Data Insights in space exploration.

SM

13-02-2023

OUTLINE

- Executive Summary
- Introduction
- Data collection
- Data wrangling
- EDA with SQL
- Completed the required predictive analysis methodology related slides (1 pt)
- Completed the required EDA with visualization results slides (6 pts)
- Completed the required EDA with SQL results slides (10 pts)
- Completed the required interactive map with Folium results slides (3 pts)
- Completed the required Plotly Dash dashboard results slides (3 pts)
- Predictive analysis
- Conclusion

EXECUTIVE SUMMARY

- Target
  - Commercialize space travel.
  - Recycle parts of specific phases of travel.
  - Optimize successful outcome.
- Result
  - Various Machine learning algorithms are used which are able to predict the outcomes with reasonably high degree of accuracy.
  - The sample size is pretty short to be bullish of such outcomes. More data is required for better inference.
  - However the outcomes can be used to sensibly eliminate certain endeavours.

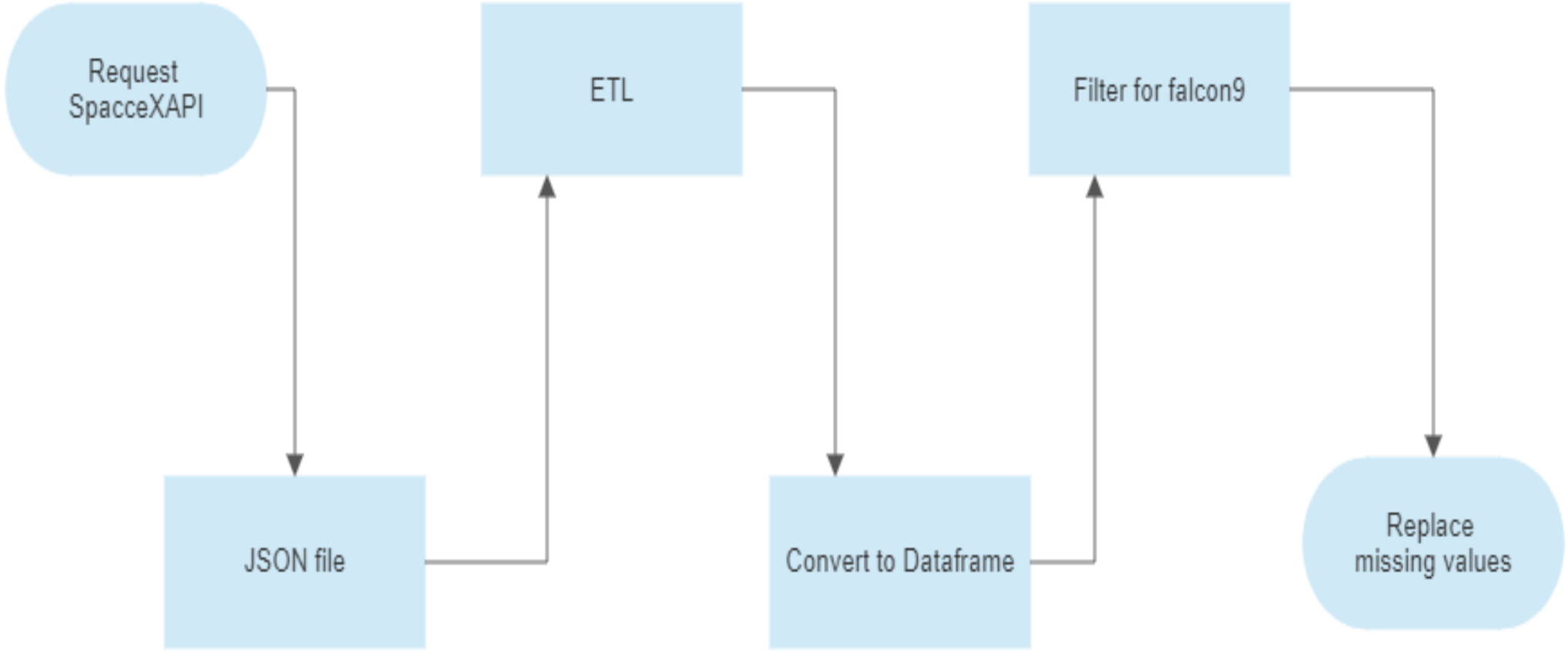
INTRODUCTION

- Future of commercial exploration includes space.
- SpaceX has the cheapest price.
- SpaceX is able to recycle rocket part of Stage1.

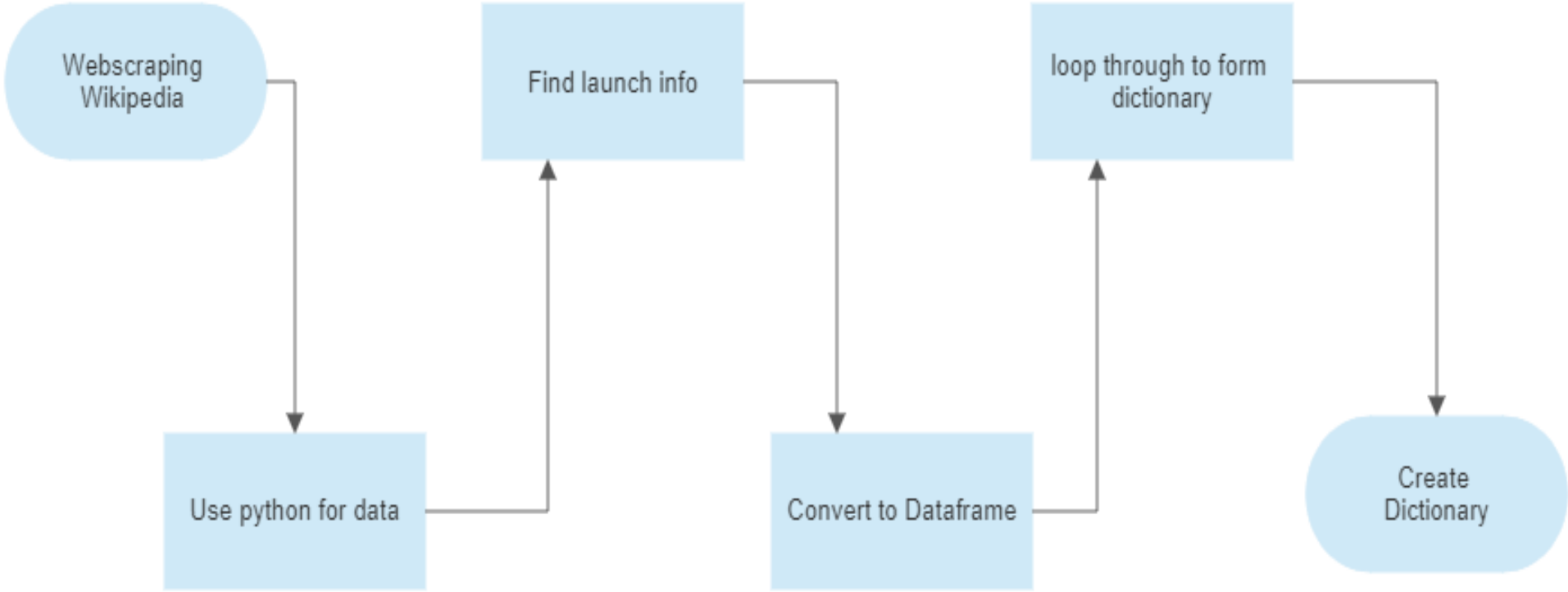
- Target
  - SpaceY wants to compete with SpaceX.
  - Predict success in Stage 1.

Data Collection

- Combine inputs from SpaceX API library and Wikipedia entry.
- Process detailed in flowchart in following slides.
- SpaceX API Columns:
  - Flight no. ,Date ,Booster-Version ,Payload-Mass ,Orbit ,Flights ,Launch-site ,Outcome ,Reused , Grid-Fins ,Legs ,Landing-Pad , Block ,Reused-Count ,Serial ,Longitude ,Latitude
- Wikipedia Web scrapped Data Columns:
  - Flight no. ,Date ,Time ,Version-Booster ,Booster-landing ,Payload-Mass ,Payload ,Orbit ,Launch-site ,Customer



https://github.com/Sanak11/DataStart/blob/master/Week1/Data-Collection/jupyter-labs-webscraping.ipynb SpaceX API Data Collection notebook repository link.



https://github.com/Sanak11/DataStart/blob/master/Week1/Data-Collection/jupyter-labs-webscraping.ipynb - Web scraping Notebook repository link.

