

# Winning Space Race with Data Science

Tanuthum Kingkaew  
23-1-2024



# Outline

---

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

# Executive Summary

---

- Methodologies:
  - Problem Definition
  - Data Collection:
    - Data Cleaning and Preprocessing
    - Exploratory Data Analysis (EDA)
  - Model Selection, Training, Evaluation and Deployment
- Summary of all results
  - Building a model for Space X Falcon 9 First Stage Landing Prediction

# Introduction

---

- Project background and context
  - The aim of the project is Space X Falcon 9 First Stage Landing Prediction
- Problems you want to find answers
  - We need to build a model in order to predict the Landing Success Rate of Space X Falcon 9



Section 1

# Methodology

# Methodology

---

## Executive Summary

- 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 model

# Data Collection

---

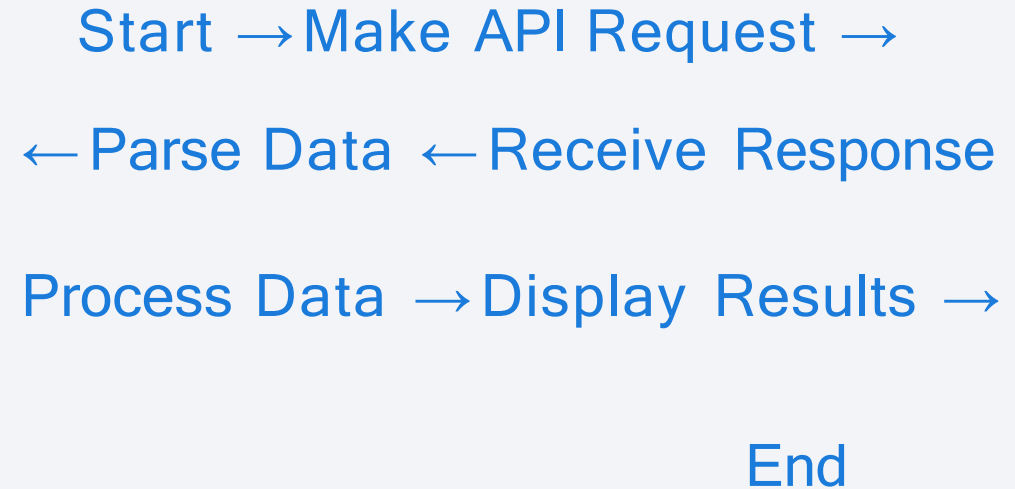
Data is collected from :

- SpaceX Rest API
- Web scraping the SpaceX website

# Data Collection - SpaceX API

---

- SpaceX provide the API to utilize its's data
- Data collection notebook in github is [here](#)





# Data Collection - Scraping

---

- Web scraping is other method that I used to collect the data
- Github link is [here](#)

Start → Choose Website → Identify Data to Scrape → Inspect HTML Structure → Scrap the Data You want → Store the Data → End

# Data Wrangling

---

- In Data Wrangling we go through the data once, how the data arranged, what are the data types present, basic statistics about the data and any addition of columns
- Github link is [here](#)

# EDA with Data Visualization

---

## ● EDA

- It is further exploration of the data by plotting different visualizations inorder to find outliers, patterns, uneven occurances etc.
- We plotted scatter plot between many variables inorder to find the relationship between variables
- And Bar graph inorder to find the count of each value in column

Github link is [here](#)

# EDA with SQL

---

- We also performed the EDA using SQL on my\_data1.db
- Github link is [here](#)

# Build an Interactive Map with Folium

---

- I used Folium to visualize the launch sites and the objects I used
  - Circle to mark the points on map
  - Markercluster, as there are many marker to make it look beautiful
  - MousePosition, to get the position of mouse on the map
  - PolyLine, to visualize the distance between two locations
- Github link is [here](#)

# Build a Dashboard with Plotly Dash

---

- I built a dashboard inorder to present the insights from data using python library dash and plotly. The graph I used are:
  - Pie chart to show the success rate of each Launch Site
  - Scatter plot to show the relation between PayloadMass and success rate
- Github link is [here](#)

# Predictive Analysis (Classification)

---

- Final I build many Classification models using the python library sklearn and found the best model using the error evaluation
- I found the best parameters for each model using GridSearchCv method.
- Github link is [here](#)

# Results

---

- In EDA we came to know that there is a correlation between payload and success rate
- Final we build a model inorder to predict the success rate of Falcon 9

