Space X
FALCON 9
and Data Science

Ing. Marbellys Campos

2023-09-30

Outline

Executive Summary

- The data is collected through data collection API to get SpaceX Data Sets
- To analyze and visualize data, we use exploratory data analysis.
- To find the best machine learning model and predict the classification of the next landing, we use the models of Decision tree, support vector machine (SVM), k nearest neighbors (KNN) and logistic regresion.
- To calculate the accuracy on the test data we use the score method.
- To visualize the best performing model we use the confusion matrix plot.

Introduction

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.

Through Data Science we will answer the following question:

Can we predict a new successful launch of the first landing stage according to historical launch data?

Section 1

Methodology

Methodology

1.- Data collection:

- Request to the SpaceX API
- Web scraping Falcon 9 and Falcon Heavy Launches Records from Wikipedia
- Clean the requested data and remove missing values.

2.- Perform data wrangling

- Generate the classification variable (Class) that represents the outcome of each landing

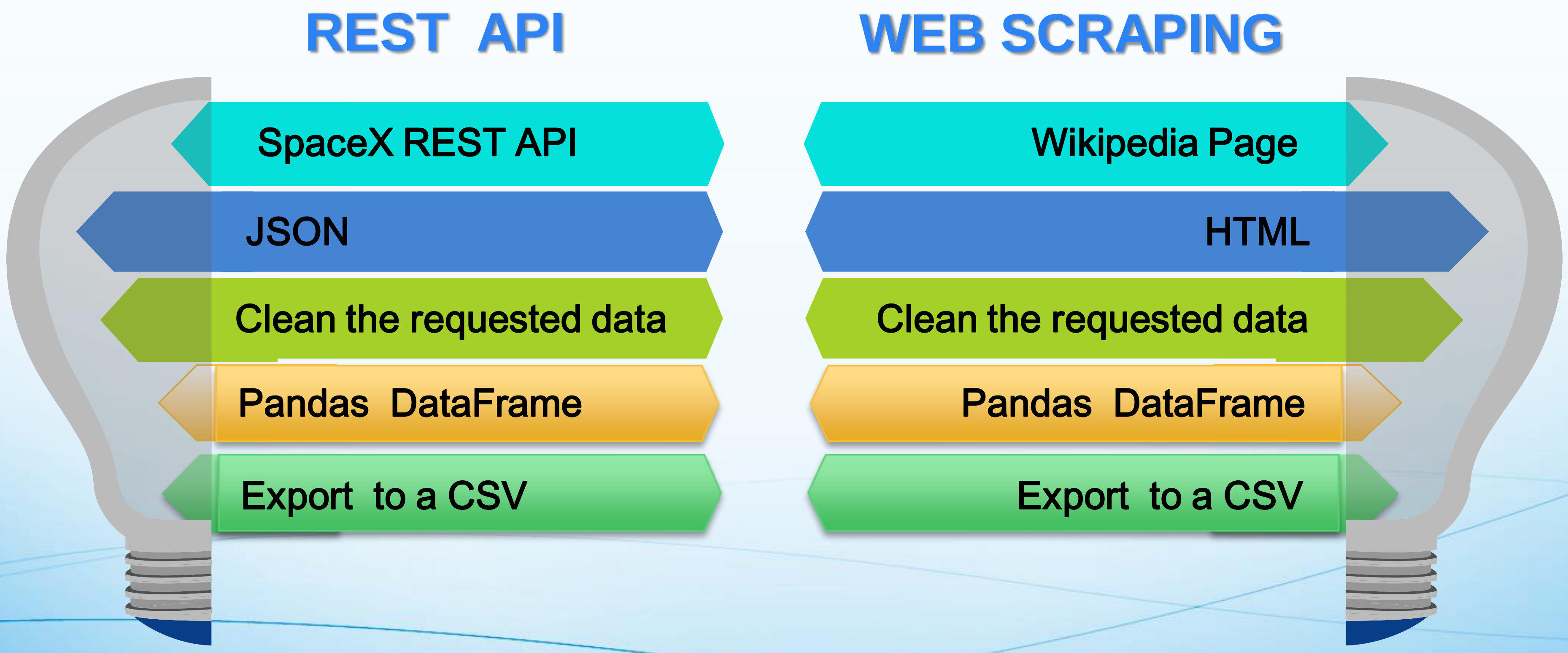
3.- Perform exploratory data analysis (EDA) using SQL, Pandasand and Matplotlib

4.- Perform interactive visual analytics using Folium and Plotly Dash

5.- Perform predictive analysis using machine learning classification models Logistics Regression – Decision tree – K nearest neighbors – Support vector machines.

6.- To obtain accuracy on the test data will be use the score method.

Data Collection



Data Collection – SpaceX API

* **SpaceX API repository**

  https://github.com/r-spacex/SpaceX-API

* **Main Endpoint**
  https://api.spacexdata.com/v4/launches/past

My Notebook (GitHub URL):

https://github.com/marbellys/ibm-data-science-capstone/blob/main/labs%200-spacex-data-collection-api.ipynb

SpaceX REST API

static_json_url='https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/API_call_spacex_api.json'

JSON

```
response.json()[:1]
```


```
[{'fairings': {'reused': False,
   'recovery_attempt': False,
   'recovered': False,
   'ships': []},
  'links': {'patch': {'small': 'https://images2.imgbox.com/94/f2/NN6Ph45r_o.png',
    'large': 'https://images2.imgbox.com/5b/02/QcxHUb5V_o.png'},
   'reddit': {'campaign': None,
```


Clean , Transform, Split Data

```
data = data[['rocket', 'payloads', 'launchpad', 'cores', 'flight_number', 'date_utc']]
```

```
data = data[data['cores'].map(len)==1]
data = data[data['payloads'].map(len)==1]
```

```
data['cores'] = data['cores'].map(lambda x : x[0])
data['payloads'] = data['payloads'].map(lambda x : x[0])
```

```
data['date'] = pd.to_datetime(data['date_utc']).dt.date
```

```
data = data[data['date'] <= datetime.date(2020, 11, 13)]
```

DataFrame

```
launchDp =pd.DataFrame(launch_dict)
```

FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Leg	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
1	2010-06-04	Falcon 9	6123.547647	LEO	CCSFS SLC 40	None None	1	False	False	Fals	None	1.0	0	B0003	-80.577366	28.561857
2	2012-05-22	Falcon 9	525.000000	LEO	CCSFS SLC 40	None None	1	False	False	Fals	None	1.0	0	B0005	-80.577366	28.561857
3	2013-03-01	Falcon 9	677.000000	ISS	CCSFS SLC 40	None None	1	False	False	Fals	None	1.0	0	B0007	-80.577366	28.561857
4	2013-09-29	Falcon 9	500.000000	PO	VAFB SLC 4E	False Ocean	1	False	False	Fals	None	1.0	0	B1003	-120.610829	34.632093
5	2013-12-03	Falcon 9	3170.000000	GTO	CCSFS SLC 40	None None	1	False	False	Fals	None	1.0	0	B1004	-80.577366	28.561857

Export CSV

```
data_falcon9.to_csv('dataset_part_1.csv', index=False)
```

Data Collection - Scraping

* **Wikipedia page - Falcon 9 historical launch records**

  https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

My Notebook (GitHub URL):

https://github.com/marbellys/ibm-data-science-capstone/blob/main/labs%201-jupyter-labs-webscraping.ipynb

Wikipedia Page

static_url = "https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922"

HTML

```
html = requests.get(static_url)
soup = BeautifulSoup(html.text)
html_tables = soup.find_all('table'
first_launch_table = html_tables[2]
```

```
<tr>
<th scope="col">Flight No.
</th>
<th scope="col">Date and<br/>time (<a href="/wiki/Coordinated_Universal_Time" title="Coordinated Universal Time">UTC</a>)
</th>
<th scope="col"><a href="/wiki/List_of_Falcon_9_first-stage_boosters" title="List of Falcon 9 first-stage boosters">Version,<br/>Booster</a> <sup class="reference" id="cite_ref-booster_11-0"><a href="#cite_note-booster-11">[b]</a></sup>
</th>
<th scope="col">Launch site
</th>
<th scope="col">Payload<sup class="reference" id="cite_ref-Dragon_12-0"><a href="#cite_note-Dragon-12">[c]</a></sup>
</th>
<th scope="col">Payload mass
</th>
<th scope="col">Orbit
</th>
<th scope="col">Customer
</th>
<th scope="col">Launch<br/>outcome
</th>
<th scope="col"><a href="/wiki/Falcon_9_first-stage_landing_tests" title="Falcon 9 first-stage landing tests">Booster<br/>landing</a>
</th></tr>
```

Clean , Transform, Split Data

```
launch_dict= dict.fromkeys(column_names)
del launch_dict['Date and time ( )']
```

```
time = datatimelist[1]
launch_dict['Time'].append(time)
```

DataFrame

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

	Flight No.	Launch site	Payload	Payload mass	Orbit	Customer	Launch outcome	Version Booster	Booster landing	Date	Time
0	1	CCAFS	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success\n	F9 v1.0B0003.1	Failure	4 June 2010	18:45
1	2	CCAFS	Dragon	0	LEO	NASA	Success	F9 v1.0B0004.1	Failure	8 December 2010	15:43
2	3	CCAFS	Dragon	525 kg	LEO	NASA	Success	F9 v1.0B0005.1	No attempt\n	22 May 2012	07:44
3	4	CCAFS	SpaceX CRS-1	4,700 kg	LEO	NASA	Success\n	F9 v1.0B0006.1	No attempt	8 October 2012	00:35
4	5	CCAFS	SpaceX CRS-2	4,877 kg	LEO	NASA	Success\n	F9 v1.0B0007.1	No attempt\n	1 March 2013	15:10

