IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Abderrahim SEBBAR>
<09.10.2021>

Outline

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

Executive Summary

- We will determine if the first stage will land and we will determine the cost of a launch :
- We use data analysis tools (RESTfulAPI,web scraping...) to load a dataset, clean it, and find out interesting insights from it.
- Guide the modeling process by using data visualization skills (SQL,FOLIUM,Plotly Dash…) to visualize data and extract meaningful models.
- As a result of all the above stages, we found that the success of the first stage of the landing was linked
- the mass of the lower payload, the orbit of the launch concerned, the booster version...,
- In addition, we analyzed several predictive models to facilitate machine learning and we found that while several models such as logistic regression, SVM and KNN provide a best precision .

Introduction

- Project background and context :

  A Space Y company wants to know the costs and information needed to bid

  against SpaceX for a rocket launch by studying data from a company Space X.

- Problems you want to find answers :

  Impact of dependants variables (payload mass,Site...) on the success of first stage landing.

  Predict the success rate of any future launch's first stage landing.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected from APIspaceX and from Wikipedia by web scraping .
- Perform data wrangling
  - Data was processed by replacing messing value with mean value
- Perform exploratory data analysis (EDA) using visualization and SQL
  - Use SQL commands to filter, sort, & summarize data to visualize the data and extract meaningful patterns to guide the modeling process.
- Perform interactive visual analytics using Folium and Plotly Dash
  - To visualize and interpret data in real-time.
- Perform predictive analysis using classification models
  - Split data into training data and test data to find the best Hyperparameter for SVM, Class Trees, and Logistic Regression

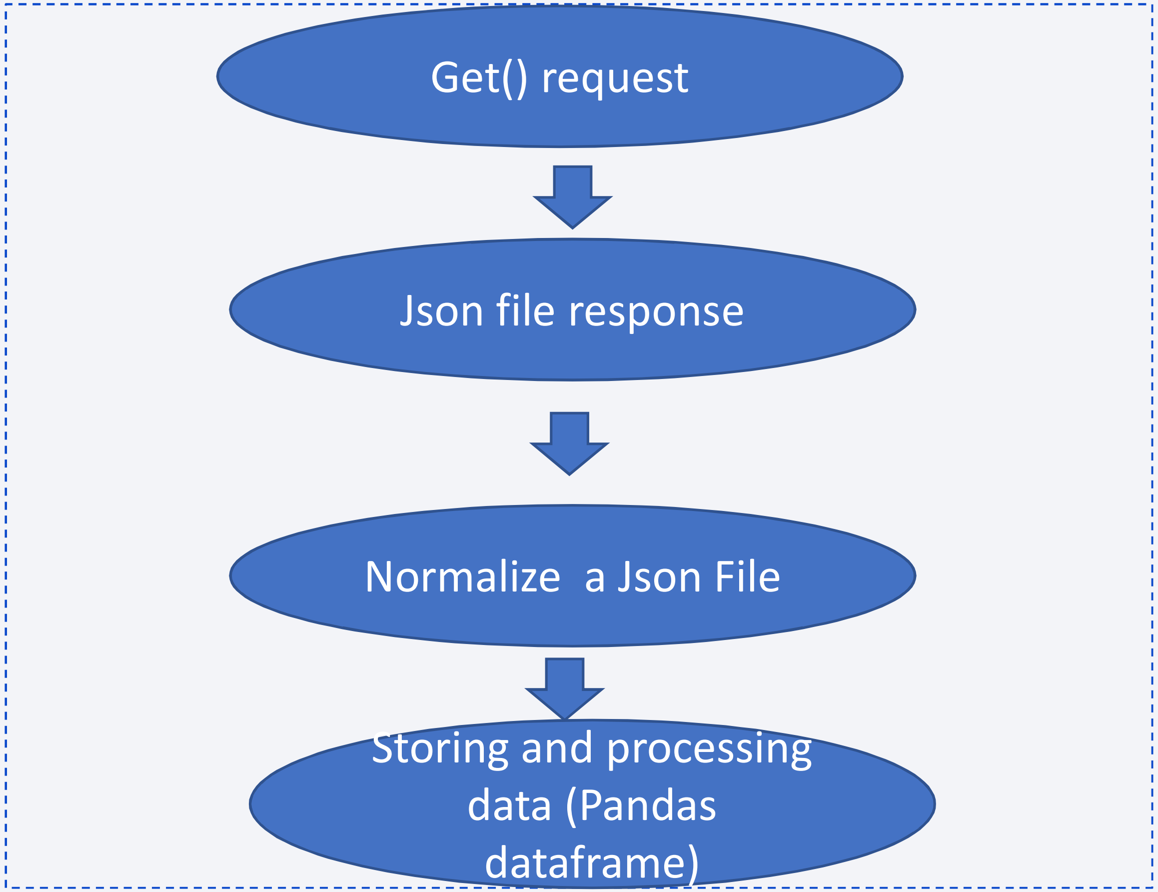
Data Collection

- Describe how data sets were collected.
- You need to present your data collection process use key phrases and flowcharts

Data Collection – SpaceX API

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

- [FirstRepo/jupyter-labs-spacex-data-collection-api.ipynb at master · abderrahim22/FirstRepo (github.com)]



Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts
- [FirstRepo/jupyter-labs-webscraping.ipynb at master · abderrahim22/FirstRepo (github.com)]

Use HTTPGET method to request the HTML page from wikipedia, as an HTTP response

▼

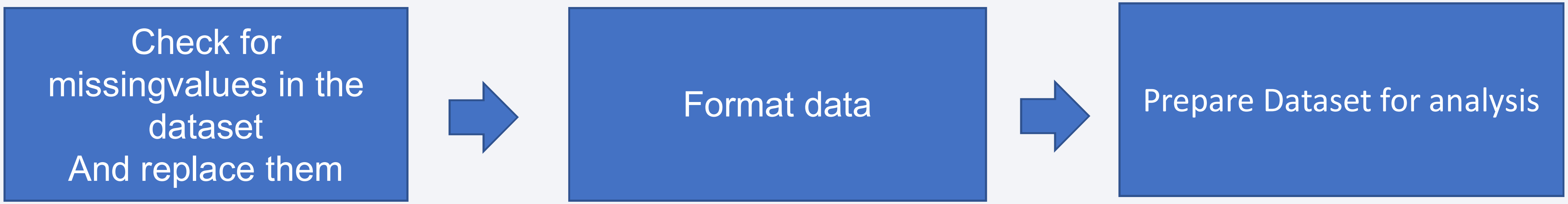
Extract all column from the HTML table via BeautifulSoup object

