IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Carl A. Gordon
October 2022

Outline

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

Executive Summary

- Summary of methodologies
  - Data collection
  - Data wrangling
  - Exploratory data analysis(EDA) with data visualization
  - EDA with SQL
  - Building an interactive map with Foilum
  - Building a Dashboard with Plotly Dash
  - Predictive analysis (Classification)
- Summary of all results
  - EDA Results
  - Interactive Analytics
  - Predictive analysis

Introduction

- Project background and context
  - Falcon 9 rockets, manufactured by SpaceX are much cheaper to produce when compared to other rocket manufacturers. The cost to produce rockets can reach up to $165 million, but Falcon 9 rockets are advertised to cost $62 million.
  - Much of these savings can be attributed to SpaceX having the ability to reuse the first stage of the rocket launch
- Problems you want to find answers
  - The task of this project is to create a predictive model that will determine if if the first stage of a SpaceX Falcon 9 rocket launch will land successfully

Section 1

Methodology

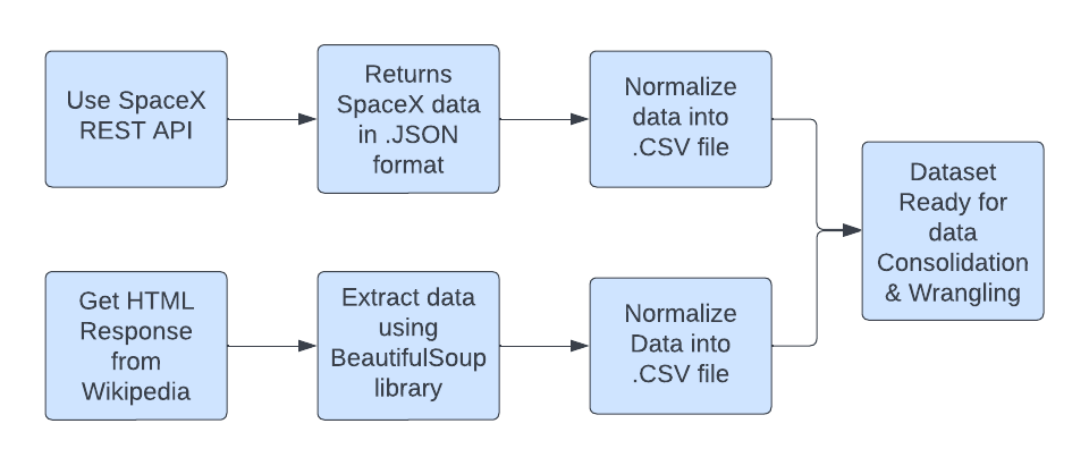
Methodology

Executive Summary

- Data collection methodology:
  - SpaceX REST API
  - Web scraping Falcon 9 information from Wikipedia
- Perform data wrangling
  - One Hot Encoding data fields for Machine Learning and data cleaning of null 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
  - Logistic Regression, K-Nearest Neighbors, Support Vector Machine, and Decision Tree models have been built and evaluated to determine the best classifier

Data Collection

- The following datasets were collected:
  - SpaceX launch data was gathered using the SpaceX REST API.
    - This provided launch information, such as rocket used, payload deliver, launch specification, landing specifications, and landing outcome.
  - Falcon 9 launch information was also scraped from its Wikipedia page using the BeautifulSoup Python library.



Data Collection – SpaceX API

- Data collection with SpaceX REST calls
  1. Getting Response from API
  2. Converting Response to a .json file
  3. Apply custom functions to clean data
  4. Assign list to dictionary, then to dataframe
  5. Filter dataframe and export to flat file (.csv)

[SpaceX API calls notebook]

1
```
spacex_url="https://api.spacexdata.com/v4/launches/past"
response = requests.get(spacex_url)
```

2
```
# Use json_normalize method to convert the json result into a dataframe
data = pd.json_normalize(response.json())
```

3
```
# Call getBoosterVersion
getBoosterVersion(data)
# Call getLaunchSite
getLaunchSite(data)
# Call getPayloadData
getPayloadData(data)
# Call getCoreData
getCoreData(data)
```

4
```
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}
```

5
```
data_falcon9 = data[data.BoosterVersion == 'Falcon 9']
data_Falcon9.to_csv('dataset_part_1.csv', index=False)
```

Data Collection - Scraping

- Falcon 9 rocket information scraped from Wikipedia page using BeautifulSoup
    1. Getting Response from HTML
    2. Creating BeautifulSoup Object
    3. Finding Tables
    4. Getting Column names
    5. Creation of dictionary
    6. Appending Data to keys (block 13 of notebook)
    7. Converting dictionary to dataframe
    8. Dataframe to .CSV

1
```
# use requests.get() method with the provided static_url
# assign the response to a object
data  = requests.get(static_url).text
```

2
```
# Use BeautifulSoup() to create a BeautifulSoup object from a response text content
soup = BeautifulSoup(data, 'html5lib')
```

3
```
# 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')
```

4
```
column_names = []
for row in first_launch_table.find_all('th'):
    name = extract_column_from_header(row)
    if (name != None and len(name) > 0):
        column_names.append(name)
```

5
```
launch_dict= dict.fromkeys(column_names)

# Remove an irrelvant column
del launch_dict['Date and time ( )']
# Let's initial the launch_dict with each value to be an empty list
launch_dict['Flight No.'] = []
launch_dict['Launch site'] = []
launch_dict['Payload'] = []
launch_dict['Payload mass'] = []
launch_dict['Orbit'] = []
launch_dict['Customer'] = []
launch_dict['Launch outcome'] = []
# Added some new columns
launch_dict['Version Booster']=[]
launch_dict['Booster landing']=[]
launch_dict['Date']=[]
launch_dict['Time']=[]
```

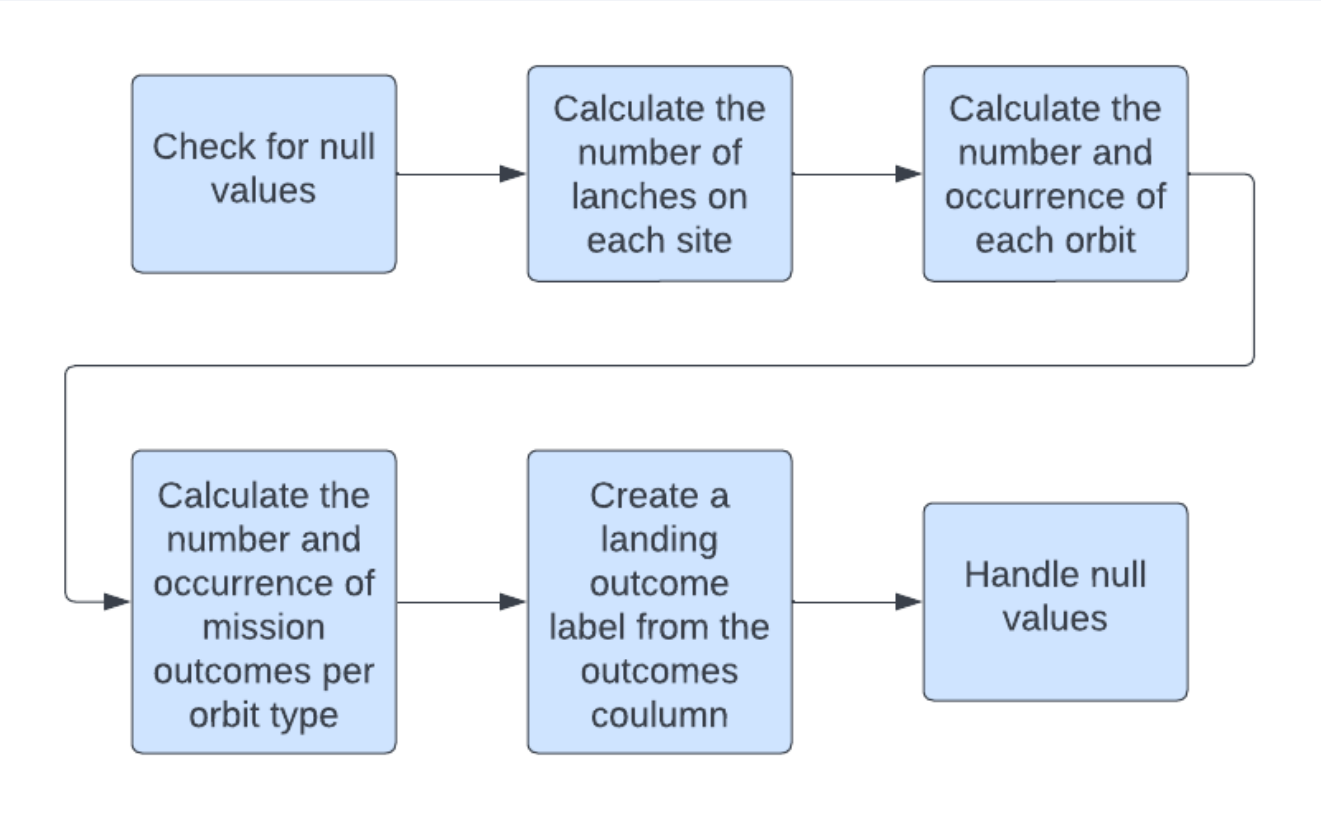
6
```
extracted_row = 0
#Extract each table
for table_number,table in enumerate(soup.find_all('table',"wikitable plainrowheaders collapsible")):
   # get table row
    for rows in table.find_all("tr"):
        #check to see if first table heading is as number corresponding to launch a number
        if rows.th:
            if rows.th.string:
                flight_number=rows.th.string.strip()
                flag=flight_number.isdigit()
```

7
```
df = pd.DataFrame.from_dict(launch_dict)
```

8
```
df.to_csv('spacex_web_scraped.csv', index=False)
```

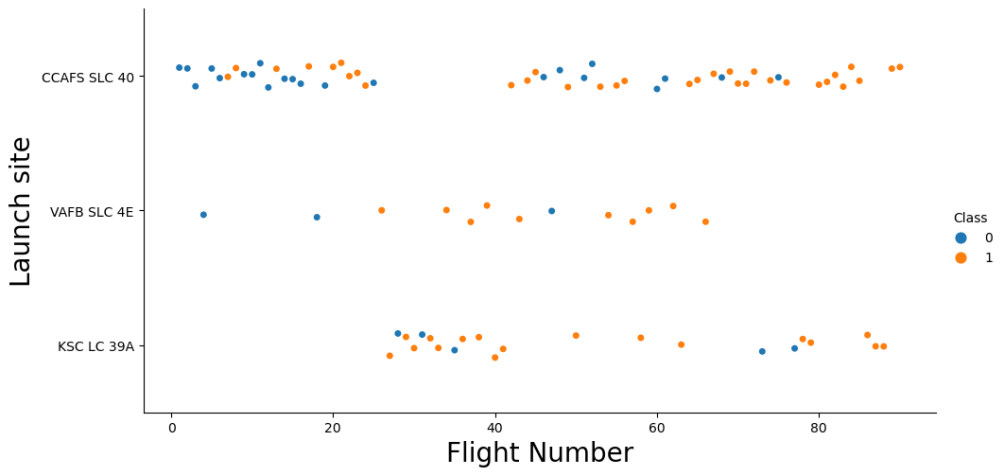
[web scraping notebook]

