IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Younes Tahri>
<24/10/2024>

Outline

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

Executive Summary

- **Summary of methodologies**
- Data Collection
- Data Wrangling
- Data Collection API(JSON) and WebScraping
- Data Visualization
- EDA SQL
- Interactive Map with Folium
- Building Dashboard with Plotly Dash
- Predictive Analysis(Classification)
- **Summary of all results**
- Data Vizualization

Introduction

- Project background and context

we will predict if the Falcon 9 first stage will land successfully. 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.

- Problems you want to find answers

What is features and variables that help to find successful landing ?

The effect each relationship with certain rocket variables will impact in determining the success rate of a successful landing.

Methodology

Executive Summary

- Data collection methodology:
  - Spacex REST API
  - Web Scrapping
- Perform data wrangling
  - Perform exploratory Data Analysis and determine Training Labels
- Perform exploratory data analysis (EDA) using visualization and SQL

Scatter and Bar graphs to show patterns of data

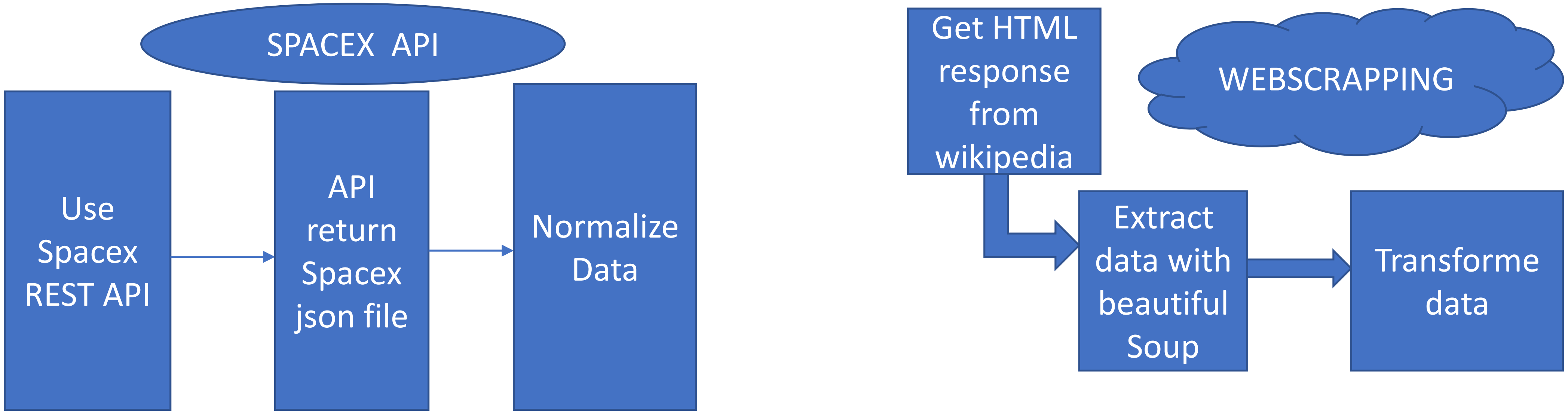
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Section 1

Methodology

Data Collection

- Data collection is process of gathering and measuring information on target variables by tools REST API and Webscrapping with BeautifulSoup



Data Collection - SpaceX API

GitHub url : https://github.com/tahri19younes98/Applied-Data-Science-Capastone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

simplified flow chart

1 .Getting Response from API

```
spacex_url="https://api.spacexdata.com/v4/launches/past"
response = requests.get(spacex_url).json()
```

2. Converting Response to a .json file

```
response = requests.get(static_json_url).json()
data = pd.json_normalize(response)
```

3. Apply custom functions to clean data

```
getLaunchSite(data)
getPayloadData(data)
getCoreData(data)
```

```
getBoosterVersion(data)
```

4. Assign list to dictionary then dataframe

```
launch_dict = {'FlightNumber': list(data['flight_number']),
'Date': list(data['date']),
'BoosterVersion':BoosterVersion,
'PayloadMass':PayloadMass,
'Orbit':Orbit,
'LaunchSite':LaunchSite,
'Outcome':Outcome,
'Flights':Flights,
'GridFins':GridFins,
'Reused':Reused,
'Legs':Legs,
'LandingPad':LandingPad,
'Block':Block,
'ReusedCount':ReusedCount,
'Serial':Serial,
'Longitude': Longitude,
'Latitude': Latitude}
```

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

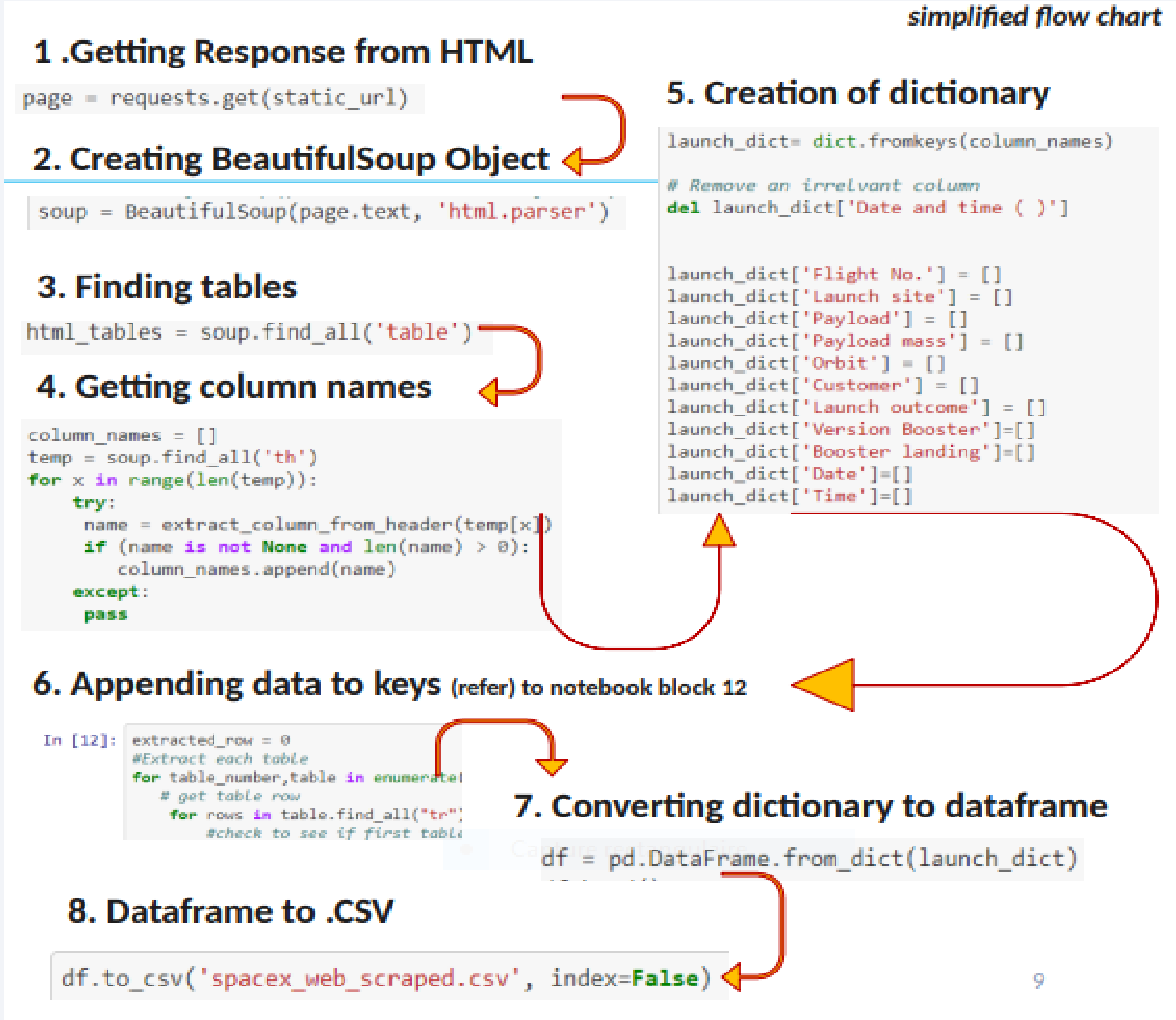
5. Filter dataframe and export to flat file (.csv)

```
data_falcon9 = df.loc[df['BoosterVersion']!="Falcon 1"]
```

