IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

S. H. Motaharpour
December-23-2022

Outline

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

Executive Summary

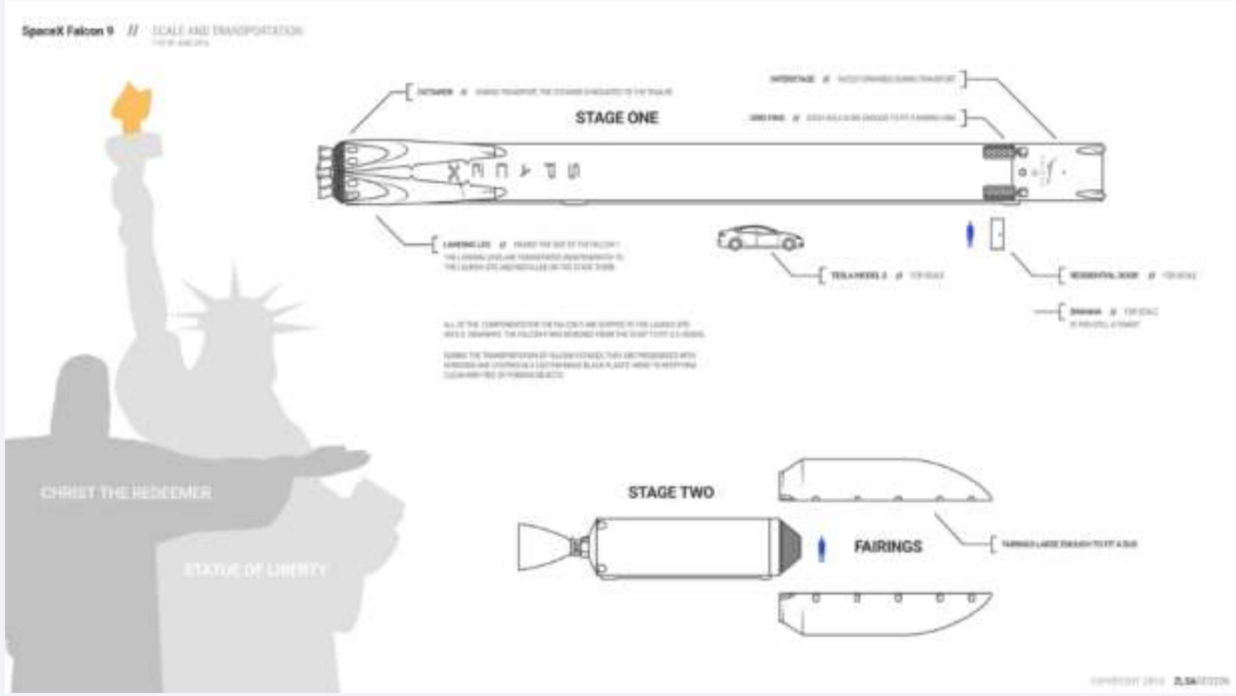
As a starting point for almost all data science projects, It needs to collect data as much as possible and relevant. Data had been collected from various sources. After raw data has been collected, Its quality is improved by performing data wrangling. Then exploring the processed data started.

We gained further insights into the data by applying some basic statistical analysis and data visualization. We will see how variables might be related to each other.

predictive models for discovering exciting insights built and evaluated.

Introduction

- **SpaceX** advertises **Falcon9** rocket launches on its website, with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can **reuse the first stage**. Therefore <u>if we can determine if the first stage will land, we can determine the cost of a launch.</u>



- This has done by **gathering information** about Space X, creating **dashboards** and train a **machine learning** model

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Collecting the Data with an API
  - Web scraping related Wiki pages
- Perform data wrangling
  - With Pandas library and some SQL query
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Test Linear-Regression, SVM, Tree-Decision and K-nearest that tuned with Cross Validation in Scikit-Learn package

Data Collection

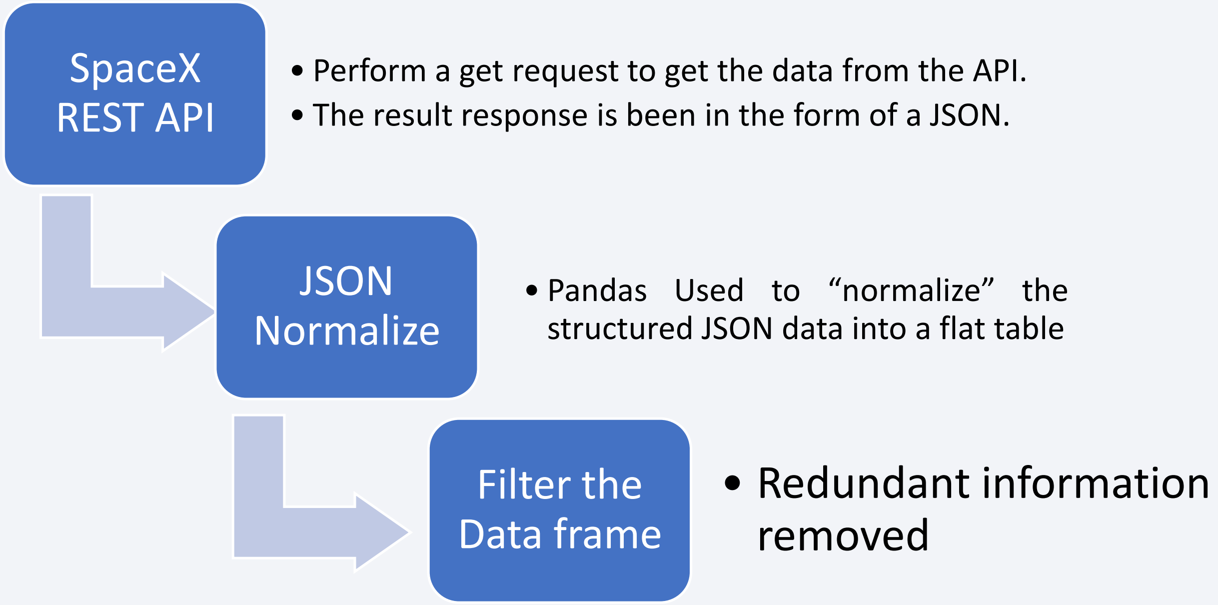
In this part we gathered data from two sources:

- the SpaceX REST API
- web scraping related Wiki pages.

