IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Jacob Ahid
June 14, 2023

Outline

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

Executive Summary

Summary of methodologies

The provided research is intended to predict the outcomes of SpaceX's Falcon 9 rocket's landing results.
The following methodologies were used as follows:

- **Data Collection** via web-scraping and SpaceX REST API
- **Data Wrangling** to preform Exploratory Data Analysis (EDA) and determine training labels
- **Exploratory Data Analysis** via SQL to gain understanding of dataset
- **Data Visualization** via Plotly and Folium to garner correlation within the dataset
- **Model** the data to predict the outcomes of the rocket flights via various Machine Learning Algorithms

Summary of all Results

- Nearly all mission outcomes were successful
- Launch sites are located near coastal regions and away from major cities, highways, and railroads
- All models resulted in similar accuracy, with decision tree model performing slightly greater than others

X

Introduction

- Project background and context
- Problems you want to find answers

Project Background and Context

Space X offers significantly lower costs for its Falcon 9 rocket launches compared to other providers, primarily because they can reuse the first stage. To determine the cost of a launch, it is essential to predict if the first stage will land. This prediction can be used by alternate companies bidding against Space X for rocket launches. To make these predictions, a machine learning pipeline will be created using data from public sources to predict the first stage landing outcome, enabling more accurate cost estimation and informed decision-making for competitive bidding.

Main Problems

What factors correlate most to the success or failure of launches?
What Machine Learning model will provide the greatest accuracy for predicting future launch results?
What are the historical results of the SpaceX missions?

Section 1

Methodology

Methodology

- Executive Summary
- Data collection methodology:
  - Data was collected from Wikipedia and SpaceX REST API
- Perform data wrangling
  - Data was filtered to only contain relevant values and was prepared for ML models via one-hot encoding
- 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

- Describe how data sets were collected.
- You need to present your data collection process use key phrases and flowcharts

Data Collection – SpaceX API

See Code Here

Call Data from REST API

Decode to .json

Convert .json to Pandas dataframe

Clean and filter Data

Export to CSV

Data Collection - Scraping

See code here

Request Falcon 9 launch Wiki page

Extract column names from Wiki table header

Create Dictionary and Parse the HTML tables to create Dataframe

Export to CSV

Data Wrangling

See code here

EDA – Identified % of null and datatypes

Calculated:

- number of launches on each site
- number and occurrences of each orbit
- number and occurrences of mission outcome per orbit type

Created a classification variable that represents the outcome of each launch (Boolean values)

Determined the success rate of Dataset

Exported to CSV

EDA with Data Visualization

See code here

Scatterplot of Flight Number vs Payload Mass and Class (outcomes) as the hue

- Plotted to identify how the variables would affect the launch outcome

Scatterplot of Flight Number vs Launch Site and Class (outcomes) as the hue

- Plotted to analyze the success rates of the different launch sites

Bar chart of the success rate of each orbit

- Plotted to identify and compare the success rates of each orbit

Scatterplot of Orbit vs Payload Mass and Class (outcomes) as the hue

- Plotted to analyze the success rates of different orbits based on the weight of the payload

Line plot of Class (outcome) by Date

- Plotted to analyze the success rate over time

EDA with SQL

See code here

- Identified the different launch sites
- Explored the total and average payload masses of different boosters
- Identified the first successful landing in ground pad
- Named the successful ground pad boosters within the payload weights of 4,000 and 6,000
- Counted the total number of successful and failed missions
- Identified the boosters who have carried the maximum payload weight
- Identified the drone ship failures in 2015
- Ranked the number of successful landing outcomes from 2010 -2017

Build an Interactive Map with Folium

See code here

Added a folium circle and marker for each launch site to identify the locations of the different sites

Added a marker cluster for each launch result to map the different launches and compare success rates

Implemented color-coding (red and green) to each launch site to further analyze and compare success rates

Drew polylines to map the distance from the launch sites to the nearest landmarks (coast, city, highway, railway)

Build a Dashboard with Plotly Dash

See code here

Created a launch site drop-down lists to filter the exhibits by different sites

Created a Pie Chart to view the percentage of successful and successful launches

Created a slider to filter exhibits by payload mass

Created a scatter plot on payload mass vs success ratio to highlight how payload mass correlates to successful launches

Predictive Analysis (Classification)

See code here

Use train_test_split to create training and testing data to evaluate model on unforeseen data

Create 4 separate ML models objects (Logistic Regression, SVM, Decision Tree, and KNN)

Create GridSearchCV objects and fit it to find the best parameters for each ML model

Calculated the accuracy of each model and plotted on a confusion matrix

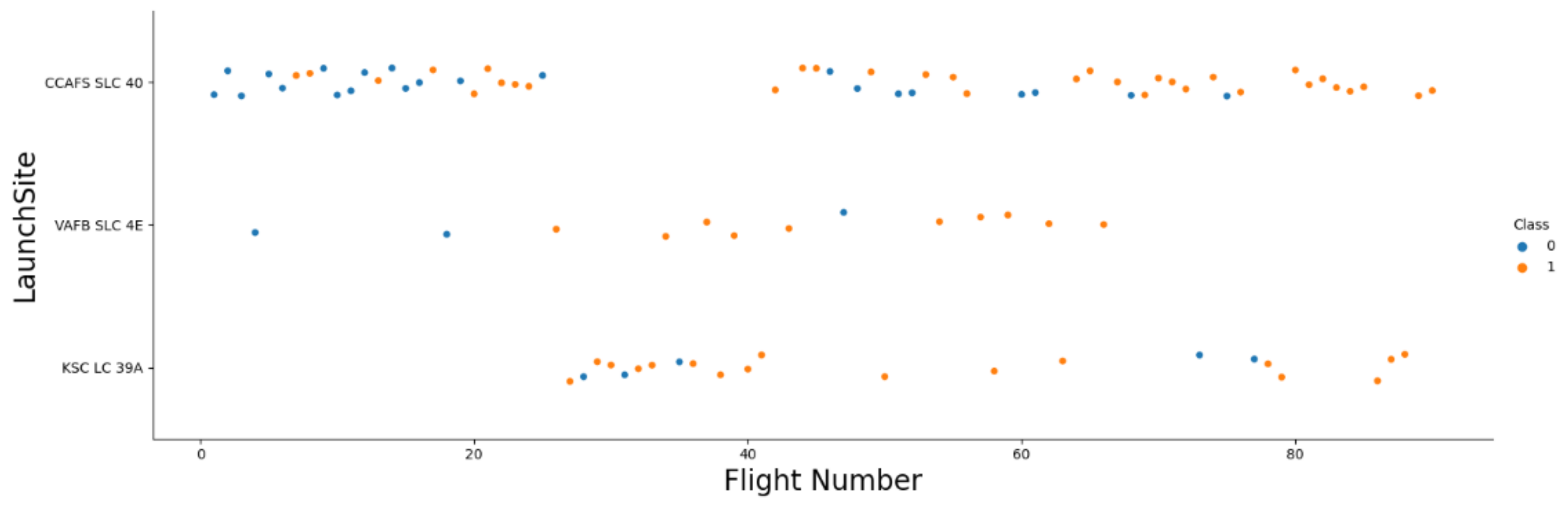
Determined the best performing model using scoring metrics

