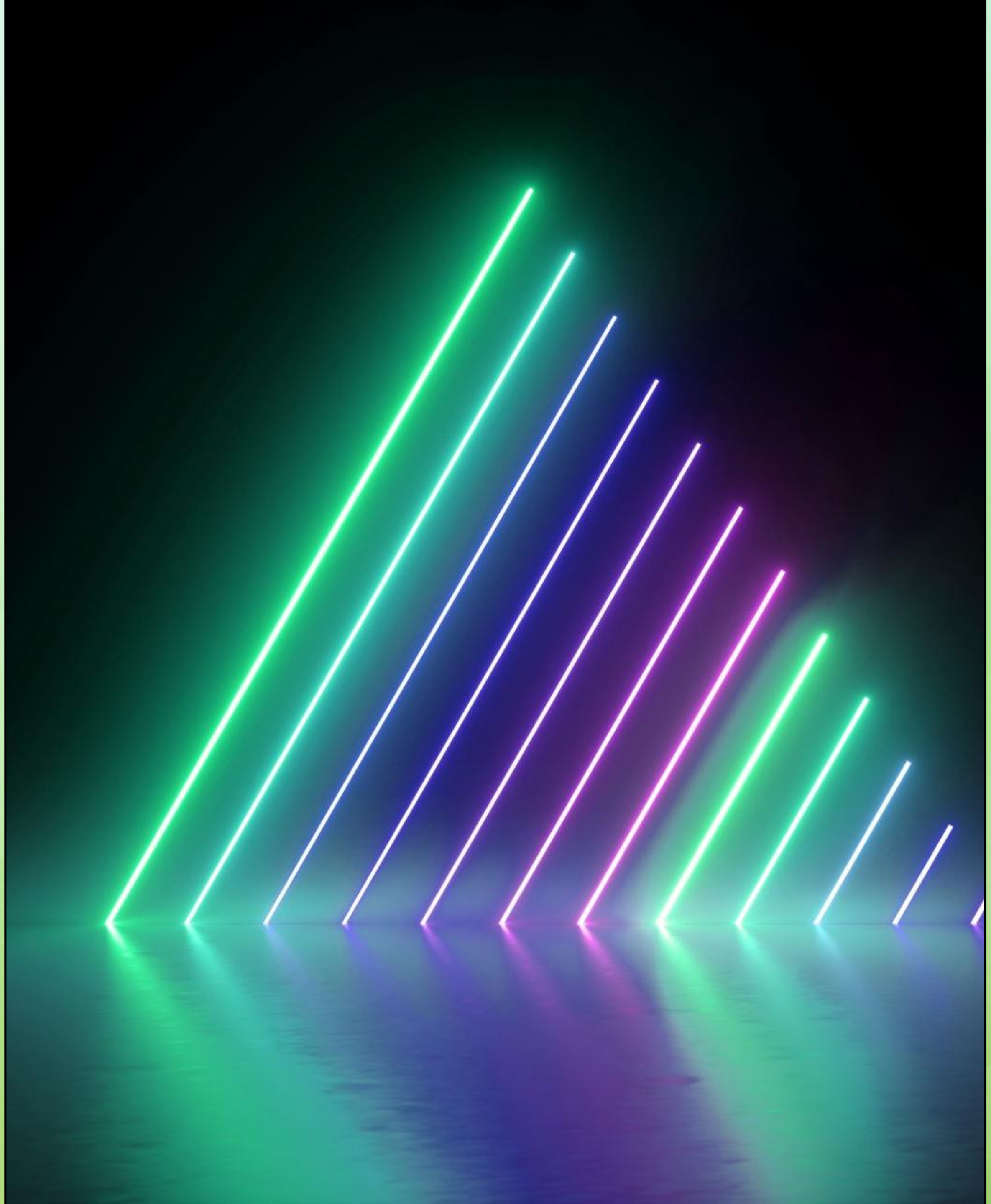
IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Bashir Abdinoor>
<25/06/2023>



Outline

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

Executive Summary

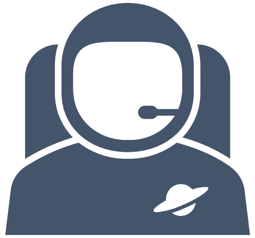
Summary of methodologies

Summary of all results

Introduction

Project background and context

SpaceX save about 100 million using spare parts from previous successful landings. The new stablished company, Space Y , wants to compete with Space X and wants to predict successful landing in the future using historical data.

