DATA SCIENCE
CAPSTONE PROJECT

NG WEI HOE
10 August 2021
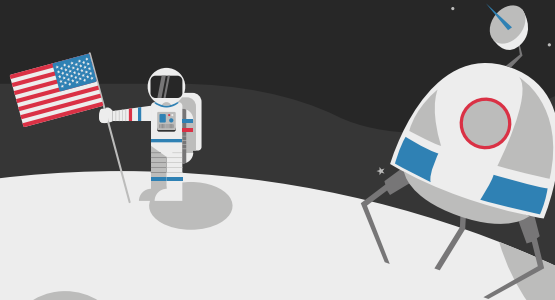
TABLE OF CONTENTS

01 Executive Summary

- Summary of methodologies
- Summary of all result

02 Introduction

- Project Background
- Research Problems

03 Methodology

- Data collection Methodology
- Data Wrangling
- EDA using visualization and SQL
- Interactive Visual Analysis using Folium and Plotly Dash
- Perform predictive analysis using classification models

04 Result

- Exploratory data analysis Results
- Interactive demo (Screen Shots)
- Predictive Analysis Results

05 Conclusion

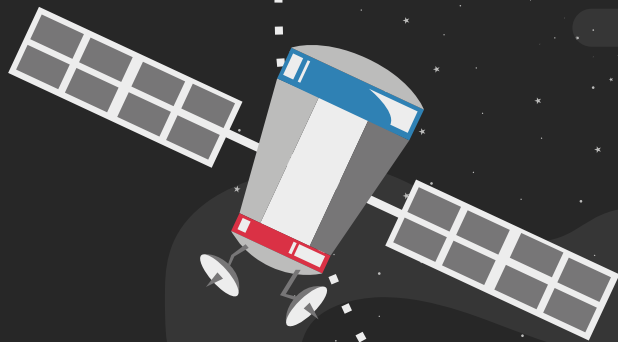- Finding in this project

06 Appendix

- Relevents tools and assets

01

Executive Summary

Executive Summary

## Summary of Methodology and Results

- The collection of data were collected from SpaceX REST API and Web scrape from wiki pages.

- The target object is Falcon 9, it has reached the farthest orbit for 27 times which is GTO orbit. The range of GTO orbit is 35,768 KM above Earth's equator and the overall success rate of landing the booster of Falcon 9 is a 66.66%

- In EDA visualization show that there is a success rate of launching Falcon 9 had increase since 2013 until 2020.

- In EDA with SQl found that Falcon 9 was launched at 3 different launch site, which is CCAFS SLC-40 , KSC LC-39A , VAFB SLC-4E. The total number mission outcomes are 100 success mission and 1 fail mission (in flight). The first success landing is on 01 -05 -2017,

Executive Summary

Summary of Methodology and Results

- In the Folium shows the location of each launch site and the distance to different city and railway.

- In ploty dash show the highest launch record with 46.4% at CCAF SLC-40 and the lowest launch record with 12.5% at CCAFS SLC-40. It also show the booster version B4 had carried a heaviest payload with 9600 KG and also lightest payload with 362 KG had launch successfully

- In the predictive analysis model shows that the prediction method that achieved the best performance is Decision Tree with 89% Accuracy and 0.944 Score.
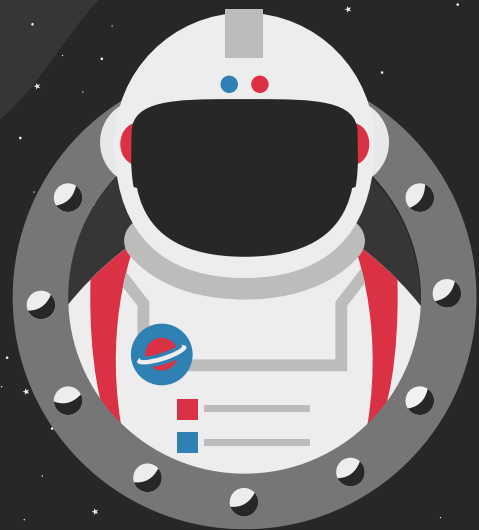
INTRODUCTION

Background

The commercial space age is here, companies are making space travel affordable for everyone. Companies like Virgin Galactic who providing suborbital space flights, Rocketlab who provide stattlite. Perhaps the most successful is SpaceX. SpaceX had accomplished sending spacecraft to the international space station. The rocket launches for SpaceX is inexpensive. 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.
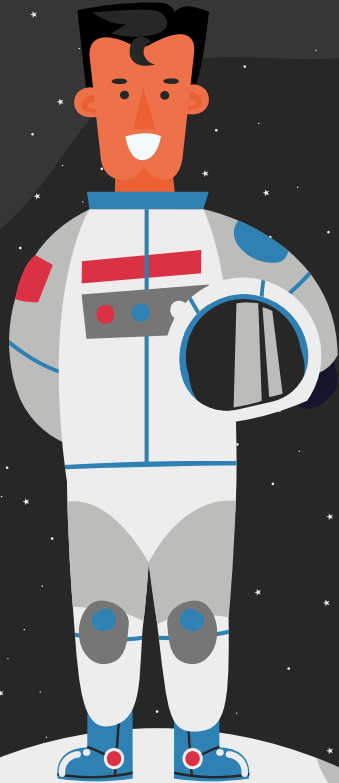
Problem

Determine successful landing booster of falcon 9.

Determine which launch site have a high success rate

Determine the relationship of the weight of the payload and the version of booster carried by falcon 9 will causes the landing

03

Methodology

DATA COLLECTION

SPACEX REST API

The following picture show how to call SpaceX API


```
[7]: spacex_url="https://api.spacexdata.com/v4/launches/past"

[8]: response = requests.get(spacex_url)
```

Check the content of the response

After request the API, decode the response using Json() and turn it into Pandas dataframe and now we have all data that provided by SpaceX


```
# Use json_normalize meethod to convert the json result into a dataframe
web_data = response.json()
web_data

data =pd.json_normalize(web_data)
data
```

	static_fire_date_utc	static_fire_date_unix	tbd	net	window	rocket	success	details	ships	capsules	payloads	launchpad	auto_update	launch_library_id	failures
0	2006-03-17T00:00:00.000Z	1.142554e+09	False	False	0.0	5e9d0d95eda69955f709d1eb	False	Engine failure at 33 seconds and loss of vehicle	[]	[]	[5eb0e4b5b6c3bb0006eeb1e1]	5e9e4502f5090995de566f86	True	None	[{'time': 33, 'altitude': None, 'reason': 'merlin engine failure'}]

Successful first

DATA COLLECTION

WEB SCRAPING

The following picture show how to web scrape from Wiki Page

```
[4]: static_url = "https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922"
response = requests.get(static_url)
response.status_code
```

After request the API, filter out the data that show in the HTML table, then append them to a empty list separately. Finally make the list into the data frame with different category

```
df=pd.DataFrame(launch_dict)
print(df)
```

	Flight No.	Launch site	Payload	Payload mass	\
0	1	CCAFS	Dragon Spacecraft Qualification Unit	0	
1	2	CCAFS	Dragon	0	
2	3	CCAFS	Dragon	525 kg	
3	4	CCAFS	SpaceX CRS-1	4,700 kg	
4	5	CCAFS	SpaceX CRS-2	4,877 kg	

	Orbit	Customer	Launch outcome	Version Booster	Booster landing	\
0	LEO	[SpaceX]	Success\n	F9 v1.0B0003.1	Failure	
1	LEO	[NASA]	Success	F9 v1.0B0004.1	Failure	
2	LEO	[NASA]	Success	F9 v1.0B0005.1	No attempt\n	

