Winning Space Race with Data Science

OCHILBEK RAKHMANOV
12-March-2024

Outline

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

Executive Summary

- In this study, we examined the successful docking rate of Space X's Falcon 9 flights. The data was formed using web scraping and data wrangling was conducted in order to get clean and useful dataset. Various statistical methods were used during the study and important findings were summarized with the help of graphics. A classification algorithm was developed as the result of the study and a training accuracy as high as 84% was achieved.

Introduction

- Sending airships to space used to be done only by states, because it is difficult and expensive. But, in the modern developing world with technology we have, this job can now be done by the private sector as well. Space X is one of the leading companies in this field. But this business is a very expensive, and companies that want to be a shareholder in this business must manage to reuse their airships. Space X has conducted many experiments on this subject and their experience can guide us. In this study, our aim is to explore the requirements for achieving a successful docking operation for Space X airships launched into space, with a particular focus on enabling the reuse of the successful Falcon 9 types.

Section 1

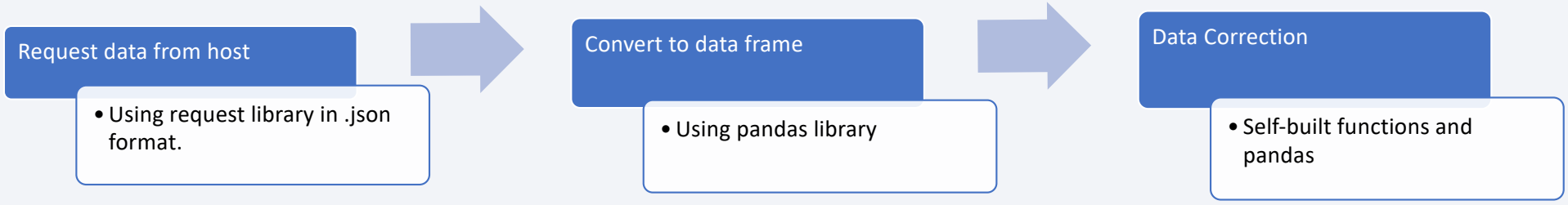
Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection



- Firstly the table was extracted from the host using request library. Using pandas we can convert the requested table into data frame using pandas library. But most of the data is not clean for ordinary user, thus, some delf-built functions should be used to extract necessary part for columns like Orbit, Payload Mass and others.

Data Collection – SpaceX API

Request data from SpaceX API

- Using request library

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

Convert to data frame

- Using pandas library

```
# Use json_normalize method to convert the json
response = requests.get(static_json_url).json()
data = pd.json_normalize(response)
```

Data Correction

- Self-built functions and pandas

```
getLaunchSite(data)
getPayloadData(data)
getCoreData(data)
```

- GitHub URL of the completed SpaceX API calls notebook: https://github.com/lachinbro/SpaceX-Falcon-9-Success-Analysis/blob/main/IBM%20SpaceX/week_1_spacex-data-collection-api.ipynb

Data Collection – Scraping

Request data from Wikipedia

- Using request library

```
# use requests.get() method with
# assign the response to a object
page = requests.get(static_url)
page.status_code
```

Parse the data using Beautiful Soap and extract the table from it.

- Using beautiful soap library

```
# Use BeautifulSoup() to create a BeautifulSoup
soup = BeautifulSoup(page.text, 'html.parser')
```

Data frame it and do correction

- Self-built functions and pandas

- GitHub URL of the completed SpaceX API calls notebook: https://github.com/lachinbro/SpaceX-Falcon-9-Success-Analysis/blob/main/IBM%20SpaceX/week_1_spacex-webscraping.ipynb

Data Wrangling

Check the launching sites

- Calculate how many successful attempt each has

Convert landing info into binary

- Fail/Success to 0/1

Save data for further use

- Dataset 2

```
# Apply value_counts() on column LaunchSite
df["LaunchSite"].value_counts()
```

```
# landing_class = 0 if bad_outcome
# landing_class = 1 otherwise
landing_class = []
for key,value in df["Outcome"].items():
     if value in bad_outcomes:
        landing_class.append(0)
     else:
        landing_class.append(1)
```

Serial	Longitude	Latitude	Class
B0003	-80.577366	28.561857	0
B0005	-80.577366	28.561857	0
B0007	-80.577366	28.561857	0

- GitHub URL to data wrangling related notebooks: https://github.com/lachinbro/SpaceX-Falcon-9-Success-Analysis/blob/main/IBM%20SpaceX/week_1_data%20wrangling.ipynb

EDA with Data Visualization

- Seaborn library was used to plot following graphics:

- ***Category plotting*** (with Class hue): Payload Mass vs Flight Number, Launch Site vs Flight Number, Payload Mass vs Launch Site

```
sns.catplot(y="PayloadMass", x="FlightNumber", hue="Class", data=df, aspect = 5)
plt.xlabel("Flight Number",fontsize=20)
plt.ylabel("Pay load Mass (kg)",fontsize=20)
plt.show()
```

- ***Bar plot***: Mean success for each Orbit

```
sns.catplot(x="Orbit",y="Class", kind="bar",data=df)
plt.xlabel("Orbit",fontsize=20)
plt.ylabel("Mean",fontsize=20)
plt.show()
```

- ***Line plot***: Success rate vs Years

- ***Scatter plot***: Payload mass vs Orbit (with Class hue)

```
sns.scatterplot(x="Orbit",y="FlightNumber",hue="Class",data = df)
plt.xlabel("Orbit",fontsize=20)
plt.ylabel("Flight Number",fontsize=20)
plt.show()
```

- GitHub URL of EDA with data visualization notebook: https://github.com/lachinbro/SpaceX-Falcon-9-Success-Analysis/blob/main/IBM%20SpaceX/week_2_EDA-viz.ipynb

EDA with SQL

- **Display the names of the unique launch sites in the space mission:**

[%sql SELECT DISTINCT "Launch_Site" FROM SPACEXTABLE **]**

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

[%sql SELECT * FROM SPACEXTABLE WHERE "Launch_Site" LIKE 'CCA%' LIMIT 5**]**

- **Display the total payload mass carried by boosters launched by NASA:**

[%sql SELECT SUM("PAYLOAD_MASS__KG_") FROM SPACEXTABLE WHERE "Customer" LIKE 'NASA%' **]**

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

[%sql SELECT AVG("PAYLOAD_MASS__KG_") FROM SPACEXTABLE WHERE "Booster_Version" LIKE 'F9 v1.1%'; **]**

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

[%sql SELECT "Date" FROM SPACEXTABLE WHERE "Landing_Outcome"=="Success (ground pad)" LIMIT 1 **]**

- **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"=="Success (drone ship)" AND "PAYLOAD_MASS__KG_" >4000 AND "PAYLOAD_MASS__KG_" <6000]

