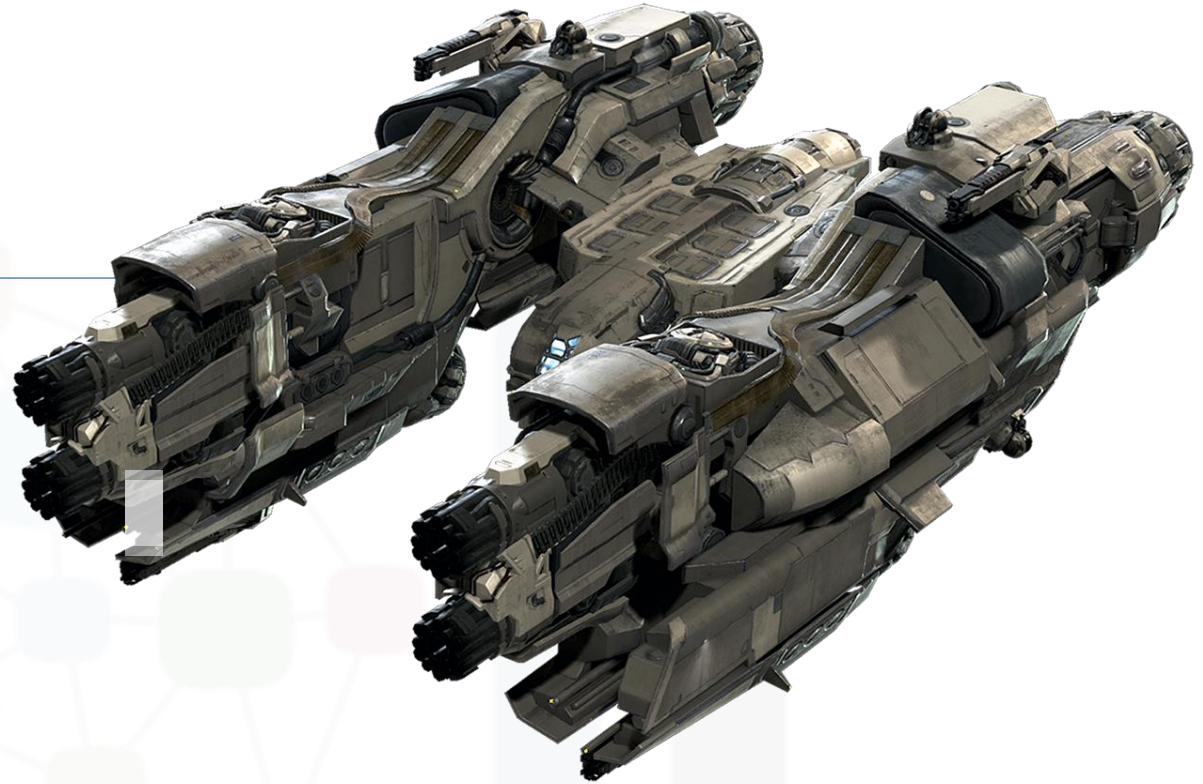
SPACEY

The Race To Space



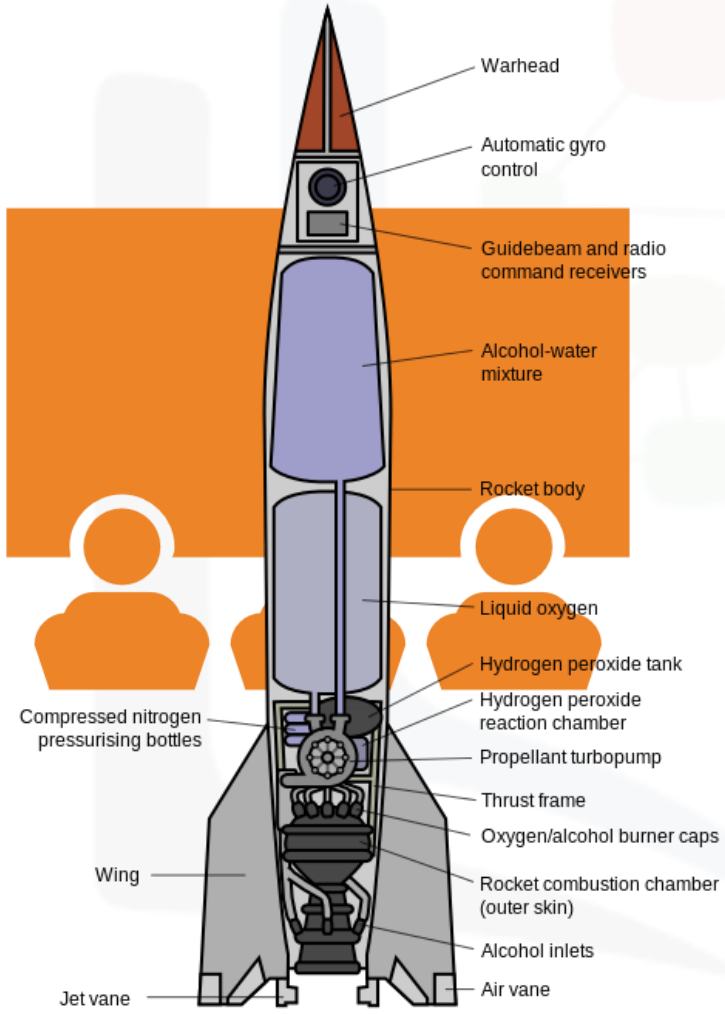
Frank Zimbello (Data Scientist)

