IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Zarik Rasool

24 November 2025

Outline

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

Executive Summary

- This project creates a machine learning pipeline to predict if the first stage of the Falcon 9 will land successfully.
- Data is collected via scraping and API, wrangled and prepared for EDA.
- Exploratory data analysis (EDA) is performed to prepare for predictive analysis.
- Predictive analysis is performed to outline the best predictive factors for first stage landing.
- Conclusive insights are provided with recommendations.

Introduction

- SpaceX Falcon 9 launches cost only 62 million USD.
- Other Space exploration companies costs 165+ million USD, nearly three times more.
- SpaceX can afford these prices because of it's first stage reuse.
- If we can determine if first stage will launch, we can determine cost of launch.
- This project uses data science methodology to obtain those insights.
- The insights can be used by competing companies who want to bid against SpaceX for rocket launches.

Section 1

Methodology

Methodology

Executive Summary

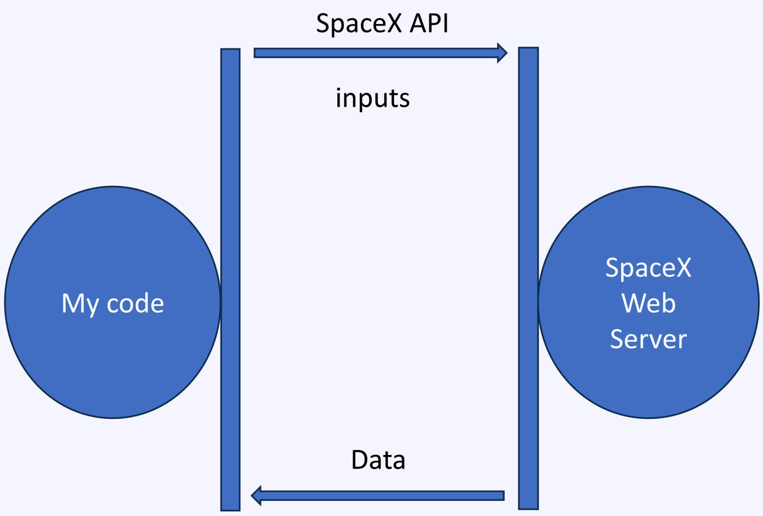
- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

- We collected data using SpaceX REST API and scraping the Wikipedia page.
- Using helper functions and the GET request on the API URL we request and parse the launches data.
- We decode the response content as a JSON object and turn it into a Pandas dataframe.
- Using a HTTP GET method, we obtain SpaceX Launch wikepedia HTML page as HTTP
- We create a beautifulsoup object from the html response and then parsed the launch records tables into a pandas dataframe

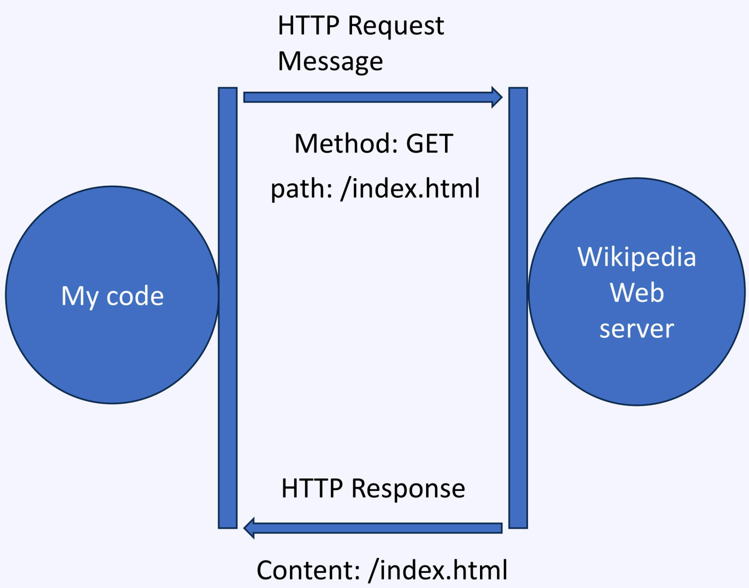
Data Collection – SpaceX API

- Diagram to show API request process
- Full notebook on github: [API Notebook]



