Vibha V.K
19/09/2021

Outline

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

Executive Summary

The goal of the next study is to determine what makes SpaceX launches successful.

For that the methodologies used are exploratory analysis via graphic visualisations or SQL queries. The exploratory data also will be presented in an interactive map and dashboard. Finally, a machine learning will be trained to distinguish between successful and failed launches.

From the study there have been determined that experience and a correct launch site and orbit has been key in order to achieve a high success rate. The results of the ML model are inconclusive, more data is needed for the training.

Introduction

For the last decades, space travel and satellite launches has been an expensive and difficult task. SpaceX has a mission to achieve cheaper space flights via reutilisations of boosters, which it has been a success. This success has make competitors to seek for information in order to take part in this new space race.

The goals of the study is to answer the next questions:

What characteristics seem determinant of a successful launch.

What is the path that SpaceX has follow to achieve their success.

Can success been predicted by historical data.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - A combination of web scrapping and REST API calls have been used to collect the required data
- Perform data wrangling:
  - Data was cleaned by selecting only the relevant information and cleaning empty values.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - The data has been prepared for the training of different ML models in order to predict the outcome of future flights.

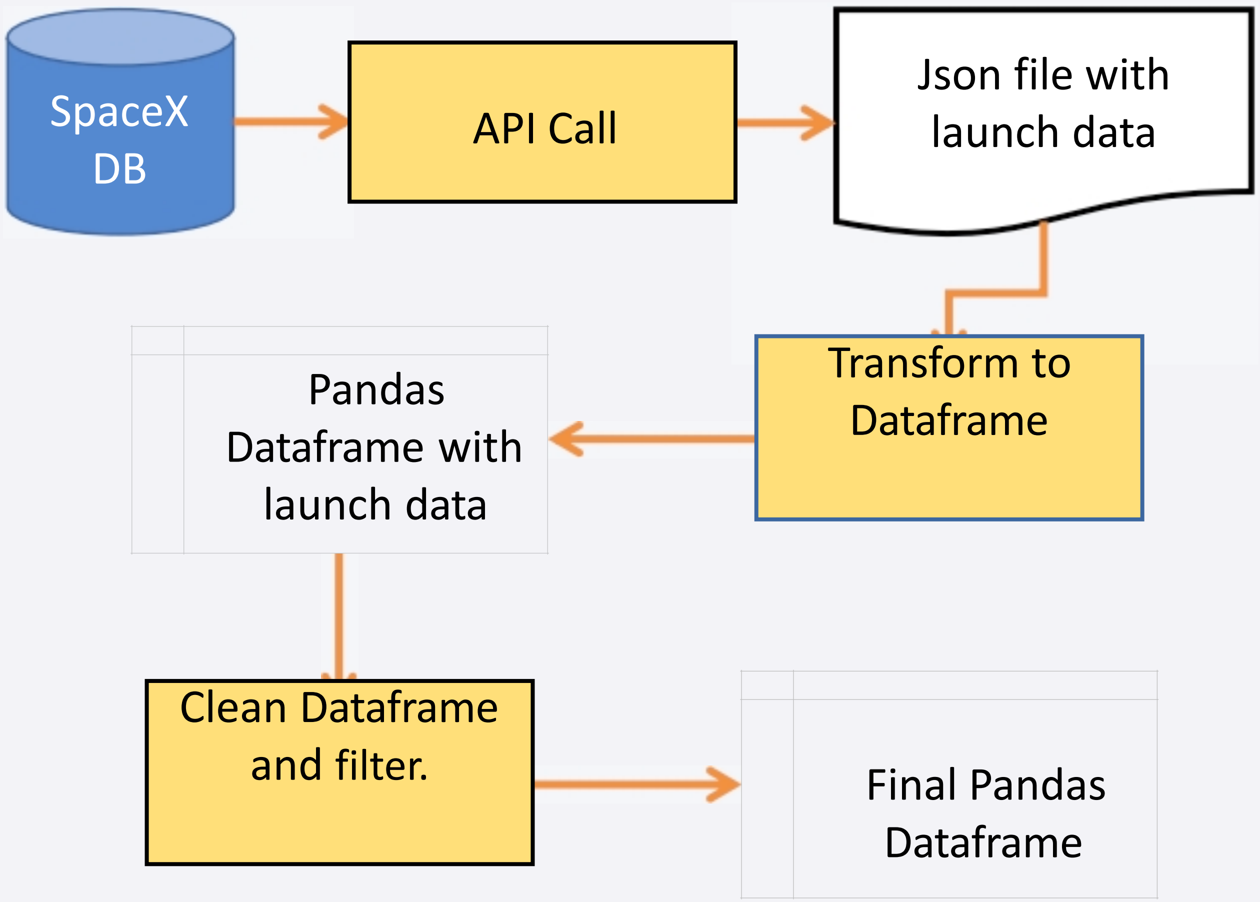
Data Collection

Data for the project will be collected from two sources:

- Directly from SpaceX: This will be achieved via SpaceX's public REST API, which allows the access to information relative to SpaceX spatial missions.
- Web scraping from Wikipedia: Historic information has been introduced in several pages of Wikipedia and it can be extracted in order to complement the information offered by SpaceX.

