SPACE-Y ANALYSIS OF FIRST STAGE RE-USE

Trevor Spraetz

1/18/2024

TABLE OF CONTENTS

2024

EXECUTIVE SUMMARY

Methodology

This research attempts to determine factors that contribute to a successful rocket landing and predict the outcome with a set of known variables. The following techniques were applied

- **Collect** data from the SpaceX REST API and web scrape Wikipedia
- **Wrangle** data into a more useful form with one-hot encoding
- **Explore** trends with data visualization across launch site, payload mass, and flight number
- **Analysis** of data in SQL to determine the total number of outcomes, total payload mass, and the maximum and minimum ranges for payload mass in successful missions
- **Explore** launch site locations and key landmarks traits of each site
- **Visualization** of the most successful launch sites, payload masses, and orbits
- **Construct Models** to predict the outcome of new missions using logistic regression, K-nearest neighbor(KNN), support vector machine (SVM), and decision tree

Results

Exploratory Data Analysis

- Site KCS LC-39A has the highest success rate
- Orbits ES-L1, GEO, HEO, and SSO have 100% success rates
- Success rates have increased over time

Folium Visualization

- Launch sites are close to the equator and coast lines

Model Outcomes

- Decision Tree modeling slightly outperformed all other options

INTRODUCTION

Background

SpaceX, a pioneer in space flight technology, has been striving to make commercial flight more readily available. By innovating the production process with old NASA technology, SpaceX had resources to research, test, and produce rockets on a quantity not seen before, allowing for rapid improvement on their first stage recovery technology. Because of this, SpaceX is able to offer launches with their Falcon 9 rocket for as low as $69 million whereas other competitors are upwards of $165 million. This competitive edge is what has put SpaceX at the forefront of the space travel industry. By determining the likelihood of recovering the first stage we can use this to determine the cost of running a launch, or whether to bid against SpaceX as a competitor. To do this, we will be using public data of past launches to feed into a machine learning model and then predict the outcome of several new cases

Objectives

- Map the effects of payload mass, launch site, flight number, and orbit destination of landing success
- Obtain the frequency of landing success over time
- Determine the best predictive model for landing outcome

METHODOLOGY

API DATA COLLECTION

- **Request** data from the SpaceX REST API and Wikipedia web scrape
- **Decode** the response json file using .json() then normalize with .json normalize()
- **Request** information from the API using custom functions
- **Create dictionary** from the API data
- **Store dictionary** in a dataframe for further processing
- **Filter Dataframe** to only contain Falcon 9 launches
- **Replace missing** payload mass values with .mean()
- **Export** the dataframe to a csv file

WIKI WEBSCRAPE

- **Request** data for Falcon 9 launches from Wikipedia
- **Use Beautiful Soup** to create an object from the HTML response
- **Collect** column names and load them into a data frame
- **Create dictionary** from parsing the HTML response
- **Store dictionary** in the dataframe for further processing
- **Export** the dataframe to a csv file

DATA WRANGLING

Steps

- **Perform EDA** and identify correlations between variables and landing outcomes
- **Calculate**
  - # and occurrence for each orbit
  - # and occurrence of mission outcomes for each orbit type
  - # of launches per site
- **Create binary** landing outcome column with one-hot encoding
- **Export** data to csv file

Landing Outcomes

Landings varied and are denoted by the following key:

- **True Ocean:** A successful landing in the ocean
- **True ASDS:** A successful landing on a drone ship
- **True RTLS:** A successful landing on a ground pad
- **False Ocean:** An unsuccessful landing in the ocean
- **False ASDS:** An unsuccessful landing on a drone ship
- **False RTLS:** An unsuccessful landing on a ground pad

In one-hot encoding, a 0 represented a failed landing whereas a 1 represented a successful landing

EDA VISUALIZATION

Charts

- Payload Mass (kg) vs Launch Site
- Payload Mass (kg) vs Orbit Destination
- Flight Number vs Payload
- Flight Number vs Launch Site

Objectives

- Identify correlations between variables using scatter plots and bar graphs. If a relationship does exist, utilize it in machine learning to construct a predictive model
- Use one-hot encoding to determine any relationships between variables such as variable fin usage with landing outcomes

EDA IN SQL

SQL Queries Outputs

- List of all launch site names
- 5 rows where the launch site began with CCA
- Total payload mass carried for Nasa
- Average payload mass for the F9 v1.1 booster

SQL Queries Outputs Continued

- Date of first successful landing on ground pad
- Booster names that succeeded landing on a drone ship with a payload between 4000-6000 kg
- Total number of successful and failed mission outcomes
- Which booster versions were launched with the maximum payload mass
- A count of landing outcomes between 6/4/2010 and 3/20/2017

FOLIUM MAPPING

Markers Designating Launch Sites

Markers placed indicating launch sites

- The Red circles indicate each **launch site** at their coordinates, including a popup label showing the site name

