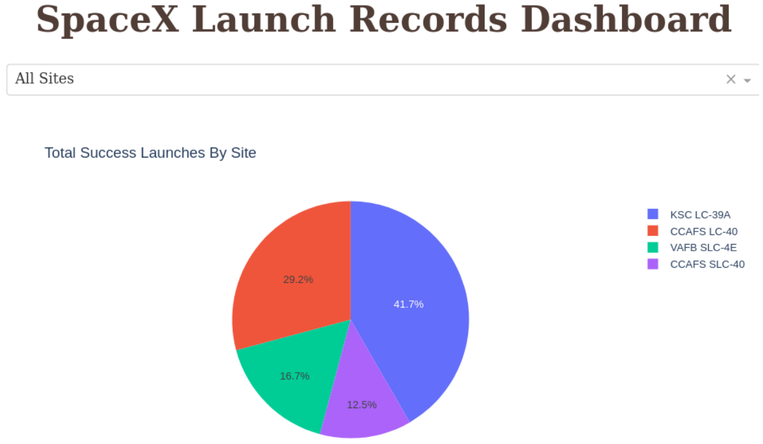
landing__outcome	qty
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

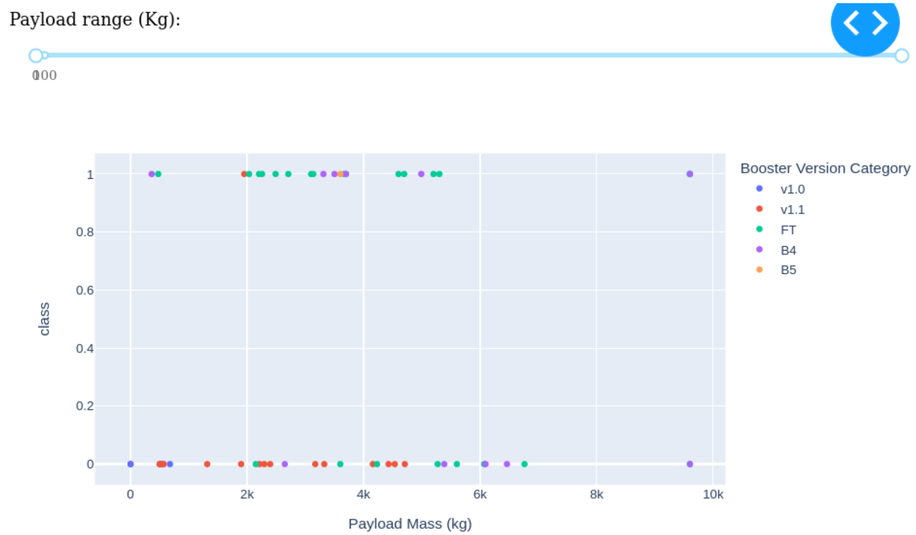
Interactive map result

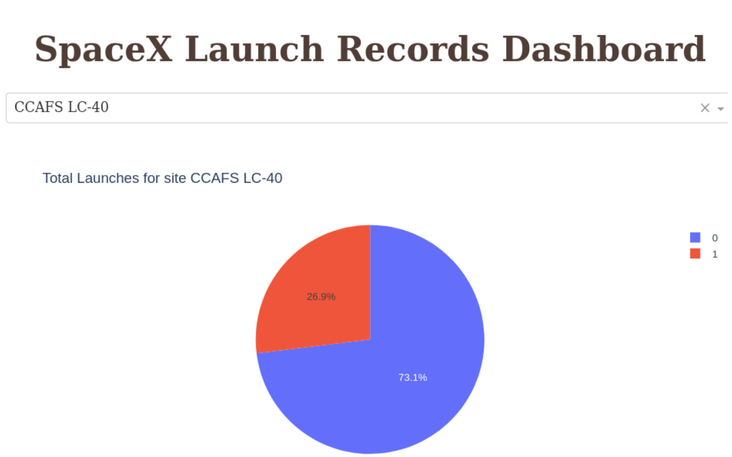


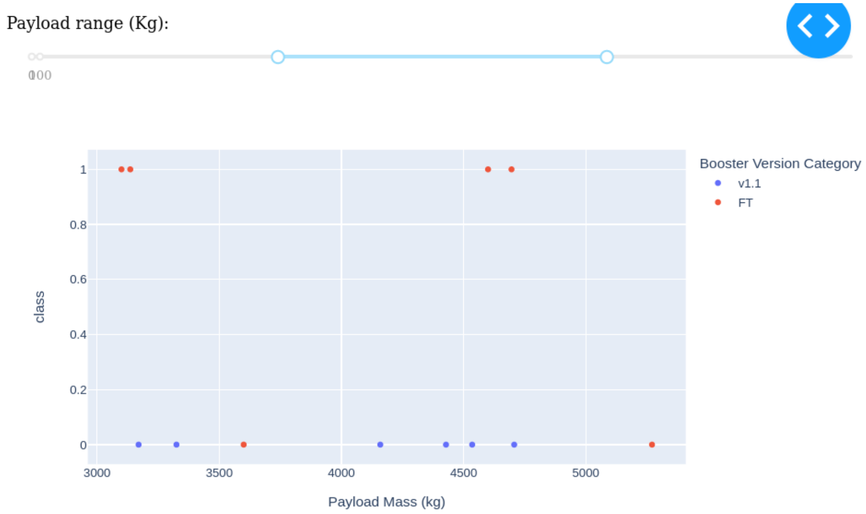


Result with interactive dashboard









Result with predictive analysis