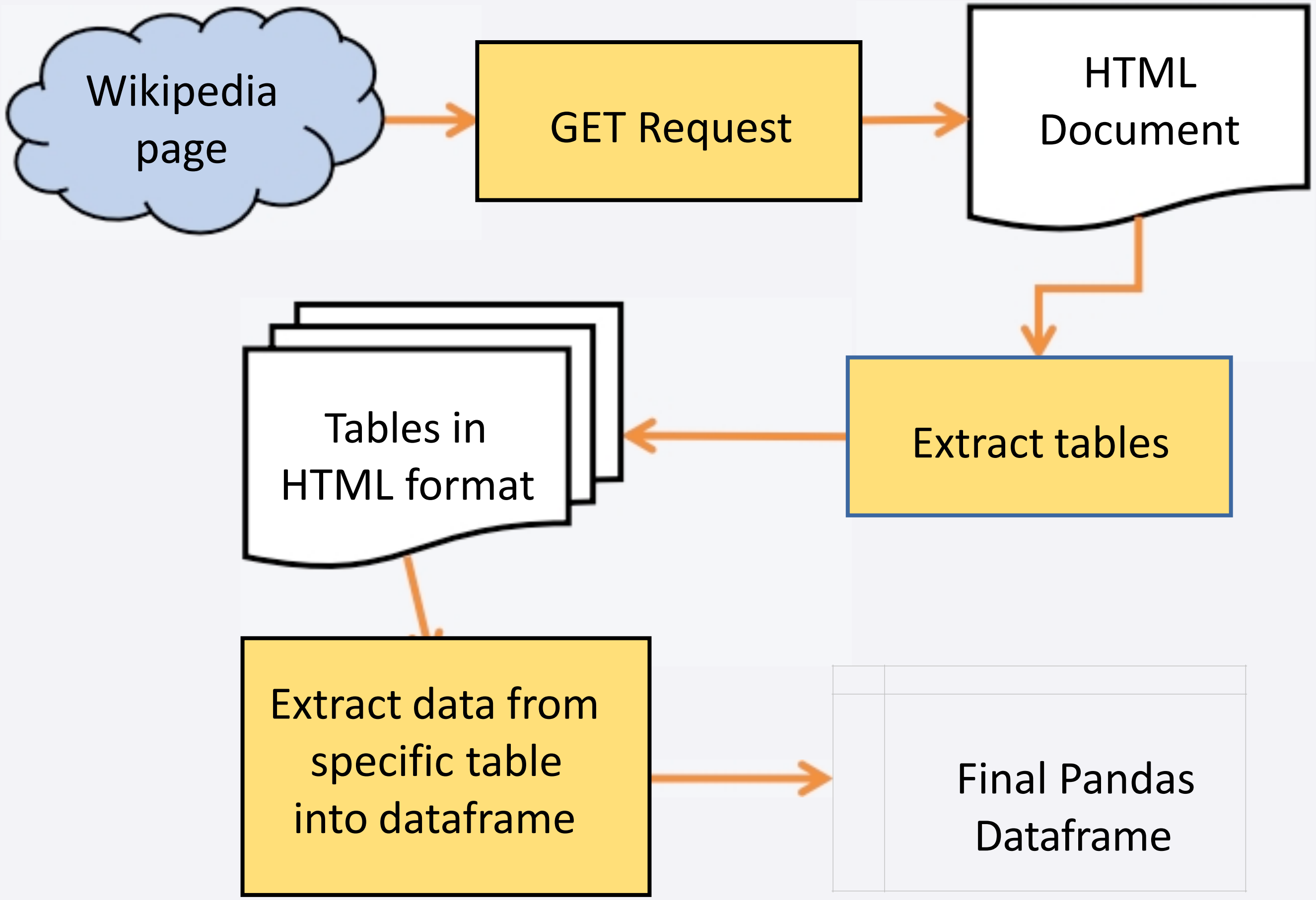
Data Collection – SpaceX API

- Data has been retrieved from SpaceX via a REST API call that returned a Json file with the information of the different flights. This file has been casted into a dataframe for the data wrangling process.



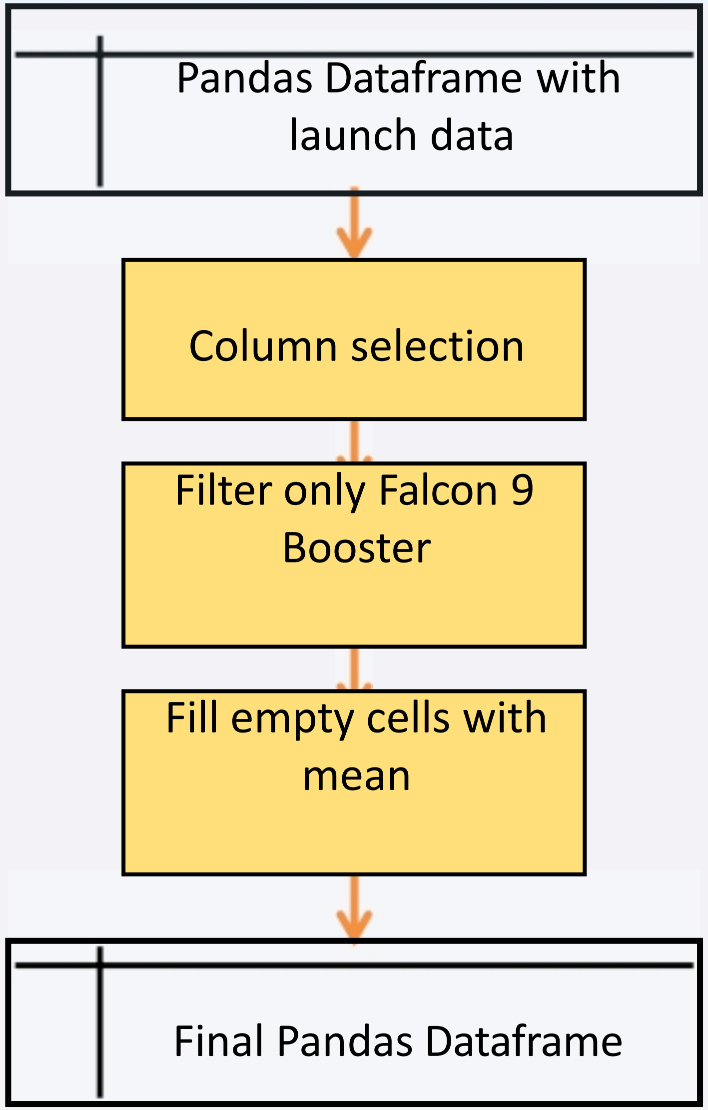
Data Collection - Scraping

- Additional data also comes from Wikipedia historic registers of the flights. This data has been extracted from the HTML document via web scrapping and loaded into a dataframe.



Data Wrangling

- From the dataframe, first the useful columns have been selected.
- Only the flights that used a Falcon 9 booster where used.
- Empty cells where filled with the mean of the column.
- Qualitative variables where transformed into a set of dummy variables in order to include them into the models.



