IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Weronika Bednarska
29.07.2022

Outline

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

Executive Summary

- Summary of methodologies
  - Collection the data
  - Data wrangling
  - EDA with SQL and with data visualization
  - Interactive map
  - Dashboard in plotly
  - Predict
- Summary of all results
  - EDA
  - analysis

Introduction

- The project background and context
  - Examine landings of rackets Falcon 9. Since 2010 Falcon's has 98.8% mission full success. It have been launched 170 times over 12 years, resulting in 168.
- Problems you want to find answers
  - The project answer is to find the answer if racket Falcon 9 will land successfully

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data have been collected SpaceX REST Api v4
- Perform data wrangling
  - Add column with successful or not successful landing outcomes
- 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

- Data set collect with SpaceX REST API v4:

  api.spacexdata.com/v4/

- Data scrapping with BeautyfulSoup library directly from wikipedia website

Data Collection – SpaceX API

- SpaceX REST calls

GitHub URL:

https://github.com/guzweronika/IBM_Data_Science_Final_Project/blob/main/1b_jupyter-labs-spacex-data-collection-api.ipynb

Get response from SpaceX REST API ¶

```
spacex_url="https://api.spacexdata.com/v4/launches/past"
```

```
response = requests.get(spacex_url)
```

Convert JSON response to data frame

```
json = response.json()
data = pd.json_normalize(json)
```

Create data frame from dictionary after wrangled data

```
launch_dict = {'FlightNumber': list(data['flight_number']),
'Date': list(data['date']),
'BoosterVersion':BoosterVersion,
'PayloadMass':PayloadMass,
'Orbit':Orbit,
'LaunchSite':LaunchSite,
'Outcome':Outcome,
'Flights':Flights,
'GridFins':GridFins,
'Reused':Reused,
'Legs':Legs,
'LandingPad':LandingPad,
'Block':Block,
'ReusedCount':ReusedCount,
'Serial':Serial,
'Longitude': Longitude,
'Latitude': Latitude}
```

Data frame only with Falcon 9 racket

```
data_falcon9 = df[df['BoosterVersion']!='Falcon 1']
```

Data Collection - Scraping

- Web scraping with BeautifulSoup from wikipedia

- GitHub URL:

  https://github.com/guzweronika/IBM_Data_Science_Final_Project/blob/main/1a_jupyter-labs-webscraping.ipynb

Get response and create BeaytifulSoup object

```python
response = requests.get(static_url)
soup = BeautifulSoup(response.content, 'html5lib')
```

Get tables from BeaytifulSoup object

```python
# Use the find_all function in the BeautifulSoup object, with element type `table`
html_tables = soup.find_all('table')
```

Create dict with table columns

```python
launch_dict= dict.fromkeys(column_names)

# Remove an irrelvant column
del launch_dict['Date and time ( )']

# Let's initial the launch_dict with each value to be an empty list
launch_dict['Flight No.'] = []
launch_dict['Launch site'] = []
launch_dict['Payload'] = []
launch_dict['Payload mass'] = []
launch_dict['Orbit'] = []
launch_dict['Customer'] = []
launch_dict['Launch outcome'] = []
# Added some new columns
launch_dict['Version Booster']=[]
launch_dict['Booster landing']=[]
launch_dict['Date']=[]
launch_dict['Time']=[]
```

Fill the dict value create a data frame

```python
extracted_row = 0
#Extract each table
for table_number,table in enumerate(soup.find_all('table',"wikitable plainrowheaders collapsible")):
   # get table row
    for rows in table.find_all("tr"):
        #check to see if first table heading is as number corresponding to launch a number
        if rows.th:
            if rows.th.string:
                flight_number=rows.th.string.strip()
                flag=flight_number.isdigit()
        else:
            flag=False
        #get table element
        row=rows.find_all('td')
        #if it is number save cells in a dictonary
        if flag:
            extracted_row += 1
            # Flight Number value
            # TODO: Append the flight_number into launch_dict with key `Flight No.`
            launch_dict['Flight No.'].append(flight_number)
            #print(flight_number)
            datatimelist=date_time(row[0])

            # Date value
            # TODO: Append the date into launch_dict with key `Date`
            date = datatimelist[0].strip(',')
            launch_dict['Date'].append(date)
            #print(date)

            # Time value
            # TODO: Append the time into launch_dict with key `Time`
            time = datatimelist[1]
            launch_dict['Time'].append(time)
            #print(time)
```

The cell has been shortened

```python
df=pd.DataFrame(launch_dict)
```

Data Wrangling

Check data state: null values and data types

Calculate the number of Outcomes mission

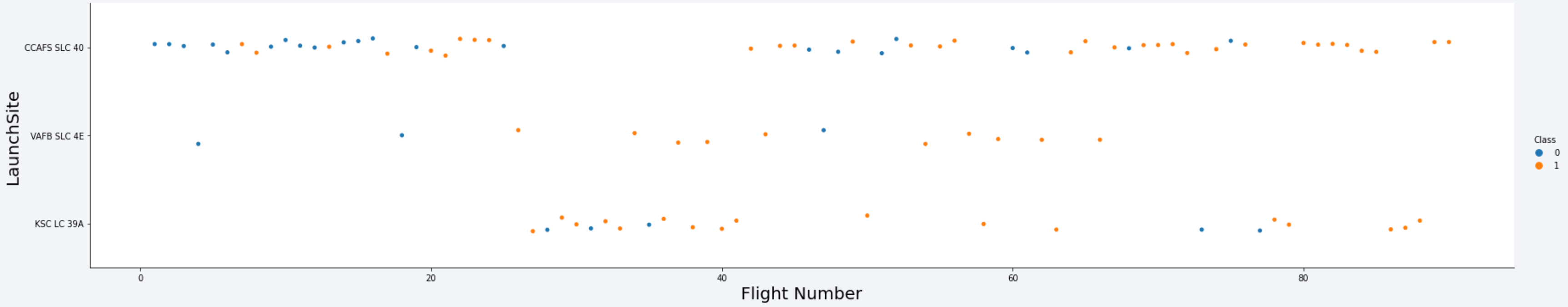
Create a list with unsuccesful landing values

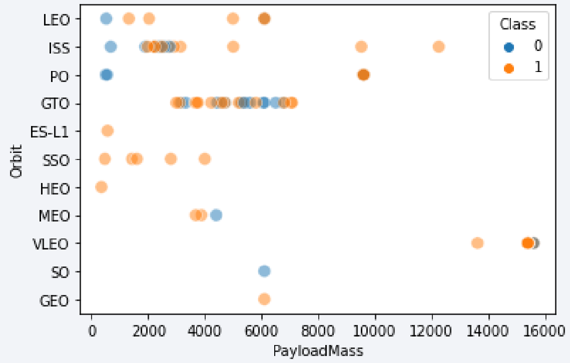
Create new feature with converted Outcome variable to successful or unsuccessful

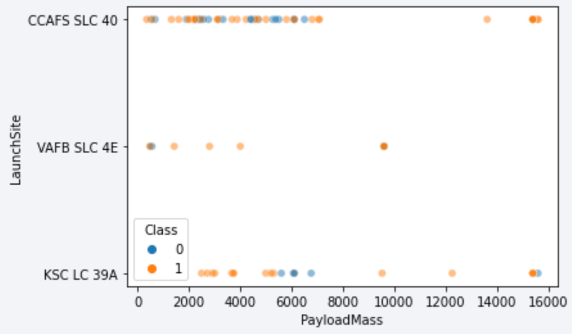
- GitHub URL:

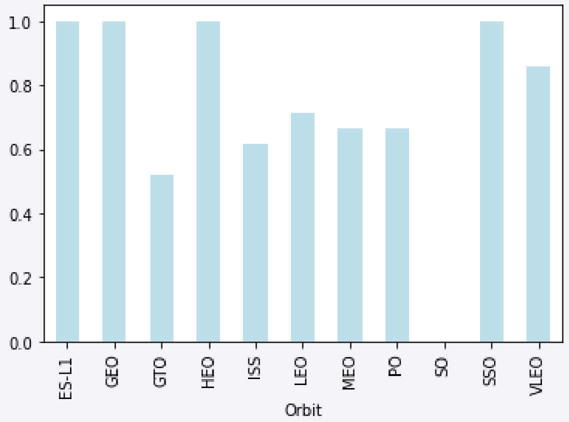
https://github.com/guzweronika/IBM_Data_Science_Final_Project/blob/main/2_labs-jupyter-spacex-Data%20wrangling.ipynb

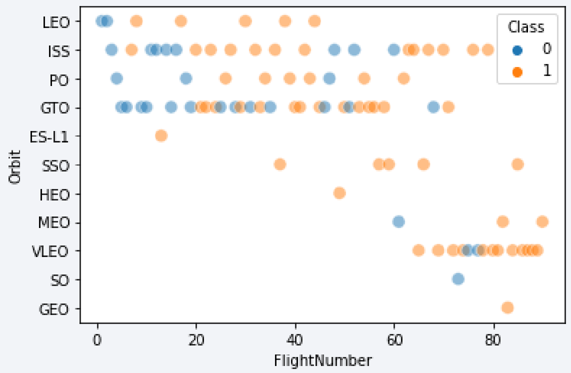
EDA with Data Visualization

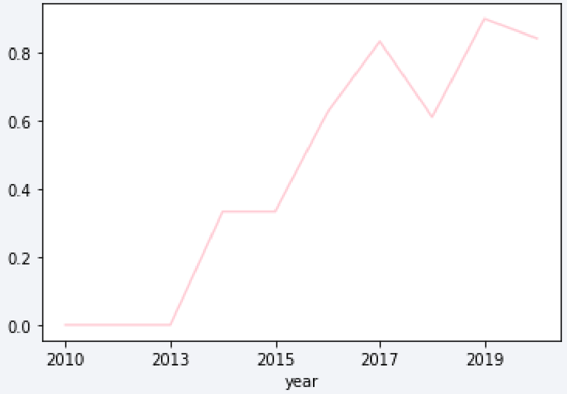












GitHub URL:

https://github.com/guzweronika/IBM_Data_Science_Final_Project/blob/main/3a_jupyter-labs-eda-dataviz.ipynb

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 acheived.
- 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.
- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad))

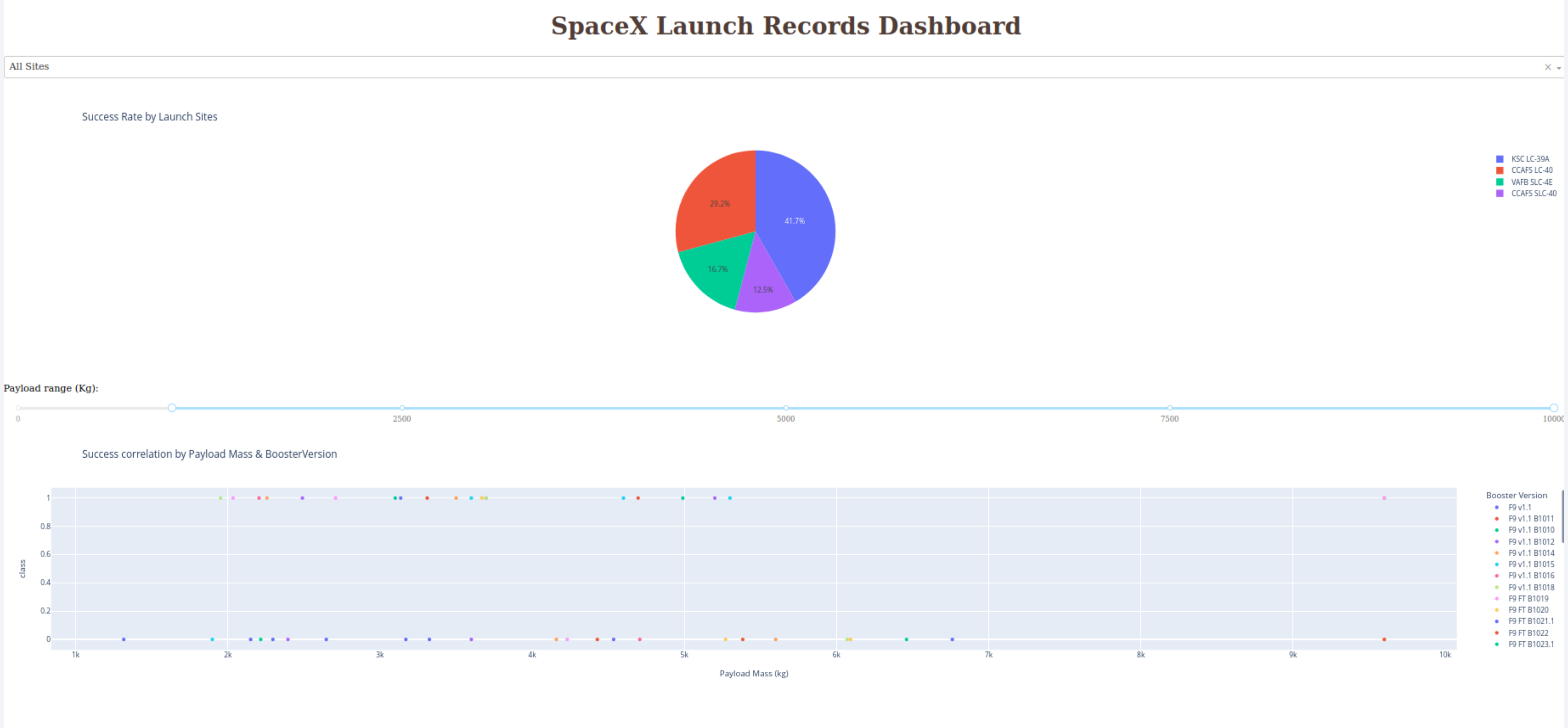
GitHub URL:
https://github.com/guzweronika/IBM_Data_Science_Final_Project/blob/main/3b_jupyter-labs-eda-sql-coursera.ipynb

Build an Interactive Map with Folium

- GitHub URL:

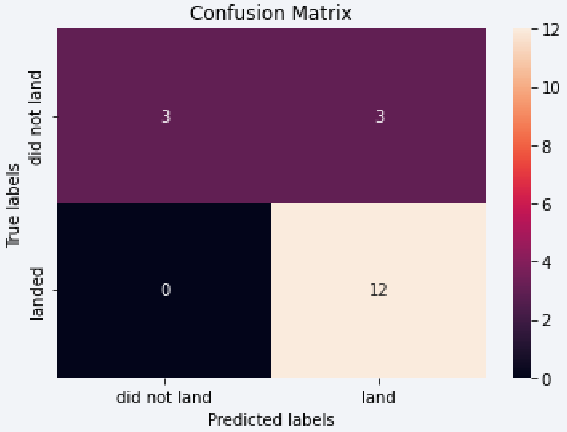
https://github.com/guzweronika/IBM_Data_Science_Final_Project/blob/main/4_lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash



https://github.com/guzweronika/IBM_Data_Science_Final_Project/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)




```
Accuracy for Logistics Regression method: 0.8333333333333334
Accuracy for Support Vector Machine method: 0.8333333333333334
Accuracy for Decision tree method: 0.8333333333333334
Accuracy for K nearsdt neighbors method: 0.8333333333333334
```

- Each of the models: Logistics Regression, Support Vector Machine, Decision tree, K nearst neighbors has the same accuracy
- GitHub URL: https://github.com/guzweronika/IBM_Data_Science_Final_Project/blob/main/5_SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results

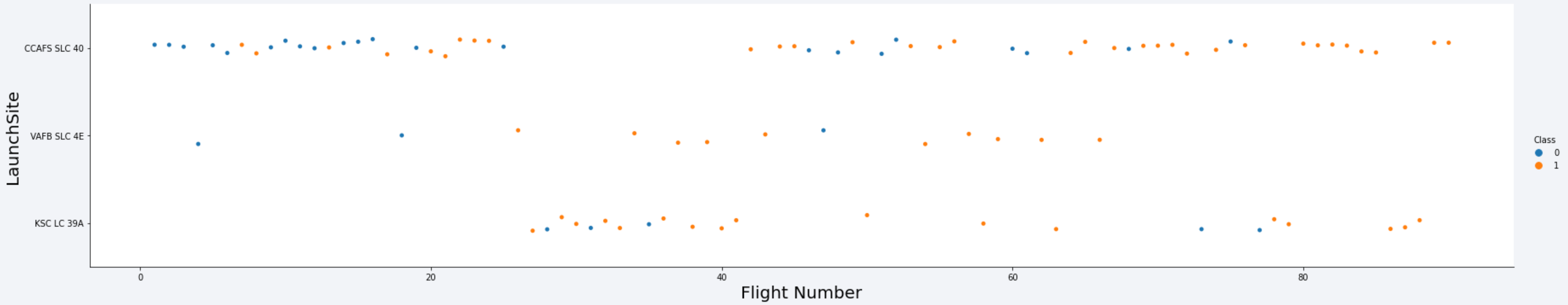
- Are Models have good accuracy to use them to prediction
- The score for each model is 83.33%

