IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Name : 韦昆宏
Date : 2022.11.05

Outline

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

Executive Summary

- **Summary of methodologies**
  - Collecting the Data
  - Data Wrangling
  - Exploratory Analysis Using SQL
  - Exploratory Analysis Using graghs for Visualization
  - Interactive Visual Analytics with Dashboard
  - Predict Success Or Failure Using Machine Learning
- **Summary of all results**
  - Find Data For The Project
  - EDA
  - Prediction and Explain Data from results

Introduction

- **Project background and context**

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, much of the savings is because SpaceX can reuse the first stage. Therefore if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

In this project, we will predict if the Falcon 9 first stage will land successfully and explain data from results.

Problems you want to find answers

Which site has the highest launch success rate?

Which payload range(s) has the highest launch success rate?

Which Booster version has the highest launch success rate?

If there any other reason influences the landing outcome?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Scrapping data from websites like wikipedia
  - Request to the SpaceX API to download data
- Perform data wrangling
  - To cleaned Nan or Null data, to transferred non-numerical data to one-hot encoding
  - Normalization data for prediction in future.

Methodology

Executive Summary

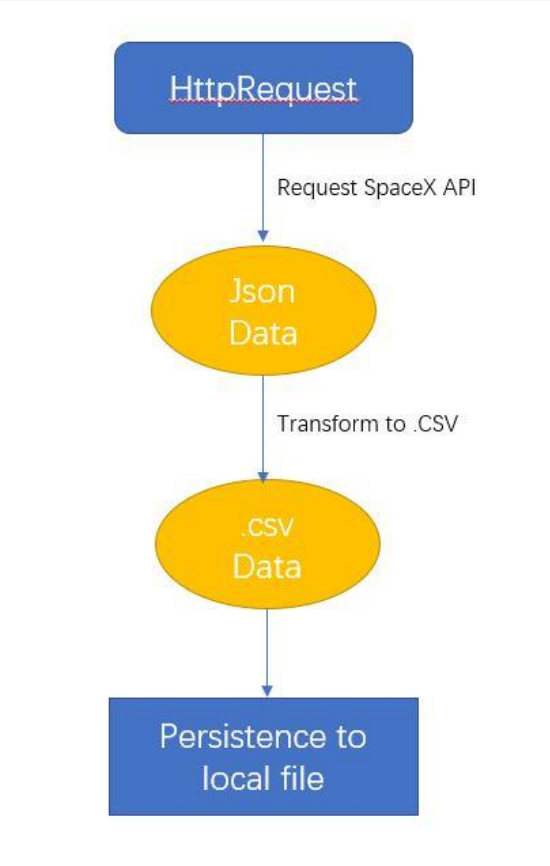
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Building models with sklearn-api.
  - Tuning parameters with grid-search
  - Evaluate models with confusion-matrix

Data Collection

- Using spacex-api to download and transfer the data to .cvs file from the json format data.
- Web scrap Falcon 9 launch records with BeautifulSoup, Extract a Falcon 9 launch records HTML table from Wikipedia and Parse the table and convert it into a Pandas data frame.

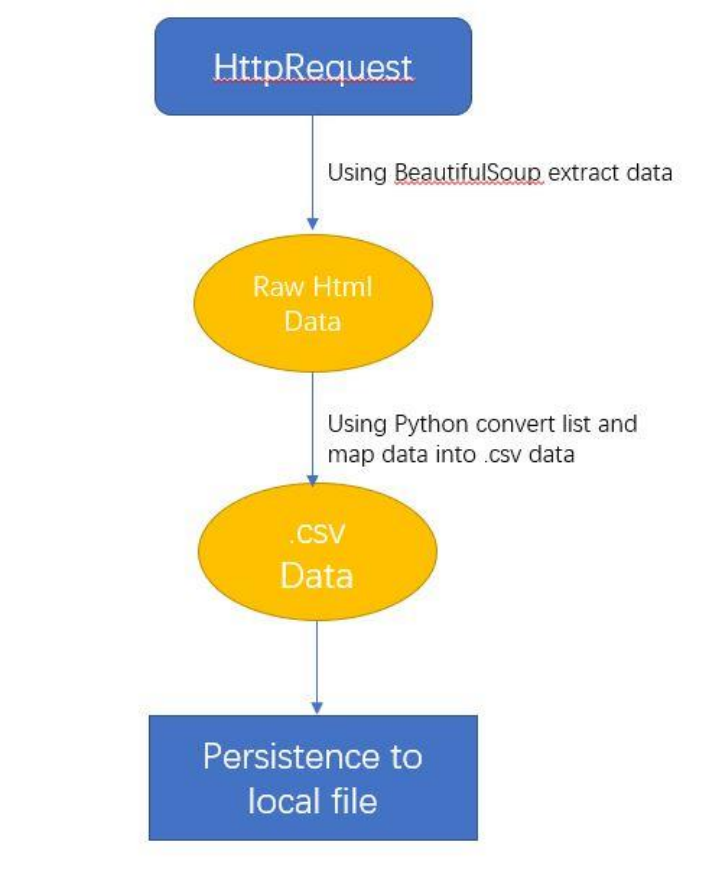
Data Collection – SpaceX API

- GitHub URL : courserahw/1.1.jupyter-labs-spacex-data-collection-api.ipynb at main · wkh1245/courserahw (github.com)



Data Collection - Scraping

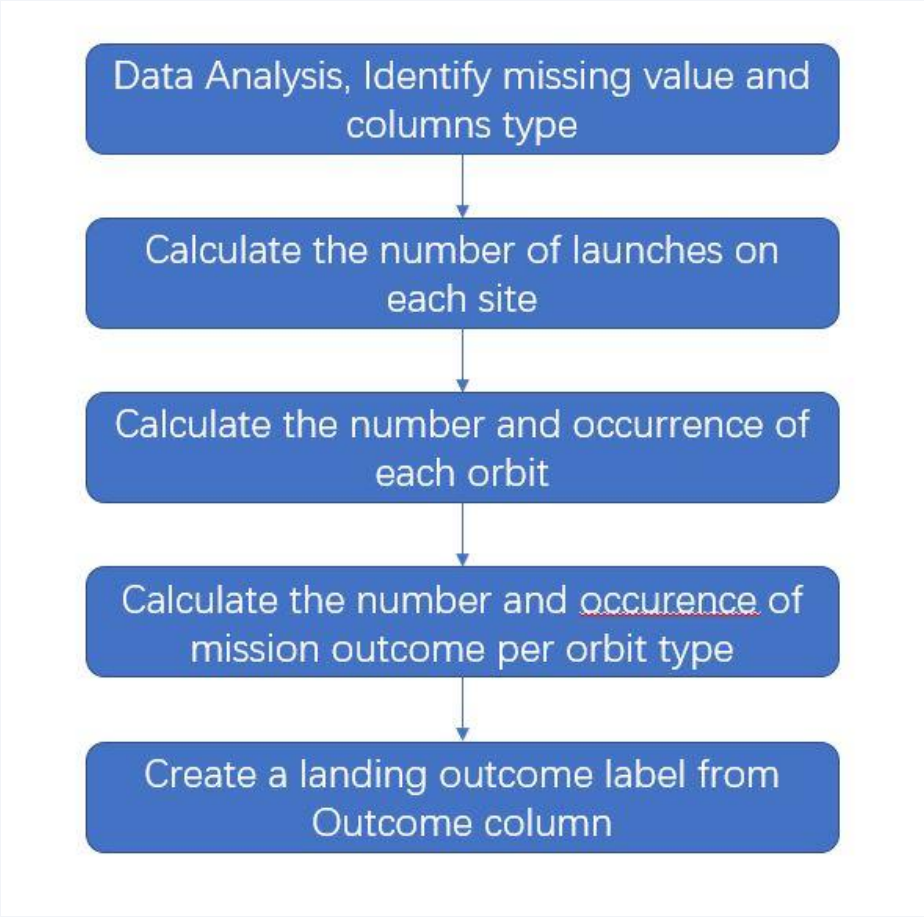
- GitHub URL : courserahw/1.2.jupyter-labs-webscraping.ipynb at main · wkh1245/courserahw (github.com)



Data Wrangling

- Explored columns and data
- Check data info using pandas including statistics
- Clean all Nan or Null data
- Label outcome into 0 or 1

- GitHub URL : courserahw/1.3.1labs-jupyter-spacex-Data wrangling.ipynb at main · wkh1245/courserahw (github.com)



EDA with Data Visualization

- Using seaborn to ploted charts to explored relations between different features and outcome, such as:

1. Plot out the FlightNumber vs. PayloadMass using scatter chart
2. Plot out the FlightNumber vs. PayloadMass using scatter chart
3. Plot out the Flight Number and Launch Site using scatter chart
4. and so on...

