SpaceX Analysis Capstone Project

WeiKai Chao 2023/01/31

Outline

1

Executive Summary

Executive Summary

- Data collection
- Data preworking
- EDA with data visualization
- EDA with SQL
- Building an interactive map with Folium
- Buliding a dashboard with plotly dash
- Predictive analysis

2

Introduction

Introduction

- SpaceX is advertising flights on their Falcon 9 rocket for a cheaper price compared to other companies. This is due to their ability to reuse parts of the rocket, which helps them save money.

3

Methodology

Methodology

- Data collection
- Data wrangling
- Data Visualiztion
- Dashboard
- Model methods

Web Scraping

- Took a table about Falcon 9 launches from Wikipedia, changed it into a computer-readable format
- Parse the table and convert it into a Pandas data frame

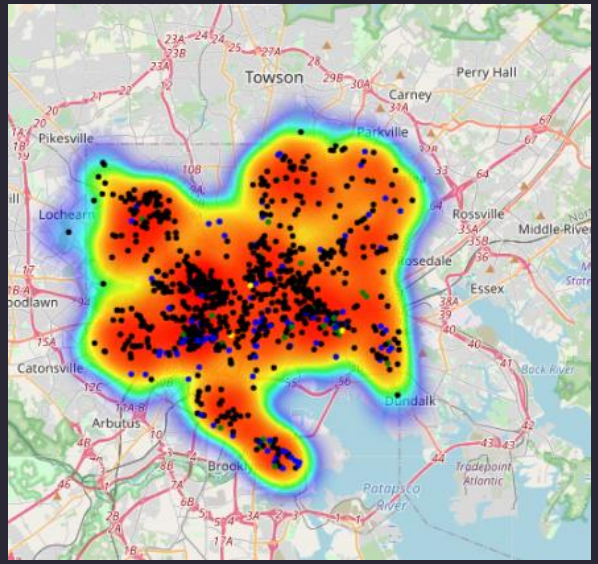
Data Wrangling

- We will examine the data to discover patterns and decide what should be the target variable for training our machine learning models.
- We will label successful landings as 1 and unsuccessful landings as 0.

EDA

- This assignment is about predicting whether the first stage of the Falcon 9 rocket will land successfully.
- SpaceX offers launches at a lower cost because they can reuse parts of the rocket.
- We will examine the data and create useful features to help us make predictions.

Analysis with Folium

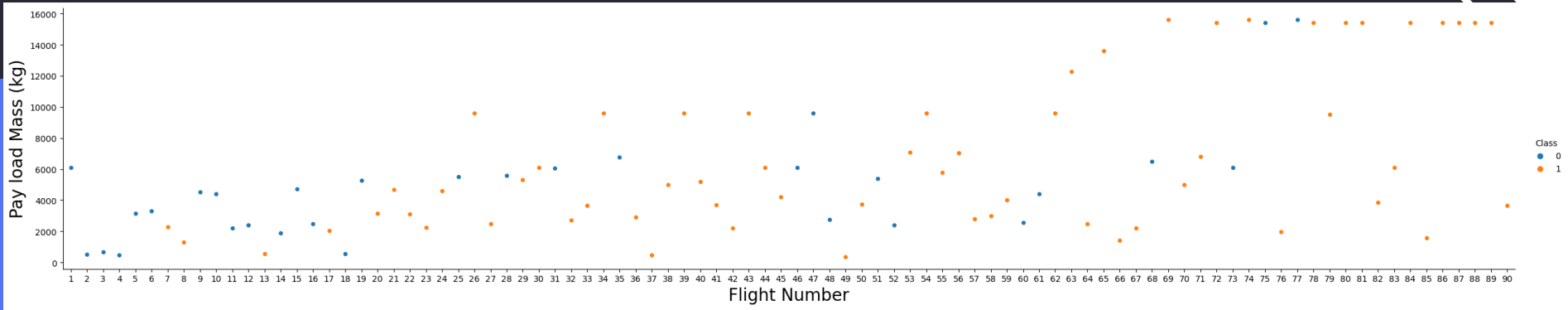


- we used matplotlib and seaborn to visualize the SpaceX launch dataset and found initial correlations between launch site and success rates.
- This lab continues the analysis with Folium, a tool that allows for interactive visual analytics to gain deeper insights into the data and make more informed predictions.

