IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Percy Dancourt
16/07/2025

Outline

Executive Summary

- Summary of methodologies
  - Data collection (API and Web Scrapping)
  - Data Wrangling
  - EDA (Visualizations and SQL)
  - Folium Interactive Visuals
  - Predictive Analysis ML
- Summary of all results
  - EDA
  - Visuals
  - Predictive Analysis ML

Introduction

- Project background and context
    - Landing the first stages of the Falcon 9 rocket provides a significant competitive advantage. It greatly reduces costs and increases the frequency of launches. SpaceX is the only company capable of doing this, which has contributed to its success
- Problems you want to find answers
    - Determine if the rocket will land successfully

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected using an API and web scrapping tables from Wikipedia.
- Perform data wrangling
  - We applied one-hot encoding to categorical variables.
- Perform exploratory data analysis (EDA) using visualization and SQL
  - We used EDA in order to understand the data.
- Perform interactive visual analytics using Folium and Plotly Dash
  - We also used interactive visuals as a form of EDA to get a better understanding
- Perform predictive analysis using classification models
  - We used different algorithms to choose the one with best results.

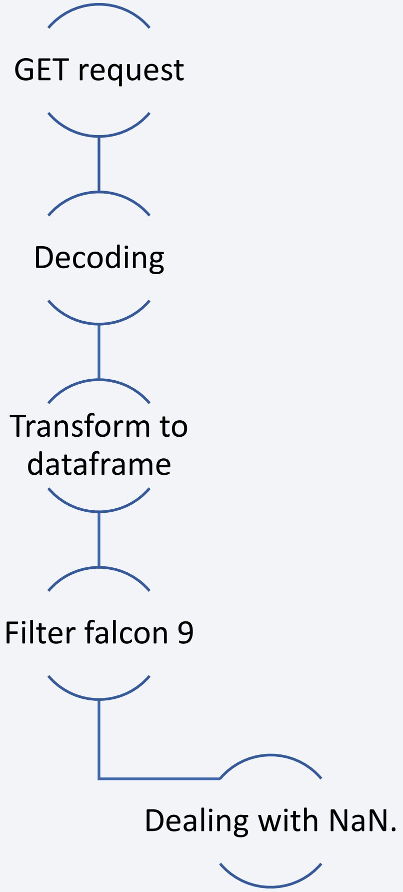
Data Collection

- Describe how data sets were collected.
  - We collected the data using an API and web scrapping from Wikipedia tables
- You need to present your data collection process use key phrases and flowcharts
  - API
    - GET request → Decoding → Transform to dataframe → Filter falcon 9 → dealing with NaN.
  - Web scraping
    - Request from URL → Create BS Object → Extract columns → Create dataframe

Data Collection – SpaceX API

- We used de API to get Space X data, after decoding the JSON we transformed it into a data frame and then filtered it, finally we dealt with the Nan.

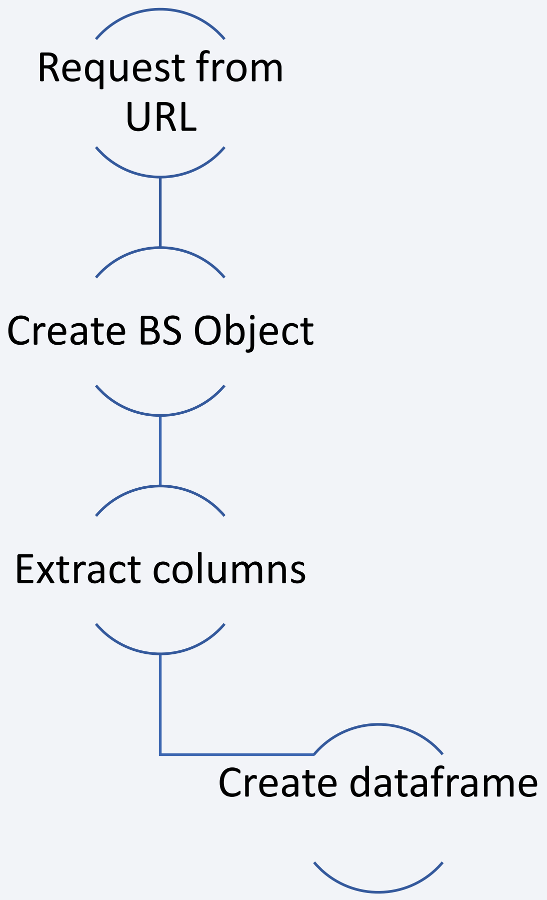
- Add the GitHub URL of the completed SpaceX API calls notebook (must include completed code cell and outcome cell), as an external reference and peer-review purpose



Data Collection - Scraping

- We parsed the table from Wikipedia previously scraped from the URL, then we transformed it to a data frame.

