IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Wing Hong Yi
July 14, 2025

Outline

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

Executive Summary

- Methodology
  - Data was collected from the SpaceX website using REST API, and also from Wikipedia using web scraping.
  - EDA was done making various charts and running various SQL queries.
  - A Folium map was made to show distances from launch sites to various population objects, and success rates of each site.
  - A Plotly Dash dashboard was made to visually compare success rates of each launch site, to find the range of payload mass with the most success, and the most successful booster.
  - Four classification models were tested to predict launch outcomes.

Executive Summary

- Results
  - The rate of successful launch outcomes increased over time.
  - Launch site CAFS SLC 40 had the most launches.
  - Orbit type GTO was used the most.
  - Orbit types ES-L1, GEO, HEO, and SSO had the highest success rate.
  - True ASDS was the most common landing outcome.
  - The names of the four launch sites are CCAFS LC-40, VAFB SLC-4E, KSC LC-39A, and CCAFS SLC-40.
  - Launch sites are located near the ocean and far away from cities and major highways.
  - The 2000 to 4000 kg payload mass range had the best success rate.
  - The FT booster version had the best booster success rate.
  - The Decision Tree Classifier model had the best built model accuracy score.

Introduction

- SpaceX Falcon 9 rocket launches cost about 62 million dollars, while other providers charge upward of 165 million dollars. The cost difference comes from SpaceX's ability to reuse the first stage.

- We want to predict if the first stage will land, so that we can determine the cost of a launch. This information will be helpful for another company to bid against SpaceX for rocket launches.

Section 1

Methodology

Methodology

Executive Summary

In this section we will go over the tasks performed, which include the following:

- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- 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 – SpaceX API

- Initial data collection was obtained from a REST API call to: https://api.spacexdata.com/v4/launches/past
- Although for this project we use a static version of that data from: https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/API_call_spacex_api.json
- **The completed SpaceX API calls notebook**: https://github.com/wing-ny/DS-Capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Obtain the static version data with a REST API call

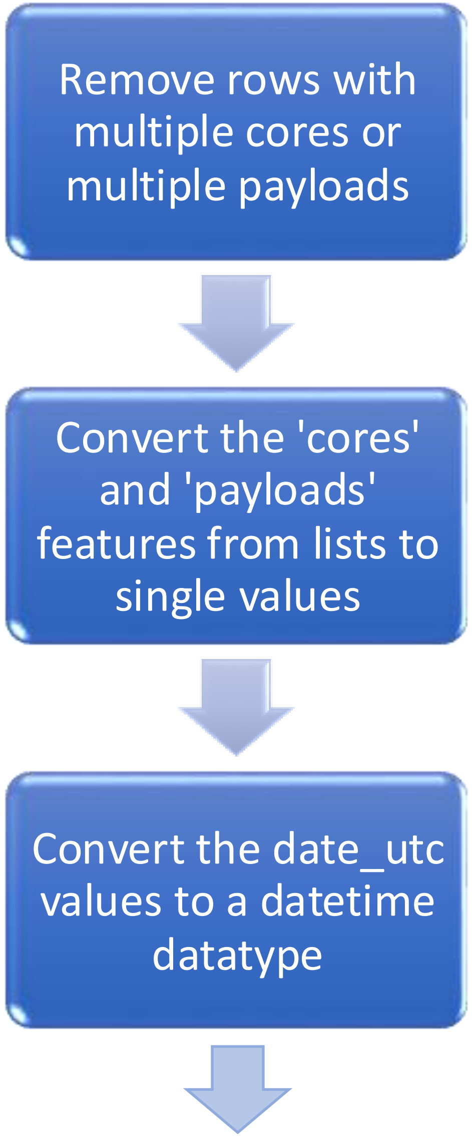
Decode the response content as JSON and turn into a pandas dataframe with pd.json_normalize

Keep only the features: 'rocket', 'payloads', 'launchpad', 'cores', 'flight_number', 'date_utc'

Remove rows with multiple cores or multiple payloads

Data Collection – SpaceX API

(flowchart continued from last slide)



Data Collection – SpaceX API

(flowchart continued from last slide)

- Some of the features are ID values we can use in REST API calls to get further information.

From the rocket column IDs, we can get the booster name with a REST API call to https://api.spacexdata.com/v4/rockets/

From the launchpad column IDs, we can get the launch site name, longitude and latitude with a REST API call to https://api.spacexdata.com/v4/launchpads/

From the payload column IDs, we can get the payload mass and the orbit used with a REST API call to https://api.spacexdata.com/v4/payloads/

Data Collection – SpaceX API

(flowchart continued from last slide)

- After getting one more batch of data, we can assemble the dataframe.

From the cores column IDs, we can get the landing outcome, landing type, number of flights with that core, if gridfins were used, if the core is reused, if legs were used, if the landing pad was used, the block of the core, the number of times a specific core has been reused, and the serial of the core with a REST API call to https://api.spacexdata.com/v4/cores/

Convert the collected data into a dataframe with features 'FlightNumber', 'Date', 'BoosterVersion', 'PayloadMass', 'Orbit', 'LaunchSite', 'Outcome', 'Flights', 'GridFins', 'Reused', 'Legs', 'LandingPad', 'Block', 'ReusedCount', 'Serial', 'Longitude', 'Latitude'

Data Collection – SpaceX API

(flowchart continued from last slide)

- Missing values in the LandingPad column are retained to indicate no landing pad used.

Remove all launches from the dataframe that are not Falcon 9

↓

With some rows now removed, reset the FlightNumber column numbering

↓

Convert the date_utc values to a datetime datatype

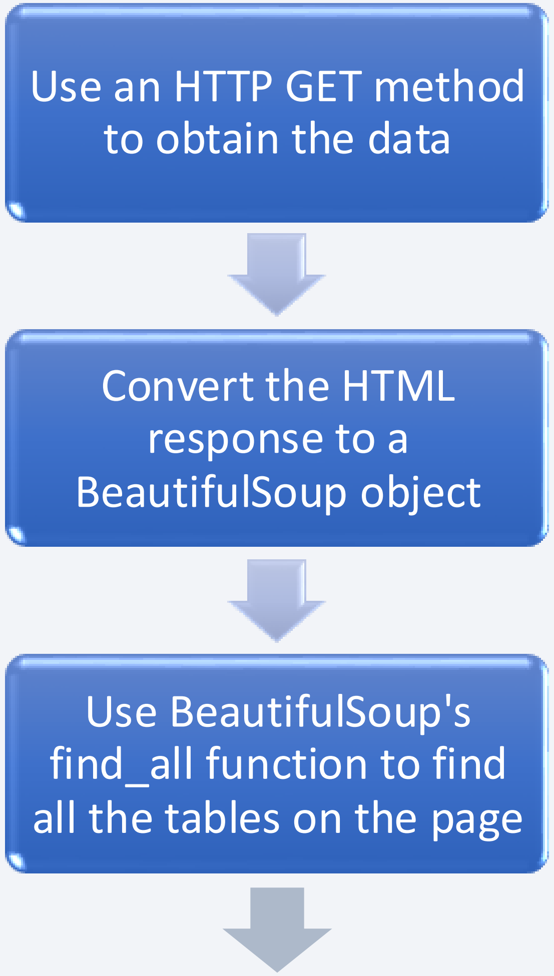
↓

Missing values in the PayloadMass column are replaced with the mean of the column values that are available.

Data Collection - Scraping

- The initial data is obtained from a snapshot of the List of Falcon 9 and Falcon Heavy launches Wikipage updated on June 9, 2021. https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922

- The completed web scraping notebook: https://github.com/wing-ny/DS-Capstone/blob/main/jupyter-labs-webscraping.ipynb



Data Collection - Scraping

- (flowchart continued from the previous slide)

- **The completed web scraping notebook:** https://github.com/wing-ny/DS-Capstone/blob/main/jupyter-labs-webscraping.ipynb

We find that the launch records we will use start from the third table

↓

Extract column names by iterating through the <th> elements of the third table, obtaining: ['Flight No.', 'Date and time ( )', 'Launch site', 'Payload', 'Payload mass', 'Orbit', 'Customer', 'Launch outcome']

↓

Remove 'Date and time ( )' from the column list, add new columns 'Version Booster', 'Booster landing', 'Date', and 'Time'

↓

Data Collection - Scraping

- (flowchart continued from the previous slide)

- **The completed web scraping notebook:** https://github.com/wing-ny/DS-Capstone/blob/main/jupyter-labs-webscraping.ipynb

For each table that is "wikitable plainrowheaders collapsible", iterate through each row

↓

For each row we extract values for 'Flight No.', 'Date', 'Time', 'Version Booster', 'Launch site', 'Payload', 'Payload mass', 'Orbit', 'Customer', 'Launch outcome', 'Booster landing'

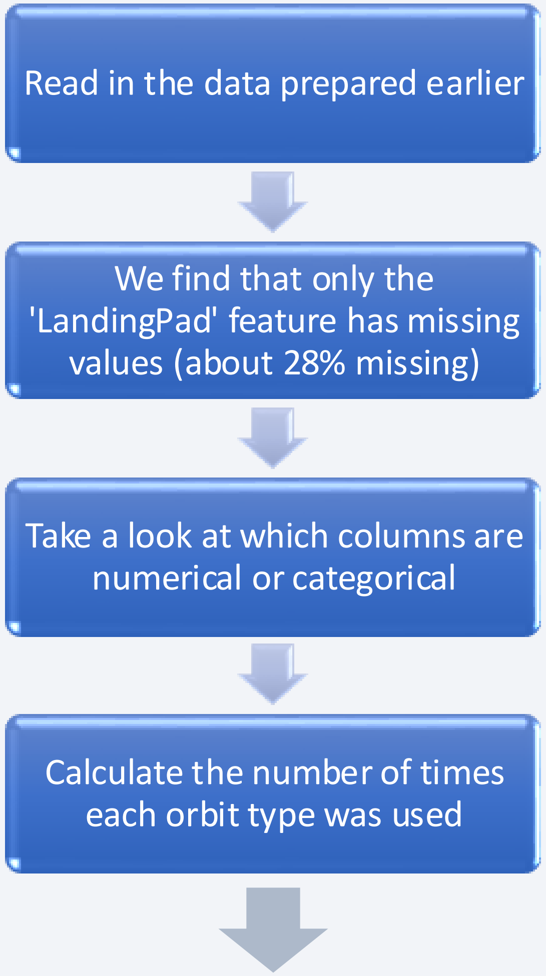
↓

Assemble the collected data into a pandas DataFrame

Data Wrangling

- We start with some EDA to look for patterns, and decide what will be the label to use for training supervised models.

