IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Sasan Sahraei
27/03 /2024

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection
  - Data Wrangling
  - Data Visualisation
  - Interactive Visual Analytics with Folium
  - Predictive Analytics using classification models
- Summary of all results
  - Correlation Variables
  - Build predictive results

Introduction

- Project background and context
    - Companies are making space travel affordable for everyone. Virgin Galactic is providing suborbital spaceflights. Rocket Lab is a small satellite provider. Blue Origin manufactures sub-orbital and orbital reusable rockets. Perhaps the most successful is SpaceX. SpaceX's accomplishments include: Sending spacecraft to the International Space Station. Starlink, a satellite internet constellation providing satellite Internet access. Sending manned missions to Space. One reason SpaceX can do this is the rocket launches are relatively inexpensive. Therefore, here we determine if the first stage will landand if the launch is cost effective.
- Problems you want to find answers
    - Factors for successful launch
    - Factors for cost effectiveness

Section
1
Methodology

Methodology

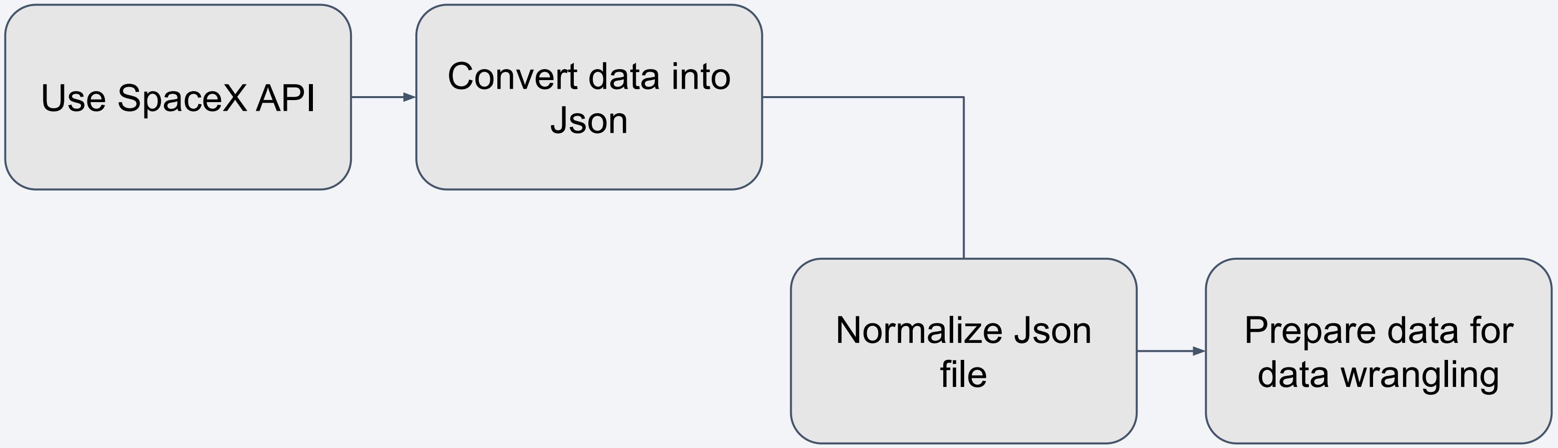
Executive Summary

- Data collection methodology:
  - SpaceX launch data that is gathered from the SpaceX REST API.
  - Falcon 9 Launch data is web scraping related Wiki pages
- Perform data wrangling
  - Exploratory Data analytics to find patterns in data and determine training data for supervised models
- Perform exploratory data analysis (EDA) using visualization and SQL
  - Exploratory Data analytics using python library and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
  - Use Folium to view previously observed correlations
- Perform predictive analysis using classification models
  - Create machine learning pipeline to predict if the first stage will land given the data from preceding observations

Data Collection

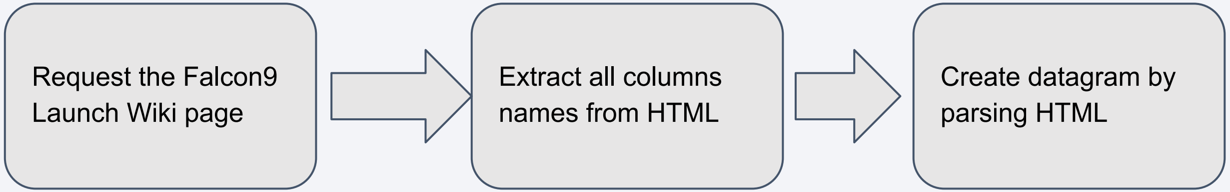
- Data collection
  - SpaceX API
  - Decoded the response as JSON using .json function
  - Normalise it using .json_normalize function
  - Clean data by checking missing values and fill in missing values
  - Webscraping from Falcon 9 launch wikipedia

Data Collection – SpaceX API



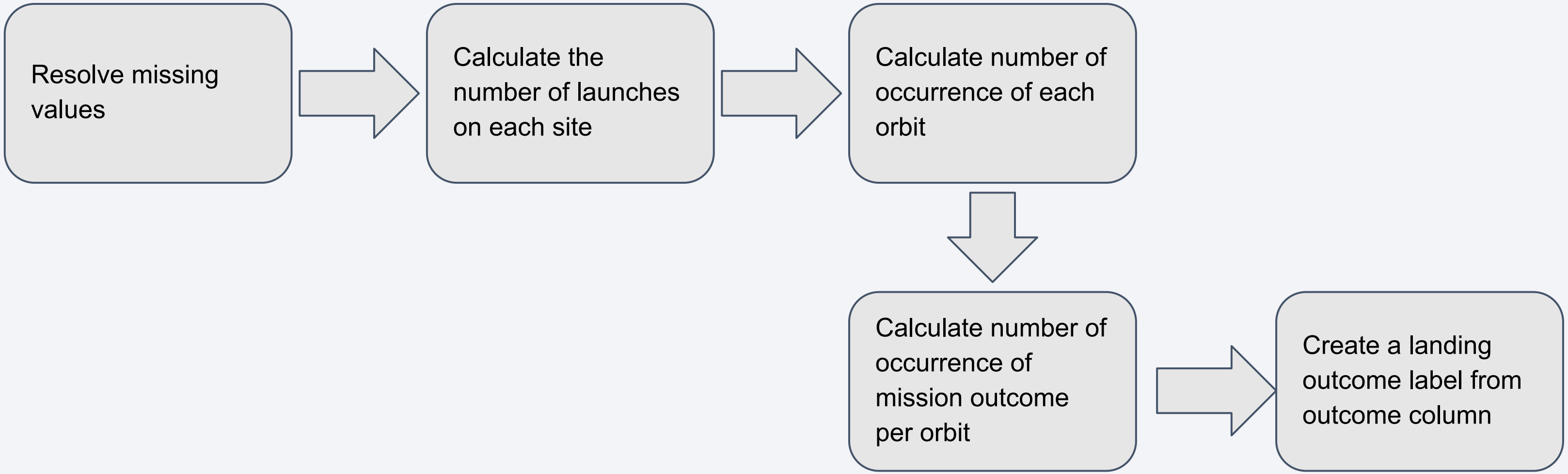
- Access details in [github]

