IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Daniel Ramirez
13-Nov-2022

Outline

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

Executive Summary

- For this development we have utilized API and webscraping to collect data about the Space X launching results, then we perform an exploratory and visual analysis of said data using SQL and Folium to generate an interactive dashboard
- Then we perform a predictive analysis using four classification models to help determine which decisions will have a better outcome.
- At the end of the work we have determined that one site have an outstanding success landing of 79%, that a VLEO orbit has the most recent success and support, that the best model to analyze these data are the decision tree and that the heavier payload mass have a greater possibility of being successful.

Introduction

- **Project background and context**

In the beginning of a new space race, now fueled by Private Public Partnerships, the space exploration now requires more data than ever, the risks involved in the launch of each rocket makes that every piece of information very valuable and mistakes are carefully studied to create better and more efficient rockets.

- **Problems you want to find answers**

In this context, we're going to evaluate the data containing information about the recent launchings of Space X to determine the best conditions of a successful launching to replicate them in our own company.

Section 1

Methodology

Methodology

Executive Summary

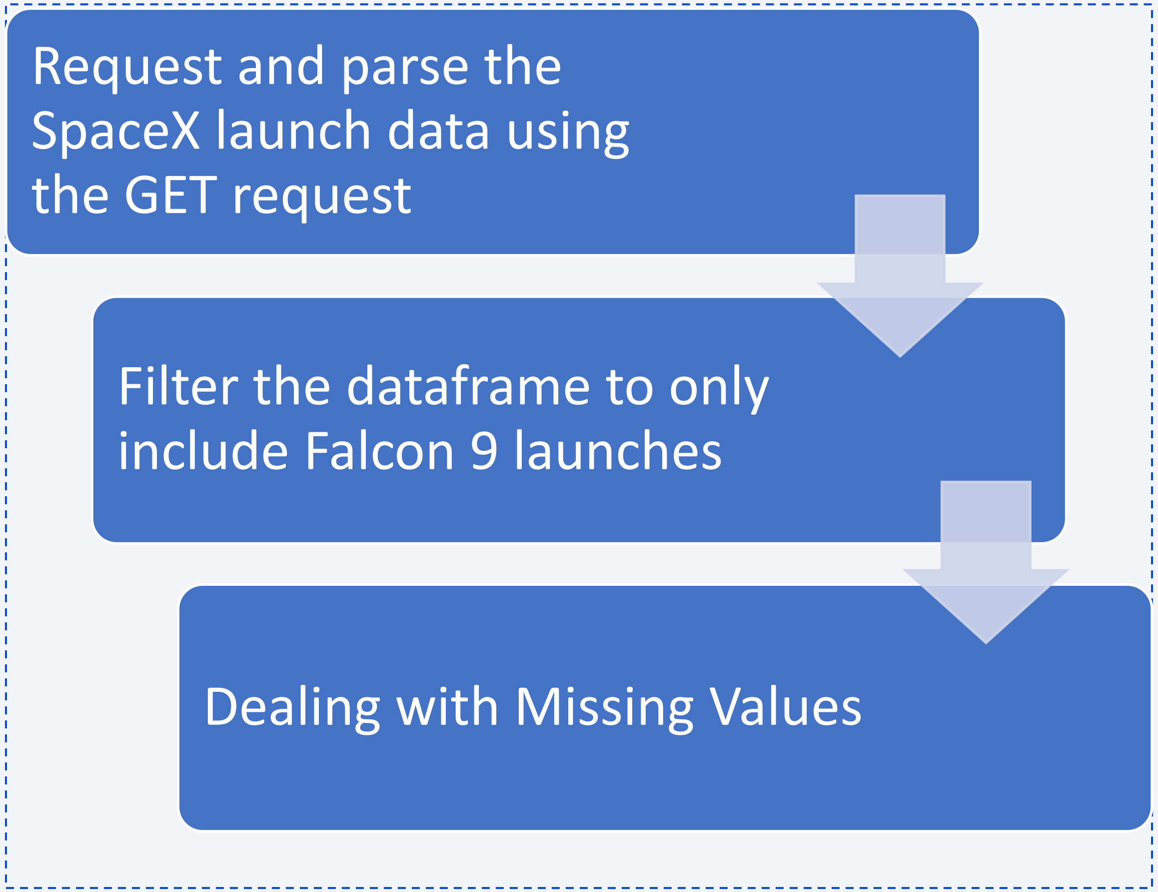
- Data collection methodology:
  - Data was obtained form two sources SpaceX API and Wikipedia webscraping
- Perform data wrangling
  - We deal with missing values, normalize and summarize the data
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - We use and evaluate four classification models

Data Collection

- The data were collected from the Space X API ("https://api.spacexdata.com/v4/launches/past") and Wikipedia (https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches)

Data Collection – SpaceX API

- Space X offers a public API where data can be obtained and used to make tests.
- The data was used according to the next flowchart
- Source code: https://github.com/dramirezfigueredo/Python-Data-Science-Capstone/blob/master/jupyter-labs-spacex-data-collection-api.ipynb
- Note: if the notebook is not opening at first, please refresh the page.