	Date	Time
0	4 June 2010	18:45
1	8 December 2010	15:43
2	22 May 2012	07:44

DATA WRANGLING

NUMBER OF LAUNCH SITE

There are three main Launch Site that launch Falcon 9

```
# Apply value_counts() on column LaunchSite
df['LaunchSite'].value_counts()

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

CCAFS SLC 40 : Cape Canaveral Space Launch Complex 40

KSC LC 39A : Kennedy Space Center Launch Complex 39A

VAFB SLC 4E :Vandenberg Air Force Base Space Launch Complex 4E (SLC-4E)

ORBIT REACHED

```
# Apply value_counts on Orbit column
df['Orbit'].value_counts()

GTO     27
ISS     21
VLEO    14
PO       9
LEO      7
SSO      5
MEO      3
HEO      1
SO       1
ES-L1    1
GEO      1
```

The data show that Falcon 9 has reach 11 different type orbit

Falcon 9 successfully reach GTO for 27 time

The range for GTO is 35,786 KM above Earth's equator

DATA WRANGLING

ASSIGN 0 AND 1 TO FAIL LANDING AND SUCCESS LANDING

```python
# landing_class = 0 if bad_outcome
# landing_class = 1 otherwise
landing_class=[]

for n in df["Outcome"]:
    if n in bad_outcomes:
        landing_class.append(0)
    else:
        landing_class.append(1)

print(landing_class)
```

```
[0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 1, 1, 1, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 0, 0, 1,
1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]
```
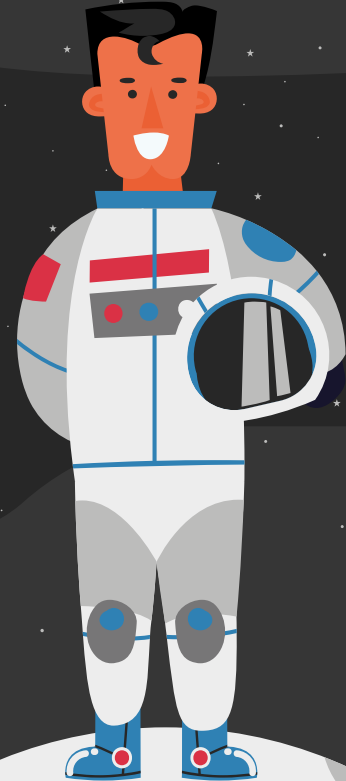
FAIL : 0   SUCCESS: 1

OVERALL SUCCESS RATE FOR LANDING

