IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Sung A Kim>
<2023-09-15>

Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

- Project background and context
- Problems you want to find answers

Section
1
Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected from the SpaceX website using HTTP request API
- Perform data wrangling
  - Missing values have been replaced with the mean
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Data has been standardized, split into train/test datasets, and then best hyperparameter was found for SVM, Classification Trees and Logistic Regression

Data Collection

- Data was collected from the SpaceX website using HTTP request API

Data Collection – SpaceX API

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

- Add the GitHub URL of the completed SpaceX API calls notebook (must include completed code cell and outcome cell), as an external reference and peer-review purpose

GET

```
requests.get(spacex_url)
```

NORMALIZE

```
data = pd.json_normalize(response.json())
```

FILL NAN

```
data_falcon9.isnull().sum()
```

Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts

- Add the GitHub URL of the completed web scraping notebook, as an external reference and peer-review purpose

GET

```
r = requests.get(static_url)
```

BEAUTIFUL SOUP PARSER

```
soup = BeautifulSoup(data, "html.parser")
```

FIND TABLES

```
html_tables = soup.find_all('table')
```

Data Wrangling

- Calculate the number of launches on each site
- Calculate the number and occurrence of each orbit
- Calculate the number and occurence of mission outcome per orbit type
- Create a landing outcome label from Outcome column

EDA with Data Visualization

- CATPLOT
  - Visualize the relationship between Flight Number and Launch Site
  - Visualize the relationship between Payload and Launch Site
  - Visualize the relationship between FlightNumber and Orbit type

- BAR CHART
  - Visualize the relationship between success rate of each orbit type

- LINE CHART
  - Visualize the launch success yearly trend

EDA with SQL

- read_sql('select * from spacexdata', conn)
- read_sql('select distinct Launch_Site from spacexdata', conn)
- read_sql("select * from spacexdata where Launch_Site like 'CCA%' limit 5", conn)
- read_sql("select avg(PAYLOAD_MASS__KG_) from spacexdata where Booster_Version='F9 v1.1'", conn)
- read_sql("select min(Date) from spacexdata where Landing__Outcome='Success (ground pad)'", conn)

Build an Interactive Map with Folium

- Markers for all launch sites
- Markers for the success/failed launches for each site
- Link of distances between a launch site to its proximities

Build a Dashboard with Plotly Dash

- Dropdown list to enable Launch Site selection
- Pie chart to show the total successful launches count for all sites
- Slider to select payload range
- Scatter chart to show the correlation between payload and launch success

Predictive Analysis (Classification)

- Standardize the data
- Split into training data and test data
- Find best Hyperparameter for SVM, Classification Trees and Logistic Regression
- Find the method performs best using test data

Results

- Scores on test data for each method
    - Logistic Regression: 0.944
    - SVM: 0.944
    - Decision Tree: 0.888
    - KNN: 0.888

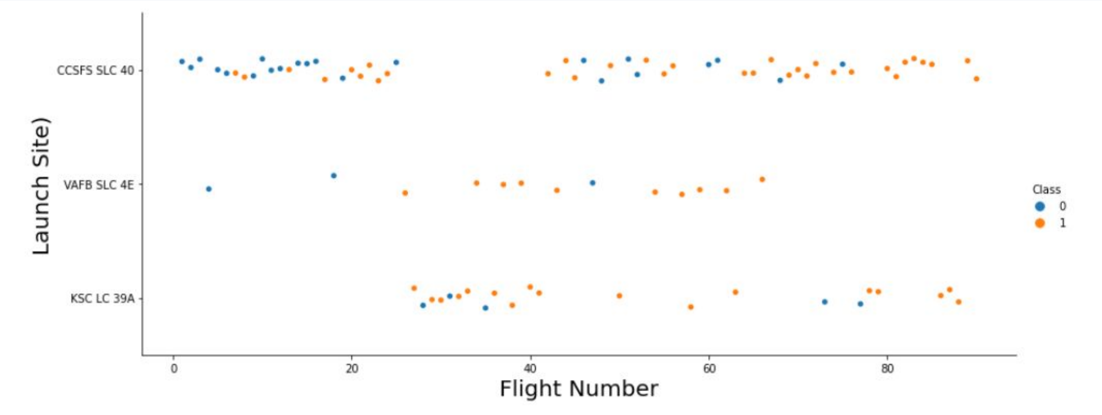
Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

