# DATA SCIENCE CAPSTONE PROJECT

DOROTHÉE-HENRI SAUBATTE

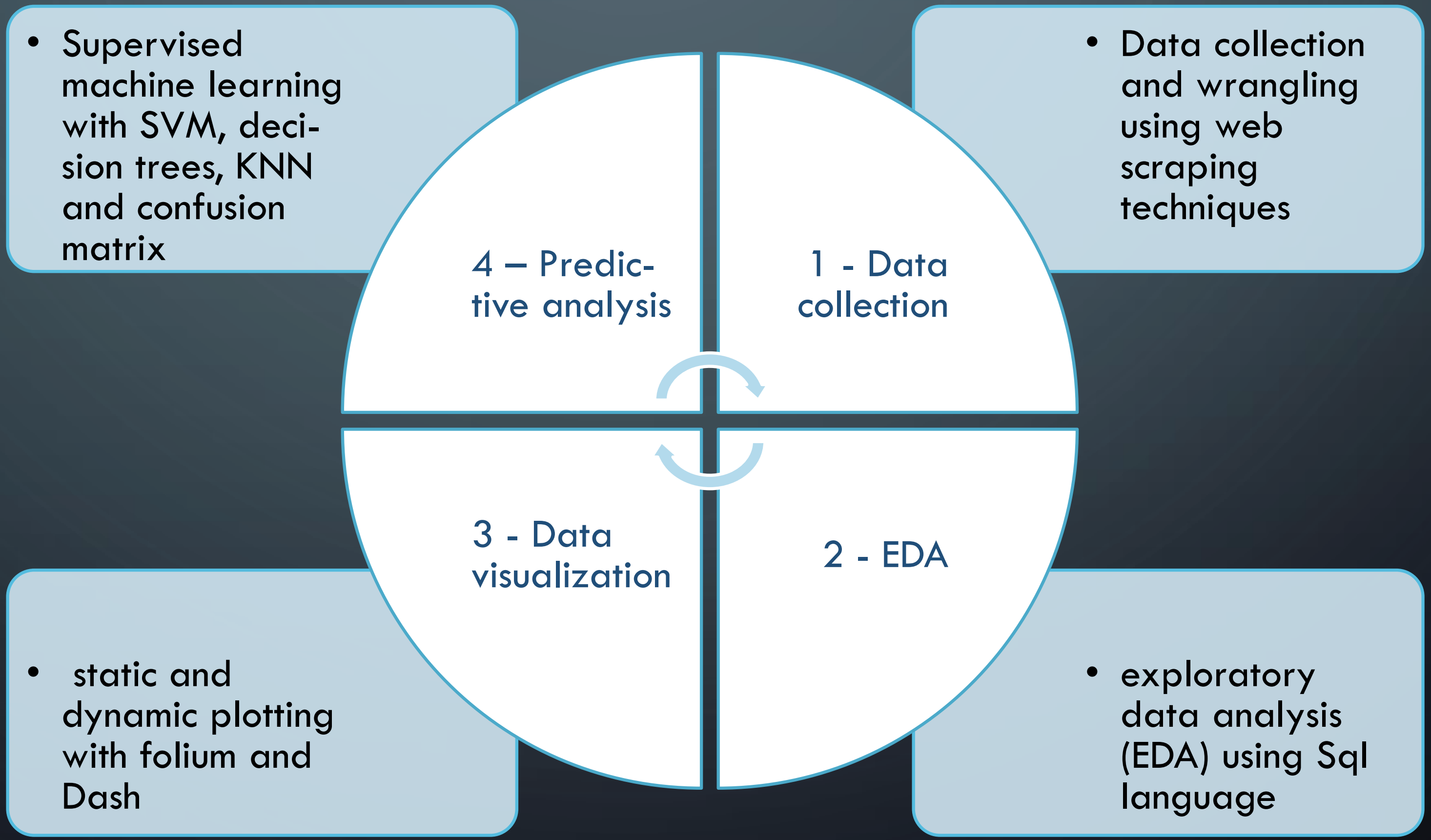
09/09/2021

# OUTLINE

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

# EXECUTIVE SUMMARY

- Summary of methodologies



1 - Data collection

- Data collection and wrangling using web scraping techniques

2 - EDA

- exploratory data analysis (EDA) using Sql language

3 - Data visualization

- static and dynamic plotting with folium and Dash

4 – Predictive analysis

- Supervised machine learning with SVM, decision trees, KNN and confusion matrix

# EXECUTIVE SUMMARY

- Summary of all results
  - **66%** of success rate for F9 type vehicles,
  - **92%** of the overall success goes to **FT and B4 booster versions**,
  - Optimal payload carried = **9.600 Kg**,
  - Payloads < 6.000 Kg have the highest chances of success,
  - ES-L1, GEO, HEO and SSO orbits recorded the highest chances of landing success,
  - Launches near equator line are cost efficient,
  - Coast proximity to launch sites, and remoteness to cities, are an important safety factors for our programs,
  - We can predict launches success rates with **91% accuracy** using decision tree method.

# INTRODUCTION

- Project background and context:
  - Many competing companies are investing in commercial spatial flights, which opens a wide range of business opportunities: spatial tourism, commercial satellites … And among them SpaceX has the most promising business model, due to its reusable and cost-killing rockets.

- Problems you want to find answers:
  - What is the likelihood of spaceX rockets first stage landing ?
  - Can we therefore determine the cost of a flight on spaceX rockets ?
  - How can we take advantage of spaceX data ?

# METHODOLOGY

- Data collection methodology:
  - We’ll describe how data were collected
- Perform data wrangling
  - We’ll describe how data were processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

# METHODOLOGY

IN THIS SECTION, WE WILL DESCRIBE IN DEPTH THE METHODOLOGIES WE USED IN OUR RESEARCH

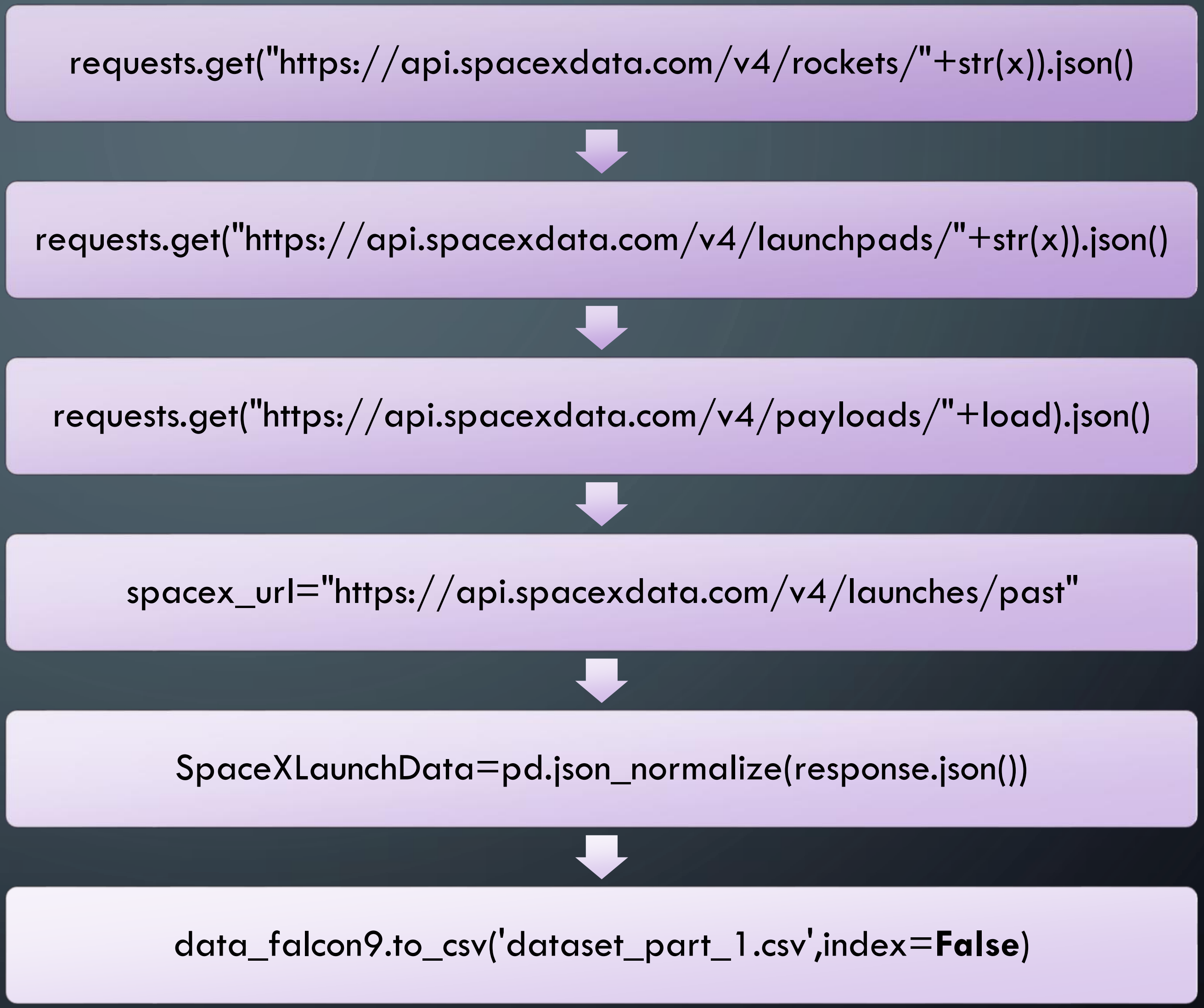
# DATA COLLECTION – SPACEX API

Using a **REST** call, we sent a request to spaceX server to collect the necessary data through an API.

We then normalize the .json content obtained, using a using **.json_normalize()** query and convert it into a dataframe.

We then clean and format the data, to have a workable dataframe. At least we convert it into a .csv table for further exploitation.

See the whole notebook for reference here:



```
requests.get("https://api.spacexdata.com/v4/rockets/"+str(x)).json()
```

```
requests.get("https://api.spacexdata.com/v4/launchpads/"+str(x)).json()
```

```
requests.get("https://api.spacexdata.com/v4/payloads/"+load).json()
```

```
spacex_url="https://api.spacexdata.com/v4/launches/past"
```

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

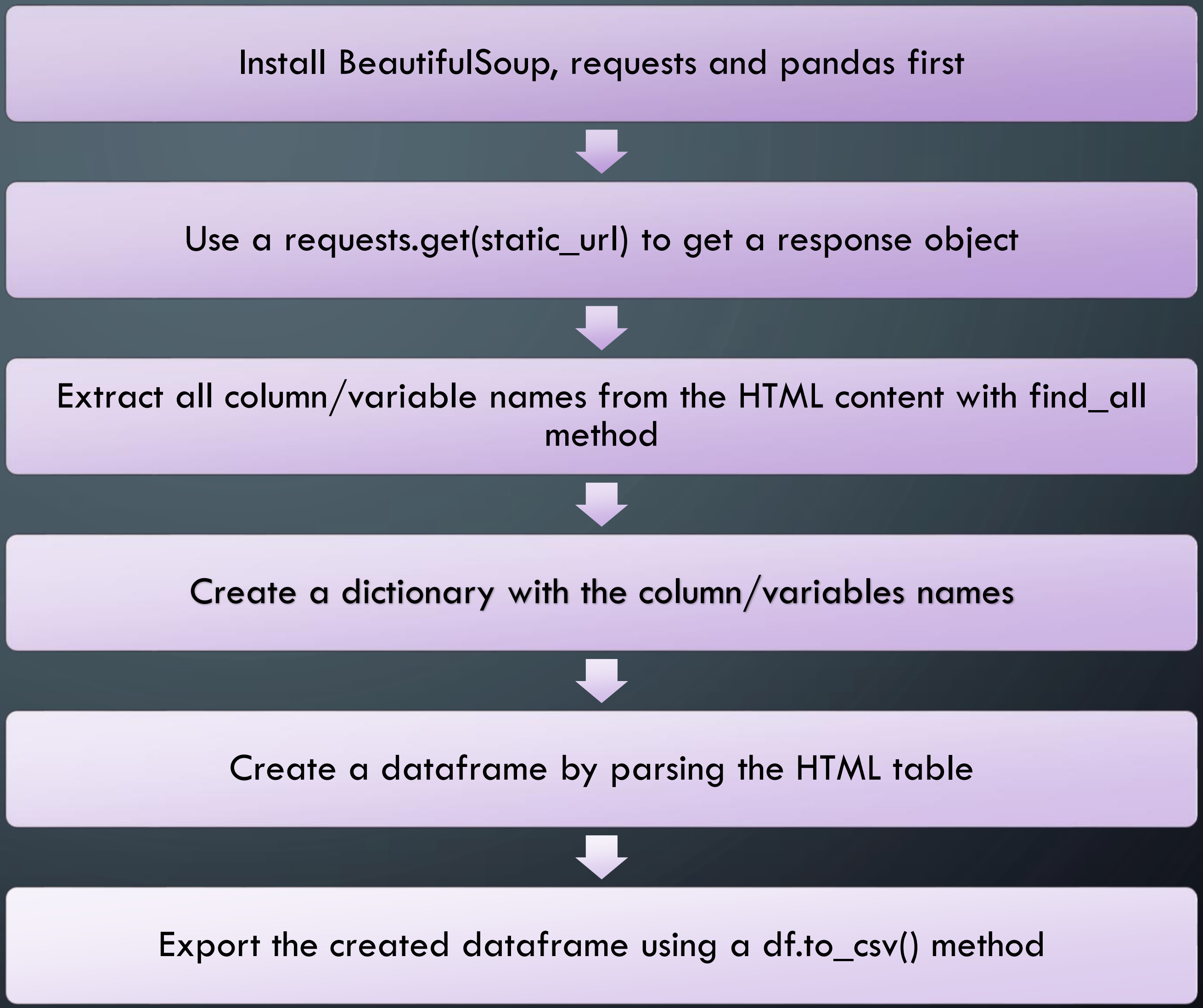
```
data_falcon9.to_csv('dataset_part_1.csv',index=False)
```

https://github.com/Henri-Saub/Capstone-project-DS-ML/blob/a67ddc365279c168c465a6ab89f0427d293fe1ce/Data%20Collection%20API%20Lab_final.ipynb

# DATA COLLECTION – WEB SCRAPING

After getting spacex static URL and performing a requests.get method on it, we used a find_all method to gather all <td> (rows) et <th> (headers) elements of the HMTL table containing the data we are looking for.

