



<Capstone-SpaceX>

<JIAN KUN MAO>

<MAY 13TH 2022>

PROJECT_OUTLINE



- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Quantum Analysis(Process)



- Data collection
- Data wrangling
- EDA with data visualization
- EDA with SQL
- Building an interactive map with Folium
- Building a Dashboard with Plotly Dash
- Predictive analysis BY machine learning

Results



- Data analysis result
- Predictive result

INTRODUCTION



- Background

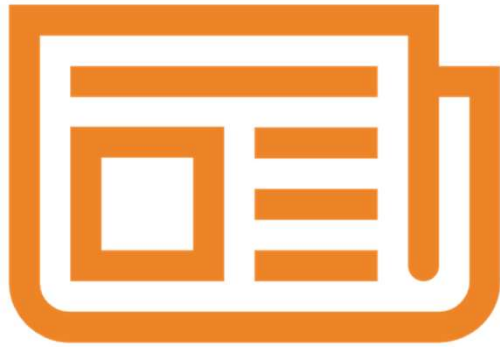
Base on the information that post on the Wikipedia, tony of the info would be found including cost of the Falcon, success Rate etc. Which help us to find the difference between Space X and other project(company).

- Question need to solve

Influence of the Success Rate

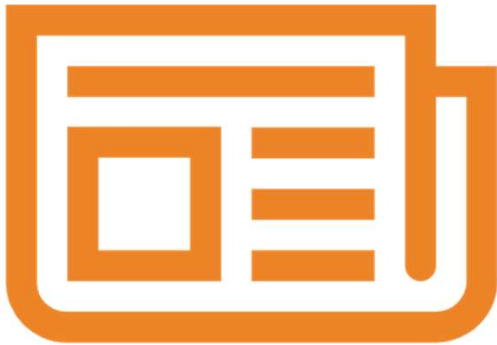
What is the advantage make of the SpaceX have highest Success Rate with lower cost.

METHODOLOGY



- Data collection & wrangling
 - data mining(beautiful Soup)
 - data transfer to DF
 - cleaning(organize and keep the variable data)
- Visualization
 - SQL(query the data w/ condition)
 - Plot(Show the relationship between variables)
 - Folium and Dash
- Classification model

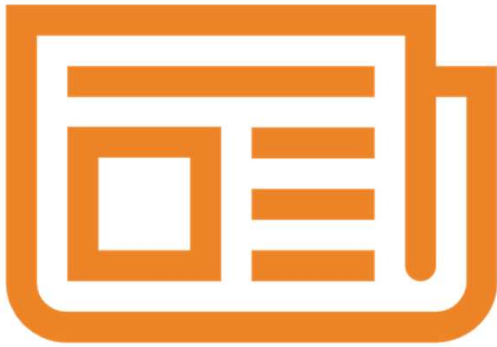
Data collection



- 1 get response from API
- 2 Covert response to j.son
- 3 cleaning the data
- 4 transfer the data to DF format
- 5 export to CSV file

[Link to github](#)

Data Scraping & Wrangling

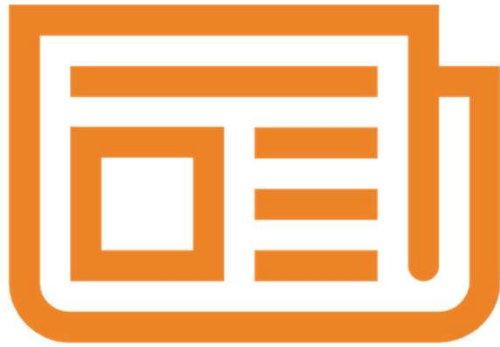


Link to github for
scaping

Link to github for
Wrangling

- 1 get response from API
- 2 create BeautifulSoup Object
- 3 find the Target table
- 4 extract the data from table
- 5 get the variable data
- 6 export to CSV file

Data visualization



- 1 Plot
- 2 SQL
- 3 Interactive map

[Link to github](#)

