IBM

APPLIED DATA SCIENCE CAPSTONE RPOJECT

OPERATION SPACEX OUTCOME PREDICTION(A Machine Learning Approach)

SETHU S
22 JULY 2023

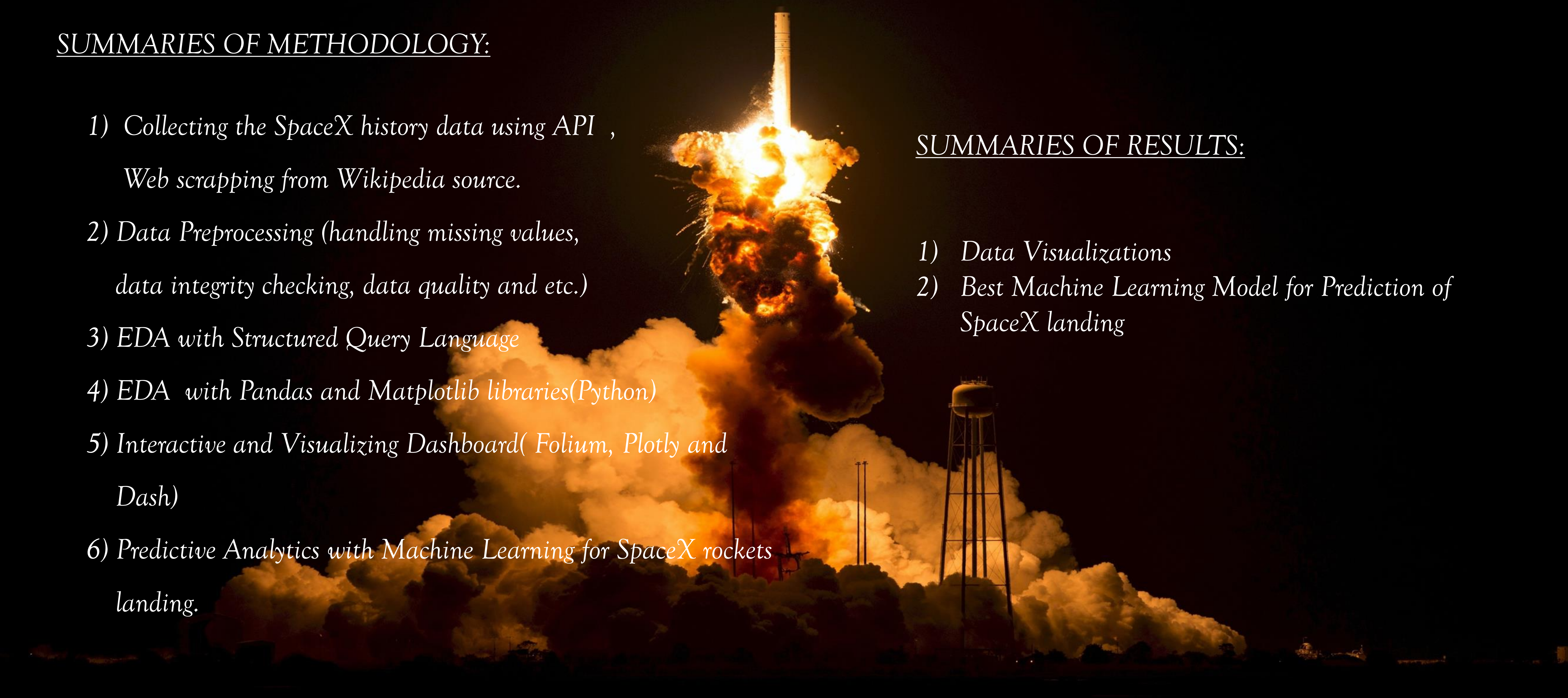
OUTLINE

- EXECUTIVE SUMMARY
- INTRODUCTION
- METHODOLOGY
- RESULTS
- CONCLUSION
- APPENDIX

GITHUB URL:

https://github.com/itsmesethus/courserac10assignments/tree/main/applied%20data%20science%20capstone%20project%20works

EXECUTIVE SUMMARY



SUMMARIES OF METHODOLOGY:

1) *Collecting the SpaceX history data using API , Web scrapping from Wikipedia source.*
2) *Data Preprocessing (handling missing values, data integrity checking, data quality and etc.)*
3) *EDA with Structured Query Language*
4) *EDA with Pandas and Matplotlib libraries(Python)*
5) *Interactive and Visualizing Dashboard(Folium, Plotly and Dash)*
6) *Predictive Analytics with Machine Learning for SpaceX rockets landing.*

SUMMARIES OF RESULTS:

1) *Data Visualizations*
2) *Best Machine Learning Model for Prediction of SpaceX landing*

INTRODUCTION

"As long as there are dreams, rockets will forever carry the hope of mankind beyond the horizon."

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.

