- BY ABIR CHAKRABORTY

IBM DATA SCIENCE CAPSTONE PROJECT

Space X Falcon 9 Landing Analysis

Executive Summary

- Data Collection

- Data Wrangling

- Exploratory Data Analysis

- Predictive Analysis (Classification Model)

Introduction

SpaceX launches Falcon 9 rockets costs $62m. This is considerably cheaper than other providers (which usually cost upwards of $165m), and much of the savings are because SpaceX can land, and then re-use the first stage of the rocket.

- If we can make predictions on whether the first stage will land, we can determine the cost of a launch, and use this information to assess whether or not an alternate company should bid and SpaceX for a rocket launch.

This project will ultimately predict if the Space X Falcon 9 first stage will land successfully through a classification model

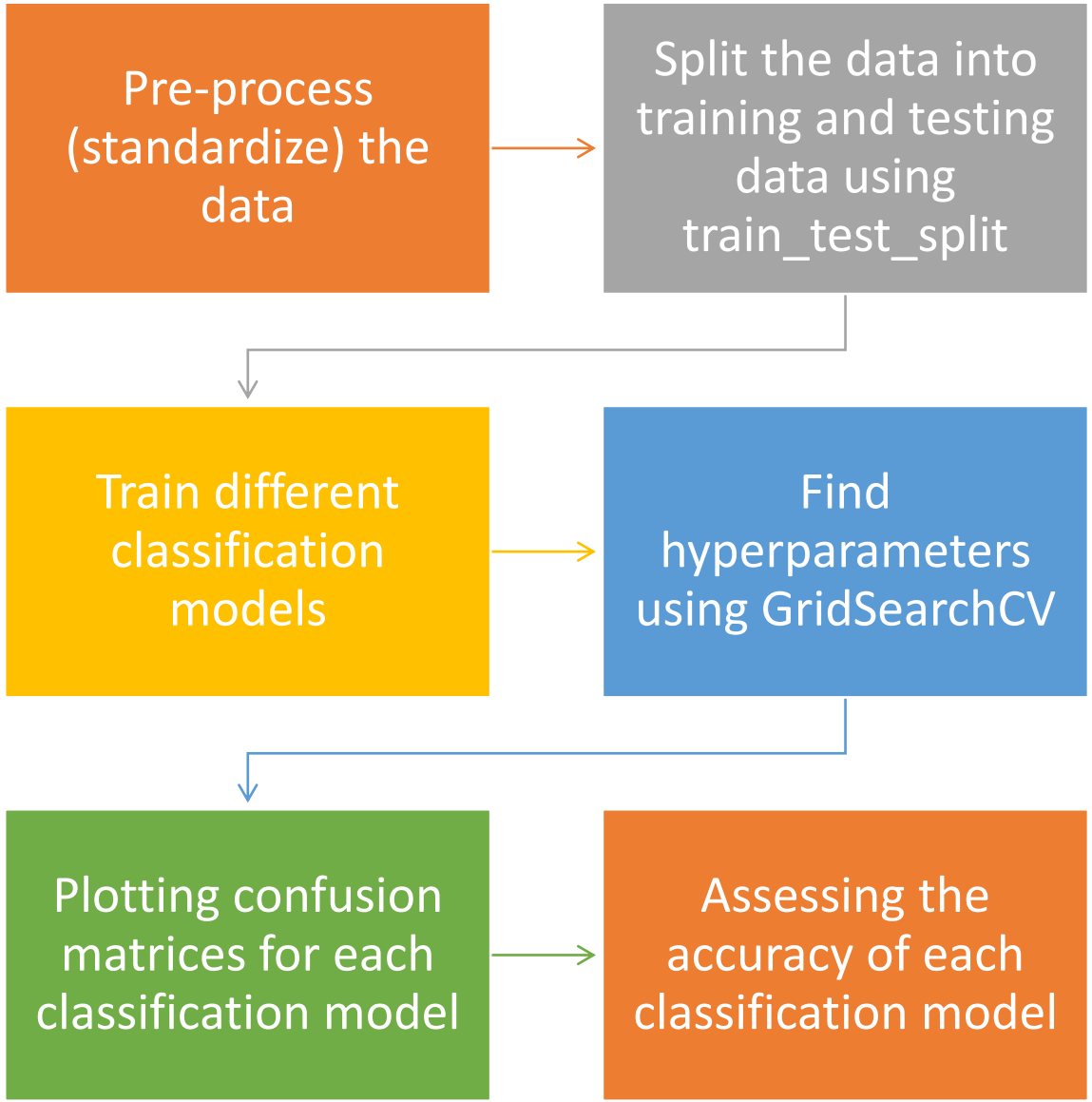
Data collection and data wrangling methodology