```python
df["Class"].mean()
```

```
0.6666666666666666
```

The Success Rate is a 66.66%

"That's one small step for a man, one giant leap for mankind"

–Neil Armstrong
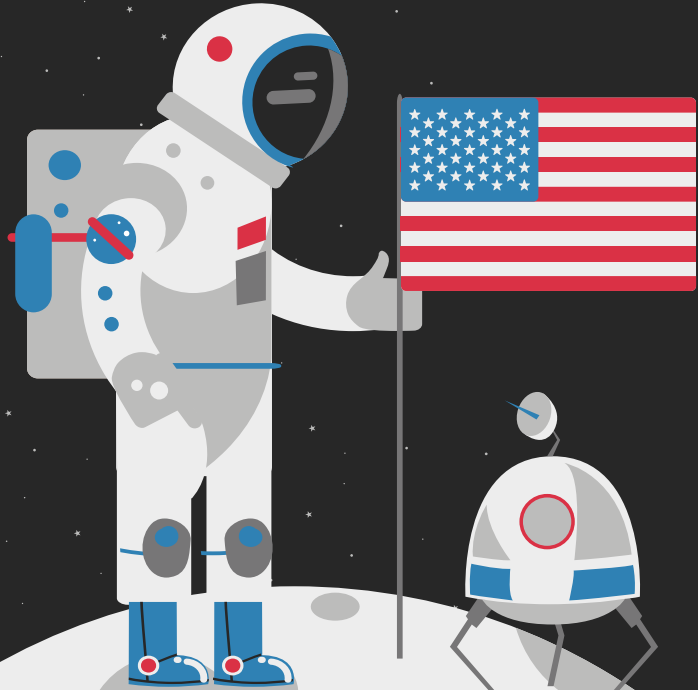
EXPLORATORY DATA ANALYSIS WITH VISUALIZATION

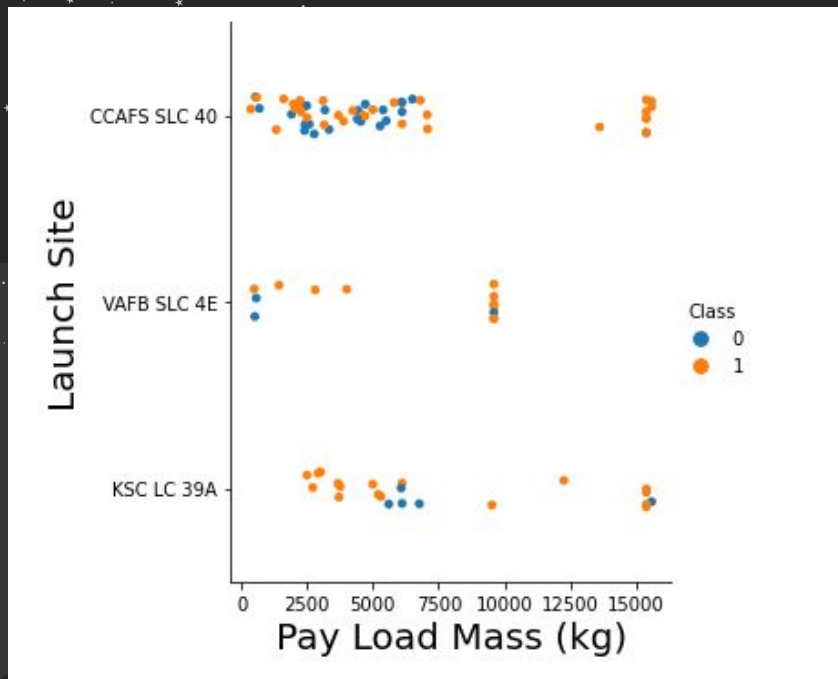# EDA with VISUALIZATION

FLIGHT NUMBER VS LAUNCH SITE
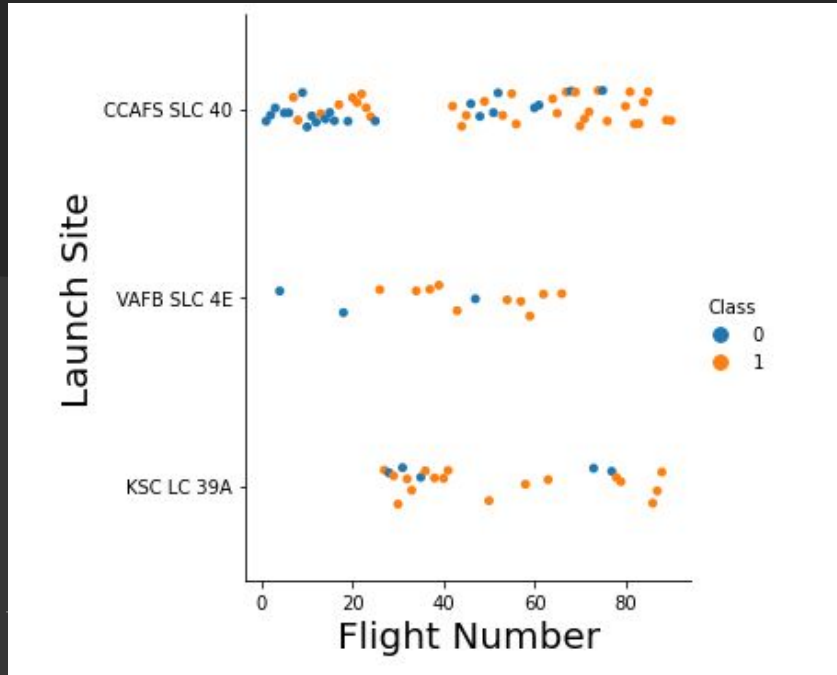
In the diagram show that the flight number increasing, the success launch rate is more higher

There had a most flight number at CCAFS SLC 40 launch site and the least flight number at VAFB SLC 4E.

There is fail mission that is below 20 flight and there is a success mission with more than 20 flight

EDA with VISUALIZATION

PAYLOAD VS LAUNCH SITE

CCAFS LC-40, has a success rate of 60%; but if the mass is above 10,000 kg the success rate is 100%

VAFB SLC 4E and KSC LC 39A has the success rate of 100% between 2500 kg and 5000 kg
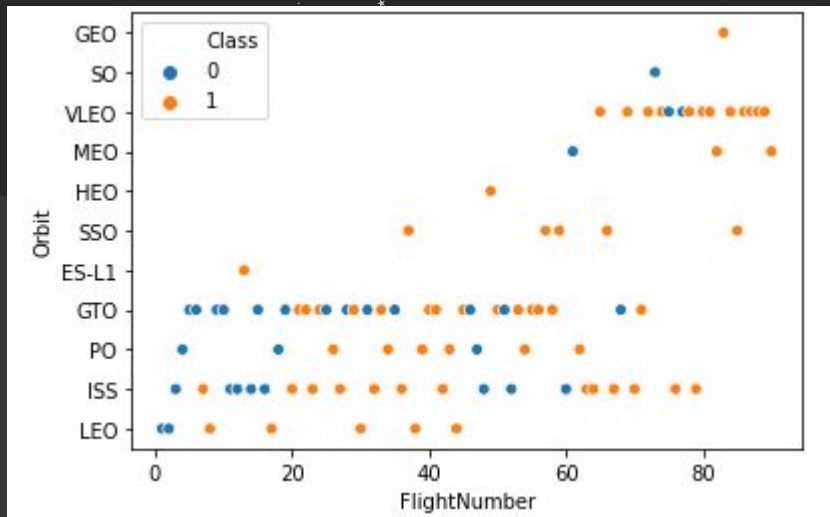
EDA with VISUALIZATION

SUCCESS RATE VS ORBIT TYPE
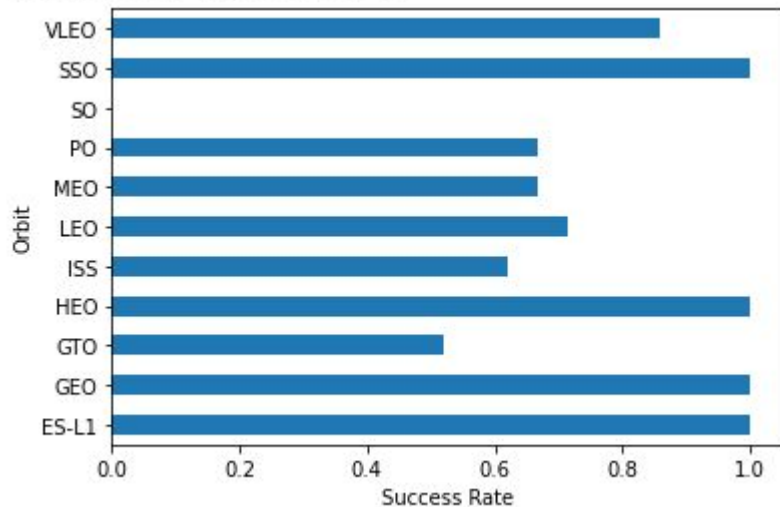
There 100% of success rate for reaching out the orbit is SSO orbit, HEO orbit, GEO orbit and ES-L1 orbit.

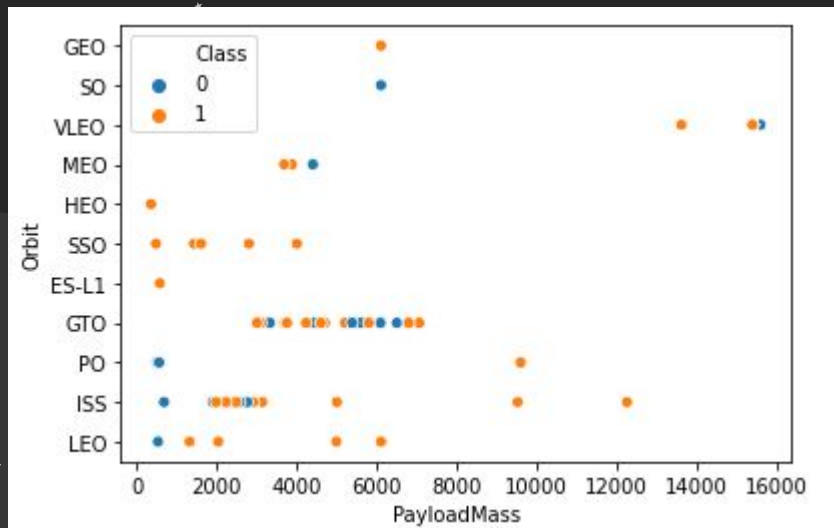# EDA with VISUALIZATION

FLIGHT NUMBER VS ORBIT TYPE

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.
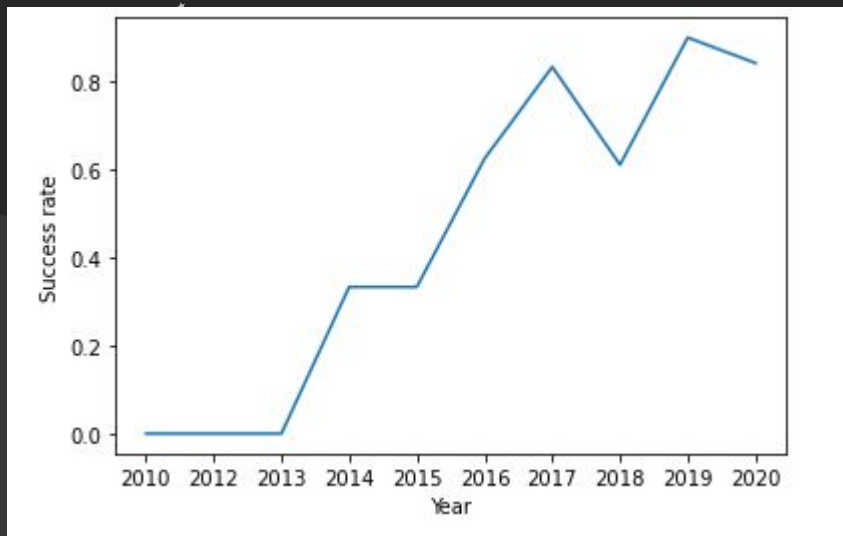
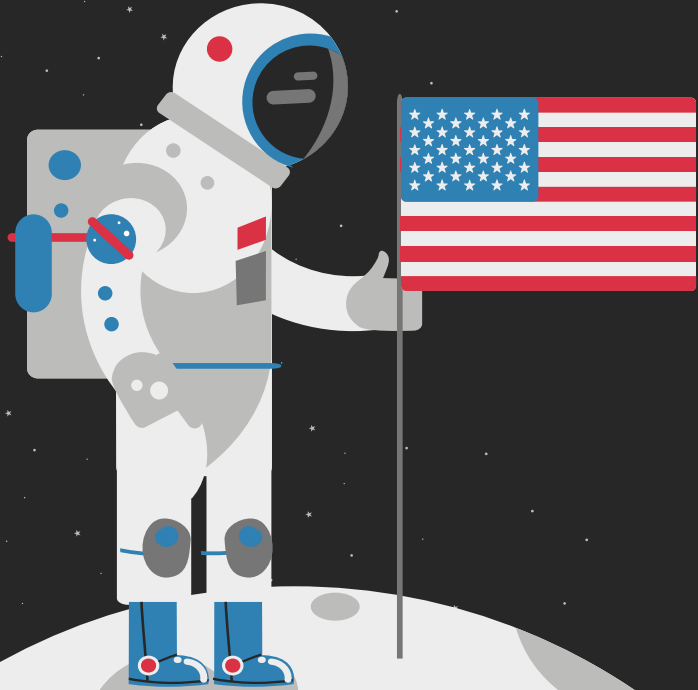# EDA with VISUALIZATION

PAYLOAD VS ORBIT TYPE

Heavy payloads have a negative influence on GTO orbits

Positive influence on GEO , Polar LEO (ISS) orbits.

EDA with VISUALIZATION

LAUNCH SUCCESS YEARLY TREND

There is a increasing success rate of launching Falcon 9 since 2013 to 2020 .

EXPLORATORY DATA ANALYSIS WITH SQL

EDA WITH SQL

ALL LAUNCH SITE NAME


