IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Mawufemor Kofi Romio Dotse
10-02-2024

Outline

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

Executive Summary

- Summary of methodologies
  - Data collection
  - Data wrangling
  - EDA with SQL
  - Building an interactive map with Folium
  - Building a Dashboard with Plotly Dash
  - Predictive analysis(Classification)
- Summary of all results
  - EDA results
  - Interactive analytics
  - Predictive analysis

Introduction

- Project background and context
    - SpaceX advertises Falcon 9 rocket launches on its website, with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage.
- Problems you want to find answers
    - The project task is to predict if the first stage of the SpaceX Falcon 9 rocket will land successfully.

Section 1

Methodology

Methodology

Executive Summary

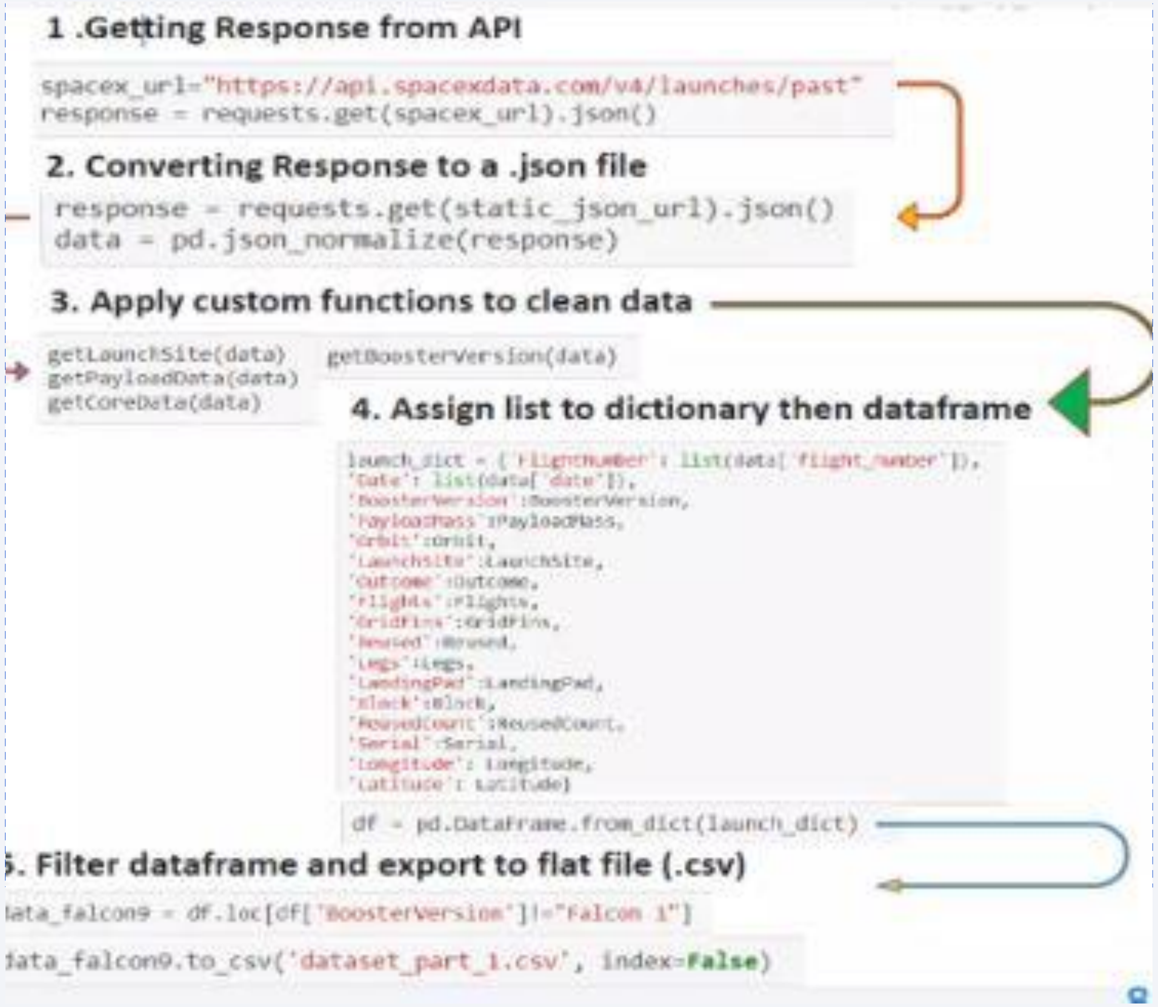
- Data collection methodology:
    - SpaceX Rest API
    - Web Scrapping from Wikipedia
- Perform data wrangling
    - One Hot Encoding data field for Machine Learning and data cleaning of null values and irrelevant columns
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
    - LR, KNN, SVM, and DT modules have been built and evaluated for the best classifier

Data Collection

- The following data sets were collected.
  - SpaceX launches data that is gathered from SpaceX REST API.
  - This API will give us data about launches, including information about the rocket used, payload delivered, launch specifications, landing specifications, and landing outcome.
  - The SpaceX REST API endpoints, or URL, starts with api.spacexdata/v4/.
  - Another popular data source for obtaining Falcon 9 Launch data is web scrapping Wikipedia using BeautifulSoup

Data Collection – SpaceX API

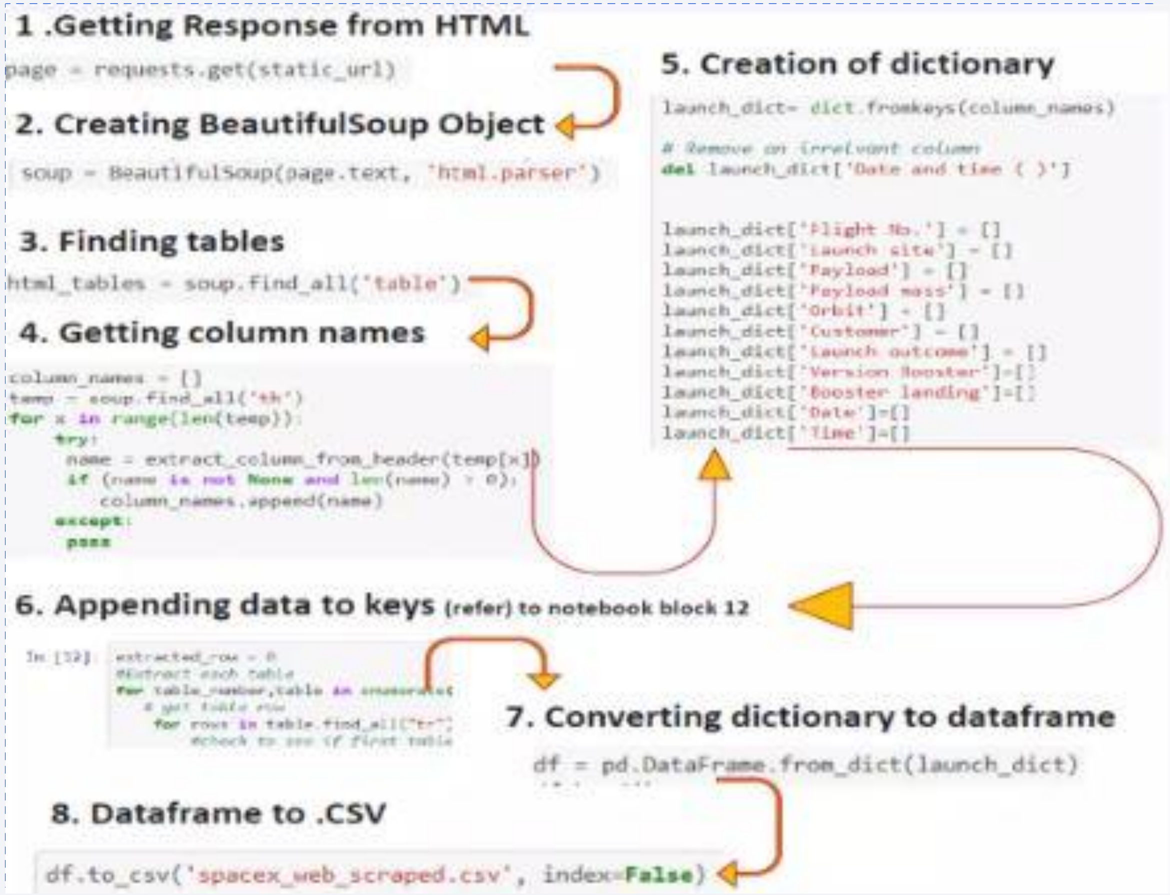
- Data collection with SpaceX REST calls
- https://github.com/KofiRomio/DataScienceEcosystem/blob/master/10-IBM%20DS%20Capstone%20project-lab2-web%20scraping.ipynb



