IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Deepak Rupakula
04.01.2026

Outline

Executive Summary

➢ Goal: SpaceX's Falcon 9 rocket first stage landing prediction using machine learning

- Initial collection of the relevant data of SpaceX using APIs and web scraping from Wikipedia
- Data cleaning to bring it into a structured understandable table format with relevant parameters
- Exploratory data analysis using SQL and pandas to understand different parameters
- Effective and high-level data visualization to investigate influence of different parameters on successful landing of the Falcon 9
  - Additional geo-visualization to find out the locations of the launch pads and transportation means in the vicinity
- Use of different machine learning algorithms for the prediction and chosing the best algorithm

✓ 84% prediction accuracy of first stage landing can be achieved with this methodology.

Introduction

- SpaceX is a highly successful rocket company due to the reusability of its rockets after the launch. Hence the rockets, such as, Falcon 9 are economical compared to other suppliers.
  - Successful landing plays a vital role here.
- The project aims to develop a model to predict the landing based on the past data from 2011 available open in the internet to understand its business model.
- Using visualization libraries, data is visualized to understand the interdependencies of different parameters on successful landing.
  - Moreover, using geo-spatial visualization, the project tries to understand the location of launching sites for the rockets.
- Finally, different machine learning algorithms are employed to select the best algorithm along with optimal parameters and predict the outcomes optimally. Also, the parameters which influence landing the most are studied and presented.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Collect data using SpaceX API by establishing connection and Wikipedia using webscraping.
- Perform data wrangling
  - Pre-process data and the outcomes (landing) to be classified as binary
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Select different models: Logistic regression, decision tree, SVM and K means.
  - Use GridSearchCV for optimal parameter selection and accuracy.

Data Collection – SpaceX API

- https://github.com/RupDeepak/Data-Science_IBM/blob/main/Capstone_Project/Capstoneproject_code/1_spacex-data-collection-api.ipynb

1. Call API
2. Establish connection
3. Get data
4. Format data

Data Collection - Scraping

- https://github.com/RupDeepak/Data-Science_IBM/blob/main/Capstone_Project/Capstoneproject_code/2-webscraping.ipynb

1. Use HTML package
2. Provide URL
3. Import relevant table from Wikipedia
4. Use data cleaning to get into right format

Data Wrangling

- https://github.com/RupDeepak/Data-Science_IBM/blob/main/Capstone_Project/Capstoneproject_code/3-spacex-Data%20wrangling.ipynb

- The main parameter was to change the landing success/failure into a binary classifier

EDA with Data Visualization

- https://github.com/RupDeepak/Data-Science_IBM/blob/main/Capstone_Project/Capstoneproject_code/5-edadataviz.ipynb

EDA with SQL

- Using bullet point format, summarize the SQL queries you performed
- https://github.com/RupDeepak/Data-Science_IBM/blob/main/Capstone_Project/Capstoneproject_code/4-eda-sql_sqllite.ipynb

Build an Interactive Map with Folium

- Markers have been created per launch pad
- Number of markers created for each launch pad based on the number of launches
- Further markers have been drawn on a coastal area, railway line, highway and a neaby city
- Distance lines between the launch pads and these markers to know how far they are
- https://github.com/RupDeepak/Data-Science_IBM/blob/main/Capstone_Project/Capstoneproject_code/6-launch_site_location.ipynb

Predictive Analysis (Classification)

- 4 models have been built using scikit learn and scipy
- GridSearchCV is used for each model for optimal parameters and classification report along with confusion matrix is generated at the end for evaluation
- https://github.com/RupDeepak/Data-Science_IBM/blob/main/Capstone_Project/Capstoneproject_code/8-SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

