Winning Space Race with Data Science

<Name> Ahmed Salama
<Date> 12/6/2020

Outline

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

Executive Summary

- we will predict if the Falcon 9 first stage will land successfully. Space X advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because Space X can reuse the first stage. Therefore if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against Space X for a rocket launch.

Introduction

- Space X advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars
- much of the savings is because Space X can reuse the first stage. Therefore if we can determine if the first stage will land, we can determine the cost of a launch
- we will predict if the Falcon 9 first stage will land successfully

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- 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 – SpaceX API

https://github.com/ahmedsalama95/DataScienceProjects/blob/main/01_spacex_data_collection_api.ipynb

Data Collection - Scraping

https://github.com/ahmedsalama95/DataScienceProjects/blob/main/02_labs_webscraping.ipynb

Data Wrangling

- https://github.com/ahmedsalama95/DataSciencePr
ojects/blob/main/02_spacex_Data_wrangling.ipynb

EDA with Data Visualization

- https://github.com/ahmedsalama95/DataScienceProjects/blob/main/04_EDA_dataviz.ipynb

EDA with SQL

https://github.com/ahmedsalama95/DataScienceProjects/blob/main/03_EDA_sql_sqllite.ipynb

Predictive Analysis (Classification)

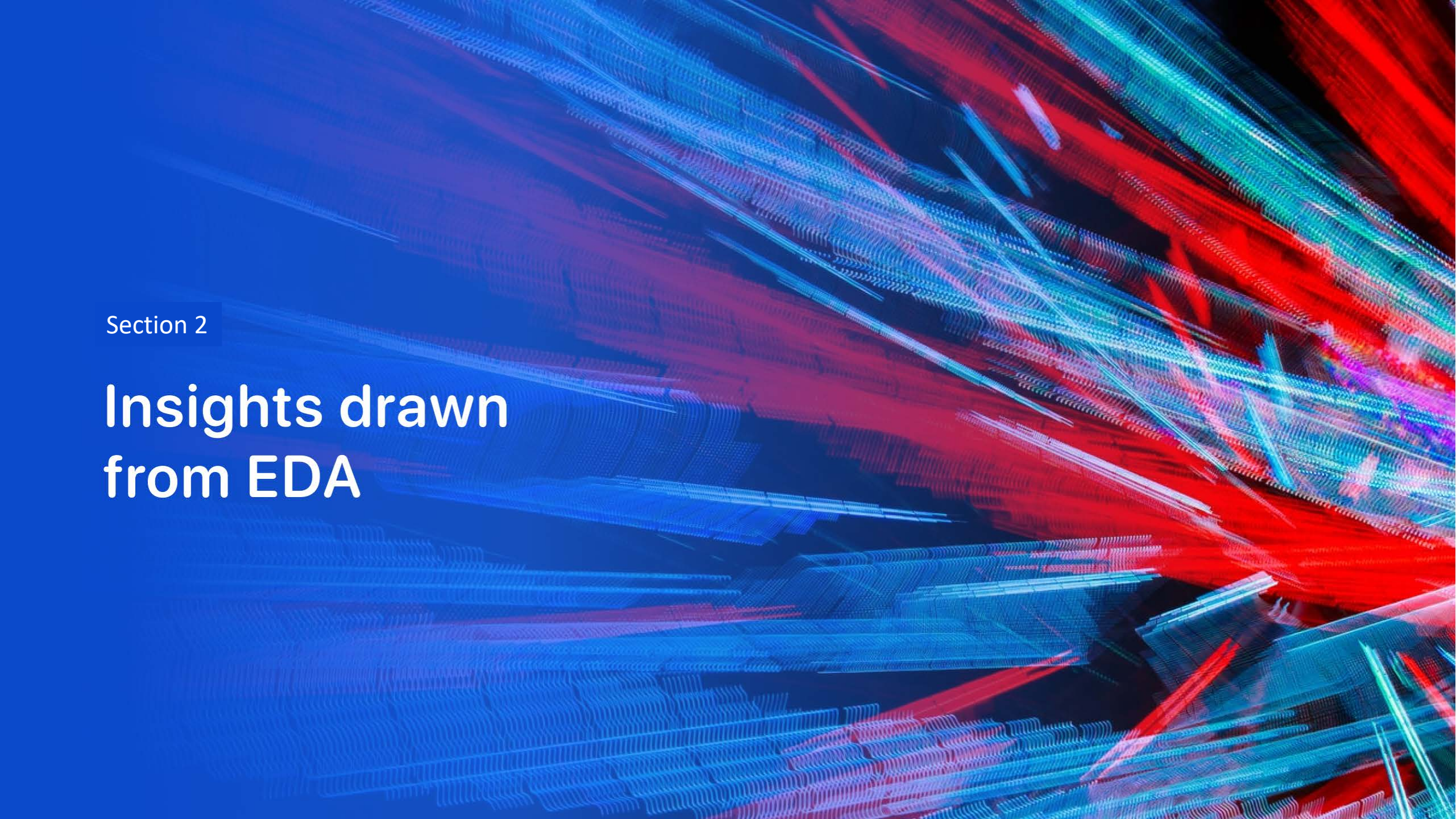
- https://github.com/ahmedsalama95/DataScienceProjects/blob/main/SpaceX_Machine_Learning_Prediction.ipynb