- **The completed data wrangling notebook:** https://github.com/wing-ny/DS-Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb



Data Wrangling

- (flowchart continued from the previous slide)
- A landing outcome is considered bad if it is one of 'False ASDS', 'False Ocean', 'False RTLS', 'None ASDS', or 'None None'
- A successful outcome would be one of 'True ASDS', 'True RTLS', or 'True Ocean'
- The completed data wrangling notebook: https://github.com/wing-ny/DS-Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

Calculate the number of times each type of landing outcome occurred

↓

Create a new column 'Class', which is assigned a value of 0 for bad landing outcomes, or 1 for good landing outcomes

↓

We look at the mean of the 'Class' column, which tells us the landing success rate is about 66.7%

EDA with Data Visualization

- To visually check if there are any relationships between certain pairs of features, we made scatter plots for the following:
  - Flight Number vs. Payload Mass (kg)
  - Flight Number vs. Launch Site
  - Payload Mass (kg) vs. Launch Site
  - Flight Number vs. Orbit
  - Payload Mass (kg) vs. Orbit
- A bar chart was made to show the success rate for each type of orbit used
- A line chart was made to show the average yearly success rate from 2010 to 2020
- The completed EDA with data visualization notebook: https://github.com/wing-ny/DS-Capstone/blob/main/edadataviz.ipynb

EDA with SQL

- SQL queries were performed to:
  - List unique launch site names used in the missions
  - List 5 records where the launch site name begins with 'CCA'
  - Display the total payload mass carried by boosters launched by NASA (CRS)
  - Display average payload mass carried by booster version F9 v1.1
  - Display the date of the first successful landing outcome on a ground pad
  - List the names of the boosters which have success in drone ship and have payload mass greater than 4000 kg and less than 6000 kg
  - List the total number of successful and failure mission outcomes

EDA with SQL

- (continued) SQL queries were performed to:
  - List the names of the booster versions which have carried the maximum payload mass
  - List the records which will display the month names, failure landing outcomes in drone ship, booster versions, launch site for the months in year 2015
  - Rank the count 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
- The completed EDA with SQL notebook: https://github.com/wing-ny/DS-Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- A map with a circle map object was created to show the location of NASA Johnson Space Center at Houston, Texas
- A map with circle map objects was created to show the locations of each of the 4 launch sites used in the dataset
- Marker clusters were added for every launch outcome, green for successful, red for failure
- Lines were added to show distances from CCAFS SLC-40 and VAFB SLC-4E to nearby highways, railways, coastlines and cities

Build an Interactive Map with Folium

- Since CCAFS LC-40 and KSC LC-39A are very close to CCAFS SLC-40, lines were not added to show the same distances for those launch sites
- The completed interactive Folium maps notebook: https://github.com/wing-ny/DS-Capstone/blob/main/lab_jupyter_launch_site_location.ipynb
- The html version when downloaded and viewed in a browser has working maps: https://github.com/wing-ny/DS-Capstone/blob/main/lab_jupyter_launch_site_location.html

Build a Dashboard with Plotly Dash

- A dropdown was added to allow for selection of all launch sites, or one of four specific launch sites.
- For the All Sites selection, it displays a pie chart to show what percentage of successful launches were made at each launch site out of all successful launches at every launch site.
    - This let us see visually which site had the most successful launches.
- For a specific launch site selection, it displays a pie chart showing the percentage of successful (1), or unsuccessful launches (0) from there.
    - This allows us to see visually how successful a particular launch site was.

Build a Dashboard with Plotly Dash

- A range slider was added to allow the selection of a payload mass range from 0 to 10000 kg.
- Based on the range selected and the launch site dropdown, a scatter plot of Payload mass vs. class (0 for failure, 1 for success) is drawn. Each point is also color coded to show what booster version category was used.
    - The range slider helps us visually find payload ranges which had the most success.
- **The completed Plotly Dash lab:** https://github.com/wing-ny/DS-Capstone/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

- The flowchart shows the steps taken to get the data ready for the classification modeling.
- The Class column from dataset_part_2.csv will be used as 'Y', the real outcomes that we want the models to be able to predict.
- The data from dataset_part_3.csv are the features, 'X', the models will use to make the predictions.

Read in the data from dataset_part_2.csv and dataset_part_3.csv.

↓

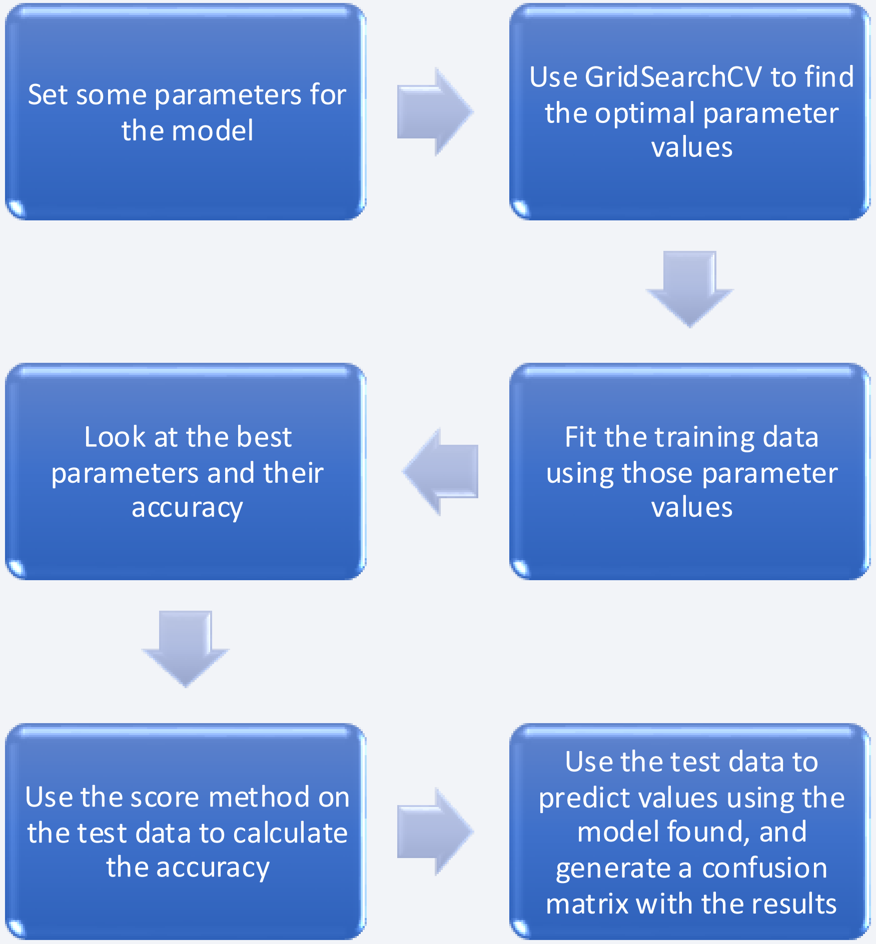
Use StandardScaler so that no columns in X have too great or too little influence due to differences in the magnitude of the values.

↓

Use train_test_split to set aside 80% of the X and Y data for training, 20% for testing. We also use random_state=2 for reproducibility.

Predictive Analysis (Classification)

- Next we repeat the steps shown on the right for each of these 4 modeling methods:
    - Logistic Regression
    - Support Vector Machine
    - Decision Tree Classifier
    - K Nearest Neighbors
- The completed predictive analysis lab: https://github.com/wing-ny/DS-Capstone/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb



Results

- Performing EDA with pandas, we found:
  - Only the LandingPad attribute had missing values (28.9% missing).
  - Launch site CAFS SLC-40 had 55 launches, KSC LC-39A had 22 launches, VAFB SLC-4E had 13 launches.
  - Of the orbit types, GTO was used 27 times, ISS 21 times, VLEO 14 times, PO 9 times, LEO 7 times. The remaining types were used 5 or fewer times: SSO, MEO, HEO, ES-L1, SO, GEO.
  - Of the landing outcomes, True ASDS occurred 41 times, "None None" 19 times, True RTLS 14 times, and the remaining occurred 6 or fewer times: False ASDS, True Ocean, False Ocean, None ASDS, False RTLS.