05/31/2023

IBM Developer

SKILLS NETWORK

CONTENTS



EXECUTIVE SUMMARY

- **The commercial space age is here and SpaceY would like to compete with SpaceX and others in the race to make space travel affordable for everyone.**
- **SpaceX is the best in this space because the rocket launches are relatively inexpensive**
- **If Space Y can understand seize their advantages, they can properly compete with SpaceX**

Methodologies

- Data Collection (API and Web Scraping)
- Data Wrangling
- Exploratory Data Analysis(SQL, Pandas, Matplotlib
- Interactive Visualization (Folium, Plotly Dash)
- Predictive Analytics

Results

- Best Parameters
- Best Predictive Model

INTRODUCTION

- The commercial space age is here, companies are making space travel affordable for everyone.
  - SpaceX – The most successful
  - Virgin Galactic is providing suborbital spaceflights.
  - Rocket Lab is a small satellite provider.
  - Blue Origin manufactures sub-orbital and orbital reusable rockets
- SpaceX advertises Falcon 9 rocket launches on its 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
- If SpaceY can determine how to obtain SpaceX's savings and launch success, SpaceY can be major competitor in this space.

INTRODUCTION

- Challenges
  - The first stage does not land sometimes
  - The first stage will crash sometimes
  - The first stage is sometimes sacrificed due to the mission parameters
    - Payload
    - Orbit
    - Customer

INTRODUCTION

- Questions for Analysis
  - Which site has the most successful launches?
  - Which site has the highest launch success rate?
  - Which payload range(s) has the highest launch success rate?
  - Which payload range(s) has the lowest launch success rate?
  - Which F9 Booster version (v1.0, v1.1, FT, B4, B5, etc.) has the highest launch success rate?

METHODOLOGY

- Data Collection
- Data Wrangling
- Exploratory Data Analysis(EDA) with SQL & Visualization
- Interact Visual Analysis with Folium and Plotly Dash
- Predictive Analysis

METHODOLOGY

Data Collection

- Request and parse the SpaceX launch data using the SpaceX API
  - Clean the requested data
  - Decode and Normalize the data
  - Gather relevant data
    - From the rocket get the booster name
    - From the payload get the mass and orbit
    - From the launchpad get launch site information
    - From cores get the outcome of the landing, type, number of flights, whether gridfins were used, whether the core is reused, whether legs were used, the landing pad used, the block of the core, number of times core has been reused, and the serial of the core
  - Save the filtered dataset

METHODOLOGY

Data Collection

- Request and parse the SpaceX launch data using Web Scraping of SpaceX Wiki
  - Clean the requested data
  - Filter the data to only include Falcon 9 launches and relevant information
  - Remove the Falcon 1 launches keeping only the Falcon 9 launches
  - Filter the data using the BoosterVersion column to only keep the Falcon 9 launches
  - Save the filtered dataset

METHODOLOGY

- Data Wrangling
  - Addressing missing values
    - Address rows missing values
    - Address NA and Nones values in our dataset.
    - Use counts, means, and other values as needed when warranted to represent the data properly
  - Calculate the number of launches on each site
  - Calculate the number and occurrence of each orbit
  - Calculate the number and occurrence of mission outcome per orbit type
  - Create a landing outcome label from Outcome column
  - Save the Dataset

METHODOLOGY

- Data Wrangling
  - Addressing missing values
    - Address rows missing values
    - Address NA and Nones values in our dataset.
    - Use counts, means, and other values as needed when warranted to represent the data properly
  - Calculate the number of launches on each site
  - Calculate the number and occurrence of each orbit
  - Calculate the number and occurrence of mission outcome per orbit type
  - Create a landing outcome label from Outcome column
  - Save the Dataset

METHODOLOGY

- Exploratory Data Analysis(EDA) with SQL & Visualization
  - 4 Sites to be analyzed -

Display the names of the unique launch sites in the space mission

```
%sql select DISTINCT(Launch_Site) from SPACEXTBL;
```

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

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

METHODOLOGY

- Exploratory Data Analysis(EDA) with SQL & Visualization
  - All Successful Landings to be analyzed -

List the total number of successful and failure mission outcomes

```
%sql SELECT Mission_Outcome, count(Mission_Outcome) FROM SPACEXTBL
```

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

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

METHODOLOGY

- Exploratory Data Analysis(EDA) with SQL & Visualization
  - Use SQL queries and various charts for each scenario below to address questions (mainly catplot and scatterplot used)
    - Flight Numbers vs Launch Sites
    - Payload Mass Vs. Launch Sites
    - Sucess Rate vs Orbit
    - Orbit vs Payload Mass
    - Year vs Success rate

