IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Name: Hamid Alibeiki

Date: 07 – 04- 2023

Github : https://github.com/hamidyazd/SpaceX-Capstone

Outline

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

Executive Summary

- **Summary of methodologies**

✓ Data Collection (API – Web Scraping)

✓ Data Wrangling

✓ Exploratory Data Analysis (SQL – Data Visualization)

✓ Interactive Visual Analytics (Folium)

✓ ML Prediction

- **Summary of all results**

  ✓ Exploratory Data Analysis result

  ✓ Interactive analytics in screenshots

  ✓ Predictive Analytics result

Introduction

- **Project background and context**
- Space X company said that Falcon 9 rocket launches by 62 million dollars cost; other companies cost is 165 million dollars. The main reason of this gap is because of reusing of the first stage by Space X. So, if we can predict the success of landing of the first stage, we can calculate the launch's cost. This results will be useful for company wants to be competitor of space X company. I want to predict the landing successfulness for the first stage by designing a machine learning method.
- **Problems you want to find answers**
- What are the items for a successful landing?
- What are the relations between various features which that affect the success rate of a successful landing?
- What are the conditions for a ensure successful landing?

Section 1

Methodology

Methodology

Executive Summary

- **Data collection methodology:**

✓ Space X Rest API

✓ Web Scrapping from Wikipedia

- **Perform data wrangling**

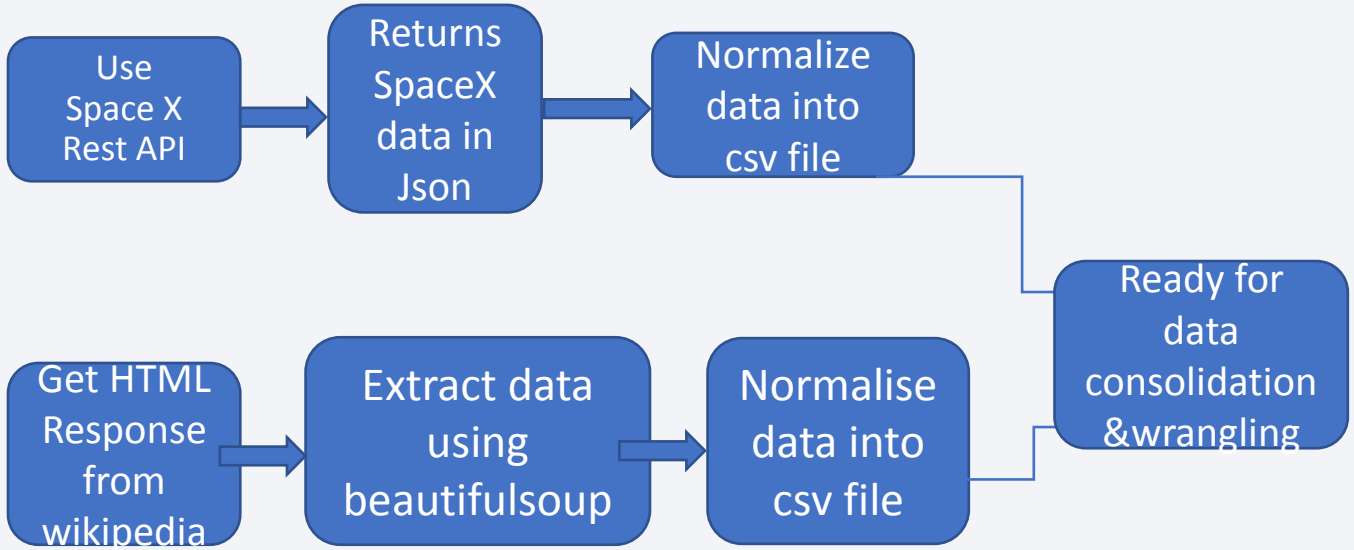
✓ One Hot Encoding data fields for ML & data cleaning of missed values and irrelevant columns.

- **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, SVM, DT and KNN models have been built & evaluated to find the best classifier.

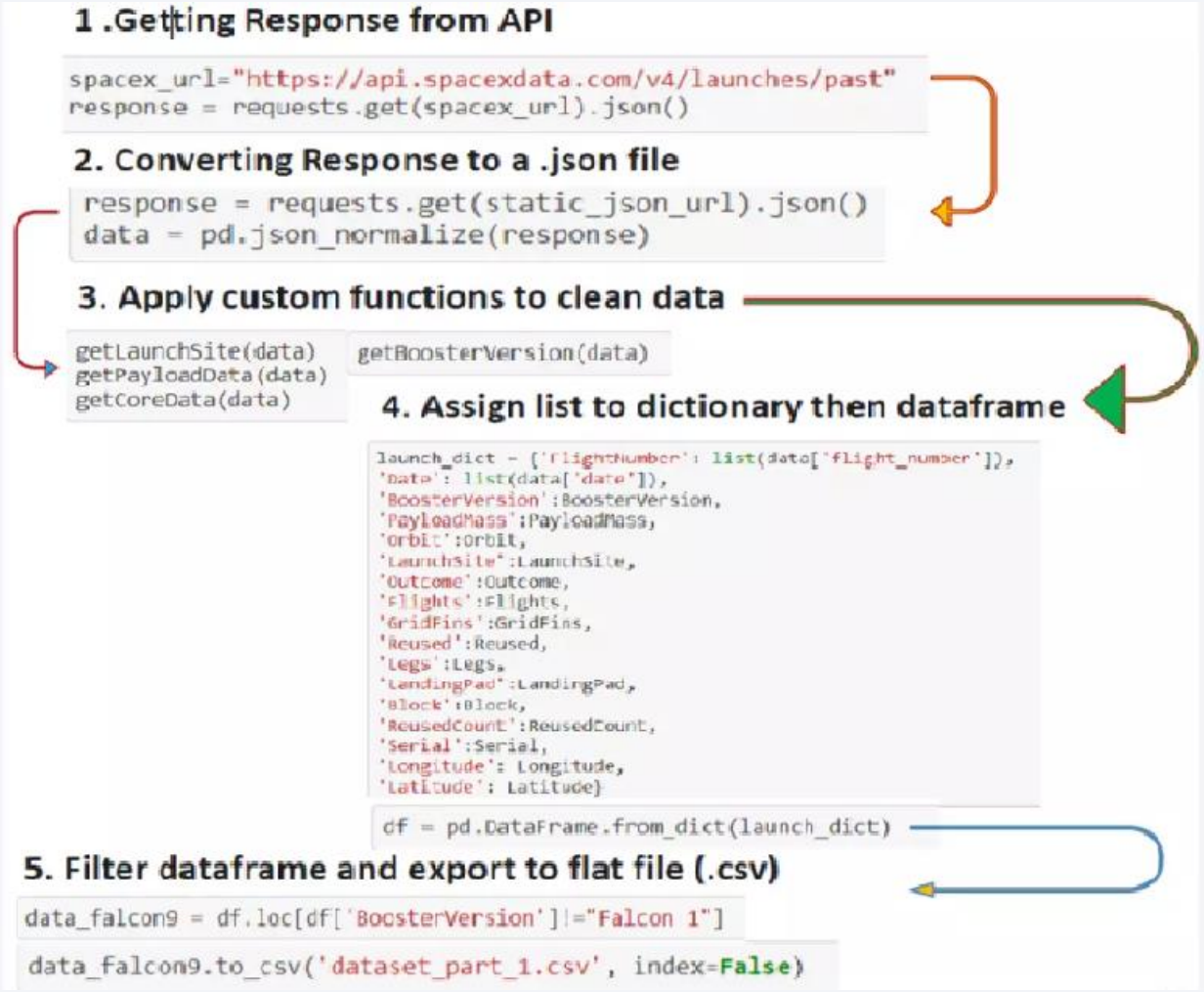
Data Collection

- **Describe how data sets were collected.**
- The datasets was collected:
- Space X launch data that is gathered from the Space X REST API.
- This API give us data for launches, included information of rocket used, payload delivered, launch and landing specifications, and landing outcome.
- The SpaceX REST API endpoints, or URL, starts here api.spacexdata.com/v4/.
- Another data source for obtaining the Falcon 9 Launch data is web scraping from Wikipedia using BeautifulSoup.



