Winning Space Race with Data Science

Zheyuan Hu
2023/3/14

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection (Web Scraping & API)
  - Exploratory Data Analysis (Data Wrangling, Visualization, Dashboard)
  - Predictive Analysis (Machine Learning)

- Summary of all results
  - Successfully collected data from public sources
  - Feature engineering and interactive dashboard
  - Mathematical model evaluation and comparison

Introduction

- Project background and context

Space X reused the first stage during Falcon 9 launches which saved large amount of expenses. By analyzing their spacecraft launching data, we want to obtain valuable insights about Space X launches and help another company Space Y to join the competition.

- Problems you want to find answers

- What are some key factors that lead to a successful landing?
- How do these factors affect the success rate?

Section 1

Methodology

Methodology

Executive Summary

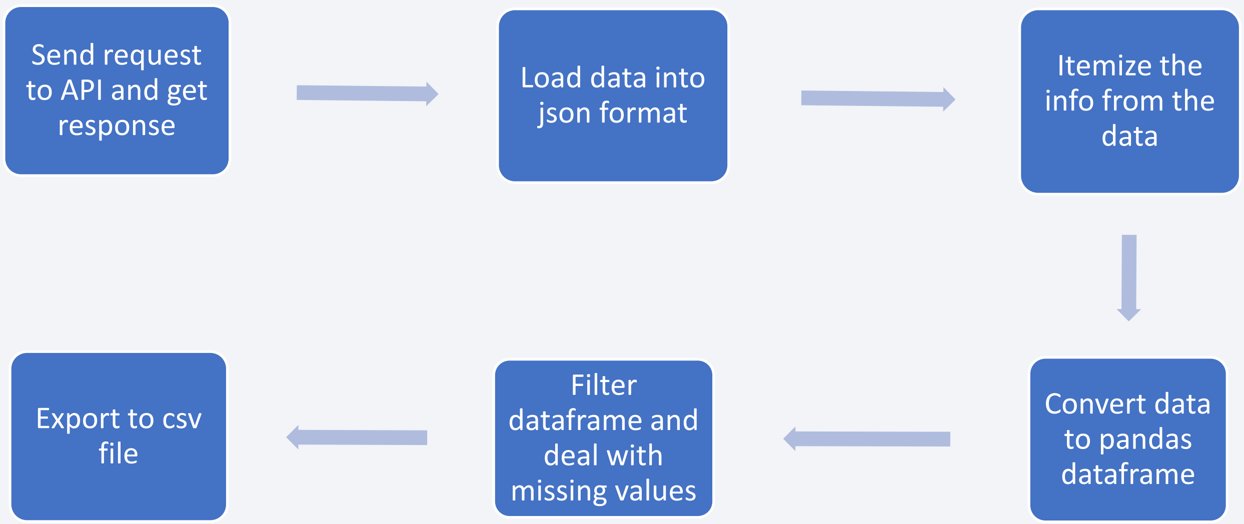
- Data collection methodology:
  - SpaceX API & Web Scraping
- Perform data wrangling
  - Data Preprocessing & Data Categorization
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Normalization & Various Model Fitting & Parameter Hypertuning

Data Collection

- Data was collected from Space X API (https://api.spacexdata.com/v4/rockets/)

- Another data source was web-scraped from Wikipedia (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches)

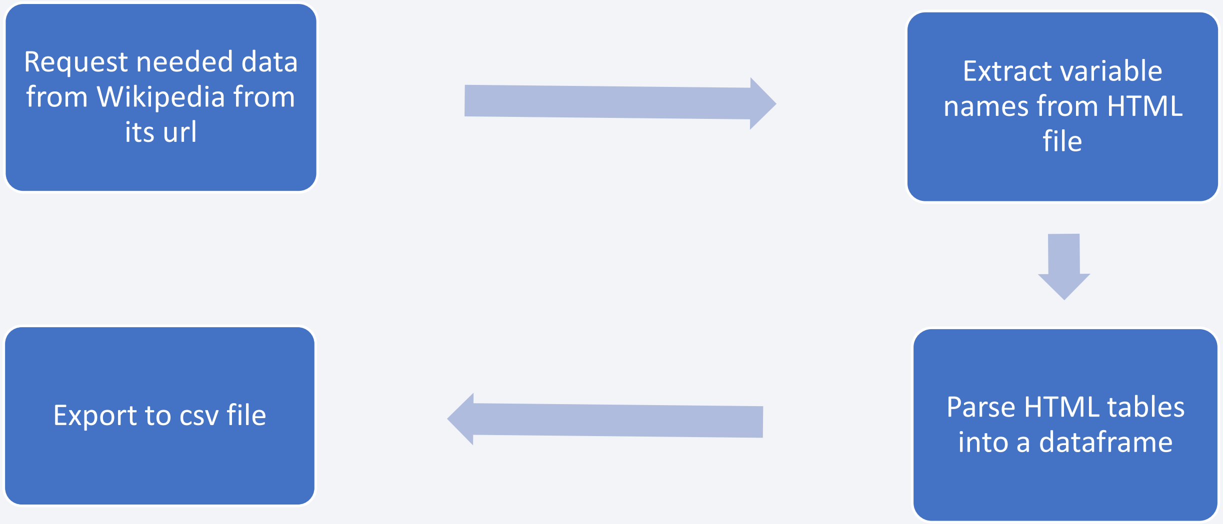
Data Collection – SpaceX API



Link to code:

https://github.com/jouran/IBM-Data-Science-Capstone/blob/2b82f629e8a273e1c2466625869f713e94c808ed/jupyter-labs-spacex-data-collection-api.ipynb

