IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Hayelom Tesfay
April 2024

Outline

01 Executive Summary

02 Introduction

03 Methodology

04 Results

05 Conclusion

06 Appendix

Executive Summary

Introduction: Nowadays, commercial rocket launches is becoming very competitive market. **T**he rocket launches in SpaceX are relatively cheap ($62 million per launch) due to its novel reuse of the first stage of its Falcon 9. If we can make predictions on whether the first stage will land, we can determine the cost of a rocket launch. Thus, to predict the success of first stage of the Falcon9 is crucial.

Objective: the goal of this Data Science capstone project is to predict if the Falcon 9 rocket first stage will land successfully.

Methodologies:

In order to achieve the objective of the this capstone project, stated above, the following methodologies where applied:

- **SpaceX REST API** and **web scraping** methods were used to **collect launch site data**.
- **Data wrangling was employed to make the datasets suitable for building machine learning models**
- **Data visualization methods were used to Explore** payload, launch site, flight number, etc.
- **SQL queries** and **data visualizations** were created to discover insights about the data set and answer questions.
- **Interactive map with Folium** was built to *explore launch site* success rates and show the launch site proximities (railway, highway, coastline, city) to geographical markers.
- **Dashboard with Plotly Dash** was built to Visualize the launch sites with the most success and successful payload ranges
- **Machine learning method** was applied to build predictive models using Logistic Regression, SVM, Decision Tree, and KNN

Results:

- The exploratory data analysis revealed that the success rate of the Falcon 9 rocket landings is 66.67%.
- The landing launch success rate increased since 2013, in spite of small dips in 2018 and 2020.
- Some Orbits like ES-L1, GEO, HEO, and SSO have the most success rate (100%) .
- KSC LC-39A has the most successful launches, with 41.7% of the total launches and with 76.9% success rate.
- Decision Tree is the best machine learning model to use for this project, as it has the highest accuracy with 94.44%.

Conclusion: from the results we can conclude that the Falcon 9 rocket first stage can be landed successfully.

Recommendations:

- Large datasets should be used to get better classification performance, more generalizable, and reliable conclusions.
- to land Falcon 9 rocket more successful, the influence of payload mass, flight number, orbit, etc should be consided.

Introduction

Project background and context

SpaceX, is an American spacecraft manufacturer and launch service provider founded in 2002 by Elon Musk with the goal of reducing space transportation costs. The company currently operates the **Falcon 9** and Falcon Heavy rockets along with the Dragon and Starship spacecraft**.**

Nowadays, commercial rocket launches is becoming very competitive market. **Falcon 9** is a **reusable**, two-stage **rocket** designed and manufactured by SpaceX for the reliable and safe transport of people and payloads into Earth orbit and beyond. Since 2017, SpaceX Falcon9 is increasingly dominating the market with better cost per kilogram.

Due to its novel **reuse of the first stage of its Falcon 9** rocket**,** the rocket launches in **SpaceX are relatively cheap** ($62 million per launch). Other providers, which are not able to reuse the first stage, cost up to $165 million each. If we can make predictions on whether the first stage will land, we can determine the cost of a launch, and use this information to assess whether or not an alternate **company should bid against SpaceX for a rocket launch**. The goal of this Data Science capstone project is to allow the company to compete with SpaceX. In order to achieve this goal, it is necessary to predict if the first stage of the Falcon 9 rocket will land successfully. Thus, **a study is proposed** with the following **Business objective:**

- to predict if the Falcon 9 rocket first stage will land successfully.

Problems we want to find answers

- Is it possible to predict whether the Falcon 9 rocket first stage will land successfully or not?
- How do payload mass, launch site, number of flights, and orbits types affect the first stage landing success of Falcon 9?
- What is the best Machine learning model to predict if the Falcon 9 first stage will land successfully?

Section 1

Methodology

Methodology

Executive Summary

Objective of this capstone project is to predict if the Falcon 9 rocket first stage will land successfully. In order to achieve this objective the following methodologies where employed:

Data Collection was employed using GET requests to the SpaceX REST API as well as using Web Scraping from Wikipedia.

Data Wrangling was used to replace missing values, one-hot encoding was applied to categorical features, to determine number of launches on each site and number occurrence on each orbit, to create a successfully/failure labels for booster landing outcome used for building models.

Exploratory Data Analysis was employed to visualize relationships between variables and to determine unseen patterns using Matplotlib , seaborn , and Pandas and libraries, also SQL queries was used to manipulate and assess the launch dataset

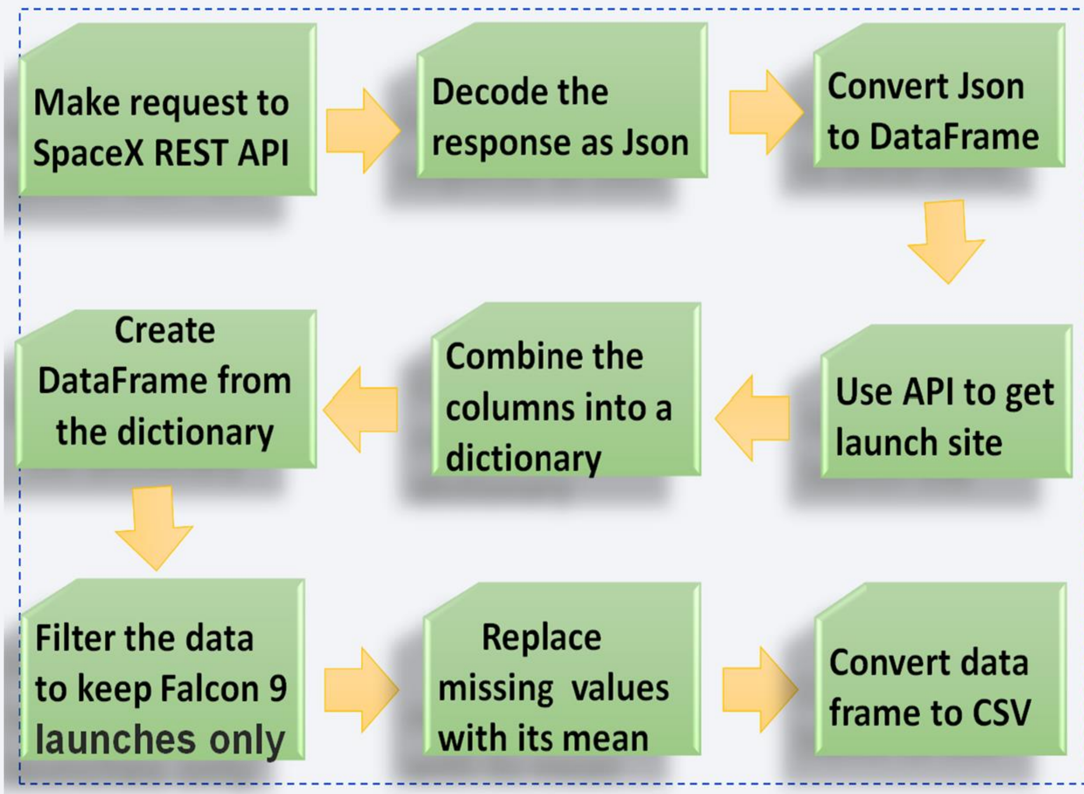
Interactive visual analytics was performed using **Folium** and **Plotly Dash**. Python's interactive visualization packages were used to answer some analytical questions. **Folium** was used for creating maps while **Plotly Dash** was used to create interactive data visualizations.

Predictive analysis using classification models was perform using Logistic Regression, SVM, Decision Tree, and KNN. Building the **Mode**l was perform using data pre-process, split data into training and test dataset, create a GridSearchCV object, Fit the object to the parameters, and train the models. **Model Evaluation** was made using the output **GridSearchCV** object, tuned hyperparameters**,** checking **the accuracy,** and by Ploting and examining the **confusion matrix. Finding the best** performing Model was done by **comparing the accuracy** scores of all models. **The Decision Tree** has the highest **accuracy,** with **94.44%,** and hence the best model.

Data Collection – SpaceX API

We follow the following steps:

- Request data from SpaceX REST API, launch data
- Convert the response to a .json file using .json() then to a Pandas DataFrame using .json_normalize()
- Request information about the launches from SpaceX REST API using custom functions
- Create dictionary from the data
- Create a Pandas DataFrame from the constructed dictionary dataset
- Filter dataframe to contain only Falcon 9 launches
- Replace missing values of Payload Mass with calculated .mean()
- Export data to csv file



