IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

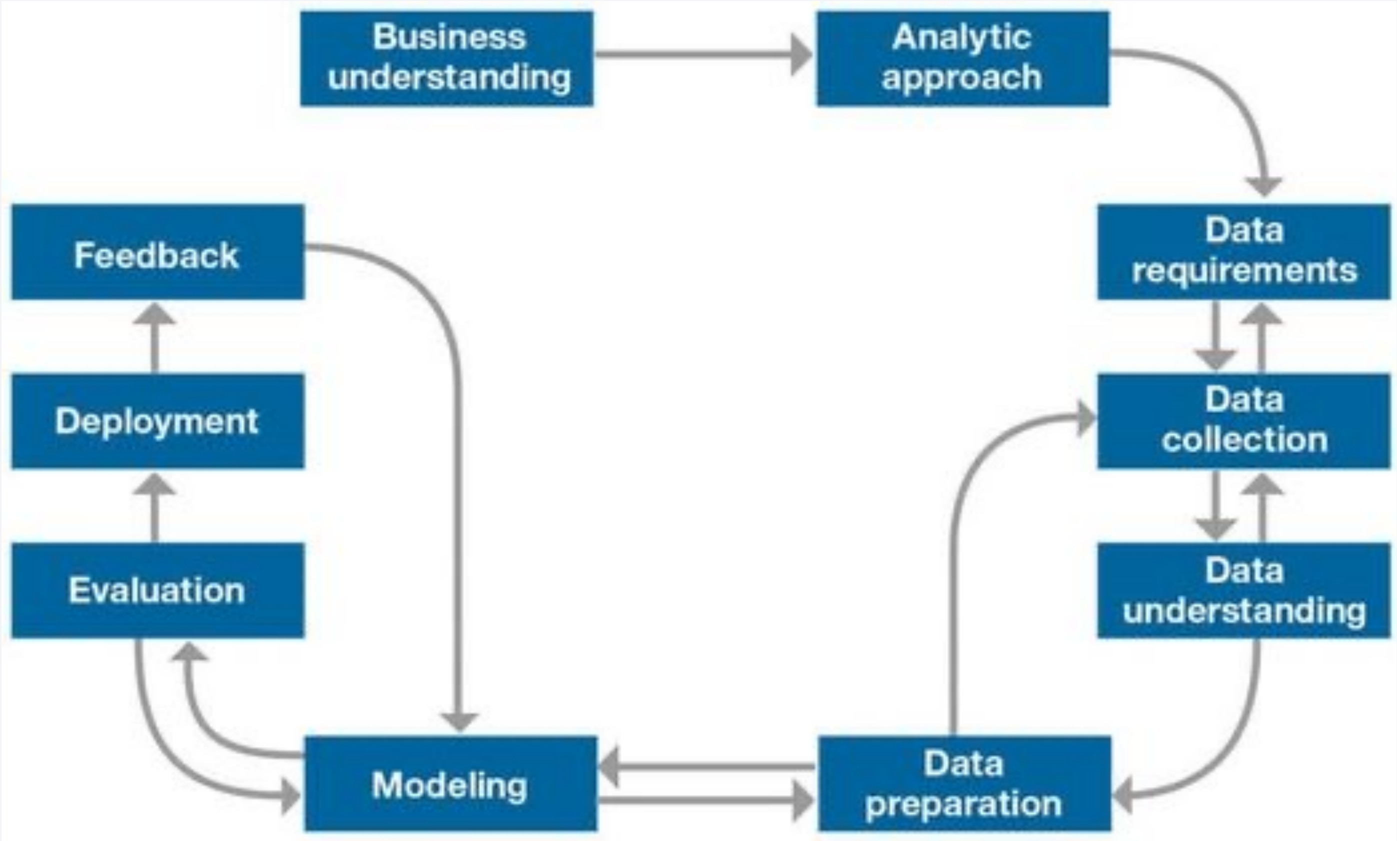
Thang Nguyen Chien
13rd Jul, 2022

Outline

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

Executive Summary

- Summary of methodologies



Executive Summary

- Summary of methodologies
    - Data Collection
    - Data Wrangling
    - EDA with SQL
    - EDA with Visualization
    - Build interactive map with Folium
    - Build a Dashboard with Plotty Dash
    - Machine Learing Prediction

Executive Summary

- Summary of all results
  - Data is collected from API and Web Scraping
  - Data is wrangled and cleaning
  - Model is built to predict Outcome with multiple algorithm.
  - Results can be view in : EDA result, Interactive analysis and Predictive analysis

Introduction

- SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upwards of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage; Therefore, if we can determine if the first stage will land, we can determine the cost of a launch.

- Instead of using rocket science to determine if the first stage will land successfully, I will train a machine learning model and use public information to predict if SpaceX will reuse the first stage.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data about Space X launches were collected via API and Web Scraping.
- Perform data wrangling
  - Data is filled nan value and do some value counts to do preliminary examination.
  - Data is one hot encoding on some fields.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash

Methodology

Executive Summary

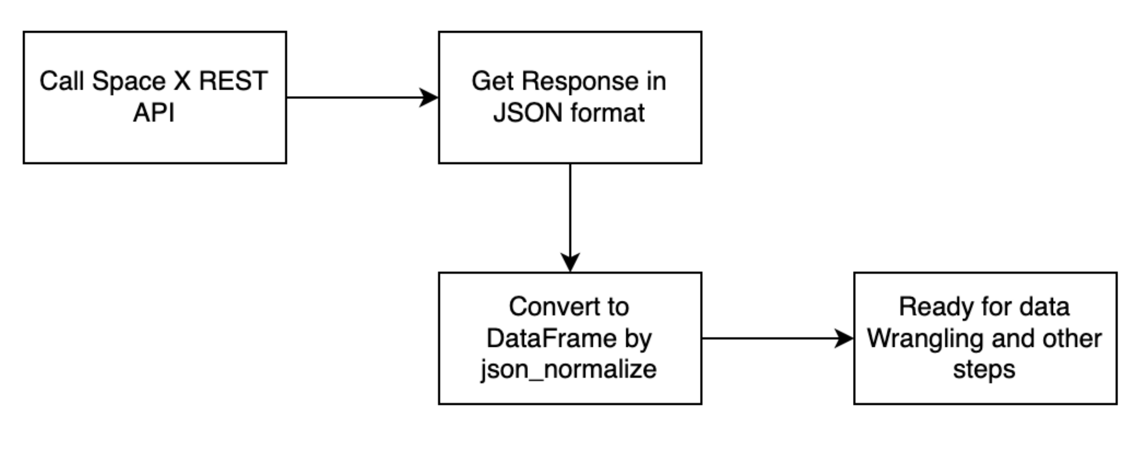
- Perform predictive analysis using classification models
    - I build classification models from multiple algorithm, such as: SVM, Decision Tree, KNN and Logistic Regression. I use GridSearch to find the best parameters for each model.
    - I evaluate on test set to find the best model to use.

Data Collection

- Dataset is collected via API and Web Scraping.
- You need to present your data collection process use key phrases and flowcharts

Data Collection – SpaceX API

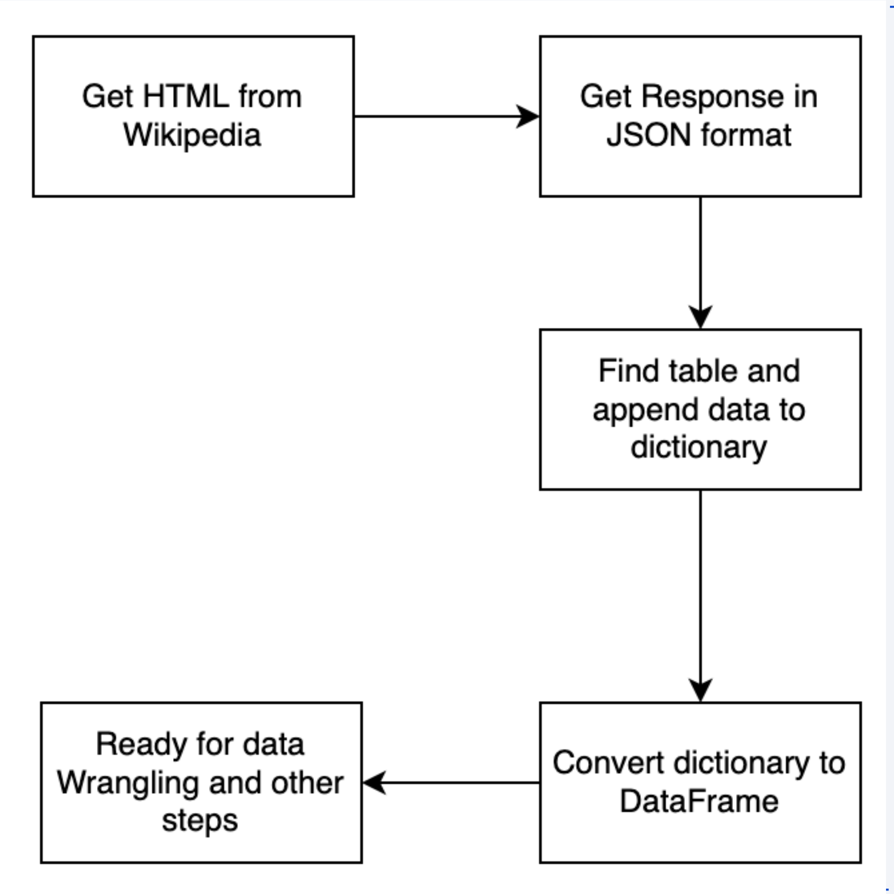
- Data is getting from SpaceX REST API
- API will give information about launches, rocket used, payload delivered, launch spec, landing spec and landing outcome.
- API is located at api.spacexdata.com/v4/.