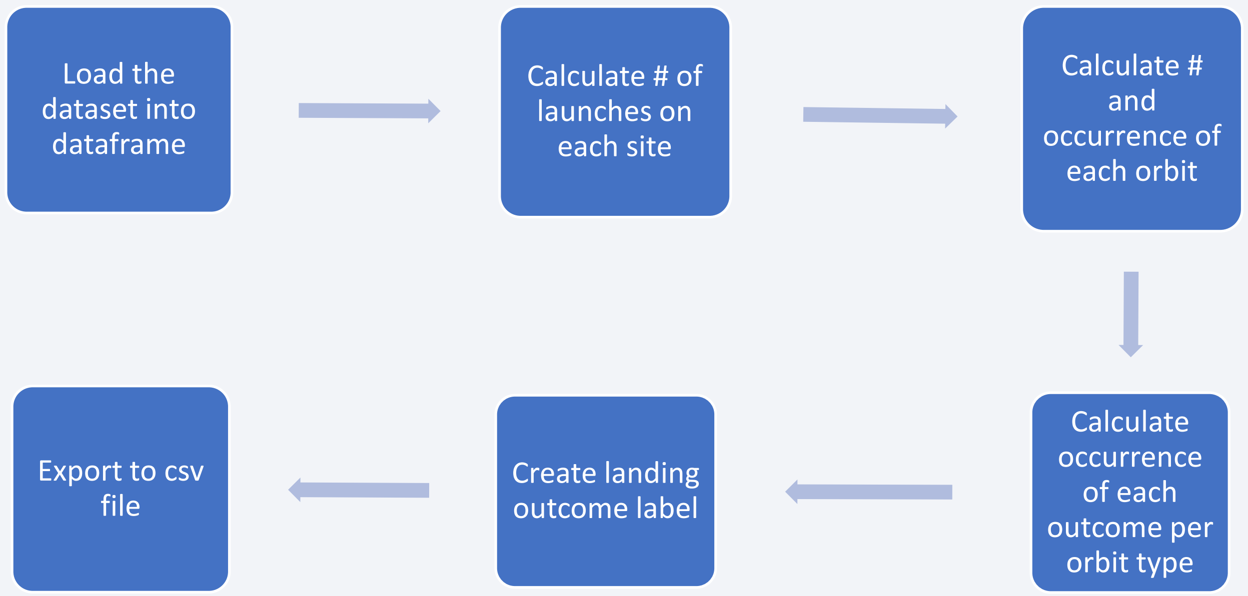
Data Collection - Scraping



Link to code:

https://github.com/jouran/IBM-Data-Science-Capstone/blob/2b82f629e8a273e1c2466625869f713e94c808ed/jupyter-labs-webscraping.ipynb

Data Wrangling



Link to code:

https://github.com/jouran/IBM-Data-Science-Capstone/blob/2b82f629e8a273e1c2466625869f713e94c808ed/IBM-DS0321EN-SkillsNetwork_labs_module_1_L3_labs-jupyter-spacex-data_wrangling_jupyterlite.jupyterlite.ipynb

EDA with Data Visualization

- Scatter plots (Flight Number vs. Payload Mass, Flight Number vs. Launch Site, Payload Mass vs. Launch Site, Flight Number vs. Orbit Type, Payload Mass vs. Orbit Type) show relationship between two variables.
- Bar chart (Orbit Type vs. Success Rate) shows comparison between multiple variables.
- Line graph (Success Rate vs. Year) shows trend.

Link to code:

https://github.com/jouran/IBM-Data-Science-Capstone/blob/2b82f629e8a273e1c2466625869f713e94c808ed/IBM-DS0321EN-SkillsNetwork_labs_module_2_jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

EDA with SQL

- Displaying the names of the unique launch sites in the space mission
- Displaying 5 records where launch sites begin with the string ‘CCA’
- Displaying the total payload mass carried by boosters launched by NASA (CRS)
- Displaying average payload mass carried by booster version F9 v1.1
- Listing the date when the first successful landing outcome in ground pad was achieved
- Listing the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- Listing the total number of successful and failure mission outcomes
- Listing the names of the booster versions which have carried the maximum payload mass
- Listing the failed landing outcomes in drone ship, their booster versions and launch site names for the months in year 2015
- Ranking 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

Link to code:

https://github.com/jouran/IBM-Data-Science-Capstone/blob/2b82f629e8a273e1c2466625869f713e94c808ed/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- I added markers with Circle, Popup Label and Text Label of all Launch Sites based on their latitude and longitude coordinates to show their geographical location.
- I added colored markers indicating success with green color and failure with red color.
- I used Marker Cluster to display multiple launches for the same site.
- I used markers to show distance between a selected launch site to a nearby coastline and a railway, and added a line between them along with the distance shown.
- By creating and marking on the interactive map, we are able to get a better understanding of the dataset. Visualizing them gives us a geological inspiration regarding potential future launch outcomes.

Link to code:

https://github.com/jouran/IBM-Data-Science-Capstone/blob/2b82f629e8a273e1c2466625869f713e94c808ed/IBM-DS0321EN-SkillsNetwork_labs_module_3_lab_jupyter_launch_site_location.jupyterlite.ipynb

