IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

A.R.J
04/08/2022

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection through API and Web Scraping
  - Exploratory Data Analysis (EDA) with SQL and Data Visualization
  - Interactive Map with Folium
  - Dashboards with Plotly Dash
  - Predictive Analysis with Machine Learning
- Summary of all results
  - Exploratory Data Analysis results
  - Interactive maps and dashboard
  - Predictive results

Introduction

- Project background and context
  - SpaceX is the only private company ever to return a spacecraft from low-earth orbit, which it first accomplished in December 2010. SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars whereas other providers cost upward of 165 million dollars each, much of the savings is because Space X can reuse the first stage. Therefore, if we can determine if the first stage will land, we can determine the cost of a launch.
  - SpaceY wants to bid against SpaceX for a rocket launch and will use this information to do so. Thus, The aim of this project is to predict if the Falcon 9 first stage will successfully land
- Problems you want to find answers
  - What determines a successful or failed landing
  - What is the relationship between the characteristics of the rocket to a successful or failed landing
  - What combination of factors allows the highest probability of a successful landing

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - SpaceX API
  - Webscraping from Wikipedia
- Perform data wrangling
  - One hot encoding
  - Created a landing outcome label from Outcome column
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Build, tune, evaluate classification models

Data Collection

Datasets were collected from:

- SpaceX API:
  - api.spacexdata.com/v4/

- Webscraping Wikipedia: https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922

Data Collection – SpaceX API

Request API and make response to a JSON file

Parse and transform the data (create dictionaries etc)

Create data frame and filter to include only relevant launches

- link

Data Collection - Scraping

Get HTML response
from Wiki web page

Create BeautifulSoup object
and create dictionary from
column names of the table

Create dataframe
from dictionary

- link

Data Wrangling

Transforming and exploring the data

One hot encoding to transform string variables into categorical variables where 1 means the mission has been successful and 0 means the mission was a failure

EDA to summarize launches per site, occurrence of each orbit and mission outcome per orbit to create landing outcome label

EDA with Data Visualization

Line Graph

- Success rate vs. Year

Bar Graph

- Success rate vs. Orbit

Scatter Graphs

- Flight Number vs. Payload Mass
- Flight Number vs. Launch Site
- Payload vs. Launch Site
- Orbit vs. Flight Number
- Payload vs. Orbit Type
- Orbit vs. Payload Mass

EDA with SQL

- names of the unique launch sites.
- 5 records where launch sites begin with the string 'CCA'
- the total payload mass carried by boosters launched by NASA (CRS).
- average payload mass carried by booster version F9 v1.1.
- the date when the first successful landing outcome in ground pad was achieved.
- the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000.
- the total number of successful and failure mission outcomes.
- the names of the booster versions which have carried the maximum payload mass.
- List the records which will display the month names, failure landing outcomes in drone ship, booster versions, launch site for the months in year 2015.
- Rank the count of successful landing outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

Build an Interactive Map with Folium

- Folium map object is a map centered on NASA Johnson Space Center at Houson, Texas
- Red circle at NASA Johnson Space Center and each launch site.
- Clustering points of the same coordinates
- Markers to show successful and failed landings.
- Markers to show distance between launch site to several key locations and plotting a line between them.
- *This allows us to easily visualize the launch sites and any relevant issues that may flag up regarding location, logistics etc.*

Build a Dashboard with Plotly Dash

- Dropdown, pie chart, rangeslider and scatter plots used to visualize data:
  - Percentage of launches by site
  - Range of payload

A quick tool to seamlessly compare variables

Predictive Analysis (Classification)

- Classification models: Logistic regression, Support vector machine, Decision trees and K nearest neighbor

- Data preparation, normalization and splitting into train and test sets

- Model training and evaluating.
  - Finding the best hyperparameters for each model
  - Finding accuracy of model on test set