- GitHub URL : courserahw/2.2.jupyter-labs-eda-dataviz.ipynb at main · wkh1245/courserahw (github.com)

EDA with SQL

- Display the names of the unique launch sites in the space mission
- Display 5 records where launch sites begin with the string 'CCA'
- Display the total payload mass carried by boosters launched by NASA (CRS)
- Display average payload mass carried by booster version F9 v1.1
- List the date when the first succesful landing outcome in ground pad was acheived
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- List the total number of successful and failure mission outcomes
- List the names of the booster_versions which have carried the maximum payload mass. Use a subquery
- List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015
- Rank the count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order
- GitHub URL : courserahw/2.1.jupyter-labs-eda-sql-coursera_sqllite.ipynb at main · wkh1245/courserahw (github.com)

Build an Interactive Map with Folium

- I used markers, circles objects and added them to a folium map
- Using Circle object for highlighted the area of coordinate
- Using Marker object for mark down all launch site belongs to the area
- GitHub URL : courserahw/3.1.lab_jupyter_launch_site_location（完成）.ipynb at main · wkh1245/courserahw (github.com)

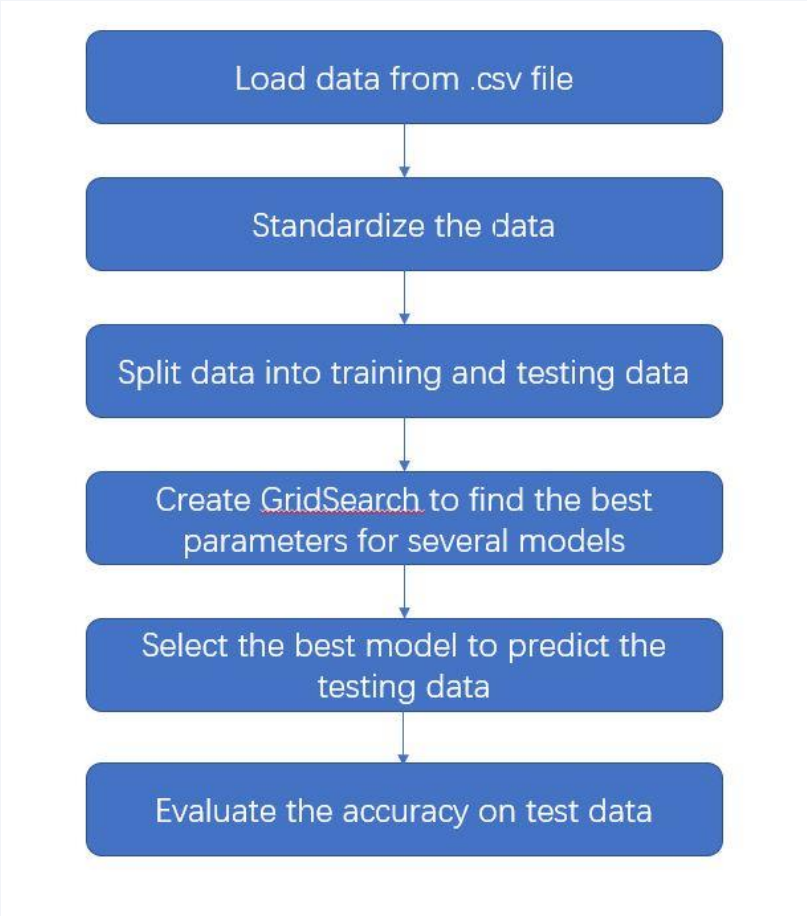
Build a Dashboard with Plotly Dash

- Using Pie chart to see the success ratio for each Launch Site
- Using Scatter chart to see the success ratios with different Payloads interval
- Interactive visual analytics enables users to explore and manipulate data in an interactive and real-time way.
- Interactive visual analytics enables users to insight data with more efficency
- GitHub URL : courserahw/3.2.1.spacex_dash_app (完成).py at main · wkh1245/courserahw (github.com)

Predictive Analysis (Classification)

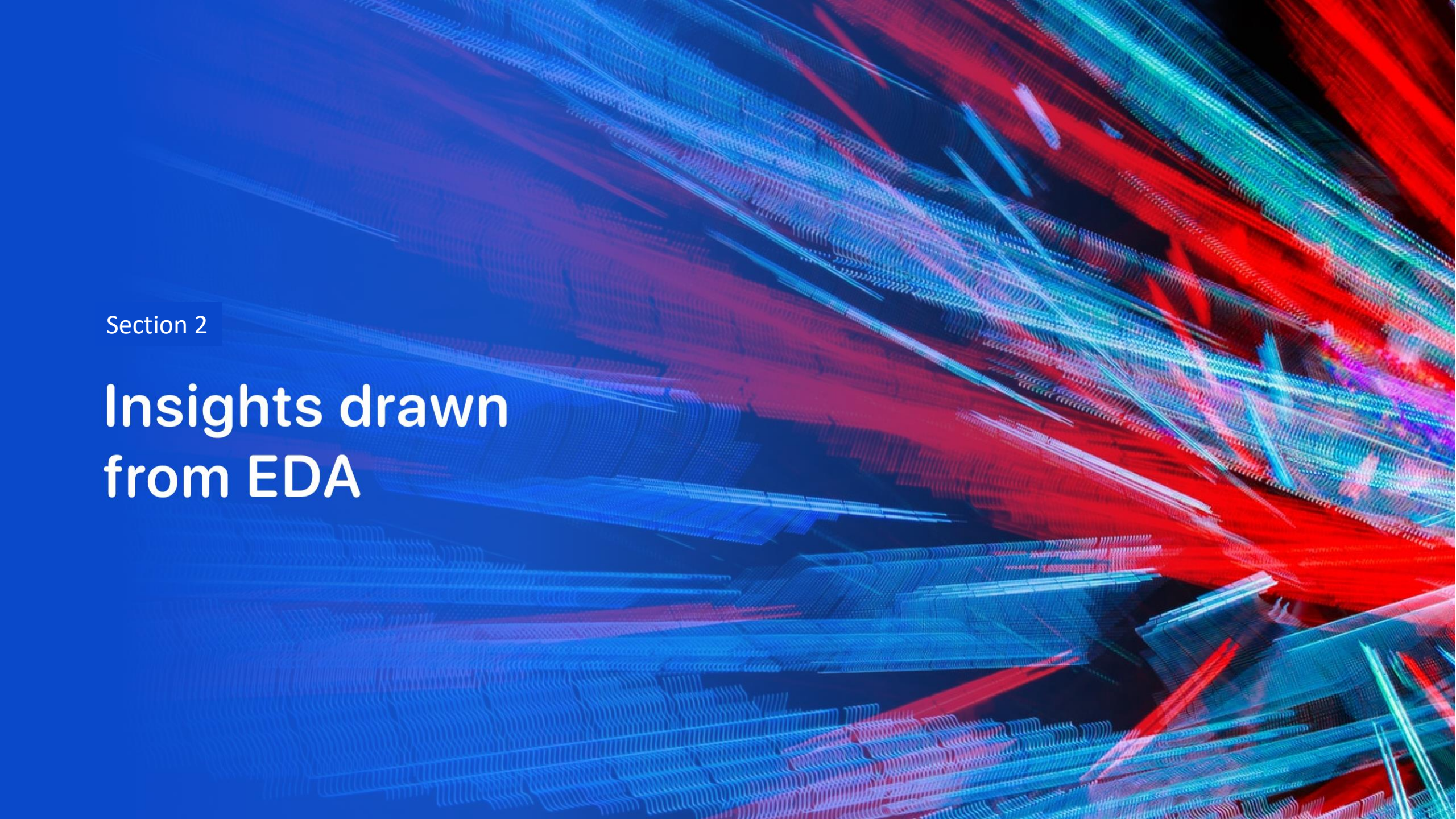
- Load data from .csv file
- Standardize the data
- Split data into training and testing data
- Create GridSearch to find the best parameters for several models
- Select the best model to predict the testing data
- Evaluate the accuracy on test data