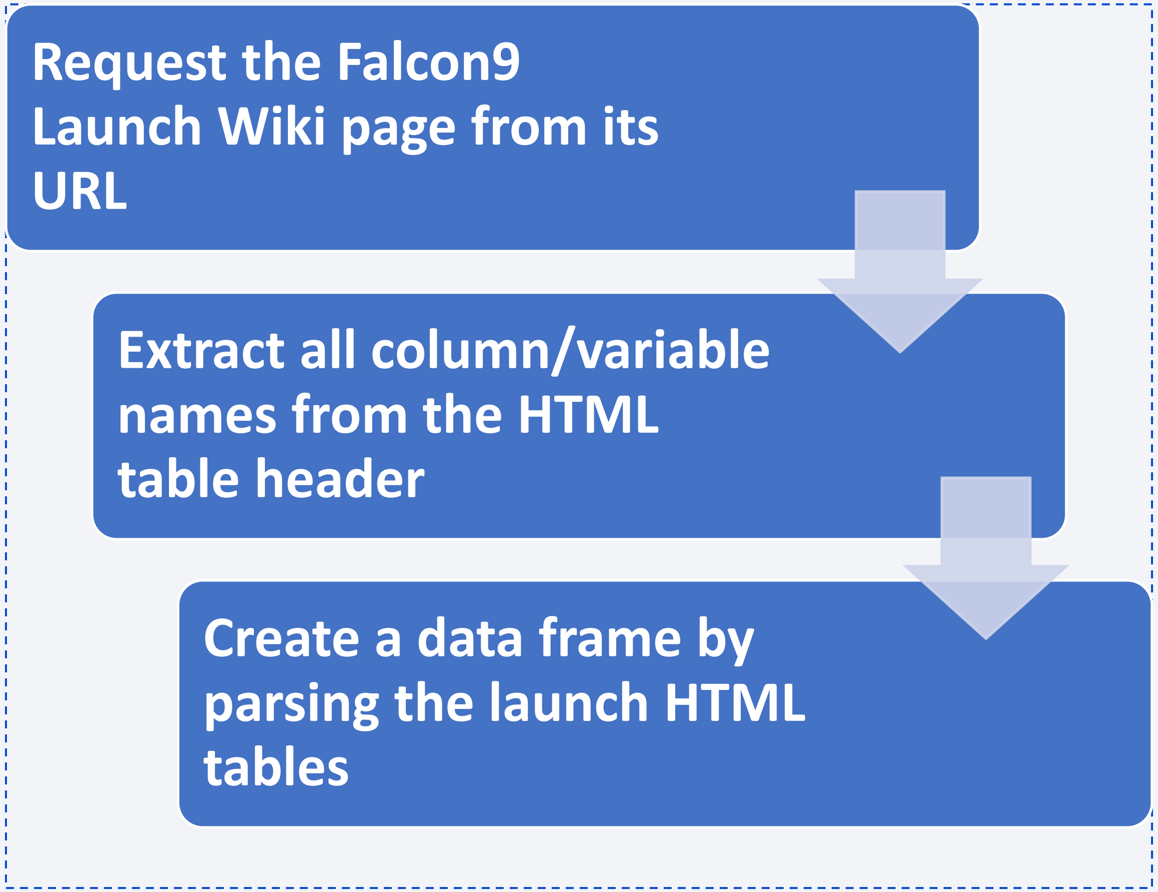


collect Falcon 9 historical launch records from a Wikipedia page titled List of Falcon 9 and Falcon Heavy launches

Data Collection - Scraping

- We collect Falcon 9 historical launch records from a Wikipedia page titled List of Falcon 9 and Falcon Heavy launches using websacrpping

- Source code: https://github.com/dramirezfigueredo/Python-Data-Science-Capstone/blob/master/jupyter-labs-webscraping.ipynb



Data Wrangling

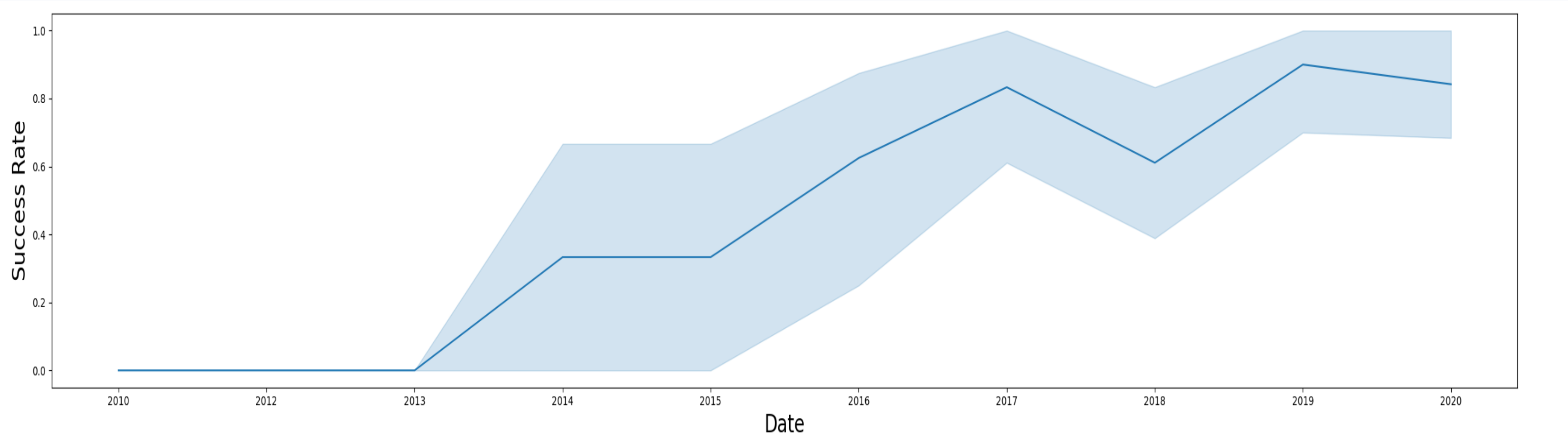
- We perform some Exploratory Data Analysis (EDA) to find some patterns in the data and determine what would be the label for training supervised models
- Then we summarized information about launches, orbits, occurrence missions and the outcome of each mission.
- The process is shown a continuation:

Calculate the number of launches on each site → Calculate the number and occurrence of each orbit → Calculate the number and occurence of mission outcome per orbit type → Create a landing outcome label from Outcome column

- Source code: https://github.com/dramirezfigueredo/Python-Data-Science-Capstone/blob/master/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- To explore data, we used scatterplots, barplots and a line graph to explain the different relationships between Flight Number and Launch Site, Flight Number and Orbit type, Payload and Launch Site, Payload and Orbit type, the success rate of each orbit type and the launch success yearly trend



- Source code: https://github.com/dramirezfigueredo/Python-Data-Science-Capstone/blob/master/module_2_jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

EDA with SQL

- The following the SQL queries were performed:

  - Display the names of the unique launch sites in the space mission
  - Display 5 records where launch sites begin with the string 'CCA'
  - Display the total payload mass carried by boosters launched by NASA (CRS)
  - Display average payload mass carried by booster version F9 v1.1
  - List the date when the first successful landing outcome in ground pad was achieved.
  - 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 successful landing outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

- Source code: https://github.com/dramirezfigueredo/Python-Data-Science-Capstone/blob/master/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Markers, circles and lines were created and added to a folium map to indicate the distance of landmarks to the launching sites, this is useful to determine the potential impact over communities in case of failure and at the same time, the minimum distance at which a successful launching affects the community.
- Marker clusters were used to indicate the launching sites and their performance.

- Source code: https://github.com/dramirezfigueredo/Python-Data-Science-Capstone/blob/master/module_3_lab_jupyter_launch_site_location.jupyterlite.ipynb

Build a Dashboard with Plotly Dash

- We have added a Success Launches by site, a payload range and a relation between the booster version, the payload mass and the success of the launching to the dashboard
- This will facilitate identify the success of a launching starting with the launching site and the mass of the rocket

- Source code: https://github.com/dramirezfigueredo/Python-Data-Science-Capstone/blob/master/spacex_dash_app.py

Predictive Analysis (Classification)

- We compared four classification models: Logistic regression, SVC, decision tree and KNN

Perform exploratory Data Analysis and determine Training Labels

Find best Hyperparameter for SVM, Classification Trees and Logistic Regression

Find the method performs best using test data

- Source code: https://github.com/dramirezfigueredo/Python-Data-Science-Capstone/blob/master/module_4_SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