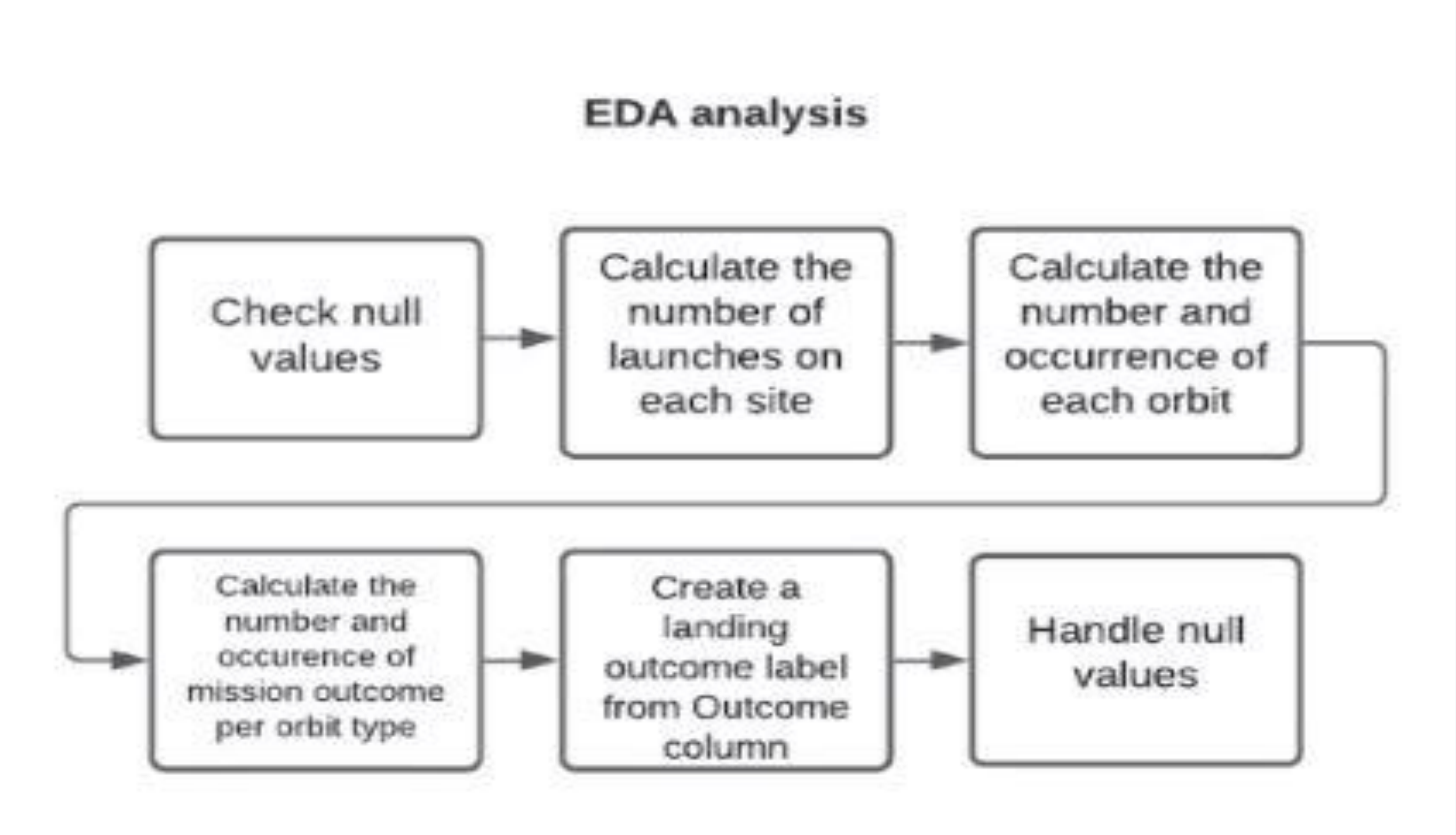
Data Collection - Scraping

- Web Scrapping from Wikipedia

- https://github.com/KofiRomio/DataScienceEcosystem/blob/master/10-IBM%20DS%20Capstone%20project-lab2-web%20scraping.ipynb

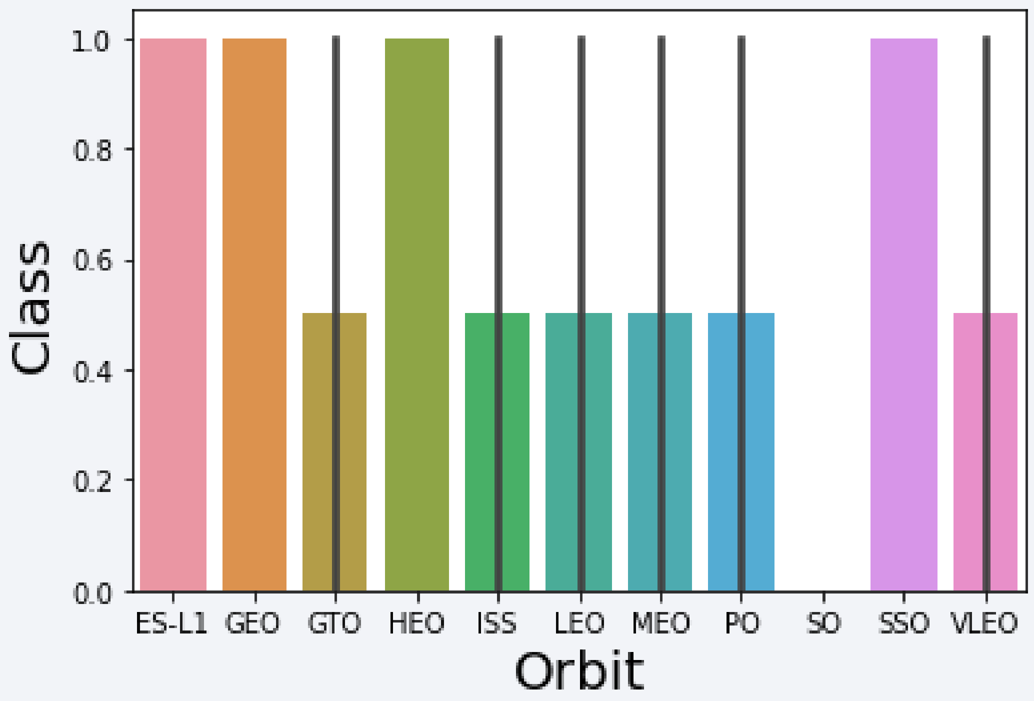


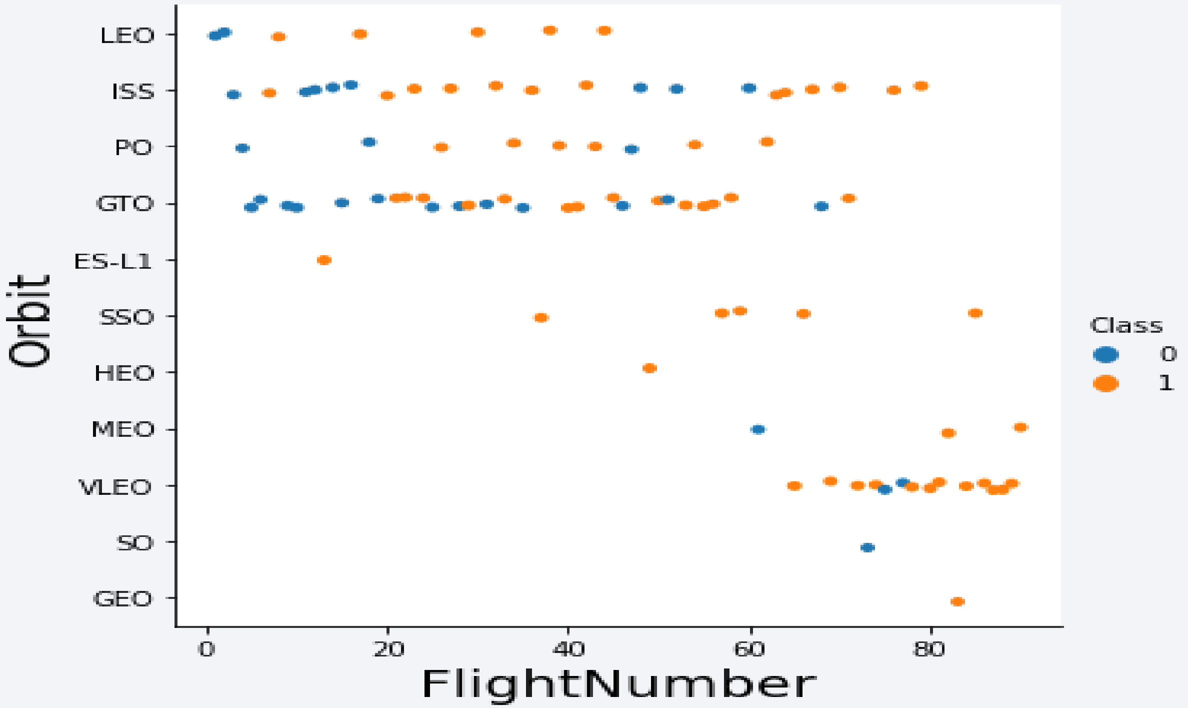
Data Wrangling



https://github.com/KofiRomio/DataScienceEcosystem/blob/master/10-IBM%20DS%20Capstone-lab4-EDA%20viz.ipynb