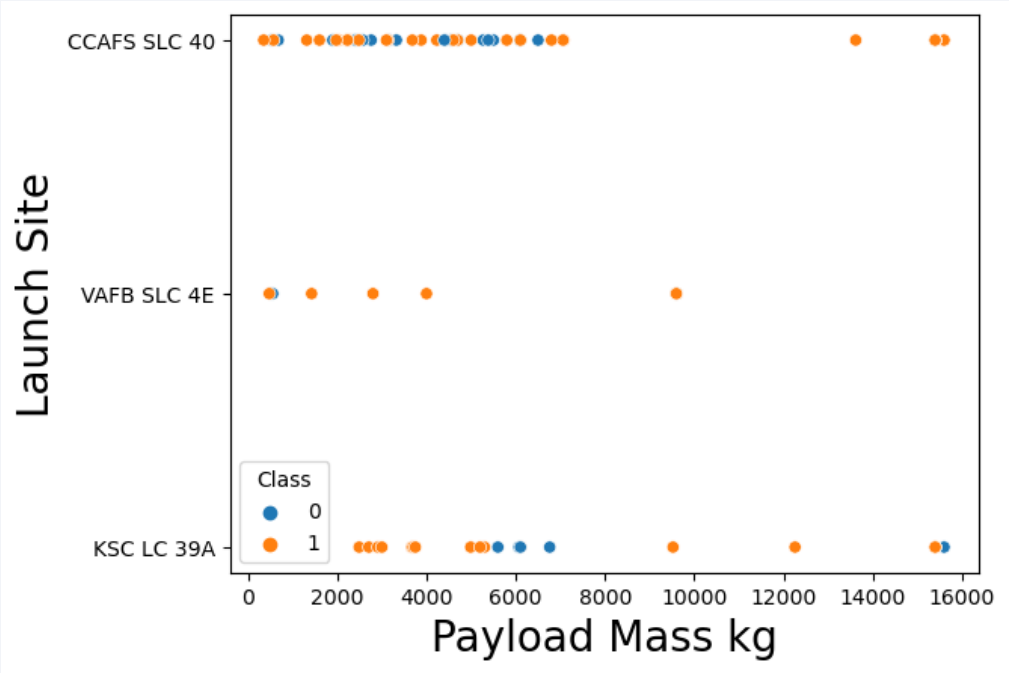
Data Wrangling



Data wrangling notebook

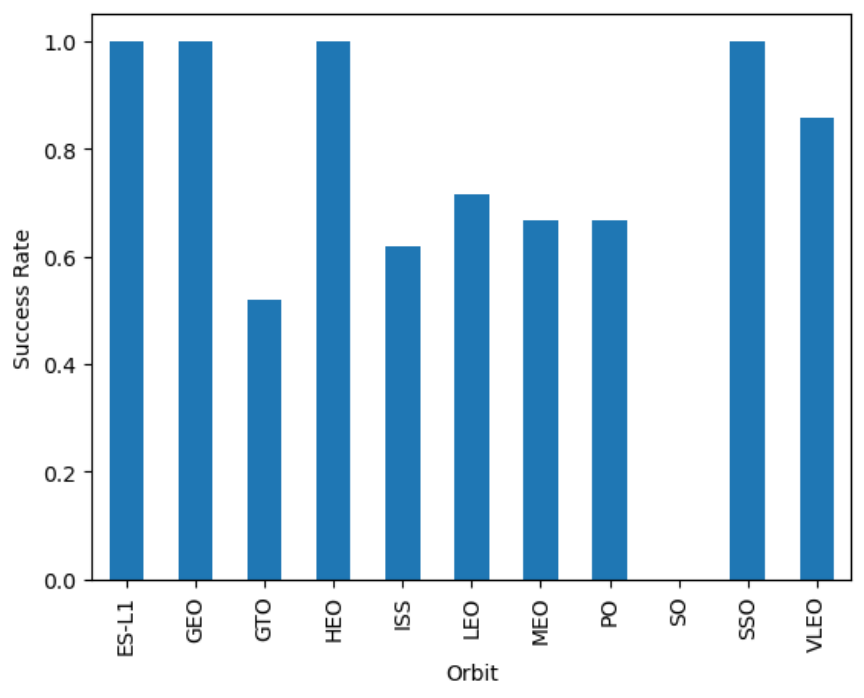
EDA with Data Visualization

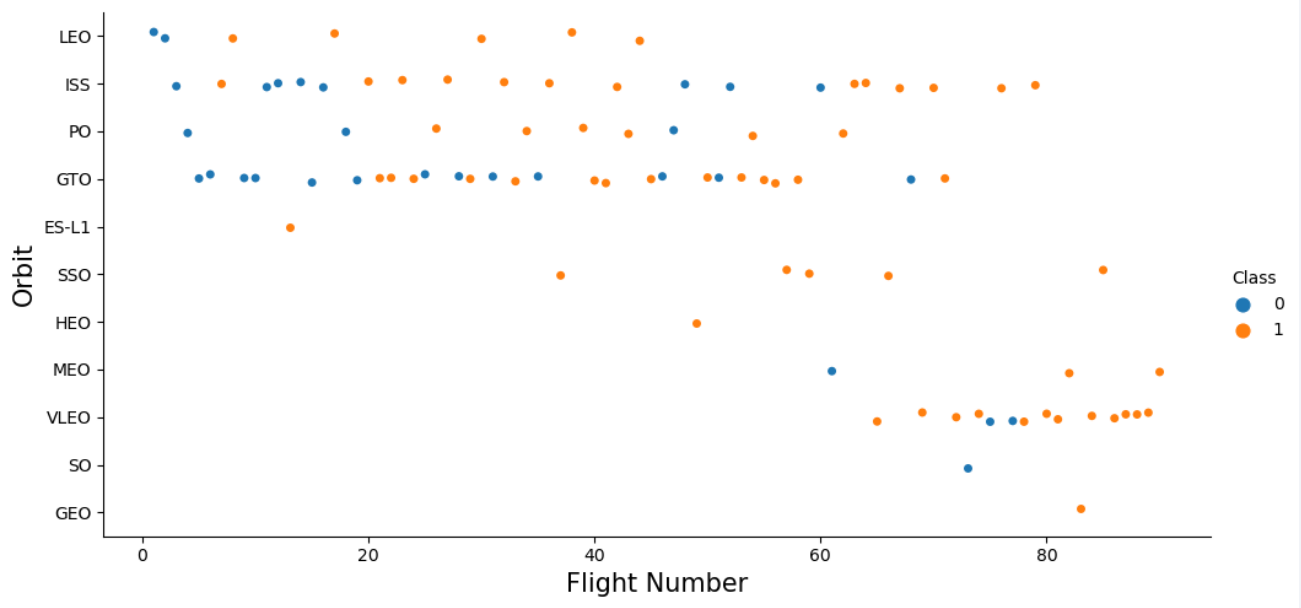




- [EDA with data visualization notebook]

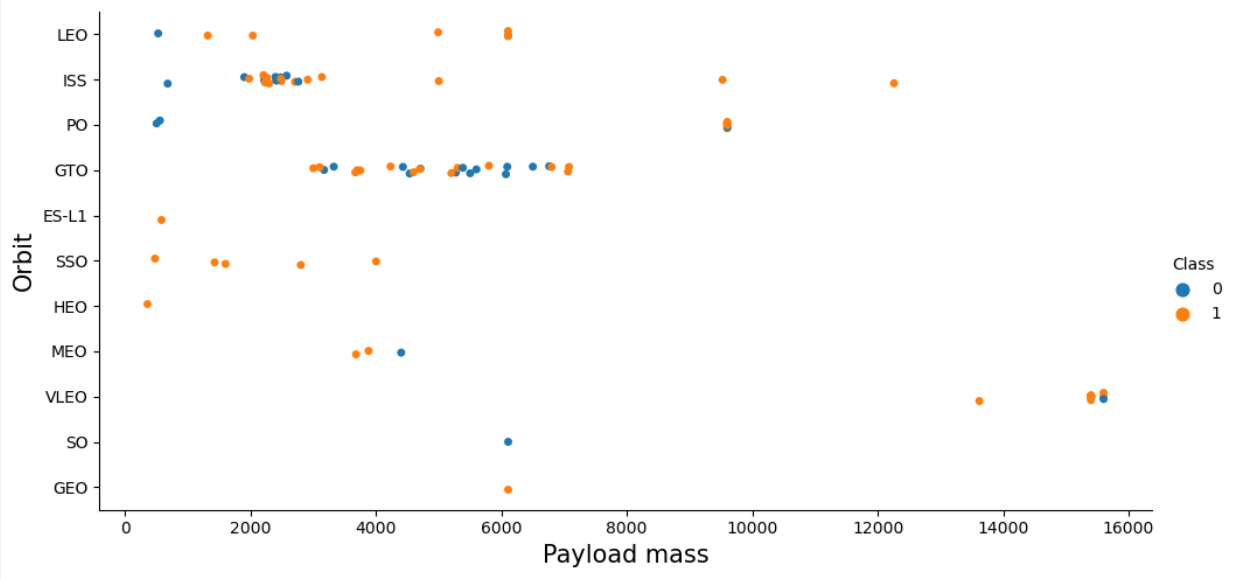
EDA with Data Visualization

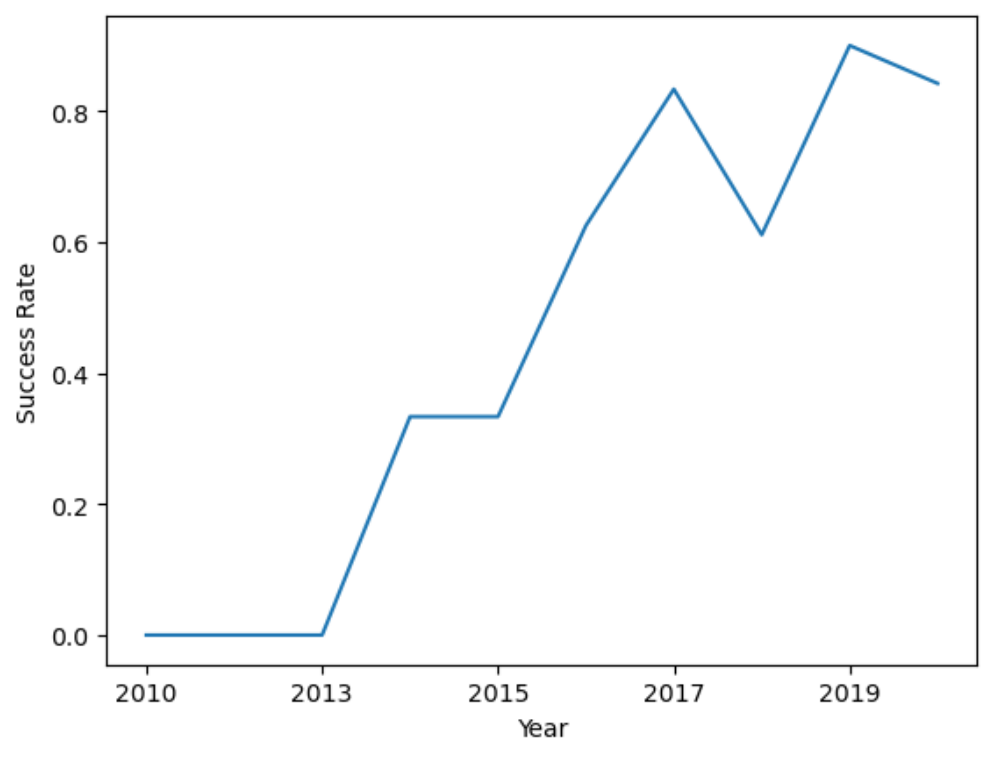




- [EDA with data visualization notebook]

EDA with Data Visualization





- [EDA with data visualization notebook]

EDA with SQL

- SQL queries performed:
  - Display the names of the unique launch sites in the space mission
  - Display 5 records where launch sites begin with the string ‘KSC’
  - Display the total payload mass carried by boosters launched by NASA (CRS)
  - Display average payload mass carried by booster version F9 v1.1
  - Listing the date where the successful landing outcome in drone ship was achieved
  - List the names of the boosters which have success in ground pad and have payload mass between 4000 and 6000 kilograms

EDA with SQL notebook

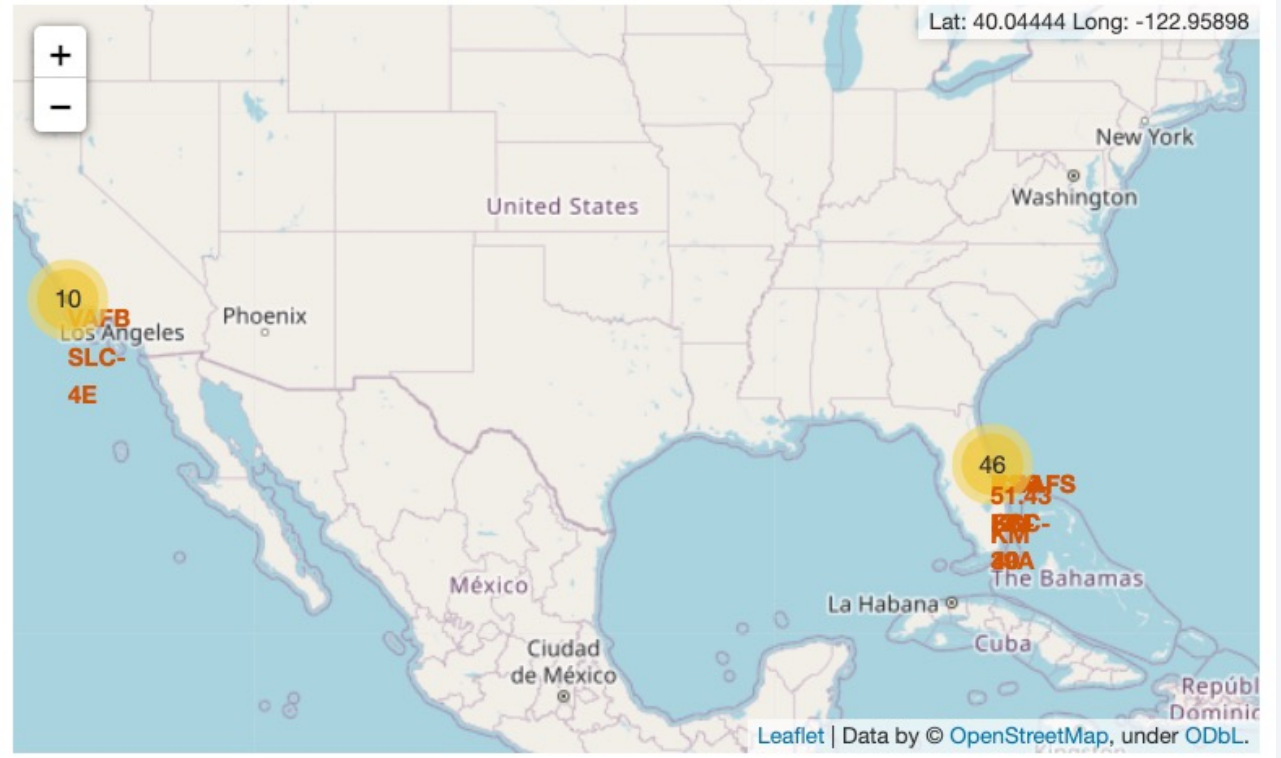
EDA with SQL

- SQL queries performed (cont.):
  - List the total number of successful and failure mission outcomes
  - List the names of the booster versions which have carried the maximum payload mass
  - List the records which will display the month names, successful landing outcomes in ground pad, booster versions, launch site for the months in 2015
  - Rank the count of successful landing outcomes between the dates 04-06-2010 and 03-20-2017 in descending order