4

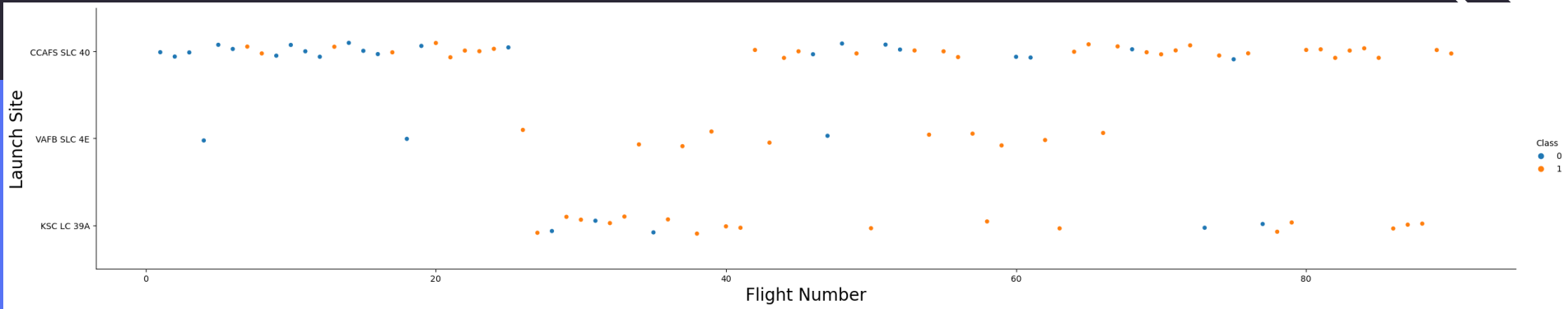
Result

EDA Visualiztion



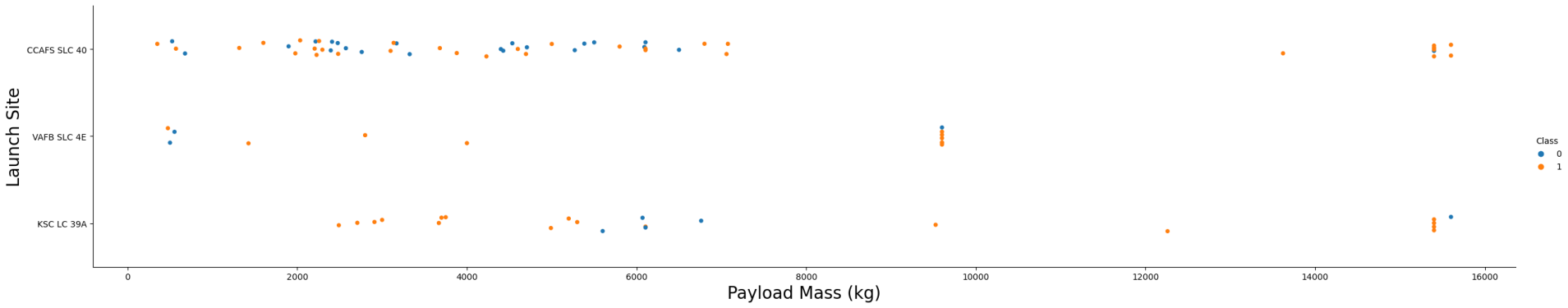
- Plot out the FlightNumber vs. PayloadMass and overlay the outcome of the launch.

EDA Visualiztion



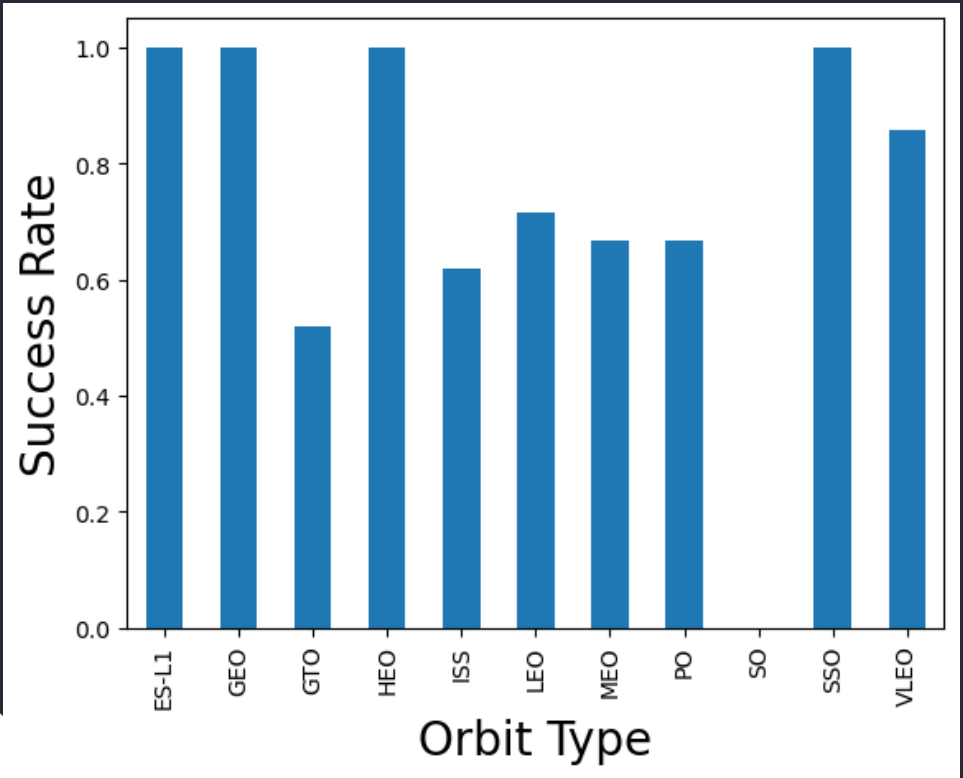
- We will create a plot to see the relationship between the flight number and payload mass, and also show the outcome of each launch. As the flight number increases, the chance of a successful first stage landing increases. But as the payload mass grows, the likelihood of a successful landing decreases.