▼

Create a data frame by parsing the HTML tables

Data Wrangling

- Check for missing values and replace them with correct value
- Format data to improve its processing
- Preparing dataset for analysis



- FirstRepo/labs-jupyter-spacex-Data wrangling.ipynb at master · abderrahim22/FirstRepo (github.com)

EDA with Data Visualization

- Summarize what charts were plotted and why you used those charts
  - Scatter plots : to observe relationships between variables with color coding
  - Bar chart : to show comparisons between categories of data
  - line chart : to visualize the trend of something over time.
- FirstRepo/jupyter-labs-eda-dataviz.ipynb at master · abderrahim22/FirstRepo (github.com)

EDA with SQL

Using bullet point format, summarize the SQL queries you performed

- select Unique launch_site **from SPACEXTBL**
- select * **from SPACEXTBL** where launch_site like 'CCA%' limit 5
- select sum(payload_mass__kg_)**as** Total **from SPACEXTBL** where customer like 'NASA (CRS)‘
- select avg(payload_mass__kg_)**as** averge **from SPACEXTBL** where booster_version like 'F9 v1.1‘
- select min(DATE) **from SPACEXTBL** where landing__outcome like 'Success (ground pad)‘
- select booster_version **from SPACEXTBL** where landing__outcome like 'Success (drone ship)' **and** payload_mass__kg_ between 4000 **and** 6000
- select DISTINCT mission_outcome, COUNT(mission_outcome)**as** Total **from SPACEXTBL** group by mission_outcome
- select booster_version **from SPACEXTBL** where payload_mass__kg_ = (select max(payload_mass__kg_) **from SPACEXTBL**)
- select booster_version **from SPACEXTBL** where payload_mass__kg_ = (select max(payload_mass__kg_) **from SPACEXTBL**)
- select landing__outcome,count(landing__outcome) **as** Count ,RANK() OVER (order by count(landing__outcome)) as ORDER **from SPACEXTBL** WHERE landing__outcome **in** ('Failure (drone ship)','Success (ground pad)') **and** Date between '2010-06-04' **and** '2017-03-20' group by landing__outcome

- FirstRepo/jupyter-labs-eda-sql-coursera.ipynb at master · abderrahim22/FirstRepo (github.com)

Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
  - Circle
  - Marker
  - MarkerCluster
  - PolyLine
  - Mouse Position
- Explain why you added those objects
  - Circle to add a highlighted circle area of sitemap
  - Marker for plotting markers on a map
  - MarkerCluster simply a map containing many markers
  - PolyLine to connect two points in the map
  - Mouse Position to define the coordinate on the map
- FirstRepo/lab_jupyter_launch_site_location_interactive.ipynb at master · abderrahim22/FirstRepo (github.com)

Build a Dashboard with Plotly Dash

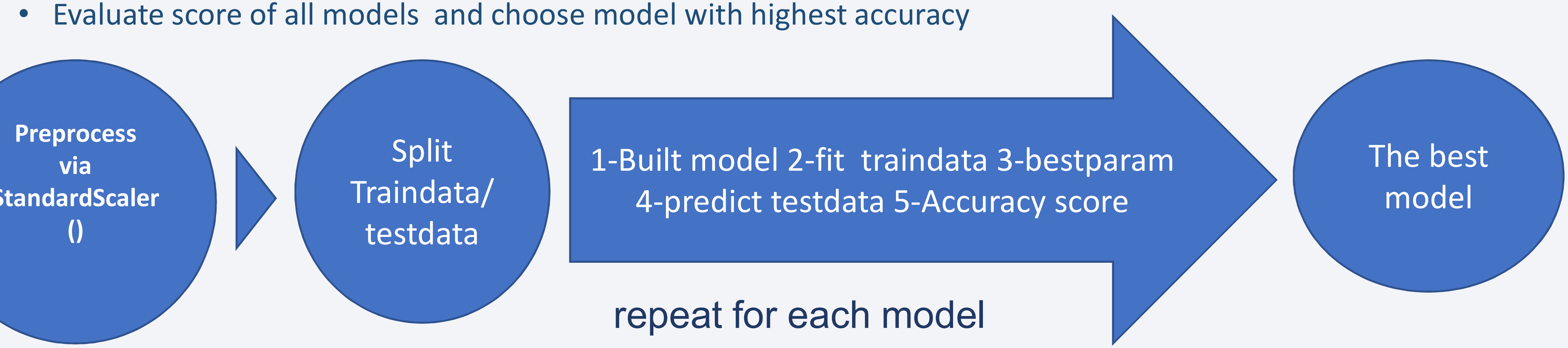
- Summarize what plots/graphs and interactions you have added to a dashboard
  - Dropdown box
  - Pie chart
  - Range Slider
  - Scatter Plot
- Explain why you added those plots and interactions
  - Dropdown box : to select the Site specific or ALL Sites
  - Pie chart : to display view of success rate for site selected
  - Range Slider : to choose the payload mass range
  - Scatter Plot :to describe the relation between payload mass and launch site on the success rate of landing

- FirstRepo/spacex_dash_app.py at master · abderrahim22/FirstRepo (github.com)

Predictive Analysis (Classification)

Summarize how you built, evaluated, improved, and found the best performing classification model

- Preprocess the data via method StandardScaler()
- Split dataset into training data and test data
- Create models, fit them via training data, predict using test data.
- Evaluate score of all models and choose model with highest accuracy



[FirstRepo/SpaceX_Machine Learning Prediction_Part_5.ipynb at master · abderrahim22/FirstRepo (github.com)]

