IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

JAGDEEP SINGH MAAN
December 2, 2022

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection
  - Data Wrangling
  - Exploratory Data Analysis With Data Visualization
  - Exploratory Data Analysis With SQL
  - Building An Interactive Map With Folium
  - Building A Dashboard With Plotly Dash
  - Predictive Analysis (Classification)
- Summary of all results
  - Exploratory Data Analysis Results
  - Interactive Analytics
  - Predictive Analysis

Introduction

- Project background and context

Space Exploration Tehcnologies Corporation is an American Spacecraft Manufacturer, Launcher and a Satellite Communications Corporation Headquartered in Hawthrone, California Founded by Elon Musk in 2002 with aim of reducing space transportation costs.

- Problems you want to find answers

This Project is looking to find answer about 'Will SpaceX Falcon Launcher will Successfully Land back to earth after spacecraft or satellite Launch?'

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data Collection using SpaceX API
  - Data Collection with Web Scrapping From Wikipedia
- Perform data wrangling
  - Changing categorical data into numeric data.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Build Better Classification Models using Model Training and Cross Validation.

Data Collection

Methods use for Data Collection:

SpaceX REST API

Data is collected using SpaceX REST API which contains detailed information about past SpaceX launches..

Web Scraping Method:

SpaceX's Wikipedia page includes a lot of information about SpaceX launches and data is collected using BeautifulSoup Python Library

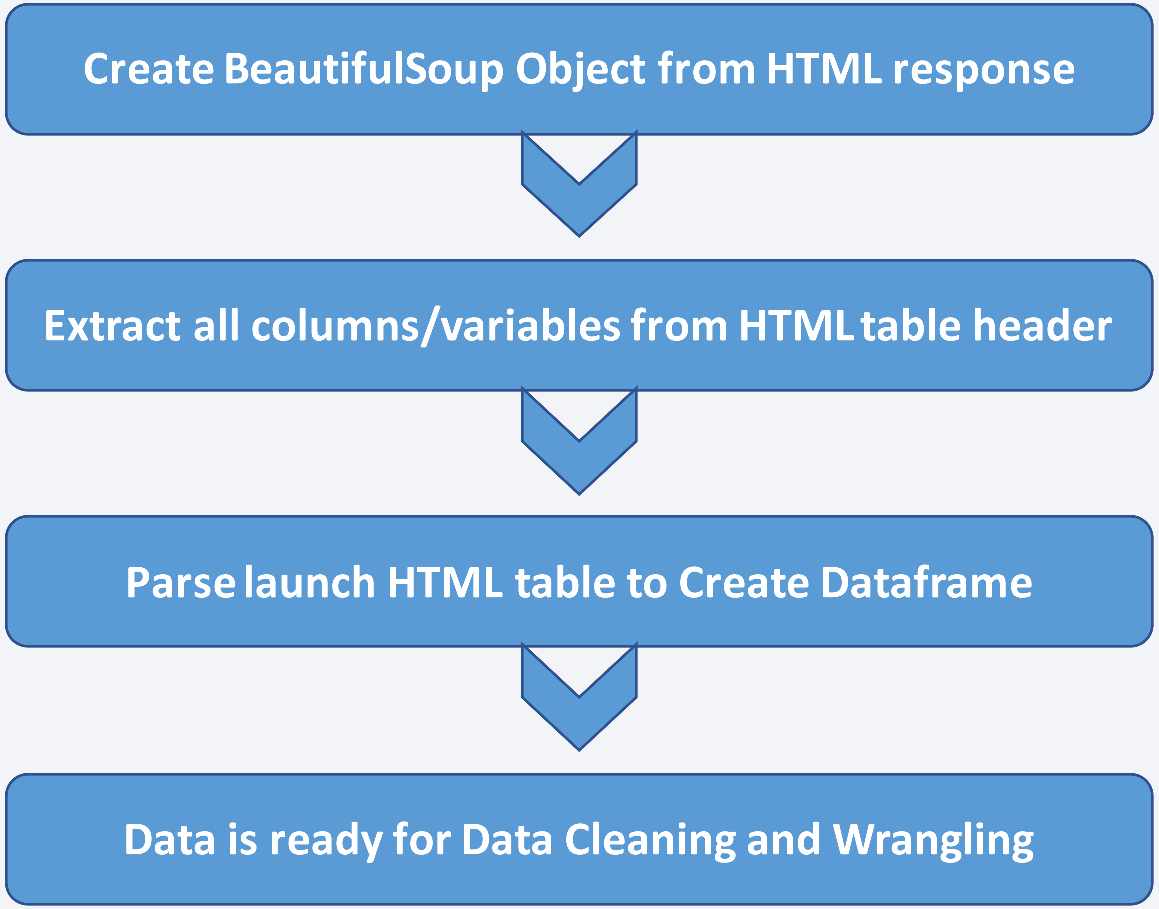
Data Collection - Scraping

SpaceX Wikipedia Page

https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922

GitHub Link For Notebook

https://github.com/JagdeepMaan/JagdeepMaan/blob/beac177fcad3b505cfe9514881aa8670ee15b279/jupyter-labs-webscraping.ipynb



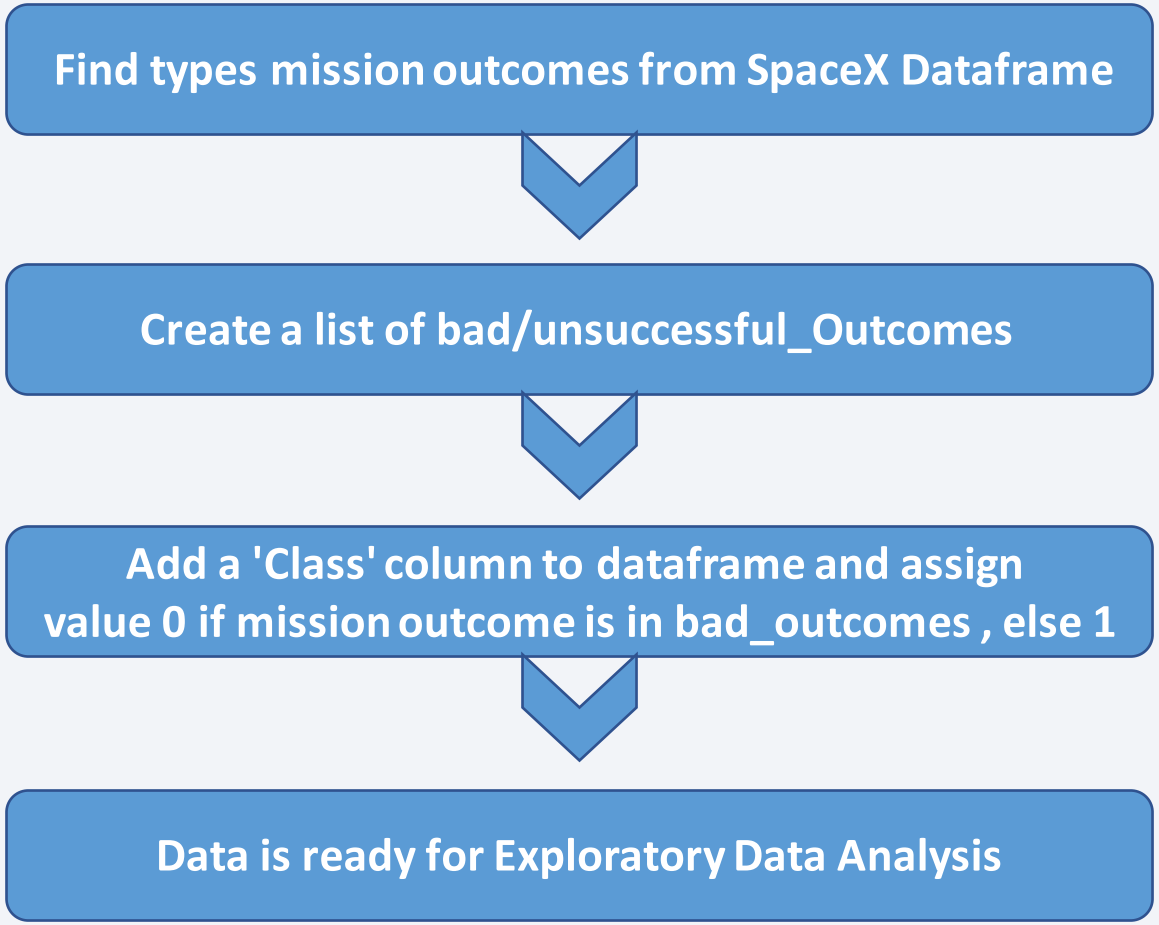
Data Wrangling

SpaceX Launch Outcomes were converted into numerical data and added 'class' column to describe if launch outcome was successful.

0 = Unsuccessful, 1 = Successful

GitHub Link For Notebook

https://github.com/JagdeepMaan/JagdeepMaan/blob/dc49aeffbfed2aa33ab38655783e71df9fdc1759/IBM-DS0321EN-SkillsNetwork_labs_module_1_L3_labs-jupyter-spacex-data_wrangling_jupyterlite.jupyterlite.ipynb