These elements are then combined into a panda dataframe, and after converted into a .csv file for further exploitation.



Install BeautifulSoup, requests and pandas first

Use a requests.get(static_url) to get a response object

Extract all column/variable names from the HTML content with find_all method

Create a dictionary with the column/variables names

Create a dataframe by parsing the HTML table

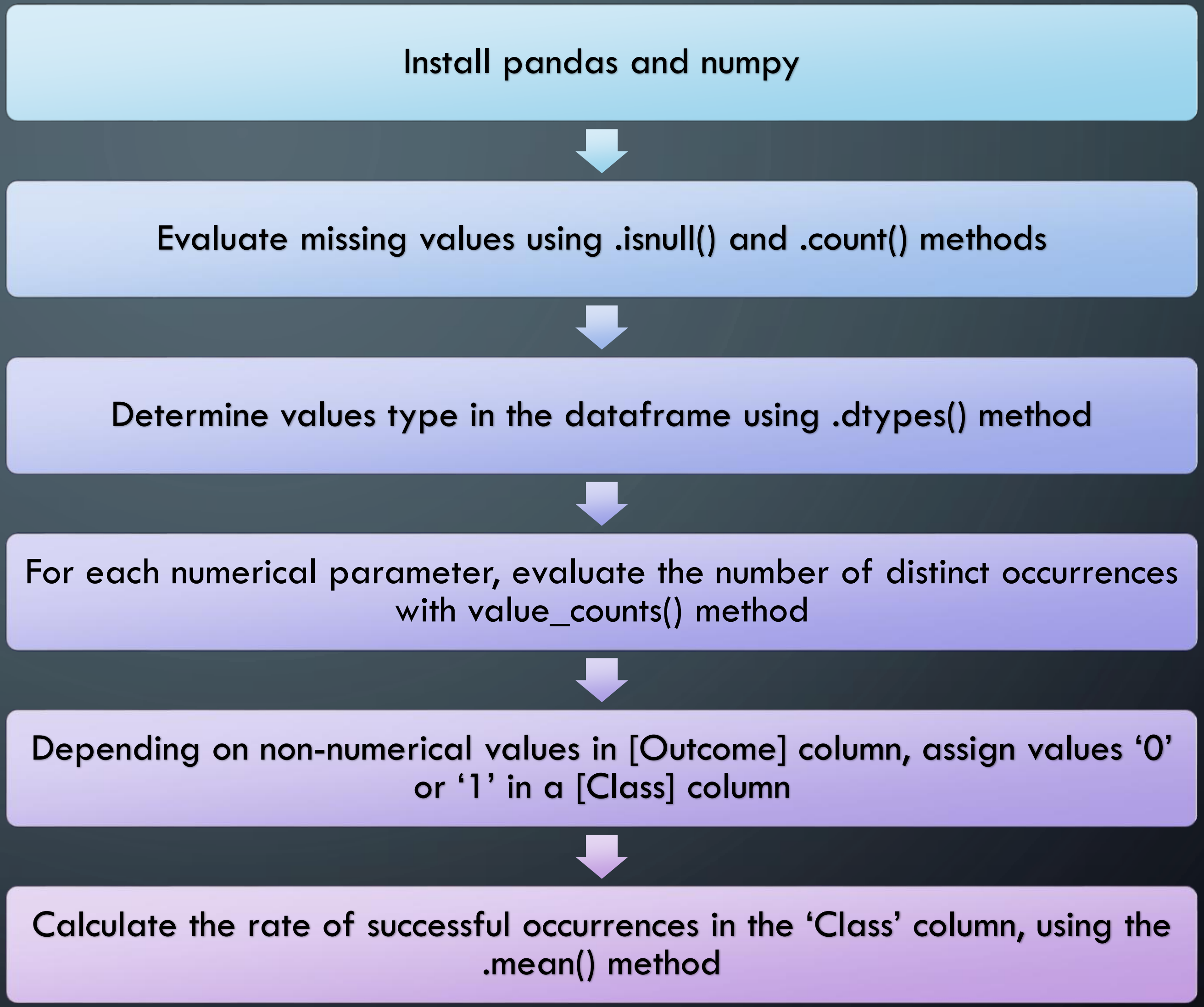
Export the created dataframe using a df.to_csv() method

See the whole notebook for reference here: https://github.com/Henri-Saub/Capstone-project-DS-ML/blob/4f5844edf3a1ad2668dc1cc575b677721b8d4ae2/Data%20Collection%20with%20Web%20Scraping%20lab_final.ipynb

# DATA WRANGLING

In this key step, the main objective is to define the data type (int, bool or object), to make sure we can use it in further calculations.

Then, we cope with null or missing data by calculating the mean() of the existing values of the targeted columns; therefore, we replace them with the value of the mean, to insure the data consistency.



Install pandas and numpy

Evaluate missing values using .isnull() and .count() methods

Determine values type in the dataframe using .dtypes() method

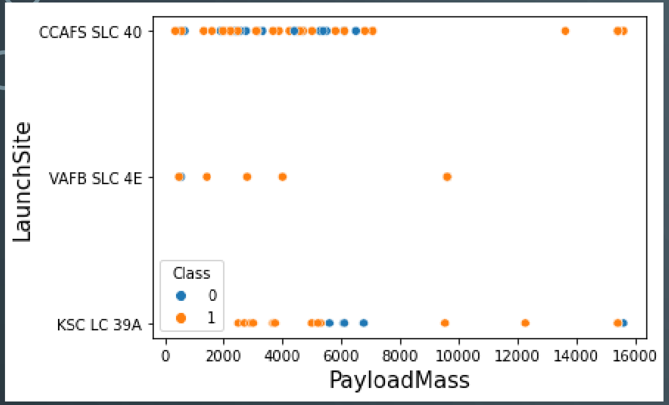
For each numerical parameter, evaluate the number of distinct occurrences with value_counts() method

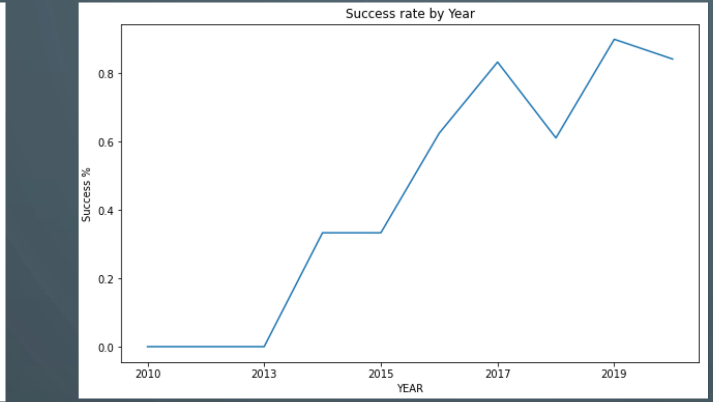
Depending on non-numerical values in [Outcome] column, assign values ‘0’ or ‘1’ in a [Class] column

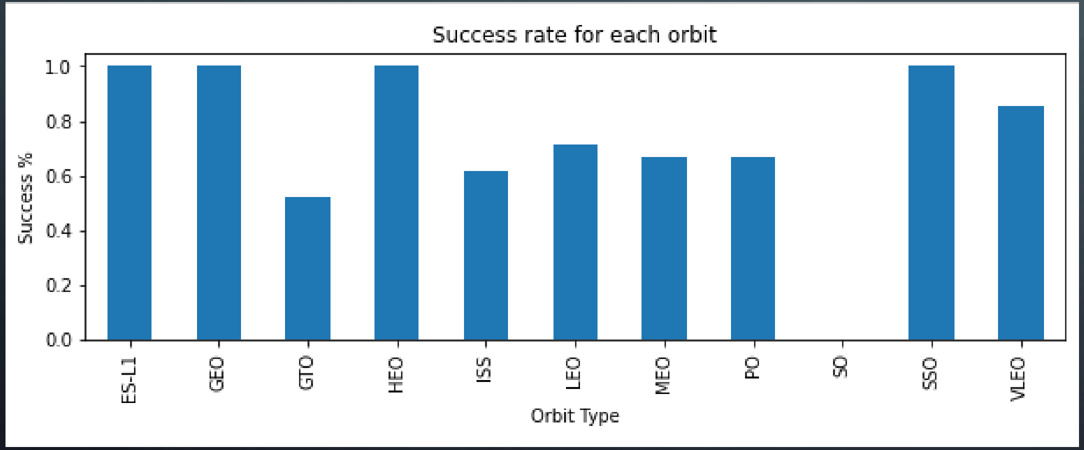
Calculate the rate of successful occurrences in the ‘Class’ column, using the .mean() method

See the whole notebook for reference here: https://github.com/Henri-Saub/Capstone-project-DS-ML/blob/4f5844edf3a1ad2668dc1cc575b677721b8d4ae2/labs-jupyter-spacex-Data%20wrangling_final.ipynb

# EDA WITH DATA VISUALIZATION





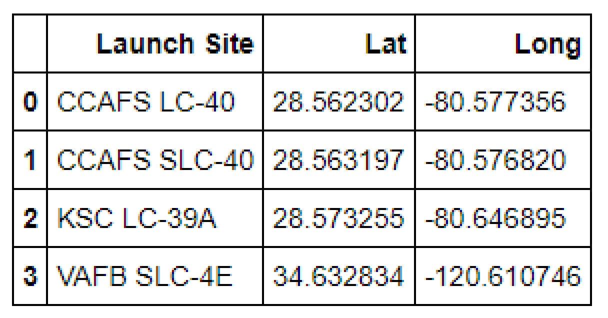


- PayloadMass vs Launch outcome plot
- LaunchSite vs Flight_Number plot
- LaunchSite vs Outcome plot
- PayloadMass vs LaunchSite scatterplot
- Class vs Orbit bar chart
- FlightNumber vs Orbit scatterplot
- PayloadMass vs Orbit scatterplot
- Class vs Date line chart
- See the whole notebook for reference here: https://github.com/Henri-Saub/Capstone-project-DS-ML/blob/4f5844edf3a1ad2668dc1cc575b677721b8d4ae2/jupyter-labs-eda-dataviz_final.ipynb

# EDA WITH SQL

- **Overview of performed SQL queries in this step:**
  - Names of the unique launch sites in the space mission,
  - 5 records where launch sites begin with the string 'KSC‘,
  - Total payload mass carried by boosters launched by NASA (CRS),
  - Average payload mass carried by booster version F9 v1.1,
  - Date where the first successful landing outcome in drone ship was achieved,
  - Names of the boosters which have success in ground pad and have payload mass > 4000 but < 6000 Kg,
  - Total number of successful and failure mission outcomes,
  - Names of the booster versions which have carried the maximum payload mass,
  - Records of the month names, successful landing outcomes in ground pad, plus booster versions, launch site for the months in year 2017,
  - Ranking of the Count of successful landing outcomes, between 2010-06-04 and 2017-03-20 in descending order.
- See the whole notebook for reference here: https://github.com/Henri-Saub/Capstone-project-DS-ML/blob/4f5844edf3a1ad2668dc1cc575b677721b8d4ae2/jupyter-labs-eda-sql-edx_final.ipynb

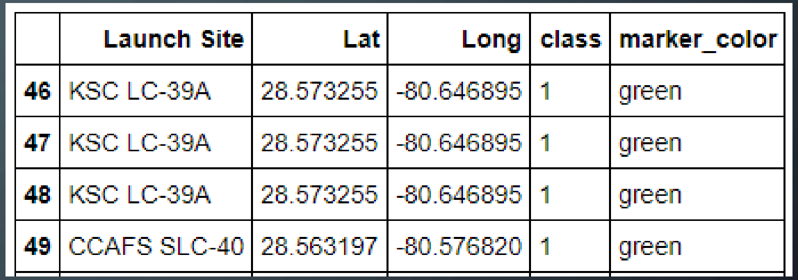
# BUILD AN INTERACTIVE MAP WITH FOLIUM



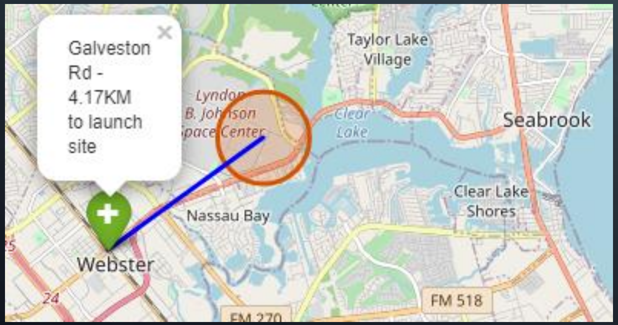
| | Launch Site | Lat | Long |
|---|---|---|---|
| 0 | CCAFS LC-40 | 28.562302 | -80.577356 |
| 1 | CCAFS SLC-40 | 28.563197 | -80.576820 |
| 2 | KSC LC-39A | 28.573255 | -80.646895 |
| 3 | VAFB SLC-4E | 34.632834 | -120.610746 |

- First, we mapped all LaunchSites coordinates using folium.Marker and folium.Circle
- Then, using the whole dataframe, we tried using folium.Marker to display each LaunchSite total number of launches, the next challenge being to visually figure out the successes and failures per site

