CAPSTONE PROJECT

Setia Mukti Azizah - 22 February 2023

URL GITHUB

https://github.com/setiamuktiazizah/capstoneProject

EXECUTIVE SUMMARY

01 INTRODUCTION

Nature of Analysis, States the problem, and the Question about Analysis

02 METODHOLOGY

The data sources that were used in the analysis and outlines the plan for the collected data

03 RESULT

Detail of the data collection, how it was organized, and how it was analyzed

04 CONCLUSION

Vous pouvez décrire ici le sujet de la section

INTRODUCTION

STATE THE PROBLEM

The commercial space age making space travel affordable for everyone namely **SpaceX**. SpaceX **advertises Falcon9** rocket launches on website with a cost of 62 million dollars; other providers cost upwards of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage.
Therefore, **if we can determine if the first stage will land, we can determine the cost of a launch.**

WHAT THE AIM FOR THIS ANALYSIS?

As a data scientist working for a new rocket company. Space Y that would like to compete with SpaceX founded by Billionaire industrialist Allon Musk. My job is **determine the price of each launch**. I will do this by gathering information about Space X and creating dashboards. I will also determine if SpaceX will reuse the first stage. Instead of using rocket science to determine if the first stage will land successfully, I will train a machine learning model and use public information to predict if SpaceX will reuse the first stage.

METHODOLOGY

DATA COLLECTION & DATA WRANGLING

SPACEX API

- Request to the SpaceX API

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	Reus
4	1	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	
5	2	2012-05-22	Falcon 9	525.0	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	
6	3	2013-03-01	Falcon 9	677.0	ISS	CCSFS SLC 40	None None	1	False	False	False	None	1.0	
7	4	2013-09-29	Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean	1	False	False	False	None	1.0	
8	5	2013-12-03	Falcon 9	3170.0	GTO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	

- Clean the requested data

```
data_falcon9.isna().sum()
```

```
FlightNumber      0
Date              0
BoosterVersion    0
PayloadMass       0
Orbit             0
LaunchSite        0
Outcome           0
Flights           0
GridFins          0
Reused            0
Legs              0
LandingPad       24
Block             0
ReusedCount       0
Serial            0
Longitude         0
Latitude          0
dtype: int64
```

WEBSCRAPPING

- Extract a Falcon9 launch records HTML table from Wikipedia
- Parse the table and convert it into a Pandas data frame

	Flight No.	Launch site	Payload	Payload mass	Orbit	Customer	Launch outcome	Version Booster	Booster landing	Date	Time
0	1	CCAFS	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success\n	F9 v1.0B0003.1	Failure	4 June 2010	18:45
1	2	CCAFS	Dragon	0	LEO	NASA (COTS)\nNRO	Success	F9 v1.0B0004.1	Failure	8 December 2010	15:43
2	3	CCAFS	Dragon	525 kg	LEO	NASA (COTS)	Success	F9 v1.0B0005.1	No attempt\n	22 May 2012	07:44
3	4	CCAFS	SpaceX CRS-1	4,700 kg	LEO	NASA (CRS)	Success\n	F9 v1.0B0006.1	No attempt	8 October 2012	00:35
4	5	CCAFS	SpaceX CRS-2	4,877 kg	LEO	NASA (CRS)	Success\n	F9 v1.0B0007.1	No attempt\n	1 March 2013	15:10

EXPLORATORY DATA ANALYSIS

Create a LandingClass to a new Column in df

DECLARE BAD OUTCOMES

```
bad_outcomes=set(landing_outcomes.keys()[[1,3,5,6,7]])
bad_outcomes
```

```
{'False ASDS', 'False Ocean', 'False RTLS', 'None ASDS', 'None None'}
```

```
for i,outcome in enumerate(landing_outcomes.keys()):
    print(i,outcome)
```

```
0 True ASDS
1 None None
2 True RTLS
3 False ASDS
4 True Ocean
5 False Ocean
6 None ASDS
7 False RTLS
```

MAKE FUNCTION LANDING CLASS

```
# landing_class = 0 if bad_outcome
# landing_class = 1 otherwise

landing_class =[]
for outcome in df['Outcome'] :
    if outcome in bad_outcomes :
        landing_class.append(0)
    else :
        landing_class.append(1)
```

terVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude	Class
Falcon 9	6104.959412	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0003	-80.577366	28.561857	0
Falcon 9	525.000000	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0005	-80.577366	28.561857	0
Falcon 9	677.000000	ISS	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0007	-80.577366	28.561857	0
Falcon 9	500.000000	PO	VAFB SLC 4E	False Ocean	1	False	False	False	NaN	1.0	0	B1003	-120.610829	34.632093	0
Falcon 9	3170.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1004	-80.577366	28.561857	0

