IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Liu
2025/02/22

Outline

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

Introduction

- Background

  SpaceX stands out due to its reuse of Falcon 9's first stage, significantly lowering launch expenses compared to competitors.

  Therefore, for Space Y, analyzing historical launch data to understand what factors influence first-stage is important.

- Goal

  To predict if SpaceX will reuse the first stage and help Space Y to make data-driven decisions to optimize rocket launches and reduce costs.

Executive Summary

- Methodologies
  - Data collection
  - Data wrangling
  - Exploratory data analysis (EDA)
  - Interactive visual analytics
  - Predictive analysis
- Results
  - Orbit and Launch Site will affect the success rate
  - All launch sites in very close proximity to the coast
  - Decision Tree has the highest score to predict the success

Section 1

Methodology

Methodology

- Data collection
- 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

- We will gather publicly available SpaceX launch data for the analysis

- We will obtain SpaceX launch data using two methods:
  - SpaceX REST API
  - Web Scraping

Data Collection – SpaceX API

Use requests.get() to fetch historical launch data from api.spacexdata.com/v4/launches/past

Convert the JSON response into a structured Pandas DataFrame using pd.json_normalize()

Extract relevant information such as Flight Number, Date, Booster Version, Payload Mass, Orbit, Launch Site, and Landing Outcome

Use additional API endpoints to enrich missing information

Data Collection - Scraping

Use BeautifulSoup to extract Falcon 9 launch data from Wikipedia tables

Parse the HTML tables and convert them into a Pandas DataFrame

Clean and filter data to remove Falcon 1 launches, keeping only Falcon 9 records

Data Wrangling

Identify null values and replace missing values with the mean of PayloadMass

Convert Landing Outcome into a binary classification variable and apply one-hot encoding to categorical features

Remove non-relevant launch records, convert all numerical data to float64 and standardize column names for consistency

EDA with Data Visualization

- Charts I use
  - the relationship between Flight Number and Launch Site
  - the relationship between Payload Mass and Launch Site
  - the relationship between success rate of each orbit type
  - the relationship between Flight Number and Orbit type
  - the relationship between Payload Mass and Orbit type
  - the launch success yearly trend

EDA with SQL

- SQL queries
  - SELELCT DISTINCT
  - LIKE
  - sum(), avg(), min(), substr(), count()
  - GROUO BY
  - Subquery
  - ORDER BY

Build an Interactive Map with Folium

- Mark all launch sites by Circle and Marker
- Mark the success/failed launches for each site by MarkerCluster
- Mark the distances between a launch site to its proximities by MousePosition and PolyLine

Build a Dashboard with Plotly Dash

- Total Success Launches By Sites
- Total Success Launches for each site
- Correlation between Payload and Success for all Sites
- Correlation between Payload and Success for each site

Predictive Analysis (Classification)

