IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Dr Sang'udi Sang'udi
31 January 2015

Outline

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

Executive Summary

- Summary of methodologies
  - Extracted a Falcon 9 launch records HTML table from Wikipedia
  - Parsed the table and convert it into a Pandas data frame
  - Performed EDA
  - Visualized the launching sites
  - Performed Predictive Analytics

- Summary of all results

  - Four Launching sites

  - The first successful launch was 2015-12-22'

  - Total number of successful mission was 60 out of 90 mission carried out

  - SVM performed better than other models

Introduction

- SpaceX offers Falcon 9 launches at **$62 million**, while competitors charge **$165 million**.
- The **key cost reduction** comes from reusing the **first-stage booster**.
- Predicting whether the first stage **lands successfully** can help estimate **launch costs**.
- This insight can be valuable for **competitors** bidding against SpaceX.
- 🔍 **Problems We Want to Solve:**
- Can we accurately predict whether **Falcon 9's first stage** will land successfully?
- What **factors influence** a successful landing?
- How can this prediction help in **cost estimation** and **market competition**?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
- Extracted a Falcon 9 launch records HTML table from Wikipedia
- Parsed the table and convert it into a Pandas data frame
- Perform data wrangling
    - Missing values was replaced by mean
- Performed exploratory data analysis (EDA) using visualization and SQL
- Performed interactive visual analytics using Folium and Plotly Dash
- Performed predictive analysis using classification models
    - KNN, Logistic regression, Decision tree, SVM

Data Collection

- Web scrap Falcon 9 launch records with `BeautifulSoup`:
- Extracted a Falcon 9 launch records HTML table from Wikipedia
- Parsed the table and convert it into a Pandas data frame

Data Collection – SpaceX API

- SpaceX API was assessed via python Library requests
- Relevant fields, including **launch dates, rocket types, payload masses, and landing outcomes**, were extracted
- As shown in the github link
- https://github.com/SanStart/IBM-Assignment/blob/main/jupyter-labs-spacex-data-collection-api-v2.ipynb

Start → Send API Request → Receive JSON Response → Extract Relevant Fields → Store in DataFrame → Save for Analysis

Data Collection - Scraping

- Ŵêč sçsắř Gắľçôŋ ľăuŋçȟ şêçôşđş xîƫȟ BêăuƫîğuľŞôuř
- Éyƫsắçƫ ă Gắľçôŋ ľắuŋçȟ şêçôşđs ȞȚŇĽ ƫăčľê ğsôŋ Ŵîlîřêđîă
- Rắşşê ƫhê ƫắčľệ ắŋđ çôŋŵêşƫ îƫ îŋƫô ă Rãŋđaş đaƫá ğsaŋê
- https://github.com/SanStart/IBM-Assignment/blob/main/jupyter-labs-webscraping.ipynb

Start → Send HTTP Request → Retrieve Web Page → Parse HTML → Extract Data → Store in DataFrame → Save for Analysis

Data Wrangling

- Retained only Falcon 9 type of rockets
- Changed outcome into class
- Replaced missing values

EDA with Data Visualization

- Visualized the following relationships ;
  - Flight Number and Launch Site
  - Payload and Launch Site
  - Success rate and orbit
  - Flight number and Orbit Type
  - PayLoad Mass and Orbit Type
  - Success rate and yearly trend

EDA with SQL

- Performed SQL Query for the following
  - Unique Launch sites
  - Total Payload for NASA (CRS)
  - Average payload mass carried by booster version F9 v1.1
  - Dates for successful landing outcome
  - Boosters which have success in drone ship
  - Successful and Failure mission
  - Booster carrying the maximum payload mass

Build an Interactive Map with Folium

- Used circles of different colors and words for markers
- To enhance better visualization

Build a Dashboard with Plotly Dash

- To be able to show longitude and latitudes in real time

Predictive Analysis (Classification)

