IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Riley Taylor
November 21, 2022

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection – API and Web Scraping
  - Exploratory Data Analysis (EDA) with SQL
  - Exploratory Data Analysis (EDA) with Data Visualization
  - Interactive Map – Folium
  - Dashboard – Plotly Dash
  - Predictive Analysis
- Summary of all results
  - Exploratory Data Analysis results
  - Interactive maps/dashboard
  - Predictive analysis results

Introduction

- Project background and context
  - SpaceX Falcon 9 rocket launches have an average cost of $62 million, compared to other providers operating at an average cost of $165 million per launch. Much of the cost savings is determined by the first stage of the launch, which SpaceX can reuse materials for.
- Problems to answer
  - What are the characteristics associated with successful or failed landings?
  - What are the effects of each relationship variable on the success of the landing?
  - What conditions allow for SpaceX to achieve a successful landing?

Section 1

Methodology

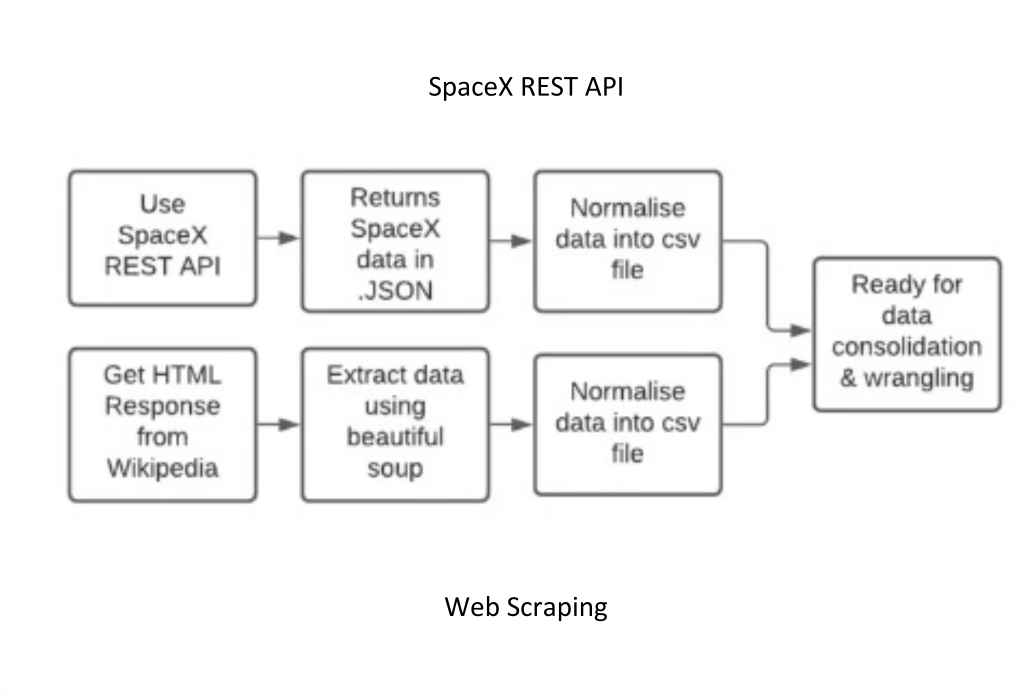
Methodology

Executive Summary

- Data collection methodology:
    - SpaceX REST API
    - Web scraping from Wikipedia
- Perform data wrangling
    - Drop unnecessary columns
    - One Hot Encoding for classification models
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
    - Linear Regression, K Nearest Neighbor, SVM, DT models have been built and evaluated

Data Collection

- Datasets were collected from SpaceX REST API and Web Scraping (Wikipedia)
- Information gathered from the SpaceX REST API
    - Rocket used, payload delivered, launch specifications, landing specifications, and landing outcome.
    - The SpaceX REST API endpoints, or URL, starts with api.spacexdata.com/v4/.
- Information gathered from Web scraping
    - Launch specifications, landing specifications, and payload information



Data Collection – SpaceX API

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

1 .Getting Response from HTML

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

2. Creating BeautifulSoup Object

```
soup = BeautifulSoup(page.text, 'html.parser')
```

3. Finding tables

```
html_tables = soup.find_all('table')
```

4. Getting column names

```
column_names = []
temp = soup.find_all('th')
for x in range(len(temp)):
    try:
     name = extract_column_from_header(temp[x])
     if (name is not None and len(name) > 0):
        column_names.append(name)
    except:
     pass
```

5. Creation of dictionary

```
launch_dict= dict.fromkeys(column_names)

# Remove an irrelvant column
del launch_dict['Date and time ( )']


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'] = []
launch_dict['Version Booster']=[]
launch_dict['Booster landing']=[]
launch_dict['Date']=[]
launch_dict['Time']=[]
```

6. Appending data to keys (refer) to notebook block 12

```
In [12]: extracted_row = 0
         #Extract each table
         for table_number,table in enumerate(
             # get table row
              for rows in table.find_all("tr");
                    #check to see if first table
```

7. Converting dictionary to dataframe

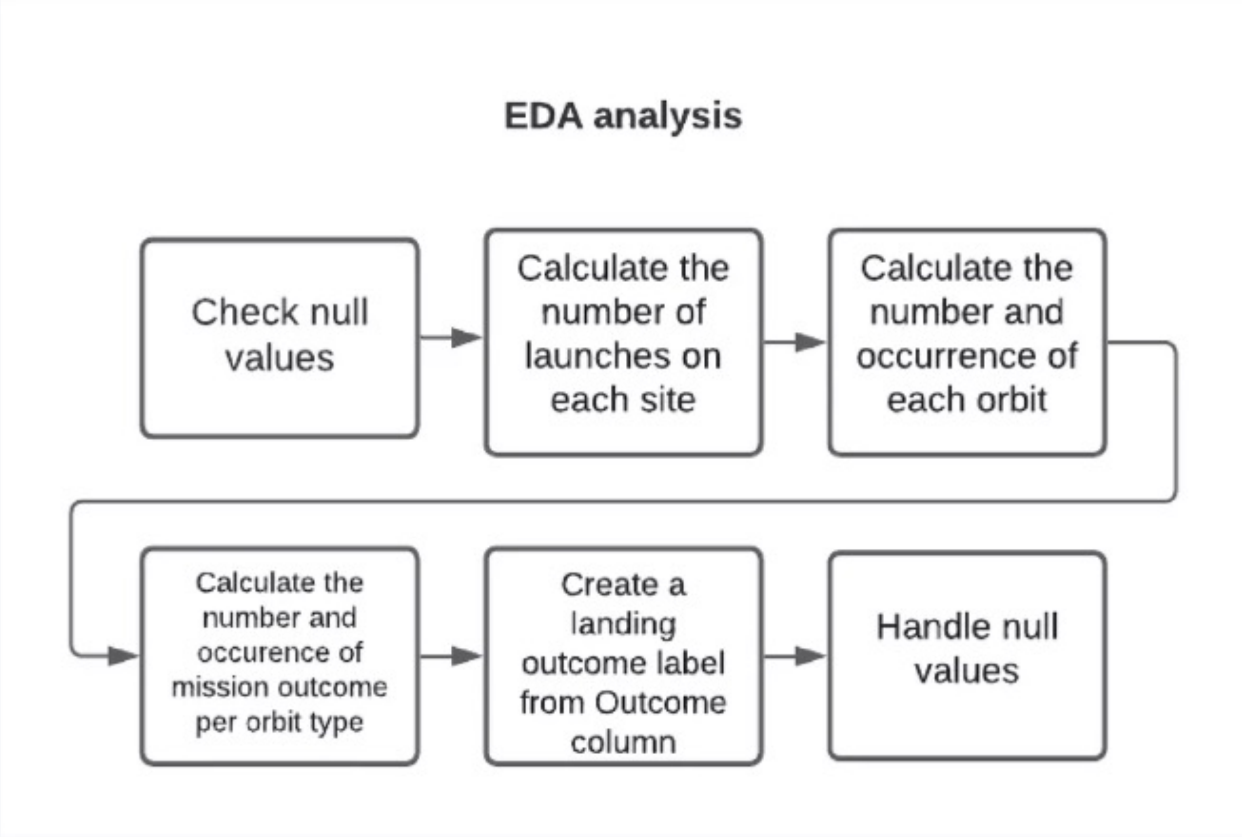
```
df = pd.DataFrame.from_dict(launch_dict)
```

