APPLIED DATA SCIENCE CAPSTONE PROJECT

EXECUTIVE SUMMARY

- Problem
  - SpaceX is a aerospace manufacturer, space transportation services and communications corporation which is a disruptive just like Tesla both founded by Elon Musk. Despite being less than 20 years old SpaceX has managed to reduce the cost by more than 50% compared to other company and said to reduce by 99% when the Starship project will be completed.
  - In this capstone we will be analyzing the data ,from wiki extracted through web scrapping and SPACEX API to get insights and predict booster landing to drone ship safely .
- Data Science Methodology
  - Data Collection
  - Data Wrangling
  - Exploratory Data Analysis
  - Predictive Analysis
- Conclusion
  - Using Existing Data and Analyzing the data, SpaceX and other rocket companies can be able to see the best way to reduce the cost of launches and evolve before there tradition costly launches lead to their absoluteness and losing their client

INTRODUCTION

- In this final course of Applied Data Science Capstone, I applied the data science skills I've learned as a Data scientist for a private space launch company in this project.
- I have collected data from various sources. After my raw data has been collected, I improved the quality by performing data wrangling. Right after, I started to explore the processed data and applied my SQL skills as I query the data and gather insights.
- Overall, I gained further insights into the data by applying some basic statistical analysis and data visualization, you'll be able to see directly how variables might be related to each other. And lastly, I built, evaluated, and refined predictive models for discovering more exciting insights.

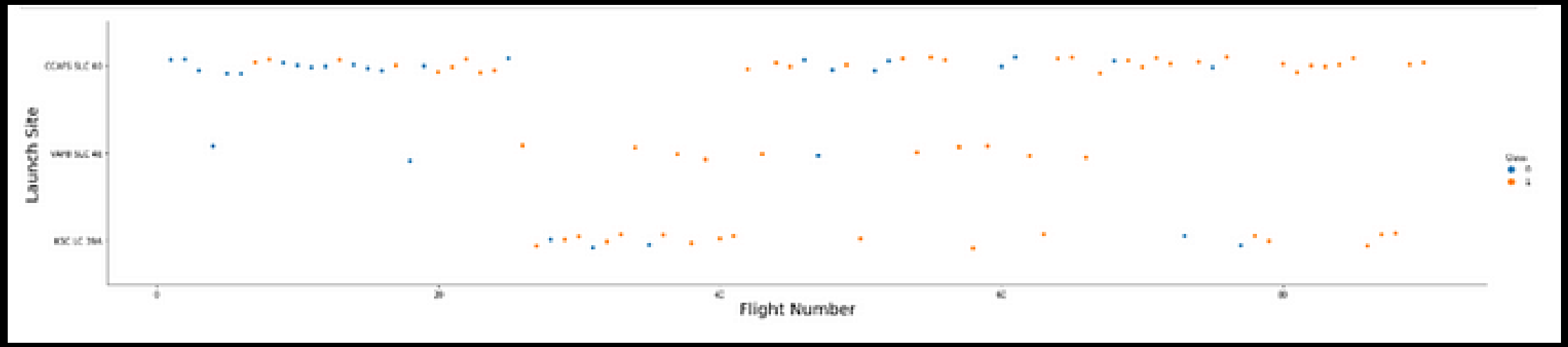
DATA COLLECTION AND WRANGLING

- Data Collection
  - Data collection through webscraping can be seen on this github link.
  - SPACEX API data collection can be seen on this github link.
- Data Wrangling
  - After collecting the data, we then check the missing data and data types and conduct data cleansing by doing the following:
    - Replace the missing data with one-Using mean or so.
    - Change data type of the data.
    - Represent categorical data using integer or float dummy numbers -one hot encoding

INTERACTIVE VISUAL ANALYTICS METHODOLOGY

Based on Exploratory Data Analysis, we can conclude that:

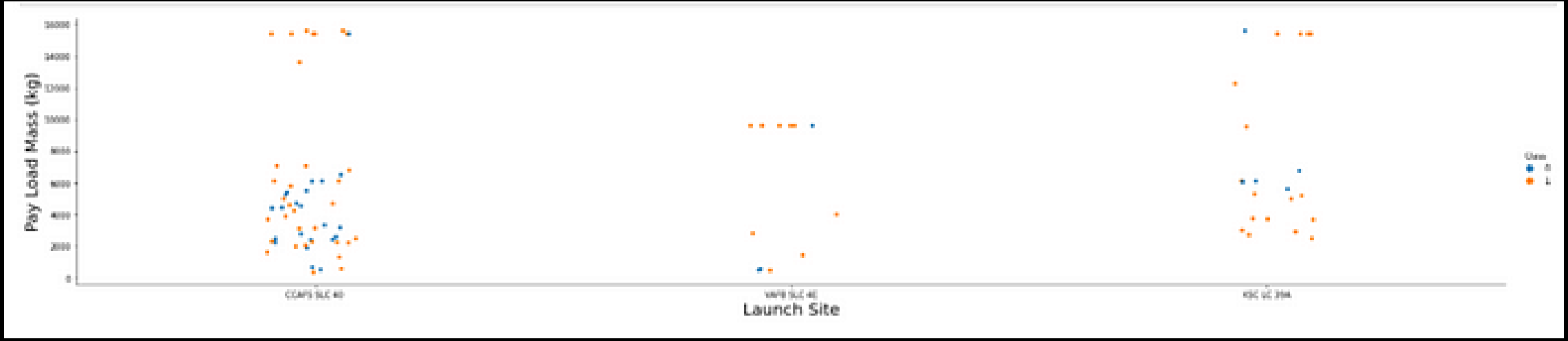
- Earlier flights launch were from CCAFS-SLC-40 site ,Followed by KSC-LC-39A
- Most Launches are Launched from CCAFS-SLC-40
- Fewer Launches from VAFB SLC 4E site



INTERACTIVE VISUAL ANALYTICS METHODOLOGY

Based on Exploratory Data Analysis, we can conclude that:

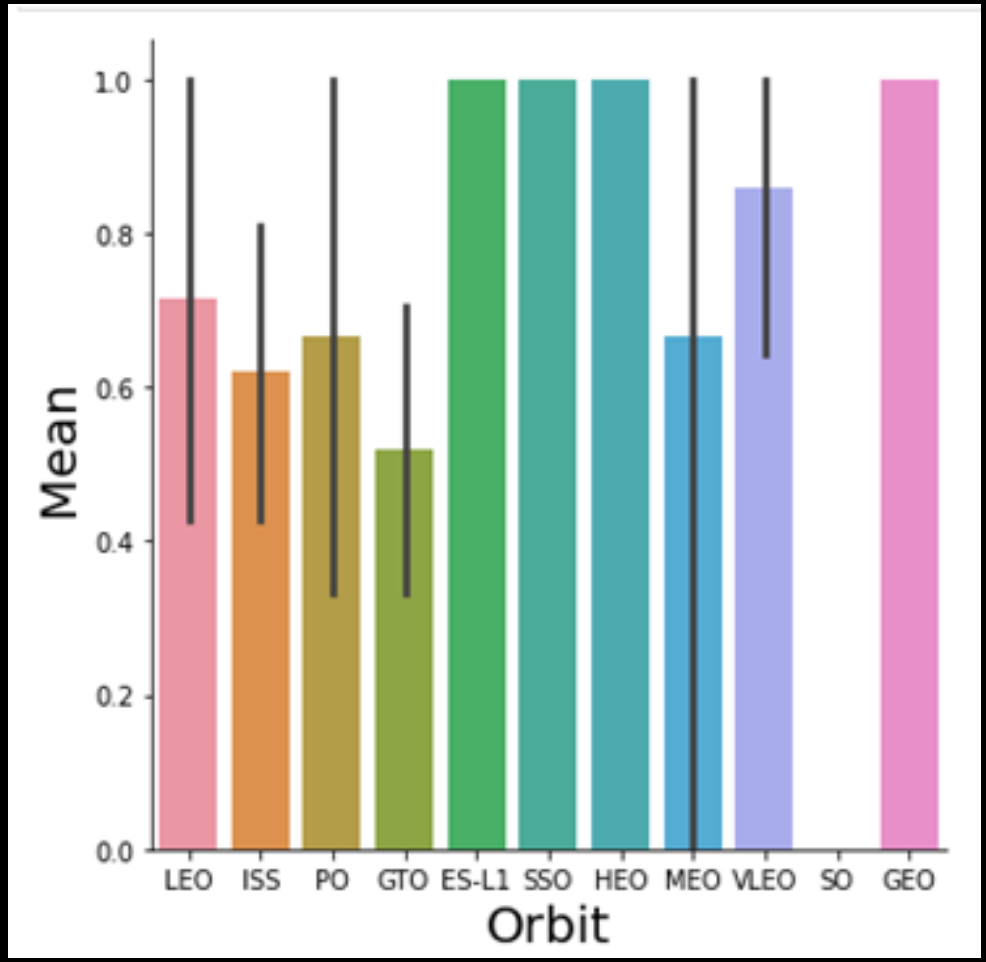
- VAFB SLC 4E has Low Payload launches
- CCAFS SLC 40 has more Higher Payload Launches and Low Payload Launches



