IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Simon Grossrieder
December 2022

Outline

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

Executive Summary

- Summary of methodologies
  - Get data from different sources (SpaceX API and Wikipedia)
  - Explore data (Visually, Interactive and Data Wrangling)
  - prepare data for further analysis
  - Train predictive classifier models (Decision Tree, Logistic regression, KNN and SVN)
- Summary of all results
  - Models allow **prediction** of first stage landings at about **83% accuracy**.

Introduction

- SpaceX is able to provide cheaper rocket launches due to reused first stages.
- For other competitors it would be highly beneficial if they knew whether SpaceX could reuse the first stage or not.
- We want to predict if the landing of Falcon 9 first stages will be successful or not using data analysis on different launching parameters.

Section 1

Methodology

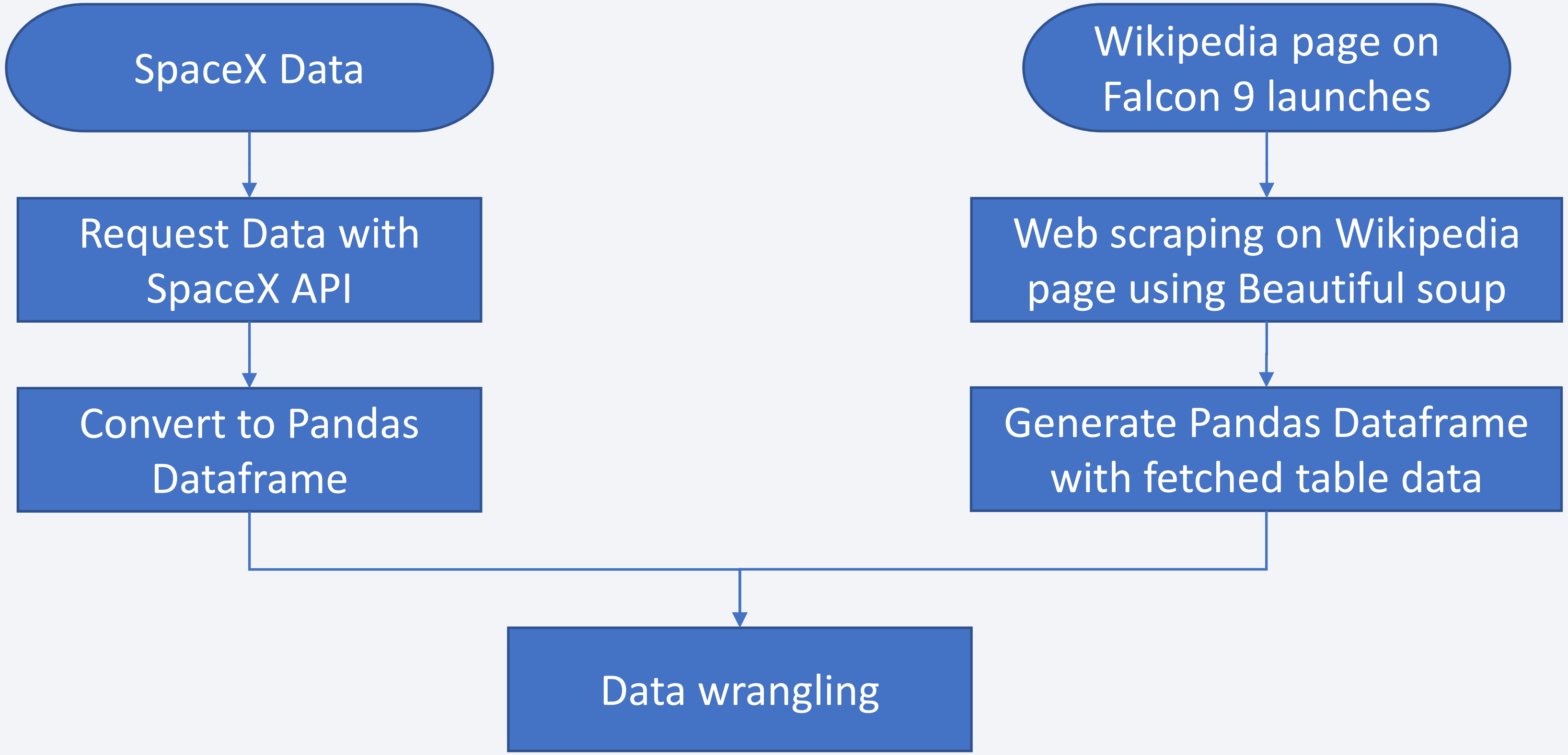
Methodology

Executive Summary

- Data collection methodology:
  - SpaceX API used to pull data on rocket launches
  - Web scraping Wikipedia page on Falcon 9 launches and landing outcomes
- Perform data wrangling
  - Create *class* data indicating mission outcome (which needs to be predicted)
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Build models with sklearn, tune them using GridSearchCV and evaluate their score and confusion matrix.

Data Collection

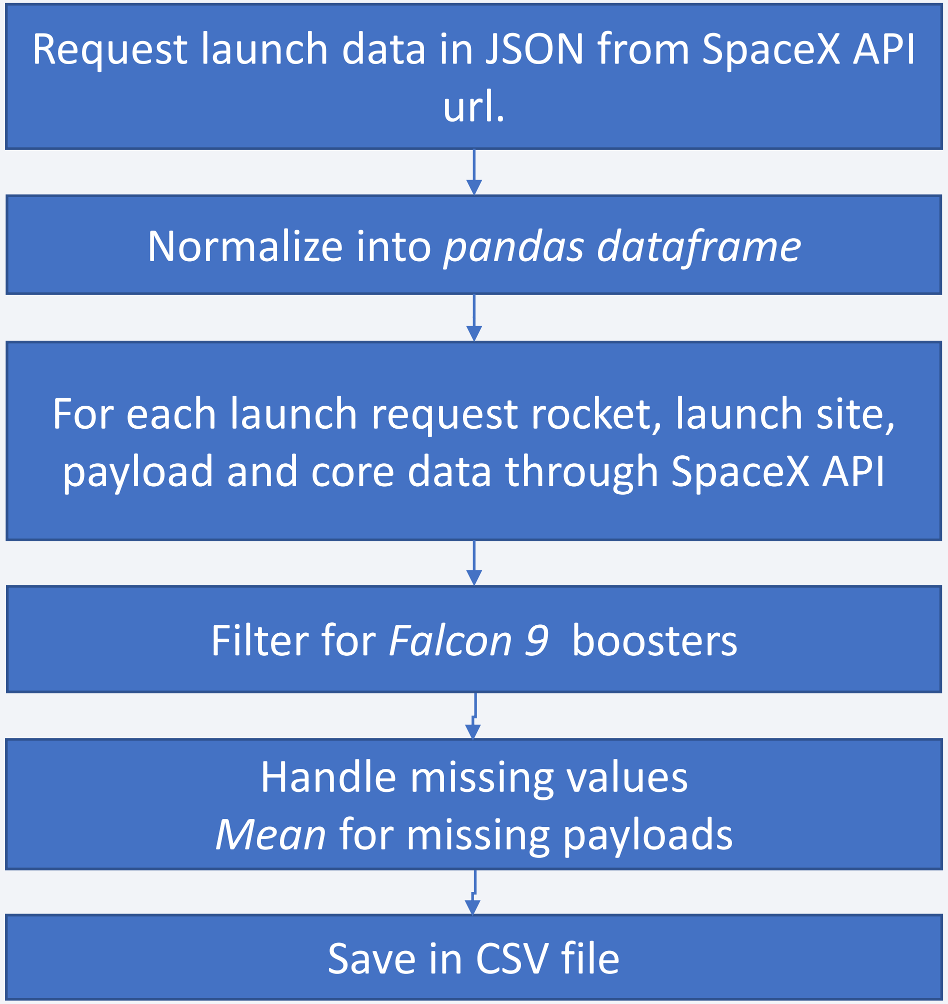
- Two datasets were separately imported and prepared for analysis



