IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Jonas Fockstedt
2022-07-24

Outline

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

Executive Summary

- Data was collected via publicly available data sets
- The data was analyzed with a variety of different methods
- The results point to that in order to maximize a successful outcome of a mission, one should use
    - Launch site KSC LC-39A
    - The FT booster
    - Have a payload between 2000kg and 6000kg

Introduction

- SpaceY
  - New space startup
  - Needs to be competitive with SpaceX
- Main question – will SpaceX reuse the first stage of the Falcon 9 rocket?

Section 1

Methodology

Methodology

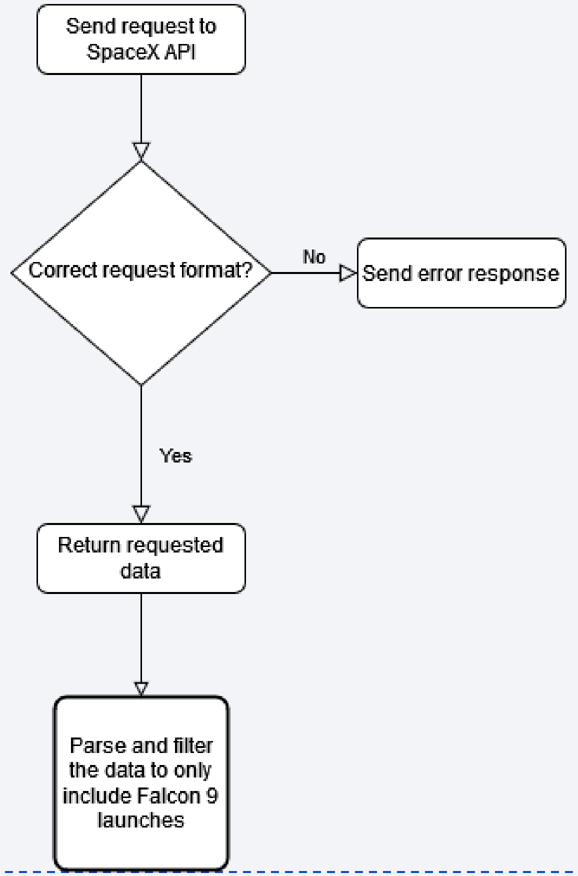
Executive Summary

- Data collection methodology:
  - Data was collected via publicly available datasets
- Perform data wrangling
  - Data was processed using Python and the Pandas library
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection – SpaceX API

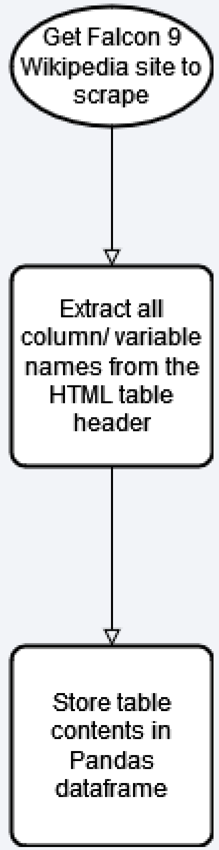
- Using the requests library for Python, the data was collected via API calls to the appropriate address
- The flow of the process is shown in the flow chart

- [Notebook]()



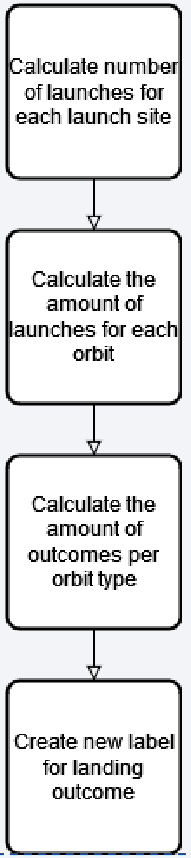
Data Collection - Scraping

- Additional data needed to be collected from the Wikipedia site, in a table containing information of the F9 rocket
- The webscraping was done using BeautifulSoup library in Python
    - The different steps are displayed in the flow chart
- Notebook



Data Wrangling

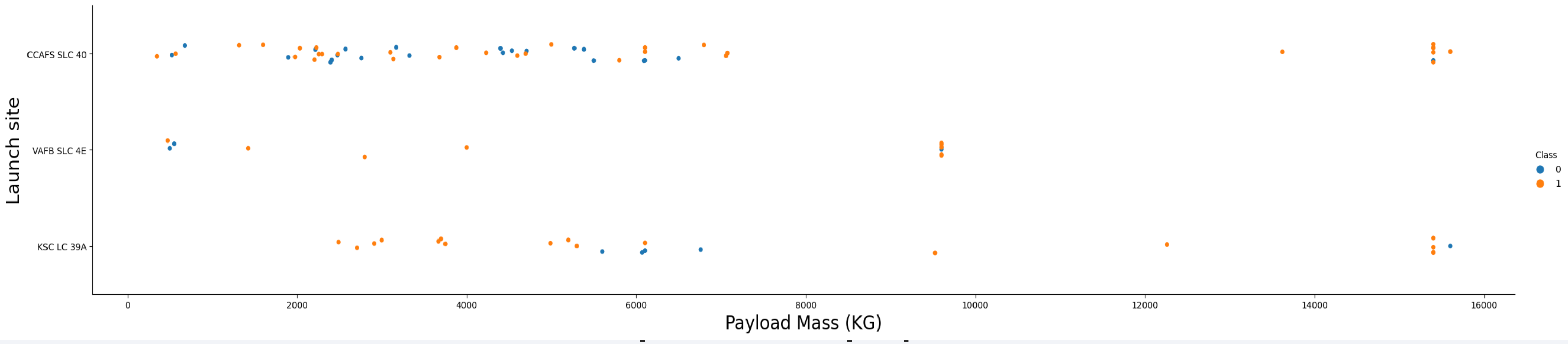
- Describe how data were processed
- Data was processed by doing calculations for each launch site as well as orbit type
- The workflow is depicted in the flow chart



