Data Science Capstone project

Alexander Hobein

02/09/2021

Outline

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

Executive Summary

SpaceX's Falcon 9 rocket convinces with low cost due to the reuse of the first stage of the rocket. Predicting whether a rocket will successfully land with a reused stage 1 can effectively reduce cost in the future. In order to perform these predictions, 4 models were developed. The data was collected using web scraping, from SpaceX API and other feature selections such as visualization. The models all performed well with the same accuracy value of 0.883.

Introduction

- SpaceX is a Space Transportation Company investing in Rockets for Space Travelling
- Especially Falcon9 Rockets seem to be promising with much lower cost
- This lower cost can be explained by the rockets ability to reuse the first stage of the rocket

→ Predicting whether a rocket will successfully land with a reused stage 1 can effectively reduce cost in the future

Methodology

- Data collection methodology:
  - Request data with SpaceX API and Web Scraping
- Perform data wrangling
  - Exploratory Data Analysis using python pandas
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform predictive analysis using classification models
  - Build 4 models, optimize hyperparameters using grid search and evaluate them

Methodology

Data collection

1. Requesting Data using the SpaceX API
   → get Requests using Python and saving data in .csv file

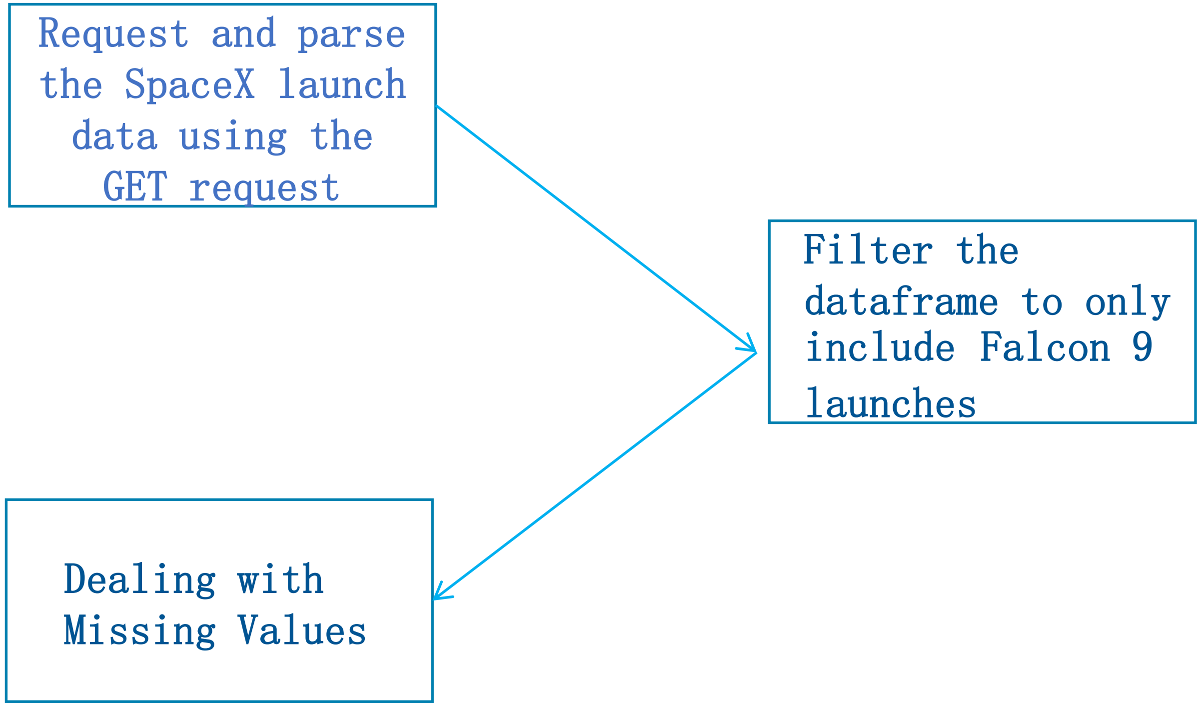
2. Web Scraping launch records from Wikipedia.com
   → BeautifulSoup using Python and saving data in .csv file