Results

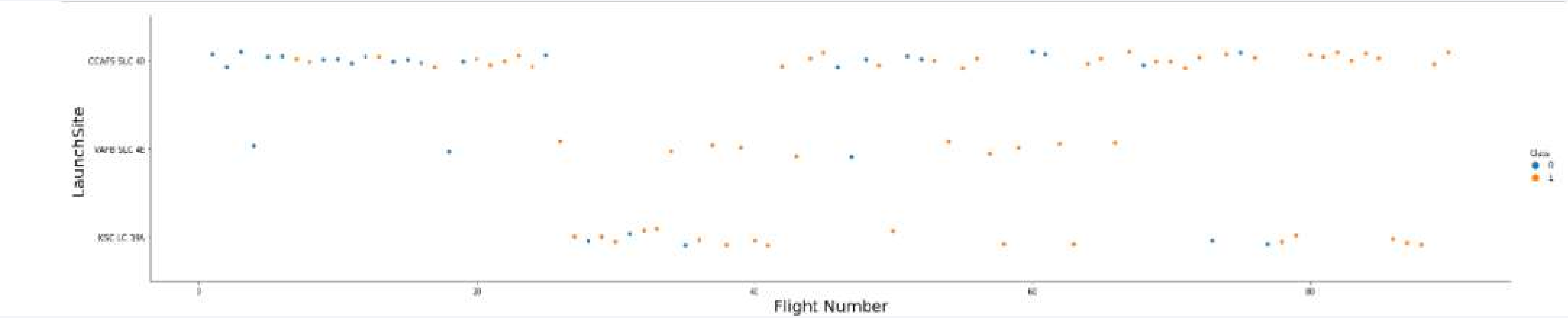
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

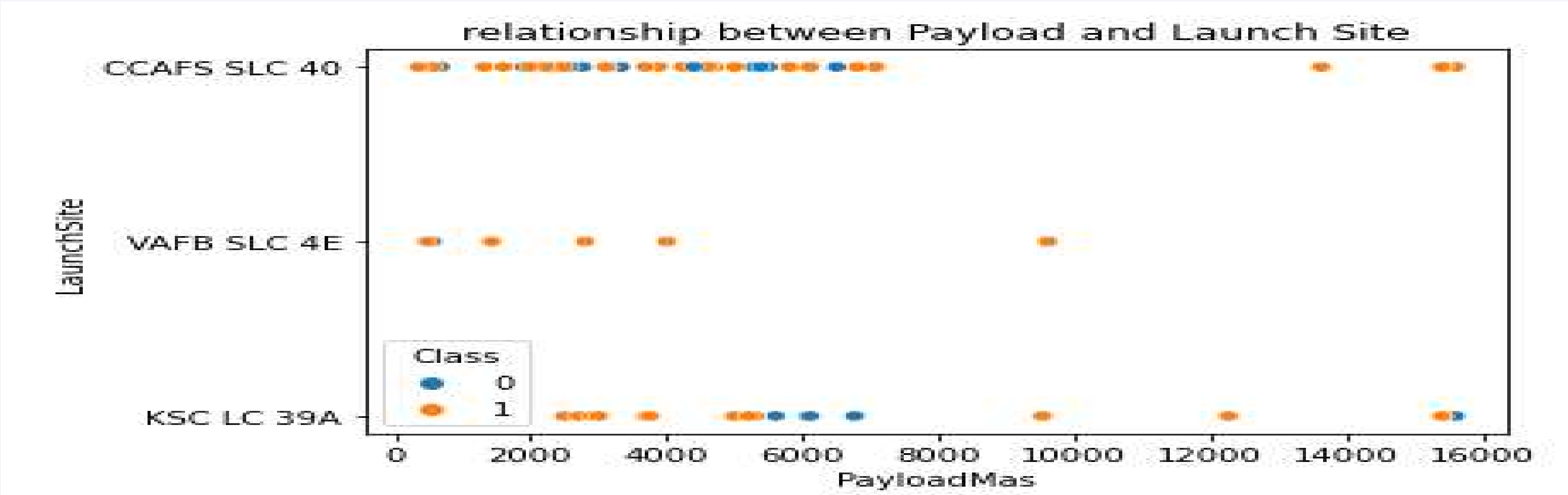
- Show a scatter plot of Flight Number vs. Launch Site



- CCAFS SLC 40 is used more as a launch site so it has a lower success rate, but the other two sites have a higher success rate and a lower number of flights.

Payload vs. Launch Site

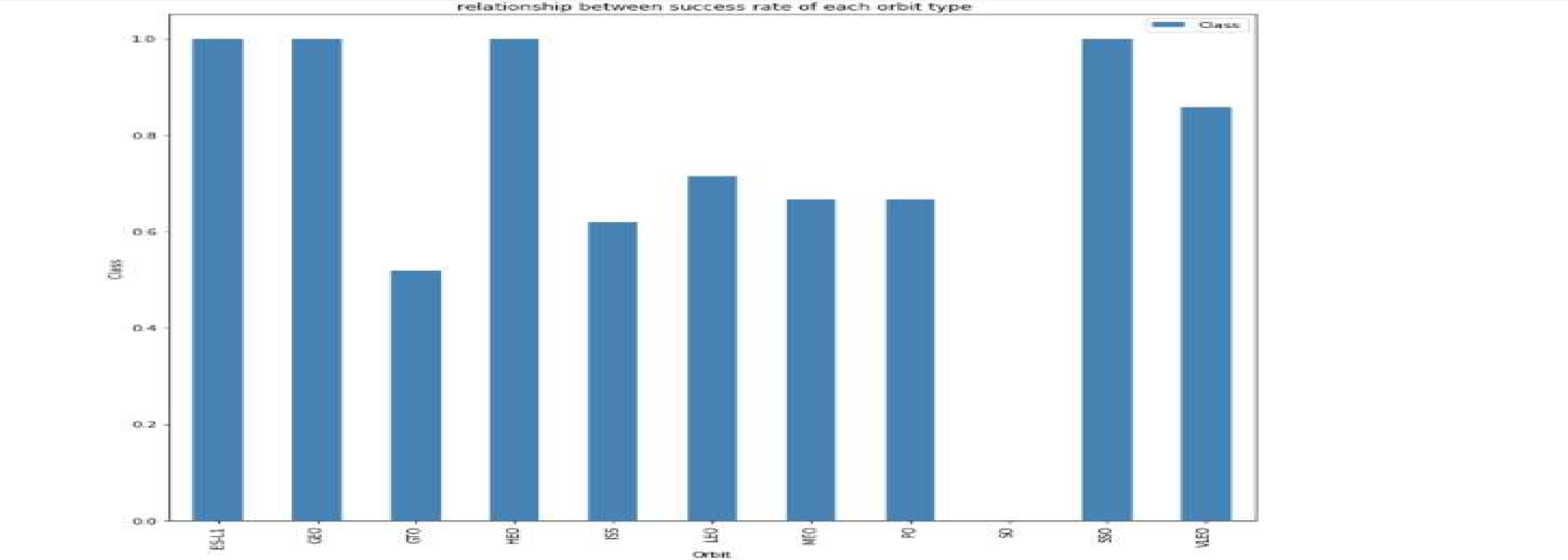
- Show a scatter plot of Payload vs. Launch Site



- The success rate appears to increase with a higher payload mass; With a mass less than 6000 kg, the KSC LC 39 A has a better success rate.