PREDICTIVE ANALYSIS

Make the target variable to numpy

```
Y = data['Class'].to_numpy()
Y
```

```
array([0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 1, 1, 1,
       1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 0, 0, 0, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1,
       1, 0, 1, 1, 1, 1, 0, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1], dtype=int64)
```

Preprocessing x variable

```
# students get this
X= preprocessing.StandardScaler().fit(X).transform(X)
```

Make train and test dataset

```
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size= 0.2, random_state=2)
```

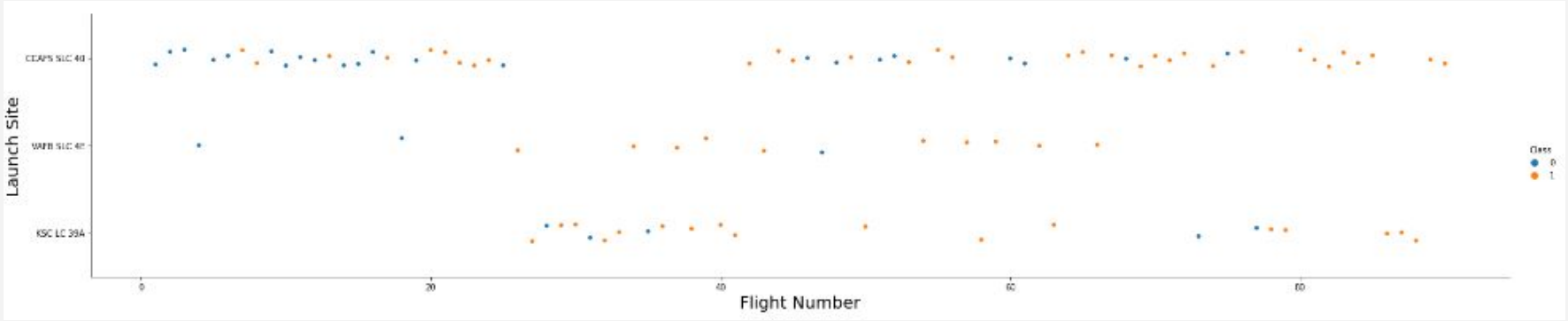
we can see we only have 18 test samples.

```
Y_test.shape
```

```
(18,)
```

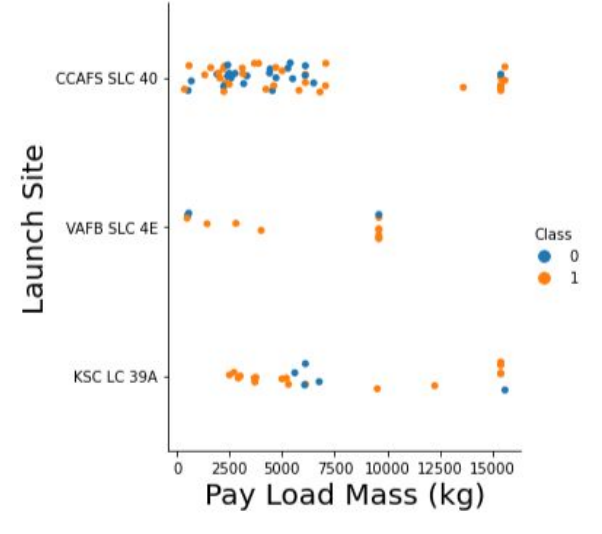
RESULT

RELATIONSHIP BETWEEN FLIGHT NUMBER AND LAUNCHSITE



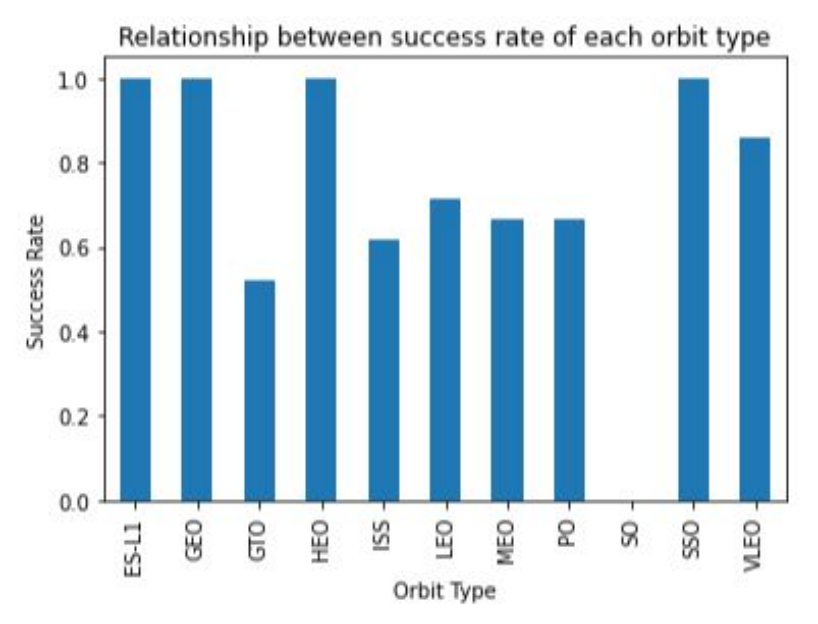
Flight numbers with number upper than 20 have good landing in all Launch Site