Export CSV

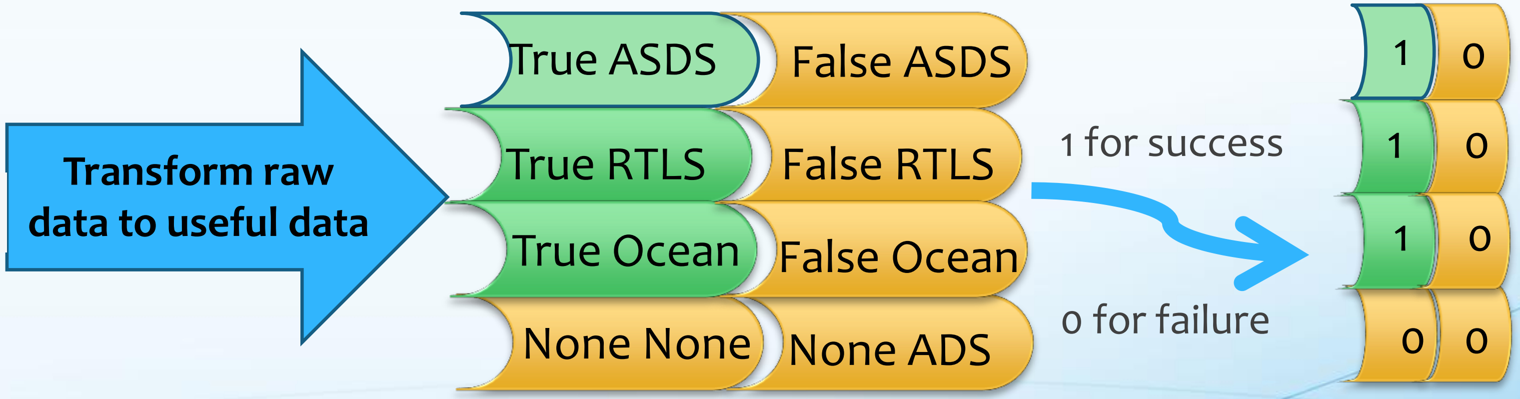
```
df.to_csv('spacex_web_scraped.csv', index=False)
```

Data Wrangling

The object is Perform exploratory Data Analysis and determine Training Labels.

In this case, the outcomes label, has different categorical values that we must transform to landing class (prediction target)

```
bad_outcomes=set(landing_outcomes.keys()[[1,3,5,6,7]])
landing_class = [0 if i in bad_outcomes else 1 for i in df['Outcome']]
```



Launch Site	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude	Class
CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0003	-80.577366	28.561857	0
CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0005	-80.577366	28.561857	0
CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0007	-80.577366	28.561857	0
VAFB SLC 4E	False Ocean	1	False	False	False	NaN	1.0	0	B1003	-120.610829	34.632093	0
CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1004	-80.577366	28.561857	0

My Notebook (GitHub URL):

https://github.com/marbellys/ibm-data-science-capstone/blob/main/labs%202-jupyter-spacex-Data%20wrangling.ipynb

EDA with SQL

- Query the names of the unique **launch sites** in the space mission.

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

- Query 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
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-08-12	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

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

sum(PAYLOAD_MASS__KG_)
45596

- Query average payload mass carried by booster version F9 v1.1

sum(PAYLOAD_MASS__KG_)
45596

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

min(date)
2015-12-22

- List the names of the booster_versions which have carried the maximum payload mass.

- List the total number of successful and failure mission outcomes

Mission_Outcome	count(*)
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

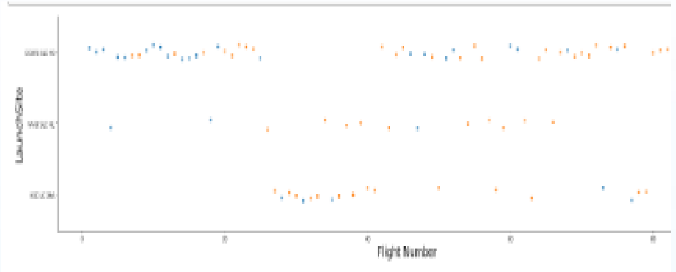
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order.

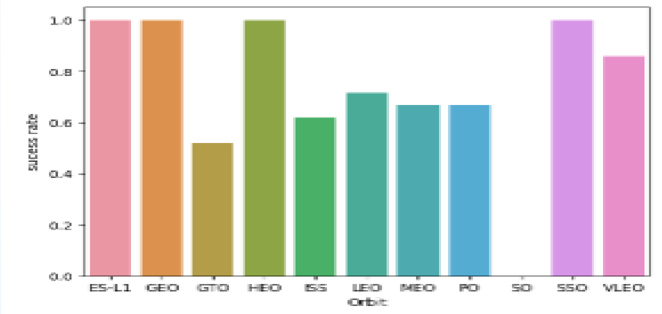
Landing_Outcome	landing
Success (ground pad)	5
Failure (drone ship)	5

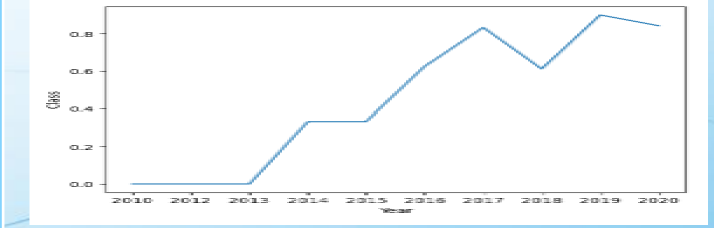
My Notebook (GitHub URL):
https://github.com/marbellys/ibm-data-science-capstone/blob/main/labs%203-eda-sql.ipynb

EDA with Data Visualization

Scatter Plots	To visualize the relationship between variables. For example: • How FlightNumber and Payload variables would affect the launch outcome. • Flight Number and Launch Site • Payload and Launch Site
Bar Plots	To visualize if there are any relationship between *success rate* and *orbit type*
Line Plots	To Visualize the launch *success yearly* trend







My Notebook (GitHub URL):

https://github.com/marbellys/ibm-data-science-capstone/blob/main/labs%204-eda-dataviz.ipynb

Build an Interactive Map with Folium

folium.Circle

- To add a highlighted circle area with a text label on a specific coordinate. Eg: Create a blue circle at NASA Johnson Space Center's coordinate with a popup label showing its name

folium.Marker

- Adds labels for Launch sites

MarkerCluster

- If a launch was successful (class=1), then we use a green marker and if a launch was failed, we use a red marker (class=0)

folium.PolyLine

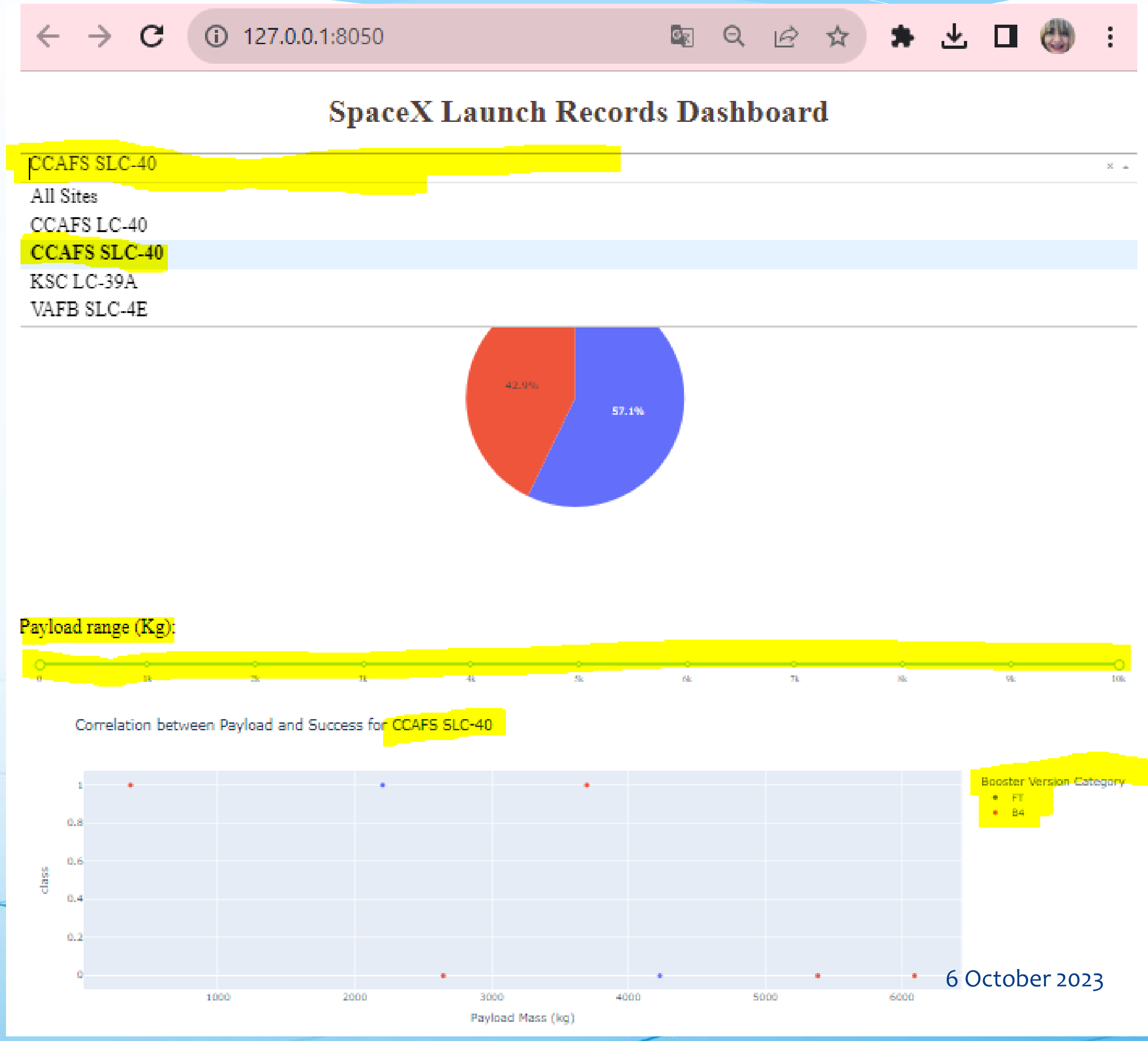
- *Draw the line, f*or calculate distance between launch sites and their proximities

