SPACEX LANDING ANALYSIS

STUART PERRY

MELBOURNE
AUSTRALIA

12/12/22

EXECUTIVE SUMMARY

Data collection methodology

Web scraping

SpaceX API

Data exploration and processing

Data cleaning

SQL databases

Dashboarding and visualizations

Mapping

Spatial analysis

Predictive analysis and Modelling

Results and discussion

INTRODUCTION

- Why are we looking at landing data from SpaceX?
- The project aims to predict a successful or failed landing of SpaceX shuttle boosters.
- Explore and optimize machine learning algorithms for predictions.
- Build visualizations and dashboards to convey the findings effectively.

DATA ACQUISITION AND WRANGLING

- Two separate datasets were collected.
  - API data from api.spacexdata.com containing launch info.
  - Web data from Wikipedia scraped using beautifulsoup4.
- API data was cleaned, and features columns reduced.
  - Some missing values imputed using the mean column value.
  - Data was labelled with binary values to represent success or failure.
- Web data was extracted from tr tags in the HTML.
  - Python logic used to filter the data and build a pandas data frame.
  - Data exported to csv then loaded into IBM db2 SQL database on cloud.

DATA EXPLORATION

- API data explored with matplotlib and seaborn visualization libraries.
  - Scatter plots to visualize correlation between variables
  - Bar chart to see the most successful orbit types
  - Line chart to describe the progression of successful launches over time.
- Web data analyzed using SQL queries.
  - Access and query the cloud db using SQL magics in Jupyter.
  - Write analysis queries to describe the dataset.
  - Combine Pandas functionality with SQL queries for more complex analysis.

MAPS AND INTERACTIVE DASHBOARD

Folium and python used with Web data to build interactive maps and spatial analysis of launch locations.

- Created child objects with html styling to add components to the map.
- Used mouse location to calculate distance between launch pads and critical infrastructure.

Plotly DASH app built using python to build an interactive analysis dashboard from API data.

- Built core components using dash's HTML package.
- Visualizations functions used the callback decorator to update data components of the dashboard.