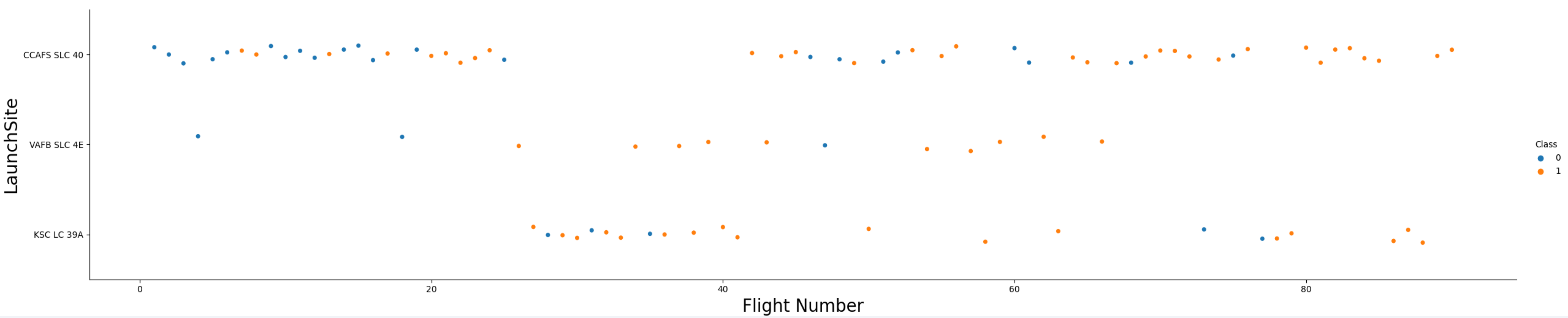
Results

- The payload mass launches over 9000 Kg have a great possibility of success.
- The VLEO orbits have gotten better with time and are being used more often, opening a business possibility
- With each launching the possibility of success is increased.
- Decision tree shows the strongest predictive analysis between the models utilized with an accuracy of 87%

Section 2

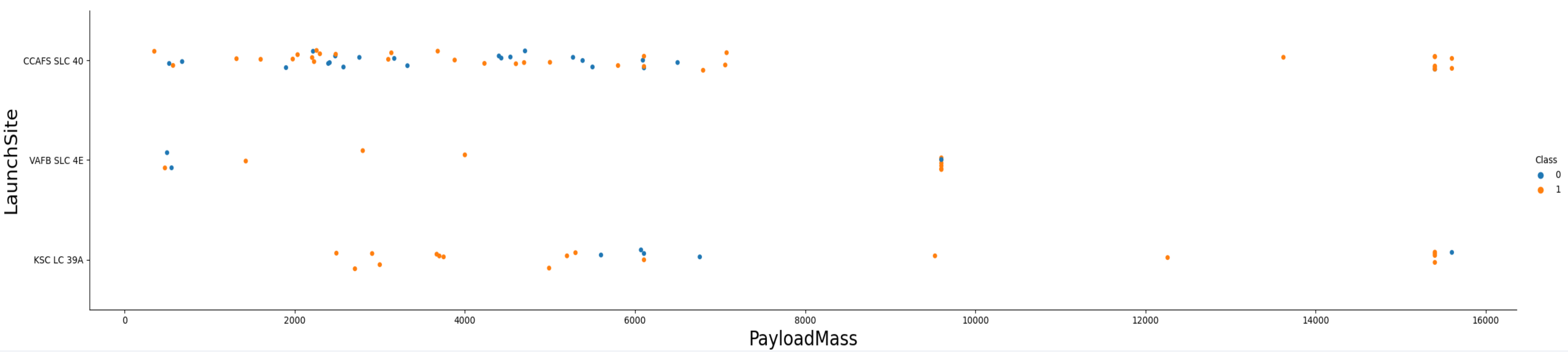
Insights drawn from EDA

Flight Number vs. Launch Site



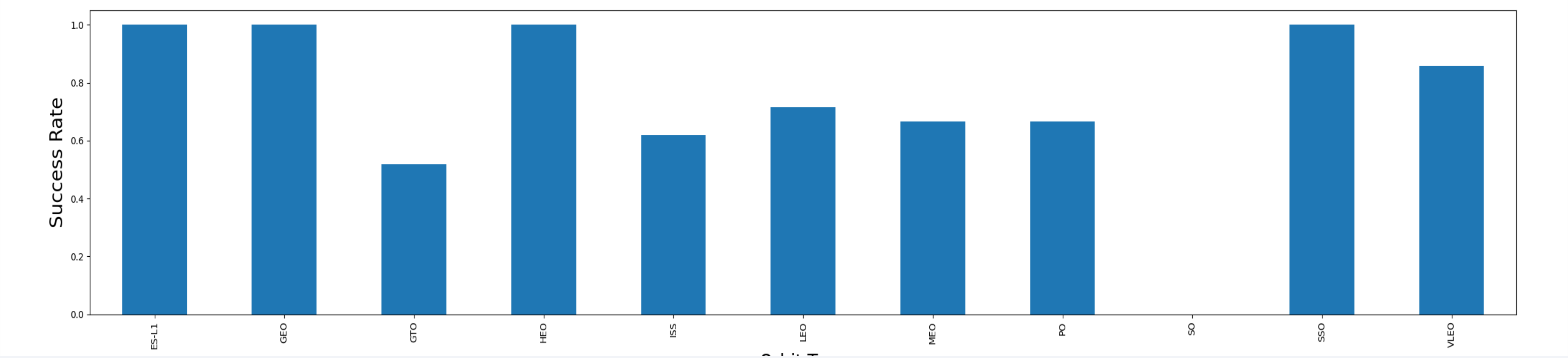
- It's possible to say that the launching success has improved with time, CCAFS SLC 40 is not only the site with the most successful landings but also the site that where the majority of launches has occurred, This could be explained by an experienced crew and is getting better results with time..
- VAFB SLC 4E has the second place in successful launches but has passed a while since the last launching occurred in the site and KSC LC 39A is the third more successful launching site.

Payload vs. Launch Site



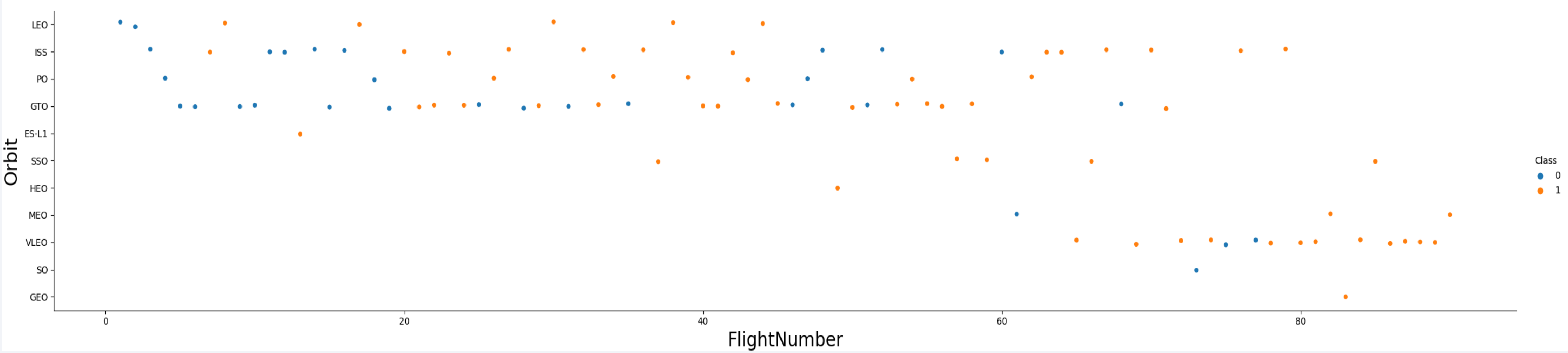
- Rockets with a payload mass over 9.000 Kg have a great success rate.
- Also, is notable that looks like you can only send rockets over 10.000 Kg from KSC LC 39A and CCAFS SLC 40, which could explain why VAFB SLC 4E has not been used in a while.

