<Capstone-SpaceX>

<JIAN KUN MAO>

<MAY 13TH 2022>

Link to github

PROJECT_OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary(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

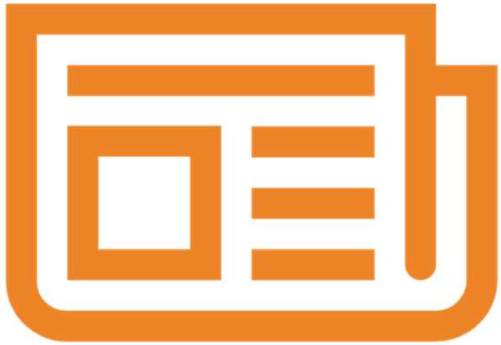
Base on the information that post on the Wikipedia, tony of the info would be found including cost of the Falcon, success Rate etc. Which help us to find the difference between Space X and other project(company).

- 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

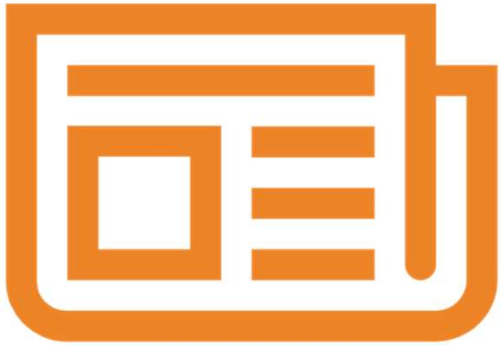
-SQL(query the data w/ condition)

-Plot(Show the relationship between variables)

-Folium and Dash

- Classification model

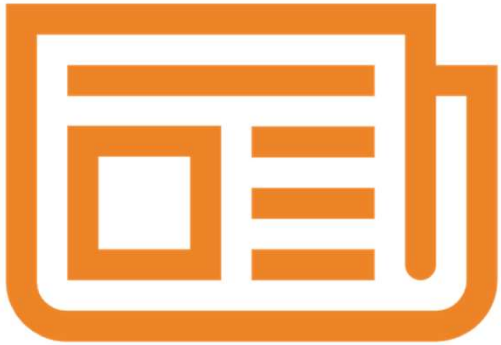
Data collection



- 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

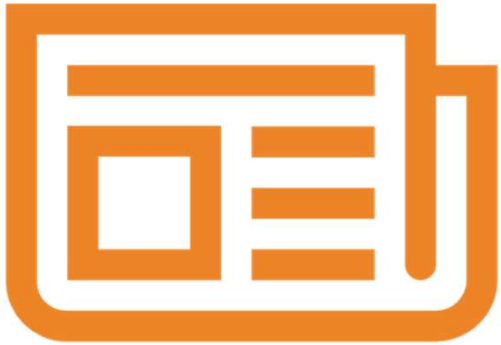


Link to github for scraping

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

SQL（key information）

In [5]:
```
%sql SELECT DISTINCT launch_site from SPACEXDATASET
```

```
 * ibm_db_sa://gbw48020:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/bludb
Done.
```

Out[5]:

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

Display average payload mass carried by booster version F9 v1.1

In [11]:
```
%sql SELECT avg(payload_mass__kg_) from SPACEXDATASET\
where booster_version ='F9 v1.1'
```

```
 * ibm_db_sa://gbw48020:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/bludb
Done.
```

Out[11]:

1
2928

List the total number of successful and failure mission outcomes

In [27]:
```
%sql SELECT count(*) from SPACEXDATASET;
```

```
 * ibm_db_sa://gbw48020:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/bludb
Done.
```

Out[27]:

1
101

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

In [61]:
```
%sql SELECT count(*) from SPACEXDATASET\
where (DATE between '2010-06-04' and '2017-03-20') and (landing__outcome like 'Failure (drone ship)')
```

```
 * ibm_db_sa://gbw48020:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/bludb
Done.
```

Out[61]:

1
5

In [62]:
```
%sql SELECT count(*) from SPACEXDATASET\
where (DATE between '2010-06-04' and '2017-03-20') and (landing__outcome like 'Success (ground pad)')
```

```
 * ibm_db_sa://gbw48020:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/bludb
Done.
```