8. Dataframe to .CSV

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

Data Wrangling

- This dataset includes several cases where the landing was unsuccessful
  - True Ocean, True RTLS, True ASDS indicates a successful mission
  - False Ocean, False RTLS, False ASDS indicates a failed mission
- String variables need to be transformed into categorical variables
  - 1 indicates a successful mission
  - 0 indicates a failed mission



EDA with Data Visualization

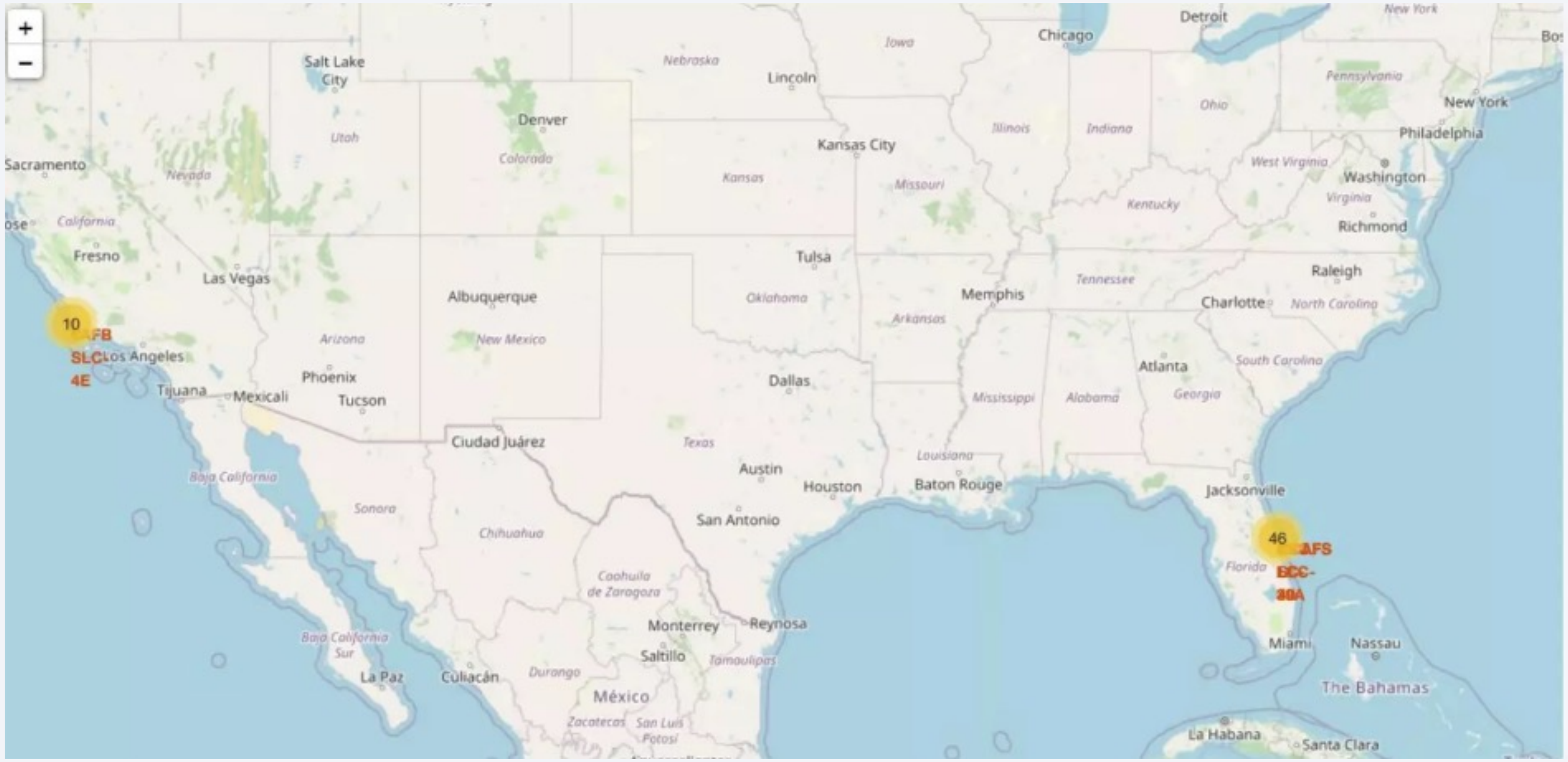
- Scatter Plots
  - Flight Number vs Payload Mass
  - Flight Number vs Launch Site
  - Payload vs Launch Site
  - Orbit vs Flight Number
  - Payload vs Orbit Type
  - Orbit vs Payload Mass
- Bar Graphs
  - Success Rate vs Orbit
- Line Graph
  - Success Rate vs Year

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 varied 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 and 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, and launch site each month in 2017
  - Ranking the count of successful landing outcomes between June 4, 2010 and March 20, 2017 in descending order

Build an Interactive Map with Folium

- Map markers have been added to the map to find the optimal locations for building launch sites



Build a Dashboard with Plotly Dash

- Dashboard Elements
  - Dropdown, Pie Chart, Range Slider, Scatter Plot
- Dropdown allows the user to select a launch site or include all launch sites
- Pie Chart shows the comparison of successful and failed launches by launch site (based on user selection)
- Range Slider allows the user to select the Payload Mass (within a fixed range)
- Scatter Plot displays the relationship between Success and Payload Mass