Data Wrangling

- Eight Different outcomes are observed from the 'Outcome' column. From it favourable and unfavourable outcome is distinguished.
- Successful and Unsuccessful outcomes are denoted by their boolean indicators 1 and 0 respectively and converted to numerical variables.
- Succesful outcome is 66.67%.
- Github repository link:
  - https://github.com/Sanak11/DataStart/blob/master/Week1/Data-Wrangling/IBM-DS0321EN-SkillsNetwork_labs_module_1_L3_labs-jupyter-spacex-data_wrangling_jupyterlite.jupyterlite.ipynb

EDA with SQL

- Various queries are performed
- Deeper idea about the dataset achieved. Some of those observations are shown in the subsequent slides.
- Github repository link:
  - https://github.com/Sanak11/DataStart/blob/master/Week2/EDA-in-SQL/jupyter-labs-eda-sql-coursera_sqllite.ipynb

SpaceX distinct Launch-Sites

Launch-Site	Launch-Site
CCAFS LC-40	Cape Canaveral Space Launch Complex 40
VAFB SLC-4E	Vandenberg Space Force Base Space Launch Complex 4E
KSC LC-39A	Kennedy Space Center Launch Complex 39A
CCAFS SLC-40	Cape Canaveral Space Launch Complex 40

Five earliest launch from Cape Canaveral Space Launch Complex 40

Date	Time (UTC)	Booster_Version	Launch_Site	Mission_Outcome	Landing _Outcome
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Success	Failure (parachute)
08-12-2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Success	Failure (parachute)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	Success	No attempt

Other Insights

- Total payload mass carried by boosters launched by NASA (CRS) is 111268Kg.
- Average payload mass carried by booster version F9 v1.1 is 2354.66667Kg.
- First succesful landing outcome in ground pad was achieved in first of March 2017.
- 24 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

Mission Outcome	Total count
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Booster Versions Carrying Maximum Payload Mass

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

Failures in 2015

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

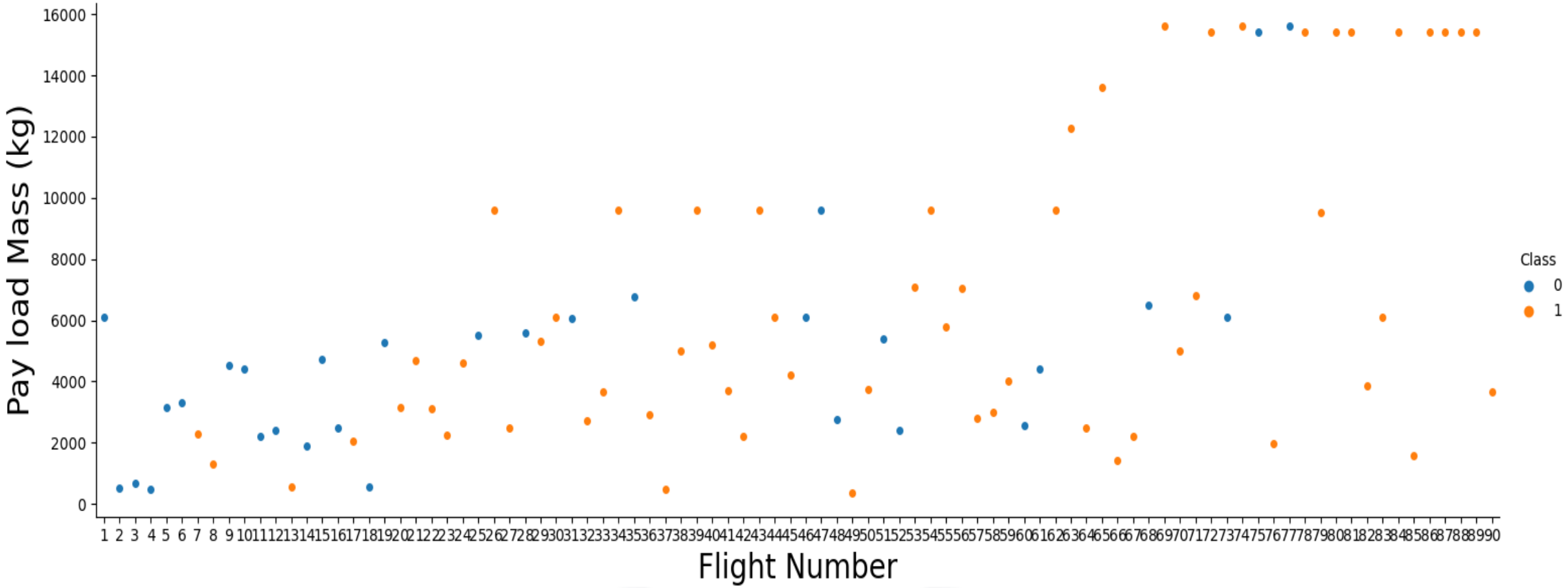
Count of successful landing outcomes between the 04-06-2010 and 20-03-2017 in descending order.

Date	Landing _Outcome	Launch_Site
19-02-2017	Success (ground pad)	KSC LC-39A
18-07-2016	Success (ground pad)	CCAFS LC-40

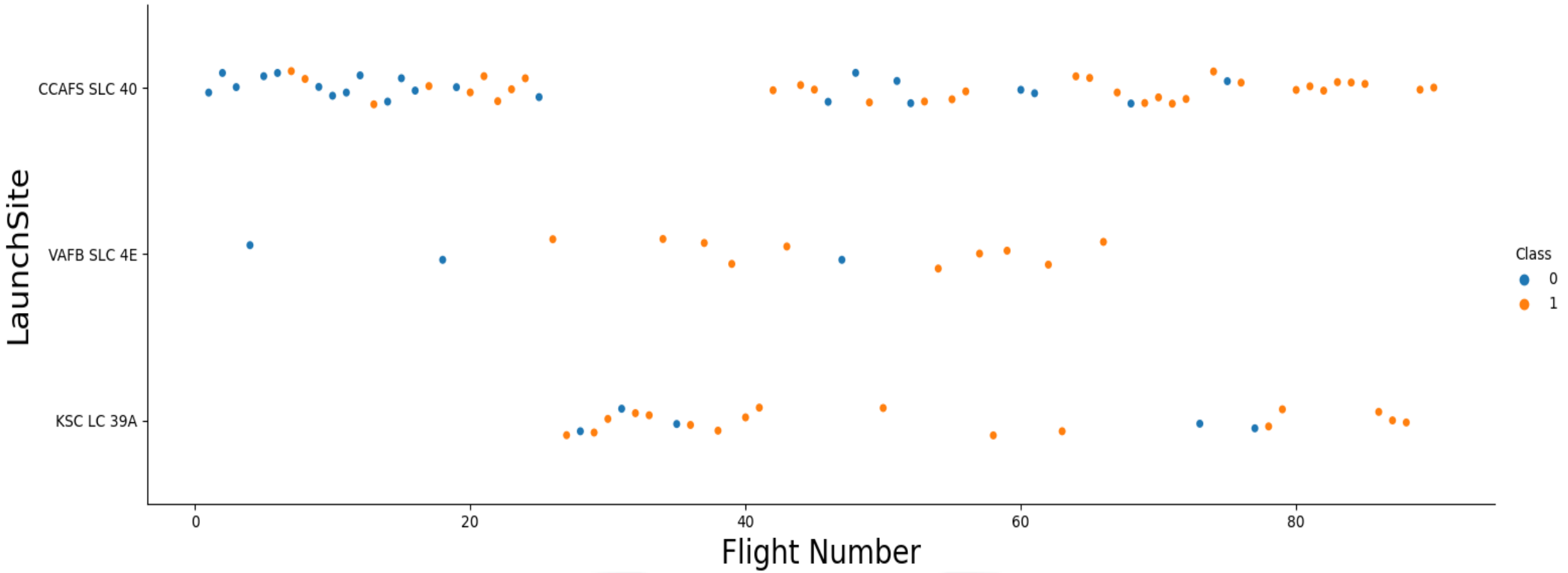
EDA and Data Visualization

- Dataset is imported to notebook and analyzed by transformation to visual plots for deeper insights.
- These insights are able to create the backbone of machine-learning model for success prediction.
- Categorical variables are converted to numerical and finally to decimal.
- Github repository link:
  - https://github.com/Sanak11/DataStart/blob/master/Week2/EDA-in-PandasMatplotlib/IBM-DS0321EN-SkillsNetwork_labs_module_2_jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