Results

- Exploratory data analysis results
  - All outcomes were successful for ES-L1, GEO, HEO and SSO orbits
  - Success Rate has increased significantly over time
  - KSC has the best success rate overall
  - CCAFS SLC 40 has the highest volume of launches
  - VAFP has the least amount of failures
- Predictive analysis results
  - All models performed with nearly identical accuracy, but based on scoring metrics the Decision Tree model was overall the best model for the data

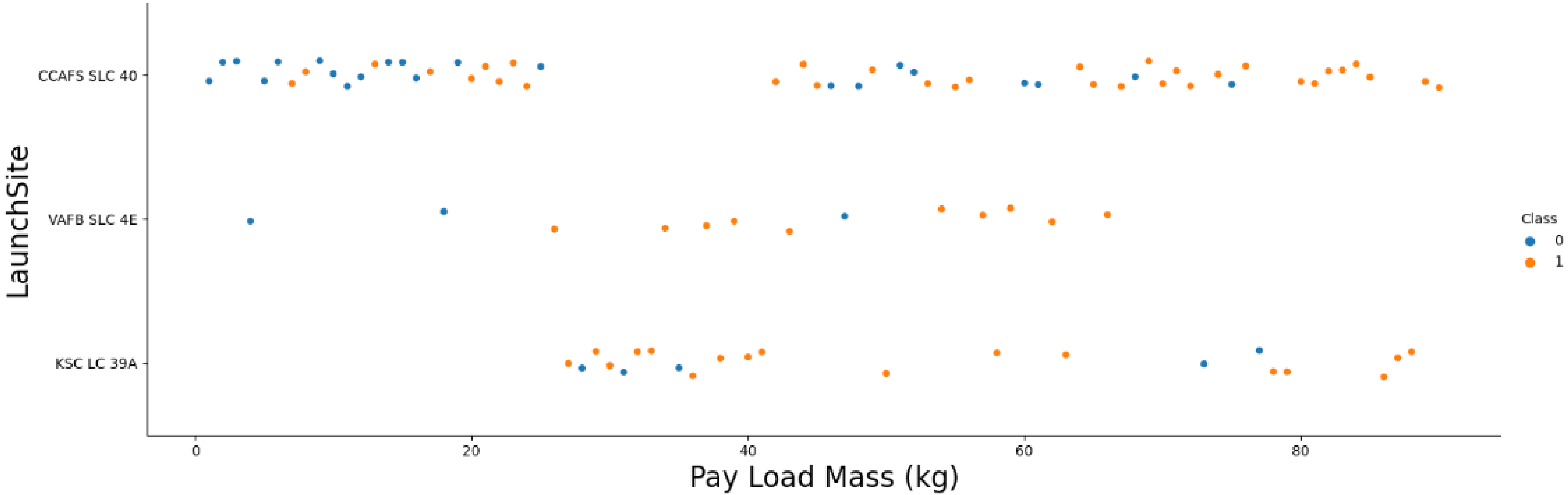
Section 2

Insights drawn from EDA



