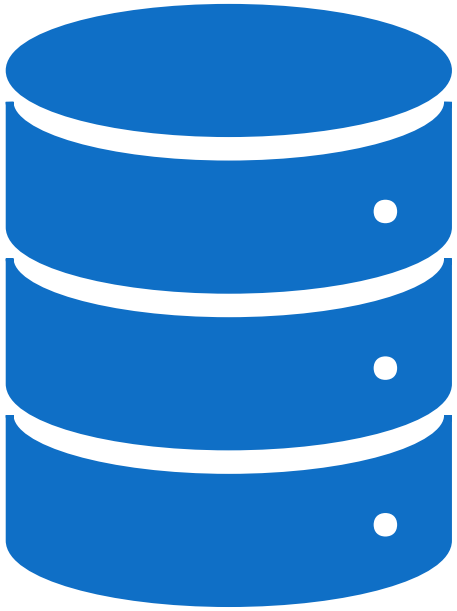
IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

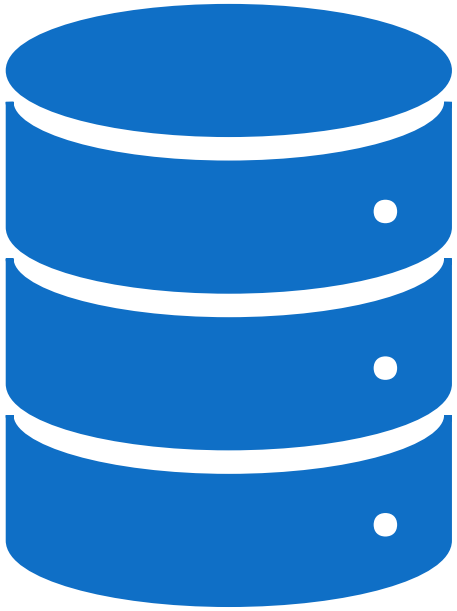
Carrie E. A. Grant, PhD.
May 4, 2025

GitHub Project Link



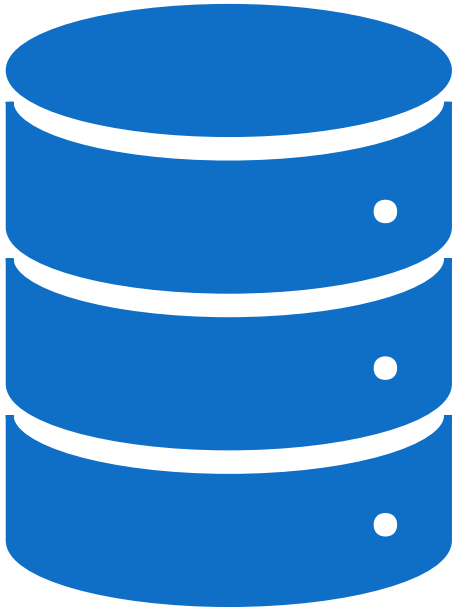
Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion



Executive Summary:

- In this capstone project, the goal is to predict if the SpaceX Falcon 9 first stage will land successfully using several machine learning classification algorithms. The main steps in this project include:
  - Data collection, wrangling, and formatting
  - Exploratory data analysis
  - Interactive data visualization
  - Machine learning prediction
- Visualizations and EDA methods found that certain features of the rocket launches have a relationship with the success or failure of a landing.
- The decision tree was found to be the best machine learning algorithm to predict if the Falcon 9 first stage will land successfully.



Introduction

Project background and context:

- Space travel is here! Private companies are competing to be in the forefront of space travel and profit from individuals who want to travel to space. Currently, SpaceX is leading the way because they have developed a way to reuse the first stage.

Questions to answers:

- In this project, data science techniques have been used to predict when a SpaceX Falcon 9 will land successfully. Thus, enabling the reuse of the first stage and ultimately saving money!

Section 1

Methodology

Methodology

- Data collection methodology
- Perform data wrangling
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

Data Collection - Overview

Step 1: Used the SpaceX API to request data.

Step 2: Decoded the response content using JSON.

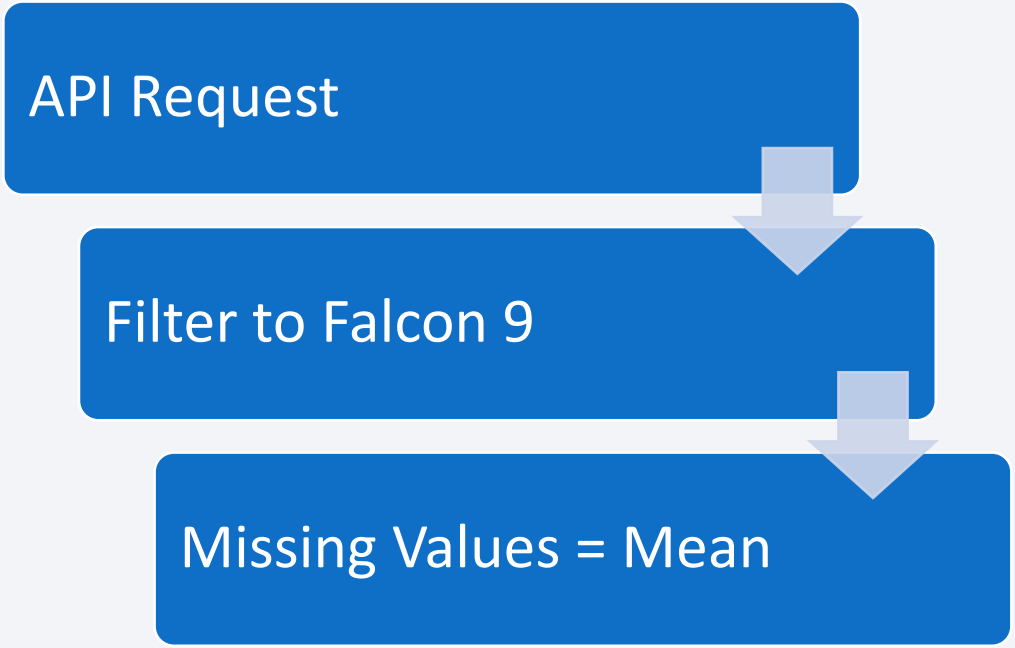
Step 3: Cleaned the data.

Step 4: Performed web scraping for Falcon 9 launch records using BeautifulSoup.

Step 5: A final dataframe for analysis!