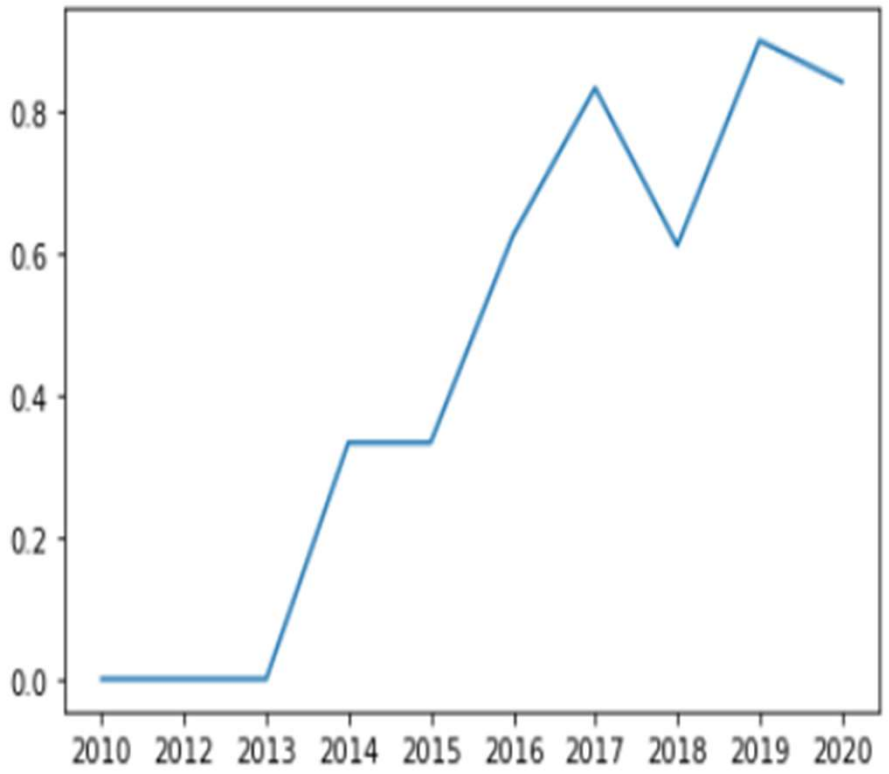
Out[62]:

1
3

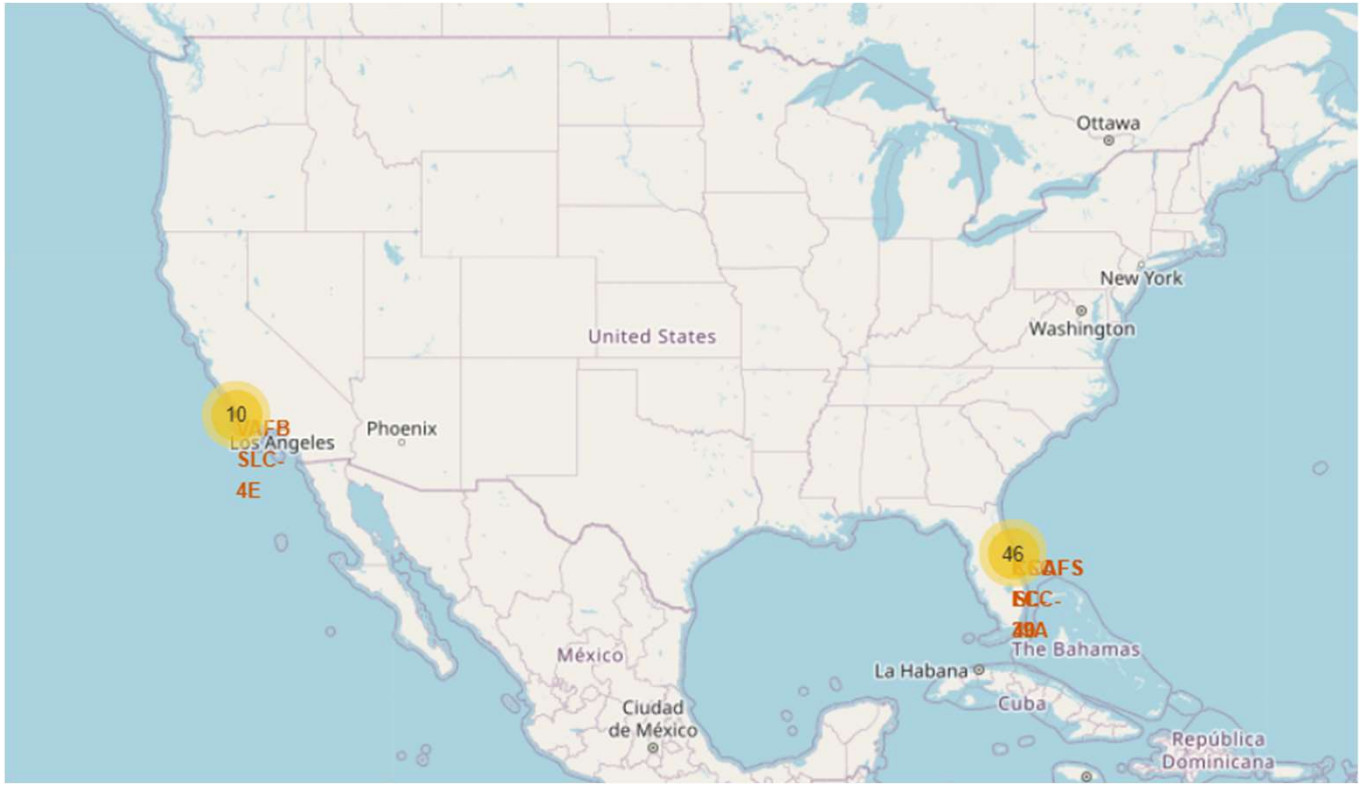
Key information in Plot

	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

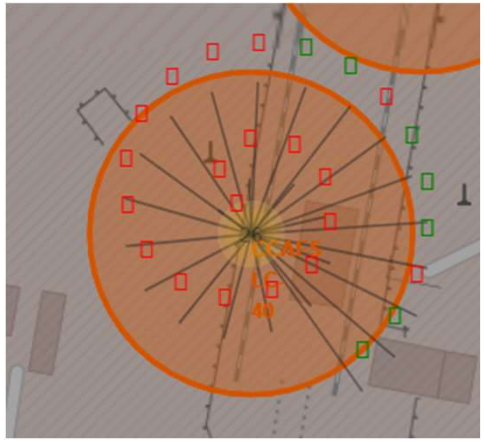


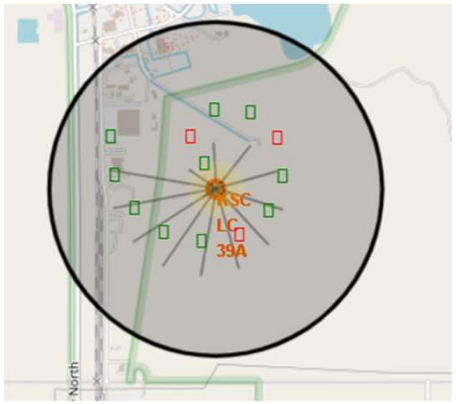
Interactive map by folium(key Information)

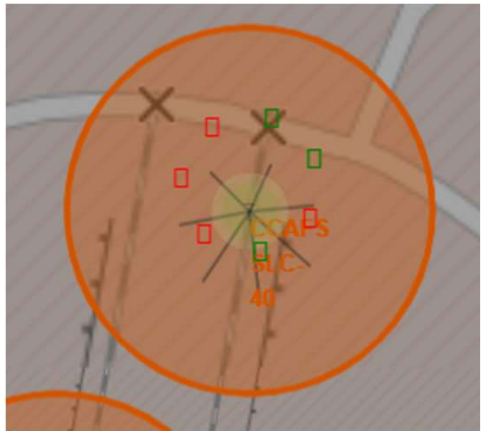


Clear to show the launch number and location
in the East and west coast of U.S.

The successful launch (green) and fail(red)