Data collection – SpaceX API

Visit the notebook for further information:

https://github.com/ahobein/FinalAssignmentSPACEX/blob/main/Complete%20the%20Data%20Collection%20API%20Lab.ipynb

Flowchart of SpaceX API calls

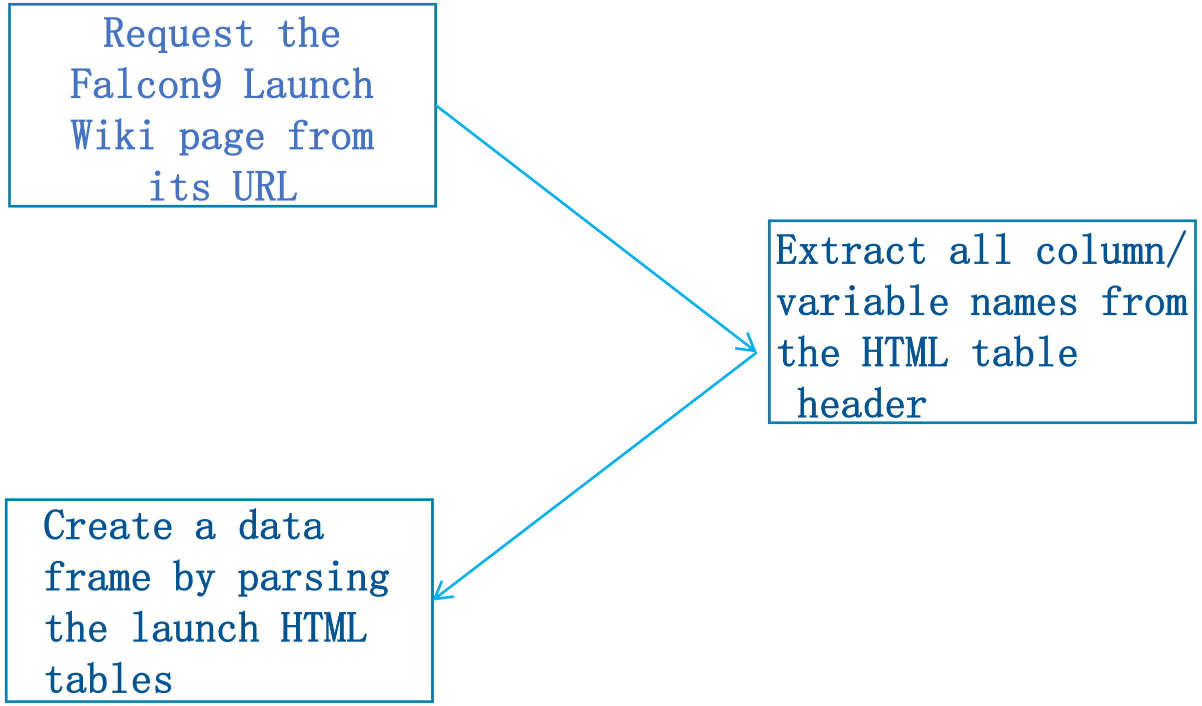


Data collection – Web scraping

Visit the notebook for further information:

https://github.com/ahobein/FinalAssignmentSPACEX/blob/main/Data%20Collection%20with%20Web%20Scraping%20lab.ipynb

Flowchart of web scraping



Data wrangling

1. Calculate the number of launches on each site
2. Calculate the number and occurrence of each orbit
3. Calculate the number and occurrence of mission outcome per orbit type
4. Create a landing outcome label from Outcome column

https://github.com/ahobein/FinalAssignmentSPACEX/blob/main/EDA.ipynb

EDA with data visualization

1. Relationship between Flight Number and Launch Site (scatter)
2. Relationship between Payload and Launch Site (scatter)
3. Relationship between success rate of each orbit type (bar)
4. Relationship between Flight Number and Orbit type (scatter)
5. Relationship between Payload and Orbit type (scatter)
6. Launch success yearly trend (line)

https://github.com/ahobein/FinalAssignmentSPACEX/blob/main/EDA%20with%20Visualization%20lab.ipynb

