IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Jan Swidzinski
25.01.2024

Executive Summary

- Summary of methodologies
  - Data collection
  - Data Wrangling
  - EDA with Data visualization
  - EDA with SQL
  - Building an Interactive map with folium
  - Building a Dashboard with Plotly Dash
  - Predictive analysis (Classification)
- Summary of all results
  - EDA results
  - Interactive analytics
  - Predictive analytics

Introduction

- Project background and context
  - While other companies' rocket launches cost more than 165 million dollars, SpaceX's rocket Falcon 9 launches, thanks to the reusage of the first stage, cost only around 62 million dollars,.
- Problems you want to find answers
  - The problem is to predict whether the first stage of the SpaceX Falcon 9 rocket will land successfully

Section 1

Methodology

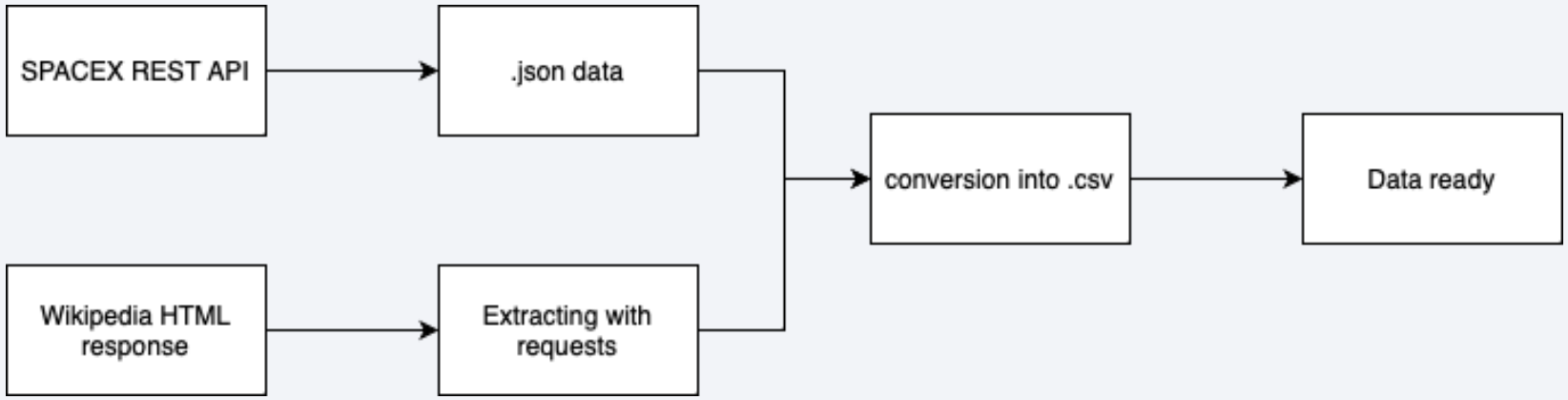
Methodology

Executive Summary

- Data collection methodology:
  - SpaceX API
  - Web scraping from Wikipedia
- Perform data wrangling
  - Data cleaning of null and unused columns and one hot encoding for categorical data in machine learning
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - KNN,SVM,LR and DT models have been used and evaluated to find the best classifier

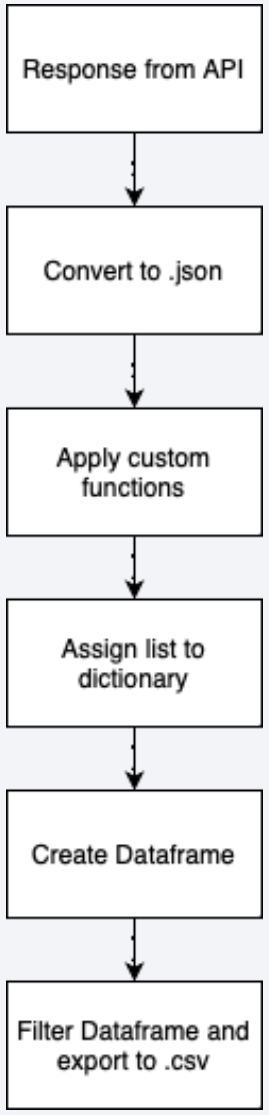
Data Collection

- Data on launches was generate from the SpaceX REST API
- API contains the data about the rocket launches, payloads, type of rockets, orbits, landing information as well as whether it was successful or not.
- The other source of data was Wikipedia page on Falcon 9 launches
- Webscraping was done using BeautifulSoup.



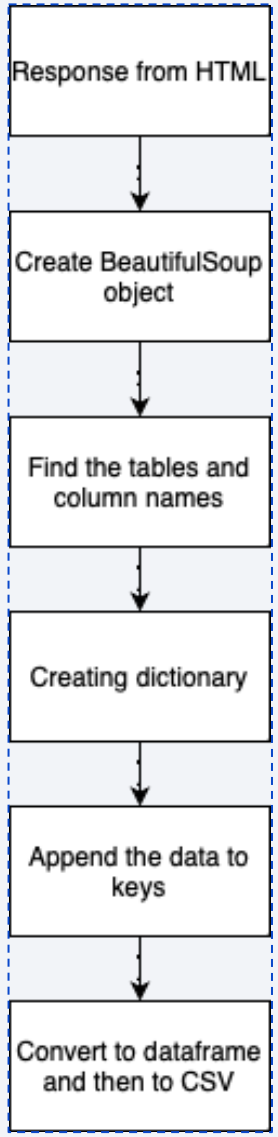
Data Collection – SpaceX API

- https://github.com/UltrOx/SpaceX-Data-Science/blob/main/jupyter-labs-spacex-data-collection-api.ipynb



