IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Name>
<Date>

Outline

Executive Summary

- Summary of methodologies:
  - Data collection via Jupyter Notebooks and Python
  - Data wrangling using Python tools
  - Dashboard building via Dash tool
  - Machine learning development using the data collection and wrangling

- Summary of all results
  - There is a very high probability the next launch will lead to a successful landing using several parameters. A classification algorithm is the preferred method of selection.

Introduction

- With an increased market for space transportation, many companies are starting to compete with government agencies like NASA. Space X has quickly emerged as one of those major organizations which has been able to cut down the cost of rocket launches thanks to its reuse of the Falcon 9 first stage. Predicting the success rate of the landing of the first stage affects the cost of a launch and the edge of Space X in the market.
- We are trying to determine how to maximize the success rate of a landing given independent variables we may control

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - We collected the data with python tools by calling Space X apis and reformatting the results obtained.
- Perform data wrangling
  - We used a set of Python libraries including pandas to reformat the data for algorithm development.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - We created and tuned classification models to better predict chances of success

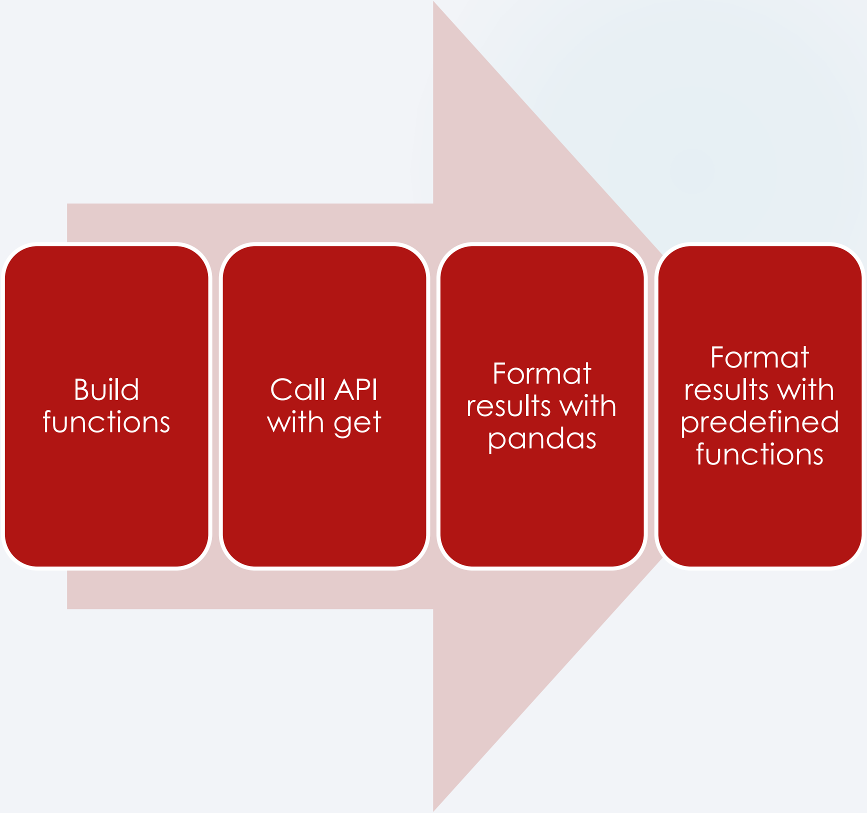
Data Collection

- Data sets collection
  - API of Space X
  - Json files reformatting
  - Python pandas, beautiful soup, and functions used to decipher the data

Data Collection – SpaceX API

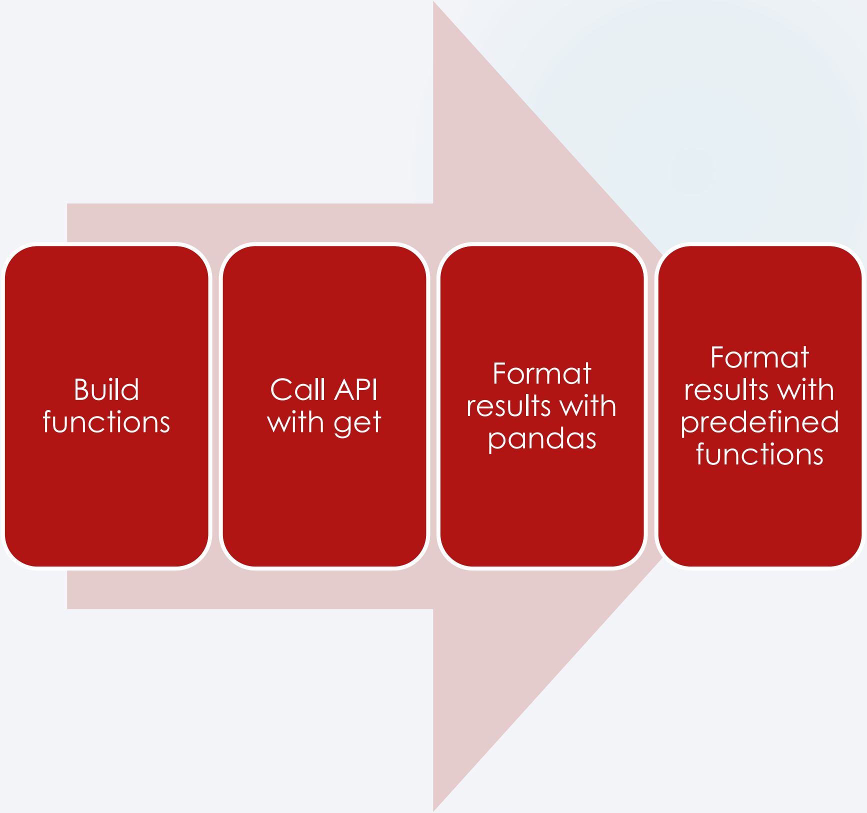
- Call API
- Get launch data as JSON
- Use pandas, functions to reformat the data