- Notebook

EDA with Data Visualization

- To get a better understanding of the relationship between different variables, scatter plots were used. These insights would lead to a better understanding of important variables
  - Below plot shows the relationship between payload mass and launch site, and whether the launch was successful or not (0=unsuccessful, 1=successful)



[Notebook](#)

EDA with SQL

- SQL queries were performed where which in summary extracted data of which
  - Which launch site had the most successful amount of landings
  - Which booster version had carried the maximum payload mass

- [Notebook]

Build an Interactive Map with Folium

- The objects added to the Folium map were
  - Circles, Markers, PolyLines, Mouse position
- These objects where necessary to add in order to mark out where the different launch sites are located on the world map as well as display information about the different launch sites. This information includes the amount of successful and unsuccessful launches the corresponding launch site had

- [Notebook]

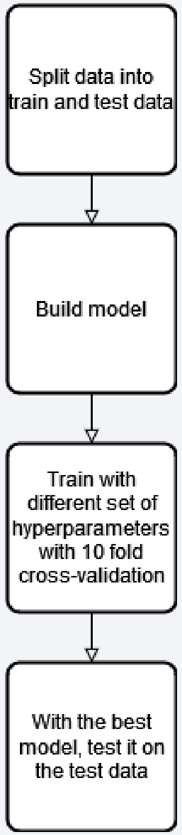
Build a Dashboard with Plotly Dash

- The Plotly Dashboard has a drop-down menu, pie chart, range slider, and a scatter plot
- The drop-down menu works as an interactive way of filtering the data displayed in the pie chart, which in turn shows the amount of successful and unsuccessful launches for each site. The range slider determines what payload mass should be displayed in the scatter plot, which shows the payload mass vs the success of a given launch

- SpaceX Dash App

Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model
- The different classification models used were developed with the Scikit learn library for Python. They were all trained with grid search where a number of different hyperparameters were tried out, with a cross-validation fold of 10. The best performing model among the given hyperparameters was later evaluated on the test set (20% of the original dataset, 80% for training and validation)
    - The models that were tested were logistic regression, decision tree, svm, and KNN
- [Notebook]

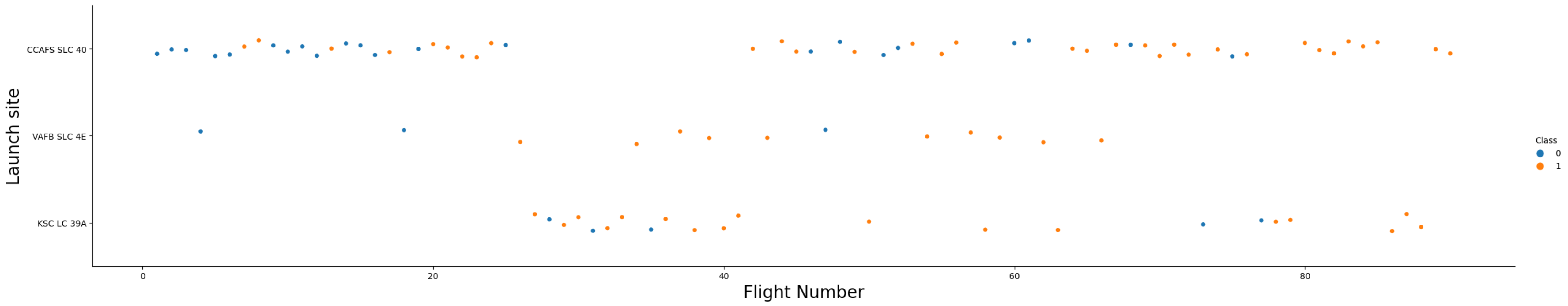


Section 2

Insights drawn from EDA

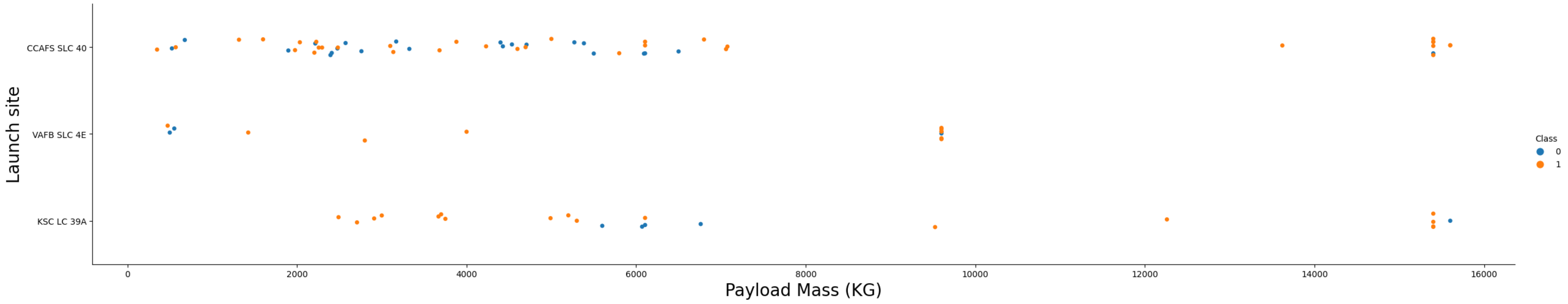
Flight Number vs. Launch Site

- A scatter plot for the Flight Number vs Launch site is shown:



- From the plot, one can conclude that all launch sites have higher success rate (class=1, unsuccessful = class 0) as the number of flights from that launch site increases.
- Launch site KSC LC 39A shows the highest landing success rate (77%)

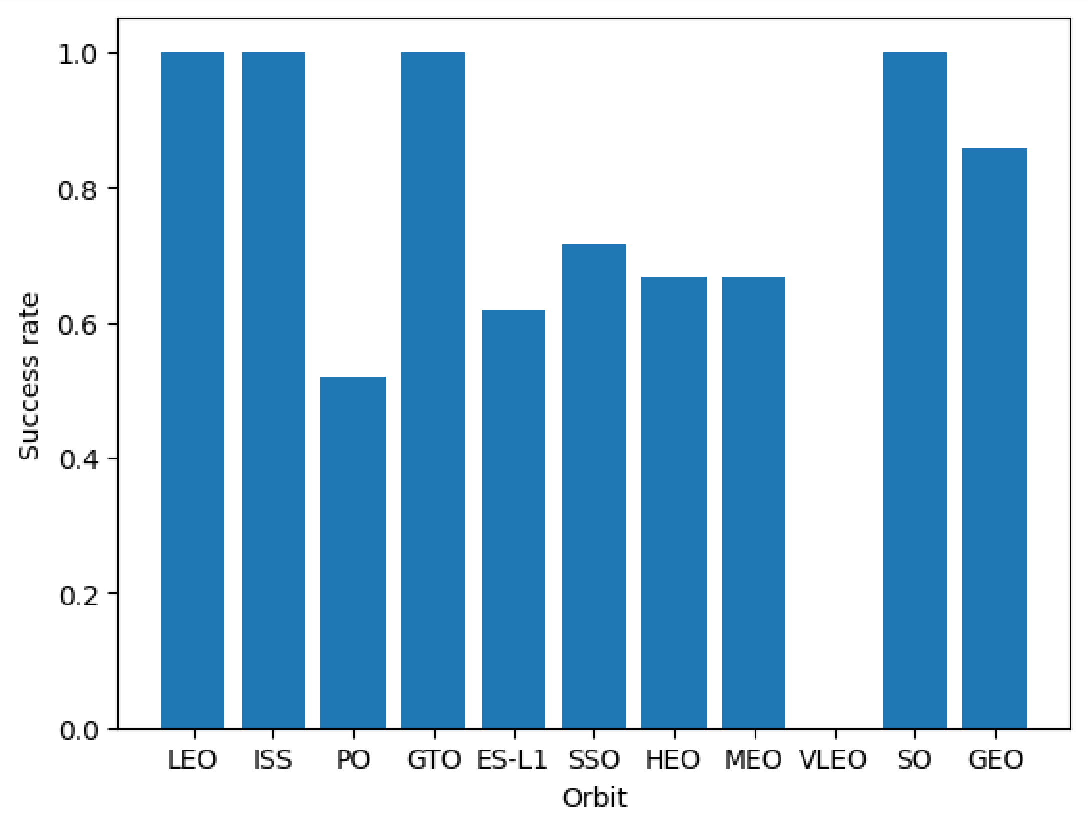
Payload vs. Launch Site



- Class = 0, unsuccessful landing, class = 1, successful landing
- The scatter plot shows the relationship between the success of a given launch site with a given payload. It tells that the success rate is generally higher for heavier payloads.
- Launch site KSC LC 39A shows a higher overall success rate (76%)

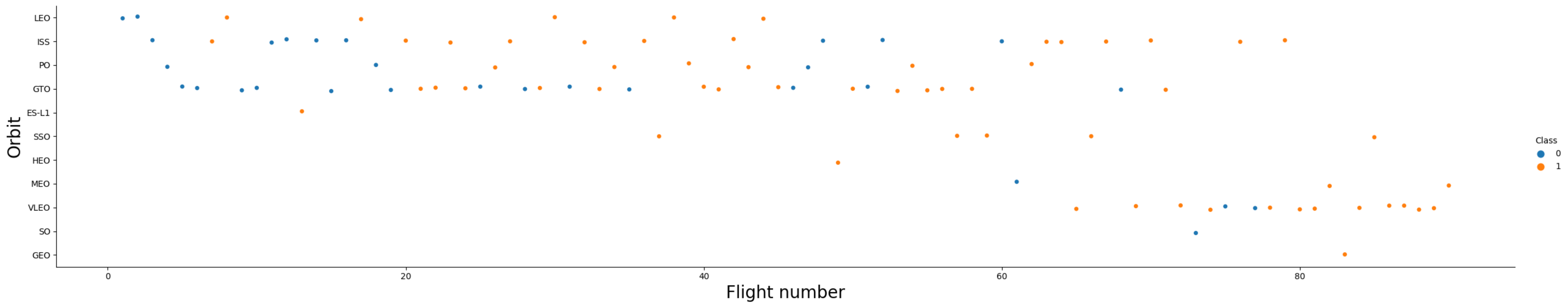
Success Rate vs. Orbit Type

- The bar chart shows the success rate for each orbit type
  - LEO, ISS, GTO, SO all have 100% success rate
  - VLEO the only orbit with 0% success rate