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

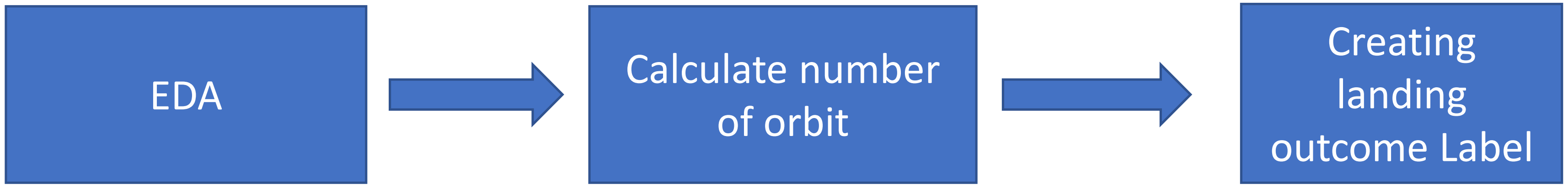
Data Collection - Scraping

GitHub url : https://github.com/tahri19younes98/Applied-Data-Science-Capastone/blob/main/jupyter-labs-webscraping.ipynb



Data Wrangling

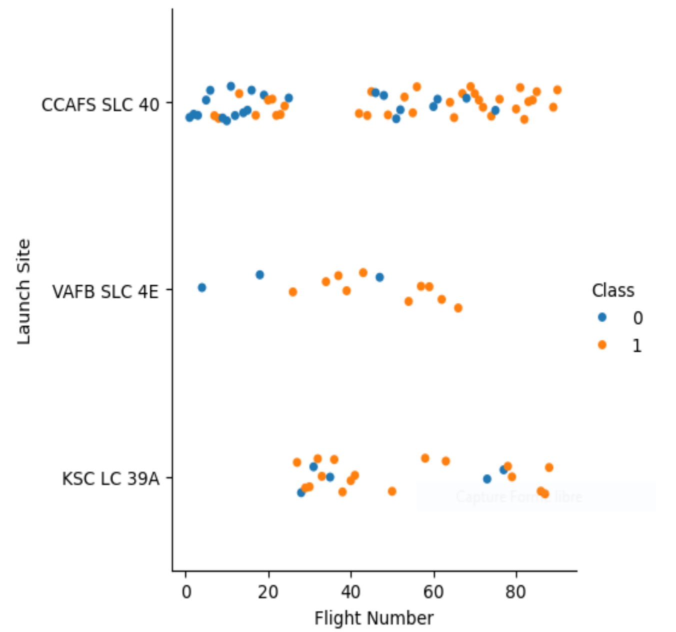
- Describe how data were processed
- Data Wrangling is process cleaning and transforming data into undertandable data to analyse
- You need to present your data wrangling process using key phrases and flowcharts

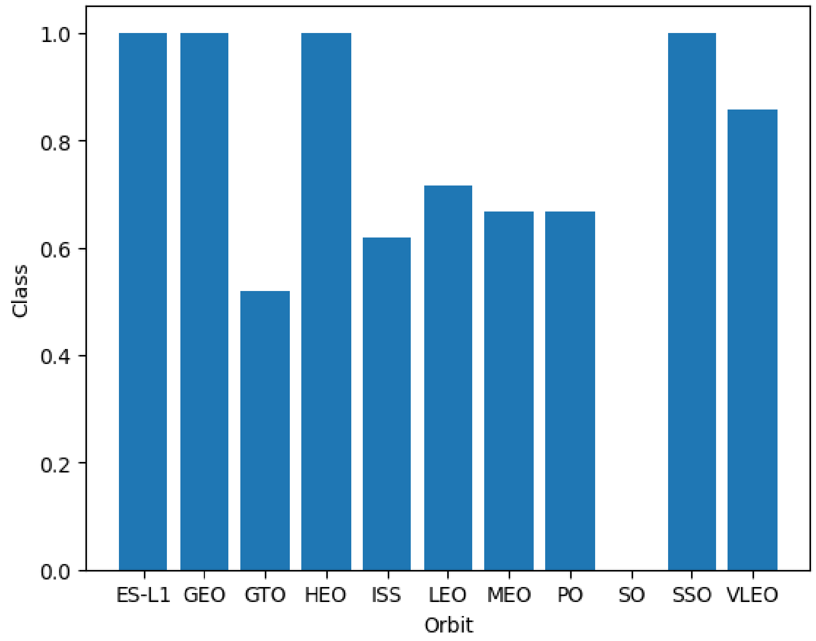


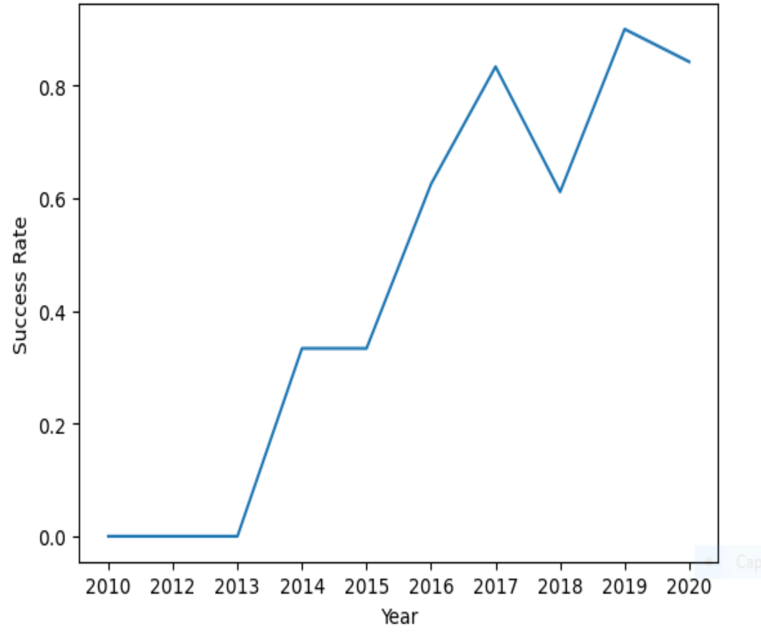
- GitHub url :[Applied-Data-Science-Capastone/labs-jupyter-spacex-Data wrangling.ipynb at main · tahri19younes98/Applied-Data-Science-Capastone]

EDA with Data Visualization

- we want to visualize data ,by scatter and bar and line plot with matplotlib and seaborn







- GitHub URL : [Applied-Data-Science-Capastone/edadataviz.ipynb at main · tahri19younes98/Applied-Data-Science-Capastone]

EDA with SQL

We performed SQL queries to gather information :

- 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)
- List the date when the first succesful landing outcome in ground pad was acheived.
- List the total number of successful and failure mission outcomes
- List the names of the booster_versions which have carried the maximum payload mass
- display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015
- 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.

