IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Alexandr Nikitin

8th June 2024

Outline

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

Executive Summary

- Data was webscraped and pre-processed in a fairly standard way.
- The visualization of the data suggests that there are features distinguishing between successful and failed launches, and ML modelling could thus be viable.
- Four different models (and various hyperparameters) were inspected.
- The model with 80 % recall seems promising, considering that the cost of the flights by the competition are about 2.6times more costly.

Introduction

- We are trying to predict whether a SpaceX Falcon 9 rocket (more precisely it's first stage) will perform a successful landing, thus significantly cutting down the costs of the flight. SpaceX offers launches from about 60 mil USD, while the competition prices start at 160 mil USD.
- Large portion of the unsuccessful landings are planned, yet sometimes a failure occurs.

The question is:

Is it possible to reliably predict whether a rocket will successfully land based on a fairly small available dataset (less then 100 records)?

Section 1

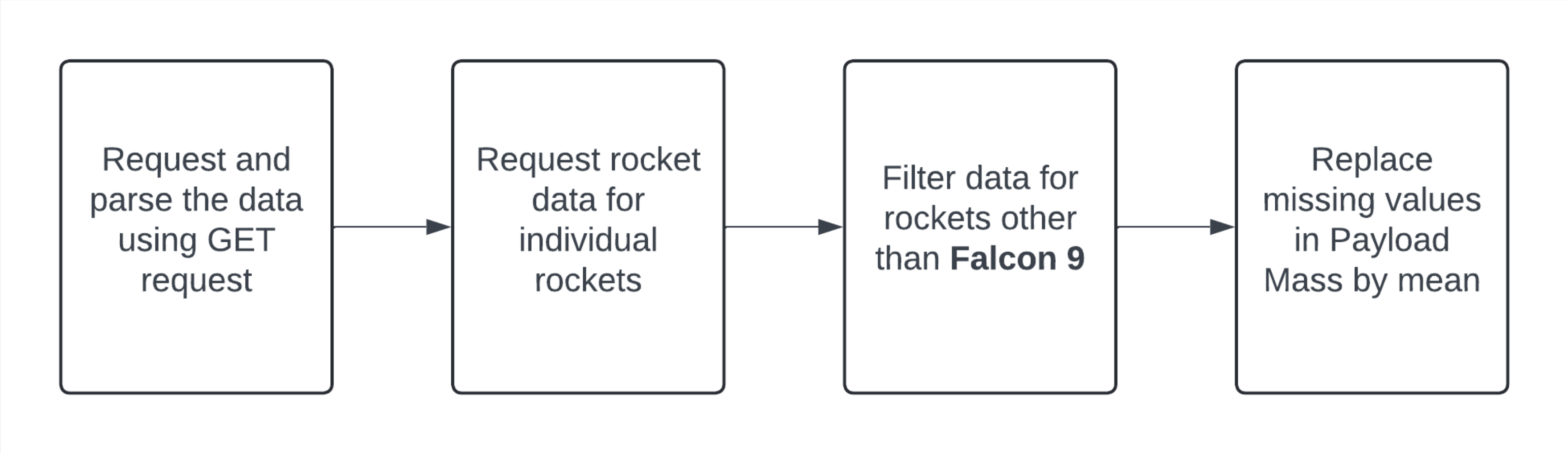
Methodology

Methodology

Executive Summary

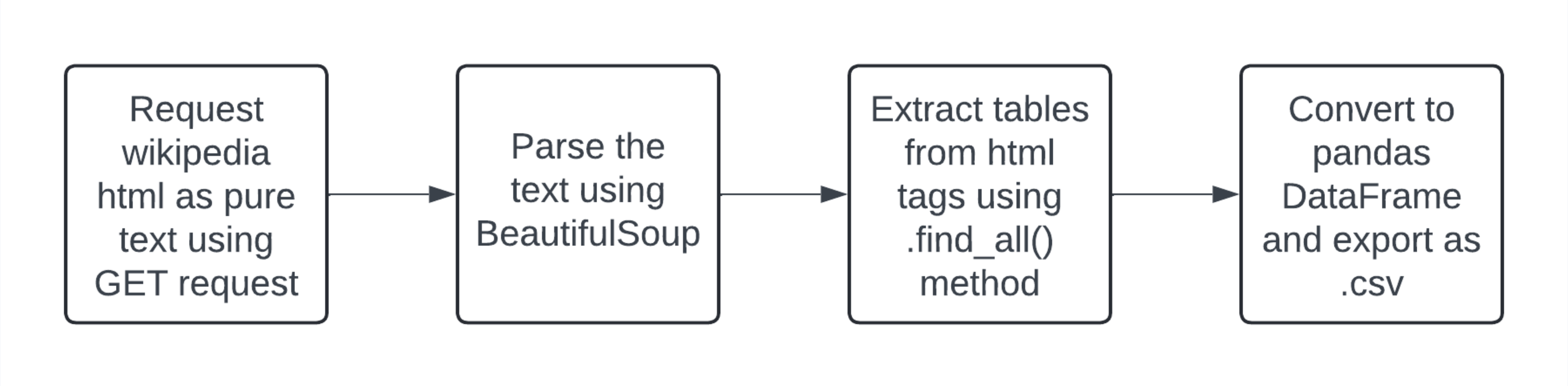
- Data collection methodology:
- Perform data wrangling
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

Data Collection – SpaceX API



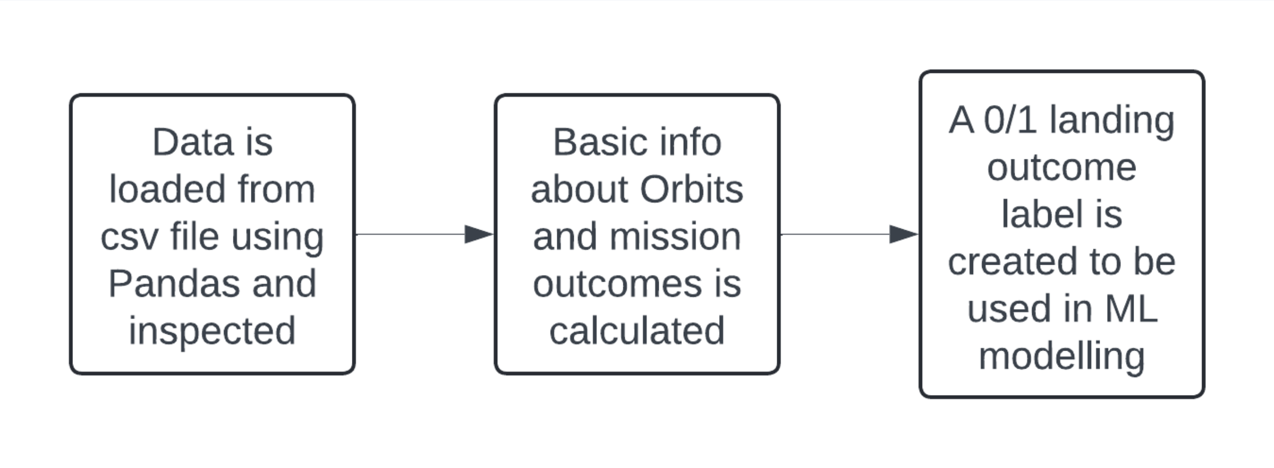
https://github.com/ValheruKirito/IBM_Data_Science_for_Professionals/blob/main/Capstone/01_Mod1_jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping



https://github.com/ValheruKirito/IBM_Data_Science_for_Professionals/blob/main/Capstone/02_Mod1_jupyter-labs-webscraping.ipynb

Data Wrangling



https://github.com/ValheruKirito/IBM_Data_Science_for_Professionals/blob/main/Capstone/03_Mod1_labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- Effect of time (and flight number) on Success of a mission with respect to Launch site and Orbit site was inspected via scatter plots
- Effect of mission payload mass on Success of a mission with respect to Launch site and Orbit site was also inspected via scatter plots
- Success rate of launches to individual orbits was inspected via barplot
- Evolution of the overall success rate in time is visualized in a lineplot

https://github.com/ValheruKirito/IBM_Data_Science_for_Professionals/blob/main/Capstone/04_Mod2_jupyter-labs-eda-dataviz.ipynb

EDA with SQL

SQL queries were used to inspect various information about the dataset, for example:

- Total payload mass delivered during missions for NASA (CRS)
- Average payload carried by Falcon 9 v1.1 boosters
- When was the first successful landing on a ground pad

https://github.com/ValheruKirito/IBM_Data_Science_for_Professionals/blob/main/Capstone/05_Mod2_jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Individual launch sites were marked in interactive map using Folium
- Each launch site was supplemented with markers marking Success, or Failure of individual mission launched from a given facility

https://github.com/ValheruKirito/IBM_Data_Science_for_Professionals/blob/main/Capstone/06_Mod3_lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- An interactive Dashboard covering Success or Failure of mission with respect to various aspects was developed
- The user can inspect data by choosing individual launch sites and payload ranges, or overall statistics
- It is possible to inspect success rate for each individual launch site in a pie chart and to inspect success with respect to booster type, launch site and carried payload mass

https://github.com/ValheruKirito/IBM_Data_Science_for_Professionals/blob/main/Capstone/07_Mod3_Dashboard.py

Predictive Analysis (Classification)

- Four classification models were inspected in order to predict successful landing of a rocket – Logistic Regression, Support Vector Classifier, Decision Tree and k Nearest Neighbours
- Grid Search Cross Validation with 10 folds was performed to identify the best model
- Accuracy, Precision, Recall and F1-scores are used to select the best model

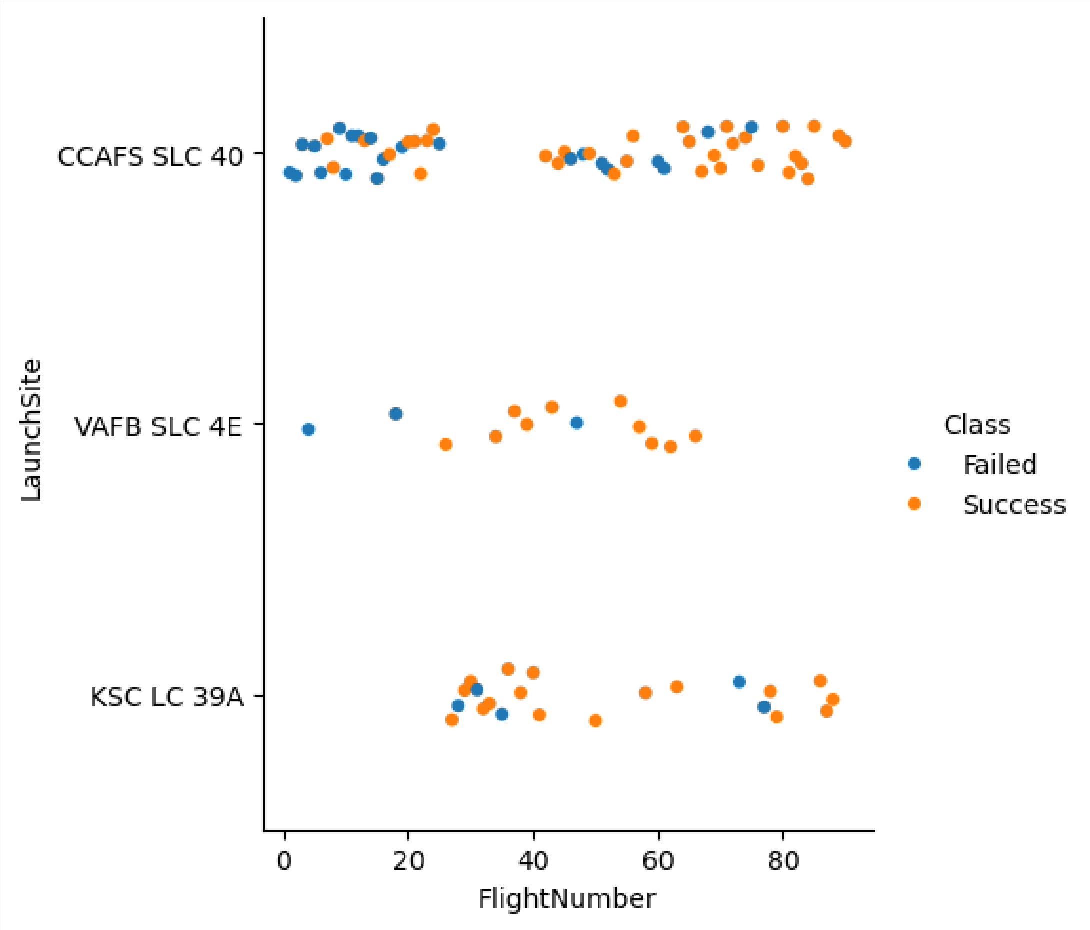
https://github.com/ValheruKirito/IBM_Data_Science_for_Professionals/blob/main/Capstone/08_Mod4_SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

