



IBM Data Science Capstone Project

SPACE X FALCON 9 LANDING
ANALYSIS

Outline

Executive Summary

Introduction

Methodology

Results

Conclusion

Appendix

Executive Summary

Summary of methodologies



Data Collection via API, Web Scraping



Exploratory Data Analysis (EDA) with Data Visualization



Exploratory Data Analysis (EDA) with SQL



Interactive Map with Folium



Dashboards with Plotly Dash



Predictive Analysis

Summary of all results

This Project produced the following outputs and visualizations

- Exploratory Data Analysis result
- Geospatial analytics
- Interactive dashboards
- Predictive analytics and classification model

Introduction

❑ Project background and context

- ❑ The aim of this project is to predict if the Falcon 9 first stage will successfully land. SpaceX says on its website that the Falcon 9 rocket launch cost 62 million dollars. Other providers cost upward of 165 million dollars each. The price difference is explained by the fact that SpaceX can reuse the first stage. By determining if the stage will land, we can determine the cost of a launch. This information is interesting for another company if it wants to compete with SpaceX for a rocket launch.

❑ Problems you want to find answers

- ❑ What are the main characteristics of a successful or failed landing ?
- ❑ What are the effects of each relationship of the rocket variables on the success or failure of a landing ?
- ❑ What are the conditions which will allow SpaceX to achieve the best landing success rate ?

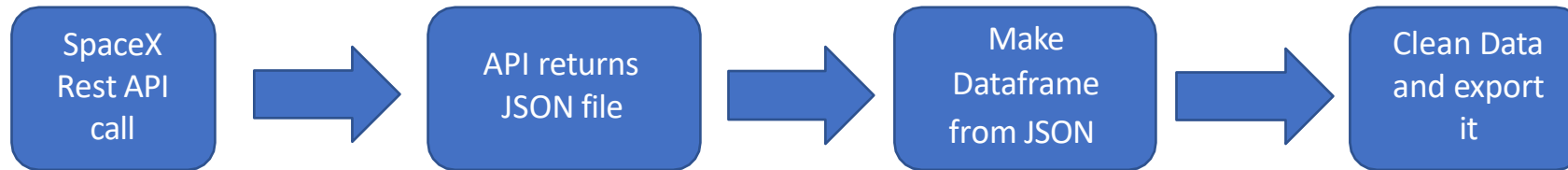
METHODOLOGY

Methodology-Summary

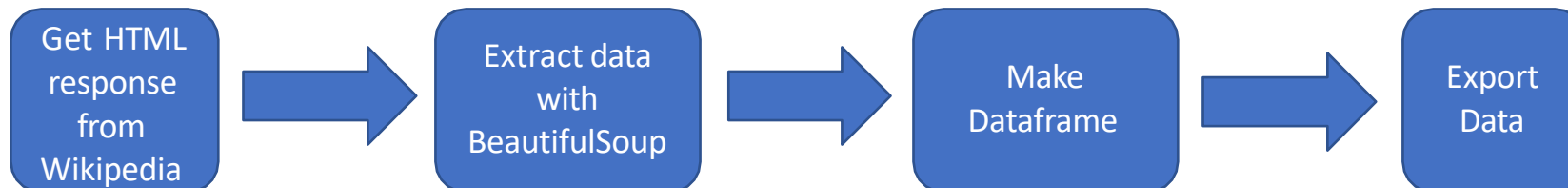
- Data collection methodology:
 - SpaceX REST API
 - Web Scrapping from Wikipedia
- Perform data wrangling
 - Dropping unnecessary columns
 - One Hot Encoding for classification models
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
 - How to build, tune, evaluate classification models

Data Collection

- Datasets are collected from Rest SpaceX API and webscrapping Wikipedia
 - The information obtained by the API are rocket, launches, payload information.
 - The Space X REST API URL is api.spacexdata.com/v4/



- The information obtained by the webscrapping of Wikipedia are launches, landing, payload information.
 - URL is https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922



Data Collection – SpaceX API

1. Getting Response from API

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

2. Convert Response to JSON File

```
data = response.json()  
data = pd.json_normalize(data)
```

3. Transform data

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

4. Create dictionary with data

```
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}
```

5. Create dataframe

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

6. Filter dataframe

```
data_falcon9 = data[data['BoosterVersion']!='Falcon 1']
```

7. Export to file

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


Data Collection - Scraping

1. Getting Response from HTML

```
response = requests.get(static_url)
```

2. Create BeautifulSoup Object

```
soup = BeautifulSoup(response.text, "html5lib")
```

3. Find all tables

```
html_tables = soup.findAll('table')
```

4. Get column names

```
for th in first_launch_table.find_all('th'):
    name = extract_column_from_header(th)
    if name is not None and len(name) > 0 :
        column_names.append(name)
```

5. Create dictionary

```
launch_dict= dict.fromkeys(column_names)

# Remove an irrelevant column
del launch_dict['Date and time ( )']

# Let's initial the launch_dict with each value to be an empty list
launch_dict['Flight No.'] = []
launch_dict['Launch site'] = []
launch_dict['Payload'] = []
launch_dict['Payload mass'] = []
launch_dict['Orbit'] = []
launch_dict['Customer'] = []
launch_dict['Launch outcome'] = []
# Added some new columns
launch_dict['Version Booster']=[]
launch_dict['Booster landing']=[]
launch_dict['Date']=[]
launch_dict['Time']=[]
```

6. Add data to keys

```
extracted_row = 0
#Extract each table
for table_number,table in enumerate(soup.find_all(
    # get table row
    for rows in table.find_all("tr"):
        #check to see if first table heading is a
        if rows.th:
            if rows.th.string:
                flight_number=rows.th.string.stri
                flag=flight_number.isdigit()
```

See notebook for the rest of code

7. Create dataframe from dictionary

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

8. Export to file

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

Data Wrangling

- In the dataset, there are several cases where the booster did not land successfully.
 - True Ocean, True RTL S, True ASDS means the mission has been successful.
 - False Ocean, False RTL S, False ASDS means the mission was a failure.
- We need to transform string variables into categorical variables where 1 means the mission has been successful and 0 means the mission was a failure.

