IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Juan Francisco Araujo Bayas

24/11/2023

Outline

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

Executive Summary

- Summary of methodologies

The different **models** used on this project, for the **reuse of the first stage prediction** are as follows: ***logistic regression, support vector machine, decision tree classifier, k nearest neighbors.***

- Summary of all results

The accuracy when testing the **different models** are as follows:

- **Logistic regression:** 0,833
- **Support vector machine:** 0,833
- **Decision tree classifier:** 0,56
- **k nearest neighbors:** 0,833

Based on those **results, we can** say that the **best prediction models** for the reuse of the first stage are: ***logistic regression, support vector machine, and k nearest neighbors*****.**

Introduction

- Project background and context
  - Companies (Virgin Galactic, Blue Origin, SpaceX), are making **space travel affordable** for everyone.
  - SpaceX rocket launches are **inexpensive due to the savings made on the reuse** of the first stage.

- Problems you want to find answers
  - By gathering information about SpaceX and creating dashboards for our team, **we will determine the price of each launch**.
  - By training **machine learning model and use public information to predict the outcome**, we will determine if SpaceX will reuse the first stage (if **successful**) or not.

Section 1

Methodology

Methodology

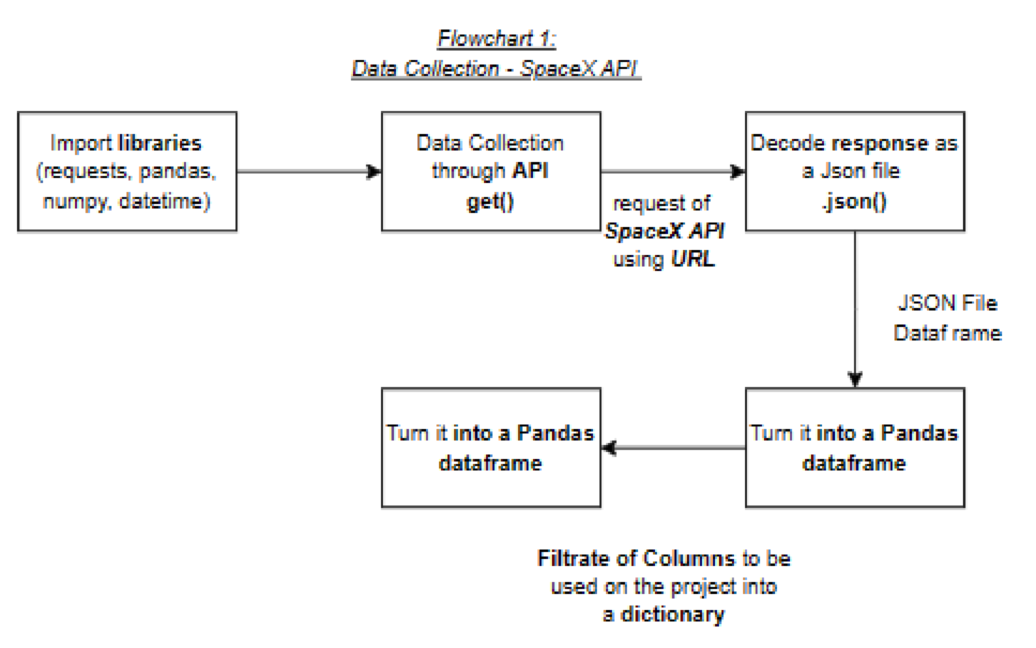
Executive Summary

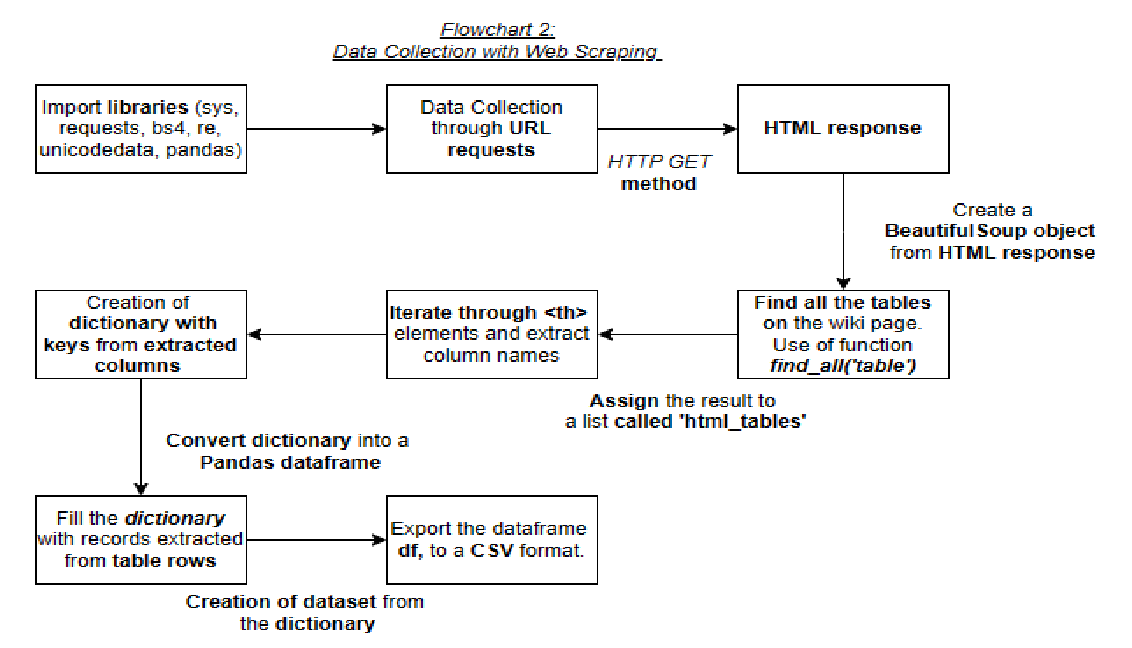
- Data collection methodology:
  - In order to collect Data, we had used two ways: first calling an API through an URL in the form of a JSON file, and secondly by doing Web scraping from wikipedia.
- Perform data wrangling
  - This step, was done through three processes: on a first hand we had filtrated some columns that will be used for this project, then we selected launches from the Falcon 9 and lastly we replace numerical missing values (not for non numerical values) by the mean of the variable column.
- 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

- Describe how data sets were collected.

In order to collect Data, we had used two ways: first calling an API through an URL in the form of a JSON file, and secondly by doing Web scraping from wikipedia.

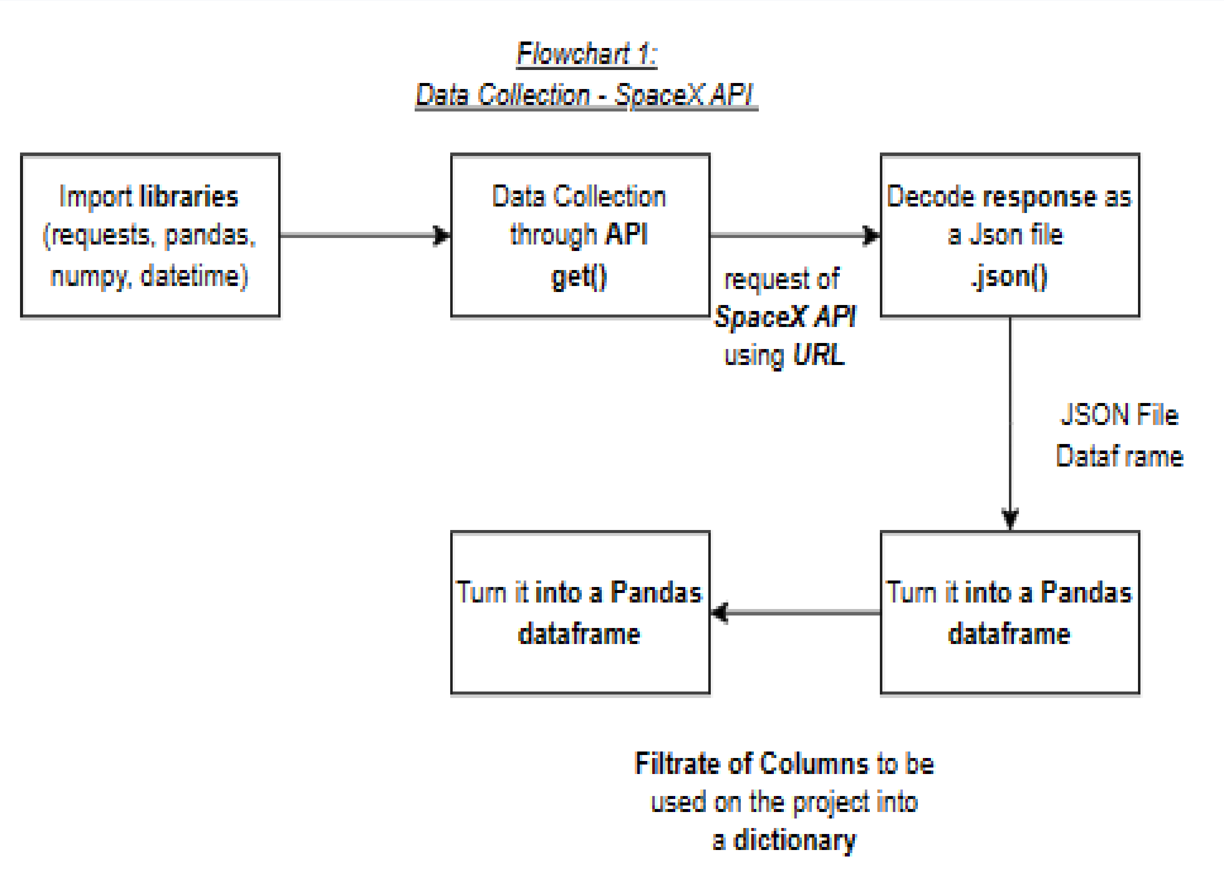




Data Collection – SpaceX API

- On this flowchart, we can see the steps done for the data collection with SpaceX REST.