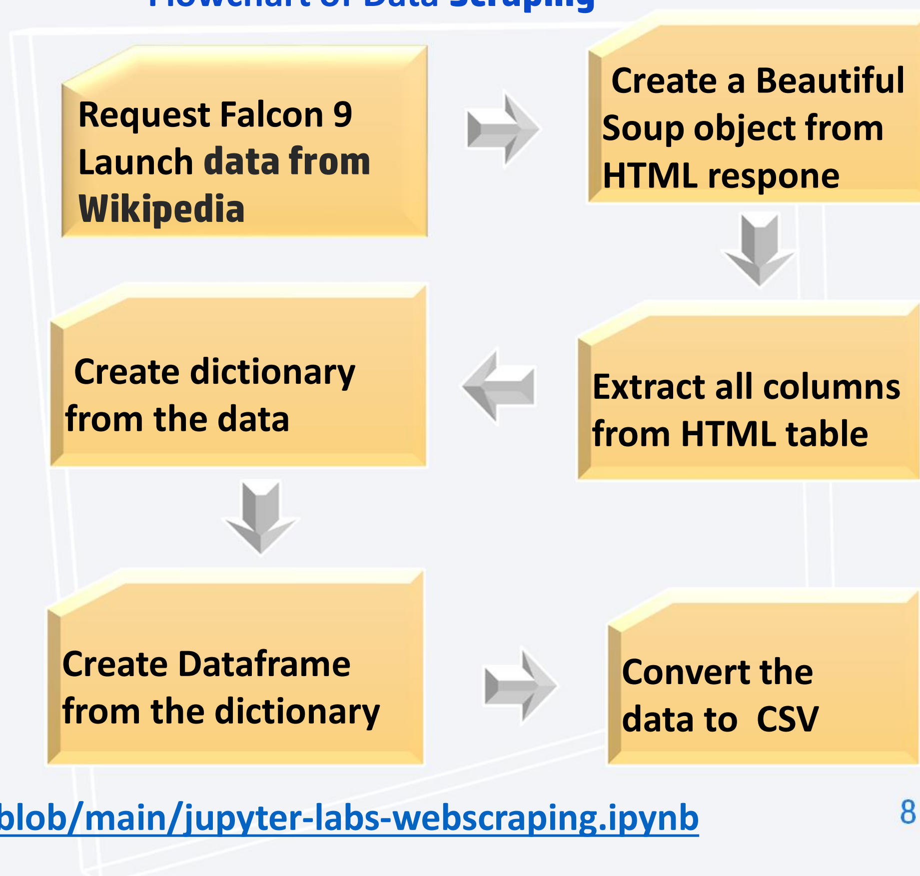
GitHub URL https://github.com/Hayethdt/thanks-cse/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping

The steps we follow in Data Collection –Scraping

- Request Falcon 9 launch data from Wikipedia. Assign the response to an object
- Create BeautifulSoup object from HTML response
- Find all tables within the HTML page
- Extract column names from HTML table header
- Collect all column header names from the tables found within the HTML page
- Create dictionary from the data
- Create dataframe from the dictionary
- Convert the dictionary to a Pandas DataFrame ready for export. Finally, export data to csv file

Flowchart of Data Scraping



GitHub URL https://github.com/Hayethdt/thanks-cse/blob/main/jupyter-labs-webscraping.ipynb

Data Wrangling

Here, we perform **exploratory Data Analysis** and determine **Training Labels**

Perform some Exploratory Data Analysis **(EDA)** to:

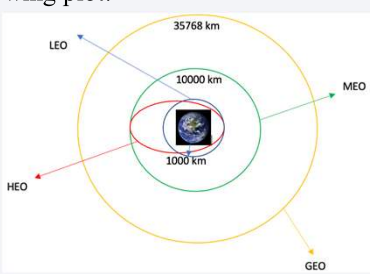
- **Find** some **patterns** in the data
- **determine** the **label** for training **models.**

We **convert** the **outcomes** into Training **Labels**:

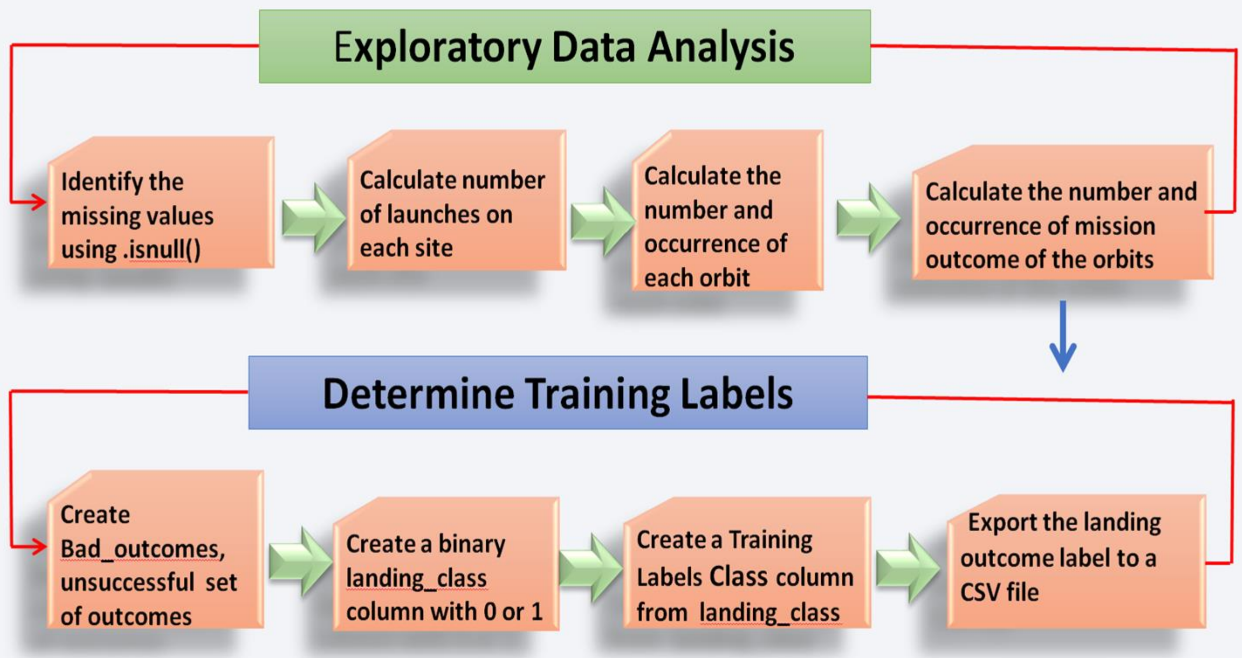
- **1** for **booster landing success**
- **0** for **booster landing failure**

The data wrangling process is presented using a flowchart.

Each launch aims to a dedicated orbit. some are shown in the following plot:



Flowchart of Data Wrangling



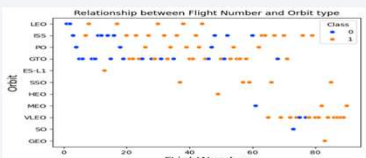
URL link https://github.com/Hayethdt/thanks-cse/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

SCATTER CHARTS

Scatter charts were plotted to visualize the relationships between:

- Flight Number and Launch Site
- Payload and Launch Site
- Flight number and Orbit type
- payload and orbit type

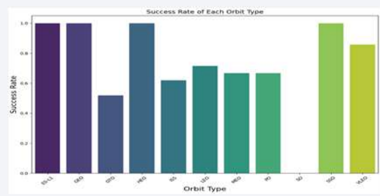


We use these scatter charts to:

- uncover hidden relationships
- identify patterns
- detect outliers
- observe unexpected gaps

BAR CHART

A bar chart was drawn to visualize the relationship between **Success Rate** and **Orbit Type**

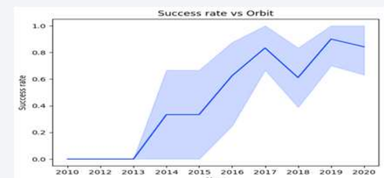


A bar chart plots numeric values for levels of a categorical feature as bars. A bar chart is used when you want to:

- show a distribution of data points,
- perform a comparison of metric values across different subgroups of your data.
- see which groups are highest or most common,
- how other groups compare against the others.

LINE CHARTS

Line charts were plotted to visualize the relationships between **Success Rate** and **Year** (launch success yearly trend)



A line chart uses points connected by line segments *to demonstrate changes in value.* Line graphs are best used to:

- show trends over time
- see how a variable changed over time
- identify any trends or patterns that may be present.

GitHub URL https://github.com/Hayethdt/thanks-cse/blob/main/edadataviz.ipynb

EDA with SQL

The **SQL queries** we performed to gather some important information about the dataset used in this Data Science capstone project are summarized as follows:

- 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)
- 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 (u*se the min function)*
- 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. Use a subquery
- 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 landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

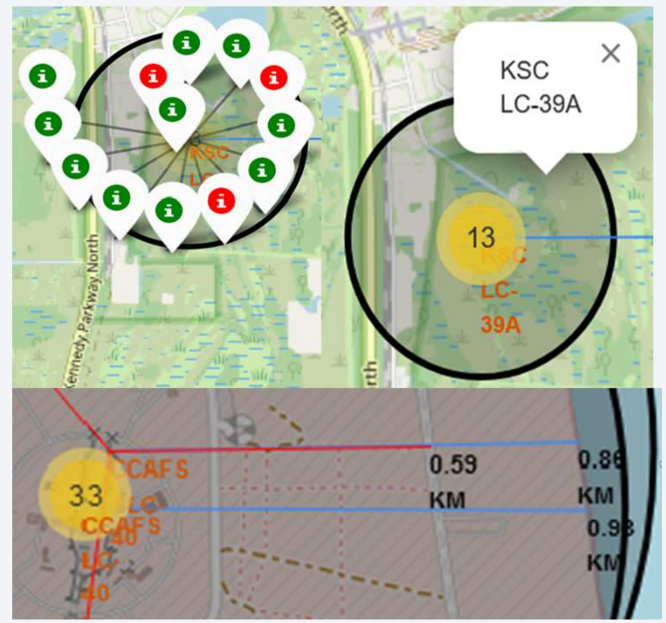
GitHub URL https://github.com/Hayethdt/thanks-cse/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

A **map marker** (pin, location pin, place marker) identifies a precise location on a map
Types of markers include: shapes (*polygons, **circles** and rectangles*), ***information windows*** (***popups***, tables and graphs), **lines** (called **polylines** *as they represent sequences of locations*) and **images.**.