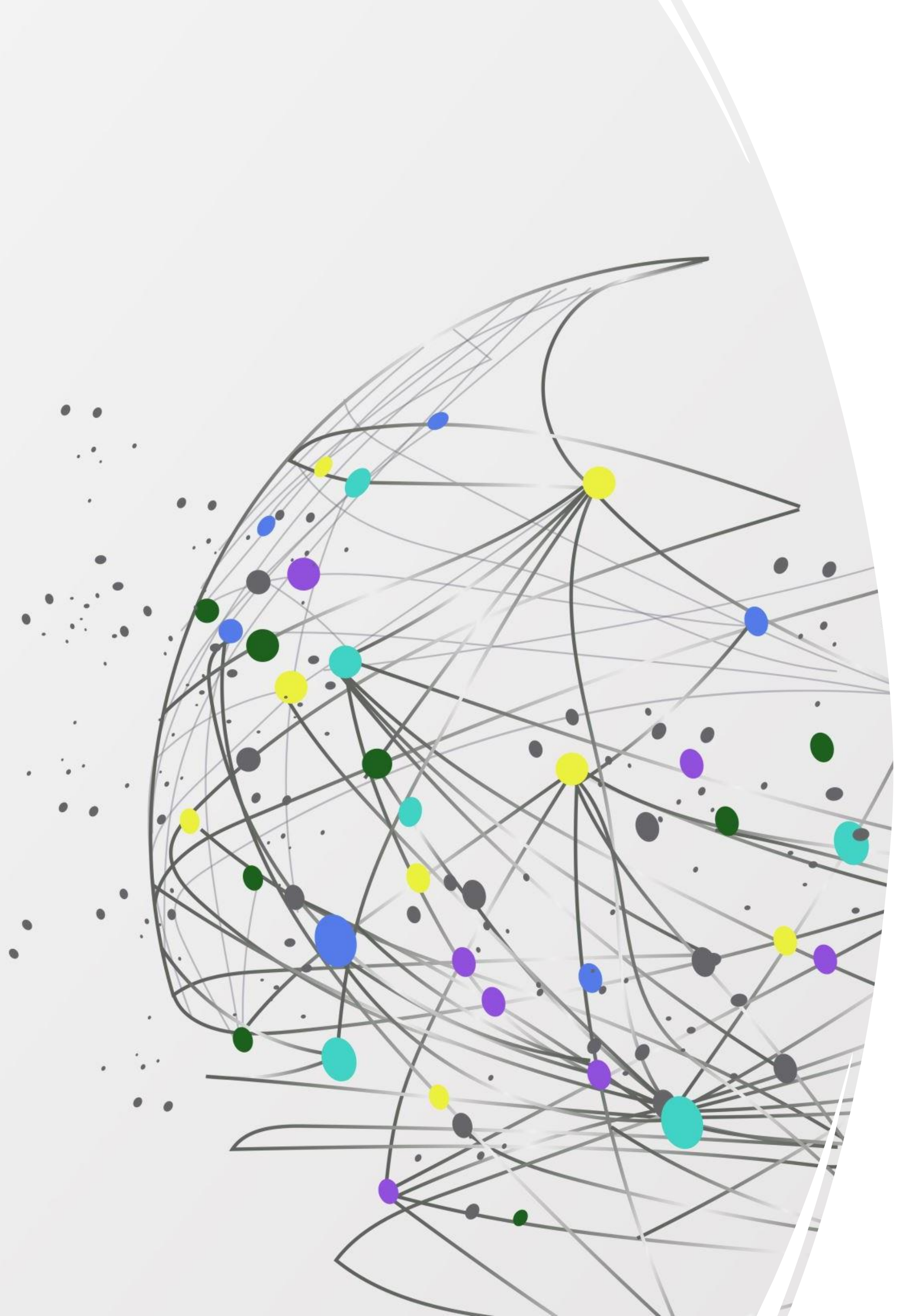


Problems to solve

- **The best model to predict successful landings in the future.**
- **How they improved in the past.**
- **The best predictors of successful landings.**

Section 1

Methodology



Methodology

- Executive Summary
- Data collection methodology:
  - Scrape from wikipidea and Space X rest api.
- Perform data wrangling
  - The data was converted in data frame to be analysed.
- 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 was collected specifically from SpaceX REST API and Wikipidea.