Data Collection – SpaceX API

- The content of the Wikipedia page on *Falcon 9* launches was wrangled using *Beautiful Soup* to extract data from the tables on the webpage.

https://github.com/SiGr27/Capstone/blob/main/SpaceX%20Data%20Collection.ipynb



Data Collection - Scraping

- SpaceX API was used to request data on past SpaceX rocket launches

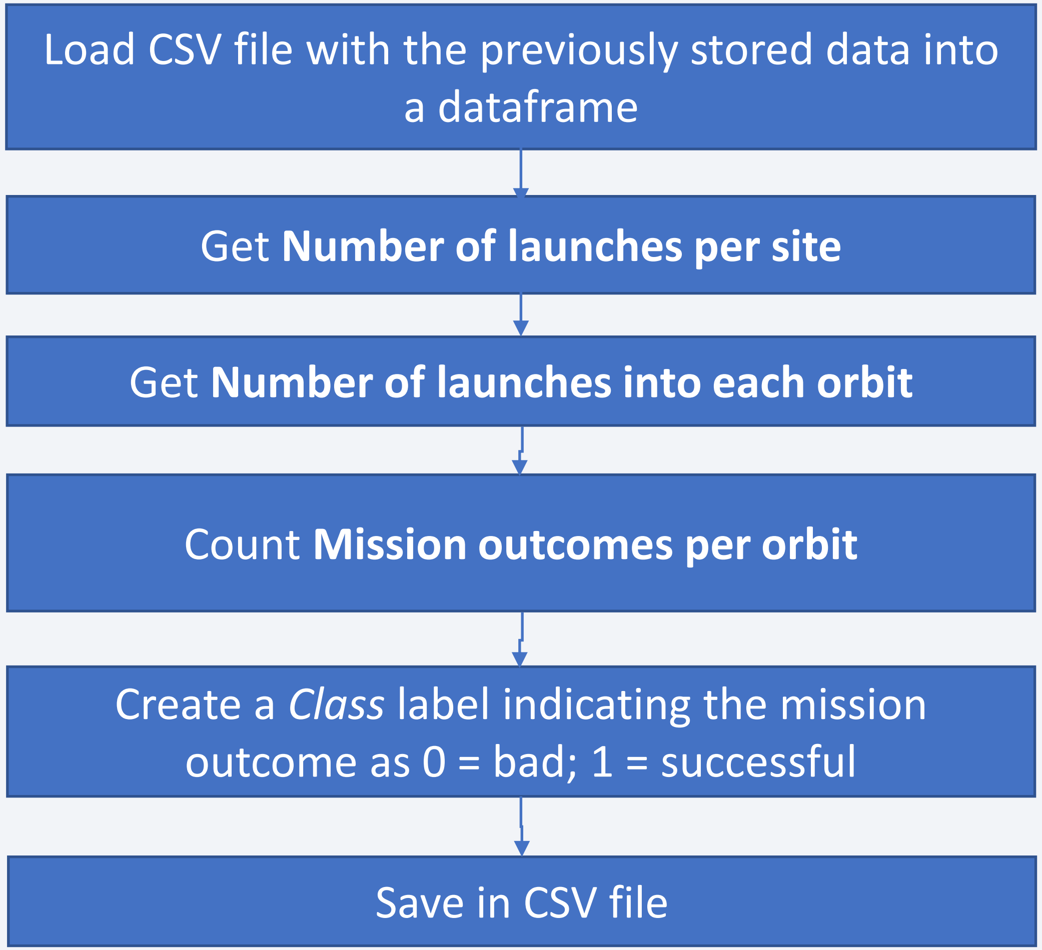
https://github.com/SiGr27/Capstone/blob/main/SpaceX_WebScraping.ipynb

1. Request HTML content of Wikipadia page on *Falcon 9* launches
2. Create a *Bueatiful Soup* object of content
3. Extract all tables
4. From the first launch table get columns and prepare launch dictionary
5. Fill dictionary going through each table and extracting the data
6. Create dataframe from dictionary and save into CSV file

Data Wrangling

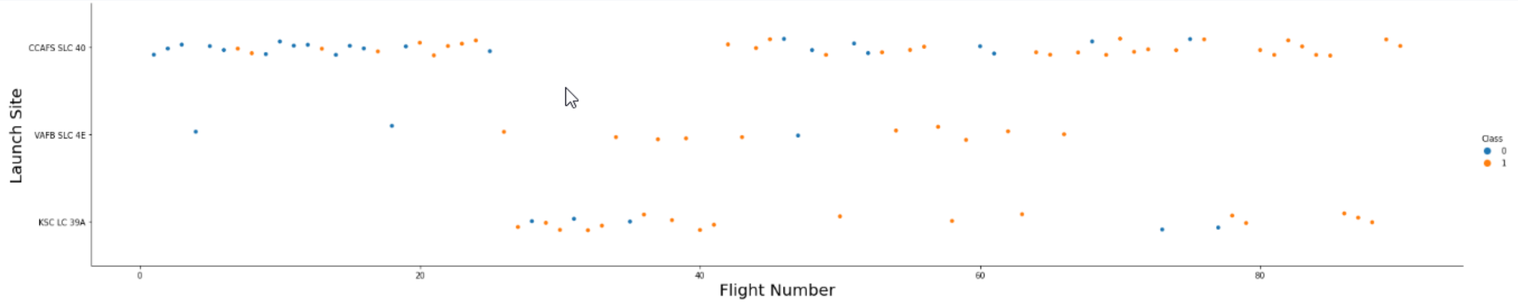
- Applying some basic operations on the dataframe to get some first insights on the *Falcon 9* launches.
- Transform data for further processing.

https://github.com/SiGr27/Capstone/blob/main/SpaceX_Data%20Wrangling.ipynb

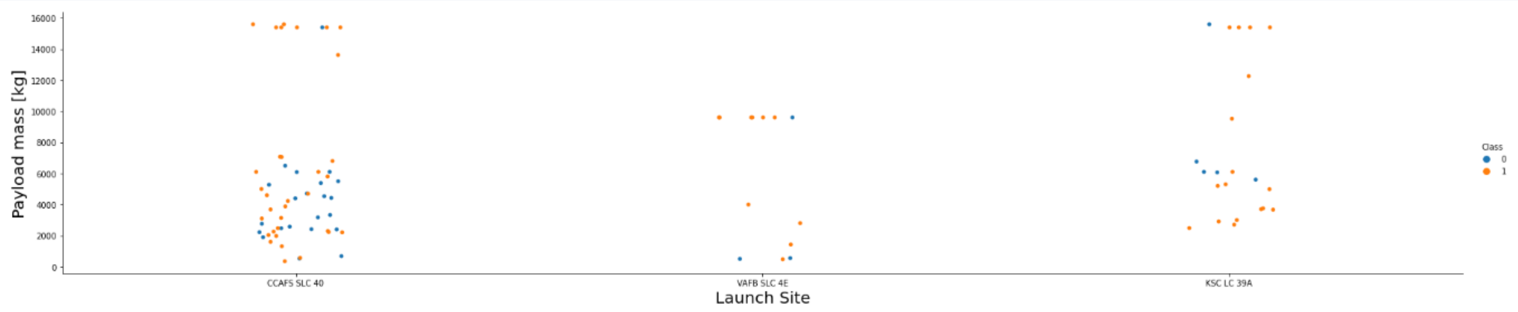


EDA with Data Visualization

Use a scatter plot to depict how the flight number and Launch sites have an impact on the mission outcome