- https://github.com/Degueu/testrepo/blob/main/jupyter-labs-spacex-data-collection-api(1).ipynb



Data Collection - Scraping

- https://github.com/Degueu/testrepo/blob/main/jupyter-labs-spacex-data-collection-api(1).ipynb



Data Wrangling

- Read csv from web with pandas
- Discover data discordance like missing values or columns misclassified
- Create derived calculations
- Fill missing data
- Calculate means

EDA with Data Visualization

- Used bar charts
- Used scatter charts
- Used line charts
- Used tables
- https://github.com/Degueu/testrepo/blob/main/jupyter-labs-eda-dataviz(1).ipynb

EDA with SQL

- Used 10 different SQL statements including:
  - **unique launch sites**
  - **NASA's share in the payloads**
  - **Booster versions used**
  - **Historical launch outcomes**
- https://github.com/Degueu/testrepo/blob/main/jupyter-labs-eda-sql-coursera(1).md

Build an Interactive Map with Folium

- We created a folium map to also underline the location of the launch sites as they are very restricted:
- https://github.com/Degueu/testrepo/blob/main/SpaceXDash(2).ipynb

Build a Dashboard with Plotly Dash

- Plots/graphs and interactions you have added to a dashboard
  - A pie chart with success outcomes per site
  - A slider to control a scatter plot
- Reason:
  - Better understand if there is a relationship between launch site and outcomes
  - Provide an easy interactive access for users to ask the data

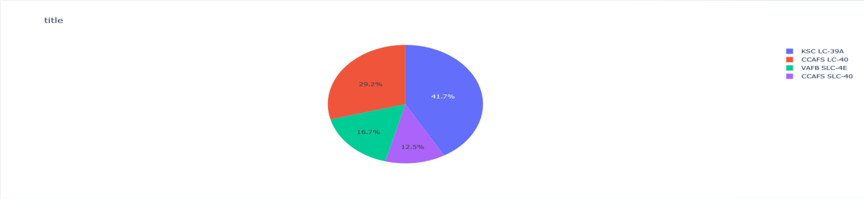
Predictive Analysis (Classification)

- We built and test a few machine learning models (via sklearn) to determine the best pattern (with gridsearch and confusion matrix) that can be used to describe the data:
  - Logistic regression
  - Vector machine
  - Decision tree
  - Kneighbor classifier

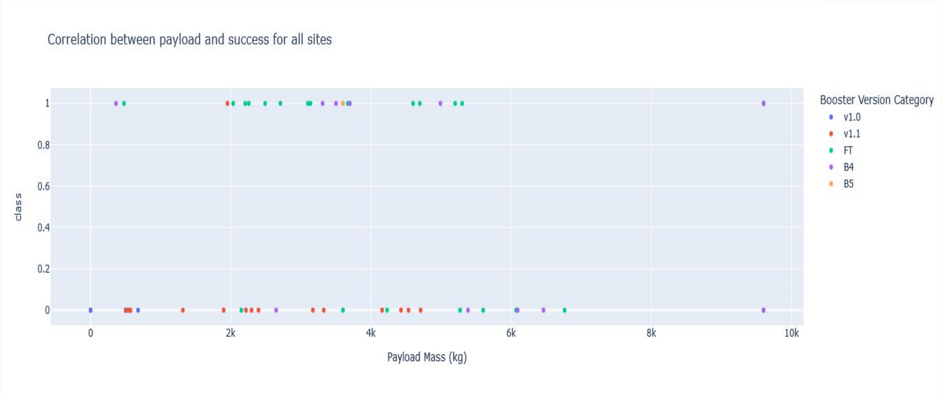
- https://github.com/Degueu/testrepo/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5(1).ipynb

Results

- Exploratory data analysis results
  - Space X has a lot success in their enterprise
  - They can probably feel confident their next launch will result in a success no matter the launch site
  - Looks like higher payloads can lead to more success
- Interactive analytics demo in screenshots



- Predictive analysis results
  - The decision tree seems to be the least performant classifier. A k Neighbor due its simplicity might be a good choice

