IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Name: Sudhanshu Prakash
Date: Jan 23, 2022

Outline

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

Executive Summary

- Summary of methodologies
    - Data collection
    - Data wrangling
    - EDA with data visualization
    - 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

- SpaceX is one of the most successful for sending spacecraft to the International Space Station. It re-uses first stage to reduce cost in second stage of the launch. Sometimes first stage crashes due to the mission parameters like payload, orbit, and customer

- The price of each launch will be determined. Whether SpaceX will reuse the first stage will be determined. Instead of using rocket science to determine if the first stage will land successfully, machine learning model and public information will be used to predict if SpaceX will reuse the first stage.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:

  SpaceX launch data was gathered from an API, specifically the SpaceX REST API (api.spacexdata.com/v4/)

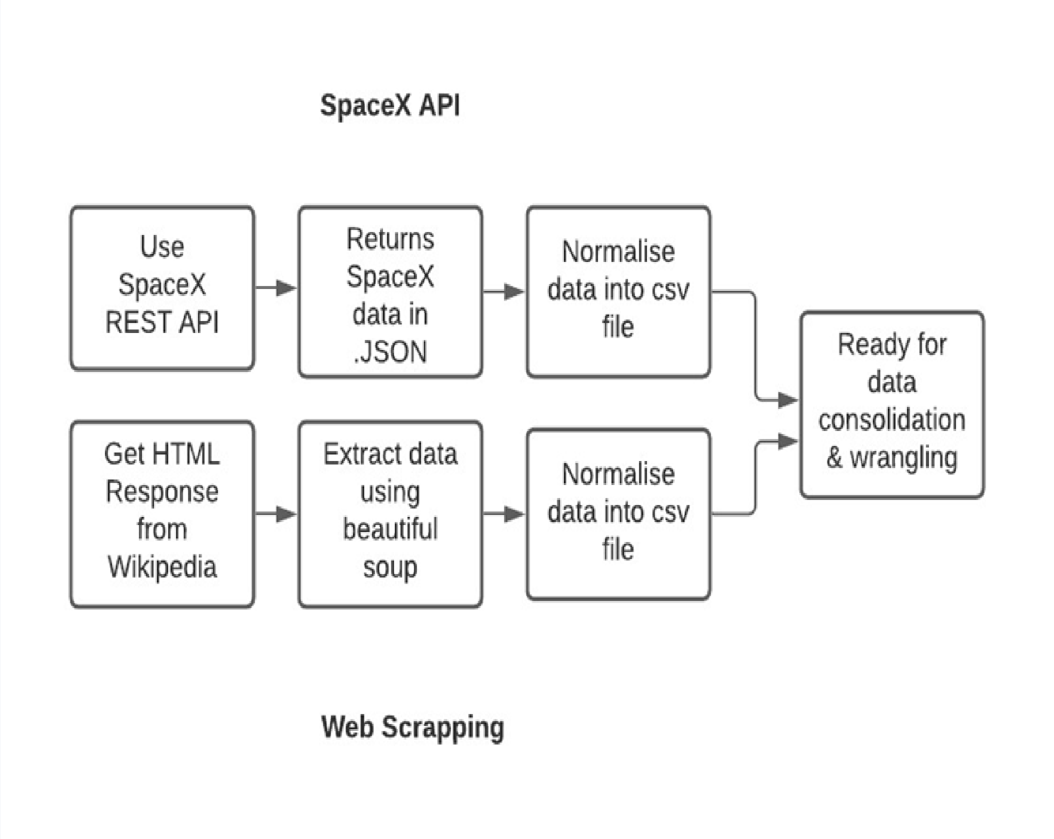
- Perform data wrangling

  Missing values replaced with mean value, correct data types and the data is normalized

- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

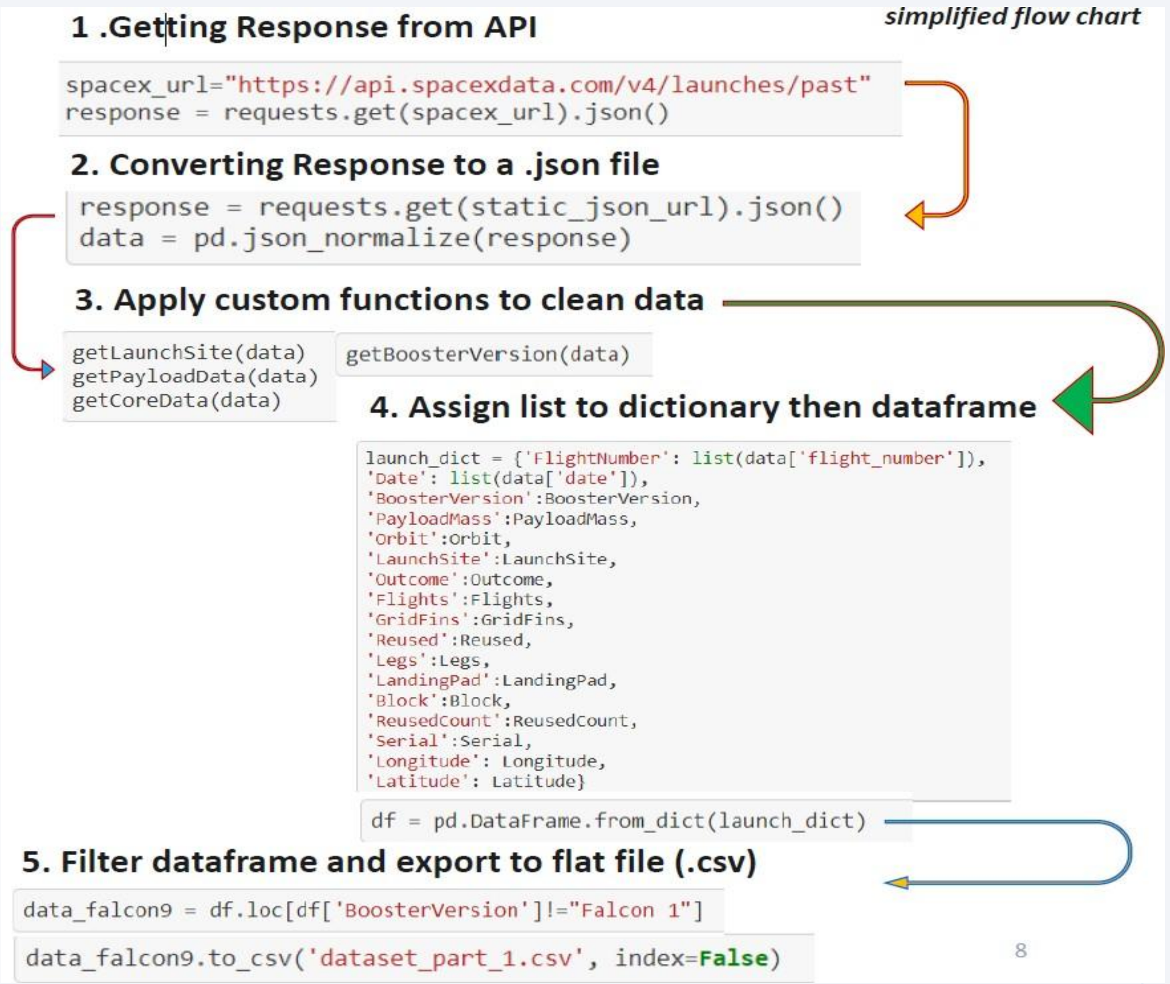
Data Collection

- The following datasets was collected:
- SpaceX launch data that is gathered from the 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.com/v4/.
- Another popular data source for obtaining Falcon 9 Launch data is web scraping Wikipedia using BeautifulSoup.



Data Collection – SpaceX API

- Data collection with SpaceX REST calls
- Github url : https://github.com/sudhanshu-prakash/Spacex-Data-Science-Project/blob/main/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts

- https://github.com/sudhanshu-prakash/Spacex-Data-Science-Project/blob/main/jupyter-labs-webscraping.ipynb

TASK 1: Request the Falcon9 Launch Wiki page from its URL

First, let's perform an HTTP GET method to request the Falcon9 Launch HTML page, as an HTTP response.

```
response = requests.get(static_url)
```

Create a `BeautifulSoup` object from the HTML `response`

```
soup = BeautifulSoup(response.content)
```

TASK 2: Extract all column/variable names from the HTML table header

Next, we want to collect all relevant column names from the HTML table header

Let's try to find all tables on the wiki page first. If you need to refresh your memory about `BeautifulSoup`, plea
the end of this lab

