Winning Space Race with Data Science

Duarte Pais
23.01.2025

Outline

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

Executive Summary

- This project aimed to assess simple data analysis, by:
    1. fetching SpaceX rocket launch data, from public REST APIs and web scraping webpages
    2. assess different ML algorithms to try to predict success of the landing of the first stage of SpaceX's rockets
    3. Create interactive maps to allow users to investigate some findings

Introduction

- SpaceX is a rocket launch company, whose competitive services are due to the fact that they recycle the first stage of their rockets. This recycling lowers the launching costs, leading to cheaper prices for their clients to place their devices into orbit.
- If, however, the first stage does not successfully land after a launch, SpaceX incurs additional costs similar to their competitor which do not recycle the first stage rocket, lowering the profit margins and potentially jeopardising their business.
- The idea is then to assess which factors influence the landing outcome of the first stage rocket.

Section 1

Methodology

Methodology

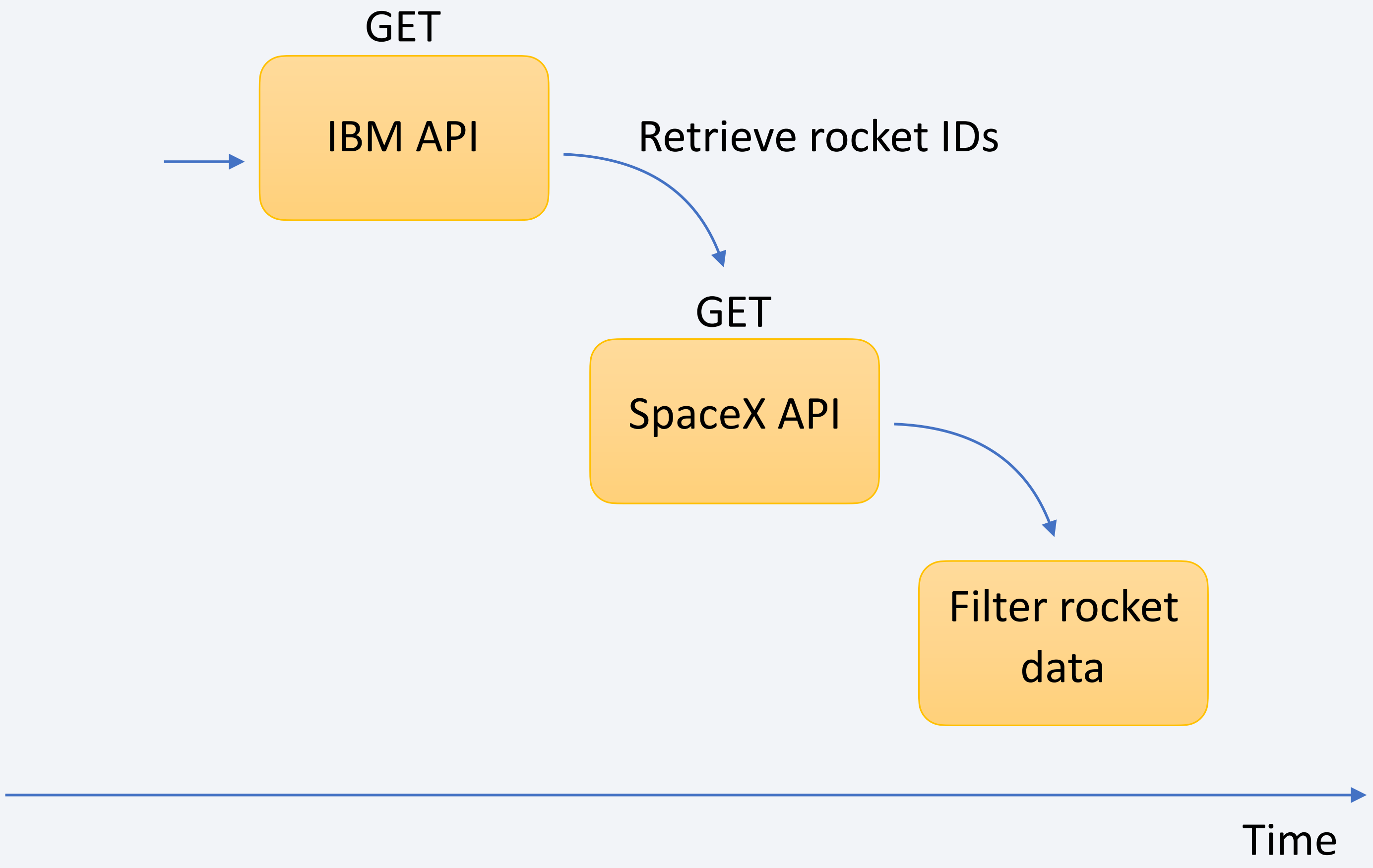
- Data collection methodology:
    - Describe how data was collected
- Perform data wrangling
    - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
    - How to build, tune, evaluate classification models

Data Collection

The data was:

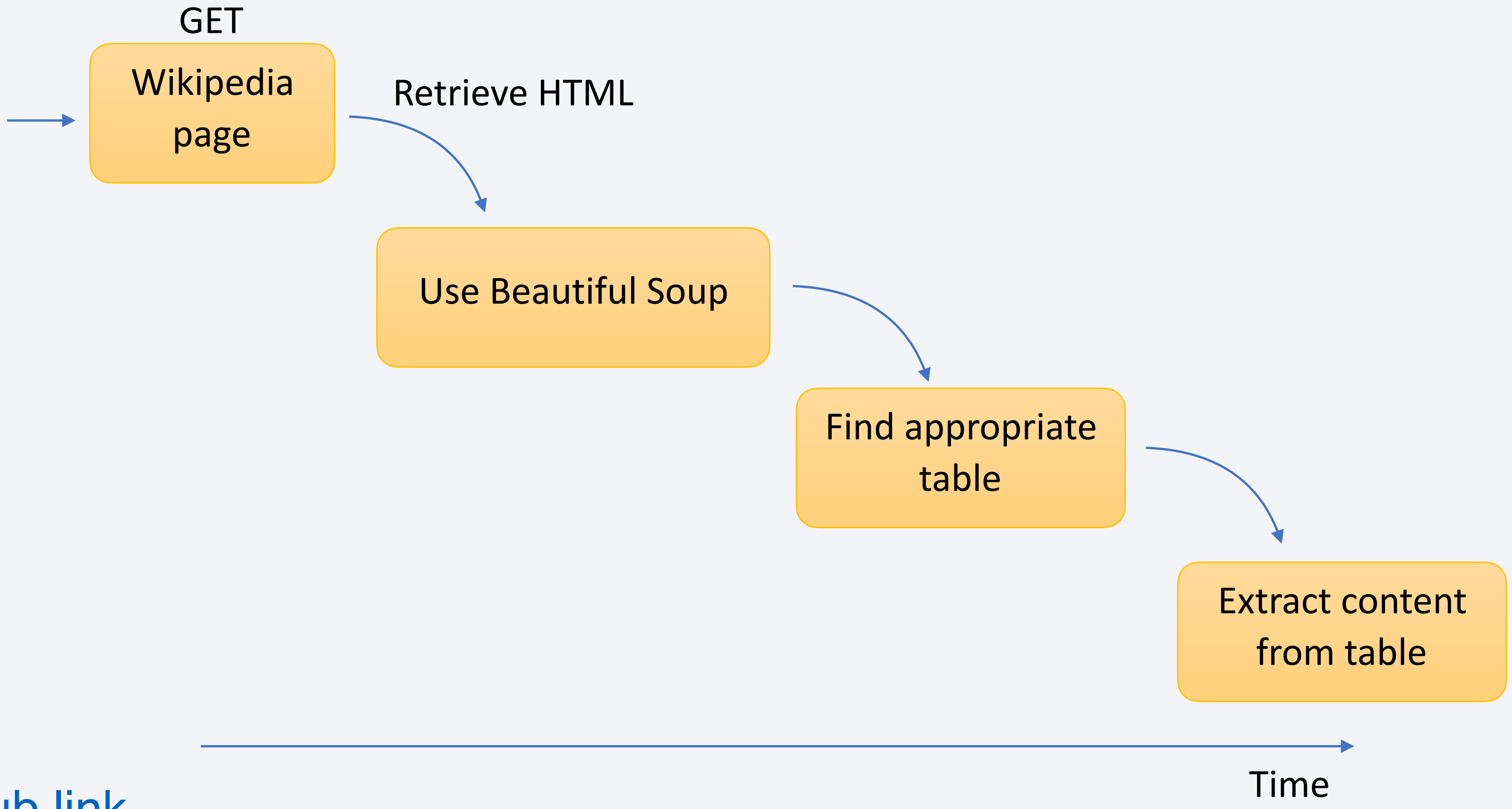
1. collected from RESTful APIs and
2. scraped from Wikipedia Webpages.

