IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Josh
12/19/23

Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

- Other space launch companies will have to compete with SpaceX which is able to reduce costs be landing and reuse the first stage. Competitors would like to know if a launch for a given mission profile and rocket configuration will successfully land to competitively price their services.

- Can we predict if the SpaceX first stage will successfully land give attributes about the rocket and mission profiles.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Launch data was pulled using a provided API. Data available included the **launch outcome, rocket, payload and launch site details**
- Data Wrangling
  - The landing outcomes were classified as either 0 for ‘bad outcomes’ or 1 for ‘good outcomes’ and added as a new column ‘Class’ to the data set
  - Null payload values were replaced with the average payload value for this analysis
  - Categorial features that were identified as having a correlation to launch success were ‘one-hot’ encoded

Methodology

Executive Summary

- Perform exploratory data analysis (EDA) using visualization and SQL
  - Using inline SQL in a Jupiter Notebook, we identified the types of landing outcomes and their count, launch sites boosters, and total and average payload per customer in kgs
  - Using plotly and seaborns libraries we plotted multiple scatter plots to see if there is a correlation between each feature in the dataset and the success of a launch
  - A lie chart of showing average success created to see if there is a trend with an increase in time
- Perform interactive visual analytics using Folium and Plotly Dash
  - Using Folium, we plotted the location of each launch site, as well as the number of both failed and success full launches, as well as show the proximity to infrastructure and cities
  - Using a Plotly Dash, we created a interactive tool, allowing the user to see the launch outcome for each site, and for a range of payloads

Methodology

Executive Summary

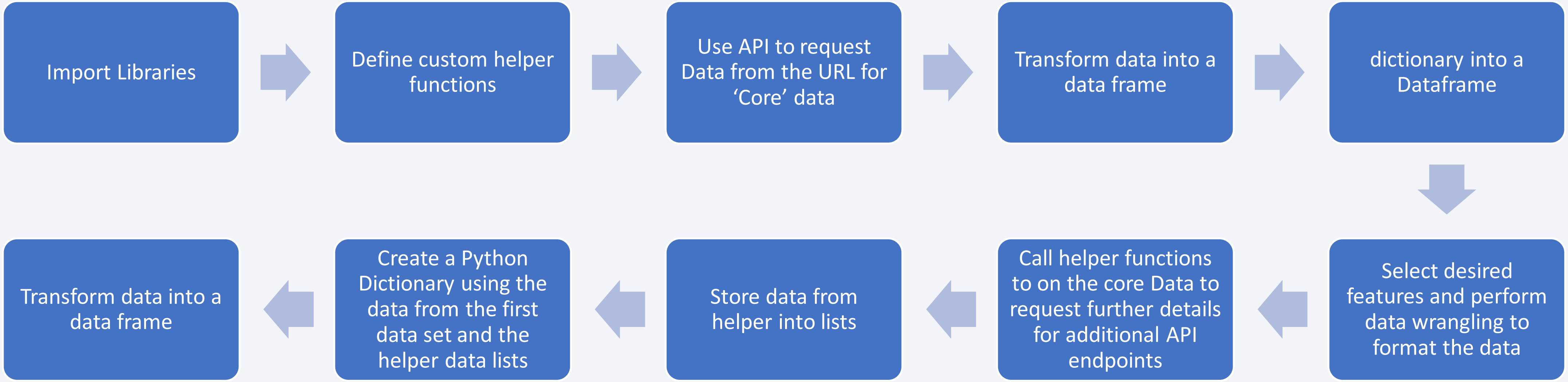
- Perform predictive analysis using classification models
  - Using the data set after feature engineering and data wrangling, we standardize the data and split the data into test and train set with test set at 20%
  - Multiple models are to be used and compared
- Using the GridSearchCV library and a selection of different models we:
  - Fitted each model using the training set to get the best parameters for each model
  - Scored each model using the test set to identify the best model for the data

Data Collection

- Data sets were collected from either and API call to retrieve a CSV File or web scrapped using a Python Library Beautiful soup to read the data from Wikipedia tables.
- A breakdown of each flow is outlined in the next 2 slides.

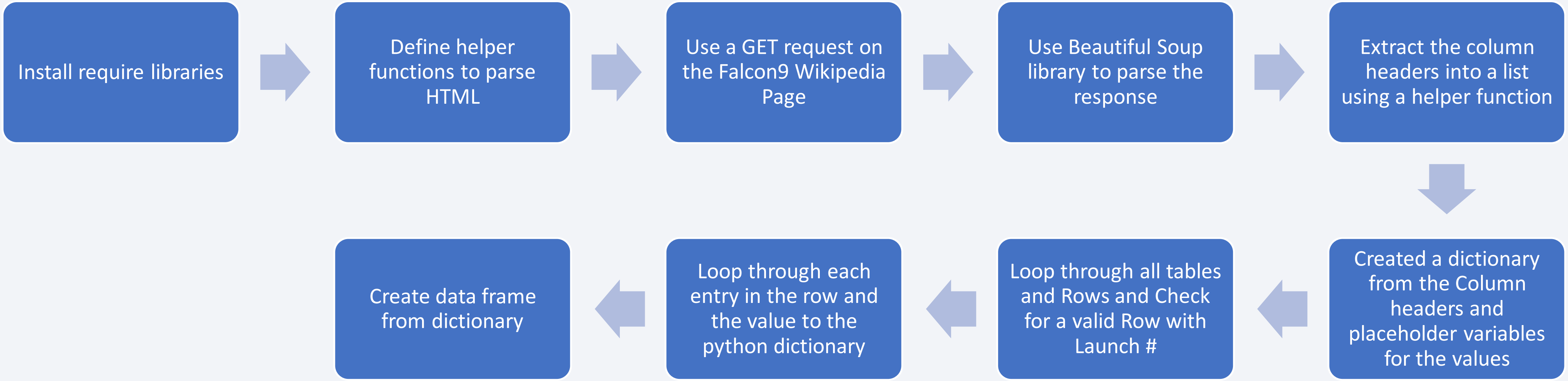
Data Collection – SpaceX API

- https://github.com/JGundisalvus/IBM_DS_Capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb



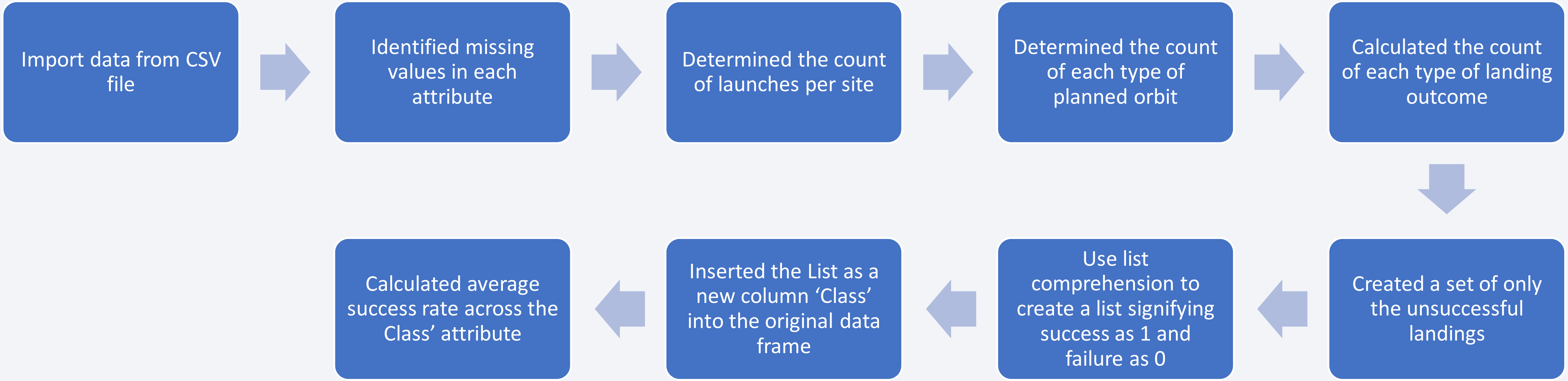
Data Collection - Scraping

- https://github.com/JGundisalvus/IBM_DS_Capstone/blob/main/jupyter-labs-webscraping.ipynb



Data Wrangling

- https://github.com/JGundisalvus/IBM_DS_Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb



EDA with SQL

- https://labs.cognitiveclass.ai/v2/tools/jupyterlab?ulid=ulid-3b2a81d2238ed2815cfd3c00243962e49176a464

- Selected Distinct Landing Outcomes
- Selected Records with Launch Sites Starting with ‘CCA’
- Found total and average mass of Payloads
- Found Boosters that lifted a given range of payload mass
- Found total number of successful and unsuccessful launches

EDA with Data Visualization

- https://github.com/JGundisalvus/IBM_DS_Capstone/blob/main/jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb
- Scatterplots were used to determine if there was a correlation between two attributes and the success of the mission
    - Flight # vs Payload Mass
    - Payload Mass vs Launch Site
    - Orbit vs Payload Mass
- Line charts showed if there is a trend between the year and the success rate
- A bar chart showed which Orbit Types had the highest success rates

Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
- Circle and Marker objects are use to denote the location of each launch site.
- For each launch site additional Markers are added and are colored coded for each landing outcome and grouped in a Marker Cluster. This allows us to see how many failure and successes occur at each site immediately.
- Marker objects displaying distance and Line objects were used to illustrate how close each launch site was to a city, rail line, highway, and the coast
- https://github.com/JGundisalvus/IBM_DS_Capstone/blob/main/lab_jupyter_launch_site_location.jupyterlite.ipynb