METHODOLOGY

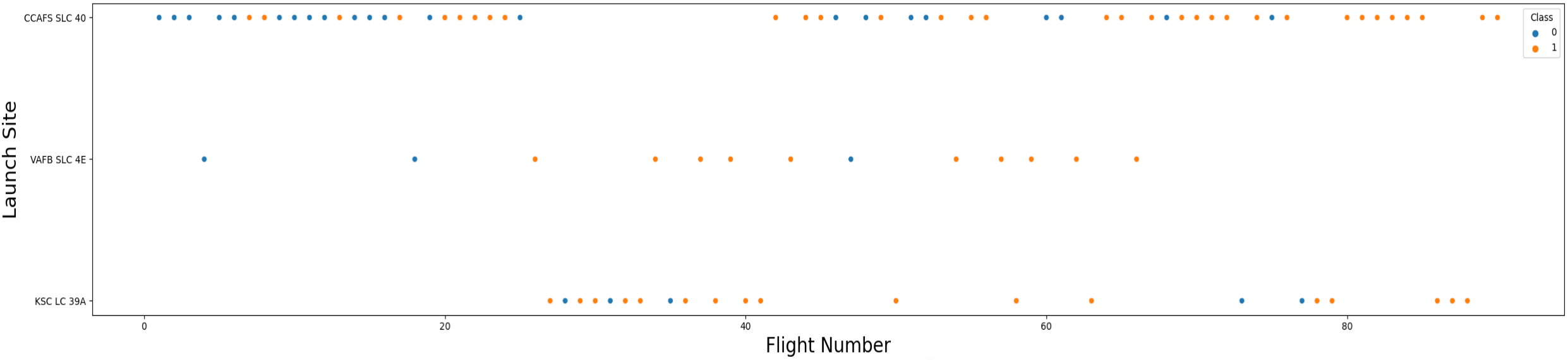
- Predictive Analysis
  - Multiple methods used to find most accurate model for successful launches
    - Decision tree method
    - K nearest neighbors method
    - Logistics Regression method
    - Support Vector Machine method

DISCUSSION & RESULTS

- The data was collected from through both Web scraping and rest API
- The data was cleaned and missing values were replaced or removed and Wrangled as needed
- The data was then analyzed using SQL, Scatterplots, Bar Chart, Maps, Pie Charts, and Dashboards Interactive Dashboards
- Predictive Analysis was used to answer questions and overcome challenges to entry and competition

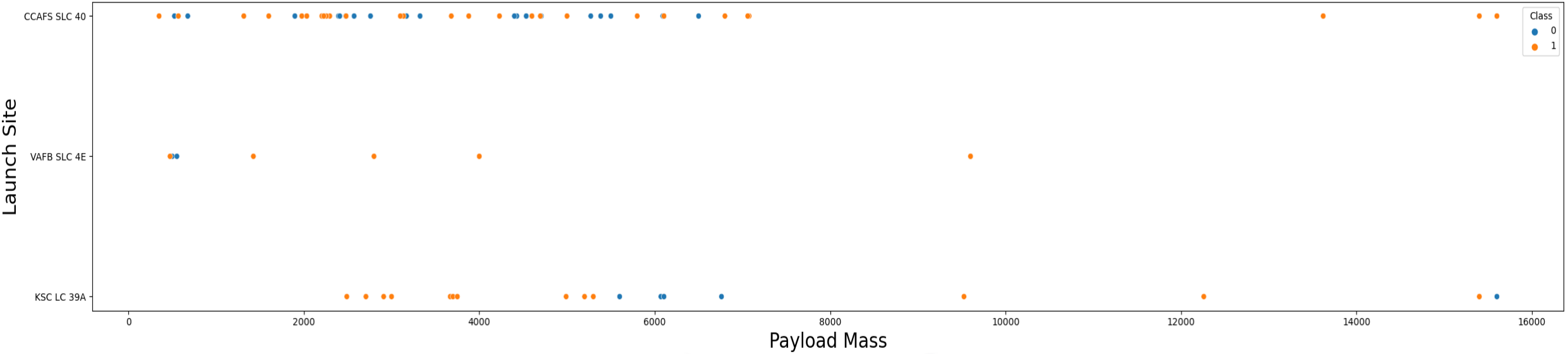
DISCUSSION & RESULTS

- Exploratory Data Analysis(EDA) with SQL & Visualization
  - Flight Numbers vs Launch Sites
    - One Flight only at a given time from only one site



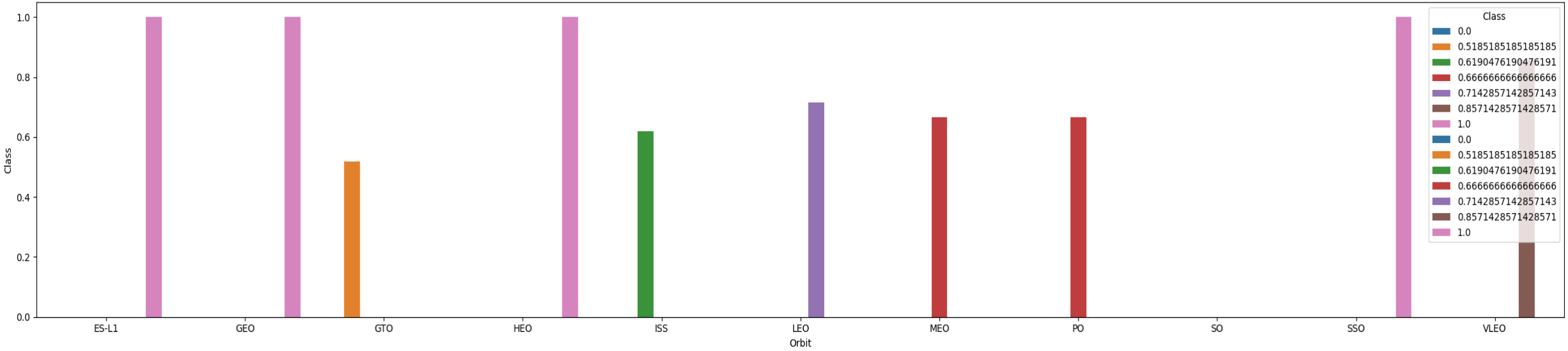
DISCUSSION & RESULTS

- Exploratory Data Analysis(EDA) with SQL & Visualization
  - Payload Mass Vs. Launch Sites
    - VAFB-SLC launch site has no rockets launched for heavy payload mass (greater than 10000).



