IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Faye Green
25/05/24

Outline

Executive Summary

- Summary of methodologies
  - Data Collection
  - Data Wrangling
  - EDA with Data Visualization
  - EDA with SQL
  - Building an Interactive Map with Folium
  - Predictive Analysis (Classification)
- Summary of all results
  - EDA Results
  - Interactive Analytics
  - Predictive Analytics

Introduction

- Project background and context

  SpaceX offers Falcon 9 rocket launches on its website for $62 million, while other providers charge over $160 million per launch. The significant cost savings are largely due to SpaceX's ability to reuse the rocket's first stage.

- Problems you want to find answers

  The project's goal is to predict whether the first stage of the SpaceX Falcon 9 rocket will successfully land.

Section 1

Methodology

Methodology

Executive Summary

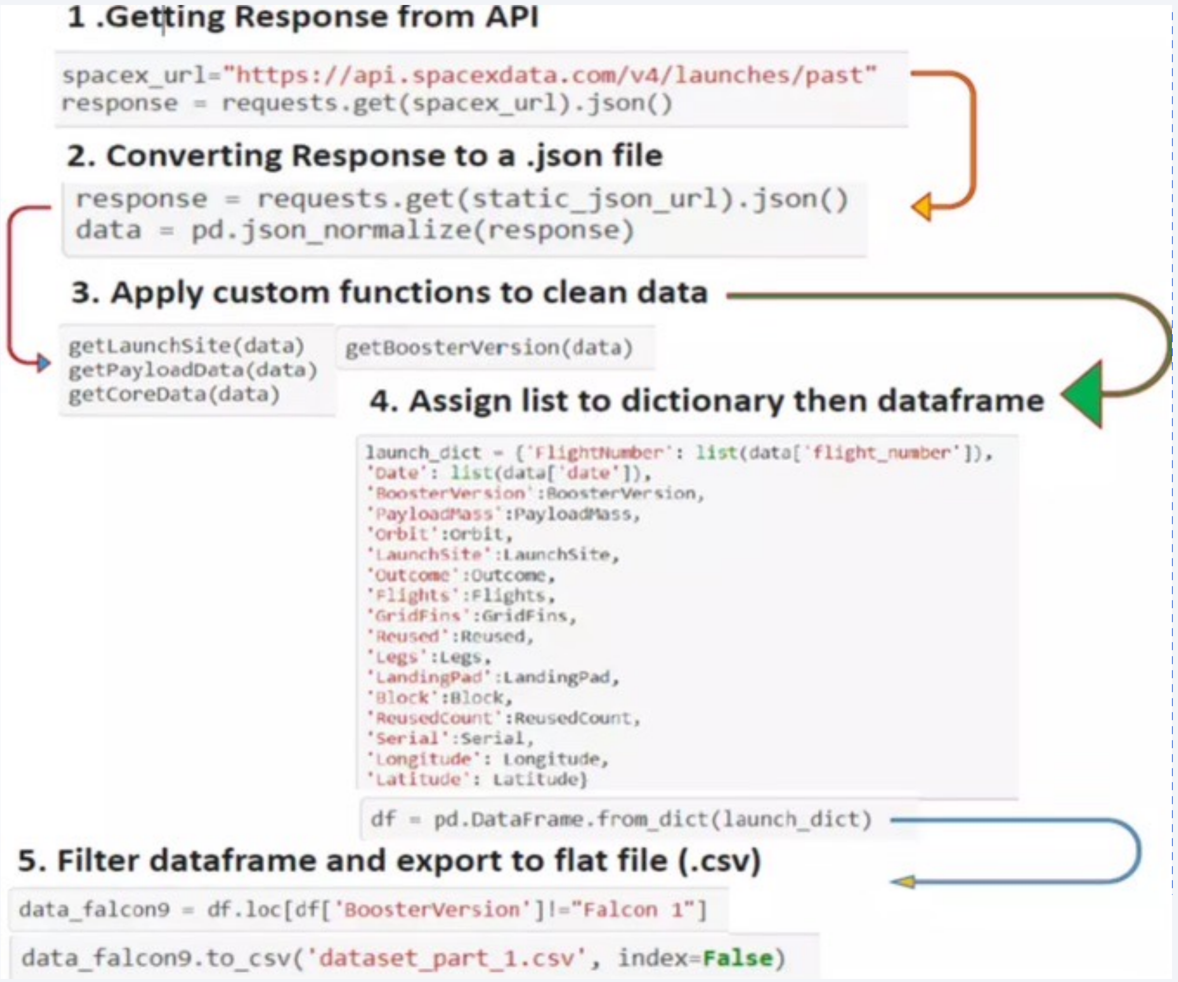
- Data collection methodology:
  - SpaceX Rest API
  - Web Scraping from Wikipedia
- Perform data wrangling
  - Encoding data fields for machine learning and cleaning the data of null values & irrelevant columns
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Use of linear regression, KNN, SVM & DT models

Data Collection

1. SpaceX launch data was gathered from the SpaceX REST API, which provides information on launches, including details about the rocket, payload, launch specifications, landing specifications, and landing outcomes. The SpaceX REST API endpoints begin with api.spacexdata.com/v4/ .
2. Another valuable data source for falcon 9 launch data is web scraping Wikipedia using BeautifulSoup.

