IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Alireza Oveisi Rastabi
28/03/2024

Outline

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

Executive Summary

- Methodologies used
  - Data was collected using SpaceX Rest API, converted, and cleaned
  - Data wrangling and Exploratory Data Analysis was performed using SQL and Visualizations
  - Several types of machine learning techniques such as logistic regression, Decision Trees, SVM, and KNN
- Summary of all results
  - It could be seen with time, number of successful landings increased.
  - Several factors such as weight and launch site affected the success of the landing.

Introduction

- This project included using many different data analysis techniques to predict the outcome of a launch using historical data collected from SpaceX Falcon9 launches.
- SpaceX has already created a huge benefit by having a re-usable base. The aim of this project is to decrease this cost even further by finding what factors can affected the failure or success rate of a launch/landing using predictive machine learning algorithms.

Section 1

Methodology

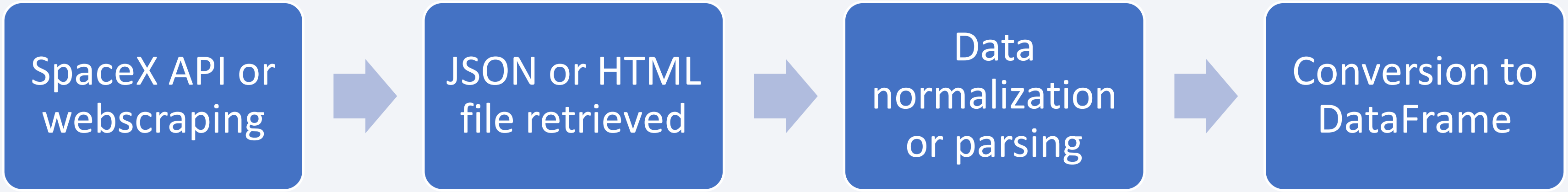
Methodology

Executive Summary

- Data collection methodology:
  - JSON file collected using SpaceX Rest API. This was then normalized and converted into a data frame.
- Perform data wrangling
  - Data was assessed to remove or average null data. An extra column ‘Class’ was also added to identify the successful landings using a Boolean variable
- 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

- Describe how data sets were collected.
  - Two methods were used to try this data collection.
  - First a JSON file was retrieved using the SpaceX rest API which was normalized and converted into a Dataframe.
  - Web scraping was also attempted using Wikipedia as a source. Data was parsed using BeautifulSoup algorithm and converted into Dataframe.



Data Collection – SpaceX API

- SpaceX API = ttps://api.spacexdata.com/v4/launches/past"
- Github Link: https://github.com/alireza-oveisi/capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Request.get

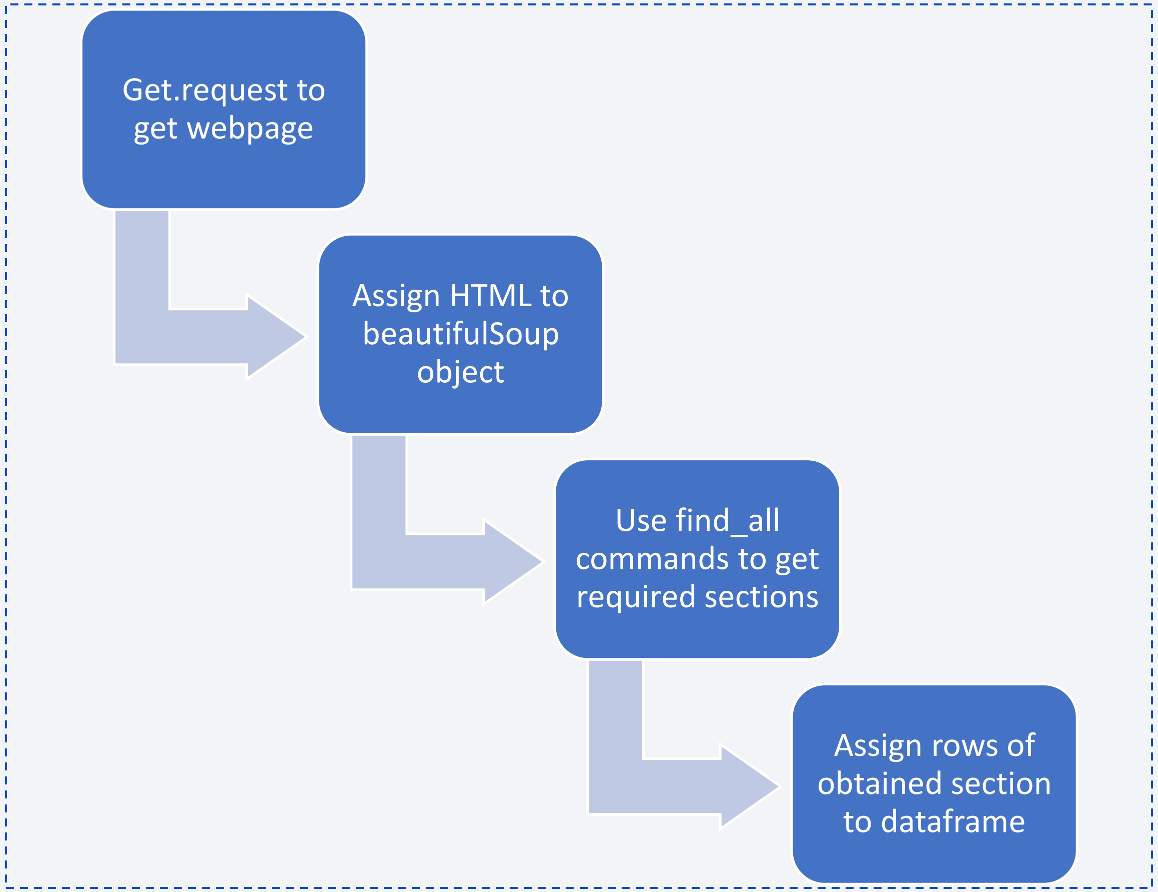
Obtain json.response

Normalize and convert using pd.normalize

Data Collection - Scraping

- Github link:

https://github.com/alireza-oveisi/capstone/blob/main/jupyter-labs-webscraping.ipynb



Data Wrangling

- Github link: https://github.com/alireza-oveisi/capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

Data type conversions (as_type command) → Separate landing outcomes into good and bad → Create new list → Loop through data → Append 0 or 1 to list based on type of landing outcome → Add the class list as a new column to data

EDA with Data Visualization

Github link:
https://github.com/alireza-oveisi/capstone/blob/main/jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

- Several graphs and charts were plotted here to explore the data available.
- These charts were to assess failure and success of flight factors such as flight number, orbit, or payload mass.
- These charts are presented later in the presentation.