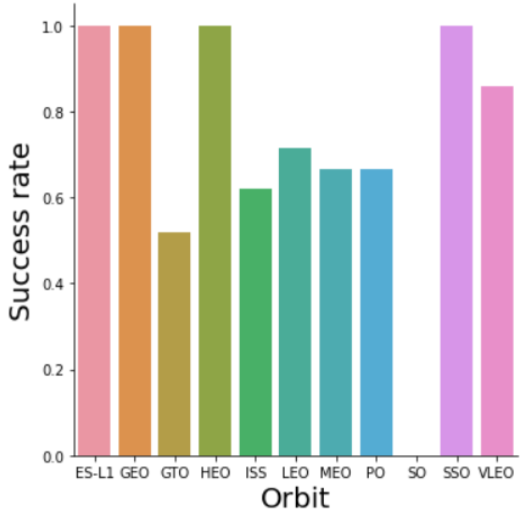
Use a scatter plot to depict how the payload and launch sites have an impact on the mission outcome



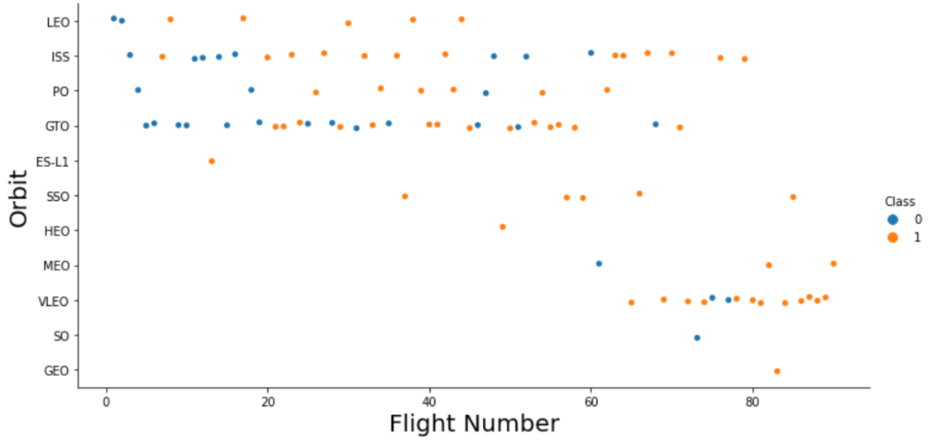
https://github.com/SiGr27/Capstone/blob/main/EDA_with_Visualization.ipynb

EDA with Data Visualization

Use a bar chart to depict the success rate for each orbit



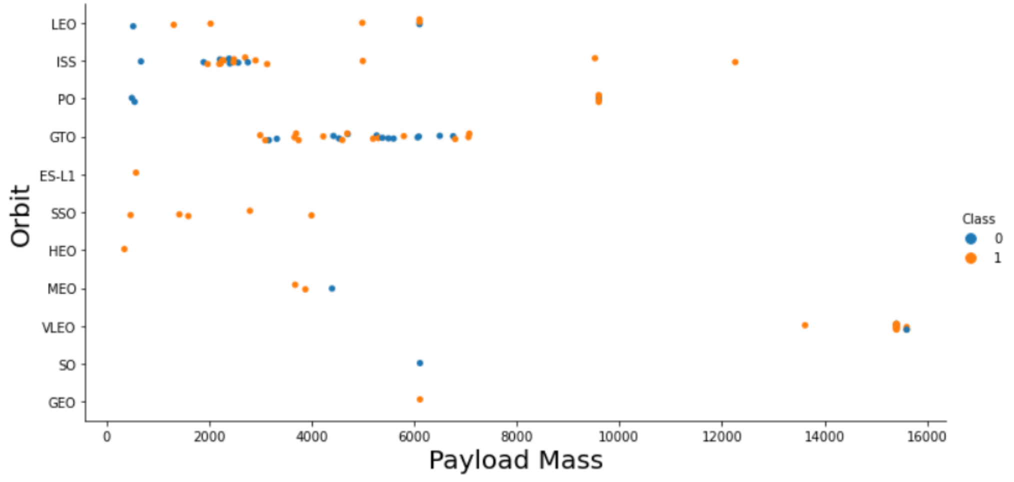
Use a scatter plot to depict which orbit was targeted along the flights and how successful the missions were



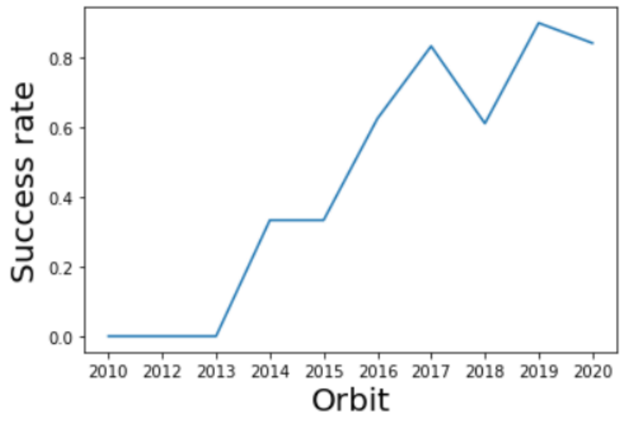
https://github.com/SiGr27/Capstone/blob/main/EDA_with_Visualization.ipynb

EDA with Data Visualization

Use a scatter plot to depict the payloads sent to each orbit and how successful the missions were.



Use a line plot to depict the evolution of the success rate along the years



https://github.com/SiGr27/Capstone/blob/main/EDA_with_Visualization.ipynb

EDA with SQL

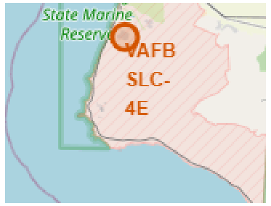
Executed Queries

- Display each distinct launch site
- Display (5) launches from sites starting with *“CCA”*
- Display the total payload carried for the customer *“NASA (CRS)”*
- Display the total payload carried by the booster *“F9 V1.1”*
- Display the date of the firs successful landing on a ground pad
- List boosters that successfully carried a payload between 4 and 6 tons and landed on a drone ship
- Count total missions grouped by their outcome
- List boosters that have carried the maximum payload mass
- List the failed landings on a drone ship in 2015 with the corresponding launch site and booster version
- Rank the landing outcome occurrences between 2010-06-04 and 2017-03-20 (descending)

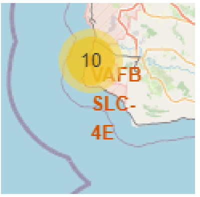
https://github.com/SiGr27/Capstone/blob/main/SpaceX_EDAwithSQL.ipynb

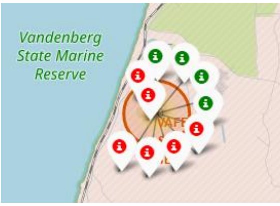
Build an Interactive Map with Folium

- Add Circles and Markers to locate the launch sites



- Add markers grouped into marker clusters for each launch. The color of the marker should indicate the mission outcome.