EDA Visualiztion



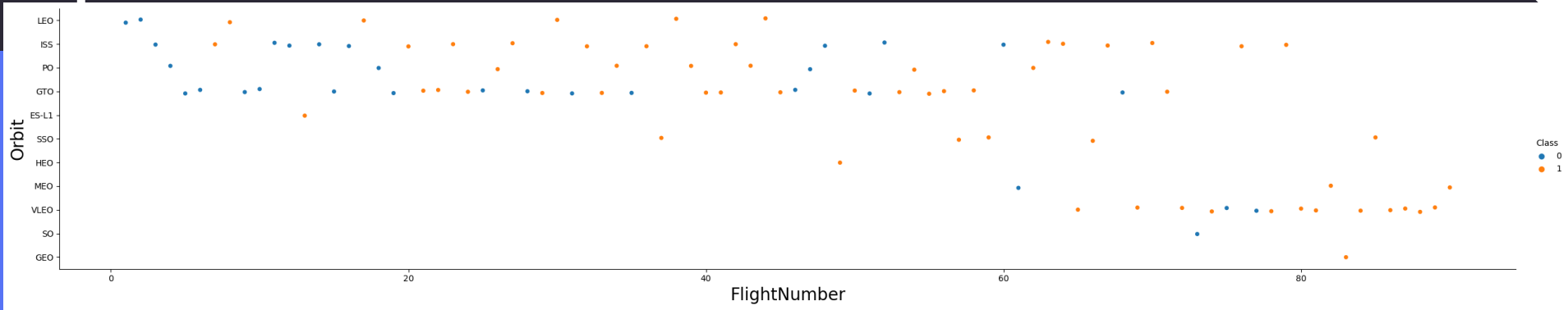
- From the picture of Payload Vs. Launch Site scatter point chart find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass.

EDA Visualiztion



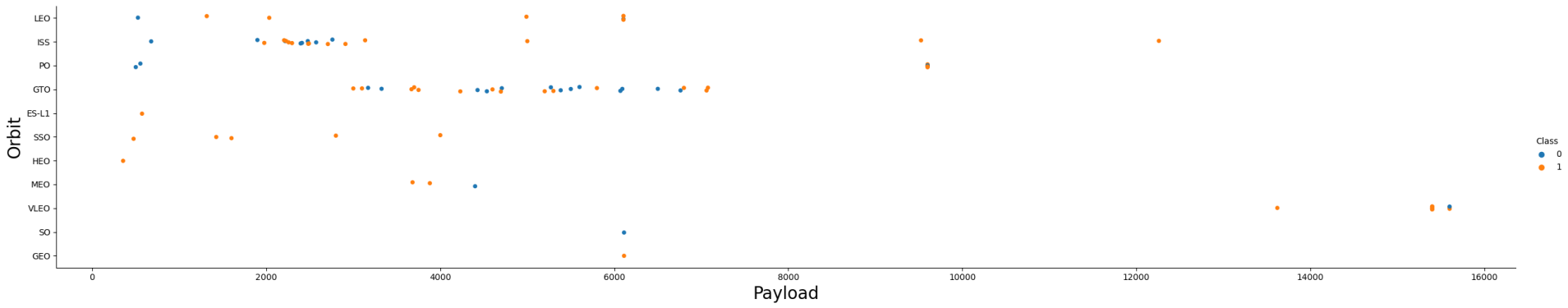
- SO is the worst orbit. It even get the 0 success rate.
- Over 80% orbits have more than 60% success rate. The result says the orbits did the incredible work.

EDA Visualiztion



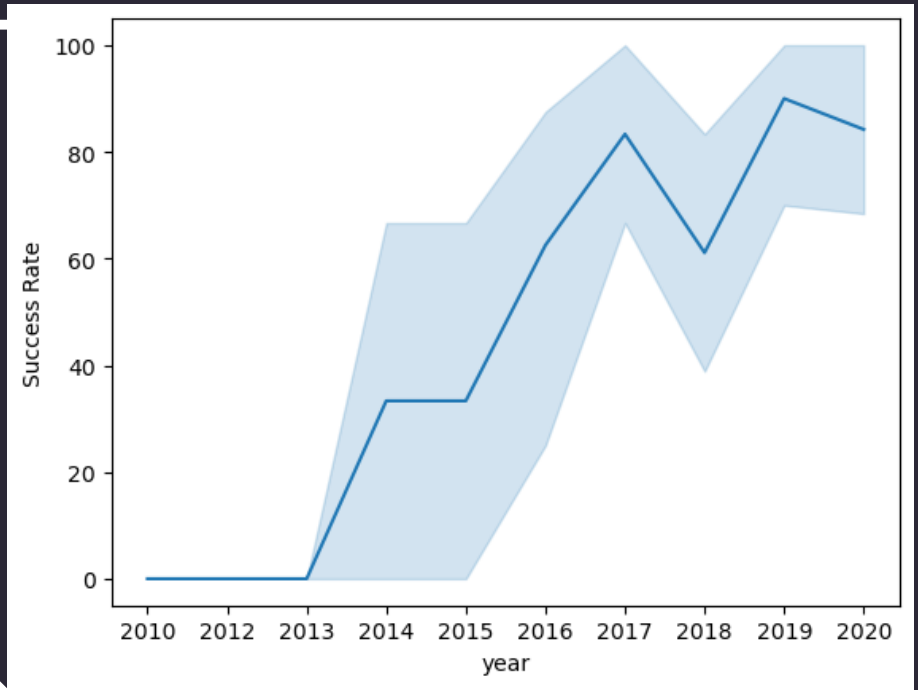
- The LEO, ISS, PO, GTO orbits got relative low number. There are range between 0 to 60.
- The SO orbit has the record between 60 to 100, and more calss 1.