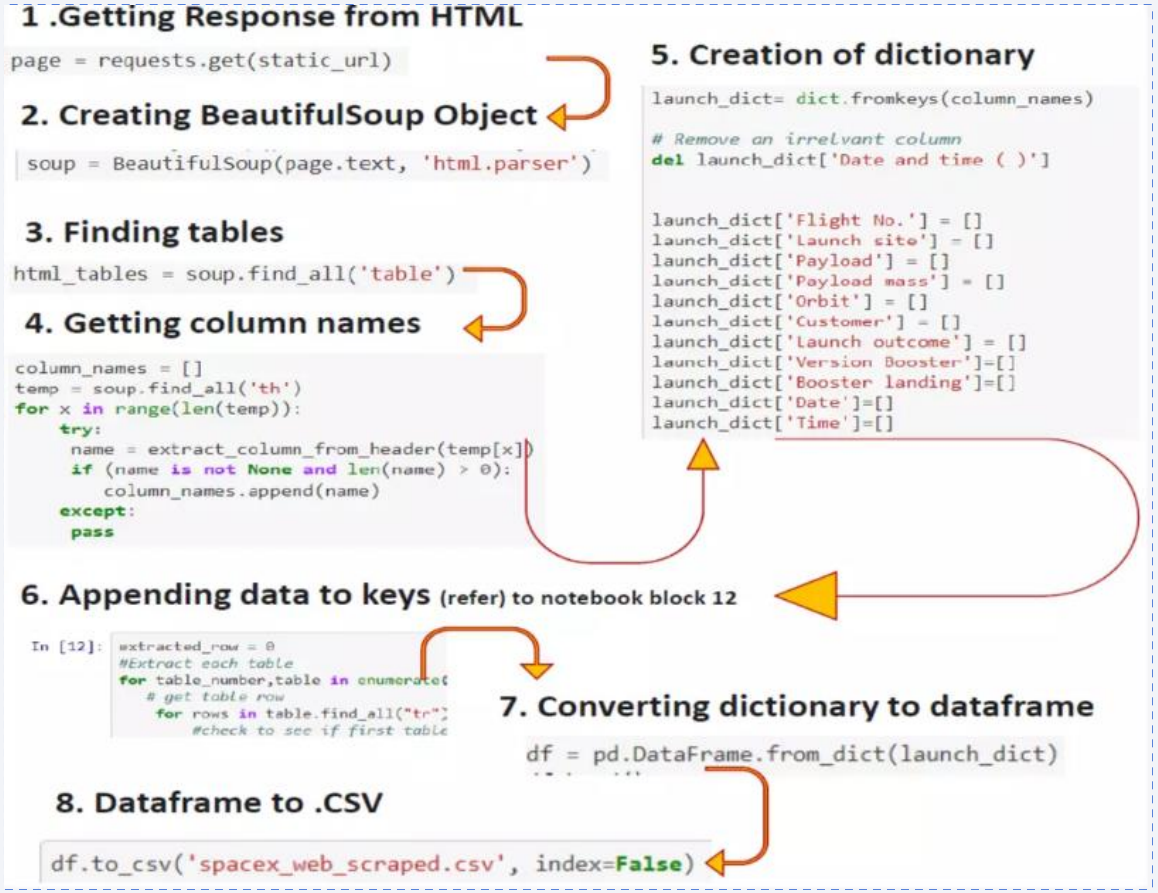
Data Collection – SpaceX API

- ✓ Data collection by Space X REST calls

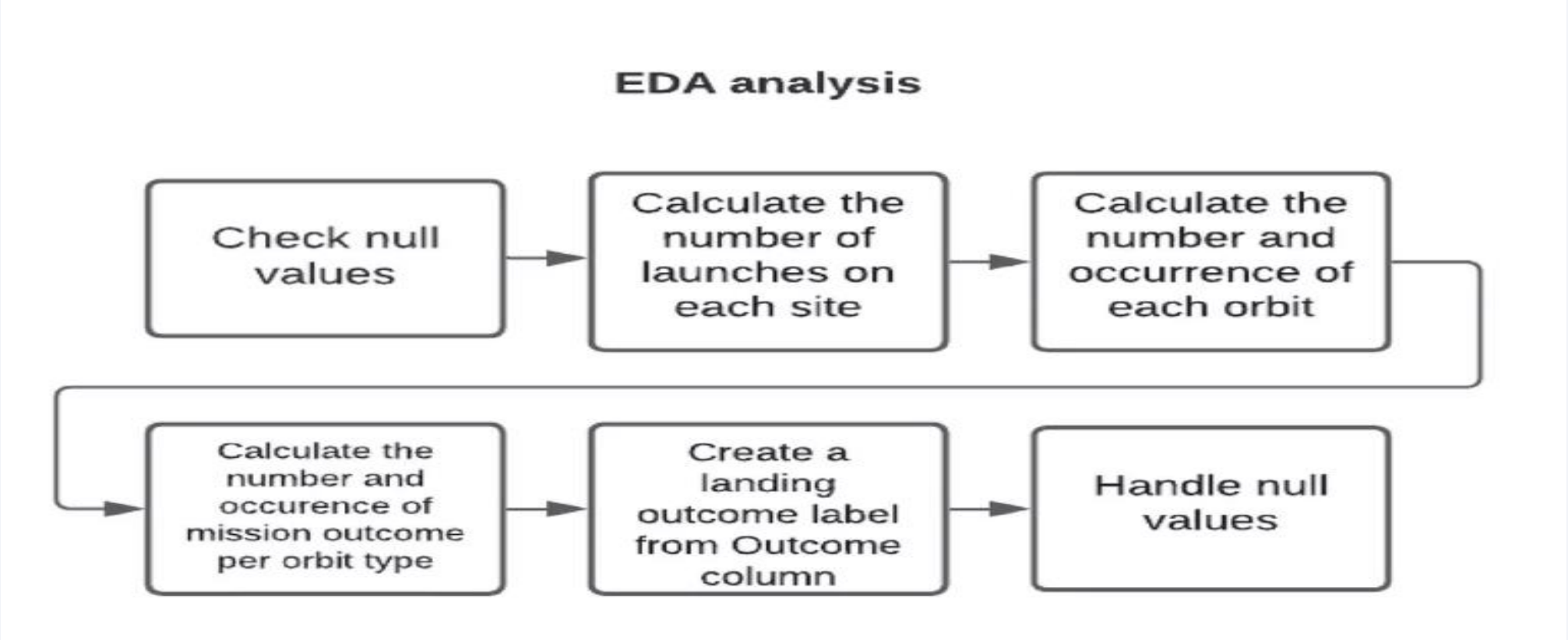


Data Collection - Scraping

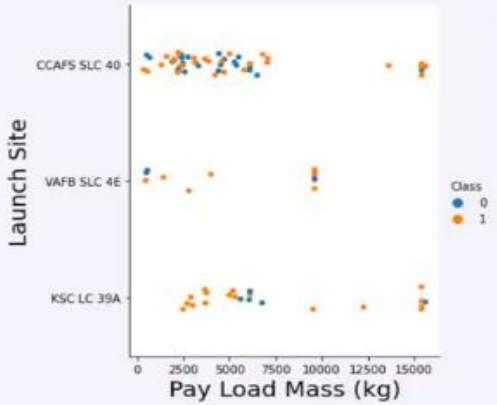
- ✓ Web scrapping: Wikipedia

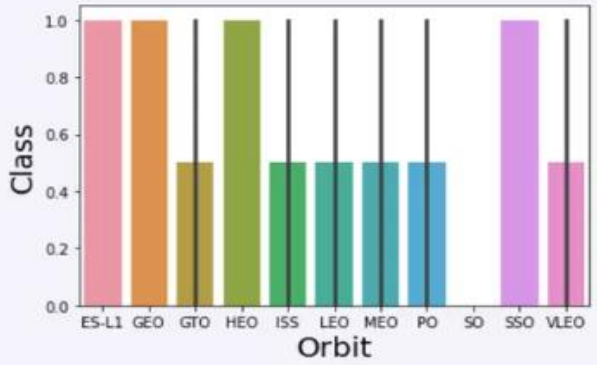


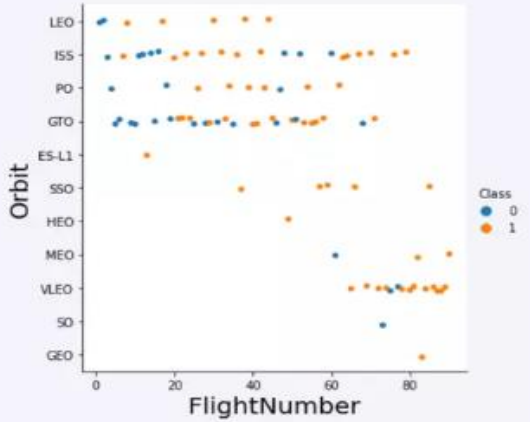
Data Wrangling

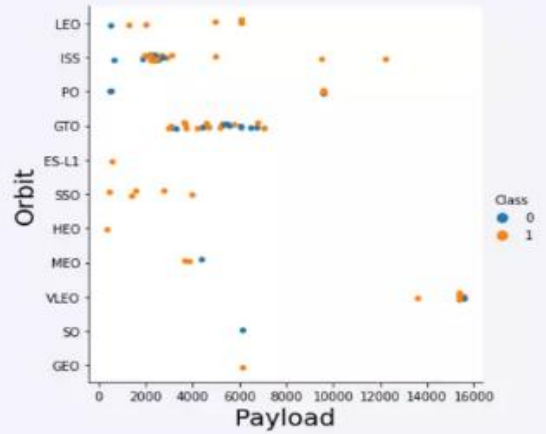


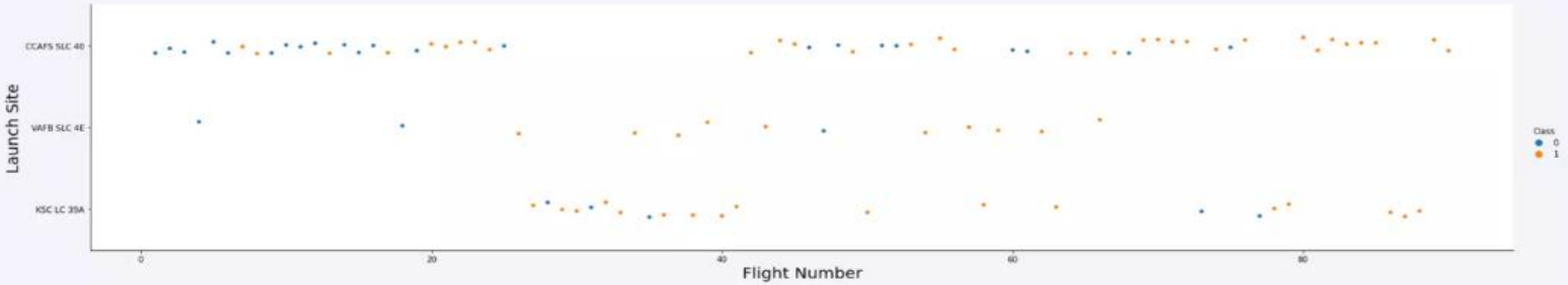
EDA by Data Visualization

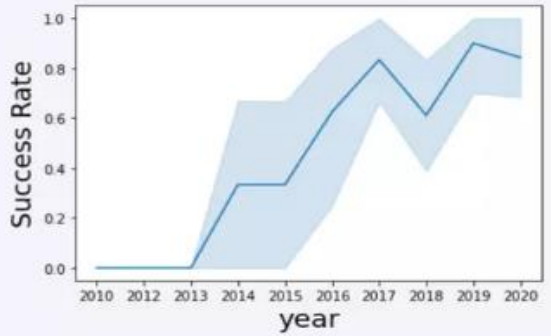












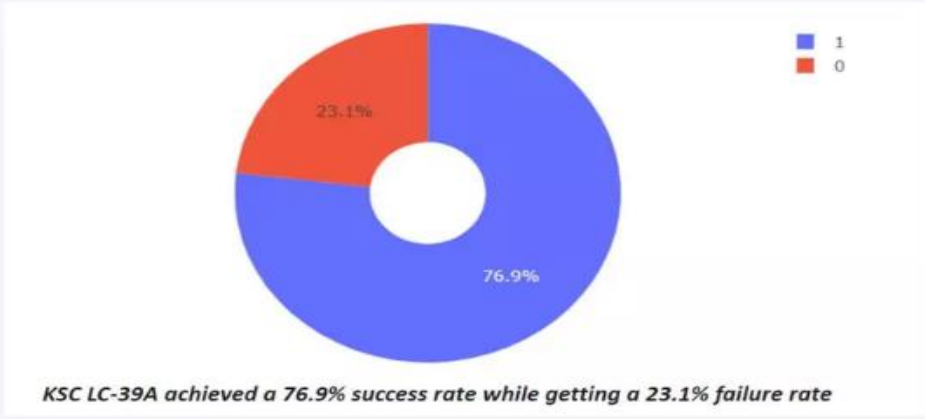
EDA with SQL

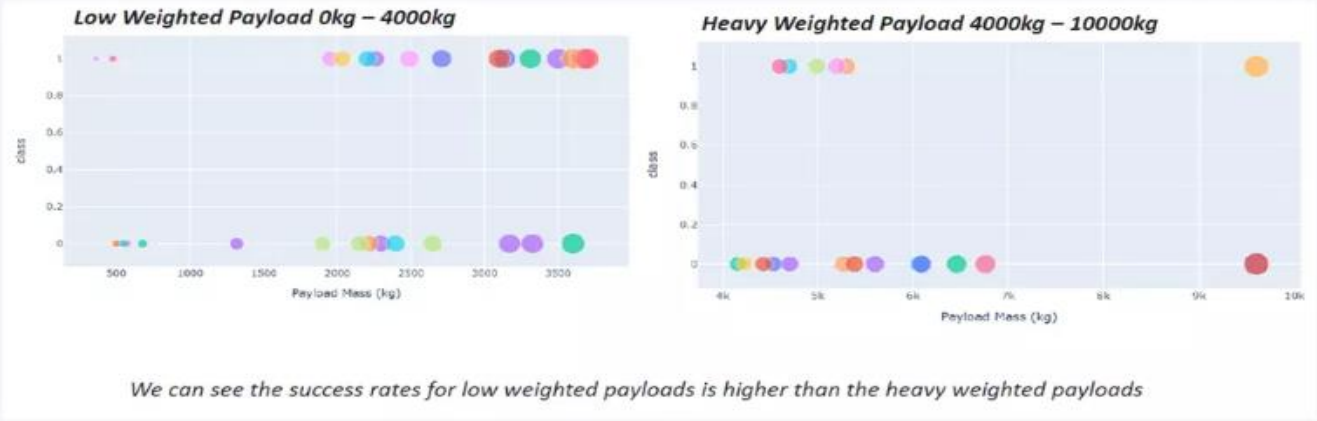
- ✓ Displaying the unique launch site's names in space mission
- ✓ Displaying 5 records where launch sites begin with the string 'KSC
- ✓ Displaying the total payload mass carried by boosters launched by NASA (CRS)
- ✓ Displaying average payload mass carried by booster version F9 v1.1
- ✓ Listing the 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 between 4000 up to 6000.
- ✓ Listing the number of successful & failure mission outcomes
- ✓ Listing the names of the booster versions carried the maximum payload mass
- ✓ Listing the records displayed the month names, successful landing outcomes in ground pad booster versions, launch site for the months in 2017
- ✓ Ranking the count of successful landing outcomes between 2010 06 04 & 2017 03 20 in descending order.