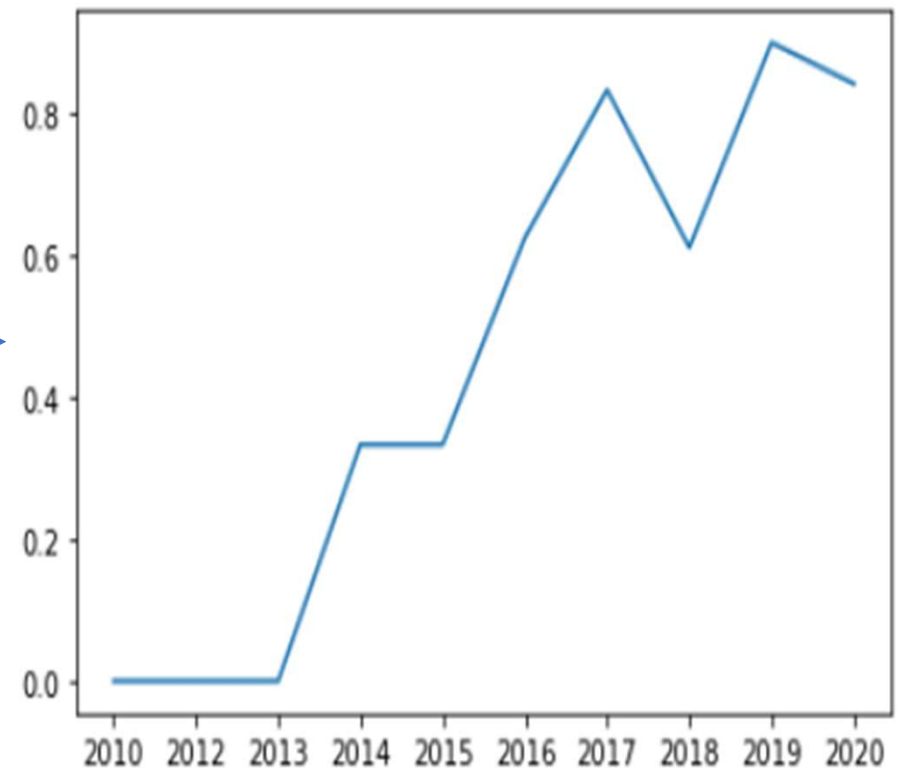
[Link to github](#)

[Link to github](#)