Standardize the data

Split the data into training and test data

Use GridSearchCV to find the best parameters

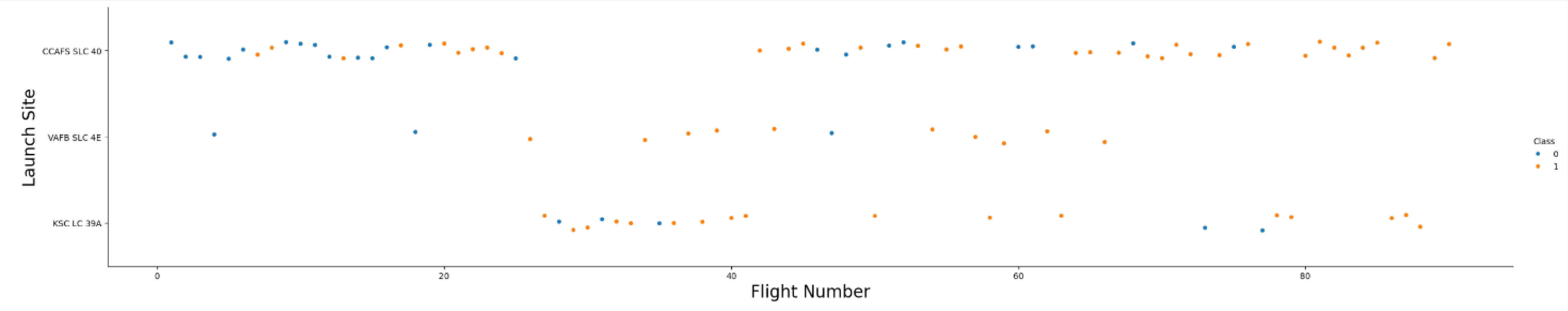
Results

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

Section 2

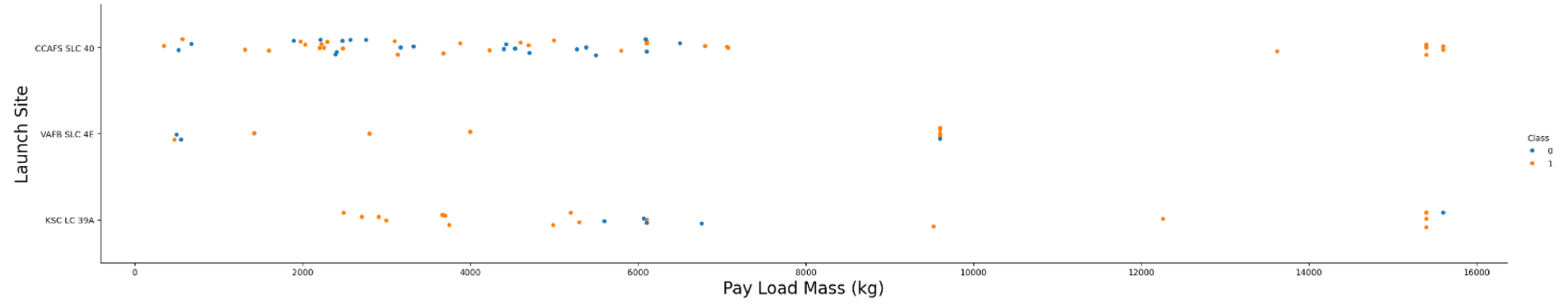
Insights drawn from EDA

Flight Number vs. Launch Site



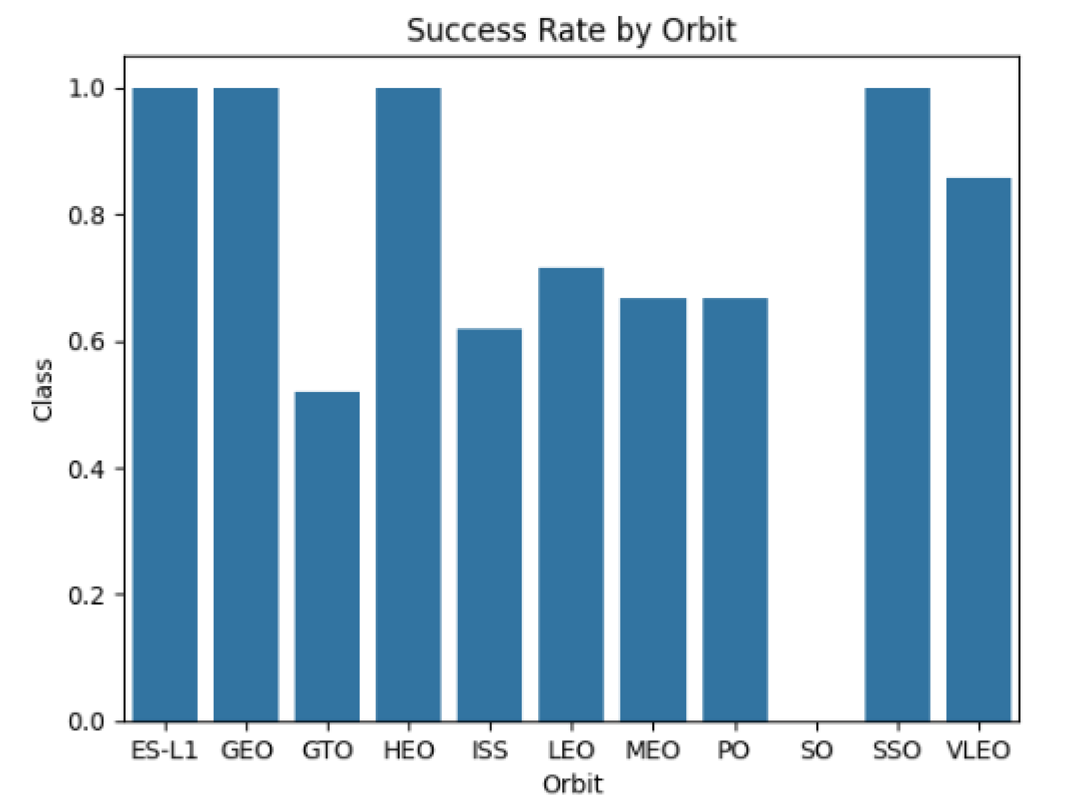
VAFB-SLC – the least

Payload vs. Launch Site



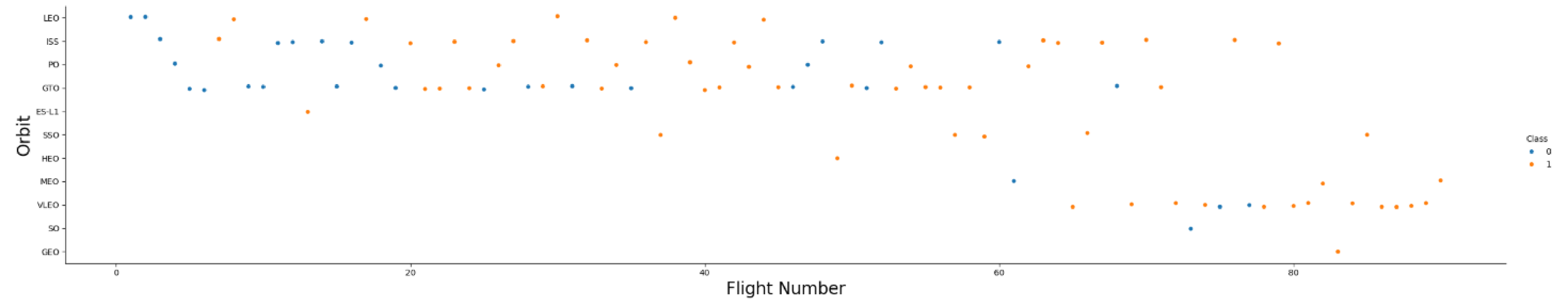
VAFB-SLC - there are no rockets launched for heavy payload mass(greater than 10000)