Flight Number vs. Orbit Type

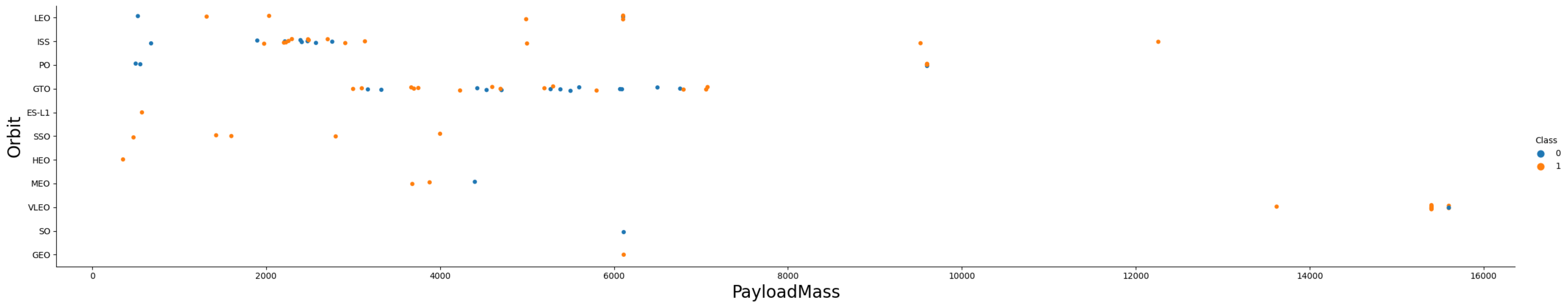
- Successful landing = Class 1, unsuccessful landing = Class 0



- The general trend shows a higher success rate as the flight number increases for all orbit types

Payload vs. Orbit Type

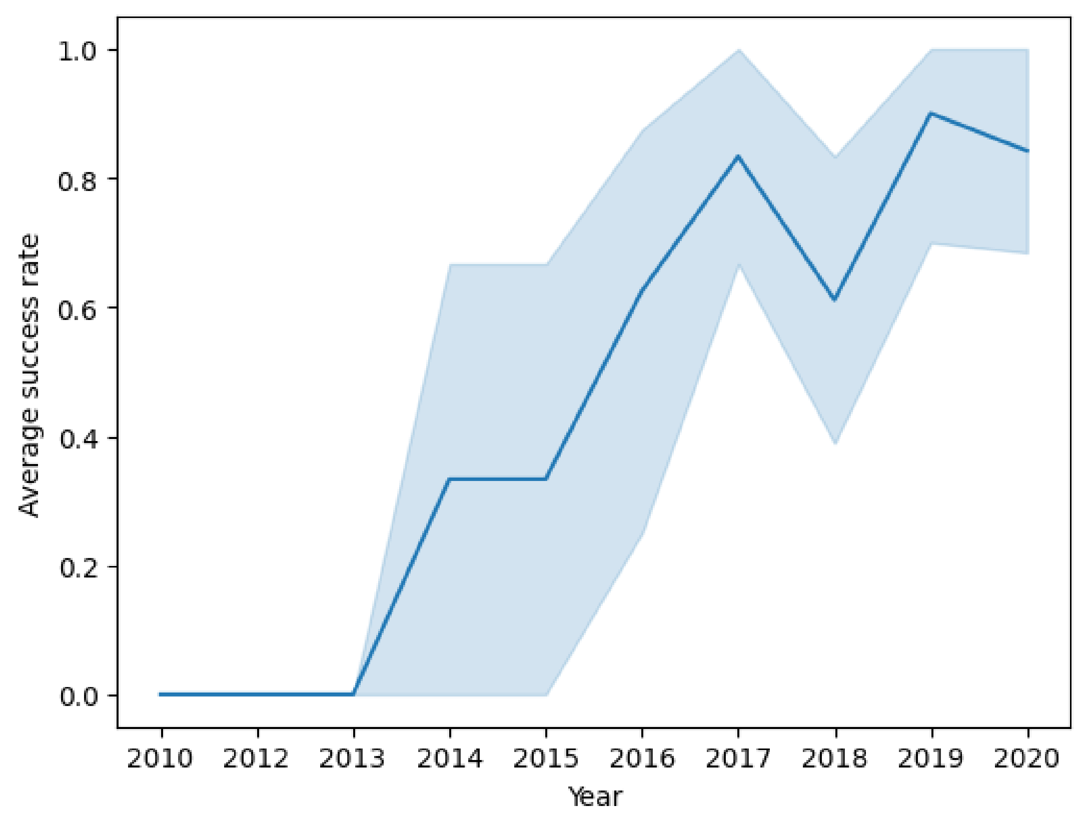
- Successful landing = Class 1, Unsuccessful landing = Class 0



- The VLEO orbit missions are the ones with the highest payload mass for the rocket
- SSO, HEO, and GEO orbit all show 100% success rate, despite having different payload masses

Launch Success Yearly Trend

- The plot shows the average success rate for a given year
- After 2013, the trend steadily increases



All Launch Site Names

- The names of the unique launch sites are displayed in the table

Launch sites
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40

Launch Site Names Begin with 'CCA'