Data Collection – SpaceX API



- [Github link]

Data Collection - Scraping



- Github link

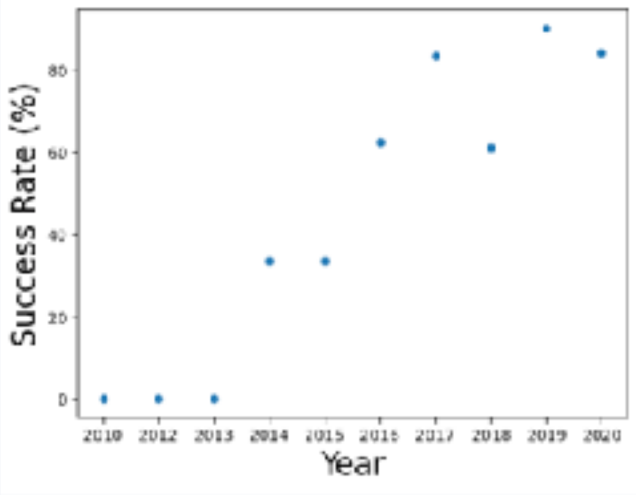
Data Wrangling

- The outcome of the first stage landings was assigned to a new column “Class" in the data frame.
- The content of the existing “Outcome" column was summarised to in the “Class" variable, to simply include Success (value 1) or Failure (value 0) information.

- Github link

EDA with Data Visualization

- An exploratory data analysis was done on the several variables that could have potentially influenced the class of the first stage landing. These were the orbit destination, launch site, payload mass and flight number.
- Below is the most relevant plot, showing that with time, SpaceX gained enough experience with the technology, approaching a 100% success rate



- [Github link]

EDA with SQL

- With SQL, the dataset was explored to return information on:
  - Different launch sites,
  - Total payload transported for a given client
  - Different landing outcomes
  - First success landing
  - Landing outcome statistics
  - Rockets taking the heaviest payloads
  - etc.
- Github link

Build an Interactive Map with Folium

- With Folium, the different launch sites were shown in a map, together with clustered information about the outcome of the landing of the first rocket stage.

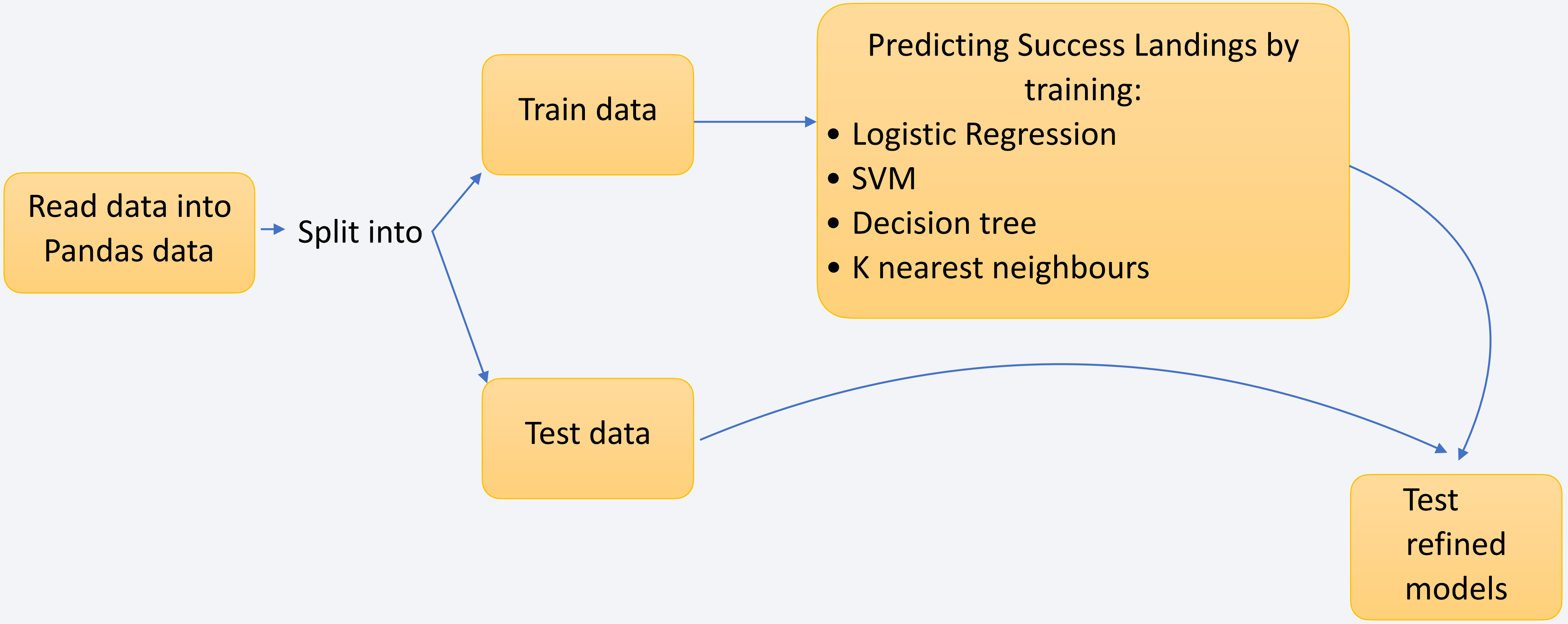
- [Github link]

Build a Dashboard with Plotly Dash

- The dashboard included:
  - A pie chart with the landing outcome per launching site
  - A scatter plot with the landing outcome vs. Payload mass for different booster versions.