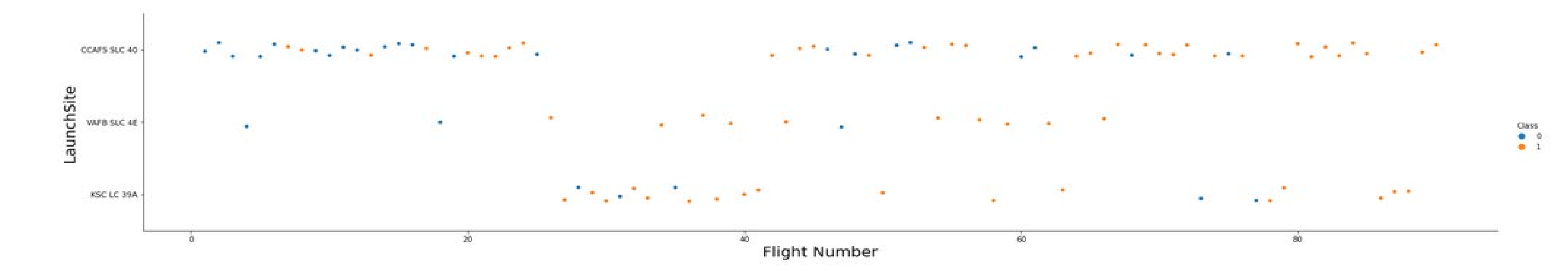


Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

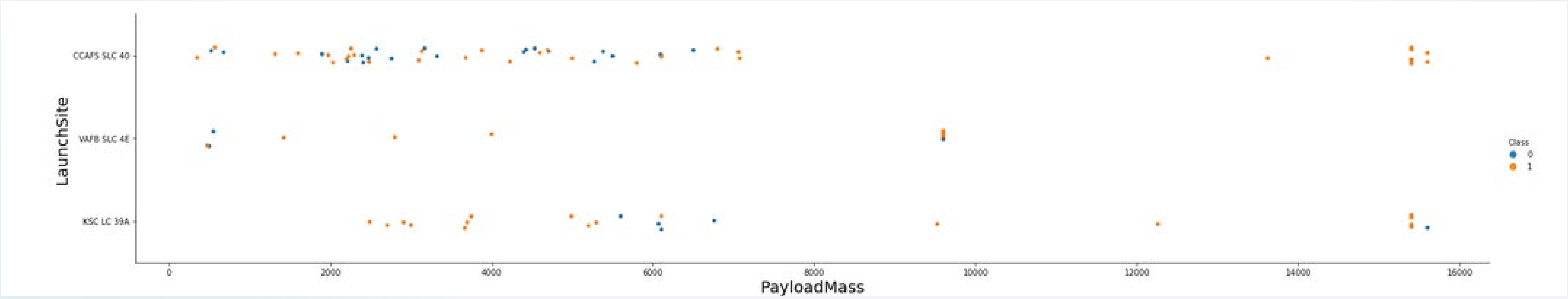
- Scatter plot of Flight Number vs. Launch Site



- Show the screenshot of the scatter plot with explanations
  - Flight number doesn’t seem to have an incidence on launch site outcome. But below flight number 40, we tend to get more failures