- Add the GitHub URL of the completed web scraping notebook, as an external reference and peer-review purpose



Data Wrangling

1. We used value_counts() to analyze the data
2. We analyzed the missing values.
3. We created a target feature.
4. Add the GitHub URL of your completed data wrangling related notebooks, as an external reference and peer-review purpose

EDA with Data Visualization

- We used visualizations like scatterplots to understand the relationship between different features, this gave us an idea of which features where important and could be useful for predictive analysis.
- Add the GitHub URL of your completed EDA with data visualization notebook, as an external reference and peer-review purpose

EDA with SQL

- Unique launch sites
- Total payload mass carried by Nasa CRS
- Average payload by F9
- Amount of Succes and Failure
- count of landing outcomes ount of landing outcomes

- Add the GitHub URL of your completed EDA with SQL notebook, as an external reference and peer-review purpose

Build an Interactive Map with Folium

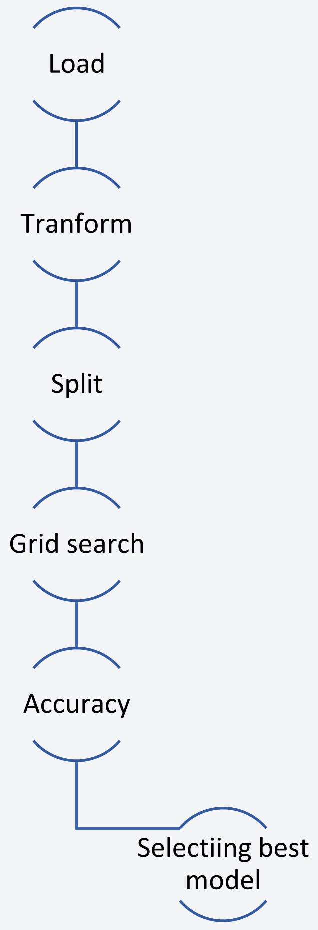
- Marked all launch sites
- Added map objects such as markers to determine the success or failure
- We identified sites with more success
- We identified what was near and far from the launching sites
- Add the GitHub URL of your completed interactive map with Folium map, as an external reference and peer-review purpose

Build a Dashboard with Plotly Dash

- Interactive dashboard using the Plotly Dash dataframe
- Pie chart with total launches filtered by sites
- Scatter plot showing the relationship between Payload and Outcome
- We added this graphs because this are important features to understand the success and failures.
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

Predictive Analysis (Classification)

- We loaded the dataframe, transformed the data in order for different algorithms to understand it, we used train test split to split the training, testing and evaluation data, and the we used grid search on differente algorithms to compare them and choose the best one.
- Add the GitHub URL of your completed predictive analysis lab, as an external reference and peer-review purpose



Results

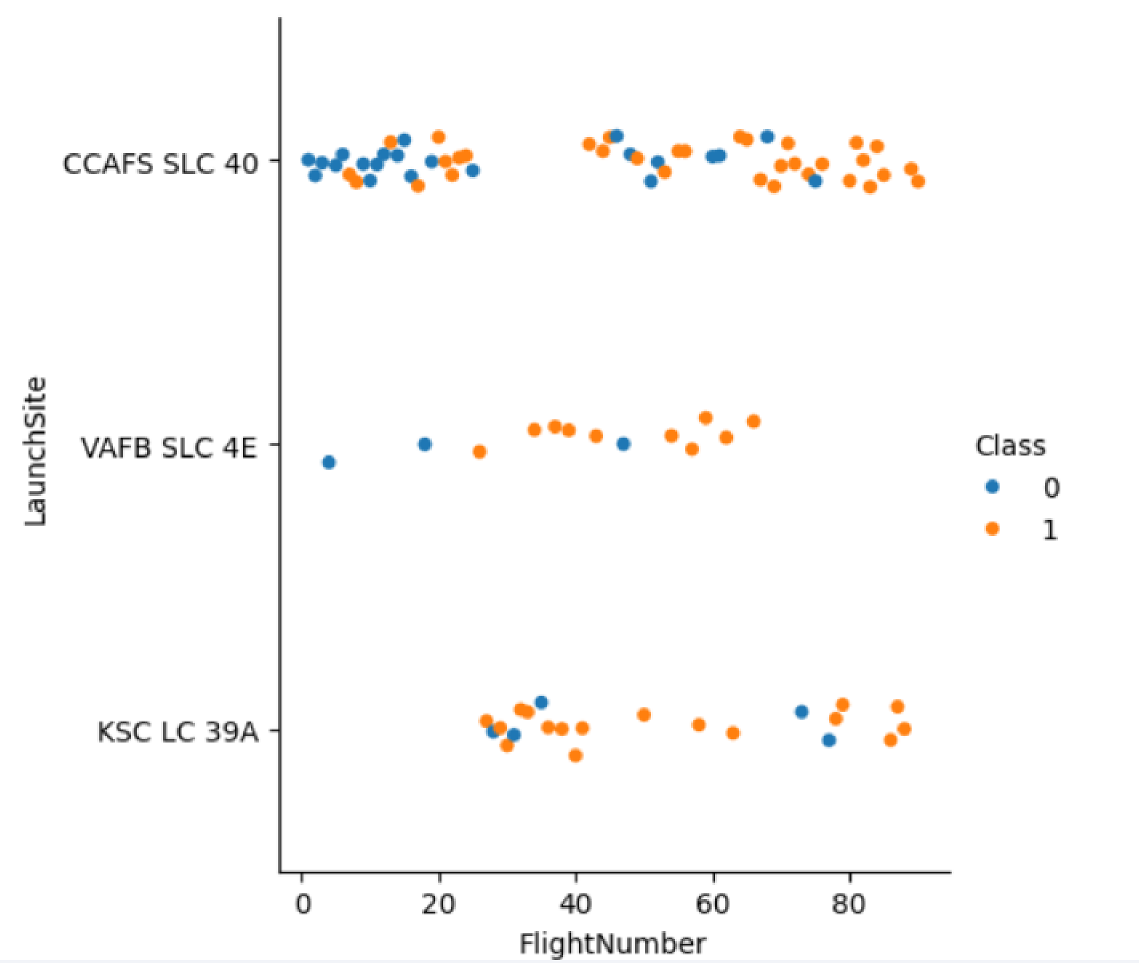
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

