IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Nicolas Zemel>
<12/31/2023>

Outline

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

Executive Summary

- Summary of methodologies
  - API Data Collection
  - Web Scraping Data Collection
  - Data Wrangling
  - SQL EDA
  - Data Visualization EDA
  - Folium Interactive Visual Analytics
  - Machine Learning Analytics
- Summary of all results
  - EDA Results
  - Folium Screenshots
  - Machine Learning Results

Introduction

- Project background and context
  - Space X has one of the lowest advertised prices for rocket launches at $62M which is over $100M less than other rocket launches. It's believed that the savings are due to the reuse of the first stage of the rocket. By analyzing rocket launches we would like to predict the likelihood of a successful first stage landing.
- Problems you want to find answers
  - What factors affect a rocket landing successfully.
  - How various features may affect a specific landing.

Section 1

Methodology

Methodology

Executive Summary

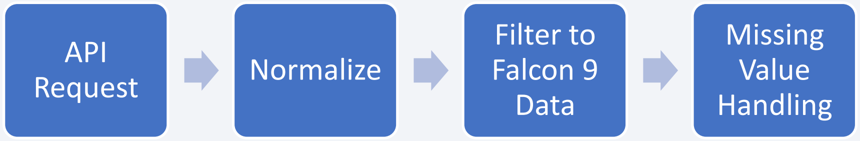
- Data collection methodology
- Perform data wrangling
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

Data Collection – SpaceX API

API data was obtained directly from api.spacexdata.com using requests.get()

- The API json data was normalized using pd.json_normalize, filtered to only include Falcon 9 launches, then the missing PayLoad masses were assumed to be the mean of the Falcon 9 rocket data.

- https://github.com/nzmel/MLResources/blob/master/module%201%20lab%201.ipynb
- https://github.com/nzmel/MLResources/blob/master/dataset_part_1.csv



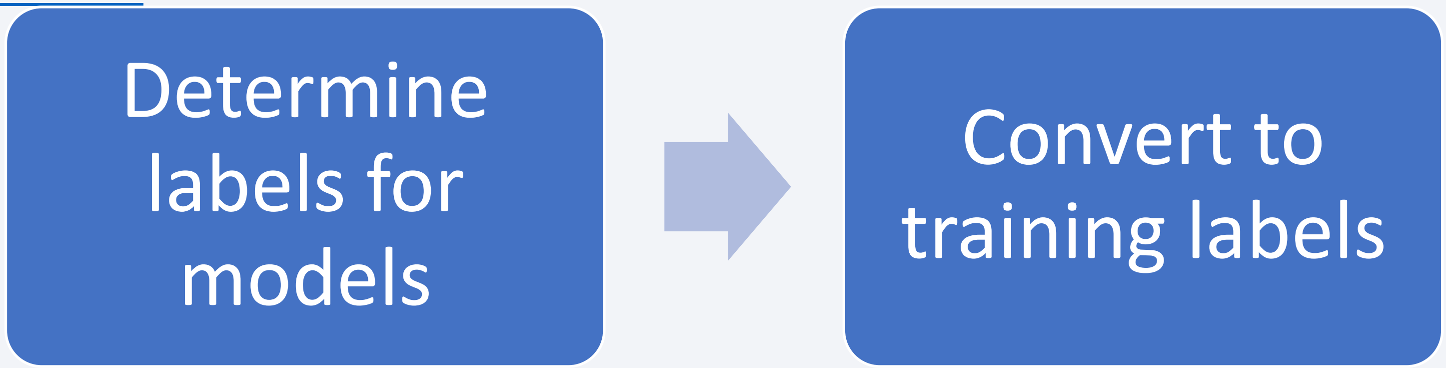
Data Collection - Scraping

Webscraped data was obtained from Wikipedia (https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches) using BeautifulSoup on a static webpage of Jun 9th 2021.

- https://github.com/nzmel/MLResources/blob/master/module%201%20lab%202.ipynb
- https://github.com/nzmel/MLResources/blob/master/spacex_web_scraped_part2.csv

Data Wrangling

- Missing values were counted
- dtypes were checked for correctness
- Counts of occurrences of each orbit were determined to make sure we had enough values for an ML model
- Analyzed mission outcomes
- https://github.com/nzmel/MLResources/blob/master/labs-jupyter-spacex-Data%20wrangling_module1_l3.ipynb
- https://github.com/nzmel/MLResources/blob/master/dataset_part_2_module1_l3.csv



EDA with Data Visualization

- A Category plot of Flight Number vs Payload mass was used to see the correlation between success and flight number.
- A Category plot of Flight Number vs Launch Site to see which launch sites are successful overall and concerning flight numbers
- A Category plot of Payload Mass vs Launch Site to see the success rate at different launch sites relative to payload.
- A Bar plot of Orbit vs the success rate to determine which Orbits are the most successful.
- A scatterplot of Flight number vs Orbit to see the success of orbits over time.
- A scatterplot of Payload mass vs orbit to determine which orbits have the most success at different payload amounts.
- Line plot of Date vs Class to determine the success rate over time.