Gihub link: https://github.com/thangnch/MiAI_Study/blob/main/01.%20Data%20Collect%20via%20API.ipynb

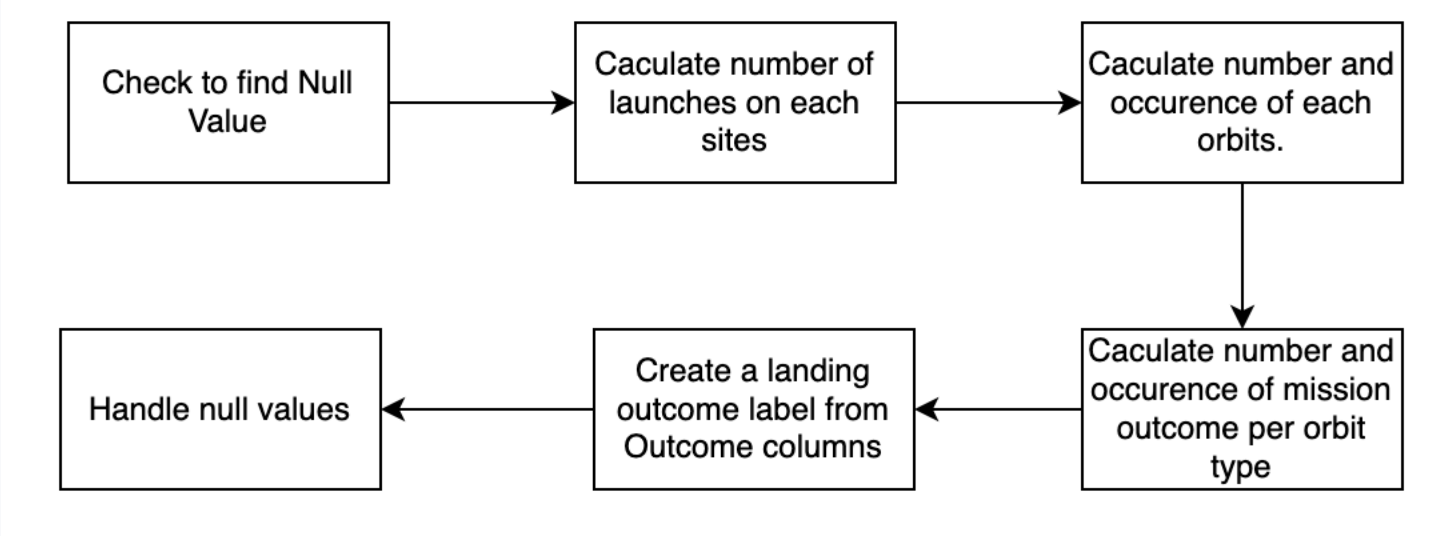
Data Collection - Scraping

- Data is getting from Web Scraping at Wikipedia using BeatifulSoup.
- Table is found on response and append data to dictionary.
- Convert dictionary to Dataframe



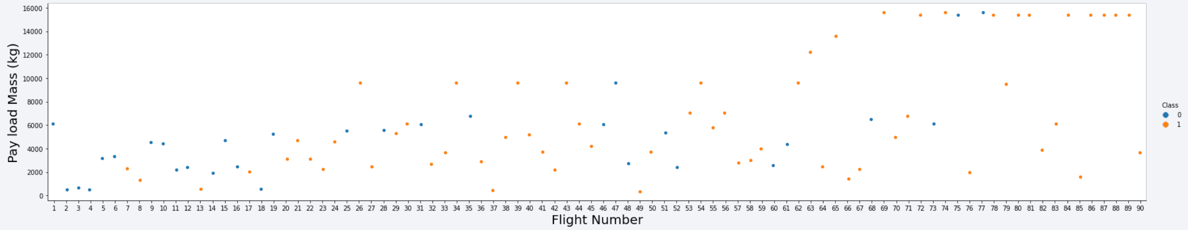
Github link: https://github.com/thangnch/MiAI_Study/blob/main/02.%20Data%20Collect%20via%20Web%20Scraping.ipynb

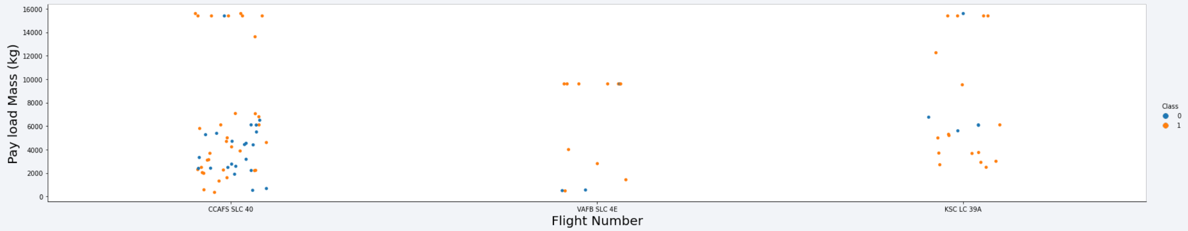
Data Wrangling

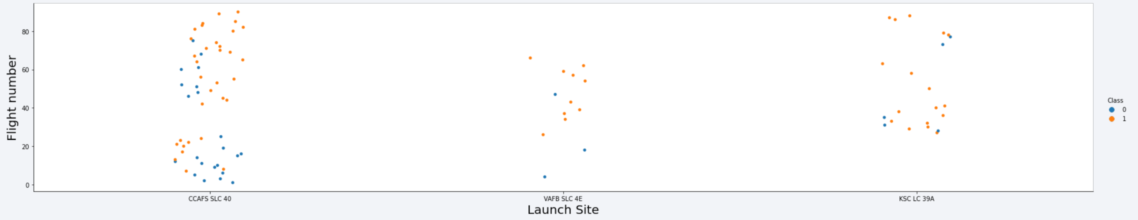


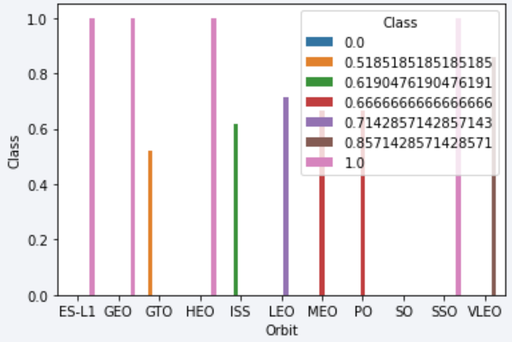
Github link: https://github.com/thangnch/MiAI_Study/blob/main/03.%20Data%20Wrangling.ipynb