Split the data into train and test sets

Assigned four models KNN, Logistic regression, Decision Tree, SVM

Evaluated each model using score accuracy

Results

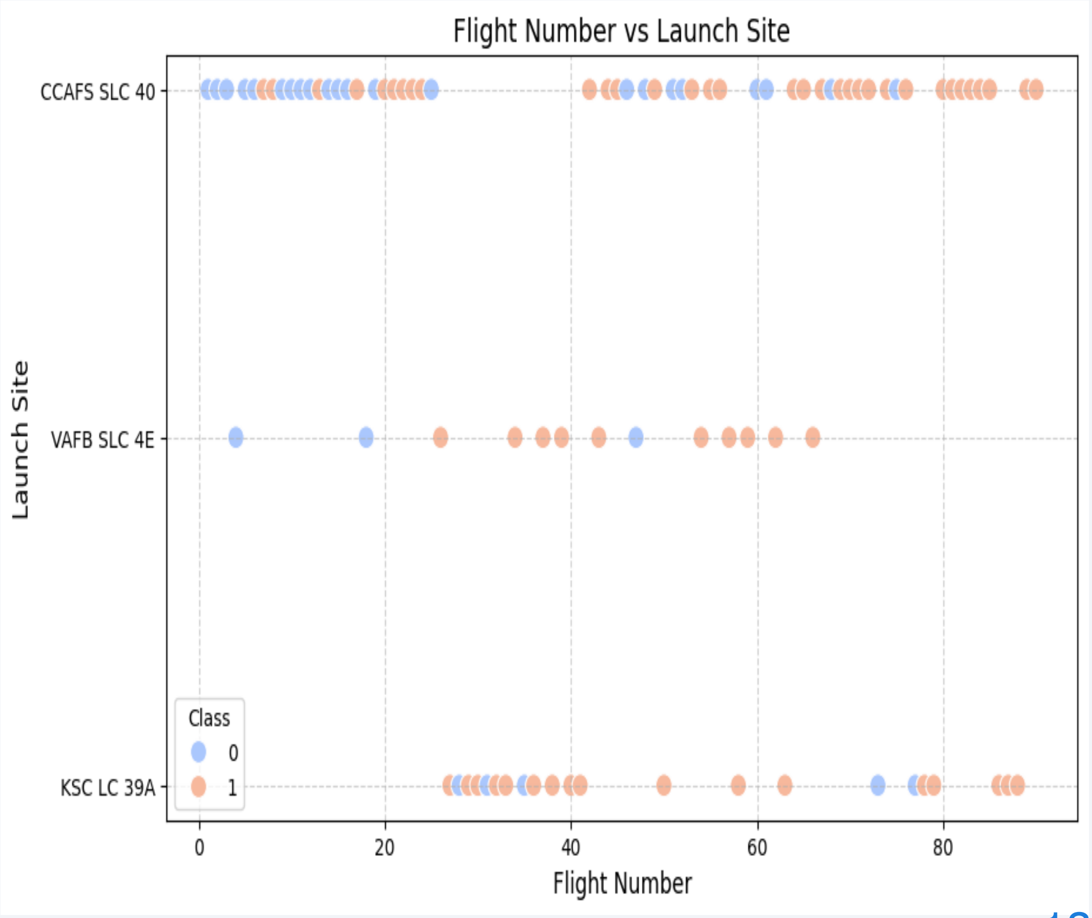
- SVM performed better compared to that models