- **List the total number of successful and failure mission outcomes**

[%sql SELECT COUNT(*) FROM SPACEXTABLE WHERE "Mission_Outcome"=="Success" **]**

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

[%sql SELECT "Booster_Version" FROM SPACEXTABLE WHERE "PAYLOAD_MASS__KG_" == (SELECT MAX("PAYLOAD_MASS__KG_") FROM SPACEXTABLE); **]**

- **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 COUNT(DISTINCT "Landing_Outcome") AS count FROM SPACEXTABLE ORDER BY count DESC; **]**

GitHub URL of EDA with SQL notebook:
https://github.com/lachinbro/SpaceX-Falcon-9-Success-Analysis/blob/main/IBM%20SpaceX/week_2_eda-sql.ipynb

Build an Interactive Map with Folium

- We used folium library for geo location maps. We used Latitude and Longitude information from Space X dataset to location the launch location.
- We used markers and circles to identify the locations. Later we marked successful experiments with green and failed ones with Red.
- We added the total number of launches to each site.
- We also did calculation of the distance of launch sites to nearest urban places.

- GitHub URL of interactive map with Folium map: https://github.com/lachinbro/SpaceX-Falcon-9-Success-Analysis/blob/main/IBM%20SpaceX/week_3_Folium-locations.ipynb

Build a Dashboard with Plotly Dash

- Firstly, we created drop-down menus for launch sites, to show user what sites Space X used.
- Next, we created a visualized pie chart for successful attempts, to show user how successful were the trials
- Next, we added range slider, to give information of payload masses carried during attempts
- Lastly, we created a scatter plot to show how successful were different payload types.

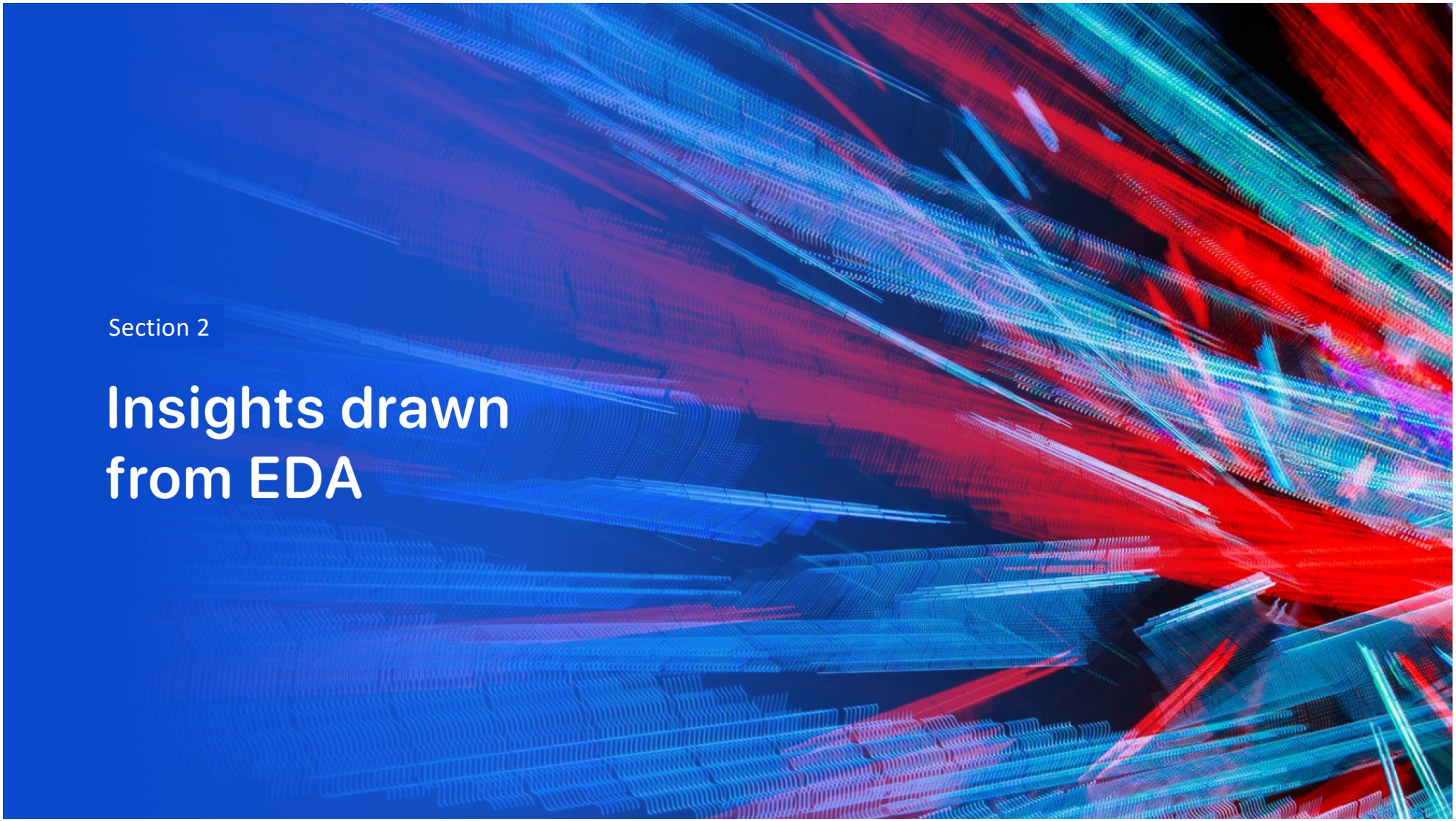
- GitHub URL of Plotly Dash lab: https://github.com/lachinbro/SpaceX-Falcon-9-Success-Analysis/blob/main/IBM%20SpaceX/week_3_Dashboard.ipynb

Predictive Analysis (Classification)

- The dataset was scaled to mean zero and split in the ratio of 80/20.
- Different parameters were tested using grid search to identify best version of Logistic regression, Support Vector Machines, Decision Tree and kNN.
- The best performers were tested with testing dataset.
- GitHub URL of predictive analysis lab: https://github.com/lachinbro/SpaceX-Falcon-9-Success-Analysis/blob/main/IBM%20SpaceX/week_4_SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