Success Rate vs. Orbit Type

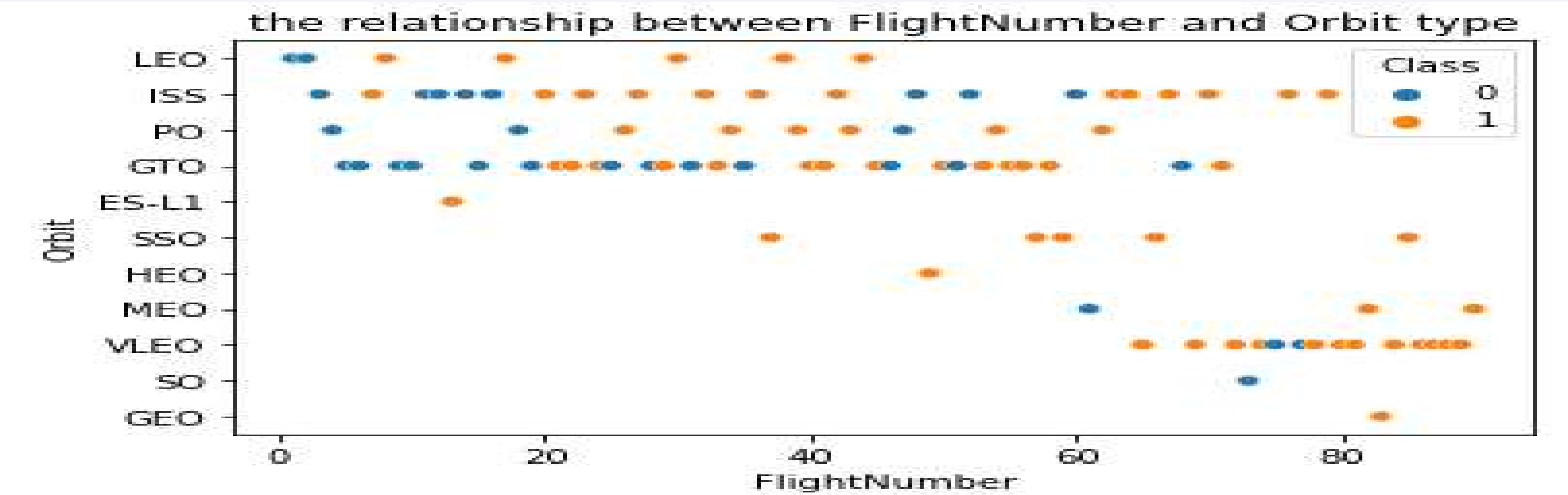
- Show a bar chart for the success rate of each orbit type



- The Orbits ES-L1, GEO, HEO and SSO have Top average success rates
- GTO orbit has lowest average success rate

Flight Number vs. Orbit Type

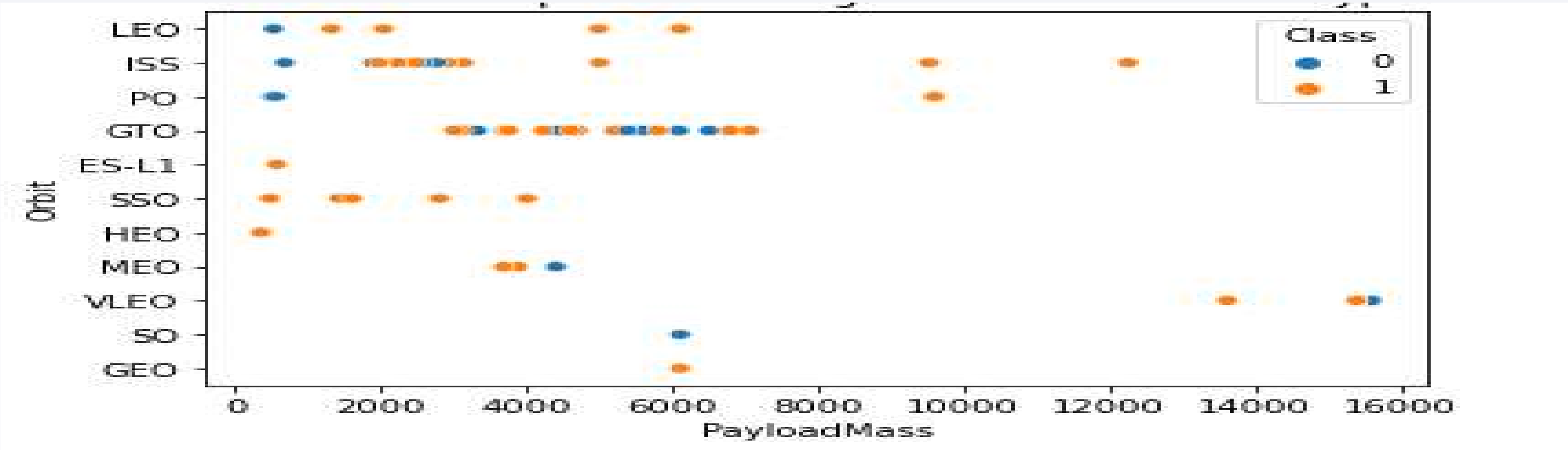
- Show a scatter point of Flight number vs. Orbit type