In general, this section of lab project contains the following **major tasks**:

- to Mark all launch sites on the map
- to Mark the success/failed launches for each site on the map
- to Calculate the distances between a launch site to its proximities



Markers Indicating Launch Sites

Map objects created and added: **folium.Circle, folium.Marke**r, and **MarkerPopup**

- Added blue circle at **NASA Johnson Space Center's coordinate** with a popup **label** showing **its name** using its latitude and longitude coordinates.
- Added red circles at **all launch sites coordinates** with a popup **label** showing **its name** using its latitude and longitude coordinates.

Colored-Labeled Markers of Launch Outcomes

Map objects created and added: **folium.Marker** and **MarkerCluster() object**

- We added Marker Cluster because many launch records have **same coordinate** due to only 4 launch sites.
- **Colored markers** of **successful** (green) and **failed** (red) **launches** are **added** at each launch site **to show** which launch sites have high **success rates.**

PolyLine to Show Distances between CCAFS SLC-40 to its Proximities

Map objects created and added: **folium.PolyLine** and **folium.Marker**

- **colored** lines to **show distance** between launch site **CCAFS SLC-40** and **its proximity** are **drawn** to the closest ***coastline, railway, highway,*** *and* ***city.***

GitHub URL: https://github.com/Hayethdt/thanks-cse/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

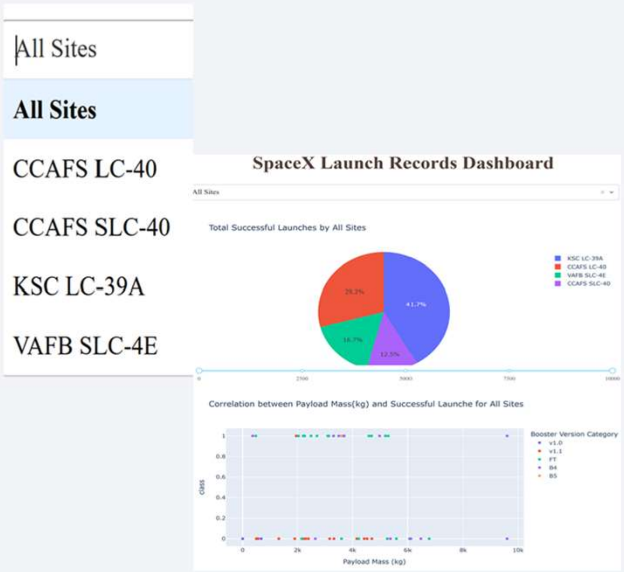
We built an **interactive** dashboard application with Plotly includes:

1. Dropdown Menu for selecting launch sites

- Allow user to select all launch sites or a specific launch site
- Show the **successful and unsuccessful** launches **as a percent of the total**

2. Range Slider for selecting range of payload mass (kg).

- Allow user to select payload mass range
- show the success/failed landing for payload mass ranges and booster versions



3. Pie chart

- display **by default** the **distribution of landing success for all launch sites**
- show the percentage of successful and failed landing for a **specific selected** site from the **dropdown**
- it makes clear to see **which sites are most successful**

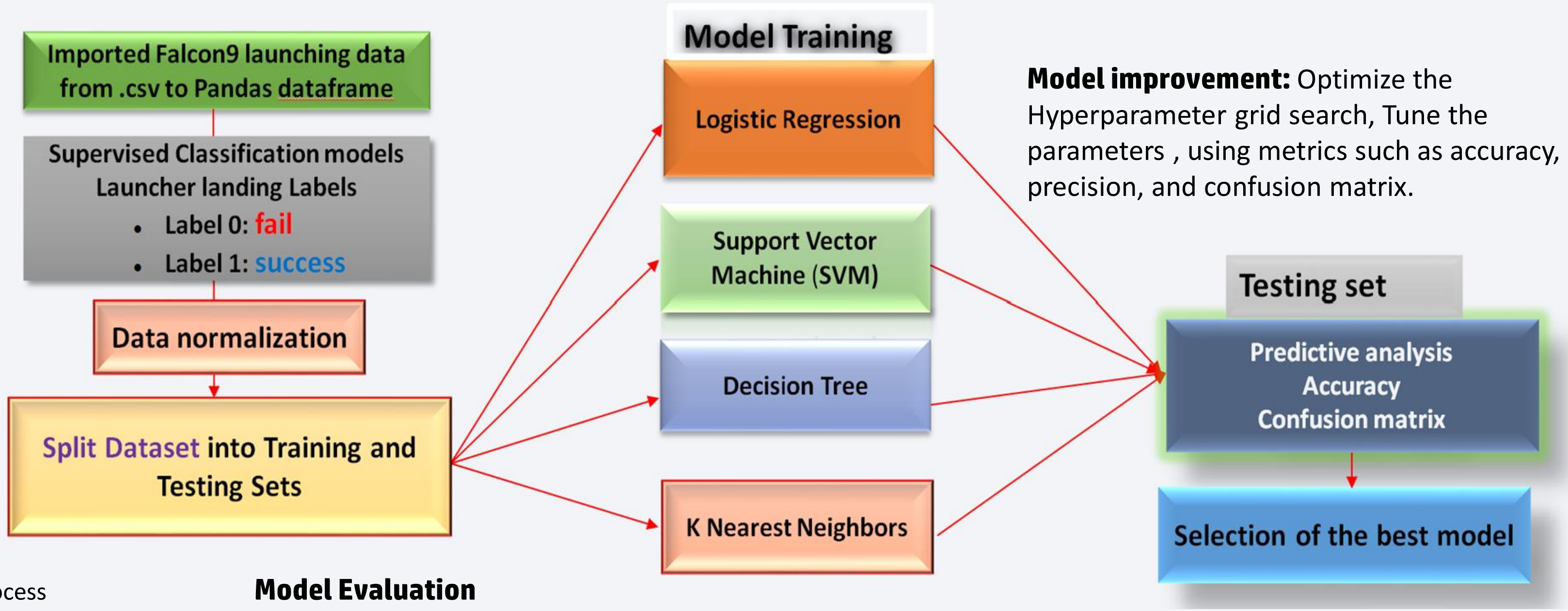
4. Scatter chart/graph

- Displays launch site, payload mass, booster version, success/failure
- Allow user to see the relationship between Payload and Launch Success
- Shows the successful and failed landing across different payload mass for different booster versions for all or any specific launch sites.

GitHub URL https://github.com/Hayethdt/thanks-cse/blob/main/Spacex_Dashboard_project.ipynb

Predictive Analysis (Classification)

Flowchart of Data Predictive Analysis



Model improvement: Optimize the Hyperparameter grid search, Tune the parameters , using metrics such as accuracy, precision, and confusion matrix.

Building the Model

- Perform data pre-process
- **Split data i**nto training and test dataset
- Select ML algorithms
- Create **a GridSearchCV object**
- Fit the object to the parameters
- **Train the ML models**

Model Evaluation

Using the output **GridSearchCV** object:

- Check the tuned hyperparameters (**best_params_)**
- Check **the accuracy** (score and best_score_)
- **Plot and examine** the Confusion Matrix

Finding the best performing Model

- **Compare the accuracy** scores for **all models**
- The model with the highest accuracy score is the best performing model

The Decision Tree has the highest **accuracy,** with **94.44%.**

GitHub URL https://github.com/Hayethdt/thanks-cse/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

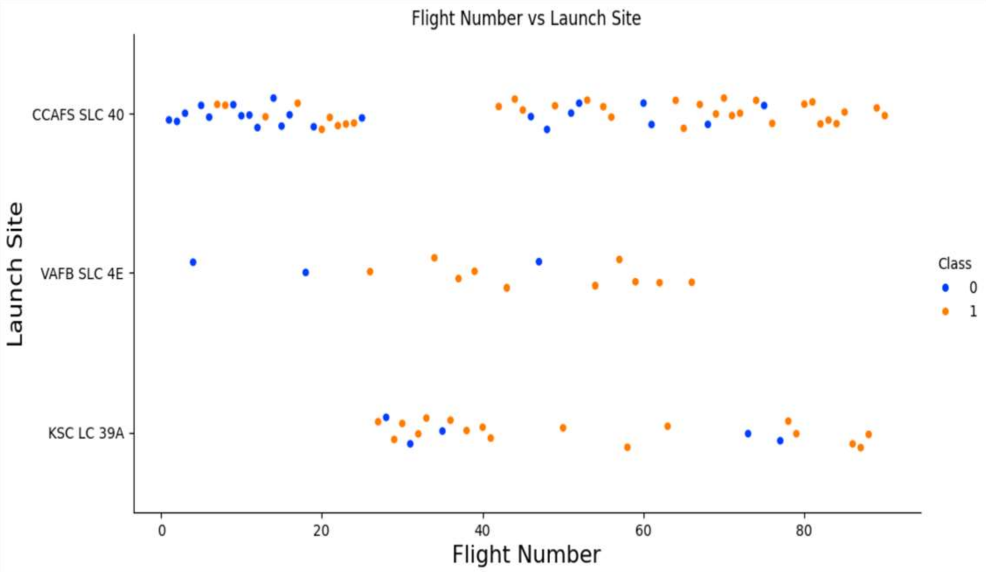
Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

From this **scatter plot**, we can conclude that:

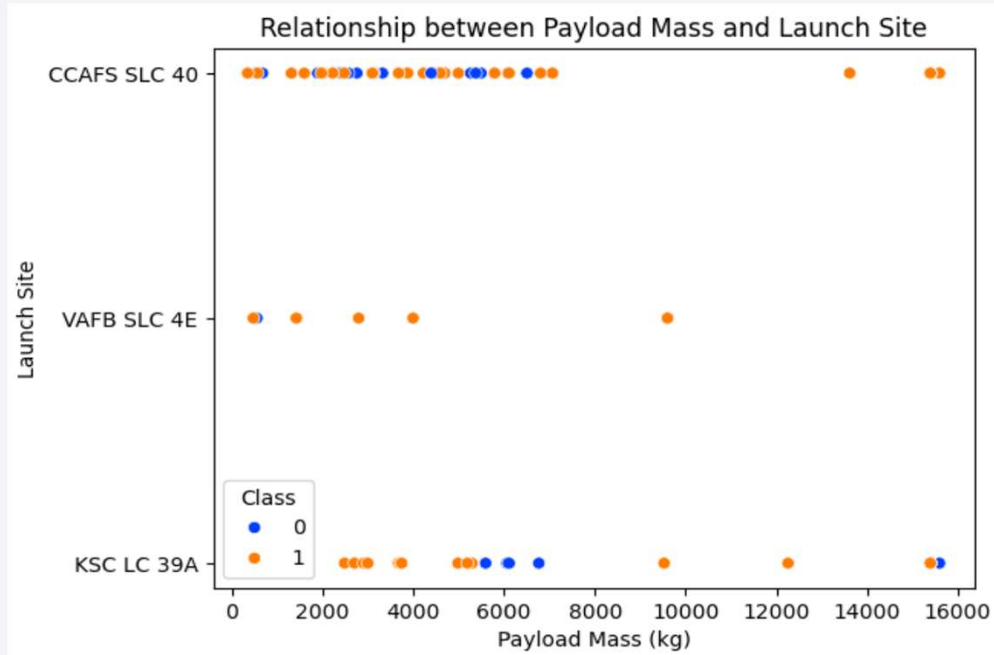
- the success rate increased as the number of flight increased.
- the site VAFB SLC 4E has the least number of attempts.
- most early flights were launched from CCAFS SLC 40, while no early flights were launched from KSC LC 39A.
- site CCAFS SLC 40 concentrates most failures, principally in the early stage.
- earlier flights were less successful, while all sites were 100% successful in the end flights.
- the launches from site KSC LC 39A were more successful, while the launches from site CCAFS SLC 40 was generally least successful.



Payload vs. Launch Site

The following conclusions can be made from the plot:

- most of the launches are from CCAFS SLC 40 site, though typically lighter payloads.
- CCAFS SLC 40 and VAFB-SLC 4E have some outliers in the heavy payloads mass.
- There seems to be no clear relationship between payload mass and Launch Site for a given site.
- Site VAFB SLC 4E has no launch for heavy payload mass greater than 12000 kg. Moreover, it has the least launches compare to other.
- KSC LC 39A was the most successful launch site.



Success Rate vs. Orbit Type

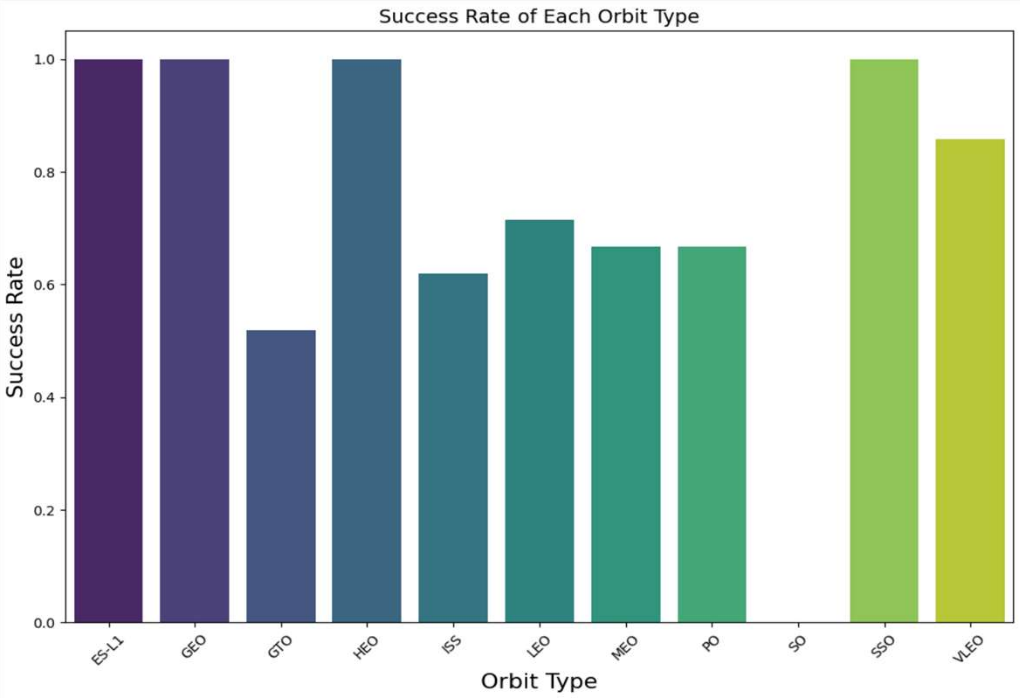
From the bar chart, we can conclude the following points:

Orbits that have a 100% success rate are:

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

Orbit that has 0% success rate is SO

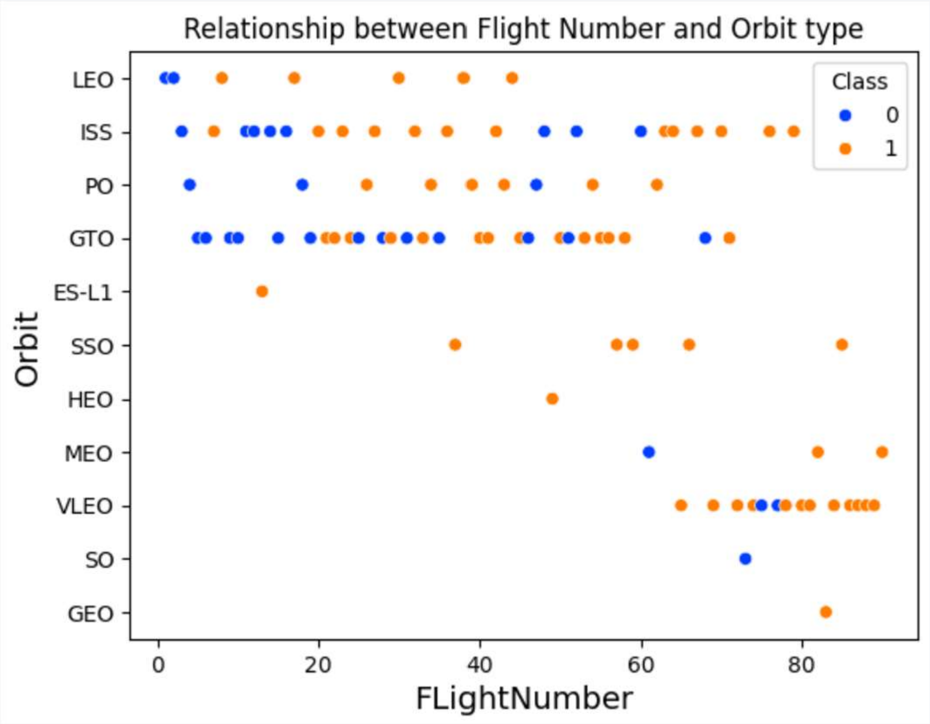
(Heliocentric Orbit).



Flight Number vs. Orbit Type

The scatter plot shows the Flight Number vs. Orbit type. From this plot we deduce the following points:

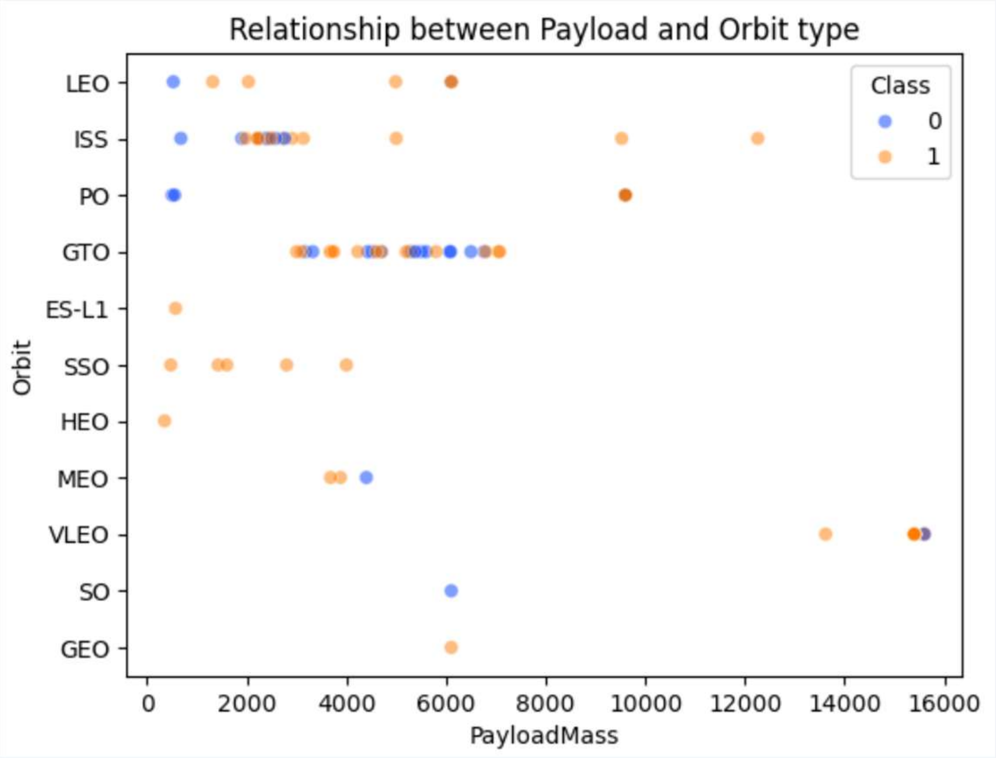
- **SSO** orbit has a 100% success rate with only five flights.
- The orbits **ES-L1, HEO, and GEO** have a 100% success rate by having one flight each. In contrast, **SO orbit** has **0% success** rate with only one unsuccessful flight.
- The number of flights for **ES-L1, HEO, GEO,** and **SO** is not weighty for concluding about their success rate.
- Unsuccessful landing occurred for **LEO** only for the early launches. For **LEO orbit**, success is related to the number of flights; whereas, in the **GTO orbit**, there is no clear relationship between flight number and the orbit.
- Generally, as Flight Number increases, the success rate increases.
- Flights numbers greater than 30 have a higher success rate than flight numbers less than 30.