- Github link

Predictive Analysis (Classification)



- [Github link]

Results

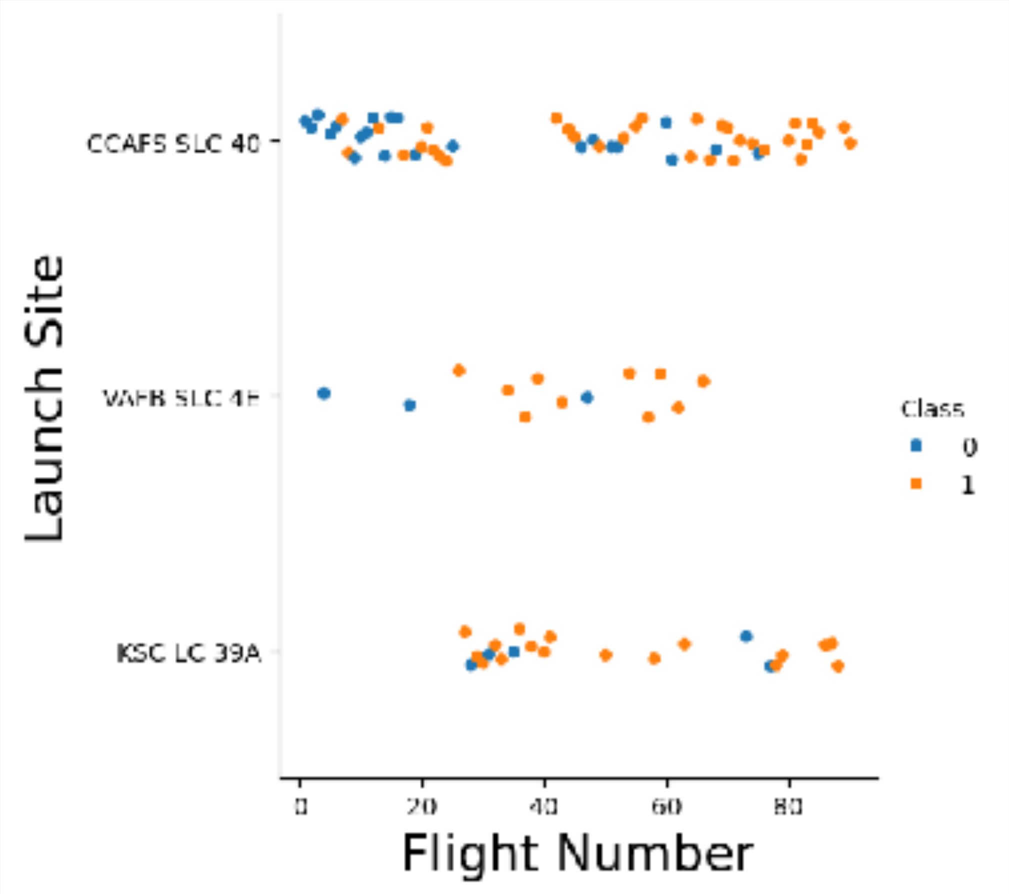
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

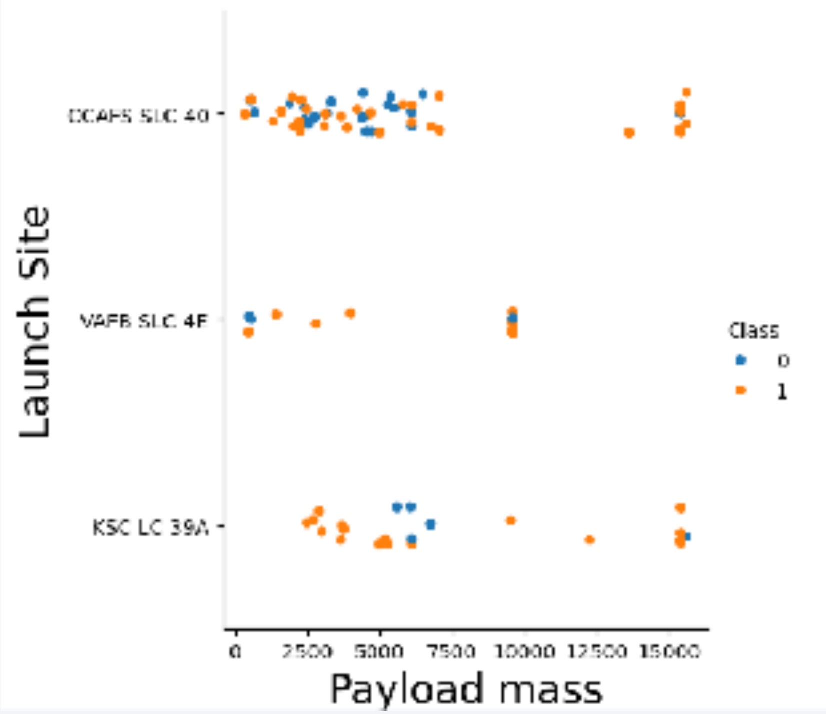
Flight Number vs. Launch Site

- It seems that during flight 20 and 40, the CCAFS SLC 40 site was not chosen by SpaceX for some reason.
- The data shows that with time, more successful landings were achieved, regardless of the launch site.