The Blue circle indicates **NASA Johnson Space Center** with a popup label displaying name

Markers Designating Launch Outcome

Markers were added in either green (**success**) or red (**failure**) to denote launch outcomes at each site and help determine overall success rates

Distance Calculations

Dashed lines were used to determine the distance between **CCAFS SLC-40** launch site and the nearest **railway**, **coastline**, **highway**, and **major city**

CONSTRUCTING A DASHBOARD WITH PLOTLY

Dropdown List with Launch Sites

- User can choose which launch site to pull data for

Pie Chart Illustrating Successful Launched

- Displays success rate of launches at selected site

Slider for Payload Mass Range

- Can specify payload mass range

Scatter plot for Payload Mass vs Success Rate by Booster

- Illustrates relationship between Payload Mass and success rate

PREDICTIVE MODEL DEVELOPMENT

1. Created a NumPy array from the class column
2. Standardized the data with StandardScalar. Fit and transformed data
3. Split the data into test/train sets
4. Created GridSearchCV object where cv=10
5. Applied the object to the following models
    1. Logistic Regression
    2. Support Vector Machine
    3. K-Nearest Neighbor
    4. Decision Tree Classifier
6. Determined the accuracy score for each model
7. Calculated the confusion matrix for each model
8. Determined the most accurate model using three models: Jaccard Score, F1 score, and overall accuracy form the confusion matrix

RESULTS

SUMMARY OF DATA ANALYSIS

General EDA

- Landing success has increased over time
- KSC LC-39A has the highest success rate among landing sites. It saw an increase in use at the beginning, followed by a decrease, then an increase in use again
- Orbits ES-L1, SSO, HEO, and GEO have a 100% success rate

Visual Analysis

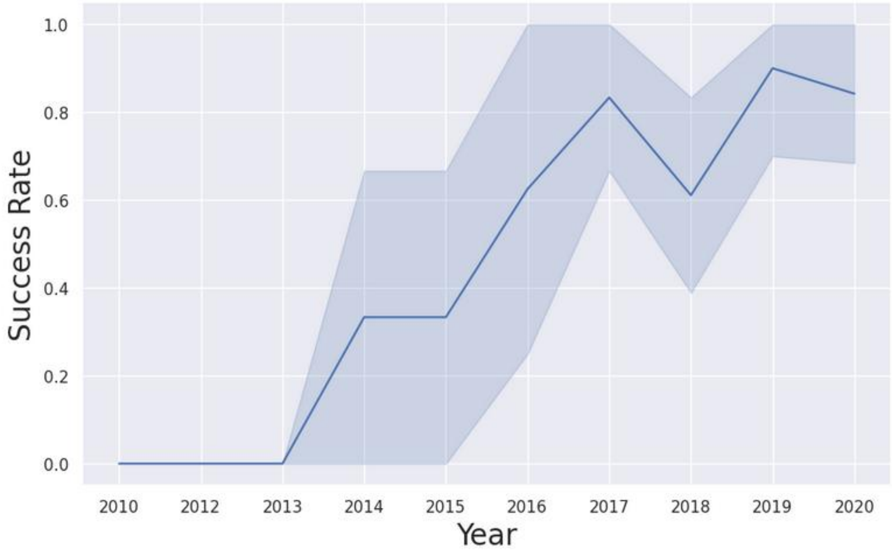
- Launches are done close to the equator, near coastlines
- Launch sites are remote enough to mitigate disaster in the event of a failed launch

Predictive Analysis

- The Decision Tree model had the highest overall performance. Logistical Regression had the lowest.

OVERALL LANDING PERFORMANCE

- The landing success rate increased from 2013-2017 and again from 2018-2019
- Decreases in landing success rate from 2017-2018 and 2019-2020
- Landing success rate is up 233% from the middle of 2013