GitHub URL : [Applied-Data-Science-Capastone/jupyter-labs-eda-sql-coursera_sqllite.ipynb at main · tahri19younes98/Applied-Data-Science-Capastone]

Build an Interactive Map with Folium

- To visualize the Launch Data into an interactive map. We took the Latitude and Longitude Coordinates at each launch site and added a Circle Marker around each launch site with a label of the name of the launch site.
- Marker indicate points like launch sites
- Circle is highlighted on spcifique coordinate
- Marker cluster is groupe markers that indicate same coordinate
- Lines are used to indicated distance between two coordinates

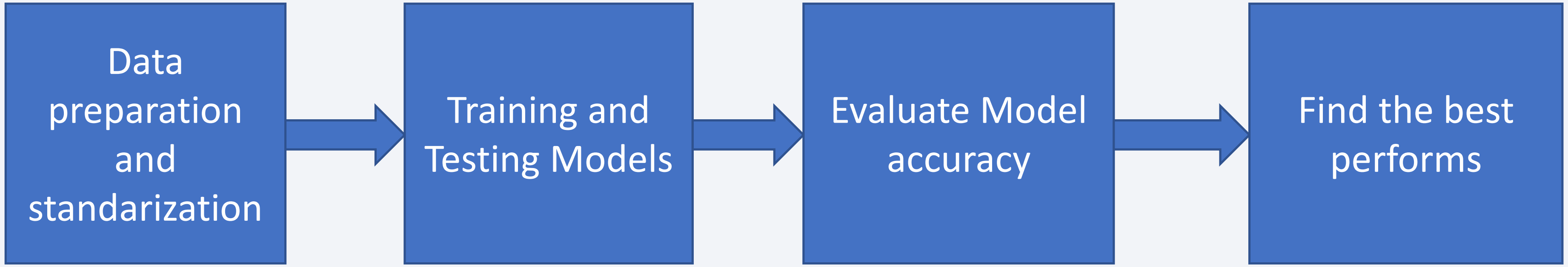
- GitHub URL :Applied-Data-Science-Capastone/lab_jupyter_launch_site_location.ipynb at main · tahri19younes98/Applied-Data-Science-Capastone

Build a Dashboard with Plotly Dash

- The following graphs and plots were used to visualize data
  - Percentage of launches by site
  - Payload range
- GitHub URL : [Applied-Data-Science-Capastone/spacex_dash_app.py at main · tahri19younes98/Applied-Data-Science-Capastone]

Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model
- Four classification models were compared: logistic regression, support vector machine, decision tree and k nearest neighbors.

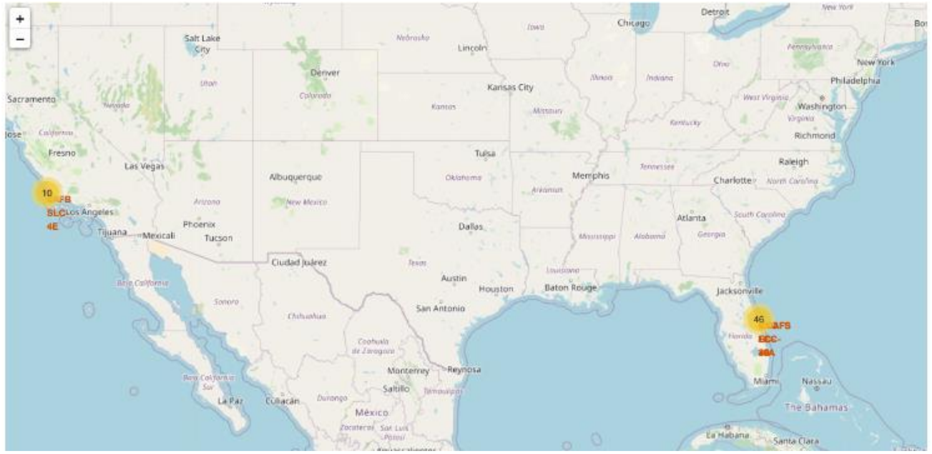


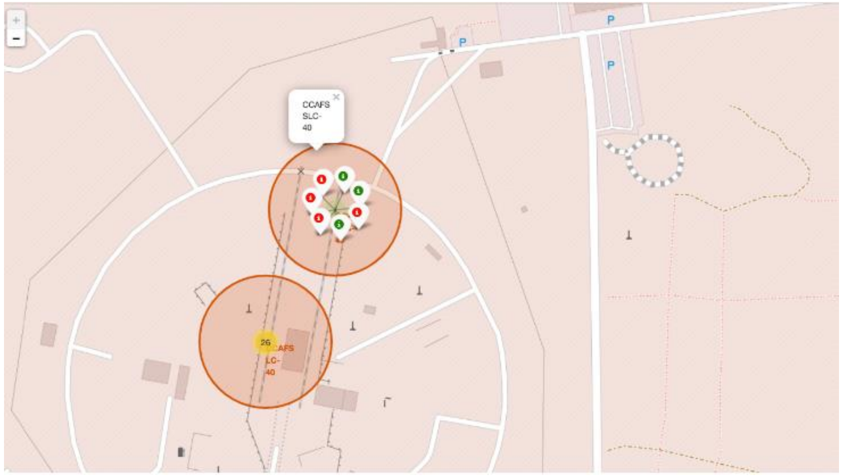
- GitHub URL : [Applied-Data-Science-Capastone/SpaceX_Machine Learning Prediction_Part_5.ipynb at main · tahri19younes98/Applied-Data-Science-Capastone]

Results

- **Exploratory data analysis results:**
- Space X uses 4 different launch sites;
- The first launches were done to Space X itself and NASA;
- The average payload of F9 v1.1 booster is 2,928 kg;
- The first success landing outcome happened in 2015 fiver year after the first launch;
- Many Falcon 9 booster versions were successful at landing in drone ships having payload above the average;
- Almost 100% of mission outcomes were successful;
- Two booster versions failed at landing in drone ships in 2015: F9 v1.1 B1012 and F9 v1.1 B1015;
- The number of landing outcomes became as better as years passed

- Interactive analytics demo in screenshots





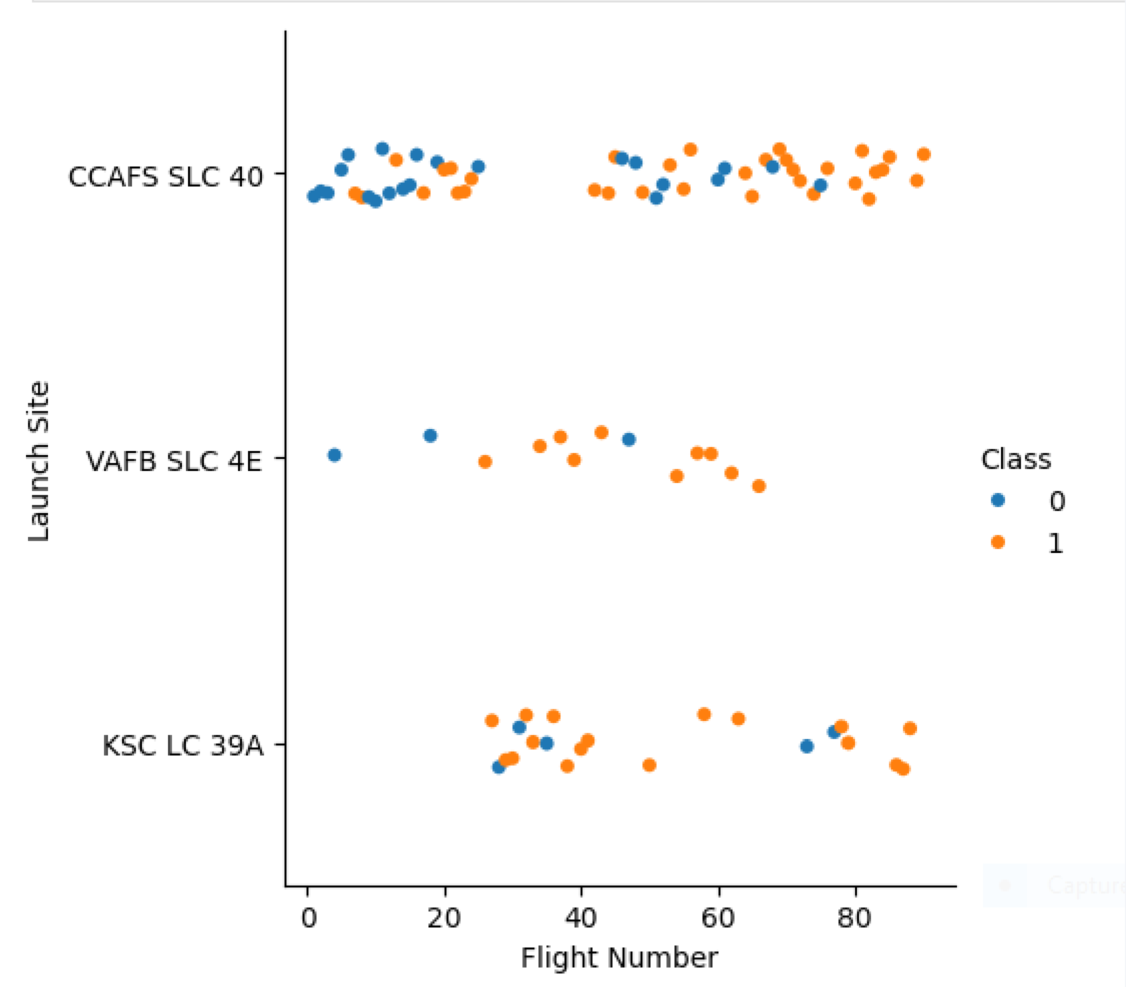
- Predictive analysis results
- GitHub url : [Applied-Data-Science-Capastone/SpaceX_Machine Learning Prediction_Part_5.ipynb at main · tahri19younes98/Applied-Data-Science-Capastone]