The goal was to use this data to predict whether SpaceX will attempt to land a rocket or not.

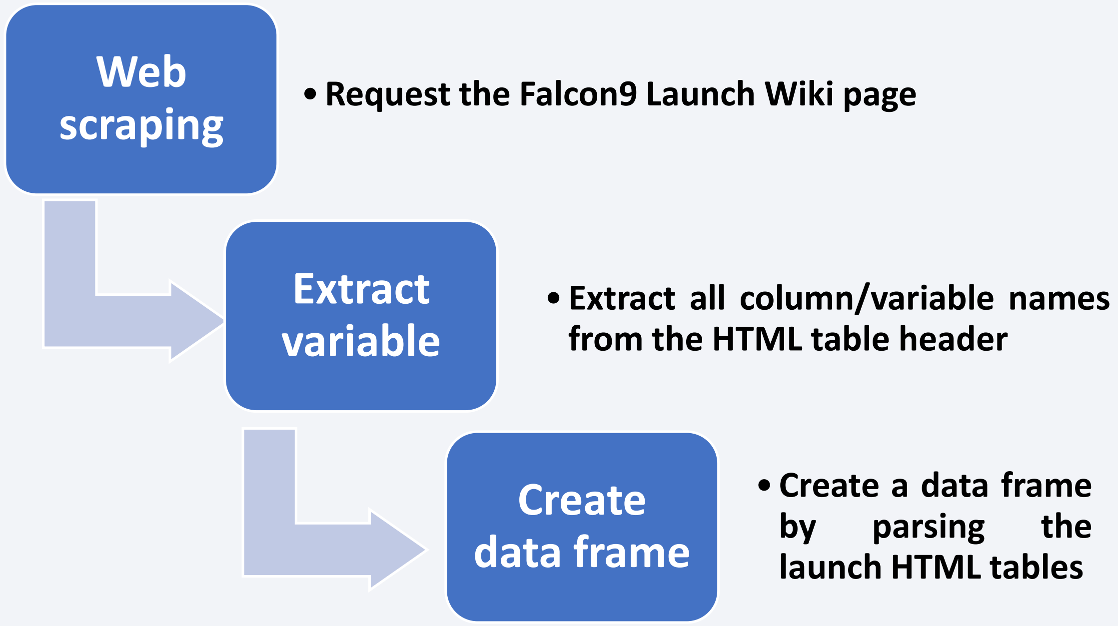
Data Collection – SpaceX API

- This API will give us data about launches, including information about the rocket used, payload delivered, launch specifications, landing specifications, and landing outcome.
- We will perform a get request using the requests library to obtain the launch data.
- This result can be viewed by calling the .json() method.
- Our response will be in the form of a JSON, specifically a list of JSON objects.
- [Notebook Link]



Data Collection - Scraping

- The Python BeautifulSoup package has been used to web scrape some HTML tables that contain valuable Falcon 9 launch records.
- Then the data from those tables parsed and converted them into a Pandas dataframe for further visualization and analysis.
- [Notebook Link]



Data Wrangling

- The raw data must to transform into a clean dataset which provides meaningful data on the situation we are trying to address:
- Wrangling Data using an API, Sampling Data, and Dealing with Nulls.
- [Notebook Link]

EDA with Data Visualization

- Visualize the relationship between:
  - Flight Number and Launch Site
  - Payload and Launch Site
  - success rate of each orbit type
  - FlightNumber and Orbit type
  - Payload and Orbit type
- Visualize the launch success yearly trend
- [Notebook Link]

EDA with SQL

- The SQL queries that performed:
  - the unique launch sites in the space mission
  - the total payload mass carried by boosters launched by NASA (CRS)
  - 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
  - List the failed landing_outcomes in drone ship, their booster versions, and launch site names for 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
- [Notebook Link]()

Build a Dashboard with Plotly Dash

- With interactive visual analytics, users could find visual patterns faster and more effectively.
- A dashboard application built with the Python Plotly Dash package.
- This dashboard application contains input components such as a dropdown list and a range slider to interact with a pie chart and a scatter point chart.
- This dashboard can be used to find more insights from the SpaceX dataset more easily than with static graphs.
- Notebook Link

Predictive Analysis (Classification)

- we will build a machine learning pipeline to predict if the first stage of the Falcon 9 lands successfully.
- This will include:
  - Preprocessing [to standardize the data]
  - Train test split
  - Train the model and perform Grid Search [to find the hyperparameters]
  - Determine the model with the best accuracy using the training data.
- Logistic Regression, Support Vector machines, Decision Tree Classifier, and K-nearest neighbors have been tested.
- Notebook Link