This is flowchart of data collecting

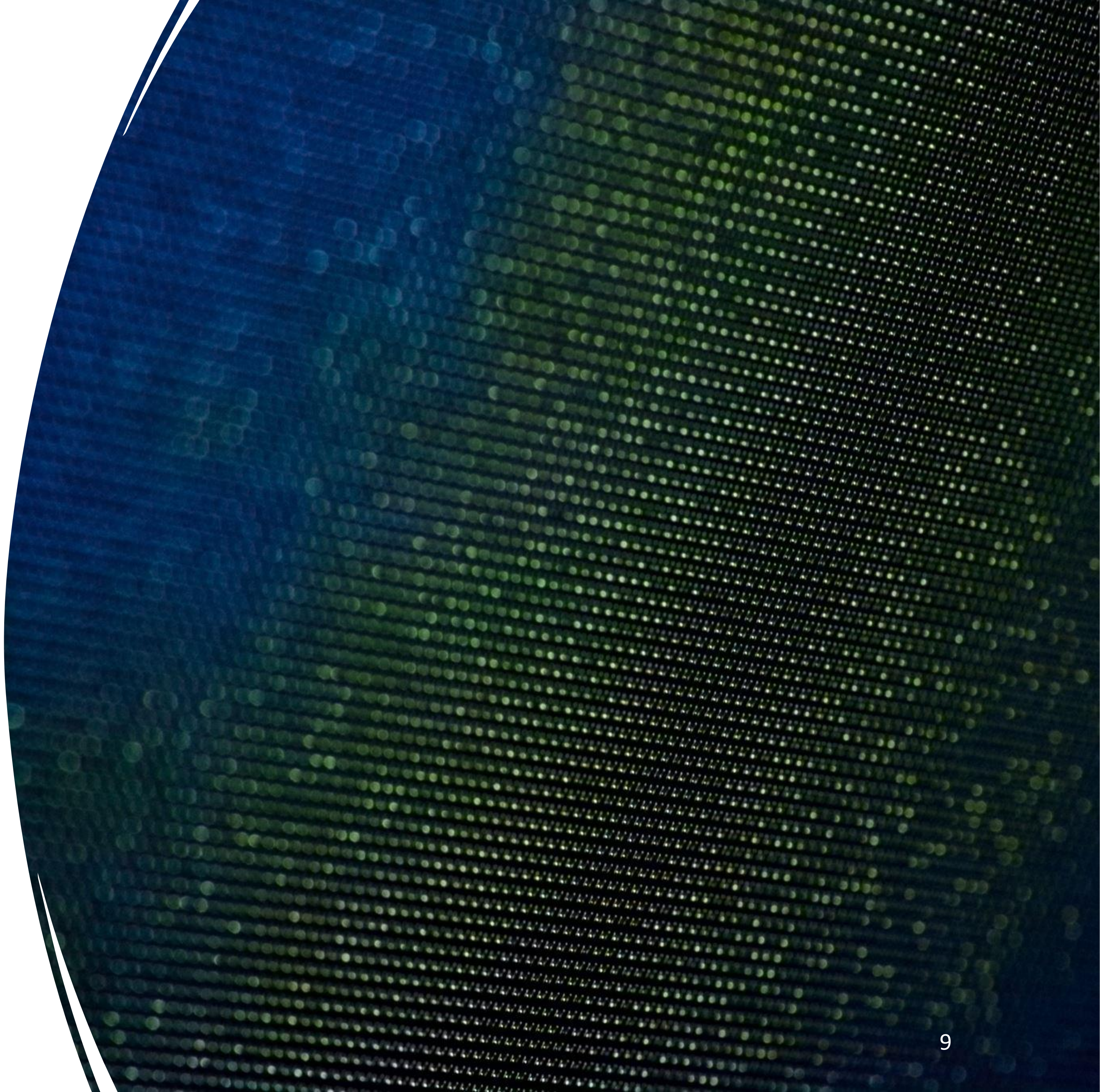
Use url → request using requests library → you get response in json

Data Collection – SpaceX API

- Request from SpaceX API ➔ parse the results using GET request ➔ Filter the data only to include Flacon 9 launches.

Data Collection - Scraping

- Request the Falcon9 Launch Wiki page from its URL ➔ Extract all column/variable names from the HTML table header ➔ Create a data frame by parsing the launch HTML tables.



Data Wrangling

- Include only Falcon9 data ➔ Replace Missing values with the mean values for the column Payload_mass_kg.

EDA with Data Visualization

- Bar chart, Scatter plots and Line graph were plotted to visualize the data and gain better understanding of it.

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 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.**
- **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 landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order.**

Build an Interactive Map with Folium

- Each launch site, circle was placed on top of it.
- Different classes were presented with different colours (green for successful landing and red for failure).
- This to clearly show the outcomes in a world map.

Build a Dashboard with Plotly Dash

Questions

- Which site has the largest successful launches?
- Which site has the highest launch success rate?
- Which payload range(s) has the highest launch success rate?
- Which payload range(s) has the lowest launch success rate?
- Which F9 Booster version (v1.0, v1.1, FT, B4, B5, etc.) has the highest launch success rate?

- Plots were pie chart & counter to answer those questions.

Predictive Analysis (Classification)

- Data → Numpy array → standardize the data → split into test and training data → create logistic regression, support vector machine, decision tree classifier and k nearest neighbour → score & compare their accuracy.