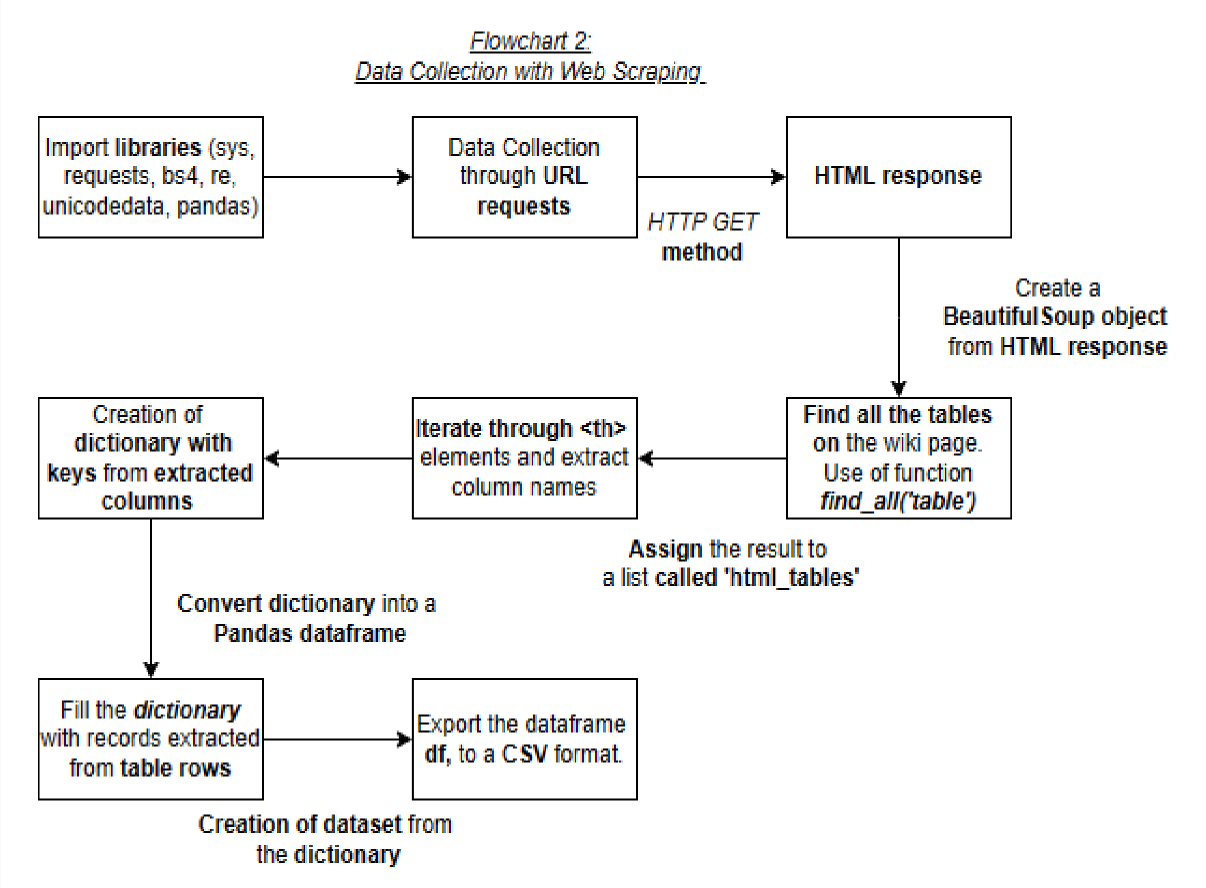
- The GitHub URL of the completed SpaceX API calls notebook is shown here: [testcapproj/Lab 1 Week 1 Capstone Project python code/Lab1 Week 1 Complete Data Collection with Collection API Lab.ipynb at main · juan15-fran/testcapproj (github.com)](github.com) . It can be used as an external reference and peer-review purpose.



Data Collection - Scraping

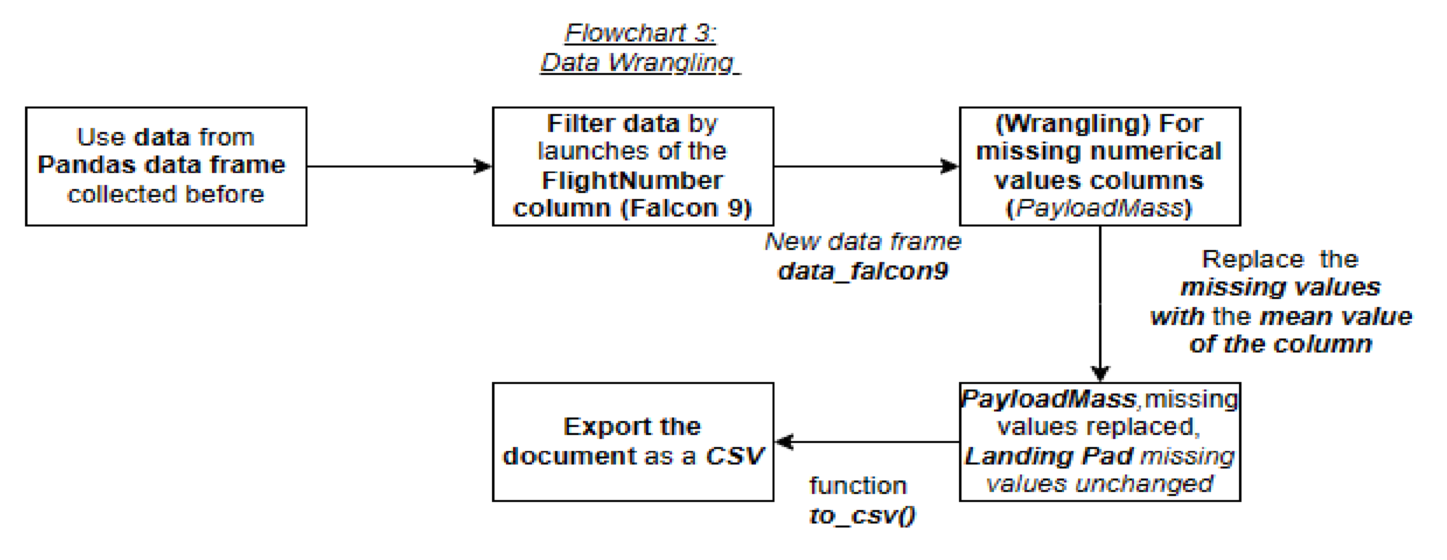
- On this flowchart, we can see the steps done for the data collection with Web scraping from wikipedia.

- The GitHub URL of the completed web scraping notebook is shown here: [testcapproj/Lab2 Week 1 Capstone Project python code/Lab 2 Week 1 Complete the Data Collection with Web Scraping Lab.ipynb at main · juan15-fran/testcapproj (github.com)]. It can be used as an external reference and peer-review purpose.



Data Wrangling

- After collecting the data, we **wrangled the data** by replacing some **missing values, from** *PayloadMass* column by it´s **mean value**. And for the ***Landing Pad*** **column values**, we **leaved without any change** as they aren’t numerical.



- The GitHub URL of the completed data wrangling related notebooks is as follows: [testcapproj/Lab3 Week 1 Capstone Project python code/Lab 3 Week 1 Data Wrangling.ipynb at main · juan15-fran/testcapproj (github.com)] . This can be used as an external reference and peer-review purpose.

EDA with Data Visualization

- The **charts** that **were plotted** are exposed as follows:

  - ***FlightNumber vs PayloadMass (kg)***: to study their ***relationship*** and the ***frequency of launching failure***
  - ***FlightNumber vs LaunchSite***: to study their ***relationship*** and the ***frequency of launching failure***
  - ***PayloadMass (kg) vs LaunchSite***: study their ***relationship*** and the ***frequency of launching failure***
  - ***Success Rate and Each Orbit Type***: to study their ***relationship***
  - ***FlightNumber and Orbit Type***: to study their ***relationship*** and the ***frequency of launching failure***
  - ***PayloadMass and Orbit Type***: to study their ***relationship*** and the ***frequency of launching failure***
  - ***Year and average success rate:*** to study the ***average launch success trend***

- The GitHub URL of the completed EDA with data visualization notebook is as follows: testcapproj/Lab2 Week 2 Falcon 9 First Stage Landing Prediction/Lab 2 Week 2 Falcon 9 First Stage Landing Prediction.ipynb at main · juan15-fran/testcapproj (github.com). It can be used for external reference and peer-review purpose.

EDA with SQL

- The SQL queries performed on the data frame **'SPACEXTABLE',** for EDA with SQL are **exposed as follows:**

Queries	*Explanation*
*select *from*	To display the columns from the data frame.
select distinct column name from	To select the different values that are on this column.
select column name from data frame where 'CCA%' like limit 5	To display 5 records where launch sites begin with 'CCA'.
select column name from data frame where Customer = 'NASA (CRS)'	To display the total payload mass carried by boosters launched by NASA (CRS).
select column name from data frame where Booster_Version = 'F9 v1.1'	To display payload mass carried by booster version F9 v1.1.
select min(Date) as min_date from data frame where Landing_Outcome = 'Success (ground pad)'	To list the date when the first successful landing outcome in ground pad was achieved.
select colum name from data frame where (Landing_Outcome = 'Success (drone ship)') and (PAYLOAD_MASS_KG> 4000 and PAYLOAD_MASS_KG< 6000);	To list the names of the boosters which have success in drone ship and have payload mass **greater than 4000 but less than 6000.**
select (column name) as successful_mission from data frame where Landing_Outcome LIKE 'Success%'	To list the total number of successful mission outcomes.
select column name , count (Landing_Outcome) as failure_mission from data frame where Landing_Outcome LIKE 'Failure%'	To list the total number of failure mission outcomes.
select column name from data frame where PAYLOAD_MASS_KG_ = (select MAX(PAYLOAD_MASS_KG_) from data frame)	To list the names of the booster_versions which have carried the maximum payload mass (Subquery).
select substr(**Date, 6,2**) *as Month, **Date**, Landing_Outcome, Booster_Version, Launch_Site from SPACEXTABLE where (Landing_Outcome LIKE 'Failure (drone ship)') and (substr(**Date,0,5**)='2015')*	To list records that will display the **month names, failure landing_outcomes in drone ship, booster versions, launch_site** for the **month in year 2015.**
*select * from SPACEXTABLE where (**Date between** '2010-06-04' and '2017-03-20') and (Landing_Outcome = 'Success (ground pad)') or (Landing_Outcome = 'Failure (drone ship)')*	To **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.**

- The GitHub URL of the completed EDA with SQL notebook is as follows: [testcapproj/Lab1 Week 2 Complete the EDA with SQL/Lab 1 Week 2 Complete the EDA with SQL .ipynb at main · juan15-fran/testcapproj (github.com)](). It can be used for external reference and peer-review purpose.

Build an Interactive Map with Folium

- The map objects ***(markers, circles, lines, etc)*** created and added to a folium map, are as follows:

➢ ***Highlighted circle area with a text label:*** to **locate each launch site** on the map,

➢ ***MarkerCluster object:*** to mark the **success/failed for each site** on the map,

➢ ***Polyline:*** to draw a **PolyLine between a launch site** to the selected **coastline point, closest city, railway and highway**. This in order to compute the **distance among the different points**.

- The GitHub URL of the completed interactive map with Folium map is as follows: testcapproj/Lab1 Week 3 Interactive Visual Analytics with Folium/Lab1 Week 3 Interactive Visual Analytics with Folium.ipynb at main · juan15-fran/testcapproj (github.com). It can be used for external reference and peer-review purpose.

Build a Dashboard with Plotly Dash

- We had created the following plots/graphs and interactions in a dashboard:

➢ ***Launch Site Drop-down Input Component***: to select the Launch Site, from which we want to display the graph information.

➢ ***Success pie-chart:*** to display the success rate from the different Launch Site selected from the dropdown.

➢ ***Range Slider:*** to select the Payload.

➢ ***Success-payload-scatter-chart scatter plot:*** to render the success payload scatter plot.