Build an Interactive Map with Folium

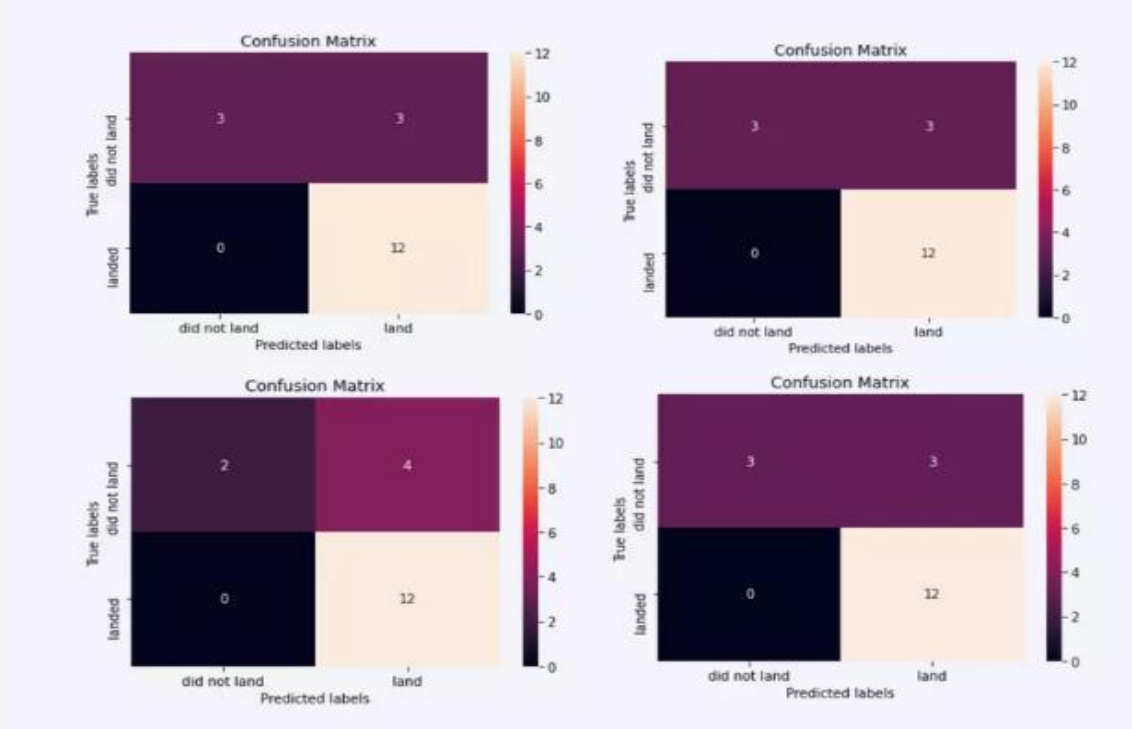
Map markers added to map to finding an optimal location for building a launch site.

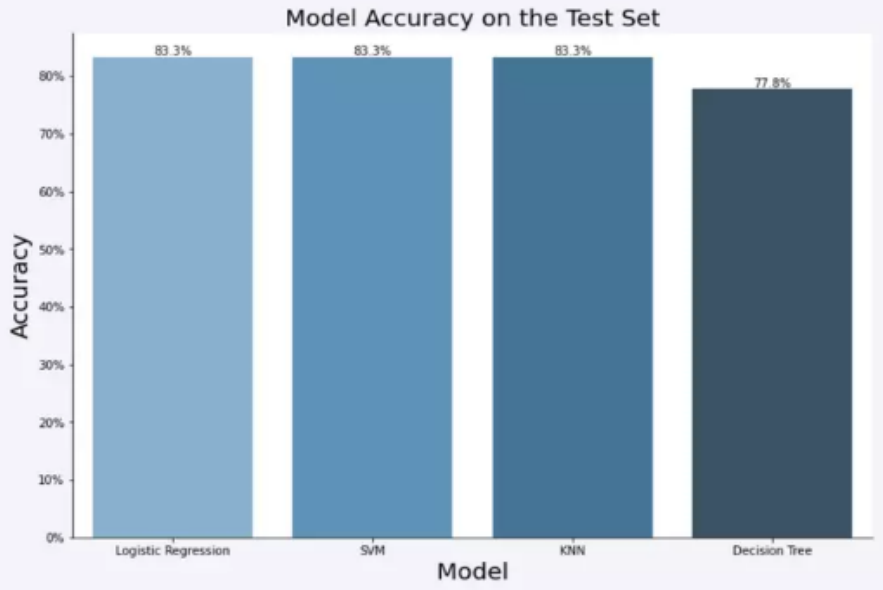
Build a Dashboard with Plot ly Dash





Predictive Analysis (Classification)

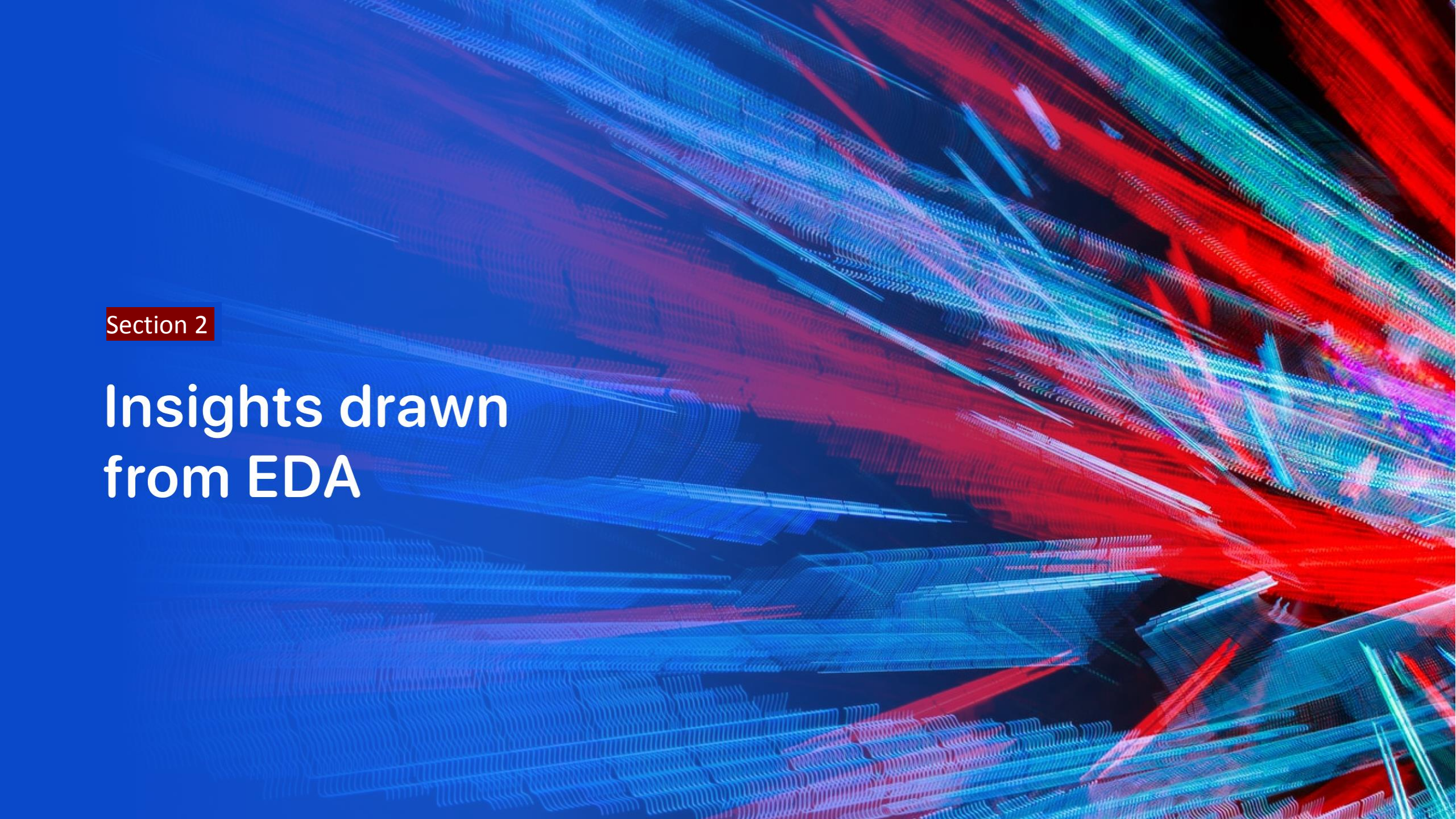




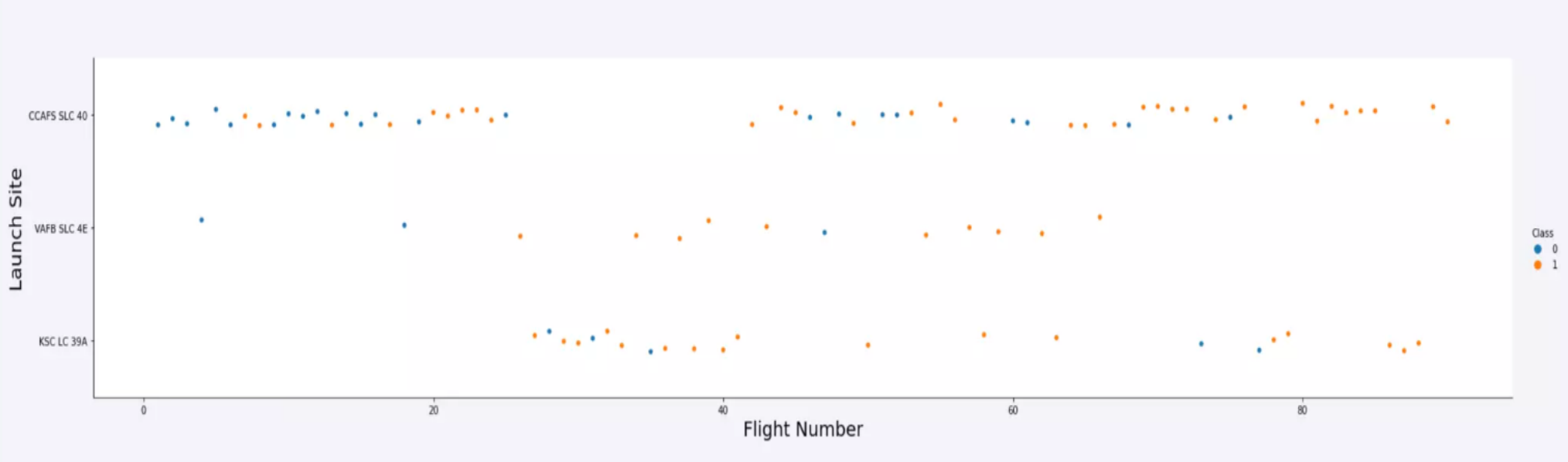
The KNN, SVM & Logistic Regression model achieved the highest accuracy at 83.3%, but SVM performs the best in terms of Area Under the Curve at 0.958!

