IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Karthick Raja B
4-11-23

Outline

- Executive Summary
- Introduction
- Methodology
- Results
  - EDA with visualization
  - EDA with SQL
  - Interactive maps with folium
  - Plotly Dash dashboard
  - Predictive analysis
- Conclusion

Executive Summary

Summary of methodologies

- The research attempts to identify the factors for a successful rocket landing. To make this determination, the following methodologies where used:
- **Collect** data using SpaceX REST API and web scraping techniques
- **Wrangle** data to create success/fail outcome variable
- **Explore** data with data visualization techniques, considering the following factors: payload, launch site, flight number and yearly trend
- **Analyze** the data with SQL, calculating the following statistics: total payload, payload range for successful launches, and total # of successful and failed outcomes
- **Explore** launch site success rates and proximity to geographical markers
- **Visualize** the launch sites with the most success and successful payload ranges
- **Build Models** to predict landing outcomes using logistic regression, support vector machine (SVM), decision tree and K-nearest neighbor (KNN)

Result

- **Exploratory Data Analysis:**
  - **Temporal Trends:** Success rates in rocket landings have shown improvement over time.
  - **Leading Landing Site:** KSC LC-39A stands out with the highest success rate among landing sites.

- **Orbital Analysis:**
  - Certain orbits like ES-L1, GEO, HEO, and SSO exhibit a 100% success rate in rocket landings.

- **Visualization and Geography:**
  - Launch sites tend to be near the equator and close to coastal areas.

- **Predictive Analytics:**
  - All models performed similarly on the test set. The decision tree model showed slightly better performance in predicting successful landings.

Introduction

Background

SpaceX, a leader in the space industry, strives to make space travel affordable for everyone. Its accomplishments include sending spacecraft to the international space station, launching a satellite constellation that provides internet access and sending manned missions to space. SpaceX can do this because the rocket launches are relatively inexpensive ($62 million per launch) due to its novel reuse of the first stage of its Falcon 9 rocket. Other providers, which are not able to reuse the first stage, cost upwards of $165 million each. By determining if the first stage will land, we can determine the price of the launch. To do this, we can use public data and machine learning models to predict whether SpaceX - or a competing company - can reuse the first stage.

Explore

- How payload mass, launch site, number of flights, and orbits affect first-stage landing success
- Rate of successful landings over time .
- Best predictive model for successful landing (binary classification)

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Collect data using SpaceX REST API and web scraping techniques
- Perform data wrangling
  - Wrangle data - by filtering the data, handling missing values and applying one hot encoding - to prepare the data for analysis and modeling
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Build Models to predict landing outcomes using classification models. Tune and evaluate models to find best model and parameters

Data Collection – SpaceX API

- Data collection with SpaceX REST calls.

https://github.com/krb647/SpaceX_Falcon9_Landing_Prediction/blob/main/SpaceX_Data_Collection_API.ipynb

1 .Getting Response from API

```
spacex_url="https://api.spacexdata.com/v4/launches/past"
response = requests.get(spacex_url).json()
```

2. Converting Response to a .json file

```
response = requests.get(static_json_url).json()
data = pd.json_normalize(response)
```

3. Apply custom functions to clean data

```
getLaunchSite(data)
getPayloadData(data)
getCoreData(data)
```

```
getBoosterVersion(data)
```

4. Assign list to dictionary then dataframe

```
launch_dict = {'FlightNumber': list(data['flight_number']),
'Date': list(data['date']),
'BoosterVersion':BoosterVersion,
'PayloadMass':PayloadMass,
'Orbit':Orbit,
'LaunchSite':LaunchSite,
'Outcome':Outcome,
'Flights':Flights,
'GridFins':GridFins,
'Reused':Reused,
'Legs':Legs,
'LandingPad':LandingPad,
'Block':Block,
'ReusedCount':ReusedCount,
'Serial':Serial,
'Longitude': Longitude,
'Latitude': Latitude}
```

```
df = pd.DataFrame.from_dict(launch_dict)
```

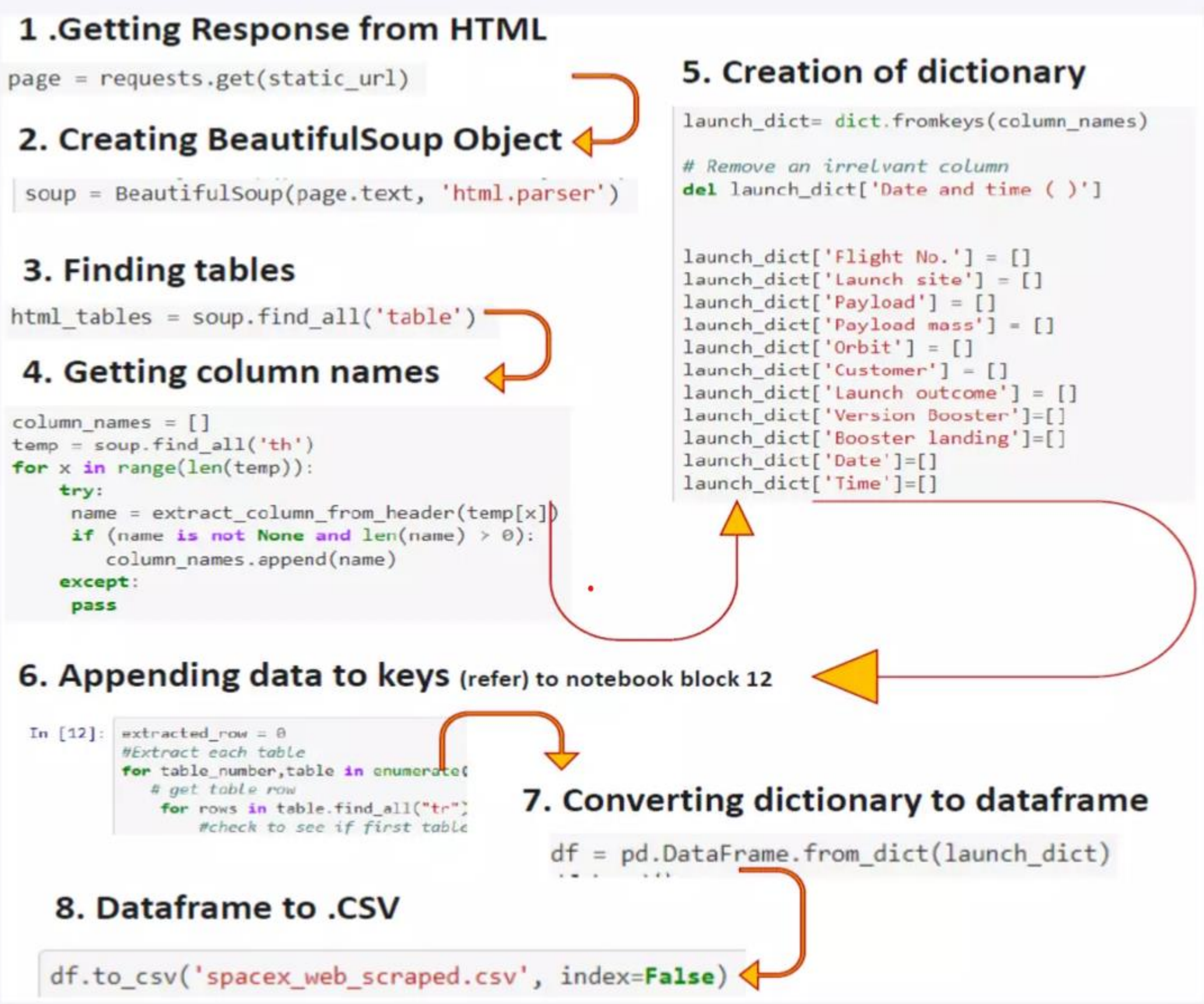
5. Filter dataframe and export to flat file (.csv)

```
data_falcon9 = df.loc[df['BoosterVersion']!="Falcon 1"]
```