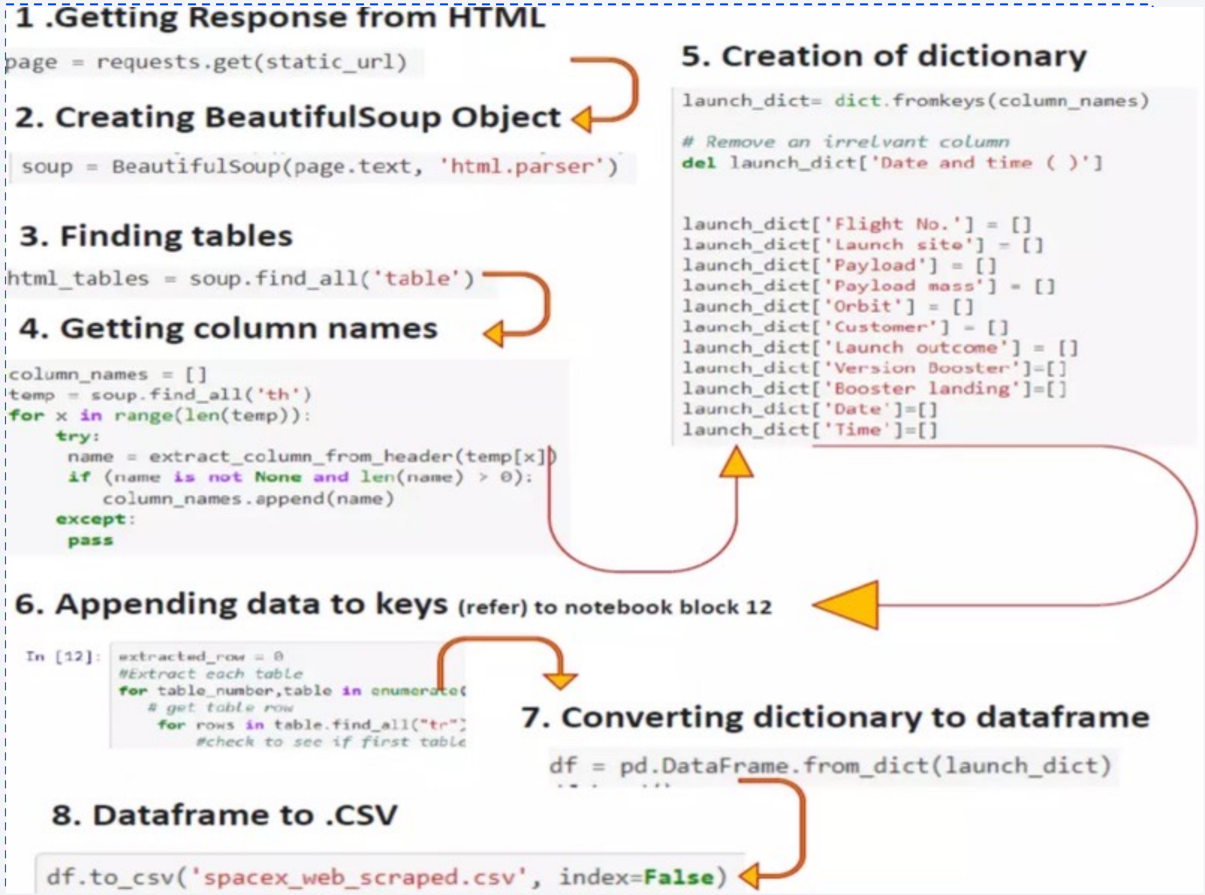
Data Collection – SpaceX API

Data collection using the SpaceX REST API



Data Collection - Scraping

- Data collection using web scraping from Wikipedia

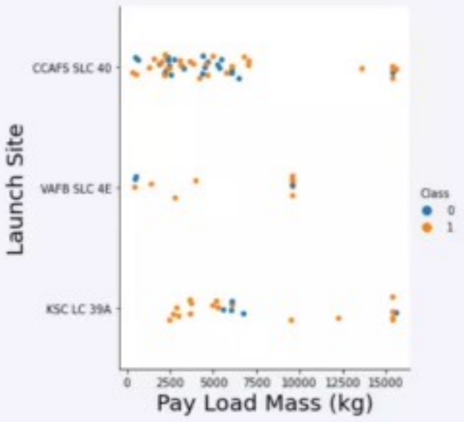


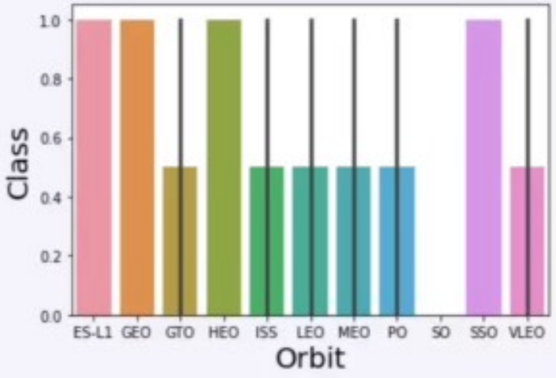
Data Wrangling

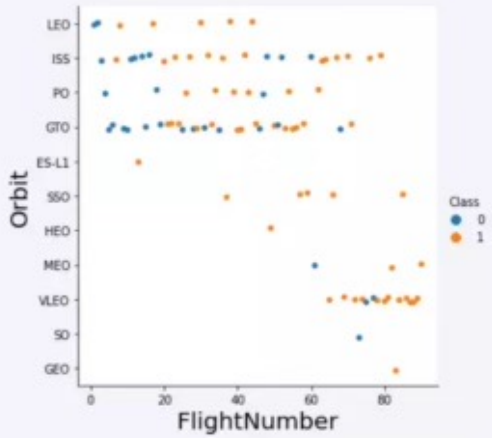
EDA Analysis

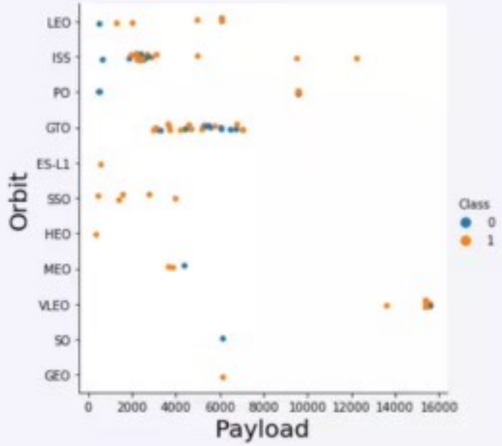
1. Check null values
2. Calculate the number of launches on each site
3. Calculate the number of occurrence of each orbit
4. Calculate the number & occurrence of mission outcome per orbit type
5. Create a landing outcome label from the “Outcome” column
6. Handle null values

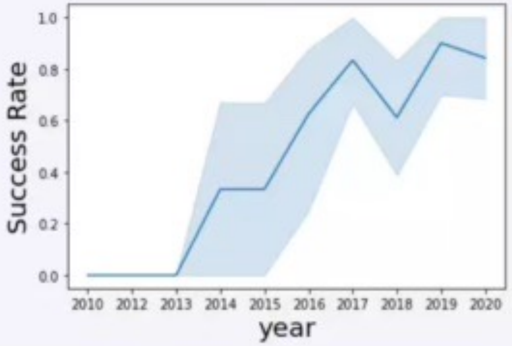
EDA with Data Visualization











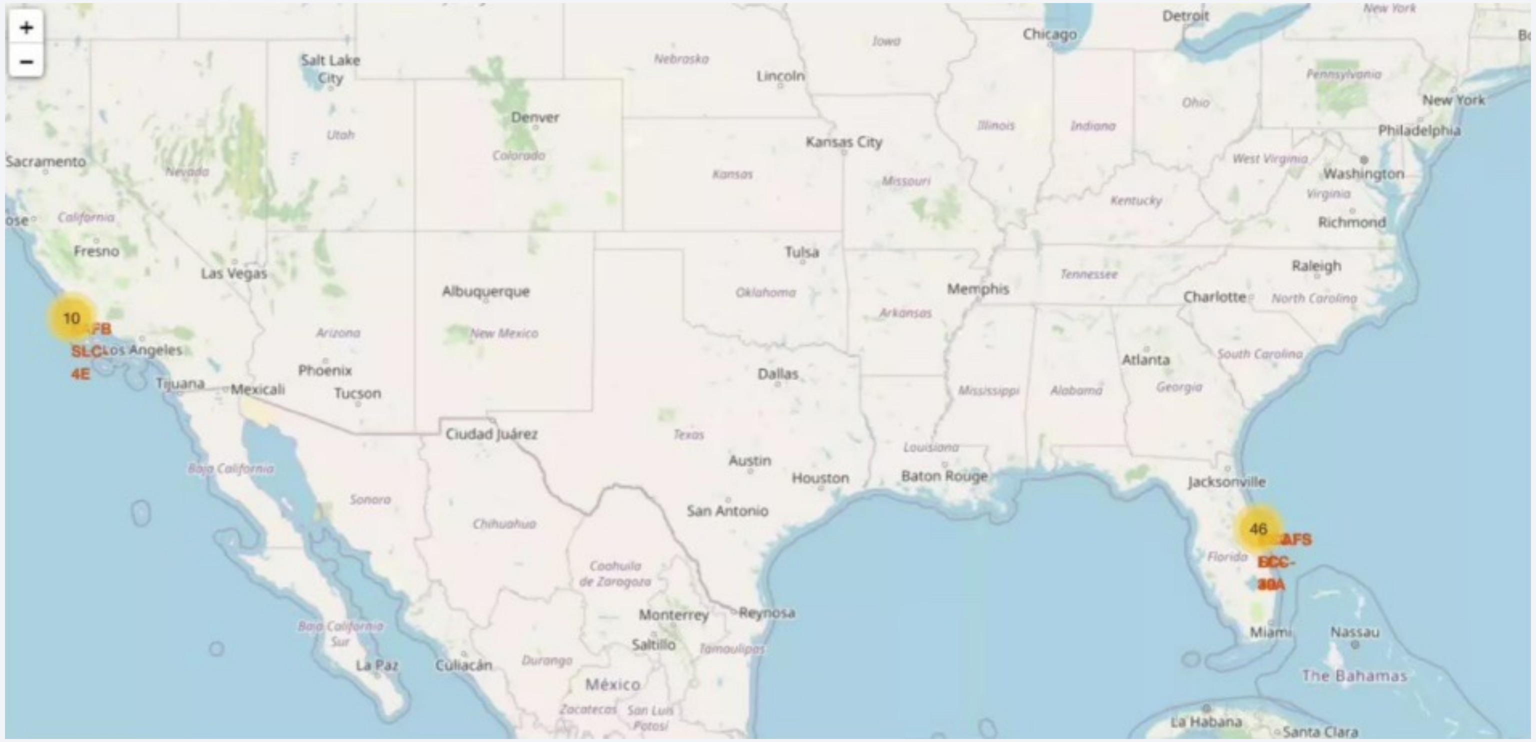
EDA with SQL

SQL Queries Used:

- 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 the average payload mass carried by booster version F9 v1.1.
- Listing the dates where successful landing outcomes on a drone ship were achieved.
- Listing the names of the boosters that succeeded in landing on ground pads and have payload masses greater than 4000 but less than 6000.
- Listing the total number of successful and failed mission outcomes.
- Listing the names of the booster versions that have carried the maximum payload mass.
- Displaying records that show the month names, successful landing outcomes on ground pads, booster versions, and launch sites for the months in the year 2017.
- Ranking the count of successful landing outcomes between the dates 2010-06-04 and 2017-03-20 in descending order.

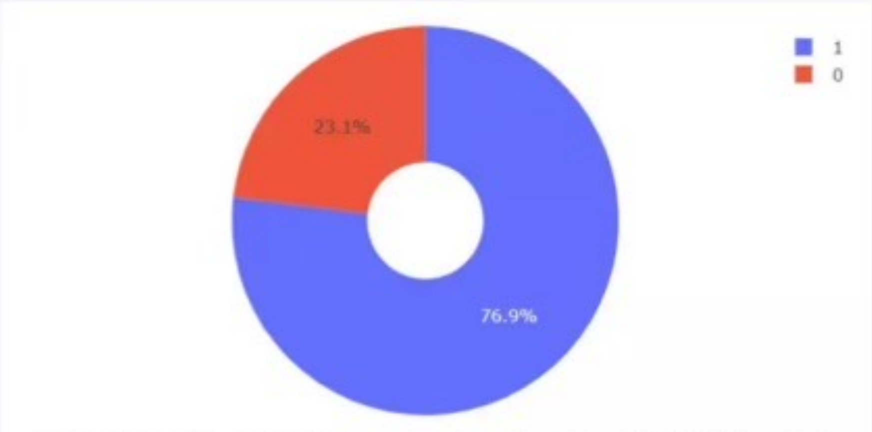
Build an Interactive Map with Folium

These markers were added to the map to show the optimal location for building a rocket launch site.