Build a Dashboard with Plotly Dash

- A **Pie chart** was added that showed either the Total proportion of successful landings between different sites OR if a specific site was selected, it would display the proportion of success vs non-successful landing
- **A Line chart** along with a **Range Slider** was used to show the breakdown of successful launches against payload sizes for each type of booster. The use is then able to narrow down their expected payload range to a specific site and see which booster would result in successful vs non-success landings

- https://github.com/JGundisalvus/IBM_DS_Capstone/blob/main/spaceX%20Dashboard.py

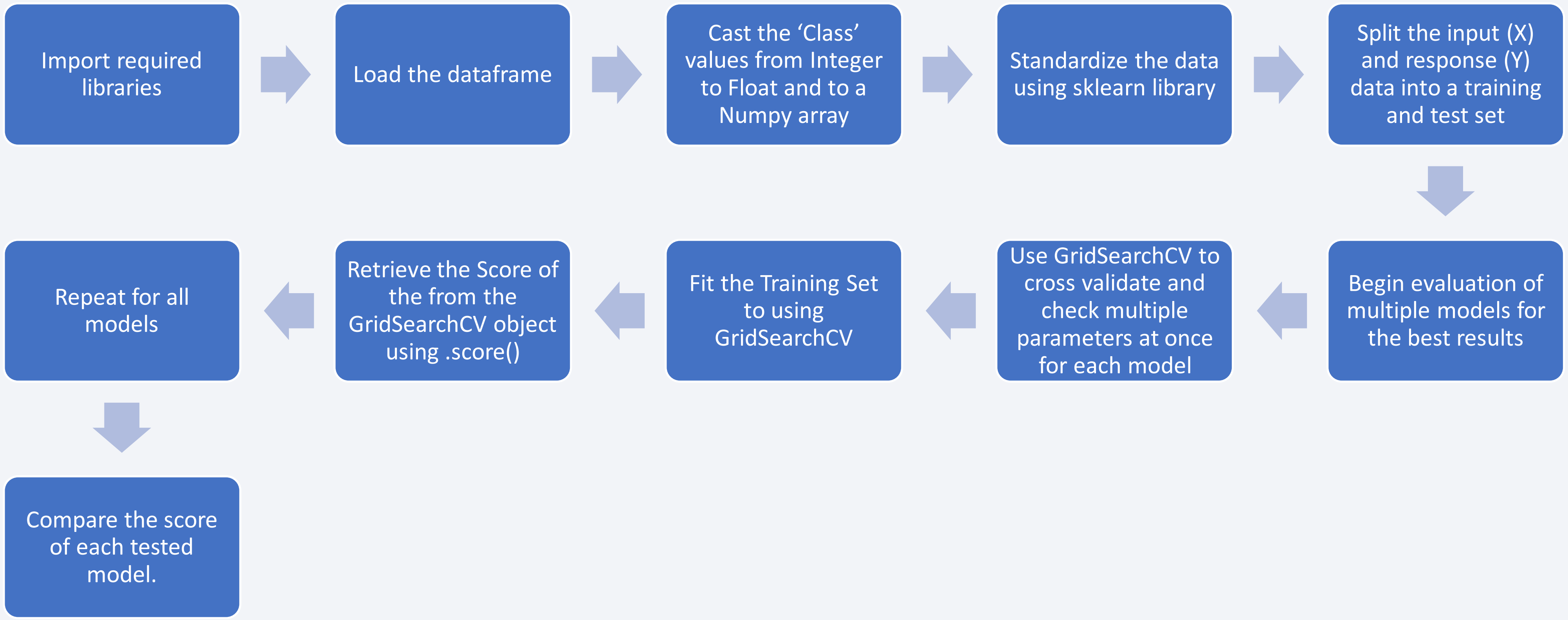
Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model
- The necessary data was imported into a pandas data frame and was standardized before use
- The data was then split into 80% training data and 20% test data
- For each model we used, (Logistic regression, SVM, KNN, Decision Tree) we used a GridSearchCV object to
    - Find the optimal parameters for the model
    - Provided the optimal model
- Each model was then fit and scored against the test data.

- https://github.com/JGundisalvus/IBM_DS_Capstone/blob/main/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

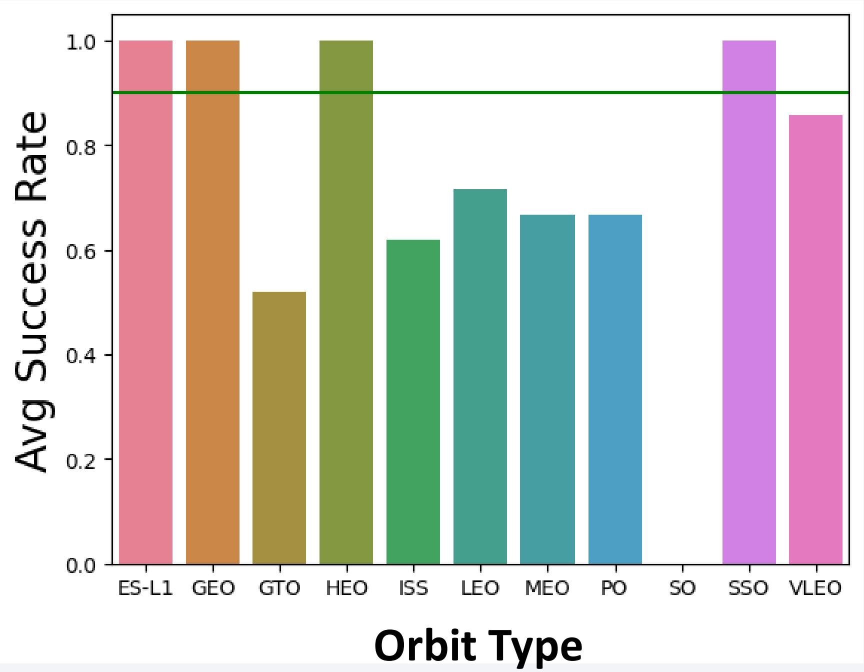
Predictive Analysis (Classification)

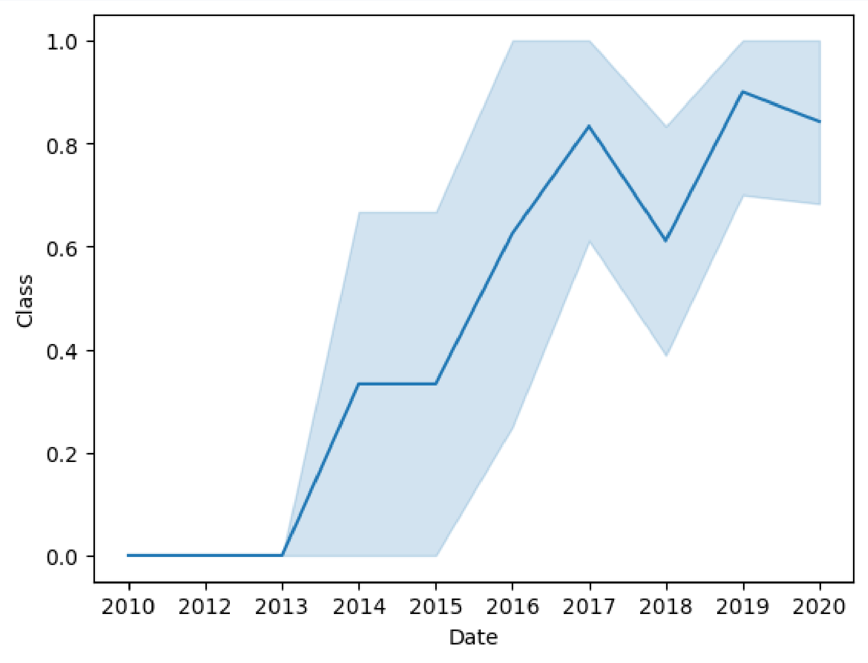
- https://github.com/JGundisalvus/IBM_DS_Capstone/blob/main/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb



Results

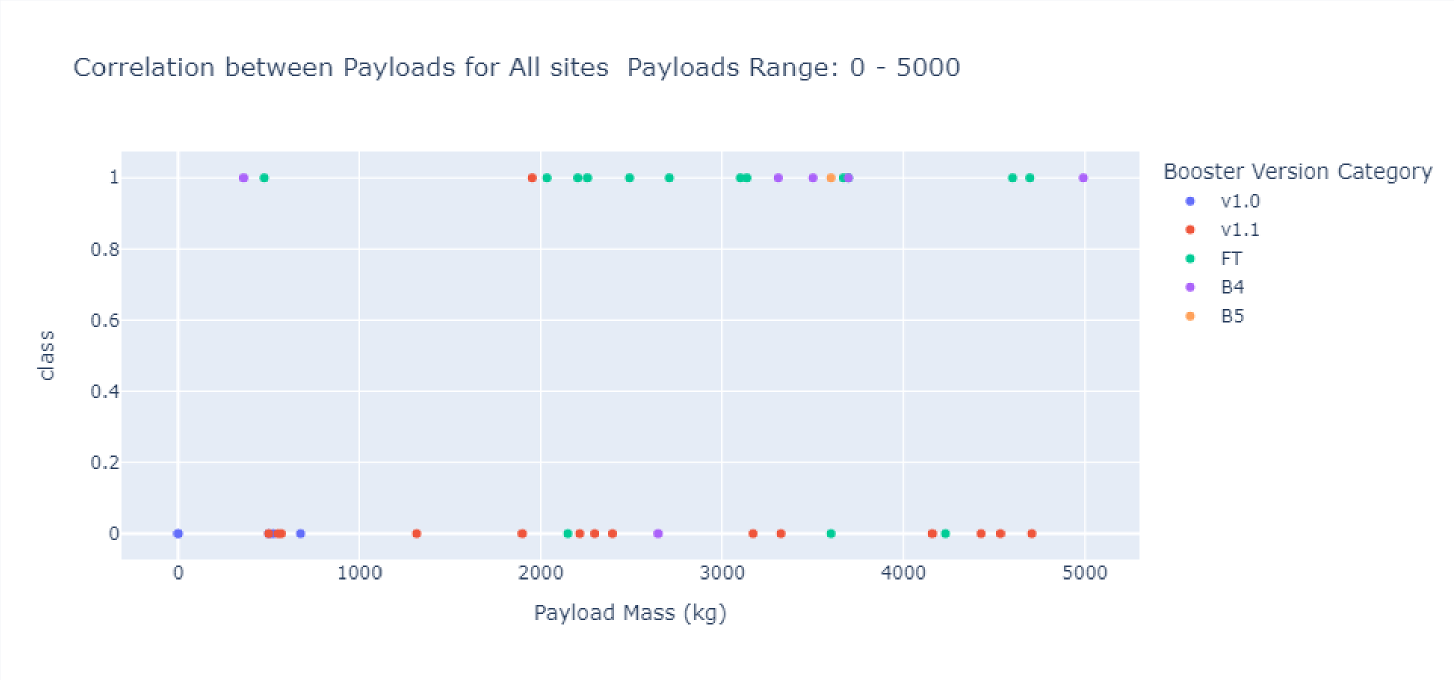
- From our exploration we found that the success rate increases with time, and that the success rate changes with respect to the orbit type of the mission profile.