EDA with SQL

- Select distinct launch sites
- Select launch sites beginning with ‘CCA’
- Find total mass of boosters launched by NASA (CRS)
- Find average mass of launches with booster version F9 v1.1
- Find date of first successful ground pad landing
- Find successful drone landings with mass between 4000 and 6000
- Sum of mission outcomes (success or fail)
- Find booster versions which carried maximum mass
- Find data for failed drone landings in a certain year (2015)
- Rank number of outcome types between specific dates

Github link: https://github.com/alireza-oveisi/capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
- Circles created for the 4 different launch sites
- Add a cluster with markers for failures and successful missions on each site
- Provided data had an issue and would give errors beyond this point, reported to coursera but not reply
- Finding distances and adding lines to nearest cities, coasts, railways or highways
- These objects were added to the map to showcase the launch sites and their mission outcomes
- It was also used to show common factors between sites such as being away from cities or close to coast
- Link: https://github.com/alireza-oveisi/capstone/blob/main/lab_jupyter_launch_site_location.ipynb

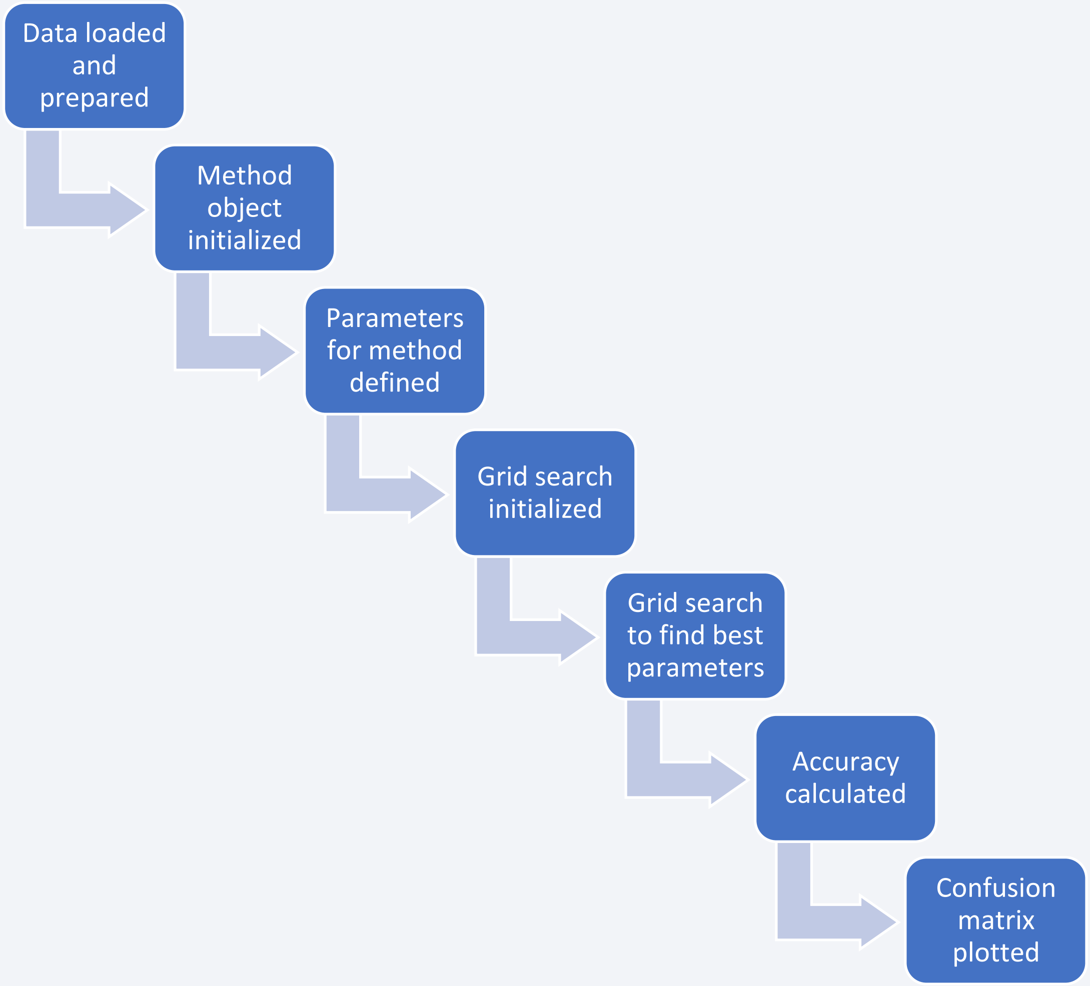
Build a Dashboard with Plotly Dash

- Elements of dashboard:
  - Drop down box to select site
  - Slider to filter payload mass
  - Pie chart to show percentage of landing outcomes
  - Scatter graph to show payload mass vs. mission outcome (color coded by booster version)
- Images of dashboard on next slide
- Github link: https://github.com/alireza-oveisi/capstone/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

- 4 different methods were tested
  - Logistic Regression
  - Decision Tree
  - Support Vector Machine
  - K Nearest Neighbor

Github link: https://github.com/alireza-oveisi/capstone/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5%20(1).ipynb



Results

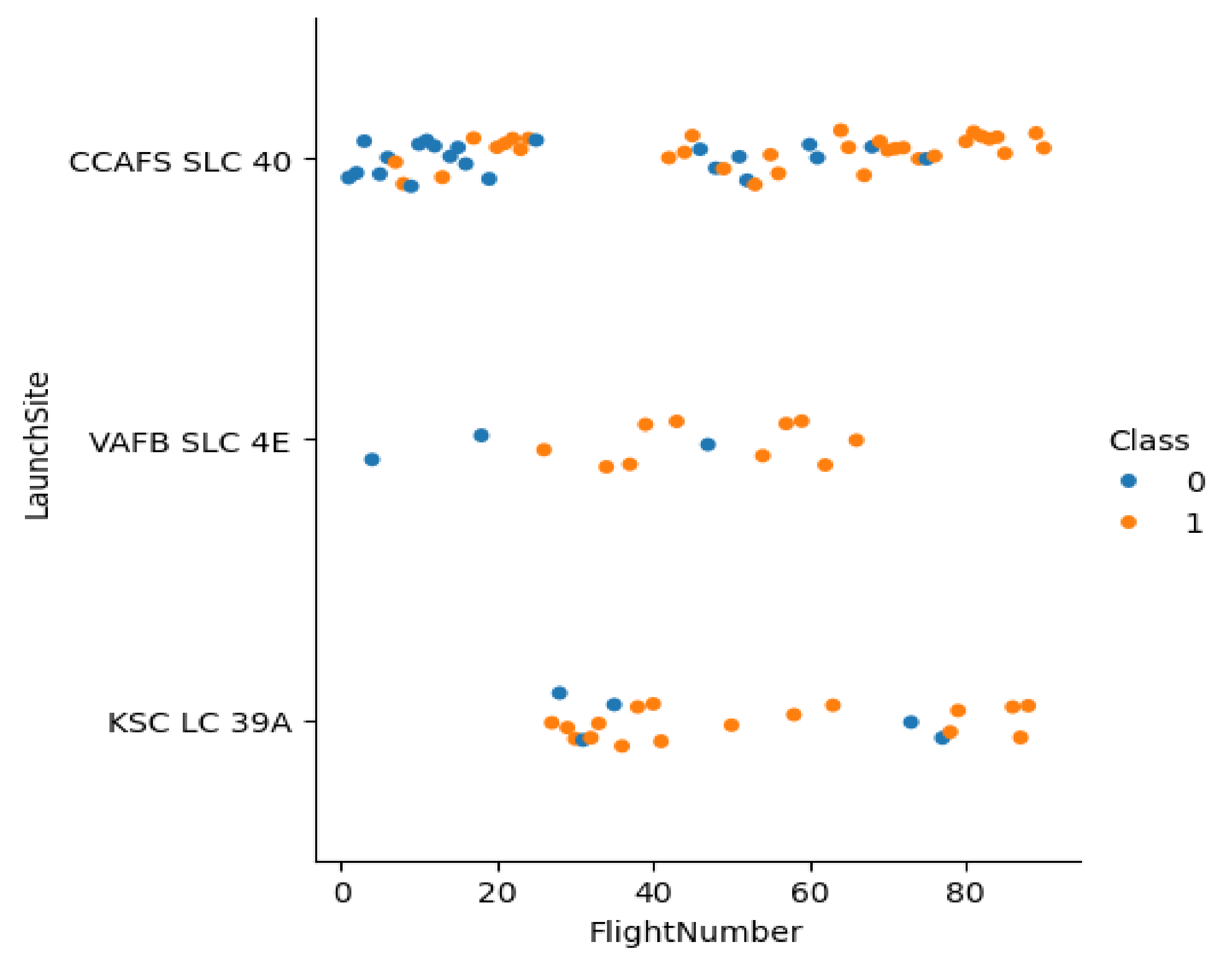
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