EDA with Data Visualization





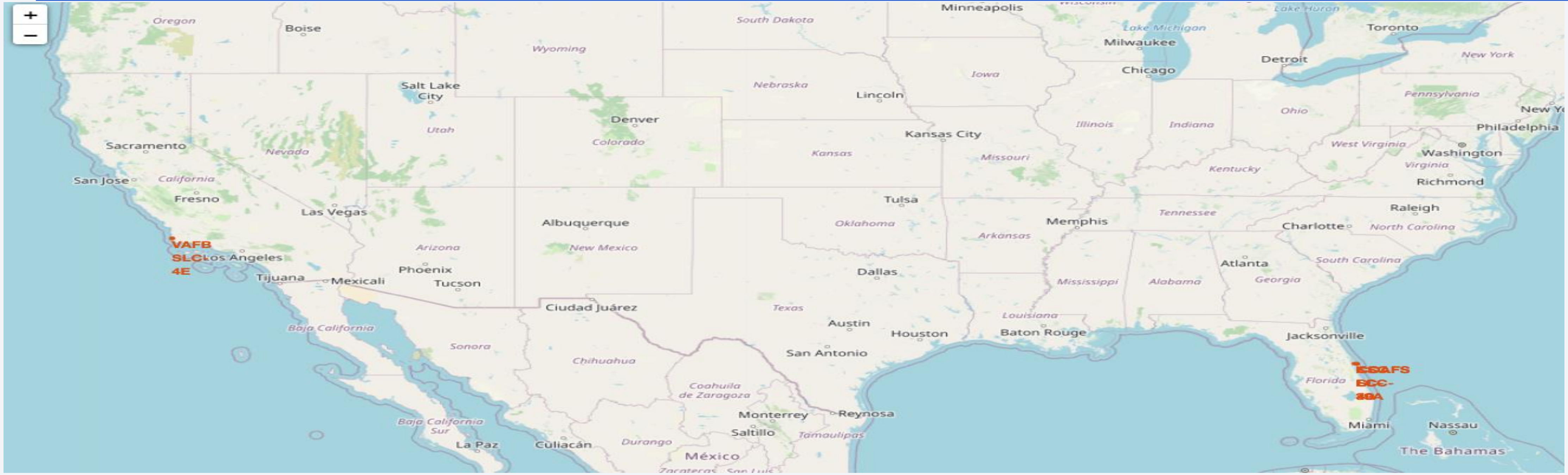
https://github.com/KofiRomio/DataScienceEcosystem/blob/master/10-IBM%20DS%20Capstone-lab4-EDA%20viz.ipynb

EDA with SQL

- SQL queries performed include:
  - Display the names of the unique launch sites in the mission
  - Display 5 records where launch sites begin with the string 'CCA‘
  - Display the total payload mass carried by boosters launched by NASA (CRS)
  - Display average payload mass carried by booster version F9 v1.1
  - List the date when the first succesful landing outcome in ground pad was acheived.
  - List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
  - List the total number of successful and failure mission outcomes
  - List the names of the booster_versions which have carried the maximum payload mass. Use a subquery
  - 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.

https://github.com/KofiRomio/DataScienceEcosystem/blob/master/10-IBM%20DS%20Capstone-lab4-EDA%20viz.ipynb

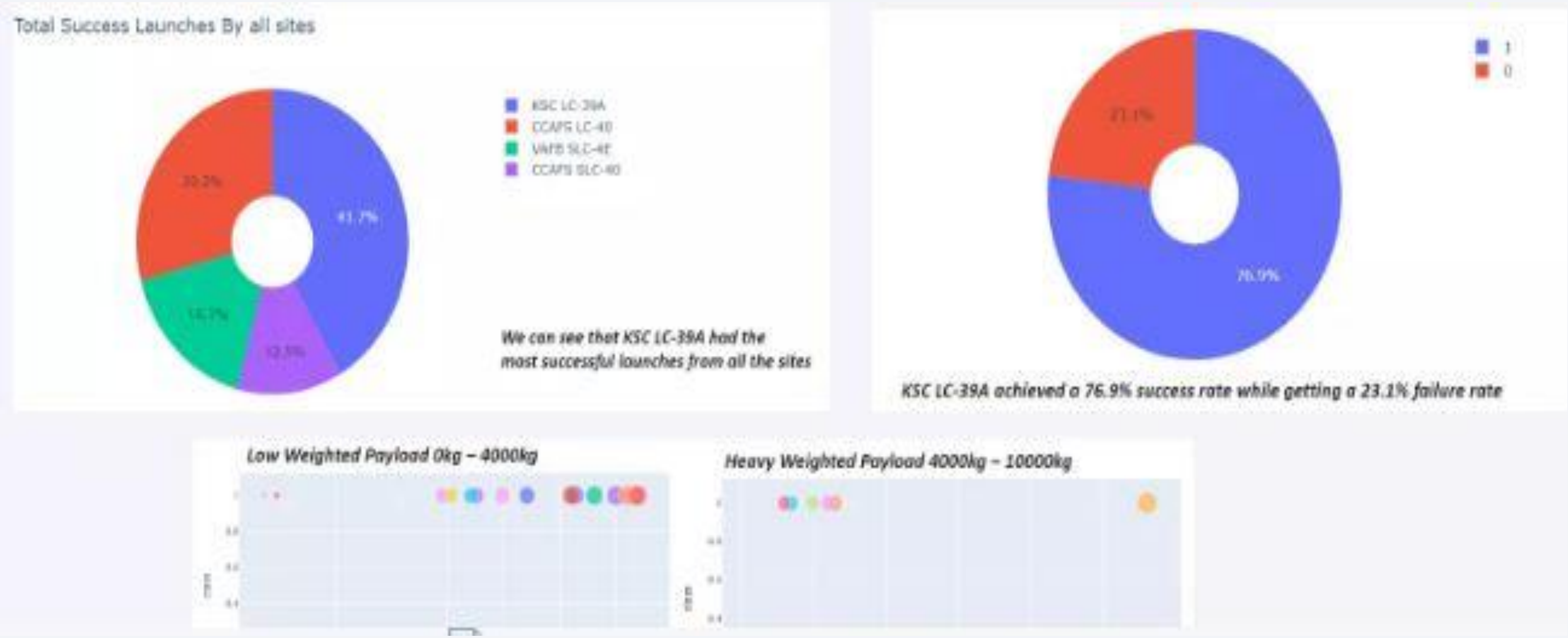
Build an Interactive Map with Folium



Map markers help to find an optimal location

https://github.com/KofiRomio/DataScienceEcosystem/blob/master/10-IBM%20DS%20Capstone-lab4-EDA%20viz.ipynb

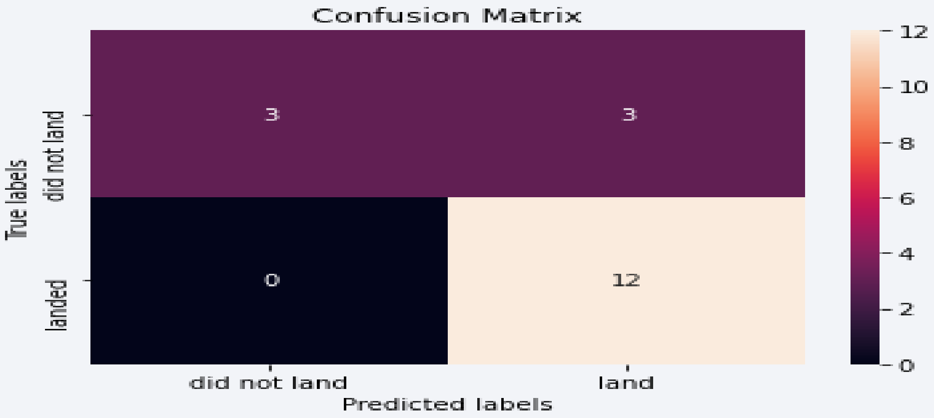
Build a Dashboard with Plotly Dash

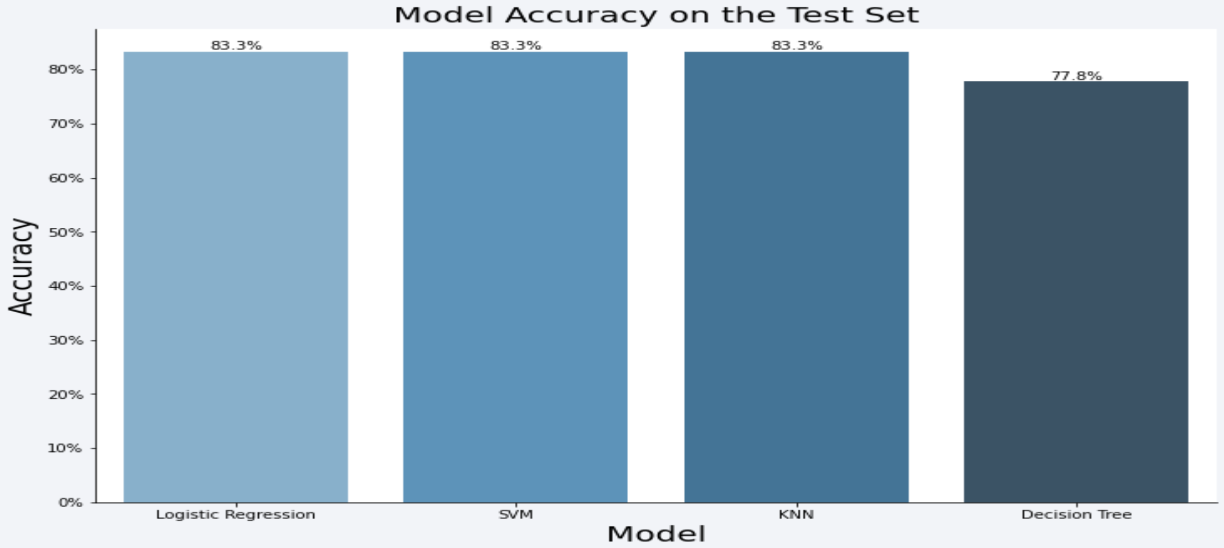


https://github.com/KofiRomio/DataScienceEcosystem/blob/master/dash_interact2_HY%20(1).py