EDA Visualiztion



- Failed landing rate is high for GTO (Geostationary Transfer Orbit) missions due to their high velocity and altitude.
- This makes the landing more challenging and less predictable compared to Polar, LEO (Low Earth Orbit), and ISS (International Space Station) missions which have a lower velocity and altitude.

EDA Visualiztion



- The success rate for launches has generally improved since 2013, with a slight dip in 2018
- In recent years, the success rate has been around 80%.

EDA with SQL

```
%%sql
SELECT DISTINCT LAUNCH_SITE
FROM SPACEX;
```

```
 * ibm_db_sa://hhk96939:***@ba99a9e
Done.
```

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

EDA with SQL

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

```
 * ibm_db_sa://hhk96939:***@ba99a9e
Done.
```

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

EDA with SQL

```
%%sql
SELECT SUM(PAYLOAD_MASS__KG_)
FROM SPACEX
WHERE Customer = 'NASA (CRS)';
```

```
 * ibm_db_sa://hhk96939:***@ba99a9e6
Done.

     1
45596
```

EDA with SQL

```
%%sql
SELECT AVG(PAYLOAD_MASS__KG_)
FROM SPACEX
WHERE Booster_Version LIKE 'F9 v1.0%';
```

```
 * ibm_db_sa://hhk96939:***@ba99a9e6-d59e-488
Done.
```

1
340

EDA with SQL

```
%%sql
SELECT MIN(Date)
FROM SPACEX
WHERE Landing__Outcome = 'Success (ground pad)';
```

```
 * ibm_db_sa://hhk96939:***@ba99a9e6-d59e-4883-8fc0-d6
Done.
```

1
2015-12-22

EDA with SQL

```
%%sql
SELECT BOOSTER_VERSION
FROM SPACEX
WHERE LANDING__OUTCOME = 'Success (drone ship)'
    AND 4000 < PAYLOAD_MASS__KG_ < 6000;
```

```
 * ibm_db_sa://hhk96939:***@ba99a9e6-d59e-4883-8fc0-c
Done.
```

booster_version
F9 FT B1021.1
F9 FT B1023.1
F9 FT B1029.2
F9 FT B1038.1
F9 B4 B1042.1
F9 B4 B1045.1
F9 B5 B1046.1

EDA with SQL

```
%%sql
SELECT MISSION_OUTCOME, COUNT(MISSION_OUTCOME) AS TOTAL_NUMBER
FROM SPACEX
GROUP BY MISSION_OUTCOME;
```

```
 * ibm_db_sa://hhk96939:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1
Done.
```

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

EDA with SQL

```
%%sql
SELECT DISTINCT BOOSTER_VERSION
FROM SPACEX
WHERE PAYLOAD_MASS__KG_ = (
    SELECT MAX(PAYLOAD_MASS__KG_)
    FROM SPACEX);
```

```
 * ibm_db_sa://hhk96939:***@ba99a9e6-d59e-488
Done.
```

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

EDA with SQL

```
%%sql
SELECT LANDING__OUTCOME, BOOSTER_VERSION, LAUNCH_SITE
FROM SPACEX
WHERE Landing__Outcome = 'Failure (drone ship)'
    AND YEAR(DATE) = 2015;
```

```
 * ibm_db_sa://hhk96939:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f
Done.
```