- GitHub URL : courserahw/4.1.SpaceX_Machine Learning Prediction_Part_5.ipynb at main · wkh1245/courserahw (github.com)



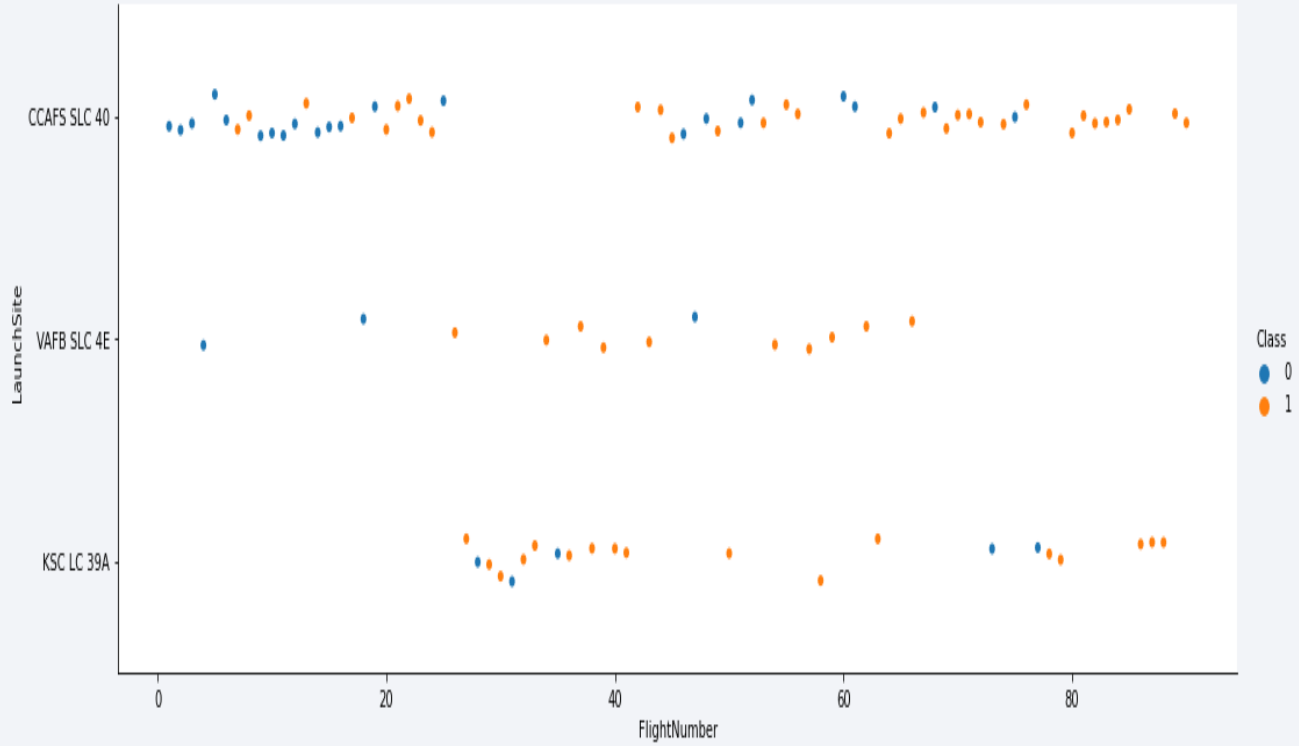
Results

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


Section 2
Insights drawn from EDA

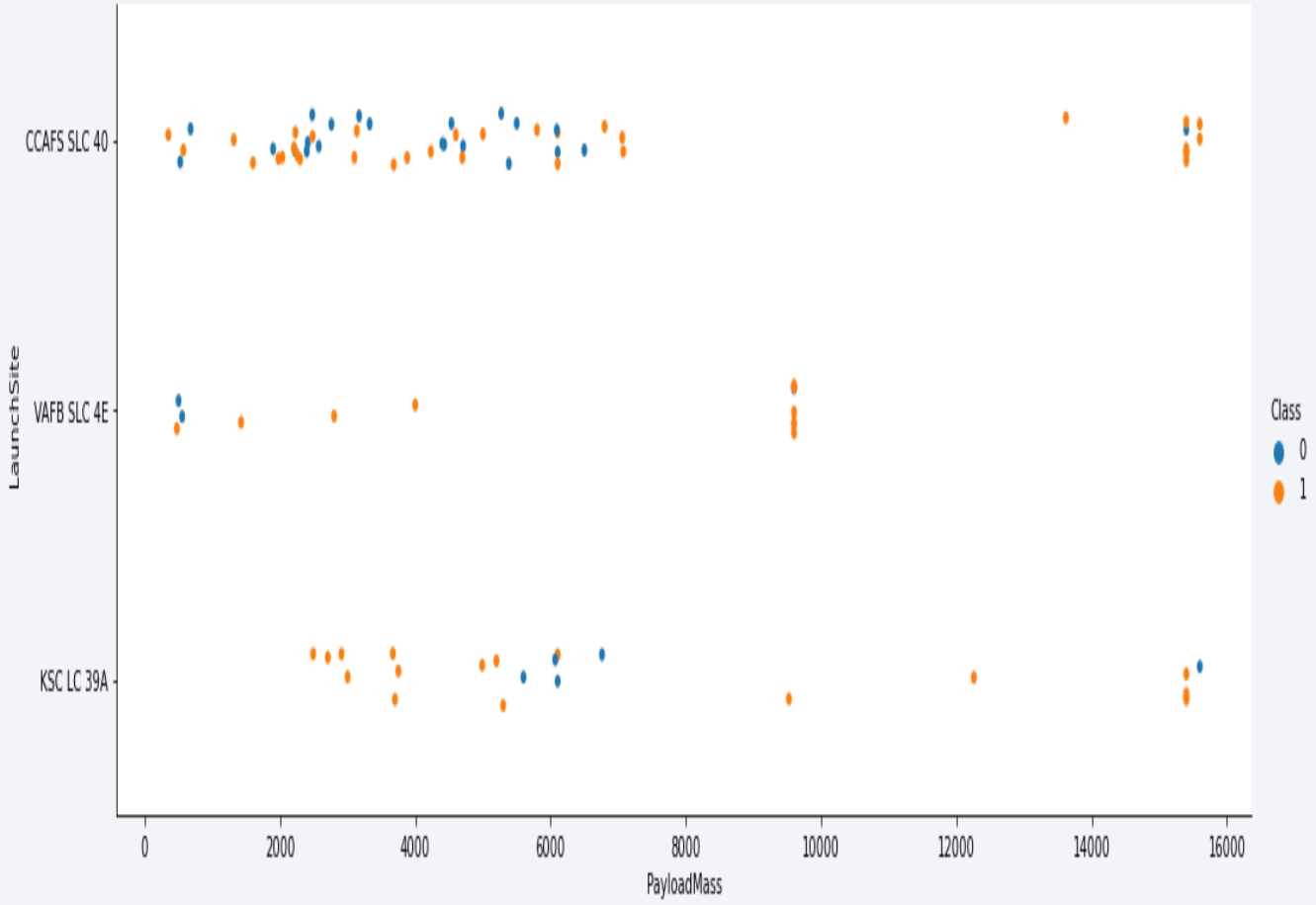
Flight Number vs. Launch Site

- Green dots means land successfully
- Blue dots means land failure
- As the flight number increases, the result is more likely to land successfully
- CCAFS SLC 40 has the highest flight number