- The orbits VLEO, SSO, MEO and GEO have a high success rates after 60 flight numbers
- The orbits LEO, ISS, PO and GTO function better between 20 and 70 flight number

Payload vs. Orbit Type

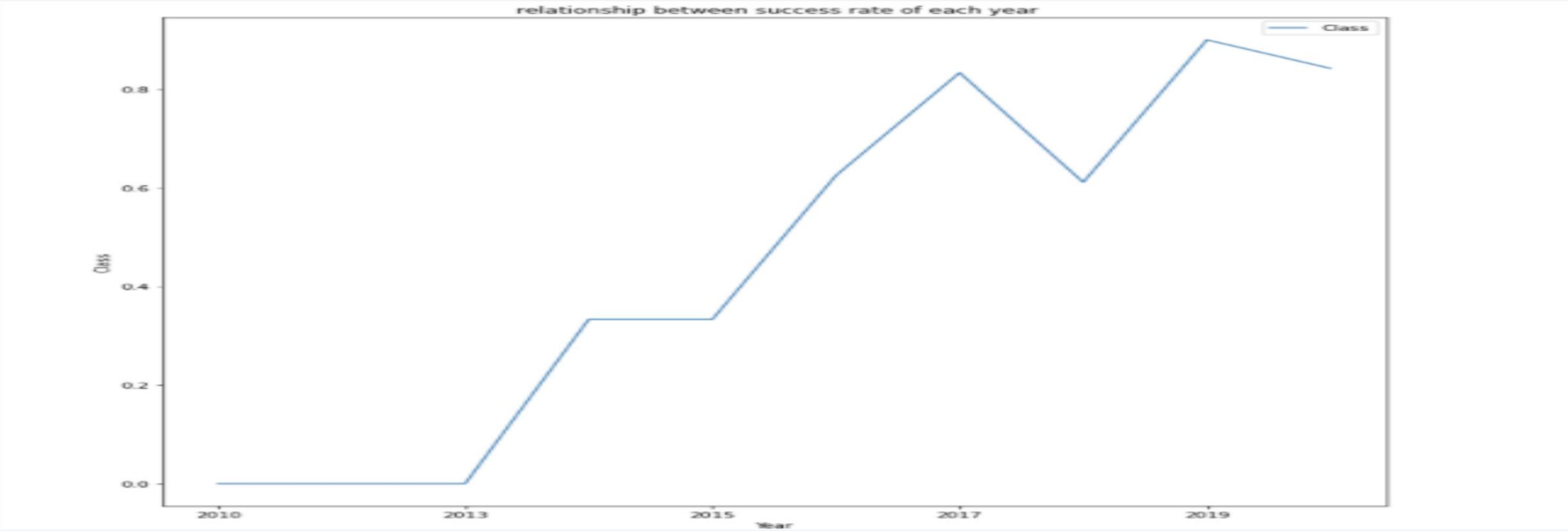
- Show a scatter point of payload vs. orbit type



- The Heavy payloads have a negative influence on orbits .

Launch Success Yearly Trend

- Show a line chart of yearly average success rate



- The sucess rate since 2013 kept increasing till 2020

All Launch Site Names

- Find the names of the unique launch sites
  - select Unique launch_site **from SPACEXTBL**
- Present your query result with a short explanation here
  - we find the unique names of All launch sites from table spacextbl

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`
  - select * **from SPACEXTBL** where launch_site like 'CCA%' limit 5
- Present your query result with a short explanation here

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	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	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)
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Display five records from table **SPACEXTBL** for only the launch sites begin with 'CCA'

Total Payload Mass

- Calculate the total payload carried by boosters from NASA
  - select sum(payload_mass__kg_)**as** Total **from SPACEXTBL** where customer like 'NASA (CRS)'
- Present your query result with a short explanation
  - Display the total of payload mass from SPACEXTBL Table just for NASA (CRS) Customer.

total
45596

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
  - select avg(payload_mass__kg_)**as** averge **from SPACEXTBL** where booster_version like 'F9 v1.1'
- Present your query result with a short explanation
  - We calculate the averge payload Mass kg from **SPACEXTBL** Table for

  Booster version F9v11

averge
2928

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad

```
%sql select min(DATE) from SPACEXTBL where landing__outcome like 'Success (ground pad)'
```

- Present your query result with a short explanation
  - We use function min() to define the first successful landing outcome on ground pad from SPACEXTBL Table

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

```
%sql select booster_version from SPACEXTBL where landing__outcome like 'Success (drone ship)' and payload_mass__kg_ between 4000 and 6000
```

- Present your query result with a short explanation
  - We select the booster version from SPACEXTBL Table with

    two conditions : 1. Successful Drone Ship Landing

    2. Payload between 4000 and 6000

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

```
1 select DISTINCT mission_outcome, COUNT(mission_outcome)as Total from SPACEXTBL group by mission_outcome
```

- Present your query result with a short explanation
  - We observe the number of successful mission outcomes

I is 99 and failure is 1 from SPACEXTBL Table

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

Boosters Carried Maximum Payload

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

```
select booster_version from SPACEXTBL where payload_mass__kg_ = (select max(payload_mass__kg_) from SPACEXTBL)
```

- Present your query result with a short explanation
  - We obtain list of the booster version from SPACEXTBL table with the Max Payload

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

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

```
%sql select Date,booster_version,launch_site,landing__outcome from SPACEXTBL where year(Date)=2015 and landing__outcome like 'Failur
one ship)'
```

- Present your query result with a short explanation

we have just two launch records in 2015

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

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

```
%sql select landing__outcome,count(landing__outcome) as Count,RANK() OVER (order by count(landing__outcome) DESC) as ORDER  from SPACEXTBL  WHERE landing__outcome in ('Failure (drone ship)','Success (ground pad)') and Date between '2010-06-04' and '2017-03-20'  group by landing__outcome
```

- Present your query result with a short explanation

We use rank function to order the count of landing outcomes between

Two dates.