Section 2

Insights drawn from EDA

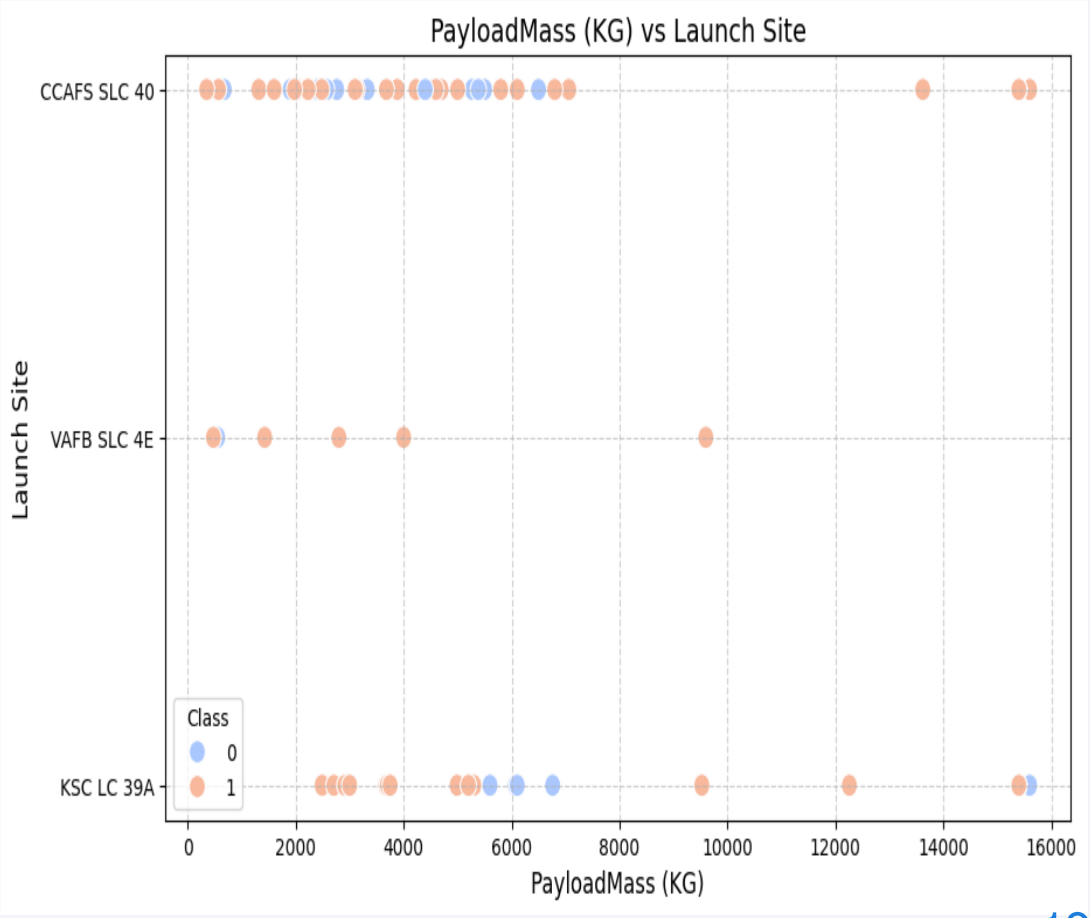
Flight Number vs. Launch Site

- CCAFS SLC 40 had more flight numbers compared to other launch sites
- Historical usage and infrastructures
- Heavy reliance on Falcon 9