- At this point, we tried to map each LaunchSite launch outcomes using folium.MarkerCluster, distinguished by color: green for successful launches and red for unsuccessful launches.



| | Launch Site | Lat | Long | class | marker_color |
|---|---|---|---|---|---|
| 46 | KSC LC-39A | 28.573255 | -80.646895 | 1 | green |
| 47 | KSC LC-39A | 28.573255 | -80.646895 | 1 | green |
| 48 | KSC LC-39A | 28.573255 | -80.646895 | 1 | green |
| 49 | CCAFS SLC-40 | 28.563197 | -80.576820 | 1 | green |

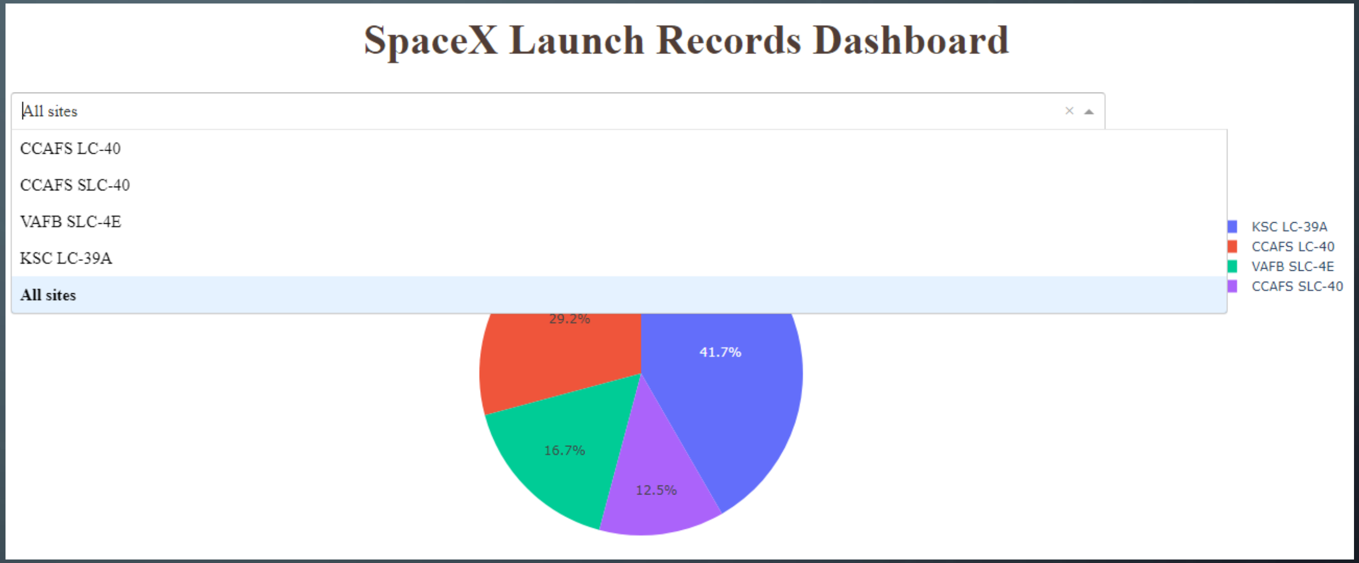


- At least, We calculated and mapped the distance between a LaunchSite and a railway, using folium.PolyLine, to illustrate how far are civil infrastructures from Launchsites, and what it means in terms of safety.
- See the whole notebook for reference here: [https://github.com/Henri-Saub/Capstone-project-DS-ML/blob/4f5844edf3a1ad2668dc1cc575b677721b8d4ae2/lab_jupyter_launch_site_location_final.ipynb](https://github.com/Henri-Saub/Capstone-project-DS-ML/blob/4f5844edf3a1ad2668dc1cc575b677721b8d4ae2/lab_jupyter_launch_site_location_final.ipynb)

# BUILD A DASHBOARD WITH PLOTLY DASH

- In this section, we created a pie chart of all successful flights by launch site, and a scatter plot of launches success rates, by payload mass

- By building such plots, we are trying to answer to some of the following questions:
    - Which site has the highest launch success rate?
    - Which payload range(s) has the highest/lowest launch success rate?



SpaceX Launch Records Dashboard

- See the whole notebook for reference here: https://github.com/Henri-Saub/Capstone-project-DS-ML/blob/4f5844edf3a1ad2668dc1cc575b677721b8d4ae2/spacex_dash_app.py

# PREDICTIVE ANALYSIS (CLASSIFICATION)

- Going from a sample of 90 observations, we splitted it into:
  - a **train set of 72** observations and
  - a **test set of 18**.

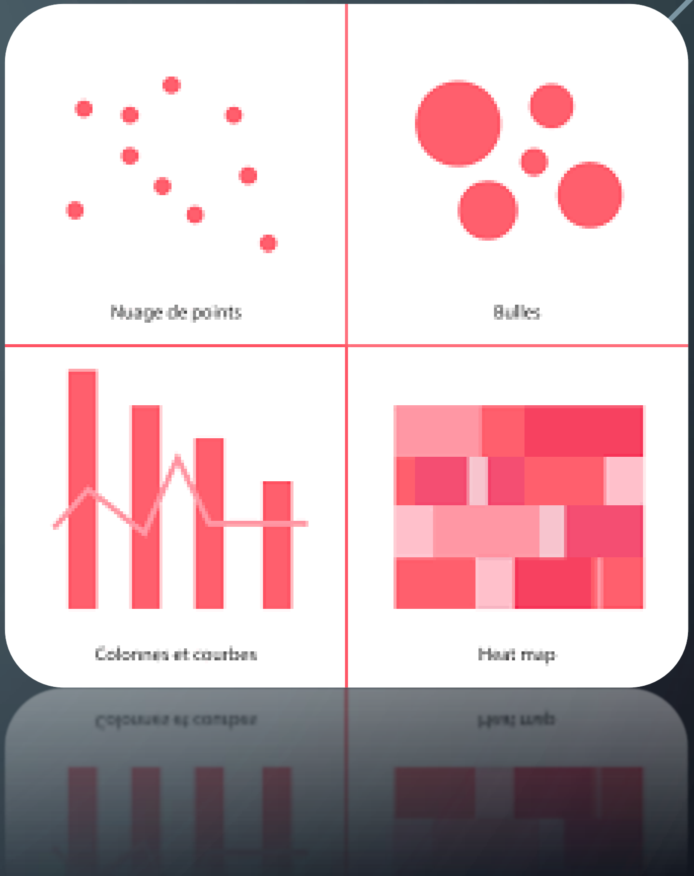
- Then we used **GridSearchCV** to pass our train set through a logistic regression, support vector machine, decision tree and K-nearest neighbor predictive models, and find which one has the best accuracy using their hyperparameters

- Then, we used the **.score()** method to pass the test sets through these models and check their accuracy scores

- We then plotted the training and test results using confusion matrix

# RESULTS

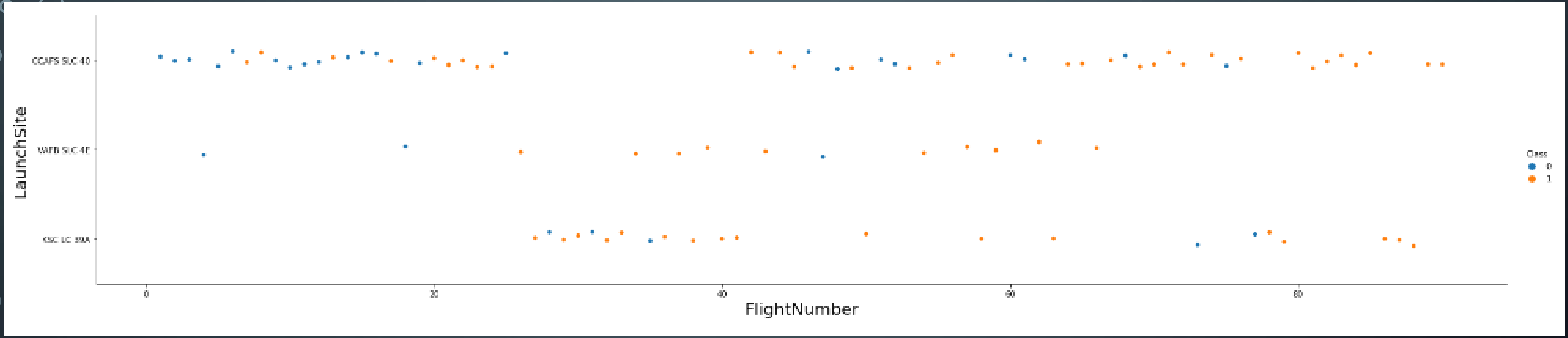
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results



# EDA WITH VISUALIZATION

IN THIS SECTION, WE WILL DISPLAY THE PLOTS GENERATED BY OUR PRIME DATA EXPLORATION

# FLIGHT NUMBER VS. LAUNCH SITE



CCAFS-SLC-40 Launshsite has the highest number of flights

**Hypothesis:** “CCAFS-SLC-40 is the 1st & main test-site for spaceX, the insights of which are used to perfect launches from the other facilities, which explains their higher success rate”

After the 41st launch, CCAFS-SLC-40 site's success rate significantly increases, which strengthen the likeliness of our upper hypothesis