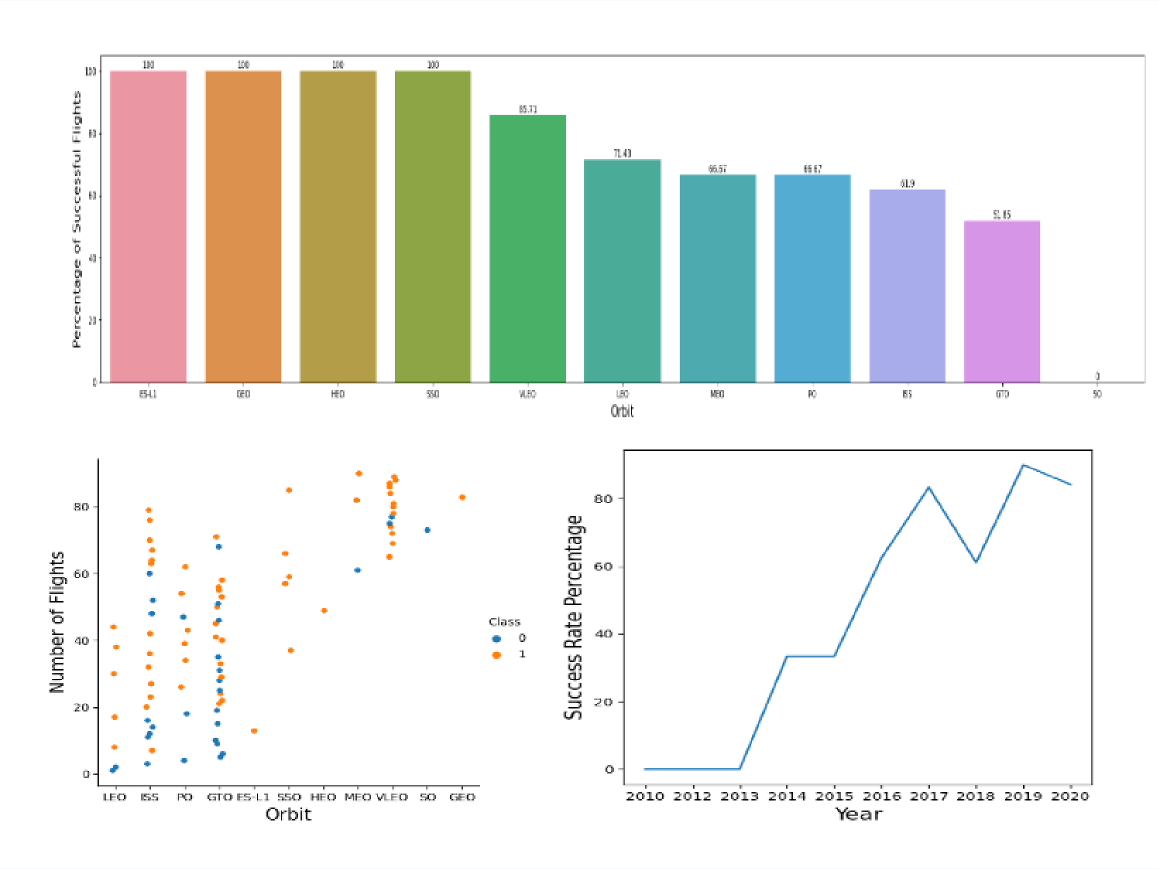


EDA with Data Visualization

In Exploratory Data Analysis collected and cleaned data is used to show relationships between variables to know more about the data points by visualization of data in charts and graphs.

GitHub Link For Notebook

https://github.com/JagdeepMaan/JagdeepMaan/blob/d1b6aa6a2317fa7895b478612bdebfa0031f15b8/eda-dataviz.ipynb.jupyterlite.ipynb



EDA with SQL

In Exploratory Data Analysis collected and cleaned data using Structured Query Language -SQL.

GitHub Link For Notebook

https://github.com/JagdeepMaan/JagdeepMaan/blob/bad2eae720995b4c1d59da8f0369b785cf403206/jupyter-labs-eda-sql-coursera_sqllite.ipynb

SQL Queries Used For Data Analysis

- Getting Launch Sites Names
- Total Payload Carried By All Spacecrafts
- Average Payload Per Flight
- Number of Successful Mission Outcomes
- First successful launch with safe booster landing on ground pad
- Name of Booster Versions

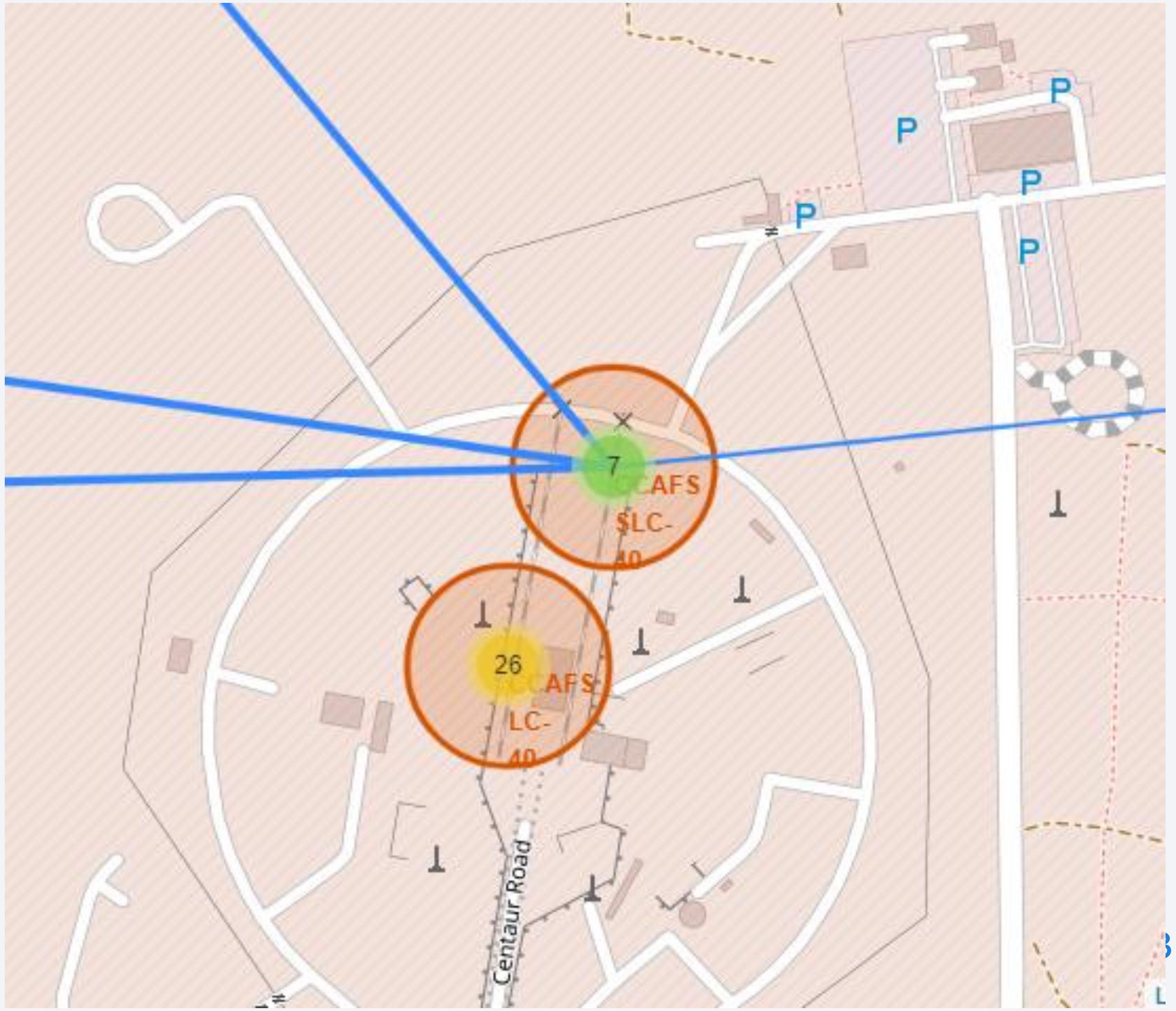
Build an Interactive Map with Folium

Interactive Map with Folium

- Adding Marker and Circle for each SpaceX Launch Site.
- Marked all Flights from each launch site using Marker Cluster object.
- Added Lines From CCAFS SLC-40 to Landmarks

GitHub Link For Notebook

https://github.com/JagdeepMaan/JagdeepMaan/blob/3ddb5aa85021aff0d5eedb221c1ec3ed79523e1f/launch_site_location.jupyterlite.ipynb



Build a Dashboard with Plotly Dash

- **Added Pie Chart**

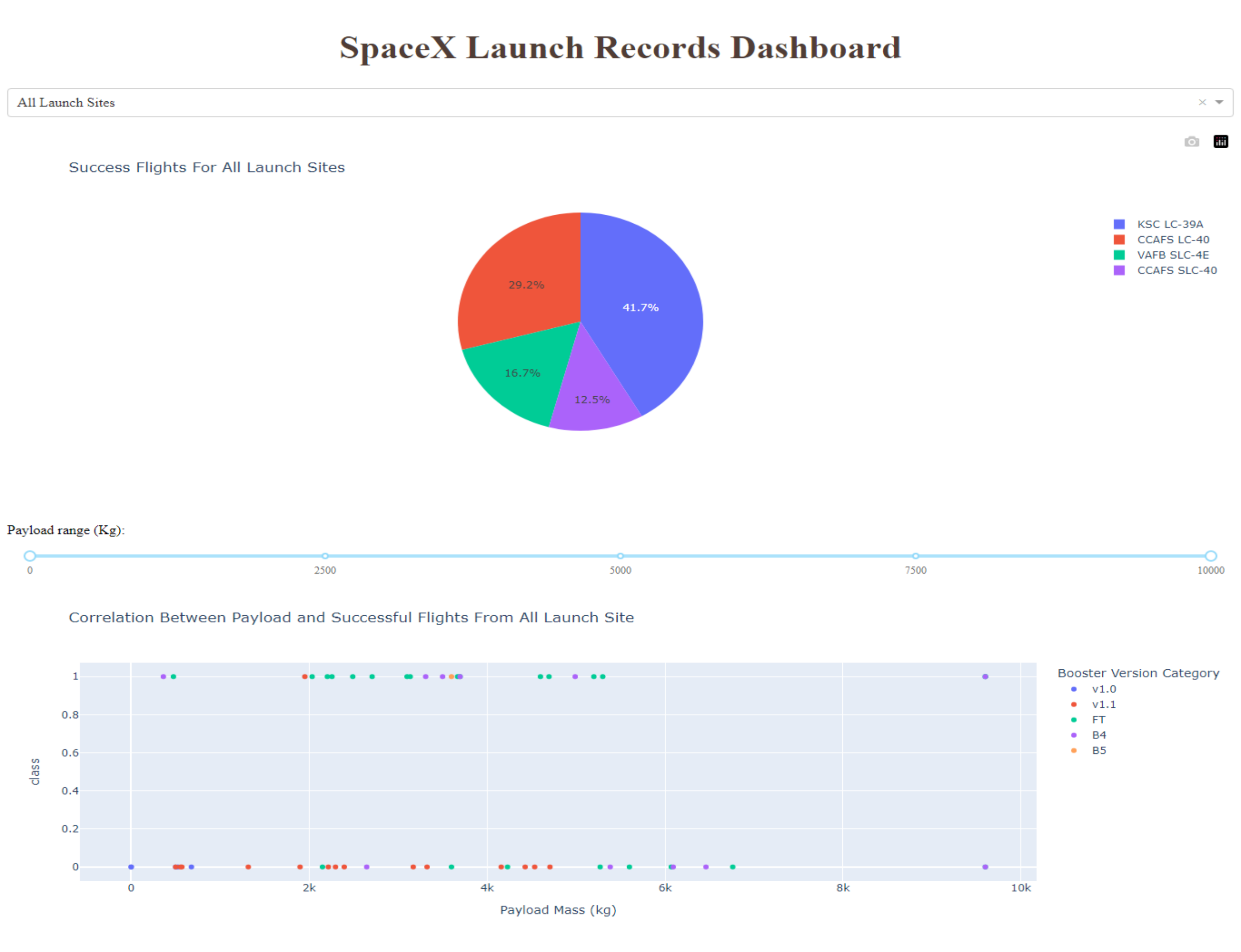
Displaying percentage number of flights from each launch site

- **Added Scatter Chart with Range Slider**

Filter using payload mass to get data points about launches with different booster versions.

