SPACEX FALCON 9 FIRST LANDING PREDICTION

Falko

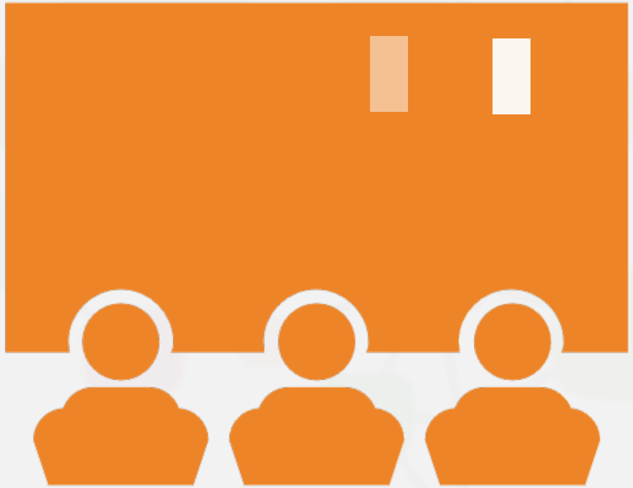
27-09-2022

IBM Developer

SKILLS NETWORK

OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
  - Visualization – Charts
  - Dashboard
- Discussion
  - Findings & Implications
- Conclusion
- Appendix



EXECUTIVE SUMMARY

- In this project, we will predict if the Falcon 9 first stage will land successfully to determine the cost. This information can be used to bid against SpaceX for a rocket launch.
- Data is collected from the SpaceX API and webscraping from Wikipedia. After data wrangling, cleaning, exploration and visualization, several classification models are build including:
  - Support Vector Machine
  - Logistic Regression
  - Decision Tree
  - KNN
- All three models have an accuracy of 83.33% on unseen test data

INTRODUCTION

Background

We will predict if Falcon 9 first stage will land successfully. Thereby, we can determine the cost of a launch. An alternative company can use this information to bid against SpaceX for rocket launches.

Research question

What variables and conditions affect the successful landing of a rocket launch?

METHODOLOGY

Outline

- Data collection
- Data wrangling
- Exploratory data analysis (EDA) with visualization and SQL
- Interactive visualization in Folium and Plotly
- Predictive analysis

METHODOLOGY: DATA COLLECTION

Getting data from SpaceX Rest API

- GET request to REST API
- Receiving data (JSON file)
- Normalizing data for a CSV file
- Cleaning data with functions
- Creating a dataframe by coding a dictionary including all columns
- Export to CSV file

METHODOLOGY: DATA COLLECTION

Web scraping Wikipedia

- GET HTML Response object from Falcon Heavy Launches Records
- Creating a 'soup' object using BeautifulSoup
- Normalizing data (CSV file)
- Assigning column header names from table
- Parsing all tables in dictionaries
- Creating a Pandas DataFrame from dictionary

METHODOLOGY: DATA WRANGLING

Data wrangling actions

- Checking null values
- Creating classes (0 = failure, 1 = successful) based on column which identifies whether and why rockets landed successfully
- Calculating number of rocket launches on each site
- Calculing the number and occurance of each orbit
- Mission outcome per orbit type
- Handling null values of dataframe

METHODOLOGY: EXPLORATORY DATA ANALYSIS – DATA VISUALIZATION

Different type of charts

- Scatter charts
  - Flight_Number vs Launch_Site
  - Payload vs Launch_Site
  - Orbit vs Flight_Number
  - Payload vs Orbit
- Bar chart
  - Success_Rate vs Orbit
- Line chart
  - Success_Rate vs Year

METHODOLOGY: EXPLORATORY DATA ANALYSIS – SQL QUERIES

SQL queries to answer following questions

- 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 succesful 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. Use a subquery
- List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015
- Rank the count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order

METHODOLOGY: INTERACTIVE VISUALIZATION - FOLIUM

Interactive visualization actions using Folium

- Taking the latitude and longitude coordinates
- Adding a Circle Marker around all launch sites
- Assigning launch_outcomes (success/failure) to classes
- Calculating the distance from launch site to other landmarks

METHODOLOGY: INTERACTIVE VISUALIZATION – PLOTLY DASH

Different types of plots are created

- Pie plot:
  - Showing the total launches by each site (+total)
- Scatter plot:
  - Showing relatingship with outcome and payload mass

METHODOLOGY: PREDICTIVE ANALYSIS

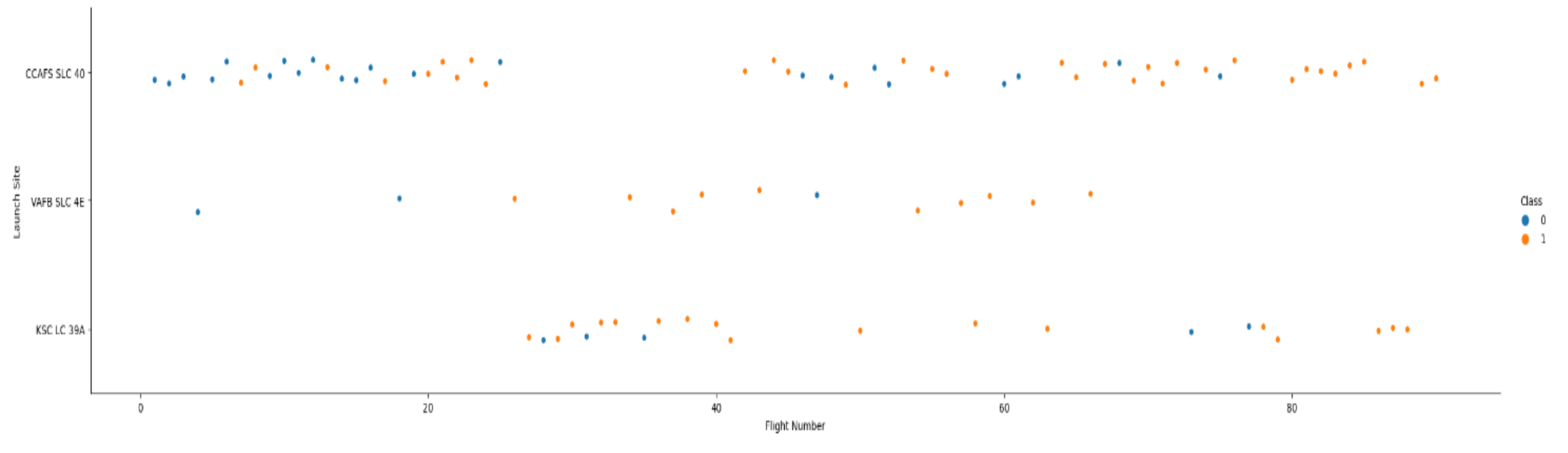
- First the data loaded, transformed and standardized
- Then the data is split into training and test data
- Training different classification models: Support Vector Machine, Classification Decision Tree, Logistic Regression
- Each model is evaluated and improved by finding the best hyperparameters using GridSearchCV
- The best performing classification model is identified and confusion matrices are plotted for each classification model
- Finally, the accuracy of each classification model is calculated