INTERACTIVE VISUAL ANALYTICS METHODOLOGY

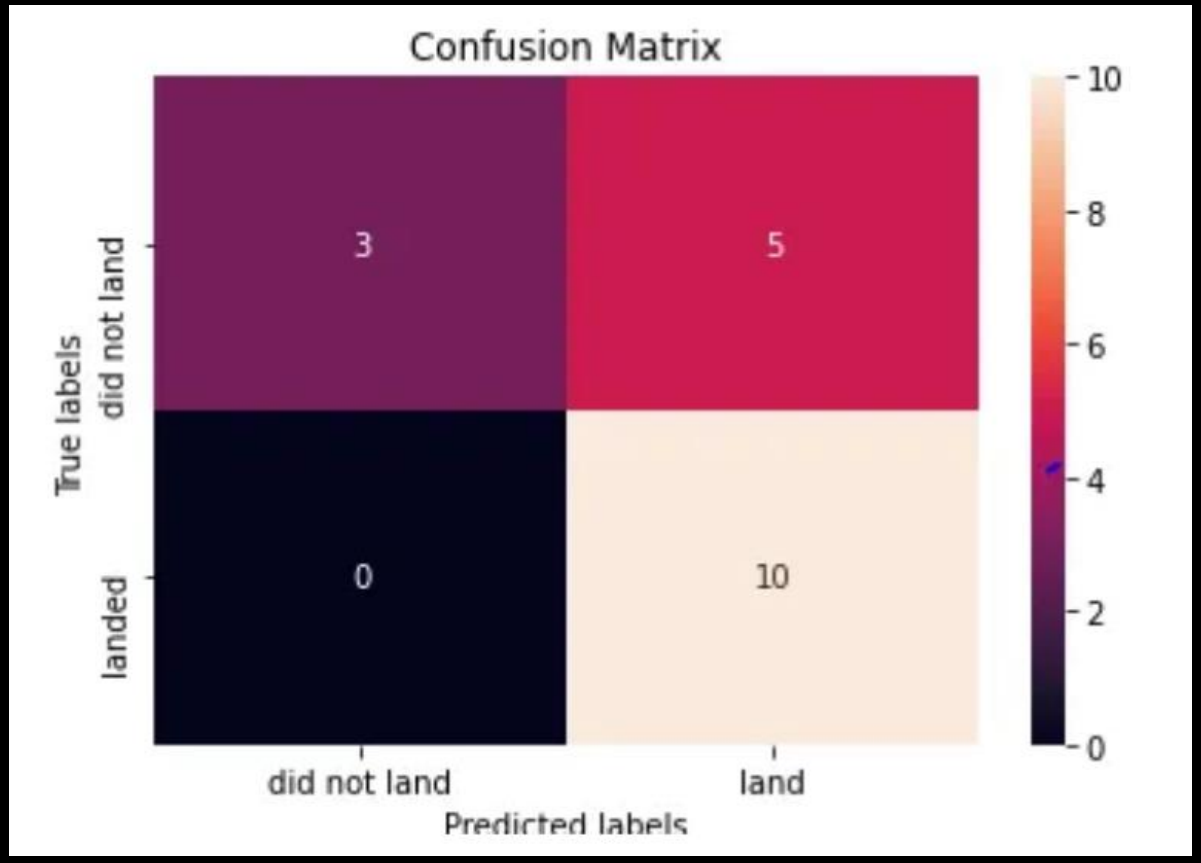
Based on Exploratory Data Analysis, we can conclude that:

- GEO,HEO & ES-L1,SS) have high success rate



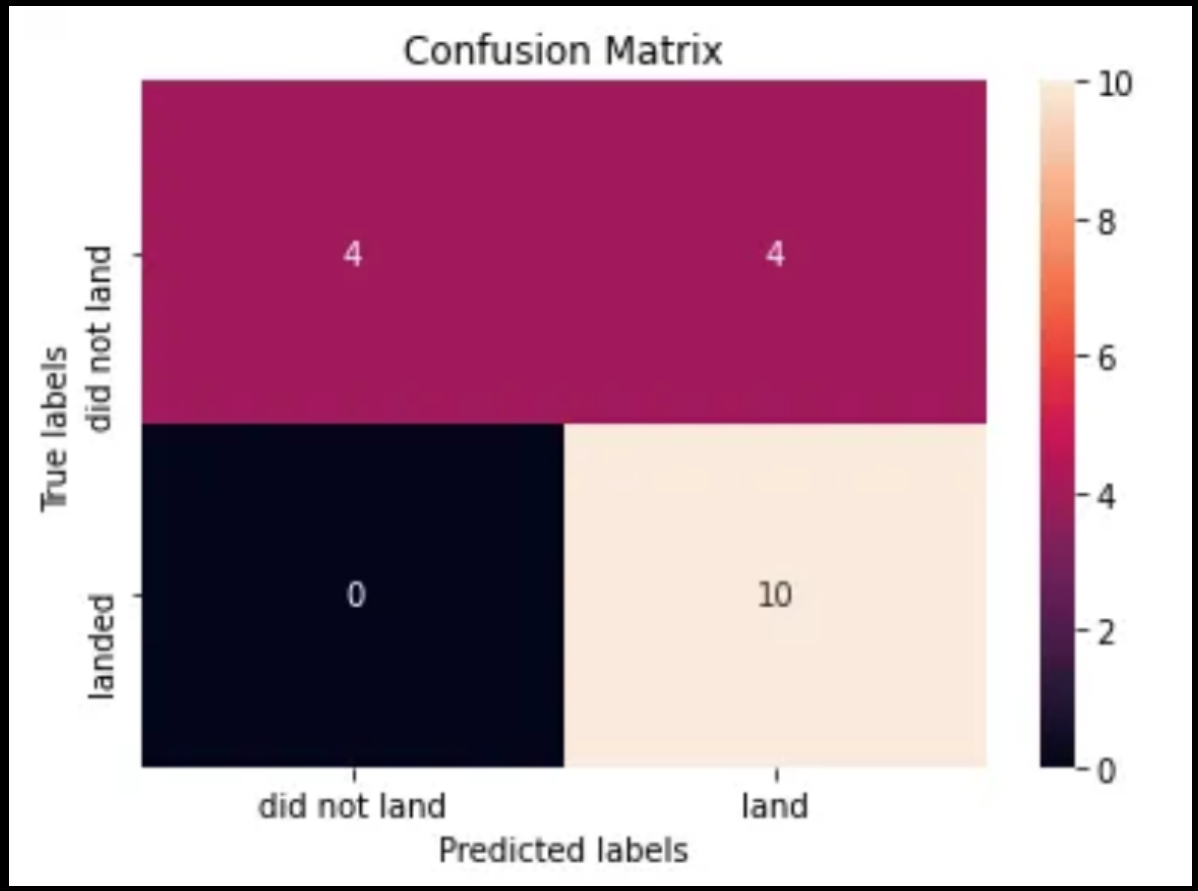
PREDICTIVE ANALYSIS METHODOLOGY

- Using the data I trained the Machine learning models such as:
  - K-Nearest Neighbor Confusion Matrix



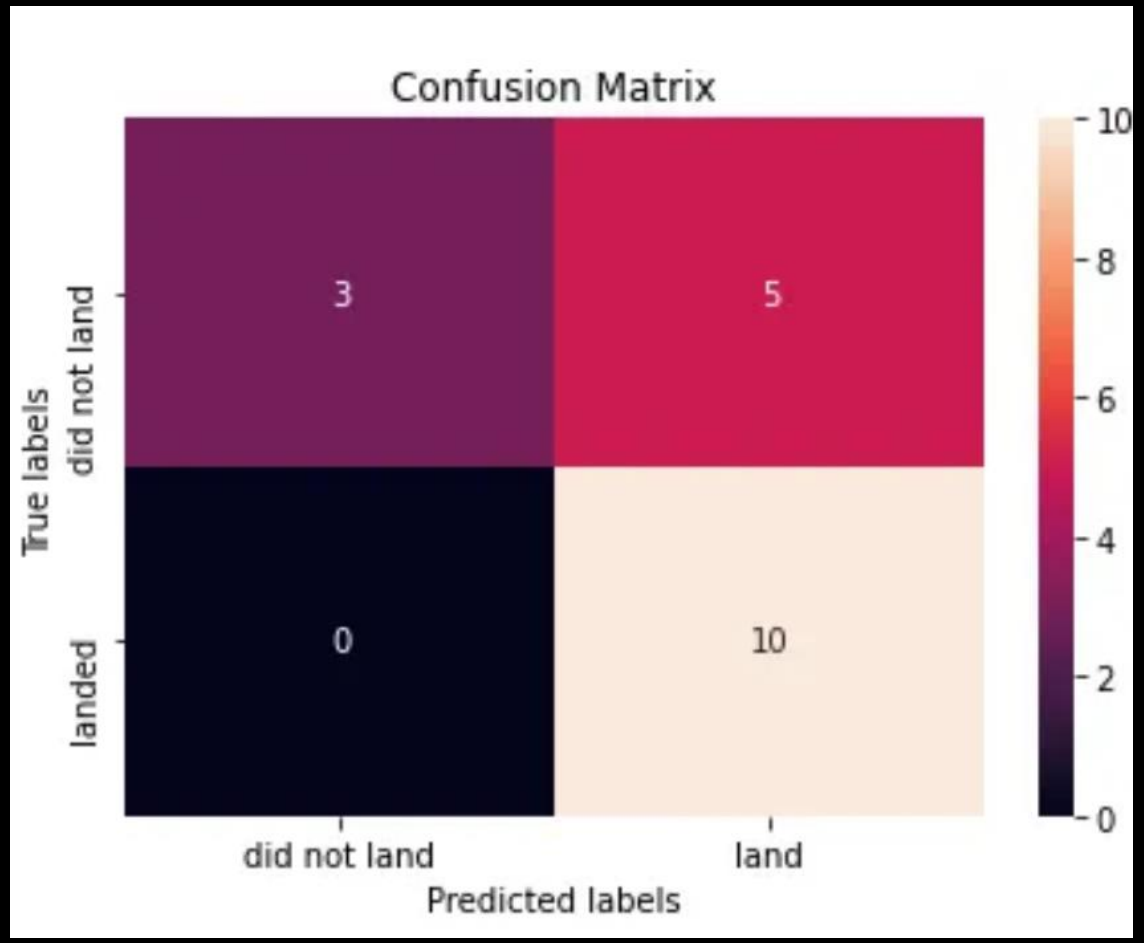
PREDICTIVE ANALYSIS METHODOLOGY

- Using the data I trained the Machine learning models such as:
  - Decision Tree Classifier