GitHub Link For Notebook

https://github.com/JagdeepMaan/JagdeepMaan/blob/80bc9896feb35dcfcd7cbefc91ecb9aa298f264c/spacex_dash_app.py



Predictive Analysis (Classification)

Models for Predictive Analysis

Linear Regression

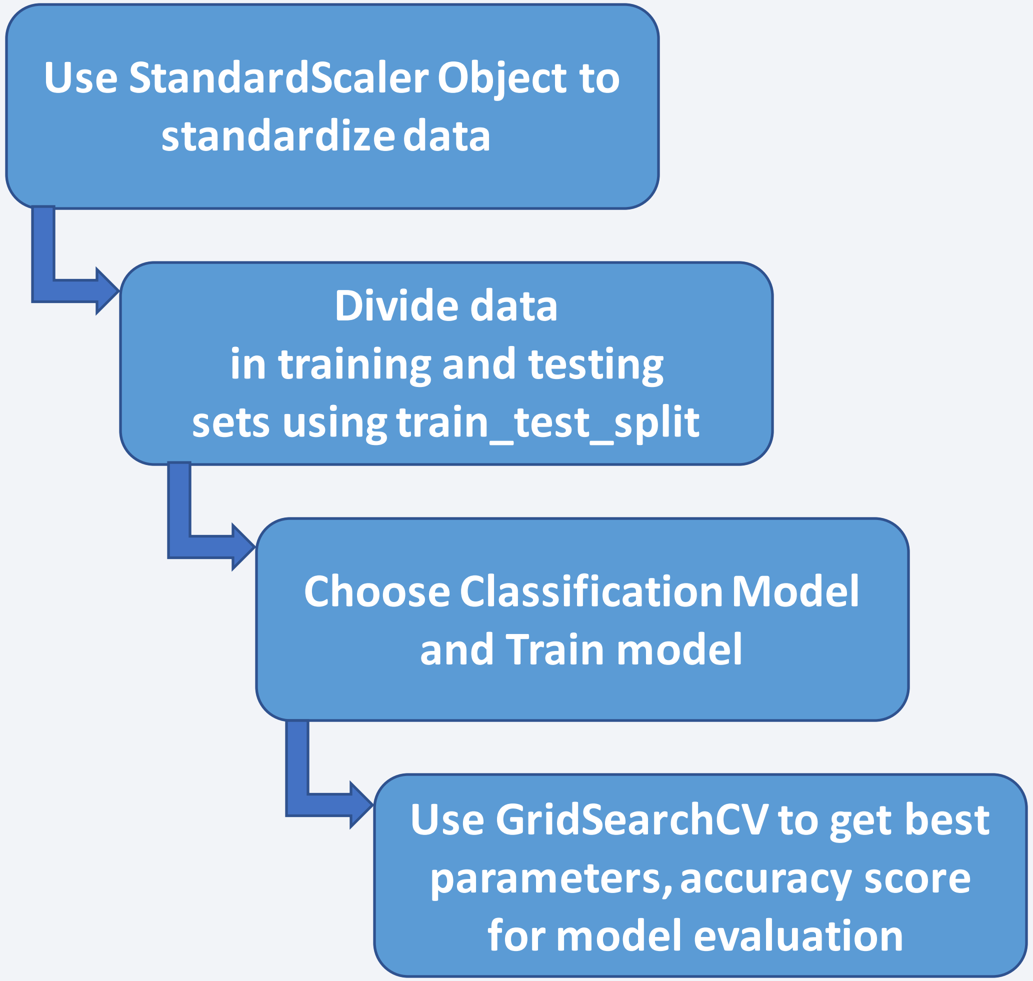
DecisionTree Classifier

Support Vector Classifier

KNearest Neighbors

GitHub Link For Notebook

https://github.com/JagdeepMaan/JagdeepMaan/blob/482eb477fb8b265eed41158686f912b3cdc187bd/IBM-DS0321EN-SkillsNetwork_labs_module_4_SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb



Results

- **Exploratory Data Analysis**

  Flight Number vs Launch Site, Payload vs Launch Site, Success Rate vs Orbit, Payload vs Orbit, Yearly Launch Success Rate, All Launch Sites Name, Total Payload Mass, Launch By Year

- **Interactive Data Visualization with Folium Map**

  SpaceX Launch Site on Folium Map With Landing Outcome, Distance from Launch Site To Important Locations