RESULTS

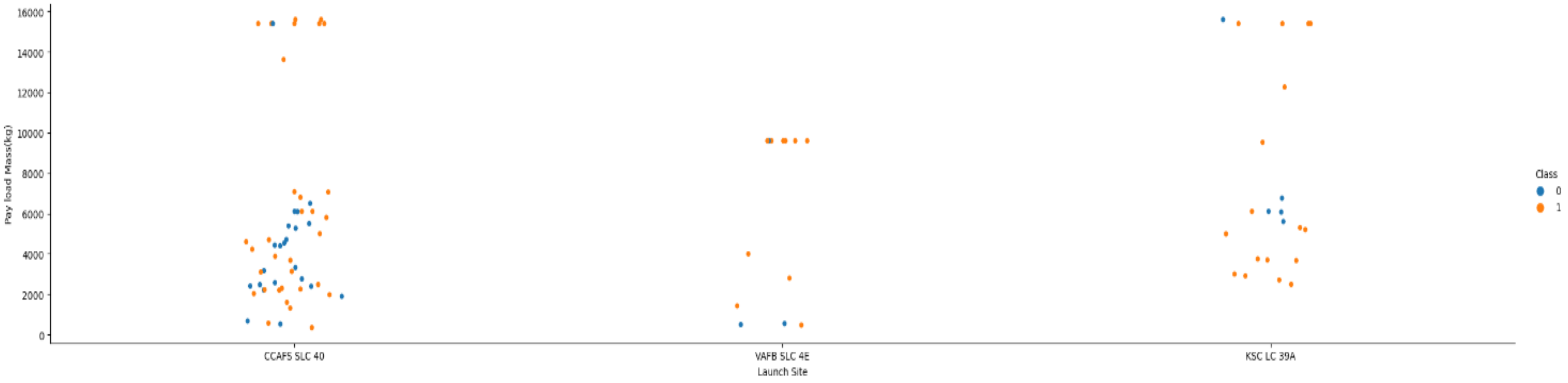
Outline

- Exploratory data analysis
  - EDA with visualization
  - EDA with SQL
- Interactive graphs
  - Folium
  - Plotly Dash
- Predictive analysis

EDA WITH VISUALIZATION: FLIGHT NUMBER VS. LAUNCH SITE



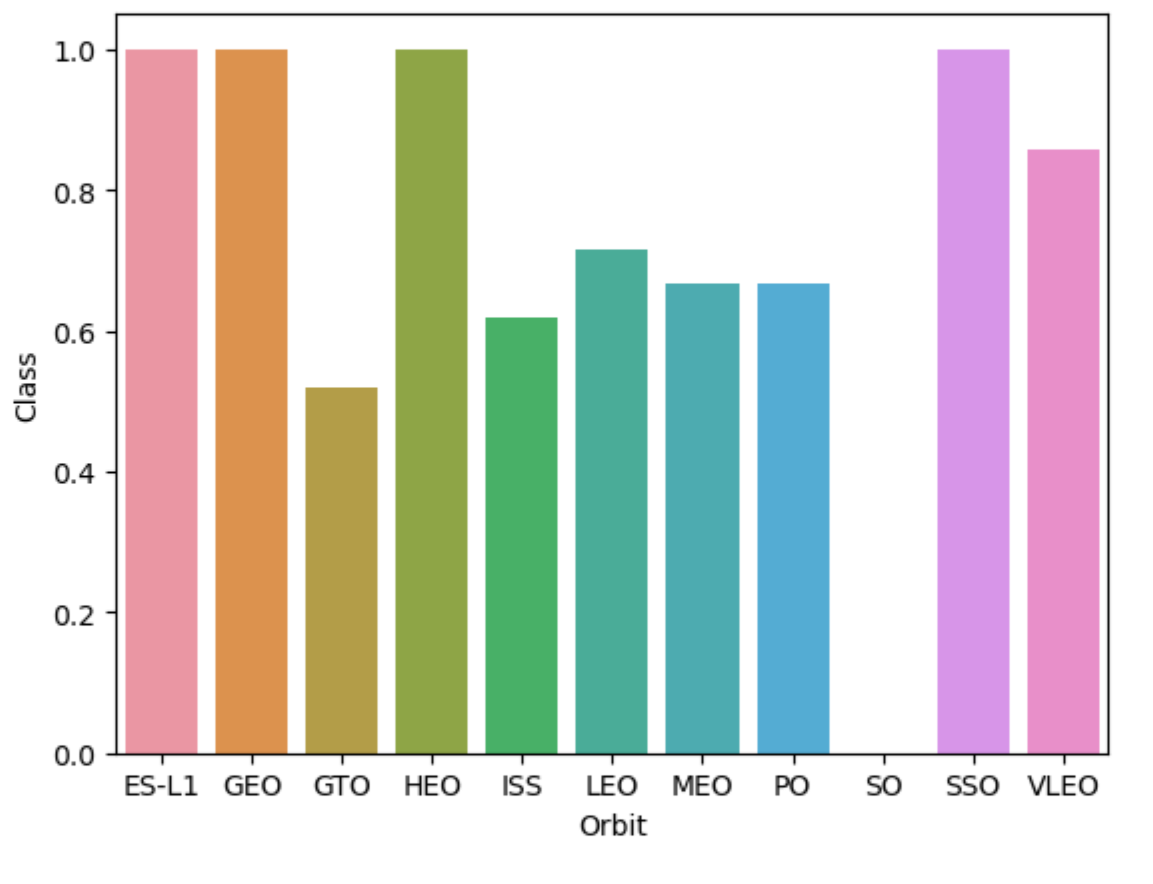
EDA WITH VISUALIZATION: PAYLOAD VS LAUNCH SITE