Results

- ✓ The KNN, SVM & Logistic Regression models are the best in terms of prediction accuracy for our dataset.
- ✓ Low weighted payloads perform better heavier payloads than!
- ✓ The success rates for Space X launches is directly proportional time in years eventually perfect the launches.
- ✓ KSC LC 39A had the most successful launches from all the sites we considered.
- ✓ Orbit GEO,HEO,SSO,ES L1 has the best Success Rate.


Section 2
Insights drawn from EDA

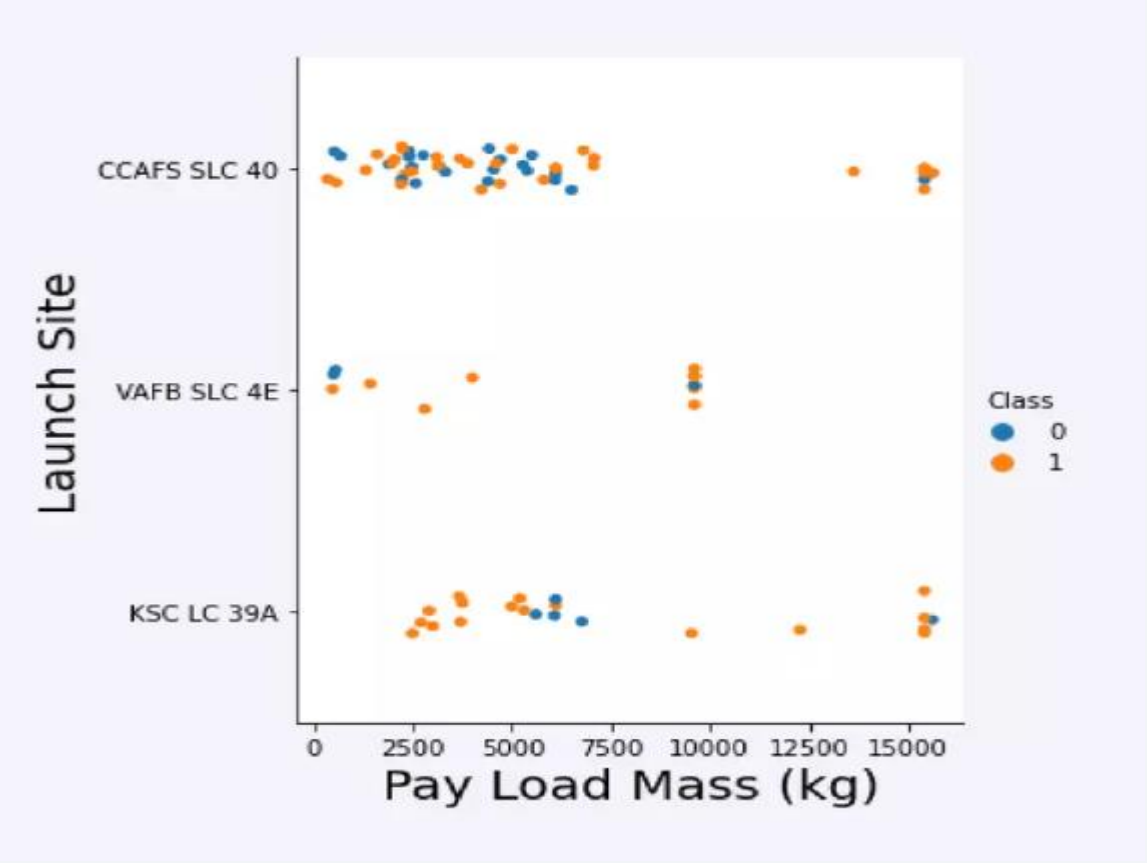
Flight Number VS Launch Site



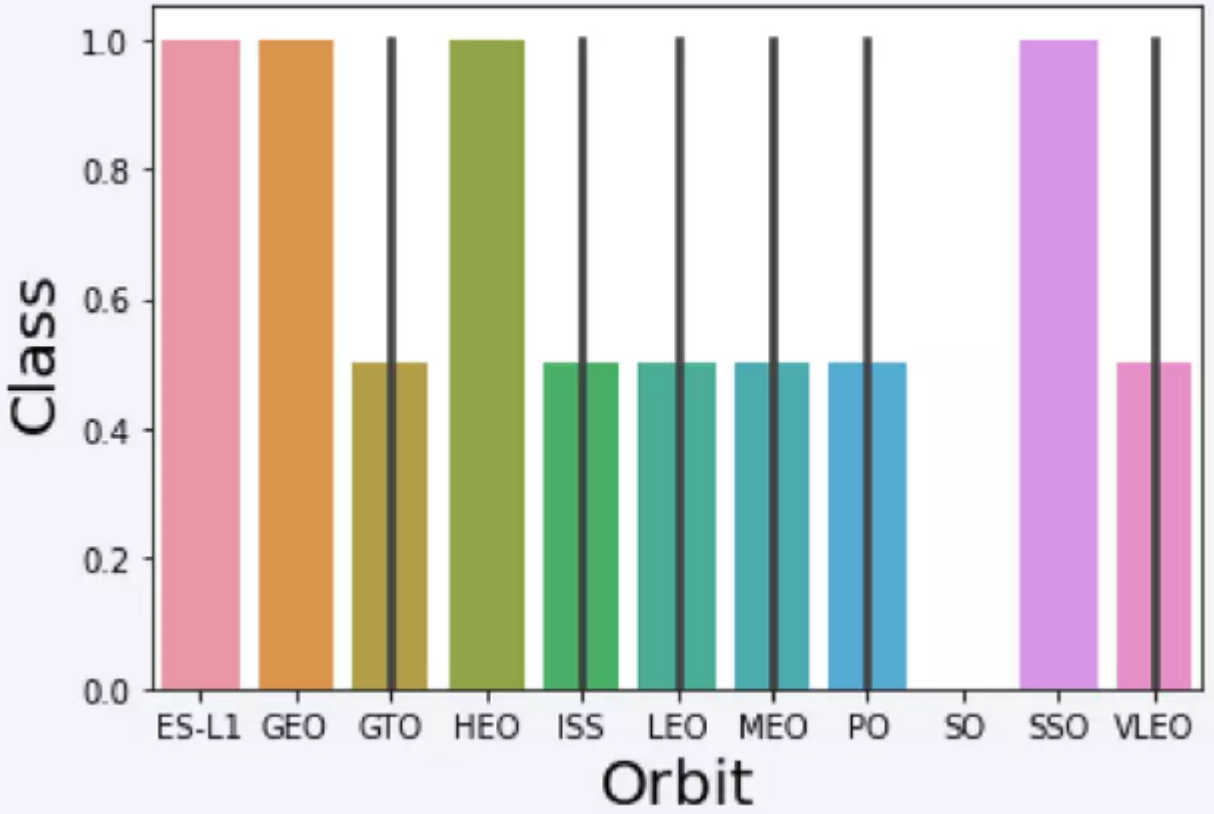
- ✓ Launches from the site of CCAFS SLC 40 are launches form other sites.
- ✓ significantly higher lunches from other sites than.

Payload VS Launch Site

✓The majority of IPay Loads with lower Mass launched from CCAFS SLC 40.

