IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Name>
<Date>

Outline

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

Executive Summary

- Summary of methodologies
  1. Data Collection
  2. Web Scrapping
  3. Data wrangling
  4. Data Analysis (EDA) using visualization
  5. Data Analysis (EDA) using SQL
  6. Interactive visualization analytics using folium
  7. Interactive visualization analytics using Plotly Dash
  8. Predictive analysis using classification models

- Summary of all results
  1. Data analysis Results
  2. Interactive analysis
  3. Predictive analysis

Introduction

- Project background and context

  Space travelling in becoming cheaper so everyone can have access to it. Satellite companies try to reduce the travelling cost by reusing parts of the rockets. Among them, SpaceX - probably the most successful one, advertises Falcon 9 rocket launches with a cost of 62 million dollars (other provides cost 165 million dollars and more). The difference in the cost is because the first stage of the rocket can be reused.

- Problems you want to find answers

  For this project, we are trying to ask some questions like:

  - Can we predict the success of the first stage landing based on the data available?
  - Does the geographic location of a launch site affects the success of the first stage landing?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - SpaceX Public Rest API
  - Web Scrapping mission data from Wikipedia
- Perform data wrangling
  - Analyze and select relevant features for the model • One Hot encoding data fields • Cleaning null values and irrelevant columns
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Test and train classification models such as: Logistic Regression, K-Nearest Neighbor, Support Vector Machine, and Decision Trees, evaluating each one for the best classifier.

Data Collection

- Requests at SpaceX **Rest API** (https://api.spacexdata.com/v4/)

- **Web Scrapping** using Beautiful Soup for Falcon 9 and Falcon Heavy Launches records (Launches)

Data Collection

- Code for data collection: [Data collection]

Data Collection – SpaceX API

- Code for data collection ucin SpaceX API: [Data collection]

Place your flowchart of SpaceX API calls here

Data Collection - Scraping

- Data collection using Web scraping: Web scraping code

Place your flowchart of web scraping here

Data Wrangling

- Data Wrangling code: Data wrangling

EDA with Data Visualization

Categorical Graphs

• Flight Number x Payload Mass

• Flight Number x Launch Site

Bar graphs

• Success Rate x Orbit

Data visualization code: [EDA Visualization]

Scatter Graphs

• Flight Number x Payload Mass

Line Graphs

• Success Rate x Year

EDA with SQL

- EDA with SQL code: SQL code

Build an Interactive Map with Folium

- Map objects used in folium map
  - **Circle object** added at NASA Johnson Space Center's coordinate
  - **Circle object** for each launch site
  - **Marker Cluster** used to group the success /failed markers
  - **Markers** to show success and failed landing outcomes for each launch site
  - **Markers and Lines** to show the distance between launch sites and key locations like: closest railway, closest highway, closest city and coastlines.

Code for Folium Map: Map with Folium

Build a Dashboard with Plotly Dash

- Charts:
  - Pie Chart – Shows the total success and failures for one or all launch points
  - Scatter Chart – Shows the relationship between the Payload Mass Carried, and the success rate.
- Inputs:
  - Dropdown – Used to allow the user to select a launch site or all launch sites to see the statistics
  - Ranger Slider – Used to select the range of the payload mass in a fixed range

Predictive Analysis (Classification)

- Data Preparation
  - Loading data
  - Normalize data
  - Split train / test data
- Model Preparation
  - Initialize Machine Learning Algorithms
  - Select parameters for each algorithm using GridSearchCV
  - Train the GridSearchCV model with the training data
  - Find the best parameters for each model
- Model Evaluation
  - Compute the accuracy of each algorithm with the best hyperparameters
  - Plot confusion matrix for evaluating results
- Model Comparison
  - Compare models based on the accuracy on a new test data
  - Examine the model with best accuracy

Code for classification: [Classification]

Results

- Exploratory data analysis results
  - The first successful launch was in 2015
  - The success rate for Falcon 9 boosters is high
  - The possibility for successful landing increases with the years
- Interactive analytics demo in screenshots

- Predictive analysis results

All models had the same accuracy of 83.33%.