EDA with SQL

- Exploratory Data Analysis was further done using SQL
- Here basic question were answered like average payload and other

Predictive analysis (Classification)

- 4 Models were built: k-Nearest Neighbors, Decision Tree, Support Vector Machine and Logistic Regression
- Hyperparameters were optimally found using Grid Search
- The Accuracy of all models were evaluated on a separate test set

Results

- Exploratory data analysis results
- Predictive analysis results

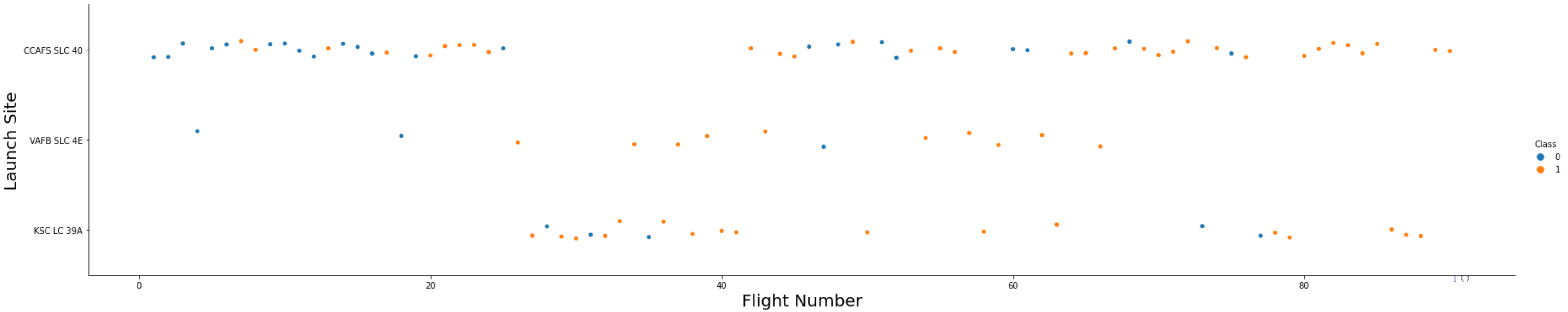
EDA with Visualization

https://github.com/ahobein/FinalAssignmentSPACEX/blob/main/EDA%20with%20Visualization%20lab.ipynb