Payload vs. Orbit Type

This scatter plot shows that:

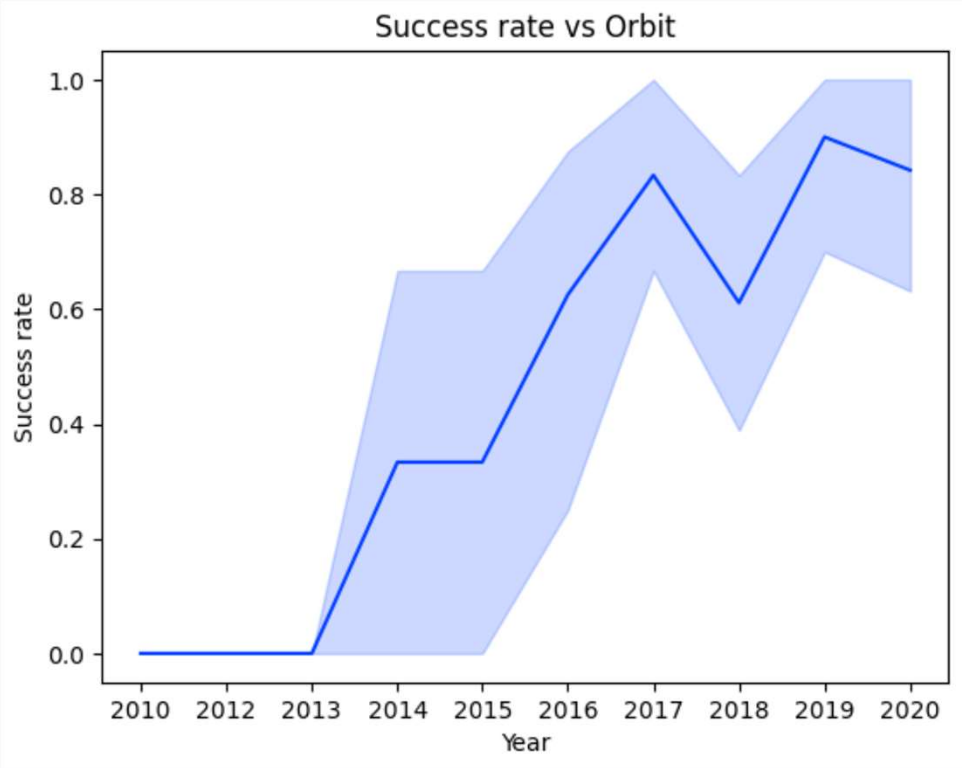
- There is no direct correlation between Orbit type and payload mass for GTO orbit as both success and failure launches equally present.
- VLEO launches are associated with heavier payloads
- With heavy payloads, the successful landing are more for PO (with 2 data points), LEO and ISS orbits.
- For the payloads lighter, the success rate increases in the SSO and HEO orbits.
- Between 2000 and 7500 kg, success rate seems to be evenly distributed for GTO.



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 value is 0.
- The **success rate** was **sharply increased** from 2013 to 2014, from 2015 to 2017, and from 2018 to 2019, though, there is little deflection started at 2016 .
- Unlike the previous year, the **success rate** was **slightly increased** from 2014 to 2015..
- The success rate was **sharply decreased** from 2017 to 2018 and slightly decreased from 2019 to 2020.
- **In general**, starting at 2013, the trend of the chart shows an increase in landing success rate, except the dip in 2018 and in 2020.



RESULTS

EDA with SQL

All Launch Site Names

SQL query

```
%%sql

SELECT DISTINCT Launch_Site
FROM SPACEXTBL;
```

Query result

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

- The key word **DISTINCT** is applied to return only **unique launch sites** name from the Launch_Site **column** of the SPACEXTBL table.
- The SQL query and its result are given above.
- The unique **launch sites resulted** from the SQL query are four in number, namely: **CCAFS LC-40, VAFB SLC-4E, KSC LC-39A, and CCAFS SLC-40**

Launch Site Names Begin with 'CCA'

SQL query

```
%%sql

SELECT *
FROM SPACEXTBL
WHERE Launch_Site LIKE 'CCA%'
LIMIT 5;
```

- **WHERE** filters records that satisfy the condition
- **LIKE** is used with the wild card '**CCA%'** to retrieve string values beginning with 'CCA'.
- **LIMIT** 5 fetches only 5 records, and
- The **5 sites name** that begin with **CCA** In the **Launch_Site column** are: **CCAFS LC-40, CCAFS LC-40, CCAFS LC-40, CCAFS LC-40**, and **CCAFS LC-40.**
- The **5 records** where the launch sites **begin with `CCA`** are displayed below.

Query result

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
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	Failure (parachute)
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

SQL query

```
%%sql

select SUM(PAYLOAD_MASS__KG_)
from SPACEXTBL
where Customer = "NASA (CRS)
```

Query result

Sum_of_PAYLOAD_MASS
45596

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

- The SUM() function calculate d the total payload mass in the **PAYLOAD_MASS__KG** column.
- **The WHERE** condition **filters** the results to **boosters only from NASA (CRS)** .
- The total payload maass carried by boosters from NASA**(CRS)** was calculated as **45596.**

Average Payload Mass by F9 v1.1

SQL query

```
%%sql

select AVG(PAYLOAD_MASS__KG_) as Average_PAYLOAD_MASS
from SPACEXTABLE
where Booster_Version = "F9 v1.1";
```

Query result

Average_PAYLOAD_MASS
2928.4

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

- The AVG() **function** is used to calculate the average of the PAYLOAD_MASS__KG_ **column**, but
- the **WHERE** condition **filters** the results to include **only F9 v1.1 booster** version.
- Hence, the average payload mass carried by booster version F9 v1.1 was **2928.4.**

First Successful Ground Landing Date

SQL query

```
%%sql

SELECT min(Date)
FROM SPACEXTBL
WHERE Landing_Outcome = "Success (ground pad)"
ORDER BY Date DESC LIMIT 1;
```

Query result

min(Date)
2015-12-22

- The MIN( ) function is used to find the first or **minimum** date from the **DATE column.**
- The WHERE clause **filters the results** to only the **successful ground pad landings.**
- The dates of the **first successful** landing outcome on ground pad was **22nd December 2015**.

Successful Drone Ship Landing with Payload between 4000 and 6000

SQL query

```
%%sql

select Booster_Version as "Booster_Success (drone ship)", PAYLOAD_MASS__KG_
from SPACEXTBL
where Landing_Outcome = "Success (drone ship)"
and
PAYLOAD_MASS__KG_ between 4000 and 6000
order by PAYLOAD_MASS__KG_;
```

Query result

Booster_Success (drone ship)	PAYLOAD_MASS__KG_
F9 FT B1026	4600
F9 FT B1022	4696
F9 FT B1031.2	5200
F9 FT B1021.2	5300

- The **WHERE** clause is used to **filter boosters successfully landed on drone ship.**
- The **BETWEEN clause** is used to **retrieve payload mass greater** than **4000** but less than **6000.**
- **The AND** is used to **determine successful landing with payload mass** greater than 4000 but less than 6000
- It appears that there **only 4 Boosters** with a payload mass between 4000 and 6000.

Total Number of Successful and Failure Mission Outcomes

SQL query

```
%%sql

select (select COUNT("MISSION_OUTCOME") from SPACEXTBL
where "MISSION_OUTCOME" LIKE '%Success%') as "Number of successful mission",
(select COUNT("MISSION_OUTCOME") from SPACEXTBL
where "MISSION_OUTCOME" LIKE '%Failure%') as "Number of failure mission"
```

Query result

Number of successful mission	Number of failure mission
100	1

The **COUNT() function** is used to calculate the total **number of mission outcomes**. We used wildcard like **'%' to filter** for **WHERE** MISSION_OUTCOME was a **success or a failure.** There have been **100 successful mission** outcomes out of 101 missions.