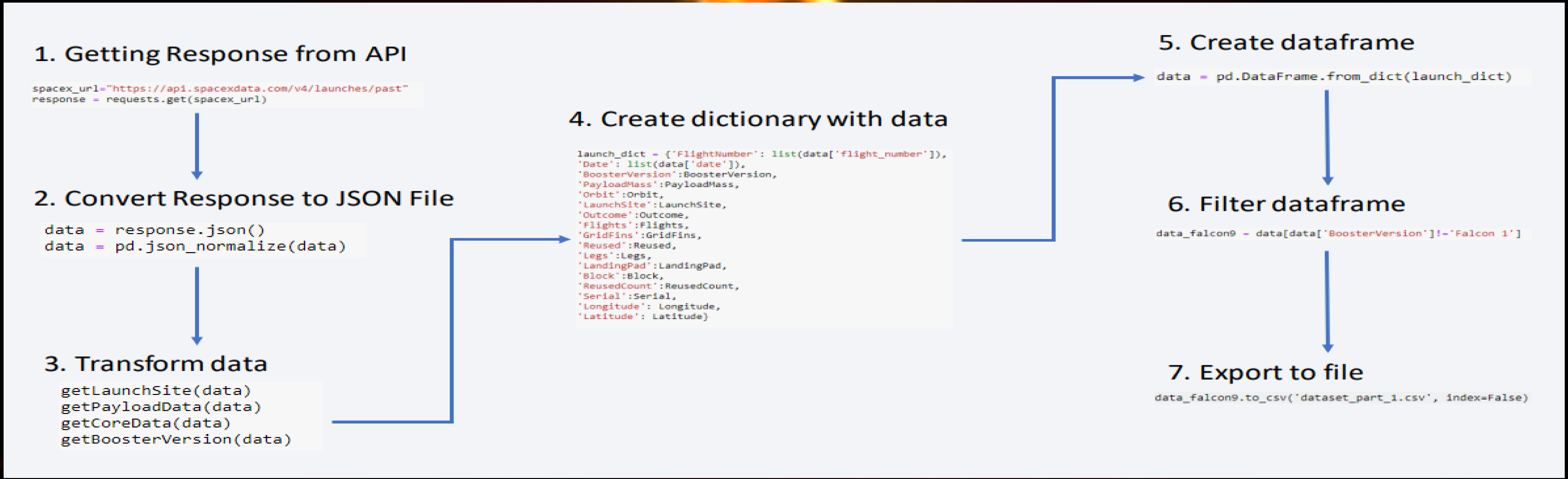
We, hereby going to travel with the process of our Data Science Team and how they have worked to solve this problem of SpaceX Organisation before their next launch of rockets into the orbit whether it will be succeful or loss for them based on their previous history launches. So, okay let's dive into the presentation of the Team.

METHODOLOGY

❖ DATA COLLECTION ----- (Rest API, Web Scrapping):

* Using the Rest API we extract the data from the source in the form of JSON and later we can easily turn that format to data frame using the help of Pandas library of Python.

* For Web Scrapping we can use the BeautifulSoup and request libraries to scrap out the data from the Wikipedia source.



https://github.com/itsmesethus/coursera-c10-assignments/blob/main/applied%20data%20science%20capstone%20project%20works/Week1%20SpaceX%20Falcon%20Data%20Collection-Wrangling.ipynb

❖ DATA PREPROCESSING :

* Here in this section we need to look for the data integrity, data quality and handling the missing values will be considered.

* Because if data is n't in correct format or any other possibilities the results may mislead to wrong predictions.

1. Calculate launches number for each site

```
df['LaunchSite'].value_counts()
```

```
CCAFS SLC 40    55
KSC LC 39A      22
VAFB SLC 4E     13
Name: LaunchSite, dtype: int64
```

2. Calculate the number and occurence of each orbit

```
df['Orbit'].value_counts()
```

```
GTO      27
ISS      21
VLEO     14
PO        9
LEO       7
SSO       5
MEO       3
SO        1
```

3. Calculate number and occurrence of mission outcome per orbit type

```
landing_outcomes = df['Outcome'].value_counts()
landing_outcomes
```

```
True ASDS      41
None None      19
True RTLS      14
False ASDS      6
True Ocean      5
None ASDS       2
False Ocean     2
False RTLS      1
Name: Outcome, dtype: int64
```

4. Create landing outcome label from Outcome column

```
landing_class = []
for key,value in df["Outcome"].items():
    if value in bad_outcomes:
        landing_class.append(0)
    else:
        landing_class.append(1)
df['Class']=landing_class
```

5. Export to file

```
df.to_csv("dataset_part_2.csv", index=False)
```

https://github.com/itsmesethus/coursera-c10-assignments/blob/main/applied%20data%20science%20capstone%20project%20works/Week1%20SpaceX%20Falcon%20Data%20Collection-Wrangling.ipynb

❖ EDA WITH STRUCTURED QUERY LANGUAGE:

* SQL is the best programming language when handling in terms of huge volumes of data. Using this we have done the EDA for the Falcon 9 rockets. And some results are,

Display the names of the unique launch sites in the space mission

```
%sql select distinct(LAUNCH_SITE) from SPACEXTBL
```

```
* ibm_db_sa://sdk38546:***@dashdb-txn-sbox-yp-lon02-07.services.eu-gb.bluemix.net:50000/BLUDB
Done.
```

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

Display the total payload mass carried by boosters launched by NASA (CRS)

```
%sql select sum(PAYLOAD_MASS__KG_) from SPACEXTBL where CUSTOMER = 'NASA (CRS)'
```

```
* ibm_db_sa://sdk38546:***@dashdb-txn-sbox-yp-lon02-07.services.eu-gb.bluemix.net:
Done.
```

1
45596

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

```
%sql select avg(PAYLOAD_MASS__KG_) from SPACEXTBL where BOOSTER_VERSION = 'F9 v1.1'
```

```
* ibm_db_sa://sdk38546:***@dashdb-txn-sbox-yp-lon02-07.services.eu-gb.bluemix.net:500
Done.
```

1
2928.400000

List the date when the first succesful landing outcome in ground pad was acheived.

Hint:Use min function

```
%sql select min(DATE) from SPACEXTBL where Landing__Outcome = 'Success (ground pad)'
```

```
* ibm_db_sa://sdk38546:***@dashdb-txn-sbox-yp-lon02-07.services.eu-gb.bluemix.net:5000
Done.
```

1
2015-12-22

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

```
%sql select BOOSTER_VERSION from SPACEXTBL where Landing__Outcome = 'Success (drone ship)' and PAYLOAD_MASS__KG_ > 4000 and PA
```

```
 * ibm_db_sa://sdk38546:***@dashdb-txn-sbox-yp-lon02-07.services.eu-gb.bluemix.net:50000/BLUDB
Done.
```

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

List the total number of successful and failure mission outcomes

```
%sql select count(MISSION_OUTCOME) from SPACEXTBL where MISSION_OUTCOME = 'Success' or MISSION_OUTCOME = 'Failure (in flight)'
```

```
 * ibm_db_sa://sdk38546:***@dashdb-txn-sbox-yp-lon02-07.services.eu-gb.bluemix.net:50000/BLUDB
Done.
```

