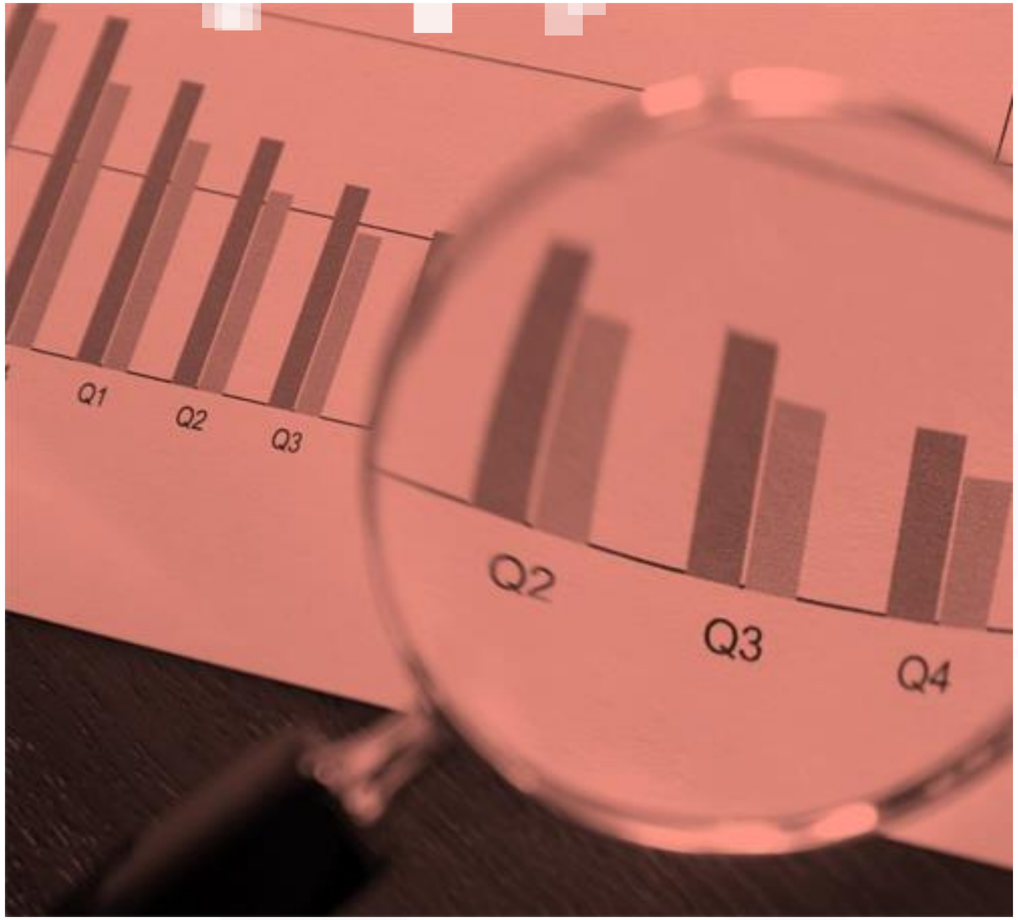


DATA ANALYSIS OF SPACEX

Ritu Vyaghrambare

04/06/2022

OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
  - Visualization – Charts
  - Dashboard
- Discussion
  - Findings & Implications
- Conclusion
- Appendix

EXECUTIVE SUMMARY

- Data collection from the SpaceX API
- Exploratory data analysis to determine the attributes responsible for landing of Falcon9 second stage
  - Landing prediction of first stage to estimate the cost of launch
  - Estimation of launch success rate with respect to payload mass, orbit type, and launch site location.
- Interaction dashboard to compare different aspects
- Predictive analysis using a machine learning model and output is presented in the form of confusion matrix

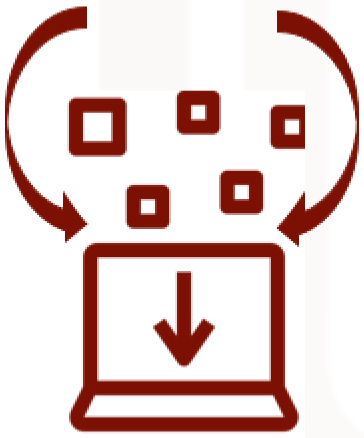
INTRODUCTION

- In the commercial space age, launching a rocket with a minimum cost is crucial. SpaceX does it with an ease. So, by analyzing the data from SpaceX API, the details can be studied.
- The exploratory data analysis is done to find the relation between different features and their affect on cost.
- Different graphs shows the relation of payload mass, orbit type, and the location of launching can help to build a machine learning model.
- An interactive dashboard is given to study the analysis.
- The findings are presented in terms of confusion matrix.

METHODOLOGY

- The data for analysis is taken from the SpaceX API using the web scrapping. Data wrangling is done to clean the data.
- Exploratory data analysis is done with the help of SQL and Python to study the relationship between the different features.
- The interactive dashboard is created to compare the features and predict if the rocket will land successfully.
- The train set and test set is taken from the data to run the different methods. The output of it is used to create the machine learning model for easy prediction.