Data Collection - Scraping

- Diagram to show scraping process
- Full notebook on github: Webscraping



Data Wrangling

- Data wrangling is a necessary step before exploring any data set.
- We select important features create new dataframe with new columns.
- We then filter dataframe for Falcon9 launches.
- We replace missing values for continuous values like payload mass with mean.
- We check to ensure correct data types.
- Create a landing outcome label for the different scenarios for landing
- Github link available: [Data wrangling notebook]

EDA with Data Visualization

- EDA with visualization was used to scope potential relationships between factors and success class.
- Scatter plots are used to gauge initial relationships such Payload Mass vs Launch Site.
- Bar chart to scope the type class relationships and distribution such as Success class vs orbit type.
- A simple line chart to gauge performance over the years.
- No link for notebook available as it was not downloadable.

EDA with SQL

- We explored the data further by querying the data using SQL
- Filtered for launch site using “distinct” and “like”
- We computed sum and averages values of payload mass
- We filtered and presented data by payload mass
- We counted the total figures for mission outcomes
- We ranked the different types mission outcomes
- Full notebook github link: SQL Notebook

Build an Interactive Map with Folium

- We used folium interactive maps to visualize the positioning of launch sites.
- We implement folium.Circle to visually see the locations of the launch sites.
- Using MarkerCluster to show in colour, successful and failed launches on the map.
- Using PolyLine lines to check proximities to other launch sites and other points of interests such as railways, coatlines etc.
- Link: the link to the full notebook is unavailable as it was not downloadable

Build a Dashboard with Plotly Dash

- We added a success-pie-chart based on a selected site drop down to see which sites have the largest successful launches
- We added a range slider to select payload to be able to gauge which payload ranges have the highest success rate.
- We added a call back function to render the success payload scatter chart to see which booster version had the highest launch success rate
- Link: the link to the full notebook is unavailable as it was not downloadable

Predictive Analysis (Classification)

1. Load SpaceX cleaned and processed Dataframe
2. Assign X,Y labels, create numpy array for outcome Class
3. Standardize data in X by transformation
4. Split data into training and testing data
5. Create model object (logistic regression, SVM etc.) & gridsearchCV to fit training data
6. Obtain parameters and apply to test set. Obtain accuracy score and compare.