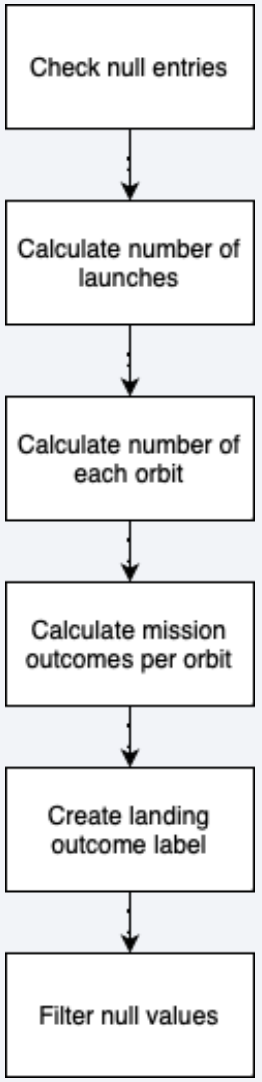
Data Collection - Scraping

- https://github.com/Ultr0x/SpaceX-Data-Science/blob/main/jupyter-labs-webscraping.ipynb



Data Wrangling

- https://github.com/UltrOx/SpaceX-Data-Science/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb



EDA with Data Visualization

- Various scatter plots that show correlations between various variables
- Bar plot that shows success rate for each orbit
- Line chart showing the average launch success rate over the years
- https://github.com/Ultr0x/SpaceX-Data-Science/blob/main/jupyter-labs-eda-dataviz.ipynb

EDA with SQL

- Task 1: Display the names of the unique launch sites in the space mission.
- Task 2: Display 5 records where launch sites begin with the string 'CCA'.
- Task 3: Display the total payload mass carried by boosters launched by NASA (CRS).
- Task 4: Display average payload mass carried by booster version F9 v1.1.
- Task 5: List the date when the first successful landing outcome on a ground pad was achieved.
- Task 6: List the names of the boosters which have landed successfully on a drone ship and have a payload mass greater than 4000 but less than 6000.
- Task 7: List the total number of successful and failure mission outcomes.
- Task 8: List the names of the booster_versions which have carried the maximum payload mass.
- Task 9: List the records which will display the month names, failure landing_outcomes on a drone ship, booster versions, launch site for the months in the year 2015.
- Task 10: 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.
- https://github.com/Ultr0x/SpaceX-Data-Science/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- I added the closes city market, the waterfront as well as train pointer to indicate geopgraphical as well as human made structures.
- https://github.com/UltrOx/SpaceX-Data-Science/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

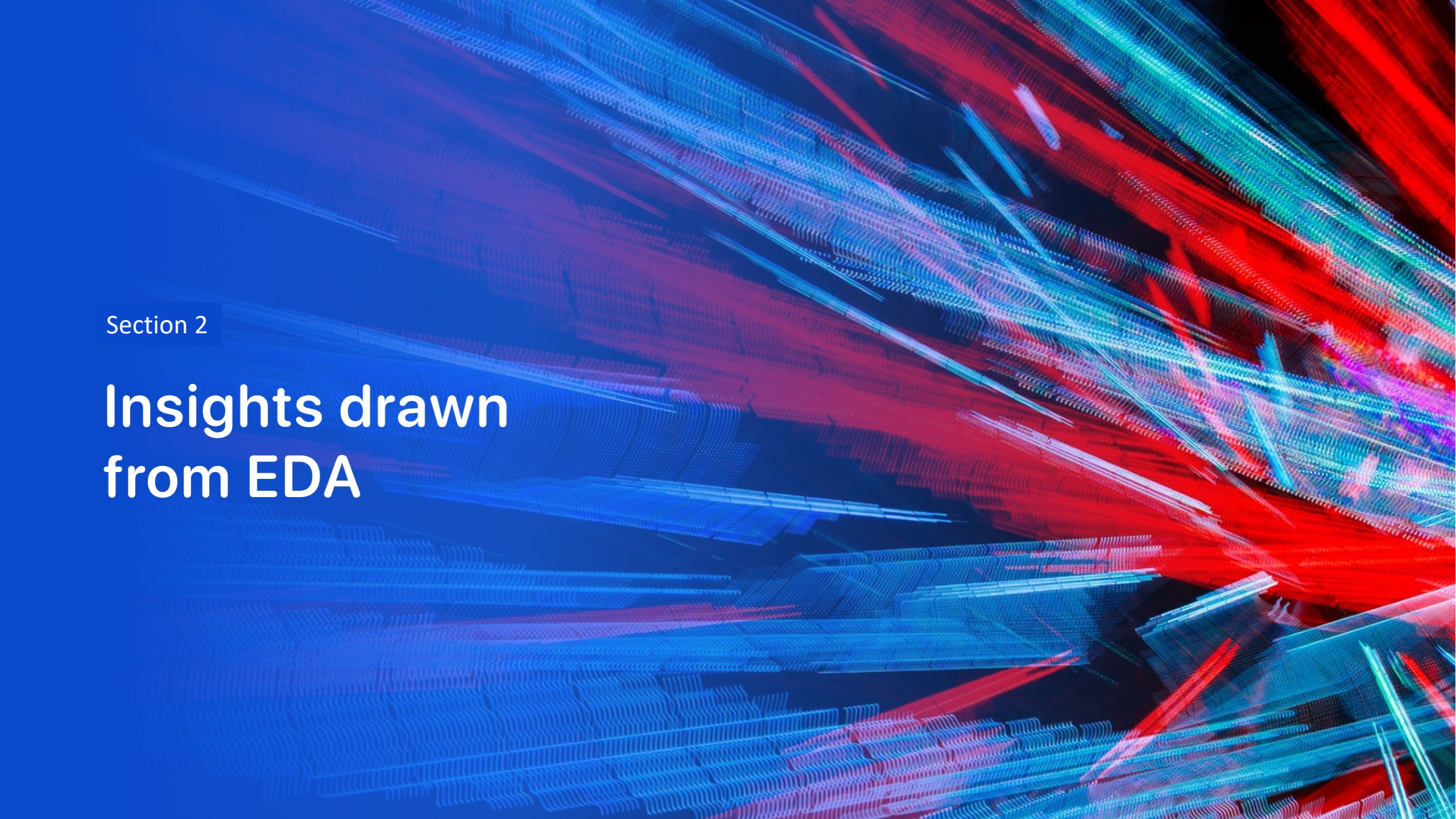
- I added Pie graphs that indicate the success rate for all and for distinct sites
- I also added scatter plots that show the payloads and their classes correlating to weights for low and heavy payloads
- https://github.com/UltrOx/SpaceX-Data-Science/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