- Add markers and polylines to indicate distances as e.g. the distance of the launch site to the next coastline.

https://github.com/SiGr27/Capstone/blob/main/Interactive_Visual_Analytics_with_Folium.ipynb

Build a Dashboard with Plotly Dash

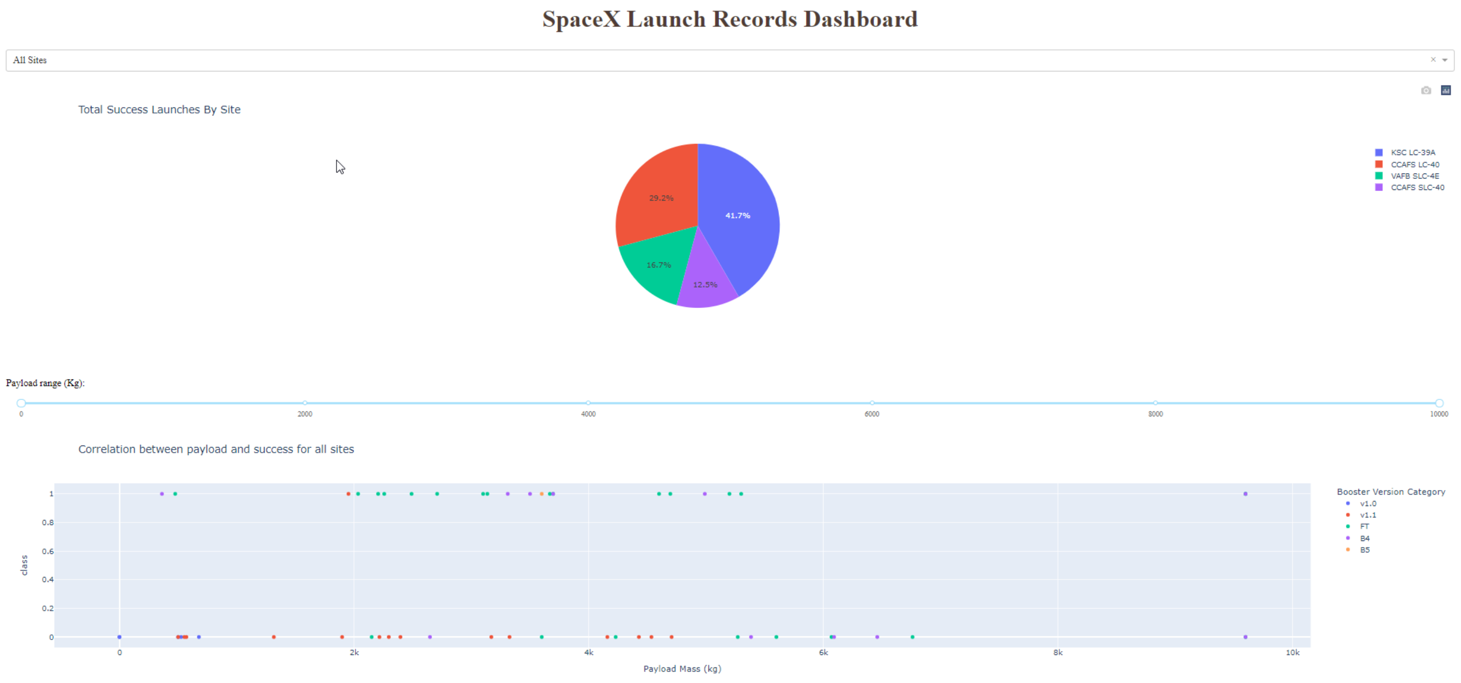
Dashboard Elements:

Dropdown selector for the launch site

Pie chart to depict success rate of selected site

Slider to select payload range

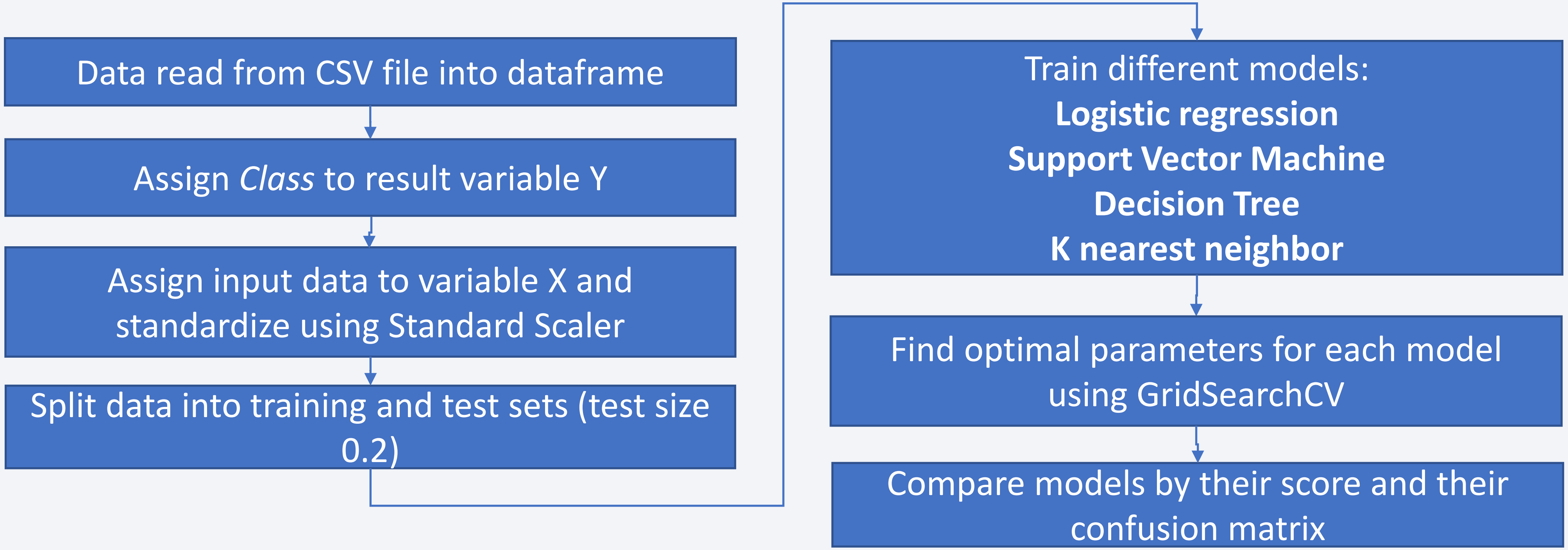
Scatter plot indicating mission outcomes for selected payload range and launch site. Marker colors differentiate the booster versions.



https://github.com/SiGr27/Capstone/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

- Train and compare different classification models to select most performant one.



https://github.com/SiGr27/Capstone/blob/main/Machine%20Learning%20Prediction.ipynb