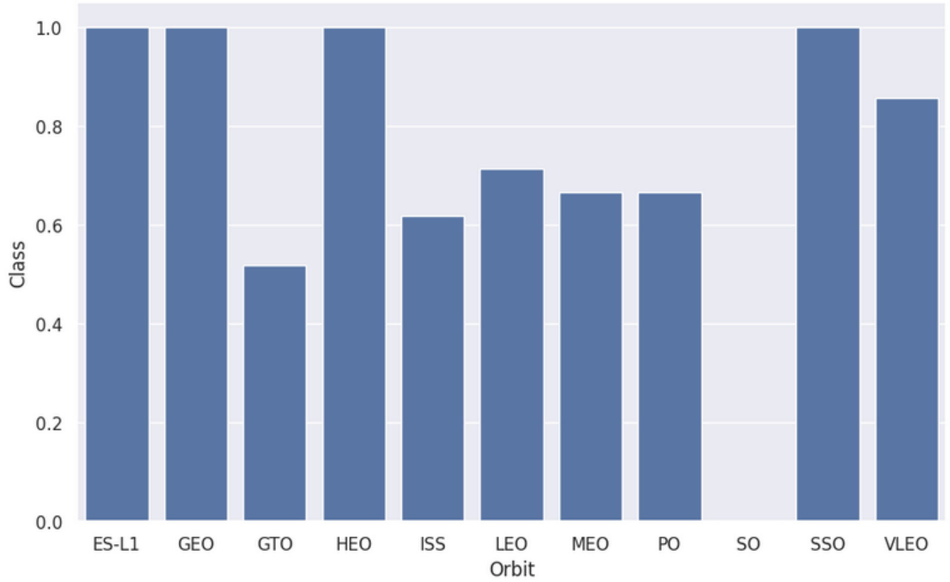


PAYLOAD MASS AND LAUNCH SITES

- **No correlation** between **launch sites** producing successful landings for any **payload mass** threshold
- Positive correlation between **payload mass** and **success rate**
- **>87% success rate** on launches with **>7000 kg** payload mass
- KSC LC 39A has a 100% landing success rate for launches under 5000 kg

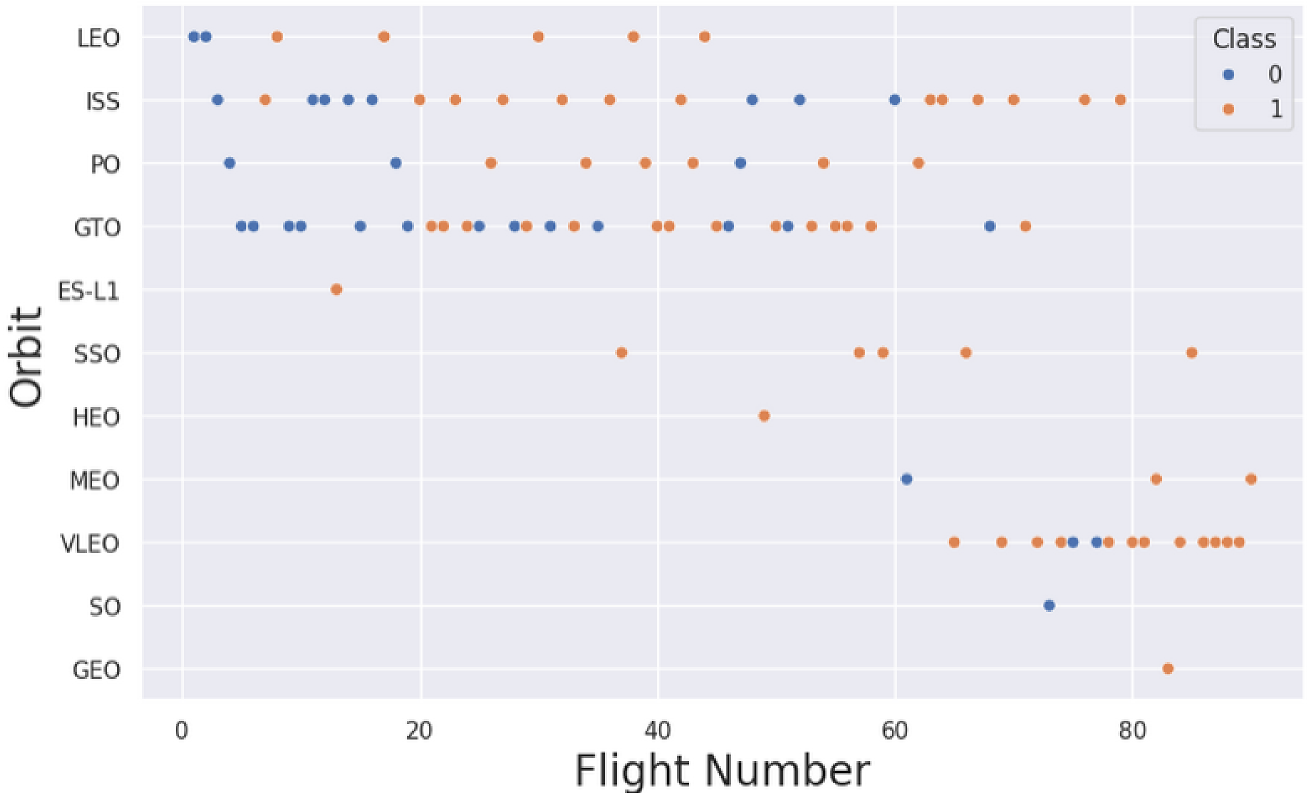
ANALYSIS OF ORBITS

- 100% Success Rate: GEO, HEO, SSO, ES-L1
- 50%-85% Success Rate: GTO, ISS, LEO, MEO, PEO
- 0% Success Rate: SO



