SpaceY - Rewriting History

C. Tsangarides

13 Feb 2024

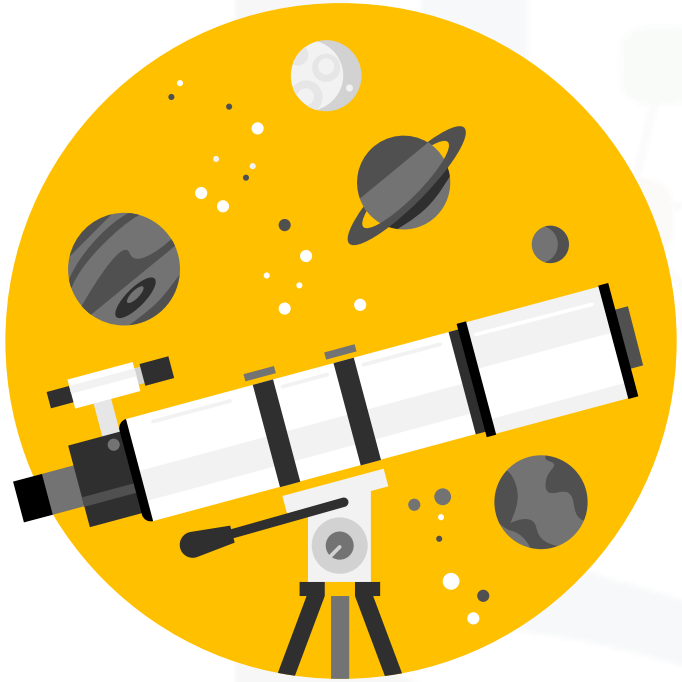
IBM Developer

SKILLS NETWORK

OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
  - Visualization – Charts
  - Interactive Maps / Dashboards
  - Predictive Analysis
- Conclusion
- Appendix

EXECUTIVE SUMMARY



- SpaceX is one of the most successful commercial-space company
- Its rocket launches are relatively inexpensive to others (i.e. Flacon 9 rocket @ 62 million dollars)
- Why: Can utilise the same rocket over and over again
- How: By recovering the first stage of rocket

- SpaceY can compete with SpaceX:
  - Public information of SpaceX launches available
  - Present visuals of feature insights that provide relations and patterns between attributes and success rates of landing
  - Demonstrate that predictive analysis through ML can help identifying the costs of future launches

INTRODUCTION

- The Falcon 9 rocket seems to be one of the most successful rockets of SpaceX
- Most of the time it is successfully recovered, others not on purpose, and others due to failures in landing
- Objective is to understand what are these situations and what accompanying variables play a crucial role in the success rate of landing
- Examples of such variables are:
  - Location of launching
  - Payload mass
  - Number of fins
- These are captured and analysed to produce further knowledge to understand possible correlations
- Finally, a ML model is created to predict with high accuracy when a new launch under a new combination of conditions is going to successfully land or not

METHODOLOGY – Part A

1. Data Collection through SpaceX (REST) API
   - API used to collect all possible data related to Flacon 9 launches with all its attributes
2. Followed by webscrapping WikiPedia Falcon 9 page
   - Aim to capture historic launches and their success rates
   - BeautifulSoup utilised to achieve this
3. DataFrame consisting above data is created
4. Data Wrangling
   - Sorting and cleaning data, dropping or filling missing data
   - Filtering only Falcon 9 related datasets

Data Collection

1. SpaceX Rest API (link)
   - It includes all data related to SpaceX rocket launches with the accompanied attributes per rocket and launch conditions, including orbit, customer and launch site

2. Webscrapping Falcon 9 launches (link)
   - It is achieved by using Beautiful Soup to read and collect data from wikipedia tables related to the above rocket from historic launches

3. This includes converting json format to a dataframe, which then is cleaned and prepared for further analysis

Data Collection - DataFrame

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	
4	1	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False	
5	2	2012-05-22	Falcon 9	525.0	LEO	CCSFS SLC 40	None None	1	False	False	False	
6	3	2013-03-01	Falcon 9	677.0	ISS	CCSFS SLC 40	None None	1	False	False	False	
7	4	2013-09-29	Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean	1	False	False	False	
8	5	2013-12-03	Falcon 9	3170.0	GTO	CCSFS SLC 40	None None	1	False	False	False	
...	...	...	...	...	...	...	...	...	...	...	...	
89	86	2020-09-03	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	2	True	True	True	5e9e3032
90	87	2020-10-06	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	3	True	True	True	5e9e3032
91	88	2020-10-18	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	6	True	True	True	5e9e3032

Screenshot of the consolidated dataframe

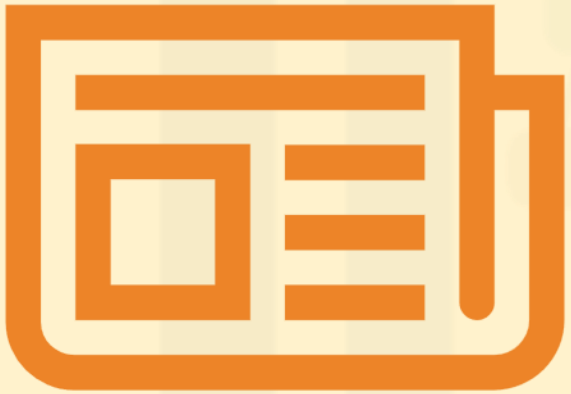
METHODOLOGY – Part B

1. Exploratory Data Analysis
   - Performed on produced dataframe from Method.-Part A
   - Important insights unveiled via statistical approaches
   - Visualisation of relationships between attributes and success-rate produced

