IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Mohamad Azrai
Mahadan
20/01/2022

Outline

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

Executive Summary

- The purpose of the project is to help the stakeholder making the right decision in determining the bid on different rockets based on successful launch. There are two methods that have been used in order to collect the data on Space X Falcon 9. The first one is by using the SpaceX API and the other method is by webscraping the Wikipedia. The data is cleaned by replacing some missing values. Then, the exploratory data analysis (EDA) is performed by using visualization and SQL to get some insights on the data. Furthermore, by performing interactive visual analytics using Folium and Plotly Dash, some pattern have been discovered to help us determining some features for the model building. Classification methods algorithm have been used including K-means, SVM and Logistic Regression to predict the success (rates) of the rocket launch. Overall, all algorithms are performed the same with KSC LC-39A has the highest rates of successful launch.

Introduction

- SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage. Therefore if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.
- We want to help making the right decision in determining the bid on different rockets based on successful launch. This will help the company to maximize the profits.

Contents

- Data collection methodology
- Perform data wrangling
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

Data Collection – SpaceX API

Request rocket launch data from SpaceX API

↓

Decode the response content as a Json using .json()

↓

Turn it into a Pandas dataframe using .json_normalize()

↓

Apply ID retrieval function method to readable data

↓

Combine the columns into a dictionary and convert to dataframe

- Reference:

https://github.com/azraimahadan/AI_ML_DL/blob/main/Data%20Collection%20API.ipynb

Data Collection - Scraping

Provide some helper functions to extract HTML

↓

Request the Falcon9 Launch HTML page

↓

Create BeautifulSoup object

↓

Iterate through the elements and apply the provided extract function to extract column name one by one

↓

Create a data frame by parsing the launch HTML tables

Reference :

https://github.com/azraimahadan/AI_ML_DL/blob/main/Data%20Collection%20with%20Webscraping.ipynb

Data Wrangling

- The data is filtered first by only including Falcon9.

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

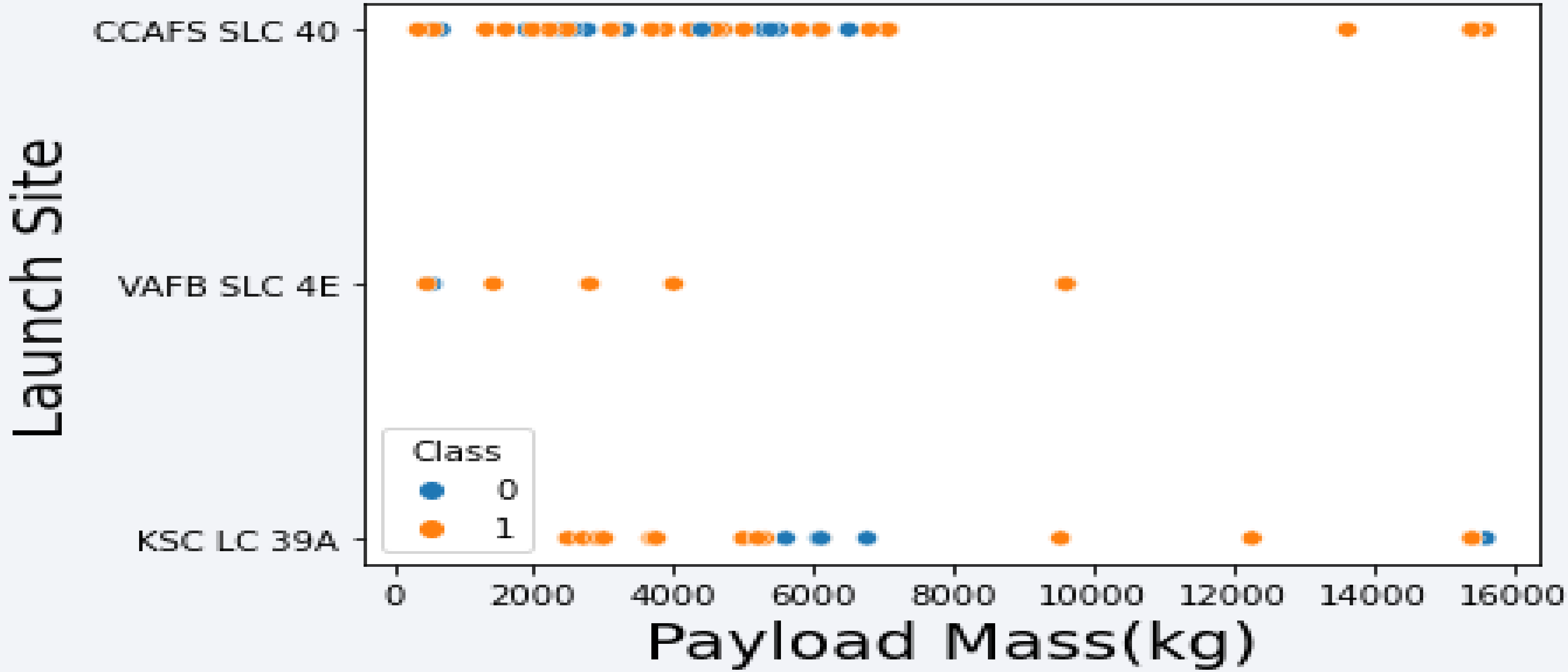
- Then, reset the FlightNumber column index.
- Then, replace the NaN values in payload column with the payload mean.

```
# Calculate the mean value of PayloadMass column
payload_mean = data_falcon9.PayloadMass.mean()
# Replace the np.nan values with its mean value
data_falcon9['PayloadMass'] =
data_falcon9['PayloadMass'].replace(np.nan,payload_mean)
```