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

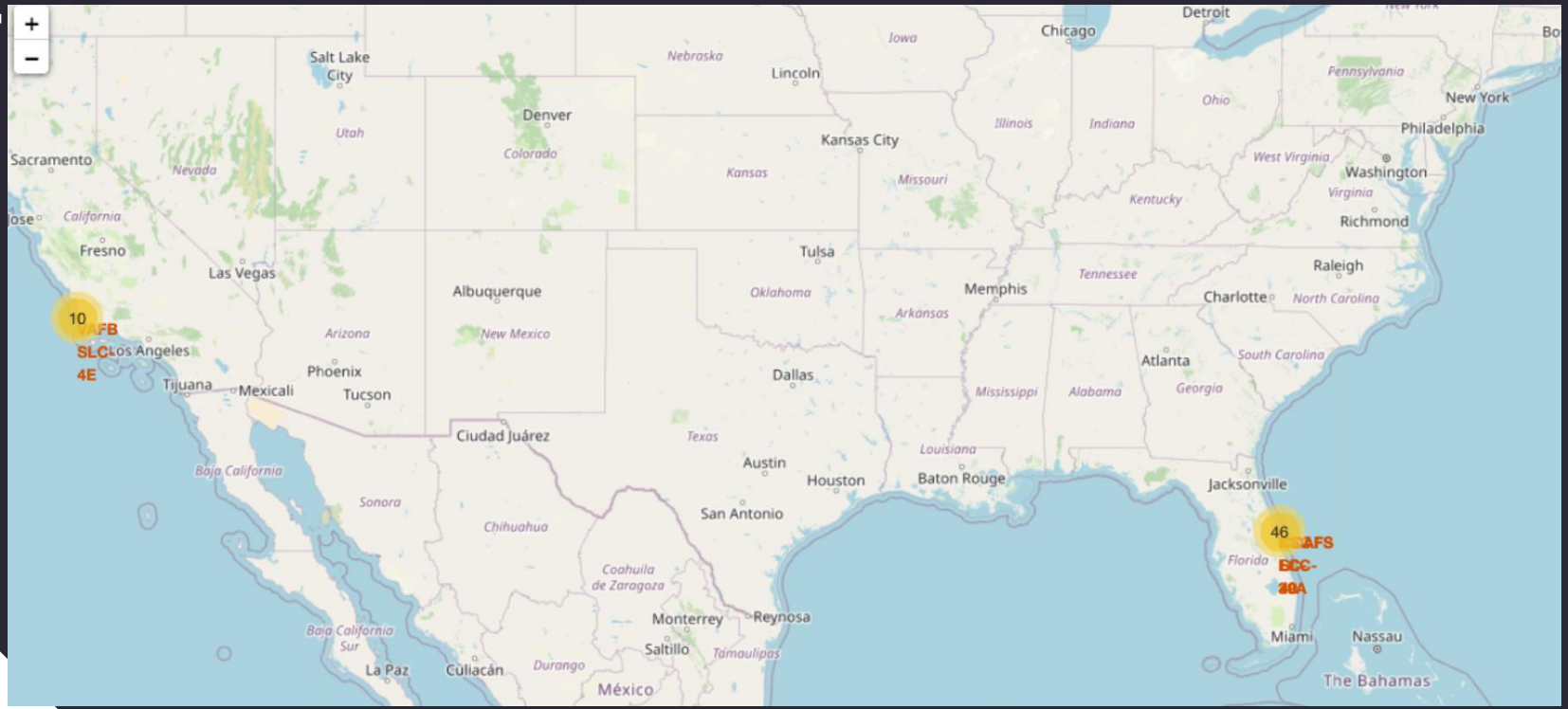
EDA with SQL

```
%%sql
SELECT LANDING__OUTCOME, COUNT(LANDING__OUTCOME) AS TOTAL_NUMBER
FROM SPACEX
WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY LANDING__OUTCOME
ORDER BY TOTAL_NUMBER DESC
```

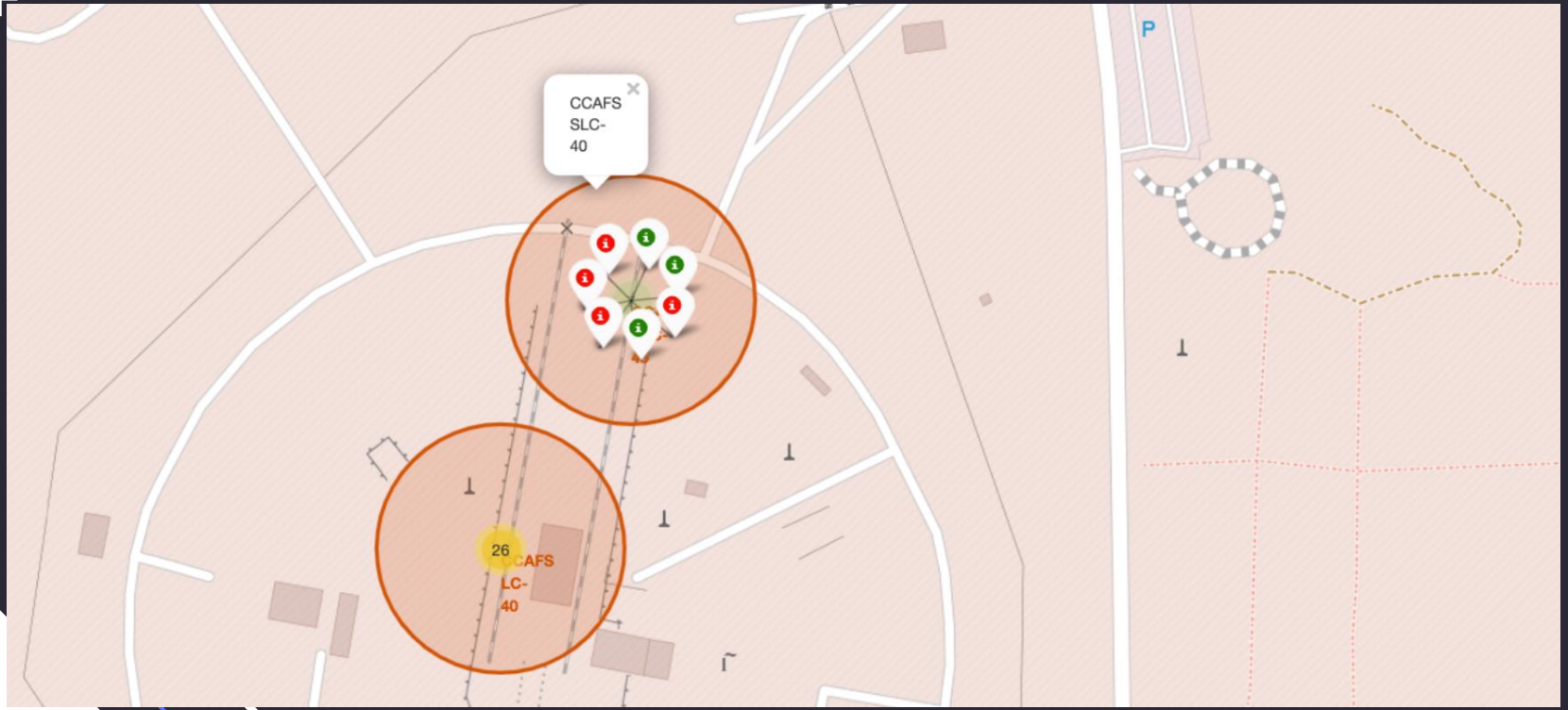
```
 * ibm_db_sa://hhk96939:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj
Done.
```