```
[5]: display_uniq_name = %sql SELECT  DISTINCT LAUNCH_SITE FROM SPACEXTBL
     display_uniq_name

      * ibm_db_sa://wtj78002:***@dashdb-txn-sbox-yp-dal09-11.services.dal.bl
     Done.
[5]:    launch_site

      CCAFS LC-40

      CCAFS SLC-40

      CCAFSSLC-40

      KSC LC-39A

      VAFB SLC-4E
```

Falcon 9 only launch at these 3 different launch site

EDA WITH SQL

LAUNCH SITE NAME BEGIN WITH 'CCA'

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

```
[17]: %sql SELECT DISTINCT LAUNCH_SITE FROM SPACEXTBL WHERE LAUNCH_SITE LIKE '%CCA%' LIMIT 5
```

 * ibm_db_sa://wtj78002:***@dashdb-txn-sbox-yp-dal09-11.services.dal.bluemix.net:50000/BLUD
Done.

[17]:

launch_site
CCAFS LC-40
CCAFS SLC-40
CCAFSSLC-40

This 3 are the launch site name begin with 'CCA'

EDA WITH SQL

TOTAL PAYLOAD MASS LAUNCH BY NASA

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

```
[7]: %sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE CUSTOMER LIKE 'NASA (CRS)'
```

 * ibm_db_sa://wtj78002:***@dashdb-txn-sbox-yp-dal09-11.services.dal.bluemix.net:50000/BLUDB
Done.

```
[7]:      1

      45596
```

45596 KG is the total payload mass that had been launch by NASA

EDA WITH SQL

AVERAGE PAYLOAD MASS BY F9 v1.1

```
[8]: %sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE BOOSTER_VERSION LIKE 'F9 v1.1'

 * ibm_db_sa://wtj78002:***@dashdb-txn-sbox-yp-dal09-11.services.dal.bluemix.net:50000/BLUDB
Done.

[8]:         1

    2928.400000
```

2928.40 KG is the average Payload Mass (KG) for booster F9 V1.1

EDA WITH SQL

FİRST SUCCESSFUL GROUND LANDING DATE


```
[9]: %sql SELECT MIN(DATE) FROM SPACEXTBL WHERE LANDING__OUTCOME ='Success (ground pad)'

 * ibm_db_sa://wtj78002:***@dashdb-txn-sbox-yp-dal09-11.services.dal.bluemix.net:50000/BLUDB
Done.

[9]:        1

     01-05-2017
```

1st of May 2017 is the first successful ground lading date

EDA WITH SQL

SUCCESSFUL DRONE SHIP LANDING WITH PAYLOAD BETWEEN 4000 AND 6000


```
[10]: %sql SELECT BOOSTER_VERSION FROM SPACEXTBL WHERE LANDING__OUTCOME = 'Success (drone ship)' AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000
```

 * ibm_db_sa://wtj78002:***@dashdb-txn-sbox-yp-dal09-11.services.dal.bluemix.net:50000/BLUDB
Done.

[10]: **booster_version**

F9 FT B1022

F9 FT B1026

F9 FT B1021.2

F9 FT B1031.2

There are 3 type of different version booster that successful landed on drone ship with payload mass between 4000 KG and 6000 KG

EDA WITH SQL

TOTAL NUMBER OF SUCCESSFUL AND FAILURE MISSION OUTCOMES

List the total number of successful and failure mission outcomes

```
[11]: %sql SELECT SUM(MISSION_OUTCOME ='Success') as Success , SUM(MISSION_OUTCOME ='Success (payload status unclear)') as Success_Payload_status_unclear,SUM(MISSION_OUTCOME = 'Failure (in flight)') as Failure FROM SPACEXTBL
      #WHERE MISSION_OUTCOME LIKE 'Success%' OR MISSION_OUTCOME LIKE 'Failure%' GROUP BY MISSION_OUTCOME


      #%sql SELECT DISTINCT MISSION_OUTCOME as Failure  FROM SPACEXTBL
       * ibm_db_sa://wtj78002:***@dashdb-txn-sbox-yp-dal09-11.services.dal.bluemix.net:50000/BLUDB
       Done.
```

```
[11]: success  success_payload_status_unclear  failure
         99                               1       1
```

There are total of 100 success of landing , 99 success landing with stated payload mass and 1 success landing without payload status unclear

There is only 1 failure for landing

EDA WITH SQL

BOOSTERS CARRIED MAXIMUM PAYLOAD

```
%sql SELECT BOOSTER_VERSION FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL) GROUP BY BOOSTER_VERSION
```

 * ibm_db_sa://wtj78002:***@dashdb-txn-sbox-yp-dal09-11.services.dal.bluemix.net:50000/BLUDB
Done.

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

These are the booster version which have carried the maximum payload mass

EDA WITH SQL

2015 LAUNCH RECORDS

```sql
%sql SELECT DATE,MONTHNAME(DATE) AS MONTH,LANDING__OUTCOME,BOOSTER_VERSION,LAUNCH_SITE FROM SPACEXTBL WHERE LANDING__OUTCOME='Failure (drone ship)' AND YEAR(DATE)= 2015
```

* ibm_db_sa://rbn61462:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.

DATE	MONTH	landing__outcome	booster_version	launch_site
2015-01-10	January	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
2015-04-14	April	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

These are the 2 fail landing at drone ship with booster version of F9 v1.1 B10102 and F9 v1.1 B1015 at CCAFS LC -40 on January and April 2015

EDA WITH SQL

RANK SUCCESS COUNT BETWEEN 2010-06-4 and 2017-03-20