EDA with SQL notebook

Build an Interactive Map with Folium

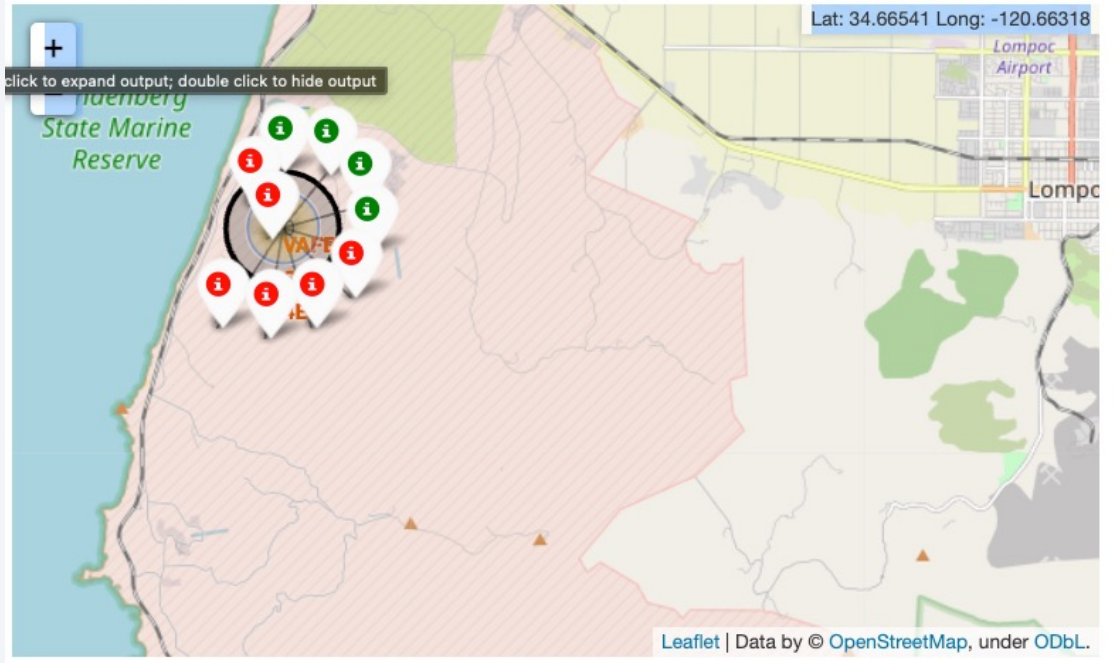
- Map markers were added to the map to find the optimal location for building a launch site based on success rate.



[Interactive map with Folium]

Build an Interactive Map with Folium

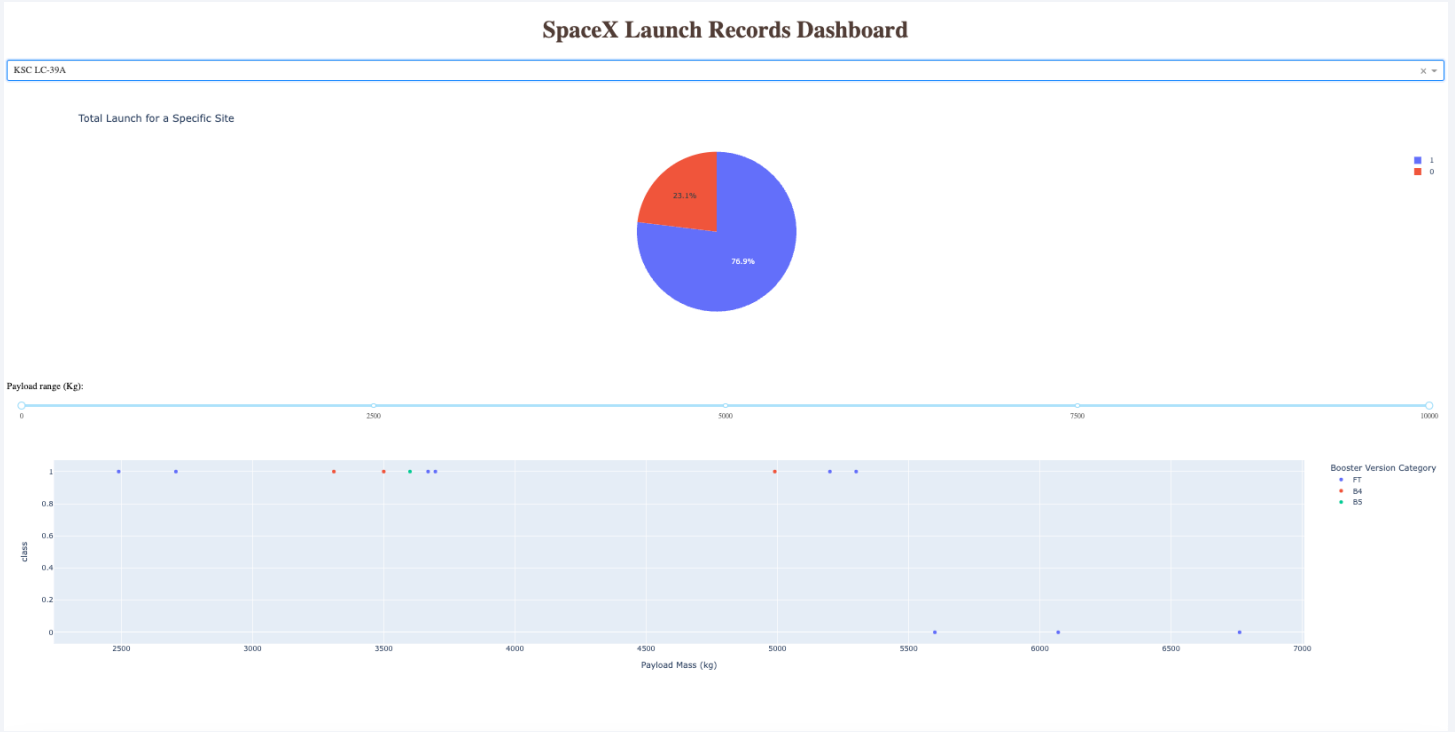
- Map markers indicate success or failure at a given launch site



[Interactive map with Folium]

Build a Dashboard with Plotly Dash

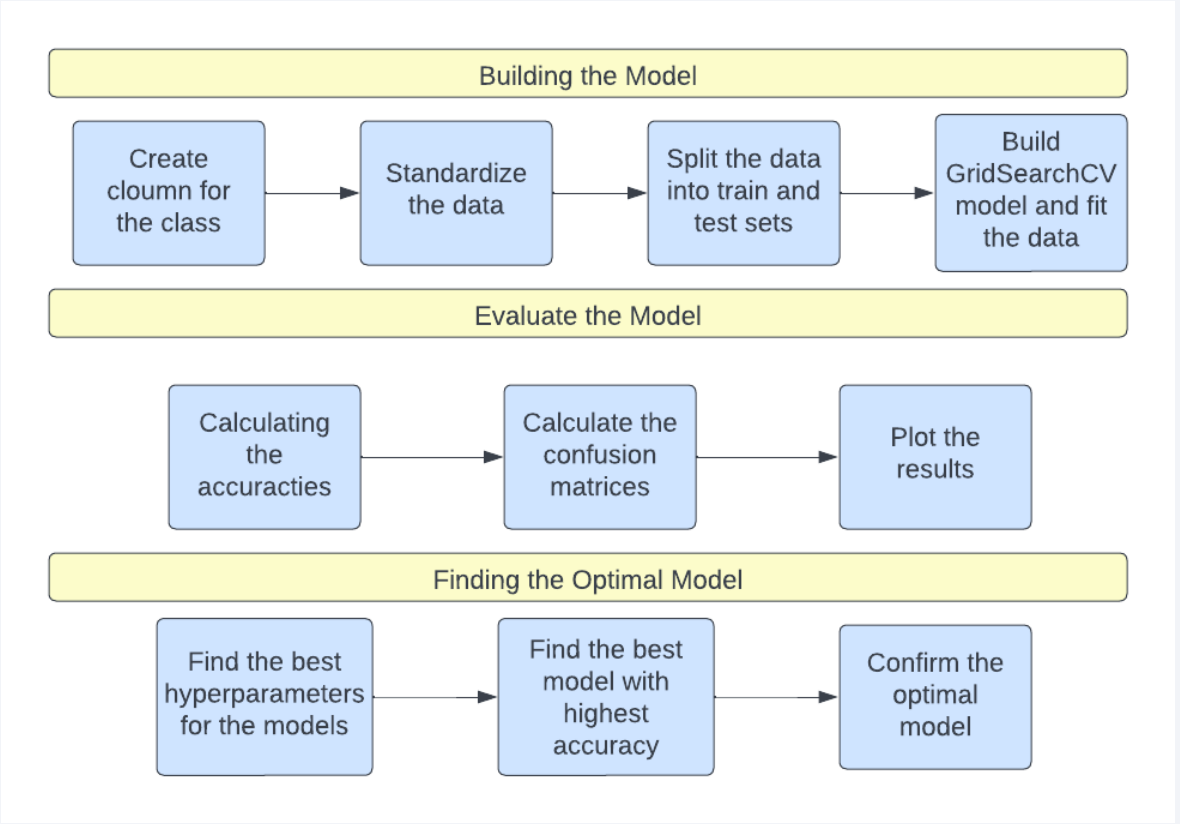
- Key Takeaways:
  - KSC LC-39A achieved a success rate of 76.9%
  - KSC LC-39A had the most successful launches when compared to all sits
  - The success Rate for low weight payloads were higher compared to heavy weighted payloads



[SpaceX Dashboard with Plotly Dash]

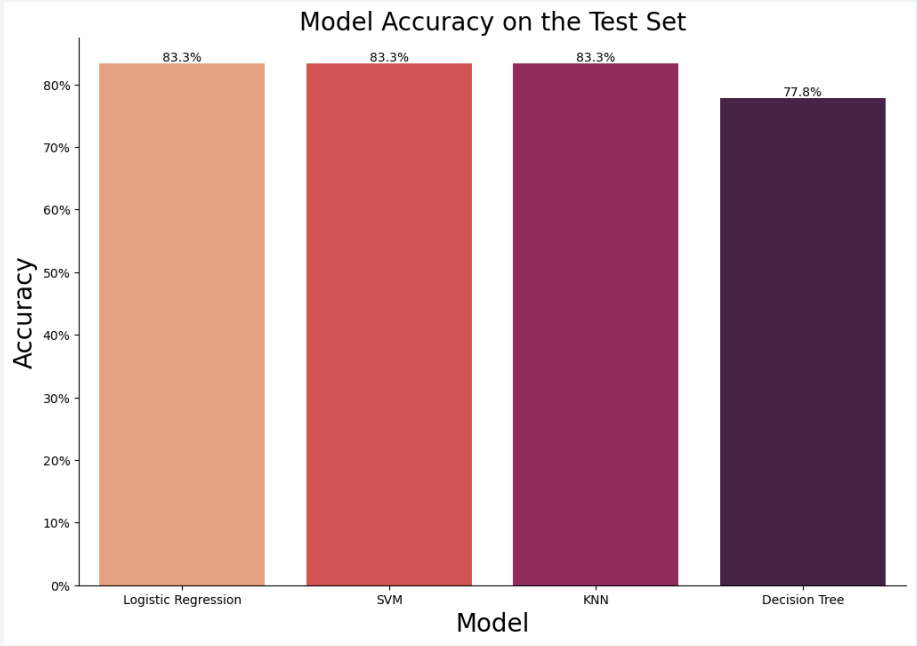
Predictive Analysis (Classification)