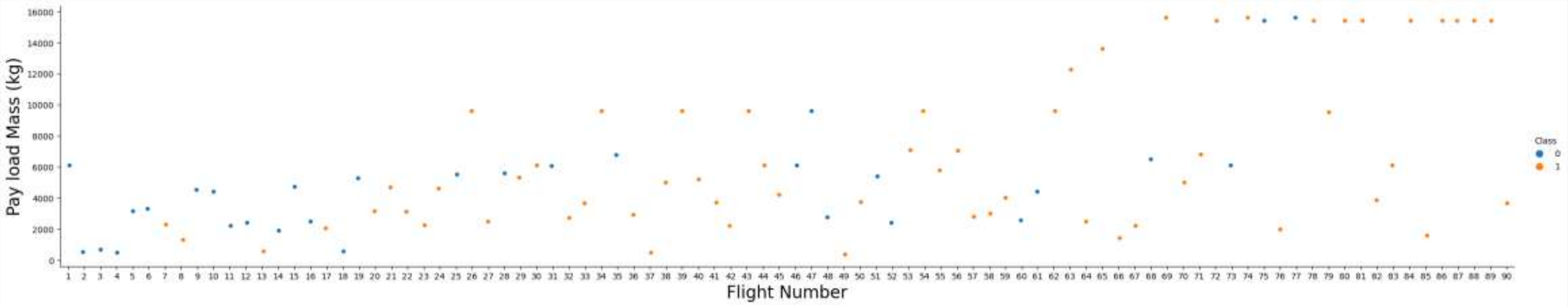
Results

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

Section 2

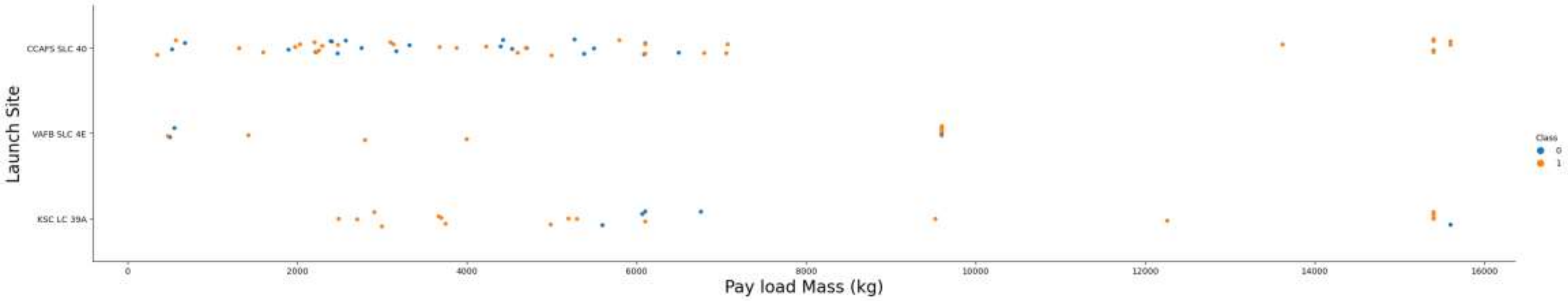
Insights drawn from EDA

Flight Number vs. Launch Site



- We see that as the flight number increases, the first stage is more likely to land successfully. The payload mass is also important; it seems the more massive the payload, the less likely the first stage will return.