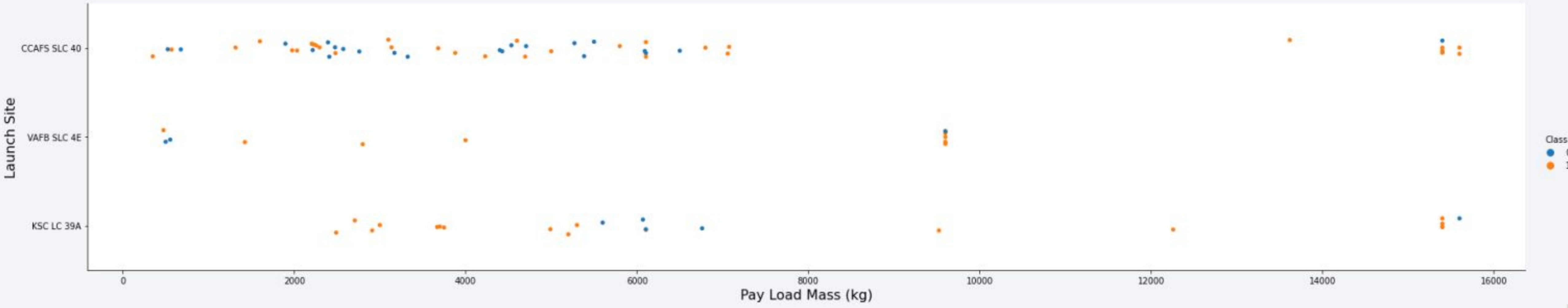
Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

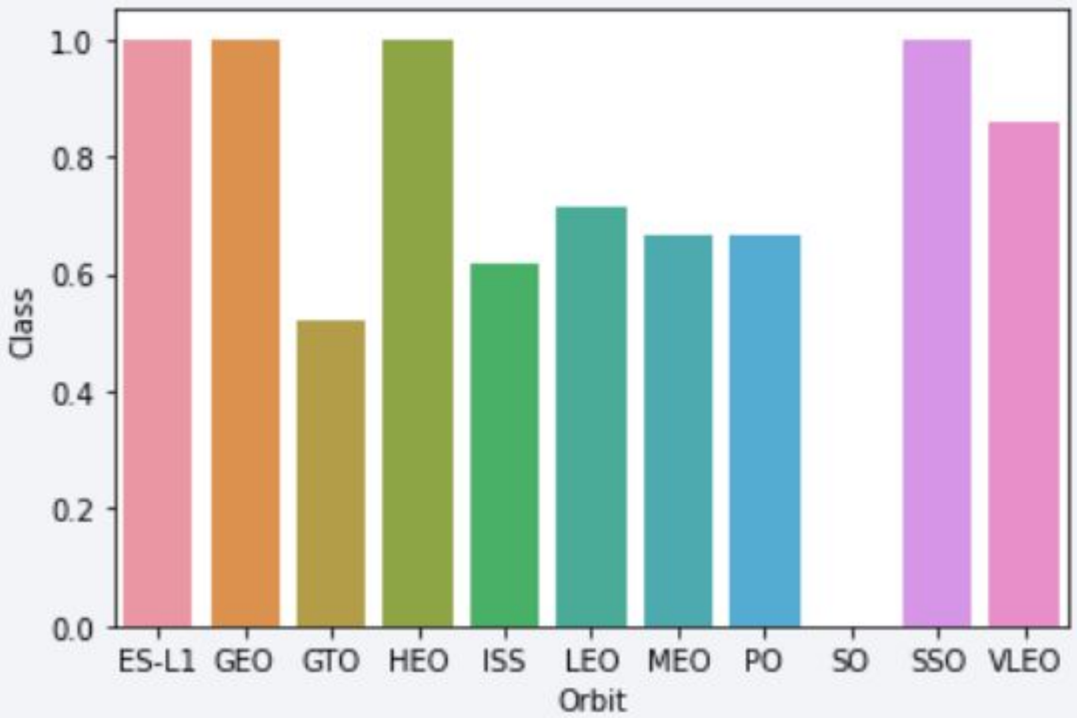
- As the number of lights increases, the success rate also increases.