Predictive Analysis (Classification)

- Data Preparation
    - Load dataset, normalize data, and split into training and testing sets
- Model Preparation
    - Select machine learning algorithms, set parameters for each to GridSearchCV, training the GridSearch model with the training set
- Model Evaluation
    - Determine the best hyperparameters for each model type, compute the accuracy for each model with the test set, plot the confusion matrix
- Model Comparison
    - Compare the models to each other based on accuracy and chose the model with the highest accuracy rate

Results

- Results of the data analysis
- Results of the interactive analytics
- Results of the predictive analysis

Section 2

Insights drawn from EDA

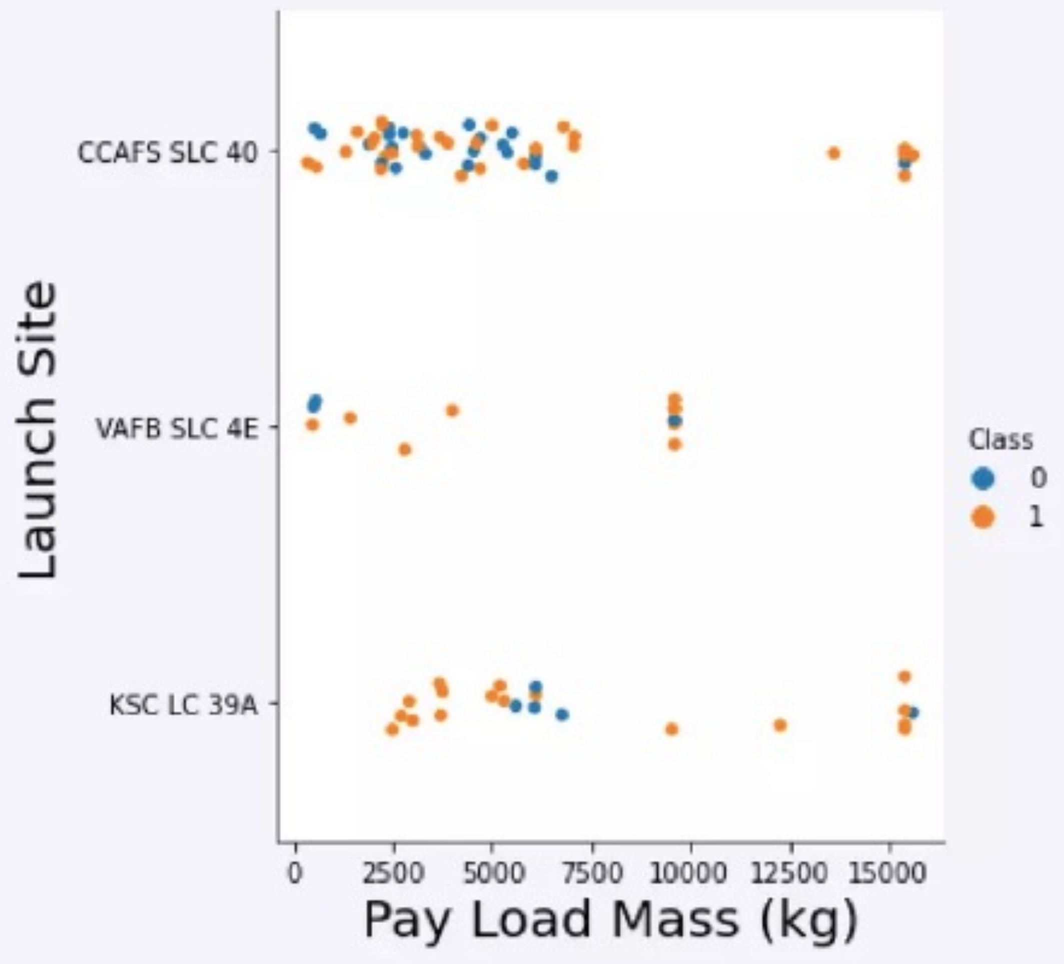
Flight Number vs. Launch Site

The success rate is increasing with the number of launches for each site

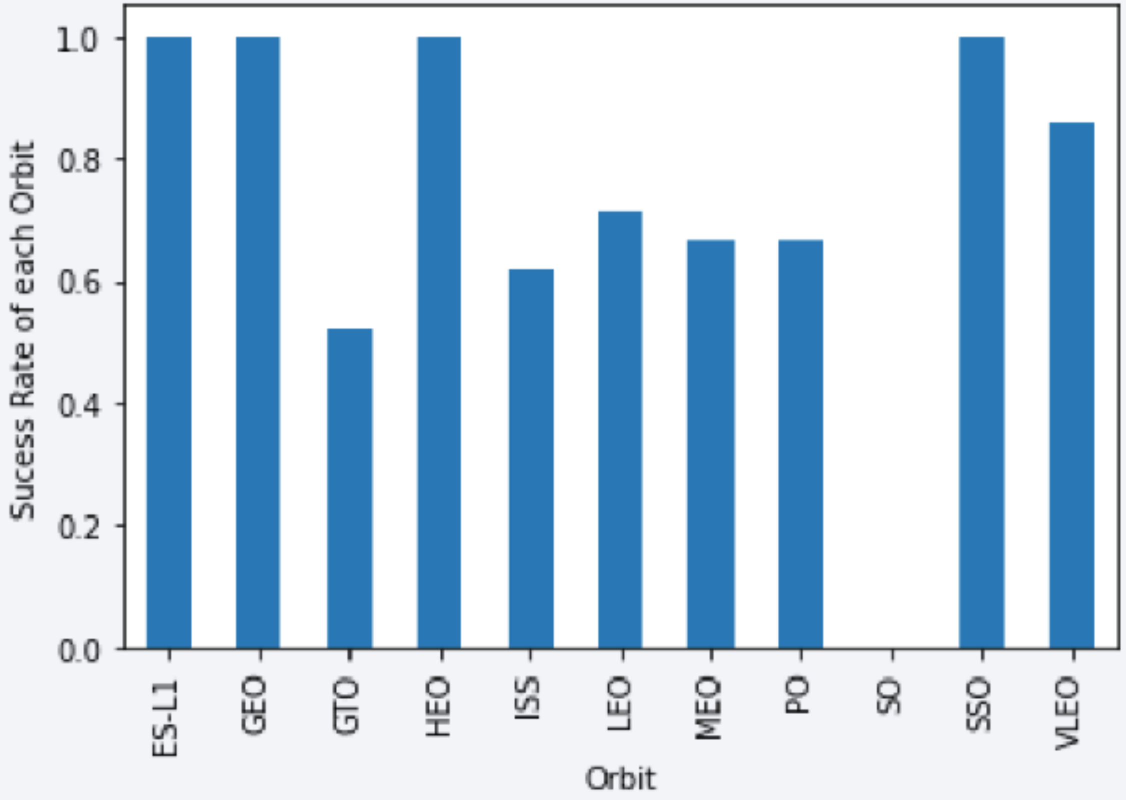
Payload vs. Launch Site

A heavier payload may cause a landing to fail.