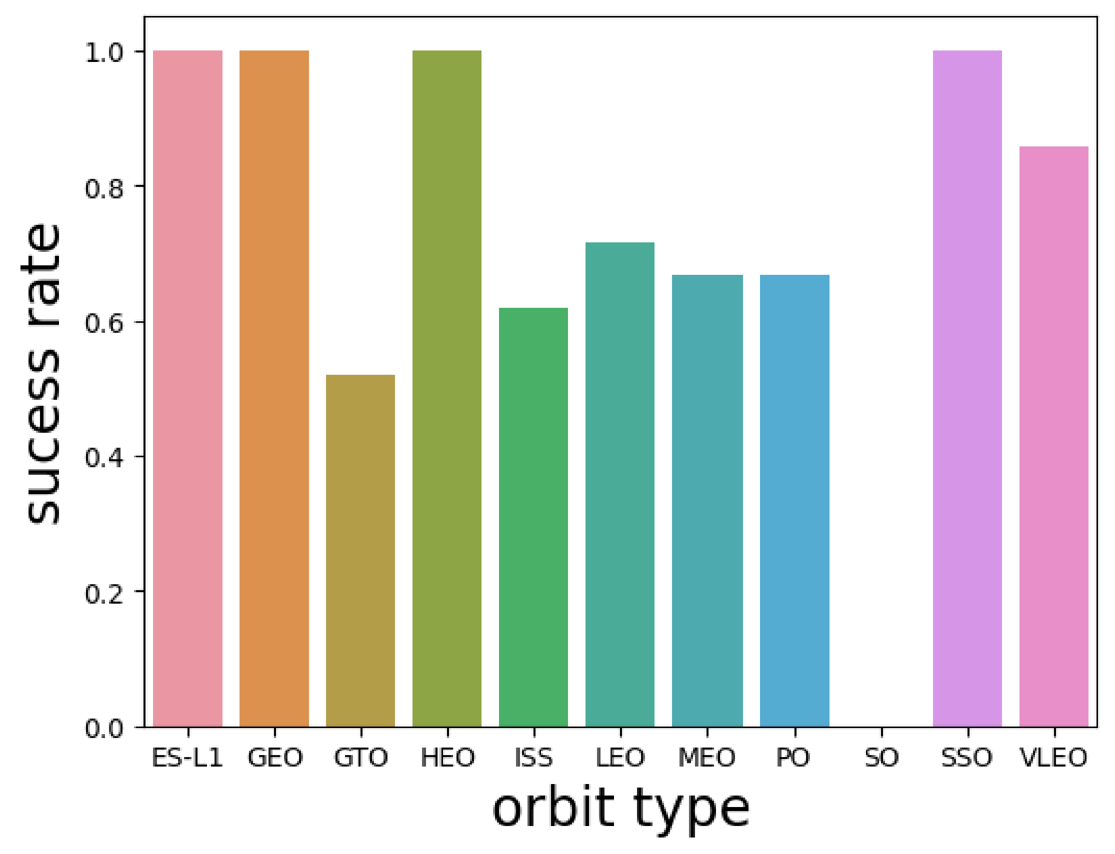
Payload vs. Launch Site



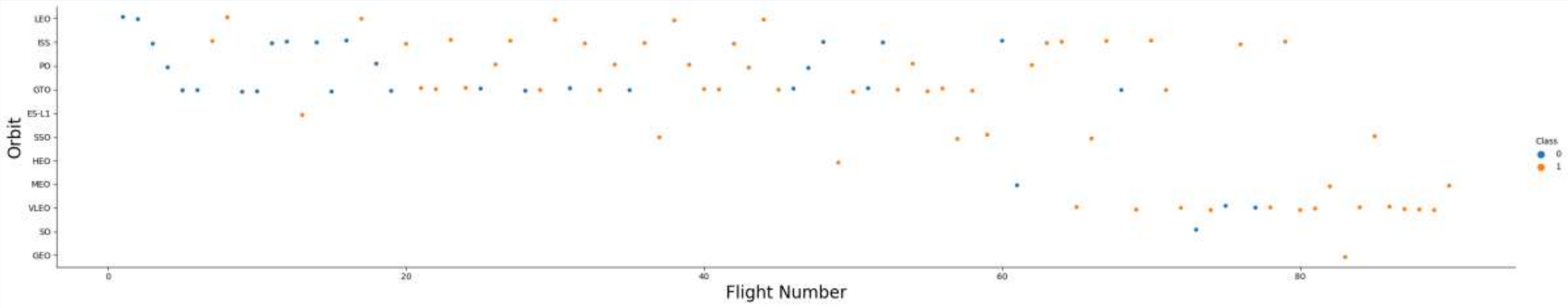
- More tests were performed in the weight range below 8 tons.
- Tests above 8 tons have been more successful.

Success Rate vs. Orbit Type

- Orbit type has effect on Success rate
- ES-L1, GEO, HEO, SSO have highest success rate

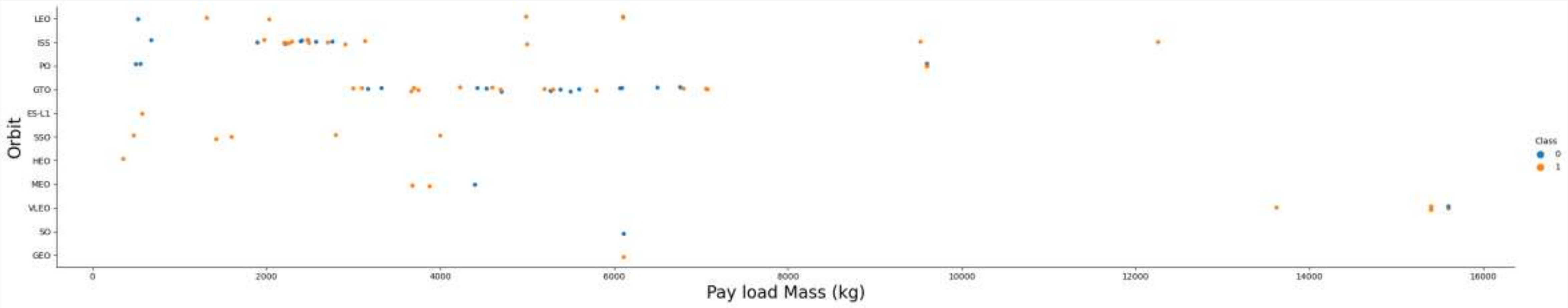


Flight Number vs. Orbit Type



- The flight number has no meaningful relationship with the Orbit type

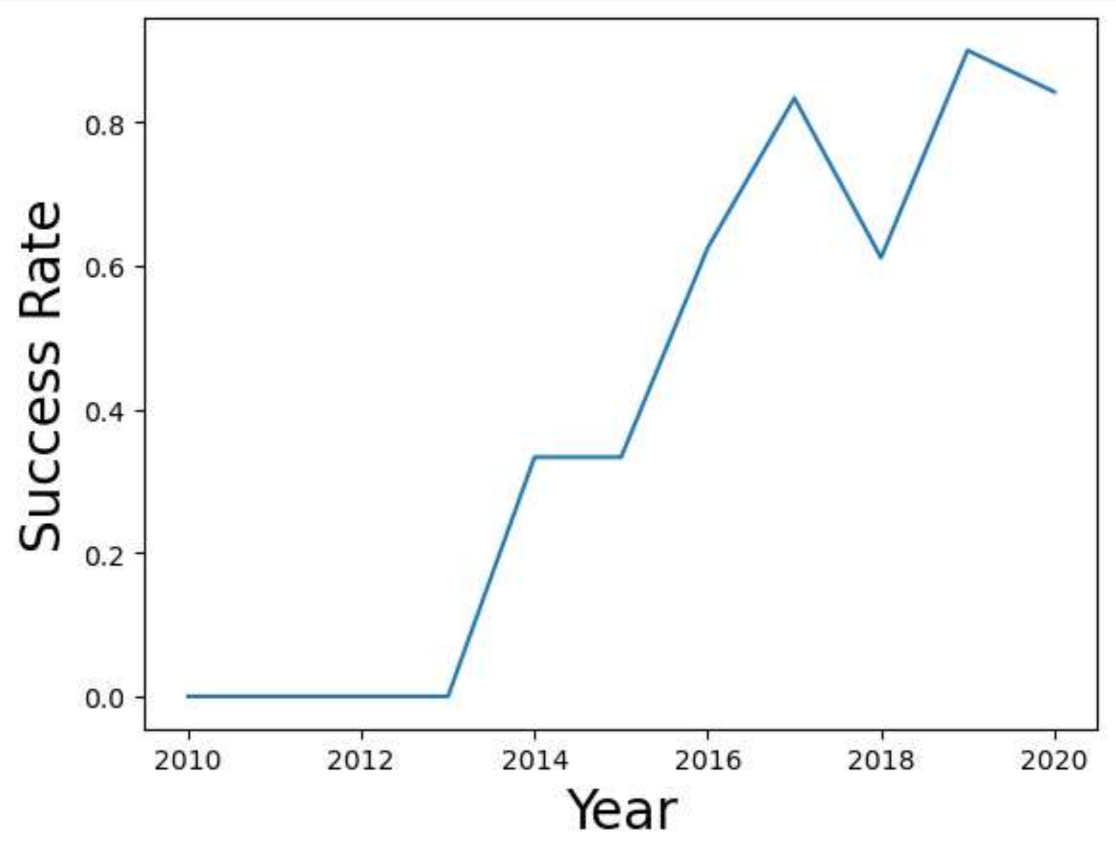
Payload vs. Orbit Type



- The two orbit GTO and ISS have been the most focused on the test.
- Weight has no effect on the performance of samples in different orbits

