IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Name>
<Date>

Outline

Executive Summary

- Summary of methodologies

In the task we use different methodologies and libraries including Python and SQL languages. Main libraries used in the project were Pandas, Numpy and Scikit Learn

- Summary of all results
- Based on the analysis we can say that Decision tree is the best model to use for the prediction of the success of the mission.

Introduction

- **Project background and context**
  - Our company SPACE Y is planning to start development of the commercial space rockets. This rockets will be alternative for other companies and provide cost effective and sustainable option for space travel and cargo delivery.

- **Problems you want to find answers**
  1. Determine the price of the rocket launch by SPACE Y
  2. Predict the use of the first stage rockets in SPACE Y using Machine Learning.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected using SPACE X API
- Perform data wrangling
  - Data was extracted from the json response anc cleaned using Pandas library. Only important information was saved for further analysis.
- 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

- Describe how data sets were collected.
    - Data was collected from the SPACE X API using the request library
    - Next step was data cleaning using Pandas Library
- You need to present your data collection process use key phrases and flowcharts

Request from SPACE X API endpoint

static_json_url='https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/API_call_spacex_api.json'

JSON normalization

data = pd.json_normalize(dictr)

Data cleaning and processing

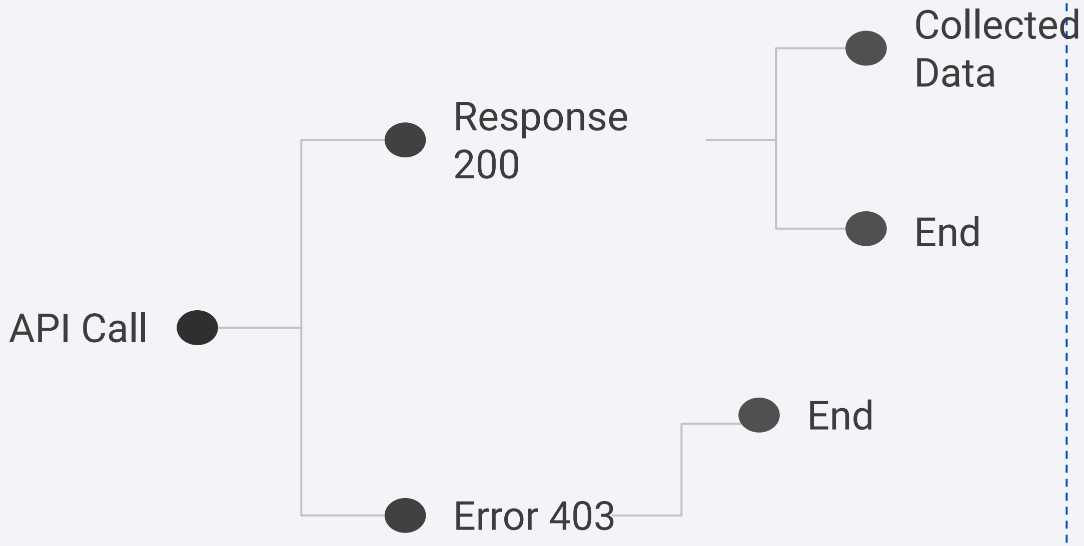
Pandas

Data Collection – SpaceX API

- Present your data collection with SpaceX REST calls using key phrases and flowcharts

- Add the GitHub URL of the completed SpaceX API calls notebook (must include completed code cell and outcome cell), as an external reference and peer-review purpose

https://github.com/PiotrKrosniak/coursera_capstone/blob/master/data_colection.ipynb

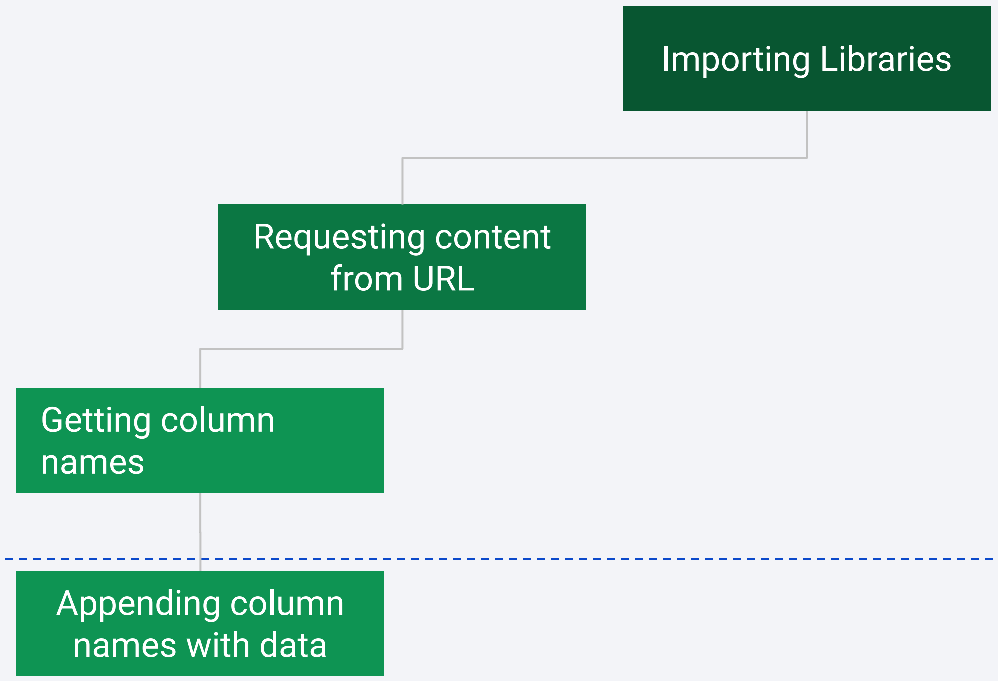


Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts

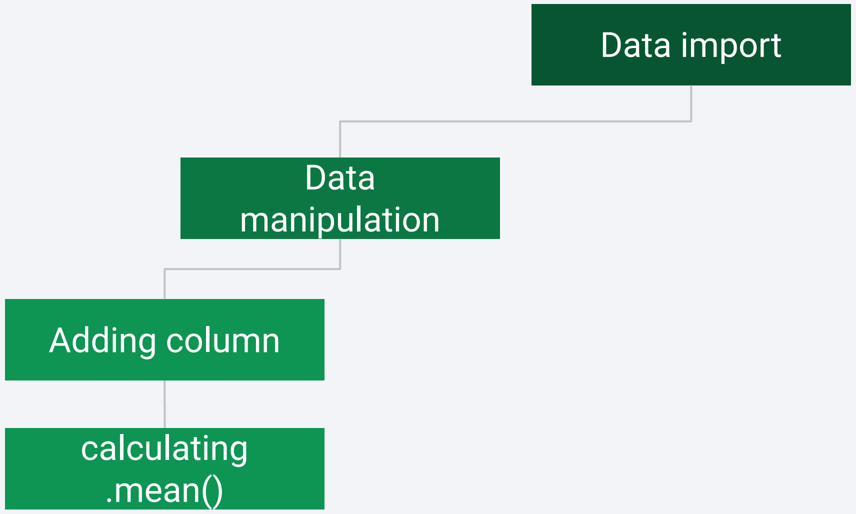
- Add the GitHub URL of the completed web scraping notebook, as an external reference and peer-review purpose

https://github.com/PiotrKrosniak/coursera_capstone/blob/master/data_collection_webscraping_python.ipynb



