IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Andrii Kostiuk
October 13, 2024

Outline

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

Executive Summary

- SpaceY is a new commercial rocket launch provider who wants to bid against SpaceX.
- SpaceX advertises launch services starting at $62 million for missions that allow some fuel to be reserved for landing the 1st stage rocket booster, so that it can be reused.
- SpaceX public statements indicate a 1st stage Falcon 9 booster to cost upwards of $15 million to build without including R&D cost recoupment or profit margin.
- Given mission parameters such as payload mass and desired orbit, the models produced in this report were able to predict the first stage rocket booster landing successfully with an accuracy level of 83.3%.
- As a result, SpaceY will be able to make more informed bids against SpaceX by using 1st stage landing predictions as a proxy for the cost of a launch.

Introduction

THIS REPORT IS A PART OF A PROJECT FOR APPLIED DATA SCIENCE CAPSTONE COURSE.

THIS IS A DATA SCIENCE PROJECT FOR FICTIONAL ROCKET COMPANY SPACEY TO ANALYZE THEIR MAIN COMPETITOR IN ROCKET BUILDING – SPACEX.

WITH THE HELP OF THE DATA SCIENCE FINDINGS AND MODELS IN THIS REPORT, SPACEY WILL BE ABLE TO MAKE MORE INFORMED BIDS AGAINST SPACEX FOR A ROCKET LAUNCH.

Section 1

Methodology

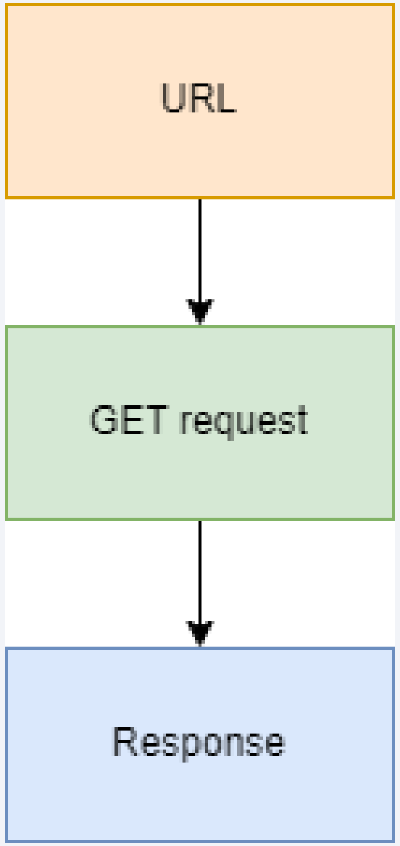
Methodology

Executive Summary

- 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

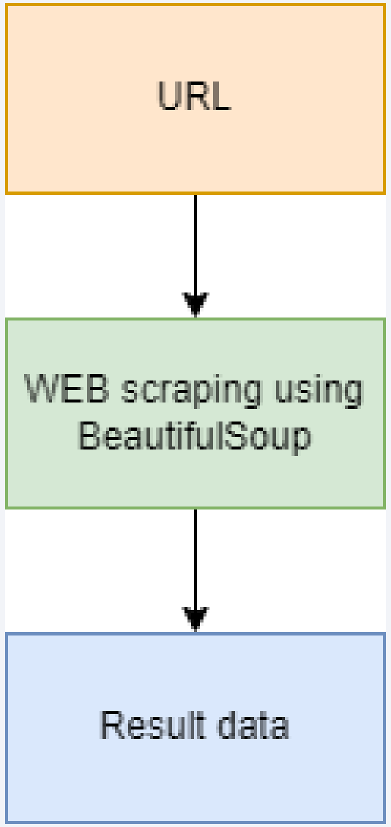
Data Collection – SpaceX API

- Acquired historical rocket launch data from Open Source REST API for SpaceX.
- Cleaned and prepared gathered data by filtering dataframe to only include Falcon 9 launches and replaced missing values.
- GitHub link

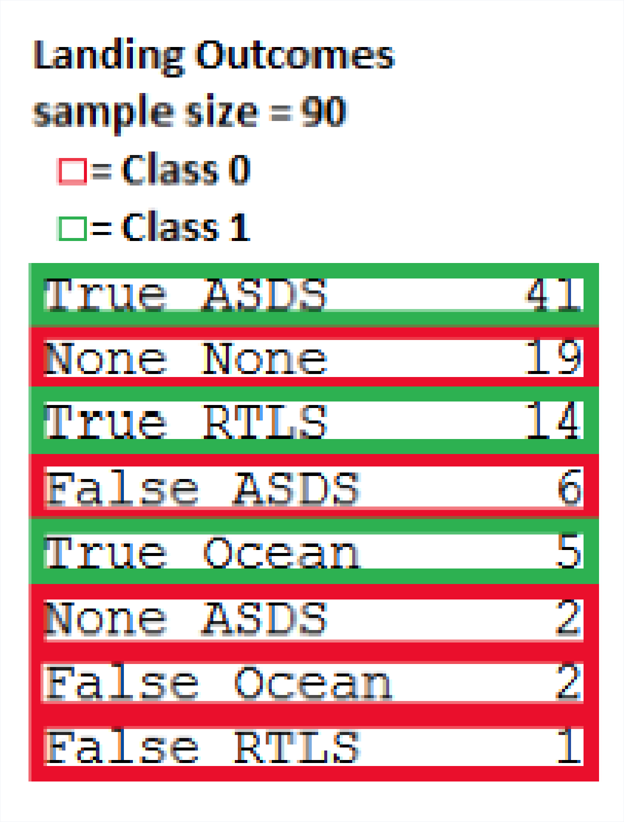


Data Collection - Scraping

- Acquired historical rocket launch data from Wikipedia page.
- Cleaned and prepared gathered data by extracting tables content and converting it into dataframe.
- GitHub link



Data Wrangling



- Processed data by exploring it to determine labels for training models
- Created new column Class with mission outcome (0 - mission failed, 1 – successed)
- [GitHub link]

