IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Arun Venkateswaran
03-12-2024

Outline

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

Executive Summary

- Summary of methodologies
  - Data was collected through web scrapping and SpaceX API
  - Data wrangling, exploratory data analysis with visualization
  - Interactive visualizations and dashboard
  - Predictive Analysis
- Summary of all results
  - EDA helped identify features impacting launch
  - Predictive analysis identified the best suited model

Introduction

- Project background and context
  - To be able to develop capability to determine if the first stage would successfully land after a launch so it can be reused based on the past launch data of SapceX and predictive model.
- Problems you want to find answers
  - How various feature impacts the occurrence for a successful landing.
  - A model that can predict successful landings of the first stage

Section 1

Methodology

Methodology

Executive Summary

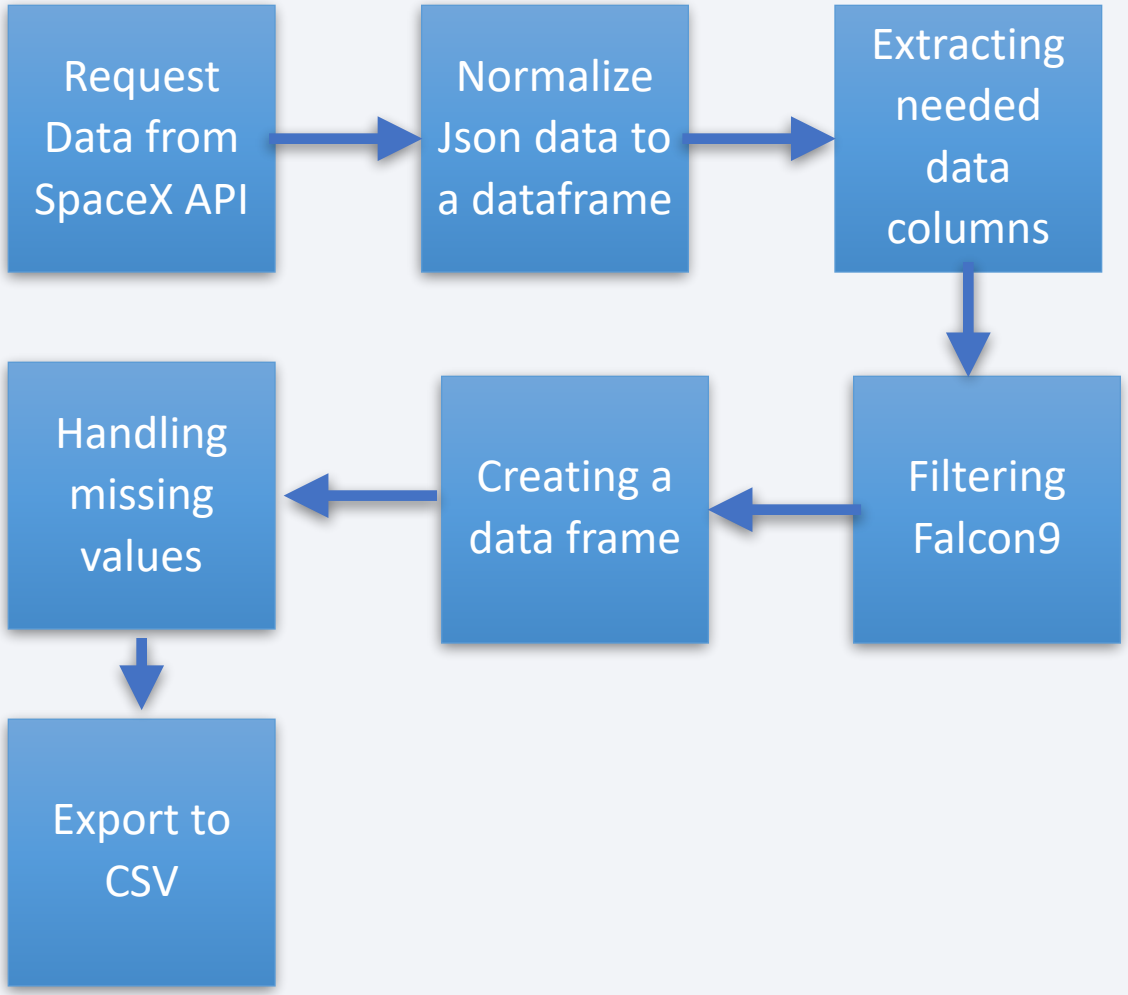
- Data collection methodology:
  - Data was collected using SpaceX rest API and from Wikipedia through web scrapping
- Perform data wrangling
  - Data was normalized, dealt with missing values, labeled data, used Hot encoding
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Data was cleaned and classified into training and test sets.Different Classifications models accuracy were evaluated based on the classified data set

Data Collection

- Data was called by using SpaceX Rest API and web scrapping SpaceX data from Wikipedia.
- Both the sources had to be used to get a complete list of all the key features needed to make a complete data set

Data Collection – SpaceX API

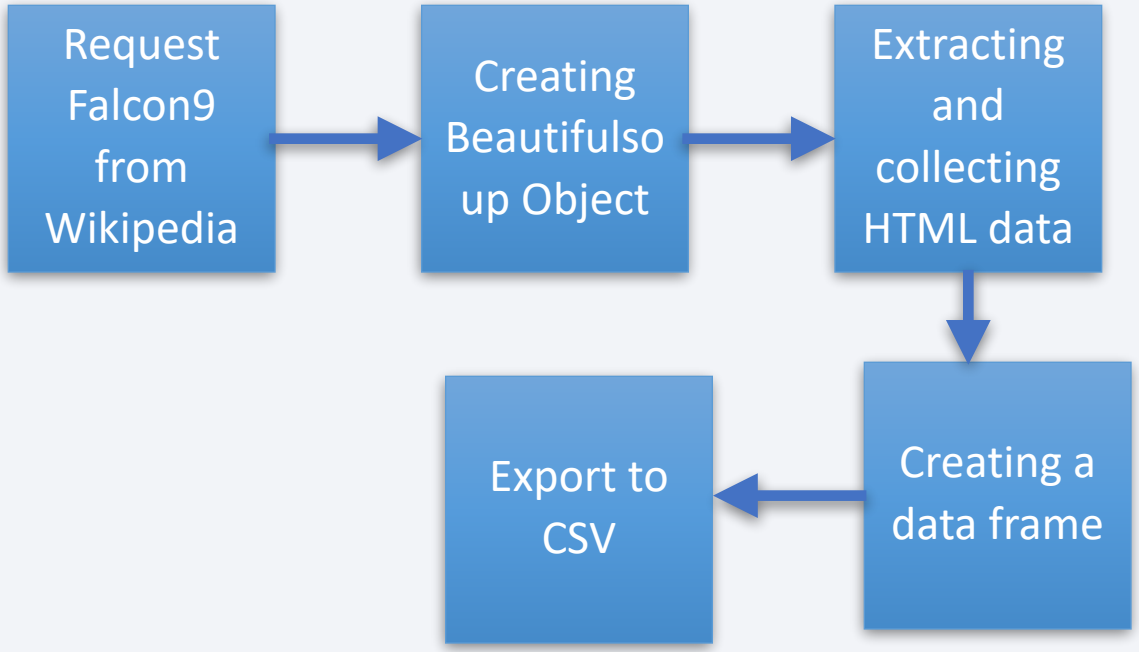
- Data from SpaceX API was normalized to a data frame and Falcon9 data was filtered.
- Missing Payload and Mass values were replaced by mean values
- Github URL-[Data-Collection-API]



Data Collection - Scraping

- Web-scrapping Falcon9 data from wikipedia through a Beautifulsoup object and converting it to a dat frame and exporting it to a CSV