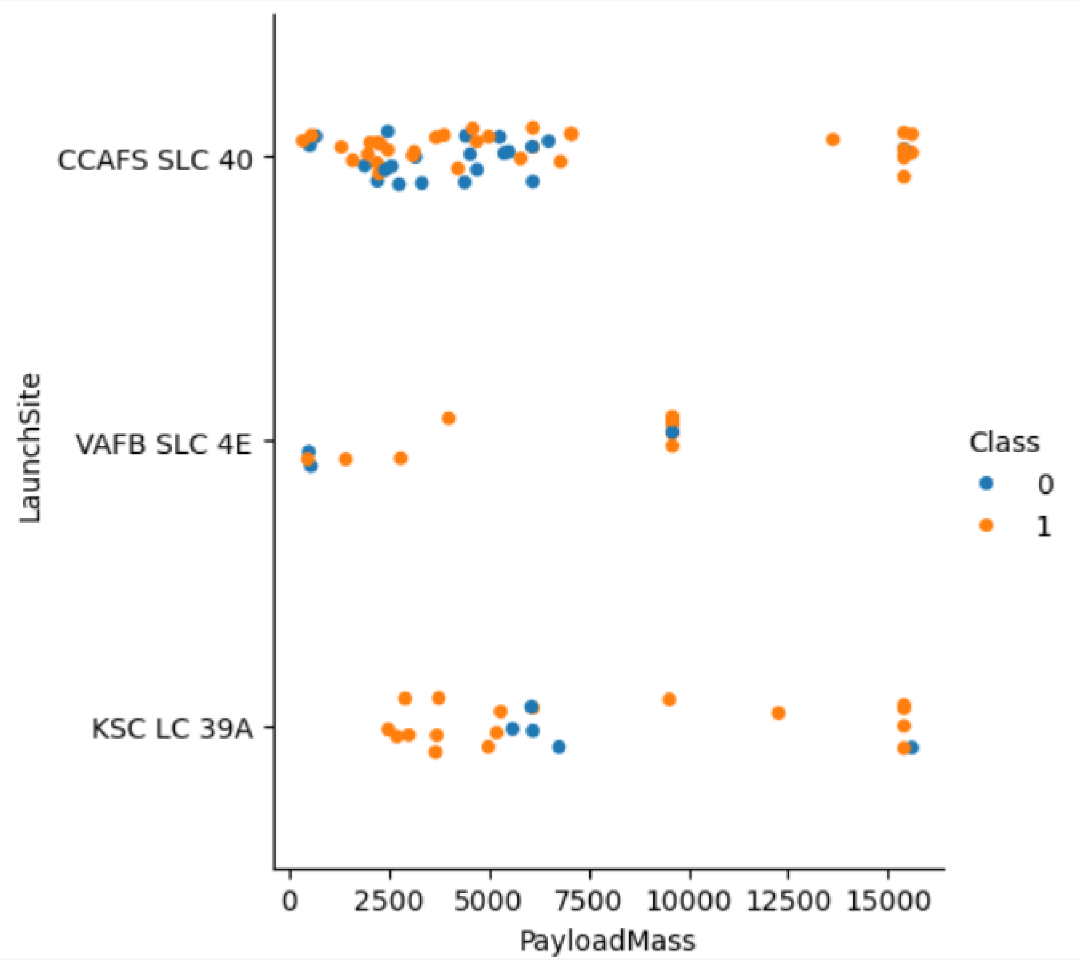
Flight Number vs. Launch Site

- CCAFS SLC 40 has more launches with mixed outcomes; other sites show fewer launches, mostly successful.



