Predicting SPACEX Launch Outcomes

Michelle Dean

November 2022

Outline

- Executive Summary
- Introduction
- Methodology
    - EDA with SQL & Visualization
    - Folium Maps
    - Dashboard Plots
- Results
- Conclusion
- Appendix

Executive Summary

SPACEX is the leader of the most affordable space travel in the USA. The cost of space travel has been significantly reduced because of the possibility to preserve the first stage of rocket launch.

By training machine learning models with SPACEX's historical data our results show that using a SVM model is most accurate when predicting future landing outcomes. With knowing the predictive outcome of launches we can understand more about cost.

Introduction

My goal is to understand what makes space travel more affordable by predicting SPACEX's rocket launch success outcomes.

- What factors determines if the first stage is able to land?
- What is the best way to accurately predict if the first stage will land?
- Using supervised machine learning models and historical outcome of SPACEX launches I aim to find the best method to accurately predict future SPACEX launch outcomes

Methodology

- Data collection methodology:
  - Web scrapping HTML tables from Wikipedia
  - Called SPACEX API get requests, parsed with JSON files into pandas data frame
- Data wrangling
  - Removed unimportant data
  - Created training labels and columns
  - Replaced missing values with means
- Exploratory data analysis (EDA) using visualization and SQL
- Interactive visual analytics using Folium and Plotly Dash
- Predictive analysis using classification models
  - Evaluation of classification models

1:

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

2:

boosterversion
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

- EDA with SQL

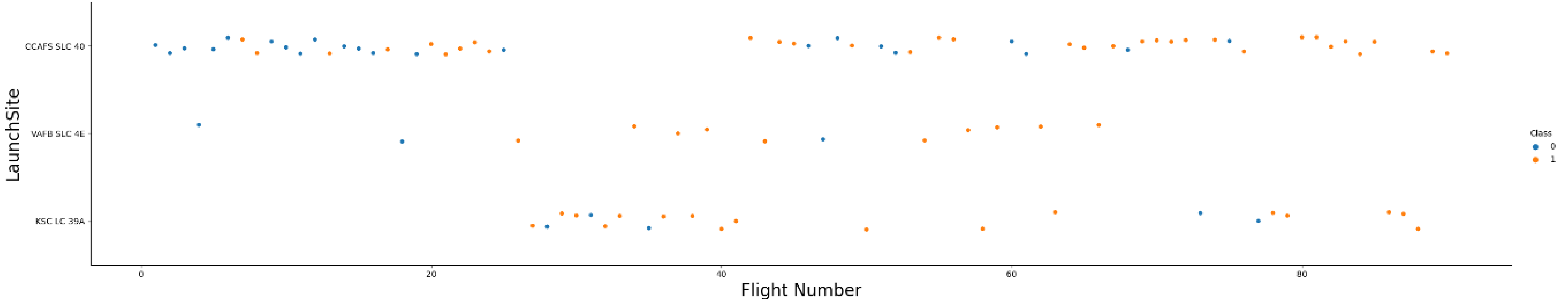
Some examples of queries used to understand more about the factors that affect the success or failure of launches