Results

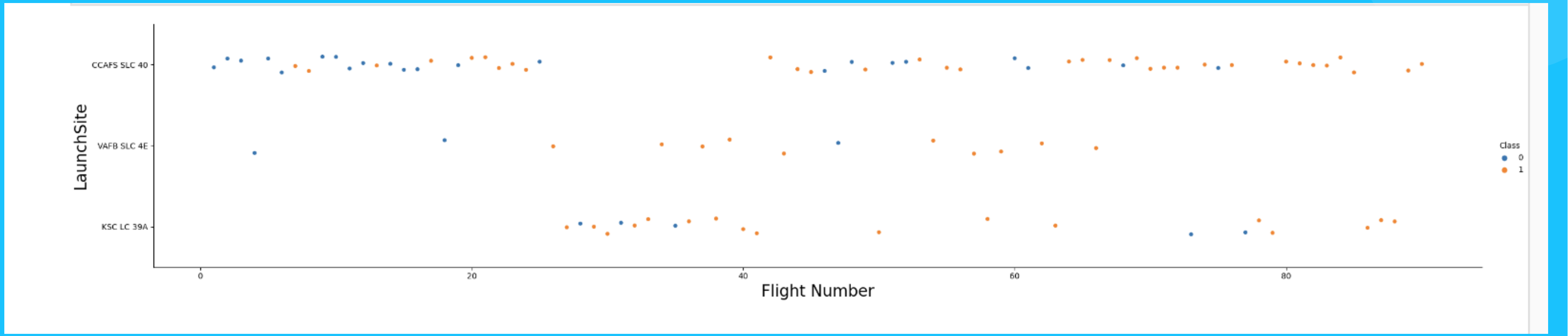
- Exploratory data analysis results
- **We annalysed 90 falcon9 launches.**
- Interactive analytics demo in screenshots
- \- **KSC LC-39A had the highest launch success count from all the sites.**
- Predictive analysis results
- \- **Tree decision algorithm produced the highest accuracy 83.33%.**