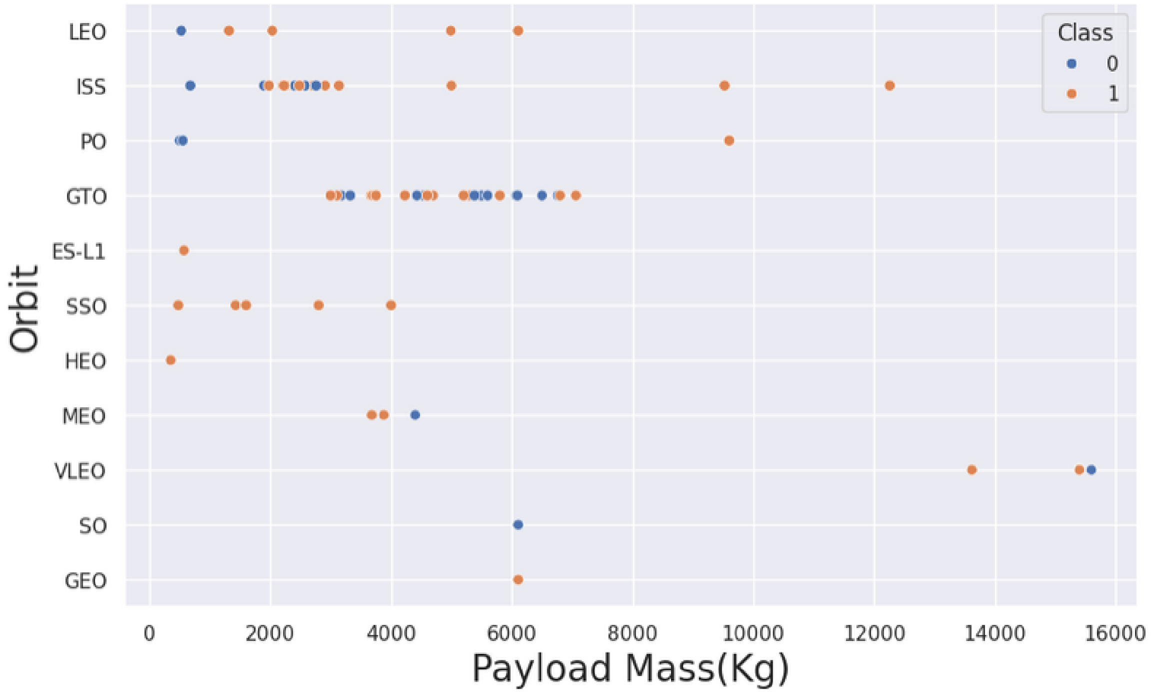
FLIGHT NUMBER AND ORBIT

- Success increases over time with each orbit
- ISS had the greatest recursion of flight failures
    - Flights with lesser payload mass tend to be associated with these failures
- SO has only been attempted once (unsuccessful)
- GTO shows a consistent rate of failures over time



PAYLOAD AND ORBIT

- ISS, PO, and LEO show increased success chance with heavier payloads
- More data is needed to determine the effects of payload mass on success for GTO, HEO, and MEO



LAUNCH SITE INFORMATION

Launch Site Names

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

SQL Launch Site Query

```
[32]: %sql SELECT DISTINCT(LAUNCH_SITE) FROM SPACEXTBL;
 * sqlite:///my_data1.db
Done.
```

[32]:

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

CCA Launch Site Records

- Showing all records containing CCA in the launch site, limited to 5 records

```
[17]: %sql SELECT * FROM SPACEXTBL WHERE LAUNCH_SITE LIKE 'CCA%' LIMIT 5;
 * sqlite:///my_data1.db
Done.
```

[17]:

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

PAYLOAD MASS STATISTICS

Total Payload Mass

- **45,596 kg** in total carried for NASA launches

```
[34]: %sql SELECT SUM(PAYLOAD_MASS__KG_) \
FROM SPACEXTBL \
WHERE Customer ='NASA (CRS)';

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

[34]:

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass

- **2,928 kg** was the **average** payload for the Falcon 8 v1.1 rocket

```
[36]: %sql SELECT AVG(PAYLOAD_MASS__KG_) \
FROM SPACEXTBL \
WHERE BOOSTER_VERSION = 'F9 v1.1';

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

[36]:

AVG(PAYLOAD_MASS__KG_)
2928.4

MISSION MARKERS & MILESTONES

First Successful Ground Pad Landing

- 22 December 2015

```
[37]: %sql SELECT MIN(DATE) \
FROM SPACEXTBL \
WHERE LANDING_OUTCOME = 'Success (ground pad)';
```

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

[37]:

MIN(DATE)
2015-12-22

Drone Ship Landing

- Successful landings with payload mass between 4000 kg and 6000 kg include JSCAT-14, JSCAT-16, SES-10, SES-11, and EchoStar 105

```
[38]: %sql SELECT BOOSTER_VERSION \
FROM SPACEXTBL \
WHERE LANDING_OUTCOME = 'Success (drone ship)' \
AND PAYLOAD_MASS__KG_ >4000 AND PAYLOAD_MASS__KG_ < 6000;
```