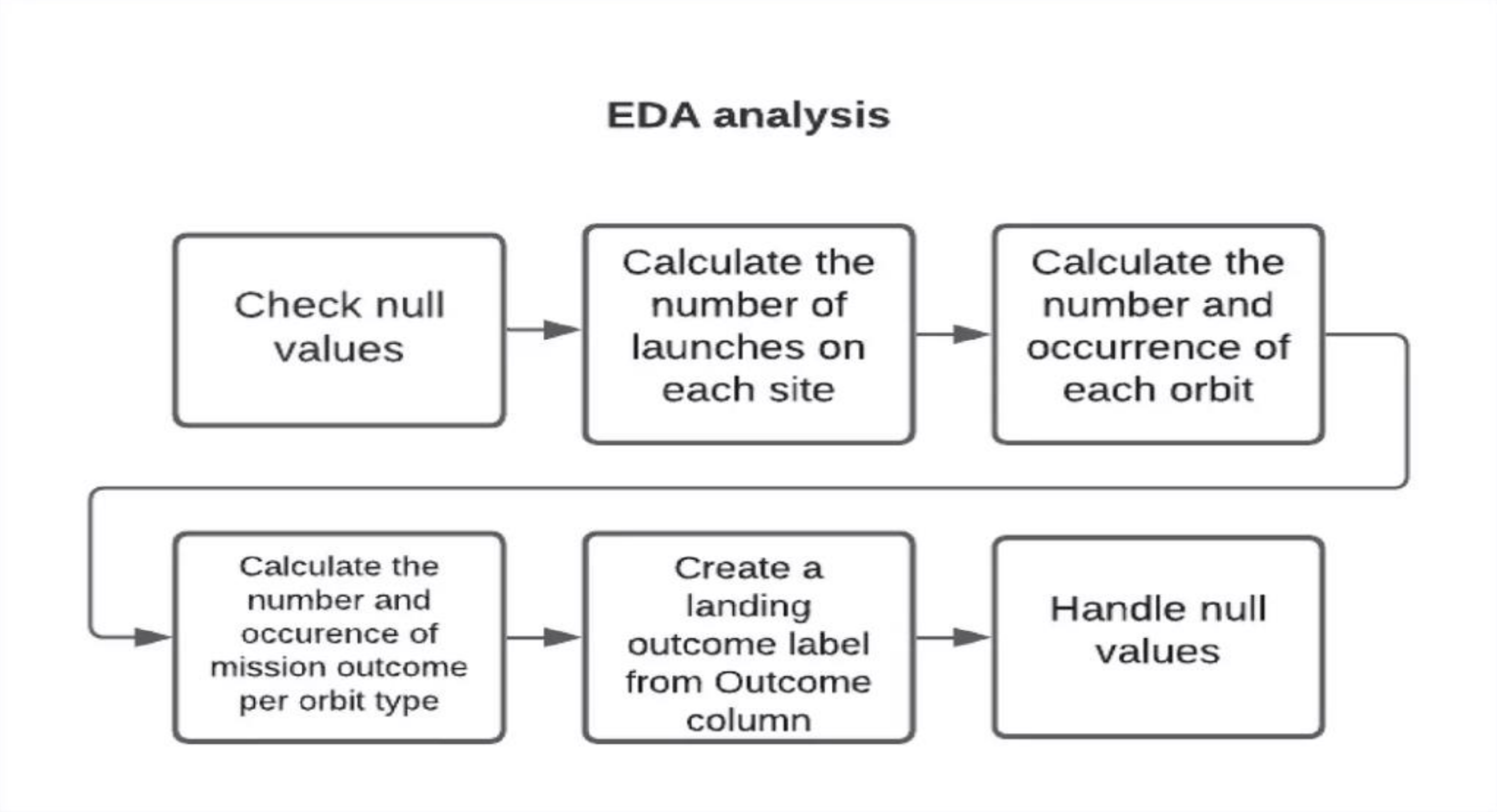
```
data_falcon9.to_csv('dataset_part_1.csv', index=False)
```

Data Collection - Scraping

- web scraping process using key phrases and flowcharts
- https://github.com/krb647/SpaceX_Falcon9_Landing_Prediction/blob/main/SpaceX_Data_Collection_WebScraping.ipynb

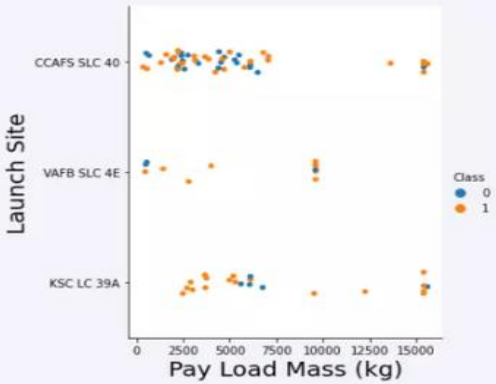


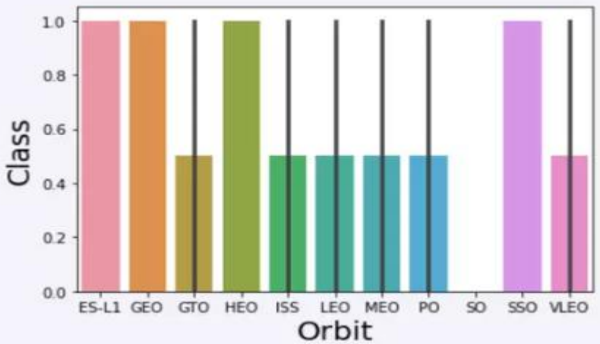
Data Wrangling

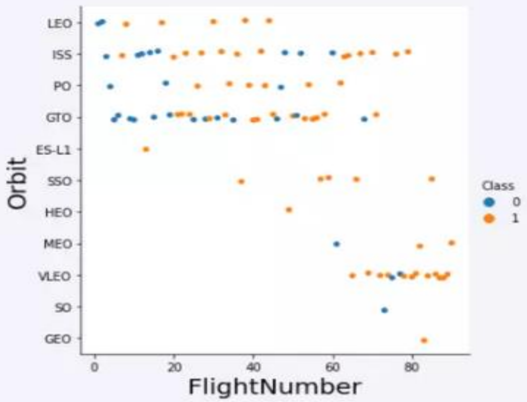


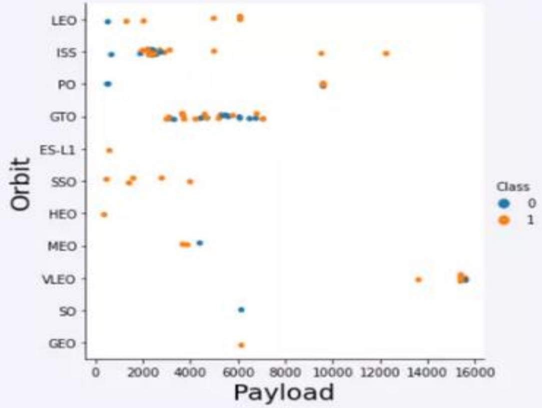
https://github.com/krb647/SpaceX_Falcon9_Landing_Prediction/blob/main/SpaceX_Data_Wrangling.ipynb

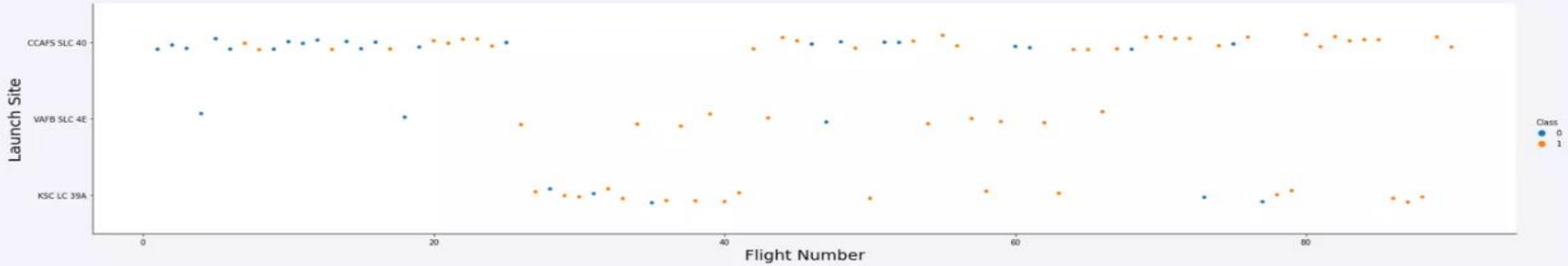
EDA with Data Visualization

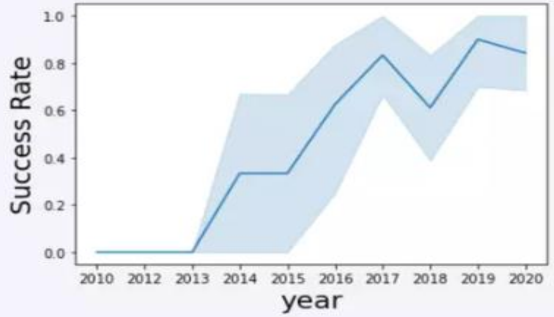












https://github.com/krb647/SpaceX_Falcon9_Landing_Prediction/blob/main/SpaceX_EDA_with_Data_Visualization.ipynb