Section 2

Insights drawn from EDA

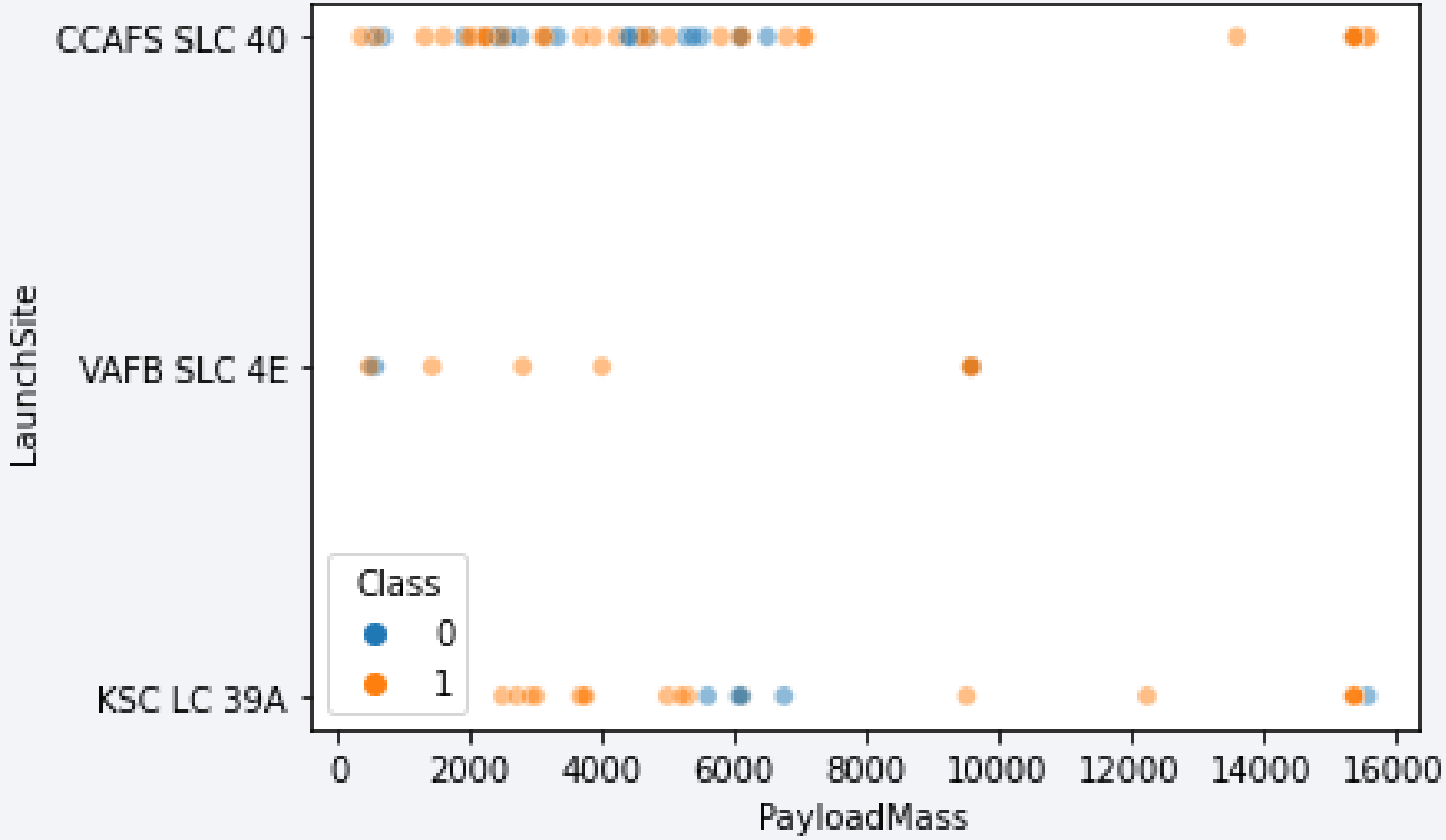
Flight Number vs. Launch Site

Launch Site CCAFS SLC 40 has the biggest amount of flights



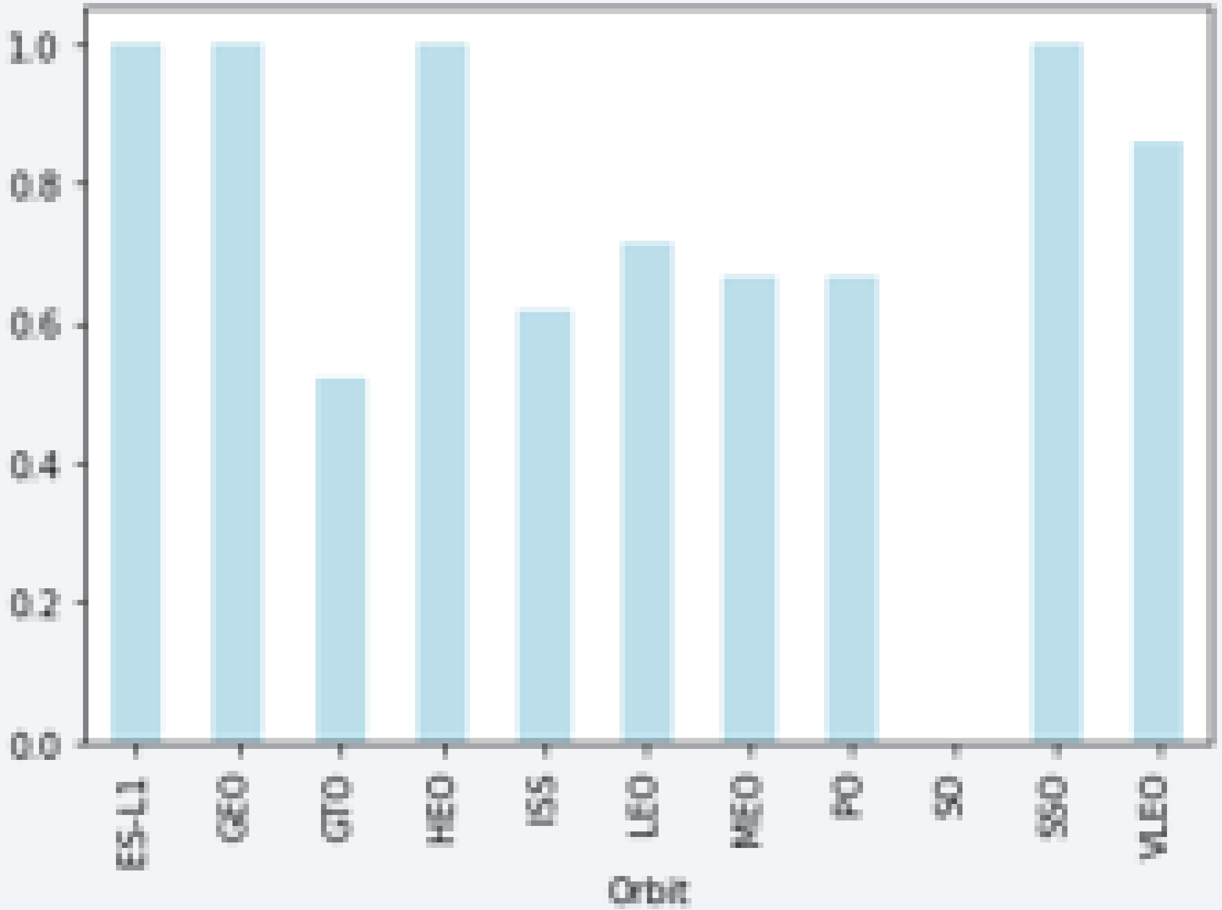
Payload vs. Launch Site

From Launch Site VAFB SLC 4E the Payload Mass never was bigger then 10,000



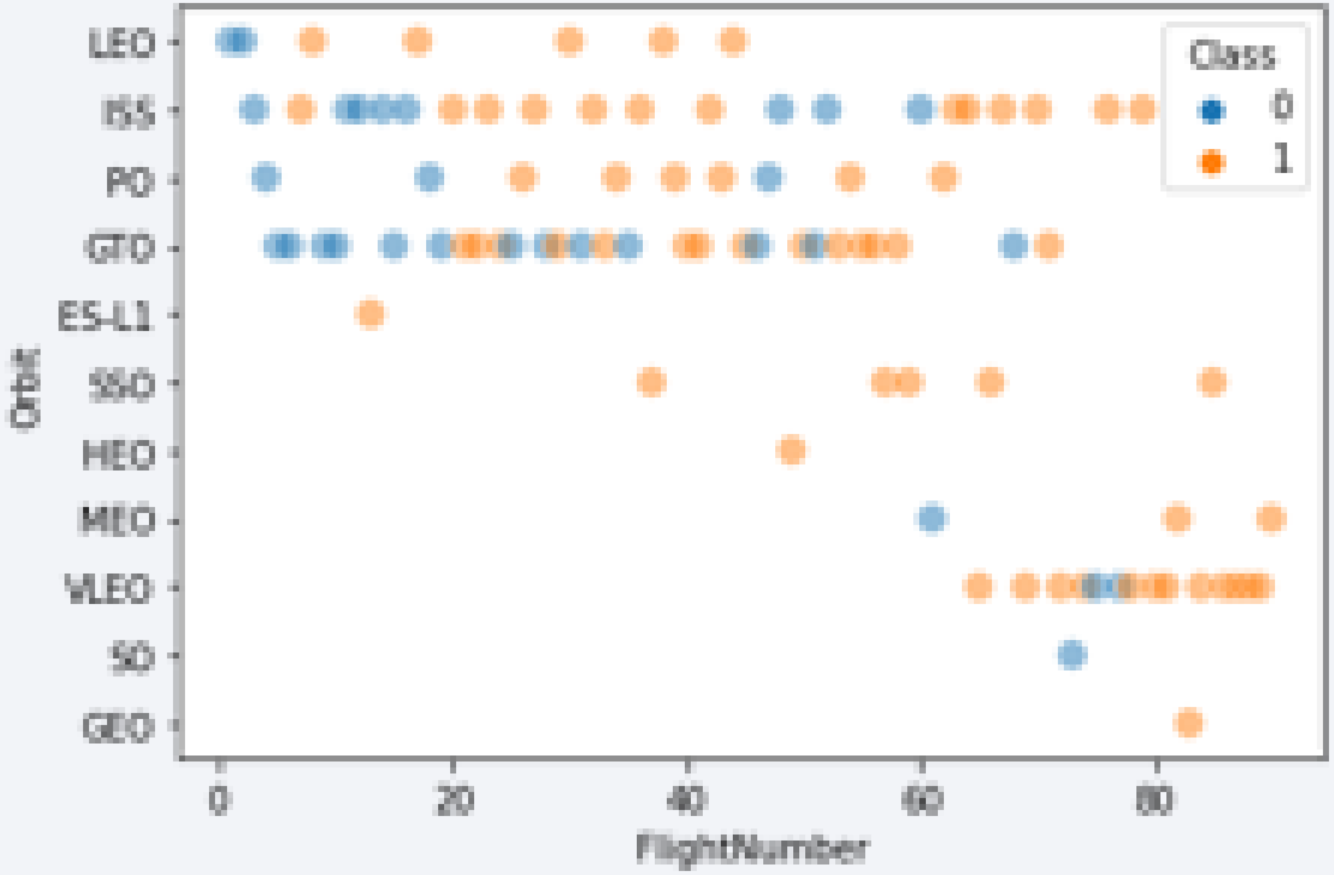
Success Rate vs. Orbit Type

Orbits ES-L1, CEO, HEO and SSO have the biggest success rate



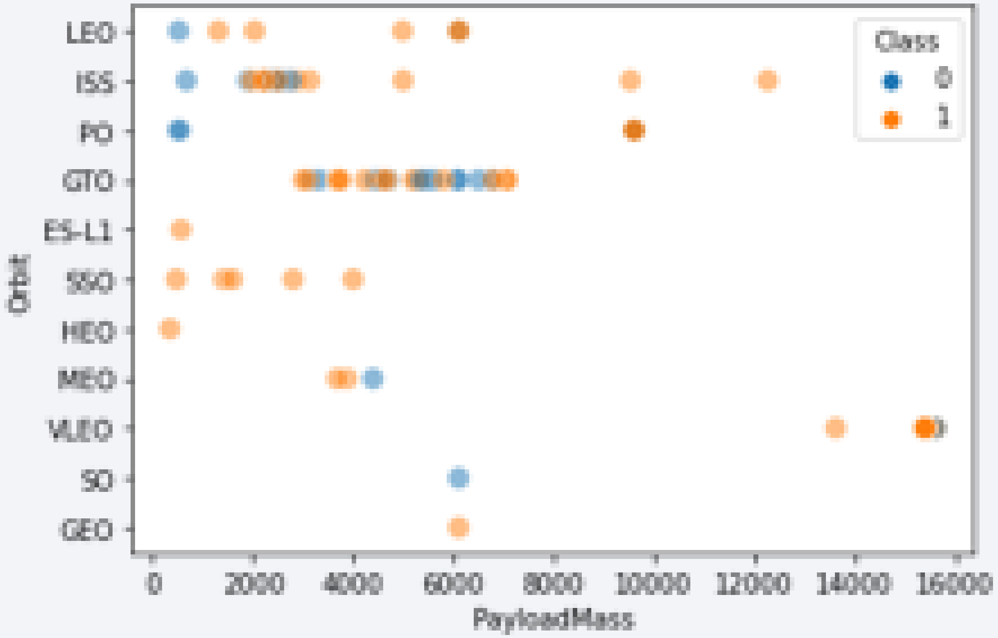
Flight Number vs. Orbit Type

Into orbit ISS flights have been taking place from the beginning until today



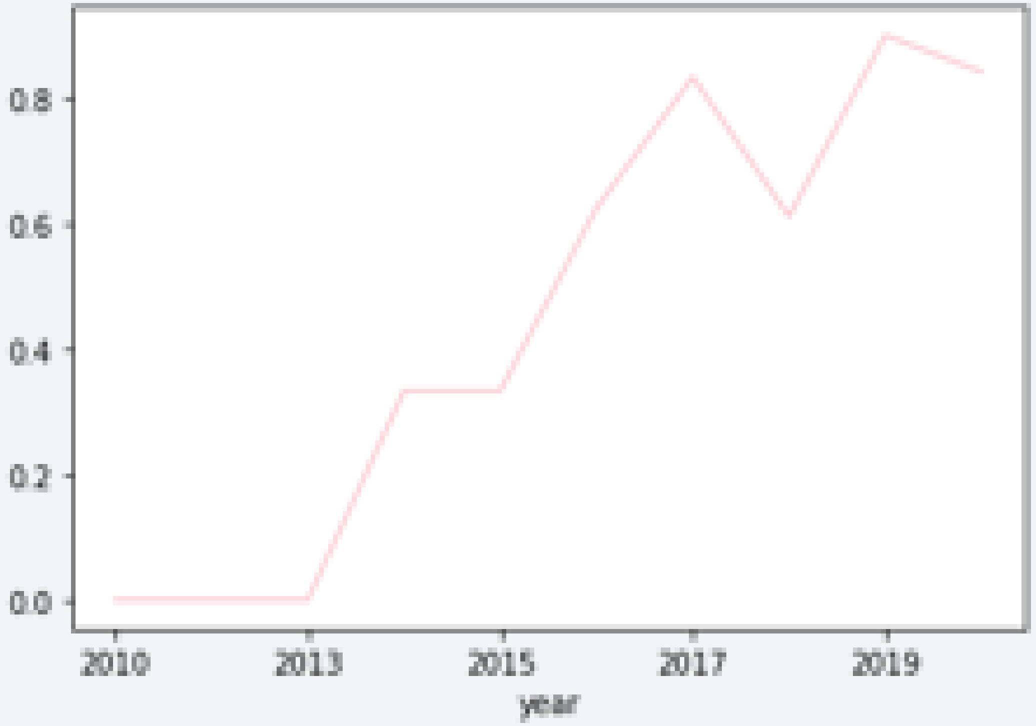
Payload vs. Orbit Type

- For orbit VUEO was the flights with the biggest payload mass



Launch Success Yearly Trend

- In the years 2010-2013 the flights were100% fails



All Launch Site Names

```
%sql SELECT DISTINCT(Launch_Site) FROM SPACEXTBL
```

```
 * sqlite:///my_db.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 SPACEXTBL WHERE Launch_Site LIKE 'CCA%' LIMIT 5
```

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	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)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	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 SPACEXTBL WHERE Customer = 'NASA (CRS)'
```

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

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

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

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

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

```
%sql SELECT Date FROM SPACEXTBL WHERE Landing_Outcome is 'Success (ground pad)' LIMIT 1
```

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

Date
22-12-2015

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%%sql
SELECT Booster_Version FROM SPACEXTBL
WHERE Landing_Outcome is 'Success (drone ship)' AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000
```

```
 * sqlite:///my_db.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

List the total number of successful and failure mission outcomes

```
%%sql
SELECT COUNT(Mission_Outcome) FROM SPACEXTBL
WHERE Mission_Outcome lIKE 'Failure (in flight)' OR Mission_Outcome LIKE 'Success'
```

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

COUNT(Mission_Outcome)
99

Boosters Carried Maximum Payload

```
%%sql
SELECT Booster_Version FROM SPACEXTBL
WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL )
```

```
 * sqlite:///my_db.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 Booster_Version, Launch_Site FROM SPACEXTBL
WHERE Landing_Outcome IS 'Failure (drone ship)' AND Date LIKE "%2015"
```

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

Booster_Version	Launch_Site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

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

