IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Name>
<Date>

Outline

Executive Summary

The methodologies utilized to analyze the data were as follows:

- Data Collection using Web Scraping and SpaceX API.
- Exploratory Data Analysis with SQL, Data Visualizations using Plotly libraries and Folium maps and other interactive visual analytics.
- Machine Learning Predictions using Logistic Regression, Support VectorMachine, Decision Tree Classifier, and K Nearest Neighbors.

Summary of all methods:

- It was easy and possible to collect all necessary data.
- Exploratory Data Analysis helped to differentiate and identify best features for Machine Learning Predictions.
- GridSearchCVwas utilized to find the best parameters for machine learning models. All four produced similar results with accuracy rate of about 92%%.

Introduction

Project background and context:

- **Overview:**
  - The commercial space industry is booming, making space travel more affordable and accessible.
  - Key players include Virgin Galactic, Rocket Lab, Blue Origin, and the most successful so far, SpaceX.
- **Key Companies:**
  - Virgin Galactic: Provides suborbital spaceflights for tourism and research.
  - Rocket Lab: Specializes in launching small satellites into orbit.
  - Blue Origin: Focuses on developing reusable rockets for suborbital and orbital missions.
  - SpaceX: Known for its achievements in space exploration, including missions to the International Space Station, the Starlink satellite internet constellation, and manned space missions.
- **SpaceX's Competitive Edge:**

  Cost-Effective Launches:

  - SpaceX offers Falcon 9 rocket launches at $62 million per launch, significantly cheaper than other providers, who charge upwards of $165 million.
  - The cost savings primarily come from the ability to reuse the first stage of their rockets

  Innovations and Achievements:

  - Reusability of rockets, reducing the overall cost of space missions.
  - Successful missions to the ISS, development of the Starlink constellation, and manned space missions.

Introduction

Project background and context:

- **Falcon 9 Overview:**

  Two Stages:

  - **First Stage:** Does most of the work during launch and is much larger than the second stage. It is also the stage that SpaceX aims to recover and reuse.
  - **Second Stage:** Helps deliver the payload to its final orbit.

  Reusability:

  - SpaceX has demonstrated the ability to recover the first stage, although it does not always succeed.
  - Reusability depends on factors like payload, orbit, and mission parameters.

Introduction

Problems you want to find answers

SpaceY aims to compete with SpaceX by determining the cost of each launch and predicting the reusability of the first stage.

- **Cost Determination:**
  - **Objective:** Determine the cost of each rocket launch for Space Y.
  - **Importance:** Understanding the cost structure is crucial for Space Y to offer competitive pricing and attract customers.

- **Outcome expected**
  - **To be able to predict the successful landing of the first stage of rockets will help to predict the total coast of launches**
  - **To know the characteristics of a launch site**

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - In this project, SpaceX launch data will be collected REST API and web scraping techniques:
    - **API Endpoint: We will gather data from:
      the endpoint api.spacexdata.com/v4/launches/past.**
    - **Web Scraping Tool: Use the Python BeautifulSoup package to scrape HTML tables from relevant Wiki pages containing Falcon 9 launch records.**

- Perform data wrangling
    - **Data wrangling involves enriching raw data obtained from the SpaceX API by resolving IDs to fetch detailed information about boosters, launchpads, and payloads.**
    - **Data were filtered** to include only Falcon 9 launches and calculate the mean of the payload masses to replace missing values.
    - **Null values in the landing pad column** are preserved for later processing using one-hot encoding.

Methodology

Executive Summary

- Perform exploratory data analysis (EDA) using visualization and SQL
  - In this project, an extensive was conducted Exploratory Data Analysis (EDA) to understand and derive insights from the SpaceX launch data:
    - **Using SQL:** We employed SQL to retrieve specific subsets of data and perform aggregations. This included calculating success rates over time since 2013 and filtering data to focus on Falcon 9 launches. SQL queries enabled us to identify trends and patterns in launch outcomes across different variables.
    - **Visualization:** Visualization played a crucial role in our EDA process. We used various graphical representations such as charts, graphs, and plots to visualize the data insights derived from SQL queries. For example, we created bar charts to compare success rates among different launch sites like CCAFS LC-40, KSC LC-39A, and VAFB SLC 4E. Scatter plots were utilized to analyze the relationship between payload mass and successful landings.

Methodology

Executive Summary

- Perform interactive visual analytics using Folium and Plotly Dash
  - In the Interactive Visual Analytics and Dashboard module, participants leverage Folium and Plotly Dash to create dynamic tools for stakeholder engagement:
    - **Geographic analysis of launch site locations and proximities** using Folium enhances understanding of spatial patterns essential for selecting optimal launch sites.
    - **Plotly Dash** facilitates the development of interactive dashboards featuring components like dropdown lists and sliders, enabling real-time data exploration and visualization.
    - **This approach transforms** data presentation from static graphs to interactive visual analytics, empowering users to discover insights more effectively and present compelling data-driven narratives.

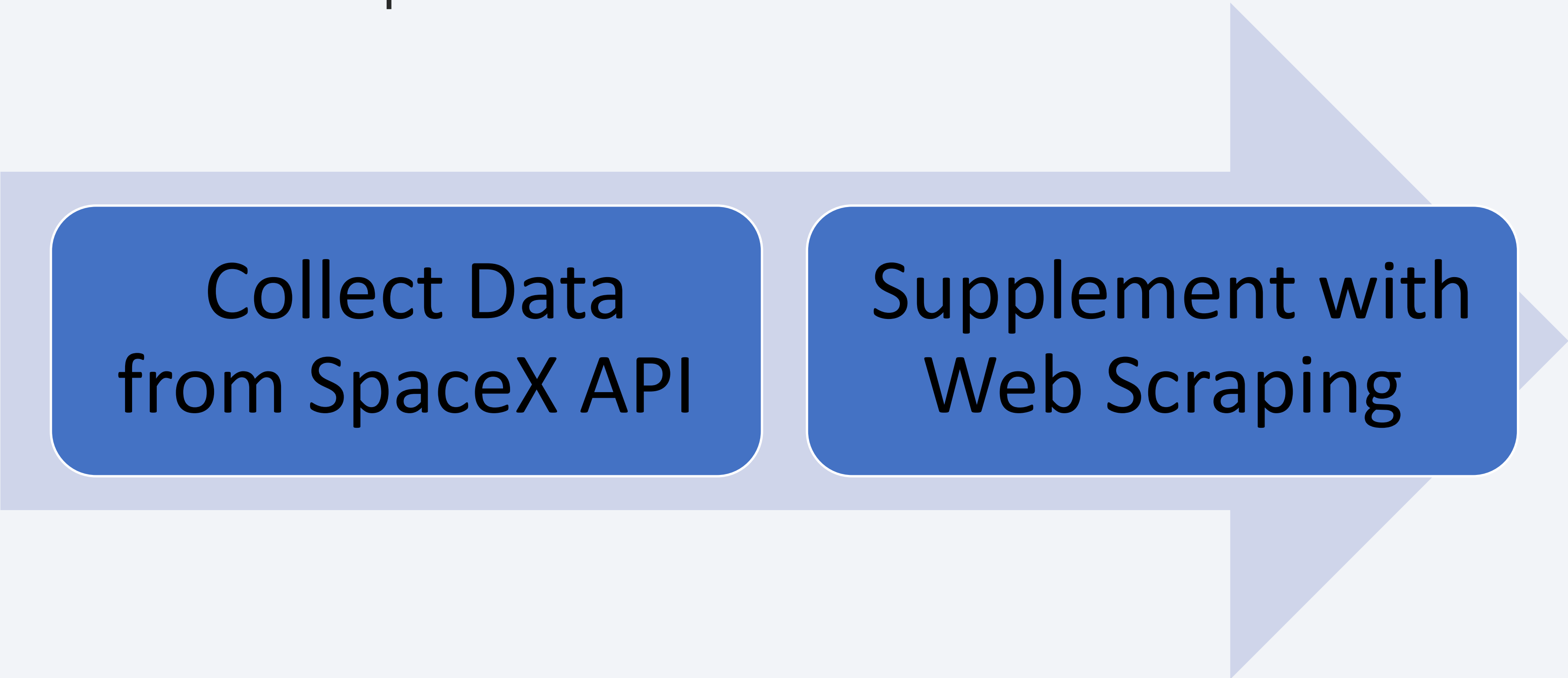
Methodology

Executive Summary

- Perform predictive analysis using classification models
  - Predictive analysis outlines the process of building, tuning, and evaluating classification models for predicting the successful landing of the Falcon 9 first stage:
    - **Preprocessing and Data Standardization:** Initial steps involve preprocessing to ensure data consistency and reliability.
    - **Model Training and Hyperparameter Optimization:** Utilizing Grid Search to fine-tune model hyperparameters for improved predictive accuracy.
    - **Algorithm Testing:** Evaluating multiple classification algorithms, including Logistic Regression, Support Vector Machines, Decision Tree Classifier, and K-nearest neighbors.
    - **Performance Evaluation:** Assessing model performance through metrics such as accuracy and generating a confusion matrix to visualize prediction outcomes.

