IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Abigail Tawiah
17th April,2024

Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

- Being intrigued about how data is able to drive the world, I was encouraged and motivated to take up a course in Data Science to enable me gain insight about how data can be manipulated to draw meaningful insights. Applied Data science Capstone project is a course that I took to complete IBM professional certification in Data Science on coursera .This capstone project enabled me practice and use the skills acquired in the previous modules to real world datasets to extract,clean,explore and analyze to get insights from it.

Methodology

Executive Summary

- 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.
- This is because spaceX has developed Technology to land the first stage booster which is 70% the cost of the rocket .By landing it safely they are able to reuse the booster and the cost of the launches .
- Using their reused boosters cost 50% less of their their cost to use a new booster and this has made spaceX the company of that dominate the market .
- 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 Collection**
- The data for this project was collected through webscrapintg and is available on this github here
- SpaceX api data collection is available on this github here
- **Data Wrangling**
- After data collection, analysis was made to fish put the missing data ,and wrong data types by doing the following to clean the data :

    Replace the missing data using various techniques
    Change data type of the data.
    Represent categorical data using integer or float dummy numbers -one hot encoding

- **Exploratory Data Analysis**
- After the data was cleaned, it was then analyzed and visualized to get some insights of the launches .

Section 1

Methodology

Data Collection - Scraping

- The SpaceX data set used for this project was scrapped from the web

- Find the link to how it was achieved here

Data Wrangling

- In the data set were null values and some missing data values. There also existed values in formats that were not appropriate

- Find the link to how it was achieved here

EDA with SQL

For the purposes of EDA, sql lite queires were used. Below are some queries and brief description:

- *%sql SELECT DISTINCT Launch_Site FROM SPACEXTABLE*: Display the names of the unique launch sites in the space mission
- *%sql SELECT Launch_Site FROM SPACEXTABLE WHERE Launch_Site like 'CCA%' LIMIT 5 :* Display 5 records where launch sites begin with the string 'CCA'
- *%sql SELECT MIN(Date) FROM SPACEXTABLE WHERE Landing_Outcome like 'Success%ground pad%':* **List the date when the first succesful landing outcome in ground pad was acheived.**
- *Queries used can be found here*

- Click here to access the notebook containing the EDA queries with sql

EDA with Data Visualization

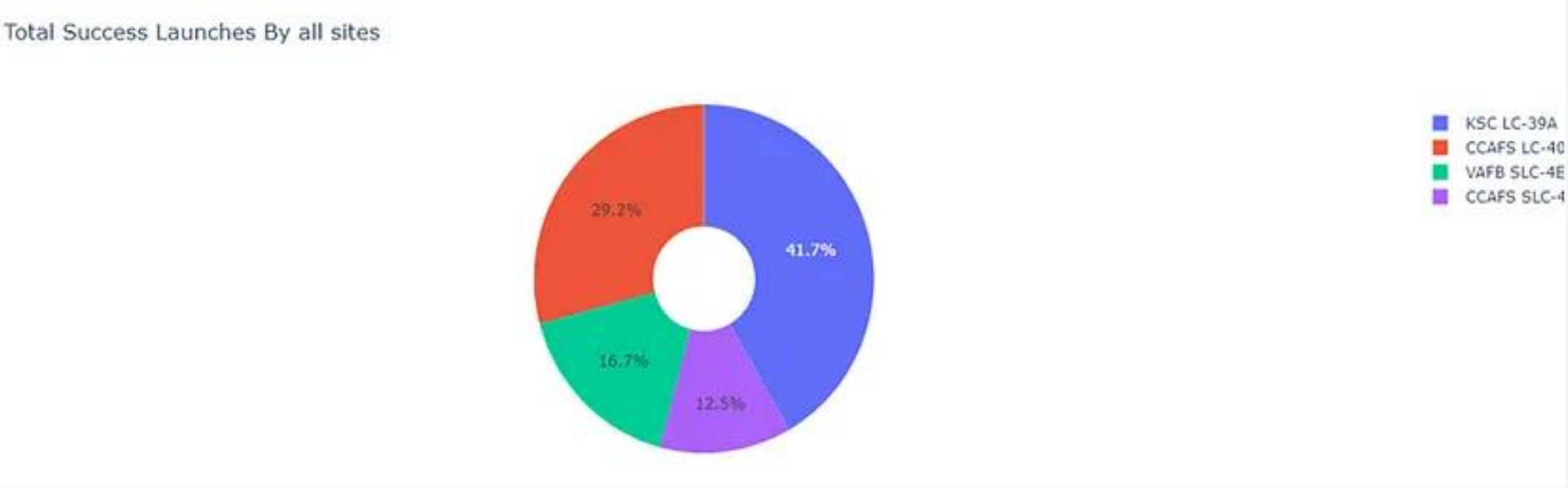
- Scatter plots were used to visualize the data. It was used because, it is ideal to display the relationship between two variables and observe the nature of such relation
- Here is a link to the notebook