EDA with Data Visualization

For the exploratory analysis the next charts where used:

- Different scatter plots with Launch Site, Payload mass, flight number and orbit as axis and success class as hue in order to identify relationships between these variables and success rate.
- Bar chart with orbit as class and success rate in the y-axis in order to identify orbits with higher success rate.
- Line chart with Year in the x-axis and success rate in the y-axis in order to see the evolution of the success rate with the years.

EDA with SQL

SQL queries where used to answer questions about:

- Launch sites and their successes and failures.
- Which boosters where used with high payloads.
- What was the distribution of outcomes at the landing.
- What is the distribution of payload mass.
- When was the first successful landing.

Build an Interactive Map with Folium

Folium maps where populated with the following items:

- Launch site markers so they could be located in the map and to been able to extract insight from their locations.
- Number of launches and outcome to each site in order to understand the distribution of the successes and witch sites were most used.
- Distance to important sites or infrastructures around the sites in order to understand what was required to have at a good launch site.

Build a Dashboard with Plotly Dash

The plots included in the dashboard are:

- Pie chart of successes per site. It transforms into a pie chart of successes vs. failures when a site is selected and it allow to study the distribution of outcomes.
- Scatter plot of payload by outcome that allow to study which payloads are used in each site and detect payloads with higher failure rates.

Predictive Analysis (Classification)

- The launch data was splitted into 2 sets for training and testing purposes each.
- With each set, 4 machine learning models where trained in order to predict outcome of the landings (logistic regression, SVM, decision tree and KNN).
- The resulting score of the models where compared in order to determine the best model of the 4.
- The confusion matrix was analyzed in order to analyze the final results.

Results

The results will be presented in three different ways:

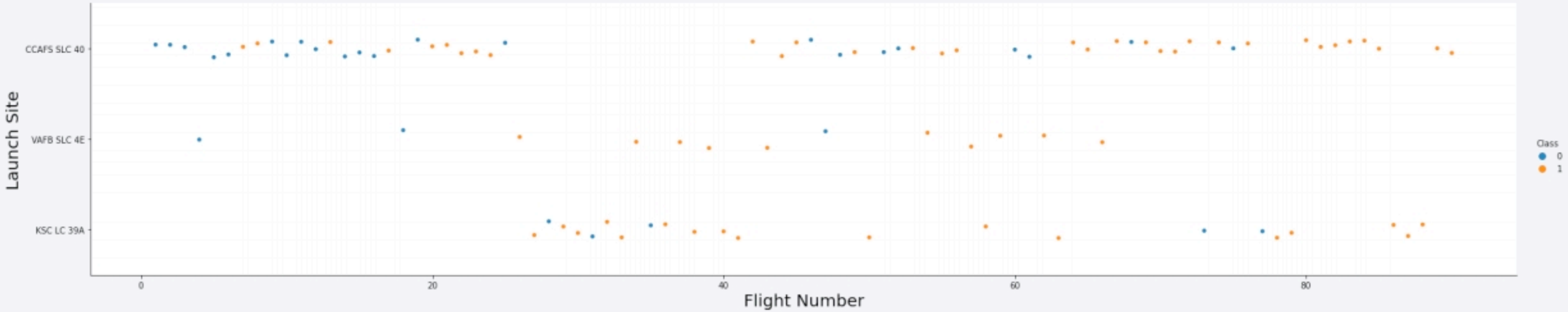
- The geographical distribution of launches will be presented in an interactive map in order to ease the understanding of the location requirements of a successful launch site.
- A dashboard will allow for interactive exploration of the underlying dataset.
- The results of the prediction process will be presented via a confussion matrix that shows the distribution of correct and incorrect predictions.

Section 2

Insights drawn from EDA

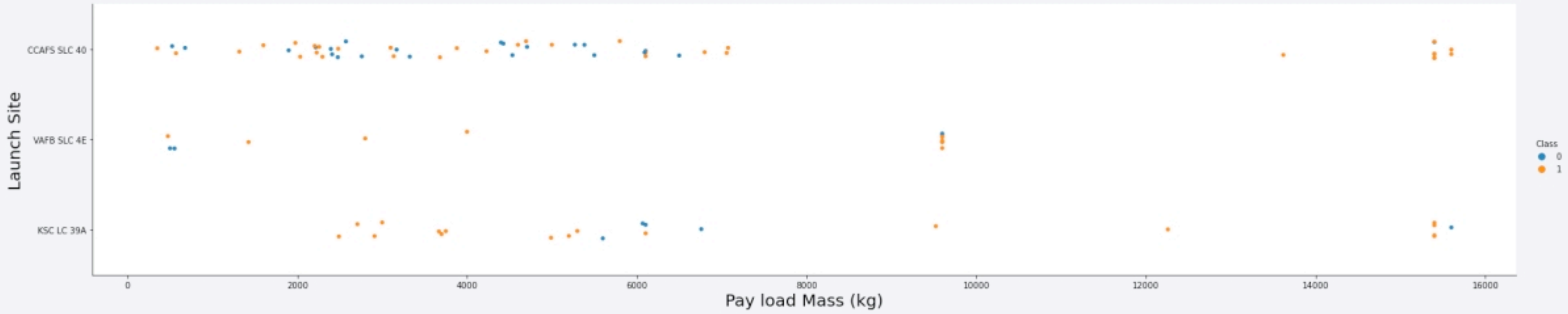
Flight Number vs. Launch Site

All of the 3 launch sites seem to have improved their success rate as times passes. The first launch site used, CCAFS SLC 40, had a very low success rate at the beginning, which was solved in the other sites and currently has a high success rate.