Section 2

Insights drawn from EDA

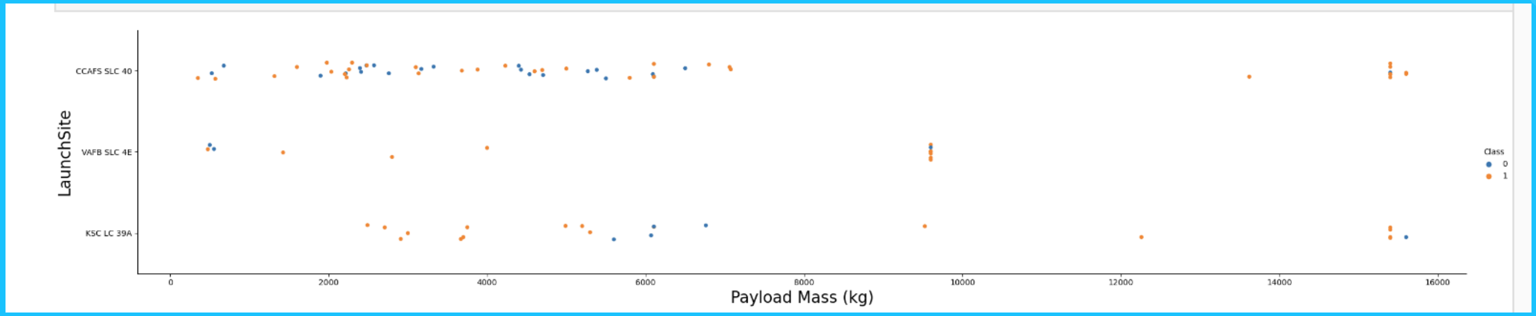
Flight Number vs. Launch Site

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



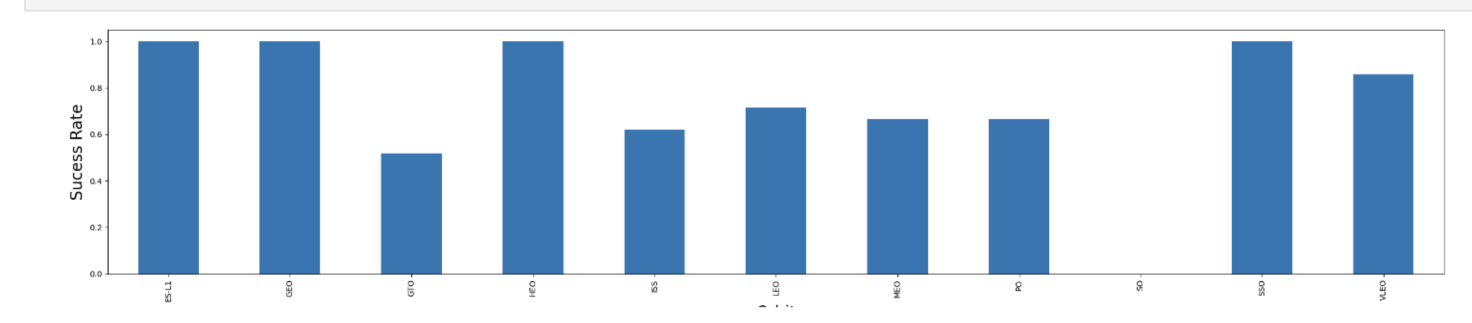
Payload vs. Launch Site

- a scatter plot of Payload vs. Launch Site



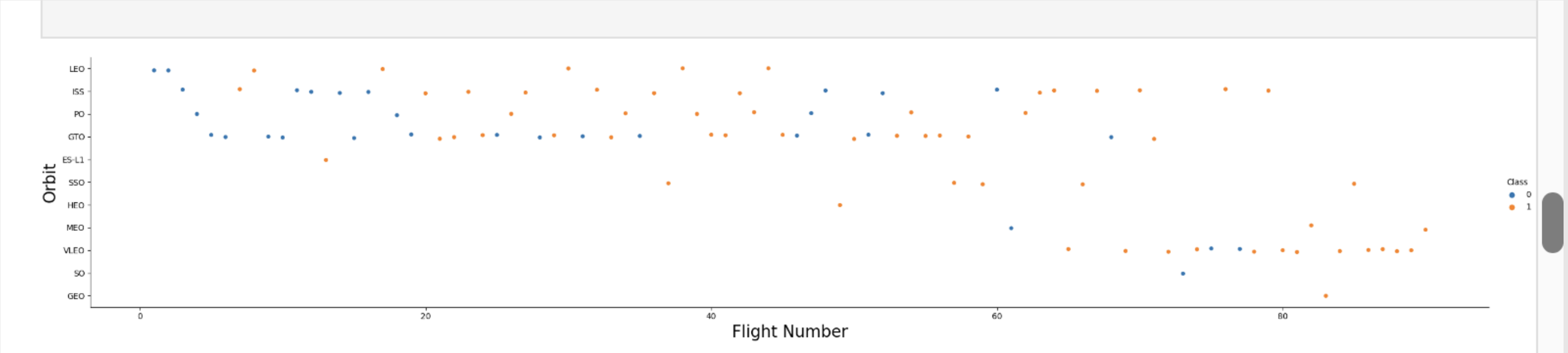
Success Rate vs. Orbit Type

- A bar graph of success rate Vs Orbit Type



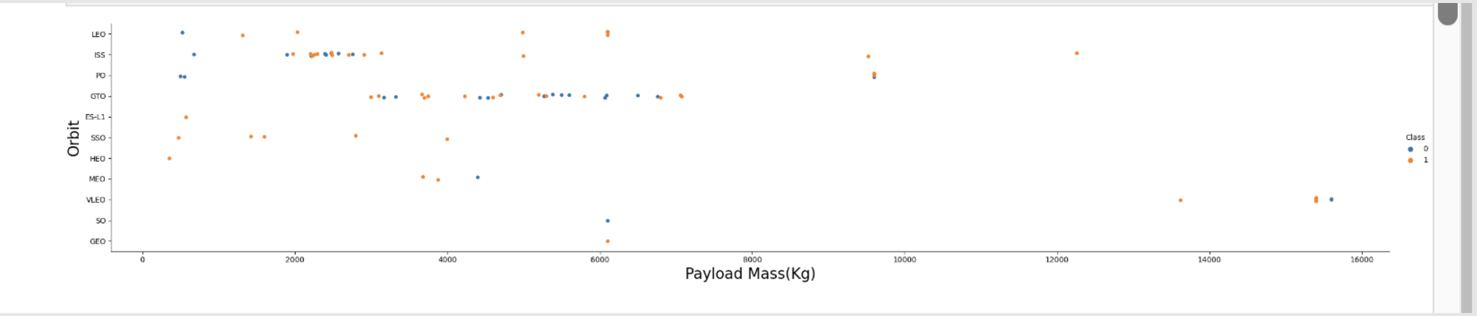
Flight Number vs. Orbit Type

- a scatter plot of Fight Number vs. Orbit Type

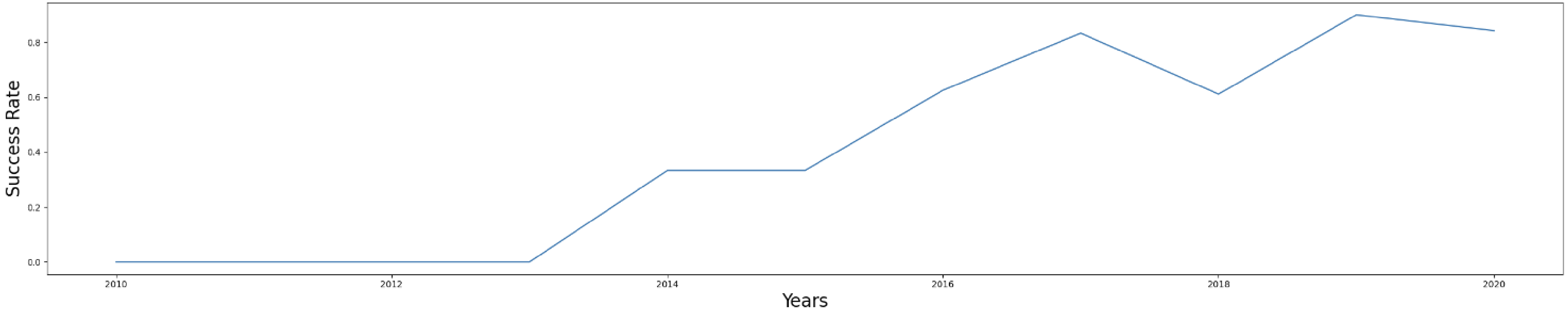


Payload vs. Orbit Type

- a scatter plot of Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

```
[20]: %sql SELECT DISTINCT LAUNCH_SITE from SPACEXTBL

 * sqlite:///my_data1.db
Done.
```