1: Successful Drone Ship Landing with Payload between 4000 and 6000

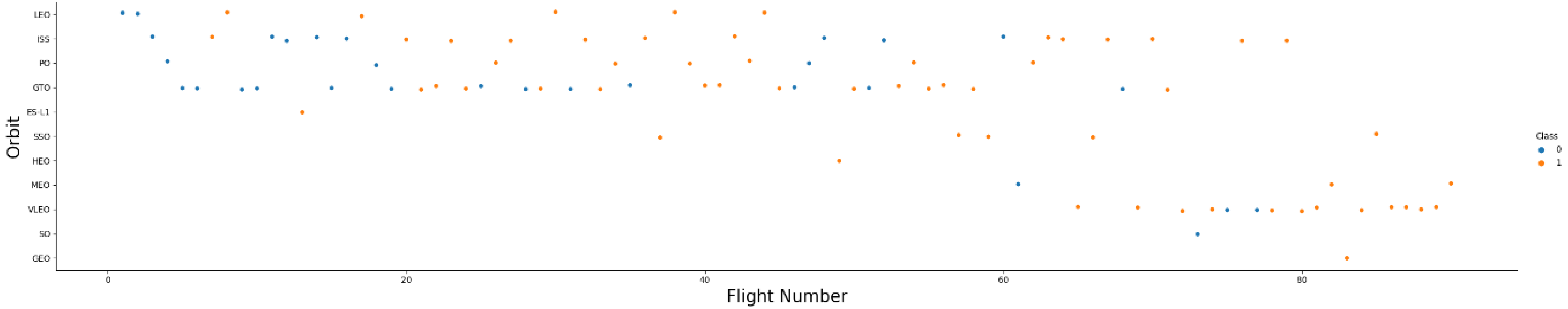
```
%sql select Booster_Version from
SPACEXTBL where "Landing _Outcome" = 'Success
(drone ship)' and PAYLOAD_MASS__KG_ BETWEEN
4000 and 6000;
```

2: Boosters Carried Maximum Payload

```
%sql select Booster_Version as boosterversion f
rom SPACEXTBL
where PAYLOAD_MASS__KG_ =(select max(PAYLOAD_MA
SS__KG_) from SPACEXTBL);
```



2:

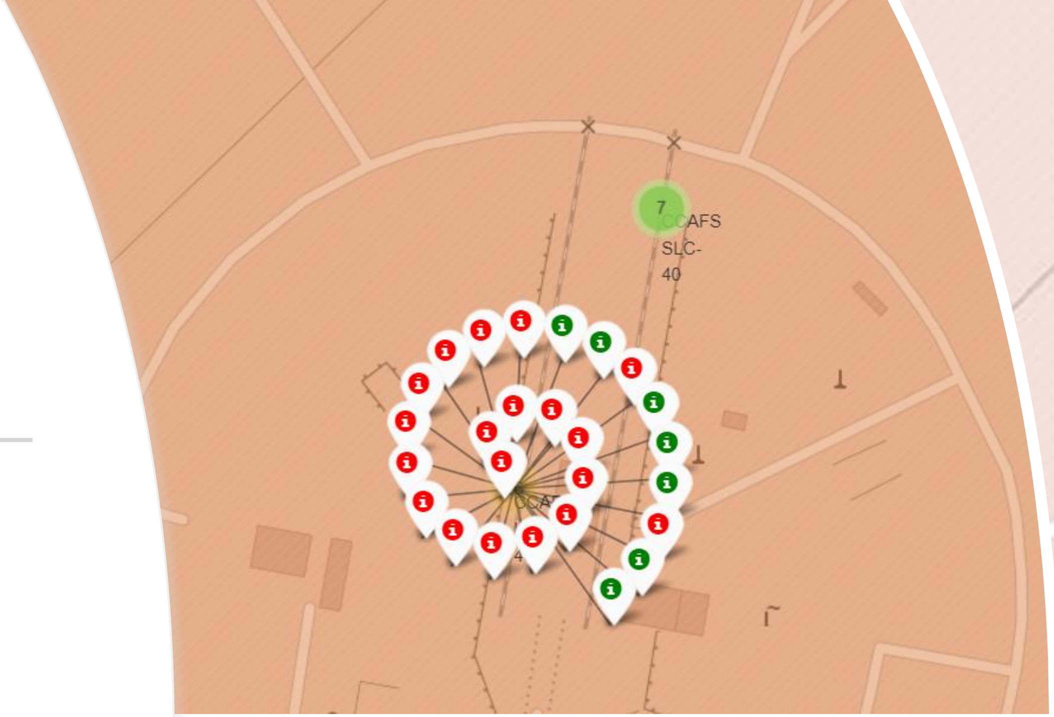


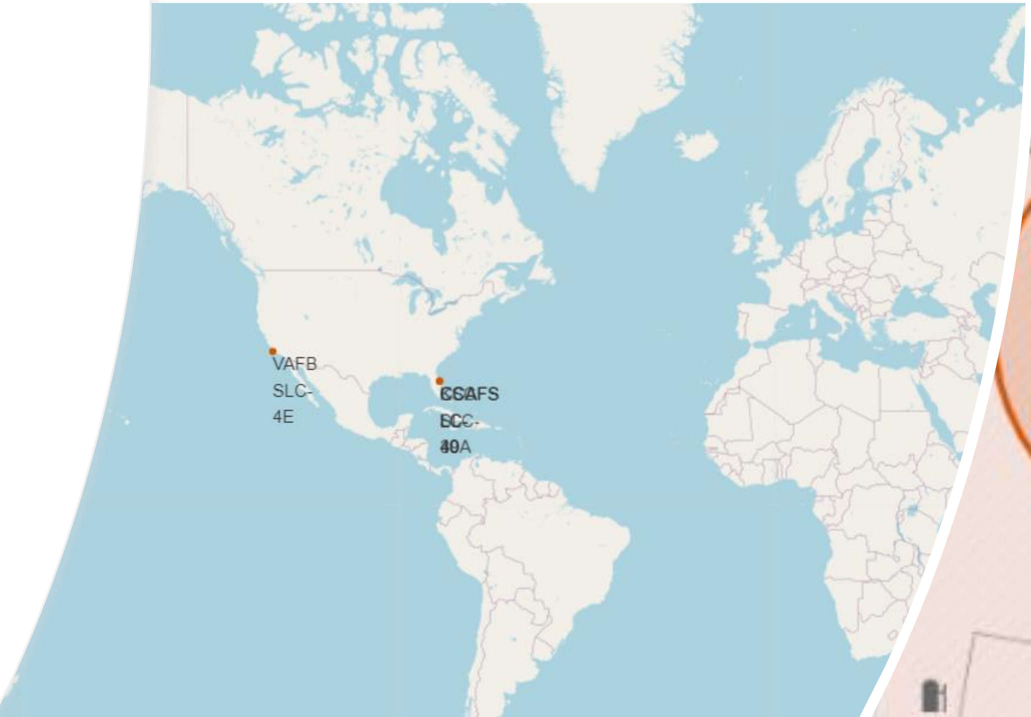
EDA with Visualization

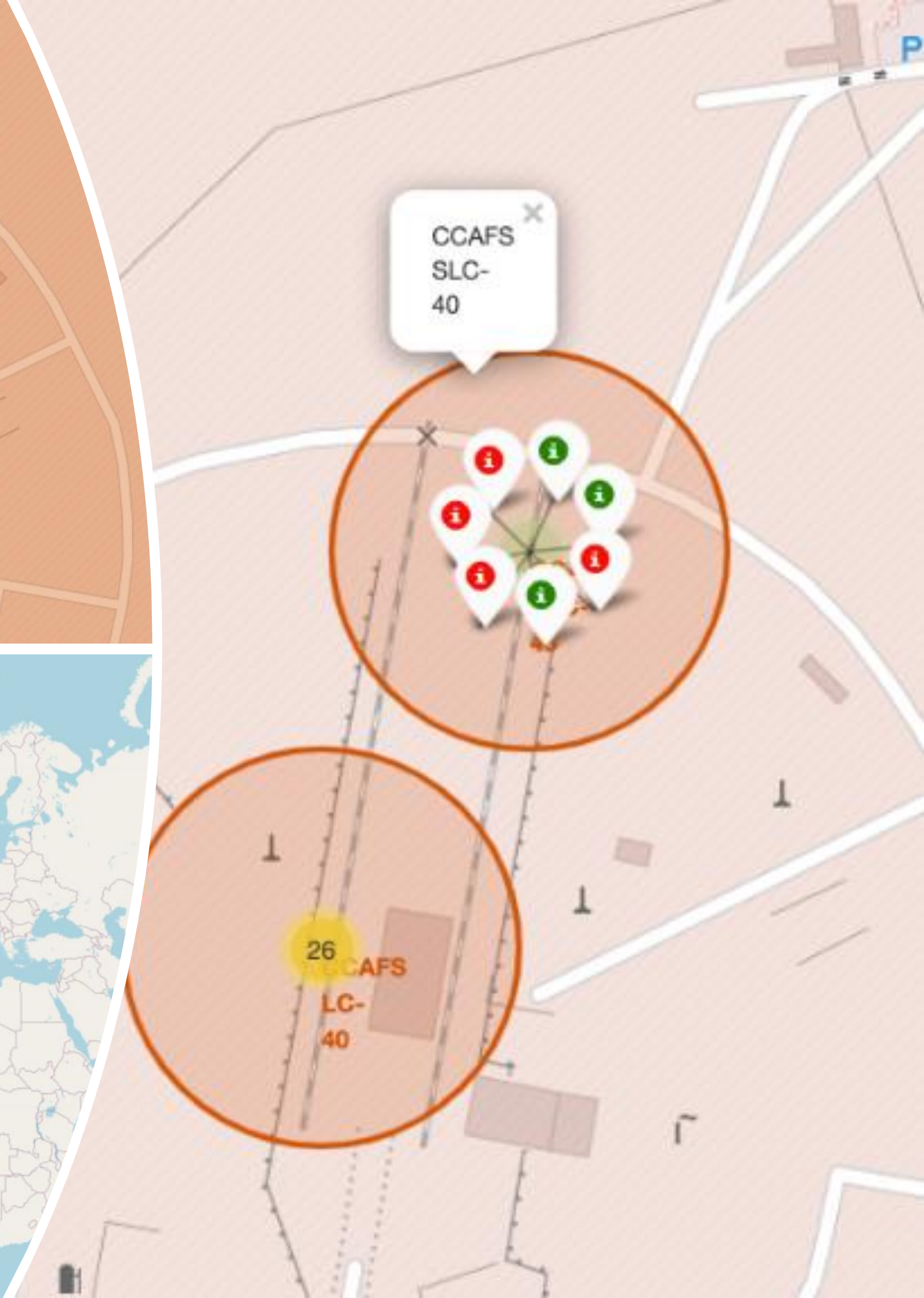
- 1: We see that different launch sites have different success rates. CCAFS LC-40, has a success rate of 60 %, while KSC LC-39A and VAFB SLC 4E has a success rate of 77%
- 2: We see that in the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit

Visualizing with Folium Maps

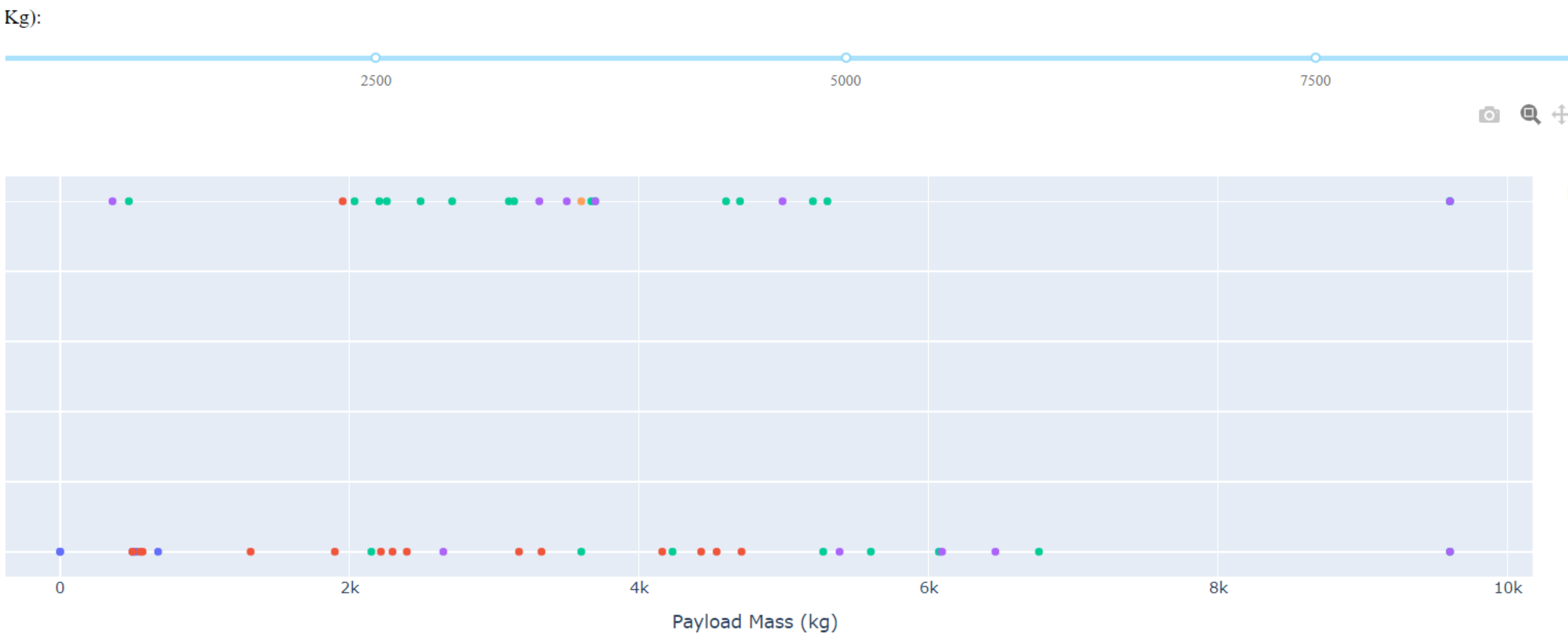
- Spiral Image with green markers for successful and red markers failed launches at the CCAFS SLC-40 launch site
- World map of all SPACEX launch location sites with markers
- Map showing marker clusters of the two launch locations at Cape Canaveral