Section 2

Insights drawn from EDA

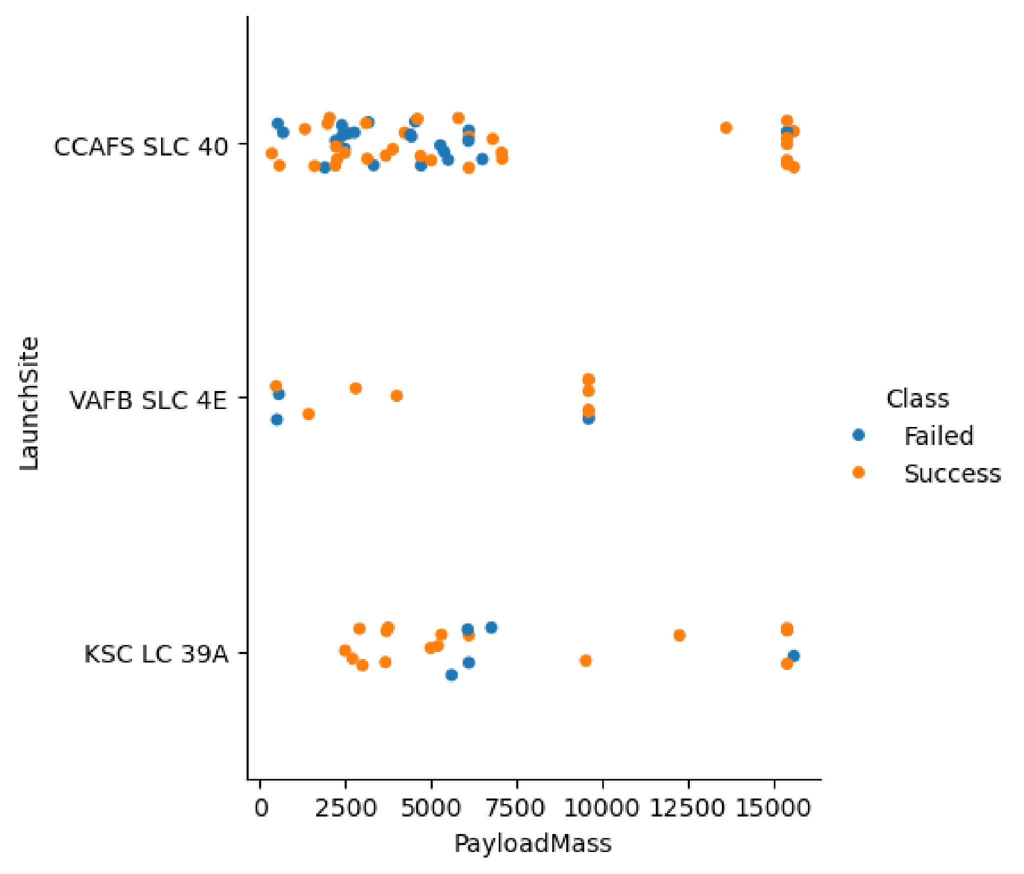
Flight Number vs. Launch Site

- Highest number of flights launched from CCAFS SLC 40
- Success of the missions is more common for launches with higher flight numbers (hence in time as presented later on)
- CCAFS SLC 40 has a lot of Failed launches in the lower Flight Number region (presumably original test flights)



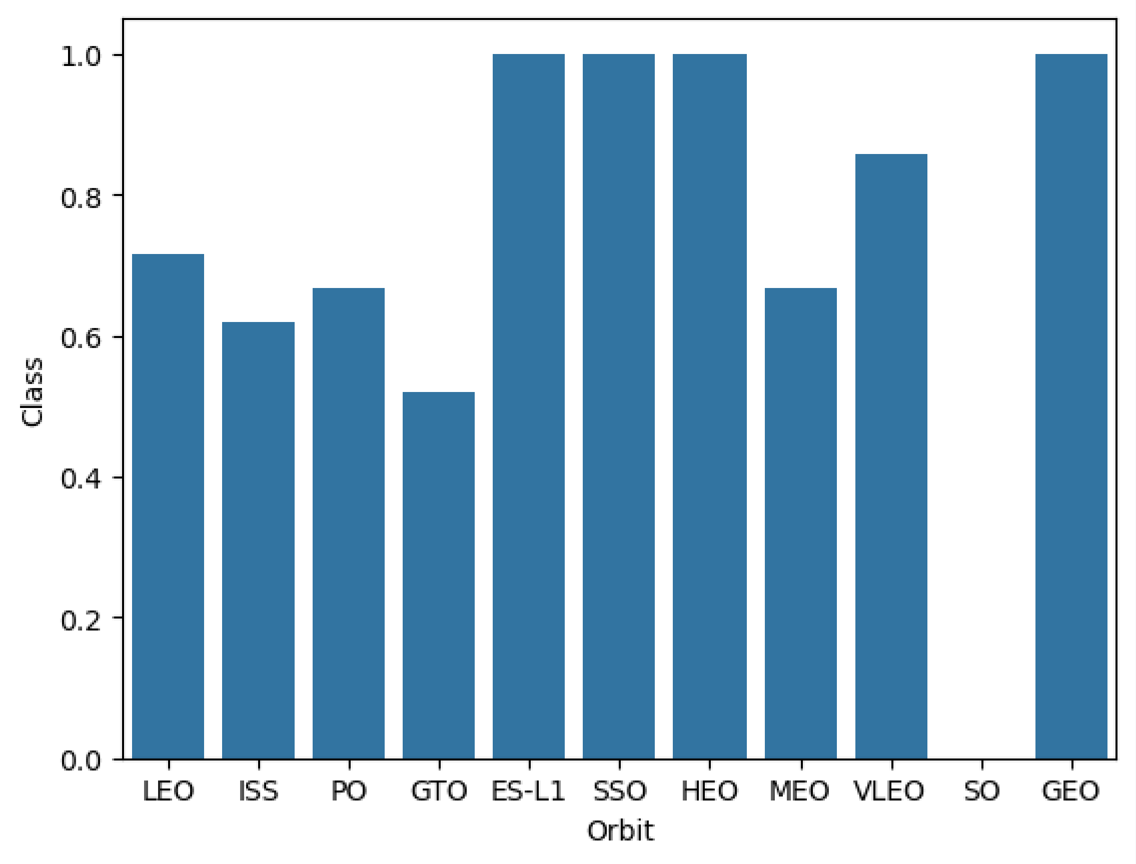
Payload vs. Launch Site

- Very high payloads seem to be quite successful.
- Payloads launched from VAFB SLC 4E are capped at 10 000 kg.