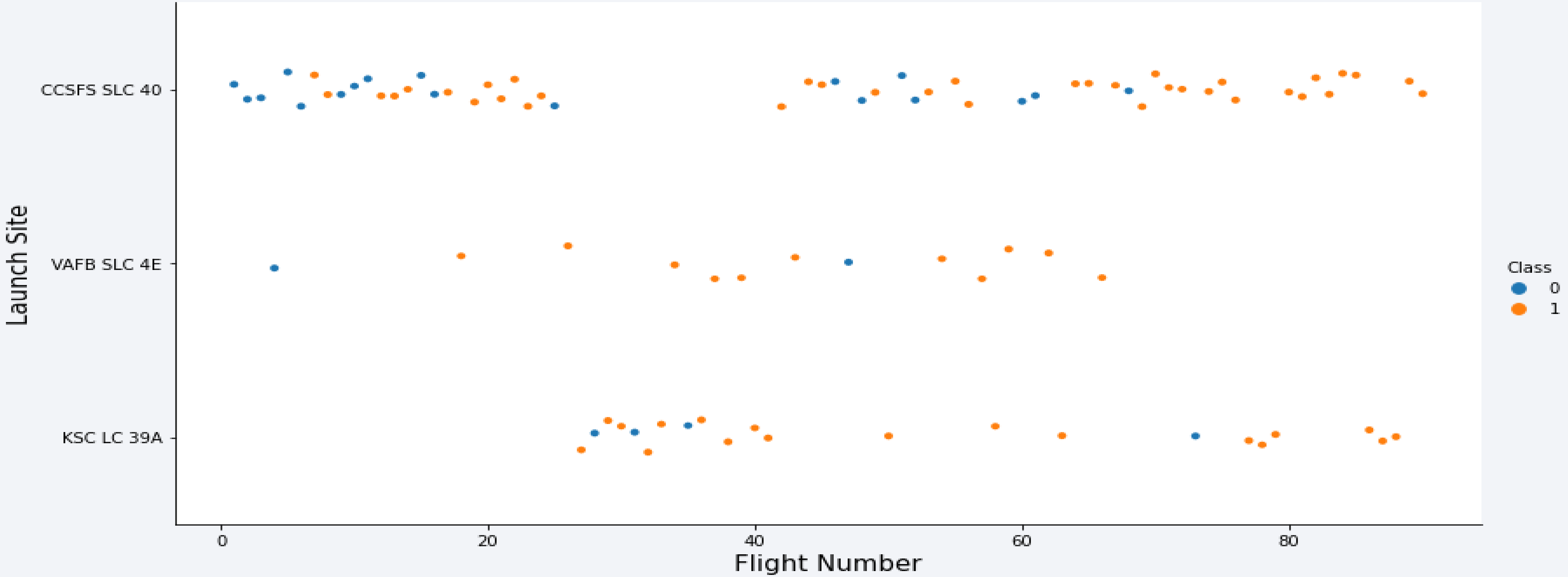
- **Predictive analysis results**

  Classification Model with Accuracy Score, Confusion Matrix

Section 2

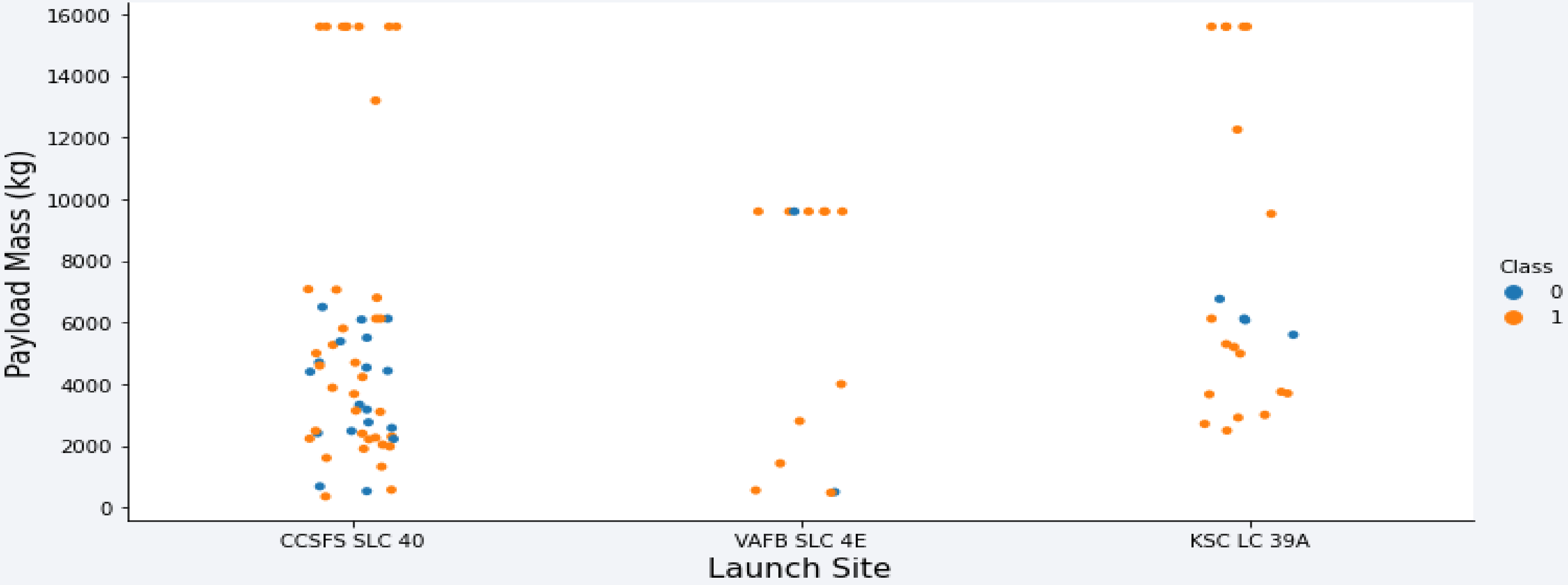
Insights drawn from EDA

Flight Number vs. Launch Site



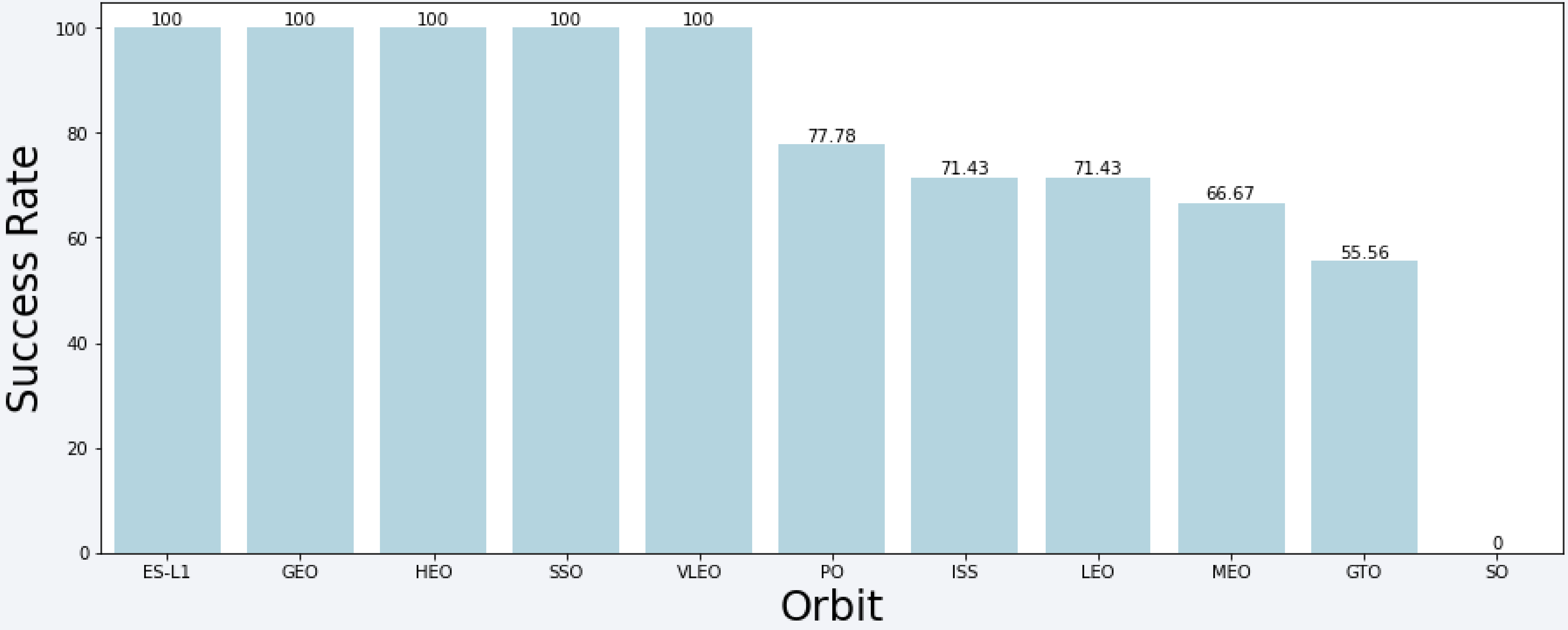
- Recent Landing Outcomes for Flights were successful as boosters safely landed back to earth.
- SpaceX used CCSFS SLC-40 and KSC LC-39A for Recent Launches.

Payload vs. Launch Site



CCSFS SLC-40 and KSC LC-39A launch sites were used for launch of spacecraft carried more than 10000 kg payload mass.