Payload vs. Launch Site

- Payload vs. Launch Site

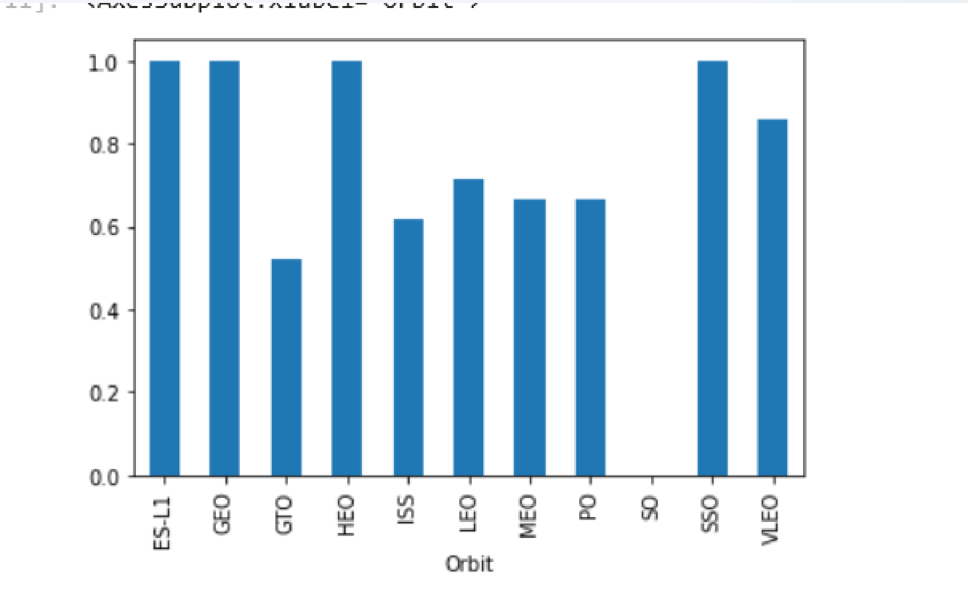


- Payload about 8000 seems to be a good idea if the launch site is not KSC

Success Rate vs. Orbit Type

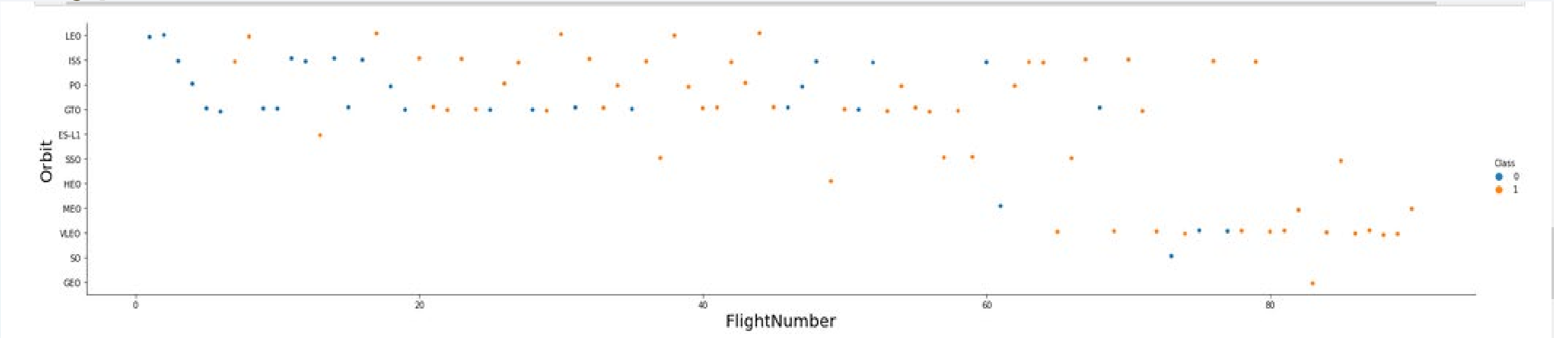
- Bar chart for the success rate of each orbit type

- 5 orbits should be the preferred targets for launches. Example: ESL1



Flight Number vs. Orbit Type

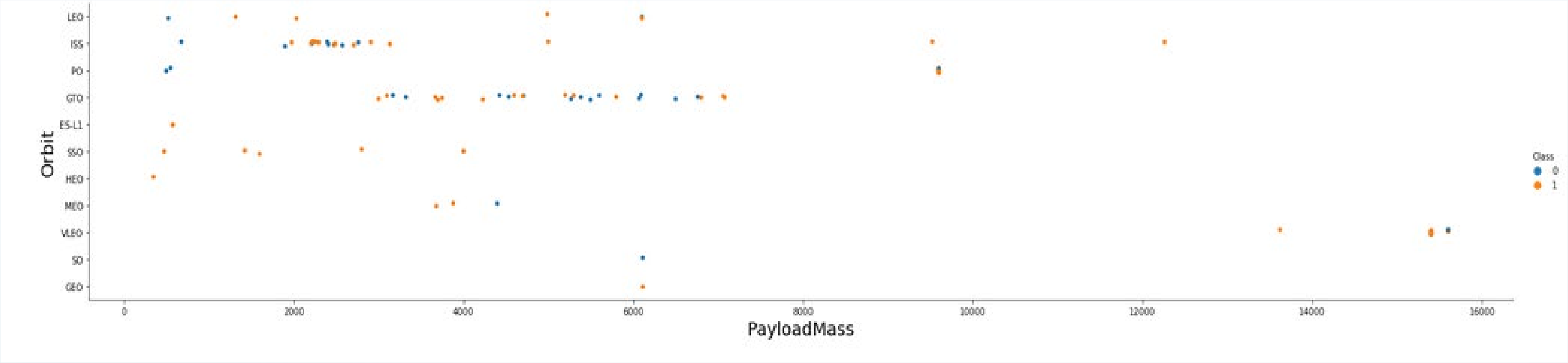
- Scatter point of Flight number vs. Orbit type



- Flight orbit and flight number seems to lead to good success combo. But we can't really point to a clear causality

Payload vs. Orbit Type

- Scatter point of payload vs. orbit type

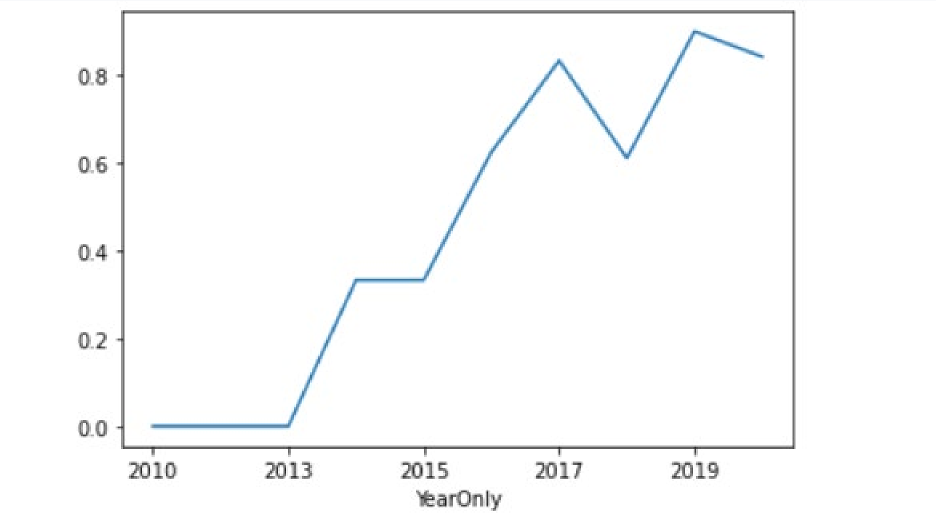


- There doesn’t seem to be a clear pattern. ESL1 and ‘below’ show better success rate under 8000. Above 8000, ISS PO do better