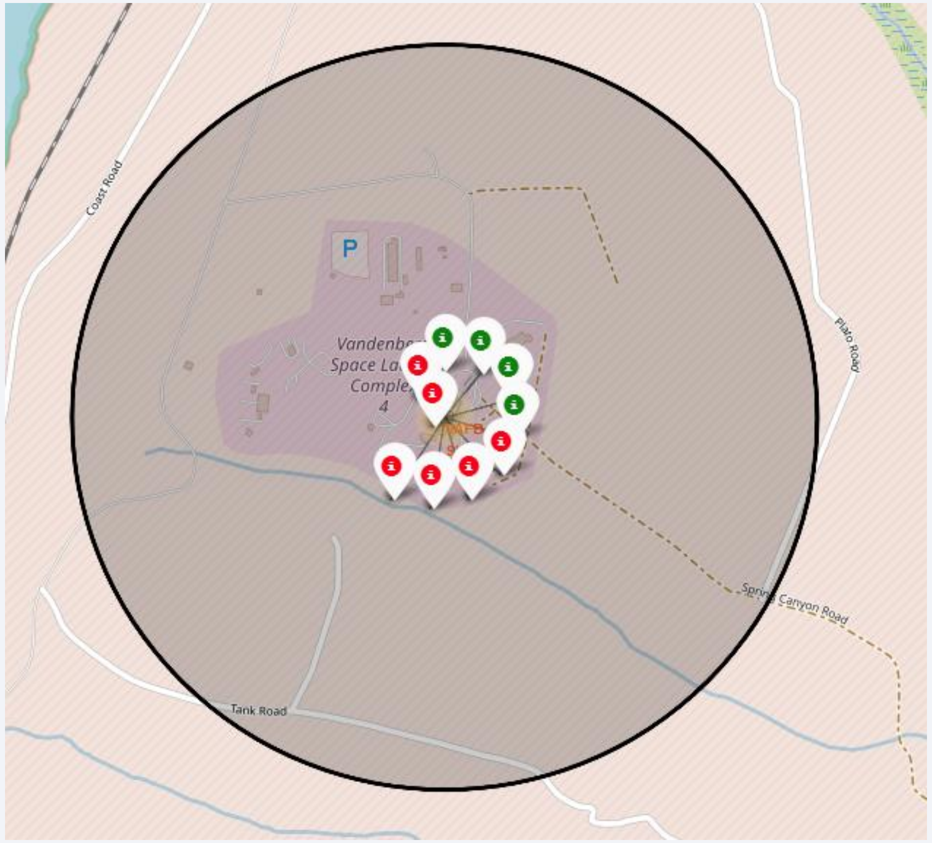
EDA with Data Visualization

- During EDA with Data Visualization were plotted:
  - scatter point charts for FlightNumber vs. PayloadMass, FlightNumber vs LaunchSite, Payload Mass vs Launch Site, FlightNumber vs Orbit type, Payload Mass vs Orbit type with overlaying outcome of the launch;
  - bar chart for relationship between success rate of each orbit type;
  - line chart for launch success yearly trend.
- GitHub link

EDA with SQL

- During EDA with SQL were used queries to list information about:
  - Launch sites
  - Booster versions
  - Booster landings
  - Payload masses
  - Mission outcomes
- GitHub link

Build an Interactive Map with Folium



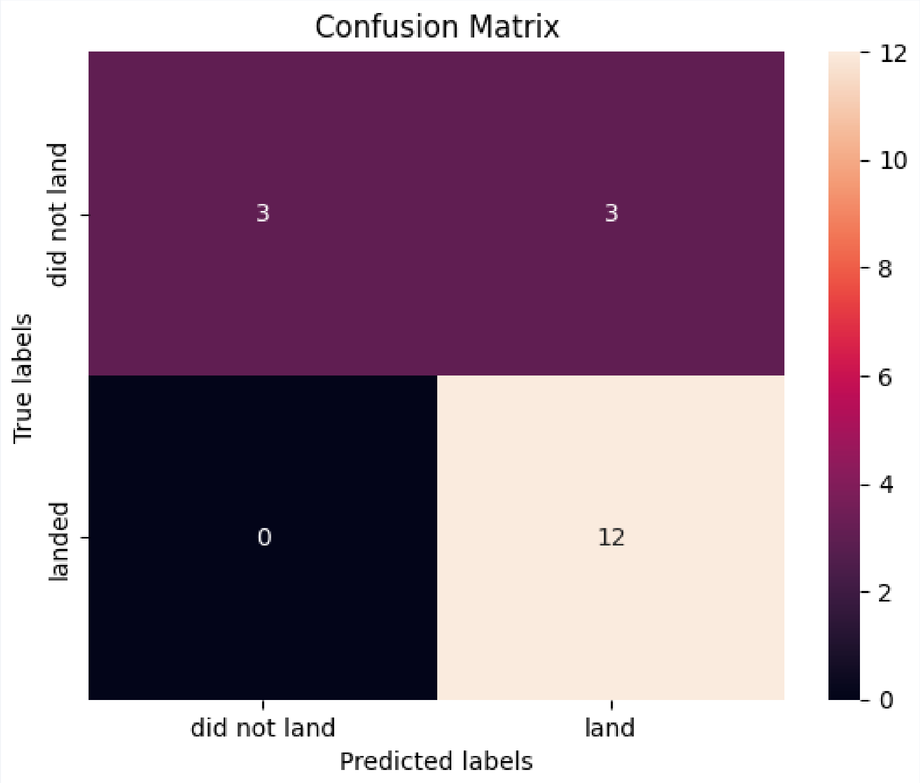
- Analyzed Launch Sites locations using interactive mapping library called Folium
- Marked all launch sites, created marker clusters, circles to represent launch sites areas
- Calculated and marked using lines distances between launch site and closest railway, highway, city, coastline
- [GitHub link]

Build a Dashboard with Plotly Dash

- Used Python interactive dashboard library called Plotly Dash to create a dashboard for stakeholders to explore and manipulate data in the interactive way
- Dashboard includes
  - Pie chart with success rate (colored based on launch site)
  - Scatter chart representing payload mass vs landing outcome (colored based on booster version)
  - Elements to control and manipulate those charts
- GitHub link

Predictive Analysis (Classification)

- For classification Predictive Analysis I did such steps:
  - Loaded data
  - Standardized it and divided into train and test groups
  - Used logistic regression for prediction with best parameters based on grid search
  - Repeated in with SVC, Decision Tree and KNN classifiers
  - Evaluated accuracy of each model using test data and confusion matrix
- [GitHub link]()



Results

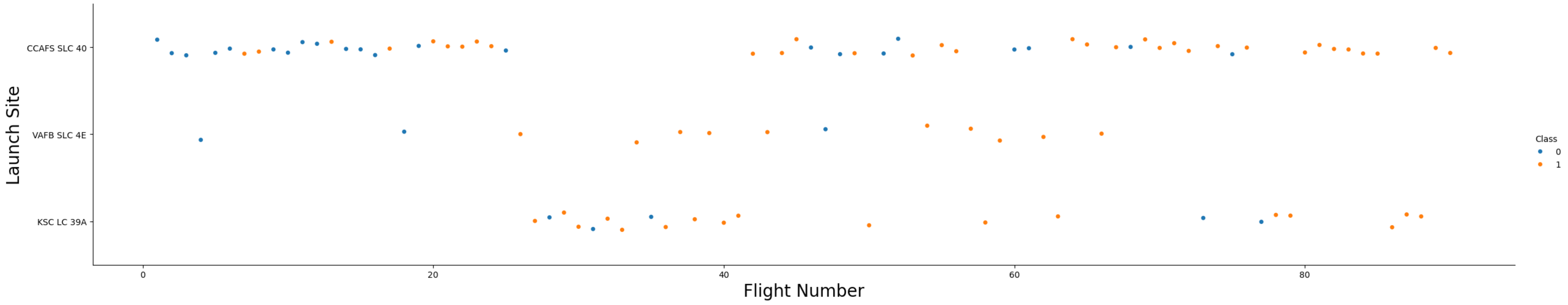
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