DATA COLLECTION

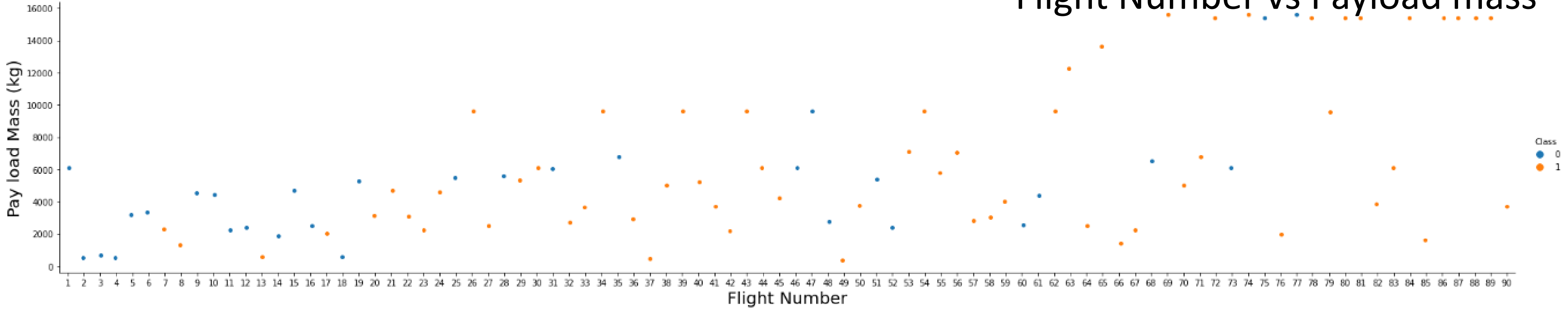


- The rocket launch data was requested from the SpaceX API. The Json content was decoded and converted to pandas data frame.
- Null values in the 'PayloadMass' column were identified and by data wrangling. The missing values were replaced by the mean of the data.

EXPLORATORY DATA ANALYSIS

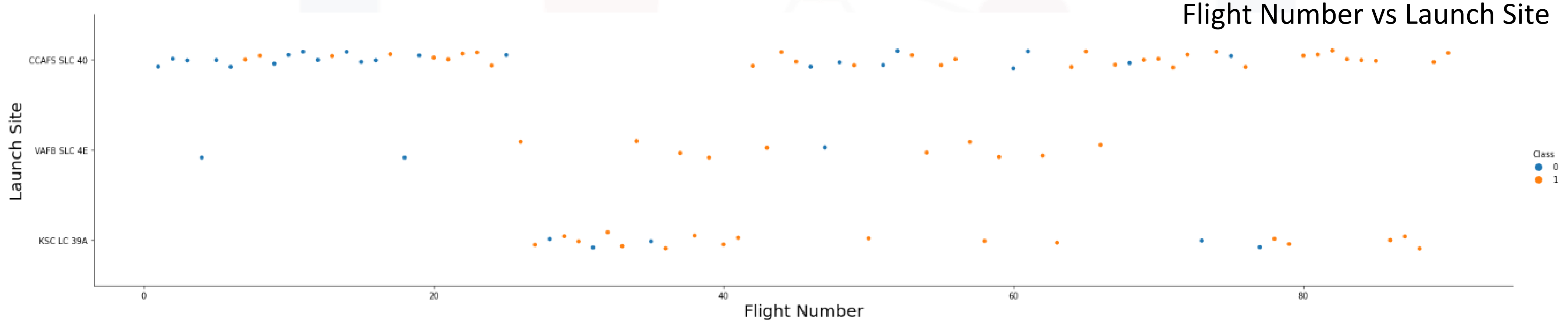
- The connection to the data base was made to query the data for important information.
- The different features were compared as shown:

Flight Number vs Payload mass

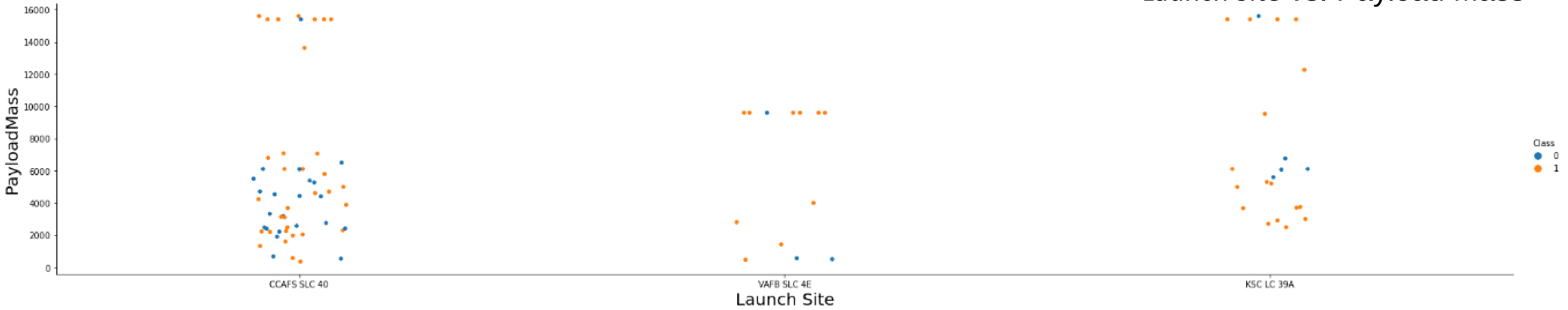


EXPLORATORY DATA ANALYSIS

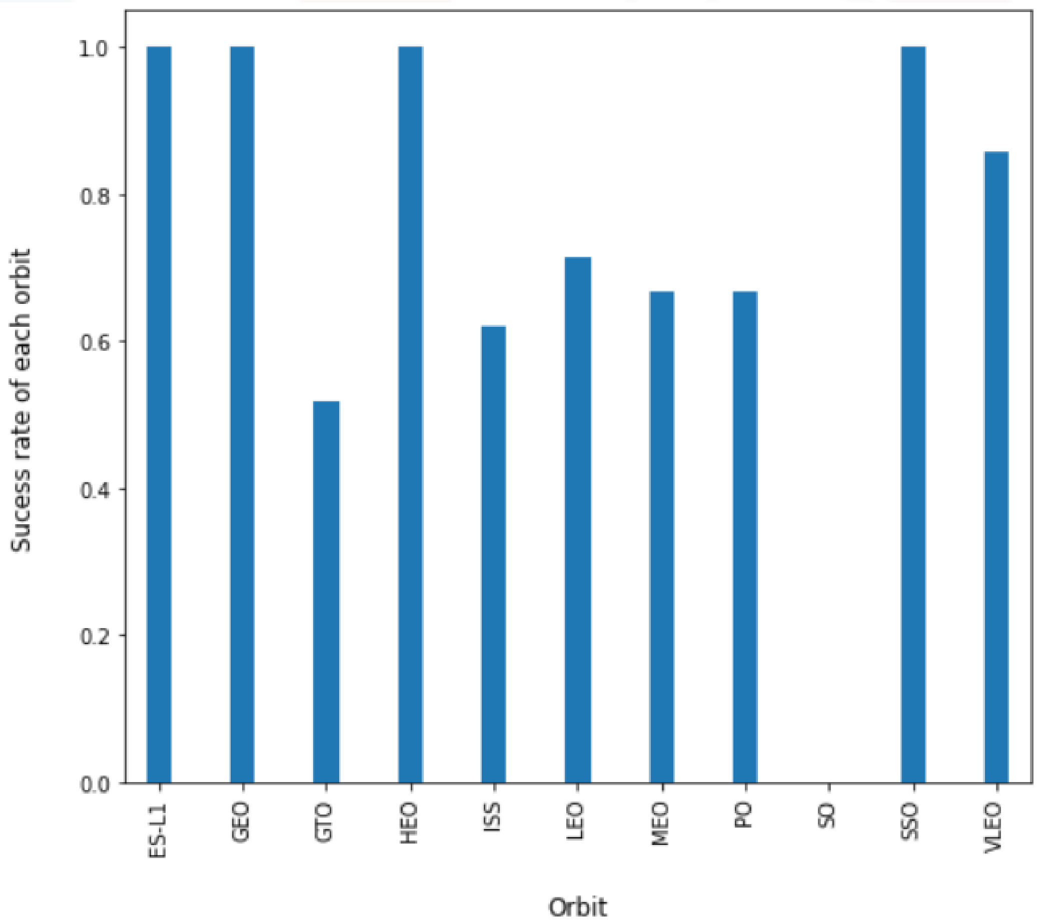
Flight Number vs Launch Site



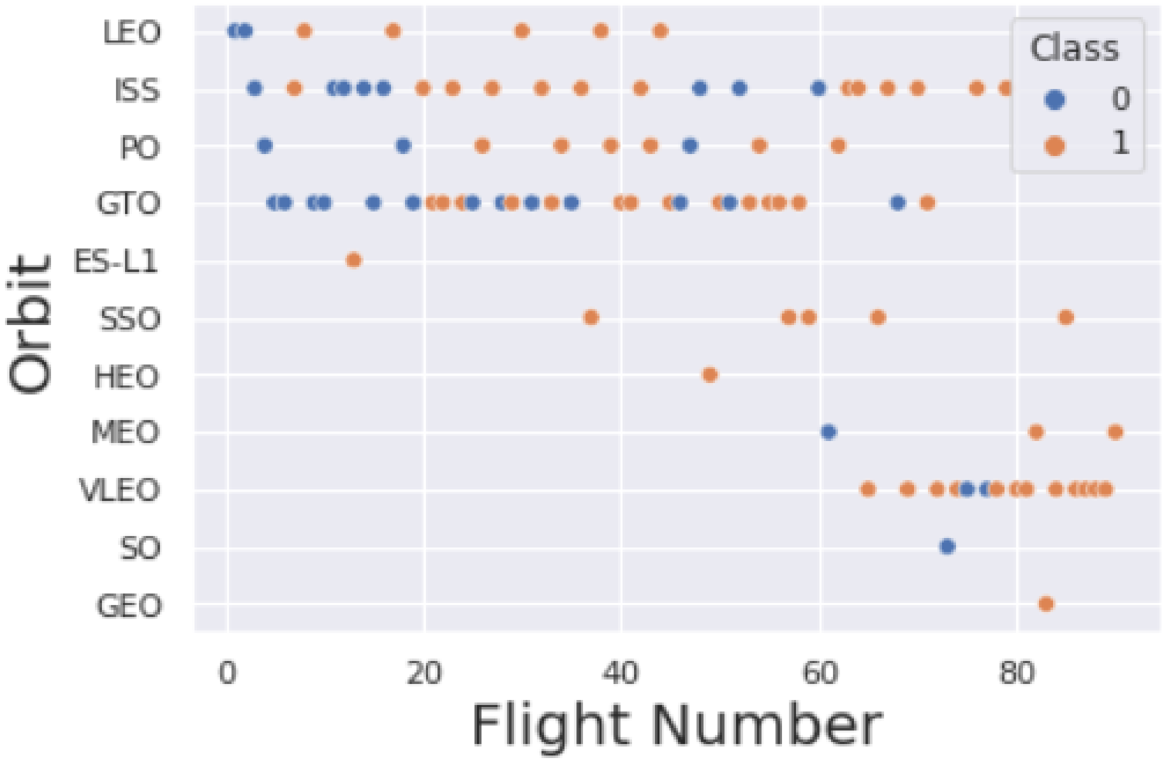
Launch Site vs. Payload Mass



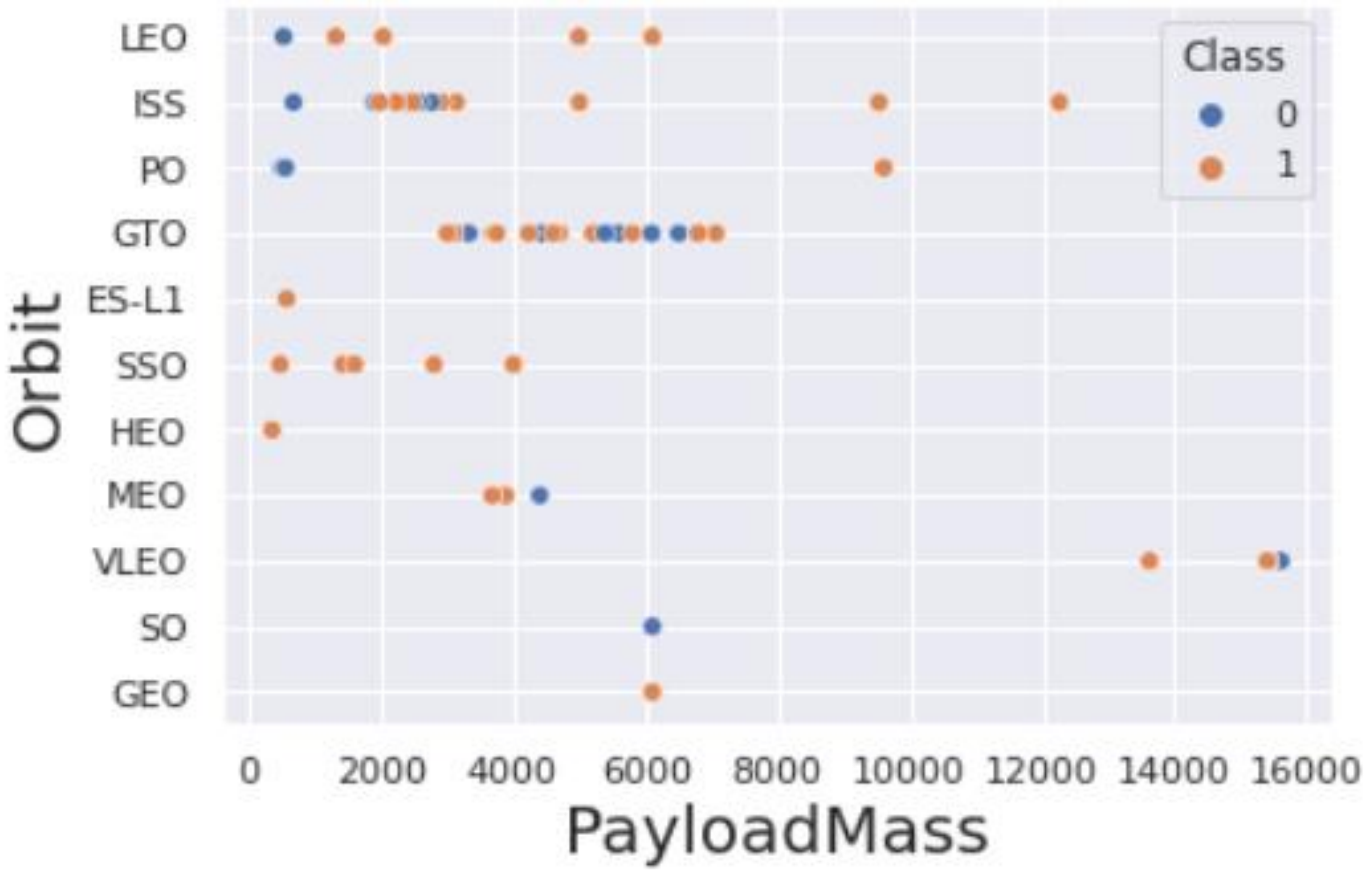
SUCCESS RATE OF EACH ORBIT



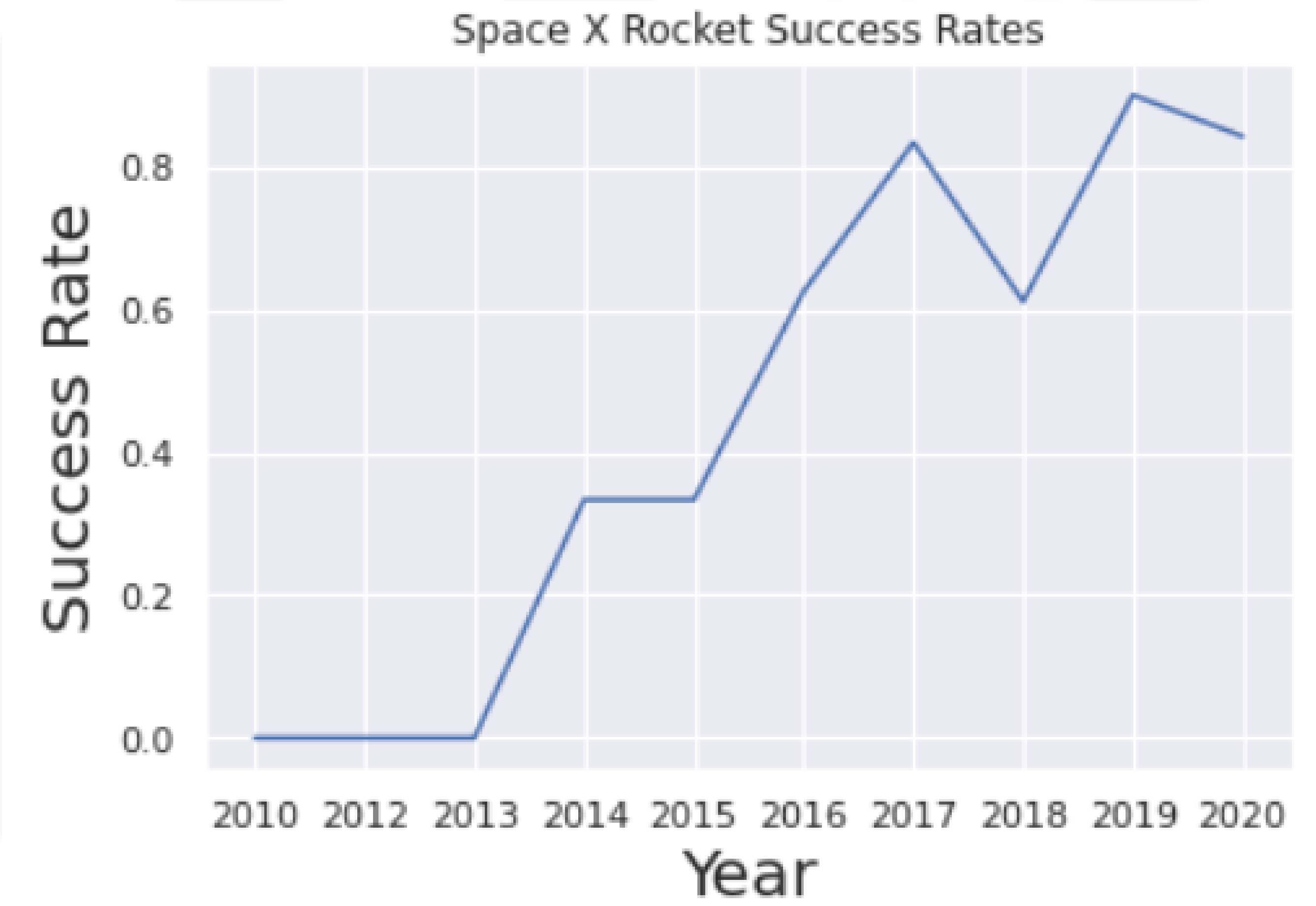
RELATIONSHIP BETWEEN FLIGHTNUMBER AND ORBIT TYPE



RELATIONSHIP BETWEEN PAYLOAD AND ORBIT TYPE



SUCCESS RATE



EDA WITH SQL

Unique Launch sites with the space mission:

```
%sql SELECT DISTINCT LAUNCH_SITE FROM SPACEXDATASET
```

```
 * ibm_db_sa://ydh08468:***@125f9f61-9715-46f9-9399-c81
Done.
```

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

EDA WITH SQL

Records where launch site begins with CCA string:

```
%sql SELECT LAUNCH_SITE FROM SPACEXDATASET WHERE (LAUNCH_SITE) LIKE 'CCA%' LIMIT 5;
```

```
 * ibm_db_sa://ydh08468:***@125f9f61-9715-46f9-9399-c8177b21803b.c1ogj3sd0tgtu0lqde00
Done.
```

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

EDA WITH SQL

Total payload mass carried by boosters launched by NASA (CRS):

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

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXDATASET WHERE CUSTOMER='NASA (CRS)';
```

```
 * ibm_db_sa://ydh08468:***@125f9f61-9715-46f9-9399-c8177b21803b.c1ogj3sd0tgtu0lqde0
Done.
```

1
45596

EDA WITH SQL

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

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXDATASET WHERE BOOSTER_VERSION='F9 v1.1';
```

```
 * ibm_db_sa://ydh08468:***@125f9f61-9715-46f9-9399-c8177b21803b.c1ogj3sd0tgtu0lqde00.database
Done.
```

