IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Tomasz Szymański
31-01-2022

Outline

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

Executive Summary

- Summary of methodologies

  a) data collection:

  - data scraping of the Space X launches of Wikipedia webpage,

  - Space X public API,

  b) Exploratory Data Analysis : SQL queries, Matplotlib and Seaborn graphs, an interactive Plotly dashboards and Folium maps

  c) few various machine learning models (logistic regression, decision tree, KNN, SVN) to predict successful landings

- Summary of all results

  a) All models give similar results about 0,78

  b) Seems that more data is required to improve accuracy as seems that results are overpredicted.

Introduction

- Project background and context
  - Space exploration became more and more common,
  - Space X becomes the cheapest way to launch rackets to space, as the Space X reuse the Stage 1 (part) of rocket.
  - Space X launch costs 62 millions $ while others cost 165 millions $
- Problem:
  - How to predict if launch will be completed successfully,
  - What machine learning models should be used,
  - What data are necessary to predict launch correctly

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was got from Space X and wikipedia
- Perform data wrangling
  - Pandas and NumPy ware used to clean and analyse 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
  - Scikit-Learn was used to create Machine Learning models and GridSearchCV for tuning

Data Collection

- SpaceX API Calls:
  - Space X has public API, where you can find information about launches coordinates, rocket, Payload mass, outcome of the mission etc.
  - to obtain data from HTTP web page, get request was used.
- Scraping SpaceX launch data Wikipedia page
  - Wikipedia contains data from past SpaceX launches.
  - to get these data, Beautiful Soup library was used.

Data Collection – SpaceX API

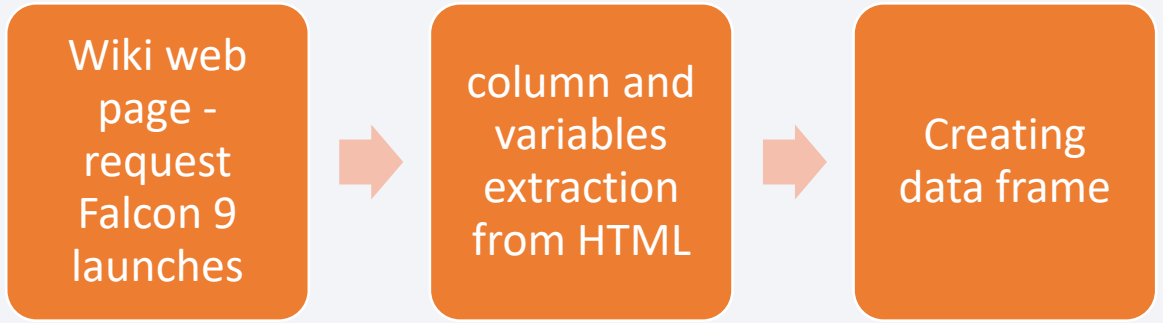
- GitHub URL of SpaceX API calls notebook:

https://github.com/monas1975/IBM_data_science_spaceX_project/blob/main/Data_Collect_API.ipynb

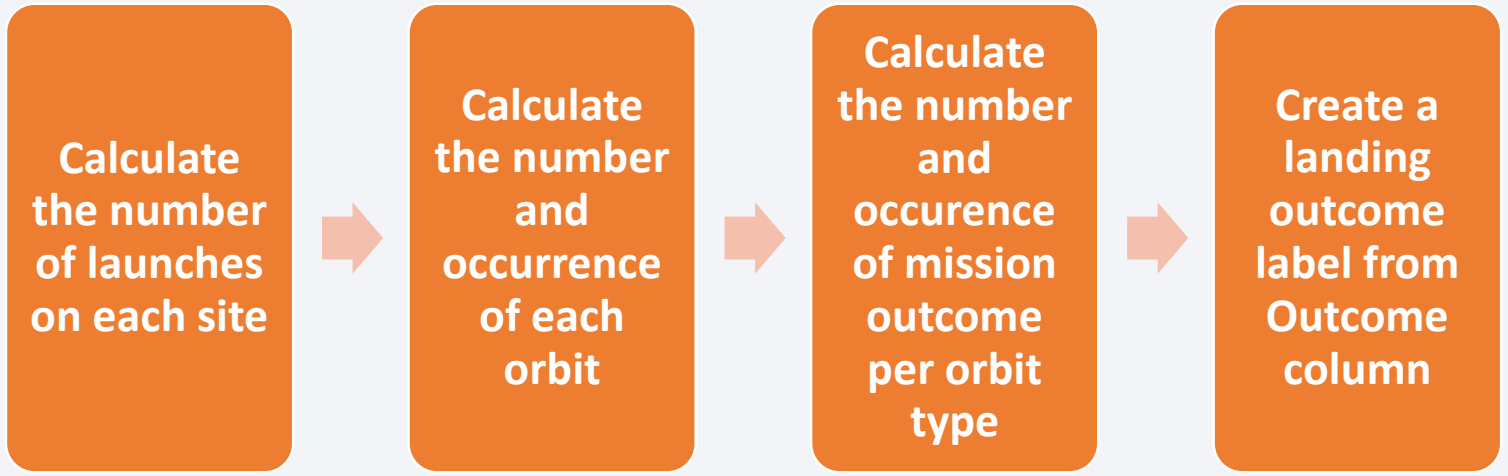
Data Collection - Scraping

- GitHub URL of SpaceX API calls notebook:

https://github.com/monas1975/IBM_data_science_spaceX_project/blob/main/Data_collection_with_Web_Scrapping.ipynb



Data Wrangling



- GitHub URL:

https://github.com/monas1975/IBM_data_science_spaceX_project/blob/main/Data_wrangling.ipynb

EDA with Data Visualization

- Graph drawn:
    - FlightNumber vs. PayloadMass (scatterplot),
    - FlightNumber vs LaunchSite (scatterplot),
    - Payload vs Launch Site (scatterplot),
    - relationship between success rate and orbit type (barplot),
    - FlightNumber and Orbit type (scatterplot),
    - Payload and Orbit type (scatterplot),
    - launch success yearly trend (lineplot)
- GitHub URL:
  https://github.com/monas1975/IBM_data_science_spaceX_project/blob/main/EDA_with_Visualization_lab.ipynb

EDA with SQL

- summarize the SQL queries performer:

  - the names of the unique launch sites in the space mission,
  - records (5) where launch sites begin with the string 'CCA',
  - total payload mass carried by boosters launched by NASA (CRS),
  - average payload mass carried by booster version F9 v1.1,
  - date when the first successful landing outcome in ground pad was achieved,
  - 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)) between the date 2010-06-04 and 2017-03-20, in descending order

- GitHub URL: https://github.com/monas1975/IBM_data_science_spaceX_project/blob/main/Exploratory_Analysis_with_SQL.ipynb

Build an Interactive Map with Folium

- Elements added to map:
    - circle and marker for each launch site, added as point of attention at map,
    - marker cluster – easy way to show number of launches at each site,
    - mouse position- to get coordinates,
    - polly lines (lines between site and coastline)

- GitHub:

https://github.com/monas1975/IBM_data_science_spaceX_project/blob/main/Interactive_Visual_Analytics_with_Folium_lab.ipynb

Build a Dashboard with Plotly Dash

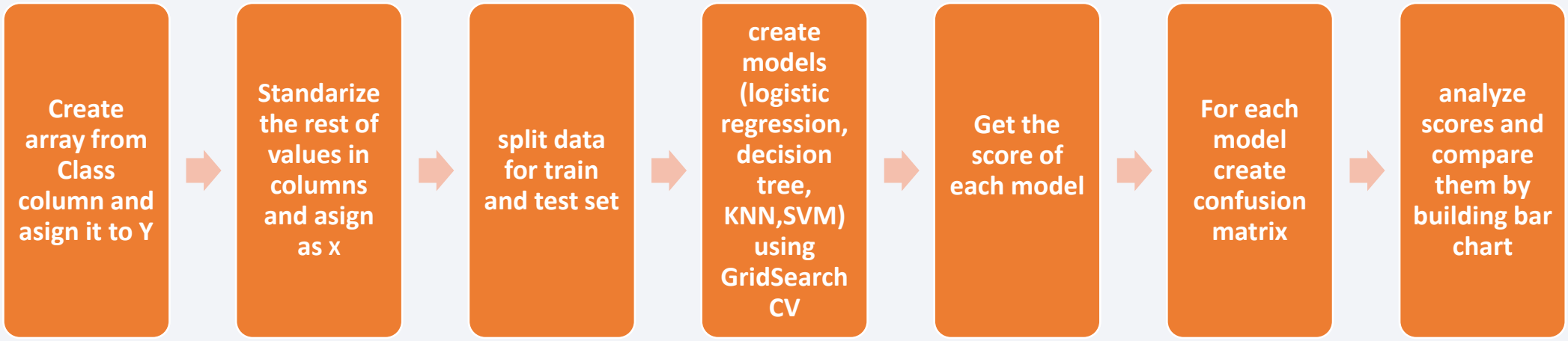
- The purpose of adding interactive dashboard was to give the users opportunity to decide which information would like to see.