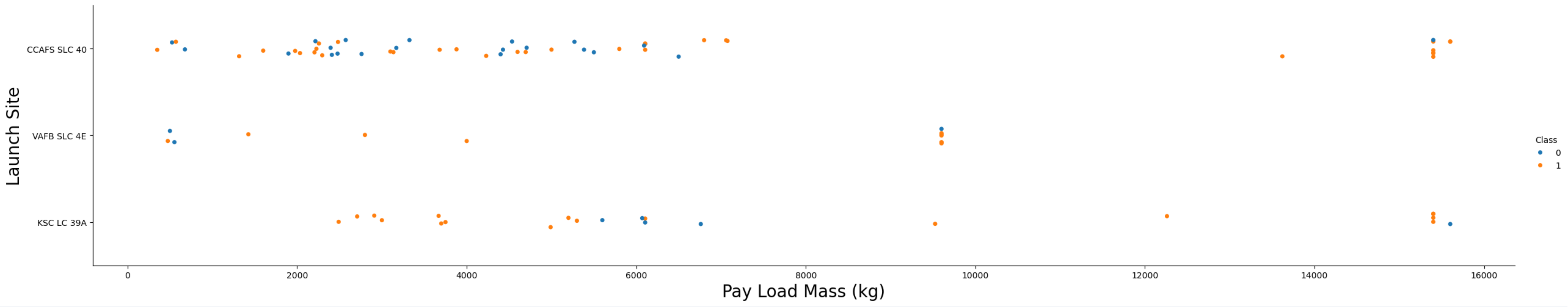
Flight Number vs. Launch Site

- Scatter plot of Flight Number vs. Launch Site
- CCAFS SLC 40 appears to have been where most of the early 1st stage landing failures took place



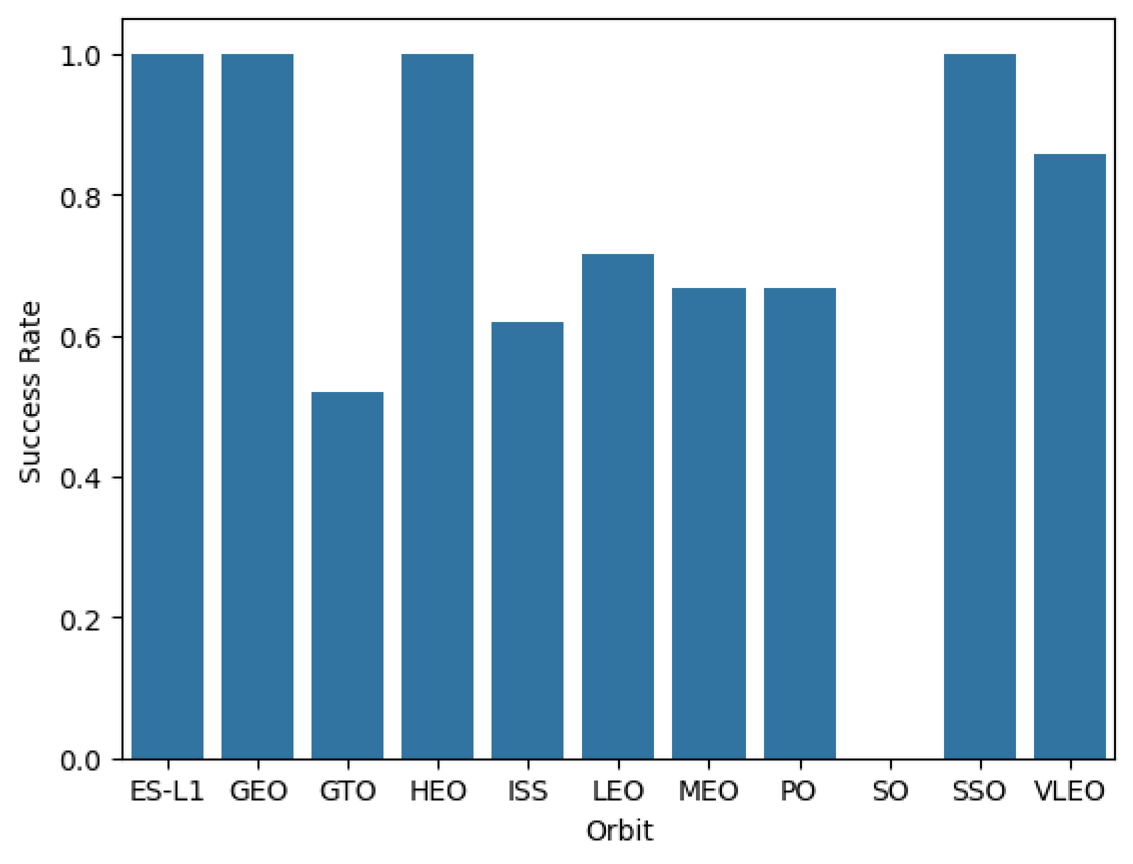
Payload vs. Launch Site

- Scatter plot of Payload vs. Launch Site
- CCAFS SLC 40 and KSC LC 39A appear to be favored for heavier payloads



Success Rate vs. Orbit Type

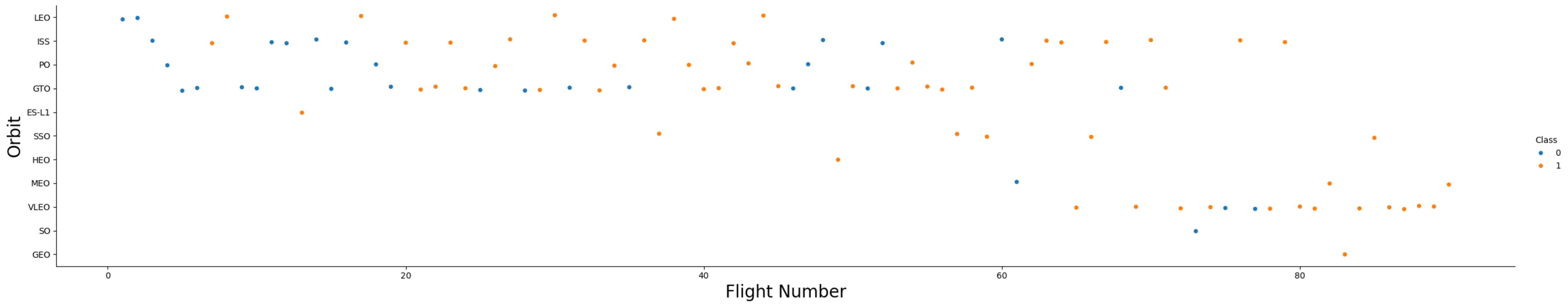
- Success rate of each orbit type
- All orbit types except 'SO' have successful 1st stage landings