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 occurrence of each orbit

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

```
GTO    27  
ISS    21  
VLEO   14  
PO      9  
LEO     7  
SSO     5  
MEO     3  
SO      1  
ES-L1   1  
HEO     1  
GEO     1  
Name: Orbit, dtype: int64
```

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 RTL S   14  
False ASDS    6  
True Ocean    5  
None ASDS     2  
False Ocean   2  
False RTL S   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
```

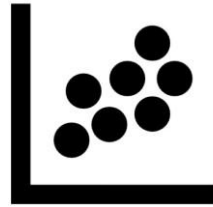
5. Export to file

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

EDA with Data Visualization

- 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



Scatter plots show relationship between variables. This relationship is called the correlation.

- Bar Graph

- Success rate vs. Orbit

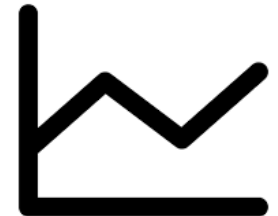
Bar graphs show the relationship between numeric and categorical variables.



- Line Graph

- Success rate vs. Year

Line graphs show data variables and their trends. Line graphs can help to show global behavior and make prediction for unseen data.



EDA with SQL

We performed SQL queries to gather and understand data from dataset:

- Displaying 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 successful landing outcome in ground pad was achieved.
- 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.
- 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.

Build an Interactive Map with Folium

Folium map object is a map centered on NASA Johnson Space Center at Houston, Texas

- Red circle at NASA Johnson Space Center's coordinate with label showing its name (*folium.Circle*, *folium.map.Marker*).
- Red circles at each launch site coordinates with label showing launch site name (*folium.Circle*, *folium.map.Marker*, *folium.features.DivIcon*).
- The grouping of points in a cluster to display multiple and different information for the same coordinates (*folium.plugins.MarkerCluster*).
- Markers to show successful and unsuccessful landings. **Green** for successful landing and **Red** for unsuccessful landing.
(*folium.map.Marker*, *folium.Icon*).
- Markers to show distance between launch site to key locations (railway, highway, coastway, city) and plot a line between them.
(*folium.map.Marker*, *folium.PolyLine*, *folium.features.DivIcon*)

These objects are created in order to understand better the problem and the data. We can show easily all launch sites, their surroundings and the number of successful and unsuccessful landings.

Build a Dashboard with Plotly Dash

- Dashboard has dropdown, pie chart, rangeslider and scatter plot components