Success Rate vs. Orbit Type



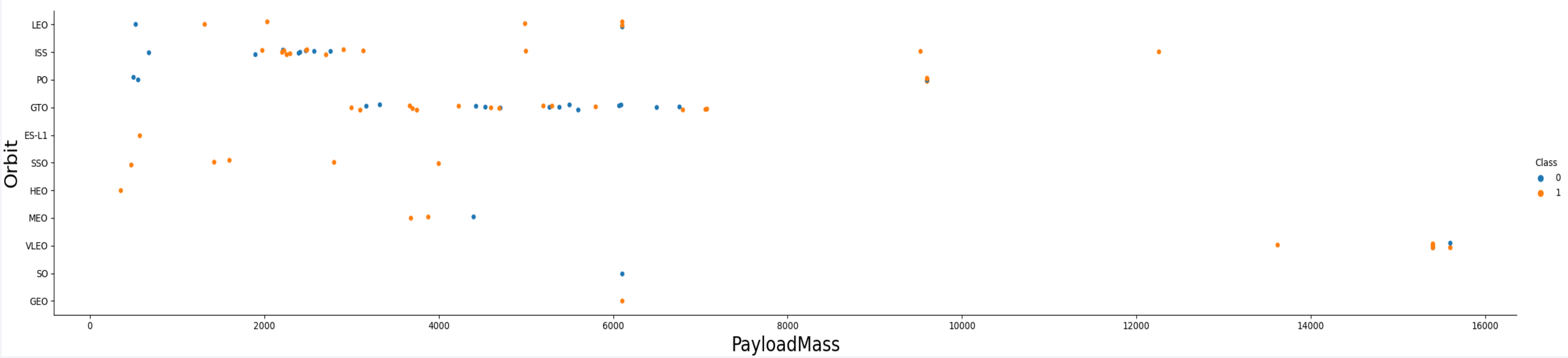
- Rockets that follows the highlly elliptycal orbit (HEO), the geostationary orbit (GEO) and the sun synchronous orbit (SSO), and the ES-L1 (Lagragian Point) have an excellent performance
- Low and medium orbits are not as succesful; but the very low earth orbit (VLEO) rockets have a great performance.

Flight Number vs. Orbit Type



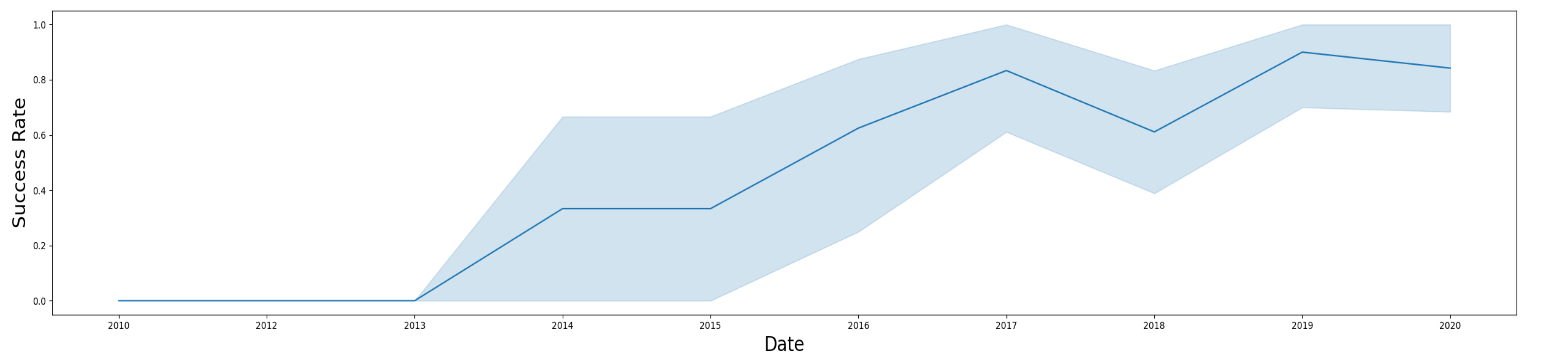
- There is a recent interest in the use of the Very Low Earth Orbits (VLEO), so there could be a business opportunity in that field.
- As we have already seen, with more flights the success rate has increased.

Payload vs. Orbit Type



- A rocket with a Payload Mass over 8.000 Kg has a great chance of success, independient of the orbit that would follow.
- The Very low Earth Orbit (VLEO) that has received a recent interest allows the launching of this heavy objects.

Launch Success Yearly Trend



- With more time, the results have been better, and now the succes rate overall is over 80%

All Launch Site Names

Launch Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40

- There are four launch sites.
- We used a query that takes every individual launch site from the data.

Launch Site Names Begin with 'CCA'

In [8]:

```
%sql select * from SPACEXDATASET where Launch_site like '%CCA%' LIMIT 5
```

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

Out[8]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS_KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

- We can see five examples of launches from the site that begins with “CCA”.

Total Payload Mass

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

In [9]:

```
%sql SELECT SUM("PAYLOAD_MASS__KG_") FROM SPACEXDATASET WHERE Customer = "NASA (CRS)"
```

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

Out[9]:

SUM("PAYLOAD_MASS__KG_")
45596

- All the boosters where the costumer has been NASA weight 45.596 Kg

Average Payload Mass by F9 v1.1

In [10]:
```
%sql SELECT AVG("PAYLOAD_MASS__KG_") FROM SPACEXDATASET WHERE Booster_version like '%F9 v1.1%'
```

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

Out[10]:

AVG("PAYLOAD_MASS__KG_")
2534.6666666666665

- The average payload mass carried by booster version F9 v1.1 is 2534, 66 Kg

First Successful Ground Landing Date

Hint:Use min function

In [27]:

```
# given that sql lite can't identify the format DD-MM-YYYY as a date, the normal min formule won't work
# %sql SELECT MIN("Date") FROM SPACEXDATASET WHERE "Landing _Outcome" = 'Success (ground pad)'
# so, we will have to do a bit of substr and limit 1

%sql SELECT date from SPACEXDATASET where substr(Date,7,4)||'-'||substr(Date,4,2)||'-'||substr(Date,1,2) between '2010-06-03' AND '2020-12-31' and "Landing _Outcome" = 'Success (ground pad)'
```

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

Out[27]:

Date
22-12-2015

- Normally, you would use the select min date query, but I have been using the skill network notebook and sql lite doesn't recognize the date format in the .csv file that uses, so I had to do a bit more but the date of the first successful landing in ground pad is 22-12-2015

Successful Drone Ship Landing with Payload between 4000 and 6000

In [29]:
```
%sql select Booster_version from SPACEXDATASET where "Landing _Outcome" = 'Success (drone ship)' and "PAYLOAD_MASS__KG_">4000 and "PAYLOAD_MASS__KG_"<6000
```

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

Out[29]:

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

- There are four boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

Total Number of Successful and Failure Mission Outcomes

```
%sql SELECT mission_outcome, count(*) as Count FROM SPACEXDATASET GROUP by mission_outcome ORDER BY mission_outcome
```

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

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

- There have been only one mission that has been classified as failure in the whole process, the other 100 have been successful in fulfilling their objectives.