Data Wrangling

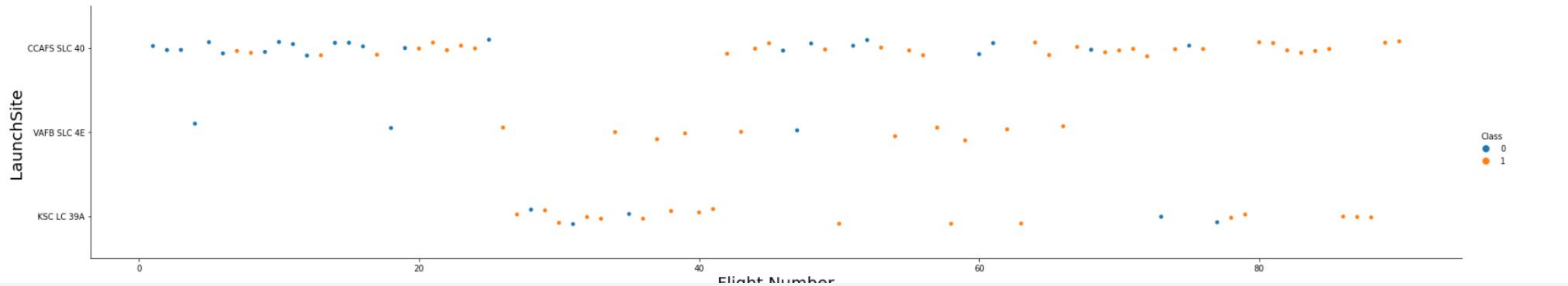
- Describe how data were processed
  - Data was extracted from the CSV file
  - cleaned from all null values
  - New column was created to capture landing class
- You need to present your data wrangling process using key phrases and flowcharts
- Add the GitHub URL https://github.com/PiotrKrosniak/coursera_capstone/blob/master/data_wrangling.ipynb



EDA with Data Visualization

- Summarize what charts were plotted and why you used those charts
  - Scatter plots for different columns in the dataframe
  - bar chart
  - line plot
- Add the GitHub URL
  https://github.com/PiotrKrosniak/coursera_capstone/blob/master/jupyter_labs_eda_dataviz.ipynb

```python
# Plot a scatter point chart with x axis to be Flight Number and y axis to be the launch site, and hue to be the class value
sns.catplot(y="LaunchSite", x="FlightNumber", hue="Class", data=df, aspect = 5)
plt.xlabel("Flight Number",fontsize=20)
plt.ylabel("LaunchSite",fontsize=20)
plt.show()
```



EDA with SQL

- Using bullet point format, summarize the SQL queries you performed
    - Extract the distinct site launch
    - count the total and average payload mass carried by the booster
    - check the mission outcomes and successful boosters

- Add the GitHub URL
  https://github.com/PiotrKrosniak/coursera_capstone/blob/master/eda_with_sql.ipynb

Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map

- Explain why you added those objects
- Add the GitHub URL
  https://github.com/PiotrKrosniak/coursera_capstone/blob/master/lab_jupyter_launch_site_location%20(1).ipynb

Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard
  - Bar Chart
  - Scatter plot
- Explain why you added those plots and interactions
  - Bar chart to display success rate for different launch sites
  - Scatter plot to display correlation between Payload Mass and mission success

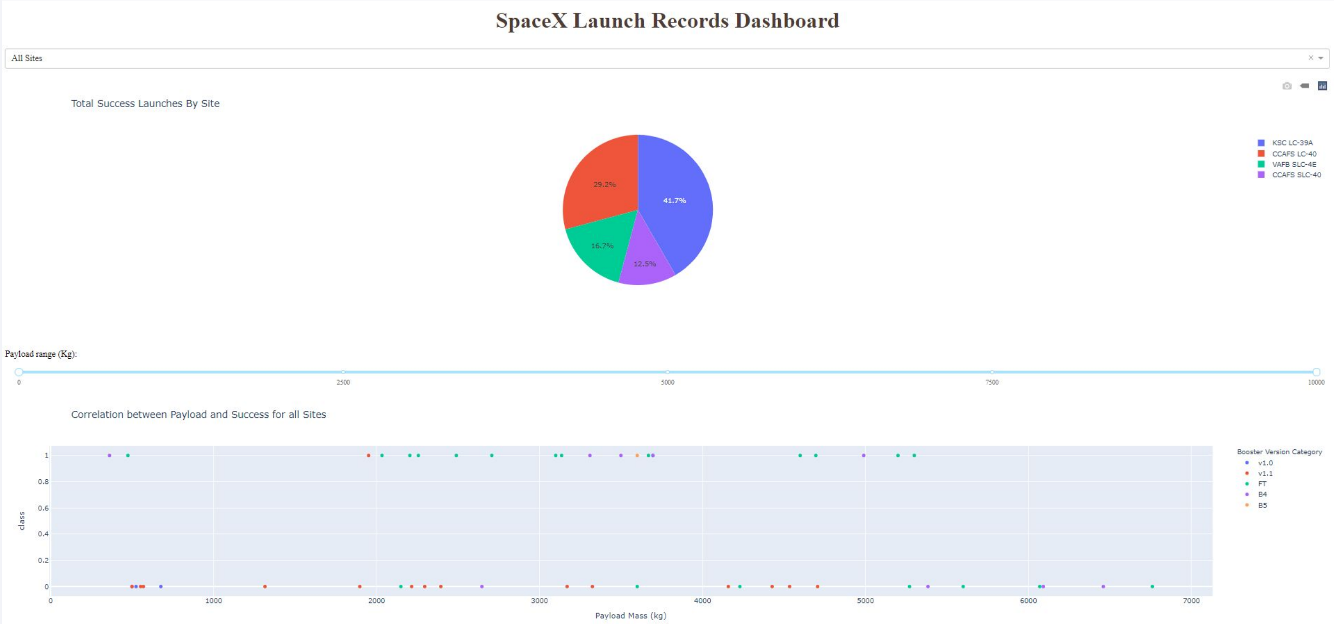
Add the GitHub URL
https://github.com/PiotrKrosniak/coursera_capstone/blob/master/spacex_dash_app.py

Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model
  - I analyzed multiple models including logistic regression, Support Vector Machines (SVM), Decision Tree, K Nearest Neighbor,
  - The best accuracy model is Decision Tree with 90% accuracy
- Add the GitHub URL https://github.com/PiotrKrosniak/coursera_capstone/blob/master/maschine_learning.ipynb

Results

- Exploratory data analysis results
    - 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%.
- Predictive analysis results
    - Decision tree model have the highest accuracy

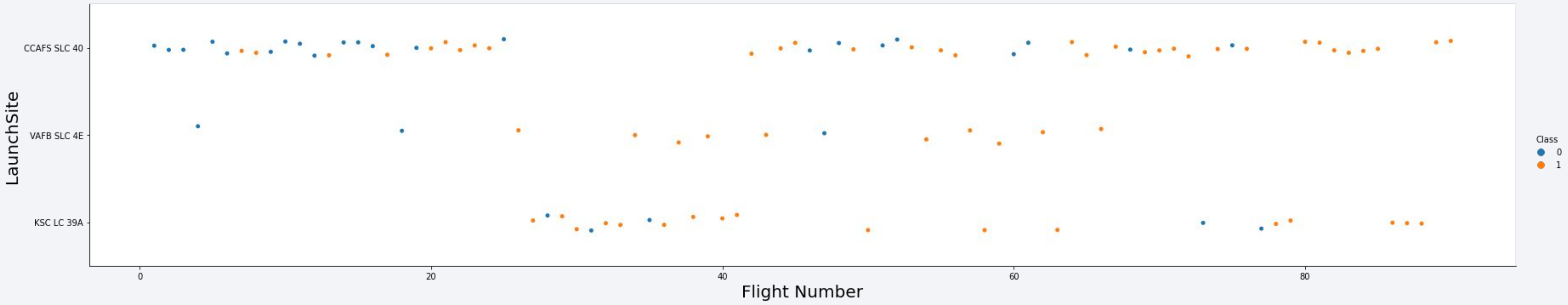


Section 2

Insights drawn from EDA

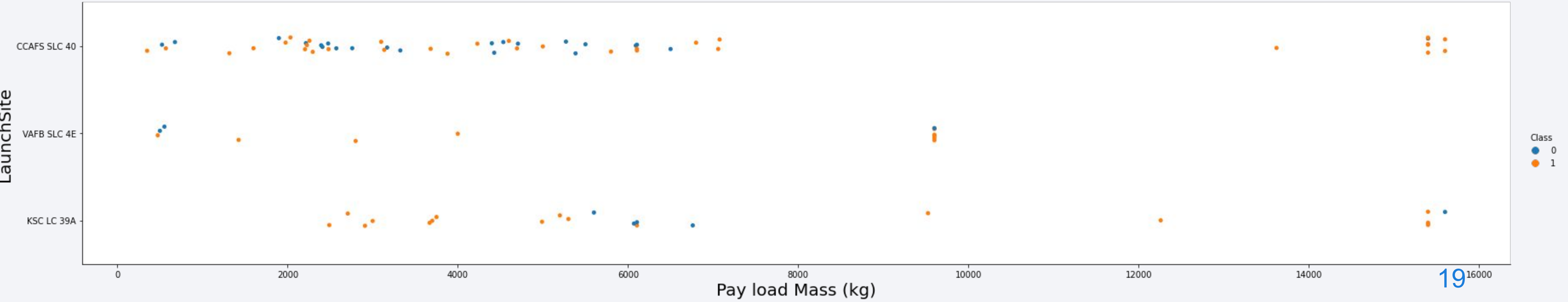
Flight Number vs. Launch Site

- We can see that some launch sites are utilized more frequently and also that have different success rates.
- Site CCAFS SLC 40 is the most utilized site for launching and site VAFB SLC 4E is the least site.