Results

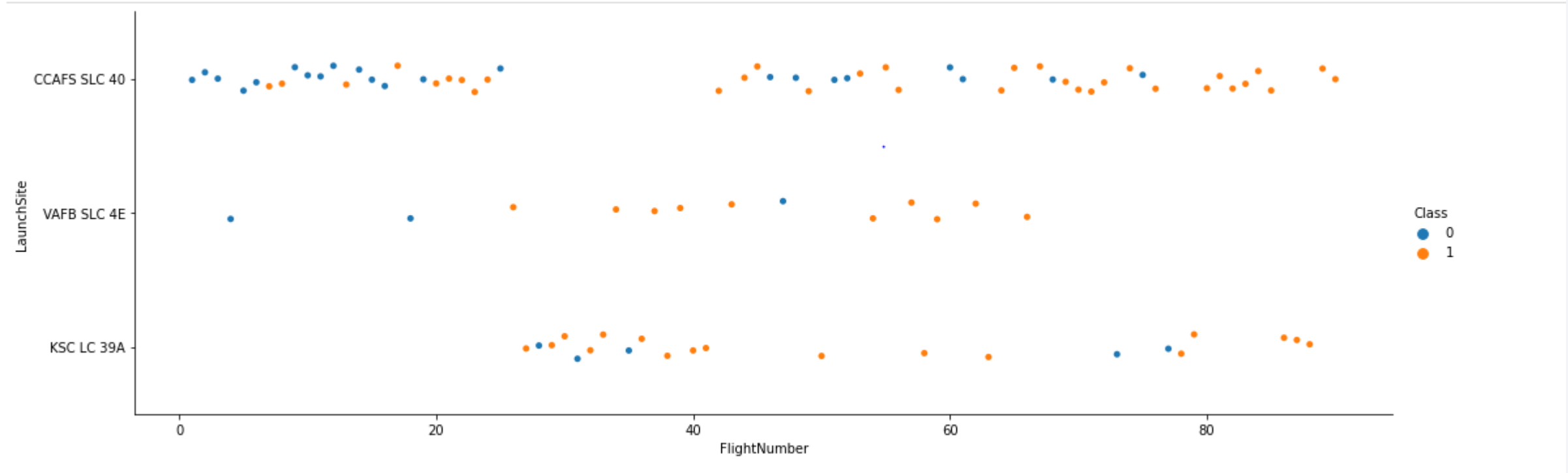
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