- Low and high flight numbers have been largely launched from CCSFS SLC 40

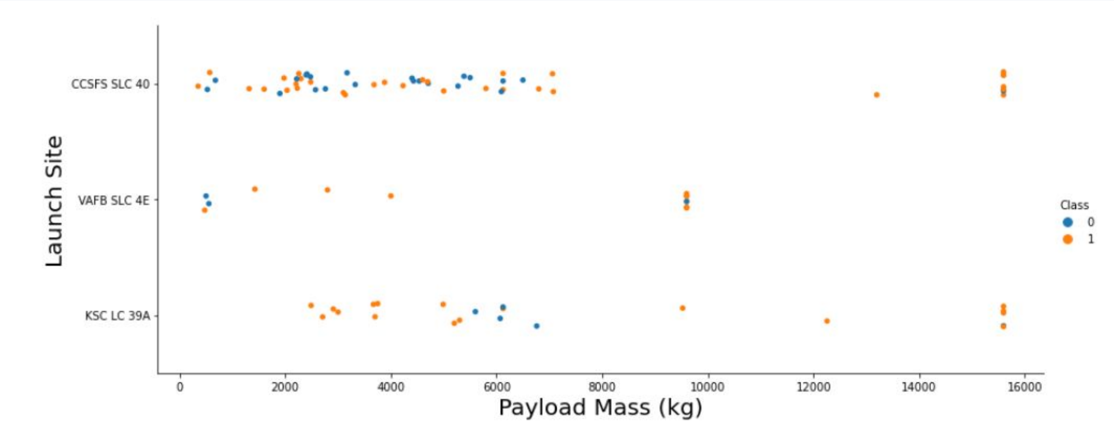
- Mid flight numbers have been launched from VAFB SLC 4E



Payload vs. Launch Site

- Lighter payloads have been largely launched from CCSFS SLC 40

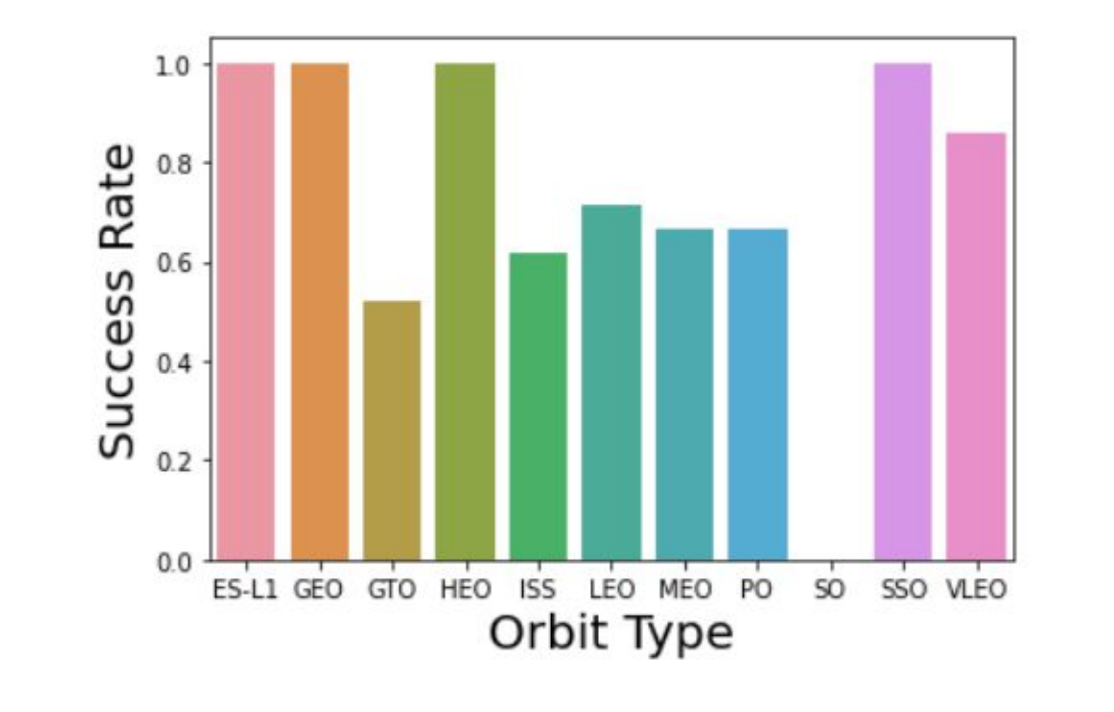
- Mid weight payloads have been launched from VAFB SLC 4E