EDA with SQL

Queries

Display:

• Names of unique launch sites

• 5 records where launch site begins with 'CCA'

• Total payload mass carried by boosters launched by NASA (CRS)

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

List:

• Date of first successful landing on ground pad

• Names of boosters which had success landing on drone ship and have payload mass greater than 4,000 but less than 6,000

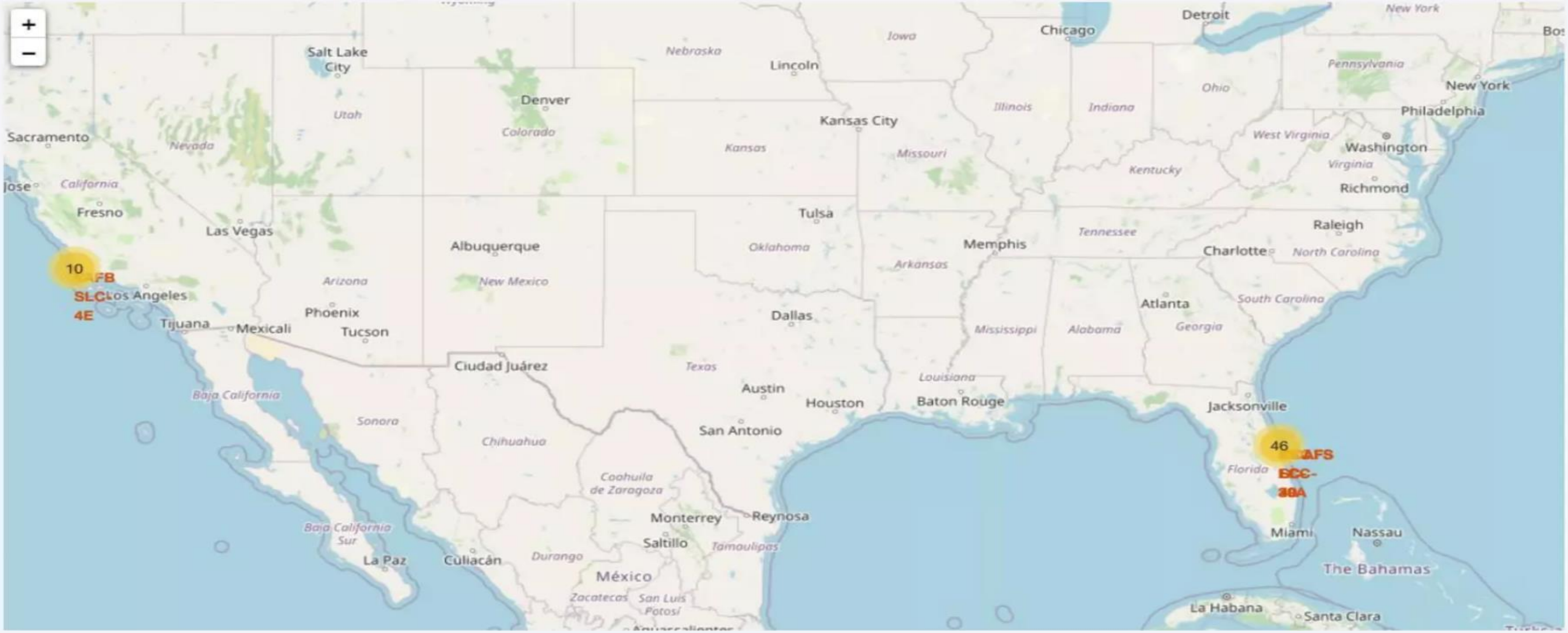
• Total number of successful and failed missions

• Names of booster versions which have carried the max payload Failed landing outcomes on drone ship, their booster version and launch site for the months in the year 2015

• Count of landing outcomes between 2010-06-04 and 2017-03-20 (desc) 2023

https://github.com/krb647/SpaceX_Falcon9_Landing_Prediction/blob/main/SpaceX_EDA_with_SQL.ipynb

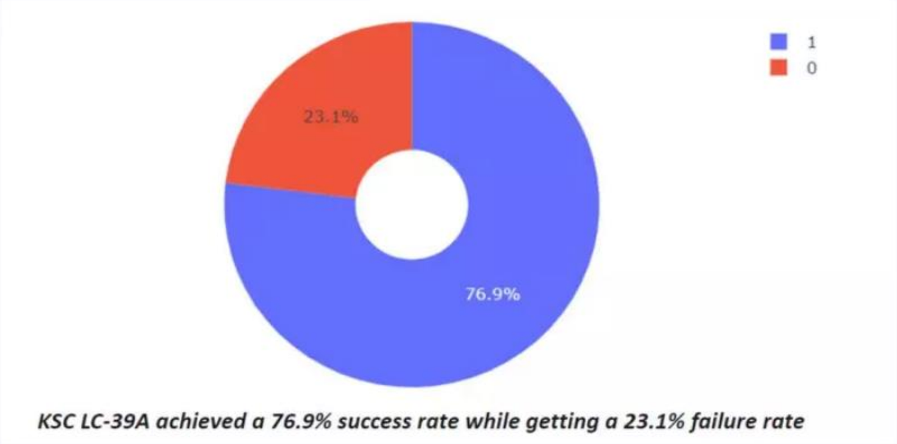
Build an Interactive Map with Folium

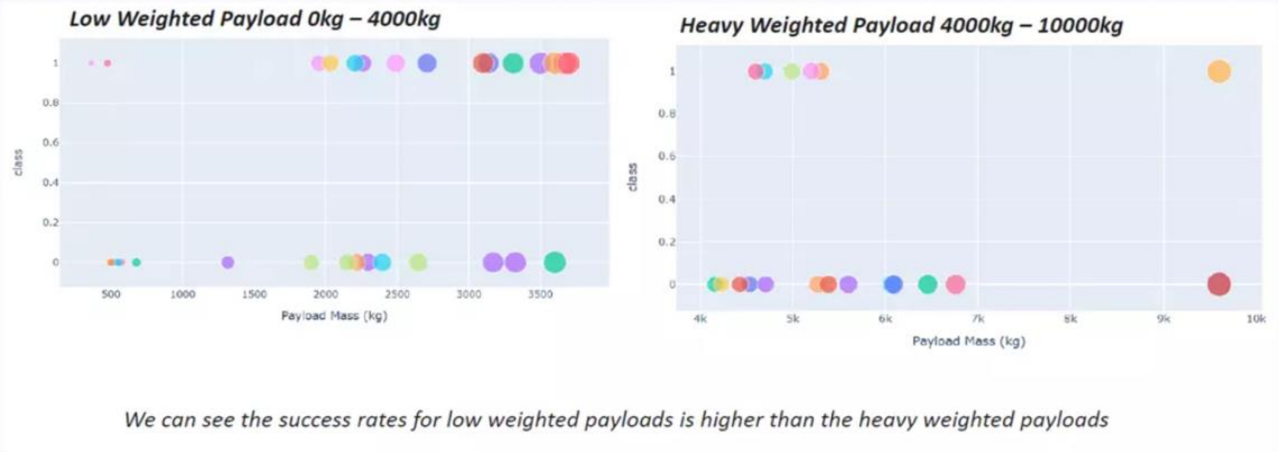


Map marker have been added to the map with aim to finding an optimal location for building a launch site

https://github.com/krb647/SpaceX_Falcon9_Landing_Prediction/blob/main/SpaceX_Interactive_Visual_Analytics_Folium.ipynb