Data Collection

- As mentioned before data were collected through API Data Collection and Web Scraping for Additional Data
- Data collection macro process



Data Collection – SpaceX API

- Data collection detailed process

Collect Data from SpaceX API

- Perform GET request using requests library
- Receive JSON response
- Normalize JSON data into structured dataframe using json_normalize

🔗 Full Data Collection code is stored in Data Collection

Data Collection – Scraping and result production

- Web scraping and result production process

Supplement with Web Scraping

- Request the Falcon9 Launch Wiki page from its URL
- Extract all column/variable names from the HTML table header using BeautifulSoup for HTML table scraping
- Create a data frame by parsing the launch HTML tables

Full scraping code is stored in Webscraping

Data Wrangling

- Data wrangling process

Data Wrangling

- Import Libraries and Define Auxiliary Functions
- Data Analysis
  - Calculate the number of launches on each site
  - Calculate the number and occurrence of each orbit
  - Calculate the number and occurence of mission outcome of the orbits
  - Create a landing outcome label from Outcome column

🔗 Full Data wrangling code is stored in Data wrangling

EDA with Data Visualization

- EDA was done by using the CSV file saved a the end of data wrangling process. The objective was to perform exploratory Data Analisys and Feature Enginneering.

- The main charts used were scatter plot in order to see the numerical and one or more categorical variables like:
    - Flight Number and PayloadMass: from the 40th flight there is a strong progression in success
    - Launch Site and Flight Number: CCAFS SLC 40 is the main launch site
    - Launch Site and Payload Mass: Heavy rockets were not launched from VAFB SLC 4E
    - Flight Number and Orbit: No link detected between the number of flights and the success in GTO orbit
    - Payload Mass and Orbit: With heavy payloads the successful/positive landing rate are more for Polar, LEO and ISS.
  - A bar chart was also used and has shown that ES-L1, GEO, HEO and SSO are the orbits having the high success rate.
  - Observation that the success rate since 2013 kept increasing was made thanks to a line

Full EDA code is stored in Data visualisation

EDA with SQL

- 10 SQL were used during this phase
  - 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 succesful landing outcome in ground pad was acheived.
  - 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.

Full EDA SQL code is stored in Data SQL

Build an Interactive Map with Folium

- Interactive Map was used to see if the SpaceX success rate depends on the location and proximities and also to discover the key factors to build an optimal launch site.
  - First we marked all site with a circle and marker to identify where the sites were located: Close to the equator to benefit of the speed of rotation of the earth and save fuel but also close to the ocean in case of launch failure
  - Some marker were added to indicate if the launch was a success or a failure. Without surprise LC-39A is the more successful launch site. Same result as in the next exercises.
  - Then mouse position was used to get the coordinate (Lat,Long) for a mouse over on the map. This help to define the coordinates of some key points, like railways, highways, coastline and cities and determine how far the launch sites are:
    - City is necessary to accommodate employees
    - Railways and highways play a key role to transport all the necessary element to create the rocket
    - And as mentioned above this is better that the rocket crash into ocean than cities

Full INTERCACTIVE VIZ code is stored in Location map

In Github the map are not display, link to the map 1 2 3 4 5 6

Build a Dashboard with Plotly Dash

- **We created a pie chart lined to drop-down list by using a callback function:**
  - The purpose was to see which site has the highest number of successes and then to select a specific site and check its detail success rate.
    - KSC LC-39A has the highest number of successes with a breakdown of 76.9% success flights vs 23.1% of failures
- **Then we created a Range Slider to Select Payload and a scatter plot liked to the dropdown list and the range slider:**
  - The purpose was to find if variable payload was correlated to mission outcome. From a dashboard point of view, we wanted to be able to easily select different payload range and see if we can identify some visual patterns.
    - When the payload is between 0 and 2 tons have a low success rate
    - Between 2 and 4 tones the FT booster has the highest number of successes
    - Between 6 and 8 tons there are only failures, above 8 tons the number of flights are very weaks and the results between failure and success are equal
    - To select a FT booster with a payload launched from CCAFS LC-40 may be a good option to succeed

Full Dash code is stored in Dash

Predictive Analysis (Classification)

- In this project, we built a machine learning pipeline to predict the successful landing of Falcon 9's first stage, involving:
  - data preprocessing,
  - train-test splitting,
  - model training,
  - hyperparameter tuning using grid search, and evaluating various classification models
- to identify the best performing model based on accuracy and confusion matrix analysis.

Full Analysis code is stored in Predictive analysis

Building the Machine Learning Pipeline

- **Goal:** Predict whether the first stage of Falcon 9 successfully lands.

Data Preprocessing

- **Standardization:**
  - Normalize the features to have a similar distribution (mean = 0, standard deviation = 1).

Data Splitting

- **Train_test_split:**
  - Split the data into training and testing sets.

Model Training

- Models to Test:
  - Logistic Regression
  - Support Vector Machines (SVM)
  - Decision Tree Classifier
  - K-Nearest Neighbors (KNN)

Hyperparameter Tuning

- Perform grid search to find the best hyperparameters for each algorithm.

Model Evaluation

- Determine the model with the best accuracy using the test data.

Results

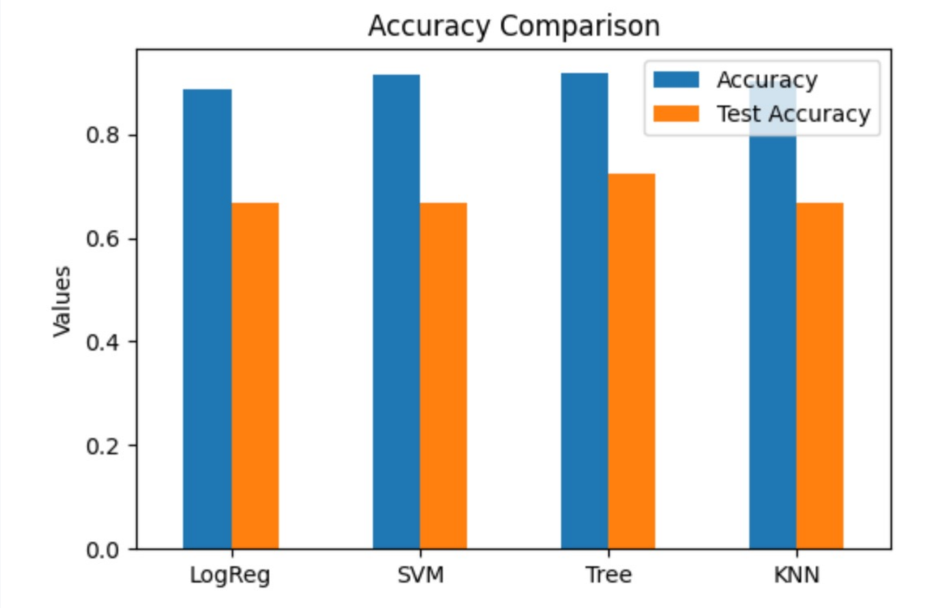
- SpaceX uses four different launch sites: one on the East Coast and three on the West Coast
- The average payload of the F9 v1.1 booster is 2,928 kg.T
- The first successful landing occurred in 2015, five years after the initial launch.
- Many Falcon 9 booster versions successfully landed on drone ships with payloads above the average.
- Almost 100% of mission outcomes were successful.
- Two booster versions failed to land on drone ships in 2015: F9 v1.1 B1012 and F9 v1.1 B1015.
- The number of successful landings improved over the years.