Success Rate vs. Orbit Type



ES-L1, GEO, HEO, SSO have the 100% success rate

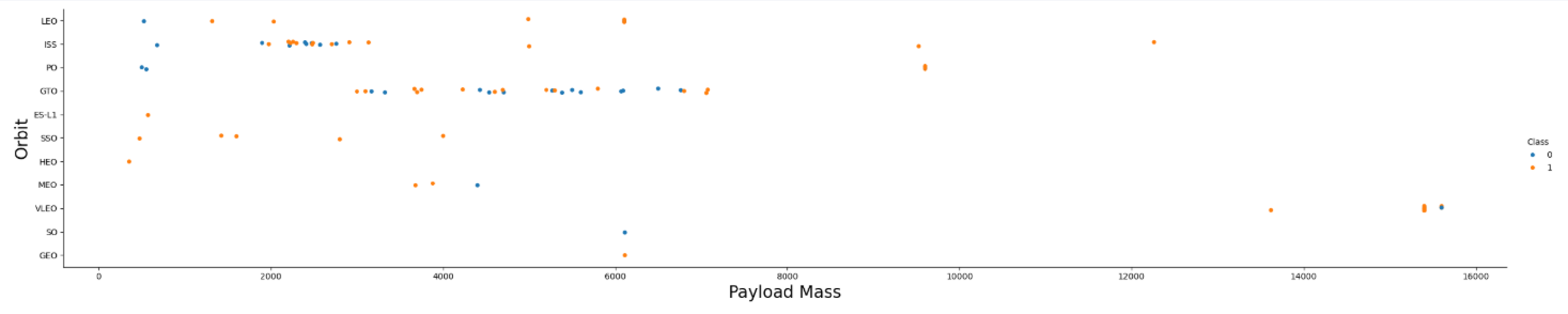
Flight Number vs. Orbit Type



LEO - success seems to be related to the number of flights

GTO - no relationship between flight number and success

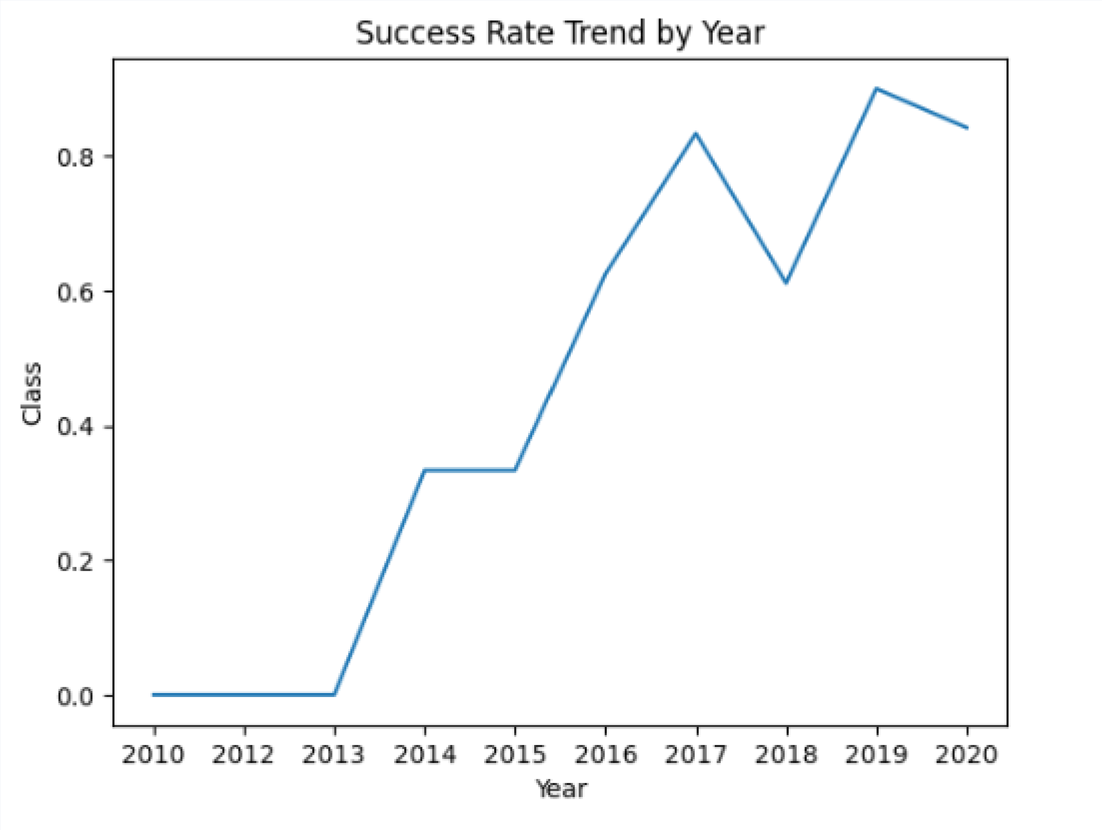
Payload vs. Orbit Type



Polar,LEO and ISS - With heavy payloads the successful rate are more

GTO – no relationship between payload and success

Launch Success Yearly Trend



the success rate was increasing

All Launch Site Names

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

There are four launch sites

Launch Site Names Begin with 'CCA'