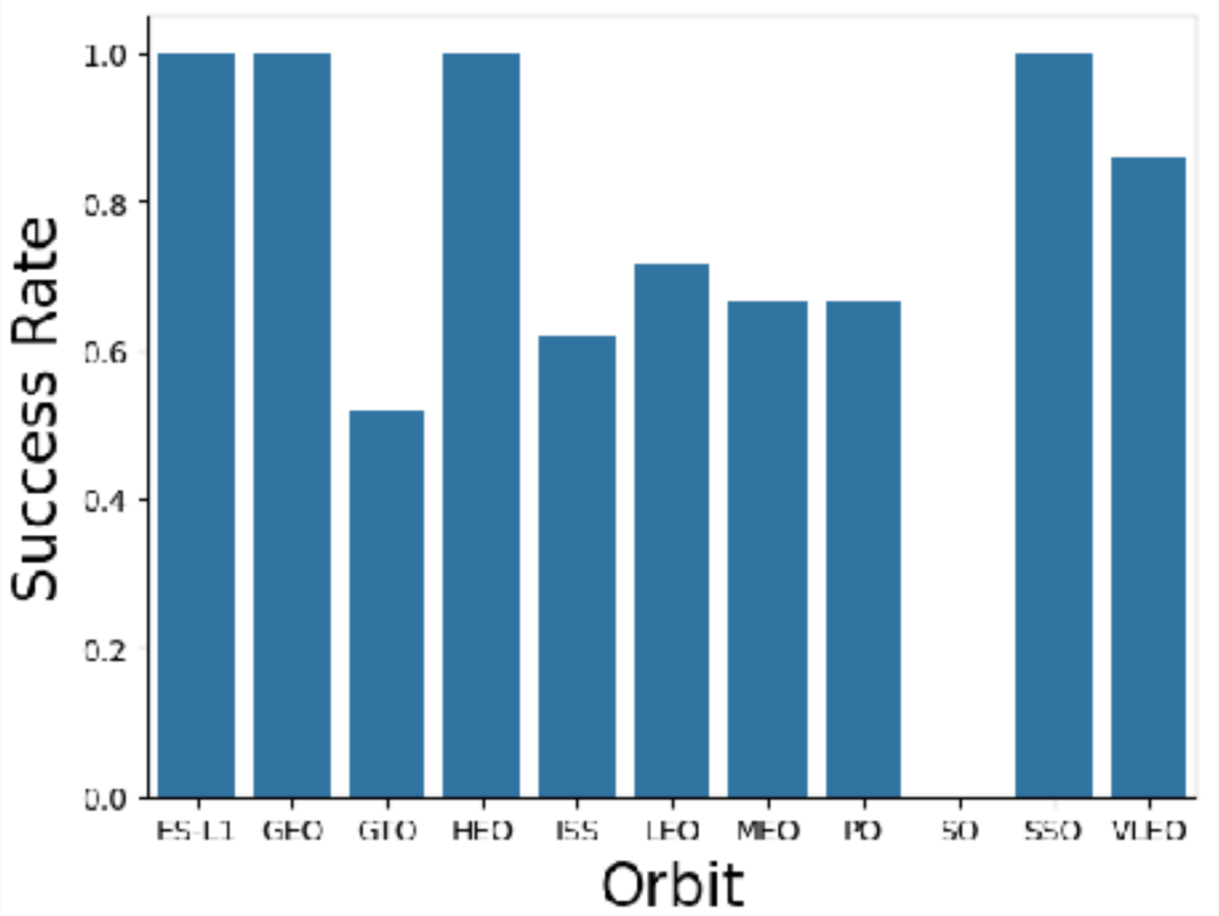
Payload vs. Launch Site

- More launches seem to have been made with lighter payloads, possibly to first test the technology.



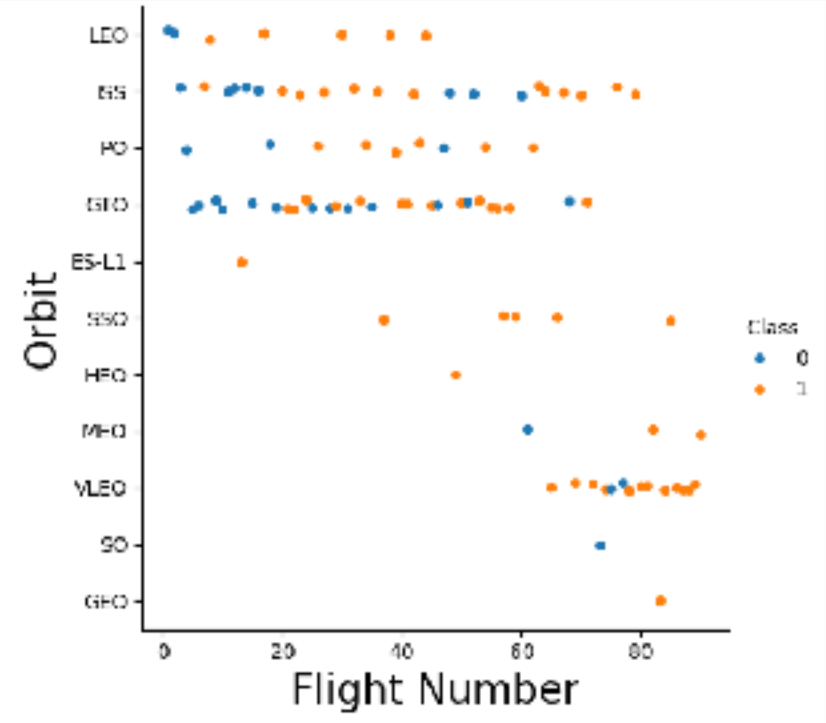
Success Rate vs. Orbit Type

- No successful landing has been made for rockets targeting SO.
- THe ES-L1, GEO, HEO and SSO have 100% success landing rates.



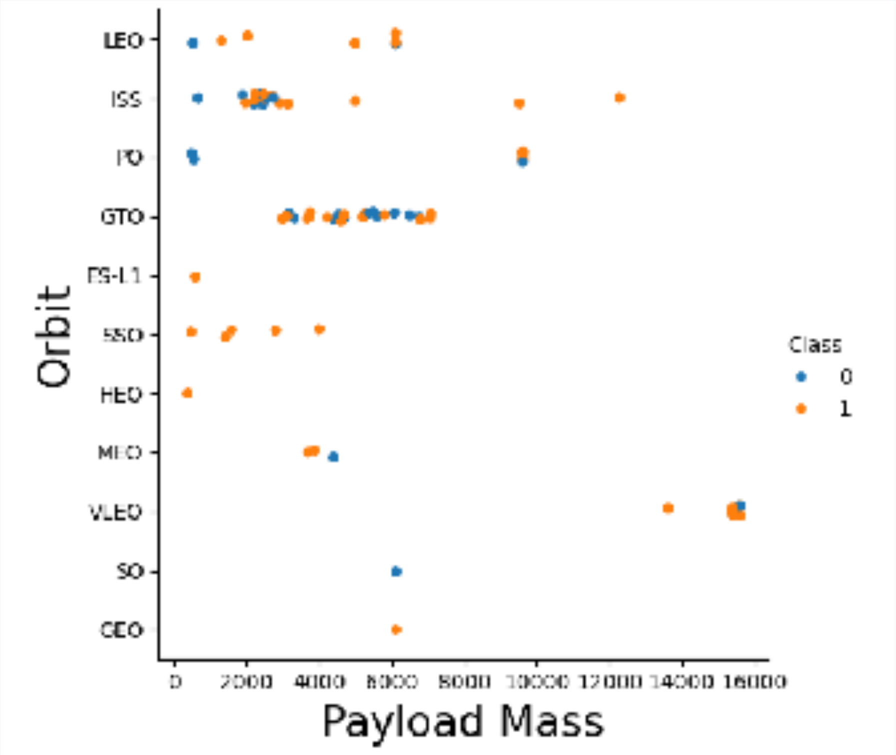
Flight Number vs. Orbit Type

- Initially, most rockets were targeting the LEO, ISS, PO and GTO, but in later times VLEO was the preferred target.