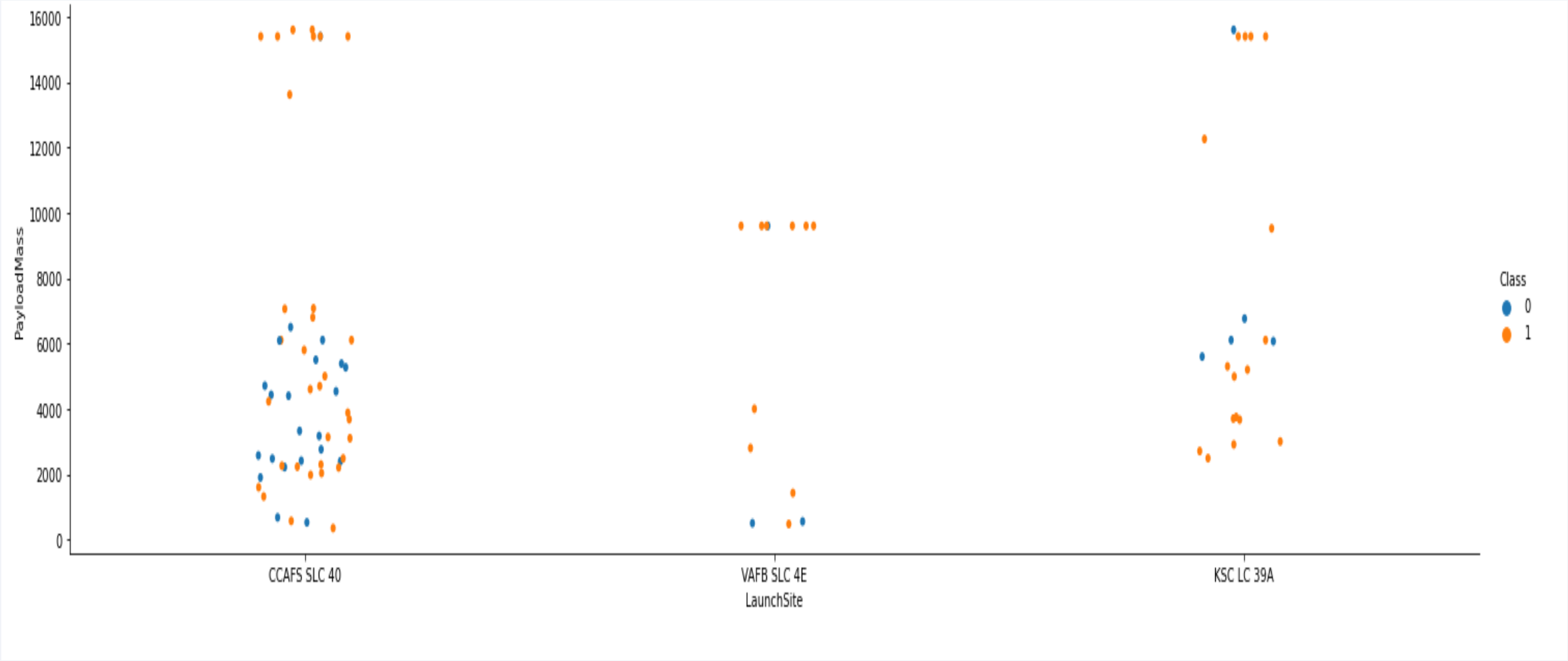
Insights drawn from EDA



Flight Number vs. Launch Site

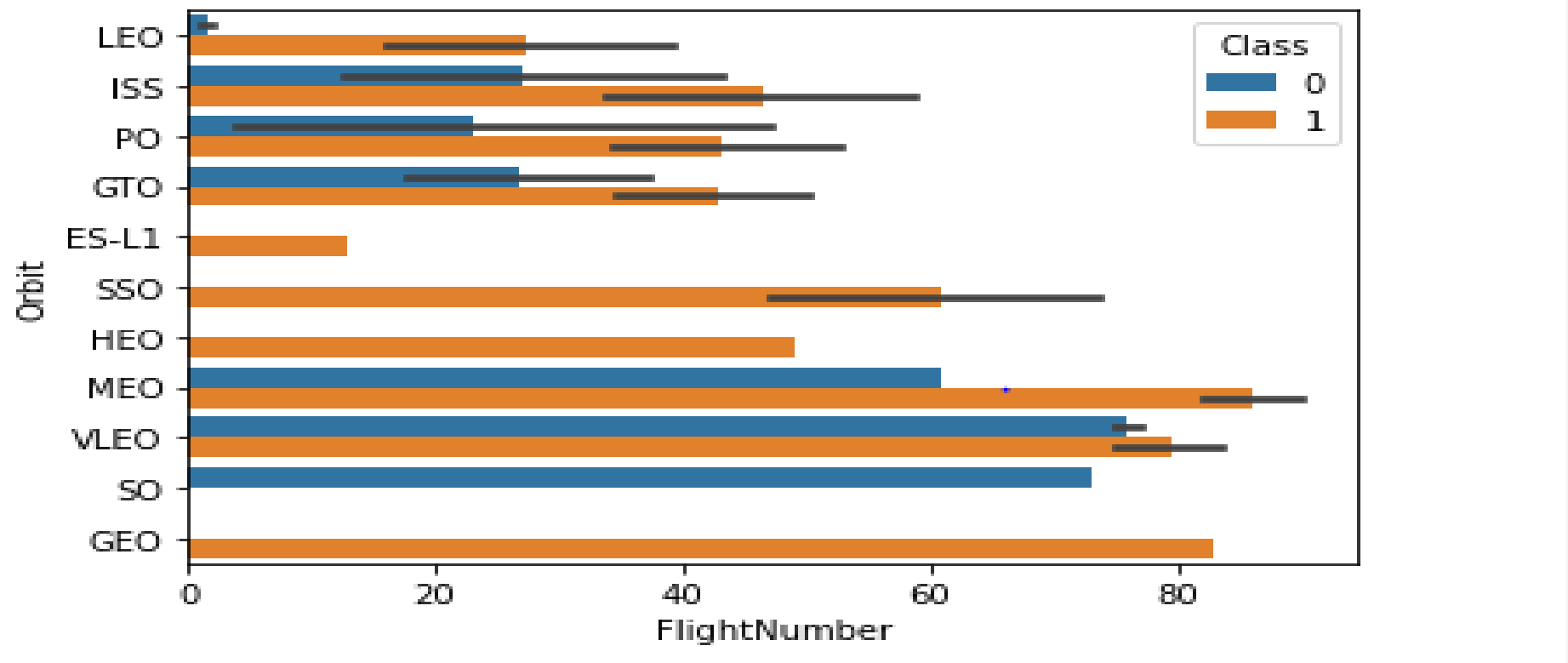


Payload vs. Launch Site