However, depending on the launch site, a heavier payload may increase the success rate.



Success Rate vs. Orbit Type



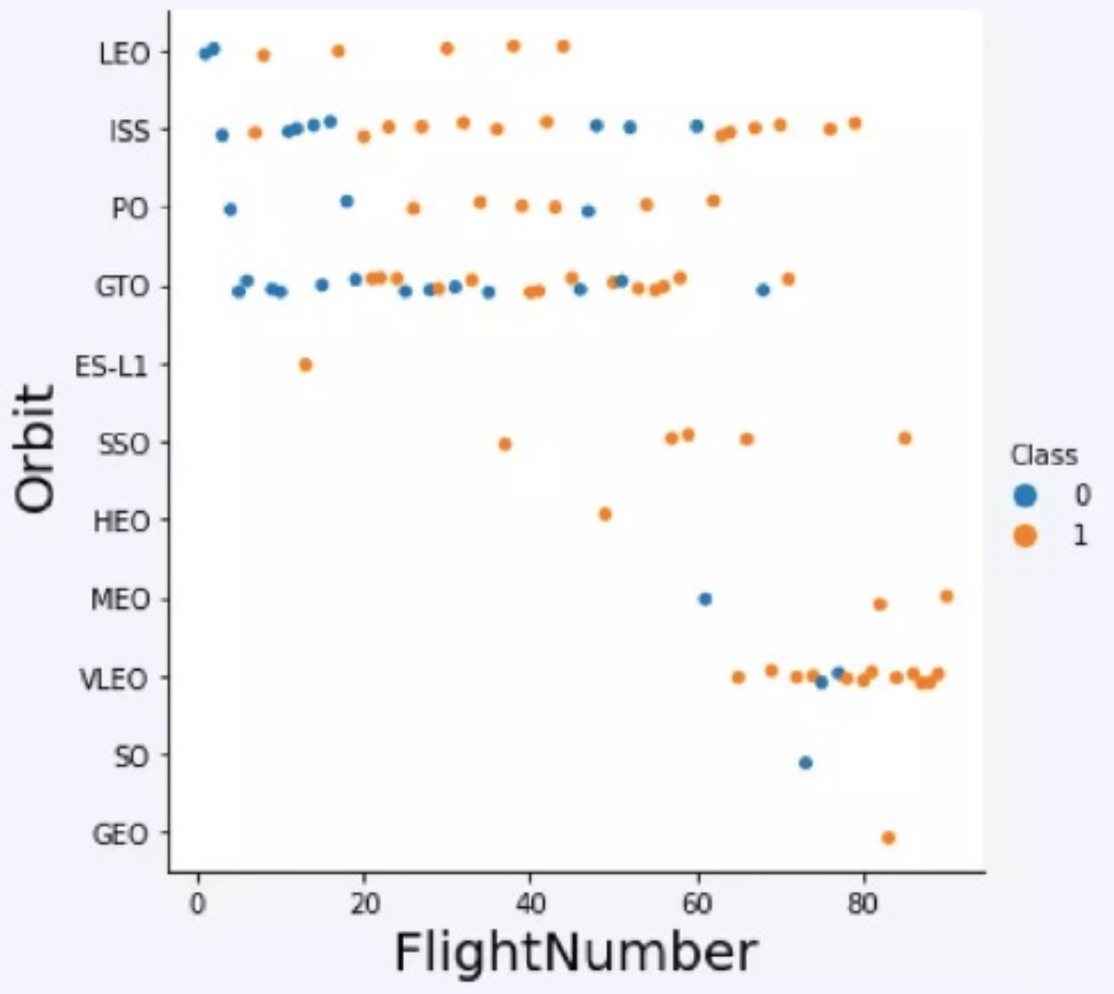
The orbit types ES-L1, GEO, HEO, SSO have the best success rates

Flight Number vs. Orbit Type

The success rate increases with the number of flights for the LEO orbit.

Other orbits show no relation between the success rate and number of flights.

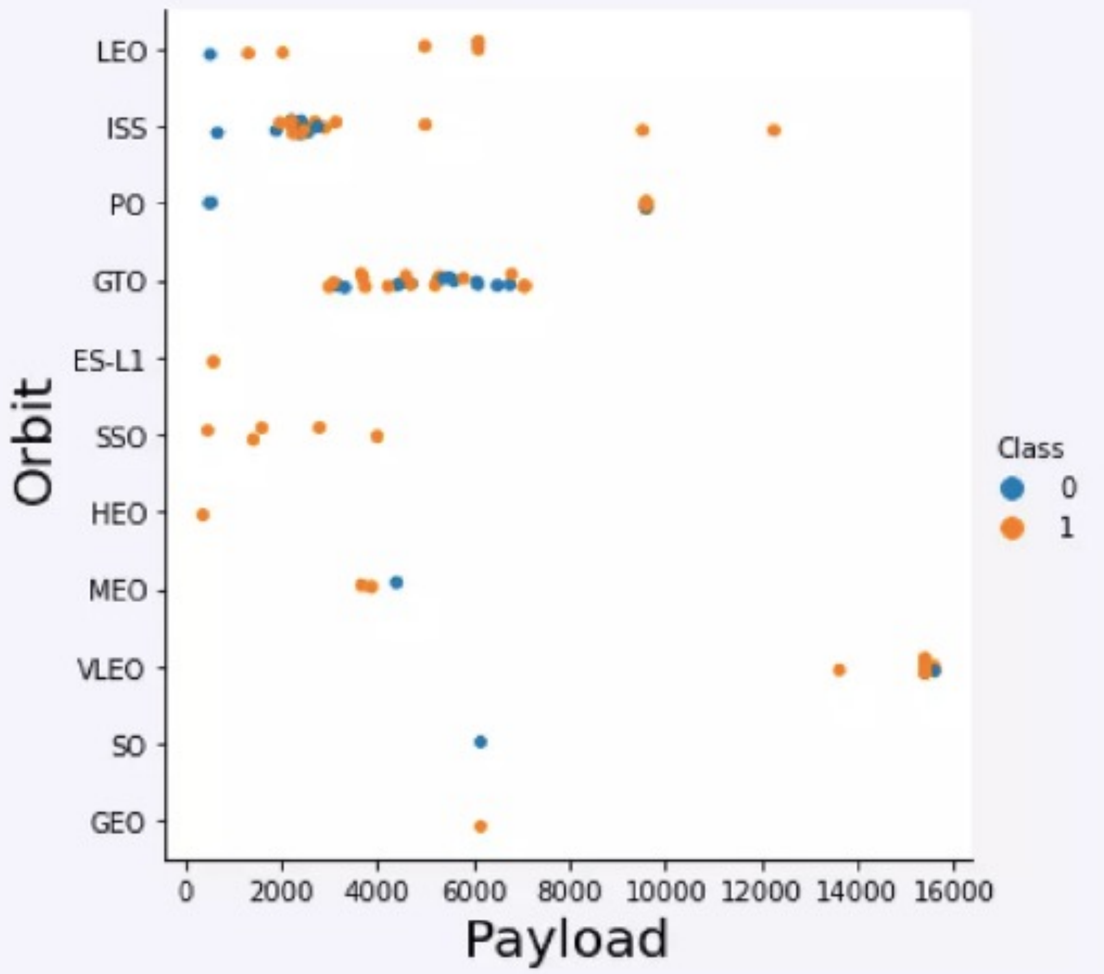
The assumption can be made that the knowledge learned during repeat launches increases the success rate over time.



