Winning Space Race with Data Science

Abigail Tawiah
17th April,2024

Outline

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

Executive Summary

This executive summary provides a comprehensive overview of the SpaceX dataset report, offering insights into key findings, trends, and recommendations for stakeholders.

1. Introduction: The report delves into the analysis of data collected from SpaceX missions, encompassing various parameters such as launch success rates, payload characteristics, mission outcomes, and more. This dataset serves as a valuable resource for understanding the performance and evolution of SpaceX's endeavors in space exploration.

2. Key Findings:

•**Launch Success Rate:** Analysis reveals a high overall success rate in SpaceX launches, demonstrating the company's proficiency in achieving mission objectives.

•**Payload Diversity:** The dataset illustrates a wide array of payloads launched by SpaceX, ranging from communication satellites to scientific instruments, reflecting the company's versatility in serving diverse markets.

•**Reusable Technology Impact:** Examination of reusable rocket technology indicates a notable reduction in launch costs over time, fostering sustainability and economic viability in space missions.

•**Market Dominance:** SpaceX emerges as a dominant player in the commercial space industry, capturing a significant share of satellite launch contracts and governmental missions.

•**Innovation and Expansion:** Continued innovation in rocket design, propulsion systems, and mission architectures underscores SpaceX's commitment to advancing space exploration capabilities and expanding its market reach.

3. Trends and Insights:

•**Cost Efficiency:** SpaceX's emphasis on reusability has led to substantial cost savings compared to traditional expendable launch systems, driving down the price of access to space.

•**Market Disruption:** The company's disruptive business model has challenged established aerospace norms, prompting competitors to adapt and innovate to remain competitive.

•**International Collaboration:** Collaboration with international space agencies and private entities has facilitated global cooperation in space exploration, opening avenues for collaborative missions and technological exchange.

•**Human Spaceflight:** SpaceX's successful crewed missions, such as the Crew Dragon flights to the International Space Station (ISS), signal a new era in human spaceflight, with implications for future space tourism and interplanetary exploration.

4. Recommendations:

•**Continued Innovation:** SpaceX should maintain its focus on innovation and technology development to sustain its competitive edge and address evolving market demands.

•**Safety and Reliability:** Prioritizing safety and reliability in all aspects of mission planning and execution is crucial to uphold SpaceX's reputation and ensure mission success.

•**Market Diversification:** Exploring opportunities in emerging markets such as small satellite launches, space tourism, and lunar exploration can diversify revenue streams and mitigate risks associated with market fluctuations.

•**Sustainability:** Investing in sustainable practices and technologies, including further advancements in reusable rocket technology and environmental impact mitigation measures, will contribute to long-term viability and environmental stewardship in space exploration.

5. Conclusion: The SpaceX dataset report underscores the company's remarkable achievements, highlighting its pivotal role in shaping the future of space exploration. By leveraging data-driven insights and strategic foresight, SpaceX is well-positioned to continue leading innovation in the aerospace industry and pushing the boundaries of human exploration beyond Earth's orbit.

This executive summary encapsulates the key insights and recommendations gleaned from the comprehensive analysis of the SpaceX dataset, providing stakeholders with valuable perspectives to inform strategic decision-making and future endeavors in space exploration.

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

Problem Statement

- 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

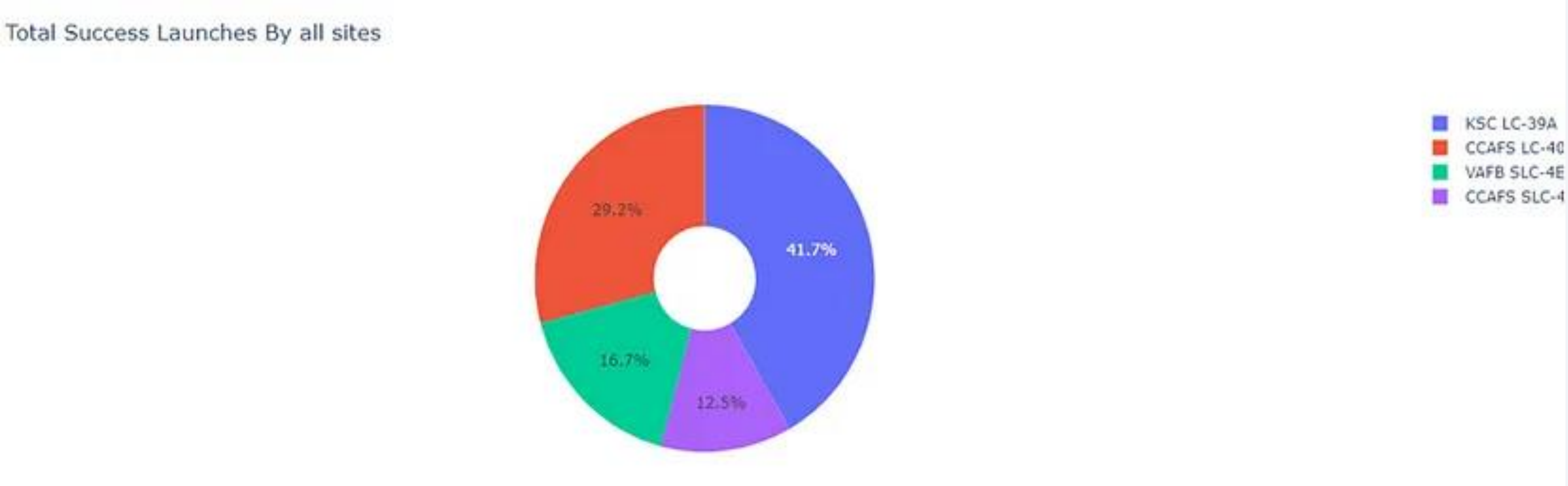
- Scatter plots were used to visualize the data. It was used because, it is ideal to display the relationship between two variables and observe the nature of such relation
- Here is a link to the [notebook]