Payload vs. Launch Site



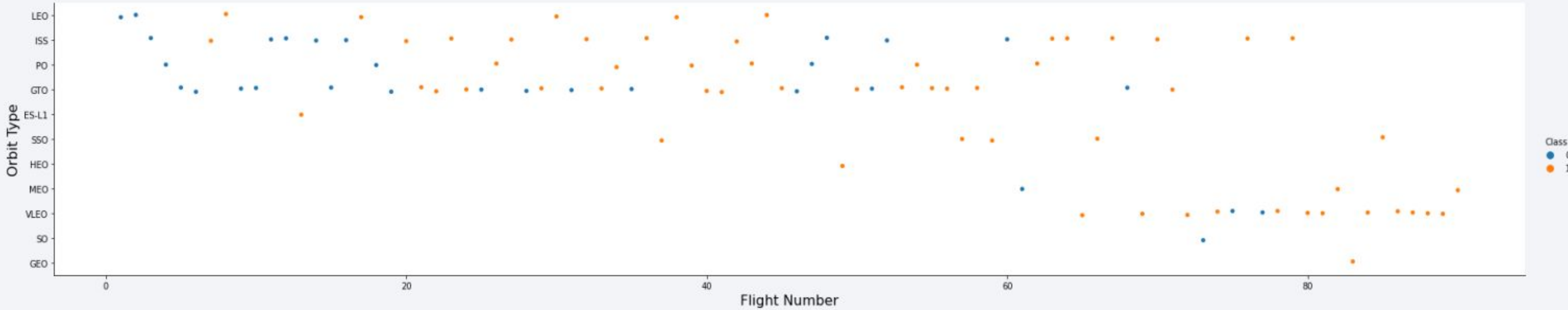
- CCAFS SLC 40 is the launch sites that have carried the heaviest payloads with the highest success rate.

Success Rate vs. Orbit Type



- ES-L1 , GEO , HEO , SSO , VLEO orbits have success rate above 80%

Flight Number vs. Orbit Type



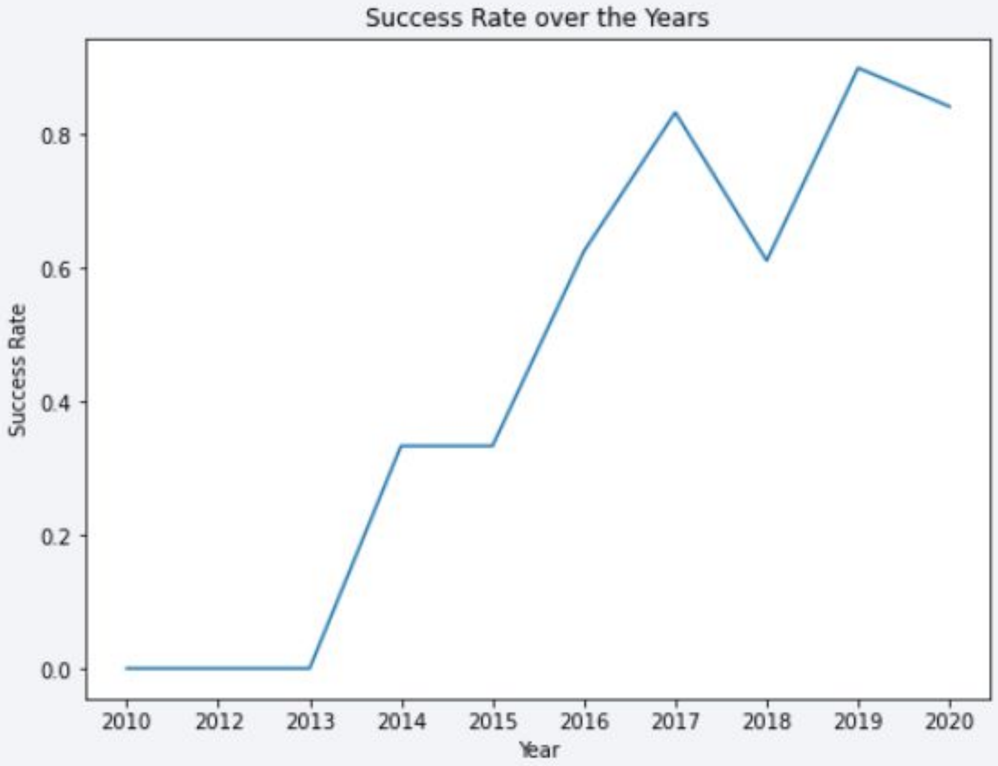
- VLEO orbit have many successful flights