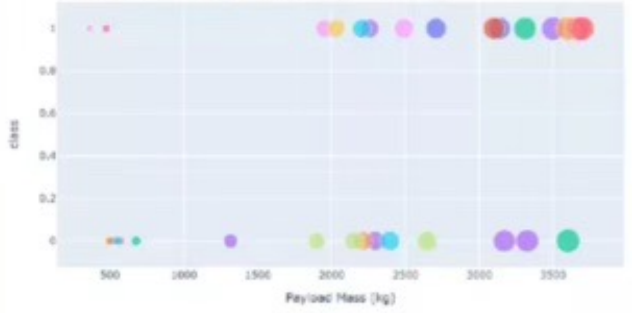
Build a Dashboard with Plotly Dash

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

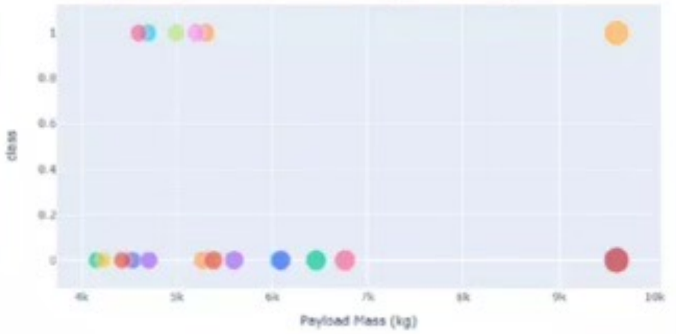


KSC LC-39A achieved a 76.9% success rate while getting a 23.1% failure rate

Low Weighted Payload 0kg – 4000kg



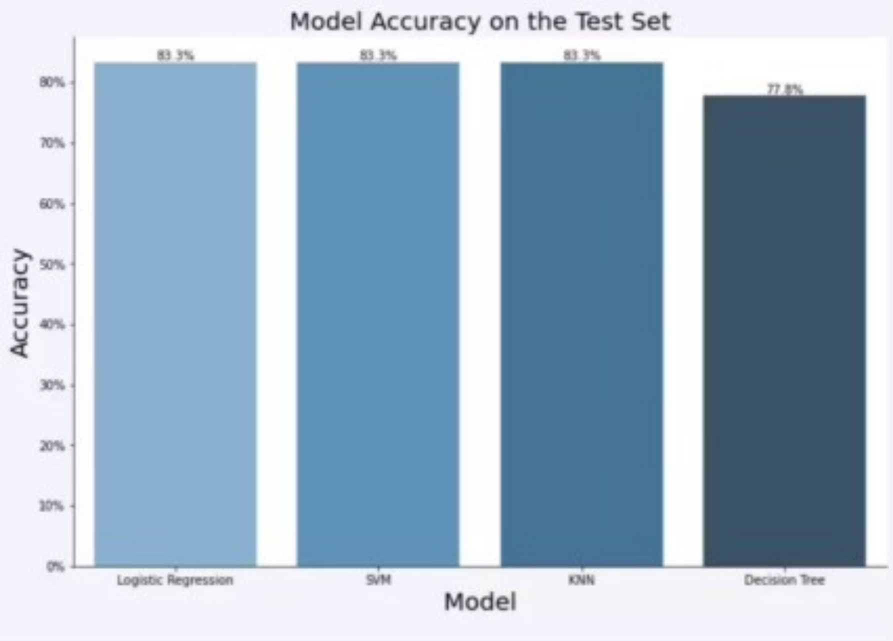
Heavy Weighted Payload 4000kg – 10000kg

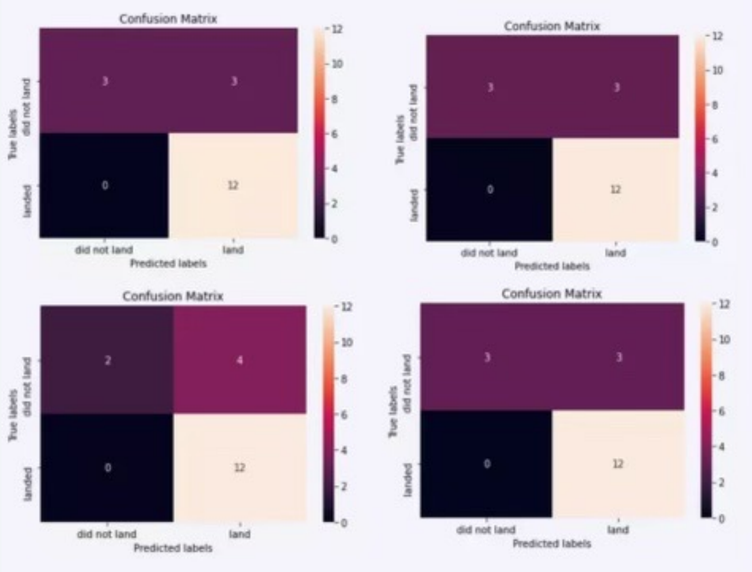


We can see the success rates for low weighted payloads is higher than the heavy weighted payloads

Predictive Analysis (Classification)

After testing multiple predictive analysis models, the logistic regression, SVM & KNN models achieved the highest accuracy with 83.3%. Furthermore, the SVM performs better according ot the area under the curve with 0.96.





Results

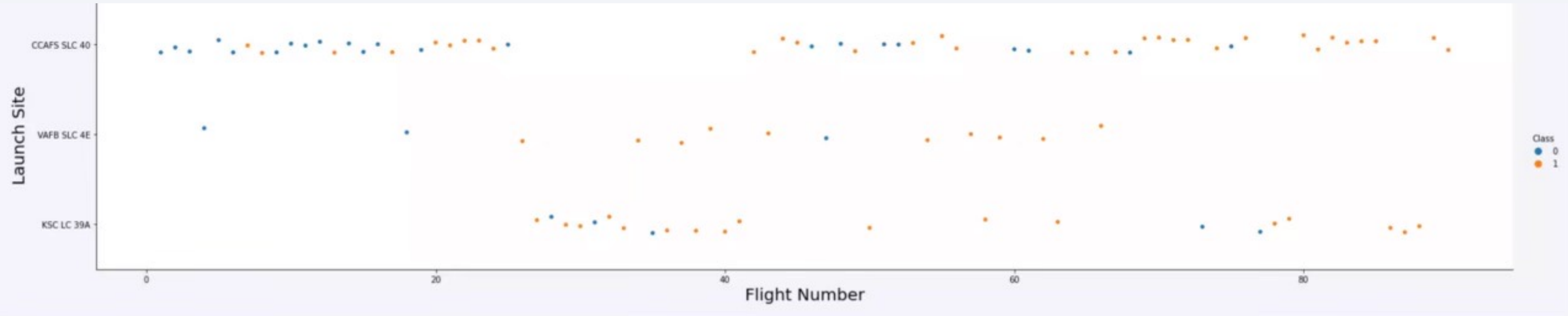
- 68% of retail investor accounts lose money when trading CFDs with this provider. You should consider whether you can afford the high risk of losing your money. Your capital is at risk.
- The SVM, KNN, and Logistic Regression models demonstrate the highest prediction accuracy for this dataset.
- Lower-weighted payloads perform better than heavier ones.
- The success rates for SpaceX launches increase over time as they perfect their techniques.
- KSC LC 39A had the most successful launches among all sites.
- Orbits GEO, HEO, SSO, and ES L1 have the highest success rates.