- https://github.com/nzmel/MLResources/blob/master/jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

EDA with SQL

- Looked at Unique Launch sites
- Looked at Unique Launch sites that begin with CCA
- Looked at payload masses specific to NASA CRS
- Looked at average payload mass carried by booster version F9 v1.1
- Looked at the date of the first successful landing of a ground pad.
- Looked at the boosters that could carry between 4000 and 6000 kg payload masses onto a drone ship while landing successfully.
- Looked at the overall mission outcomes.
- Looked at the boosters that have carried the maximum payload mass.
- Looked at the failures in 2015 for drone ship landings.
- Looked at the count of each landing outcome.
- https://github.com/nzmel/MLResources/blob/master/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- We added circles, lines, markers, and marker clusters to Folium Maps to indicicate specific coordinates, indicate distances, indicate launch points, and group events, respectively.
- https://github.com/nzmel/MLResources/blob/master/lab_jupyter_launch_site_location.jupyterlite.ipynb

Build a Dashboard with Plotly Dash

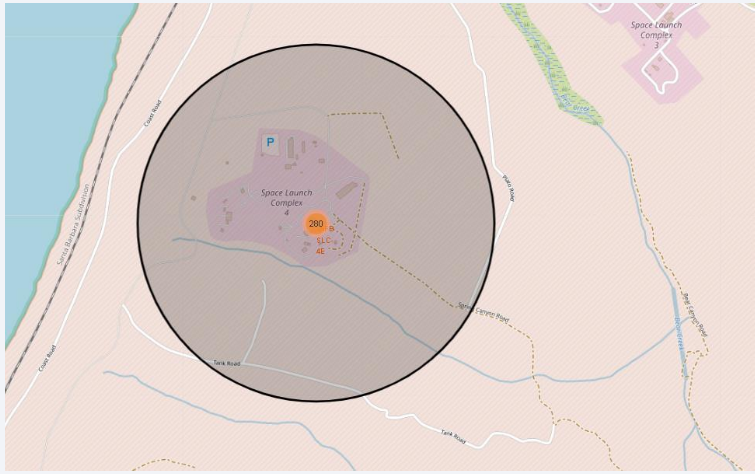
- Pie charts were plotted to show the launch count at different sites and overall.
- Scatter plots were used to show the relationship between Payload mass and Outcome for different rocket boosters.
- https://github.com/nzmel/MLResources/blob/master/spacex_dash_app.py

Predictive Analysis (Classification)

- The data was standardized so that it could be used in different models.
- Train/Test splitting was then done to ensure overfitting/underfitting doesn't occur.
- Grid Search cross-validation was done to ensure that the best hyperparameters were chosen and the best model was chosen.
- Models were evaluated and chosen using the accuracy of the test set and a confusion matrix to analyze where the accuracy was lower.
- https://github.com/nzmel/MLResources/blob/master/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

Results

- EDA showed us:
  - The success rate has been increasing dramatically over time, starting in 2014 and peaking in 2019.
  - Payloads are most successful with orbits LEO, ISS, and SSO, however only the ISS orbit has a solid success rate with heavy payloads (>8000 kg)
  - Most payloads are less than 10000 kg.
  - SSO, GEO, HEO, and ES-L1 all have perfect success rates; however, they are low samples, except for SSO.
  - There are 3 launch sites, and 2 of them do all the heavy loads.
- Decision trees were found the be the most effective at differentiating between successful landings and failures. The data performed as well on the training and test set.