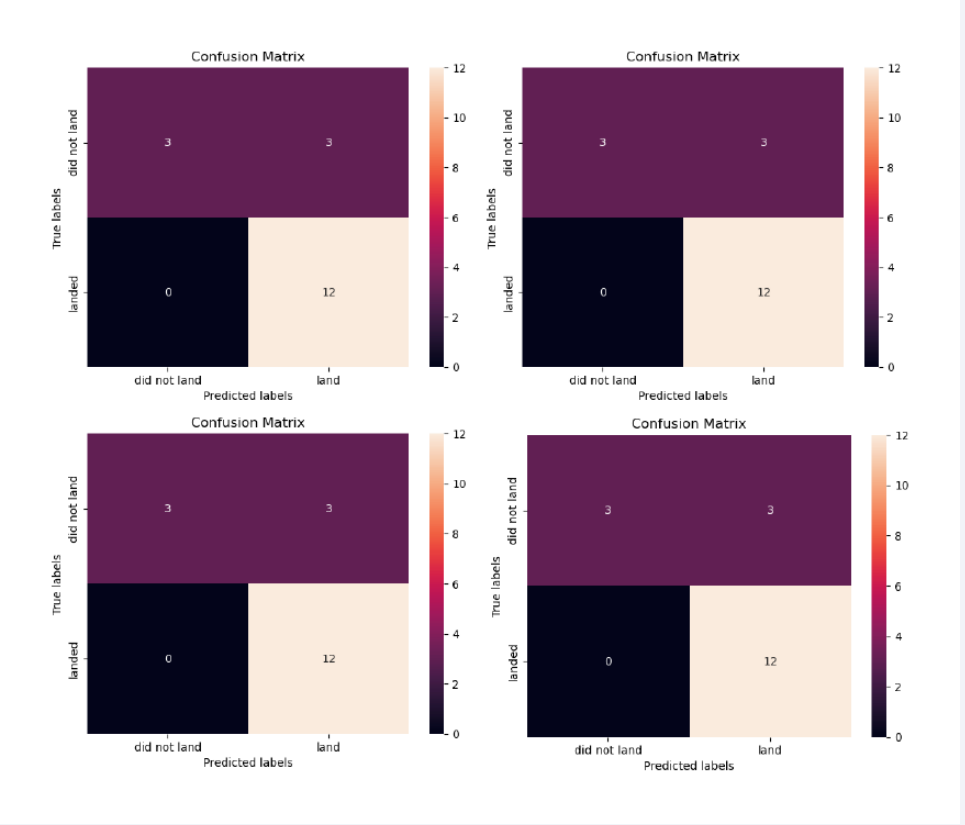
- Classification models were developed to determine the best predictor of knowing the success of the first launch landing, thus determining the cost of the launch
- The models produced were the following :
  - Logistic Regression
  - K-Nearest Neighbors (KNN)
  - Decision Trees
  - Support Vector Machines (SVM)



Predictive analysis lab

Predictive Analysis (Classification)

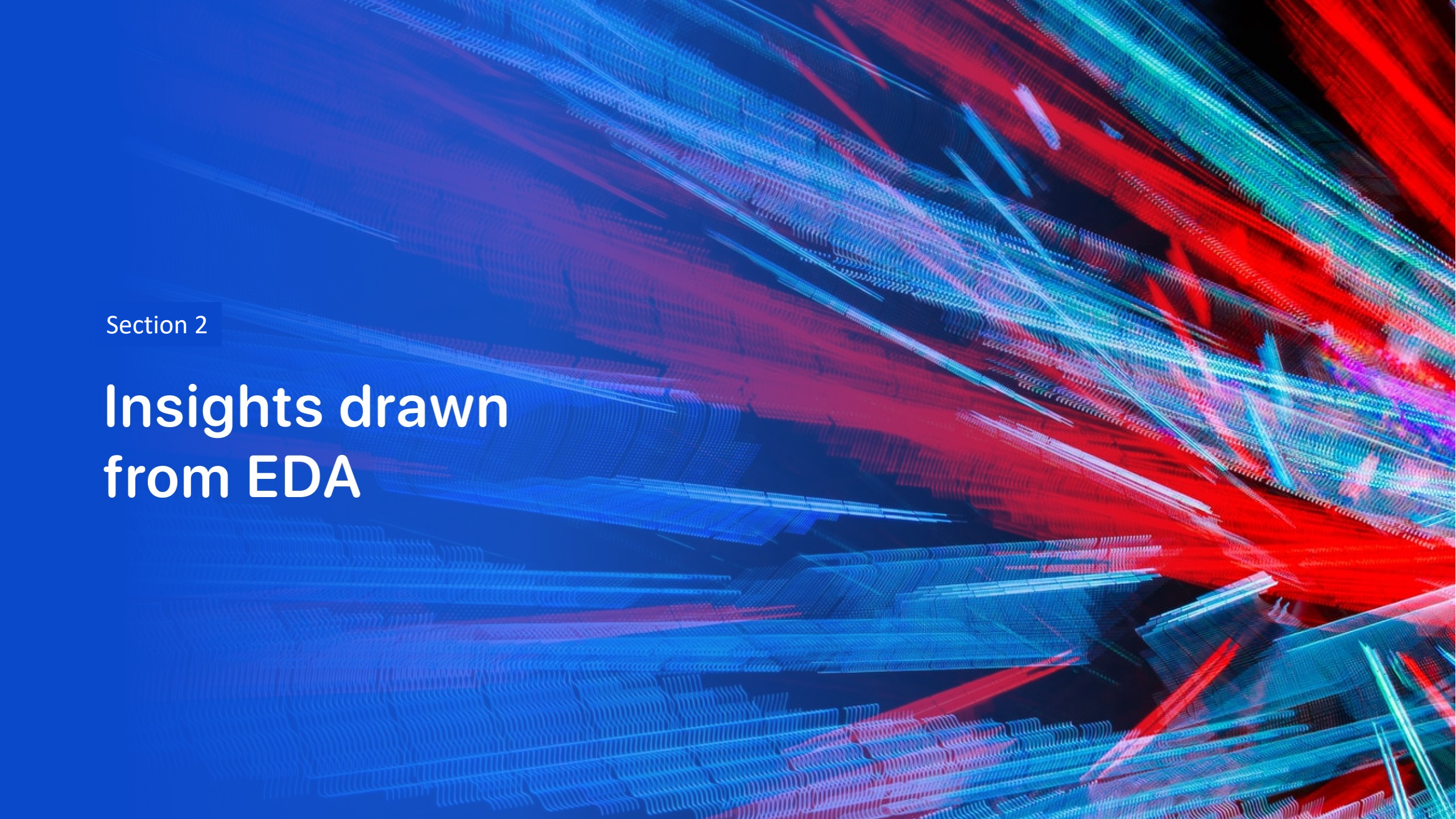




[Predictive analysis lab]

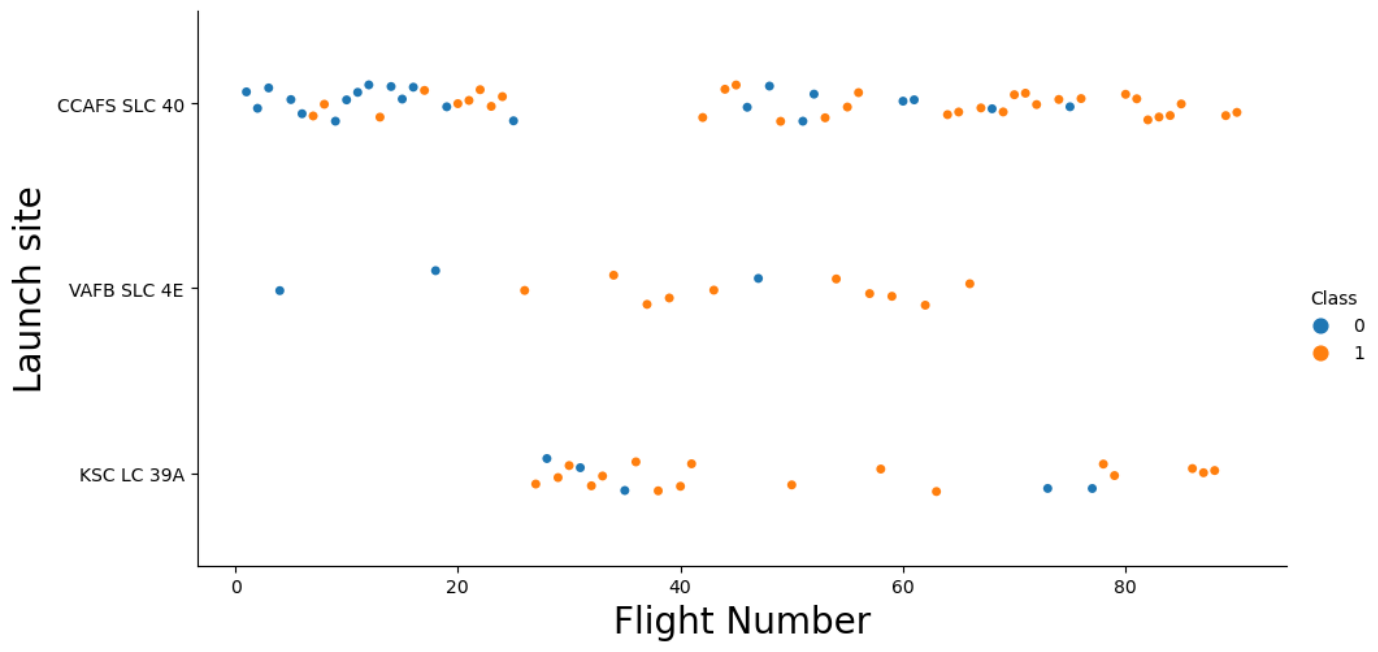
Results

- The Logistics Regression, SVM, and KNN 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
- GEO, HEO, SSO, ES L1 orbits had the best success rate.


Section 2
Insights drawn from EDA

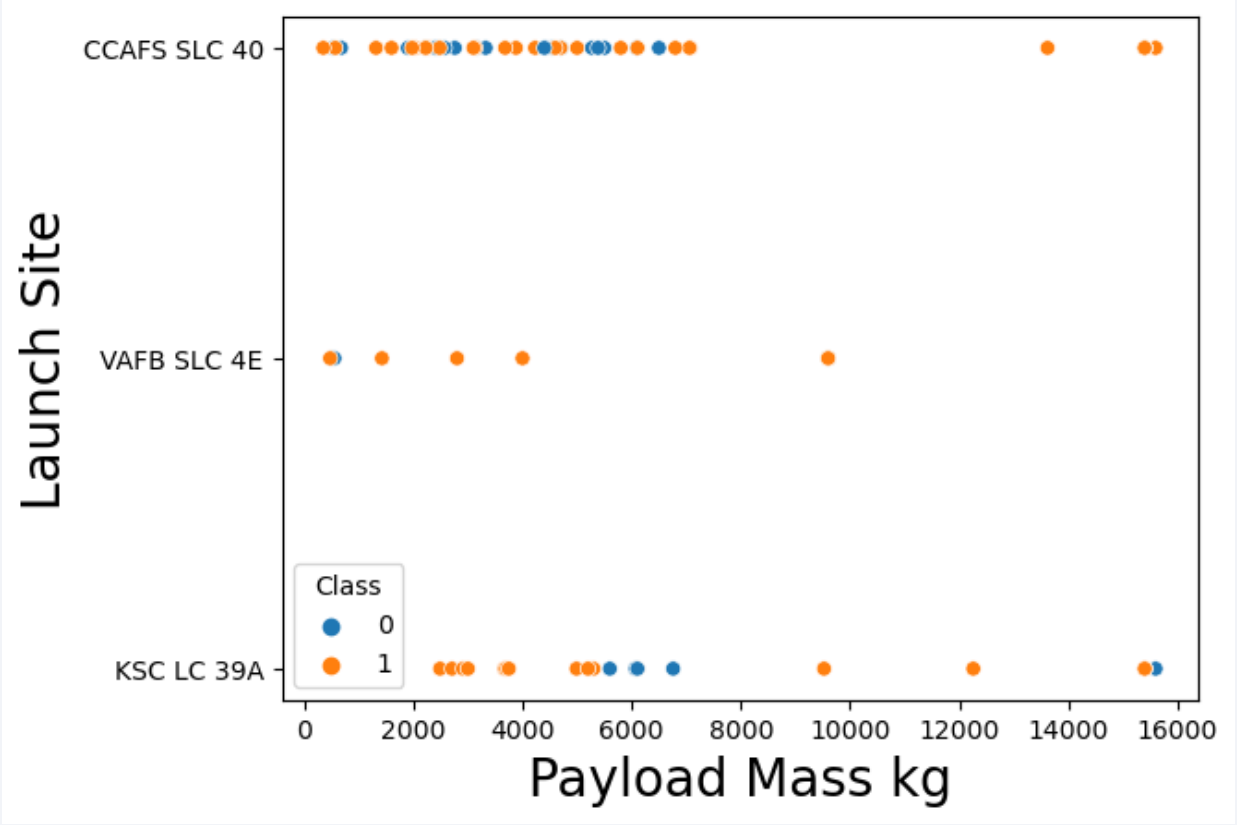
Flight Number vs. Launch Site

- Launches from the site of CCAFS SLC 40 occur more frequently when compared to other launch sites