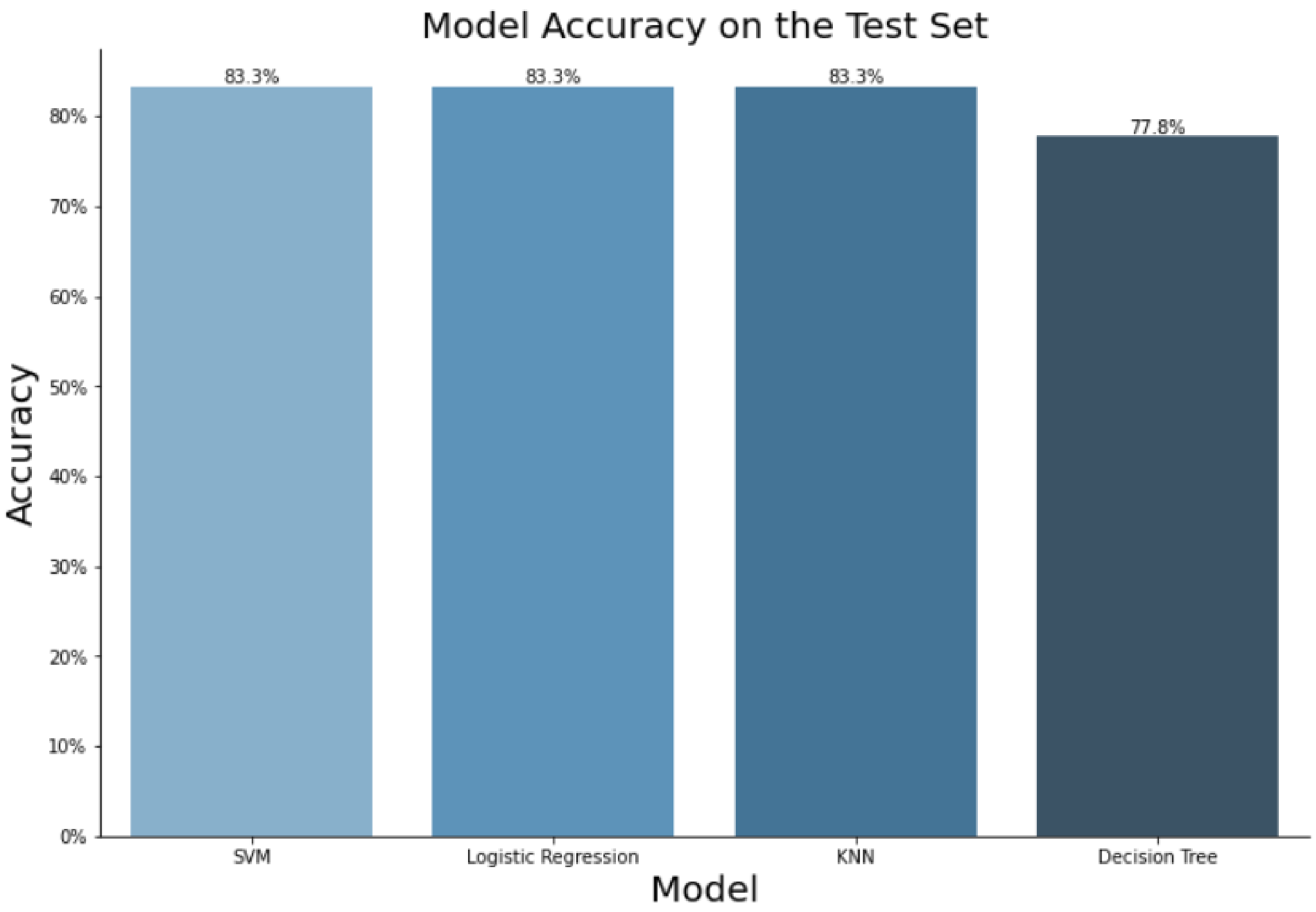
Scatters Plot of Payload vs Launch Outcome at various payloads



- From these scatter plots we can see as the payload increases there are fewer booster versions that have success. The booster version with the most success at higher payloads are FT and B4. Overall the boosters with the most success is FT.