Results

- Launch sites criteria
  - Launch sites are locating close to:
    - The Equator:
      - For launches intended for geostationary orbits, being closer to the equator can provide a velocity boost due to the Earth's rotation. While not all coastal sites are near the equator, many key sites (like those in Florida) are situated to take advantage of this benefit.
    - The sea for several important reasons:
      - Safety: Launching rockets over the ocean minimizes the risk to human life and property. In case of an accident during launch or an early-stage failure, the debris is more likely to fall into the sea, reducing the potential for damage and casualties on land
      - Flight Path Clearance: Rockets need to follow specific trajectories to reach their intended orbits. Launching over the ocean provides a clear, uninhabited flight path, which reduces the risk of the rocket flying overpopulated areas during its ascent.
    - The highway and railway:
      - Logistics and Transportation: Proximity to railways and highways facilitates the efficient transport of materials, heavy equipment, supplies, and personnel needed for launch operations.
      - Accessibility and Safety: These transportation infrastructures provide quick access for workers, suppliers, and emergency responders, ensuring smooth operations and enhanced safety.
    - A little further from the cities
      - Safety and Environmental Concerns: Maintaining a distance of at least 20 km from cities ensures public safety by minimizing the risk of injury or damage from potential launch failures, explosions, and exposure to hazardous materials.
      - Noise and Vibrations: Keeping cities far from launch sites reduces the impact of extreme noise and vibrations, preventing disruptions and potential harm to urban areas.

Results

- Predictive analysis:
  - Among the four models tested, logistic regression, SVM, decision trees and KNN, the decision tree gives the best results with an accuracy rate of 92%



]:

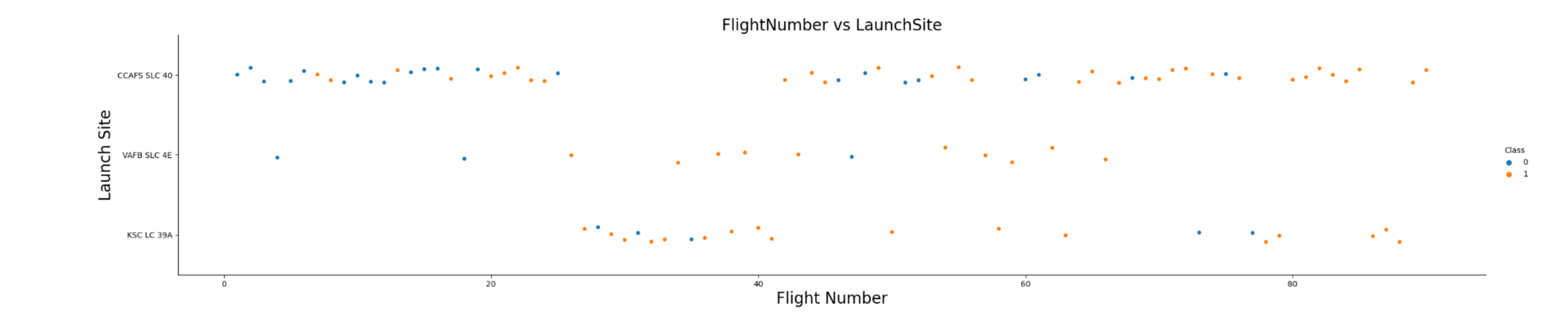
	Accuracy	Test Accuracy
LogReg	0.887500	0.666667
SVM	0.916071	0.666667
Tree	0.917857	0.722222
KNN	0.901786	0.666667

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

- Scatter plot of Flight Number vs. Launch Site

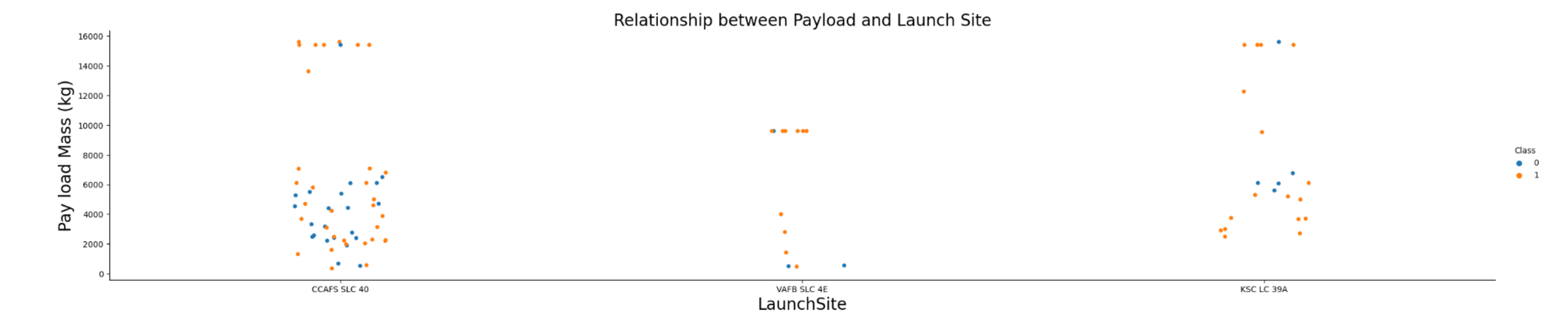


Blue unsuccessful launch; Orange successful launch.

- CCAFS SLC 40 is the most used launch site, SpaceX rented it from 2007 to US Air Force.
- The cessation of use of this site between flights 30 and 40/45 can be justified by the explosion in November 2016 of the Falcon 9-29 launcher. CCAFS SLC 40 became operational in December 2017 after a year of works. Space X needed $50 million to rehabilitate it
- From the 20th flight the successes follow one another, since flight 80 there has been no failure

Payload vs. Launch Site

- Scatter plot of Payload vs. Launch Site



Blue unsuccessful launch; Orange successful launch.

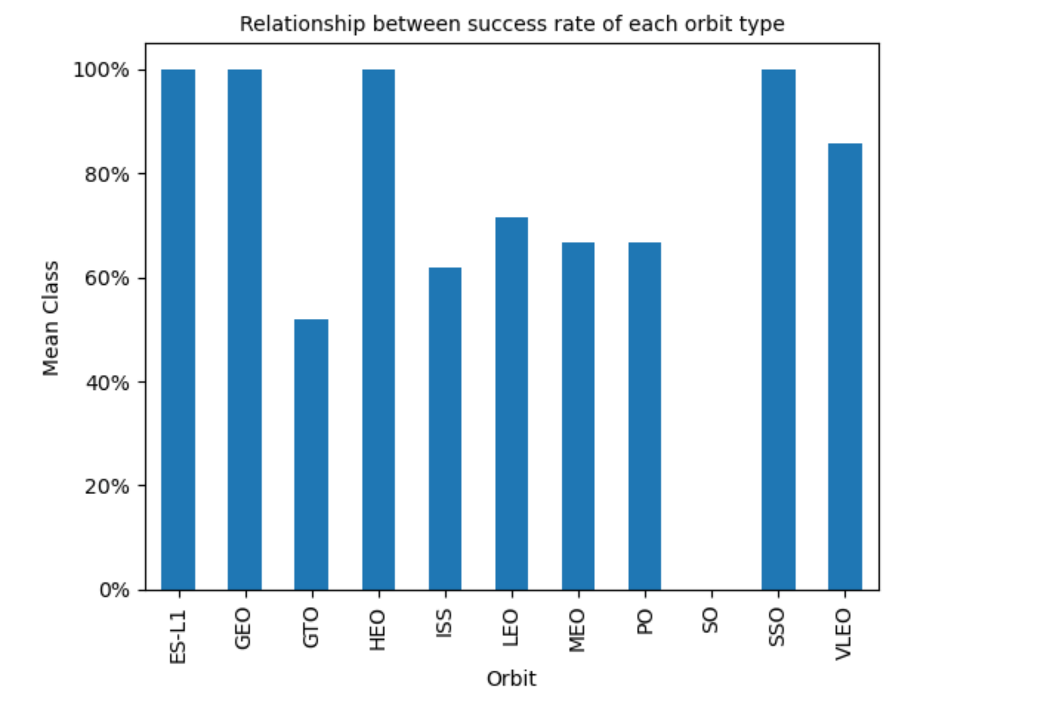
- The most common range for payload mass is between 0 – 6500 kg
- Only CCFAS SLC 40 and KSC LC 39A can managed a payloads over 12000kg

Success Rate vs. Orbit Type

```
Entrée [6]: # Apply value_co
            df['Orbit'].valu

   Out[6]: Orbit
           GTO      27
           ISS      21
           VLEO     14
           PO        9
           LEO       7
           SSO       5
           MEO       3
           ES-L1     1
           HEO       1
           SO        1
           GEO       1
```