IBM Developer

SKILLS NETWORK

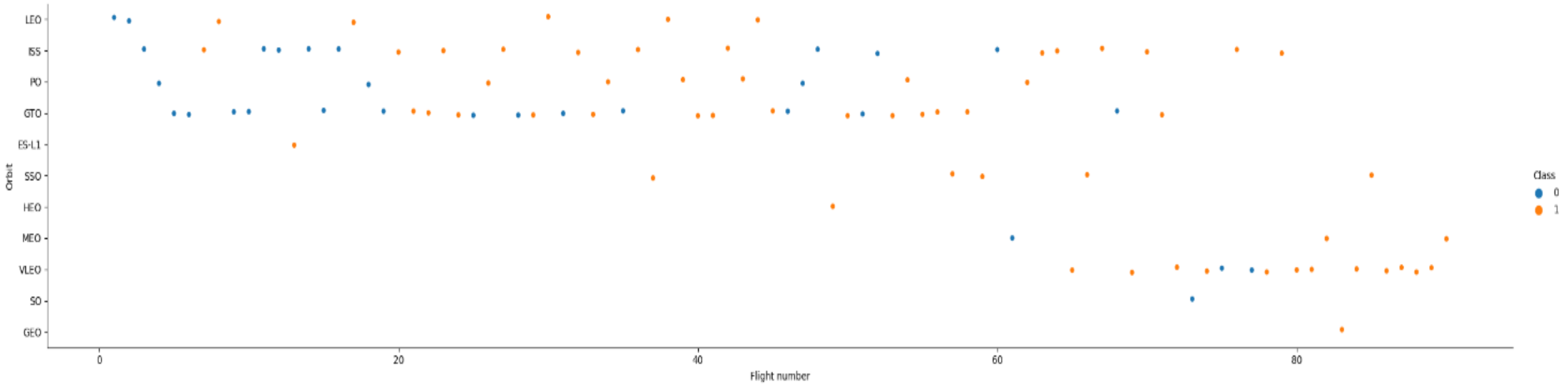
EDA WITH VISUALIZATION: SUCCESS RATE OF EACH ORBIT



IBM Developer

SKILLS NETWORK

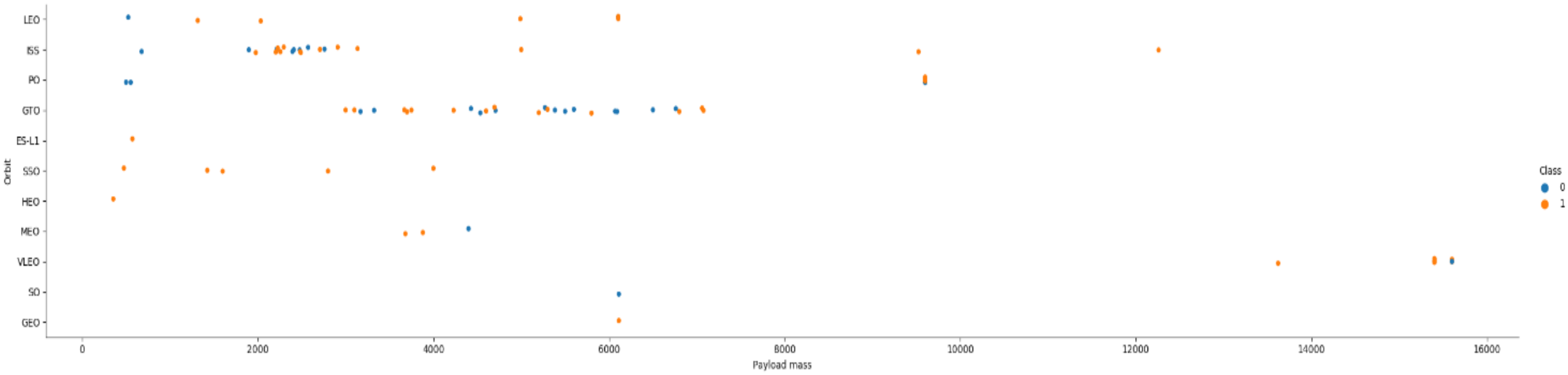
EDA WITH VISUALIZATION: FLIGHT NUMBER VS ORBIT TYPE