Launch Success Yearly Trend

- Line chart of yearly average success rate

- Space X is encountering more and more success as years pass



All Launch Site Names

- Names of the unique launch sites
- We can see we have 4 main unique sites (maybe 3 as 2 names look very similar CCAFS) we are launching from

[5]:

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

Launch Site Names Begin with 'CCA'

- 5 records where launch sites begin with `CCA`

```
Done.
```

[7]:

	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	None	0	LEO	SpaceX	Success	Failure (parachute)
	2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	None	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
	2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	None	525	LEO (ISS)	NASA (COTS)	Success	No attempt
	2012-10-08	00:35:00	F9 v1.0 B0006	CCAFS LC-40	None	500	LEO (ISS)	NASA (CRS)	Success	No attempt
	2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	None	677	LEO (ISS)	NASA (CRS)	Success	No attempt

- This was generated via a self-managed db2 database using python. It gives us more details about the launch sites

Total Payload Mass

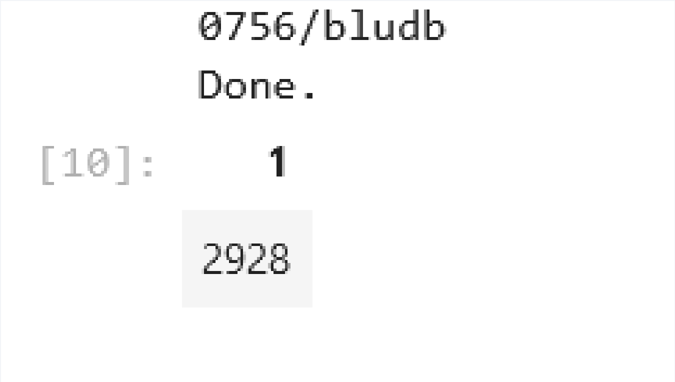
- Below is the total payload carried by boosters from NASA

[9]:

1
107010

Average Payload Mass by F9 v1.1

- Below is the average payload mass carried by booster version F9 v1.1


0756/bludb
Done.

[10]:

1
2928

First Successful Ground Landing Date

- Below is the date of the first successful landing outcome on ground pad

2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Below are the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

[17]:

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

Total Number of Successful and Failure Mission Outcomes

- Below is the total number of successful and failure mission outcomes

Done.

[18]:

mission_outcome	2
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- Below are names of the booster which have carried the maximum payload mass

[19]:

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

2015 Launch Records

- Below are the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

[20]:

booster_version	launch_site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

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

- Below are the counts of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

Done.

[30]:

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

Section 4

Launch Sites Proximities Analysis

Launch sites (folium)

- East Coast: 2 West Coast: 1 U.S.: 3

Launch sites are located in coastal areas to take advantage of weather patterns and the ocean in case of an emergency

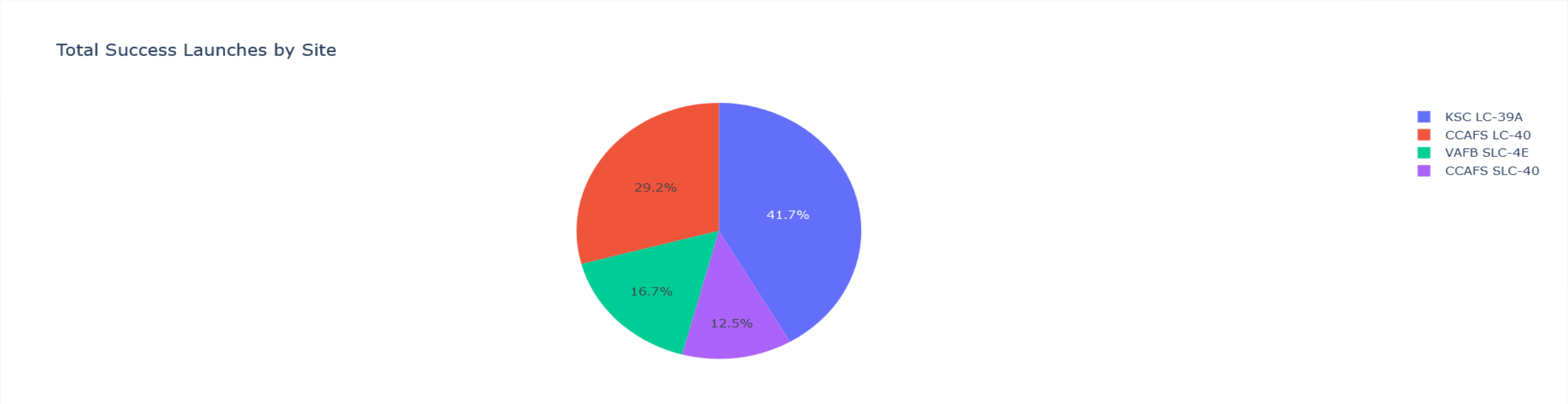
Folium github link:
https://github.com/Degueu/testrepo/blob/main/Space%20X%20map%20info.ipynb