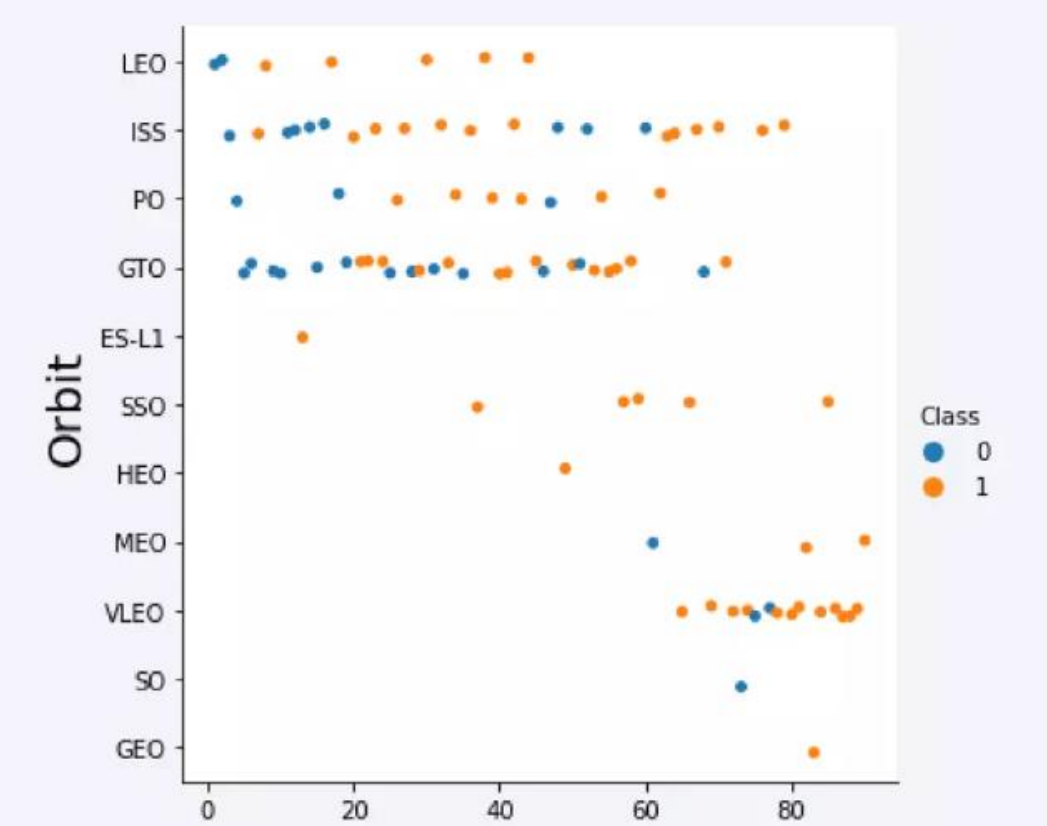


Success Rate vs. Orbit Type



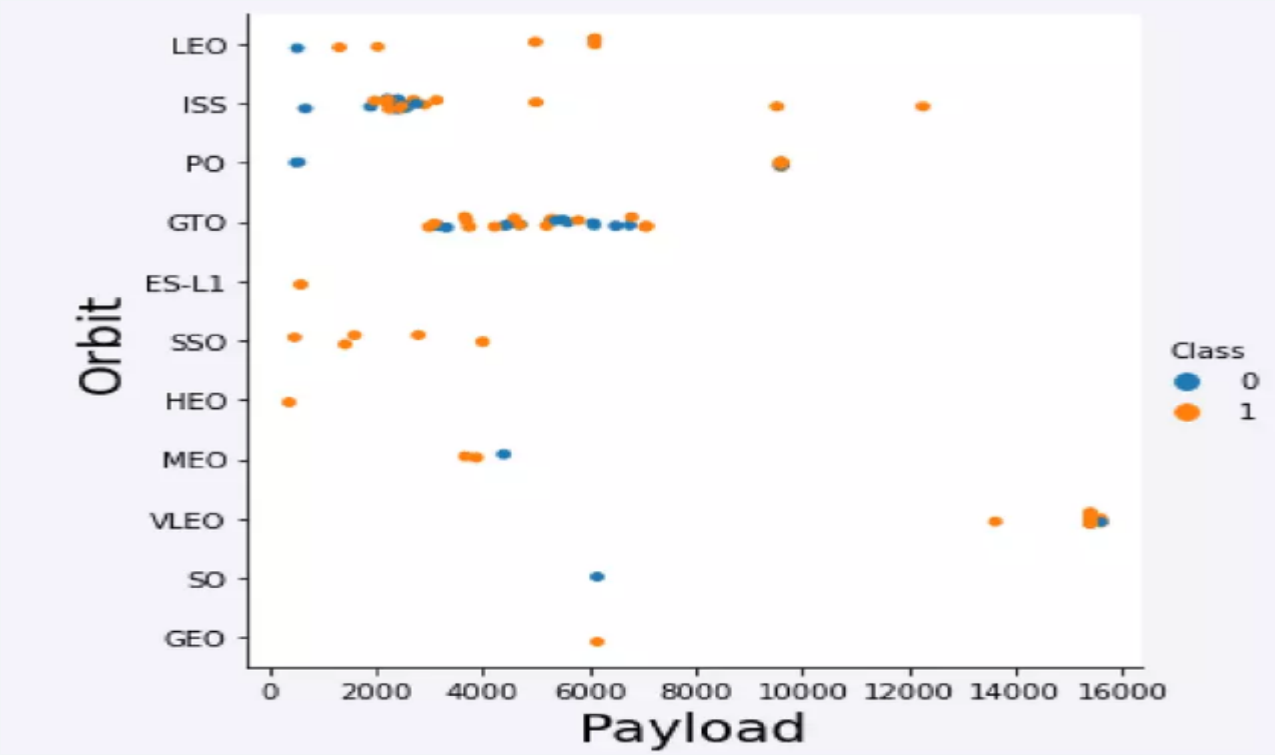
✓ The orbit types SSO, HEO, ES-L1 & GEO have the highest success rate.

Flight Number vs. Orbit Type



✓A trend shows a shifting to VLEO launches in recent years.

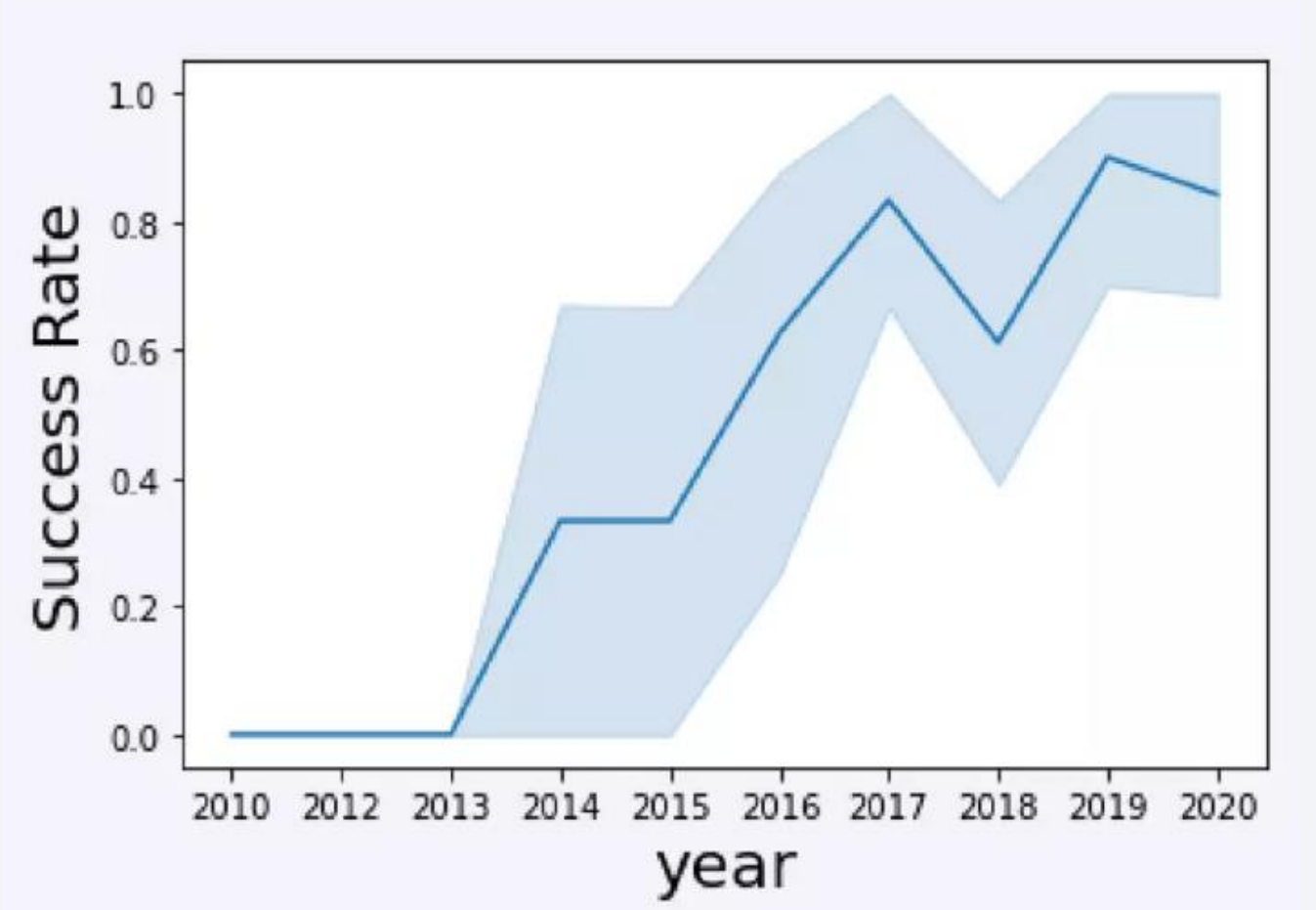
Payload vs. Orbit Type



✓ Strong correlation between ISS & Payload at the range around 2000, and GTO at the range of 4000-8000 as well.

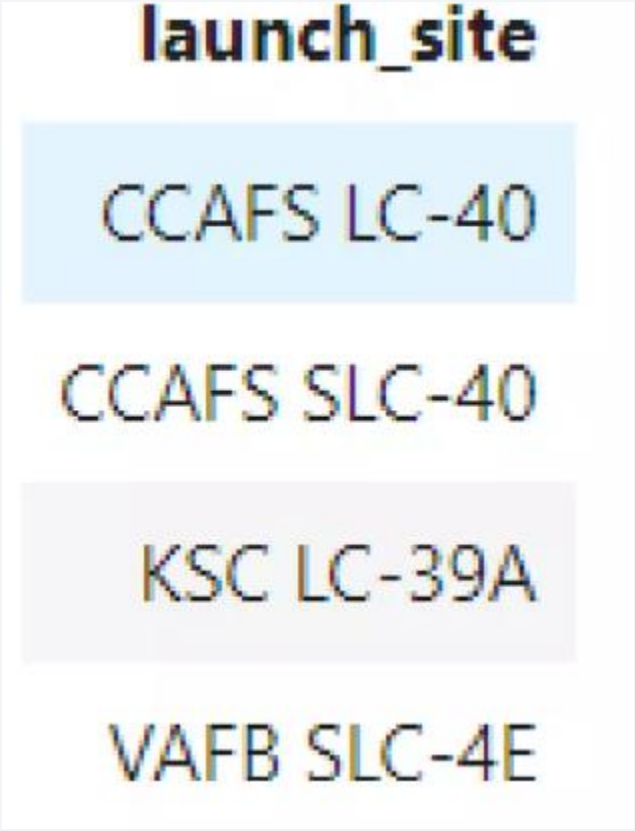
Launch Success Yearly Trend

- ✓ Launch success rate has increased significantly since 2013 & established since 2019, because of advance in technology as well as lessons learned.



All Launch Site Names

✓%sql select distinct (LAUNCH_SITE) from SPACEXTBL



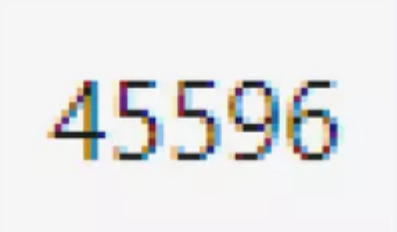
Launch Site Names Begin with 'CCA'

✓ %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)'



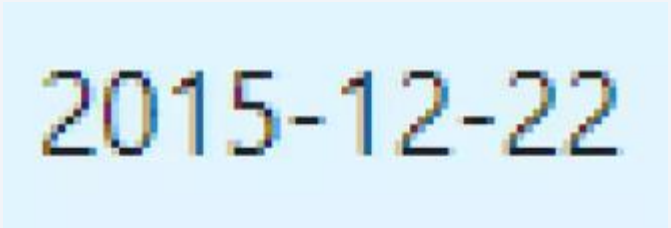
Average Payload Mass by F9 v1.1

✓ %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)'



Successful Drone Ship Landing with Payload 4000-6000

%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)

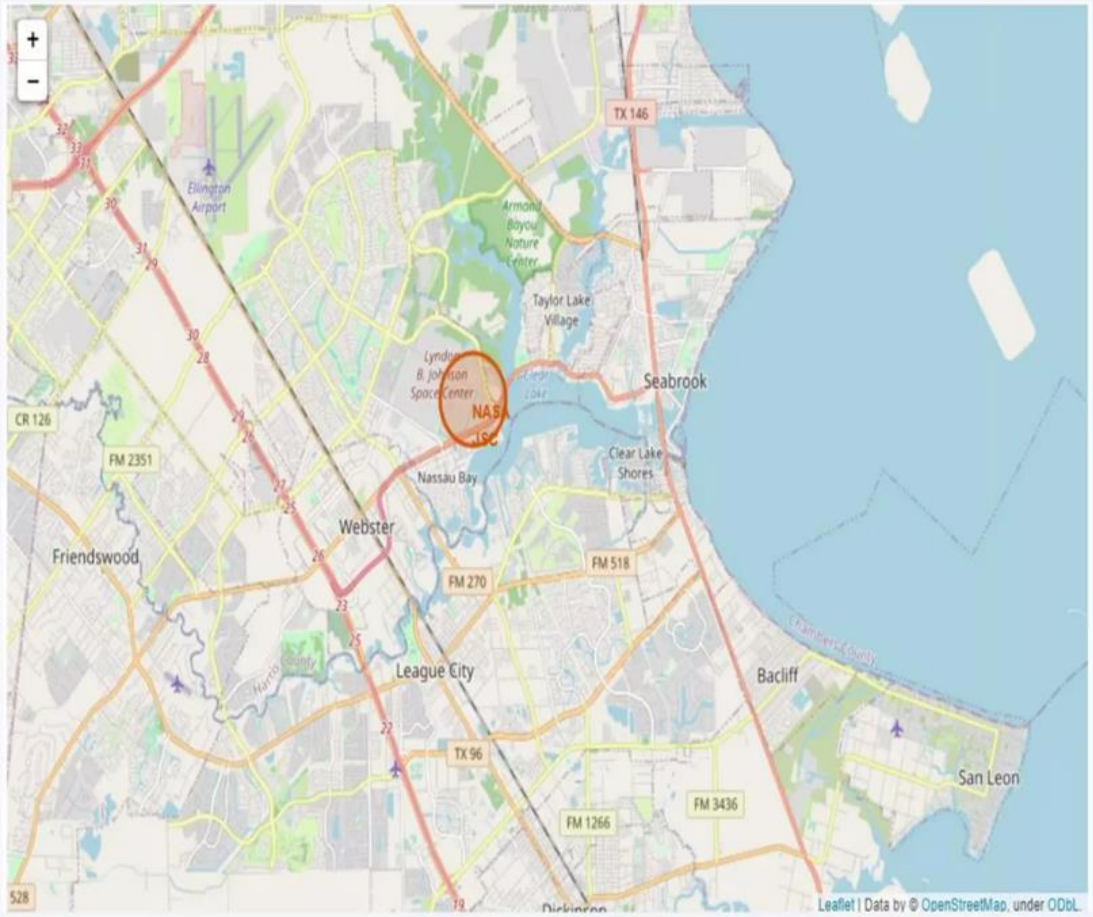
Rank Landing Outcomes Between 2010-06-04 & 2017-03-20