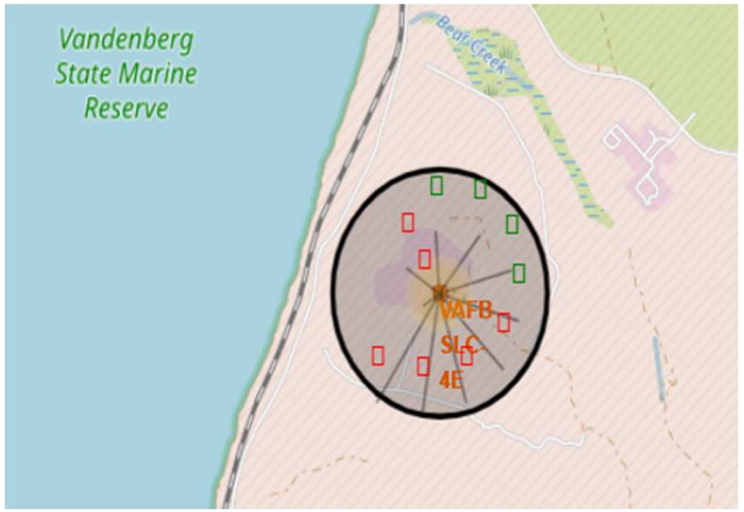




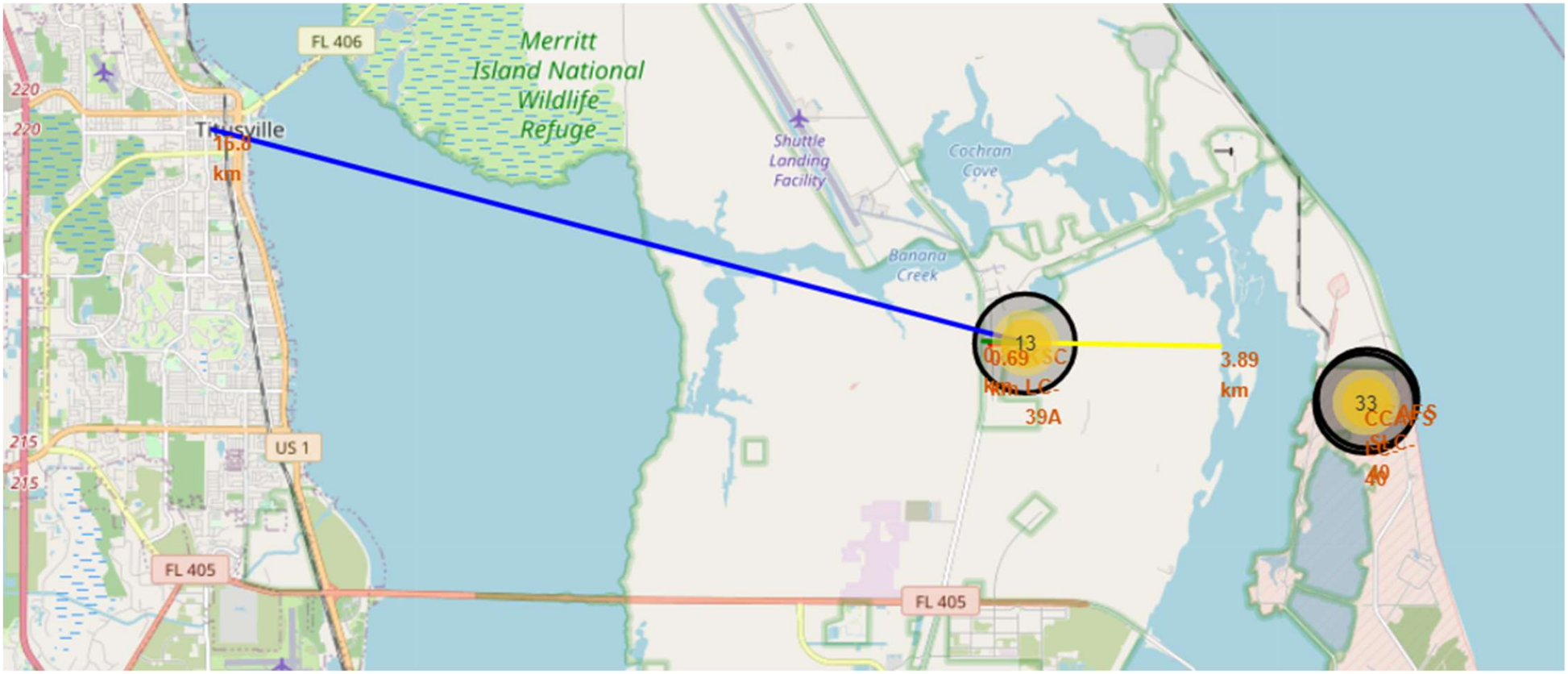


EAST

WEST

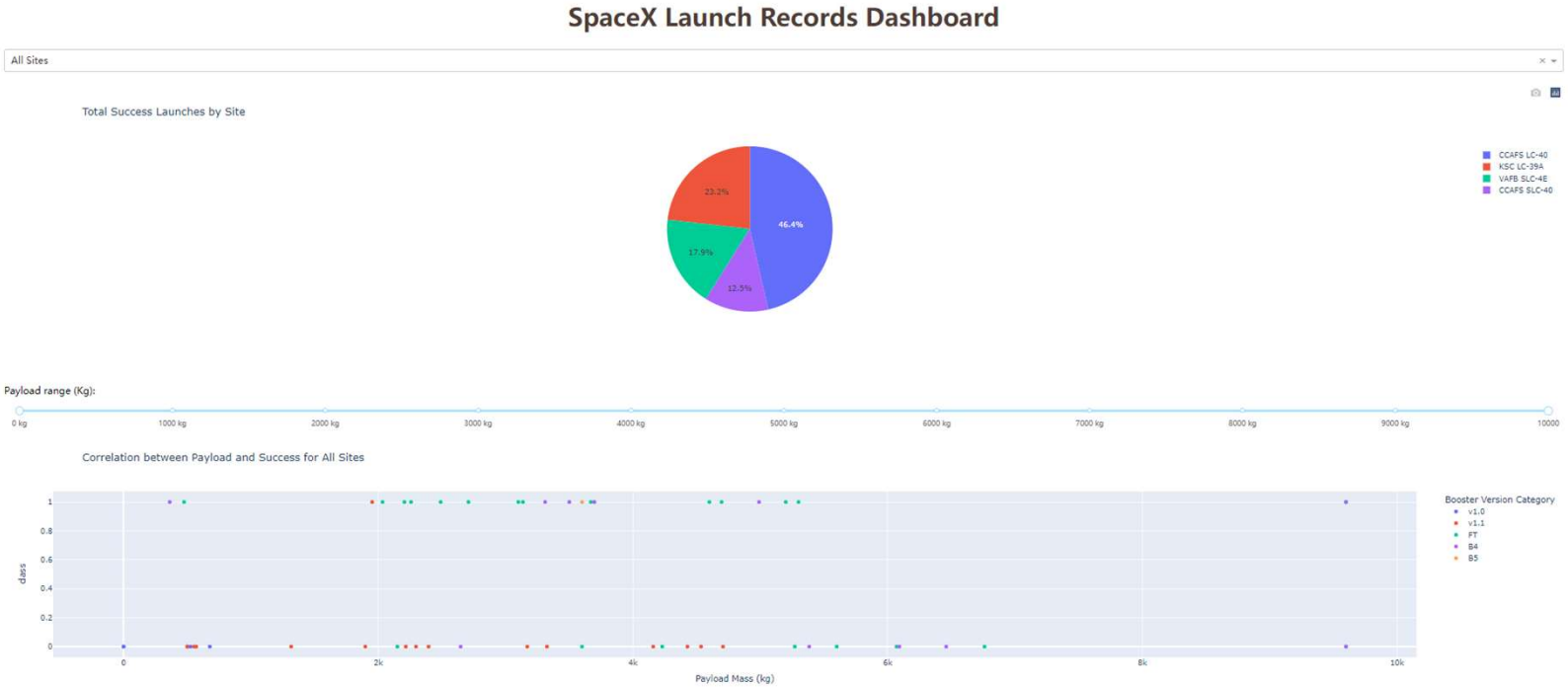


Launch site Traffic Condition

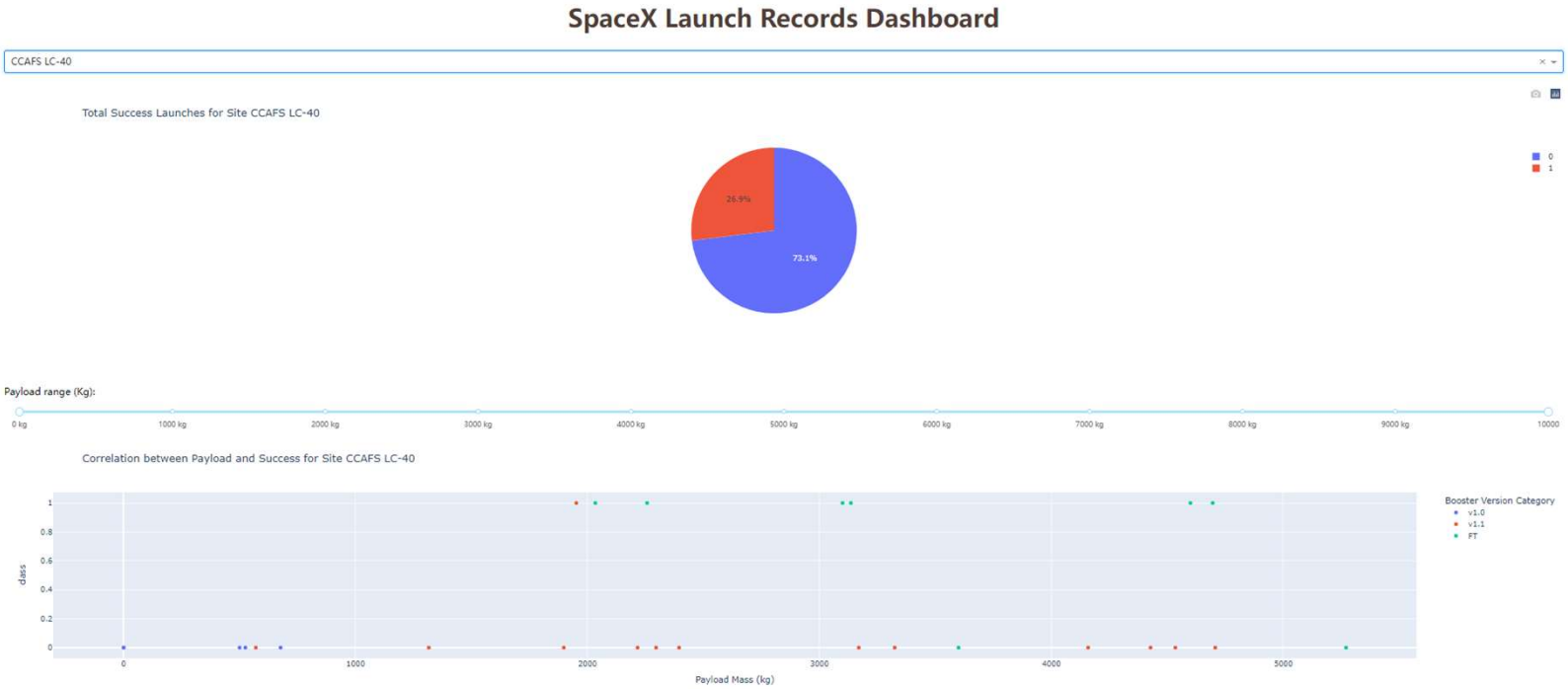