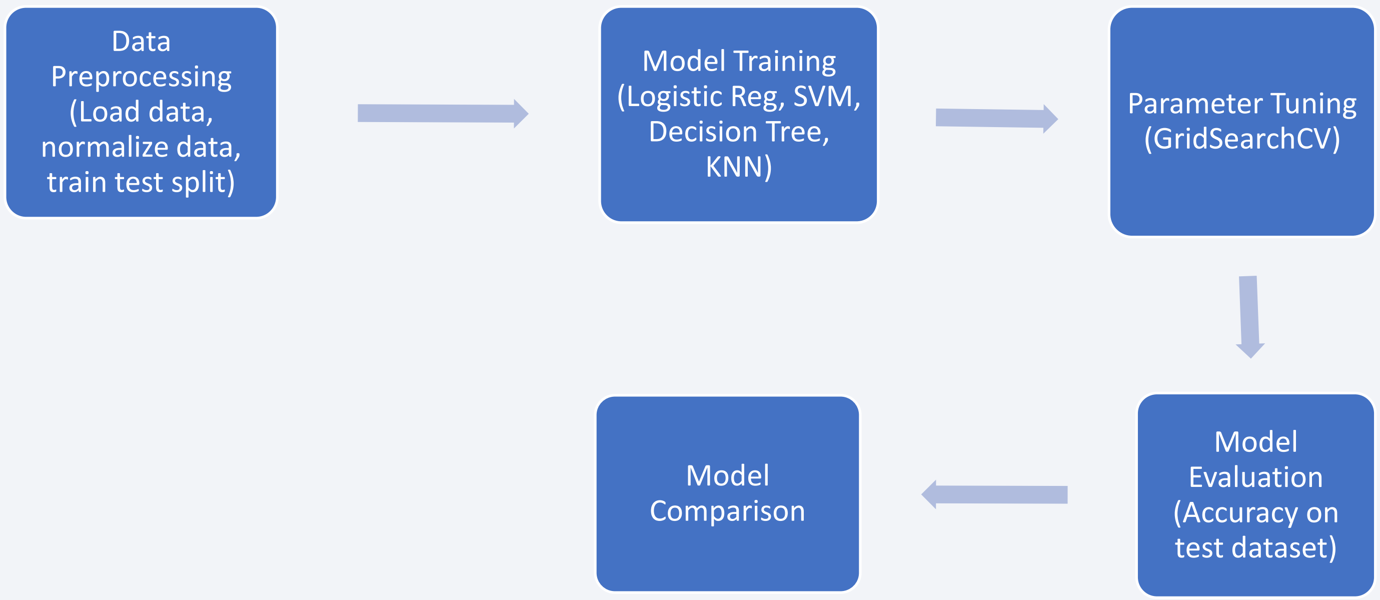
Build a Dashboard with Plotly Dash

- Graph 1: Pie Chart showing percentage of launches by site, with a dropdown that allows user to choose launch site
- Graph 2: Scatter Chart showing the relationship of successful outcome versus payload mass, with a slider that allows user to choose the range of payload mass.
- This dashboard grants flexibility and user-friendliness in exploring the relationships between payload mass, launch sites, and successful outcome. It could assist users to experiment and discover what helps achieve a successful launch.

Link to code:

https://github.com/jouran/IBM-Data-Science-Capstone/blob/2b82f629e8a273e1c2466625869f713e94c808ed/spacex_dash_app.py

Predictive Analysis (Classification)



Link to code:

https://github.com/jouran/IBM-Data-Science-Capstone/blob/2b82f629e8a273e1c2466625869f713e94c808ed/IBM-DS0321EN-SkillsNetwork_labs_module_4_SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

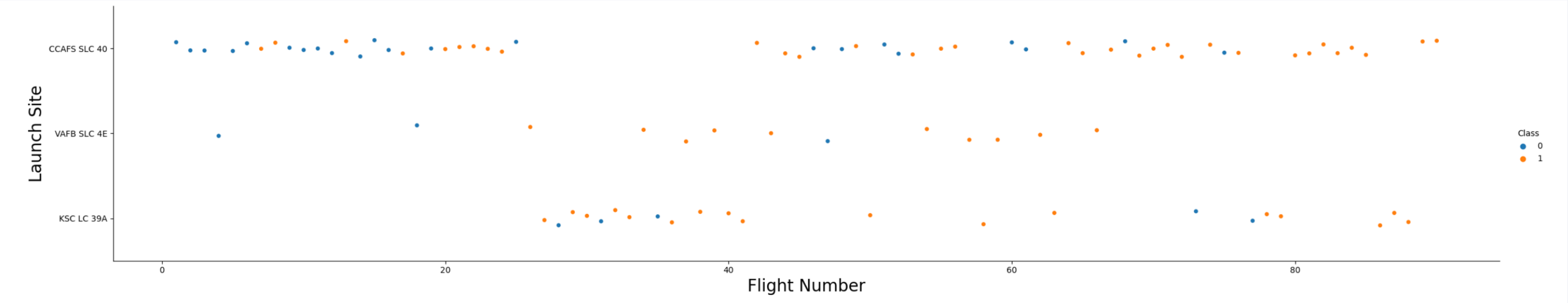
Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

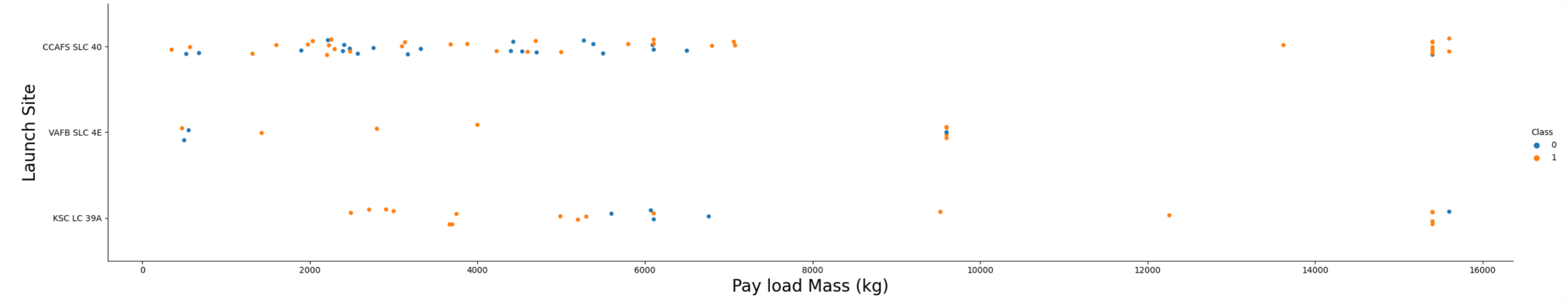
Insights drawn from EDA

Flight Number vs. Launch Site



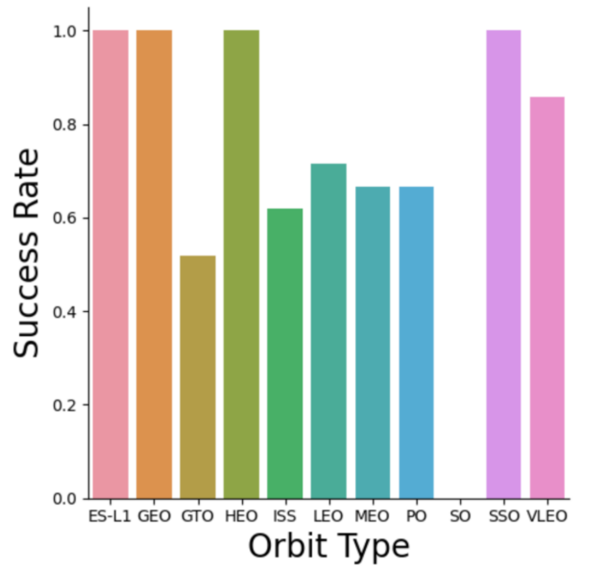
- First couple flights all failed and last few flights are all successful
- CCAFS SLC-40 site has way more flights than others
- Flight success rate appears to be growing over time