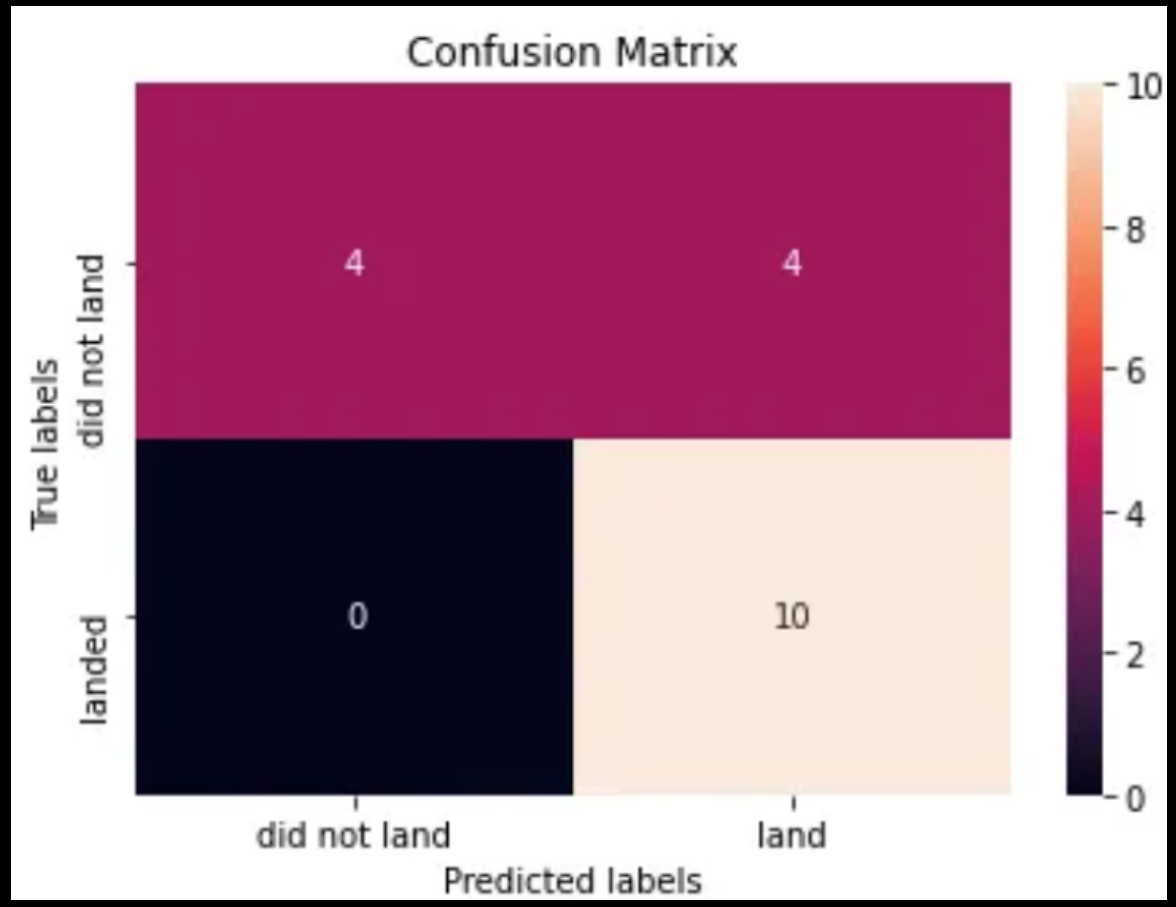
PREDICTIVE ANALYSIS METHODOLOGY

- Using the data I trained the Machine learning models such as:
  - Logistic Regression