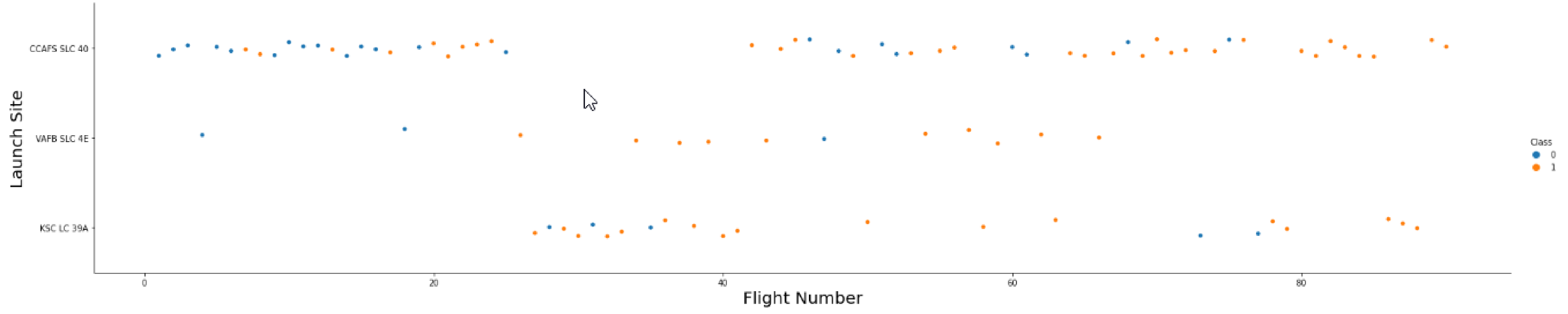
Results

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

Section 2

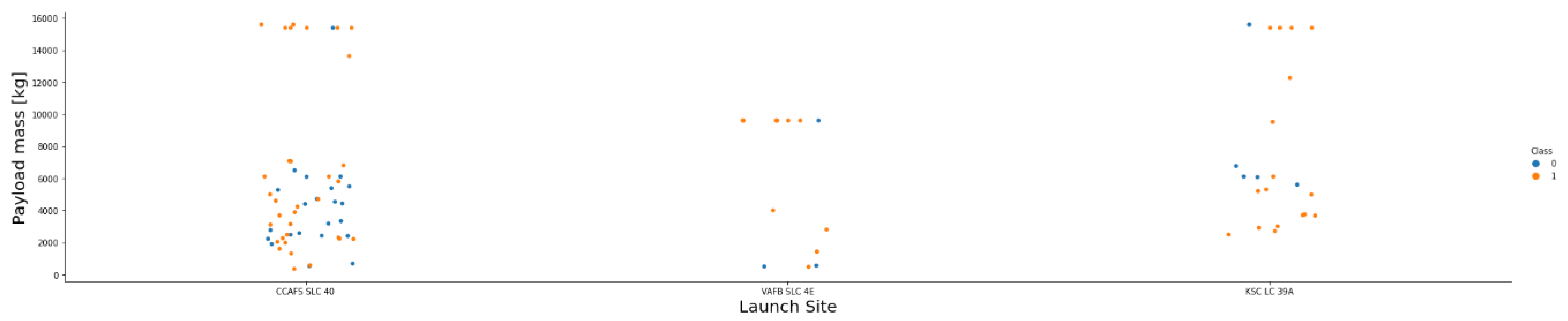
Insights drawn from EDA

Flight Number vs. Launch Site



- In the beginning there were quite a lot of unsuccessful missions. Up from the 20th mission the success rate became quite good.
- Most of the first missions were launched from CCAFS SLC40. As soon most of the missions were successful the other sites were used as well

Payload vs. Launch Site

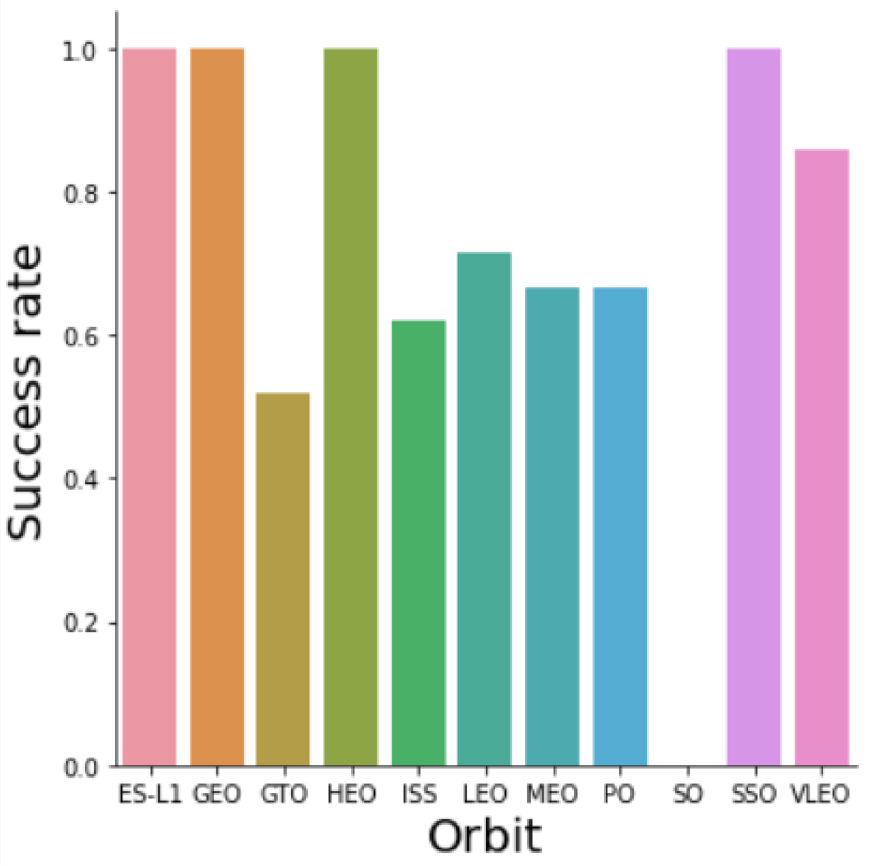


- VAFB SLC 4E was not used for high payload launches
- The payload does not seem to have an impact on the success of the mission.

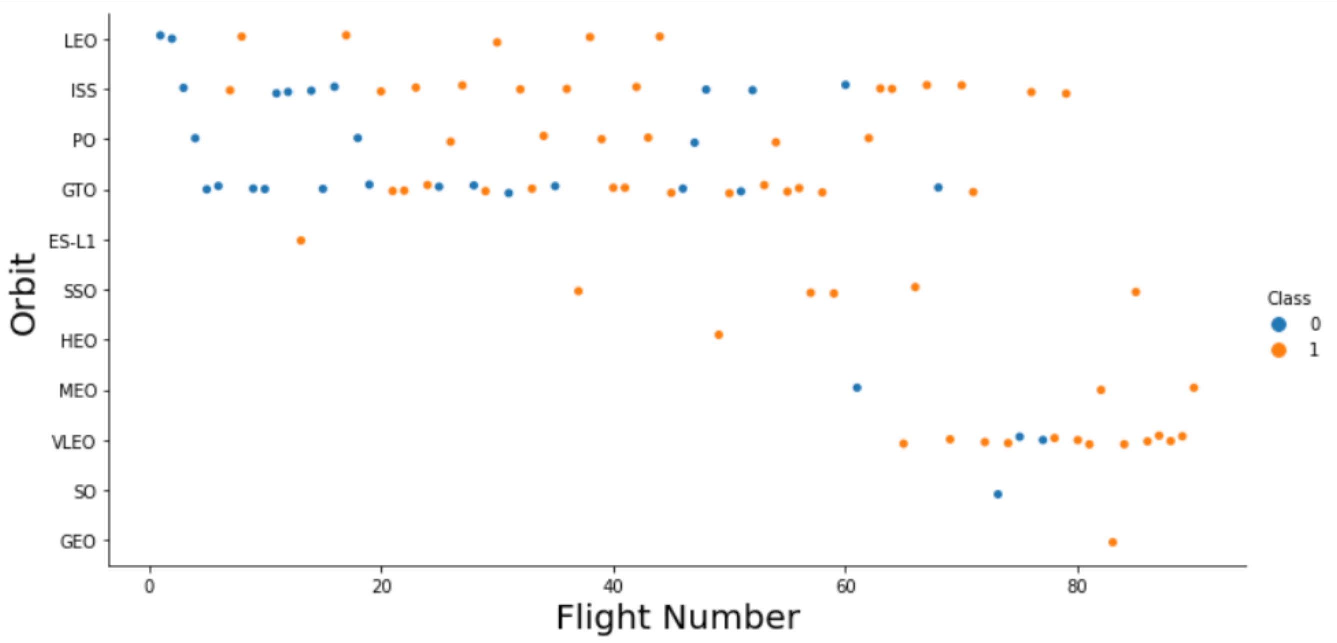
Success Rate vs. Orbit Type

- Some of the orbit have 100% success so far, but these were rarely targeted.
- The more often targeted orbits have comparable rates around 50% to 70%