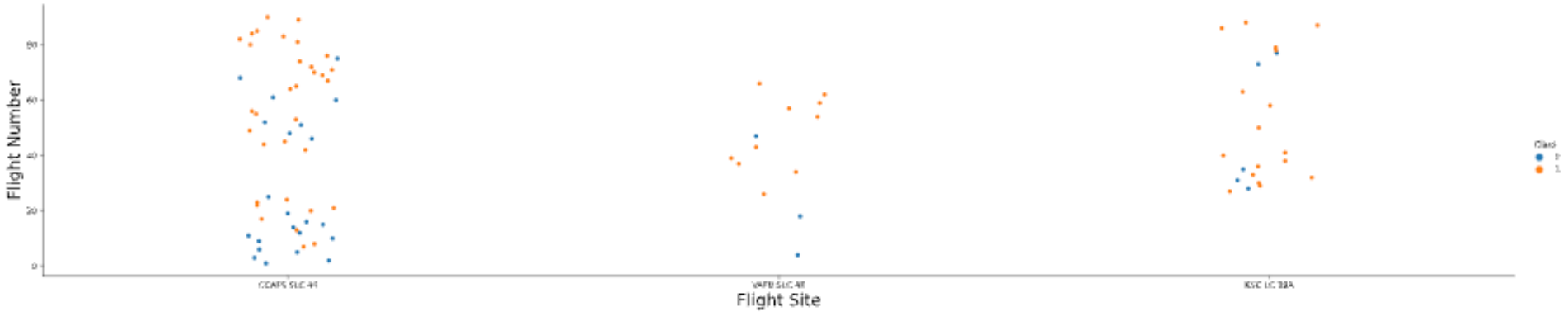
Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results


Section 2
Insights drawn
from EDA

Flight Number vs. Launch Site

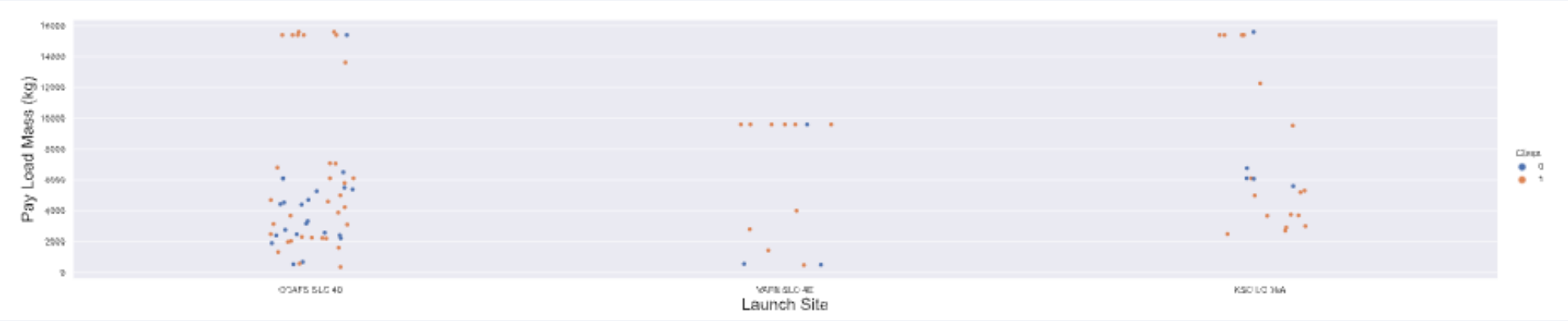
- Show a scatter plot of Flight Number vs. Launch Site



- The graphs shows that later trials were more successful and most of initial trials failed on CCAFS LC-40 site.

Payload vs. Launch Site

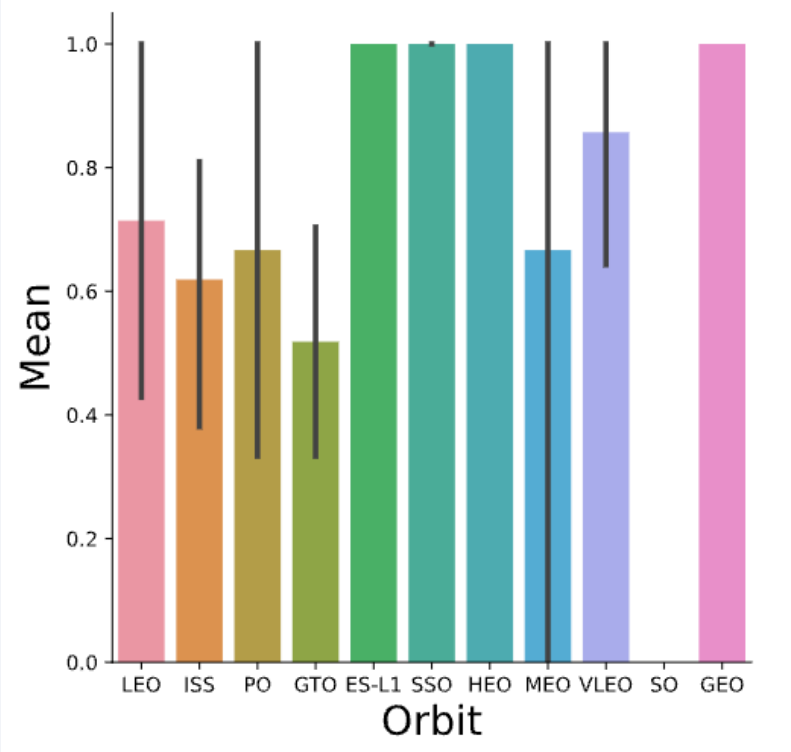
- Show a scatter plot of Payload vs. Launch Site



- The graph show that most of the higher loads were more successful

Success Rate vs. Orbit Type

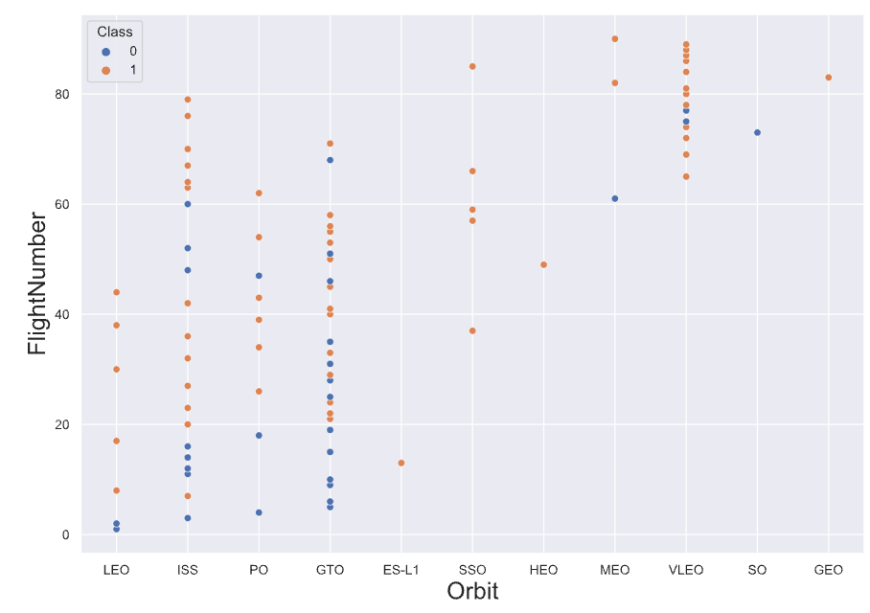
- Show a bar chart for the success rate of each orbit type



- The graph shows that 4 of the orbit types has %100 success rate.