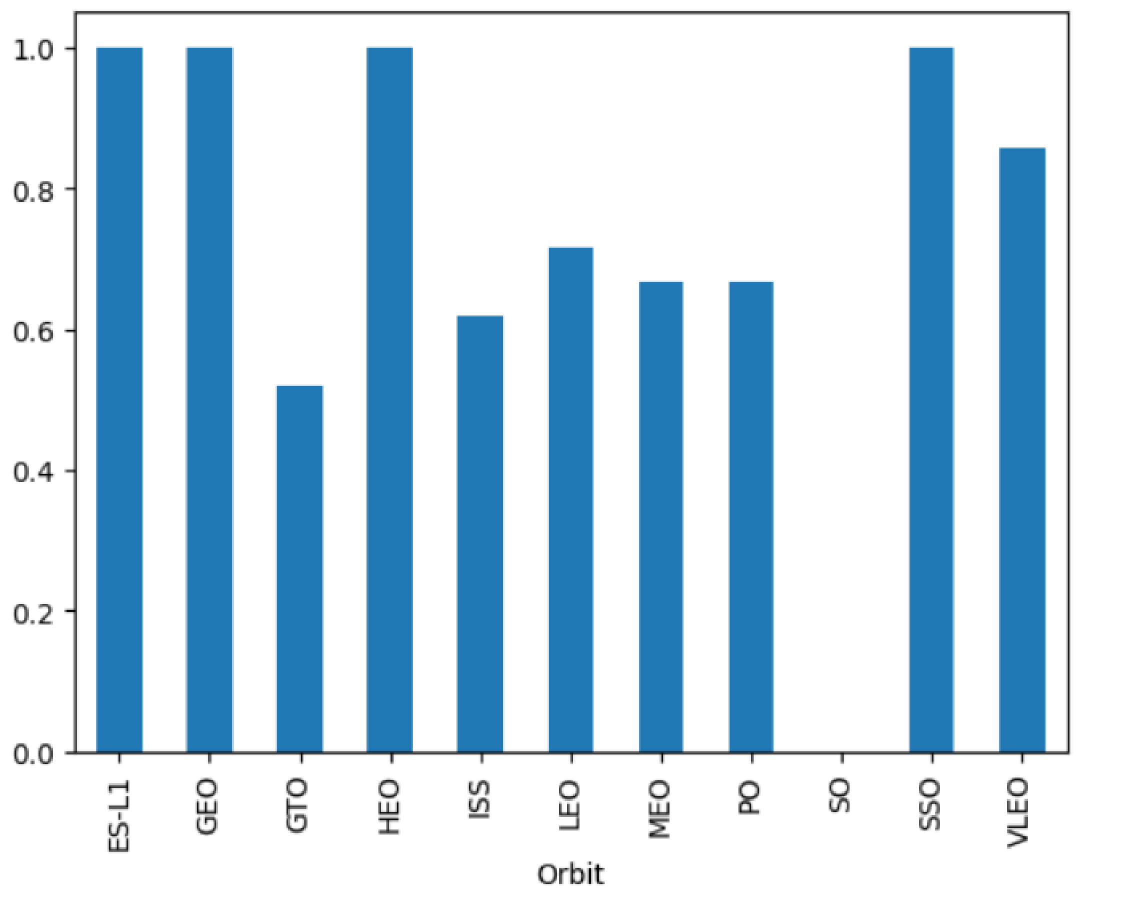
Payload vs. Launch Site

- Higher payloads are mostly launched from KSC LC 39A and CCAFS SLC 40, with a higher success rate at all sites.