- SVM, KNN and logistic regression had all accuracy rate of 83.3 percent while Decision trees had highest accuracy of 88.8 percent having only one FP and one FN entries.
- https://github.com/UltrOx/SpaceX-Data-Science/blob/main/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

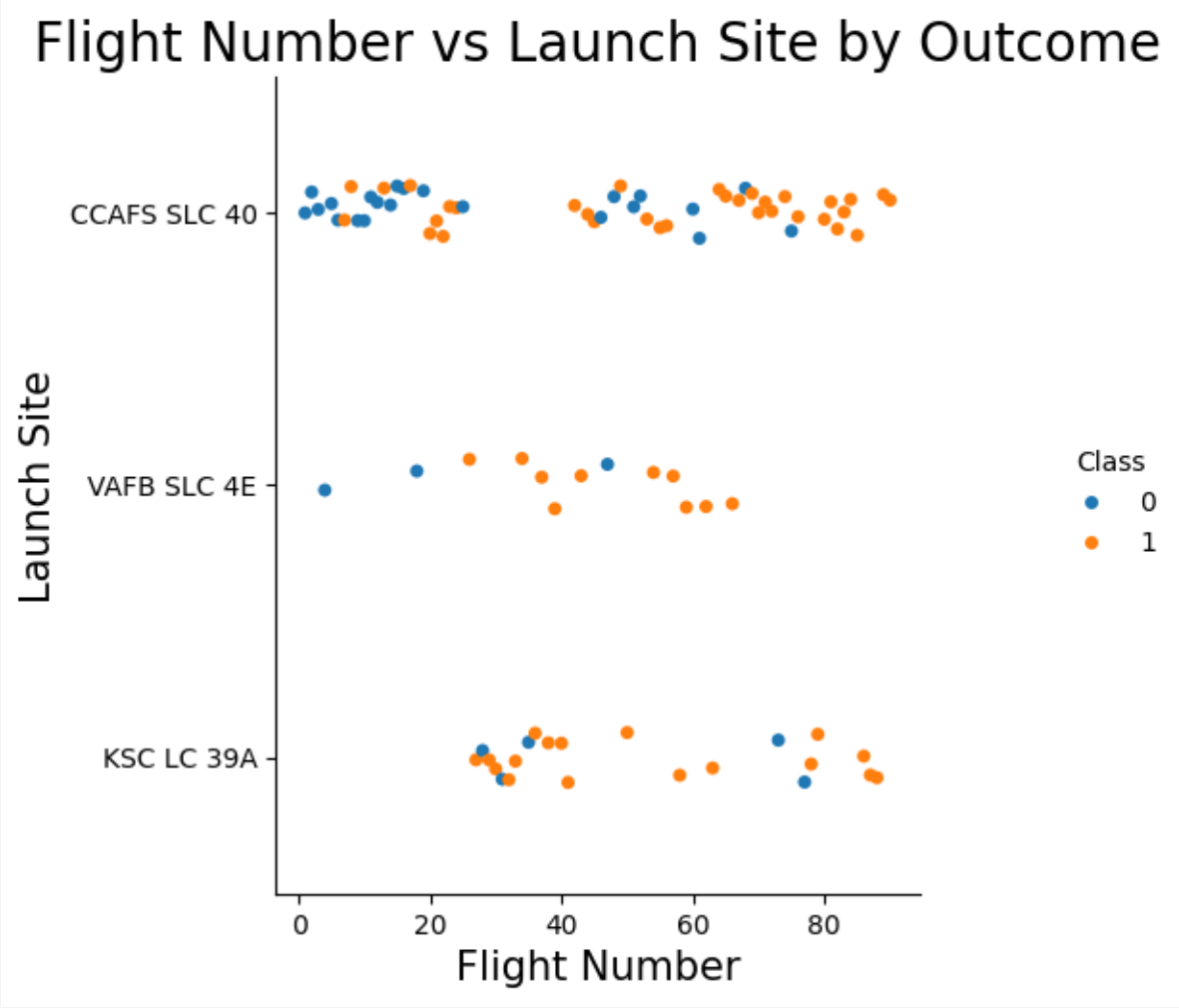
Results

- Payloads with lower masses had higher performance rates than heavier payloads
- Average launch success rate is directly proportional to time, becoming higher each year
- KSC LC 39A had largest amount of successful launches
- Orbit GEO,HEO.SSO.ES L1 has the largest success rate
- Decision tree model had highest classification accuracy