Data Collection - Scraping



Access github for more detail

Data Wrangling



More information available in [github]

EDA with Data Visualization

- Following graphs are used:
  - Scatter plot
  - Bar chart
  - Line chart

- More detailed information on [github]

EDA with SQL

- Display the names of the unique launch sites in the space mission
- Display 5 records where launch sites begin with the string 'CCA'
- Display the total payload mass carried by boosters launched by NASA (CRS)
- Display average payload mass carried by booster version F9 v1.1
- List the date when the first successful landing outcome in-ground pad was achieved
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- List the total number of successful and failure mission outcomes
- List the names of the booster versions which have carried the maximum payload mass. Use a subquery
- List the failed landing outcomes in drone ship, their booster versions, and launch site names for the in year 2015
- 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

Detailed information in [github]

Build an Interactive Map with Folium

- folium.Marker() was used to create marks on the maps.
- folium.Circle() was used to create a circles above markers on the map.
- folium.Icon() was used to create an icon on the map.
- folium.PolyLine() was used to create polynomial line between the points.
- folium.plugins.AntPath() was used to create animated line between the points.
- markerCluster() was used to simplifv the maps which contain several markers with identical coordination

More information in [github]

Build a Dashboard with Plotly Dash

- Dash and html components were used as they are the most important thing and
- almost everything depends on them, such as graphs, tables, dropdowns, etc.
- Pandas was used to simplifying the work by creating dataframe.
- Plotly was used to plot the graphs.
- Pie chart and scatter chart were used to for plotting purposes.
- Rangeslider was used for payload mass range selection.
- Dropdown was used for launch sites.

Predictive Analysis (Classification)

- Building the model
  - Create column for the class
  - Standardize the data
  - Split the data info train and test sets
  - Build GridSearchCV model and fit the data
- Evaluating the model
  - Calculating the accuracies
  - Calculating the confusion matrix
  - Plot the results
- Finding the optimal model
  - Find the best hyperparameters for the models
  - Find the best model with highest accuracy
  - Confirm the optimal model

More information in github

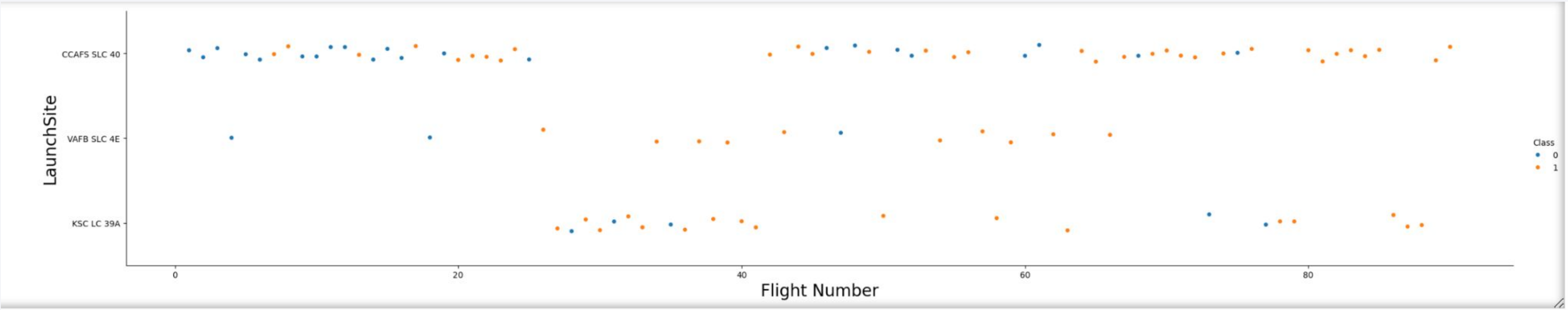
Results

- Exploratory data analysis results
  - Both API and web scraping are capable to collect Xspace data
- Interactive analytics demo in screenshots
  - EDA with SQL is effective for data filtering
  - EDA with interactive visualization provides informative information
  - Plotly Dash is powerful to show instant data change
- Predictive analysis results
  - Decision Tree Classifier Algorithm has the best accuracy of predicting.