Flight Number vs. Orbit Type

- Scatter point of Flight number vs. Orbit type

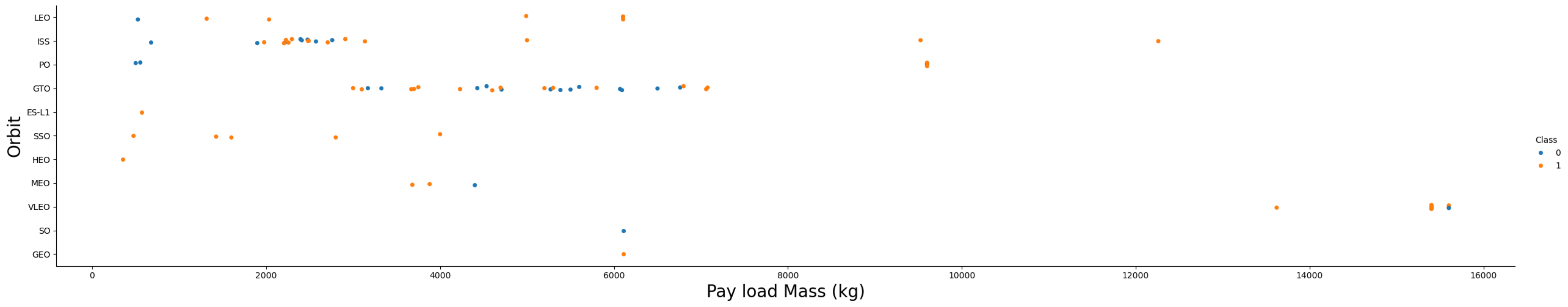
- Flight number positively correlated with 1st stage recovery for all orbit types



Payload vs. Orbit Type

- Scatter point of payload vs. orbit type

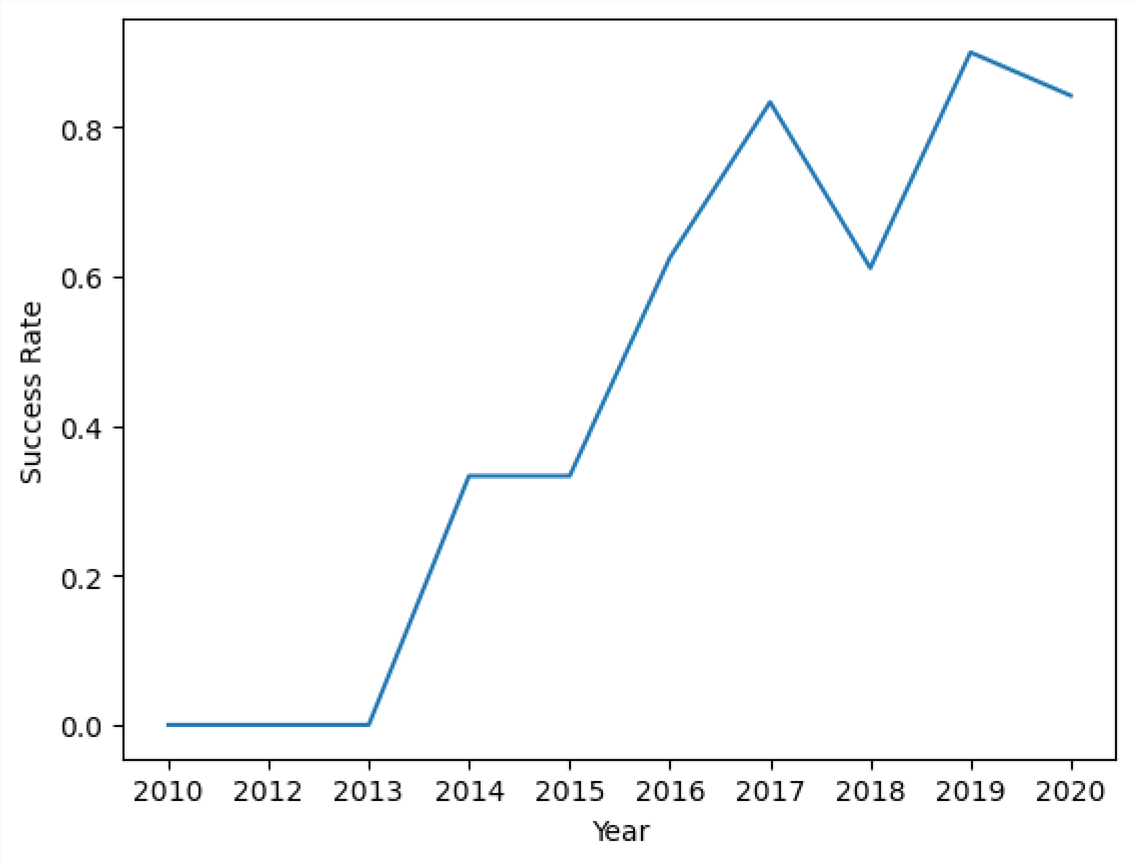
- Heavier payloads have a negative influence on GTO orbits and positive influence on ISS orbits



Launch Success Yearly Trend

- Line chart of yearly average success rate

- Success rate trending positively on a yearly basis since 2013



All Launch Site Names

Unique launch sites

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

Launch Site Names Begin with 'CCA'

- 5 records where launch sites 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

Total Payload Mass

- The total payload carried by boosters from NASA

Customer	sum("PAYLOAD_MASS__KG_")
NASA (CRS)	45596

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1

Booster_Version	avg("PAYLOAD_MASS__KG_")
F9 v1.1	2928.4

First Successful Ground Landing Date

- Date of the first successful landing outcome on ground pad

min("Date")
2015-12-22

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

Booster_Version	PAYLOAD_MASS__KG_
F9 FT B1022	4696
F9 FT B1026	4600
F9 FT B1021.2	5300
F9 FT B1031.2	5200

Total Number of Successful and Failure Mission Outcomes

- Total number of successful and failure mission outcomes

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

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass
- Present your query result with a short explanation here

F9 B5 B1048.3	4850
F9 B5 B1048.4	15600
F9 B5 B1048.5	15600
F9 B5 B1049.2	9600
F9 B5 B1049.3	13620
F9 B5 B1049.4	15600
F9 B5 B1049.5	15600
F9 B5 B1049.6	15440
F9 B5 B1049.7	15600