- Use the method value_counts() on the column LaunchSite to determine the number of launches on each site
- Calculate the number and occurrence of each orbit
- Reference:

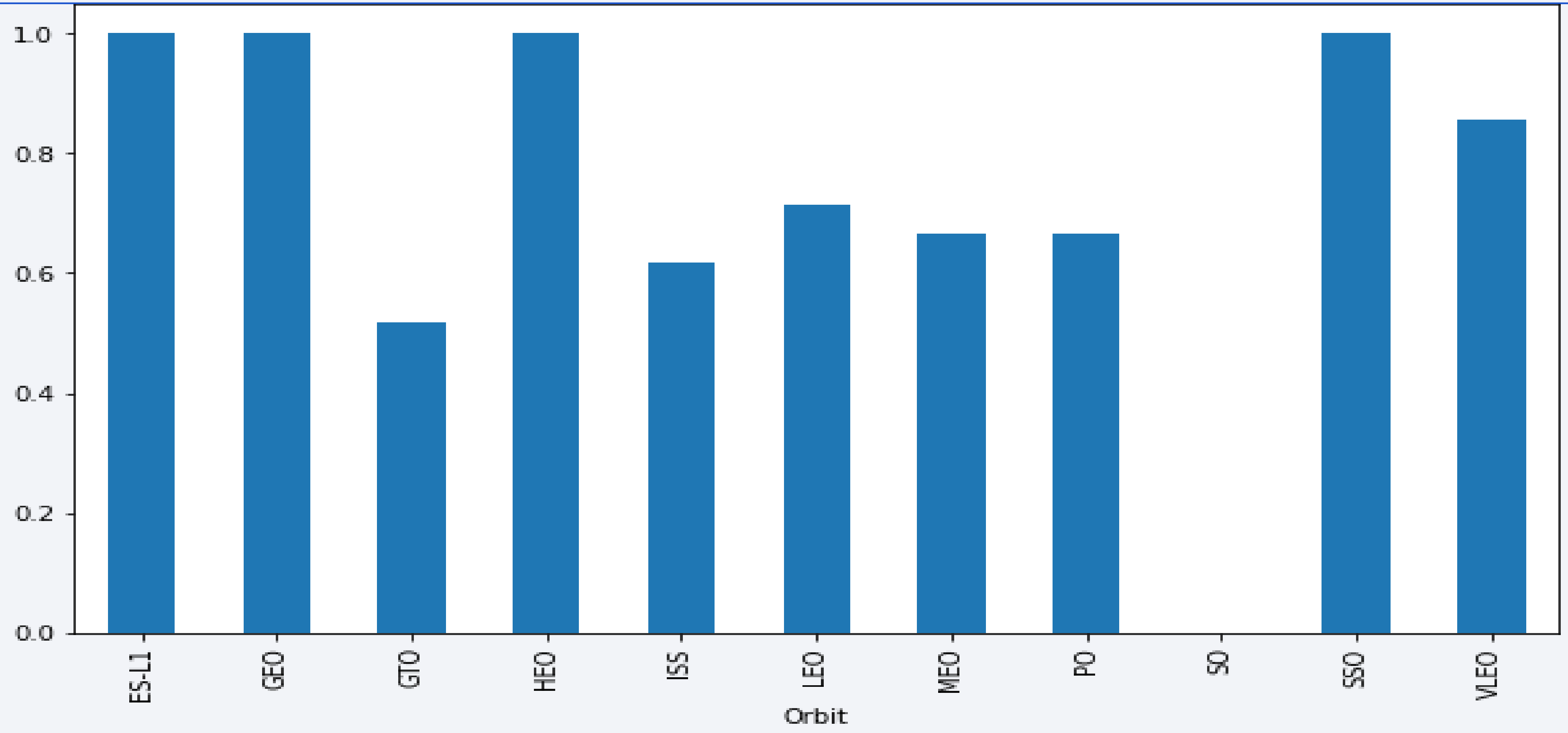
https://github.com/azraimahadan/AI_ML_DL/blob/main/Data%20Collection%20API.ipynb