2. Map locations
   - Folium used to produce maps of the launch sites
   - Identify possible correlations between number of launches per site and the location benefits

3. Interactive Visualisations
   - Dashboards used to create interactive plots and charts
   - Allowing stakeholders getting a better first-hand intuition

METHODOLOGY – Part C



1. Predictive Analysis
   - Data curated into training and testing format
   - 4 different ML Classifier models tested
2. Optimisation
   - GridSearchCV utilised to identify optimum hyperparameters per ML model
3. Evaluation
   - Each model is evaluated by calculating the accuracy score of the test portion of datasets
4. Deployment
   - Ideal model is selected

RESULTS

- It was found that there 17 different attributes and these are as follows:

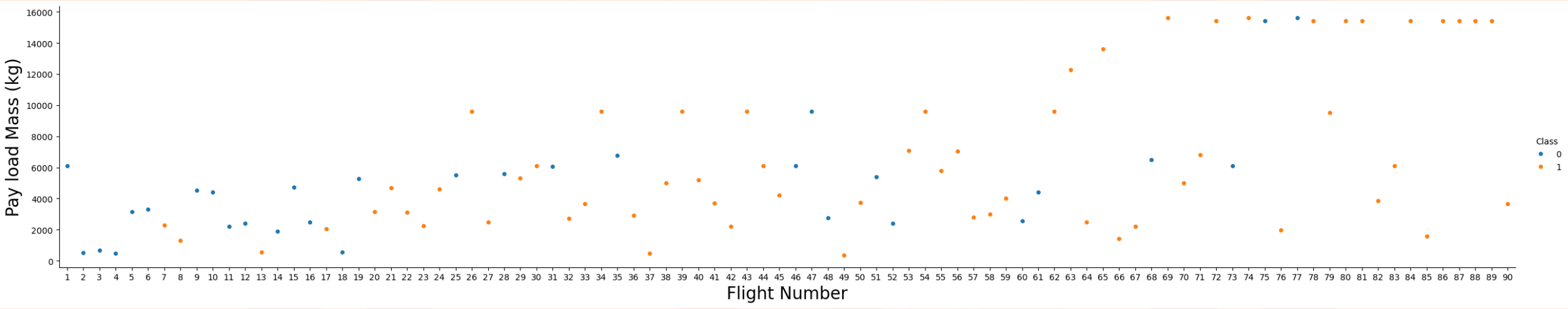
  FlightNumber,Date, BoosterVersion, PayloadMass, Orbit, LaunchSite, Outcome, Flights, GridFins, Reused, Legs, LandingPad, Block, ReusedCount, Serial, Longitude, Latitude

- Different launch sites were used for a number of launches:
  CCAFS LC-40, VAFB SLC-4E, KSC LC-39A, CCAFS SLC-40

- Depending on the combination of the above attributes and the launch-site, the result of success rate will be different

- Some of the attributes show a stronger effect than others and this is visible in the next slides

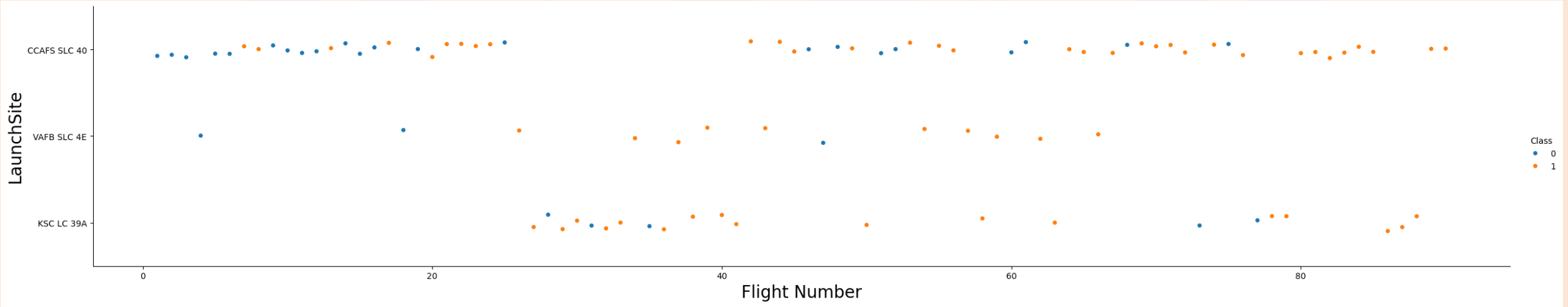
RESULTS – Payload mass effect



As the flight number increases, the 1st stage is more likely to land successfully.