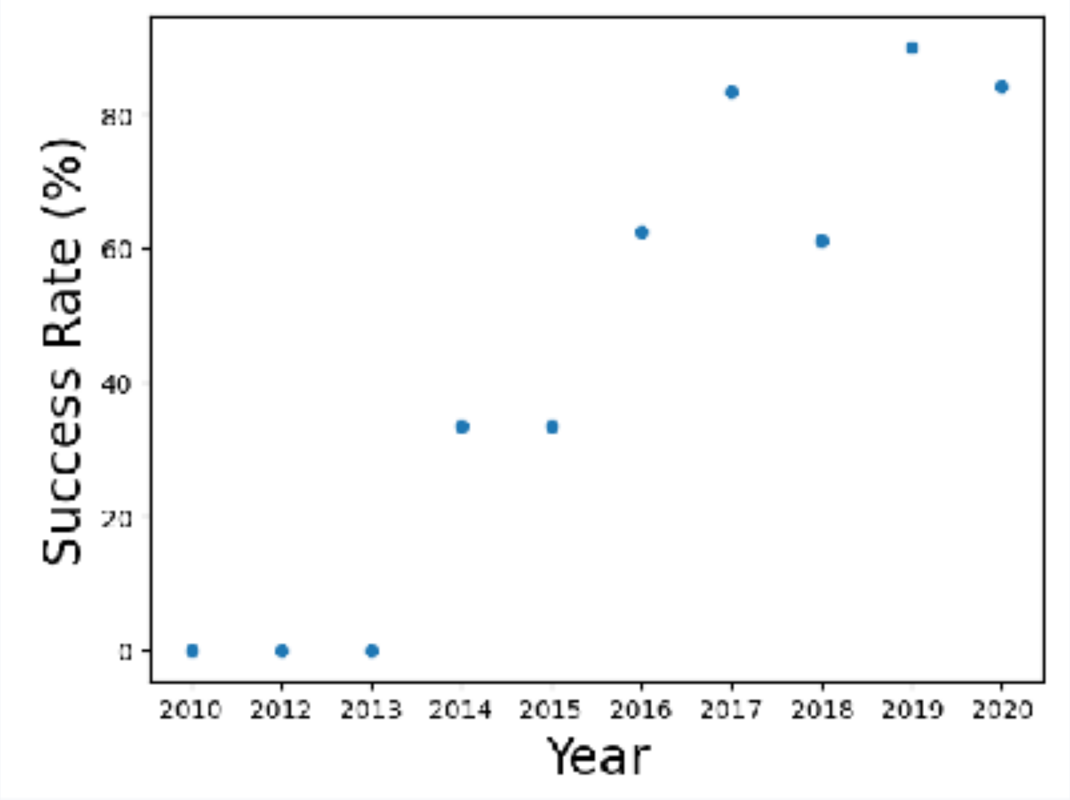
Payload vs. Orbit Type

- For almost all target orbits, most of the payloads are under 6000 kg, whereas larger Masses than 12000 kg seem to target VLEO.



Launch Success Yearly Trend

- The success rate increased as the years passed.



All Launch Site Names

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

Launch Site Names Begin with 'CCA'

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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	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 carried by boosters from NASA is 45596 kg

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 is 2928.4 kg

First Successful Ground Landing Date

- The date of the first successful landing outcome on ground pad is 2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- The boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000 are:
  - F9 FT B1022
  - F9 FT B1026
  - F9 FT B1021.2
  - F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- The total number of successful is 61 and of failure mission outcomes is 10

Boosters Carried Maximum Payload

- The booster versions which have carried the maximum payload mass are:
  - 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

Month	Booster_Version	Launch_Site	Landing_Outcome
1	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
2	F9 v1.1 B1013	CCAFS LC-40	Controlled (ocean)
3	F9 v1.1 B1014	CCAFS LC-40	No attempt
4	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)
4	F9 v1.1 B1016	CCAFS LC-40	No attempt
6	F9 v1.1 B1018	CCAFS LC-40	Precluded (drone ship)

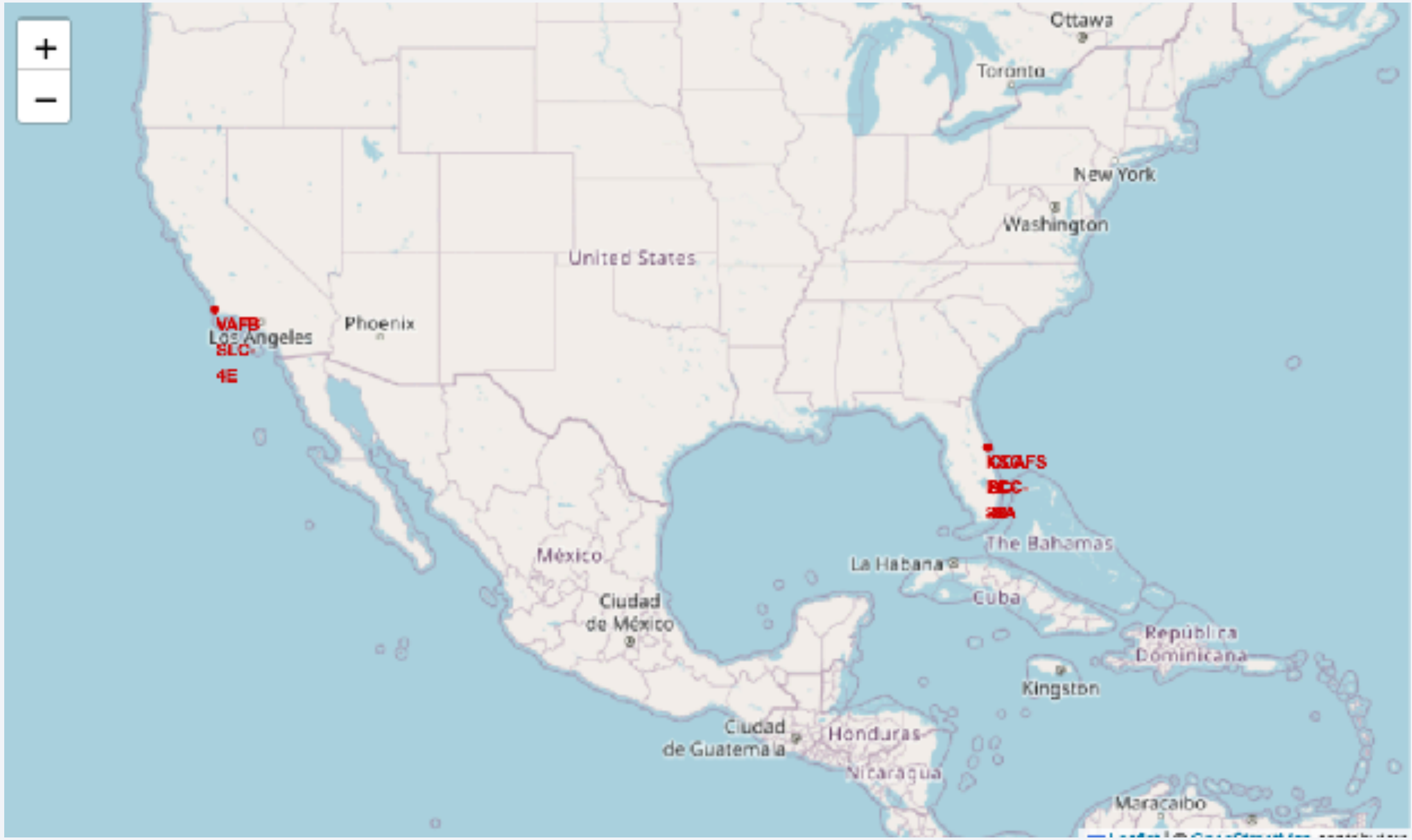
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