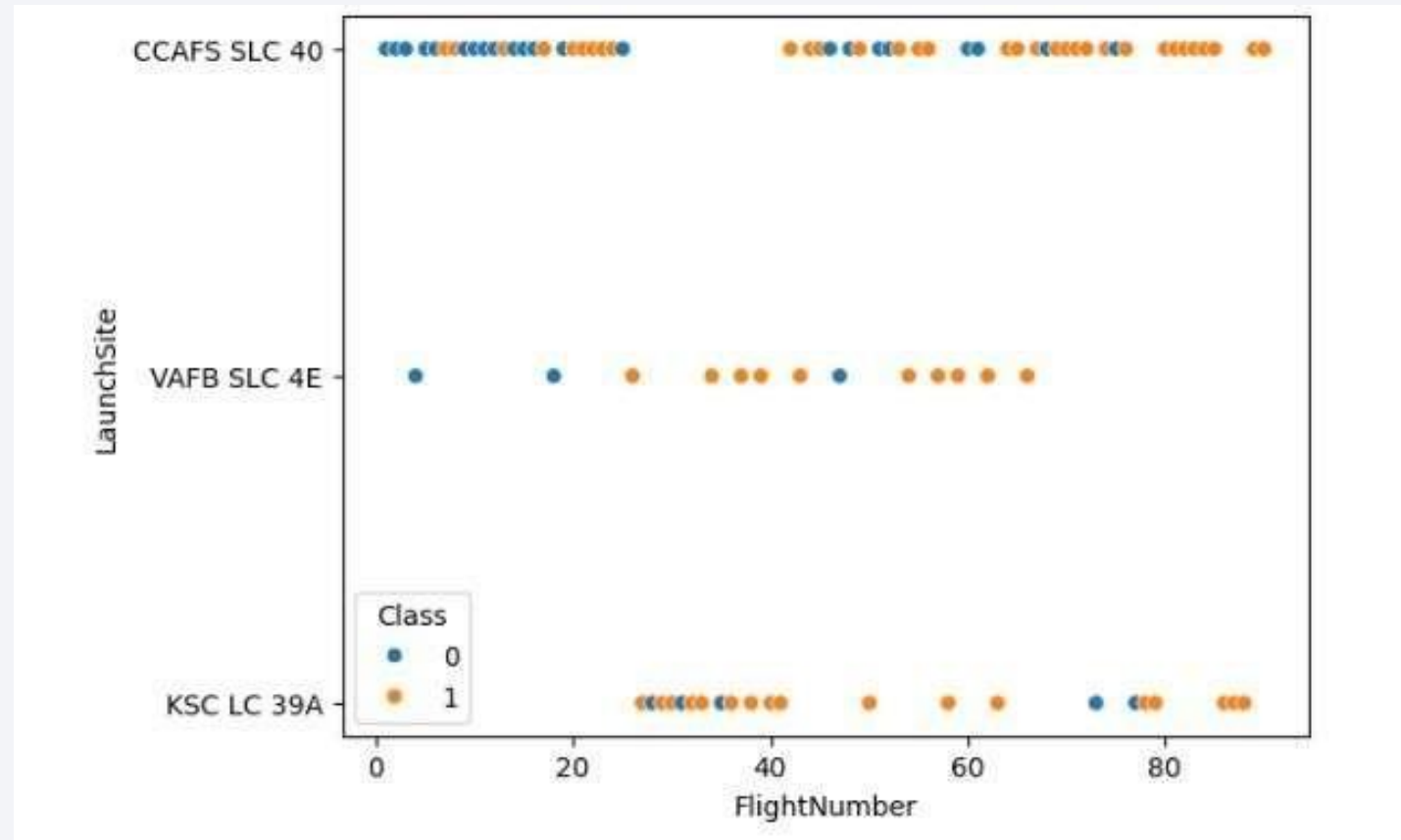


The background of the slide is an abstract composition of numerous thin, overlapping lines and streaks in shades of blue, red, and cyan. These lines are oriented diagonally, creating a sense of motion and depth. The overall effect is a vibrant, digital-looking texture.

## Section 2

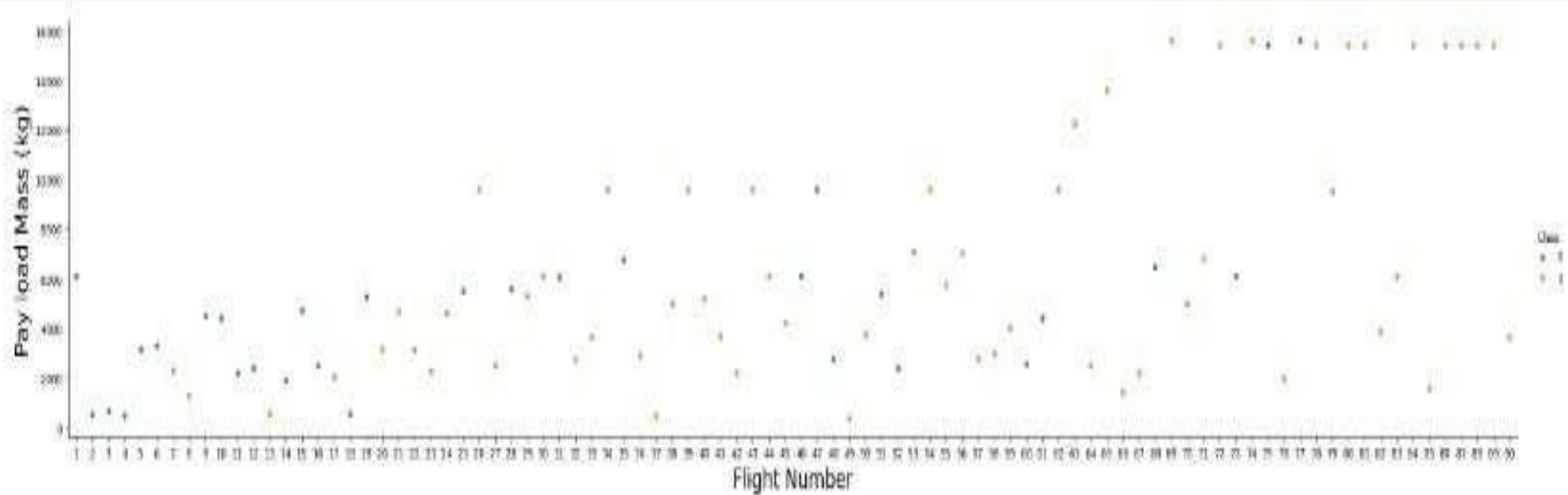
# Insights drawn from EDA

# Flight Number vs. Launch Site



We can see few flights were launched from VAFB SLC site, most of the flight launches from KSC LC are successful and most of the failed launches occurred at CCAFS SLC