Section 5

Build a Dashboard with Plotly Dash

SpaceX Launch Records Dashboard

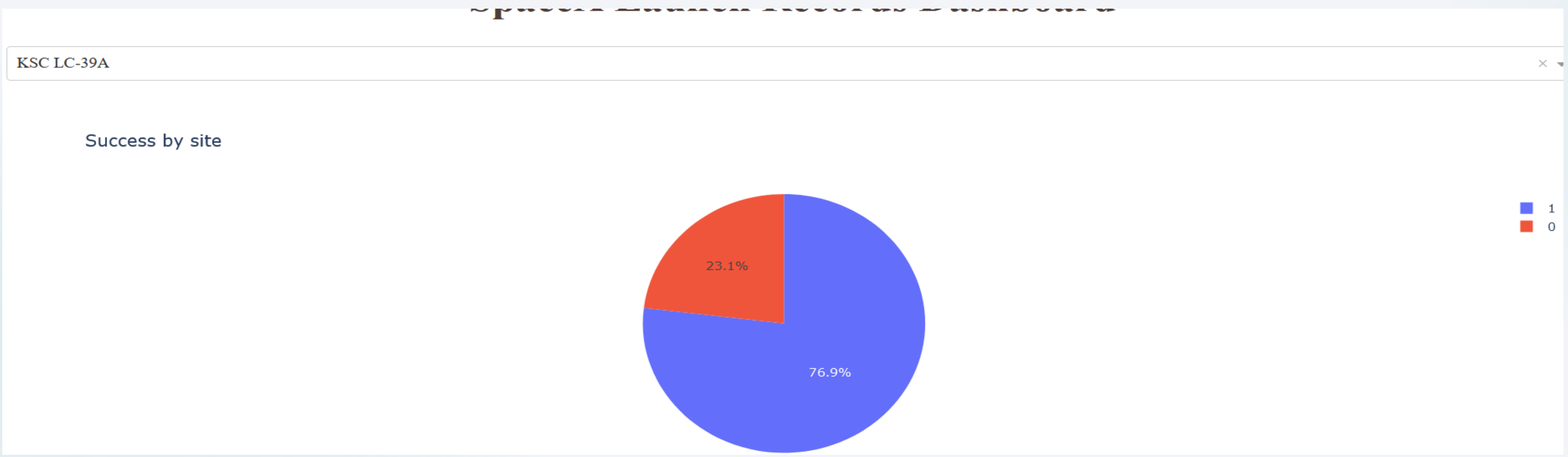
- Below is the screenshot of the launch success count for all sites, in a pie chart