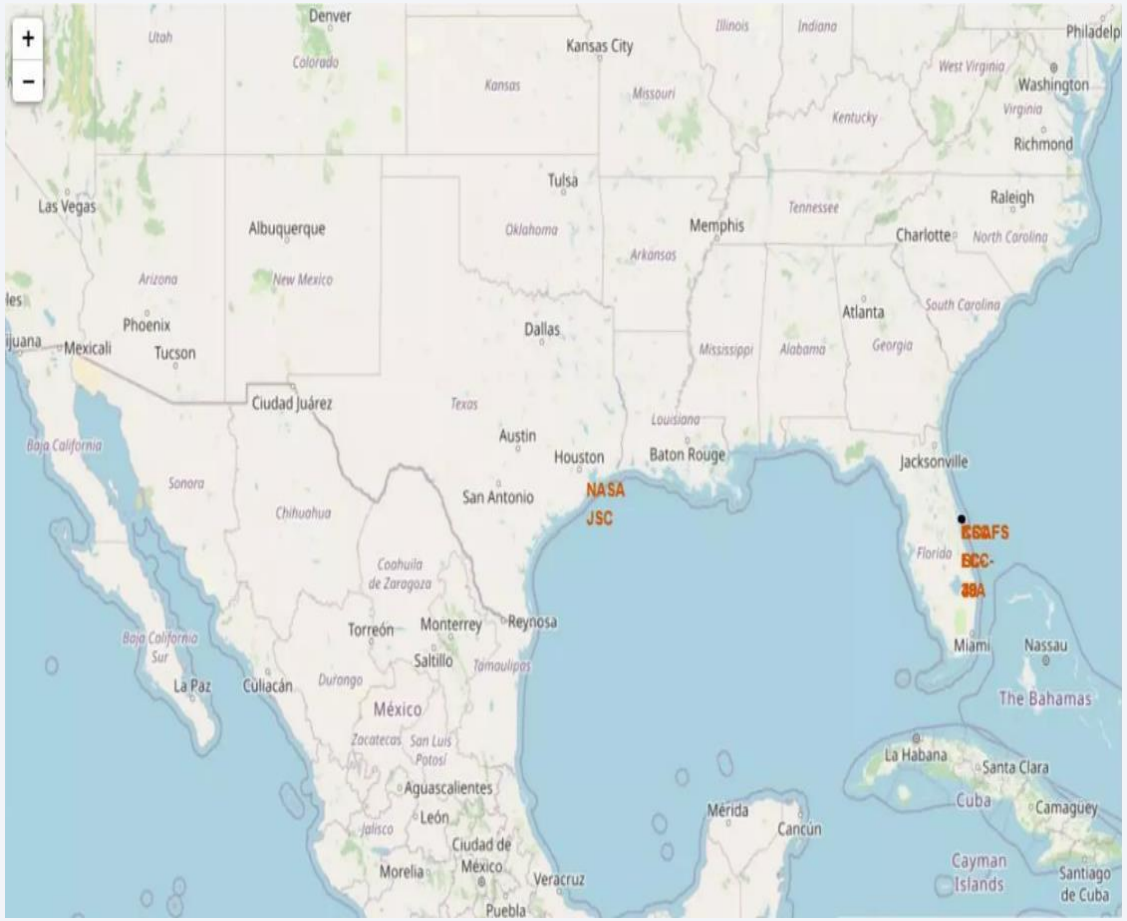
✓ %sql select * from SPACEXTBL where Landing Outcome like 'Success%' and (DATE between '2010-06-04' & '2017-03-20') ordered by date desc…

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)

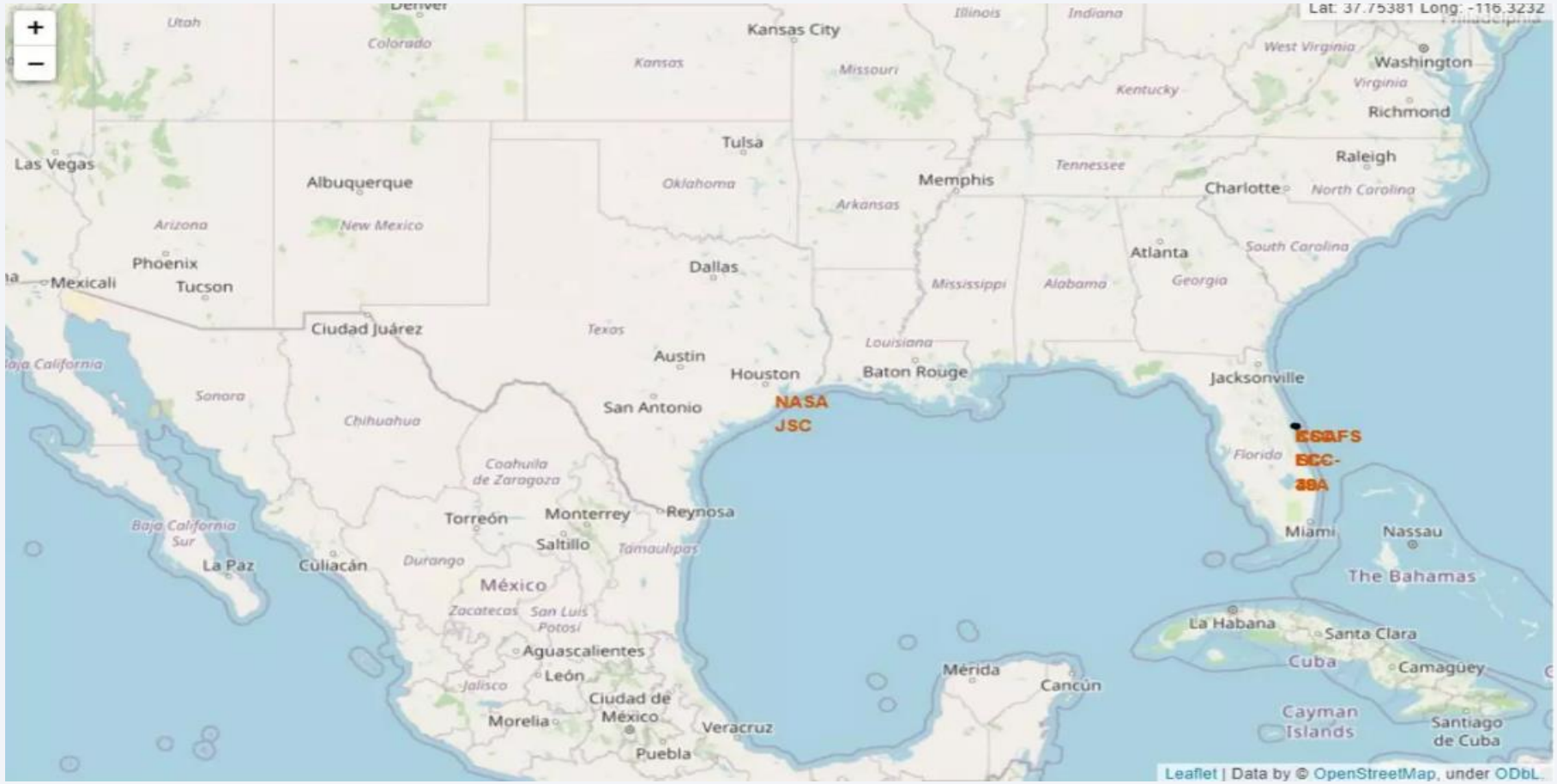

Section 3
Launch Sites
Proximities Analysis

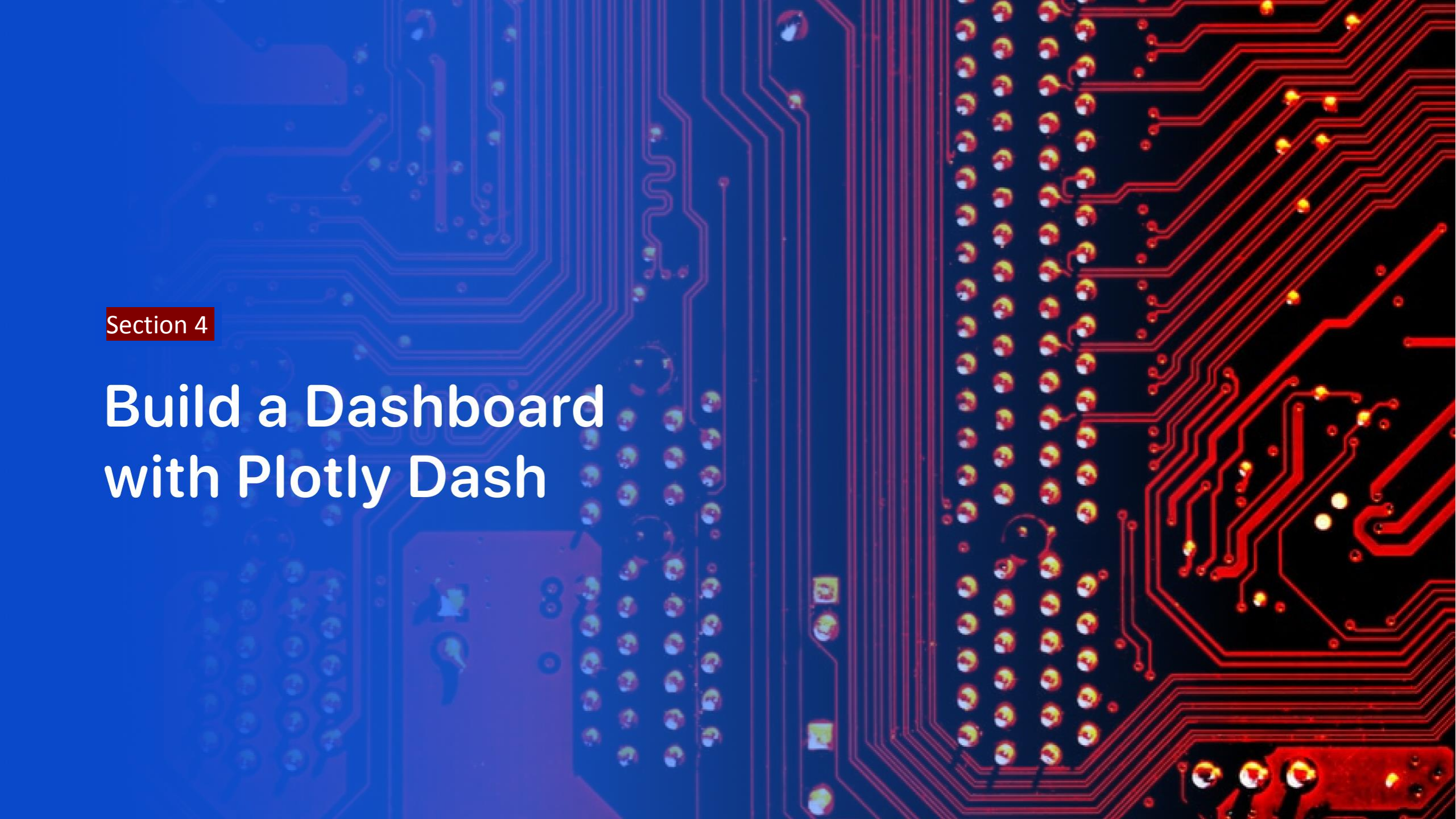
Folium Map 1


TX 146
Ellington Airport
Armand Bayou Nature Center
Taylor Lake Village
Seabrook
NASA JSC
Clear Lake Shores
Nassau Bay
CR 126
FM 2351
Webster
Friendswood
FM 270
FM 518
League City
Bacliff
TX 96
FM 1266
FM 3436
San Leon
528
Leaflet | Data by © OpenStreetMap, under ODbL