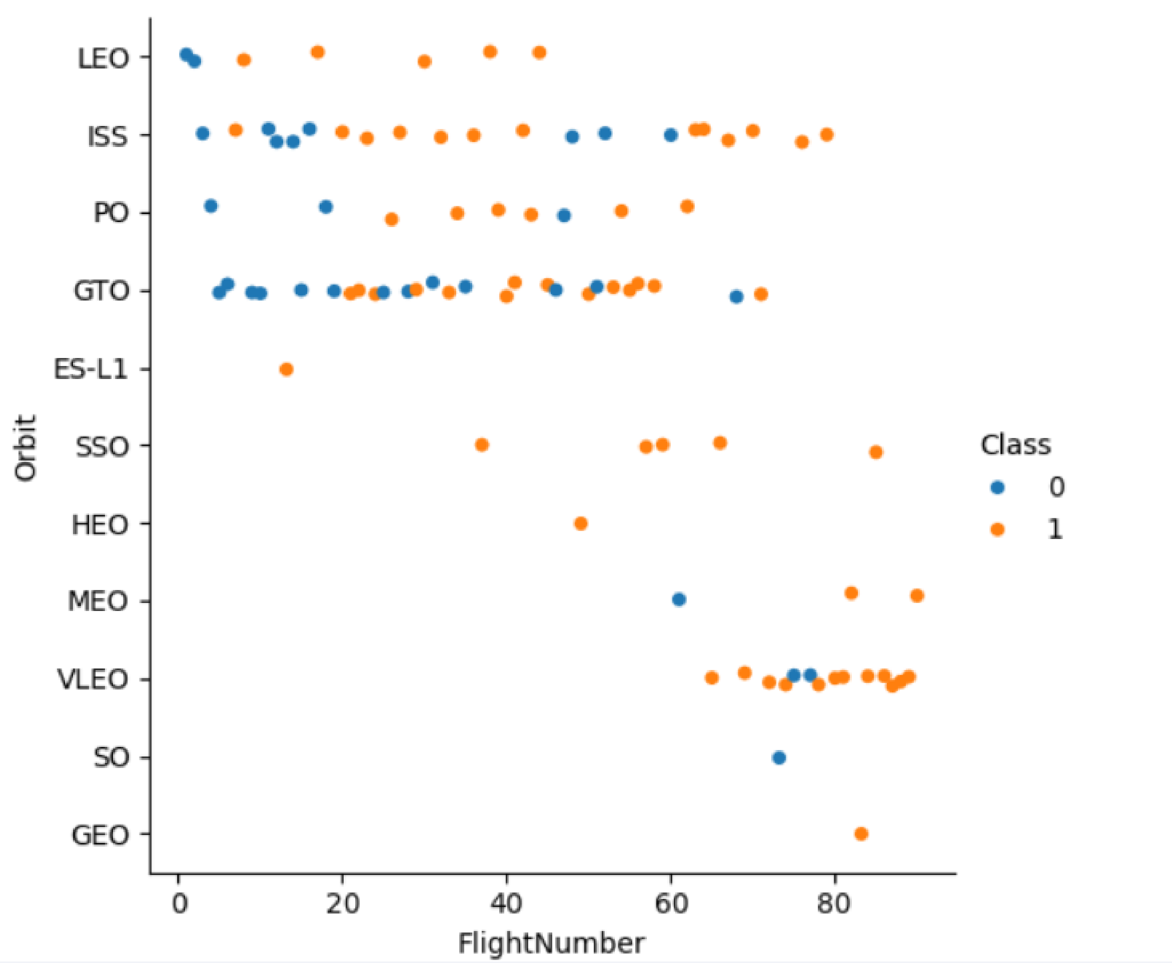
Success Rate vs. Orbit Type

- We can see ES-L1, GEO, HEO, SSO had the most succes rate



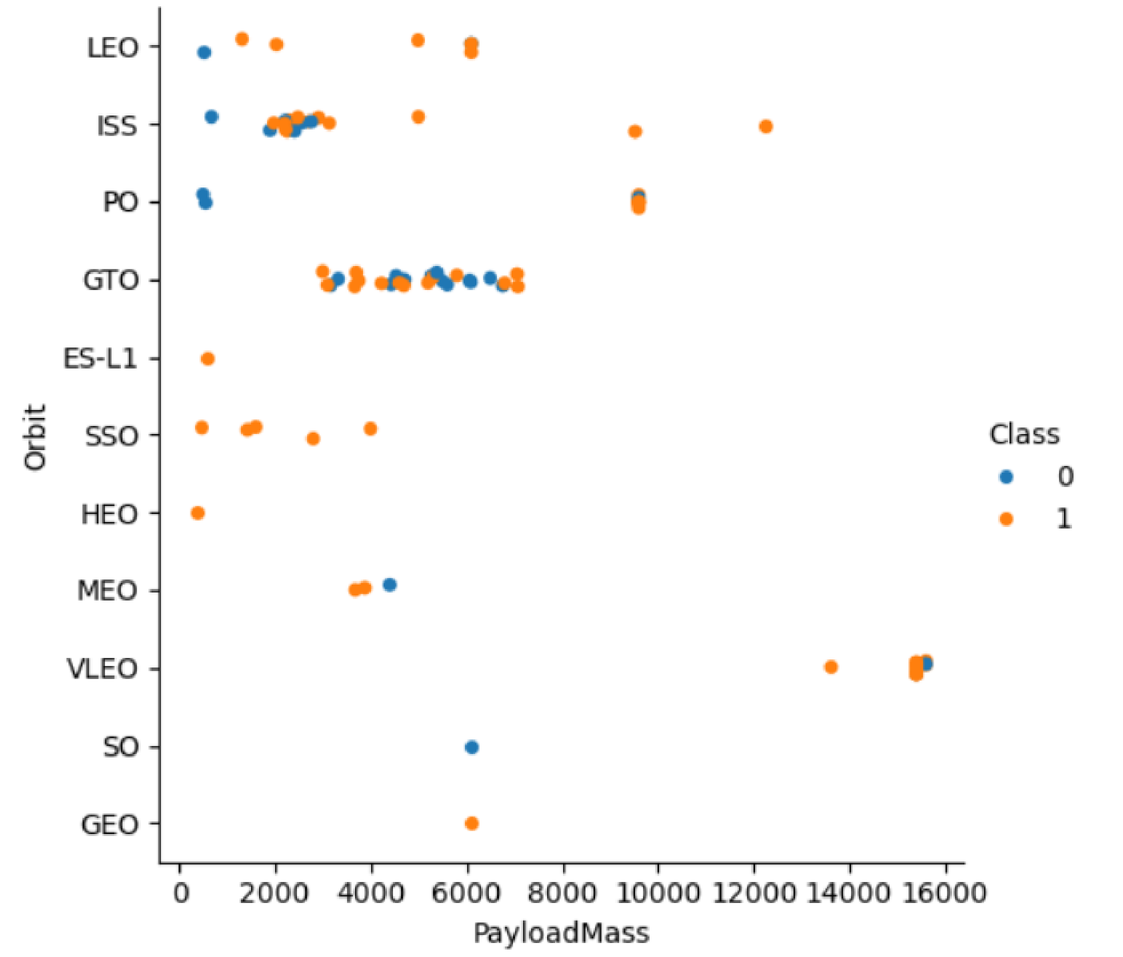
Flight Number vs. Orbit Type

- In LEO, more flights tend to mean more success. In GTO, flight number doesn't seem to affect success.



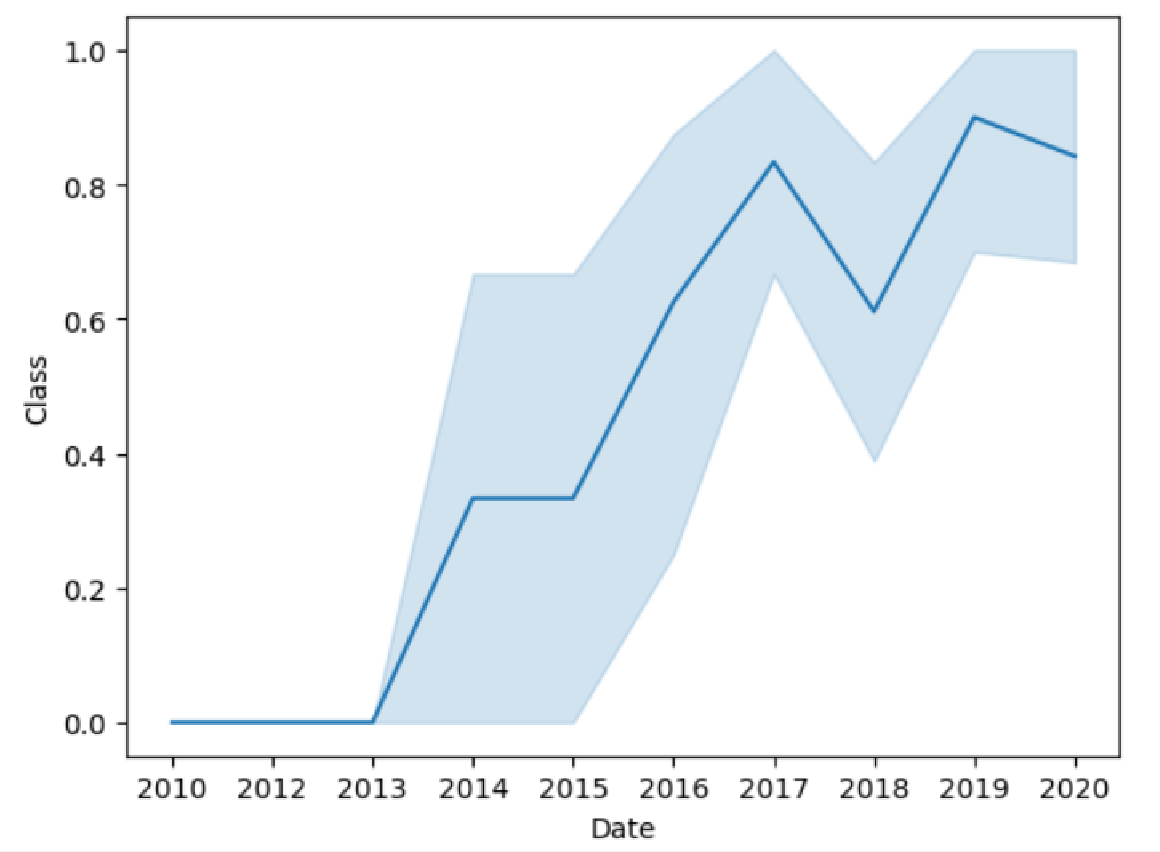
Payload vs. Orbit Type

- Heavy payloads usually land successfully in Polar, LEO, and ISS orbits. In GTO, both success and failure happen often, so it's harder to tell a pattern.



Launch Success Yearly Trend

- Since 2013 succes rates went up but then in 2018 succes rates drop. Finally in 2019 succes rate spiked back.