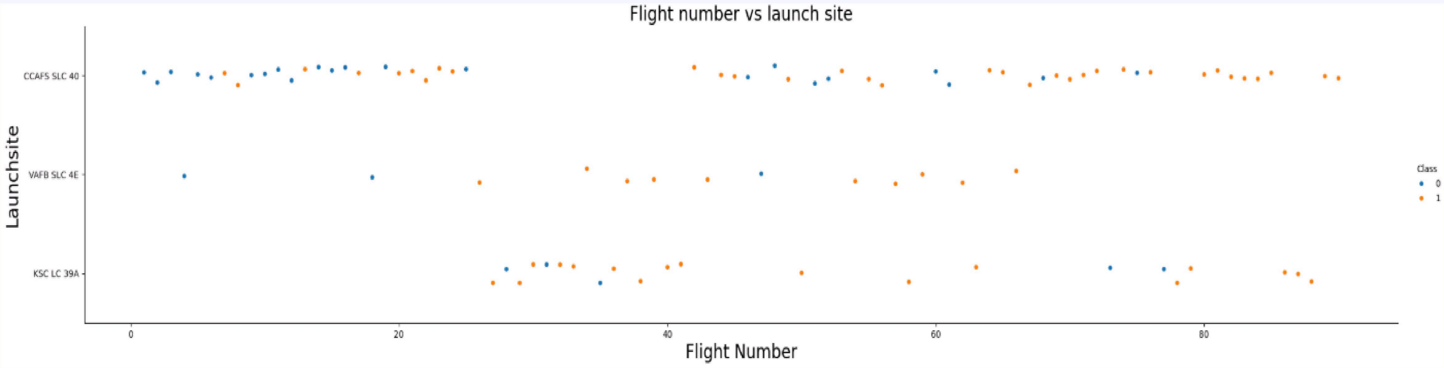
Results

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

Section 2

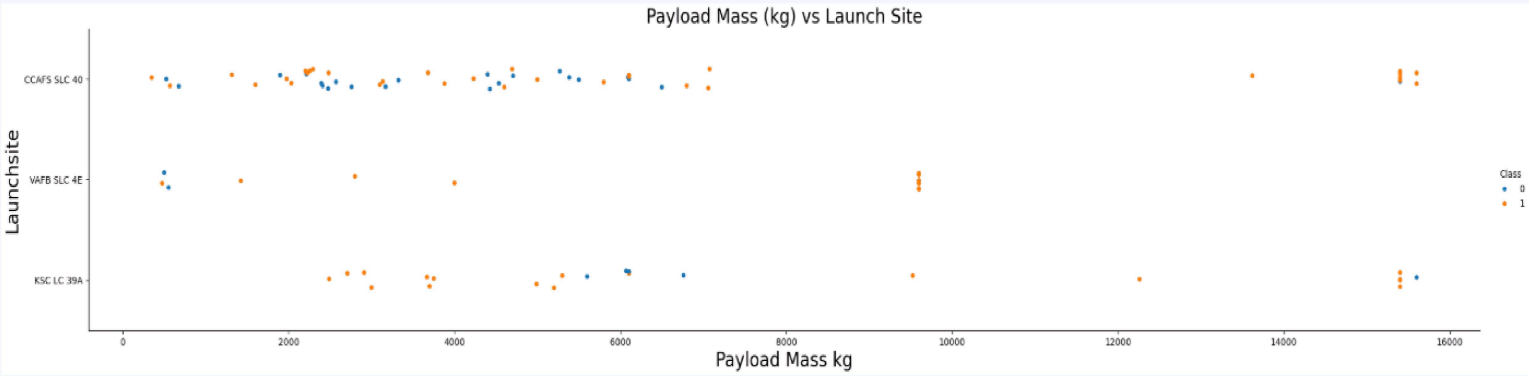
Insights drawn from EDA

Flight Number vs. Launch Site



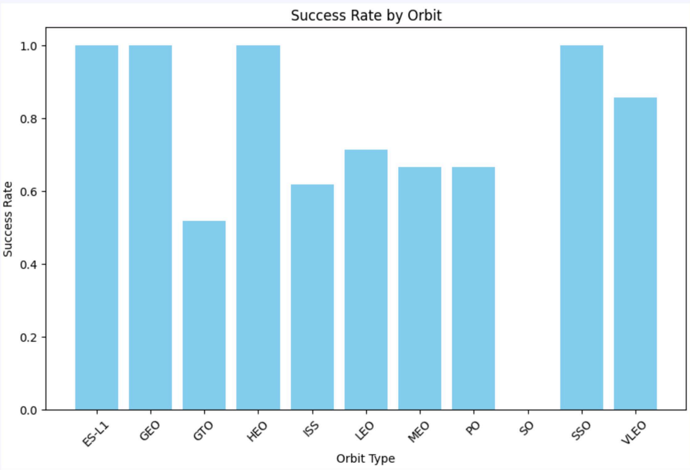
- Each launch site has a similar density of successful and unsuccessful landings.
- From this figure we obtain that the most used launch site is the CCAFS SLC 40 with the slightly more reliable site being VAFB SLC 4E.