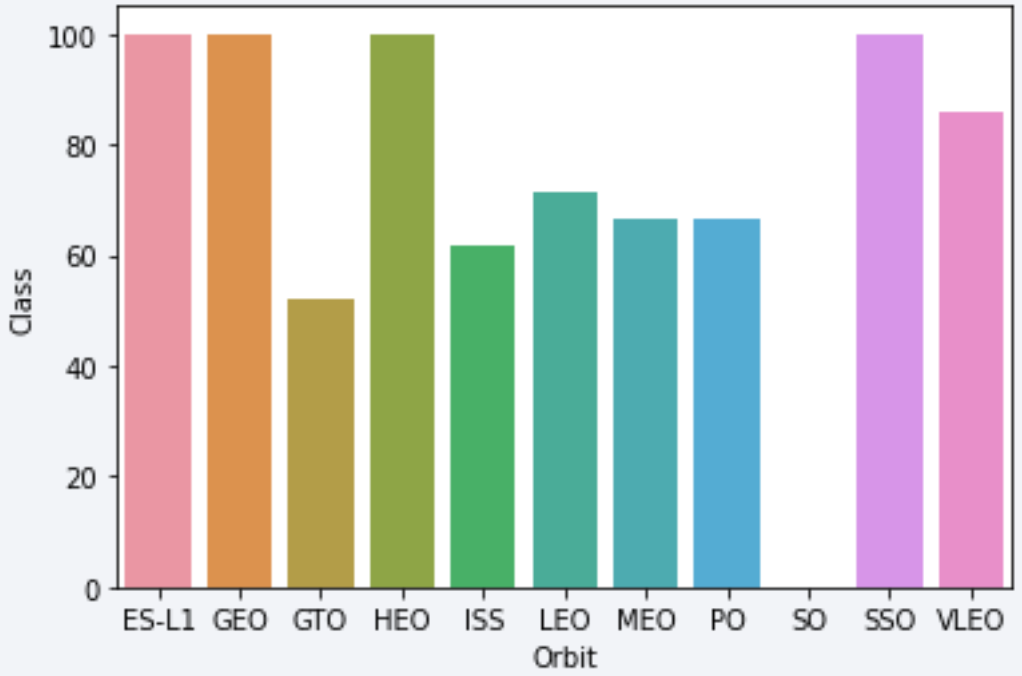
Payload vs. Launch Site

- VAFB SLC 4E have no rockets launched for heavypayload mass(greater than 10000).
- For CCAFS SLC 40, the more payload heavier, the higher success ratio
- KSC LC 39A is not clear trending that success ratio would increase with higher payload



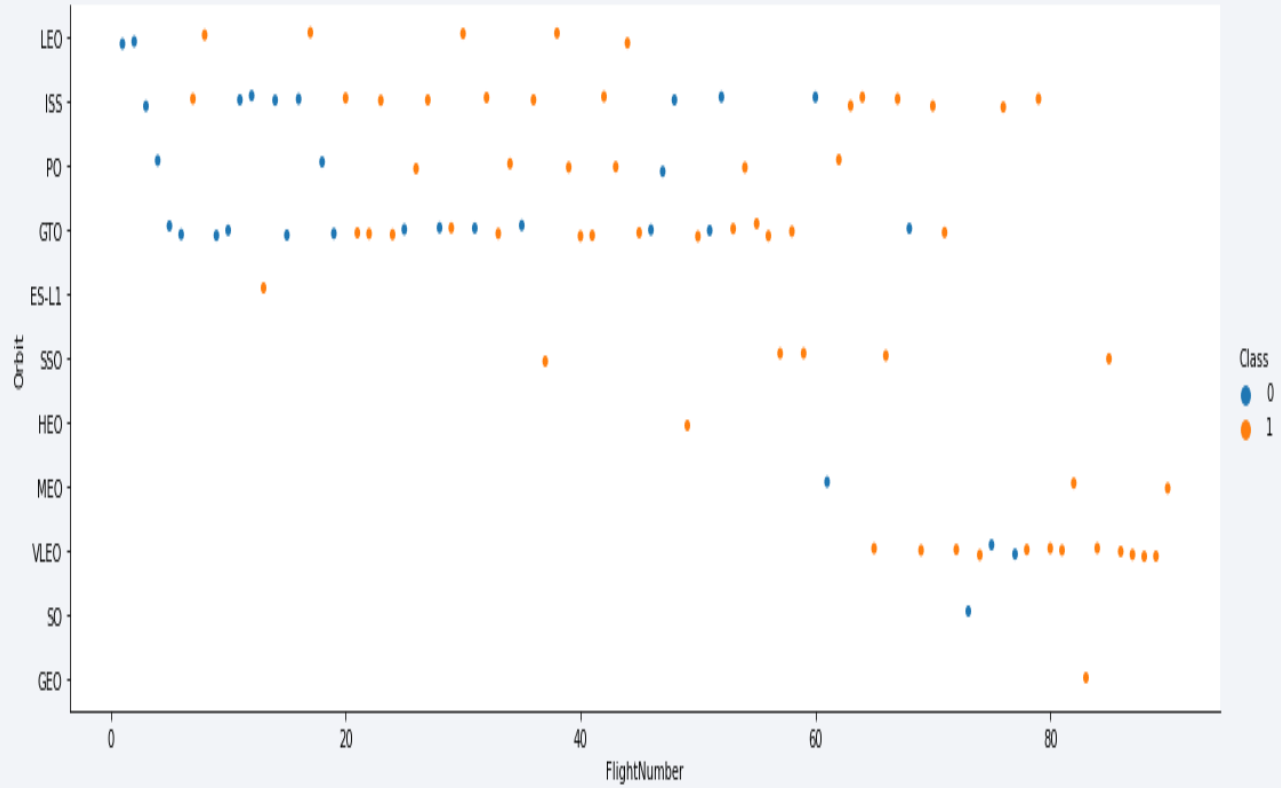
Success Rate vs. Orbit Type

- It is indicated that ES-L1, GEO,HEO and SSO are 100% success ratio
- The lowest success ratio is 0%, orbit type is SO
- Others are higher than 60% except GTO