[20]:

Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40
None

Launch Site Names Begin with 'CCA'

```
]: %sql SELECT * FROM SPACEXTBL WHERE LAUNCH_SITE LIKE "%CCA%" LIMIT 5;

 * sqlite:///my_data1.db
Done.
]:
```

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Custo
06/04/2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0.0	LEO	Spa
12/08/2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0.0	LEO (ISS)	N (CC
22/05/2012	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525.0	LEO (ISS)	N (CC
10/08/2012	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500.0	LEO (ISS)	N (
03/01/2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677.0	LEO (ISS)	N (

Total Payload Mass

Display the total payload mass carried by boosters launched by NASA (CRS)

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE CUSTOMER LIKE "%NASA%"
```

```
 * sqlite:///my_data1.db
Done.
```

SUM(PAYLOAD_MASS__KG_)
107010.0

Average Payload Mass by F9 v1.1

```
33]: %sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE Booster_Version LIKE "%F9 V1.1%"
 * sqlite:///my_data1.db
Done.
```

33]:

AVG(PAYLOAD_MASS__KG_)
2534.6666666666665

First Successful Ground Landing Date

```
[57]: %sql SELECT Date FROM SPACEXTBL WHERE Mission_Outcome LIKE "%Success%" LIMIT 1
 * sqlite:///my_data1.db
Done.
```

[57]:

Date
06/04/2010

Successful Drone Ship Landing with Payload between 4000 and 6000

[44]:

Booster_Version
F9 v1.1
F9 v1.1 B1011
F9 v1.1 B1014
F9 v1.1 B1016
F9 FT B1020
F9 FT B1022
F9 FT B1026
F9 FT B1030
F9 FT B1021.2
F9 FT B1032.1
F9 B4 B1040.1
F9 FT B1031.2
F9 B4 B1043.1
F9 FT B1032.2
F9 B4 B1040.2
F9 B5 B1046.2
F9 B5 B1047.2
F9 B5 B1046.3
F9 B5B1054
F9 B5 B1048.3
F9 B5 B1051.2
F9 B5B1060.1
F9 B5 B1058.2
F9 B5B1062.1

Total Number of Successful and Failure Mission Outcomes

```
[51]: %sql SELECT COUNT(Date) FROM SPACEXTBL
 * sqlite:///my_data1.db
Done.
```

[51]:

COUNT(Date)
101

Boosters Carried Maximum Payload

[55]:
```
%%sql
SELECT DISTINCT Booster_Version FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ == (
    SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL)
```

```
 * sqlite:///my_data1.db
Done.
```

[55]:

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 SUBSTR(Date, 4,2), LANDING_OUTCOME, BOOSTER_VERSION, LAUNCH_SITE
FROM SPACEXTBL
WHERE substr(Date,7,4)='2015'
```

```
 * sqlite:///my_data1.db
Done.
```

]:

SUBSTR(Date, 4,2)	Landing_Outcome	Booster_Version	Launch_Site
10	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
11	Controlled (ocean)	F9 v1.1 B1013	CCAFS LC-40
02	No attempt	F9 v1.1 B1014	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40
04	No attempt	F9 v1.1 B1016	CCAFS LC-40
06	Precluded (drone ship)	F9 v1.1 B1018	CCAFS LC-40
12	Success (ground pad)	F9 FT B1019	CCAFS LC-40