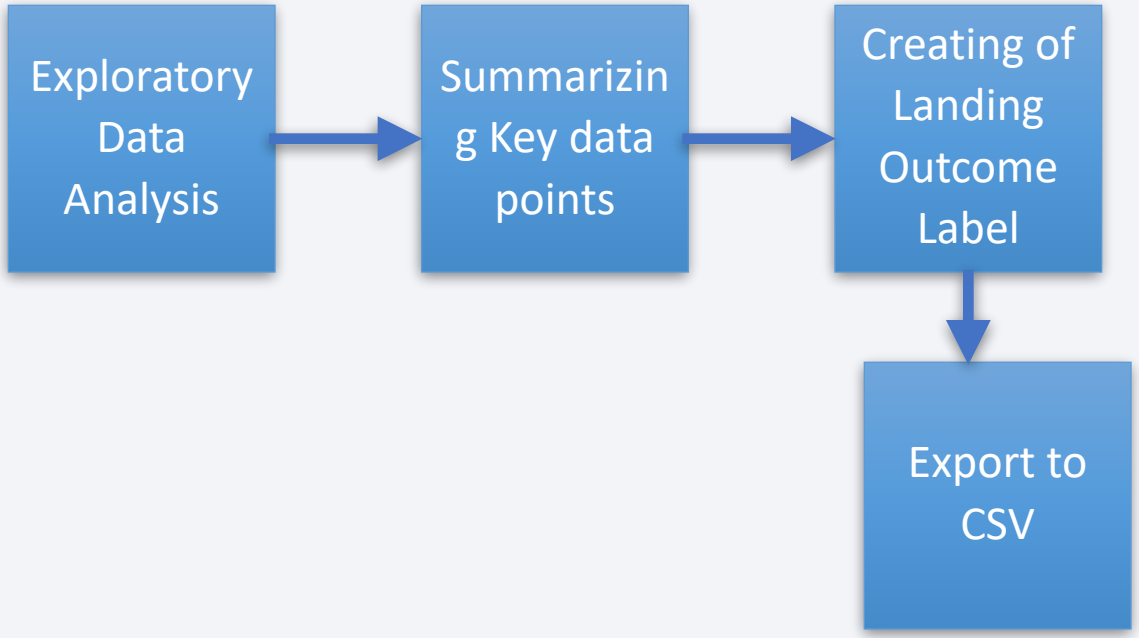
- Github Url-[Data Collection-Web-Scrapping]



Data Wrangling

- A preliminary exploratory analysis was done.
- Some key items were summarized- launches per site, number and occurrence of mission outcome of the orbits, landing outcome label from Outcome column
- Finally data was exported to a CSV

- GitHub URL- [Data Wrangling]



EDA with Data Visualization

- Below scatter plots were plotted to view the relationship between features
  - Flight Number vs Launch Site ,Flight Number vs Payloadmass, Payloadmass vs Launch Site, Flight Number vs Orbit, Payloadmass Vs Orbit.
- Bar Chart displaying the relationship between Orbit and the corresponding mean success rate within each Orbit
- Line chart displaying the success rate over the years
- GitHub URL- [EDA-Data-Vizualization]

EDA with SQL

- Following SQL Queries were performed
  - Names of the unique launch sites in the space mission
  - Top 5 records where launch sites begin with the string ‘CCA'
  - Total payload mass carried by boosters launched by NASA (CRS)
  - Average payload mass carried by booster version F9 v1.1
  - Listing the date when the first successful landing outcome in ground pad was achieved
  - Listing the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
  - Total number of successful and failure mission outcomes
  - Names of the booster versions which have carried the maximum payload mass
  - Failed landing outcomes in drone ship, their booster versions and launch site names for the months in year 2015
  - Ranking the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date2010-06-04 and 2017-03-20 in descending order
- GitHub URL- EDA-SQL

Build an Interactive Map with Folium

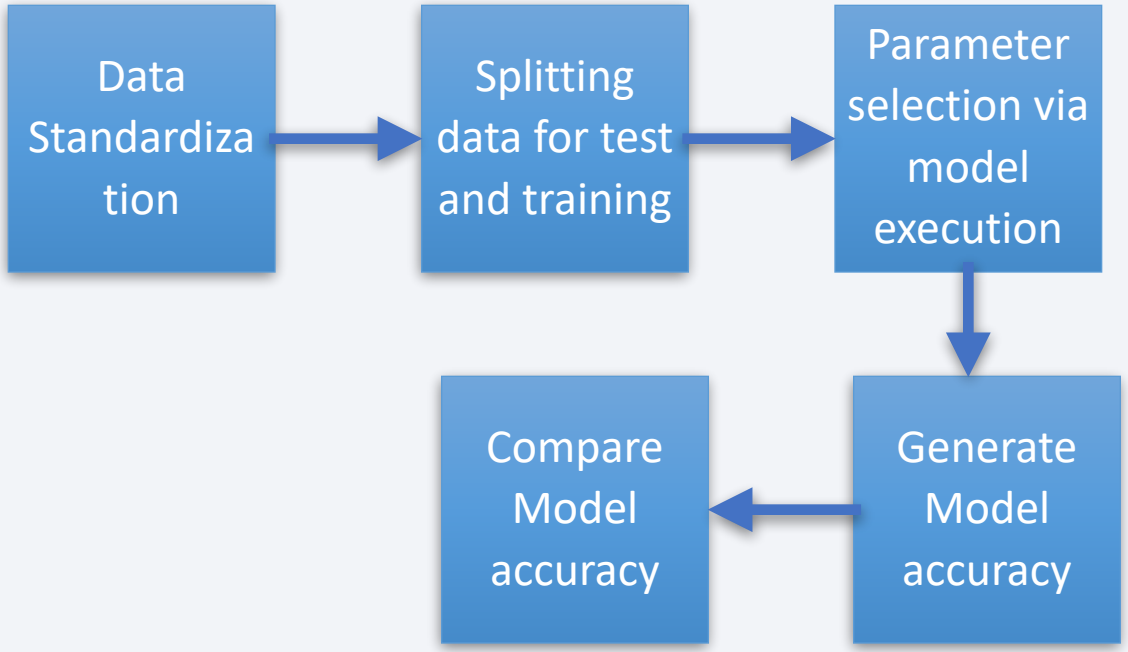
- Added Marker with circle, Pop up and text labels of Nasa space center and SpaceX launch sites using the geographical coordinates
- Highlighted the launch sites using green and red colored marked to provide a visual inference of success rates relative to the launch sites.
- Added proximity lines between a launch site and a city, a landmark near the launch site
- The plotting of the sites in map provides a insight into the success rates of launches and the sites geographical location and proximity to other landmarks
- GitHub URL- Folium-Interactive-Map

Build a Dashboard with Plotly Dash

- Dashboard was built with below options
  - Dropdown list to select the sites to analyze them
  - Pie chart showing the total success rates and success vs failed count for a site based on drop down selection.
  - A slider was included to change the Payloadmass
  - Scatter chart showing the relation between launch success and payload
- Helps to establish the success rates between launch sites and payload mass visually through the plots
- GitHub URL- Plotly-Dash

Predictive Analysis (Classification)

- Data was standardized and prepared, post that it was split for test and training.Different models were tested with parameter selection and accuracy were compared
- GitHub URL-[Predictive-Analysis]