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

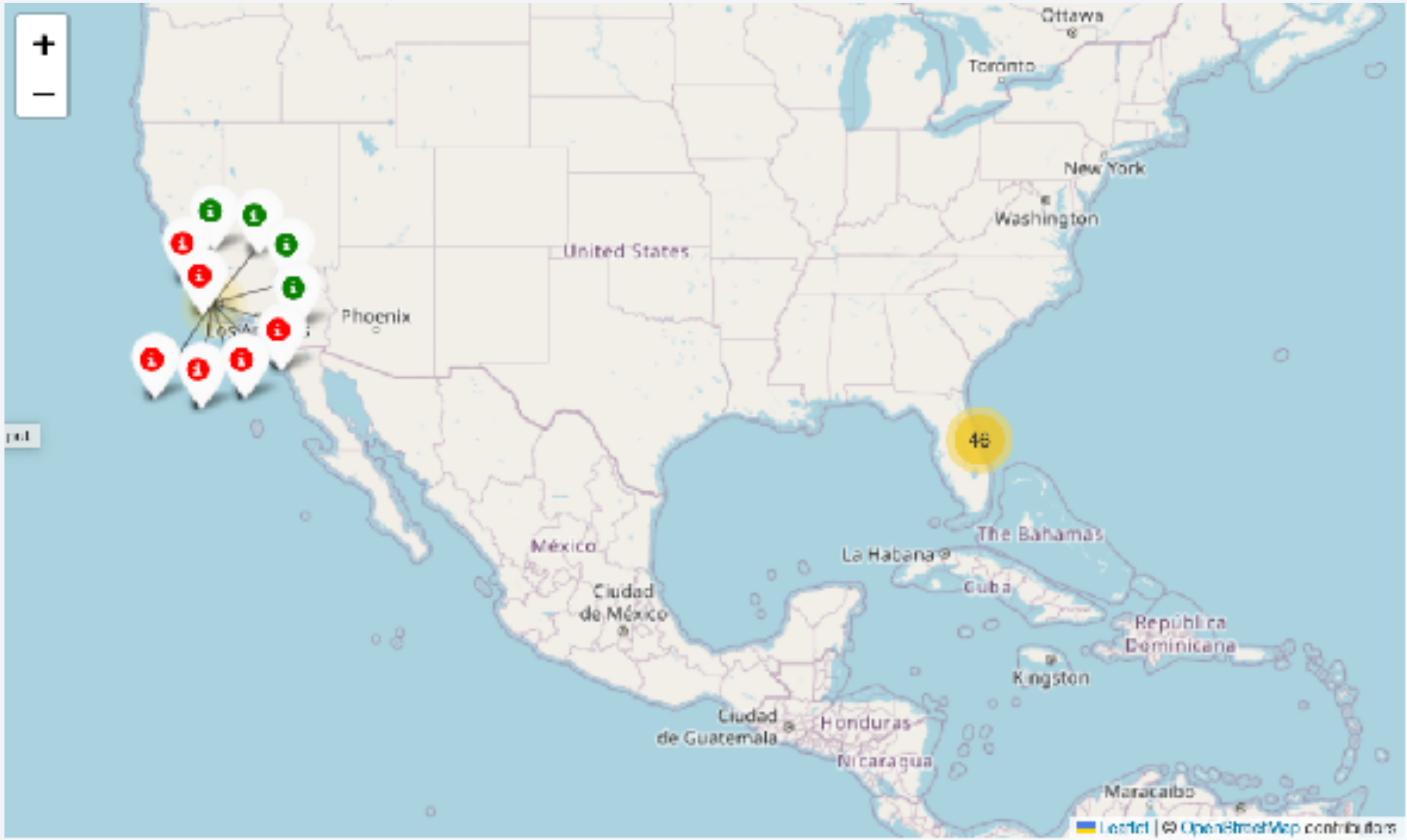
Section 3

Launch Sites Proximities Analysis

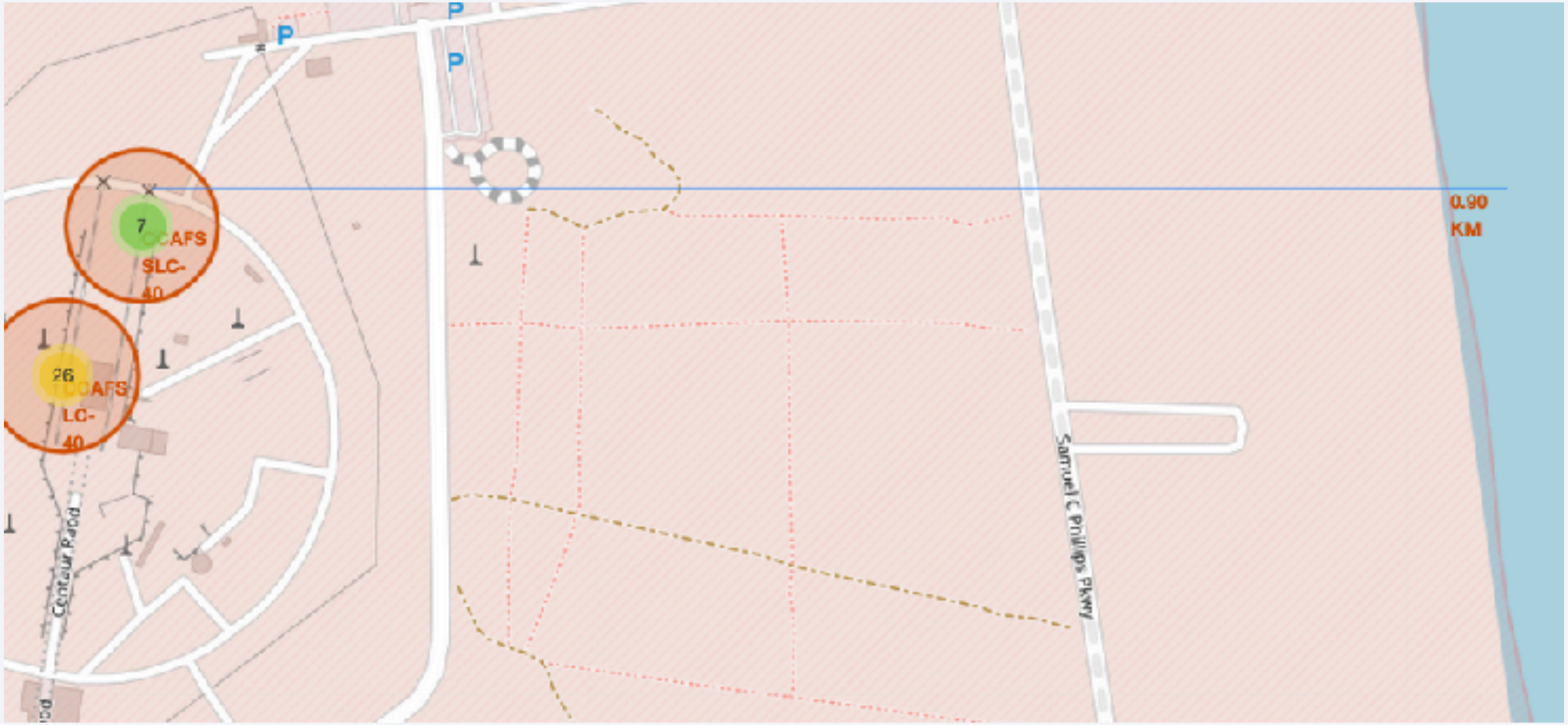
Locations of launching sites



Success rates in Californian launching site



Distance from Floridian Launching site to coast

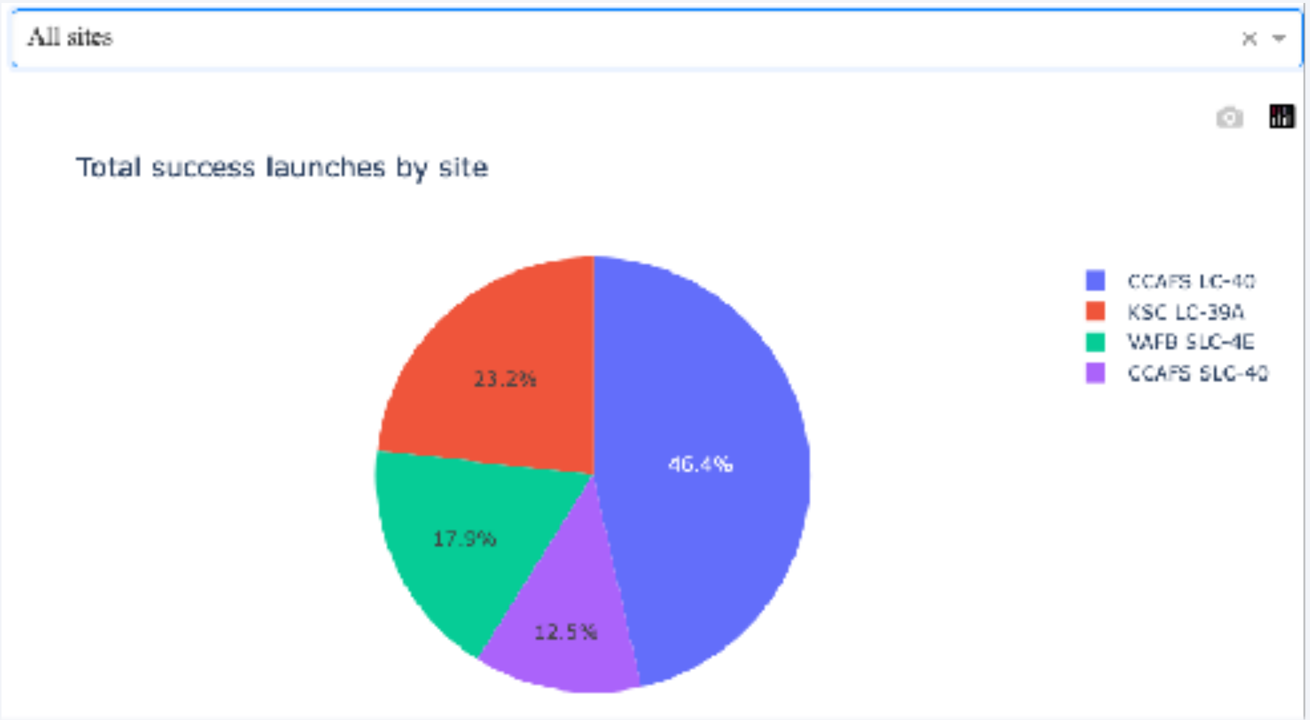


Section 4

Build a Dashboard with Plotly Dash

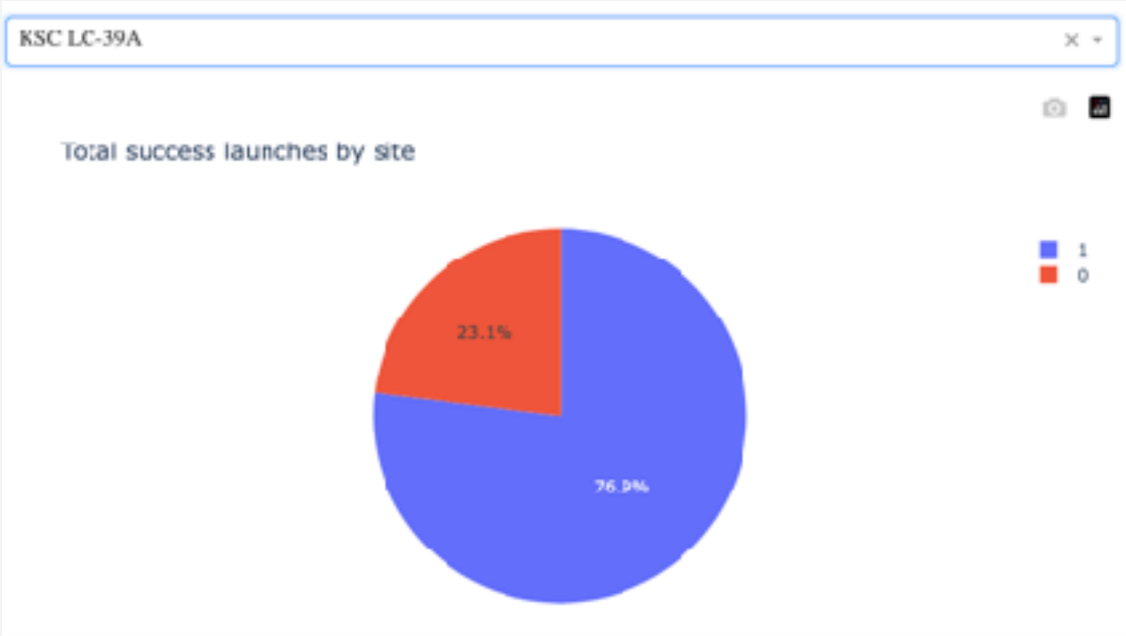
Total launches per site

- Almost half of the launches took place in CCAFS LC-40



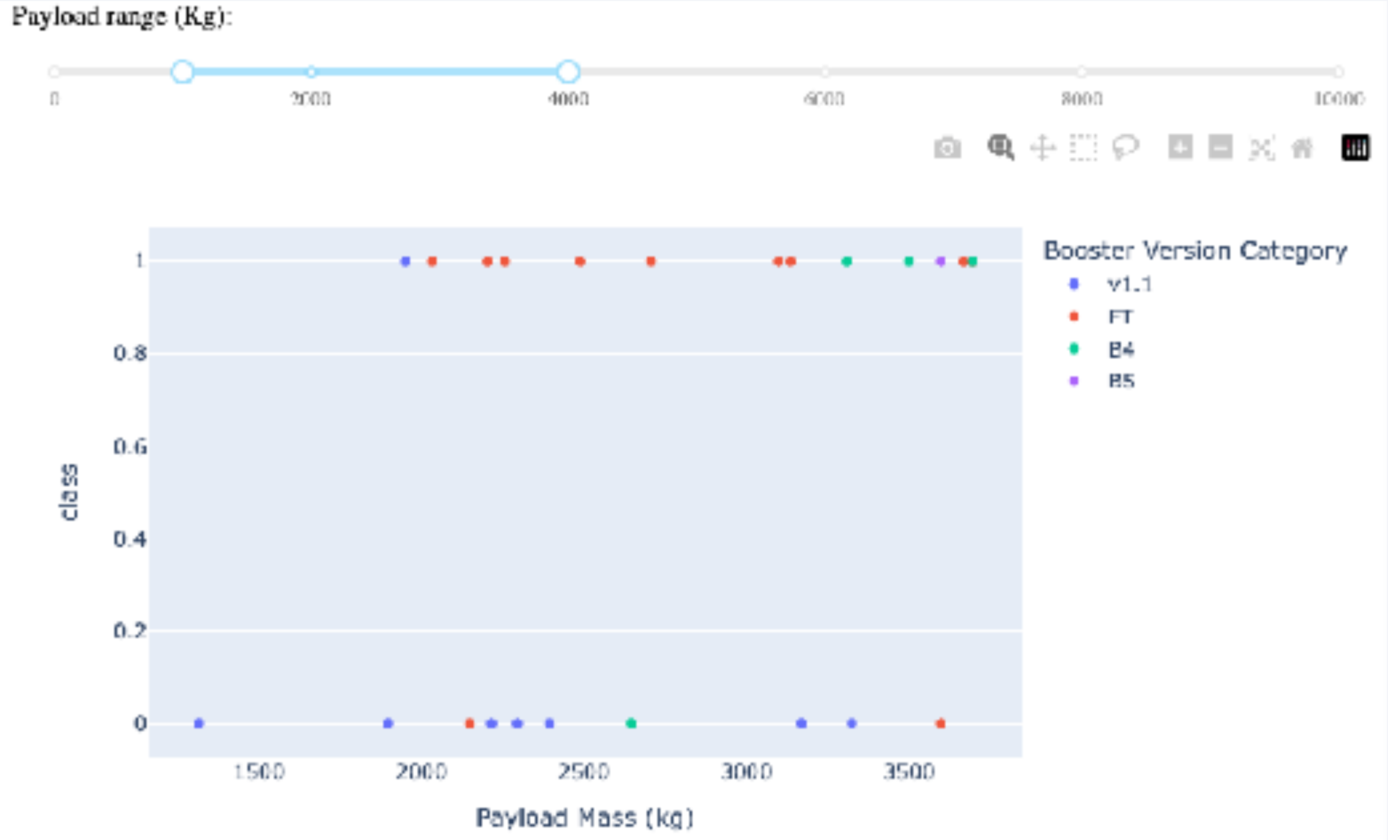
Most successful launch site

- KSC LC-39A has just over 3/4 of success rate.