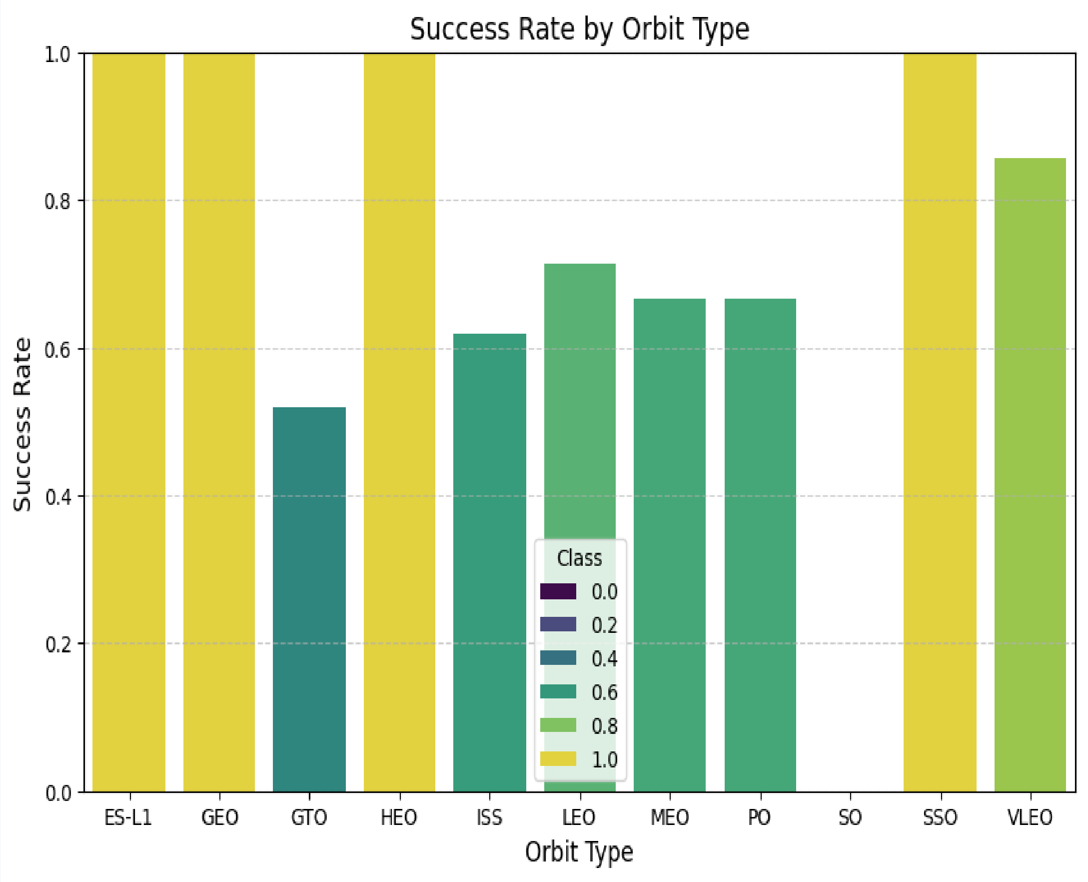
Payload vs. Launch Site

- The VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).