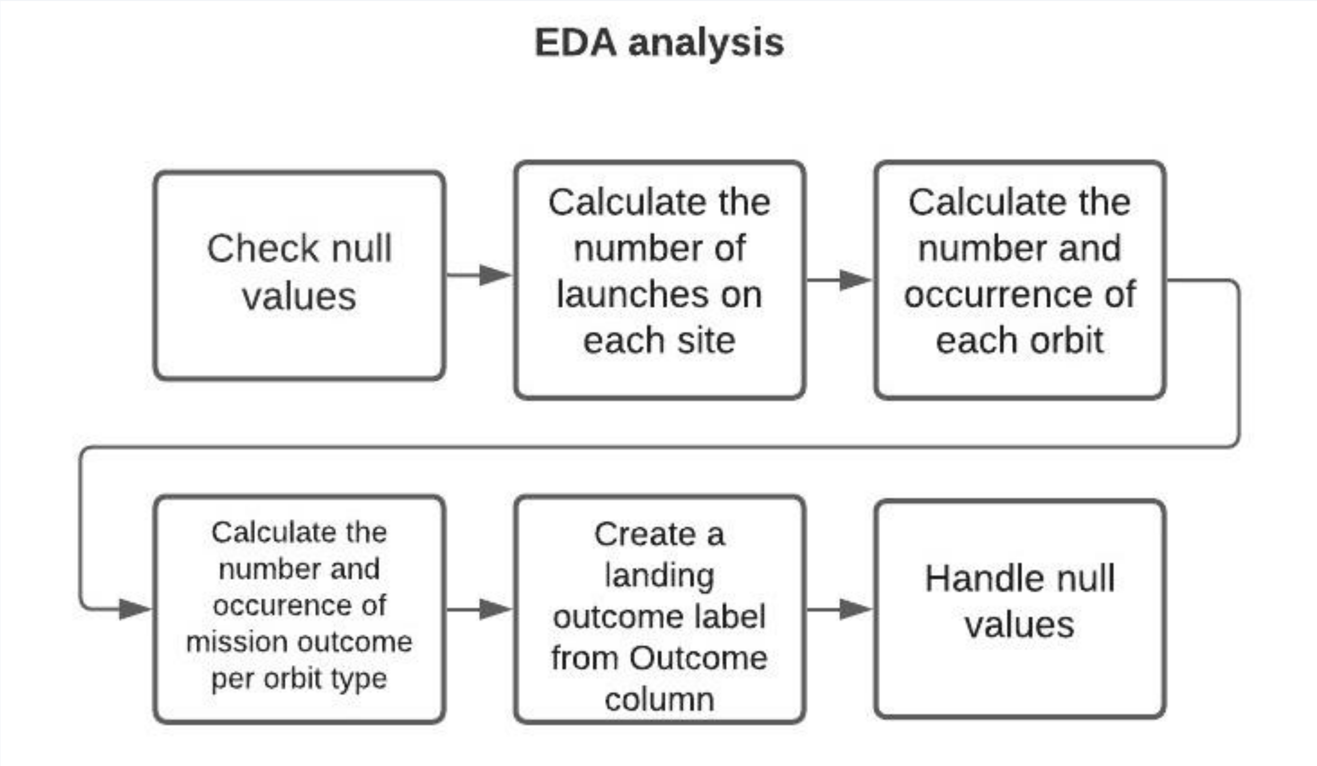
```
# Use the find_all function in the BeautifulSoup object, with element type `table`
# Assign the result to a list called `html_tables`
html_tables = soup.find_all('table')
len(html_tables)
```

TASK 3: Create a data frame by parsing the launch HTML tables

We will create an empty dictionary with keys from the extracted column names in the previous task. Later, this c
dataframe

```
launch_dict= dict.fromkeys(column_names)
```

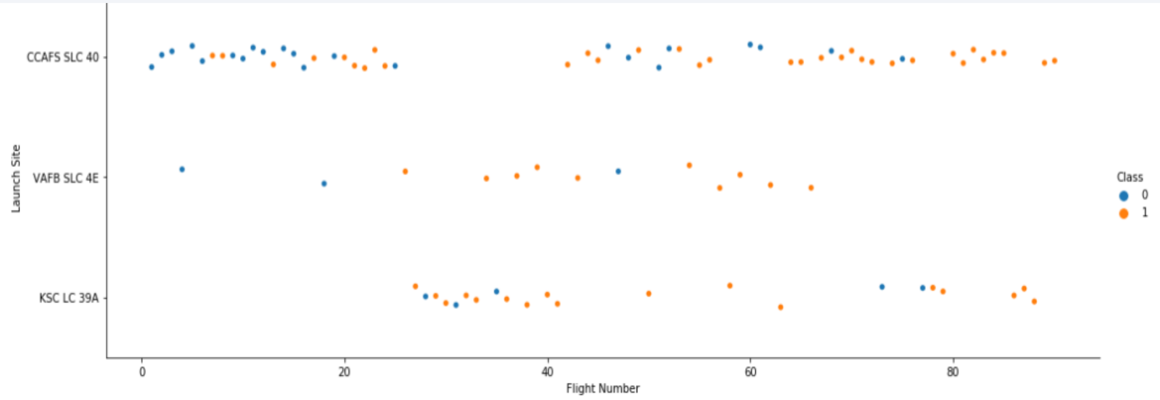
Data Wrangling



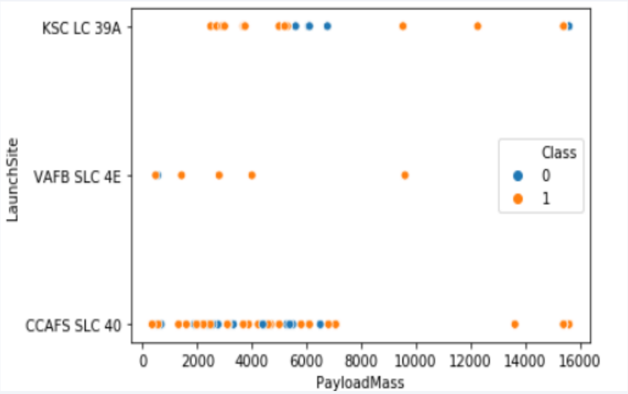
Github URL: https://github.com/sudhanshu-prakash/Spacex-Data-Science-Project/blob/main/jupyter-labs-webscraping.ipynb

EDA with Data Visualization

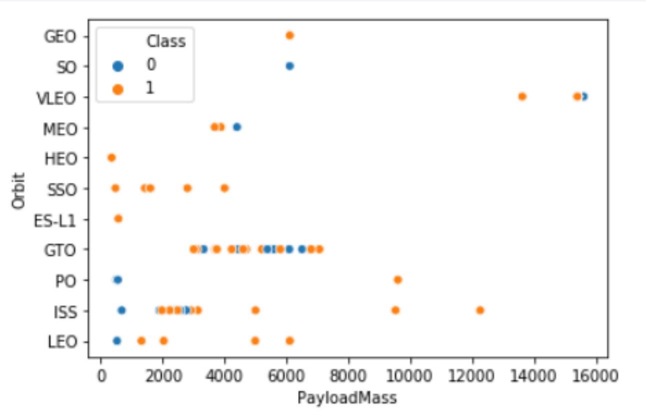
Visualize the relationship between Flight Number and Launch Site



Visualize the relationship between Payload and Launch Site

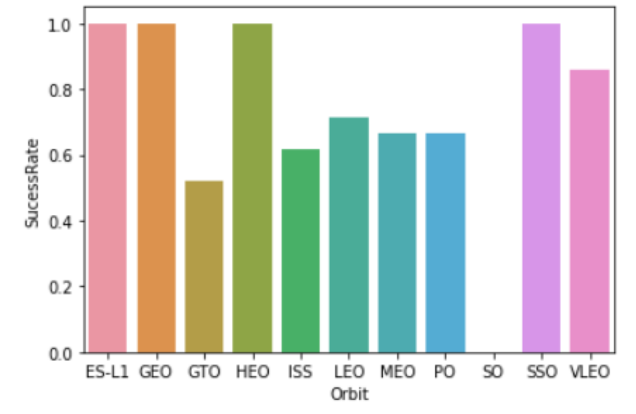


Visualize the relationship between Payload and Orbit type

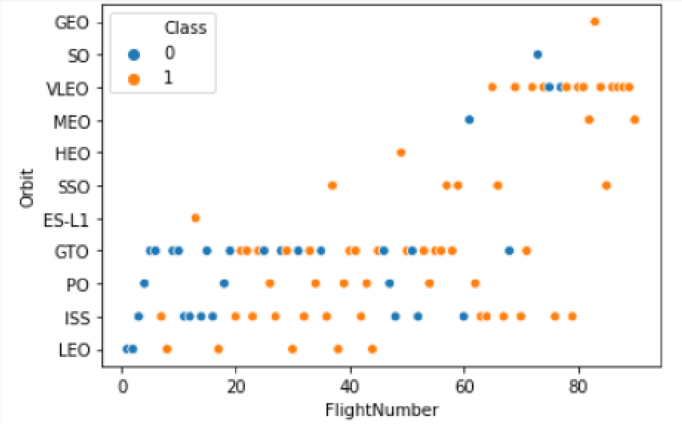


Github URL: https://github.com/sudhanshu-prakash/Spacex-Data-Science-Project/blob/main/jupyter-labs-eda-dataviz.ipynb