Section 2

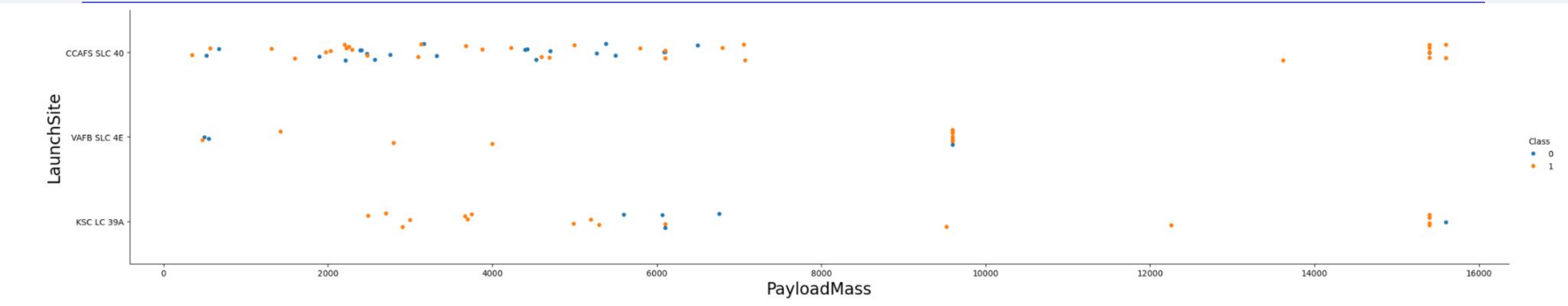
Insights drawn from EDA

Flight Number vs. Launch Site



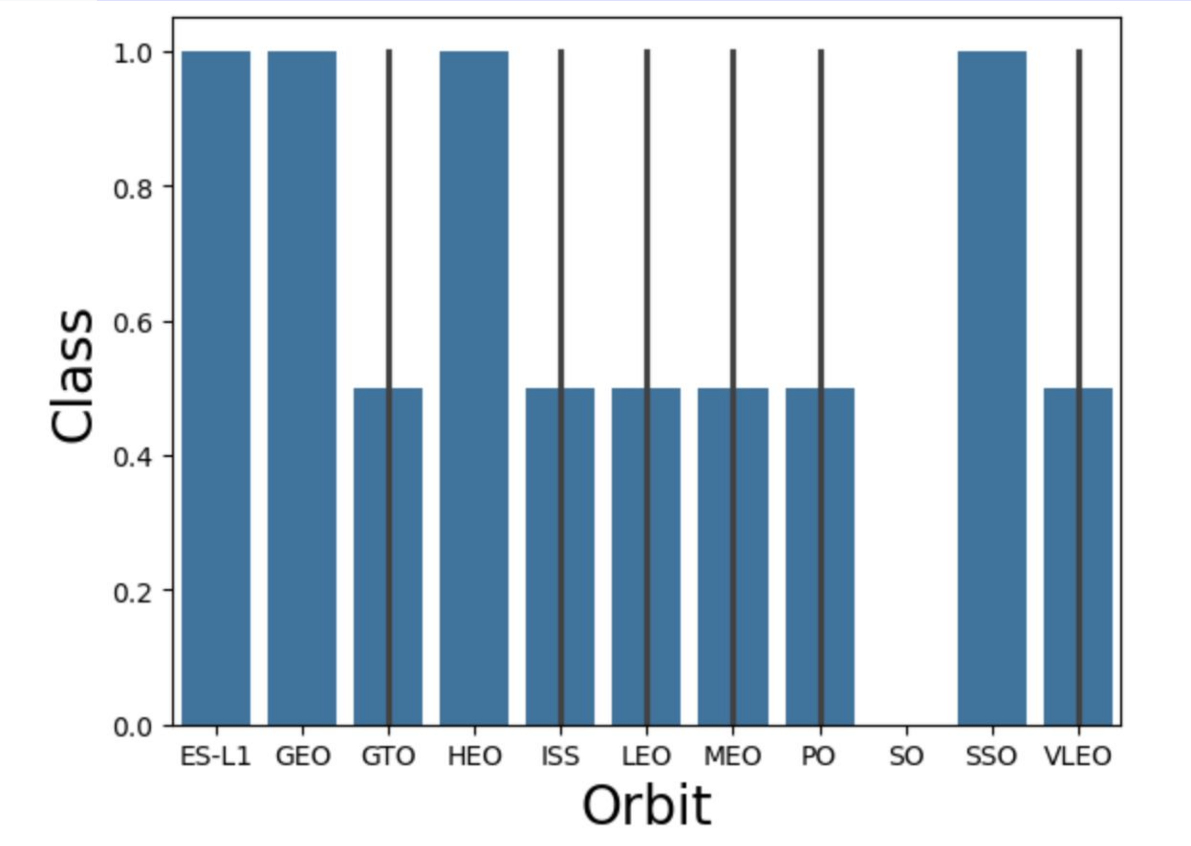
From the plot we found the larger the flight amount at a launch site the greater the success rate