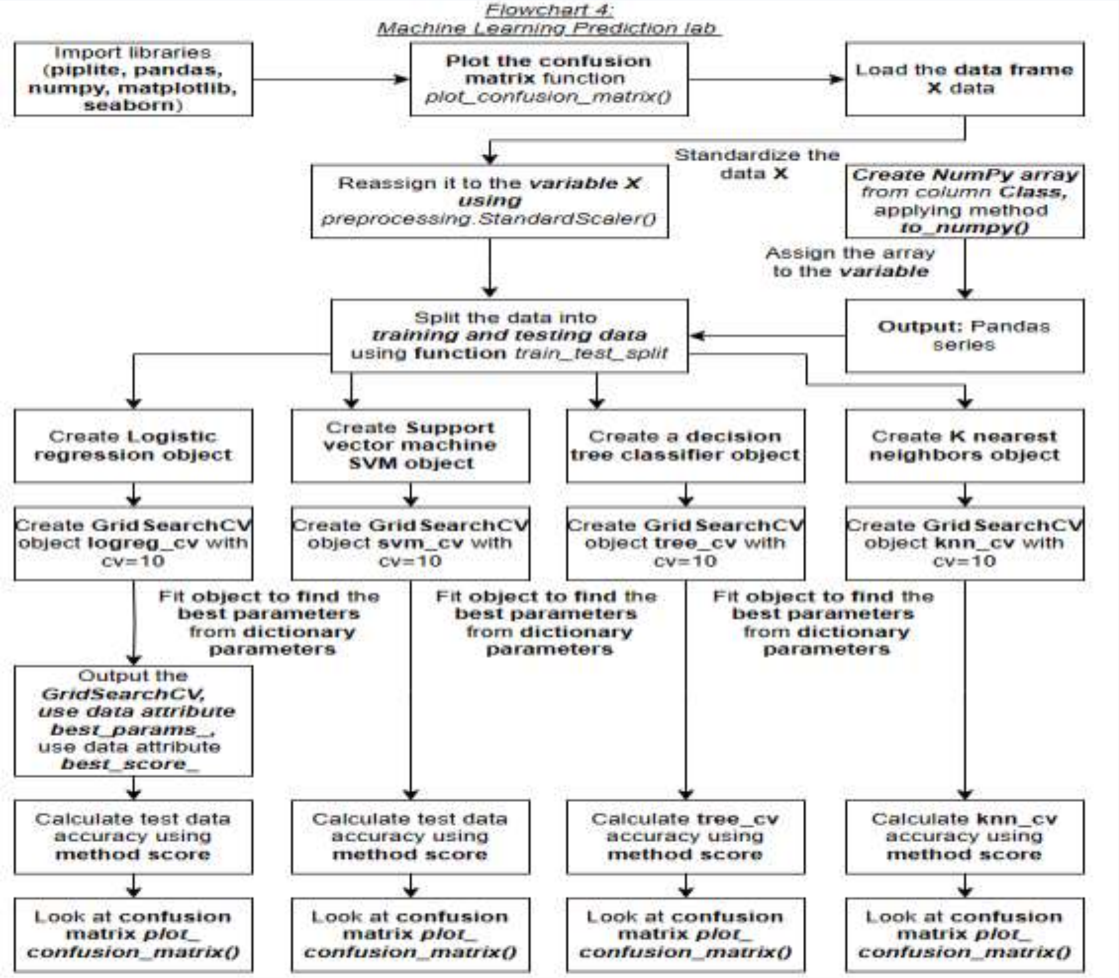
- The GitHub URL of the completed Plotly Dash lab is as follows: testcapproj/Lab2 Week 3 Build an Interactive Dashboard with Plotly Dash/Lab 2 Week 3 spacex_dash_app.py at main · juan15-fran/testcapproj (github.com). It can be used as an external reference and peer-review purpose.

Predictive Analysis (Classification)

- In order to predict whether the first stage of Falcon 9 will land successfully, we will use 4 models ***(Logistic Regression, Support Vector Machine, Decision Tree Classifier and K-nearest Neighbors)*** which will then be selected taking into account their **individual accuracy.**

- **In the following chart we can see the steps** done in order to come up with the **models.**

- The GitHub URL of the completed predictive analysis lab is as follows: [testcapproj/Lab1 Week 4 Hands-On Lab Complete the Machine Learning Prediction Lab/Lab 1 Week 4 Hands-On Lab Complete the Machine Learning Prediction Lab.ipynb at main · juan15-fran/testcapproj (github.com)](). It can be used as an external reference and peer-review purpose.



Results

- From the ***Exploratory data analysis*** we have the following results:

➢ failure mission outcomes are 10 (**14,08%**), and successful mission outcomes are 61 **(85,92%);**

➢ the **success rate** from the different Launch Sites are ***CCAPS SLC 40*** (60%), ***VAFB SLC 4E*** (76,92%), ***KSC LC 39A*** (77,27%)**;**

➢ the **Payload Mass (kg)** with the **higher success rate** are among the following Masses: 1700 (kg) - 2500 (kg), 3800 (kg) - 4300 (kg), 8000 (kg) – 15700 (kg)**;**

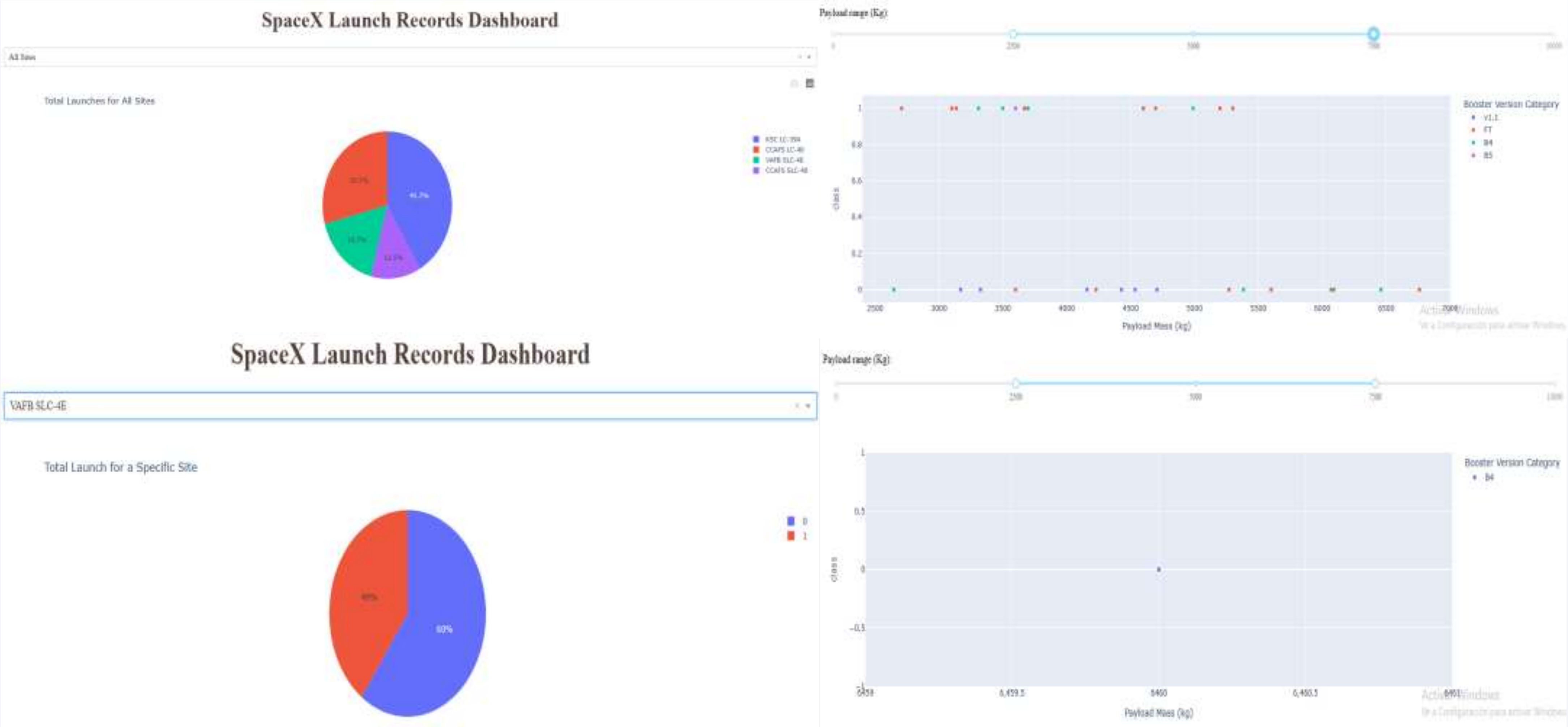
➢ The Orbits Type with most of successes are, **ES-11, GEO, HEO and SSO**. This could be linked to ***the launch site from where the rocket*** is launched. Due to the launch aims to a dedicated Orbit.

- Interactive analytics demo in screenshots

- Predictive analysis results

Results

- Interactive analytics demo in screenshots



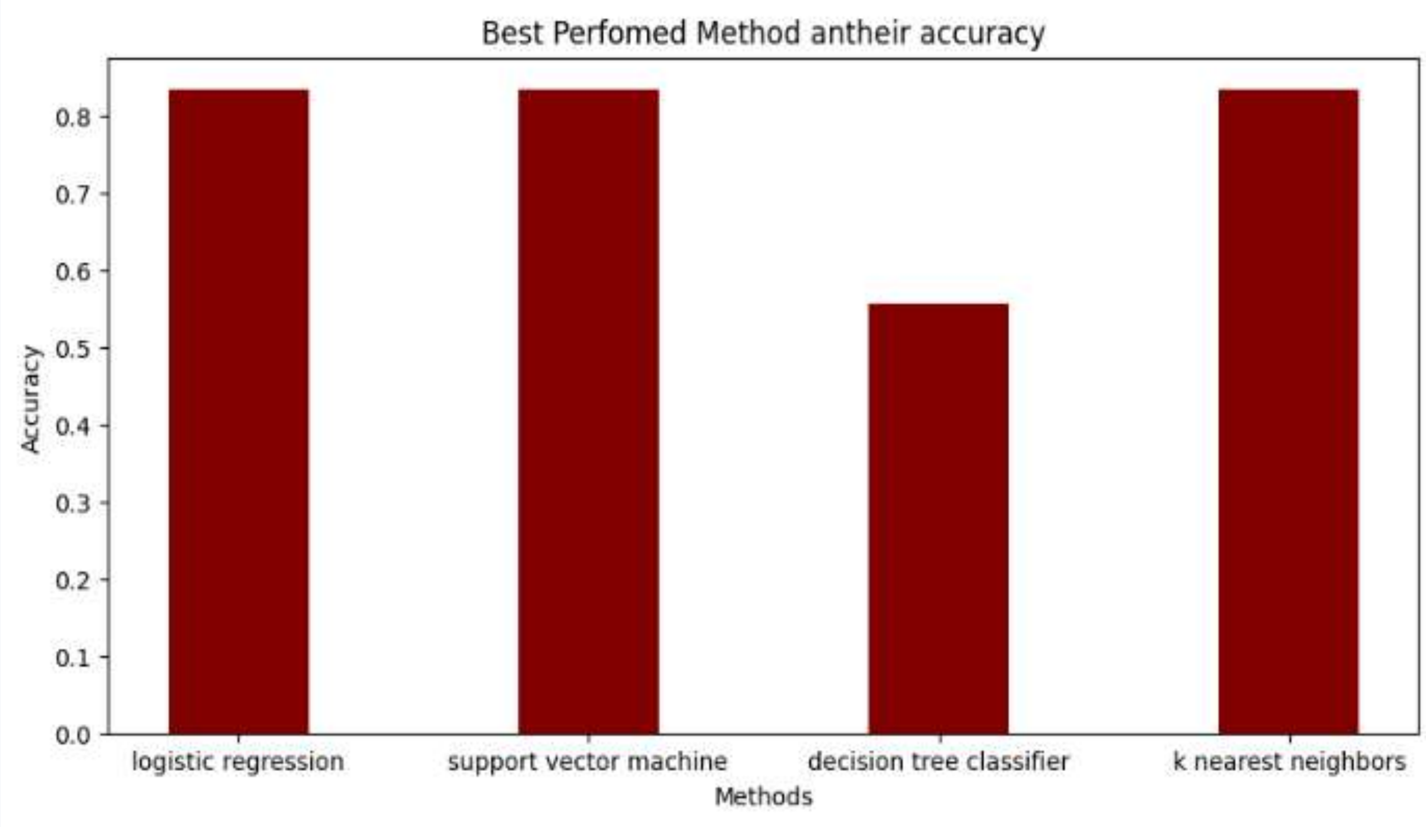
Results

- Predictive analysis results

The accuracy when testing the **different models** are as follows:

- **Logistic regression:** 0,833
- **Support vector machine:** 0,833
- **Decision tree classifier:** 0,56
- **k nearest neighbors:** 0,833

Based on those **results, we can** say that the **best prediction models** for the reuse of the first stage are: ***logistic regression, support vector machine, and k nearest neighbors*.**

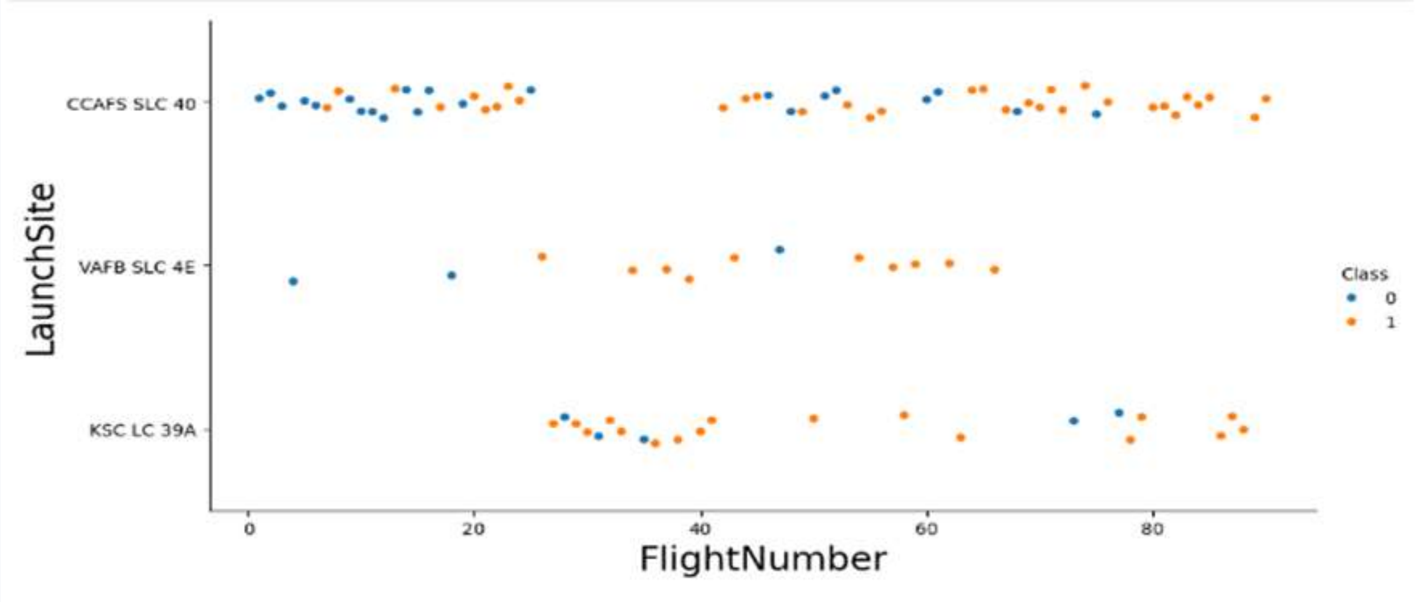


Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

<u>**Scatter plot of Flight Number vs. Launch Site**</u>

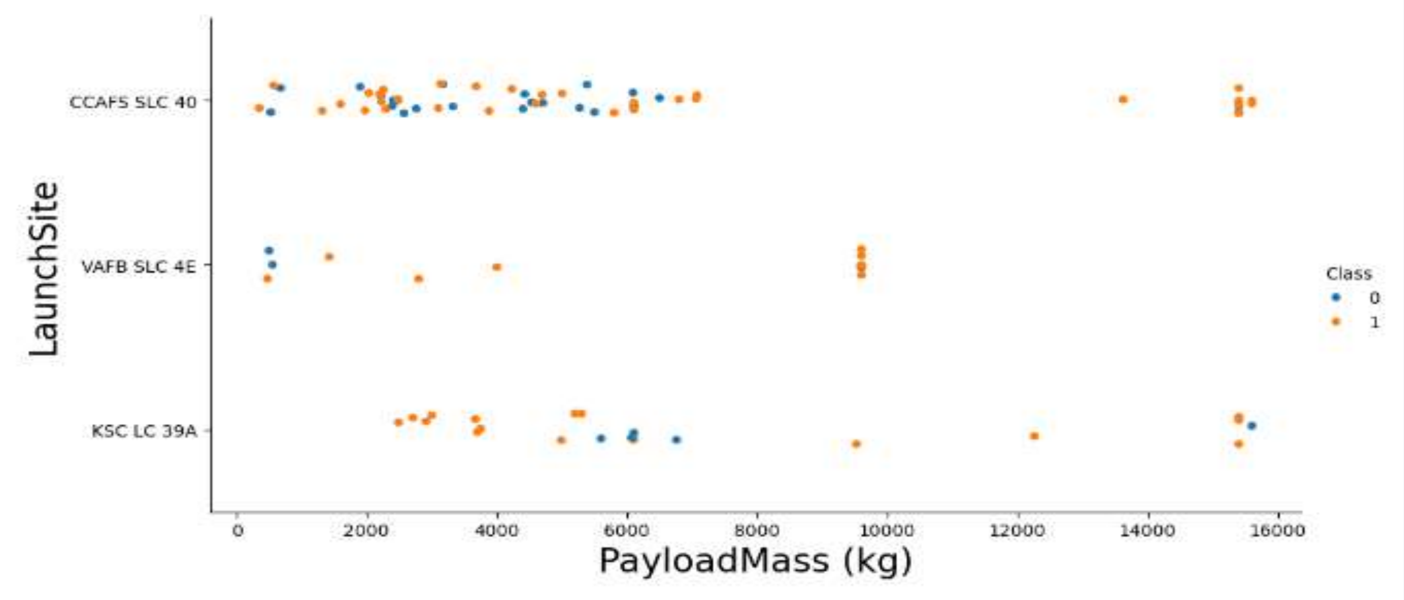


In this scatter plot we can see through the data of class (1= success; 0= fail) the **success rate from the different Launch Sites are**:

- CCAPS SLC 40 (60%),
- VAFB SLC 4E (76,92%),
- KSC LC 39A (77,27%).

Payload vs. Launch Site

Scatter plot of Payload vs. Launch Site



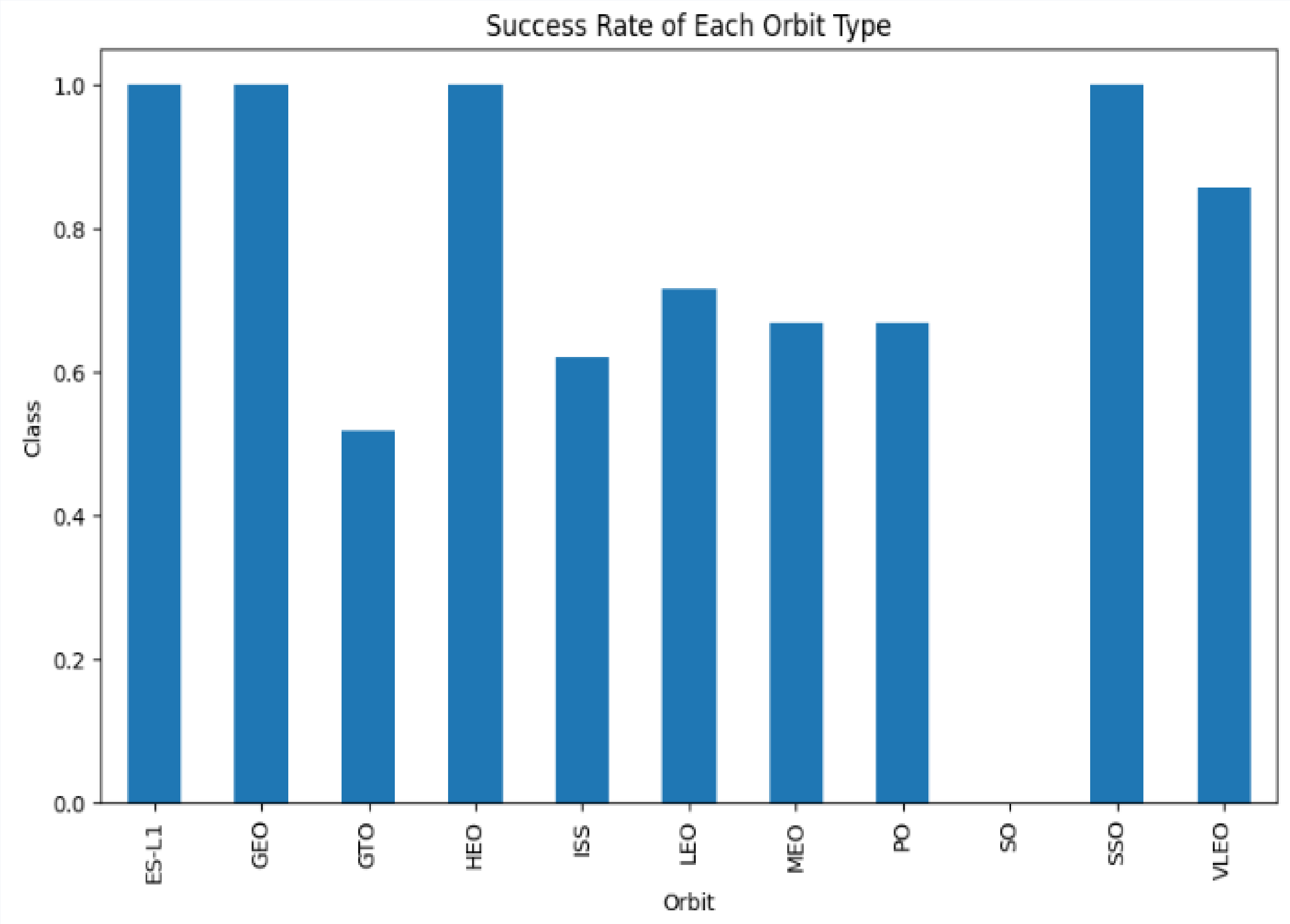
In this scatter plot we can see through the data of class (1= success; 0= fail), that the **Payload Mass (kg) with the higher success rate** are among the following Masses:

- 1700 (kg) - 2500 (kg)
- 3800 (kg) - 4300 (kg)
- 8000 (kg) – 15700 (kg)