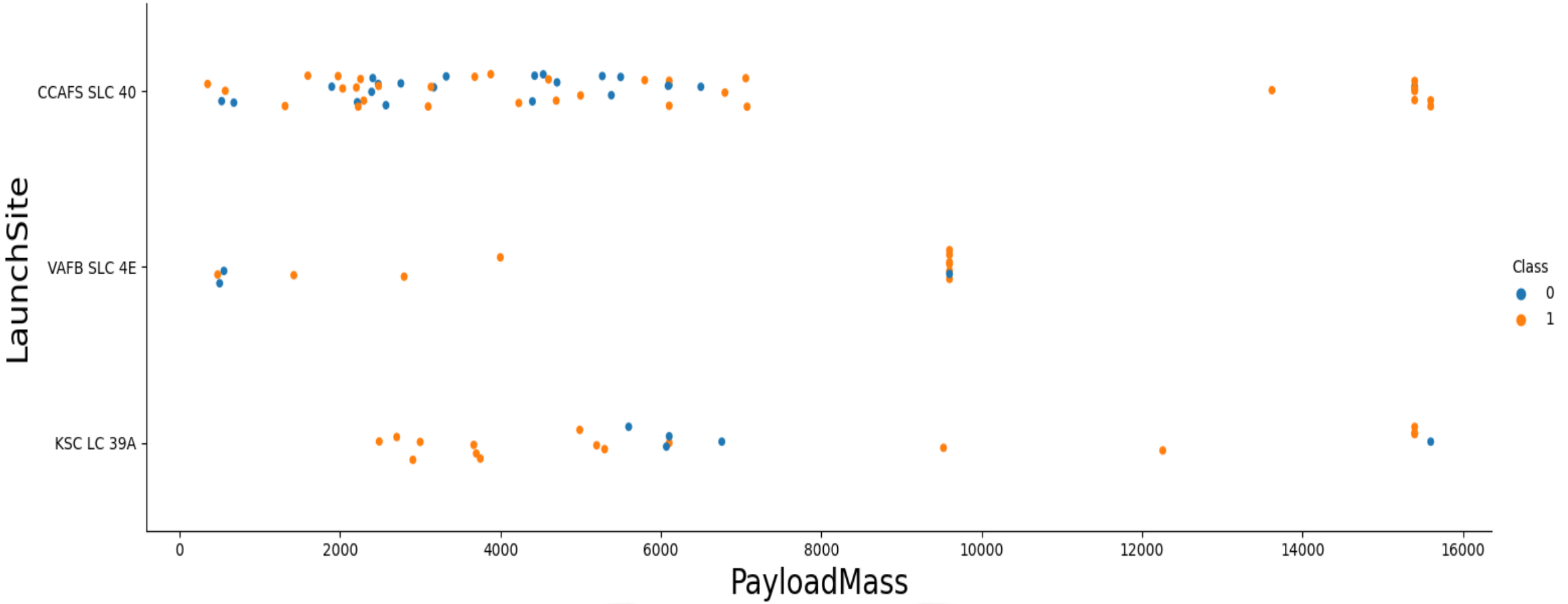
Relationship between Flight Number and Payload-Mass



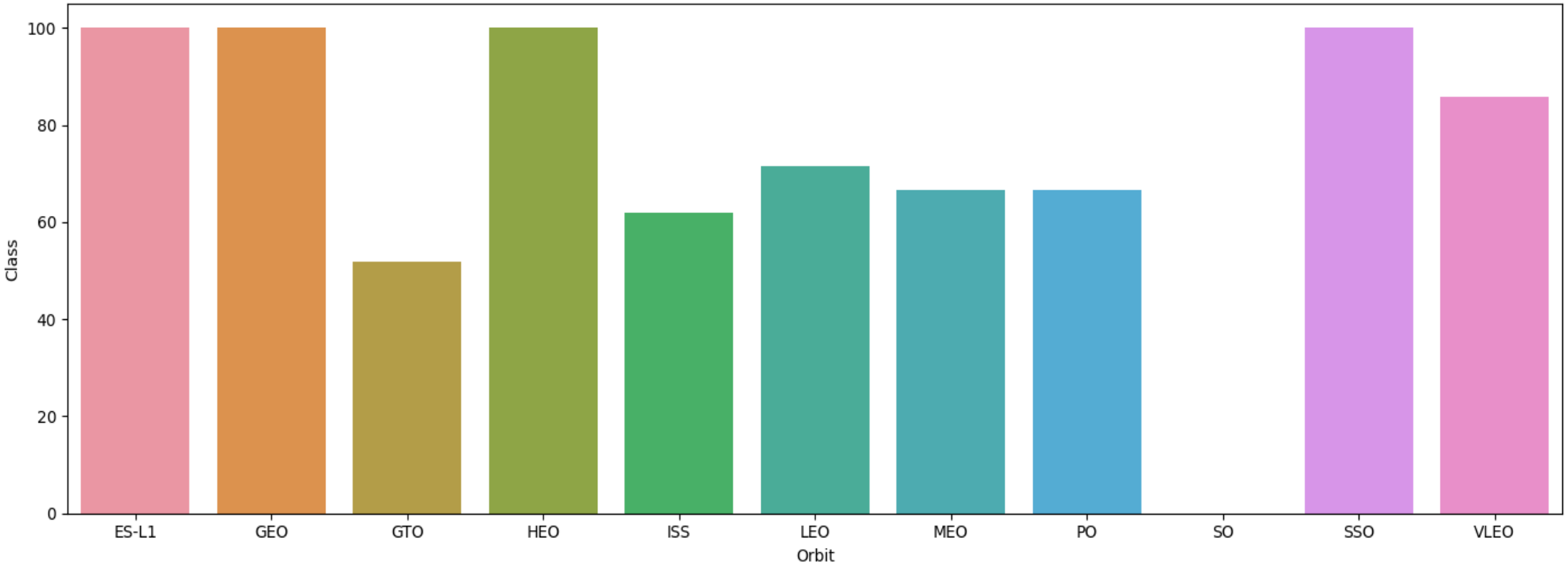
Relationship between Flight Number and Launch Site



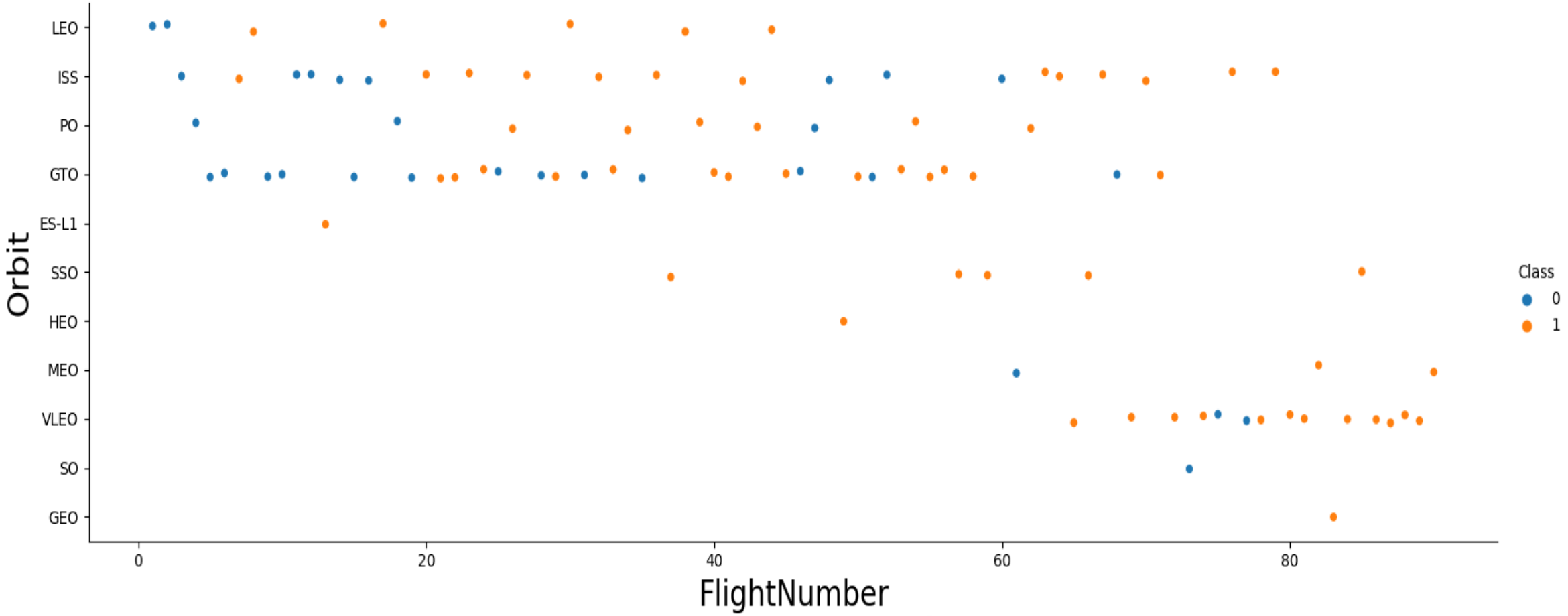
Relationship between Payload-mass and Launch Site