Flight Number vs. Orbit Type

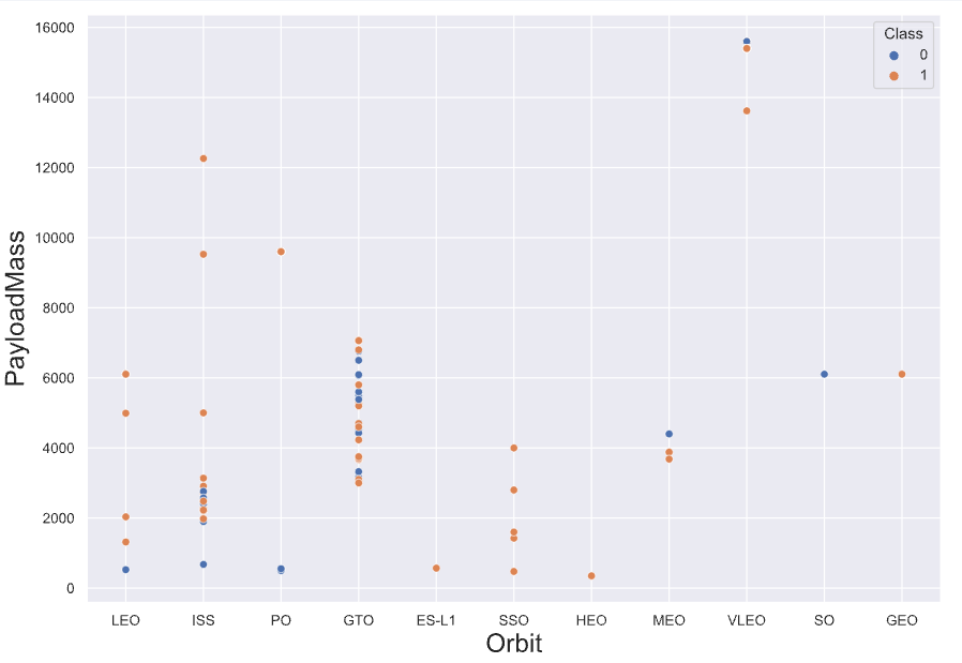
- Show a scatter point of Flight number vs. Orbit type



- Initial flight were likely to be unsuccessful, while later ones are more successful

Payload vs. Orbit Type

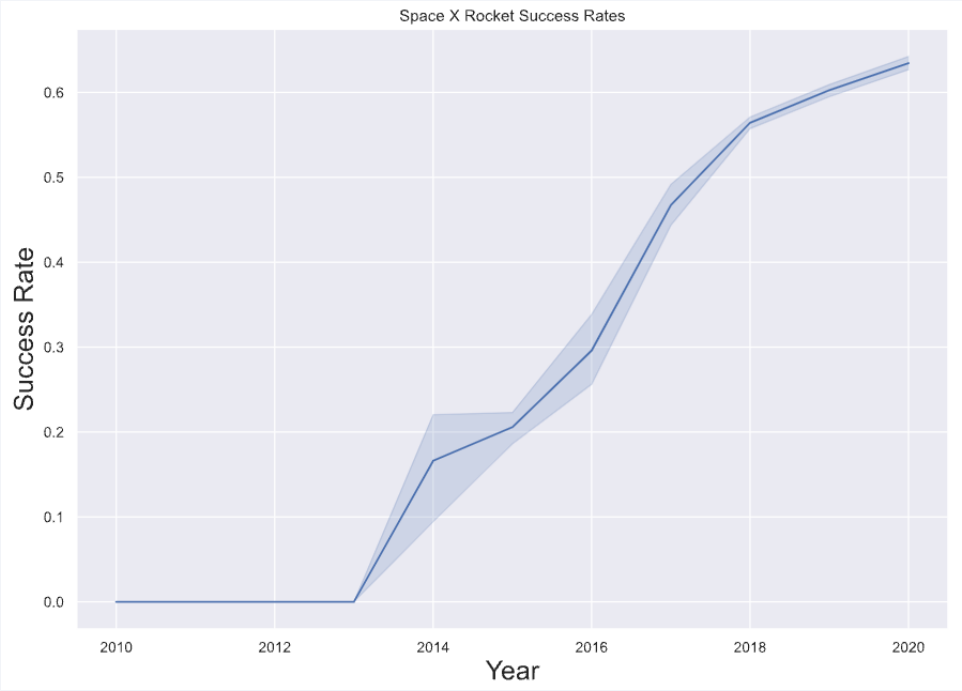
- Show a scatter point of payload vs. orbit type



- SSO flights were successful regardless the payload mass

Launch Success Yearly Trend

- Show a line chart of yearly average success rate



- Success rate increase with time

All Launch Site Names

- Find the names of the unique launch sites. Present your query result with a short explanation here

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'

- Find 5 records where launch sites begin with `CCA`. Present your query result with a short explanation here

Display 5 records where launch sites begin with the string '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
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX
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
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)

Total Payload Mass

- Calculate the total payload carried by boosters from NASA. Present your query result with a short explanation here

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

```
%sql SELECT SUM("PAYLOAD_MASS__KG_") FROM SPACEXTABLE WHERE "Customer" LIKE 'NASA%'
```

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

SUM("PAYLOAD_MASS__KG_")
99980

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1.
  Present your query result with a short explanation here

Display average payload mass carried by booster version 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

- Find the dates of the first successful landing outcome on ground pad.
  Present your query result with a short explanation here

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

Hint:Use min function

```
%sql SELECT "Date" FROM SPACEXTABLE WHERE "Landing_Outcome"=="Success (ground pad)" LIMIT 1
```

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

Date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000. Present your query result with a short explanation here

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"=="Success (drone ship)" AND "PAYLOAD_MAS
```

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

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

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes. Present your query result with a short explanation here

List the total number of successful and failure mission outcomes

```
%sql SELECT COUNT(*) FROM SPACEXTABLE WHERE "Mission_Outcome"=="Success"
```

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

COUNT(*)
98

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass. Present your query result with a short explanation here

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

```
%sql SELECT "Booster_Version" FROM SPACEXTABLE WHERE "PAYLOAD_MASS__KG_" == (SELECT MAX("PAYLOAD_MASS__KG_") FR
```

```
 * 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

- List the failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015. Present your query result with a short explanation here

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 COUNT(DISTINCT "Landing_Outcome") AS count  FROM SPACEXTABLE ORDER BY count DESC;
```

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

Landing_Outcome	count
Failure (parachute)	11

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. Present your query result with a short explanation here

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 COUNT(DISTINCT "Landing_Outcome") AS count  FROM SPACEXTABLE ORDER BY count DESC;
```