1
2928

EDA WITH SQL

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

Hint:Use min function

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

```
 * ibm_db_sa://ydh08468:***@125f9f61-9715-46f9-9399-c8177b21803b.c1ogj3sd0tgtu0lqde00.
Done.
```

1
2015-12-22

EDA WITH SQL

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_ > 4
```

```
 * ibm_db_sa://ydh08468:***@125f9f61-9715-46f9-9399-c8177b21803b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:3
Done.
```

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

EDA WITH SQL

List the total number of successful and failure mission outcomes

```
%sql SELECT sum(case when MISSION_OUTCOME LIKE '%Success%' then 1 else 0 end) AS "Successful Mission", sum(case when MISSION_OUTCOME LIKE '%Failure%'
```

```
 * ibm_db_sa://ydh08468:***@125f9f61-9715-46f9-9399-c8177b21803b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30426/bludb
Done.
```

Successful Mission	Failure Mission
100	1

EDA WITH SQL

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

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

```
 * ibm_db_sa://ydh08468:***@125f9f61-9715-46f9-9399-c8177b21803b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30426/bludb
Done.
```

booster_version
None

EDA WITH SQL

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

```
%sql SELECT {fn MONTHNAME(DATE)} as "Month", BOOSTER_VERSION, LAUNCH_SITE FROM SPACEX WHERE year(DATE) = '2015' AND LANDING_OUTCOME = 'Failure (drone
```

```
 * ibm_db_sa://ydh08468:***@125f9f61-9715-46f9-9399-c8177b21803b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30426/bludb
Done.
```

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

EDA WITH SQL

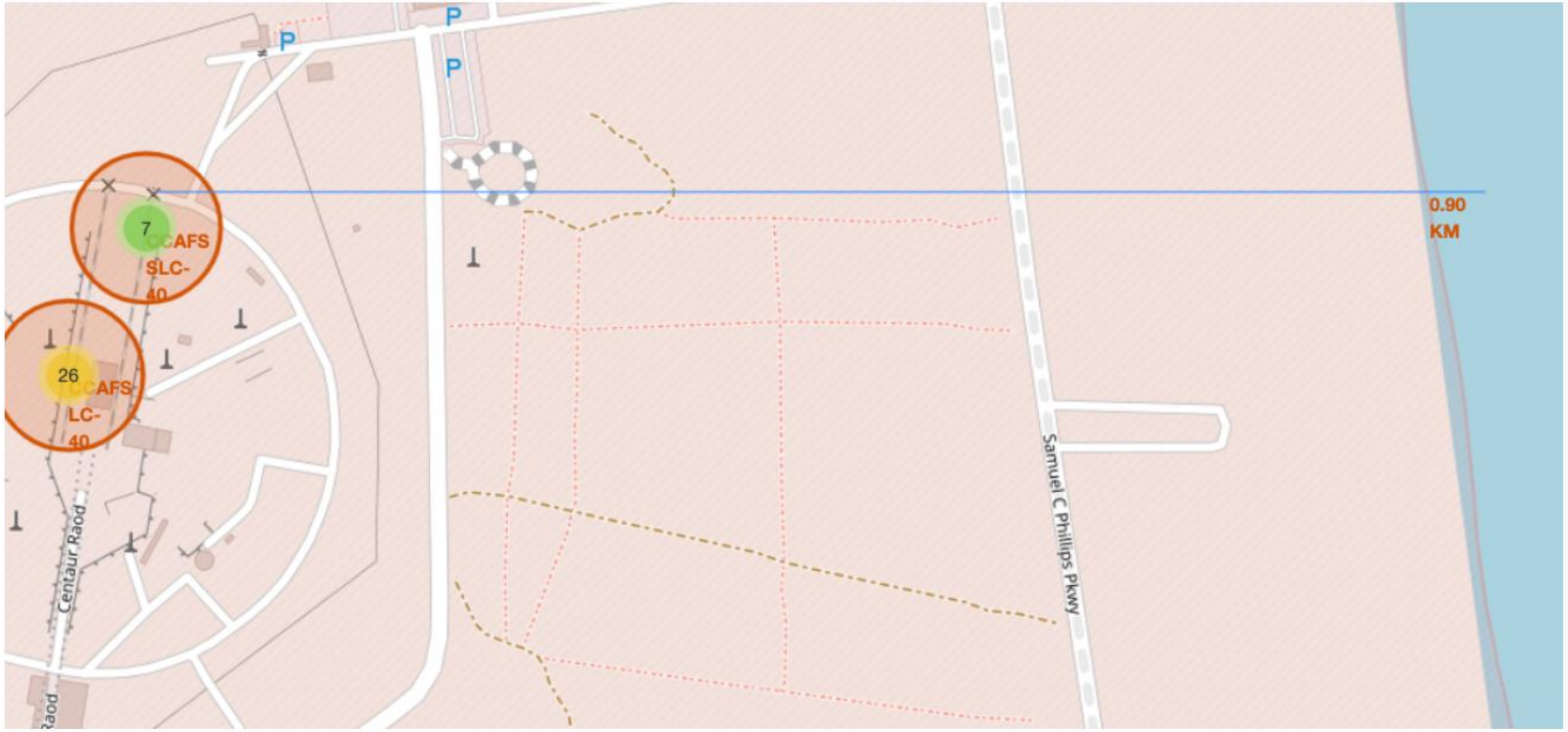
Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