Boosters Carried Maximum Payload

```
%sql SELECT Booster_Version, Payload_Mass__KG_ FROM SPACEXDATASET WHERE Payload_Mass__KG_ = (SELECT MAX(Payload_Mass__KG_) FROM SPACEXDATASET) ORDER BY I
```

* sqlite:///my_data1.db
Done.

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

- Twelve boosters have carried the maximum payload mass: 15600 Kg

2015 Launch Records

```
%sql SELECT Booster_Version, Launch_Site, "Landing _Outcome", substr(Date, 4, 2) AS Month, substr(Date, 7,4) AS Year  FROM SPACEXDATASET WHERE "Landing _Outcome" LIKE 'Failure (dr
```

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

Booster_Version	Launch_Site	Landing _Outcome	Month	Year
F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)	01	2015
F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)	04	2015

- There are two boosters that have failed their landing in a drone ship in 2015.

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

```
# As showed before, sql lite doesn't work with the date format of the archive
# So the basic formule won't work

# %sql  SELECT "Landing _Outcome", COUNT("Landing _Outcome") FROM SPACEXDATASET WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY "Landing _Outcome" ORDE

# But we will do a bit of substring

%sql  SELECT "Landing _Outcome", COUNT("Landing _Outcome") FROM SPACEXDATASET WHERE substr(Date,7,4)||'-'||substr(Date,4,2)||'-'||substr(Date,1,2) BETWEEN '2010-06-04' AND '2017-
```

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

Landing _Outcome	COUNT("Landing _Outcome")
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

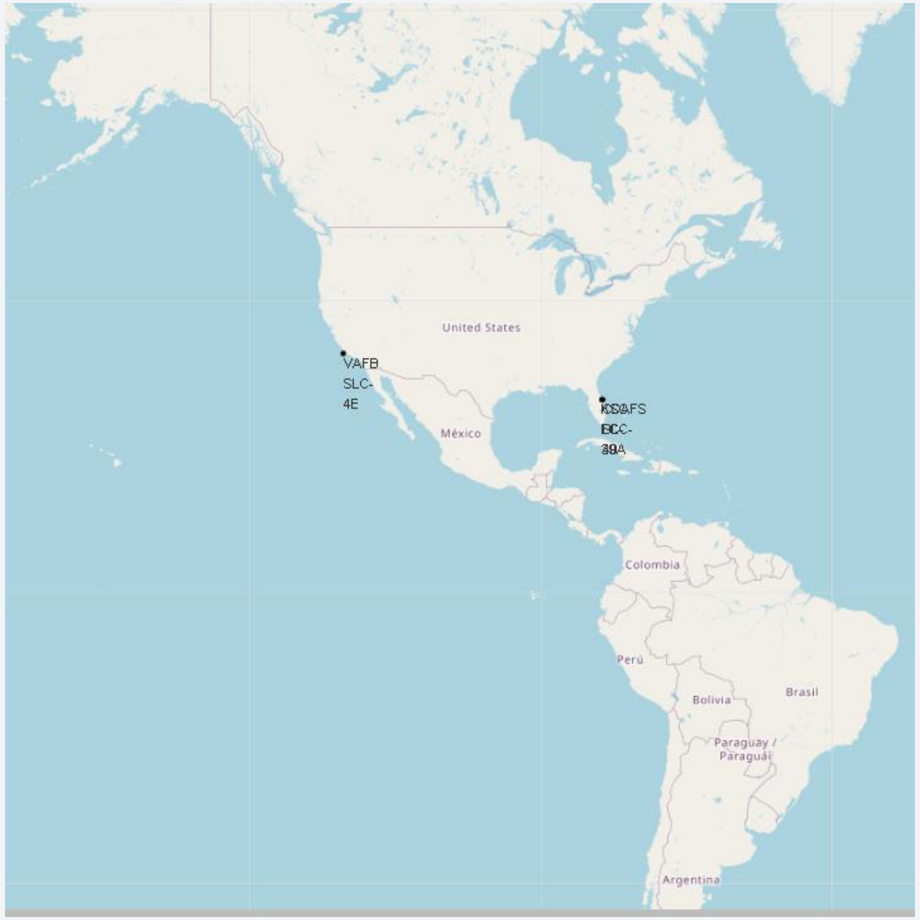
- Between 2010-06-04 and 2017-03-20, there were ten (10) no attempt at landing, followed by five (5) success of landing into a drone ship and five (5) failure at landing into a drone ship

Section 3

Launch Sites Proximities Analysis

Location of lauching sites

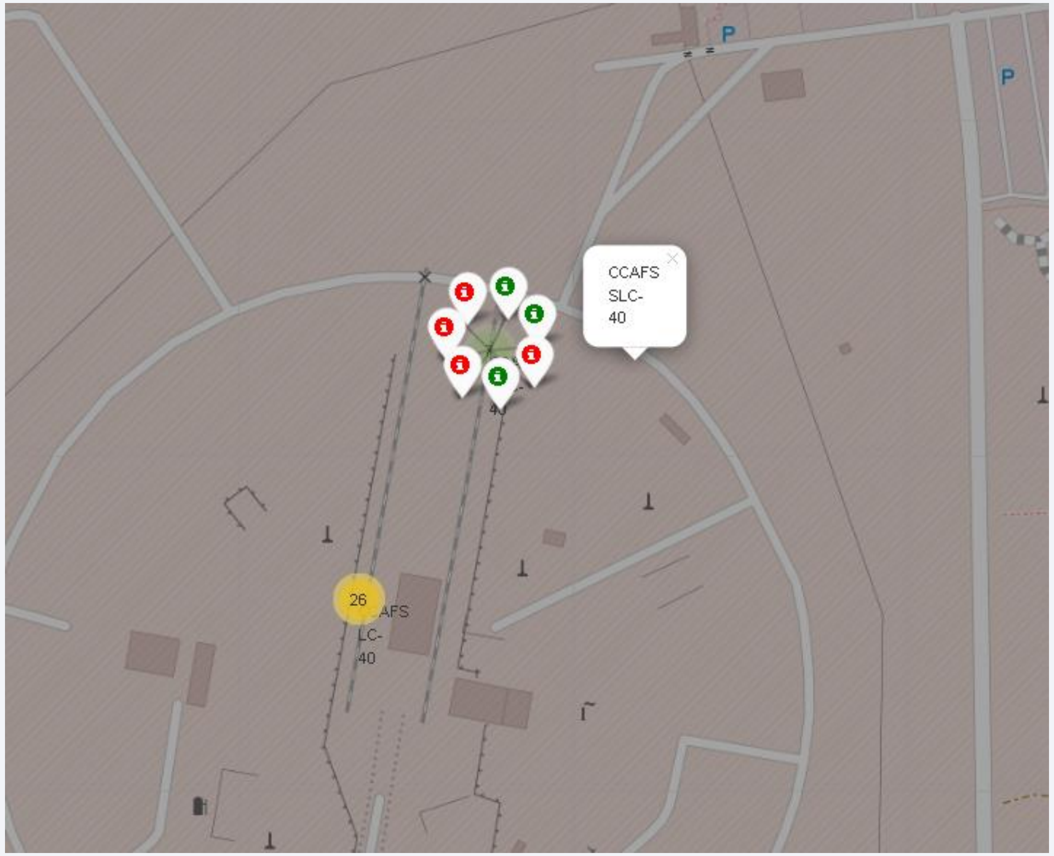
- The location sites are near the coastline, to reduce risks related to populate areas and facilitate the retrieving of ships fallen to sea.
- They are near the equator to have a stable climate through the year.
- At the same time they are located in USA to take advantage of the existent infrastructure.