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

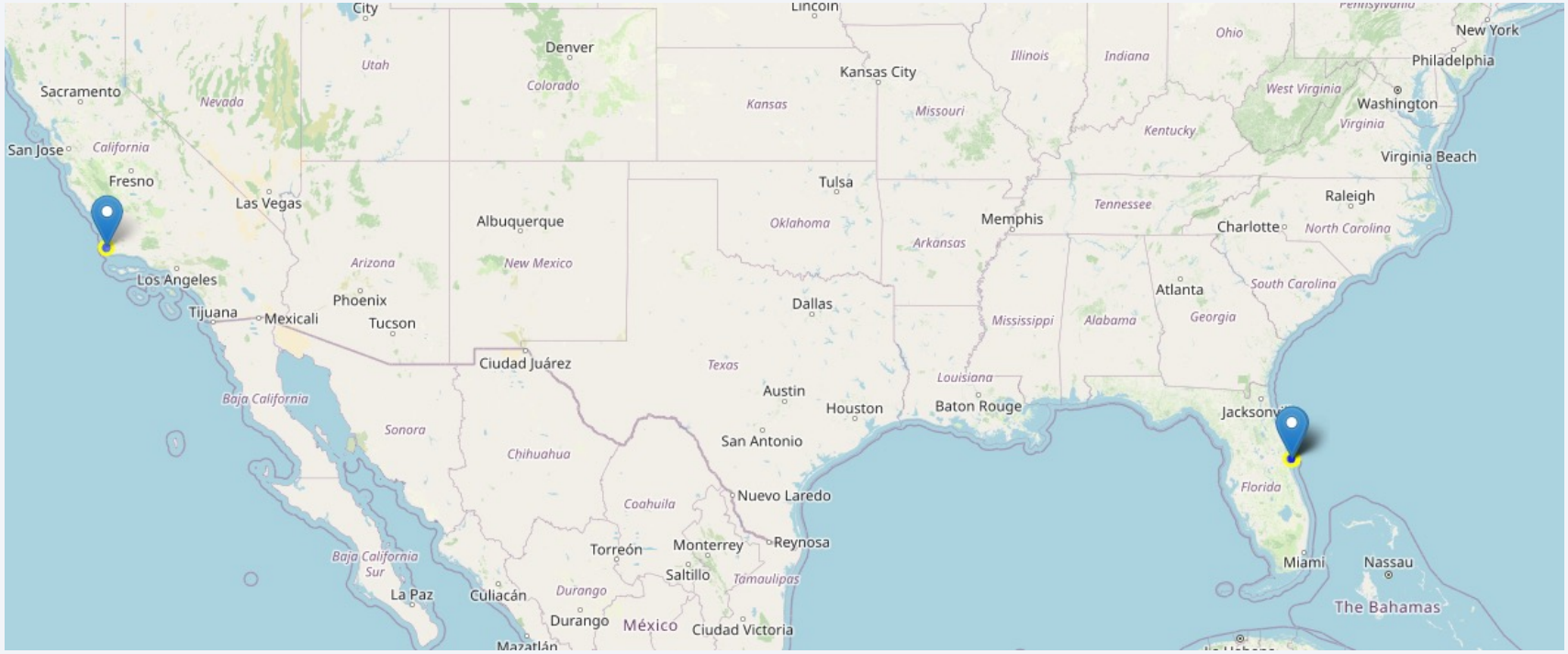
Landing_Outcome	count
Failure (parachute)	11

Section 3

Launch Sites Proximities Analysis

Launch sites on map

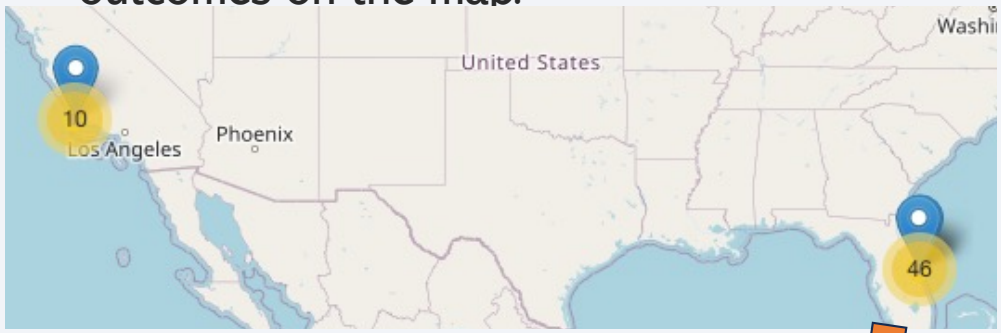
Explore the generated folium map and make a proper screenshot to include all launch sites' location markers on a global map.