	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

Top 5 record of launch site names begin with CCA

Total Payload Mass

sum(PAYLOAD_MASS__KG_)
45596

Total payload mass is 45596

Average Payload Mass by F9 v1.1

avg(PAYLOAD_MASS__KG_)
2928.4

Average payload mass by F9 v1.1 is 2928.4

First Successful Ground Landing Date

min("Date")
2015-12-22

- The first successful landing outcome on ground pad is on 2015/12/22

Successful Drone Ship Landing with Payload between 4000 and 6000

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

The list of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

Total Number of Successful and Failure Mission Outcomes

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

- Successful mission outcomes - 100
- Failure mission outcomes - 1

Boosters Carried Maximum Payload

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

The list of the booster which have carried the maximum payload mass

2015 Launch Records

month	Booster_Version	Launch_Site

No data

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

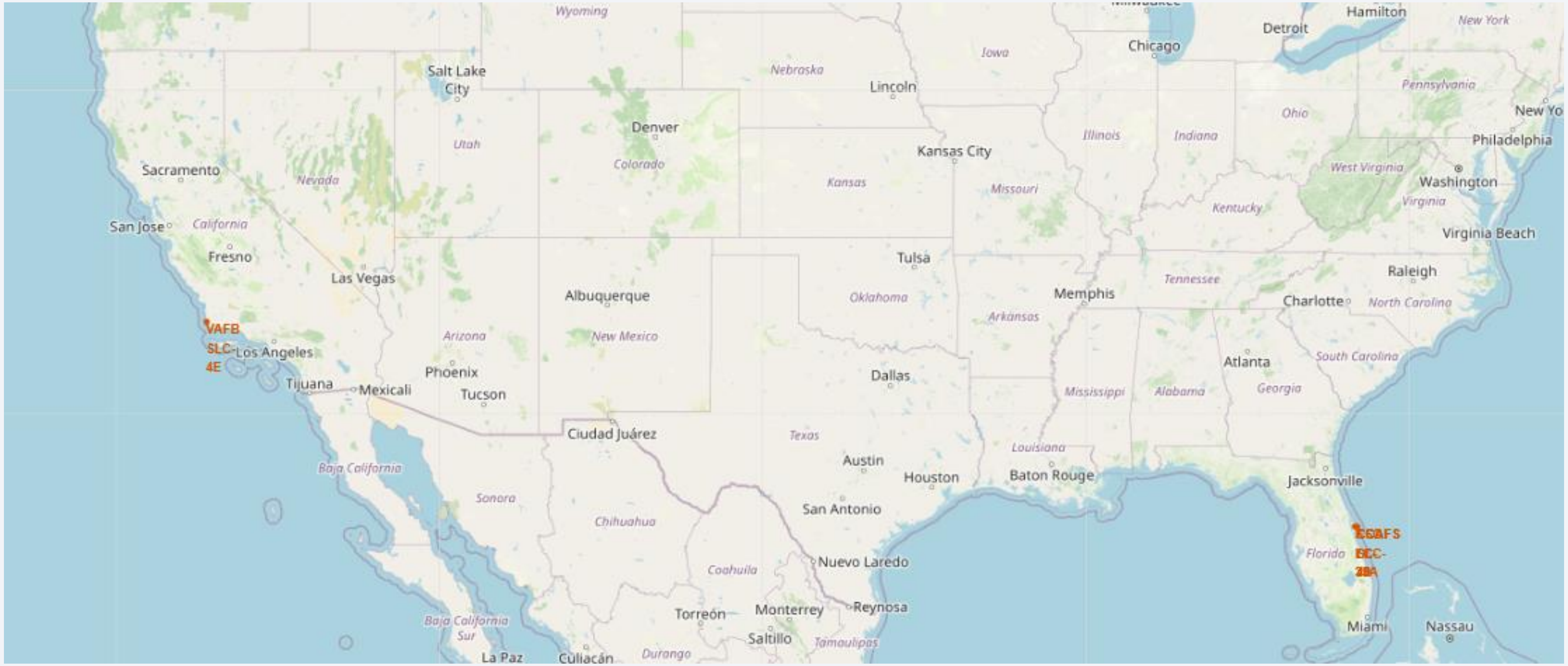
Mission_Outcome	count(*)
Success	31

Success count is 31

Section 3

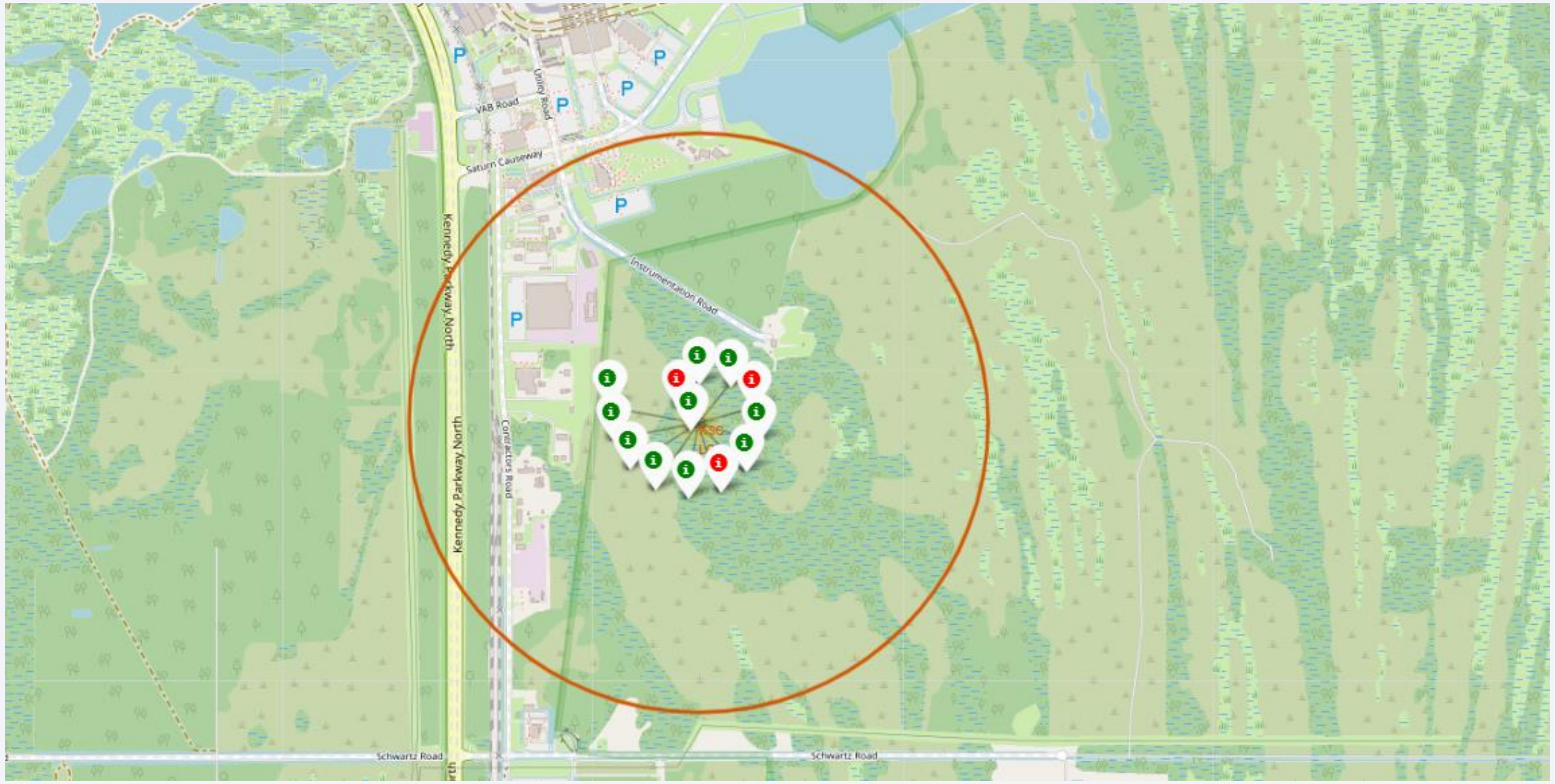
Launch Sites Proximities Analysis

all launch sites on a map



all launch sites in very close proximity to the coast

the success/failed launches for each site



KSC LC-39A has relatively high success rates

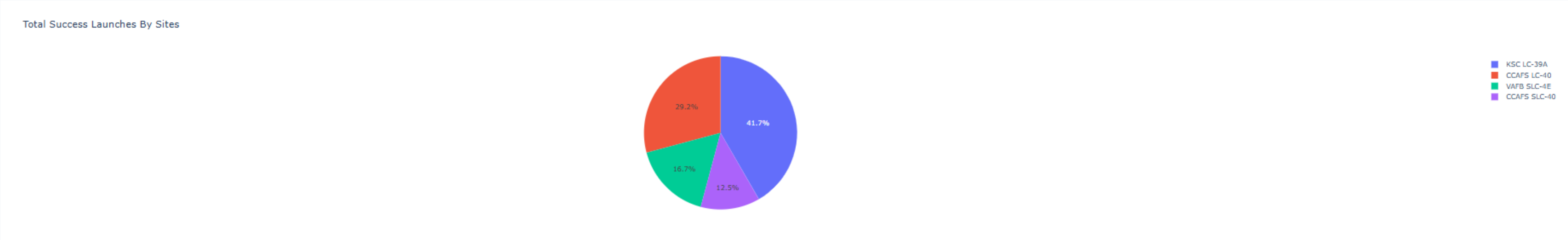