Build an Interactive Map with Folium

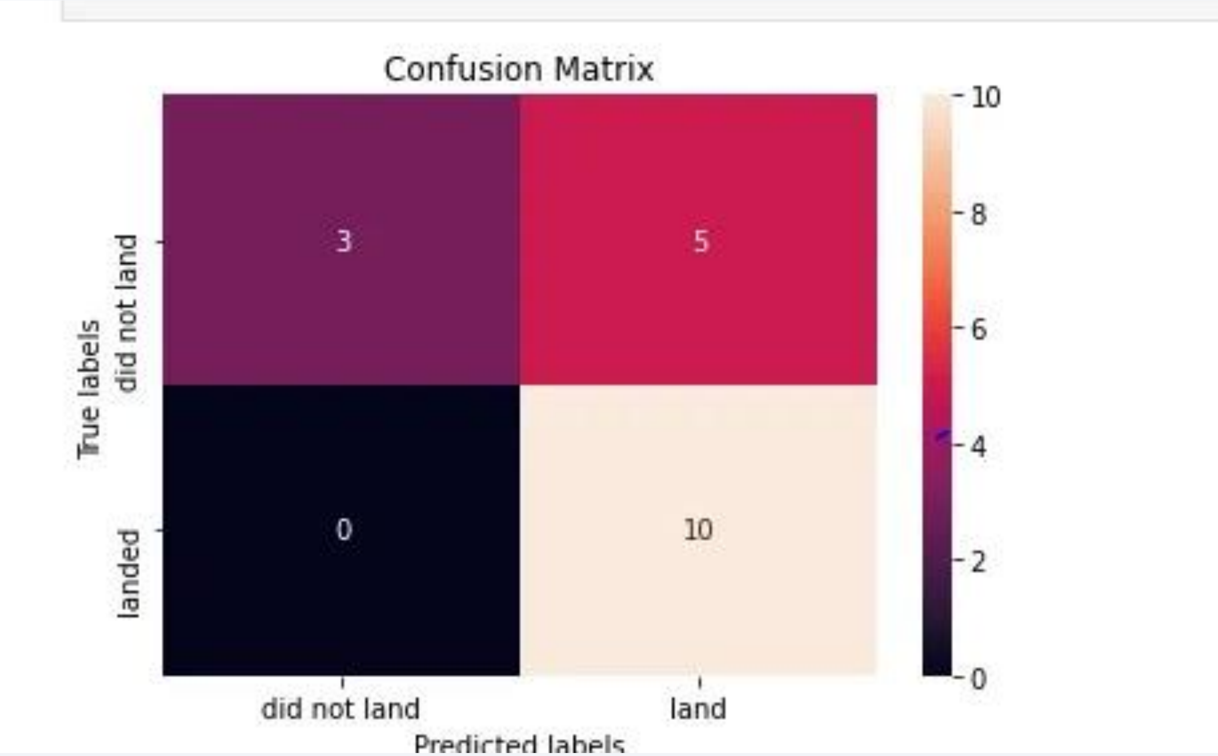
- Interractive maps are useful for data exploration and communicating research
- SpaceX launches from different site and as such had to display the information of failed and successful launches as a cluster on the map. The ability to zoom in and out aids one to easily spot sites on the map.