1
100

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

```
%sql select BOOSTER_VERSION from SPACEXTBL where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_) from SPACEXTBL)
```

```
 * ibm_db_sa://sdk38546:***@dashdb-txn-sbox-yp-lon02-07.services.eu-gb.bluemix.net:50000/BLUDB
Done.
```

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

Rank the count of successful landing_outcomes between the date 2010-06-04 and 2017-03-20 in descending order.

```
%sql select * from SPACEXTBL where Landing__Outcome like 'Success%' and (DATE between '2010-06-04' and '2017-03-20') order by
```

```
 * ibm_db_sa://sdk38546:***@dashdb-txn-sbox-yp-lon02-07.services.eu-gb.bluemix.net:50000/BLUDB
Done.
```

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing_o
2017-02-19	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success
2017-01-14	17:54:00	F9 FT B1029.1	VAFB SLC-4E	Iridium NEXT 1	9600	Polar LEO	Iridium Communications	Success	Succes
2016-08-14	05:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Succes
2016-07-18	04:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEO (ISS)	NASA (CRS)	Success	Success
2016-05-27	21:39:00	F9 FT B1023.1	CCAFS LC-40	Thaicom 8	3100	GTO	Thaicom	Success	Succes
2016-05-06	05:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Succes
2016-04-08	20:43:00	F9 FT B1021.1	CCAFS LC-40	SpaceX CRS-8	3136	LEO (ISS)	NASA (CRS)	Success	Succes

❖ EDA WITH MATPLOTLIB AND PANDAS:

* Matplotlib and Pandas are the most versatile libraries in Python for handling the visualizations and data frames.

Scatter Graphs :

- Flight Number vs. Payload Mass
- Flight Number vs. Launch Site
- Payload vs. Launch Site
- Orbit vs. Flight Number
- Payload vs. Orbit Type
- Orbit vs. Payload Mass

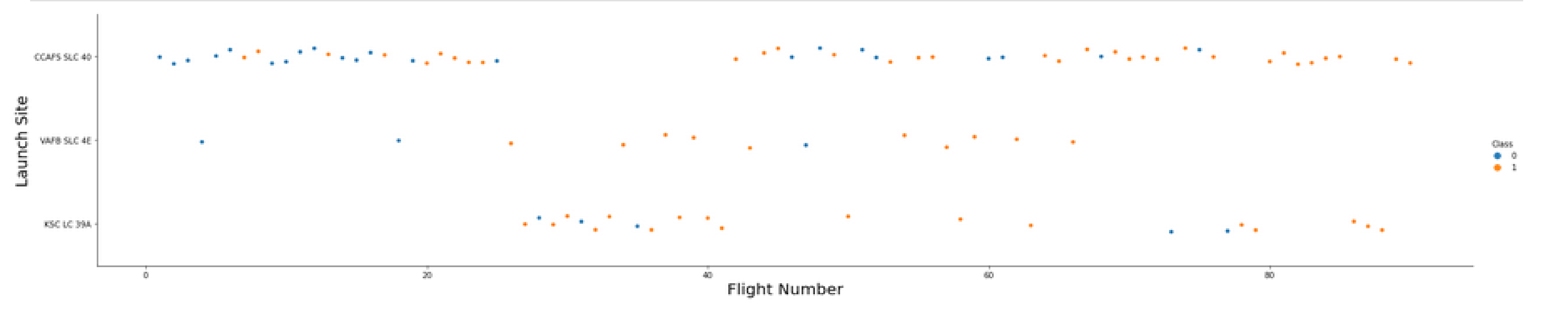
Line Graph:

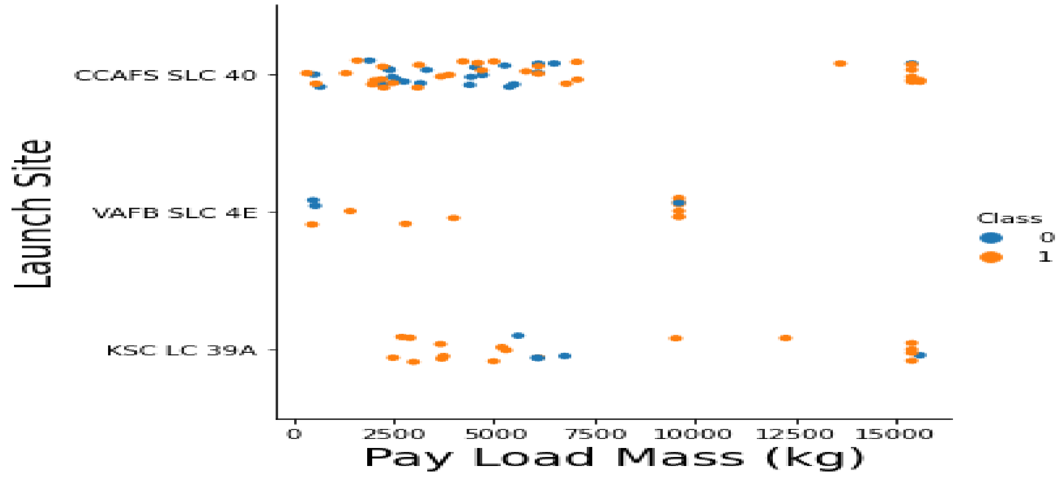
- Success rate vs. Year

Bar Graph :