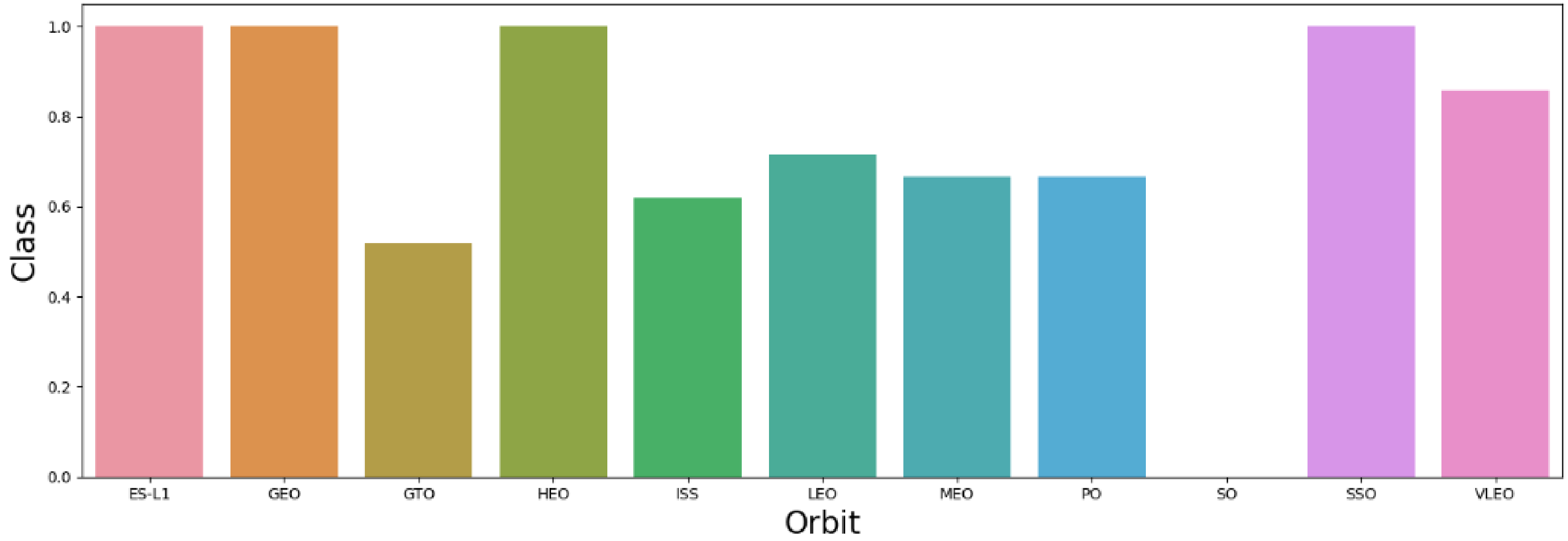
Flight Number vs. Launch Site

- Orange = Success, Blue = Failure
- Higher flight numbers tend to have greater success
- CCAFS has the highest volume
- VAFB has the contains the lowest number of failures
- KSC has the best overall success rate



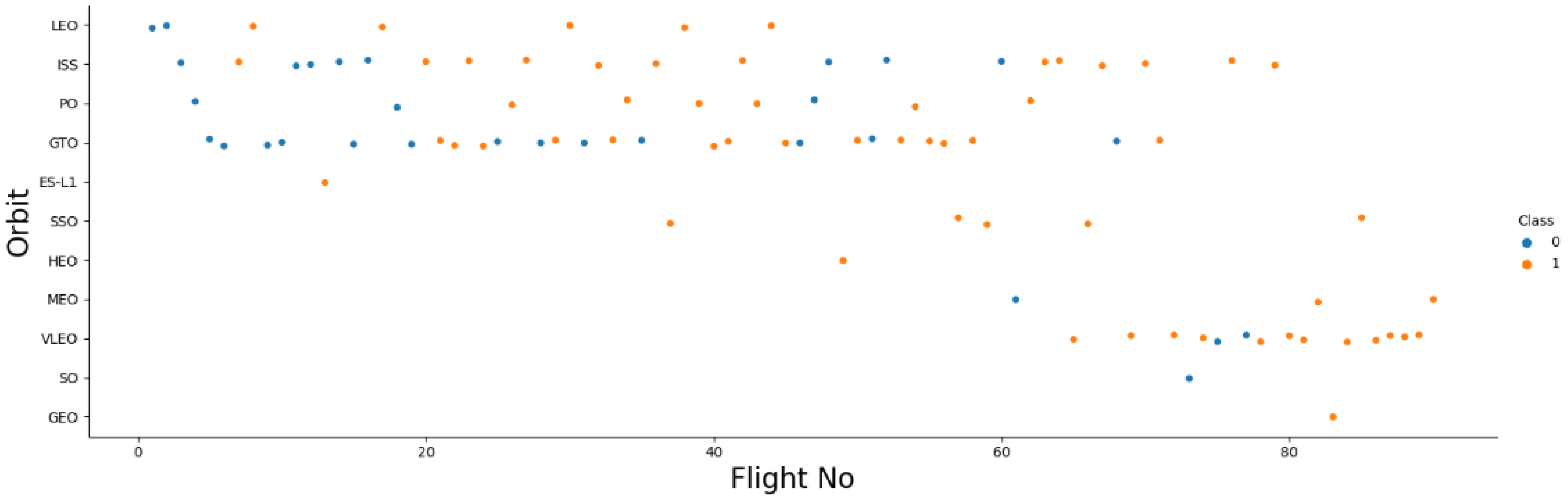
Payload vs. Launch Site

- Orange = Success, Blue = Failure
- VAFB-SLC launchsite had no rockets launched for heavy payload mass(greater than 10000)
- Heavier payload mass correlates to greater success
- Launches about 80k had a success rate of 100%