Payload vs. Launch Site

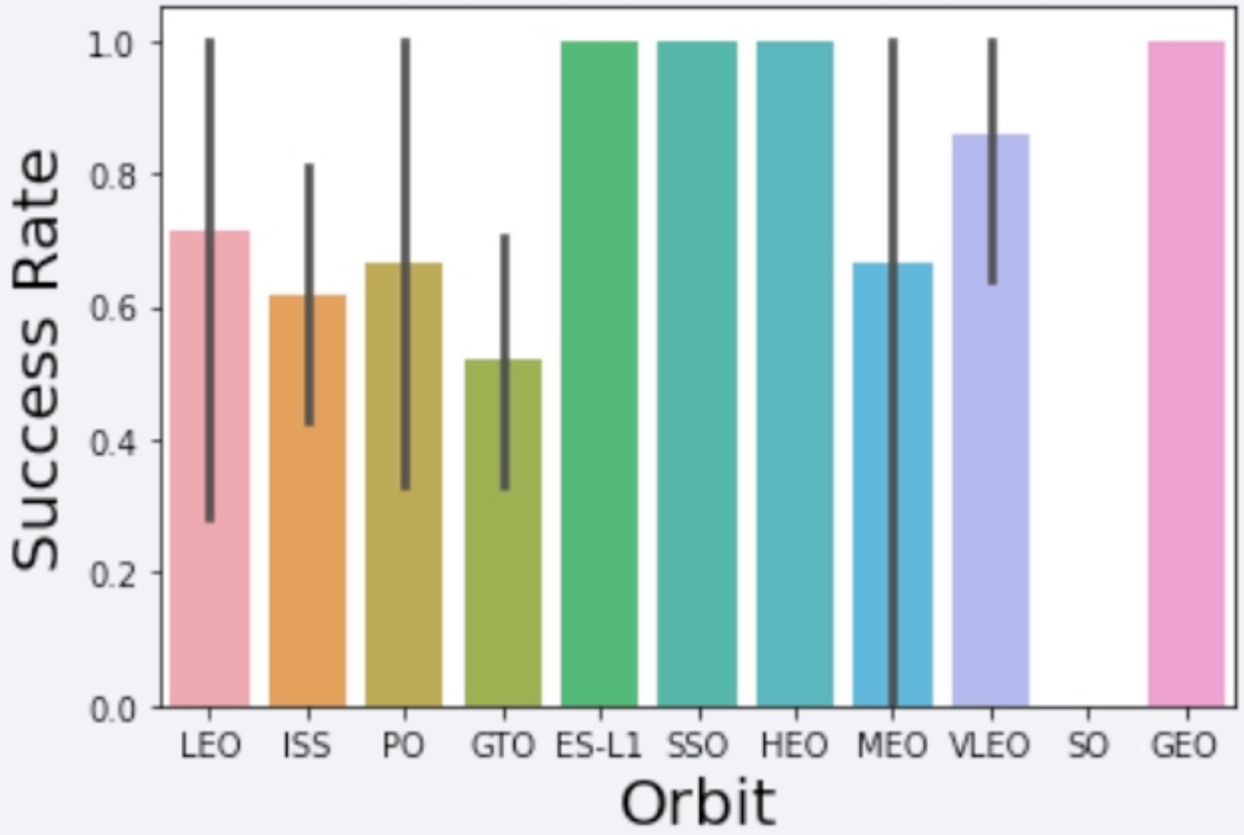
Launch site VAFB SLC 4E seems to not allow payloads as heavy as the other sites. Most of the failures concentrate in the low payload range, probably these launches where initial trials with low loads.



Success Rate vs. Orbit Type •

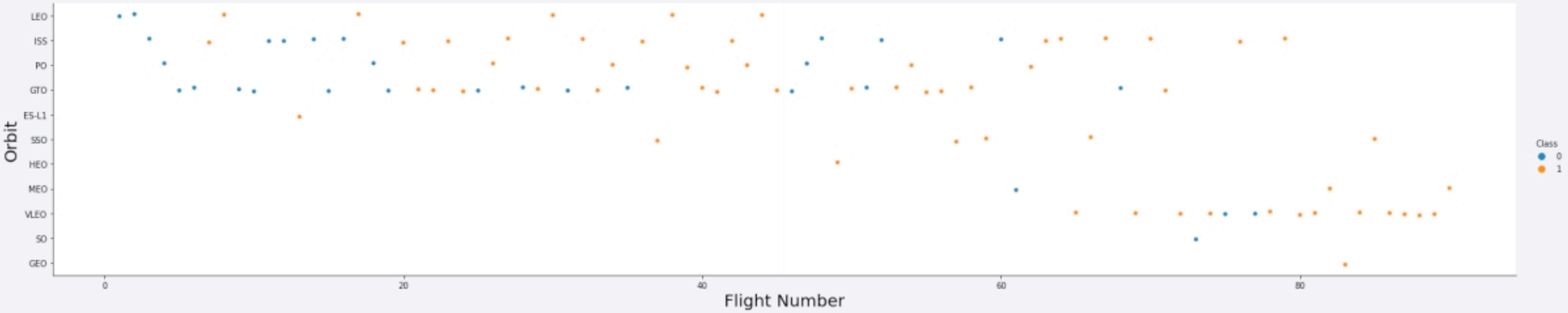
ES-L1, SSO, HEO and GEO are the orbits with the highest success rate.

- The rest of the orbits have mixed results, with GTO, SO and ISS being the worst orbits.
- Probably the orbits with bad results were also initial trials.



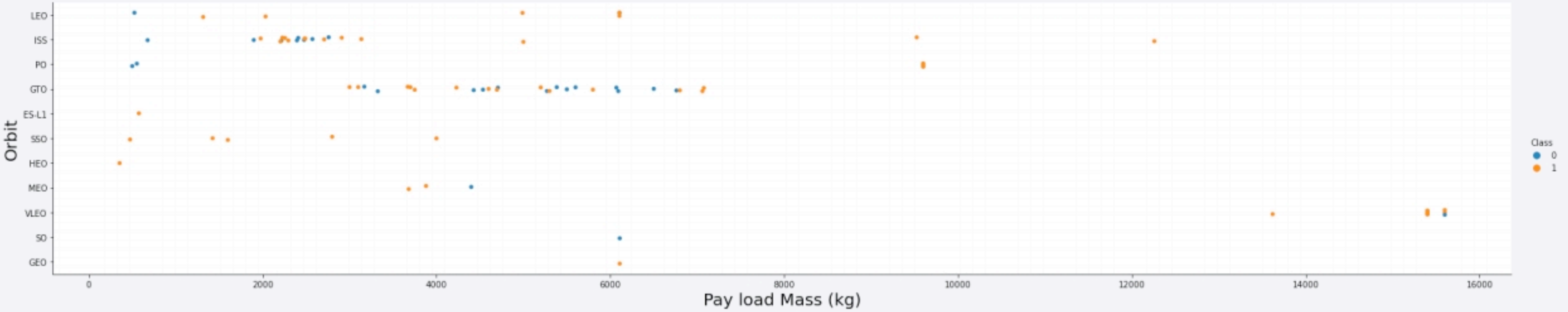
Flight Number vs. Orbit Type

The scatter plot reveals that SSO, HEO, MEO, SO and GEO results are not reliable, due to the low launch number. VLEO seems to be the most used now, while the worse results of the other 4 orbits can be explained by the initial trials results.