Flight Number vs. Launch Site

This graph shows the successful and unsuccessful launches per launch sites

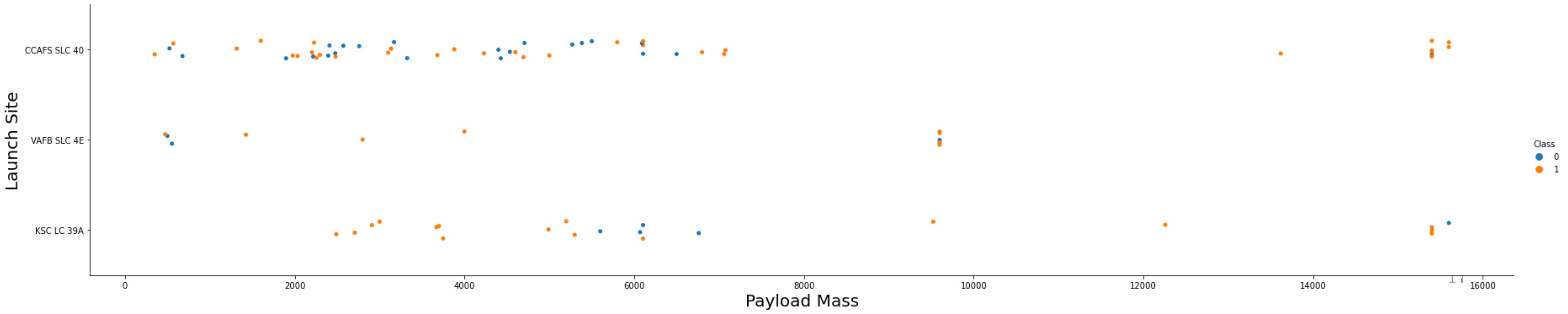
→ KSC LC 39A has the best ratio



Payload vs. Launch Site

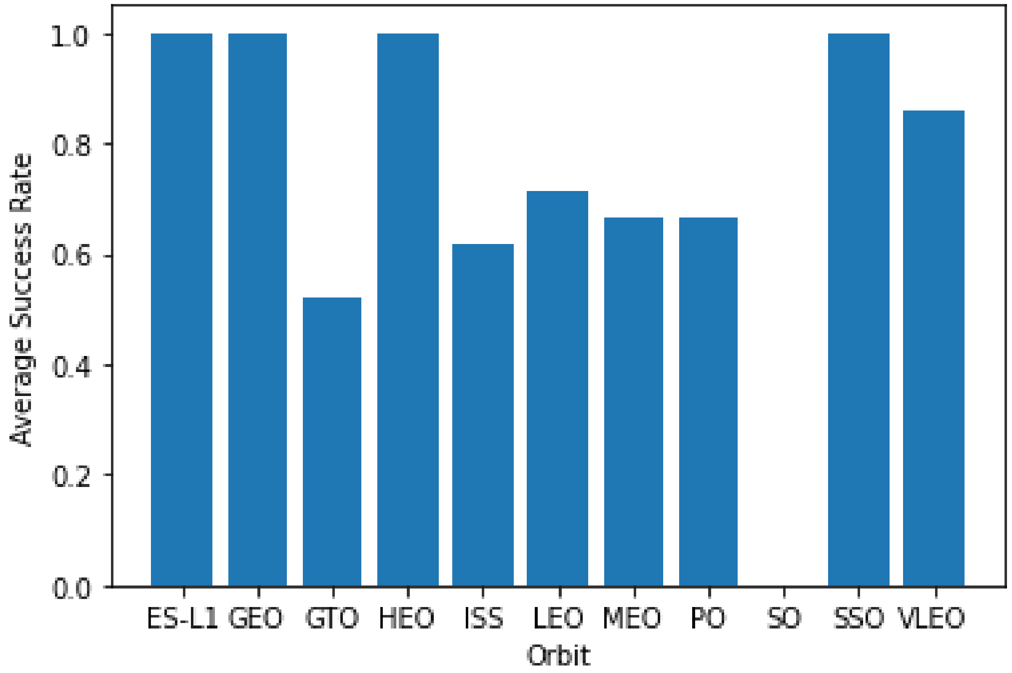
This graph shows the successful and unsuccessful launches of the Launch Site depending on the Payload mass