Build a Dashboard with Plotly Dash

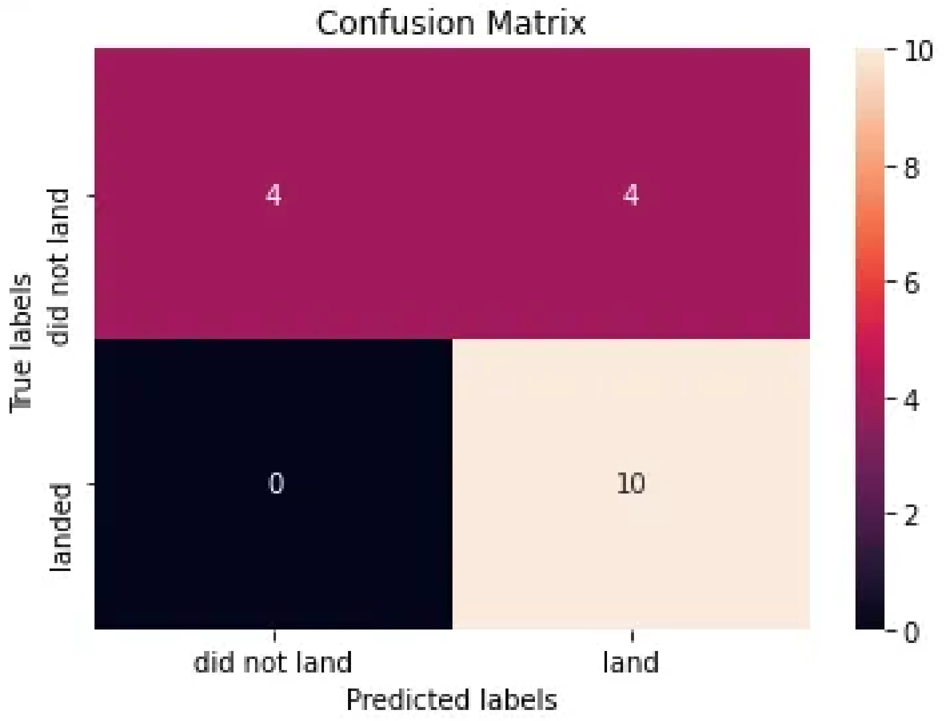
- Plotly Dash delivers interactive, customizaeable data apps
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose



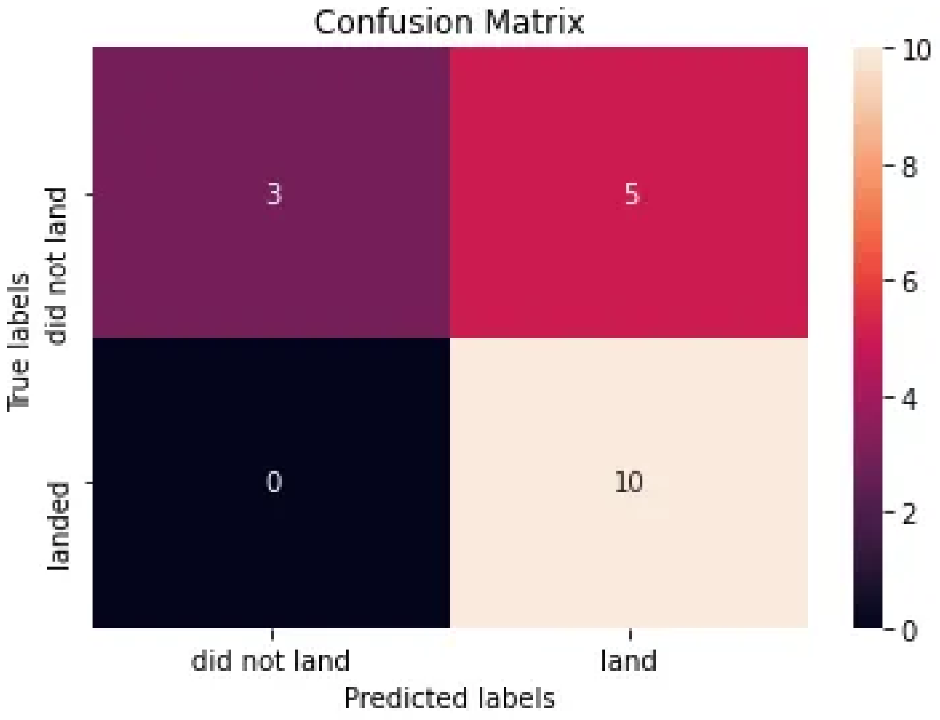
Predictive Analysis (Classification)



Using the data I trained the Machine learning models such as:
KNeighborsClassifier