Success Rate vs. Orbit Type

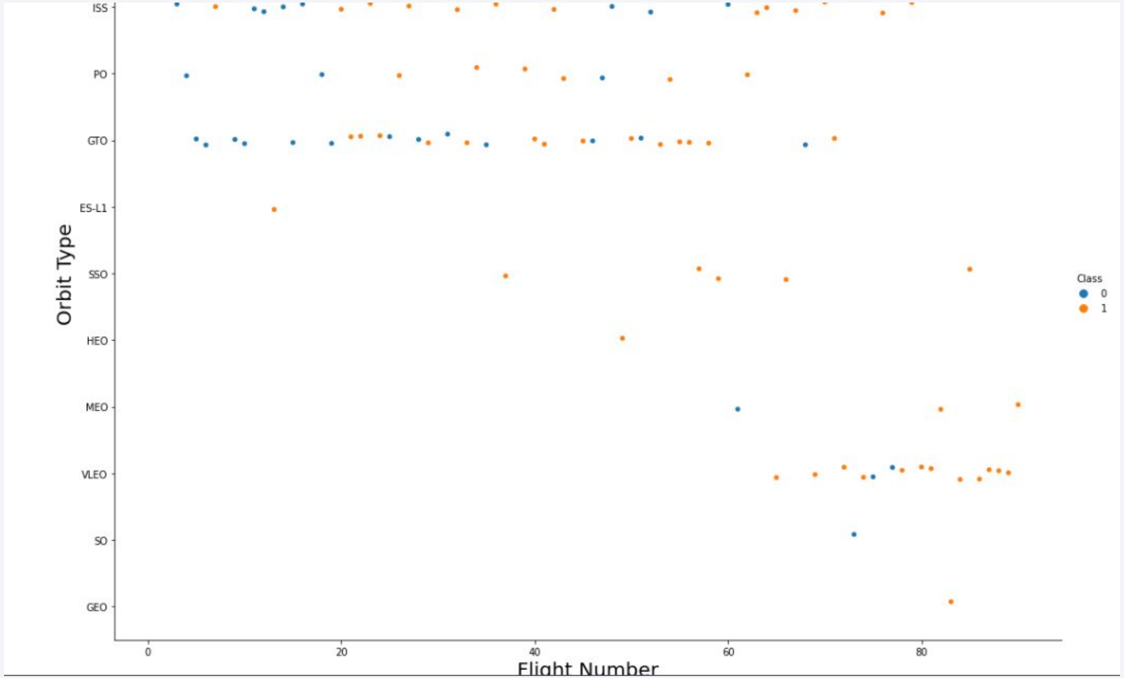
- Almost all orbit types except SO had success rate higher than 50%