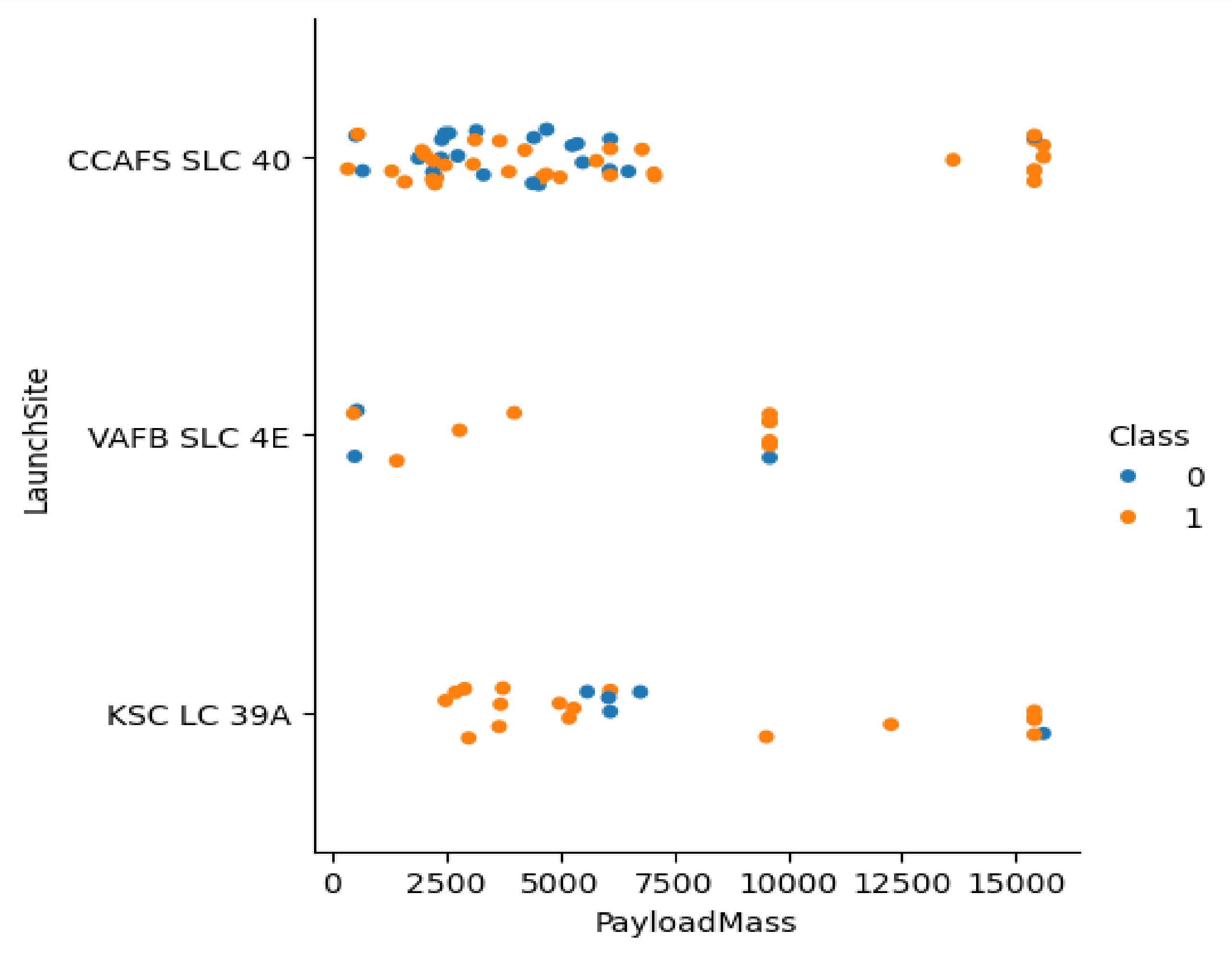
- Shows that mostly flight success rate increased for all sites as flight number went up