Predictive Analysis (Classification)

- The SVM , KNN, and Logistic Regression model achieved the highest accuracy at 83.3%, while the SVM performs the best in terms of Area under the curve at 0.958





https://github.com/KofiRomio/DataScienceEcosystem/blob/master/10-IBM%20DS%20Capstone-lab4-EDA%20viz.ipynb

Results

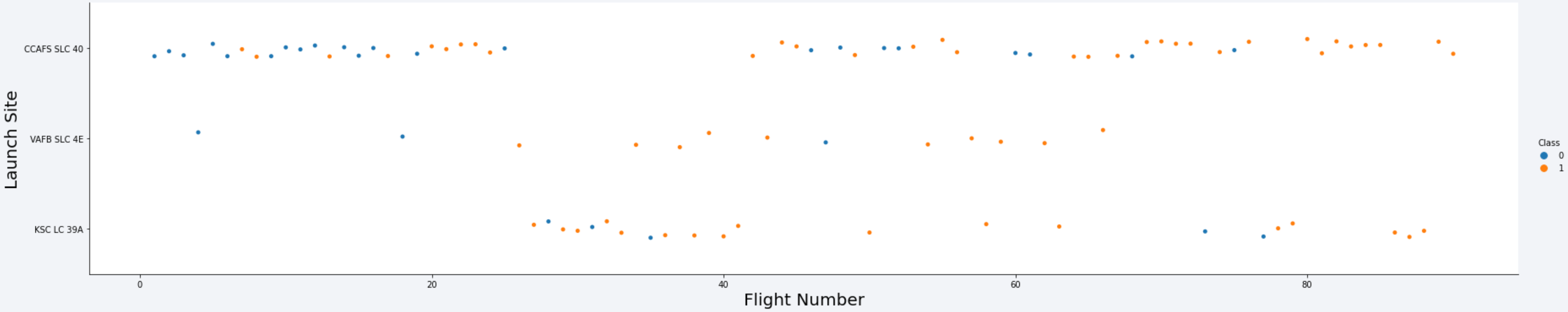
- The SVM, KNN, and Logistic Regression models are the best in terms of prediction accuracy for this dataset.
- Low weighted payloads perform better than the heavier payloads.
- The success rates for SpaceX launches is directly proportional to time in years they will eventually perfect the launches.
- KSC, LC, 39A had the most successful launches from all the sites.
- Orbit GEO, HEO,SSO, ES L1 has the best success rate.

Section 2

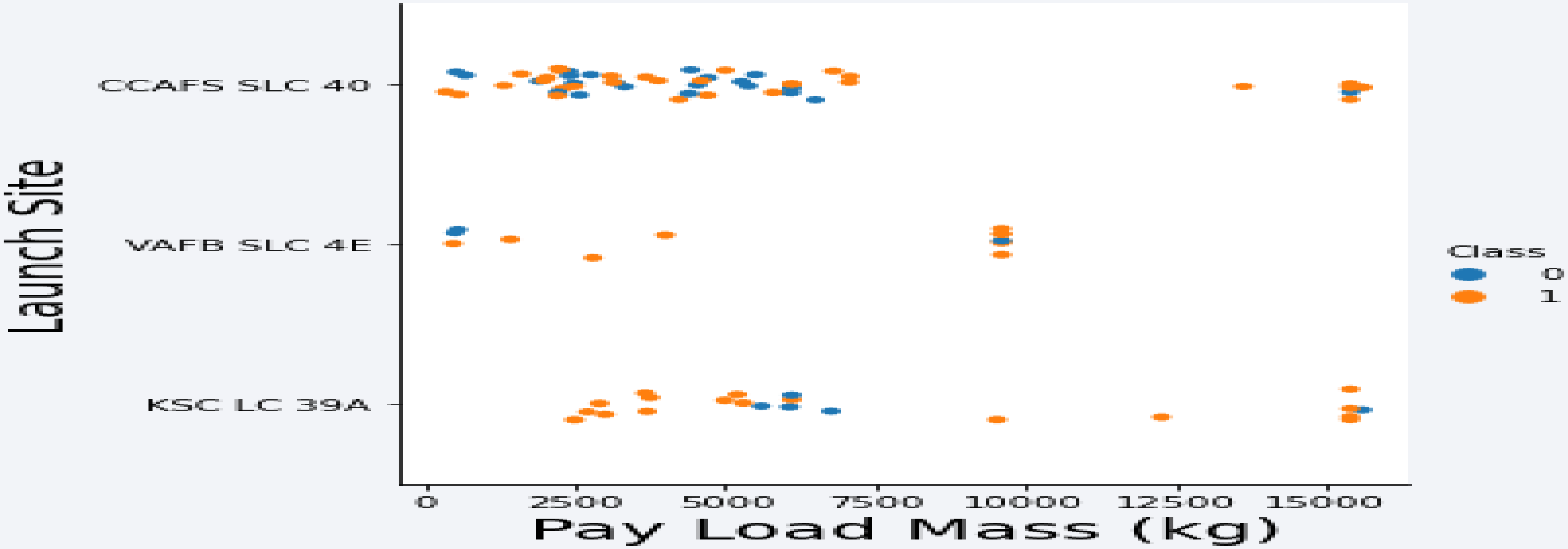
Insights drawn from EDA

Flight Number vs. Launch Site

- Launches from the site of CCAFS SLC 40 are significantly higher than launches from other sites