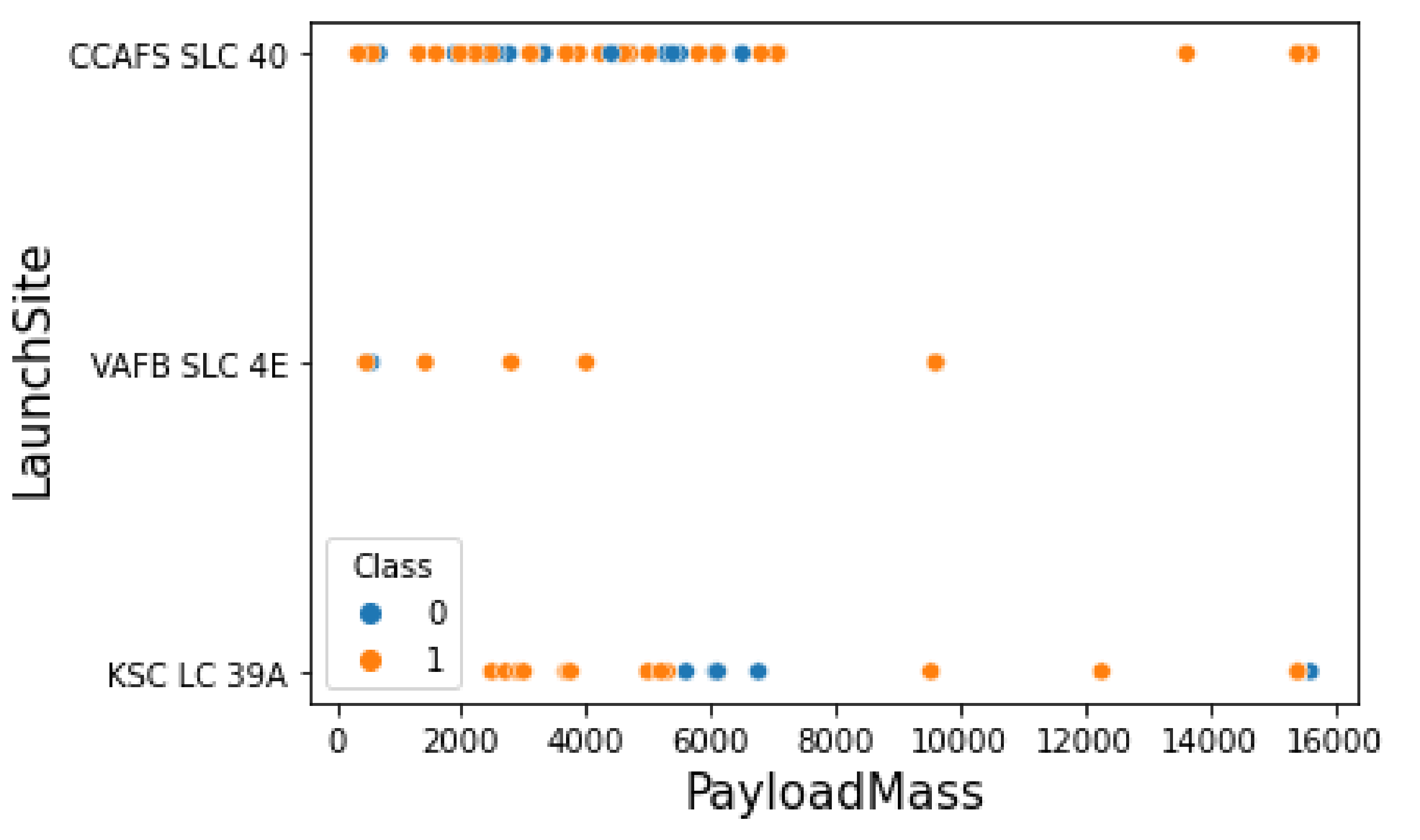
# PAYLOAD VS. LAUNCH SITE

1/ Most of the launches carried a payload_Mass <= 7000 Kg,

2/ 08 launches with a payload >9000 Kg done, 1 failure

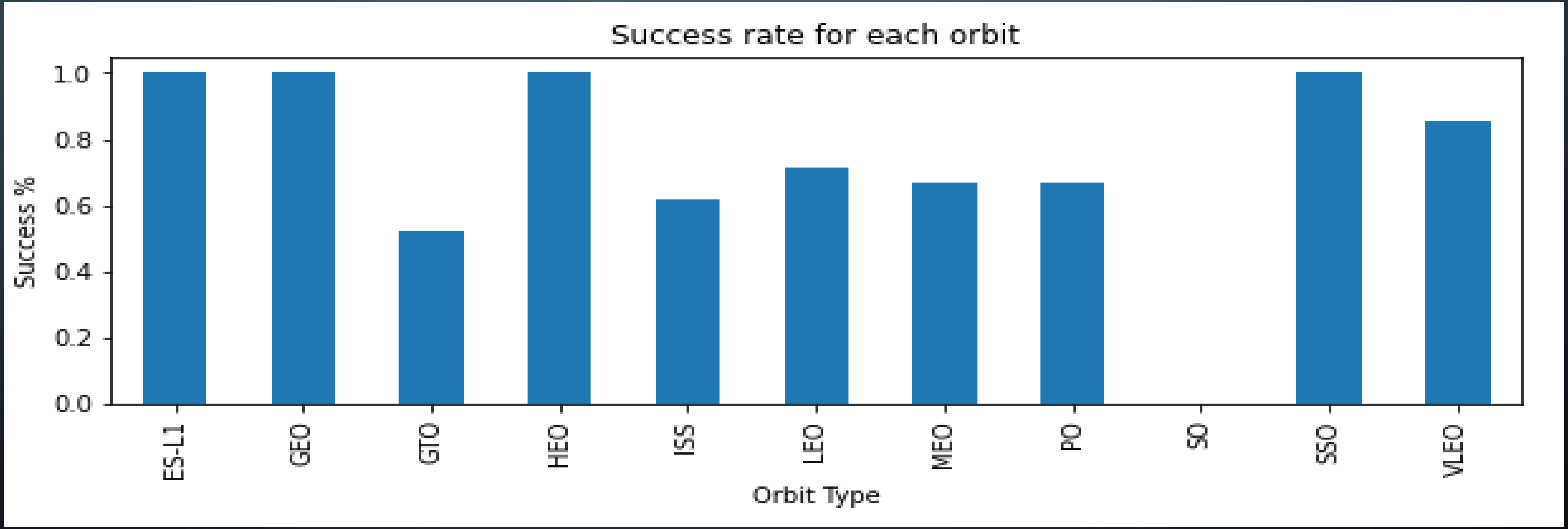
3/ VAFB-SLC-4E site has the highest success rate for payload_mass <= 6000, followed by KSC-LC-39A.

**Hypothesis:** " for intensive rockets use, the lesser the payload mass, the higher the probability of success"



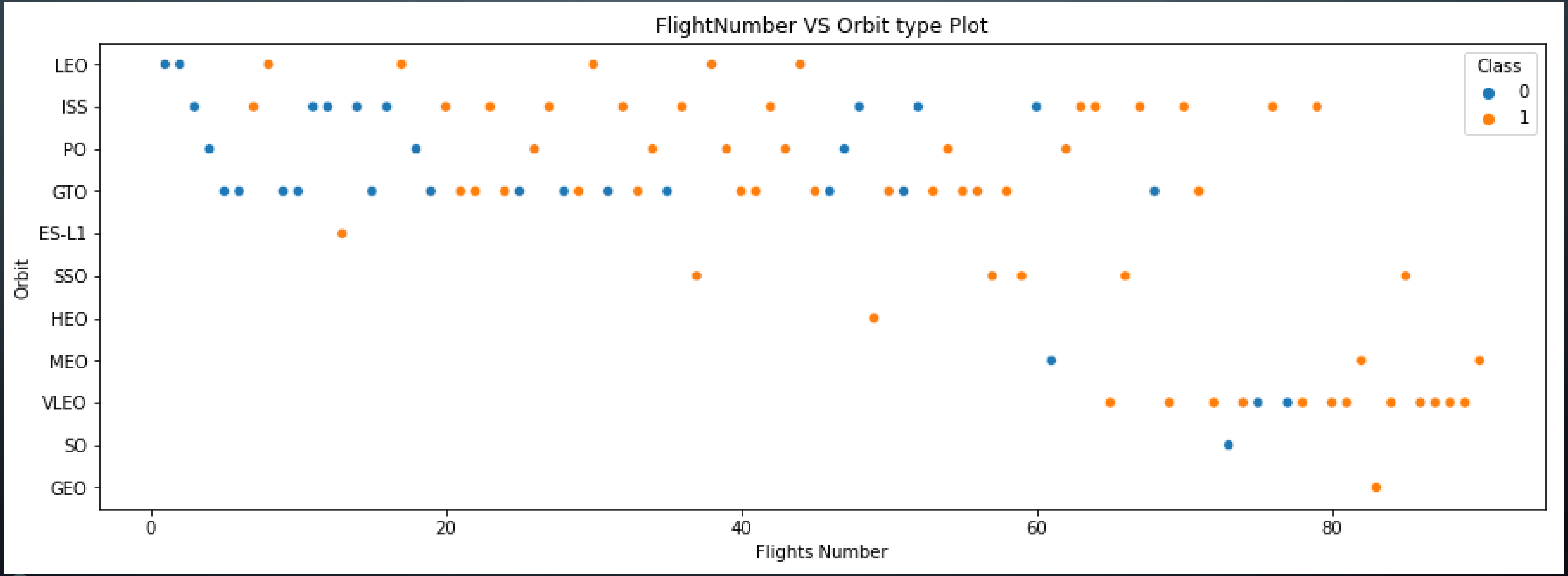
# SUCCESS RATE VS. ORBIT TYPE

**ES-L1, GEO, HEO and SSO** orbits have the **highest chances of landing success**. At these altitudes, satellites gain higher stability due to the geosynchronous nature. Which is crucial for telecommunication, military, and spatial exploration businesses.



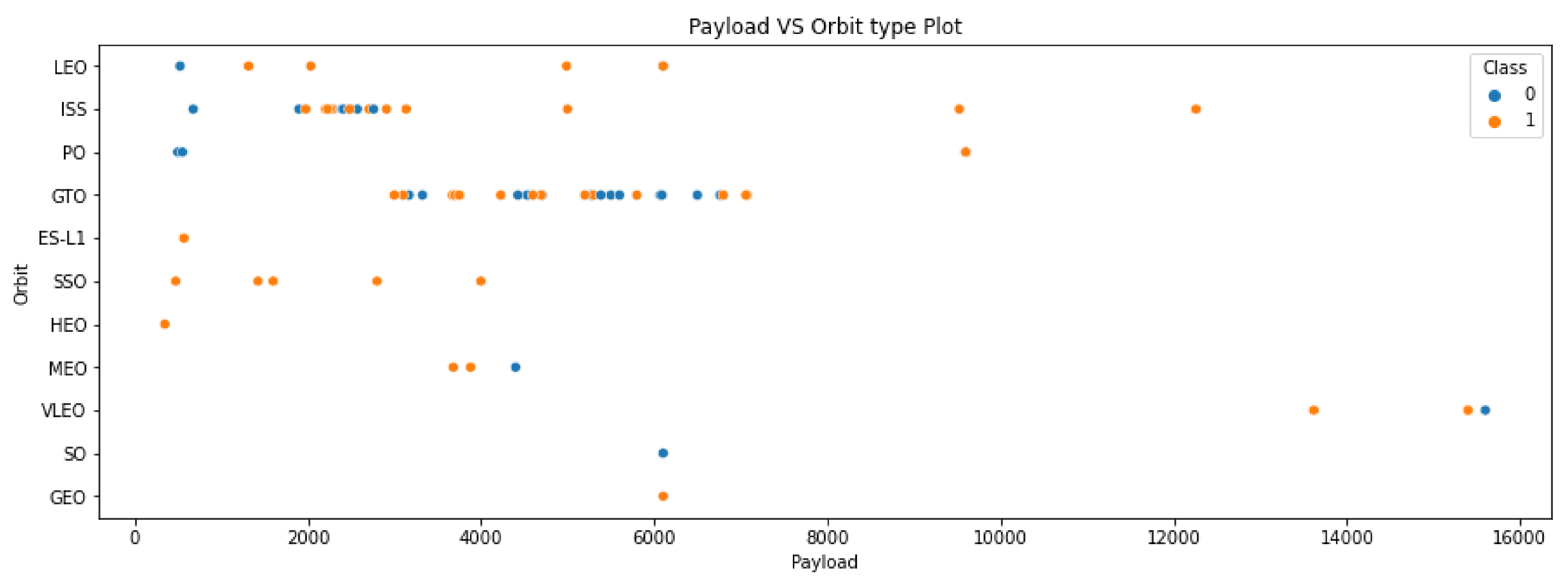
# FLIGHT NUMBER VS. ORBIT TYPE

Despite LEO orbit's success rate, appears related to the number of flights; there seems to be no positive correlation between flight number and the success rate.

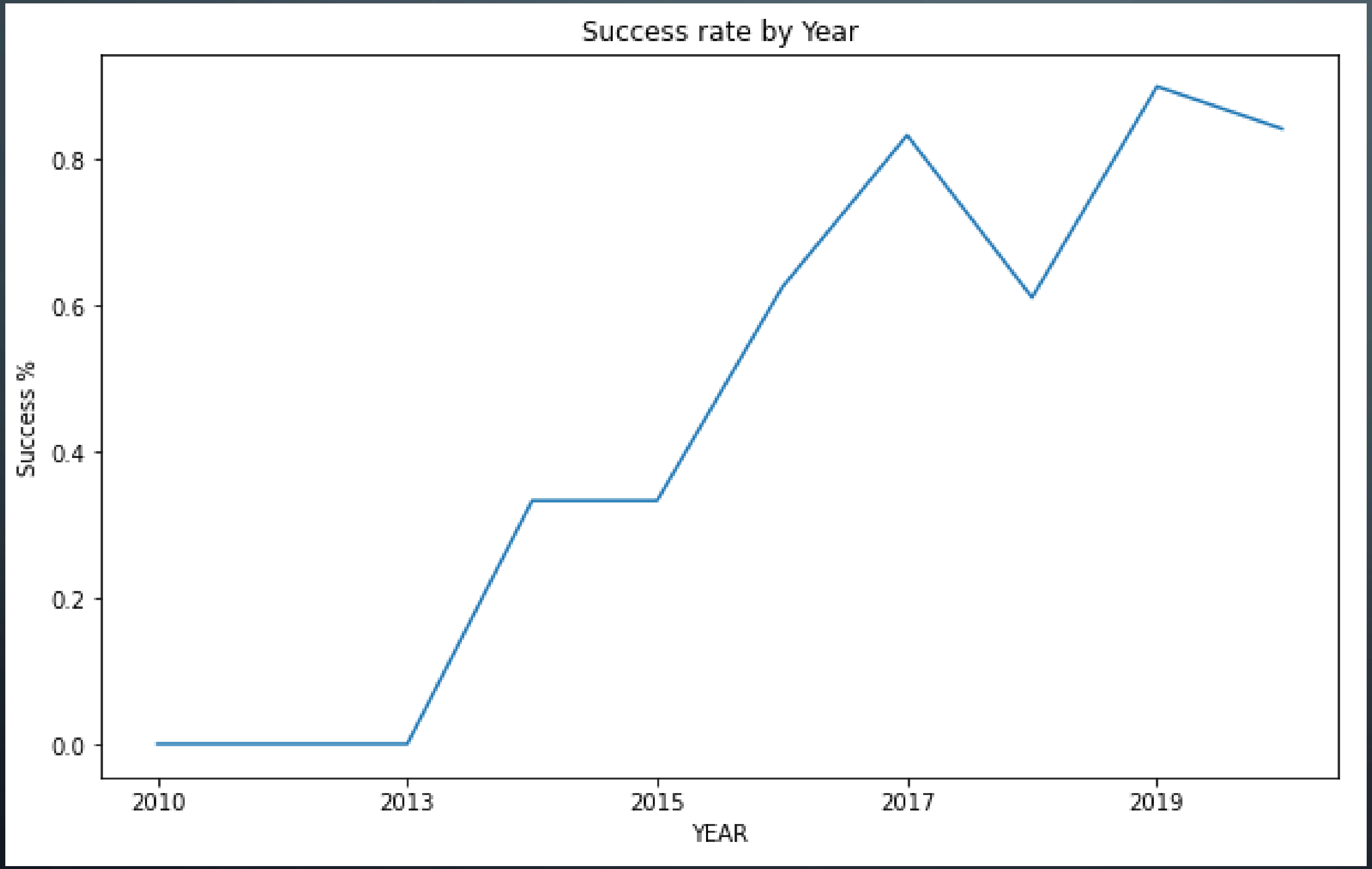


# PAYLOAD VS. ORBIT TYPE

Heavy payloads have a negative influence on GTO orbits, but positive Polar LEO (ISS) orbits



# LAUNCH SUCCESS YEARLY TREND



The Line chart shows an increasing trend of the yearly success rate over the years, which seems to sustain our hypothesis in the 'Flight Number vs. Launch Site' chart.