- Success rate was exceptional for orbit type ES-L1, GEO, HEO, SSO with 100%



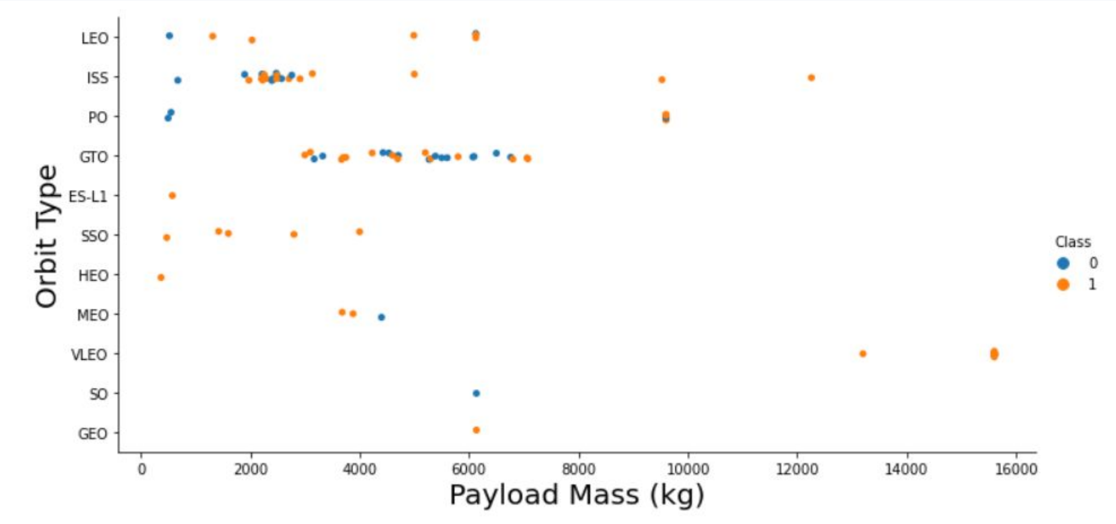
Flight Number vs. Orbit Type

- For LEO, higher flight number resulted in higher success rate
- For GTO, there is no such relationship



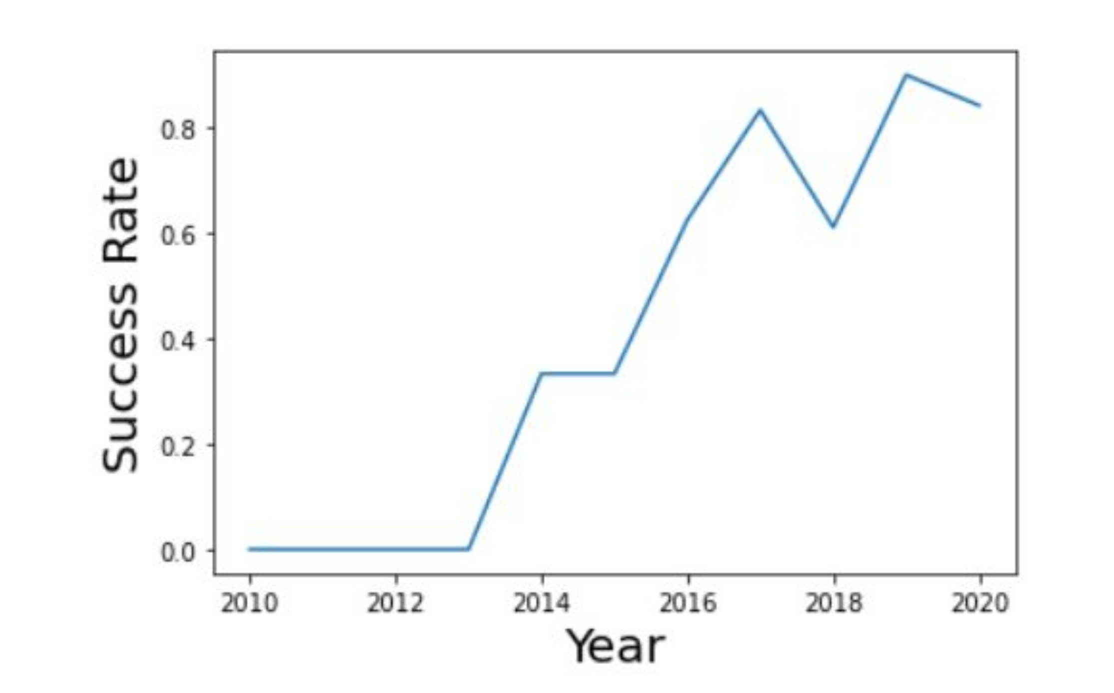
Payload vs. Orbit Type

- For Polar,LEO and ISS, higher payload resulted in higher success rate



Launch Success Yearly Trend

- Success rate has been increasing since 2013



All Launch Site Names

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

- pd.read_sql('select distinct Launch_Site from spacexdata', conn)

Launch Site Names Begin with 'CCA'

- pd.read_sql("select * from spacexdata where Launch_Site like 'CCA%' limit 5", conn)

	index	Date	Time_(UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer
0	0	2010-06-04 00:00:00	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX
1	1	2010-12-08 00:00:00	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
2	2	2012-05-22 00:00:00	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)
3	3	2012-10-08 00:00:00	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)
4	4	2013-03-01 00:00:00	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)