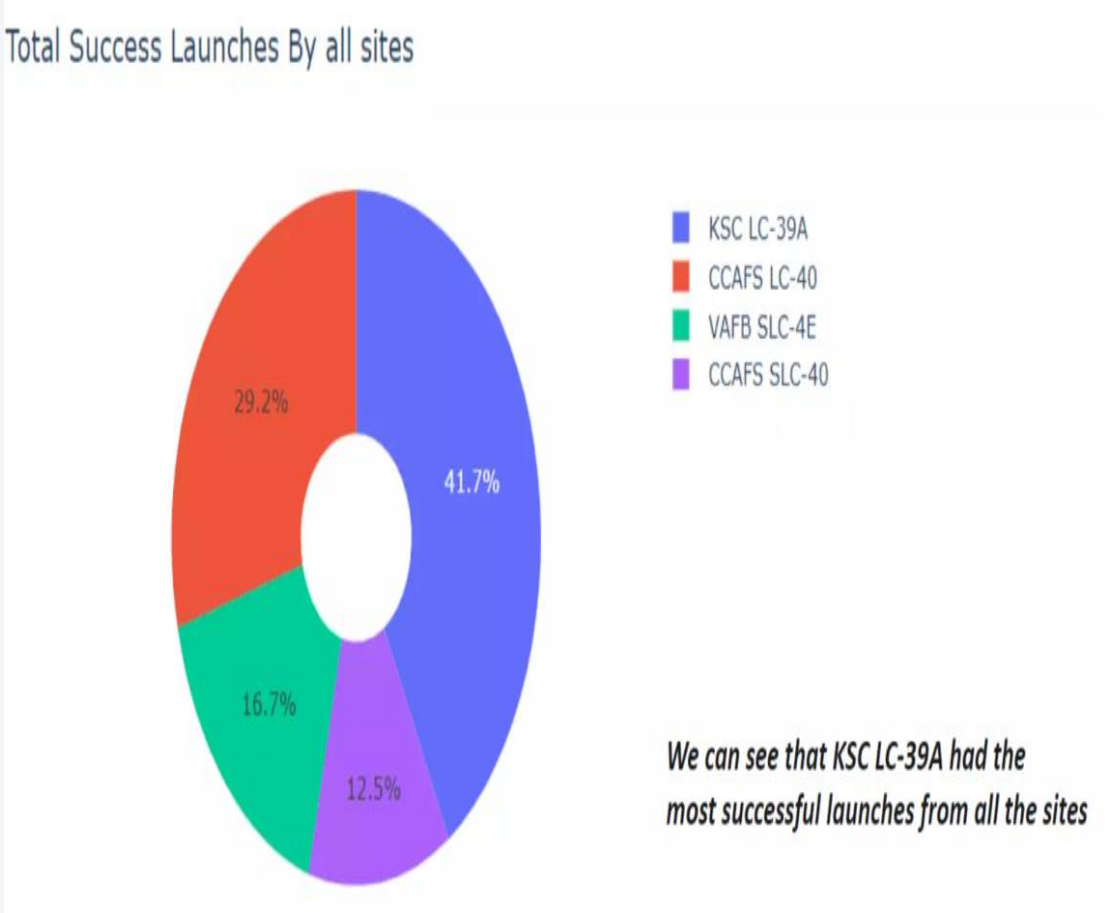

Denver
Kansas City
Las Vegas
Albuquerque
Phoenix
Tucson
Mexicali
Ciudad Juárez
Dallas
Austin
Houston
San Antonio
NASA JSC
Baton Rouge
Jacksonville
Memphis
Atlanta
Charlotte
Raleigh
Richmond
Washington
Philadelp
Tulsa
Miami
Nassau
The Bahamas
La Habana
Santa Clara
Cuba
Camaguey
Cayman Islands
Santiago de Cuba
Mérida
Cancún
Monterrey
Reynosa
Torreón
Saltillo
Culiacán
La Paz
México
Ciudad de México
Veracruz
Puebla
Morelia
León
Aguascalientes
Utah
Colorado
Kansas
Missouri
Illinois
Indiana
West Virginia
Virginia
Kentucky
Tennessee
North Carolina
South Carolina
Georgia
Mississippi
Alabama
Arkansas
Oklahoma
Arizona
New Mexico
Texas
Louisiana
Florida
Baja California
Sonora
Chihuahua
Coahuila de Zaragoza
Durango
Zacatecas
San Luis Potosí
Tamaulipas
Jalisco

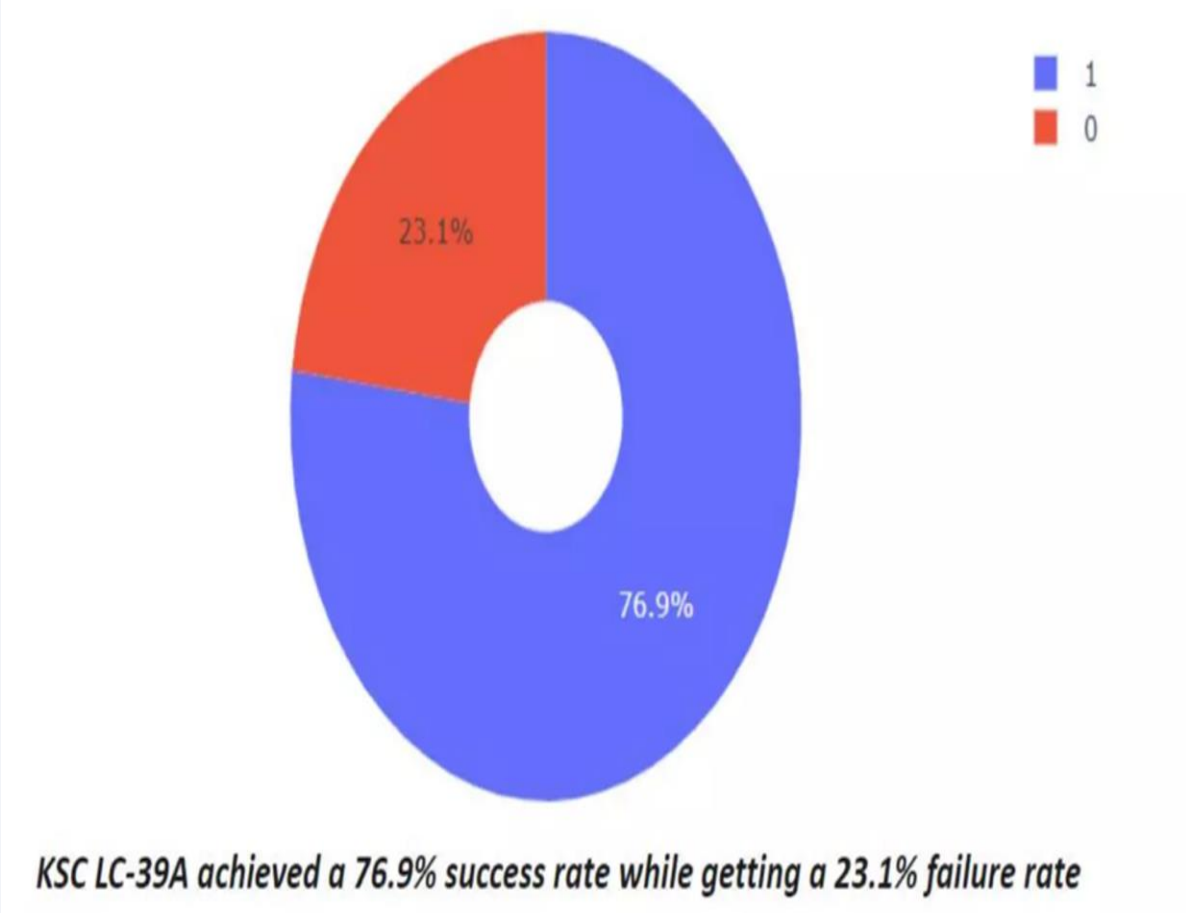
Folium Map 2


Lat: 37.75381 Long: -116.3232
Kansas City
Tulsa
Las Vegas
Albuquerque
Memphis
Charlotte
Raleigh
Washington
Richmond
Atlanta
Dallas
Phoenix
Tucson
Mexicali
Ciudad Juárez
Austin
Houston
Baton Rouge
Jacksonville
San Antonio
NASA
JSC
CCAFS
SLC-
40
Miami
Nassau
The Bahamas
Monterrey
Reynosa
Torreón
Saltillo
La Paz
Culiacán
México
La Habana
Santa Clara
Cuba
Camaguey
Mérida
Cancún
Cayman
Islands
Santiago
de Cuba
Aguascalientes
León
Ciudad de
México
Morelia
Veracruz
Puebla
Leaflet | Data by © OpenStreetMap, under ODbL.