explore the proximities of launch sites

launch sites are close to coastline

Section 4

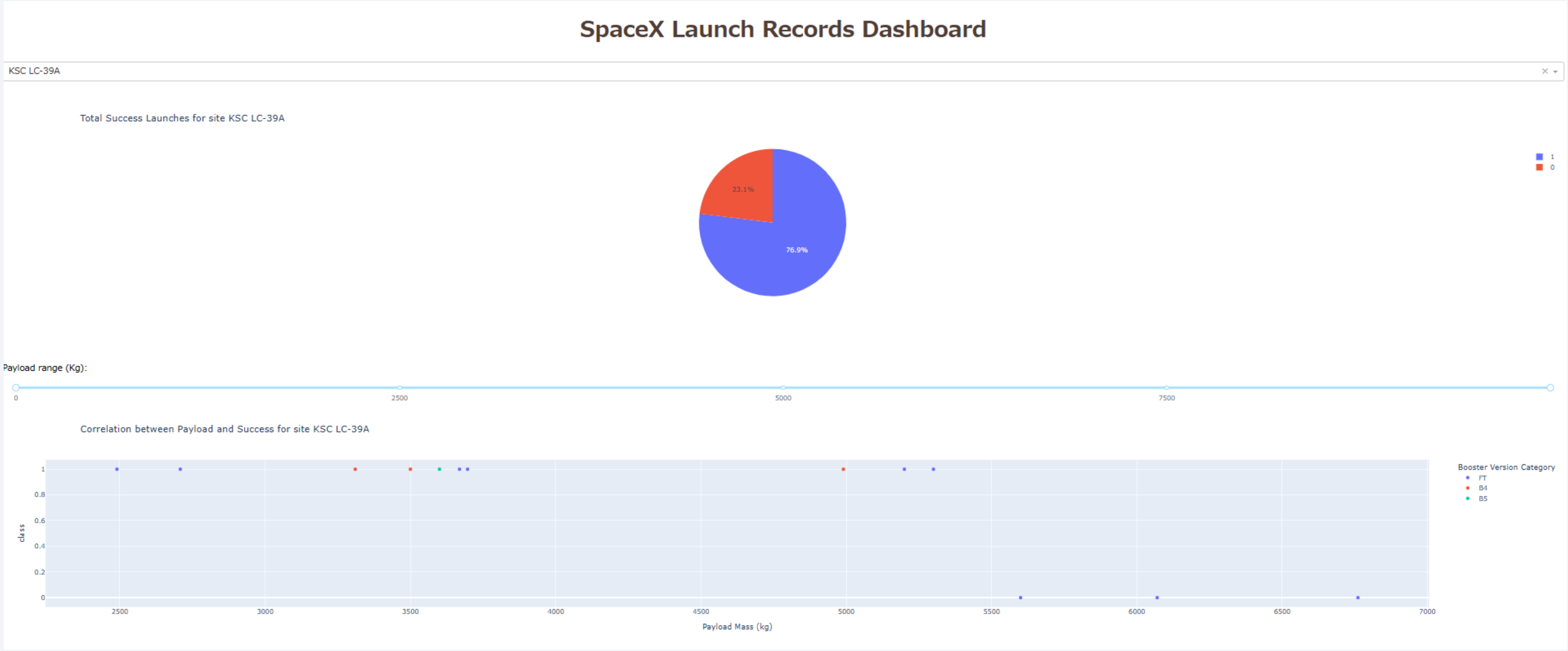
Build a Dashboard with Plotly Dash

Total Success Launches By Sites



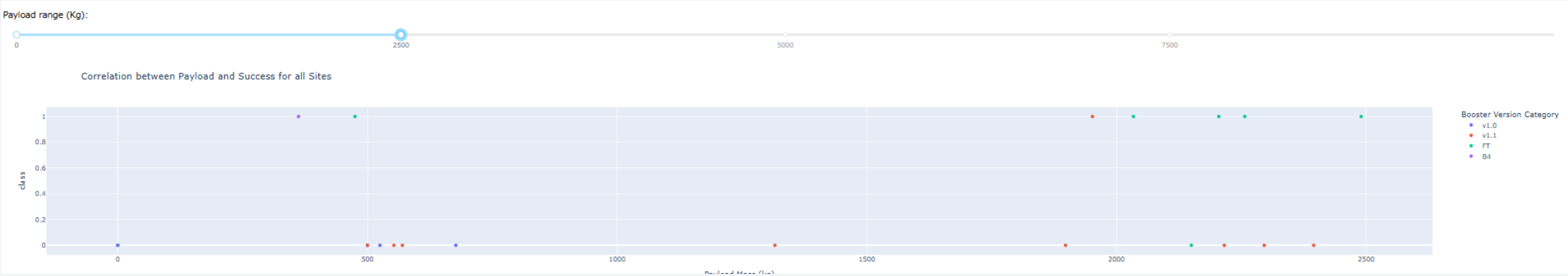
KSC LC-39A has the largest successful launches

Total Success Launches for site KSC LC-39A



KSC LC-39A has the highest launch success rate

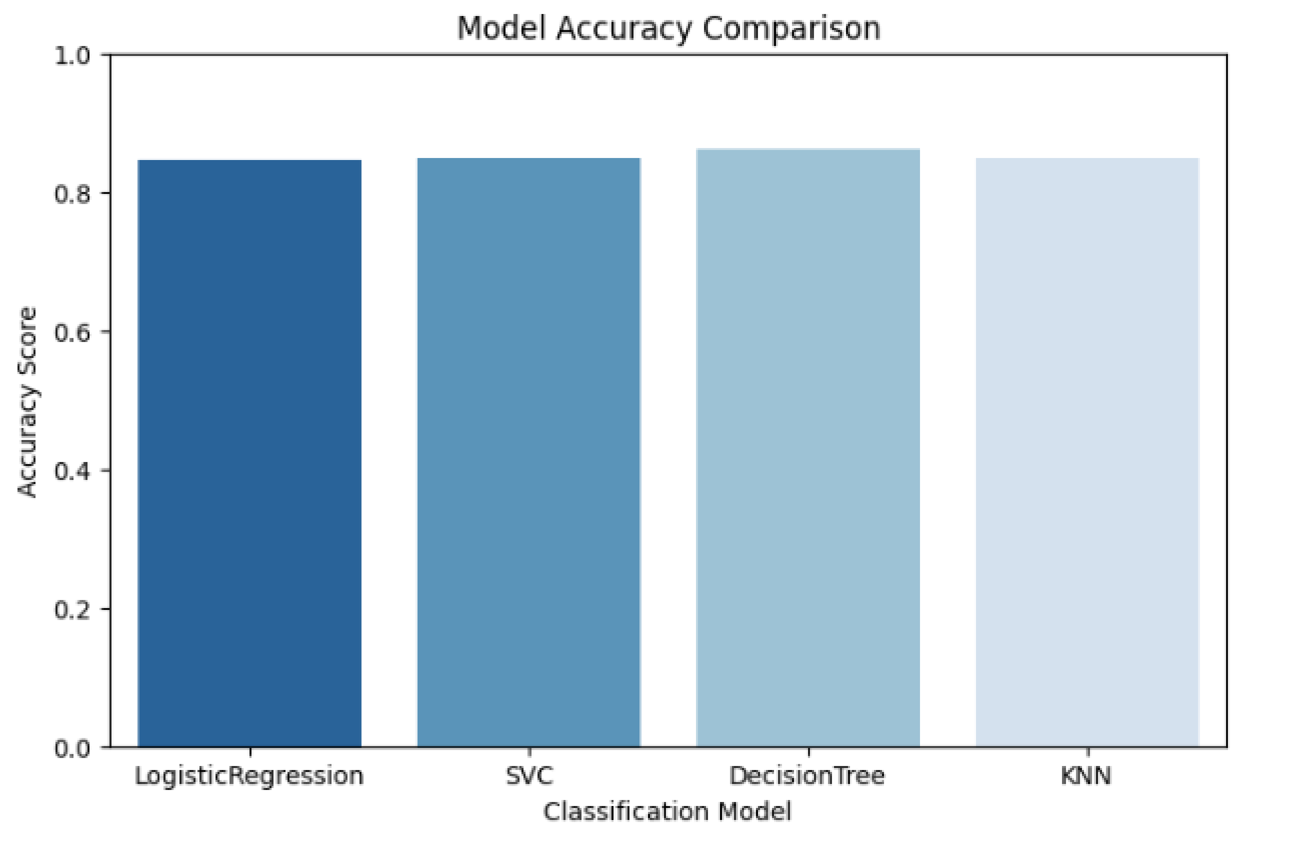
Payload vs. Launch Outcome scatter plot for all sites