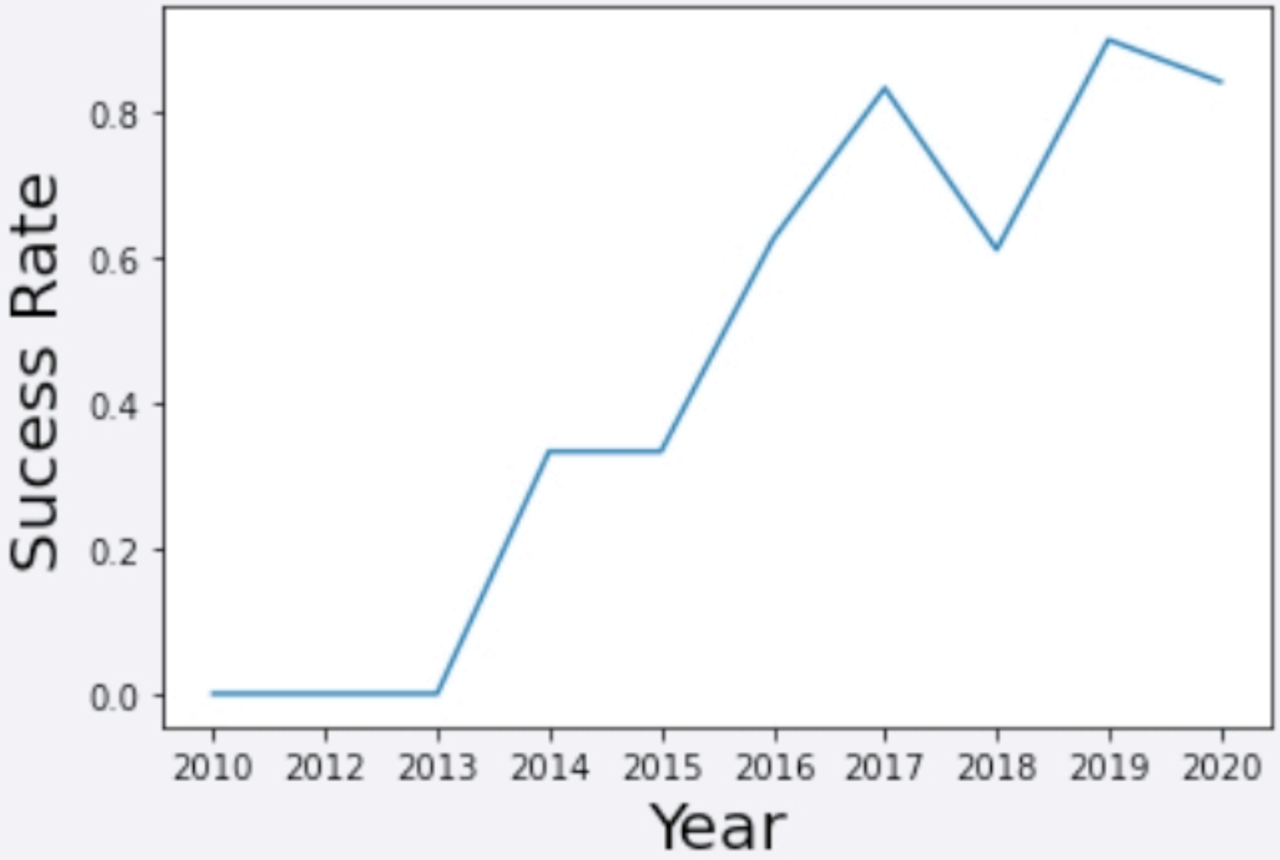
Payload vs. Orbit Type

The scatter plot reveals VLEO, the most used orbit in the last launches, is also the one with the higher payload. The rest of the orbits have lower payloads, with the first 4 being used with medium payloads.



Launch Success Yearly Trend

The overall evolution of success rate through the years shows how SpaceX has improved their rockets and launch protocols in the last decade.



All Launch Site Names

- There are only 4 launch sites used by SpaceX.
- From the data visualization we also learned that only 3 of these are used in the Falcon 9 booster version launches.
- The 4th one, CCAFS LC-40 might have been abandoned in favor of the similar named one, CCAFS SLC-40.

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

Launch Site Names Begin with 'CCA'

The first 5 launches from the CCAFS LC-4a site had successful outcomes, but did not achieve a landing. These where at the start of the SpaceX launches, so they where still trial launches, with 2 of the payload shown been called ‘demo’.

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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	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

The total payload sent to space for NASA is over 45 Tons. With the firt payloads weighting slightly over 500 Kg, as seen previously in the first 5 launches.

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

 * ibm_db_sa://cps31401:***@dashdb-txn-sbox-yp-lon02-13.services.eu-gb.bluemix.net:50000/BLUDB
Done.
```

1
45596

Average Payload Mass by F9 v1.1

The average payload is over 2500 Kg. The payload distribution is skewed to the right, as there where initial launches with 0 payload but the most recent can reach over 15 Tons.

```
%sql SELECT AVG("PAYLOAD_MASS__KG_") FROM spacex WHERE "Booster_Version" LIKE 'F9 v1.1%'
```

```
 * ibm_db_sa://cps31401:***@dashdb-txn-sbox-yp-lon02-13.services.eu-gb.bluemix.net:50000/BLUDB
Done.
```

1
2534.666666

First Successful Ground Landing Date

The first successful landing in a ground pad was at the end of 2015, over 5 years later than the first flights of SpaceX. This shows that it took a lot of time to achieve the desired outcome of recoverable boosters.