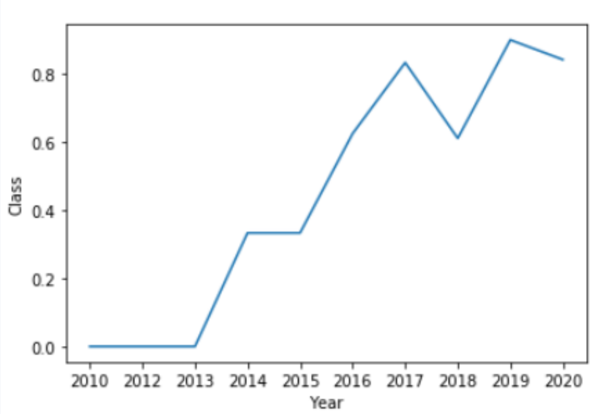
Visualize the relationship between success rate of each orbit type



Visualize the relationship between FlightNumber and Orbit type



Visualize the launch success yearly trend



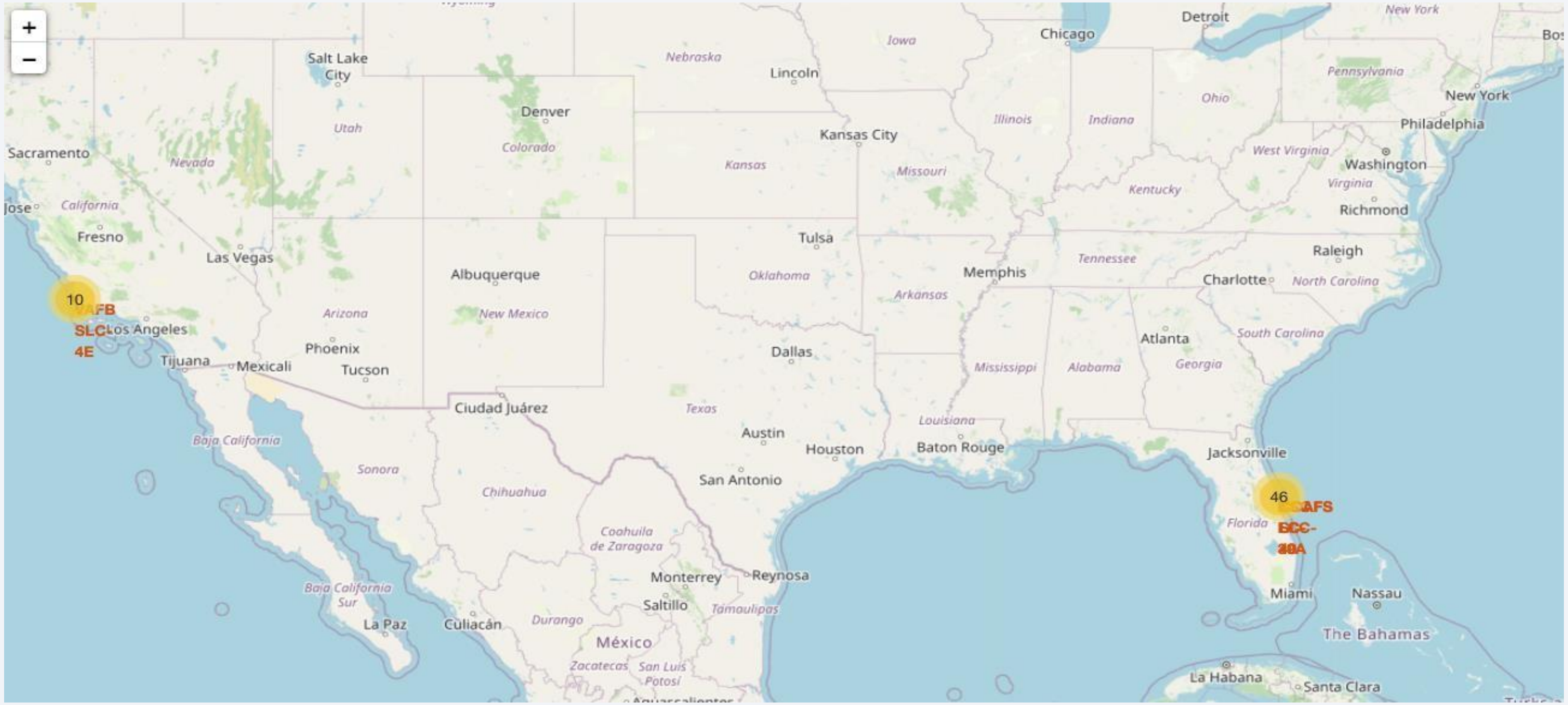
EDA with SQL

SQL queries performed include:

- 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 versions, launch site for the months in year 2017
- Ranking the count of successful landing outcomes between the date 2010-06-04 and 2017-03-20 in descending order.

- Github URL: https://github.com/sudhanshu-prakash/Spacex-Data-Science-Project/blob/main/jupyter-labs-eda-sql-coursera.ipynb

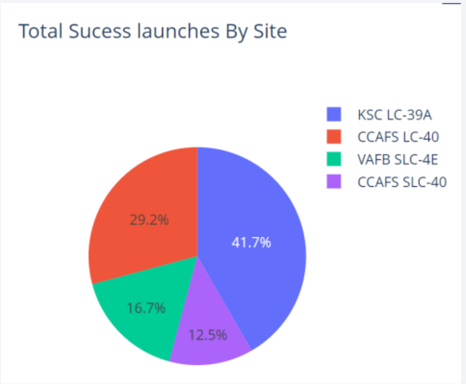
Build an Interactive Map with Folium



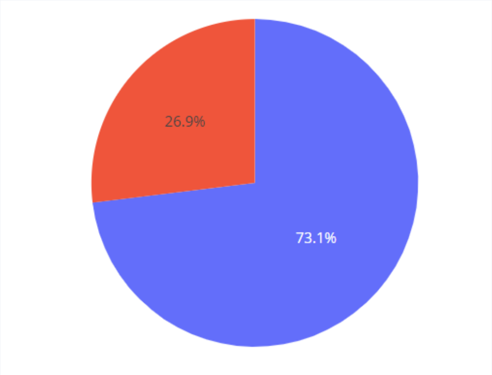
Github URL: https://github.com/sudhanshu-prakash/Spacex-Data-Science-Project/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

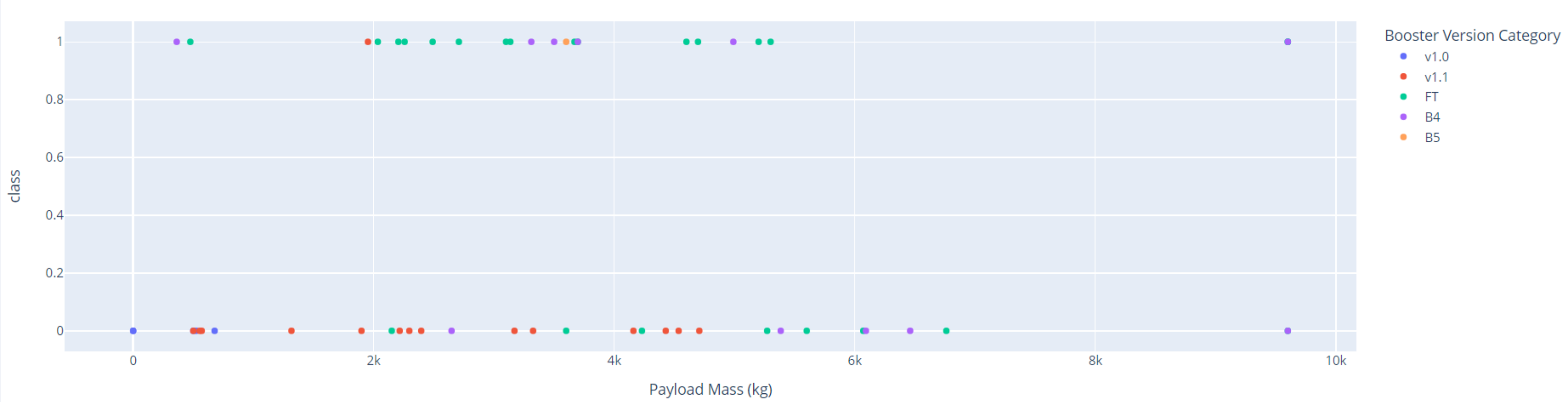
Success launches by all Site



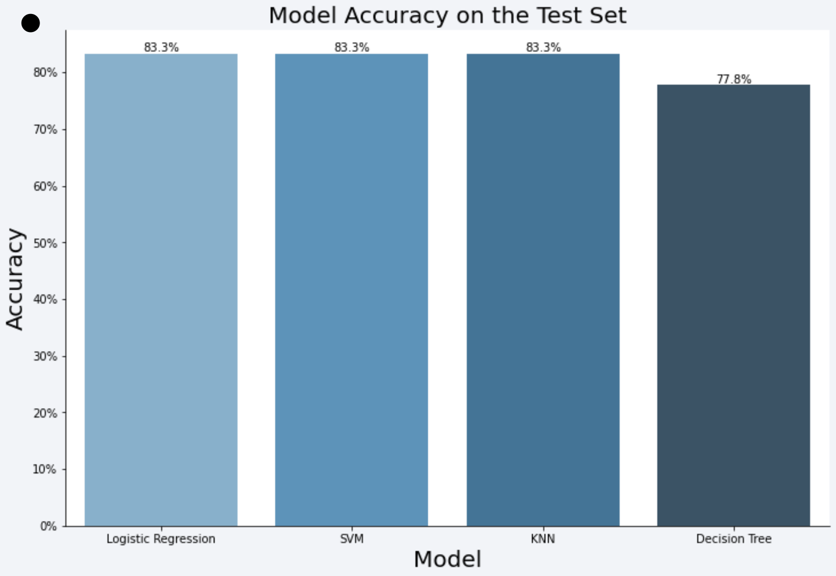
Total Sucess launches for Site CCAFS LC-40