Flight outcome based on site and flight number

Payload vs. Launch Site

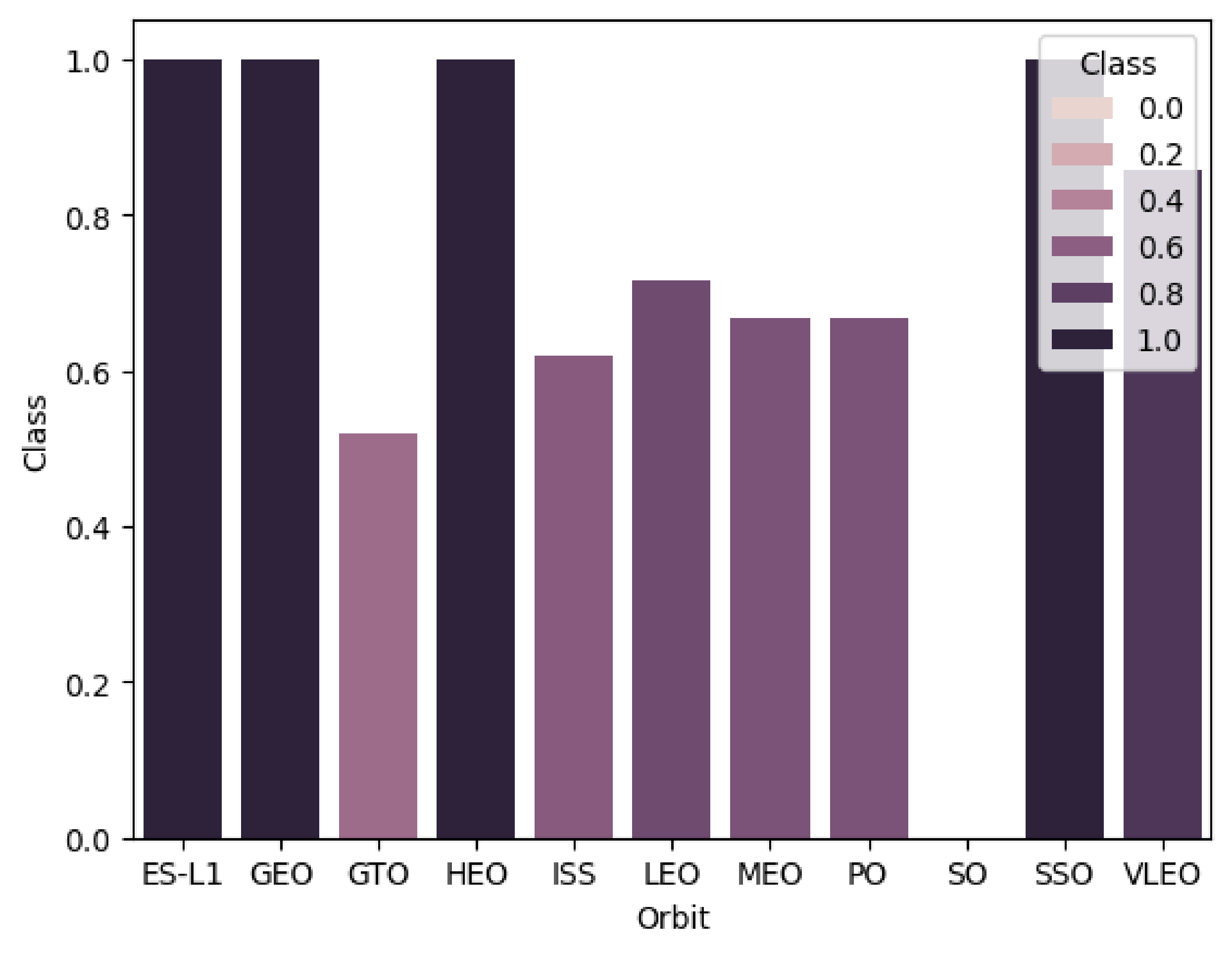
- Some small correlation can be seen for payload mass increase to success rate



Flight outcome based on site and mass

Success Rate vs. Orbit Type

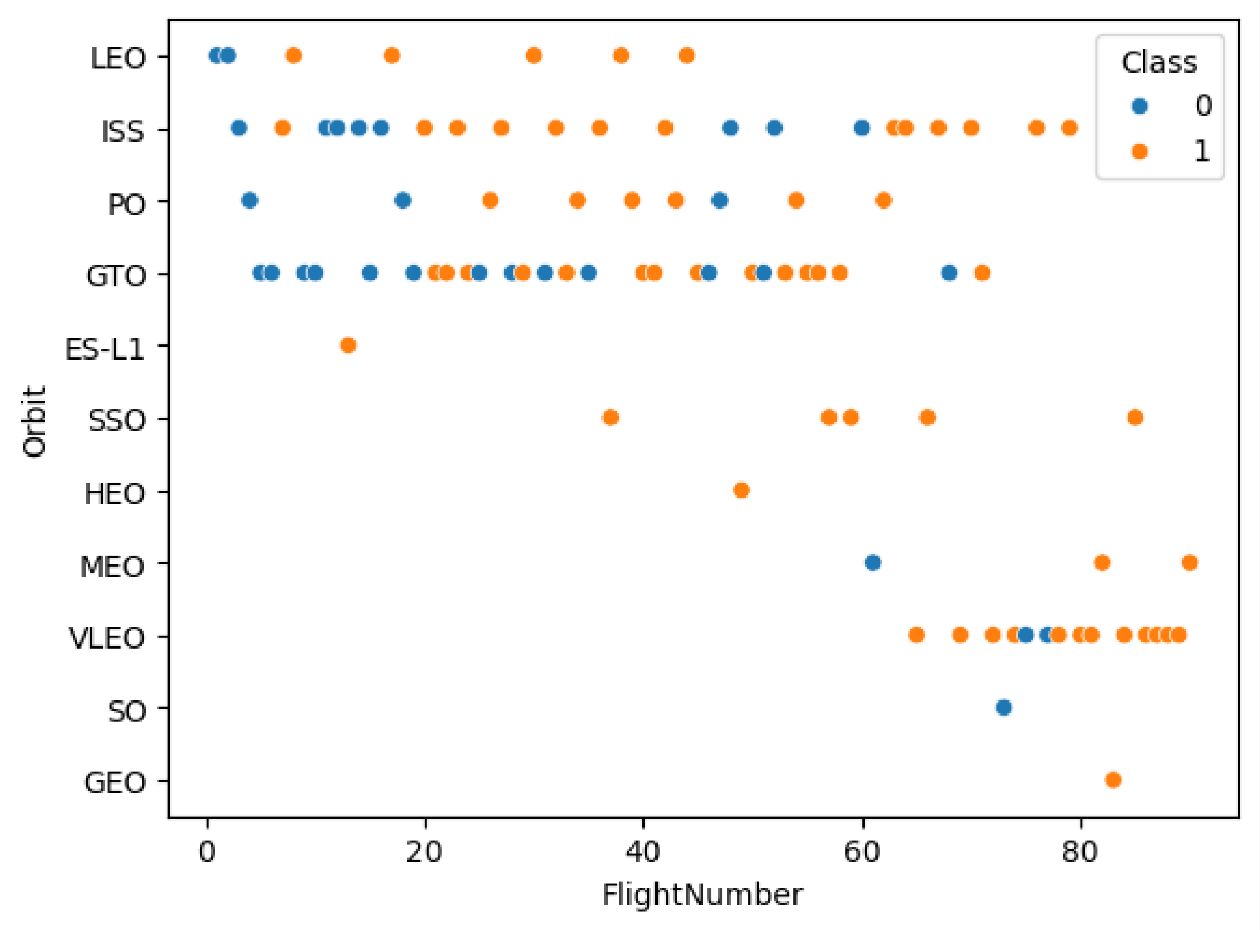
- Some orbits have a higher success rate than others (ES-L1, GEO, HEO, and SSO)
- Otherwise, similar rates.