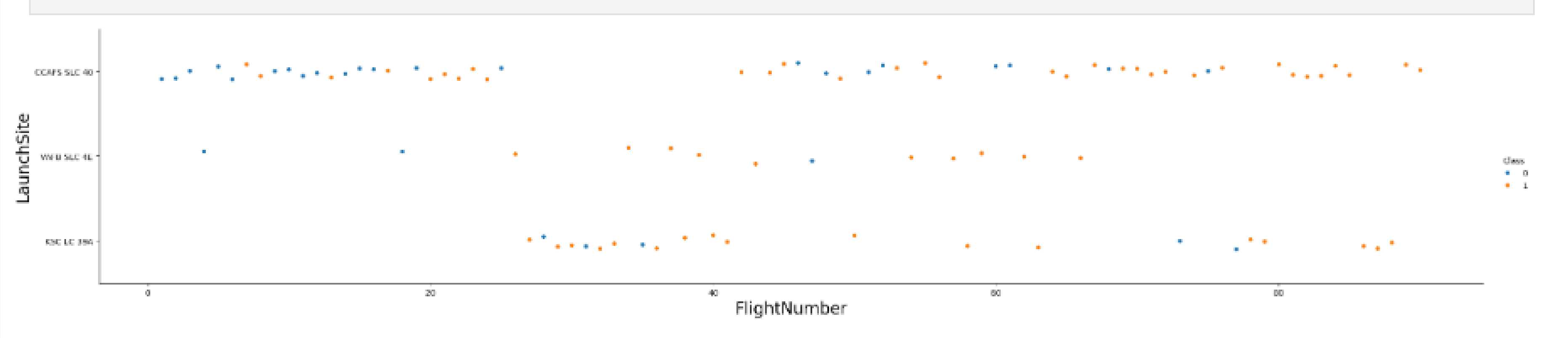


Section 2

Insights drawn from EDA

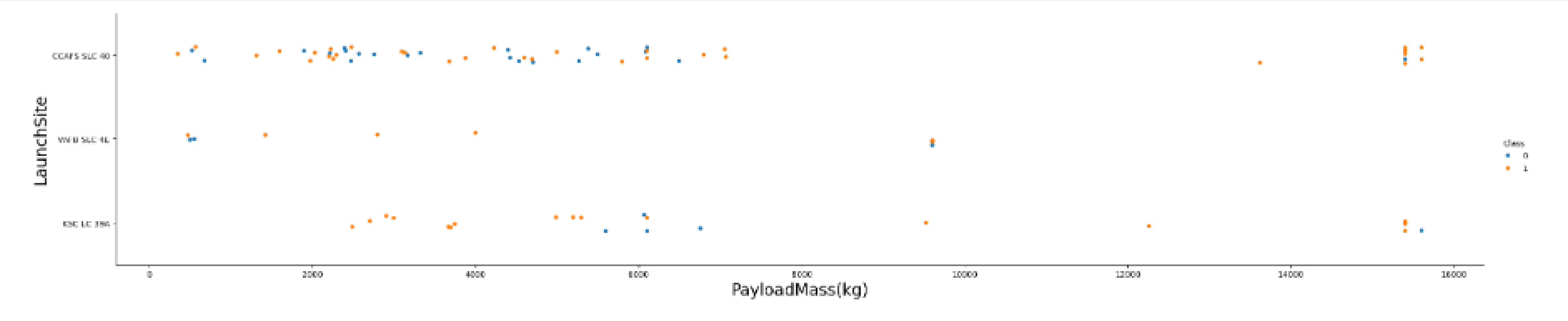
Flight Number vs. Launch Site

- 3 launch sites are seen, and as the flight number increases at a launch site, the more likely a successful outcome will occur.