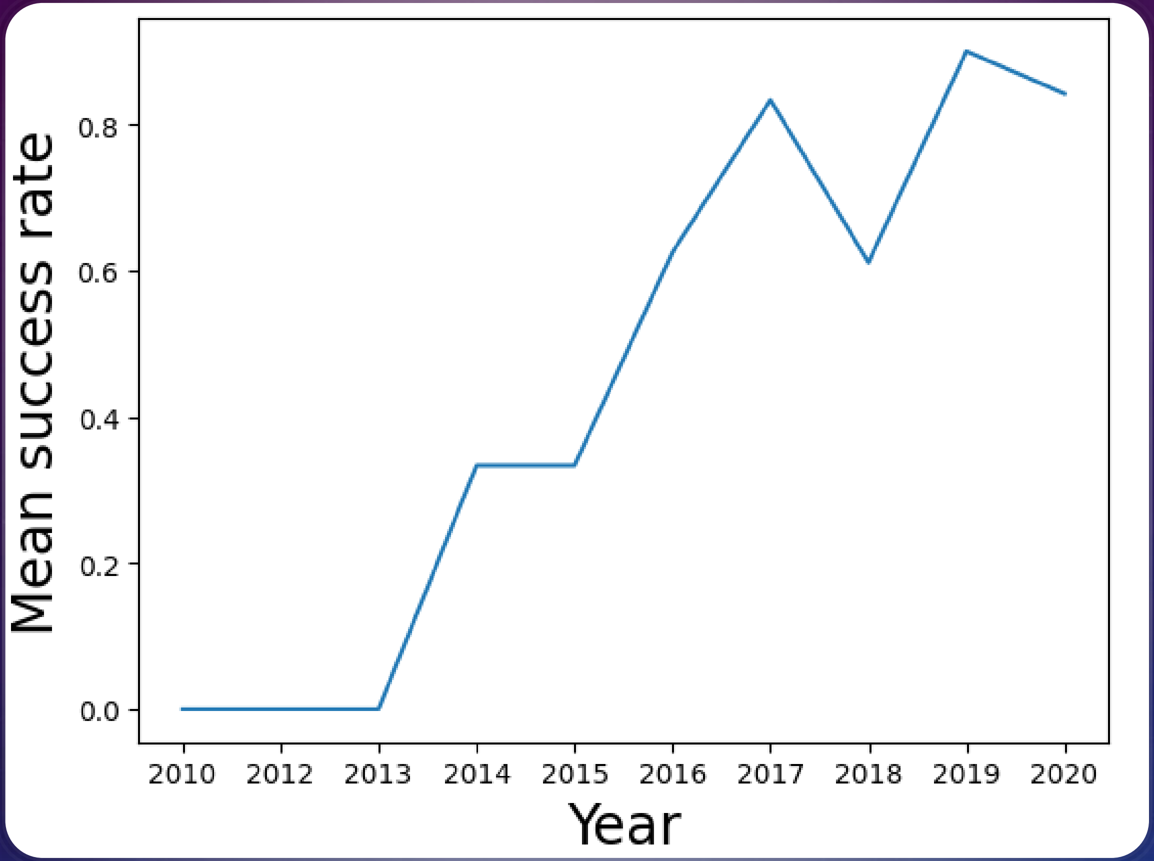
MODELLING AND PREDICTIVE ANALYSIS

Data was encoded and standard scaler applied to obtain a mean of 0 and std dev of 1.

A training and testing dataset were split into 80% and 20% respectively.

Four model types were fit, Grid Search was used for optimized hyperparameters.

- Logistic regression with the lbfgs solver
- Support Vector Classifier
- Decision Tree
- K-nearest Neighbours



DATA ANALYSIS AND VISUALISATION RESULTS

- Landing success rate over time has improved from 0% to ~80% within 7 years, following this trend, SpaceX should average >90% landing success within a few years.
- These values are driven by a few select orbit successes, and more development will have to happen for better landing rates in the SO, GTO and ISS orbit types.

IBM Developer

SKILLS NETWORK

SQL DATA ANALYSIS RESULTS

SpaceX launched from four unique launch sites

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

Total payload mass carried for the was 45,596Kg

Average payload mass carried by F9 v1.1 booster was 2,928Kg

Boosters which were successful at landing on drone ships with payloads between 4,000Kg and 6,000Kg

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

SQL DATA ANALYSIS RESULTS

Of all mission outcomes from the web dataset, there was only 1 confirmed mission failure with a total sample size of 101.

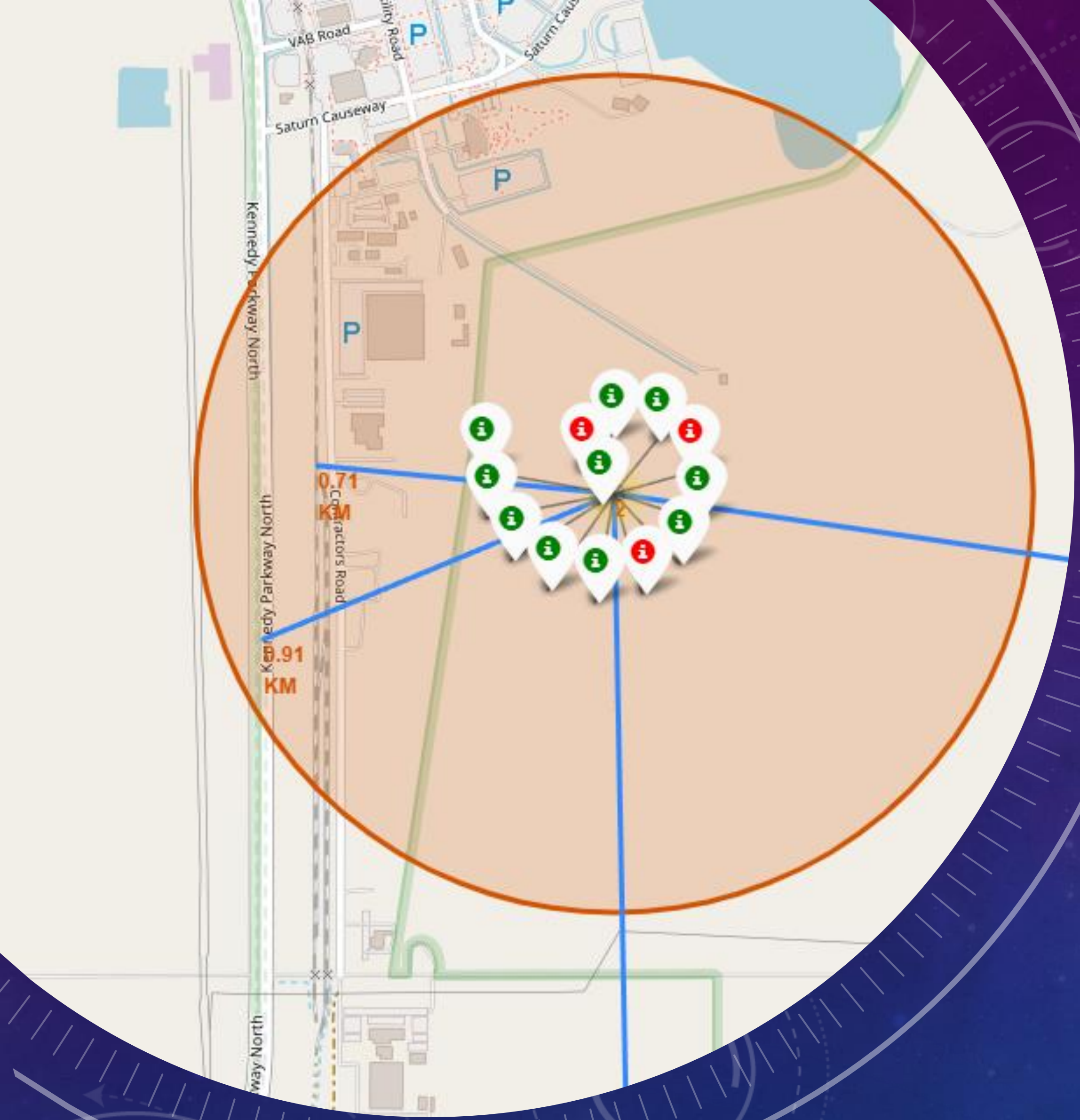
- Failed landings can still be deemed as a successful mission