landing__outcome	COUNT	ORDER
Failure (drone ship)	5	1
Success (ground pad)	3	2

Section 4

Launch Sites Proximities Analysis

The map with all launch sites

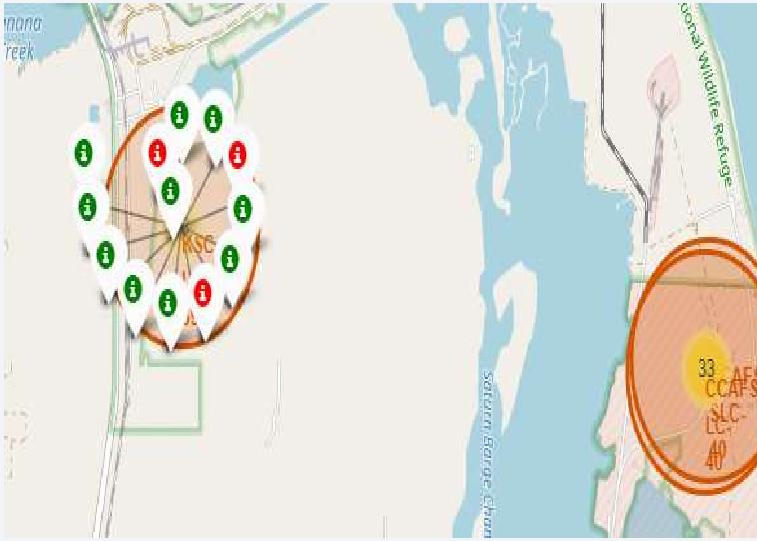
- The map with all launch sites

- Explain the important elements and findings on the screenshot :

we can explore the map by zoom-in/out the marked areas , and try to observe all proximities(coast…)

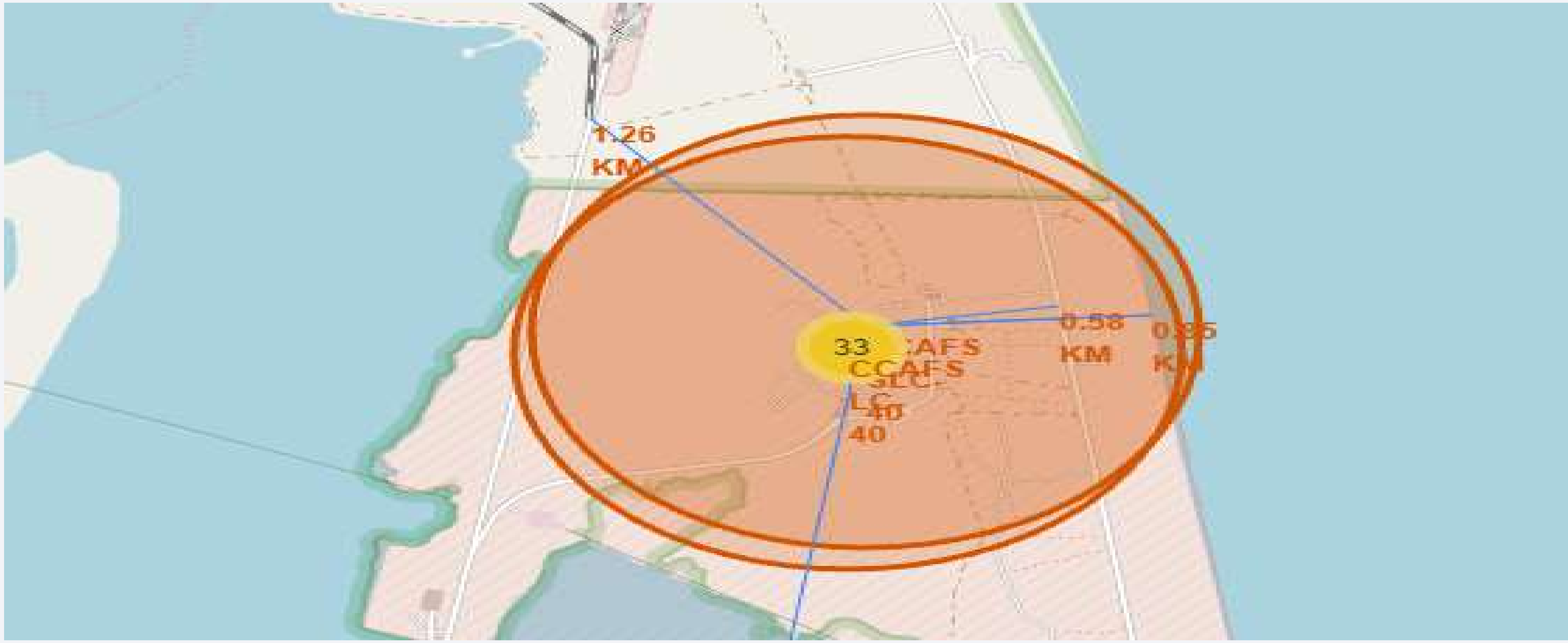
The success/failed launches for each site on the map