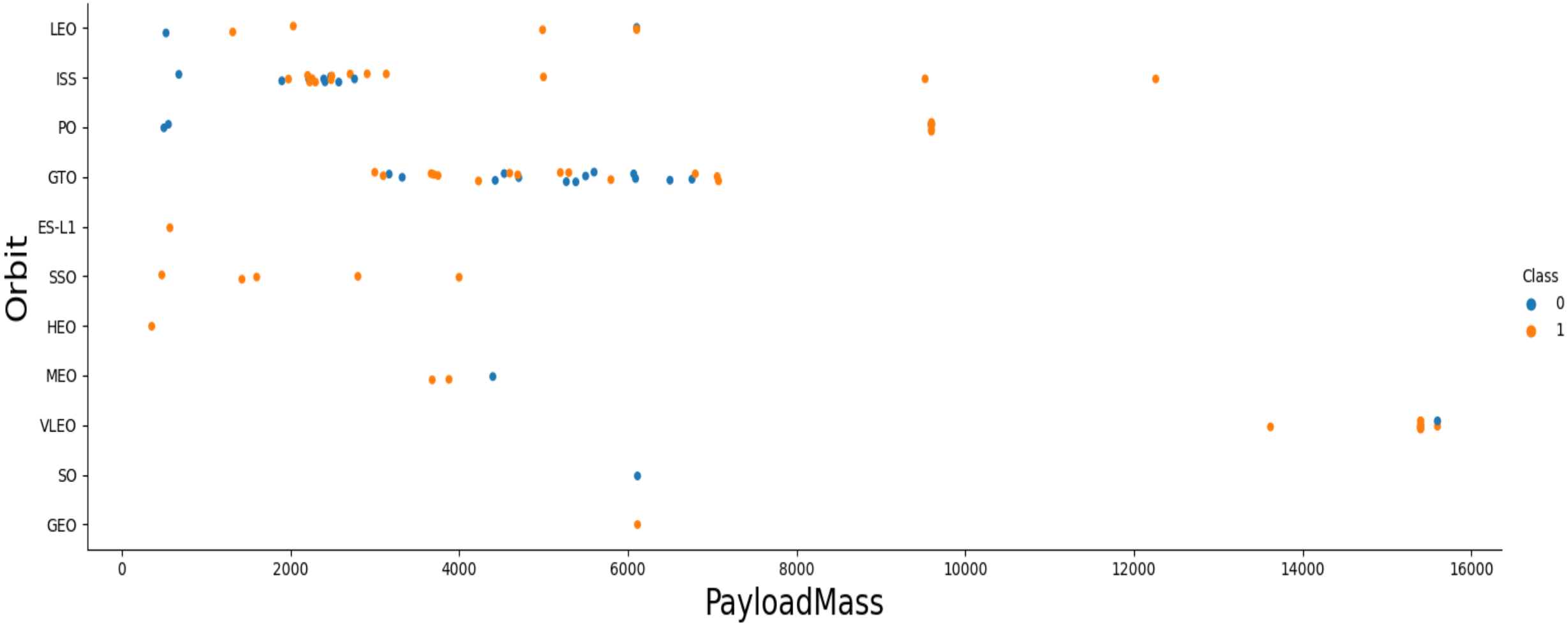
Relationship between success rate of each orbit type



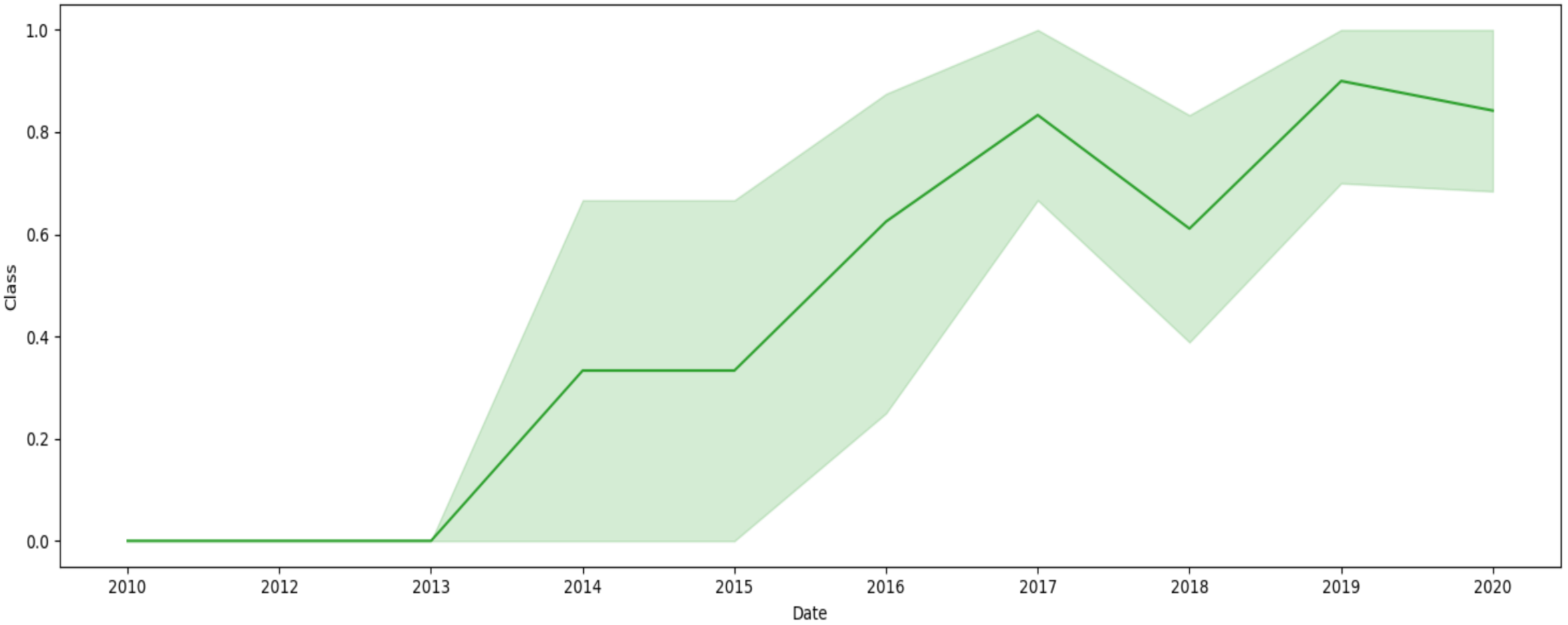
Relationship between Flight-Number and Orbit type



Relationship between Payload and Orbit type



Growth in success Yearly



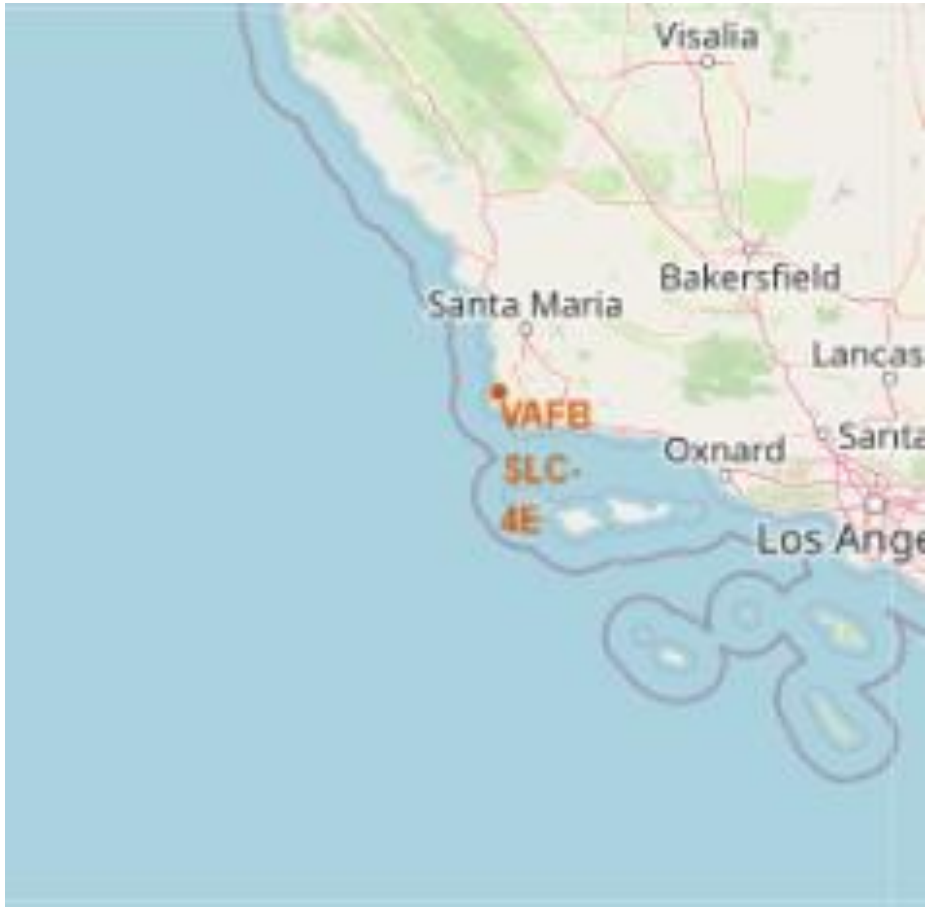
Launch-site Analysis

Launch Site	Lat	Long
CCAFS LC-40	28.562302	-80.577356
CCAFS SLC-40	28.563197	-80.576820
KSC LC-39A	28.573255	-80.646895
VAFB SLC-4E	34.632834	-120.610745