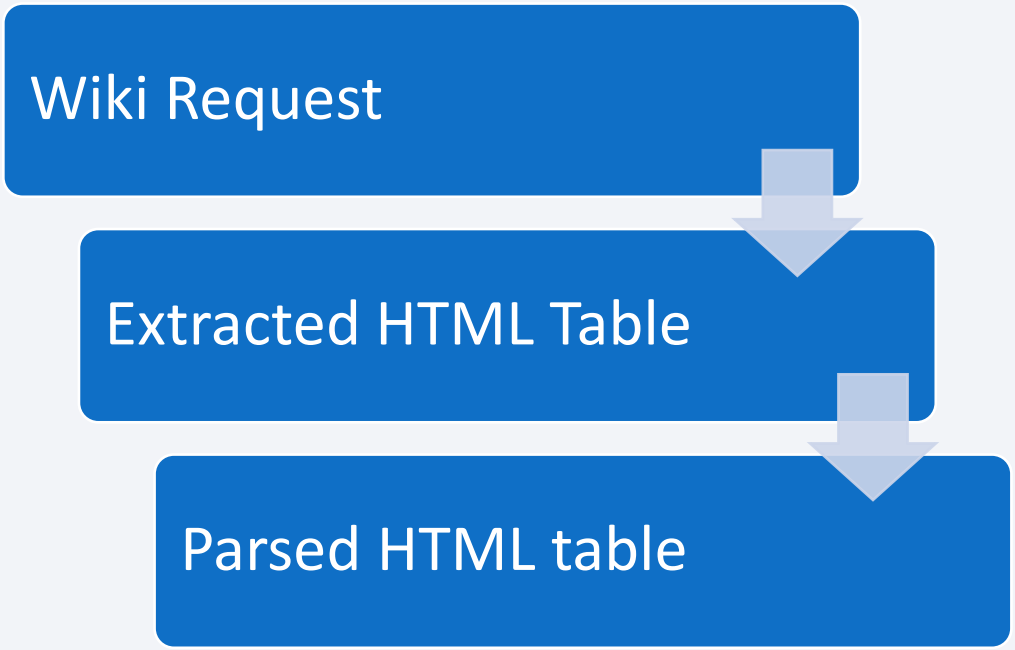
Data Collection – SpaceX API

- SpaceX API:
  - Used spacex_url="https://api.spacexdata.com/v4/launches/past"
  - Filtered downloaded data to include only Falcon 9 launches.
  - Replaced missing values with mean
  - Resulted in a dataframe with 90 rows and 17 columns.
- GitHub Link:
  - Lab 1_ Data Collection API



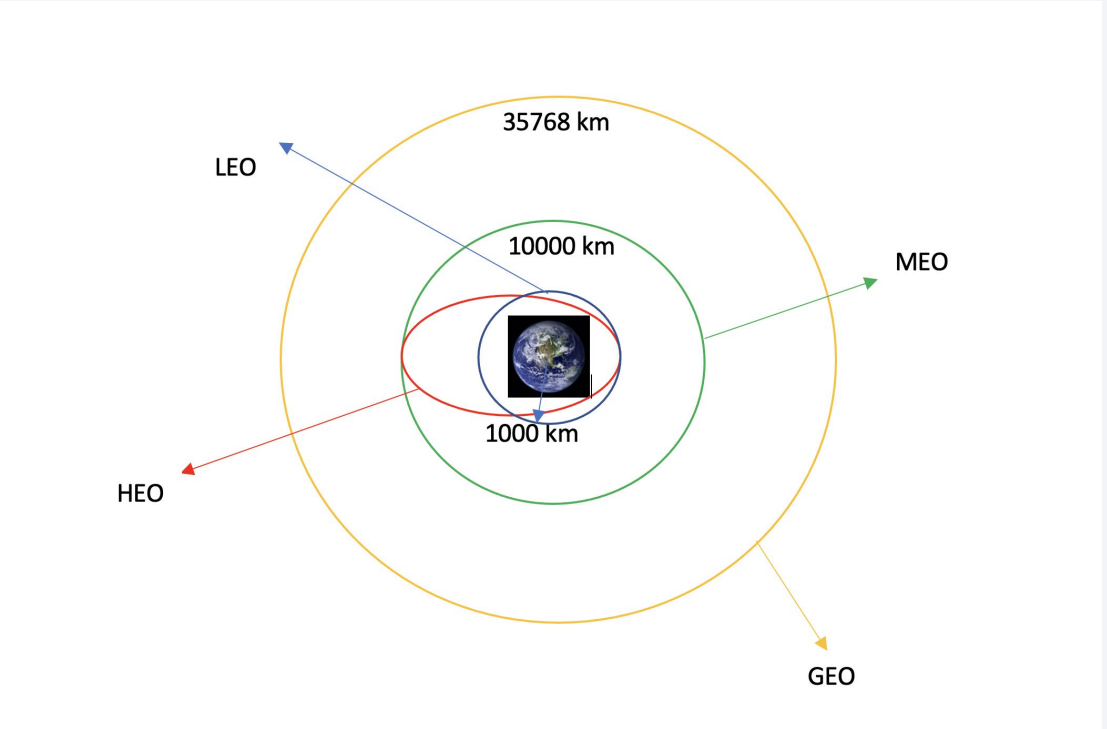
Data Collection – Web Scraping

- Web Scraping:
  - Requested the Falcon9 Launch Wiki page from its URL
  - Extracted all column/variable names from the HTML table header
  - Created a data frame by parsing the launch HTML tables
  - Resulted in a dataframe with 90 rows and 17 columns.
- GitHub Link:
  - Lab 2_Web Scraping



Data Wrangling

- Data Wrangling:
  - Calculated the number of launches on each site.
  - Calculated the number and occurrence of each orbit.
  - Calculated the number and occurrences of mission outcome of the orbits.
  - Created a landing outcome label from Outcome column.
- GitHub Link:
  - Lab 3 Data Wrangling



EDA with Data Visualization

- EDA with Data Visualization:
  - Functions from the Pandas and NumPy libraries are used to derive basic information about the data collected.
    - Number of launches on each launch site
    - Number of occurrence of each orbit
    - Number and occurrence of each mission outcome
    - Success rate of each orbit type

- GitHub Link:
  - Lab 5_EDA with Data Visualization

EDA with Data Visualization

- EDA with Data Visualization
  - Functions from the Matplotlib and Seaborn libraries are used to visualize the data through scatterplots, bar charts, and line charts.
  - Visualizations were created to understand the relationships between the following pairs of data.
    - Success rate and orbit type
    - Flight Number and Launch Site
    - Payload Mass and Launch Site

- GitHub Link:
  - Lab 5_EDA with Data Visualization

EDA with SQL

- EDA with SQL:
  - Analyzed the data by creating lists of attributes that contributed to successful landings.
  - Recorded dates and names of successes
  - Found total counts of successes.
  - Found total and average payload mass.
  - Ranked the data.

- GitHub Link:
  - Lab 4_ EDA with SQL

Build an Interactive Map with Folium

- Build an Interactive Map with Folium
  - Functions from the Folium libraries are used to visualize the data through interactive maps.
  - The Folium library is used to mark the following:
    - All launch sites on a map.
    - The launches that were successful and that failed at each launch site.
    - The distances between each launch site to the nearest coastline, railway, highway and city.

- GitHub Link:
  - Lab 6_Interactive Map with Folium

Build a Dashboard with Plotly Dash

- Build a Dashboard with Plotly Dash
  - Functions from Dash are used to generate an interactive site.
  - Users can change in site location using a dropdown menu.
  - Users can also change the range of the payload using a slider.
  - A pie chart and a scatterplot were created to show:
    - The total success launches from each launch site
    - The relationship between payload mass and mission outcome (success or failure) for each launch site.