Payload vs. Orbit Type

- Orbits like VLEO and ISS are the only ones that had a high payload mass

- SO and GE have few launches

Launch Success Yearly Trend



- Success rate increases since 2013
- The biggest success rate was in 2019

All Launch Site Names

- Search all the launch sites and present each one once

Task 1

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

```
%sql SELECT UNIQUE(launch_site) FROM SPACEX;
```

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

Launch Site Names Begin with 'CCA'

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

Task 2

Display 5 records where launch sites begin with the string 'CCA'

```
%sql SELECT launch_site FROM SPACEX WHERE launch_site LIKE 'CCA%' LIMIT 5
```

- Display the first 5 sites that the name begins with ‘CCA’

Total Payload Mass

- The total payload mass in KG carried by boosters launched by NASA:

total_kg
45596

Task 3

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

```
%sql SELECT SUM(payload_mass__kg_) as total_kg FROM SPACEX WHERE customer='NASA (CRS)'
```

Average Payload Mass by F9 v1.1

- The average payload mass in KG carried by booster F9 v1.1

total_kg
2928

Task 4

Display average payload mass carried by booster version F9 v1.1

```
%sql SELECT AVG(payload_mass__kg_) as total_kg FROM SPACEX WHERE booster_version='F9 v1.1'
```

First Successful Ground Landing Date

Task 5

List the date when the first successful landing outcome in ground pad was acheived.

Hint:Use min function

```
%sql SELECT MIN(date) as date_ FROM SPACEX WHERE landing__outcome='Success (ground pad)'
```

date_
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- List of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

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

Task 6

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

```
%sql SELECT booster_version from SPACEX where \
  landing__outcome='Success (drone ship)' and \
  payload_mass__kg_ BETWEEN 4000 AND 6000
```

Total Number of Successful and Failure Mission Outcomes

- Total number of successful and failure mission outcomes

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

Task 7

List the total number of successful and failure mission outcomes

```
]: %sql SELECT mission_outcome, COUNT(*) as total FROM SPACEX GROUP BY mission_outcome
```

Boosters Carried Maximum Payload

- Boosters which have carried the maximum payload mass

booster_version
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
F9 B5 B1051.3
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

Task 8

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

```
%sql SELECT booster_version FROM SPACEX WHERE\
    payload_mass__kg_=(SELECT MAX(payload_mass__kg_) FROM SPACEX)
```

2015 Launch Records

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

booster_version	launch_site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

Task 9

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

```
%sql SELECT booster_version, launch_site FROM SPACEX WHERE \
    landing__outcome='Failure (drone ship)' AND\
    YEAR(date)=2015
```