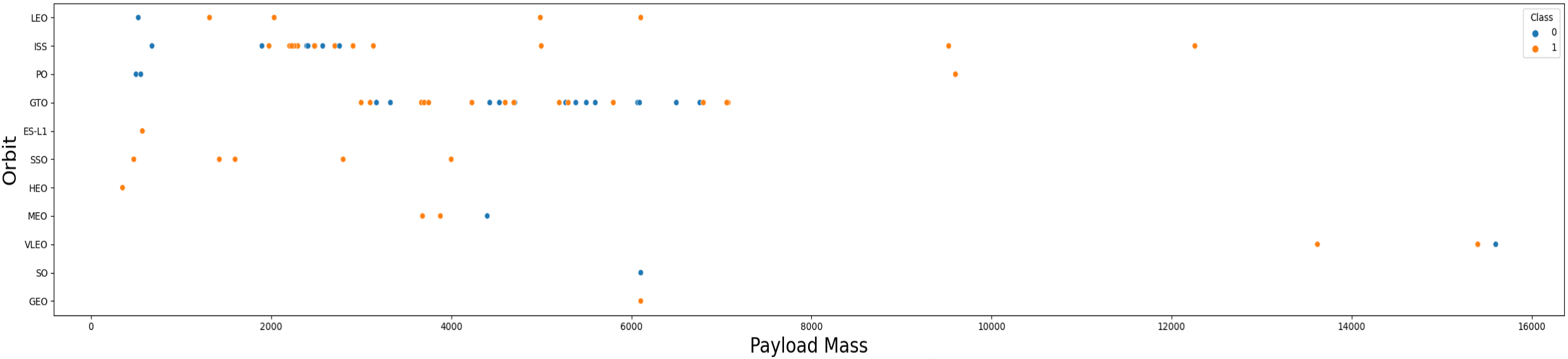
DISCUSSION & RESULTS

- Exploratory Data Analysis(EDA) with SQL & Visualization
  - Sucess Rate vs Orbit
    - LEO orbit success appears related to the number of flights but there seems to be no relationship between flight number when in GTO orbit



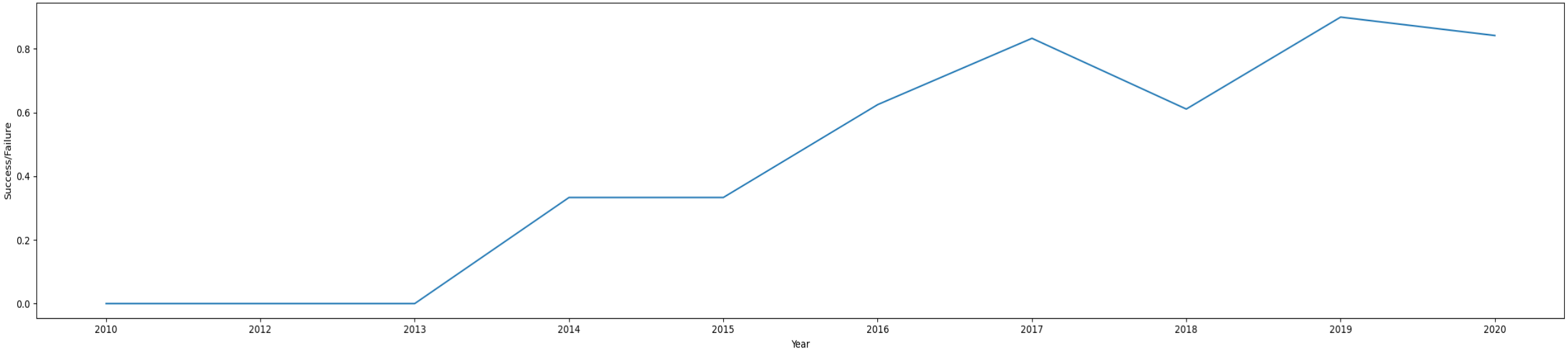
DISCUSSION & RESULTS

- Exploratory Data Analysis(EDA) with SQL & Visualization
  - Orbit vs Payload Mass
    - 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



DISCUSSION & RESULTS

- Exploratory Data Analysis(EDA) with SQL & Visualization
  - Year vs Success rate
    - From 2013 success rates have greatly increased until 2020