Section 2

Insights drawn from EDA

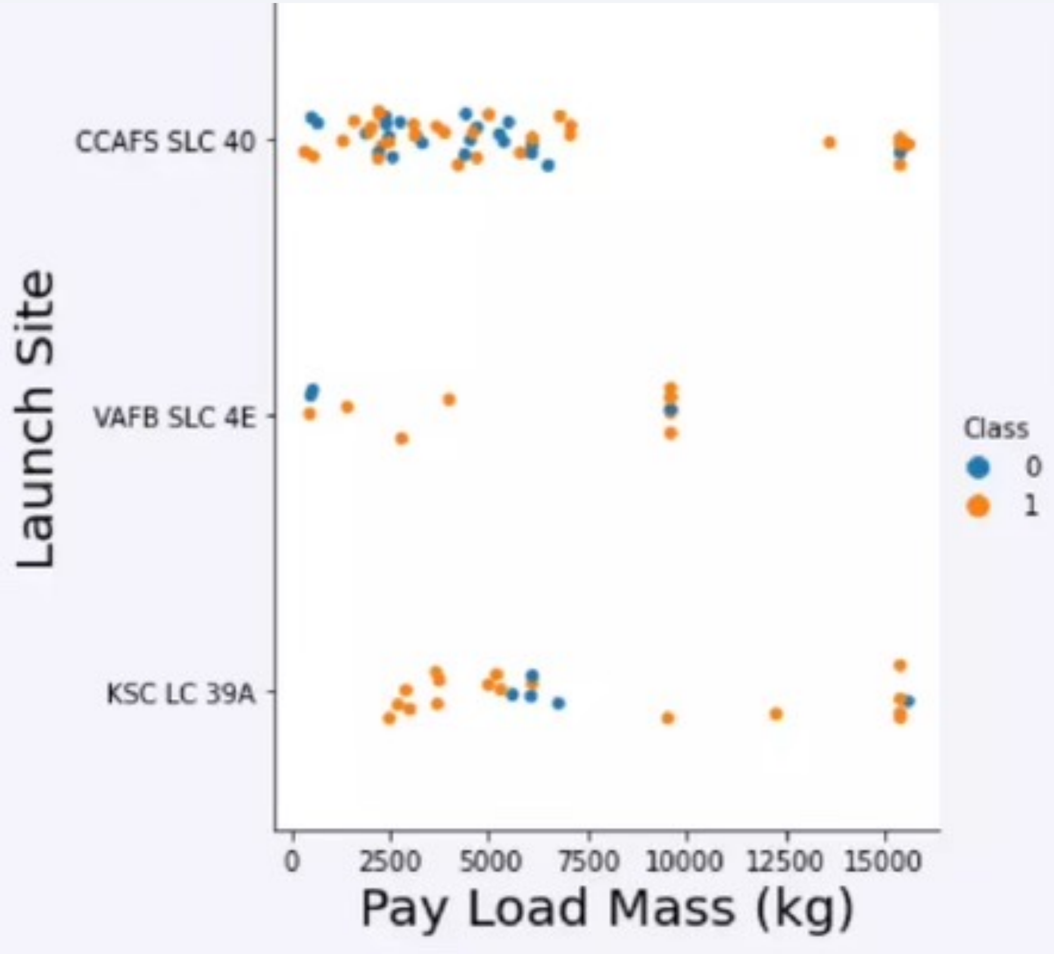
Flight Number vs. Launch Site

Rocket launches from the CCAFS SLC 40 site are significantly higher compared to other launch sites.



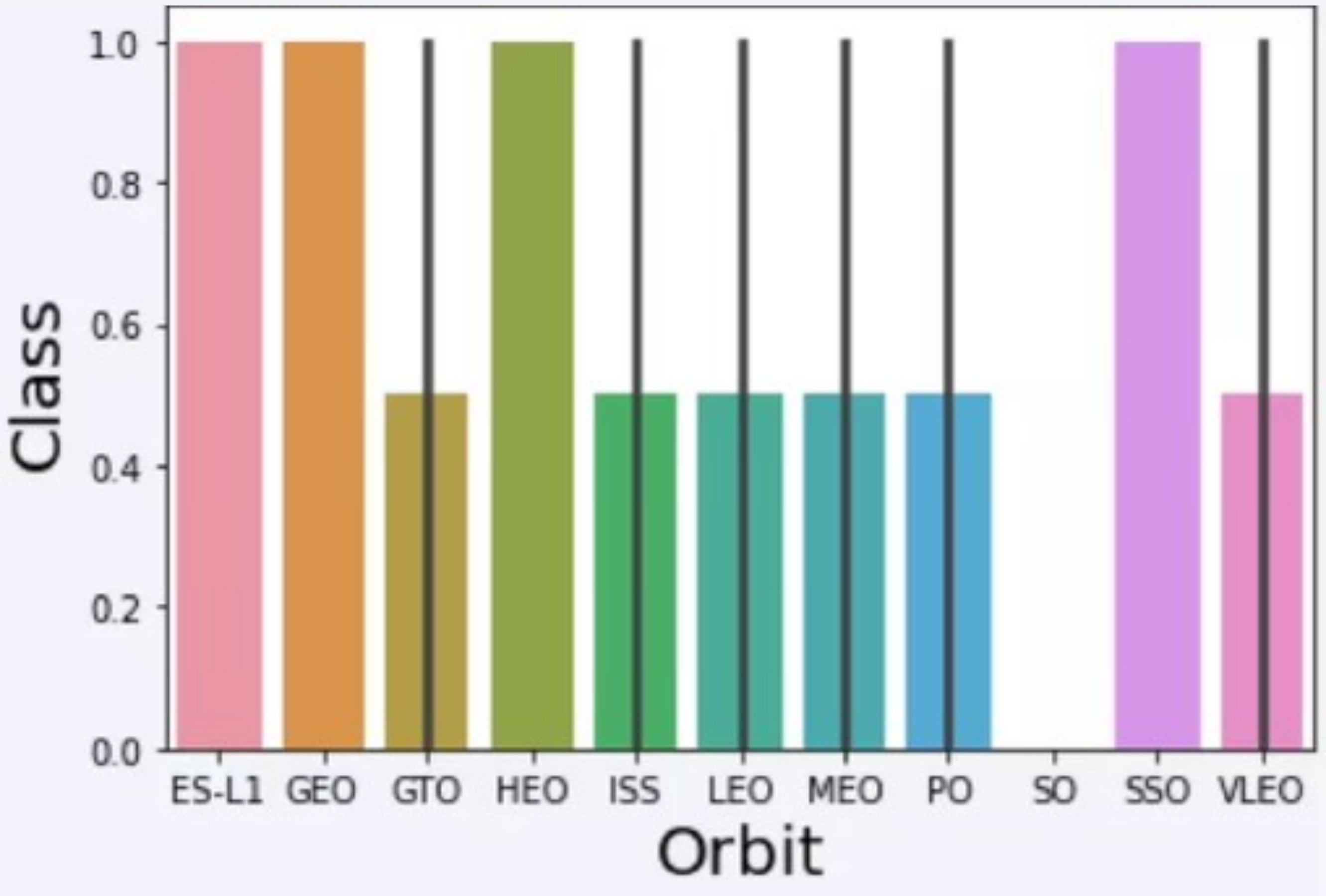
Payload vs. Launch Site

Most pay loads with a lower mass have been launched from CCAFS SLC 40.



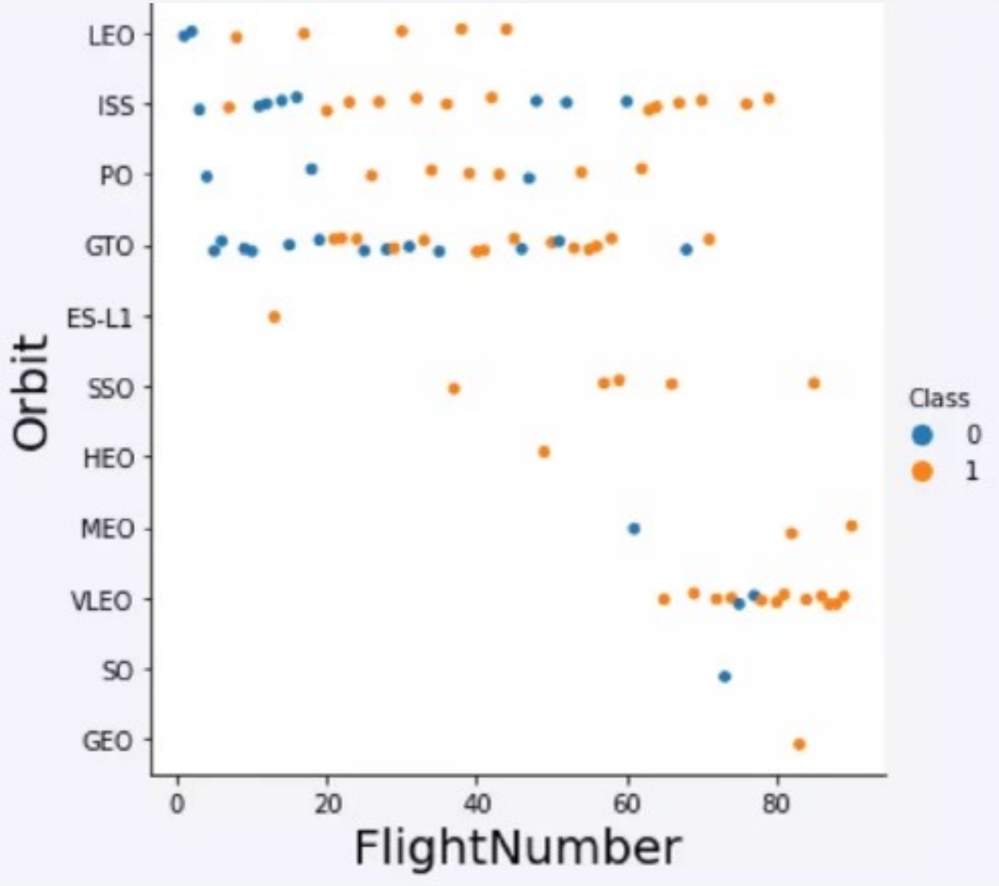
Success Rate vs. Orbit Type

The ES-L1, GEO, HEO and SSO orbit types have the highest success rates.