My Notebook (GitHub URL):

https://github.com/marbellys/ibm-data-science-capstone/blob/main/labs%205-Launch%20Sites%20Locations%20Analysis%20with%20Folium.ipynb

Build a Dashboard with Plotly Dash

- **Dropdown list** to enable Launch Site selection.
- **Pie chart** to show **the total successful launches** count for all sites, if dropdown list is **All Sites**. If a specific launch site was selected, show the **Success vs. Failed** counts for the site.
- **Slider** to select payload range.
- **Scatter chart** to show the correlation between **payload and launch success.**

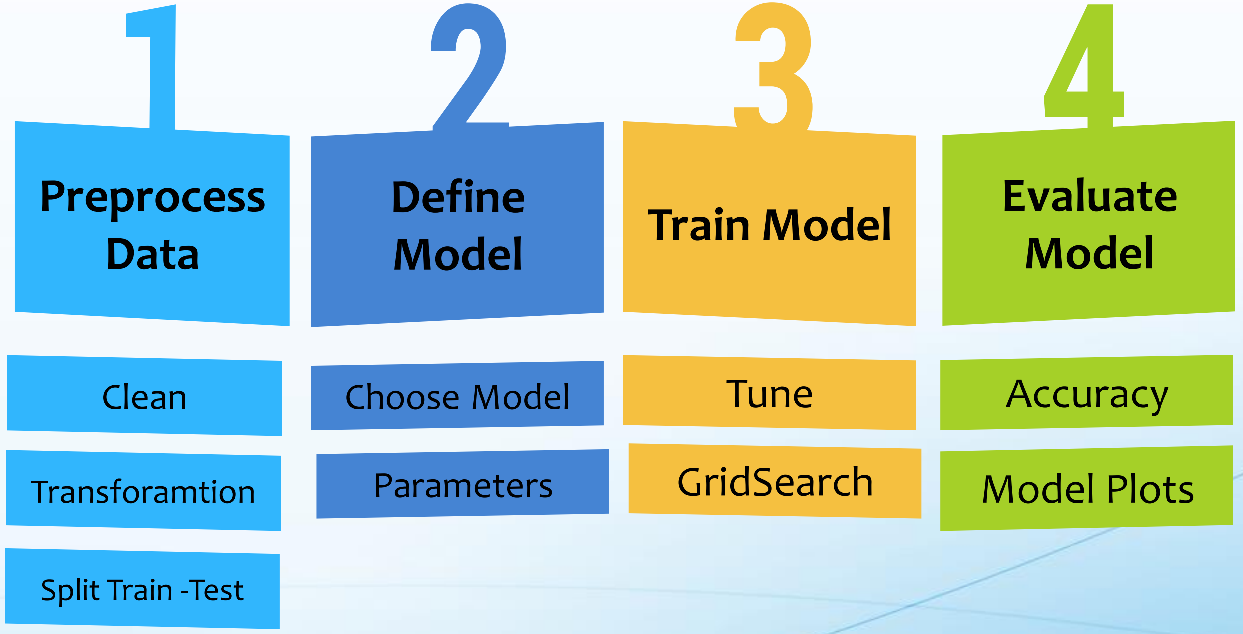


My Dashboard Python Code (GitHub URL):

https://github.com/marbellys/ibm-data-science-capstone/blob/main/labs%206-Dash_interactivity.py

Predictive Analysis (Classification)

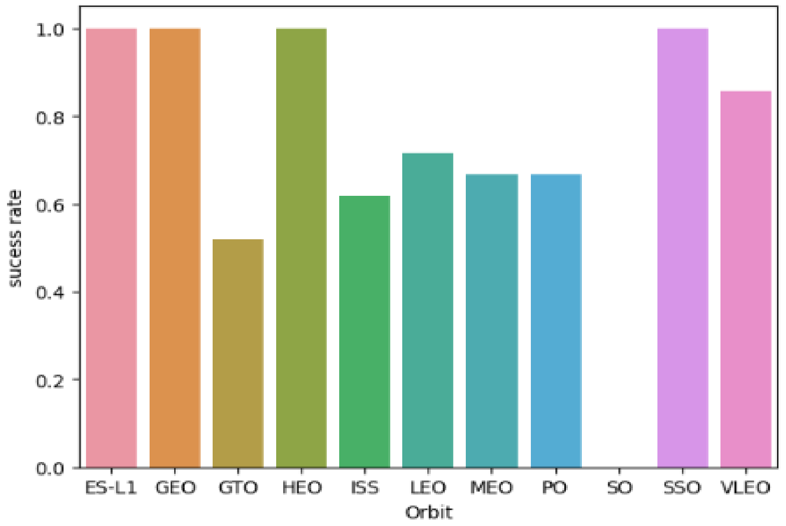
- ✓ Load the data (cleaned & transformed)
- ✓ Perform exploratory Data Analysis to determine Training Labels.
  - ✓ create a column for the class (Y target variable)
  - ✓ Standardize the data (X training variables)
  - ✓ Split into training data and test data (20% test data)
- ✓ Fit SVM, Classification Trees and Logistic Regression to find best Hyperparameter, using GridSearchCV.
- ✓ Calculate the accuracy on the test data using the method **score** to find the method performs best

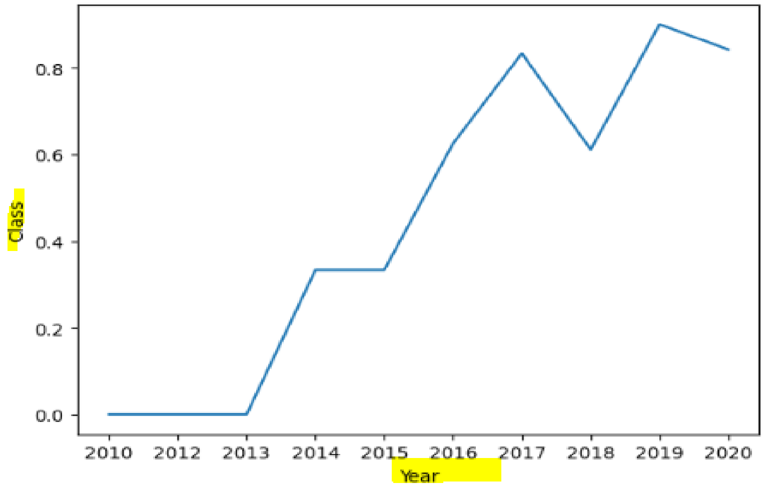


My Notebook (GitHub URL):

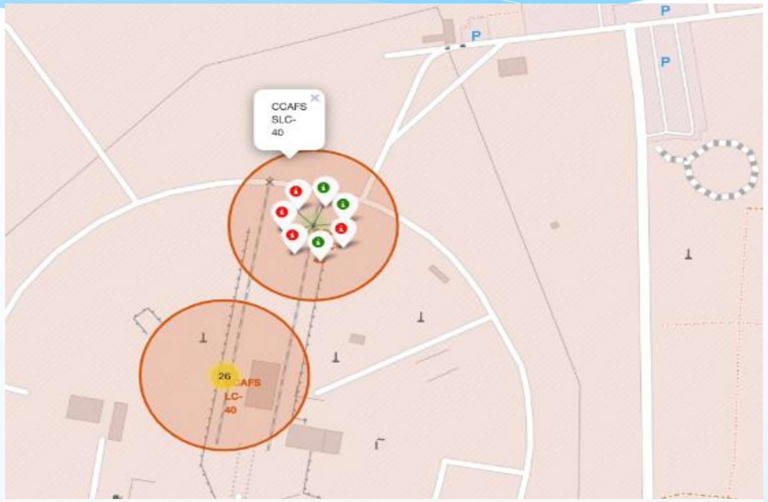
https://github.com/marbellys/ibm-data-science-capstone/blob/main/labs%207%20-SpaceX_Machine_Learning_Prediction_Part_5_MARBELLYS.ipynb