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

[38]:

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

Total Outcomes

- 1 failure in flight
- 99 successes
- 1 success with unclear payload status

```
[40]: %sql SELECT MISSION_OUTCOME, COUNT(*) as total_number \
FROM SPACEXTBL GROUP BY MISSION_OUTCOME;
```

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

[40]:

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

BOOSTER PERFORMANCE

Maximum Payload Mass Achieved

- 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

```
[42]: %sql SELECT BOOSTER_VERSION \
FROM SPACEXTBL \
WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) \
FROM SPACEXTBL);
```

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

[42]:

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

FAILED LANDINGS ON DRONE SHIP

- There were two failed landing attempts on drone ships in 2015
  - SQL was used to obtain the month, date, booster version, and launch site of the failed landings

```
[43]: %sql SELECT substr(Date,6,2) as month, DATE, LANDING_OUTCOME, LAUNCH_SITE, BOOSTER_VERSION \
FROM SPACEXTBL WHERE LANDING_OUTCOME='Failure (drone ship)' AND substr(Date,0,5)='2015'

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

[43]:

month	Date	Landing_Outcome	Launch_Site	Booster_Version
01	2015-01-10	Failure (drone ship)	CCAFS LC-40	F9 v1.1 B1012
04	2015-04-14	Failure (drone ship)	CCAFS LC-40	F9 v1.1 B1015

COUNT OF LANDING OUTCOMES

Ranked Descending

- Landing outcomes between June 4th, 2010, and March 20th, 2017, were ranked in the SQL database in descending order

```
[46]: %sql SELECT LANDING_OUTCOME, COUNT(*) as Total_Counts \
FROM SPACEXTBL WHERE DATE between '2010-06-04' and '2017-03-20' \
GROUP BY Landing_Outcome order by Total_Counts DESC;

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

[46]:

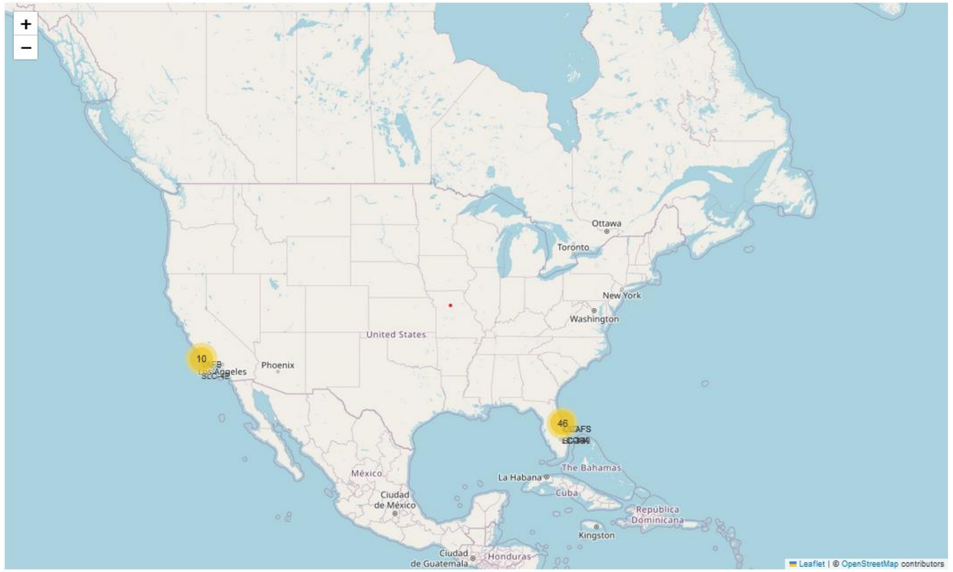
Landing_Outcome	Total_Counts
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

LAUNCH SITE ANALYSIS

LAUNCH SITES

Key Features

- **Near the equator** to maximize initial velocity about the earth's axis. This positioning **provides a boost** at take-off, **reducing fuel** consumption