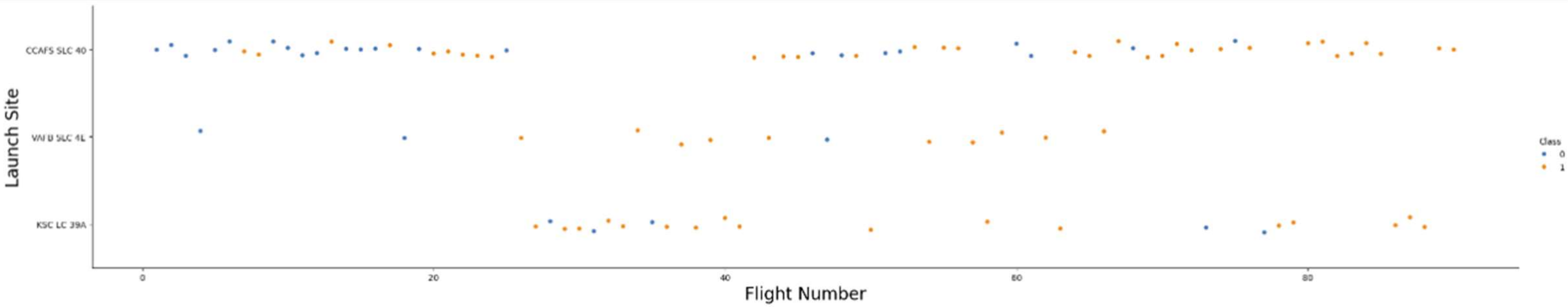
Results

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

Section 2

Insights drawn from EDA

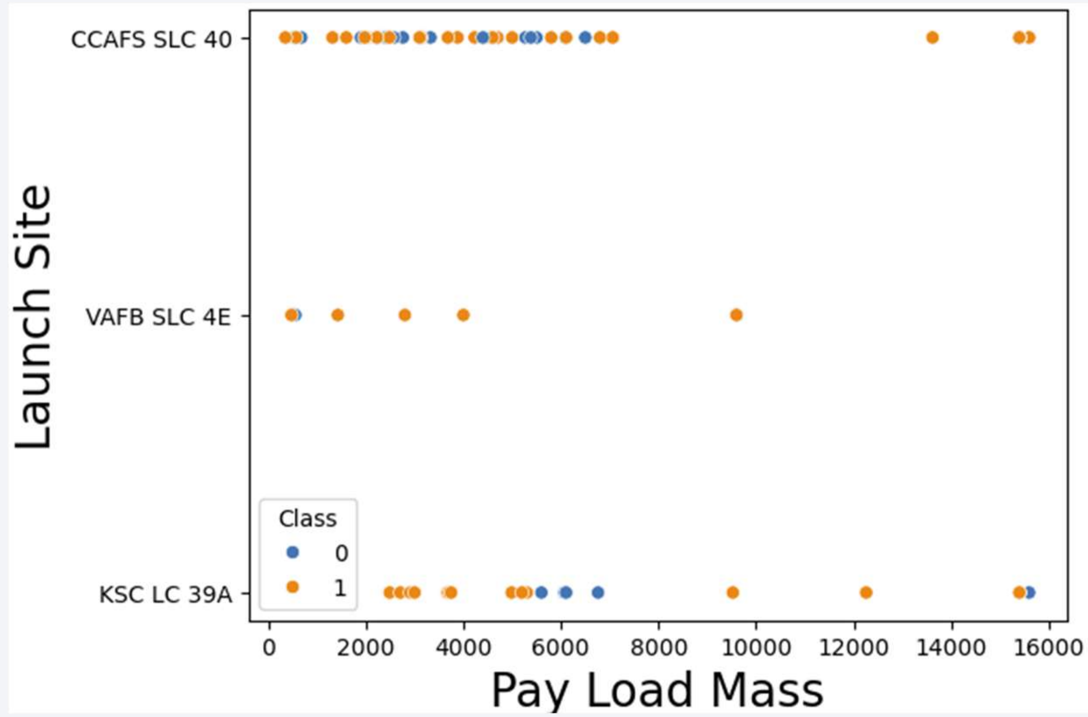
Flight Number vs. Launch Site



- Not many effects can be found here.
- For the 2nd launch site, the failure rate is the least especially after flight number 20.