Section 2
Insights drawn from EDA

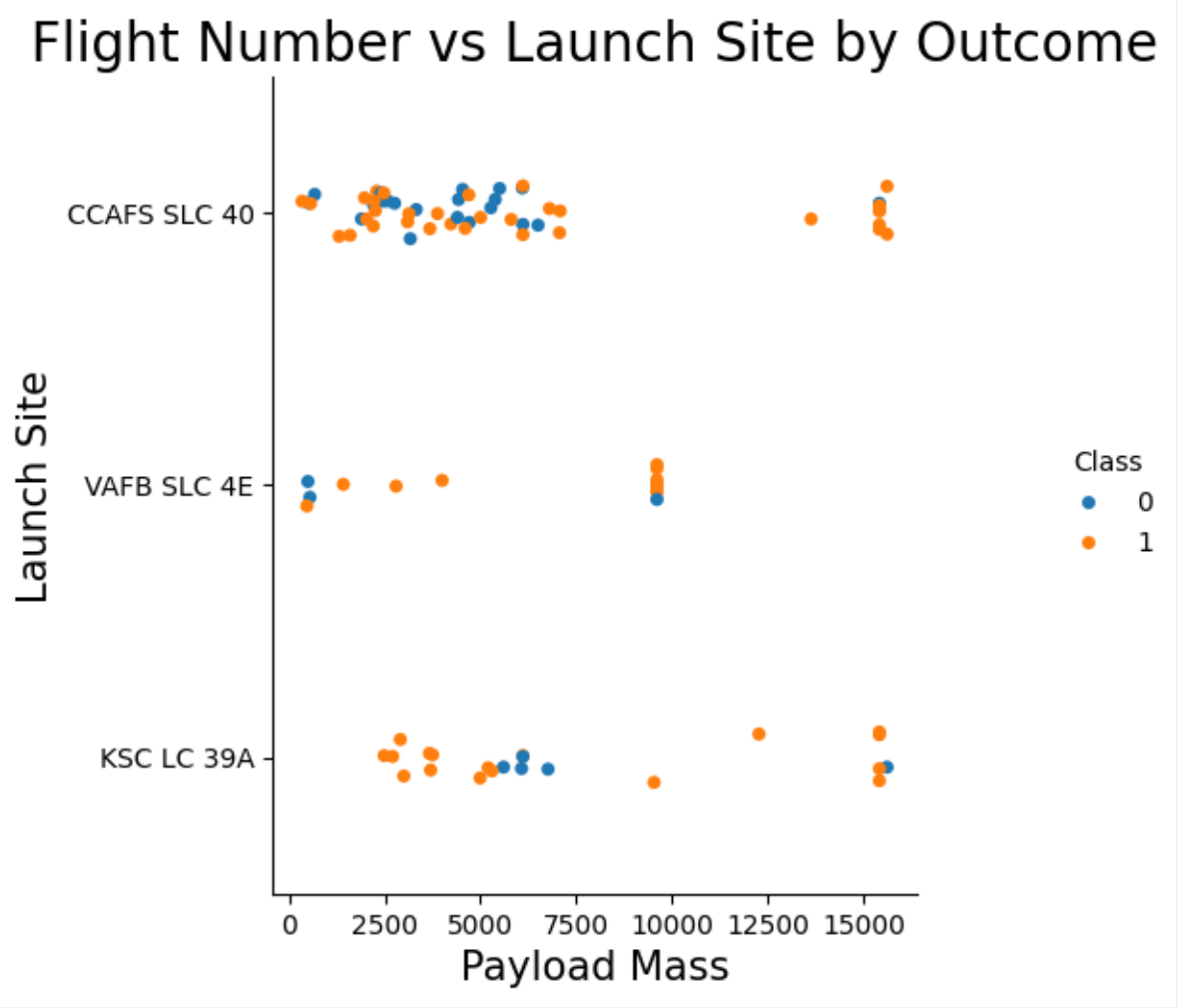
Flight Number vs. Launch Site

- CCAFS SLC 40 flight Numbers ahve the highest range and most high glihts
- VAFB SLC 4E has the smalles amount of Flights
- ALL Launch sites have both classes



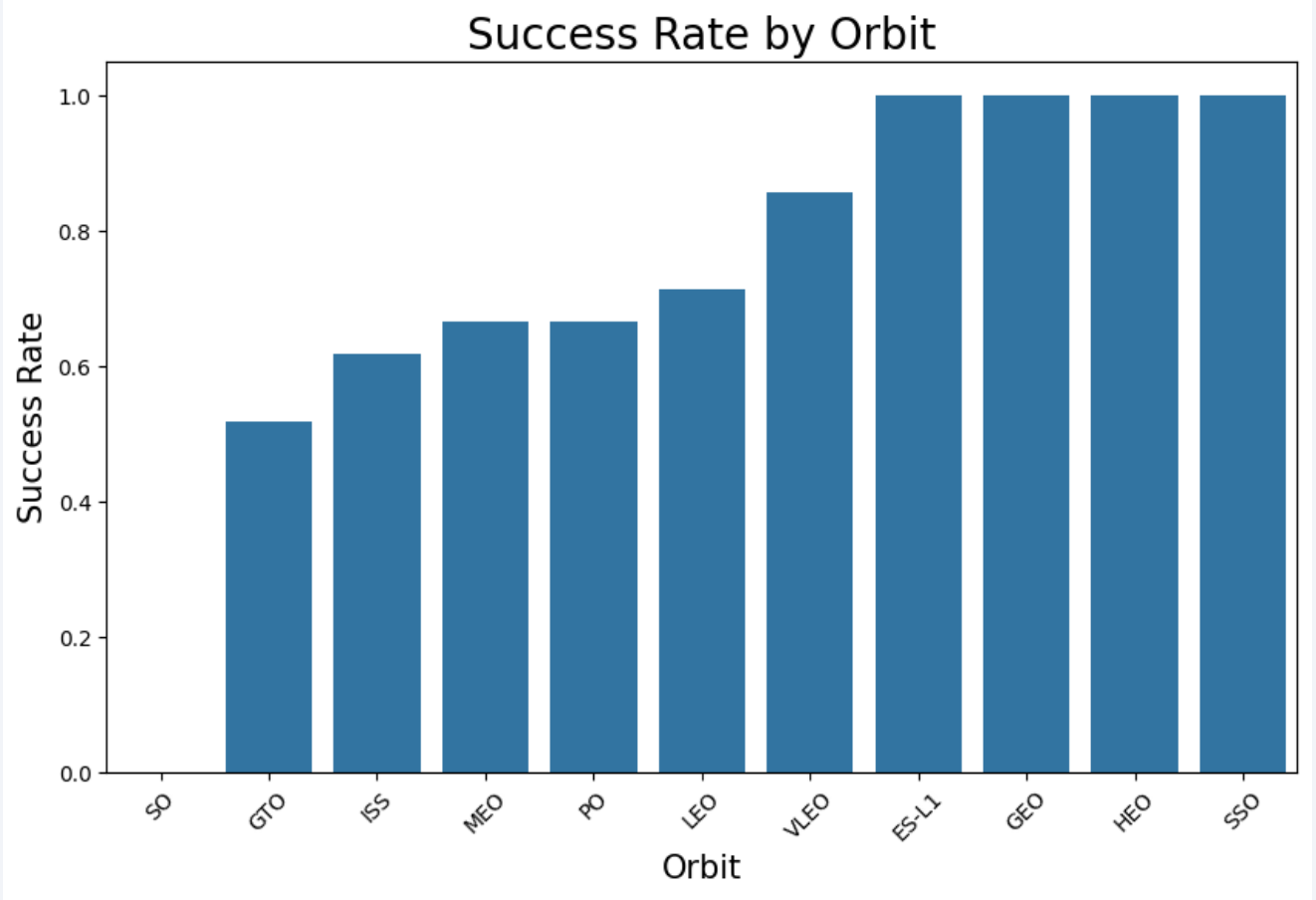
Payload vs. Launch Site

- Most Payloads with low mass have been launched from CCAFS SLC 40
- VAFV SLC 4E launches have smallest maximum payload mass