LAUNCH MARKERS

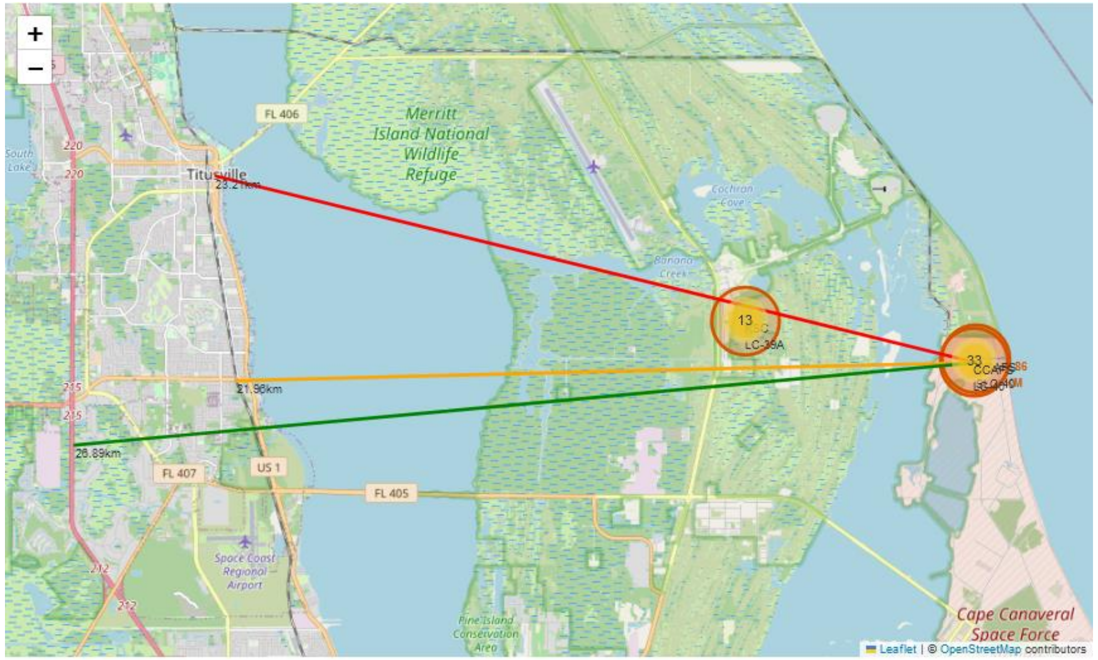
At Launch Sites

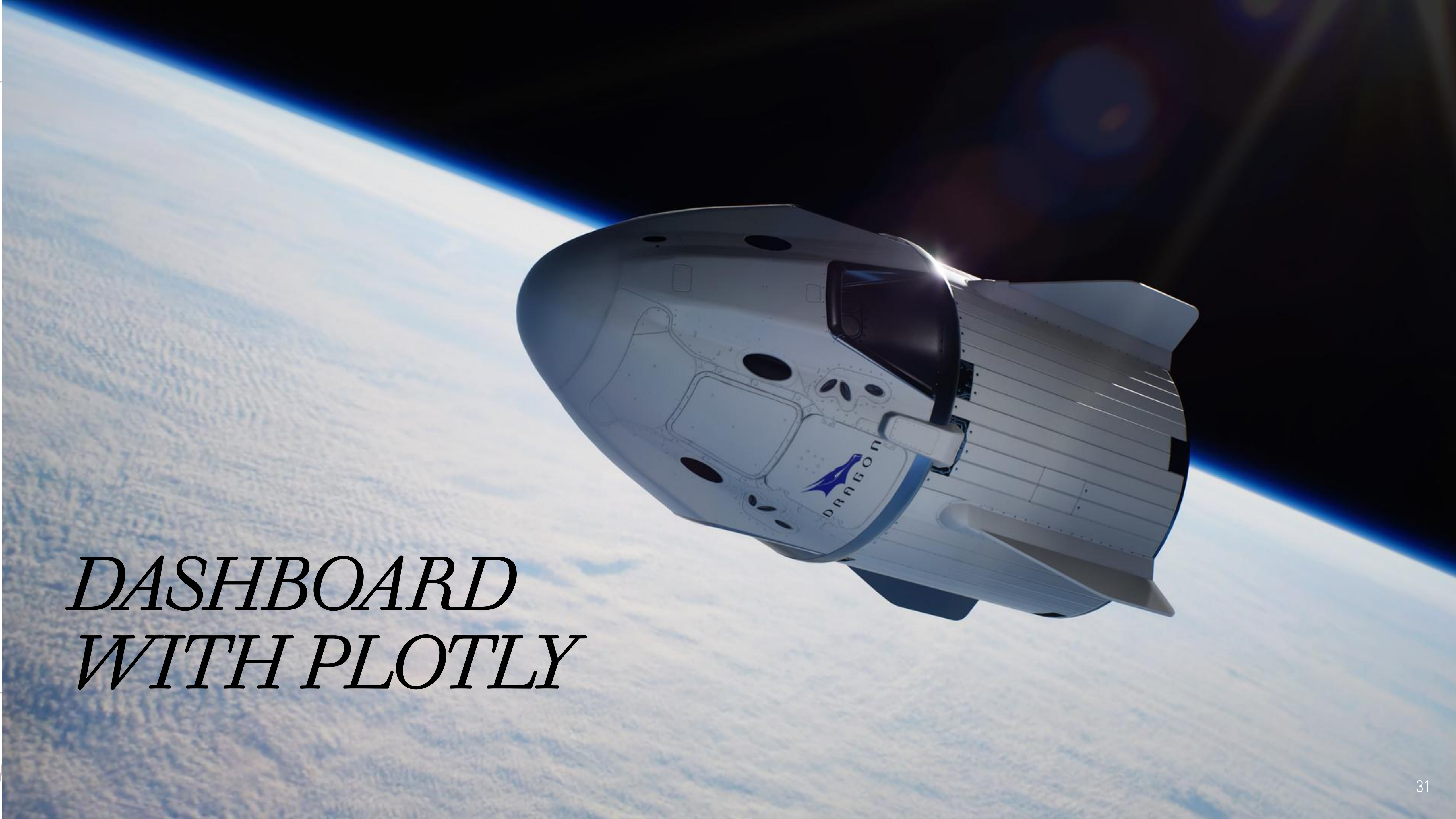
- Green markers represent successful launches
- Red markers represent unsuccessful launches
- Illustrated below, CCAFS SLC-40 has had 3 out of 7 successful launches (42.9%)