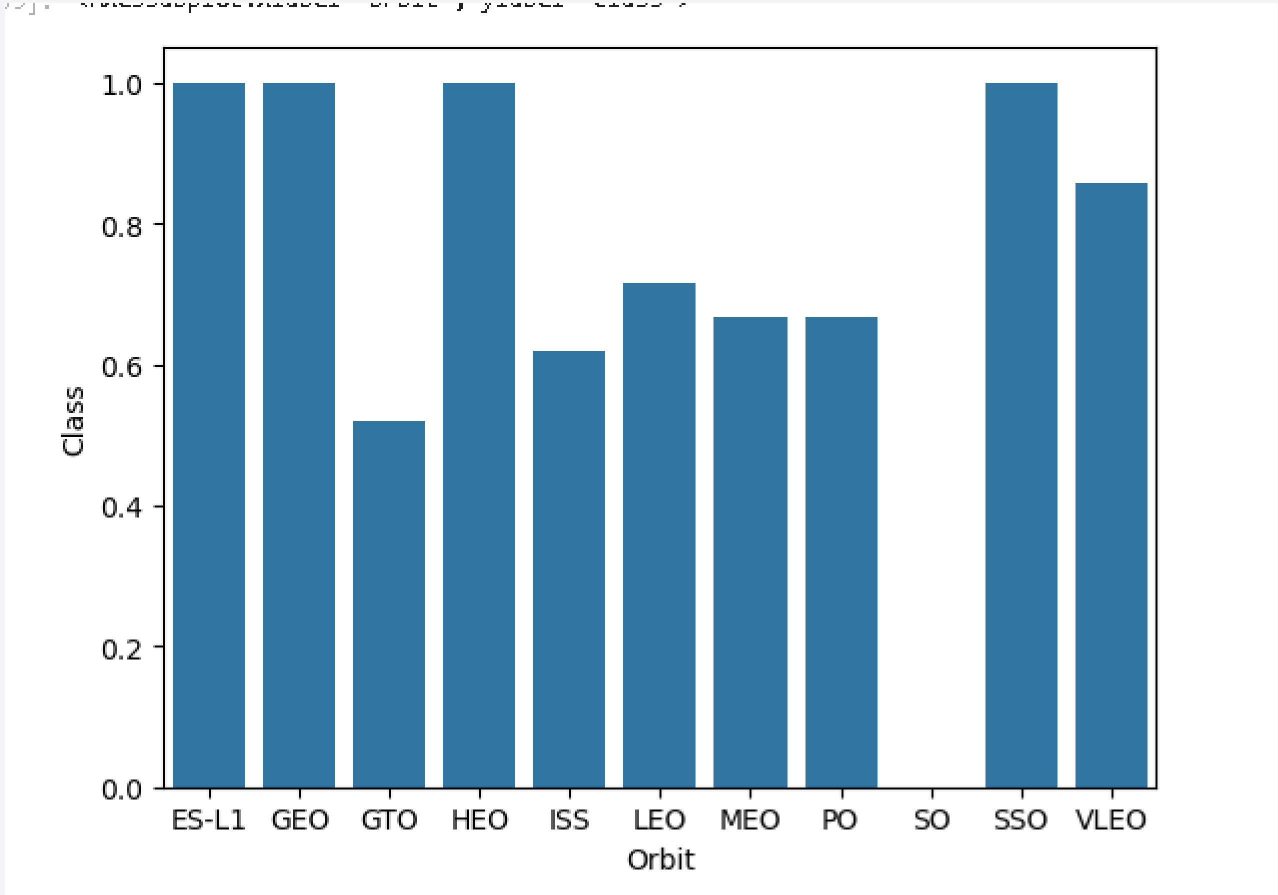
Payload vs. Launch Site

Only 2 launch sites handle any variety of payload. The last launch site handles mid-range, at most.



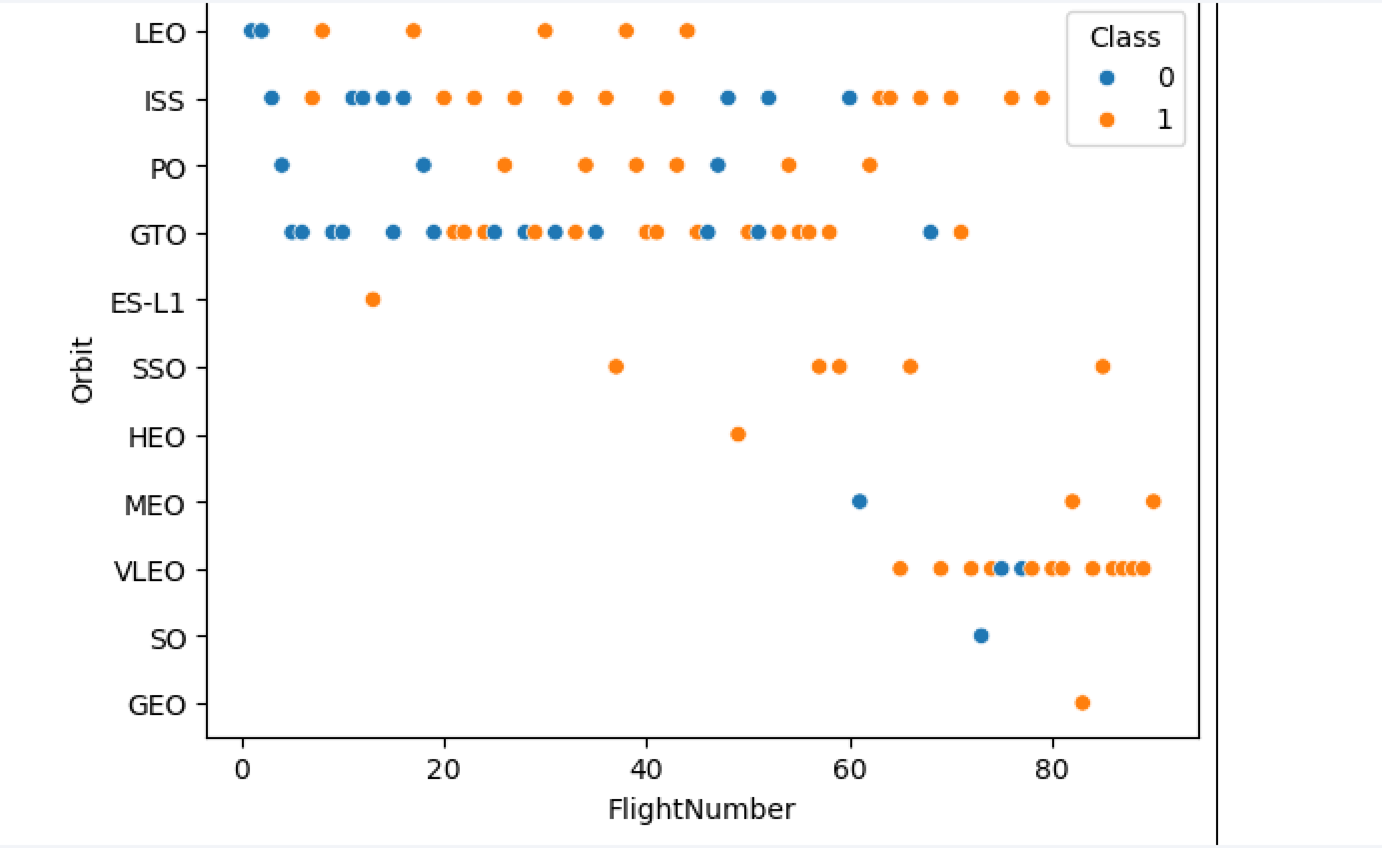
Success Rate vs. Orbit Type

- The success rate of all of the perfect success rates and 0 success rates are based on a low number of samples and are not indicative of what the true success rate is.