DASHBOARD

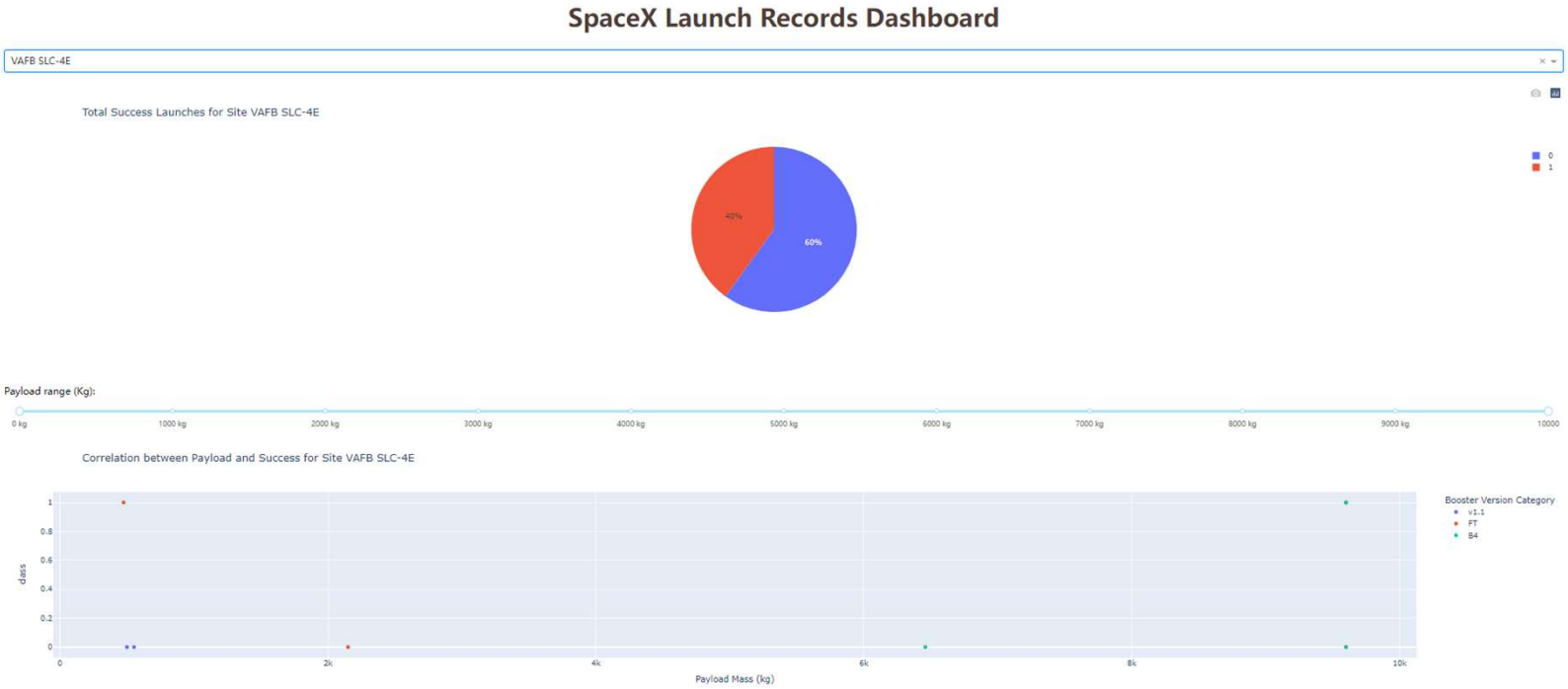
Overall


SpaceX Launch Records Dashboard
All Sites
Total Success Launches by Site
CCAFS LC-40
KSC LC-39A
VAFB SLC-4E
CCAFS SLC-40
23.2%
46.4%
17.9%
12.5%
Payload range (Kg):
0 kg
1000 kg
2000 kg
3000 kg
4000 kg
5000 kg
6000 kg
7000 kg
8000 kg
9000 kg
10000
Correlation between Payload and Success for All Sites
class
Payload Mass (kg)
Booster Version Category
v1.0
v1.1
FT
B4
B5