EDA with Data Visualization



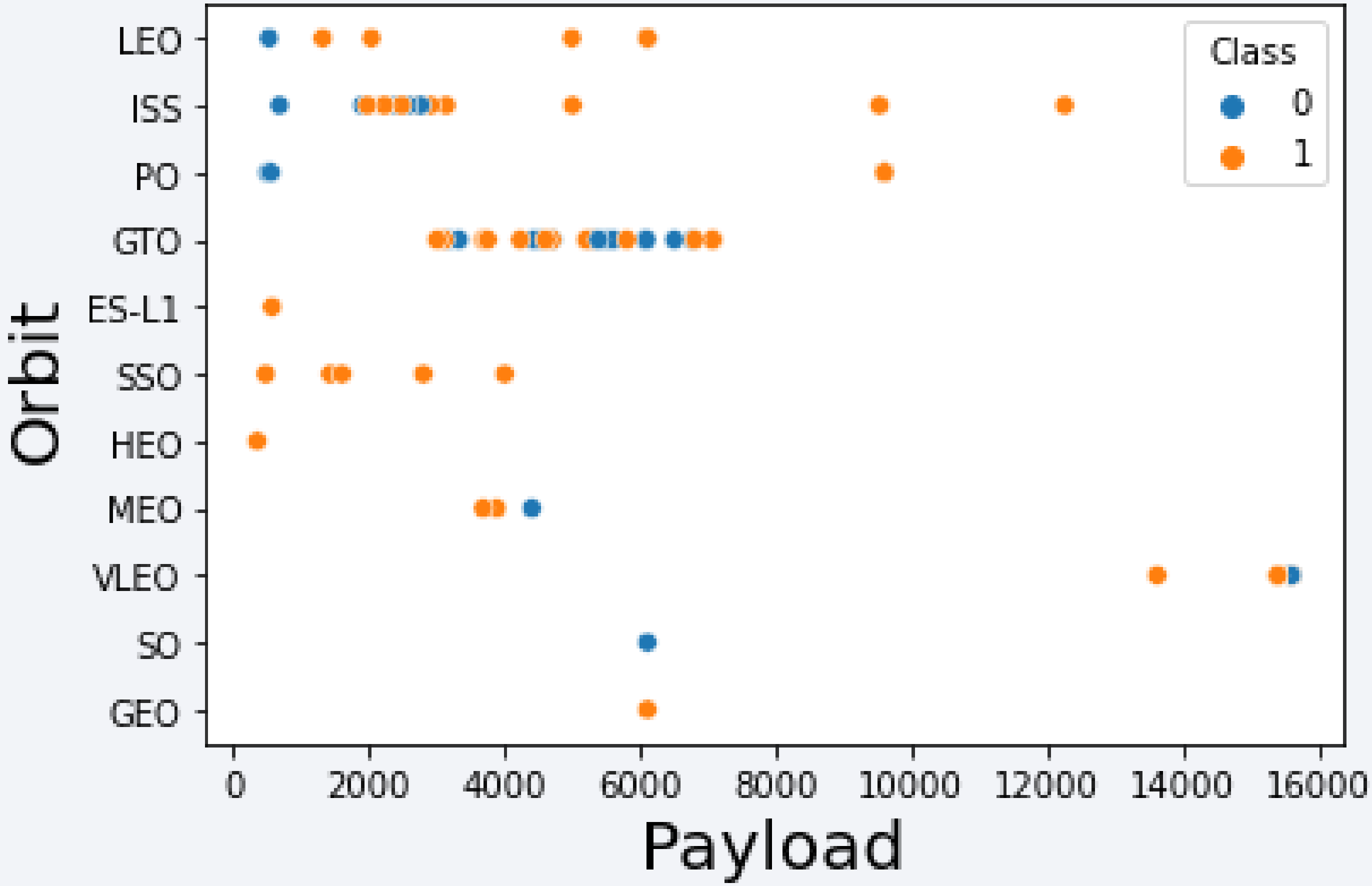
There are no rockets launched for heavypayload mass(greater than 10000) for the VAFB-SLC launchsite.