DISCUSSION & RESULTS

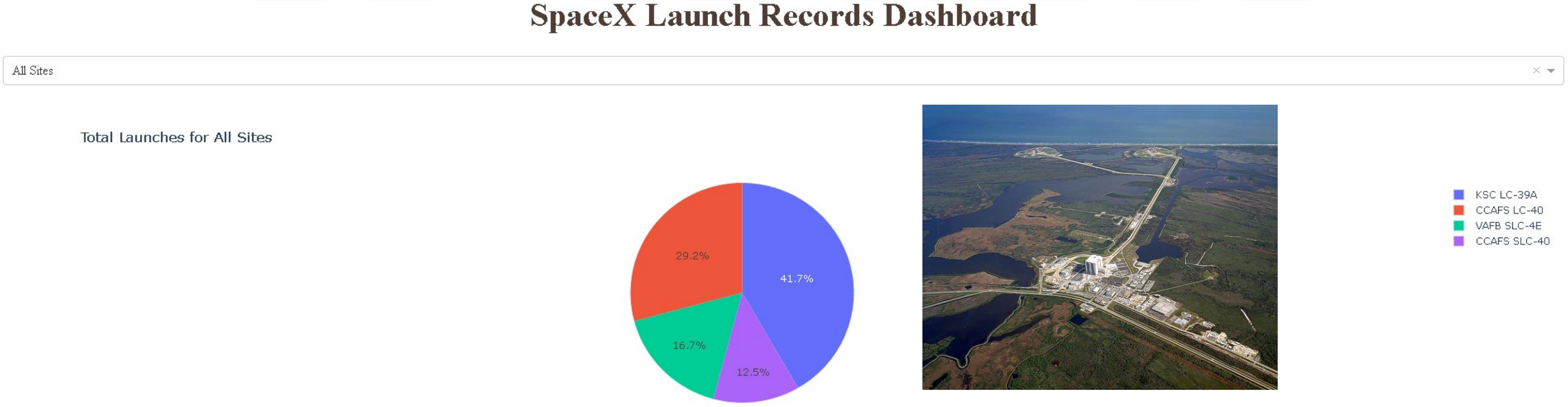
- Interact Visual Analysis with Folium and Plotly Dash
  - All Launch sites are in proximity to the Equator line
    - This will mean less fuel needed due to gravity and the Earth's rotation near the equator
  - All launch sites in very close proximity to the coast
    - For safety reasons in case of an explosion or crash, etc.

DISCUSSION & RESULTS

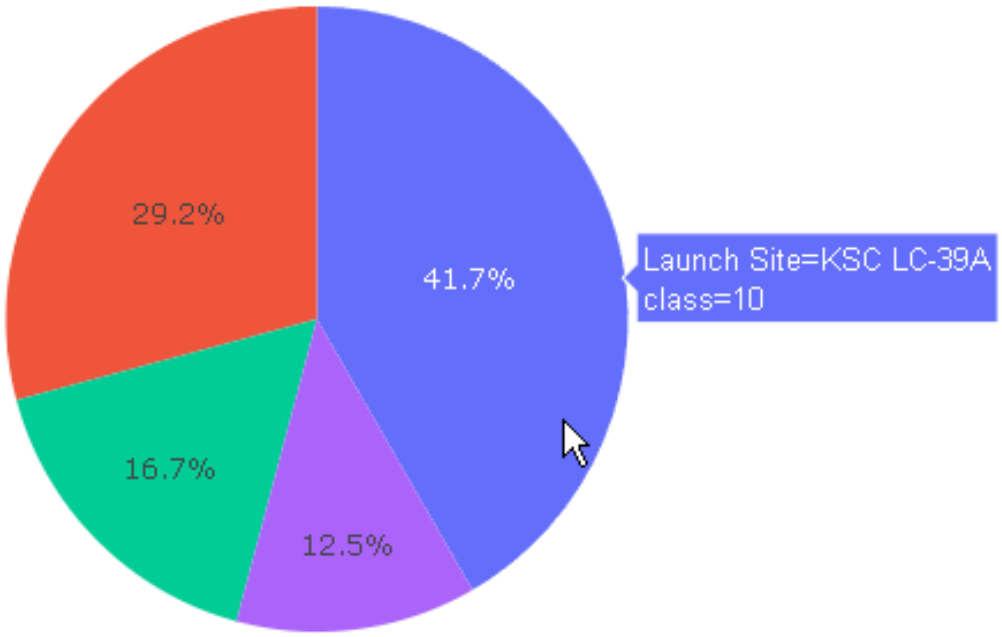
- Interact Visual Analysis with Folium and Plotly Dash
  - All launch sites in close proximity to railways
    - Rockets, parts, equipment, people, etc. needs to be moved to and from the launch site
  - All launch sites in close proximity to highways
    - Rockets, parts, equipment, people, etc. needs to be moved to and from the launch site
  - All launch sites are distant from cities
    - For safety if there are any crash landings, explosions, etc.

DISCUSSION & RESULTS

- Interact Visual Analysis with Folium and Plotly Dash
  - Kennedy Space Center Launch Complex 39 appears to have the most successful launches