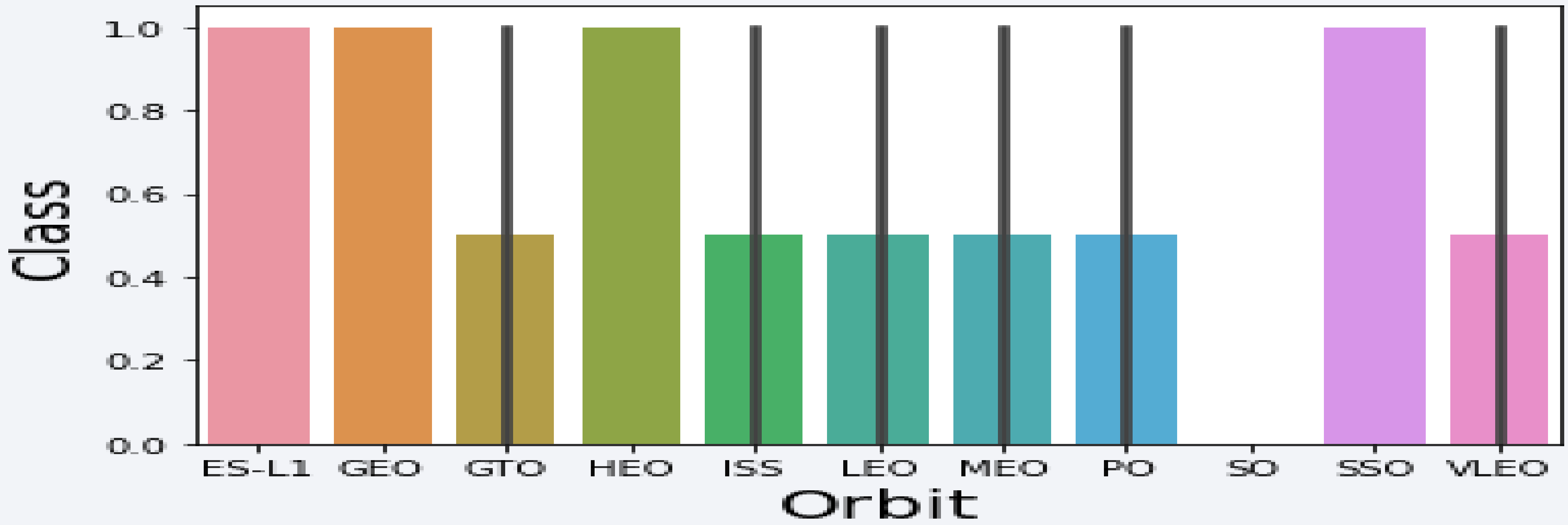


Payload vs. Launch Site



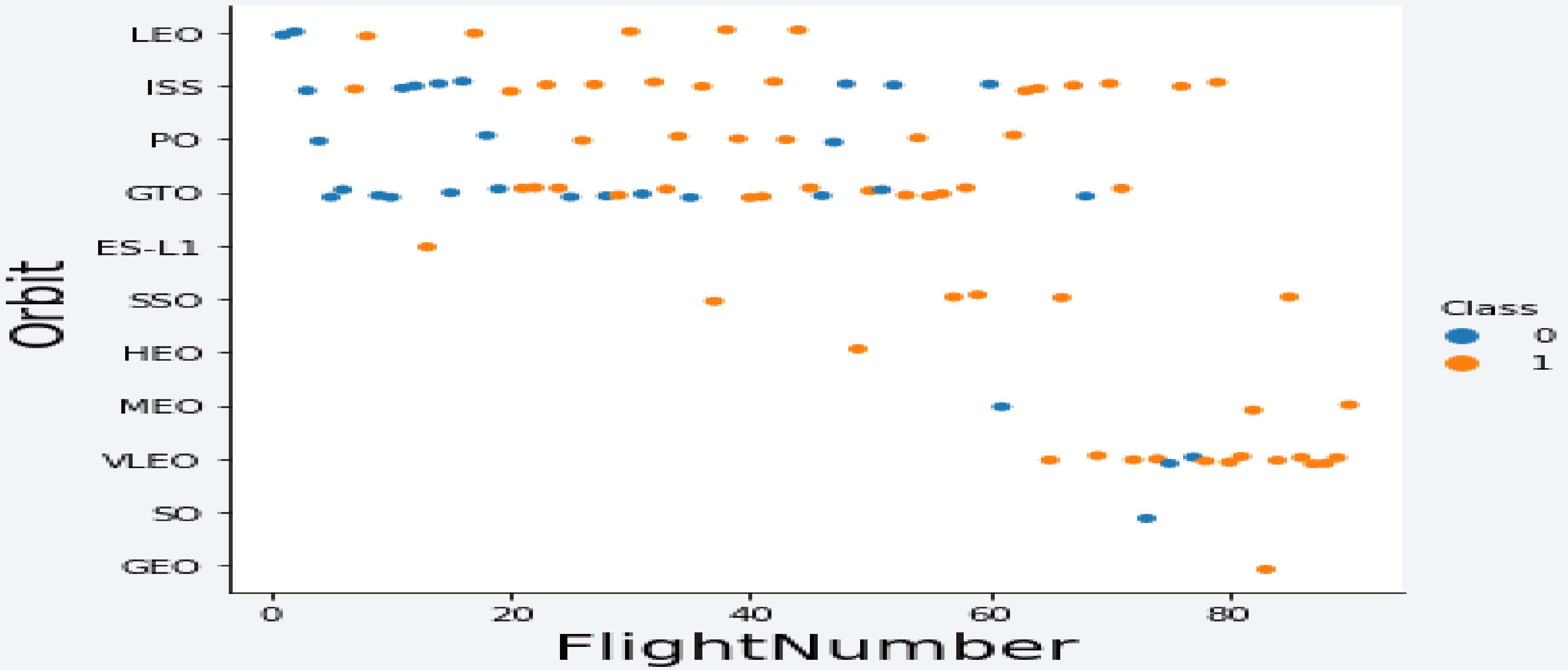
Show the screenshot of the scatter plot with explanations

Success Rate vs. Orbit Type



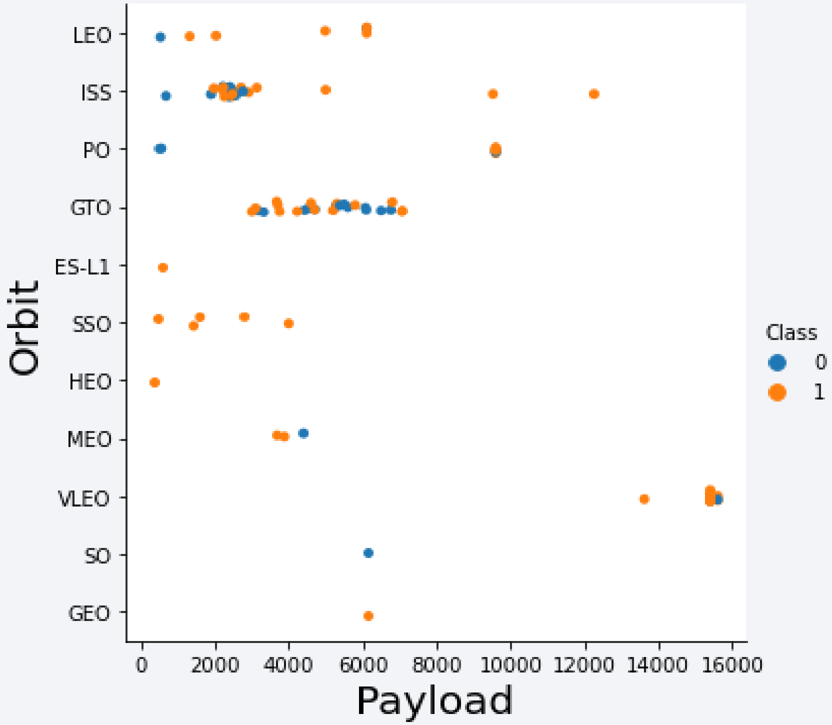
Orbit types of ES L1, HEO, GEO, SSO are among the highest success rate

Flight Number vs. Orbit Type



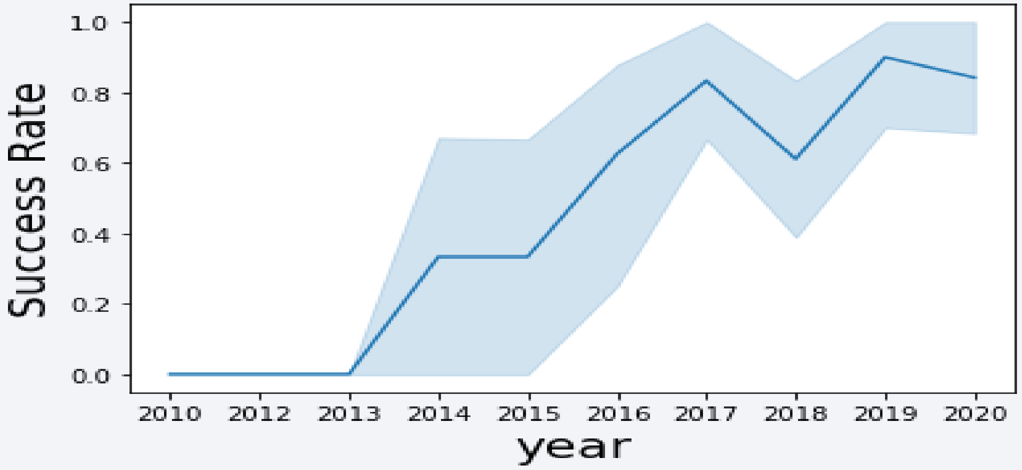
A trend can be observed of shifting to VLEO launches in recent years

Payload vs. Orbit Type



There are strong correlation between ISS and Payload at the range around 2000, and between GTO and the range of 4000-8000

Launch Success Yearly Trend



The successful launch rate has increased significantly since 2013 and has been stable since 2019 due to advances in technology and lessons learned

All Launch Site Names

- %sql select distinct (LAUNCH_SITE) from SPACEXTBL

launch_site
CCAFS LC-40
CCAFS SLC-40
CCAFSSLC-40
KSC LC-39A
VAFB SLC-4E

Launch Site Names Begin with 'CCA'

- %sql select * from SPACEXTBL where LAUNCH_SITE like ‘CCA%’ limit 5

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

- %sql select sum(PAYLOAD_MASS_KG_) from SPACEXTBL where CUSTOMER = ‘NASA (CRS)’

45596

Average Payload Mass by F9 v1.1

- %sql select avg(PAYLOAD_MASS_KG_) from SPACEXTBL where BOOSTER_VERSION = ‘F9 v1.1’

2928.4

First Successful Ground Landing Date

- %sql select min (DATE) from SPACEXTBL where Landing_Outcome = ‘Success(ground pad)’

2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- %sql select BOOSTER_VERSION from SPACEXTBL where Landing_Outcome = ‘Success(drone ship)’ & PAYLOAD_MASS_KG_ > 4000 & PAYLOAD_MASS_KG_ < 6000

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

Total Number of Successful and Failure Mission Outcomes

- %sql select count(MISSION_OUTCOME) from SPACEXTBL where MISSION_OUTCOME = ‘Success’ / MISSION_OUTCOME = ‘Failure (in flight)’

100

Boosters Carried Maximum Payload

- %sql select BOOSTER_VERSION from SPACEXTBL where PAYLOAD_MASS_KG_ = (select max(PAYLOAD_MASS_KG_) from SPACEXTBL)