EDA with Data Visualization



Orbit ES-L1,GEO,HEO, and SSO have 100% success rates

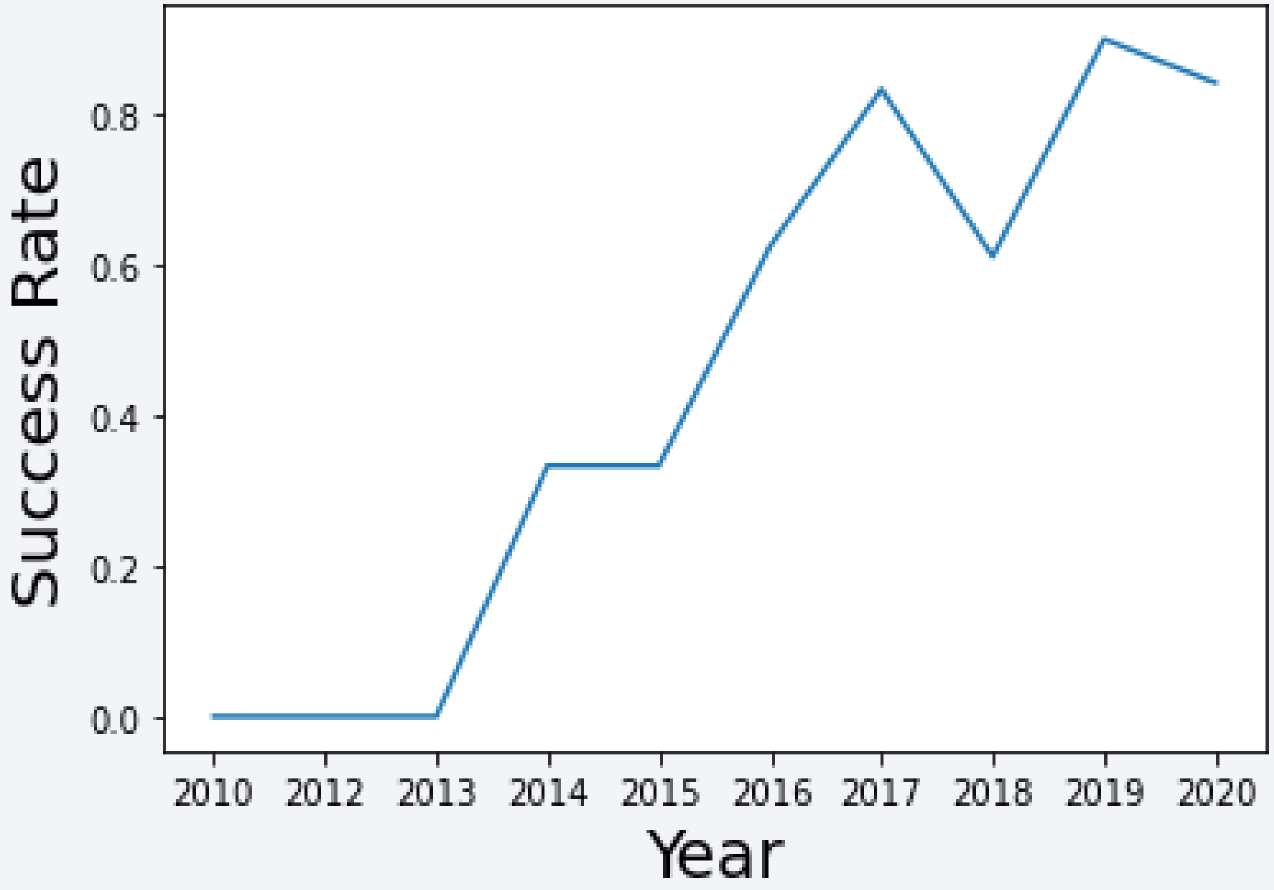
EDA with Data Visualization



With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.

However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here

EDA with Data Visualization



The success rate since 2013 kept increasing till 2020

EDA with SQL

Task 1

Display the names of the unique launch sites in the space mission

In [6]:
```
%%sql
select unique(launch_site) from SPACEXTBL
```

```
 * ibm_db_sa://jrp31646:***@98538591-7217-4024-b027-8baa776ffad1.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30875/bludb
Done.
```

Out[6]:

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

Task 2

Display 5 records where launch sites begin with the string 'CCA'

In [8]:
```
%sql select * from spacextbl where launch_site like 'CCA%' limit 5
```

```
 * ibm_db_sa://jrp31646:***@98538591-7217-4024-b027-8baa776ffad1.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30875/bludb
Done.
```

Out[8]:

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

- https://github.com/azraimahadan/AI_ML_DL/blob/main/EDA%20with%20SQL.ipynb

EDA with SQL

Task 3

Display the total payload mass carried by boosters launched by NASA (CRS)

```
In [15]: %sql select sum(payload_mass__kg_) from spacextbl where customer='NASA (CRS)'

 * ibm_db_sa://jrp31646:***@98538591-7217-4024-b027-8baa776ffad1.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30875/bludb
Done.
```

Out[15]:

1
45596

Task 4

Display average payload mass carried by booster version F9 v1.1

```
In [25]: %sql select avg(payload_mass__kg_) from spacextbl where booster_version='F9 v1.1'

 * ibm_db_sa://jrp31646:***@98538591-7217-4024-b027-8baa776ffad1.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30875/bludb
Done.
```

Out[25]:

1
2928

EDA with SQL

Task 5

List the date when the first successful landing outcome in ground pad was acheived.

Hint:Use min function

In [17]:
```
%sql select * from spacextbl
```

```
 * ibm_db_sa://jrp31646:***@98538591-7217-4024-b027-8baa776ffad1.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30875/bludb
Done.
```

Out[17]:

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
2013-09-29	16:00:00	F9 v1.1 B1003	VAFB SLC-4E	CASSIOPE	500	Polar LEO	MDA	Success	Uncontrolled (ocean)
2013-12-03	22:41:00	F9 v1.1	CCAFS LC-40	SES-8	3170	GTO	SES	Success	No attempt
2014-01-06	22:06:00	F9 v1.1	CCAFS LC-40	Thaicom 6	3325	GTO	Thaicom	Success	No attempt
2014-04-18	19:25:00	F9 v1.1	CCAFS LC-40	SpaceX CRS-3	2296	LEO (ISS)	NASA (CRS)	Success	Controlled (ocean)
2014-07-14	15:15:00	F9 v1.1	CCAFS LC-40	OG2 Mission 1 6 Orbcomm-OG2 satellites	1316	LEO	Orbcomm	Success	Controlled (ocean)
2014-08-05	08:00:00	F9 v1.1	CCAFS LC-40	AsiaSat 8	4535	GTO	AsiaSat	Success	No attempt

EDA with SQL

Task 6

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

In [43]:

```
%sql select booster_version,"Landing _Outcome",payload_mass__kg_ from spacextbl where "Landing _Outcome" = 'Success (drone ship)' \
and payload_mass__kg_ between 4000 and 6000
```

```
 * ibm_db_sa://jrp31646:***@98538591-7217-4024-b027-8baa776ffad1.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30875/bludb
Done.
```

Out[43]:

booster_version	Landing _Outcome	payload_mass__kg_
F9 FT B1022	Success (drone ship)	4696
F9 FT B1026	Success (drone ship)	4600
F9 FT B1021.2	Success (drone ship)	5300
F9 FT B1031.2	Success (drone ship)	5200

Task 7

List the total number of successful and failure mission outcomes

In [42]:

```
%sql select count(*) as total from spacextbl
```

```
 * ibm_db_sa://jrp31646:***@98538591-7217-4024-b027-8baa776ffad1.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30875/bludb
Done.
```