```sql
%sql select landing__outcome, count(*) as count from SPACEXTBL where landing__outcome like '%Success%' and DATE between '2010-06-04' and '2017-03-20' group by landing__outcome order by count(*) de
```

* ibm_db_sa://rbn61462:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.

landing__outcome	COUNT
Success (drone ship)	5
Success (ground pad)	3

These are the 3 success landing at ground pad and 5 success landing at drone ship between 06-04-2010 and 20-03-2017

INTERACTIVE MAP
WITH FOLIUM

FOLIUM MAP WITH MARKER

VAFB SLC- 4E

KSC LC 39A | CCAFS SLC -40 | CCAFS LC-40
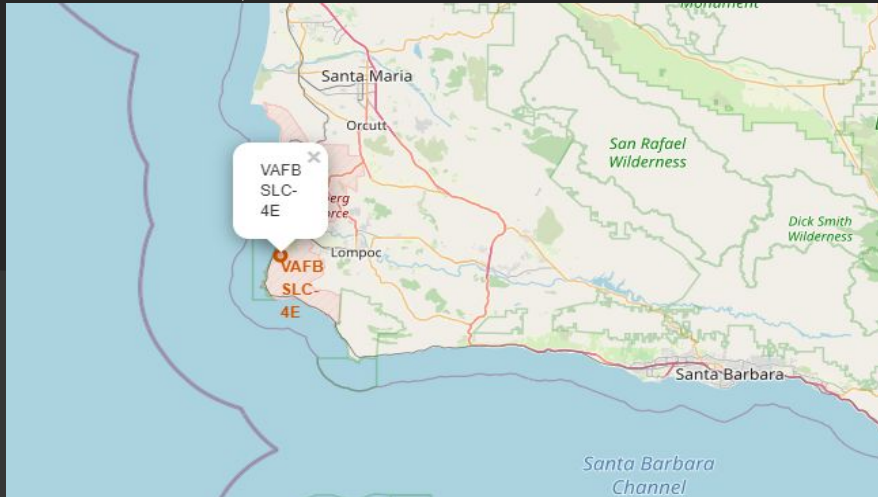
On the left side we have VAFB SLC -4E launch where located near Lompoc airport
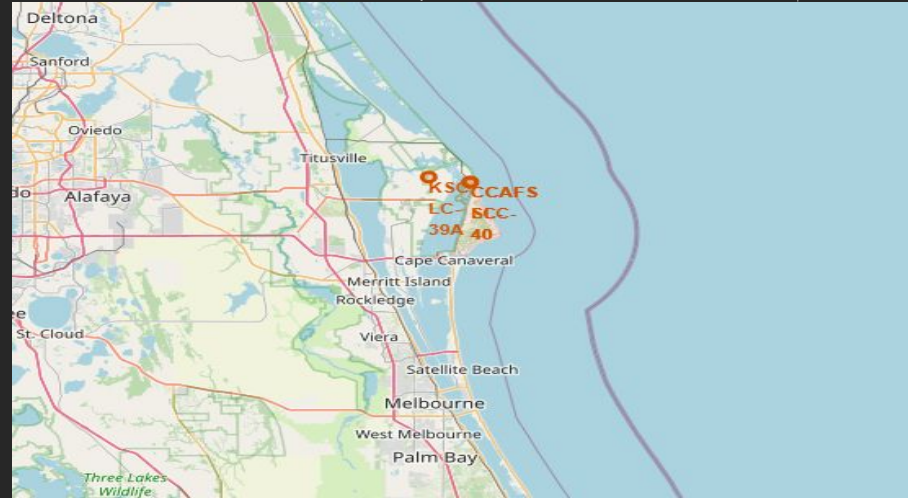
On the right side we have 3 different launch site which is KSC LC 39 A , CCAFS SLC-40 and CCAFS LC-40

FOLIUM MAP WITH COLOR LABEL
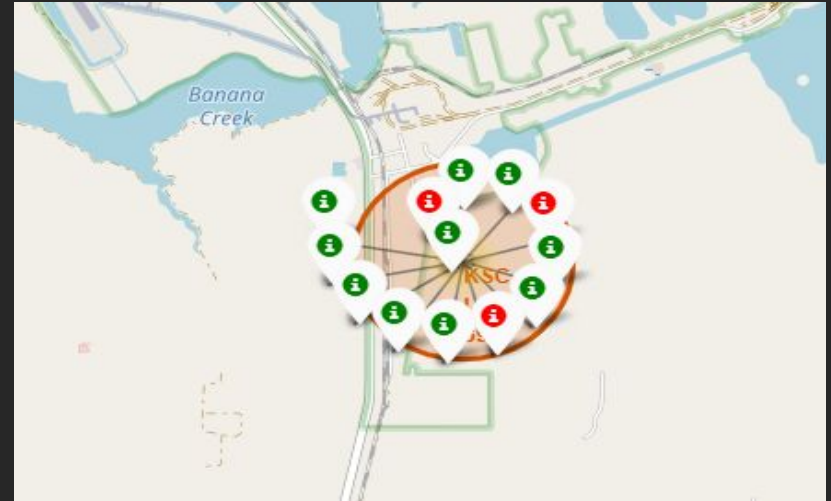
Green - Success Mission
Red - Fail Mission

VAFB SLC- 4E
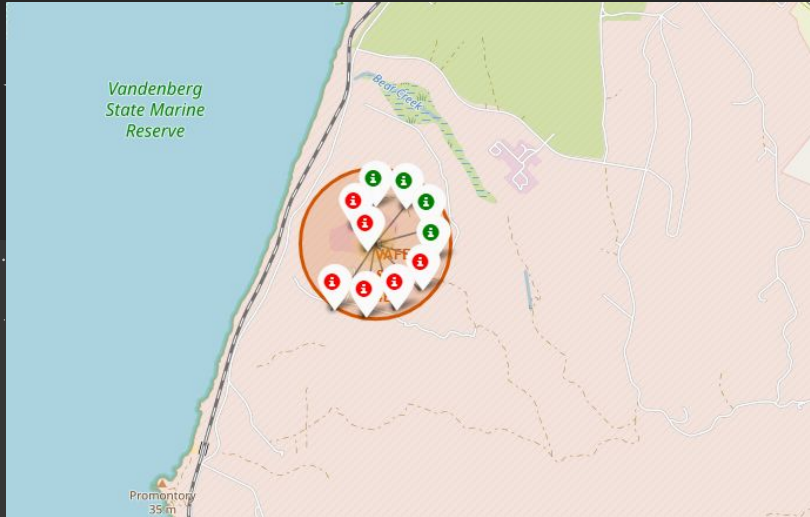
KSC LC 39A

At VAFB SLC -4E launch site have total number of 10 launch mission, 4 had a successful landing and 6 had a fail mission

At KSC LC 39A launch site have total number of 13 launch mission, 10 had a successful landing and 3 had a fail mission

FOLIUM MAP WITH COLOR LABEL

Green - Success Mission
Red - Fail Mission

CCAFS SLC -40

CCAFS LC - 40

At CCAFS SLC -40 launch site have total number of 7 launch mission, 3 had a successful landing and 4 had a fail mission