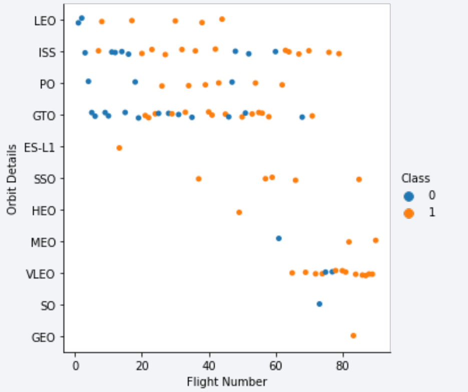
EDA with Data Visualization

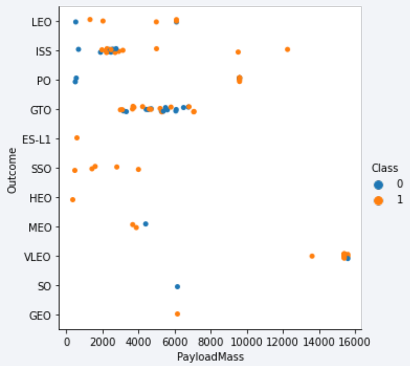


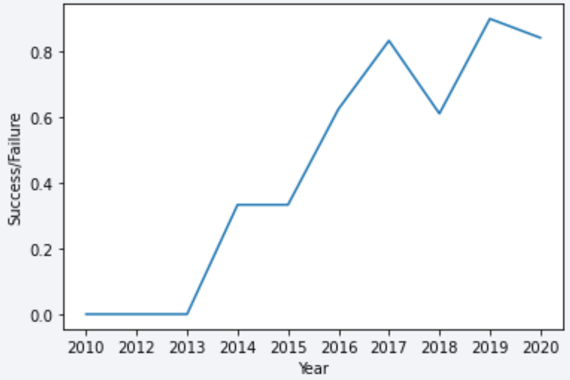












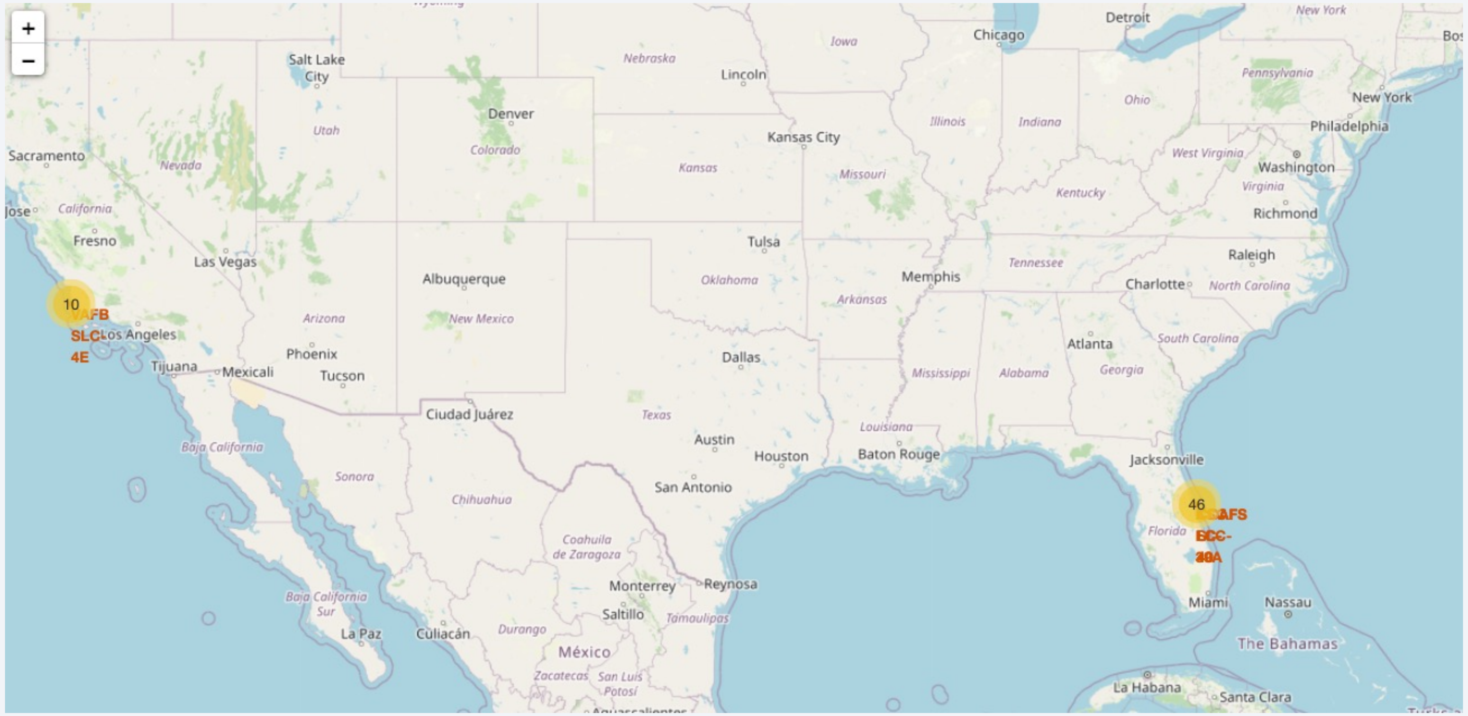
Github link: https://github.com/thangnch/MiAI_Study/blob/main/05.%20EDA%20with%20Visualization.ipynb

EDA with SQL

- Displaying the names of the unique launch sites in the space mission
- Displaying 5 records where launch sites begin with the string 'KSC'
- Displaying the total payload mass carried by boosters launched by NASA (CRS)
- Displaying average payload mass carried by booster version F9 v1.1
- Listing the date where the successful landing outcome in drone ship was achieved.
- Listing the names of the boosters which have success in ground pad and have payload mass greater than 4000 but less than 6000
- Listing the total number of successful and failure mission outcomes
- Listing the names of the booster versions which have carried the maximum payload mass.
- Listing the records which will display the month names, successful landing outcomes in ground pad ,booster

Github link: https://github.com/thangnch/MiAI_Study/blob/main/04.%20EDA%20with%20SQL.ipynb

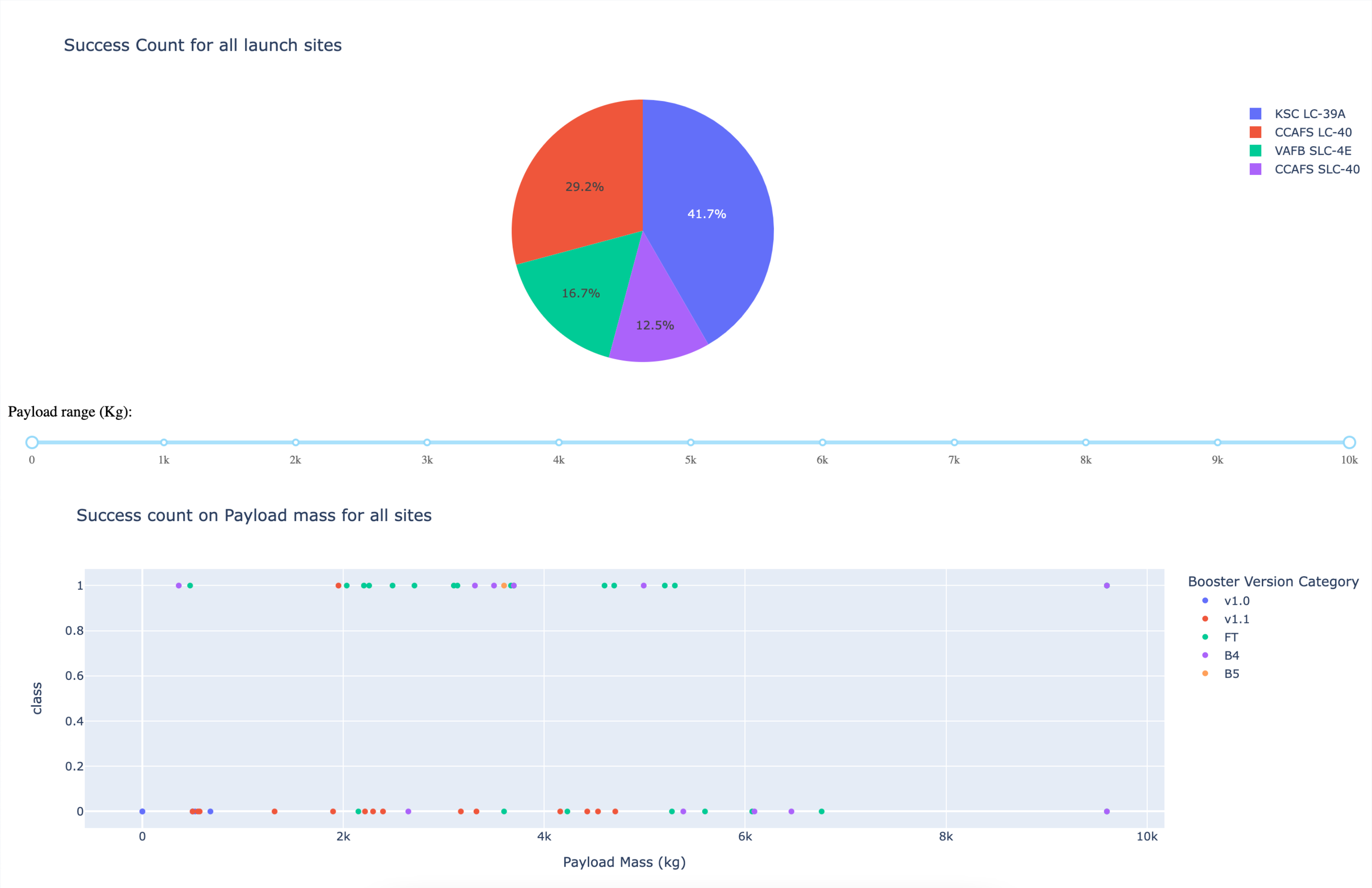
Build an Interactive Map with Folium



Markers are added to the map to find the best location for building a launch site.