Mean success rate based on destination orbit

Flight Number vs. Orbit Type

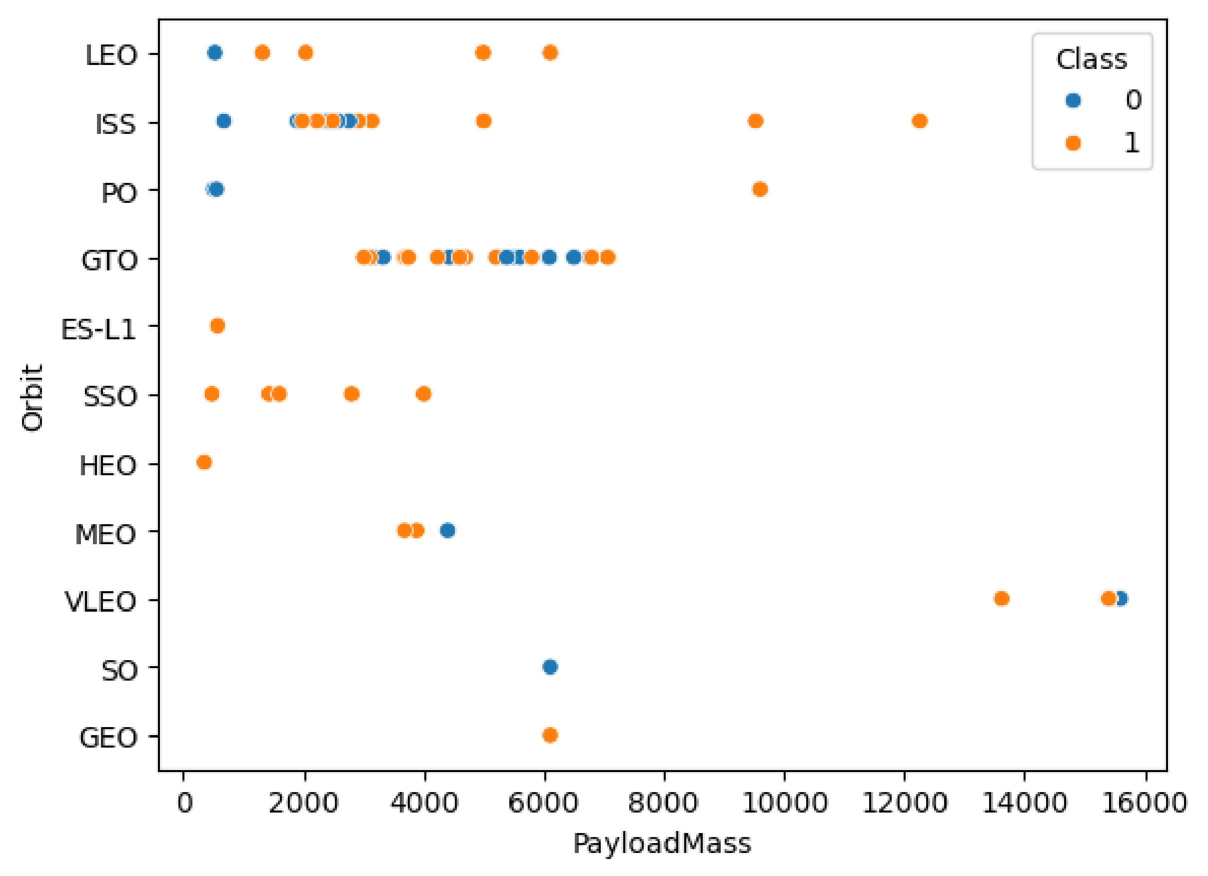
- Most orbits have better success rates as flight numbers went up
- Some failures occasionally during later flights



Flight outcome based on flight number and orbit

Payload vs. Orbit Type

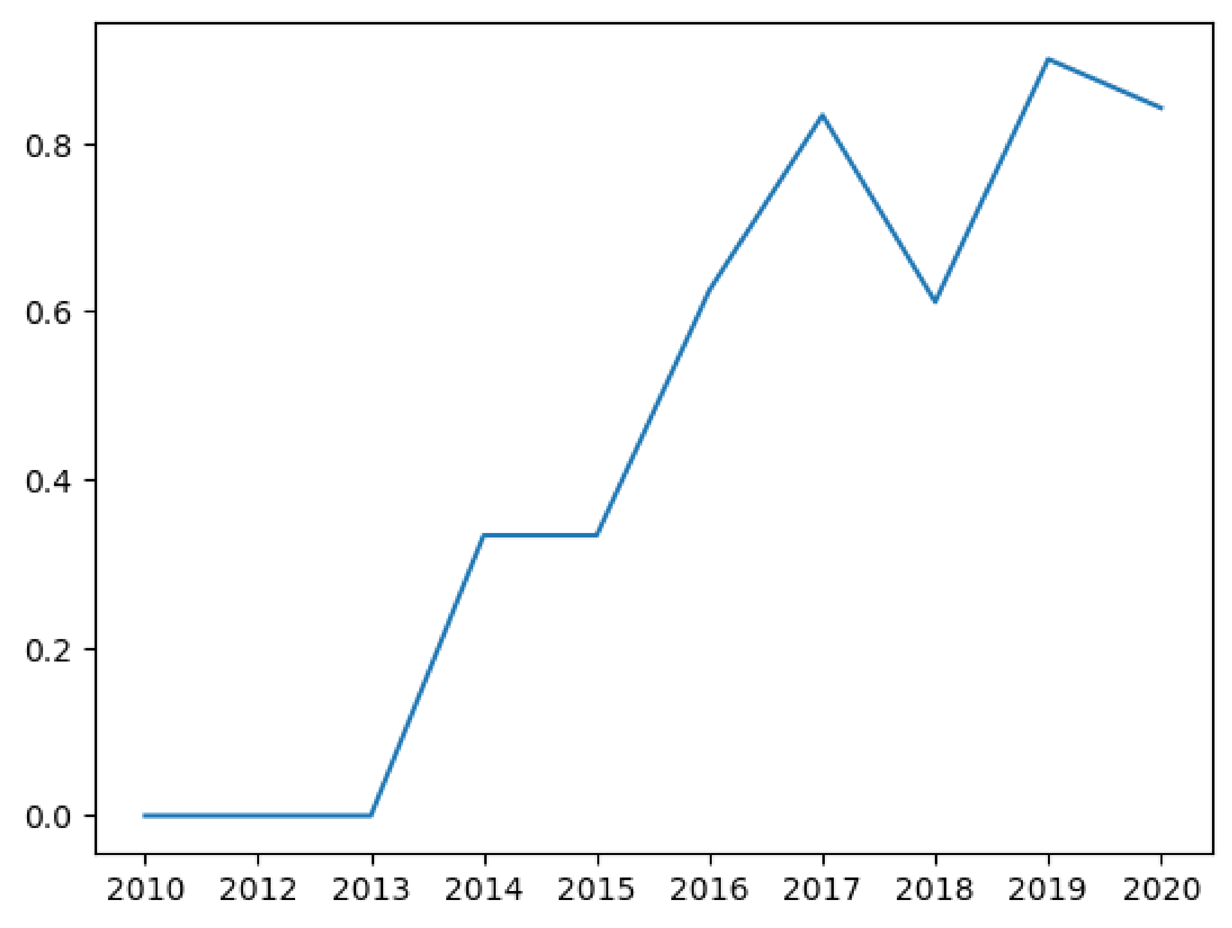
- Larger payload mass correlated to success of mission
- True for most orbits except GTO



Flight outcome based on mass and orbit

Launch Success Yearly Trend

- Significant yearly increase in mission success rate
- Small dip between 2018 and 2019, could have been technical difficulties



Mean flight outcome vs time

All Launch Site Names

- 4 Unique launch sites

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

Launch Site Names Begin with 'CCA'

Done.

16]:

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 mass of 45596 KG carried by NASA missions

```
]: %%sql
   select sum(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE CUSTOMER = 'NASA (CRS)'

 * sqlite:///my_data1.db
Done.
```

sum(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

- Average payload mass of 2928.4KG for F9 v1.1 boosters

```
%%sql
SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE BOOSTER_VERSION = 'F9 v1.1'
```

```
 * sqlite:///my_data1.db
Done.
```

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

- First Successful ground pad landing was on 22 December 2015

```
%%sql
SELECT MIN(DATE) FROM SPACEXTBL WHERE LANDING_OUTCOME = 'Success (ground pad)'
```

```
 * sqlite:///my_data1.db
Done.
```

MIN(DATE)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Successful drone ship landings with mass between 4000 to 6000 KG

Done.