Section 2

Insights drawn from EDA

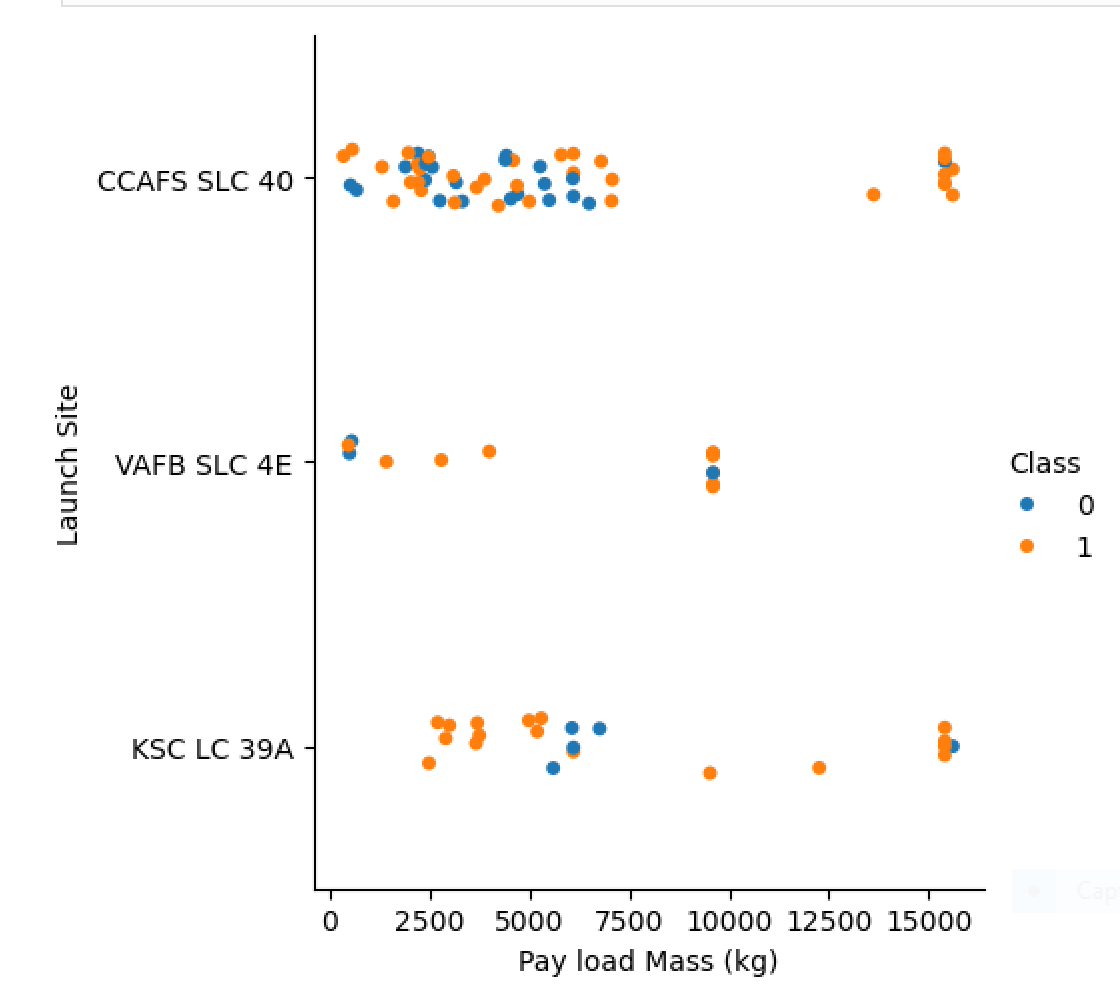
Flight Number vs. Launch Site

- The more amount of flights at a launch site the greater the success rate at a launch site



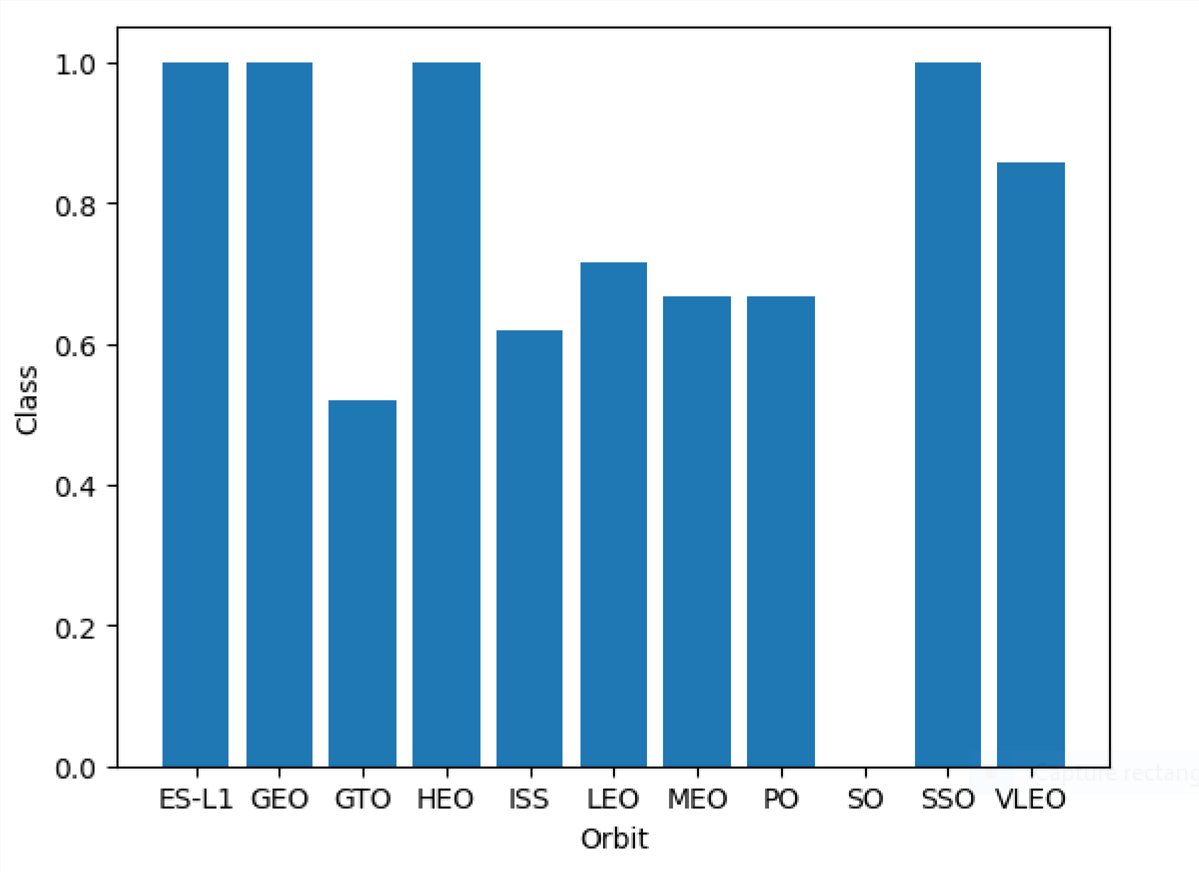
Payload vs. Launch Site

- The greater the payload mass for Launch Site CCAFS SLC 40 the higher the success rate for the Rocket