Launch outcomes in CCAFS SLC-40

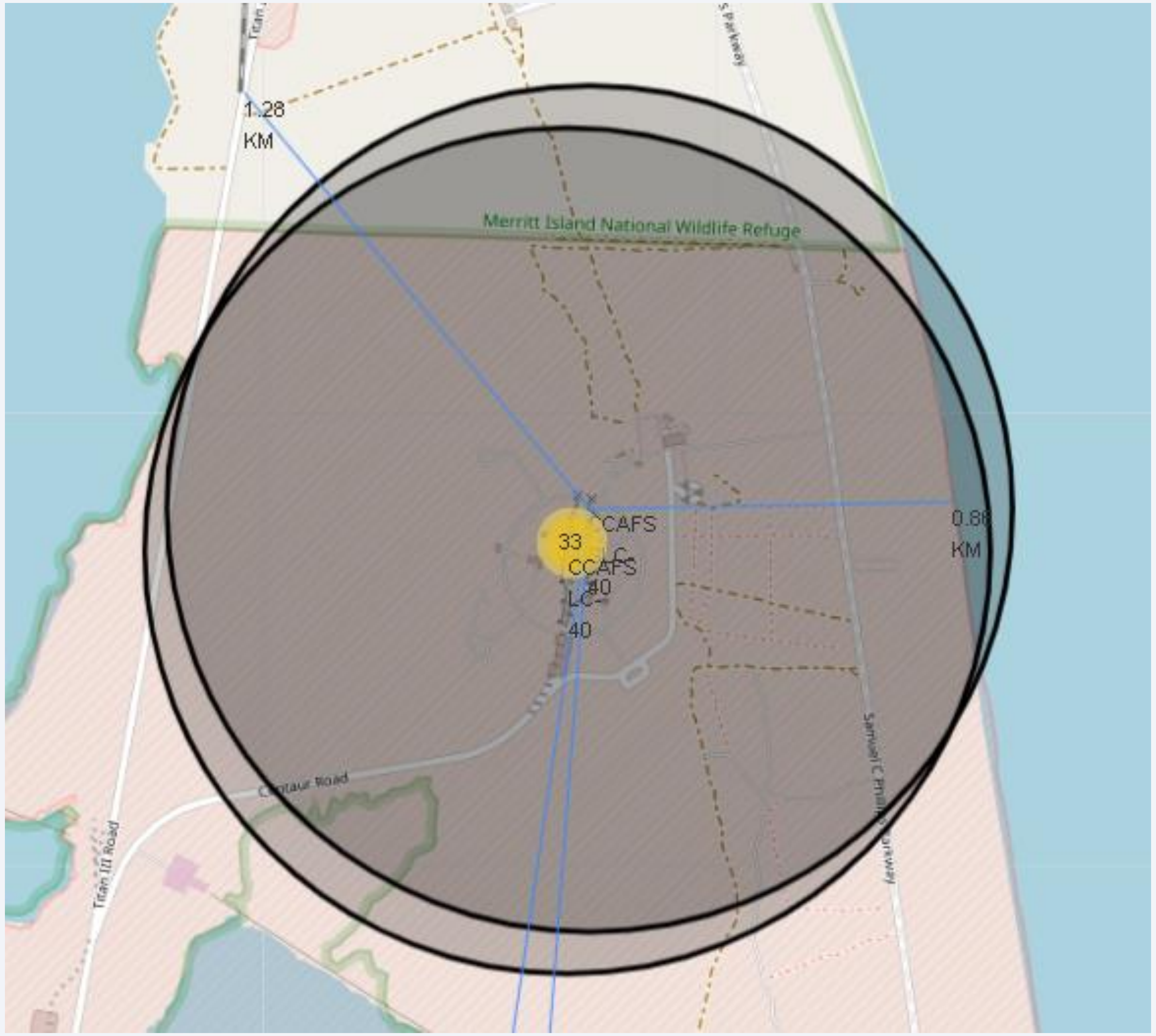
- In the map we can see that from the Cape Canaveral Space Launch Complex 40 (CCAFS SLC-40) have been launched three successful outcomes and four failed outcomes in seven missions.

- At south of the complex you can see the Cape Canaveral Launch Complex 40 (CCAFS LC-40) from where have been launched 26 missions.