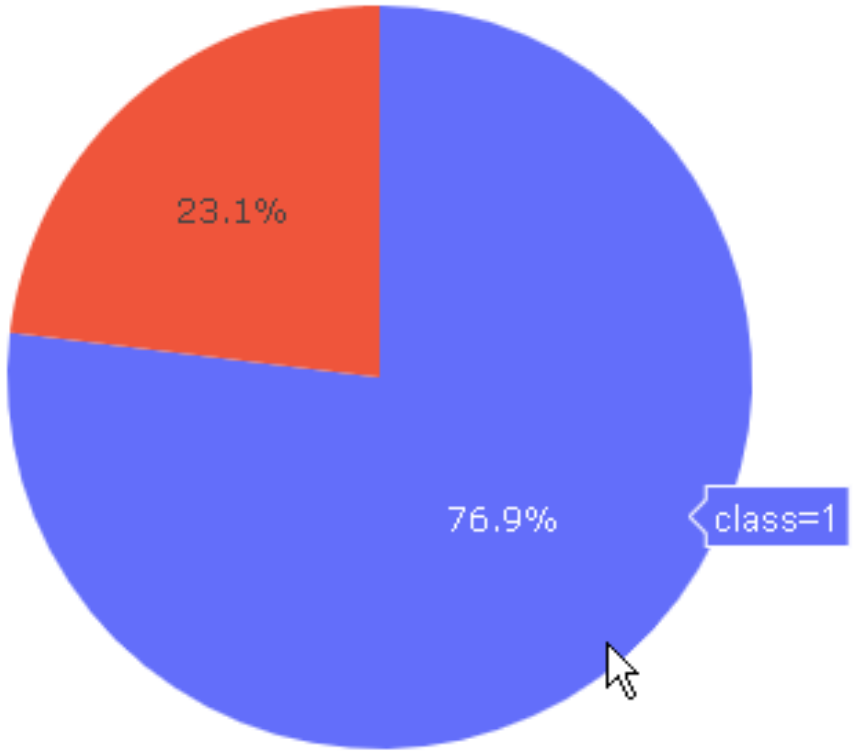


DISCUSSION & RESULTS



- Which site has the most successful launches?
  - Kennedy Space Center Launch Complex 39 appears to have the most successful launches with
    - 10 successful launches
    - Accounts for 41.7% of the successful launches

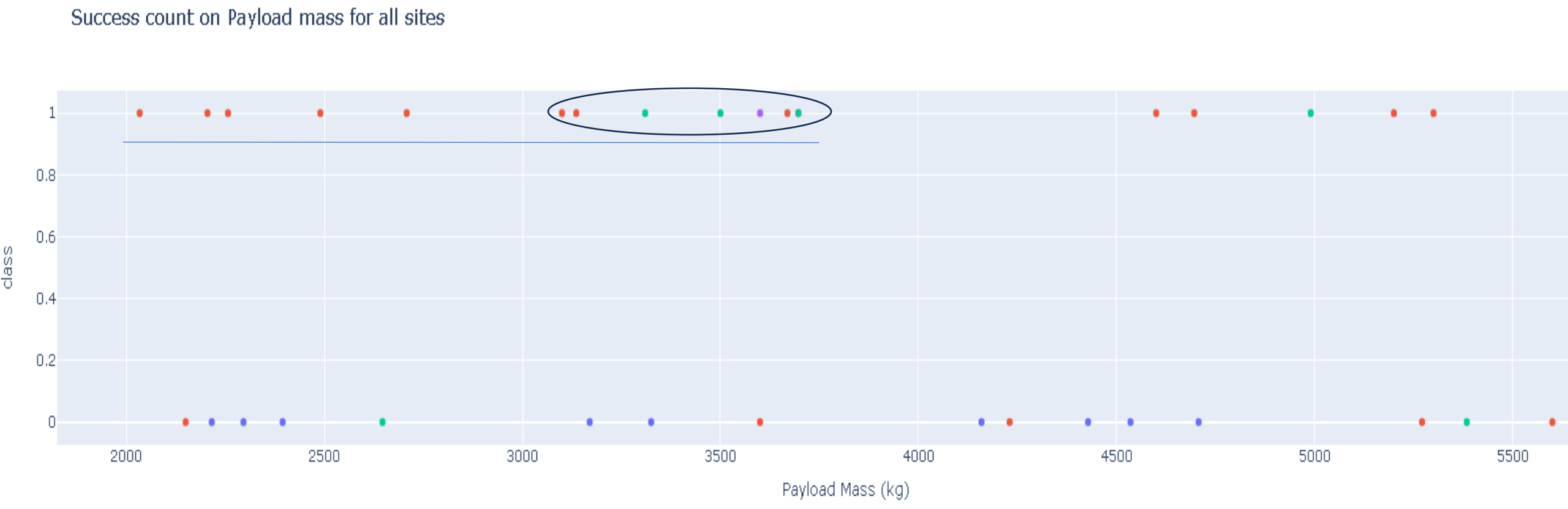
DISCUSSION & RESULTS



- Which site has the highest launch success rate?
  - Kennedy Space Center Launch Complex 39 appears to also have the highest launch success rate
    - Success Rate of 76.9% compared to next highest of less than 43%