- GitHub Link:
  - Lab 7 Dash

Predictive Analysis (Classification)

- Predictive Analysis (Classification)
  - Functions from the Scikit-learn library are used to create machine learning models.
  - The machine learning prediction phase include the following steps:
    - Standardized the data
    - Split the data into training and test data
    - Created machine learning models
    - Fit the models on the training set
    - Found the best combination of hyperparameters for each model
    - Evaluated the models based on their accuracy scores and confusion matrix

- GitHub Link:
  - Lab 8_ Predictive Analysis

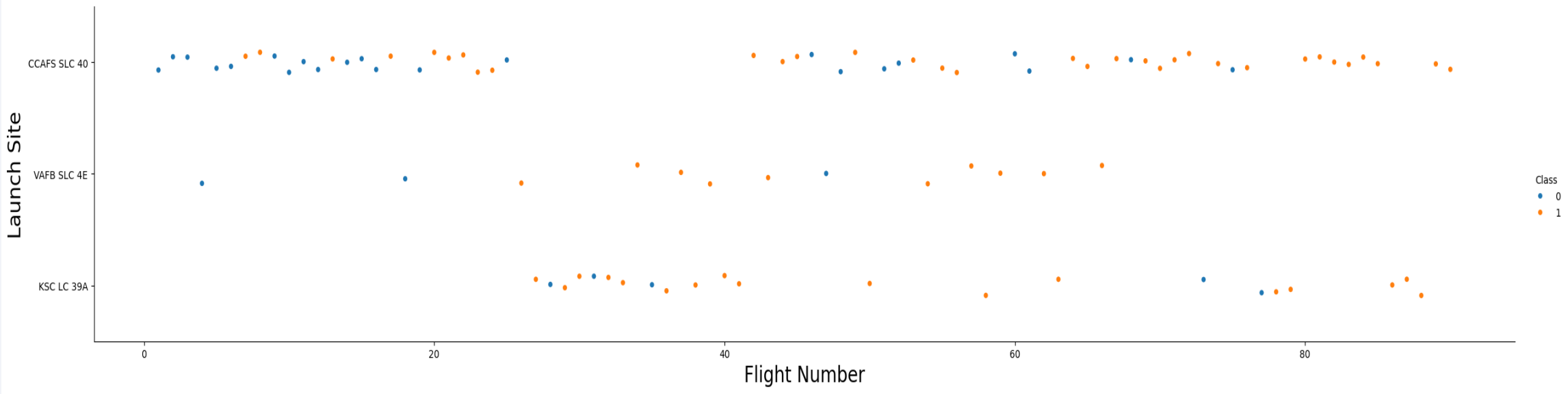
Results

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

Scatter plot of Flight Number vs. Launch Site

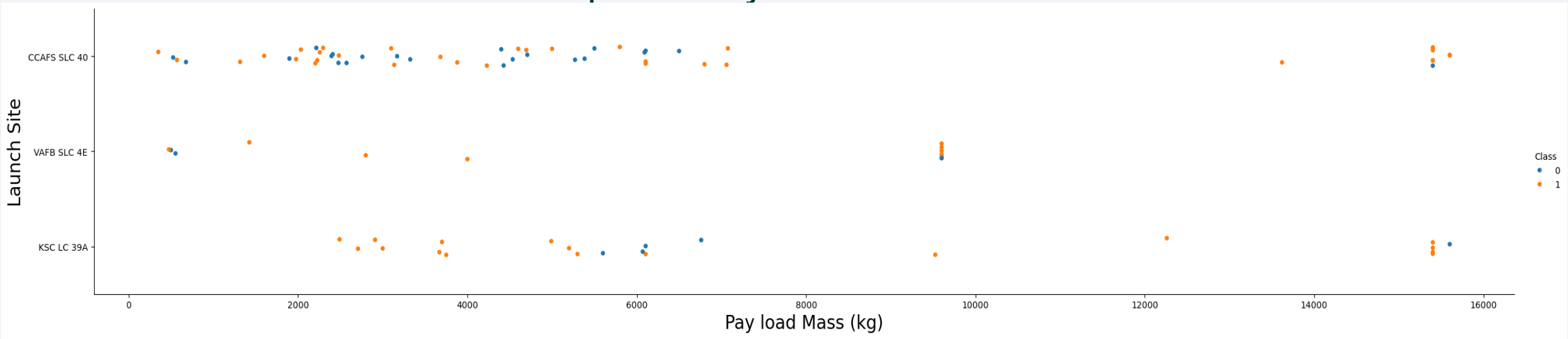


Observations:

- The most flights launched from CCAFS SLC 40.
- The least number of flights launched from VAFB SLC 4E.
- The greater the flight number the higher chance of success.

Payload vs. Launch Site

Scatter plot of Payload vs. Launch Site

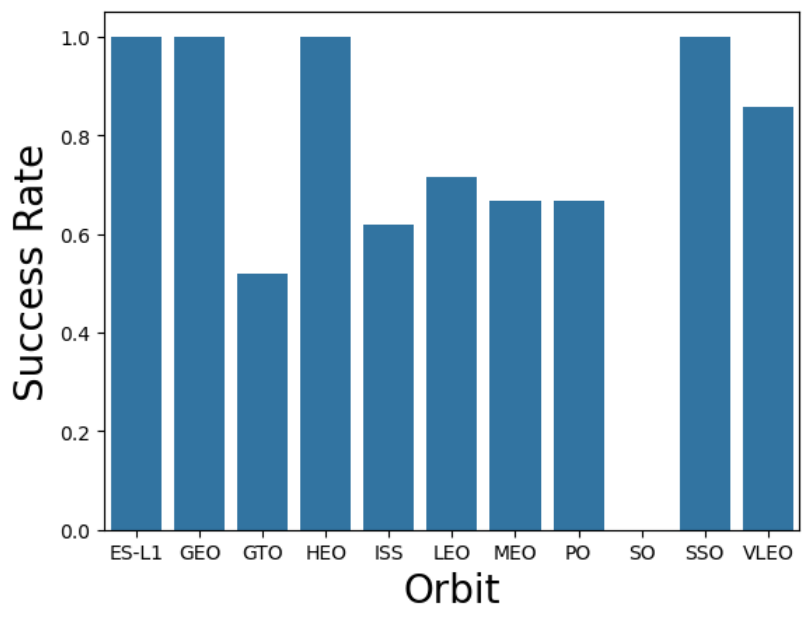


Observation:

- The VAFB-SLC launch site there are no rockets launched with heavy payload mass(greater than 10000).