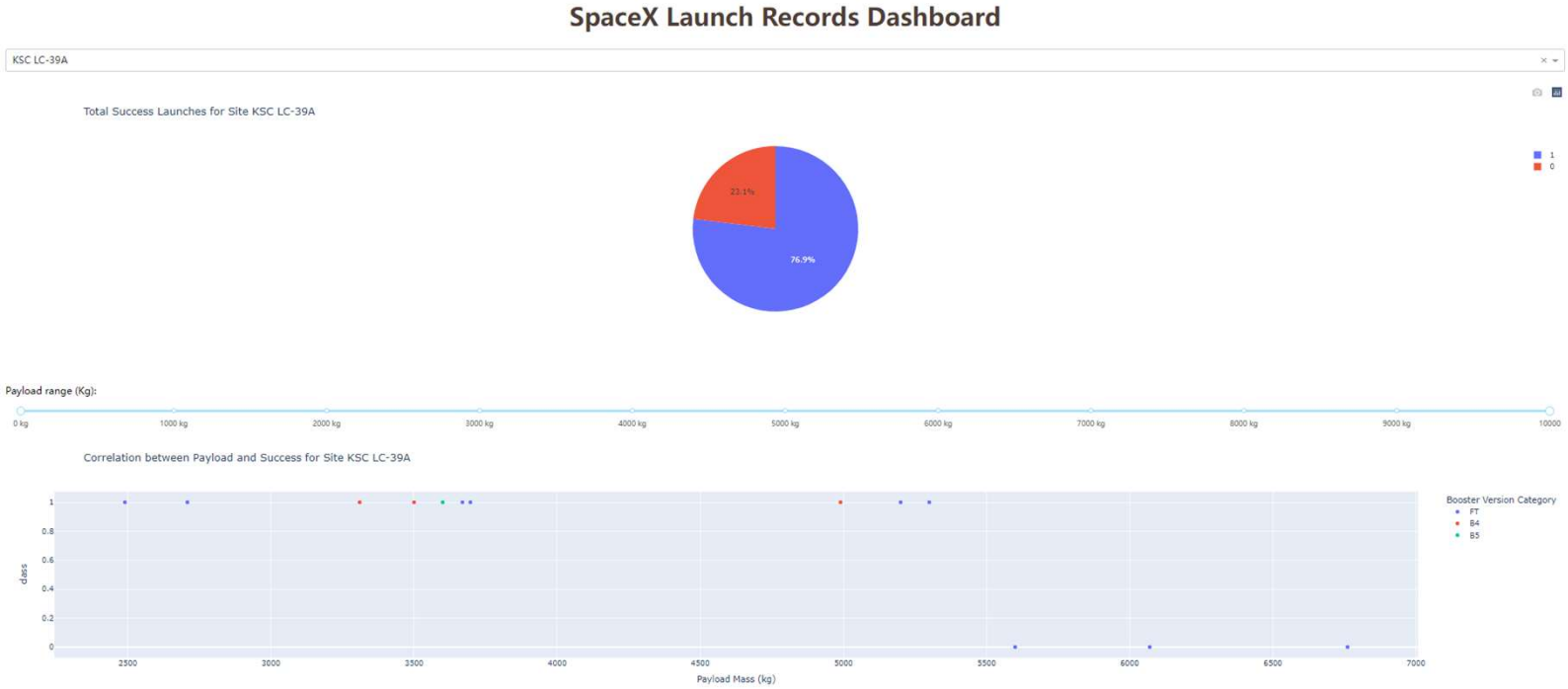
CCAFS LC-40


SpaceX Launch Records Dashboard
CCAFS LC-40
Total Success Launches for Site CCAFS LC-40
26.9%
73.1%
0
1
Payload range (Kg):
0 kg
1000 kg
2000 kg
3000 kg
4000 kg
5000 kg
6000 kg
7000 kg
8000 kg
9000 kg
10000
Correlation between Payload and Success for Site CCAFS LC-40
class
Payload Mass (kg)
Booster Version Category
v1.0
v1.1
FT