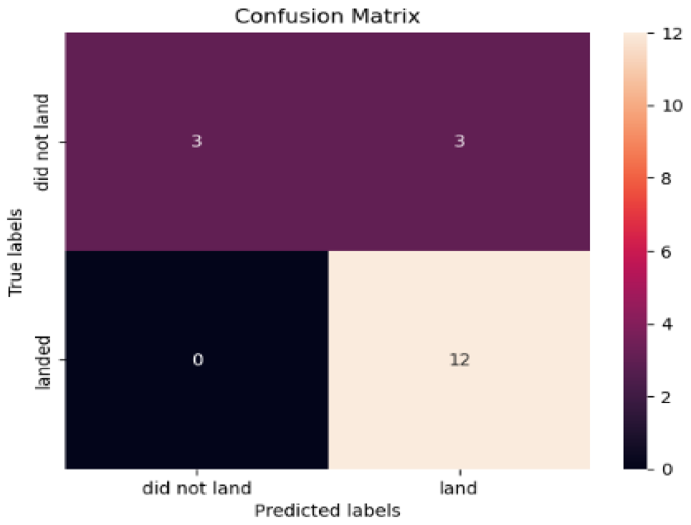
Results


1.0
0.8
0.6
0.4
0.2
0.0
sucess rate
ES-L1
GEO
GTO
HEO
ISS
LEO
MEO
PO
SO
SSO
VLEO
Orbit


0.8
0.6
0.4
0.2
0.0
Class
2010
2012
2013
2014
2015
2016
2017
2018
2019
2020
Year


SpaceX Launch Records Dashboard
All Sites
Total Success Launches by Site
41.7%
29.2%
16.7%
12.5%
KSC LC-39A
CCAFS LC-40
VAFB SLC-4E
CCAFS SLC-40
Payload range (Kg):
Correlation between Payload and Success for all Sites
Booster Version Category
v1.0
v1.1
FT
B4
B5
class
Payload Mass (kg)


CCAFS
SLC-
40
26

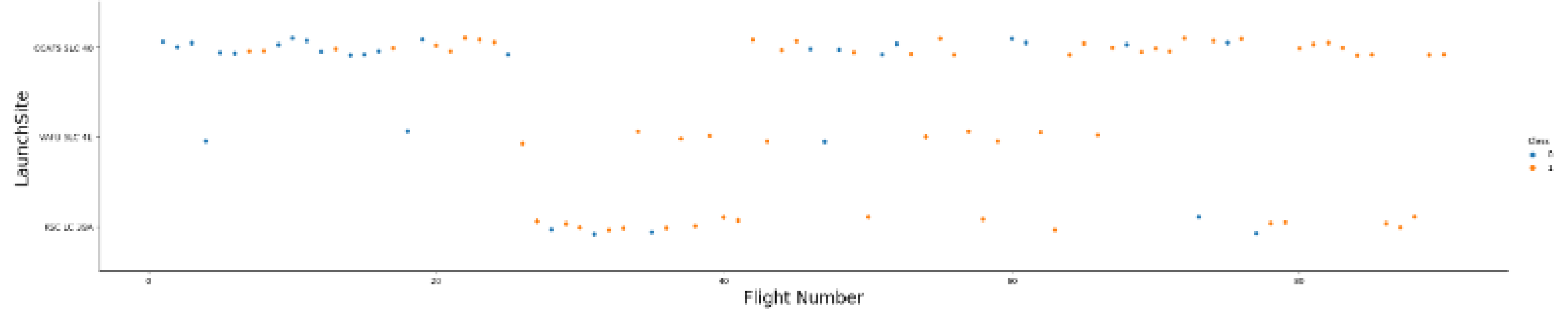

Confusion Matrix
3
3
0
12
True labels
did not land
landed
did not land
land
Predicted labels
12
10
8
6
4
2
0

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

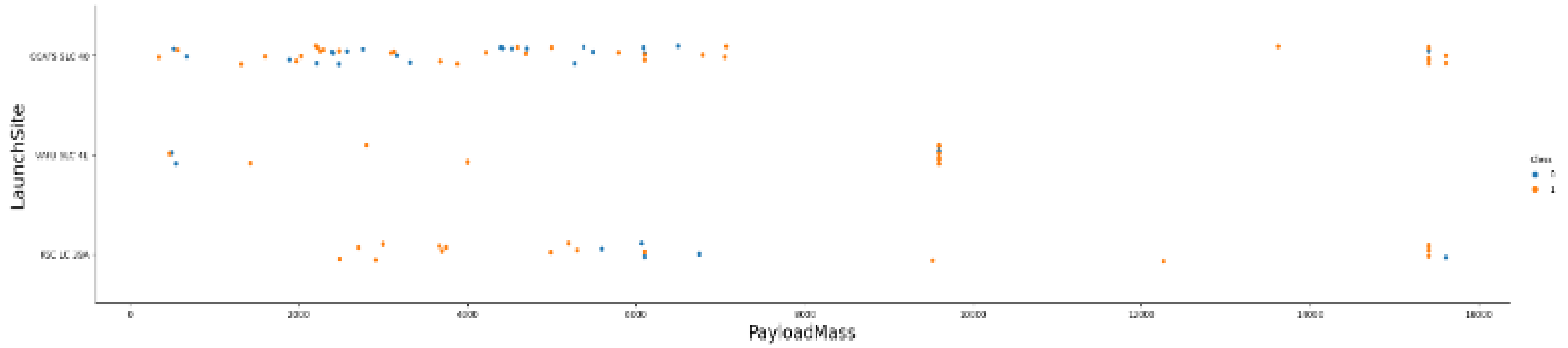
```
sns.catplot(y="LaunchSite", x="FlightNumber",  hue="Class",data=df, aspect = 5)

# Plot a scatter point chart with x axis to be Flight Number and y axis to be the launch site, and hue to be the class value
plt.xlabel("Flight Number",fontsize=20)
plt.ylabel("LaunchSite",fontsize=20)
plt.show()
```



EXPLANATION: We can see from the scatter plot that as the number of flights increases, there are more successful landings in the first stage. The first launches were made mostly in CCAFS SLC 40, with little probability of success. However, in the latest launches the probability of success is quite high. At the VAFB SLC 4E and KSC LC 39A sites, although fewer launches are seen, much higher success rates are observed.

Payload vs. Launch Site

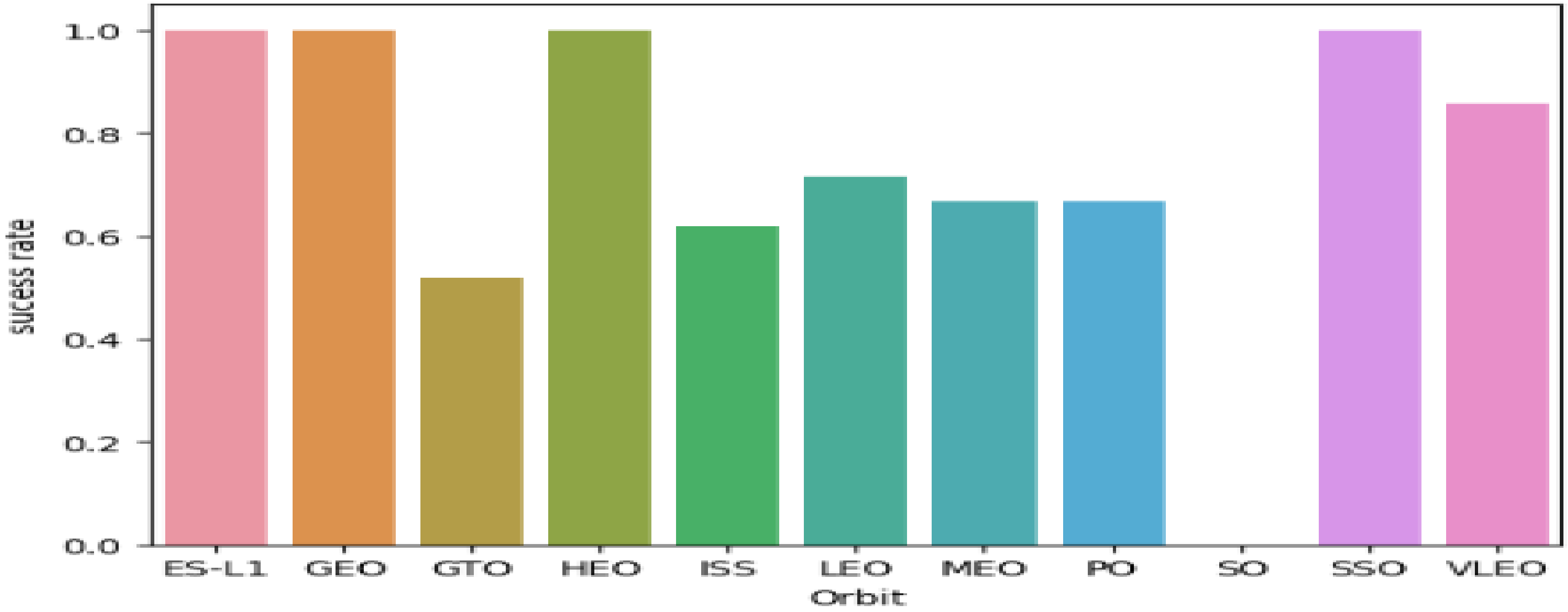
```
### TASK 2: Visualize the relationship between Payload and Launch Site
sns.catplot(y="LaunchSite", x="PayloadMass", hue="Class",data=df, aspect = 5)
plt.ylabel("LaunchSite",fontsize=20)
plt.xlabel("PayloadMass",fontsize=20)
plt.show()
```



EXPLANATION: For the VAFB-SLC launchsite there are no rockets launched for heavypayload mass (greater than 10000). When the Payload is high, the successs rate is much higher. In KSC LC39A launchsite is much higher successs rate with low Payload whereas this rate is much lower in CCAFS SCL 40 lauchsite.