Build an Interactive Map with Folium

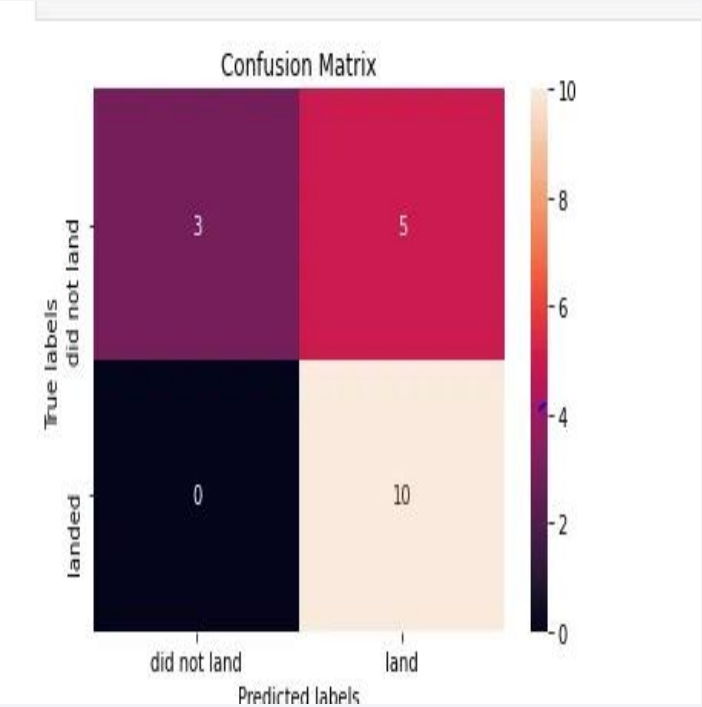
- Interractive maps are useful for data exploration and communicating research
- SpaceX launches from different site and as such had to display the information of failed and successful launches as a cluster on the map. The ability to zoom in and out aids one to easily spot sites on the map.