Payload vs. Launch Site



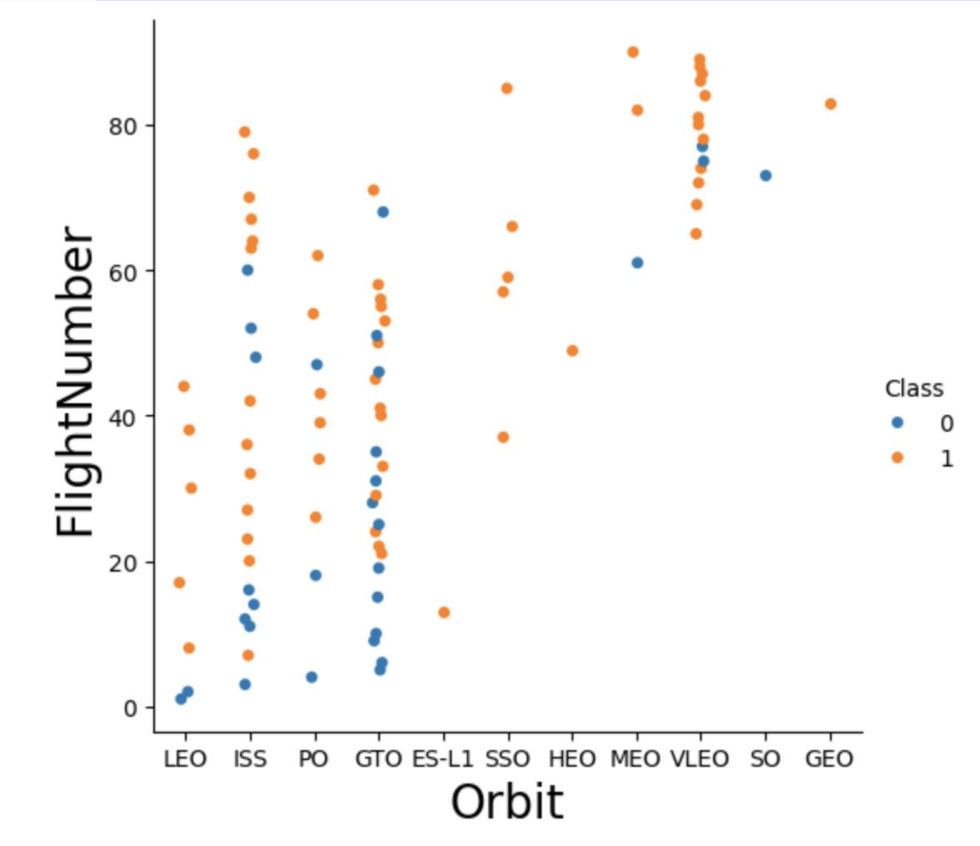
We did not see strong correlation

Success Rate vs. Orbit Type



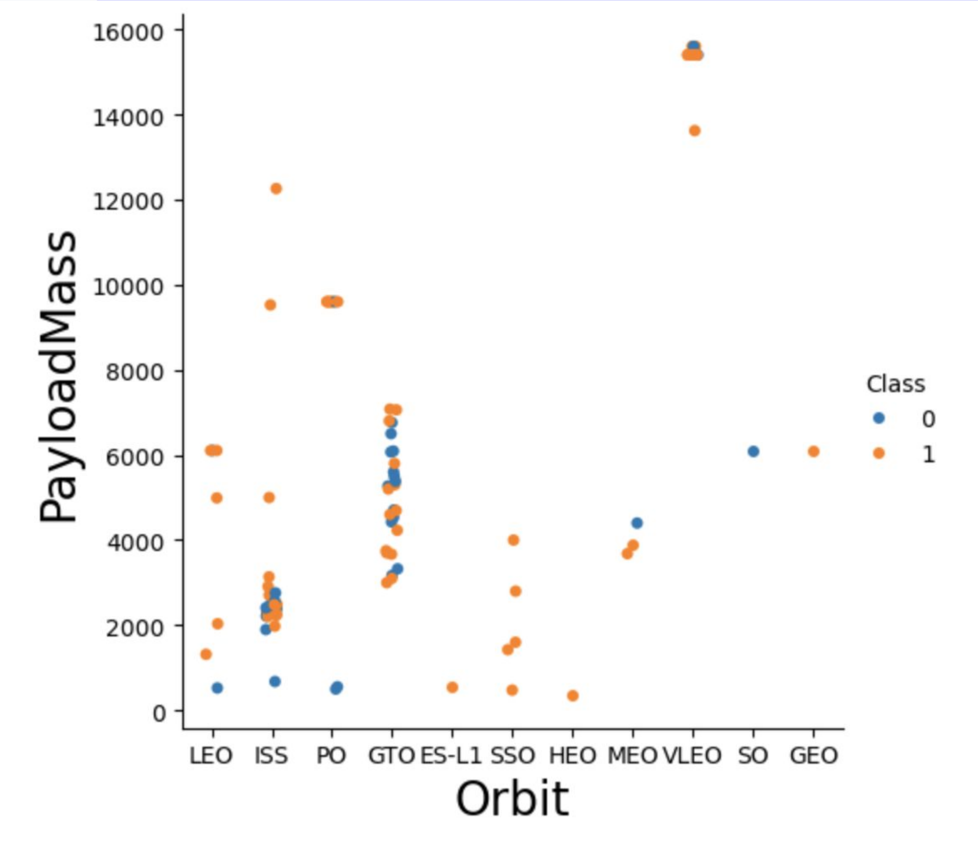
ES-L1, GEO, HEO and SSO have success rate of %100

Flight Number vs. Orbit Type



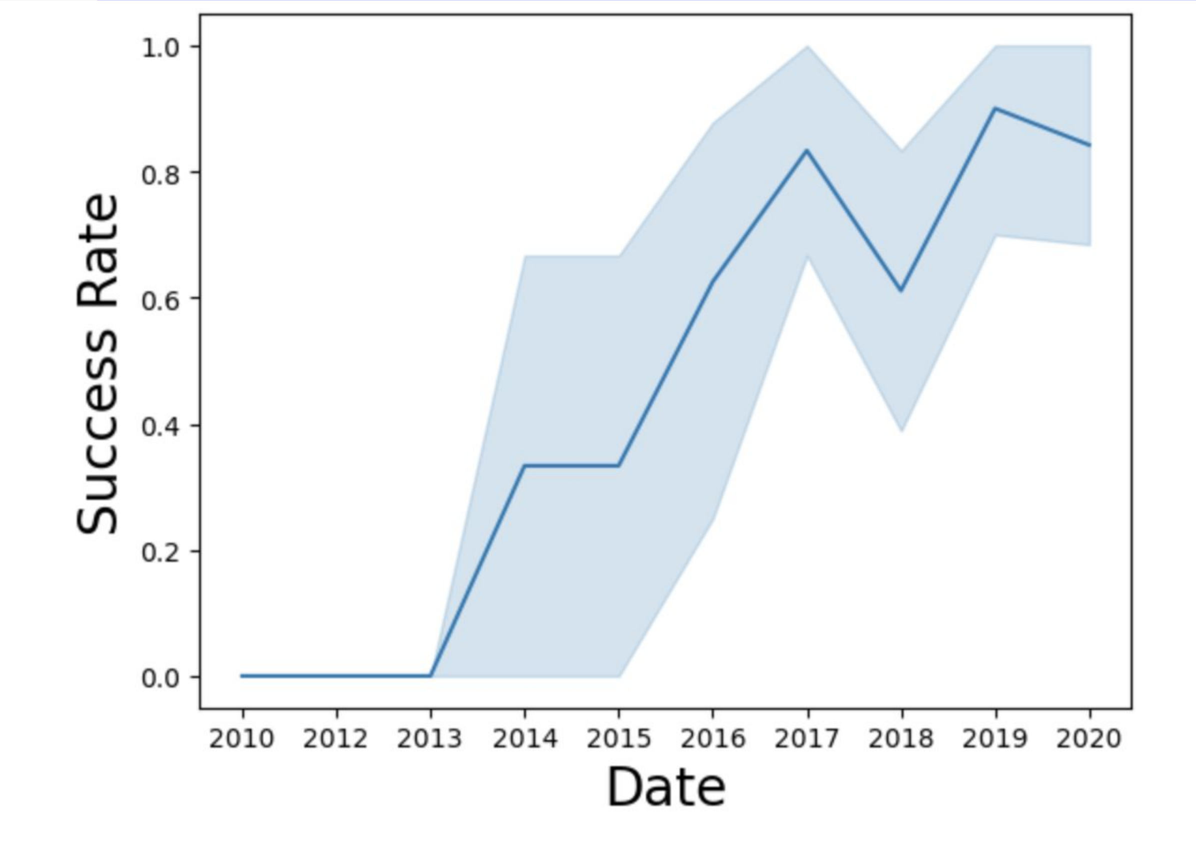
There does not seem to be relationship between flight number and GTO

Payload vs. Orbit Type



We observe different behaviours for payload mass between 2000 to 3000 and 3000 to 8000

Launch Success Yearly Trend



Launch success rate has increased significantly since 2013 and has stabilized since 2019, potentially due to advance in technology and lessons learned.