Success Rate vs. Orbit Type

Bar Chart for the Success Rate of each Orbit Type

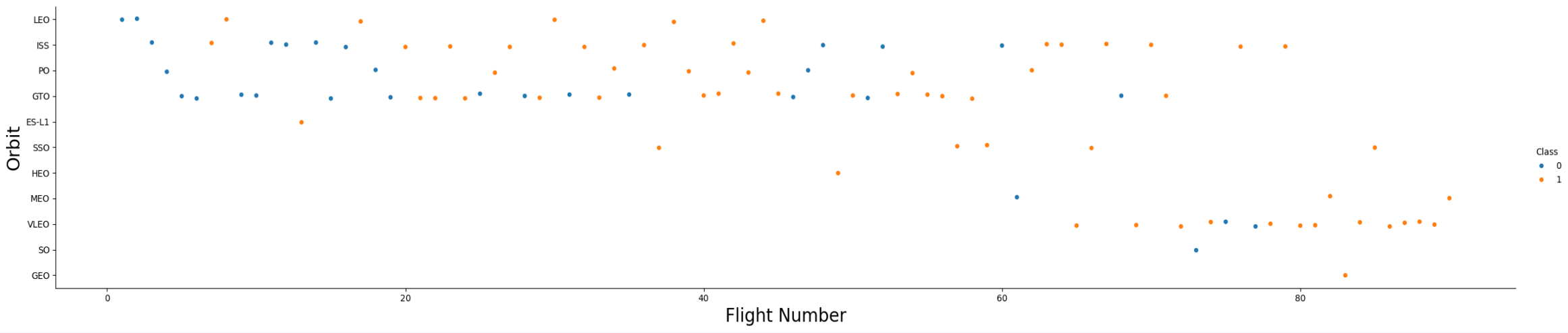


The orbits with 100% success rate are:

- ES-L1
- GEO
- HEO
- SSO

Flight Number vs. Orbit Type

Scatter plot of Flight number vs. Orbit type

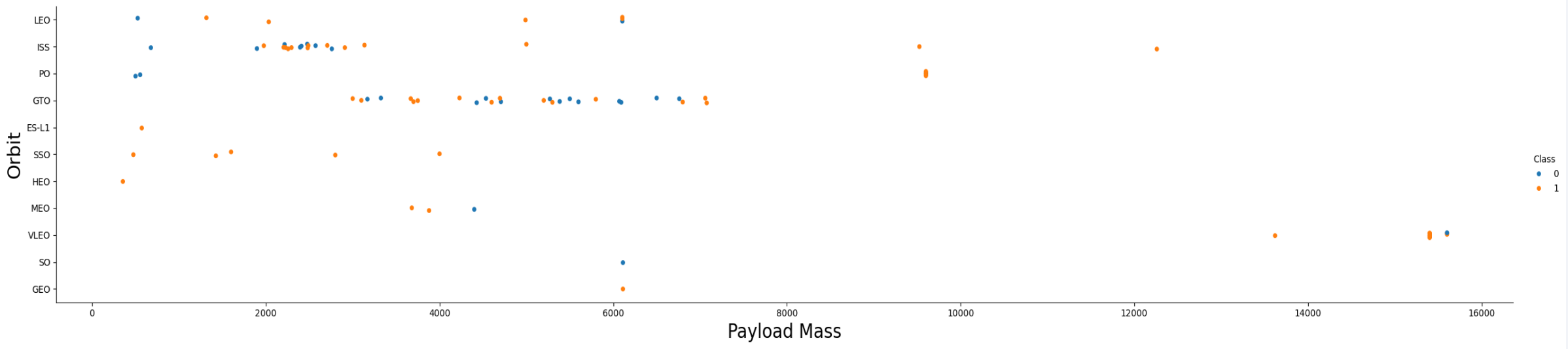


Observations:

- In the LEO orbit, success seems to be related to the number of flights.
- Conversely, in the GTO orbit, there appears to be no relationship between flight number and success.

Payload vs. Orbit Type

Scatter plot of Payload vs. Orbit Type

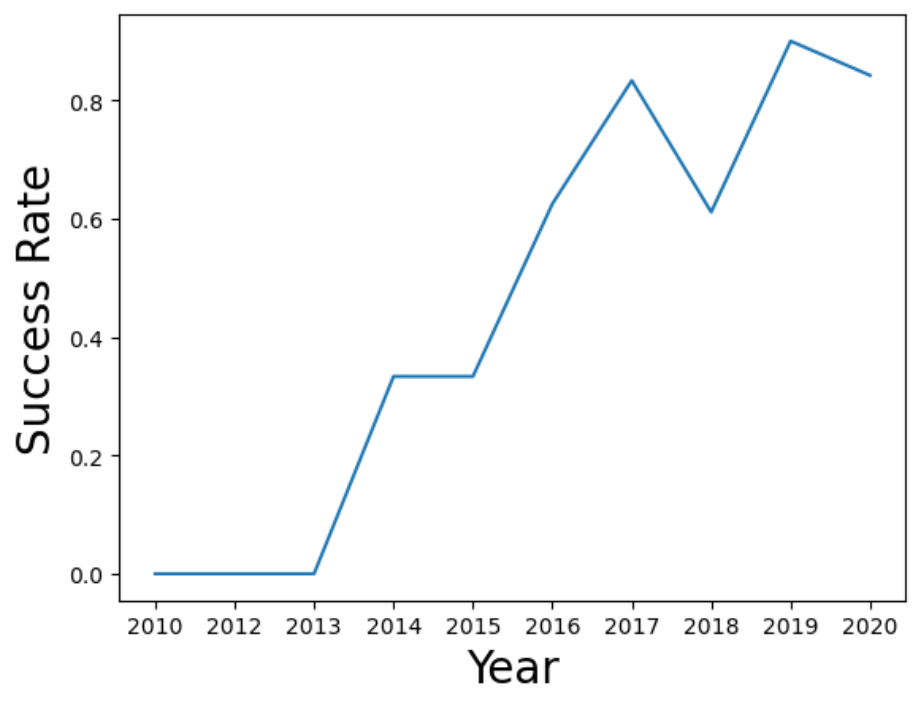


Observations:

- With heavy payloads, PO, VLEO, and ISS have a larger number of successful landings.
- For GTO, both successful and unsuccessful landings are present at all payloads.

Launch Success Yearly Trend

Line Chart of Yearly Average Success Rate



Observations:

- The success rate steadily increased from 2013 to 2014 ending with a 30% chance of success.
- The rate leveled off in 2014.
- In 2015 until 2017, the success rate continued to increase to over 80%.
- In 2017, there was a 20% decrease in success rate.
- From 2018 to 2019, the success grew over 20% ending with over 80% success!
- In 2019, there was a slight decrease of 5% or so.
- Overall, the success rate increased from 2013 to 2020.

All Launch Site Names

List of the Unique Launch Sites

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

There are four unique launch sites, three in Florida and one in California.

Florida:
- CCAFS LC-40
- CCAFS SLC-40
- KSC LC-39A

California:
- VAFB SLC-4E

Launch Site Names Begin with 'CCA'

List of 5 records where Launch Sites begin with `CCA`

Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Observations:

- All records listed are from one launch site, CCAFS LC-40 in Florida.
- One record was SpaceX and four were NASA.
- All flights completed there missions successfully.
- There were two failed landings, but the other flights did not attempt to land so they were not necessarily successes.
- The two flights with failed landings had 0 KG of payload mass, while the other three had at least 500 KG.

Total Payload Mass

Total Payload Carried by Boosters from NASA

SUM(PAYLOAD_MASS__KG_)
45596

Overall, the total payload carried by boosters from NASA for the years in our study was 45,596 KG.

Average Payload Mass by F9 v1.1

Average Payload Mass Carried by Booster Version F9 v1.1

AVG(PAYLOAD_MASS__KG_)
2928.4

The average payload mass carried by Booster Version F9 v1.1 for the years in our study was 2,928.4 KG.