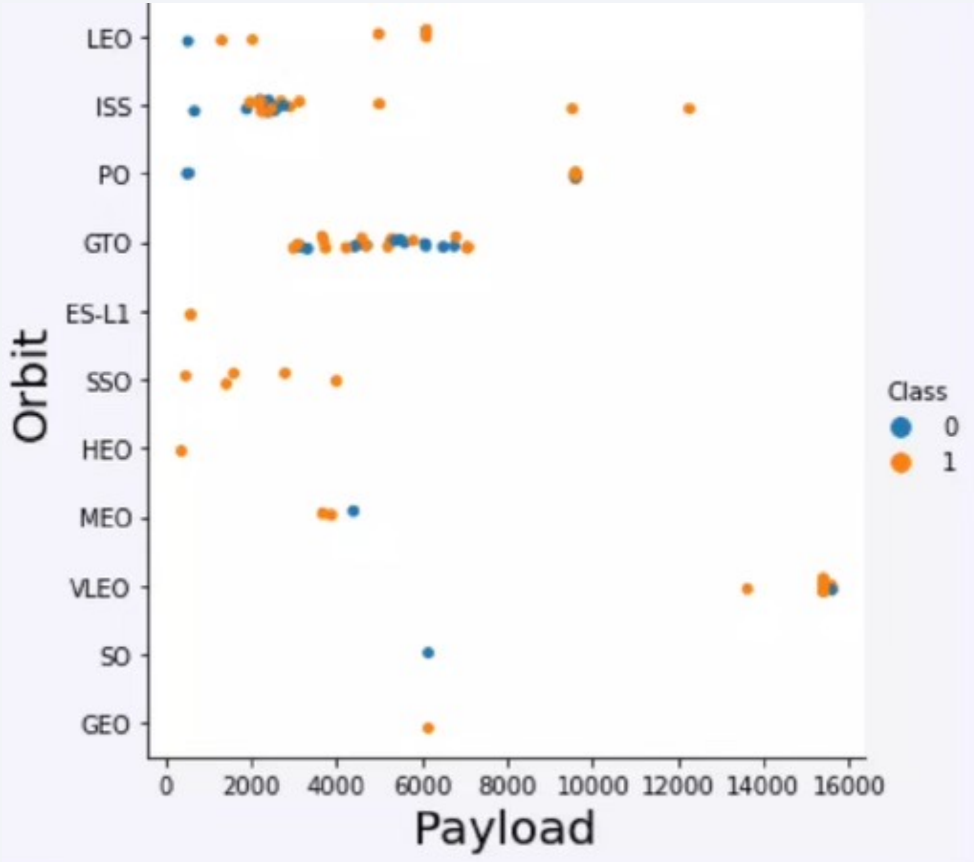
Flight Number vs. Orbit Type

There is a trend of launches shifting to VLEO recently.



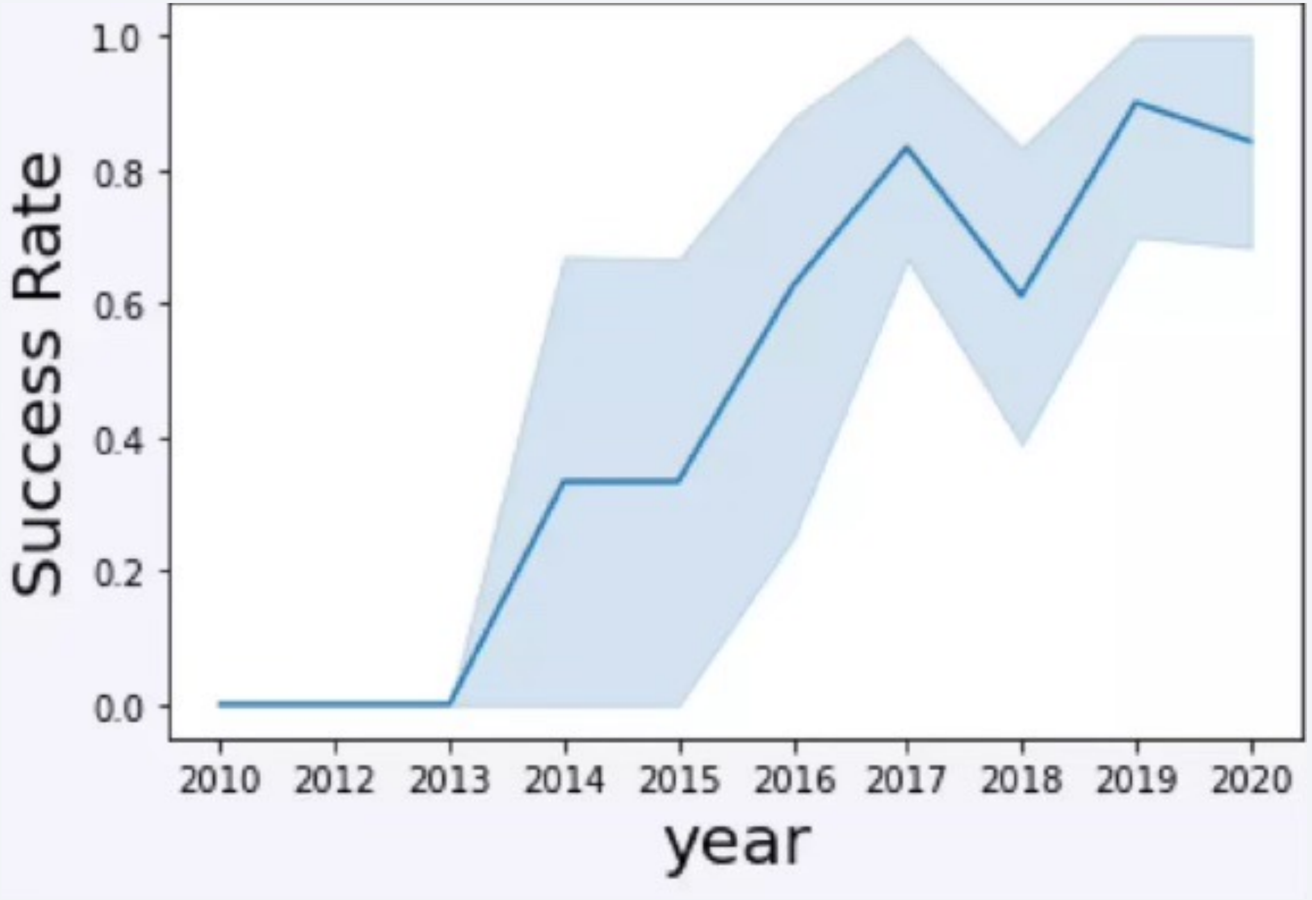
Payload vs. Orbit Type

Around 2000, there is a high correlation with ISS and payload. Likewise, there is a high correlation between GTO and payload throughout 4000 to 8000.



Launch Success Yearly Trend

Successful launches has significantly increased since 2013 and has started to slowly stabalised since 2019, which could be due to advances in technology.



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’

2928.4

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)' 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 and 2017-03-20

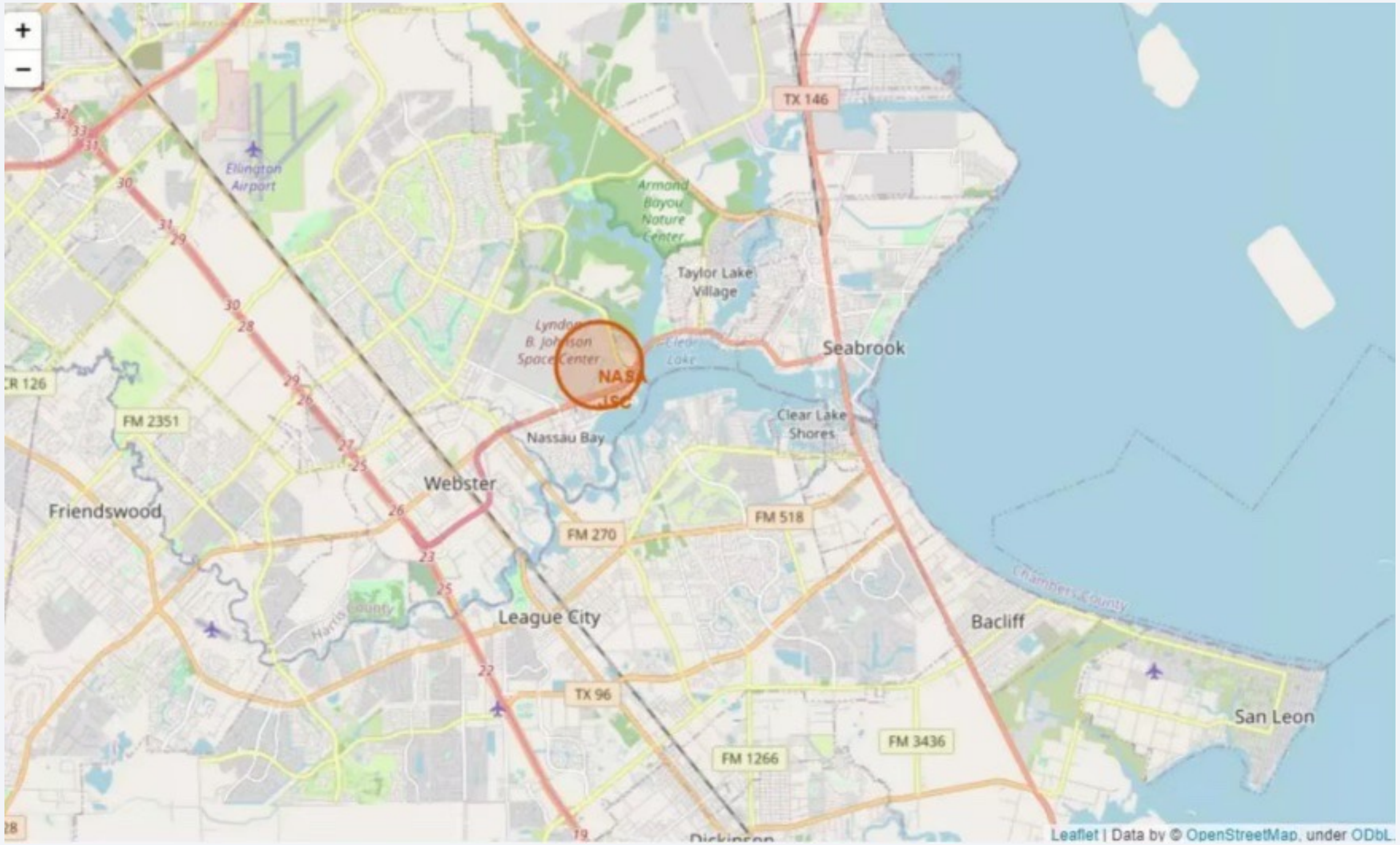
%sql select * from SPACEXTBL where Landing_Outcome like 'Success%' and (DATE between '2010-06-04' and '2017-03-20') order by date desc