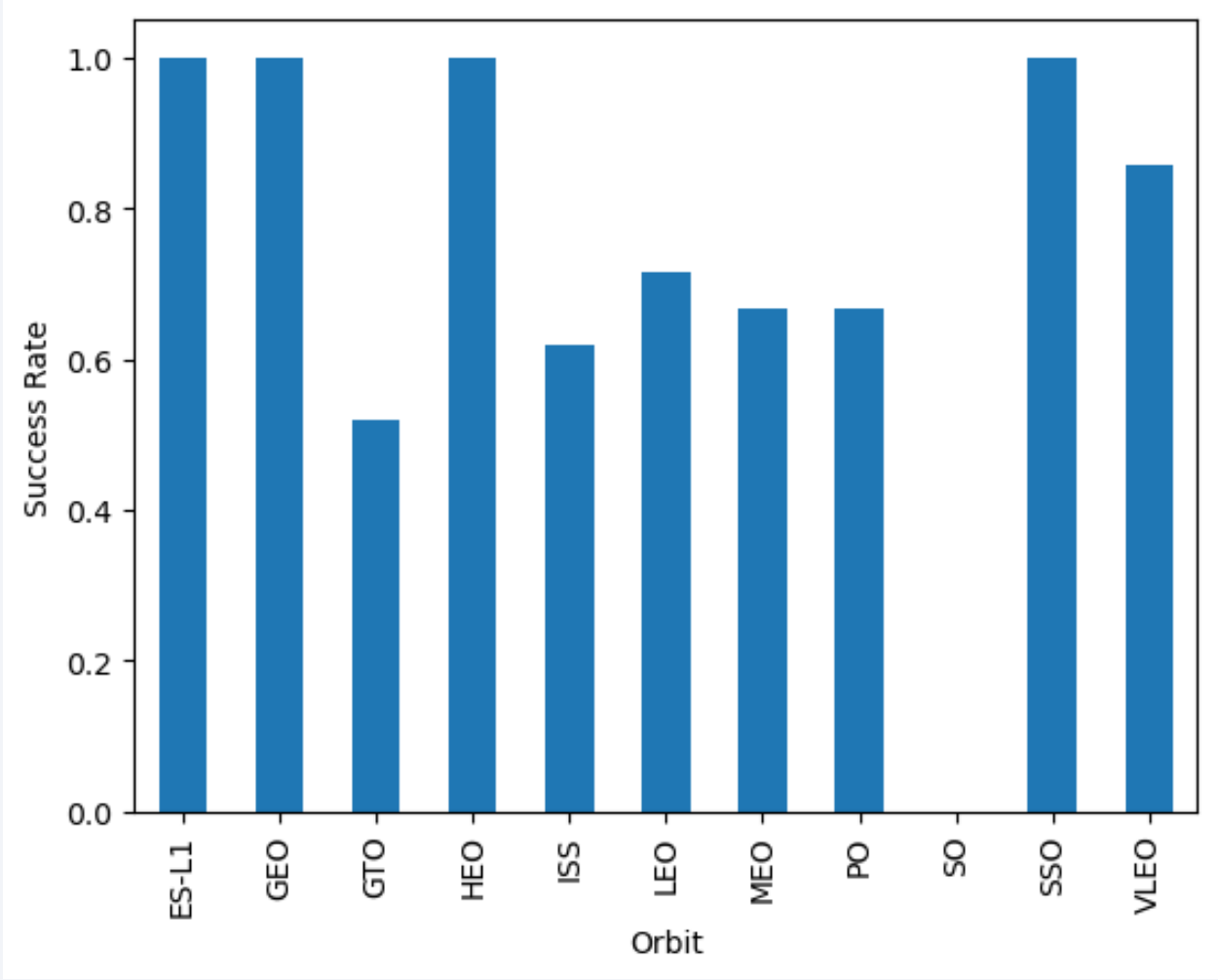
Payload vs. Launch Site

- Most of the payloads with lower mass have been launched from the CCAFS SLC 40 launch site



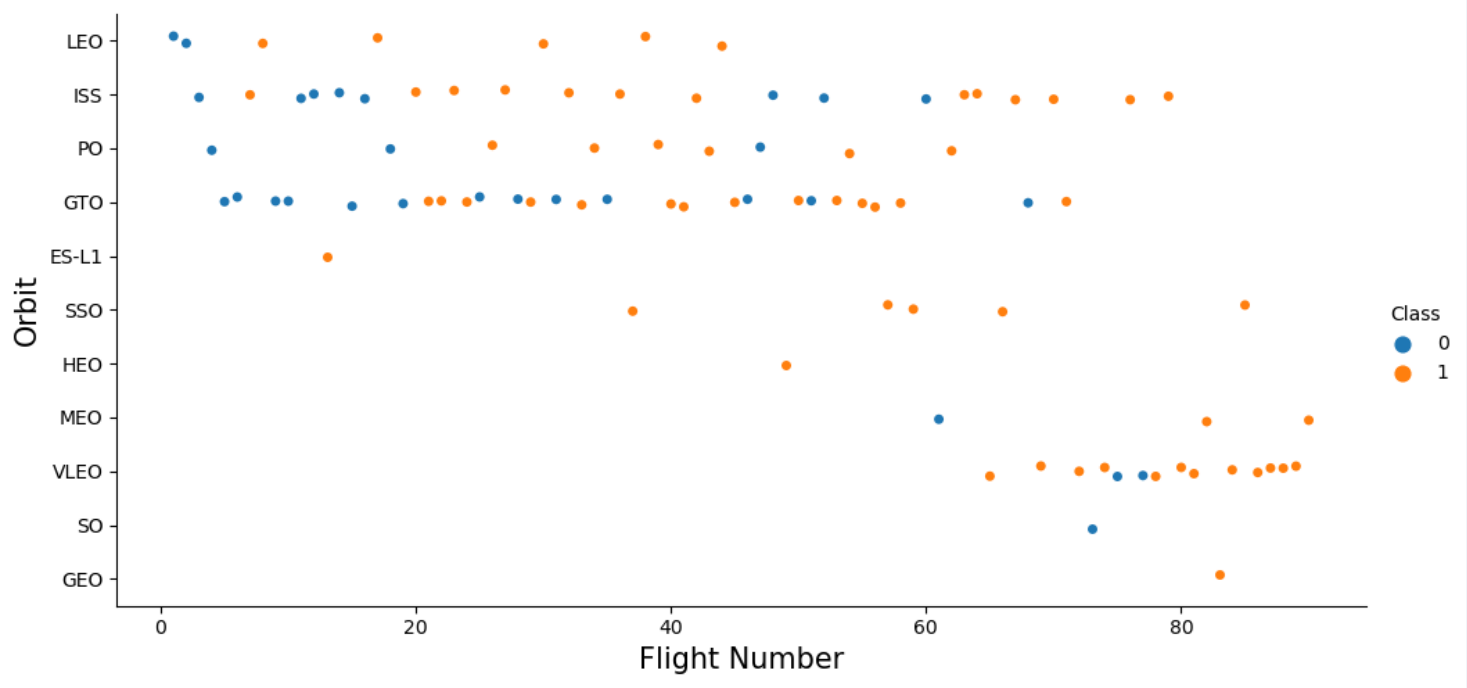
Success Rate vs. Orbit Type

- Success rates were high among the GEO, HEO, SSO, and ES-L1 orbit types.



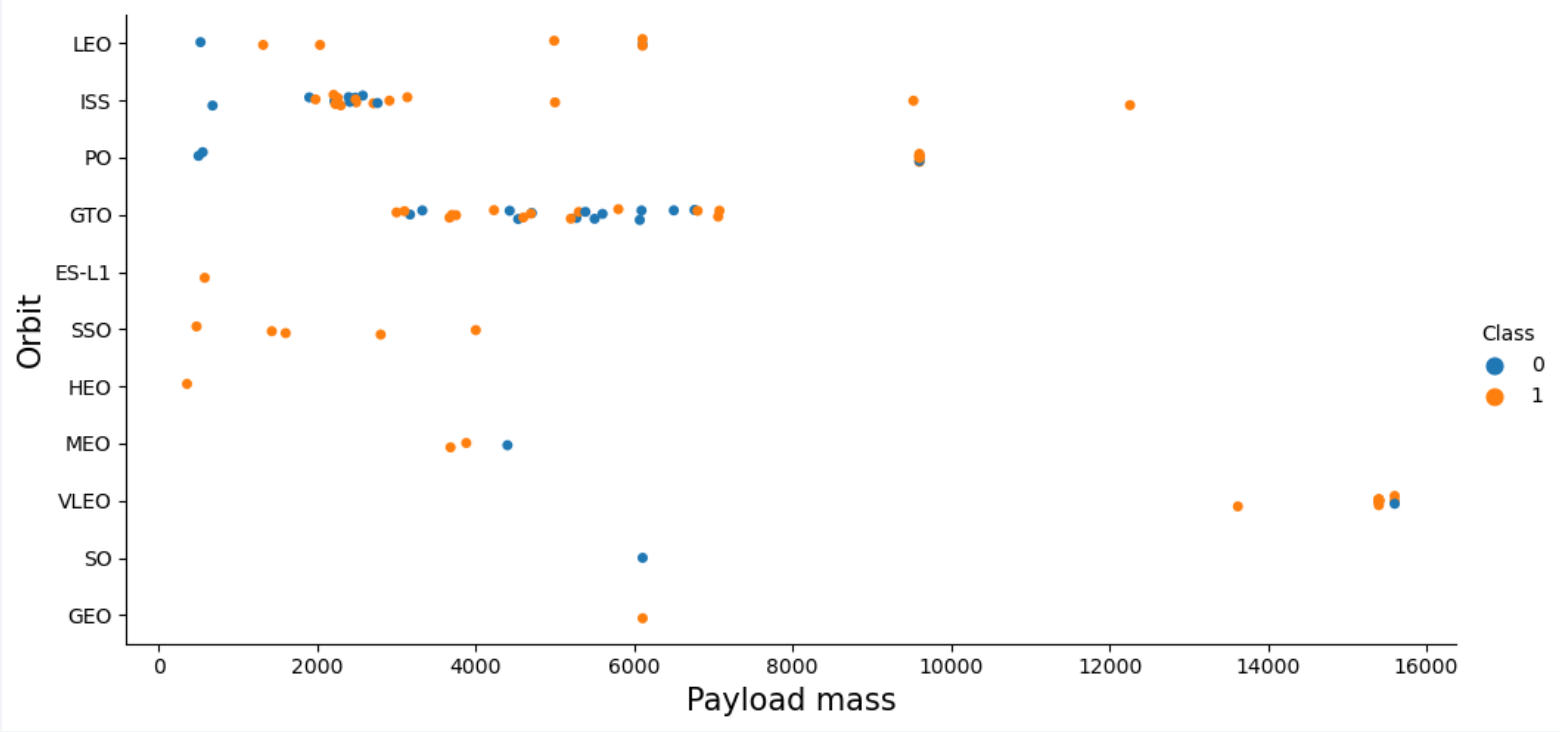
Flight Number vs. Orbit Type

- VLEO orbit type launches have occurred more frequently in recent launches.