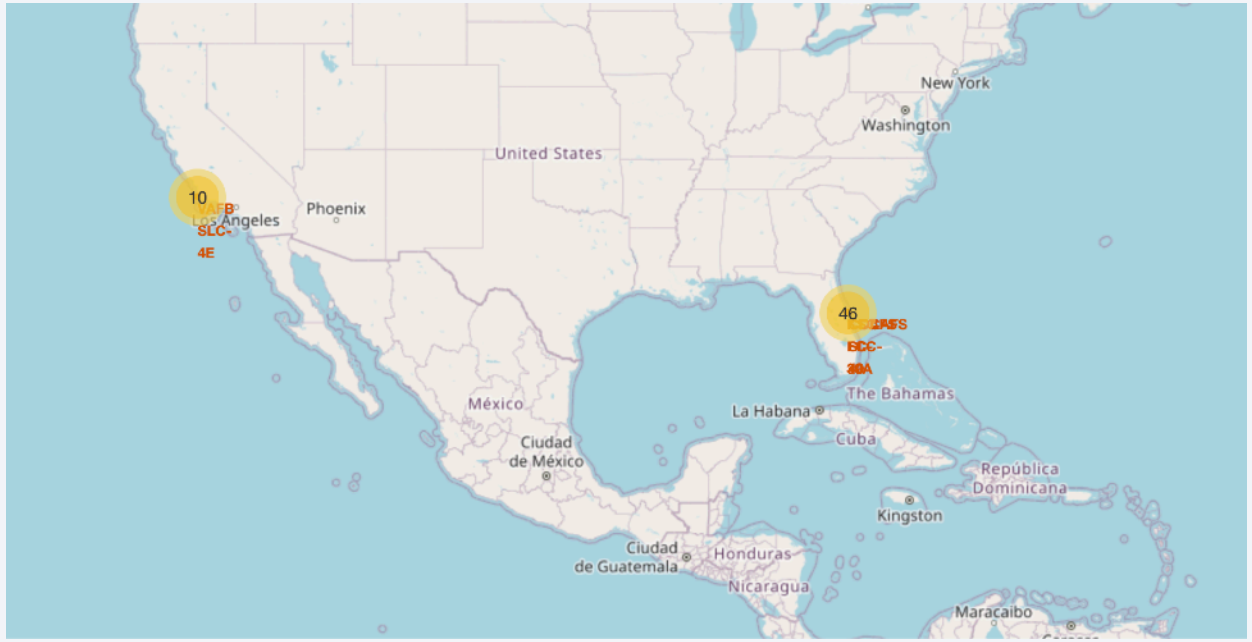


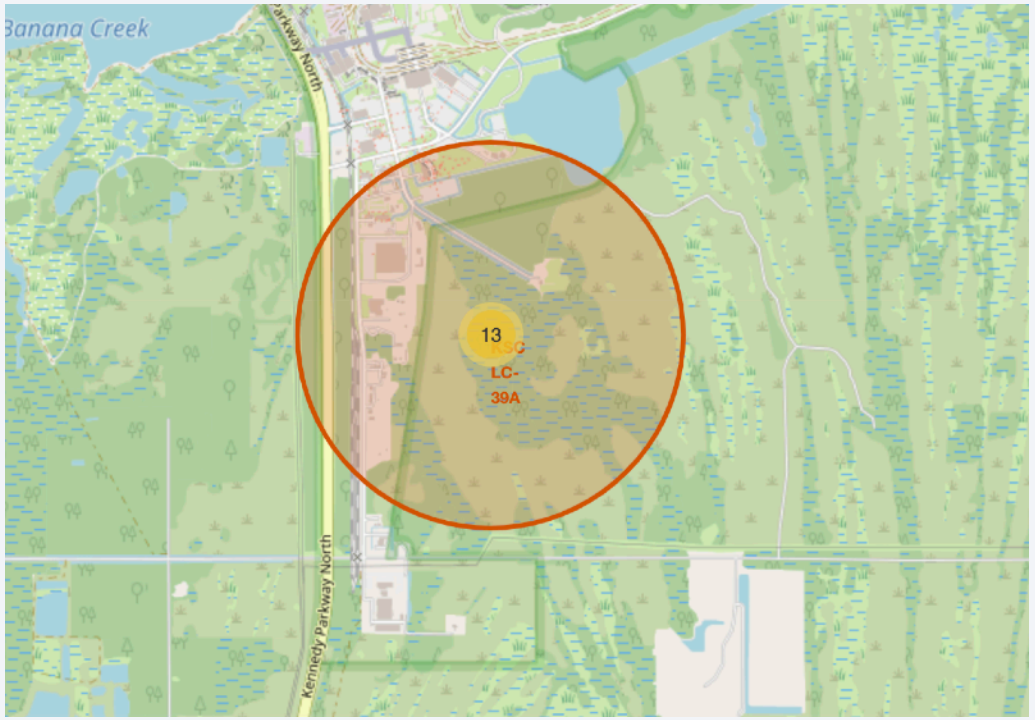
Results

- Exploratory data analysis results
  - Four different Launch sites are used
  - Average payload mass is 2928.4
  - First successful launch was in year 2015
  - Drone ship landing has a high success rate
  - Overall success rate has increased over the past decade.
- Interactive analytics demo in screenshots
- Predictive analysis results

Results

- Interactive analytics demo in screenshots
  - Majority of launches has occurred on the east coast
  - Launch sites are in close proximity to sea and are well connected for logistics support.





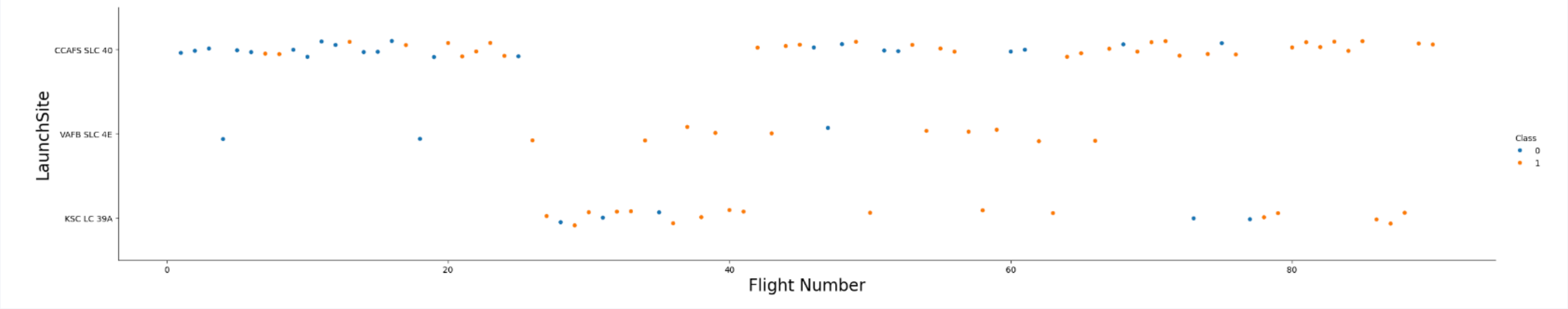
Results

- Predictive analysis results
  - Decision Tree model is the best suited and has a high accuracy of 88 % . Also the model has high accuracy of 94 % is test data

Section 2

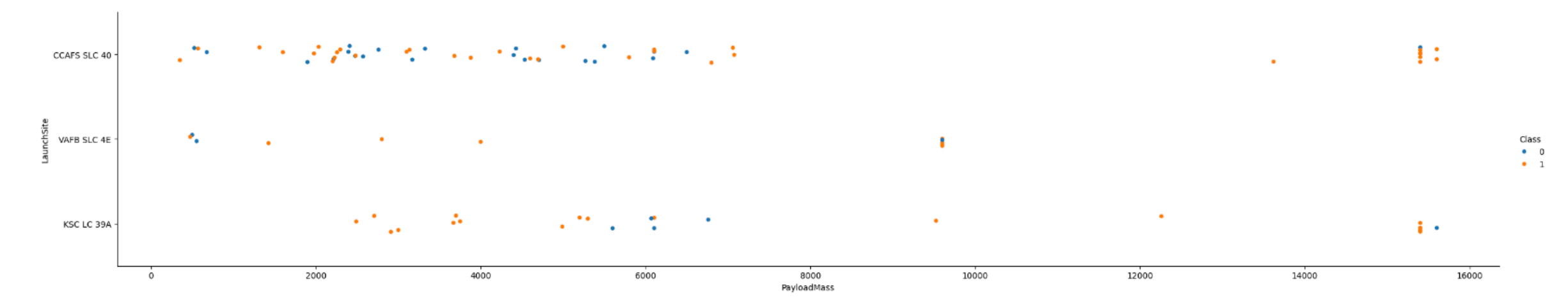
Insights drawn from EDA

Flight Number vs. Launch Site



- With increase in flight numbers the success rate of the launch sites improved
- Most of the launch has happened from CCAFS SLC40
- CCAFS SLC40 success rate has improved with the flight number