IBM Developer

SKILLS NETWORK

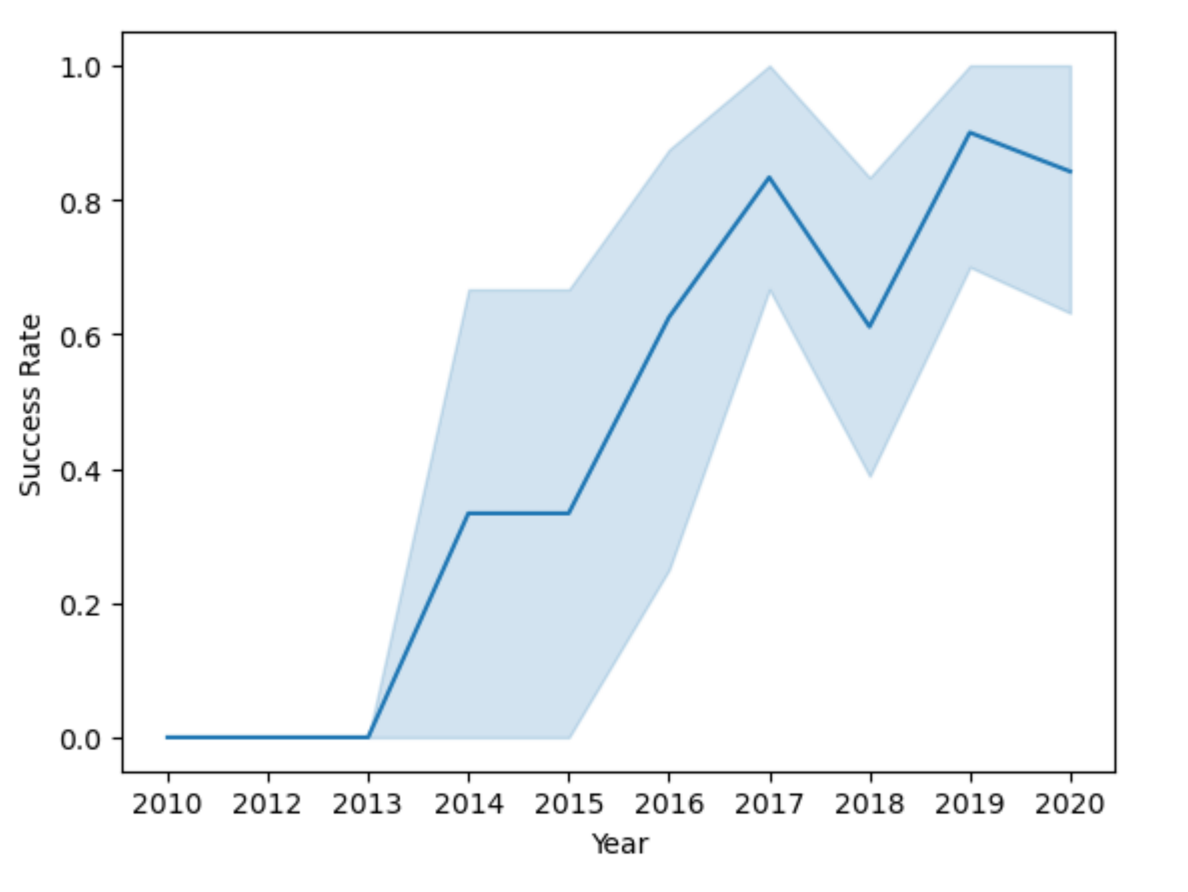
EDA WITH VISUALIZATION: PAYLOAD VS ORBIT TYPE



IBM Developer

SKILLS NETWORK

EDA WITH VISUALIZATION:
YEARLY LAUNCH SUCCESS RATE



EDA WITH SQL:
LAUNCH SITE NAMES

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

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

EDA WITH SQL:
5 LAUNCH SITES BEGINNING WITH 'CCA'

- %sql SELECT * FROM SPACEXTBL WHERE Launch_Site LIKE 'CCA%' LIMIT 5;

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

EDA WITH SQL: TOTAL PAYLOAD MASS + AVERAGE PAYLOAD MASS

- %sql SELECT SUM(PAYLOAD_MASS__KG_) AS TOTAL_PAYLOAD_MASS FROM SPACEXTBL WHERE CUSTOMER = 'NASA (CRS)';

TOTAL_PAYLOAD_MASS
45596

- %sql SELECT AVG(PAYLOAD_MASS__KG_) AS AVERAGE_PAYLOAD_MASS FROM SPACEXTBL WHERE BOOSTER_VERSION = 'F9 v1.1';

AVERAGE_PAYLOAD_MASS
2928.4

EDA WITH SQL: DATE OF FIRST SUCCESSFUL LANDING OUTCOME IN GROUND PAD

- %sql SELECT MIN(DATE) FROM SPACEXTBL WHERE [Landing _Outcome] = 'Success (ground pad)';

2015-12-22

EDA WITH SQL: NAMES OF BOOSTERS WITH SUCCESS IN DRONE SHIP AND PAYLOAD_MASS BETWEEN 4000 AND 6000

- %sql SELECT BOOSTER_VERSION FROM SPACEXTBL WHERE LANDING__OUTCOME = 'Success (drone ship)' AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000;

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

EDA WITH SQL:TOTAL NUMBER OF SUCCESSFUL AND FAILURE MISSIONS

- %sql SELECT COUNT(MISSION_OUTCOME) FROM SPACEXTBL WHERE MISSION_OUTCOME LIKE '%Success%' or MISSION_OUTCOME LIKE '%Failure%'

COUNT(MISSION_OUTCOME)
101

EDA WITH SQL: BOOSTER VERSIONS WHICH CARRIED MAXIMUM PAYLOAD MASS

- %sql SELECT DISTINCT(BOOSTER_VERSION) FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL);

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

EDA WITH SQL: MONTH NAMES, FAILURE LANDING OUTCOMES IN DRONE SHIP, BOOSTER VERSIONS, LAUNCH SITE FOR MONTHS IN 2015

- %sql SELECT substr(Date, 4,2) as Month_names, [Landing _Outcome], Booster_Version FROM SPACEXTBL WHERE substr(Date, 7, 4) = '2015' AND [Landing _Outcome] = "Failure (drone ship)";

month_name	landing__outcome	booster_version	launch_site
JANUARY	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
APRIL	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

EDA WITH SQL: COUNT OF SUCCESSFUL LANDING_OUTCOMES IN DESC ORDER BETWEEN 04-06-2010 AND 20-03-2017

- (only successful)

%%sql SELECT [Landing _Outcome], COUNT([Landing _Outcome]) AS TOTAL_OUTCOME FROM SPACEXTBL WHERE DATE BETWEEN '04-06-2010' and '20-03-2017'AND [Landing _Outcome] LIKE "%Success%"GROUP BY [Landing _Outcome] ORDER BY TOTAL_OUTCOME DESC;