→ High payloads turned out to be more successful



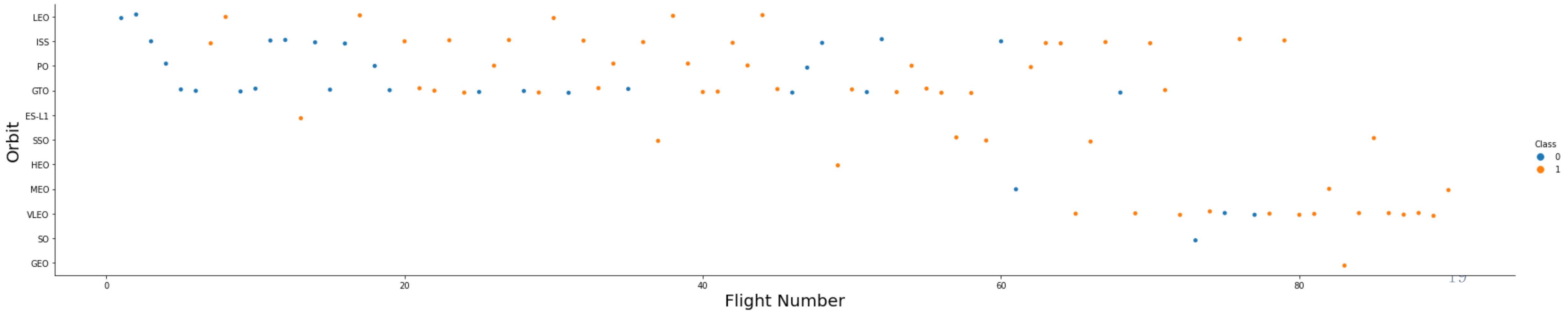
Success rate vs. Orbit type

→ ES-L1, GEO, HEO and SSO launches were always successful!