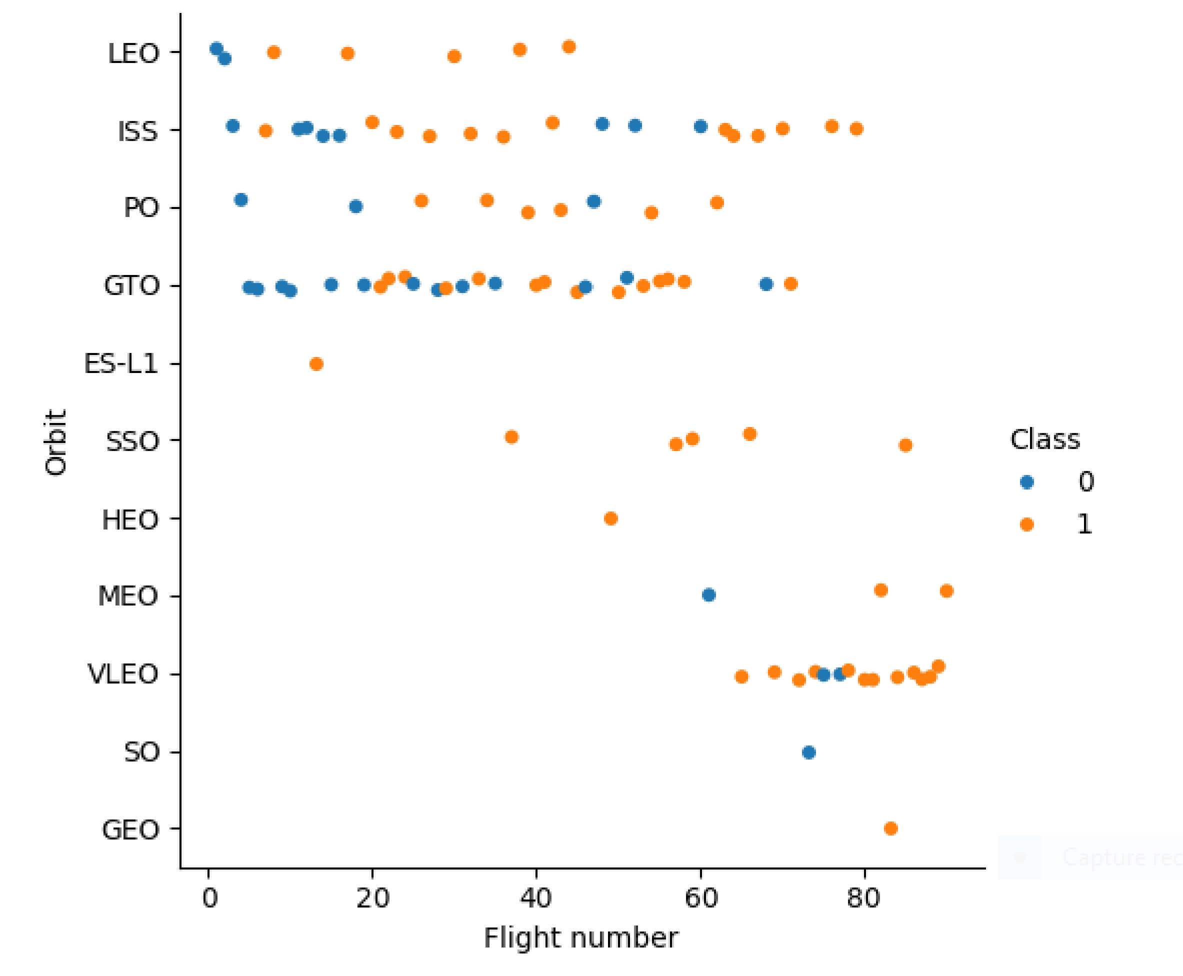
Success Rate vs. Orbit Type

- bar chart to identify which orbits have the highest success rates.