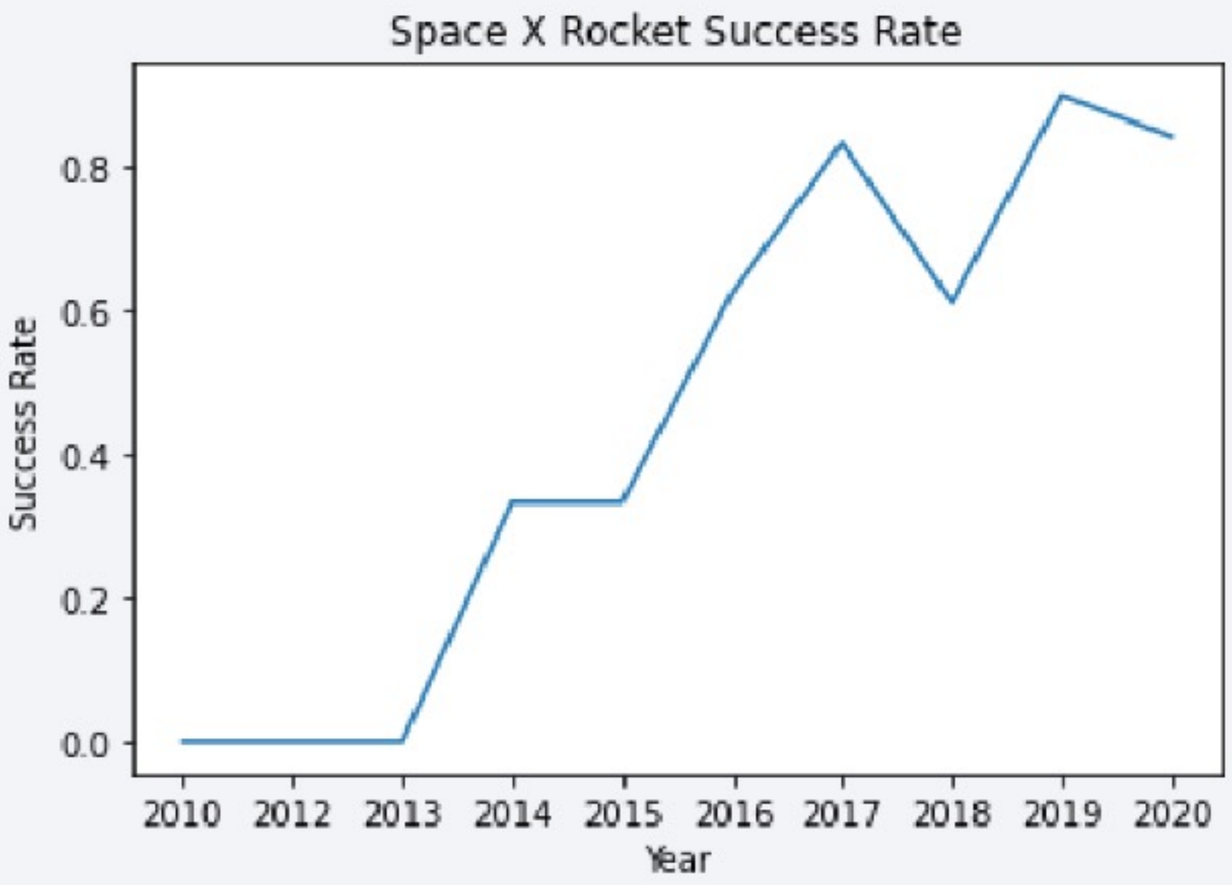
Payload vs. Orbit Type

The payload weight has a great influence on the success of the launch.

Heavier payloads improve the success rate of the LEO orbit, but decreasing the payload weight for GTO improves its respective success rate.



Launch Success Yearly Trend



The launch success rate has significantly increased since 2013

All Launch Site Names

- Query to find unique launch site names:
  - %sql select distinct(LAUNCH_SITE) from SPACEXTBL
- Launch sites:
  - CCAFS LC-40
  - CCAFS SLC-40
  - KSC LC-39A
  - VAFB SLC-4E

Launch Site Names Begin with 'CCA'

- Query to find launch sites beginning with 'CCA':
  - %sql select * from SPACEXTBL where LAUNCH_SITE like 'CCA%' limit 5
- Result:

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

- Query to calculate the total payload carried by boosters from NASA:
  - %sql select sum(PAYLOAD_MASS__KG_) from SPACEXTBL where CUSTOMER = ‘NASA (CRS)’
- Result:
  - 45596

Average Payload Mass by F9 v1.1

- Query to calculate the average payload mass carried by booster version F9 v1.1:
  - %sql select avg(PAYLOAD_MASS__KG_) from SPACEXTBL where BOOSTER_VERSION = 'F9 v1.1'
- Result:
  - 2928.40

First Successful Ground Landing Date

- Query to find the dates of the first successful landing outcome on ground pad:
  - %sql select min(DATE) from SPACEXTBL where Landing__Outcome = 'Success (ground pad)'
- Result:
  - 2015-12-22 (December 22, 2015)

Successful Drone Ship Landing with Payload between 4000 and 6000

- Query to list the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 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 Results:
  - F9 FT B1022
  - F9 FT B1026
  - F9 FT B1021.2
  - F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- Query to calculate the 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)'
- Result:
  - 100

Boosters Carried Maximum Payload

- Query to list the names of the booster which have carried the maximum payload mass:
  - %sql select BOOSTER_VERSION from SPACEXTBL where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_) from SPACEXTBL)
- Booster Version Results:
  - 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

- Query to list the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015:
  - %sql select * from SPACEXTBL where Landing__Outcome like ‘Success%’ and (DATE between ‘2015-01-01’ and ‘2015-12-31’) order by date desc
- Results:

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

- Query to rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order:
  - %sql select * from SPACEXTBL where Landing__Outcome like ‘Success%’ and (DATE between ‘2010-06-04’ and ’2017-03-20’) order by date desc
- Results:

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

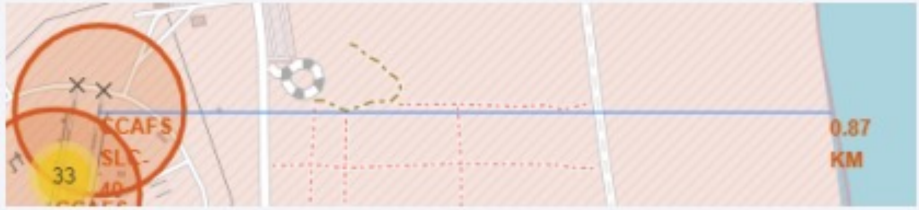
Map of Launch Sites

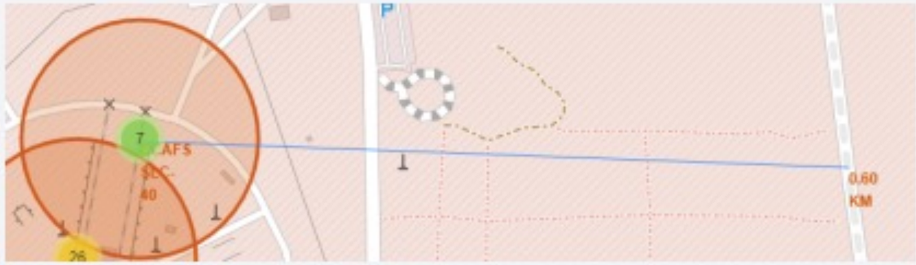
The launch sites are located along the southern coasts of the United States

Successful/Failed Launches

The Green markers represent successful launches while Red markers represent failed launches. KWC LC-39A has the highest launch success rate.

Proximities of Launch Sites



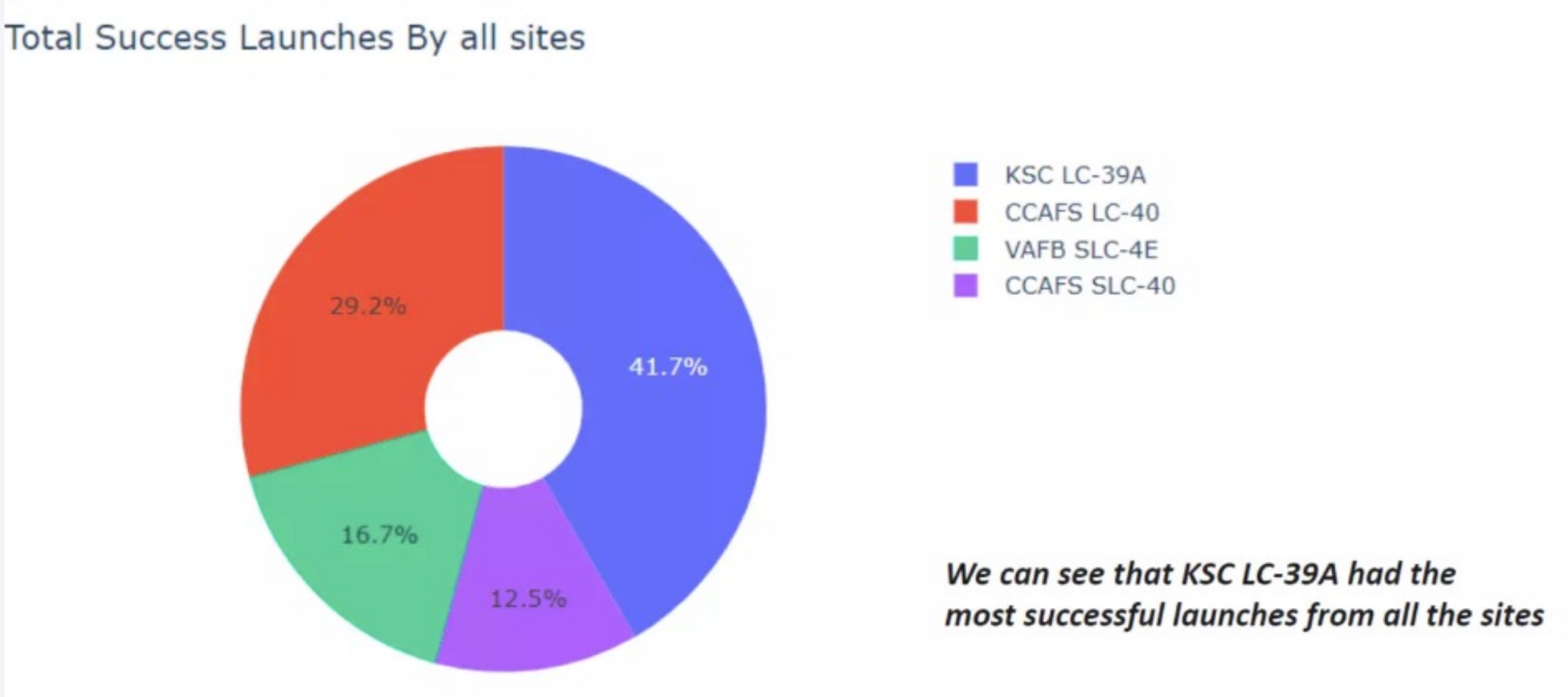


The map shows the proximity of the launch site CCAFS SLC-40 to railways, highways, coastline, and the distance to the nearest cities.