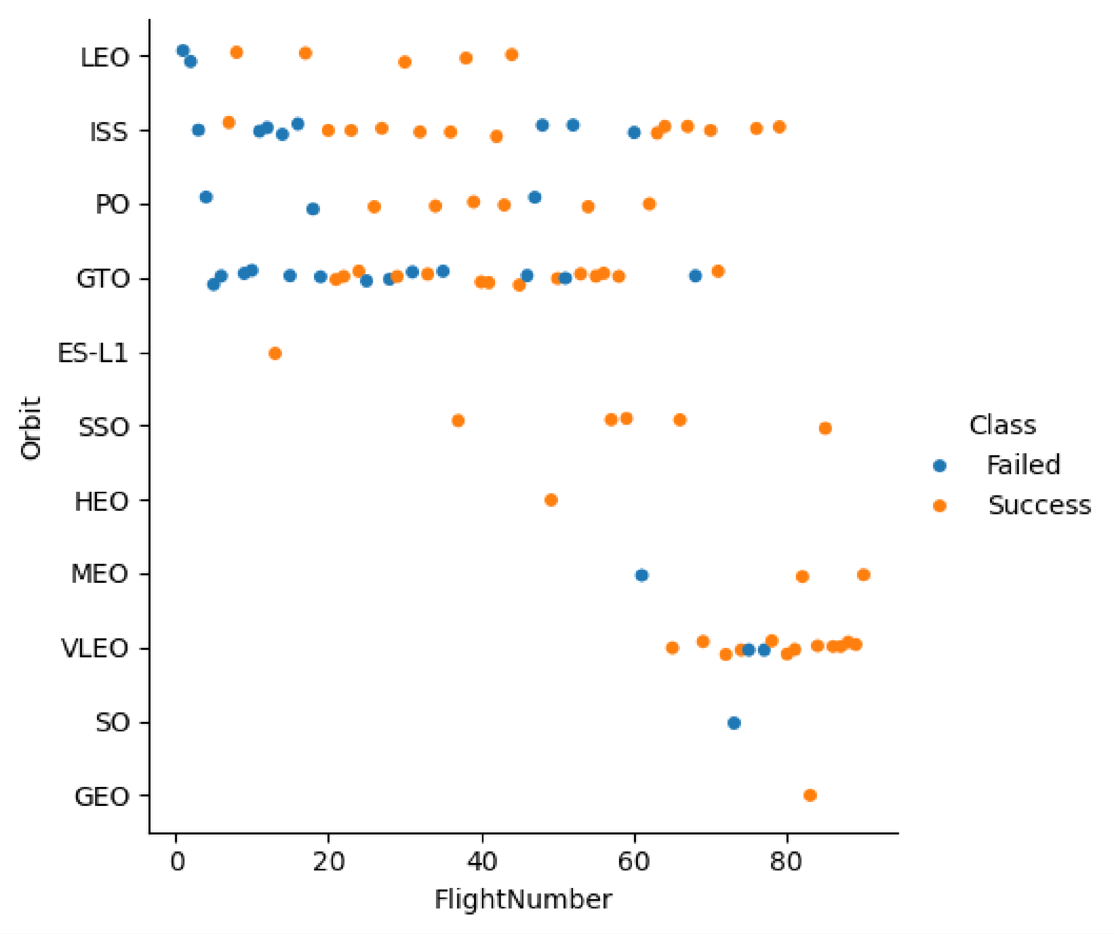
Success Rate vs. Orbit Type

Orbits ES-L1, SSO, HEO and GEO seem to be very successful, although the number of launches to this orbit can vary, influencing reliability of these results (e.g. there is only one mission to each of ES-L1, HEO, SO and GEO).