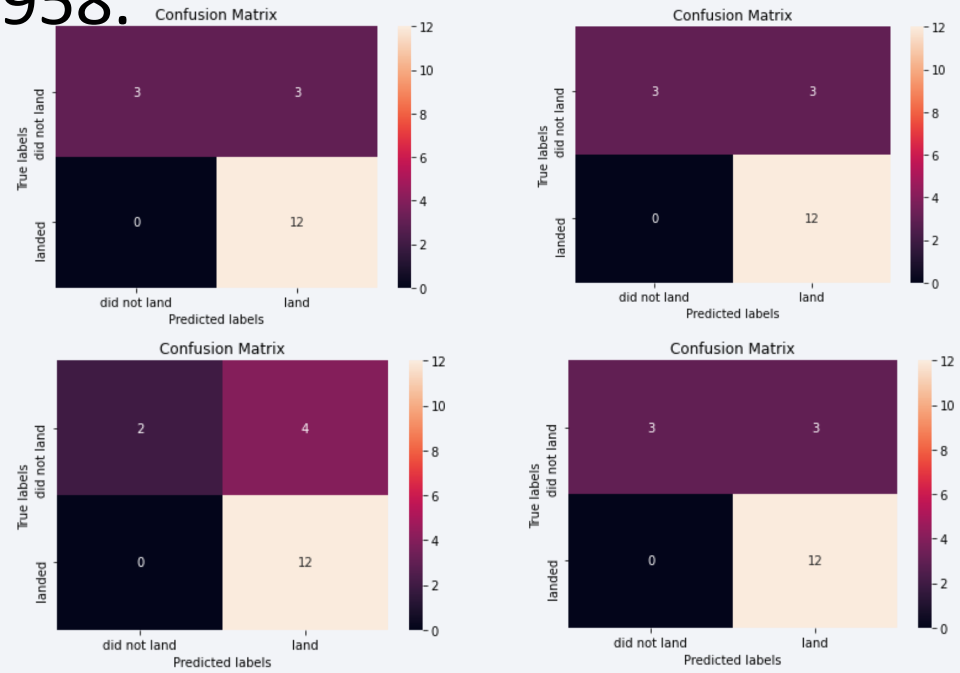


https://github.com/sudhanshu-prakash/Spacex-Data-Science-Project/blob/main/spacex_dash_app.py



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/sudhanshu-prakash/Spacex-Data-Science-Project/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

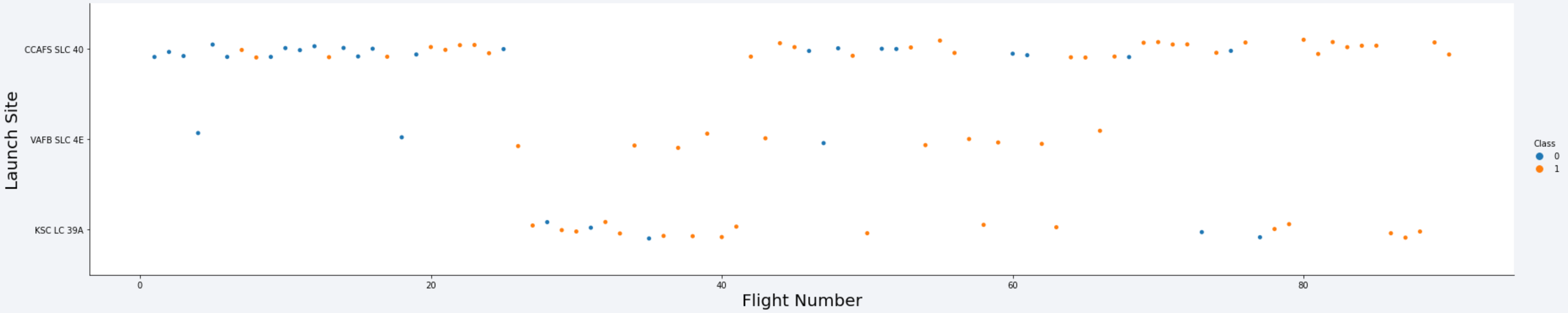
Results

- 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 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