First Successful Ground Landing Date

Date of the First Successful Landing Outcome on Ground Pad

MIN(Date)
2015-12-22

The first successful landing outcome on a ground pad was December 22, 2015.

Successful Drone Ship Landing with Payload between 4000 and 6000

Names of Boosters which have Successfully Landed on Drone Ship (4000 < payload mass < 6000)

Booster_Version
F9 FT B1021.1
F9 FT B1022
F9 FT B1023.1
F9 FT B1026
F9 FT B1029.1
F9 FT B1021.2
F9 FT B1029.2
F9 FT B1036.1
F9 FT B1038.1
F9 B4 B1041.1
F9 FT B1031.2
F9 B4 B1042.1
F9 B4 B1045.1
F9 B5 B1046.1

Fourteen boosters have successfully landed on a drone ship between 2010 and 2020!

Total Number of Successful and Failure Mission Outcomes

Total Number of Successful and Failure Mission Outcomes

Successful

COUNT (Mission_Outcome)
100

Failure

COUNT (Mission_Outcome)
1

The success rate of missions is extremely high! Only one of the flights was not successful in completing the mission.

Boosters Carried Maximum Payload

Boosters with Maximum Payload Mass

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

Twelve boosters have carried the maximum payload of 15,600 KG. This is equivalent to 34,393 pounds!

2015 Launch Records

List of Failed Landing Outcomes in Drone Ship in 2015

(Including Booster Versions & Launch Site)

SUBSTR(Date,6,2)	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

There were only two failed landings on drone ships in 2015. Both flights launched from CCAFS LC-40 in Cape Canaveral, Florida.

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

Rank of Landing Outcomes

Landing_Outcome	COUNT(Landing_Outcome)
Success	38
No attempt	21
Success (drone ship)	14
Success (ground pad)	9
Failure (drone ship)	5
Controlled (ocean)	5
Failure	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1
No attempt	1

Overall, there are more successes than failures in the landing outcome. There are also many times when a landing was not attempted.


Section 3
Launch Sites
Proximities Analysis

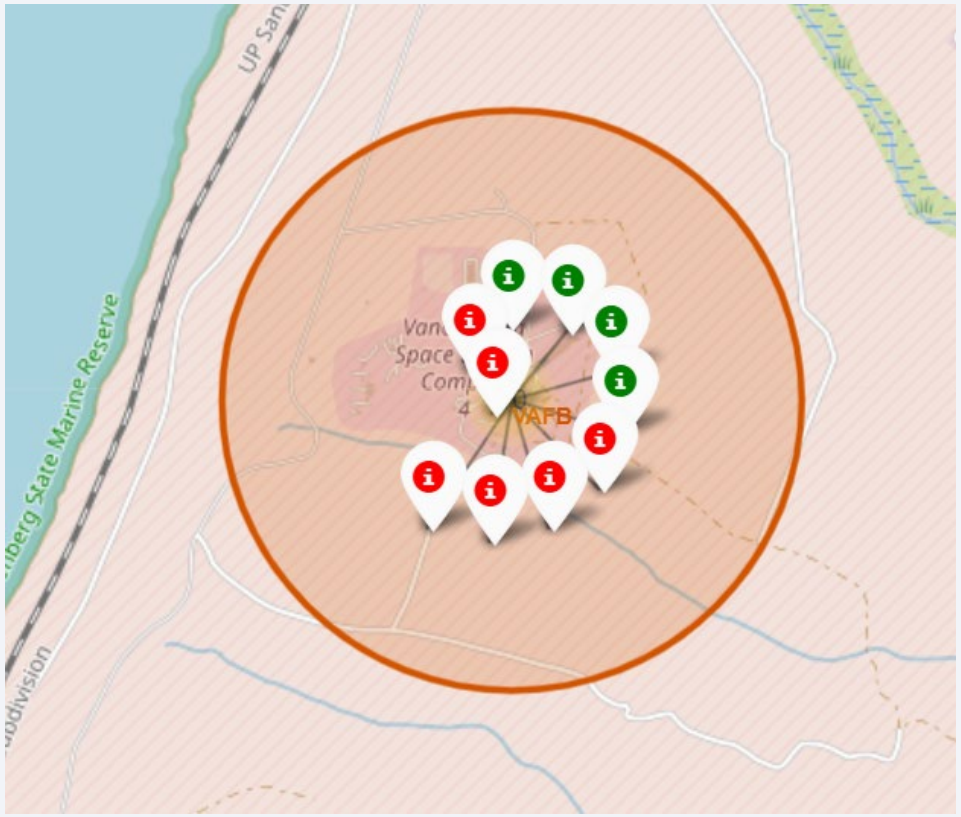
Global Map including All Launch Sites

The global map illustrates that all the launch sites are in the United States of America.

Three are in the state of Florida and one is California.

All sites are close to an ocean.

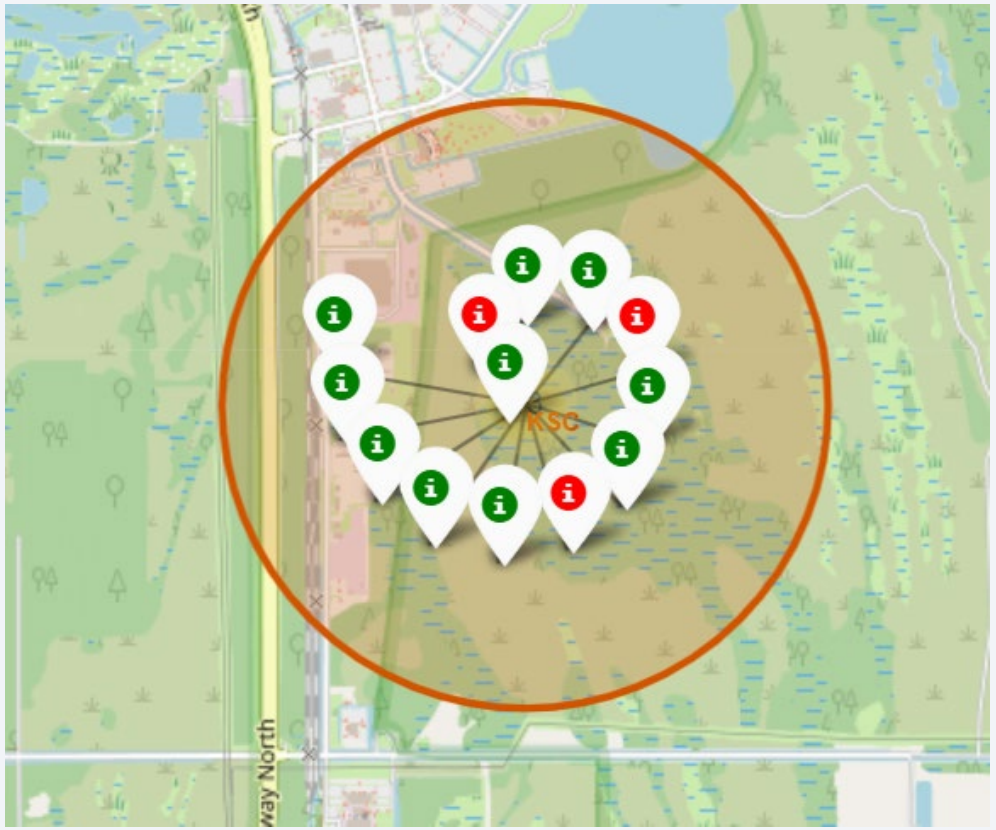
Vandenberg Space Force Base, California



Observations:

- There were ten launches from VAFB in California.
- Sixty percent of the launches failed to land successfully as indicated in red.
- Forty percent of the launches successfully landed as indicated in green.