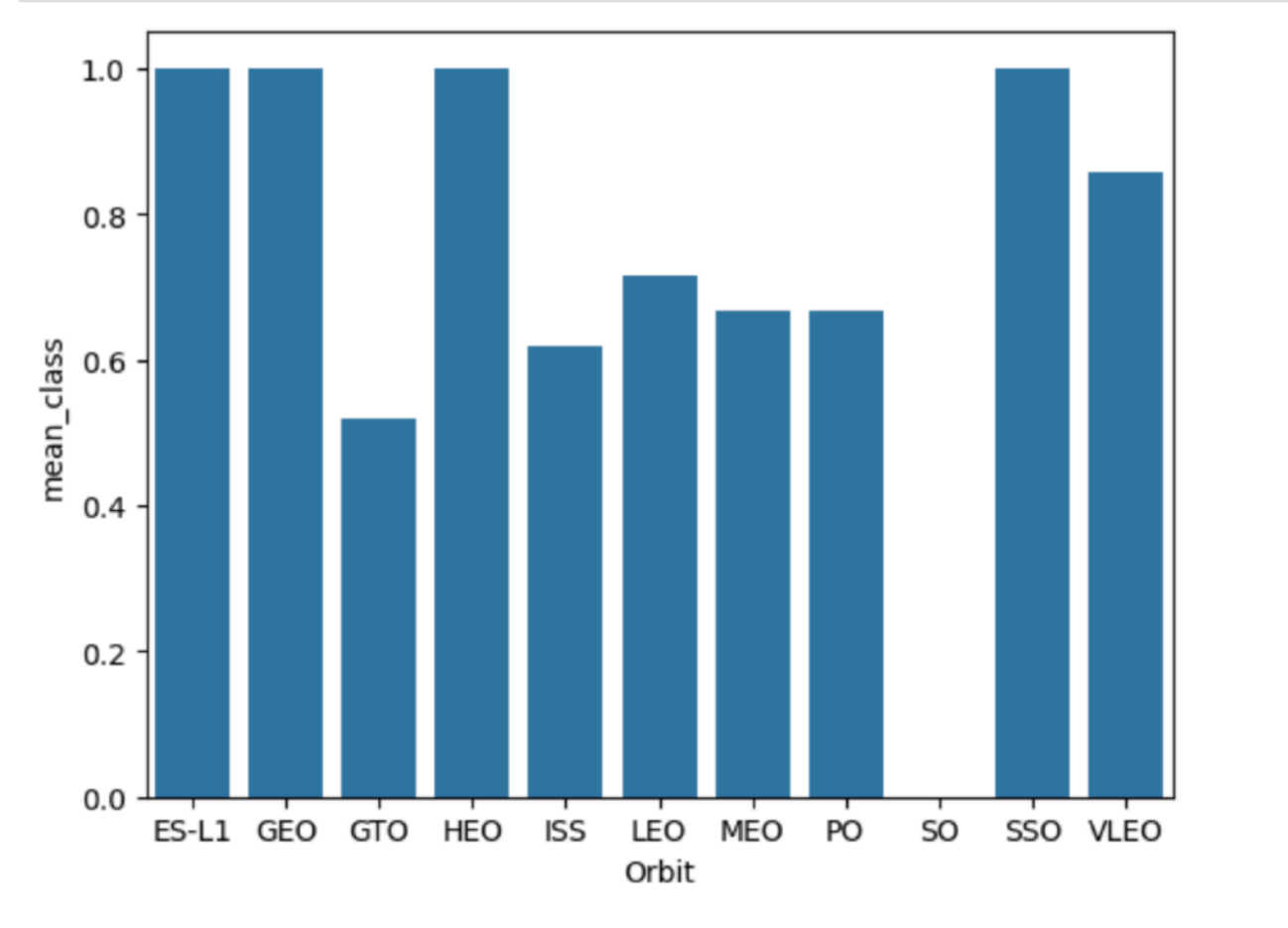
Payload vs. Launch Site



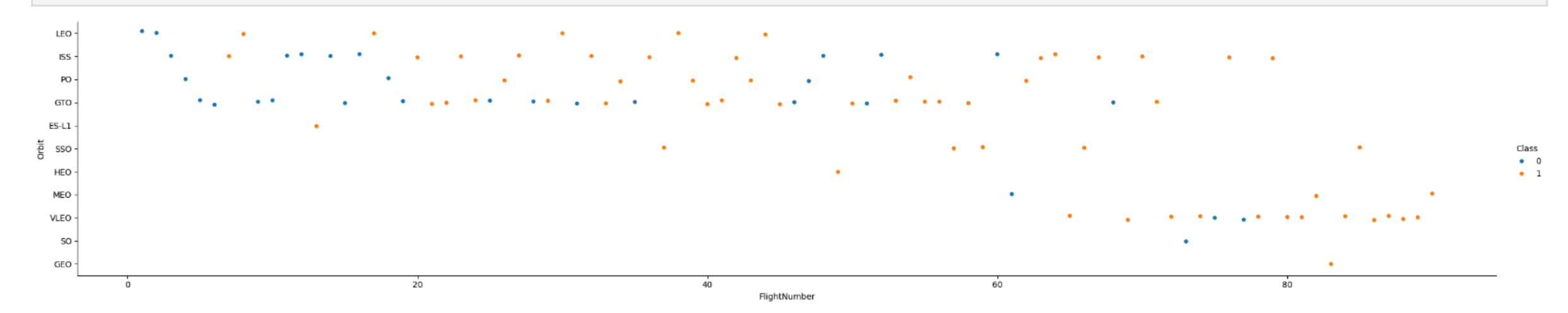
- High payload mass has a high success rate across the launch sites
- Success rates increase on payload over 9000
- VAFB SLC 4E does have launch with payload mass over 10000

Success Rate vs. Orbit Type

- ES-L1, GEO, HEO and SSO has high success rates.
- SO has zero success rate.
- All of the orbit except SO have at least 50%

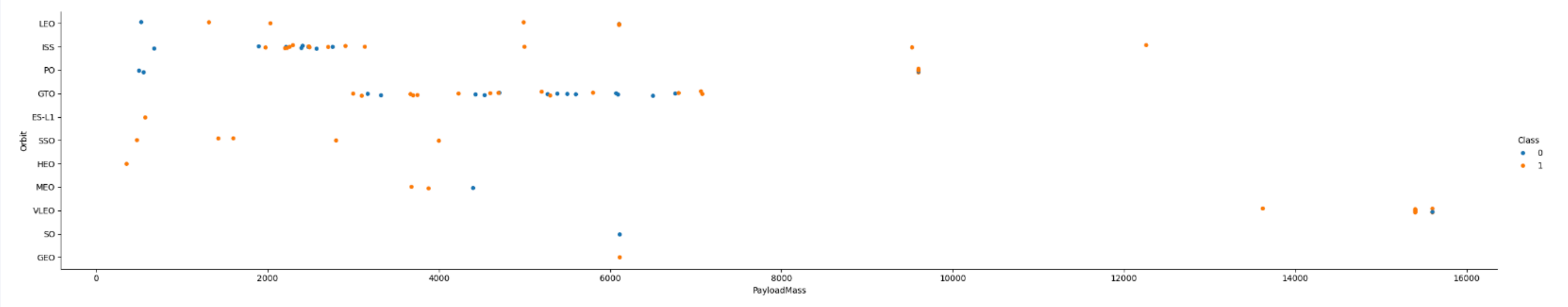


Flight Number vs. Orbit Type



- VLEO has picked up pace in recent past and has high success
- LEO success has increased over the flight number
- SSO, HEO,MEO,GEO does not contribute much to the flight numbers