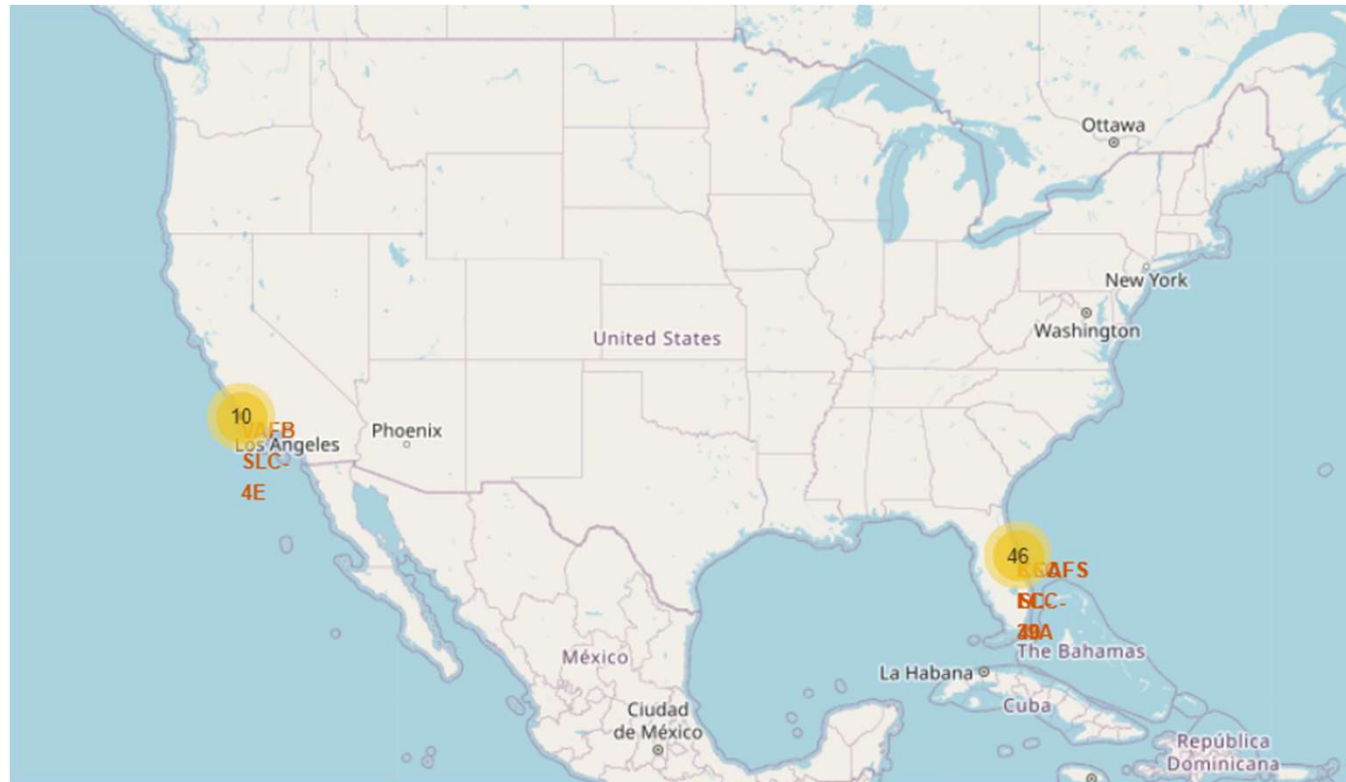
Key information in Plot

	Orbit	Class
0	ES-L1	1.000000
1	GEO	1.000000
2	GTO	0.518519
3	HEO	1.000000
4	ISS	0.619048
5	LEO	0.714286
6	MEO	0.666667
7	PO	0.666667
8	SO	0.000000
9	SSO	1.000000
10	VLEO	0.857143

Success Rate of
Each Orbit
In 2010-2020

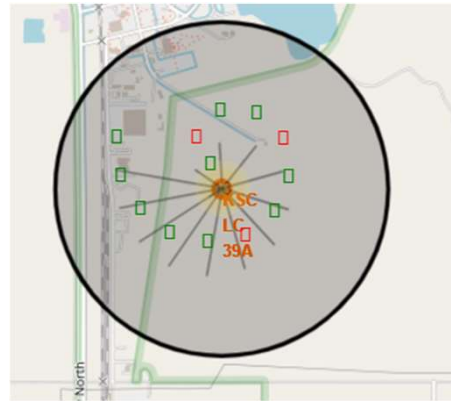
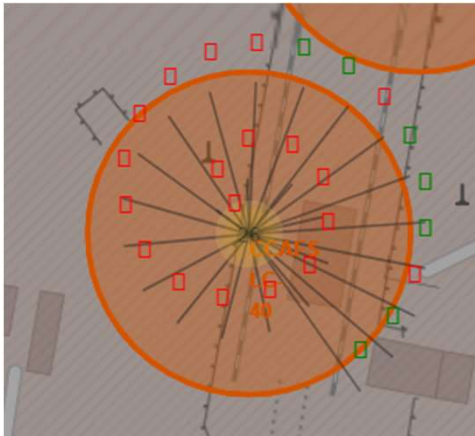


Interactive map by folium(key Information)

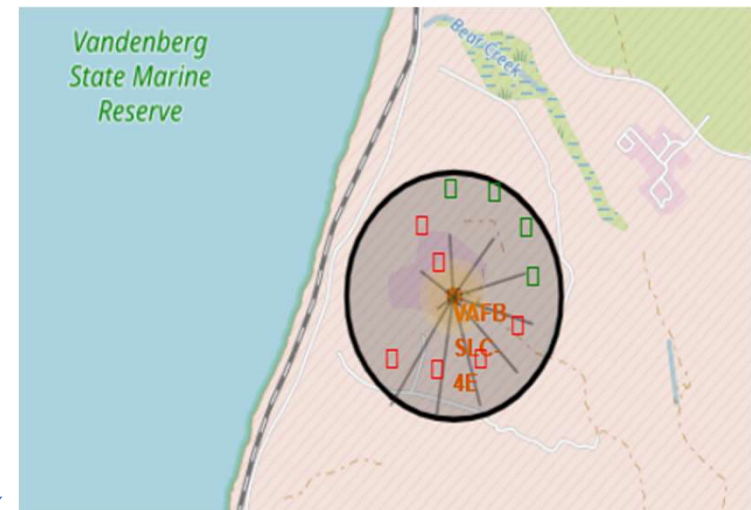


Clear to show the launch number and location
in the East and west coast of U.S.