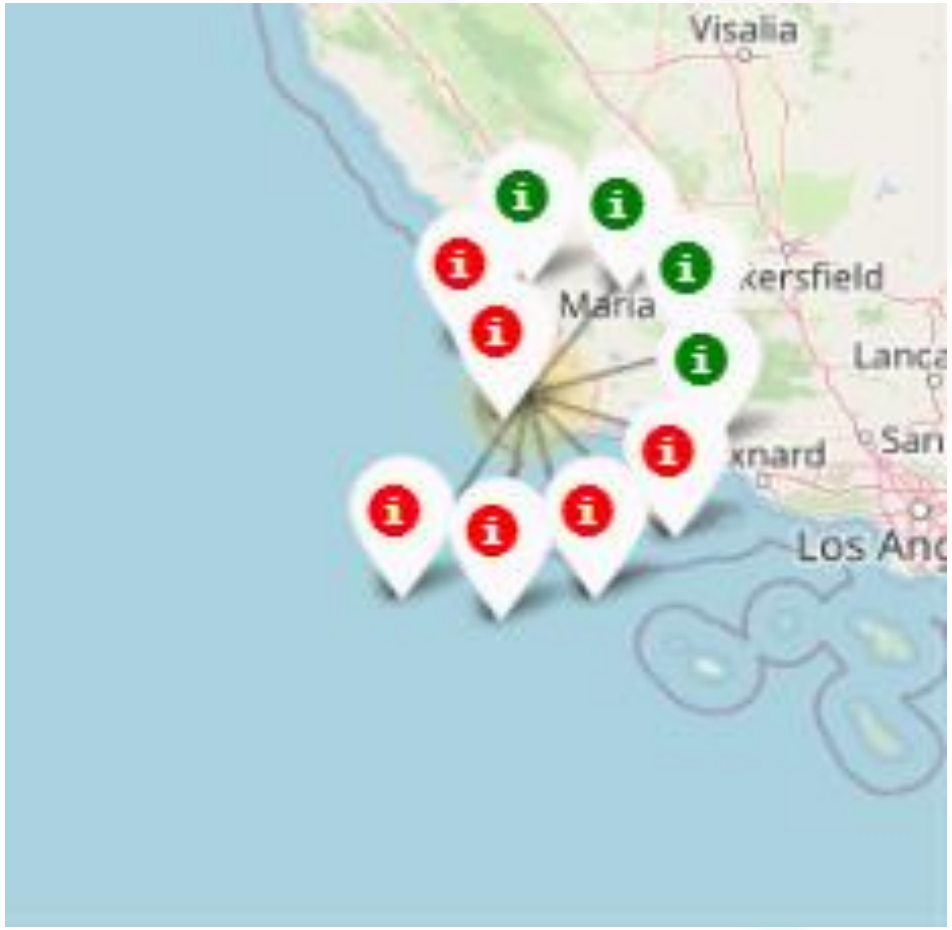
CCAFS LC-40 and CCAFS SLC-40 are in the same position on East Coast U.S.
VAFB SLC-4E is on the West coast U.S.
KSC LC-39A is near the East coast U.S. but west to CCAFS.
The maps are shown in the following slides.

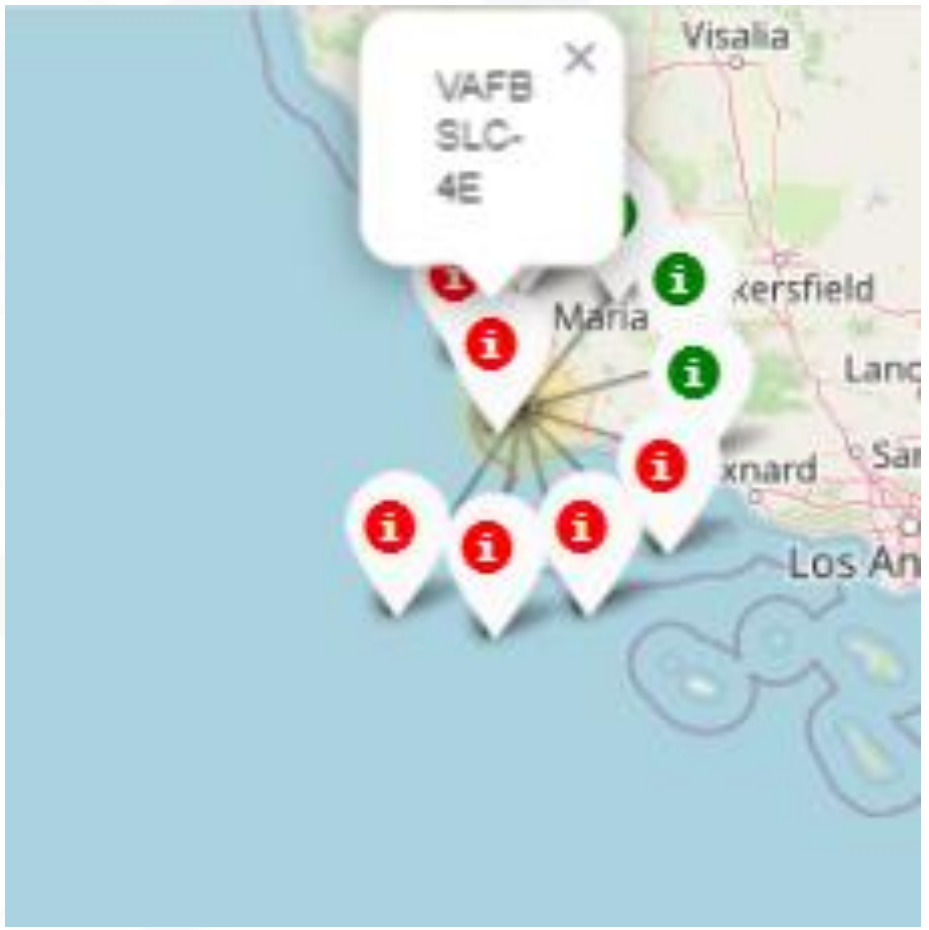

New York
Washington
United States
VAFB
Los Angeles
Phoenix
4E
CCAFS
The Bahamas
La Habana
México
Cuba
Ciudad
de México
Kingston
Ciudad
de Guatemala
Honduras
Nicaragua