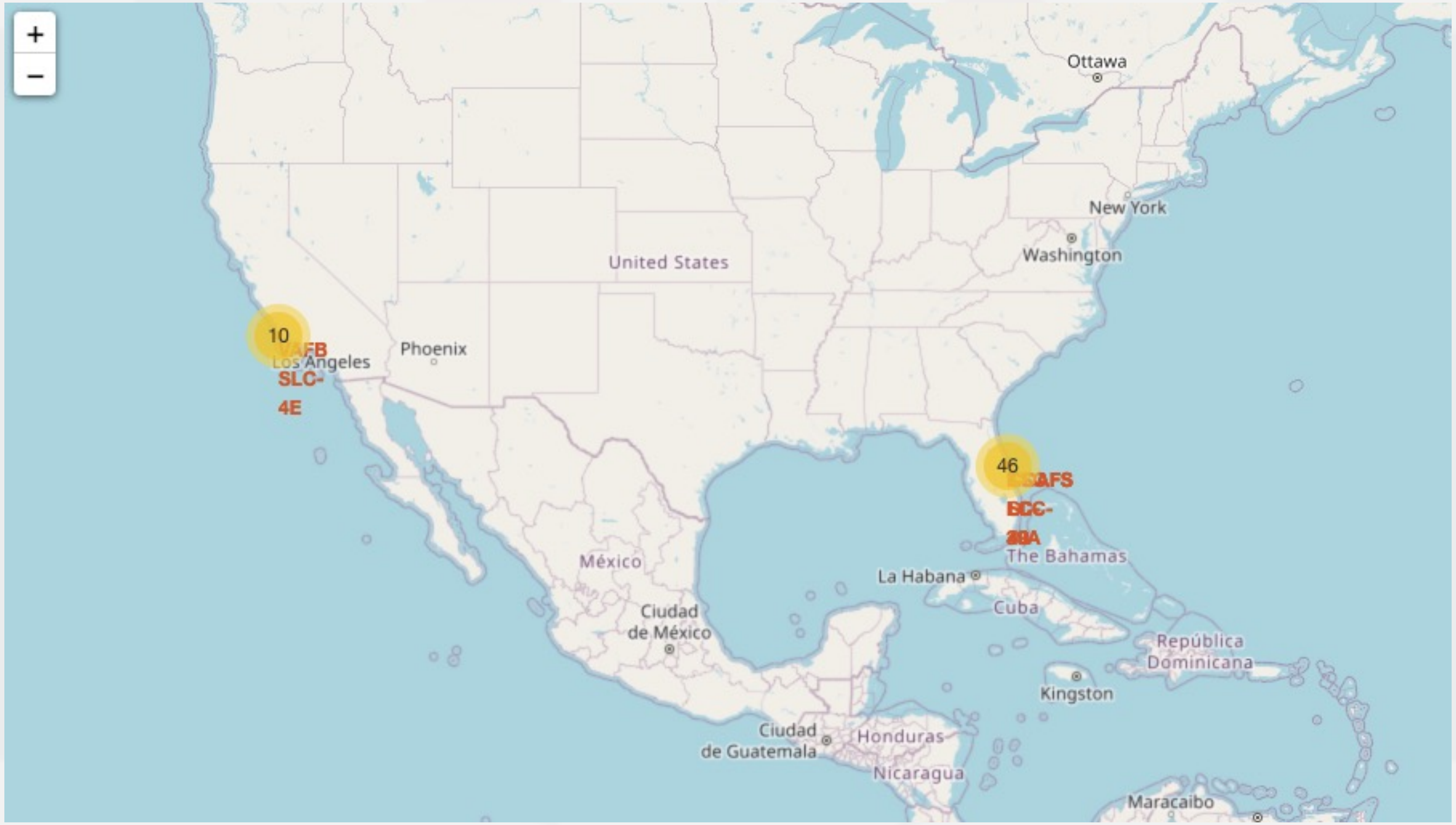
Landing _Outcome	TOTAL_OUTCOME
Success	20
Success (drone ship)	8
Success (ground pad)	6

- All

%%sql SELECT [Landing _Outcome], COUNT([Landing _Outcome]) AS TOTAL_OUTCOME
FROM SPACEXTBL
WHERE DATE BETWEEN '04-06-2010' and '20-03-2017'
GROUP BY [Landing _Outcome]
ORDER BY TOTAL_OUTCOME DESC;

Landing _Outcome	TOTAL_OUTCOME
Success	20
No attempt	10
Success (drone ship)	8
Success (ground pad)	6
Failure (drone ship)	4
Failure	3
Controlled (ocean)	3
Failure (parachute)	2
No attempt	1