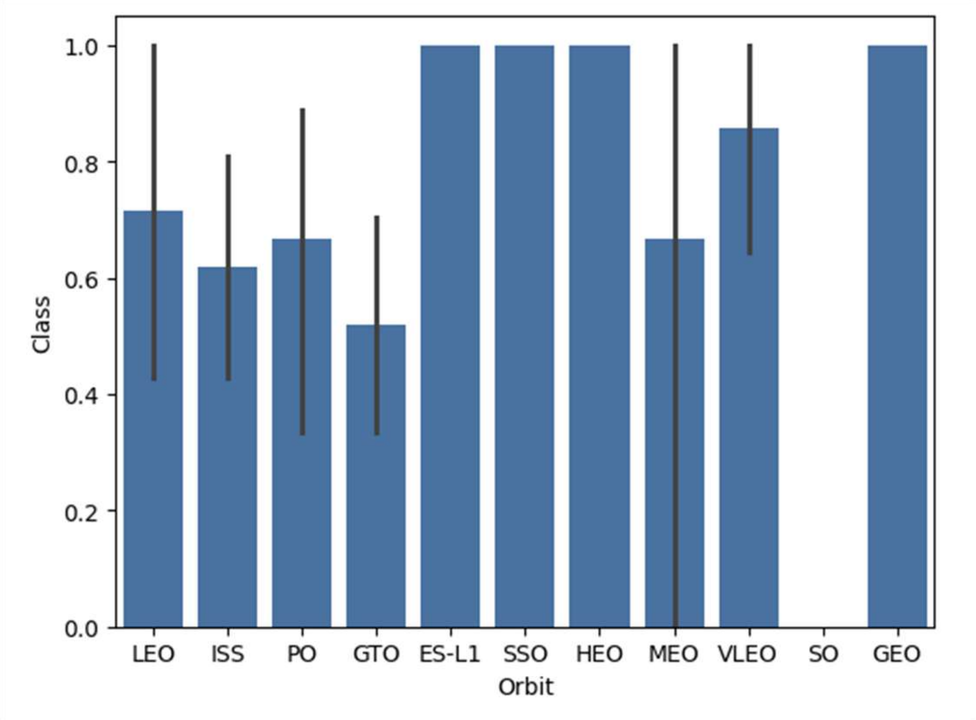
Payload vs. Launch Site

- VAFB SLC 4E launch site have no rockets with pay load mass over 10000 units
- At CCAFS SLC 40, all rockets with over 8000 units payload have successfully landed