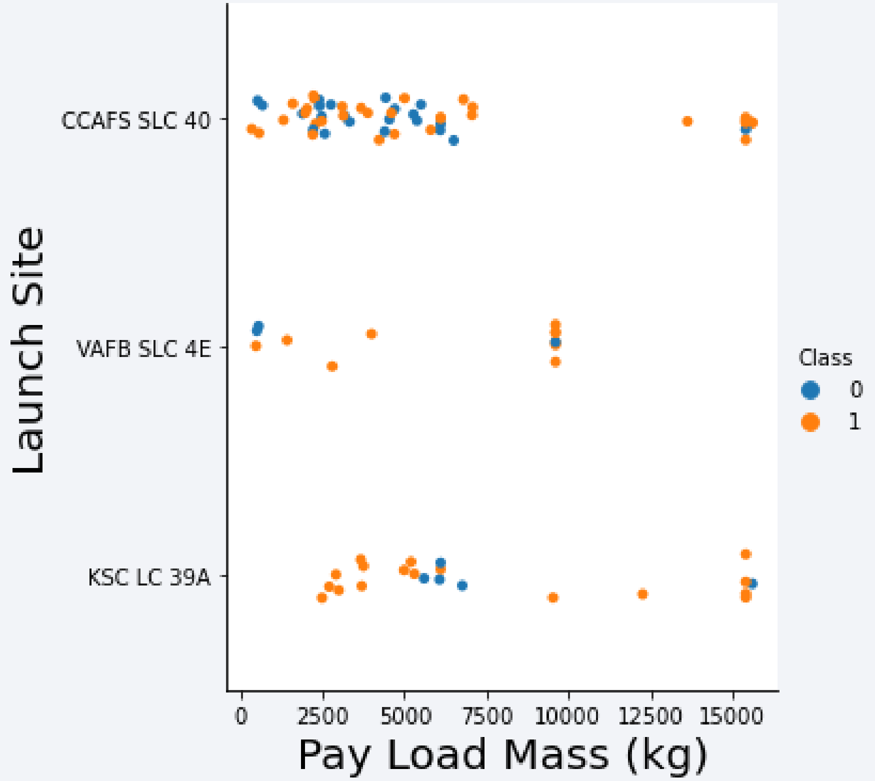
Flight Number vs. Launch Site



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

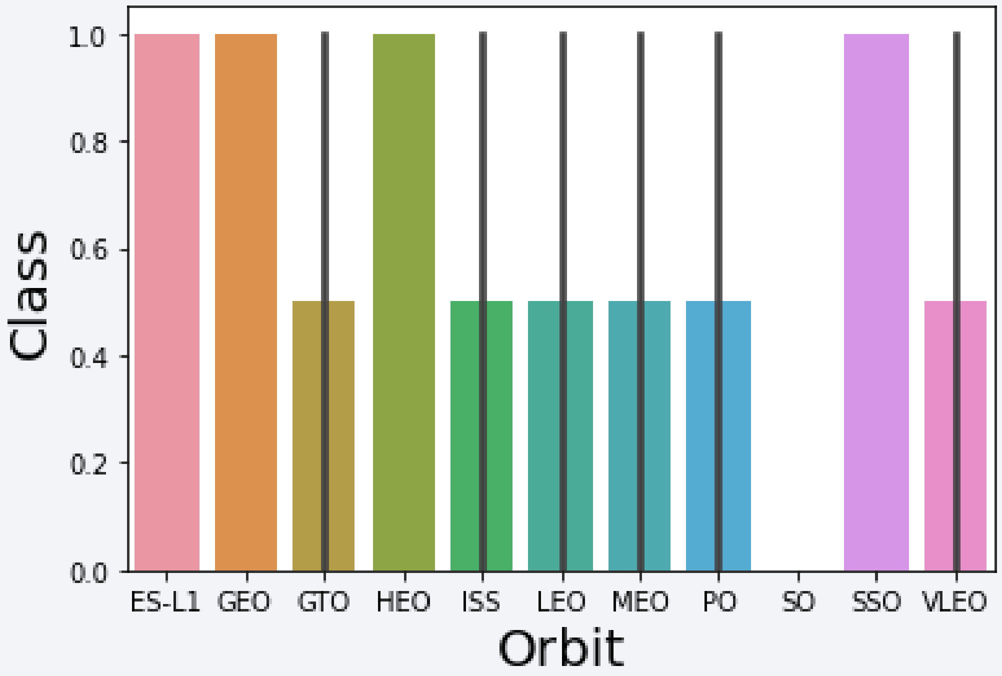
Payload vs. Launch Site

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



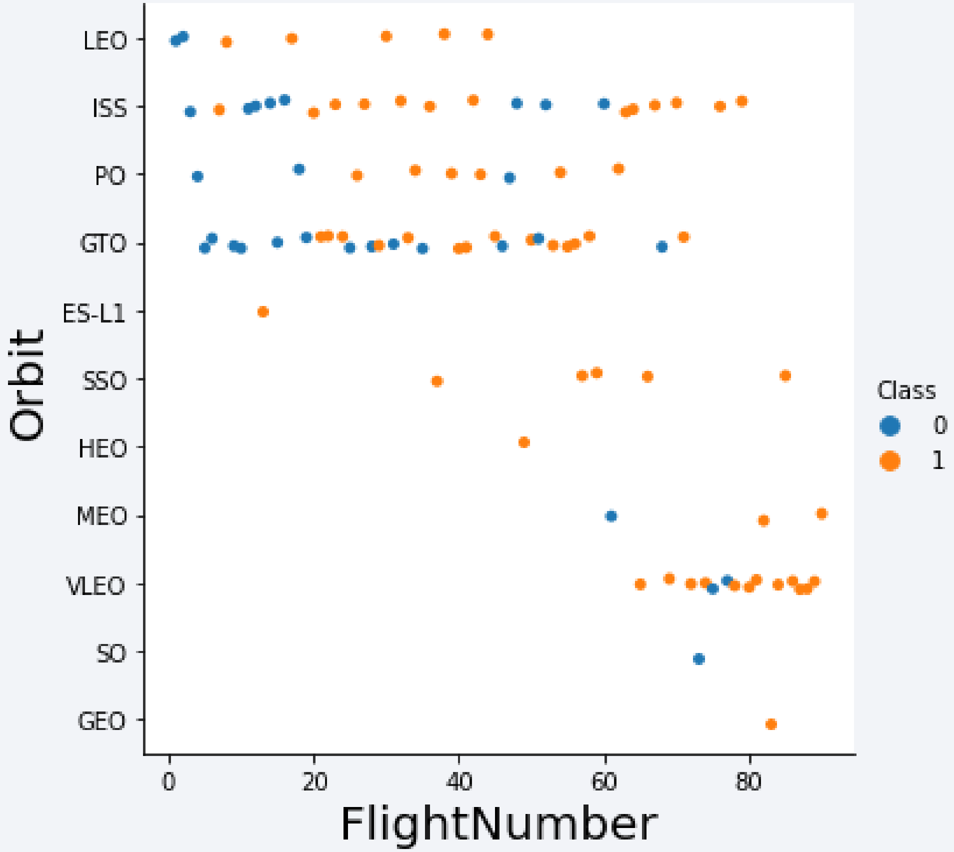
Success Rate vs. Orbit Type

- The orbit types of ES-L1, GEO, HEO, SSO are among the highest success rate.