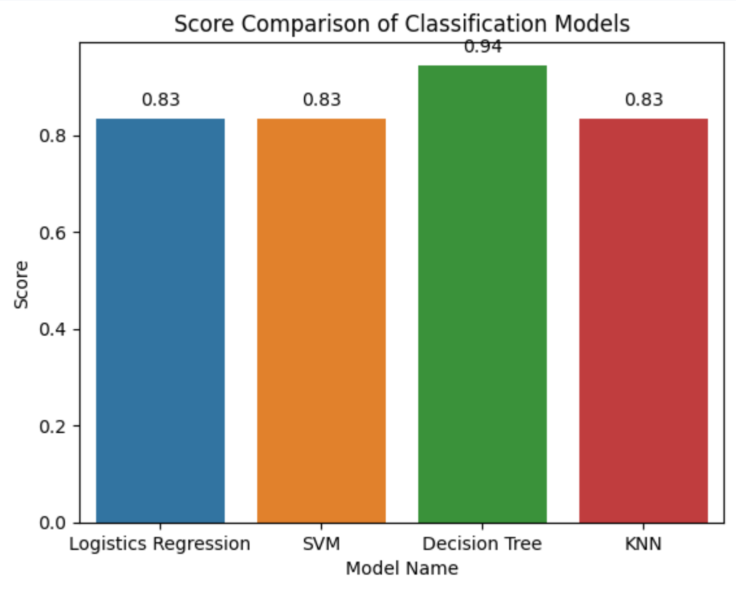
Results

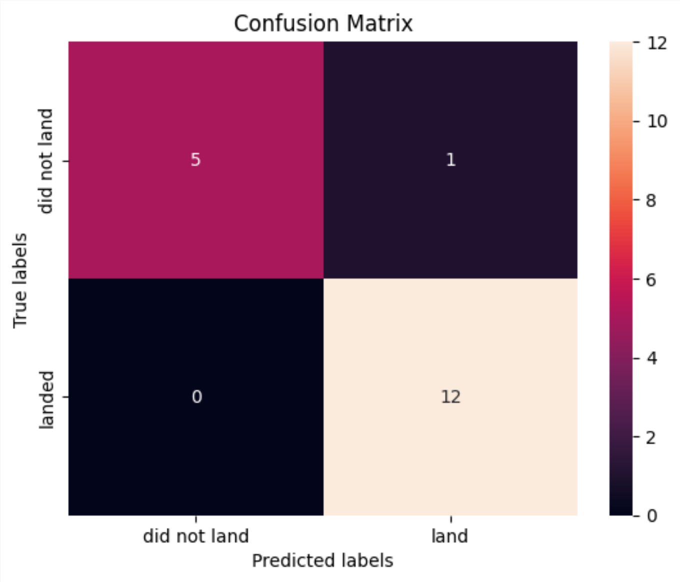
- Using the data captured on SpaceX missions, we can display to our customer the success and non-successful 1st stage landings based on: launch site, payload ranges, and booster configuration



Results

- From our classification model we found that a decision tree classifier could reliably determine if a 1st stage landing will be success ful



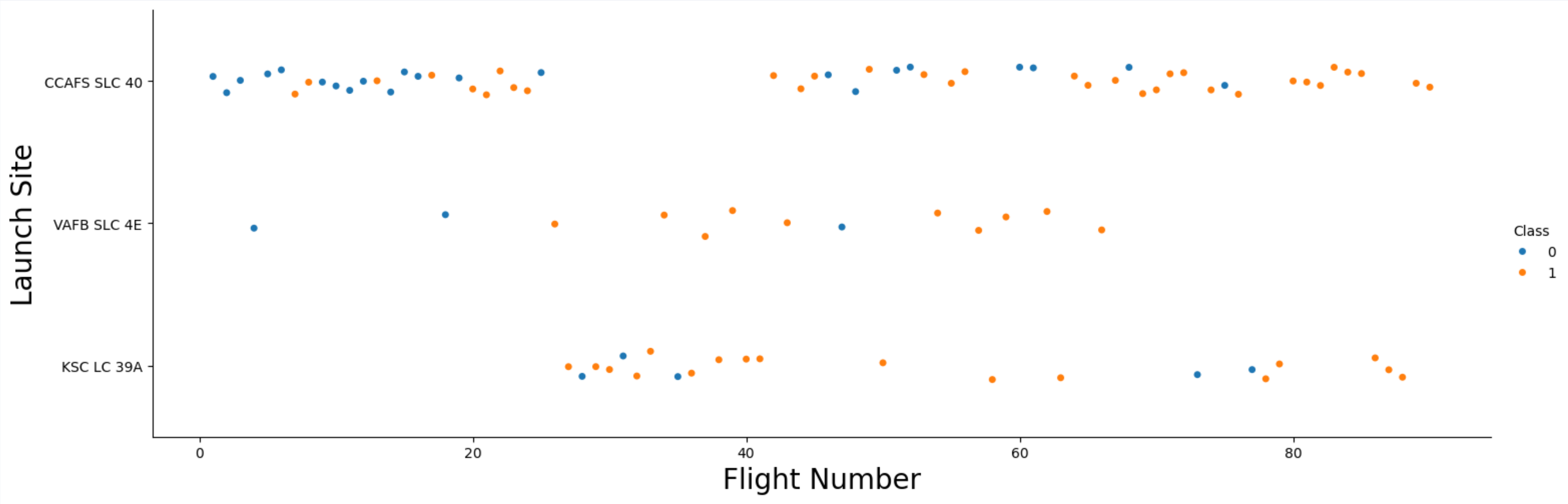


Section 2

Insights drawn from EDA

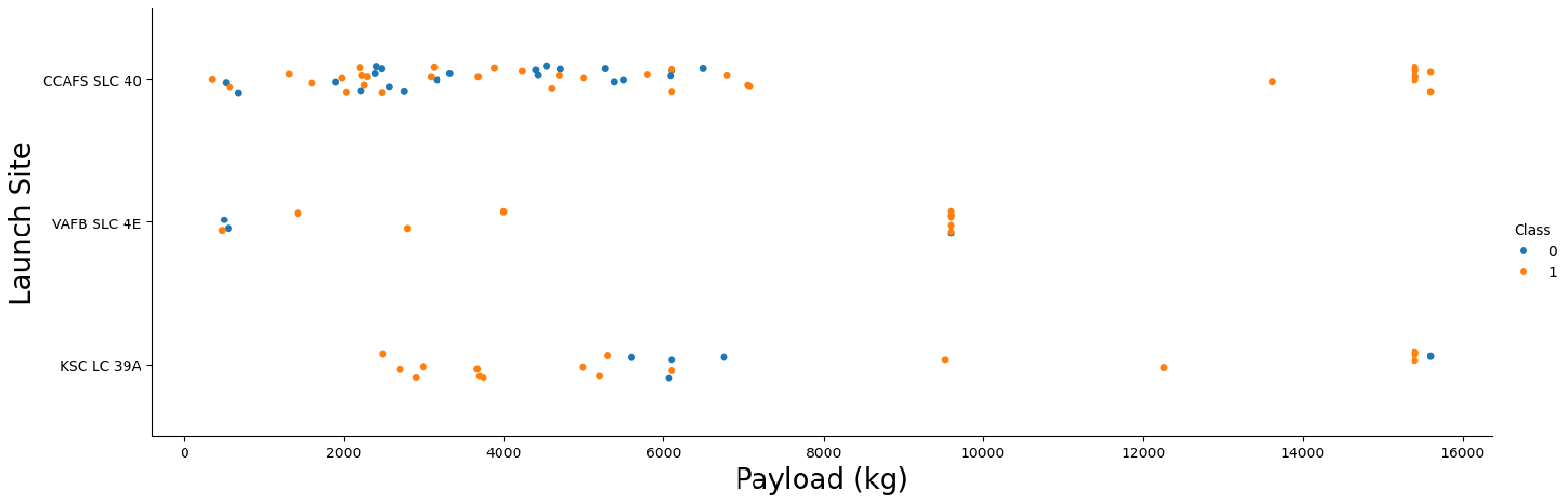
Flight Number vs. Launch Site

- In general, we see more initial failures for CCAFS SLC40 and VAFB SLC 4E, but seemingly improve over time.
- For KSC LC 39A the correlation between Flight number and success is less obvious