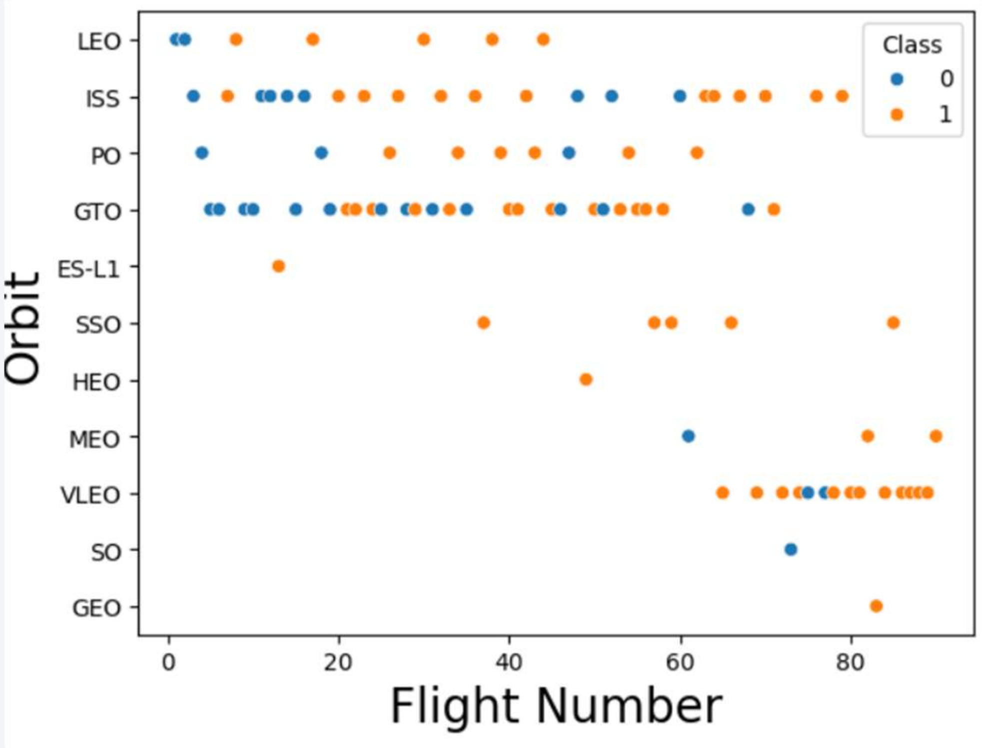
Success Rate vs. Orbit Type

- SSO, HEO, ES-L1 and GEO have 100% success rates
- GTO has the least rate



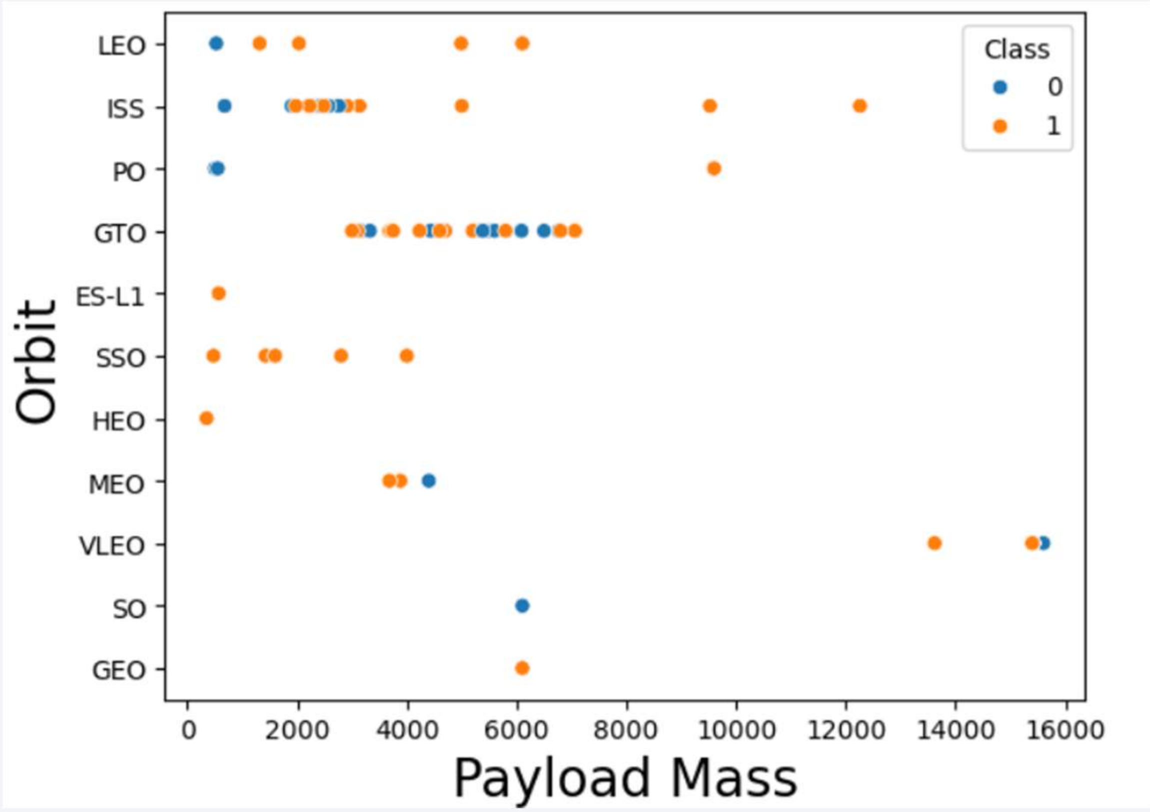
Flight Number vs. Orbit Type

- High flight numbers for the orbits GEO until SSO.
- From GTO, all ranges of flight numbers exist