 - Dropdown allows a user to choose the launch site or all launch sites (*dash_core_components.Dropdown*).
 - Pie chart shows the total success and the total failure for the launch site chosen with the dropdown component (*plotly.express.pie*).
 - Rangeslider allows a user to select a payload mass in a fixed range (*dash_core_components.RangeSlider*).
 - Scatter chart shows the relationship between two variables, in particular Success vs Payload Mass (*plotly.express.scatter*).

Predictive Analysis (Classification)

- Data preparation
 - Load dataset
 - Normalize data
 - Split data into training and test sets.
- Model preparation
 - Selection of machine learning algorithms
 - Set parameters for each algorithm to GridSearchCV
 - Training GridSearchModel models with training dataset
- Model evaluation
 - Get best hyperparameters for each type of model
 - Compute accuracy for each model with test dataset
 - Plot Confusion Matrix
- Model comparison
 - Comparison of models according to their accuracy
 - The model with the best accuracy will be chosen (see Notebook for result)

Space X Falcon 9 Landing Analysis - Results

Exploratory Data Analysis

Interactive Analytics

Predictive Analysis



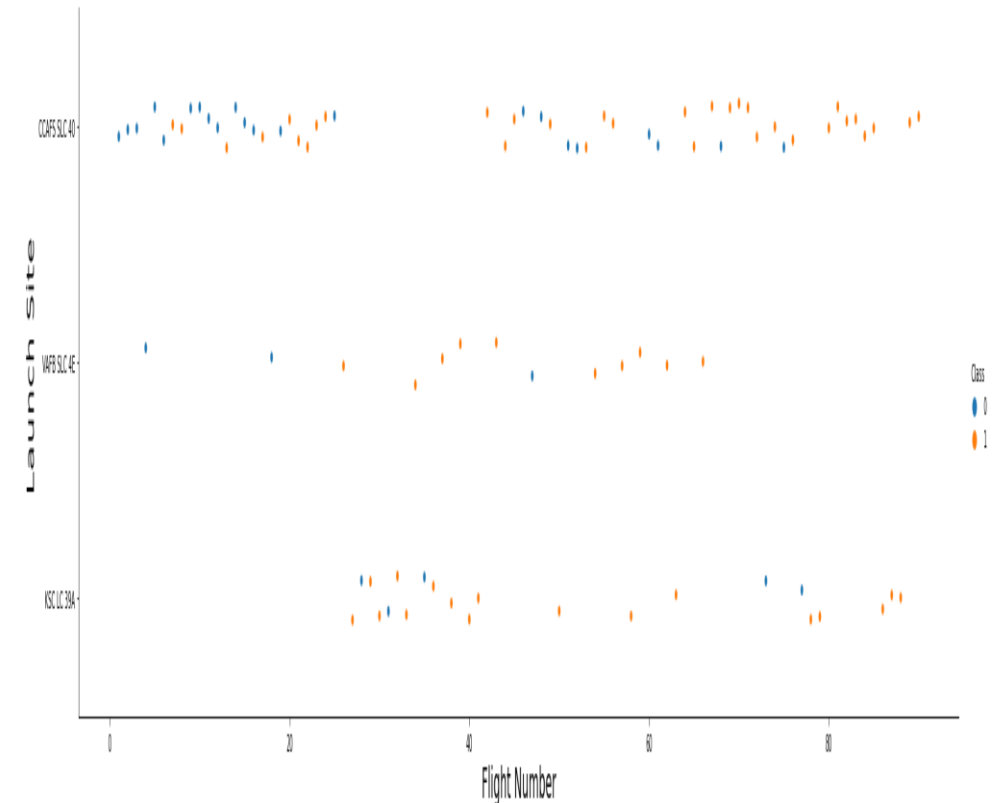
EDA - WITH VISUALIZATION

Flight Number vs. Launch Site

The scatter plot of Launch Site vs. Flight Number shows that:

- As the number of flights increases, the rate of success at a launch site increases.
- Most of the early flights (flight numbers < 30) were launched from CCAFS SLC 40, and were generally unsuccessful.
- The flights from VAFB SLC 4E also show this trend, that earlier flights were less successful.
- No early flights were launched from KSC LC 39A, so the launches from this site are more successful.
- Above a flight number of around 30, there are significantly more successful landings (Class = 1).

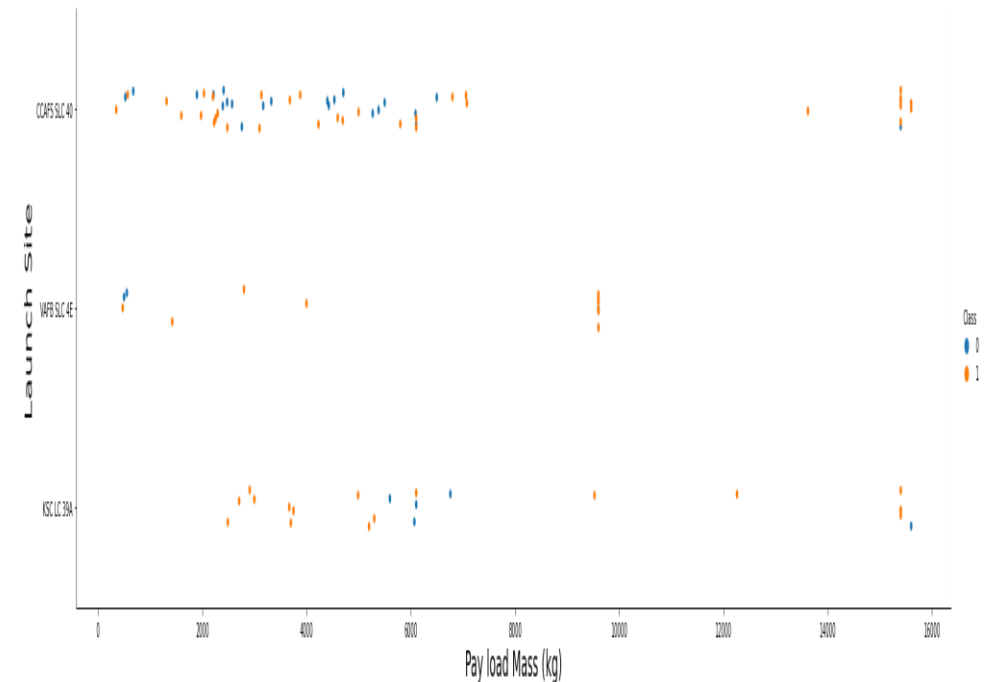
We observe that, for each site, the success rate is increasing.



Payload vs. Launch Site

The scatter plot of Launch Site vs. Payload Mass shows that:

- Above a payload mass of around 7000 kg, there are very few unsuccessful landings, but there is also far less data for these heavier launches.
- There is no clear correlation between payload mass and success rate for a given launch site.
- All sites launched a variety of payload masses, with most of the launches from CCAFS SLC 40 being comparatively lighter payloads (with some outliers).



Depending on the launch site, a heavier payload may be a consideration for a successful landing. On the other hand, a too heavy payload can make a landing fail.

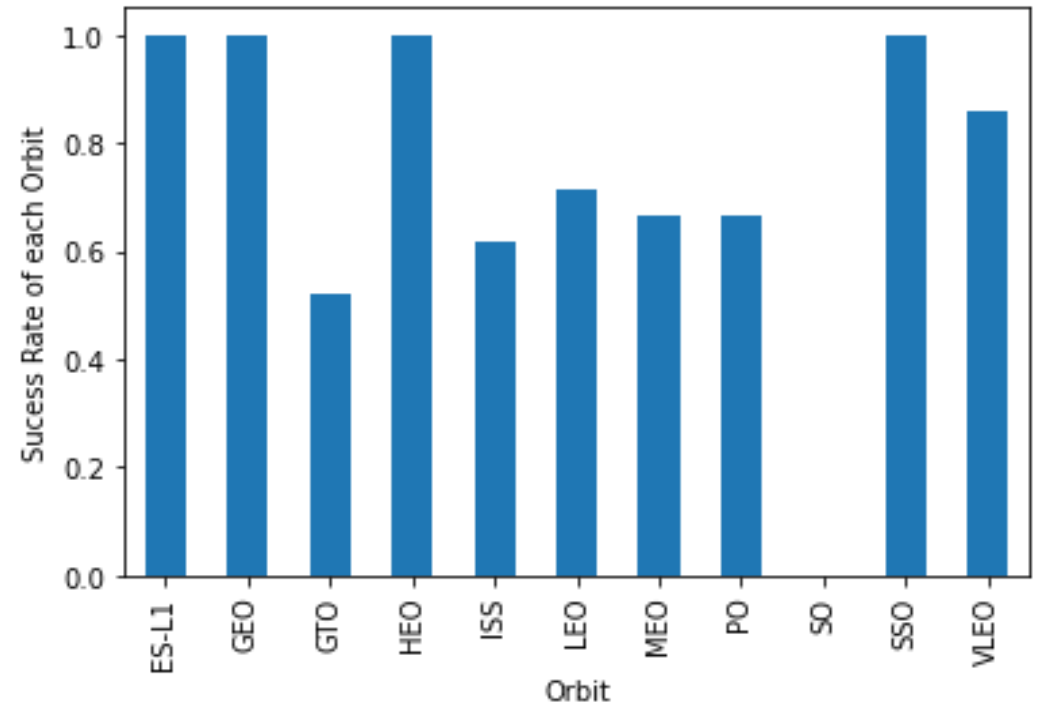
Success Rate vs. Orbit Type