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)

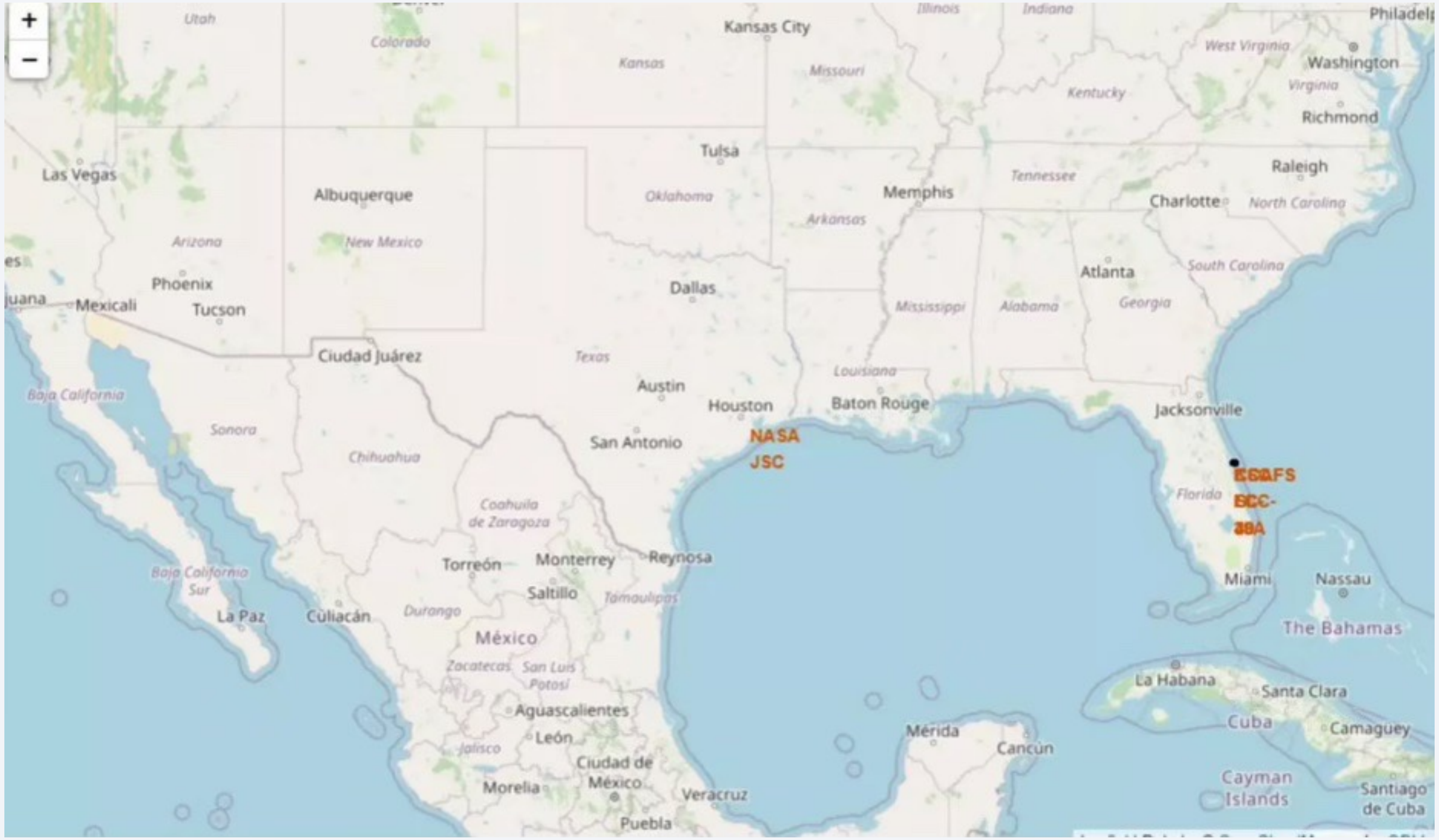
Section 3

Launch Sites Proximities Analysis

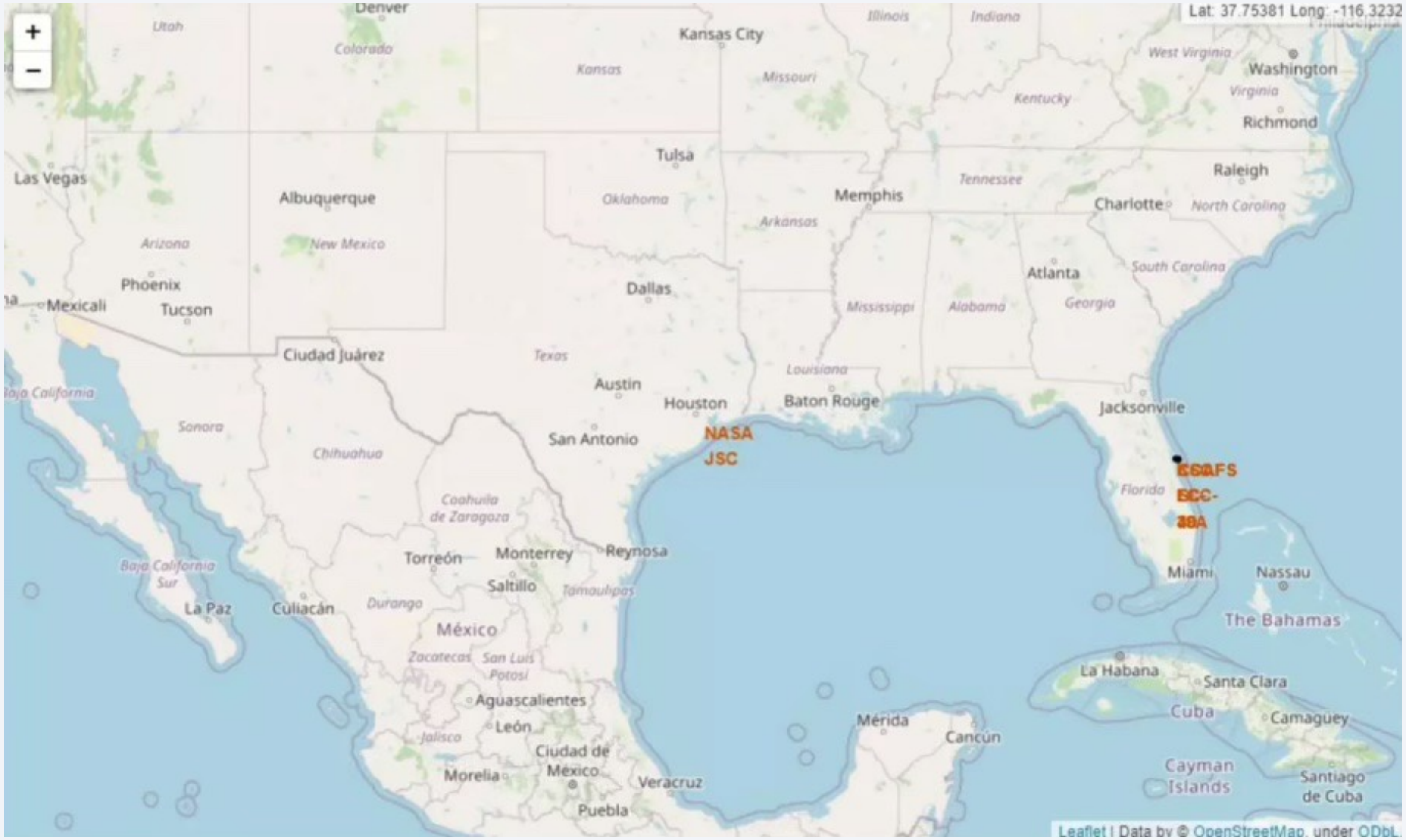
Launch Sites Map



Successful and Failed Launches Map



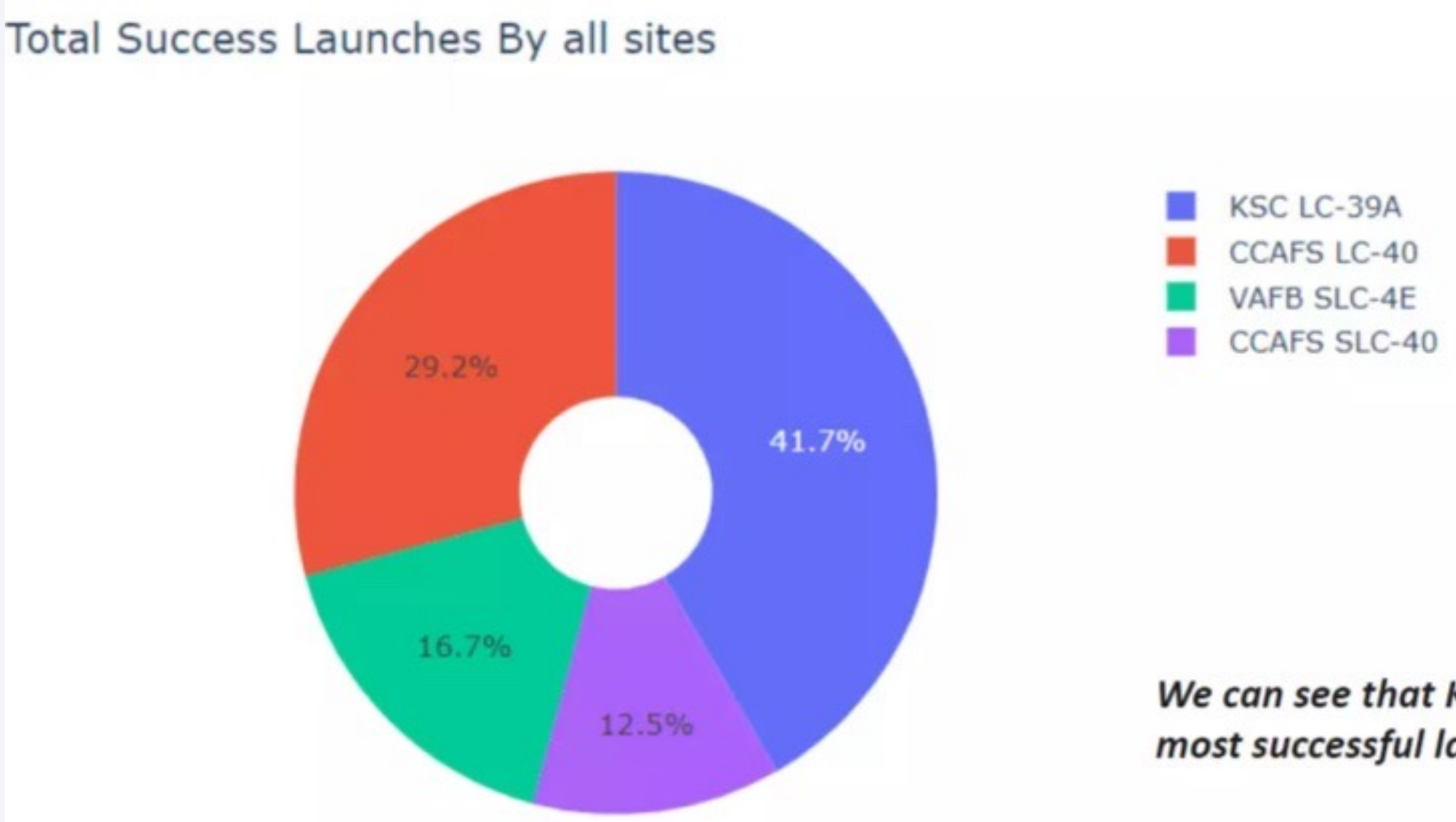