NOTABLE LANDMARKS

CCAFS SLC-40

- The **coast** is **0.86 km** away and provides a wide area for falling debris in the event of an accident or stage failure, or simply a place for a non-reusable stage to land
- The nearest **highway** is **21.96 km** and **railroad** is **26.89 km** away while the nearest **city** is **23.21 km** away
    - It's important for infrastructure to be a minimum distance away to reduce the risk of damage from falling rocket debris
    - This also minimizes the chance of harm to individuals

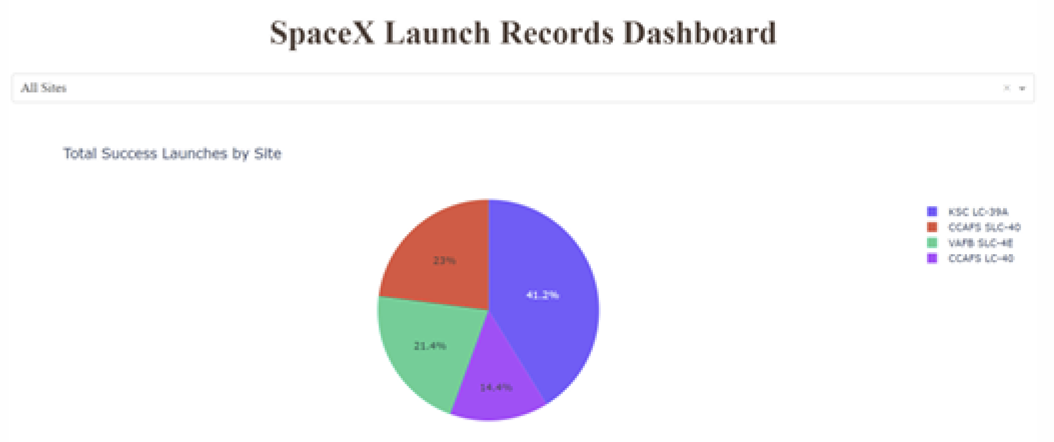




DASHBOARD WITH PLOTLY

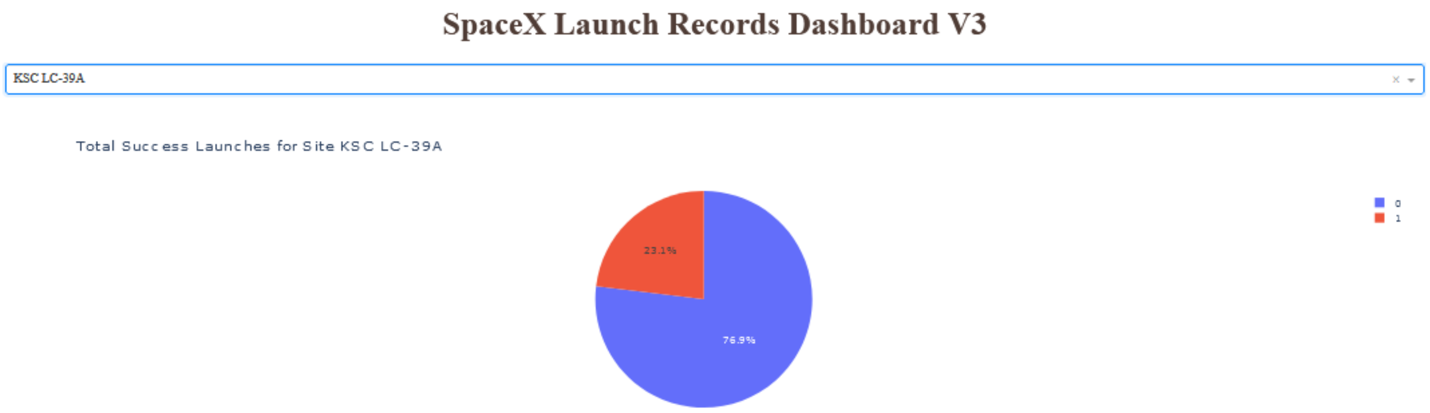
SUCCESS RATE BY SITE

- **KSC LC-29A** had the highest rate of landing success rate, constituting **41.2%** of total successful landings