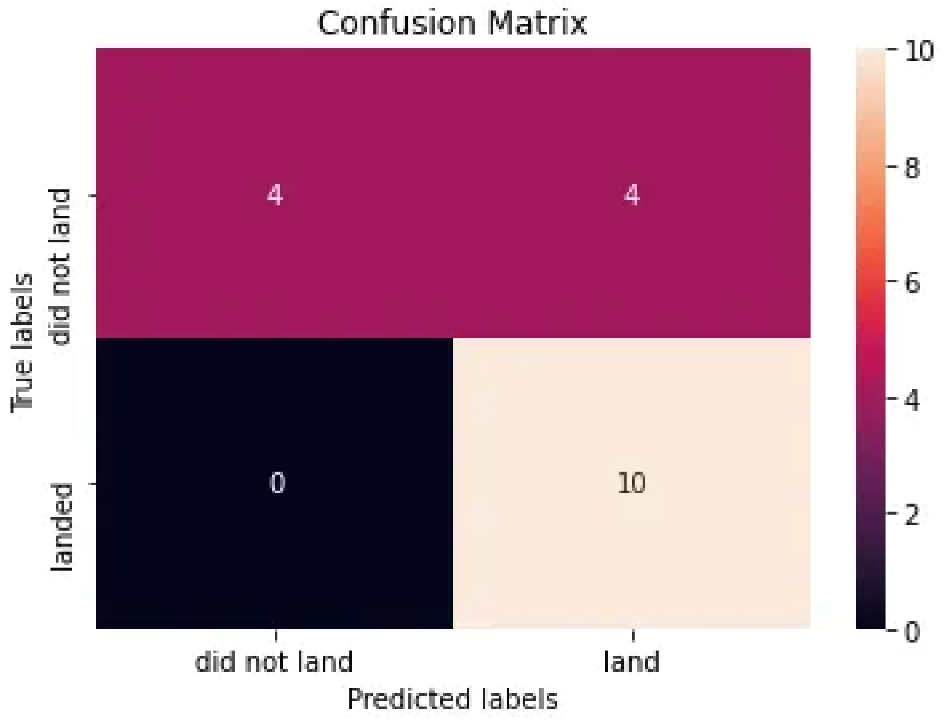


Decision Tree classifier



Logistic Regression

Support Vector Machine



I tuned the models to obtain the most accurate model of all of these evaluating their score ,best score and confusion matrix plot And concluded that KNN model has the best score ,accuracy and least bias confusion matrix

```python
parameters = {'n_neighbors': [1, 2, 3, 4, 5, 6, 7, 8, 9, 10],
              'algorithm': ['auto', 'ball_tree', 'kd_tree', 'brute'],
              'p': [1,2]}

KNN = KNeighborsClassifier()
gscv=GridSearchCV(KNN,parameters,scoring="accuracy",cv=10)
KNN_cv=gscv.fit(X_train,y_train)
```

```python
print("Accuracy",KNN_cv.score(X_test,y_test))
```

```
Accuracy 0.7777777777777778
```

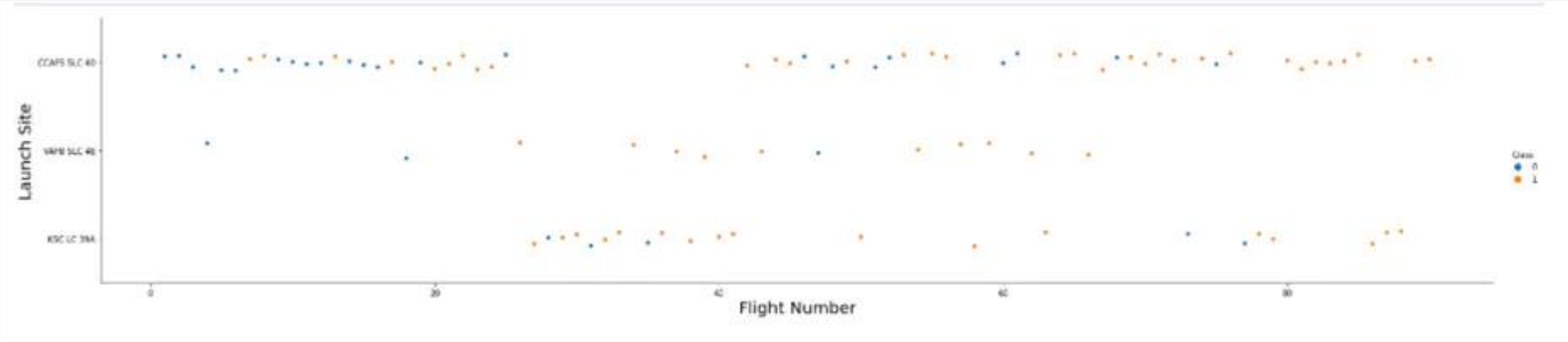
```python
print("tuned hpyerparameters :(best parameters) ",KNN_cv.best_params_)
print("accuracy :",KNN_cv.best_score_)
```

```
tuned hpyerparameters :(best parameters)  {'algorithm': 'auto', 'n_neighbors': 4, 'p': 1}
accuracy : 0.8767857142857143
```