Build a Dashboard with Plotly Dash

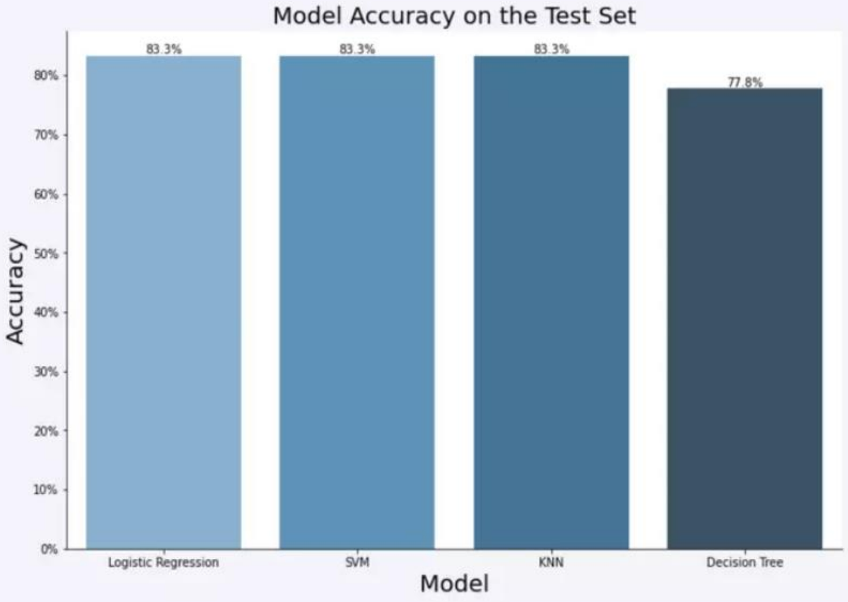


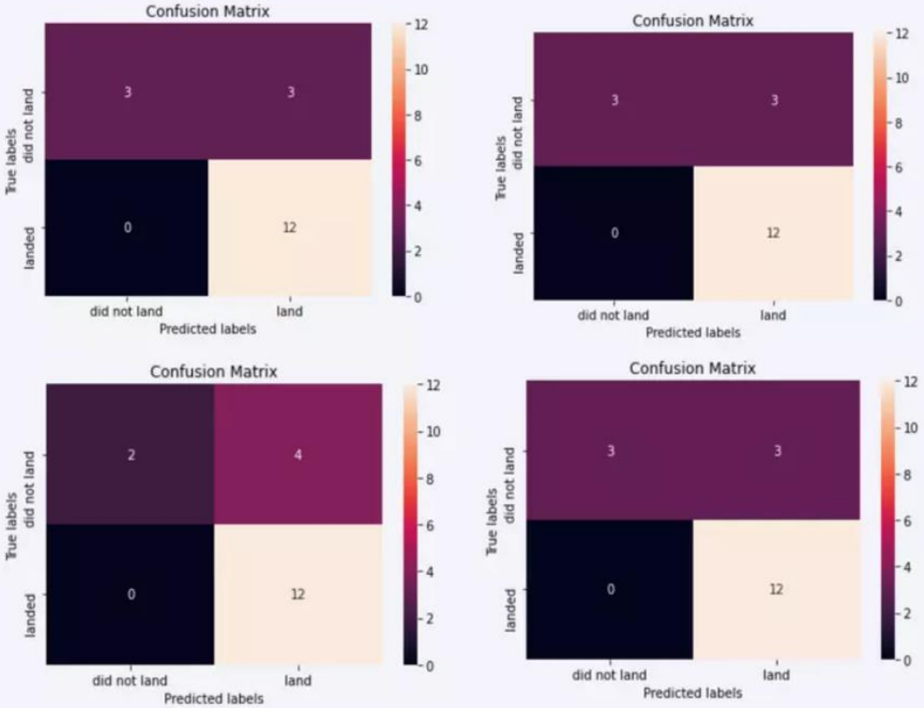


https://github.com/krb647/SpaceX_Falcon9_Landing_Prediction/blob/main/SpaceX_Interactive_Visual_Analytics_with_Plotly_Dash.py

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/krb647/SpaceX_Falcon9_Landing_Prediction/blob/main/SpaceX_Predictive_Analytics.ipynb

Results

Exploratory Data Analysis

- Launch success has improved over time
- KSC LC-39A has the highest success rate among landing sites Orbits ES-L1, GEO, HEO and SSO have a 100% success rate

Visual Analytics

- Most launch sites are near the equator, and all are close to the coast
- Launch sites are far enough away from anything a failed launch can damage (city, highway, railway), while still close enough to bring people and material to support launch activities

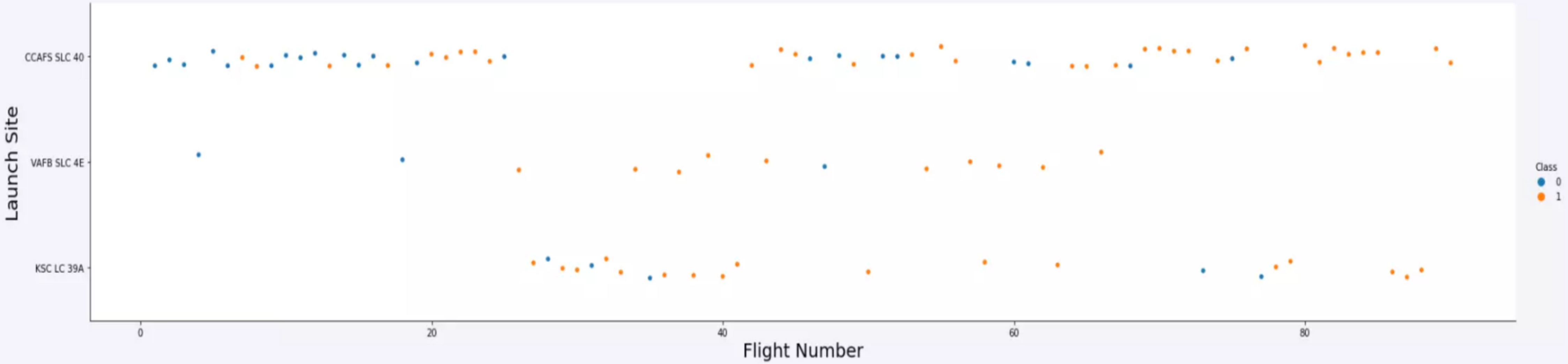
Predictive Analytics .