Launch Success Yearly Trend

- Over time, the success rate has increased, which shows the efforts and learning of the construction team.
- In 2018, an unusual thing happened



All Launch Site Names

- Find the names of the unique launch sites

```
%sql select distinct Launch_Site from spacextbl
```

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`

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
2010-06-04 00:00:00	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08 00:00:00	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 00:00:00	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 00:00:00	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01 00:00:00	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

```
%sql select sum(PAYLOAD_MASS__KG_) as 'total payload mass [kg]' from spacextbl
```

- total payload is 619967 kg

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1

```
%%sql select avg(PAYLOAD_MASS__KG_) as 'average payload mass for (F9 v1.1) [kg]'
from spacextbl where Booster_Version like 'F9 v1.1%'
```

- the average payload mass is 2534.6 kg

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad

```
%%sql select min(date) as 'first successful landing date'
from spacextbl
where "Landing _Outcome" like 'Success%'
```

- first successful landing date is 2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

```
%%sql select distinct Booster_Version from spacextbl
where "Landing _Outcome" = "Success (drone ship)" and PAYLOAD_MASS__KG_ between 4000 and 6000
```

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

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

```
%%sql select "Landing _Outcome", count(*) as count from spacextbl
group by "Landing _Outcome"
```

Landing _Outcome	count
Controlled (ocean)	5
Failure	3
Failure (drone ship)	5
Failure (parachute)	2
No attempt	21
No attempt	1
Precluded (drone ship)	1
Success	38
Success (drone ship)	14
Success (ground pad)	9
Uncontrolled (ocean)	2

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass

```
%%sql select distinct Booster_Version from spacextbl
where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_) from spacextbl)
```

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

2015 Launch Records

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

```
%%sql select Booster_Version , Launch_Site, strftime('%Y', DATE) as year from spacextbl
where "Landing _Outcome" = "Failure (drone ship)" and strftime('%Y', DATE) = '2015'
```

Booster_Version	Launch_Site	year
F9 v1.1 B1012	CCAFS LC-40	2015
F9 v1.1 B1015	CCAFS LC-40	2015

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

- 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

```
%%sql select "Landing _Outcome", count(*) as count from spacextbl
where date between date('2010-06-04') and date('2017-03-20')
group by "Landing _Outcome"
order by count desc
```

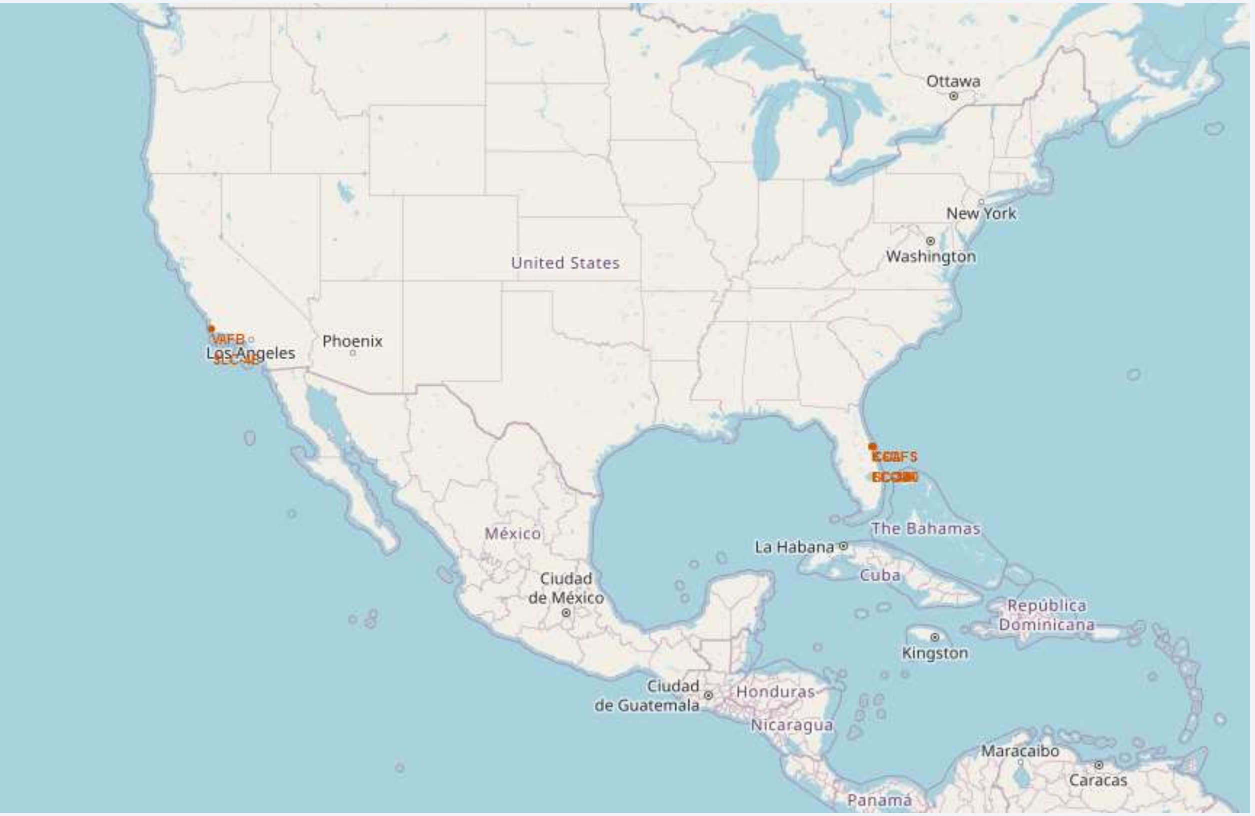
Landing _Outcome	count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