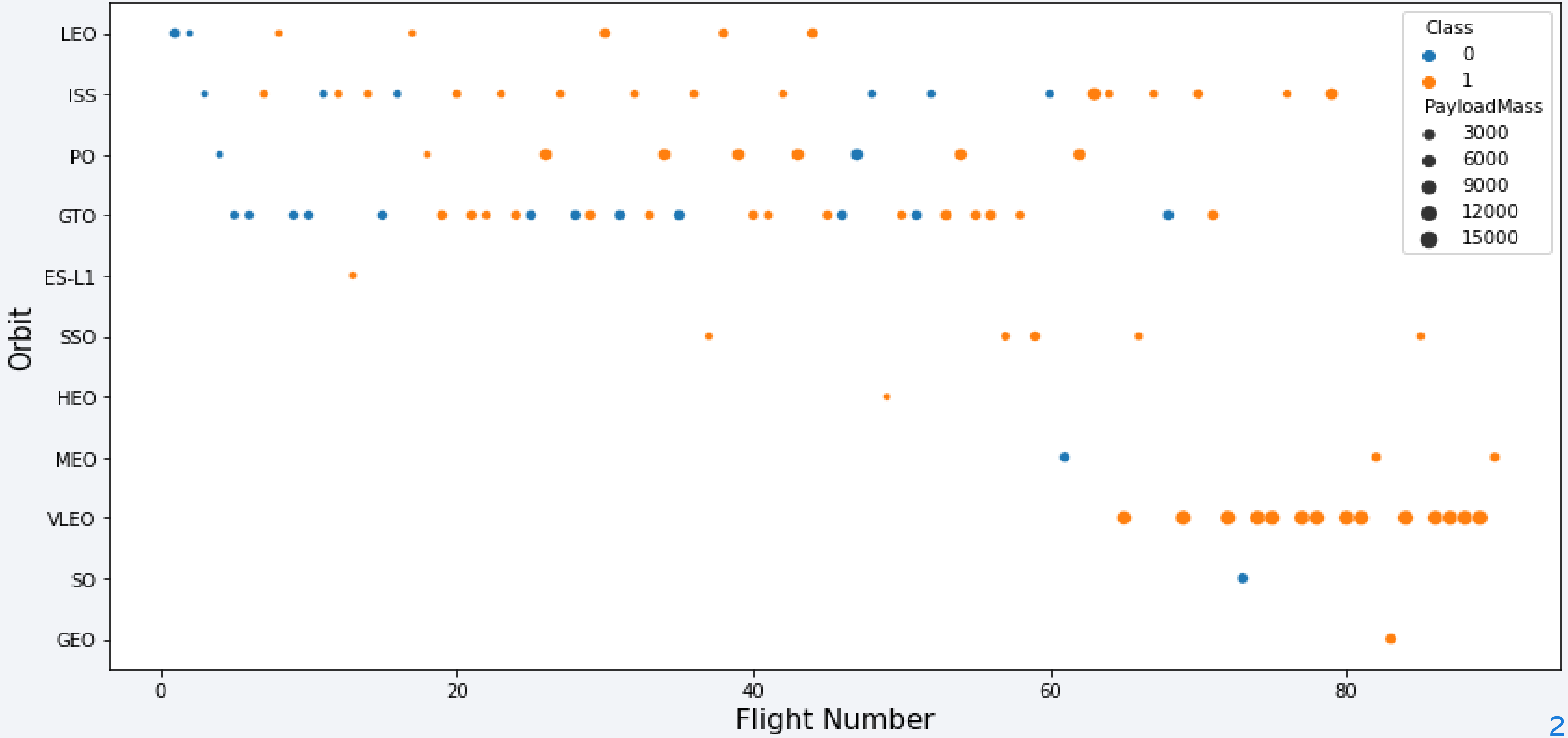
Success Rate vs. Orbit Type



SpaceX has 100 percentage Success Rate for placing Satellites in

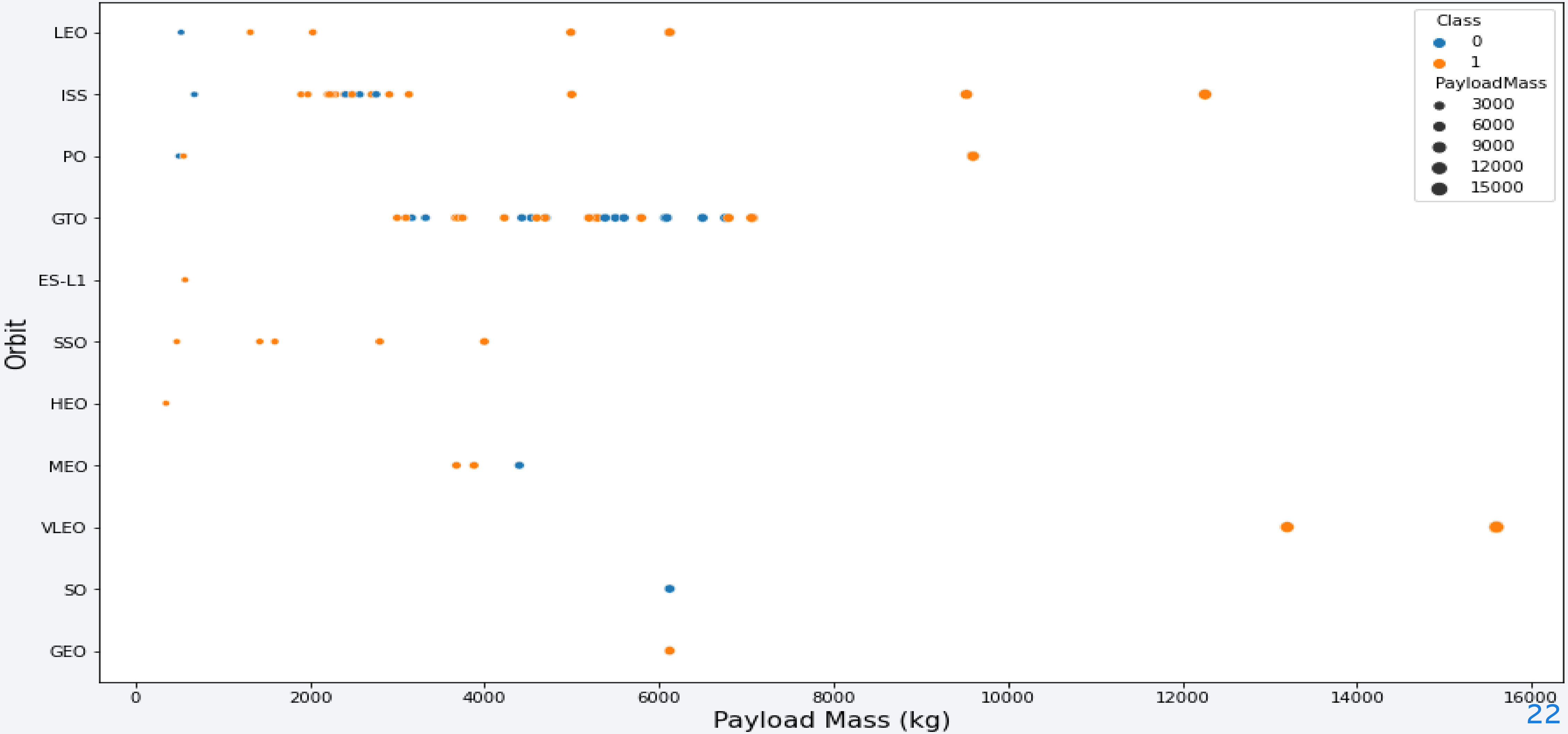
ES-L1, GEO, HEO, SSO and VLEO Orbits

Flight Number vs. Orbit Type



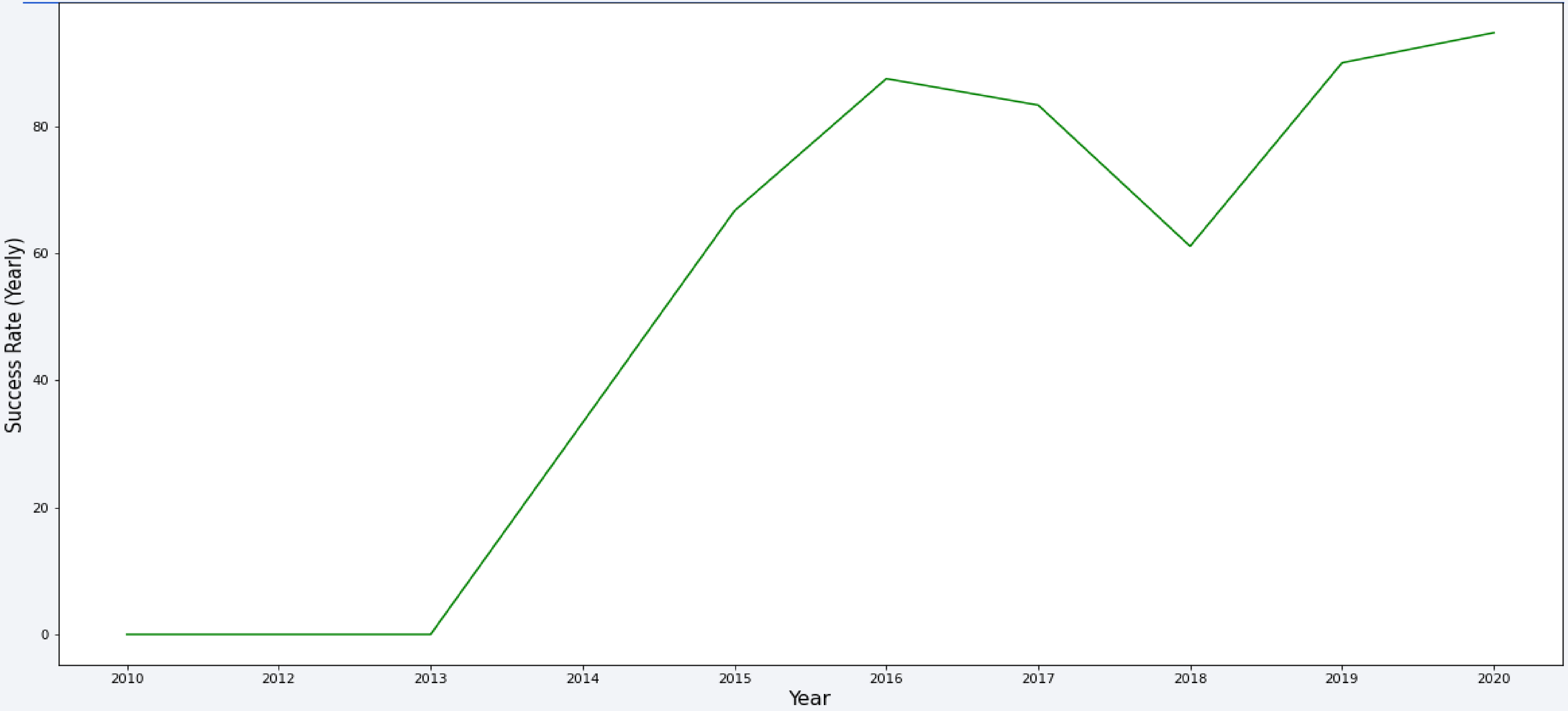
VLEO Orbit is very popular as SpaceX has placed many heaviest satellites in recent flights

Payload vs. Orbit Type



Heaviest Satellite is placed in VLEO Orbit

Launch Success Yearly Trend



Succes Rate increased sharply after year 2013

All Launch Site Names

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

%sql SELECT DISTINCT(Launch_Site) FROM SPACEXTBL

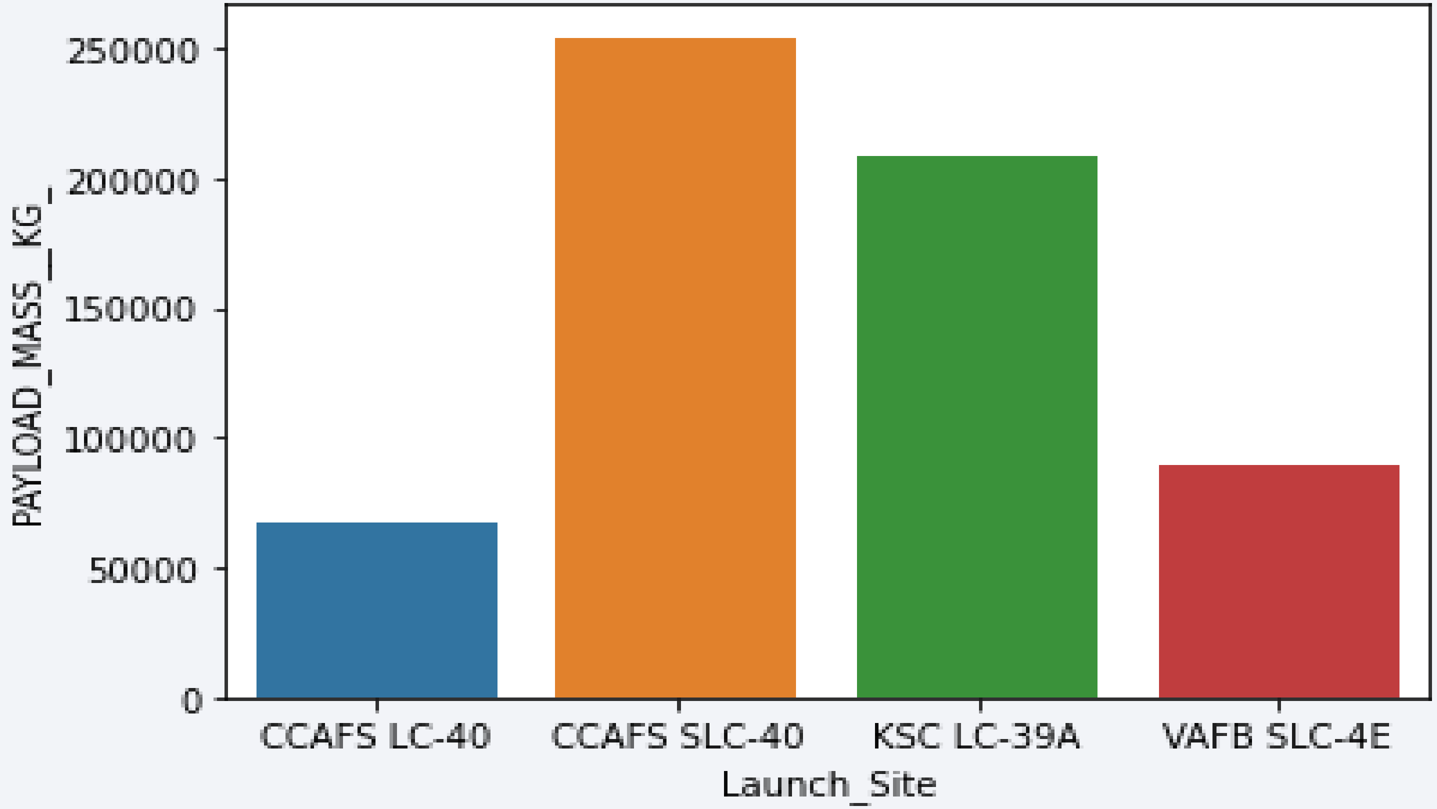
Launch Site Names Begin with 'CCA'

Launch_Site
CCAFS LC-40
CCAFS SLC-40

%sql SELECT * FROM SPACEXTBL WHERE Launch_Site LIKE 'CCA%' GROUP BY Launch_Site LIMIT 5