Section 4

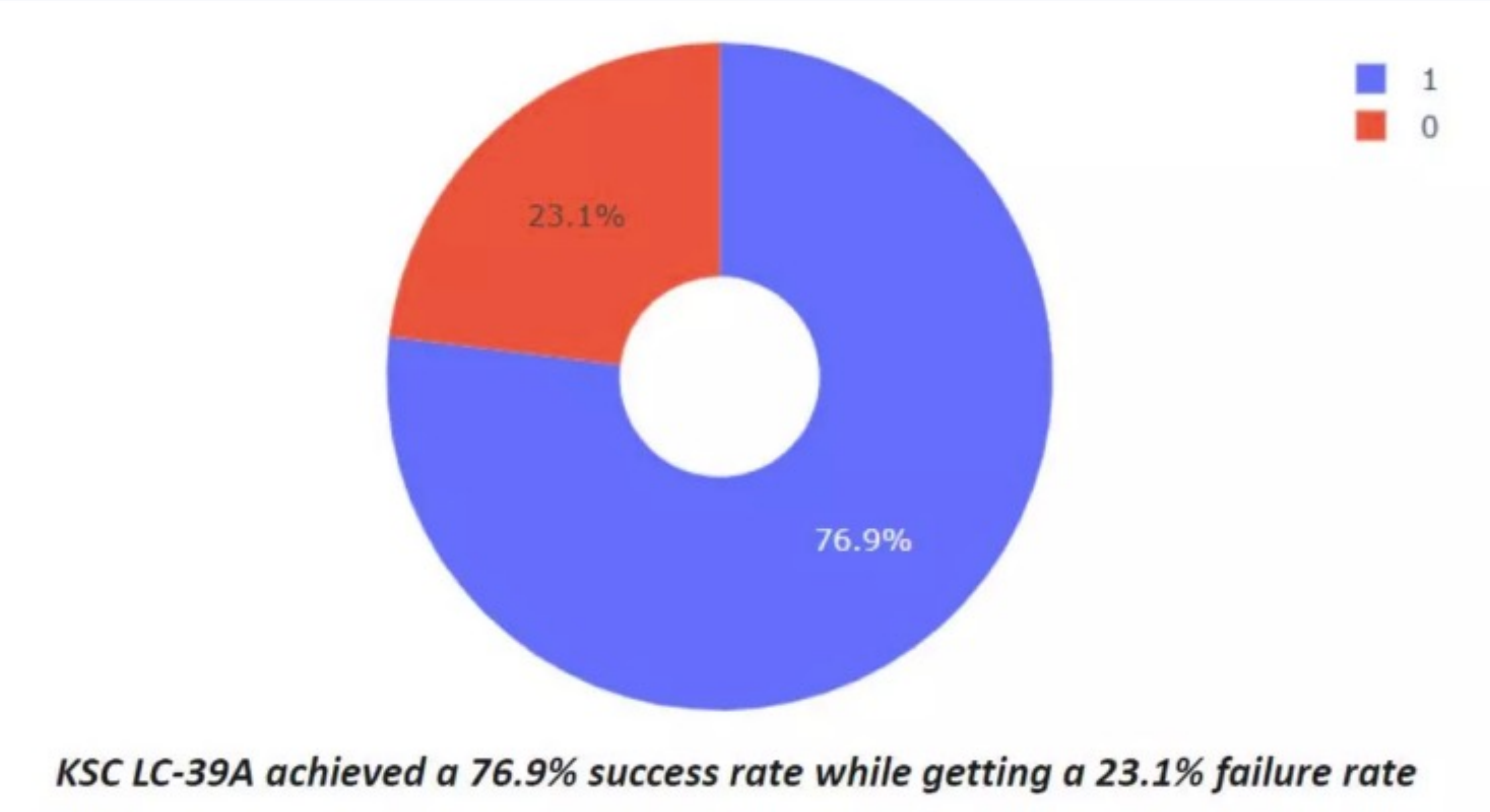
Build a Dashboard with Plotly Dash

Total Successful Launches



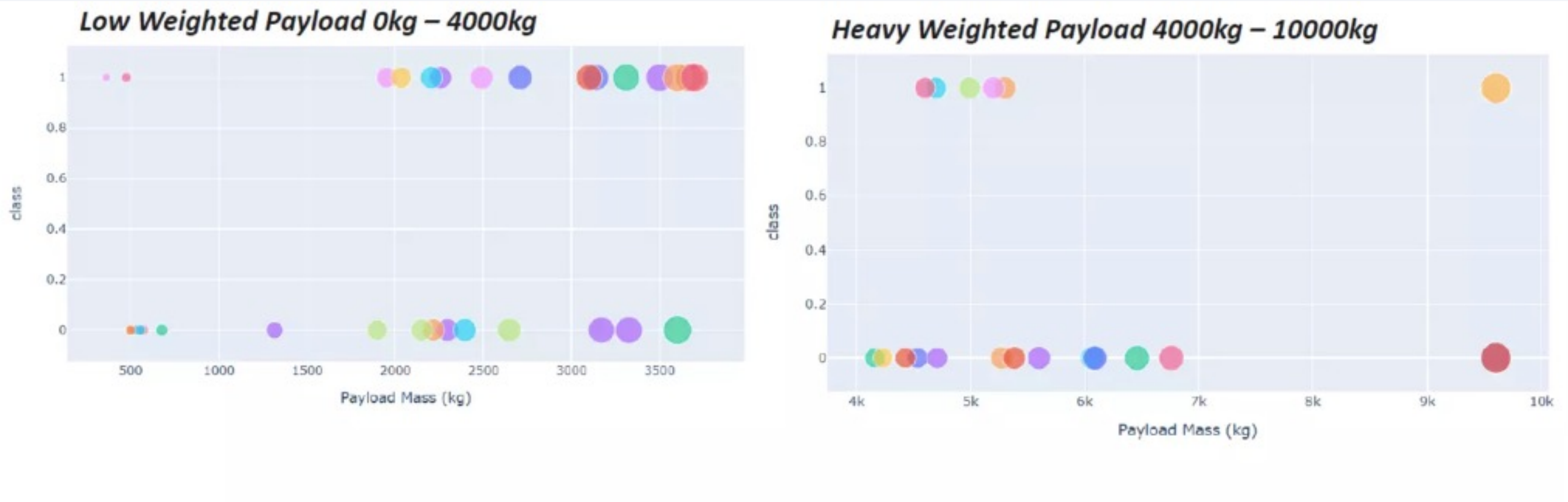
We can see that KSC LC-39A had the best success rate