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

- 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

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

Task 10

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

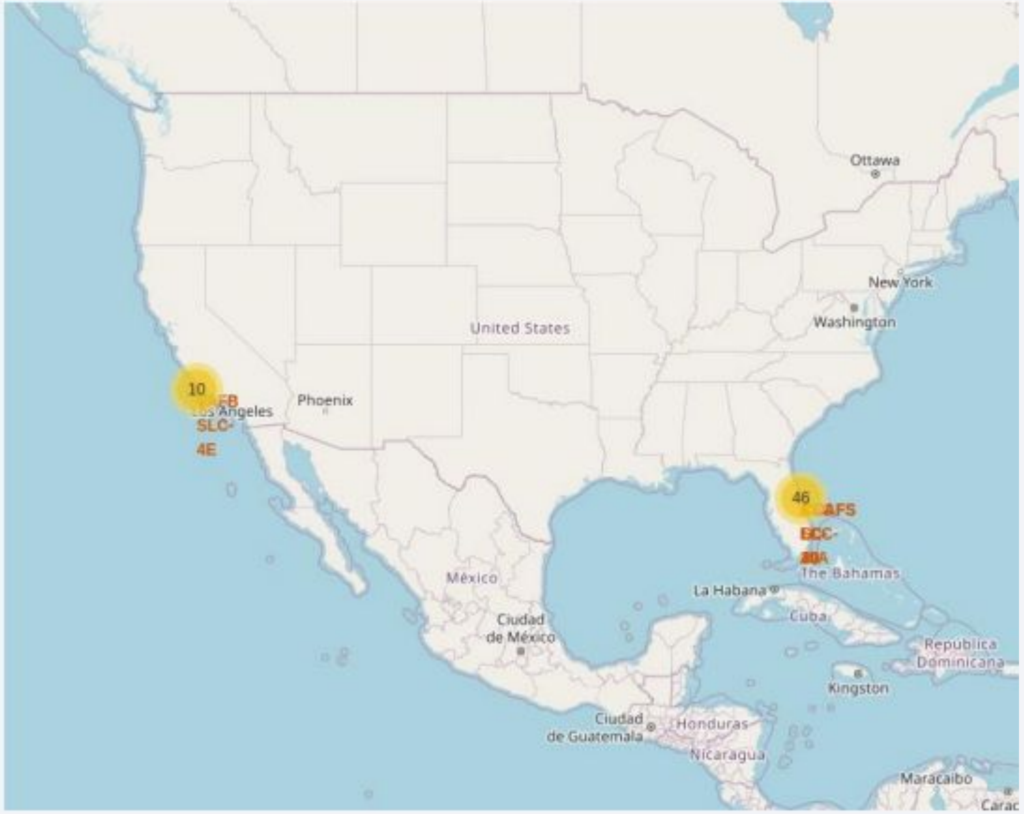
```
%sql SELECT landing__outcome as landing_outcome, COUNT(1) as total FROM \
    (SELECT landing__outcome FROM SPACEX WHERE date BETWEEN '2010-06-04' AND '2017-03-20')\
GROUP BY landing__outcome ORDER BY total DESC
```

Section 3

Launch Sites Proximities Analysis

Launch Sites Location

- Global view of the location of all launch sites.
- All launch sites are near a coastline

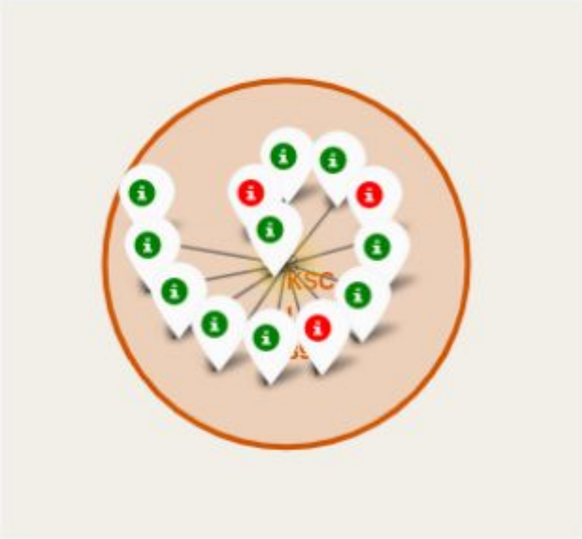


Success / Failure rate for Launch Sites

- Green: successful launches
- Red: unsuccessful launches
- KSC LC-39A has high launch success rate