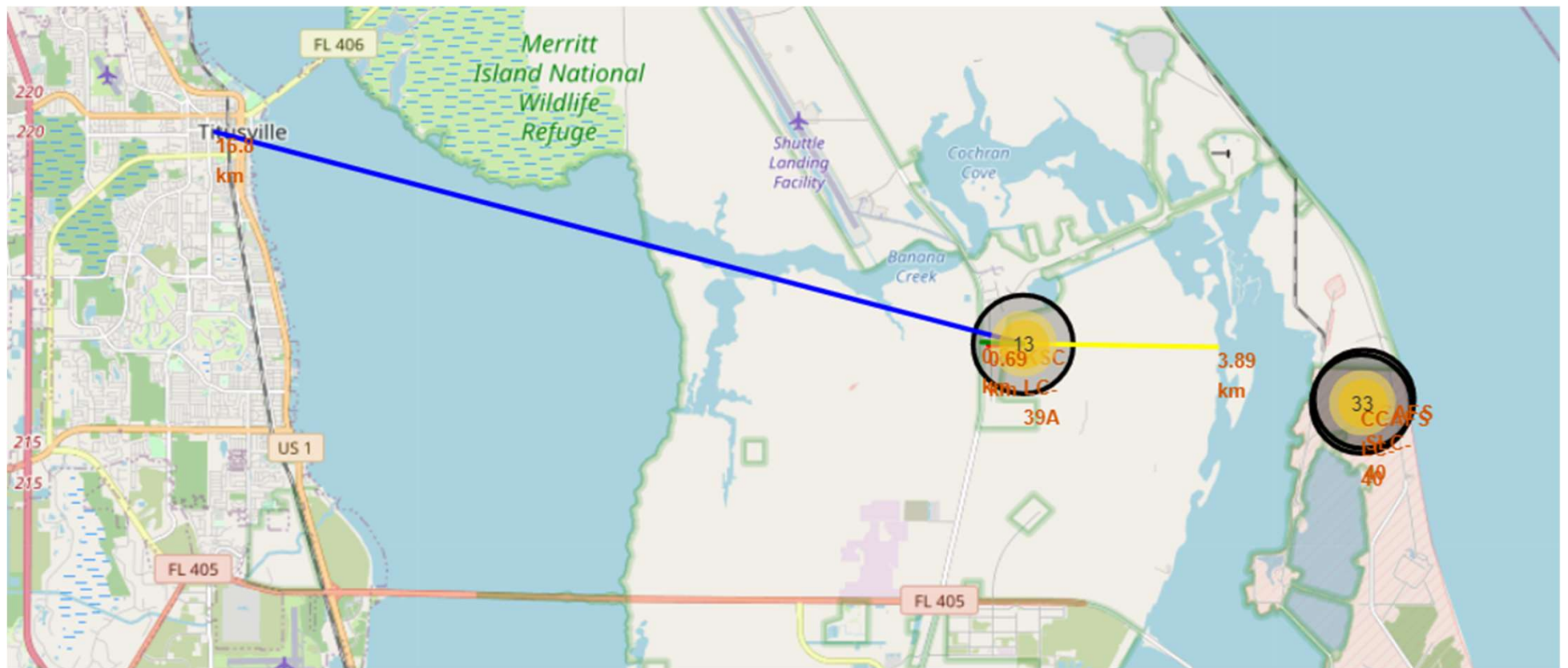
The successful launch (green) and fail(red)