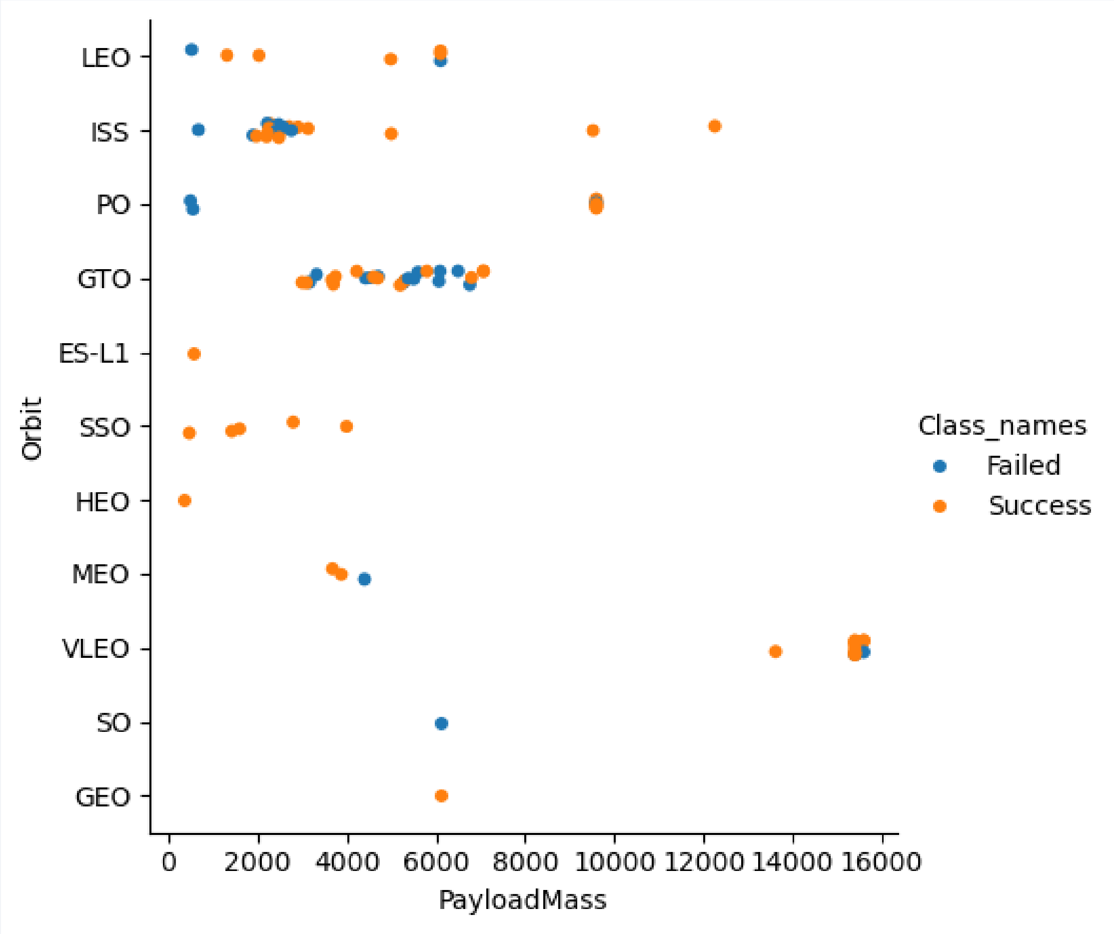
Flight Number vs. Orbit Type

- Original missions with lower Flight numbers were limited to LEO, ISS, PO and GTO orbits.
- SSO appears to be very ‘successful’ orbit.
- The commonly used orbits (ISS, PO, GTO, VLEO) have Failed missions interspersed throughout the Flight Number range.



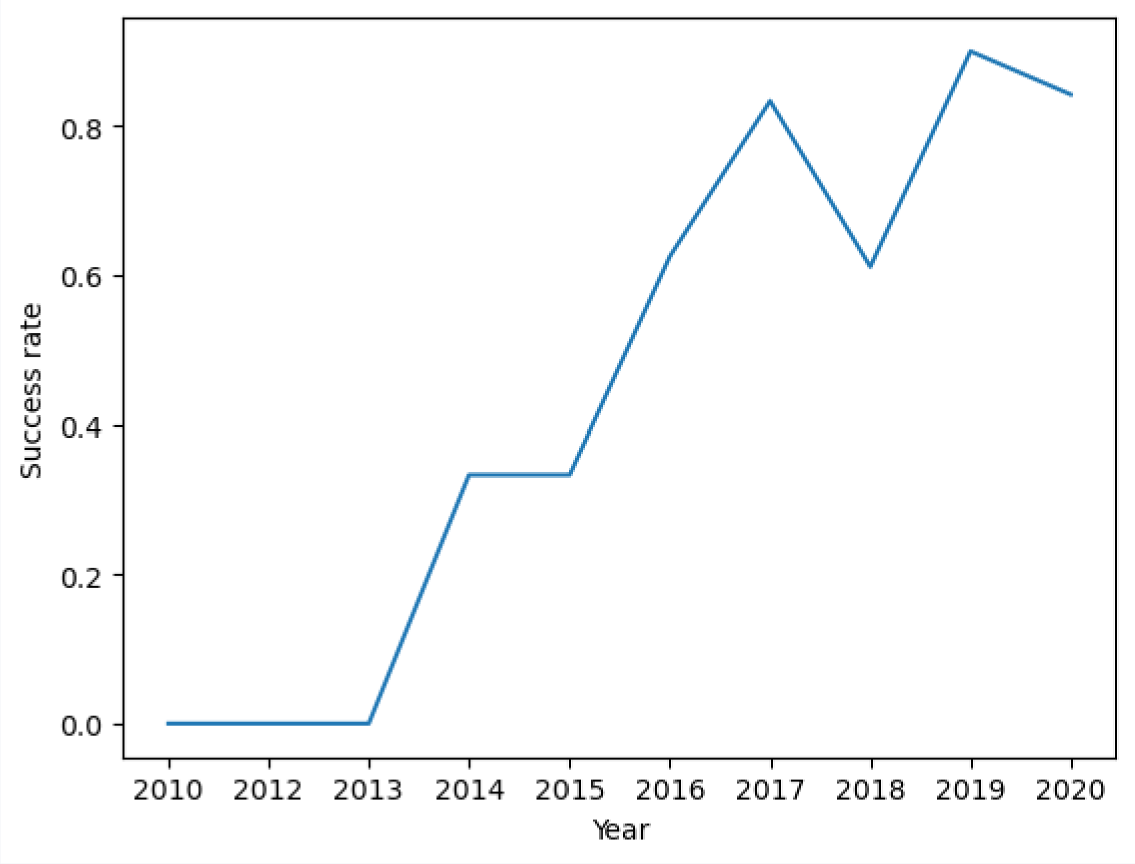
Payload vs. Orbit Type

- Majority of payloads sent to ISS is from the 2000-4000 kg range
- Ultra-heavy payloads (13 000 +) are sent exclusively to VLEO



Launch Success Yearly Trend

- The average success rate is increasing in time, suggesting an improvement in the rocket technology itself
- It corresponds to the trends observed Success with respect to Flight Number



All Launch Site Names

There are four available Launch Sites.

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

Launch Site Names Begin with 'CCA'

Overview of first couple missions Launched from Cape Canaveral.

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	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

Total Payload Launched for NASA (CRS) is over 45 tons.

Customer	TOTAL PAYLOAD
NASA (CRS)	45596

Average Payload Mass by F9 v1.1

Average Payload mass carried by F9 v1.1 booster rockets is 2535 kg.

Average
2534.67

First Successful Ground Landing Date

First successful landing on a ground pad happened 22nd of December 2015

Date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

All booster version carrying payload between 4000 and 6000 kg, that have successfully landed on a drone ship, are:

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

Total Number of Successful and Failure Mission Outcomes

- Absolute majority of the missions were deemed to be successful
- Considering, that some of the rockets did not manage to successfully land only proves, that a lot of the recorded flights were only test missions

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

Boosters Carried Maximum Payload

The maximal payload was carried by 12 different boosters.

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

There were two failed landings on a drone ship in 2015.