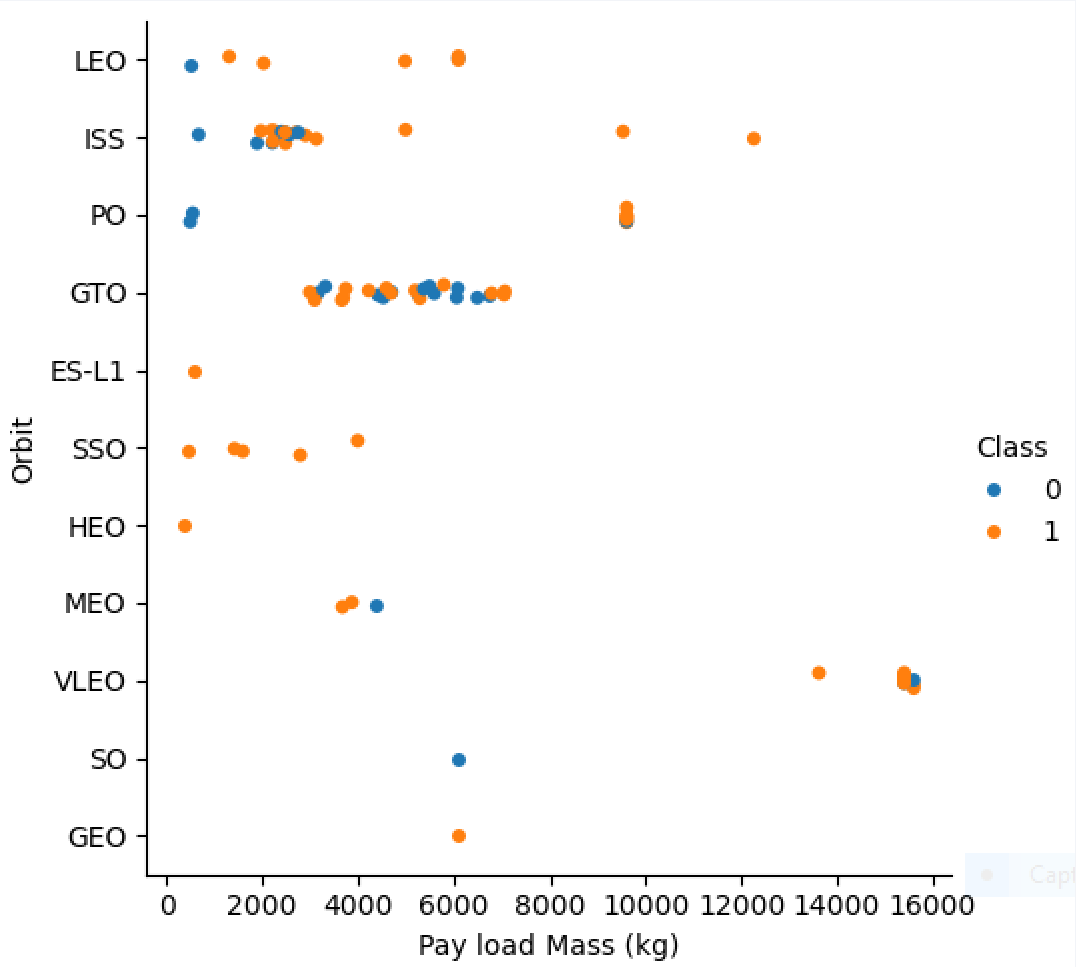
Flight Number vs. Orbit Type

- 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



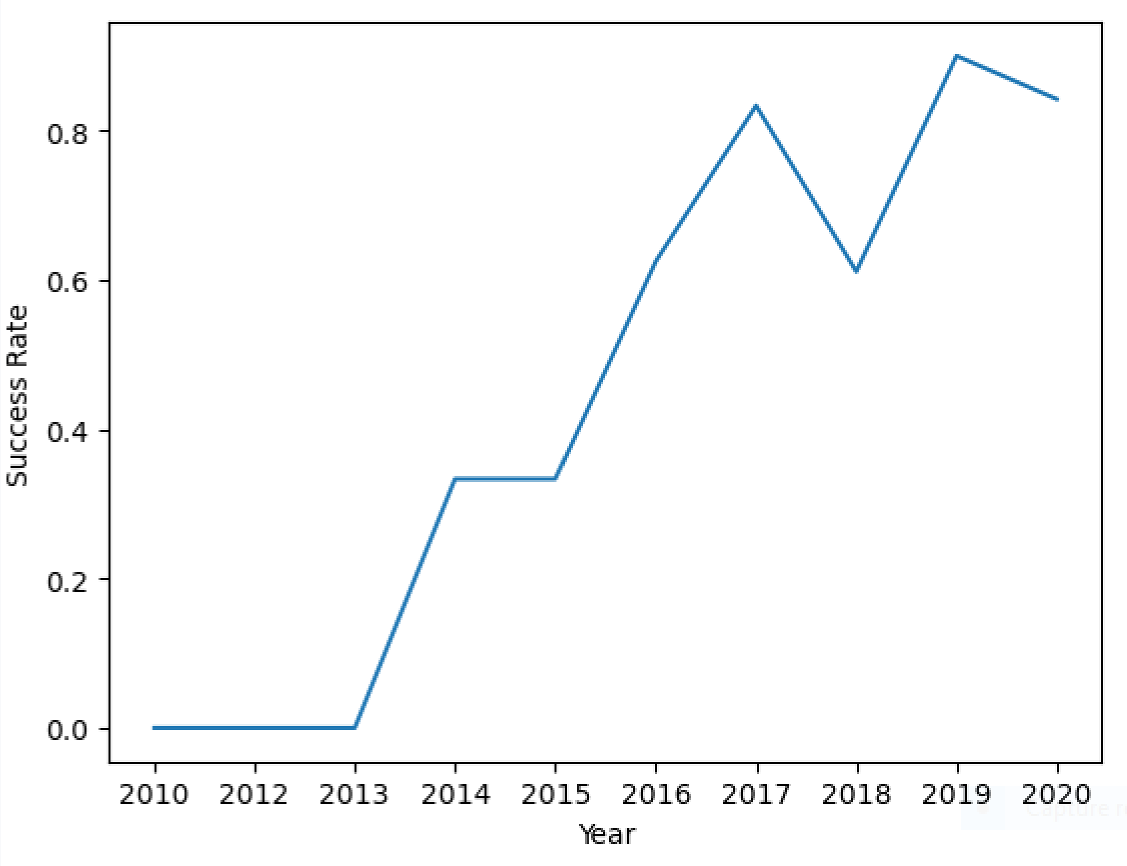
Payload vs. Orbit Type

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



Launch Success Yearly Trend

- you can observe that the success rate since 2013 kept increasing till 2020



All Launch Site Names

- Present your query result with a short explanation here
- **SELECT DISTINCT** Launch_Site **FROM** SPACEXTABLE
- Using the word DISTINCT in the query means that it will only values in the Launch_Site column from tablSpaceX

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

Launch Site Names Begin with 'CCA'

- Using the word TOP 5 in the query means that it will only show 5 records from tblSpaceX and LIKE keyword has a wild card with the words 'KSC%' the percentage in the end suggests that the Launch_Site name must start with KSC

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success
2010-12-08	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
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success

Total Payload Mass

```
SELECT SUM(PAYLOAD_MASS__KG_) AS TotalPayloadMass FROM
SPACEXTABLE WHERE Customer = 'NASA (CRS)';
```

..

Using the function SUM summates the total in the column PAYLOAD_MASS_KG_ The WHERE clause filters the dataset to only perform calculations on Customer NASA (CRS)

Out[12]:

TotalPayloadMass
45596

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
- SELECT AVG(PAYLOAD_MASS__KG_) AS AveragePayloadMass FROM SPACEXTABLE WHERE Booster_Version = 'F9 v1.1'
- Present your query result with a short explanation here
- Usign the average PayloadMassKg and selecting

Only Boost_Version F9 v1.1

AveragePayloadMass
2928.4

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad

- Function Min to extacte

The first date

And selecting Landing_Outcome

```
%%sql
SELECT MIN(DATE) AS first_successful_landing_date
FROM SPACEXTABLE
WHERE Landing_Outcome = 'Success (ground pad)'
```

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

first_successful_landing_date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

```
%%sql
SELECT Booster_Version
FROM SPACEXTABLE
WHERE Landing_Outcome = 'Success (drone ship)' AND (PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000)
```

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

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

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

```
%%sql
SELECT  MISSION_OUTCOME, COUNT(*) AS total_number
FROM SPACEXTABLE
GROUP BY MISSION_OUTCOME
```

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

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

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass

```sql
SELECT DISTINCT Booster_Version ,PAYLOAD_MASS__KG_
FROM SPACEXTABLE
WHERE PAYLOAD_MASS__KG_ = (
    SELECT MAX(PAYLOAD_MASS__KG_)
    FROM SPACEXTABLE
    );
```

Booster_Version	PAYLOAD_MASS__KG_
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

- These are the boosters which have carried the

maximum payload massregistered in the dataset.

2015 Launch Records

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

```
%%sql
SELECT  substr(Date, 6,2) , Landing_Outcome , Booster_Version ,Launch_Site
FROM SPACEXTABLE
Where Landing_Outcome = 'Failure (drone ship)' and substr(Date,0,5)='2015'
```

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

substr(Date, 6,2)	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

- 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

```
%%sql
SELECT Landing_Outcome,count(Landing_Outcome) AS Rank
FROM SPACEXTABLE
WHERE Date BETWEEN ' 2010-06-04' AND '2017-03-20'
GROUP BY Landing_Outcome
ORDER BY Rank DESC
```

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

Out[25]:

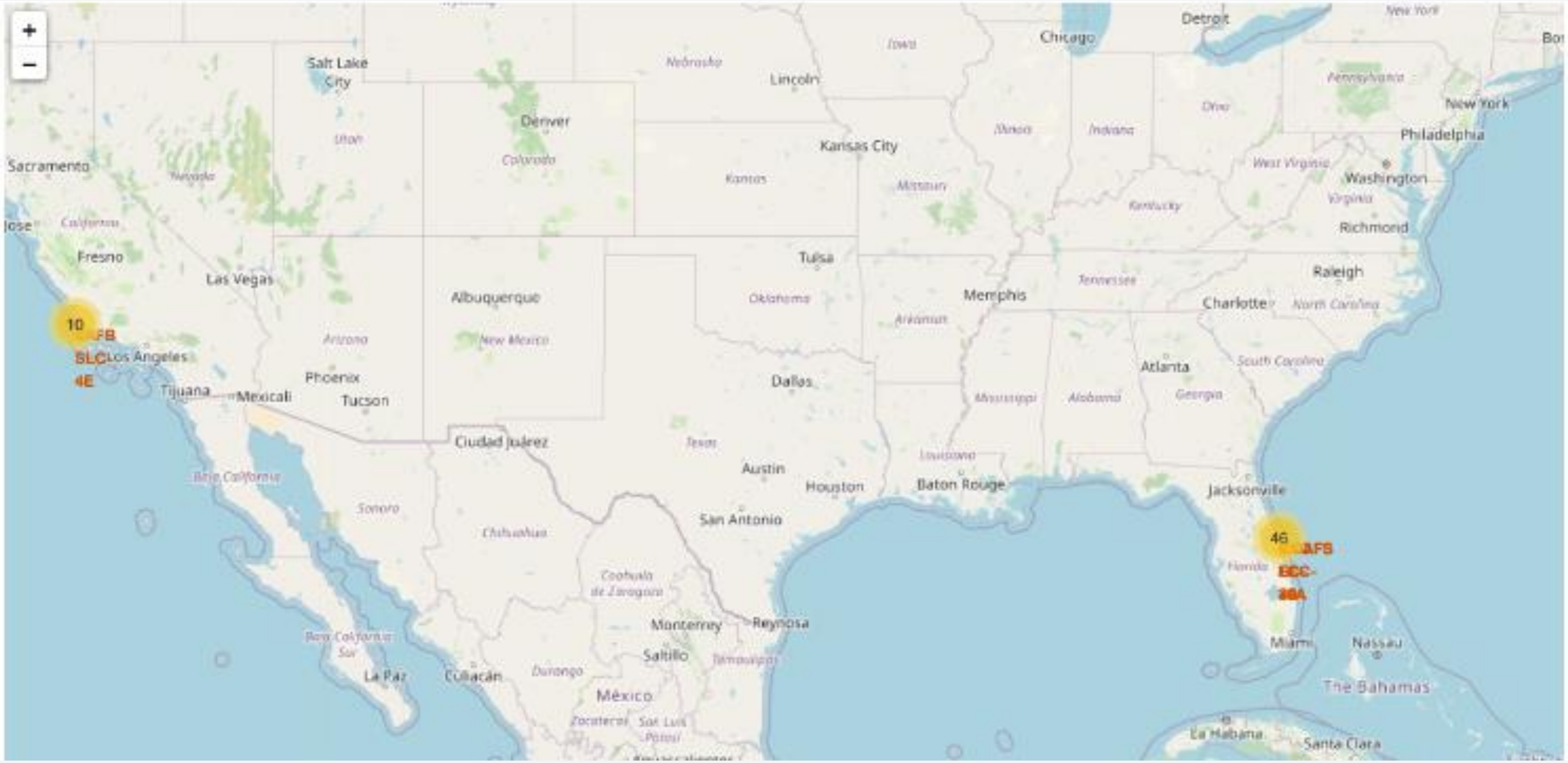
Landing_Outcome	Rank
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

- Countiing number of landing outcome
- between the date 2010-06-04 and 2017-03-20,
- in descending order

Section 3

Launch Sites Proximities Analysis

All launch sites global map markers



Colour Labelled Markers

- Green Marker shows successful Launches and Red Marker shows Failures

<Launche sites between cities and coast lines>

Explain the important elements and findings on the screenshot

- Distance to coast 0.90km

Section 4

Build a Dashboard with Plotly Dash

<Dashboard Screenshot 1>

- Replace <Dashboard screenshot 1> title with an appropriate title
- Show the screenshot of launch success count for all sites, in a piechart
- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 2>

- Replace <Dashboard screenshot 2> title with an appropriate title
- Show the screenshot of the piechart for the launch site with highest launch success ratio
- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 3>

- Replace <Dashboard screenshot 3> title with an appropriate title
- Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider
- Explain the important elements and findings on the screenshot, such as which payload range or booster version have the largest success rate, etc.

Section 5

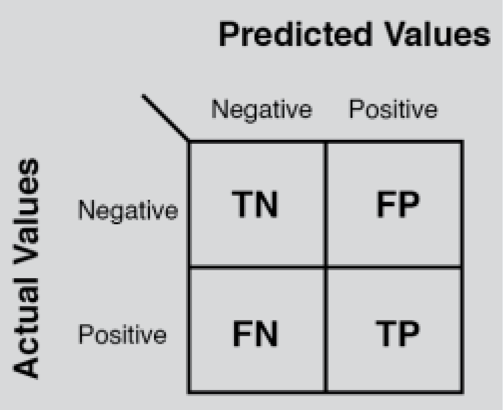
Predictive Analysis (Classification)

Classification Accuracy

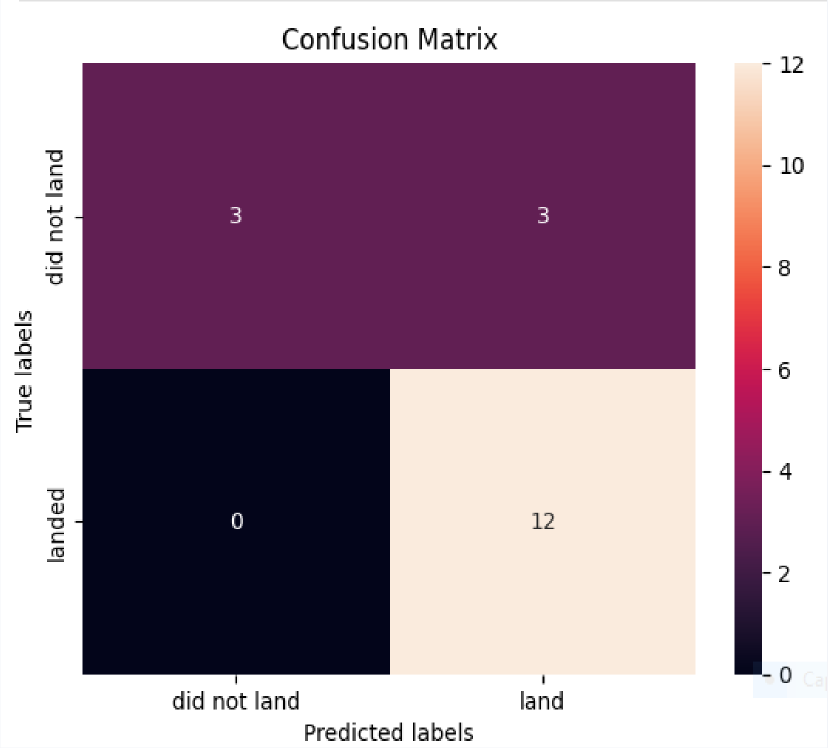
- Visualize the built model accuracy for all built classification models, in a bar chart
- Find which model has the highest classification accuracy