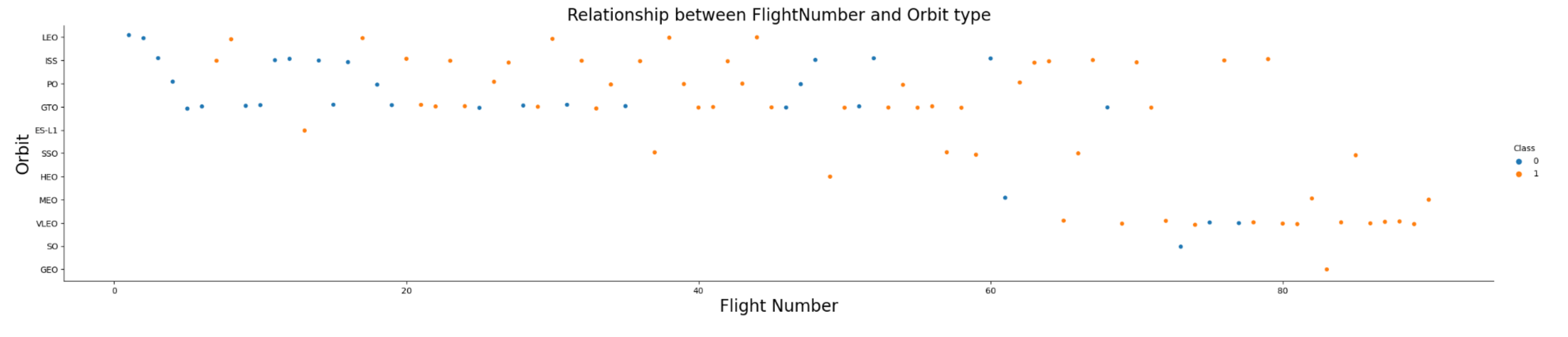
- Bar chart for the success rate of each orbit type



- Full success for the orbits:
  - ES-L1
  - GEO
  - HEO
  - SSO
- Mixed success for the ISS and GTO orbits with respective rates of 50/55% and 60/65%
- These results must be put into perspective because the most successful orbits are those for which there has been only one launch except for SSO which has 5 launches.
- Conversely, ISS and GTO are the orbits which have the highest number of launches
- The success for VELO is interesting, around 80% success for 14 launches

Flight Number vs. Orbit Type

- Scatter point of Flight number vs. Orbit type



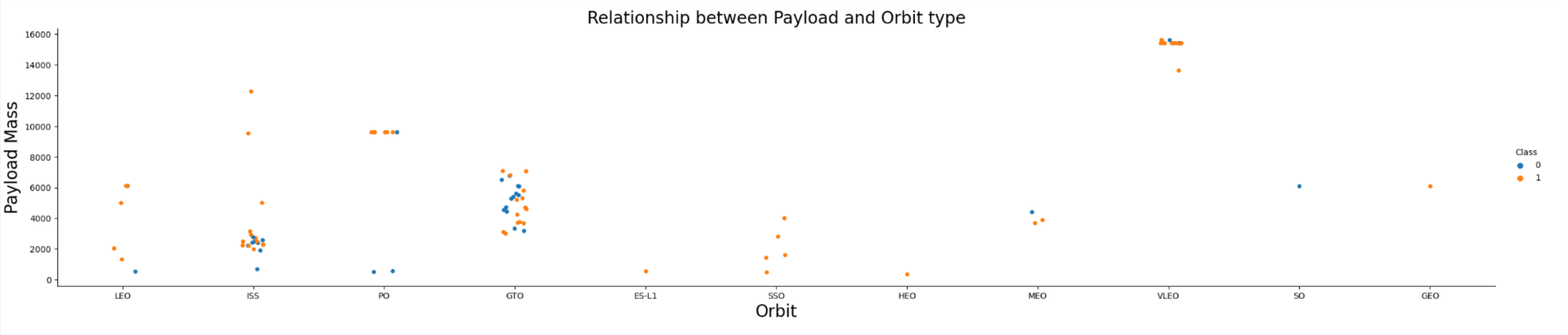

```
Entrée [6]: # Apply value_co
            df['Orbit'].valu

Out[6]: Orbit
        GTO      27
        ISS      21
        VLEO     14
        PO        9
        LEO       7
        SSO       5
        MEO       3
        ES-L1     1
        HEO       1
        SO        1
        GEO       1
```

- Launches to the ISS are constant
- It is interesting to note that since the flight #60 the orbit mainly targeted is VLEO, this must be seen considering the emergence of New Space (emergence of a space industry of private initiative which seeks to conquer space)
- This interest in the VÉLO orbit results in a significant drop towards GTO

Payload vs. Orbit Type

- Show a scatter point of payload vs. orbit type

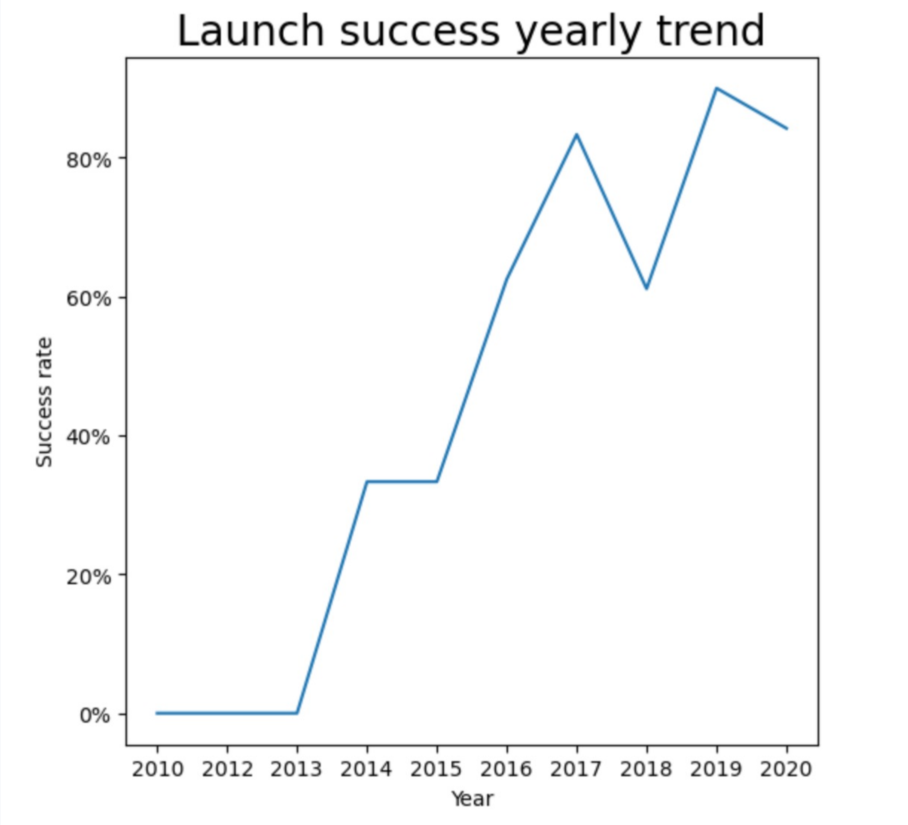


- The largest payloads are those destined for the closest orbits (VLEO and its 16,000 kg)
- Depending on the orbits, the payloads are in a contained range:
  - GTO - 3000 to 7000 Kgs
  - SSO - 0 to 4000kgs
- Si la charge utile a destination de l'ISS se situe entre 2000 et 4000 kg, cette charge peut augmenter jusqu'a 12 tonnes en fonction des besoins

Launch Success Yearly Trend

- Before 2013 the number of flights per year is really low
- Between 2013 and 2014 the success rate increases from 0 to 35%
- From 2014 to 2015 the success rate remains stable
- Since 2015 the success rate has increased considerably, peaking at 90% in 2019.