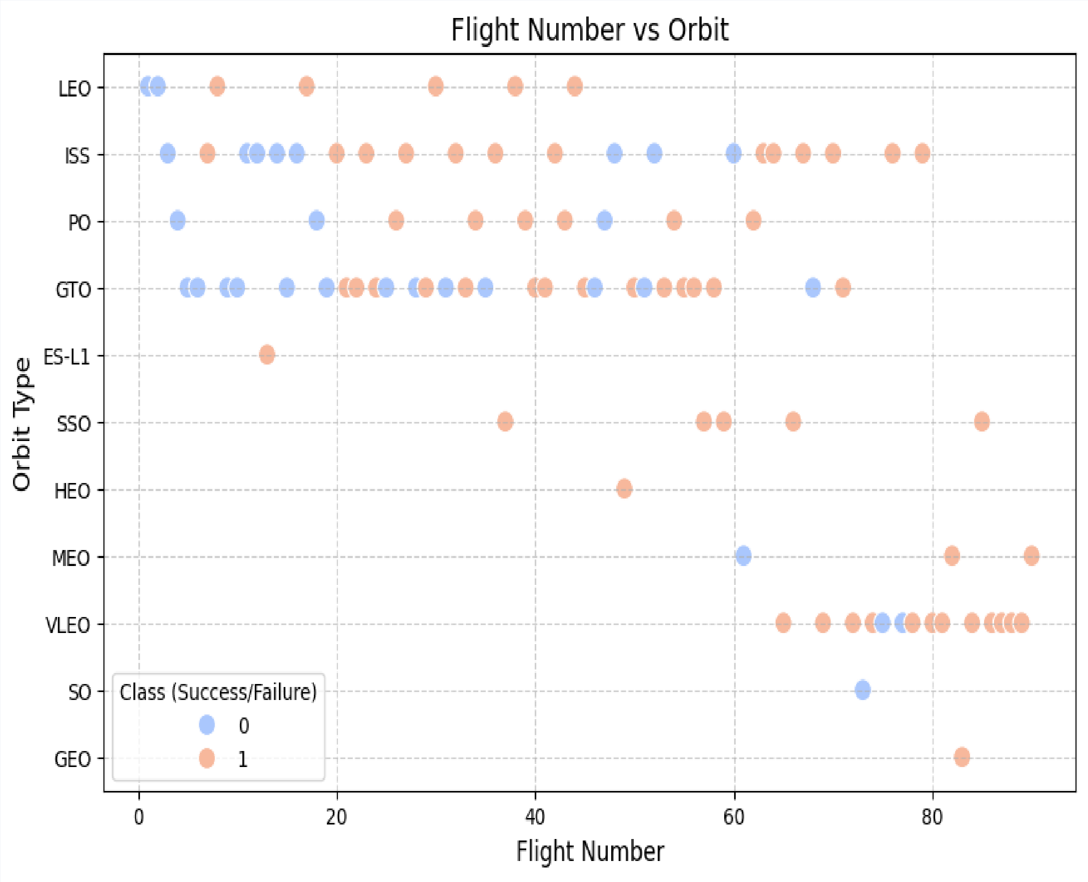
Success Rate vs. Orbit Type

- ES-L1, GEO, HEO, SSO have high success rate of 1



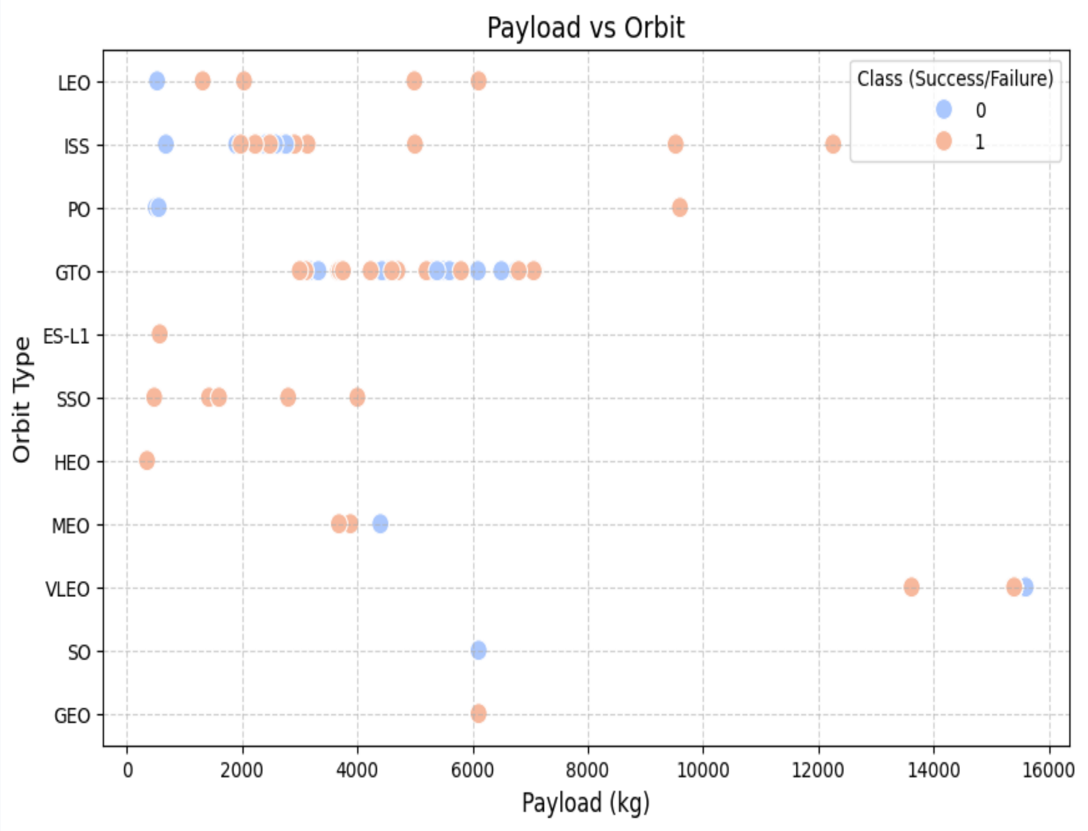
Flight Number vs. Orbit Type

- The LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.
-



Payload vs. Orbit Type

- With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.
- However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.



Launch Success Yearly Trend

- Show a line chart of yearly average success rate

- Show the screenshot of the scatter plot with explanations

All Launch Site Names

- The unique launch sites

  CCAFS LC-40

  VAFB SLC-4E

  KSC LC-39A

  CCAFS SLC-40

- `%sql SELECT DISTINCT Launch_Site FROM SPACEXTABLE;`

Launch Site Names Begin with 'CCA'

- SELECT * FROM launches WHERE Launch_Site LIKE 'CCA%' LIMIT 5
- This is the query

	Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS_KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
0	2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
1	2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of...	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2	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
3	2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
4	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

- The total payload mass carried by boosters launched by NASA (CRS) was 45596 kg
- SELECT SUM(PAYLOAD_MASS__KG_) AS total_payload_mass FROM launches WHERE Customer = 'NASA (CRS)';

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 was 2928.4 KG
- `%sql    SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTABLE WHERE booster_Version = 'F9 v1.1';`

First Successful Ground Landing Date

- The dates of the first successful landing outcome on ground pad was `'2015-12-22'`
- SELECT MIN(Date) AS earliest_date FROM launches WHERE landing_class = 1 AND Outcome = 'True RTLS'

Successful Drone Ship Landing with Payload between 4000 and 6000

- The boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000 were Falcon 9

- SELECT BoosterVersion FROM launches WHERE landing_class = 1 AND Outcome = 'True ASDS' AND PayloadMass > 4000 AND PayloadMass < 6000;