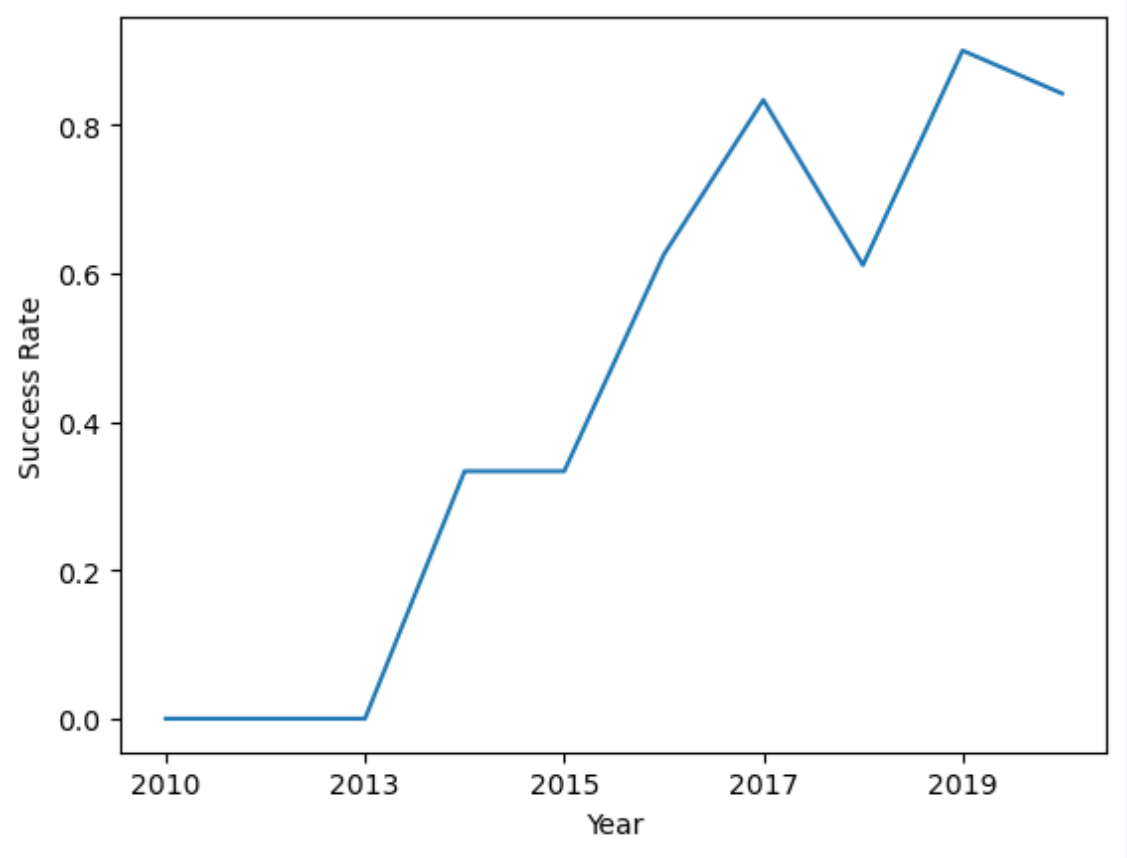
Payload vs. Orbit Type

- There are successful launches with rockets in the ISS orbit type around the 2000-4000kg payload mass, and with rockets in the GTO orbit type with a payload in the 4000-8000kg
- There is a positive landing rate for Polar, LEO, and ISS at heavy payloads



Launch Success Yearly Trend

- Since 2013, launch success rate increased and has stabilized in 2019
- This is potentially due to lessons learned from launches and technological advancements



All Launch Site Names

- Displaying the names of the unique launch sites in the space mission.

Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`
- The query result returned the CCAFS LC-40 launch site

Launch_Site
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40

Total Payload Mass

- Calculate the total payload carried by boosters from NASA
- The query result returns the total payload mass carried by boosters launched by CRS, totaling to 619,967kg

```
%sql select sum(PAYLOAD_MASS__KG_) as payloadmass from SPACEXTBL;
```

```
 * sqlite:///my_data1.db
Done.
```

payloadmass
619967

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 is 6,138kg

```
%sql select avg(PAYLOAD_MASS__KG_) as payloadmass from SPACEXTBL;
```

```
 * sqlite:///my_data1.db
Done.
```

payloadmass
6138.287128712871

First Successful Ground Landing Date

- The dates of the first successful landing outcome on ground pad is January 3rd, 2013

```
%sql select min(DATE) from SPACEXTBL;
```

```
 * sqlite:///my_data1.db
Done.
```

min(DATE)
01-03-2013

Successful Drone Ship Landing with Payload between 4000 and 6000

- The names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000 are the following:
  - F9 FT B1022
  - F9 FT B1026
  - F9 FT B1021.2
  - F9 FT B1031.2

booster_version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- There are 101 successful and failure mission outcomes with Falcon 9 rockets

Mission_Outcome	count
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- The names of the booster which have carried the maximum payload mass are given by the provided screenshot

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

2015 Launch Records

- The list of failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015 is given by the provided screenshot

booster_version	launch_site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

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

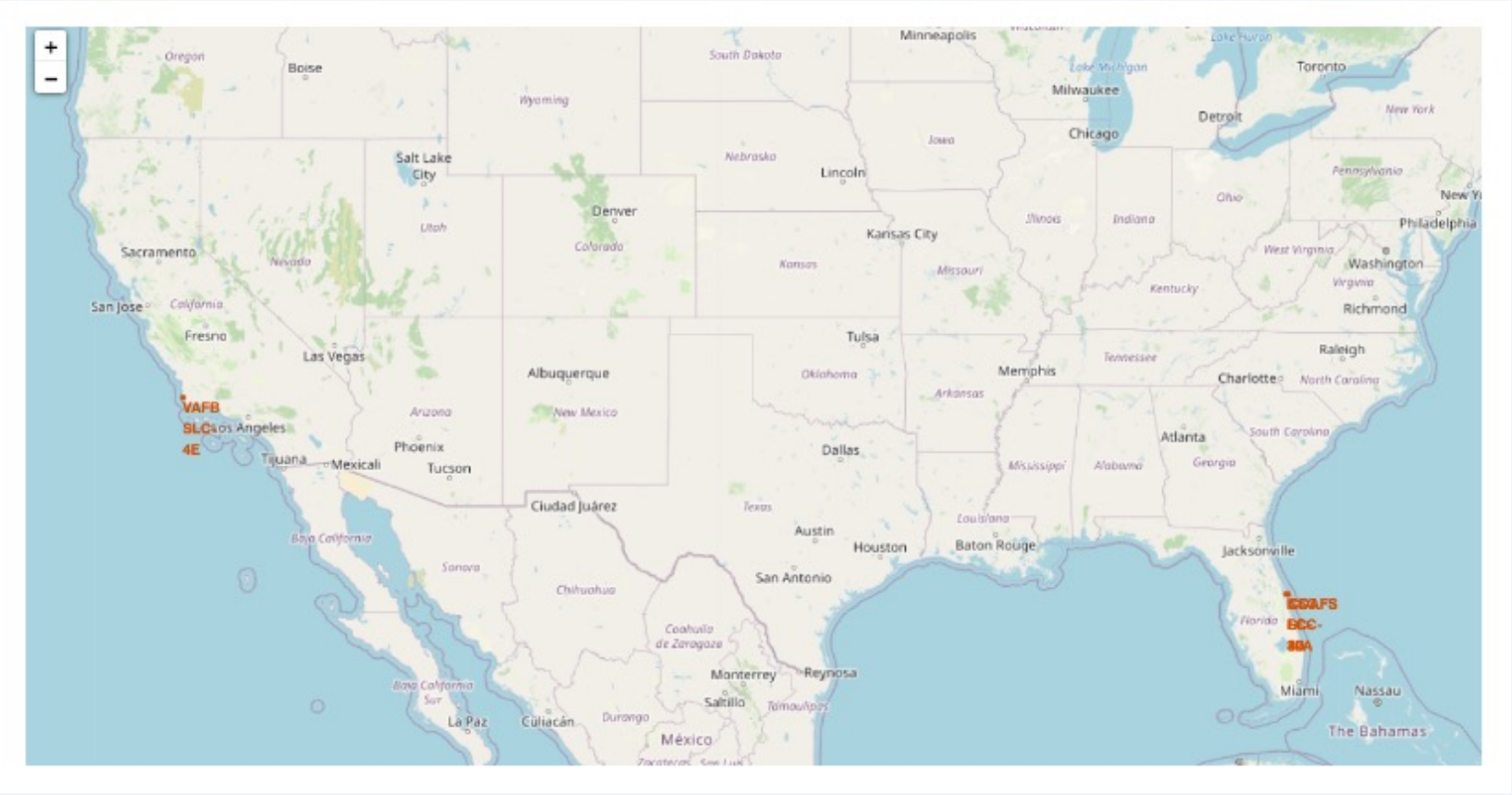
- The count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, is provided in descending order in the screenshot below.

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

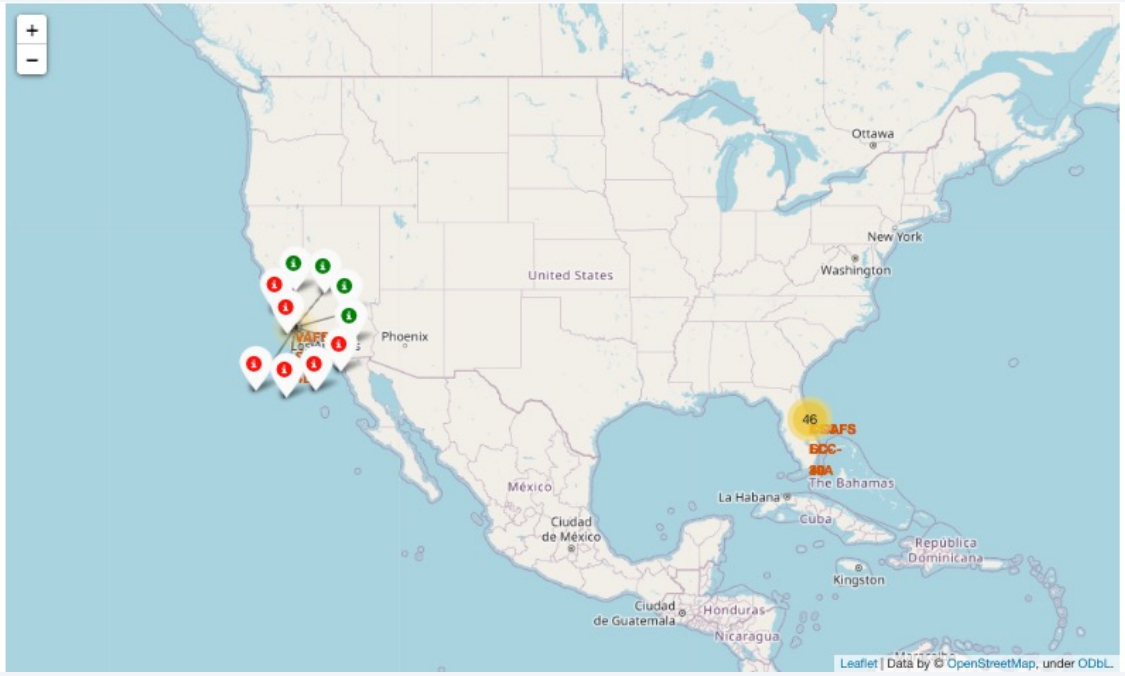
Rocket Launch Sites

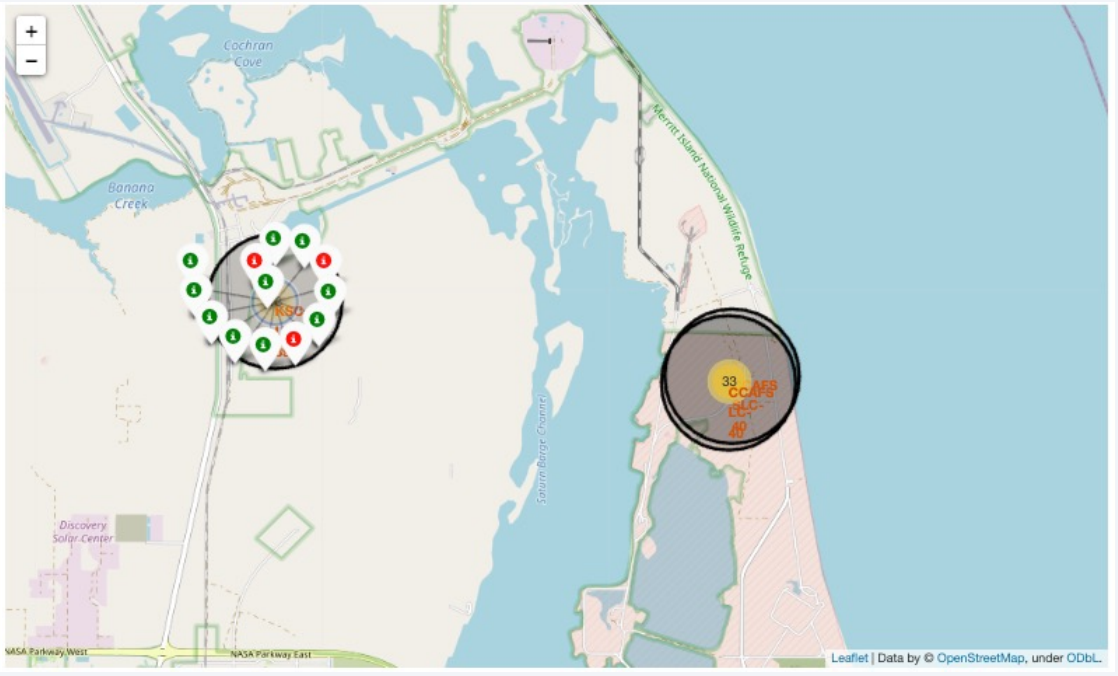
Using Folium, these are the locations of the rocket launch sites in California and Florida