Success Rate vs. Orbit Type

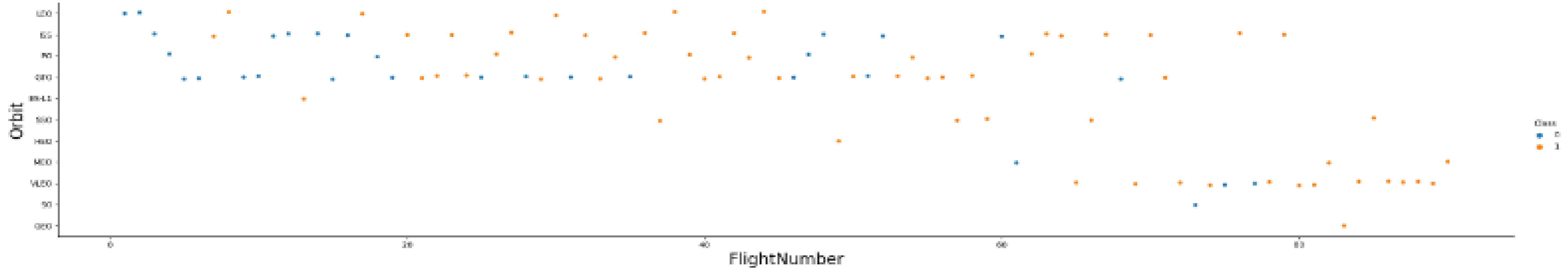
```
sns.barplot(x='Orbit' , y='Class'  , data=df_success)
plt.xlabel("Orbit",fontsize=10)
plt.ylabel("sucess rate",fontsize=10)
plt.show()
```



EXPLANATION: The Orbit types ES-L1, GEO, HEO and SSO have the more more higher success rate (100%), whereas SO orbit is null.

Flight Number vs. Orbit Type

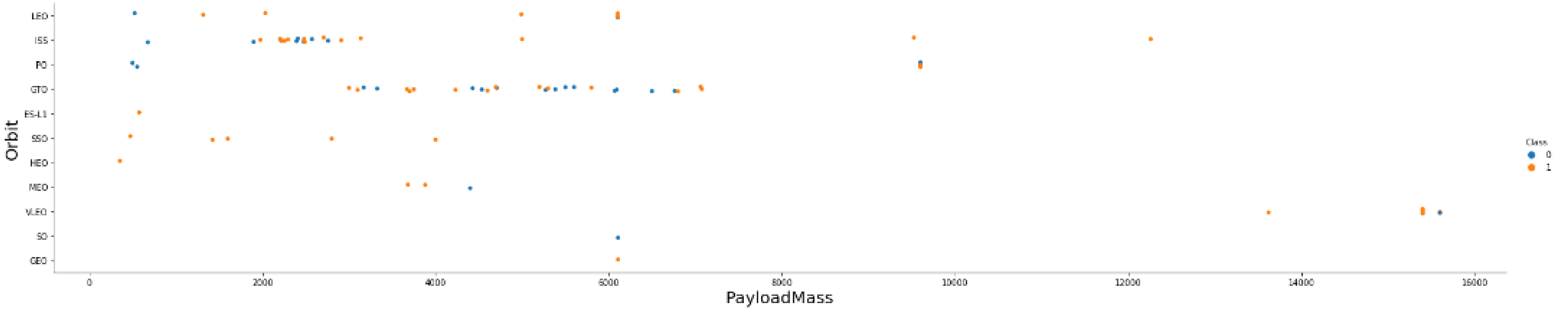
```
### TASK  4: Visualize the relationship between FlightNumber and Orbit type
sns.catplot(x='FlightNumber', y='Orbit', hue='Class', data= df,aspect=5)
plt.xlabel("FlightNumber",fontsize=20)
plt.ylabel("Orbit",fontsize=20)
plt.show()
```



EXPLANATION: all launches are successful in ES-L1, GEO, HEO and SSO orbits. There is a relationship between the number of flights and the success rate in LEO orbit, when the number of flights increases, the success rate increases. Oppositely, there is no such obvious relationship in the GTO orbit.

Payload vs. Orbit Type

```
### TASK  5: Visualize the relationship between Payload and Orbit type

sns.catplot(x='PayloadMass',y='Orbit',hue='Class', data=df, aspect=5)
plt.xlabel('PayloadMass',fontsize=20)
plt.ylabel('Orbit',fontsize=20)
plt.show()
```

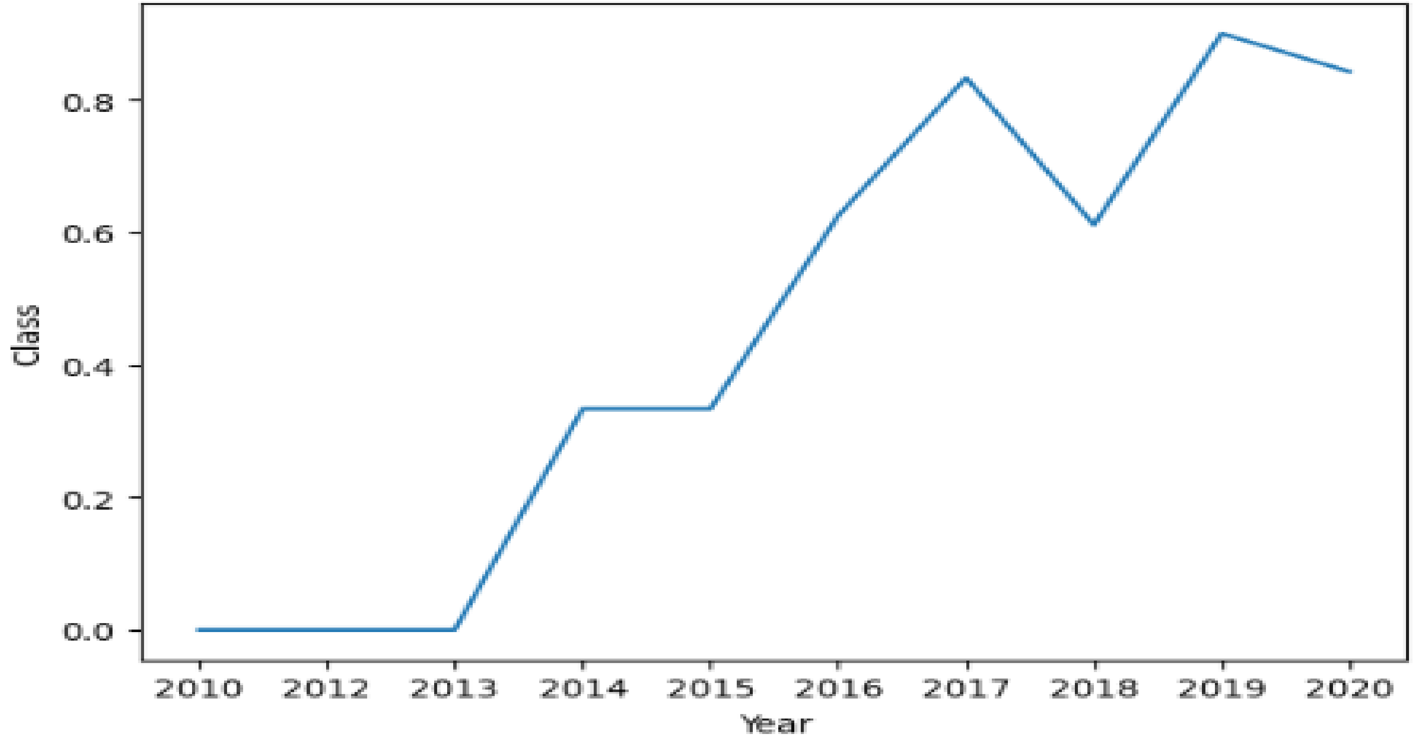


EXPLANATION: the successful landing rate are higher for Polar,LEO and ISS with heavy payloads.

whereas for GTO is dificult see this well as both positive landing rate and negative landing.

Launch Success Yearly Trend

```
# Plot a line chart with x axis to be the extracted year and y axis to be the success rate
sns.lineplot(x='Year',y='Class',data=df_years_success)
plt.show()
```



EXPLANATION: The sucess rate since 2013 kept increasing till 2020

All Launch Site Names

```
%sql select distinct Launch_Site from SPACEXTABLE
```

```
 * sqlite:///my_data1.db
Done.
```

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

EXPLANATION: There are four launch sites in the space mission.

Launch Site Names Begin with 'CCA'

```
%sql select *  from SPACEXTABLE where Launch_Site like 'CCA%' LIMIT 5
```

```
 * sqlite:///my_data1.db
Done.
```

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-08-12	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

EXPLANATION: there are five launches in LEO Orbit. All launches resulted success. There are four launches whith NASA Customer.