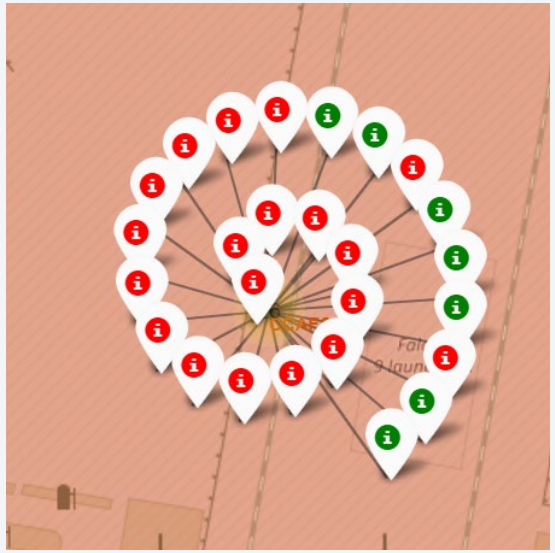
Kennedy Space Center, Florida

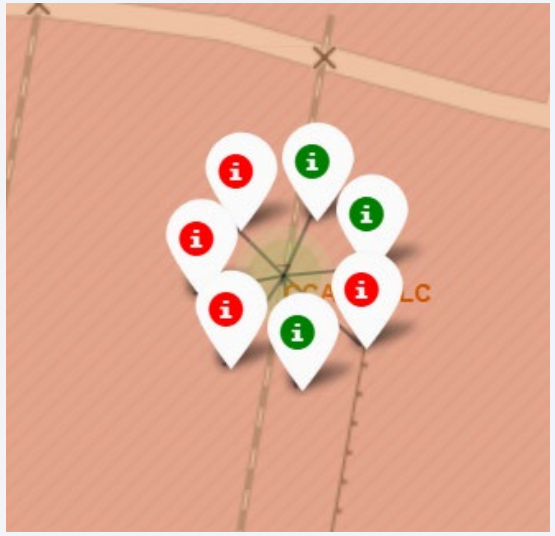


Observations:

- There were thirteen launches from KSC in Florida.
- Three of the 13 launches (23%) failed to land successfully as indicated in red.
- Ten of the 13 launches (77%) successfully landed as indicated in green.

Cape Canaveral Space Launch Complex 40, Florida

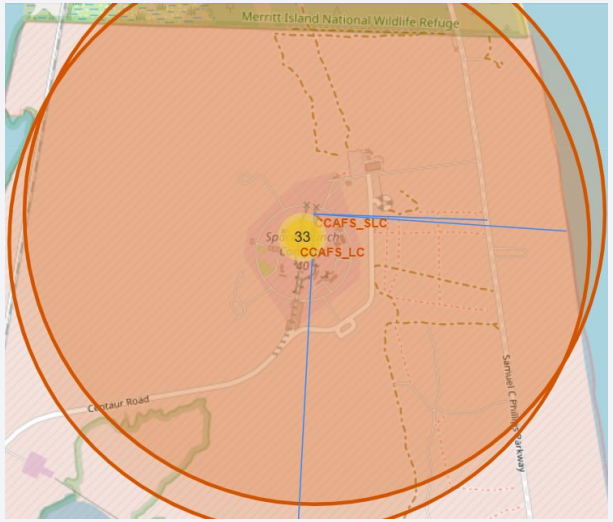




Observations:

- There are two launch sites at the Cape Canaveral Space Launch Complex.
- These sites are approximately 16 miles from the Kennedy Space Center.
- There were 33 launches from CCAFS in Florida.
- The launching site on the top had 26 launches while the site on the bottom had 7.
- For the site on the top, 7 of the 26 launches failed (27%) and 19 of the 26 launches (73%) were successful.
- For the site on the bottom, 3 of the 7 launches (43%) failed and 4 of the 7 launches (57%) were successful.

Distances from Launch Sites



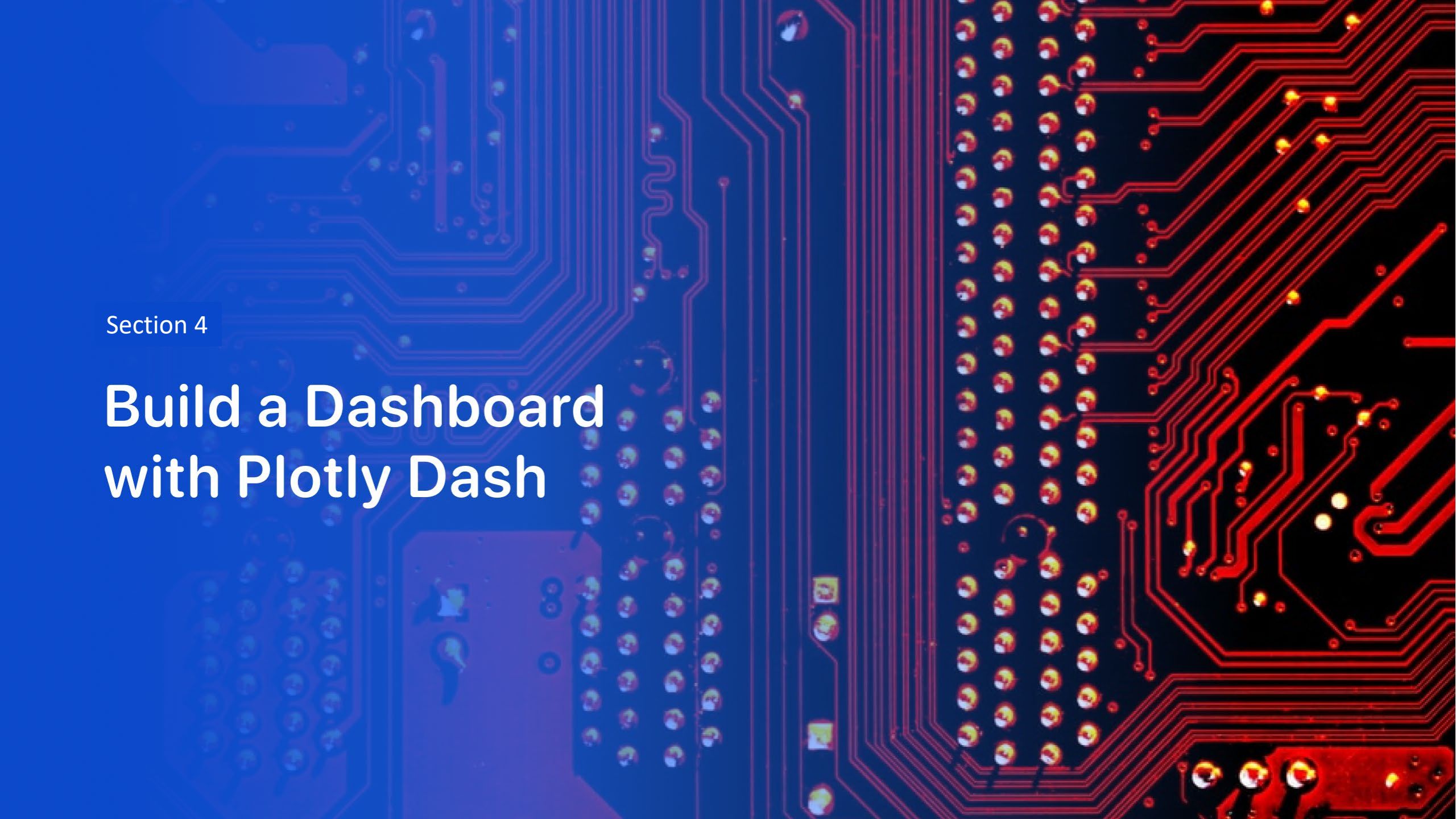
In this analysis of the distance, my focus was on the launch sites in Florida.