landing__outcome	total_number
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

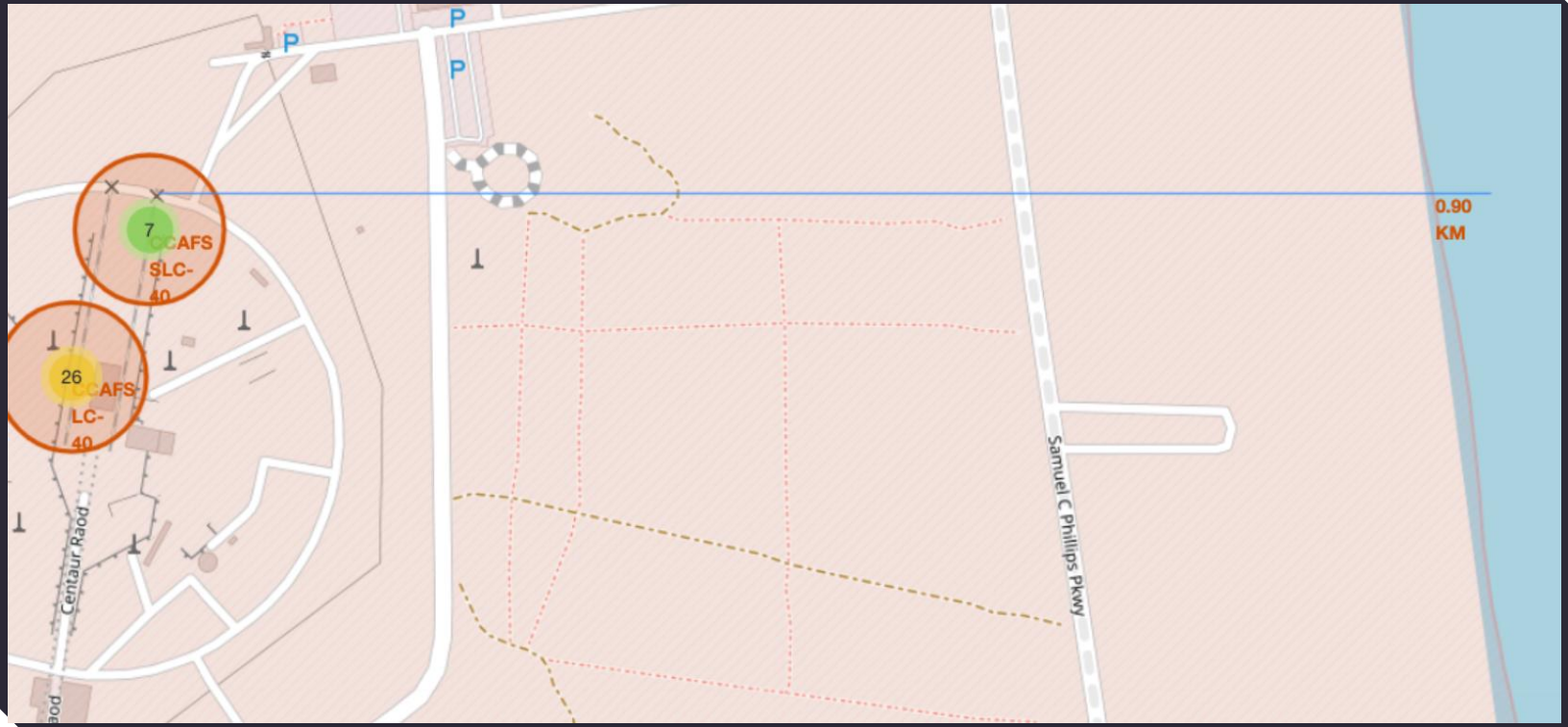
EDA with Folium



EDA with Folium

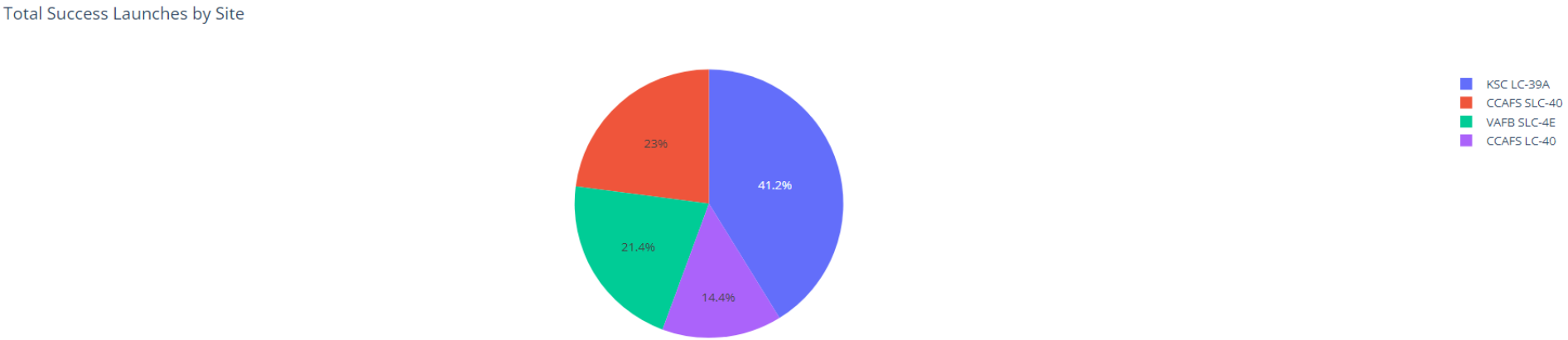


EDA with Folium



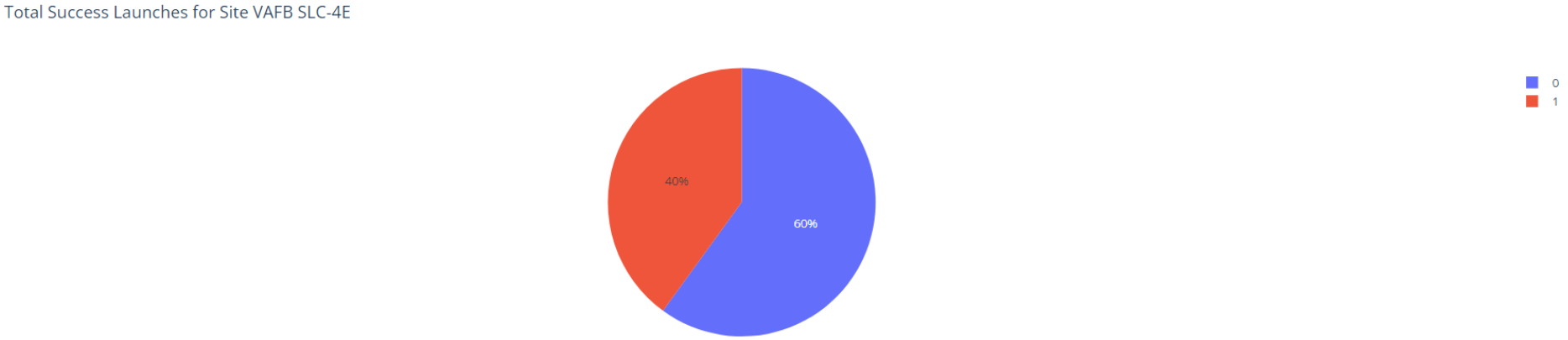