Total Number of Successful and Failure Mission Outcomes

- The total number of successful mission were 60 and failure 30
- SELECT landing_class, COUNT(*) AS count FROM launches GROUP BY landing_class;

Boosters Carried Maximum Payload

- The names of the booster which have carried the maximum payload mass were Falcon 9
- SELECT BoosterVersion FROM launches WHERE PayloadMass = (SELECT MAX(PayloadMass) FROM launches);

2015 Launch Records

- List of the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
- `11 January Falcon 9 CCAFS SLC 40 0`
- `13 April Falcon 9 CCAFS SLC 40 0`

- SELECT strftime('%B', Date) AS month_name, BoosterVersion, LaunchSite, landing_class FROM launches WHERE landing_class = 0 AND Outcome = 'False ASDS' AND strftime('%Y', Date) = '2015';

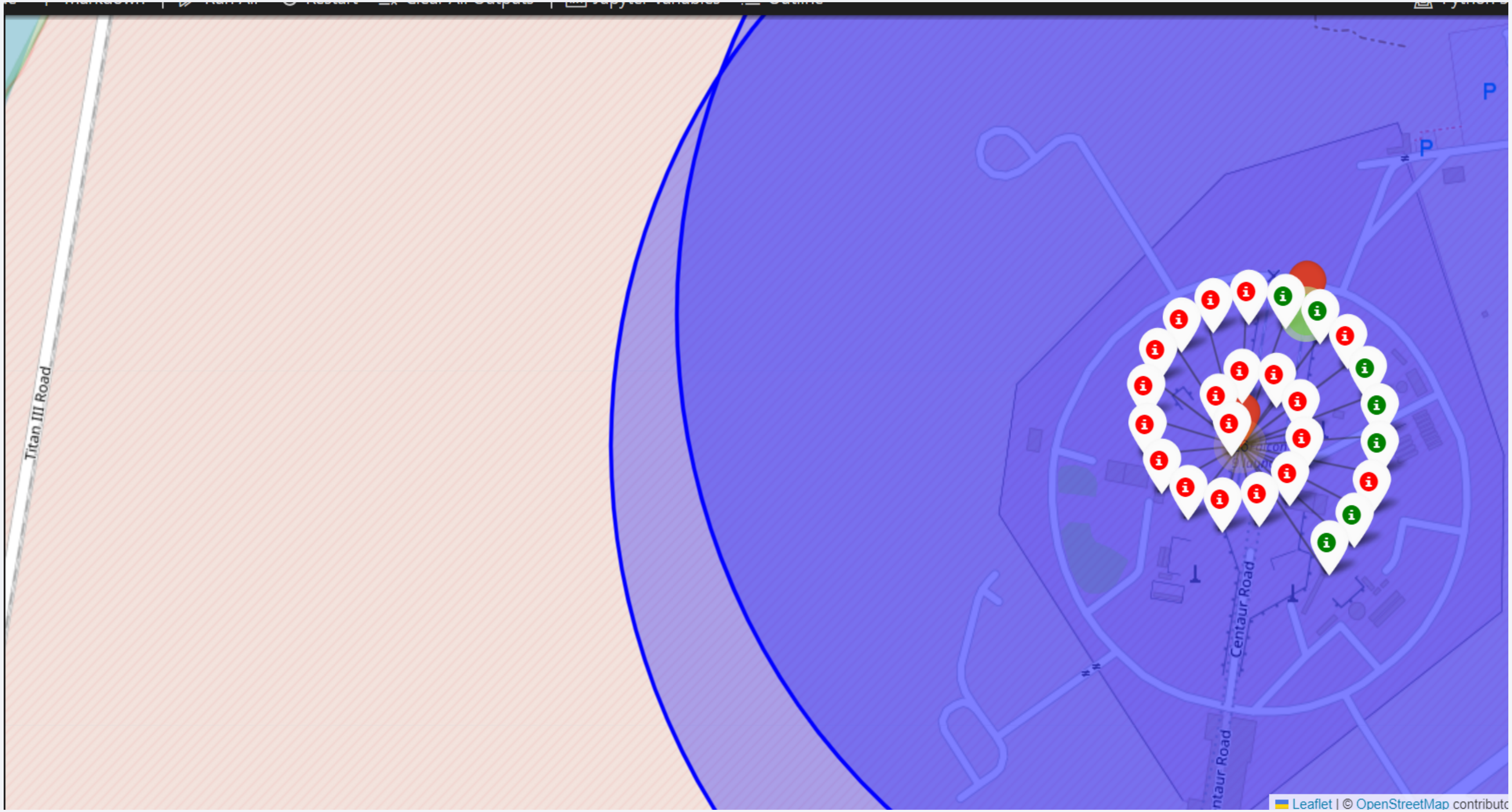
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- 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 success 11, Failure 17
- SELECT landing_class, COUNT(*) AS outcome_count FROM launches WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY landing_class ORDER BY outcome_count DESC;