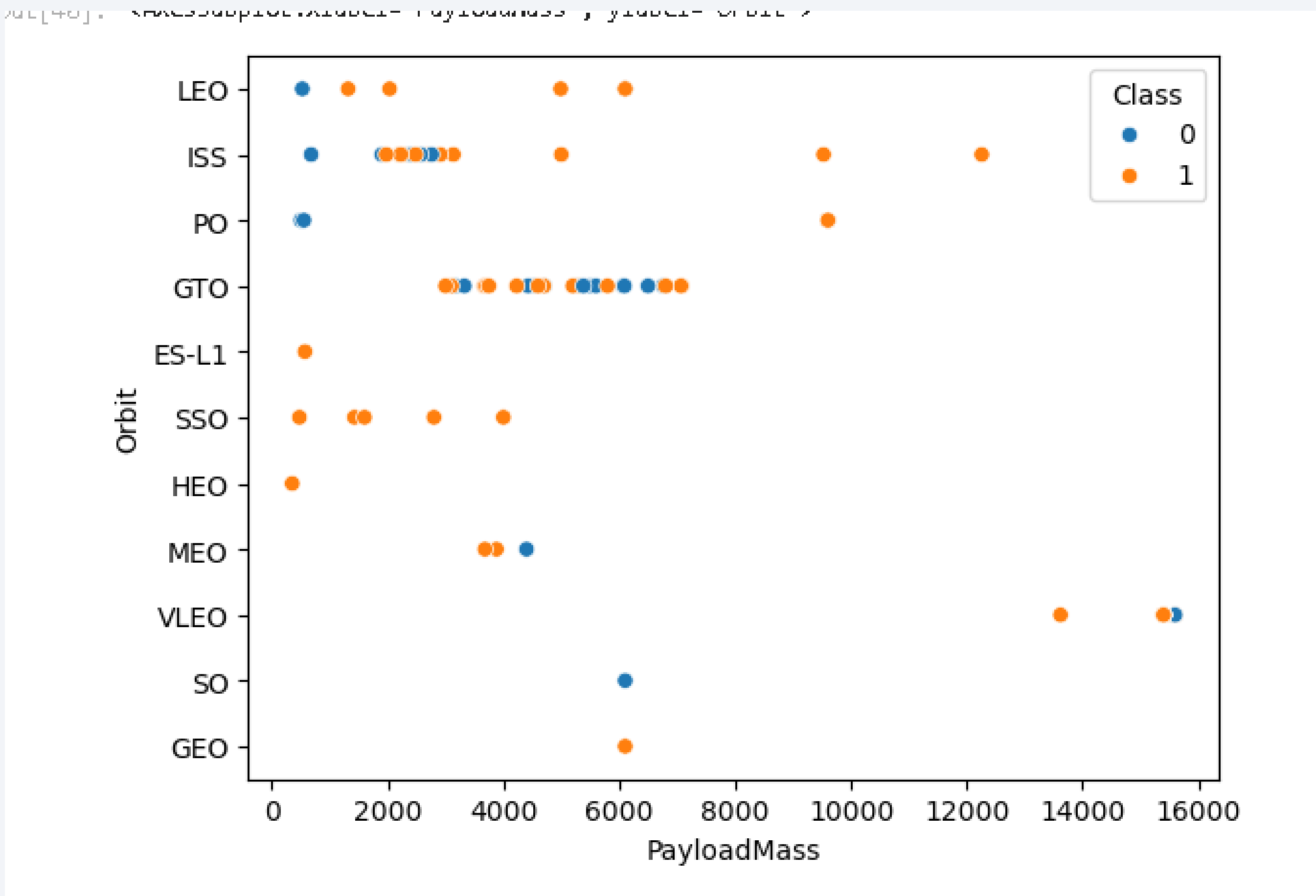
Flight Number vs. Orbit Type

- The more flights that were conducted in a particular orbit increases the success rate, except for ISS.