Payload vs. Launch Site



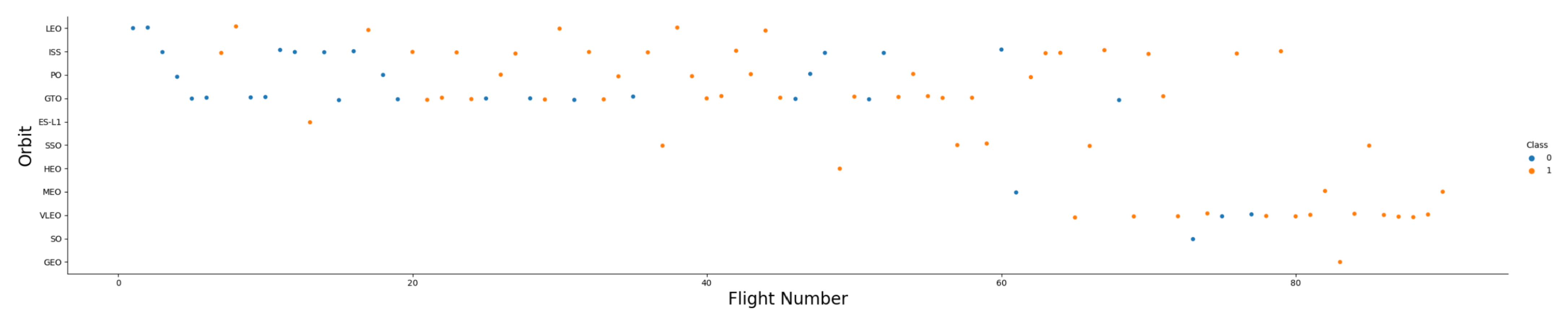
- KSC LC-39A is more likely to succeed if payload is less than 5500 kg
- For all sites, payload mass larger than 9000 kg will be more likely to succeed

Success Rate vs. Orbit Type



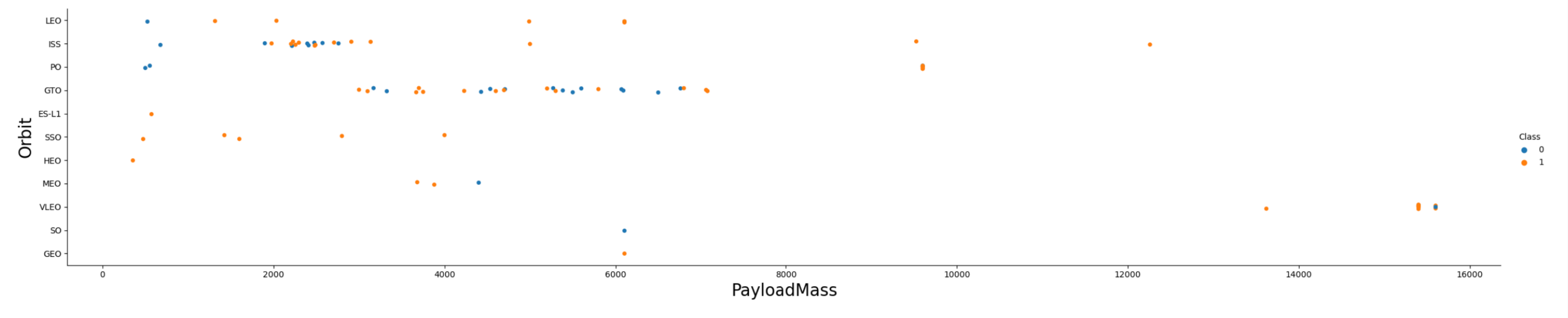
- ES-L1, GEO, HEO, SSO have 100% success rate
- SO has 0% success rate
- All other orbits have at least 50% success rate

Flight Number vs. Orbit Type



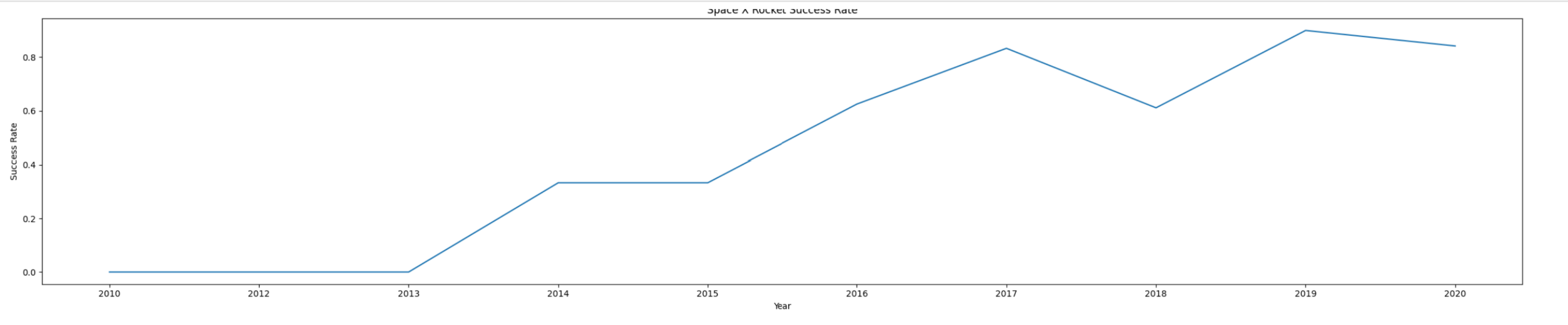
- It appears that the success rate increases over time for nearly all orbit