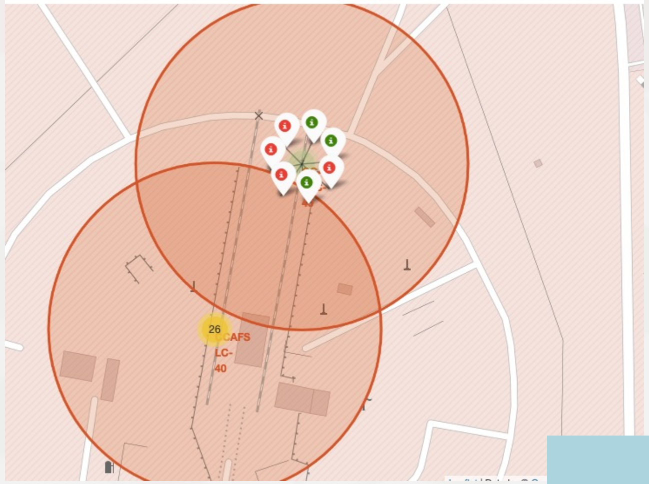
FOLIUM: LAUNCH SITES LOCATIONS

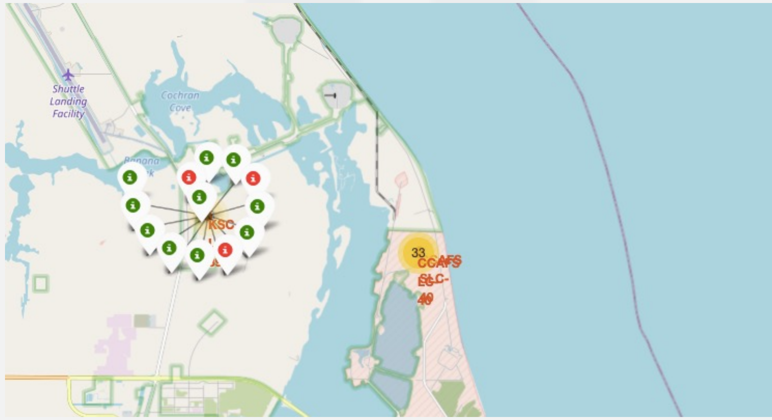


IBM Developer

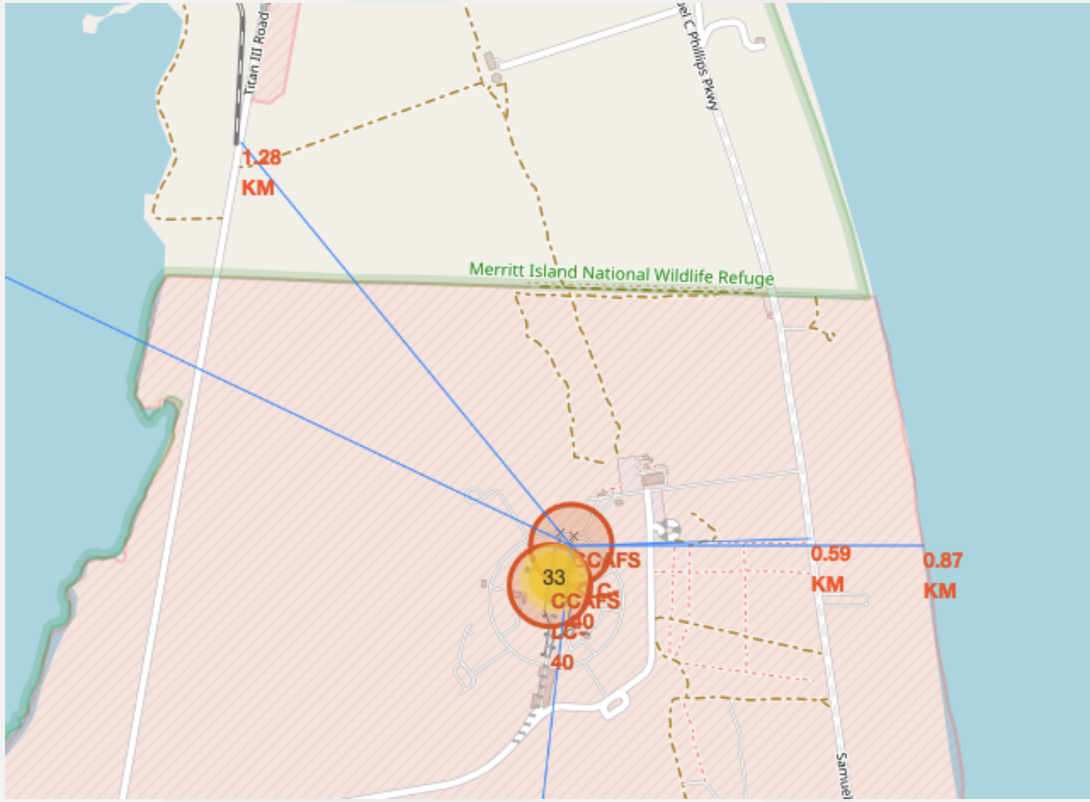
SKILLS NETWORK

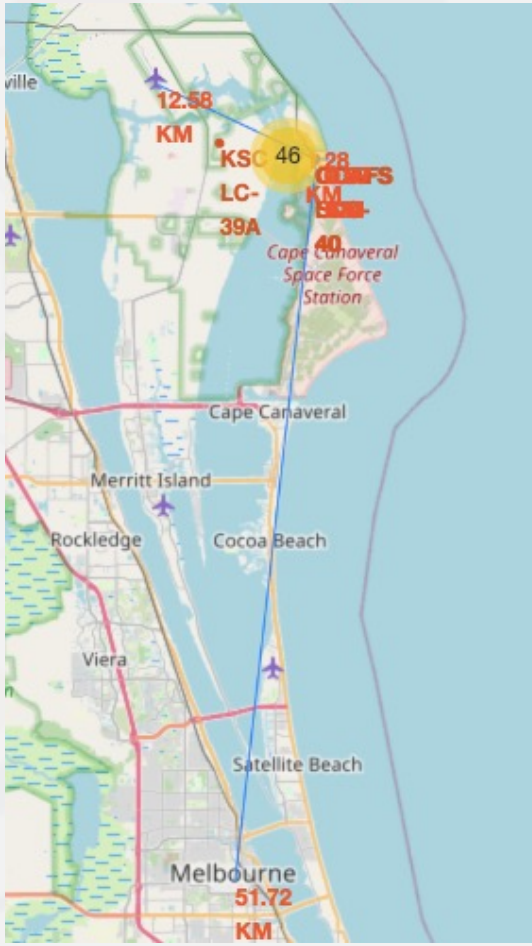
FOLIUM: LAUNCH-OUTCOMES





FOLIUM: PROXIMITY OF LAUNCH SITES





IBM Developer

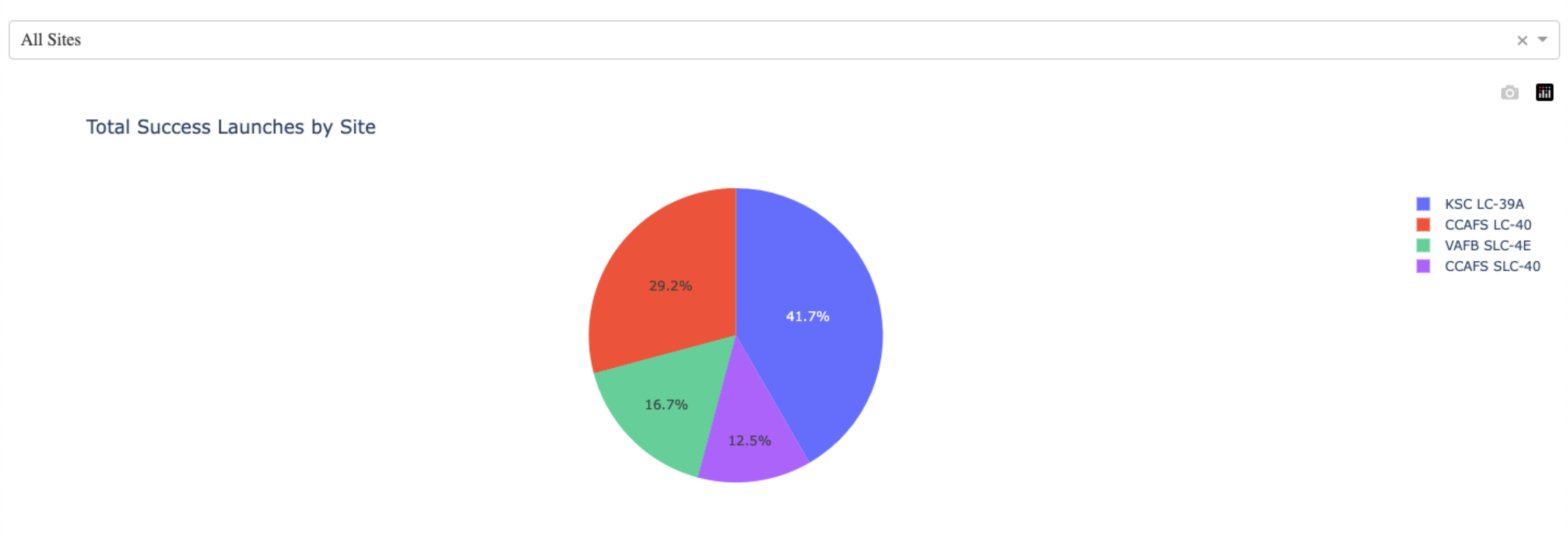
SKILLS NETWORK

FOLIUM: ANSWER QUESTIONS

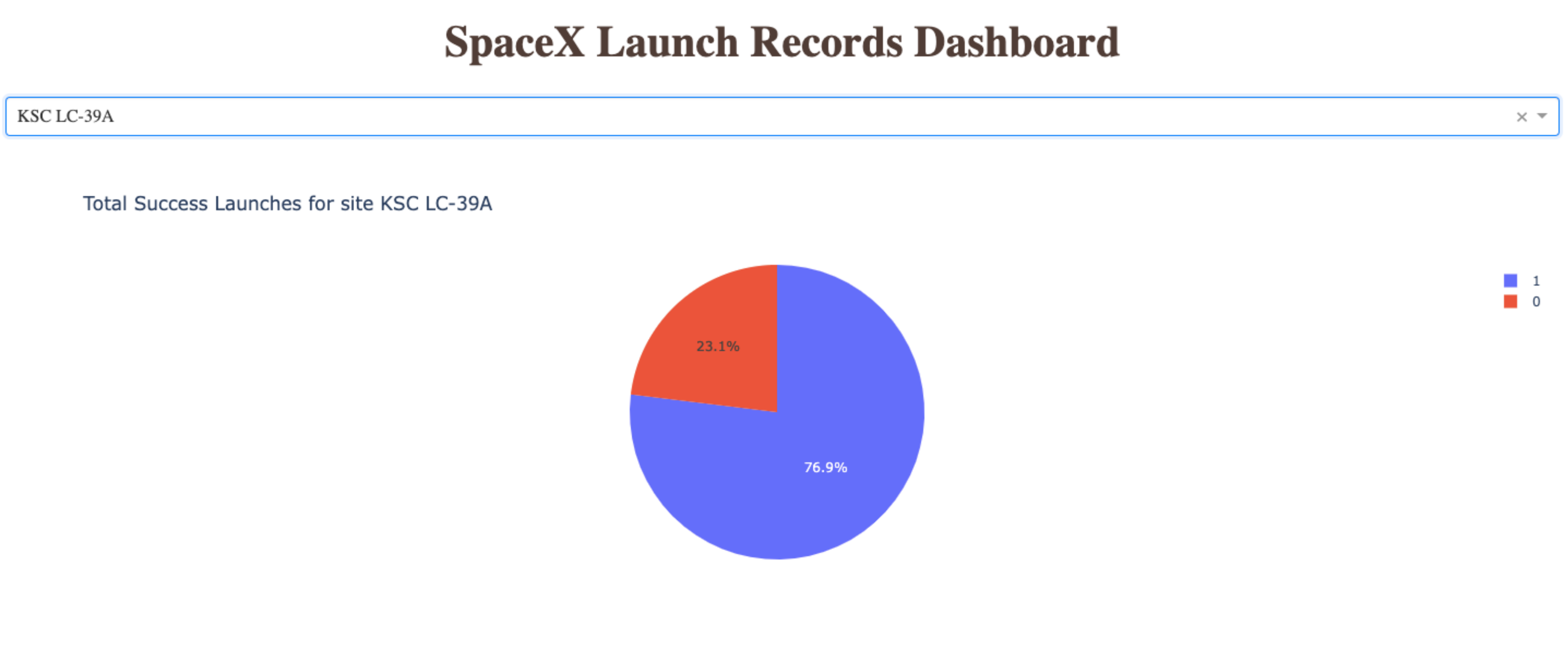
Based on the previous slide we can answer the following questions:

- Are launch sites in close proximity to railways? **Yes 1.28km**
- Are launch sites in close proximity to highways? **Yes 0.59km**
- Are launch sites in close proximity to coastline? **Yes 0.87km**
- Are launch sites in close proximity to airports? **Yes, 12.58km**
- Do launch sites keep certain distance away from cities? **Yes 51.72km**