LAUNCH SITE KSC LC-29A DETAILS

- **KSC LC-29A** had a **76.9%** success rate
  - 10 out of 13 launches succeeded



SUCCESS BY BOOSTER AND MASS

- Payloads between **2,000 kg** and **5,000 kg** had the highest chance of success
- Success rate increased across iteration of boosters

PREDICTIVE MODELING

CLASSIFICATION ACCURACY

- All models performed to the same degree with identical scores across the F1 test, the Jaccard test, and overall accuracy
- The best_score average of all three models was slightly better for the **decision tree** model

	LogReg	SVM	Tree	KNN
Jaccard_Score	0.800000	0.800000	0.800000	0.800000
F1_Score	0.888889	0.888889	0.888889	0.888889
Accuracy	0.833333	0.833333	0.833333	0.833333

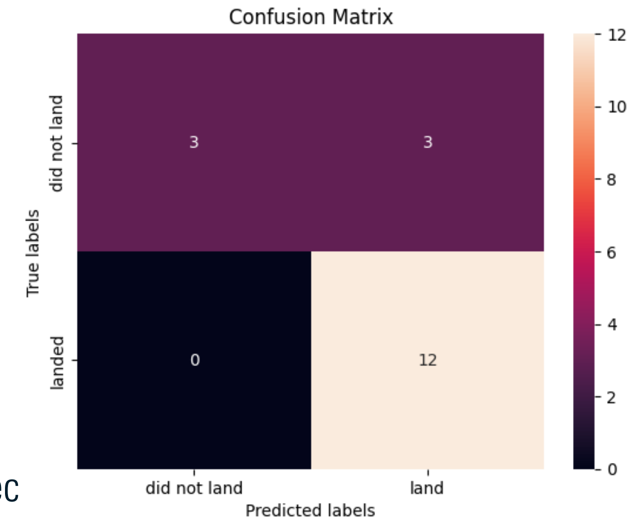
```
models = {'KNeighbors':knn_cv.best_score_,
              'DecisionTree':tree_cv.best_score_,
              'LogisticRegression':logreg_cv.best_score_,
              'SupportVector': svm_cv.best_score_}

bestalgorithm = max(models, key=models.get)
print('Best model: ', bestalgorithm,'\n','Score: ', models[bestalgorithm])
if bestalgorithm == 'SupportVector':
    print('Best parameters is :', svm_cv.best_params_)
if bestalgorithm == 'DecisionTree':
    print('Best parameters is :', tree_cv.best_params_)
if bestalgorithm == 'KNeighbors':
    print('Best parameters is :', knn_cv.best_params_)
if bestalgorithm == 'LogisticRegression':
    print('Best parameters is :', logreg_cv.best_params_)
```

```
Best model:  DecisionTree
 Score:  0.875
Best parameters is : {'criterion': 'entropy', 'max_depth': 4, 'max_features': 'sqrt', 'min_samples_leaf': 2, 'min_samples_split': 10, 'splitter': 'random'}
```

CONFUSION MATRIX

Performance Summary

- The matrix totals each outcome for a predictive model (true positive, true negative, false positive, false negative)
- 72 training rows, 18 test rows
- Confusion matrices for each model were identical
- Confusion Matrix Results:
    - 12 True Positive
    - 3 Ture Negatives
    - 3 False Positives*
    - 0 False Negative



- **Recall** = TP / (TP + FN)
    - 12 / 12 = 1
- **Precision** = TP / (TP + FN)
    - 12 / 15 = 0.80
- **F1 Score** = 2 * (Recall * Precision) / (Recall + Prec
    - 2*(1*0.8)/(1+0.8) = 0.89
- **Accuracy** = (TP + TN) / (TP + TN + FP + FN)
    - (12 + 3)/(12+3+3+0)=0.833

- *False Positives present a risk for the model to incorrectly assume a successful landing (see more in conclusion)

CONCLUSION

Research

- **Model Performance** is relatively similar amongst all models with decision tree slightly outperforming the rest
- **Location** for the launch sites is best along the equator to provide an optimal amount of initial kinetic energy
- **Launch Success** has increased over time with technology improvement
- **KSC LC-29A** had the highest success rate among any launch site
- **Orbits** ES-L1, GEO, HEO, and SSO have a 100% success rate
- **Payload Mass** typically increased with increasing landing success chances

CONCLUSION

Recommendations

Dataset: More data will always lead to more accurately trained models. This project used 72 training points and 18 testing points

Feature Analysis: More factors could be used to more accurately predict rocket landing outcomes. Private data not released by Tesla could make meaningful contributions to model accuracy

Model Accuracy: It is concerning to have any false positives in the predictive model as it could lead a bidder to over-compete with Tesla or for Tesla to under-project costs for a launch bid