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

2015 Launch Records

- %sql select * from SPACEXTBL where Landing_Outcome like ‘Success%’ & (DATE between 2015-01-01 & 2015-12-31) order by DATE desc

time_utc_	booster_version	launch_site	payload	payload_mass_kg_	orbit	customer	mission_outcome	landing_outcome
14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
17:54:00	F9 FT B1029.1	VAFB SLC-4E	Iridium NEXT 1	9600	Polar LEO	Iridium Communications	Success	Success (drone ship)
05:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
04:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
21:39:00	F9 FT B1023.1	CCAFS LC-40	Thaicom 8	3100	GTO	Thaicom	Success	Success (drone ship)
[illegible]	[illegible]	CCAFS LC-	[illegible]	[illegible]	[illegible]	SKY Perfect JSAT	[illegible]	[illegible]

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

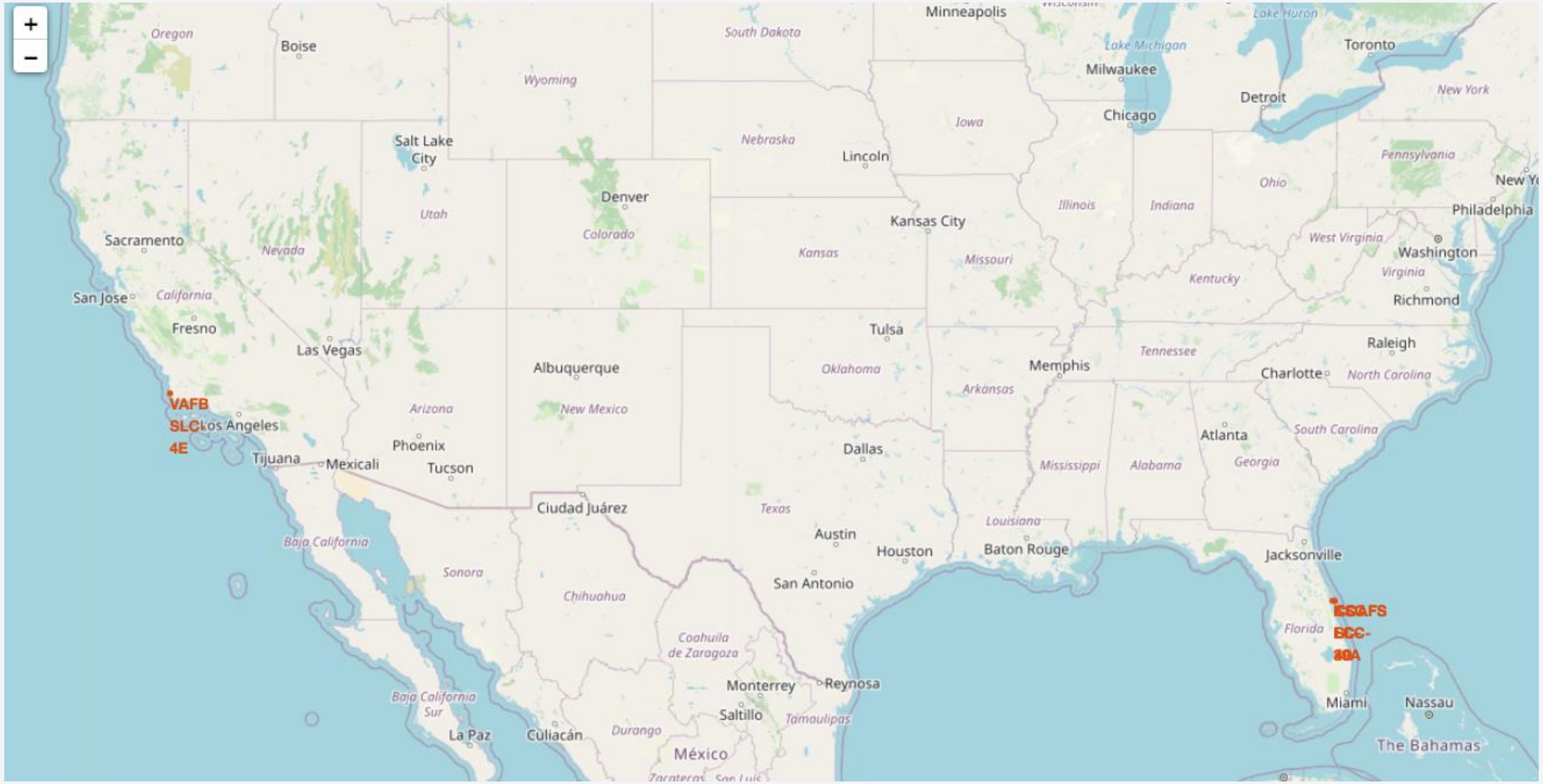
- %sql select * from SPACEXTBL where Landing_Outcome like ‘Success%’ & (DATE between 2010-06-04 & 2017-03-20) order by DATE desc

2016-05-27	21:39:00	F9 FT B1023.1	CCAFS LC-40	Thaicom 8	3100	GTO	Thaicom	Success	Success (drone ship)
2016-05-06	05:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2016-04-08	20:43:00	F9 FT B1021.1	CCAFS LC-40	SpaceX CRS-8	3136	LEO (ISS)	NASA (CRS)	Success	Success (drone ship)
2015-12-22	01:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)