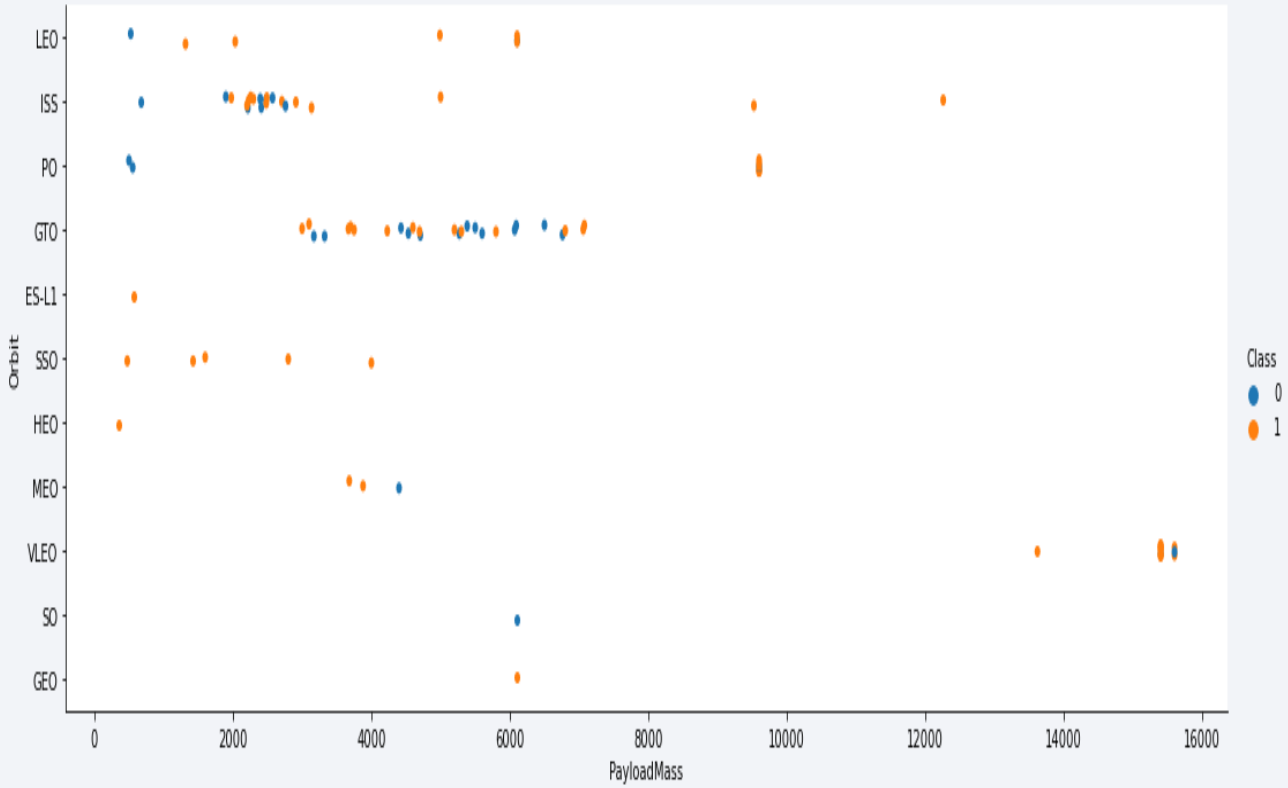
Flight Number vs. Orbit Type

- Only LEO indicated that success ratio related to FlightNumber



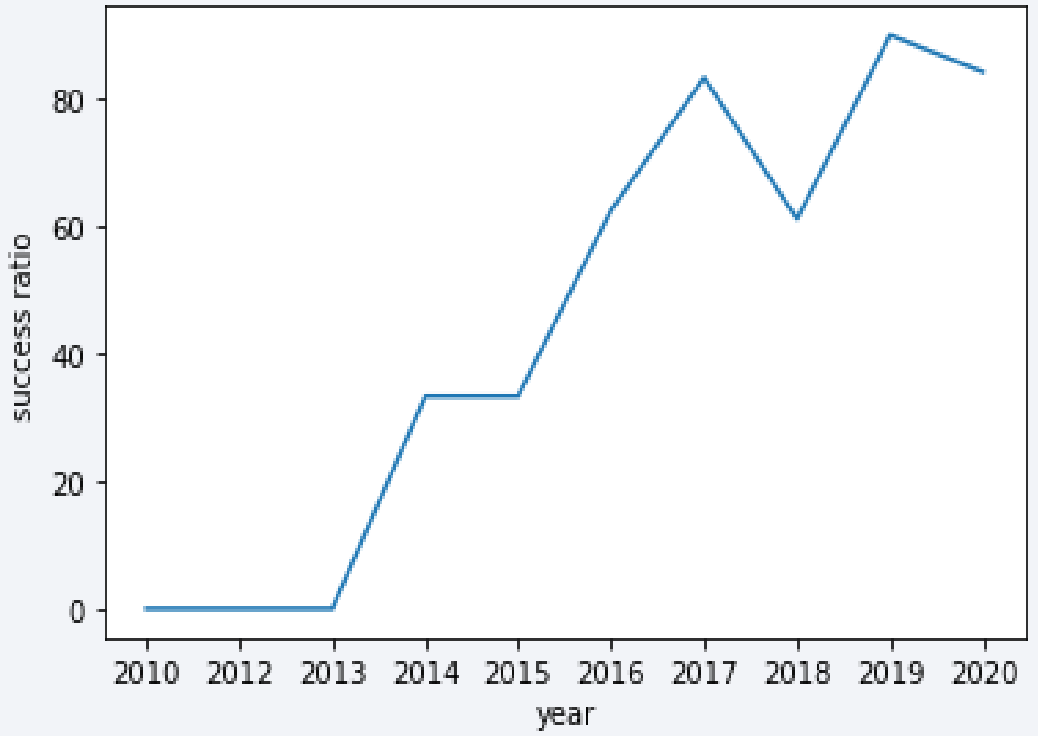
Payload vs. Orbit Type

- Only LEO indicated that success ratio related to Payload
- SSO have the highest success ratio with low payload



Launch Success Yearly Trend

- There are a clear trending that with years increase, the success ratio is higher after 2013



All Launch Site Names

Names of the unique launch sites : CCAFS LC-40, CCAFS SLC 40, KSC LC 39A , VAFB SLC 4E

Launch Site Names Begin with 'CCA'

- 5 records where launch sites begin with the string 'CCA'

	Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
0	08-12-2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of...	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
1	22-05-2012	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2	08-10-2012	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
3	01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt
4	03-12-2013	22:41:00	F9 v1.1	CCAFS LC-40	SES-8	3170	GTO	SES	Success	No attempt

Total Payload Mass

- **total payload mass carried by boosters launched by NASA (CRS) is 48213**

```
cur.execute('select sum("PAYLOAD_MASS__KG_") from SPACEXTBL where "Customer" like "%NASA (CRS)%"')
values = cur.fetchall()
values
```

```
[(48213,)]
```

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1 is 2534.67

```
cur.execute('select AVG("PAYLOAD_MASS__KG_") from SPACEXTBL where "Booster_Version" like "%F9 v1.1%"')
values = cur.fetchall()
values
```

```
[(2534.6666666666665,)]
```

First Successful Ground Landing Date

- The date of the first successful landing outcome in ground pad is 2015-12-22

```
# cur.execute('select min(date("Date")) from SPACEXTBL where "Landing _Outcome" = "Success (ground pad)"')
cur.execute('select min(strftime(SUBSTR(Date, 7, 4)||"-"||SUBSTR(Date, 4,2)||"-"||SUBSTR(Date, 1, 2))) from SPACEXTBL where "Landing _Outcome"
values = cur.fetchall()
values
```

```
[('2015-12-22',)]
```

Successful Drone Ship Landing with Payload between 4000 and 6000

- 4 booster has payload between 4000 and 6000

```
cur.execute('select * from SPACEXTBL where "PAYLOAD_MASS__KG_" between 4000 and 6000 and "Landing _Outcome" = "Success (drone ship)"')
values = cur.fetchall()
```

```
df_tmp = pd.DataFrame(data=values, columns=df.columns)
df_tmp
```

	Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
0	06-05-2016	5:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
1	14-08-2016	5:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2	30-03-2017	22:27:00	F9 FT B1021.2	KSC LC-39A	SES-10	5300	GTO	SES	Success	Success (drone ship)
3	11-10-2017	22:53:00	F9 FT B1031.2	KSC LC-39A	SES-11 / EchoStar 105	5200	GTO	SES EchoStar	Success	Success (drone ship)