Distance between each Launch Site Map



Section 4

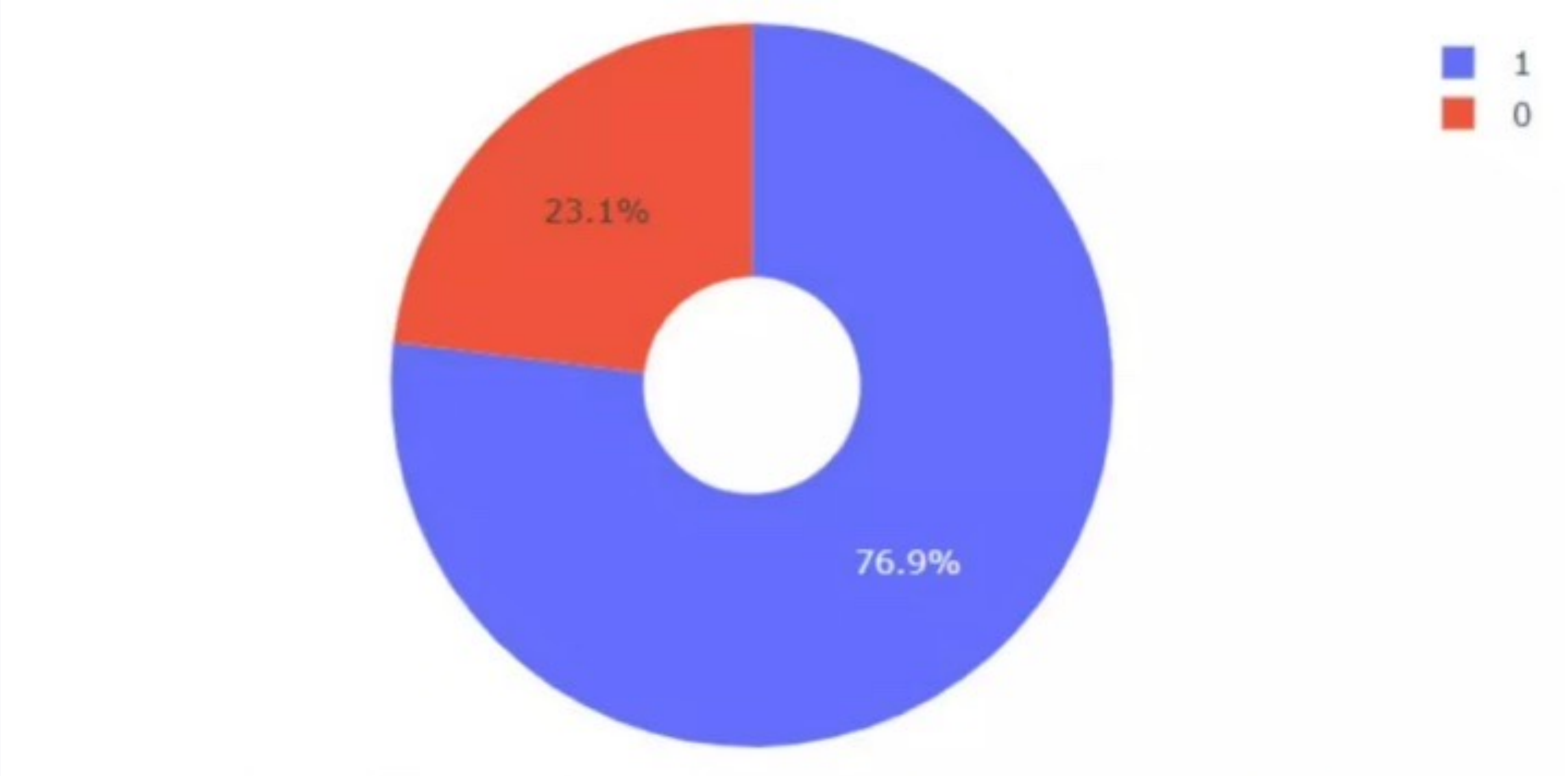
Build a Dashboard with Plotly Dash

Successful Launches according to Site



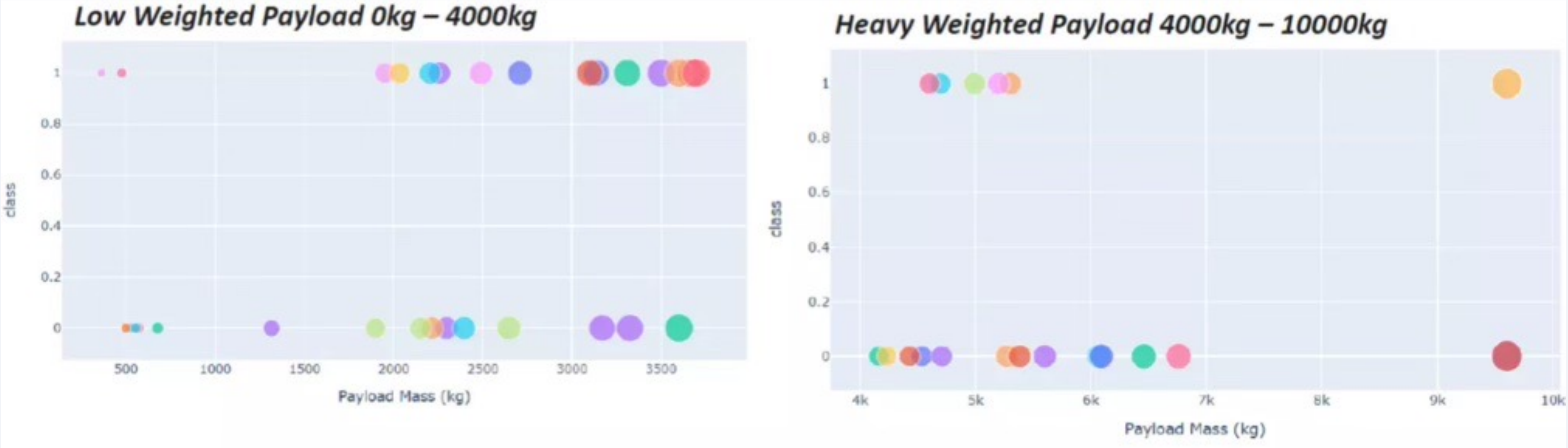
We can see that KSC LC-39A had the most successful launches from all the sites