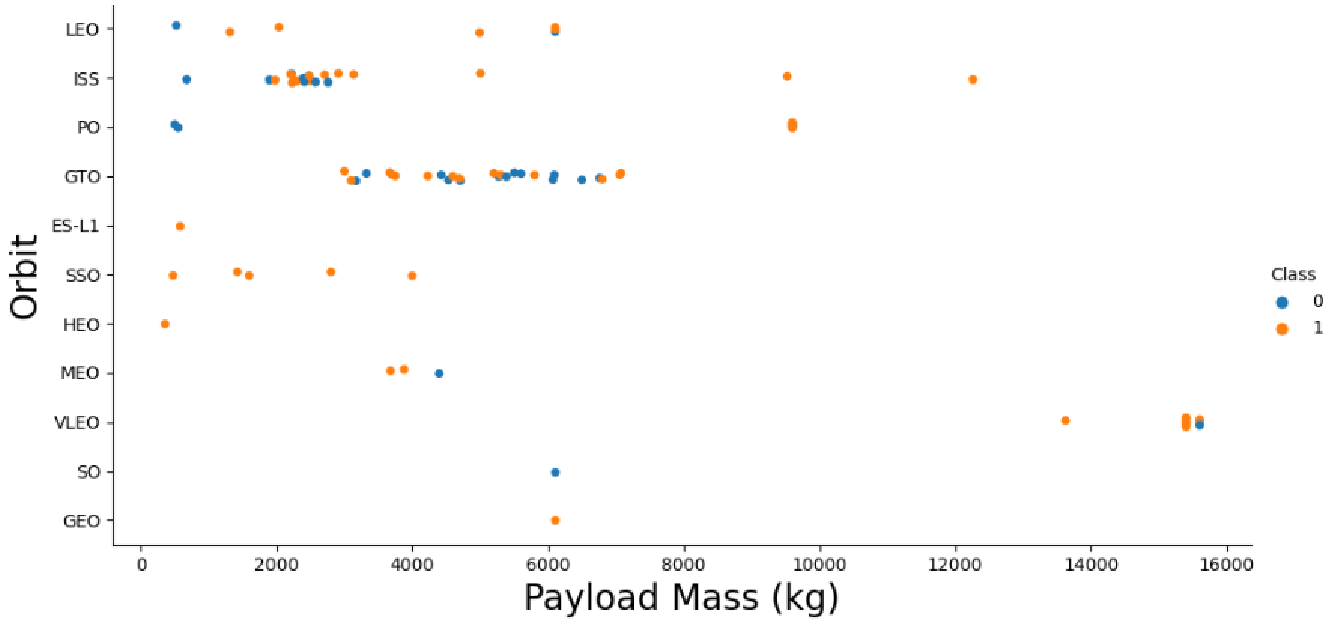
Success Rate vs. Orbit Type

- ES-L1, GEO, HEO, and SSO all maintain a 100% success rate
- SO has a 0% success rate
- The remaining orbits have a success rate between around 45% and 75%



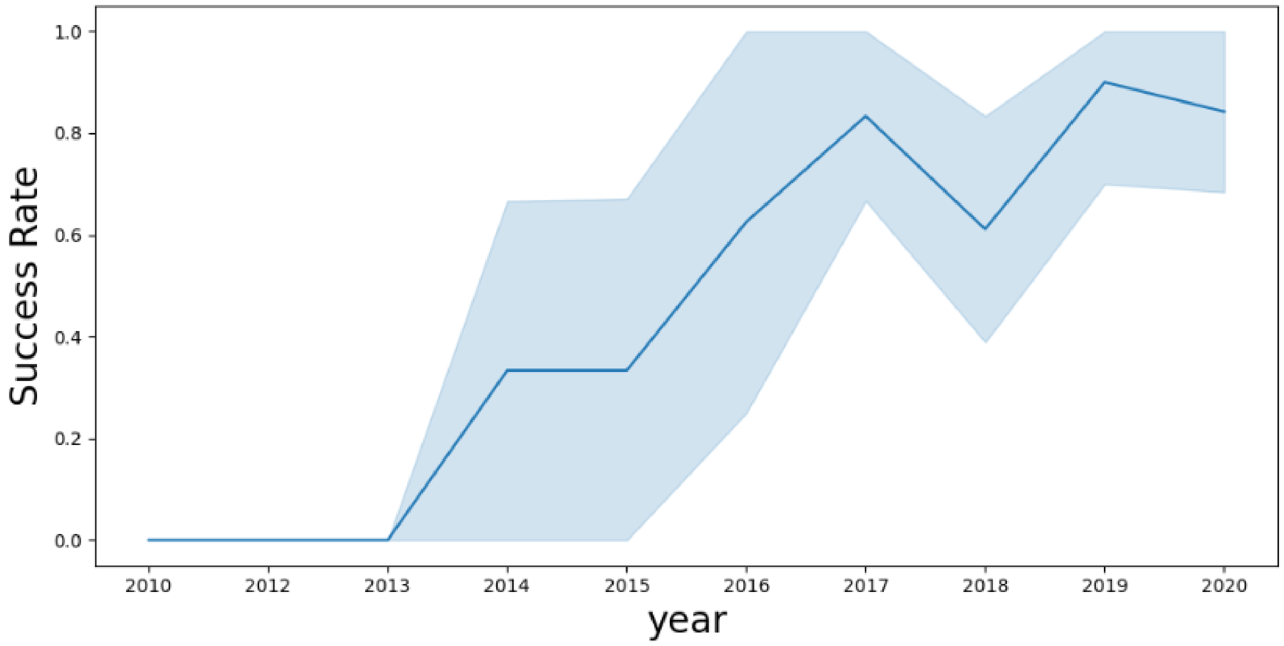
Flight Number vs. Orbit Type

- Higher flight numbers correlate to higher success rates for most orbits
- VLEO has a 100% success rate above around 78%



Payload vs. Orbit Type

- Successful landing rates are high for Polar, LEO and ISS when payloads are greater
- No valuable insights can be gained from the GTO results as a large mix of successful and failed landings are both grouped in the same payload range



Launch Success Yearly Trend

- Successful landings consistently increased from 2013-2020, with a slight dip in 2018
- The highest success rate occurred in 2019

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

All Launch Site Names

- Simple select distinct statement to identify all of the launch sites in the database

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome
06/04/2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0.0	LEO	SpaceX	Success
12/08/2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0.0	LEO (ISS)	NASA (COTS) NRO	Success
22/05/2012	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525.0	LEO (ISS)	NASA (COTS)	Success
10/08/2012	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500.0	LEO (ISS)	NASA (CRS)	Success
03/01/2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677.0	LEO (ISS)	NASA (CRS)	Success

Launch Site Names Begin with 'CCA'

- Launch info on 5 CCAFS LC-40 launch sites

```
%sql select sum(PAYLOAD_MASS__KG_) total_payload_mass \
     from spacextbl \
     where customer = 'NASA (CRS)' ;
```

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

total_payload_mass
45596.0

```
%sql select AVG(PAYLOAD_MASS__KG_) avg_payload_mass \
     from spacextbl \
     where booster_version =  'F9 v1.1' ;
```

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

avg_payload_mass
2928.4

Total and Average Payload Mass

- The total payload mass for NASA is 45,596 KG
- The average payload mass for the F9 v1.1 is 2824KG

```
%sql select min(date) first_successful_landing_date  from spacextbl \
    where landing_outcome = 'Success (ground pad)';
```

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

first_successful_landing_date
01/08/2018

First Successful Ground Landing Date

- The first successful ground landing date was January 8th, 2018

```
%sql select payload  from spacextbl \
     where landing_outcome = 'Success (drone ship)'\
     and PAYLOAD_MASS__KG_ between 4000 and 6000;
```

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

Payload
JCSAT-14
JCSAT-16
SES-10
SES-11 / EchoStar 105

Successful Drone Ship Landing with Payload between 4000 and 6000

- The successful drone landing ships with payload masses between 4,000 and 6,000 kg were the JCSAT-14 , the JCSAT-16, the SES-10, and the SES-11 / EchoStar 105

```
%sql select mission_outcome, count(*) as no_of_missions from spacextbl\
    where mission_outcome <> 'None' \
    group by mission_outcome order by count(*) desc
```