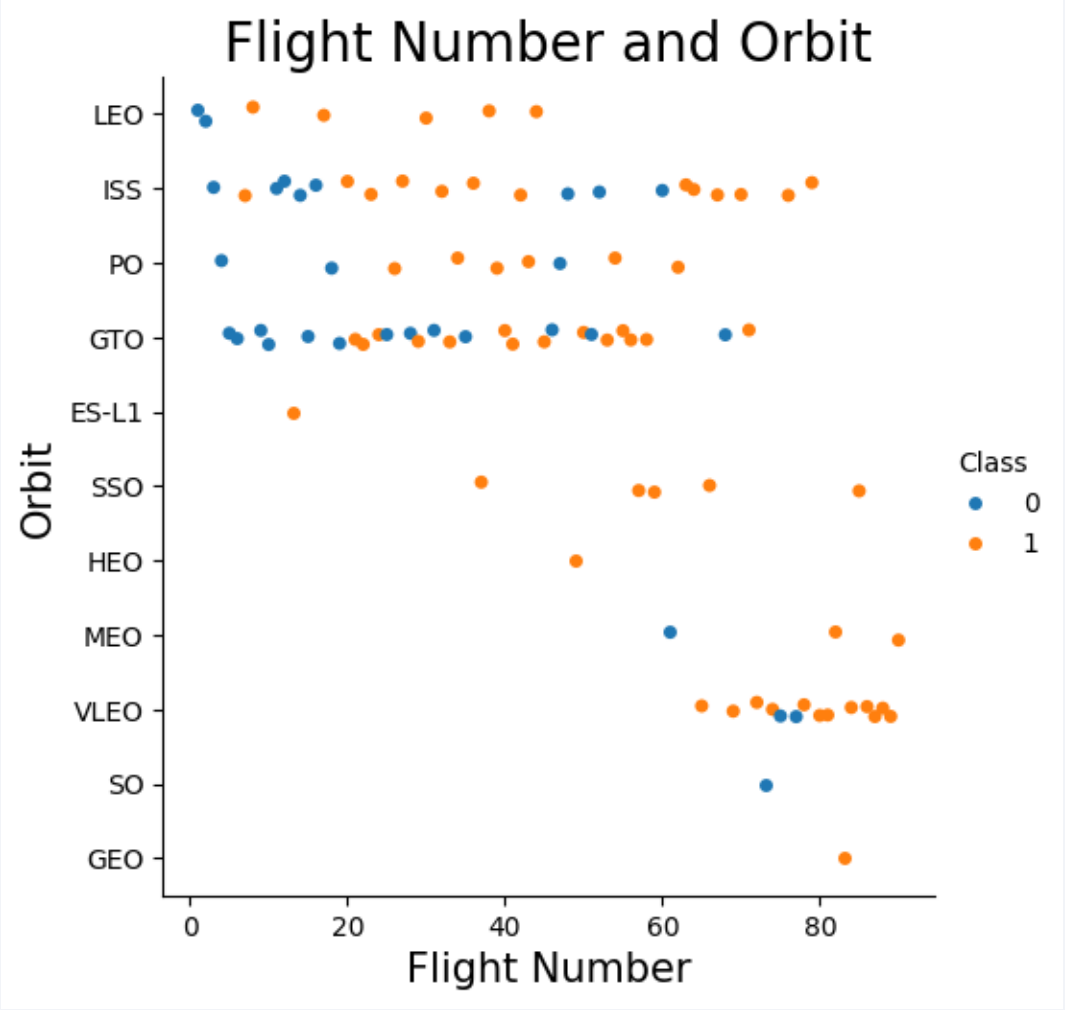
Success Rate vs. Orbit Type

- ES-L1, GEO, HEO and SSO had highest,100% success rate



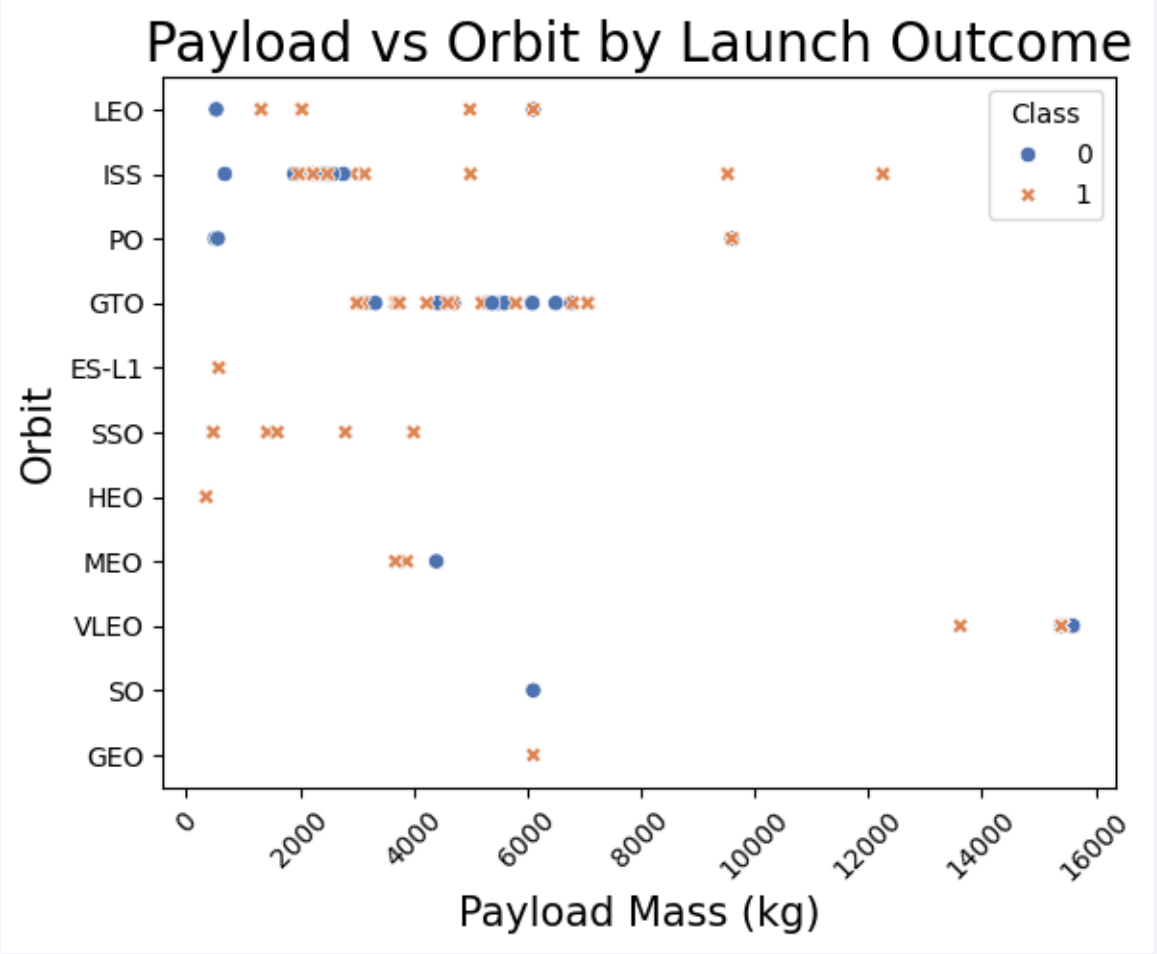
Flight Number vs. Orbit Type

- It can be seen that LEO, ISS, PO and GTO