Month	Landing_Outcome	Booster_Version	Launch_Site
January	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
April	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

- 31 missions occurred during the given period
- 10 did not attempt to land
- Success of the remaining missions is listed in the table on the right

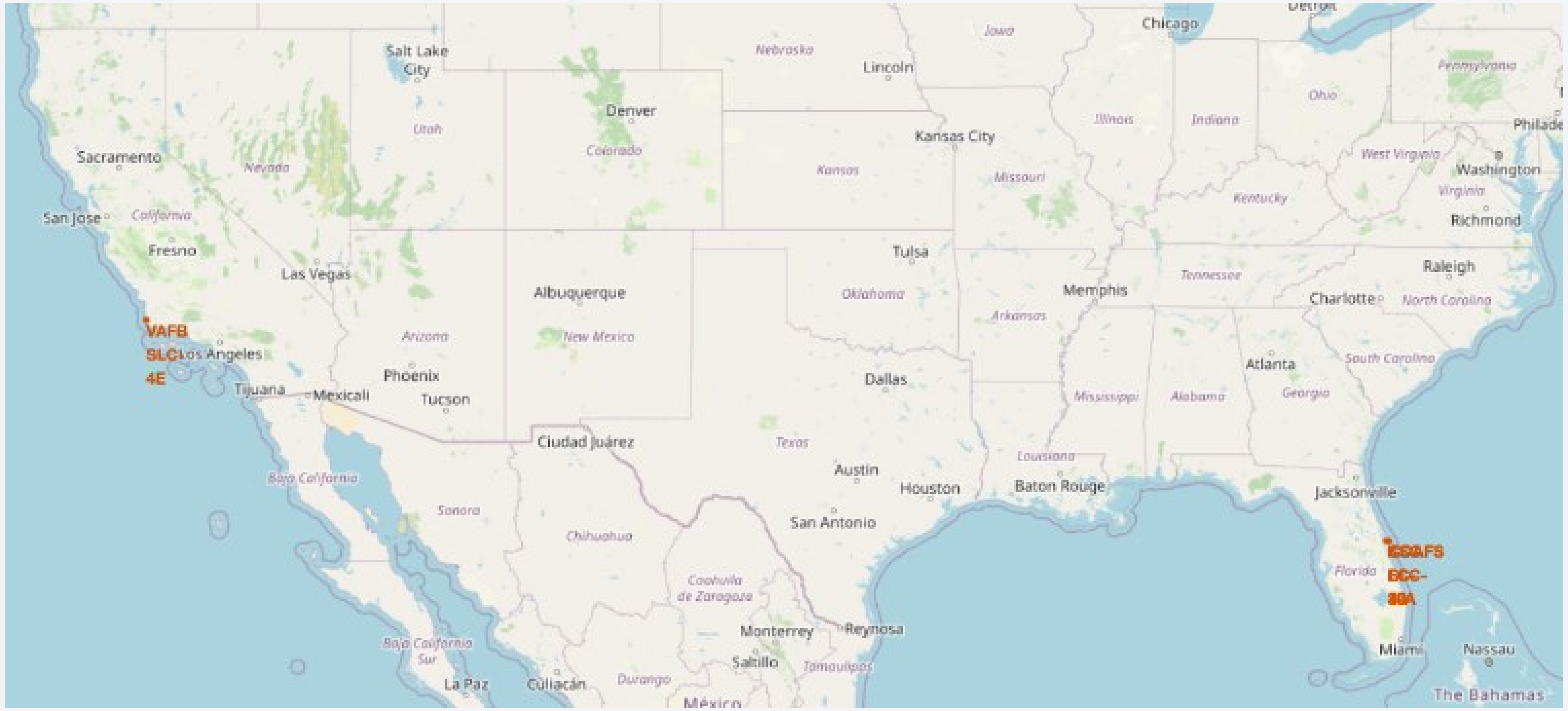
Landing_Outcome	Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

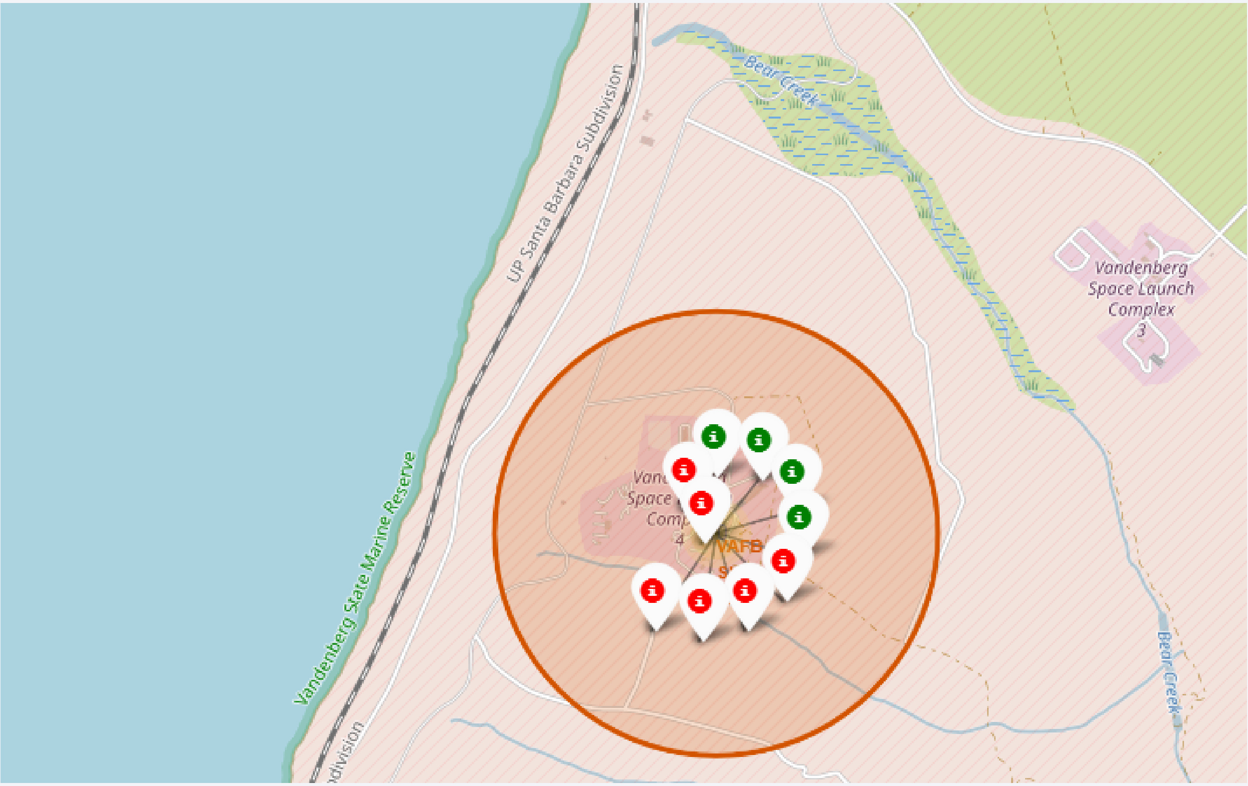
Map of all available launch sites

- Three out of four sites are located on Florida.
- Its proximity to equatorial provides an advantage in regards of launch costs due to higher centrifugal force.