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

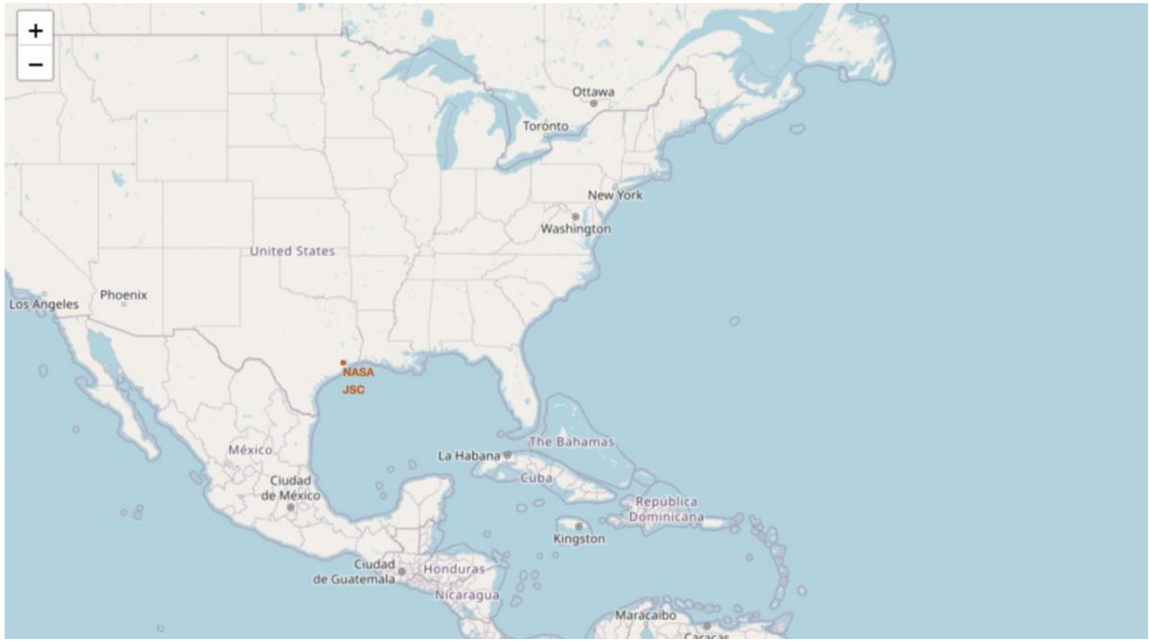
[63]:

Landing_Outcome
Failure (parachute)
Failure (parachute)
No attempt
No attempt
Controlled (ocean)
Controlled (ocean)
No attempt
No attempt
Failure (drone ship)
Failure (drone ship)
Success (drone ship)
Success (drone ship)
Failure (drone ship)
Success (ground pad)
Success (drone ship)
Success (drone ship)
Success (ground pad)
No attempt
Success (ground pad)
No attempt
Success (ground pad)
No attempt
Success (ground pad)
Success (ground pad)
Success (drone ship)