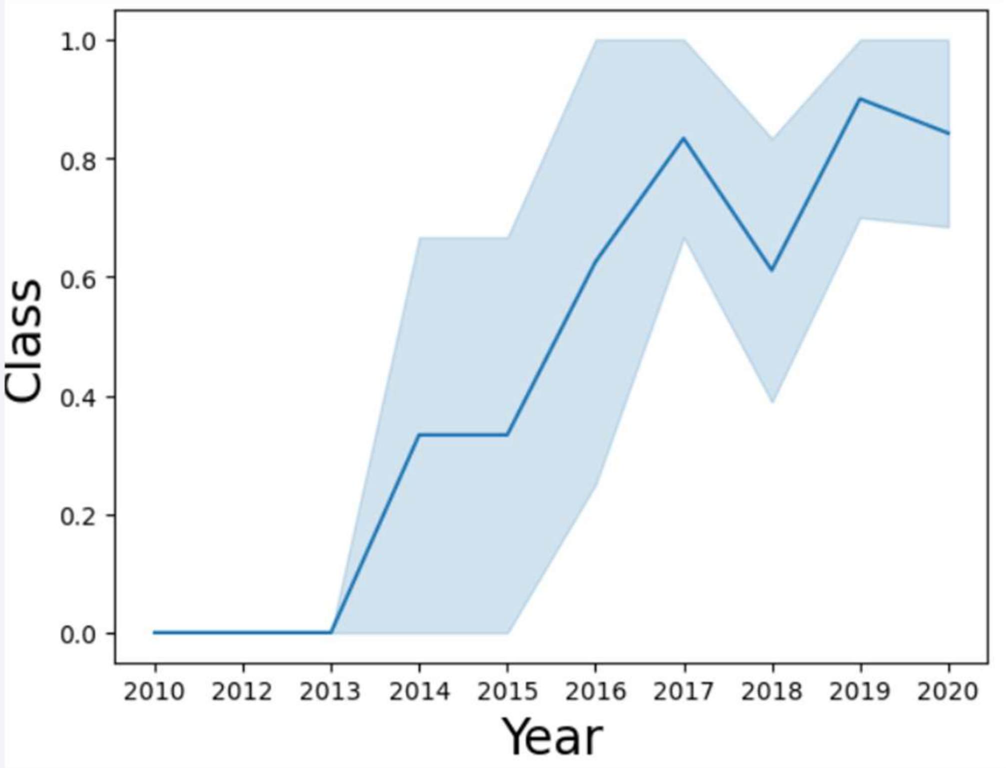
Payload vs. Orbit Type

- Heavier payloads have a high success rate of landing
  - Especially for Polar, LEO and ISS
- GTO has mixed results



Launch Success Yearly Trend

- Higher success rate over the years before COVID



All Launch Site Names

- Four distinct launch sites from Spacex
  - CCAFS LC-40
  - VAFV SLC-4E
  - KSC LC-39A
  - CCAFS SLC-40

```
%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'

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 attemp
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attemp
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attemp

- Above are the launch site
- NASA as the main customer

Total Payload Mass

- NASA as the biggest customer with a total payload mass over 9990 kg

```
%sql select sum(payload_mass__kg_) from SPACEXTABLE where customer like 'NASA%'
```

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

sum(payload_mass__kg_)
99980

Average Payload Mass by F9 v1.1

- An average payload mass of 2534 kg

```
%sql select avg(payload_mass__kg_) from SPACEXTABLE where booster_version like 'F9 v1.1

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

avg(payload_mass__kg_)
2534.6666666666665

First Successful Ground Landing Date

- First successful landing outcome in late December 2015
- Gradual increase later on as shown in slide 21

Done.

min(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Four different boosters in this payload range

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

Total Number of Successful and Failure Mission Outcomes

- Totally 71 mission outcomes

count(*)
61

Boosters Carried Maximum Payload

- 12 B5 boosters carried the maximum payload

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

- Most landing outcomes in 2015 were either unattempted or failed.
- The first successful landing out come was by F9 FT B1019 in December as also explained in slide 26

month_name	Landing_Outcome	Booster_Version
january	Failure (drone ship)	F9 v1.1 B1012
february	Controlled (ocean)	F9 v1.1 B1013
march	No attempt	F9 v1.1 B1014
april	Failure (drone ship)	F9 v1.1 B1015
april	No attempt	F9 v1.1 B1016
june	Precluded (drone ship)	F9 v1.1 B1018
december	Success (ground pad)	F9 FT B1019

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

- With majority of landing outcomes as successful, SpaceX turned to be the most profitable rocket company.

outcome	cnt
Success	38
No attempt	21
Success (drone ship)	14
Success (ground pad)	9
Failure (drone ship)	5
Controlled (ocean)	5
Failure	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1
No attempt	1

Section 3

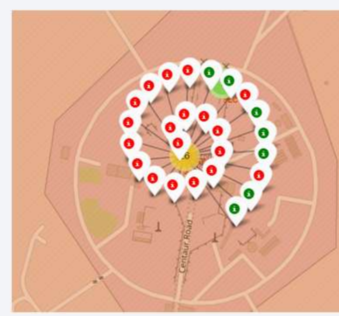
Launch Sites Proximities Analysis

Launch site locations in USA

- The four unique launch sites are marked with red markers
  - Mostly coastal regions and away from big cities