The bar chart of Success Rate vs. Orbit Type shows that the following orbits have the highest (100%) success rate:

- ES-L1 (Earth-Sun First Lagrangian Point)
- GEO (Geostationary Orbit)
- HEO (High Earth Orbit)
- SSO (Sun-synchronous Orbit)

The orbit with the lowest (0%) success rate is:

- SO (Heliocentric Orbit)

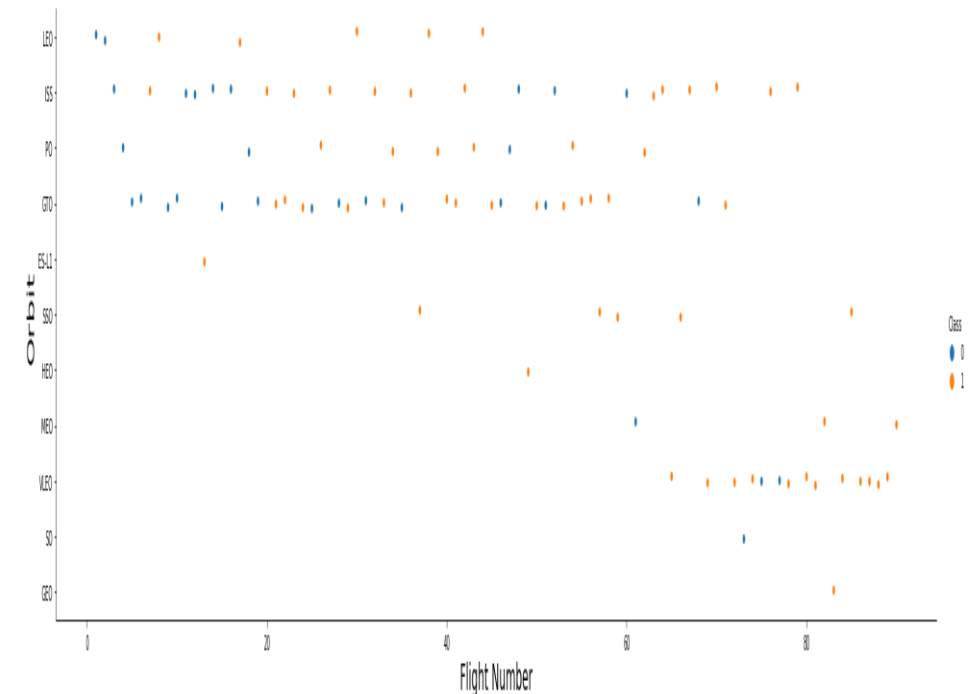


With this plot, we can see success rate for different orbit types. We note that ESL1, GEO, HEO, SSO have the best success rate.

Flight Number vs. Orbit Type

This scatter plot of Orbit Type vs. Flight number shows a few useful things that the previous plots did not, such as:

- The 100% success rate of GEO, HEO, and ESL-1 orbits can be explained by only having 1 flight into the respective orbits.
- The 100% success rate in SSO is more impressive, with 5 successful flights.
- There is little relationship between Flight Number and Success Rate for GTO.
- Generally, as Flight Number increases, the success rate increases. This is most extreme for LEO, where unsuccessful landings only occurred for the low flight numbers (early launches).

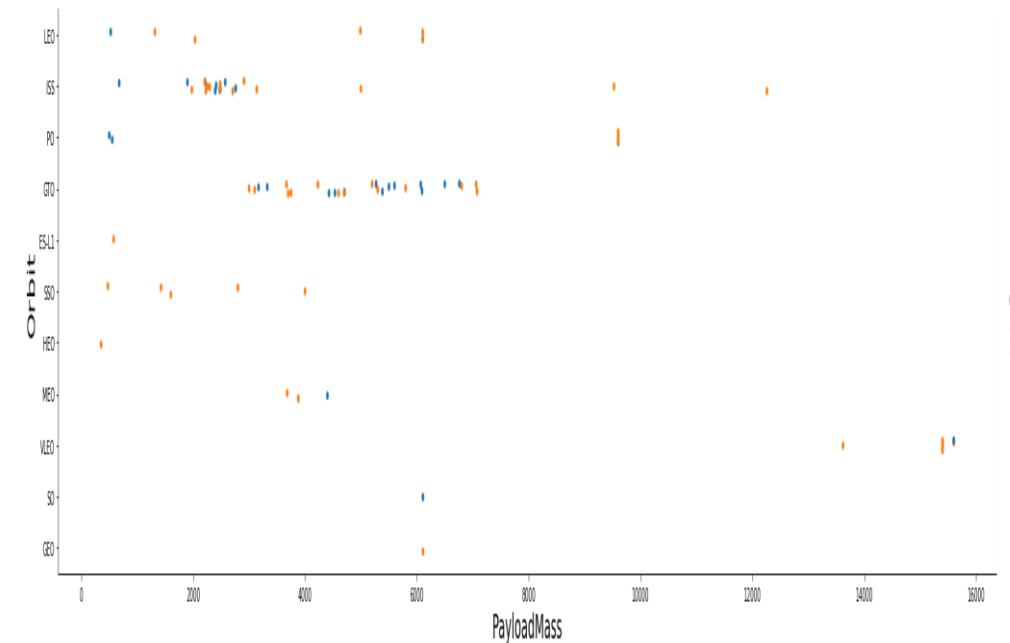


We notice that the success rate increases with the number of flights for the LEO orbit. For some orbits like GTO, there is no relation between the success rate and the number of flights. But we can suppose that the high success rate of some orbits like SSO or HEO is due to the knowledge learned during former launches for other orbits.

Payload vs. Orbit Type

This scatter plot of Orbit Type vs. Payload Mass shows that:

- The following orbit types have more success with heavy payloads:
- PO (although the number of data points is small)
- ISS
- LEO
- For GTO, the relationship between payload mass and success rate is unclear.
- VLEO (Very Low Earth Orbit) launches are associated with heavier payloads, which makes intuitive sense.

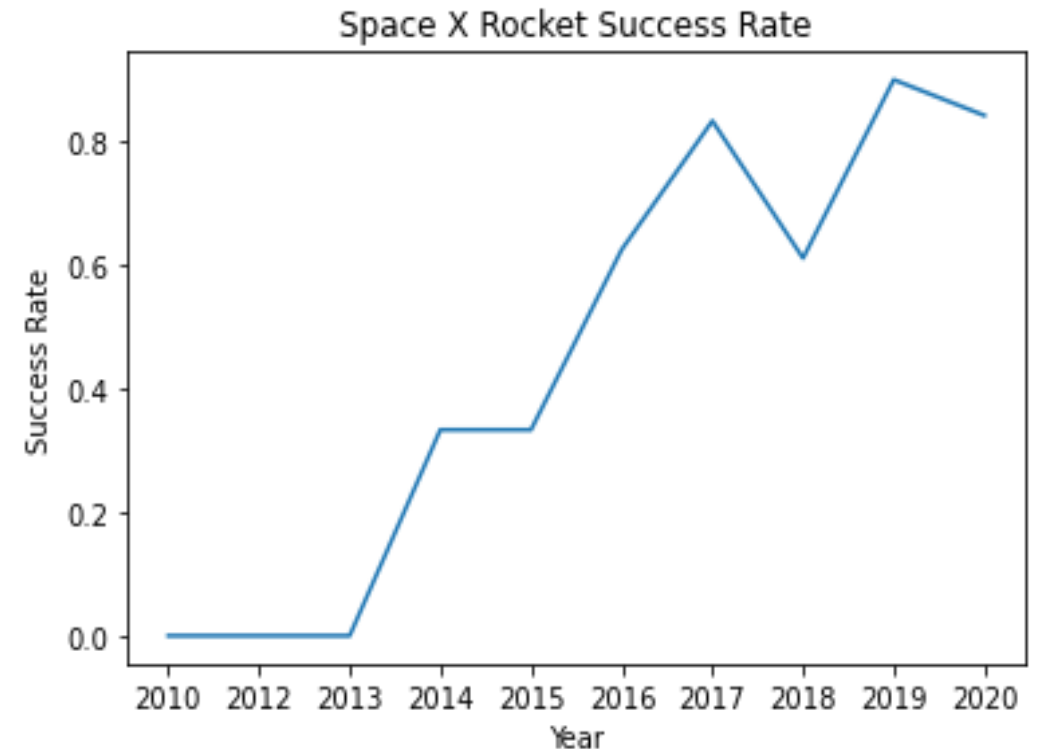


The weight of the payloads can have a great influence on the success rate of the launches in certain orbits. For example, heavier payloads improve the success rate for the LEO orbit. Another finding is that decreasing the payload weight for a GTO orbit improves the success of a launch.

Launch Success Yearly Trend

The line chart of yearly average success rate shows that:

- Between 2010 and 2013, all landings were unsuccessful (as the success rate is 0).
- After 2013, the success rate generally increased, despite small dips in 2018 and 2020.
- After 2016, there was always a greater than 50% chance of success.



Since 2013, we can see an increase in the Space X Rocket success rate.

EDA - WITH SQL

All Launch Site Names

Find the names of the unique launch sites

SQL Query