Payload vs. Launch Site



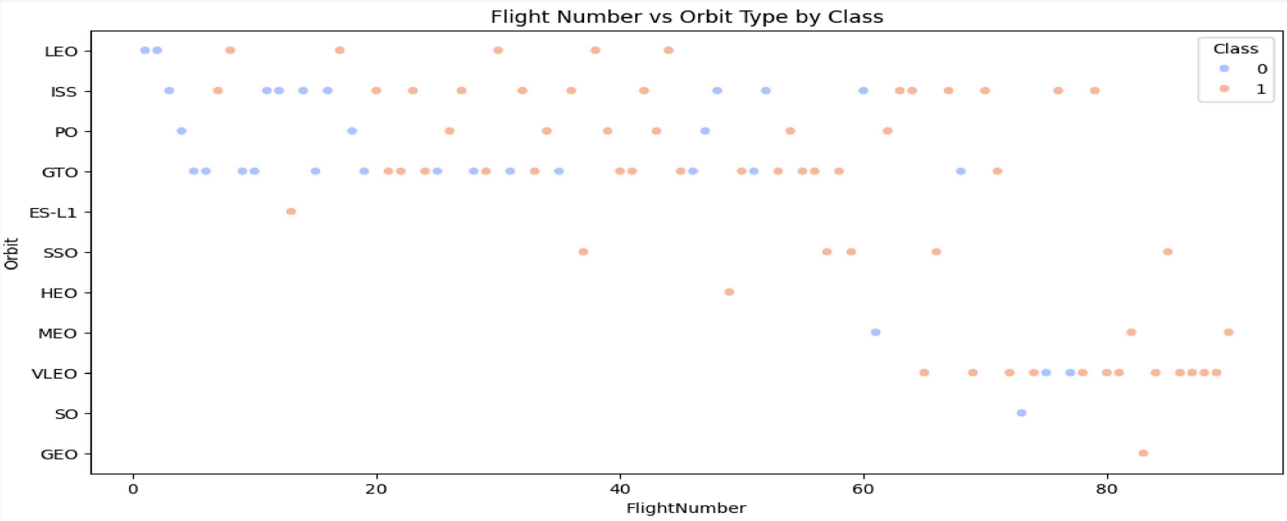
- The first observation is that it appears lower mass has higher failure rates than higher mass, but no solid trend yet.
- Standardized masses appear to be a good strategy as they have high success rates

Success Rate vs. Orbit Type



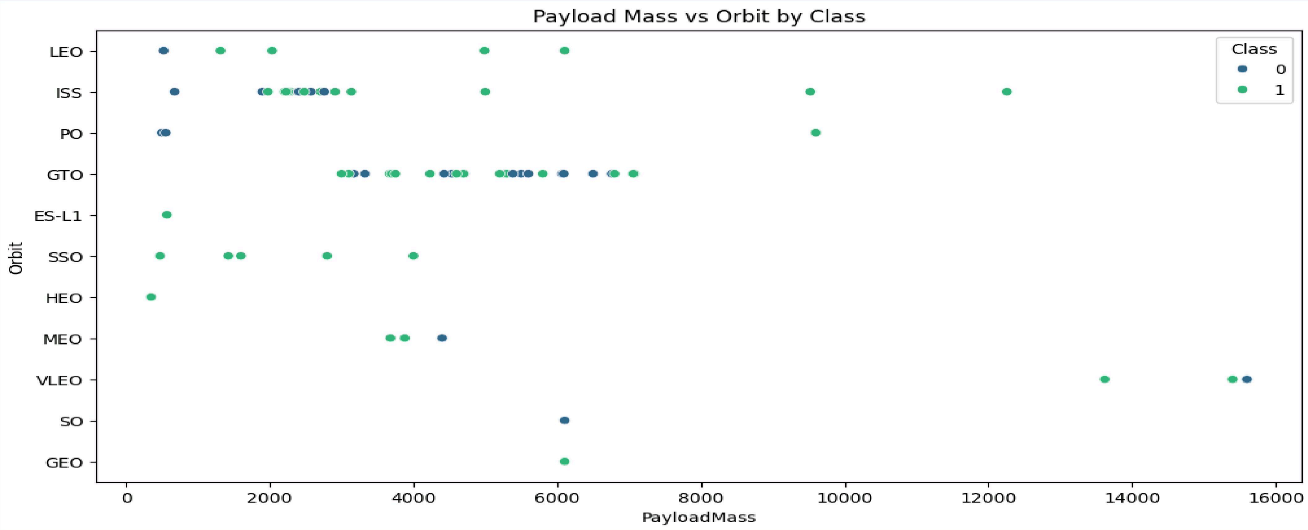
- ES-L1, GEO, HEO & SSO orbit lands have perfect scores on success and the poorest performing is PO and GTO.
- There appears no apparent trend except the middle placed orbit types are fail most.