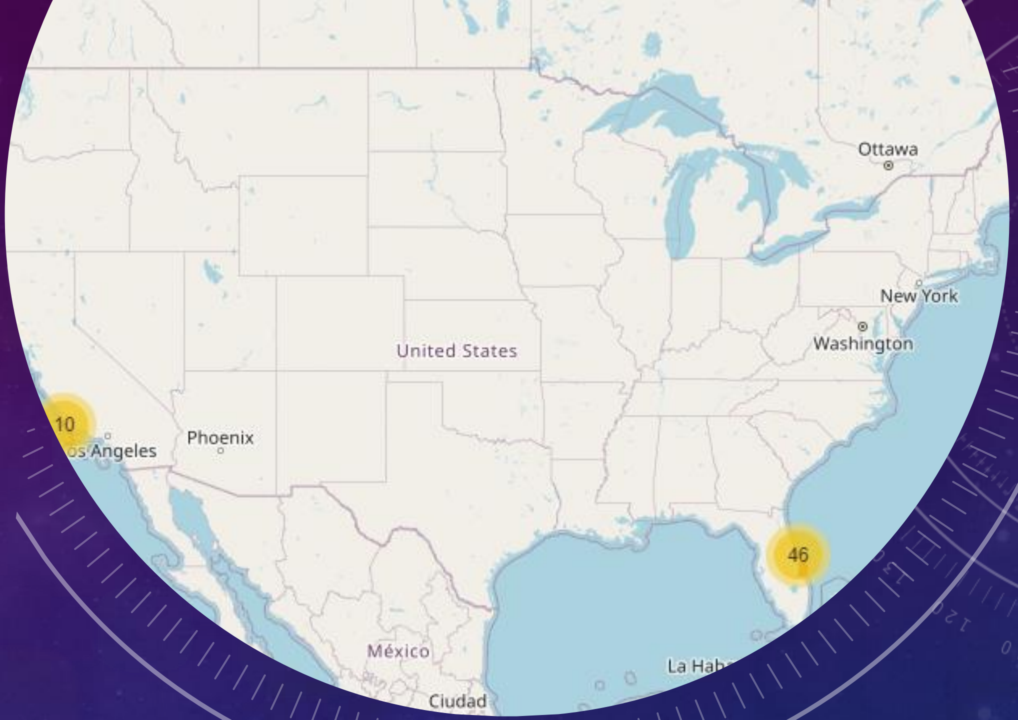
There were 12 individual booster versions which carried the maximum payload capacity.