EAST
WEST



Launch site Traffic Condition

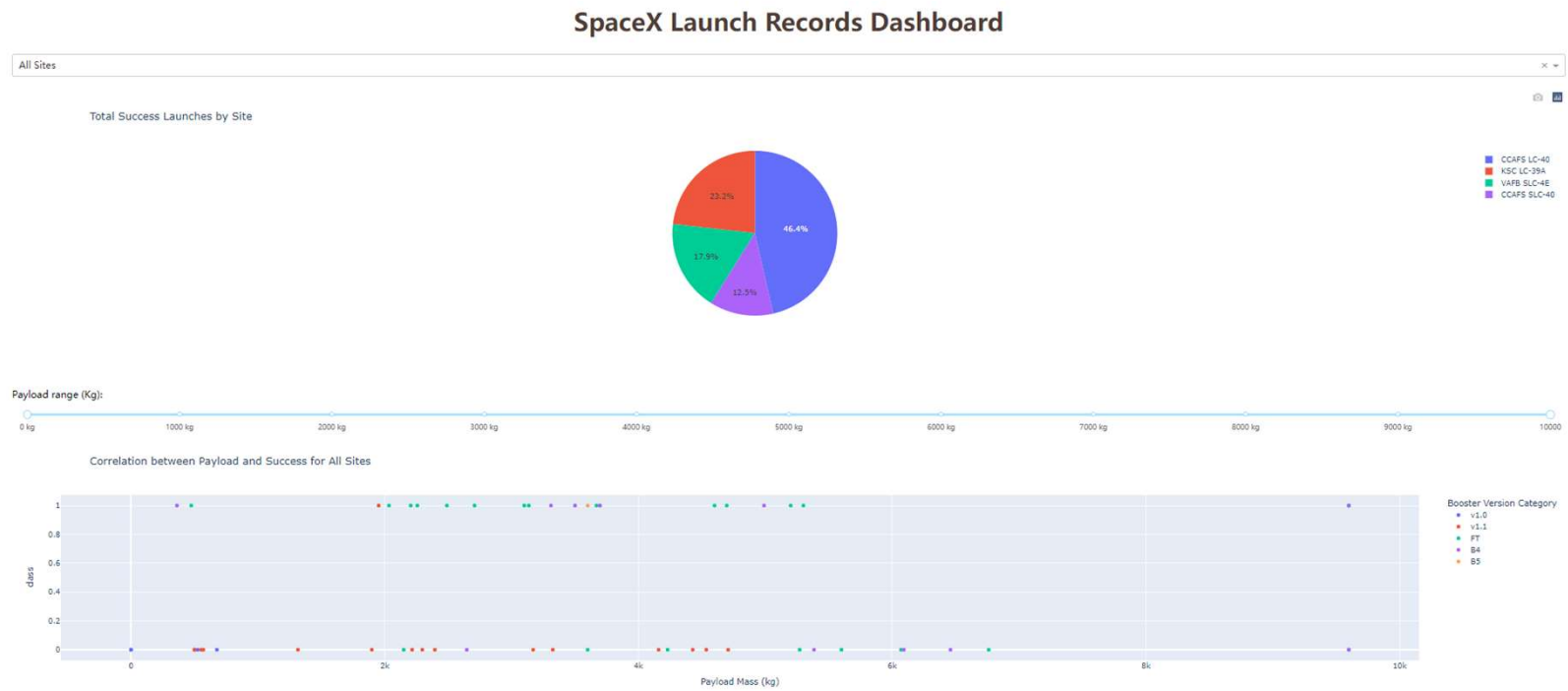


DASHBOARD



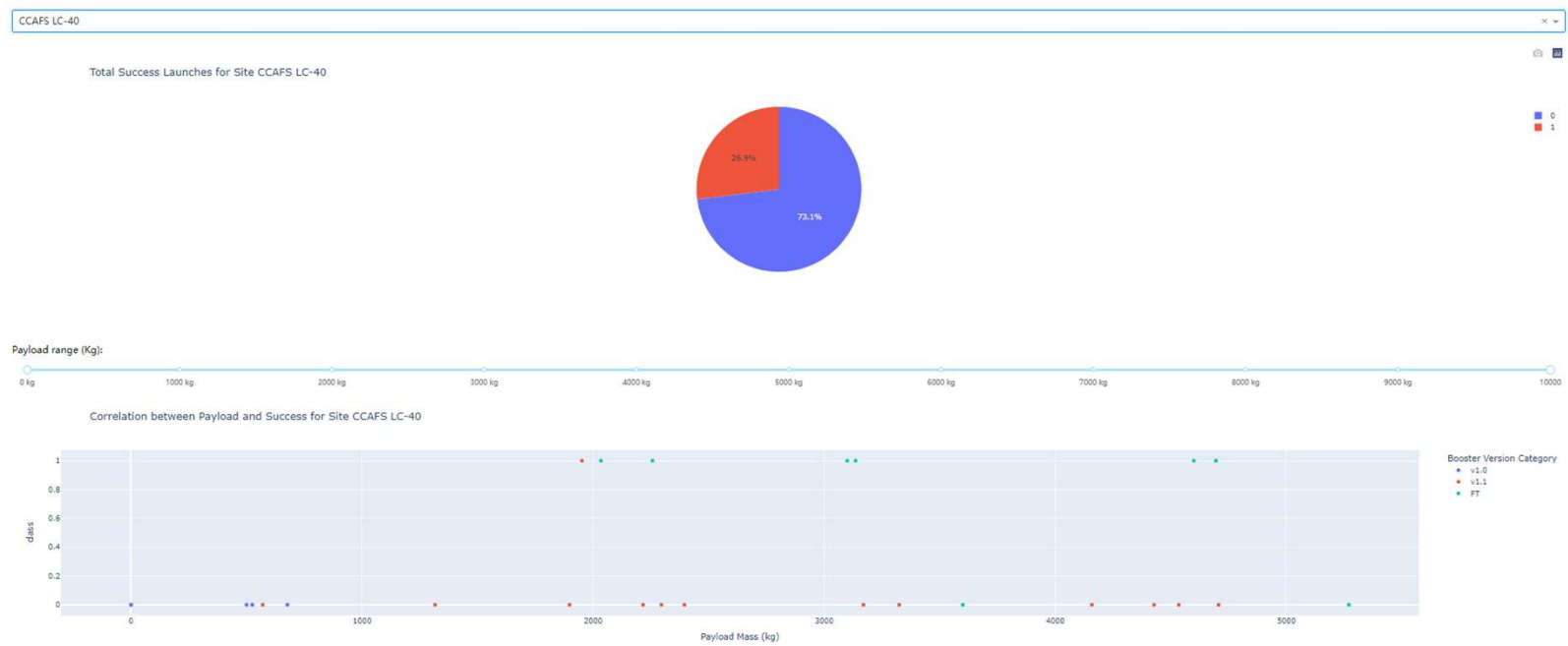
[Link to github](#)