- Decision Tree model is the best predictive model for the dataset

Section 2

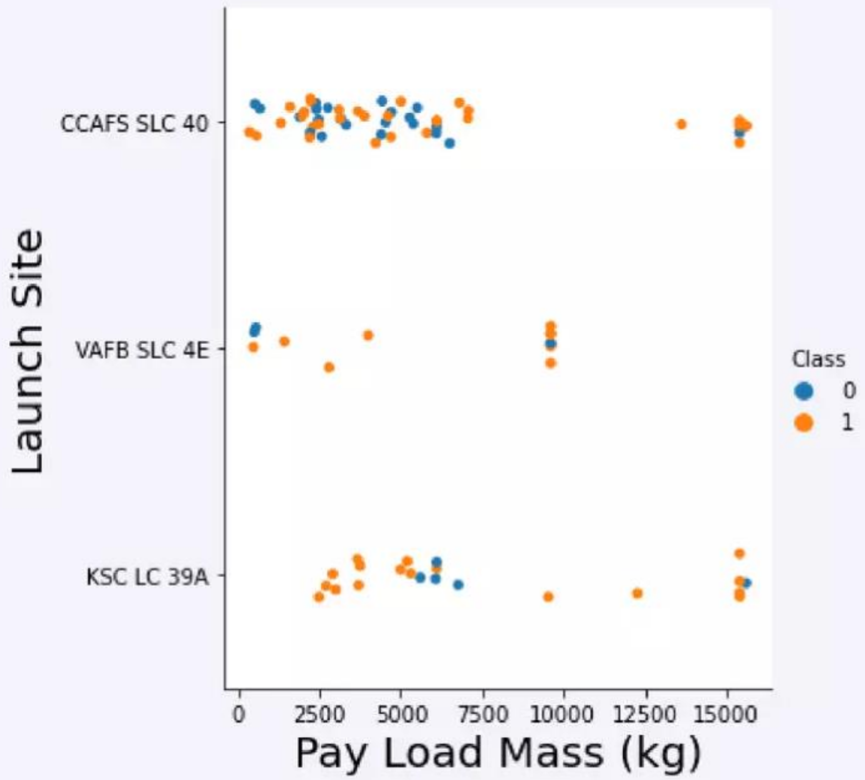
Insights drawn from EDA

Flight Number vs. Launch Site



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

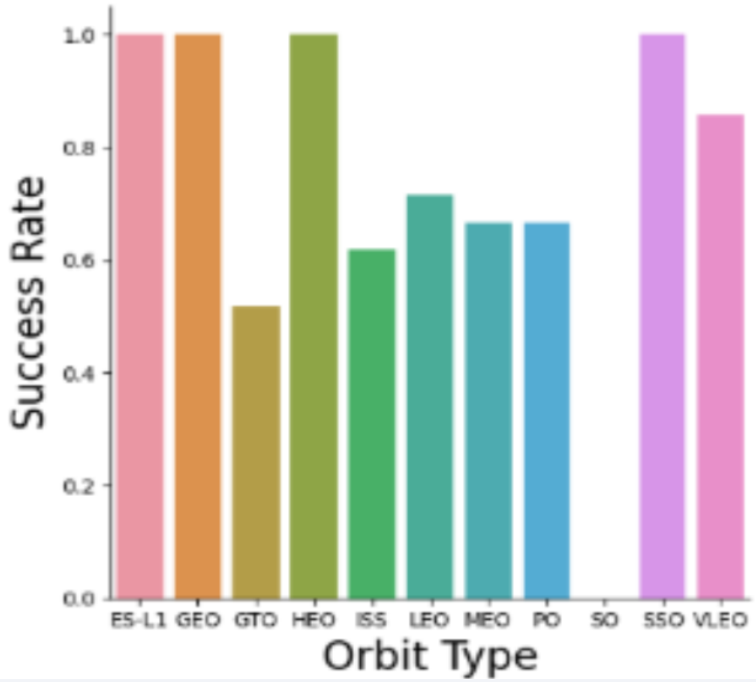
Payload vs. Launch Site



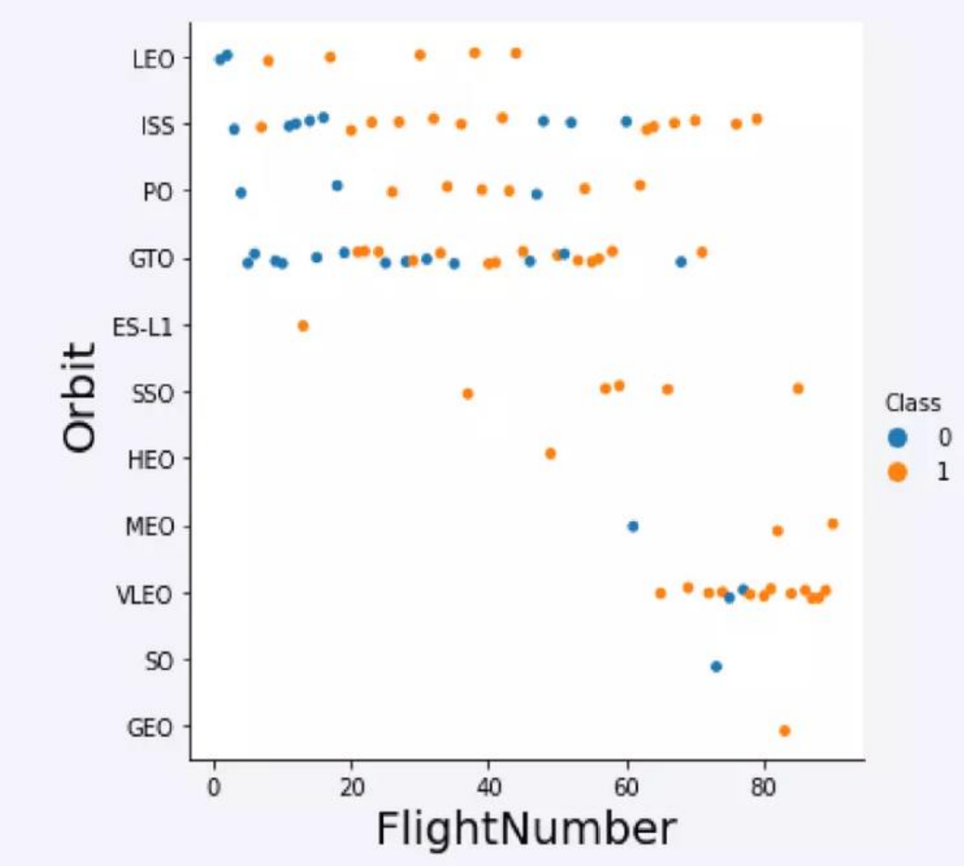
- The majority of IPay Loads with lower Mass have been launched from CCAFS SLC 40.

Success Rate vs. Orbit Type

- **100% Success Rate**: ES-L1, GEO, HEO and SSO
- **50%-80% Success Rate**: GTO, ISS, LEO, MEO, PO
- **0% Success Rate**: SO