- Model comparison

Results

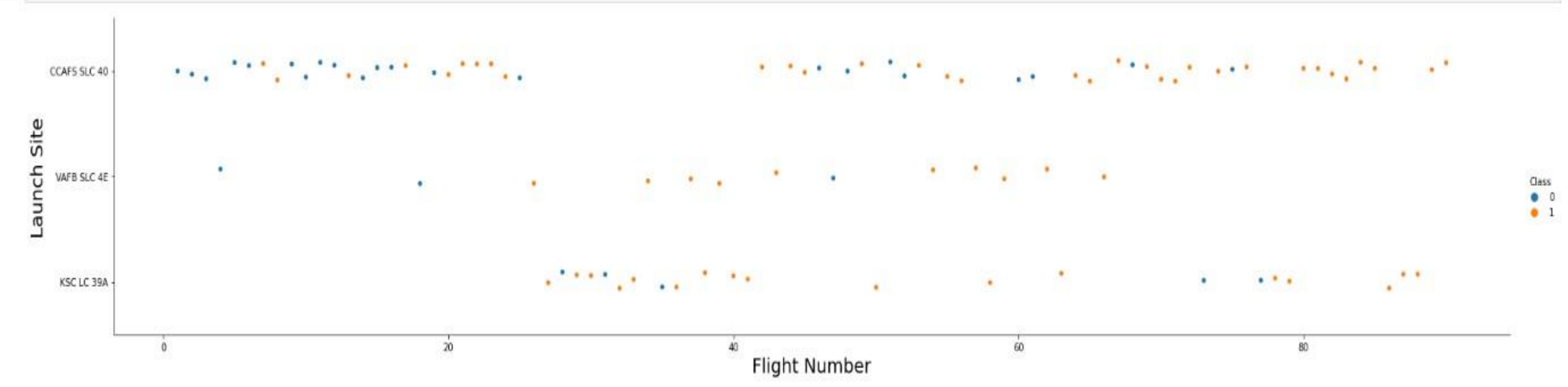
Incoming:

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

Section 2

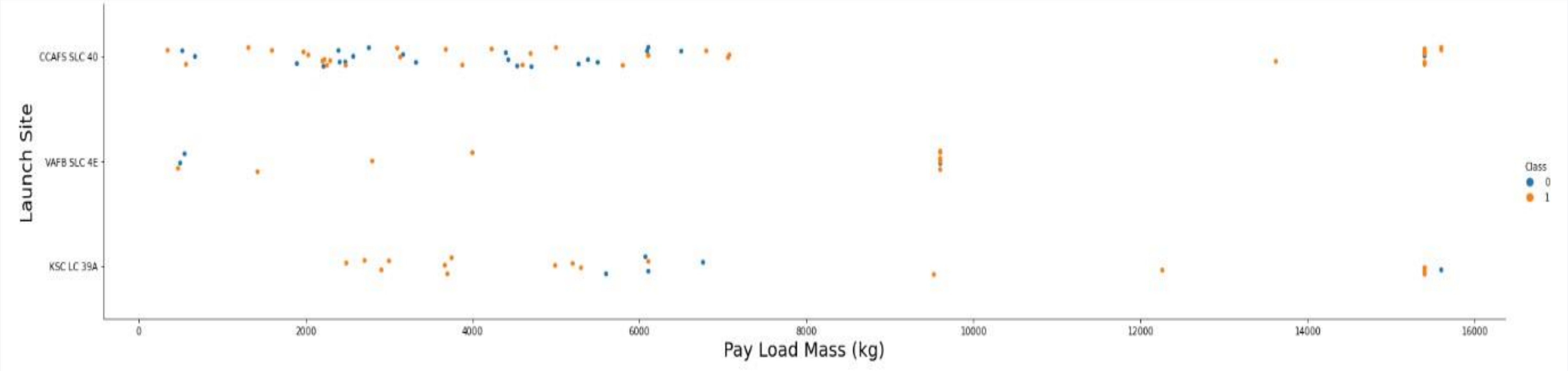
Insights drawn from EDA

Flight Number vs. Launch Site



- *There is a general increase in success rate for each launch site*

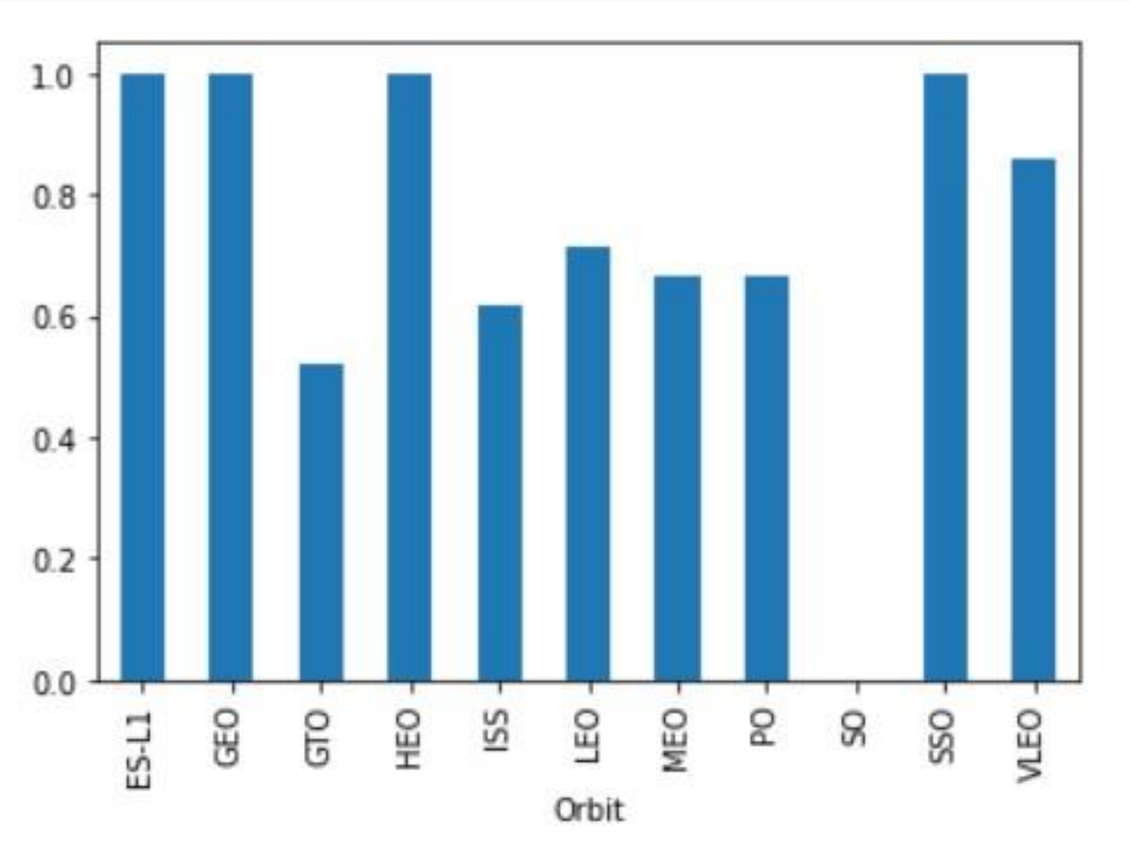
Payload vs. Launch Site



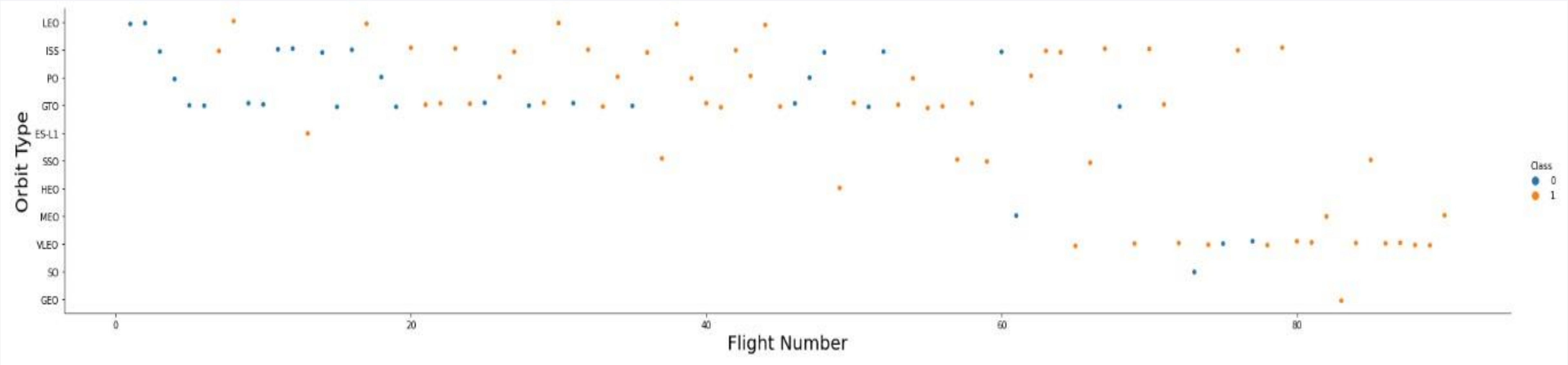
- Heavier Payload Mass has higher success rates generally
- CCAFS has mixed outcomes at lighter mass and VAFB site has no launches at masses over 9000Kg.

Success Rate vs. Orbit Type

- ES-L1, GEO, HEO and SSO have best of class success rates
- VLEO orbit type has above 80% success rate, placing it just below the best of class
- The remaining orbit types have less than 80% success rate