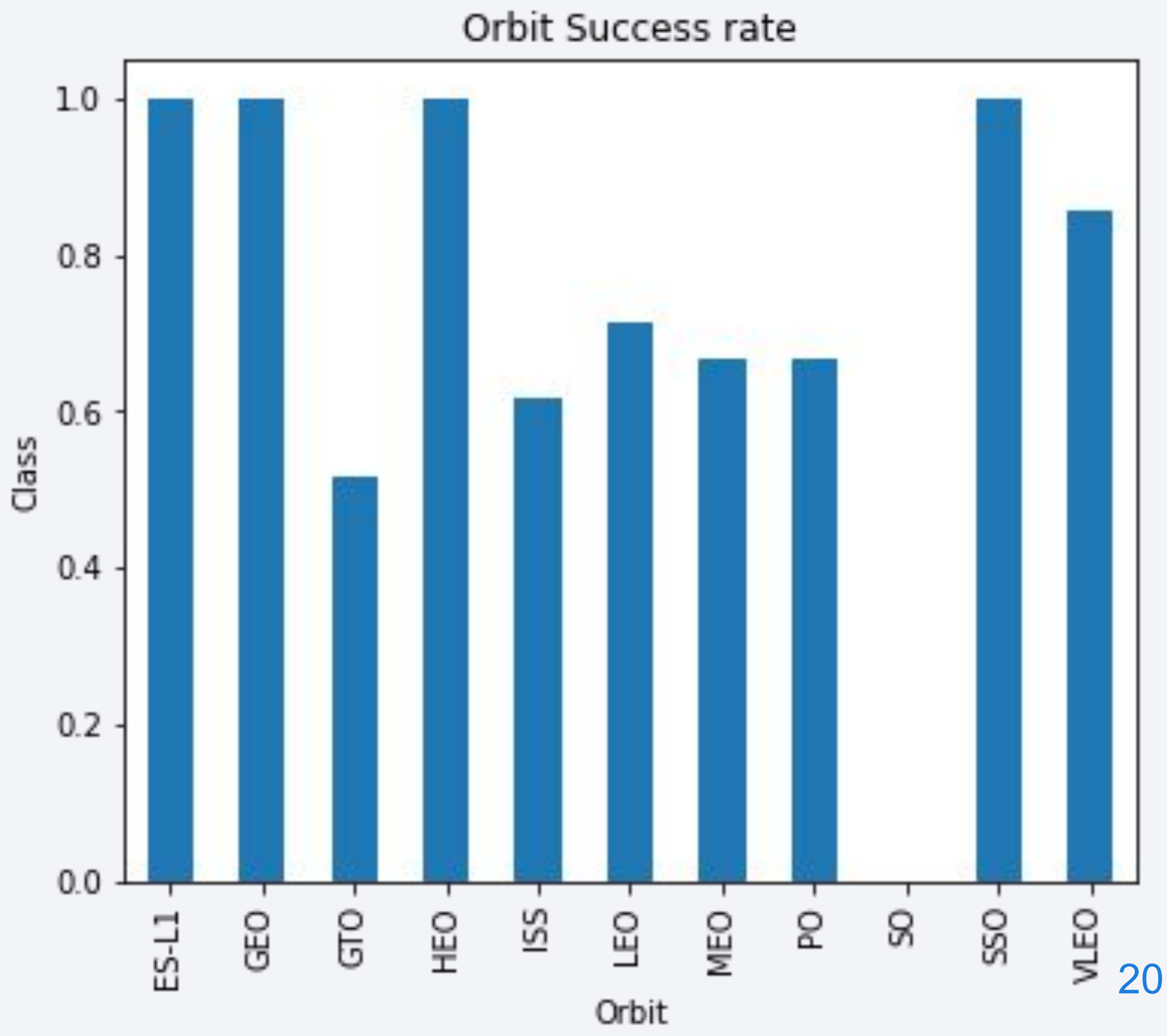
Payload vs. Launch Site

- We can see that different site launch also are utilized for different Payload volume.
- Sites CCAFS SLC 40 is utilized for smaller volume between 0-7000 kg and sites KSC LC 39A for cargo between 2000 - 7000 and finally site VAFB SLC 4E foir cargo from 0 - 10000



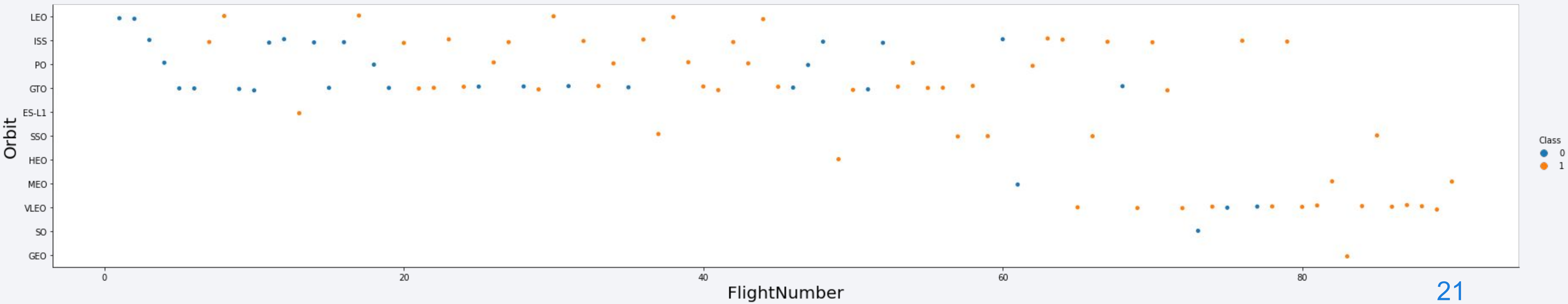
Success Rate vs. Orbit Type

- We can see that different orbit that SpaceX rockets are going to have different success rates.
- For orbit SSO, ISS, GEO and ESL1 success is very high for this is dropping below.



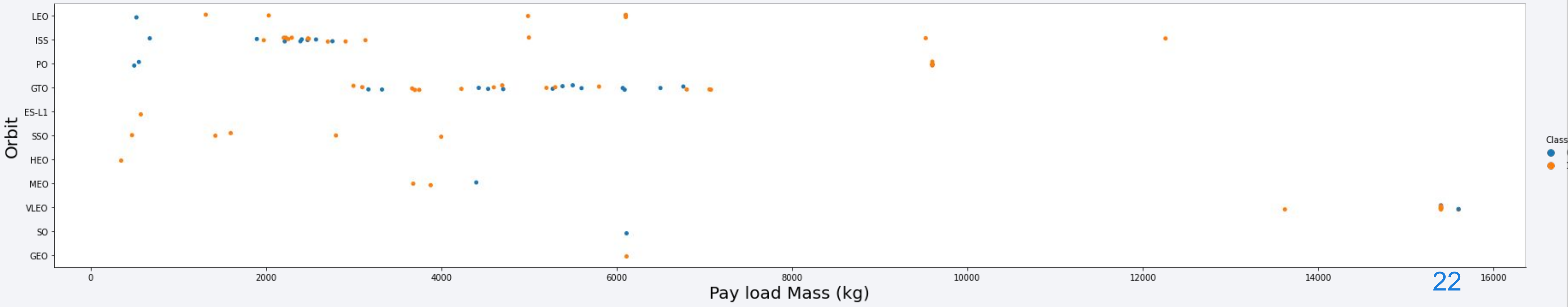
Flight Number vs. Orbit Type

- We can see that number of flight was also send to different orbit.
- Over 60% of flight was send to orbit LEO, ISS, PO and GTO with only few flight to other orbits.