Results

- Performing EDA with SQL, we found:
  - The names of each launch site used: CCAFS LC-40, VAFB SLC-4E, KSC LC-39A, CCAFS SLC-40.
  - The total payload mass launched for the customer NASA (CRS): 45596 kg
  - The average payload mass carried by booster version F9 v1.1: 2534.7 kg
  - The date of the first successful landing outcome in ground pad: 2015-12-22
  - The booster versions with successful landing in drone ship and payload mass greater than 4000 kg and less than 6000 kg: F9 FT B1022, F9 FT B1026, F9 FT B1021.2, F9 FT B1031.2
  - The total number of successful mission outcomes (100), and failure outcomes (1).

Results

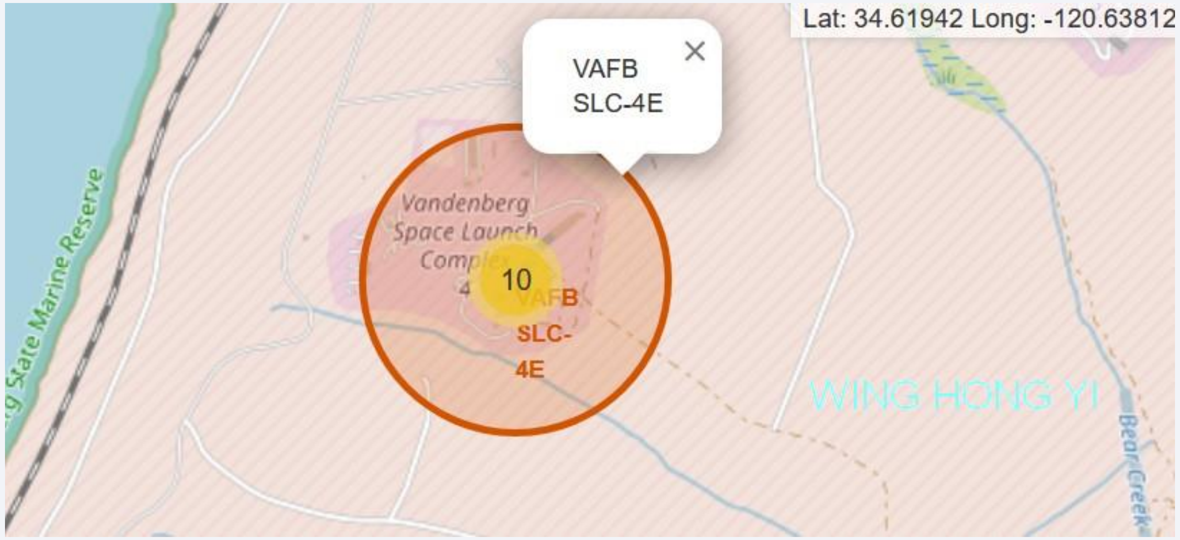
- (continued) Performing EDA with SQL, we found:
  - The booster versions that have carried the maximum payload mass. (see Task 8 in the lab notebook: https://github.com/wing-ny/DS-Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb)
  - There were two records where the landing outcome was failure in drone ship in 2015.
  - The total landing outcomes between 2010-06-04 and 2017-03-20 in descending order were: No attempt 10 times, Success (drone ship) 5 times, Failure (drone ship)5 times, Success (ground pad) 3 times, Controlled (ocean) 3 times, Uncontrolled (ocean) 2 times, Failure (parachute) 2 times, Precluded (drone ship) 1 times.

Results

- A Folium map was created, with one launch site on the west coast and 3 sites on the east coast.

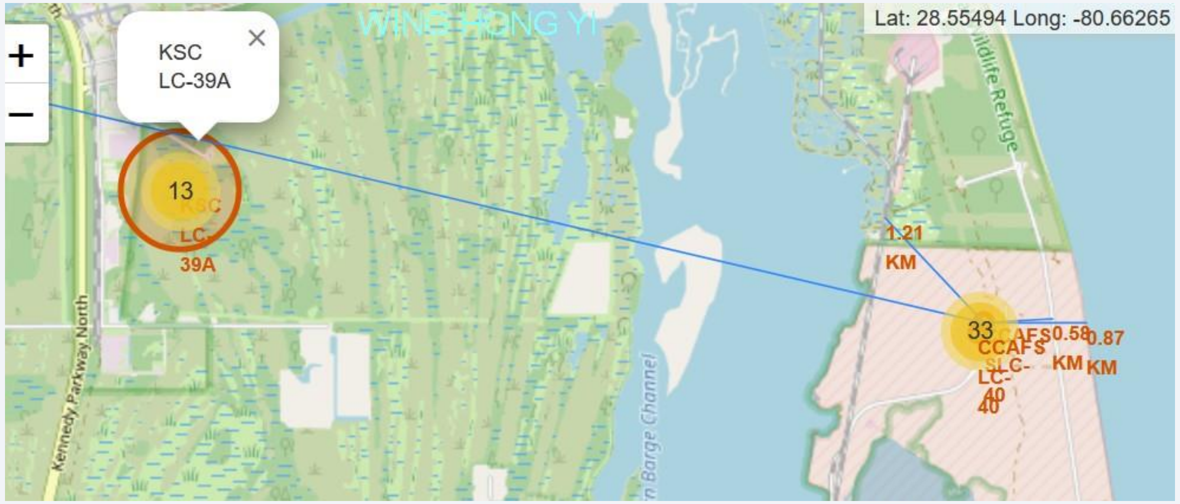
Results

- This shows the launch site on the west coast, VAFB SLC-4E. The green markers indicate successful launches, red markers indicate failures.



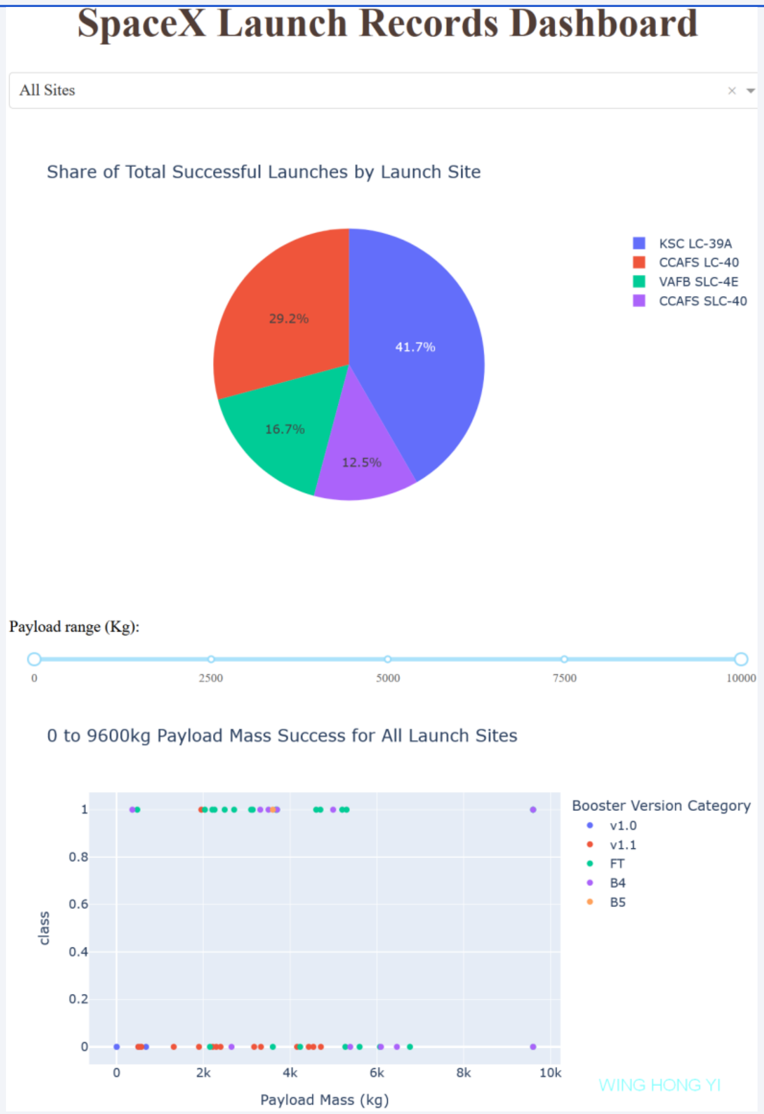
Results

- This shows the 3 launch sites on the east coast, KSC LC-39A on the left, CCAFS SLC-40 and CCAFS LC-40 very close to each other on the right.