- Success rate vs. Orbit

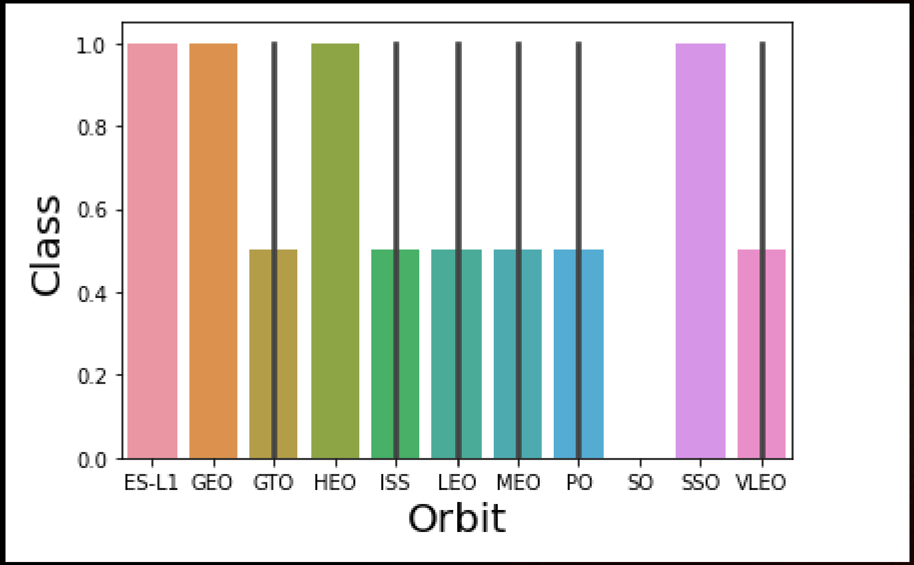
❖ EDA REASULTS:

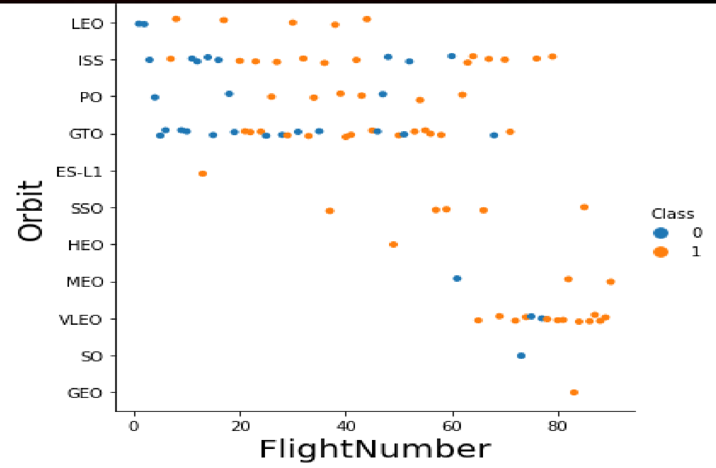




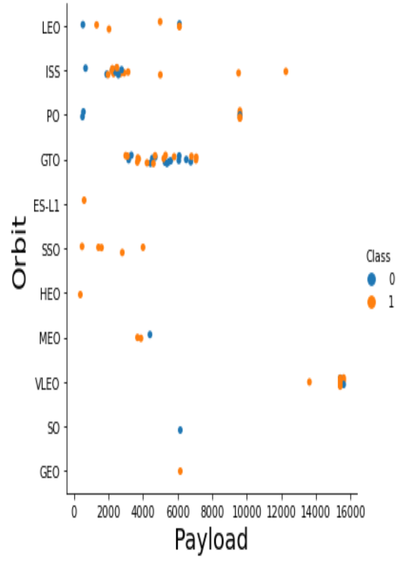
Now try to explain any patterns you found in the Payload Vs. Launch Site scatter point chart.

We can infer that Launch Site-CCAFS SLC 40 is suitable for launching rockets with payload mass varying from low to very high. Launch Site VAFB SLC 4E is preferred for medium payload mass.

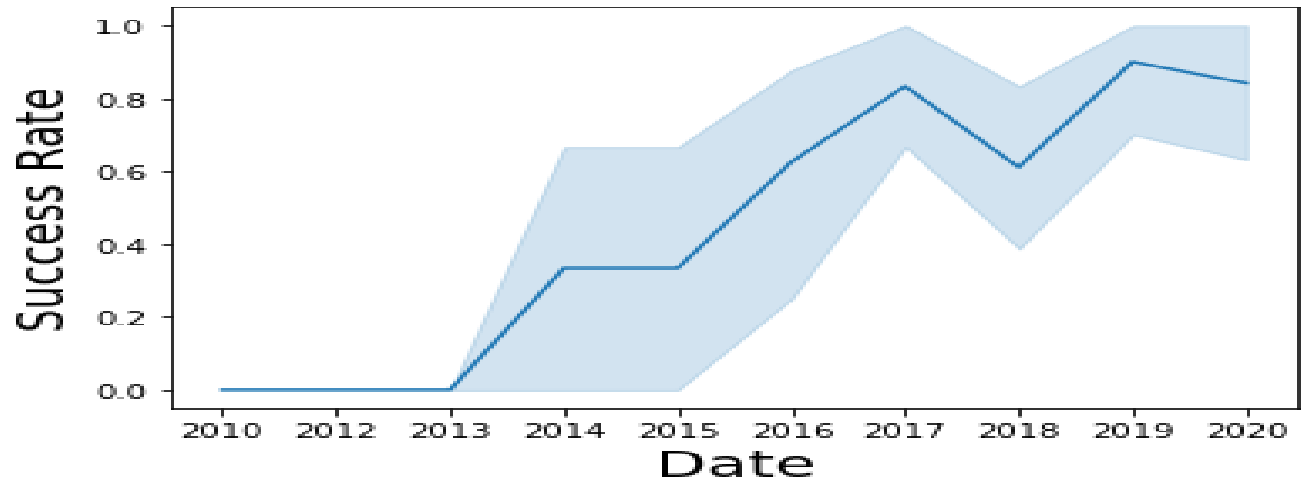




You should see that in the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.



You should observe that Heavy payloads have a negative influence on GTO orbits and positive on GTO and Polar LEO (ISS) orbits.