Launches at Vandenberg Space Launch Complex 4

The Vandenberg Launch Complex in California sports 4 successful missions and 6 unsuccessful ones.

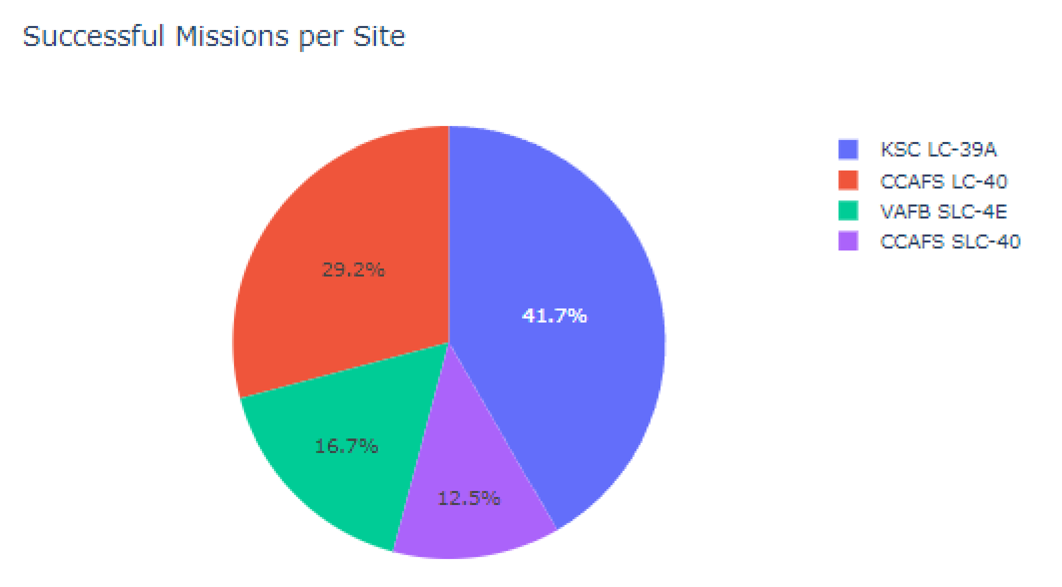


Section 4

Build a Dashboard with Plotly Dash

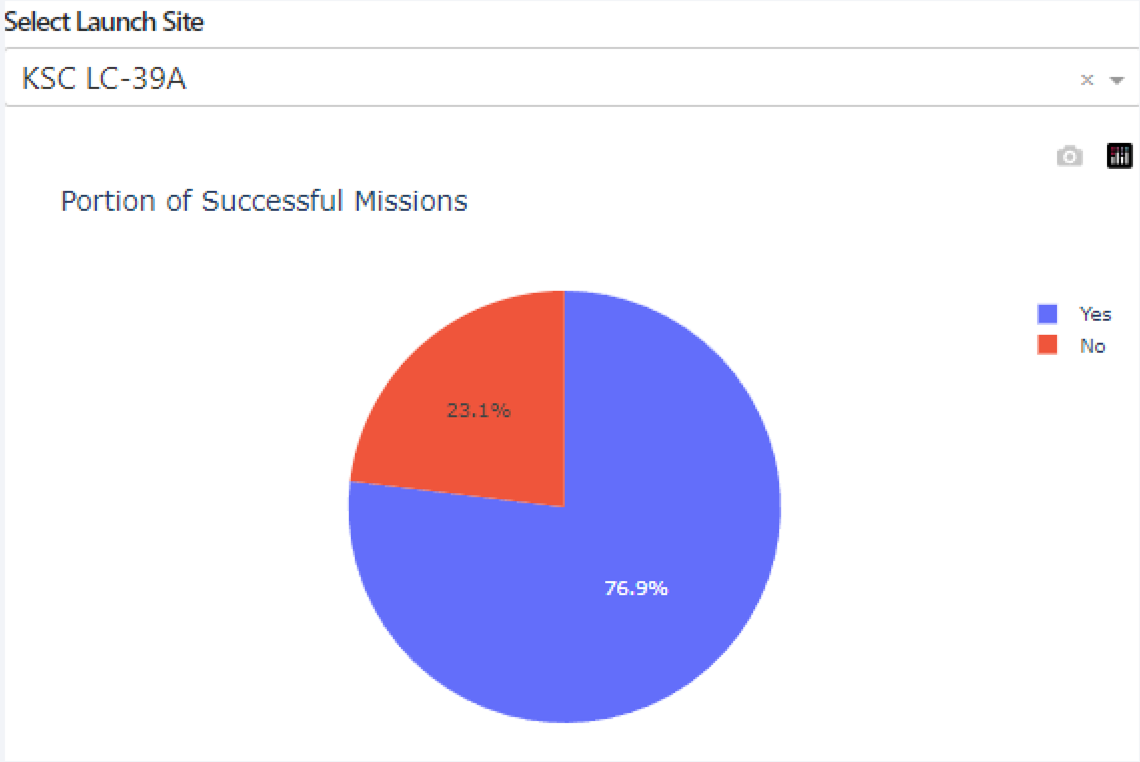
Successful missions per launch site

- Largest number of successful missions was Launched from KSC LC-39A constituting about 42 % of all successful launches
- CCAFS LC-40 accounts for another 29 % of all successful launches



Success rate of KSC LC-39A site

KSC LC-39A not only accounts for the highest number of successful launches overall, it also has the highest success rate of about 77 %.