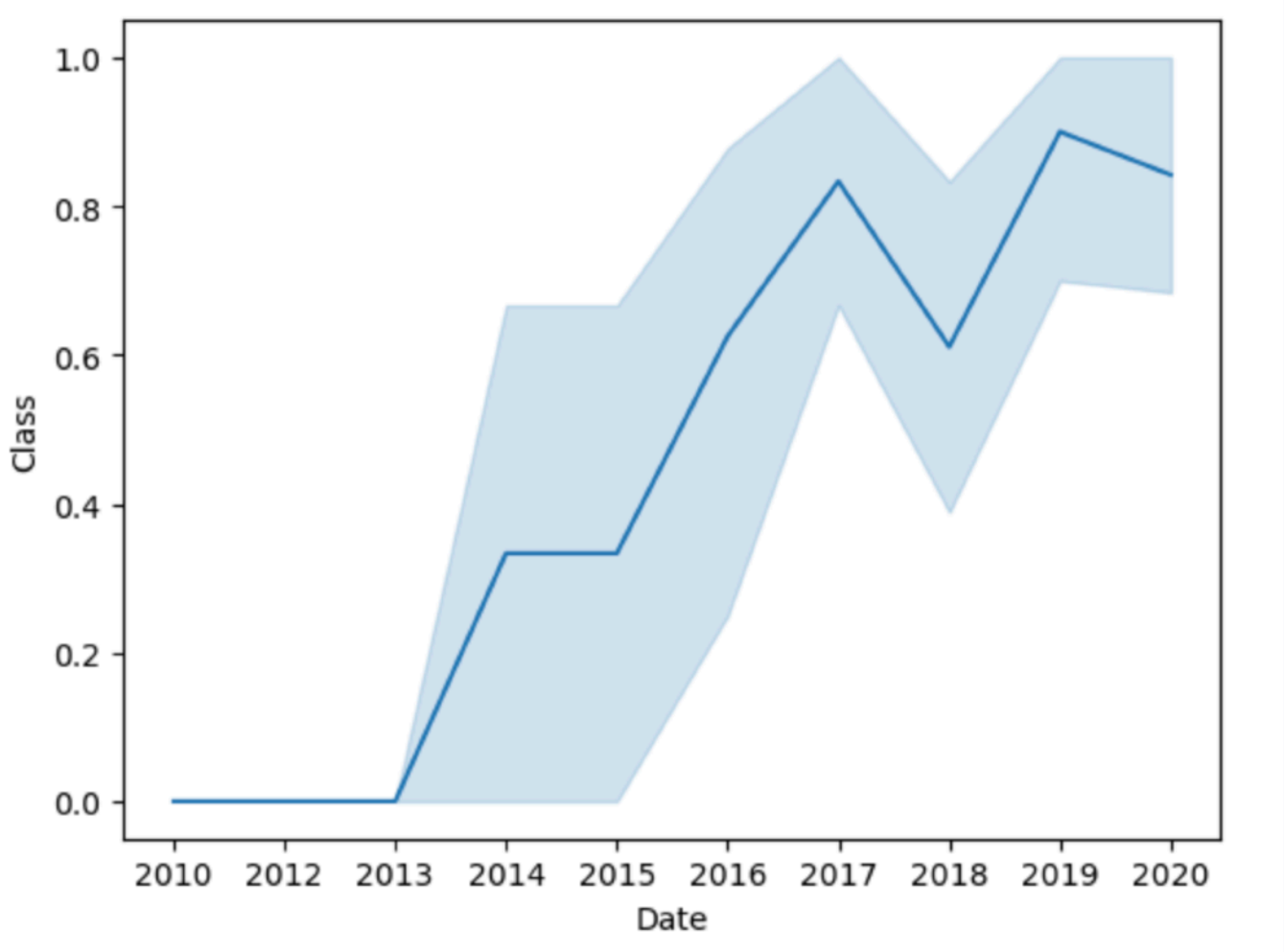
Payload vs. Orbit Type



- VLEO alone is used for very high payload mass with good success rate
- GTO has high mid range payload mass with considerable success and failures
- Except for VLEO and GTO other orbits focus on small payload more

Launch Success Yearly Trend

- Overall Success rate has increased starting from year 2013.
- Success rate remained same between 2014 and 2015
- Year on Year success rate dropped in 2018



All Launch Site Names

- Four different launch sites as listed in the query output

Task 1

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

```
%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'

- Queried records with launch site starting with “CCA” and restricted the number of records to 5

```
%sql Select * From SPACEXTABLE  Where (lower(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_Outcom
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachut
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 (parachut
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attem
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attem
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attem

Total Payload Mass

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

```
%sql Select SUM(All PAYLOAD_MASS__KG_) AS TotalPayloadMass From SPACEXTABLE  Where (Customer like 'NASA (CRS)')
```

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

Done.
```

TotalPayloadMass
45596

- Total mass is 45596 for boosters launched by NASA

Average Payload Mass by F9 v1.1

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

```
%sql Select Booster_Version,Avg(All PAYLOAD_MASS__KG_) AS AvgPayloadMass From SPACEXTABLE
    Where (Booster_Version like 'F9 v1.1')
```

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

Booster_Version	AvgPayloadMass
F9 v1.1	2928.4

- Average value of the booster version F9 v1.1 is queried.

First Successful Ground Landing Date

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

Hint:Use min function

```
%sql Select Min (Date) From SPACEXTABLE  Where (Landing_Outcome like 'Success (ground pad)')
```

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

Min (Date)
2015-12-22

- Dec-22nd of 2015 was the first successful launch

Successful Drone Ship Landing with Payload between 4000 and 6000

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

```
%sql Select Booster_Version From SPACEXTABLE
  Where (Landing_Outcome like 'Success (drone ship)') AND ( PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000)
```

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

Done.
```

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

- Four boosters have had success in drone ship landing that has a payload mass between 4000 and 6000

Total Number of Successful and Failure Mission Outcomes

List the total number of successful and failure mission outcomes

```
%sql SELECT MISSION_OUTCOME, COUNT(*) AS QTY FROM SPACEXTABLE GROUP BY MISSION_OUTCOME;
```

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

Done.
```

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

- Queried the success and failure outcomes

Boosters Carried Maximum Payload

- Booster versions with maximum payload mass listed

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

```
%sql Select booster_version,PAYLOAD_MASS__KG_ From SPACEXTABLE Where PAYLOAD_MASS__KG_ = (Select Max(PAYLOAD_MASS__KG_)
```

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

Done.
```

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

List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.

Note: SQLLite does not support monthnames. So you need to use substr(Date, 6,2) as month to get the months and substr(Date,0,5)='2015' for year.

```
%sql Select substr(Date, 6,2) AS Month,substr(Date,0,5) As Year, booster_version,Launch_site from SPACEXTABLE Where (Lan
```

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

Month	Year	Booster_Version	Launch_Site
01	2015	F9 v1.1 B1012	CCAFS LC-40
04	2015	F9 v1.1 B1015	CCAFS LC-40