- **Mark the success/failed launches for each site on the map**

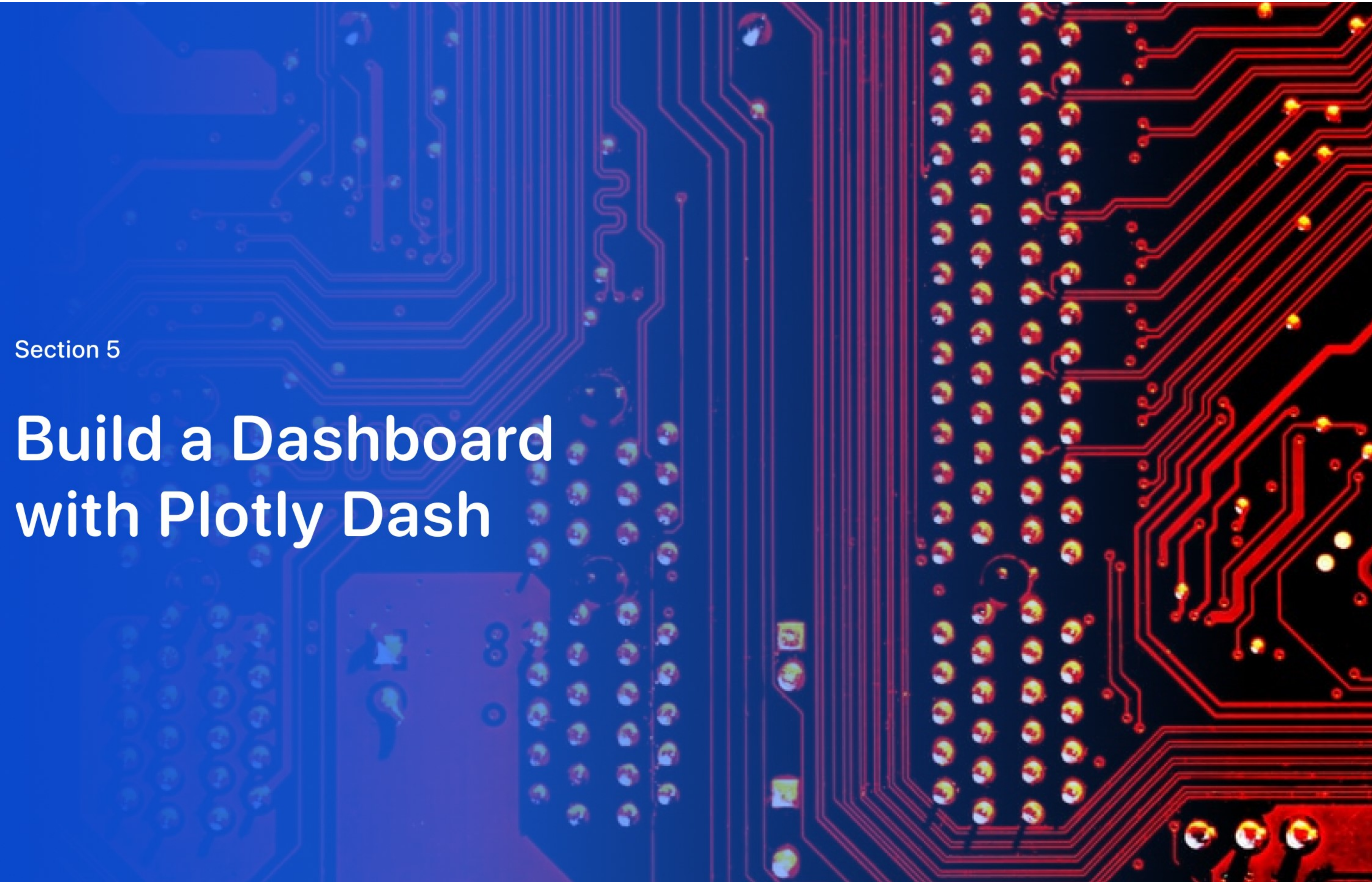


- From the color-labeled markers in marker clusters, we be able to easily identify which launch sites have relatively high success rates.

he distances between a launch site to its proximities



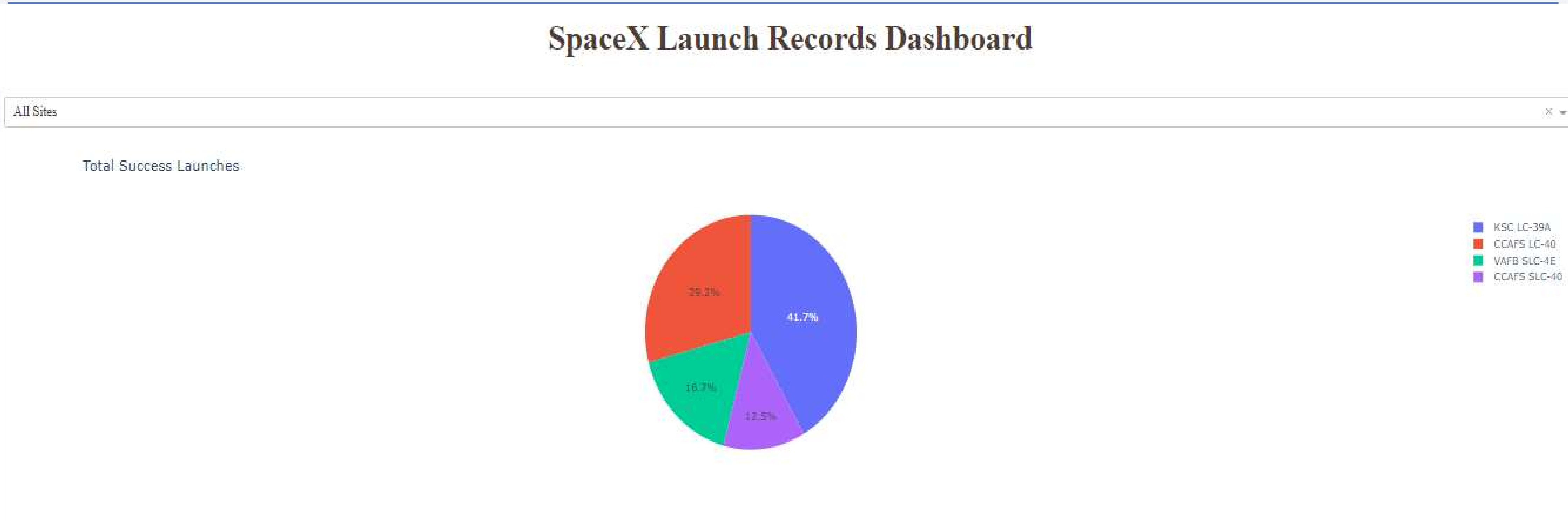
The important element is the possibility to determinate the proximities for each launch sites