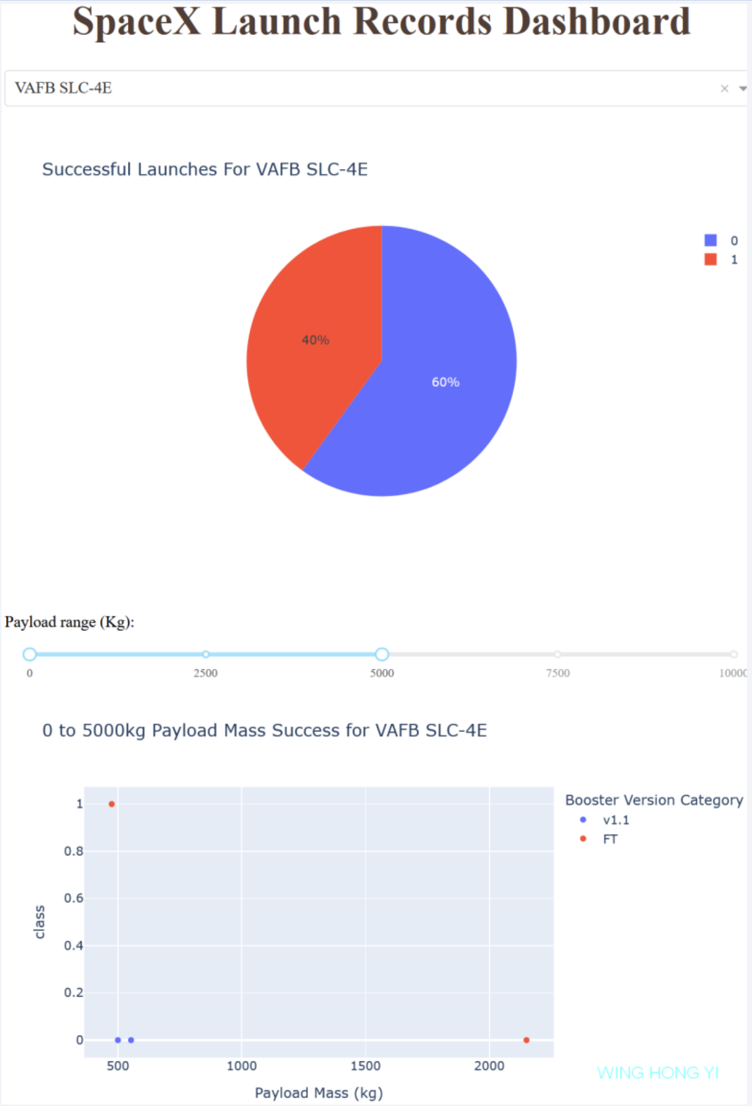
Results

- The dashboard allows you to select All Sites, or a specific launch site.
- A range slider allows a range to be selected for the payload mass, which is shown in a scatter plot versus launch success (0 for failure, 1 for success).
- A color coded booster version is also shown as the scatter plot points.



Results

- When a specific launch site is selected, the pie chart shows the percentage of successful and unsuccessful launches for that site.



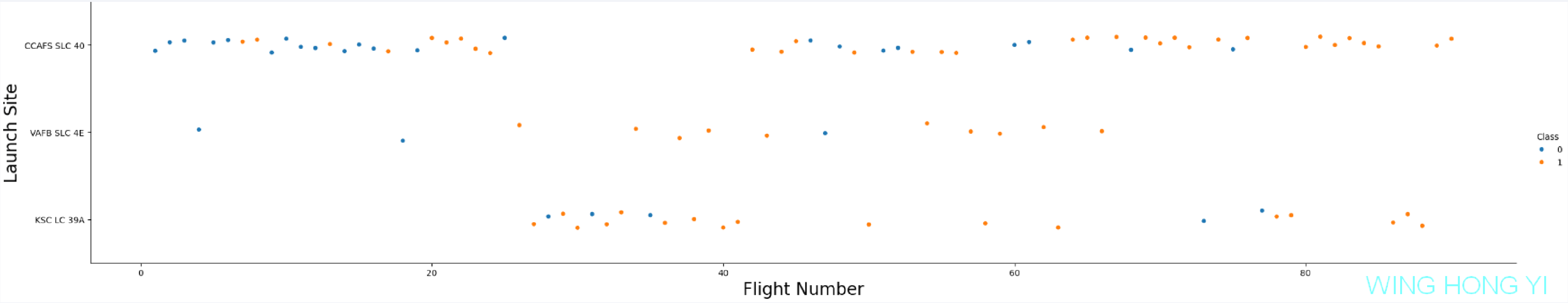
Results

- Predictive analysis results
  - The four models used, Logistic Regression, SVM, Decision Tree Classifier, and K Nearest Neighbors, performed with the same accuracy. Since the Decision Tree Classifier is fairly simple, we can choose to use that one if we must pick only one.

Section 2

Insights drawn from EDA

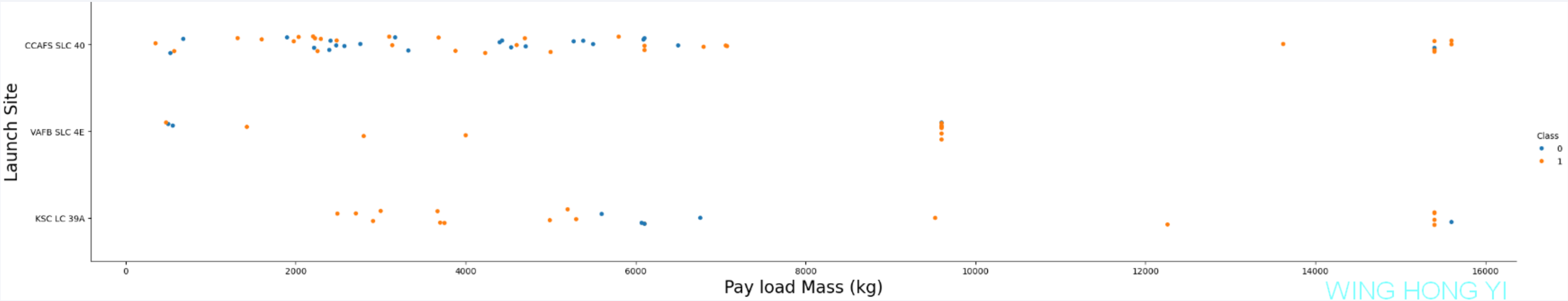
Flight Number vs. Launch Site



- With the darker colored points representing unsuccessful outcomes, lighter representing success, the scatter plot shows more unsuccessful outcomes for earlier flights, more successful outcomes for the later flights.

- The launch site CCAFS SLC 40 appears to have the most launches.