Orbit	Class	
ES-L1	1	1
GEO	1	1
GTO	1	14
	0	13
HEO	1	1
ISS	1	13
	0	8
LEO	1	5
	0	2
MEO	1	2
	0	1
PO	1	6
	0	3
SO	0	1
SSO	1	5
VLEO	1	12
	0	2

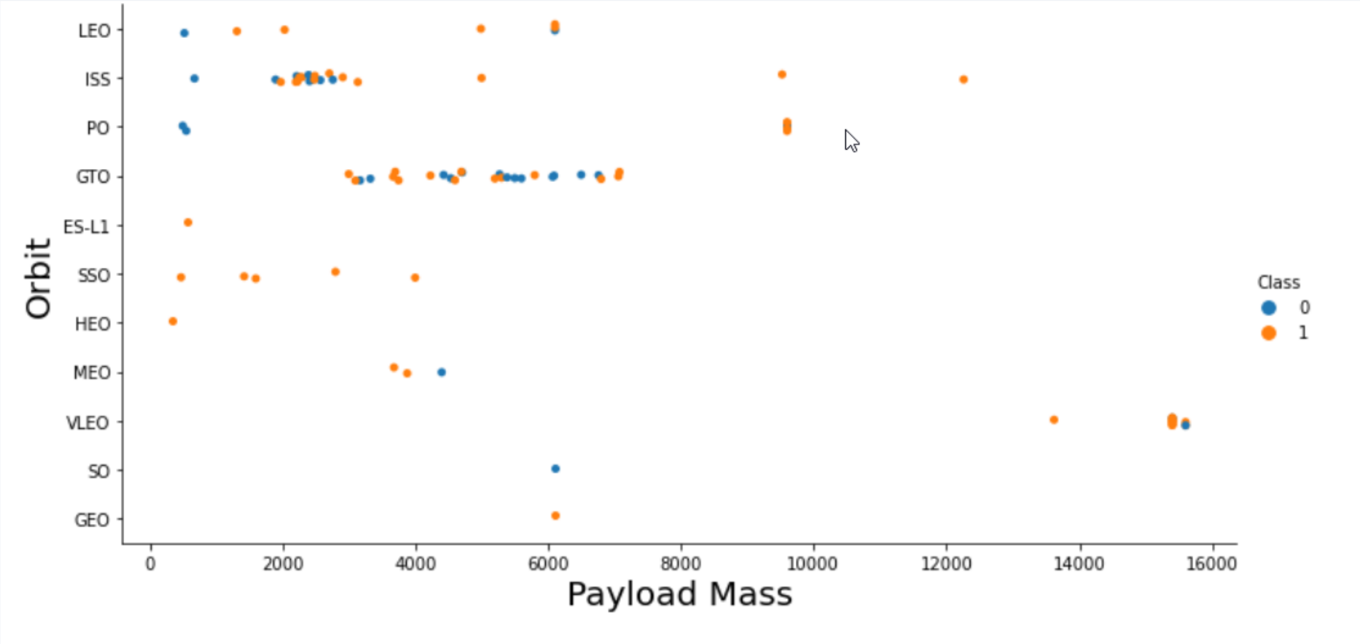


Flight Number vs. Orbit Type



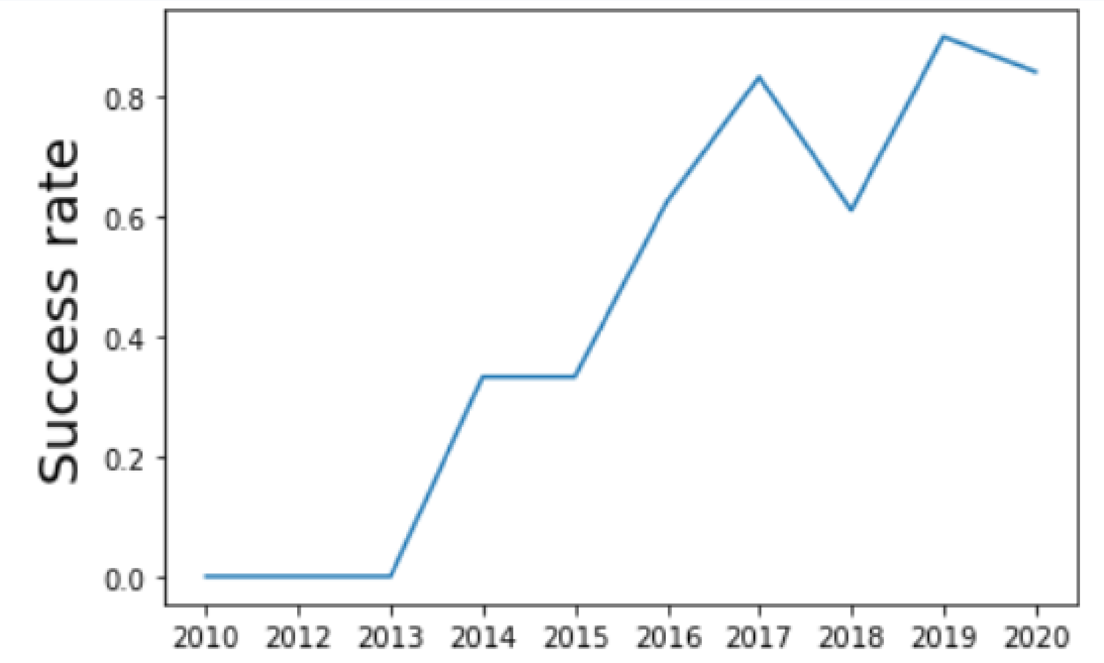
- In the starting phase, mostly LEO, ISS, PO and GTO were targeted.
- During the last flights the VLEO was regularly targeted.
- The success is more impacted by the flight number than the orbit.

Payload vs. Orbit Type



- ISS orbit has a higher success rate with high payloads.
- GTO orbit has a lower success rate with high payloads
- For the other orbits there is no obvious correlation observable.

Launch Success Yearly Trend



- A positive trend of the success rate over the years is clearly visible.
- The only remarkable decrease of the average success rate happened in 2018.

All Launch Site Names

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

- **4 launch sites** were used for SpaceX launches

Launch Site Names Begin with 'CCA'

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

- These 5 launches from CCAFS LC-40 happened in the early years of the SpaceX program
- The missions were all successful, but none of the landings.

Total Payload Mass

45596

- SpaceX transported about **46 tons** of material and equipment in the space for **NASA CRS**.

Average Payload Mass by F9 v1.1

- An **average payload** transported by F9 v1.1 booster weights **2928 kg**

First Successful Ground Landing Date

- The **first successful landing** on a ground pad happened on **December 22th, 2015**

Successful Drone Ship Landing with Payload between 4000 and 6000

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

- The above listed boosters have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

Total Number of Successful and Failure Mission Outcomes

mission_outcome	count_mission
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

- With **1 only failure** in flight, SpaceX has a very high mission success rate.