```
%sql SELECT COUNT(LANDING_OUTCOME) FROM SPACEX WHERE LANDING_OUTCOME LIKE '%Success%' AND DATE > '2010-06-04' AND DATE < '2017-03-20'
```

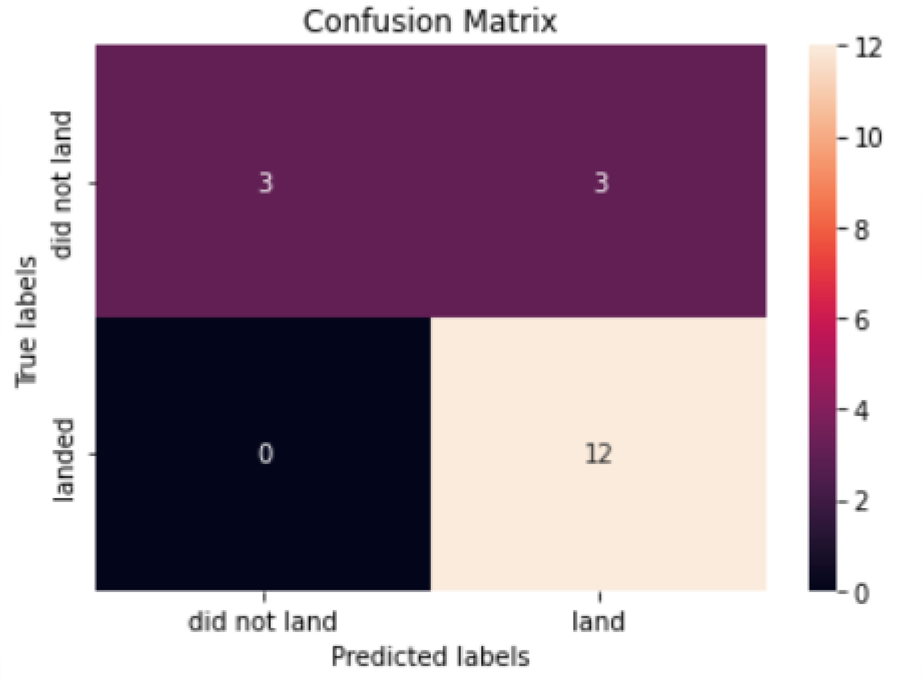
```
 * ibm_db_sa://ydh08468:***@125f9f61-9715-46f9-9399-c8177b21803b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30426/bludb
Done.
```

1
10

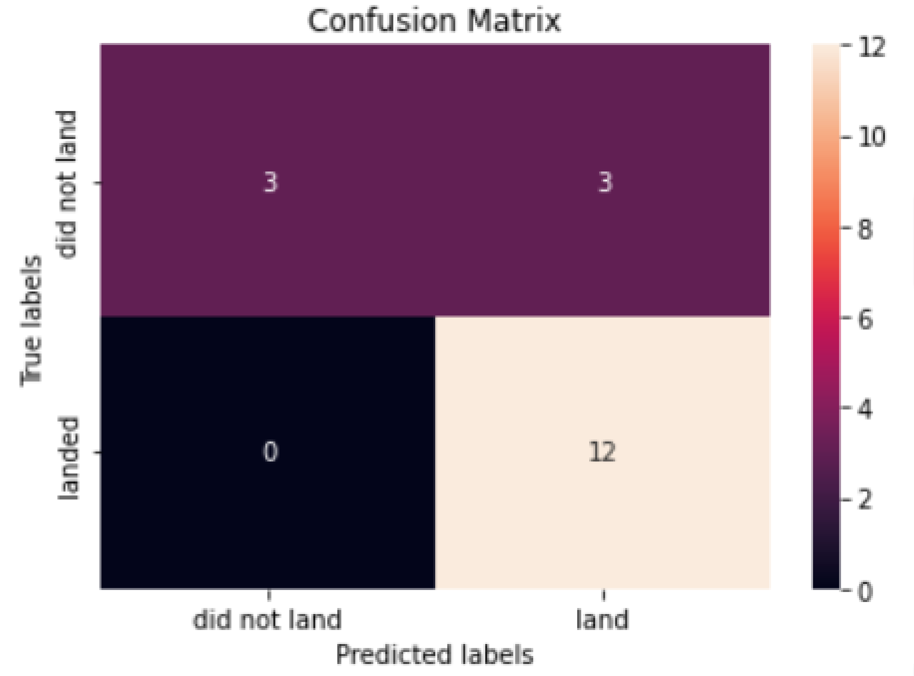
LAUNCHING SITES



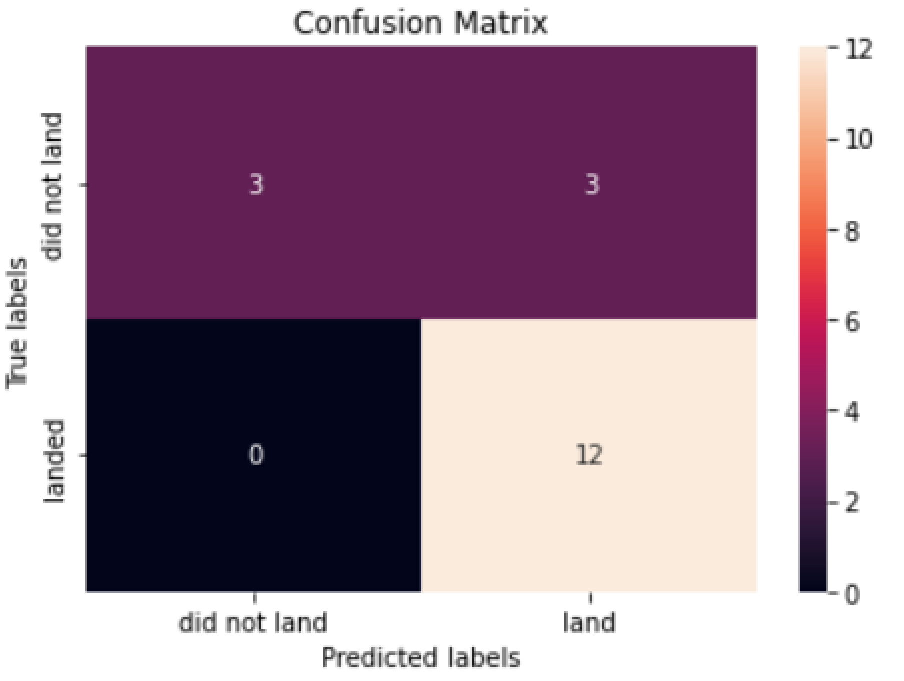
PREDICTIVE ANALYSIS (LOGISTIC REGRESSION)



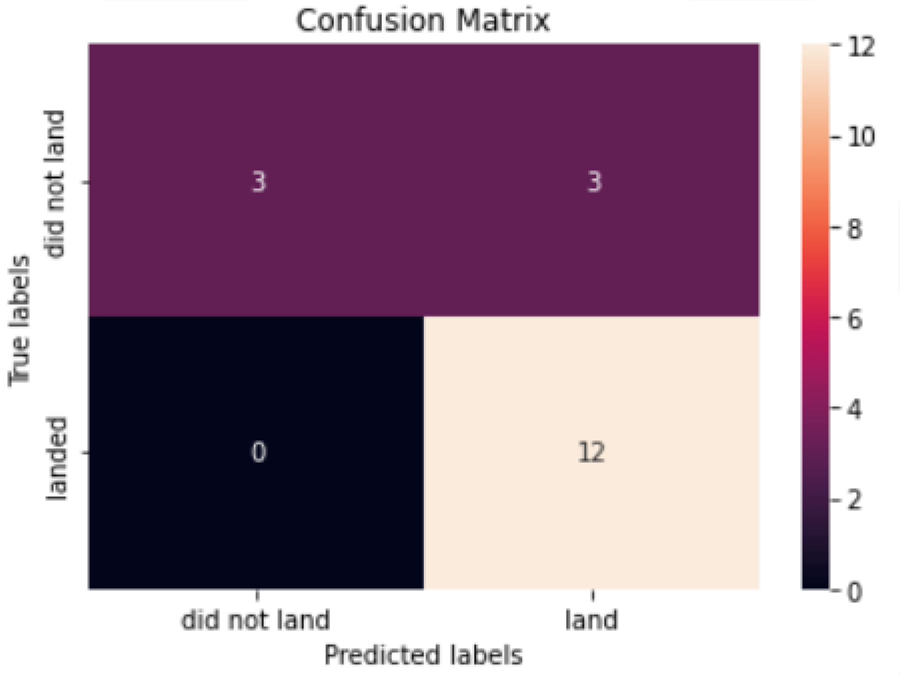
PREDICTIVE ANALYSIS (SUPPORT VECTOR MACHINE OBJECT)



PREDICTIVE ANALYSIS (DECISION TREE CLASSIFIER OBJECT)



PREDICTIVE ANALYSIS (K NEAREST NEIGHBORS OBJECT)



RESULTS

	Model	Accuracy	Prediction score
0	LogisticRegression()	0.8464285714285713	0.8333333333333334
1	SVC()	0.8482142857142856	0.8333333333333334
2	DecisionTreeClassifier()	0.8910714285714286	0.7222222222222222
3	KNeighborsClassifier()	0.8482142857142858	0.8333333333333334

All the opted models give us the considerable accuracy.

CONCLUSION

- The retrieval of the first stage successfully can make the mission cost efficient.
- For a successful landing, the appropriate launch site, payload mass, and orbit type plays a significant role.
- Success rate is directly proportional to payload mass.
- CCAFS SLC 40 launching site has higher success rate for the payload masses below 6000 kg and above 14000 kg.
- Orbit type VLEO and ISS can be accessed for the launches.
- ISS orbit type is more suitable for the payload masses between 2000 kg and 4000 kg.
- Decision tree classifier is the best fitted machine learning model for the most accurate predictions.