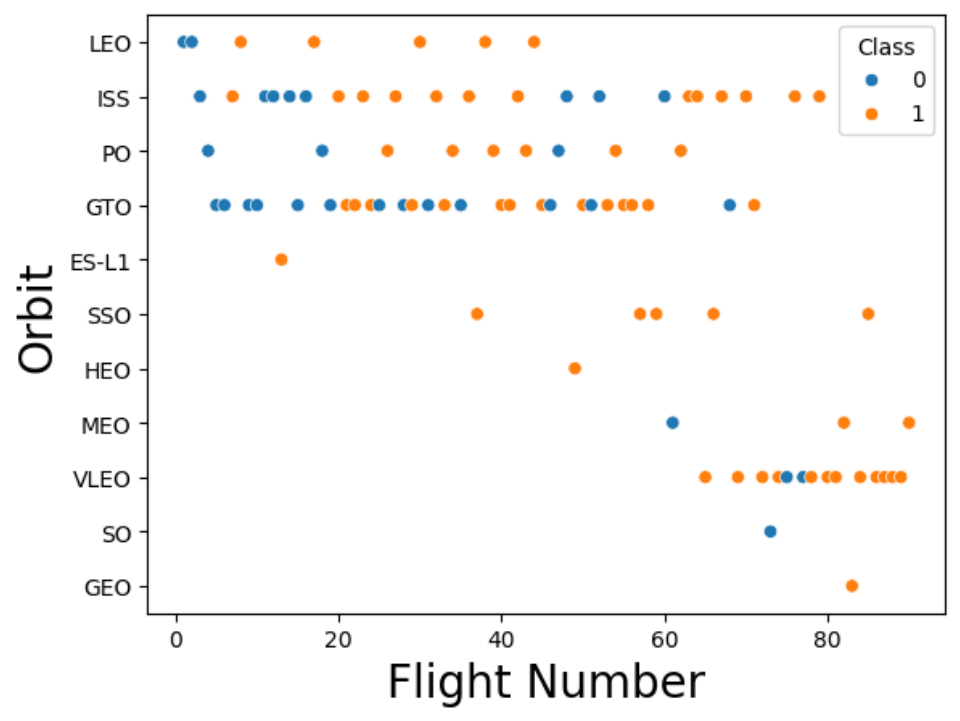
The payload mass is also important; it seems the more massive the payload, the less likely the first stage will return.

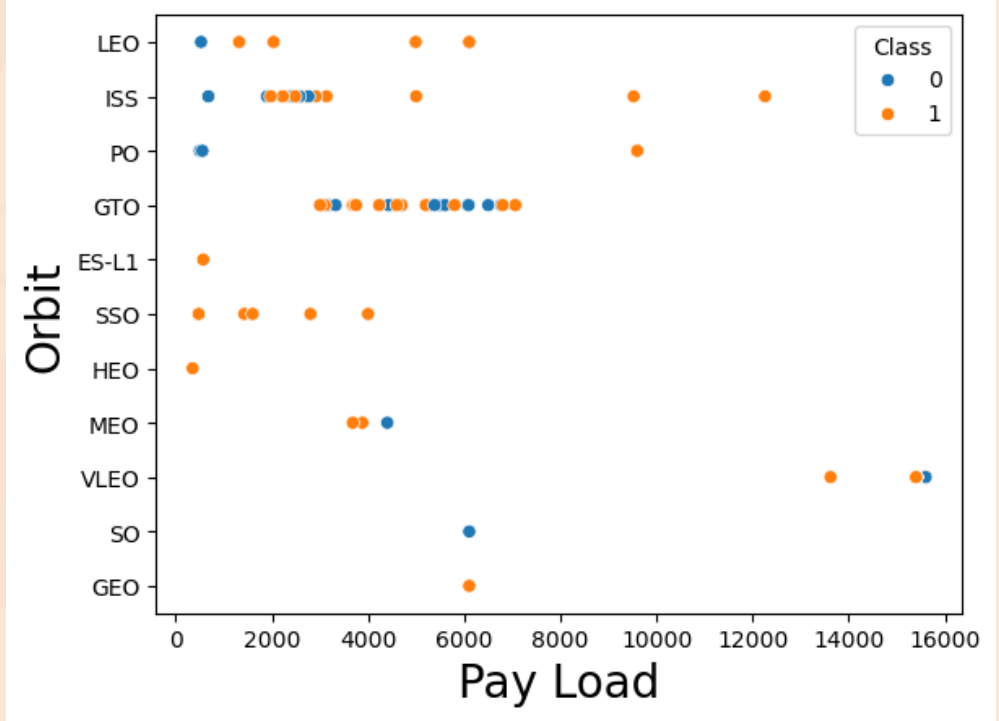
RESULTS – Launch site effect



Different launch sites have different success rates:
CCAFS LC-40, has a success rate of 60 %, while KSC LC-39A and VAFB SLC 4E has a success rate of 77%.