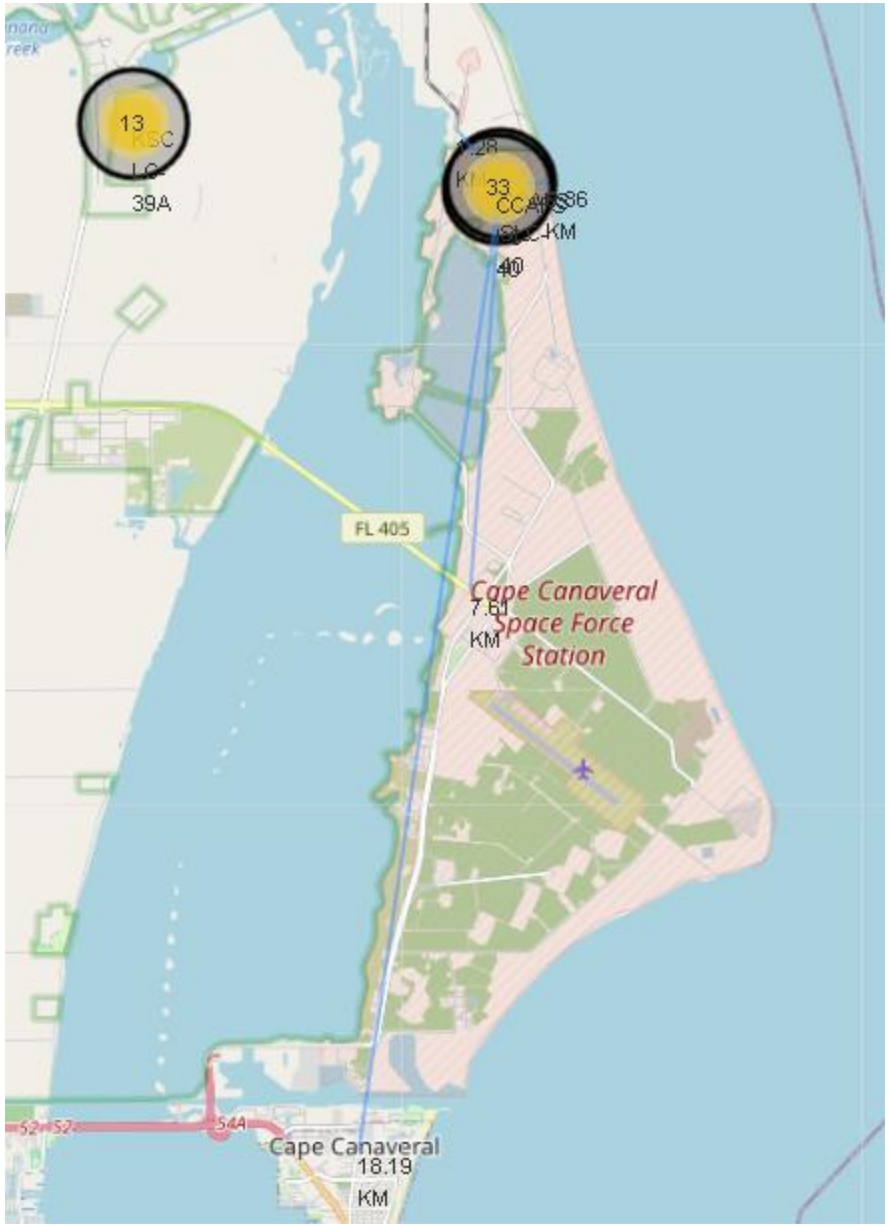
Landmarks near CCAFS SLC-40

- The Cape Canaveral Space Launch Complex 40 (CCAFS SLC-40) is at 1,28 Km from a railway and at 840m from the coastline.



Landmarks near CCAFS SLC-40

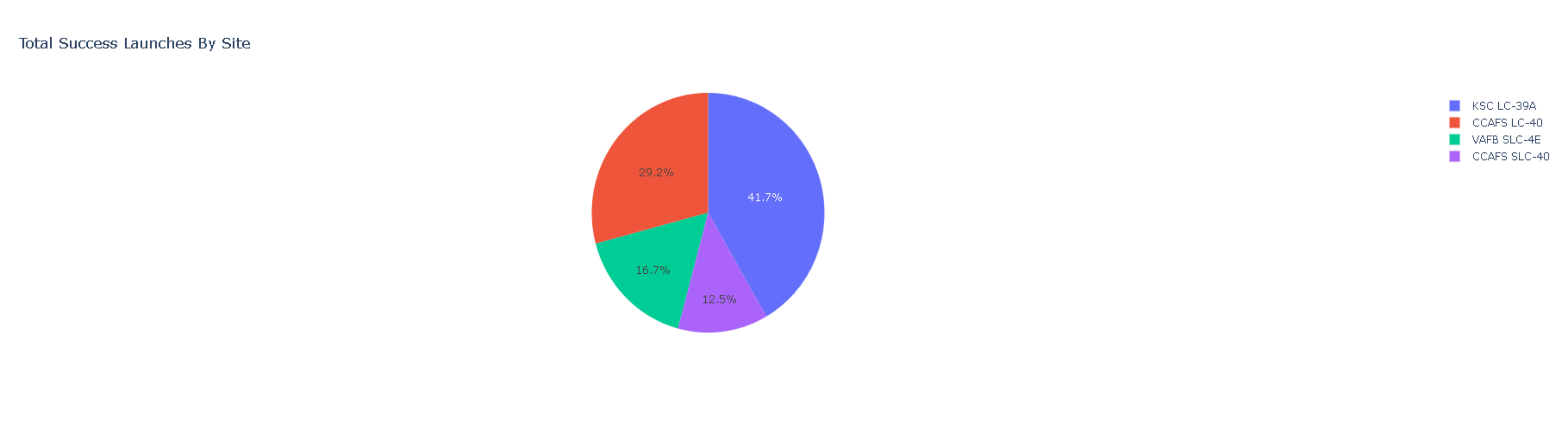
- The Cape Canaveral Space Launch Complex 40 (CCAFS SLC-40) is at 7,61 Km from a highway and at 18.19 Km from Cape Canaveral.



Section 4

Build a Dashboard with Plotly Dash

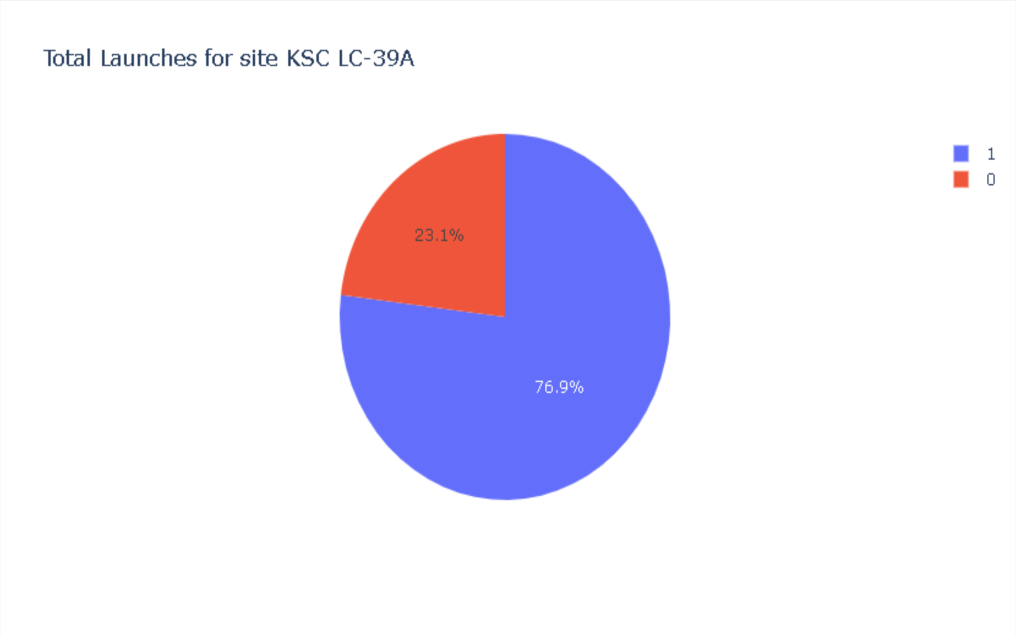
Success Launches by site



- KSC LC 39A accounts for the 41,2% of the successful launches, followed by the Cape Canaveral Launch Complex 40 (CCAFS LC-40) with the 29,2% of the successful missions