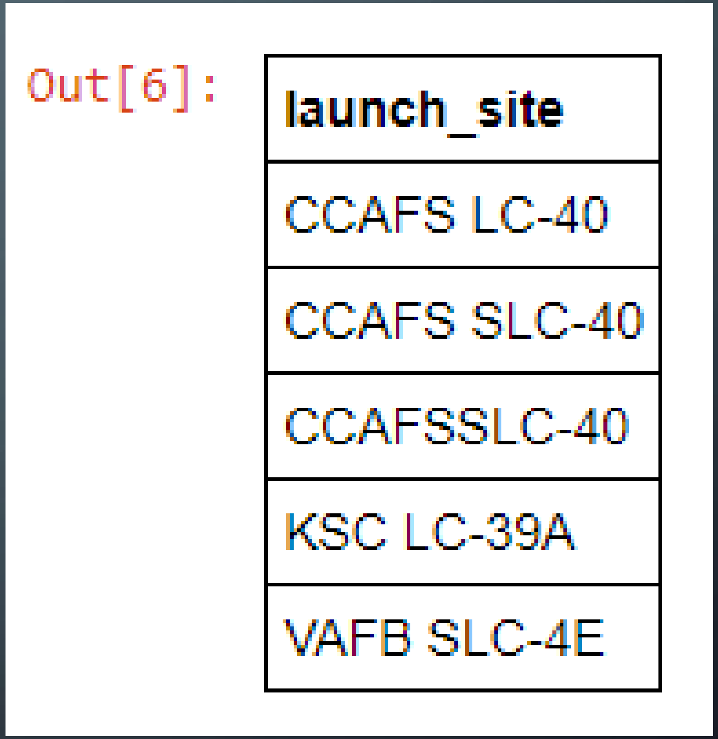
# EDA WITH SQL

IN THIS SECTION, WE WILL SHOW THE RESULTS OF OUR EXPLORATORY ANALYSIS USING SQL FRAMEWORK

# ALL LAUNCH SITE NAMES

- Using magic Sql we found 5 different sites names which, in reality, are **four.**

- This highlights the necessity to clean up the data for further exploration

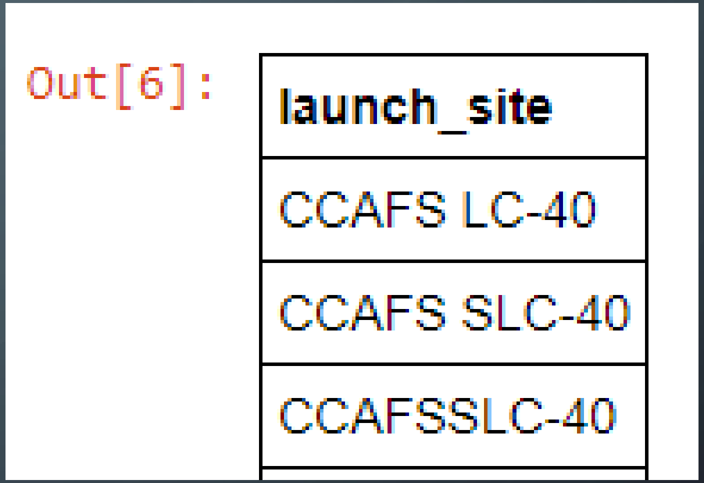


Out[6]:

| launch_site |
|---|
| CCAFS LC-40 |
| CCAFS SLC-40 |
| CCAFSSLC-40 |
| KSC LC-39A |
| VAFB SLC-4E |

# LAUNCH SITE NAMES BEGIN WITH `CCA`

- Here are the names of the sites beginning by ‘CCA’; as you can see there is a space missing in the third, to match the second.



Out[6]:

| launch_site |
|---|
| CCAFS LC-40 |
| CCAFS SLC-40 |
| CCAFSSLC-40 |

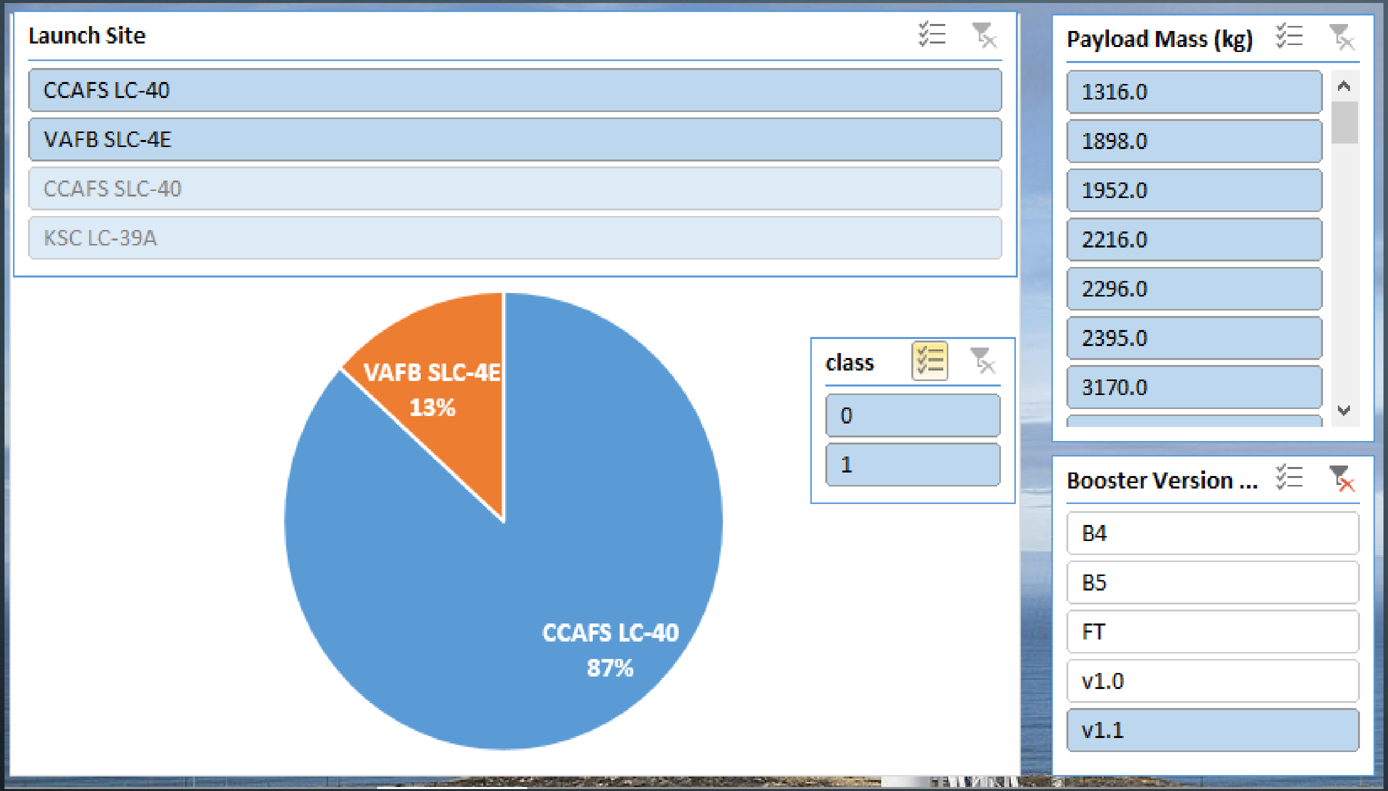
# TOTAL PAYLOAD MASS

- The total payload carried by boosters from NASA is **48213 Kg**

- It is likely that, upon receipt of a NASA's operational long duration flights certification if the current mission on ISS succeeds, more contracts will come for deep exploration of the Moon and Mars.

*Credit: https://www.newsbytesapp.com*

# AVERAGE PAYLOAD MASS BY F9 V1.1

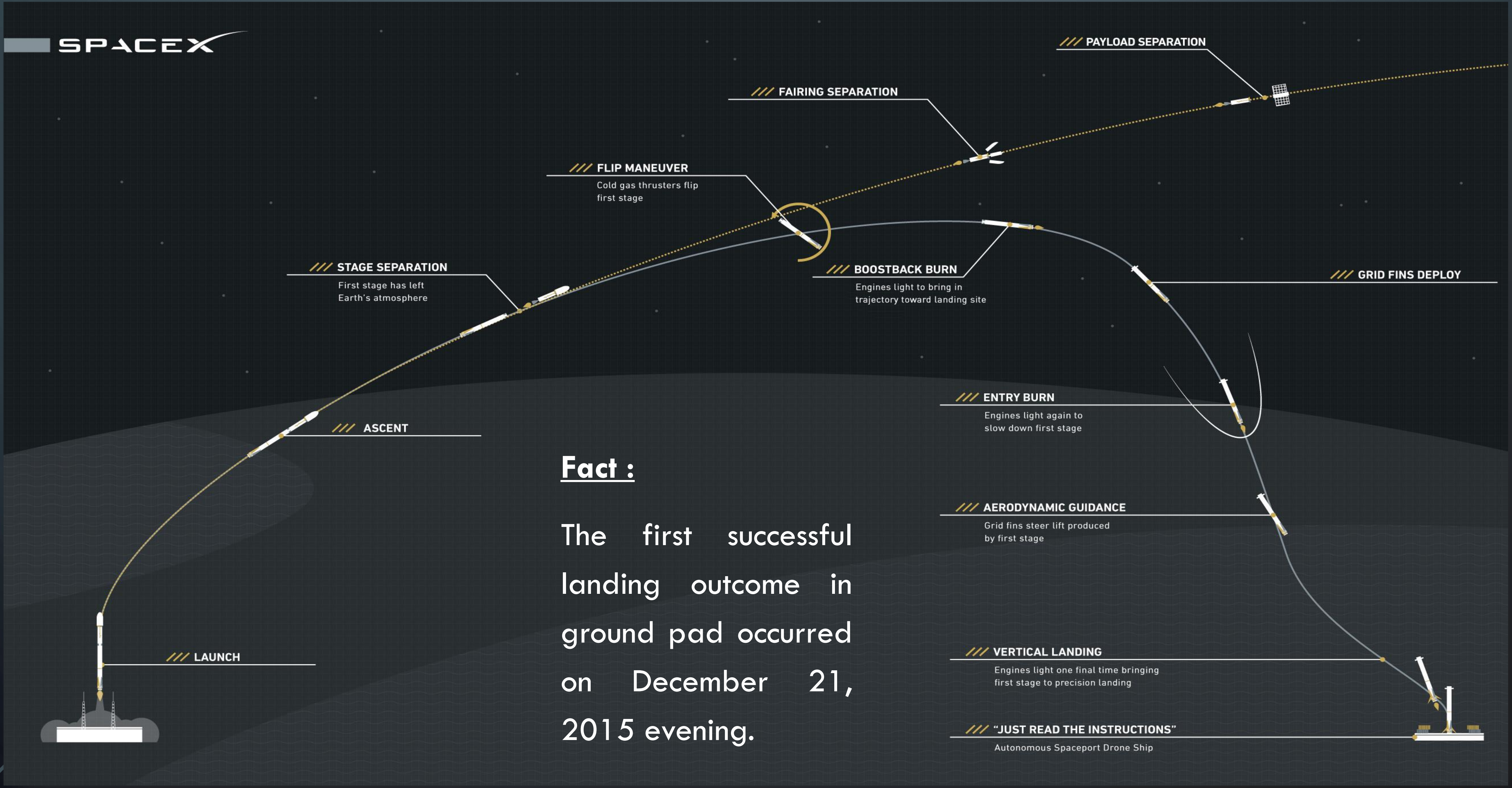


*Pivot table pie chart of the Payload range, by booster version, launch site and class*

- The average payload mass carried by booster version F9 v1.1 is **2928,40 Kg**
- This model was mainly used for payloads ranging from [500-4707Kg], and has a high success rate (14 over 15 flights).
- **Hypothesis:** its main purpose was the development and enhancement of reusable rockets technology

As recorded by spaceX, the only failure on this model was due to due to an overpressure event in the second stage oxygen tank.

# FIRST SUCCESSFUL GROUND LANDING DATE


SPACEX
/// PAYLOAD SEPARATION
/// FAIRING SEPARATION
/// FLIP MANEUVER
Cold gas thrusters flip first stage
/// STAGE SEPARATION
First stage has left Earth's atmosphere
/// BOOSTBACK BURN
Engines light to bring in trajectory toward landing site
/// GRID FINS DEPLOY
/// ENTRY BURN
Engines light again to slow down first stage
/// ASCENT
/// AERODYNAMIC GUIDANCE
Grid fins steer lift produced by first stage
/// LAUNCH
/// VERTICAL LANDING
Engines light one final time bringing first stage to precision landing
/// "JUST READ THE INSTRUCTIONS"
Autonomous Spaceport Drone Ship

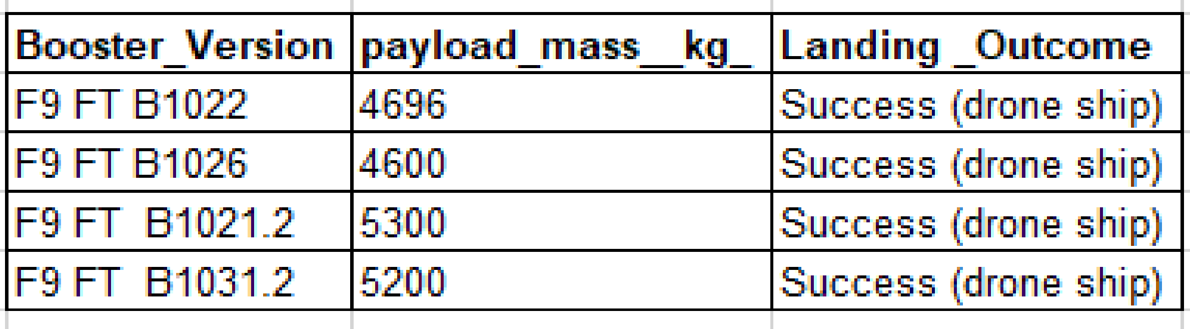
**Fact :**

The first successful landing outcome in ground pad occurred on December 21, 2015 evening.