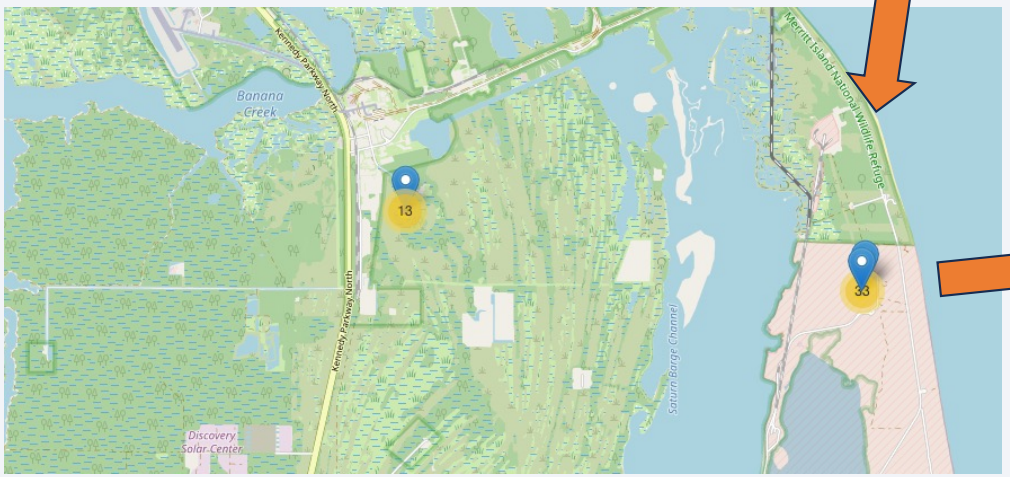


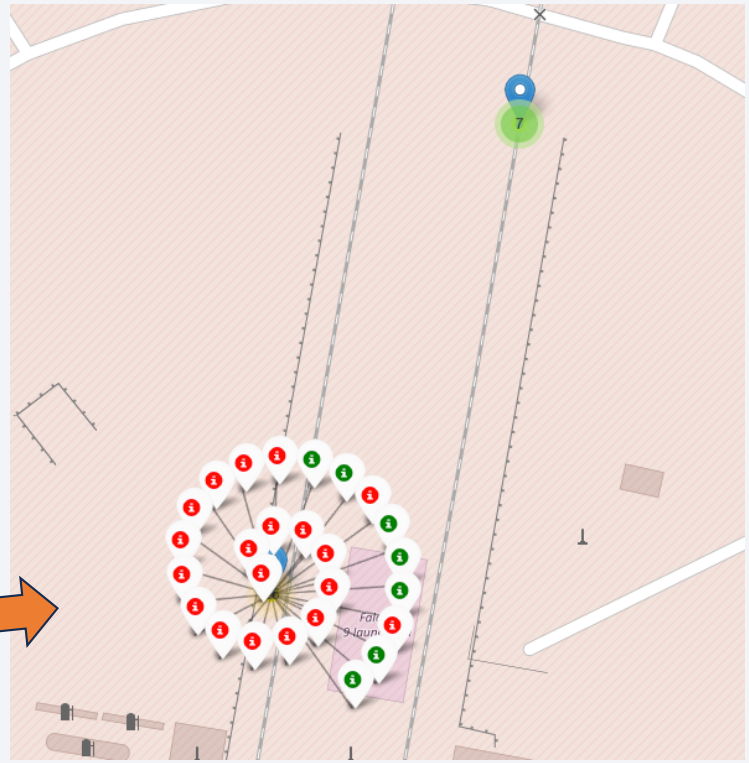
It is clear from map that launches were conducted in two major sites, one in CA, other in FL.

Launch sites and number of successes

Explore the folium map and make a proper screenshot to show the color-labeled launch outcomes on the map.