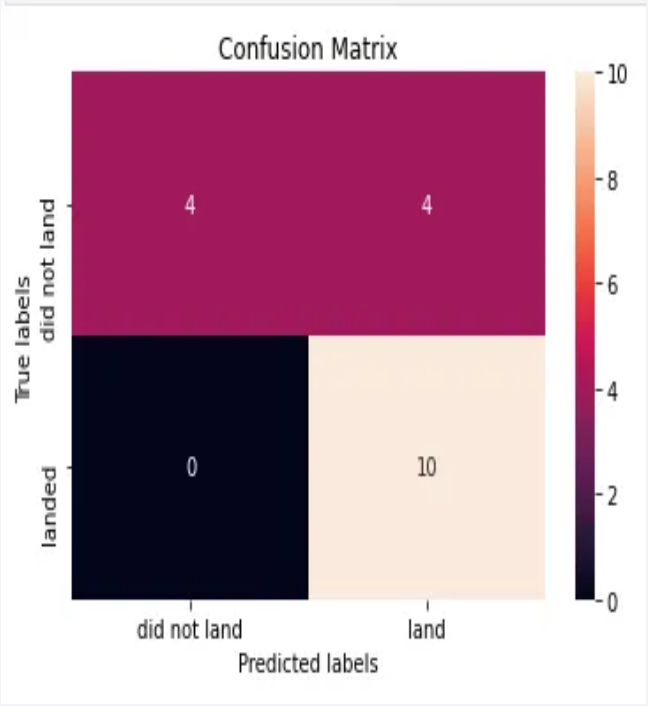
Build a Dashboard with Plotly Dash

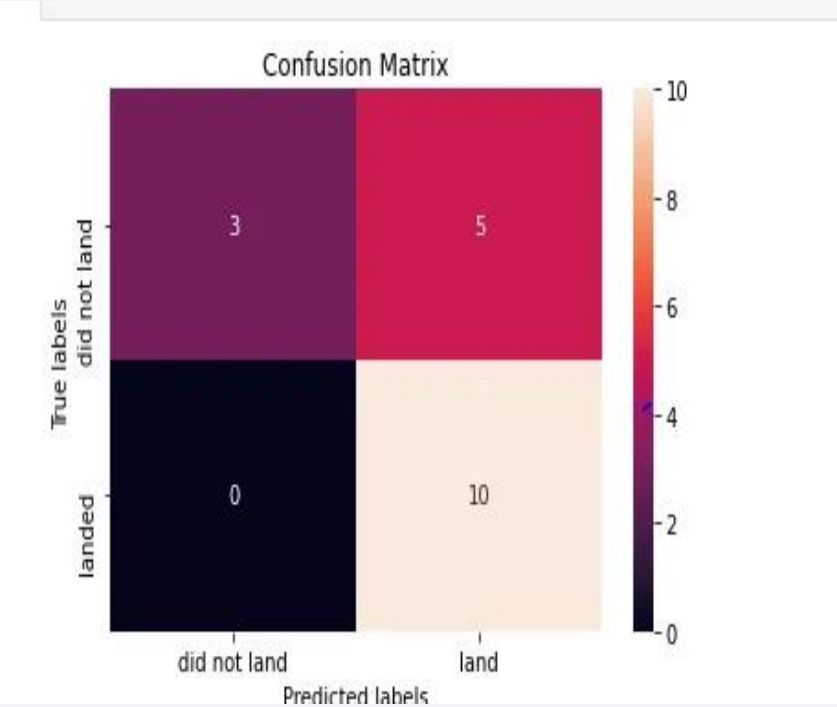
- Plotly Dash delivers interactive, customizaeable data apps
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose



Predictive Analysis (Classification)