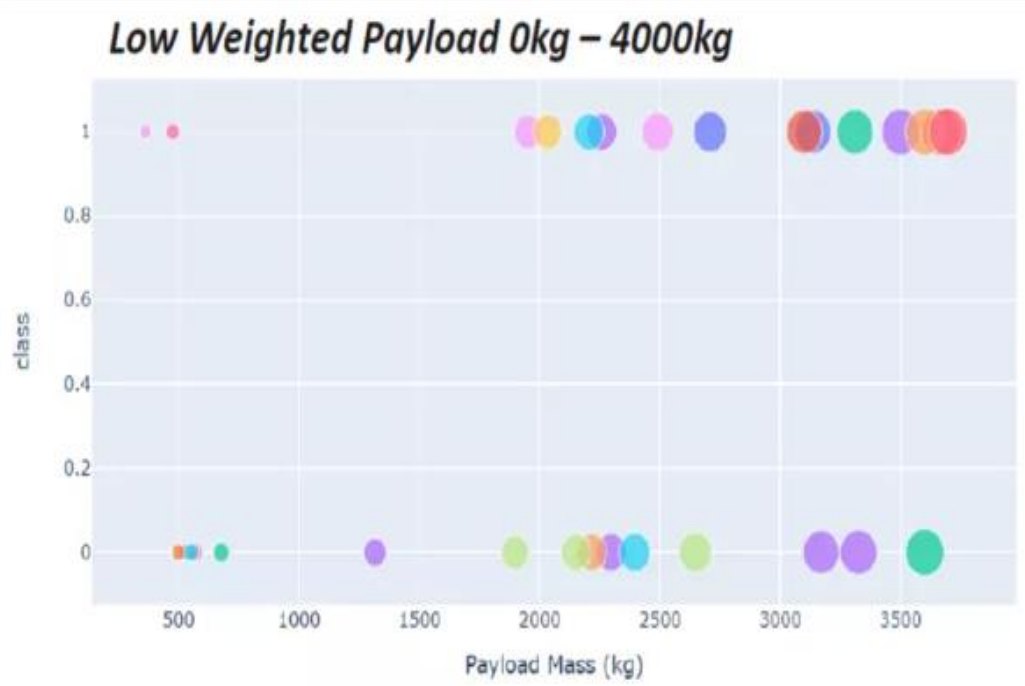

Section 4
Build a Dashboard with Plotly Dash

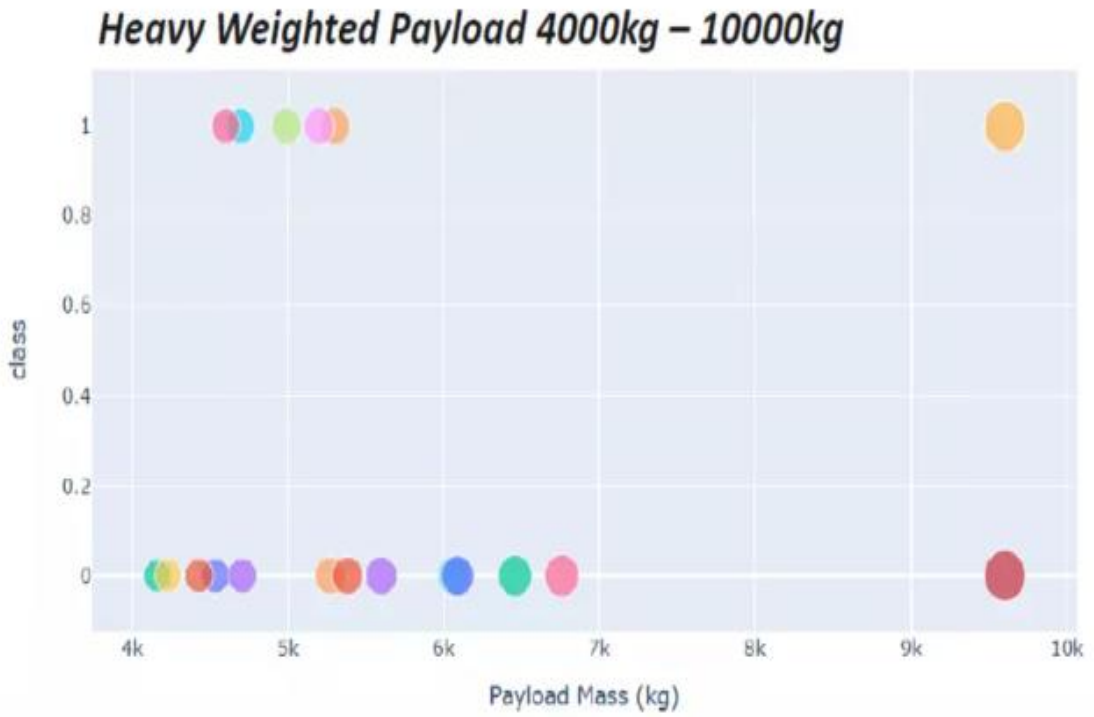
Dashboard


Total Success Launches By all sites
KSC LC-39A
CCAFS LC-40
VAFB SLC-4E
CCAFS SLC-40
41.7%
29.2%
16.7%
12.5%
We can see that KSC LC-39A had the most successful launches from all the sites


1
0
23.1%
76.9%
KSC LC-39A achieved a 76.9% success rate while getting a 23.1% failure rate

Dashboard



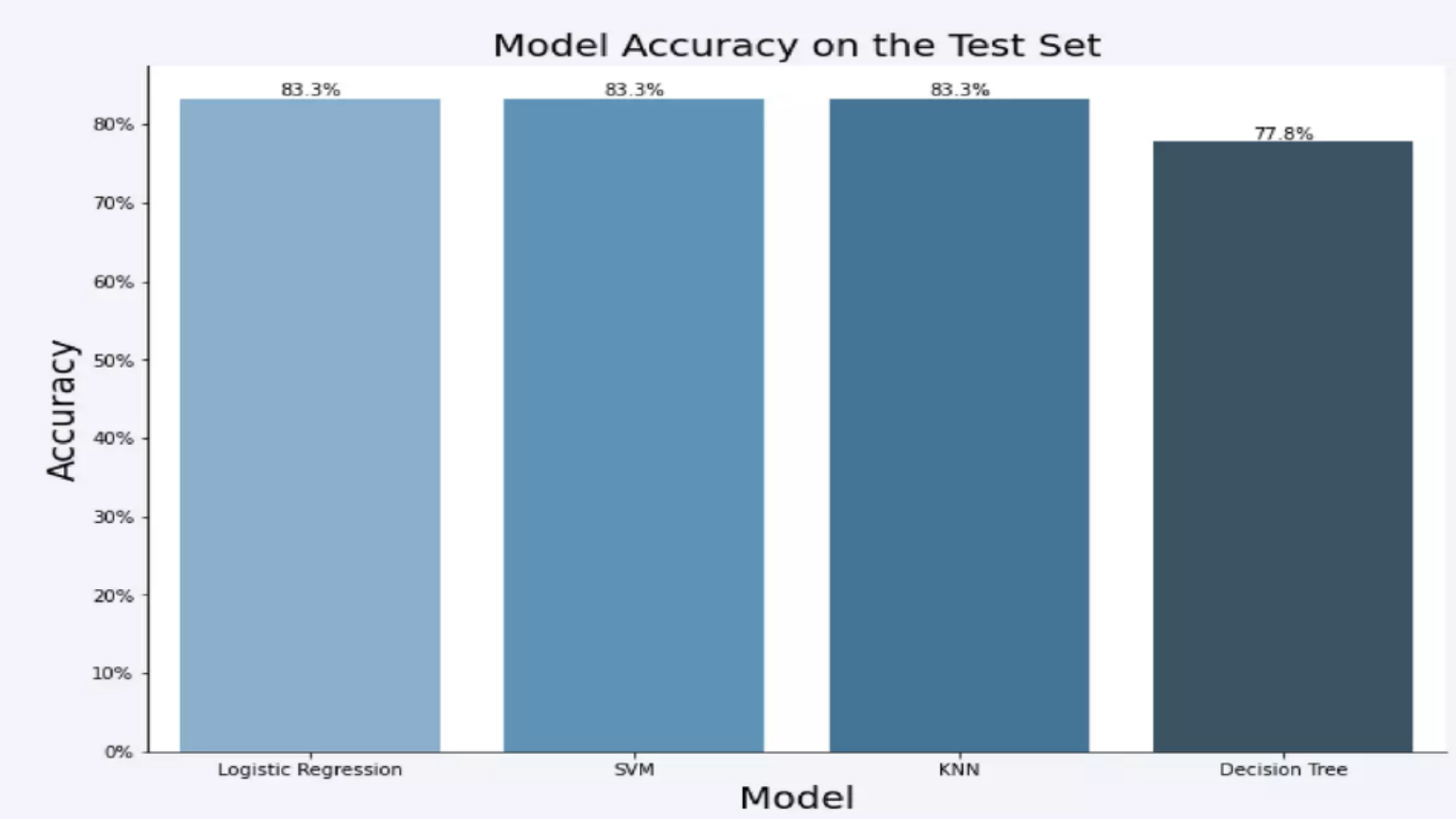


We can see the success rates for low weighted payloads is higher than the heavy weighted payloads

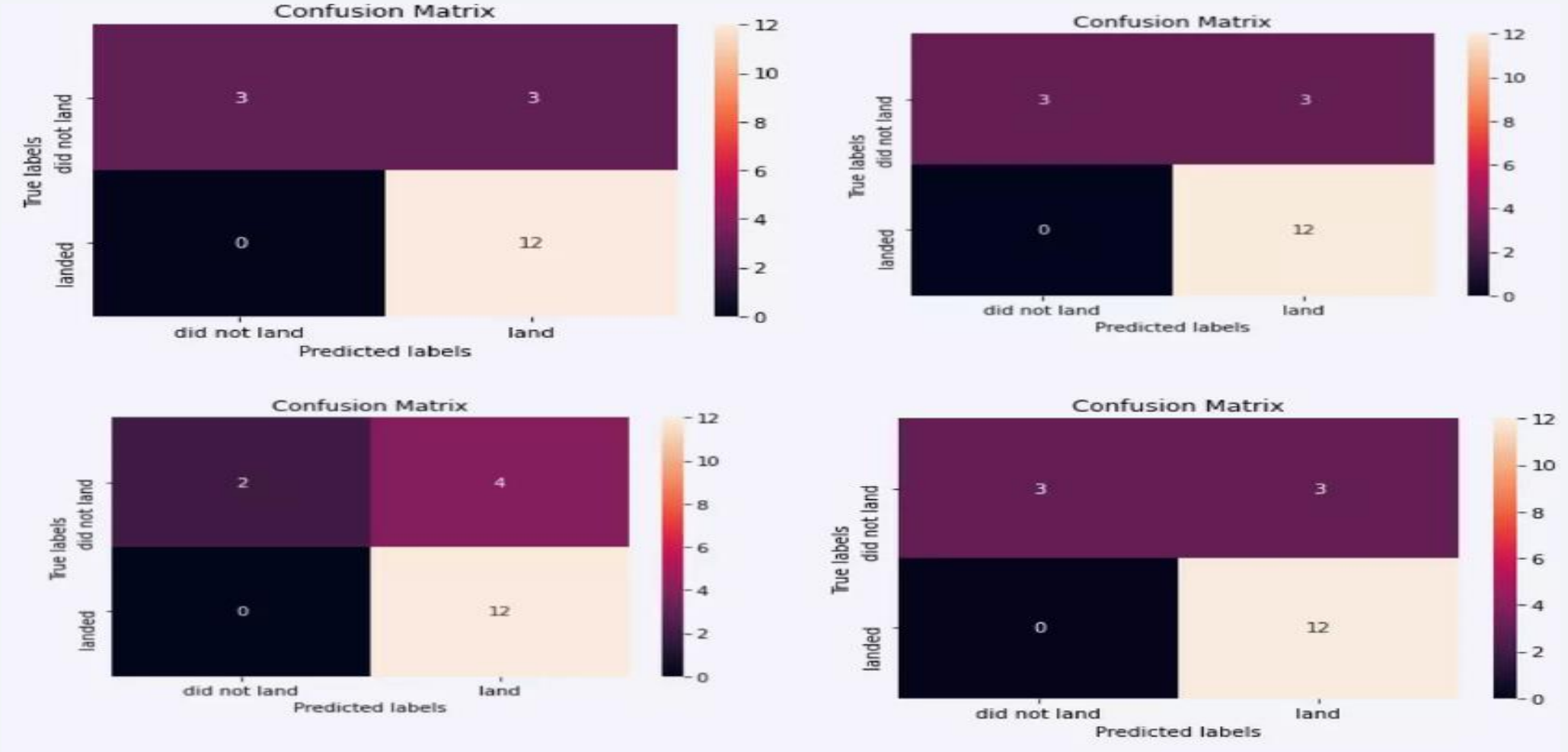
Section 5

Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix



Conclusions

- ✓ The KNN, SVM & Logistic Regression models are the best in prediction accuracy for this dataset.
- ✓ Low weighted payloads perform better heavier payloads than.
- ✓ The success rates for Space X launches is directly proportional time in years they will eventually perfect the launches.
- ✓ The most successful launches from all the sites was by KSC LC 39A.
- ✓ Orbit HEO, SSO, GEO & ES L1 has the best Success Rate.

Thank you!