Total Number of Successful and Failure Mission Outcomes

- Total number of successful and failure mission outcomes is 71

```
cur.execute('select count(*) from SPACEXTBL where "Landing _Outcome" like "Failure%" or "Landing _Outcome" like "Success%"')
values = cur.fetchall()
values
```

```
[(71,)]
```

Boosters Carried Maximum Payload

- The maximum payload mass is 15600

```
cur.execute('select "Booster_Version", "PAYLOAD_MASS__KG_" from SPACEXTBL where "PAYLOAD_MASS__KG_" = (select max("PAYLOAD_MASS__KG_") from SPA
values = cur.fetchall()
values
```

```
[('F9 B5 B1048.4', 15600),
 ('F9 B5 B1049.4', 15600),
 ('F9 B5 B1051.3', 15600),
 ('F9 B5 B1056.4', 15600),
 ('F9 B5 B1048.5', 15600),
 ('F9 B5 B1051.4', 15600),
 ('F9 B5 B1049.5', 15600),
 ('F9 B5 B1060.2 ', 15600),
 ('F9 B5 B1058.3 ', 15600),
 ('F9 B5 B1051.6', 15600),
 ('F9 B5 B1060.3', 15600),
 ('F9 B5 B1049.7 ', 15600)]
```

2015 Launch Records

- Picture belows indicated the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

```
cur.execute('select substr(Date,4,2), substr(Date,7,4), "Landing _Outcome", "Booster_Version", "Launch_Site" from SPACEXTBL where "Landing _Out
values = cur.fetchall()
values
```

```
[('01', '2015', 'Failure (drone ship)', 'F9 v1.1 B1012', 'CCAFS LC-40'),
 ('04', '2015', 'Failure (drone ship)', 'F9 v1.1 B1015', 'CCAFS LC-40')]
```

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

- Ranking of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20

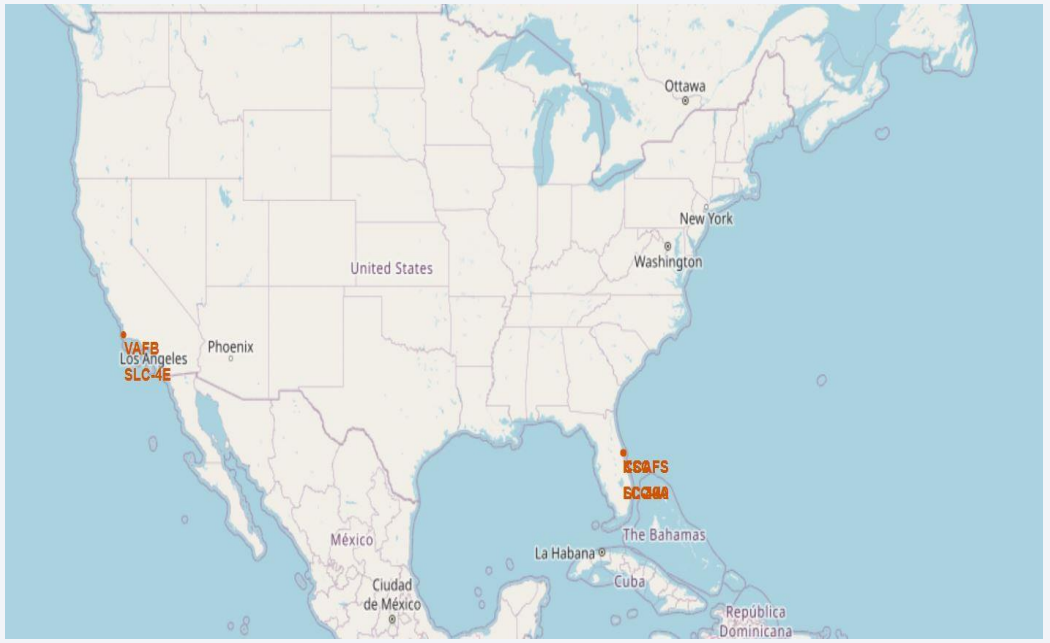
	Date	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
0	2017-02-19	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
1	2017-01-14	F9 FT B1029.1	VAFB SLC-4E	Iridium NEXT 1	9600	Polar LEO	Iridium Communications	Success	Success (drone ship)
2	2016-08-14	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
3	2016-07-18	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
4	2016-05-27	F9 FT B1023.1	CCAFS LC-40	Thaicom 8	3100	GTO	Thaicom	Success	Success (drone ship)
5	2016-05-06	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
6	2016-04-08	F9 FT B1021.1	CCAFS LC-40	SpaceX CRS-8	3136	LEO (ISS)	NASA (CRS)	Success	Success (drone ship)
7	2015-12-22	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)