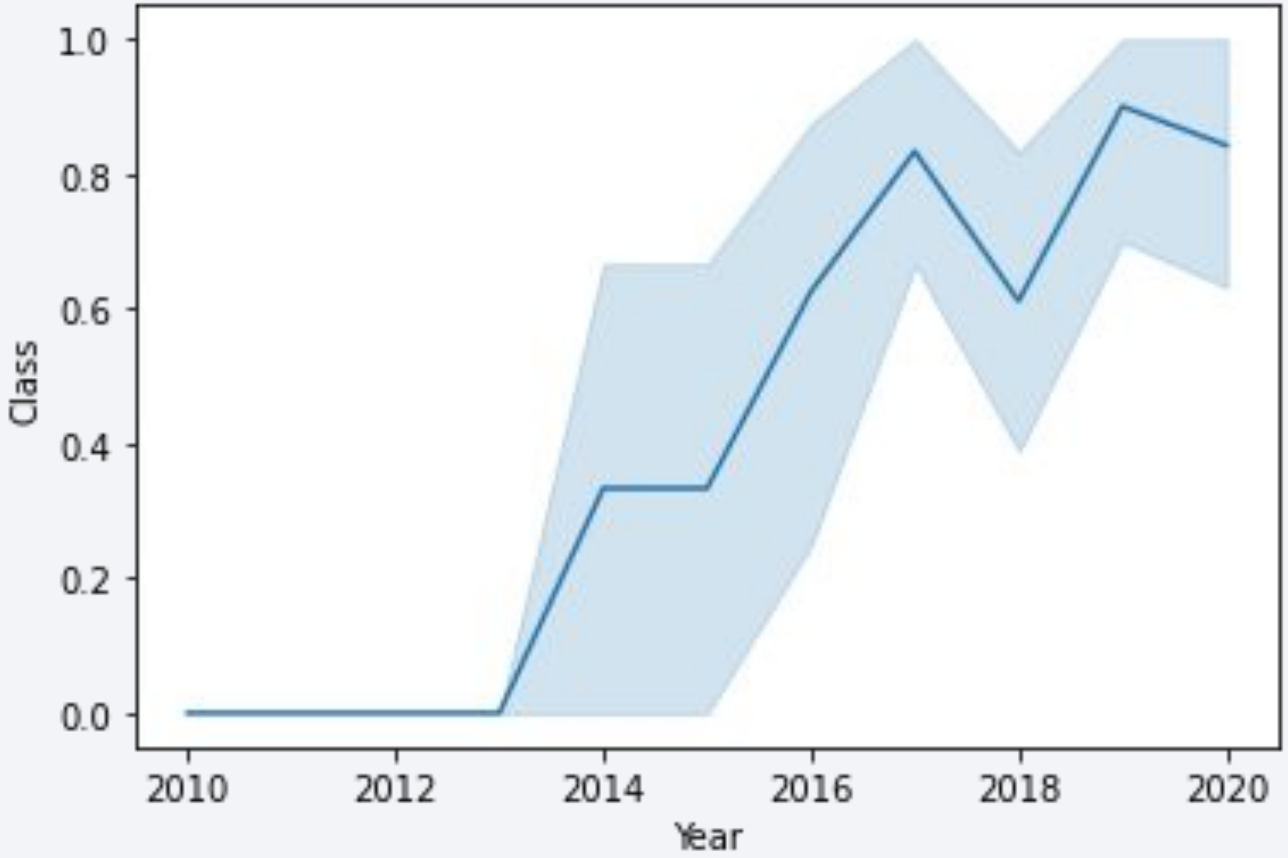
Payload vs. Orbit Type

- We can see that different payload was send to different Orbit.
- The highest Payload was send to orbit VLEO and for other payload was between 0 -1000 KG



Launch Success Yearly Trend

- We can see that the success rate since 2013 kept increasing till 2020



All Launch Site Names

- SpaceX is utilizing 4 different site launch for their rockets.

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

Launch Site Names Begin with 'CCA'

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

launch_site
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40

Total Payload Mass

- Total volume of cargo carried by the SpaceX booster is 619967 KG

```
In [7]: %sql Select SUM (PAYLOAD_MASS__KG_) from spacex

 * ibm_db_sa://zsk32467:***@8e359033-a1c9-4643-82ef-8ac06f5107eb.bs2io90108kqblod8lcg.databases.appdomain.cloud:30120/bludb
Done.

Out[7]:
```

1
619967

Average Payload Mass by F9 v1.1

- Average payload mass carried by the booster version F9 v1.1 is 2928 KG

1
2928

First Successful Ground Landing Date

- First successful landing on the ground was on 12-22-2015

1
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

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

Boosters Carried Maximum Payload

- List the names of the
  mass
- Present your query re

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

2015 Launch Records

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

landing__outcome	booster_version	launch_site
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
Controlled (ocean)	F9 v1.1 B1013	CCAFS LC-40
No attempt	F9 v1.1 B1014	CCAFS LC-40
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40
No attempt	F9 v1.1 B1016	CCAFS LC-40
Precluded (drone ship)	F9 v1.1 B1018	CCAFS LC-40

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

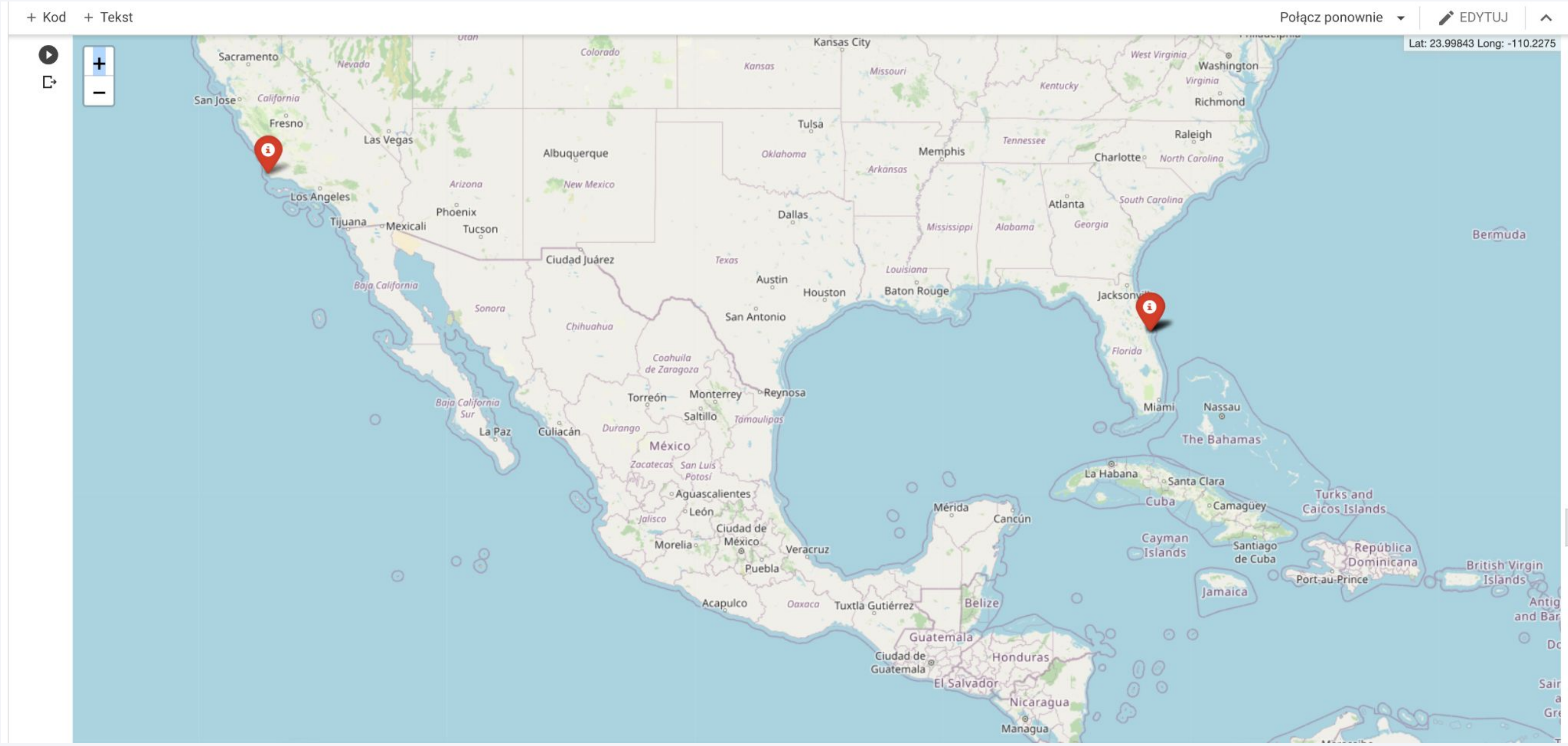
- Present your query result with a short explanation here