Observations:

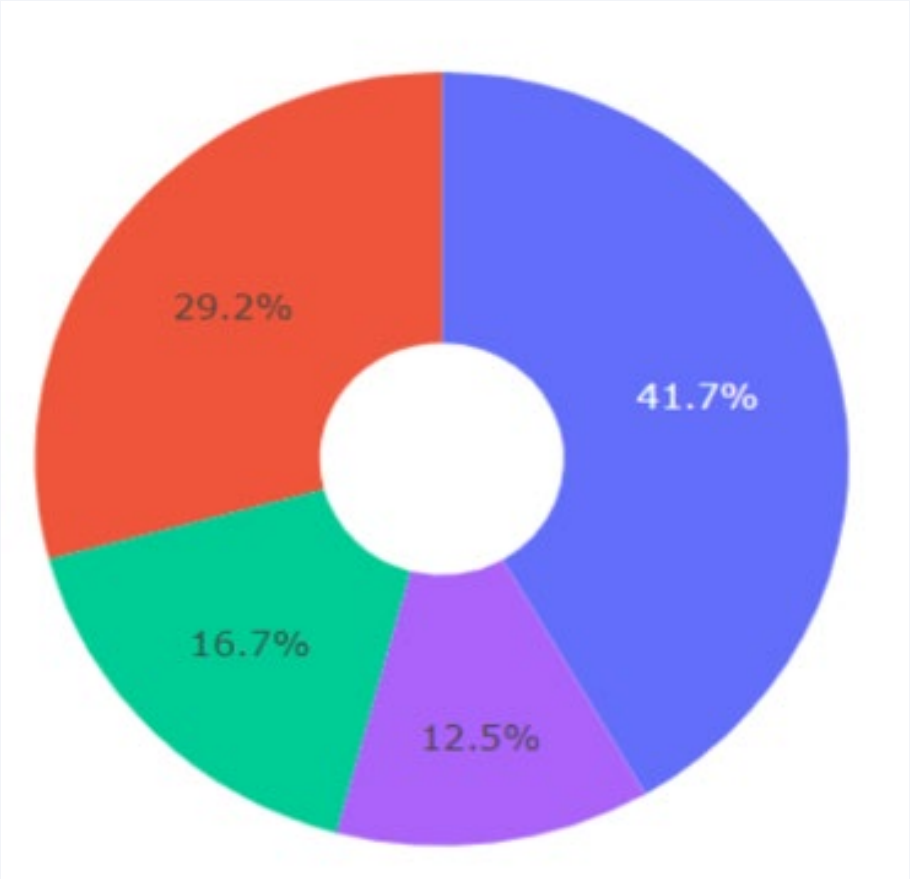
- The Kennedy Space was closest to a railroad. The distance was 22.8 km = 14.12 miles.
- The Cape Canaveral site was closest to a highway and a coastline.
- The closest highway (not an interstate) is only 0.596 km = 0.370 miles.
- The distance from the coastline to the launch site CCAFS_SLC is 0.8705 km. This is about 0.54 miles!
- The closest 'large' city is 54.007 km = 33.558 miles.

Conclusions:

- Launch sites are close to coastlines
- Launch sites are far from interstates, railroads, and large cities.


Section 4
Build a Dashboard with Plotly Dash

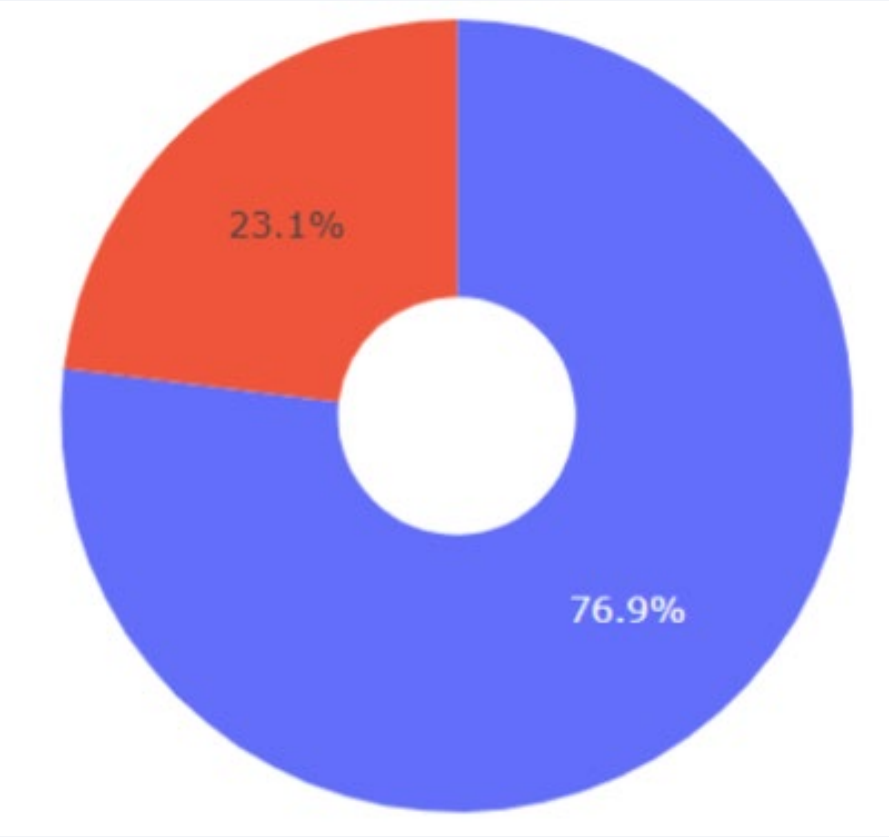
Total Launch Success Rate for All Sites



Observation:

- The Kennedy Space Center was the launch site where that resulted in the most successes!