Overall

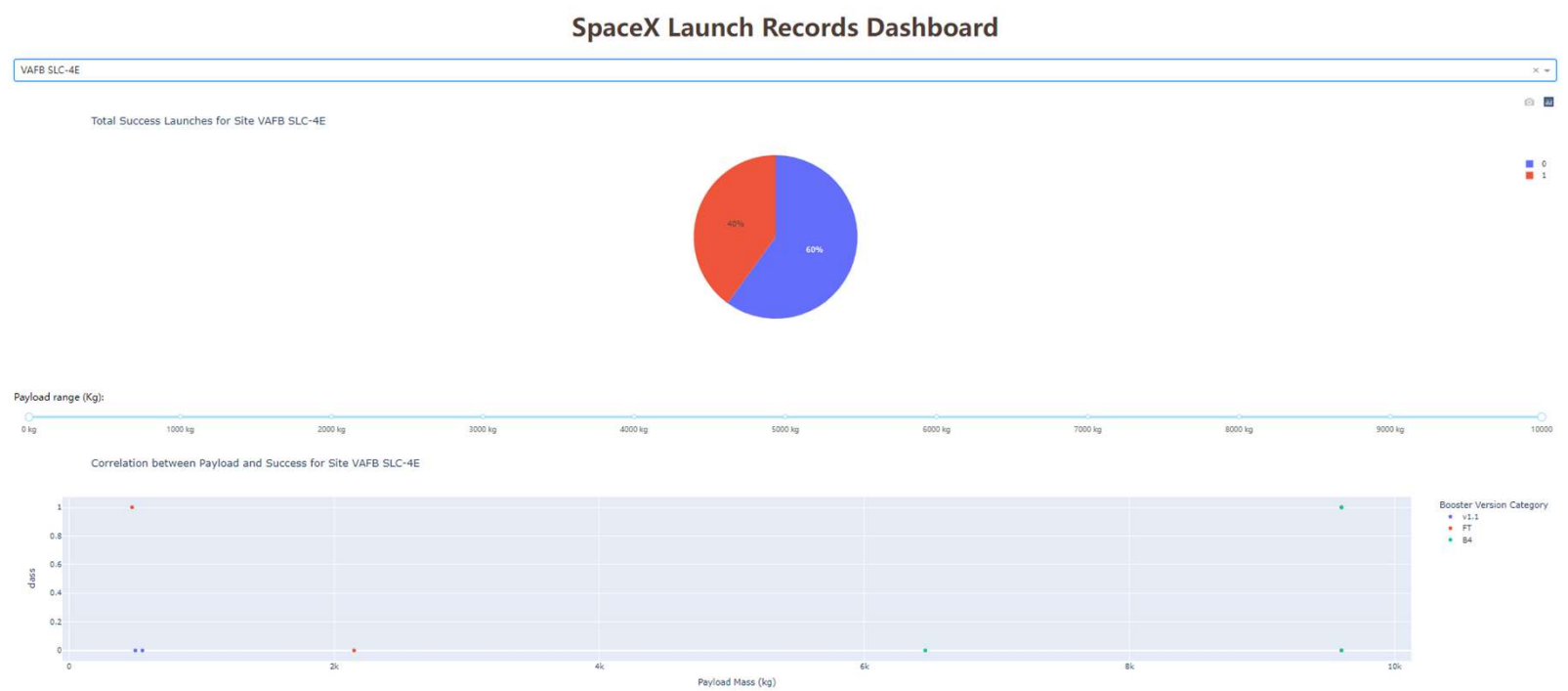


CCAFS LC-40

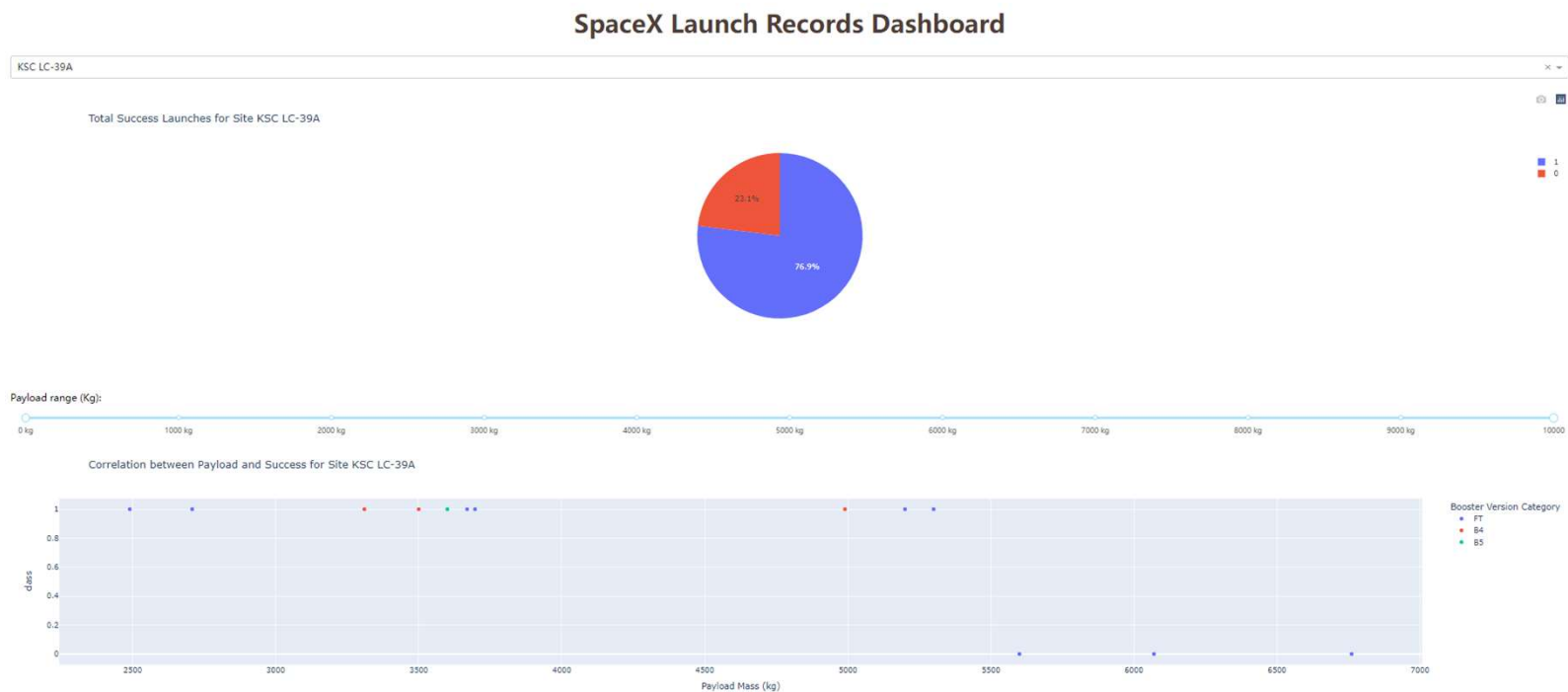
SpaceX Launch Records Dashboard



VAFB SLC-4E

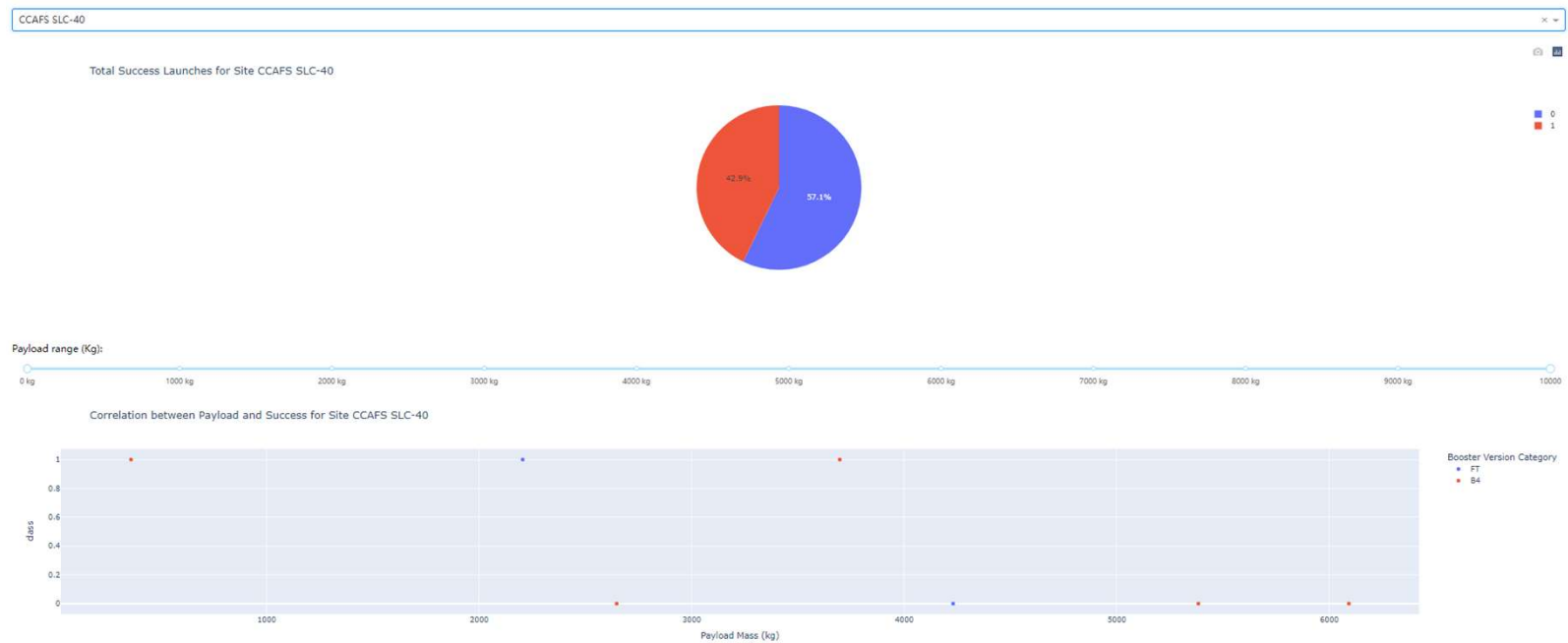


KSC LC-39A(Highest Success Rate)



CCAFS SLC-40


SpaceX Launch Records Dashboard



Classification model by machine learning

Model Choosing

Accuracy of 4 different models