All Launch Site Names

- This SQL query selects all unique launch sites from the table named SPACEXTABLE

```
%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'

- This query retrieves up to 5 launch site names from SPACEXTABLE that contain "CCA" anywhere in their name

```
%sql SELECT Launch_Site FROM SPACEXTABLE WHERE Launch_Site LIKE "%CCA%" LIMIT 5
```

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

Launch_Site
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40

Total Payload Mass

- This query calculates the total payload mass (in kg) from all records in the SPACEXTABLE

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTABLE
```

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

SUM(PAYLOAD_MASS__KG_)
619967

Average Payload Mass by F9 v1.1

- This query returns the average payload mass (in kg) from the SPACEXTABLE

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTABLE
```

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

AVG(PAYLOAD_MASS__KG_)
6138.287128712871

First Successful Ground Landing Date

- This query finds the earliest date of a launch where the landing was successful, by searching for entries with `'Succes'` in the `Landing_Outcome` column, sorted in ascending order.

```
%sql SELECT DATE FROM SPACEXTABLE WHERE Landing_Outcome LIKE '%Succes%' ORDER BY Date ASC LIMIT 1
```

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

Date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- This query retrieves all unique booster versions from SPACEXTABLE where:
  - The landing was successful on a drone ship, and
  - The payload mass was between 4000 and 6000 kg.
- %sql SELECT DISTINCT Booster_Version FROM SPACEXTABLE WHERE Landing_Outcome = 'Success (drone ship)' AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000

```
%sql SELECT DISTINCT Booster_Version FROM SPACEXTABLE WHERE Landing_Outcome = 'Success (drone ship)' AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_
```

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

- This query counts how many times each mission outcome appears in `SPACEXTABLE`, showing the total number of missions per outcome.

```
%sql SELECT Mission_Outcome, COUNT(*) as Total FROM SPACEXTABLE GROUP BY Mission_Outcome
```

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

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

Boosters Carried Maximum Payload

- This query retrieves the booster version and payload mass for the mission with the heaviest payload in the `SPACEXTABLE`.

```
%%sql
SELECT Booster_Version, Payload_Mass__kg_
FROM SPACEXTABLE
WHERE Payload_Mass__kg_ = (
    SELECT MAX(Payload_Mass__kg_)
    FROM SPACEXTABLE
);
```

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

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

2015 Launch Records

- This query returns the month, landing outcome, booster version, and launch site for all missions in 2015 where the landing failed on a drone ship.

```
%%sql
SELECT
  SUBSTR(Date, 6, 2) AS Month,
  Landing_Outcome,
  Booster_Version,
  Launch_Site
FROM SPACEXTABLE
WHERE
  Landing_Outcome LIKE 'Failure (drone ship)%'
  AND SUBSTR(Date, 1, 4) = '2015';
```

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

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

- This query counts how many times each landing outcome occurred between June 4, 2010 and March 20, 2017, then lists them from most to least frequent.

```
%%sql
SELECT
  Landing_Outcome,
  COUNT(*) AS Outcome_Count
FROM SPACEXTABLE
WHERE Date BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY Landing_Outcome
ORDER BY Outcome_Count DESC;
```

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

Landing_Outcome	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

Launching sites

- It's important to analyze where the launchings are taking place, this could lead to meaningful insights.

<Folium Map Screenshot 2>

- Replace <Folium map screenshot 2> title with an appropriate title
- Explore the folium map and make a proper screenshot to show the color-labeled launch outcomes on the map
- Explain the important elements and findings on the screenshot

<Folium Map Screenshot 3>

- Replace <Folium map screenshot 3> title with an appropriate title
- Explore the generated folium map and show the screenshot of a selected launch site to its proximities such as railway, highway, coastline, with distance calculated and displayed
- Explain the important elements and findings on the screenshot

Section 4

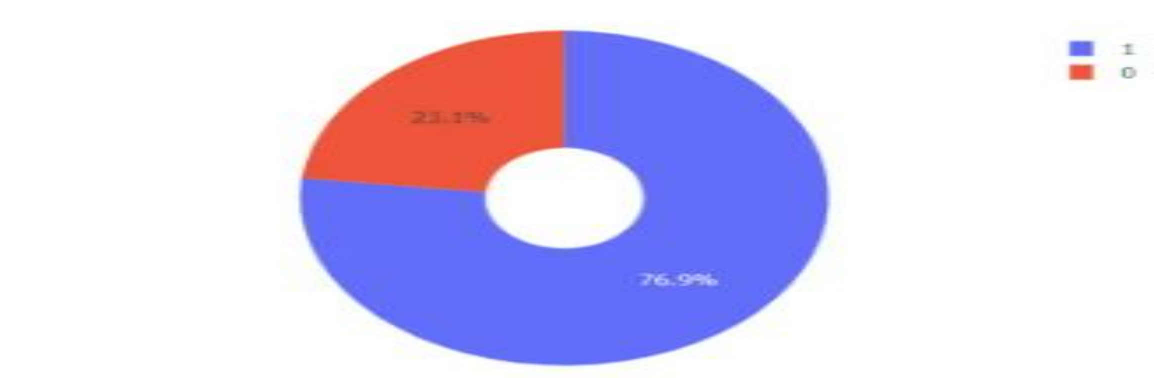
Build a Dashboard with Plotly Dash

<Dashboard Screenshot 1>

- KSC LC-39A is the most successful

<Dashboard Screenshot 2>

- 76.9% success rate
- 23.1% failure.



<Dashboard Screenshot 3>

- Replace <Dashboard screenshot 3> title with an appropriate title
- Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider
- Explain the important elements and findings on the screenshot, such as which payload range or booster version have the largest success rate, etc.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

- All models have the same accuracy