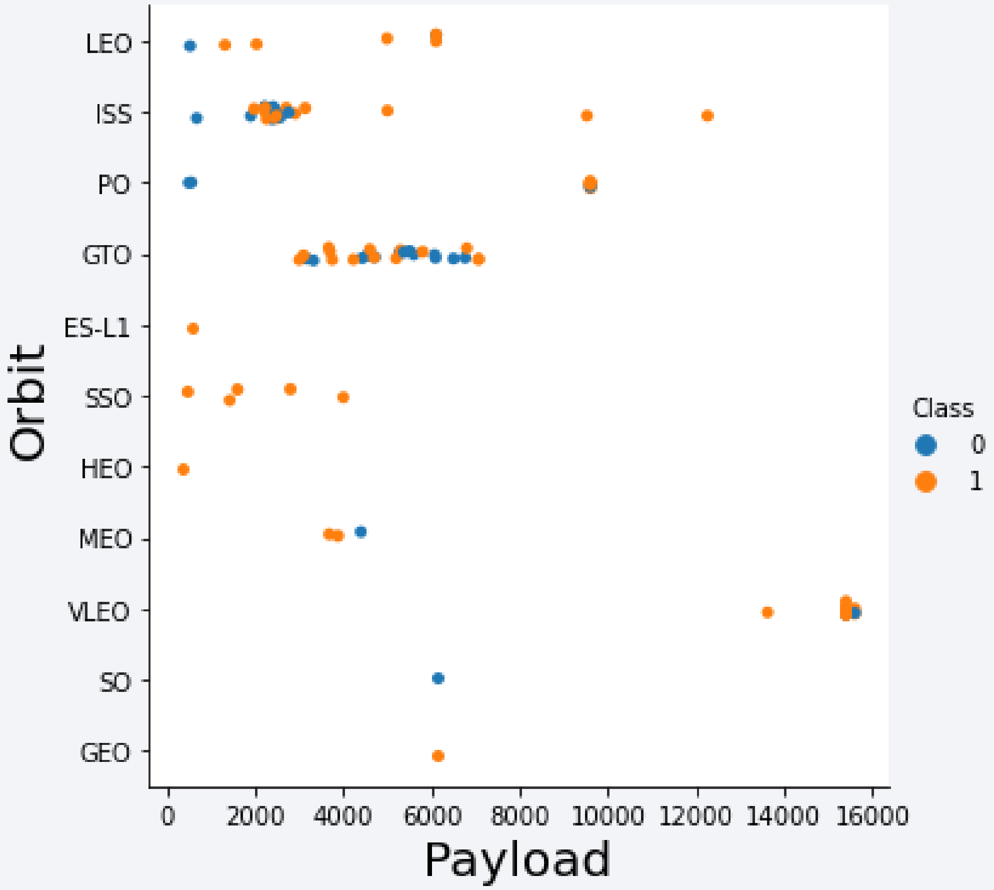
Flight Number vs. Orbit Type

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



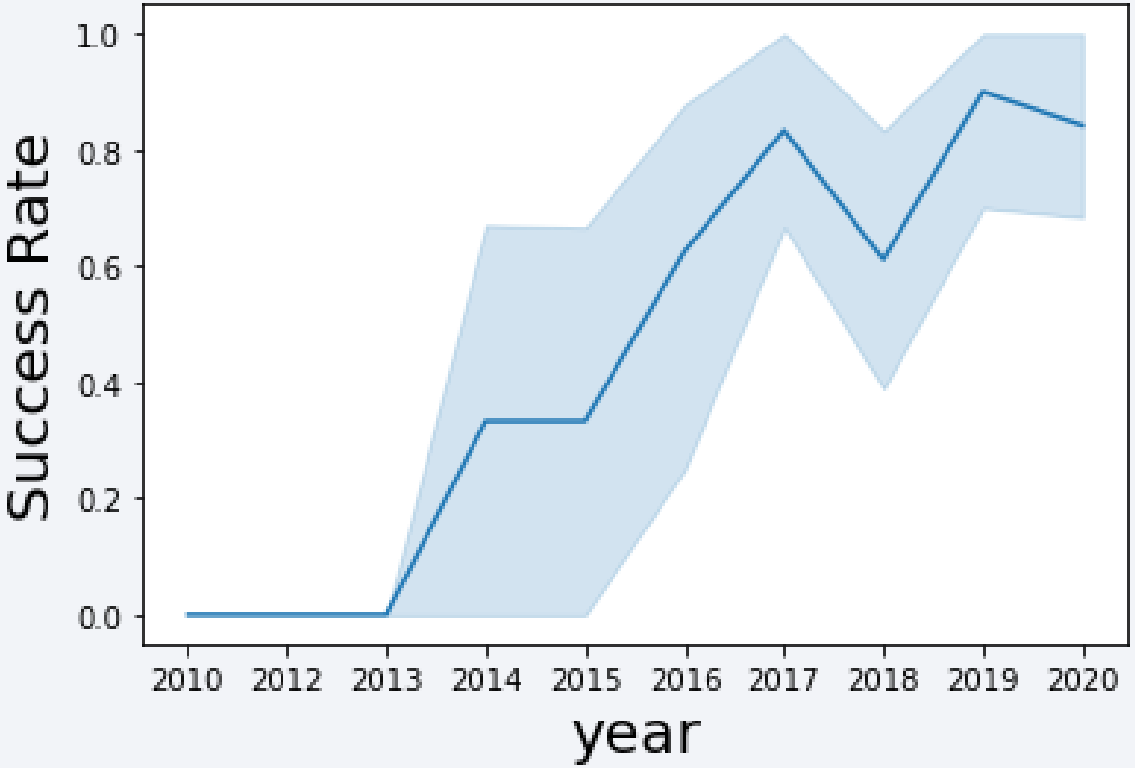
Payload vs. Orbit Type

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

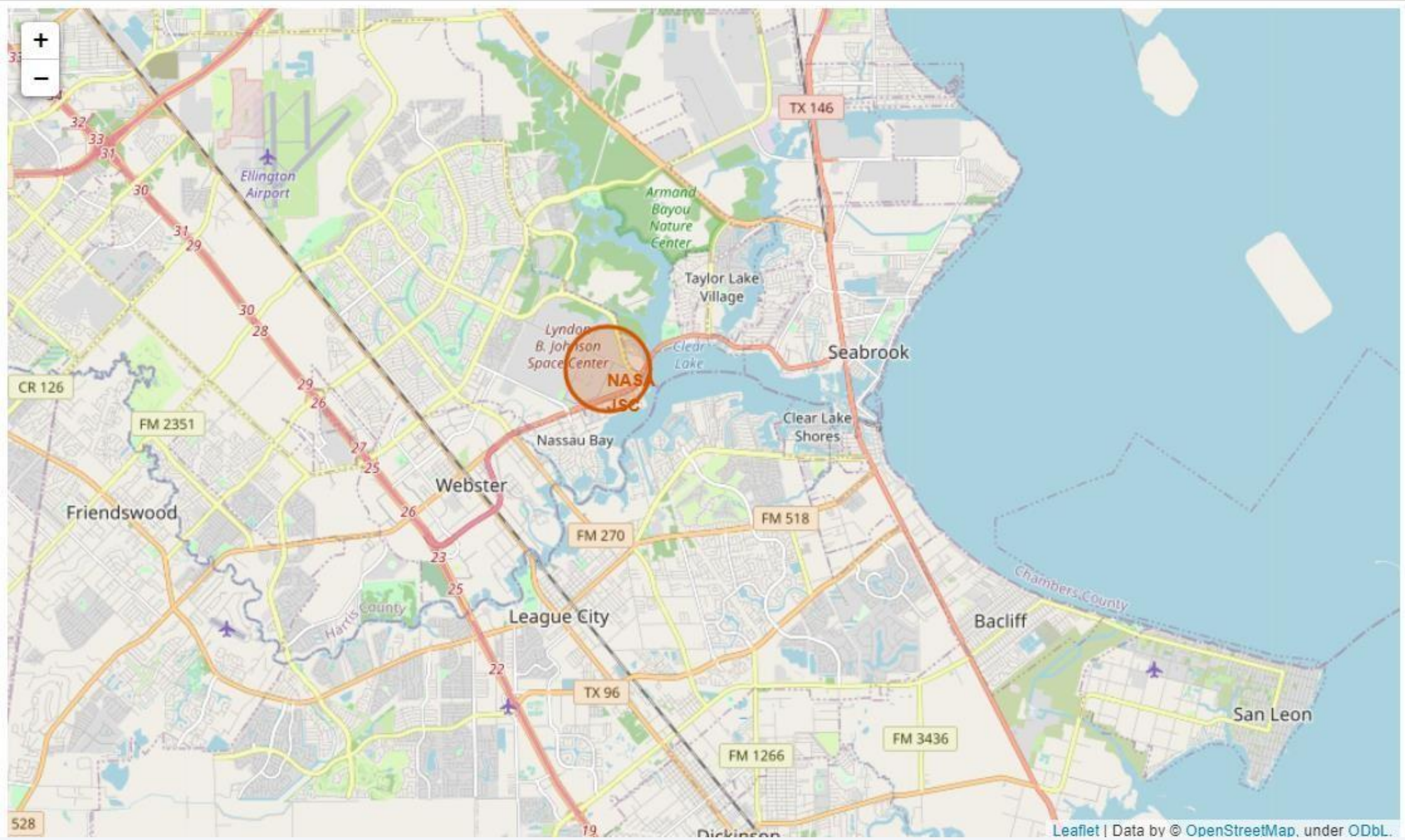
Section 4

Launch Sites Proximities Analysis

Folium Map Screenshot 1: All launch sites marked on a map

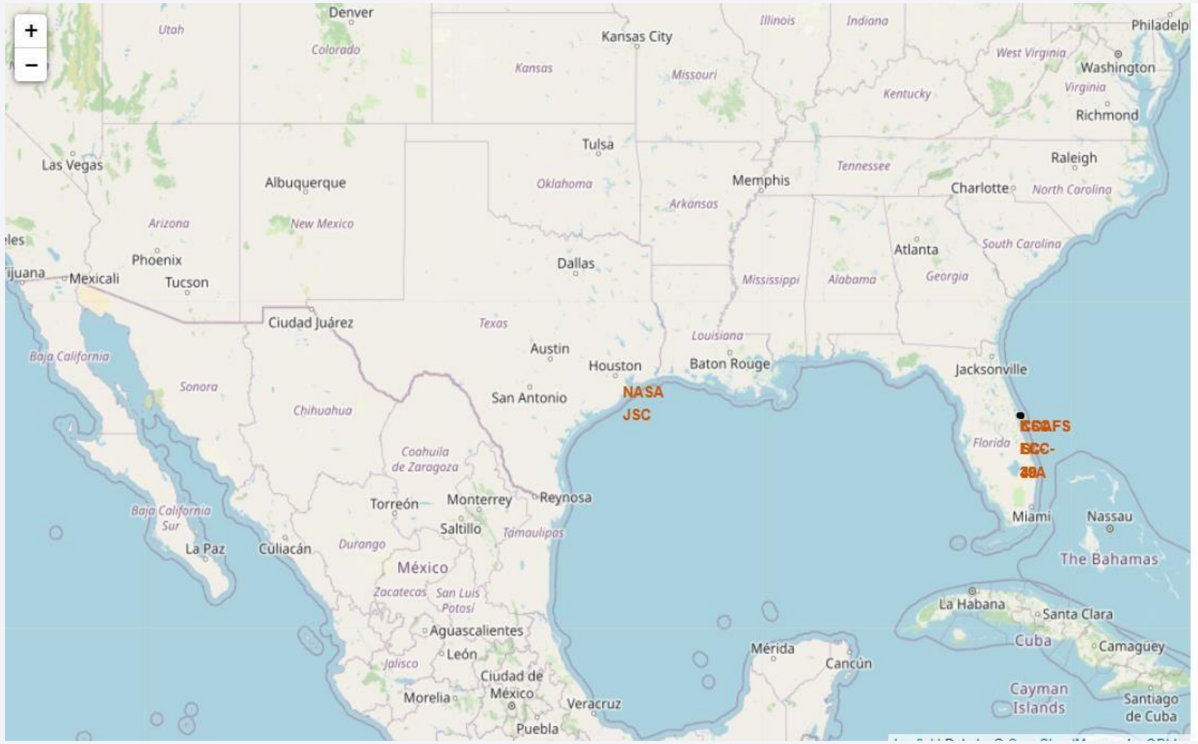
- All launch sites are in proximity to the Equator line.

- Yes , all launch sites are in very close proximity to the coast.