- KSC and CCAFS hold the lion share of launches

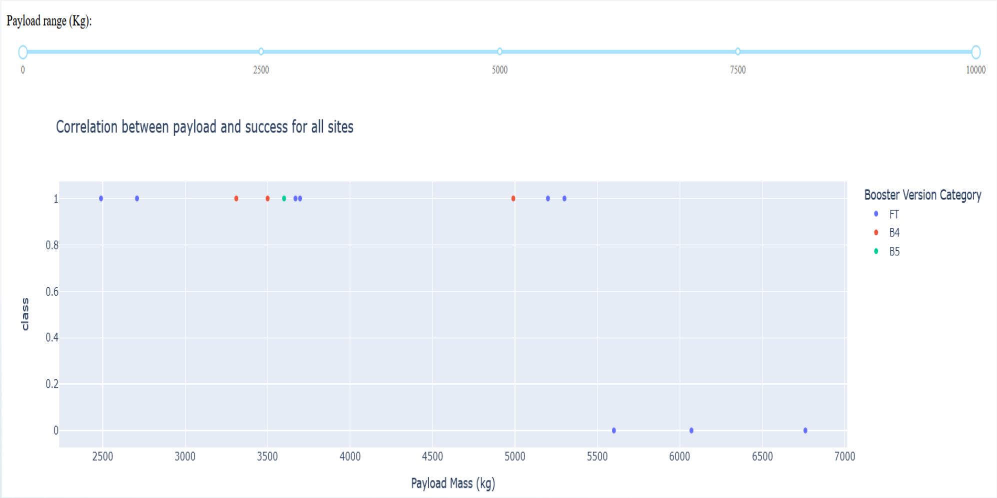
Most successful site

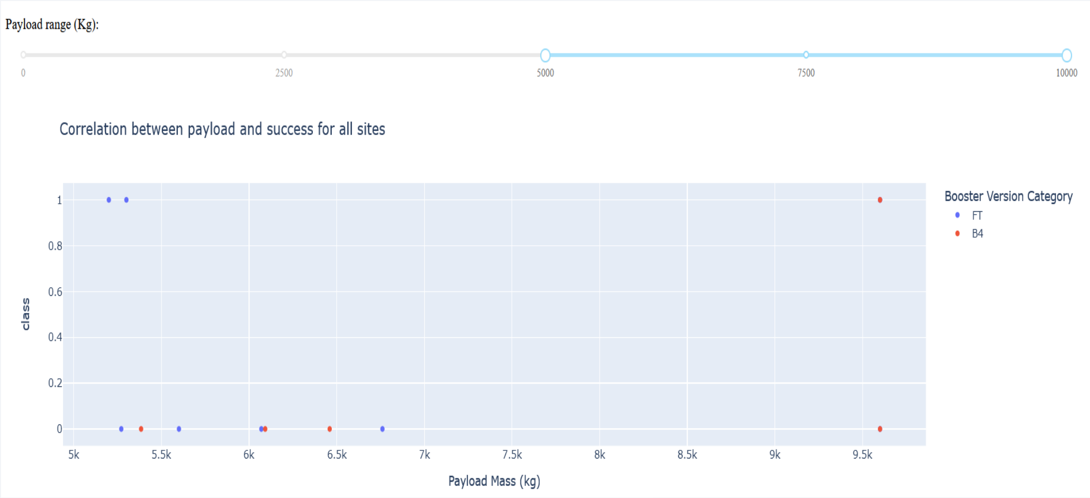
- Below is a screenshot of the pie chart for the launch site with highest launch success ratio



- KSC seems to display the highest success rate.

Correlation between payload and success for all sites





- FT seems to be a better option for launches. Above 8000 it is 50% chance of success

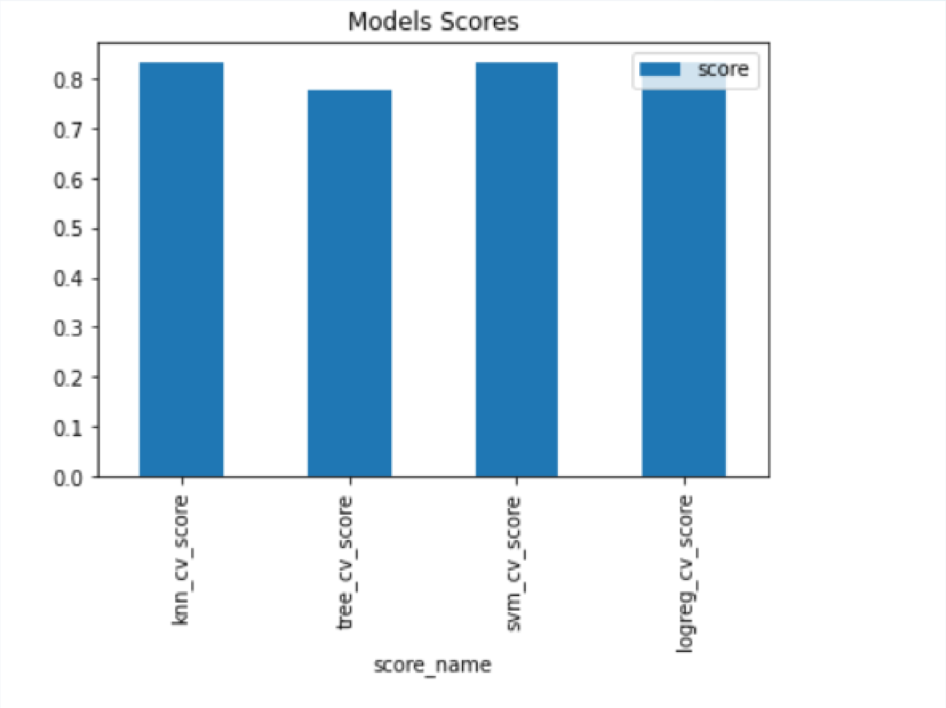
Section 6

Predictive Analysis (Classification)

Classification Accuracy

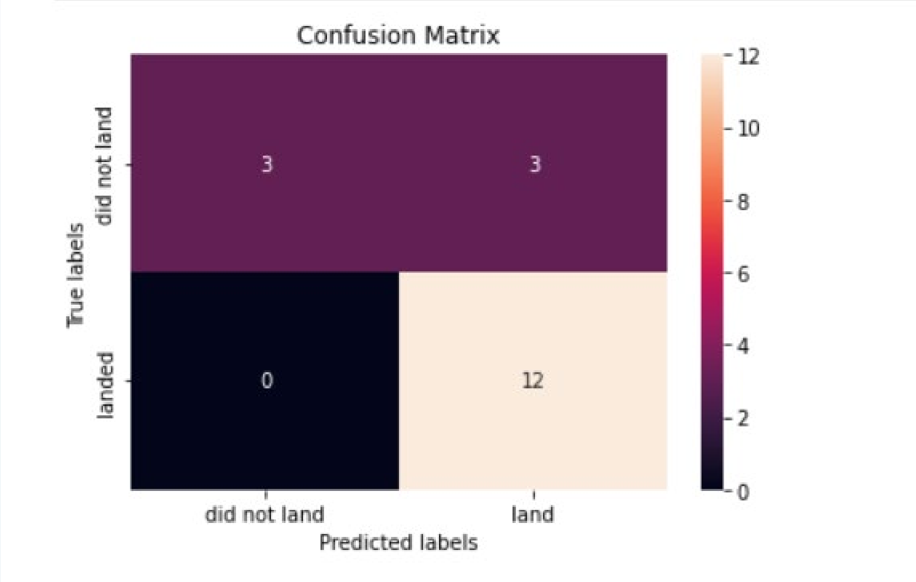
- Below is the built model accuracy for all built classification models, in a bar chart

- A KNN machine may be a good choice based on its easiness.