are being replaced, especially to VLEO launch



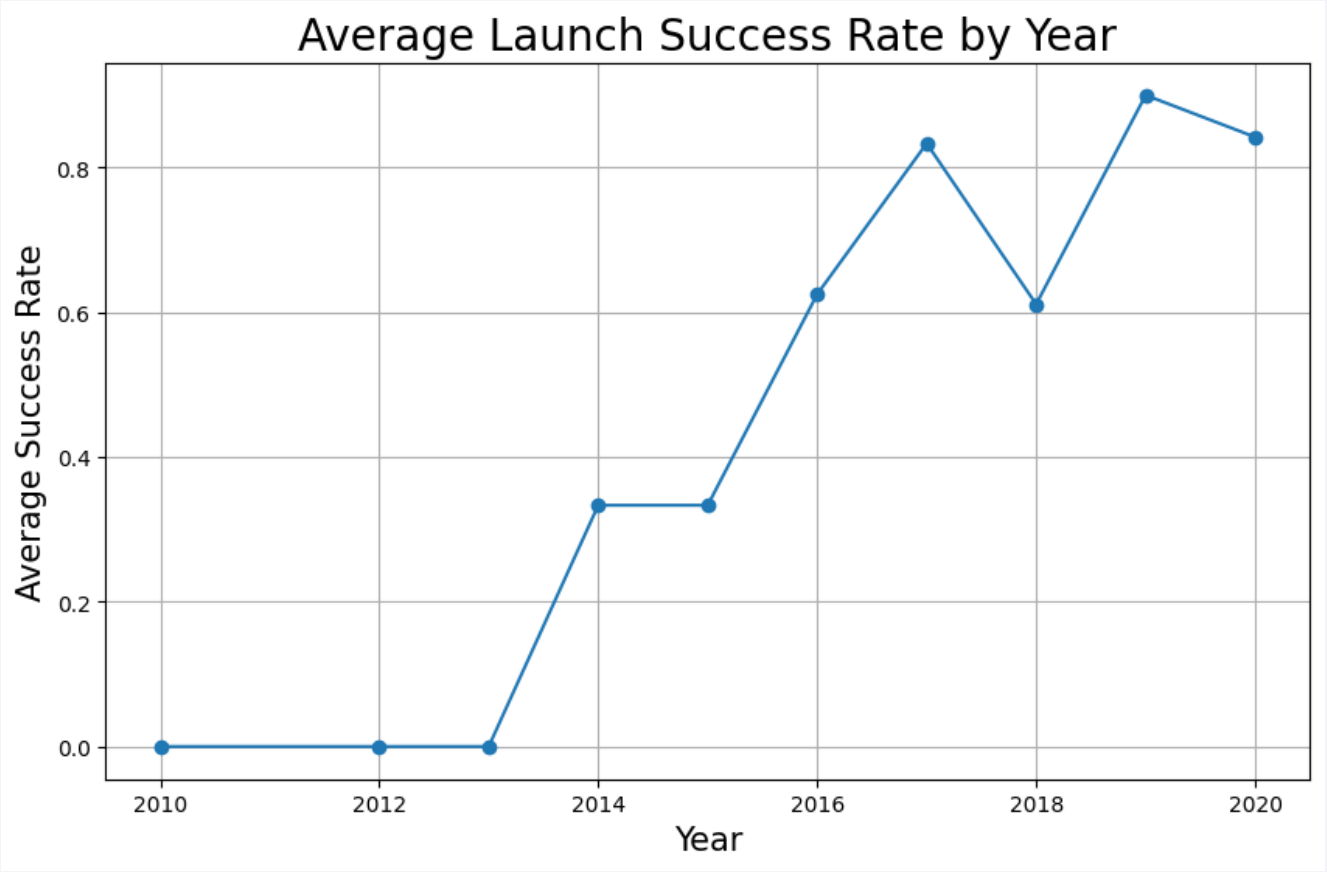
Payload vs. Orbit Type

- A strong correlation between ISS and ranges 2000-4000 and between GTO and ranges 3000-8000 as well as VLEO and around 16000.