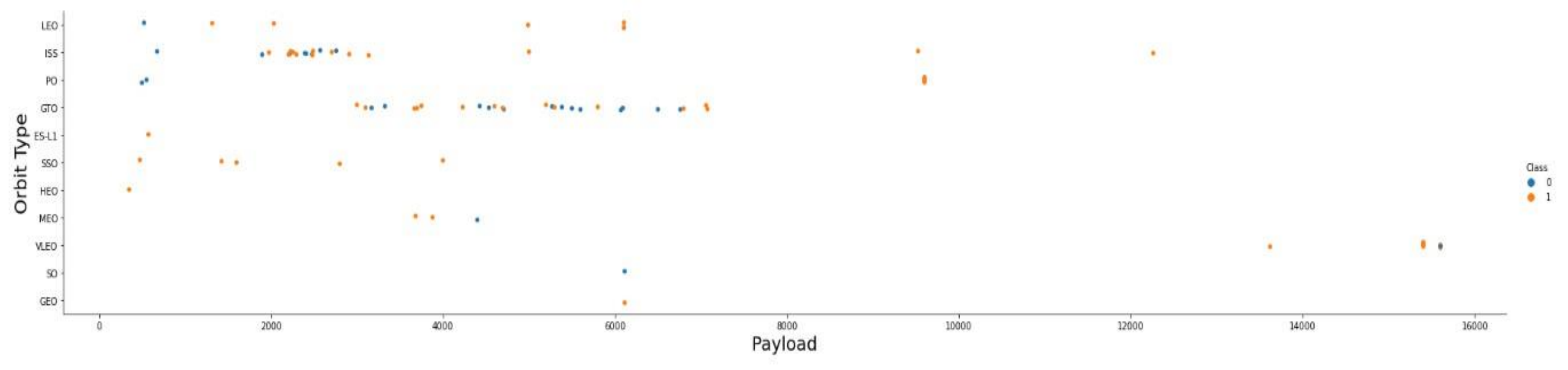


Flight Number vs. Orbit Type



- There is a general increase in success rate with the exception of GTO orbit type which shows mixed outcomes throughout
- SSO, HEO and VLEO are relatively newer orbit types with high success rates

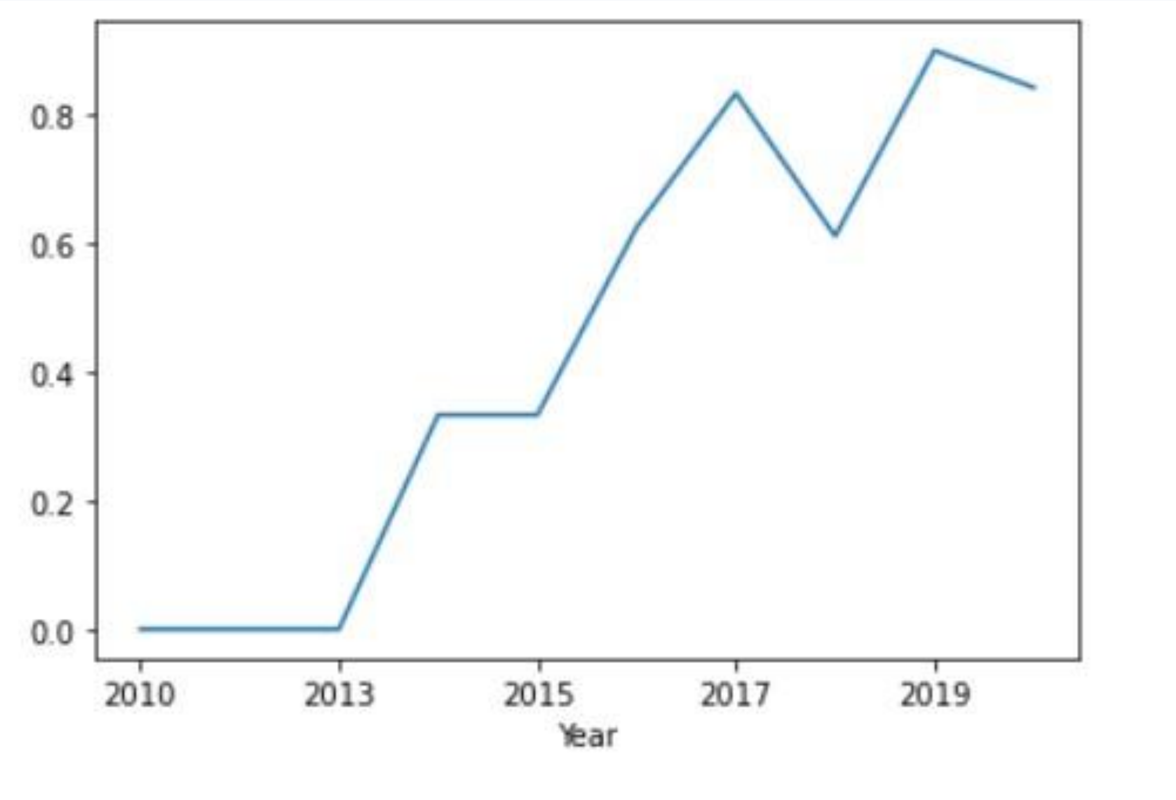
Payload vs. Orbit Type



- ISS orbit has the greatest range of Payload mass with excellent success rate at heavier mass however there aren't enough launches at those payload weights to be confident.
- SSO orbit type has had excellent success rate and lighter payload mass
- So payload mass seems to differ between orbit types