Section 5

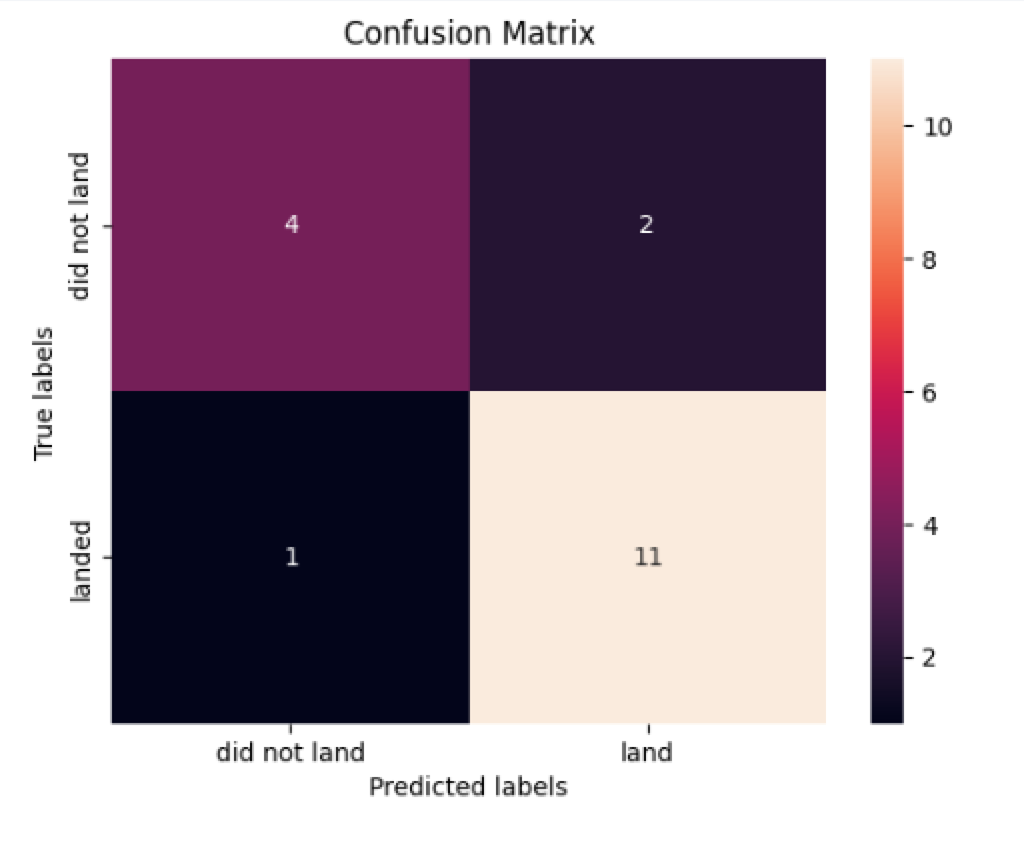
Predictive Analysis (Classification)

Classification Accuracy



Decision Tree has the highest score

Confusion Matrix



Confusion Matrix of Decision Tree

Conclusions

- Orbit ES-L1, GEO, HEO, SSO have the 100% success rate
- all launch sites in very close proximity to the coast
- KSC LC-39A has relatively high success rates
- KSC LC-39A has the largest successful launches
- Decision Tree has the highest score to predict the success

Appendix

- Github https://github.com/yuka617/spacex

Thank you!