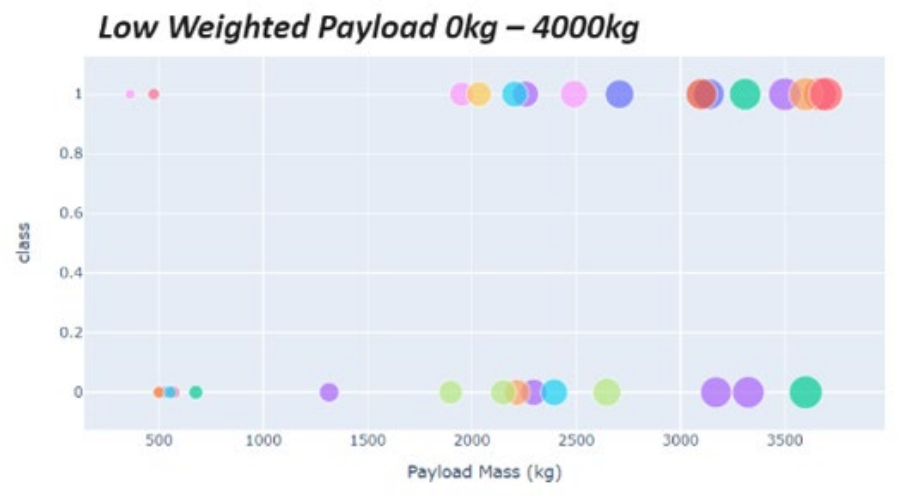
Launch Site with the highest launch Success Ratio

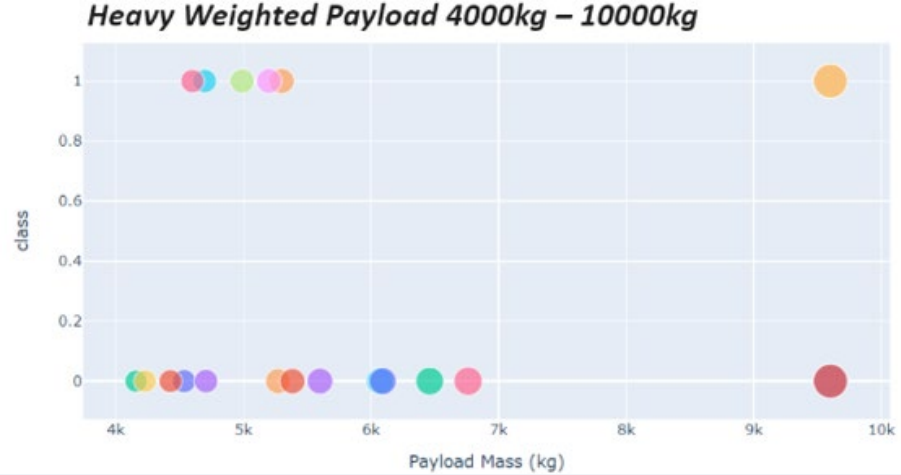


Observation:

- The Kennedy Space Center achieved the highest success rate of 76.9%!

Payload vs. Launch Outcome Scatter Plot





Observation:

- The success rate for low weighted payload was higher than the success rate of heavier weighted payload.

Section 5

Predictive Analysis (Classification)

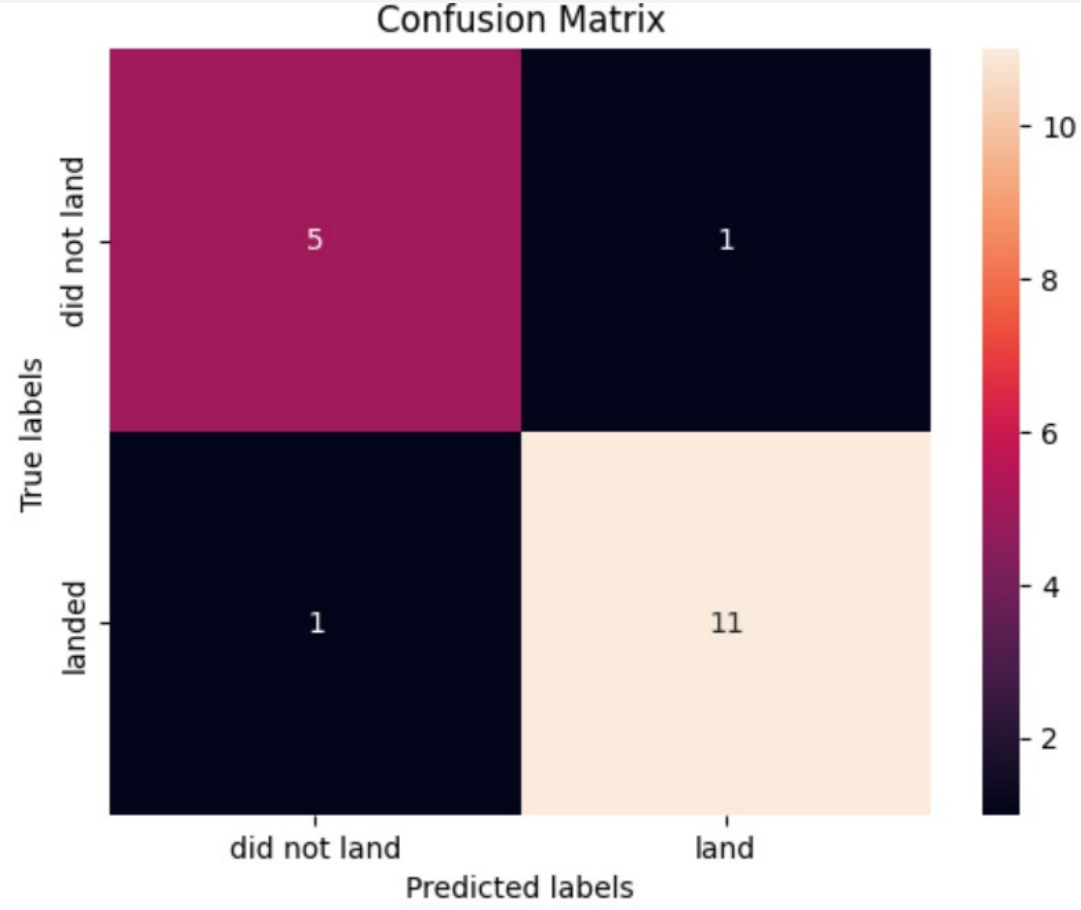
Classification Accuracy

Method	Test Data Accuracy
Logistic_Reg	0.833333
SVM	0.833333
Decision Tree	0.888889
KNN	0.833333

Observations:

- Four machine learning models were developed and run to predict the success rate.
  - Logistic Regression
  - Support Vector Machine
  - Decision Tree
  - K Nearest Neighbor
- The Decision Tree had the highest accuracy of 88.9%.

Confusion Matrix



The decision tree confusion matrix illustrates that this model resulted in:

- One false positive.
- One false negative.
- Sixteen correction predictions.
- An overall accuracy of 88.9%.

Conclusions

- There are many conclusions that were highlighted throughout this presentation. Some of the main points are listed below:
- The larger the flight amount at a launch site, the greater the success rate at a launch site.
- In 2013, the success rate was 0%. In 2020, the rate was approximately 80%. Thus, the success rate increased significantly over the 7 years.
- The success rate of orbits ES-L1, GEO, HEO, SSO, VLEO was the highest.
- The Kennedy Space Center had the highest success rate at 77% of launches successfully landed.
- The best machine learning algorithm for predicting a success/failure was the decision tree classifier.

Thank you!