Launch Success Yearly Trend

- 2013 sees the initial rise in success rates with a continuous growth
- 2018 sees the largest drop in success rates however issues seem to have completely addressed and the success rate then continues to grow higher after 2019



All Launch Site Names

- Use of SELECT DISTINCT to remove duplicates and find unique entries

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

Launch Site Names Begin with 'CCA'

- Use of LIKE to find entries containing the ‘CCA’ string followed by LIMIT 5 to show 5 entries.

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	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)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- Sum of Payload Mass WHERE customers where NASA (CRS)

total payload mass carried by boosters launched by NASA (CRS)
45596

Average Payload Mass by F9 v1.1

- AVG payload mass WHERE booster version F9 v1.1

average payload mass carried by booster version F9 v1.1
2928.4

First Successful Ground Landing Date

- Date of the first successful ground pad landing using the MIN of DATE WHERE the Landing Outcome was successful

MIN(DATE)
01-05-2017

Successful Drone Ship Landing with Payload between 4000 and 6000

- Names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000
- Using the WHERE and AND clauses

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

Total Number of Successful and Failure Mission Outcomes

- Creating a SUBQUERY to COUNT mission outcomes for successful and failed outcomes

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

Boosters Carried Maximum Payload

- SELECT DISTINCT booster version WHERE subquery MAX(payload mass) is returned

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

- The Month, Booster version and launch site of failed outcomes in 2015
- Due to sqlLITE version: Substr(DATE, 4, 2) for month. Substr(DATE,7, 4) for year

MONTH	Booster_Version	Launch_Site
01	F9 v1.1 B1012	CCAFS LC-40
04	F9 v1.1 B1015	CCAFS LC-40

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

- Ranking the count of Successful landing between the date 2010-06-04 and 2017-03-20, in descending order
- ORDER BY DESC

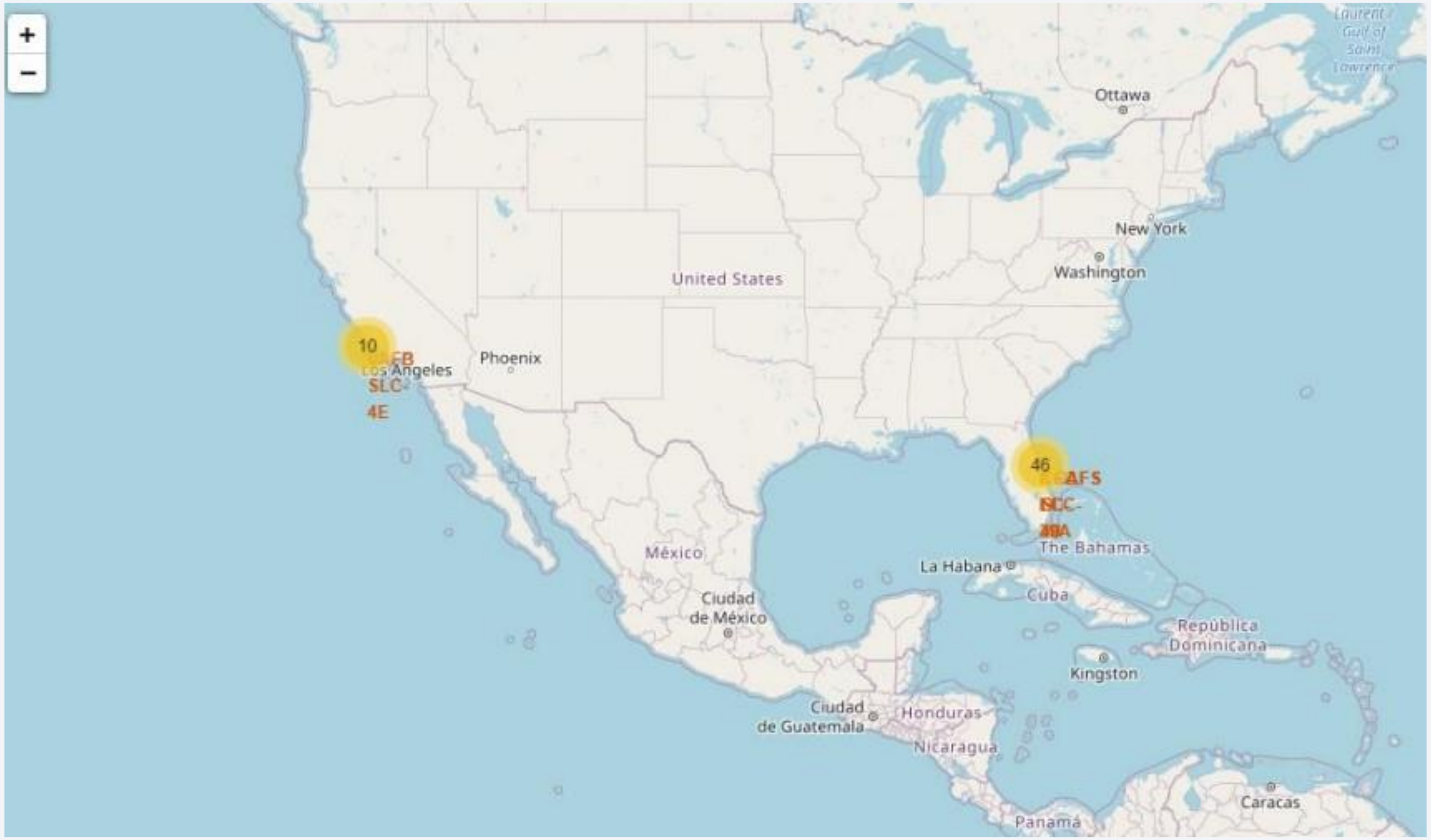
Landing _Outcome	COUNT("LANDING _OUTCOME")
Success	20
Success (drone ship)	8
Success (ground pad)	6