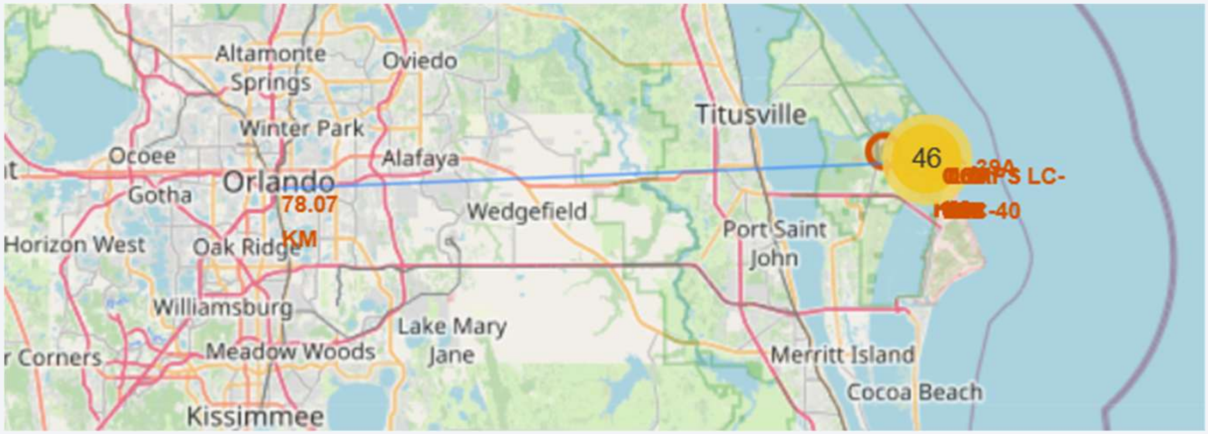
Number of launches along with coloured markers

- Number of launches per site is also visualized
  - Most of the launches in the east coast compared to the west
- Further differencitation in the colours of the markers based on successful landing
  - Green: Success
  - Red: Failed/not attempted



Proximity to cities and transport means

- Further analysis shows that the launch sites are very far from a city
    - Over 70 km away from Orlando



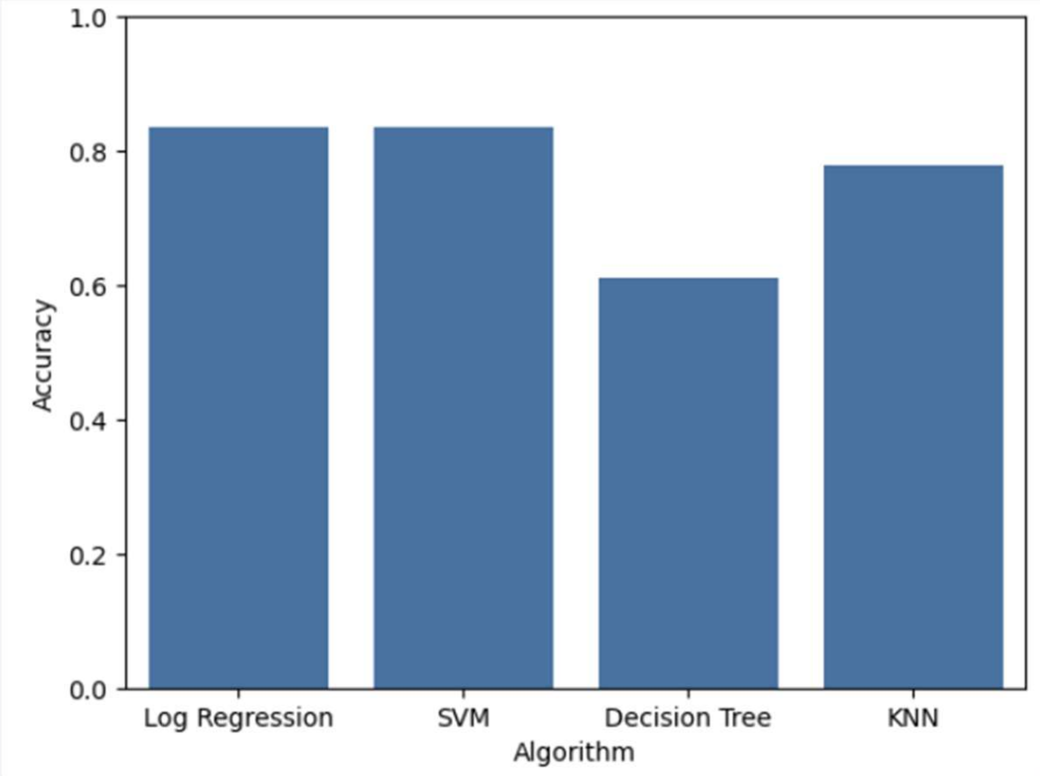
- However, the nearest railway line, coastal region and highways are within the vicinity in 2km range