```
SELECT DISTINCT "LAUNCH_SITE" FROM SPACEXTBL
```

Results

Launch_Site

CCAFS LC-40

VAFB SLC-4E

KSC LC-39A

CCAFS SLC-40

*The word **DISTINCT** in the query allows to remove duplicate LAUNCH_SITE column of the SPACEXTBL table.*

Launch Site Names Begin with 'CCA'

Find 5 records where launch sites begin with 'CCA'

SQL Query

```
SELECT * FROM SPACEXTBL WHERE "LAUNCH_SITE" LIKE '%CCA%' LIMIT 5
```

Results

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS_KG_	Orbit	Customer
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX
08-12-2010	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
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)

The WHERE clause followed by LIKE clause filters launch sites that contain the substring CCA. LIMIT 5 shows 5 records from filtering.

Total Payload Mass

Calculate the total payload carried by boosters from NASA

SQL Query

```
SELECT SUM("PAYLOAD_MASS_KG_") FROM SPACEXTBL WHERE "CUSTOMER" = 'NASA (CRS)'
```

Results

```
SUM("PAYLOAD_MASS_KG_")
```

45596

*The **SUM** keyword is used to calculate the total of the LAUNCH column, and the SUM keyword (and the associated condition) filters the results to only boosters from NASA (CRS).*

Average Payload Mass by F9 v1.1

Calculate the average payload mass carried by booster version F9 v1.1.

SQL Query

```
SELECT AVG("PAYLOAD_MASS_KG_") FROM SPACEXTBL WHERE "BOOSTER_VERSION" LIKE '%F9 v1.1%'
```

Results

```
AVG("PAYLOAD_MASS_KG_")
```

```
2534.6666666666665
```

*The **AVG** keyword is used to calculate the average of the **PAYLOAD_MASS_KG_** column, and the **WHERE** keyword (and the associated condition) filters the results to only the **F9 v1.1** booster version.*

First Successful Ground Landing Date

Identify the dates of the first successful landing on the ground pad.

SQL Query

```
SELECT MIN("DATE") FROM SPACEXTBL WHERE "Landing_Outcome" LIKE '%Success%'
```

Results

```
MIN("DATE")
```

```
01-05-2017
```

*The **MIN** keyword calculates the earliest date in the **DATE** column, representing the first landing date, while the **WHERE** keyword (with its condition) filters results to include only successful ground pad landings.*

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 Query

```
%sql SELECT "BOOSTER_VERSION" FROM SPACEXTBL WHERE "LANDING_OUTCOME" = 'Success (drone ship)' \
AND "PAYLOAD_MASS_KG_" > 4000 AND "PAYLOAD_MASS_KG_" < 6000;
```

Results

Booster_Version

F9 FT B1022

F9 FT B1026

F9 FT B1021.2

F9 FT B1031.2

*The **WHERE** keyword is used to filter the results to include only those that satisfy both conditions in the brackets (as the **AND** keyword is also used). The **BETWEEN** keyword allows for $4000 < x < 6000$ values to be selected.*

Total Number of Successful and Failure Mission Outcomes

Calculate the total number of successful and failure mission outcomes.

SQL Query

```
%sql SELECT (SELECT COUNT("MISSION_OUTCOME") FROM SPACEXTBL WHERE "MISSION_OUTCOME" LIKE '%Success%') AS SUCCESS, \
(SELECT COUNT("MISSION_OUTCOME") FROM SPACEXTBL WHERE "MISSION_OUTCOME" LIKE '%Failure%') AS FAILURE
```

Results

SUCCESS	FAILURE
100	1

*The **COUNT** keyword is used to calculate the total number of mission outcomes, and the **GROUPBY** keyword is also used to group these results by the type of mission outcome.*

Boosters Carried Maximum Payload

List the names of the boosters that have carried the maximum payload mass.

SQL Query

```
%sql SELECT DISTINCT "BOOSTER_VERSION" FROM SPACEXTBL \
WHERE "PAYLOAD_MASS_KG_" = (SELECT max("PAYLOAD_MASS_KG_") FROM SPACEXTBL)
```

Results

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

*We used a subquery to filter data by returning only the heaviest payload mass with **MAX** function. The main query uses **subquery** results and returns unique booster version (**SELECT DISTINCT**) with the heaviest payload mass.*

2015 Launch Records

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

SQL Query