Success Rate by Site



Metrics for KSC LC 39A

Payload vs Launch Outcome

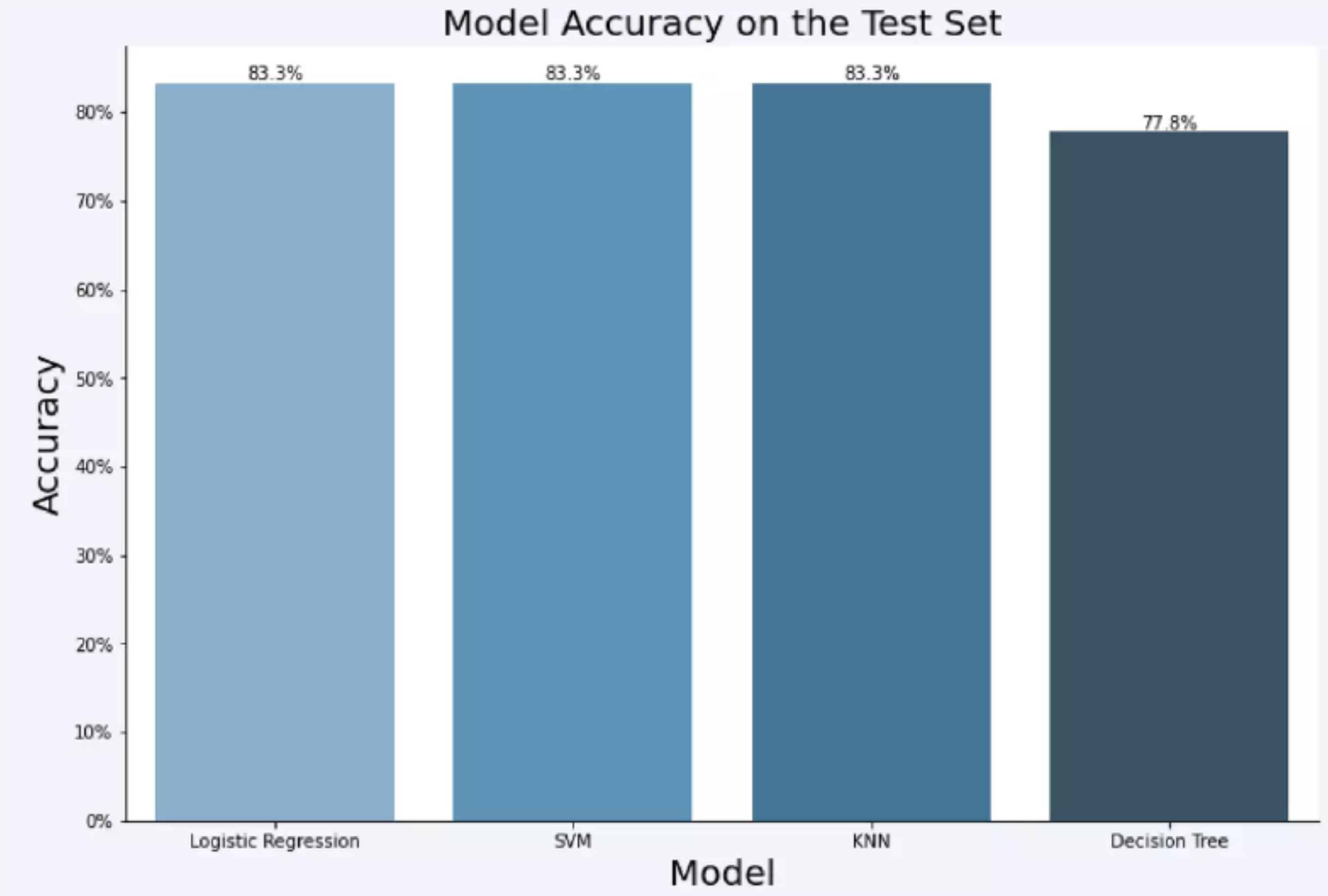


The success rate for lower payload weights is higher than the rate for heavy payloads

Section 5

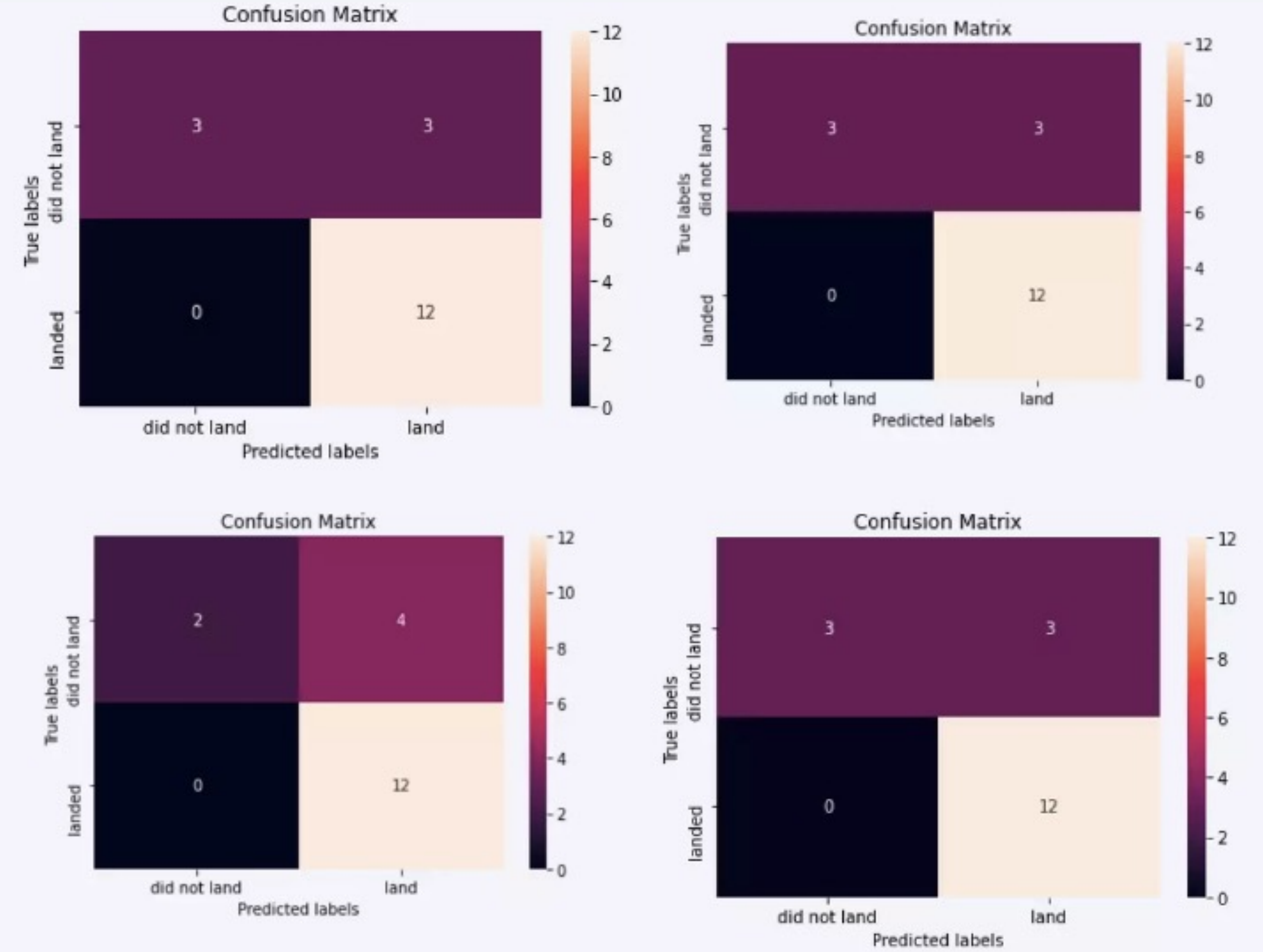
Predictive Analysis (Classification)

Classification Accuracy



Logistic Regression has the highest Accuracy Rate

Confusion Matrix



The confusion matrix doesn't provide any valuable insights since the test accuracy for each method is so similar.

Conclusions

- The success rate is best explained by several factors (ex. Launch Site, Orbit, and Previous Launches). We can assume that the number of previous launches is the most important factor.
- Generally, low weighted payloads perform better than heavier payloads.
- We can't determine exactly why some launch sites perform with a higher success rate. Inferences can be made, but additional data is needed.
- Orbits GEO, HEO, SSO, ES L1 had the best success rates.

Thank you!