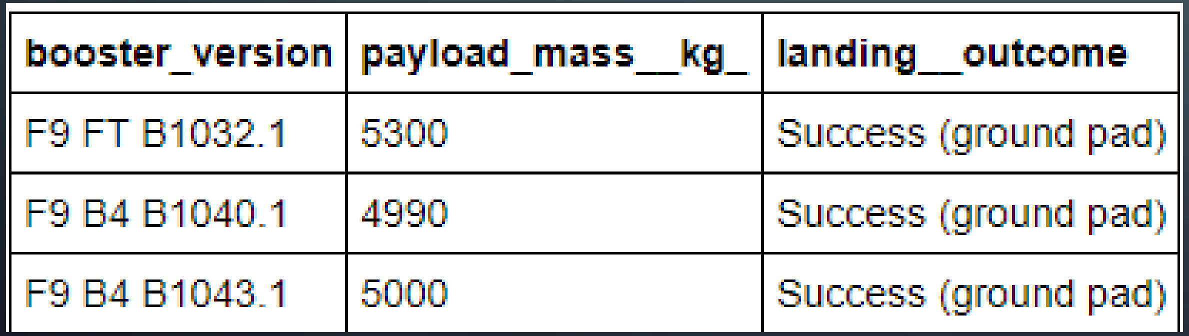
*https://commons.wikimedia.org/wiki/File:Falcon_9_First_Stage_Reusability_Graphic.jpg*

# SUCCESSFUL DRONE SHIP LANDING WITH PAYLOAD BETWEEN 4000 AND 6000 KG

- For similar payload mass range, the F9 ‘FT’ booster version is best fitted for drone ship landing …



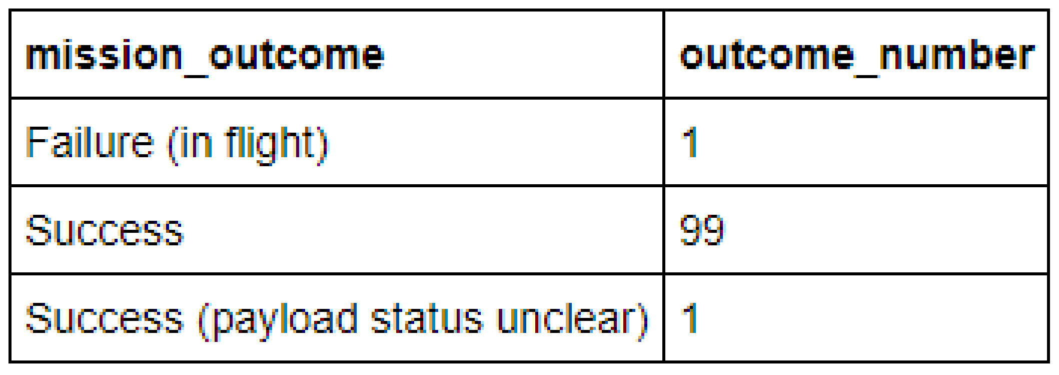
| Booster_Version | payload_mass__kg_ | Landing _Outcome |
|---|---|---|
| F9 FT B1022 | 4696 | Success (drone ship) |
| F9 FT B1026 | 4600 | Success (drone ship) |
| F9 FT  B1021.2 | 5300 | Success (drone ship) |
| F9 FT  B1031.2 | 5200 | Success (drone ship) |



| booster_version | payload_mass__kg_ | landing__outcome |
|---|---|---|
| F9 FT B1032.1 | 5300 | Success (ground pad) |
| F9 B4 B1040.1 | 4990 | Success (ground pad) |
| F9 B4 B1043.1 | 5000 | Success (ground pad) |

- On the contrary, the F9 ‘B4’ seems a better fit for ground pad landing

# TOTAL NUMBER OF SUCCESSFUL AND FAILURE MISSION OUTCOMES



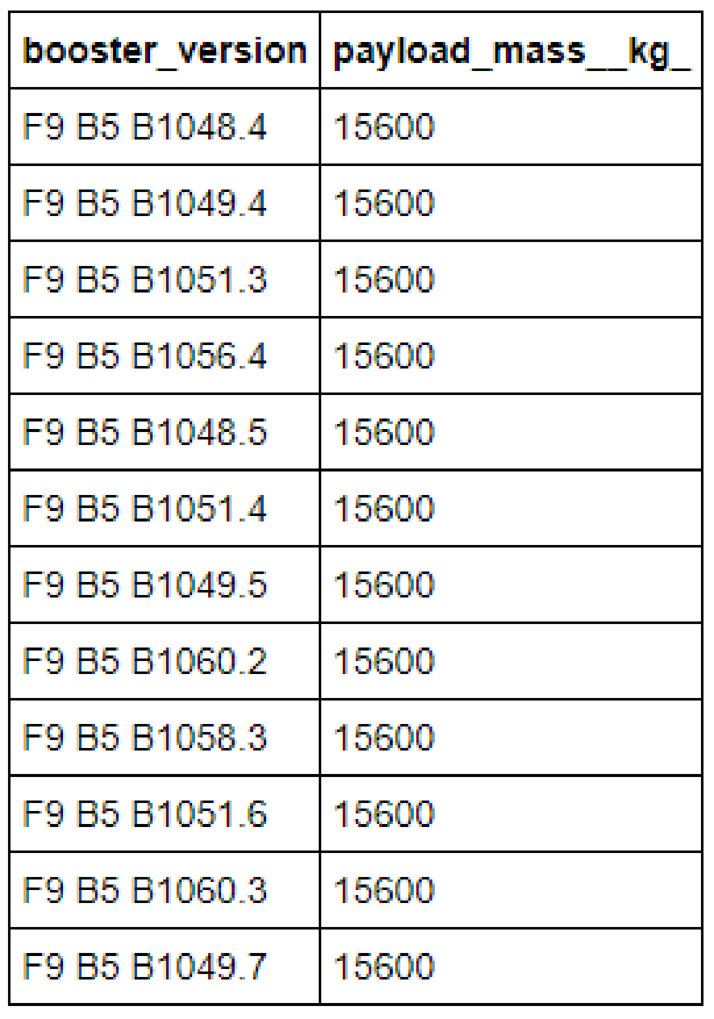
| mission_outcome | outcome_number |
|---|---|
| Failure (in flight) | 1 |
| Success | 99 |
| Success (payload status unclear) | 1 |

- Here is a chart of all missions outcome count.
- The only failure is related to the case we described in slide 28

# BOOSTERS CARRIED MAXIMUM PAYLOAD

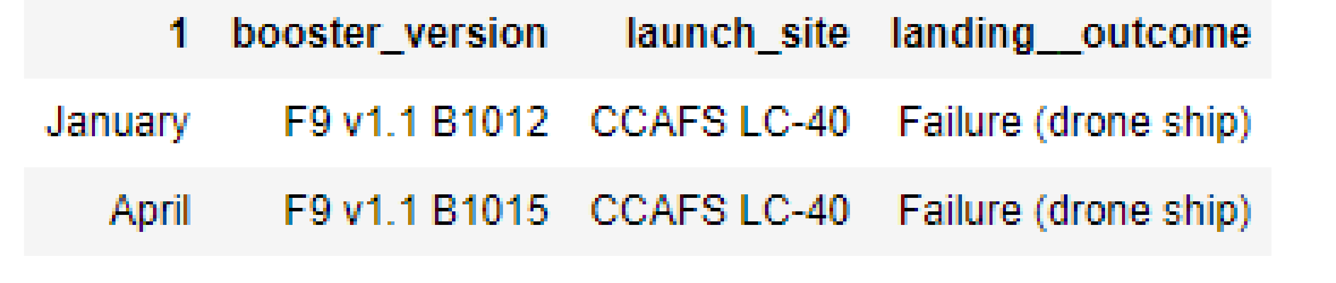
- Using a Sql subquery, we found the following list of boosters that carried a maximum payload of 15600 Kg :

- The **Falcon9 B5**, also known as “Falcon-Heavy B5” specializes in medium to heavy size missions.



| booster_version | payload_mass__kg_ |
|---|---|
| F9 B5 B1048.4 | 15600 |
| F9 B5 B1049.4 | 15600 |
| F9 B5 B1051.3 | 15600 |
| F9 B5 B1056.4 | 15600 |
| F9 B5 B1048.5 | 15600 |
| F9 B5 B1051.4 | 15600 |
| F9 B5 B1049.5 | 15600 |
| F9 B5 B1060.2 | 15600 |
| F9 B5 B1058.3 | 15600 |
| F9 B5 B1051.6 | 15600 |
| F9 B5 B1060.3 | 15600 |
| F9 B5 B1049.7 | 15600 |

# 2015 LAUNCH RECORDS

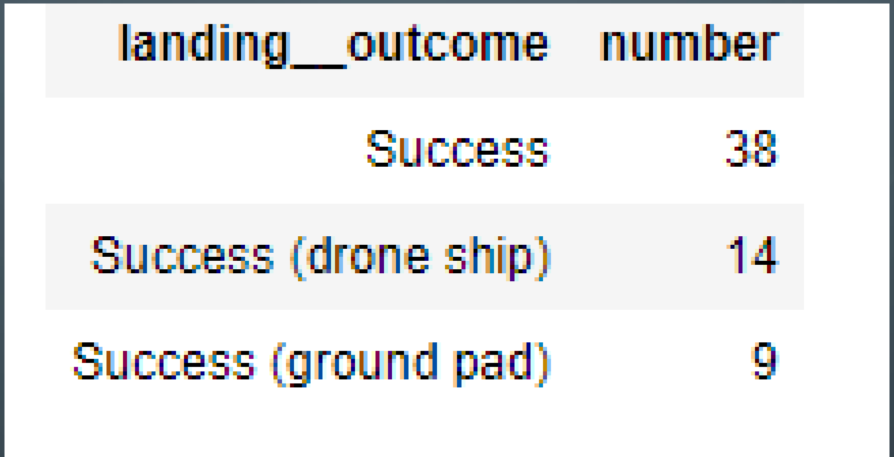


| 1 | booster_version | launch_site | landing__outcome |
|---|---|---|---|
| January | F9 v1.1 B1012 | CCAFS LC-40 | Failure (drone ship) |
| April | F9 v1.1 B1015 | CCAFS LC-40 | Failure (drone ship) |

*Records of month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015*

- Records say that during the first attempt, the rocket “became destabilized and exploded as it hit the barge because of a shortage of hydraulic fluid”; while during the second one, it landed vertically and on target, “but lateral motion caused it to tip over at the last second”

# RANK SUCCESS COUNT BETWEEN 2010-06-04 AND 2017-03-20



| landing__outcome | number |
|---|---|
| Success | 38 |
| Success (drone ship) | 14 |
| Success (ground pad) | 9 |

*Successful landing outcomes between 2010-06-04 and 2017-03-20 in descending order.*

# INTERACTIVE MAP WITH FOLIUM

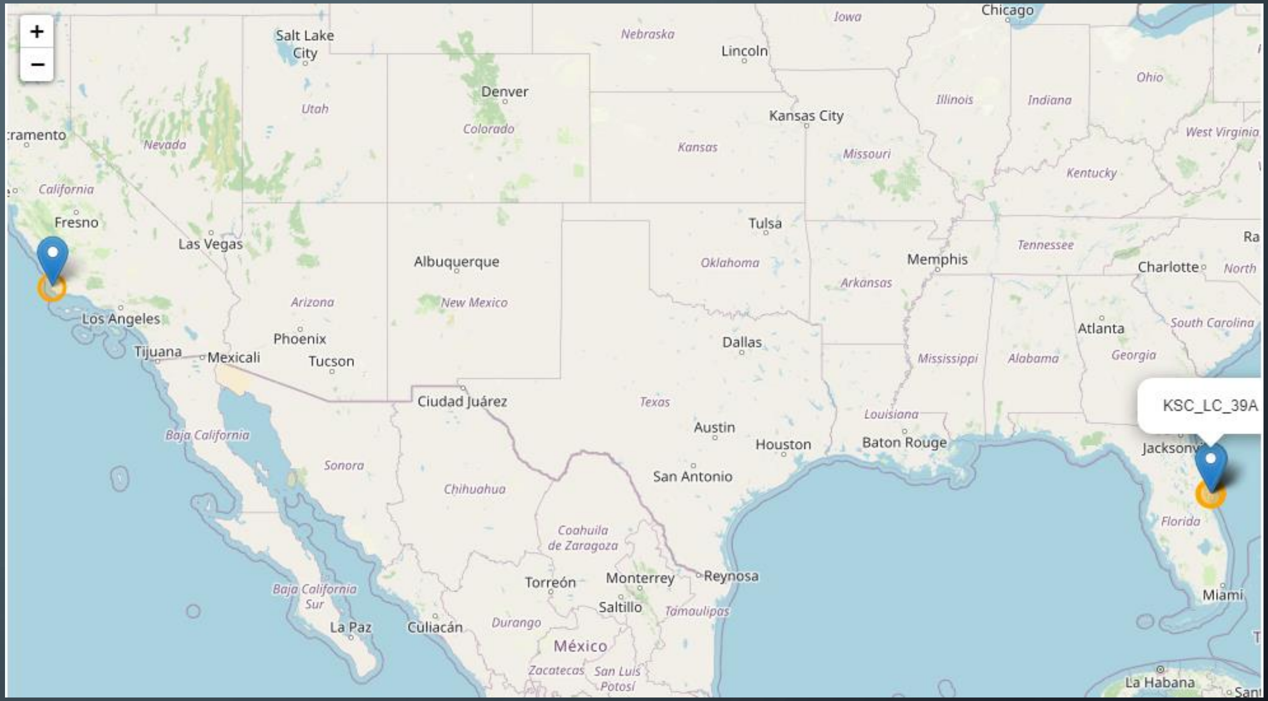
IN THIS SECTION, WE WILL PRESENT THE DIFFERENT MAPS WE OBTAINED USING FOLIUM IN OUR RESEARCH.

# MAP OF ALL LAUNCH SITES

**Observation_One:**

All launch sites are in proximity to the Equator line :

- Maximum earth spine,
- Faster rockets, due to supplement of kinetic energy, with maximum payload,
- The higher the celerity, the lesser the time to complete the mission distance,
- More the combustible saved: cost-efficient flight.