Payload vs. Orbit Type



- If larger payload mass, LEO, ISS, and PO will be more likely to succeed
- If smaller payload mass, SSO and MEO will be more likely to succeed

Launch Success Yearly Trend



- The success rate started to increase since 2013
- It can be predicted that the success rate will keep increasing in the long term

All Launch Site Names

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

- Here is the query result showing 4 unique launch sites

Launch Site Names Begin with 'CCA'

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

- Here is the query result showing 5 records where launch site names begin with ‘CCA’

Total Payload Mass

SUM(PAYLOAD_MASS__KG_)
45596

- Here is the query result showing the total payload mass is 45596 kg

Average Payload Mass by F9 v1.1

AVG(PAYLOAD_MASS__KG_)
2928.4

- Here is the query result showing the average payload mass by F9 v1.1 is 2928.4 kg

First Successful Ground Landing Date

MIN(Date)
01-05-2017

- Here is the query result showing the first successful ground landing date is 2017-1-5

Successful Drone Ship Landing with Payload between 4000 and 6000

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

- Here is the query result showing the booster version which have successful drone ship landing with payload between 4000 and 6000

Total Number of Successful and Failure Mission Outcomes

Successful	Failure
100	1

- Here is the query result showing the number of successful and failure outcomes

Boosters Carried Maximum Payload

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

- Here is the query result showing the booster versions with maximum payload

2015 Launch Records

Month	Booster_Version	Launch_Site
01	F9 v1.1 B1012	CCAFS LC-40
04	F9 v1.1 B1015	CCAFS LC-40

- Here is the query result showing launch records in 2015 where landing outcome is failure

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

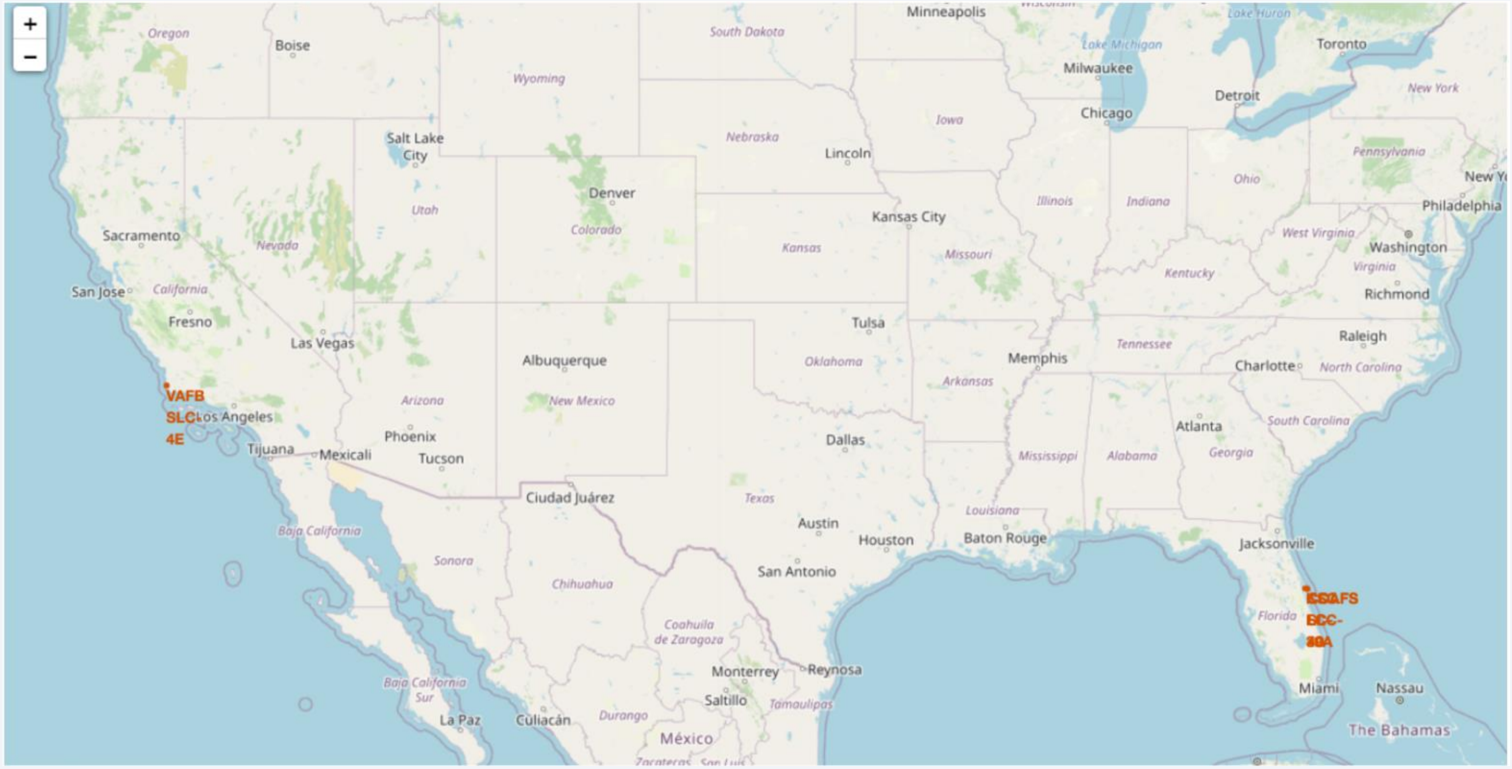
Landing _Outcome	COUNT("Landing _Outcome")
Success	20
Success (drone ship)	8
Success (ground pad)	6

- Here is the query result showing the ranking of landing outcomes between 2010-6-4 and 2017-3-20