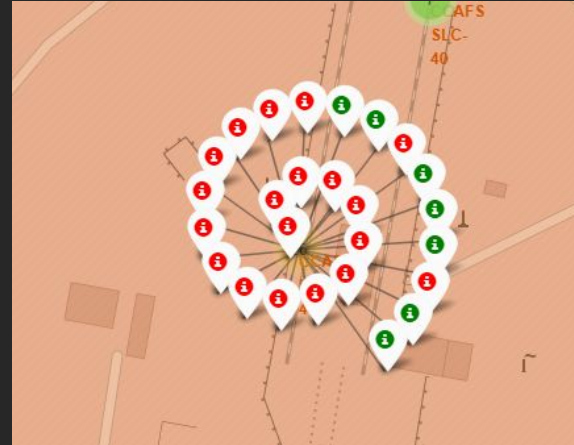
At KSC LC 39A launch site have total number of 26 launch mission, 7 had a successful landing and 9 had a fail mission

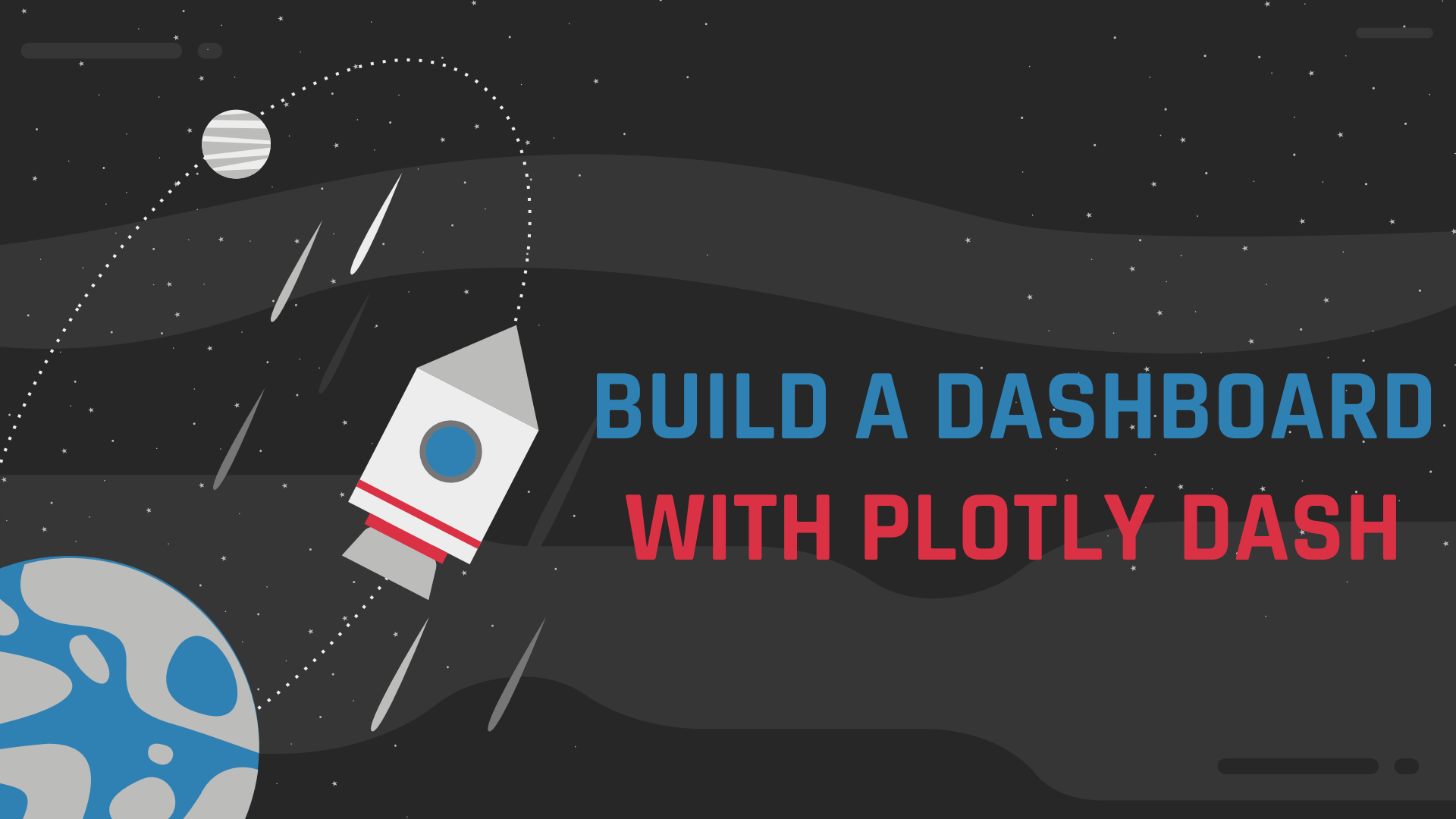# FOLIUM MAP DISTANCE CALCULATION

VAFB SLC4E TO RAILWAY
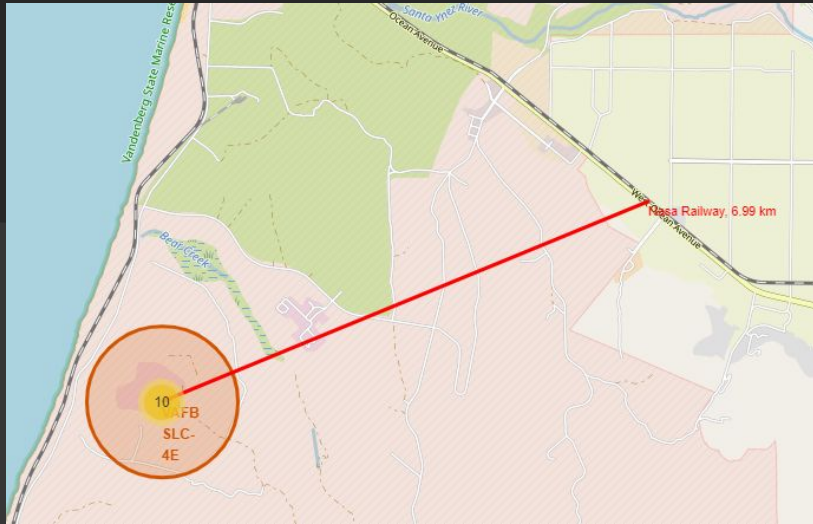
The Distance from launch site VAFB SLC4E to railway is 6.99KM

VAFB SLC4E TO SANTA MARIA & SANTA BARBARA
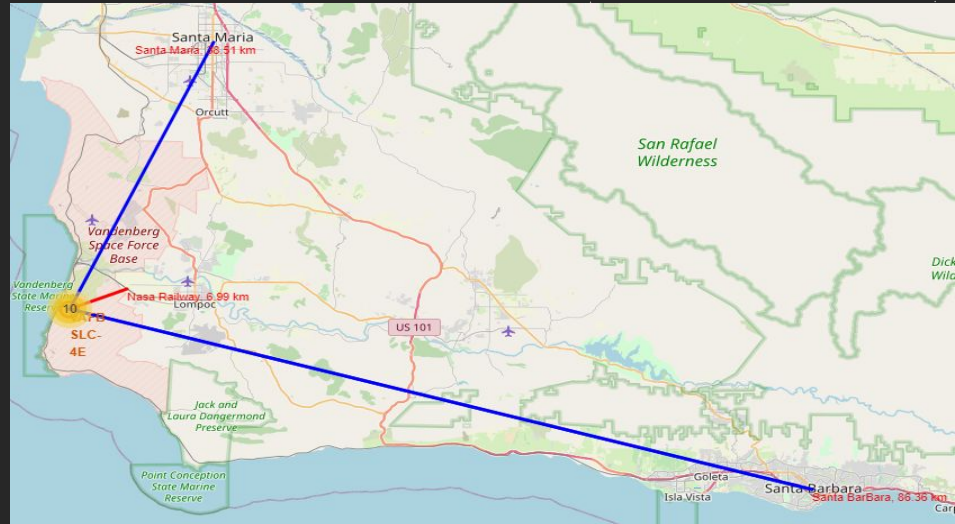
The Distance from launch site VAFB SLC4E to Santa Maria and Santa Barbara are 38.51KM and 86.83 KM
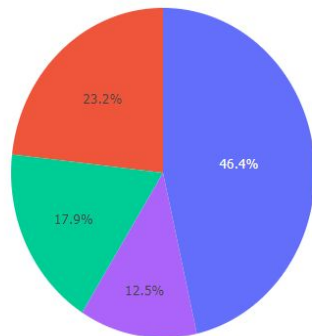
DASHBOARD

ALL LAUNCH SITE RATIO

SpaceX Launch Records Dashboard

All Site

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

46.4%
23.2%
17.9%
12.5%

The pie chart shows the percentage that SpaceX launch at different launch site.

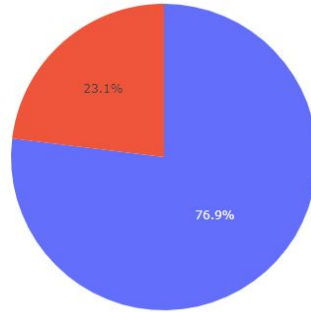CCAFS LC 40 - 46.4 %   VAFB SLC -4E - 17.9%

KSC LC-39A - 23.3%   CCAFS SLC-40 - 12.5%

DASHBOARD

LAUNCH SITE WITH THE HIGHEST LAUNCH SUCCESS RATIO

SpaceX Launch Records Dashboard

KSC LC-39A

23.1%

76.9%

■ 1
■ 0

The pie chart shows KSC LC-39A has a high launch success ratio with 76.9 % and 23.1% launch fail ratio
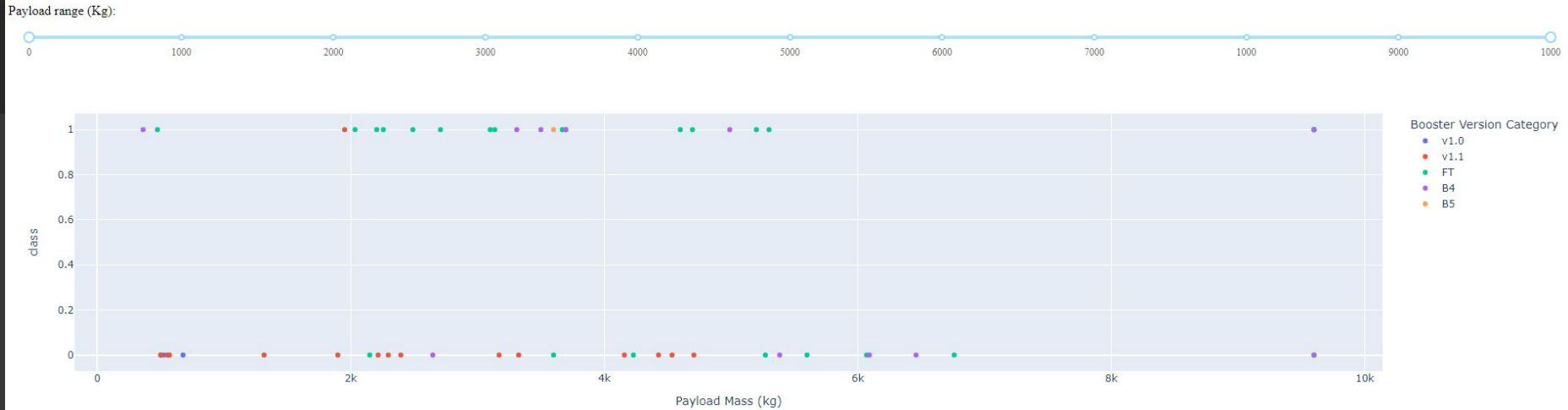
Class 1 present success (Purple)

Class 0 present failure (Red)

DASHBOARD
SELECT DIFFERENT PAYLOAD IN RANGE SLIDER
## [BEFORE SELECT PAYLOAD MASS]

Below Diagram show a scatter plot with All launch site, Success rate with different Payload Mass and different Booster Version