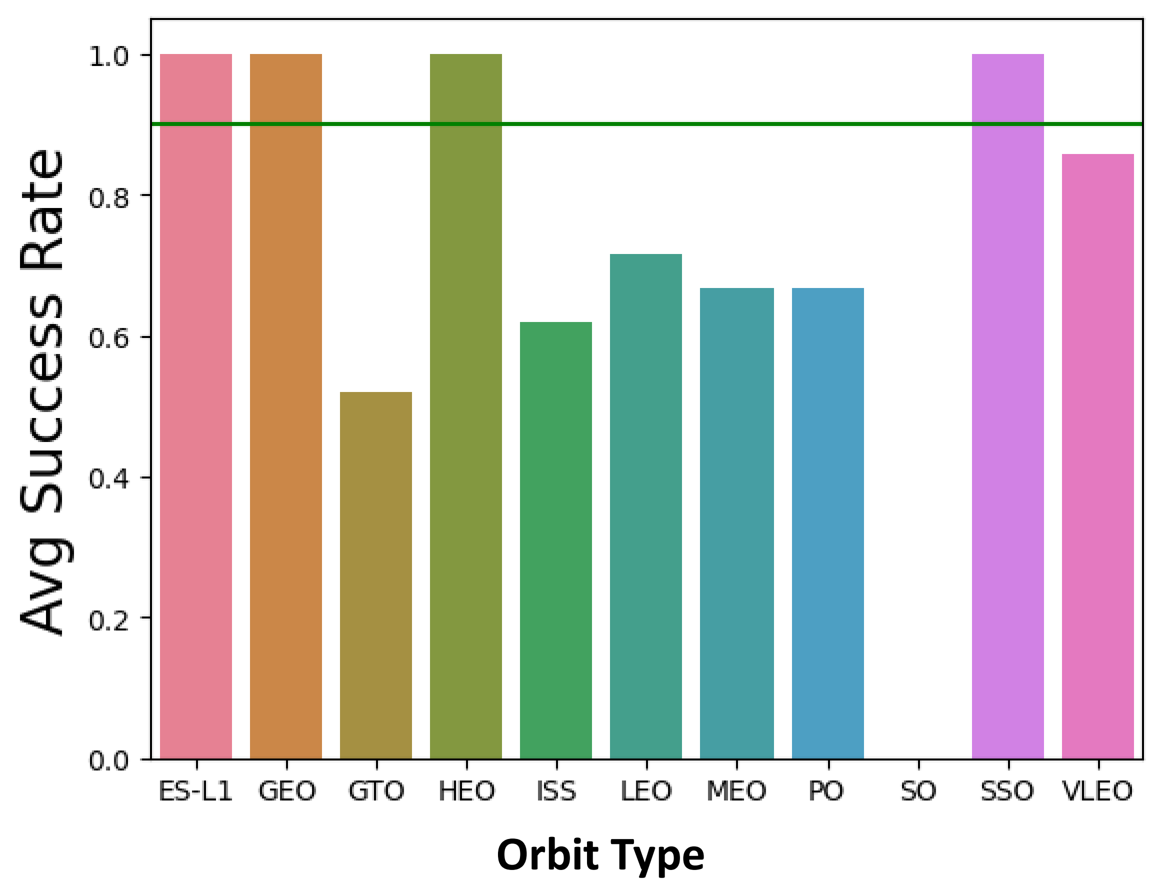
Payload vs. Launch Site

- We do not see a string correlation between payload size and the Launch site, perhaps except for VAFB SLC 4E, which had several successes launching 10,000 kg payloads.
- The maximum payload for VAFB SLC 4E appears to be close to 10,000kg as we have no data past this.



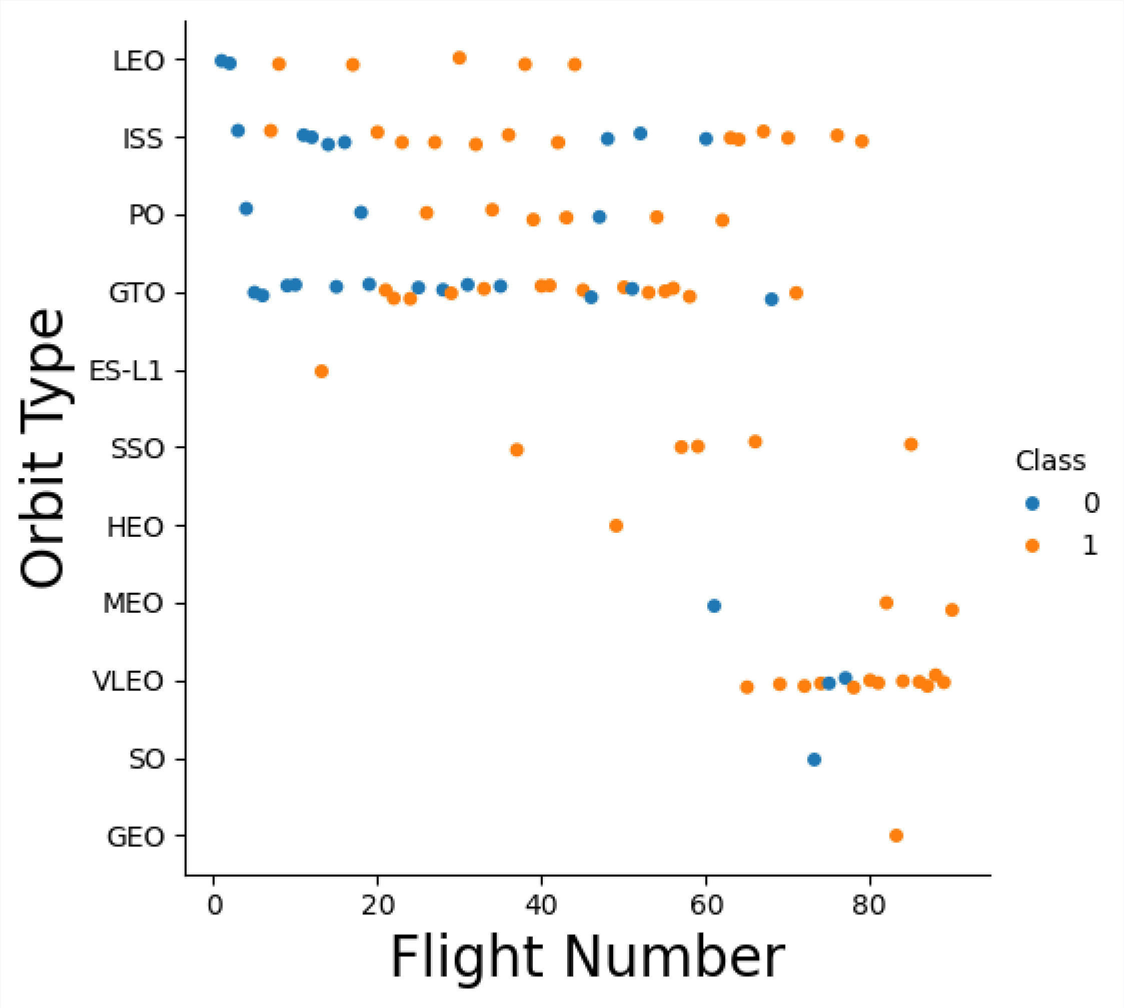
Success Rate vs. Orbit Type

- ES-L1, GEO, HEO, and SSO have very high landing success rates of nearly 100% with VLEO success rate at 85%
- All other Orbit Types have relatively Low success rates
- With SO having either no successful or planned first stage landing



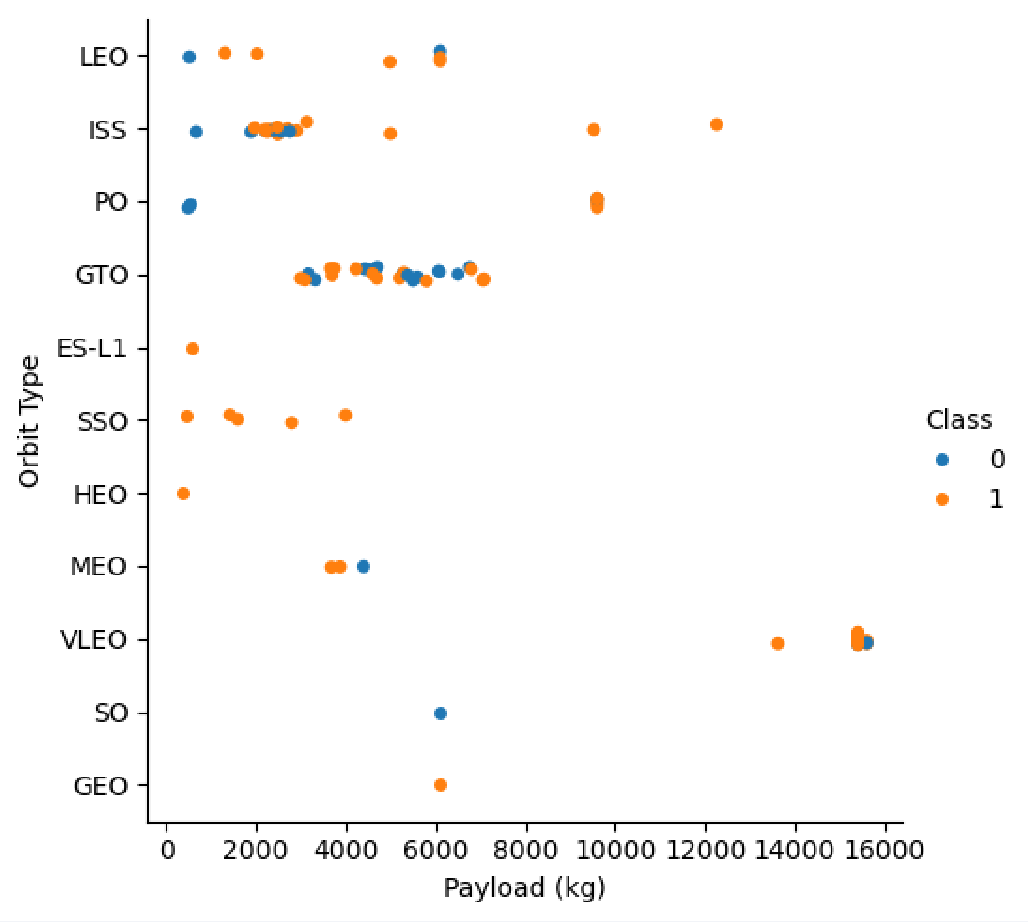
Flight Number vs. Orbit Type

- For LEO, we see correlation between flight number and Orbit Type
- However there seems to be a much lower correlation for any other Orbit Type. Possible indicating that LEO is easier to achieve a fist stage landing for any site.