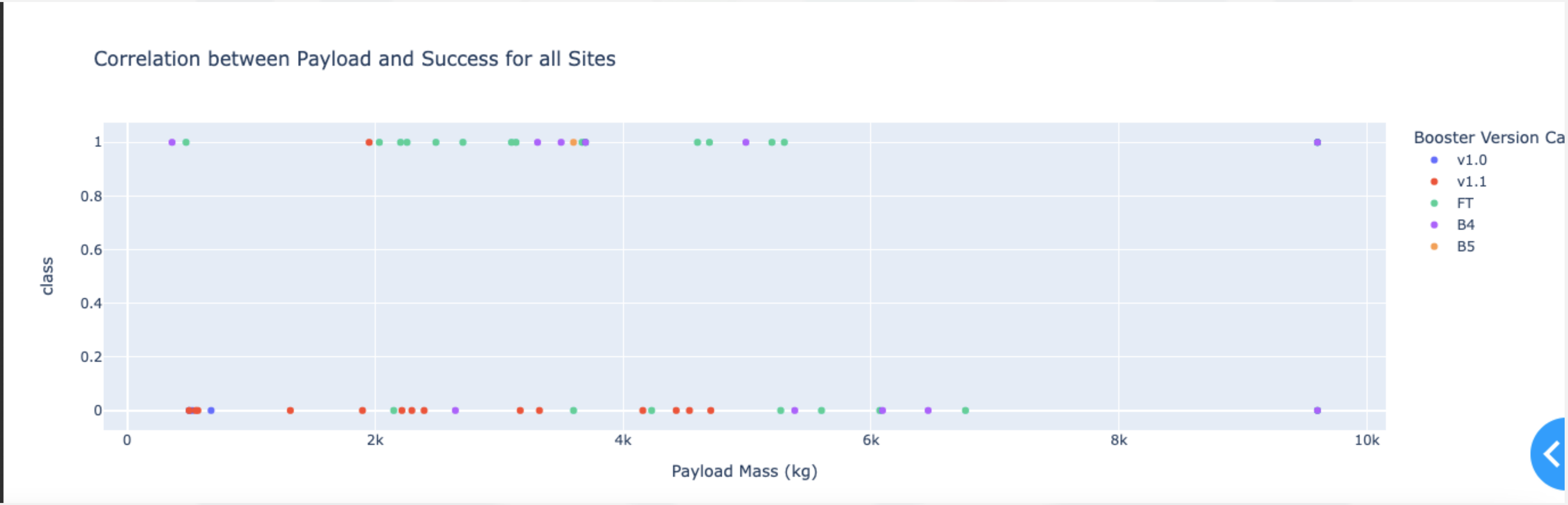
PLOTLY: TOTAL SUCCESS OF LAUNCHES



PLOTLY: HIGHEST SCORING LAUNCH SITE



PLOTLY: PAYLOAD MASS AND LAUNCH OUTCOME (ALL SITES)



PREDICTIVE ANALYSIS: ACCURACY OF PREDICTIONS OF THE MODELS ON TEST DATA

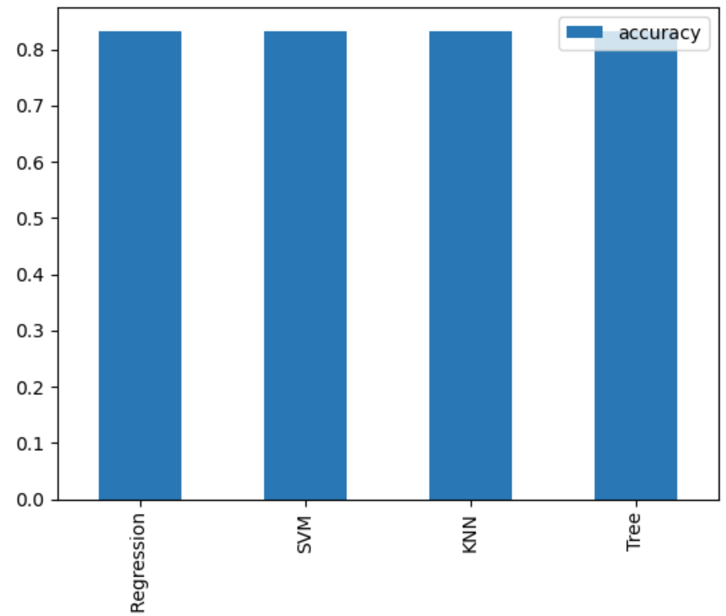
- All algorithms accuracy on test data are similar (0.833)

→But problem is: (!) very small test size (*n* = 18)

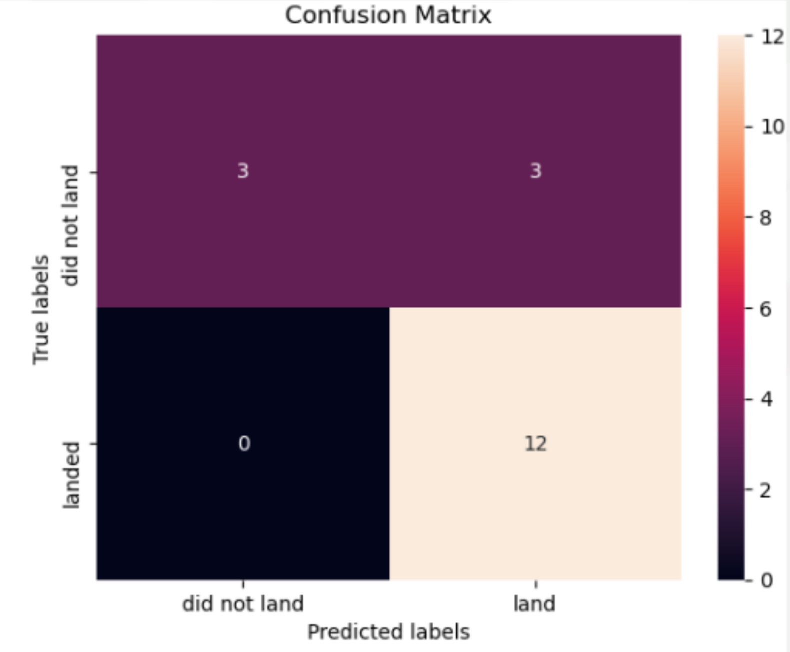
```
df = pd.DataFrame({'algoritm':  ['Logistic Regression','SVM','KNN','Tree'],
                   'accuracy':[logreg_cv.score(X_test,Y_test),
                               svm_cv.score(X_test,Y_test),
                               knn_cv.score(X_test, Y_test),
                               tree_cv.score(X_test, Y_test)]})

df.plot.bar(x='algoritm', y='accuracy')
```

```
<AxesSubplot:xlabel='algoritm'>
```



PREDICTIVE ANALYSIS: CONFUSION MATRIX



- Confusion matrix the same for all models

IBM Developer

SKILLS NETWORK

CONCLUSION

- Total number of successful and failure launches: 101
- First successful launch: 22-12-2015
- Success rate is higher the larger the flight amount at a launch site
- Increasing success rates over the years
- Orbit types that have highest success rates:
  ES-L1, GEO, HEO and SSO
- Launch sites are close to railway, highway, aiport and coastline, while further away from cities
- Launch site KSC LC-39A has most successful launches (76.9%)
- Most flights have a payload mass of approx. 2000-6000
- All developed models deliver the same accuracy in classifying success rates on unseen test data

→ Note: All preliminary conclusions since the dataset is relatively small

APPENDIX

- Github link to code:

THANK YOU!