Boosters Carried Maximum Payload

SQL query

```
%%sql

select Booster_Version, PAYLOAD_MASS__KG_
from SPACEXTBL
where PAYLOAD_MASS__KG_ =(select MAX(PAYLOAD_MASS__KG_)
                          from SPACEXTBL)
ORDER BY Booster_Version;
```

Query result

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1048.5	15600
F9 B5 B1049.4	15600
F9 B5 B1049.5	15600
F9 B5 B1049.7	15600
F9 B5 B1051.3	15600
F9 B5 B1051.4	15600
F9 B5 B1051.6	15600
F9 B5 B1056.4	15600
F9 B5 B1058.3	15600
F9 B5 B1060.2	15600
F9 B5 B1060.3	15600

The SELECT key word within the brackets finds the maximum payload. The **MAX() function** was **used in a subquery** to retrieve a list of boosters which have carried the maximum payload mass. This value is used in the WHERE condition to filter the Payload. It is found that 12 boosters have carried the maximum payload mass of 15600 kg.

2015 Launch Records

SQL query

```
%%sql

SELECT substr(Date, 6,2) as Month, Booster_Version, Launch_Site, Landing_Outcom
FROM SPACEXTBL
WHERE Landing_Outcome = 'Failure (drone ship)' AND substr(Date,0,5)='2015';
```

Query result

Month	Booster_Version	Launch_Site	Landing_Outcome
01	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

The **WHERE** keyword is used to filter the results for only failed landing outcomes, AND only for the year of 2015. It appears that **2 boosters failed** to land in drone ship in the year 2015.

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

SQL query

```
%%sql

select Landing_Outcome, COUNT(Landing_Outcome) as Count
from SPACEXTBL
where DATE BETWEEN '2010-06-04' and '2017-03-20'
group by Landing_Outcome
order by Count DESC;
```

COUNT() function was used to count the different *landing outcomes*. The WHERE clause is used with the BETWEEN keyword to filter the results to dates only between 2010-06-04 and 2017-03-20. Then, the results are grouped using GROUP BY to ensure that the counts were grouped by their outcome, and ordered using ORDER BY. Moreover, DESC is used to sort the results by descending order.

Query result

Landing_Outcome	Count
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

Section 3

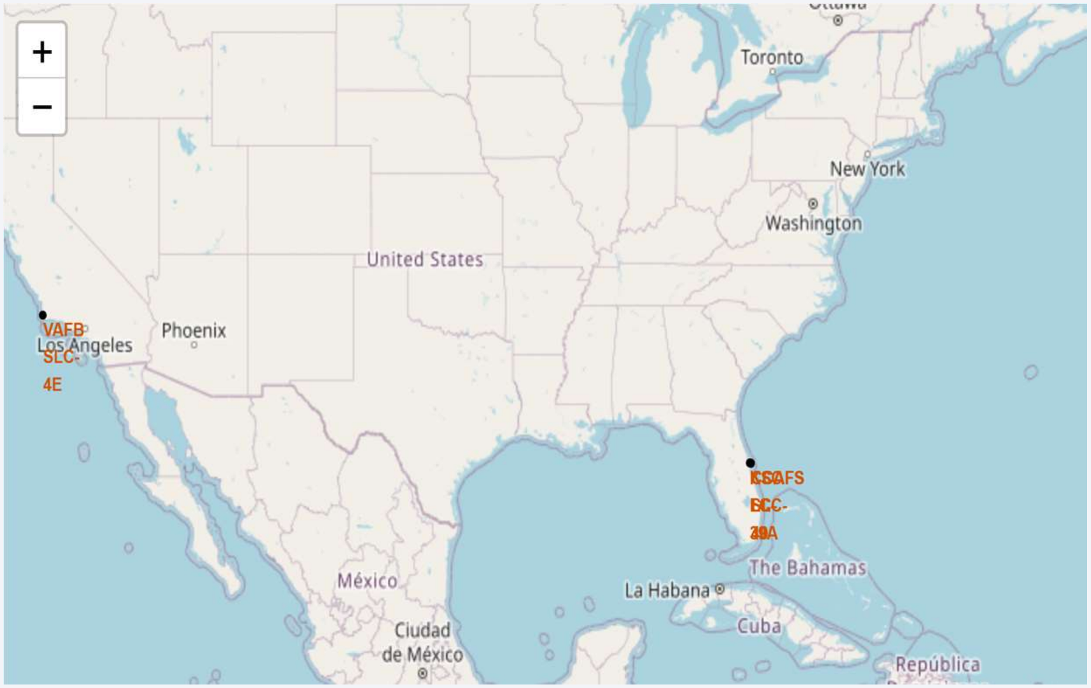
Launch Sites Proximities Analysis

All launch Sites Global Map Markers

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

The **yellow** markers are indicators of where the **locations of all the SpaceX launch sites** are situated in the US. The launch sites have been strategically placed near the coast

We can see that all launch sites are in very close proximity to the coast and they are also a couple thousand kilometers away from the equator line.



The four launch sites and their locations can be given as:

- **VAFB SLC-4E:** Vandenberg Space Launch Complex 4, **California**
- **KSC-LC29A**: Kennedy Space Center - **Merritt Island, Florida**
- **CCAFS-LC40**: Cape Canaveral Launch Complex **40, Florida**
- **CCAF-SLC40:** Cape Canaveral Space Launch Complex **40, Florida**

Folium Map Color-Labeled Launch Outcome

We create a **column** called **marker_color** to store the marker colors based on the **class value**

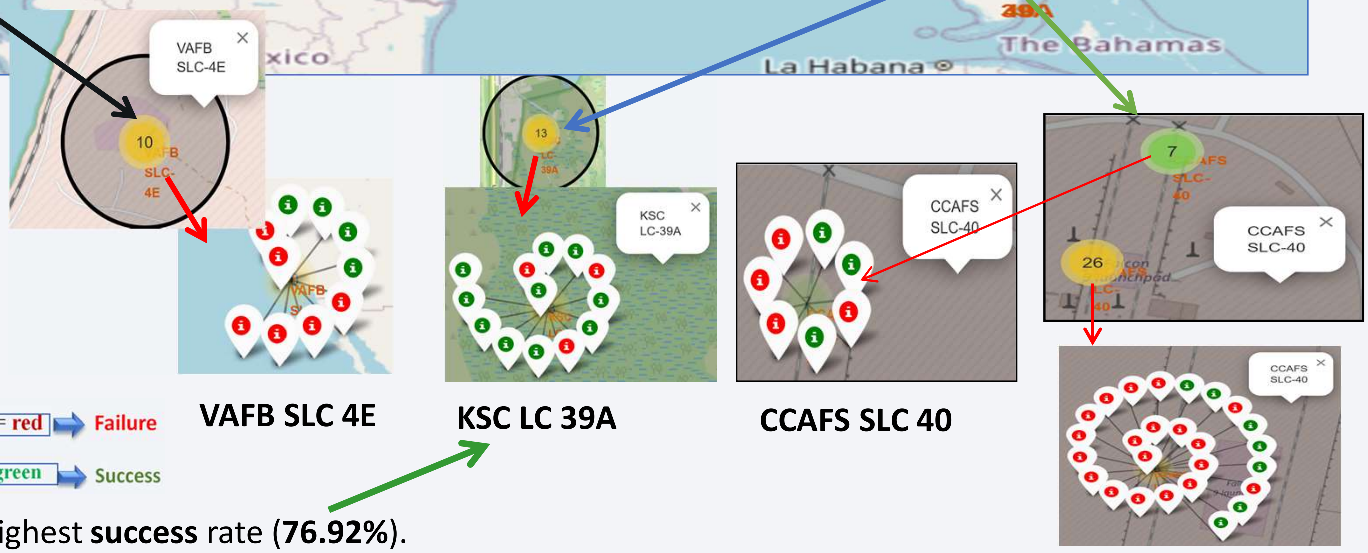
Launch Outcomes

- Red color indicates Failure
- Green color indicates Success

class	marker_color
0	red
1	green

class=0 → marker_color = red → Failure

class=1 → marker_color = green → Success



CCAFS SLC 40

- The site **KSC LC 39A** has the highest **success** rate (**76.92%**).
- As shown, it has **10 successful** launches (**green color**) and **3 failed** launches(**red color**).

Distances between Site CCAFS SLC40 and Its Proximities

Launch site selected: CCAFS SLC40.
Close:

- Railway selected: NASA Railroad,
- Highway selected: Samuel C. Phillips Parkway
- City selected: Cape Canaveral

The screenshot of the selected site with its proximities: railway, highway, coastline, and distance calculated is shown.

Distance from Site **CCAFS SLC40** to the **Closest:**

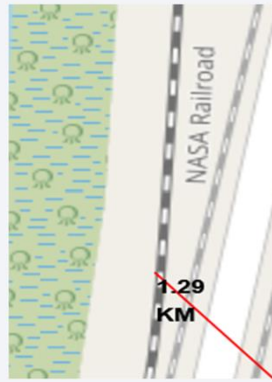
- **Coast** is **0.86 km**
- **Railway** is **1.29 km**
- **Highway** is **0.59 km**
- **Cape Canaveral City is 18.54 km**

It is clear that the **Launch site** is:

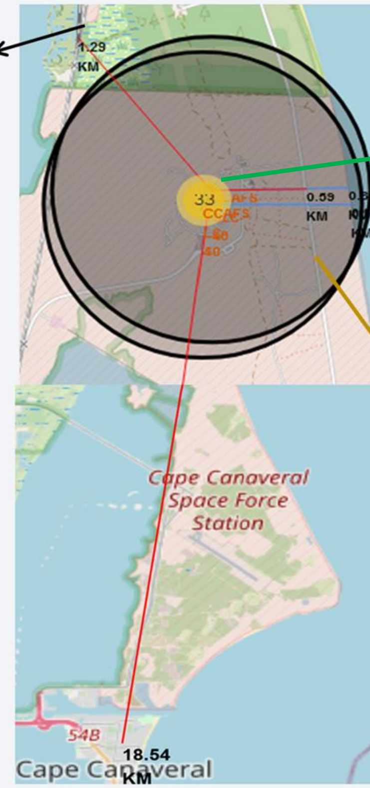
- **close** to the **coast** (**0.86 km**)
- **far** from the **city** (**18.54 km**)

After you plot distance lines to the proximities, you can answer the following questions easily:
Are launch sites in close proximity to railways**? Yes**
Are launch sites in close proximity to highways? **Yes**
Are launch sites in close proximity to coastline? **Yes**
Do launch sites keep certain distance away from cities? **Yes**

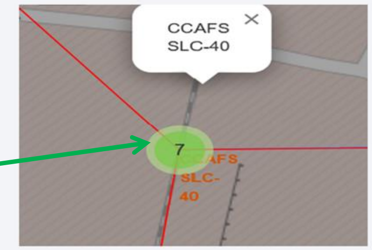
We can **conclude** that the launch sites are strategically located near highways and railways for transportation of cargo and personnel, but also far away from nearby cities for safety reason.