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

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

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 Landing_Outcome, COUNT(*) AS QTY FROM SPACEXTABLE WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20'
       GROUP BY Landing_Outcome ORDER BY QTY DESC;
```

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

]:

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

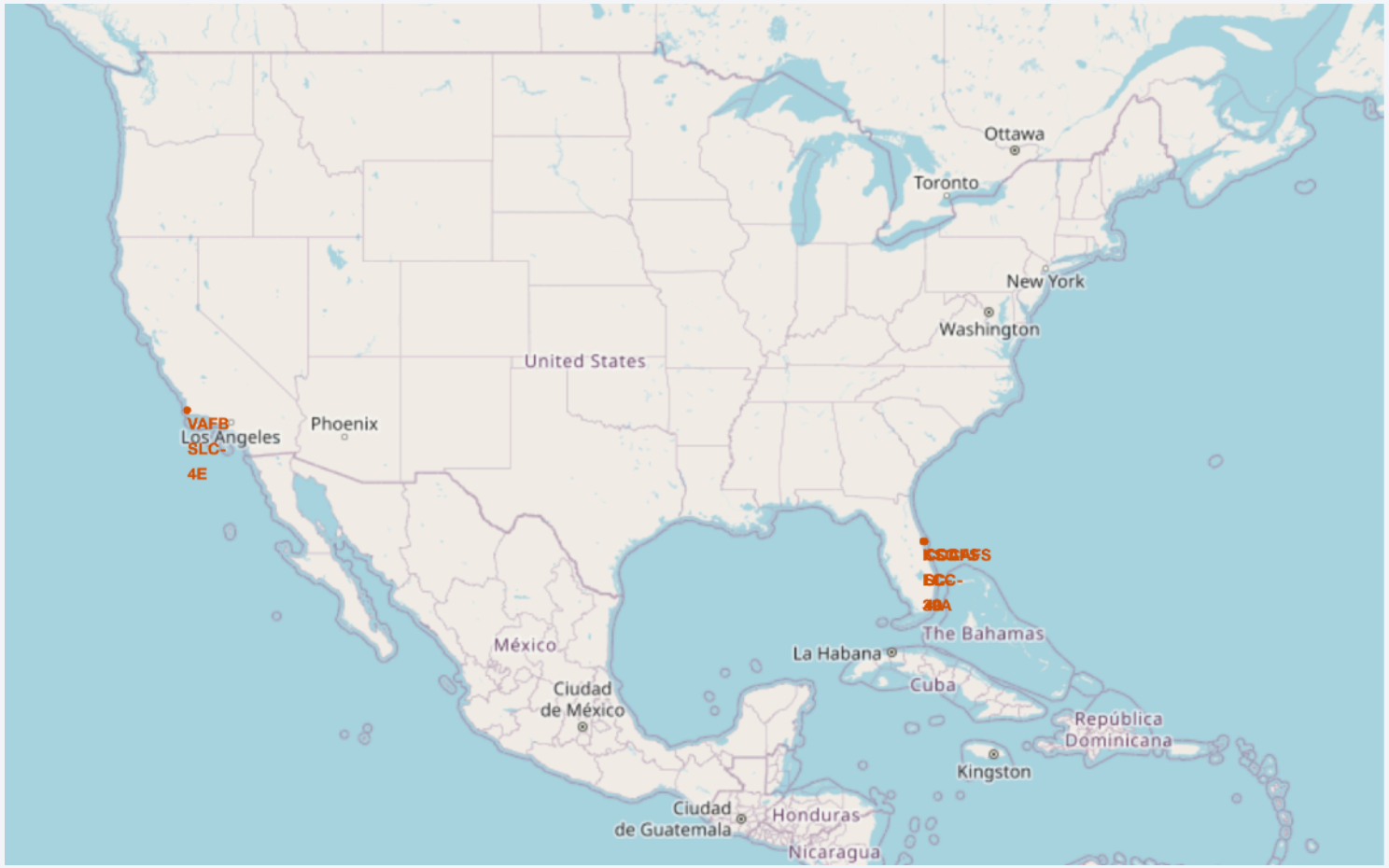
- Ranked the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

Section 3

Launch Sites Proximities Analysis

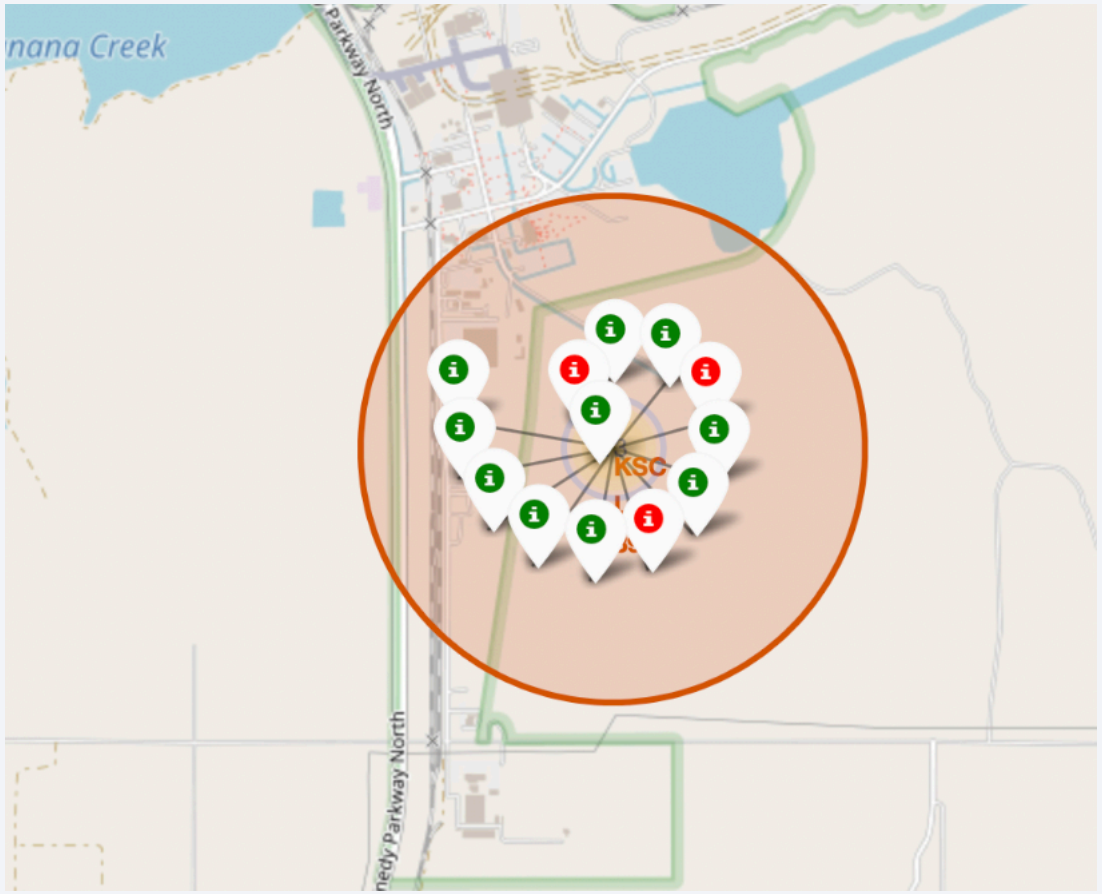
All Launch Sites Marked in Map

- Launch site are situated on either coast of America
- All of the launch sites are located in the souther half of the country
- Sites are in close proximity to sea and
- Sites are located close to the equator



Launch Record Map

- Launch sites indicated with success and failure launches.