Section 3

Launch Sites Proximities Analysis

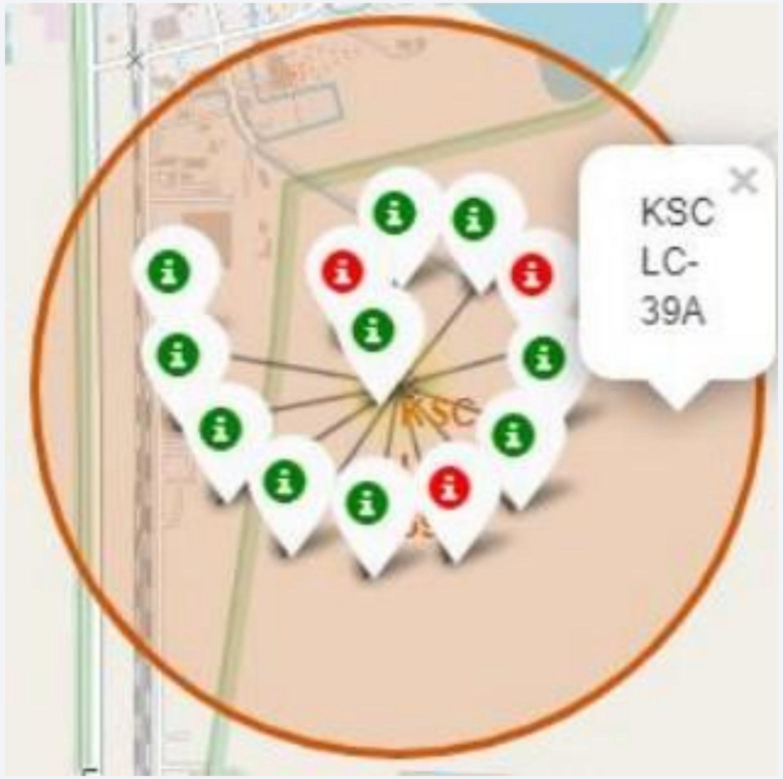
Launch Sites

- Launch sites located by the coast of US

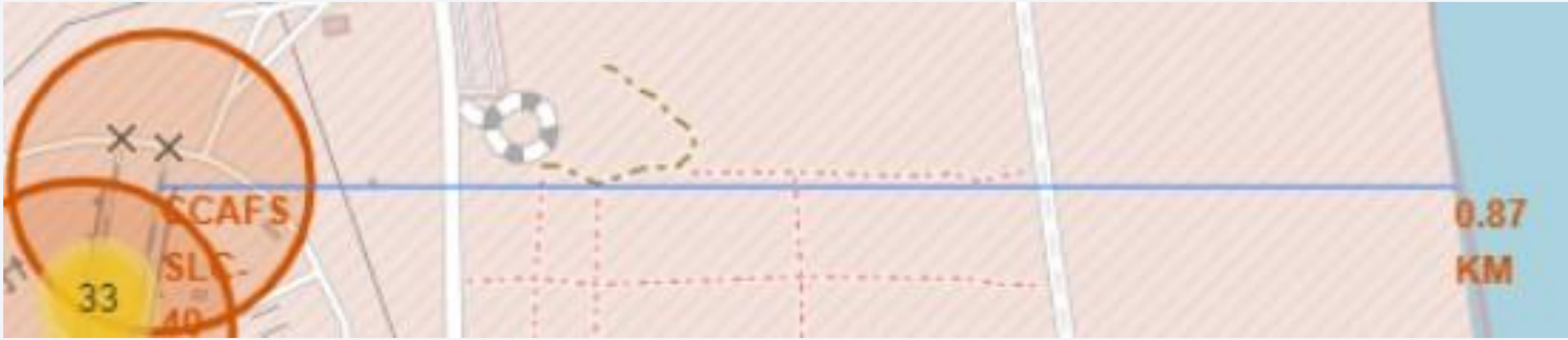


Outcome color labeled

- GREEN marks successful launch outcomes
- RED marks failed launch outcomes



Distance to key locations



- Distance of launch site to the coastline

Section 4