- The table is the result shown from a query which returns 5 records where the launch site of the mission begins 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

Total Payload Mass

- The table shows the total payload which has been carried by NASA as a customer of SpaceX
  - Shown in KG

TOTAL_PAYLOAD_KG
45596

Average Payload Mass by F9 v1.1

- The table is a result of averaging the payload carried by the F9 version 1.1 booster, shown in KG

AVG_PAYLOAD_KG
2928.4

First Successful Ground Landing Date

- The date of the first successful landing on the ground pad is shown in the table

First_Successful_Landing
01-05-2017

Successful Drone Ship Landing with Payload between 4000 and 6000

- The table shows the names of the different boosters which has had a payload mass greater than 4000kg and less than 6000kg

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

Total Number of Successful and Failure Mission Outcomes

- The table shows the amount of failed and successful mission outcomes, respectively
  - Successes dominates failures

Failures	Successes
1	100

Boosters Carried Maximum Payload

- The table displays the boosters which have carried the heaviest payload mass

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

- The table shows the missions with a failed landing outcome in the year 2015
  - The month column states the month number (e.g., 01 = January)

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

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

- The table shows the records of which occurred between 2010-06-04 and 2017-03-20 and also had a successful outcome, in descending date order

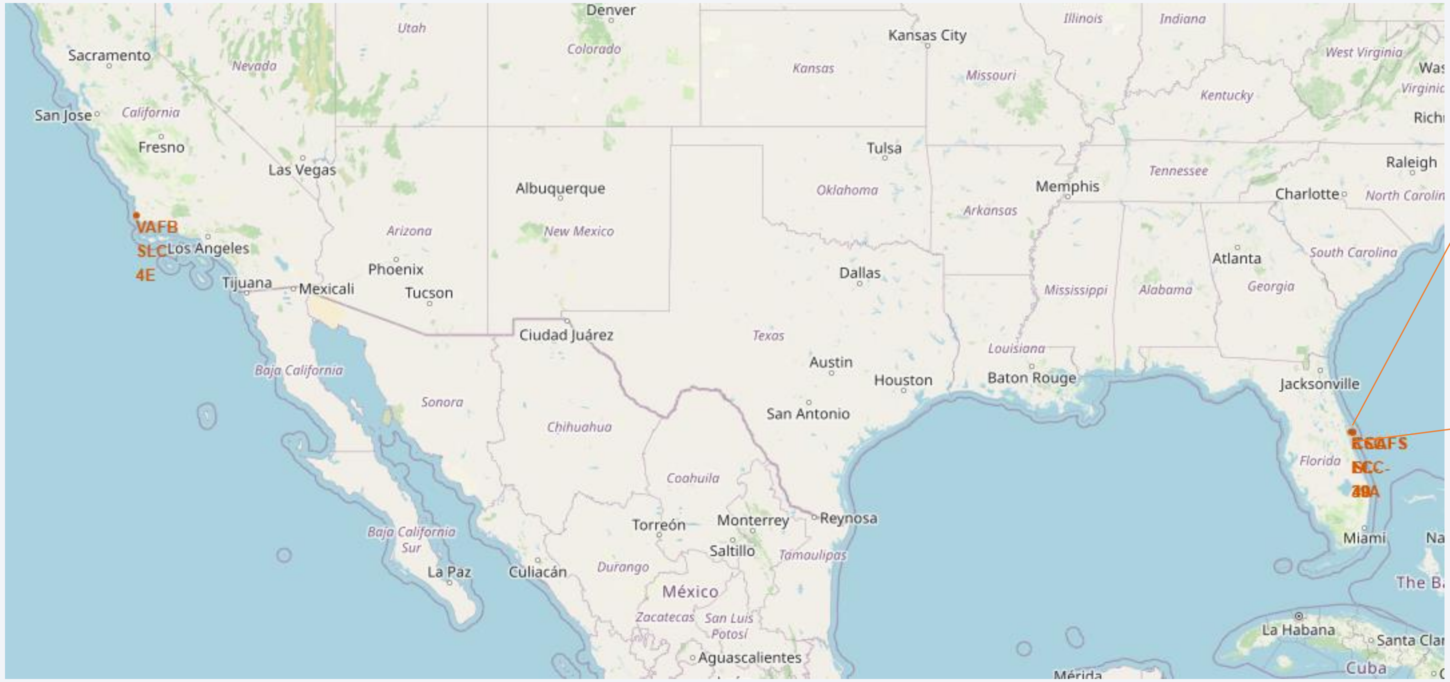
Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
22-12-2015	01:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)
08-04-2016	20:43:00	F9 FT B1021.1	CCAFS LC-40	SpaceX CRS-8	3136	LEO (ISS)	NASA (CRS)	Success	Success (drone ship)
06-05-2016	05:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
27-05-2016	21:39:00	F9 FT B1023.1	CCAFS LC-40	Thaicom 8	3100	GTO	Thaicom	Success	Success (drone ship)
18-07-2016	04:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
14-08-2016	05:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
14-01-2017	17:54:00	F9 FT B1029.1	VAFB SLC-4E	Iridium NEXT 1	9600	Polar LEO	Iridium Communications	Success	Success (drone ship)
19-02-2017	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)