  - Red -Failure
  - Green- Success



Launch Site Proximity to Landmarks

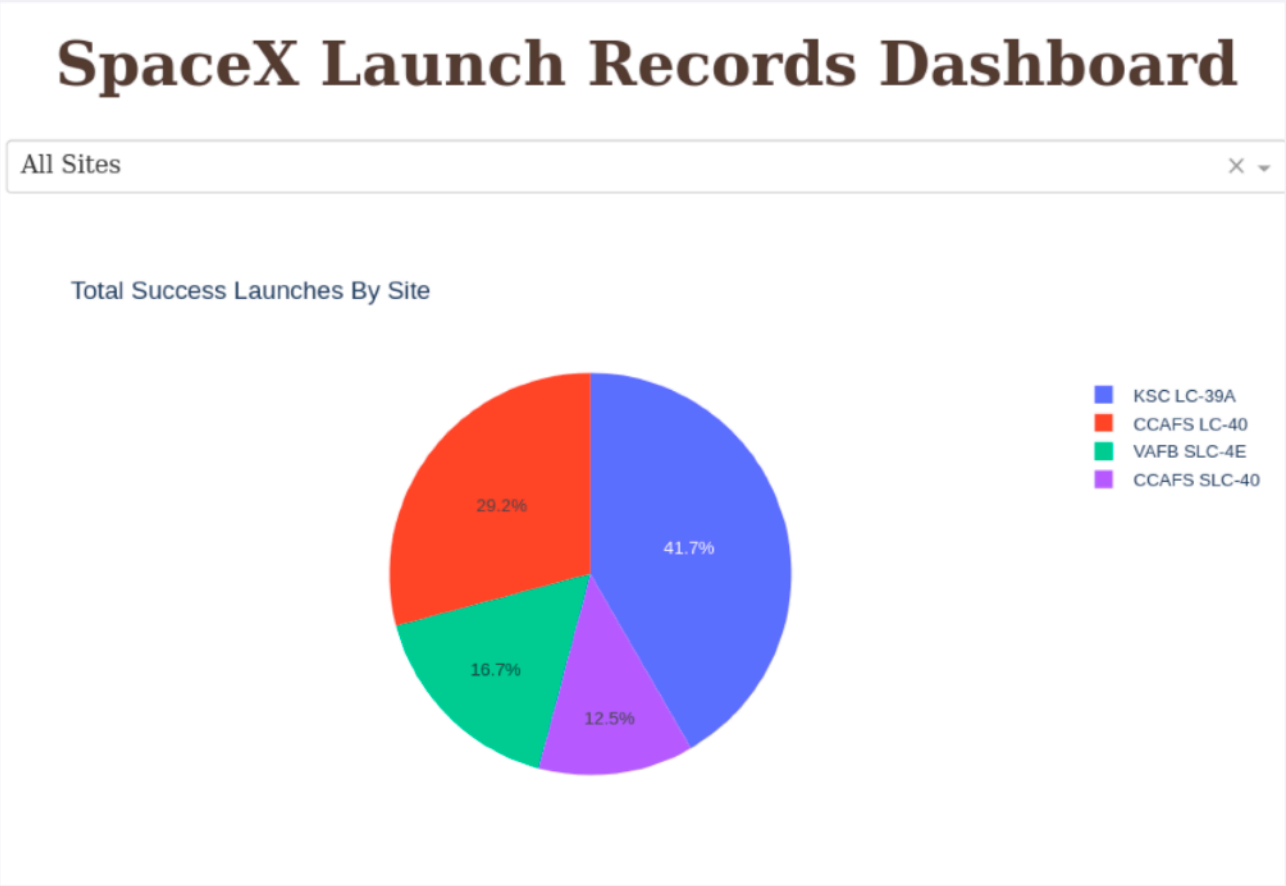
- Launch sites are located in close proximity to railway lines and road ways making it possible for good connectivity for supporting the launch
- The site are also located in close proximity to cities, creating a potential issue of rocket debris falling over populated areas

Section 4

Build a Dashboard with Plotly Dash

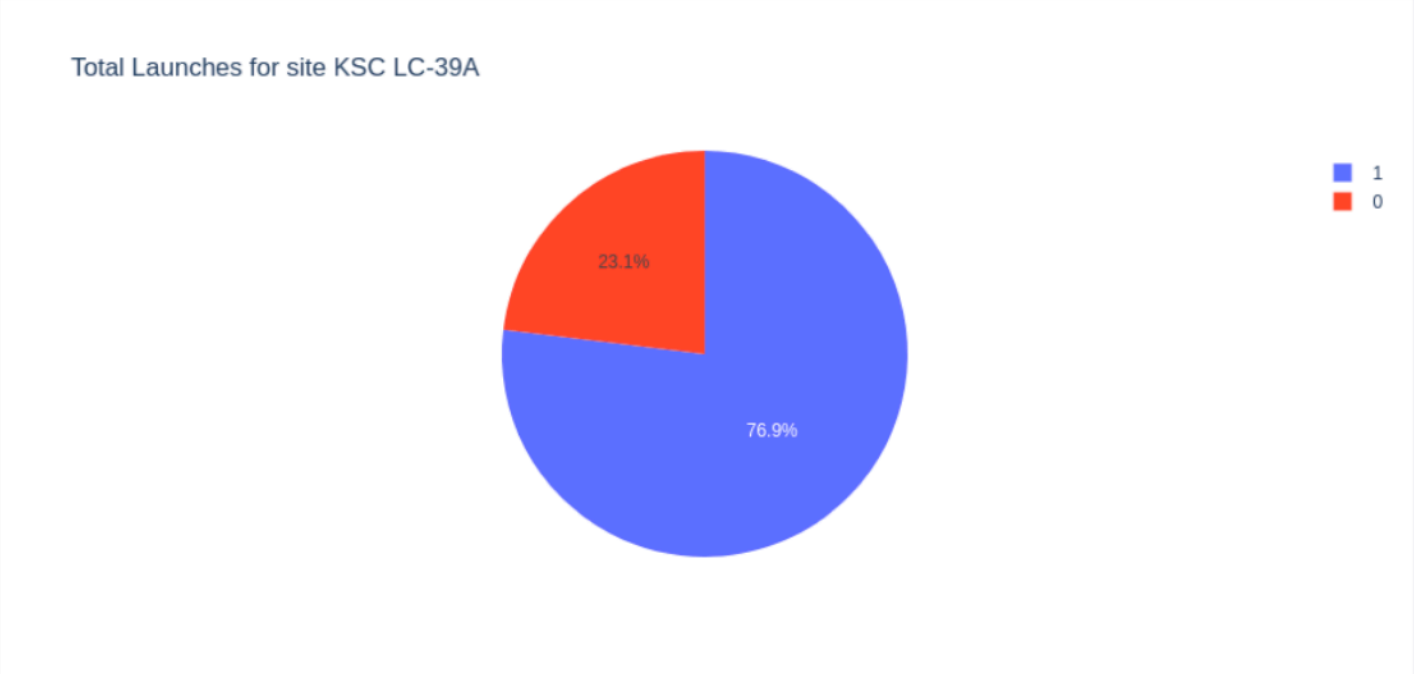
Success Rate by Launch site

- Pie Chart displaying the success rate of each launch site
- KSC LC-39A has the highest success rate

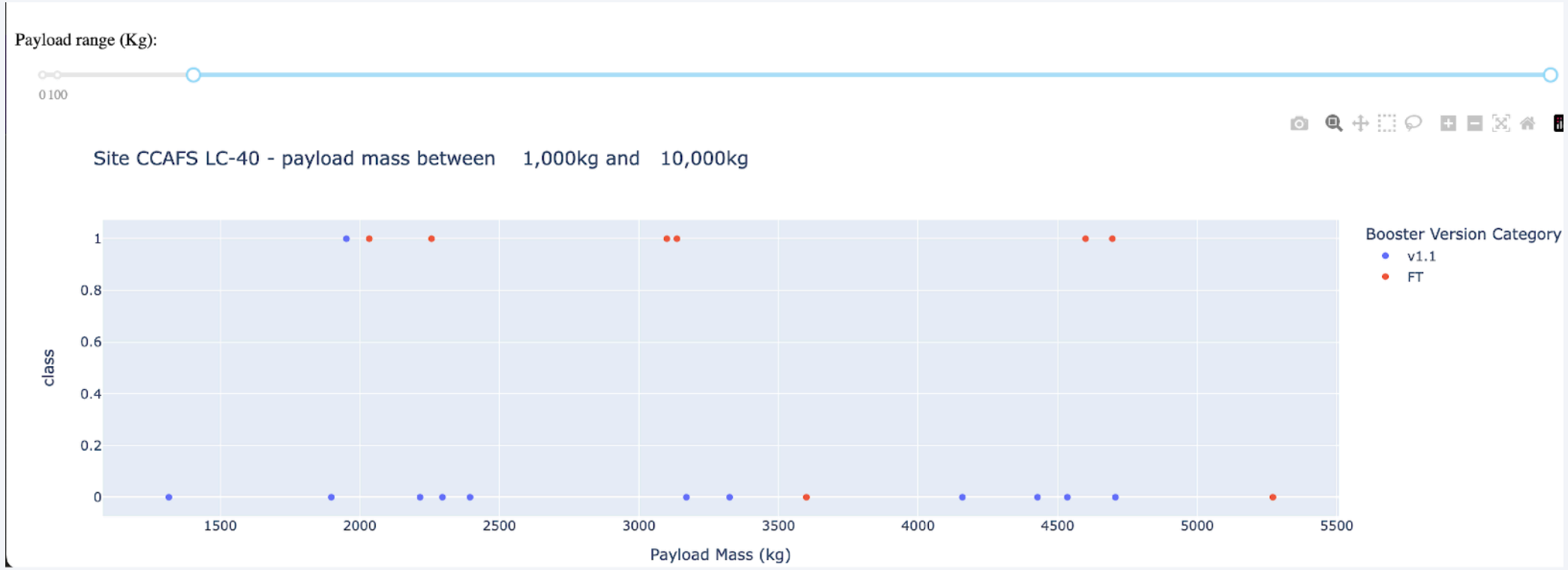


Success Rate Of site with Highest launch success

- Pie chart displaying the success rate of the site with the highest success rate
- Site has the success rate
- 76% of the launch are success



Payload vs Launch Outcome



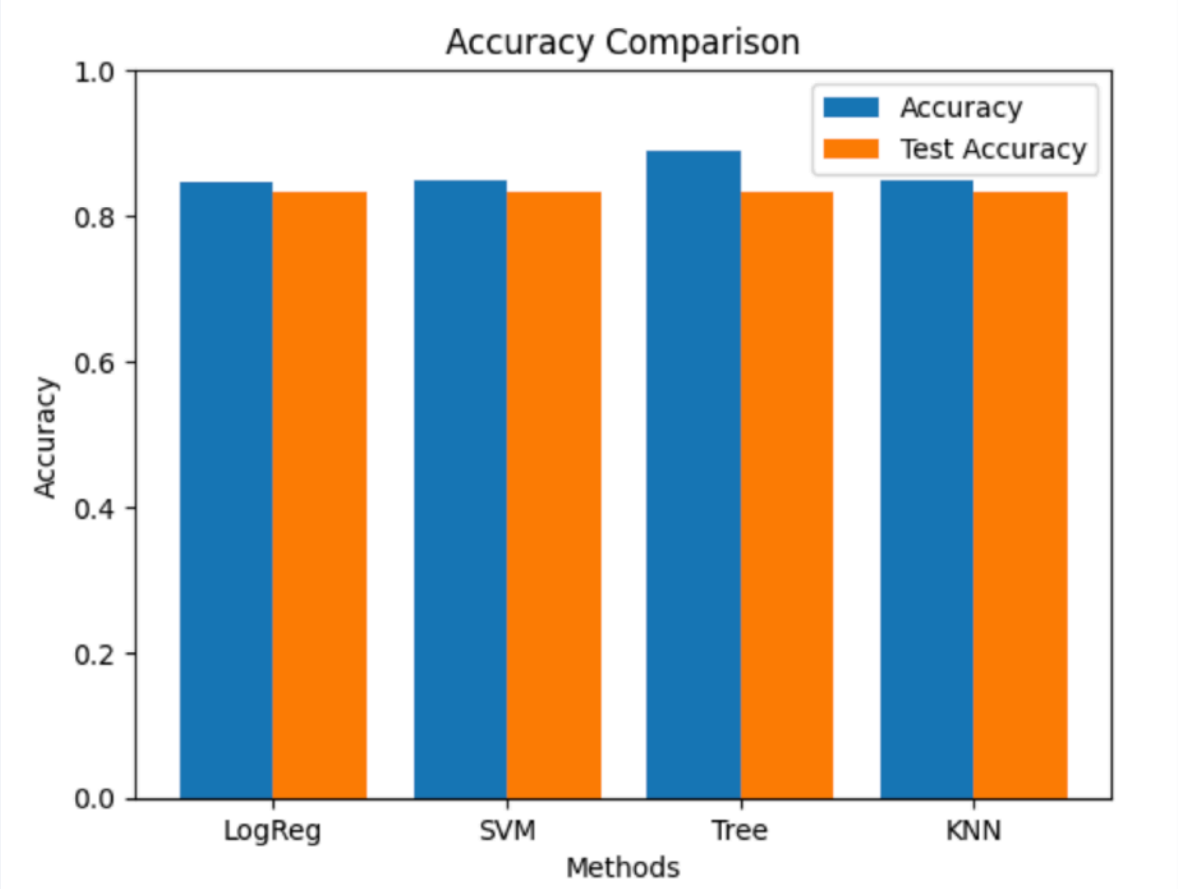
- Payloads under 5500 have high success rate