```
%%sql
SELECT Booster_Version, Launch_Site FROM SPACEXTBL
WHERE Landing_Outcome IS 'Failure (drone ship)' OR Landing_Outcome IS 'Success (ground pad)'
```

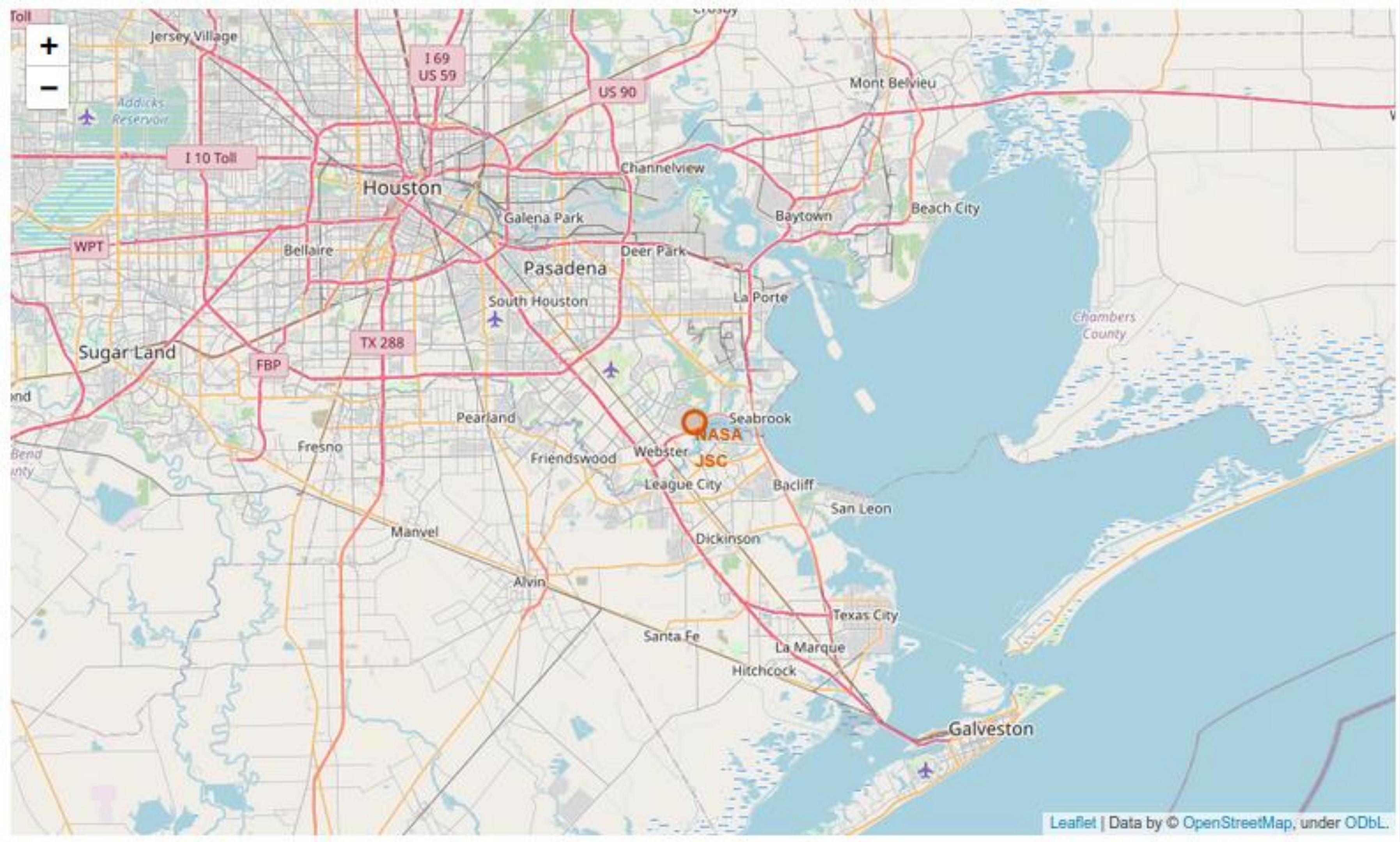
```
 * sqlite:///my_db.db
Done.
```

Booster_Version	Launch_Site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40
F9 FT B1019	CCAFS LC-40
F9 v1.1 B1017	VAFB SLC-4E
F9 FT B1020	CCAFS LC-40
F9 FT B1024	CCAFS LC-40
F9 FT B1025.1	CCAFS LC-40
F9 FT B1031.1	KSC LC-39A
F9 FT B1032.1	KSC LC-39A
F9 FT B1035.1	KSC LC-39A
F9 B4 B1039.1	KSC LC-39A
F9 B4 B1040.1	KSC LC-39A
F9 FT B1035.2	CCAFS SLC-40
F9 B4 B1043.1	CCAFS SLC-40

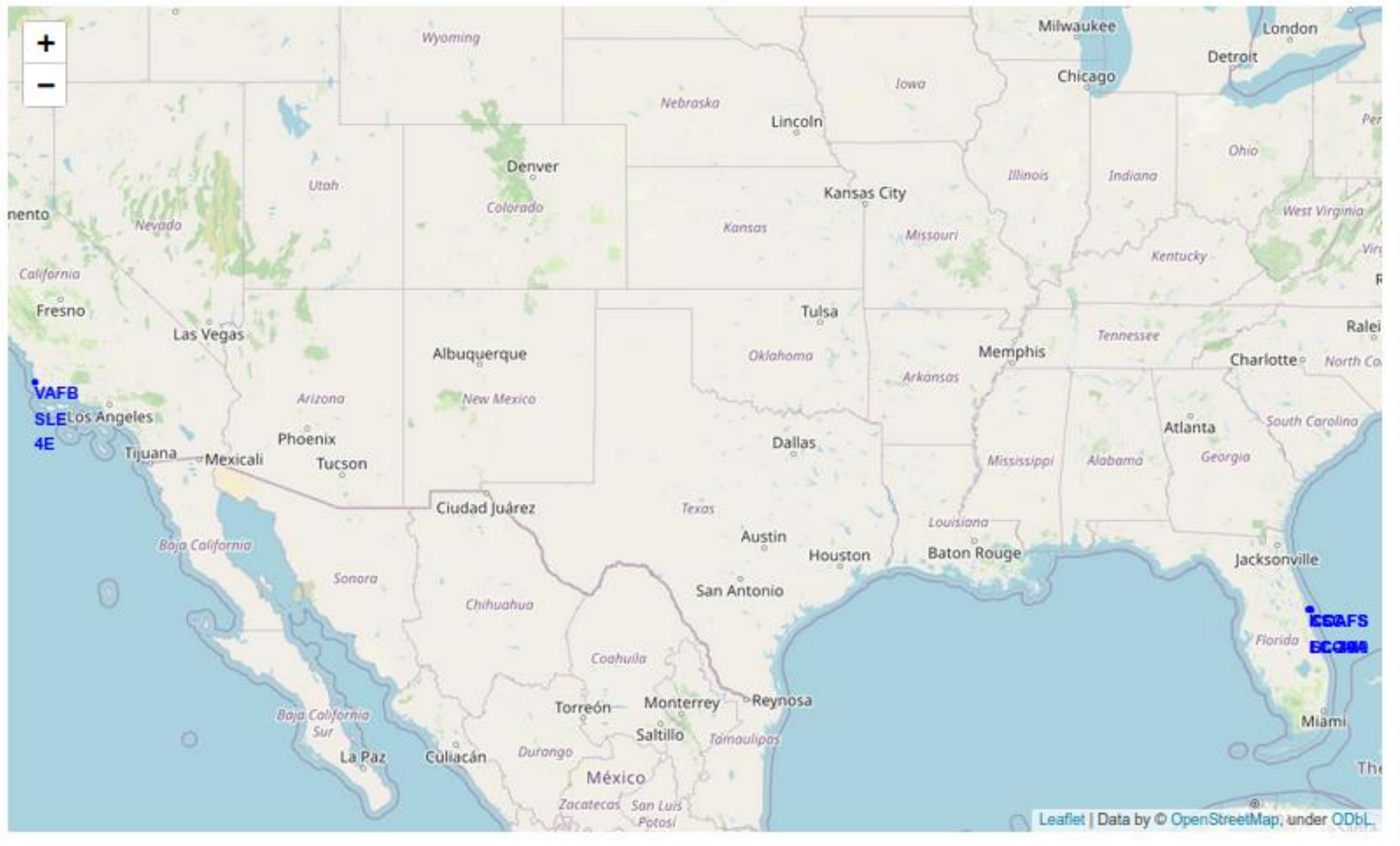