Github link: https://github.com/thangnch/MiAI_Study/blob/main/06.%20Data%20Visualization%20with%20Folium.ipynb

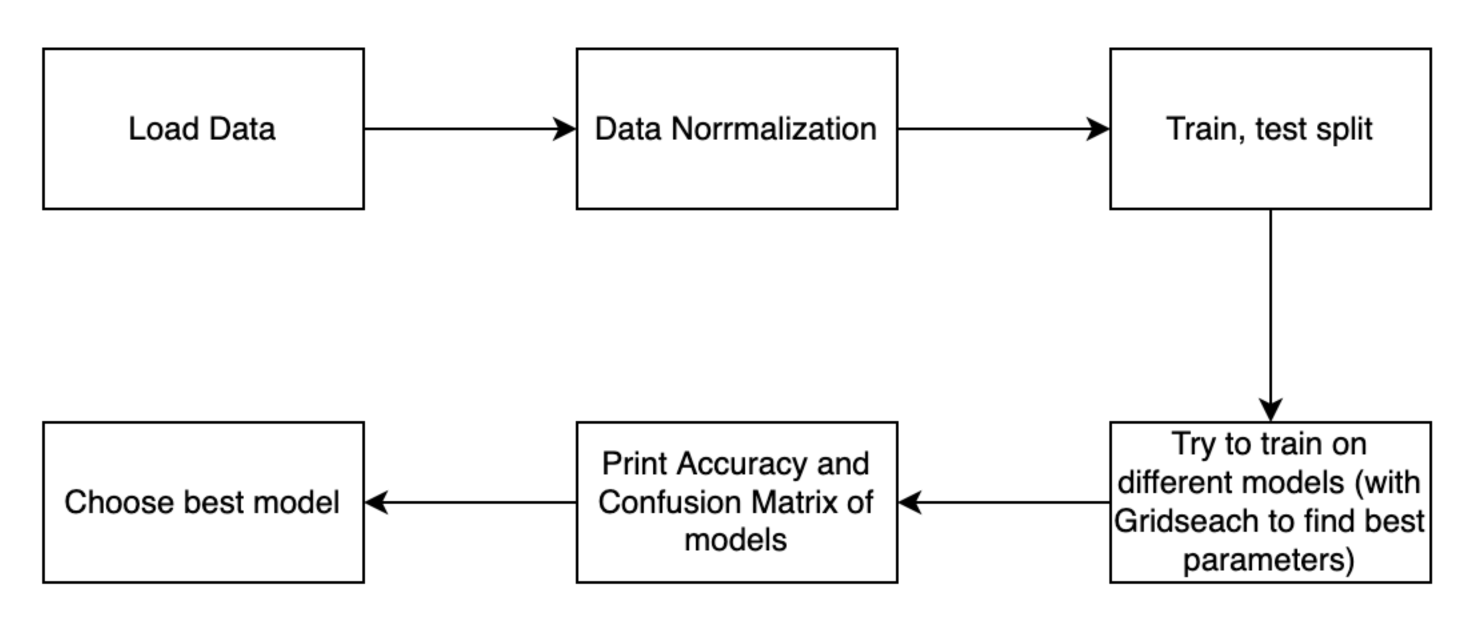
Build a Dashboard with Plotly Dash



Github link: https://github.com/thangnch/MiAI_Study/blob/main/06a.%20Building%20Ploty.py

Predictive Analysis (Classification)

- Models are built using multiple algorithm, such as: Logistic Regression, SVM, Decision Tree and KNN.
- Key findings: SVM, KNN and Logistic Regression achieved best accuracy (83%)



Github link: https://github.com/thangnch/MiAI_Study/blob/main/07.%20Machine%20Learning%20Prediction.ipynb

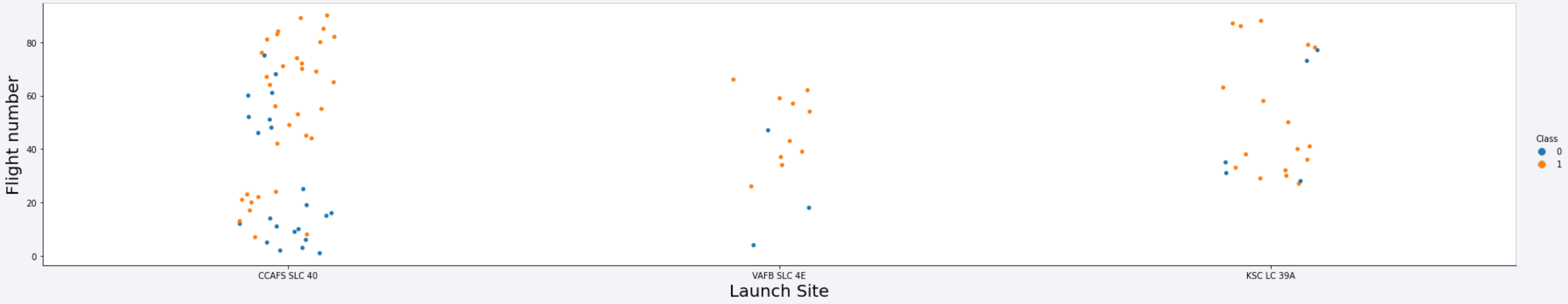
Results

- The SVM, KNN, and Logistic Regression models achieved best prediction accuracy for this dataset.
- Find out that low weighted payloads perform better than the heavier payloads.
- The success rates for SpaceX launches is directly proportional 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

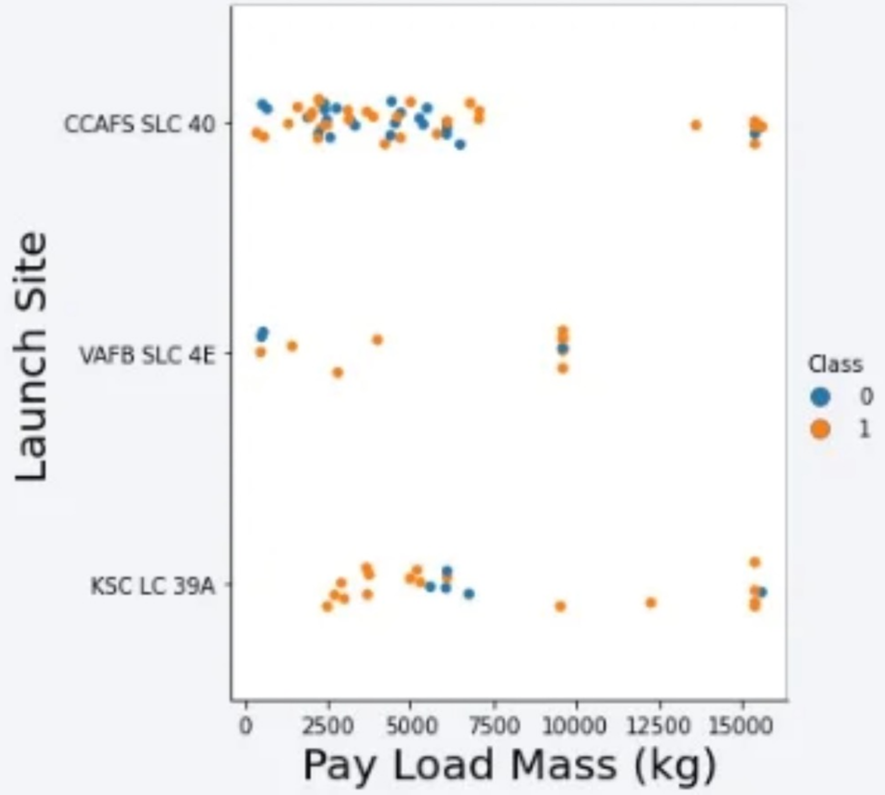
Insights drawn from EDA

Flight Number vs. Launch Site



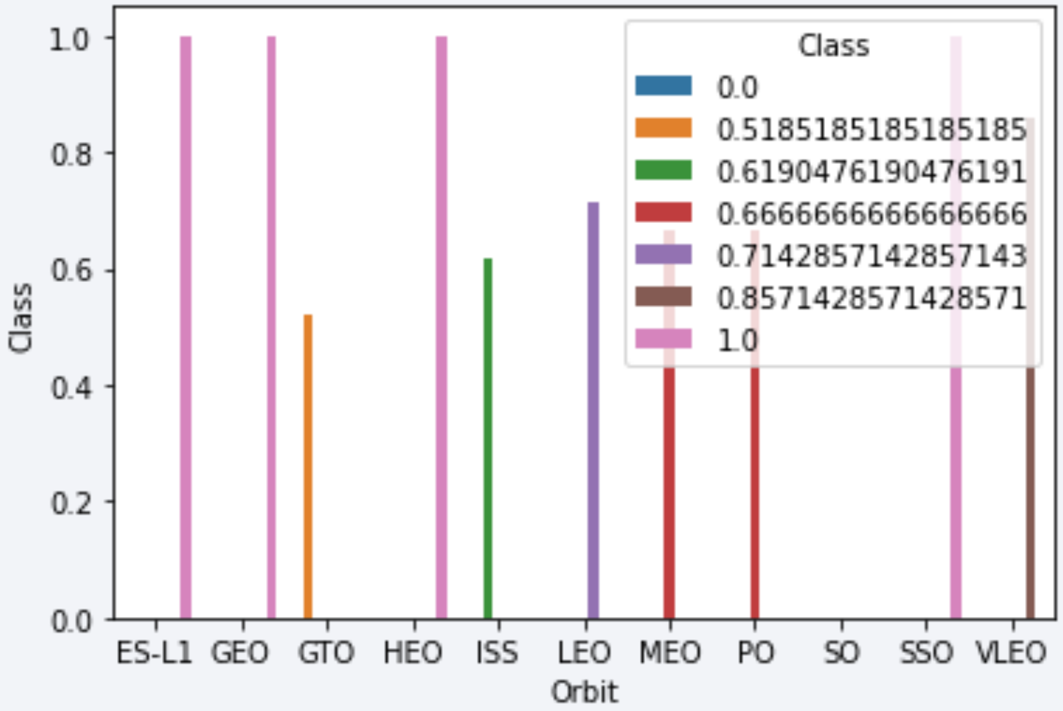
Number of lauch from CCAFS SLC40 is higher than others.