- Features add to dashboard:
  - dropdown menu – to let select launch site and information related,
  - interactable bar for payload mass,
  - scatter plot that show correlation between payload mass and landing result,
  - piechart – to show proportion between successful landing and unsuccessful.

- GitHub URL:

https://github.com/monas1975/IBM_data_science_spaceX_project/blob/main/spacex_dash_app.py

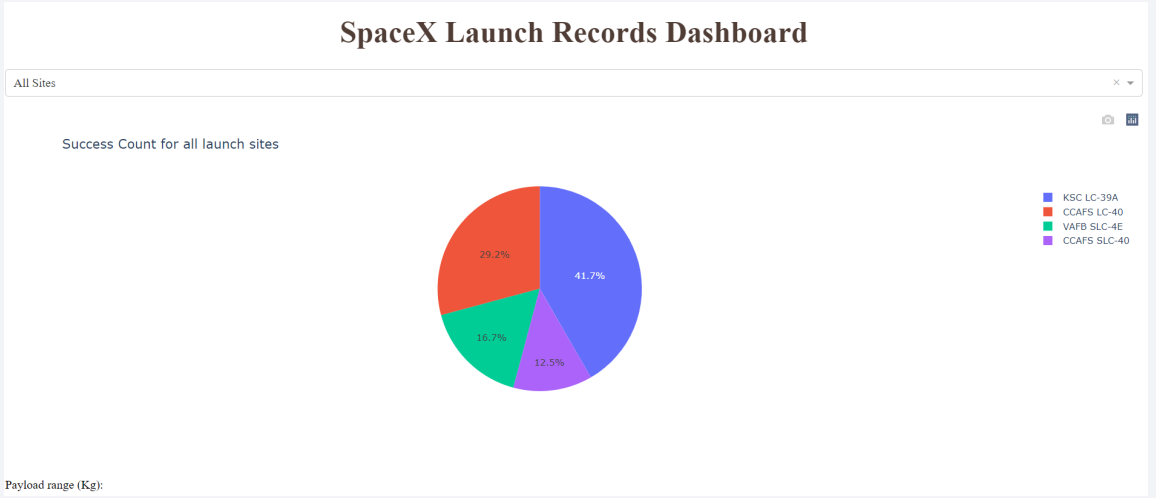
Predictive Analysis (Classification)

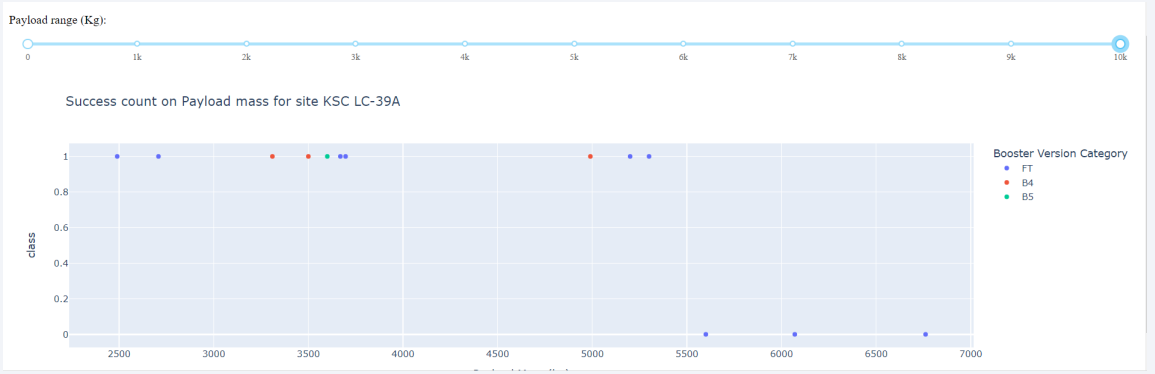


- GitHub URL: https://github.com/monas1975/IBM_data_science_spaceX_project/blob/main/Machine_Learning_Prediction_lab.ipynb

Interactive analytics demo in screenshots

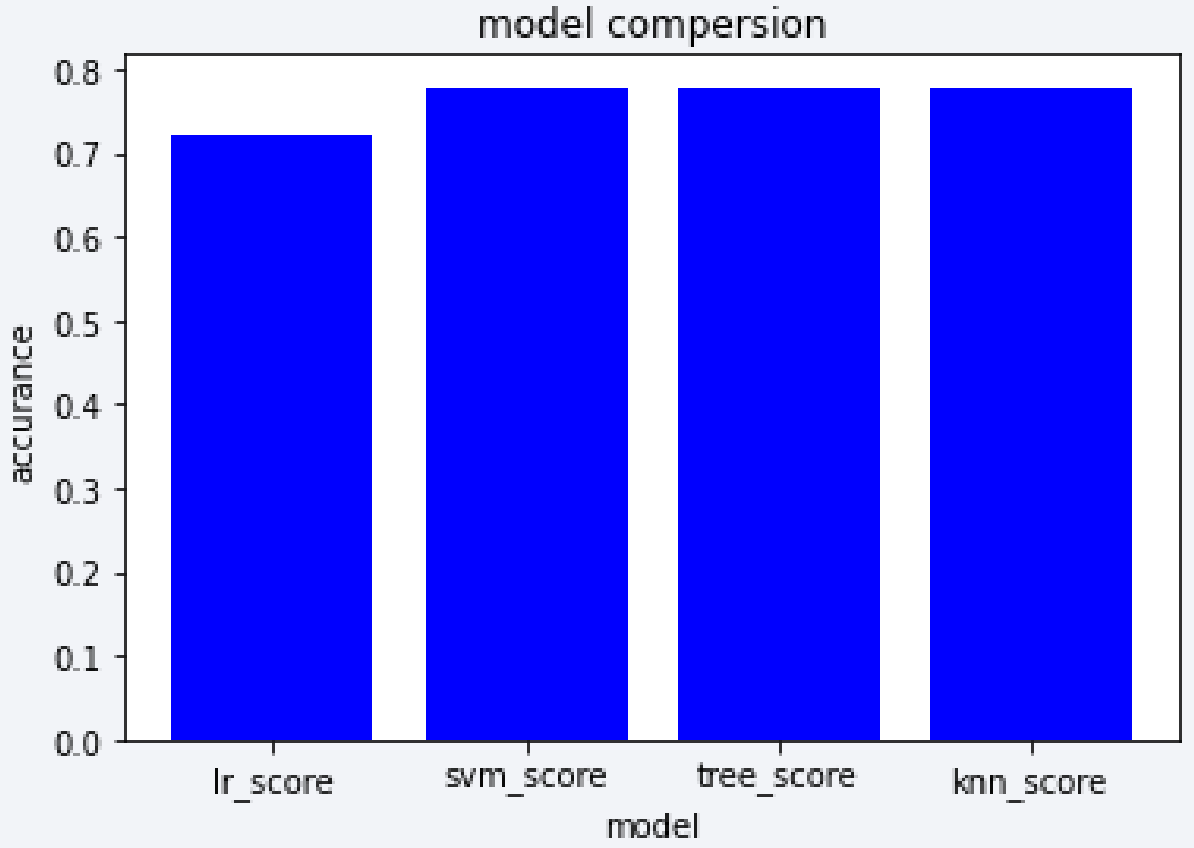
- Results from interactive dashboard depends on users' inputs,
- Pie chart shows proportion between successful and unsuccessful landings for sites,
- Scatter plot shows how payload mass impact on outcome of landings.