Success Rate vs. Orbit Type

- On the right hand side, we can see a **bar chart** of **the success rate of Each Orbit Type.**

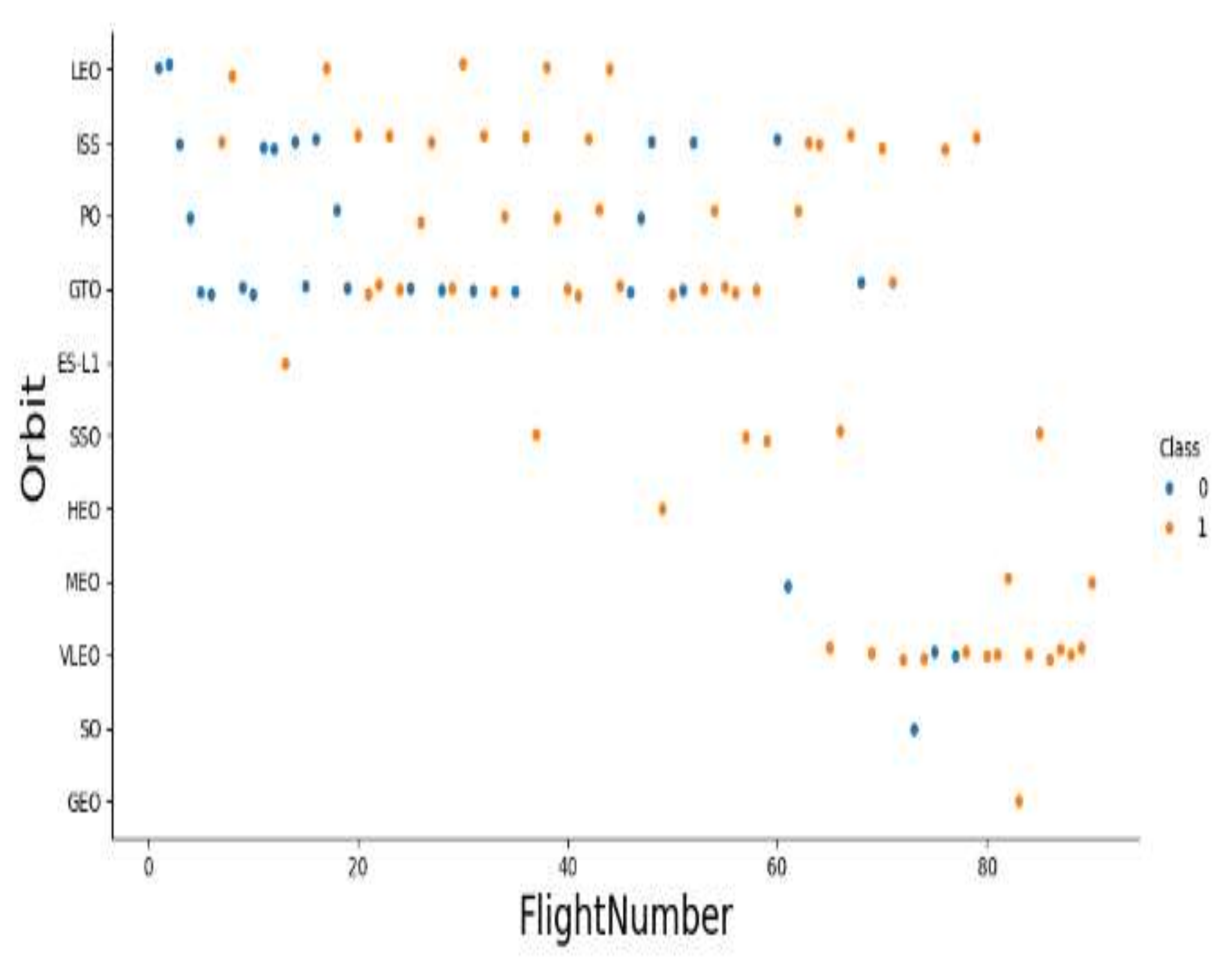
- The Orbits Type with most of successes are, **ES-11, GEO, HEO and SSO**. This could be linked to ***the launch site from where the rocket*** is launched. Due to the launch aims to a dedicated Orbit.



Flight Number vs. Orbit Type

- On the right hand side, we can see the ***scatter plot of the Flight Number vs Orbit Type***. Where **blue points shows fail launches** and **orange points shows success launches**.

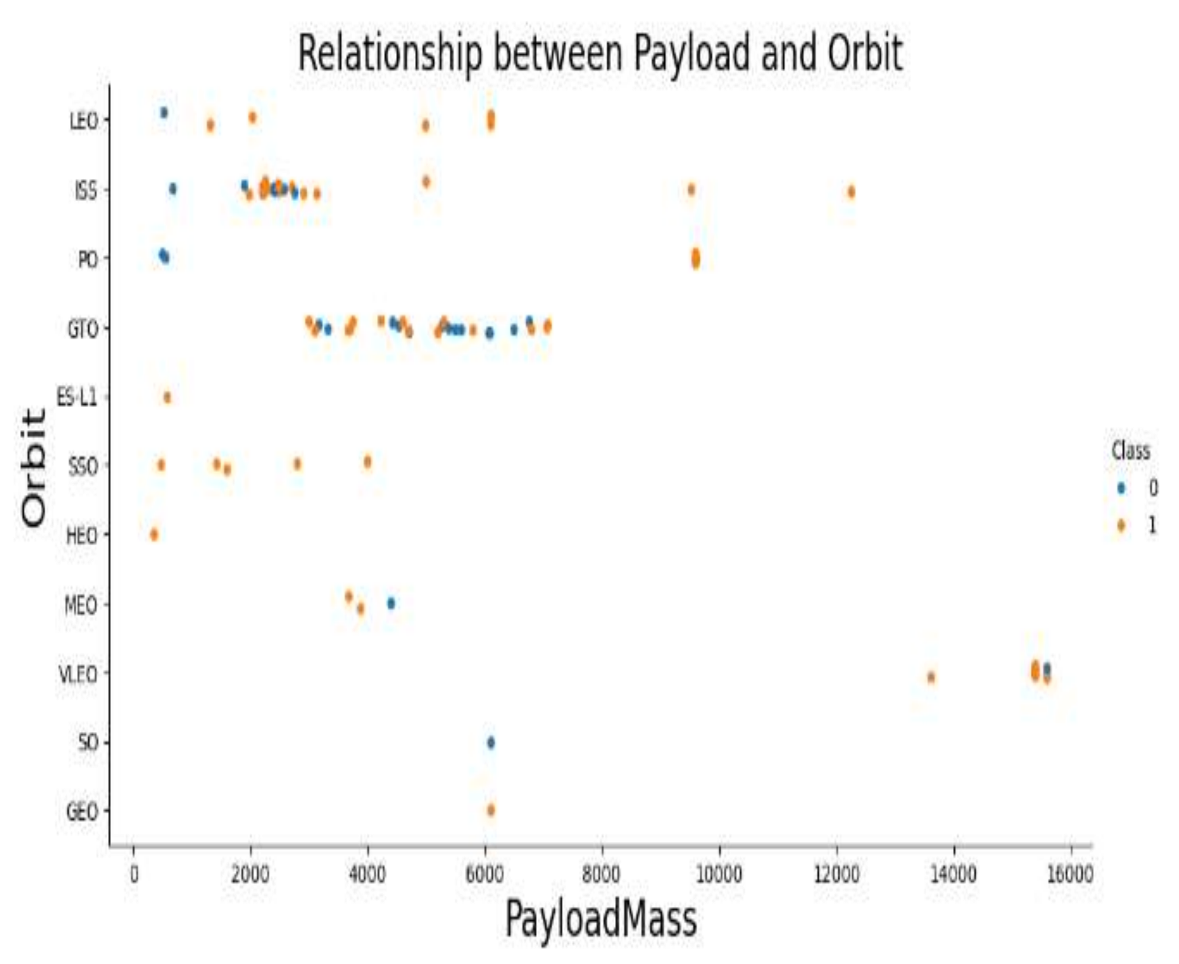
- On this ***scatter plot*** we can see that the Orbit type with the most failure rate are: **SO, GTO, ISS, VLEO and LEO. The rate,** could change with **more launches made** to each Orbit Type.



Payload vs. Orbit Type

- On the right hand side, we can see the ***scatter plot of the PayloadMass vs Orbit Type.*** Where **blue points shows fail launches** and **orange points shows success launches**.

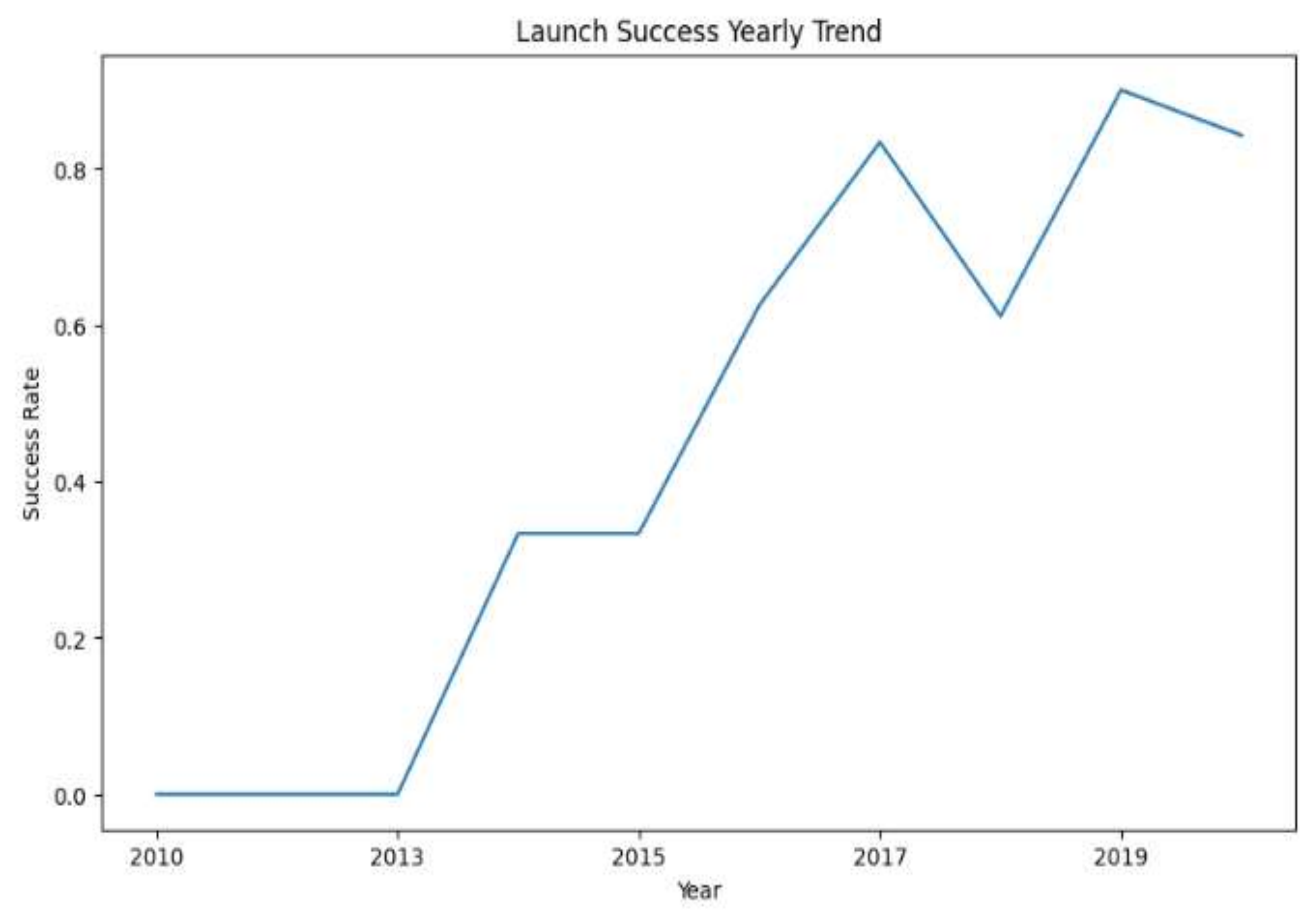
- The Orbits Type with most of successes are, **ES-11, GEO, HEO and SSO**. This could be linked to ***the launch site from where the rocket*** is launched. Due to the launch aims to a dedicated Orbit. We can say that is more linked to the launch site, than to the PayloadMass, as with the **same PayloadMass** we **have failures for other** Orbit Types of **rocket launches**.