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

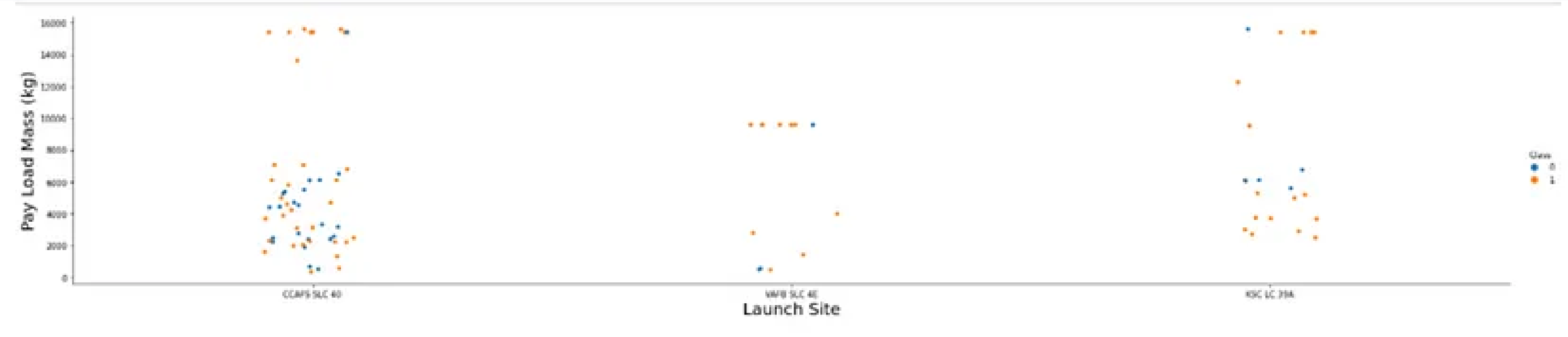


From the chart, it can be inferred that:

●VAFB SLC 4E has Low Payload launches

●CCAFS SLC 40 has more Higher Payload Launches and Low Payload Lauches .

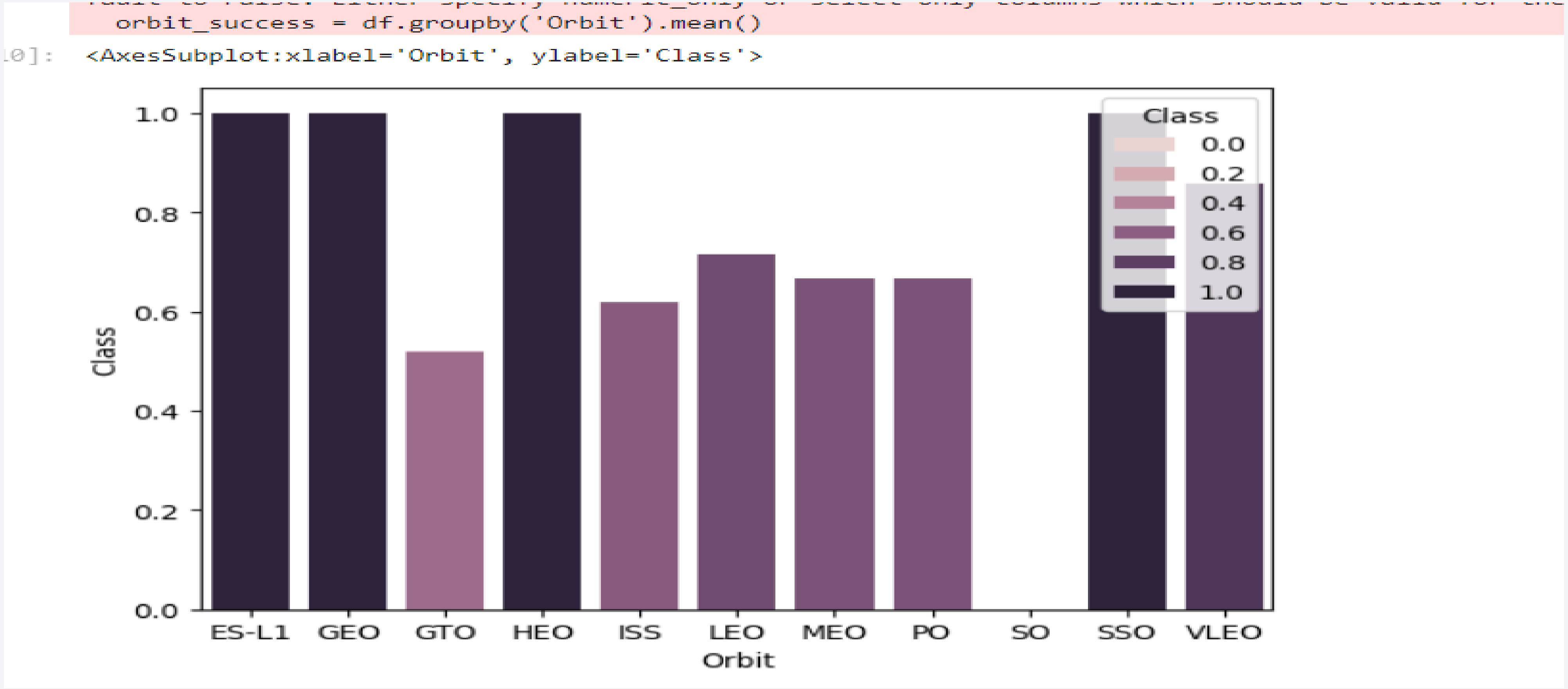
Payload vs. Launch Site



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 .