- Making GET requests to the SpaceX REST API
- Web Scraping
- Using the .fillna() method to remove NaN values
- Using the .value_counts() method to determine the following:
- Number of launches on each site
- Number and occurrence of each orbit
- Number and occurrence of mission outcome per orbit type

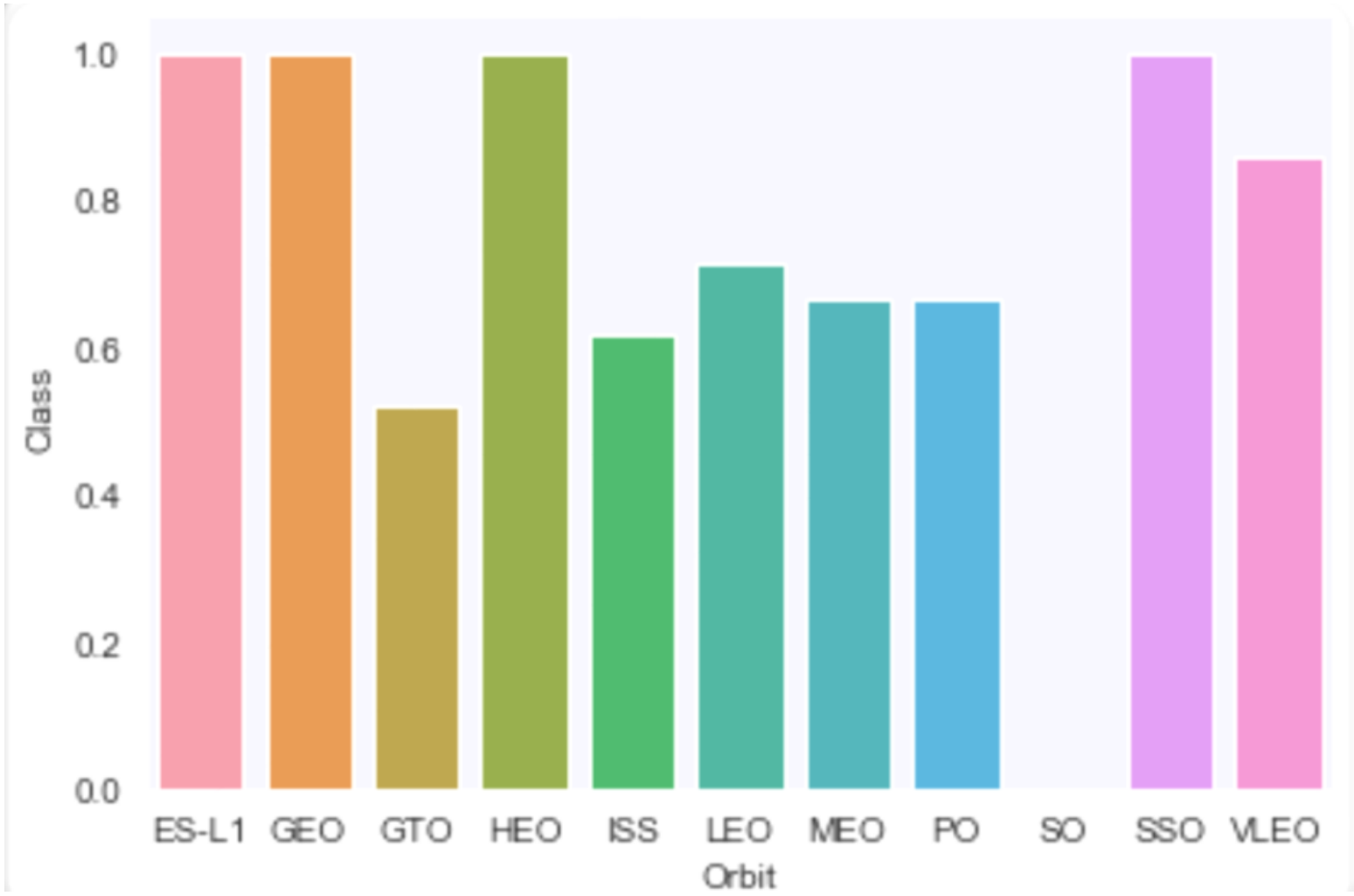
EDA and interactive visual analytics methodology

- Using SQL queries to manipulate and evaluate the SpaceX dataset
- Using Pandas and Matplotlib to visualize relationships between variables, and determine patterns
- Geospatial analytics using Folium
- Creating an interactive dashboard using Plotly Dash