Most Successful Launch Site



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

Payload and Launch Outcomes

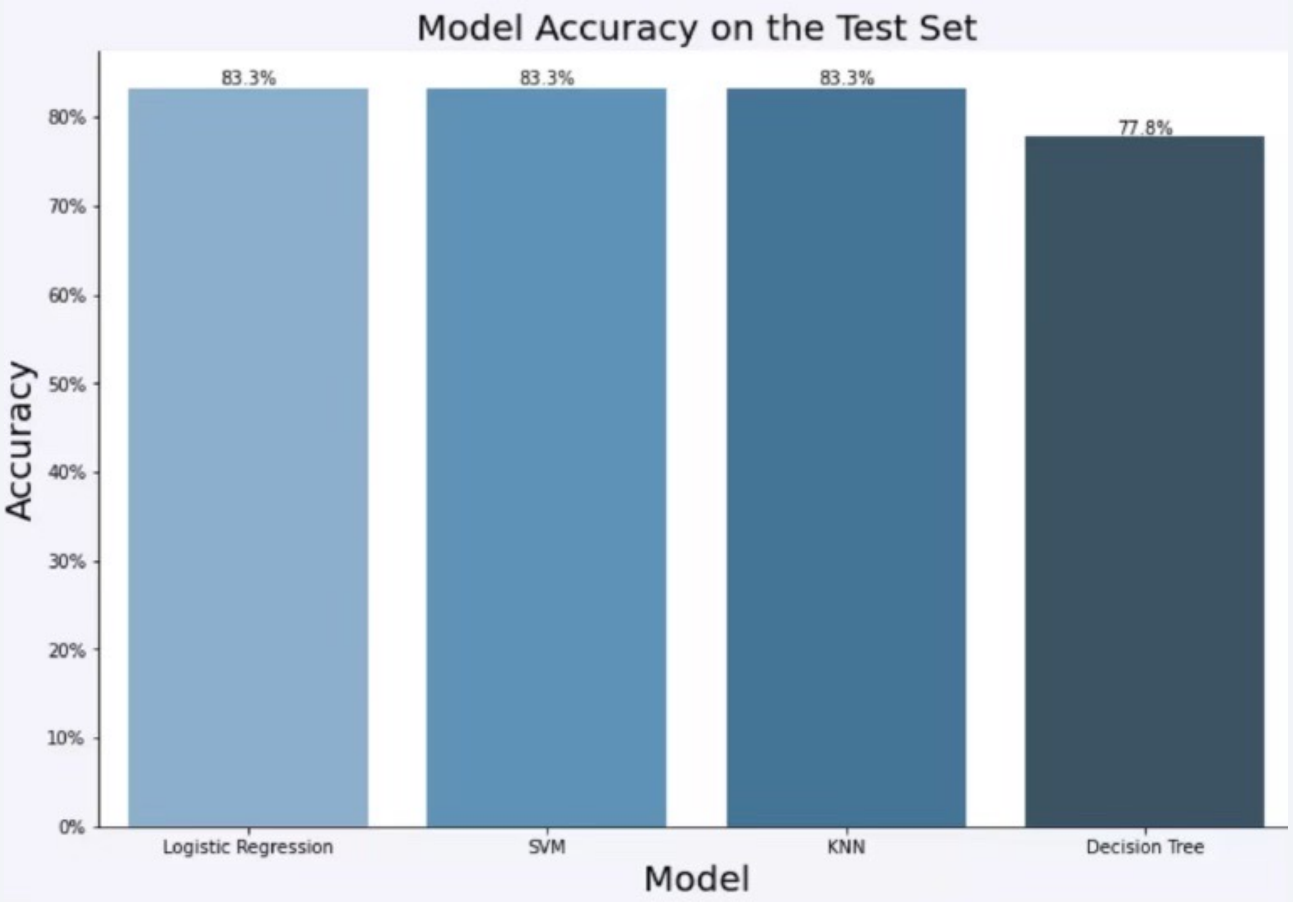


We can see the success rates for low weighted payloads is higher than the heavy weighted payloads

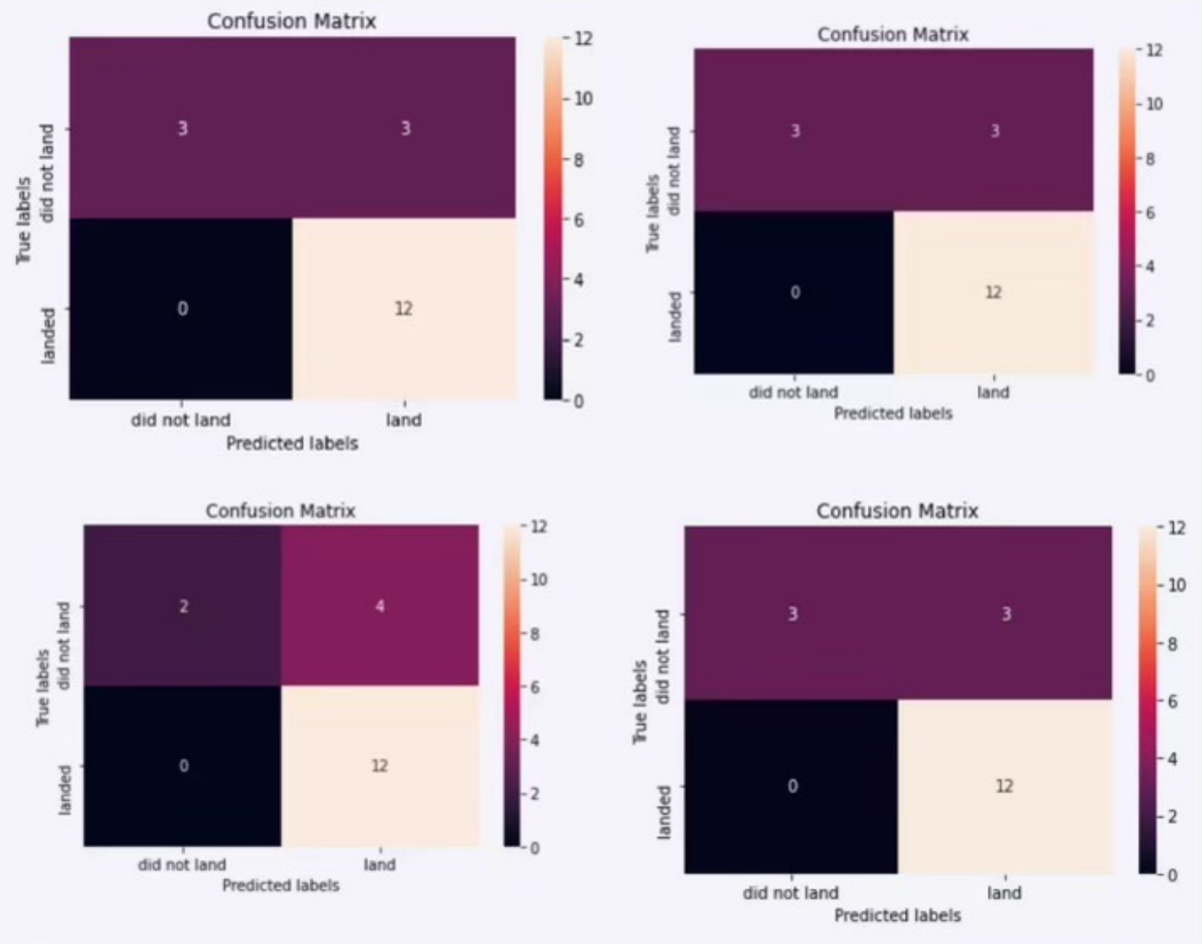
Section 5

Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix



Conclusions

- The SVM, KNN, and Logistic Regression models provide the highest prediction accuracy for this dataset.
- Lower-weighted payloads perform better than heavier ones.
- The success rates for SpaceX launches improve over time as they refine their processes.
- KSC LC 39A has the highest number of successful launches among all sites.
- Orbits GEO, HEO, SSO, and ES L1 have the highest success rates.

Thank you!