PolyLine between a launch site to the selected

Dashboard with Plotly Dash



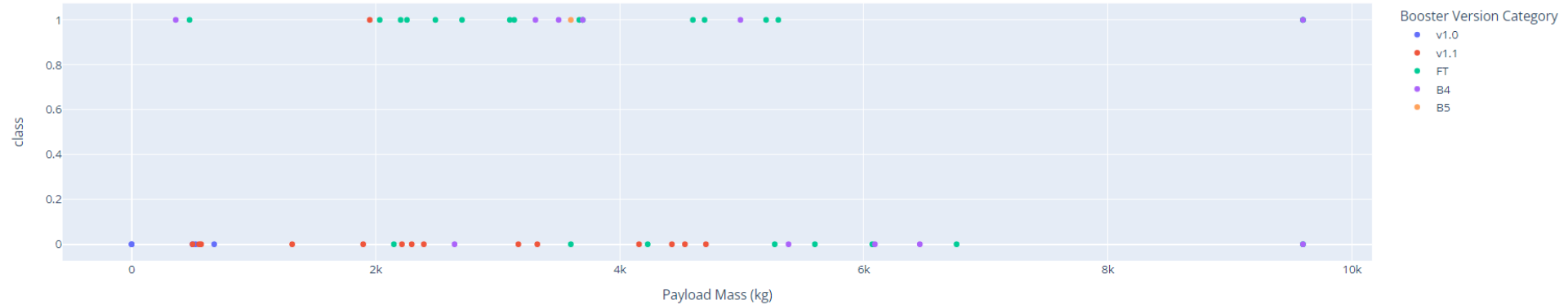
- The KSC LC-39A has the most success launches in the four sites.
- The last three sites have almost same part.