Section 5

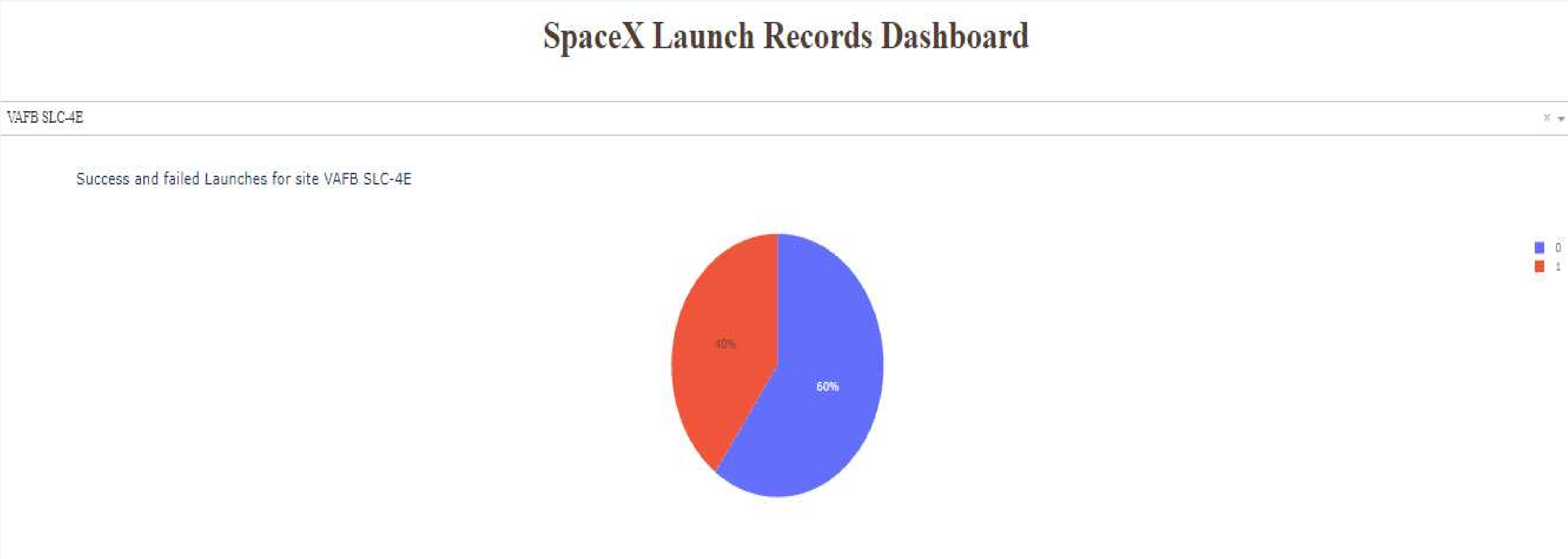
Build a Dashboard with Plotly Dash

Launch success count for all sites



The important element is showing the total launch success for all launch sites and we can order all sites via pie char

Success and Failed launch for a specific site



- Explain the important element is the possibility to observe in real time by showing the Success and Failed launch for a specific Site.

Correlation between Payload and Success for all Sites

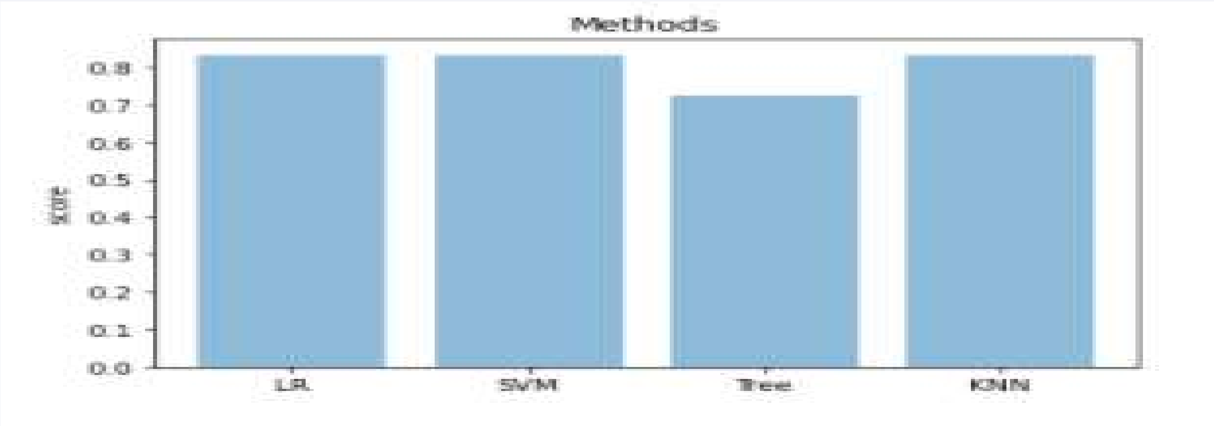
- The booster version FT has the largest success rate.

Section 6

Predictive Analysis (Classification)

Classification Accuracy

- Visualize the built model accuracy for all built classification models, in a bar chart



- Find which model has the highest classification accuracy

Based on the Bar char We have three highest classification accuracy (LR,SVM,KNN) with score =0.83

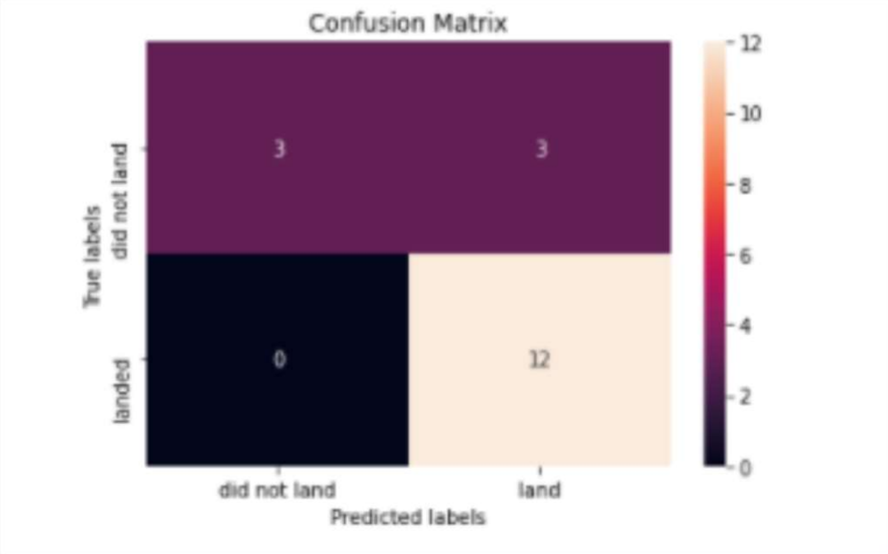
Confusion Matrix

Show the confusion matrix of the best performing model with an explanation

e confusion matrix for the SVM ,KNN,LR models

edicts all 12 True positive answers for

ccessful landings.



Conclusion

We have achieved the objective presented at the beginning of this project, the company spaceY can see in the results the strategy used by the company SpaceX to make a successful investment in Space race.

Thank you!