Cochran
Cove
KSC
LC-
39A
CCAFS
SLC-
40


Visalia
Bakersfield
Santa Maria
VAFB
SLC-
4E
Oxnard

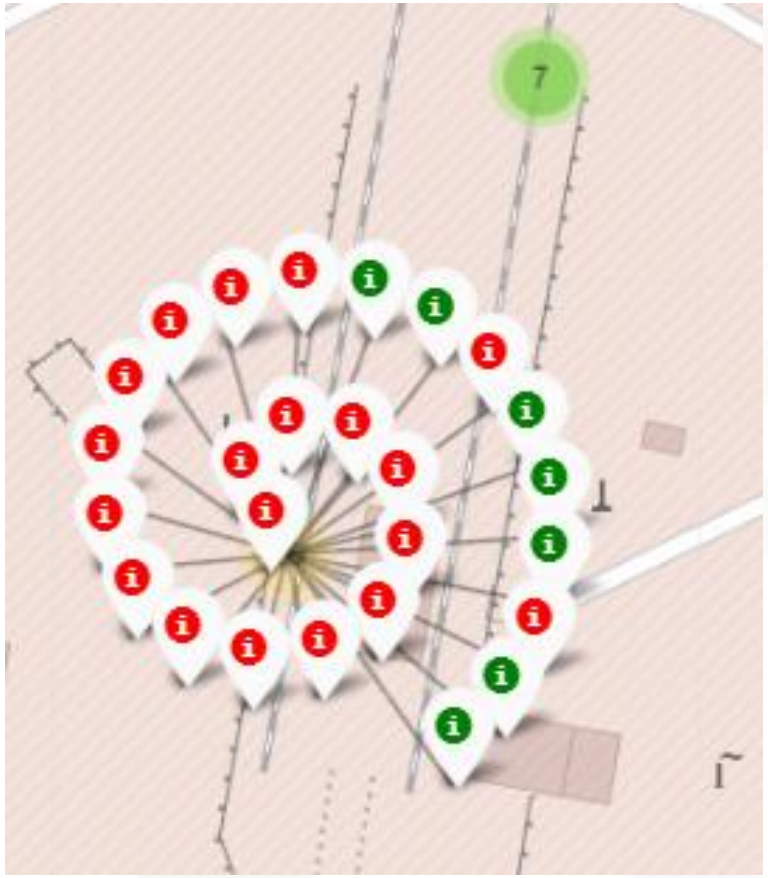
Launches at VAFB SLC-4E



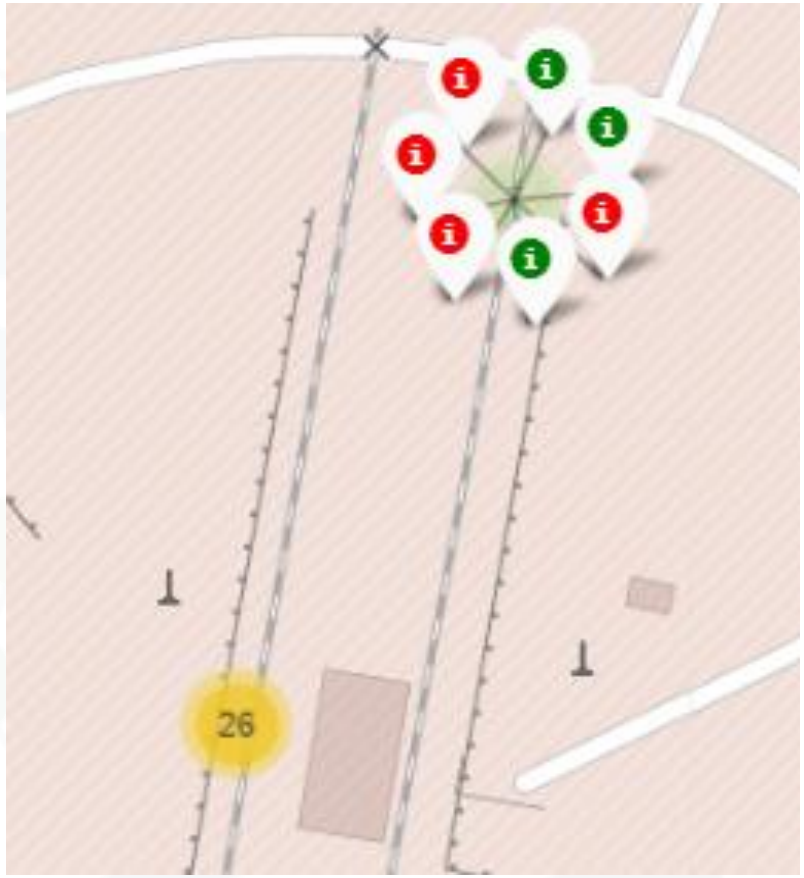


Launches at KSC LC-39A

Launches at CCAFS LC-40 and CCAFS SLC-40

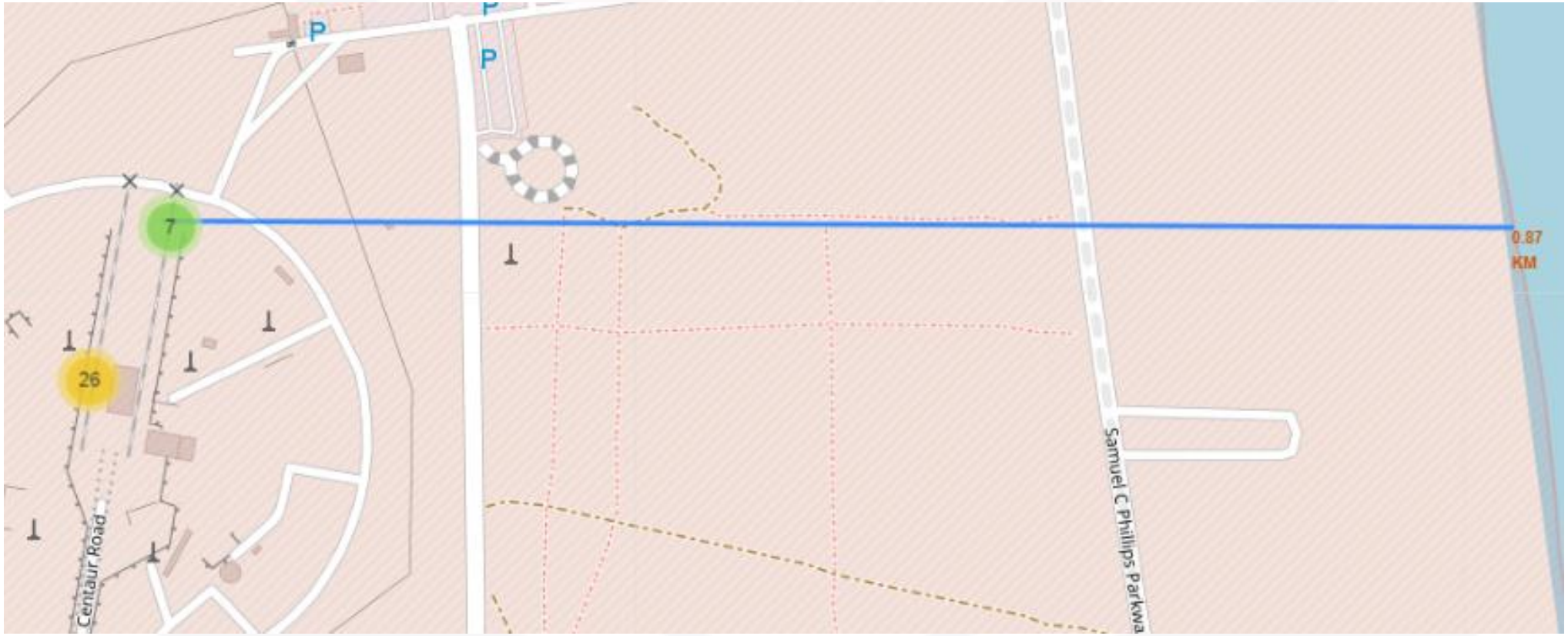


CCAFS LC-40

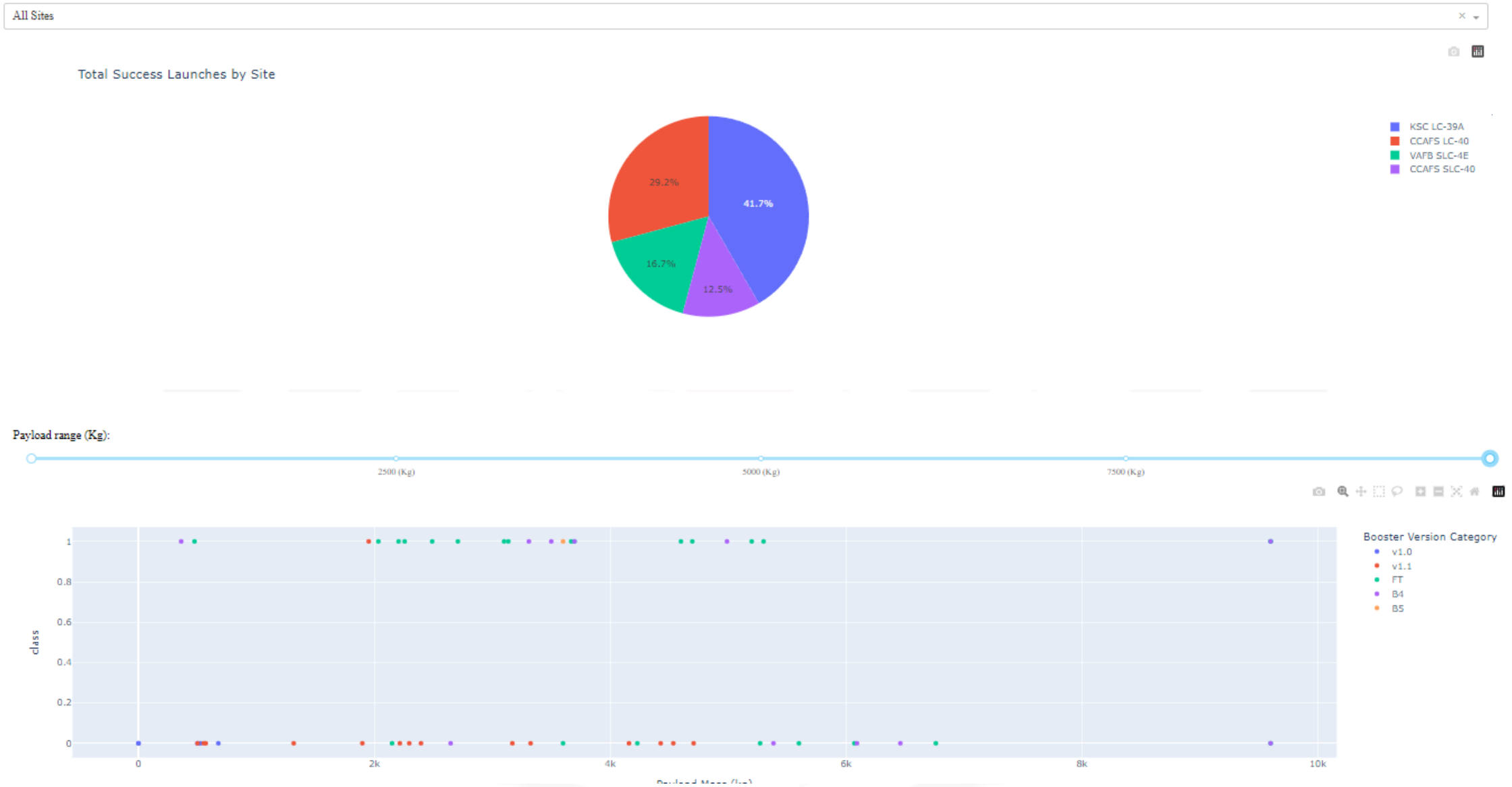


CCAFS SLC-40

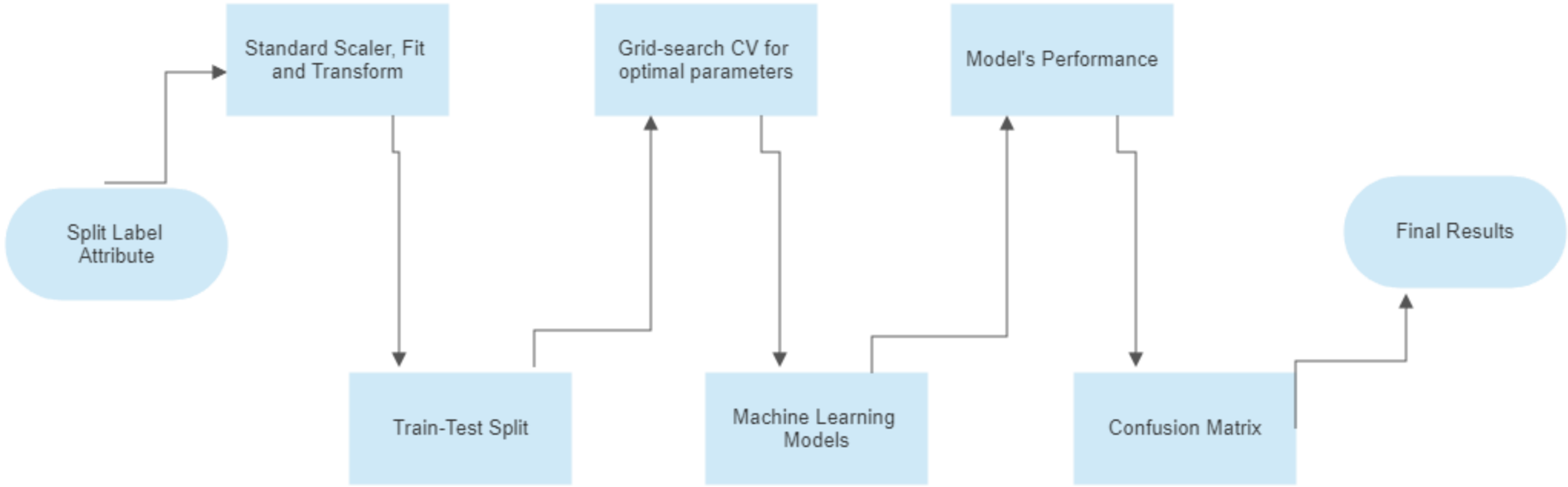