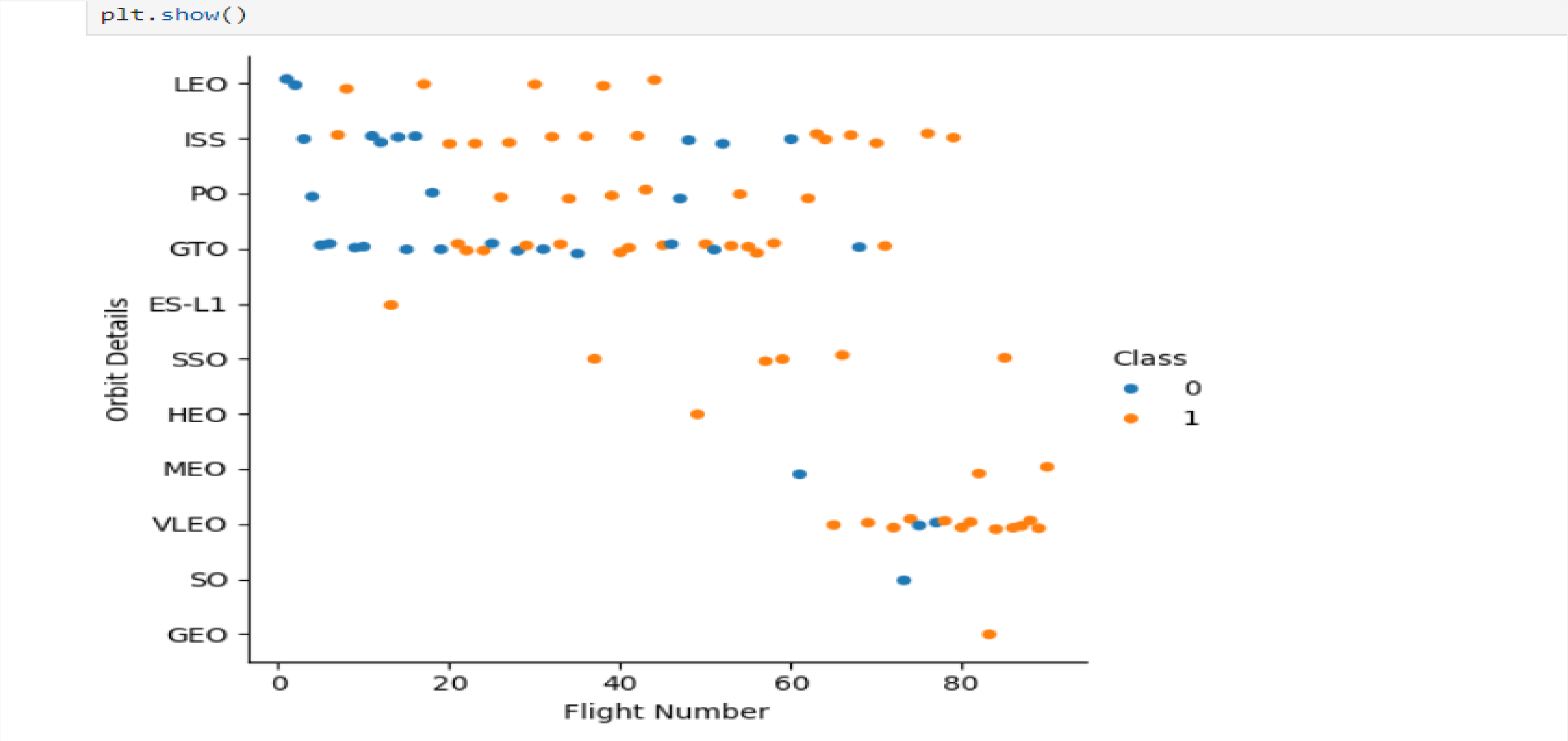
Success Rate vs. Orbit Type



From the chart, it can be inferred that:

●ES- LIO,HEO and GEO has the highest success rate

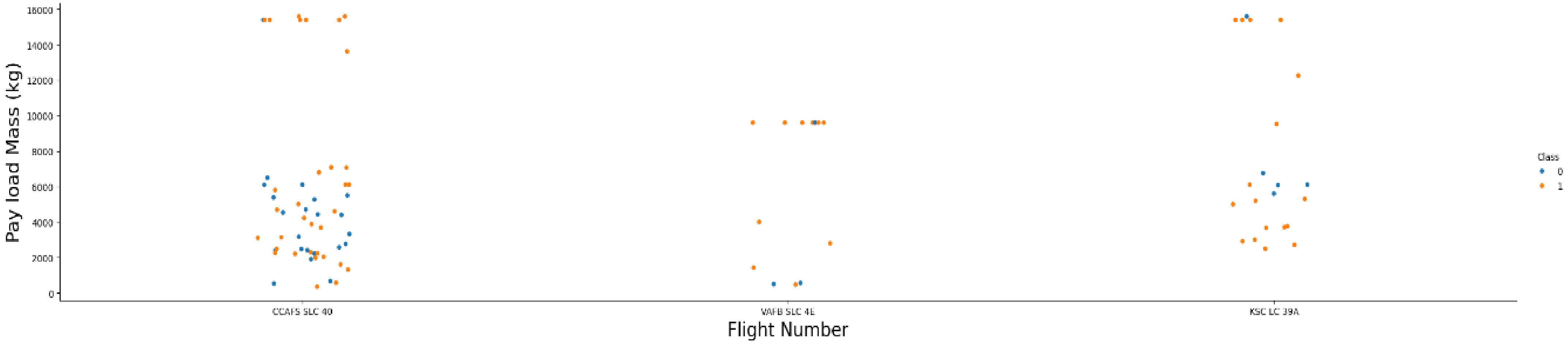
Flight Number vs. Orbit Type



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.