All Launch Site Names

```
%sql select distinct(Launch_Site) from SPACEXTABLE
```

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

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

Launch Site Names Begin with 'CCA'

```
%sql select * from SPACEXTABLE where Launch_Site like 'CCA%' limit 5
```

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

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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	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 SPACEXTABLE where Customer = 'NASA (CRS)'
```

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

sum(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

```
%sql select avg(PAYLOAD_MASS__KG_) from SPACEXTABLE where Booster_Version like '%F9 v1.1%'
```

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

avg(PAYLOAD_MASS__KG_)
2534.6666666666665

First Successful Ground Landing Date

```
%sql select Date from SPACEXTABLE where Landing_Outcome like 'Success (ground pad)' order by Date ASC limit 1
```

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

Date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql select customer from SPACEXTABLE where PAYLOAD_MASS__KG_ between 4000 and 6000 and mission_outcome = 'Success'
```

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

Customer
AsiaSat
AsiaSat
ABS Eutelsat
Turkmenistan National Space Agency
SES
SKY Perfect JSAT Group
SKY Perfect JSAT Group
EchoStar
SES
NRO
U.S. Air Force
SES EchoStar
SES
SES
Telkom Indonesia
Es hailSat
Spaceflight Industries
PSN, SpaceIL / IAI
Canadian Space Agency (CSA)
U.S. Space Force
Republic of Korea Army, Spaceflight Industries (BlackSky)
USSF

Total Number of Successful and Failure Mission Outcomes

```
%sql select count(*) from SPACEXTABLE where Mission_Outcome like '%Failure%'
%sql select count(*) from SPACEXTABLE where Mission_Outcome like '%Success%'
```

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

count(*)
100

Boosters Carried Maximum Payload

```
%sql select Booster_Version from SPACEXTABLE where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_) from SPACEXTABLE )
```

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

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 substr(Date, 6,2) from SPACEXTABLE where substr(Date,0,5)='2015' and Landing_Outcome like '%Failure%'
```

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

substr(Date, 6,2)
01
04

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
```

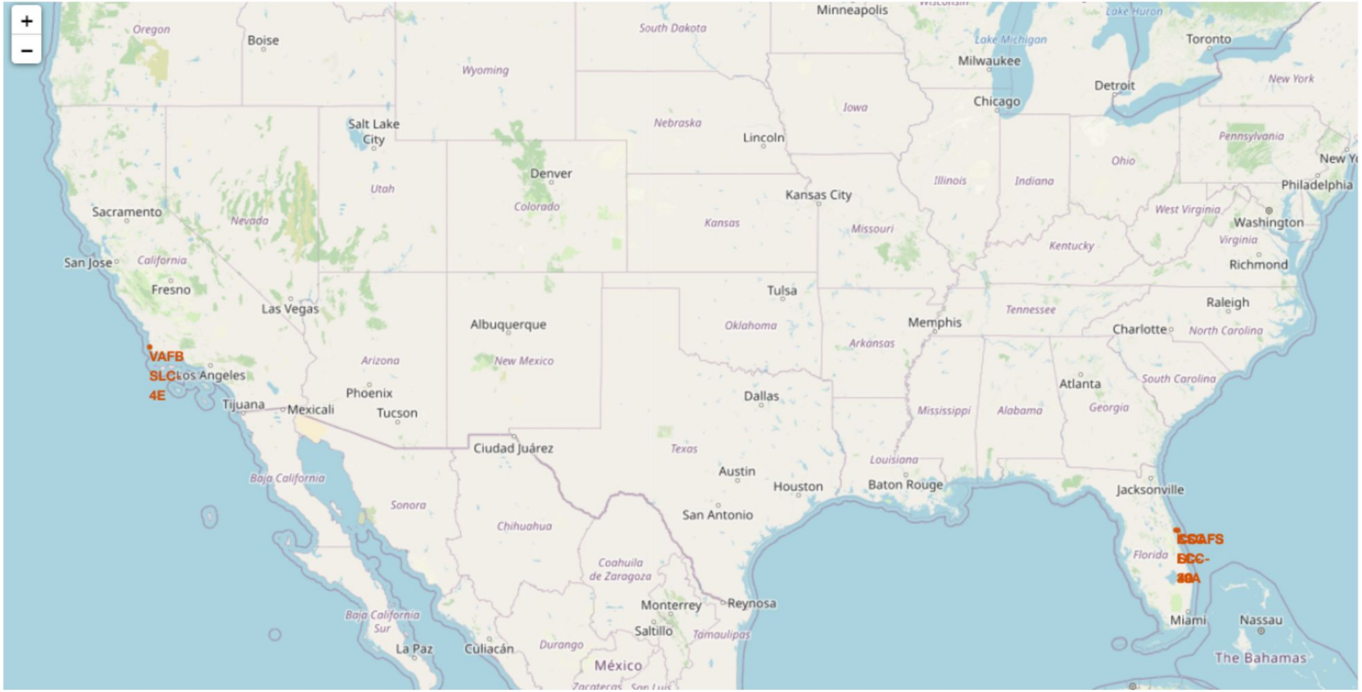
```
 * sqlite:///my_data1.db
Done.
```

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2017-02-19	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
2017-01-14	17:54:00	F9 FT B1029.1	VAFB SLC-4E	Iridium NEXT 1	9600	Polar LEO	Iridium Communications	Success	Success (drone ship)
2016-08-14	5:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2016-07-18	4:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
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	5:21:00	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	1:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)