Section 3

Launch Sites Proximities Analysis

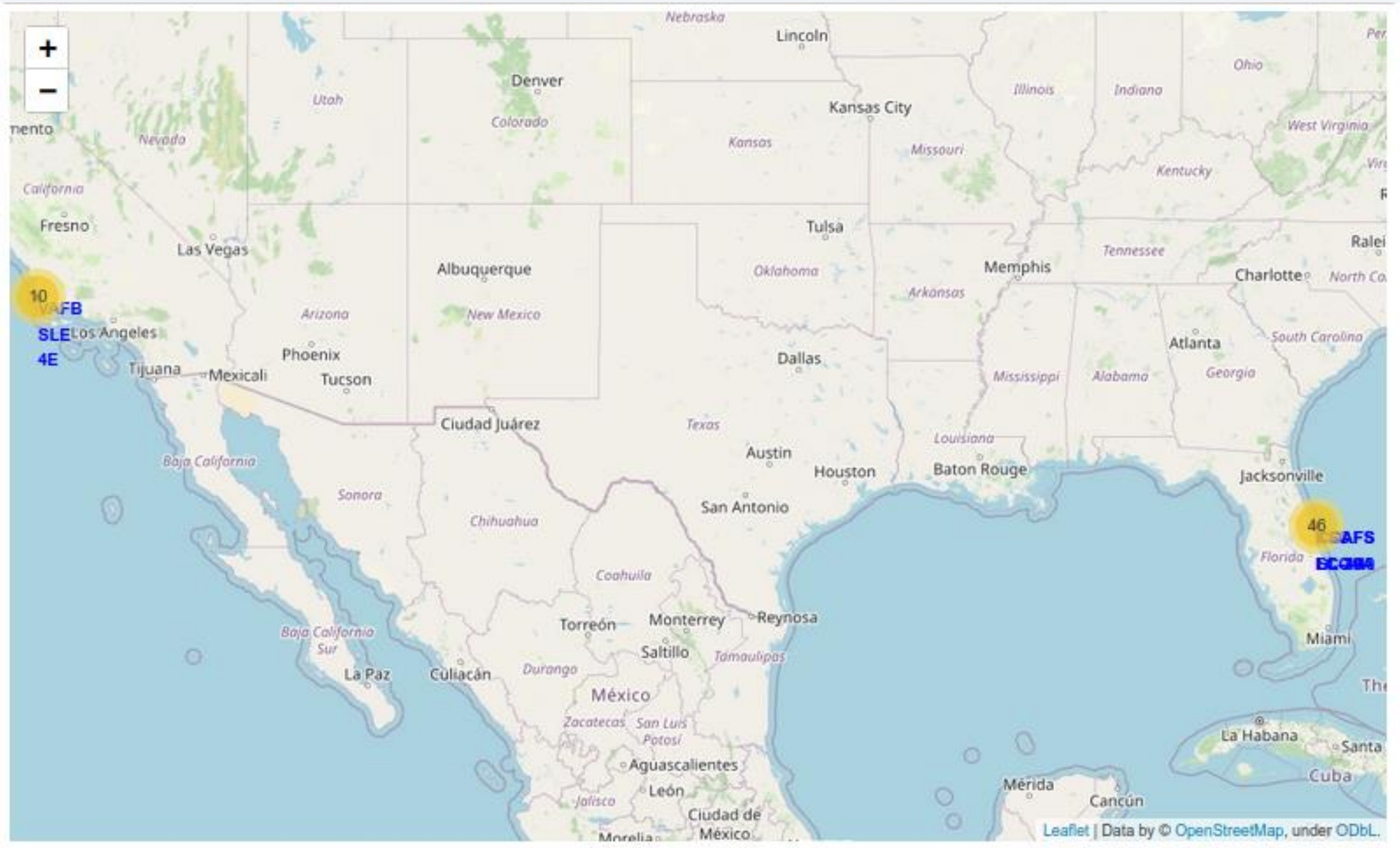
NASA location



Launch Sites locations



Launch Sites with flights numbers

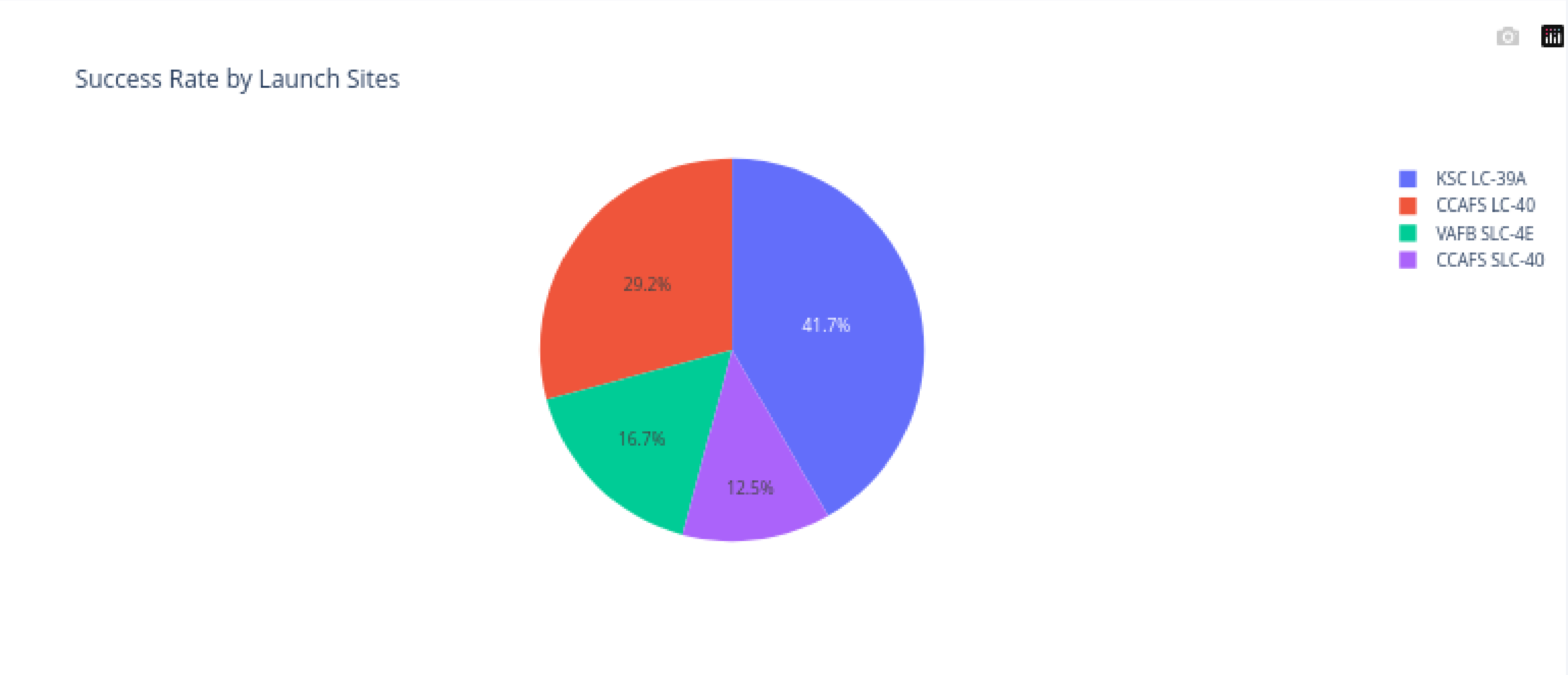


Section 4

Build a Dashboard with Plotly Dash

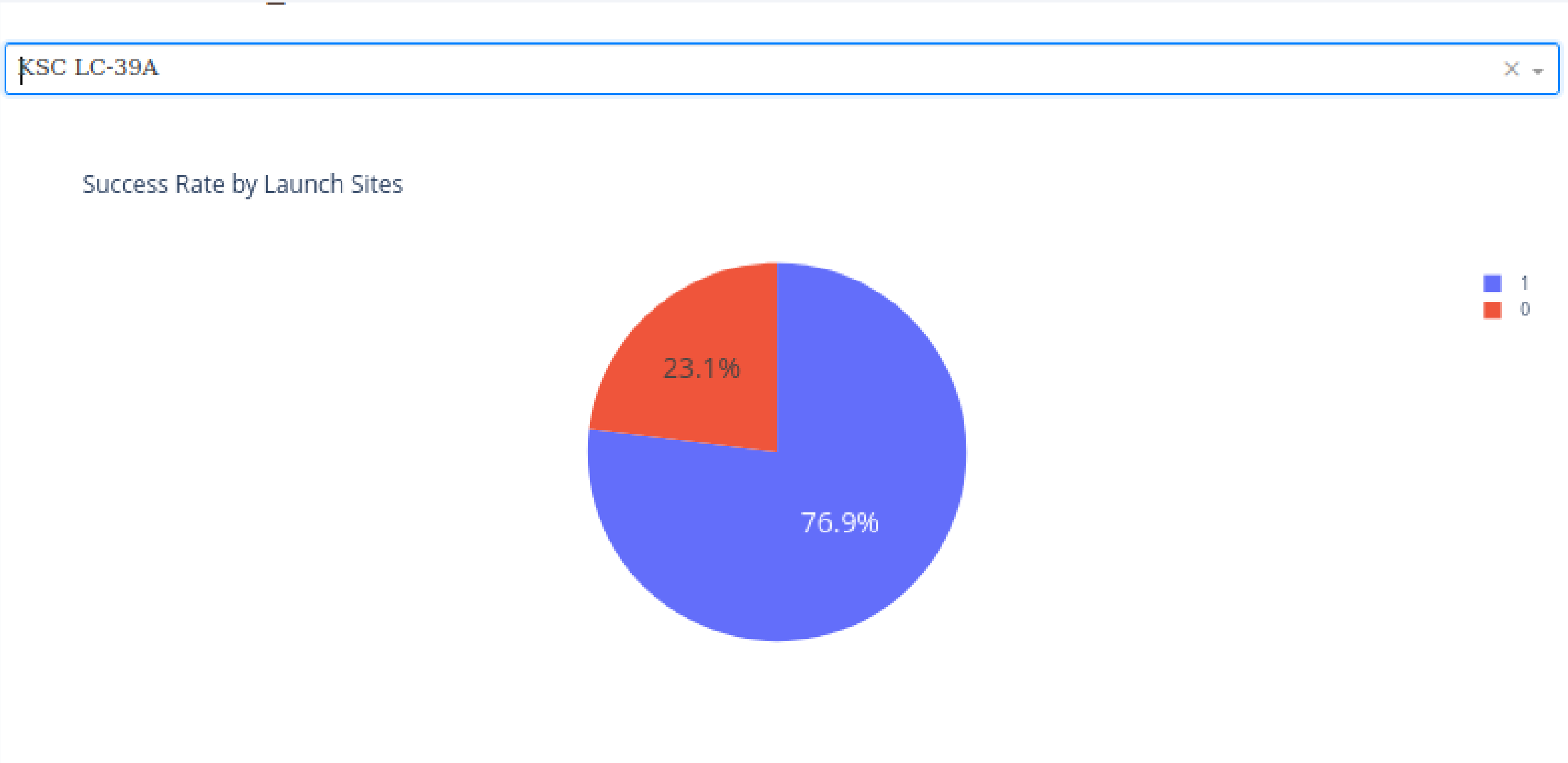
Success Rate by Launch Site

KSC LC-39 has 41.7% number of success launched



Rate for KSC LC-39A

KSC LC-39 has 79.9% success in total



Payload and Booster Version

Fails doesn't depend of payload mass

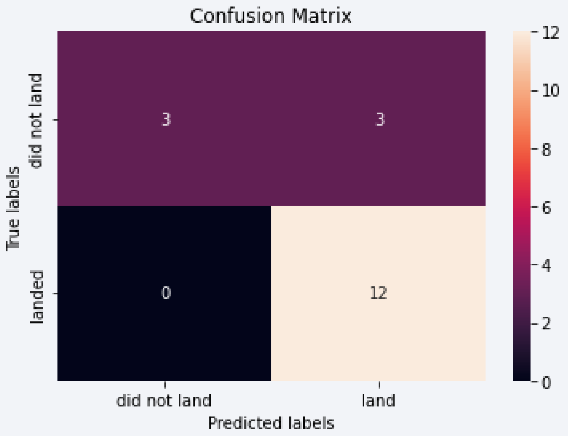
Section 5

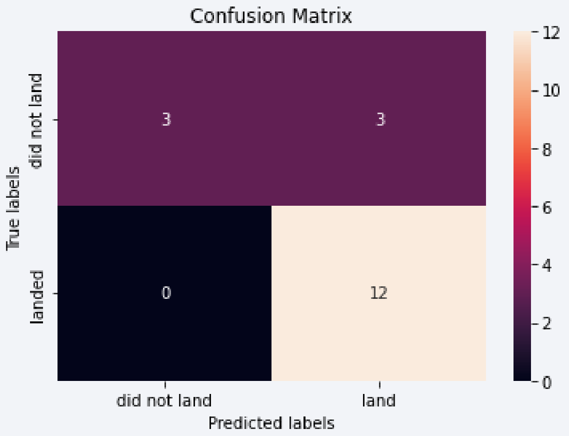
Predictive Analysis (Classification)