Distance between CCAFS LC-40 and coast



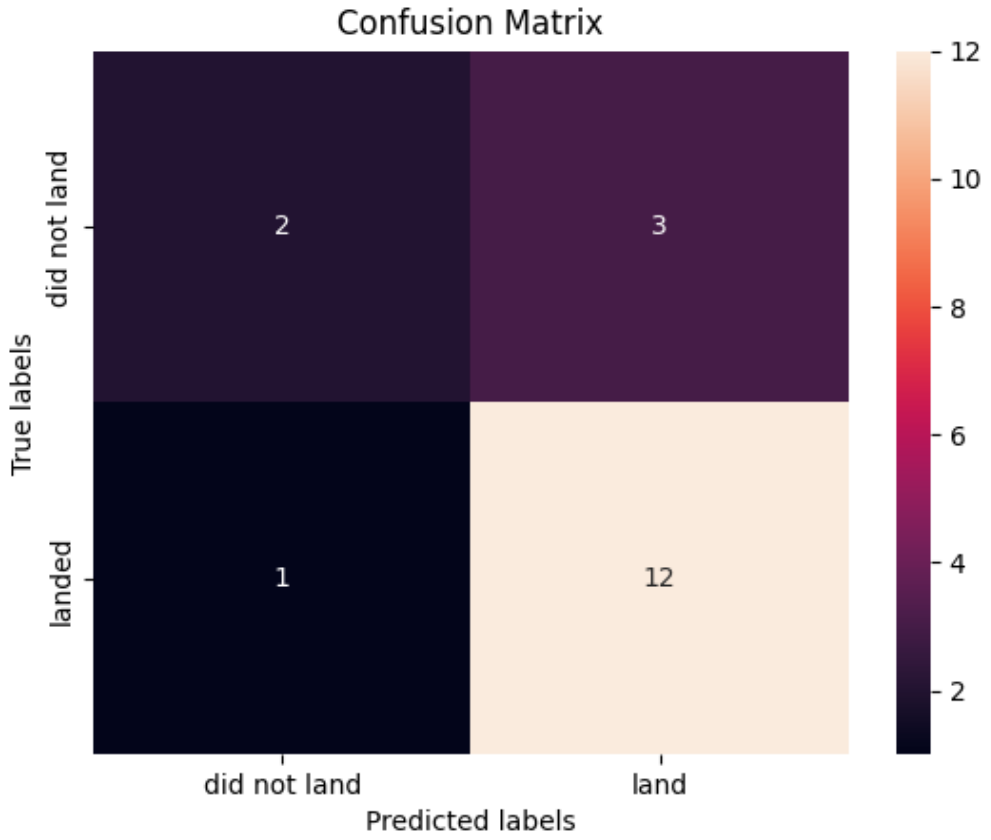
SpaceX Launch Records Dashboard



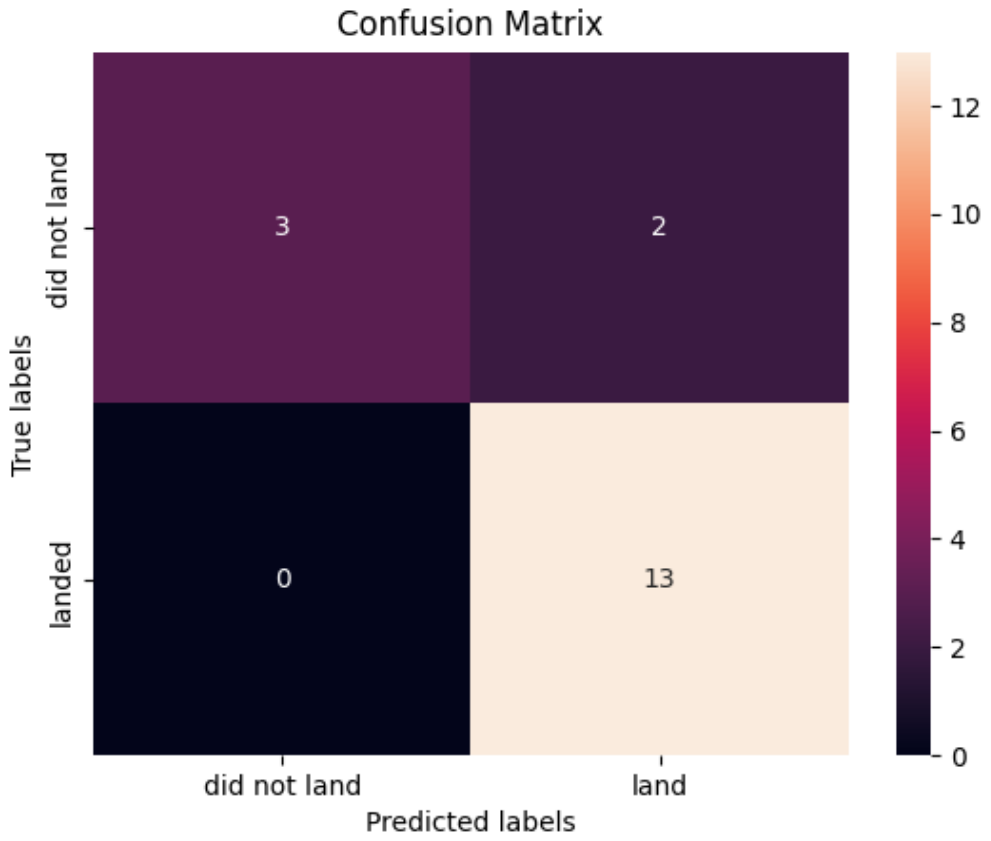
Predictive Analysis



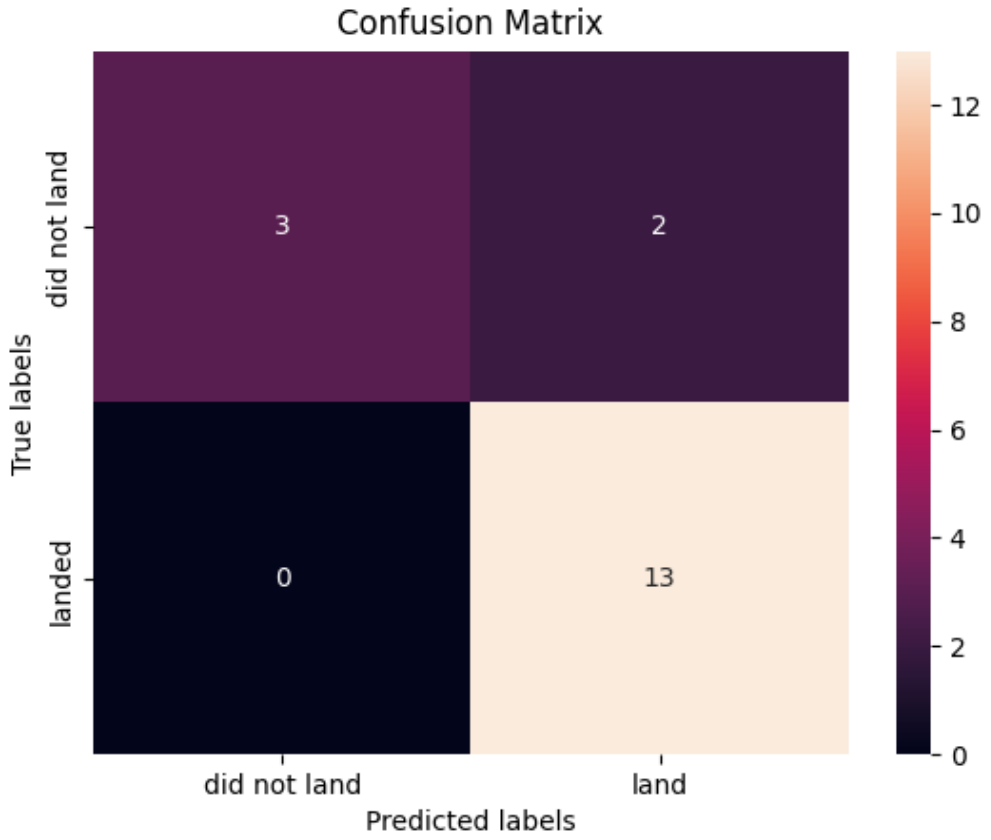
Linear regression predictions



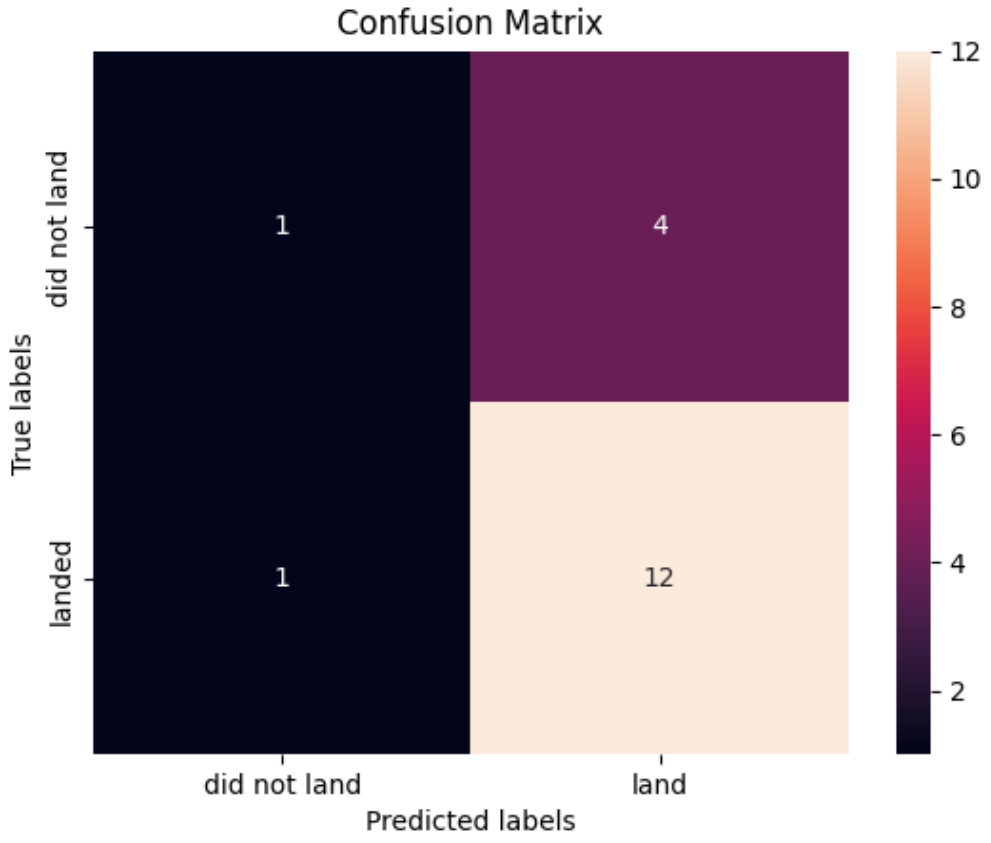
SVM predictions



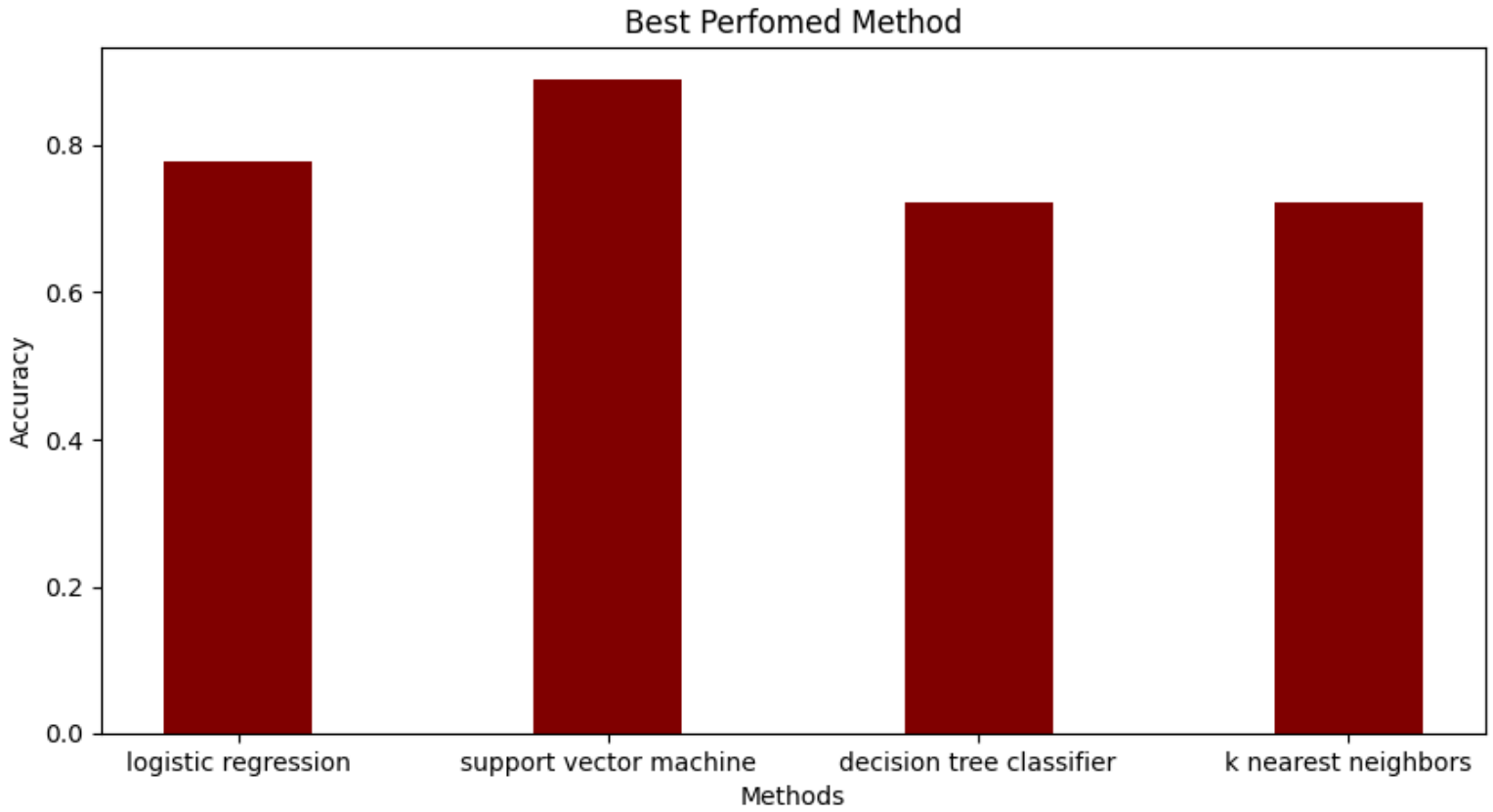
Decision tree predictions



K-N-N predictions



Models Comparison



Conclusions

- The best model is able to predict with a high degree of accuracy the success of the model.
- Initial success accuracy was 66.67%.
- Predicted final success of specific ways to perform stage 1 was 88.89%.
- More data will provide more detailed analysis but the results are enough to test it out.

Appendix

- Special thanks to all the instructors who made this course possible.
- Github Repository:
  - https://github.com/Sanak11/DataStart/tree/master

Thank You