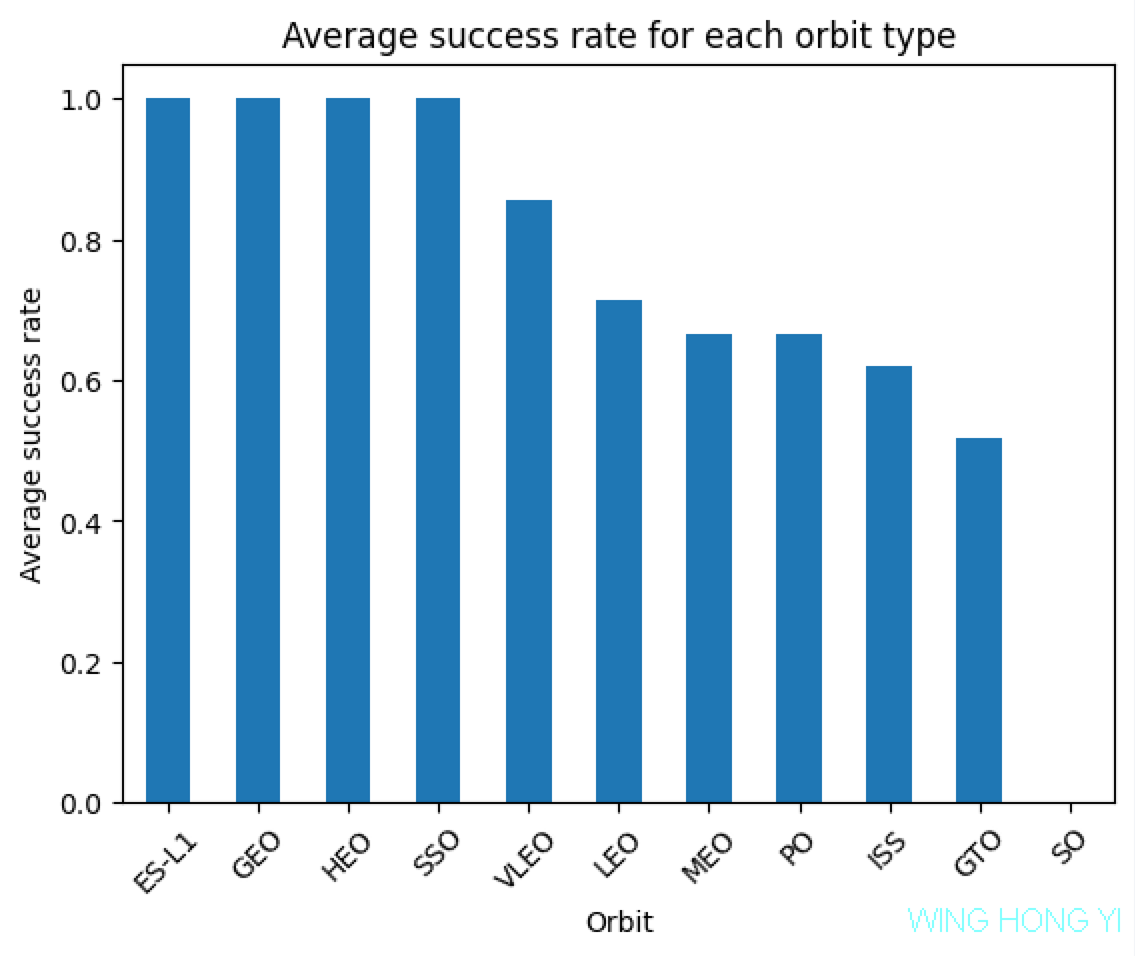
Payload vs. Launch Site



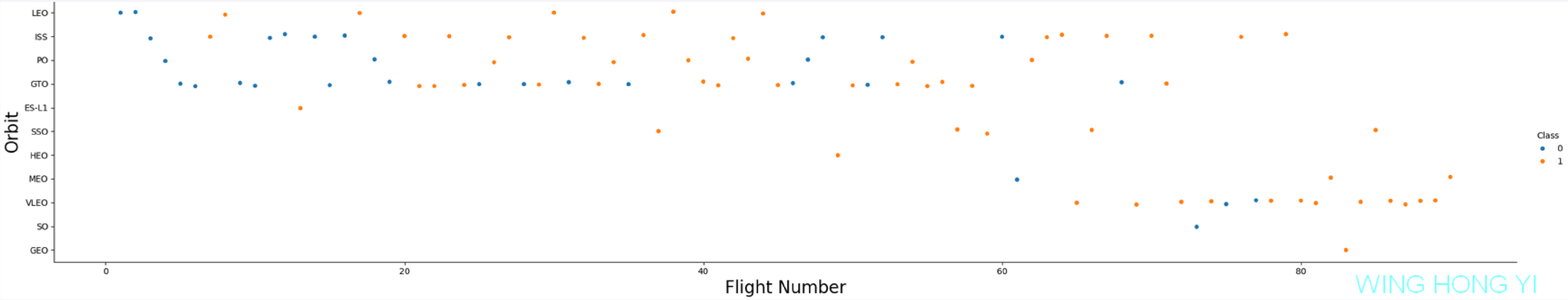
- Most of the launches had payload mass under 8000 kg.
- Outcomes were mixed at CCAFS SLC 40, while KSC LC 39A started having failures as the mass approached 6000 kg.
- Most launches with mass above 8000 kg had successful outcomes.

Success Rate vs. Orbit Type

- Orbit types ES-L1, GEO, HEO, and SSO had the highest success rate.

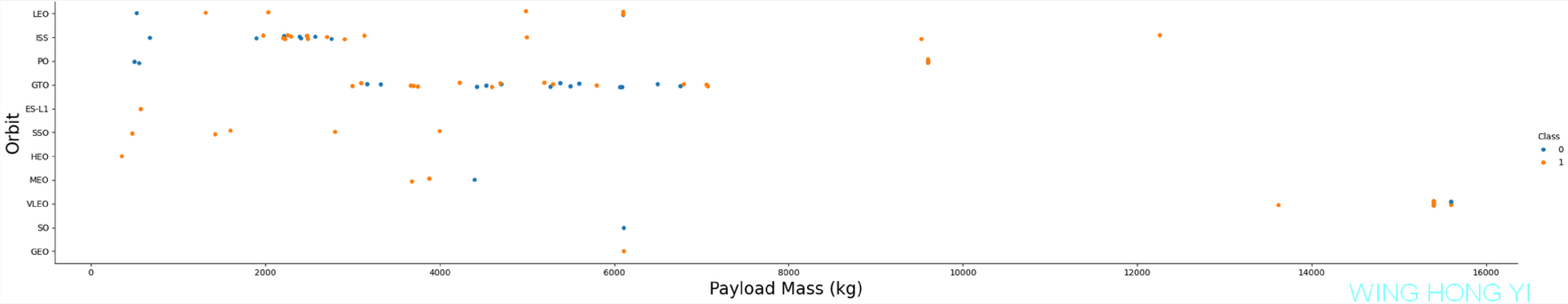


Flight Number vs. Orbit Type



- Overall, the higher the flight number the more successful outcomes tended to be
- Of the orbit types with more than a few launches, GTO looks like it had the worst success rate, which we saw from the bar graph earlier

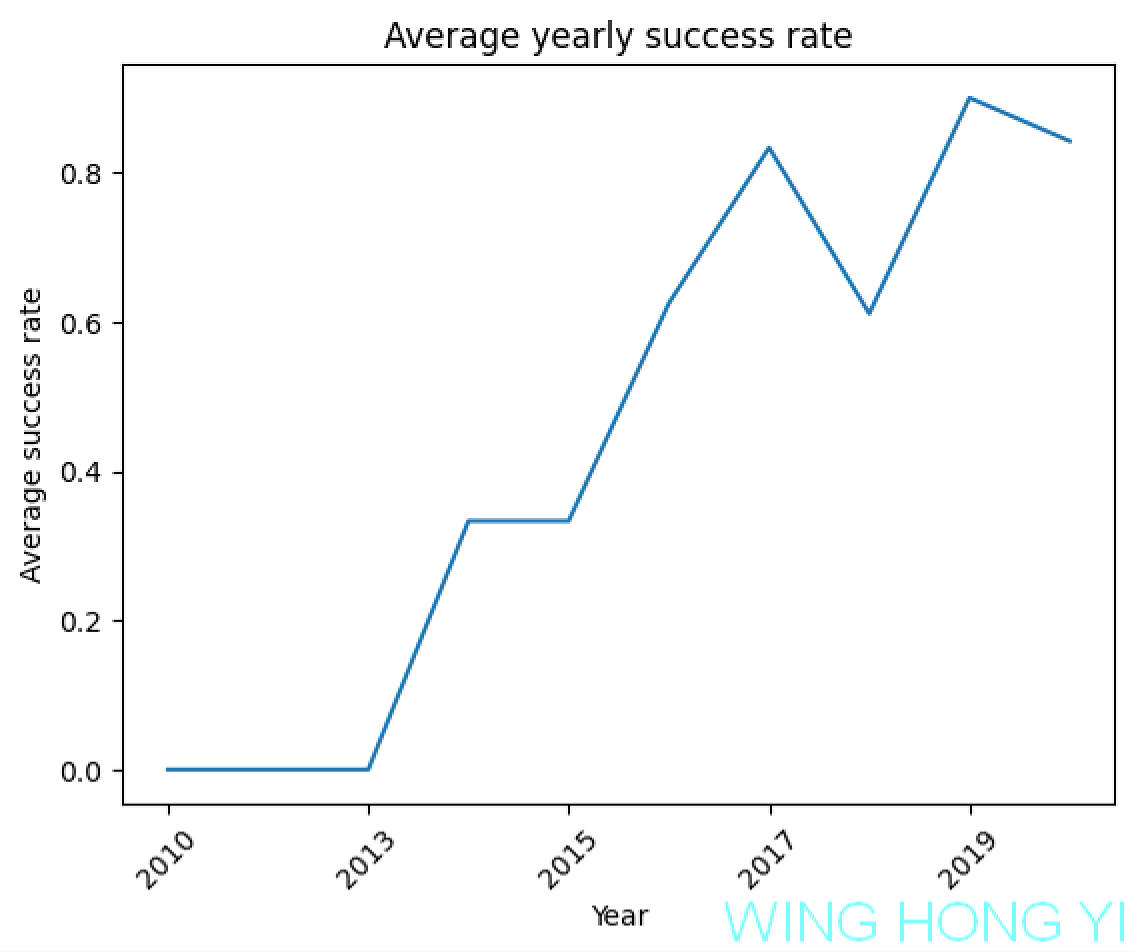
Payload vs. Orbit Type



- Most launch payloads were under 8000 kg, with ISS and GTO the most common orbit types
- ISS and GTO also have the lowest outcome success rates among the orbit types used more than a few times

Launch Success Yearly Trend

- The average success rate generally improved over time



All Launch Site Names

- We found the names of the unique launch sites using the SQL query: `SELECT DISTINCT "Launch_Site" FROM SPACEXTABLE`
    - CCAFS LC-40
    - VAFB SLC-4E
    - KSC LC-39A
    - CCAFS SLC-40

Launch Site Names Begin with 'CCA'

- We found 5 launch sites names beginning with 'CCA' using the SQL query: `SELECT * FROM SPACEXTABLE WHERE "Launch_Site" LIKE "CCA%" LIMIT 5`

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

- We found total payload carried by boosters for NASA (CRS) using the SQL query: `SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTABLE WHERE Customer = 'NASA (CRS)'`
    - 45596 kg

Average Payload Mass by F9 v1.1

- We found the average payload mass carried by booster version F9 v1.1 using the SQL query: `SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTABLE WHERE "Booster_Version" LIKE "F9 v1.1%"`
  - 2534.6666666666665 kg

First Successful Ground Landing Date

- We found the date of the first successful landing outcome on ground pad using the SQL query: `SELECT MIN("Date") FROM SPACEXTABLE WHERE "Landing_Outcome" = "Success (ground pad)"`
    - 2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- We found the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000 using the SQL query: `SELECT "Booster_Version" FROM SPACEXTABLE`
- `WHERE "Landing_Outcome" = "Success (drone ship)"`
- `AND PAYLOAD_MASS__KG_ > 4000`
- `AND PAYLOAD_MASS__KG_ < 6000`
  - F9 FT B1022
  - F9 FT B1026
  - F9 FT B1021.2
  - F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- We found the total number of successful mission outcomes using the SQL query: `SELECT COUNT("Mission_Outcome") FROM SPACEXTABLE WHERE "Mission_Outcome" LIKE "Succ%"`
    - 100
- For failures we used: `SELECT COUNT("Mission_Outcome") FROM SPACEXTABLE WHERE "Mission_Outcome" LIKE "Fail%"`
    - 1