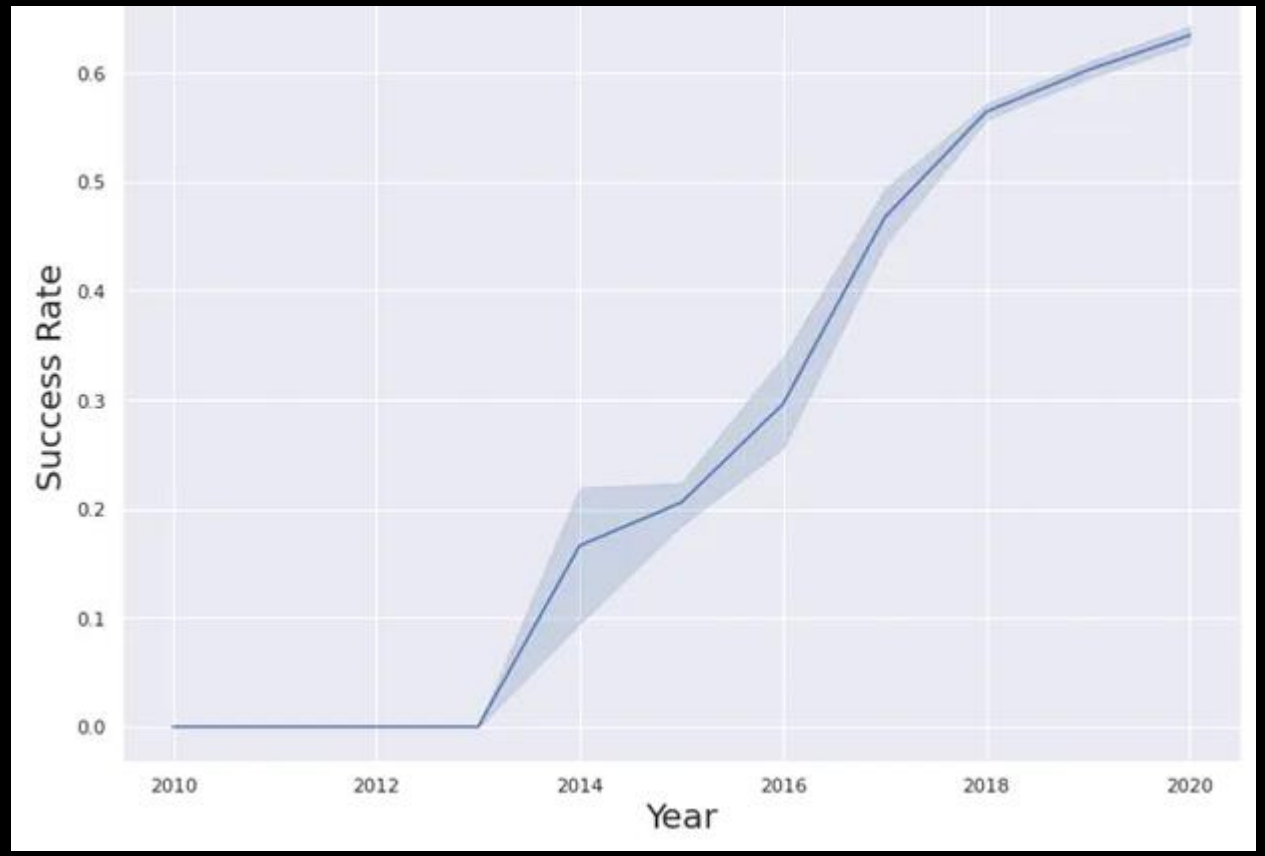
PREDICTIVE ANALYSIS METHODOLOGY

- Using the data I trained the Machine learning models such as:
  - Support Vector Machine