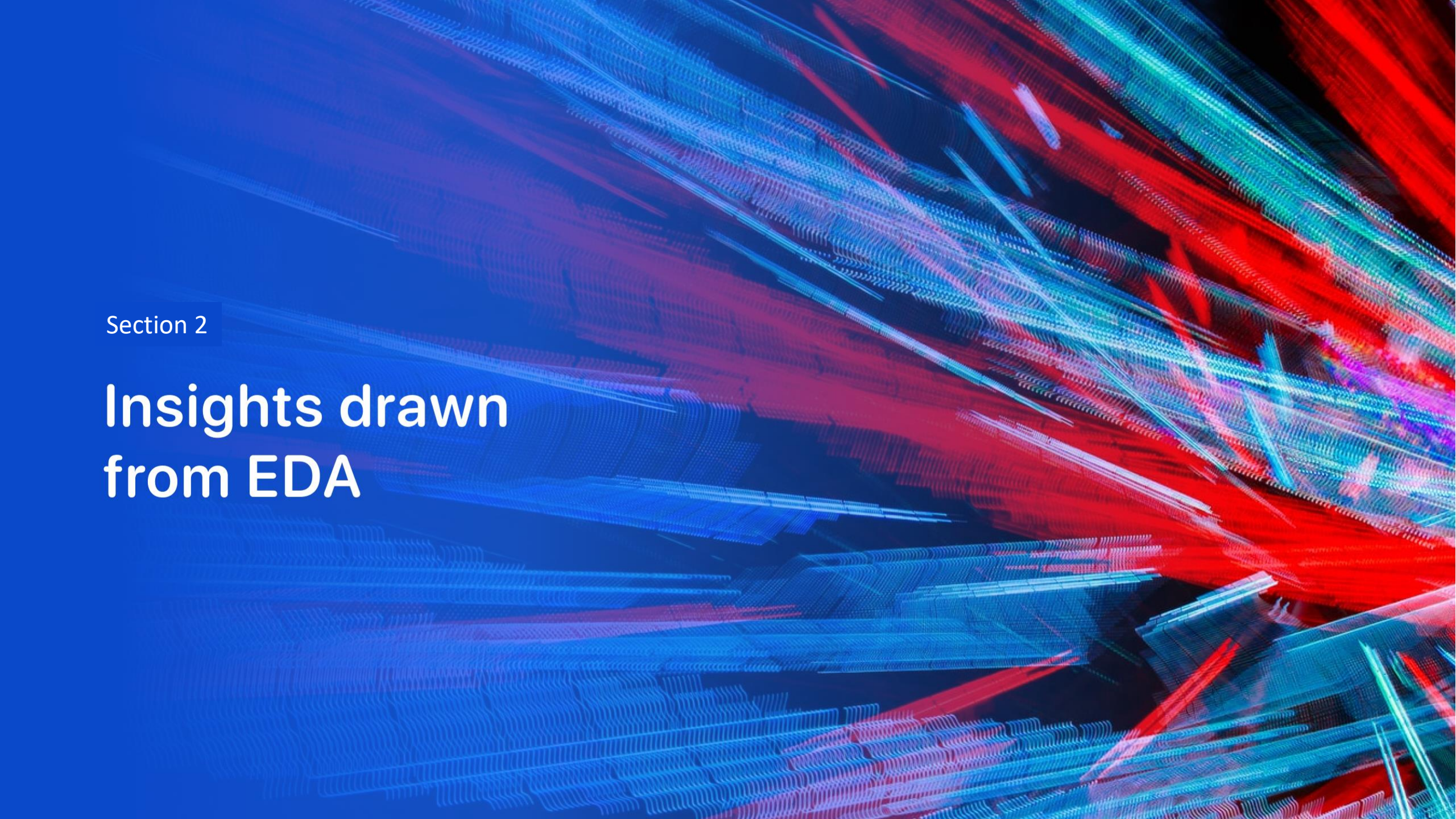




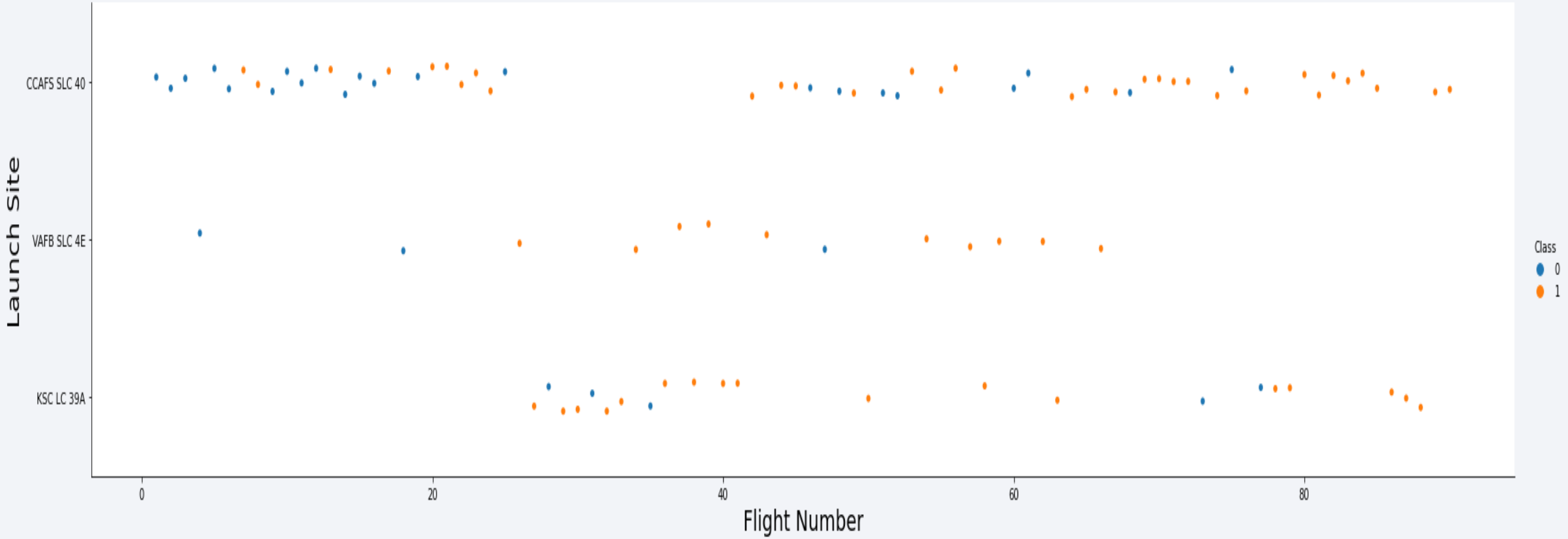
Predictive analysis results

- Predictive analysis results : all models have similar accuracy 0.78, but logistic regression has the worst 0.72