Total Payload Mass

```
5]: %sql select sum(PAYLOAD_MASS__KG_) from SPACEXTABLE where Customer ='NASA (CRS)'
```

```
 * sqlite:///my_data1.db
Done.
```

5]:

sum(PAYLOAD_MASS__KG_)
45596

EXPANATION: the total payload mass carried by boosters launched by NASA (CRS) is 45596

Average Payload Mass by F9 v1.1

In [53]:
```
%sql select avg(PAYLOAD_MASS__KG_) from SPACEXTABLE where Booster_Version ='F9 v1.1'
```

```
 * sqlite:///my_data1.db
Done.
```

Out[53]:

avg(PAYLOAD_MASS__KG_)
2928.4

EXPANATION: the average payload mass carried by booster version F9 v1.1 is 2928.4

First Successful Ground Landing Date

```
%sql select min(date)  from SPACEXTABLE where Landing_Outcome = 'Success (ground pad)'
```

```
 * sqlite:///my_data1.db
Done.
```

min(date)
2015-12-22

EXPANATION: the date when the first succesful landing outcome in ground pad was acheived in 2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

```
]: %%sql
select Booster_Version
from SPACEXTABLE
where Landing_Outcome='Success (drone ship)'
and (PAYLOAD_MASS__KG_ >4000 and PAYLOAD_MASS__KG_<6000)
```

```
 * sqlite:///my_data1.db
Done.
```

]:

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

```
]: EXPANATION: the names of the boosters which have success in drone ship
and have payload mass greater than 4000 but less than 6000 are four:
    F9 FT B1022
    F9 FT B1026
    F9 FT B1021.2
    F9 FT B1031.2
```

Total Number of Successful and Failure Mission Outcomes

```
%%sql

select count(*)
from SPACEXTABLE
where  Mission_Outcome like "Success%"
```

```
 * sqlite:///my_data1.db
Done.
```

count(*)
100

```
%%sql

select count(*)
from SPACEXTABLE
where  Mission_Outcome Not like "Success%"
```

```
 * sqlite:///my_data1.db
Done.
```

count(*)
1

```
EXPANATION: total number of successful mission outcomes is 100

            total number of failure mission outcomes is 1
```

Boosters Carried Maximum Payload

```
EXPANATION: the names of the booster_versions which have carried the maximum payload mass are:
    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
```

```
%%sql
select Booster_Version
from SPACEXTABLE
where PAYLOAD_MASS__KG_ in (select max(PAYLOAD_MASS__KG_)
                             from SPACEXTABLE)
```

```
 * sqlite:///my_data1.db
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

2015 Launch Records

```
%%sql
select date, substr(Date, 6, 2) Month,Landing_Outcome,Booster_Version,launch_site
from SPACEXTABLE
where Landing_Outcome='Failure (drone ship)'
and substr(Date,1,4)='2015'
```

```
 * sqlite:///my_data1.db
Done.
```

Date	month	Landing_Outcome	Booster_Version	Launch_Site
2015-10-01	10	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
2015-04-14	04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

EXPANATION:Failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015 are:

```
2015-10-01    10    Failure (drone ship)    F9 v1.1 B1012    CCAFS LC-40
2015-04-14    04    Failure (drone ship)    F9 v1.1 B1015    CCAFS LC-40
```

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

```
%%sql

select Landing_Outcome, count(*) landing
from SPACEXTABLE
where Landing_Outcome in ('Failure (drone ship)','Success (ground pad)')
and date between '2010-06-04' and '2017-03-20'
group by Landing_Outcome
order by landing desc
```

```
 * sqlite:///my_data1.db
Done.
```

Landing_Outcome	landing
Success (ground pad)	5
Failure (drone ship)	5

EXPLANATION: 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 are:

```
Success (ground pad)    5
Failure (drone ship)    5
```

Section 3

Launch Sites Proximities Analysis

Locations of Launch Sites on Maps

	Launch Site	Lat	Long
0	CCAFS LC-40	28.562302	-80.577356
1	CCAFS SLC-40	28.563197	-80.576820
2	KSC LC-39A	28.573255	-80.646895
3	VAFB SLC-4E	34.632834	-120.610746

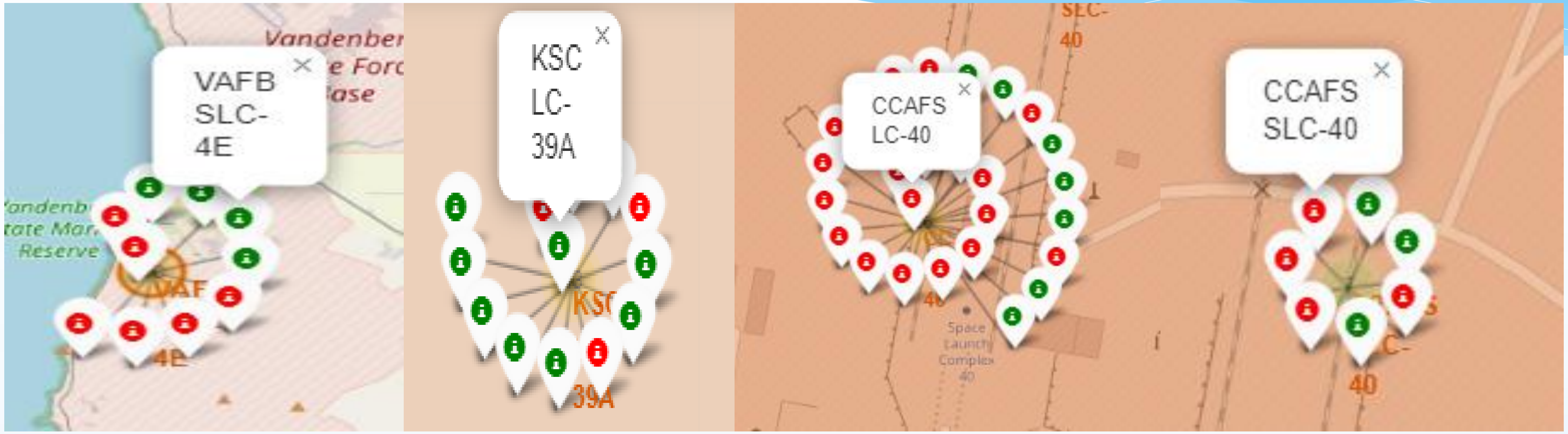
EXPLANATIONS:

All launch sites are not in proximity to the Equator line.

All launch sites are very close proximity to the coast.

There are three launch sites in the east and one in the west.

Display Launch Outcome by Color



EXPLANATION:

In green color we can see than KSC LC-39A launch site has the highest rate of success land.

Whereas CCAFS LC-40 has the lowest rate of unsuccess land

Show Distance to Proximities

: EXPLANATION:

The proximity from KSC LC-39A launch site to coastline is 6.51KM

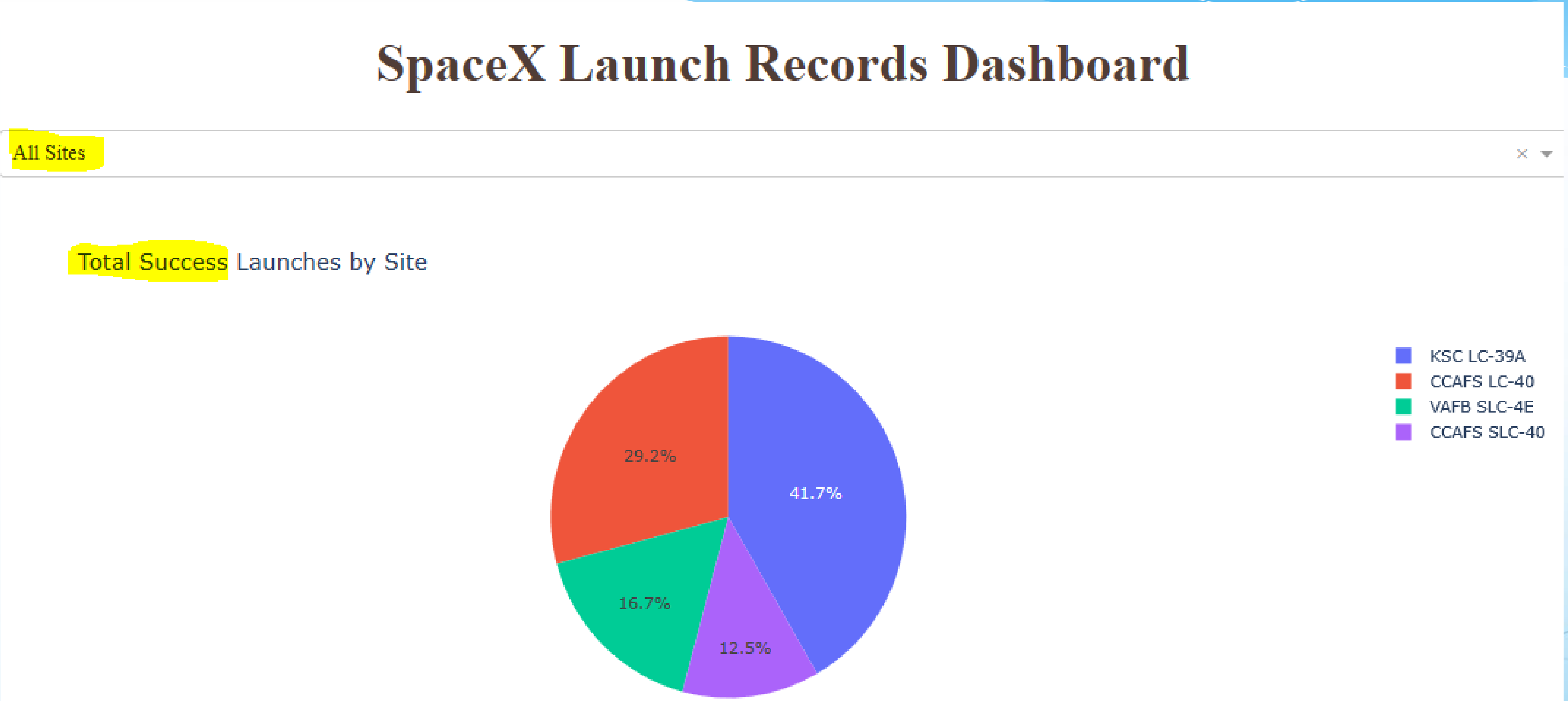
The proximity from KSC LC-39A launch site to highways is around 0.83KM

The distance from KSC LC-39A launch site to the nearest of launch and landing facility is about 4.16KM

Section 4

Build a Dashboard With Plotly Dash

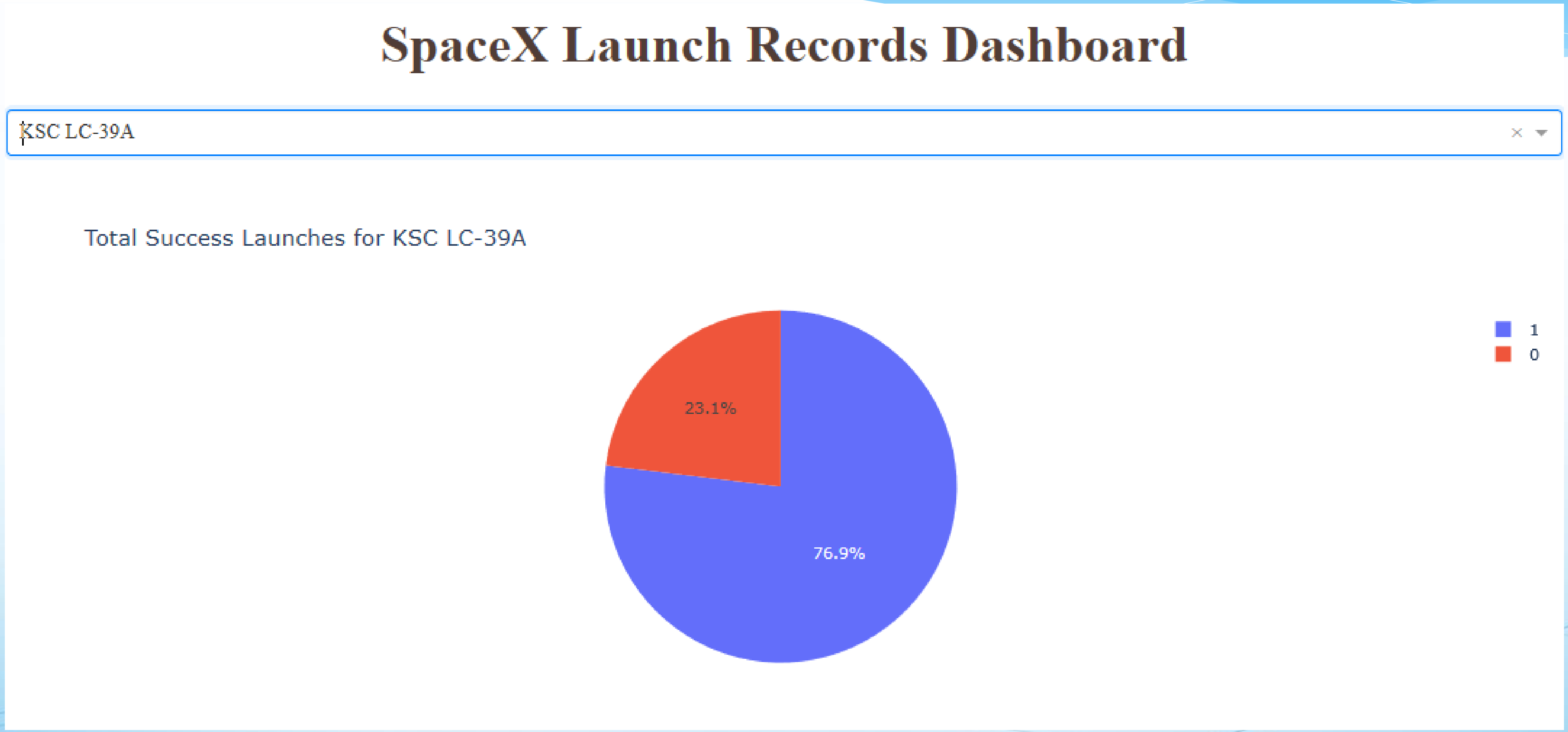
Total Success Launches for All Sites



EXPLANATIONS:

- The site with highest launch success ratio is KSC LC – 39 A (41.7 %)
- followed by CCAFS LC-40 (29.2%)
- The site wiht the lowest success lauches is VAFB SCL – 4E (16.7%)

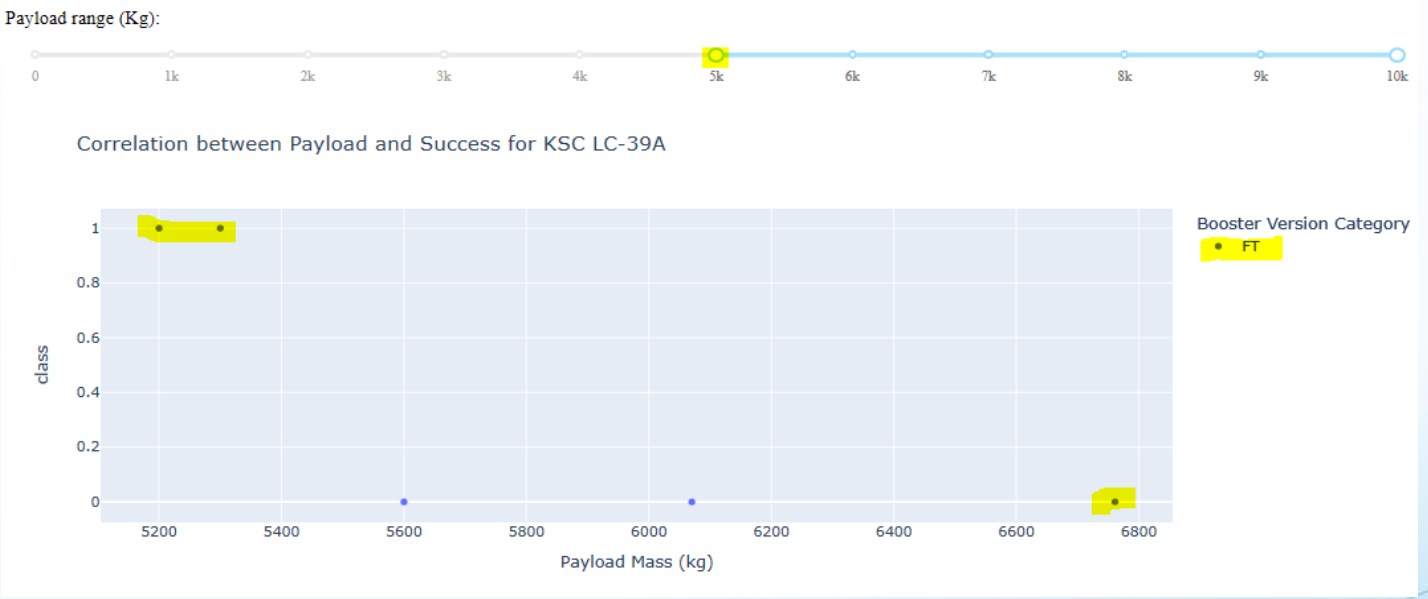
Launch site with highest launch success



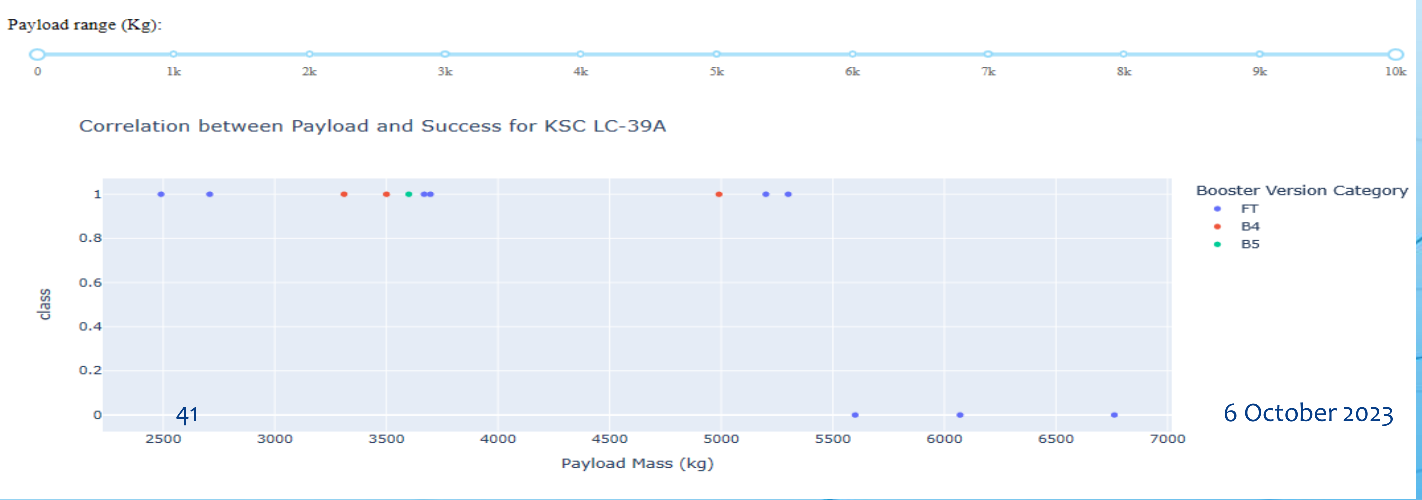
EXPLANATIONS:

- The launch site with highest launch success ratio is **KSC LC-39A**.
- At **KSC LC-39A**, 76.9% of launches were successful and 23.1% were unsuccessful.

Correlation Between Payload and Success



EXPLANATIONS:
Booster version **FT** has the most successful launches with the highest PayloadMass of 5300kg, although it has tried with greater aylοadMass (E.g. 6761Kg) but they have not been satisfactory.



EXPLANATIONS:
The highest Payload Mass of Booster version **B4** was 4990 kg for success launch.

Booster version **B5 has only one** success launch.

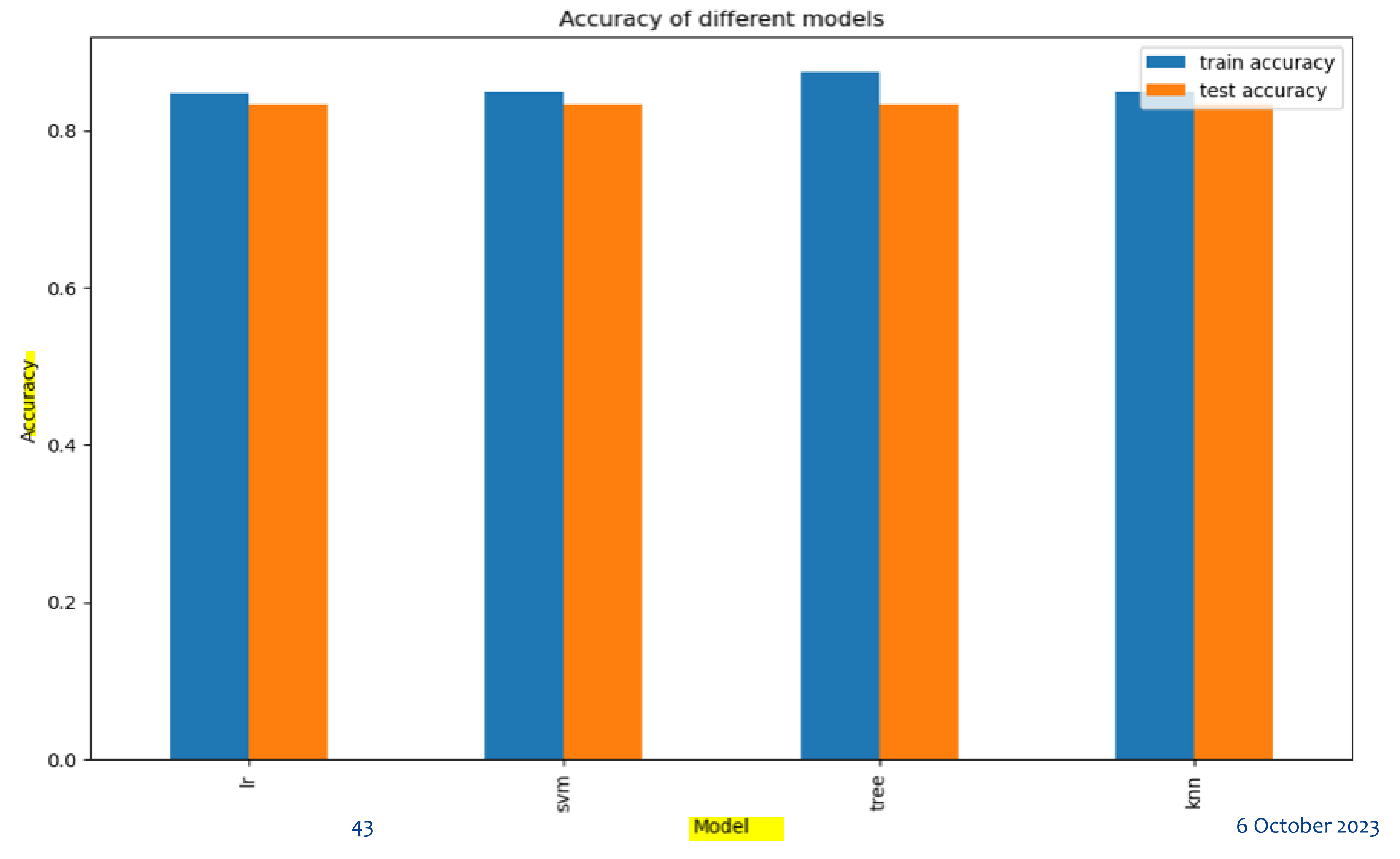
Section 5

Predictive Analysis (Classification)

Classification Accuracy

EXPLANATIONS:

- **Decision Tree model** has the highest classification **train** accuracy
- **All models** have classification **test** accuracy 0.833333

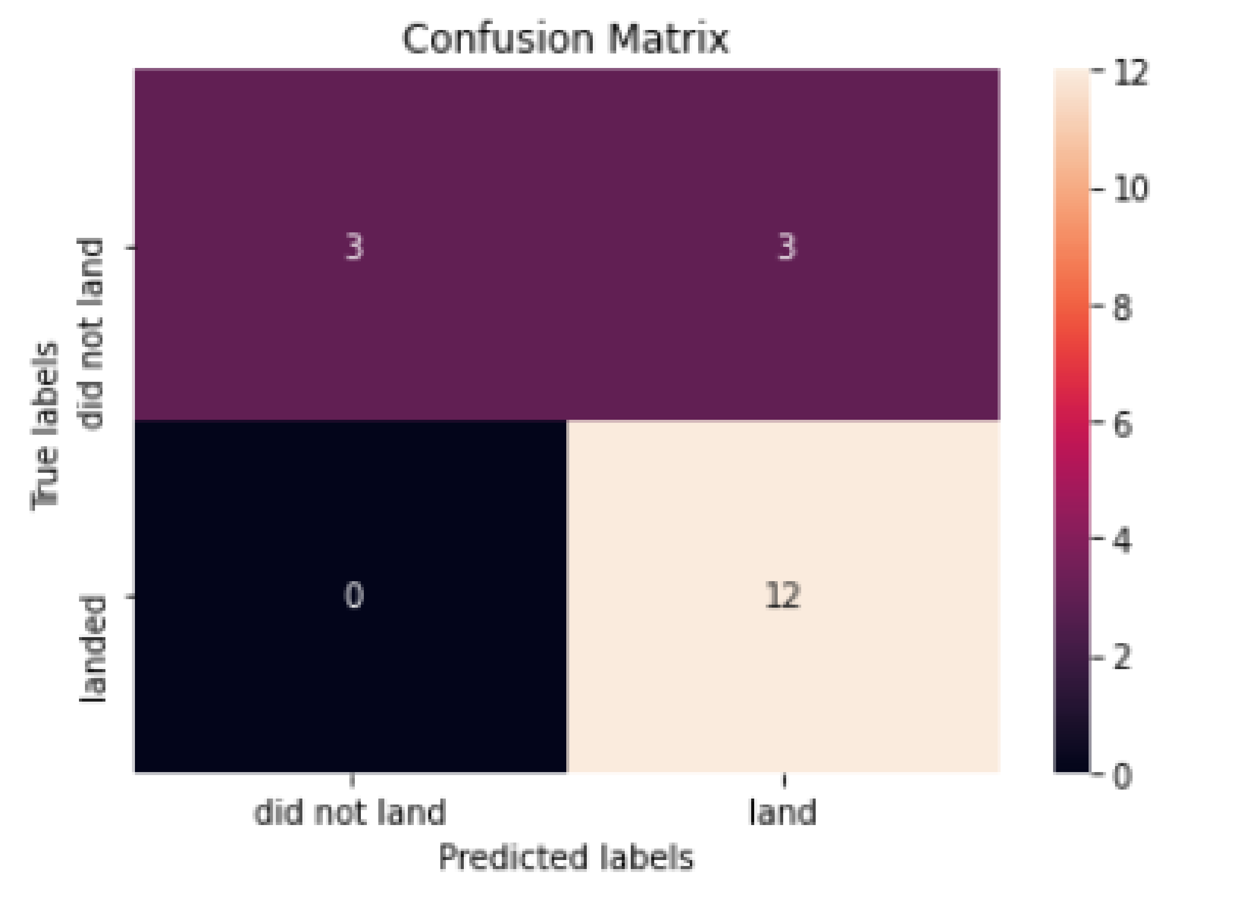


Confusion Matrix best performing model

EXPLANATIONS:

- Of the set of test, of a total of 18 launches, 12 of them were satisfactory and 6 of them were Unsatisfactory. Of these 12 success, the classifier predicted 12 to be success

- OF the 6 that landed unsatisfactorily, the classifier predicted that 3 would be successful and 3 would not be.



The best method

All four methods have the same accuracy hence they all perform the same at 83.3%

Conclusions

With every machine learning analysis carried out, the question that is the subject of this project can be answered….**Can we predict a new successful launch of the first landing stage according to historical launch data?**

Variables launch sites , payload mass, orbit type and launch year, associated with the success rate of Falcon 9 rocket launches determined that :

KSC LC-39A Launch sites had the highest rate of success land.
KSC LC39A launch site had much higher successs rate with low Payload mass.
The Orbit types ES-L1, GEO, HEO and SSO have the more higher success rate (100%).
The successful landing rate are higher for Polar, LEO and ISS with heavy payloads.
The sucess rate since 2013 kept increasing till 2020.

Therefore, it can be concluded that these findings will help Space X to reuse the Falcon 9 first stage and save costs.

Appendix

Share Links:

My Jupyter Notebook : https://github.com/marbellys/ibm-data-science-capstone

Reference Links:

Confussion Matrix: https://telefonicatech.com/blog/como-interpretar-la-matriz-de-confusion-ejemplo-practico

Muchas Gracias