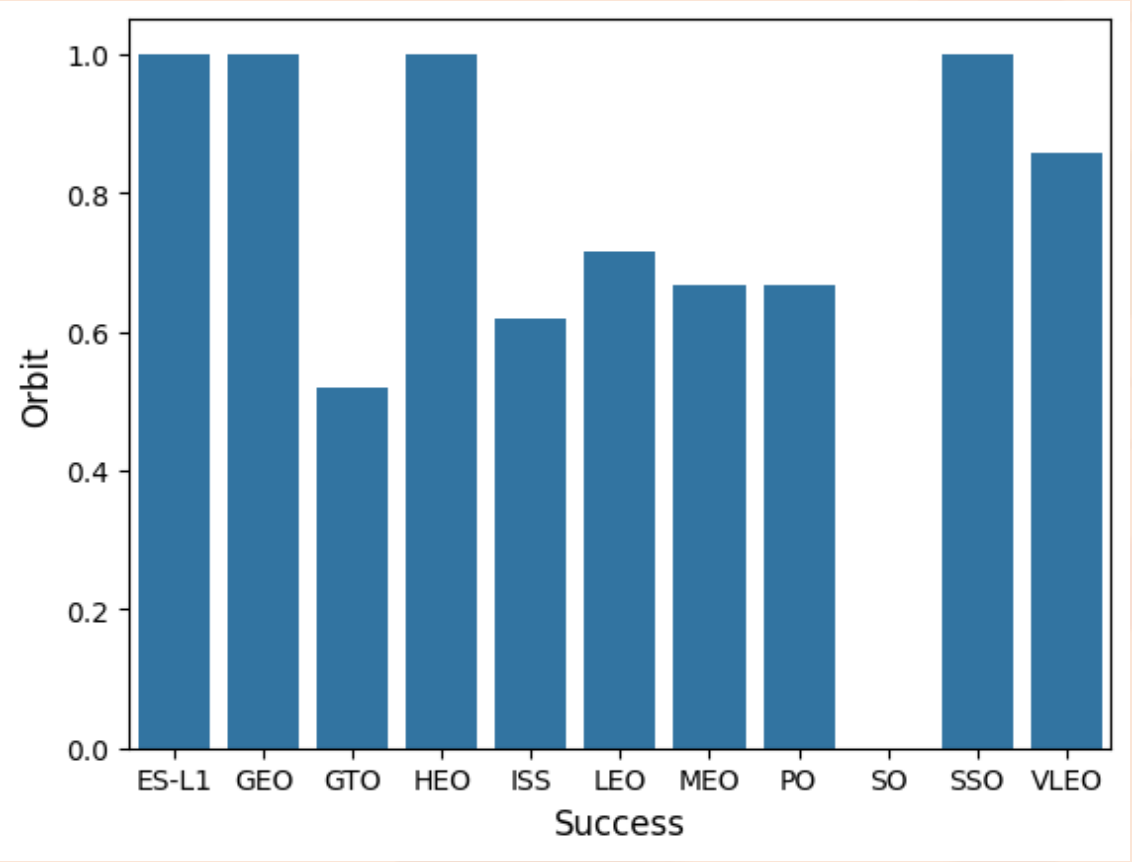
RESULTS – Orbit effect

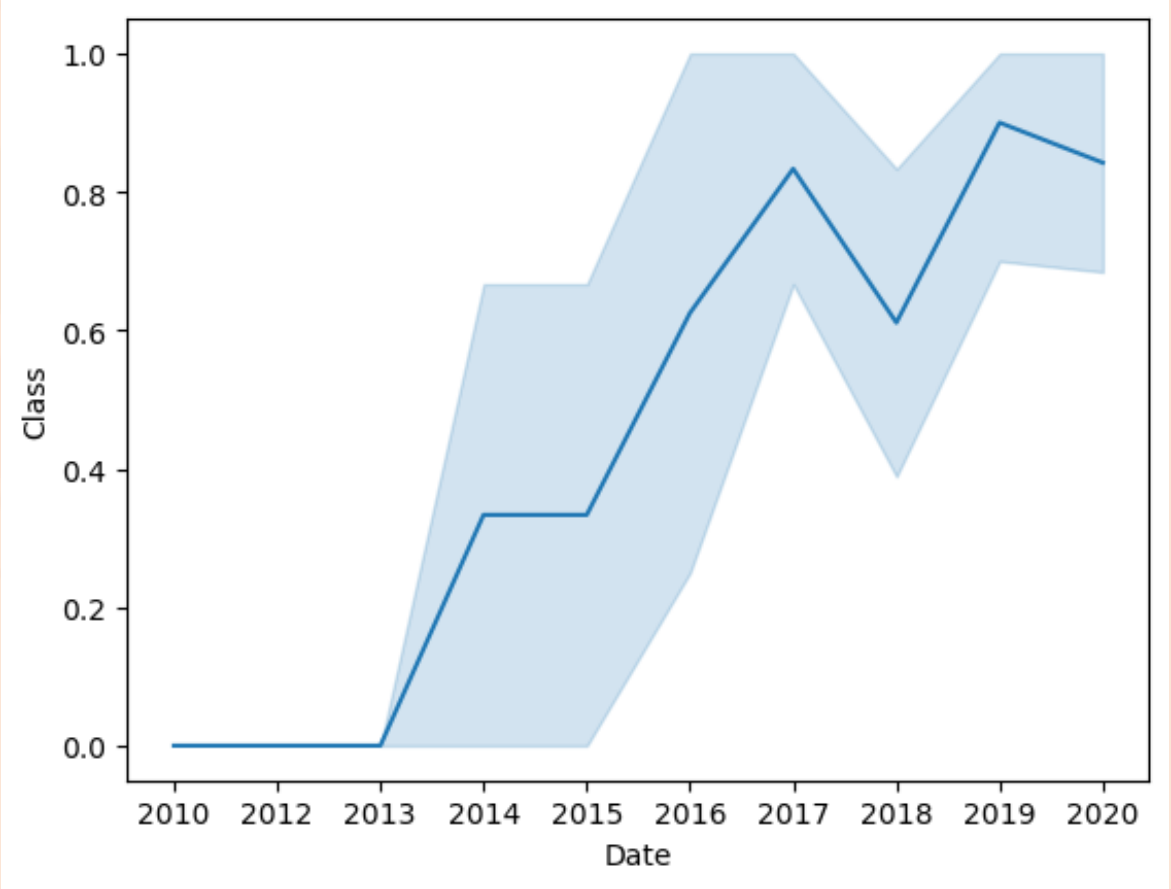




- LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit
- With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS.
- However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.