Section 4

Launch Sites Proximities Analysis

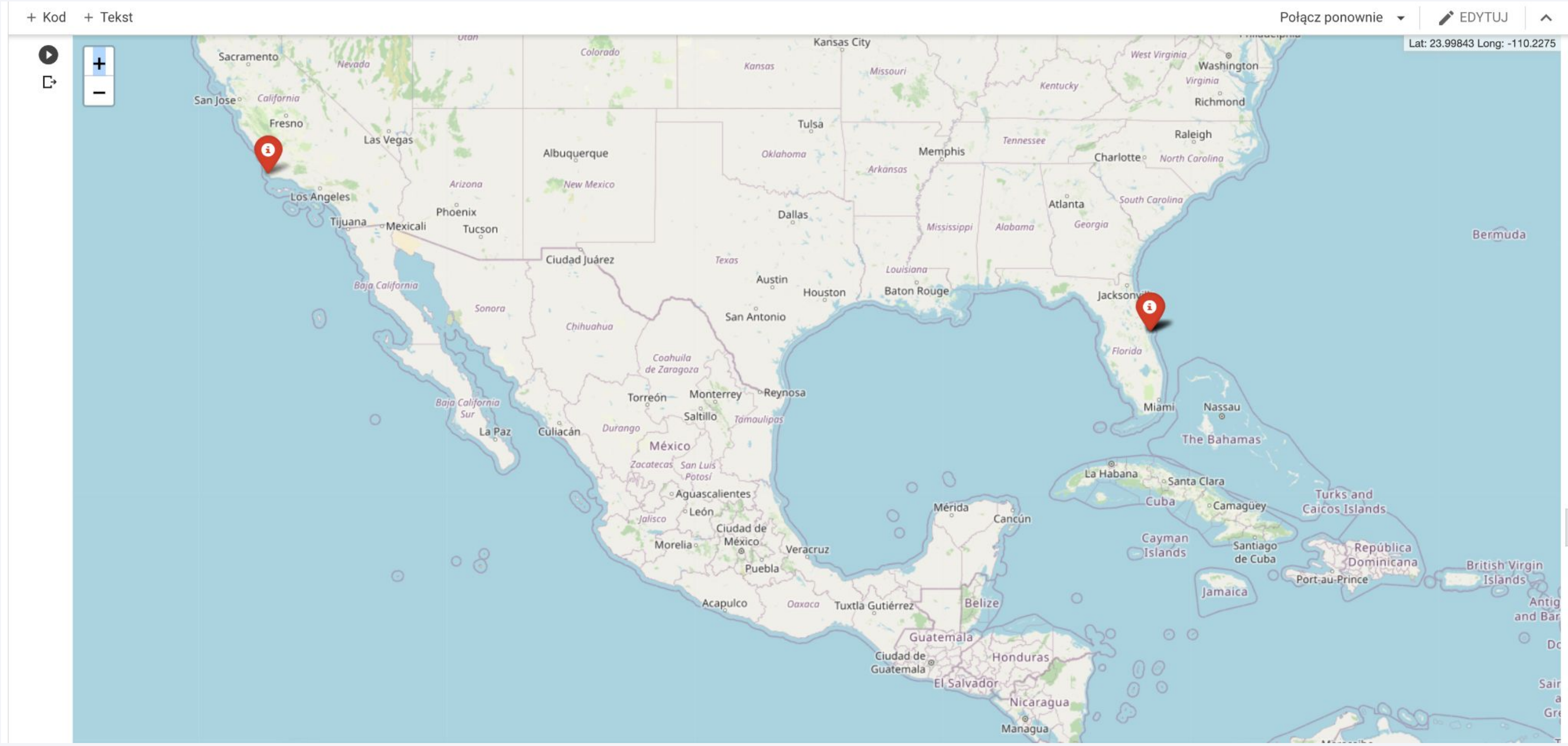
Launch site locations on map

- SpaceX is using 4 different launch sites two in USA and two in Canada



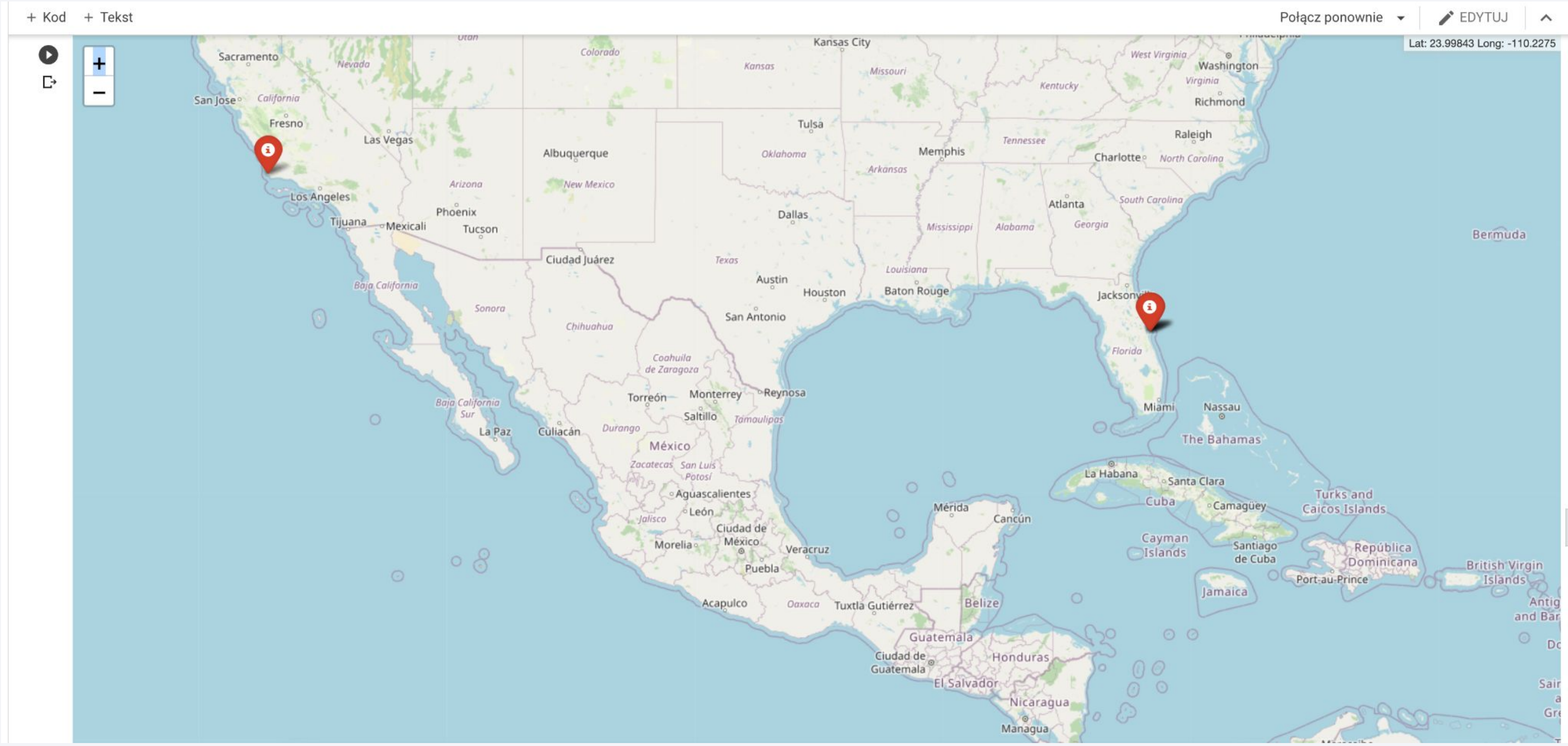
Folium map with markers cluster

- Marker for the sites



Distance between coastline

- Distance between launch site and coastline

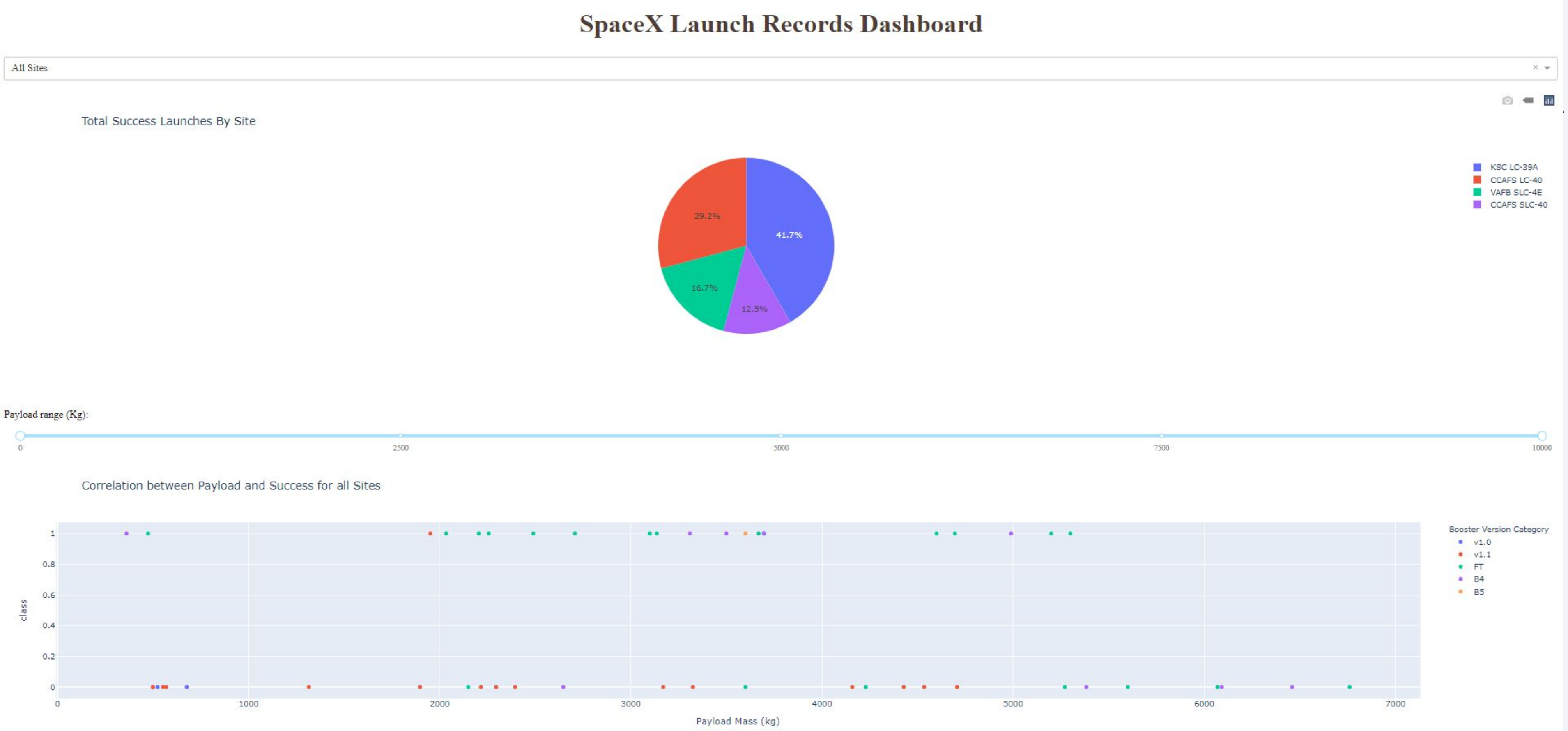


Section 5

Build a Dashboard with Plotly Dash

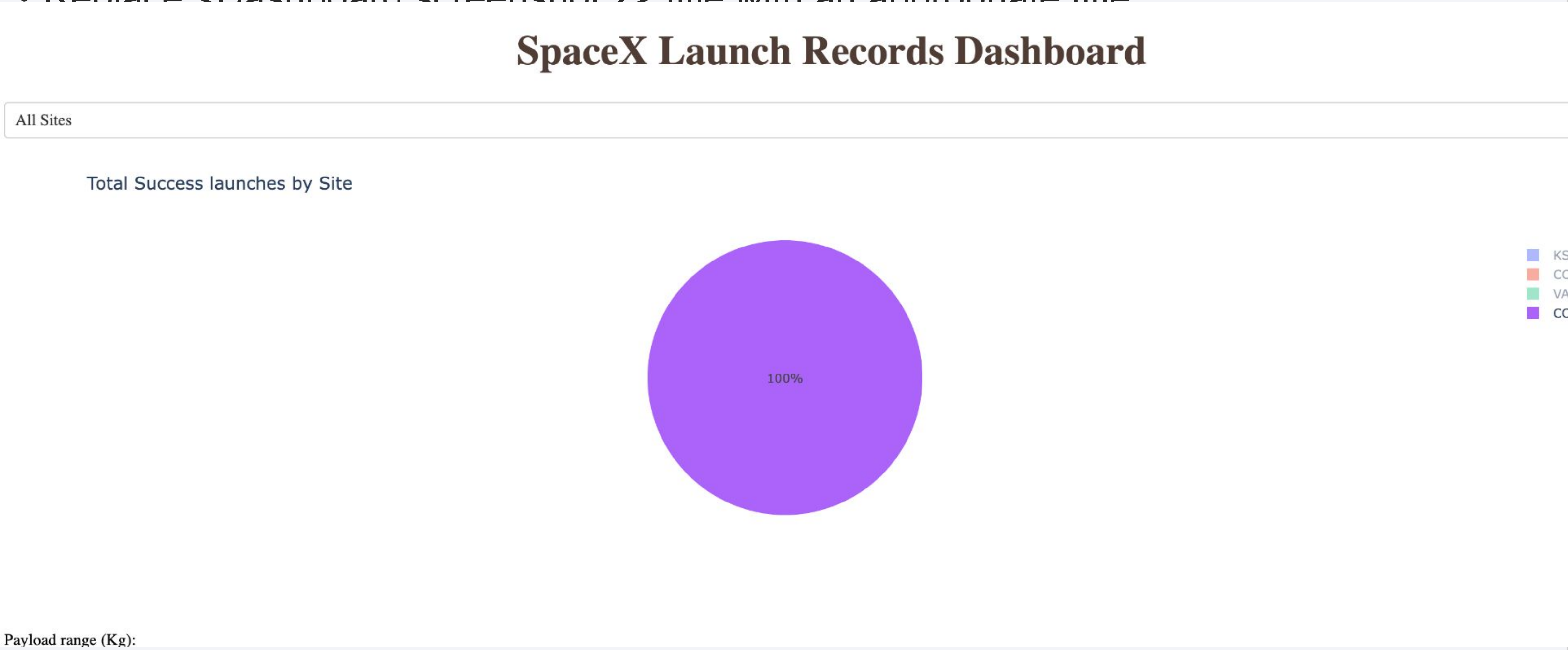
Space X launch records dashboard

- Dashboard is presenting different sites and succes rates for these sites