CCAFS LC-40

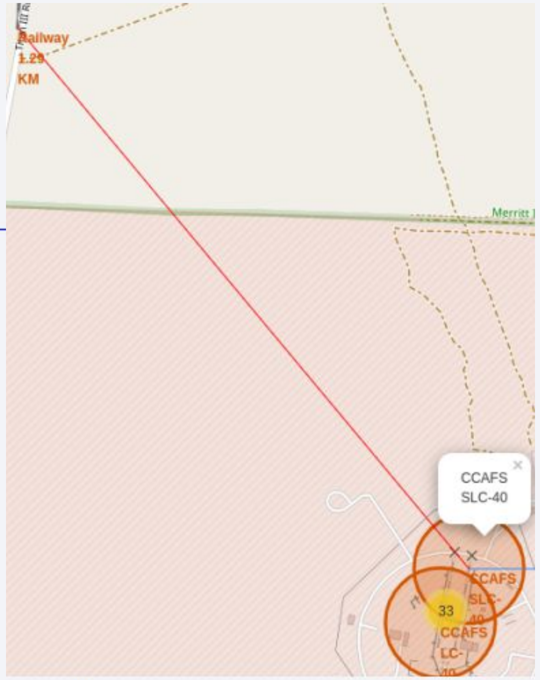
CCAFS SLC-40

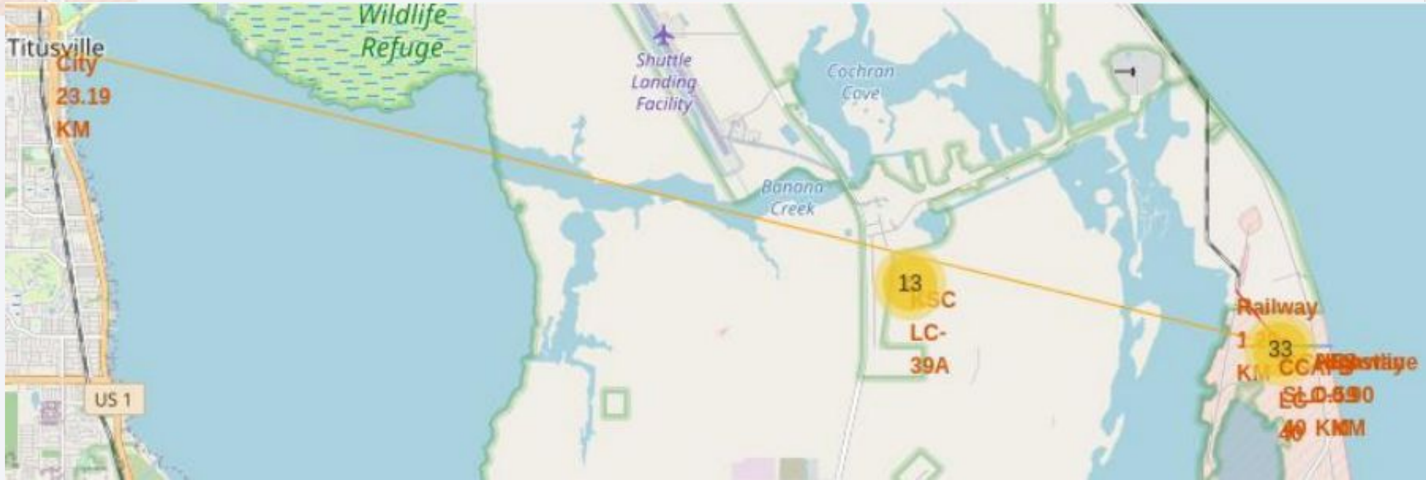


KSC LC-39A

VAFB SLC-4E

Launch Sites Logistic Location





Section 4

Build a Dashboard with Plotly Dash

<Dashboard Screenshot 1>

- Replace <Dashboard screenshot 1> title with an appropriate title

- Show the screenshot of launch success count for all sites, in a piechart

- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 2>

- Replace <Dashboard screenshot 2> title with an appropriate title
- Show the screenshot of the piechart for the launch site with highest launch success ratio
- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 3>