# Payload vs. Launch Site

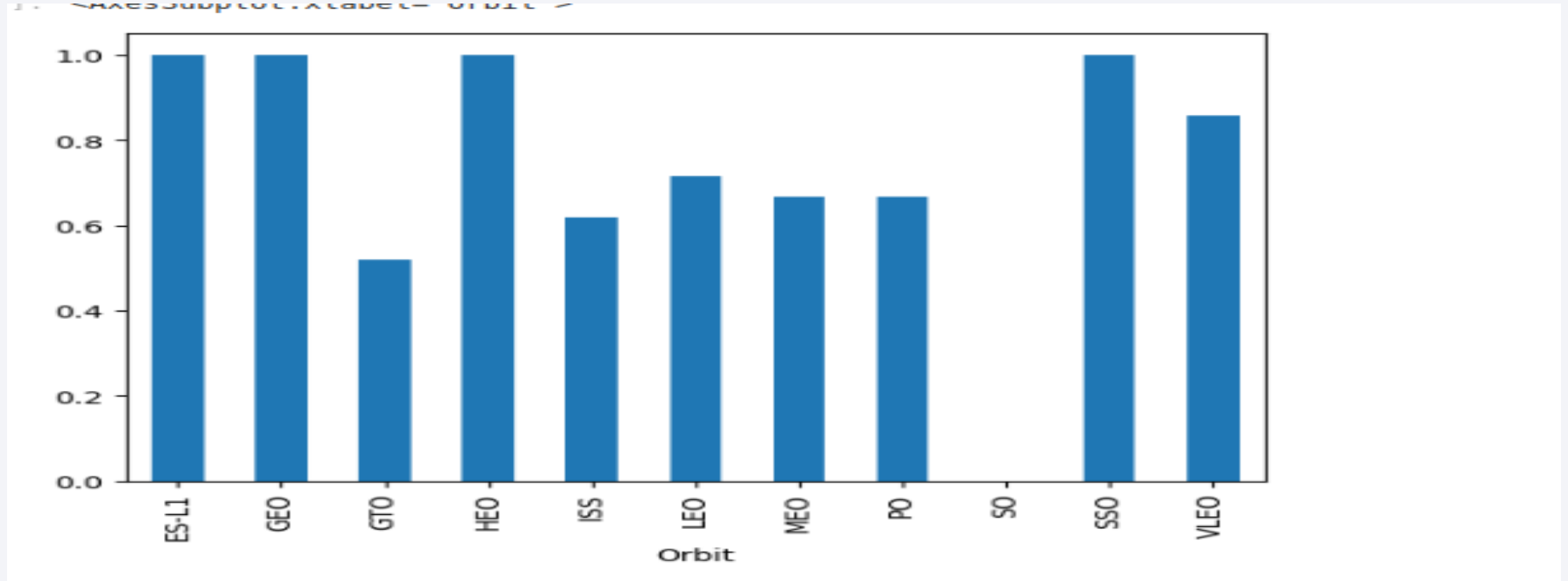


We see that different launch sites have different success rates. CCAFS LC-40 , has a success rate of 60 %, while KSC LC-39A and VAFB SLC 4E has a success rate of 77%.

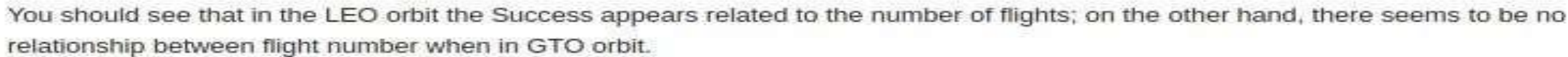


# Success Rate vs. Orbit Type

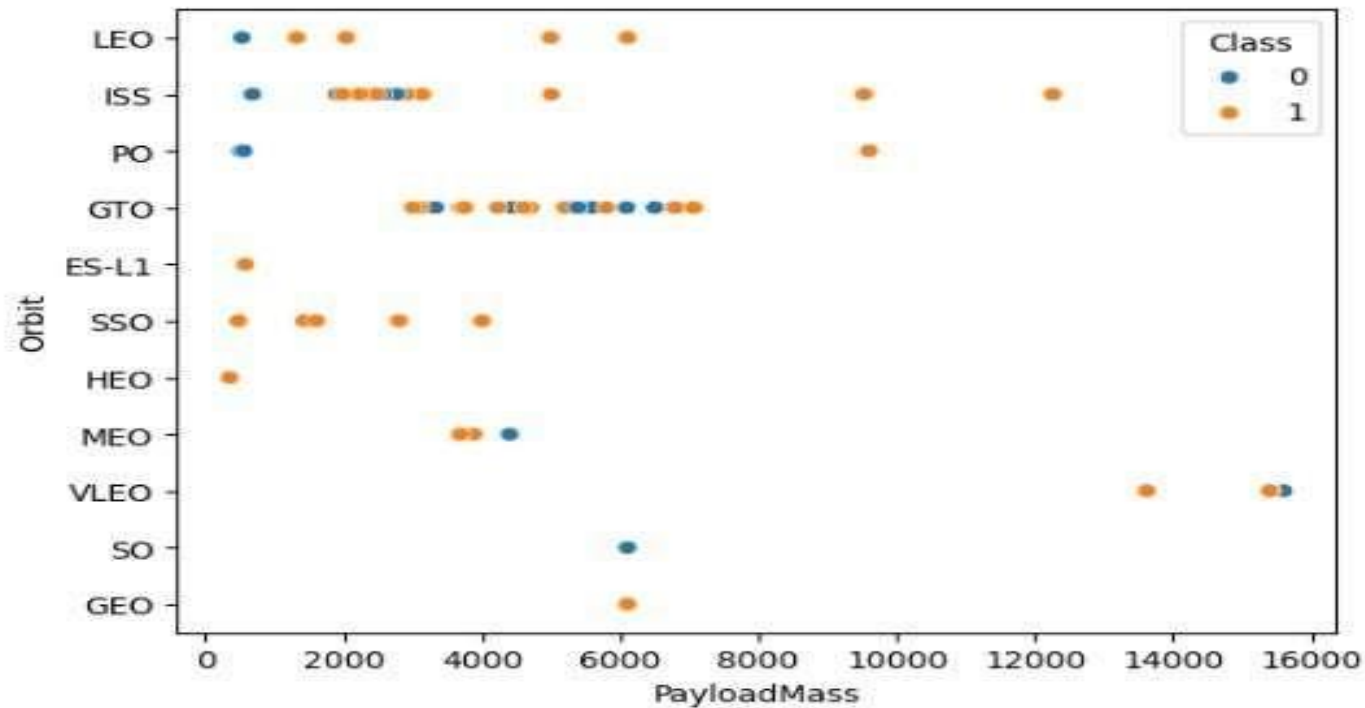
---



We can see that ES-L1, GEO, HEO, SSO type of orbits have good success rate than other orbits