Best launching sites

- Replace <Dashboard screenshot 2> title with an appropriate title



Payload and launch ortucome - dashboard

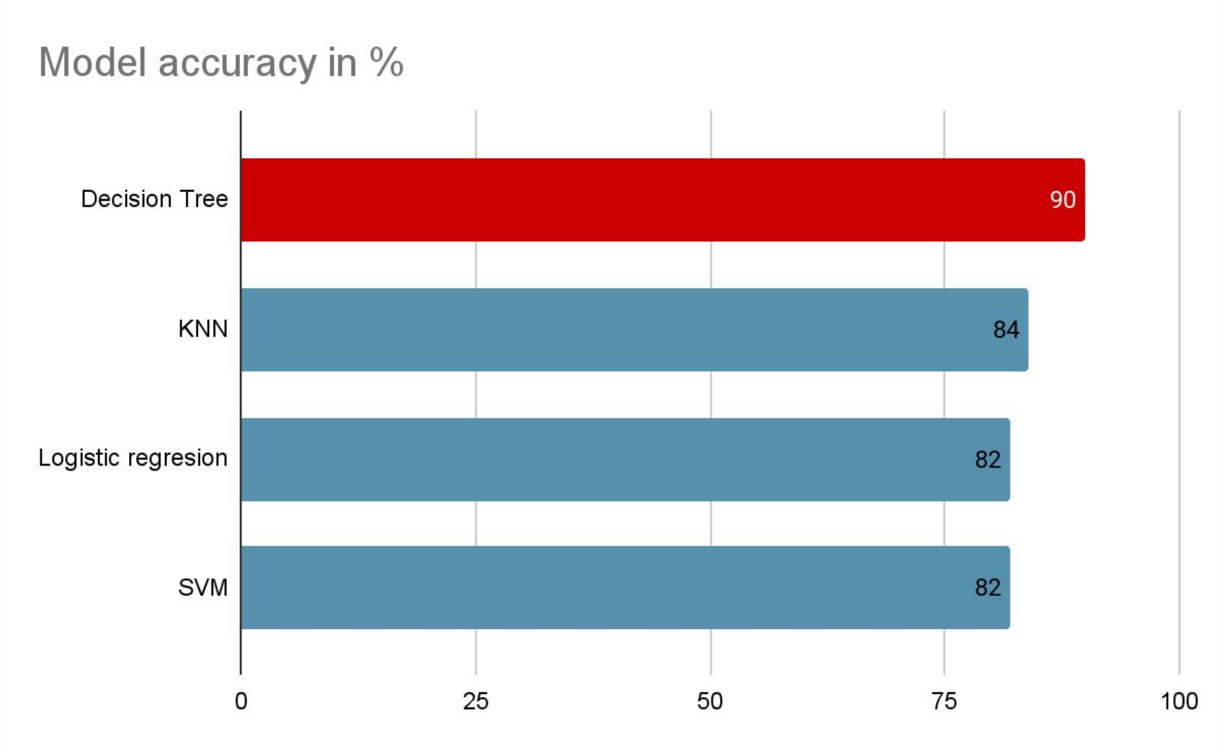
- Different payload have different launch outcome the best outcome have middle payload launch between 2000 and 4000 KG

Section 6

Predictive Analysis (Classification)

Classification Accuracy

- The highest accuracy was achieved using Decision Tree model with 90% accuracy based on the test data .



Confusion Matrix

- The highest accuracy was achieved using Decision tree model 90% based on the training data. Concussion matrix is showing that major problem is false positive.

Conclusions

- The best launch site to use for our rockets is CCAFS SLC 40 with highest succes rate among all launching sites.
- Payload of the flights is usually medium between 2000-4000 KG and our rockets should focus on this volume
- We should send our flight to SSO, ISS, GEO and ESL1 orbits that most of the flight have successful
- For prediction of the flight outcome we will use Decision Tree classifier as the the one with highest accuracy

Thank you!