- Show the screenshot of the scatter plot with explanations



All Launch Site Names

- There are four unique launch sites:
  - **CCAFS LC-40**(Old name for CCAFS SLC-40): Cape Canaveral Air Force Station Launch Complex 40
  - **CCAFS SLC-40** (previously named LC-40): Cape Canaveral Air Force Station Space Launch Complex 40. The site is used by SpaceX since 2007
  - **KSC LC-39A** (Kennedy Space Center Launch Complex 39): The site is rented by SpaceX since 2013
  - **VAFB SLC-4E**: Vandenberg Space Launch Complex 4. This site is used since 2011 by SpaceX, the first launch occurred in Sept.2013

```
%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'

- Find 5 records where launch sites begin with `CCA`
    - We used the clause where like to find sites starting by CCA and limited the result to 5

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

- The result is a table of the 5 sites beginning by CCA.

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

- Calculate the total payload carried by boosters from NASA CRS
  - CRS stands for: Commercial Resupply Services (CRS) are a series of flights awarded by NASA for the delivery of cargo and supplies to the International Space Station (ISS) on commercially operated spacecraft (***Wikipedia source***)

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

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

sum(PAYLOAD_MASS__KG_)
45596

  - The answer is 45,596 Kg meaning that the all the boosters launched by NASA (CSR) carried a global payload mass close to 46 tones

Average Payload Mass by F9 v1.1

- The request calculate the average payload mass carried by booster version F9 v1.1

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

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

avg(PAYLOAD_MASS__KG_)
2534.6666666666665

  - The payload mass for Falcon 9 V 1.1 is in a range of 500 – 4700 Kg which explains this low average

First Successful Ground Landing Date

- The first successful landing on a ground platform took place in 2015. The launch site was CCAFS LC-40. It is from this year 2015 that the successes will follow one another.

```
%sql select min(Date) from SPACEXTABLE where Landing_Outcome like '%Success%'
```

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

min(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- The names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000 are:
  - F9 FT B1021.2
  - F9 FT B1031.2
  - F9 FT B1022
  - F9 FT B1026

```
%sql select distinct Booster_Version from SPACEXTABLE where Landing_Outcome\
== 'Success (drone ship)' and PAYLOAD_MASS__KG_ between 4000 and 6000
```

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

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

Total Number of Successful and Failure Mission Outcomes

- SpaceX has a success rate of 98:
  - Out of a total of 100 flights only 1 experienced a failure
  - The uncertain status corresponds to a flight on January 8, 2018, carried out on behalf of the US American Government. Unconfirmed reports suggest that the Zuma satellite may have been lost, but nothing is confirmed. Some people suggest Zuma is in orbit and operating secretly.

```
%sql select Mission_Outcome, count(*) as total_count from\
SPACEXTABLE GROUP BY Mission_Outcome
```

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

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

Boosters Carried Maximum Payload

- The names of the booster which have carried the maximum payload mass, 15,600 kg are:

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

  - B5 is the latest generation of booster launched in 2018. Moreover, the results of the request correspond to flights carried out from 2019

```
%sql select distinct Booster_Version,PAYLOAD_MASS__KG_,Date from SPACEXTABLE\
where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_) from SPACEXTABLE)

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

2015 Launch Records

- The failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015
  - We created a dedicated table (MONTH) to get the month name based on the month number

```
df_m=pd.read_csv('month.csv',delimiter=';',dtype='str')

df_m.to_sql("MONTH",con, if_exists='replace', index=False,method="multi")
```

```
12
```

```
%sql select month,substr(Date,0,5),Landing_Outcome,Booster_Version,Launch_Site from MONTH as m, SPACEXTABLE as s\
where n_month =(select substr(s.Date, 6,2)) and substr(Date,0,5)='2015'\
and Landing_Outcome='Failure (drone ship)'
```

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

month	substr(Date,0,5)	Landing_Outcome	Booster_Version	Launch_Site
January	2015	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
April	2015	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

- List of all landings between 2010-06-04 and 2017-03-20 inclusively.
  - 8 successful landing in total
  - 7 unsuccessful landing in total
  - An important number of no attempt that also need to be take into consideration

```
%sql select Landing_Outcome, count(*) as count_outcomes from SPACEXTABLE\
where Date between '2010-06-04' and '2017-03-20' group by Landing_Outcome order by count_outcomes DESC
```

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

Landing_Outcome	count_outcomes
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

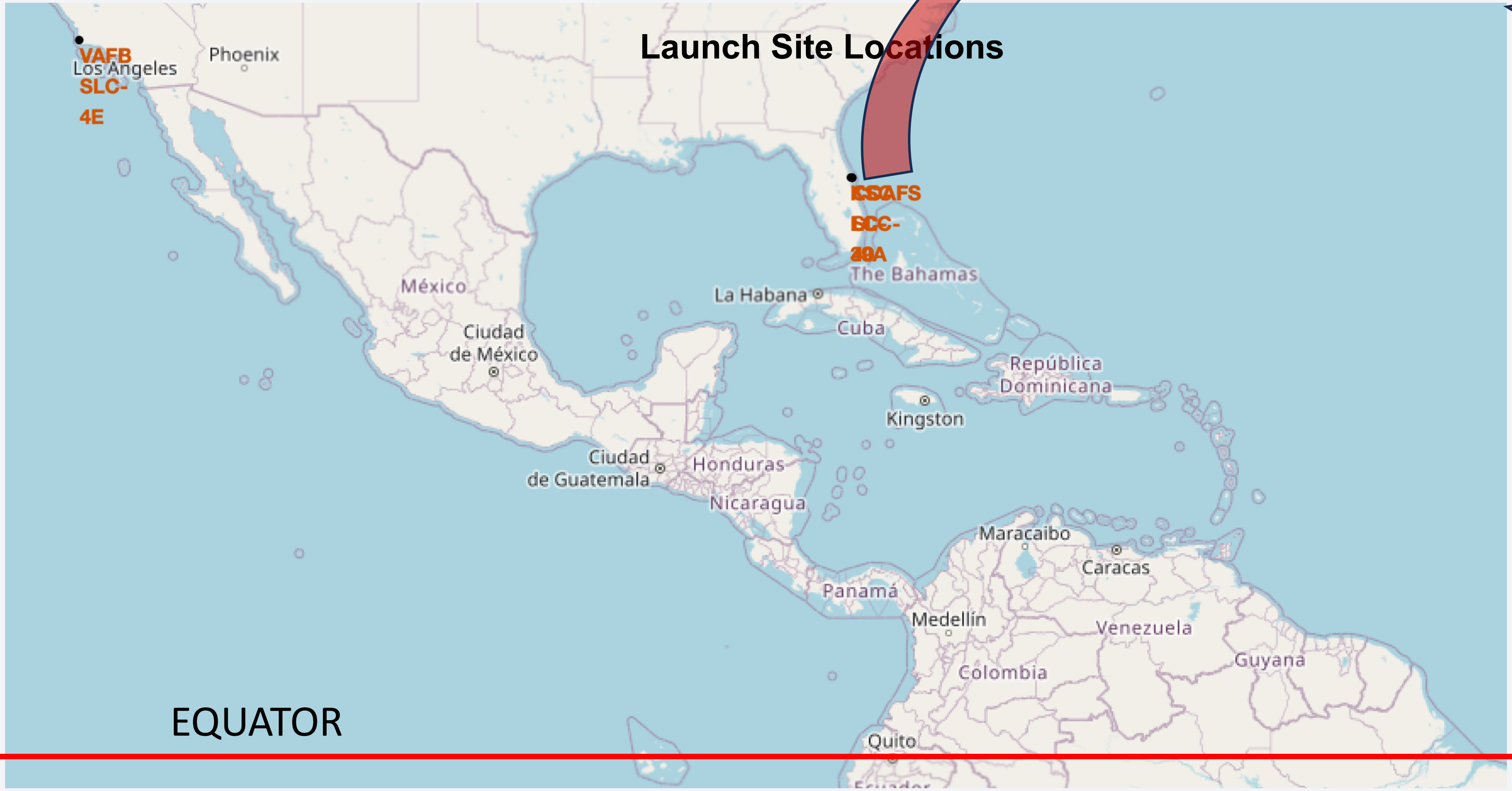
Launch Site Locations

- Replace <Folium map screenshot 1> title with an appropriate title

- Explore the generated folium map and make a proper screenshot to include all launch sites' location markers on a global map