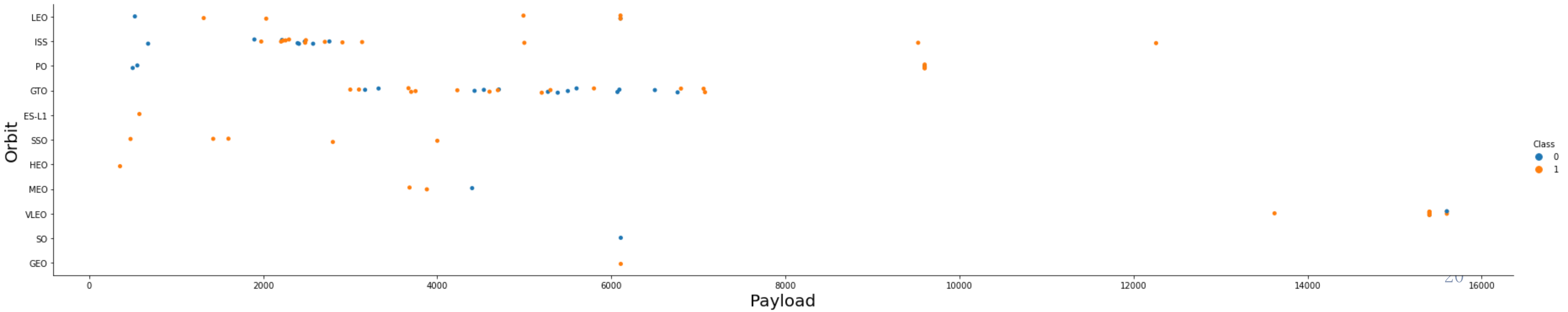
Flight Number vs. Orbit type

→ LEO orbit the Success appears related to the number of flights



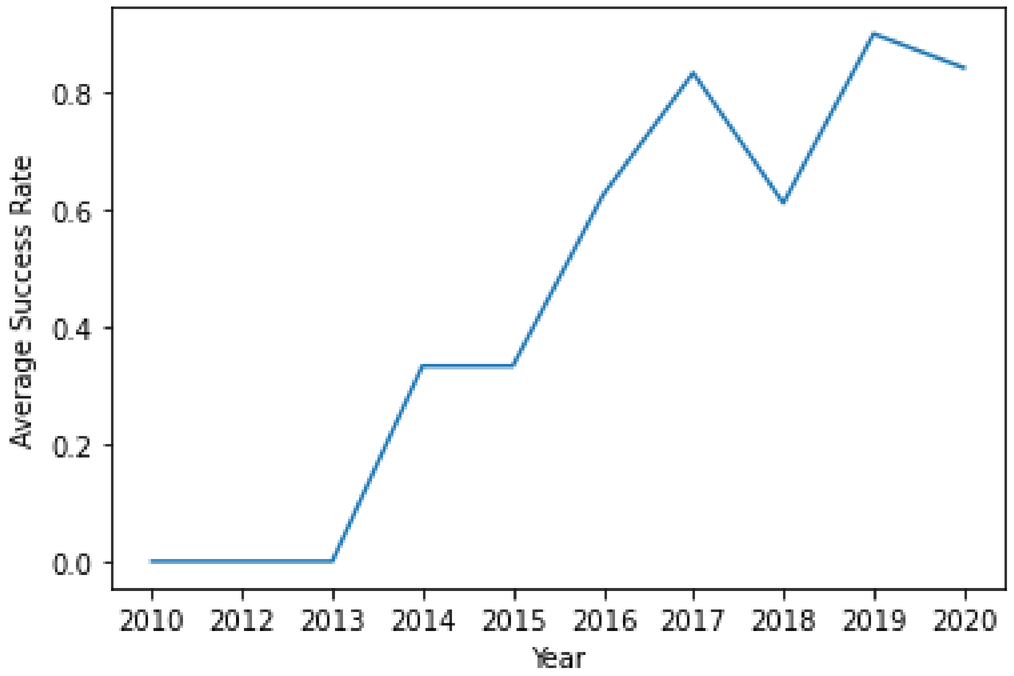
Payload vs. Orbit type

→ Heavy payloads have a negative influence on GTO orbits and positive on GTO and Polar LEO (ISS) orbits.



Launch success yearly trend

→ The success rate of launches increased from 2010 to 2020



EDA with SQL

https://github.com/ahobein/FinalAssignmentSPACEX/blob/main/EDA%20with%20SQL.ipynb

All launch site names

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

There are 4 unique launch sites in the database

Launch site names begin with `CCA`

launch_site
CCAFS LC-40
CCAFS SLC-40

There are 2 sites beginning with 'CCA'

Total payload mass

customer	2
NASA (CRS)	45596

The total payload mass of NASA (CRS) was 45596 kg

Average payload mass by F9 v1.1

1	booster_version
2928	F9 v1.1

The average payload of a F9 was 2928 kg

First successful ground landing date

landing__outcome	earliestdate
Success (ground pad)	2015-12-22

← first record of successful landing

Successful drone ship landing with payload between 4000 and 6000

booster_version	landing__outcome	payload_mass__kg_
F9 FT B1022	Success (drone ship)	4696
F9 FT B1026	Success (drone ship)	4600
F9 FT B1021.2	Success (drone ship)	5300
F9 FT B1031.2	Success (drone ship)	5200

All successful drone ship landings with the payload between 4000 and 6000 were of type F9 FT

Total number of successful and failure mission outcomes

landing__outcome	2
Controlled (ocean)	5
Failure	3
Failure (drone ship)	5
Failure (parachute)	2
No attempt	22
Precluded (drone ship)	1
Success	38
Success (drone ship)	14
Success (ground pad)	9
Uncontrolled (ocean)	2

Most of the mission outcomes were successes

Boosters carried maximum payload

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

All boosters are of type F9 B5

2015 launch records

launch_site	booster_version	landing__outcome	DATE
CCAFS LC-40	F9 v1.1 B1012	Failure (drone ship)	2015-01-10
CCAFS LC-40	F9 v1.1 B1015	Failure (drone ship)	2015-04-14

Both failed landing attempts in 2015 were of type F9 v1.1 and both starting from CCAFS LC-40