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

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

Total Number of Successful and Failure Mission Outcomes

- 100 successes (with one having an unclear payload status
- 1 failure in flight

```
%sql select booster_version from spacextbl\
    where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_)from spacextbl);
```

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

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

Boosters Carried Maximum Payload

- 12 boosters carried the max payload

```
%sql select substr(Date, 4, 2) as monthnames, landing_outcome, booster_version, launch_site from spacextbl \
    where landing_outcome = 'Failure (drone ship)' and substr(Date,7,4)='2015';
```

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

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

2015 Launch Records

- 2 failed drone ships (F9 v1.1 B1012 and B10125) at launch site CCAFS LC-40 in April and October of 2015

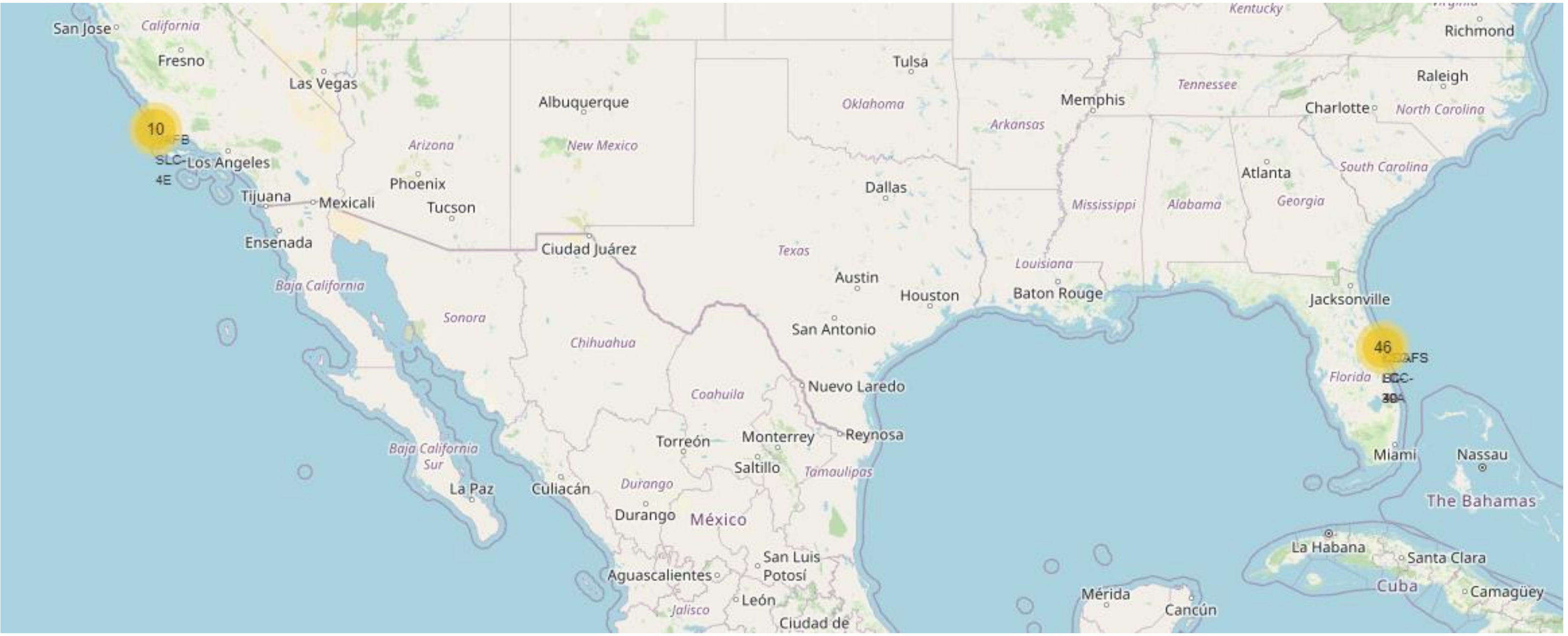
Landing_Outcome	num_of_outcomes
Success	20
No attempt	10
Success (drone ship)	8
Success (ground pad)	7
Failure (drone ship)	3
Failure	3
Failure (parachute)	2
Controlled (ocean)	2
No attempt	1

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

- Total number of outcomes by the landing outcome (35 total successes, 8 failures, 11 without attempts)