Dashboard with Plotly Dash



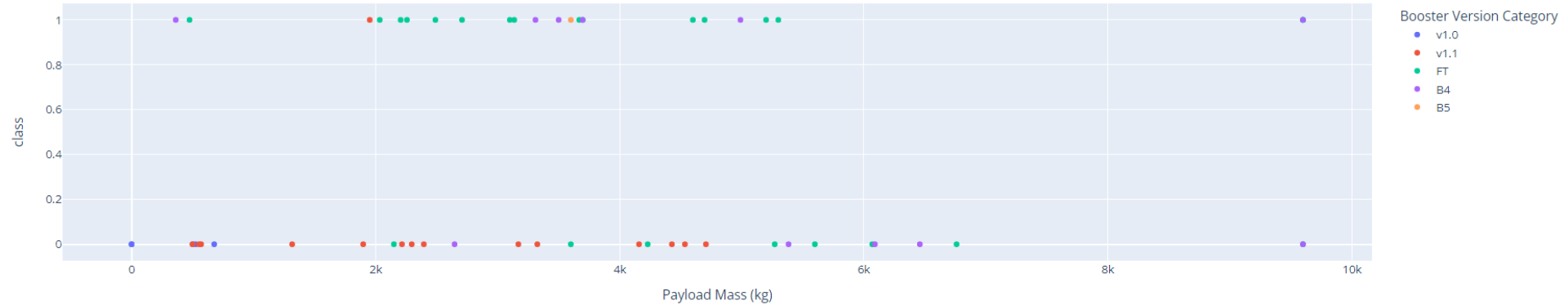
The class 0 is 60% in site VAFB SLC-4E. It is over the half of them.

Dashboard with Plotly Dash



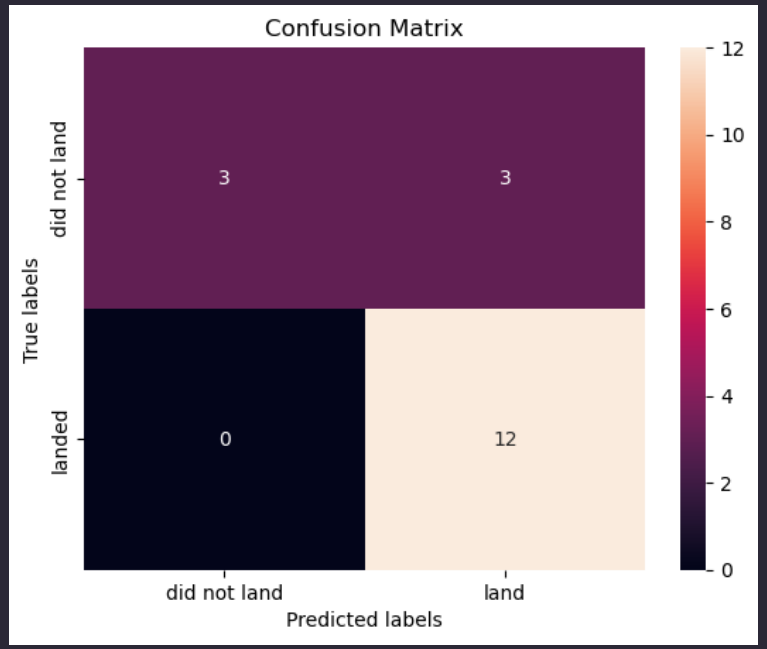
- B4 has some of outlier.
- V1.1 almost in the lower payload mass.

Dashboard with Plotly Dash



- B4 has some of outlier.
- V1.1 almost in the lower payload mass.

Machine Learning



- Use SVM
- The accuracy is 0.8333

Machine Learning

	LogReg	SVM	Tree	KNN
Jaccard_Score	0.833333	0.845070	0.805556	0.819444
F1_Score	0.909091	0.916031	0.892308	0.900763
Accuracy	0.866667	0.877778	0.844444	0.855556

- F1_score is the good metrics to evaluate the model.
- SVM is the best model, and then is the LogReg
- SVM in the three metrics all the best model.

5

Conclusion

Machine Learning

- We created a machine learning model to forecast successful Stage 1 landings for Space Y to compete with SpaceX.
- Information was gathered from SpaceX Wikipedia and API, a dashboard was created for visualization, and data was entered into a SQL database.
- Our model had 83% accuracy and can be used by Allon Mask of Space Y to make launch decisions. To improve accuracy, more data should be gathered and the optimum model should be selected.

THE END