Total Payload Mass

- 45596
- pd.read_sql("select sum(PAYLOAD_MASS__KG_) from spacexdata where Customer='NASA (CRS)'", conn)

Average Payload Mass by F9 v1.1

- 2928.4
- pd.read_sql("select avg(PAYLOAD_MASS__KG_) from spacexdata where Booster_Version='F9 v1.1'", conn)

First Successful Ground Landing Date

- 2015-12-22 00:00:00

- pd.read_sql("select min(Date) from spacexdata where Landing__Outcome='Success (ground pad)'", conn)

Successful Drone Ship Landing with Payload between 4000 and 6000

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

- pd.read_sql("select distinct Booster_Version from spacexdata where Landing__Outcome='Success (drone ship)' and PAYLOAD_MASS__KG_ between 4000 and 6000", conn)

Total Number of Successful and Failure Mission Outcomes

- Failure 1
- Success 100

- pd.read_sql("select substr(Mission_Outcome,1,7) as Mission_Outcome, count(*) from spacexdata group by 1", conn)

Boosters Carried Maximum Payload

- 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

- pd.read_sql("select distinct Booster_Version from spacexdata where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_) from spacexdata)", conn)

2015 Launch Records

Landing__Outcome	Booster_Version	Launch_Site
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40
Failure (drone ship)	F9 v1.1 B1017	VAFB SLC-4E
Failure (drone ship)	F9 FT B1020	CCAFS LC-40
Failure (drone ship)	F9 FT B1024	CCAFS LC-40

- pd.read_sql("select distinct Landing__Outcome, Booster_Version, Launch_Site from spacexdata where Landing__Outcome='Failure (drone ship)'", conn)

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

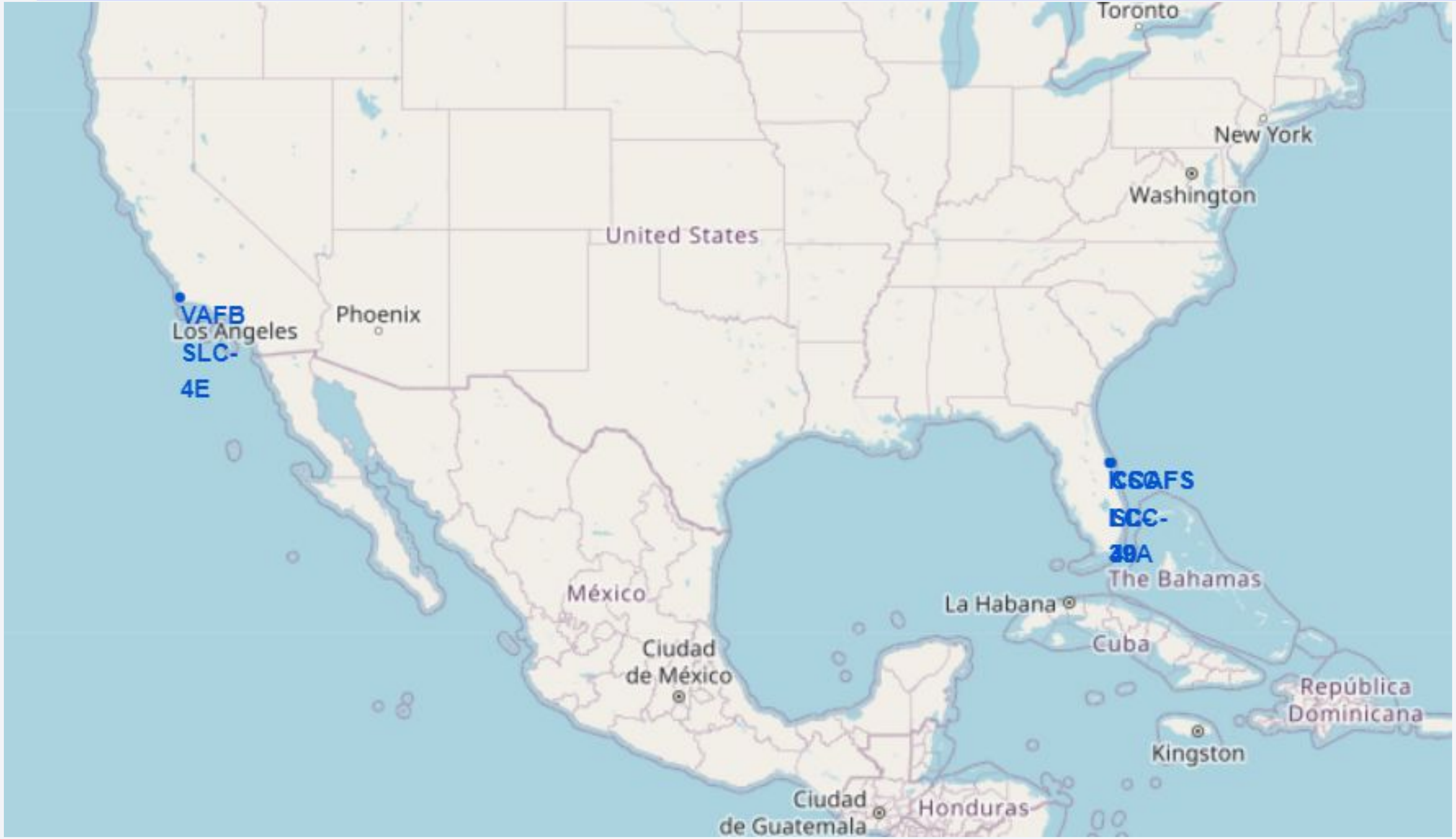
- No attempt 10
- Success (drone ship) 5
- Failure (drone ship) 5
- Success (ground pad) 3
- Controlled (ocean) 3
- Uncontrolled (ocean) 2
- Precluded (drone ship) 1