Section 5

Predictive Analysis (Classification)

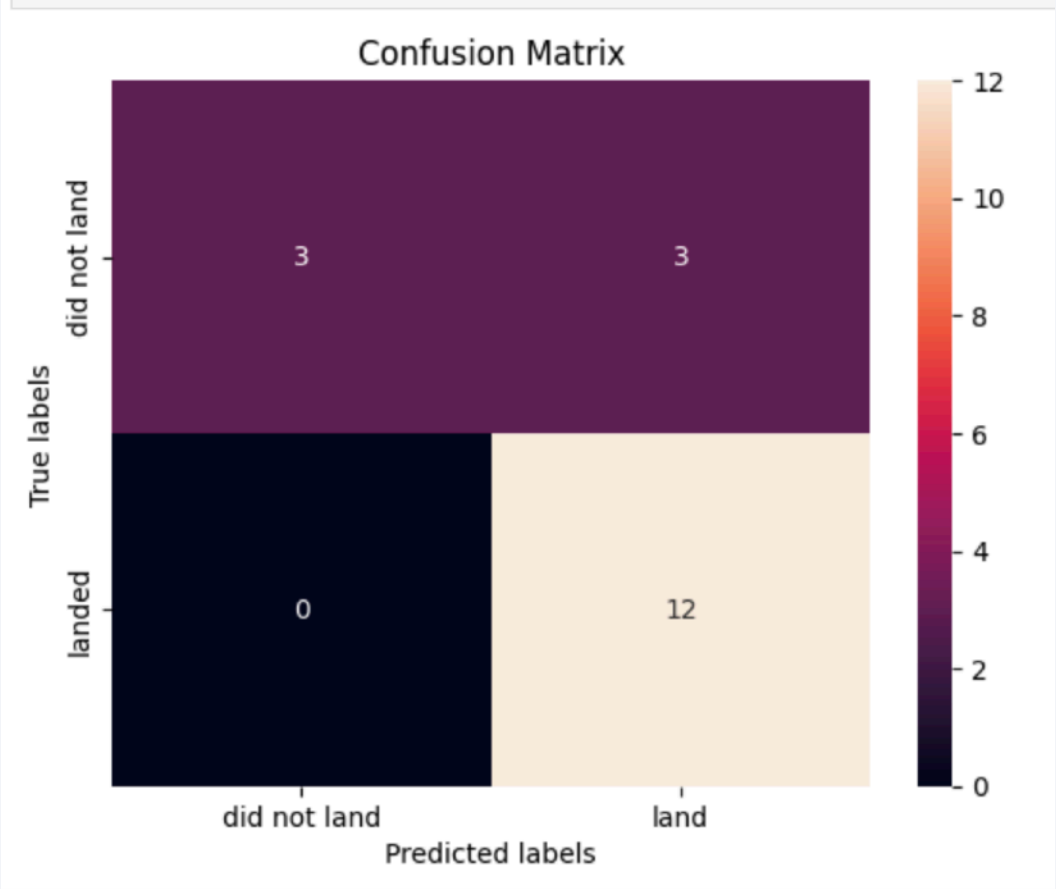
Classification Accuracy

- Decision Tree Model has the highest accuracy



Confusion Matrix of Decision Tree

- The True positive label has a higher number



Conclusions

- Success rate has increased with flight numbers in certain launch sites
- Payload over 9000 have relatively high success rate.
- SO orbit has zero success rate
- LEO orbit success rate has increased with flight numbers
- Overall success rate has increased over the years starting form 2013
- All the launch sites are located in the souther half of the country in close proximity to sea
- KSC LC-39A site has the highest success rate
- Decision Tree is able predict the success rate accurately compared to other models

Appendix

- Codes for each section can be accessed via the link provided in each slide.
- Folium map does not load in files uploaded to Github

Thank you!