- Replace <Dashboard screenshot 3> title with an appropriate title

- Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider

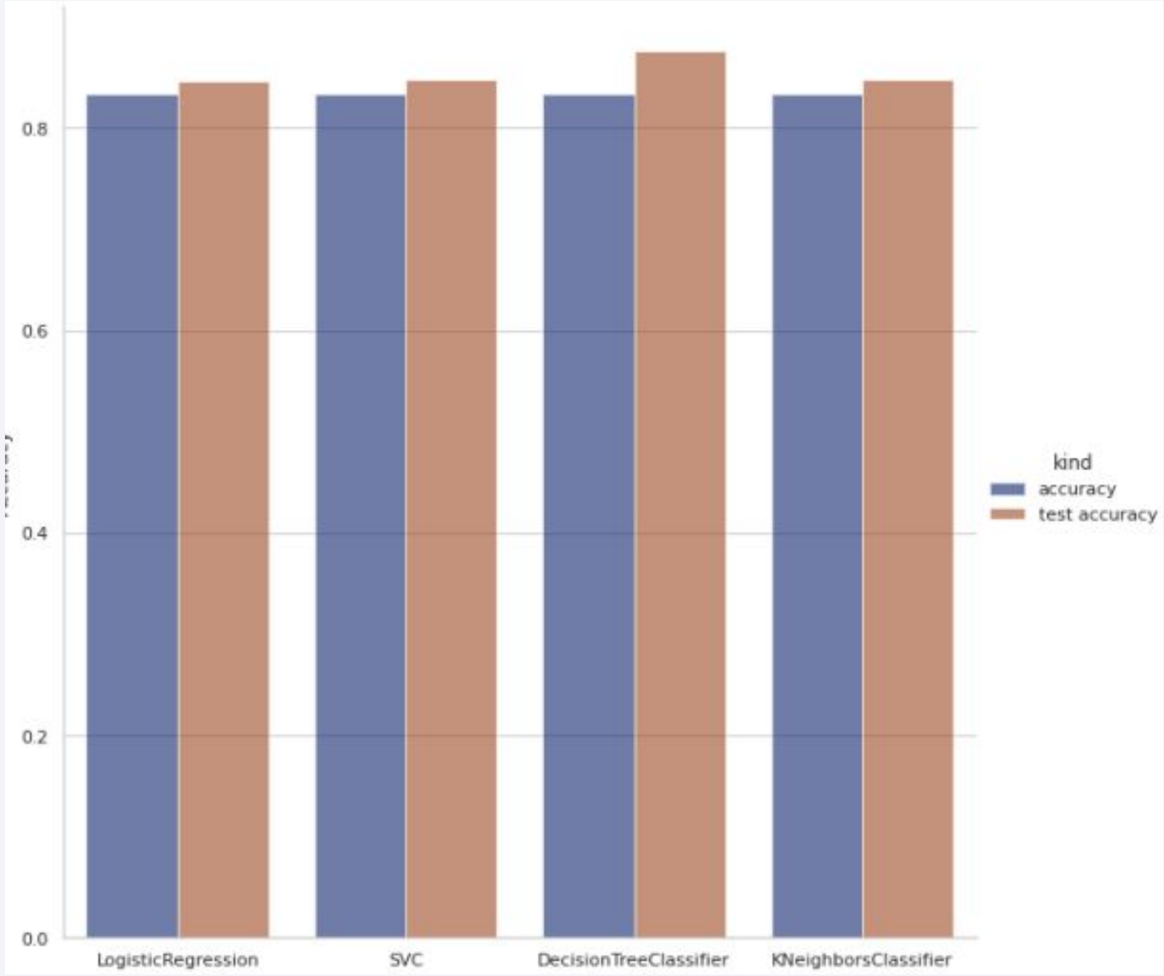
- Explain the important elements and findings on the screenshot, such as which payload range or booster version have the largest success rate, etc.

Section 5

Predictive Analysis (Classification)