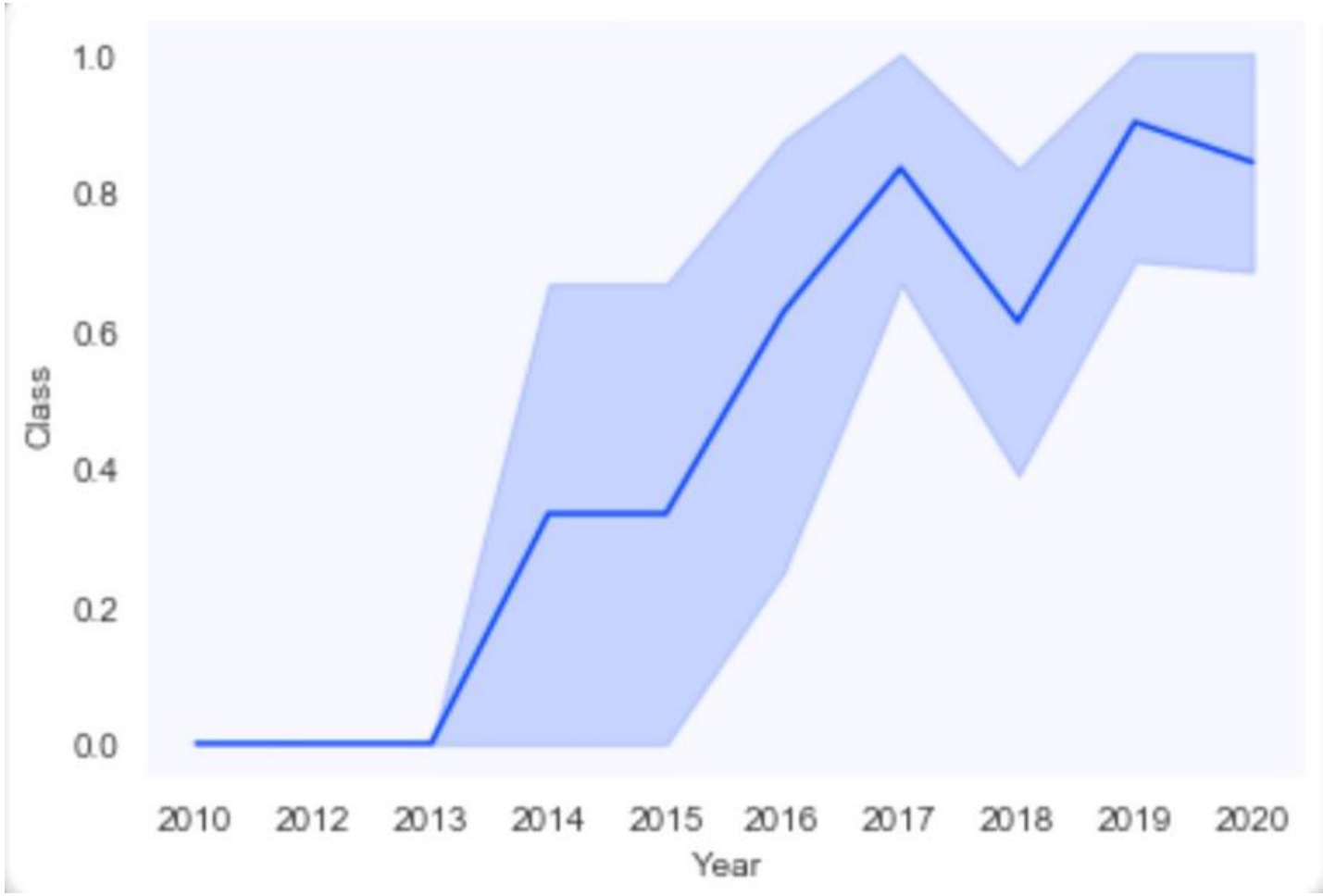
Predictive analysis methodology



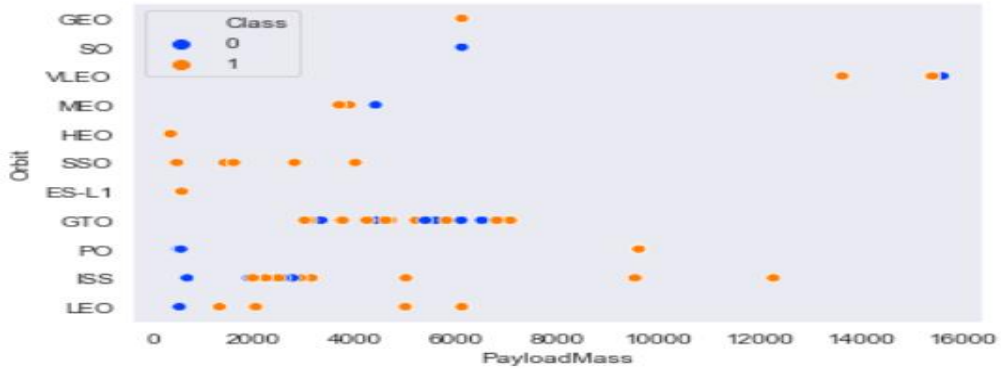
SUCCESS RATE VS. ORBIT TYPE

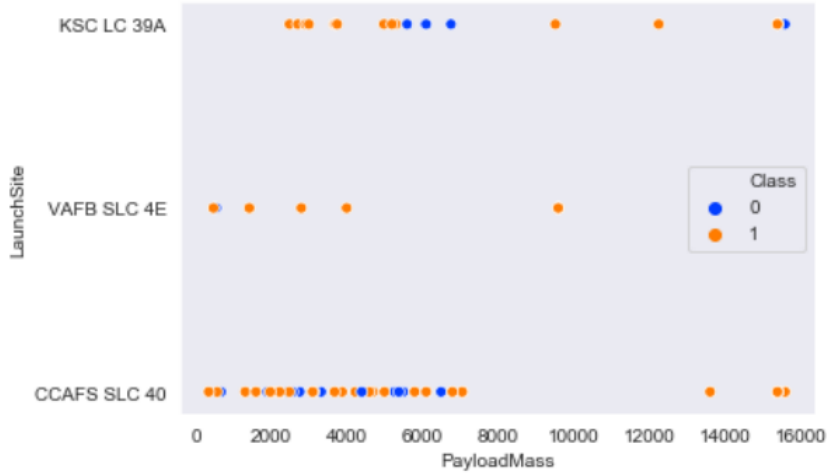


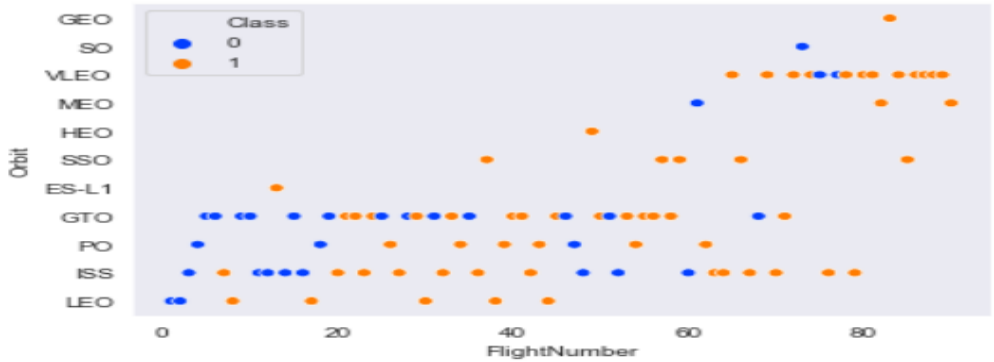
Launch Success Yearly Trend

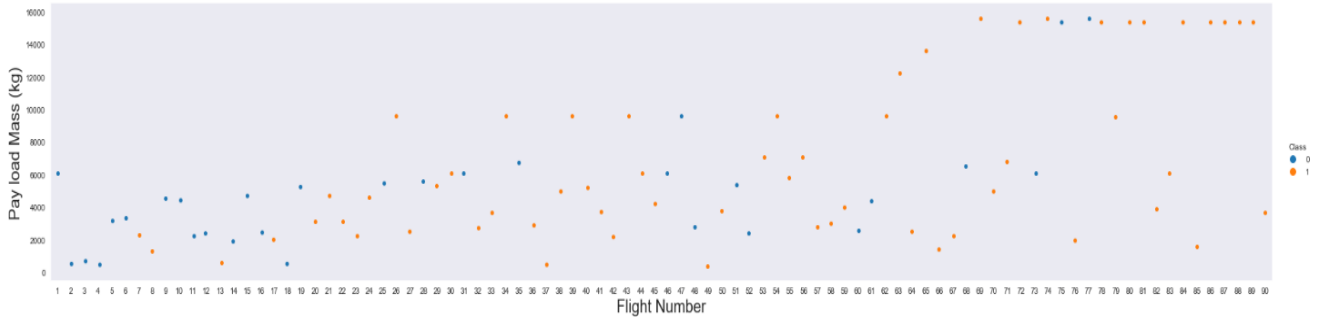


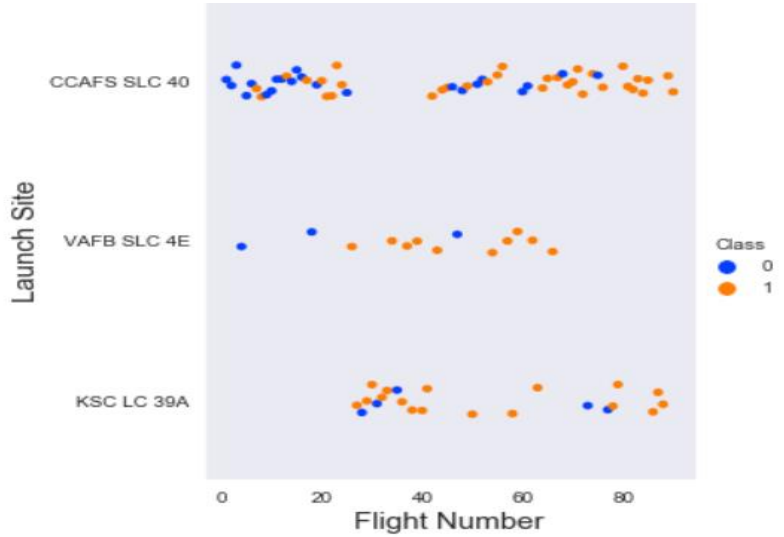
EDA


Class
0
1
GEO
SO
VLEO
MEO
HEO
SSO
ES-L1
GTO
PO
ISS
LEO
Orbit
0
2000
4000
6000
8000
10000
12000
14000
16000
PayloadMass


KSC LC 39A
VAFB SLC 4E
CCAFS SLC 40
LaunchSite
Class
0
1
0
2000
4000
6000
8000
10000
12000
14000
16000
PayloadMass


Class
0
1
GEO
SO
VLEO
MEO
HEO
SSO
ES-L1
GTO
PO
ISS
LEO
Orbit
0
20
40
60
80
FlightNumber


Pay load Mass (kg)
Flight Number


CCAFS SLC 40
VAFB SLC 4E
KSC LC 39A
Launch Site
Class
0
1
0
20
40
60
80
Flight Number

EDA with SQL 1-3

```
%sql SELECT UNIQUE(LAUNCH_SITE) FROM SPACEXTBL;
```

```
 * ibm_db_sa://kfm42587:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

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

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) AS TOTAL_PAYLOAD_MASS FROM SPACEXTBL \
    WHERE CUSTOMER = 'NASA (CRS)';
```

```
 * ibm_db_sa://kfm42587:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

total_payload_mass
45596

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

```
 * ibm_db_sa://kfm42587:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

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

EDA with SQL 4-7

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) AS AVERAGE_PAYLOAD_MASS FROM SPACEXTBL \
    WHERE BOOSTER_VERSION = 'F9 v1.1';
```

```
 * ibm_db_sa://kfm42587:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

average_payload_mass
2928

```
%sql SELECT MIN(DATE) AS FIRST_SUCCESSFUL_GROUND_LANDING FROM SPACEXTBL \
    WHERE LANDING__OUTCOME = 'Success (ground pad)';
```

```
 * ibm_db_sa://kfm42587:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

first_successful_ground_landing
2015-12-22

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

```
 * ibm_db_sa://kfm42587:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

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

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

```
 * ibm_db_sa://kfm42587:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

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

EDA with SQL 8-10

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

```
 * ibm_db_sa://kfm42587:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
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