Boosters Carried Maximum Payload

- We found the names of the boosters which have carried the maximum payload mass using the SQL query: `SELECT "Booster_Version" FROM SPACEXTABLE`
- `WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTABLE)`
  - Results listed on the right

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

- A `month_name` column (Jan, Feb, Mar, etc.) was added first to the data in order to satisfy the requirement to display the month names in the results. (see Appendix)
- We found the list of failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015 using the query:

```
SELECT month_name,
"Landing_Outcome", "Booster_Version", "Launch_Site"
 FROM SPACEXTABLE
 WHERE substr(Date, 0, 5) = '2015'
 AND "Landing_Outcome" = "Failure (drone ship)"
```

month_name	Landing_Outcome	Booster_Version	Launch_Site
Jan	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
Apr	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

- We ranked the count 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 using the SQL query: `SELECT "Landing_Outcome", COUNT(*) as count`
- `FROM SPACEXTABLE`
- `WHERE "Date" BETWEEN "2010-06-04" AND "2017-03-20"`
- `GROUP BY "Landing_Outcome"`
- `ORDER BY count DESC`
  - The result is shown on the right

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 Map – All Launch Site Locations

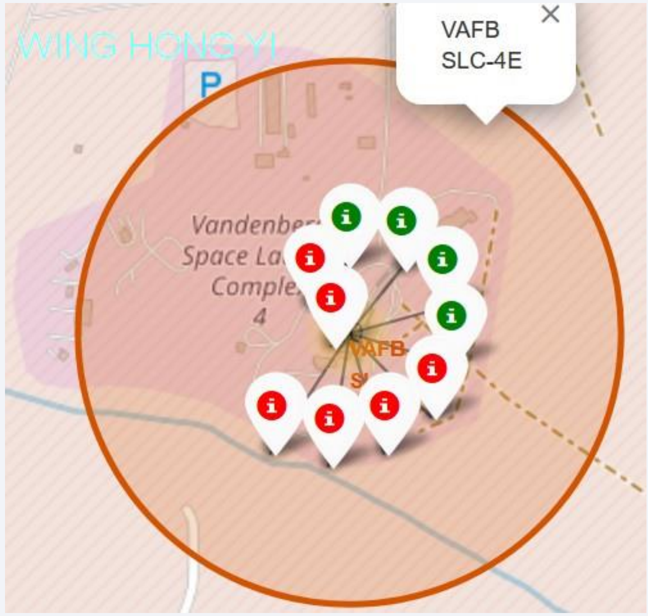
- VAFB SLC-4E is on the west coast
- KSC LC-39A, CCAFS LC-40, CCAFS SLC-40 are in very close proximity to each other on the east coast
- All the launch sites are near the ocean, to minimize damage to populations in case of failure

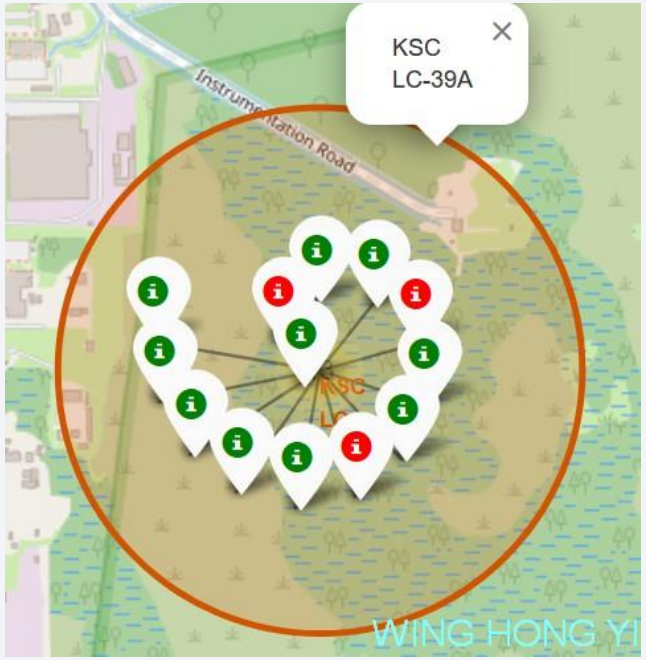
Folium Map – Outcome Marker Clusters

- Marker clusters were added to the launch sites to show successful outcomes (green), and unsuccessful outcomes (red).

- VAFB SLC-4E had more unsuccessful outcomes than successful

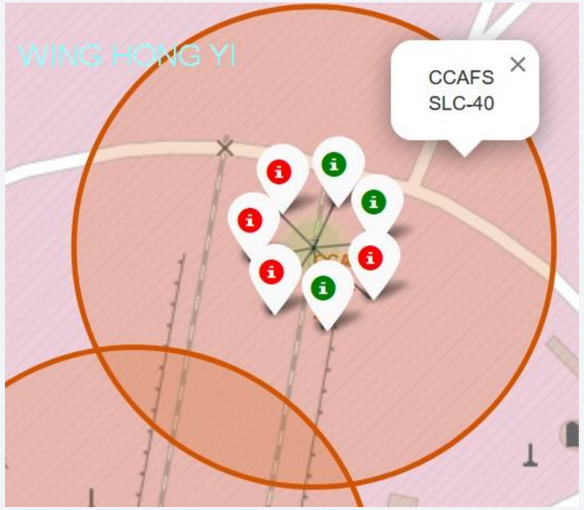
- KSC LC-39A had almost all successful outcomes



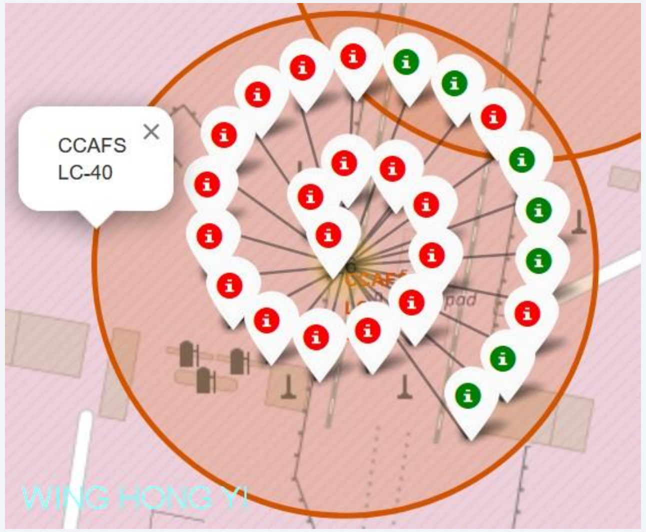


Folium Map – Outcome Marker Clusters

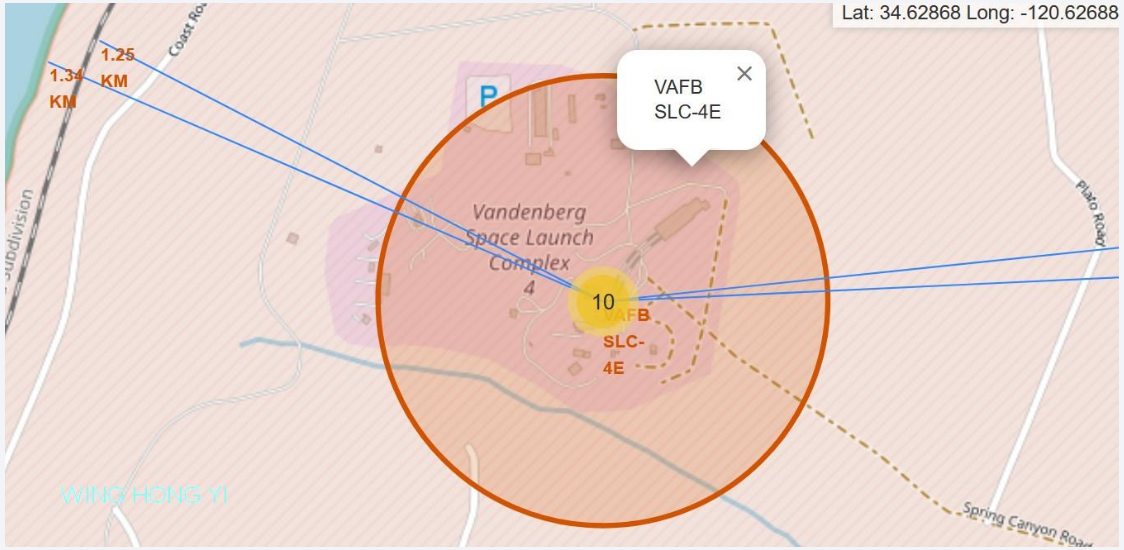
- These are the two launch sites that are very close to each other.



- CCAFS LC-40 has the most launches, with more unsuccessful outcomes than successful.
- CCAFS SLC-40 has only a few launches, with the results being almost even.