Section 3

Launch Sites Proximities Analysis

Each launch site on the site map

- All launch site are near the sea coast line

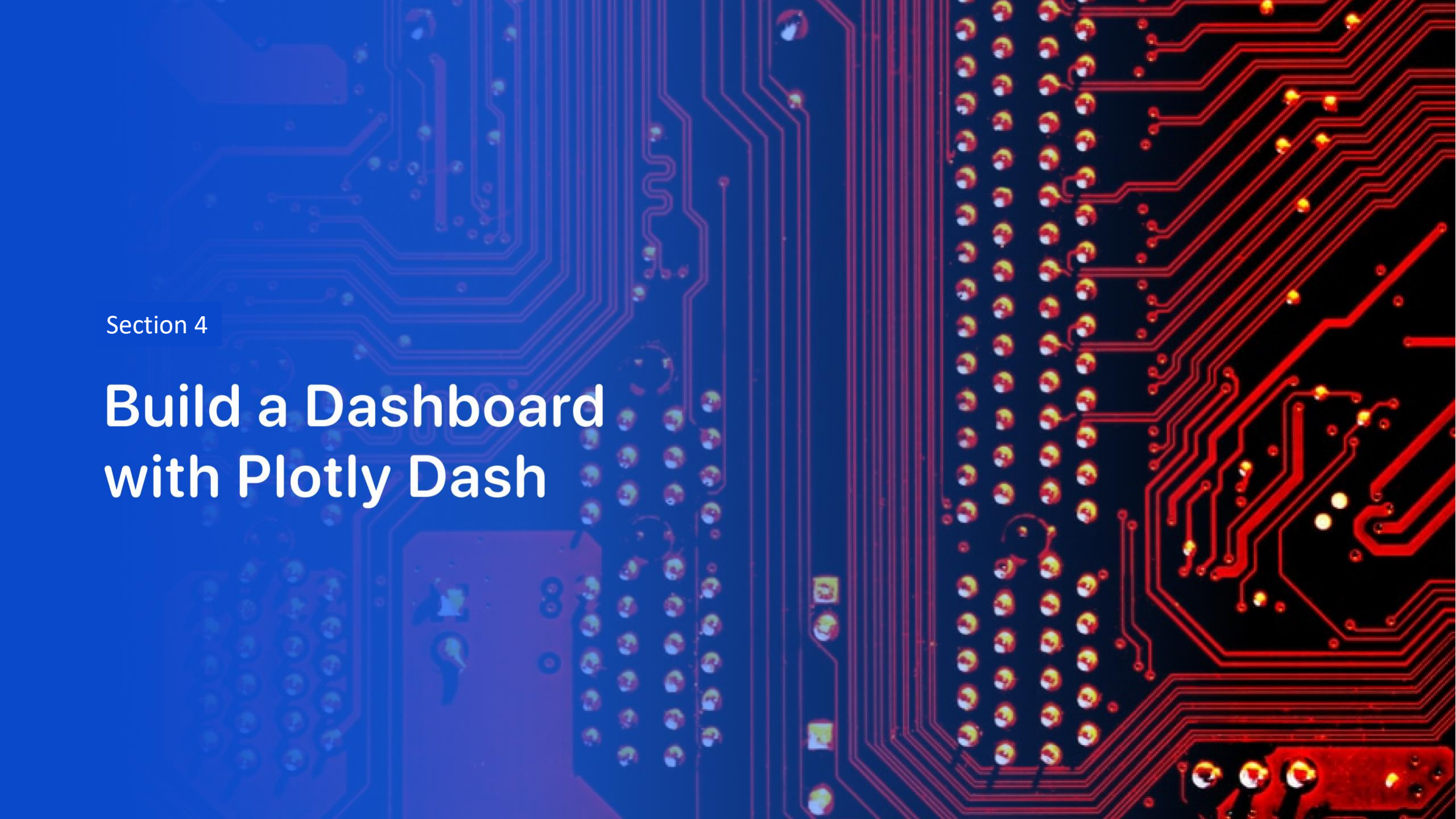


Launch outcomes on the map

- Green label means successful
- Red label mean failure

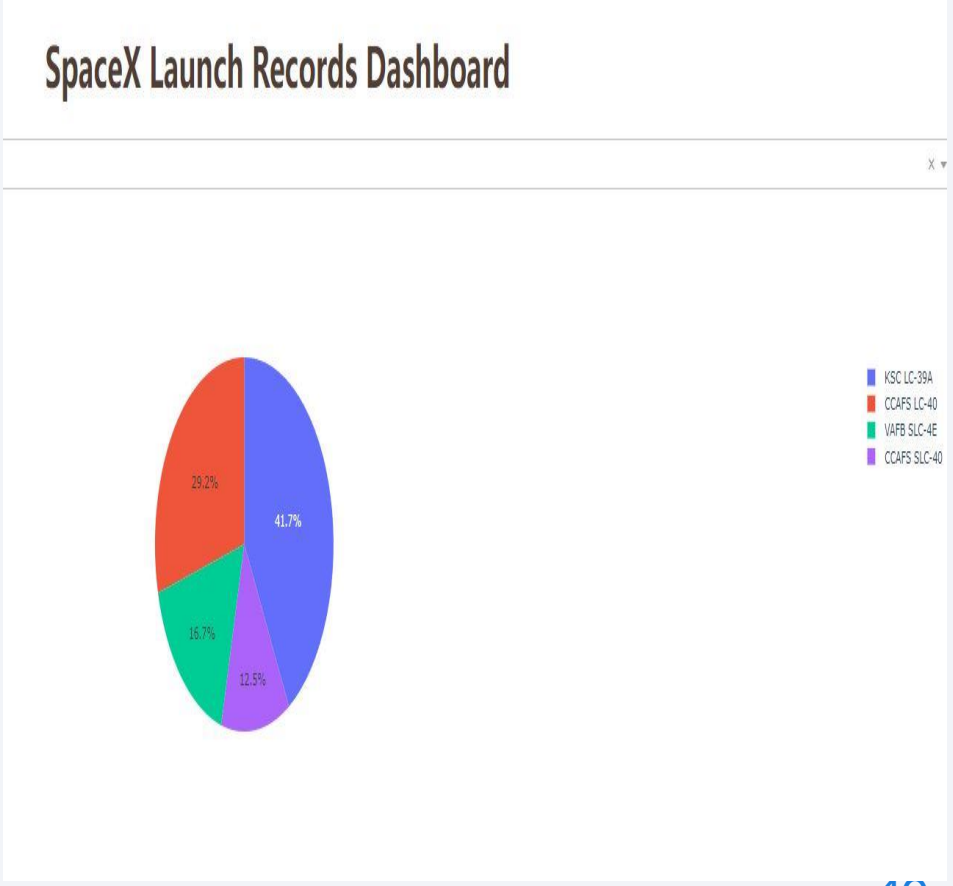
Distance between launch site and other facilities

- Picture indicated distance between launch site and sea coast


Section 4
Build a Dashboard with Plotly Dash

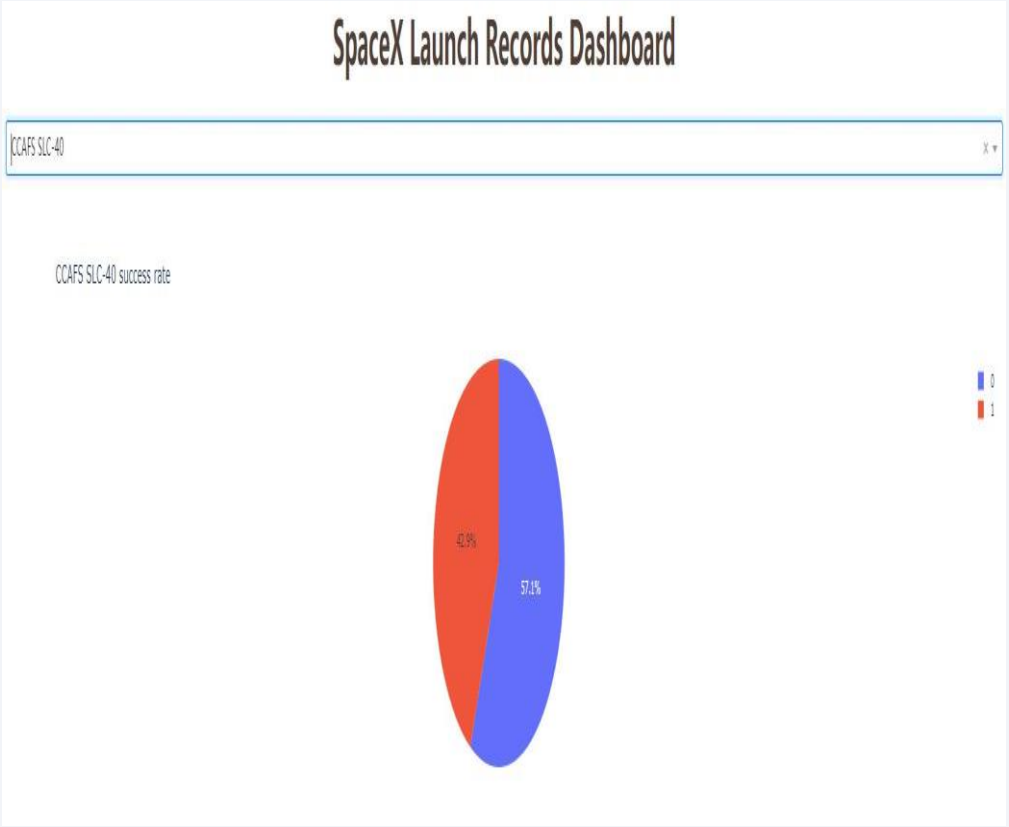
Launch success of all sites

- Pie chat indicated that KSC LC-39A is the highest success ratio among other sites.
- The lowest success ratio site of them is VAFB SLC-4E



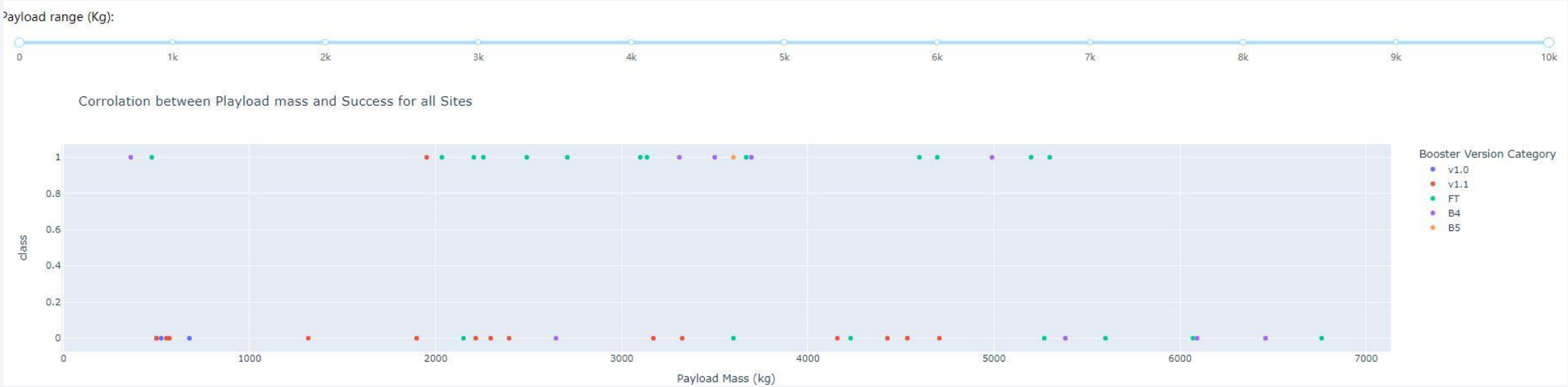
The highest success ratio of Launch site