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

```
 * ibm_db_sa://kfm42587:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

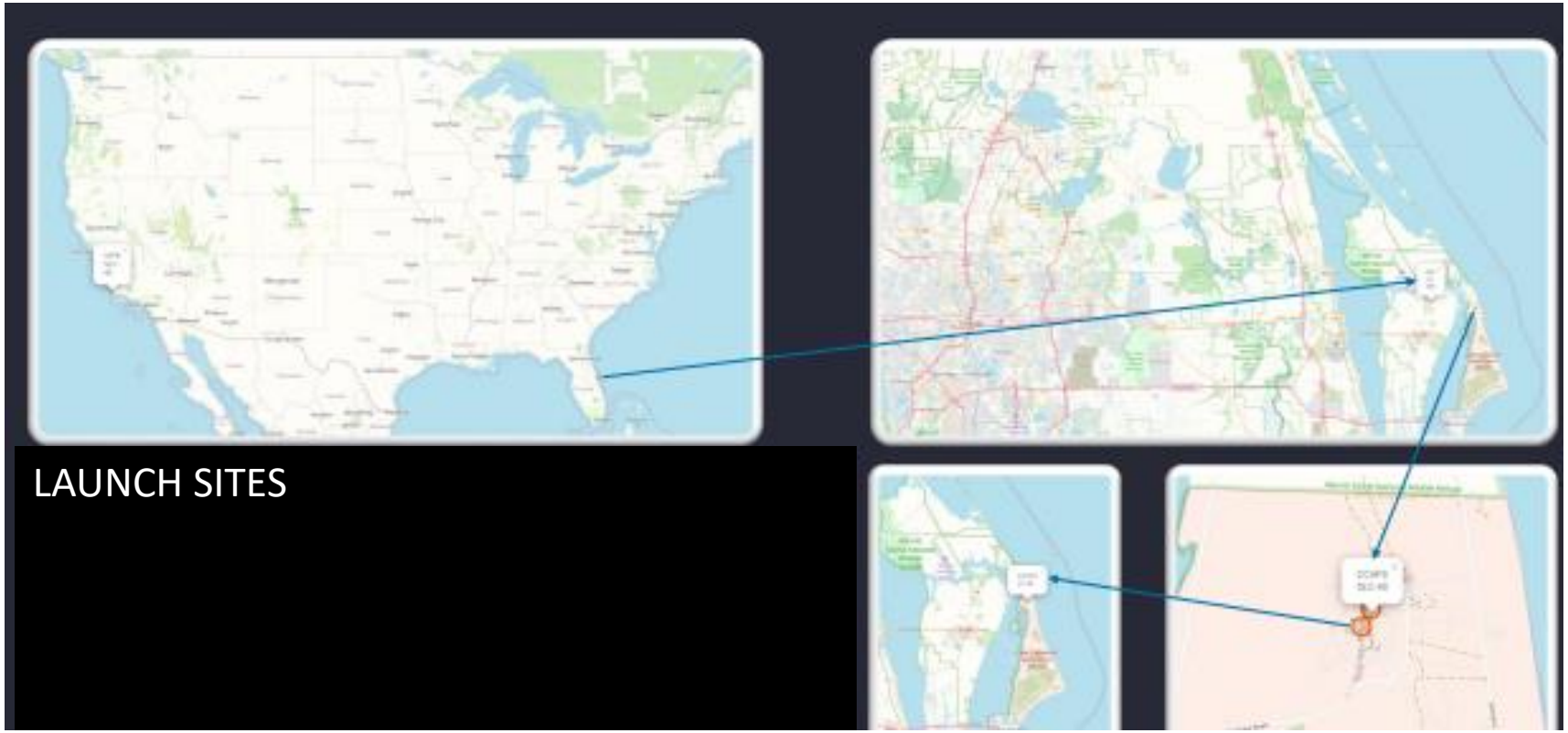
booster_version	launch_site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

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

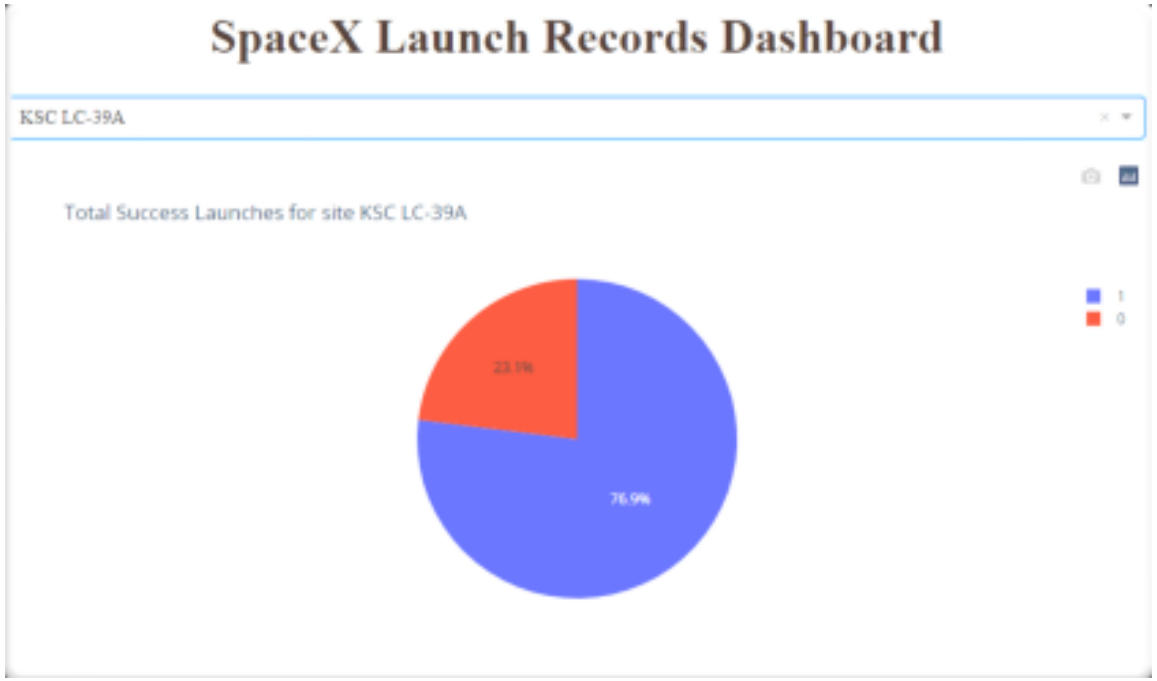
```
 * ibm_db_sa://kfm42587:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