- pd.read_sql("select Landing__Outcome, count(*) from spacexdata where Date between '2011-06-04' and '2017-03-20' group by Landing__Outcome order by 2 desc", conn)

Section
3

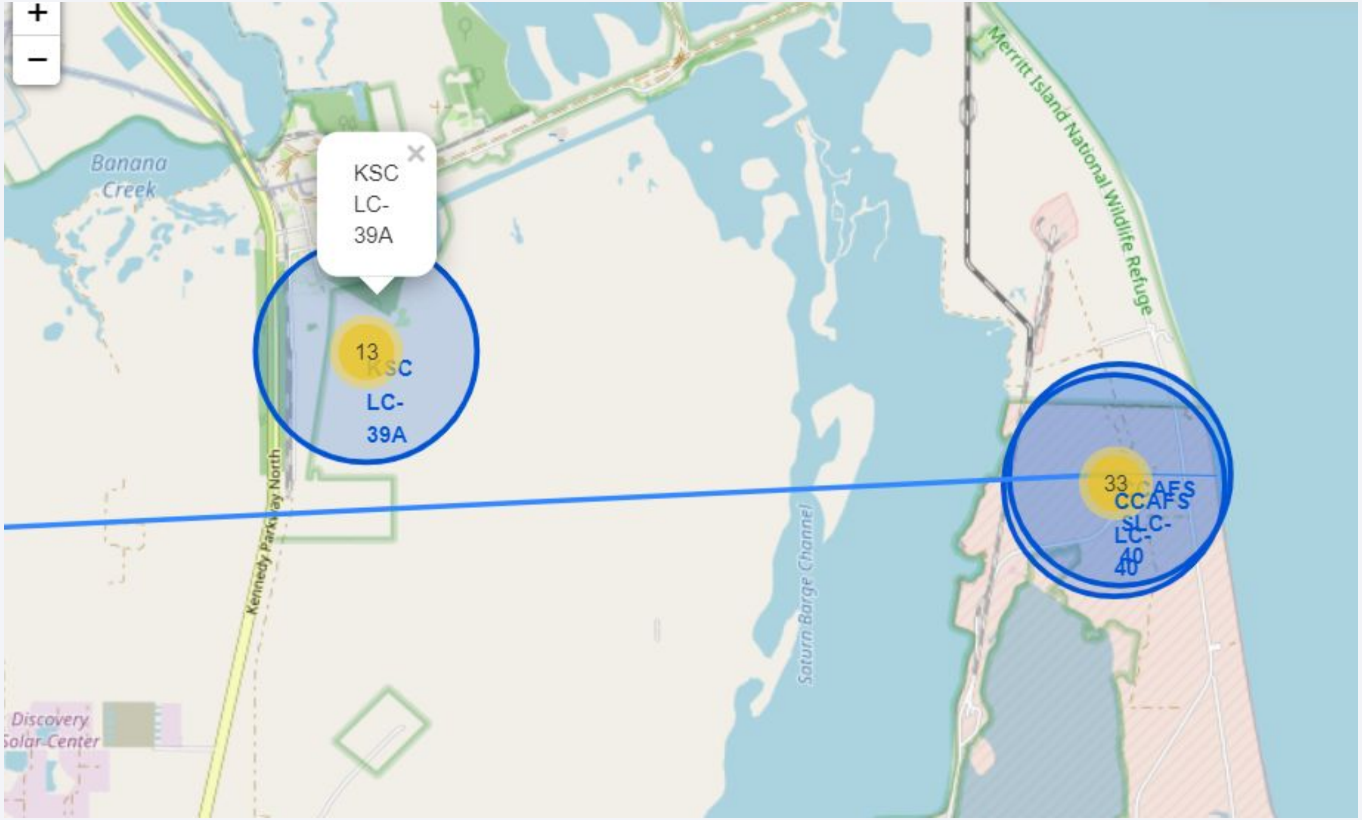
Launch Sites Proximities Analysis

<Folium Map Screenshot 1>



<Folium Map Screenshot 2>

<Folium Map Screenshot 3>



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)

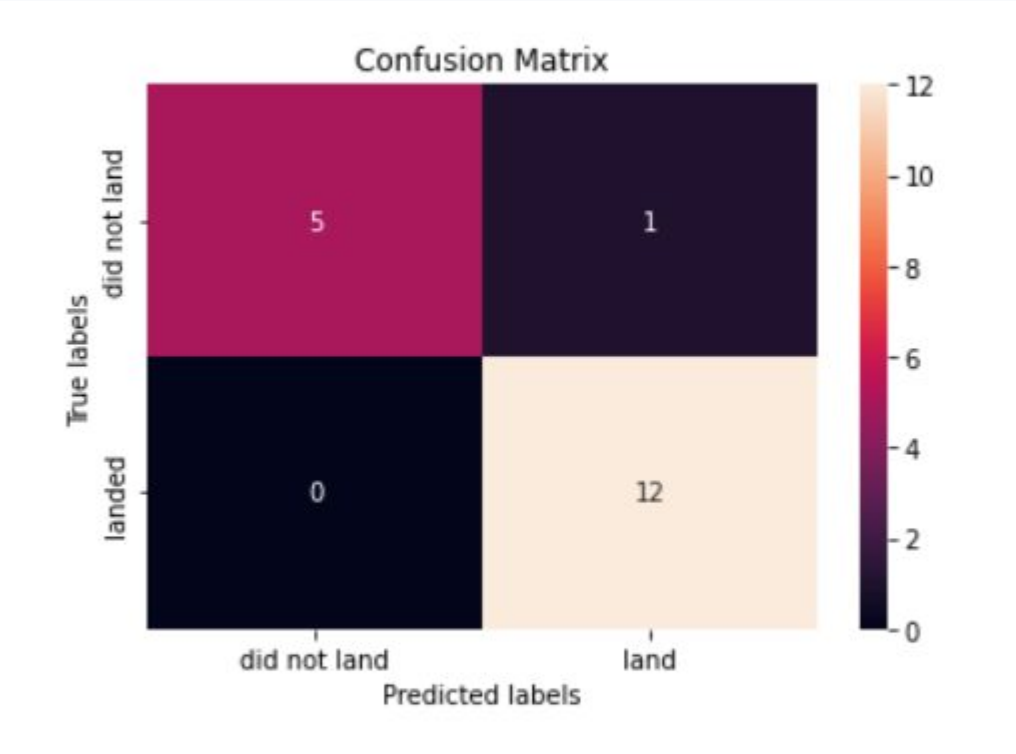
Classification Accuracy

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

- Find which model has the highest classification accuracy

Confusion Matrix

- SVM performed well with the score of 0.944



Conclusions

- SpaceX data was collected, analyzed as well as visualized to evaluate different relationships among the variables and the success rate
- SVM performed well among the classification models tested with the score of 0.944

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!