SQL RESULTS

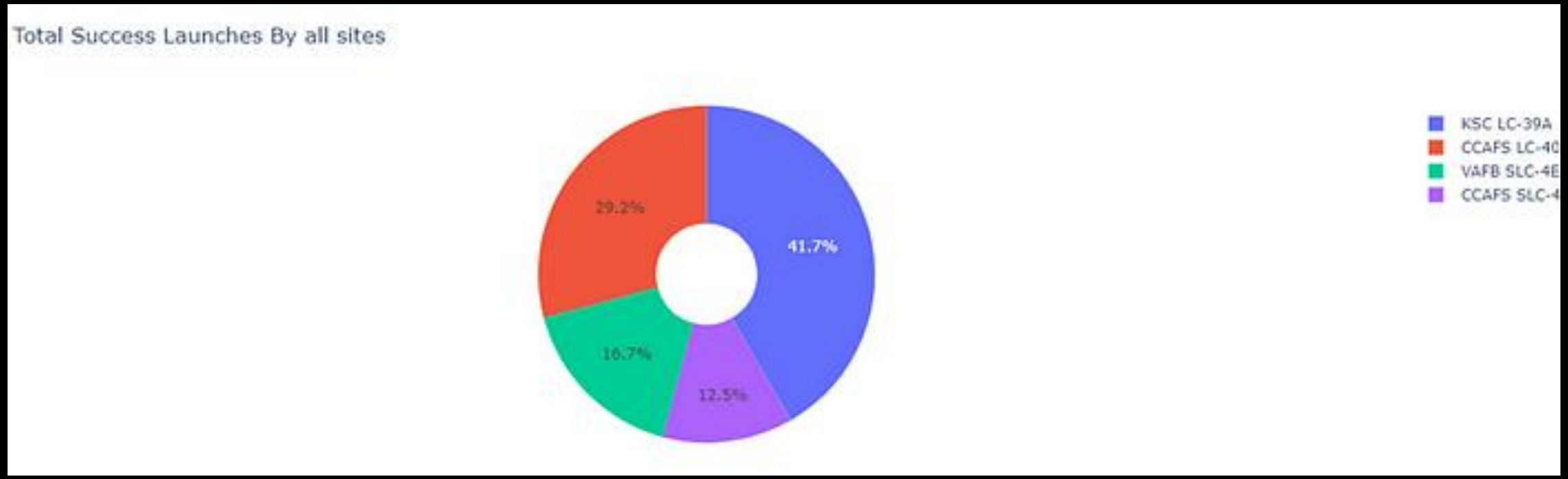
- Exploratory analysis from DB2 database using SQL can be found through this GitHub link
- Based on the SQL results, we can conclude that the success rate shows increase in landing success probability



FOLIUM: INTERACTIVE MAP

- Since SPACEX launches come from different launch sites I displayed the information of failed and successful launches as a cluster on the map.
- Through zooming in and out you can observe the clusters of success launches and failed launches .

PLOTLY DASH


Total Success Launches By all sites
29.2%
41.7%
16.7%
12.5%
KSC LC-39A
CCAFS LC-40
VAFB SLC-4E
CCAFS SLC-4

PLOTLY DASH


Payload range (Kg):
0 kg
1000 kg
2000 kg
3000 kg
4000 kg
5000 kg
6000 kg
7000 kg
8000 kg
9000 kg
10000
class
1
0.8
0.6
0.4
0.2
0
0
1000
2000
3000
4000
5000
6000
7000
Payload Mass (kg)
Booster Version
F9 v1.0 B0005
F9 v1.0 B0006
F9 v1.0 B0007
F9 v1.1
F9 v1.1 B1011
F9 v1.1 B1010
F9 v1.1 B1012
F9 v1.1 B1013
F9 v1.1 B1014
F9 v1.1 B1015
F9 v1.1 B1016
F9 v1.1 B1018
F9 FT B1019
F9 FT B1020
F9 FT B1021.1

CONCLUSION

- Using Existing Data and Analyzing the data, SpaceX and other rocket companies can be able to see the best way to reduce the cost of launches and evolve before there tradition costly launches lead to their absoluteness and losing their client
  - Please see detailed report through this GitHub link
  - For GitHub repository of all notebooks, see this GitHub link