There were two LEO missions launched from CCAFS LC-40 in 2015 that failed to land on the drone ship.

There were 8 missions that had a successful landing between the years of 2010 and 2017.

- Payloads for these missions ranged between 2,000Kg and 9,000Kg


VAB Road
Saturn Causeway
Kennedy Parkway North
0.71 KM
Contractors Road


Ottawa
New York
Washington
United States
Phoenix
10
46
México
Ciudad

FOLIUM MAP

FOLIUM ANALYSIS RESULTS

It was observed that launch sites have been placed in proximity to coastal locations.

- To lower the risk of collateral damage in the event of rocket failure
- For ease of access when drone ships return a payload.

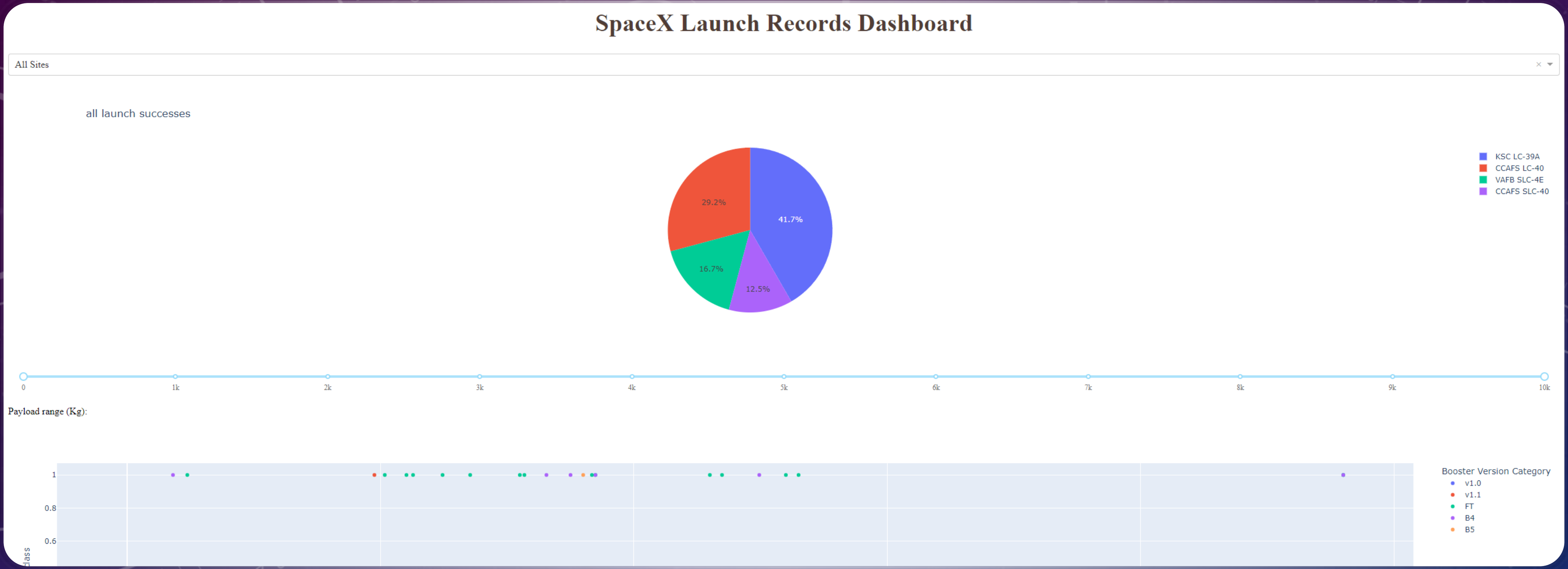
Launch sites are located closer to the equator to take advantage of earth's spin during launch.

- Allowing rockets to spend less fuel.

Launch sites appear to be close to critical infrastructure such as roads, railways and airports.

- Providing access for necessary launch equipment and personnel.

Launch sites maintain a minimum distance of ~50Km from major cities.