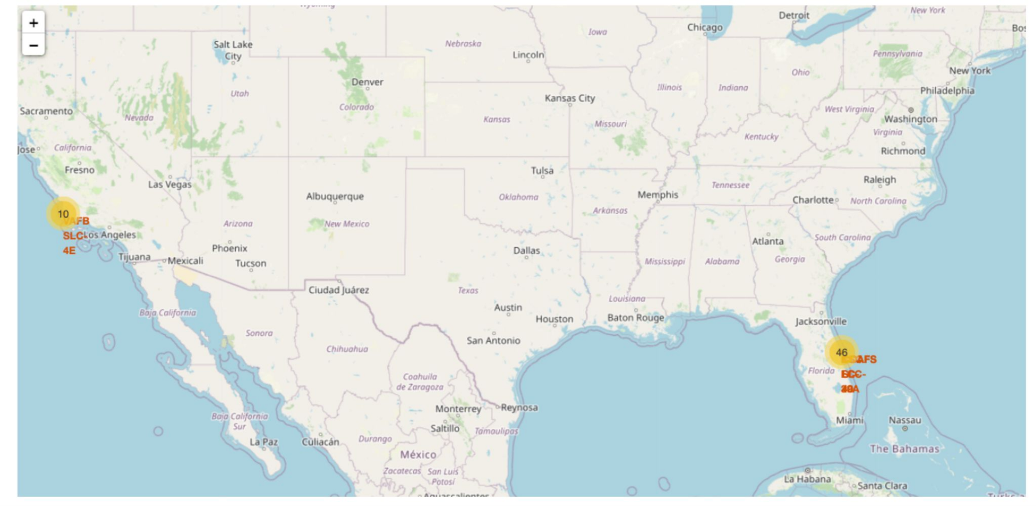
Section 3

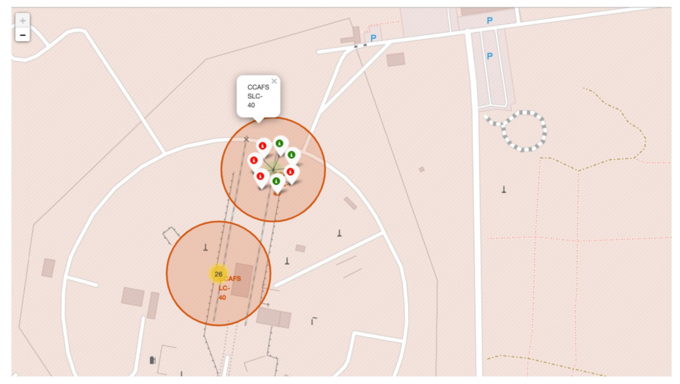
Launch Sites Proximities Analysis

All Launch Sites marked on map

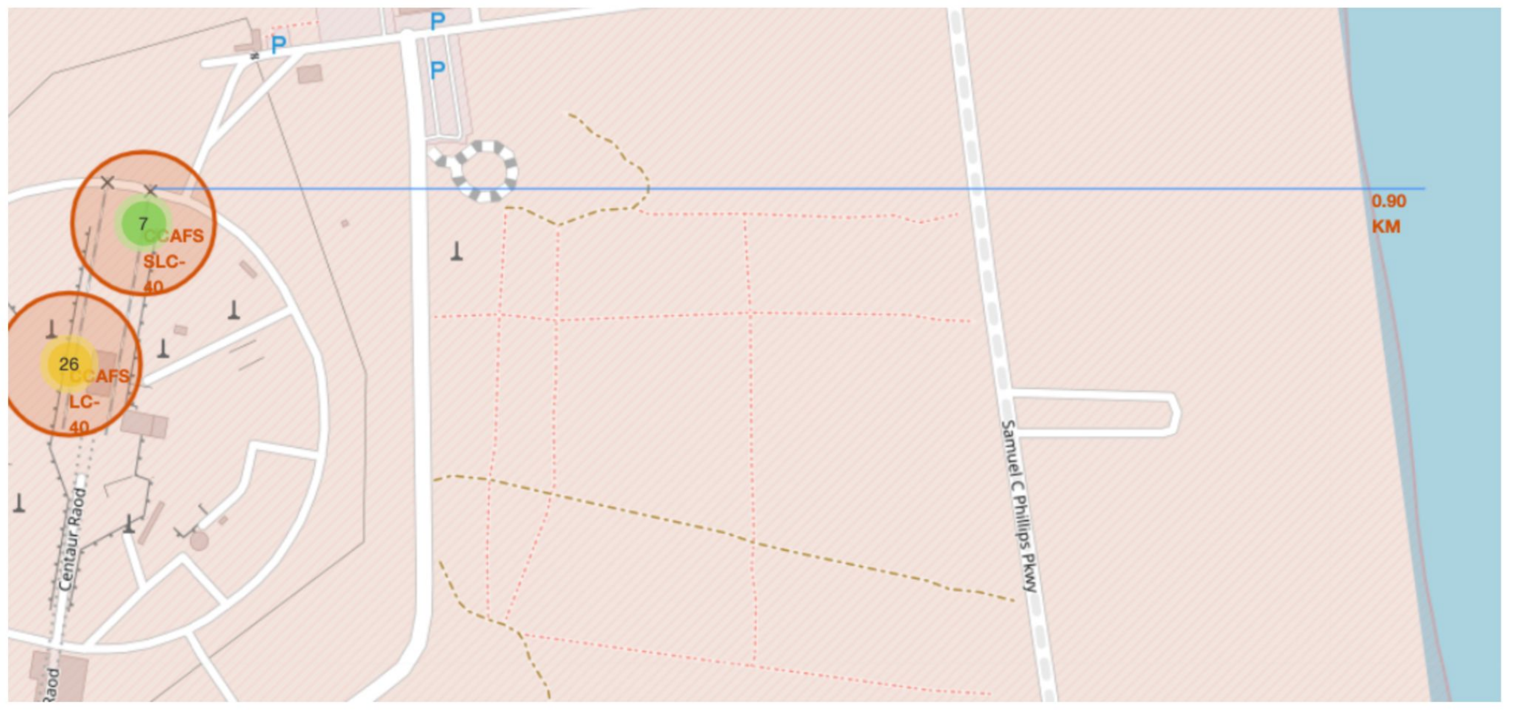


Success/Failure marked on map





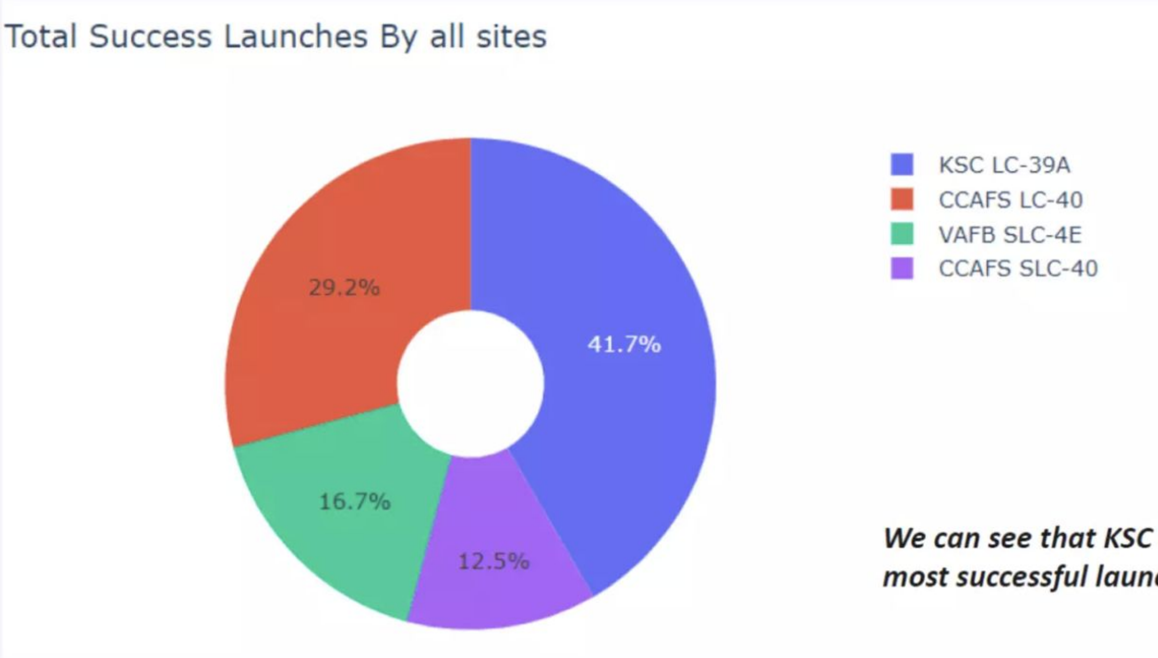