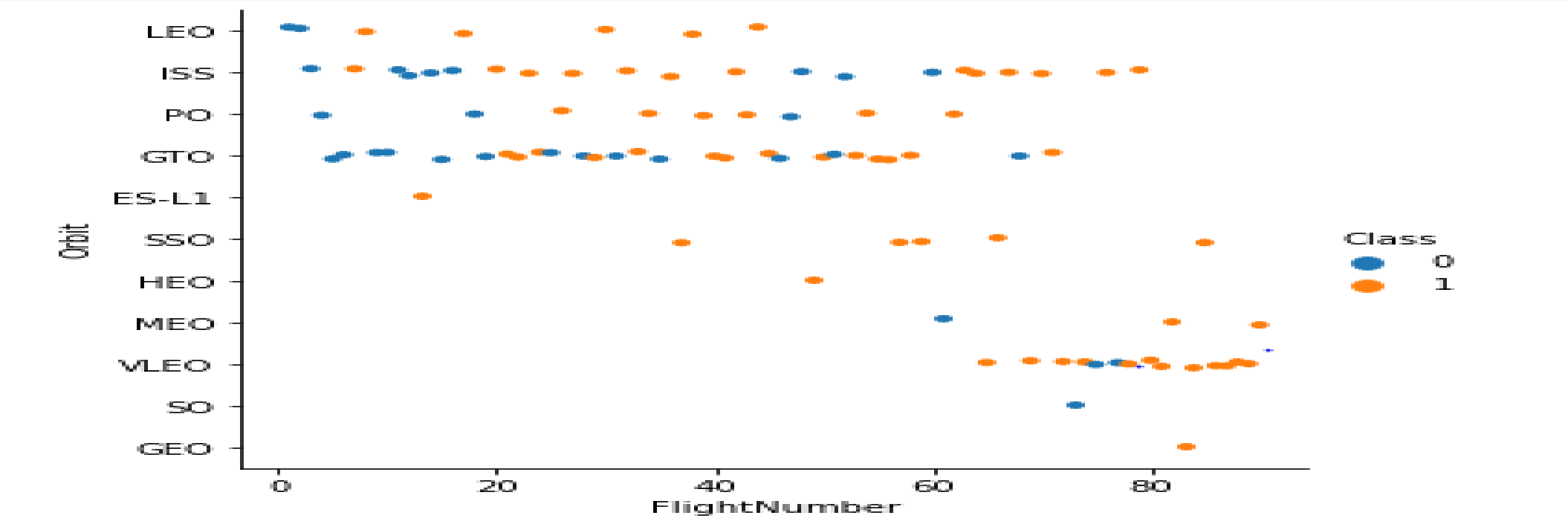


Now if you observe Payload Vs. Launch Site scatter point chart you will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000)