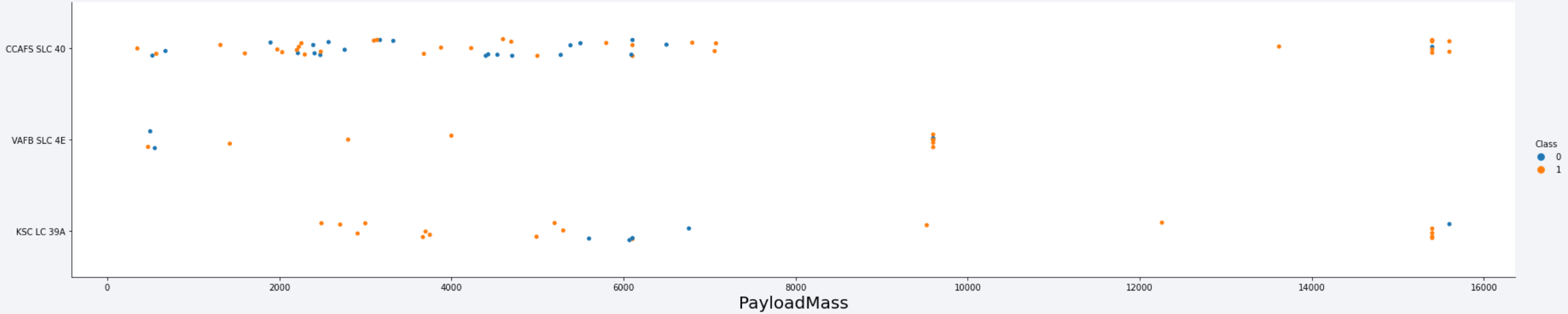

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



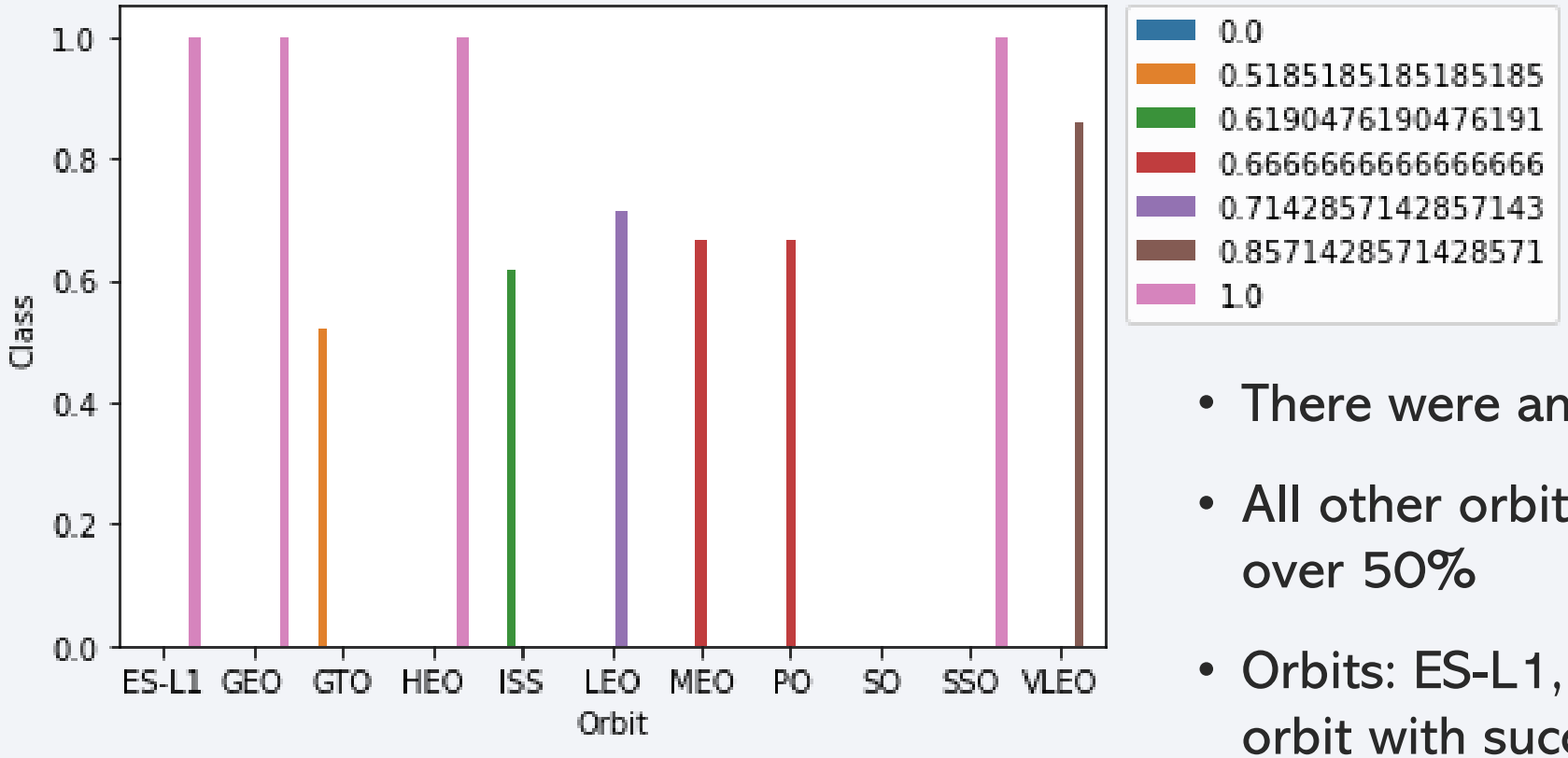
- “CCAFS SLC 40” is the most often used platform. Amount of successful landing increase as flight numbers increases

Payload vs. Launch Site



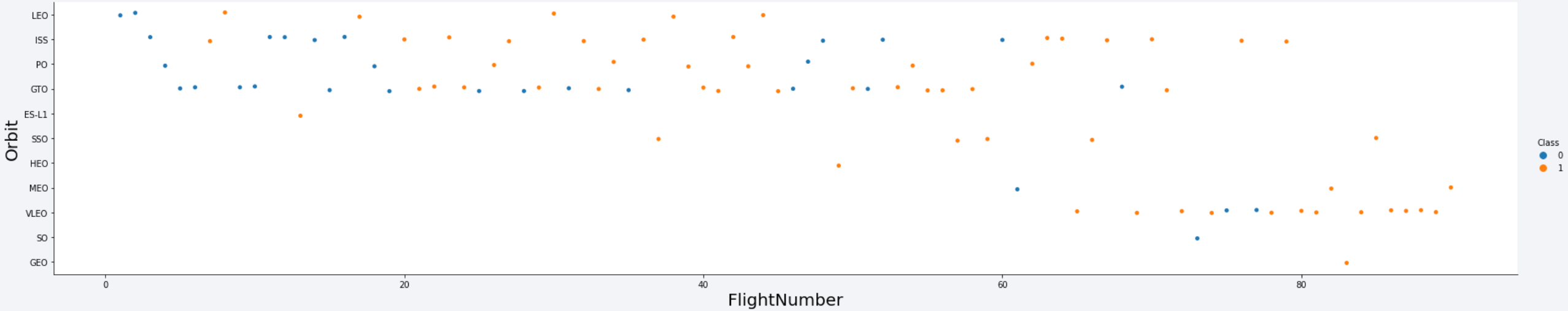
- Most of the launches have payload below 7500kg
- Most of the launches with payload over 8000 were successful

Success Rate vs. Orbit Type



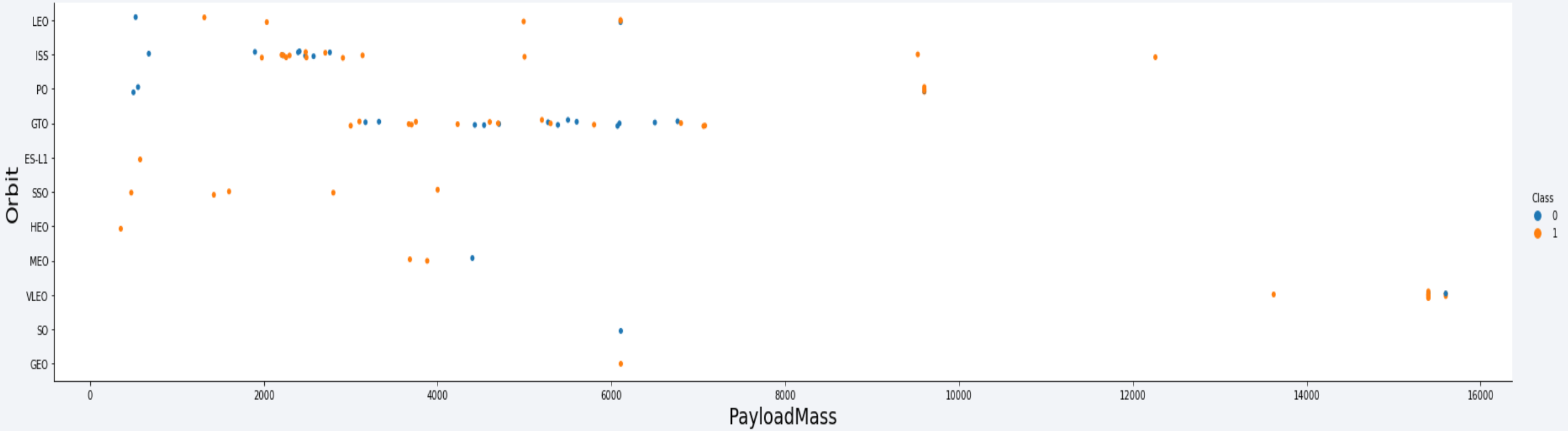
- There were any launch to SO orbit,
- All other orbits have success rate over 50%
- Orbits: ES-L1, GEO,HEO,SSO are the orbit with success rate =100%
- GTO is the orbit with the lower success rate