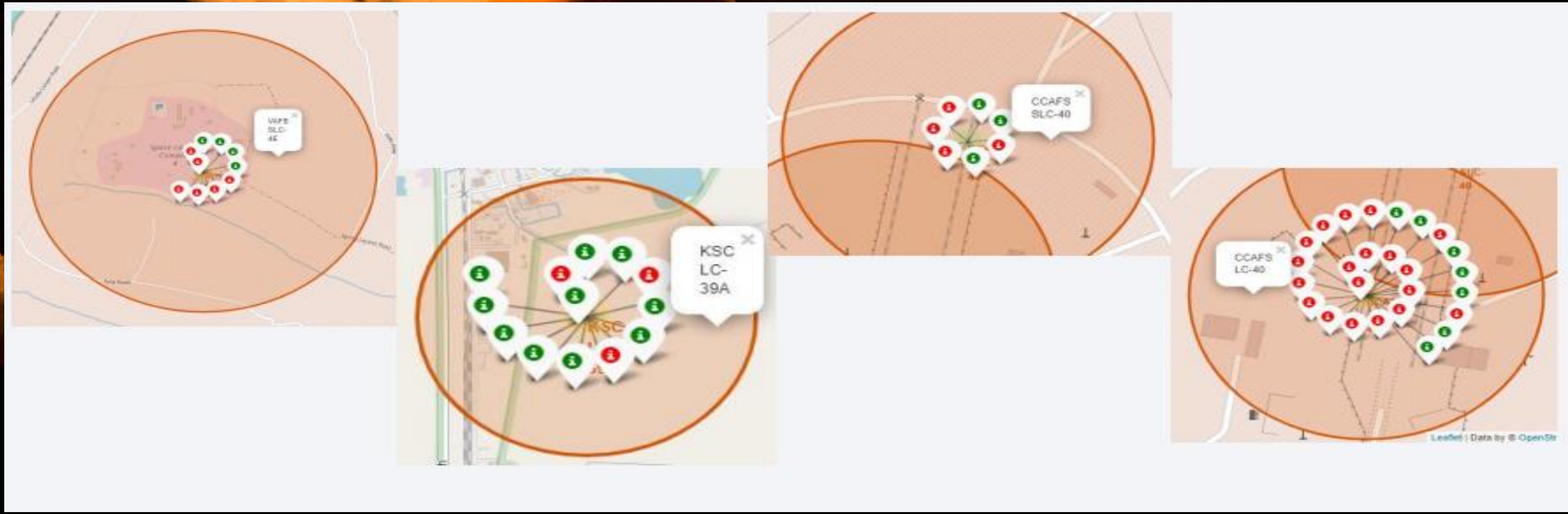


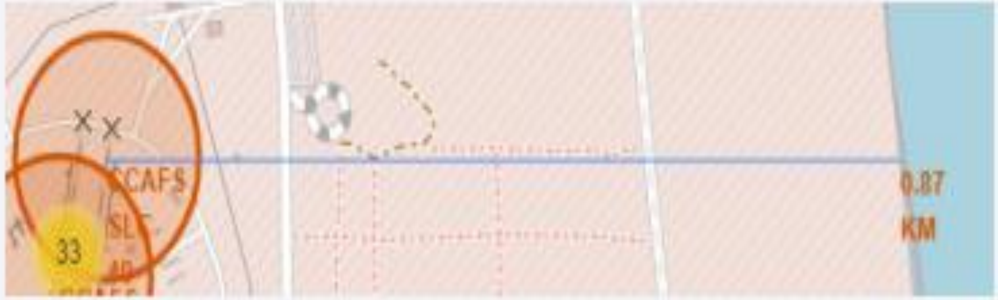
YEAR 2019 have the high no of rocket success rates than other years.

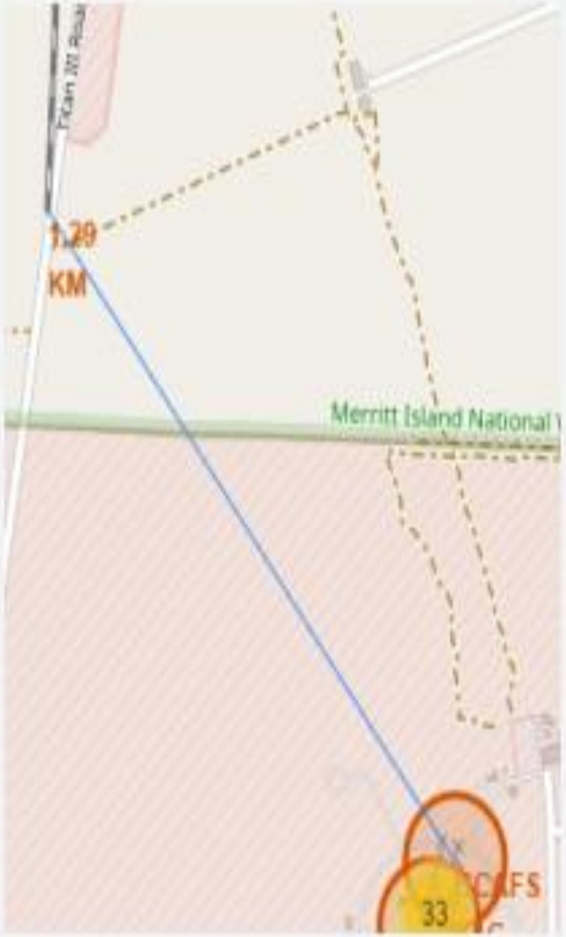
❖ VISUALIZATIONS (FOLIUM MAPS)