Payload vs. Launch Site



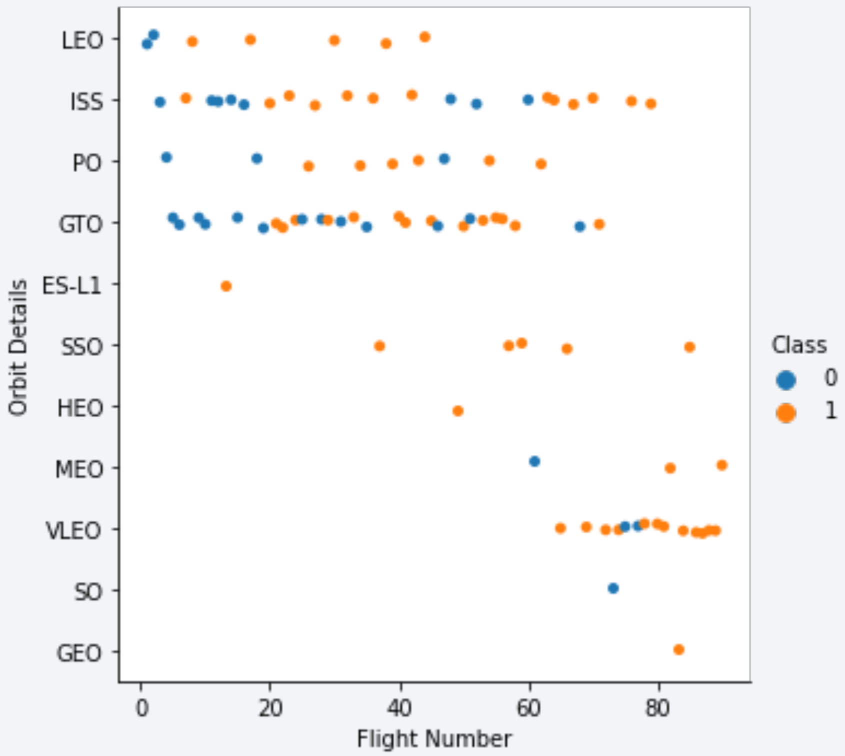
Lower payload mass have been launched from CCAFS SLC40

Success Rate vs. Orbit Type



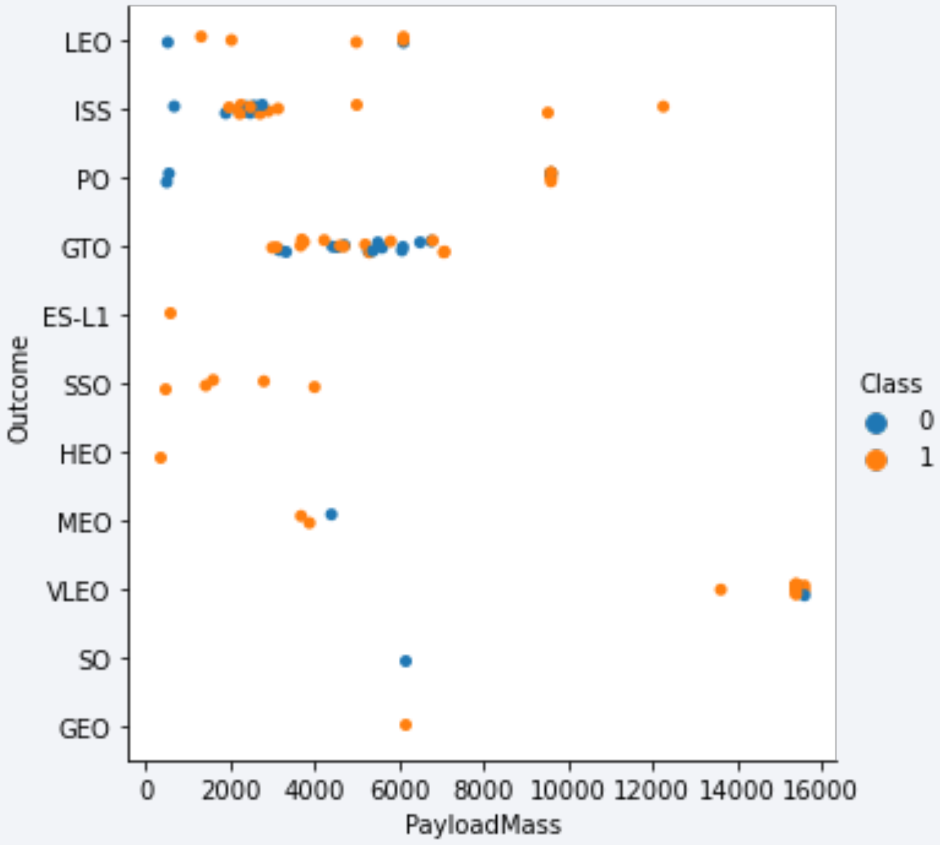
ES-L1, GEO, HEO and SSO have higher success rate.

Flight Number vs. Orbit Type



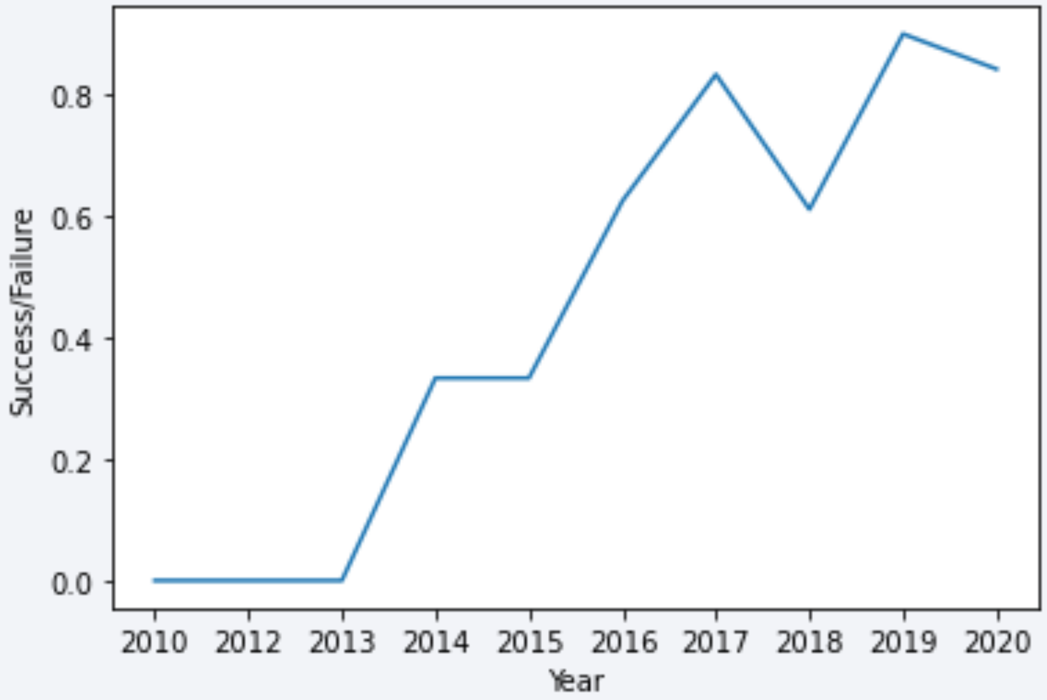
VLEO have high flight number than others

Payload vs. Orbit Type



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

Launch Success Yearly Trend



Launch success rate has increased significantly since 2013 and has stablised since 2019, potentially due to advance in technology and lessons learned.