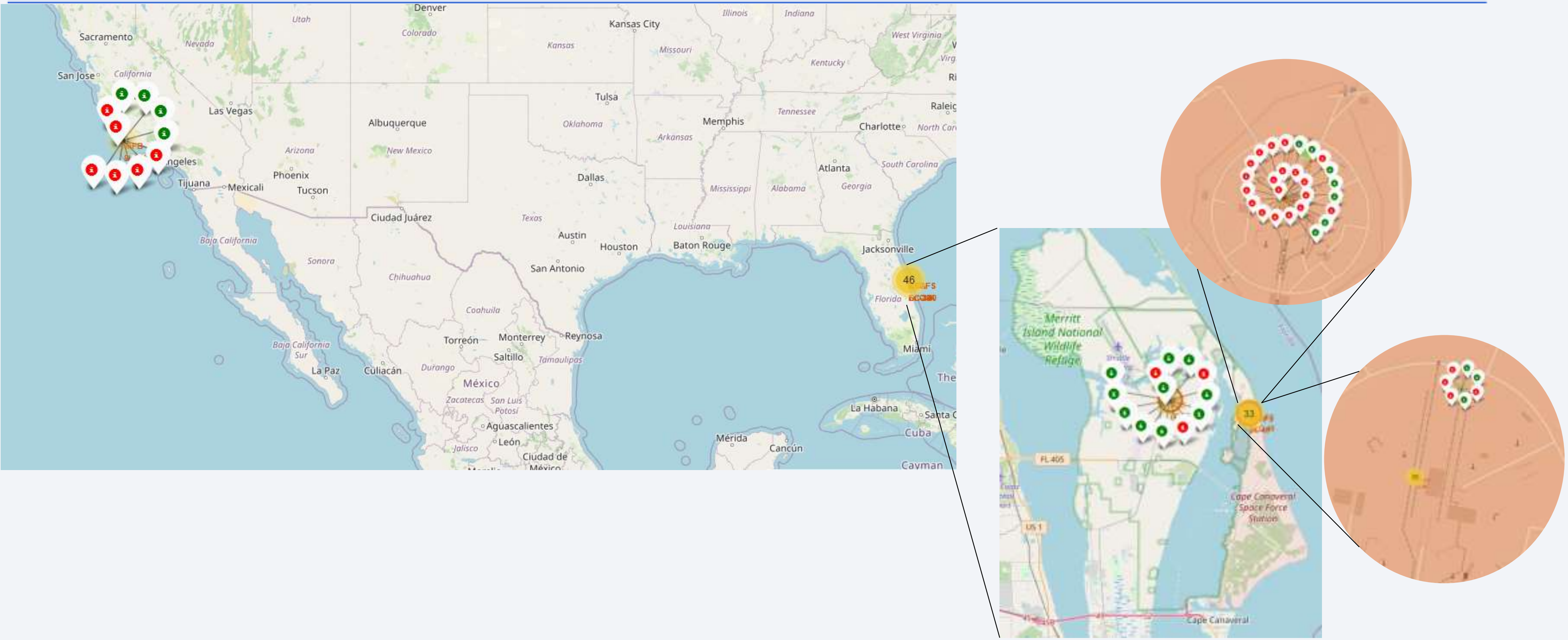
Launch Sites Proximities Analysis

launch sites’ location

- You see SpaceX launch sites in USA



the success/failed launches for each site



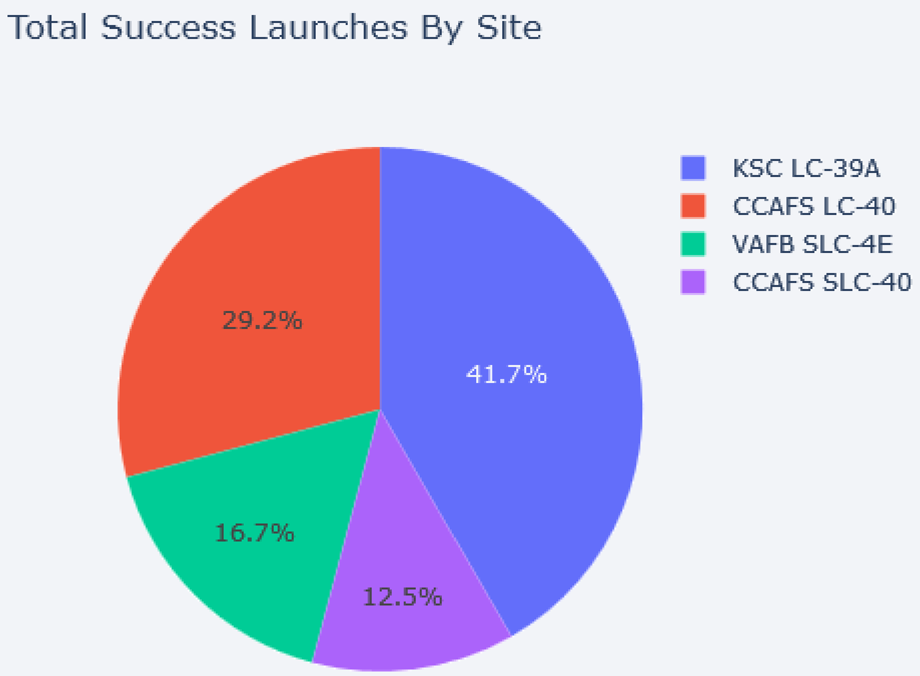
Section 4

Build a Dashboard with Plotly Dash

Success rate of sites

- Site `KSC LC-39A` has highest rate of success

- Other as order are:
  - CCAFS LC-40
  - VAFB SLC-4E
  - CCAFS SLC-40

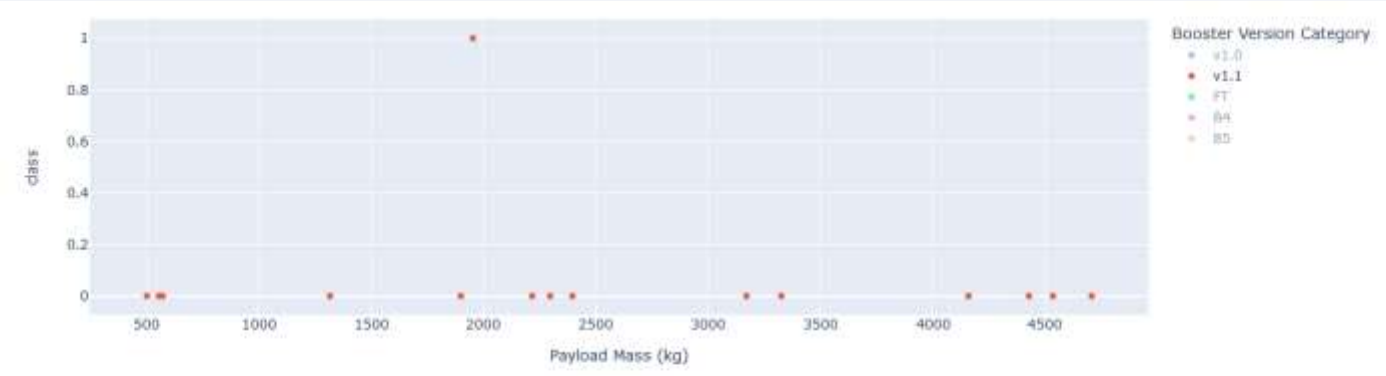


KSC LC-39A success ratio

- 77% of launches from this land side are successful.

Booster Version analysis