Launch Success Yearly Trend

- On the right hand side, we can see the ***plot line chart of the Launch Success Yearly Trend from 2013 to 2020.***

- The increase on improvement on **success rate,** through the years **could be explained** due to the improvements made on the rocket launch from old failed launches.



All Launch Site Names

- In order to display the **names from the different launch sites in** the **space mission,** we had used the following query:

➢ **select distinct** *Launch_Site* ***from*** *SPACEXTABLE.*

```
%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

Launch Site Names Begin with 'CCA'

- In order to display**, 5 records where launch sites** begin with **`CCA`.** We had used the following query:

➢ ***select** Launch_Site **from** SPACEXTABLE **where** Launch_Site **like** ‘CCA%’ **limit5.***

```
%sql select Launch_Site from SPACEXTABLE where Launch_Site like'CCA%' limit 5;
```

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

Launch_Site
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40

Total Payload Mass

- In order to compute the **total payload mass** carried by boosters from NASA (CRS). We have used the following **query:**

➢ ***select*** *sum(PAYLOAD_MASS_KG_)* ***from*** *SPACEXTABLE* ***where*** *Customer = ‘NASA (CRS)’.*

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

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

sum(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

- In order to compute the **average payload mass** carried by **booster version F9 v1.1**

➢ ***select*** *avg(PAYLOAD_MASS_KG)* ***from*** *SPACEXTABLE* ***where*** *Booster_Version = ‘F9 v1.1’.*

```
%sql select avg(PAYLOAD_MASS__KG_) from SPACEXTABLE where Booster_Version = 'F9 v1.1';
```

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

avg(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

- In order to **find the dates** of the **first successful landing outcome** on **ground pad,** we had used the following query:

➢ ***select** min(**Date**) **as** min_date **from** SPACEXTABLE **where** Landing_Outcome='Success (ground pad)'.*

```
%sql select min(Date) as min_date from SPACEXTABLE where Landing_Outcome='Success (ground pad)'  ;
```

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

min_date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- In order to **list the names of boosters which have successfully landed on drone ship** and had **payload mass greater than 4000 but less than 6000**, we had used the following query:

➢ ***select*** *Booster_Version* ***from*** *SPACEXTABLE* ***where*** *(Landing_Outcome=‘Success (drone ship)’)* ***and*** *(PAYLOAD_MASS_KG_>****4000 and*** *PAYLOAD_MASS_KG_<****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

Total Number of Successful and Failure Mission Outcomes

- In order to **compute the total number of successful and failure mission outcomes**, we had used the following queries:

➢ ***Select*** *Landing_Outcome* ***,count*** *(Landing_Outcome)* ***as*** *successful_mission* ***from*** *SPACEXTABLE* ***where*** *Landing_Outcome* ***LIKE*** *'Success%'.*

➢ ***Select*** *Landing_Outcome* ***,count*** *(Landing_Outcome)* ***as*** *failure_mission* ***from*** *SPACEXTABLE* ***where*** *Landing_Outcome* ***LIKE*** *'Failure%'.*

```
%sql select count(Landing_Outcome) as successful_mission from SPACEXTABLE where Landing_Outcome LIKE 'Success%';
```

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

successful_mission
61

```
%sql select Landing_Outcome ,count(Landing_Outcome) as failure_mission from SPACEXTABLE where Landing_Outcome LIKE'Failure%';
```

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

Landing_Outcome	failure_mission
Failure (parachute)	10

Boosters Carried Maximum Payload

- In order to list the names of the booster which have carried the maximum payload mass, we had used the following query:

➢ ***select*** *Booster_Version* ***from*** *SPACEXTABLE* ***where*** *PAYLOAD_MASS_KG = (****select*** *MAX(PAYLOAD_MASS_KG_)* ***from*** *SPACEXTABLE.*

```
%sql select Booster_Version from SPACEXTABLE where PAYLOAD_MASS__KG_ = (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

- To list the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015, we had used the following query:

➢ ***select*** *substr(****Date, 6,2****)* ***as*** *Month****, Date,*** *Landing_Outcome,Booster_Version,Launch_Site* ***from*** *SPACEXTABLE* ***where*** *(Landing_Outcome* ***LIKE*** *'Failure (drone ship)')* ***and*** *(substr(****Date,0,5****)* ***='2015'****).*

```
%sql select substr(Date, 6,2) as Month, Date, Landing_Outcome,Booster_Version,Launch_Site from SPACEXTABLE where (Landing_Outco
```

```
,Booster_Version,Launch_Site from SPACEXTABLE where (Landing_Outcome LIKE'Failure (drone ship)') and (substr(Date,0,5)='2015');
```

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

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

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

- In order to 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. We had used the following query:

➢ ***select*** ** **from** SPACEXTABLE **where** (**Date between** '2010-06-04' **and** '2017-03-20') **and** (Landing_Outcome = 'Success (ground pad)') **or** (Landing_Outcome= 'Failure (drone ship)').*

```
%sql select * from SPACEXTABLE where (Date between '2010-06-04' and '2017-03-20') and (Landing_Outcome = 'Success
```

```
' and '2017-03-20') and (Landing_Outcome = 'Success (ground pad)') or (Landing_Outcome = 'Failure (drone ship)');
```

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landin
2015-10-01	09:47:00	F9 v1.1 B1012	CCAFS LC-40	SpaceX CRS-5	2395	LEO (ISS)	NASA (CRS)	Success	Failure (
2015-04-14	20:10:00	F9 v1.1 B1015	CCAFS LC-40	SpaceX CRS-6	1898	LEO (ISS)	NASA (CRS)	Success	Failure (
2015-12-22	01:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Succ
2016-01-17	18:42:00	F9 v1.1 B1017	VAFB SLC-4E	Jason-3	553	LEO	NASA (LSP) NOAA CNES	Success	Failure (


Section 3
Launch Sites
Proximities Analysis

Folium Map of all launch sites’ location on a global map

- In the **following image**, we illustrate the **generated folium map** and **include all launch sites’ location markers** on a global map:

- The ***four Launch Site locations*** are **near by the ocean**, which can be explained due to the fact that after launching a rocket, many pieces are thrown and recuperated on the ocean.

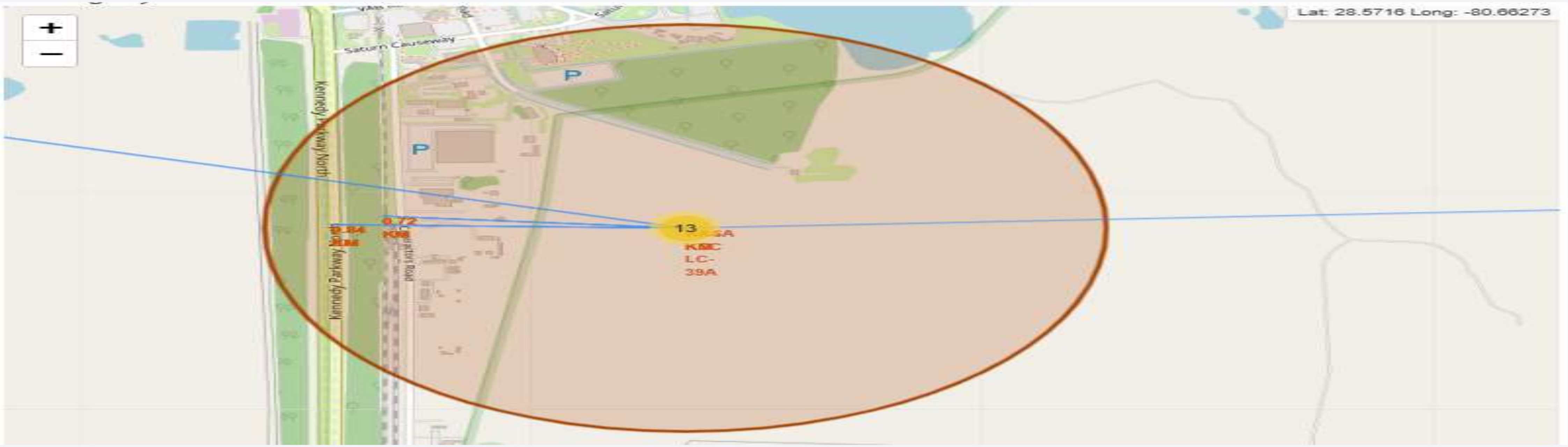
Folium Map of the colored-labeled launch outcomes

- In the **following image**, we illustrate the ***color-labeled launch outcomes on the map:***

- The **color-labeled launch outcomes on the map**, shows on **green the successful launches** and on **red the failed launches,** from CCAFS SLC-40 Launch Site**.**

Folium Map from proximities of KSC LC-39A Launch Site

- In the **following image**, we illustrate the **generated folium map** of **KSC LC-39A** launch site to its proximities such as **railway, highway, coastline**, with **distance calculated and displayed.**