Success Rate vs. Orbit Type

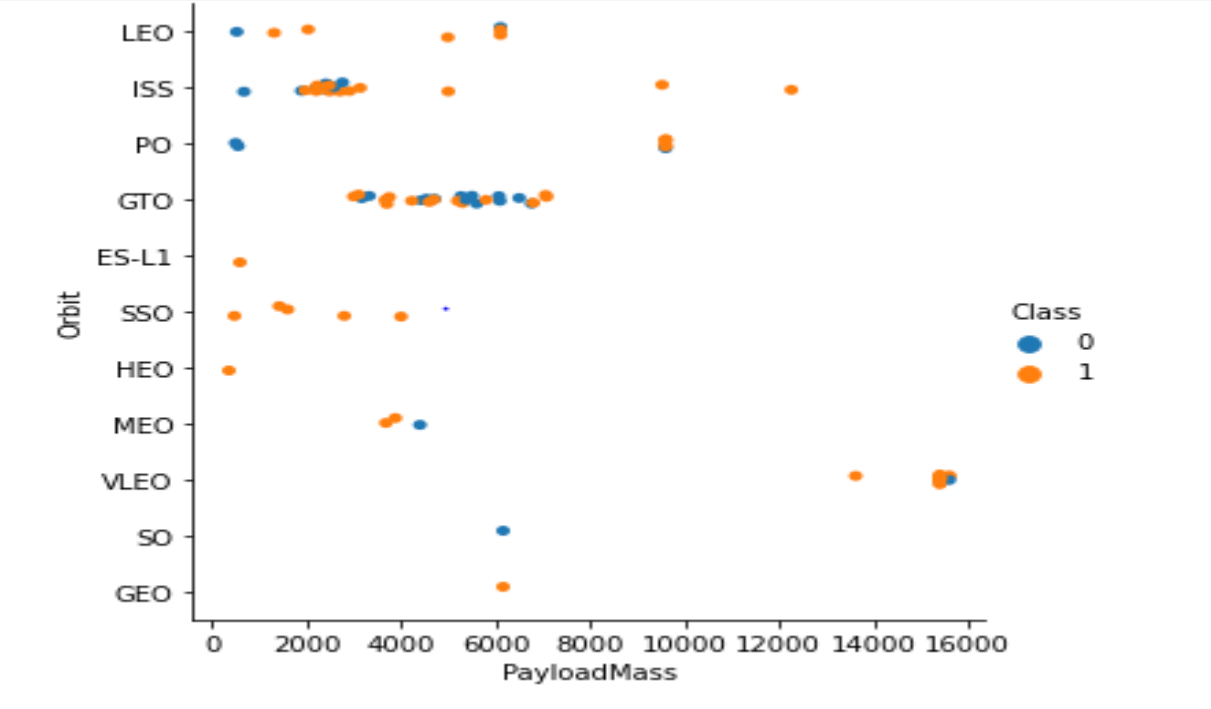


Flight Number vs. Orbit Type



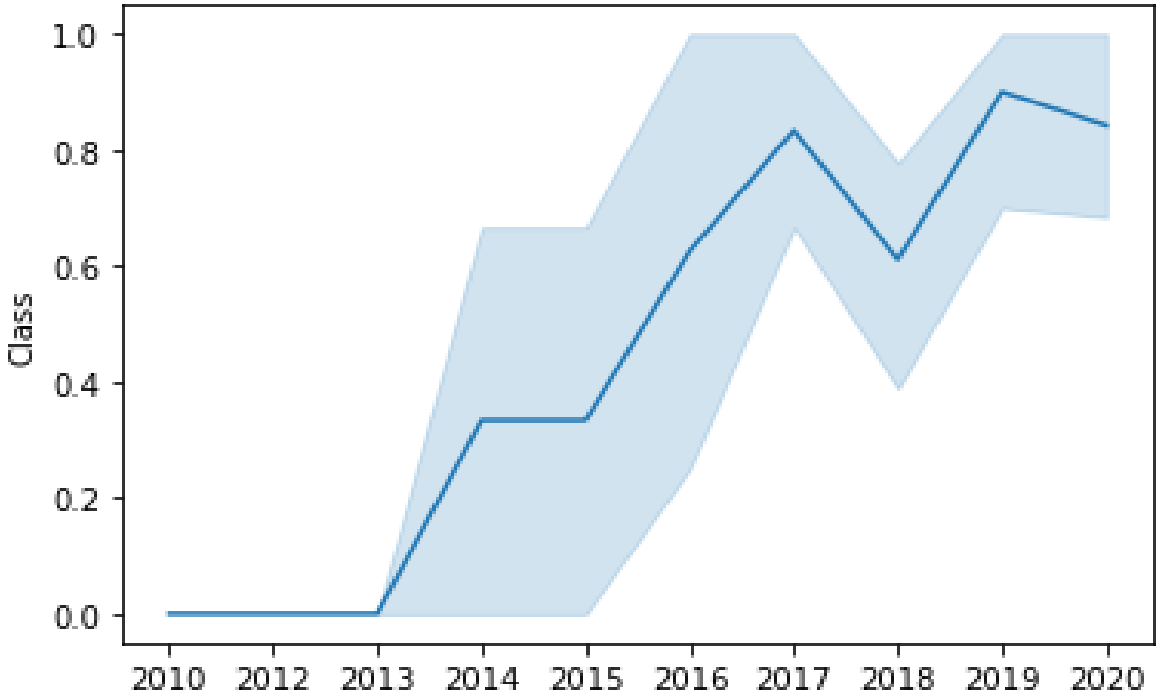
You should see that in the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