Launch Sites

Red : faliures
Green: Success







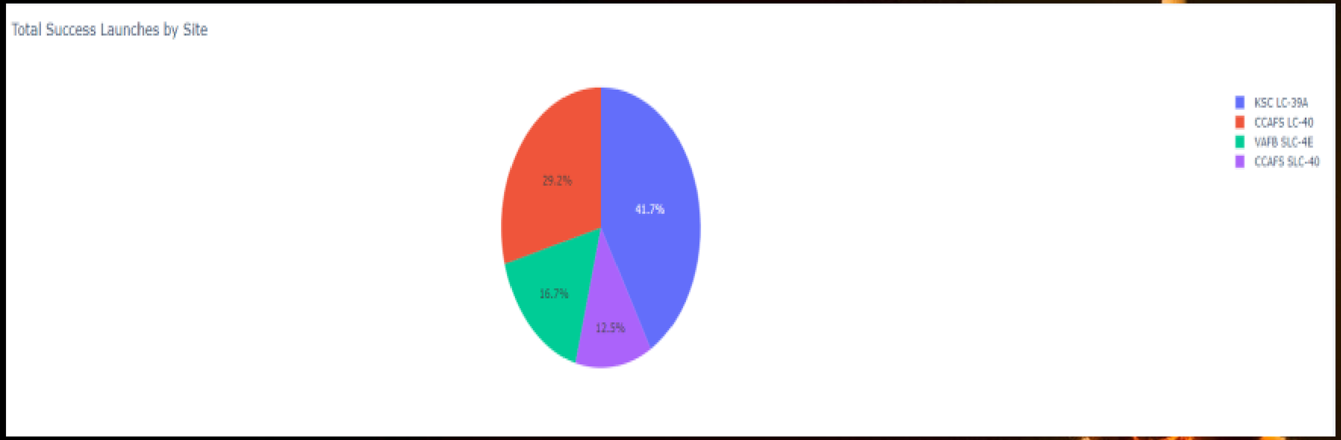
Is CCAFS SLC-40 in close proximity to railways ? Yes

Is CCAFS SLC-40 in close proximity to highways ? Yes

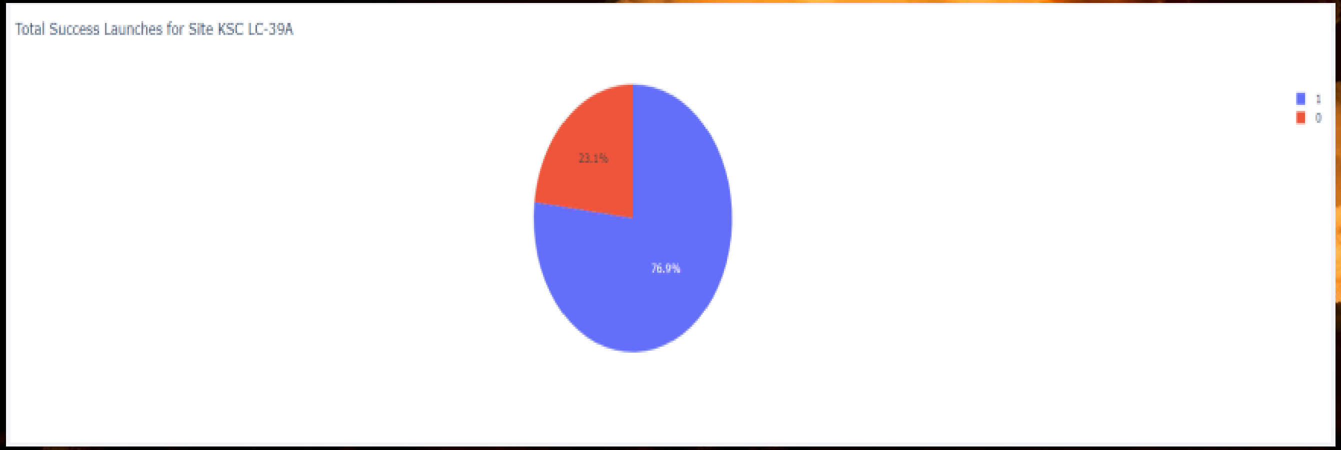
Is CCAFS SLC-40 in close proximity to coastline ? Yes

Do CCAFS SLC-40 keeps certain distance away from cities ? No

❖ PLOTLY:

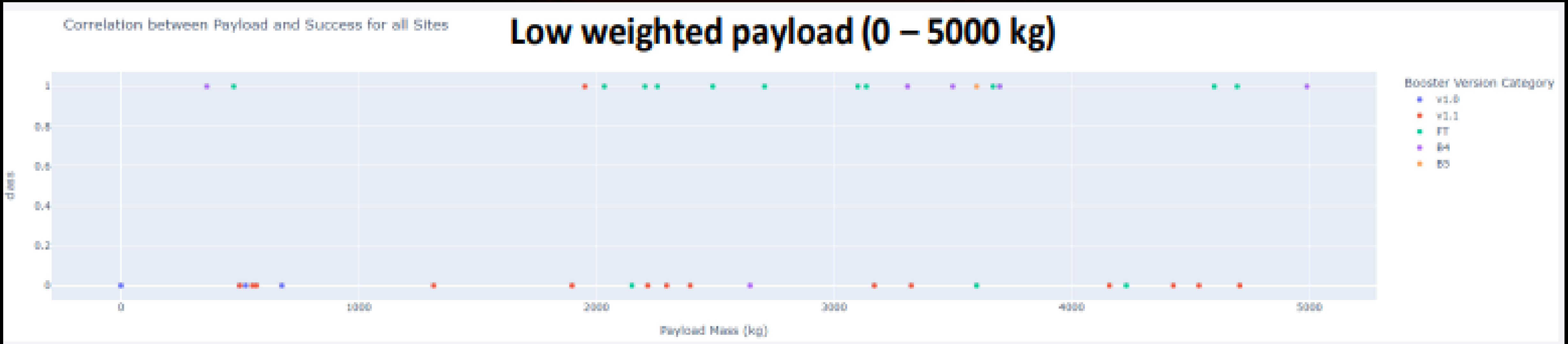


KSC LC-39A has the best success rate of launches.