# Payload vs. Orbit Type

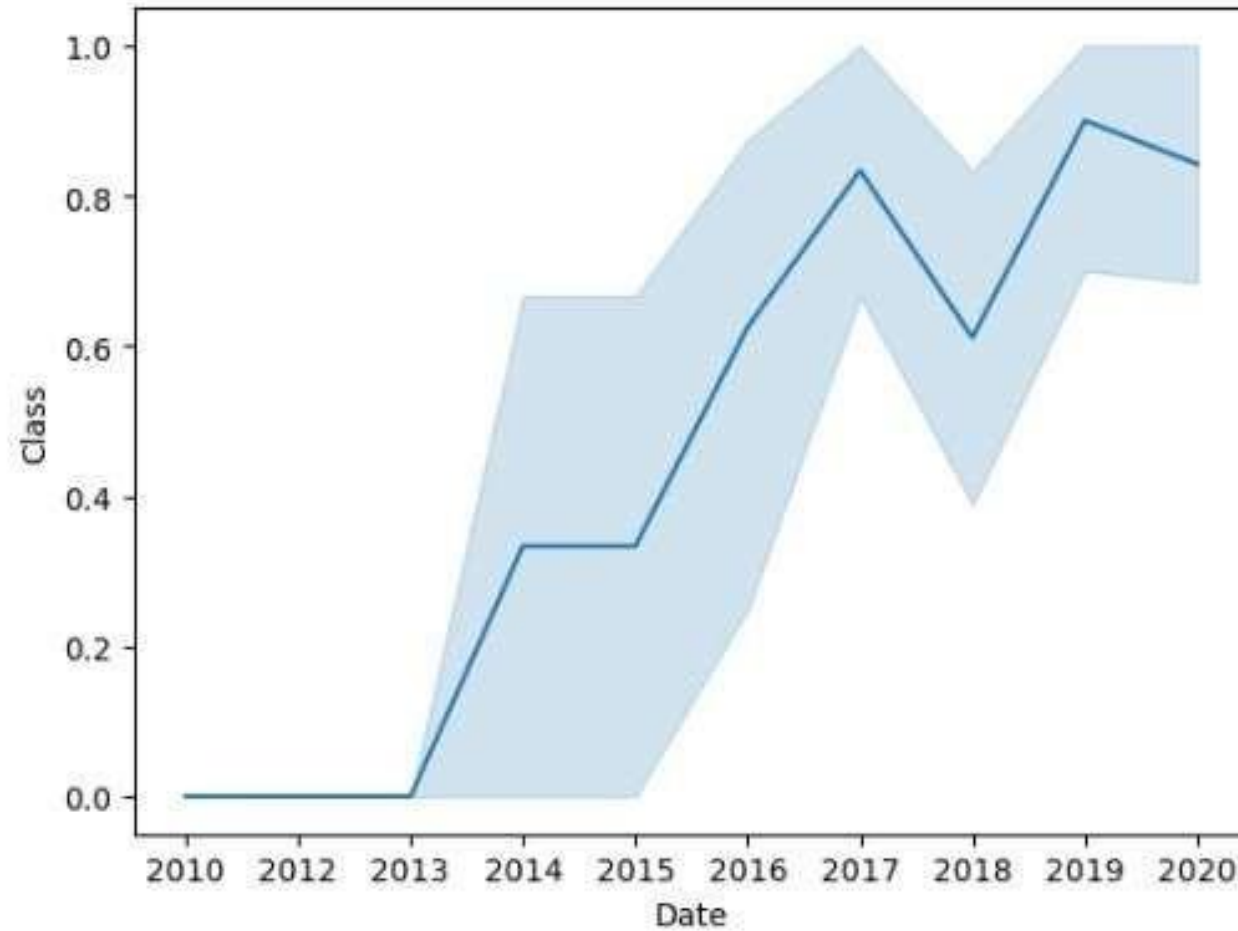


With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS.

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

# Launch Success Yearly Trend

---



you can observe that the success rate since 2013 kept increasing till 2020

# All Launch Site Names

---

```
[49]: %sql select DISTINCT("Launch_Site") from SPACEXTBL
```

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

```
[49]: Launch_Site
```

```
CCAFS LC-40
```

```
VAFB SLC-4E
```

```
KSC LC-39A
```

```
CCAFS SLC-40
```



# Launch Site Names Begin with 'CCA'

```
%sql select * from SPACEXTBL where "Launch_Site" like "CCA%" limit 5
```

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

```
Done.
```

Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS_KG	Orbit	Customer	Mission_Outcome	Landing_Outcome
18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
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)
7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

# Total Payload Mass

---

```
53]: %sql select sum(PAYLOAD_MASS_KG_) as "Load", "Customer" from SPACEXTBL group by "Customer" having Customer = "NA"
```

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

```
Done.
```

```
53]:
```

Load	Customer
45596	NASA (CRS)

# Average Payload Mass by F9 v1.1

---

```
6|: %sql select avg(PAYLOAD_MASS_KG_) as "Avg_pay_load" , "Booster_Version" from SPACEXTBL group by "Booster_Version"
```

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

```
Done.
```

```
6|: Avg_pay_load  Booster_Version
      2928.4         F9 v1.1
```

# First Successful Ground Landing Date

---

```
%sql select min(Date) as "first successful landing" from SPACEXTBL where Landing_Outcome = "Success (ground pad)"
```

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

```
Done.
```

<u>first successful landing</u>
---------------------------------

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

# Total Number of Successful and Failure Mission Outcomes

---

```
%%sql select count(*) as "Total count", CASE
      WHEN Mission_Outcome like "%Failure%" then "Failure"
      else "Success"
    end as simplified_outcome
from SPACEXTBL group by simplified_outcome
```

\* sqlite:///my\_data1.db

Done.