KSC LC 39A Launches

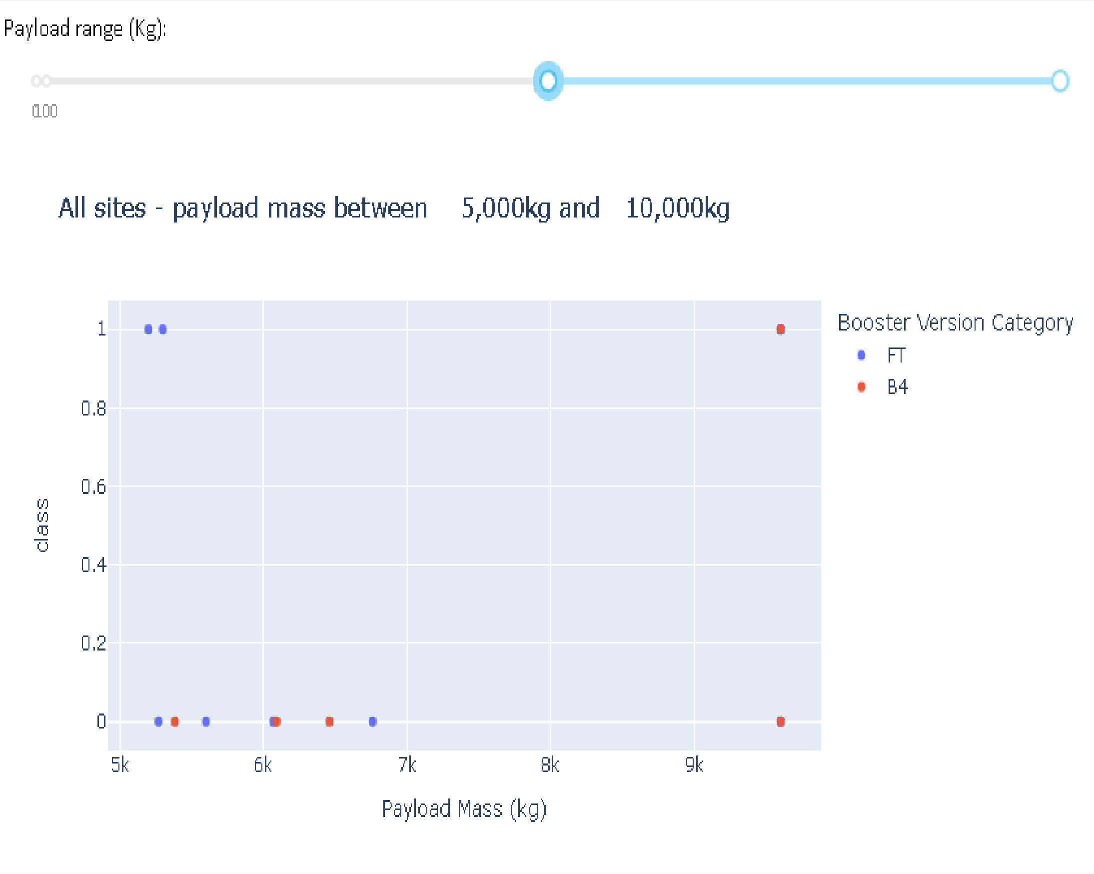
- With 10 of 13 missions successfully landing (76,9%) KSC LC 39A is the most adequate site to attempt a launching.



The Booster FT vs the Booster B4

- Between 5,000 Kg and 10,000 Kg there are just two booster versions: the FT and the B4.

- The FT has shown some success around 5000-5500 Kg, but in every try over 6000 Kg has failed.
- Meanwhile, the B4 is the opposite, his best performance was near the 10,000 Kg but in lesser weight has failed.

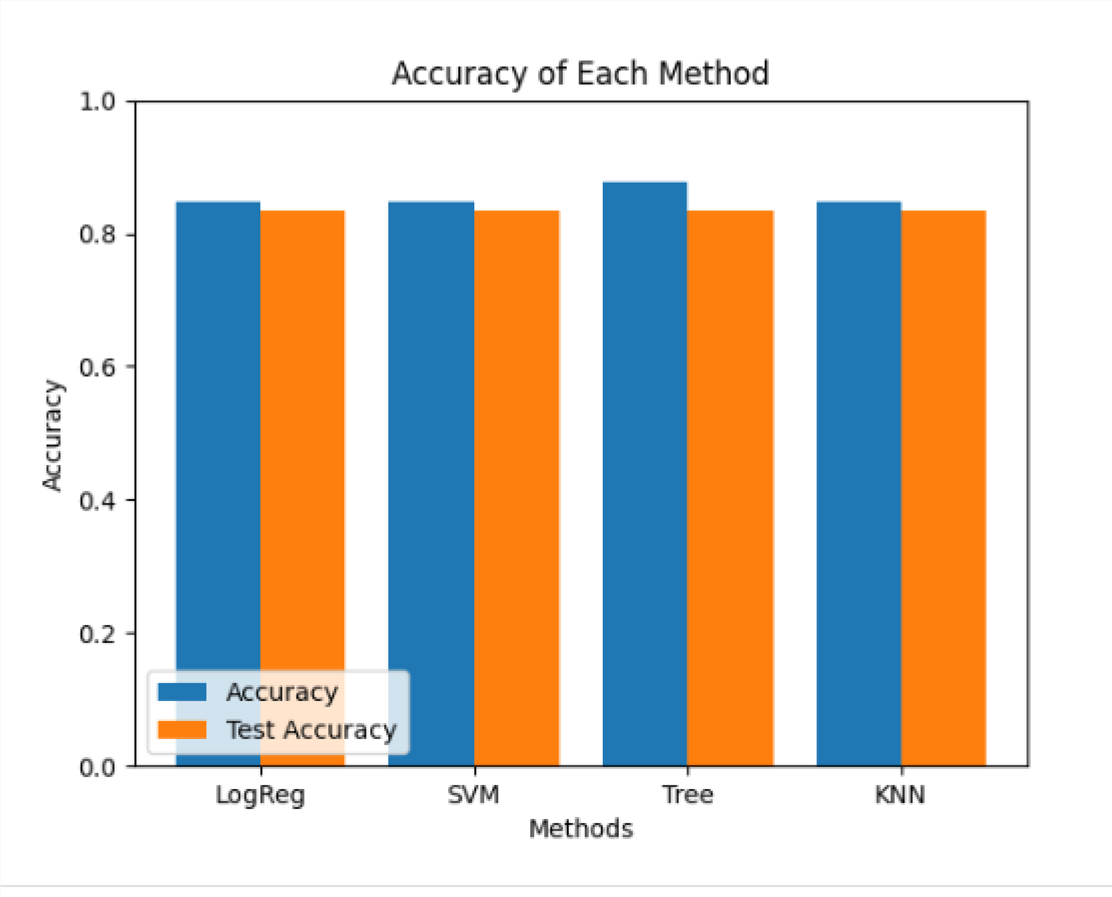


Section 5

Predictive Analysis (Classification)

Classification Accuracy

- Four models were tested: KNN, Decision Tree, SVC and Logistic Regression.
- According to the graph, all models are adequate but Decision Tree is stronger in its results.



Confusion Matrix

- For some reason, the confusion matrix is not working in the Skill Network Labs at the moment.

Conclusions

- The highest successful landing is KSC LC 39A
- If the rocket is over 9000 Kg, the landing could be a success
- The best opportunity at the moment is attempt a launching into the (VLEO) Orbit
- As time passes is safer the attempt on successful landing
- …

Appendix

- Note: This project is different, I have used only the skill labs notebooks so all the databases are into their fixed links.
- The project content are the seven skill labs modules and the spacex_dash_app.
- Everything is included in my github account according to the instructions of the course and I hope the best.

Thank you!