Rank success count between 2010-06-04 and 2017-03-20

landing__outcome	number
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

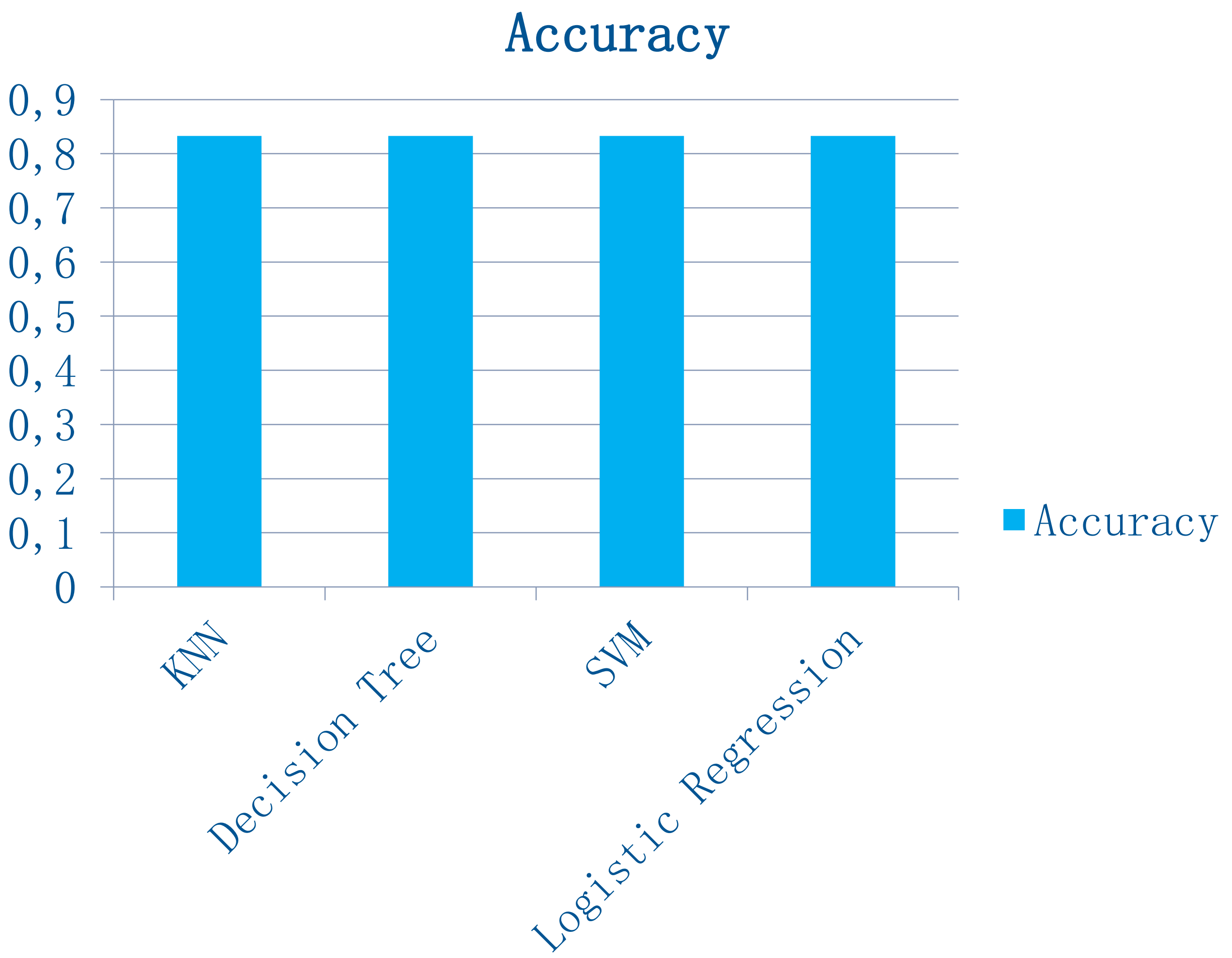
Successes and Failures are about even

Predictive analysis (Classification)

https://github.com/ahobein/FinalAssignmentSPACEX/blob/main/Machine%20Learning%20Prediction%20lab.ipynb

Classification Accuracy

All 4 models which were trained classified the test set with the same level of accuracy



Confusion Matrix

Since all models performed the same, the confusion matrix of all 4 models can be plotted like this

CONCLUSION

- All models were able to predict the landing outcome of the Falcon 9 Rocket using multiple features
- There are no performance differences in the models

APPENDIX