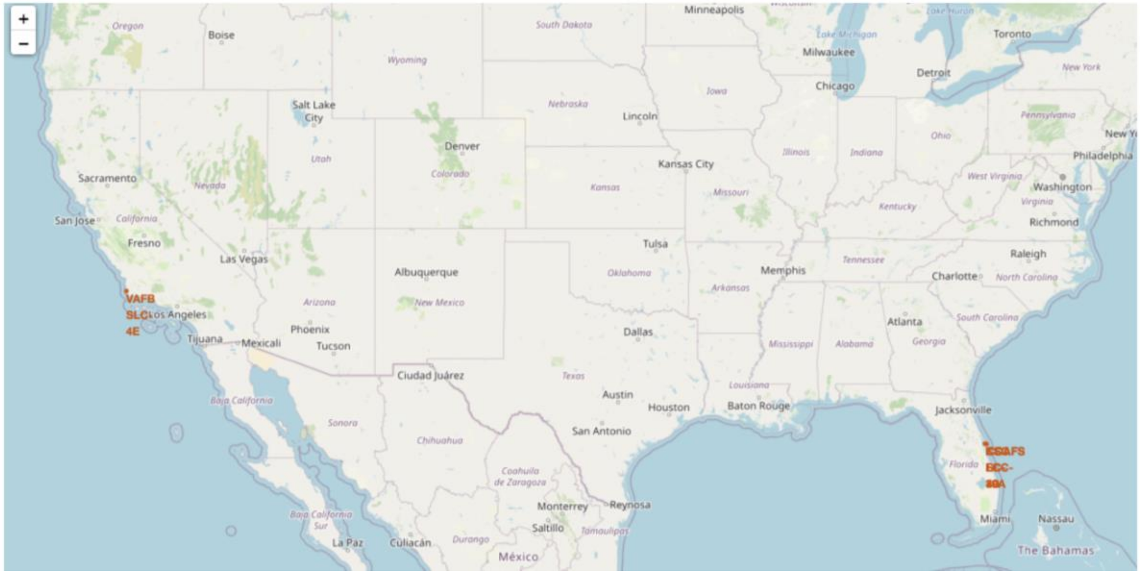
Section 3

Launch Sites Proximities Analysis

Initial location: NASA Johnson Space Center at Houston, Texas



Launch site locations

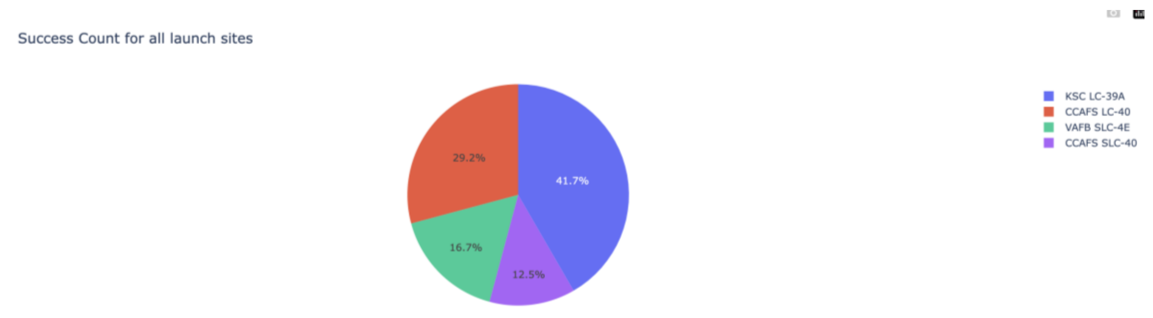


Locations relative to the globe

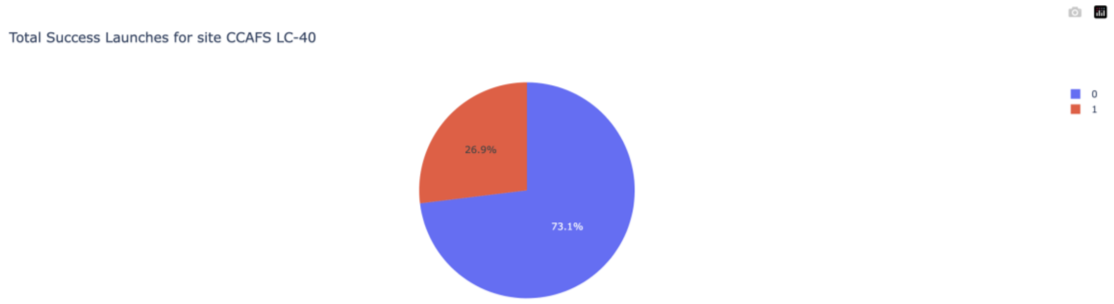
Section 4

Build a Dashboard with Plotly Dash

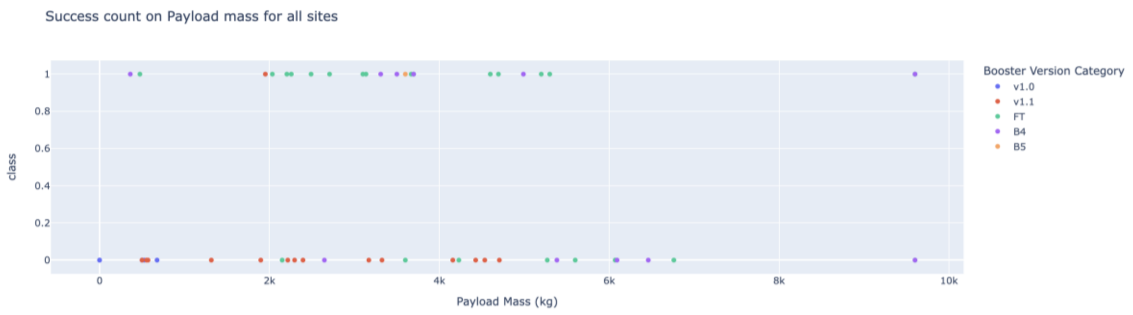
Lauch success count for all sites



Highest launch success ratio (CCAFS LC-40)



Payload vs. Launch Outcome scatter plot for all sites



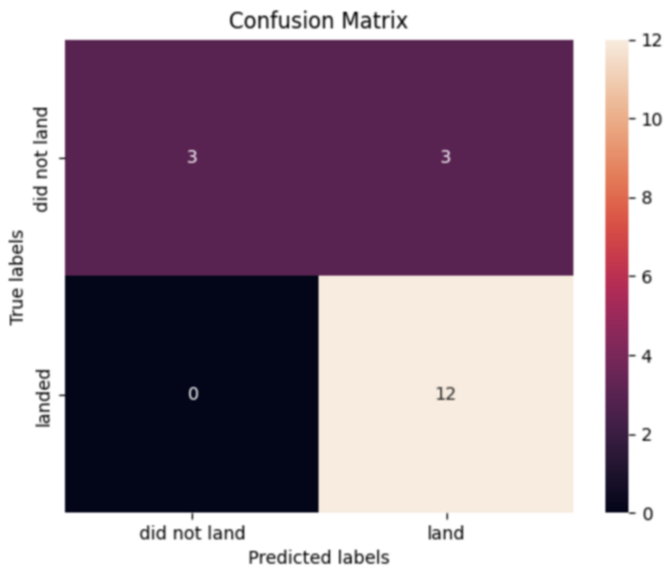
Section 5

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 for tree decision making



Conclusions

Point 1: Launch success are increasing as the years increase for spaceX, therefore, as time progresses the cost decreases.

Point 2: Tree decision making algorithm can predict if a given launch will be successful or not by 83.3% accuracy.

Point 3: Geologically, SpaceX activities are aggregated in the US.

Thank you!