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

Total Number of Successful and Failure Mission Outcomes

- Mission outcomes mostly successful except one failure in flight

```
%%sql
SELECT MISSION_OUTCOME,COUNT(*) FROM SPACEXTBL GROUP BY MISSION_OUTCOME
```

```
 * sqlite:///my_data1.db
Done.
```

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

Boosters Carried Maximum Payload

- All booster versins with max payload shown below

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

- 2 Failed drone landings occurred in 2015, shown below.

MONTH	Date	Booster_Version	Launch_Site	Landing_Outcome
01	2015-01-10	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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

- Ranking of landing outcomes between the date 2010-06-04 and 2017-03-20

27]:

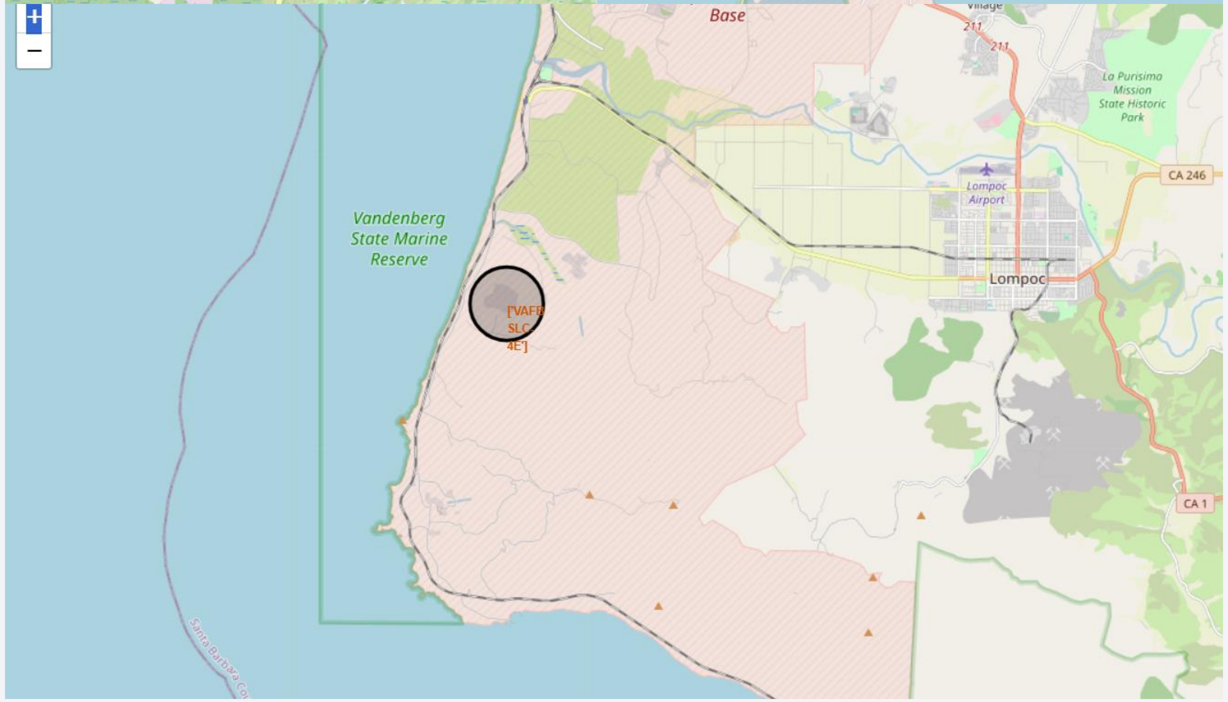
Landing_Outcome	COUNTER
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

Site locations

- 4 Launch Site locations marked on a folium map with circles (3 in top photo, 1 in bottom)



<Folium Map Screenshot 2>

- Jupyter notebook showed an error when trying to complete this, could not achieve results

<Folium Map Screenshot 3>

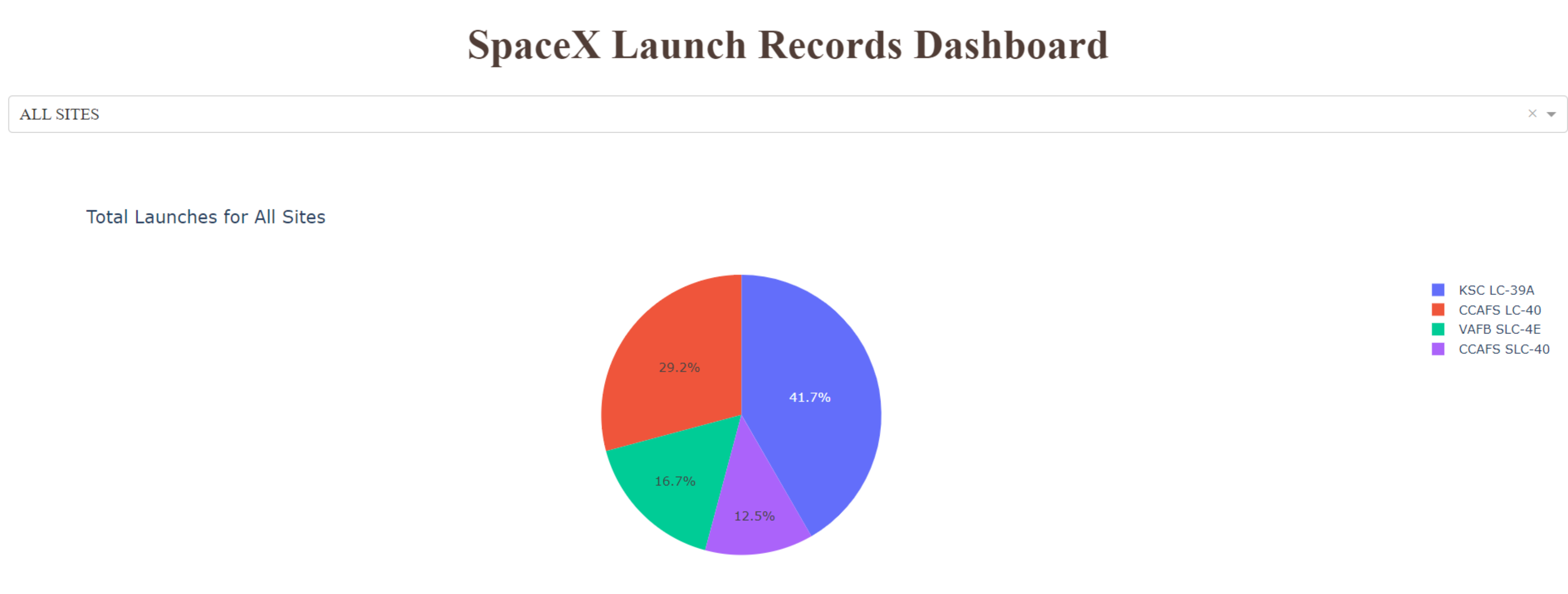
- Jupyter notebook showed an error when trying to complete this, could not achieve results