Total Payload Mass

Total_Payload_Mass (kg)
619967

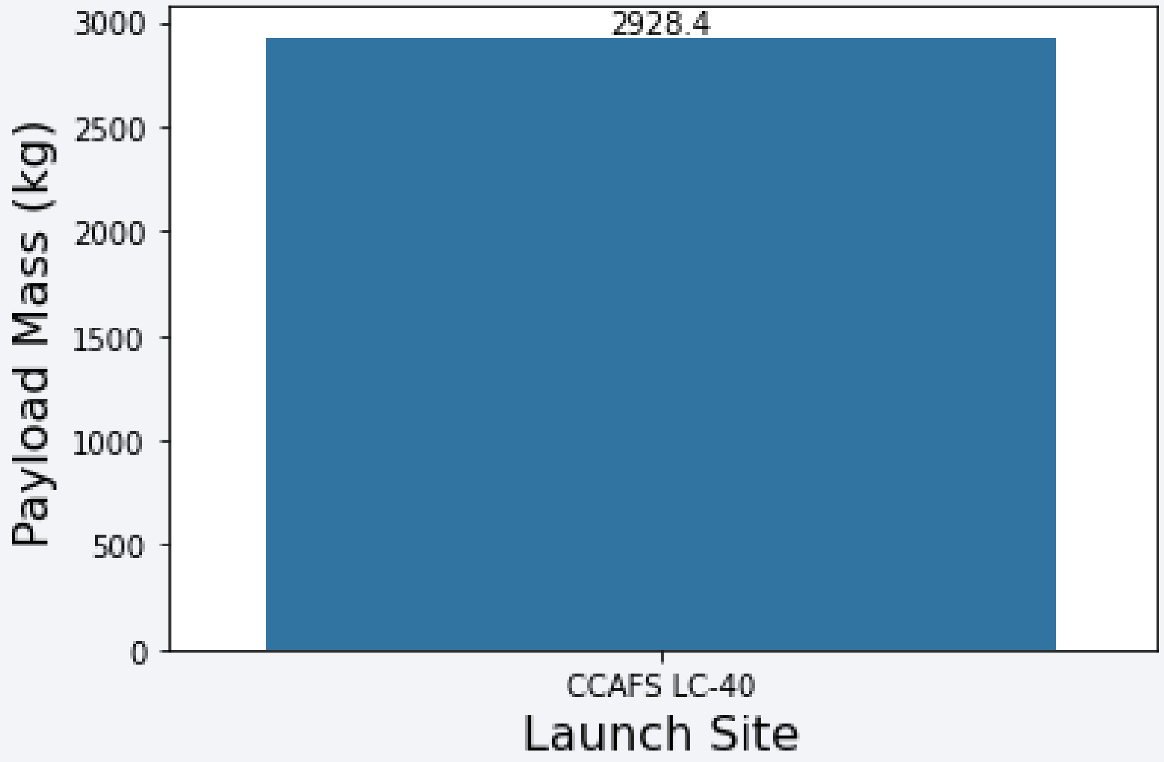


%sql SELECT SUM(PAYLOAD_MASS__KG_) As Total_Payload_Mass (kg)

FROM SPACEXTBL

Average Payload Mass by F9 v1.1

Average_Payload_Mass_(kg)
2928.4



%sql SELECT AVG(PAYLOAD_MASS__KG_) AS Average_Payload FROM SPACEXTBL WHERE Booster_Version == 'F9 v1.1'

First Successful Ground Landing Date

Date	Time_UTC	Booster_Version	Launch_Site	Landing_Outcome
22-12-2015	01:29:00	F9 FT B1019	CCAFS LC-40	Success (ground pad)

```
%sql SELECT Date, Time_UTC, Booster_Version, Launch_Site, Landing_Outcome
FROM SPACEXTBL WHERE Landing_Outcome == 'Success (ground pad)'
ORDER BY Date DESC LIMIT 1;
```

Successful Drone Ship Landing with Payload between 4000 and 6000

Date	Time_UTC	Booster_Version	Launch_Site	Payload	Payload_Mass_KG	Orbit	Customer	Mission_Outcome	Landing_Outcome
06-05-2016	05:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
14-08-2016	05:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
30-03-2017	22:27:00	F9 FT B1021.2	KSC LC-39A	SES-10	5300	GTO	SES	Success	Success (drone ship)
11-10-2017	22:53:00	F9 FT B1031.2	KSC LC-39A	SES-11 / EchoStar 105	5200	GTO	SES EchoStar	Success	Success (drone ship)

```
%sql SELECT * FROM SPACEXTBL
WHERE Landing_Outcome == "Success (drone ship)" AND
Payload_Mass_KG BETWEEN 4000 AND 9000
```

Total Number of Successful and Failure Mission Outcomes

Mission_Outcome	Flights
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

%sql SELECT DISTINCT(Mission_Outcome), COUNT(*) AS Flights

FROM SPACEXTBL GROUP BY Mission_Outcome;

Boosters Carried Maximum Payload

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

```
%sql SELECT DISTINCT(Booster_Version), Payload_Mass_KG FROM SPACEXTBL
WHERE Payload_Mass_KG = (SELECT MAX(Payload_Mass_KG) FROM SPACEXTBL)
```

2015 Launch Records ' Failed Landing Outcomes (Drone Ship) '

Date	Booster_Version	Launch_Site	Landing_Outcome
10-01-2015	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
14-04-2015	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

```
%sql SELECT Date, Booster_Version, Launch_Site, Landing_Outcome FROM SPACEXTBL
WHERE Landing_Outcome == 'Failure (drone ship)' AND SUBSTRING(Date,7,4) == '2015'
```

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

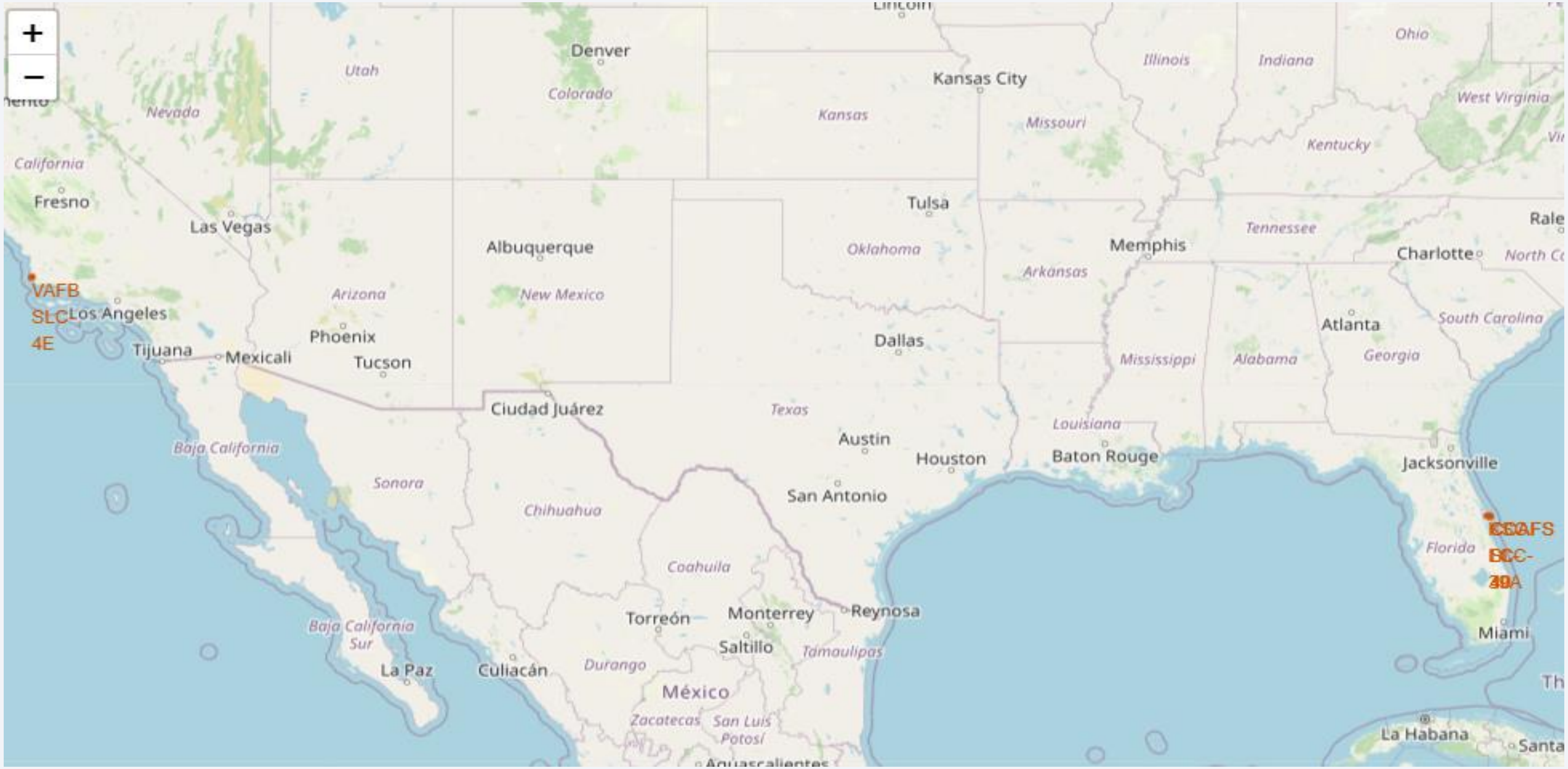
Date	Landing_Outcome	Flights
2012-05-22	No attempt	10
2015-12-22	Success (ground pad)	5
2016-08-04	Success (drone ship)	5
2015-10-01	Failure (drone ship)	5
2014-04-18	Controlled (ocean)	3
2013-09-29	Uncontrolled (ocean)	2
2015-06-28	Precluded (drone ship)	1
2010-08-12	Failure (parachute)	1

%sql SELECT Date, Landing_Outcome, COUNT(Landing_Outcome) AS Flights FROM SPACEXTBL WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY Landing_Outcome ORDER BY Flights DESC;