Confusion Matrix

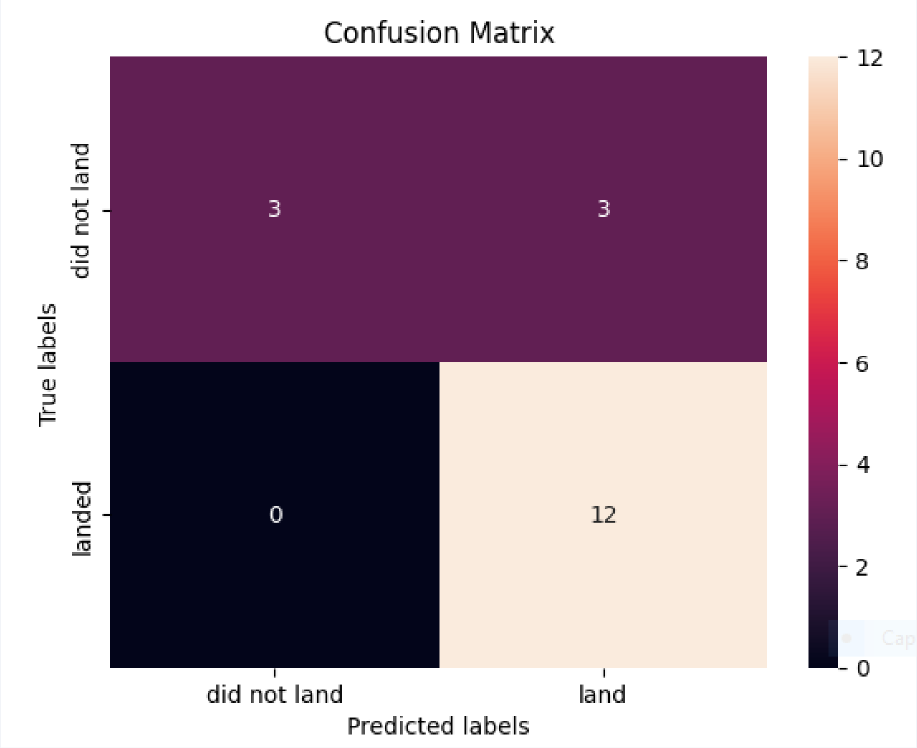
Examining the confusion matrix, we see that Tree can distinguish between the different classes. We see that the major problem is false positives.



Logistic Regression

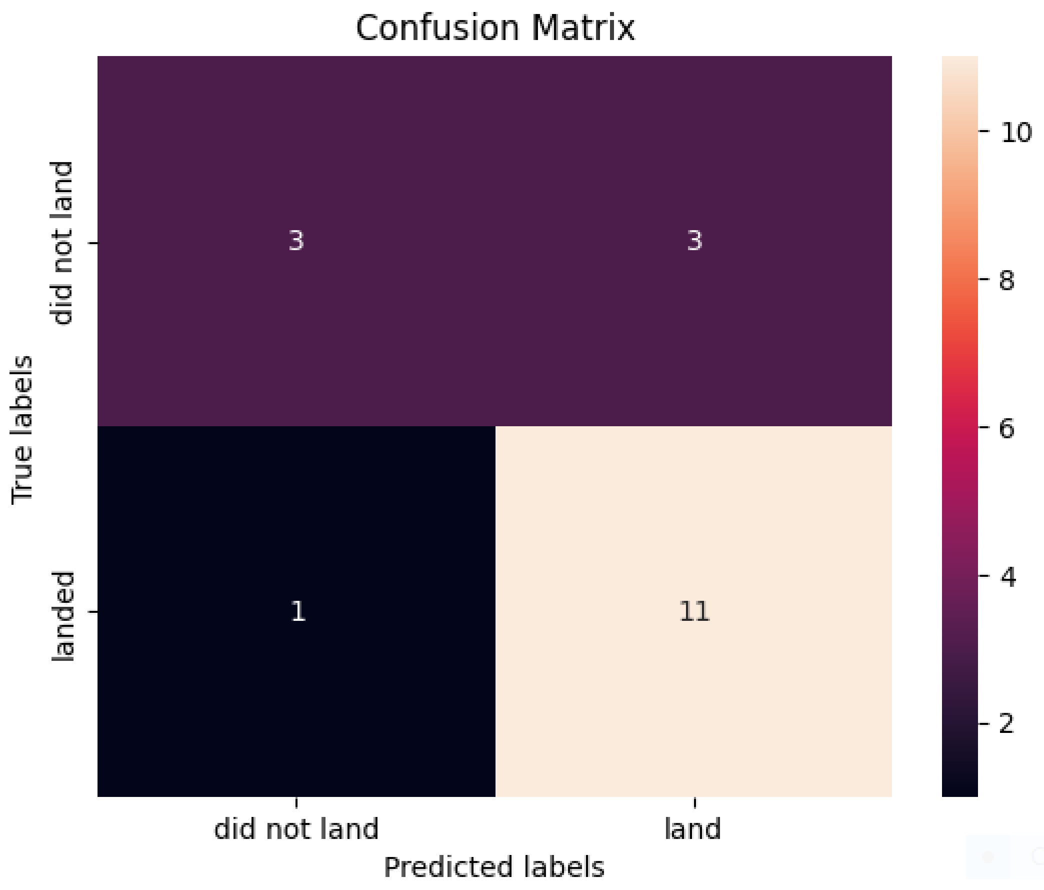


SVM

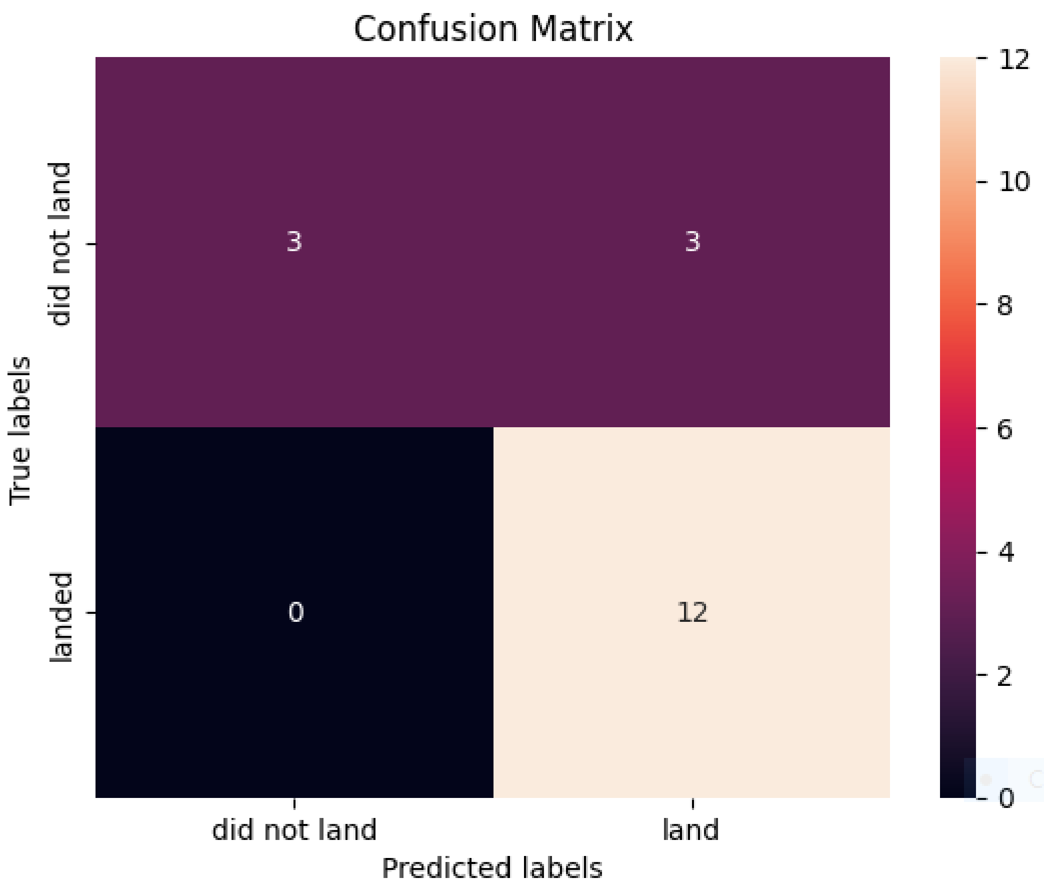


Confusion Matrix

Decision Tree



KNN



Conclusions

- The Tree Classifier Algorithm is the best for Machine Learning for this dataset
- Orbit GEO,HEO,SSO,ES-L1 has the best Success Rate
- The success rates for SpaceX launches increase over the time
- Low weighted payloads perform better than the heavier payloads
- the best launch site is KSC LC-39A;

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!