Launch Success Yearly Trend

- Launch rate has incrasead dramatically from 2013 to 2019 coming at more than 85 percent showing the possible advancment in tehchnology and experience gained.



All Launch Site Names

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

Launch Site Names Begin with 'CCA'

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (p
2010-12-08	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 (p
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	N
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	N
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	N

Total Payload Mass

total_payload_mass
48213

Average Payload Mass by F9 v1.1

2534.6666666666665

First Successful Ground Landing Date

first_successful_landing_date
2015-12-22

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

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

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

2015 Launch Records

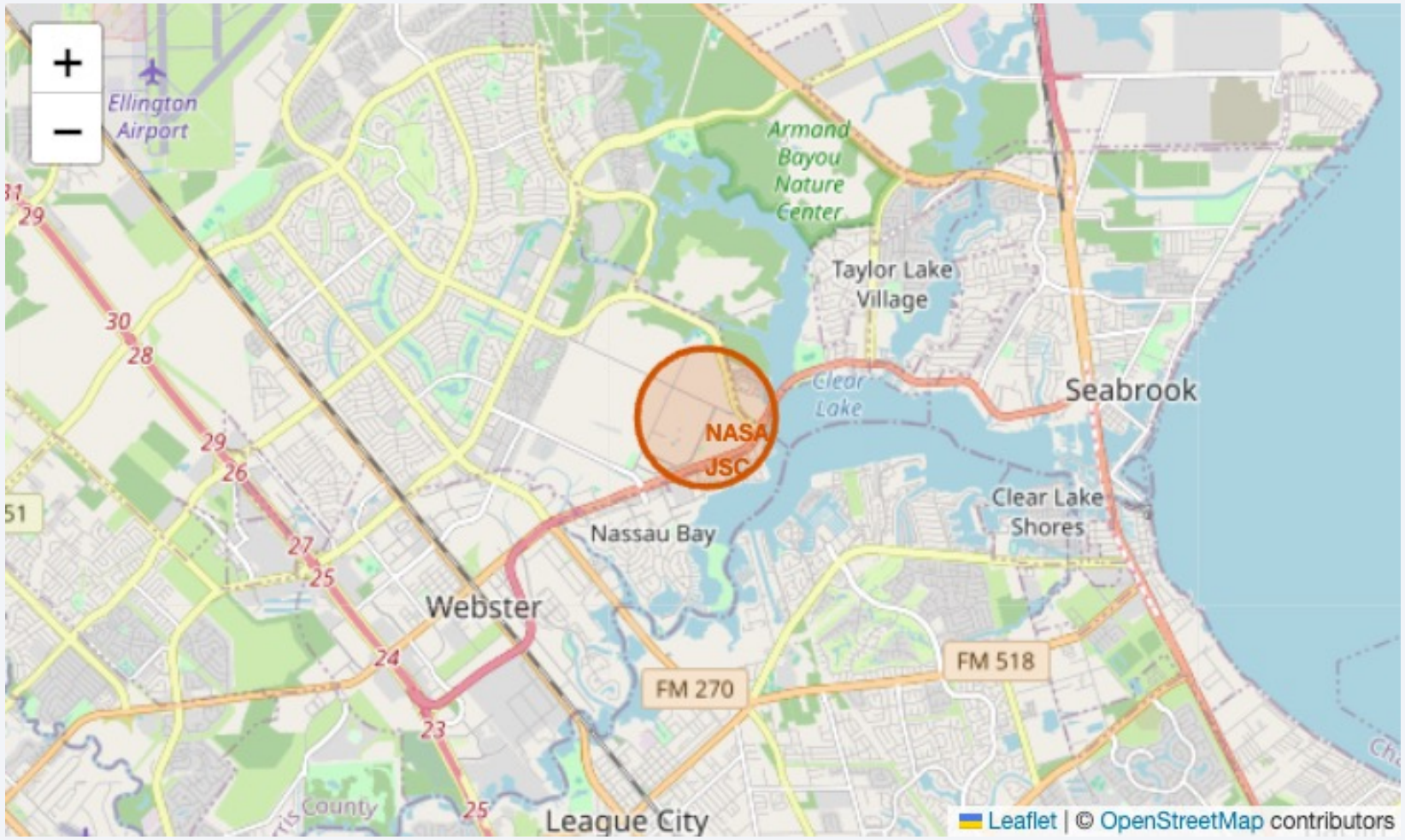
month	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