Distances between a launch site to its proximities



Section
4

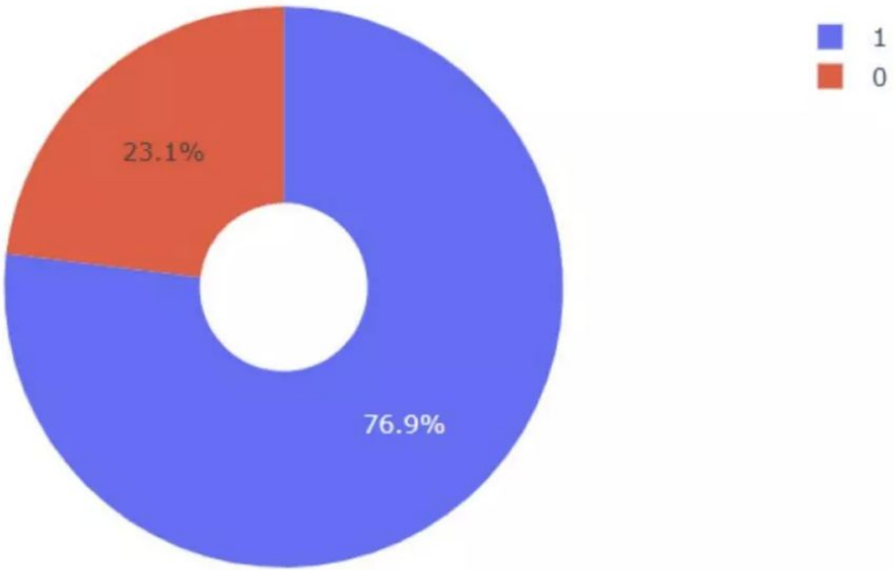
Build a Dashboard with Plotly Dash

Total success launches by all sites



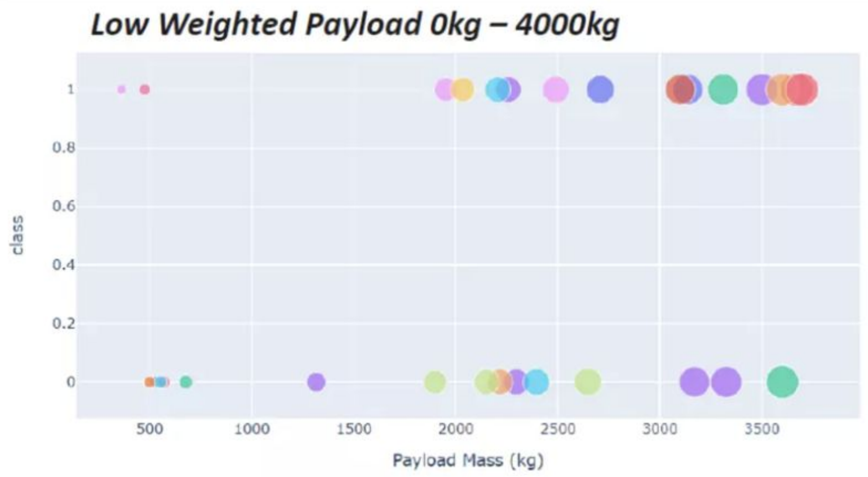
We can see that KSC LC-39A had the most successful launches from all the sites