**Observation_Two:**

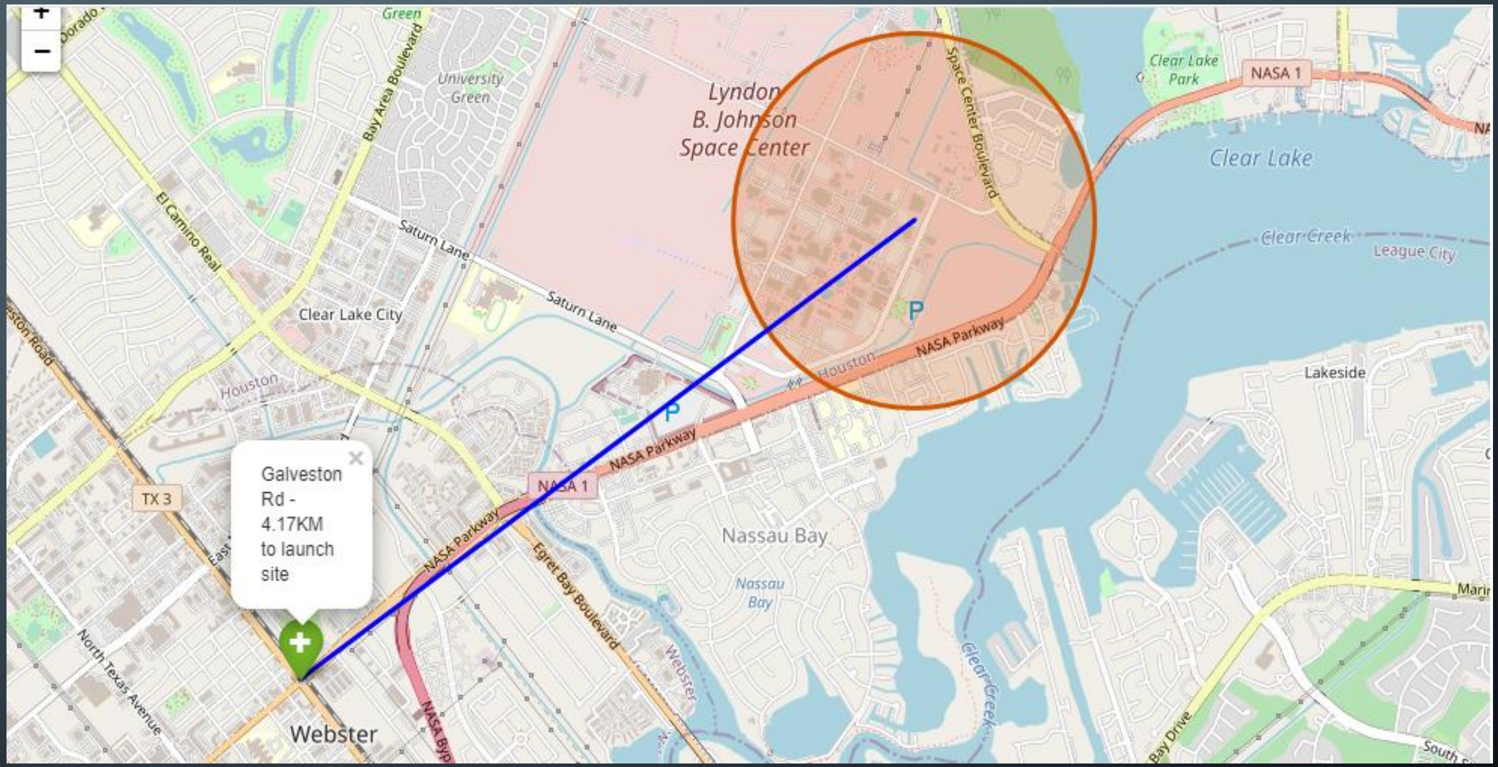
All launch sites are in very close proximity to the coast :

- water proximity in case of fire,
- Reduction of the wave blast impact created by motors ignition,
- Easier access to debris from the first stages which may be harder in the inner land,
- Safety for population in case of rocket disintegration in the lower atmosphere.

# DISTANCE TO CIVIL INFRASTRUCTURES

- Launch sites are mostly built as far away as possible from civil infrastructures such as railways, highways, cities … for safety in case of a catastrophic failure

- However, they seem close to bodies of water or coastlines to mitigate the risk of components falling over populated areas

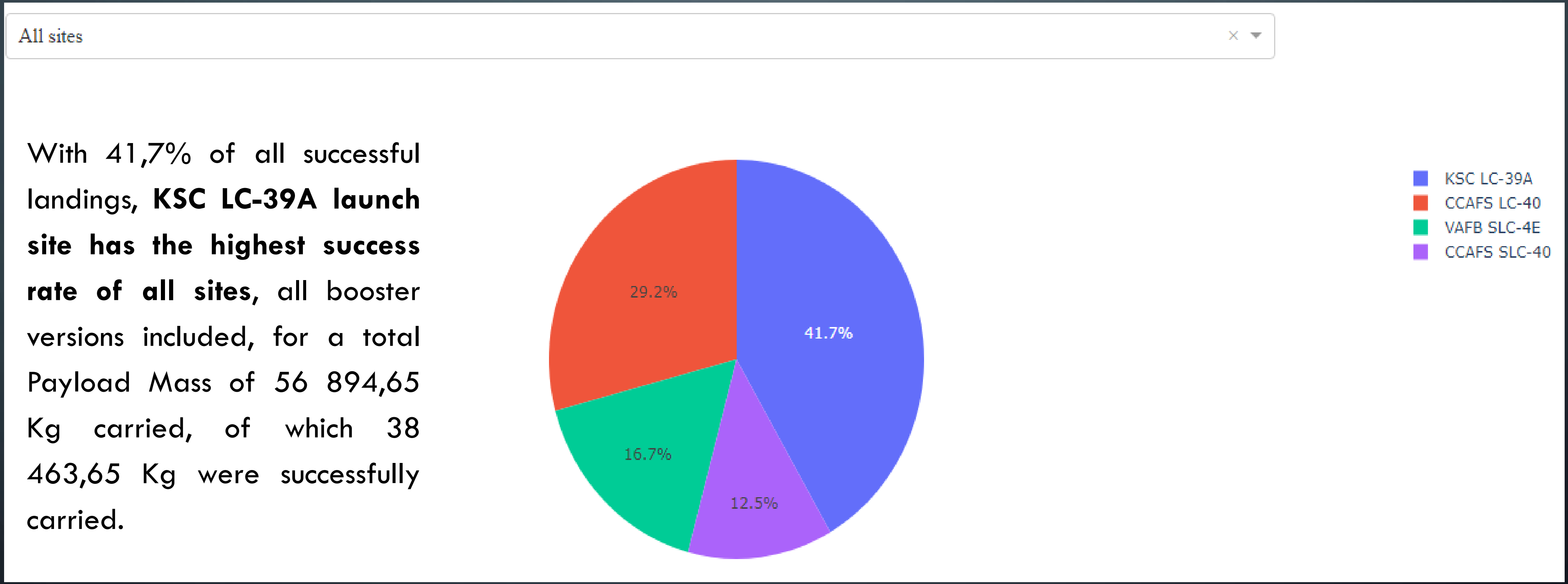


*Folium map of the distance between Houston Launch Site and nearest buildings and infrastructures*

# BUILD A DASHBOARD WITH PLOTLY DASH

IN THIS SECTION, WE EXPOSE THE RESULTS OF OUR DASH APP, TO VISUALIZE THE OUTPUTS OF OUR RESEARCH.

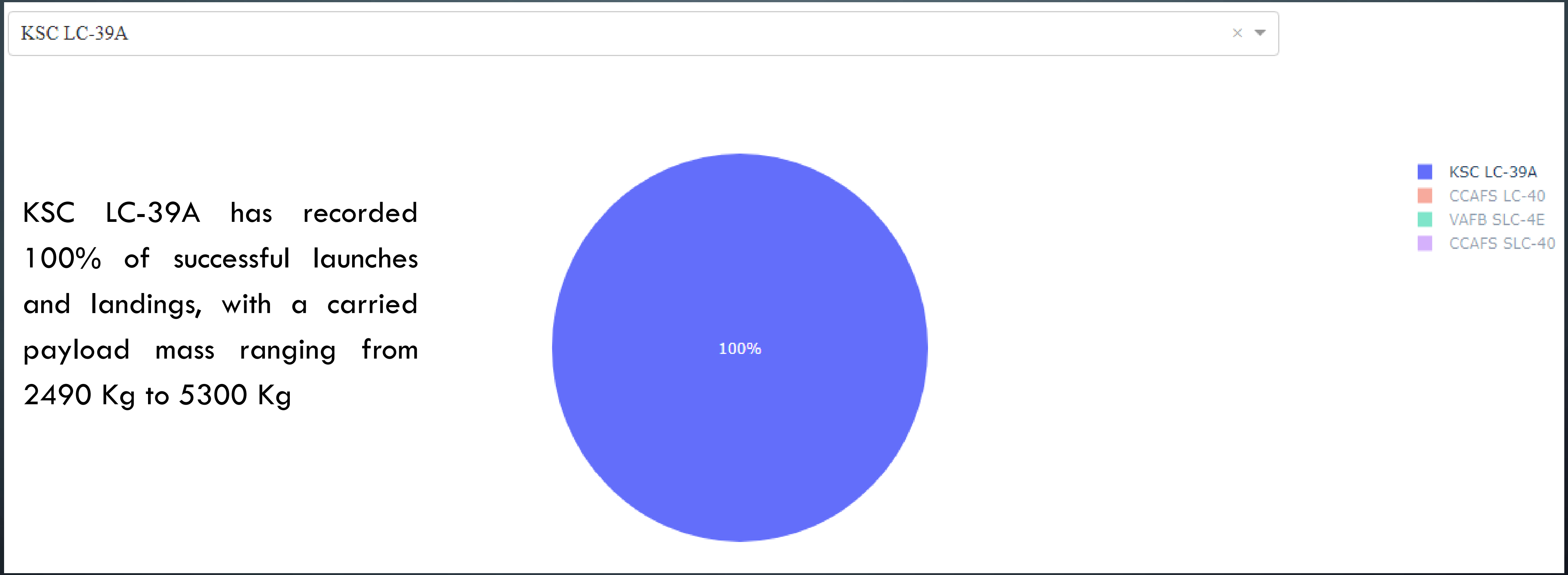
# SPACEX LAUNCH SUCCESS RATIO PER SITE



With 41,7% of all successful landings, **KSC LC-39A launch site has the highest success rate of all sites,** all booster versions included, for a total Payload Mass of 56 894,65 Kg carried, of which 38 463,65 Kg were successfully carried.

CCAFS LC-40, VAFB SLC-4E and CCAFS SLC-40 launch sites come next, with a total payload mass of 21 775 Kg, 29 275 Kg and 6 263,6 Kg successfully carried respectively

# KSC LC-39A LAUNCH SUCCESS RATIO



KSC LC-39A × ▾

KSC LC-39A has recorded 100% of successful launches and landings, with a carried payload mass ranging from 2490 Kg to 5300 Kg

*Dashboard screenshot of KSC LC-39A site success rate on Dash app*

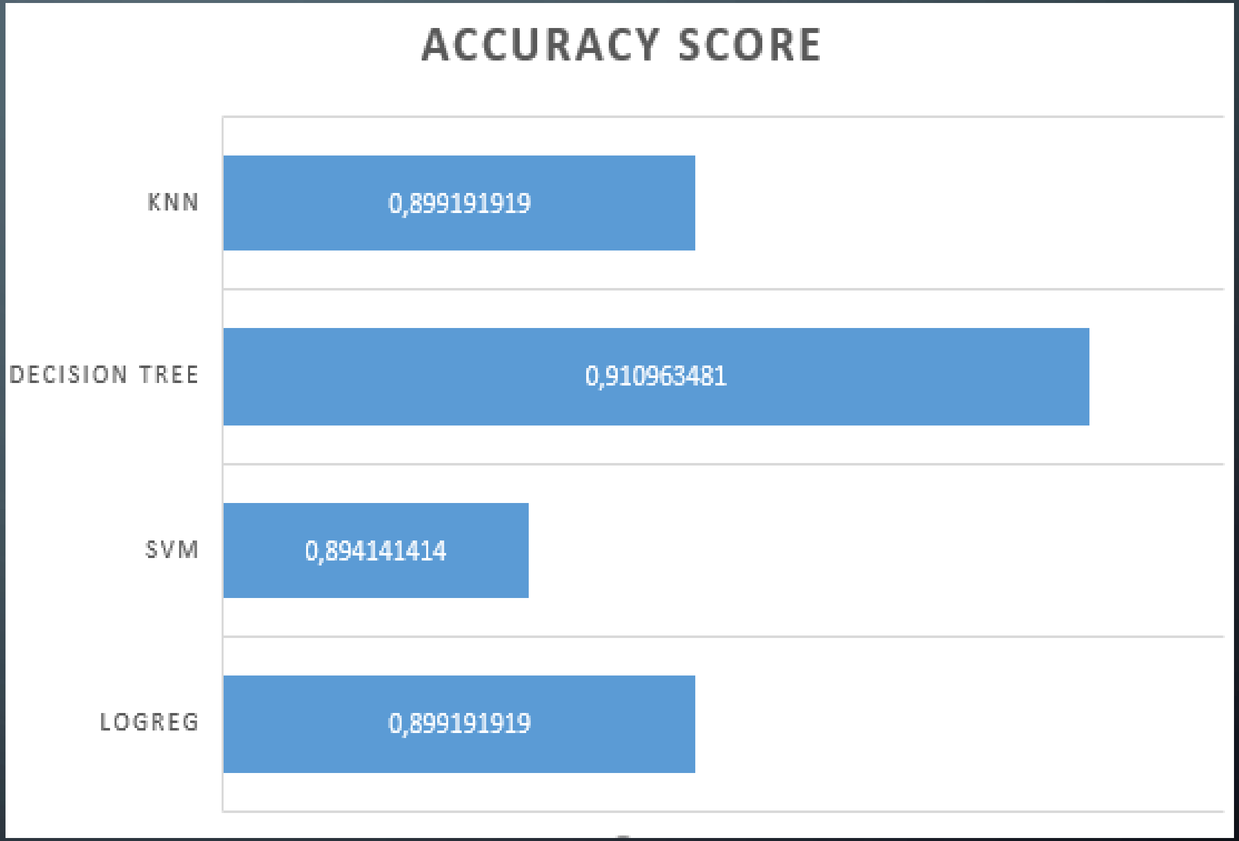
# PREDICTIVE ANALYSIS (CLASSIFICATION)

IN THIS SECTION, WE EXPOSE THE RESULTS OF OUR PREDICTIVE ANALYSIS, BASED ON “TRAIN-TEST” ITERATIONS USING DECISION TREE, KNN, SUPPORT VECTOR MACHINE AND LOGISTIC REGRESSION METHODS FOR MACHINE LEARNING.