Section 4

Build a Dashboard with Plotly Dash

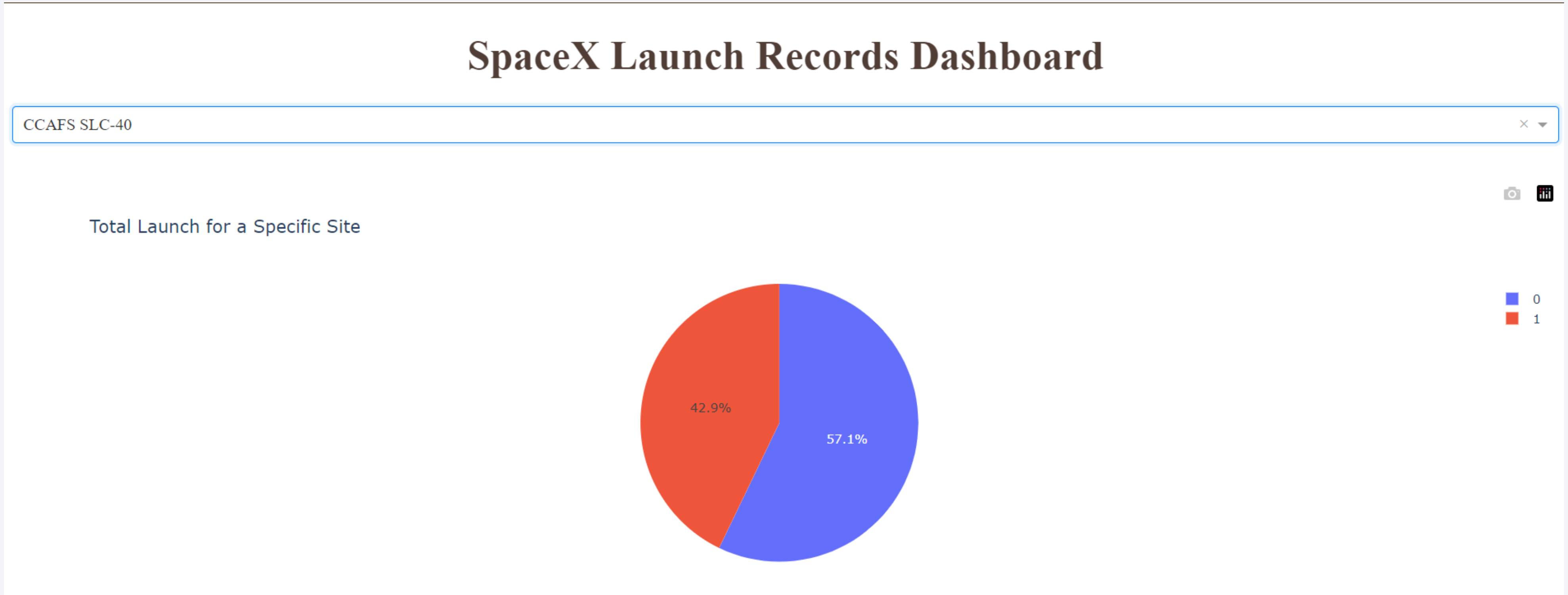
All site launches

- Pie chart showing the percentage of launches for each site



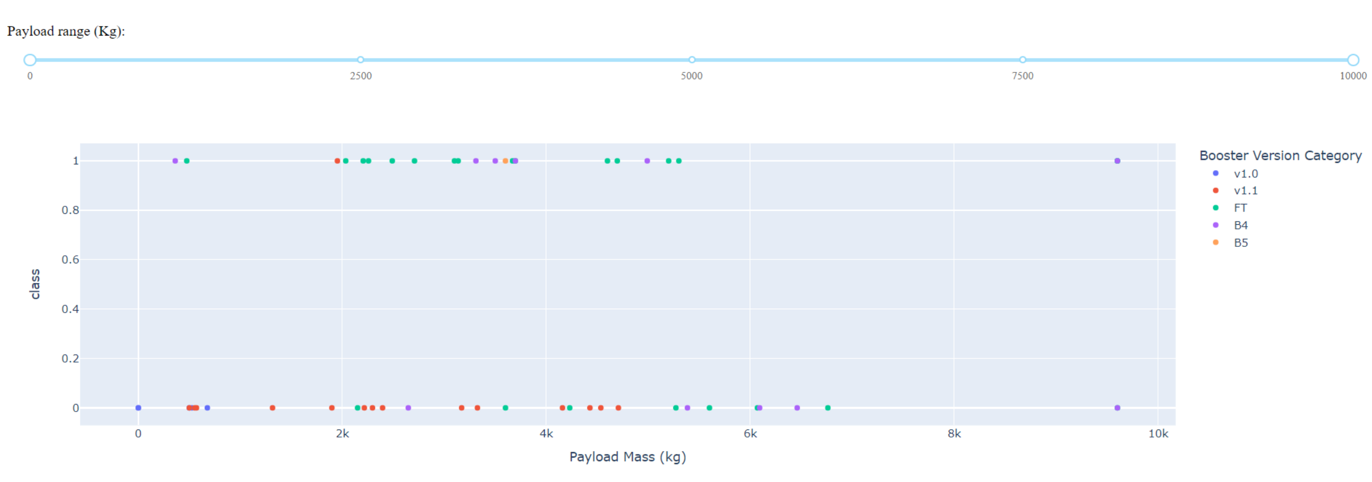
Launch site with highest success

- Site with the highest success ratio for launches. This was found to be the CCAFS SLC-40 site with 42.9% success rate

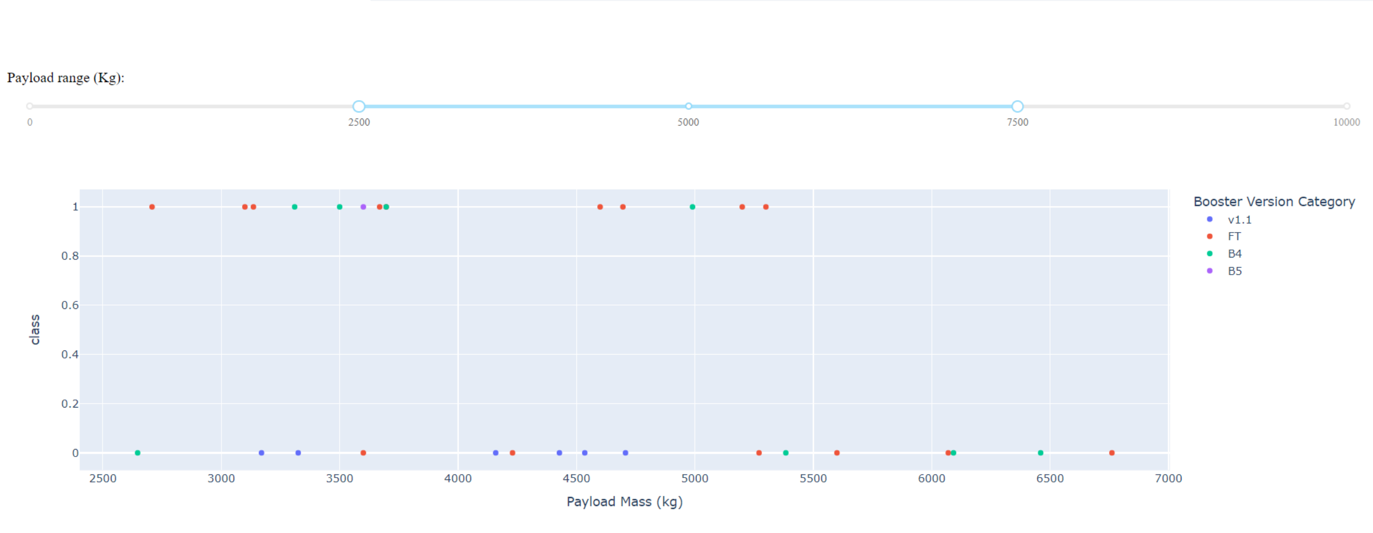


Slider and scatter graph

- Image shows slider that can filter payload mass and show a scatter graph of successes color coded by booster version