Confusion Matrix

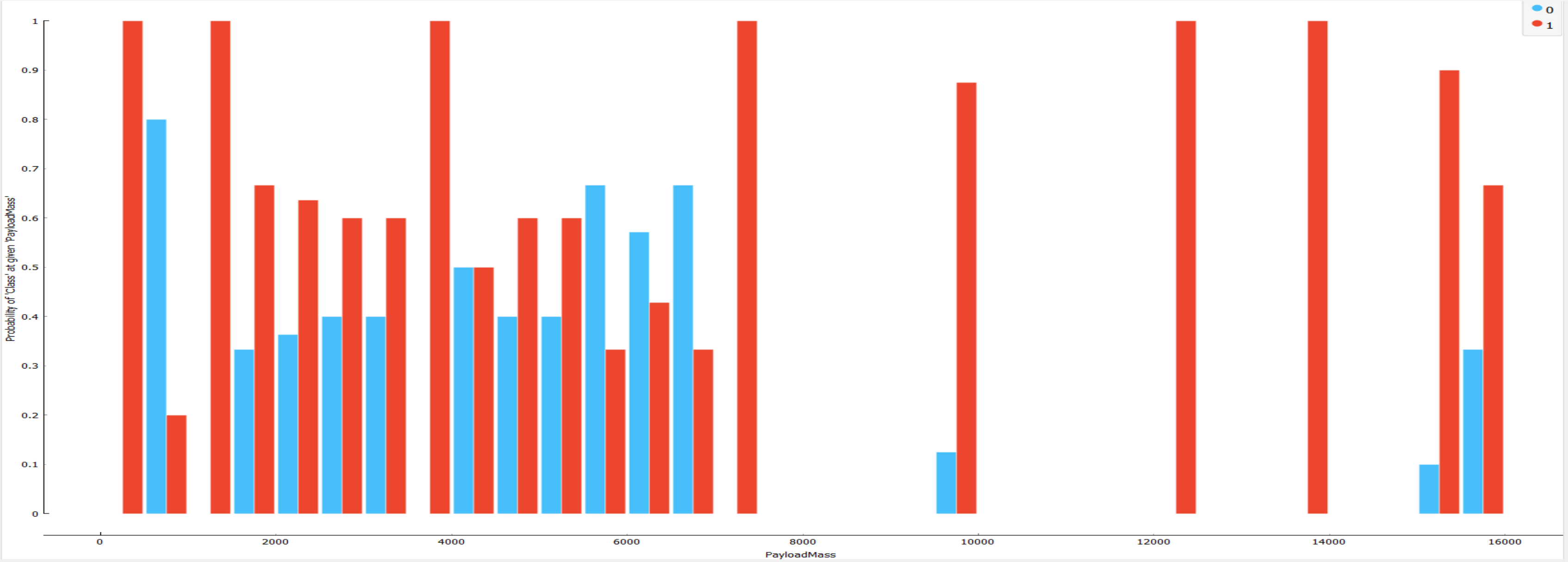
- Below is the confusion matrix for the KNN model. We have 3 false positives and 0 false negative in the testing data. This is a descent result

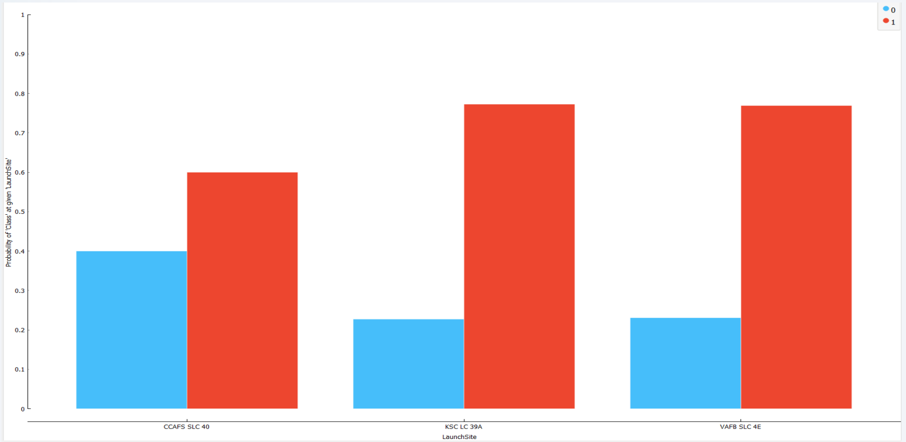


Conclusions

- Space X seems to be on the right trajectory when it comes down to lowering costs of launches by reusing the stage 1
- The company has very little failure rates as time passes by
- The next launch has more than 50% chance of success based on the evolutionary design of the Falcon 9
- The Primary Component Analysis may put the flight number, date and orbits as good candidates
- A naïve Bayes model may also be a good candidate

Appendix


Probability of 'Class' at given 'PayloadMass'
PayloadMass
0
1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
2000
4000
6000
8000
10000
12000
14000
16000


Probability of 'Class' at given 'LaunchSite'
LaunchSite
CCAFS SLC 40
KSC LC 39A
VAFB SLC 4E
0
1

Thank you!