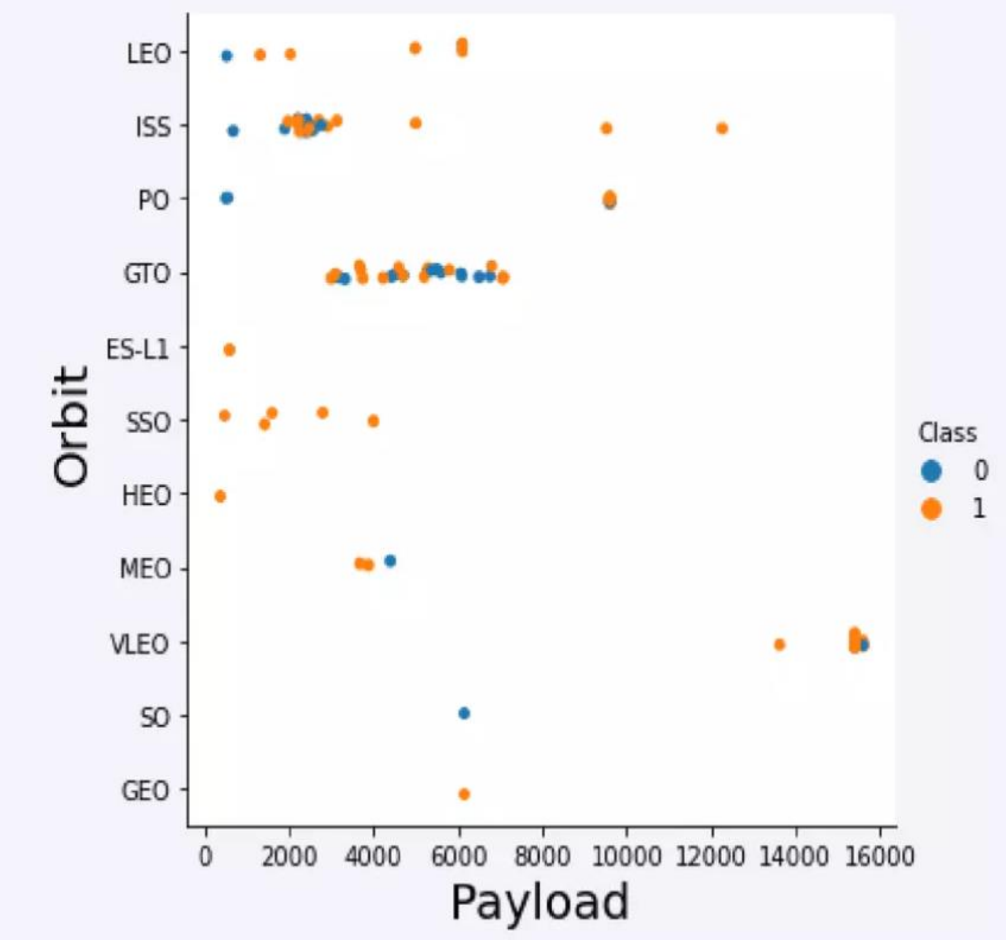


Flight Number vs. Orbit Type



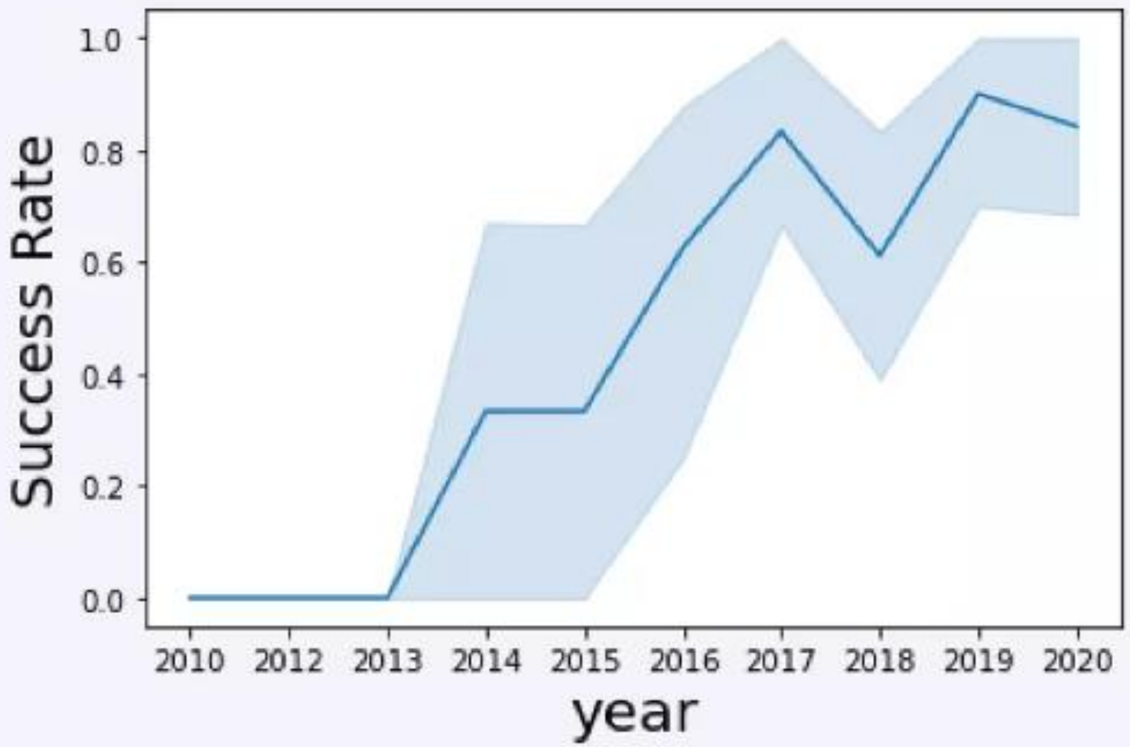
- A trend can be observed of shifting to VLEO launches in recent years.

Payload vs. Orbit Type



- There are strong correlation between ISS and Payload at the range around 2000, 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

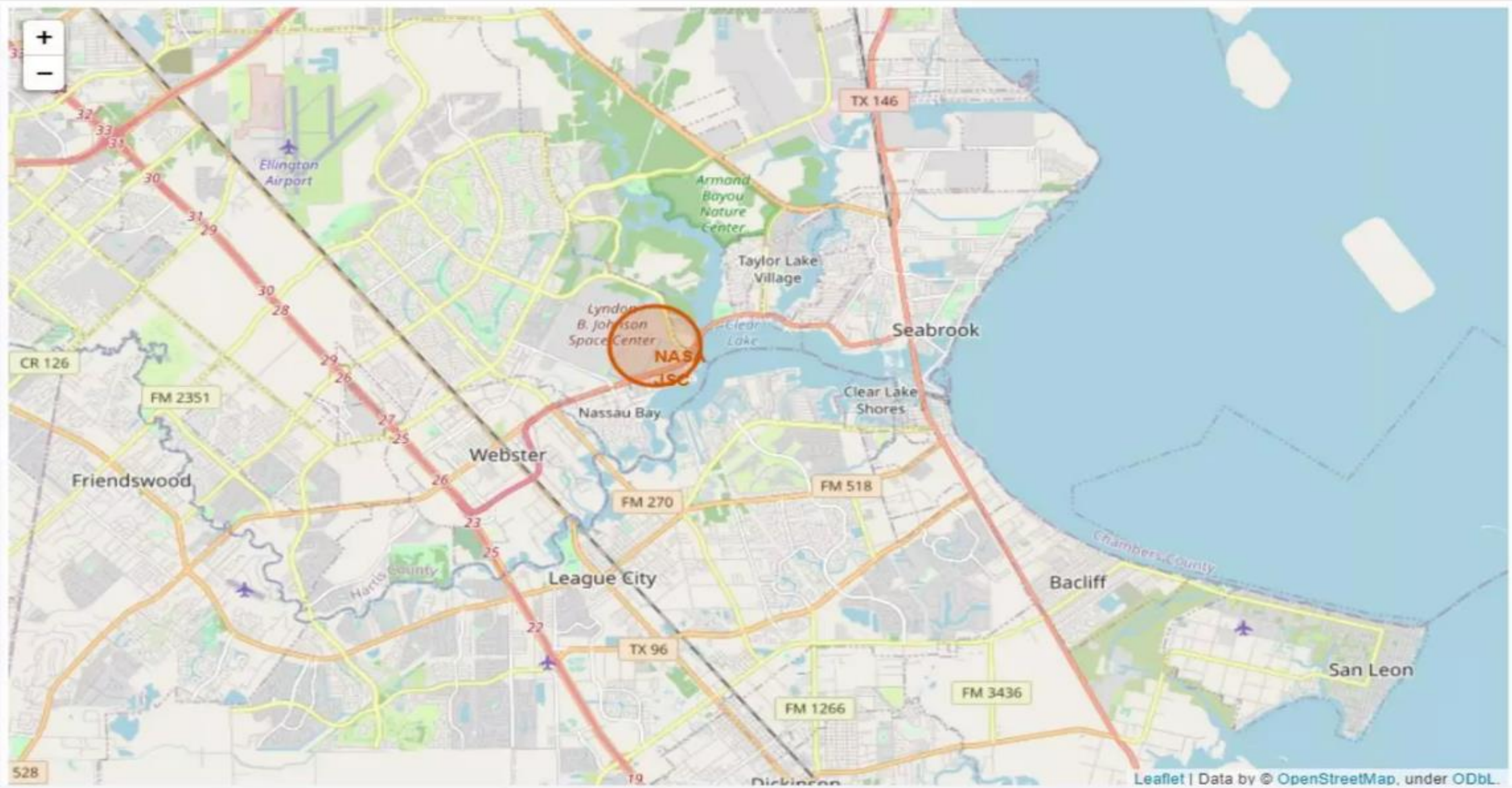
- %sql select * from SPACEXTBL where Landing__Outcome like 'Success%' and (DATE between '2010-06-04' and '2017-03-20') order by date desc