All Launch Site Names

- %sql select distinct(LAUNCH_SITE) from SPACEXTBL

launch_site
CCAFS LC-40
CCAFS SLC-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
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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	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

- %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'

2928.400000

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)' and PAYLOAD_MASS__KG_ > 4000 and 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' or 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 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

- %sql select * from SPACEXTBL where Landing__Outcome like 'Success%' and (DATE between ‘2015-01-01' and ‘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

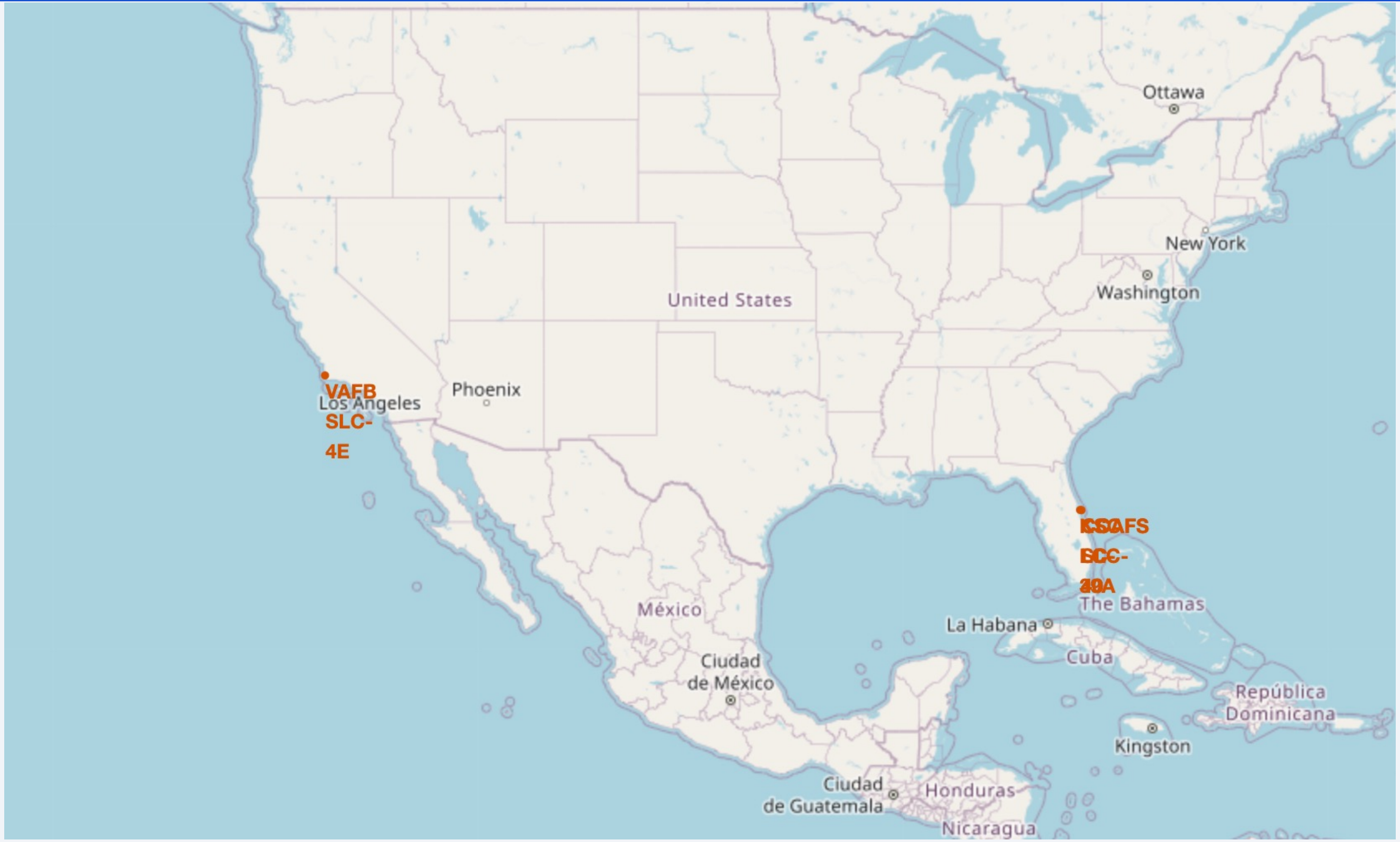
- %sql select * from SPACEXTBL where Landing__Outcome like 'Success%' and (DATE between '2010-06-04' and '2017-03-20') order by date desc