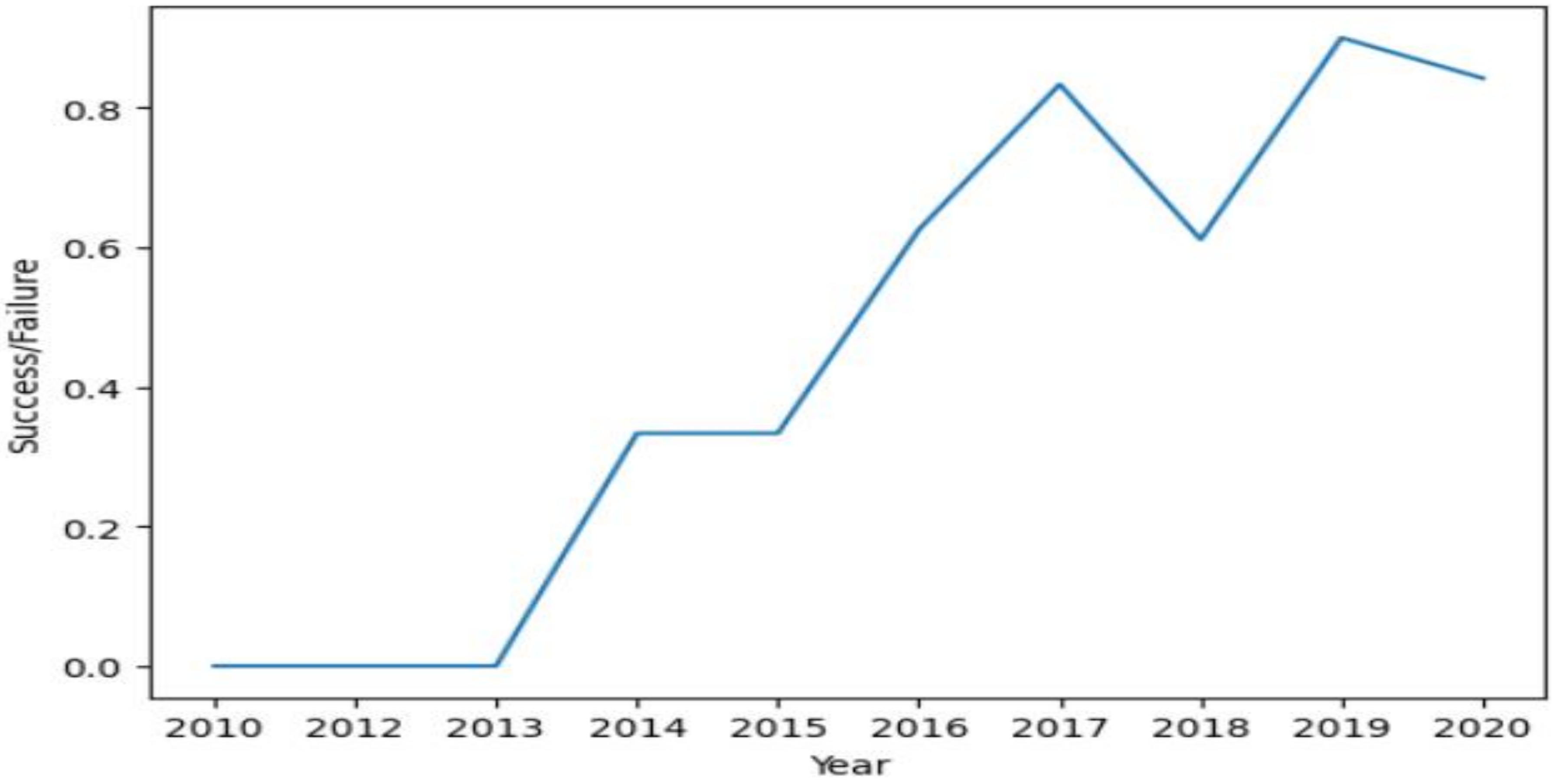
Payload vs. Orbit Type

```
# Plot a scatter point chart with x axis to be Pay Load Mass (kg) and y axis to be the launch site, and hue to be the class value
sns.catplot(y="PayloadMass", x="LaunchSite", hue="Class", data=df, aspect = 5)
plt.xlabel("Flight Number",fontsize=20)
plt.ylabel("Pay load Mass (kg)",fontsize=20)
plt.show()
```



Now if you observe Payload Vs. Launch Site scatter point chart you will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).