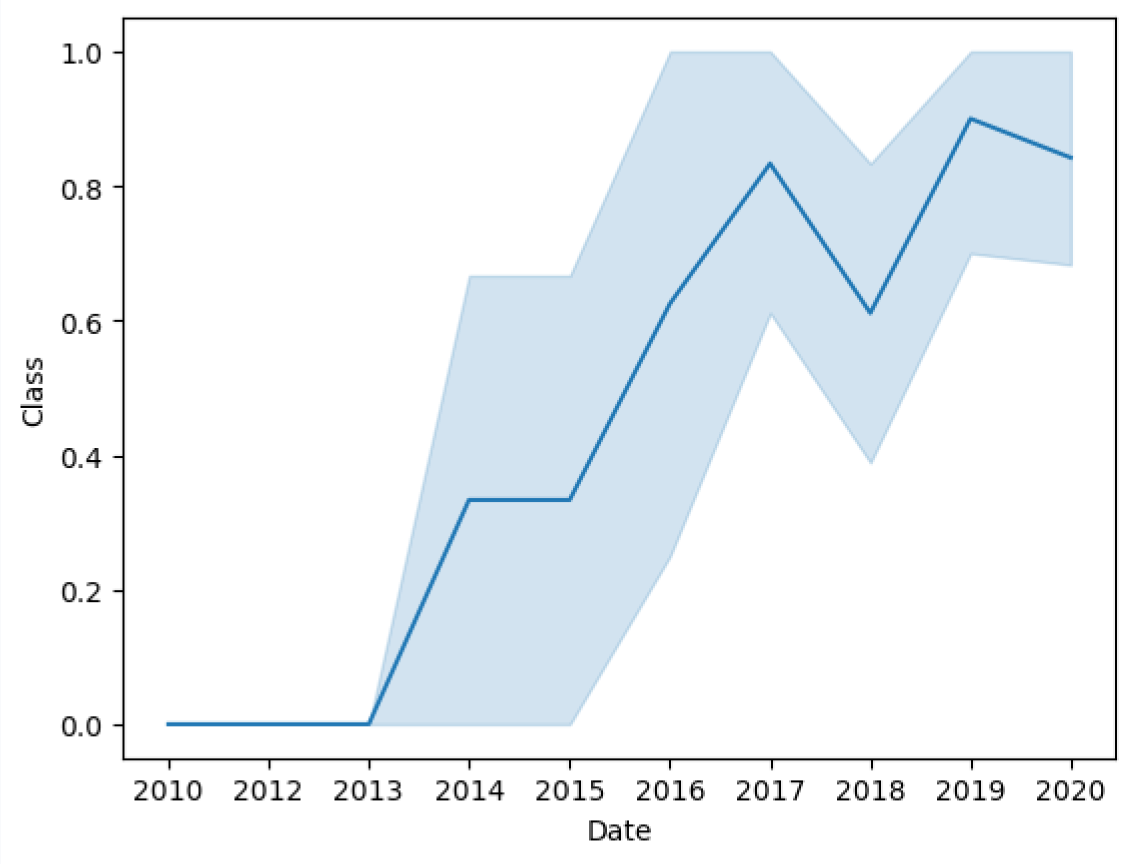
Payload vs. Orbit Type

- There is a noticeable correlation with higher payloads and successful landings for ISS, PO and VLEO Orbit Types
- Success for GTO orbit types have little correlations with Payload mass



Launch Success Yearly Trend

- There is an obvious trend that the success rate increase as time elapsed since 2013 increases



All Launch Site Names

- Selecting Distinct values from the 'Launch_Site' column resulted in:
  - CCAFS LC-40
  - VAFB SLC-4E
  - KSC LC-39A
  - CCAFS SLC-40

Task 1

Display the names of the unique launch sites in the space mission

```
%sql select Distinct Launch_Site from SPACEXTABLE
```

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

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

Launch Site Names Begin with 'CCA'

- %sql SELECT * FROM SPACEXTABLE WHERE launch_site LIKE 'CCA%' LIMIT 5;
- Selected all rows from the SPACEXTable where the launch_site started 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	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

- Applied the Sum function on the PAYLOAD_MAS__KG_ column and filtered for NASA (CRS) as the customer. Total payload mass carried was 45596 kg

Task 3

Display the total payload mass carried by boosters launched by NASA (CRS)

```
%%sql select sum(PAYLOAD_MASS__KG_) as 'NASA CRS Total Payload (kg)'
from SPACEXTABLE
where customer = 'NASA (CRS)'
```

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

NASA CRS Total Payload (kg)
45596

Average Payload Mass by F9 v1.1

- Applied the avg function on the PAY_LOAD_MASS__KG_ columns and filtered for the F9 v1.1 booster version.
- Total payload mass carried was 2928.4 kg

Task 4

Display average payload mass carried by booster version F9 v1.1

```
%%sql Select avg(PAYLOAD_MASS__KG_) as 'Avg Mass'
from SPACEXTABLE
where Booster_version = 'F9 v1.1'
```

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

Avg Mass
2928.4

First Successful Ground Landing Date

- The first successful Launch date was 12-22-2015
- Queried the SPACEXTABLE for the minimum of the ‘Date’ column where the mission outcome was a ‘Success’ and the landing outcome was on a ground pad

```
%%sql Select min(Date) from SPACEXTABLE
where ( mission_outcome = 'Success'
      AND landing_outcome like'%ground pad%')
```

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

min(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Booster Versions
  - F9 FT B1022
  - F9 FT B1026
  - F9 FT B1021.2
  - F9 FT B1031.2

- These 4 booster versions carried payloads between 4000 and 6000 kg and successfully land on a drone ship.

```
%%sql select distinct BOOSTER_VERSION from SPACEXTABLE
where ( PAYLOAD_MASS__KG_ between 4000 and 6000
       AND LANDING_OUTCOME like '%Success (drone ship)%')
```

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

[No Title]

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

Total Number of Successful and Failure Mission Outcomes

- The breakdown of mission outcomes
  - Failure 1
  - Success 100
- Using the query shown right, we determined that there was only 1 mission failure and 100 mission successes

```
%%sql select
    CASE
        WHEN Mission_Outcome LIKE '%success%' Then'Success'
        ELSE 'Failure'
    END AS Outcomes,
    Count(*) as count
from SPACEXTABLE
group by Outcomes
```

Boosters Carried Maximum Payload

- These Booster Versions carried the maximum payload mass
- By using a sub query find the max payload mass, it was then used the filter the payload masses carried by the boosters

```
%%sql Select booster_version
from SPACEXTABLE
where PAYLOAD_MASS__KG_ =
    (select max(PAYLOAD_MASS__KG_)from SPACEXTABLE)
```

- 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

- In 2015 there were 2 failures for Drone ship landings one in January and one in April.

Month	Year	Booster_Version	Landing_Outcome	Mission_Outcome
01	2015	F9 v1.1 B1012	Failure (drone ship)	Success
04	2015	F9 v1.1 B1015	Failure (drone ship)	Success

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

- No Landing attempt was the largest type of landing outcomes, followed by drone ship landings.
- Ground pad landings showed a 100% success between these dates.