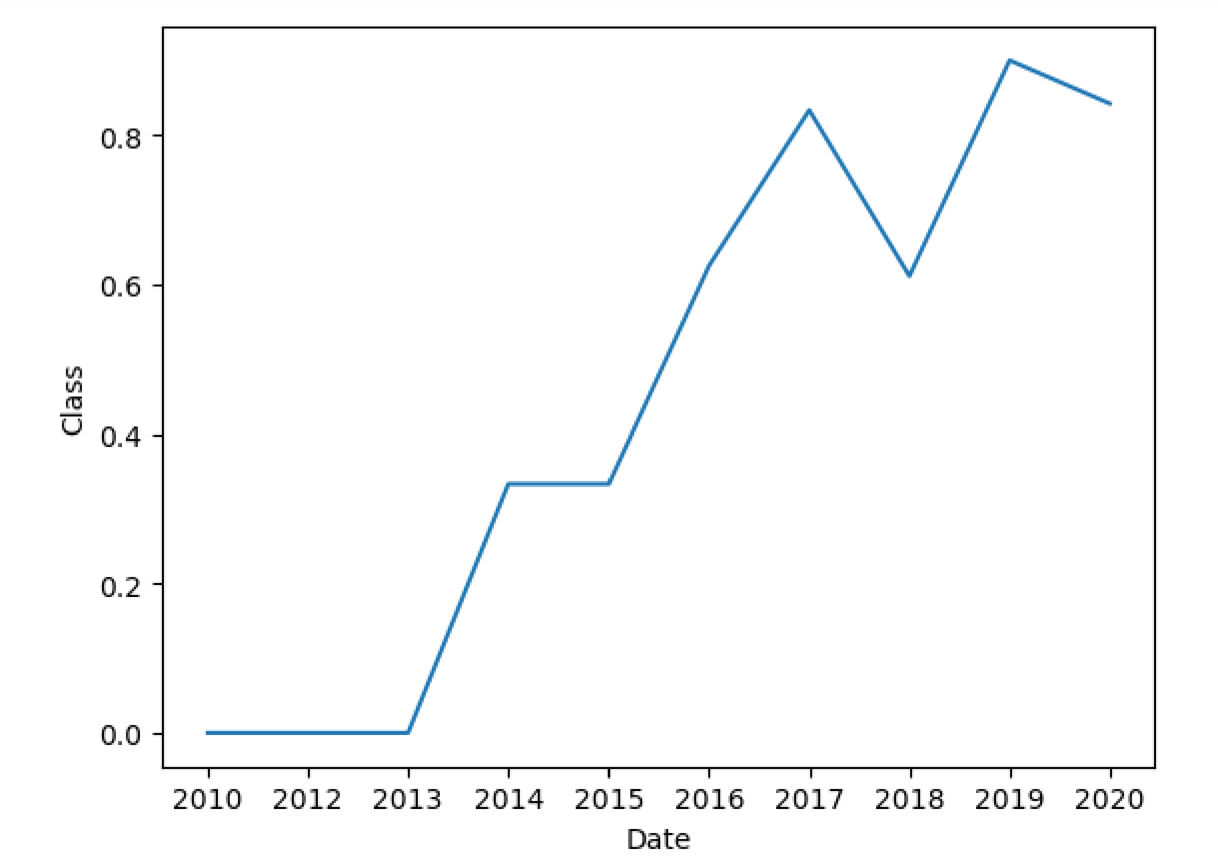
Payload vs. Orbit Type

- More data points should be obtained, but generally, the payload mass related to success isn't correlated when considering payload mass.



Launch Success Yearly Trend

- As you can clearly see, the success rate over time increases starting from 2013 and peaking in 2019.



All Launch Site Names

- 4 launch sites were found. Some of these launch sites are used for higher payloads and some are used for lower payloads

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

Launch Site Names Begin with 'CCA'

As you can see, CCA launch sites have a wide variety of customers and generally don't attempt a landing, based on the first 5 data points.

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

- Calculate the total payload carried by boosters from NASA
- When you sum these numbers, you get 45596 kg total payload carried by boosters launched by NASA (CRS).
- What's interesting to note is that loads are generally 2000 kg, suggesting

NASA (CRS) likes to launch big loads.

PAYLOAD_MASS_KG_
500
677
2296
2216
2395
1898
1952
3136
2257
2490
2708
3310
2205
2647
2697
2500
2495
2268
1977
2972

Average Payload Mass by F9 v1.1

- The average payload for the F9 v1.1 is 2535 kg.

avg(PAYLOAD_MASS__KG_)
2534.6666666666665

First Successful Ground Landing Date

- Below is the first successful ground pad mission, which makes sense since the success rate was at 0% prior to and during 2013.

min(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- These are some of the boosters that can handle relatively large payloads with a successful landing on a drone ship. What's interesting is that NASA (CRS) didn't use these in 2010-2013, likely due to them not using heavy payload rockets.

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

Total Number of Successful and Failure Mission Outcomes

- Most missions were a success; however, success is not based on recapture of the device.

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

Boosters Carried Maximum Payload

- Since Falcon 9 has a variety of boosters that can handle the theoretical maximum pay load, then choosing the one with the highest success rate may be influential in the outcome of success.

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

2015 Launch Records

- Only 2 drone ships landings failed in 2015. We should compare this to the total drone ship launches that year for a drone ship to see the success rate over time.

month	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

Generally, no attempt is the most common way for a landing outcome to occur, with drone ship coming in second with a 50/50 chance of landing.

The best chance of landing seems to be a ground pad.

Done.

Landing_Outcome	count(landing_outcome)
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