Launch Success Yearly Trend



you can observe that the sucess rate since 2013 kept increasing till 2020

All Launch Site Names

```
# df["Launch_Site"].unique()
%sql SELECT DISTINCT Launch_Site FROM SPACEXTABLE
```

```
 * sqlite:///my_data1.db
Done.
```

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

The query displays the unique Launch sites available in the Spacextable

Launch Site Names Begin with 'CCA'

Display 5 records where launch sites begin with the string 'CCA'

```
# df[df.Launch_Site.str.startswith('CCA')]
# df.query("Launch_Site.str.contains('CCA')", engine='python')
%sql SELECT Launch_Site FROM SPACEXTABLE WHERE Launch_Site like 'CCA%' LIMIT 5
```

```
 * sqlite:///my_data1.db
Done.
```

Launch_Site
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40

The query displays five records where launch site begins with the string 'CCA'

Total Payload Mass

```
]: # df['PAYLOAD_MASS__KG_'].sum()
   %sql SELECT SUM(PAYLOAD_MASS__KG_) AS total_payload_mass FROM SPACEXTABLE

    * sqlite:///my_data1.db
   Done.
```

total_payload_mass
619967

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

Average Payload Mass by F9 v1.1

```
# f['PAYLOAD_MASS__KG_'].meabn()
%sql SELECT AVG(PAYLOAD_MASS__KG_) AS average_payload_mass FROM SPACEXTABLE
```

```
 * sqlite:///my_data1.db
Done.
```

average_payload_mass
6138.287128712871

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

First Successful Ground Landing Date

```
%sql SELECT MIN(Date) FROM SPACEXTABLE WHERE Landing_Outcome like 'Success%ground pad%'
```

```
 * sqlite:///my_data1.db
Done.
```

MIN(Date)
2015-12-22

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

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql SELECT * FROM SPACEXTABLE WHERE Landing_Outcome like '%drone%' LIMIT 10
```

```
 * sqlite:///my_data1.db
Done.
```

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2015-01-10	9:47:00	F9 v1.1 B1012	CCAFS LC-40	SpaceX CRS-5	2395	LEO (ISS)	NASA (CRS)	Success	Failure (drone ship)
2015-04-14	20:10:00	F9 v1.1 B1015	CCAFS LC-40	SpaceX CRS-6	1898	LEO (ISS)	NASA (CRS)	Success	Failure (drone ship)
2015-06-28	14:21:00	F9 v1.1 B1018	CCAFS LC-40	SpaceX CRS-7	1952	LEO (ISS)	NASA (CRS)	Failure (in flight)	Precluded (drone ship)
2016-01-17	18:42:00	F9 v1.1 B1017	VAFB SLC-4E	Jason-3	553	LEO	NASA (LSP) NOAA CNES	Success	Failure (drone ship)
2016-03-04	23:35:00	F9 FT B1020	CCAFS LC-40	SES-9	5271	GTO	SES	Success	Failure (drone ship)
2016-04-08	20:43:00	F9 FT B1021.1	CCAFS LC-40	SpaceX CRS-8	3136	LEO (ISS)	NASA (CRS)	Success	Success (drone ship)
2016-05-06	5:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2016-05-27	21:39:00	F9 FT B1023.1	CCAFS LC-40	Thaicom 8	3100	GTO	Thaicom	Success	Success (drone ship)
2016-06-15	14:29:00	F9 FT B1024	CCAFS LC-40	ABS-2A Eutelsat 117 West B	3600	GTO	ABS Eutelsat	Success	Failure (drone ship)
2016-08-14	5:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

Total Number of Successful and Failure Mission Outcomes

```
%sql SELECT Mission_Outcome, COUNT(Mission_Outcome) as total_outcomes FROM SPACEXTABLE GROUP BY Mission_Outcome
```

```
 * sqlite:///my_data1.db
Done.
```

Mission_Outcome	total_outcomes
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

List the total number of successful and failure mission outcomes

Boosters Carried Maximum Payload

```
%sql SELECT Booster_Version FROM SPACEXTABLE WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTABLE)
```

```
 * sqlite:///my_data1.db
Done.
```

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

2015 Launch Records

```
%sql SELECT substr(Date, 6, 2) AS Month, Landing_Outcome, Booster_Version, Launch_Site FROM SPACEXTABLE
WHERE substr(Date, 0, 5) = '2015' AND Landing_Outcome like 'Failure%'
```

```
 * sqlite:///my_data1.db
Done.
```

Month	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

Task 10

List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.

Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

```
%sql SELECT Landing_Outcome, COUNT(Landing_Outcome) AS landing_outcomes FROM SPACEXTABLE
WHERE (Date BETWEEN '2010-06-04' AND '2017-03-20')
AND (Landing_Outcome = 'Failure (drone ship)' OR Landing_Outcome = 'Success (ground pad)') GROUP BY Landing_Outcome Order BY landing_outcomes DESC
```

```
 * sqlite:///my_data1.db
Done.
```

Landing_Outcome	landing_outcomes
Failure (drone ship)	5
Success (ground pad)	3

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.

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 .

Thank you!