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

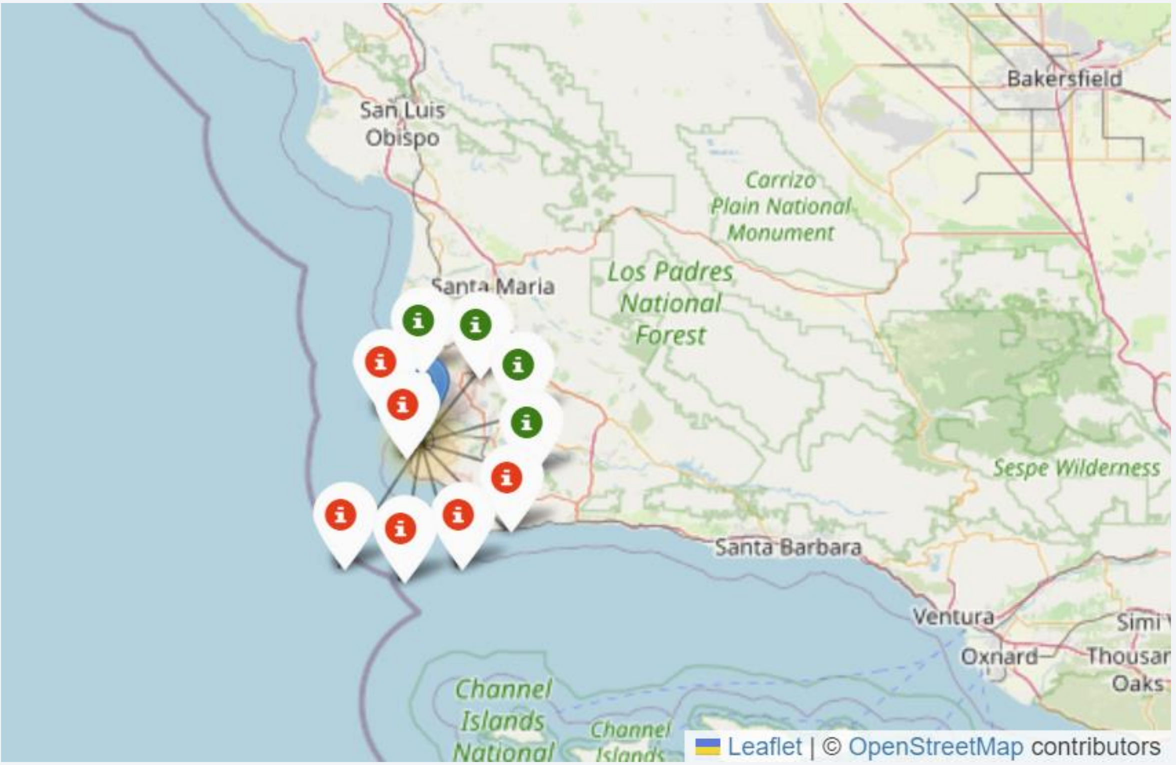
Launch Sites Proximities Analysis

Folium – World Map of Launch Sites

- From the map of the 4 launch sites, we found these 2 important factors about their locations
  - They are close to a coastline, allowing for water landings or drone ship landings
  - They are closer to the equator, as opposed to being position far north in the united states.

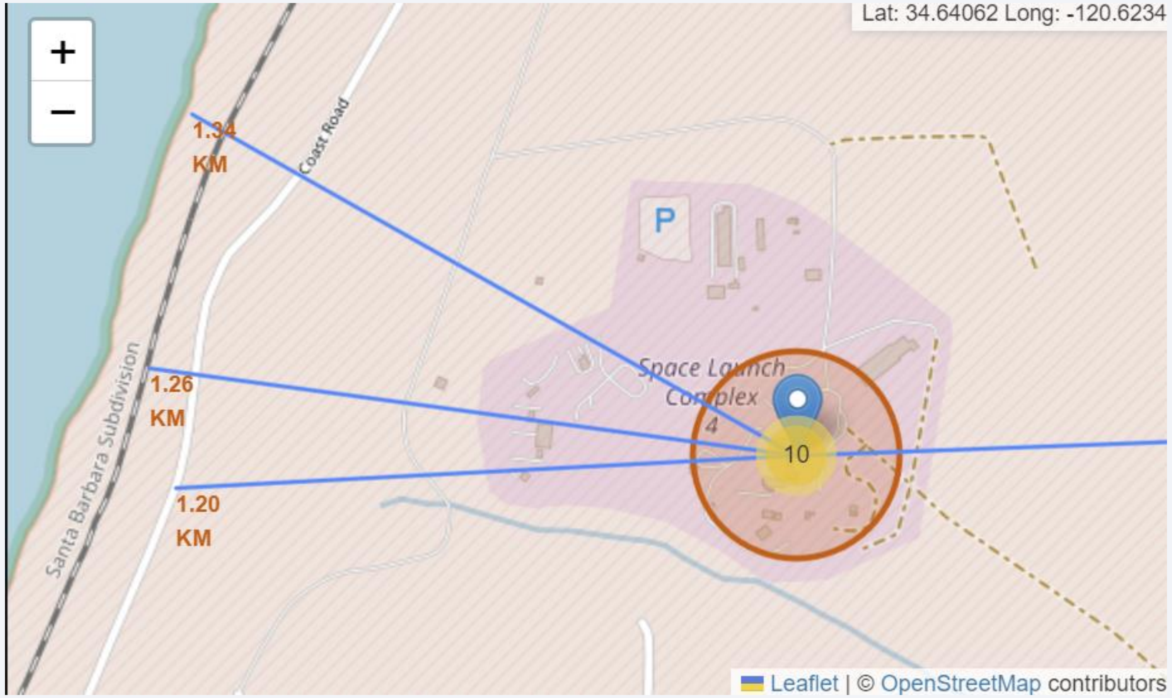
Folium – Local Map and Outcome Plotting

- From the map located right we can quickly visualize the individual mission success
- Each white marker signifies either a failure in red, or a success in green.



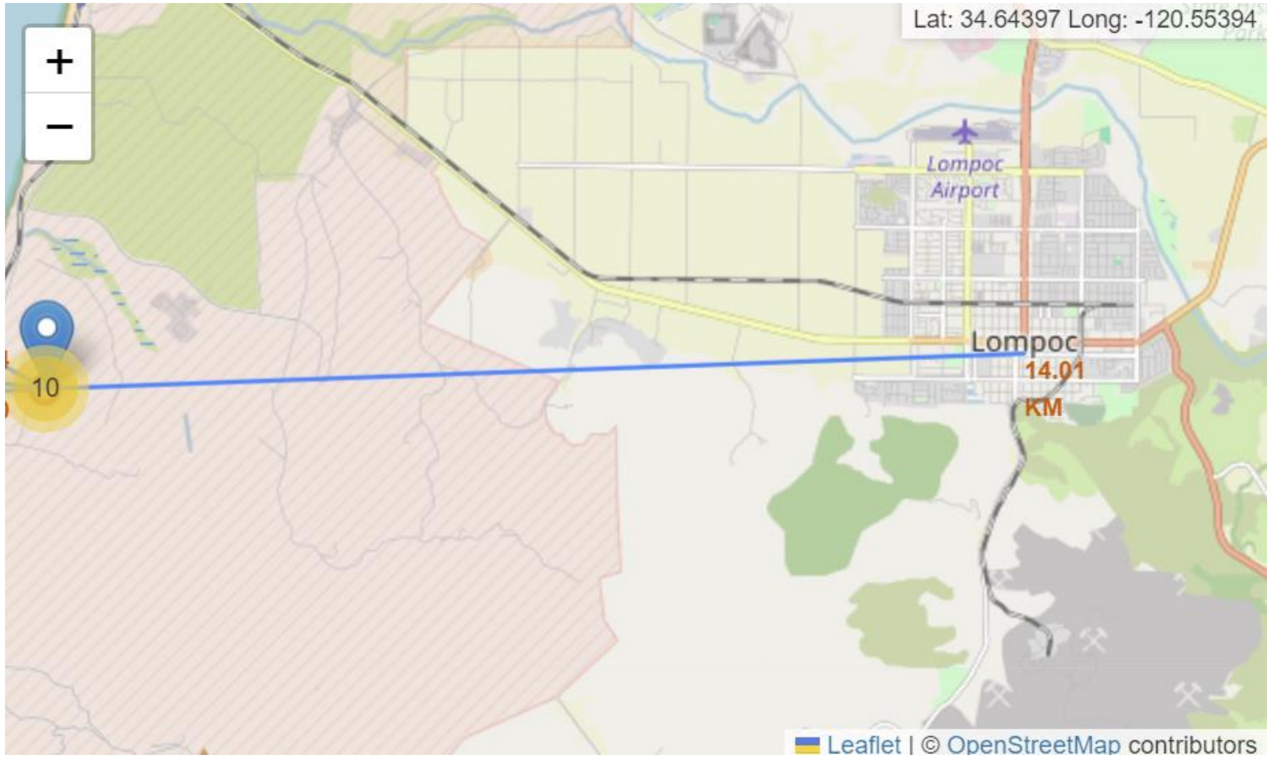
Folium map – Proximity to Infrastructure, Coastline and Cities

- Majority of the Launch sites are in proximity (roughly <2km) to a major body of water, a rail line, or highway



Folium map – Proximity to Infrastructure, Coastline and Cities

- Launch sites are situated farther from populated city centers
- Evident since VAFB Launch site is roughly 14 km from the nearest town.