Build a Dashboard with Plotly Dash

Success by launch site

Pie chart showing proportion of successful outcomes by launch site

KSC LC-39A is best of class

Success rate for KSC LC-39A

Pie chart displaying the ratio of successful to failed outcomes for the KSC LC-39A launch site

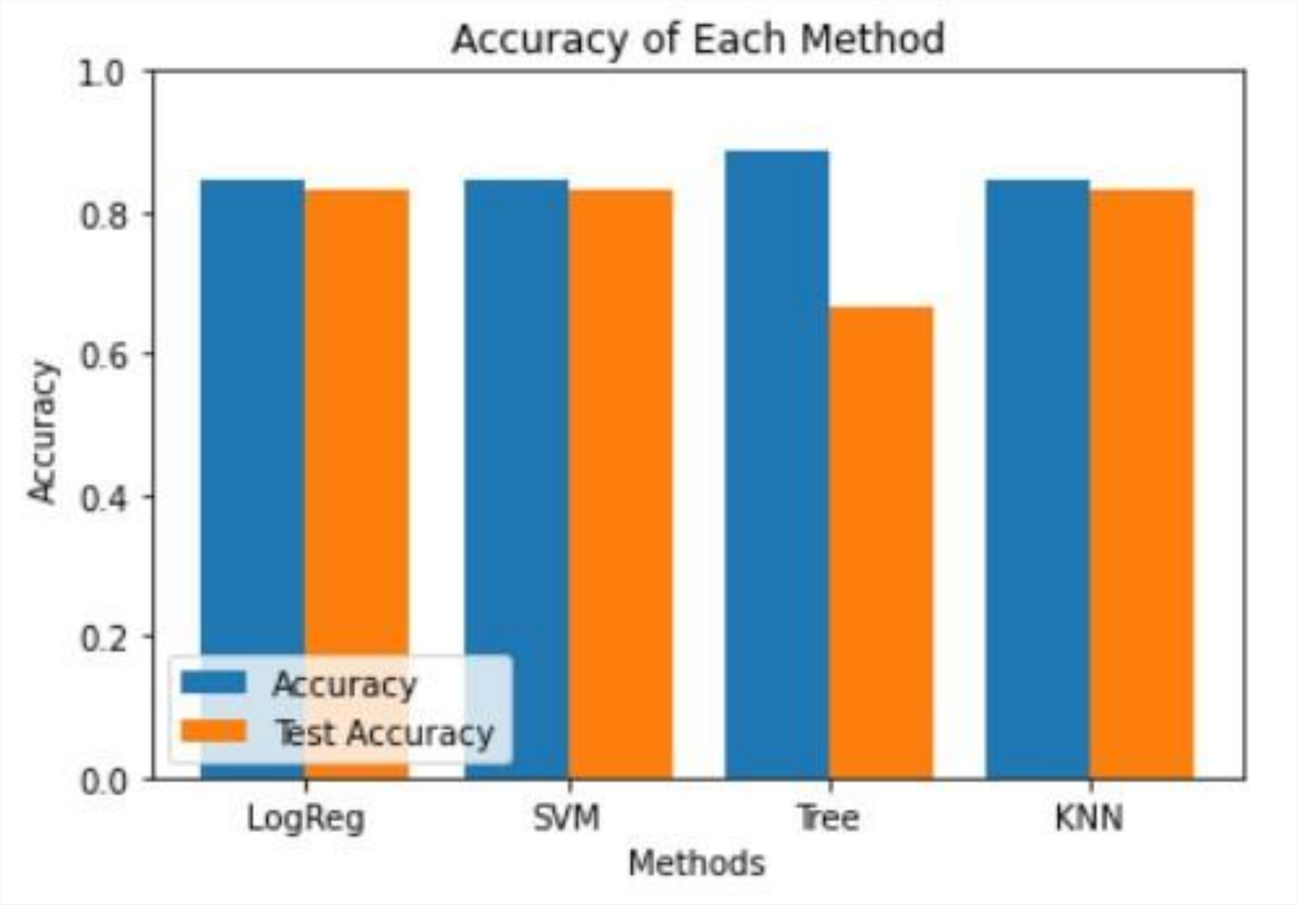
76.9% success rate

Section 5

Predictive Analysis (Classification)