Boosters Carried Maximum Payload

booster_version	payload_mass__kg_
F9 B5 B1048.4	15600
F9 B5 B1048.5	15600
F9 B5 B1049.4	15600
F9 B5 B1049.5	15600
F9 B5 B1049.7	15600
F9 B5 B1051.3	15600
F9 B5 B1051.4	15600
F9 B5 B1051.6	15600
F9 B5 B1056.4	15600
F9 B5 B1058.3	15600
F9 B5 B1060.2	15600
F9 B5 B1060.3	15600

- Boosters F9 B5 were used for max payload flights.
- The maximum carried payloads is 15.6 tons

2015 Launch Records

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

- In 2015 two landing attempts on a drone ship failed.
- Both were launched from CCAFS LC-40 with a booster F9 v1.1

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

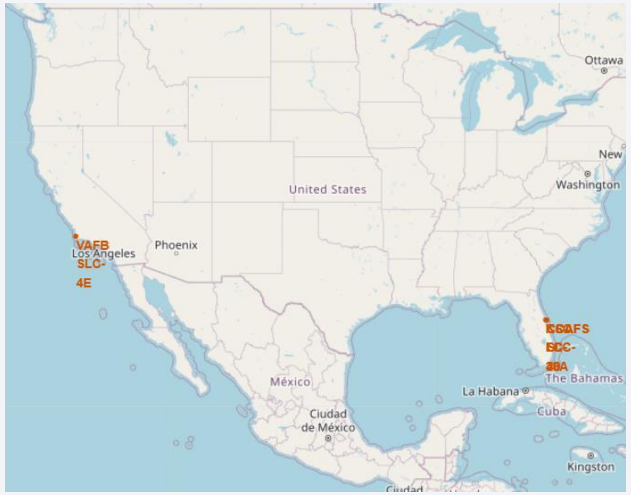
landing__outcome	2
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

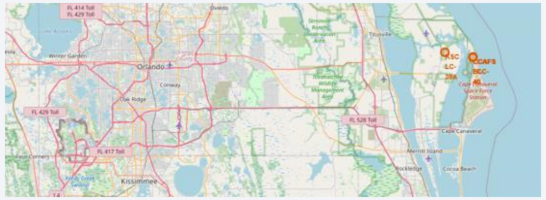
- In the mentioned time period
  - 1 third of the boosters had no landing attempts
  - 1 third of the boosters were attempted to land on a drone ship

Section 4

Launch Sites Proximities Analysis

Locations of the launch sites

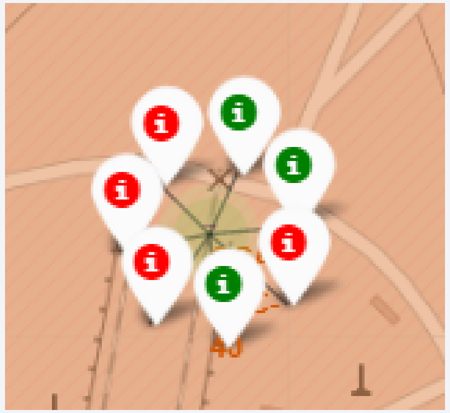




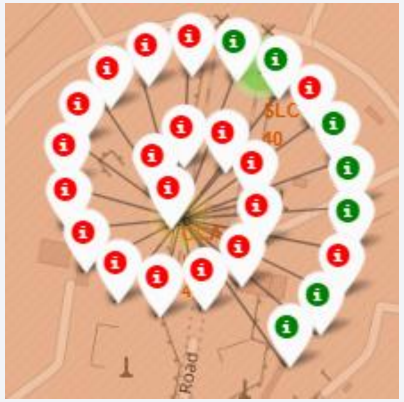
- All launch sites are located in the United States of America
- 1 launchsite is near LA California
- 3 launch sites are near Orlando Florida

Launch outcomes of each site

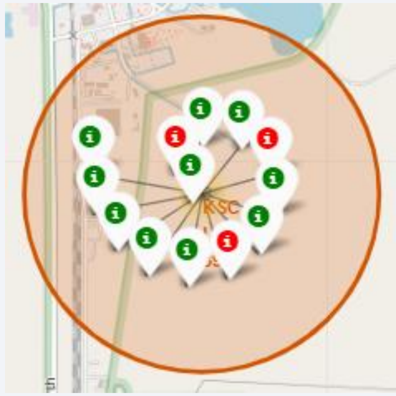
VAFB SLC-4E



CCAFS LC-40



CCAFS SLC-40



KSC LC-39A

- KSC LV-39A has the highest success rate
- Most launches happened from CCAFS SLC-40

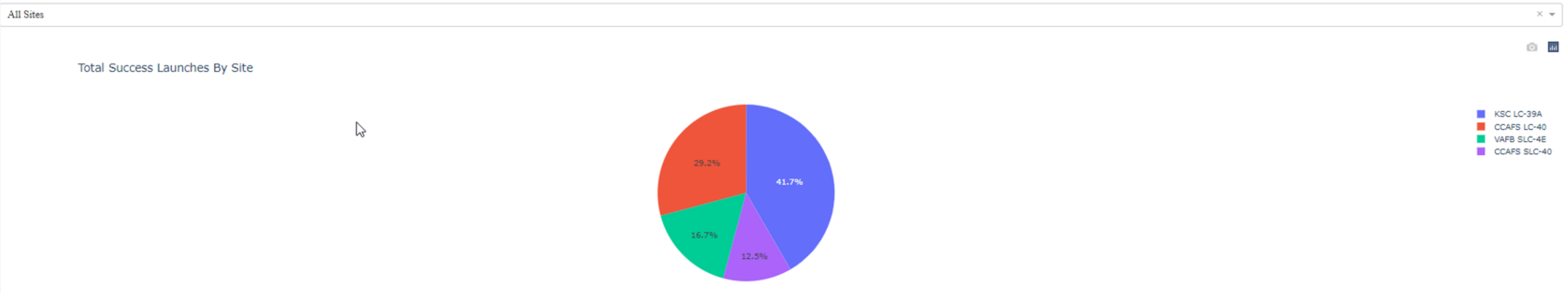
Coastline distance of VAFB SLC-4E

- VAFB SLC-4E is located 1.36 km apart from the ocean.