- Explain the important elements and findings on the screenshot

Launch Site Locations

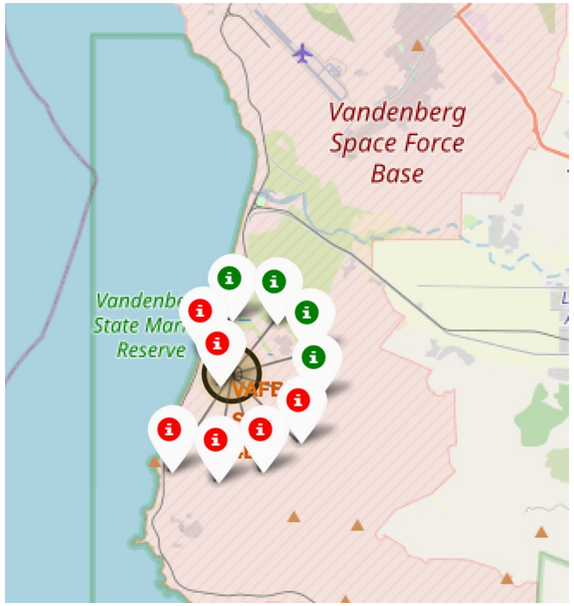


- 1 Launch site on the West coast VAFB SLC-4E
- 3 Launch sites on the East coast CCAFS LC-40, CCAFS SLC-40, KSC LC-39A
- LC-40 and SLC-40 share the same coordinates

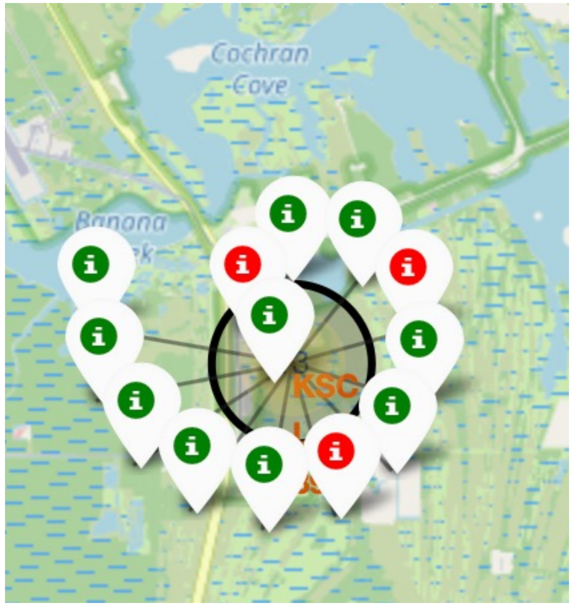
- Launch sites are relatively close to the equator (around 3000 km) because the Earth's rotation speed is faster at the equator (1675 km/h), providing a "catapult" effect that saves tons of fuel.
- Additionally, being near oceans ensures that, in case of an accident, rockets fall into the ocean, minimizing damage and risk to populated areas.

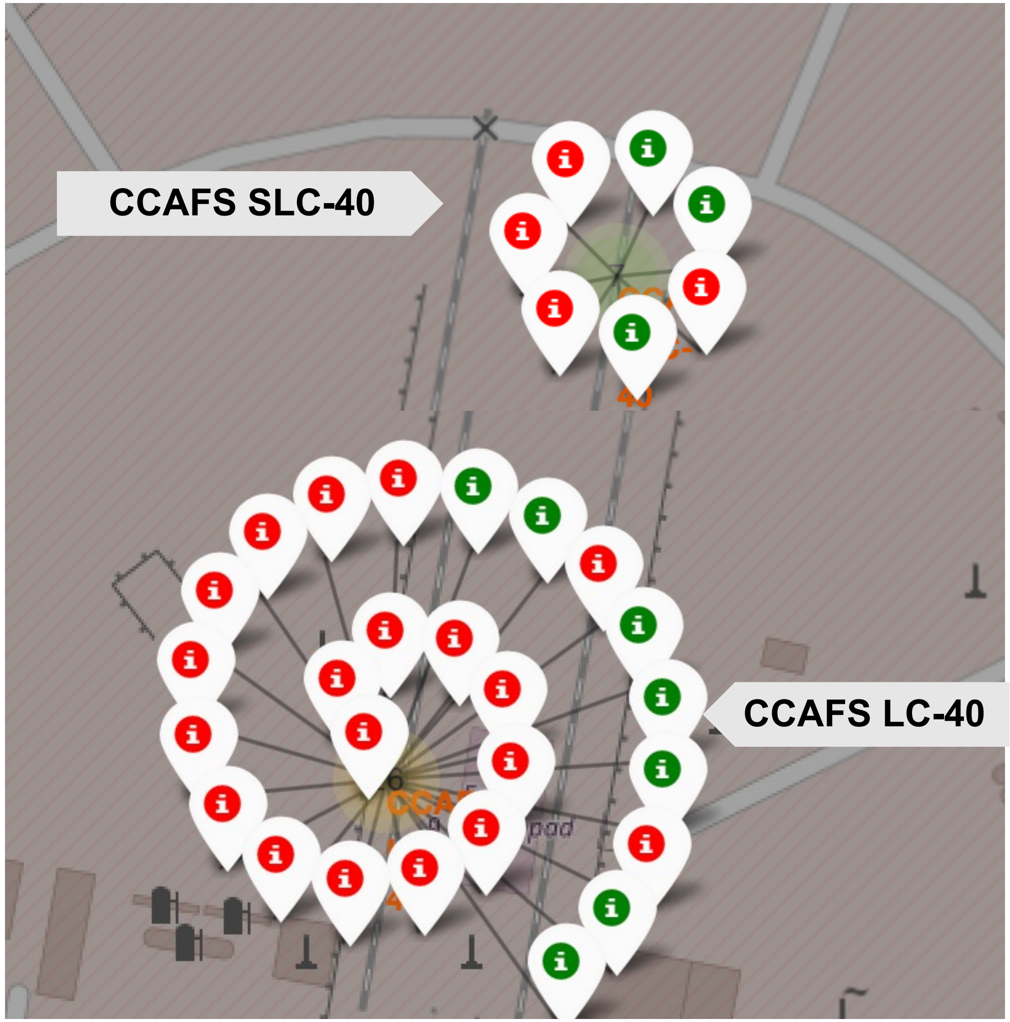
Unsuccessessfull/Successfull Markers By Site

VAFB SLC-4E



KSC LC-39A





Successful landing (green icon) and failed landing (red icon).

- KSC LC-39A is the most successful site
- Despite its numerous flights CCAFS LC-40 accumulates very little success

Key Location Proximities

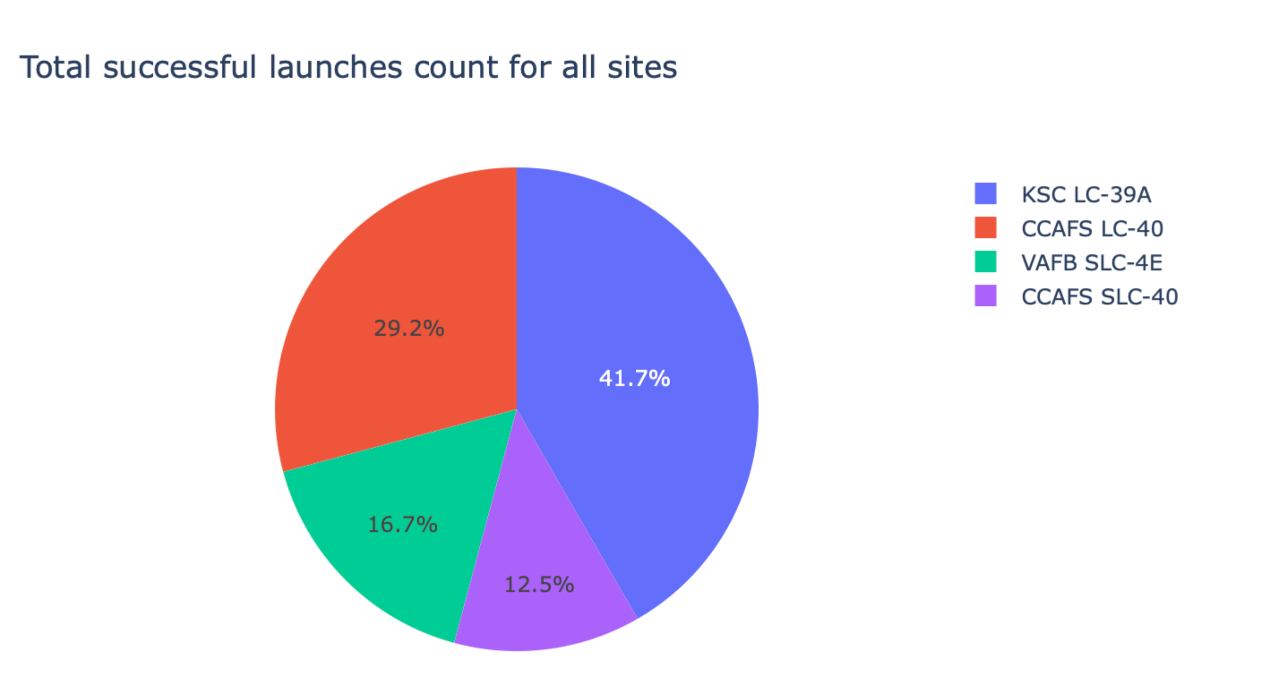
- With KSC LC-39A as an example, it can be noted that highway and railways are very closed. This makes it possible to transport the supplies and materials needed to build and launch the rockets.
- As already indicated, the seaside is really close, in the event of an accident the rocket falls into the ocean and not on the population.
- Finally, the fact that the town is a little further away also makes it possible to protect the inhabitants but also to accommodate the staff.