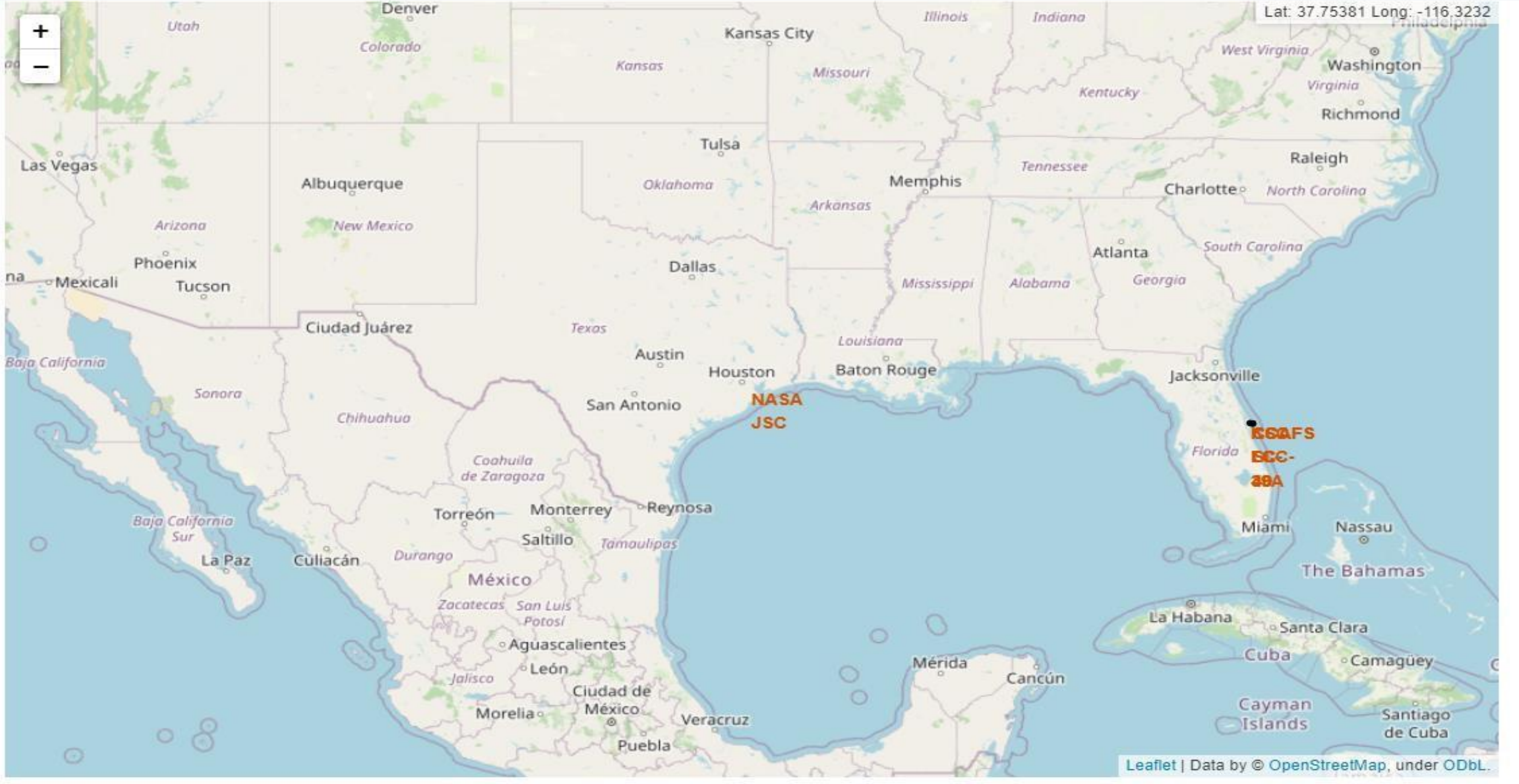


Folium Map Screenshot 2: `Success/failed launches marked on the map`

- From the color-labeled markers in marker clusters, are able to easily identify which launch sites have relatively high success rates.



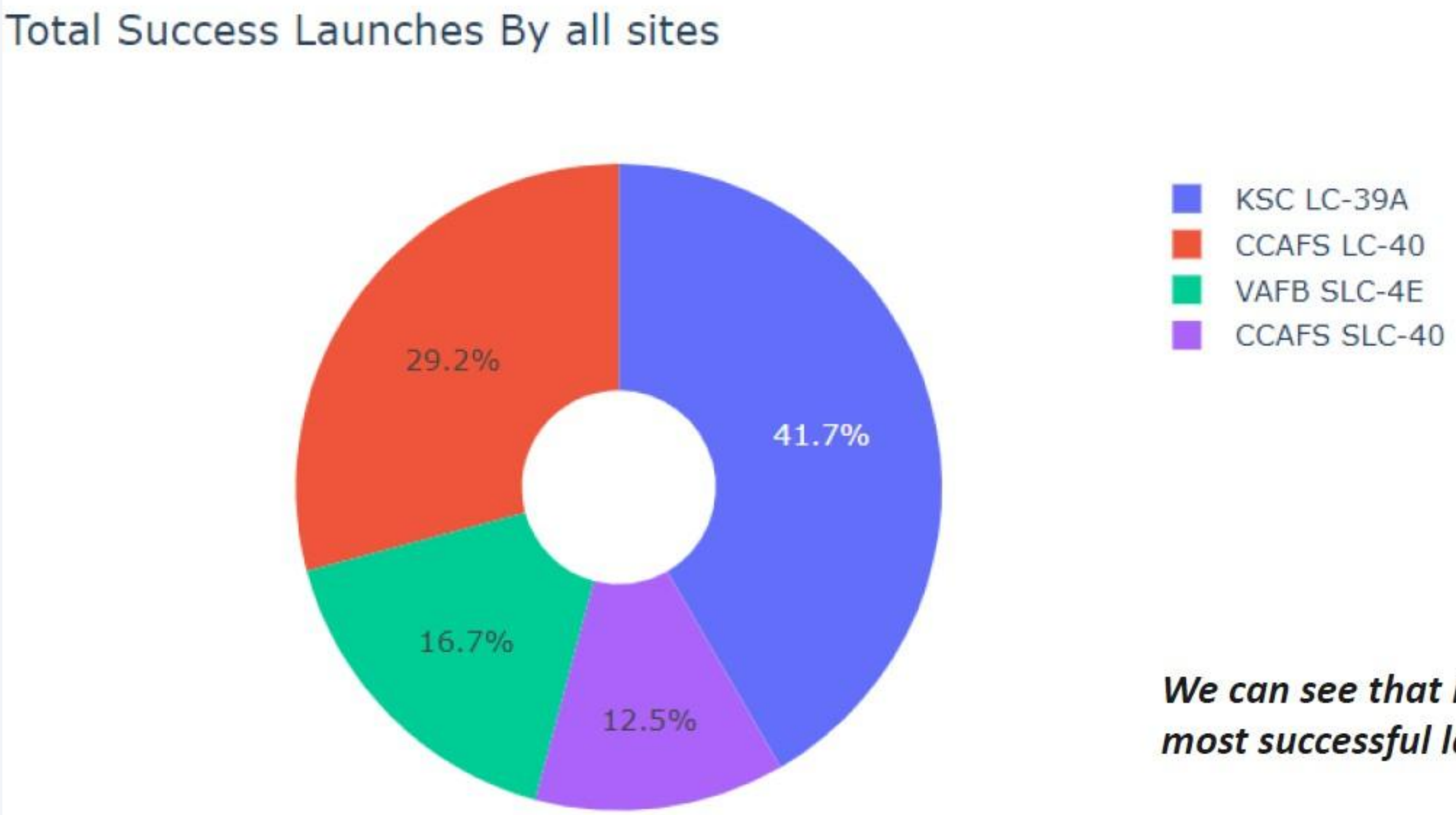
Folium Map Screenshot 3: Distances between a launch site to its proximities