Launch Sites with Color Markers

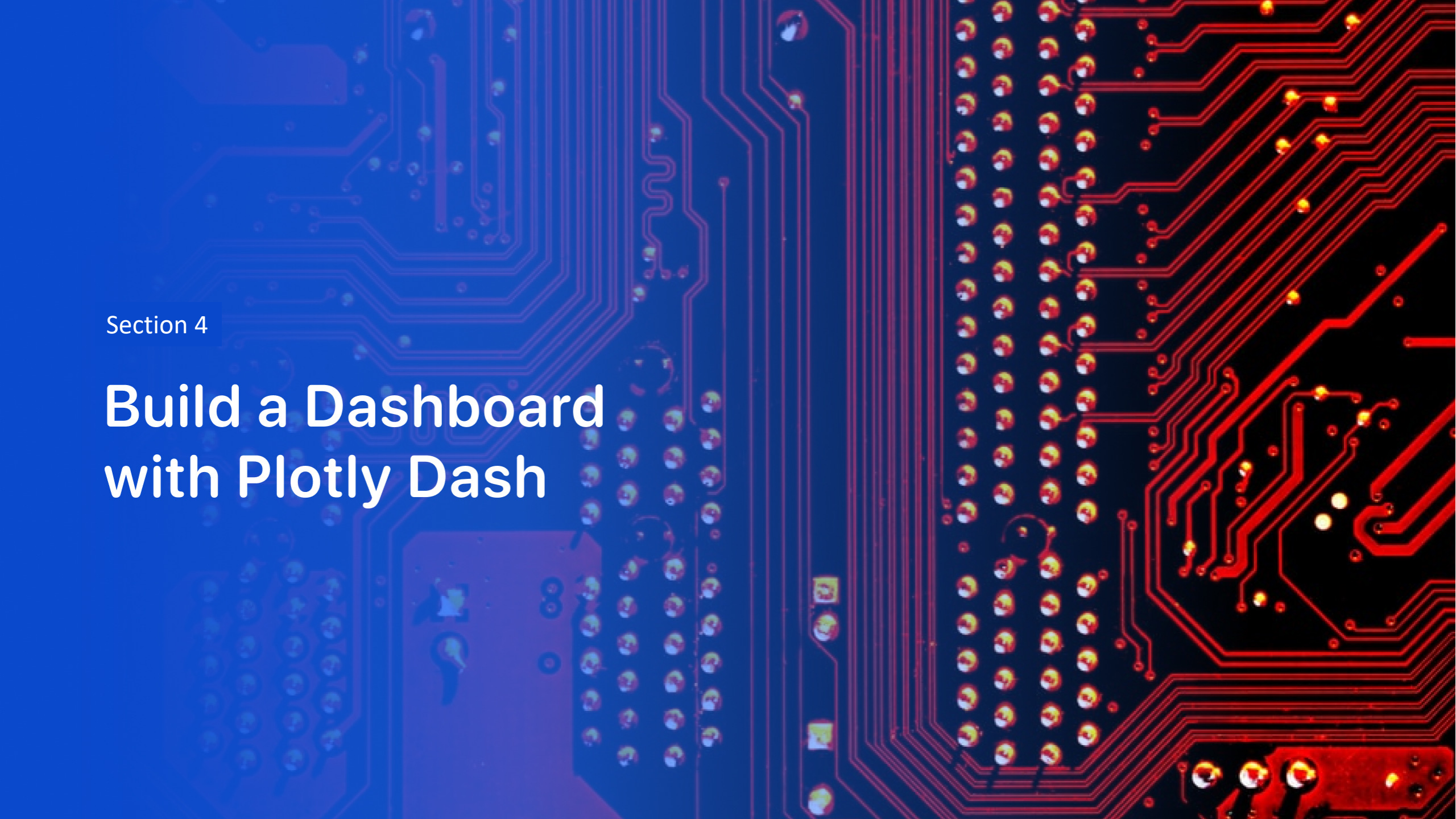
- At each launch site, color markers indicate the **success** or **failure** of a rocket launch





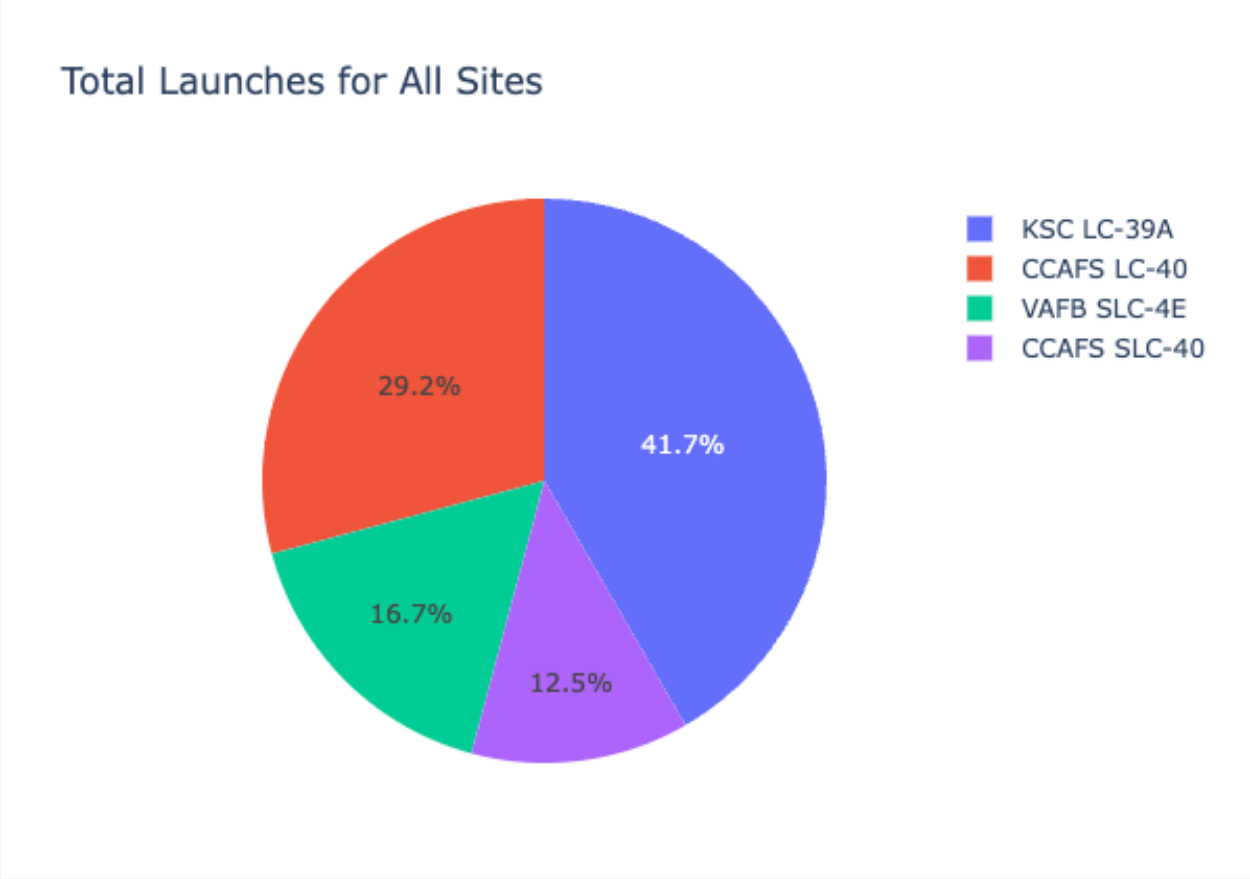
Launch Site with Proximity to Coastline

- Plotline showing the CCAFS SLC-40 launch site 0.86kg away from the coastline


Section 4
Build a Dashboard with Plotly Dash

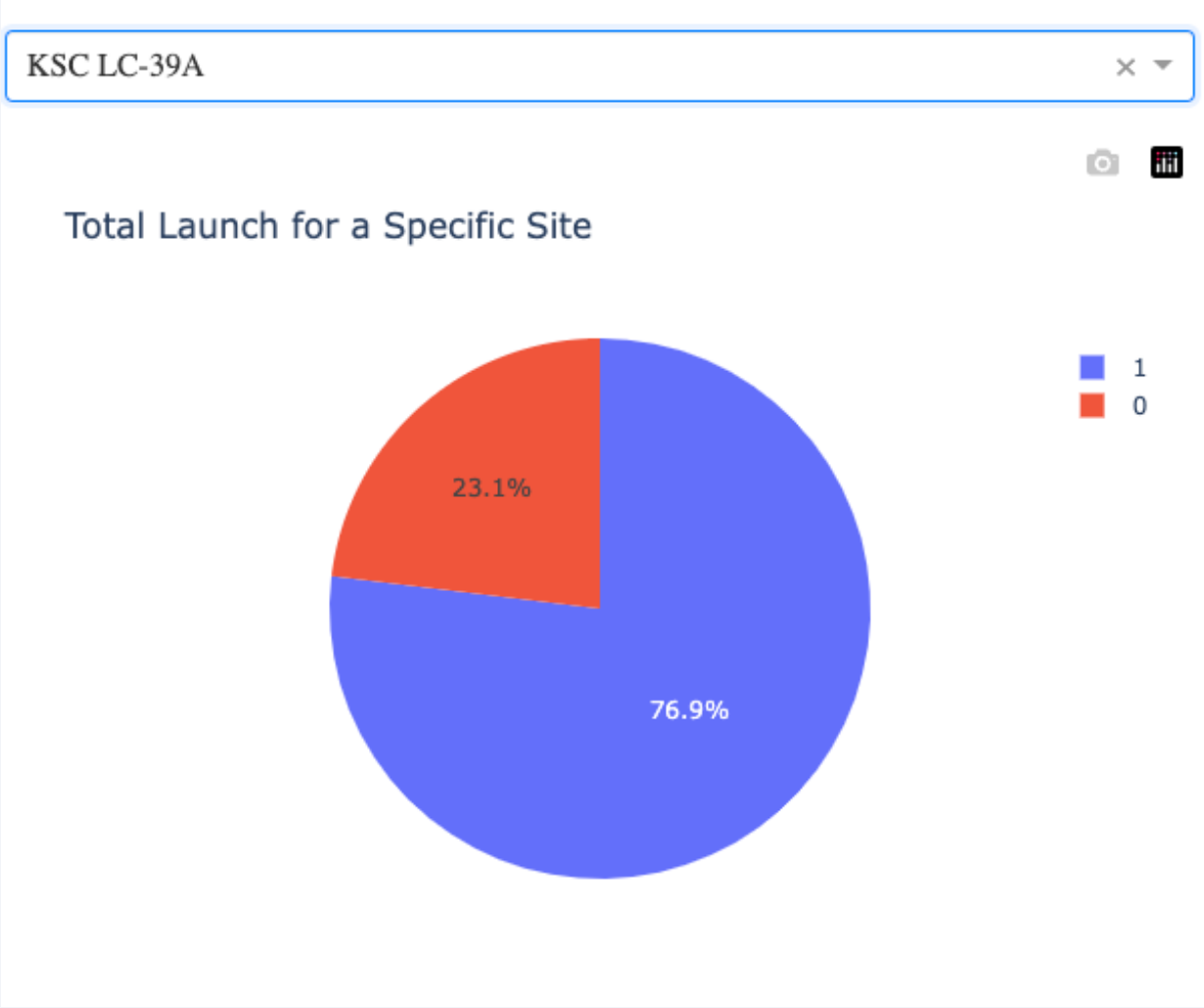
Total Success Launches for All Sites

KSC LC-39A had the most successful launches amongst all other launch sites at 41.7%