Out[42]:

total
101

EDA with SQL

Task 8

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

In [49]:
```
%sql select booster_version,payload_mass__kg_ from spacextbl where payload_mass__kg_ = (select MAX(payload_mass__kg_) from spacextbl)
```

```
 * ibm_db_sa://jrp31646:***@98538591-7217-4024-b027-8baa776ffad1.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30875/bludb
Done.
```

Out[49]:

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

Task 9

List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

In [53]:
```
%sql select booster_version, launch_site from spacextbl where "Landing _Outcome" like 'Failure (drone ship)' and year(date) = 2015
```

```
 * ibm_db_sa://jrp31646:***@98538591-7217-4024-b027-8baa776ffad1.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30875/bludb
Done.
```

Out[53]:

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

EDA with SQL

Task 10

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

In [69]:
```
#%sql select unique("Landing _Outcome") from spacextbl
%sql select "Landing _Outcome",count(*) as Total from spacextbl where date between '2010-06-04' and '2017-03-20' group by "Landing _Outcome" order by To
```

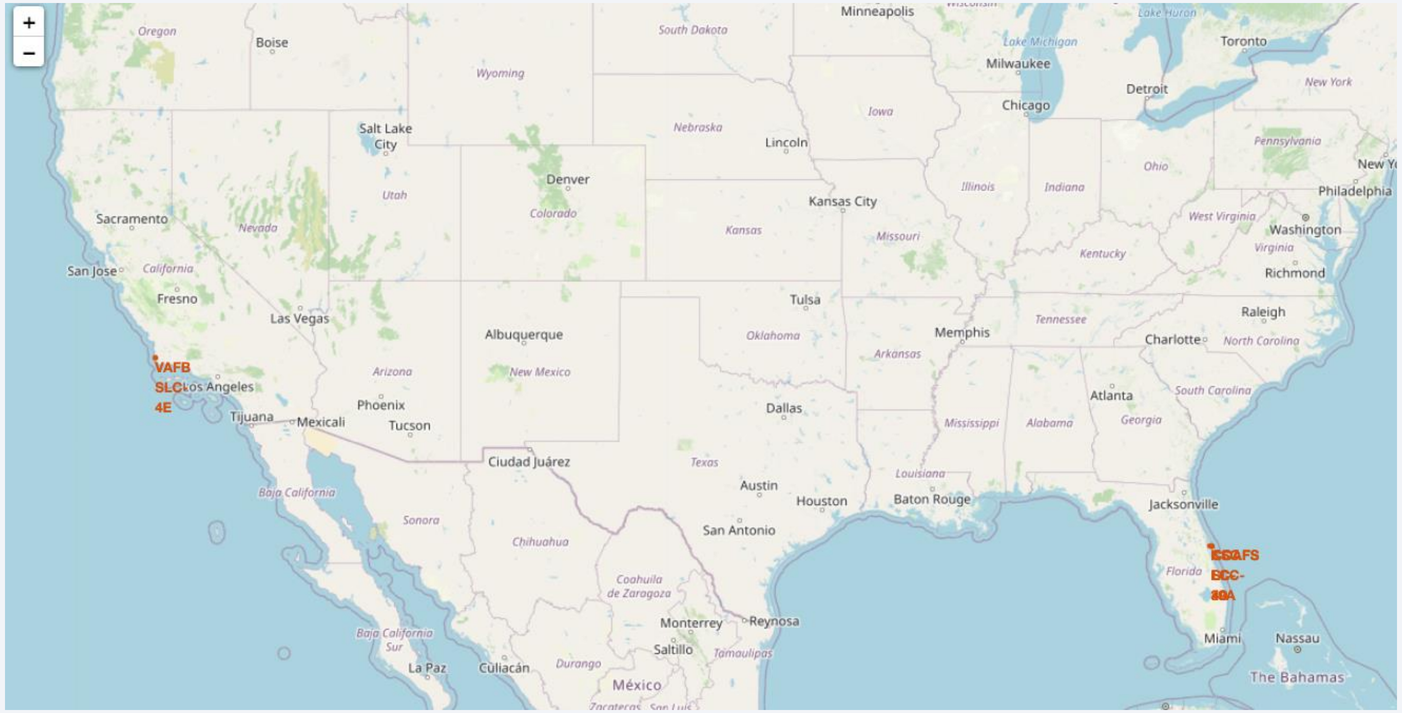
```
 * ibm_db_sa://jrp31646:***@98538591-7217-4024-b027-8baa776ffad1.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30875/bludb
Done.
```

Out[69]:

Landing _Outcome	total
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

Build an Interactive Map with Folium

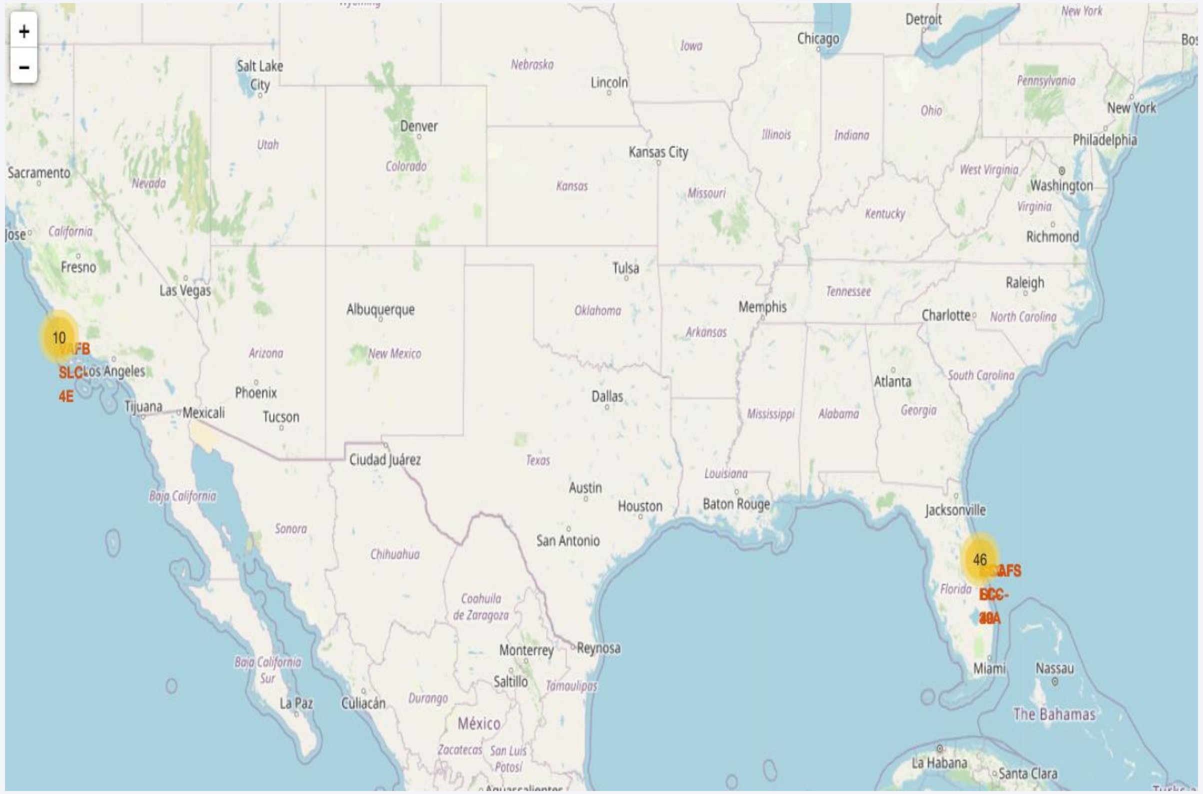
- Create a folium Map object, with an initial center location to be NASA Johnson Space Center at Houston, Texas.
- Use folium.Circle to add a highlighted circle area with a text label on a specific coordinate.

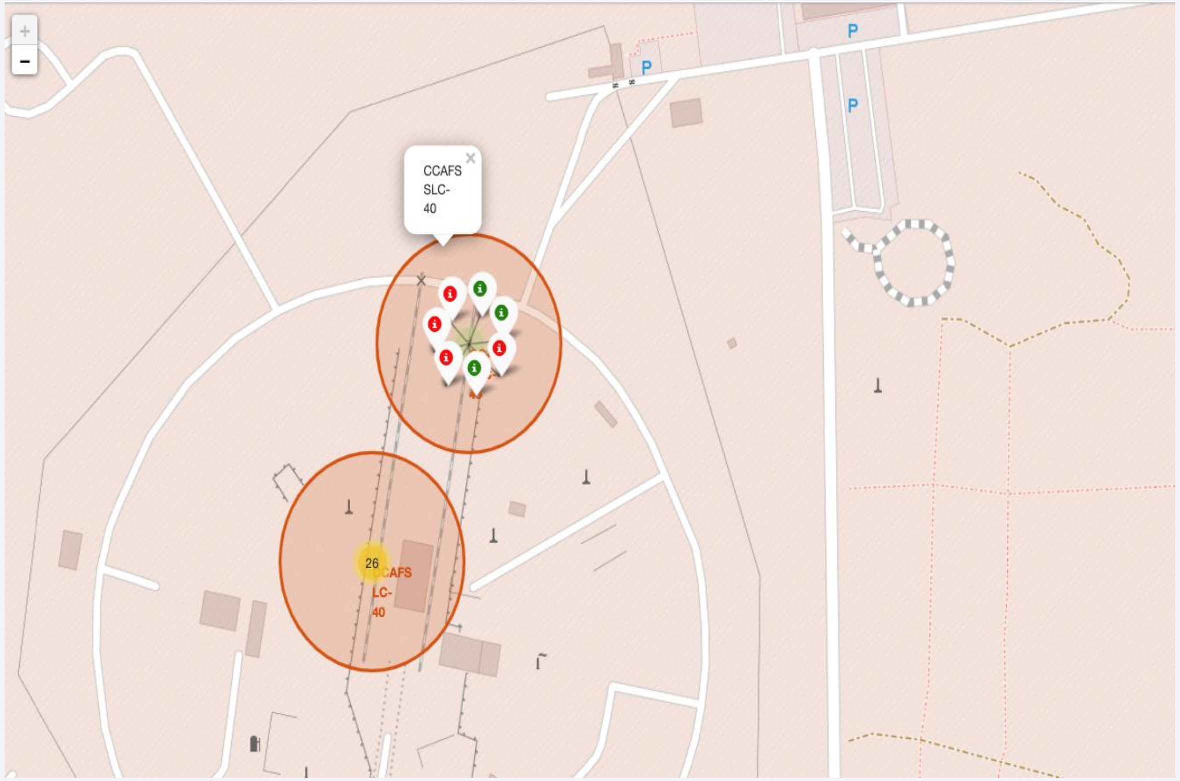


- https://github.com/azraimahadan/AI_ML_DL/blob/main/Interactive%20Visual%20Analytics%20and%20Dashboard.ipynb