```
%sql SELECT substr("DATE", 4, 2) AS MONTH, "BOOSTER_VERSION", "LAUNCH_SITE" FROM SPACEXTBL\
WHERE "LANDING_OUTCOME" = 'Failure (drone ship)' and substr("DATE",7,4) = '2015'
```

Results

MONTH	Booster_Version	Launch_Site
01	F9 v1.1 B1012	CCAFS LC-40
04	F9 v1.1 B1015	CCAFS LC-40

*This query returns month, booster version, launch site where landing was unsuccessful and landing date took place in 2015. **Substr** function process date in order to take month or year. Substr(DATE, 4, 2) shows month. Substr(DATE, 7, 4) shows year.*

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 Query

```
%sql SELECT "LANDING_OUTCOME", COUNT("LANDING_OUTCOME") FROM SPACEXTBL\
WHERE "DATE" >= '04-06-2010' and "DATE" <= '20-03-2017' and "LANDING_OUTCOME" LIKE '%Success%\
GROUP BY "LANDING_OUTCOME" \
ORDER BY COUNT("LANDING_OUTCOME") DESC ;
```

Results

Landing_Outcome	COUNT("LANDING_OUTCOME")
Success	20
Success (drone ship)	8
Success (ground pad)	6

*This query returns landing outcomes and their count where mission was successful and date is between 04/06/2010 and 20/03/2017. The GROUP BY clause groups results by landing outcome and **ORDER BY COUNT DESC** shows results in decreasing order.*

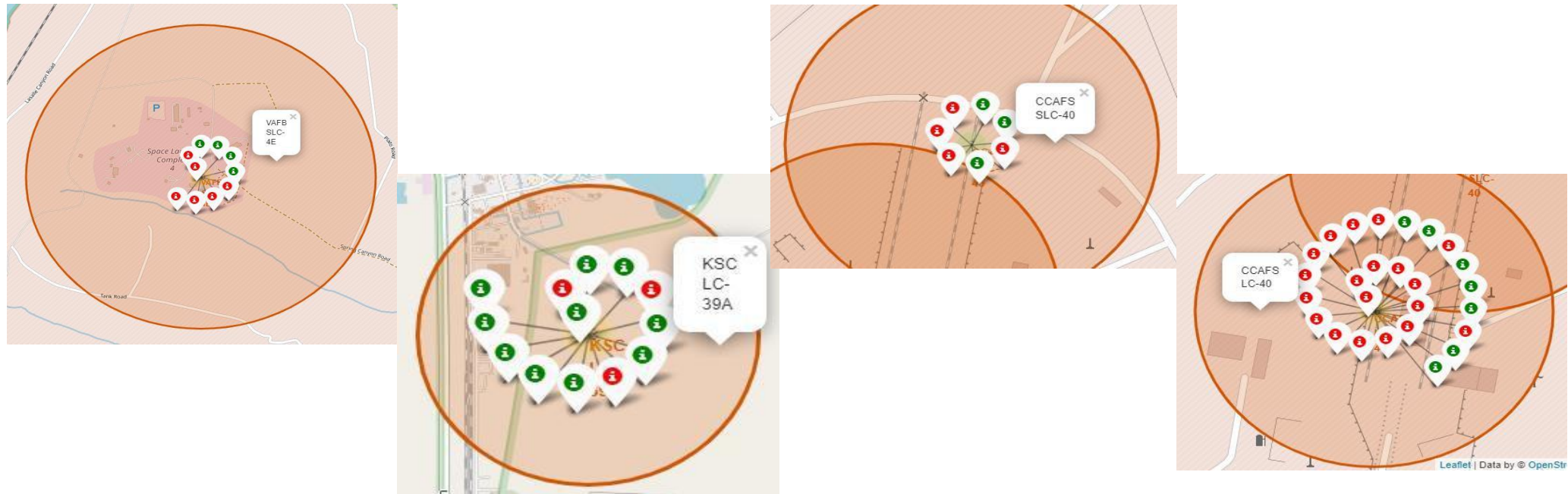
LAUNCH SITES PROXIMITY ANALYSIS – FOLIUM INTERACTIVE MAP

Folium map – Ground stations



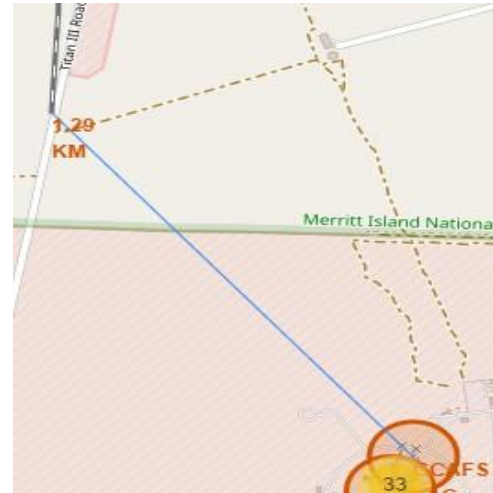
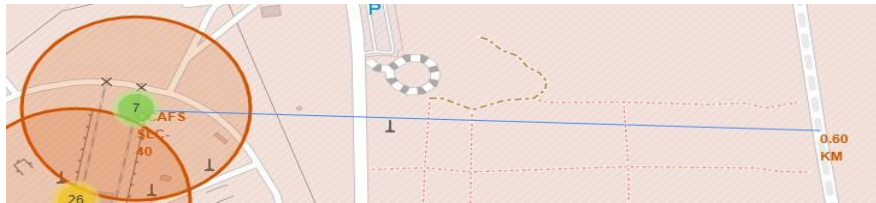
All SpaceX launch sites are on coasts of the United States of America, specifically Florida and California.

Folium map – Color Labeled Markers



Green marker represents successful launches. Red marker represents unsuccessful launches. We note that KSC LC39A has a higher launch success rate.

Folium Map – Distances between CCAFS SLC-40 and its proximities



Are launch sites in close proximity to railways?

- **YES.** The coastline is only 0.87 km due East. Are launch sites in close proximity to highways?