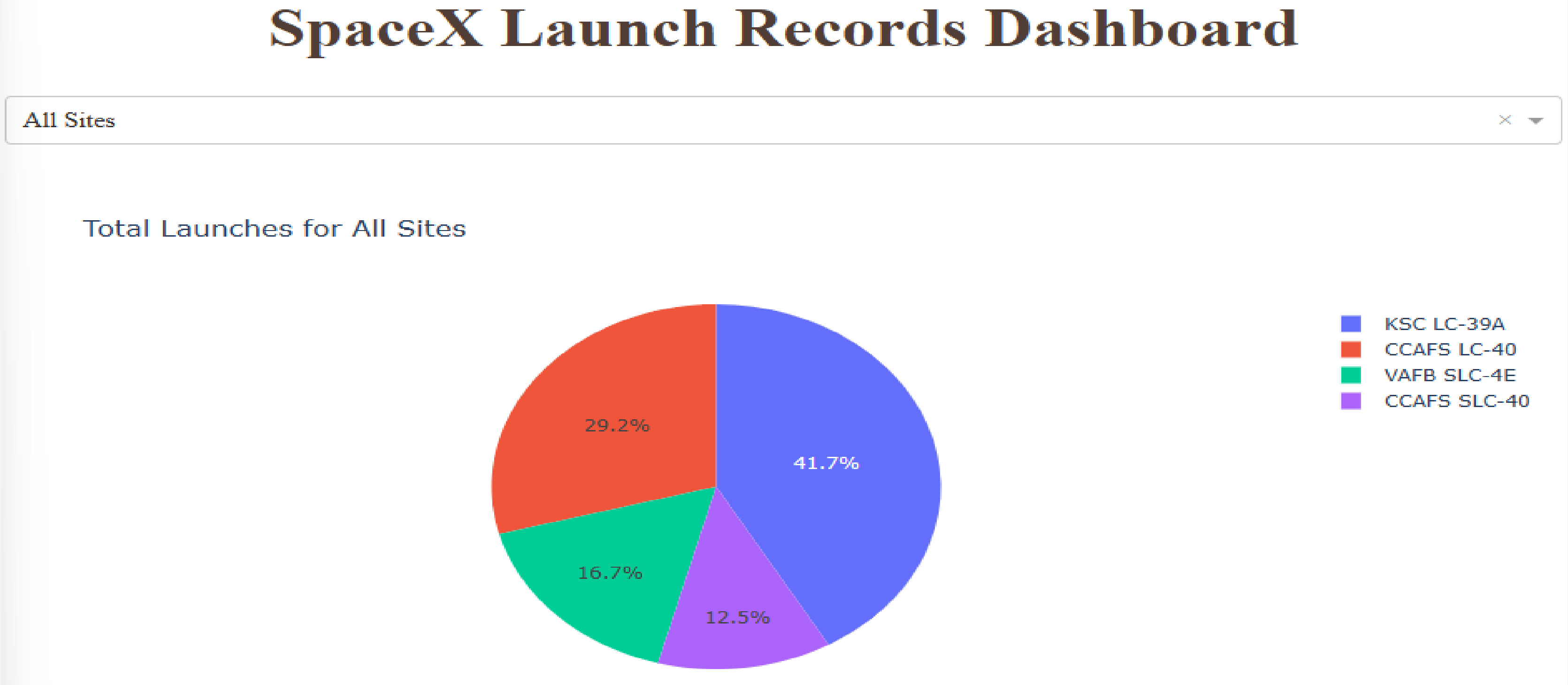
- As we can see **it's important that the Launch Site**, it´s **near railway, highway and the coastline** in order to ***manage the supply chain of the rocket launch***. However **it´s away from different cities** due to ***risks management reasons*** (sound pollution…).

Section 4

Build a Dashboard with Plotly Dash

Interactive Dashboard of Launches from all Launch Sites

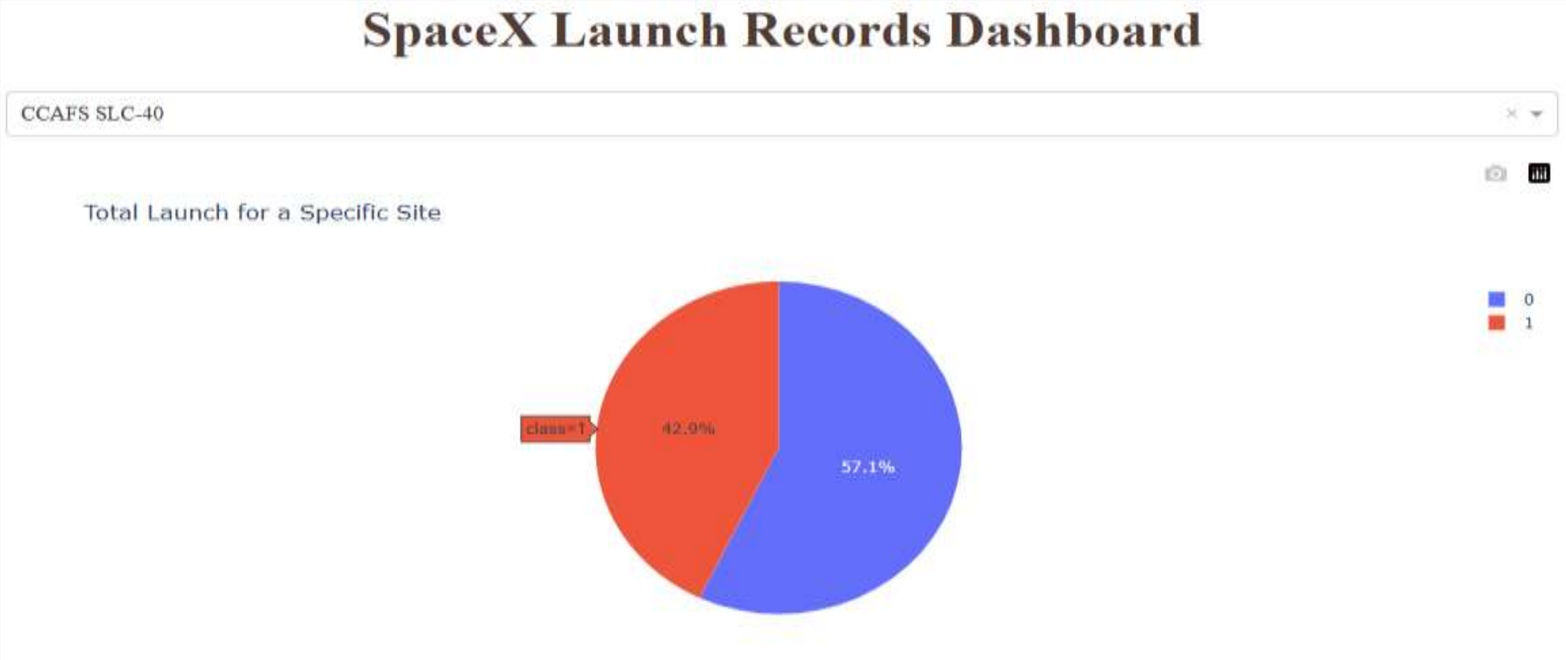
- In the **following piechart**, we can ***see the percentage of launches from all sites***.



- On the **following piechart, we see** that the Launch Site were there are more launches is ***KSC LC-39A,*** then it is followed by: ***CCAFS LC-40, VAFB SLC-4E and CCAFS SLC-40.***

PIECHART FOR THE LAUNCH SITE WITH HIGHEST LAUNCH SUCCESS RATIO

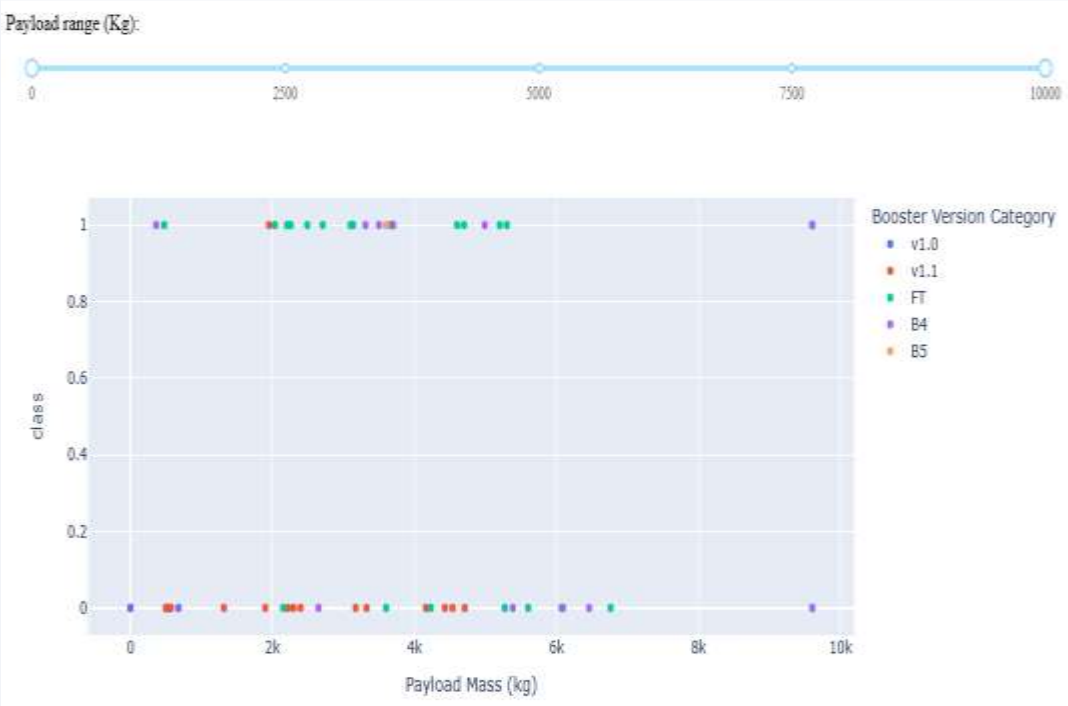
- In the **following piechart,** we see the piechart from **CCAFS SLC-40 launch site** with the **highest launch success ratio.**

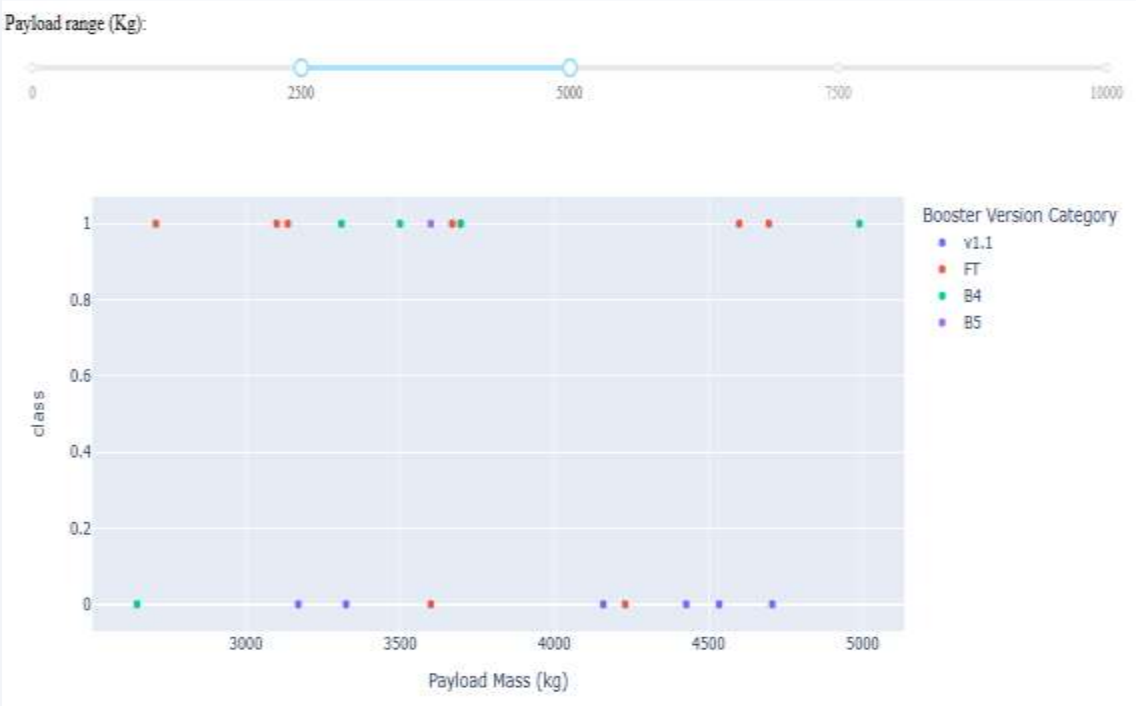


- On this **piechart** we see that from the rocket launched from **CCAFS SLC-40,** we **have *42,9% of success*** and ***57,1% of failure***.