```
%sql SELECT MIN("Date") FROM spacex WHERE "Landing__Outcome" = 'Success (ground pad)'

 * ibm_db_sa://cps31401:***@dashdb-txn-sbox-yp-lon02-13.services.eu-gb.bluemix.net:50000/BLUDB
Done.
```

1
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

Booster Falcon 9 seems to be the one that can be recovered after the launch of high payloads.

Booster_Version
F9 B4 B1040.1
F9 B4 B1043.1
F9 FT B1032.1

Total Number of Successful and Failure Mission Outcomes

The overall mission outcome of the flights are successes. The failures tend to happen in the landing stages of the flight, not in the delivery of the payload.

Mission_Outcome	2
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

From the list of boosters that have achieved the max payload it seems clear that the latest generation of Falcon 9 boosters is the most indicated for the delivery of bigger payloads.

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

2015 Launch Records

The year 2015 was the year of the first success, but it also had several landing failures. In this year, the Falcon 9 boosters where already been used which later would become the best performing booster in SpaceX.

Landing__Outcome	Booster_Version	Launch_Site
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

In general, landing outcome show mixed results between years 2010 and 2017, with most successes occurring in the later years, according to the previous line graph.

Landing__Outcome	Number of Outcomes
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Uncontrolled (ocean)	2
Failure (parachute)	1
Precluded (drone ship)	1

Section 4

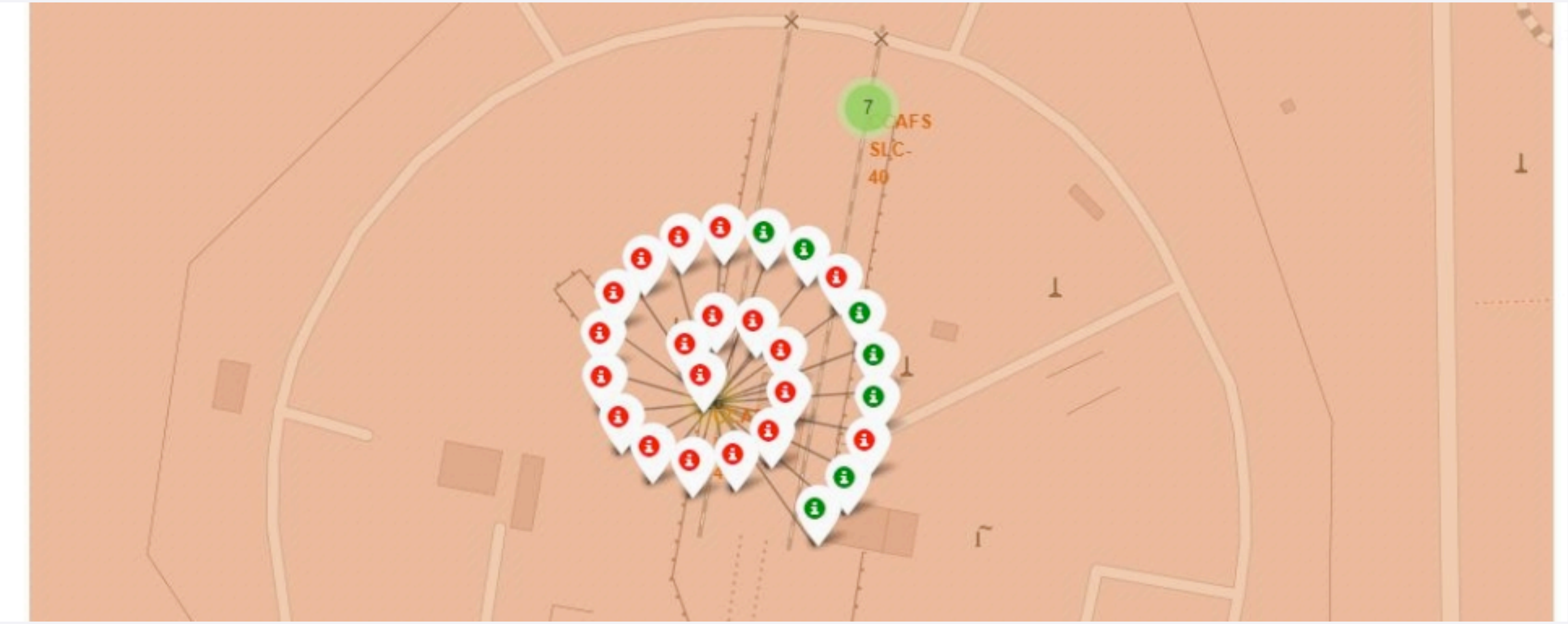
Launch Sites Proximities Analysis

Launch Site Locations

Launch sites tend to be placed as near to the equator as possible, in order to use the rotation of the Earth. This strategy was used by early NASA flights.

Map of Launch Outcomes at Cape Canaveral

Initial launches took place at Cape Canaveral. Where NASA used to launch their own flights. The high failure rate here is due to the initial trials.



Launch Sites Surroundings

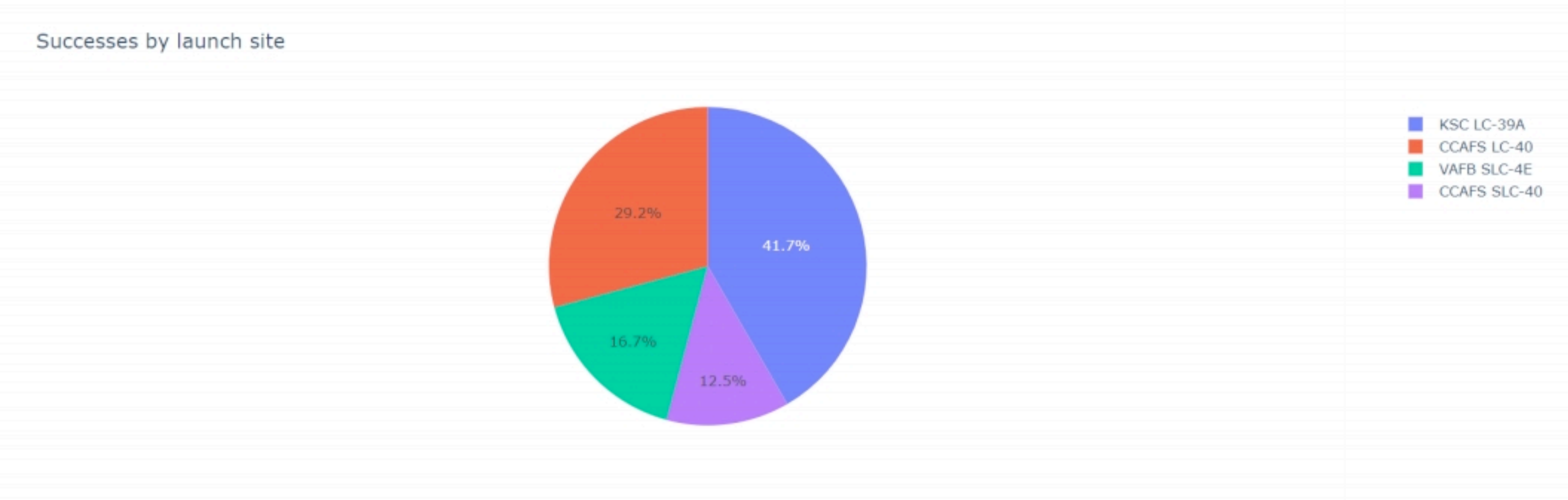
Launch sites seem to be located near to cities and railroads in order to simplify the transport of material to the sites and the allocation of workers.

Section 5

Build a Dashboard with Plotly Dash

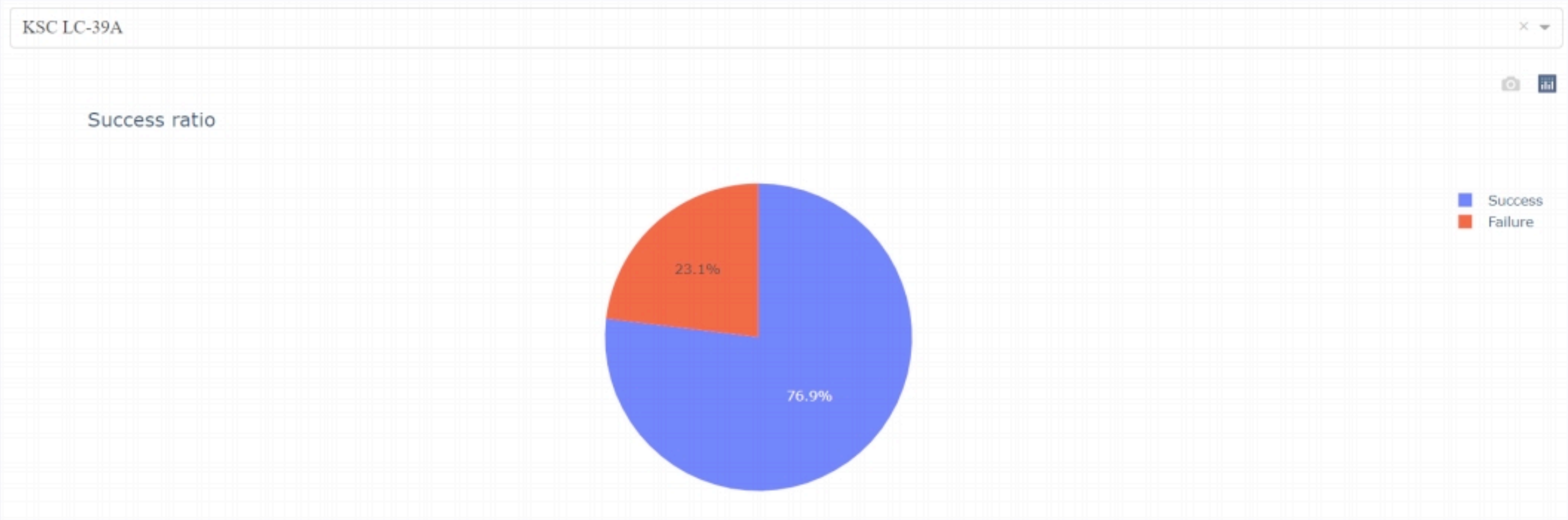
Successes by Launch Site

Out of all the sites, the one that represents the majority of all successes is KSC LC-39, despite not being the one with the most launches, as we saw in the flight number vs launch site scatter plot.



Highest Success

The site with the higher success is KSC LC-39A, but it could be due to initial trials taking place in other sites.



Distribution of Payload and Outcome by Launch Site