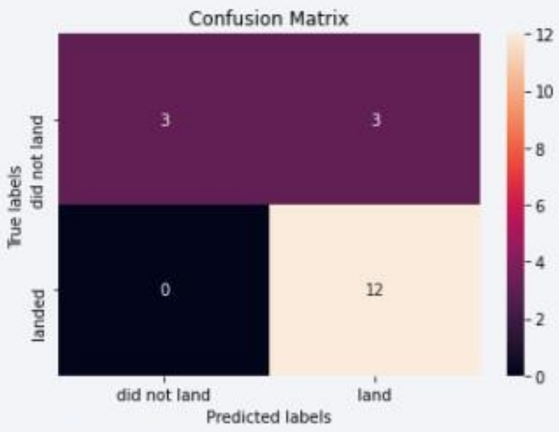
Classification Accuracy

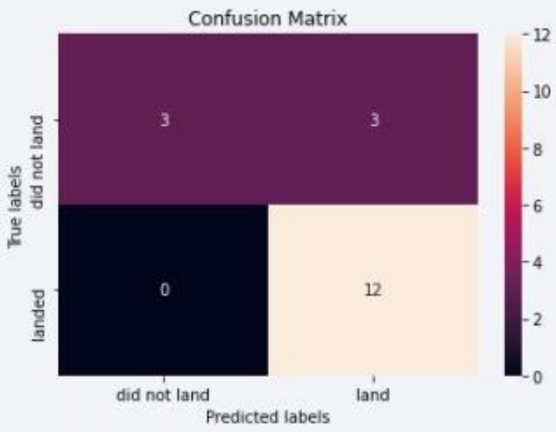
- Four classification models trained and tested
- Accuracy on train set and test sets for each model compared to other models
- Shows a similar accuracy except for Decision Tree model which performs worst on test accuracy

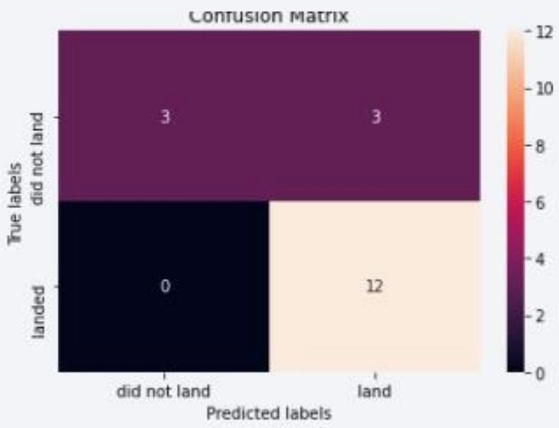


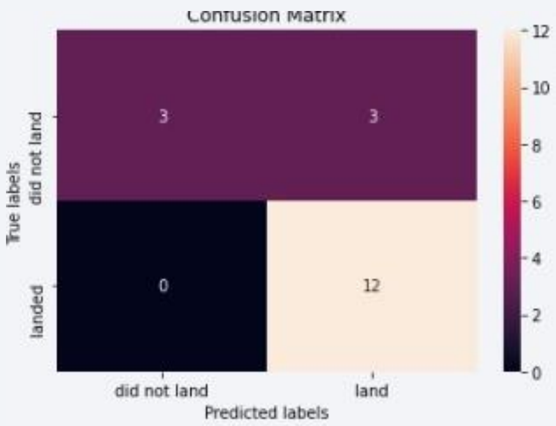
Confusion Matrix

- Identical Confusion Matrix plotted for each model
- Every model has an issue of false positive as seen on the top right quadrant
- First Row, Second column









Conclusions

- The success rate of launches are affected by variables such as launch site, payload mass, orbit types etc
- The positive liner relationship between the number of missions, time and the success rate indicates a continued learning from launches and combining variables common to successful landings
- KSC LC-39A is the best launch site
- ES-L1, GEO, HEO and SSO are the best orbit types
- Payload mass is orbit type dependent as both lighter and heavier launches succeeded at KSC LC-39A
- Model can be selected to evaluate these variable and select the best combination with the lowest risk and highest chance of success

Thank you!