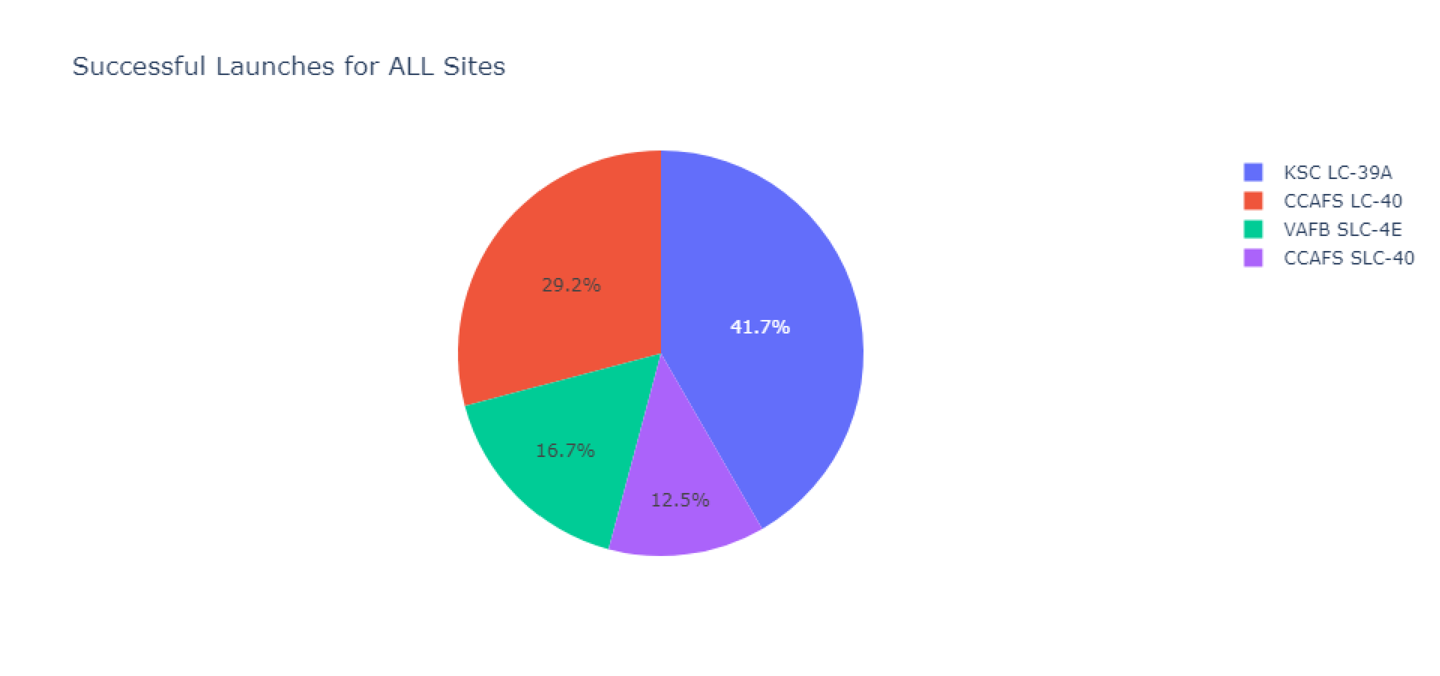


Section 4

Build a Dashboard with Plotly Dash

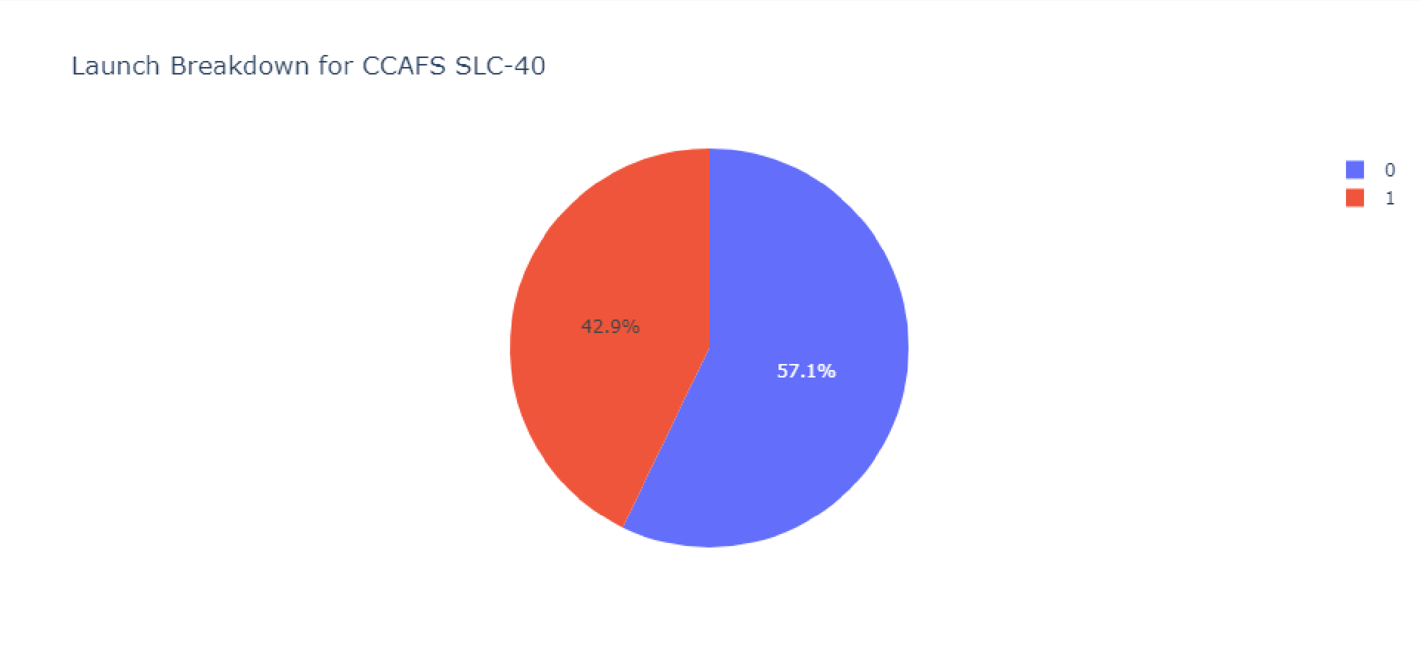
Plotly Dashboard – Pie Chart

- Launch Sites KSC LC-39A and CCAFS LC-40 Have the majority of success launches
- With VAFB SLC-4E and CCADS-SLC-40 have a minority of the successful launches



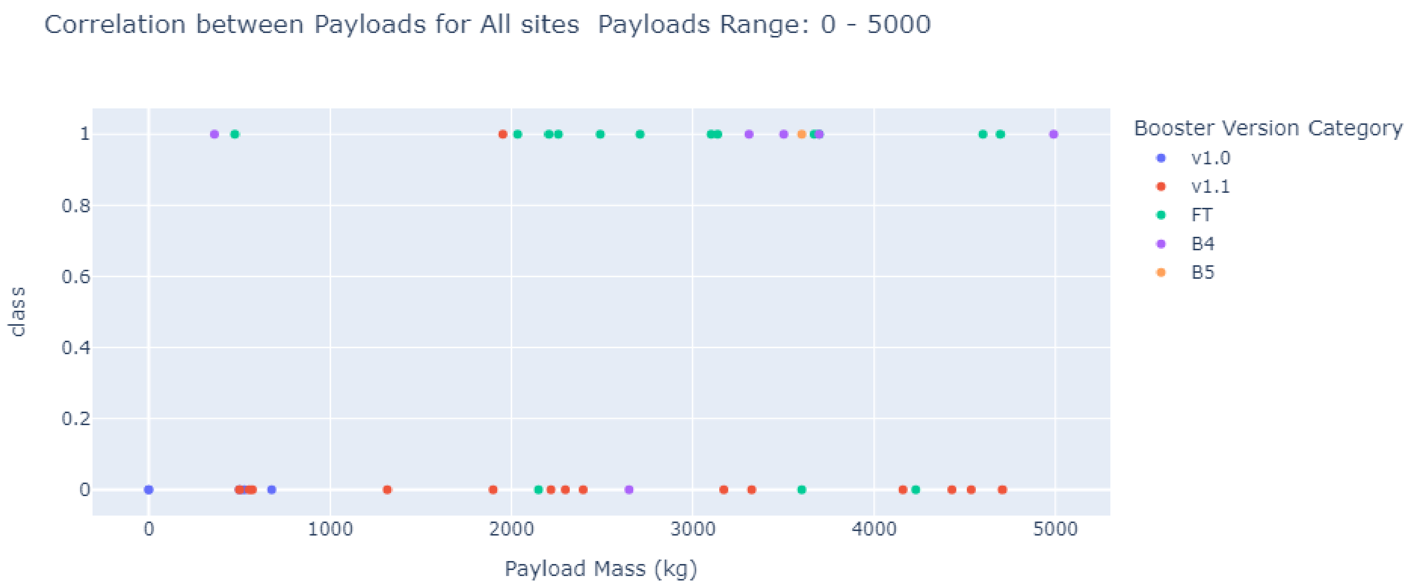
Plotly Dashboard – Site Specific Pie Chart

- CCAFS SLC-40 has the greatest proportion of success full launches of any site at roughly 42%