Total count	simplified_outcome
1	Failure
100	Success

# Boosters Carried Maximum Payload

---

```
%sql select DISTINCT(Booster_Version) from SPACEXTBL 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

---

```
substr(Date,6,2) = 2015 for year
```

```
] %sql select substr(Date,6,2) as "month",Landing_Outcome,Booster_Version,Launch_Site from SPACEXTBL where substr(D
* sqlite:///my_data1.db
Done.
```

```
] month Landing_Outcome Booster_Version Launch_Site
01 Failure (drone ship) F9 v1.1 B1012 CCAFS LC-40
04 Failure (drone ship) F9 v1.1 B1015 CCAFS LC-40
```



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

```
%sql select count(*) as "Count",Landing_Outcome from SPACEXTBL group by Landing_Outcome having Date Between "2010-06-04" and "2017-03-20"
```

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

```
Done.
```

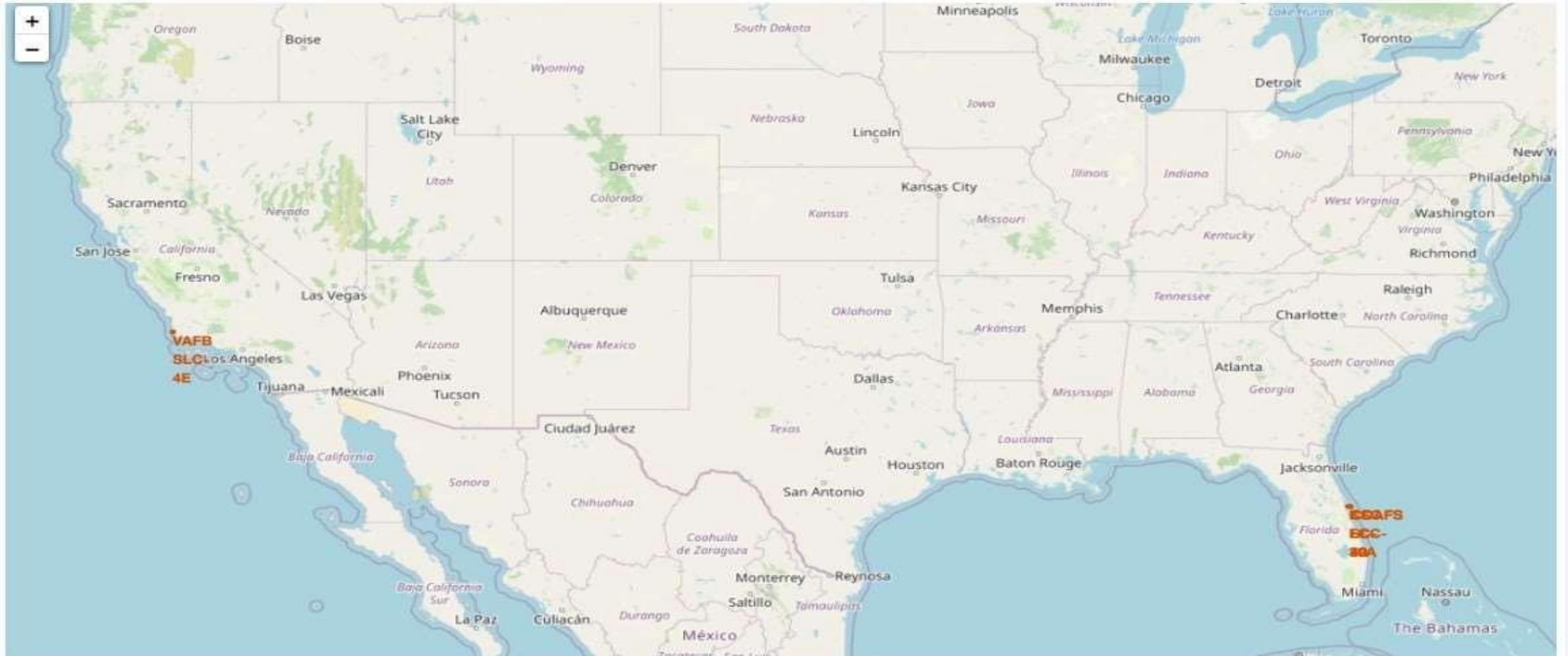
Count	Landing_Outcome
21	No attempt
14	Success (drone ship)
9	Success (ground pad)
5	Failure (drone ship)
5	Controlled (ocean)
2	Uncontrolled (ocean)
2	Failure (parachute)
1	Precluded (drone ship)

A satellite view of Earth from space, showing the curvature of the planet and city lights at night. The image is a composite of a solid blue background on the left and a satellite photograph of Earth on the right. The Earth's surface is dark, with numerous bright yellow and orange lights representing cities and urban areas. The horizon line of the Earth is visible, separating the dark surface from the deep blue of the upper atmosphere and space.

Section 3

# Launch Sites Proximities Analysis

# Mark all launch sites on a map



We can observe that all the four Launch sites are near to the coast, road, railways and cities