2015 Launch Records

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

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

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

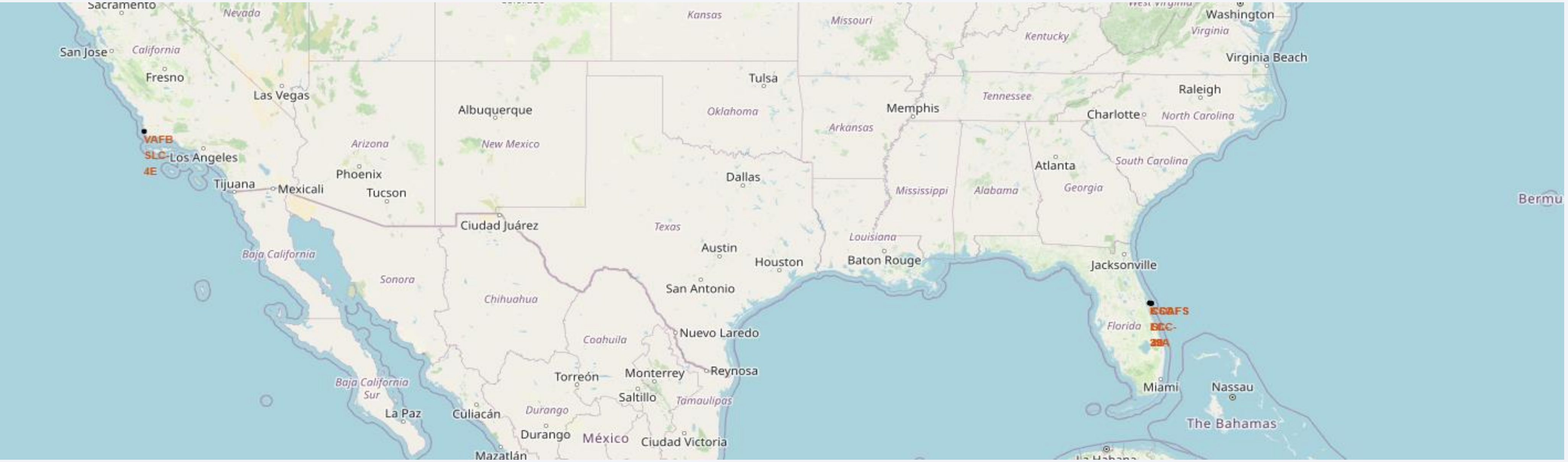
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

Landing_Outcome	count(*)	RANK () OVER (order by count(*) DESC)
No attempt	10	1
Success (drone ship)	5	2
Failure (drone ship)	5	2
Success (ground pad)	3	4
Controlled (ocean)	3	4
Uncontrolled (ocean)	2	6
Failure (parachute)	2	6
Precluded (drone ship)	1	8