Payload vs. Orbit Type



With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.
However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.

Launch Success Yearly Trend



you can observe that the sucess rate since 2013 kept increasing till 2020

All Launch Site Names

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

Launch Site Names Begin with 'CCA'

```
%sql SELECT "Launch_Site" FROM "SPACEXTBL" WHERE "Launch_Site" LIKE 'CCA%' LIMIT 5;
```

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

Launch_Site
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40

Total Payload Mass

```
%sql SELECT SUM("PAYLOAD_MASS__KG_") FROM "SPACEXTBL" WHERE "Customer"== "NASA (CRS)" ;
```

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

SUM("PAYLOAD_MASS__KG_")
45596

Average Payload Mass by F9 v1.1

```
%sql SELECT AVG("PAYLOAD_MASS__KG_") FROM "SPACEXTBL" WHERE "Booster_Version" LIKE "F9 v1.1%" ;
```

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

AVG("PAYLOAD_MASS__KG_")
2534.6666666666665

First Successful Ground Landing Date

```
[ ]  %sql SELECT MIN("Date") FROM "SPACEXTBL" WHERE "Landing _Outcome" =="Success (ground pad)"  ;
```

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

MIN("Date")
01-05-2017

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql SELECT "Booster_Version" FROM "SPACEXTBL" WHERE "Landing _Outcome" =="Success (drone ship)" AND 4000<"PAYLOAD_MASS__KG_" <6000  ;
```

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

Booster_Version
F9 FT B1021.1
F9 FT B1022
F9 FT B1023.1
F9 FT B1026
F9 FT B1029.1
F9 FT B1021.2
F9 FT B1029.2
F9 FT B1036.1
F9 FT B1038.1
F9 B4 B1041.1
F9 FT B1031.2
F9 B4 B1042.1
F9 B4 B1045.1
F9 B5 B1046.1

Total Number of Successful and Failure Mission Outcomes

```
%sql  SELECT Count( "Mission_Outcome") FROM "SPACEXTBL" WHERE "Mission_Outcome"== "Success" ;
```

```
 * sqlite:///my_data1.db
Done.
Count( "Mission_Outcome")
98
```

```
%sql  SELECT Count( "Mission_Outcome") FROM "SPACEXTBL" WHERE "Mission_Outcome"== "Failure (in flight)" ;
```

```
 * sqlite:///my_data1.db
Done.
Count( "Mission_Outcome")
1
```

Boosters Carried Maximum Payload

```
%sql SELECT "Booster_Version" FROM "SPACEXTBL" WHERE "PAYLOAD_MASS__KG_"== (SELECT MAX("PAYLOAD_MASS__KG_") FROM "SPACEXTBL") ;
```

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

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 "Landing _Outcome" == "Failure (drone ship)" AND "Date" Like "%2015";
```

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

substr("Date",4,2)	Landing _Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