DASHBOARD

SELECT DIFFERENT PAYLOAD IN RANGE SLIDER
[AFTER SELECT PAYLOAD MASS]

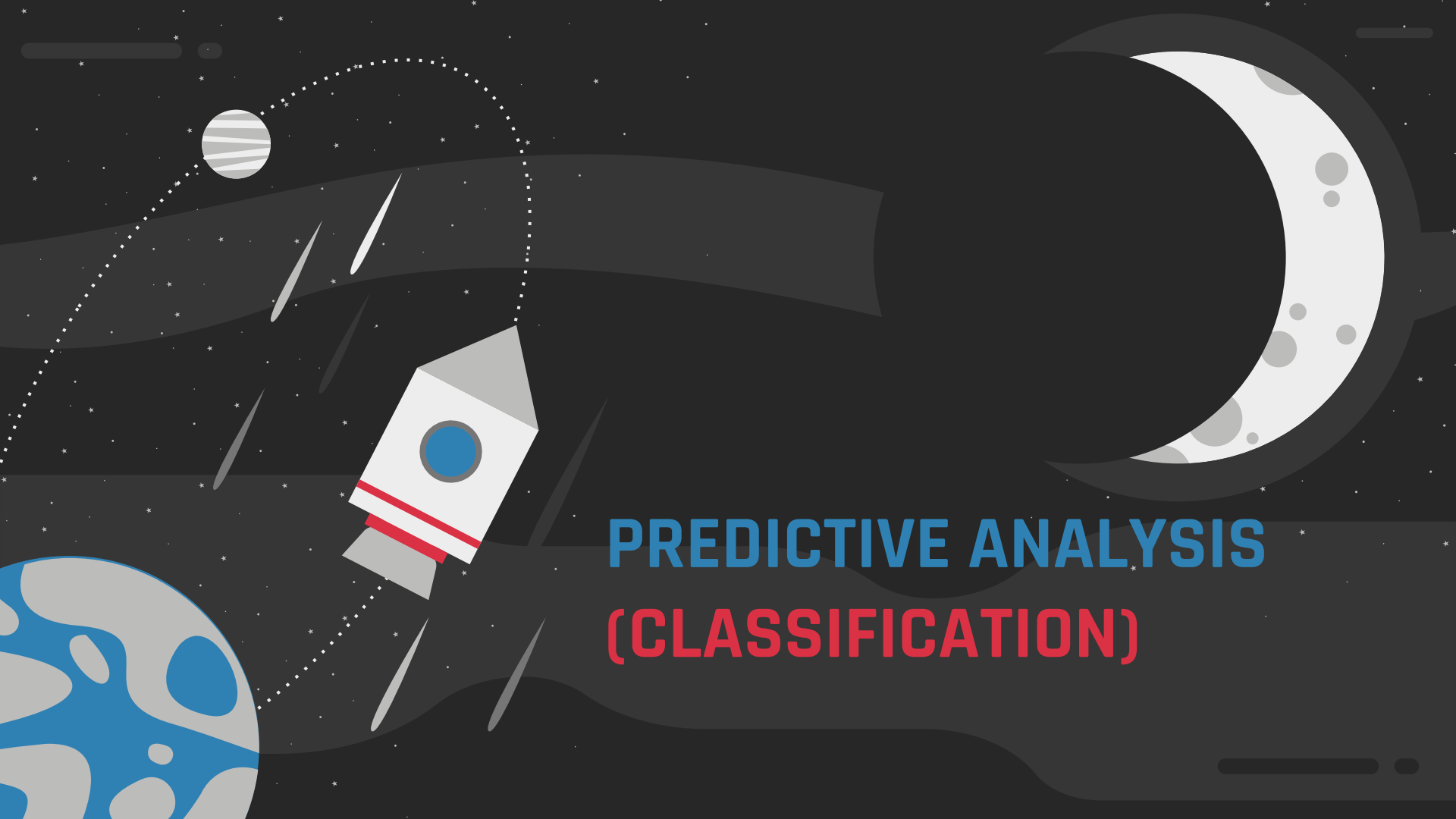Below Diagram show a scatter plot with All launch site, Success rate with different Payload Mass and different Booster Version. It only show the results of success rate with a payload mass between 2000 kg and 7000 kg
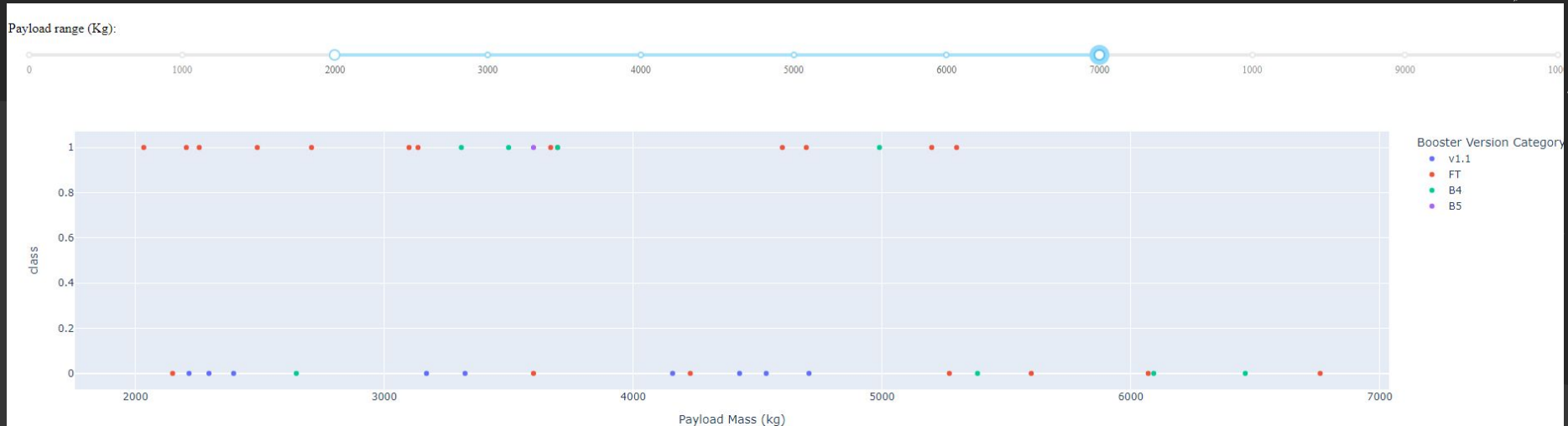
PREDICTIVE ANALYSIS (CLASSIFICATION)

CLASSIFICATION MODEL

CLASSIFICATION ACCURACY

Right side is showing the different type of classification model will be testing the data

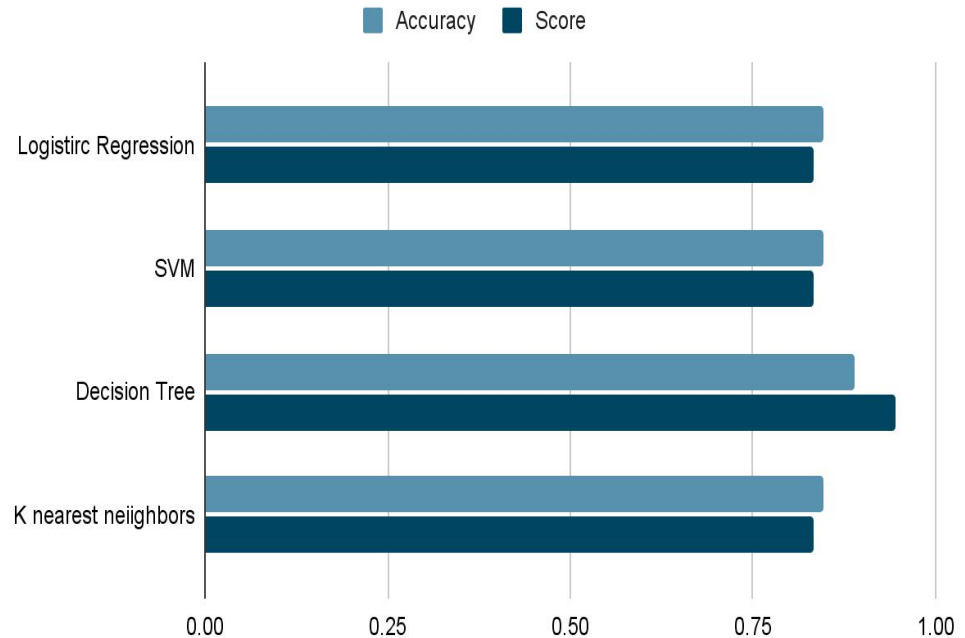Logistic regression has the accuracy of 84.72 % and 0.8333 score

Support Vector Machine has the accuracy of 84.72 and 0.8333 score

Decision Tree Classifier has the accuracy of 89% and 0.9444score

K nearest neighbors has the accuracy of 84.72% and 0.8333 score

So, Decision Tree Classifier has the best performance model

Classification Model

CONFUSION MATRIX

Right side is showing the confusion matrix of Decision Tree Classifier

Decision Tree Classifier is the best performing model

Confusion Matrix is showing that it has accurately predict 15 out of 18 of the launch mission.

It correctly classified of 3 of the rocket did not land and 12 of the rocket had landed

The algorithm misclassified that 3 of the rocket had landed by saying that the did not land
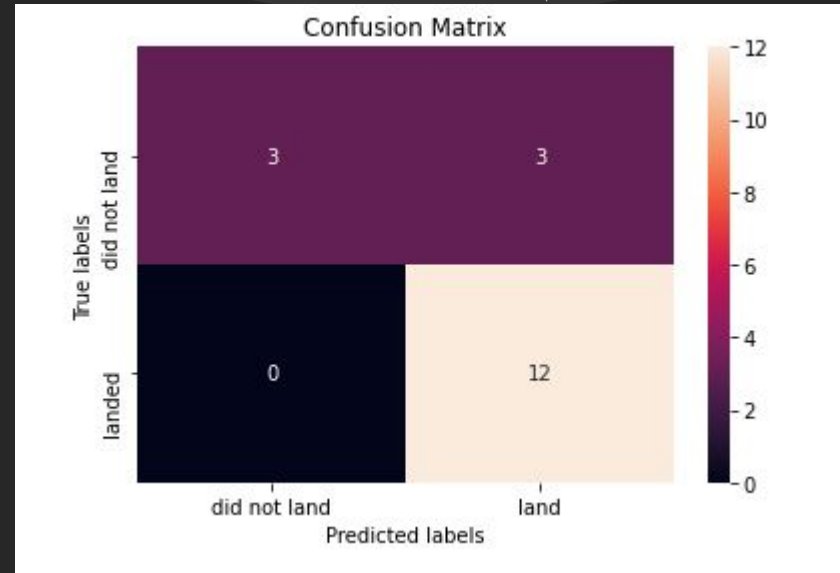
CONCLUSION

CONCLUSION