Section 3

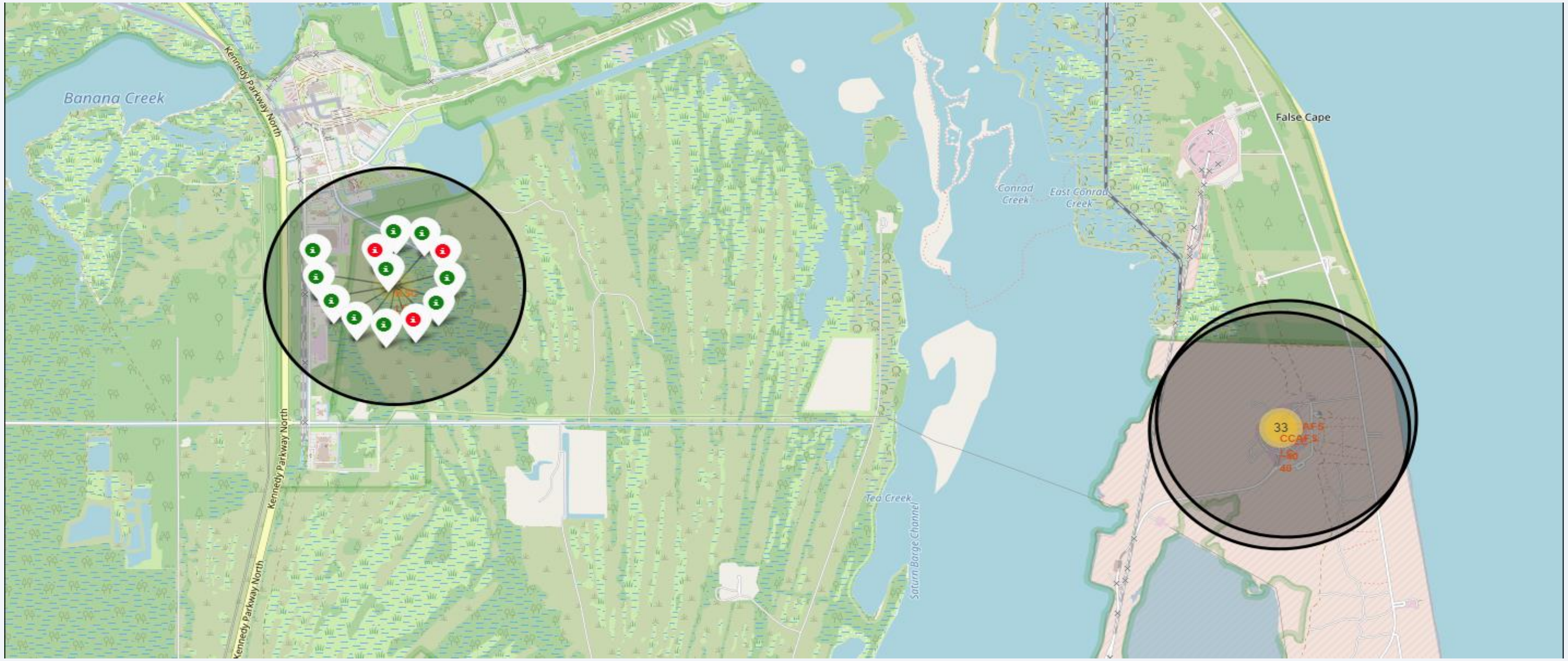
Launch Sites Proximities Analysis

Launch Sites Locations Map



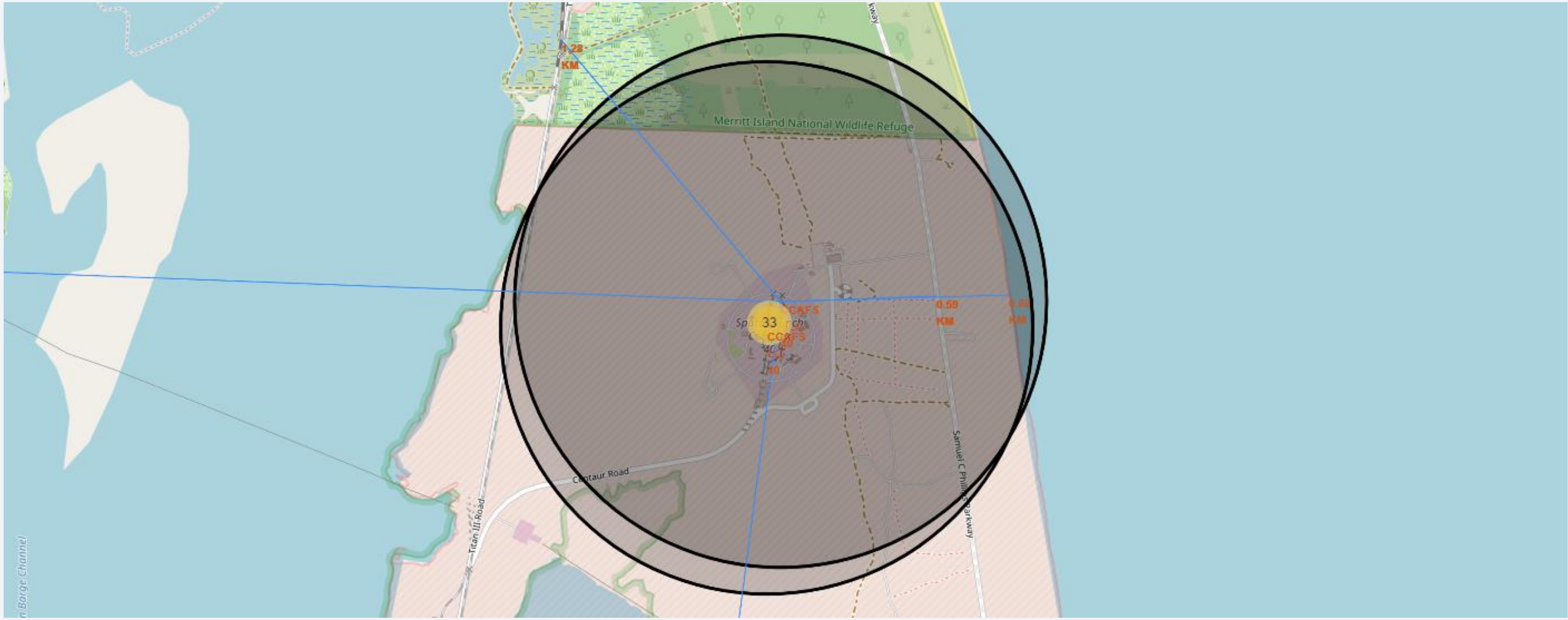
- Visualizing launch sites on a map highlights the importance of launch site geographic location

Launches clusters



- Visualizing launches clusters on launches sites marked by color based on their result (green – successful, red – not)

Distances Visualization

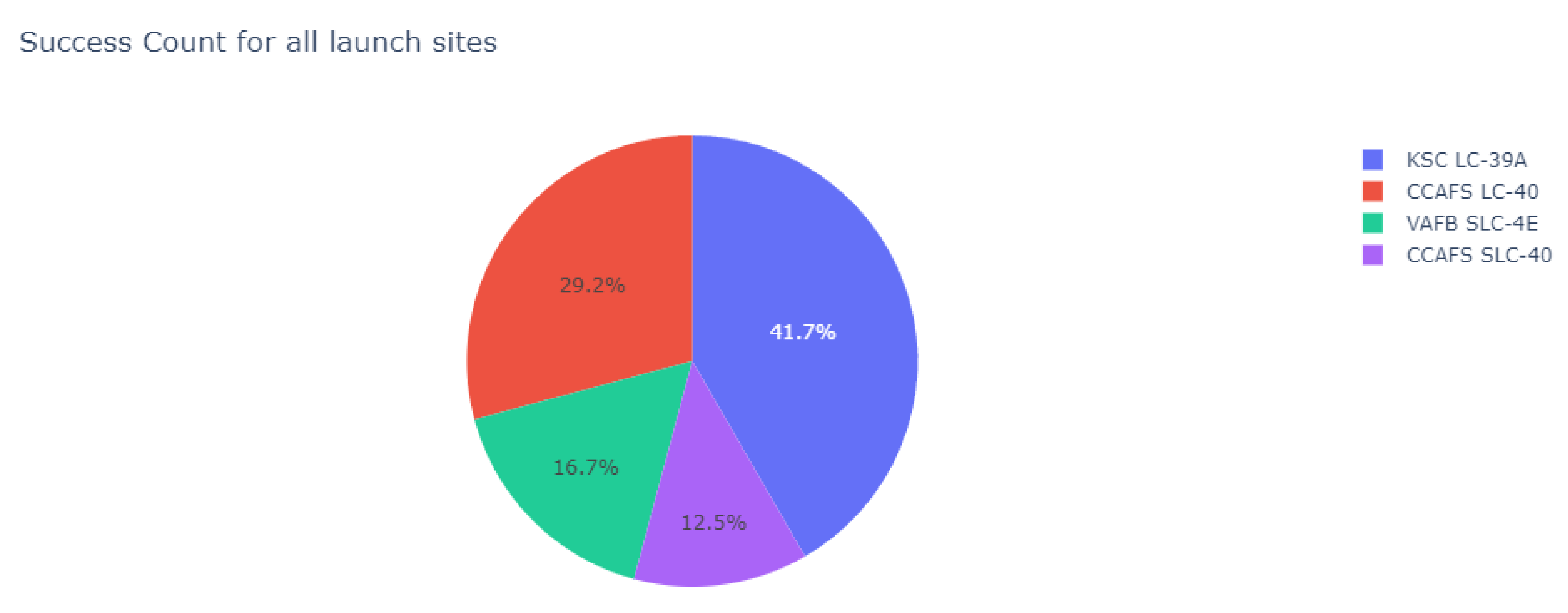


- Analyzing distances between launch sites and closest town, railway, highway, coastline