Flight Number vs. Orbit Type



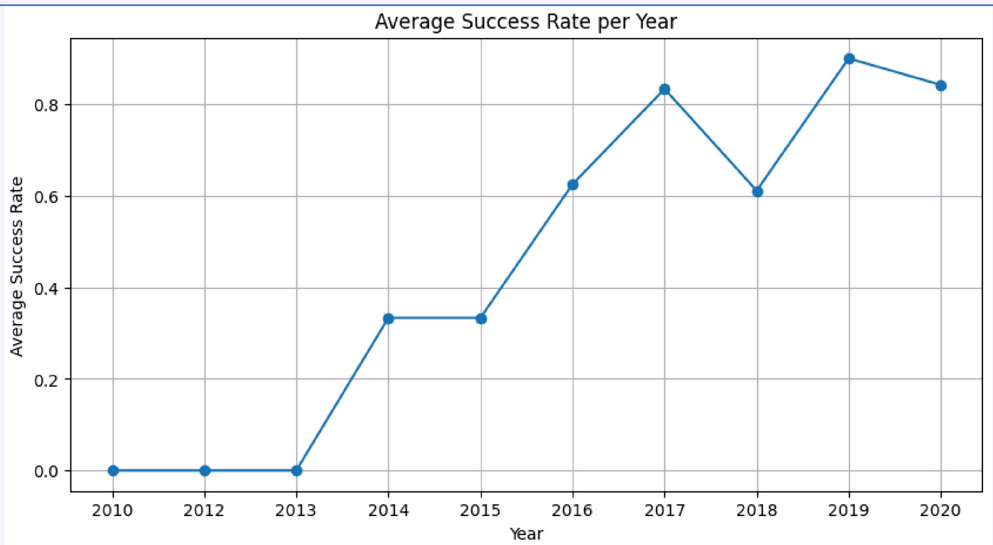
- Orbit SSO has 4/4 attempt being successful making it one of the most reliable followed by VLEO with12/14 successful, then LEO 5/7 lands making them the top orbit types.

Payload vs. Orbit Type



- The scatter plot shows that payload mass and orbit type can both impact chances of failure and success, but the distribution is not biased. In the low earth orbit, lower payload mass perform badly, in the middle orbits, medium mass perform badly.

Launch Success Yearly Trend



- The average success rate per year shows us an increasing trend in success as years progressed from 2010.

All Launch Site Names

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

- 4 launch sites are in different parts of the country

Launch Site Names Begin 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

- SpaceX has 2 sites named CCAFS, CCAFS-LC & CCAFS SLC-40 which has no records before 2013 meaning it came after LC-40

Total Payload Mass

Total_Payload_Mass
48213

- Total payload mass from NASA, a prominent customer is 48213 kg

Average Payload Mass by F9 v1.1

Avg_Payload_Mass
2928.4

- The average payload mass carried by booster version F9 v1.1 is approximately 2928.4 kg

First Successful Ground Landing Date

First_Successful_Ground_Land
2015-12-22

- The first ground land was in 22 December 2015, almost 6 years since the first launch

Successful Drone Ship Landing with Payload between 4000 and 6000

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

- The boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000 are presented above.
- The mass between 4000 & 6000 had the most failures but these are the top performing boosters in that range.

Total Number of Successful and Failure Mission Outcomes

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

- SpaceX have an exceptional mission outcome rate with only 1 failure in flight

Boosters Carried Maximum Payload

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

- The boosters which have carried the maximum payload mass are listed above. This range of boosters have a high success rate when used.

2015 Launch Records

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

- Listed above, are the failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015.
- These occurred 3 months and booster versions apart at the same site due to the same reason.

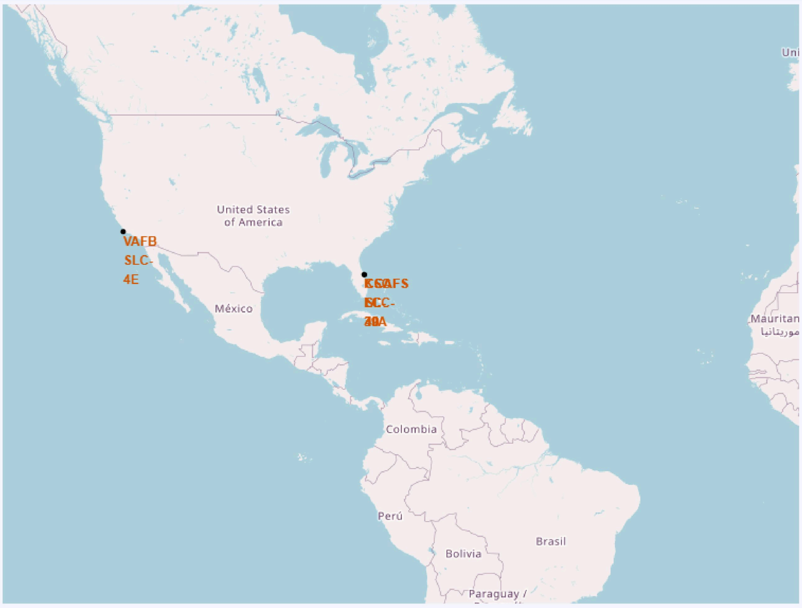
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