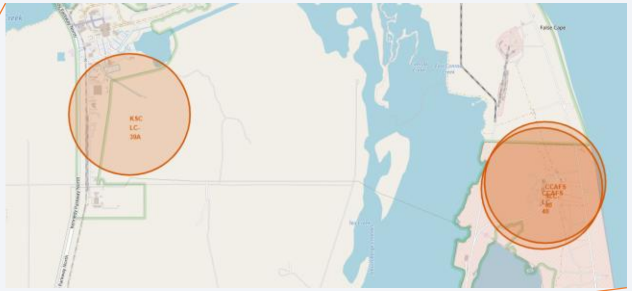
Section 3

Launch Sites Proximities Analysis

Folium Map – Launch Site Locations

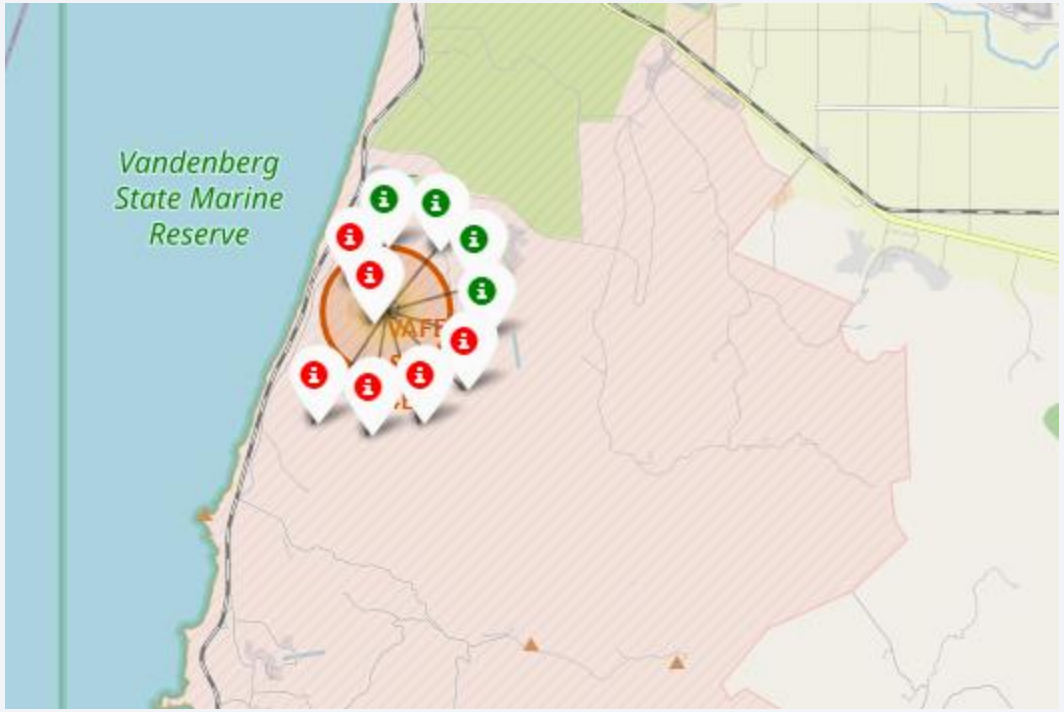
- All four launch sites are marked out on the map. Three of them, CCAFS LC-40, KSC LC-39A, and CCAFS SLC-40 are close to each other and therefore are not distinguishable in the zoomed-out view
- All launch sites are along the coast of the US





Folium Map – Launch Site Outcomes

- The screenshot shows the mission outcomes (successful or unsuccessful landing) for the VFAB SLC-40 launch site
  - Successful = green
  - Unsuccessful = red
- Gives a quick and easily interpreted overview of the success rate of a given launch site



Folium Map – Launch Site Proximities

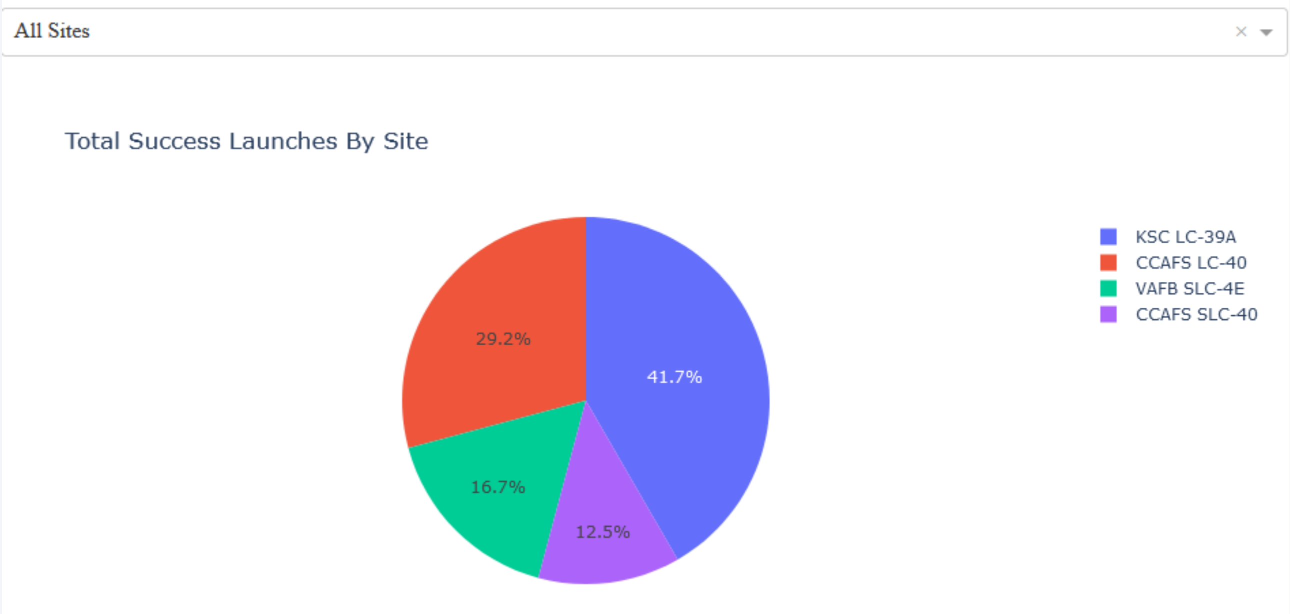
- The screenshot shows the closest relevant resources to the launch site CCAFS SLC-40, drawn with blue lines
  - Railroad in top left
  - Coastline in top right
  - Closest town in bottom
- For transportation purposes, the launch site is close to a railway
- For safety purposes, the launch site is stationed far away from the closest city and close to a coastline