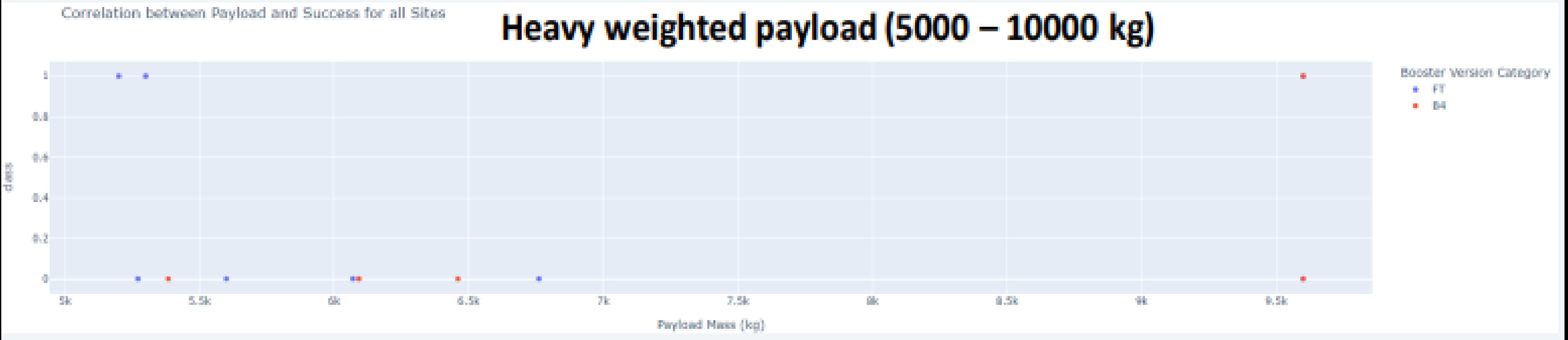


KSC LC-39A has achieved a 76.9% success rate while getting a 23.1% failure rate.

Low weighted payload (0 – 5000 kg)



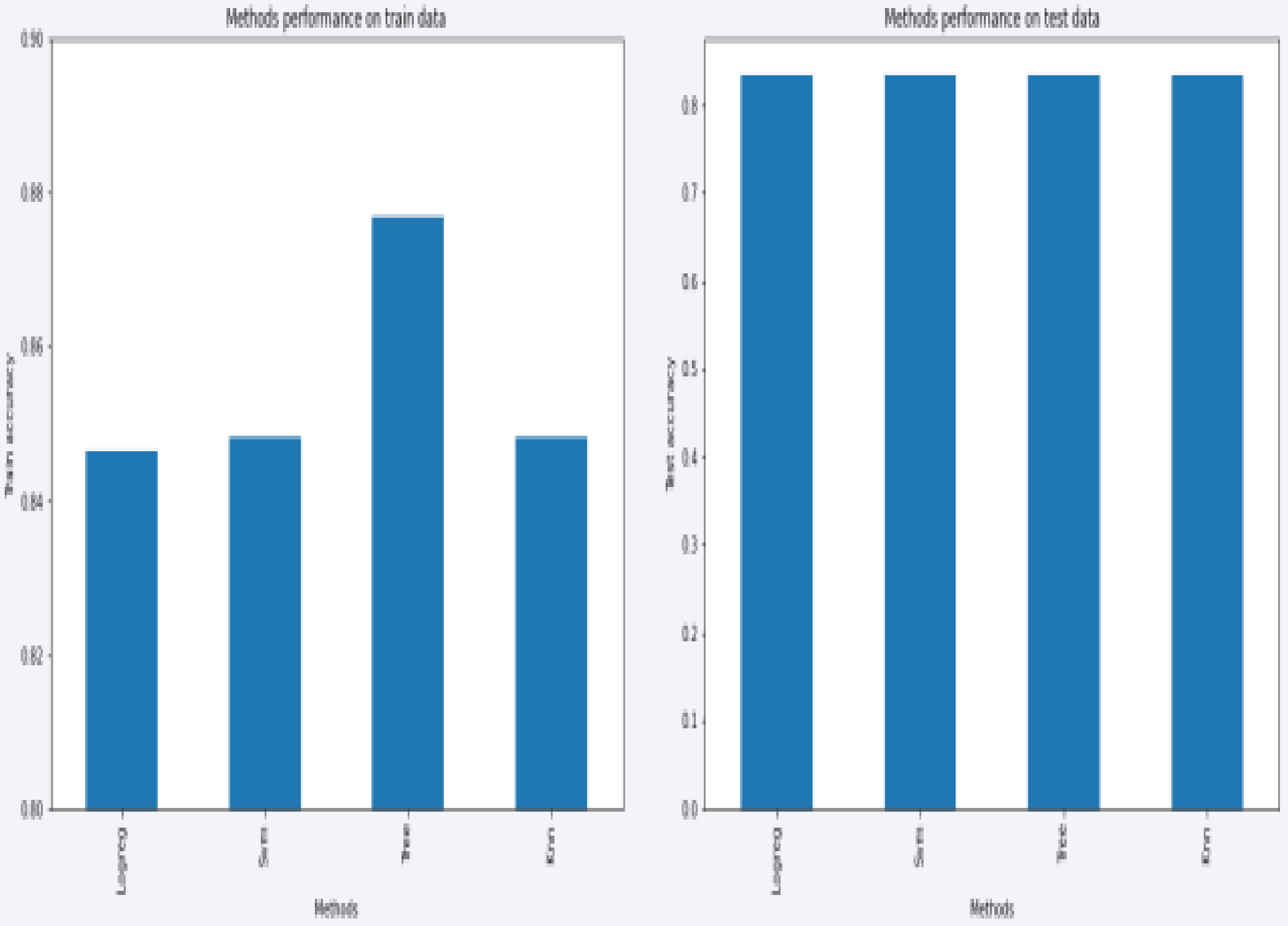
Heavy weighted payload (5000 – 10000 kg)



Low weighted payloads have a better success rate than the heavy weighted payloads.