Section 4

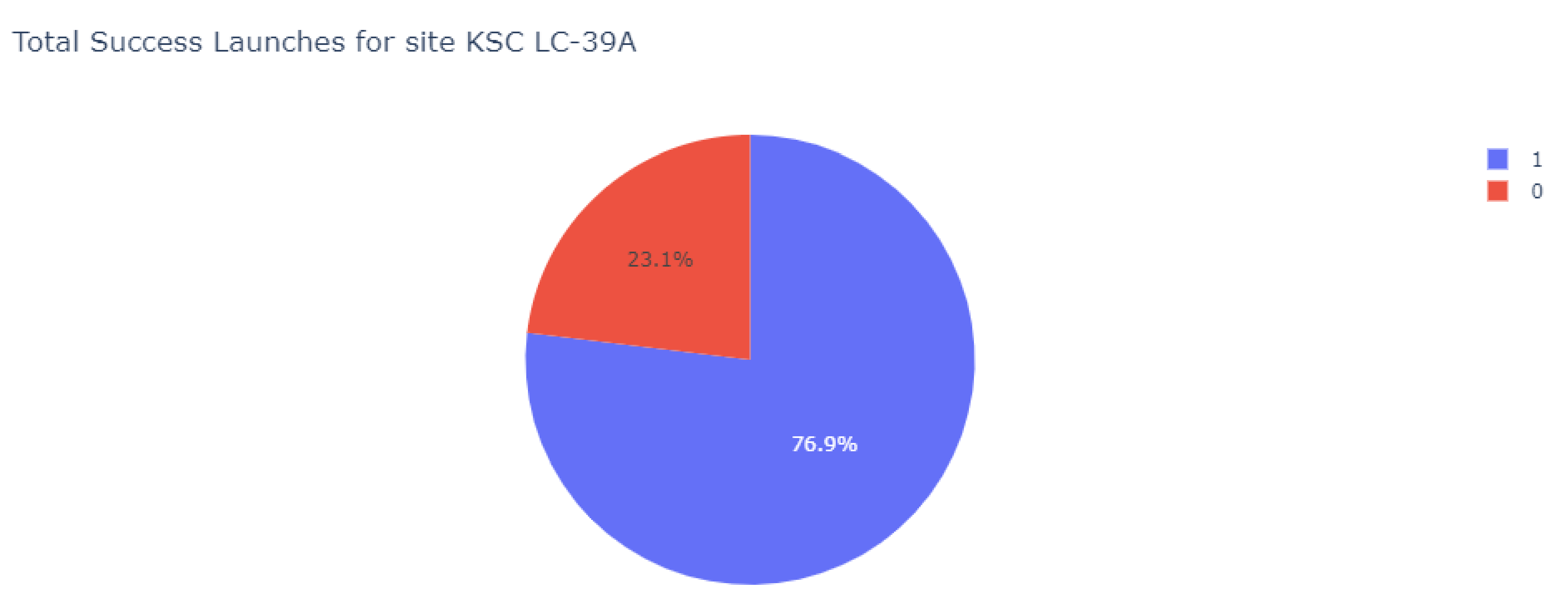
Build a Dashboard with Plotly Dash

Success Count for all launch sites



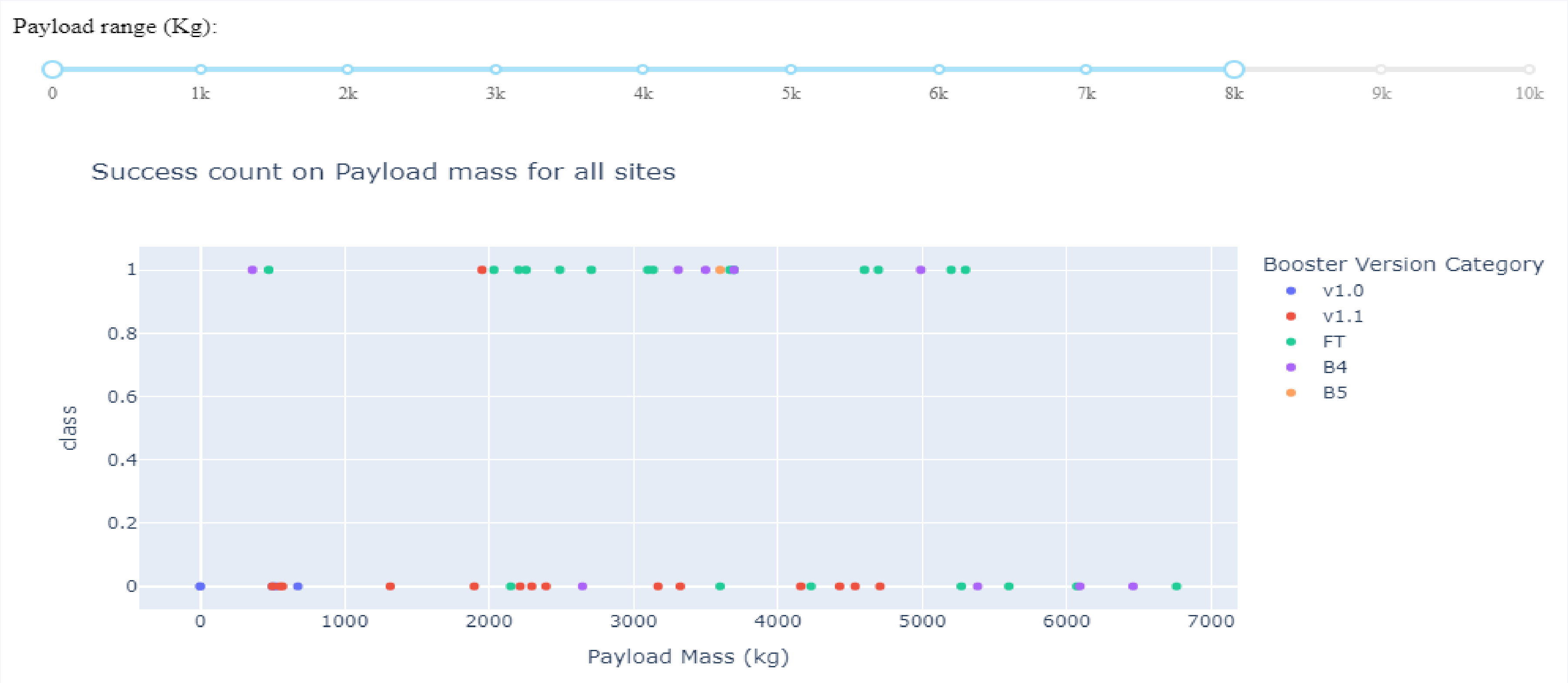
- Pie chart showing booster landing success rate colored based on launch site

Launch Site with highest success rate



- KSC LC-39A is a launch site with highest success rate of booster landing – 76.9%