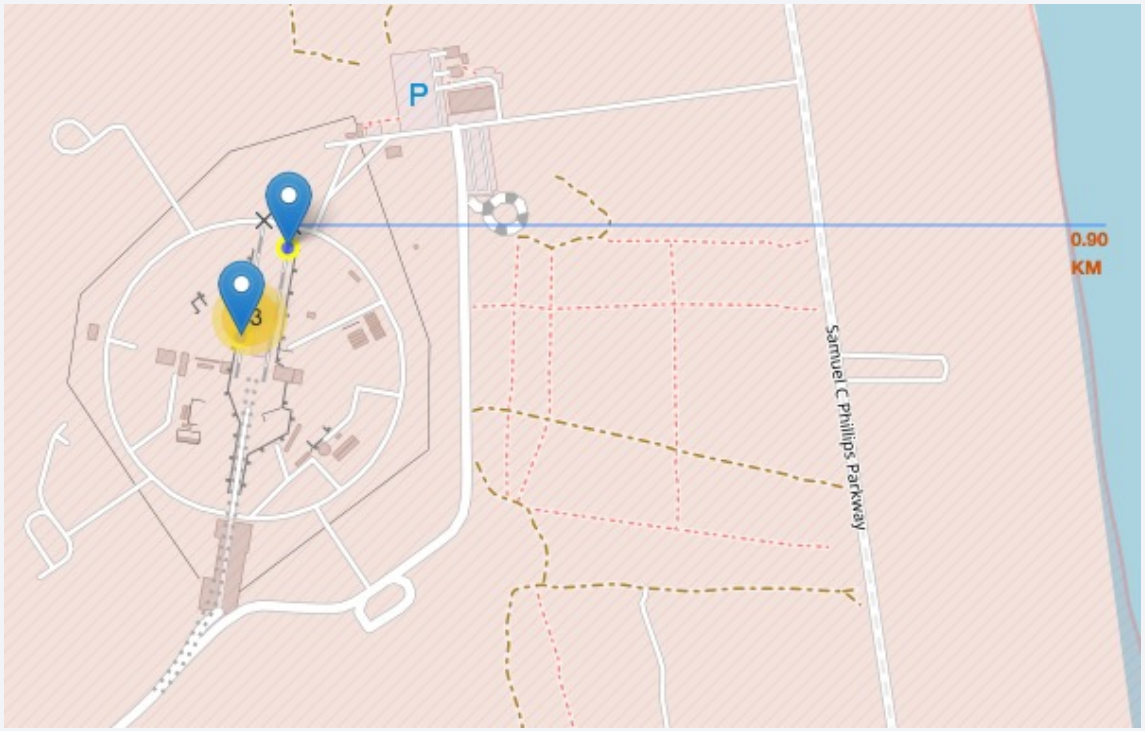


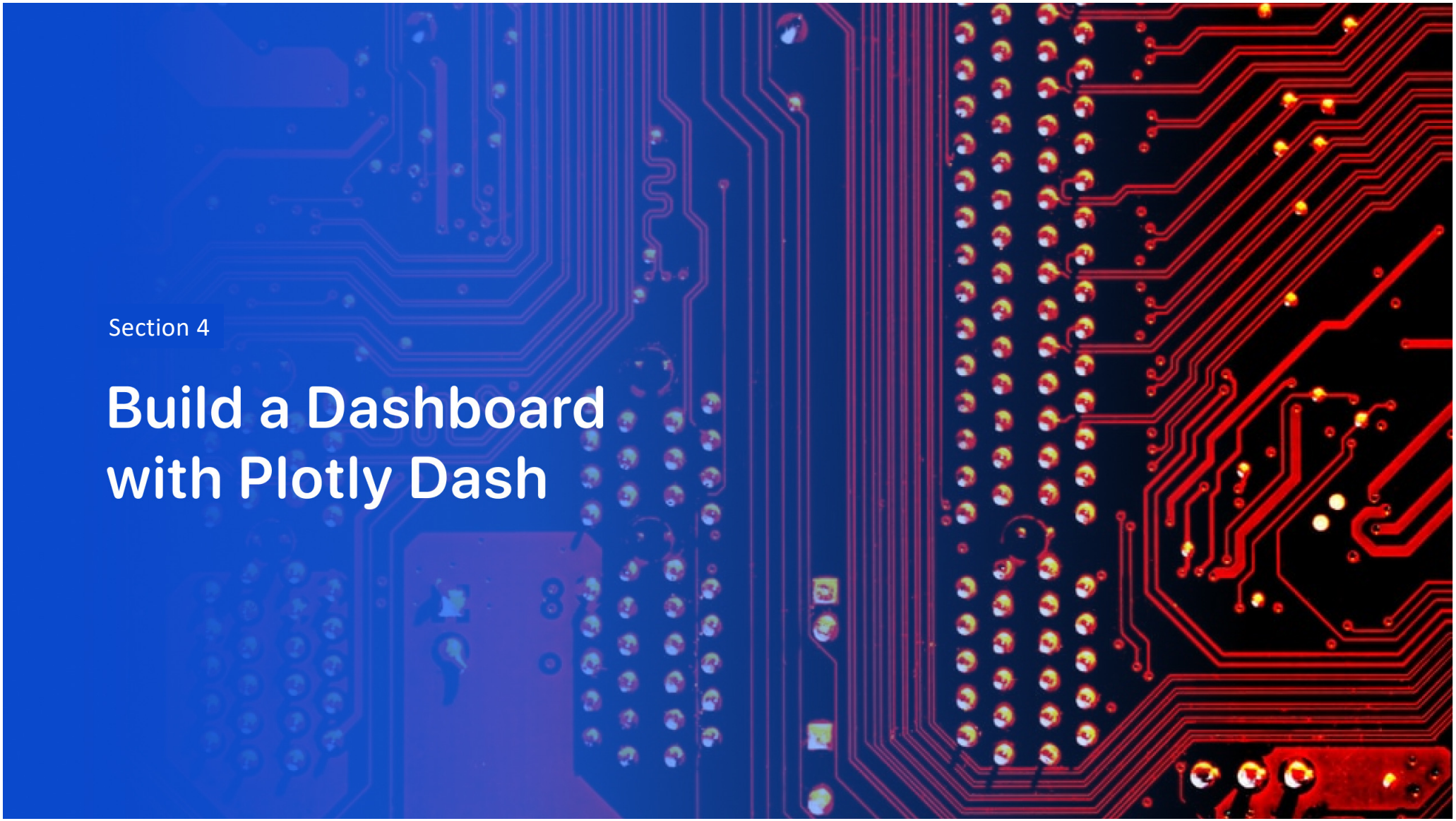




Distance to proximities

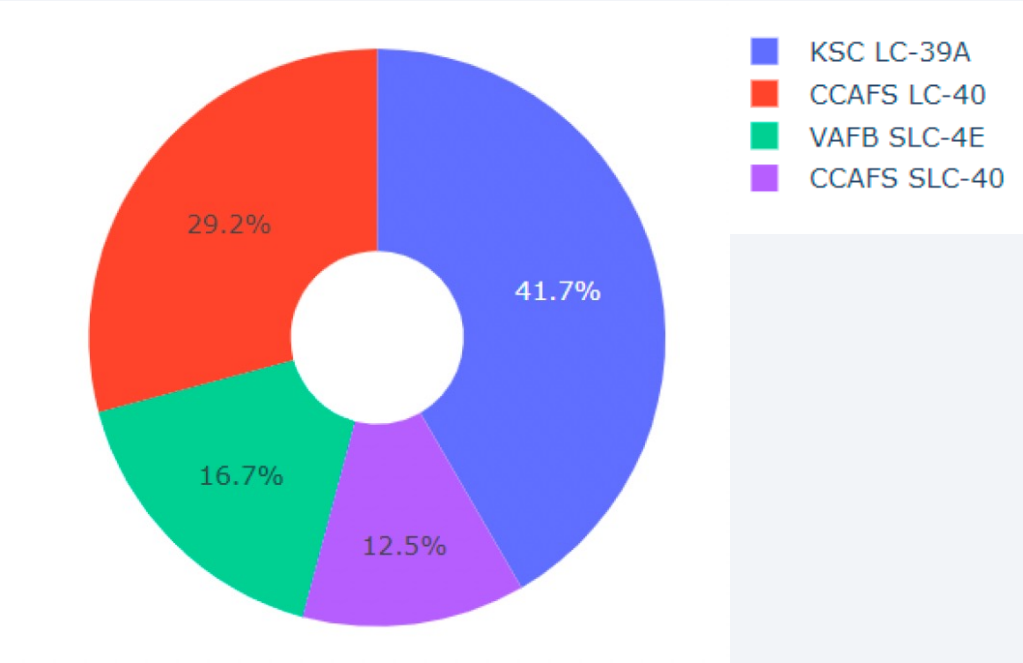
Explore the generated folium map and show the screenshot of a selected launch site to its proximities such as railway, highway, coastline, with distance calculated and displayed




Section 4
Build a Dashboard with Plotly Dash

Attempts Pie chart

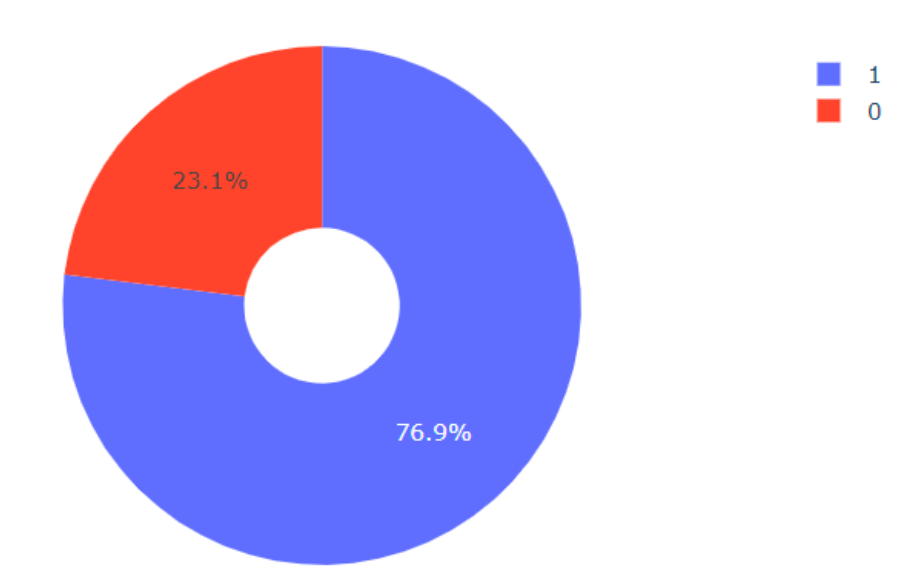
Show the screenshot of launch success count for all sites, in a piechart



- The graph shows that KSC LC-39A has highest rate

Successful launch site's success pie chart

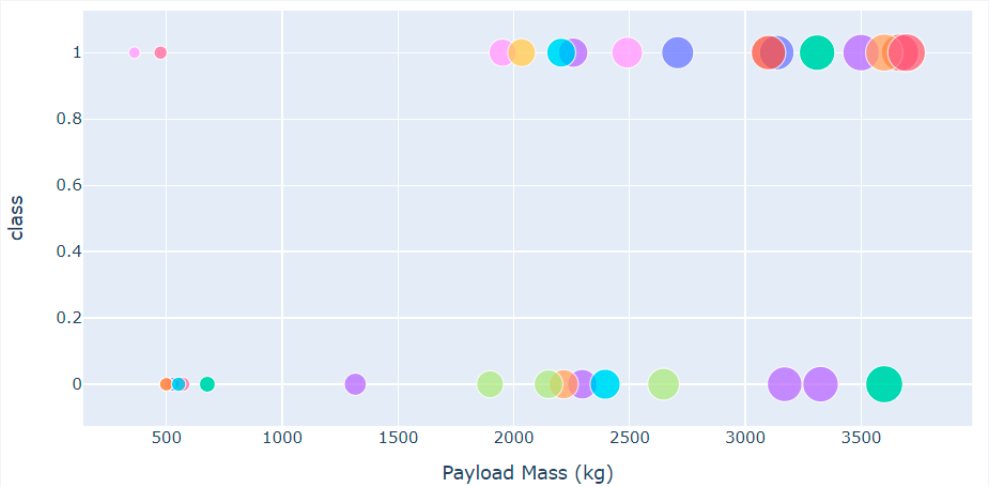
Show the screenshot of the piechart for the launch site with highest launch success ratio