Build an Interactive Map with Folium

- Mark the success/failed launches for each site on the map using MarkerCluster object





- https://github.com/azraimahadan/AI_ML_DL/blob/main/Interactive%20Visual%20Analytics%20and%20Dashboard.ipynb

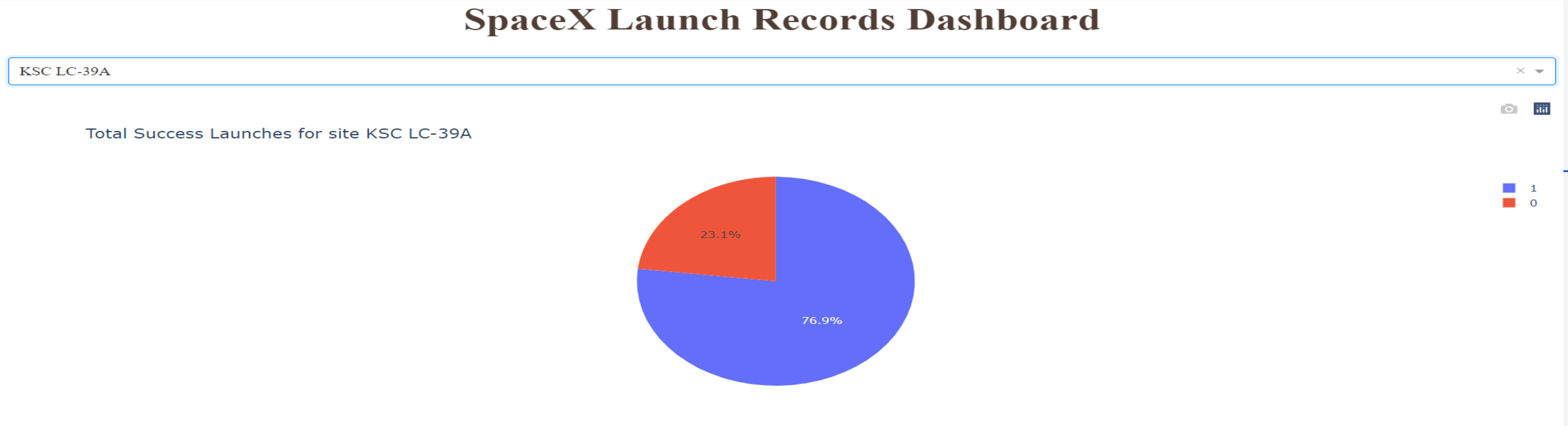
Build a Dashboard with Plotly Dash

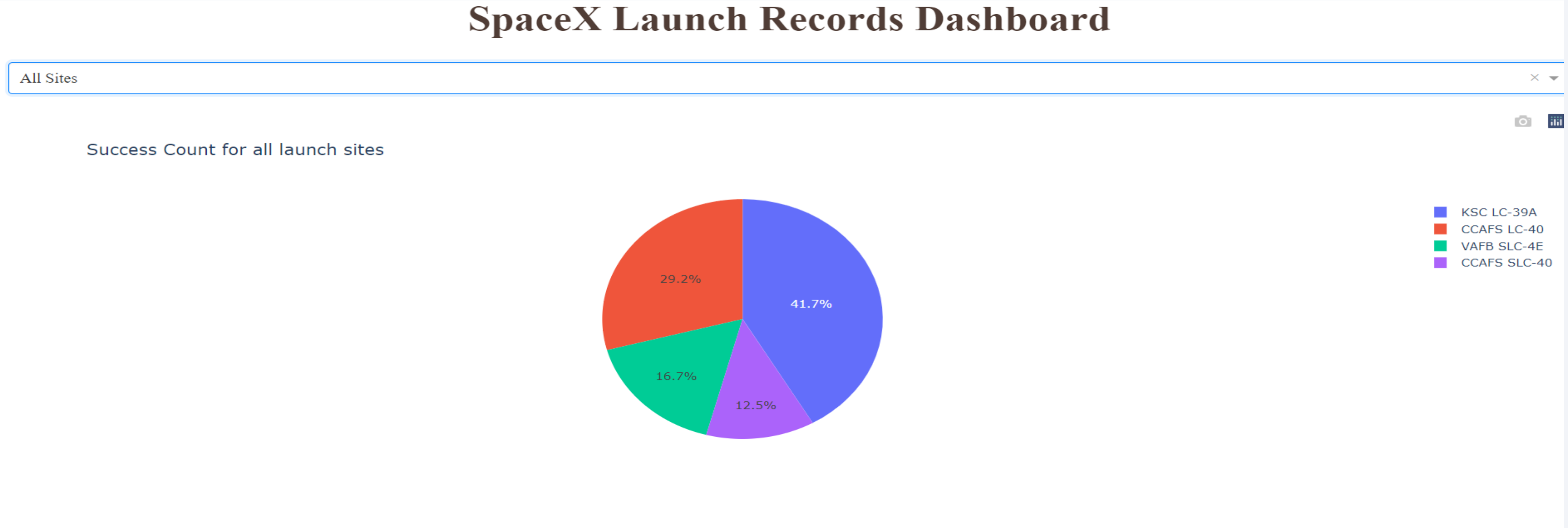
Scatter plot is used to see the relationship between the booster version and the payload mass.
We can see that booster version B5 have payload mass less than 3500kg and 100% success rate.