Section 3

Launch Sites Proximities Analysis

SpaceX Launch Sites Folium Map

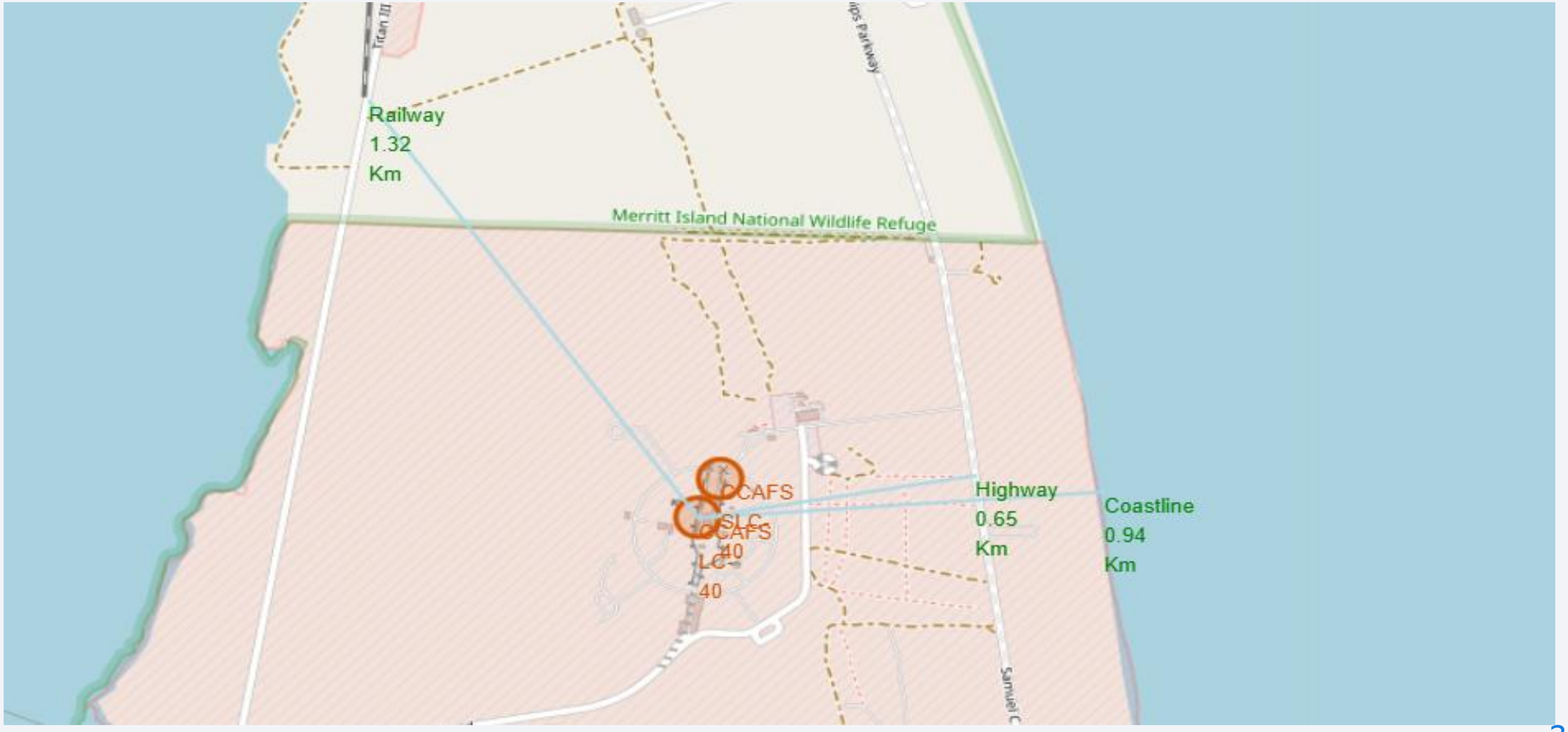


SpaceX Launch Sites Landing Outcomes

Successful Flight

Unsuccessful

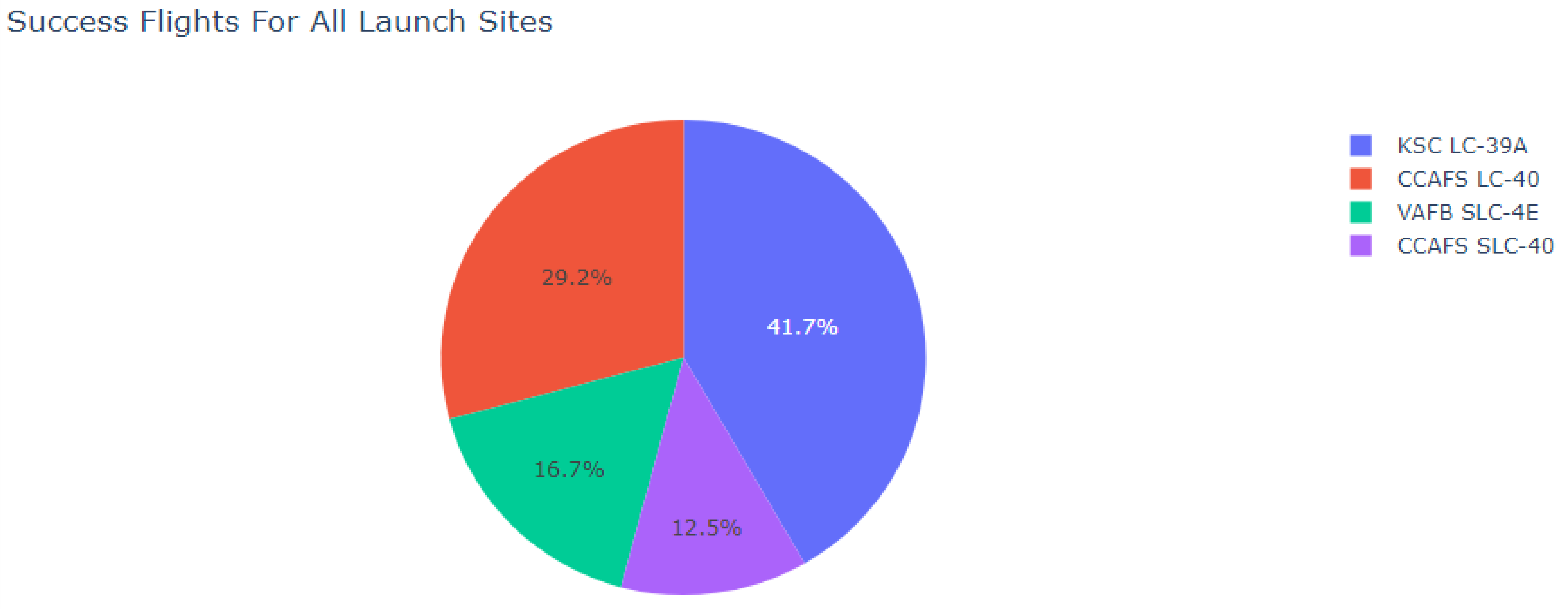
Distances from Launch Site to Important Locations