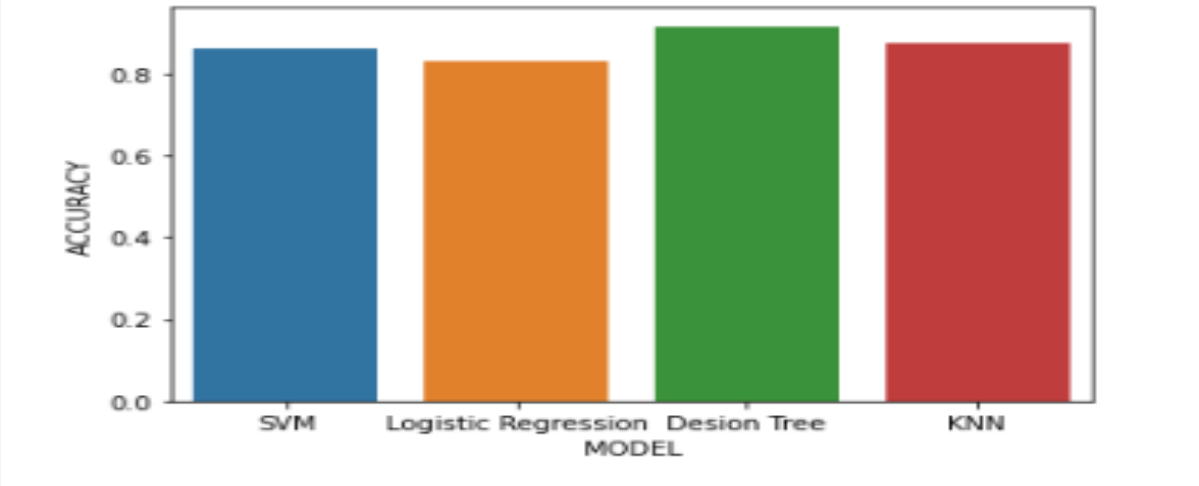
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

Date	Landing _Outcome
22-07-2018	Success
25-07-2018	Success
07-08-2018	Success
10-09-2018	Success
08-10-2018	Success
15-11-2018	Success
03-12-2018	Success
11-01-2019	Success
22-02-2019	Success
02-03-2019	Success
04-05-2019	Success
24-05-2019	Success
12-06-2019	Success
25-07-2019	Success
11-11-2019	Success
05-12-2019	Success
17-12-2019	Success
07-01-2020	Success
29-01-2020	Success
07-03-2020	Success
22-04-2020	Success
30-05-2020	Success
04-06-2020	Success
13-06-2020	Success
30-06-2020	Success
20-07-2020	Success
07-08-2020	Success
18-08-2020	Success
30-08-2020	Success
03-09-2020	Success
06-10-2020	Success
18-10-2020	Success
24-10-2020	Success
05-11-2020	Success
16-11-2020	Success
21-11-2020	Success
25-11-2020	Success
06-12-2020	Success

Section 5

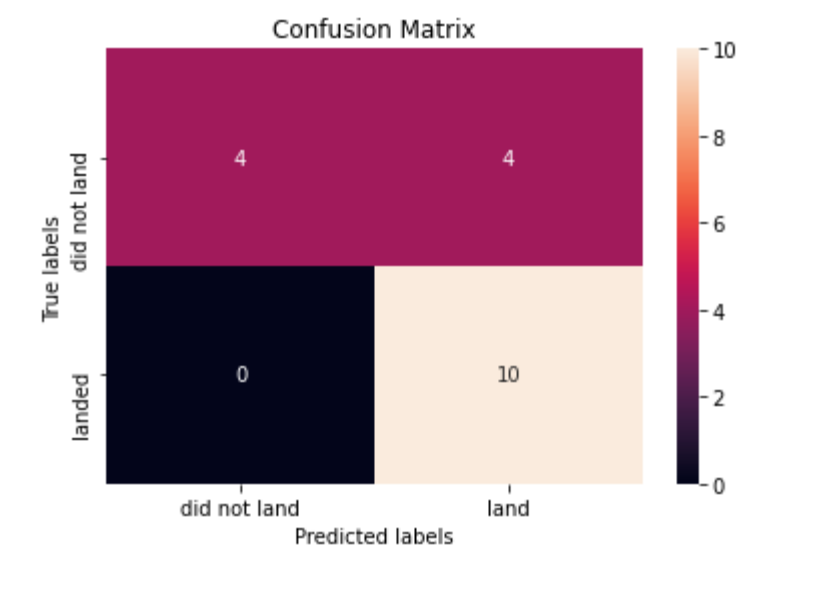
Predictive Analysis (Classification)

Classification Accuracy



Decision Tree Classifier is The Highest Accuracy =92%

Confusion Matrix



Appendix

https://github.com/ahmedsalama95/DataScienceProjects

Thank you!