https://github.com/azraimahadan/AI_ML_DL/blob/main/IBM%20Data%20Science/spacex_dash_app.py

Build a Dashboard with Plotly Dash

FT has higher success rate than other booster version





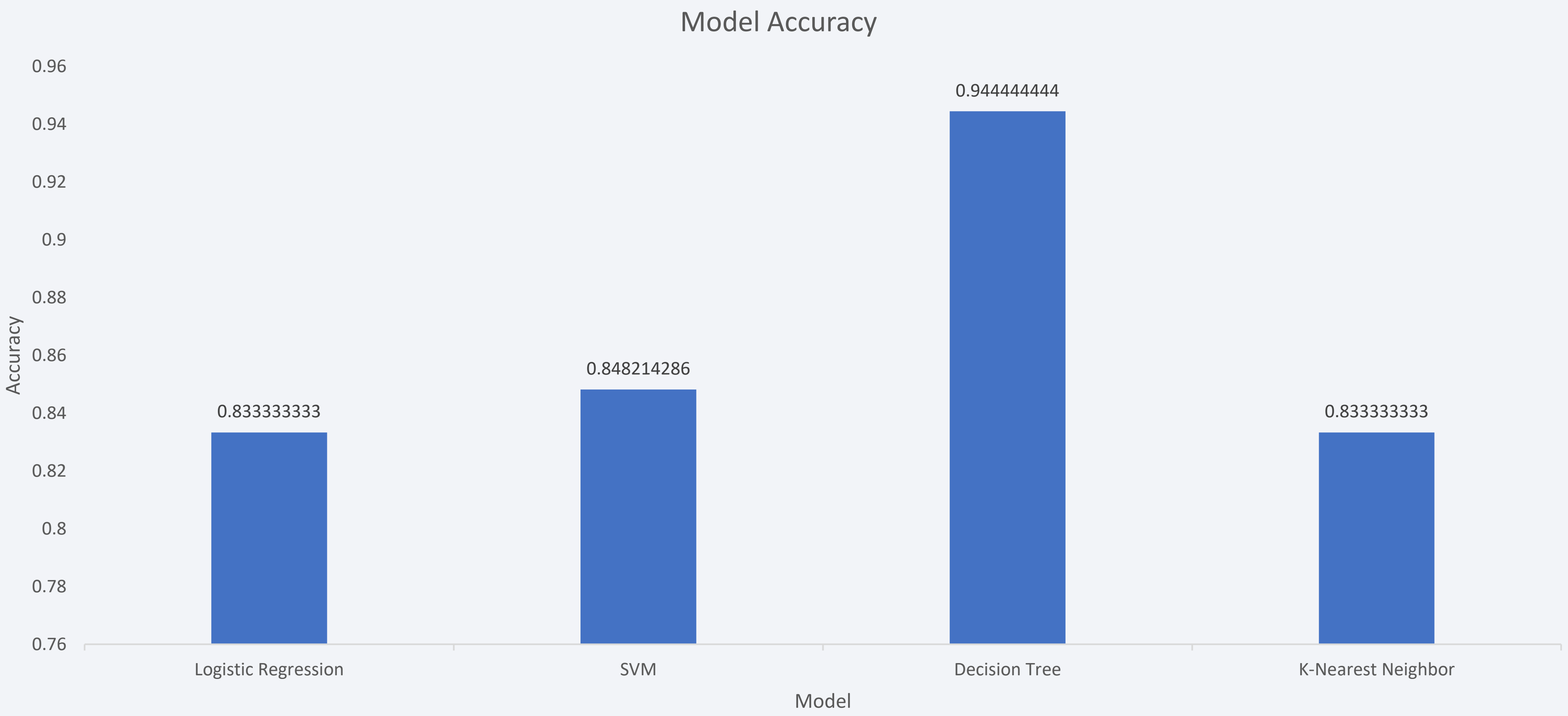
KSC LC-39A has the highest rates of successful launch.

Predictive Analysis (Classification)

- Create a classification machine learning pipeline to predict if the first stage will land given the data prepared before
- Perform exploratory Data Analysis and determine Training Labels
- Create a column for the class
- Standardize the data
- Split into training data and test data using train_test_split function
- Find best Hyperparameter for SVM, Classification Trees and Logistic Regression using GridSearchCV.

https://github.com/azraimahadan/AI_ML_DL/blob/main/Machine%20Learning.ipynb

Classification Accuracy



Decision tree model has the highest accuracy.

Confusion Matrix

- Decision tree model has the highest accuracy among all other models
- It successfully classified almost all test data

Conclusions

- Low weight payload rocket perform better than high weight payload mass
- Decision tree classifier is the best model for prediction of successful launch
- The success rate of launch kept increasing since 2013 until 2020
- Orbit ES-L1,GEO,HEO, and SSO have 100% success rates
- KSC LC-39A has the highest rates of successful launch

Appendix

Successful launch

Unsuccessful launch

- https://en.wikipedia.org/wiki/List_of_Falcon%5C_9%5C_and_Falcon_Heavy_launches
- Data Set 1 (Data Wangling)
- Data set 2 (Processed)

Thank you!