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

- The summary above shows ranks 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.
- This shows the that every landing outcome type has a 50/50 ratio of success and failure.

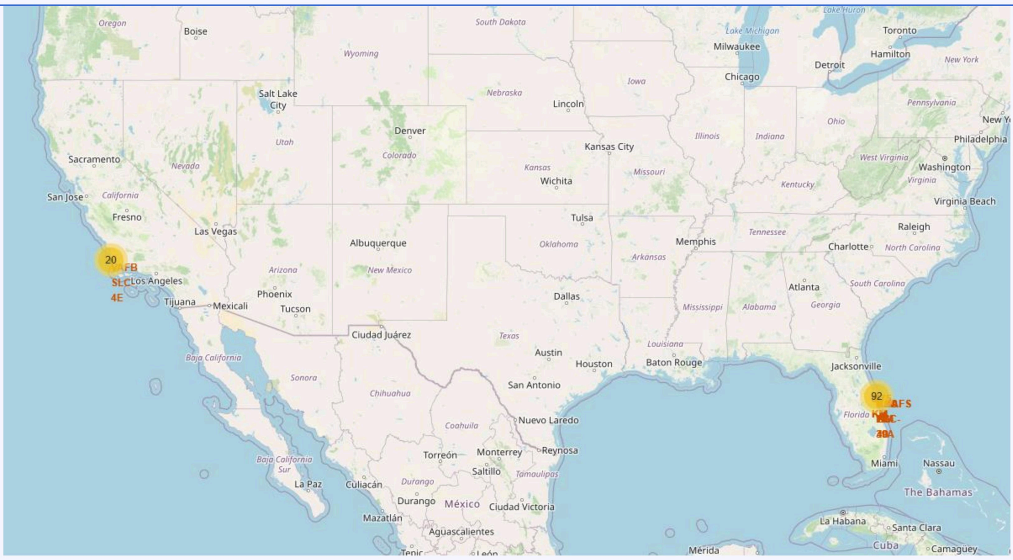

Section 3
Launch Sites
Proximities Analysis

Launch Site Locations



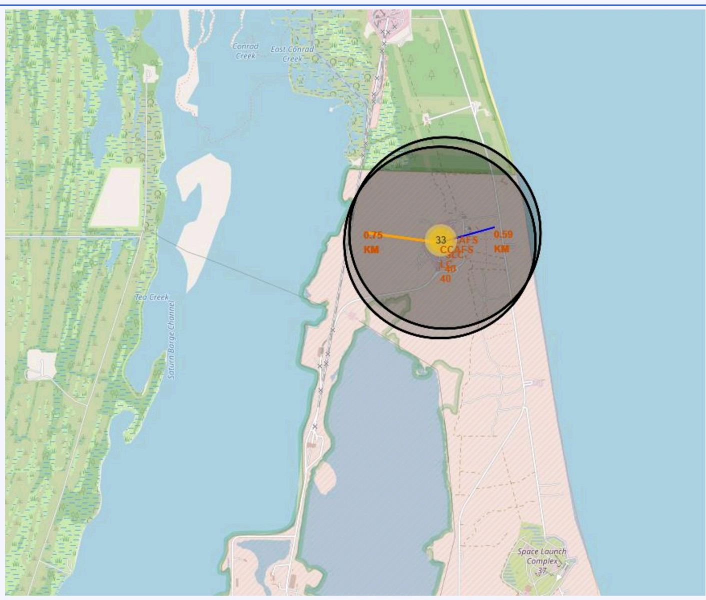
- Launch sites located west and east coast, with the CCAFS LC-40, SLC-40 and KSC LC-39A located in the same area