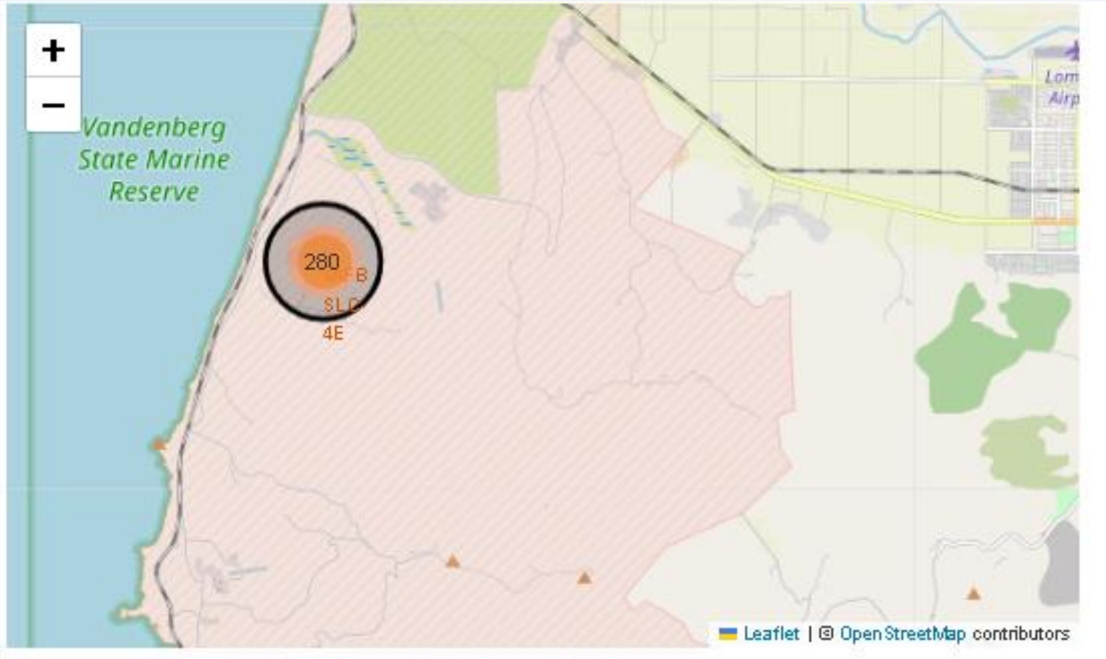
Launch Sites Proximities Analysis

Launch Site Analysis

- As you can see, there are two main hubs, however, the hub in Florida has 2 different launch sites. What is shown on the map are called markers.
- All launch sites are within proximity to the ocean.

Map Marker Clusters

- This shows 3 map marker clusters, which when zoomed out become 2 map marker clusters.



Distance to Coastline

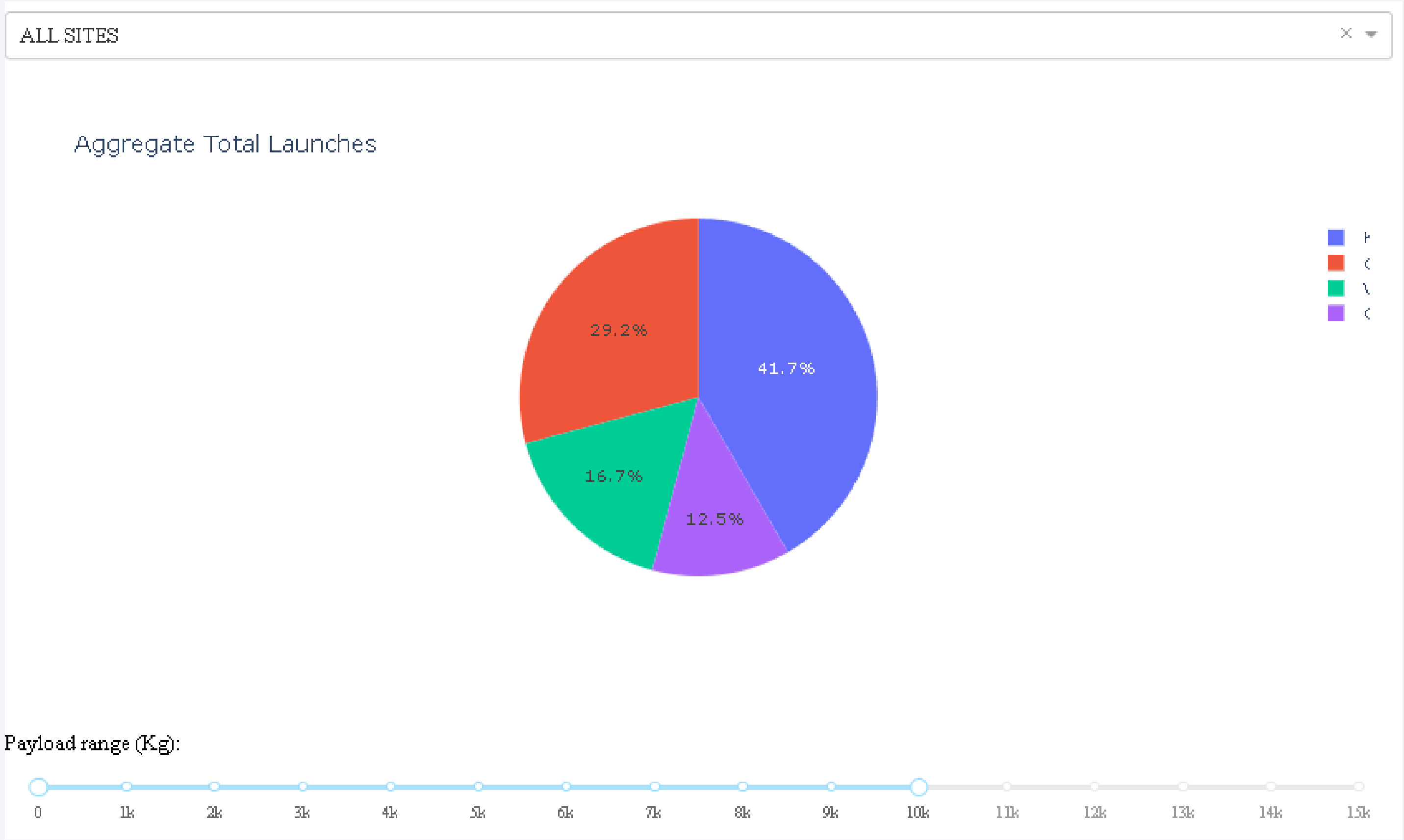
- You can see that this site is within 1 km of the coastline and if you zoom out, it's also located far from any residential buildings.