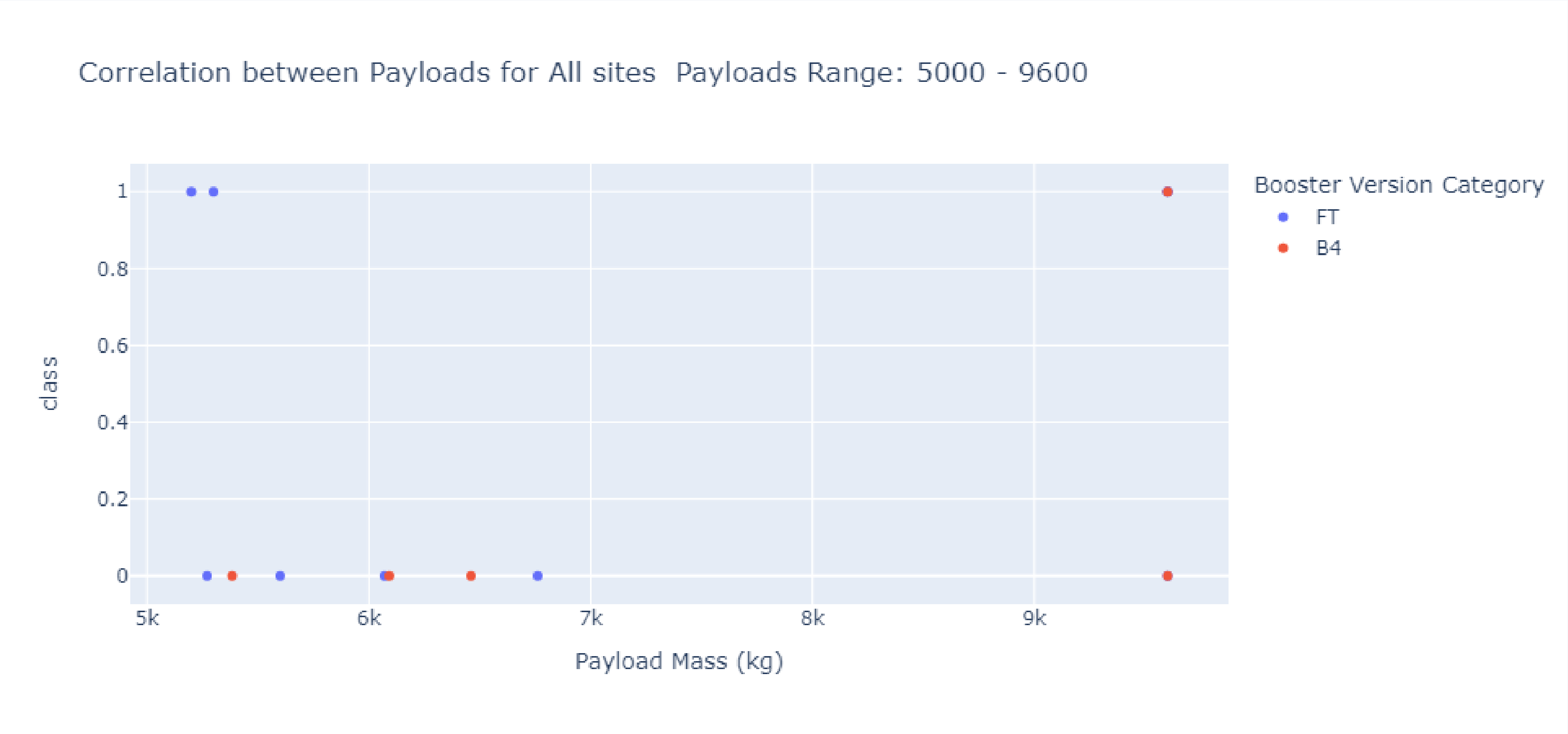
Plotly Dashboard – Payload Range 0-5000kg

- Noticeably in the 0-5000 kg range, FT booster has a large proportion of successes
- V1.1 Booster on the other hand shows an increased number of unsuccessful landings



Plotly Dashboard – Payload Range 5000k -9600g

- For larger payload masses, there are fewer success overall

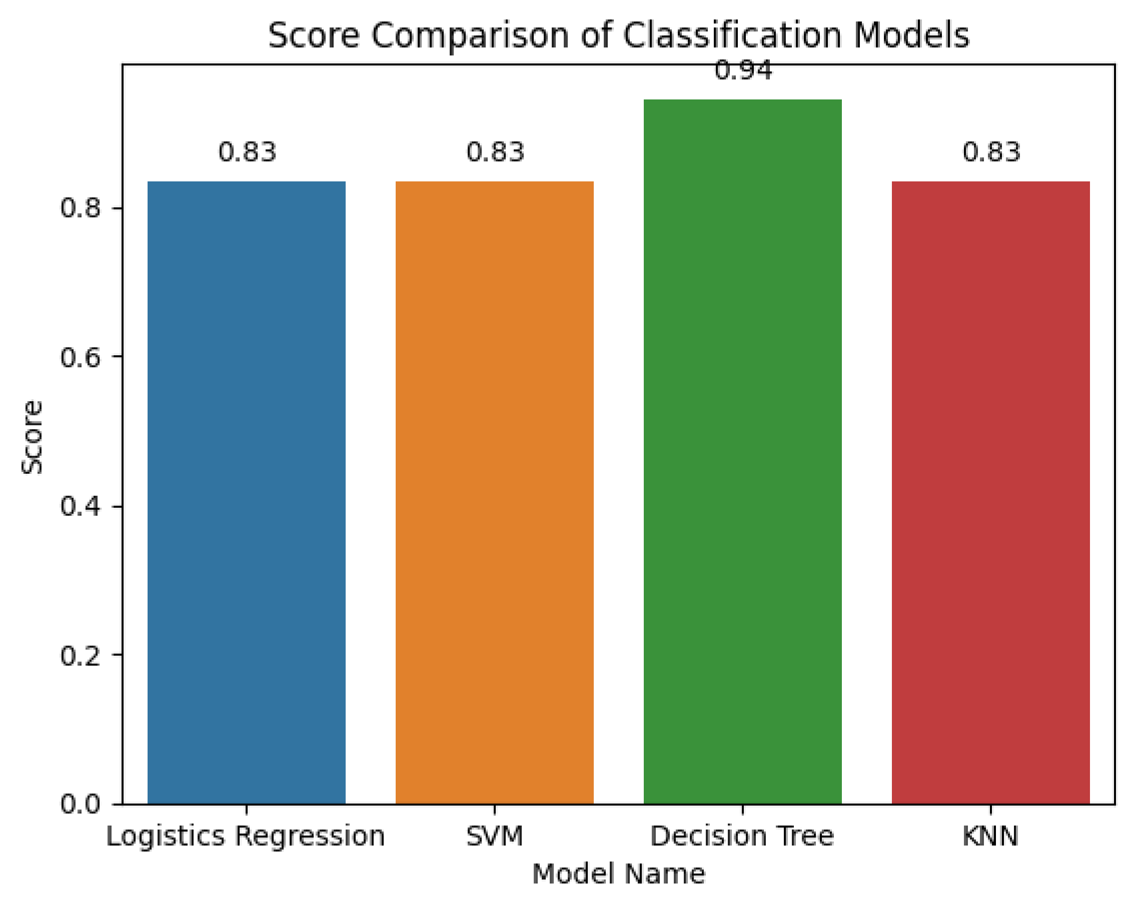


Section 5

Predictive Analysis (Classification)

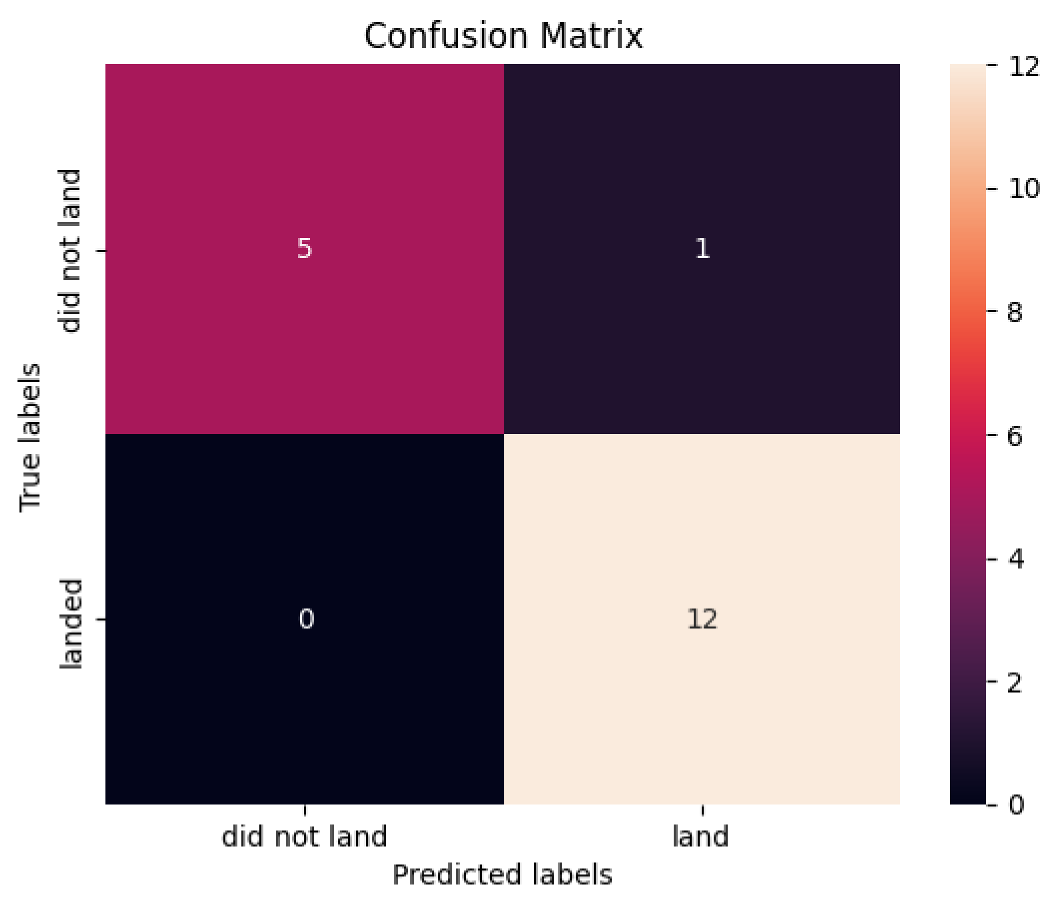
Classification Accuracy

- The Decision Tree classifier had the highest score at 94%
- All other classifiers performed equally at 83% accuracy



Confusion Matrix

- The confusion matrix for the Decision Tree shows that the model is accurate, except for one false negative show in the top right square as a value of 1.
- No false positives exists for this model, evident by the 0 value in the bottom left square.



Conclusions

- Through data exploration we determined that there are strong correlation between attributes and the success of the launches
- An interactive dashboard will allow our customer to gauge if a mission will have a successful 1st stage landing for different payloads and sites
- Using a Decisions Tree classification model, we can with 94% accuracy determine if the 1st stage will land

Thank you!