Site Launch Outcomes

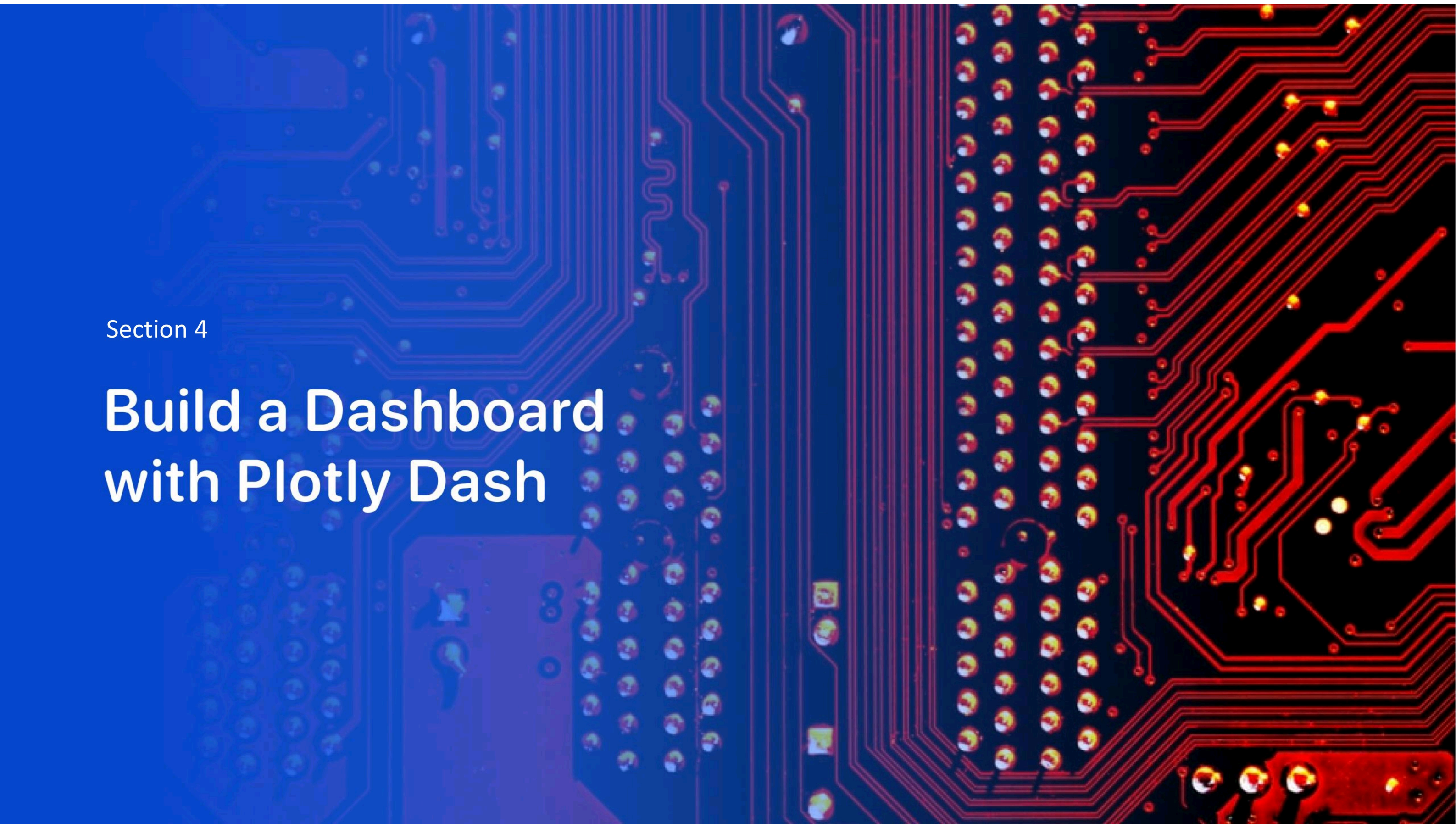


- Zooming into the map, one can visually see success and failure rate for each site.

Launch Site Proximity



- Map outlines how close in proximity sites are to railways and the ocean


Section 4
Build a Dashboard
with Plotly Dash

Launch Success Rate

- Code failure, could not generate pie chart.