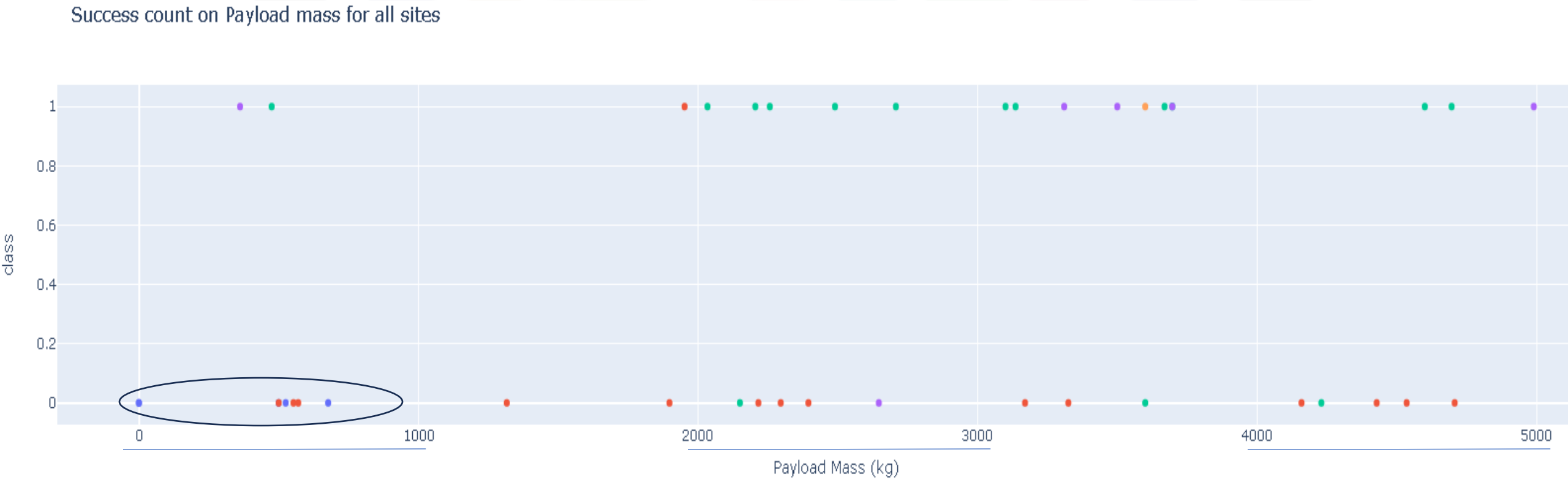
DISCUSSION & RESULTS

- Which payload range(s) has the highest launch success rate?
  - The range of 2k – 4k appears to have the most success
  - More specifically between over 3K and under 4k has the highest success rate



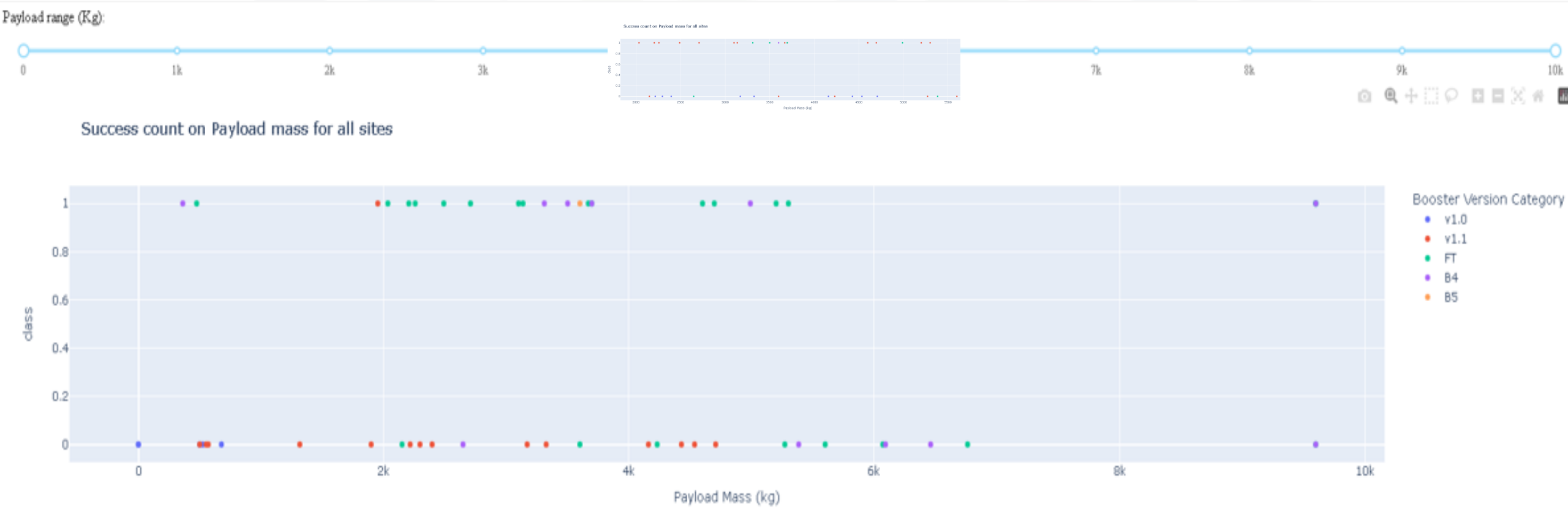
DISCUSSION & RESULTS

- Which payload range(s) has the lowest launch success rate?
  - The ranges of 0-1K, 2k–3k, and 4k-5k all appear to have low success
  - More specifically between over 0 and under 1k has the lowest success rate