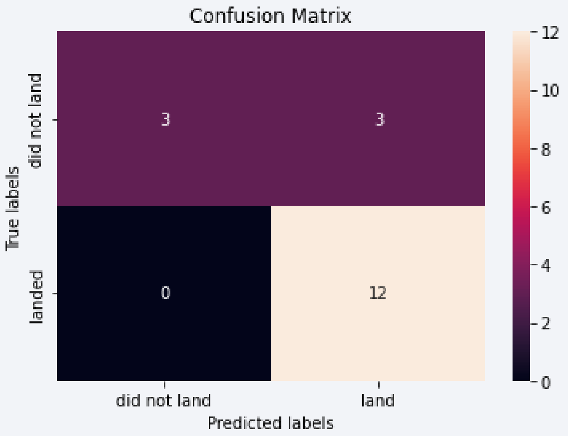
Classification Accuracy

```
Accuracy for Logistics Regression method: 0.8333333333333334
Accuracy for Support Vector Machine method: 0.8333333333333334
Accuracy for Decision tree method: 0.8333333333333334
Accuracy for K nearsdt neighbors method: 0.8333333333333334
```

Confusion Matrix







Conclusions

- In the years 2010-2013 the flights were100% fails
- Into orbit ISS flights have been taking place from the beginning until today
- Orbits ES-L1, CEO, HEO and SSO have the biggest success rate
- From Launch Site VAFB SLC 4E the Payload Mass never was bigger then 10,000
- Fails doesn't depend of payload mass

Thank you!