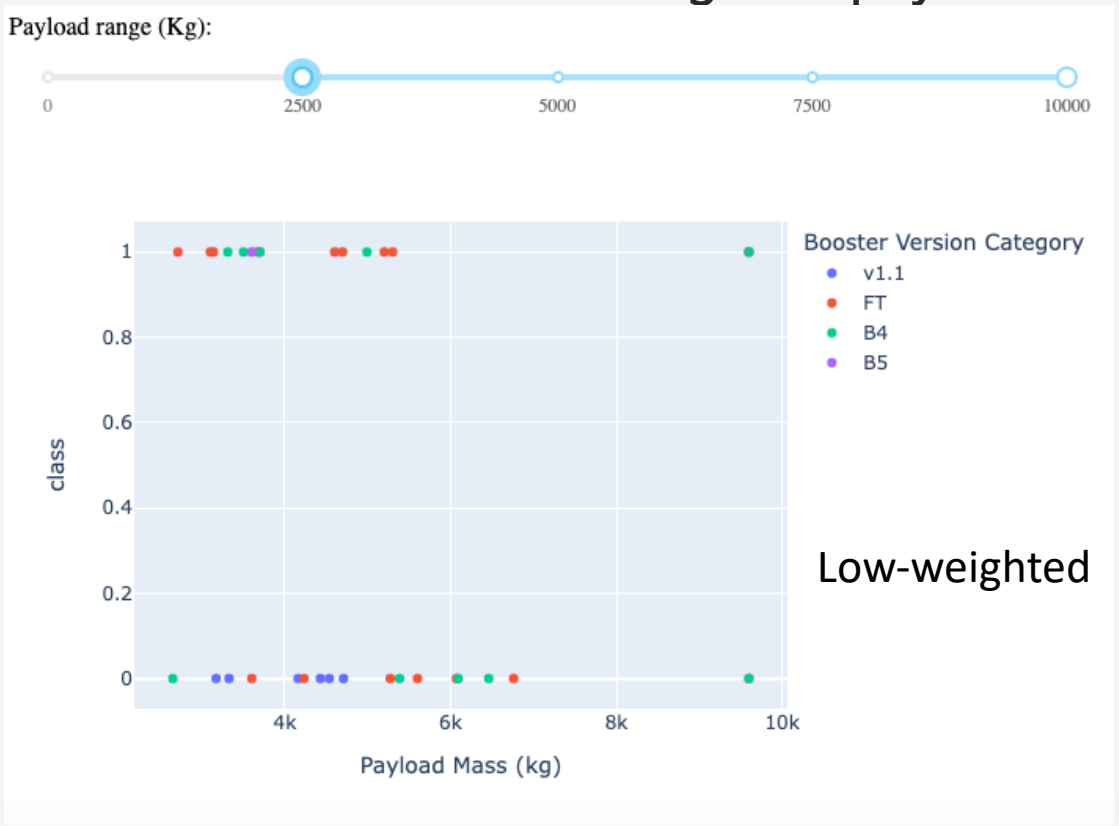
Launch Site with Highest Success Ratio

- Launch site KSC LC-39A achieved a 76.9% success rate

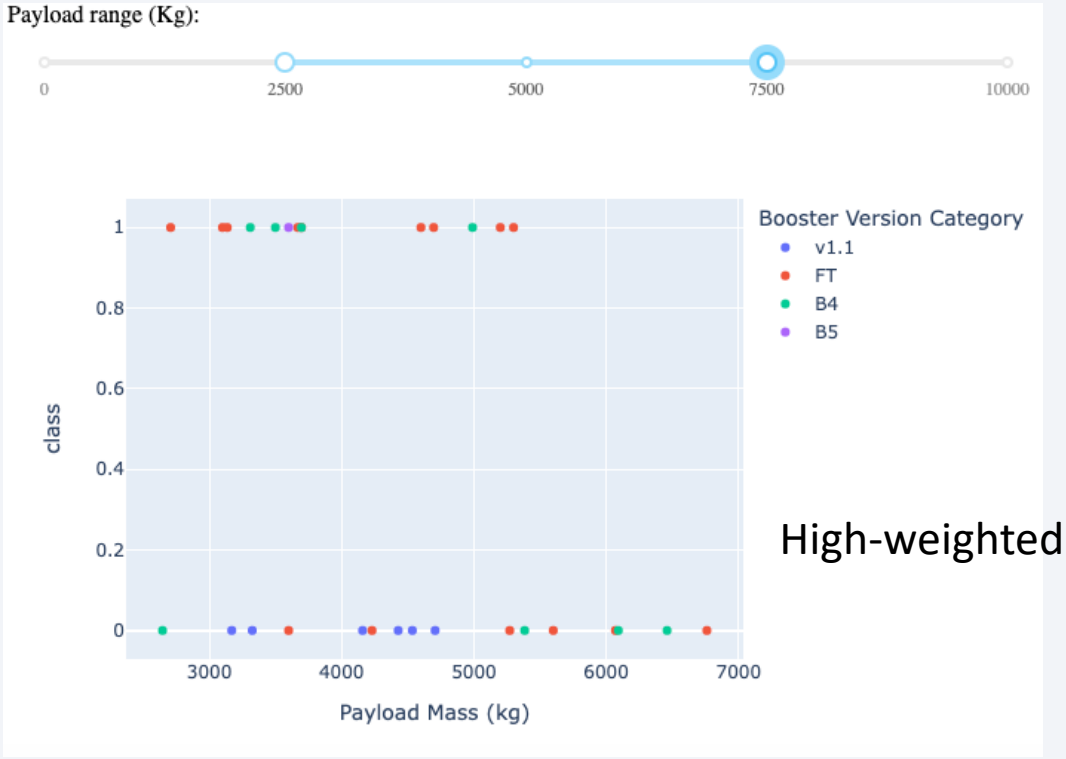


Payload vs Launch Outcome (All Sites)

When looking at the Payload vs. Launch Outcome scatter plot , we can see that the success rate for lower weighted payloads are higher than heavy weighted payloads



Low-weighted



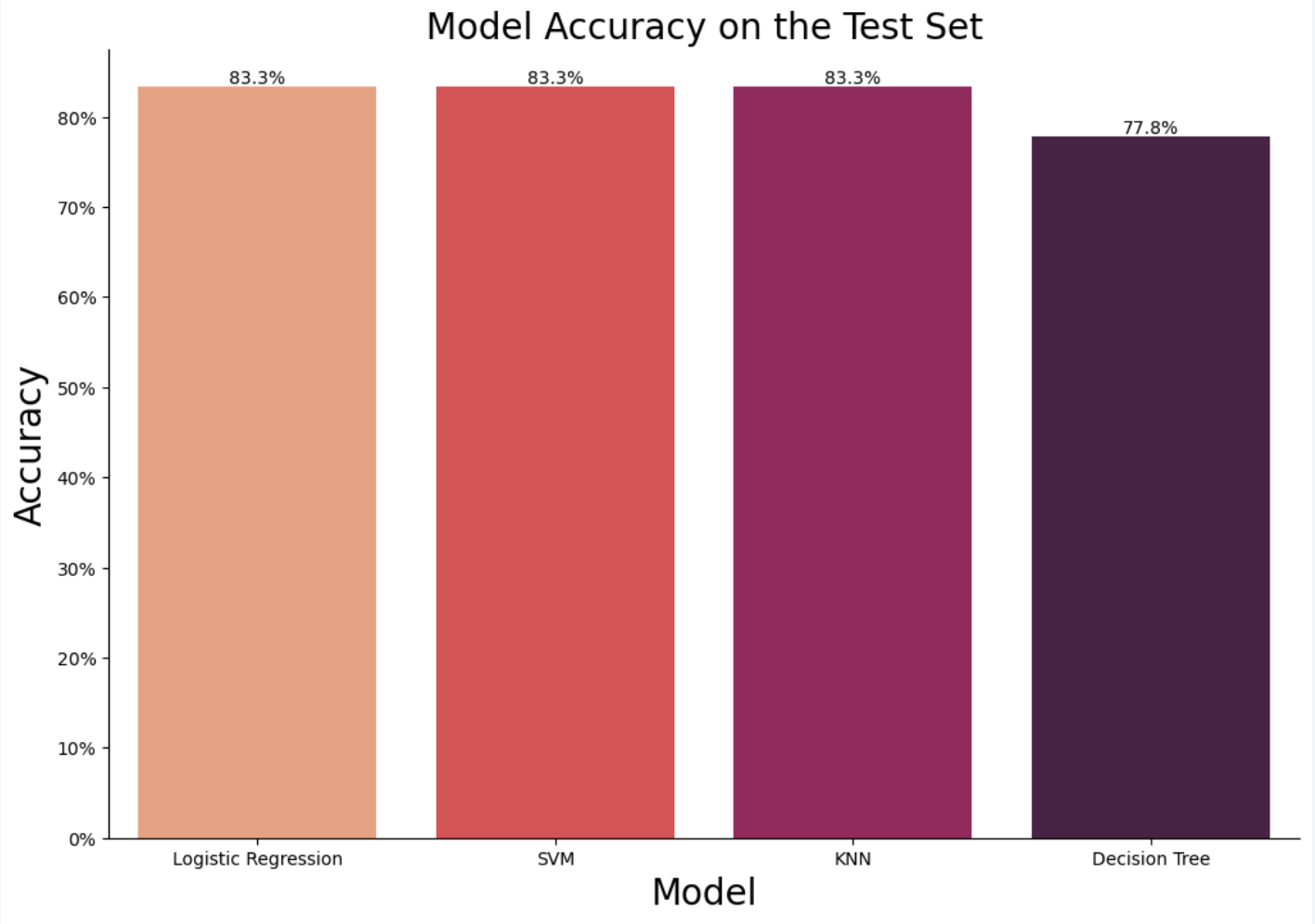
High-weighted

Section 5

Predictive Analysis (Classification)

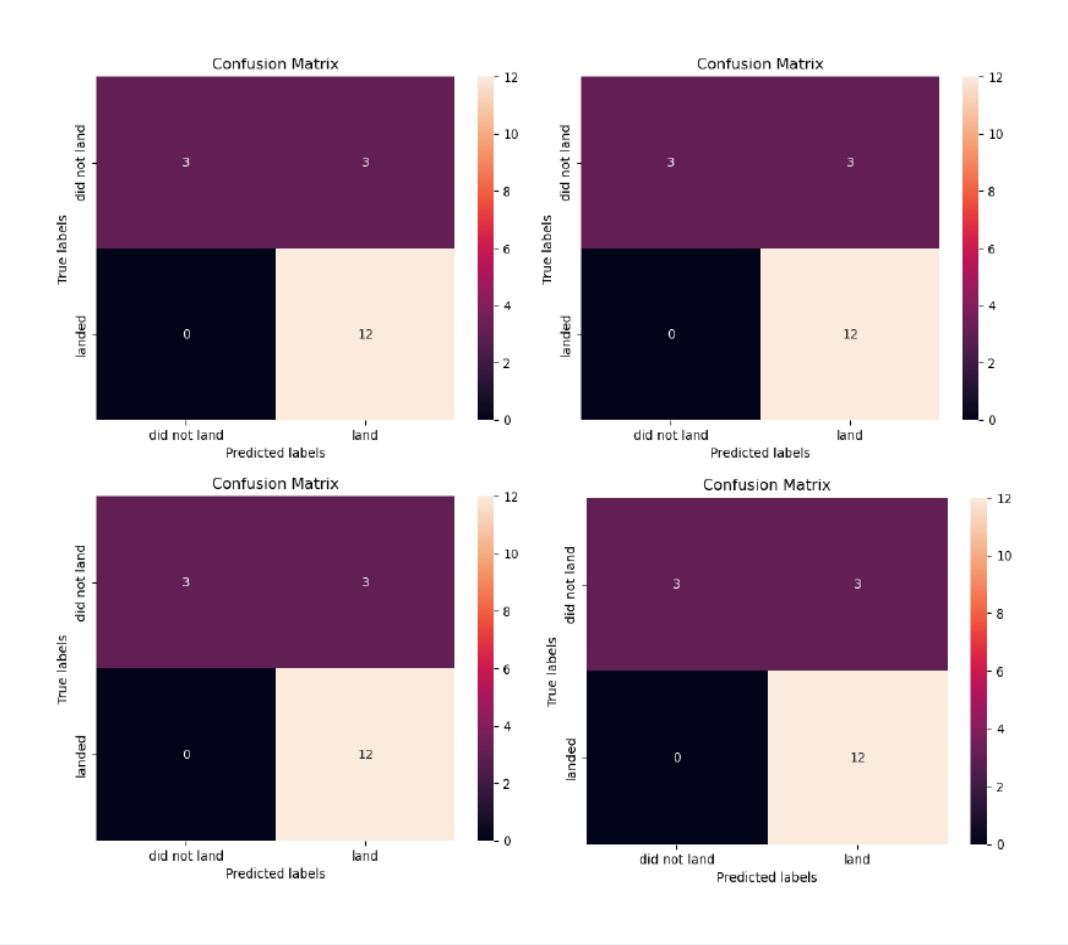
Classification Accuracy

- Logistic Regression, Support Vector Machines, and K-Nearest Neighbor classification models tied for the highest accuracy at 83.3%



Confusion Matrix

- The confusion matrices of Logistic Regression, SVM, and KNN had similar results, being able to distinguish between different classes (Success and Failure) of Launches



Conclusions

- The Logistic Regression, SVM, and KNN models appear to be the best classifiers in terms of prediction accuracy for this SpaceX dataset
- Low weighted payloads had a higher success rate than heavier ones
- The success rate of SpaceX launches got better over time, accounting for better technology and lessons learned from launches
- Launches that had GEO, HEO, SSO, and ES-L1 orbit types had the best success rate
- Launch site KSC LC-39A had the most successful Launches amongst all other launch sites

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!