Section 3

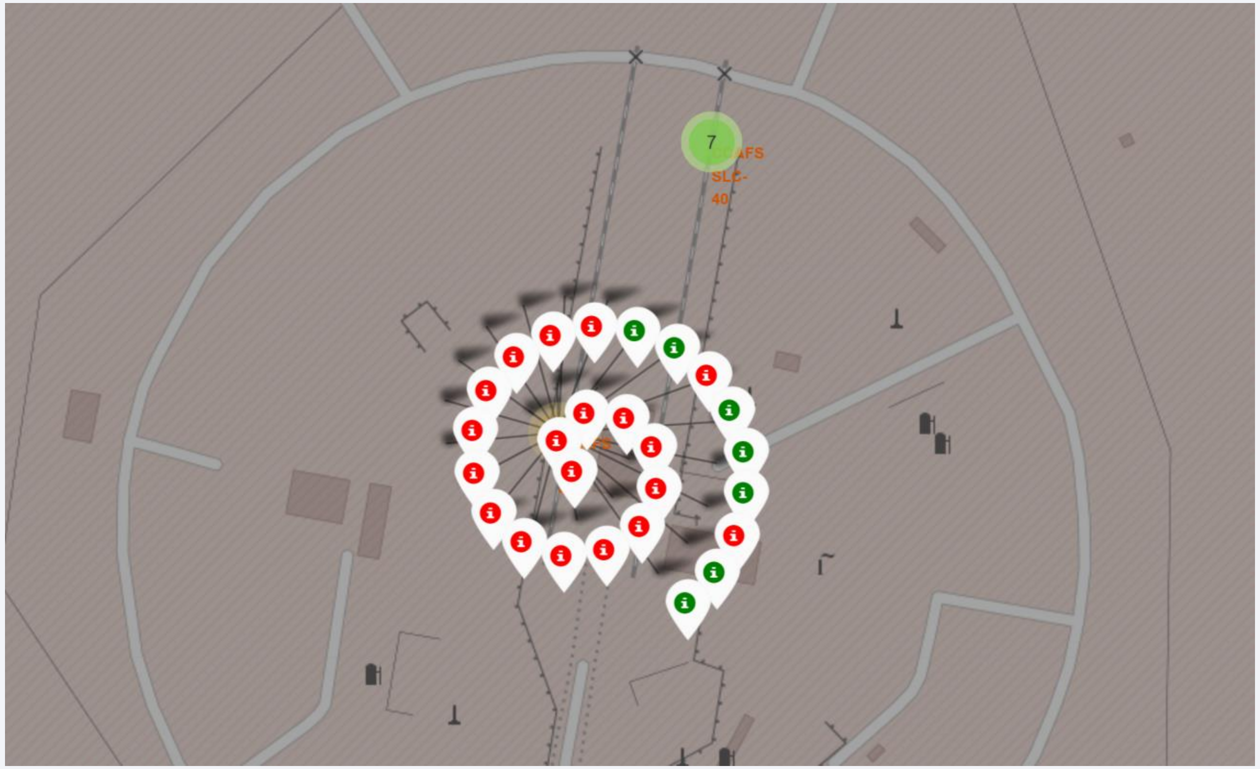
Launch Sites Proximities Analysis

All Launch Sites



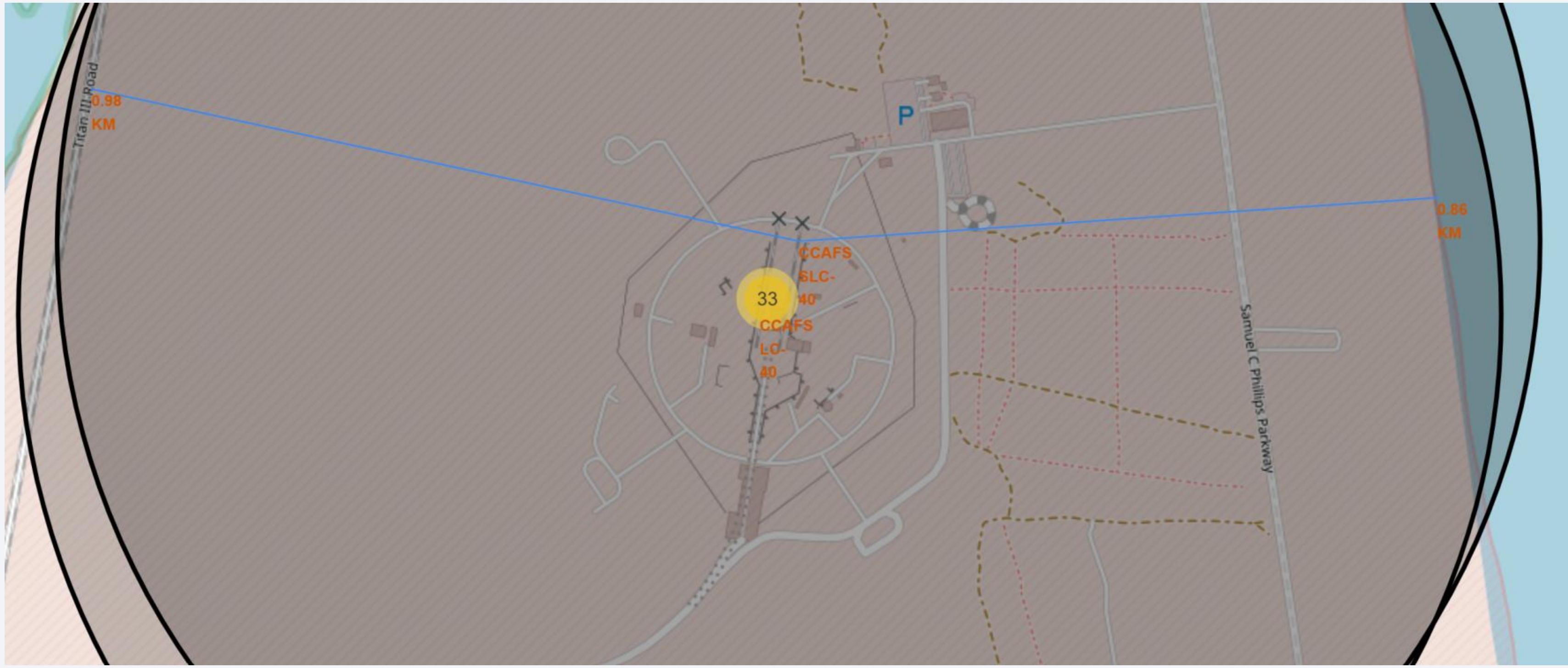
- All launch sites are relatively close to a major city
- All launch sites are near coastlines