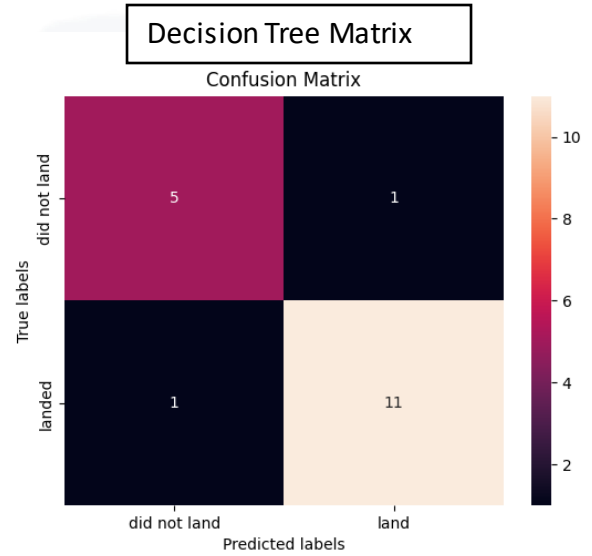
DISCUSSION & RESULTS

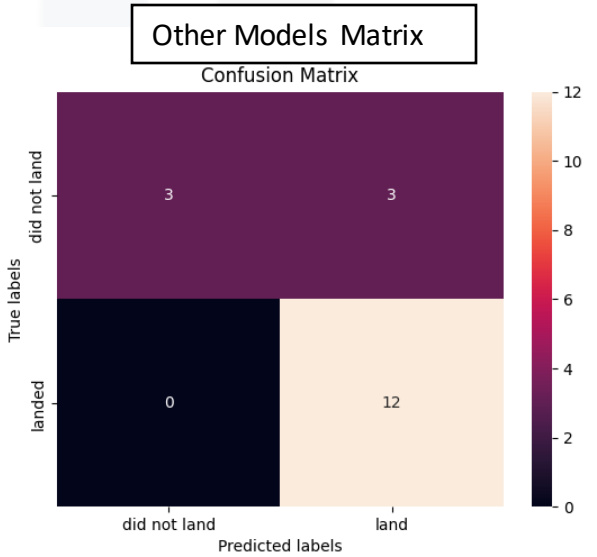
- Which F9 Booster version (v1.0, v1.1, FT, B4, B5, etc.) has the highest launch success rate?
  - The F9 Booster version FT had 13 of 21 success and twice that of the next highest B4 Booster version



DISCUSSION & RESULTS

- Predictive Analysis
  - Multiple machine learning methods and models were use. Train and Test sets were used and Hyperparameters were tuned. The Decision tree model showed the most accuracy in our analysis and we may want to use that moving forward to determine future launch success given all the factors we have reviewed and used above.
    - Accuracy for Decision tree method: 88.88%
    - Accuracy for K nearest neighbors method: 86.11%
    - Accuracy for Logistics Regression method: 83.33%
    - Accuracy for Support Vector Machine method: 83.33%
  - You can also see the Confusion matrix showing better accuracy for the Tree as True Positives + True Negatives are higher and False Positives + False Negatives are lower with the Decision tree Confusion Matrix than the other Models





OVERALL FINDINGS & IMPLICATIONS

Findings

- Kennedy Space Center Launch Complex 39 has the most successful launches and highest success rate
- Cape Canaveral Force Station Launch Complex 40 is the second best site to launch from

Implications

- Launch from this site whenever possible for highest chance of success
- When KSC above is not available or if wanting to do multiple flights at same time, this is the second best option overall

OVERALL FINDINGS & IMPLICATIONS

Findings

- The F9 Booster version FT has the highest success rate The F9
- The F9 Booster version B4 Booster version is the next most successful booster version

Implications

- This booster version should be used whenever available
- This booster version should be used whenever the FT is notavailable

OVERALL FINDINGS & IMPLICATIONS

Findings

- Payload Mass 3k-4k Kg is optimal for success and also greatly decreases the failure rate
- The FT Booster version can help mitigate failure in the 2-3k Kg Payload Mass bracket

Implications

- Managing payload to stay withing that mass range will highly improve success rates
- If you cannot manage to the above mass be sure to stay above 2K and use the FT booster to help ensure success

OVERALL FINDINGS & IMPLICATIONS

Findings

- It is probably best to avoid the VAFB SLC due unless heavy payload mass ear 10K Kg

Implications

- This site appears to be the only one with success near 10K payload but otherwise not very successful

CONCLUSION

- The commercial space age is here and SpaceY would like to compete with SpaceX and others in the race to make space travel affordable for everyone.
- If SpaceY can determine how to obtain SpaceX's savings and launch success, SpaceY can be major competitor in this space.

CONCLUSION

- SpaceY can compete with SpaceX in this space base on our findings
  - Our Data Analysis has revealed several factors for success
    - Choose launch sites with highest success rate
      - KSC SLC39 (CAFS SLC40 if not available)
    - Choose F9 Booster version with most success
      - FT Booster version (B4 if not available)
    - Manage Payload properly to help guarantee success
      - 3000-4000 Kg Payload optimal for success
      - 2000-3000 Kg Payload (used FT Booster)
      - Avoid under 2K or over 4 or 5k payloads
      - Use VAFB SLC if heavy payload necessary(10K)

CONCLUSION

- The Analysis was succesful and provided data to allow SpaceY to compete in this space
- SpaceY should look at not only the major findings but findings as they may increase their business and competiveness in this space

CONCLUSION

- Further recommendations
  - SpaceX only launches one flight at a time so if there is demand and opportunity SpaceY may look at launching mutliple flights further market share and revenue opportunities
  - SpaceX looks like their success has had a downturn from 2019 to 2020 and SpaceY may be able to capitalize on this using the factors of this analysis to move ahead

APPENDIX

- Resources
  - Coursera and Coursera courses
  - IBM Instructors and Network Skills Labs
  - SpaceX Wikipedia for Web Scraping
  - SpaceX API for needed launch data

APPENDIX

- Acknowledgements
  - IBM Company – Producers of the course, labs, network skills environment
    - Yan Luo – Instructor - Ph.D., Data Scientist and Developer for IBM
    - Joseph Santarcangelo – Instructor - Ph.D., Data Scientist at IBM
  - Coursera – course providers and online education service