```
accuracy_test = logreg_cv.score(X_test, Y_test)
print("Test accuracy:", accuracy_test)
```

```
Test accuracy: 0.8333333333333334
```

```
accuracy_test = svm_cv.score(X_test, Y_test)
print("Test accuracy:", accuracy_test)
```

```
Test accuracy: 0.8333333333333334
```

```
print("Test accuracy:", tree_cv.score(X_test, Y_test))
```

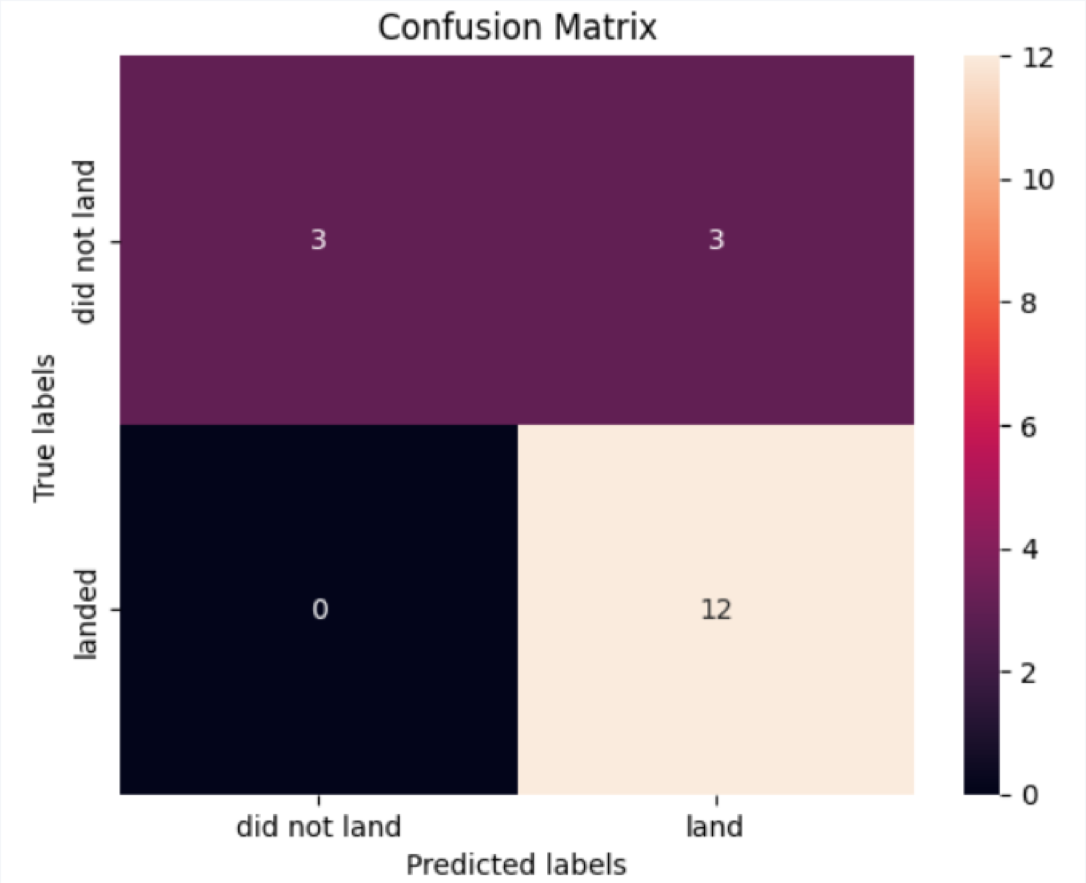
```
Test accuracy: 0.8333333333333334
```

```
print("Test accuracy:", knn_cv.score(X_test, Y_test))
```

```
Test accuracy: 0.8333333333333334
```

Confusion Matrix

- Model can classify correctly, it has 3 false positives



Conclusions

- Launch success since 2013
- Higher payloads are mostly launched from KSC LC 39A and CCAFS SLC 40
- Heavy payloads usually land successfully in Polar, LEO, and ISS orbits. In GTO, both success and failure happen often, so it's harder to tell a pattern.
- We can see ES-L1, GEO, HEO, SSO had the most succes rate

Appendix

Thank you!