RELATIONSHIP BETWEEN PAYLOAD AND LAUNCH SITE



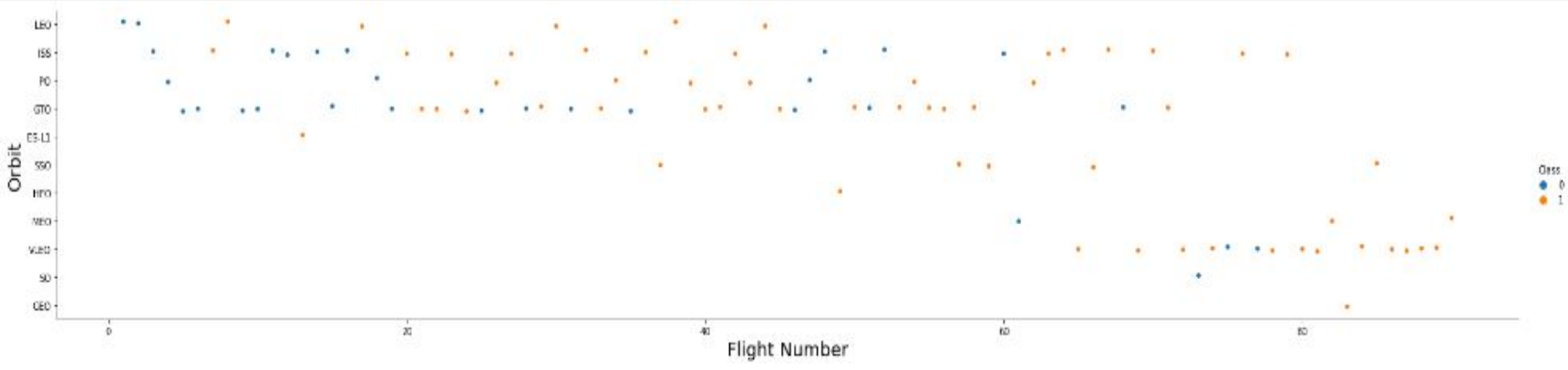
VAFB-SLC launchsite there are no rockets launched for heavy payload mass(greater than 10000).

RELATIONSHIP SUCCESS RATE OF EACH ORBIT TYPE



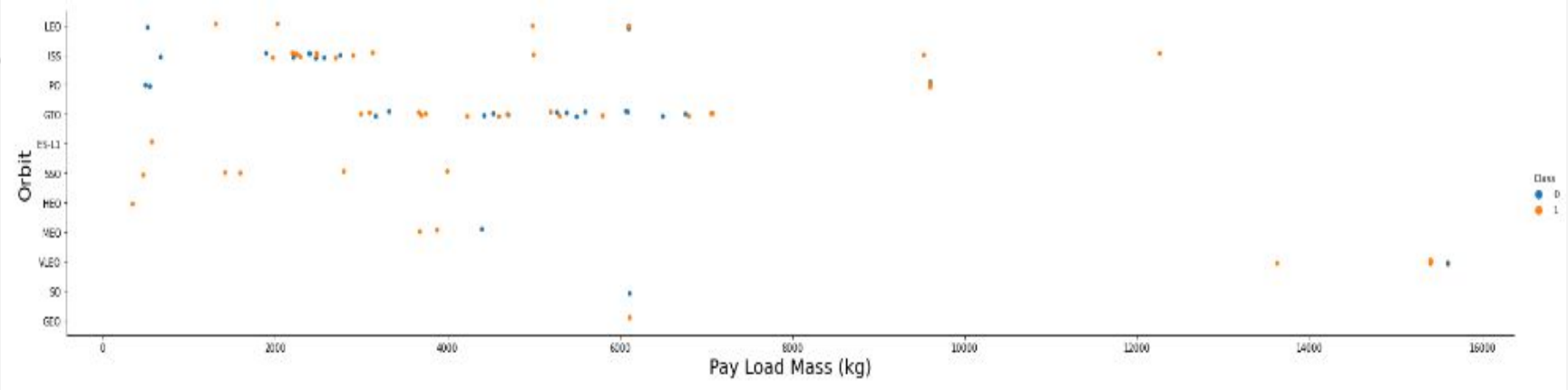
Orbit Type ESL-1, GEO, HEO, and SSO have 100% success rate.