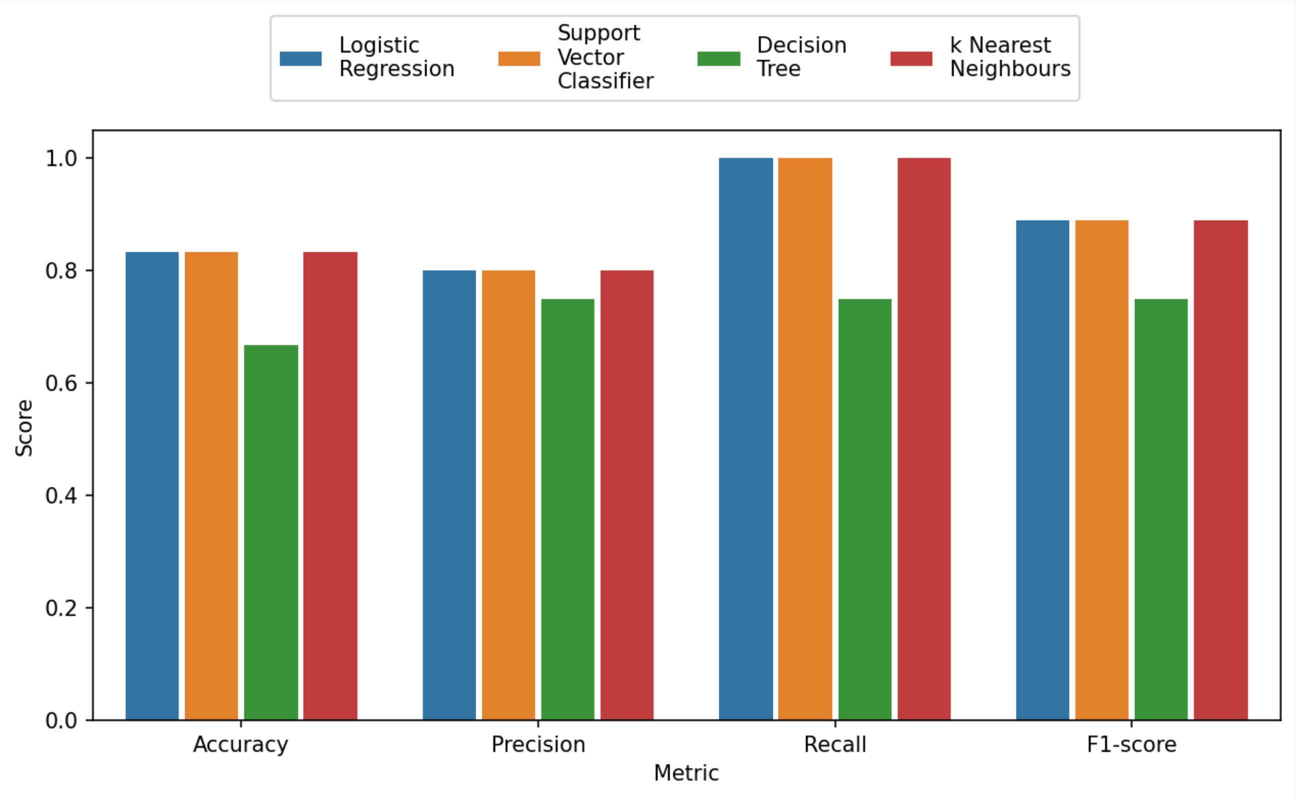
Mission success for payload's mass range 5000-10000 kg

- Payloads between 5500 and 7000 kg appear to be unsuccessful
- All payloads larger than 5000 kg are launched using boosters FT and B4

Section 5

Predictive Analysis (Classification)

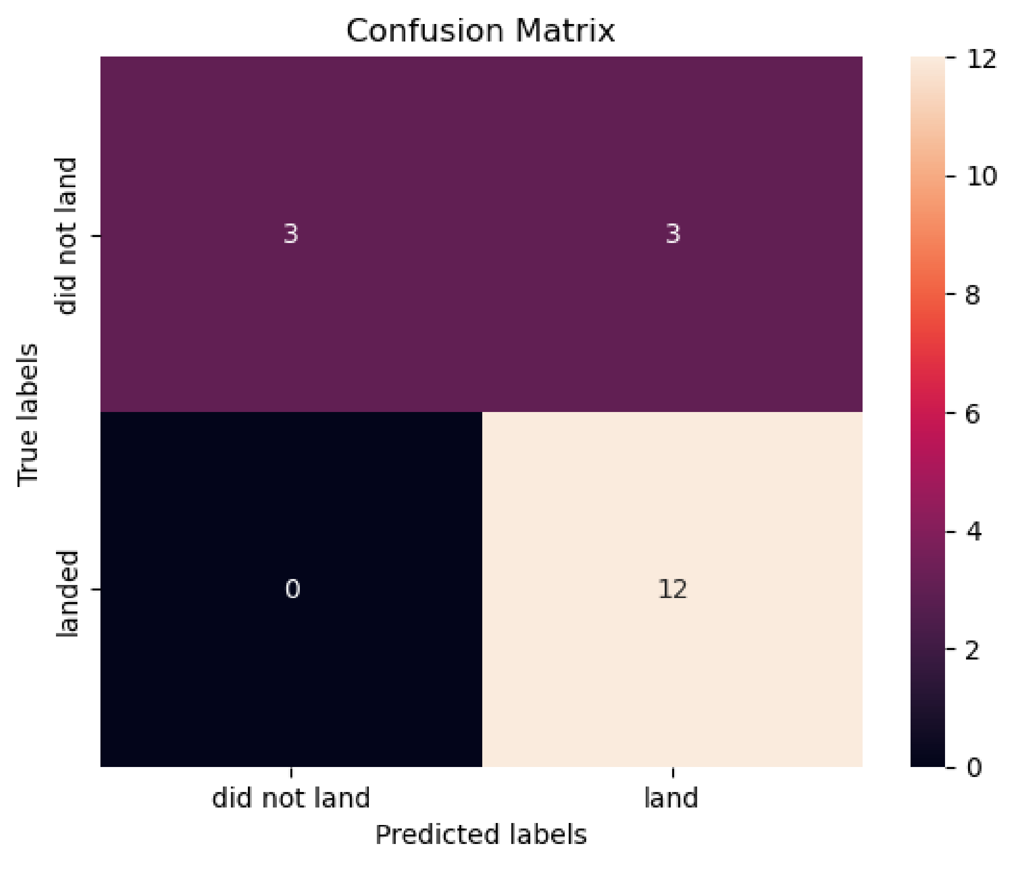
Classification Accuracy



- Logistic regression, SVM and KNN yield identical results.
- Since Logistic regression is computationally the least demanding, we have decided to deploy this model.

Confusion Matrix

- If the model predicts that the rocket will successfully land, there is an 80 % chance, that the prediction is correct.
- If the model predict that the rocket will **NOT land**, the model is **very certain** in its prediction.



Thank you!