Interactive dashboard of Payload Mass vs Launch Outcome for all sites

- In **the following scatter plot** for **all sites**, we see different payload mass selected in the range slider.





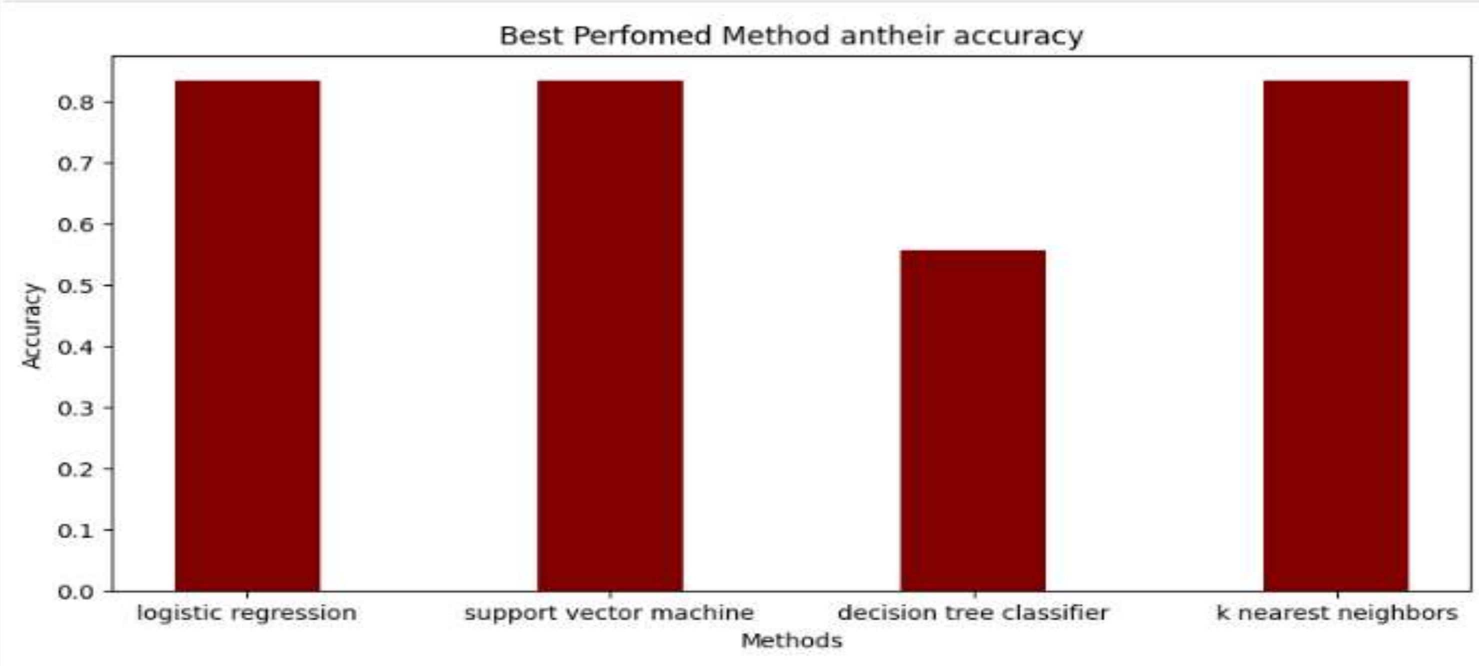
- By changing the ***Payload Mass (Kg), we*** can see that the Booster Version Category v1.0 don´t appear anymore. Also the ***success rate increase among the different Payload Mass (Kg)***.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

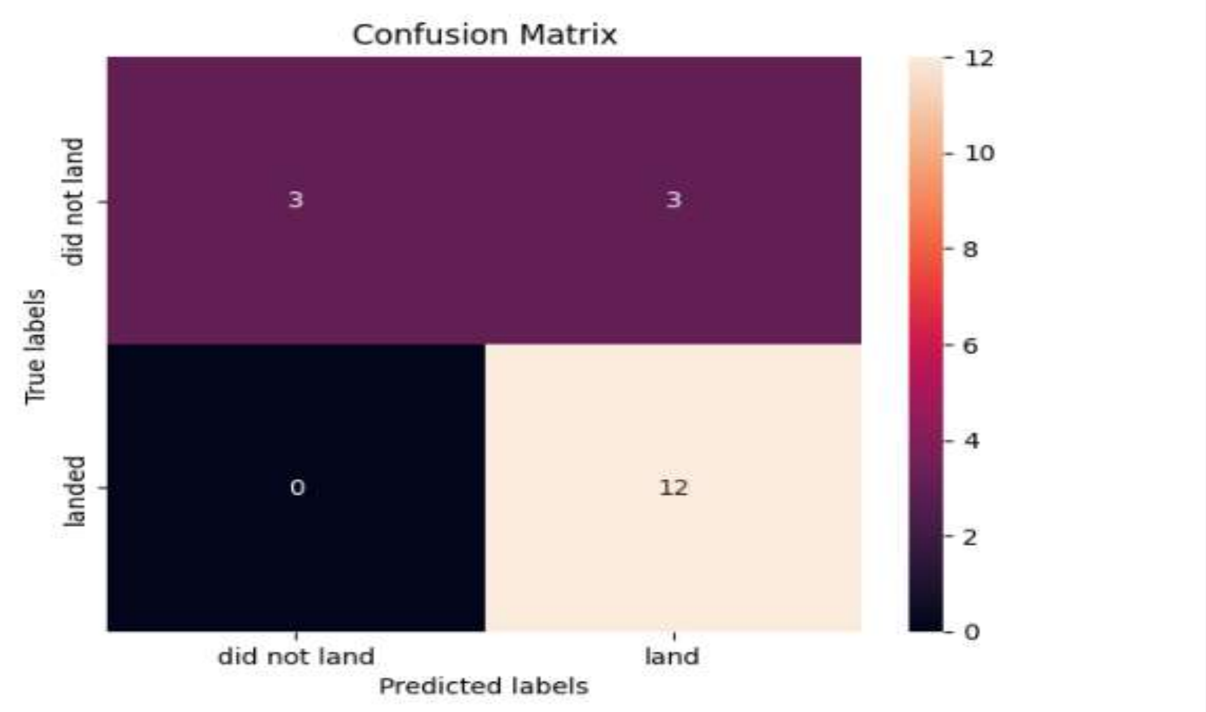
- In the **bar chart,** we can **visualize the model accuracy** for **all built classification models.**



- We have three models with the best accuracy, that are: ***logistics regression, support vector machine and knearest neighbors.***

Confusion Matrix

- The following ***confusion matrix*** represent the results from ***the 4 best performing models***. Here we can see that **from the 18 predicted labels, 15 had the same output from the True labels,** which is different from **appendix 1 that shows** the **confusion matrix from K nearest neighbors model**. Those tests and matrix can help to measure the accuracy of the ***predictive model***.



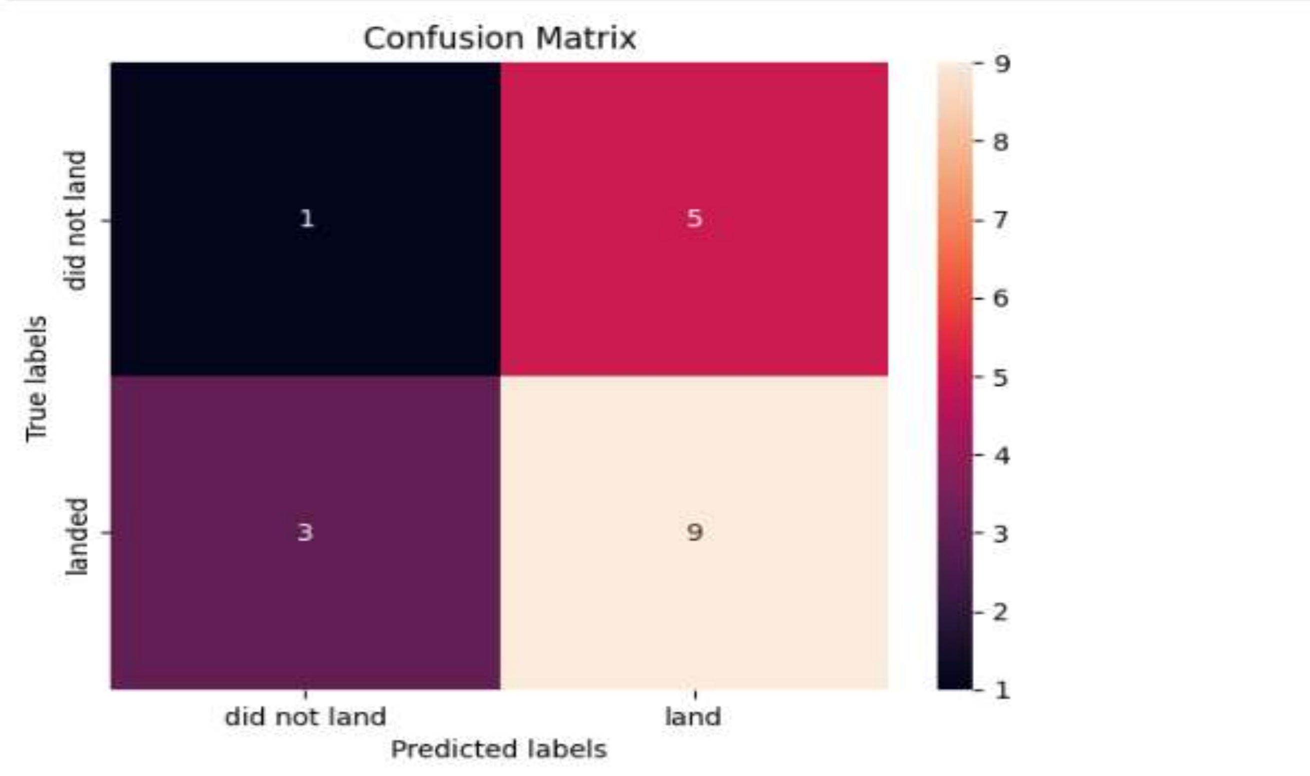
Conclusions

- ***On a first hand***, after going ***through different processes to build our predictive models*** we had found through the dashboard created, ***different variables*** that ***had an impact on the outcome*** of failure or not of a rocket launch from SpaceX. Those are: **Launch Site, Orbit Type, Payload Mass (Kg) and Booster Version Category**. The failure of the rocket launch, has ***a negative direct impact on the reusage of the first stage and on the cost of rocket launches as well***. This because, the **first stage is expensive.**

- ***On a second hand***, regarding the **predictive model**, we had **found that 3 of them could be useful due to their accuracy,** to predict whether a **rocket launch will be successful or not**. The rocket launch **success prediction** will be helpful to **define whether the first stage could be reused or not**, which **impact positively** the **definition of the price of each launch**.

Appendix 1 Confusion Matrix from k nearest neighbors model

- The following ***confusion matrix*** represent the results from the model less accurate, that is k nearest neighbors model.



Thank you!