Launch Site Success Rate

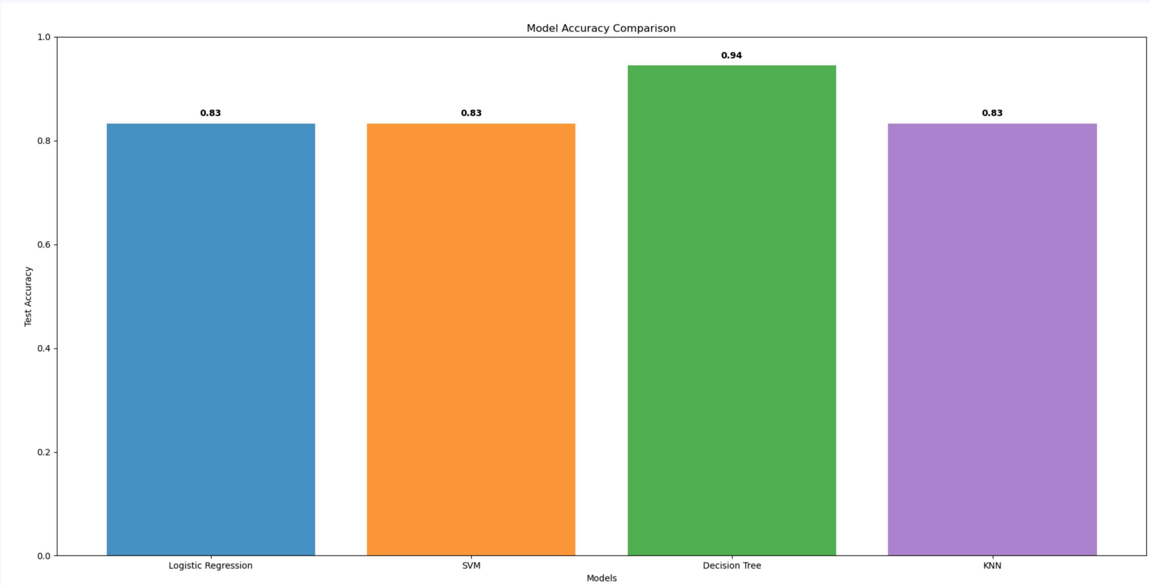
- Code failure, could not generate pie chart.

Payload vs Launch Outcome

- Code failure, could not generate pie chart.

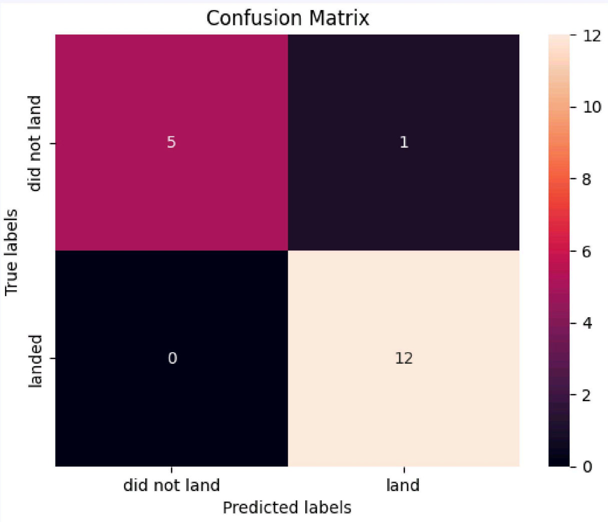

Section 5
Predictive Analysis
(Classification)

Classification Accuracy



- The decision tree classifier performs the best with an accuracy score of 0.94.

Confusion Matrix



- The decision tree classifier decision matrix shows 1 false positive, 0 false negatives, and mostly correctly predicts landing and did not land outcomes. This model performs extremely well.

Conclusions

- The applied data science methods can be used to predict whether the first stage will land or not as evidenced by the model accuracies.
- From the EDA process, it can be seen the most important factors to consider are booster version, payload mass and the orbit type.
- Other factors such as launch sites, landing type, etc. are not powerful predictors.
- The success rate per year shows that time has been the most important factors to the improvement of landing outcomes.
- This can be credited to improved techniques and researched methods which are a factor of time and past failures.

Appendix

Link to github repository:

References:

IBM Data Science Professional Certificate course via Coursera


Thank you!