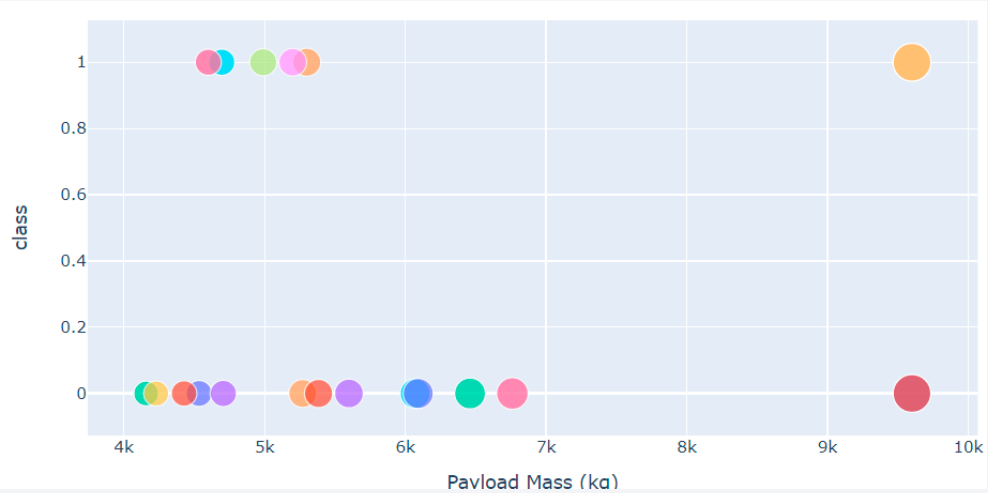


- The success rate at KSC LC-39A is 76.9%

Payload graphic

Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider



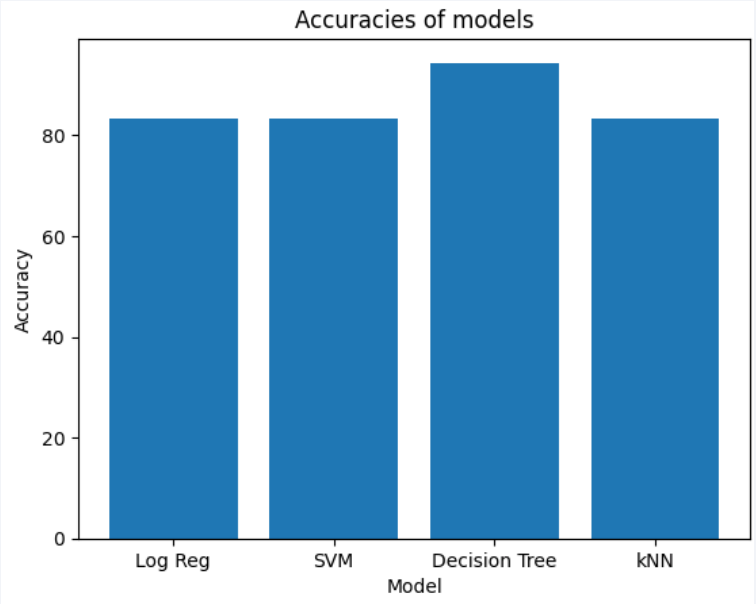


Section 5

Predictive Analysis (Classification)

Classification Accuracy

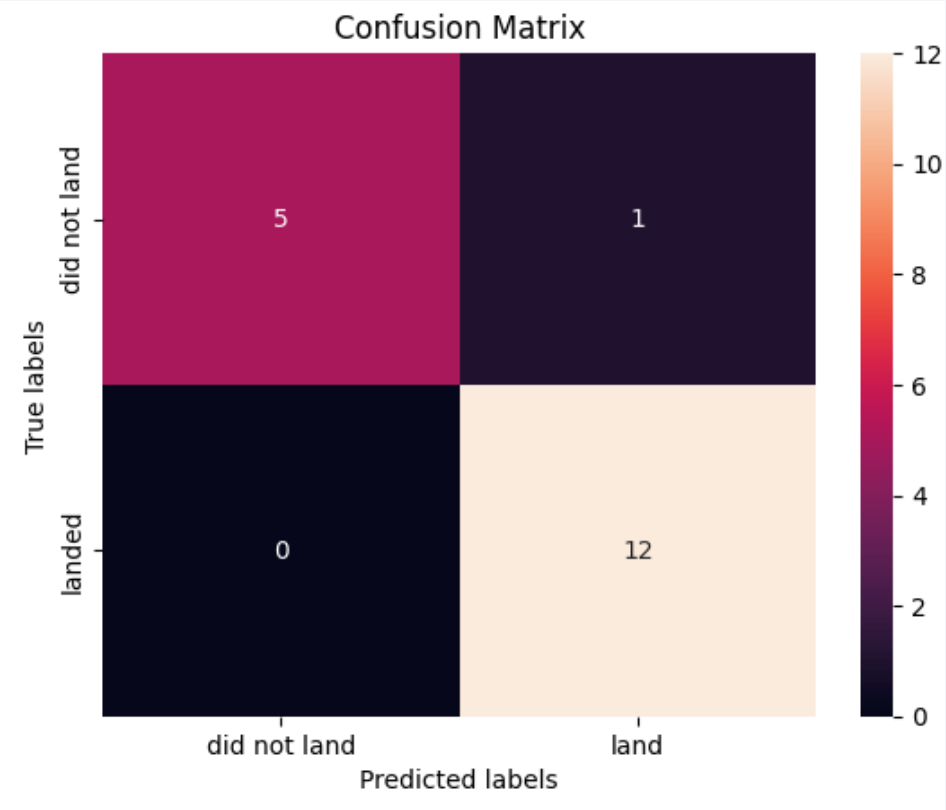
- Visualize the built model accuracy for all built classification models, in a bar chart



- Decision tree has the highest classification accuracy

Confusion Matrix

- Show the confusion matrix of the best performing model. Only one input was misclassified



Conclusions

- With proper data processing even the complex dataset like Falcon 9 launches can be converted to useful dataset to develop an classification model.
- Graphical representation could give deep insight about the data.
- Interactive graphical representation helped to understand which launch sites were more successful comparing to others.
- SQL queries can easily show top performers we need to learn
- Experiments with machine learning models show that Decision Tree model can understand the data classification better than other models with a high classification accuracy of 93.3%.

Thank you!