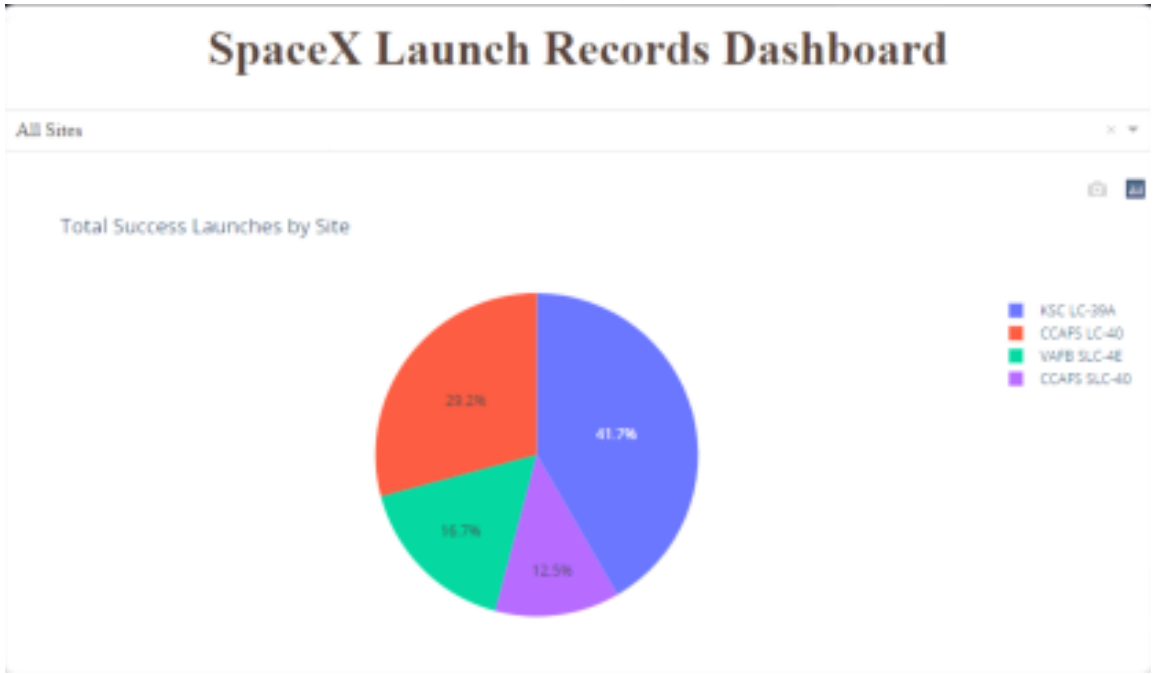
booster_version
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
F9 B5 B1051.3
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