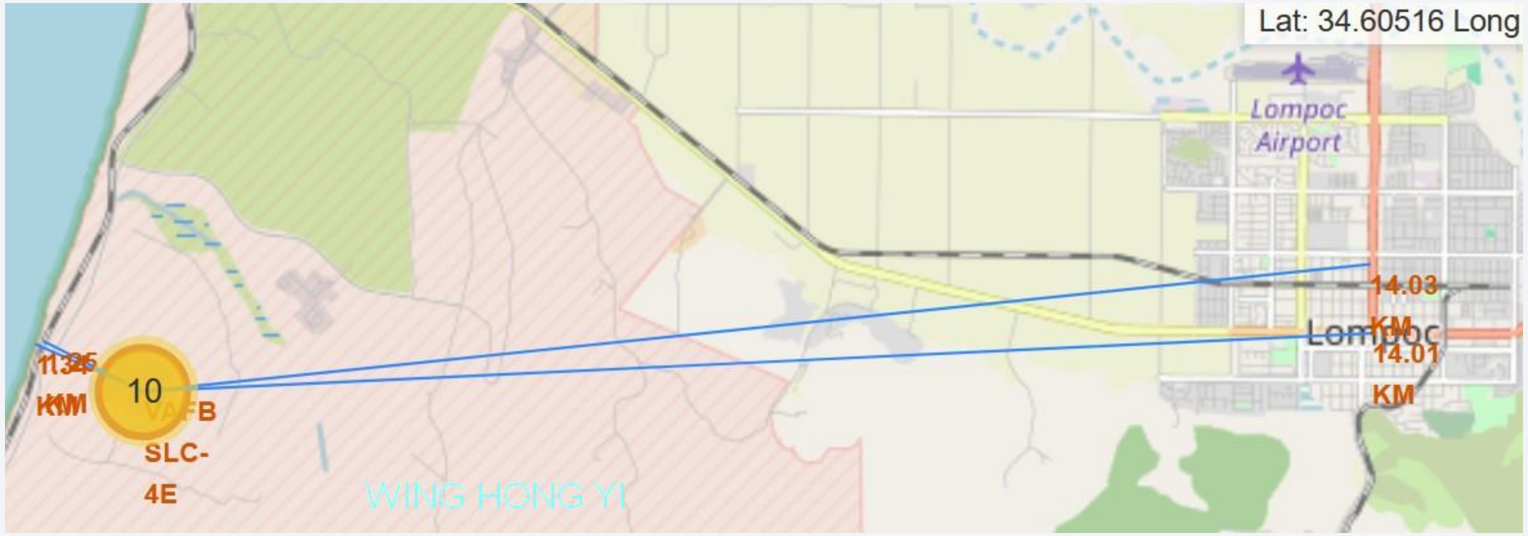


Folium Map - Launch Site Proximities



- VAFB SLC-4E is about 1.25 km from a railway and about 1.34 km from the coastline.
- The site is near the coastline so debris can fall into the water in case of launch failure.

Folium Map - Launch Site Proximities



- VAFB SLC-4E is about 14 km from the nearest city and major highway
- The site is far from the city and major highway so as to minimize the possible damage to people in case of launch failure.
- The other launch sites are similarly close to a coastline and railway, and far from a city and major highway.

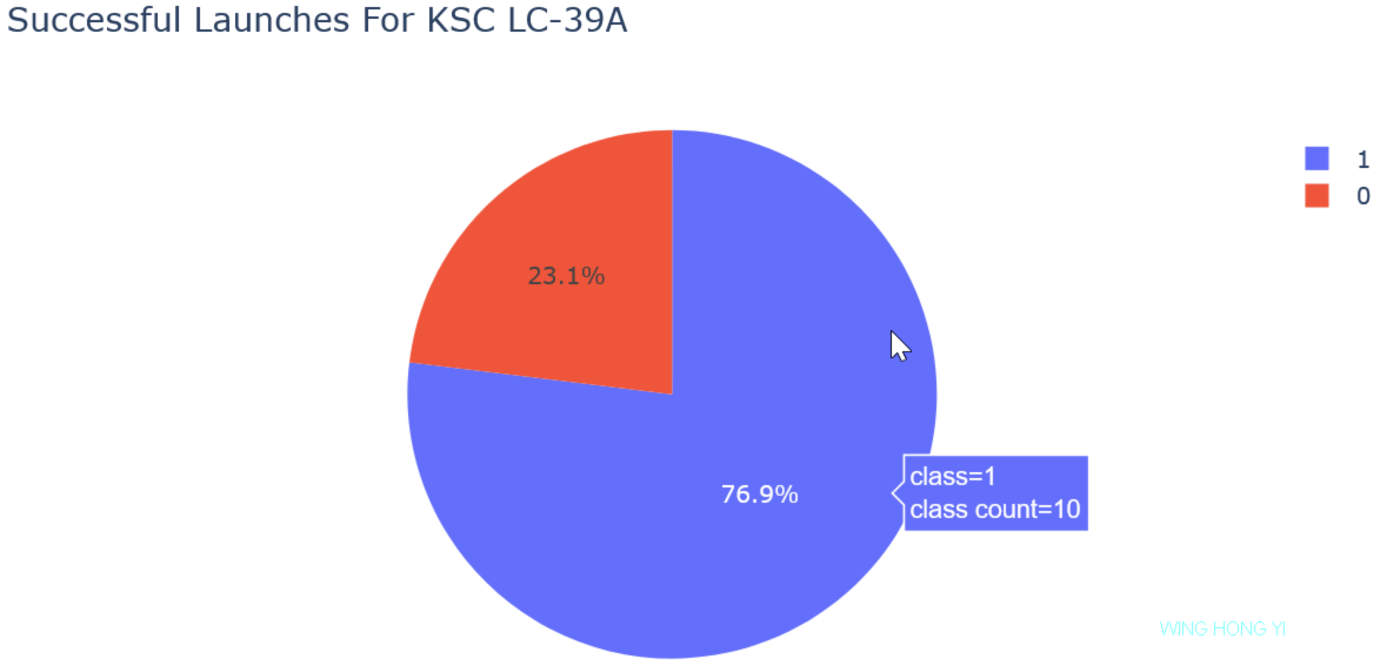
Section 4

Build a Dashboard with Plotly Dash

Share of Total Successful Launches by Site

- KSC LC-39A has the largest share of successful Launches, followed by CCAFS LC-40, VAFB SLC-4E, and CCAFS SLC-40

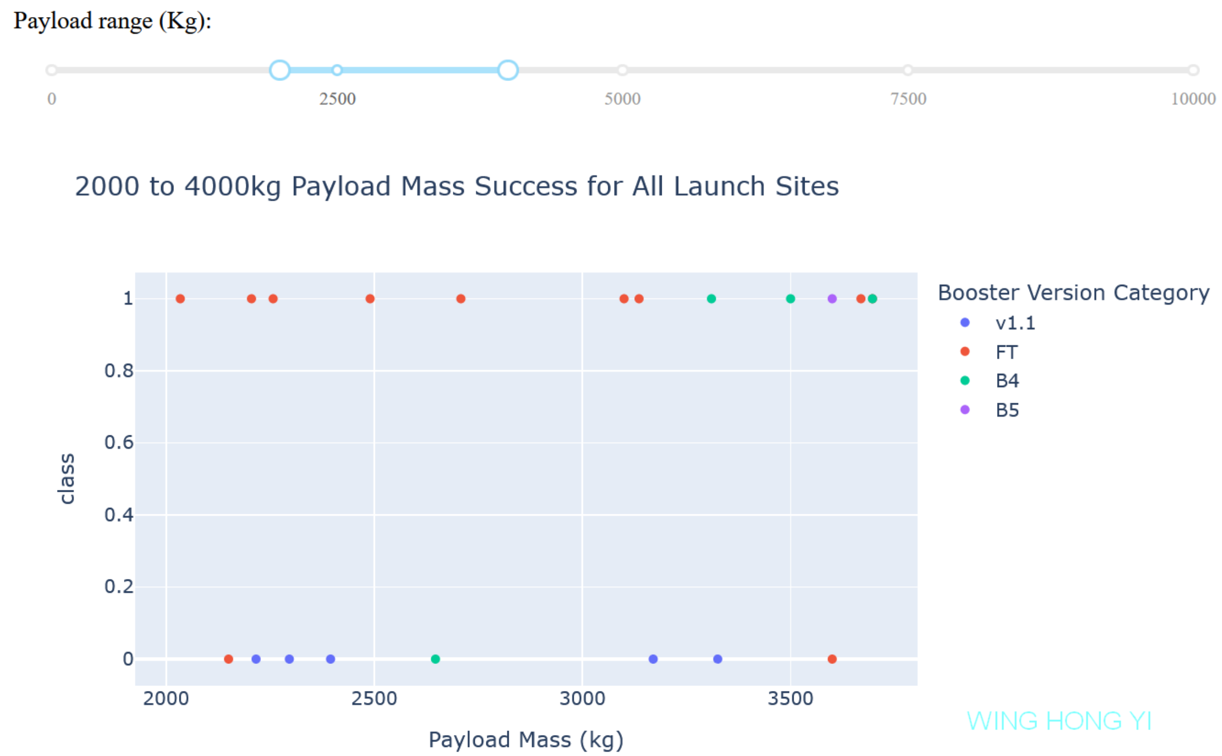
Highest Success Ratio: KSC LC-39A



- KSC LC-39A has a 76.9% success rate (success=1, failure=0)
- There were 10 successes in the data, so then there were 3 failures