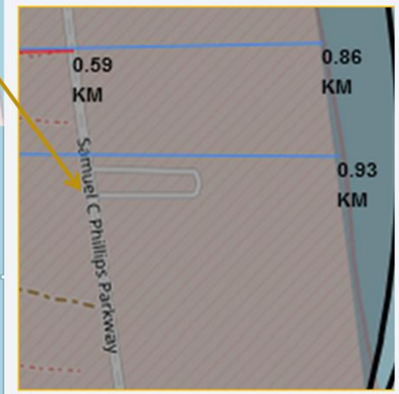


NASA Railroad



Cape Canaveral City, Florida





Samuel C. Phillips Parkway

Section 4

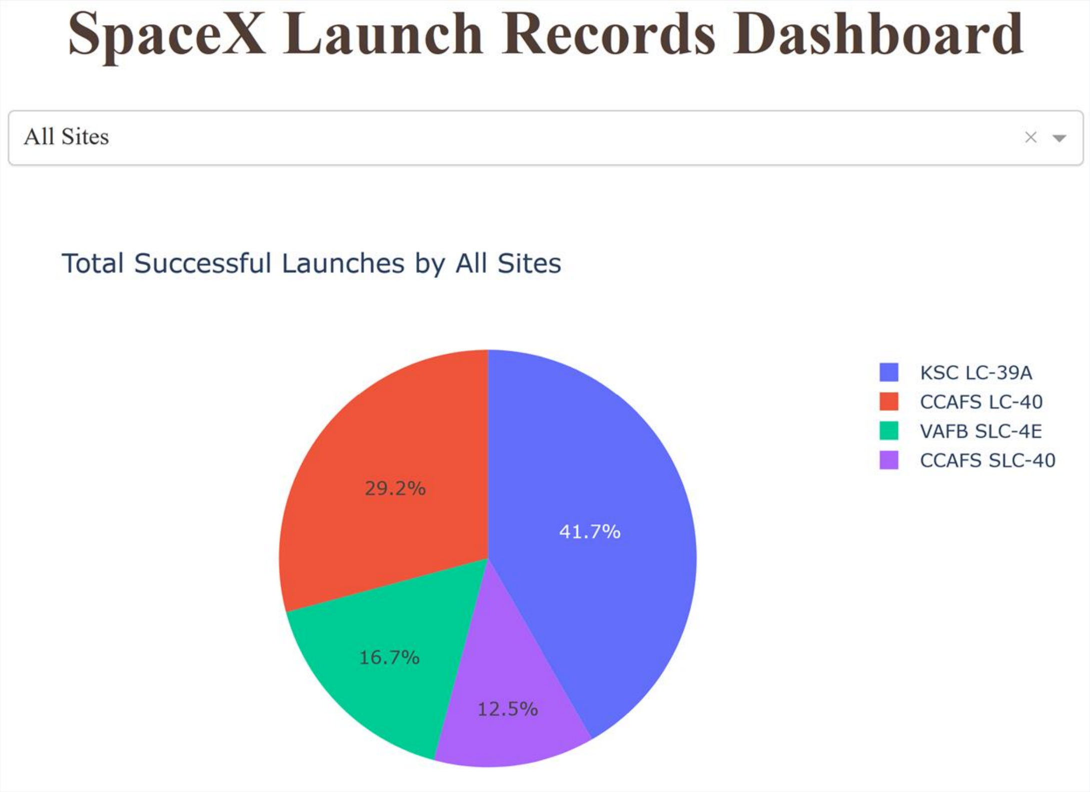
Build a Dashboard with Plotly Dash

Launch Success Count for All Sites

The **pie chart** show the launch success count for **all sites.**

From the pie chart , we can identify the launch site with highest success rate as well as with least success rate of all sites. We find that:

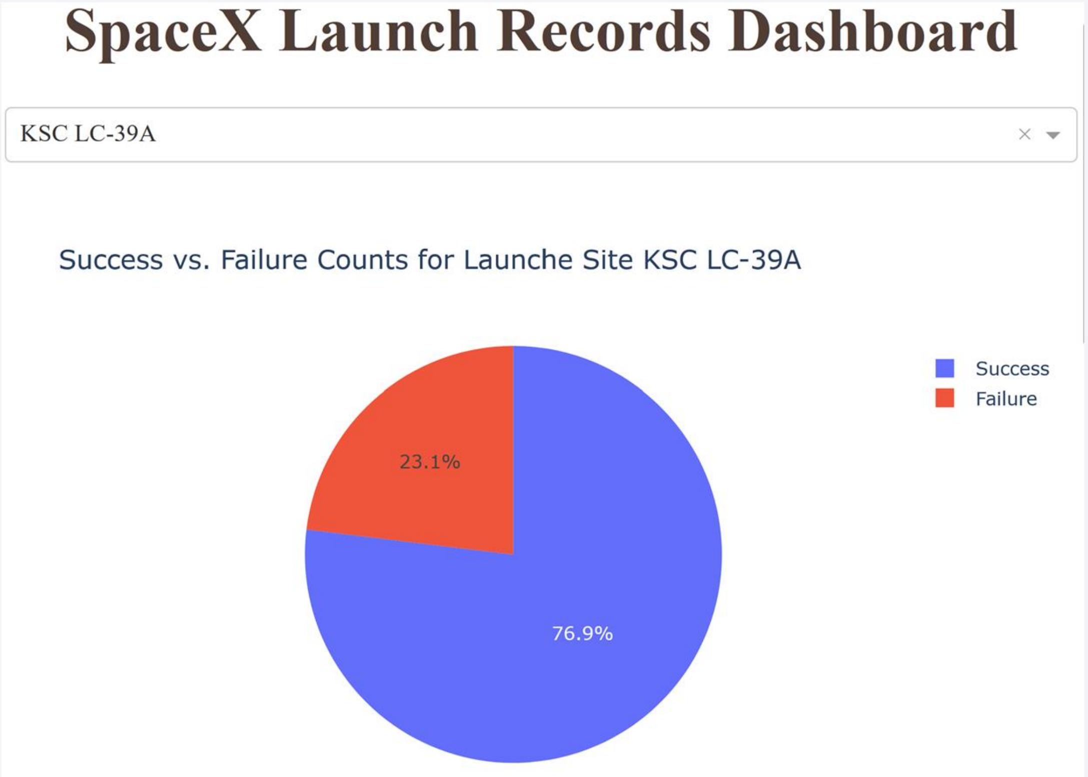
- The launch site **KSC LC-39 A** has the **most successful** launches, with **41.7%** of the total successful launches.
- The **CCAF-SLC40** Launch site has the **least successful** launches of all, with **12.5%.**



KSC LC-39A with the Highest Launch Success Ratio

The pie chart **shows Success vs. Failure countts for KSC LC-39A**, which has the highest launch success ratio of all launch sites.

- As we can see from the pie chart , **76.9% of the total launches** at site **KSC LC-39A** were **successful**.
- This result is the highest success rate of all the launch sites.

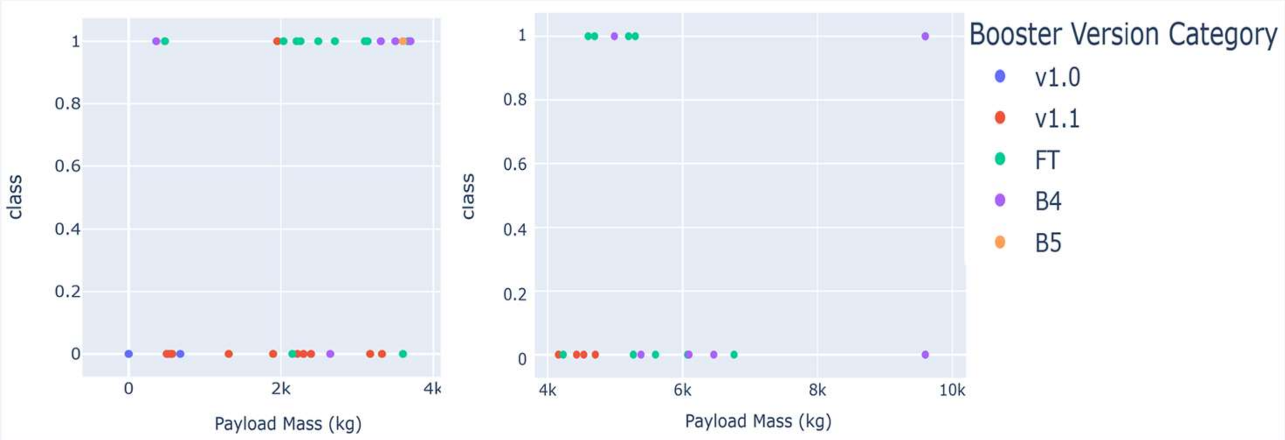


Payload vs. Launch Outcome Scatter Plot for All Sites

The scatter plot show the Payload vs. Launch Outcome for all sites, with different payload selected in the range slider.