Section 4

Build a Dashboard with Plotly Dash

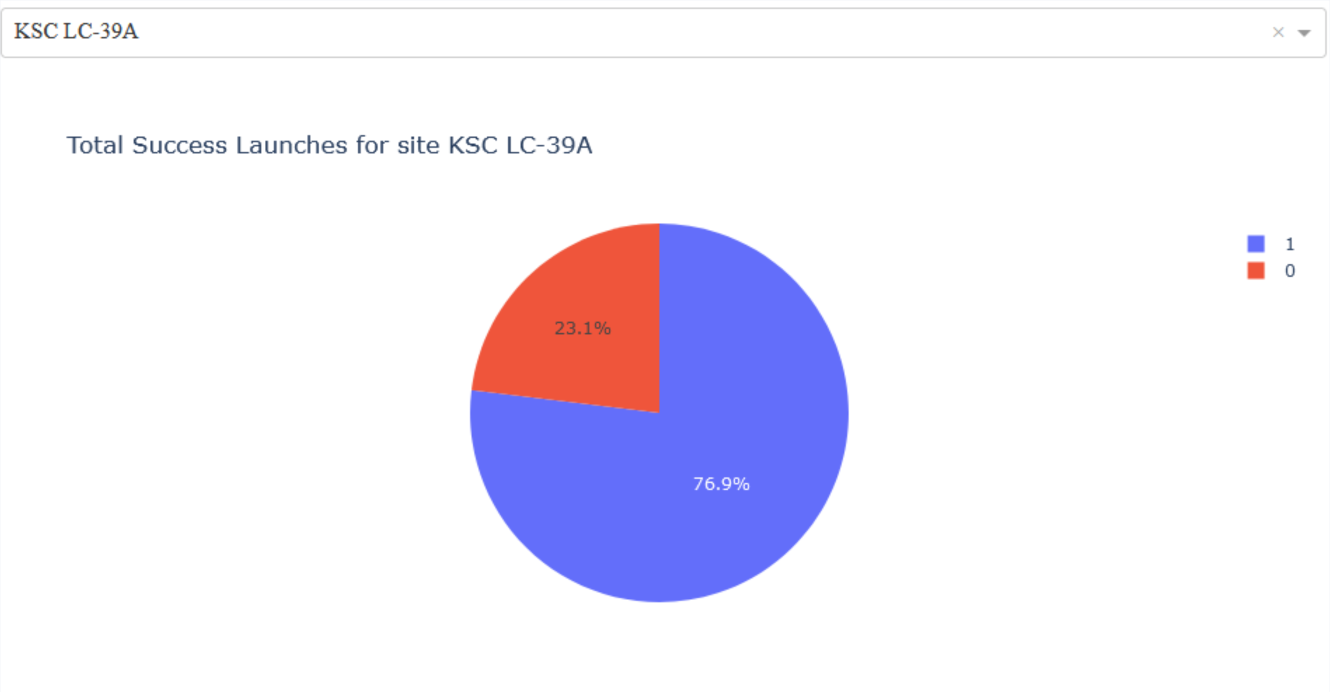
Dashboard – Successful Launches By Site

- The pie chart displays the success rate for all of the different launch sites, by all total launches
- KSC LC-39A has the highest success rate of 41.7%
- CCAFS SLC-40 with the lowest success rate of 12.5%



Dashboard – Most Successful Launch Site

- The screenshot shows the distribution between successful and unsuccessful launches for the KSC LC-39A launch site
- 76.9% of the total launches had a successful outcome (label = 1)