site	points	Dist
KSC LC-39A	Seaside	7.383858
KSC LC-39A	Railway	6.032779
KSC LC-39A	Highway	7.156426
KSC LC-39A	City	20.994232

Section 4

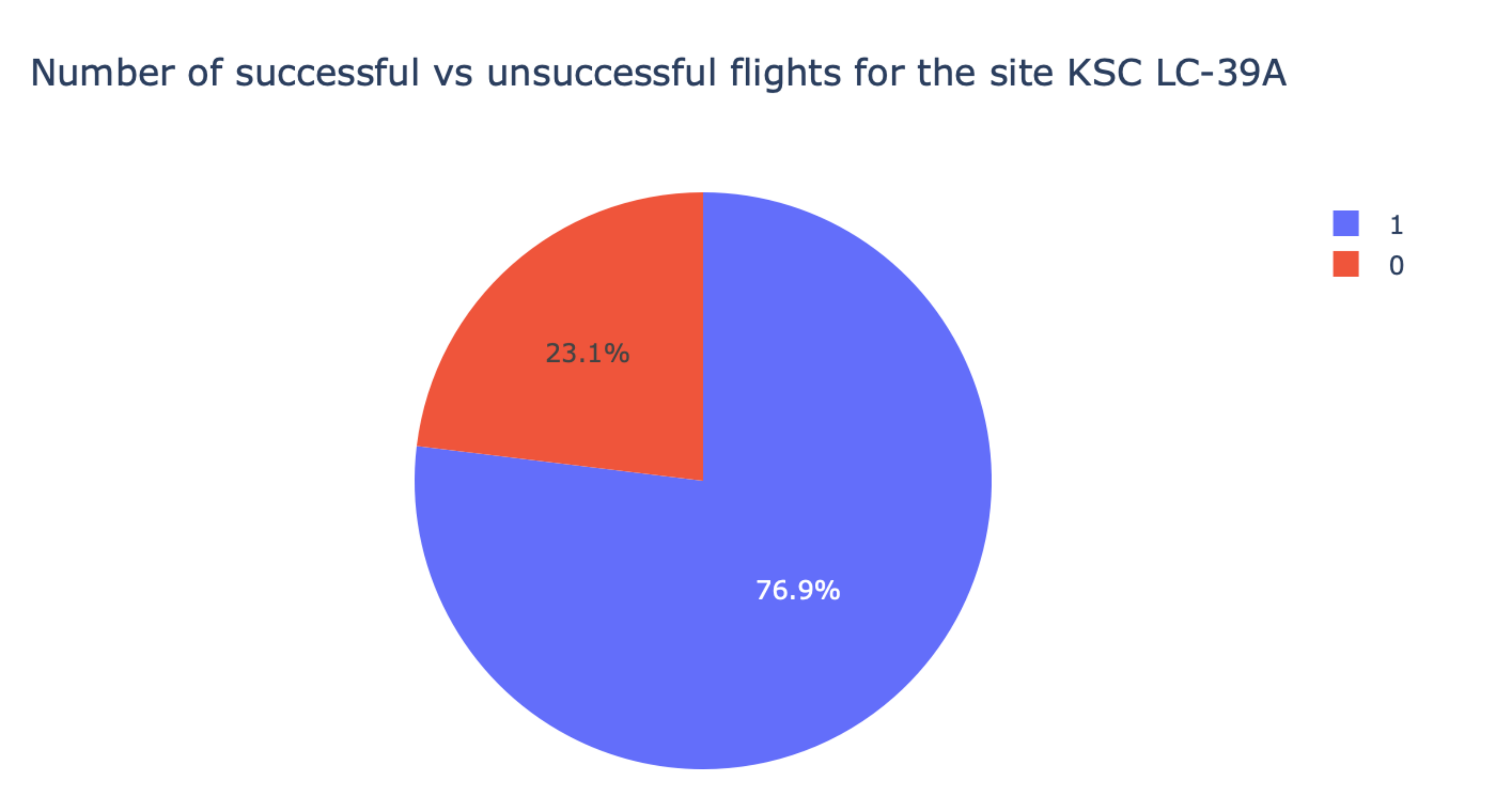
Build a Dashboard with Plotly Dash

Total successful launches by launch site



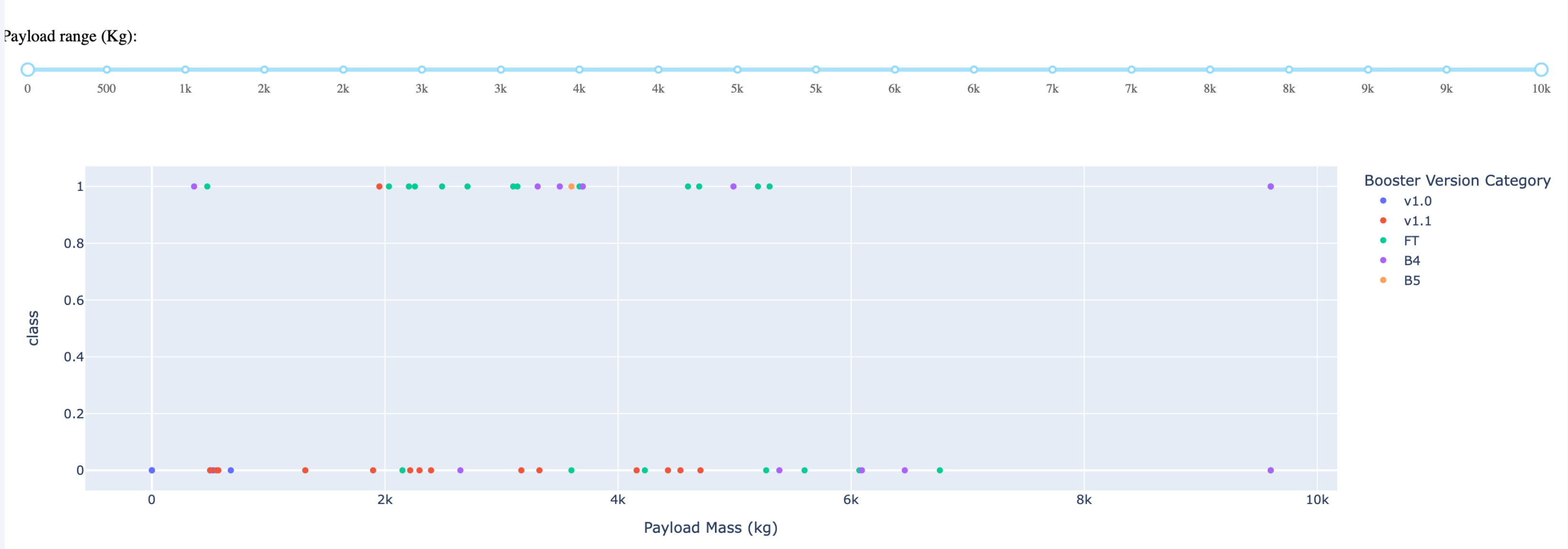
- As CCAFS LC-40 and CCAFS SLC-40 are the same site (CCAFS LC is the old name of CCAFS SLC). The conclusion is that CCAFS CL and SCL have the same number of success than KSC LC 39