Success vs. Payload mass for different booster versions

- Across the chosen payload mass interval, the success rate seems to be unaffected by it.
- It seems that for the B4 booster version, higher payloads lead to higher success rates.

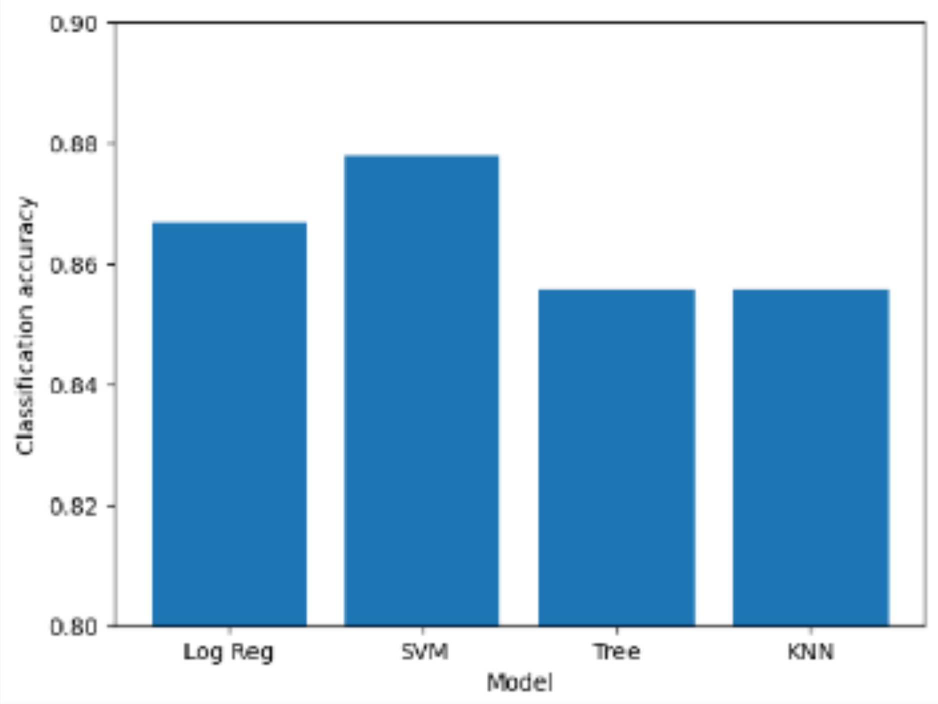


Section 5

Predictive Analysis (Classification)

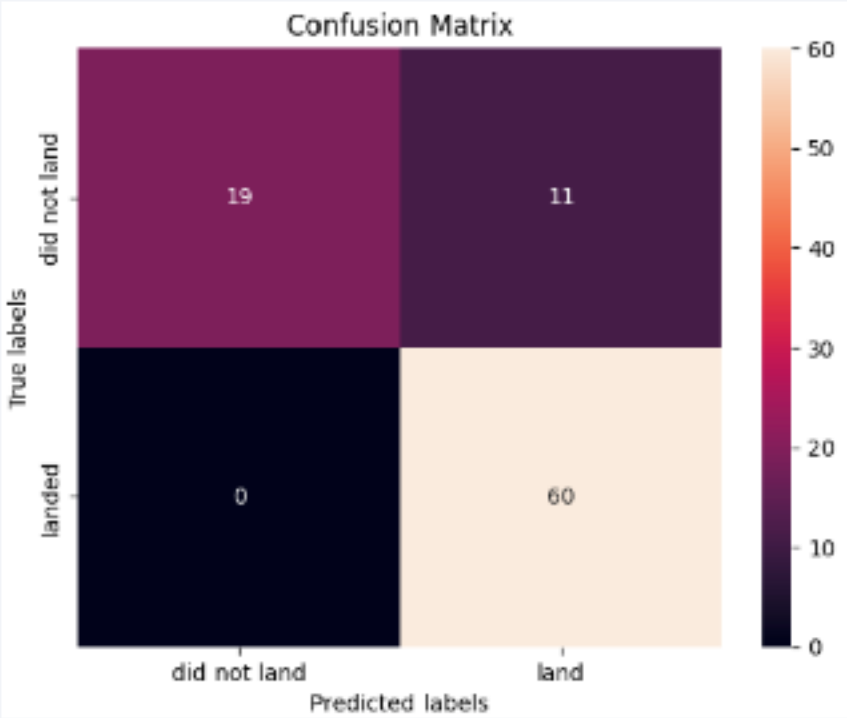
Classification Accuracy

- The optimised SVM model has the best accuracy of all models.



Confusion Matrix

- The SVM predicts no False Negatives, that is, for all landing attempts, it correctly predicts landings for all real landings.
- However, it is not perfect and there are some False Positives, i.e. it predicts some successful landings for some real failed landings attempts.



Conclusions

- This was a comprehensive exercise to play around with real world data and explore the different visualisation tools that are available and to investigate the SK-Learn library
- However, the dataset is actually quite small and incomplete to successfully develop models that can comprehensively predict the different success variables for a rocket launch.
- A more complete attempt would need better data sources.

Appendix

- Nothing to add.

Thank you!