Section 3

Launch Sites Proximities Analysis

Launch sites



Section 4

Build a Dashboard with Plotly Dash

<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)

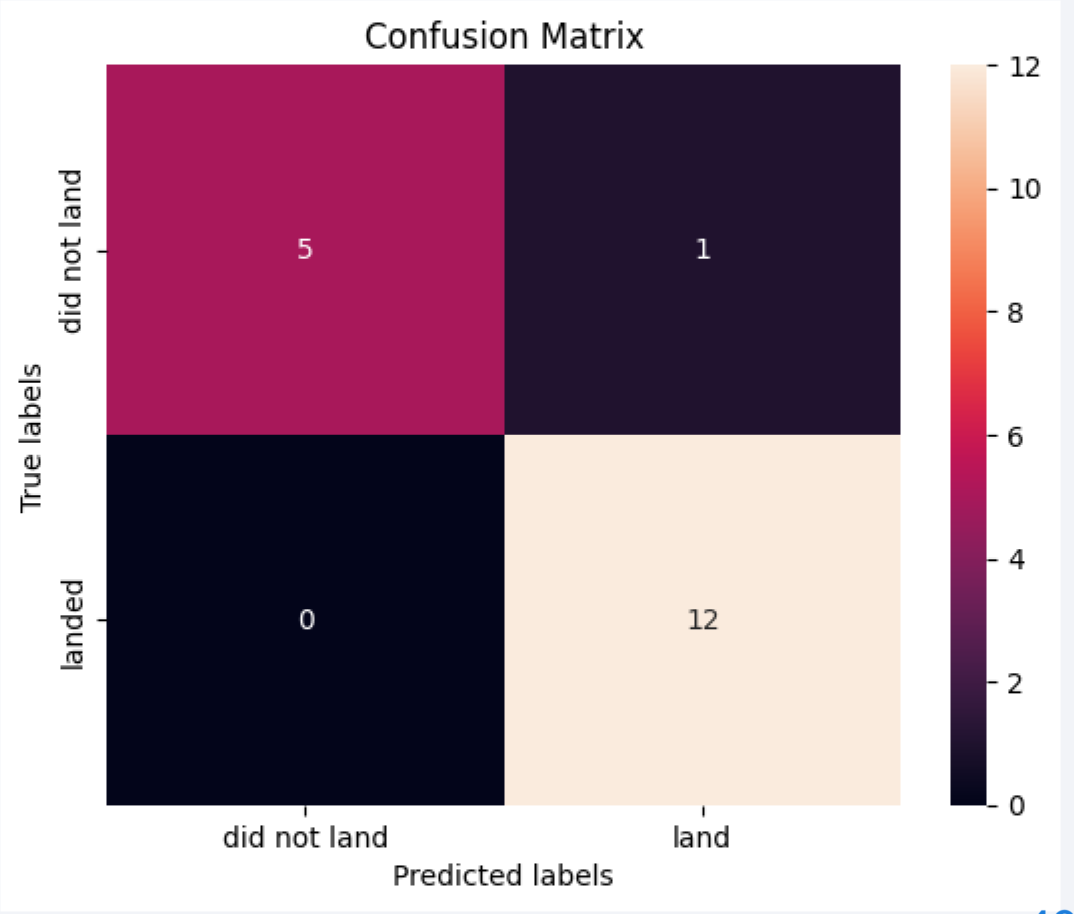
Classification Accuracy

- Visualize the built model accuracy for all built classification models, in a bar chart

- Find which model has the highest classification accuracy

Confusion Matrix

- Show the confusion matrix of the best performing model with an explanation



Conclusions

- Logistic Regression achieved an accuracy of 94.44% on the test set.

- SVM also achieved an accuracy of 94.44% on the test set.

- Both Logistic Regression and SVM showed high accuracy but had some false positives.

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!