interactive map with Folium results



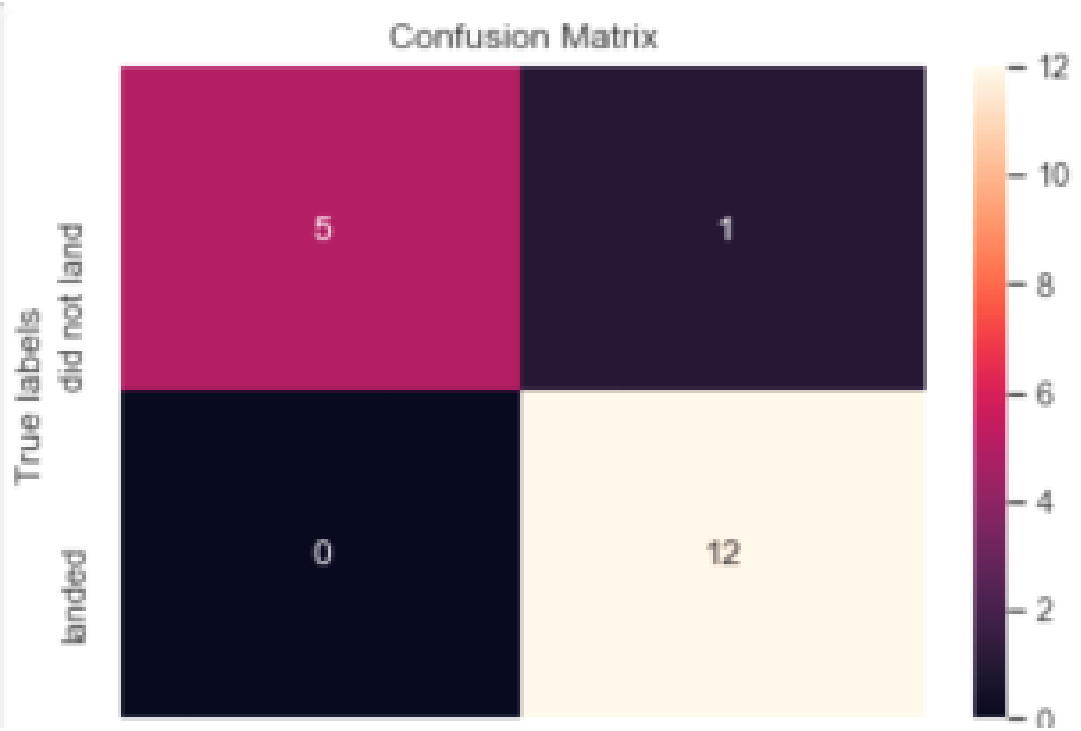
Plotly Dash dashboard results

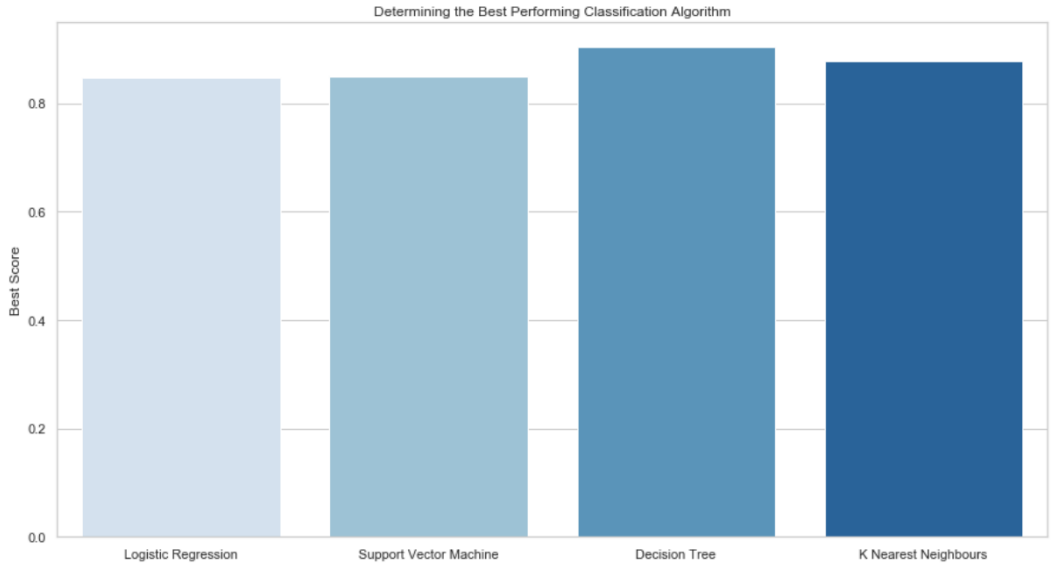




predictive analysis (classification) results

	Algorithm	Accuracy Score	Best Score
0	Logistic Regression	0.833333	0.846429
1	Support Vector Machine	0.833333	0.848214
2	Decision Tree	0.944444	0.903571
3	K Nearest Neighbours	0.888889	0.876786





Conclusions

As the number of flights increases, the rate of success at a launch site increases, with most early flights being unsuccessful. I.e. with more experience, the success rate increases.

The success for massive payloads (over 4000kg) is lower than that for low payloads.

The best performing classification model is the Decision Tree model, with an accuracy of 94.44%