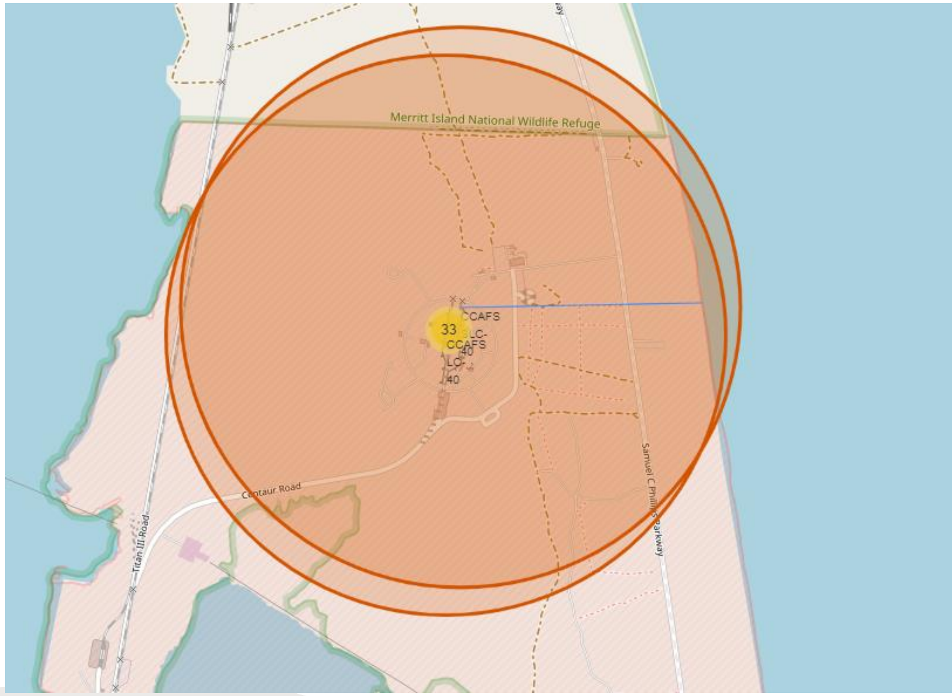
Section 3

Launch Sites Proximities Analysis



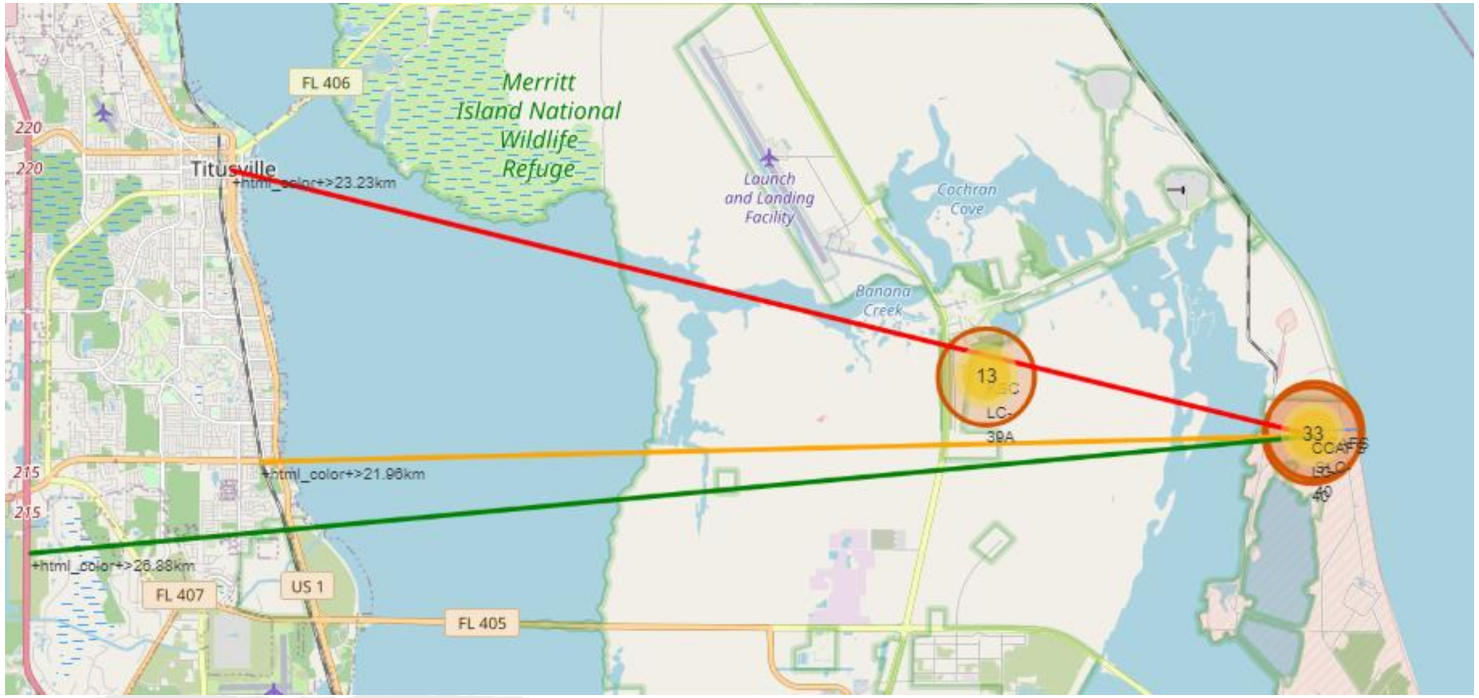
Launch Site Locations

- Launch sites are found in coastal regions and close to the equator



Launch Site Proximity to Coastline

- Launch site is extremely close (0.863km) to coastline



Launch Site Proximity to Other Landmarks

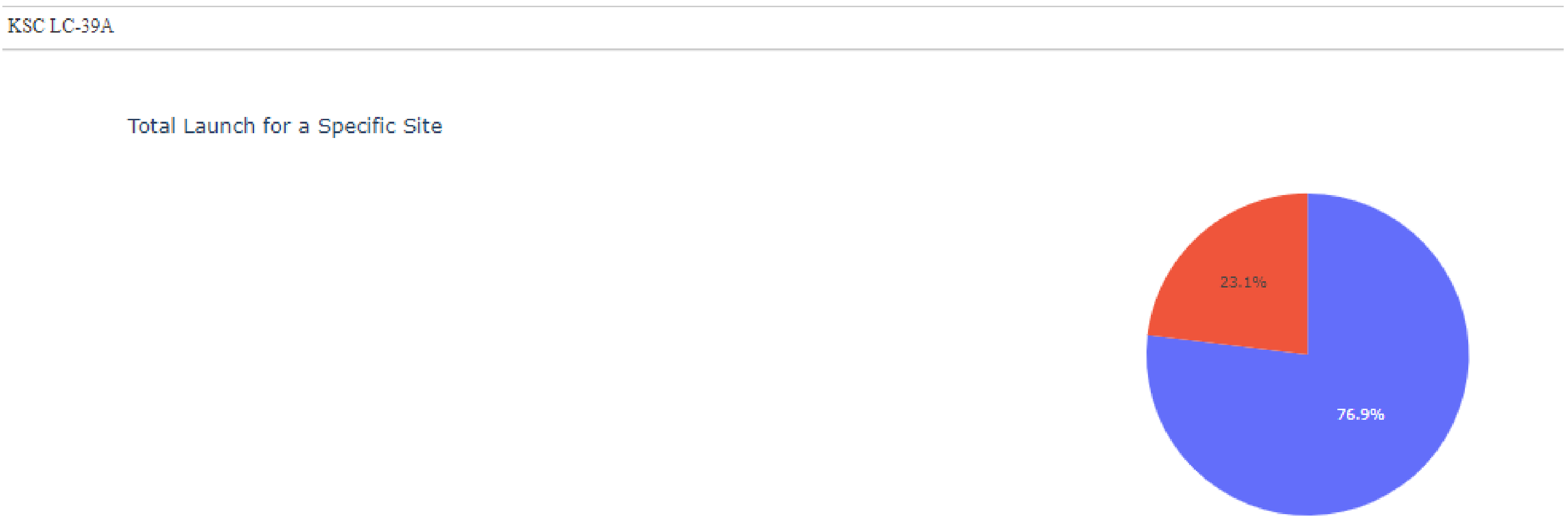
- Launch sites are not in close proximity to nearest railway
  - ex CCAFS is 22km from nearest railway
- launch sites are not in close proximity to highways?
  - ex CCAFS is 26.88km from nearest highway
- launch sites are not in close proximity to cities?
  - CCAFS isex. 23 km from Titusville