- **YES.** The nearest highway is only 0.59 km away. Are launch sites in close proximity to railways?
- **YES.** The nearest railway is only 1.29 km away. Do launch sites keep certain distance away from cities?
- **YES.** The nearest city is 51.74 km away.

Using the CCAFS SLC-40 launch site as an example site, we can understand more about the placement of launch sites.

INTERACTIVE DASHBOARD– PLOTLY DASH

Dashboard – Total success by Site

Total Success Launches by Site



The launch site KSC LC-39A had the most successful launches, accounting for 41.7% of the total successful launches.

Dashboard – PIE CHART FOR THE LAUNCH SITE WITH HIGHEST LAUNCH SUCCESS RATIO

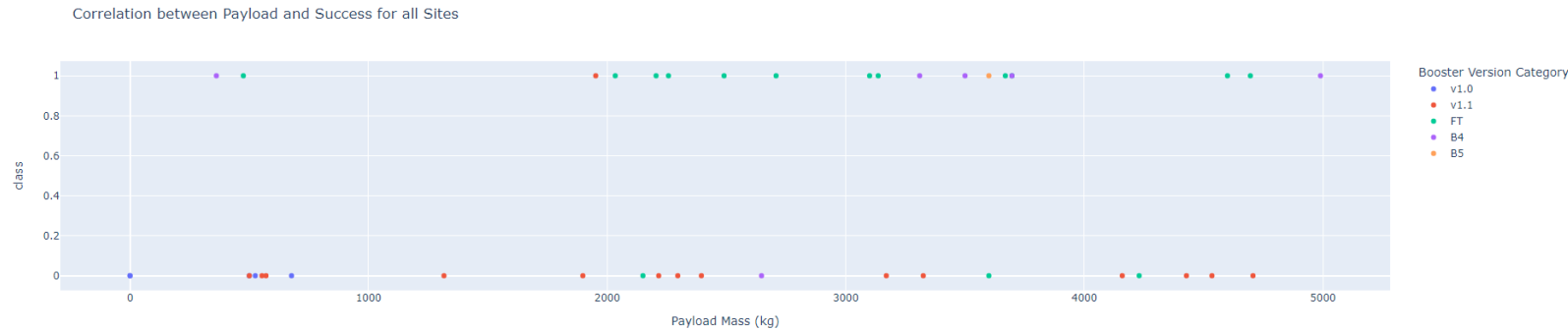
Total Success Launches for Site KSC LC-39A



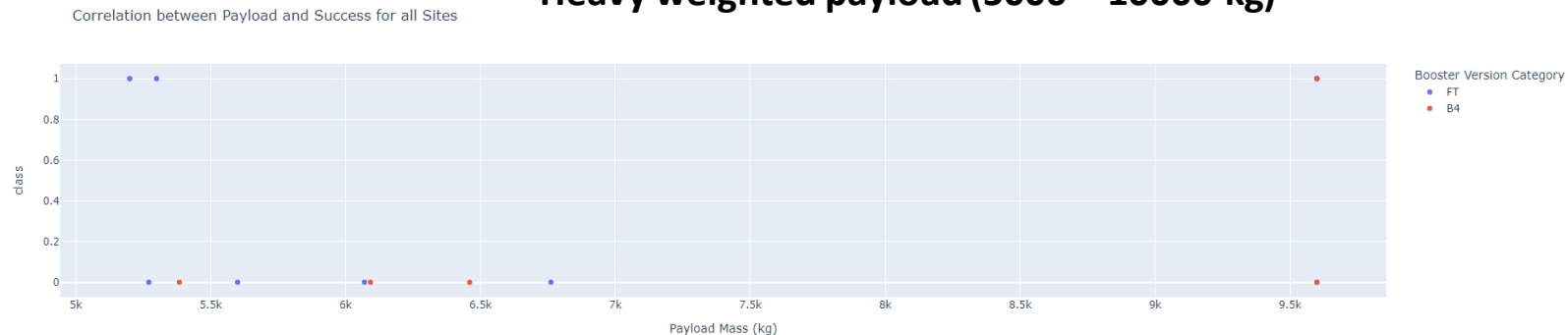
The launch site KSC LC-39A also had the highest rate of successful launches, with a 76.9% success rate..

Dashboard – PIE CHART FOR THE LAUNCH SITE WITH HIGHEST LAUNCH SUCCESS RATIO

Low weighted payload (0 – 5000 kg)



Heavy weighted payload (5000 – 10000 kg)



- Plotting launch outcomes against payloads for all sites reveals a gap near 4000 kg, suggesting the data be divided into two ranges:

- 0 – 4000 kg (low payloads)
- 4000 – 10000 kg (massive payloads)

- These two plots indicate that success rates for massive payloads are lower than those for low payloads.

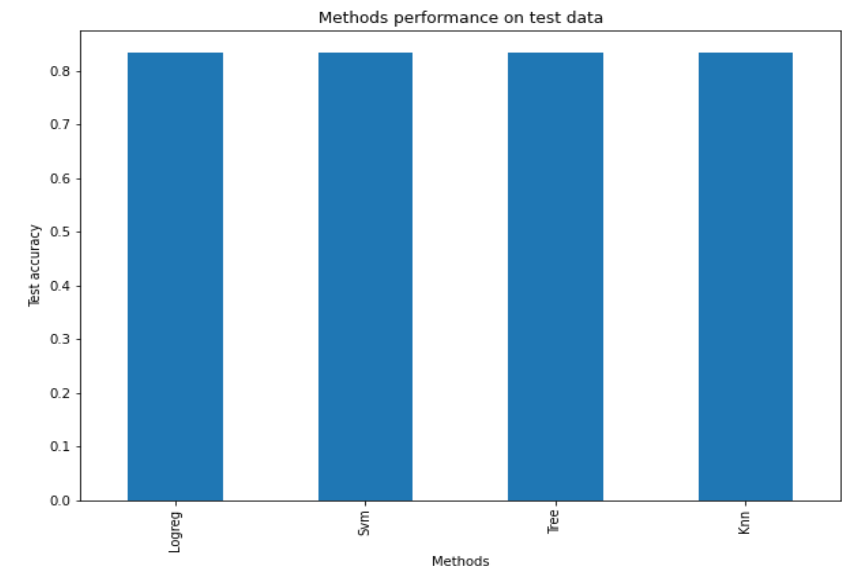
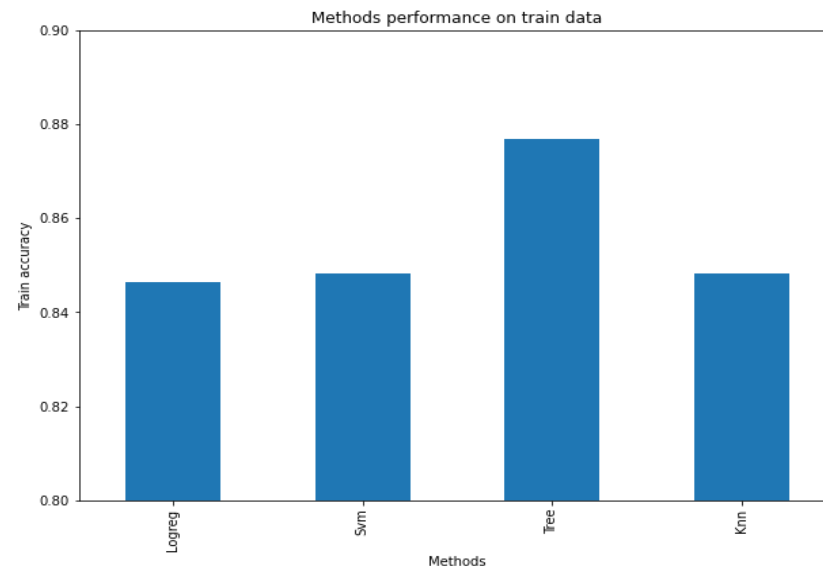
- Additionally, some booster types (v1.0 and B5) have never been launched with massive payloads.

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

IPREDICTIVE ANALYSIS – CLASSIFICATION

Classification Accuracy

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



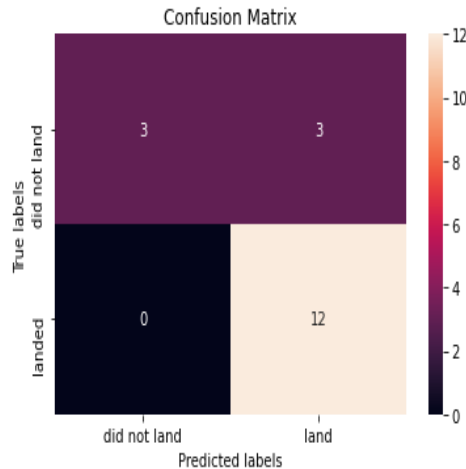
For accuracy test, all methods performed similar. We could get more test data to decide between them. But if we really need to choose one right now, we would take the decision tree.

Decision tree best parameters

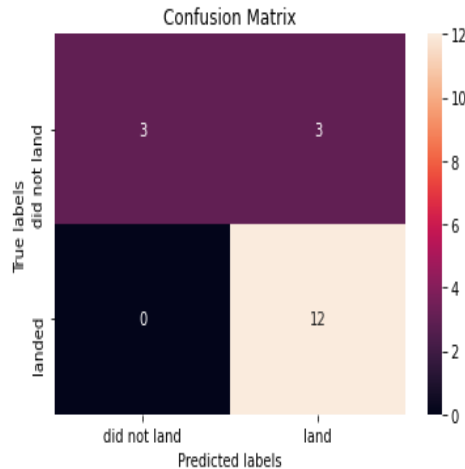