RESULTS – Historic success trend





- You can observe that the sucess rate since 2013 kept increasing till 2020

RESULTS – SQL

Launch sites:

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

Launch sites beginning with CCA:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (p
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 (p
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	N
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	N
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	N

Booster version with highest payload mass::

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

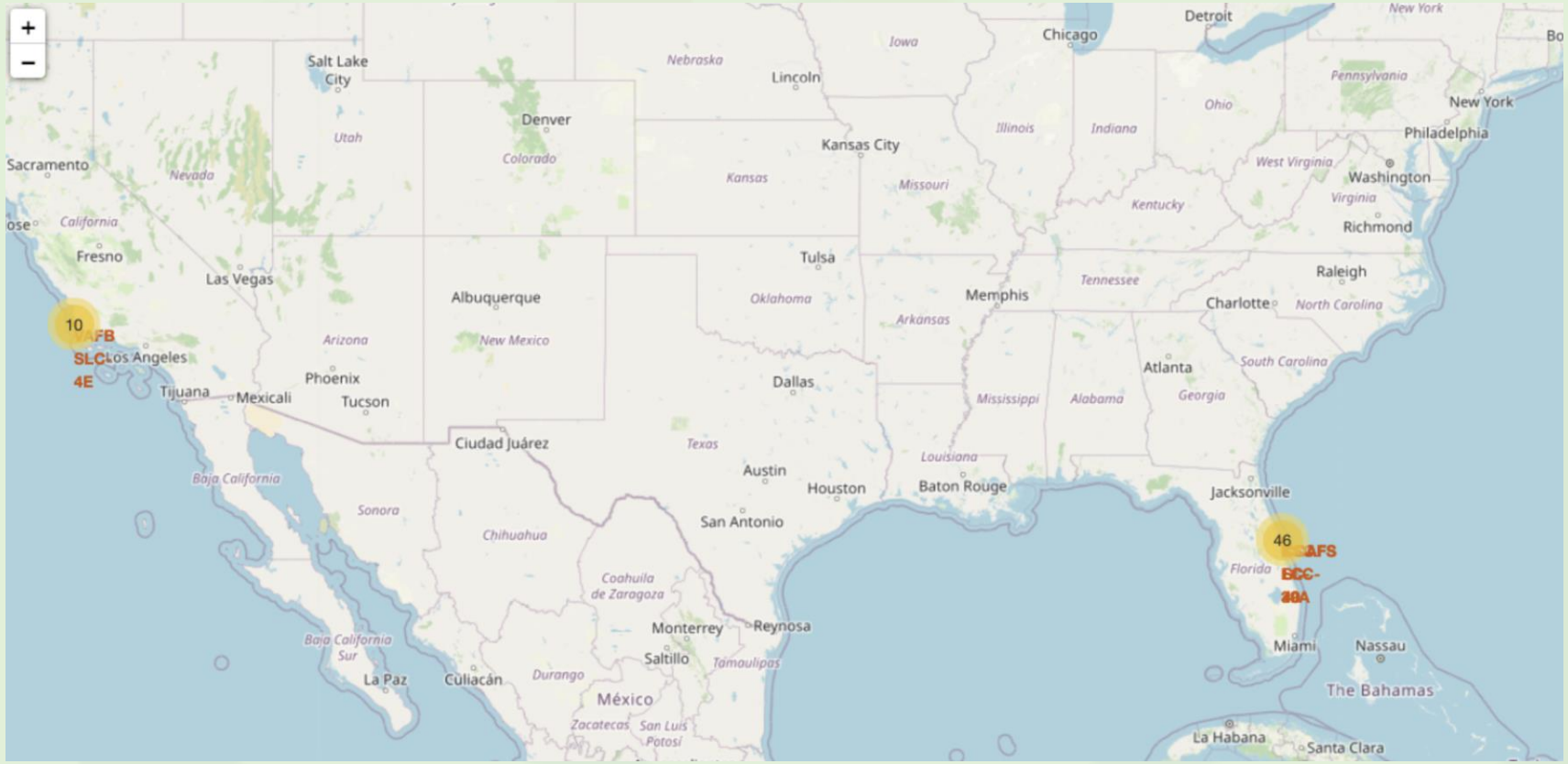
Booster version F9 v1.1:

AVG(PAYLOAD_MASS__KG_)
2928.4

1st historic launch:

Min(Date)
2010-06-04

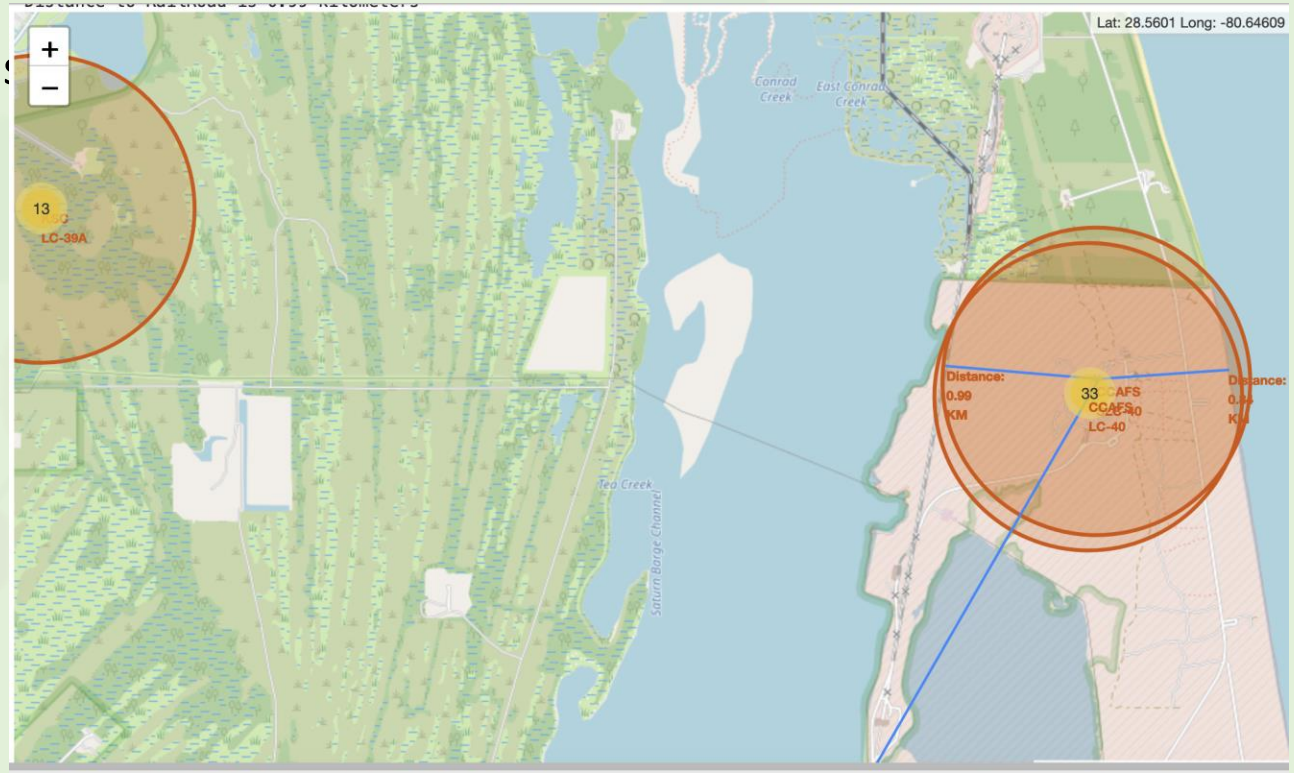
Interactive Maps



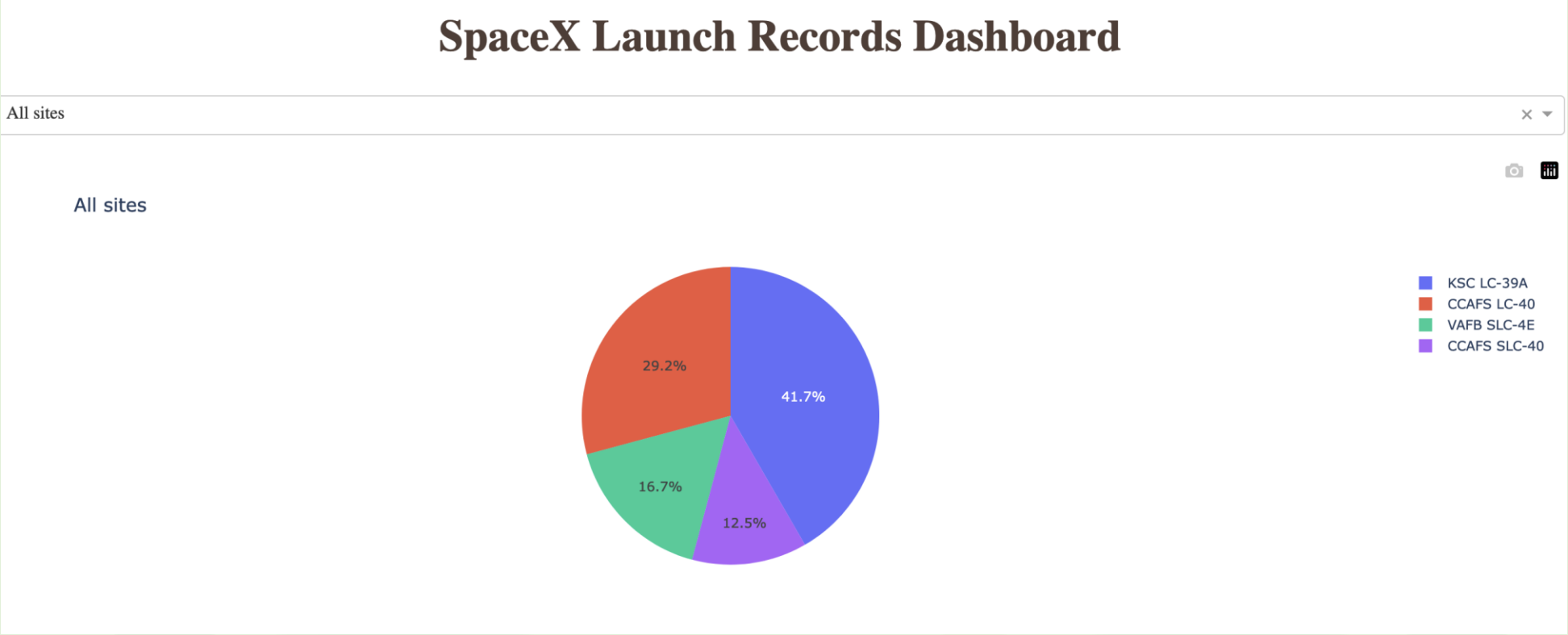
Total 56 launches per 3 major launch sites with all of them being close to the coastline

Interactive Maps

- All launch sites are in close proximity to railroads (<1km), but not close to the cities (33km).
- In addition, they are always next to the coast (0.85km) but not that close to highways.
- Coloured markers show the successful (green) landings and unsuccessful (red) ones