Payload vs. Launch Outcome for All Sites

- The 2000 to 4000 kg payload range appears to have the best success rate
- The FT booster version (red dots) appear more often for class 1 (success) than they do for class 0 (failure), and appears to have a better success rate than the other booster versions

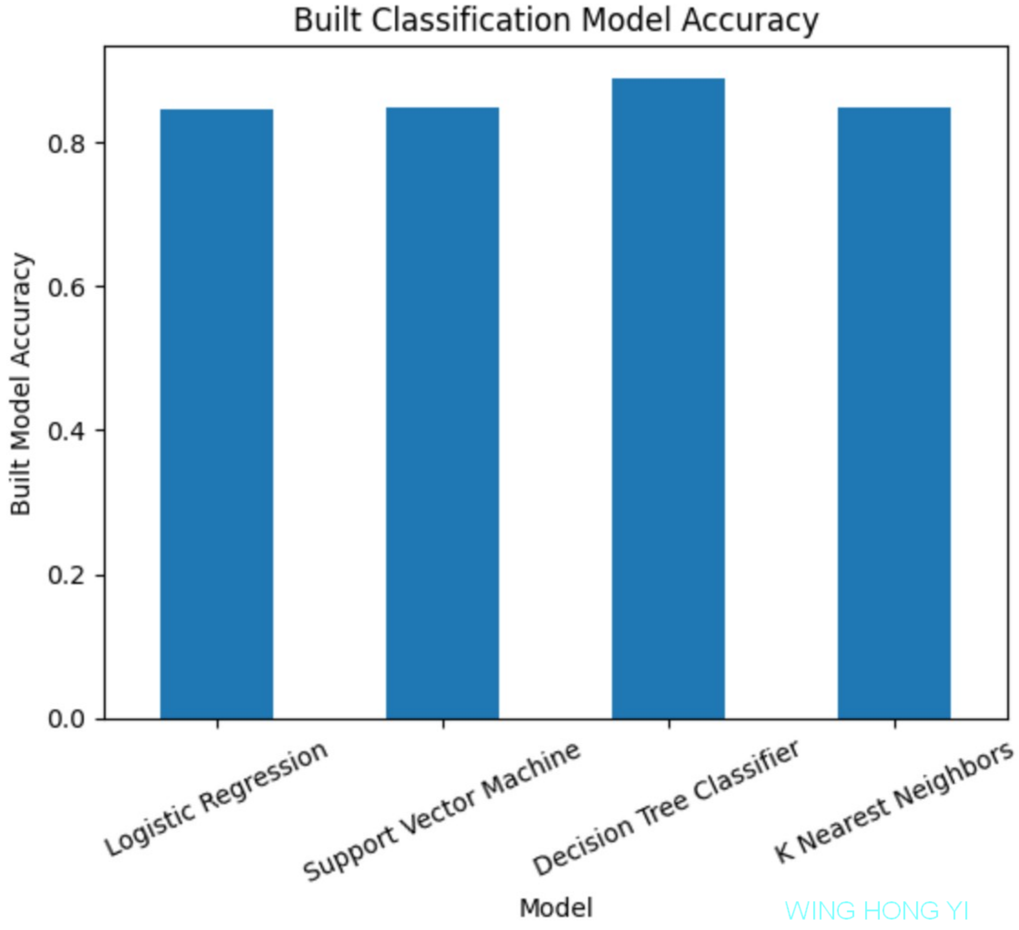


Section 5

Predictive Analysis (Classification)

Classification Accuracy

- The bar chart shows the best_score_ accuracy for each model after using GridSearchCV to search for the best parameters.
- The Decision Tree Classifier model has the highest classification accuracy among the four.

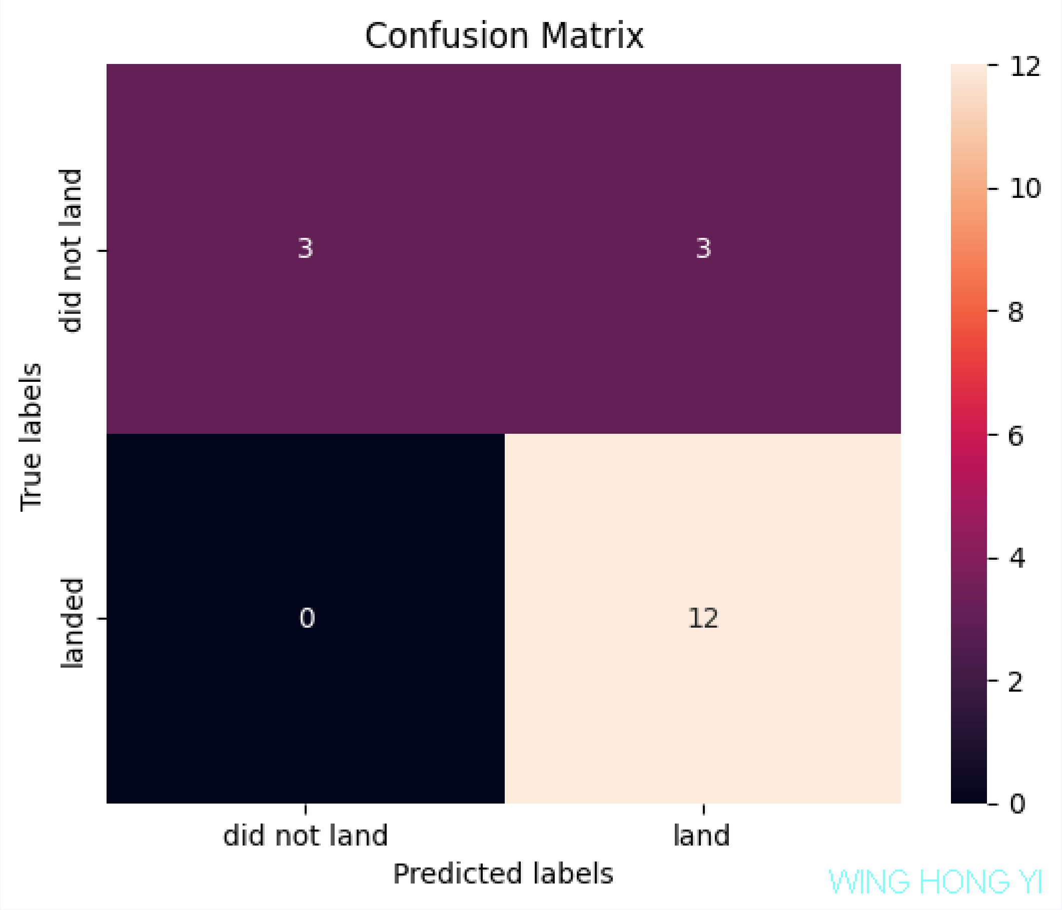


WING HONG YI

Confusion Matrix

- When testing each model with the test data the results were actually the same for each model. So while Decision Tree Classifier had the best accuracy score, the other models also have the same confusion matrix.

- The models correctly predicted 12 successful landings and 3 unsuccessful landings (upper left). There were 0 false negatives, but 3 false positives (upper right).



Conclusions

- The rate of successful launch outcomes increased over time.
- Orbit types ES-L1, GEO, HEO, and SSO had the highest success rate.
- Launch sites are near the ocean and away from cities and major highways for safety reasons.
- Site KSC LC-39A had the highest success rate.
- The 2000 to 4000 kg payload mass range had the best success rate.
- The FT booster version had the best booster success rate.
- We found the Decision Tree Classifier model to have the highest built model accuracy score, though the other 3 models tried were not far behind. They also all performed identically when used with the testing data.

Appendix

- A month name column was created in the Jupyter notebook using the following:

```
In [18]: # Add a column of month names
         %sql ALTER TABLE SPACEXTABLE ADD COLUMN month_name TEXT

          * sqlite:///my_data1.db
         Done.

Out[18]:  []
```



```
In [19]: %%sql UPDATE SPACEXTABLE
         SET month_name = CASE
             WHEN substr(Date, 6,2) = '01' THEN 'Jan'
             WHEN substr(Date, 6,2) = '02' THEN 'Feb'
             WHEN substr(Date, 6,2) = '03' THEN 'Mar'
             WHEN substr(Date, 6,2) = '04' THEN 'Apr'
             WHEN substr(Date, 6,2) = '05' THEN 'May'
             WHEN substr(Date, 6,2) = '06' THEN 'Jun'
             WHEN substr(Date, 6,2) = '07' THEN 'Jul'
             WHEN substr(Date, 6,2) = '08' THEN 'Aug'
             WHEN substr(Date, 6,2) = '09' THEN 'Sep'
             WHEN substr(Date, 6,2) = '10' THEN 'Oct'
             WHEN substr(Date, 6,2) = '11' THEN 'Nov'
             WHEN substr(Date, 6,2) = '12' THEN 'Dec'
         END;
```

Appendix

- The data to make the classification built model accuracy bar chart was assembled with the following Python code:

```
accuracy = [0.8464285714285713, 0.8482142857142856, 0.8892857142857142,
          0.8482142857142858]

model = ['Logistic Regression', 'Support Vector Machine', 'Decision Tree Classifier',
        'K Nearest Neighbors']

modelData = pd.Series(data=accuracy, index=model)
```

Thank you!