Section 5

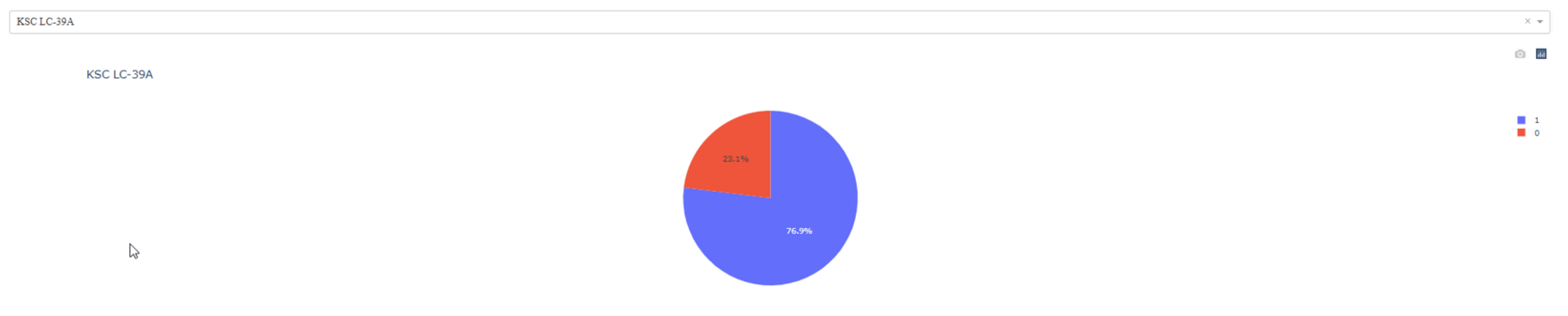
Build a Dashboard with Plotly Dash

Successful Launches per Site



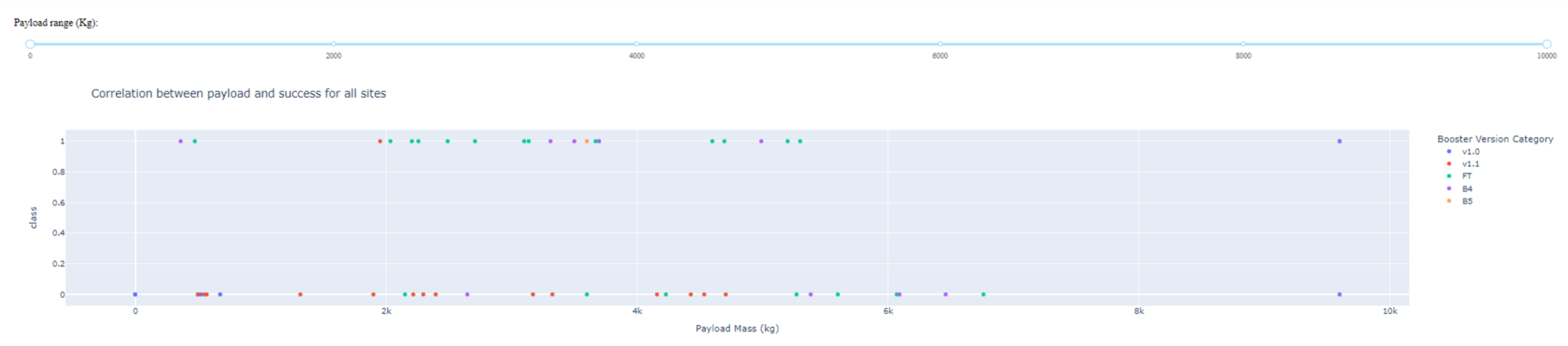
Most of the successful launches have been launched from KSC-LC-39A

Most successful launch site



- More than 3 out of 4 launches from KSC LC-39A are successful

Payload vs. Launch Outcome

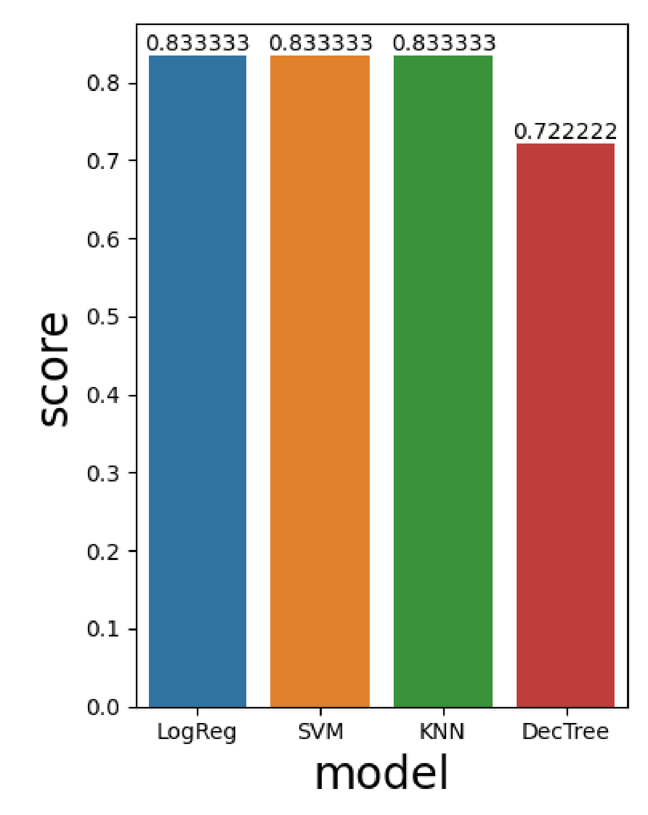


- Payload does not have a direct impact on a successful mission
- Booster V1.0 and V1.1 were still quite unsuccessful where the other versions became more reliable.

Section 6

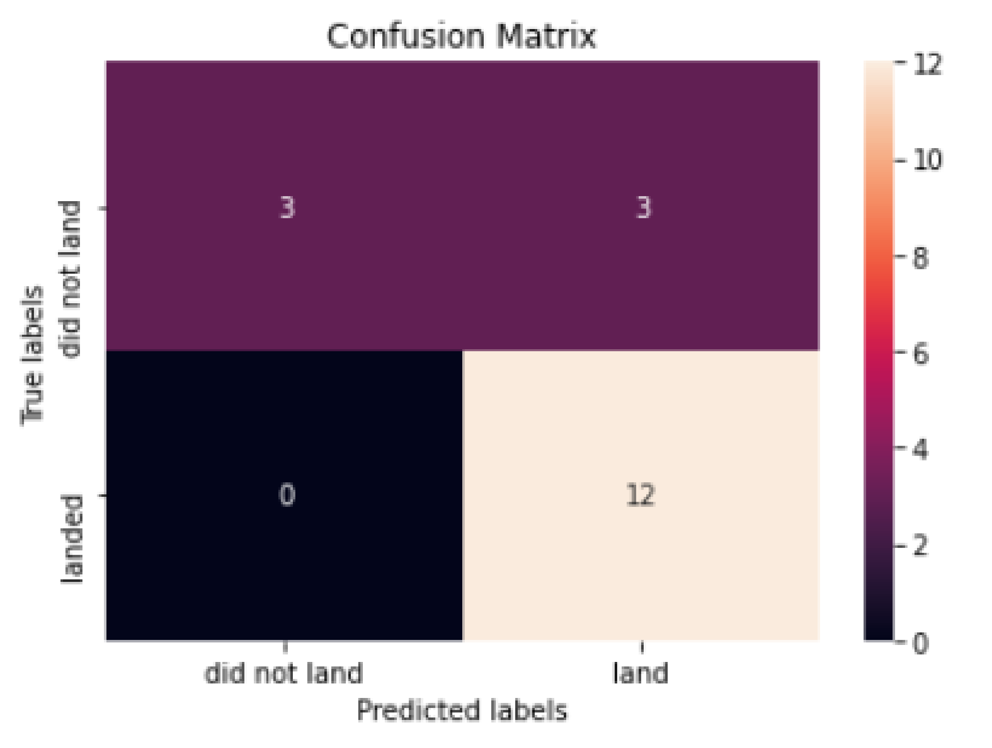
Predictive Analysis (Classification)

Classification Accuracy



- Decision tree performed the worst
- All other models performed equally well on the test data.

Confusion Matrix



- This is the confusion matrix of the logistic regression model.
- There are **3 false positives** in the predictions of the testdata

Conclusions

- About 83% prediction accuracy can be expected.
- The models all tend to predict some false positives.
- More samples would allow to tune the models.

Appendix

- Notebooks, data and code are available on GitHub:

Data Collection	https://github.com/SiGr27/Capstone/blob/main/SpaceX%20Data%20Collection.ipynb
Web Scraping	https://github.com/SiGr27/Capstone/blob/main/SpaceX_WebScraping.ipynb
Data Wrangling	https://github.com/SiGr27/Capstone/blob/main/SpaceX_Data%20Wrangling.ipynb
EDA with data visualiztaion	https://github.com/SiGr27/Capstone/blob/main/EDA_with_Visualization.ipynb
EDA with SQL	https://github.com/SiGr27/Capstone/blob/main/SpaceX_EDAwithSQL.ipynb
Machine learning prediction	https://github.com/SiGr27/Capstone/blob/main/Machine%20Learning%20Prediction.ipynb

Thank you!