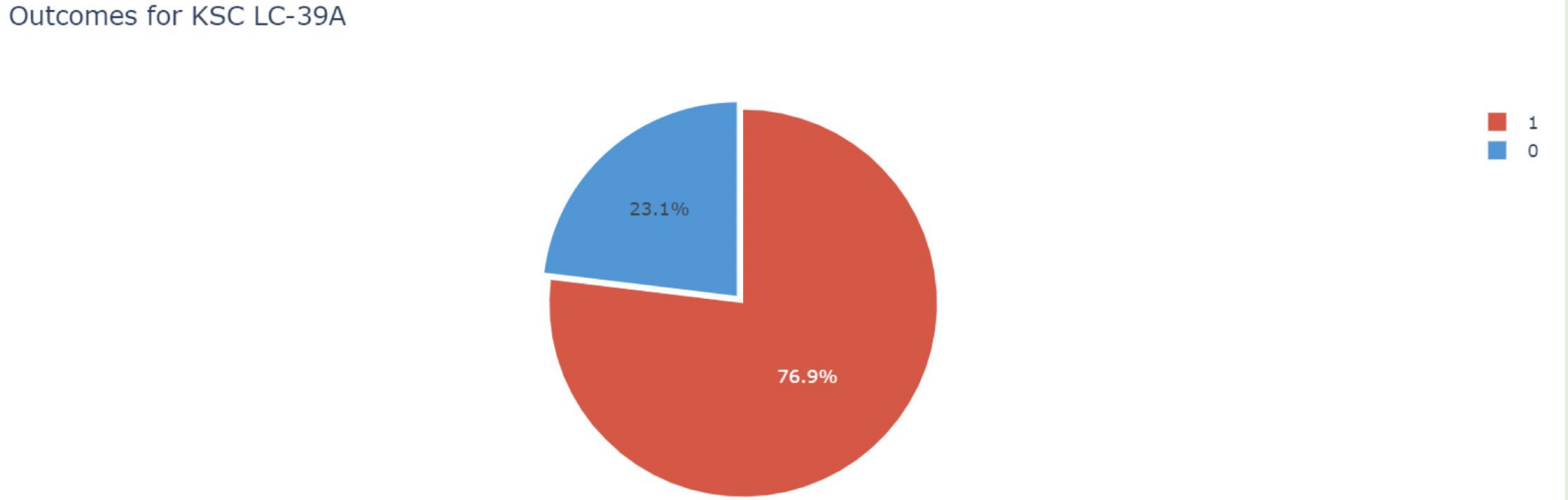


Net launches per site



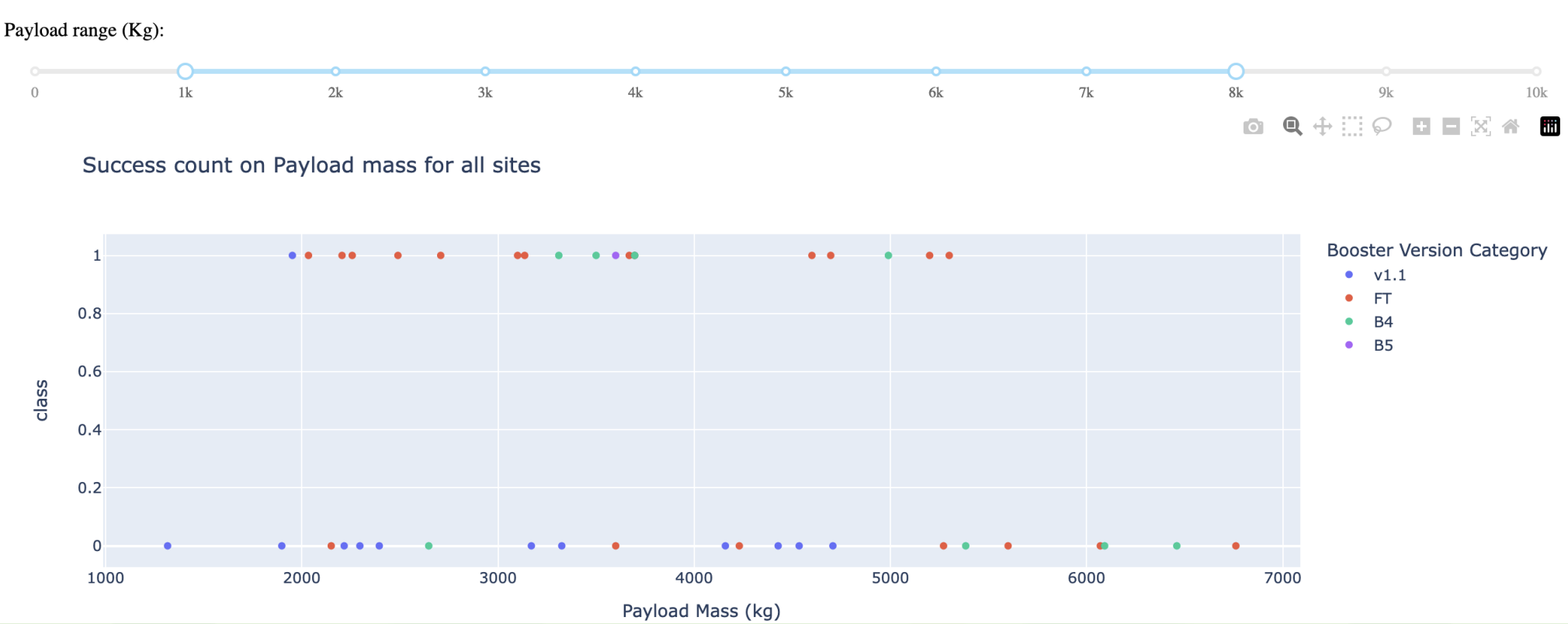
- A set of dashboards created to provide a quick understanding of the success rate per launch site
- For example, CCAFS SLC-40 demonstrates the lowest success rate of all sites