The distribution of successes seem to be similar for every payload. On the other hand, site KSC LC-39A seems to have the higher success rate loading middle range loads.

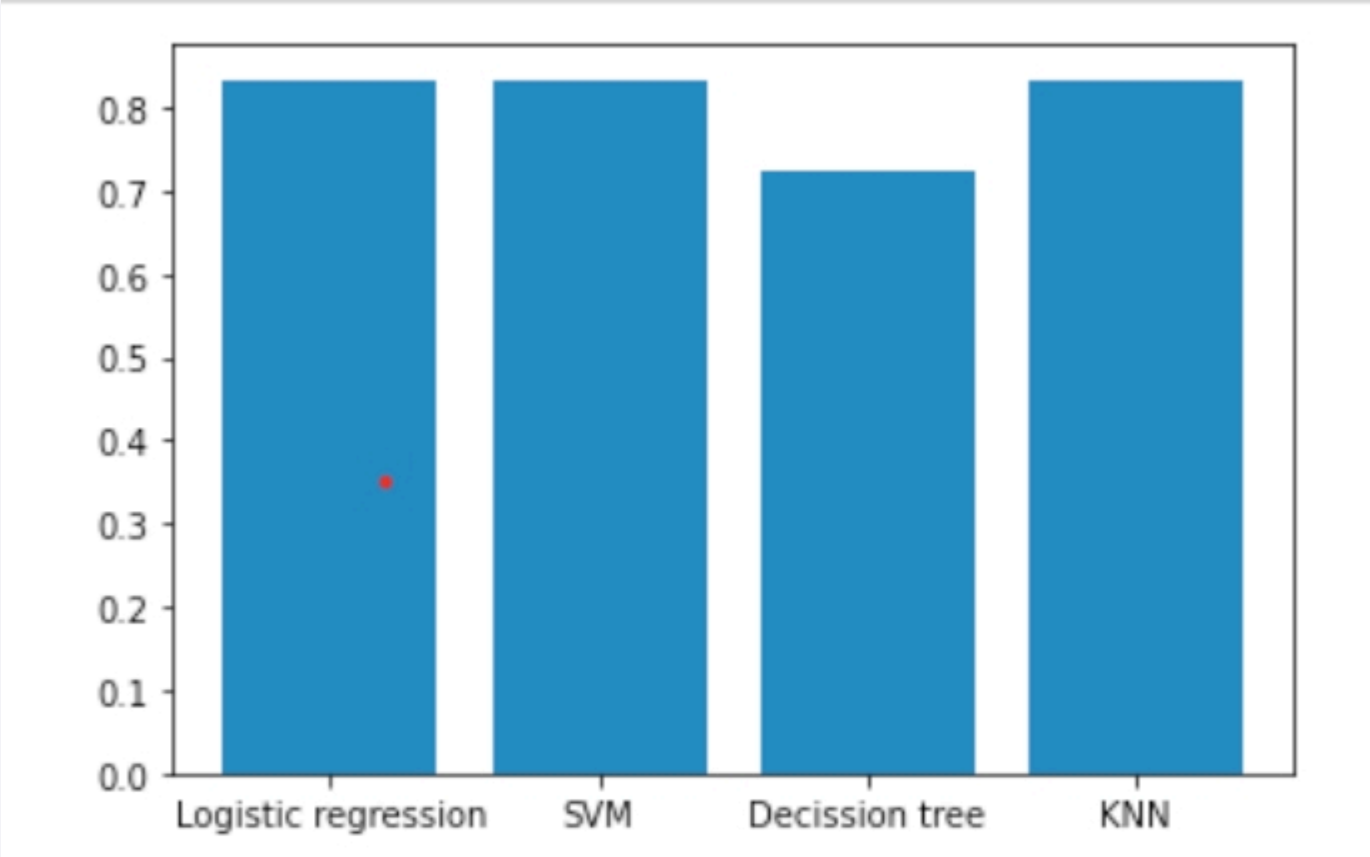
Section 6

Predictive Analysis (Classification)

Classification Accuracy

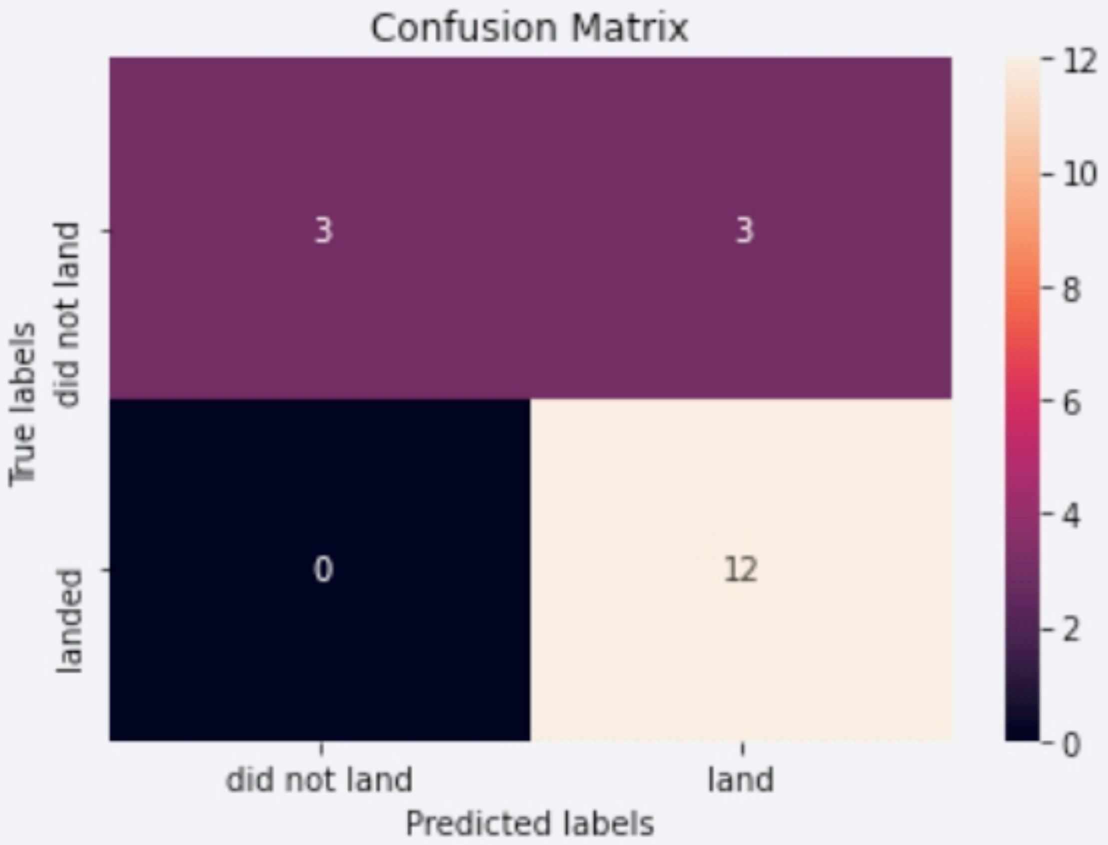
Logistic Regression, SVM and KNN all achieved a score of 0.8334 in the same test set.

The similar score might have been due to the small sample of data, as it had less that 80 samples to train the models.



Confusion Matrix

All these models where able to detect when the booster would land, but due to unbalance in the sampling method, it fails at identifying failures.



Conclusions

The conclusions that we can derive are:

- SpaceX recent success comes from a lot of development in the technology used in the boosters and the adjustment in the orbits used for the flights.
- The improvement in the boosters has allowed to SpaceX to launch higher payloads, this trend is expected to continue in the future.
- The launch sites seem to require near communication infrastructure such as railroad and cities. They also are located as close to the equator as possible.
- The machine learning models where unable to clearly differentiate between both classes, mostly due to small sample size and unbalanced classed in the set.

Thank you!