Flight Number vs. Orbit Type



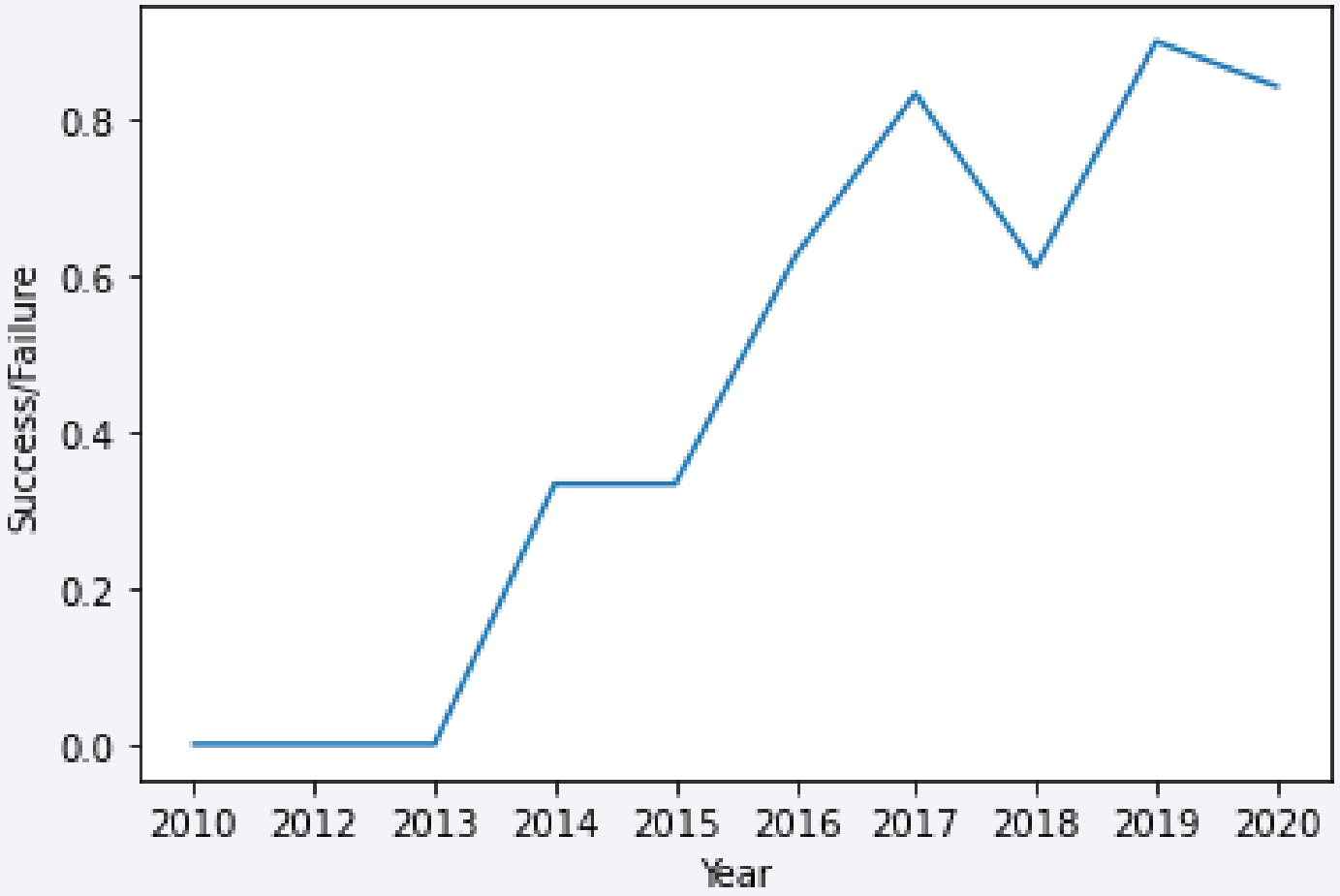
- There was an orbit type changes with flight number increase

Payload vs. Orbit Type



- There is the correlation between the payload and orbit chosen.
- The launches with very high payload mass was to VLEO orbit

Launch Success Yearly Trend



- Success rate increase year by year
- Launches between 2010 and 2013 completed unsuccessfully
- The maximum of success rate was in 2019
- In 2018 the success rate drop down to 0.6

All Launch Site Names

Display the names of the unique launch sites in the space mission

In [12]:
```
%sql select unique launch_site from RHF71790.SPACEXDATASE;
```

```
 * ibm_db_sa://rhf71790:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536/bludb
Done.
```

Out[12]:

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

- In dataset there are 4 unique launch sites names

Launch Site Names Begin with 'CCA'

```
In [20]: %sql SELECT * FROM  RHF71790.SPACEXDATASE WHERE launch_site like 'CCA%' LIMIT 5;
 * ibm_db_sa://rhf71790:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536/bludb
Done.
```

Out[20]:

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	None	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	None	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	None	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	F9 v1.0 B0006	CCAFS LC-40	None	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	None	677	LEO (ISS)	NASA (CRS)	Success	No attempt

- 5 records where launch sites begin with `CCA`

Total Payload Mass

```
In [23]: %sql SELECT SUM(payload_mass__kg_) FROM RHF71790.SPACEXDATASE WHERE customer LIKE 'NASA (CRS)' ;
          * ibm_db_sa://rhf71790:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536/bludb
         Done.
Out[23]:  1
          45596
```

- The total payload carried by boosters from NASA is 45596 kg

Average Payload Mass by F9 v1.1

```
In [24]: %sql SELECT AVG(payload_mass__kg_) FROM RHF71790.SPACEXDATASE WHERE booster_version LIKE 'F9 v1.1%' ;
 * ibm_db_sa://rhf71790:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536/bludb
Done.
Out[24]:
1
2534
```

- The average payload mass carried by booster version F9 v1.1 is 2534 kg

First Successful Ground Landing Date

```
In [25]: %sql SELECT MIN(date) FROM RHF71790.SPACEXDATASE WHERE landing__outcome LIKE 'Success (ground pad)';
          * ibm_db_sa://rhf71790:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536/bludb
         Done.
```

Out[25]:

1
2015-12-22

- The first successful landing outcome on ground pad was on 22nd of December 2015

Successful Drone Ship Landing with Payload between 4000 and 6000

```
In [34]: %sql SELECT DISTINCT booster_version FROM RHF71790.SPACEXDATASE WHERE landing__outcome LIKE 'Success (drone ship)' AND (payload_mass__kg_>4000 AND payload_mass__kg_<6000);
 * ibm_db_sa://rhf71790:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536/bludb
Done.
```

Out[34]:

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

- There are four boosters version with payload mass between 4000 and 6000 kg that have successfully landed on drone ship

Total Number of Successful and Failure Mission Outcomes

```
In [33]: %sql SELECT mission_outcome, COUNT(mission_outcome) FROM RHF71790.SPACEXDATASE GROUP BY mission_outcome;
 * ibm_db_sa://rhf71790:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536/bludb
Done.
```

Out[33]:

mission_outcome	2
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

- There were 99 successful mission and only 1 failure.