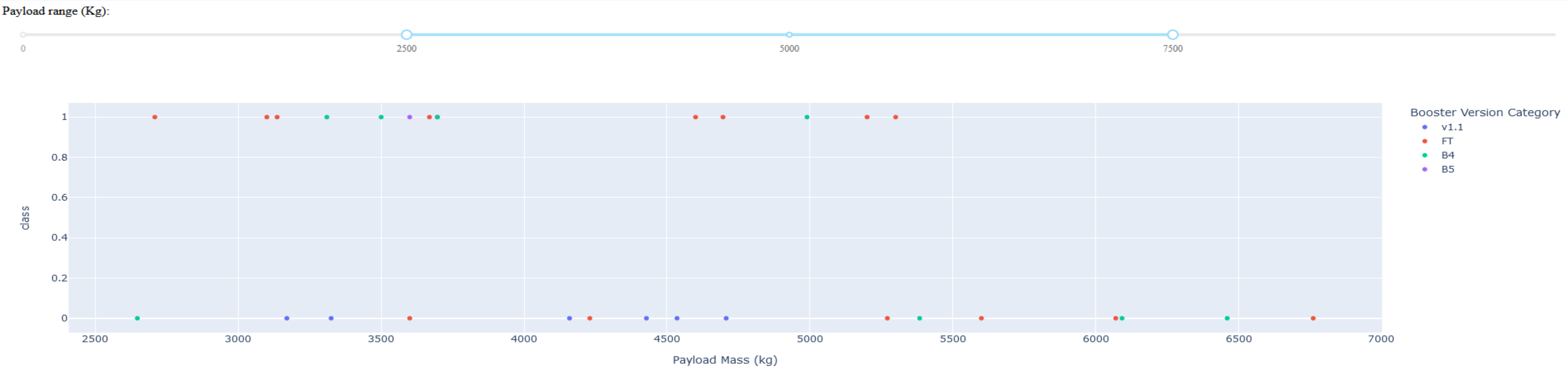


Dashboard – Payload vs. Success Rate For Booster

- The screenshot shows the mission outcome for the different boosters for all payload masses (class 1 = successful outcome, class 0 = unsuccessful outcome)
- The FT booster shows the highest amount of successful mission outcomes (specifically in the range of 2000kg and 6000kg)

Dashboard – Payload vs. Success Rate For Booster

- This screenshot instead shows the mission outcomes for boosters which carried a payload mass between 2500kg and 7500kg
- Also shows that the FT booster seems beneficial for a successful mission outcome

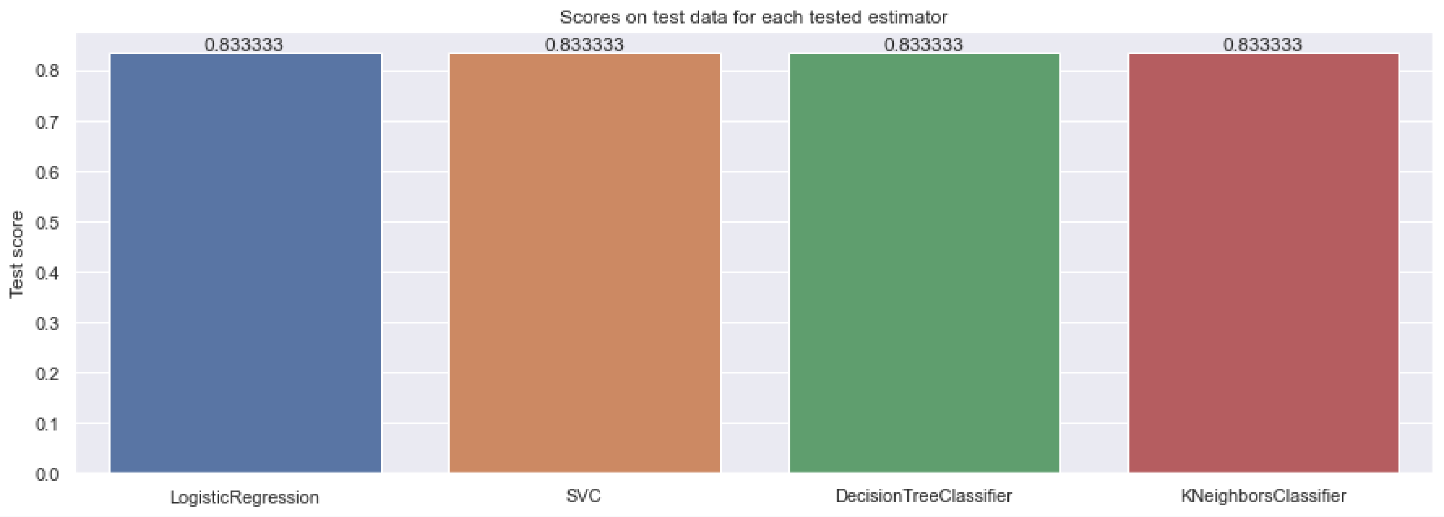


Section 5

Predictive Analysis (Classification)

Classification Accuracy

- The bar chart shows the score on the test set for the four different ML models
- All models performed equally as good on this test set



Confusion Matrix

- Since all models performed the same on the test set, the confusion matrix represents the performance of all models
- All predictions of a positive landing outcome were correct (100% recall rate)
- The false negatives and true negatives showed a 50% accuracy respectively
- The most important prediction to accurately predict is whether the mission will have a positive outcome

Conclusions

- Seems to be important for a launch site to be close to the coast, railway for transportation of the rocket itself, and to have good distance from closest city
- According to the data, launch site KSC LC-39A is the site to pick in order to have the highest probability of a successful mission outcome
    - Best to pick a payload between 2000kg and 6000kg
    - Use the FT booster since it has a historical record of the most successful mission outcomes
    - The LEO, ISS, GTO, and SO orbit types all have a 100% success rate
- The data set is probably too small in order to draw any major conclusions from a machine learning perspective, as the different models tell the same story
    - But the models are 100% sure when the mission will have a successful outcome

Appendix - Notebooks

- SpaceX API notebook
- Webscraping notebook
- Data wrangling notebook
- EDA Data visualization notebook
- EDA SQL notebook
- Launch Sites Location Analysis notebook
- SpaceX Dash App
- Landing Prediction notebook

Thank you!