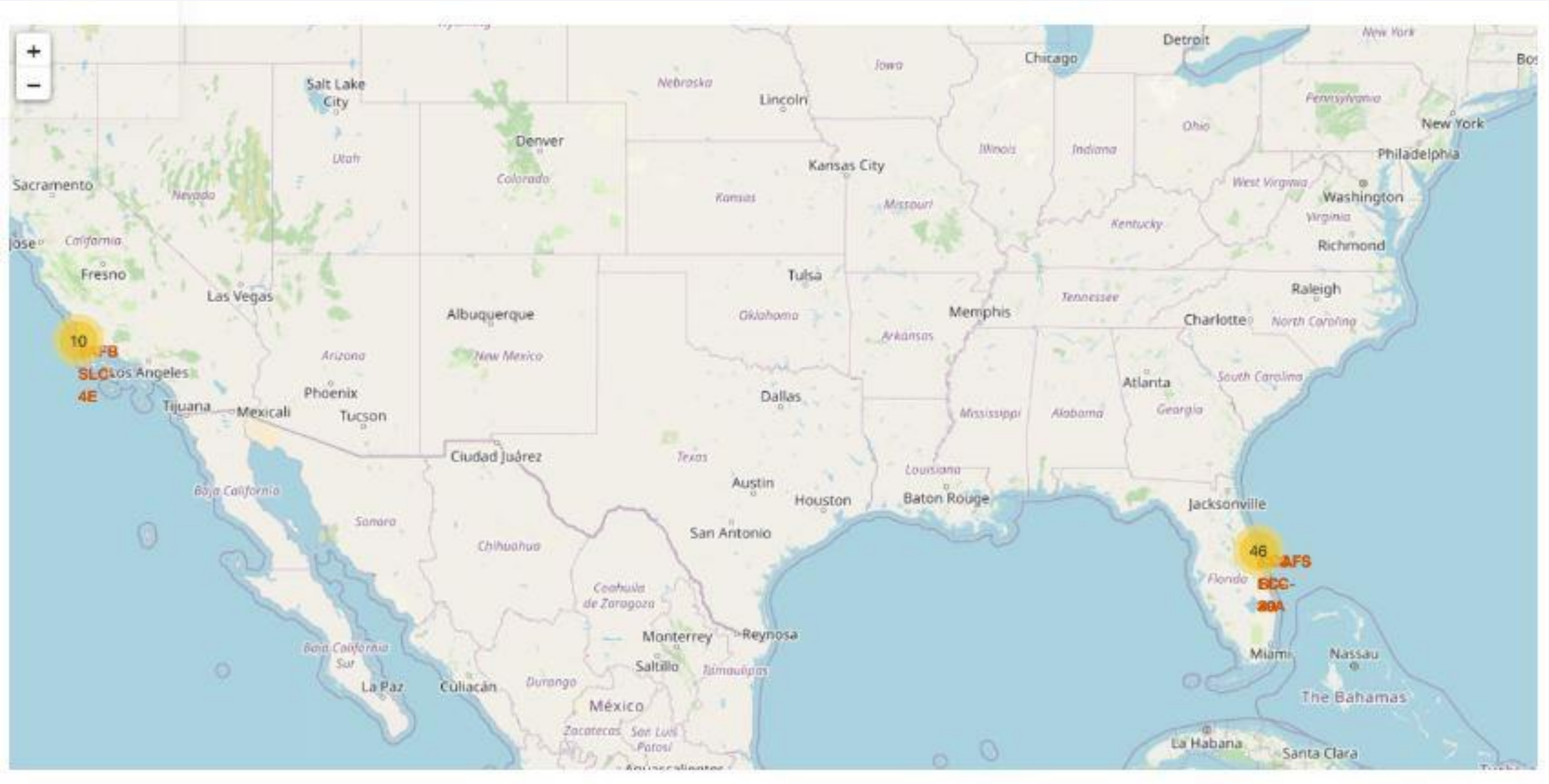
Section 3

Launch Sites Proximities Analysis

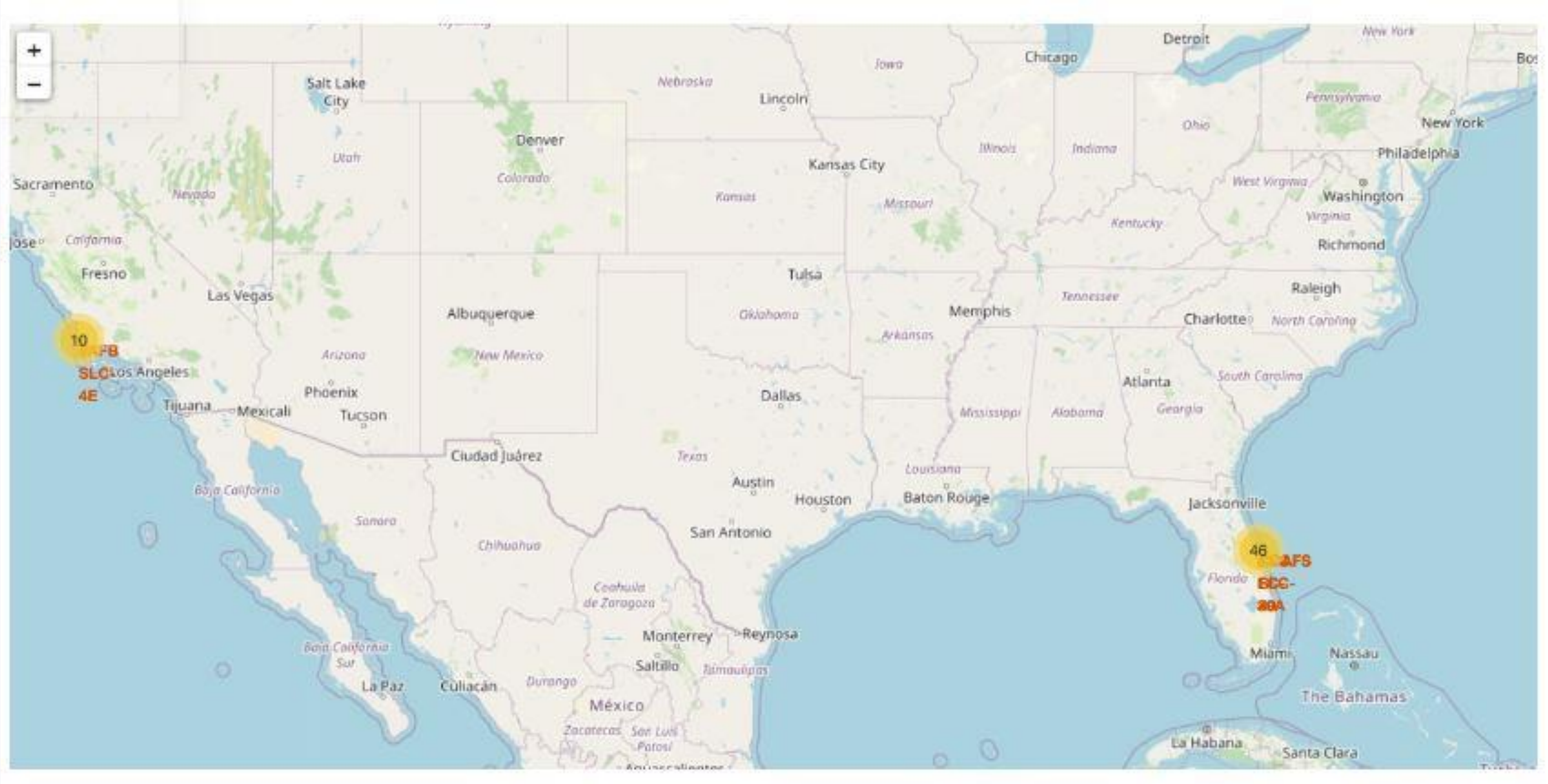
All launch sites marked on map



Success/Failed launches marked on the map



Distance between a launch site to its proximities



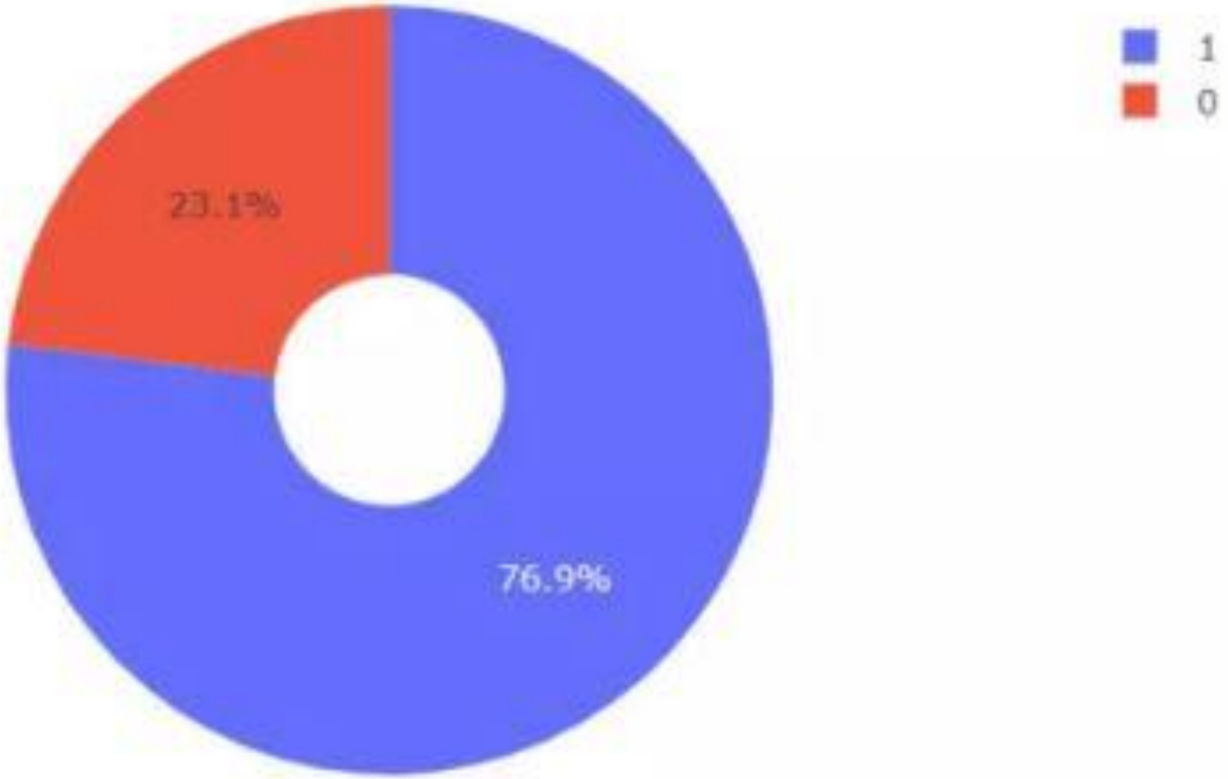
Section 4

Build a Dashboard with Plotly Dash

Total success launches by all sites

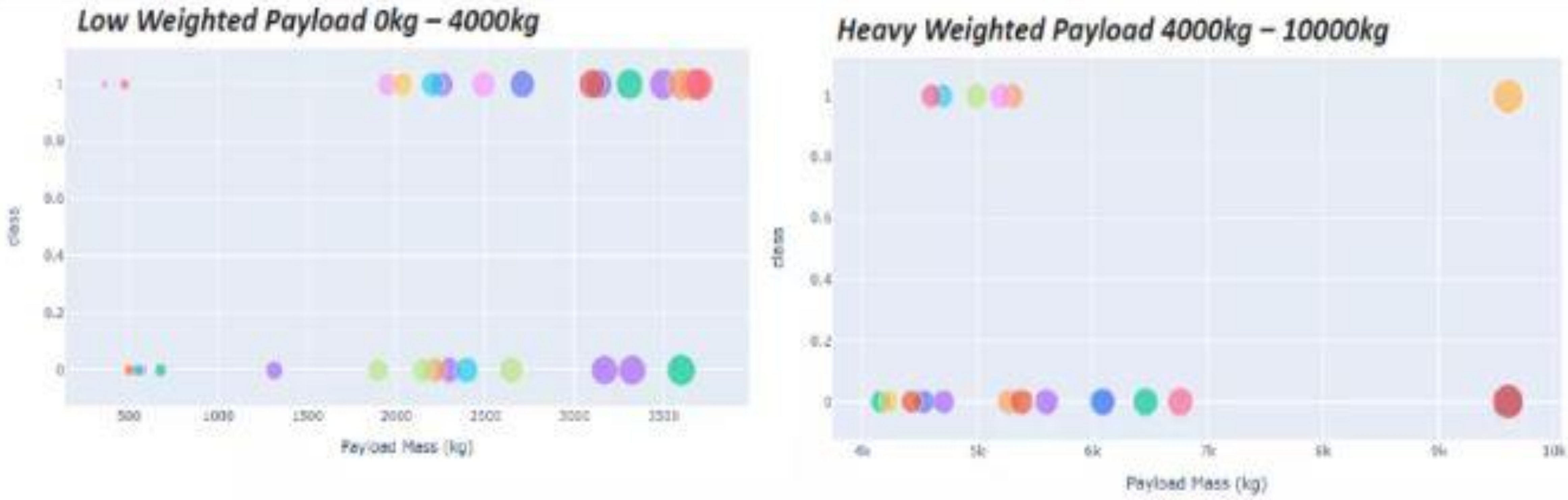
We can see that KSC LC-39A had the most successful launches from all the sites