LAUNCH RECORDS DASHBOARD

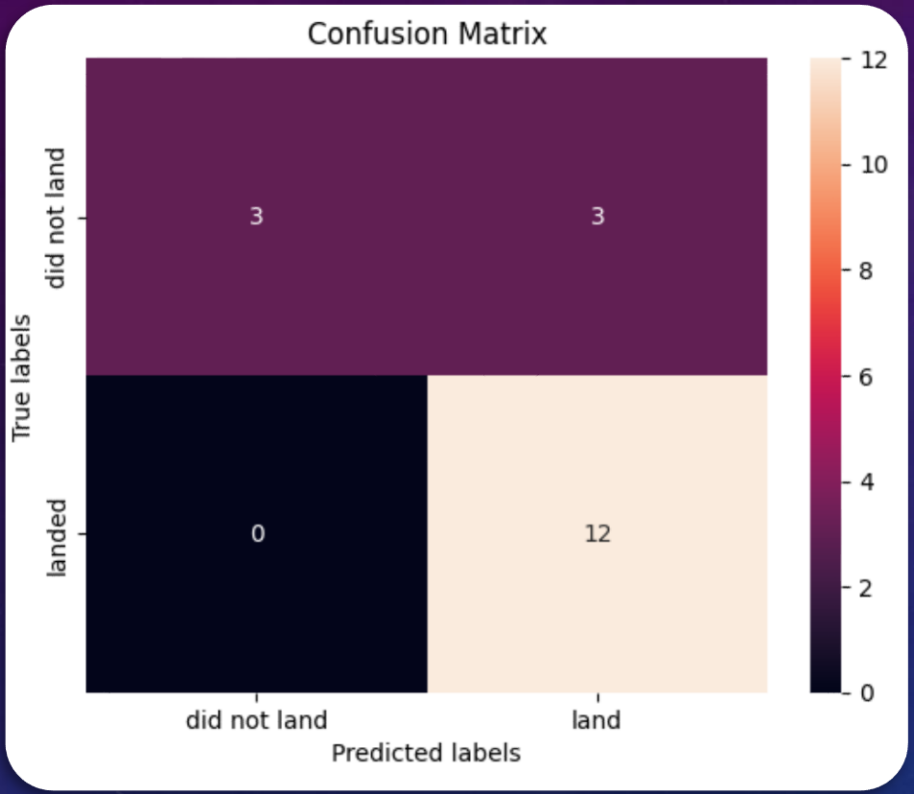
IBM Developer

SKILLS NETWORK

DASHBOARD ANALYSIS RESULTS

- Booster versions 1.0 and 1.1 have an exceedingly high failure rate with only 1 booster successfully landing out of 18 launches.
- Launches from KSC LC-39 launch site with the B4 booster variant saw a 100% success rate over 3 launches.
- All launches from KSC LC-39 below 5500Kg had a successful landing outcome.
  - 10 out of the 13 launches were successful.

MODELLING AND PREDICTIVE ANALYSIS RESULTS



All models performed relatively similarly, the testing accuracy was highly analogous for all models and the confusion matrix displayed the same results for all models except the Decision Tree.

Model	Training Accuracy	Testing Accuracy
Logistic regression	0.846	0.833
SVC	0.848	0.833
Decision Tree	0.888	0.666
K-Nearest Neighbours	0.848	0.833

MODELLING AND PREDICTIVE ANALYSIS RESULTS

- Because of the limited testing dataset, all trained models were tested again on the entire dataset to obtain an accuracy for each model.
- The best performing overall model appeared to be the SVC. This models also outperformed the others regarding f1 score and overall accuracy.

```
SVC()
 Whole dataset score:0.8777777777777778
 Best Testing Score: 0.8333333333333334
 Best Training Score:0.8482142857142856

              precision    recall  f1-score   support

           0       1.00      0.63      0.78        30
           1       0.85      1.00      0.92        60

    accuracy                           0.88        90
   macro avg       0.92      0.82      0.85        90
weighted avg       0.90      0.88      0.87        90
```

CONCLUSION

- Data collection and wrangling of SpaceX data was manageable and produced high quality clean data.
- Visualizations were able to communicate trends in the data, leaving the end user more informed about SpaceX.
- An interactive dashboard proved invaluable for slicing and viewing the data in different forms.
- Predictive models were able to capture the complexity the data and build models to predict landing success with appropriate accuracy.