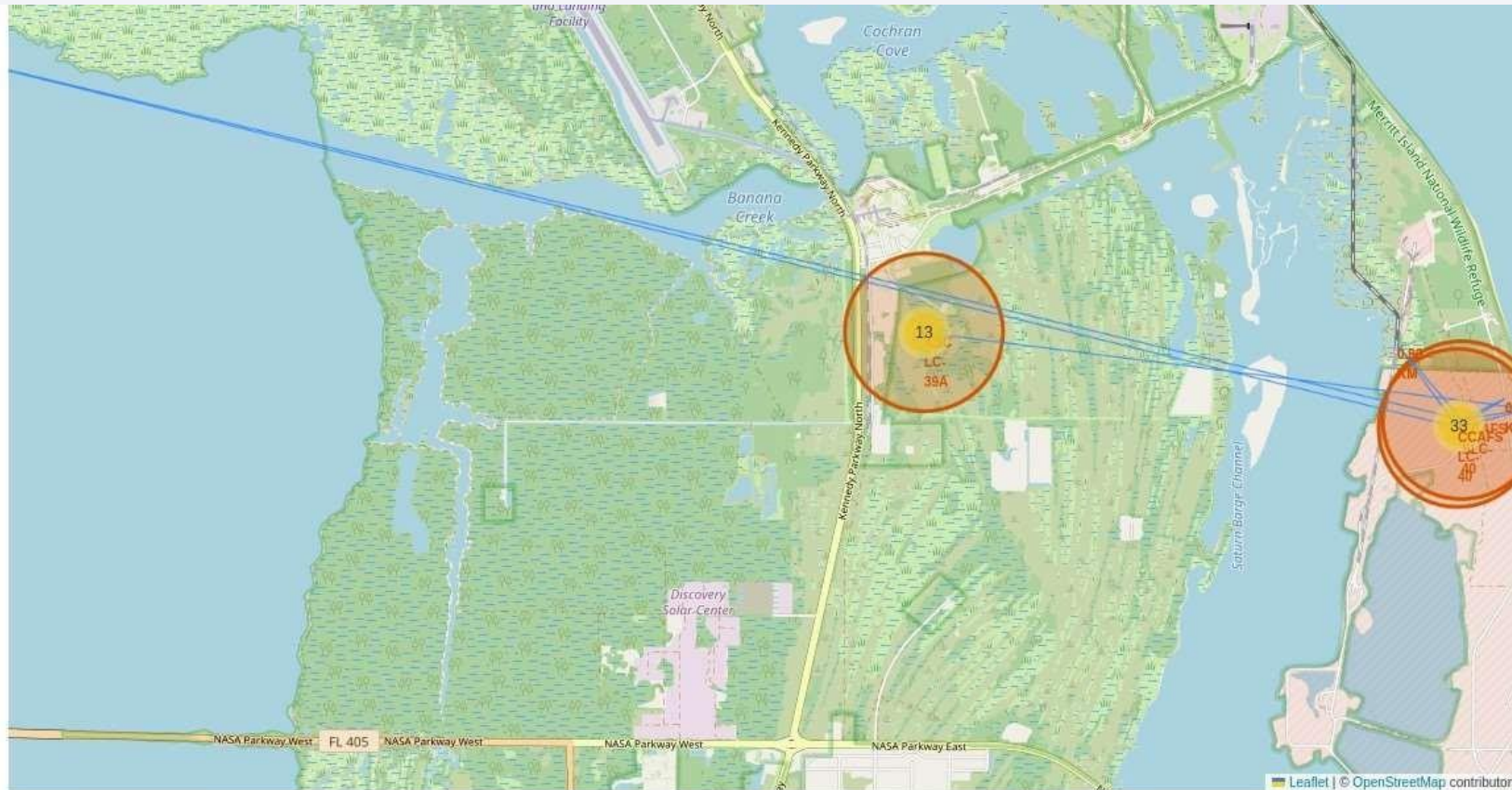


# Success/Failed launches for each site



From the color-labeled markers in marker clusters, you should be able to easily identify which launch sites have relatively high success rates.

# Distances btw a launch site to its proximities



All the four are near to the railways, highways, coastline and from cities But the VAFB SLC-4E Station is far from highway and city







Section 4

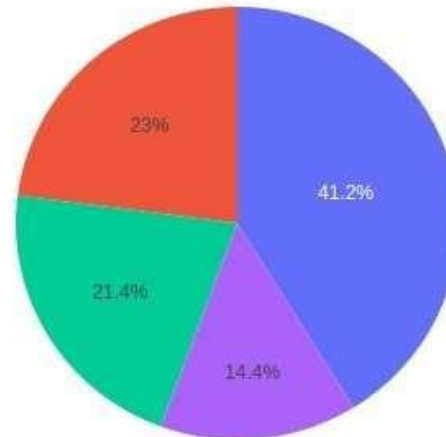
# Build a Dashboard with Plotly Dash

# Pie chart for ALL sites

## SpaceX Launch Records Dashboard

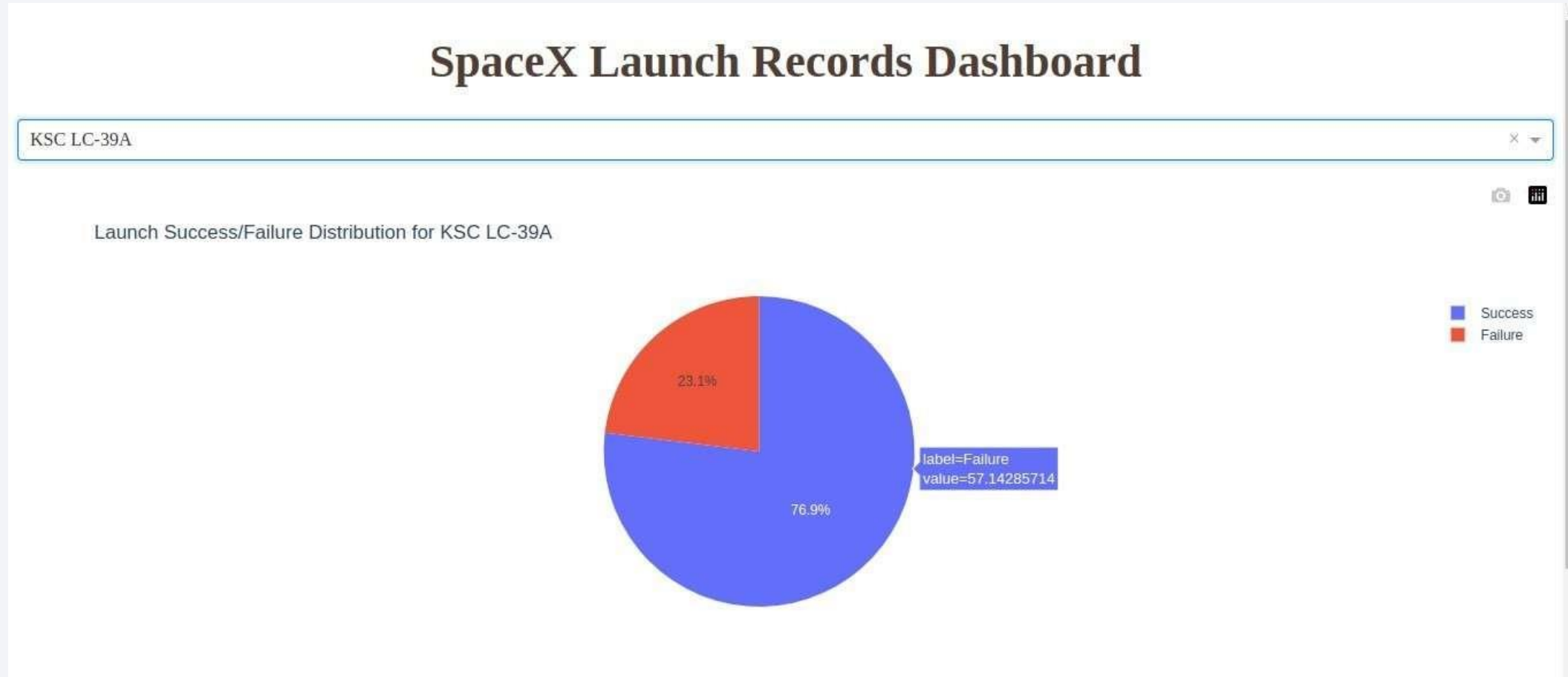
All Sites

Success Percentage for Each Launch Site



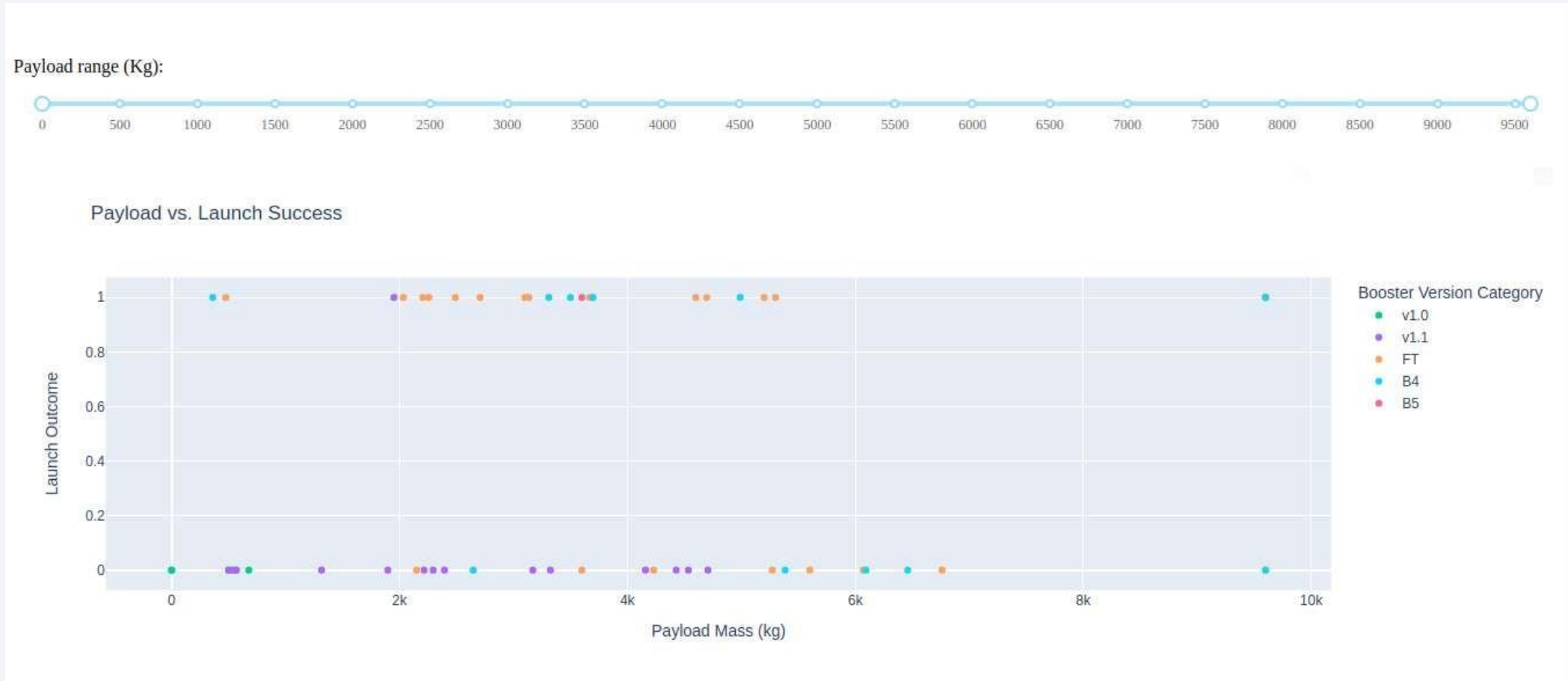
■ KSC LC-39A  
■ CAFS SLC-40  
■ VAFB SLC-4E  
■ CAFS LC-40

# Pie chart for Launch site with max success





# Scatter plot btw payloadmass and success rate





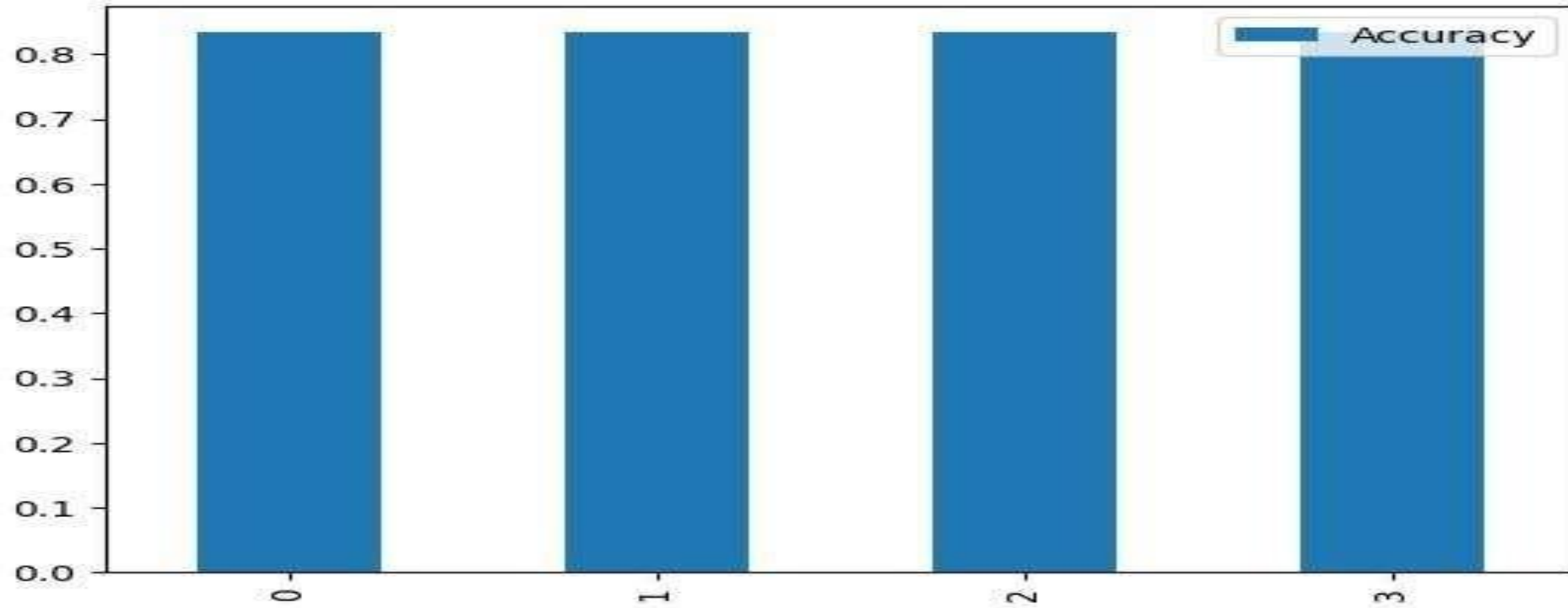