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

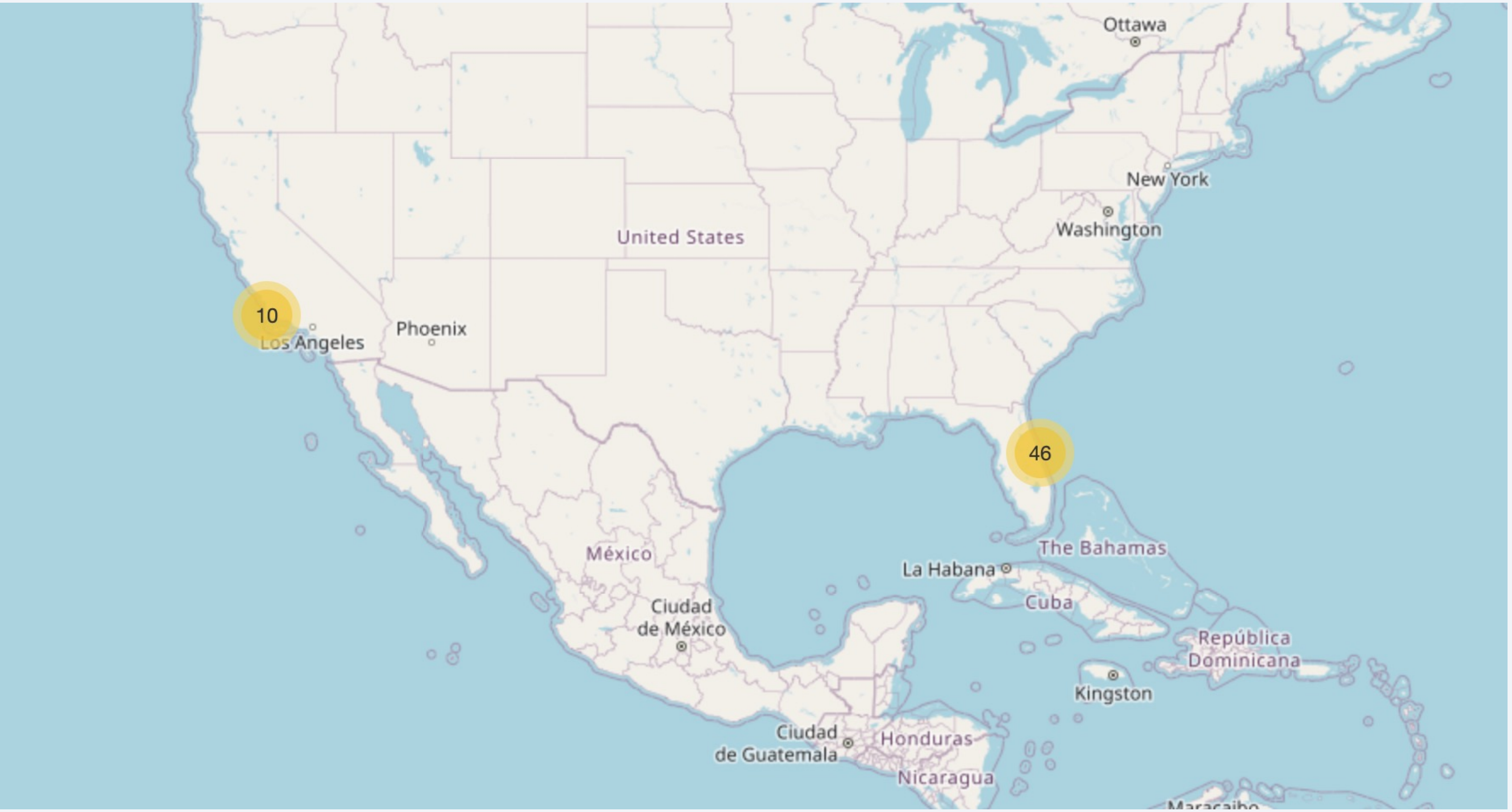
Launch Sites Proximities Analysis

All launch site on map



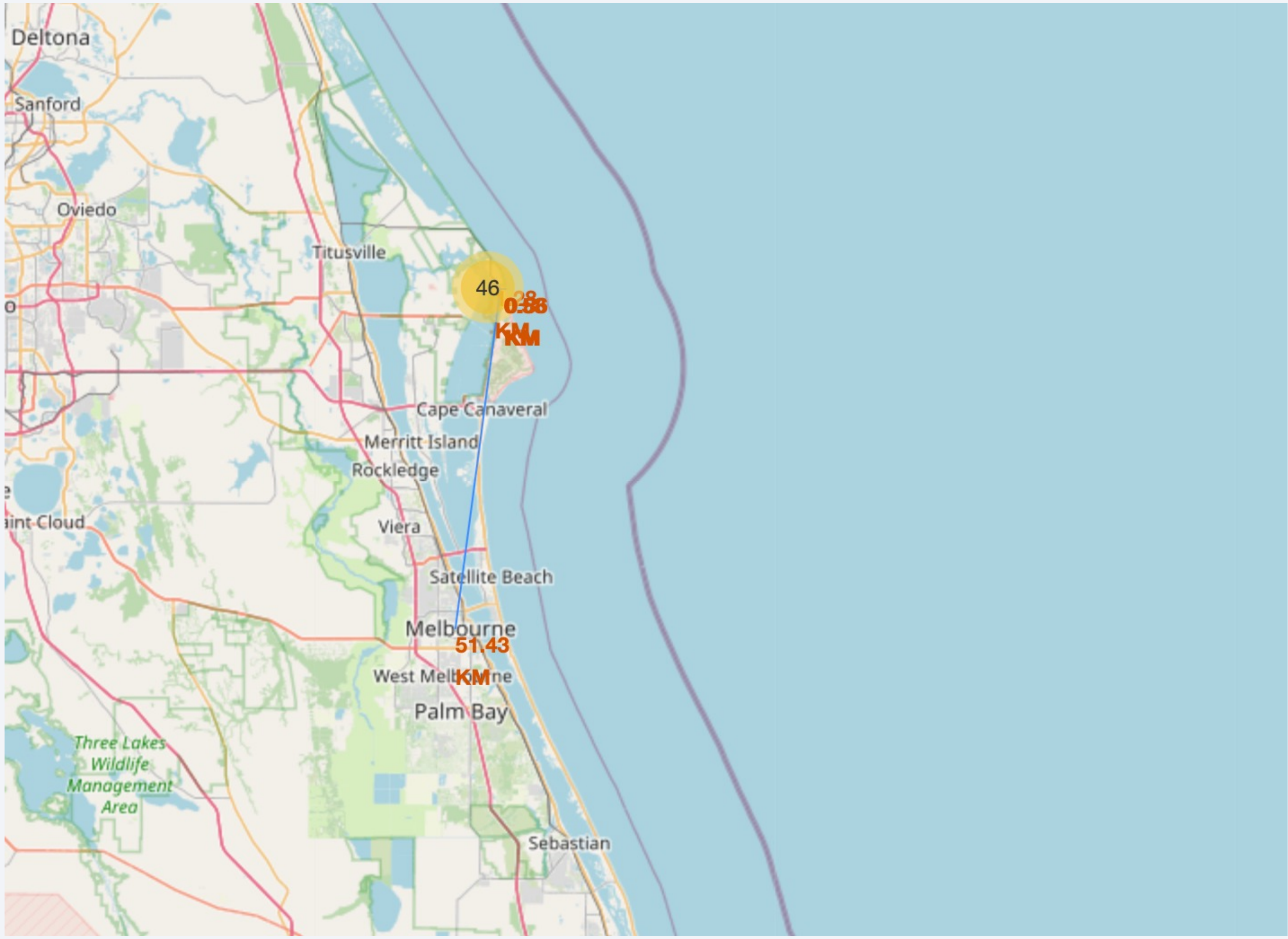
Add each site's location on a map using site's latitude and longitude coordinates.

Sucess/failed launches for each site on the map



Adding the launch outcomes for each site, and see which sites have high success rates

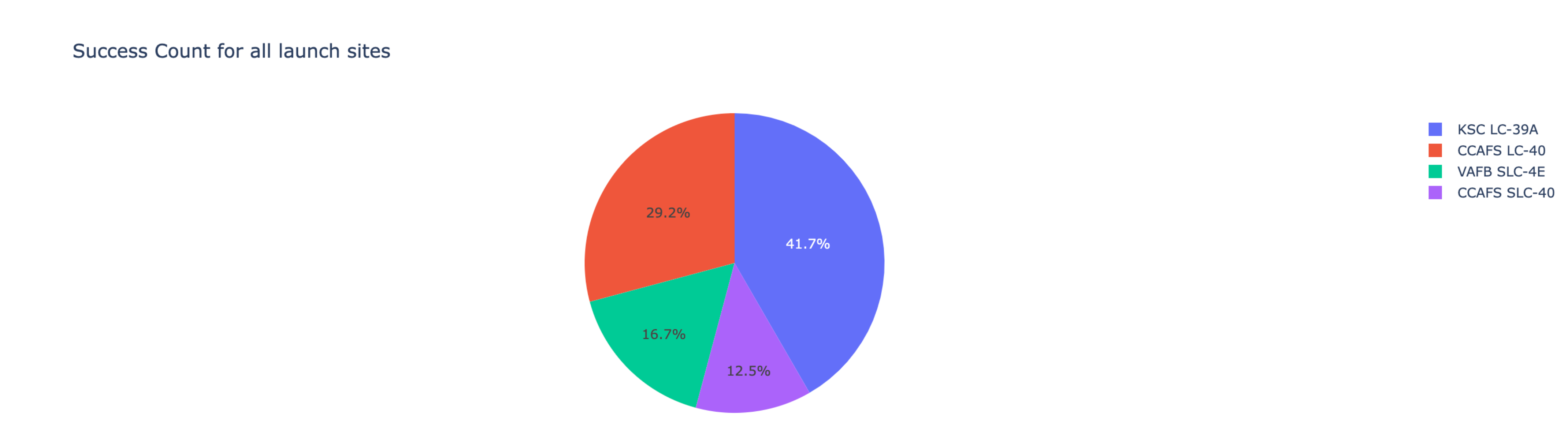
Calculate the distances between a launch site to its proximities



Section 4

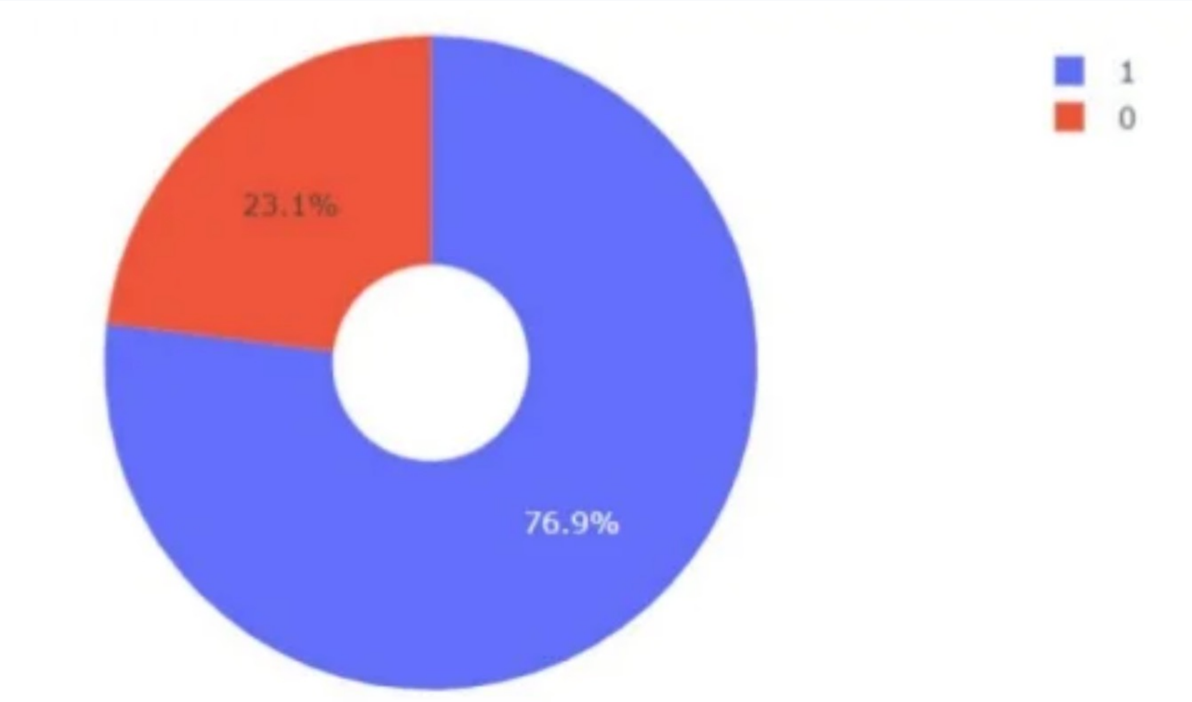
Build a Dashboard with Plotly Dash

Success Count for all launch sites



KSC LC-39A had most successful launches from all the sites.