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 sites marked on the map



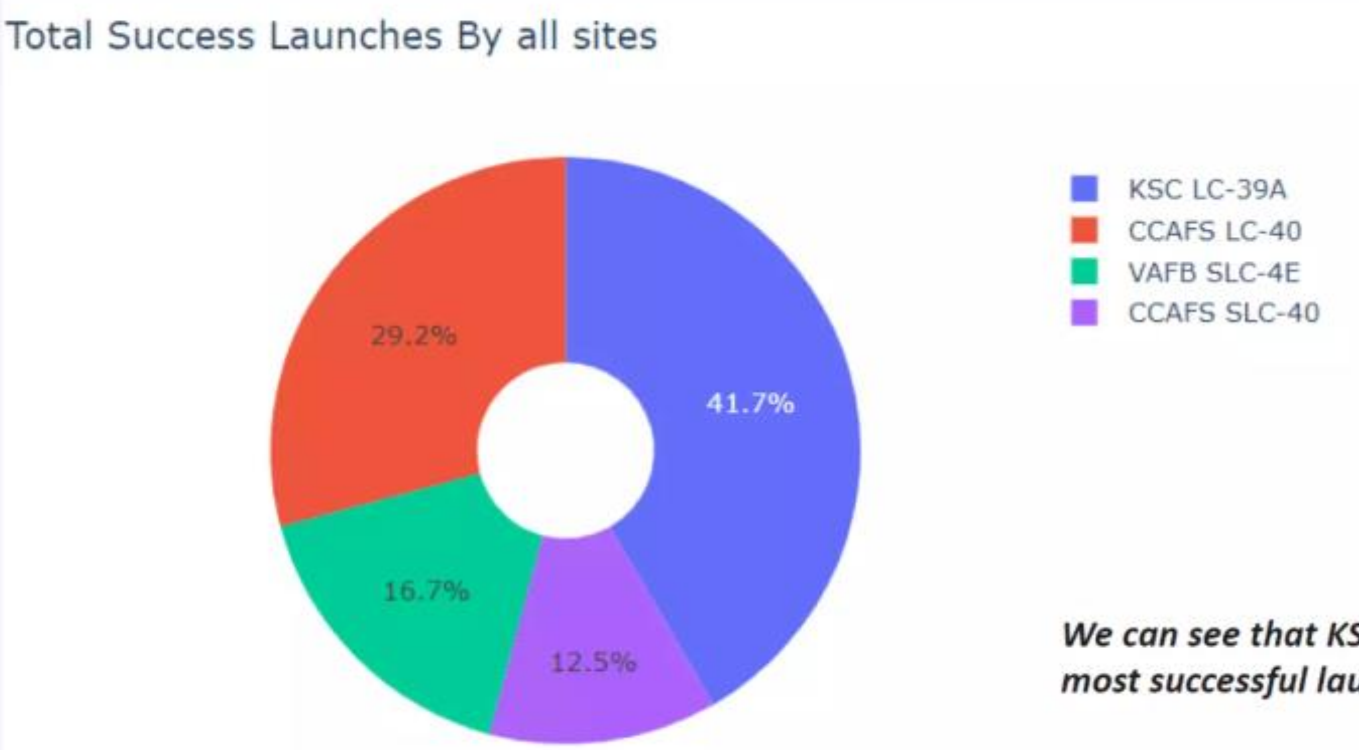
Success/failed marked on the map

Distance between launch sites and its proximities

Section 4

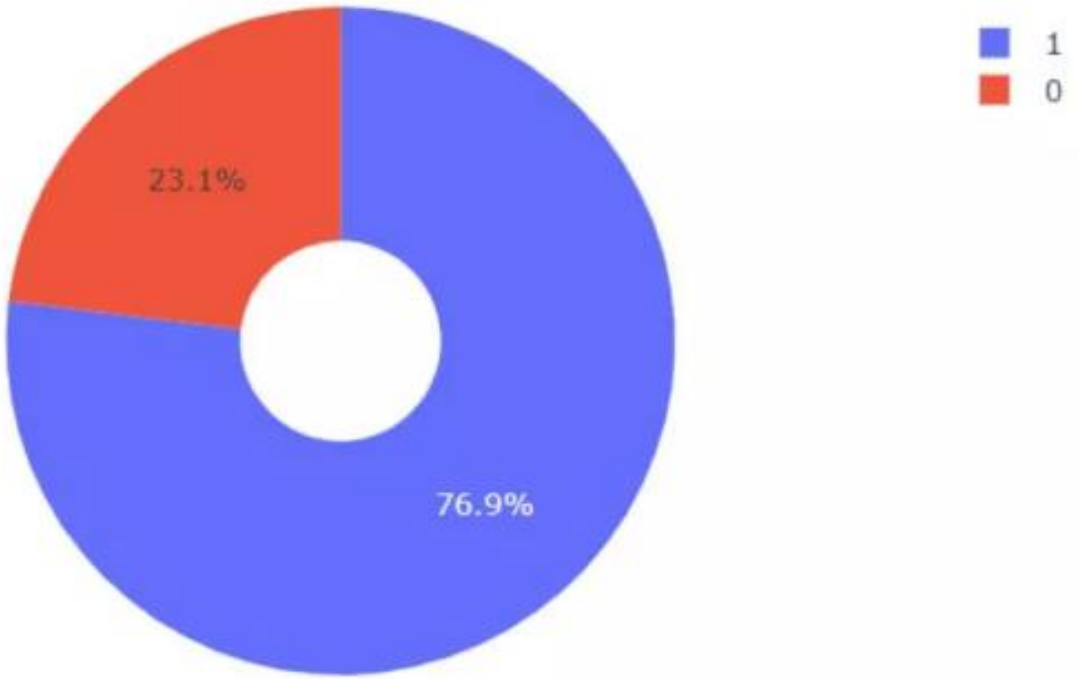
Build a Dashboard with Plotly Dash

Total success launches by all sites



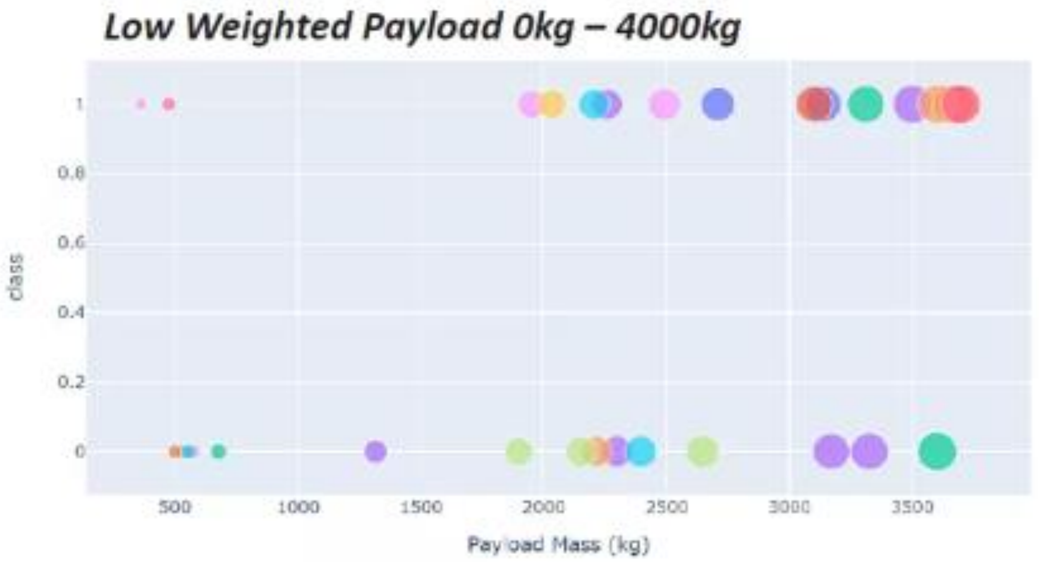
We can see that KSC LC-39A had the most successful launches from all the sites