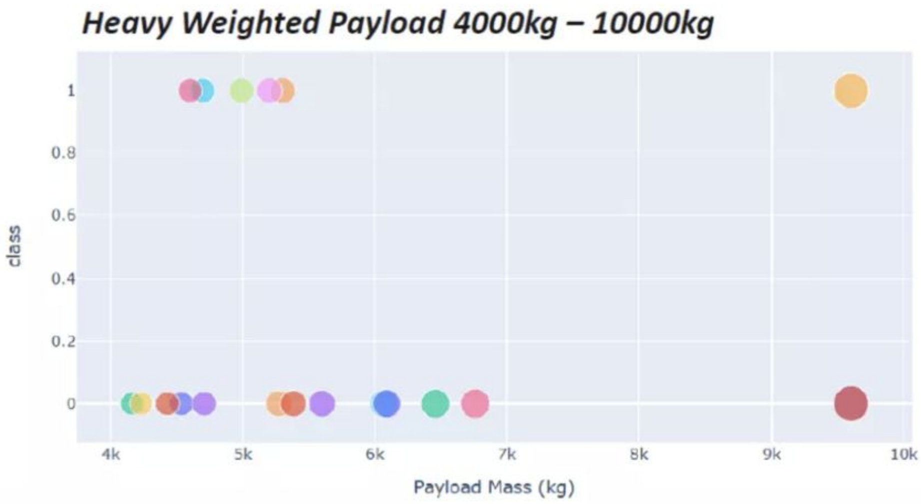
Success Rate by Site



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

Payload vs Launch Outcome



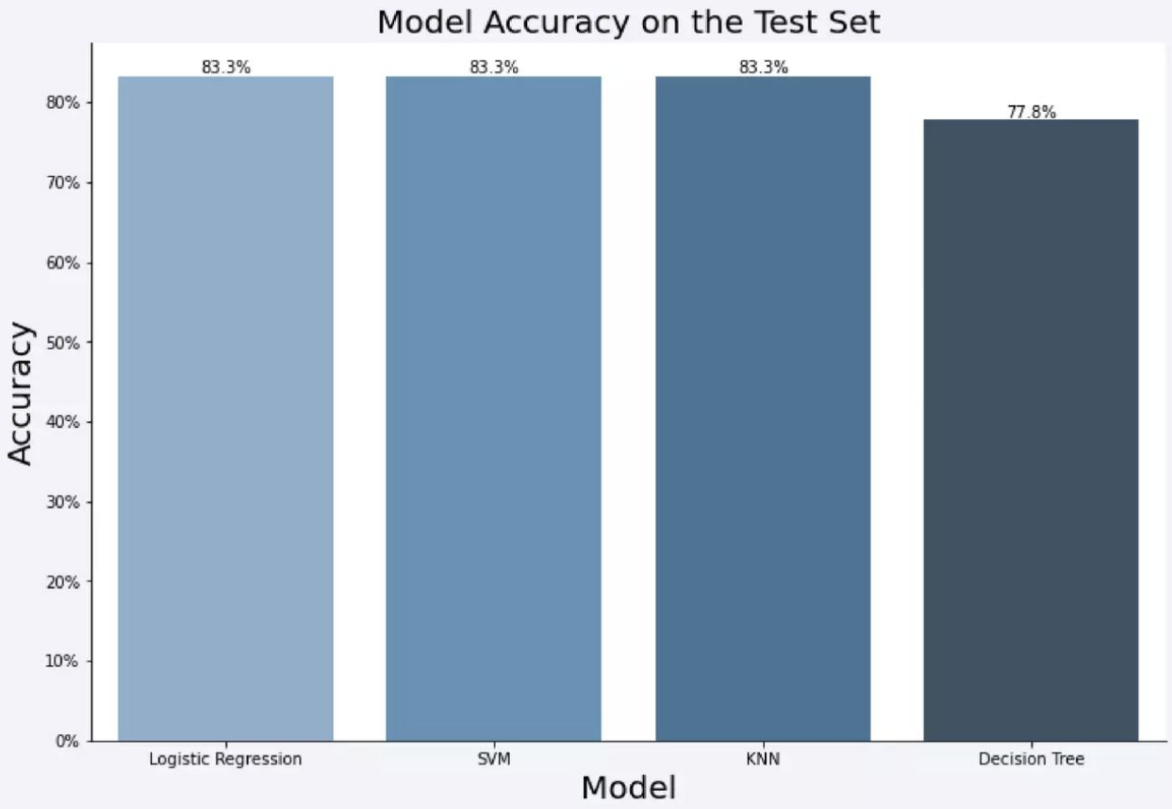


We can see the success rates for low weighted payloads is higher than the heavy weighted payloads

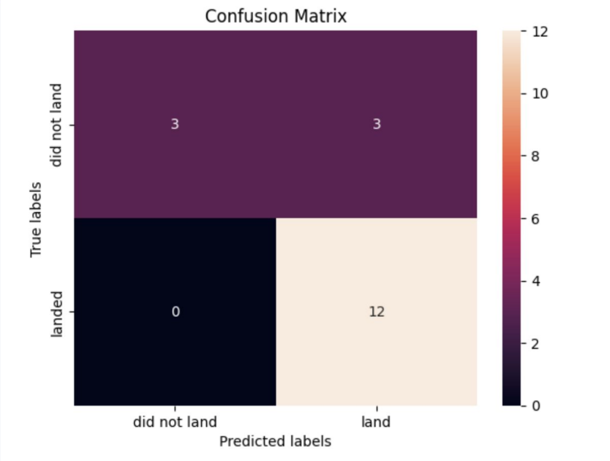
Section
5

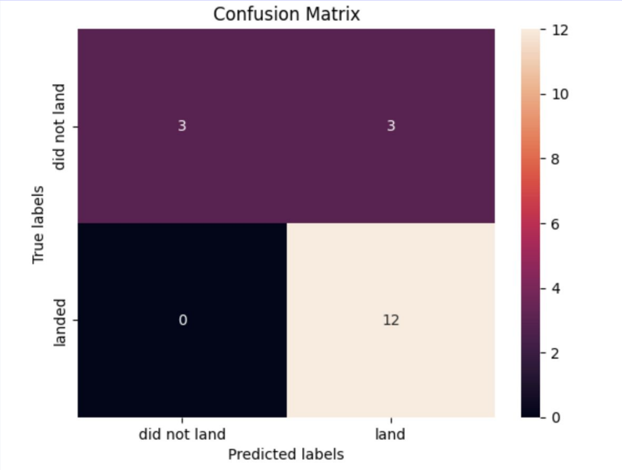
Predictive Analysis (Classification)

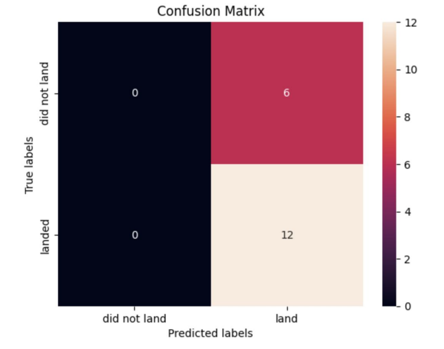
Classification Accuracy

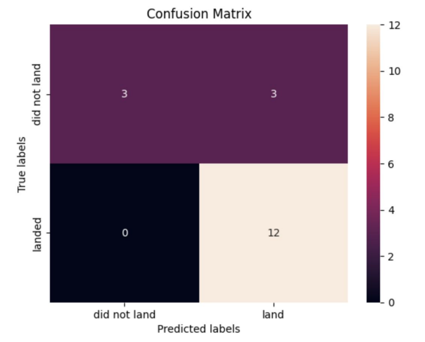


Confusion Matrix









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 Space 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

All information can be found in github

Thank you!