Launches for KSC LC-39A



- This is the KSC LC-39A individual site, showing a very high success rate of more than 75%

Payload effect and Booster Version



- It is clearly seen that there is more success with a lower payload with FT showing dramatically more counts than others

Predictive Analysis

```
tuned hpyerparameters :(best parameters)  {'C': 1.0, 'gamma': 0.03162277660168379, 'kernel': 'sigmoid'}
accuracy : 0.8482142857142856
```

- Four different models were created: LogisticRegression, Support Vector Machine, DecissionTree and K-Nearest Neighbours

- All of them gave the same level of accuracy at about 83.33%

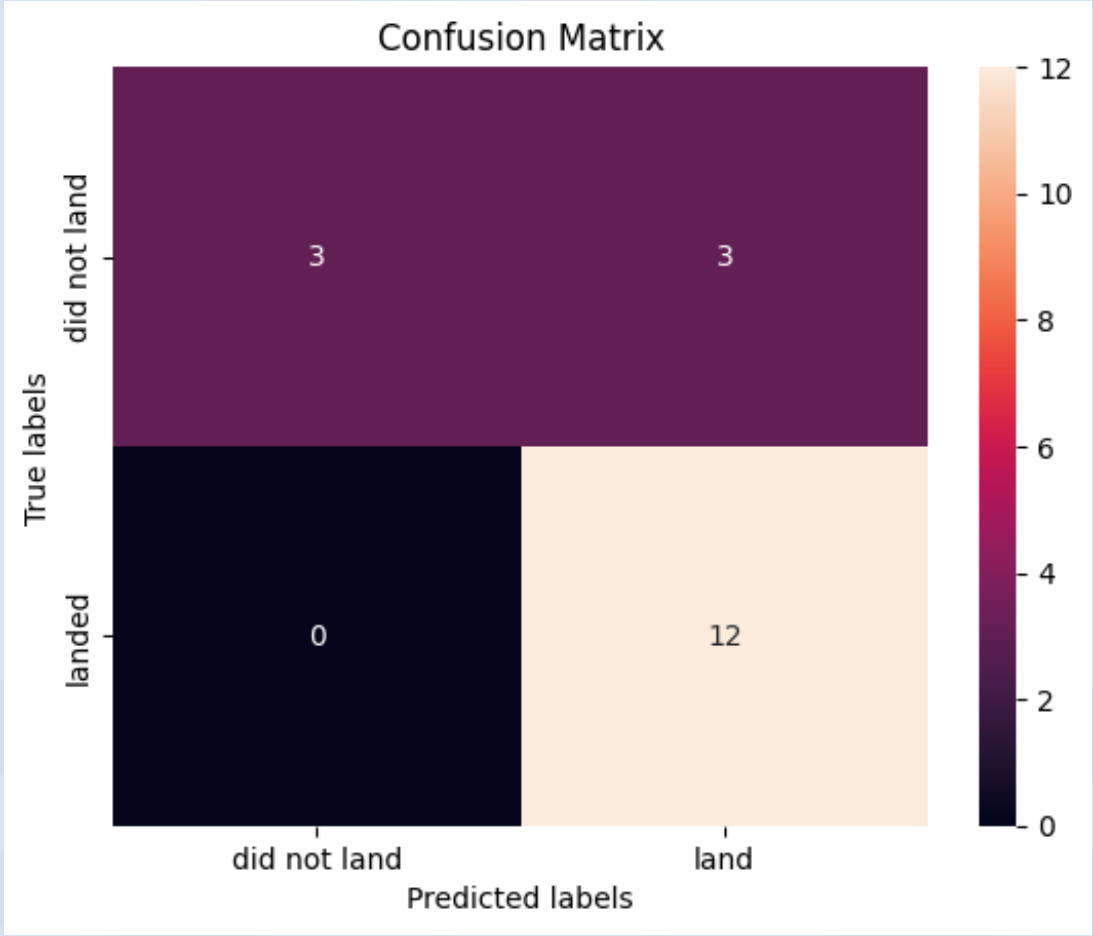
- However, the DecissionTree model showed a higher score in training dataset at 86% when compared to the rest

```
tuned hpyerparameters :(best parameters)  {'C': 1.0, 'gamma': 0.03162277660168379, 'kernel': 'sigmoid'}
File display : 0.8482142857142856
```

```
tuned hpyerparameters :(best parameters)  {'criterion': 'gini', 'max_depth': 4, 'max_features': 'sqrt', 'min_sample
s_leaf': 1, 'min_samples_split': 5, 'splitter': 'best'}
accuracy : 0.8607142857142858
```

```
tuned hpyerparameters :(best parameters)  {'algorithm': 'auto', 'n_neighbors': 10, 'p': 1}
accuracy : 0.8482142857142858
```

Predictive Analysis



- The confusion matrix was the same for all models

CONCLUSION

- According to the above results, it is seen that each attribute has a different effect on the success rate
- Consistently, launch site CCAFS SLC-40 showed the lowest rate, whereas KSC LC-39A showed the highest
- Typically, the higher the flight number, the higher the probability of having a successful launch
- Orbits SO followed by GTO have the lowest success rate
- Historically, the success rates keep increasing year by year and this seems to continue for the following ones
- All launch sites are located near a coast line and further away from the nearest city, which is expected
- All predictive models gave almost the same result of score with a poor prediction on non-successful counts, probably because they are of a smaller number in total dataset

IBM Developer

SKILLS NETWORK