Section 4

Build a Dashboard with Plotly Dash

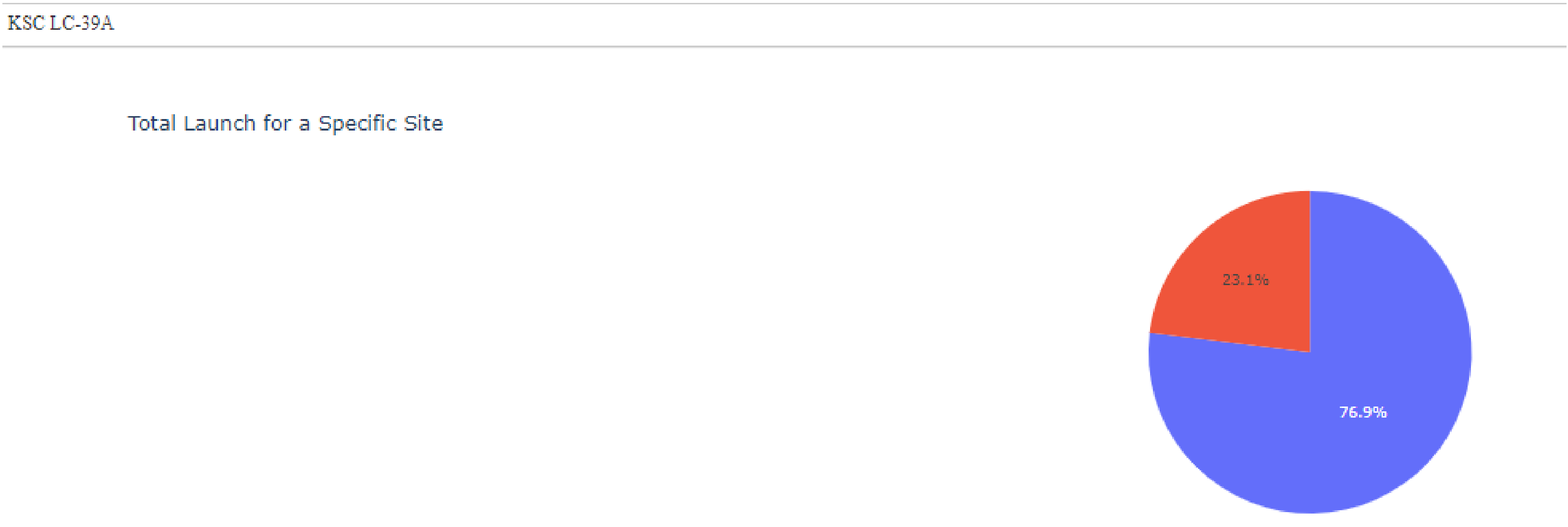
Launch Successes by All Sites

- At around 41%, KSC LC-39A had the most successful launches of the cohort



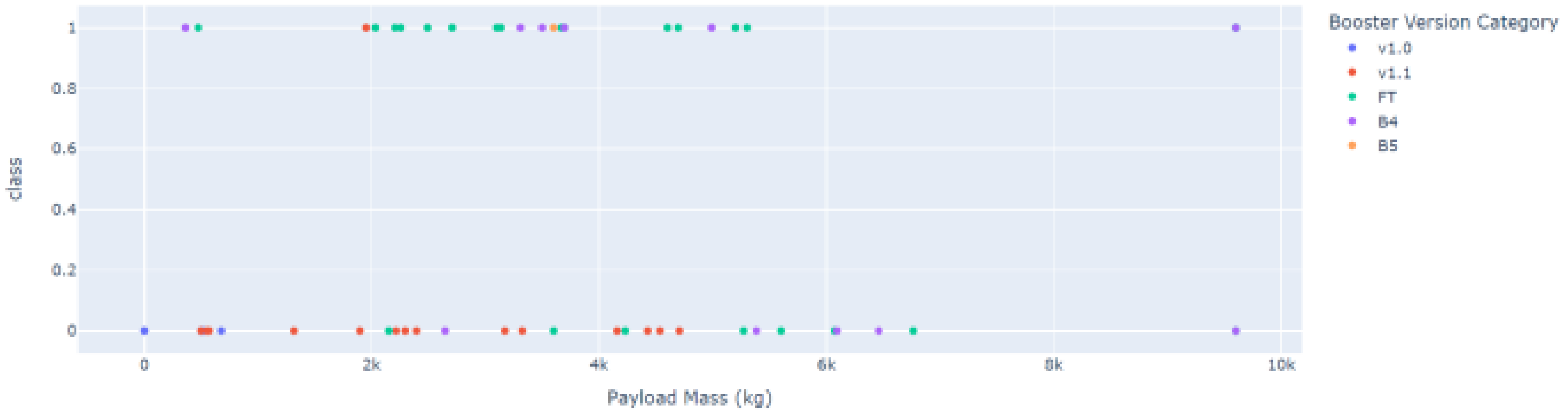
Launch Successes by All Sites

- At around 77%, KSC LC-39A had the highest success rate



Launch Successes by All Sites

- At around 77%, KSC LC-39A had the highest success rate

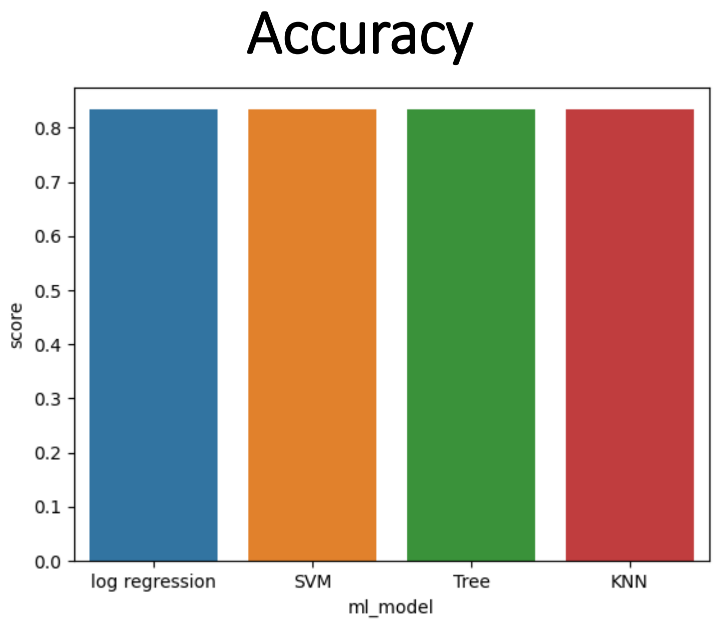


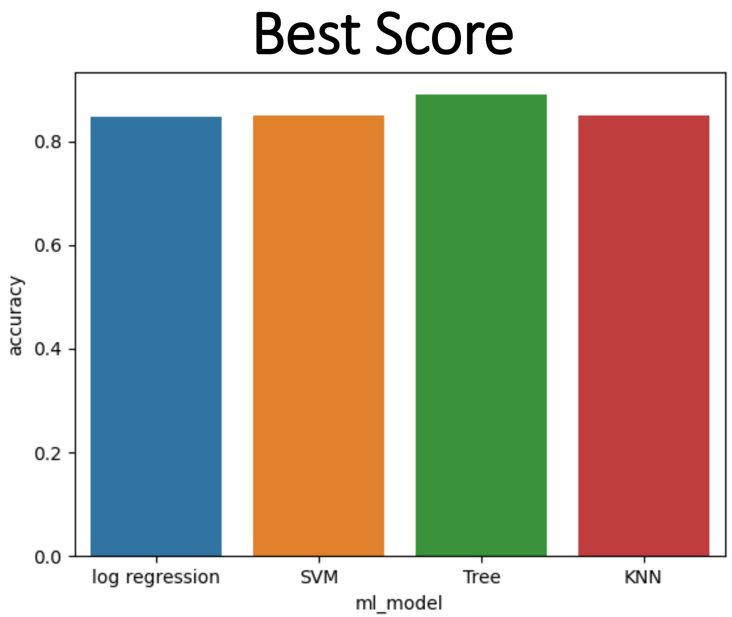
Launch Successes by All Sites

- Payload masses between 2k and 6k kg's had the highest success rates

Section 5

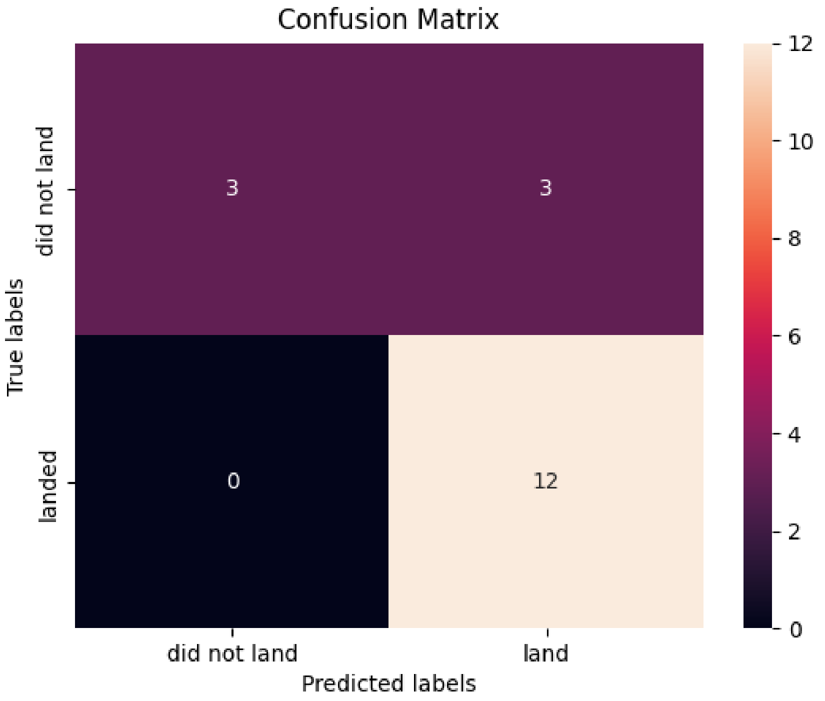
Predictive Analysis (Classification)





Classification Accuracy

- The 4 ML models all resulted in the same accuracy score
- Decision Tree had the highest “Best Score” (the greatest performance score achieved in modeling)



Confusion Matrix

- All models had the same resulting confusion matrix
- The 12 true positives suggest the model is performing reasonably well, but the 3 false positives (instances being incorrectly predicted as positive) is a concerning outcome

Conclusions

- The Orbits with the greatest success rates are the SSO, HEO, GEO, ES-L1
- Success Rates have increased marginally from 2013 to 2020
- Heavier payload mass correlates to higher success rates
- The KSC LC-39A has the highest success rate at 77%
- The Tree Classifier is the Machine Learning Model with the greatest success and the model that will be used to predict future launch results

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!