# CLASSIFICATION ACCURACY

From the initial sample of 90 observations, we made a training set of 72 observations and a testing set of 18.

Then using a GridSearchCV object for regression, we calculated each model accuracy, reaching the conclusion that the Decision Tree model is the best fit to predict the landing success rate of spaceX rockets, with **91% of accuracy**.



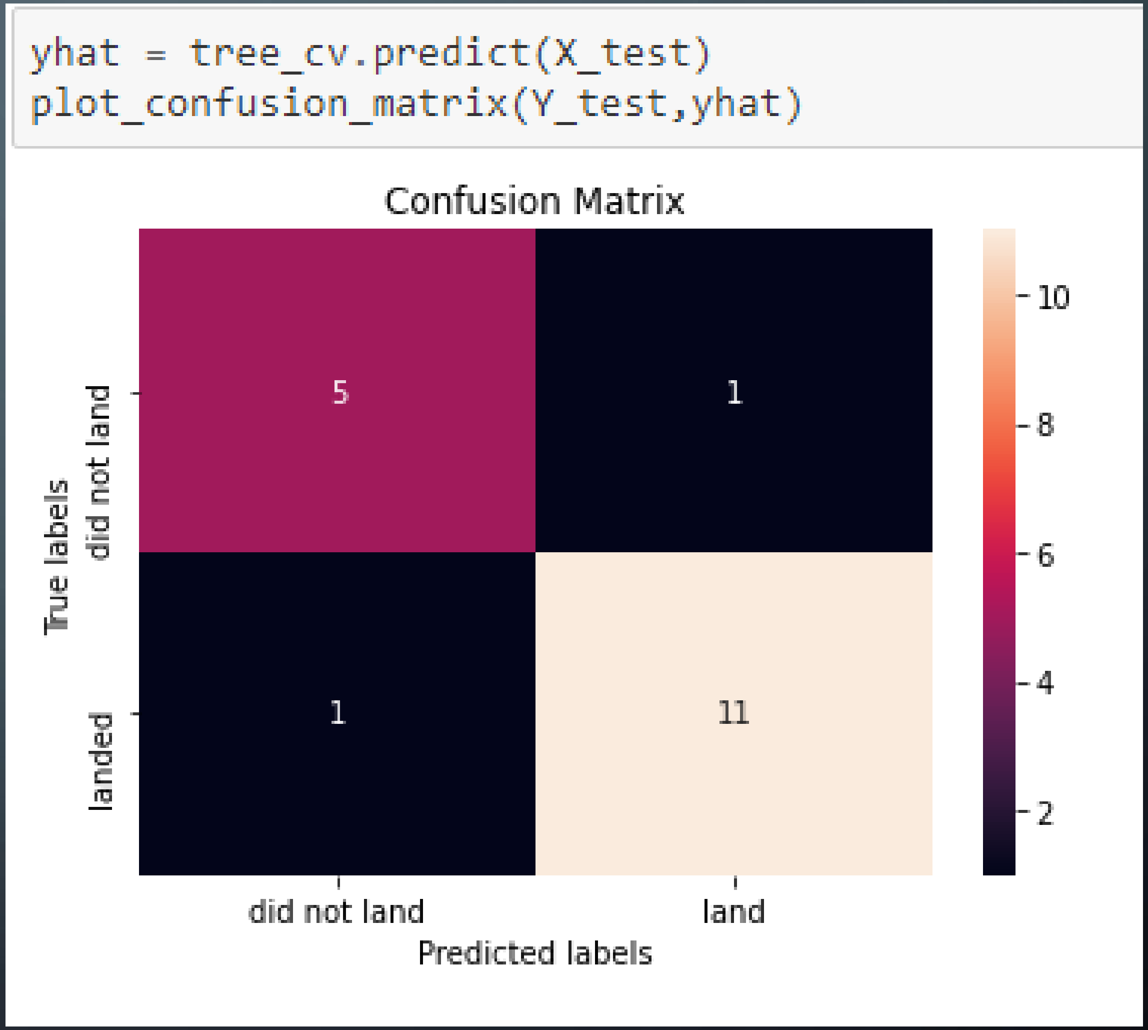
*NB: the results displayed in this chart might differ from the actual notebook, as we ran the model several times, which lead to lightly different accuracy scores. But the trend does not change.*

# CONFUSION MATRIX

Here is the confusion matrix plot of the Decision Tree model.

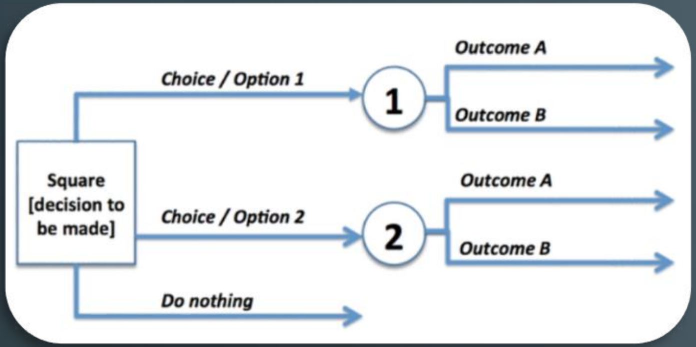
Over the testing sample of 18, our model predicted that 12 successfully land while 6 other do not.

In reality, 11 predictions over the 12 success are correct, while the we were correct for the unsuccessful landings: 6 rockets did not land in reality.



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

# CONCLUSIONS



- Based on the data provided, we found that the optimum cost is reached when we launch from KSC LC-39A site, using the F9_B5 booster model

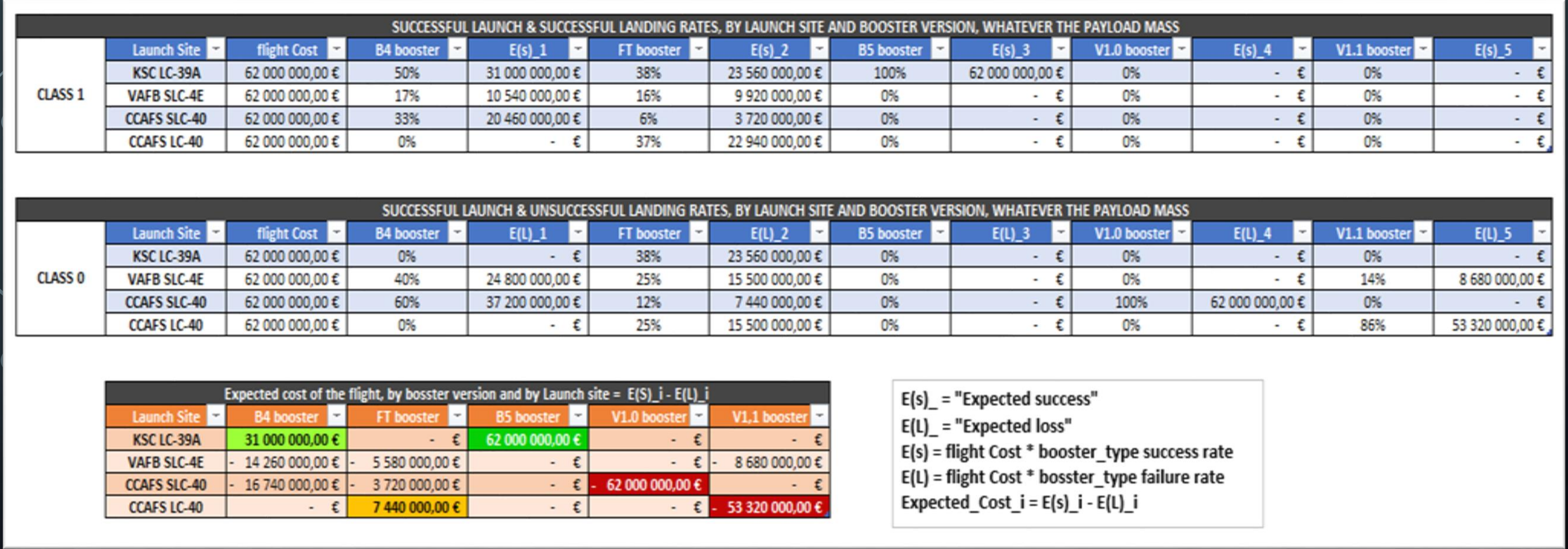
- The F9_B4 and F9_FT boosters models are the most performant in terms of success rate and payload carrying

- For Payload Mass < or = 6000 Kg, KSC LC-39A and CCAFS LC-40 sites share the same landing success rate of 43% when using a F9_FT boosters. But when it comes to F9_B4 boosters, KSC LC-39A success rate rises up to 60% against 40% for CCAFS SLC-40 launch site, this time

- For payloads heavier than 6000 Kg, we better launch our payloads from VAFB SLC-4E site and rely on F9_FT type boosters. We'll likely have a 100% success rate in this configuration, according to the provided data

- The more the launch sites are far from cities and near the coast, the better for our launching programs safety

# CONCLUSIONS



SUCCESSFUL LAUNCH & SUCCESSFUL LANDING RATES, BY LAUNCH SITE AND BOOSTER VERSION, WHATEVER THE PAYLOAD MASS

|  | Launch Site | flight Cost | B4 booster | E(s)_1 | FT booster | E(s)_2 | B5 booster | E(s)_3 | V1.0 booster | E(s)_4 | V1.1 booster | E(s)_5 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CLASS 1 | KSC LC-39A | 62 000 000,00 € | 50% | 31 000 000,00 € | 38% | 23 560 000,00 € | 100% | 62 000 000,00 € | 0% | - € | 0% | - € |
|  | VAFB SLC-4E | 62 000 000,00 € | 17% | 10 540 000,00 € | 16% | 9 920 000,00 € | 0% | - € | 0% | - € | 0% | - € |
|  | CCAFS SLC-40 | 62 000 000,00 € | 33% | 20 460 000,00 € | 6% | 3 720 000,00 € | 0% | - € | 0% | - € | 0% | - € |
|  | CCAFS LC-40 | 62 000 000,00 € | 0% | - € | 37% | 22 940 000,00 € | 0% | - € | 0% | - € | 0% | - € |

SUCCESSFUL LAUNCH & UNSUCCESSFUL LANDING RATES, BY LAUNCH SITE AND BOOSTER VERSION, WHATEVER THE PAYLOAD MASS

|  | Launch Site | flight Cost | B4 booster | E(L)_1 | FT booster | E(L)_2 | B5 booster | E(L)_3 | V1.0 booster | E(L)_4 | V1.1 booster | E(L)_5 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CLASS 0 | KSC LC-39A | 62 000 000,00 € | 0% | - € | 38% | 23 560 000,00 € | 0% | - € | 0% | - € | 0% | - € |
|  | VAFB SLC-4E | 62 000 000,00 € | 40% | 24 800 000,00 € | 25% | 15 500 000,00 € | 0% | - € | 0% | - € | 14% | 8 680 000,00 € |
|  | CCAFS SLC-40 | 62 000 000,00 € | 60% | 37 200 000,00 € | 12% | 7 440 000,00 € | 0% | - € | 100% | 62 000 000,00 € | 0% | - € |
|  | CCAFS LC-40 | 62 000 000,00 € | 0% | - € | 25% | 15 500 000,00 € | 0% | - € | 0% | - € | 86% | 53 320 000,00 € |

Expected cost of the flight, by bosster version and by Launch site = E(S)_i - E(L)_i

| Launch Site | B4 booster | FT booster | B5 booster | V1.0 booster | V1,1 booster |
|---|---|---|---|---|---|
| KSC LC-39A | 31 000 000,00 € | - € | 62 000 000,00 € | - € | - € |
| VAFB SLC-4E | - 14 260 000,00 € | - 5 580 000,00 € | - € | - € | - 8 680 000,00 € |
| CCAFS SLC-40 | - 16 740 000,00 € | - 3 720 000,00 € | - € | - 62 000 000,00 € | - € |
| CCAFS LC-40 | - € | 7 440 000,00 € | - € | - € | - 53 320 000,00 € |

E(s)_ = "Expected success"
E(L)_ = "Expected loss"
E(s) = flight Cost * booster_type success rate
E(L) = flight Cost * bosster_type failure rate
Expected_Cost_i = E(s)_i - E(L)_i

Calculation table of the expected flight cost for spaceX, by outcome class, Launch site, and booster type, regardless of the payload mass and the orbit type.

# APPENDIX

- Machine learning prediction notebook: https://github.com/Henri-Saub/Capstone-project-DS-ML/blob/4f5844edf3a1ad2668dc1cc575b677721b8d4ae2/SpaceX_Machine%20Learning%20Prediction_Part_5_final.ipynb
- Capstone project processing and visualization on Excel : https://github.com/Henri-Saub/Capstone-project-DS-ML/blob/69d8227f1a91660864e470d981daae6c587601ef/Spacex%20(Autosaved).xlsx