Color-labeled Launch Outcomes



- Green marker indicates successful launch
- Red marker indicates failed launch

Proximities of Launch Sites

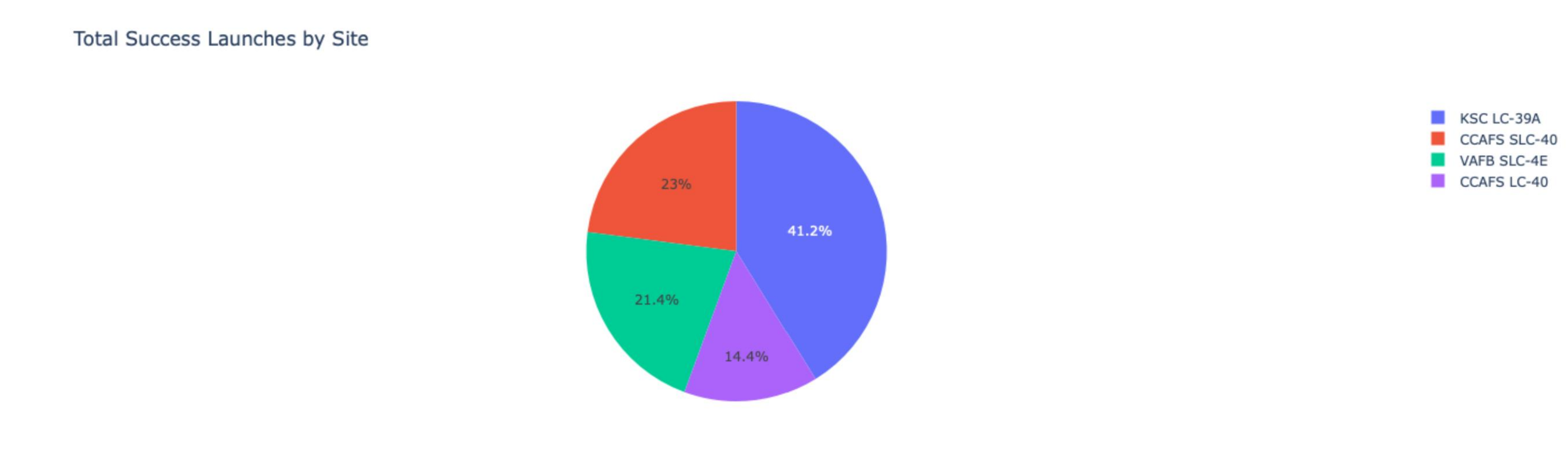


- This site has logistics advantage due to being close to railroad, highway, and coastline
- All launch sites are near coastlines