Boosters Carried Maximum Payload

```
In [36]: %sql SELECT booster_version FROM RHF71790.SPACEXDATASE WHERE payload_mass__kg_ = (SELECT MAX(payload_mass__kg_) FROM RHF71790.SPACEXDATASE );
 * ibm_db_sa://rhf71790:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536/bludb
Done.
```

Out[36]:

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

- List the names of the booster versions which have carried the maximum payload mass

2015 Launch Records

```
In [39]: %sql SELECT date, booster_version, launch_site, landing__outcome FROM  RHF71790.SPACEXDATASE WHERE landing__outcome LIKE 'Failure (drone ship)' And YEAR(DATE) =2015;
 * ibm_db_sa://rhf71790:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536/bludb
Done.
```

Out[39]:

DATE	booster_version	launch_site	landing__outcome
2015-01-10	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

- The failed landing outcomes in drone ship in 2015

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

```
In [40]: %sql SELECT landing__outcome, COUNT(*) FROM  RHF71790.SPACEXDATASE WHERE DATE > '2010-06-04' AND DATE < '2017-03-20' GROUP BY landing__outcome ORDER BY landing__outcome;
 * ibm_db_sa://rhf71790:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536/bludb
Done.
```

Out[40]:

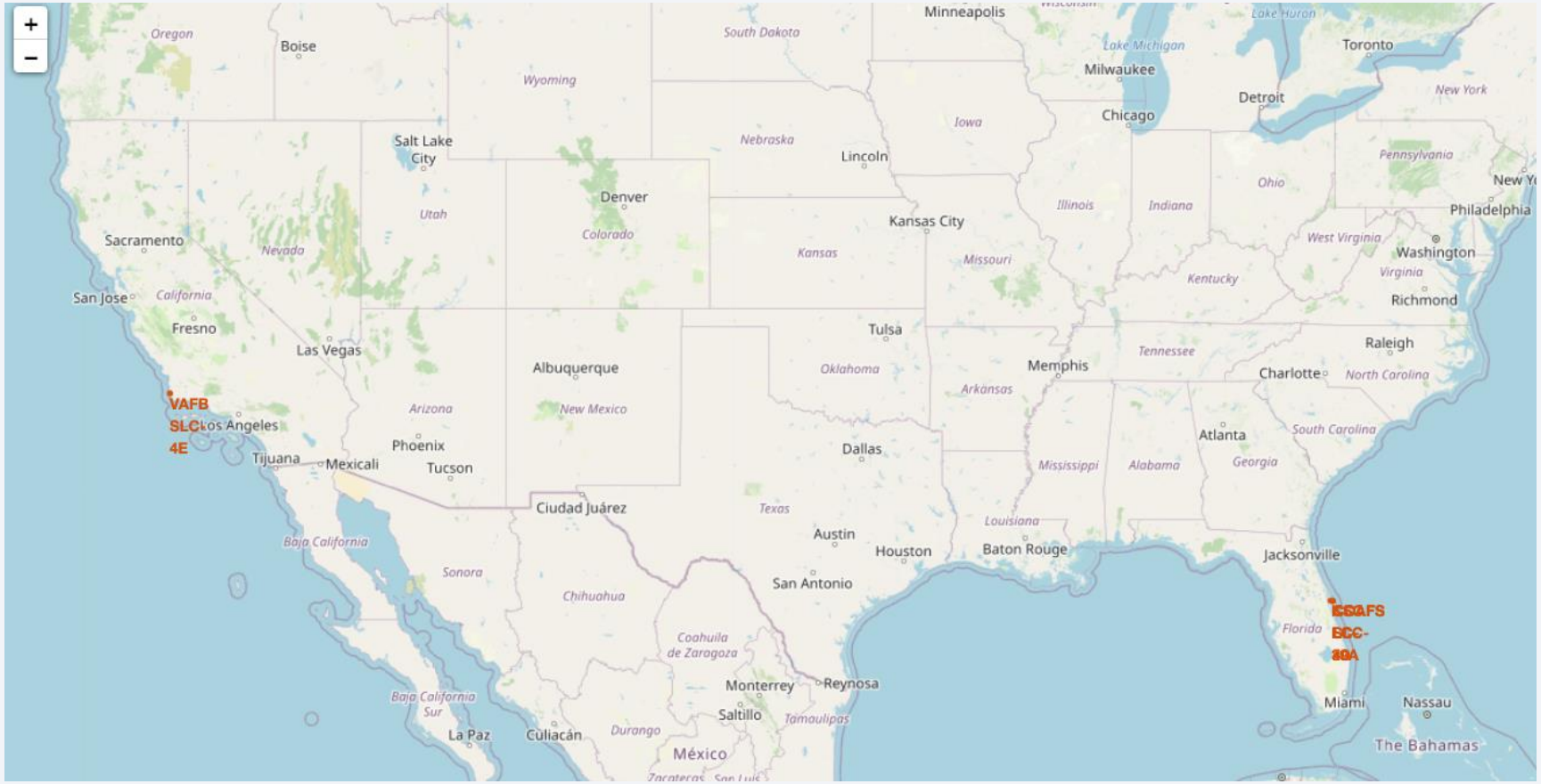
landing__outcome	2
Controlled (ocean)	3
Failure (drone ship)	5
Failure (parachute)	1
No attempt	10
Precluded (drone ship)	1
Success (drone ship)	5
Success (ground pad)	3
Uncontrolled (ocean)	2

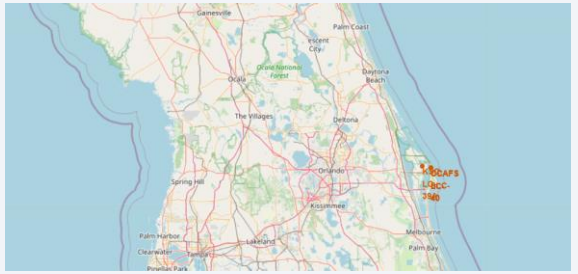
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20