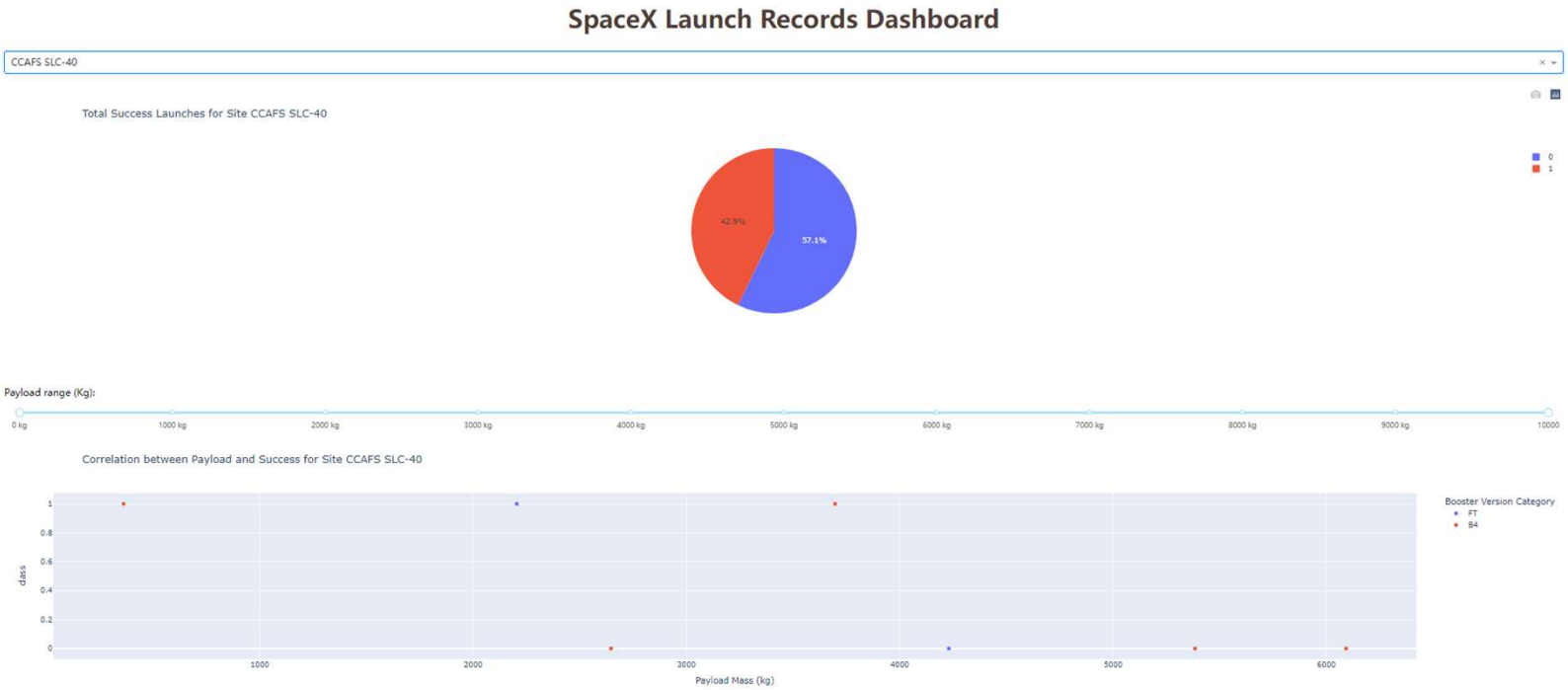
VAFB SLC-4E


SpaceX Launch Records Dashboard
VAFB SLC-4E
Total Success Launches for Site VAFB SLC-4E
40%
60%
0
1
Payload range (Kg):
0 kg
1000 kg
2000 kg
3000 kg
4000 kg
5000 kg
6000 kg
7000 kg
8000 kg
9000 kg
10000
Correlation between Payload and Success for Site VAFB SLC-4E
class
Payload Mass (kg)
Booster Version Category
v1.1
FT
B4

KSC LC-39A(Highest Success Rate)



CCAFS SLC-40


SpaceX Launch Records Dashboard
CCAFS SLC-40
Total Success Launches for Site CCAFS SLC-40
42.9%
57.1%
0
1
Payload range (Kg):
0 kg
1000 kg
2000 kg
3000 kg
4000 kg
5000 kg
6000 kg
7000 kg
8000 kg
9000 kg
10000
Correlation between Payload and Success for Site CCAFS SLC-40
class
Payload Mass (kg)
Booster Version Category
FT
B4

Classification model by machine learning

Model Choosing
Accuracy of 4 different models

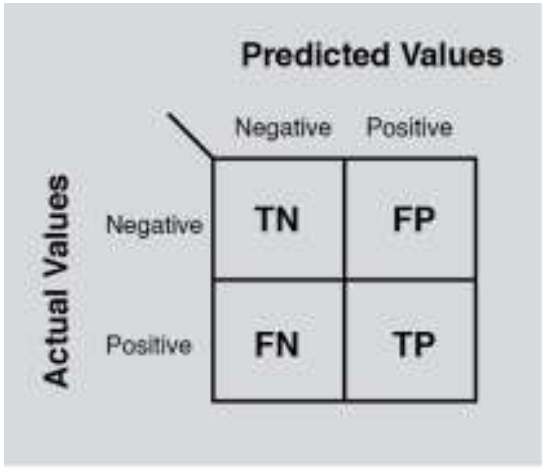
Link to github

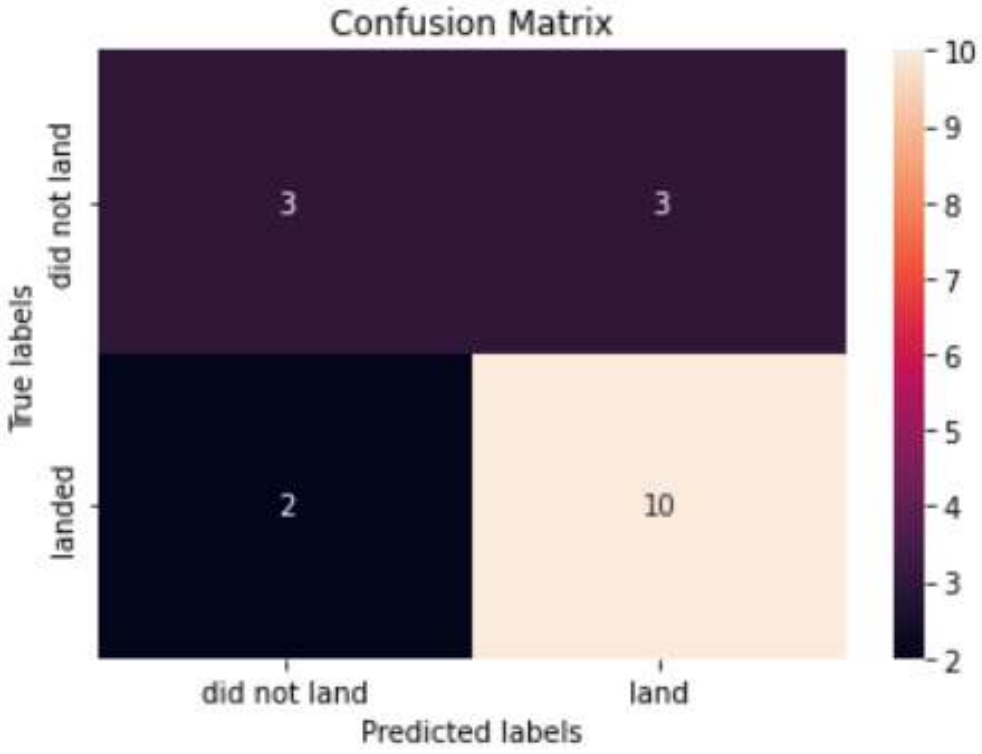
MODEL	Accuracy
LogisticRegression	0.84642857
SVN	0.84821429
TREE	0.87678571
KNN	0.84821429

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 convivence I put a guide figure to help audiences have understand the Confusion Matrix





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.