RELATIONSHIP BETWEEN FLIGHT NUMBER AND ORBIT TYPE



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.

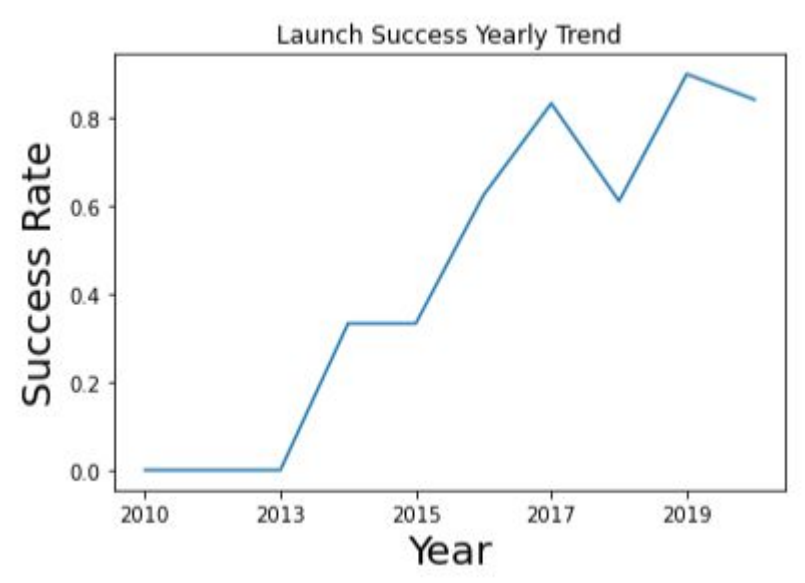
RELATIONSHIP BETWEEN PAYLOAD AND ORBIT TYPE



With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.

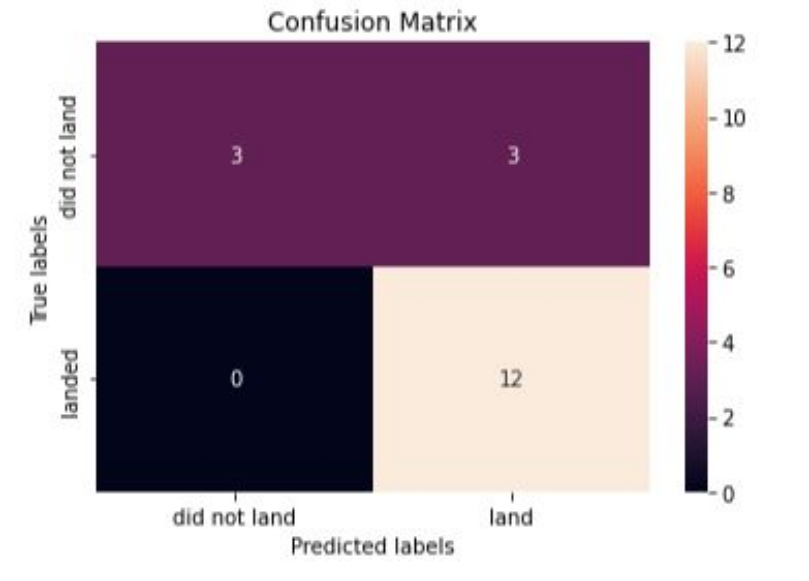
However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.

RELATIONSHIP LAUNCH SUCCESS YEARLY TREND



The success rate since 2013 kept increasing till 2020

PREDICTIVE ANALYSIS



	AUC	Accuracy
SVM	0.958	0.833
Decision Tree	0.896	0.833
KNN	0.896	0.833
Logistic Regression	0.889	0.833

SVM is the best model for this prediction because have the bes score AUC

```
tuned hpyerparameters :(best parameters)  {'C': 1.0, 'gamma': 0.03162277660168379, 'kernel': 'sigmoid'}
accuracy : 0.8482142857142856
```

DISCUSSION ?

CONCLUSION

Based on the analysis, I have **predict that Falcon9 first stage will land successfully.** I predict use Support Vector Machine as the best model with AUC score 95.8%.