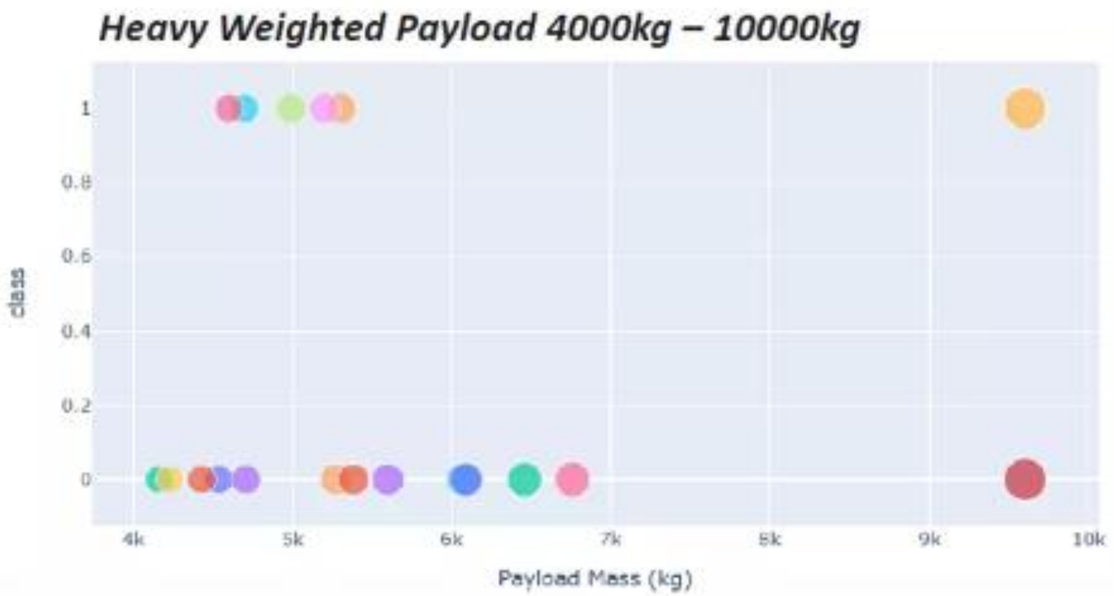
Success rate by sites



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

Payload vs Launch outcome



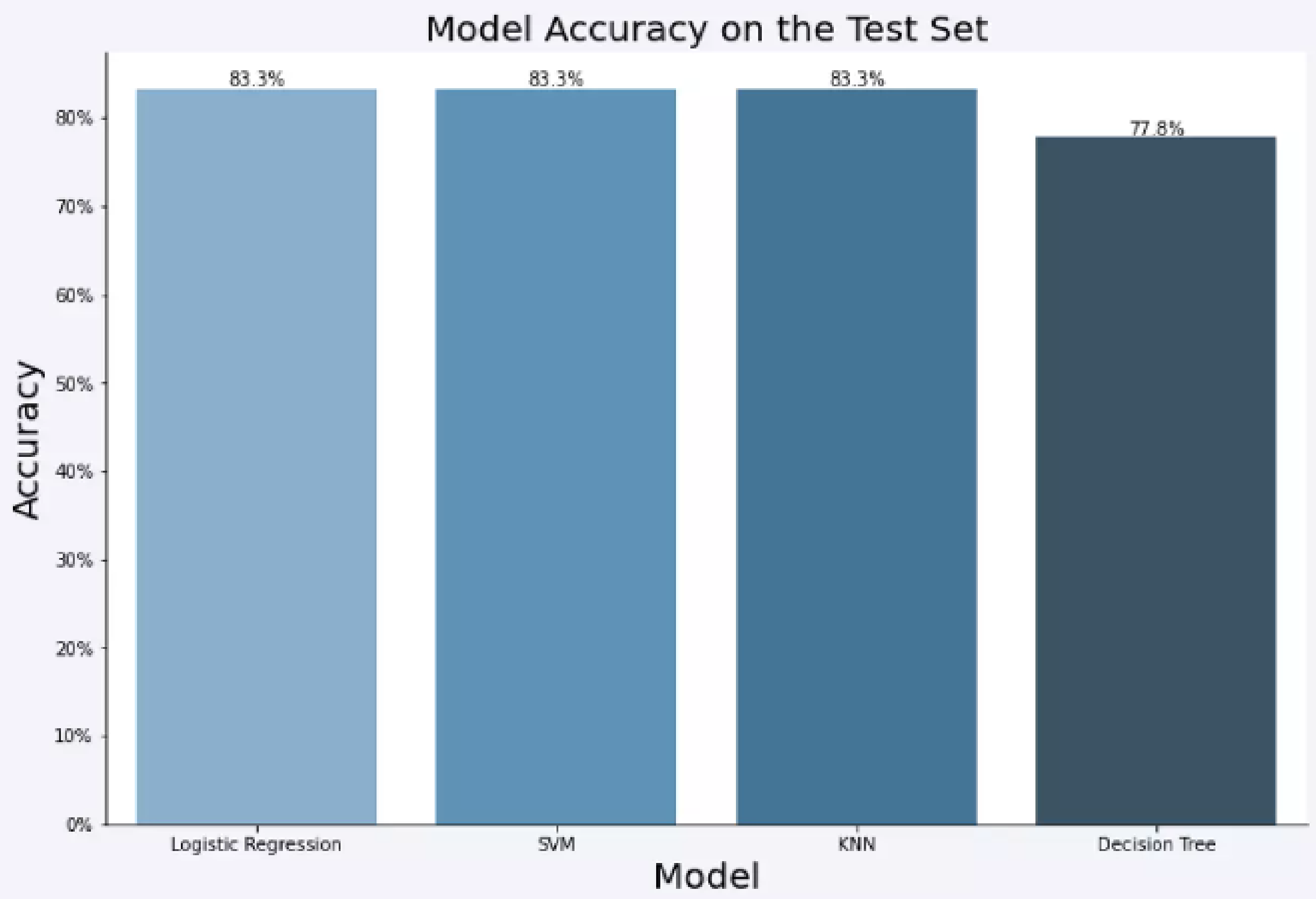


We can see the success rates for low weighted payloads is higher than the heavy weighted payloads

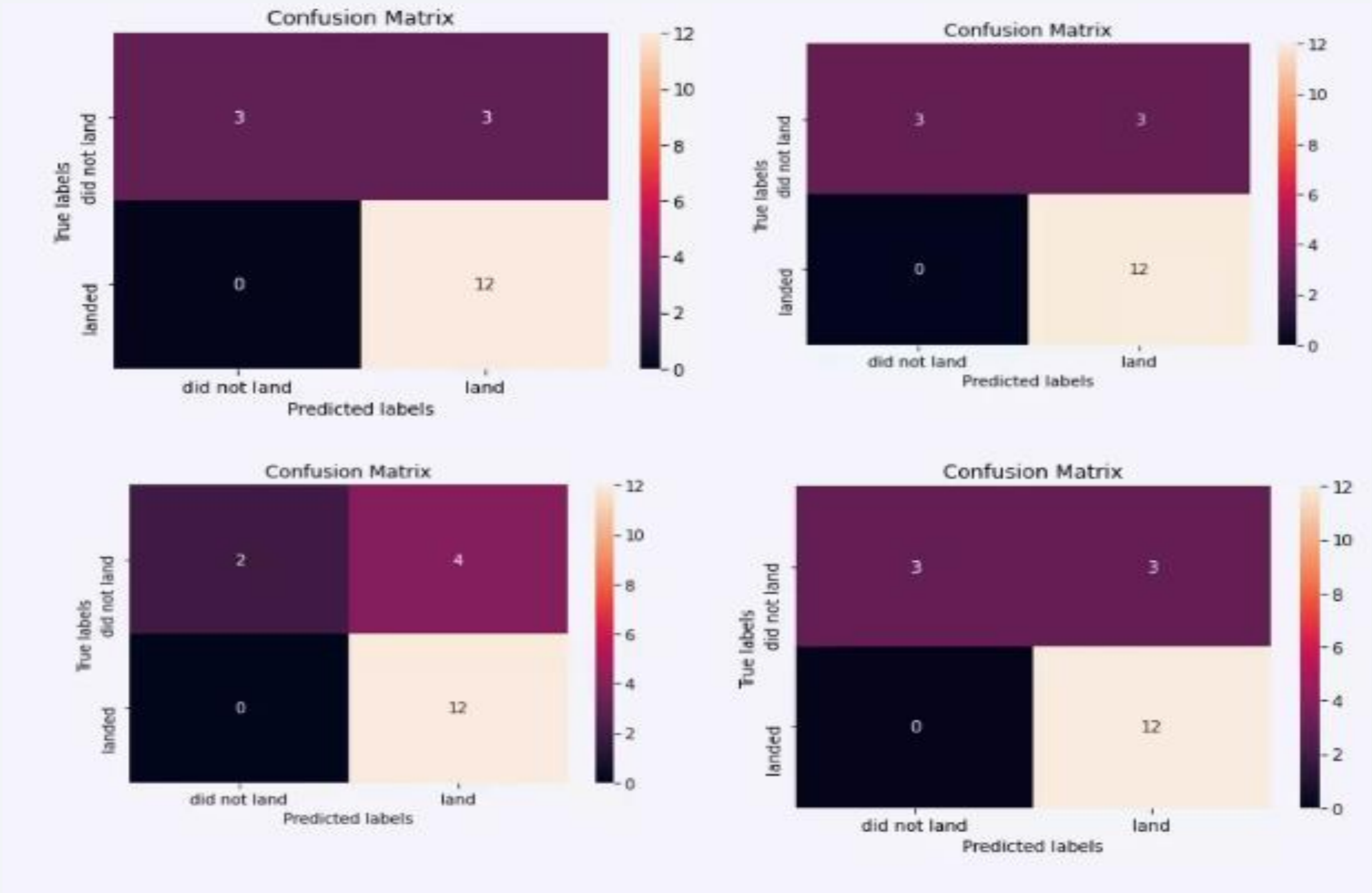
Section 5

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 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!