Link here

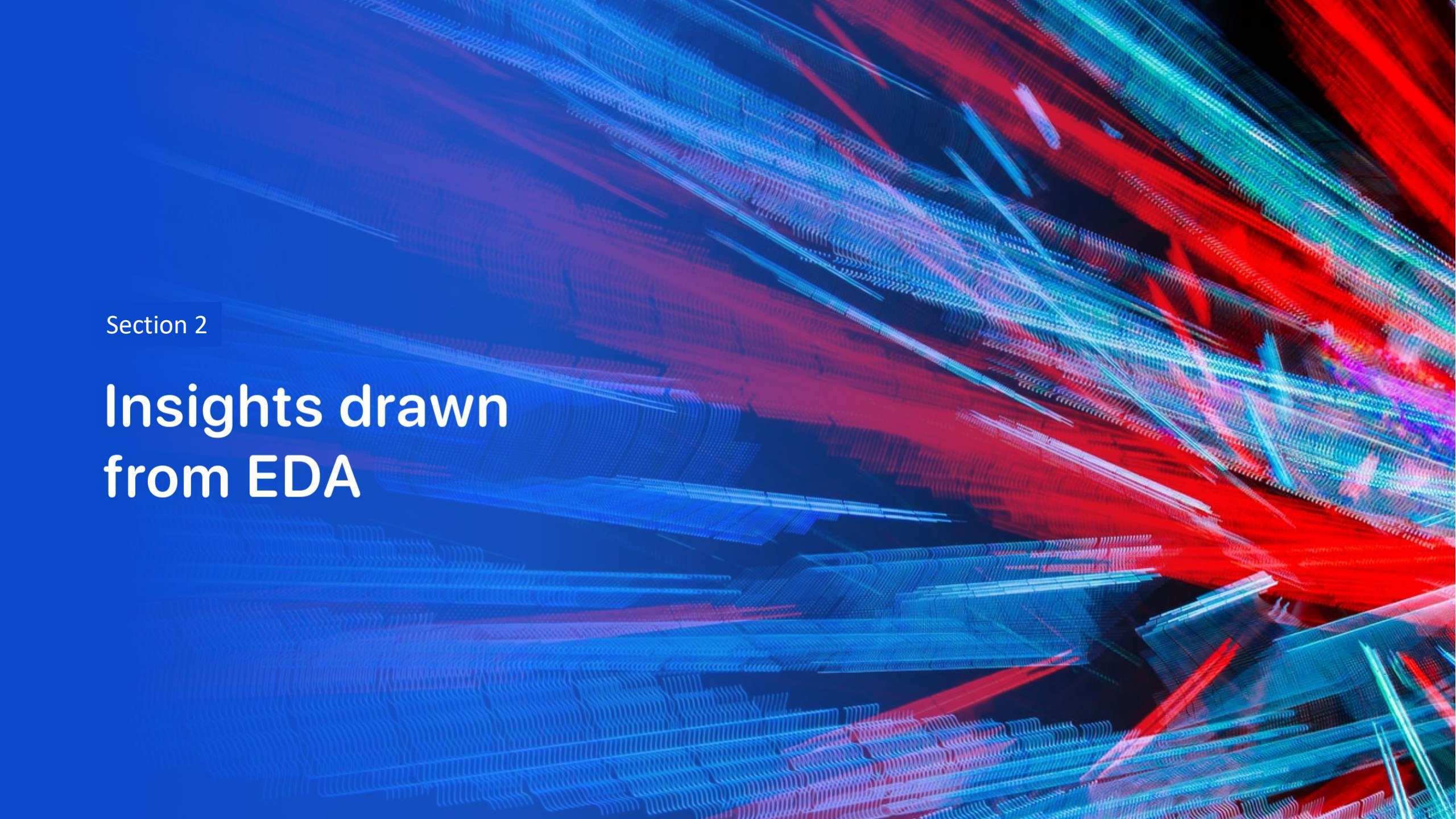
Results

Insights from the analysis depicts that most launches were made from Kennedy Space center

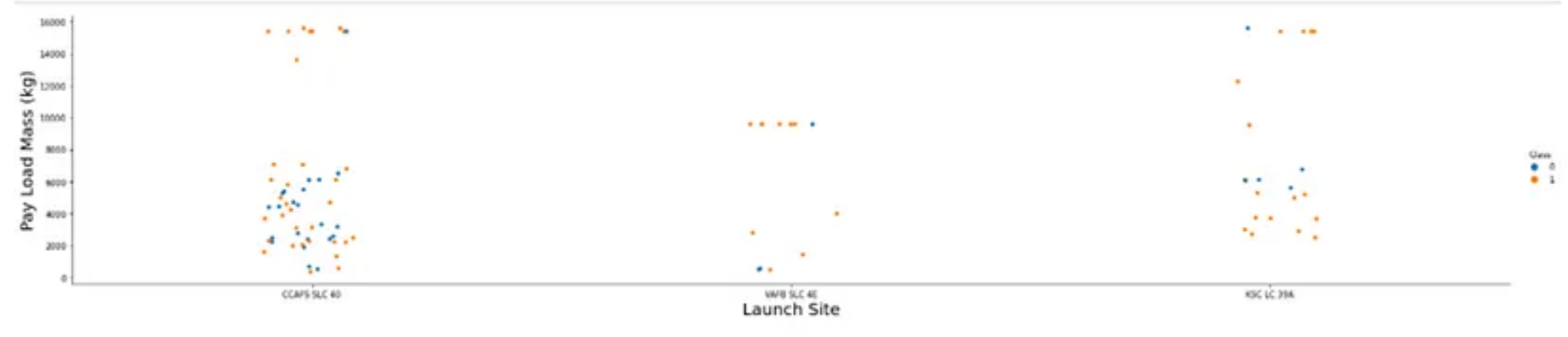
Most launches were from KSC PAD 39A since most were close to VLEO, GEO and ISS

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 .


Section 2
Insights drawn from EDA

Payload vs. Launch Site



From the chart, it can be inferred that:

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

Thank you!