```
tuned hyperparameters :(best parameters) {'criterion': 'entropy', 'max_depth': 12, 'max_features': 'sqrt', 'min_samples_leaf': 4, 'min_samples_split': 2, 'splitter': 'random'}
```

Confusion Matrix

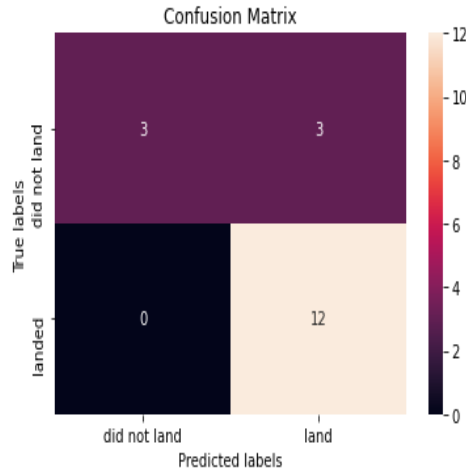
Logistic regression



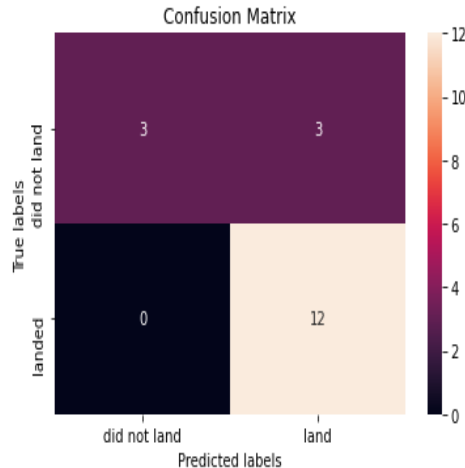
Decision Tree



kNN



SVM



- As previously shown, the Decision Tree model is the best performing classification model, achieving an accuracy of 94.44%.
- This is supported by the confusion matrix, which indicates only 1 out of 18 total predictions was incorrect (a false positive, located in the top-right corner).
- The remaining 17 predictions were correct (5 did not land, 12 did land).

		Actual values	
		1	0
Predicted values	1	TP	FP
	0	FN	TN

CONCLUSION

Conclusions

The success rate at a launch site increases with the number of flights, as most early flights tend to fail. This means that experience improves success rates.

- Between 2010 and 2013, all landings were unsuccessful, resulting in a 0% success rate.
- After 2013, the success rate generally improved, despite slight decreases in 2018 and 2020.
- Since 2016, the probability of success has consistently exceeded 50%.

Orbit types ES-L1, GEO, HEO, and SSO have the highest success rates, each at 100%.

- The 100% success rates for GEO, HEO, and ES-L1 orbits are due to only one flight each.
- The 100% success rate for SSO is notable, with five successful flights. • Orbit types PO, ISS, and LEO show greater success with heavy payloads.
- VLEO (Very Low Earth Orbit) launches are associated with heavier payloads, which is logical.

The launch site KSC LC-39 A had the most successful launches, making up 41.7% of all successful launches, and also had the highest success rate at 76.9%.

Success rates for very heavy payloads (over 4000kg) are lower than for lighter payloads. The Decision Tree model is the best performing classification model, with an accuracy of 94.4%

THANK YOU