- All booster V1.0 tests failed
- All booster V1.1 tests failed except for one
- In other booster, there is no correlation between Payload mass and success

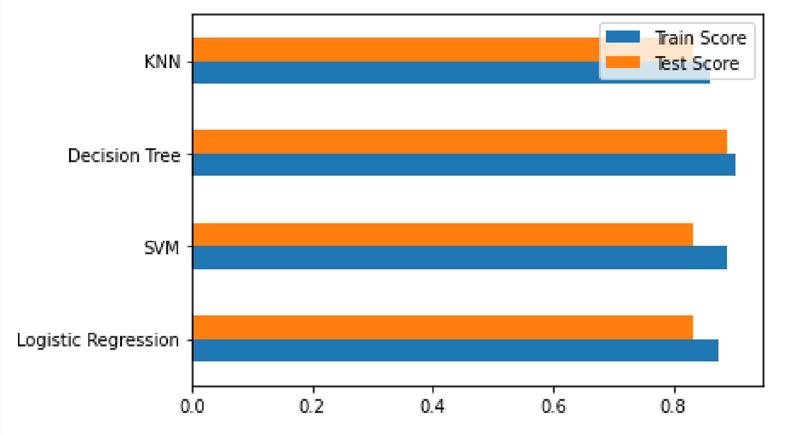


Section 5

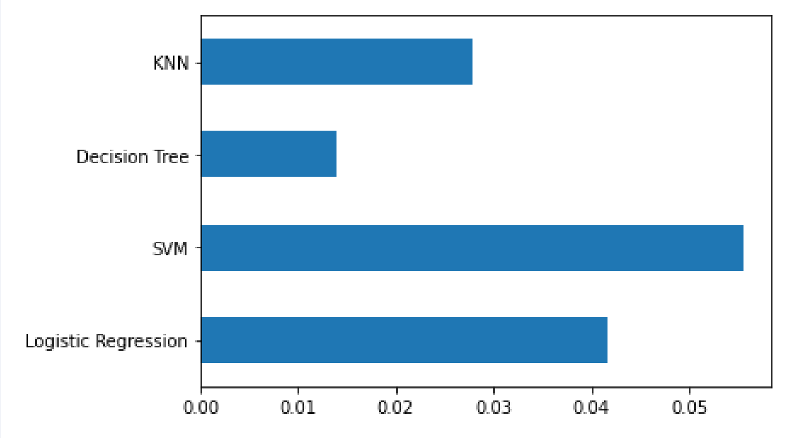
Predictive Analysis (Classification)

Classification Accuracy

- The best results from cross-validation model training are presented for each model type.
- From these results, the Decision Tree has the highest accuracy in training and test data and the lowest Differentiation between them.