Section 3

Launch Sites Proximities Analysis

Location of launch sites



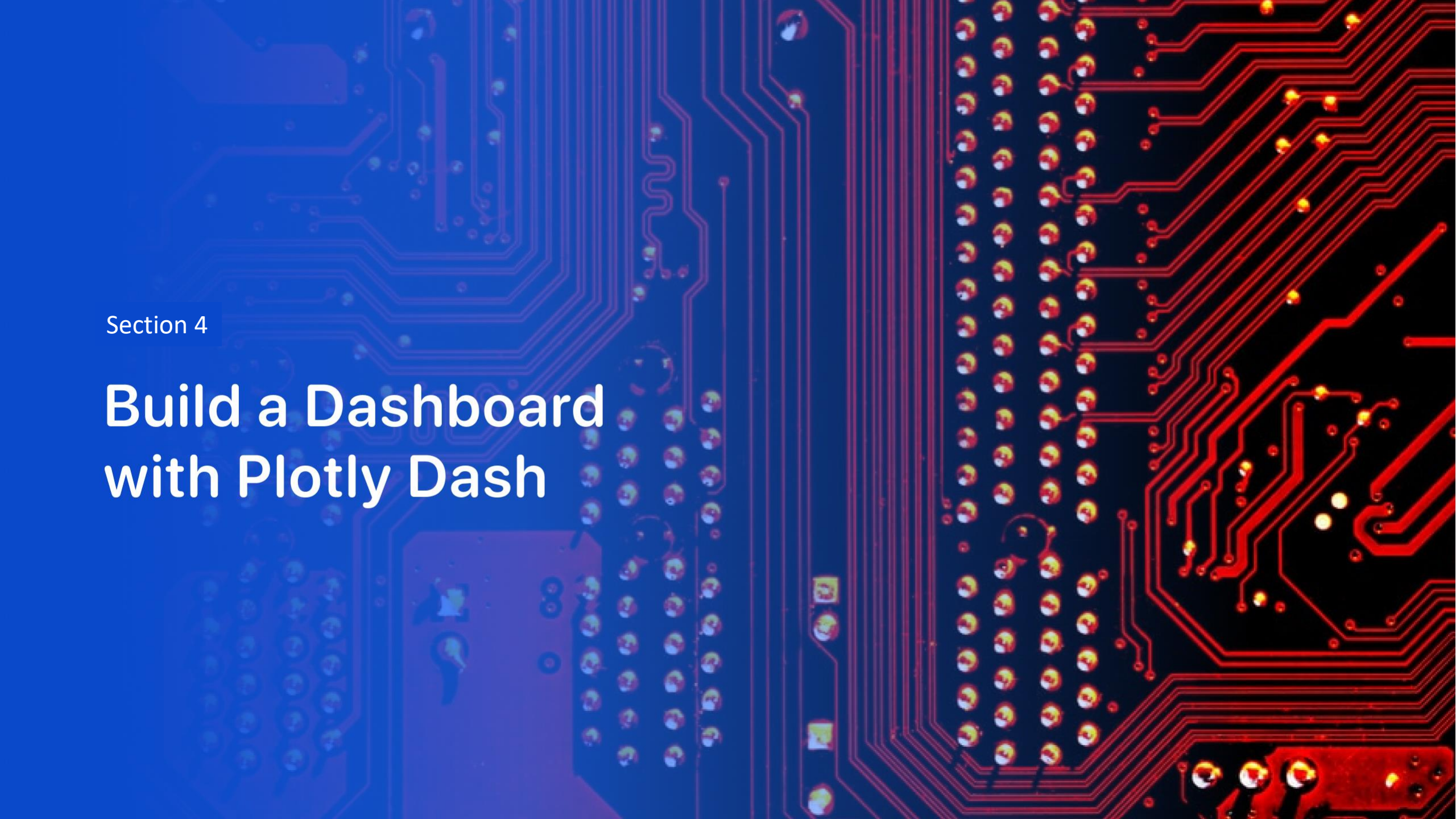


- All launch sites are located at coastline

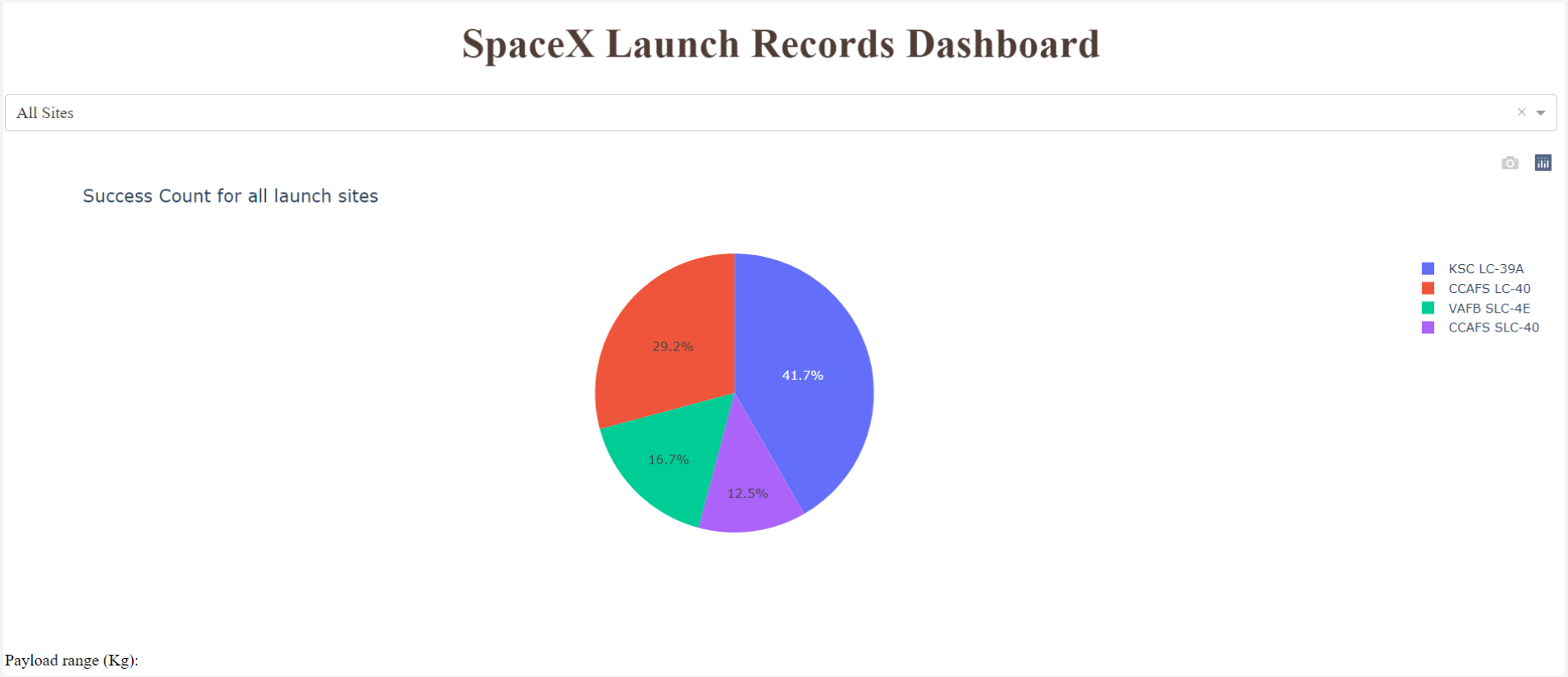
Folium Map – landing outcomes

- Florida – on the right screen shoot we see: 3 successful outcomes and 4 unsuccessful