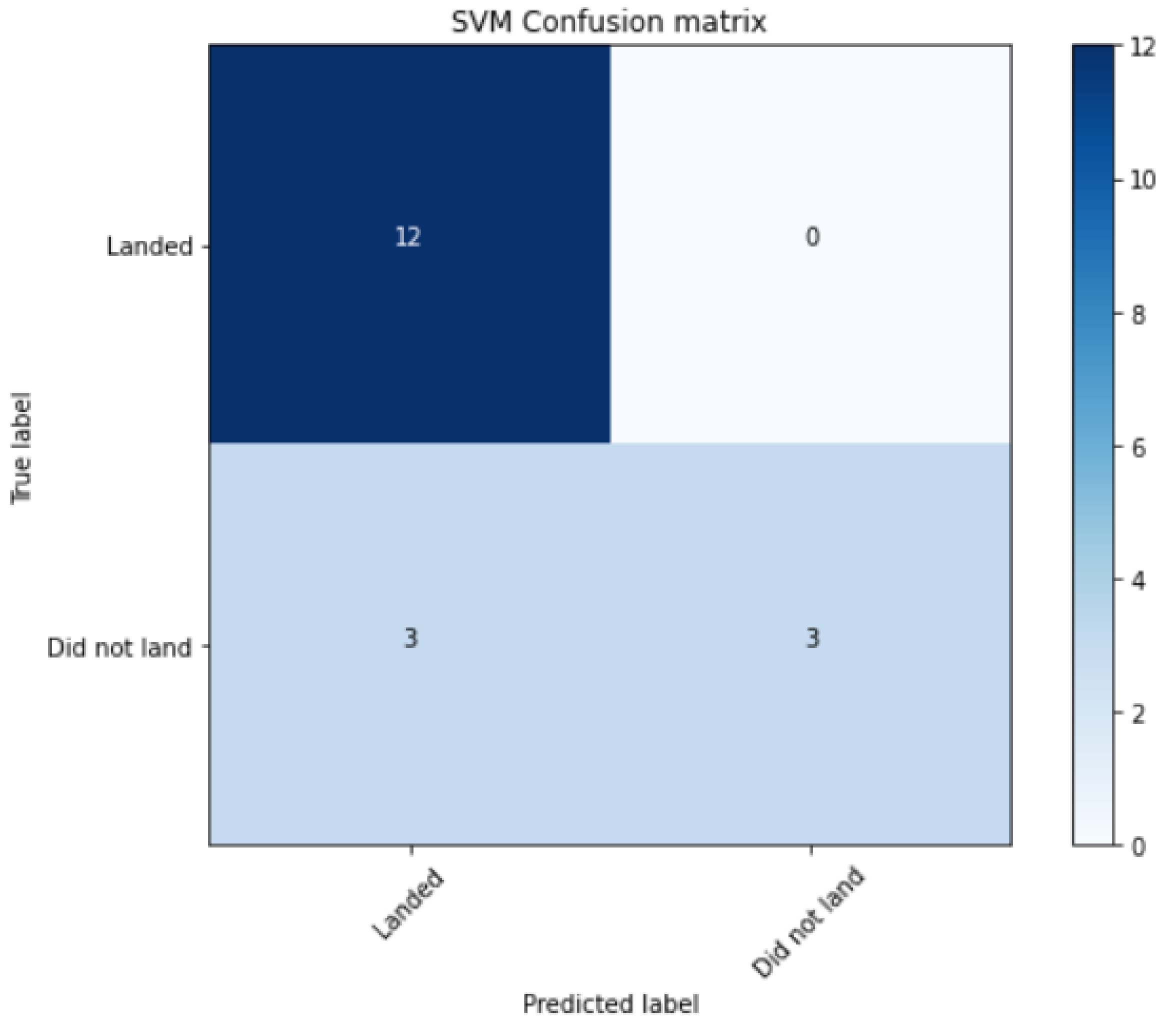
Classification Accuracy

- This bar chart shows the accuracy for each machine learning algorithm we tested.
- SVM, Logistic Regression and KNN classification have the highest classification accuracy.



Confusion Matrix

- This confusion matrix shows the best performing model with zero false positives as SVM.
- Though SVM, KNN, and Logistic have identical accuracy and precision. SVM has the best AUC predictive accuracy making it the best model.



Results

- SVM, KNN and Logistic Regression models are the best for at predictive analysis
- SPACEX has increased the amount of successful launches over the years
- The most successful launches have been from the KSC LC 339A launch site
- CCAFS LC-40 launch site has a success rate of 60 %, while KSC LC-39A and VAFB SLC 4E has a success rate of 77%
- Payloads of less weight will perform better overall
- FT boosters have the highest success rate for Falcon 9

Conclusion

- With our supervised machine learning model we know more about how to predict a successful launch for SPACEX
- From exploratory data analysis know more about what launch sites, orbit types, and booster versions have the highest success rate
- Now we can determining future launch outcomes which will help reduce cost on future space flights

Appendix

[GitHub Repository: All Labs and Notebooks for Capstone Presentaion]