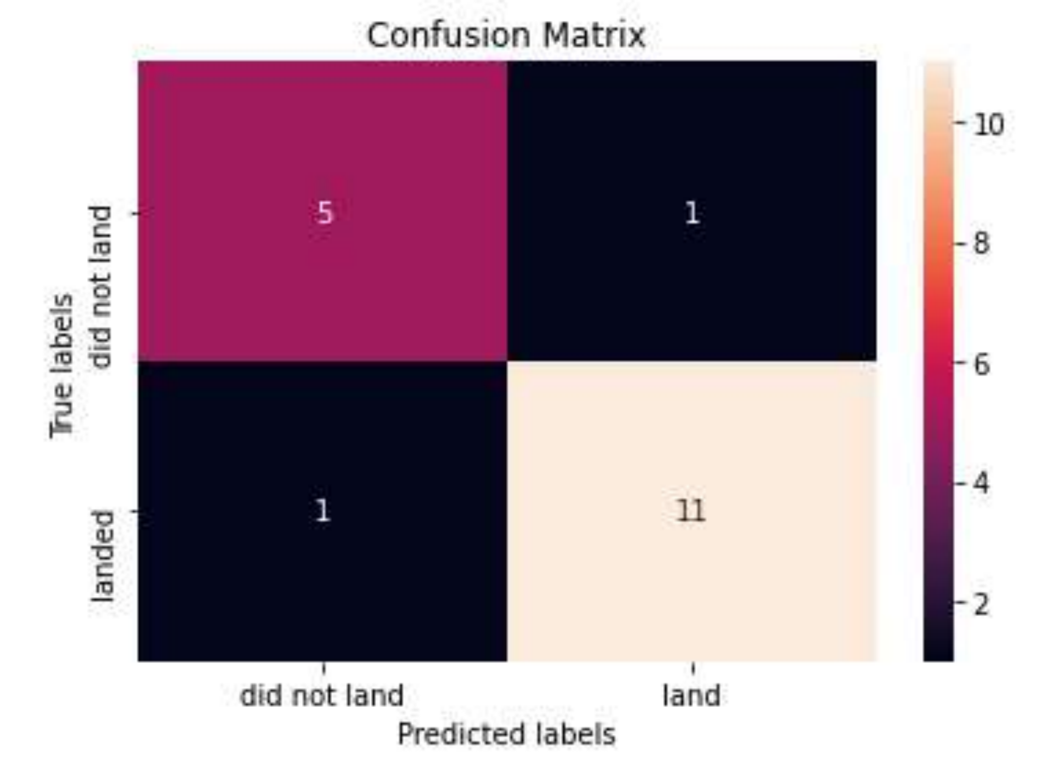
Models accuracy for train and test data



Differentiation of train and test data accuracy

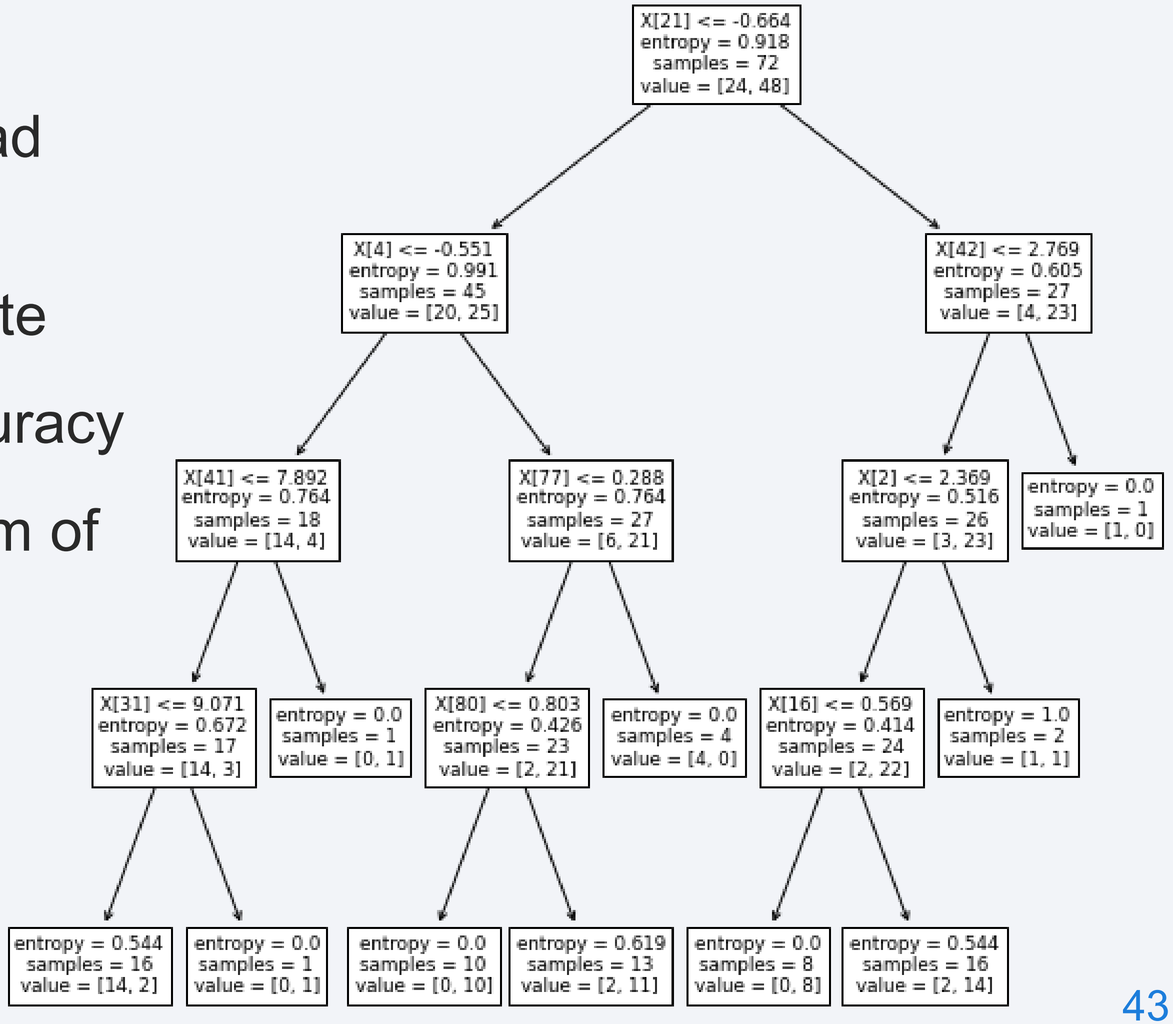
Confusion Matrix

- The confusion matrix of the Decision Tree has been shown.
- Based on this matrix the model just predict two sample wrong. One for each class.



Conclusions

- there is no correlation between Payload mass and success
- Success rates are different for each site
- Decision Tree model has highest Accuracy
- The decision tree results in a maximum of 4 decisions



Thank you!