Section 5

Predictive Analysis (Classification)

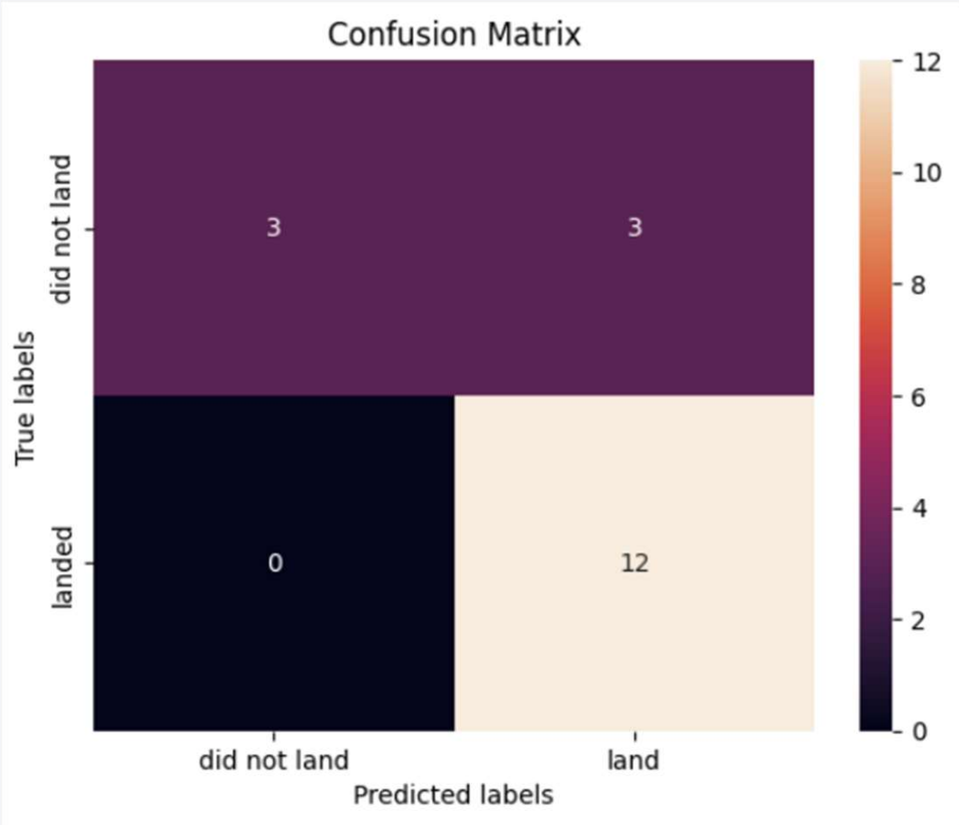
Classification Accuracy

- From the analysis, SVM and Logreg have almost equal accuracy of around 85%.
- SVM performed slightly better with higher accuracy using training data.
- Decision tree had the least accuracy here.



Confusion Matrix

- This algorithm provided high number of true positives- 12
- False positive is an important parameter in deciding the algorithm
  - SVM performed the best with only 3 false preductions of successful landing.



Conclusions

- A prediction model has been successfully developed to predict landing of SpaceX's Falcon 9 based on various parameters.
- An accuracy of 85% has been achieved with least number of false positives without overfitting.
- Many insights regarding launch pads, success rates over the years, orbit type and its dependance on success rate, payload mass and launch outcomes with booster versions for certain years have been gained throughout.
- A geospatial analysis has been furthermore helped in understanding launch pad locations, number of launches per location and its distance to the coast, railway line, highway and a city.

➢ A successful landing of the rocket can thus be predicted with high accuracy using a set of parameters and past data without rocket science knowledge. Highly important for the rocket price to stay economical and competitive in the market.

Thank you!