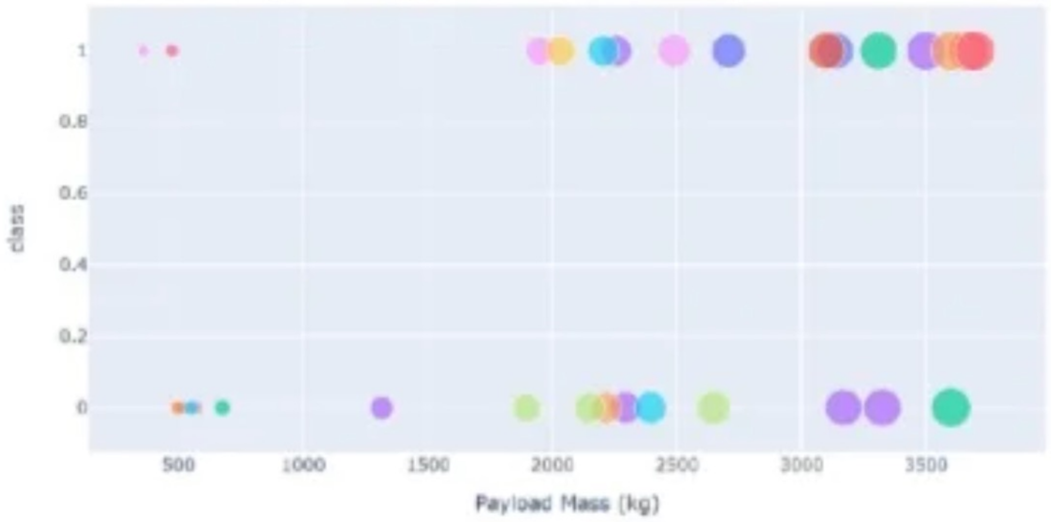
Success rate by site



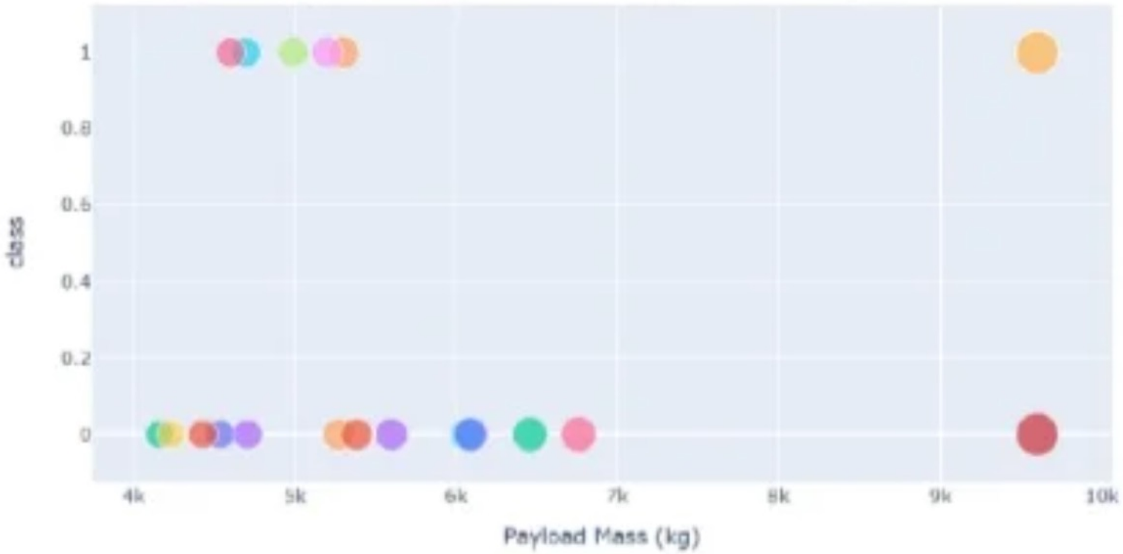
KSC LC-39A achieved a 76,9% success rate while getting a 23,1% failure rate.

Payload vs launch outcome

Low Weighted Payload 0kg – 4000kg



Heavy Weighted Payload 4000kg – 10000kg

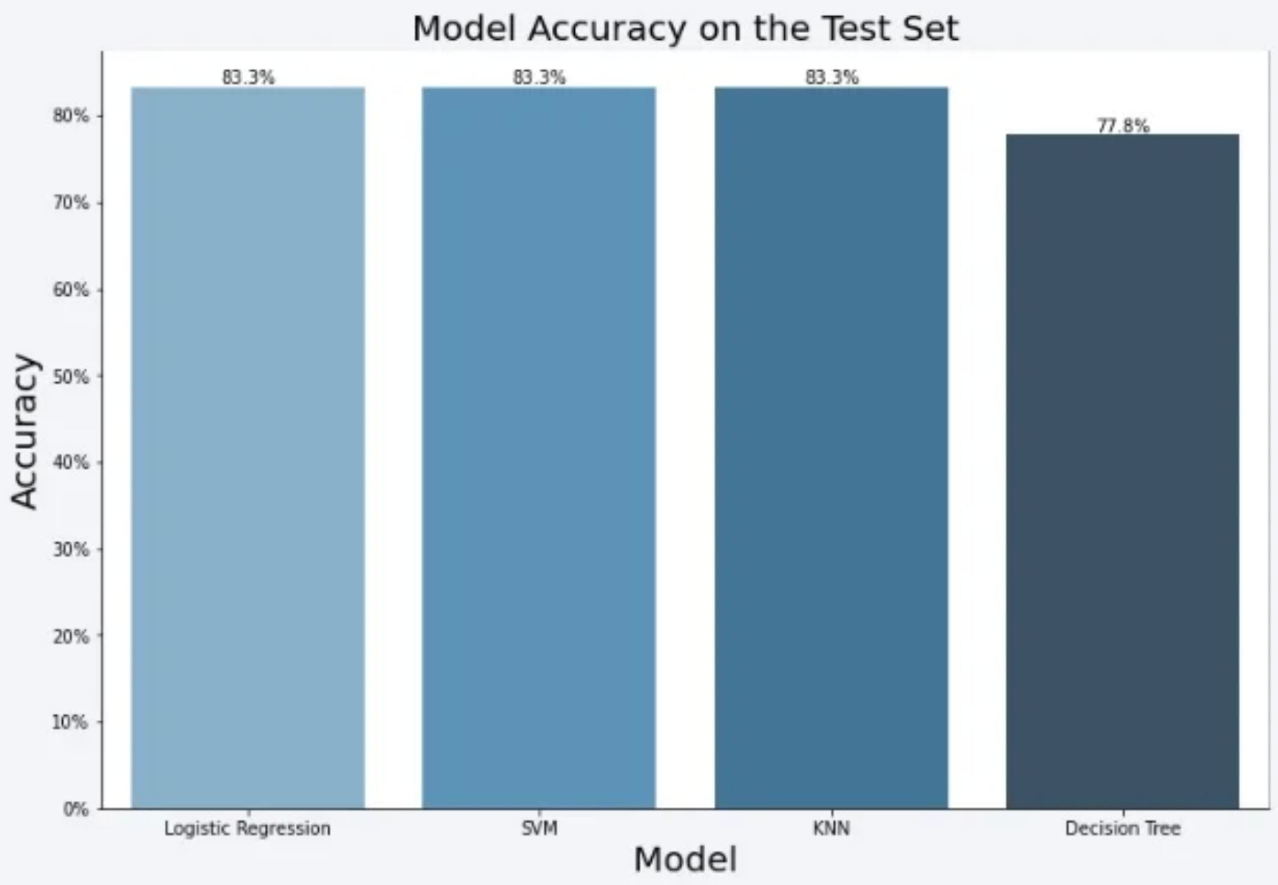


We can see the success rates for low weighted payloads is higher than the heavy weighted payloads

Section 5

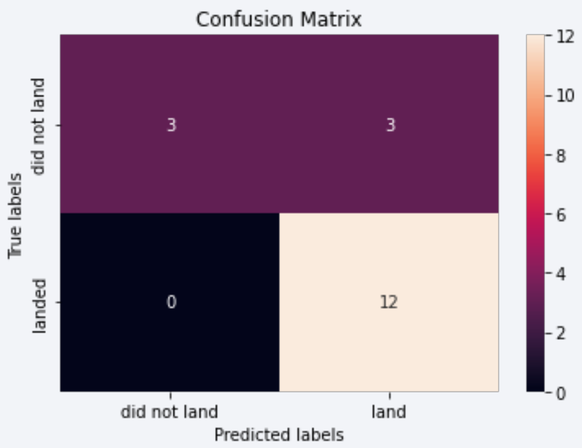
Predictive Analysis (Classification)

Classification Accuracy



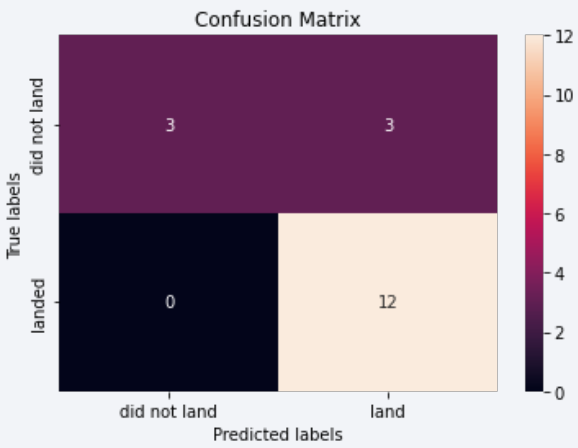
Logistic Regression, SVM and KNN have best accuracy

Confusion Matrix



SVM

Logistic Regression



KNN

- 2 models have same confusion matrix.
- All landed target values (12) are predicted correctly.
- 50% of "Do not land" values are predicted incorrectly.

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 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!