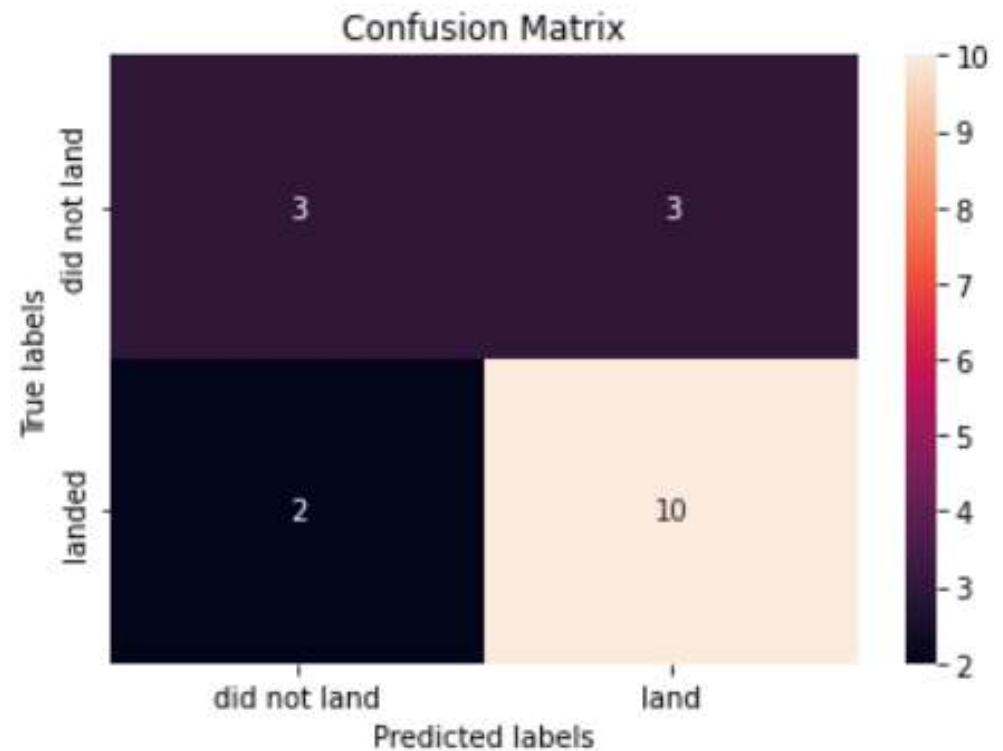
MODEL	Accuracy
LogisticRegression	0.84642857
SVN	0.84821429
TREE	0.87678571
KNN	0.84821429

Obviously Tree has the highest score in the table

Result

In here, we can see the predictive result which base on the tree algorithm, for more convience I put a guide figure to help audiences have understand the Confusion Matrix

		Predicted Values	
		Negative	Positive
Actual Values	Negative	TN	FP
	Positive	FN	TP



DISCUSSION



- In this case
- we found
 1. Orbit GEO,HEO,SSO,ES-L1 have the best Success Rate
 2. Success Rate is generally increasing in 2010-2020
 3. KSC LC-39A has the most successful launches
 4. Tree Classifier algorithm is the best predictor.