Section 4

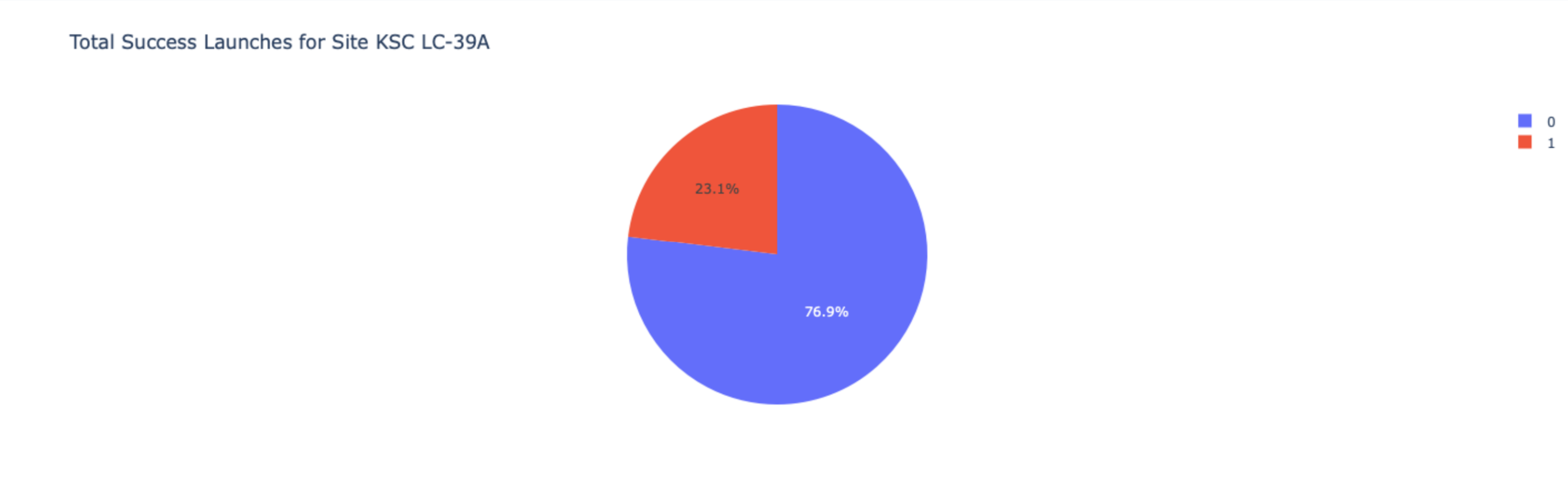
Build a Dashboard with Plotly Dash

Total Success Launches for All Sites



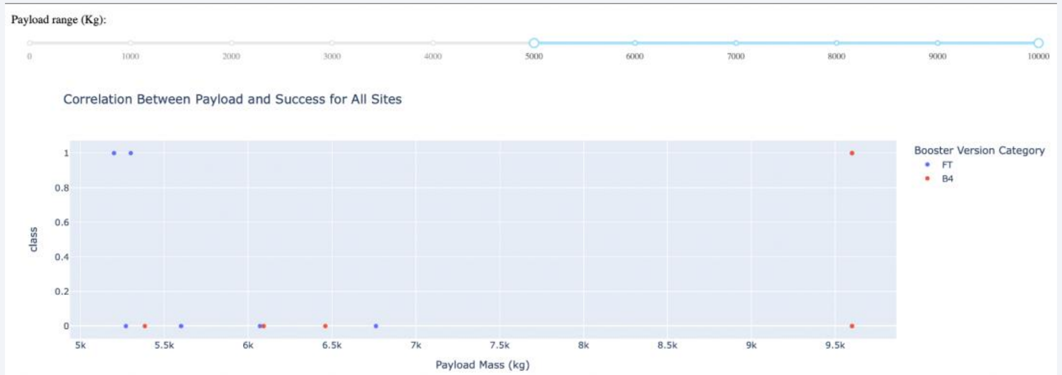
- KSC LC-39A has highest success launches
- CCAFS LC-40 has lowest success launches

Launch Site with Highest Success Ratio



- KSC LC-39A has over ¾ success rate which is quite high

Payload vs. Launch Outcome Scatter Plot

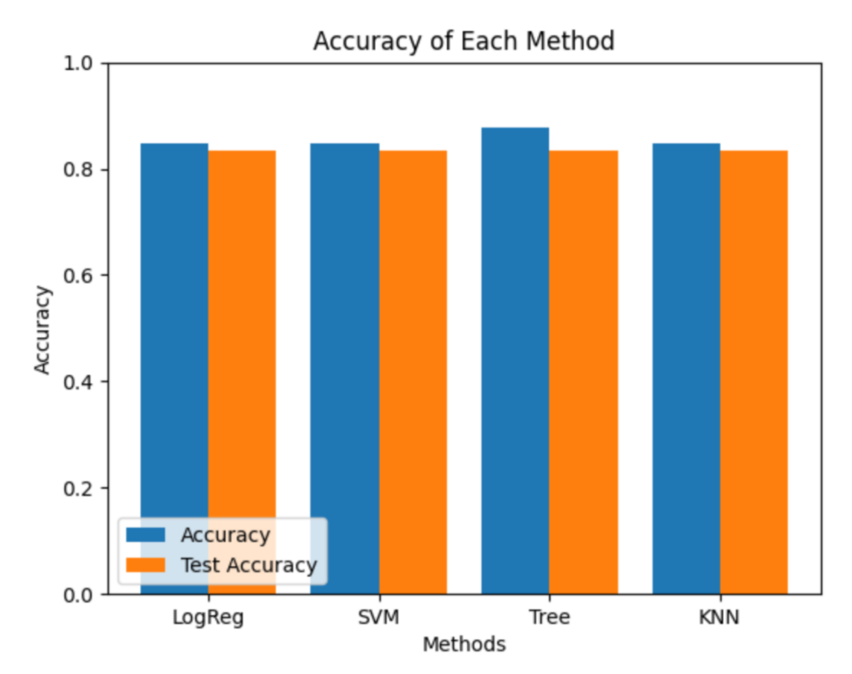


- Below 5000 kg, it is almost equally likely to succeed or fail
- Over 5000 kg, it is more likely to fail

Section 5

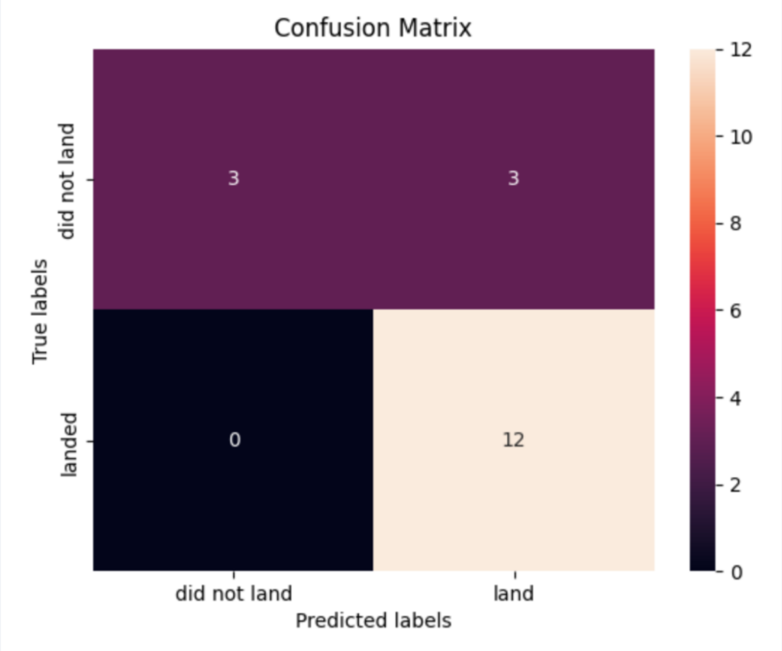
Predictive Analysis (Classification)

Classification Accuracy



- Four models were trained and the train / test accuracy was displayed above
- They all have the same test accuracy, possibly because the dataset sample size is too small
- Decision Tree model has slightly higher training accuracy

Confusion Matrix



- This is the confusion matrix for Decision Tree model
- The majority cases are true positives
- The problem in this model is mainly in the cases of false positives

Conclusions

- KSC LC-39A might be the best option among all launch sites
- Launch sites with convenient logistics are preferred
- Low payload mass is preferred
- Success rate improves over the years
- ES-L1, GEO, HEO, SSO orbits are preferred
- For future analysis, decision tree might be preferred

Appendix

Special thanks to IBM, Coursera, instructors, and peers who help review my work

Thank you!