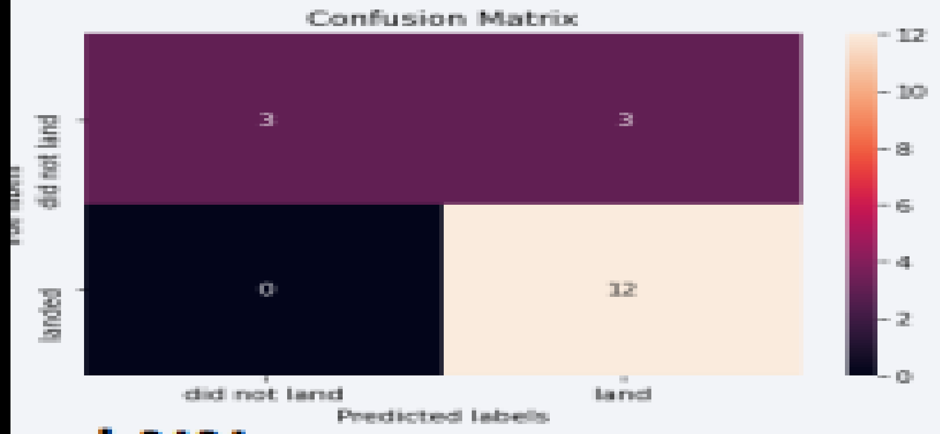
❖ PREDICTIVE ANALYSIS:

	Accuracy Train	Accuracy Test
Tree	0.876786	0.833333
Knn	0.848214	0.833333
Svm	0.848214	0.833333
Logreg	0.846429	0.833333

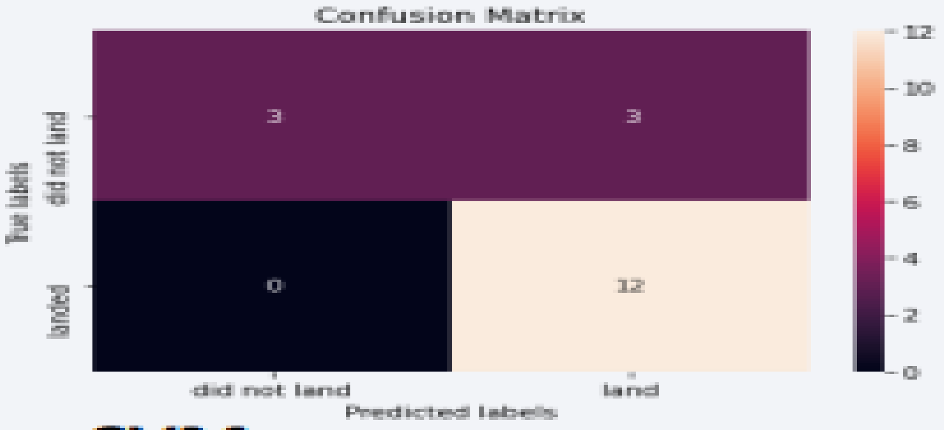


For accuracy test, all methods performed similar. We could get more test data to decide between them. But if we really need to choose one right now, we would take the decision tree.

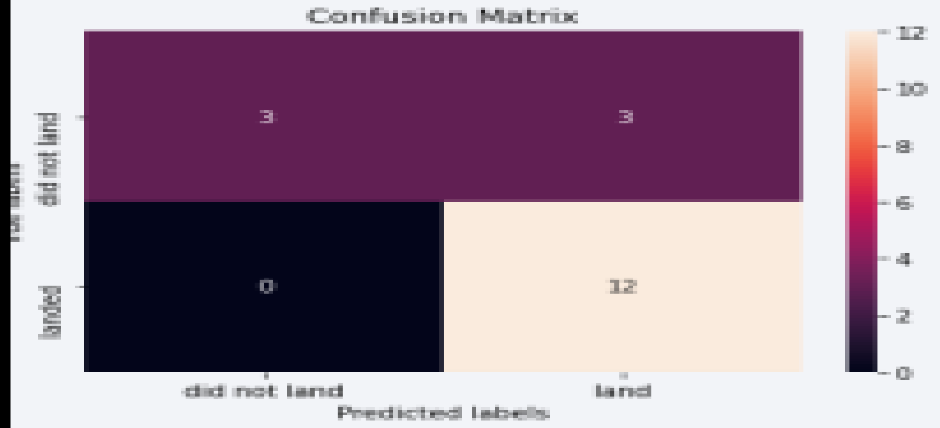
Logistic regression



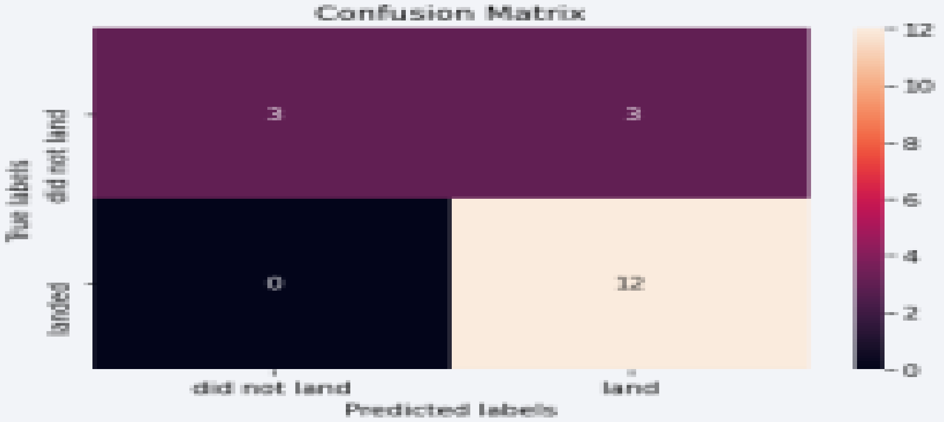
Decision Tree



kNN



SVM



❖ CONCLUSION:

*The success of a mission can be explained by several factors such as the launch site, the orbit and especially the number of previous launches. Indeed, we can assume that there has been a gain in knowledge between launches that allowed to go from a launch failure to a success.

- The orbits with the best success rates are GEO, HEO, SSO, ES-L1.

- Depending on the orbits, the payload mass can be a criterion to take into account for the success of a mission. Some orbits require a light or heavy payload mass. But generally low weighted payloads perform better than the heavy weighted payloads.

- With the current data, we cannot explain why some launch sites are better than others (KSC LC-39A is the best launch site). To get an answer to this problem, we could obtain atmospheric or other relevant data.

- For this dataset, we choose the Decision Tree Algorithm as the best model even if the test accuracy between all the models used is identical. We choose Decision Tree Algorithm because it has a better train accuracy.


THANK YOU!