Section 5

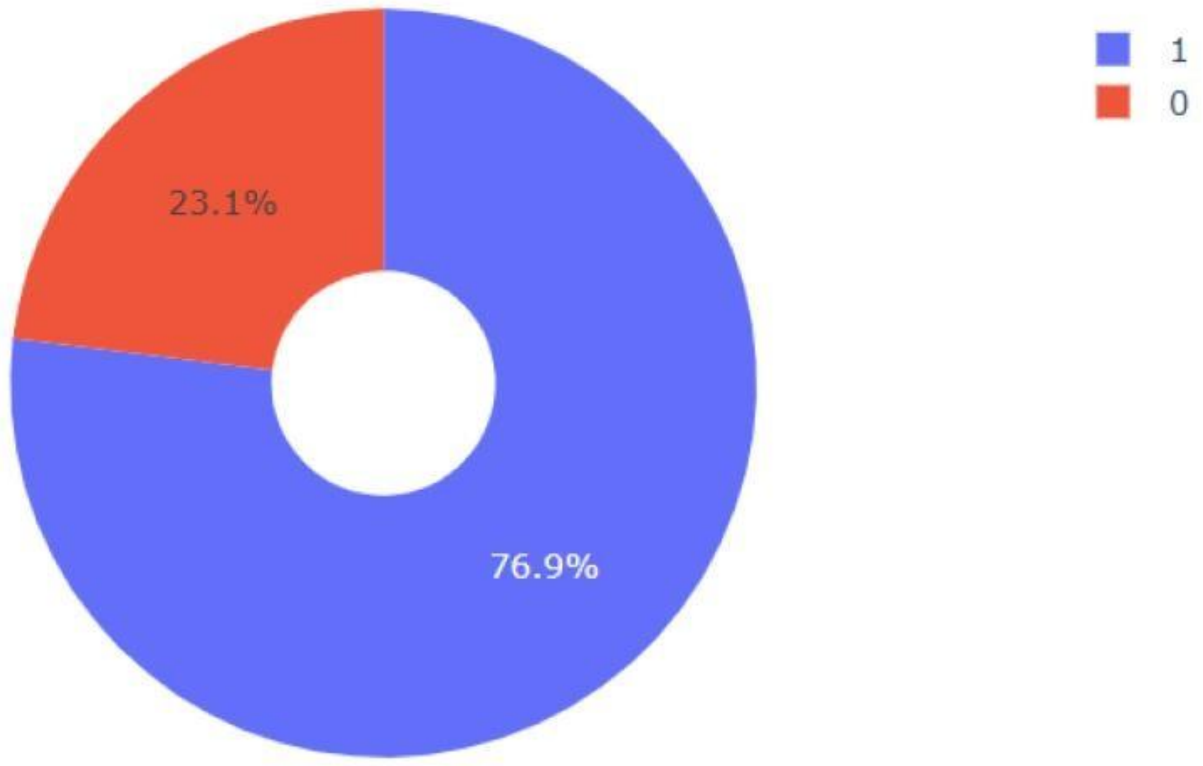
Build a Dashboard with Plotly Dash

Total success launches by all sites



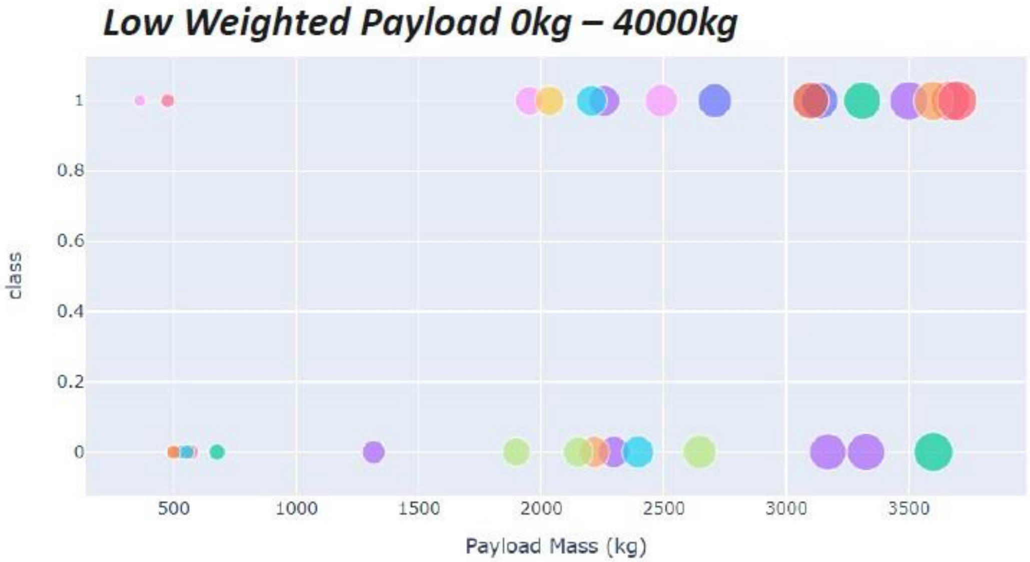
We can see that KSC LC-39A had the most successful launches from all the sites

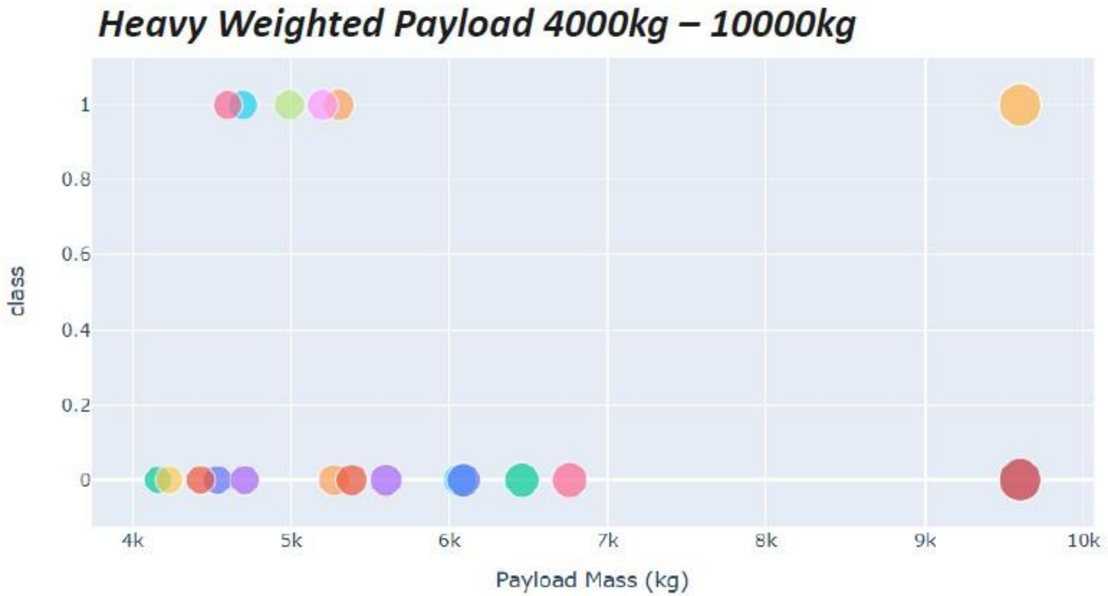
Success rate by site



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

Payload vs launch outcome



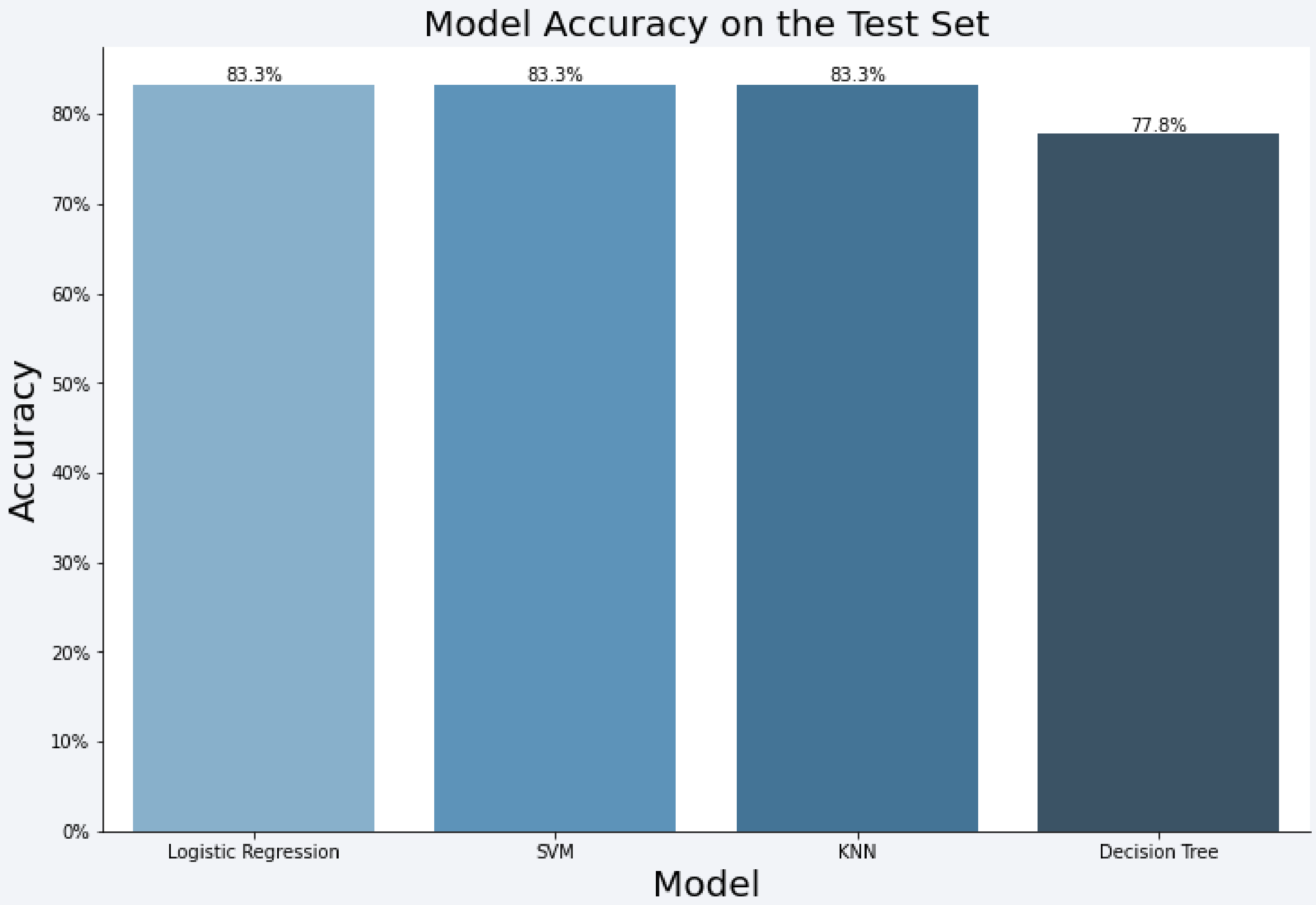


We can see the success rates for low weighted payloads is higher than the heavy weighted payloads

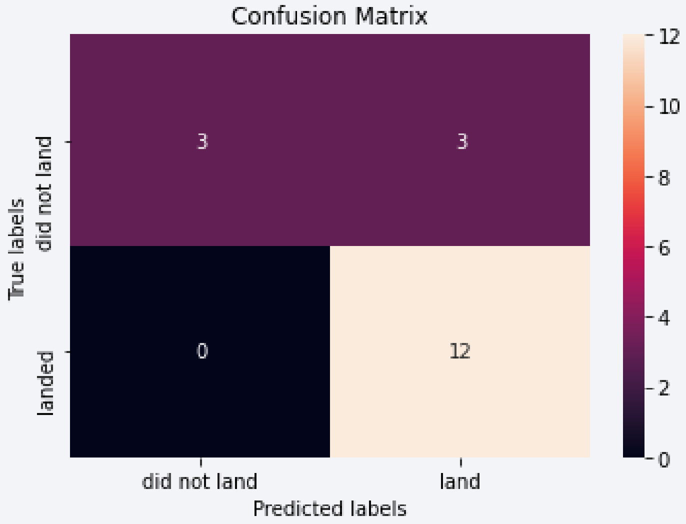
Section 6

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.

Appendix

- **Github project :**

  https://github.com/sudhanshu-prakash/Spacex-Data-Science-Project

Thank you!