- Slider set to full range



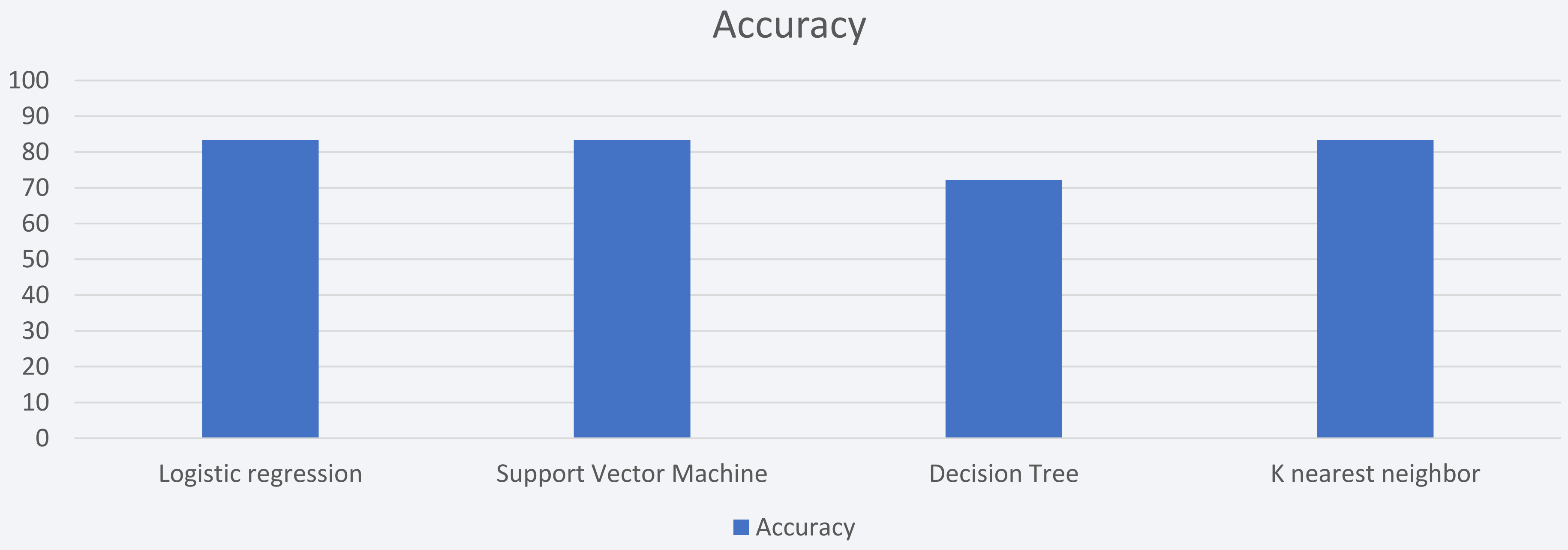
- Slider set to 2500 to 7500 KG

Section 5

Predictive Analysis (Classification)

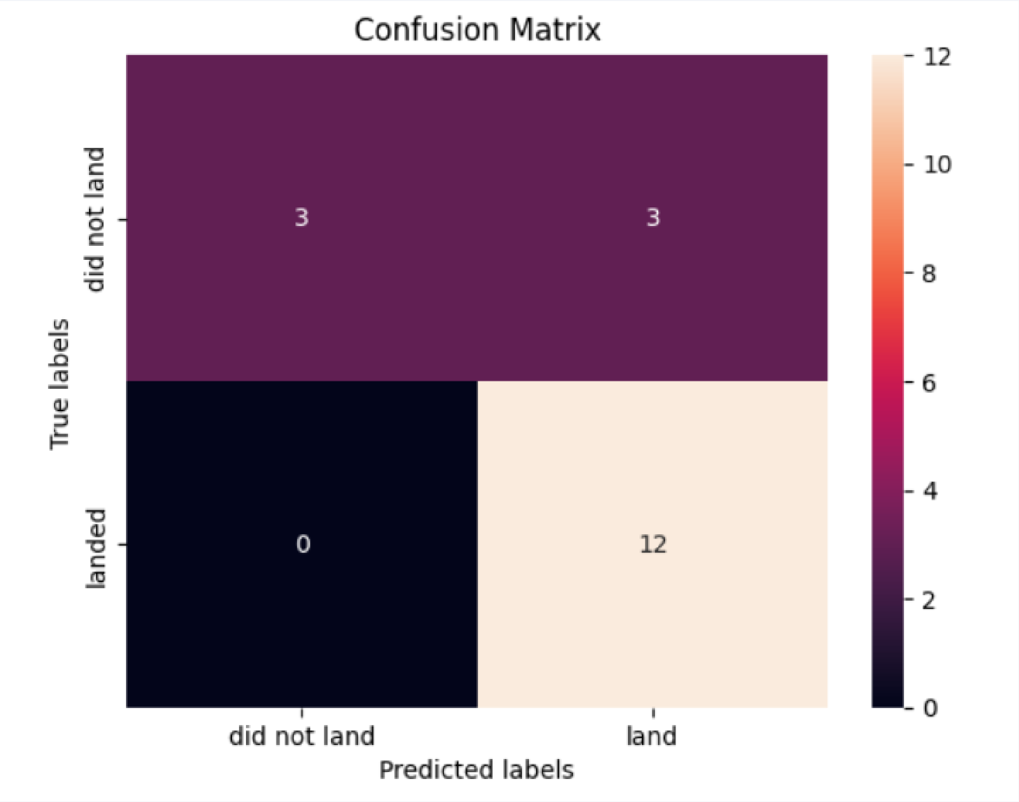
Classification Accuracy

- All models showed similar accuracy after grid search process (83.3%) except decision tree which was (72%)



Confusion Matrix

- The three methods with similar accuracy also showed the same confusion matrix



Conclusions

- Significant increase in mission success rates after 2013
- Highest launch rates has been on KSC-LC 39A site
- CCAFS SLC-40 has the highest success mission rates
- Predictive analysis mechanisms used have similar accuracy and confusion matrix with their weakness being the false positives
- Most successful landing types between 2010 and 2017 were drone ship landings
- Most successful orbit for missions was SSO

Appendix

- All codes written can be seen in the github links.

Thank you!