- 57.1% is the highest success ratio of all site, names CCAFS SLC-40



<Dashboard Screenshot 3>

- B5 has the highest launch success rate



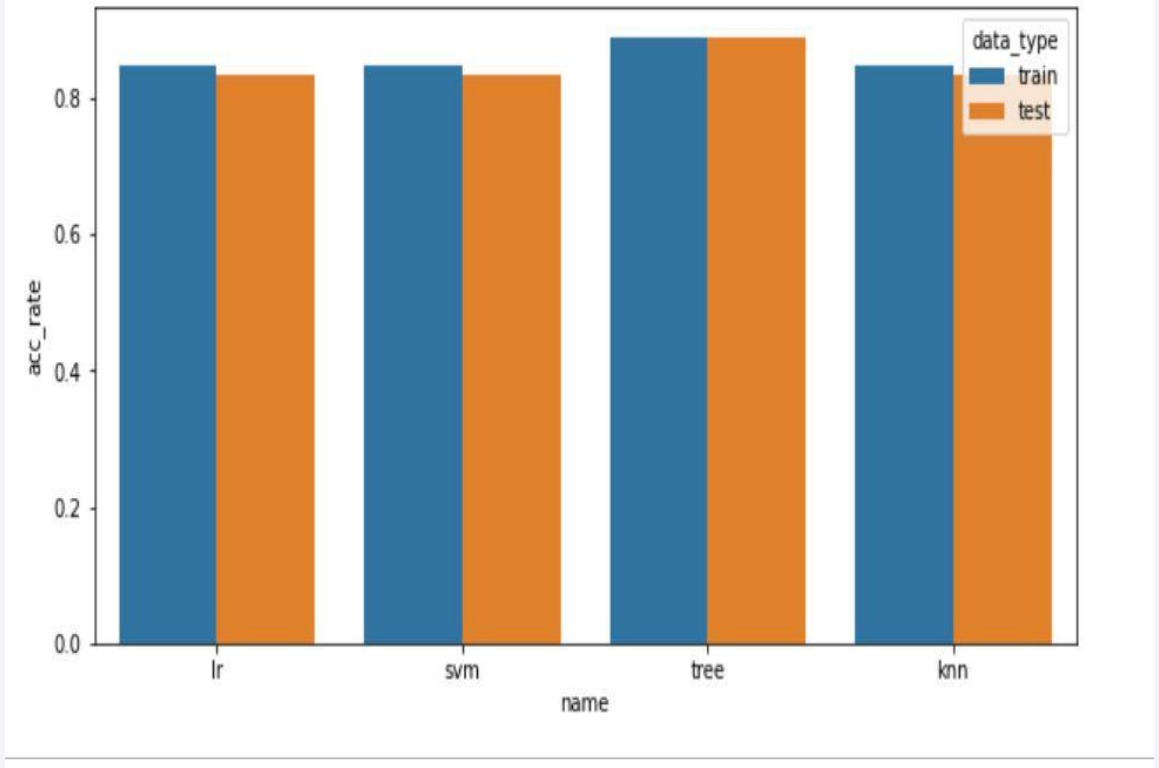
Section 5

Predictive Analysis (Classification)

Classification Accuracy

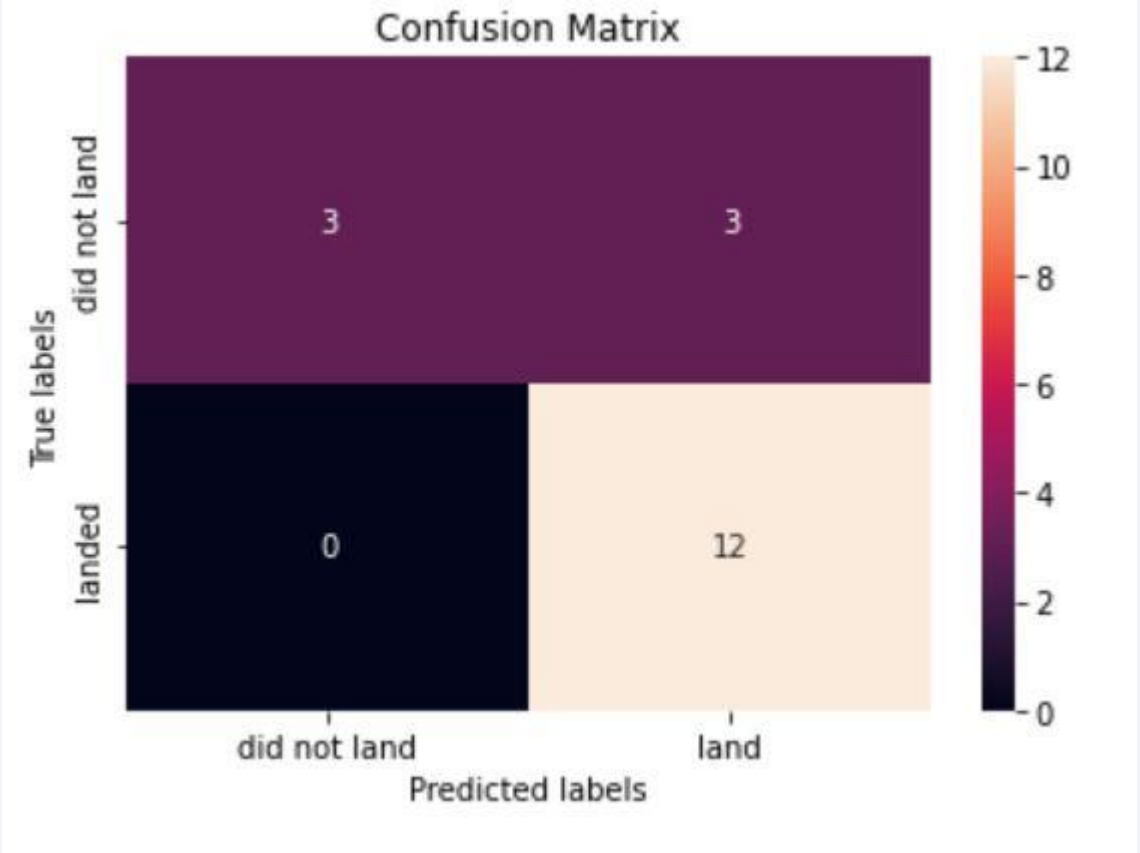
- DecisionTreeClassifier has the best accurancy in train set and test set which is 88.75% and 88.88%

```
plt.figure(figsize=(10,5))
sns.barplot(data=df_result, y='acc_rate', x='name',hue='data_type')
plt.show()
```



Confusion Matrix

- Position of False Postive shows 3, that means the machine learning model judged 3 False value into True value
- Even if there are 3 FP in the confusion matrix, the result shows a good prediction with the model



Conclusions

- KSC LC 39A has the highest success ratio which is 77.27%
- Launch sites are all closed to the sea coast
- There are not every launch site carry heavy payload mass
- SSO have the highest success ratio with low payload
- The success ratio trending increasing after 2013

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!