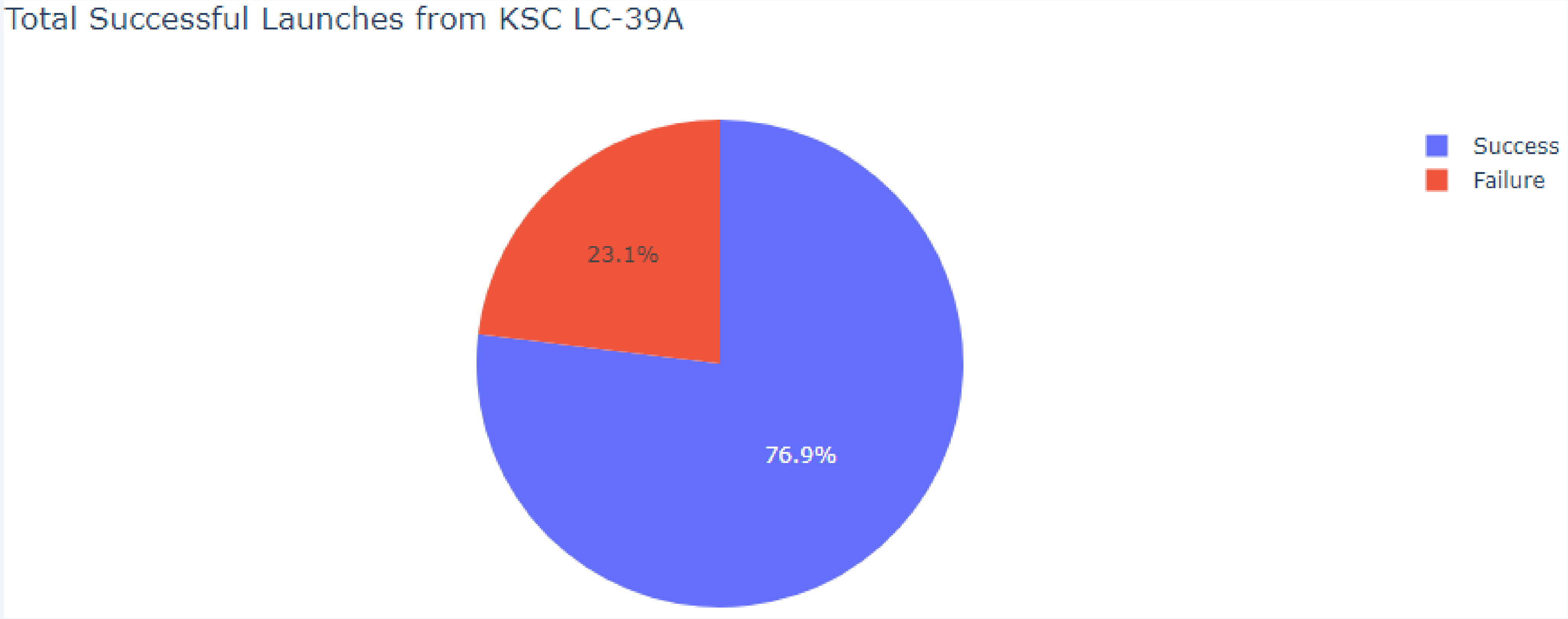
Section 4

Build a Dashboard with Plotly Dash

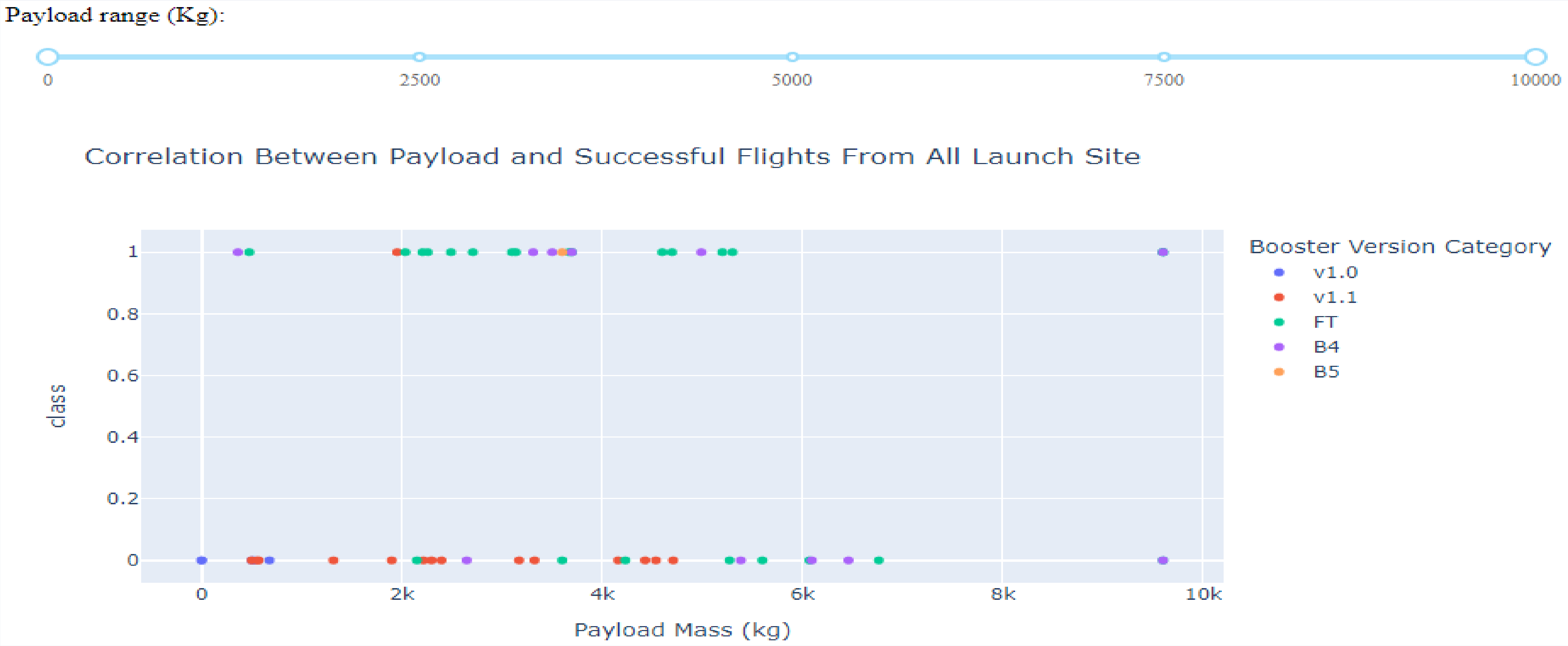
SpaceX Successful Launch Records



Most Successful Launch Site KSC LC-39A



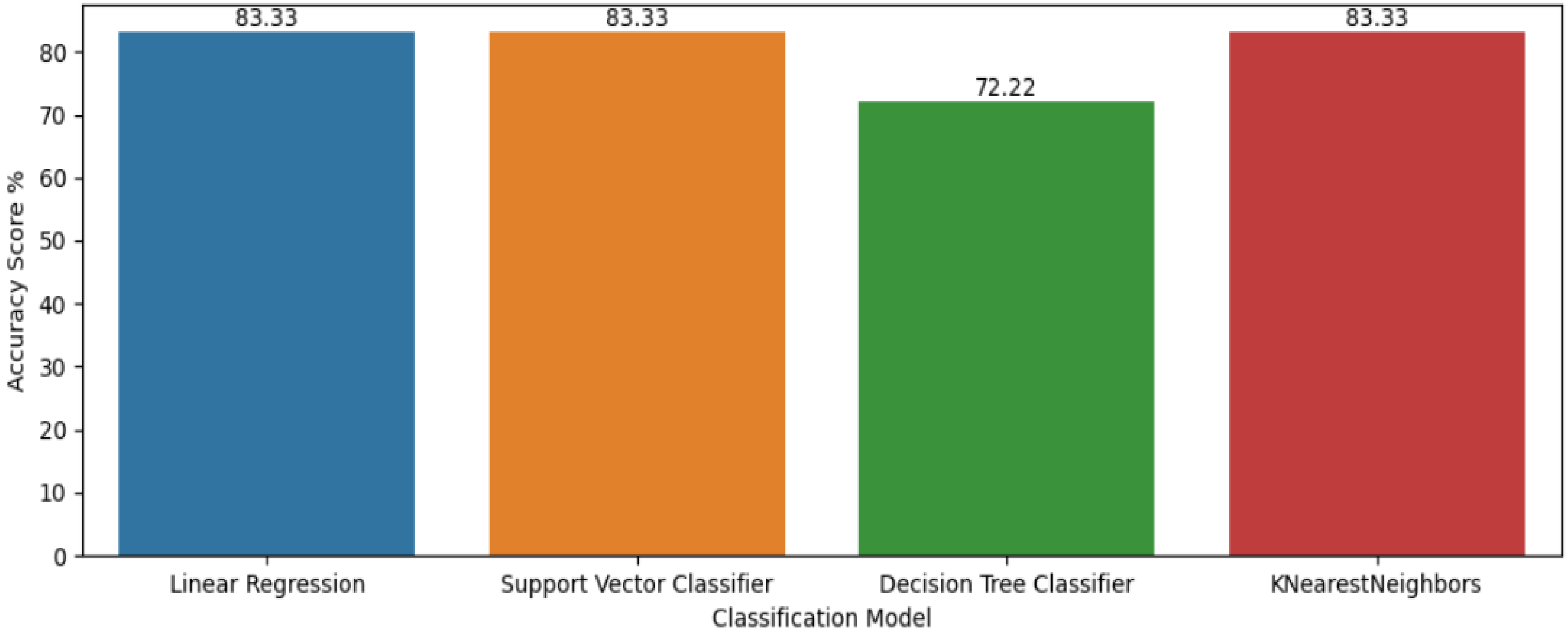
Payload vs. Launch Outcome



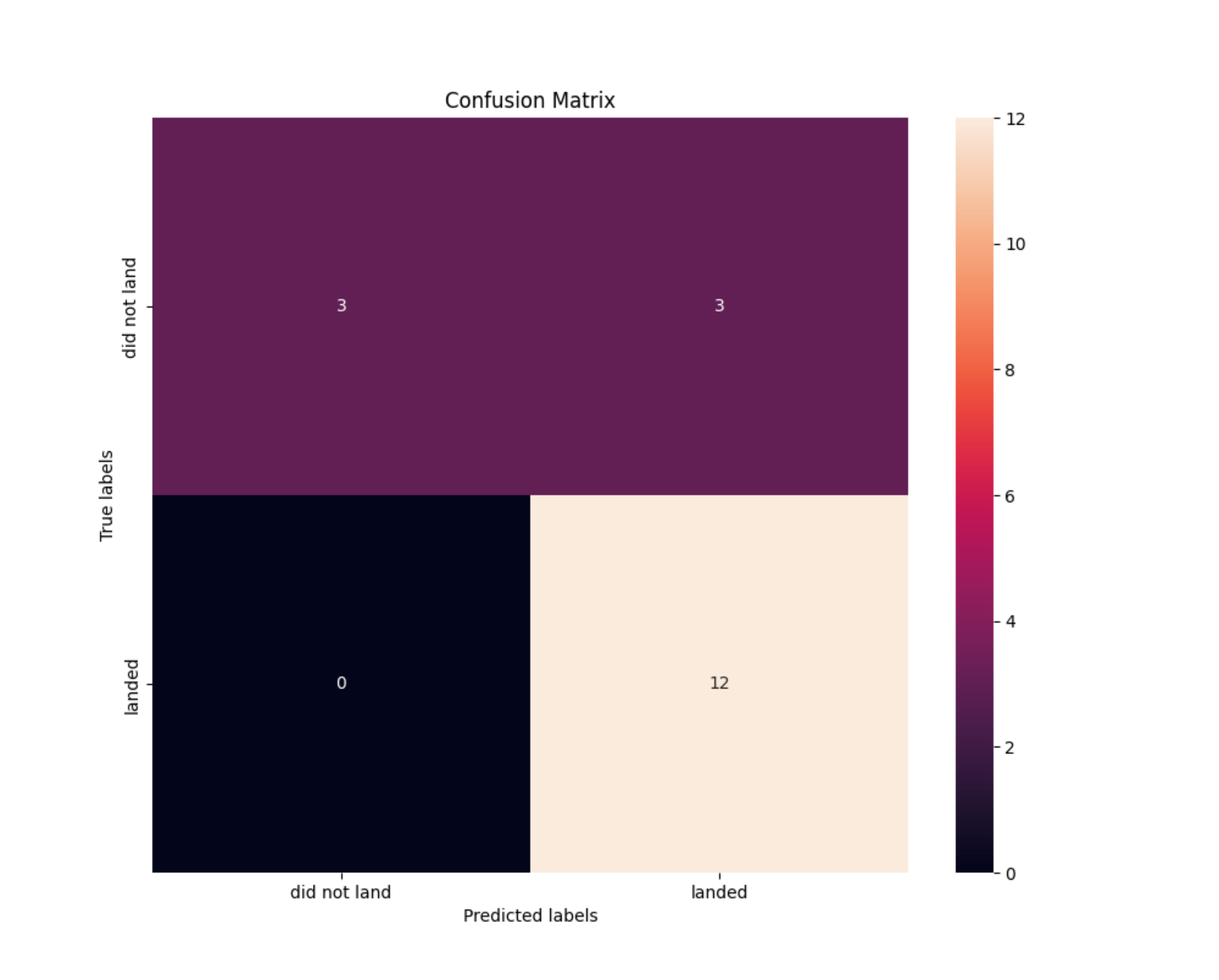
Section 5

Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix (Best Classification Model)



Conclusions

- Launch Site with Most Flights is CCSFS SLC-40 (**60 Flights**)
- Most Successful Flights Launch Site is KSC LC-39A (**76.9% Success Rate**)
- Launch Success Rate has increased Since **2013**.
- SpaceX has 100 percentage Success Rate for placing Satellites in

  ES-L1, GEO, HEO, SSO and VLEO

- Classification Model With Highest Achieved Accuracy Score (**83.33%**)

  Linear Regression, Support Vector Classifier and KNearestNeighbors

Appendix

SpaceX Website - https://www.spacex.com/

SpaceX Rest API - https://api.spacexdata.com/v4/launchpads/

SpaceX Wikipedia - https://en.wikipedia.org/wiki/SpaceX

GitHub Repository - https://github.com/JagdeepMaan/JagdeepMaan

Thank you!