- The collection of data were collected from SpaceX REST API and Web scrape from wiki pages.

- The target object is Falcon 9, it has reached the farthest orbit for 27 times which is GTO orbit. The range of GTO orbit is 35,768 KM above Earth's equator and the overall success rate of landing the booster of Falcon 9 is a 66.66%

- In EDA visualization show that there is a success rate of launching Falcon 9 had increase since 2013 until 2020.

- In EDA with SQl found that Falcon 9 was launched at 3 different launch site, which is CCAFS SLC-40 , KSC LC-39A , VAFB SLC-4E. The total number mission outcomes are 100 success mission and 1 fail mission (in flight). The first success landing is on 01 -05 -2017,

CONCLUSION

- In the Folium shows the location of each launch site and the distance to different city and railway.

- In ploty dash show the highest launch record with 46.4% at CCAF SLC-40 and the lowest launch record with 12.5% at CCAFS SLC-40. It also show the booster version B4 had carried a heaviest payload with 9600 KG and also lightest payload with 362 KG had launch successfully

- In the predictive analysis model shows that the prediction method that achieved the best performance is Decision Tree with 89% Accuracy and 0.944 Score.

APPENDIX

- Python
- SQL
- Plotly Dash
- Folium
- Pandas
- Numpy
- Matplotlib
- IBM Watson
- IBM DB2
- Scikit Learn
- Line Chart
- Wikipedia : https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches
- SpaceX REST API : https://api.spacexdata.com/v4/launches/past

THANK YOU