Section 5

# Predictive Analysis (Classification)

# Classification Accuracy

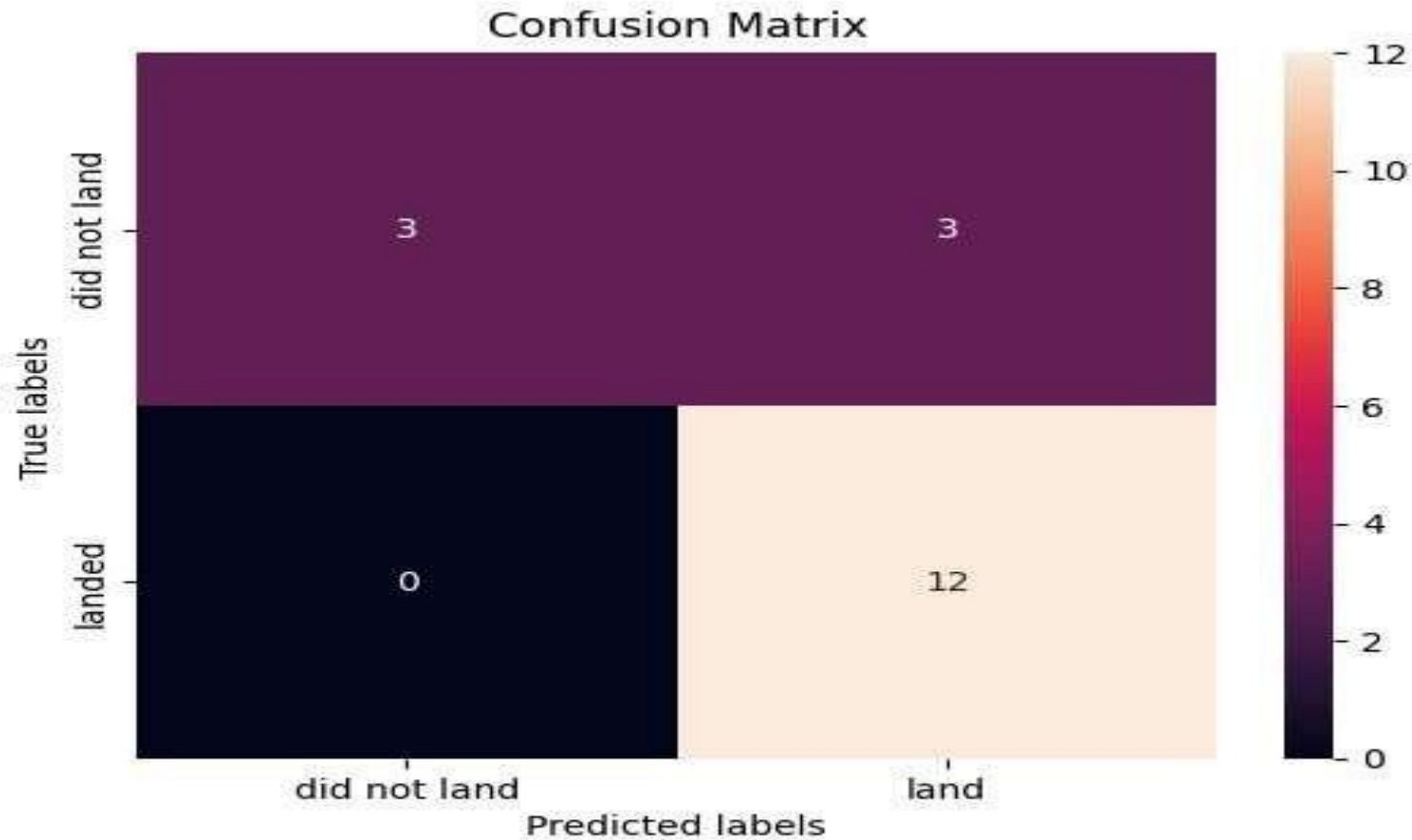
---

```
In: df.plot(kind="bar")  
plt.show()
```



We can observe that all model have same accuracy score

# Confusion Matrix



All the models have same accuracy score i.e all models are best. But now I am showing confusion matrix for KNN model

# Conclusions

---

- Except DecisionTree method all other perform the same and best than Decision Tree
- Finally we built a model inorder to predict the success rate of First Stage Landing of SpaceX Falcon 9
- The model is able to predict with 83.333 accuracy which sounds good to hear.

# Appendix

---

For full access to all the python notebooks and files you can visit the github. Link is [here](#).

Thank you!