Launch site with highest launch success ratio

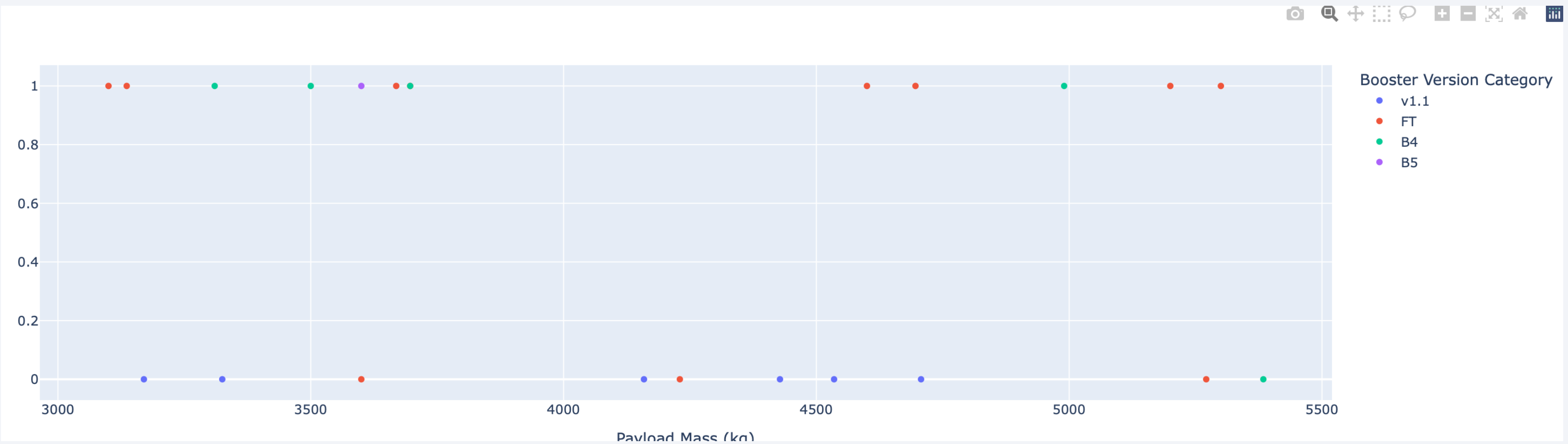


- KSC LC39A has the highest launch success ratio: 76,9%
- Out of 13 flights carried out only 3 failed

Payload vs. Launch Outcome scatter



Class indicates 1 for successful landing and 0 for unsuccessful landing

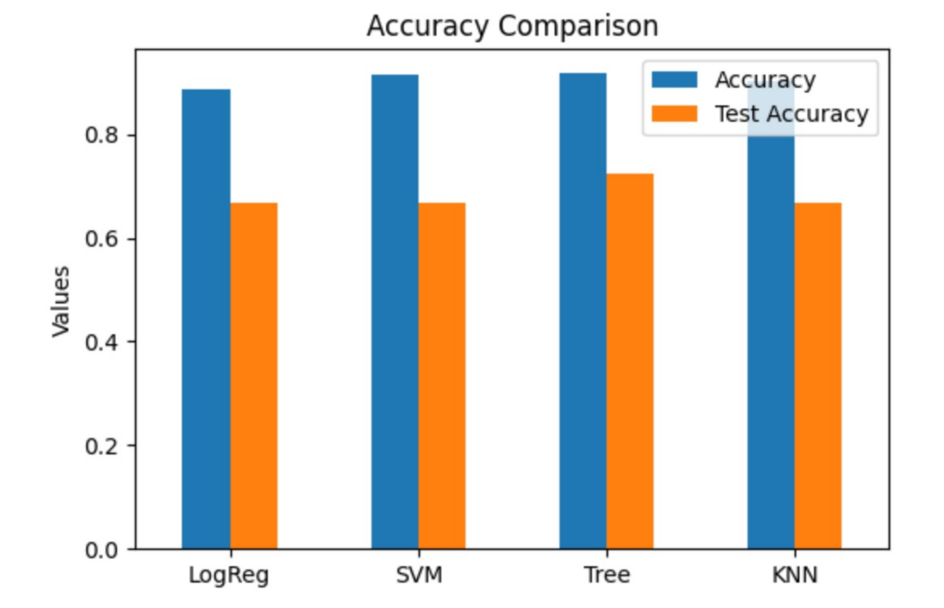


- The combination of payloads less than 6000 kg and FT booster achieves a high success rate.
- Flights with payloads greater than 6000 kg are less important and less successful (1 success for 5 failures)

Section 5

Predictive Analysis (Classification)

Classification Accuracy



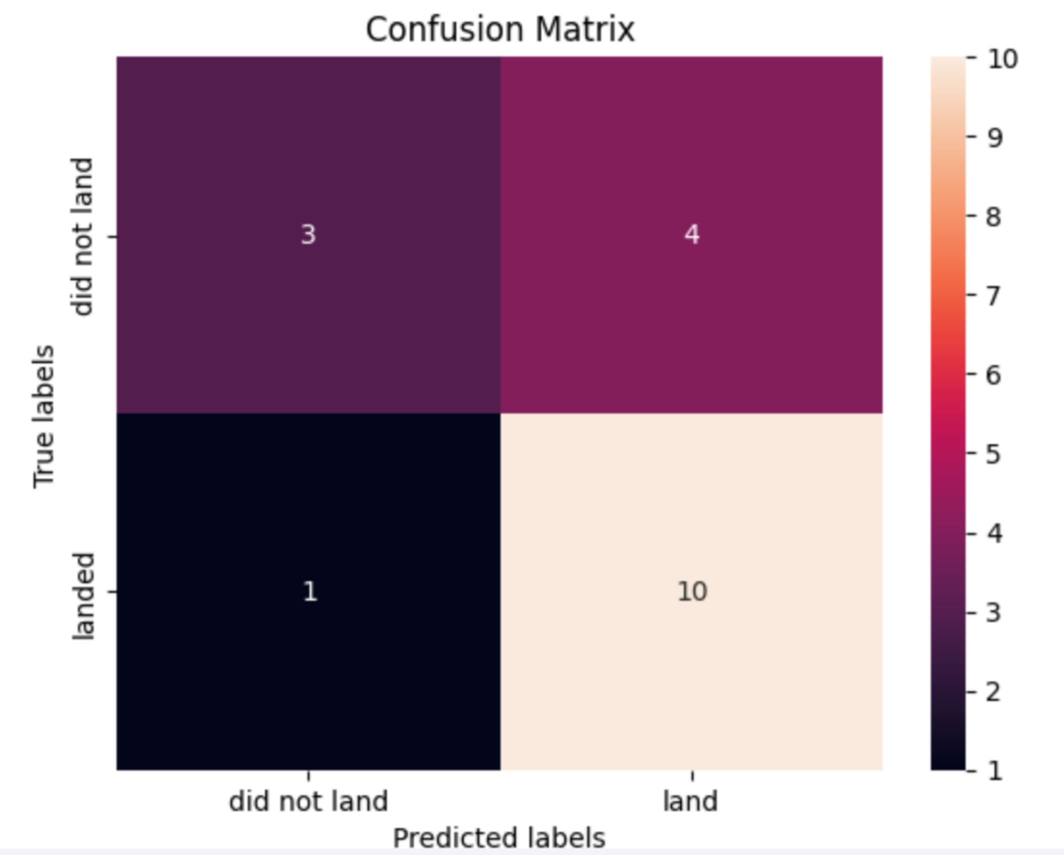
]:

	Accuracy	Test Accuracy
LogReg	0.887500	0.666667
SVM	0.916071	0.666667
Tree	0.917857	0.722222
KNN	0.901786	0.666667

- Among the four models tested, logistic regression, SVM, decision trees and KNN, the decision tree gives the best results with an accuracy rate of 92%
- The results remain quite close, this may be due to the fairly small sample size 18

Confusion Matrix

```
yhat = tree_cv.predict(X_test)
plot_confusion_matrix(Y_test,yhat)
```



- The models predicted 10 successful landings when the true label was successful landing.
- The models predicted 1 unsuccessful landings when the true label was landed.
- The models predicted 3 unsuccessful landings when the true label was unsuccessful landings
- The model predicted 4 successful landings when true label was unsuccessful landing.

Conclusions

- Based on the several analysis we performed we can conclude:
    - The main characteristics that a launch site must meet are:
        - Being located near the equator
        - Near the ocean coast
        - The site must also be close to major road and rail routes.
        - It must also be close to a city
        - KSC LC 39-A is the best launch site
- To target a Very Low Earth Orbit looks like a good solution as it is the destination with the most success and for which the payload is the largest
- Launches above payloads 7000kg are more successful.
- The decision tree is a good model to predict successful landing

Appendix

- Github repository: [ici]

Thank you!