Success count based on Payload mass

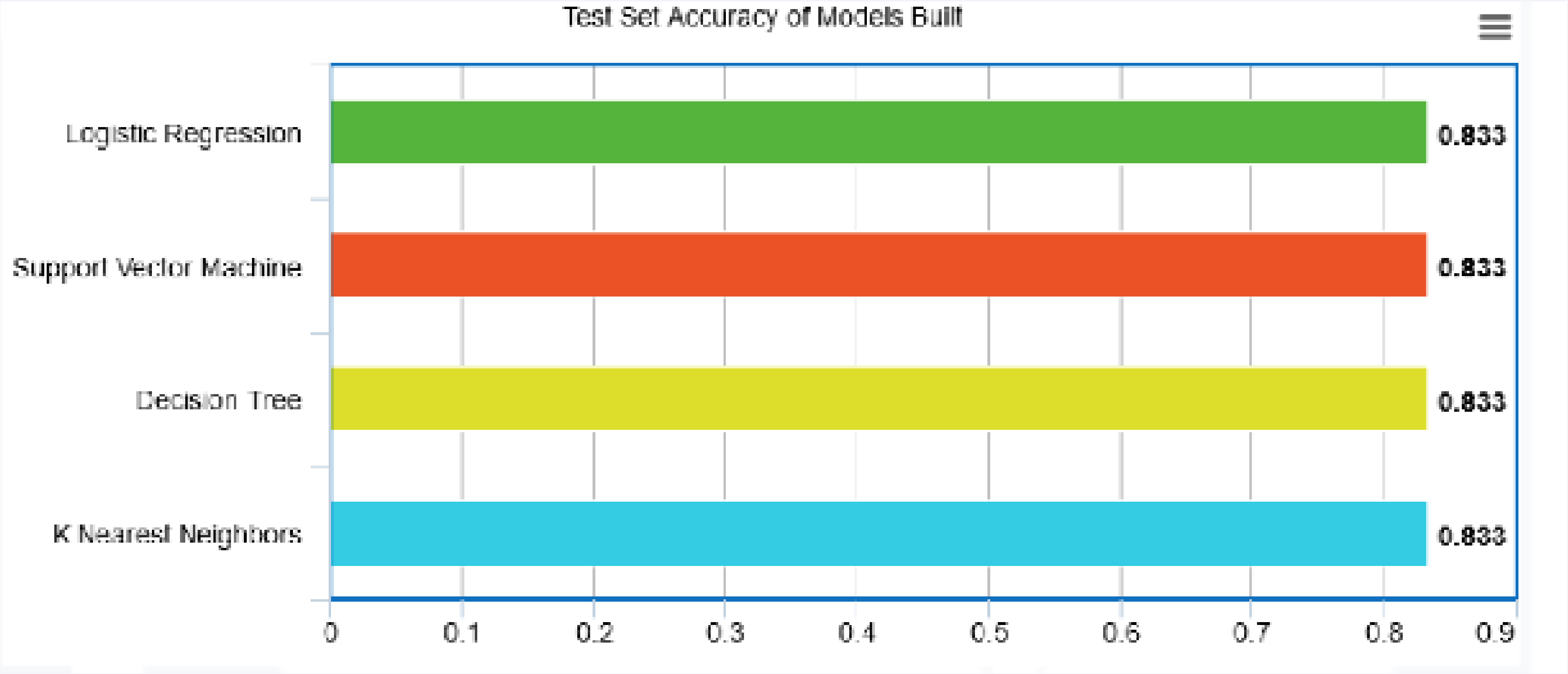


- Success count for each launch site based on payload mass with payload mass range between 0 and 8000 kg

Section 5

Predictive Analysis (Classification)

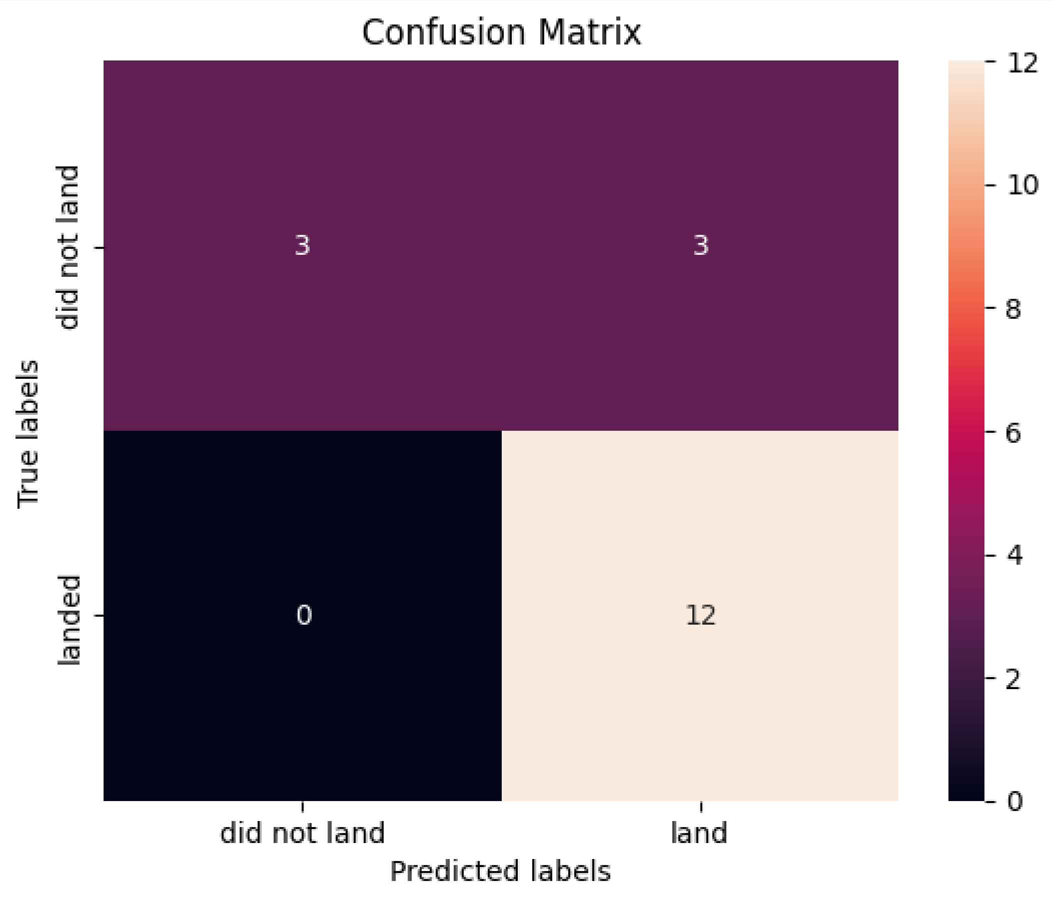
Classification Accuracy



- All models have the same accuracy score of 83.33%

Confusion Matrix

- All models perform the same with such confusion matrix



Conclusions

- Using the models from this report Spacey can predict when SpaceX will successfully land the 1st stage booster with 83.3% accuracy
- SpaceX public statements indicate the 1st stage booster costs upwards of $15 million to build
- This will enable Spacey to make more informed bids against SpaceX, since they will have a good idea when to expect the SpaceX bid to include the cost of a sacrificed 1st stage booster
- With a list price of $62 million per launch, sacrificing the $15+ million 1st stage, would put the SpaceX bid at upwards of $77 million

Appendix

- Project repository
  - https://github.com/KostiukAS/Course-DS-Capstone/tree/main

Thank you!