The Plot shows unexpected gap at nearby 4000 kg. We split the data into two ranges at 4000 kg, i.e.,

- **light** payloads (0kg – 4000 kg)
- **weighty** payloads (4000 – 10000 kg)



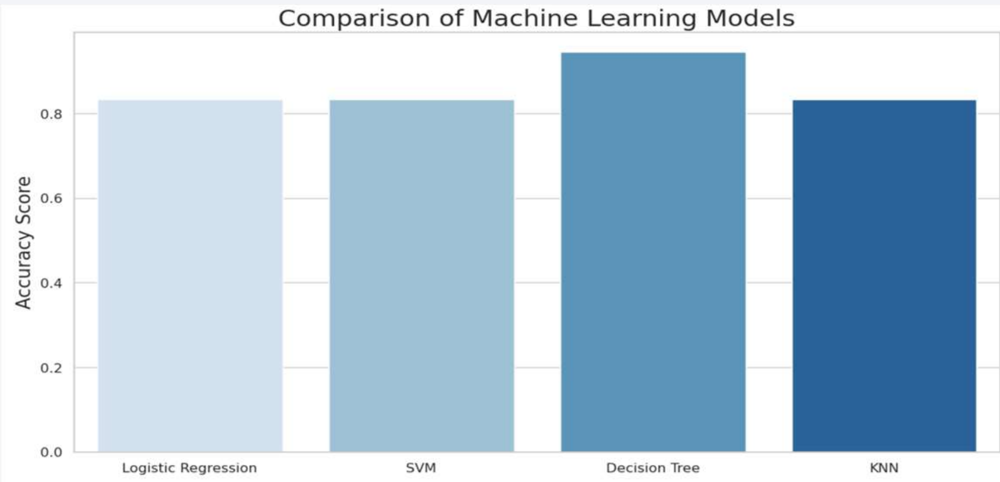
As seen from the plots:

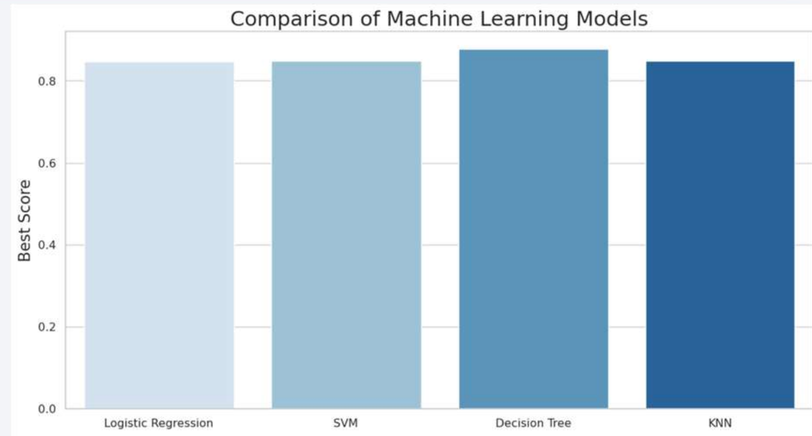
- the success for **light** payloads is **greater** than the **weighty** ones.
- the range between 2000 kg and 4000 kg has the highest success in the **light** payload.
- the range between 6000 kg and 10000 kg has the lowest success in the **weighty** payload.
- The success was low in the light payload range between 0kg and 2500kg
- **The Booster** version **FT** has the largest success than other booster versions.

Section 5

Predictive Analysis (Classification)

Classification Accuracy





Model	Accuracy Score	Best Score
Logistic Regression	0.833333	0.846429
SVM	0.833333	0.848214
Decision Tree	0.944444	0.876786
KNN	0.833333	0.848214

The bar chart and the Table show the **Accuracy Score** and the **Best Score** for all built classification models (*Decision Tree, K Nearest Neighbors, Support Vector Machine and Logistic Regression*) . All the models except the Decision Tree have same accuracy score with 83.33%.

The Decision Tree model has the highest classification accuracy with an **Accuracy Score** of **94.44%** and **Best Score** with **87.68%.**

Confusion Matrix

The confusion matrix of the best performing model, Decision Tree is shown. From the confusion matrix we have:

True **Positive: 5**

False Positive (**Error type I**): **1**

True **Negative: 12**

False Negative (Error type II) : **0**

Number of **True labels** landed: **12**

Number of **True labels** did not **land: 6**

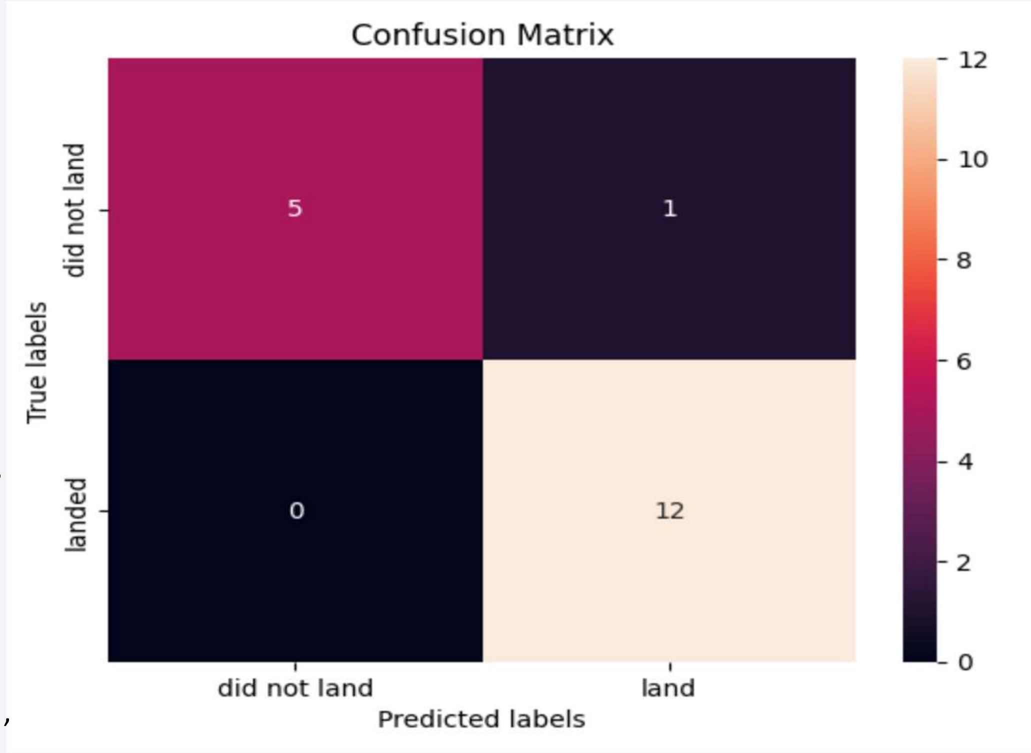
Accuracy : ratio of total correctly predicted labels to total labels

= 17/18 = 0.9444

The matrix gives same **accuracy with 94.44%** as expected!

In general, the **confusion matrix shows :**

- **18 True labels**, of which **12** landed, **6 did not land**,
- **1 label** is **wrongly classified** (a **false positive** or **Error type I**) ,
- **17 labels** are correctly classified, **12** landed, **5 did not land**,
- The decision tree is the **best model** with **accuracy 94.44%.**



Conclusions

The main objective of this Capstone project was to predict if the Falcon 9 first stage will land successfully. Based on the results obtained, the following conclusions can be drawn:

1. There is a positive correlation between the number of flights and the success rate. As the number of flights increases, the success rate of landing also increased.
2. The exploratory data analysis revealed that the success rate of the SpaceX Falcon 9 rocket landings is 66.67%.
3. The landing launch success rate increased since 2013, in spite of small dips in 2018 and 2020.
4. Some Orbits like ES-L1, GEO, HEO, and SSO have the most success rate (100%) than all other orbits.
5. The success rate of the Falcon 9 landing can be correlated with the payload mass as the lighter payloads are generally proved to be more successful than the heavier ones.
6. The launch site KSC LC-39 A had the most successful launches, with 41.7% of the total successful launches, and also the highest rate of successful launches, with a 76.9% success rate.
7. The Decision Tree classifier is the best machine learning model to use for this project, as it has the highest accuracy with 94.44%.
8. The launch sites are strategically located near highways and railways for transportation of cargo and personnel, but also far away from nearby cities for safety reason.

Appendix

- https://github.com/Hayethdt/thanks-cse/blob/main/edadataviz.ipynb
- https://github.com/Hayethdt/thanks-cse/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb
- https://github.com/Hayethdt/thanks-cse/blob/main/jupyter-labs-spacex-data-collection-api.ipynb
- https://github.com/Hayethdt/thanks-cse/blob/main/jupyter-labs-webscraping.ipynb
- https://github.com/Hayethdt/thanks-cse/blob/main/lab_jupyter_launch_site_location.ipynb
- https://github.com/Hayethdt/thanks-cse/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb
- https://github.com/Hayethdt/thanks-cse/blob/main/Spacex_Dashboard_project.ipynb
- https://github.com/Hayethdt/thanks-cse/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Thank you!