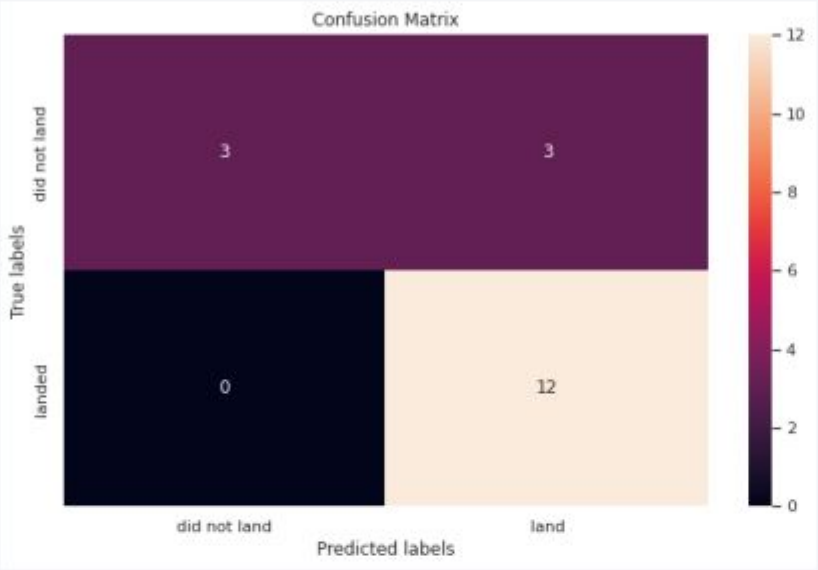
Classification Accuracy

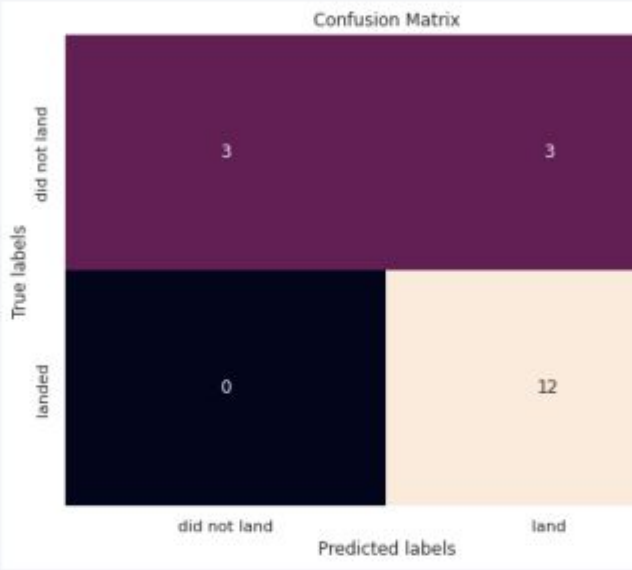
- Decision Trees has a higher accuracy on test data
- The accuracy for all the models is 83.33%

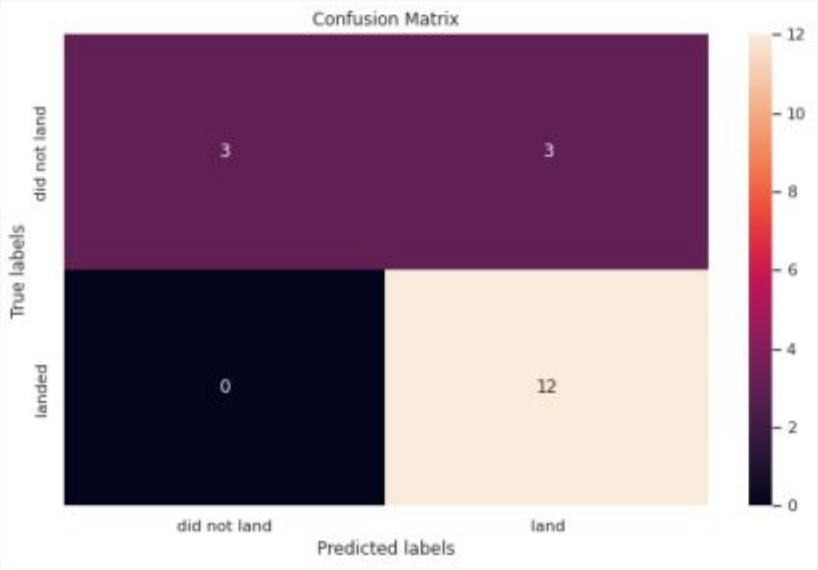


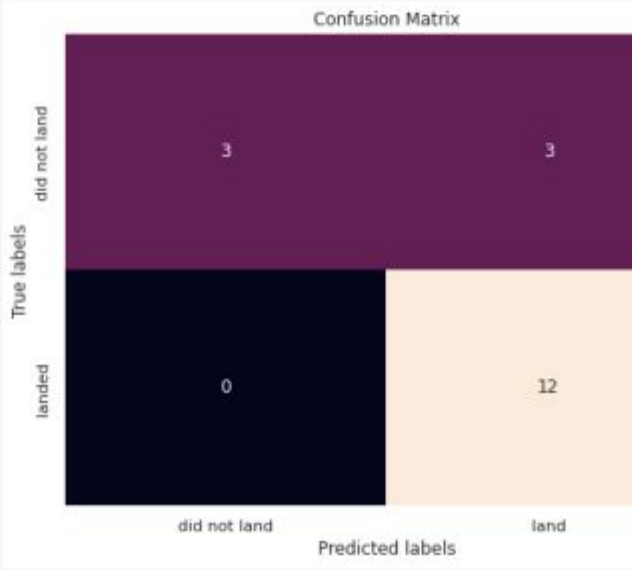
Confusion Matrix

- As the accuracy ia the same for all models , the confusion matrix is also the same.









Conclusions

- The success of the first stage landing is increasing over the years
- All the launch sites are strategically located, near coastlines and far away from cities and populated areas.
- Launch site, payload mass, and number of previous launchers are the factors for successful mission

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!