Success rate by site



KSC LC-39A achieved a 76.9% success rate while getting a 23.1% failure rate

Payload vs launch outcome

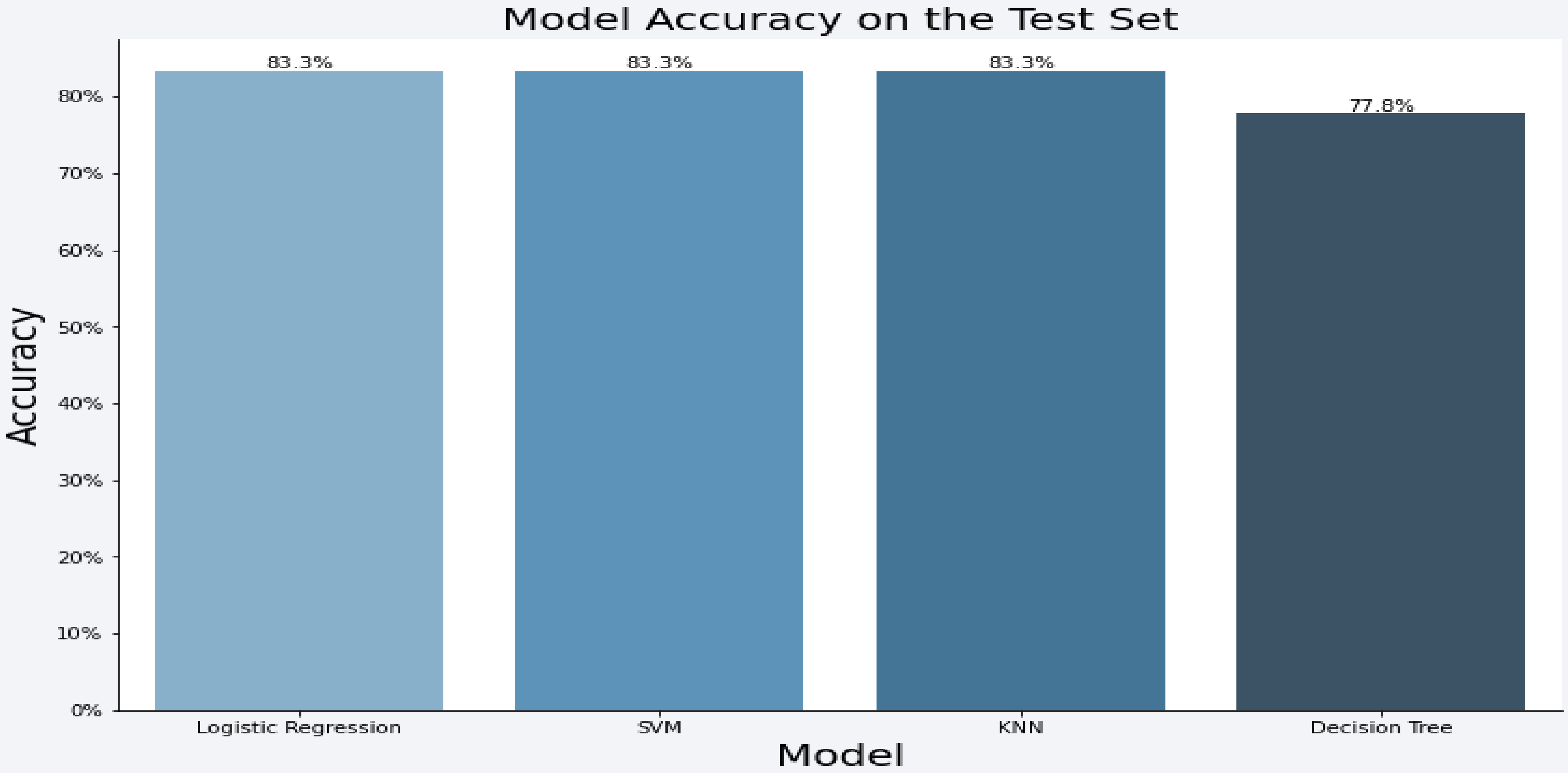


We can see the success rates for low weighted payloads is higher than the heavy weighted payloads

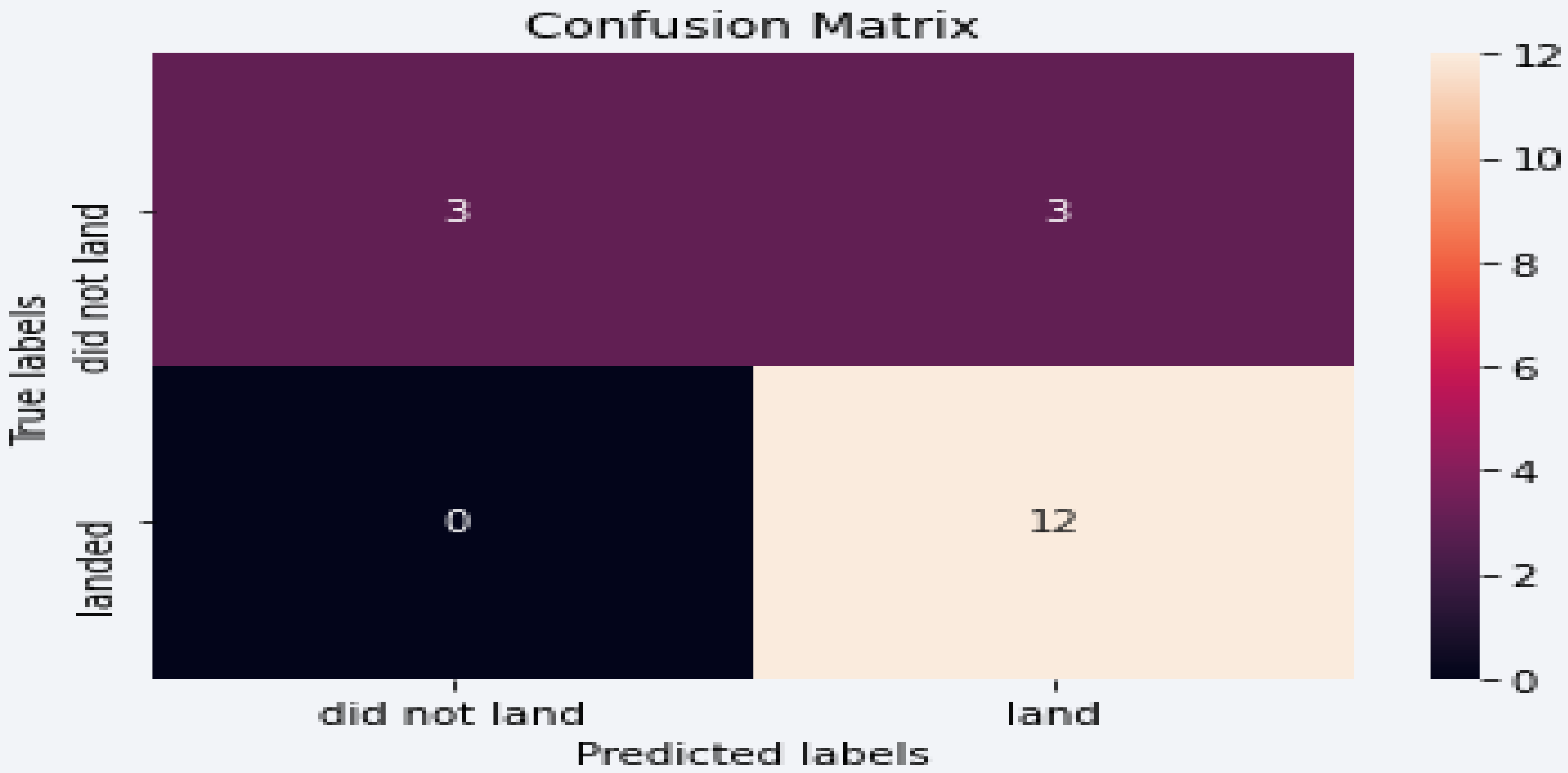
Section 5

Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix



Conclusions

- The SVM,KNN, and Logistic Regression models are the best in terms of prediction accuracy for this dataset.
- Low weighted payloads perform better than the heavier payloads.
- The success rates for SpaceX launches is directly proportional to time in years they will eventually perfect the launches.
- KSC, LC 39A had the most successful launches from all the sites.
- Orbit GEO, HEO, SSO, ES L1 has the best success rate.

Thank you!