Folium Map – distance to coastline

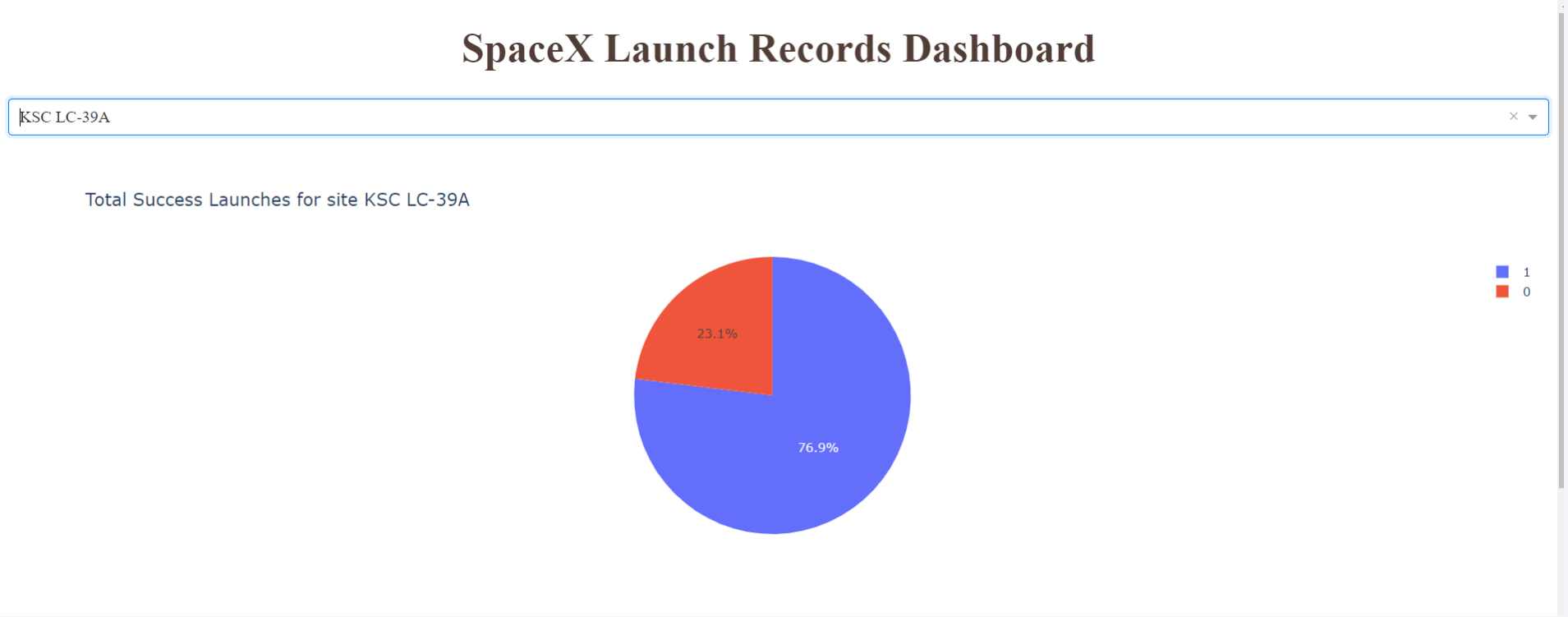

Section 4
Build a Dashboard with Plotly Dash

Successful lunches –all sites



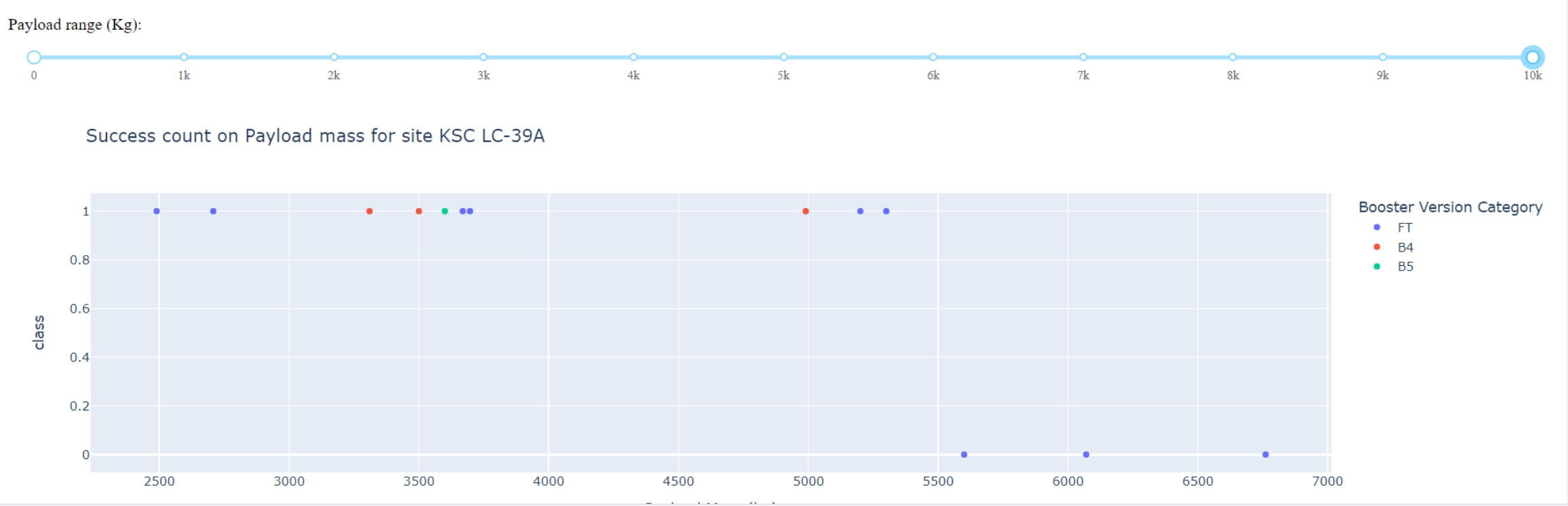
KSC LC-39A is the site with the highest number of successful launches ~42%

KSC LC-39A – site with highest success ratio



~77% of launches at KSC LC-39A was complited successfuly.

Payload impact on successful launches



Payload mass over 5500kg has negative impact on successful launches, any completed successfully.

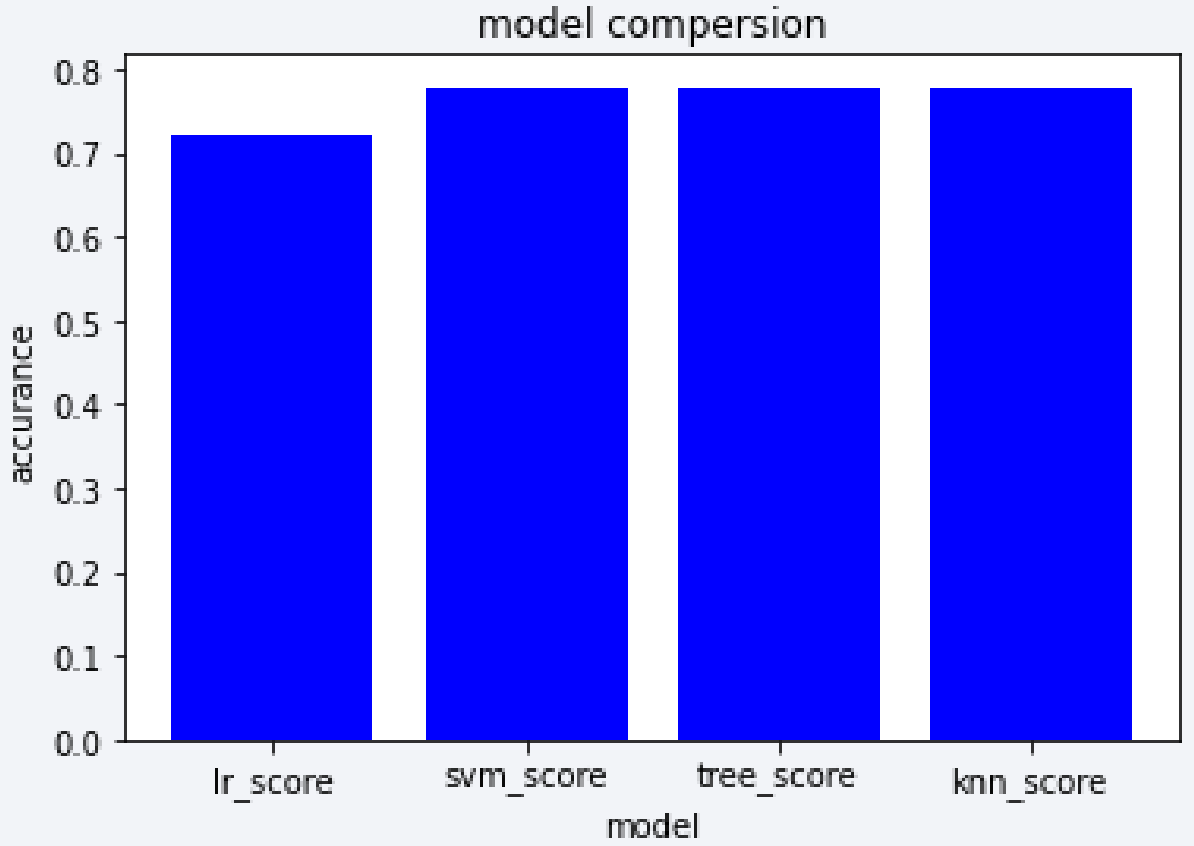
In payload range between 2500 – 5500 kg, the biggest number of successful launches was for FT booster.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

- Models SVM, Tree and knn have same accuracy, equal 0.78
- Model lr (logistic regression) has the worse accuracy 0.72
- All models have quite high accuracy



Confusion Matrix

- Knn - confusion matrix,
- All unsuccessfull landing (4) was predicted correctly,
- 4 successfull landing was predicted wronlgy,
- 10 successfull landings was predicted correctly

Conclusions

- The purpose of this project was to predict if the Stage 1 of the Falcon rocket land successfully or unsuccessfully,
- These information are required by Space Y (competitor of Sapce X) that want to send rocket cheaper than Space X,
- Launches data was received from Wikipedia and API of Space X
- Models build in this project are able to predict outcome of landings with accurance about 0.78

Appendix

- Git hub:

https://github.com/monas1975/IBM_data_science_spaceX_project

- IBM data science coure by Coursera:

https://www.coursera.org/professional-certificates/ibm-data-science

Thank you!