Section 4

Build a Dashboard with Plotly Dash

Aggregate total launches

- You can see a place to select callbacks, as well as a legend to identify which section belongs to which area.



Section 5

Predictive Analysis (Classification)

Classification Accuracy

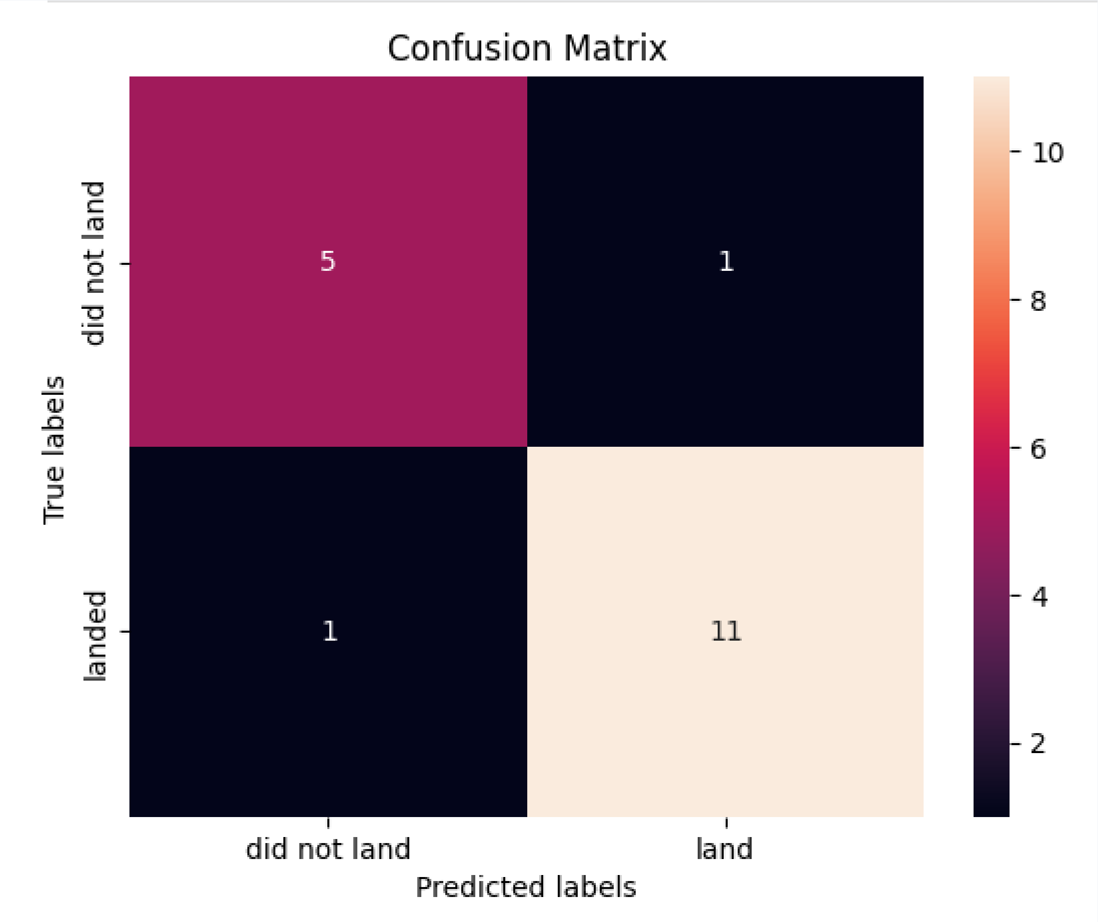
- The model with the highest accuracy is the tree method.

```
tree_cv.score(X_test,Y_test)

0.8333333333333334
```

Confusion Matrix

- This is the confusion matrix for the decision tree. The black boxes are incorrectly labeled labels. As you can see, most of the labels are correct.



Conclusions

- More data points are needed since the decision tree produces the best and worst results due to a lack of observations.
- There is a correlation between launch site, and payload, which suggests that location affects success.
- The success rate has increased over time since 2013 from 0% to a height of 80%.
- Launch sites should be close to the ocean.

Appendix

- See links provided presented in the summary of each step taken.

Thank you!