Landing_Outcome	count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

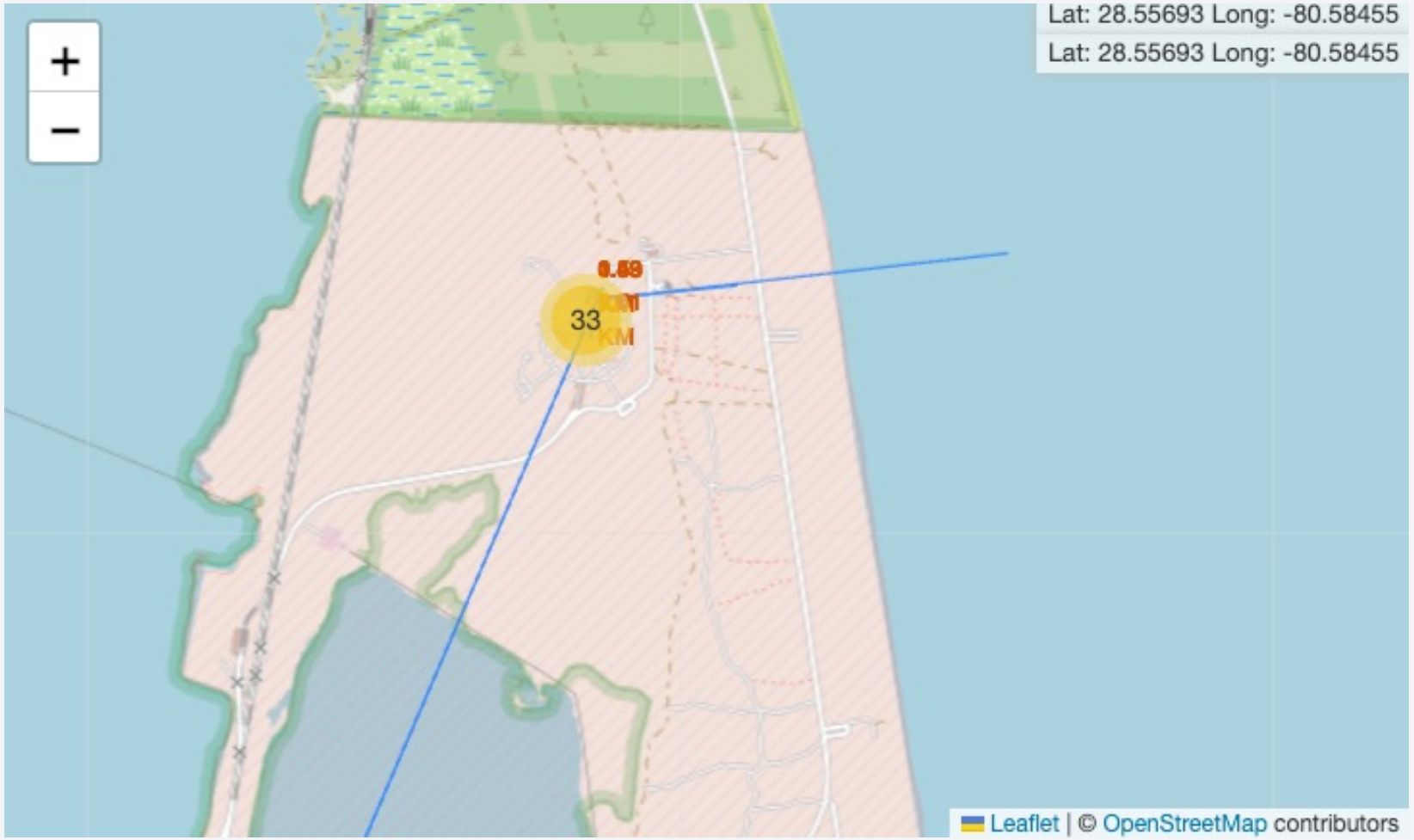

Section 3
Launch Sites
Proximities Analysis

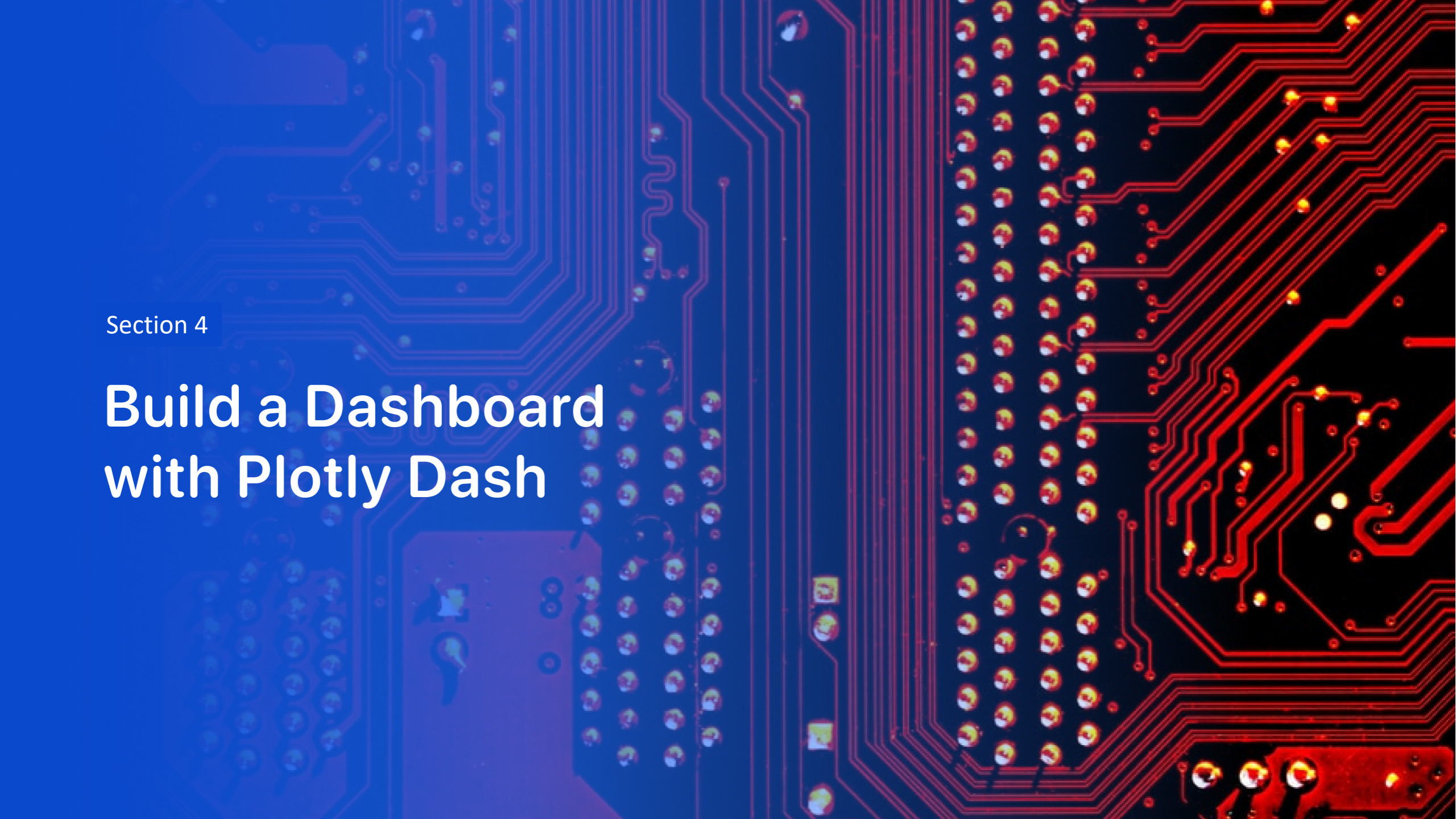
Launch site marked on the map



All the launch sites marked on the map

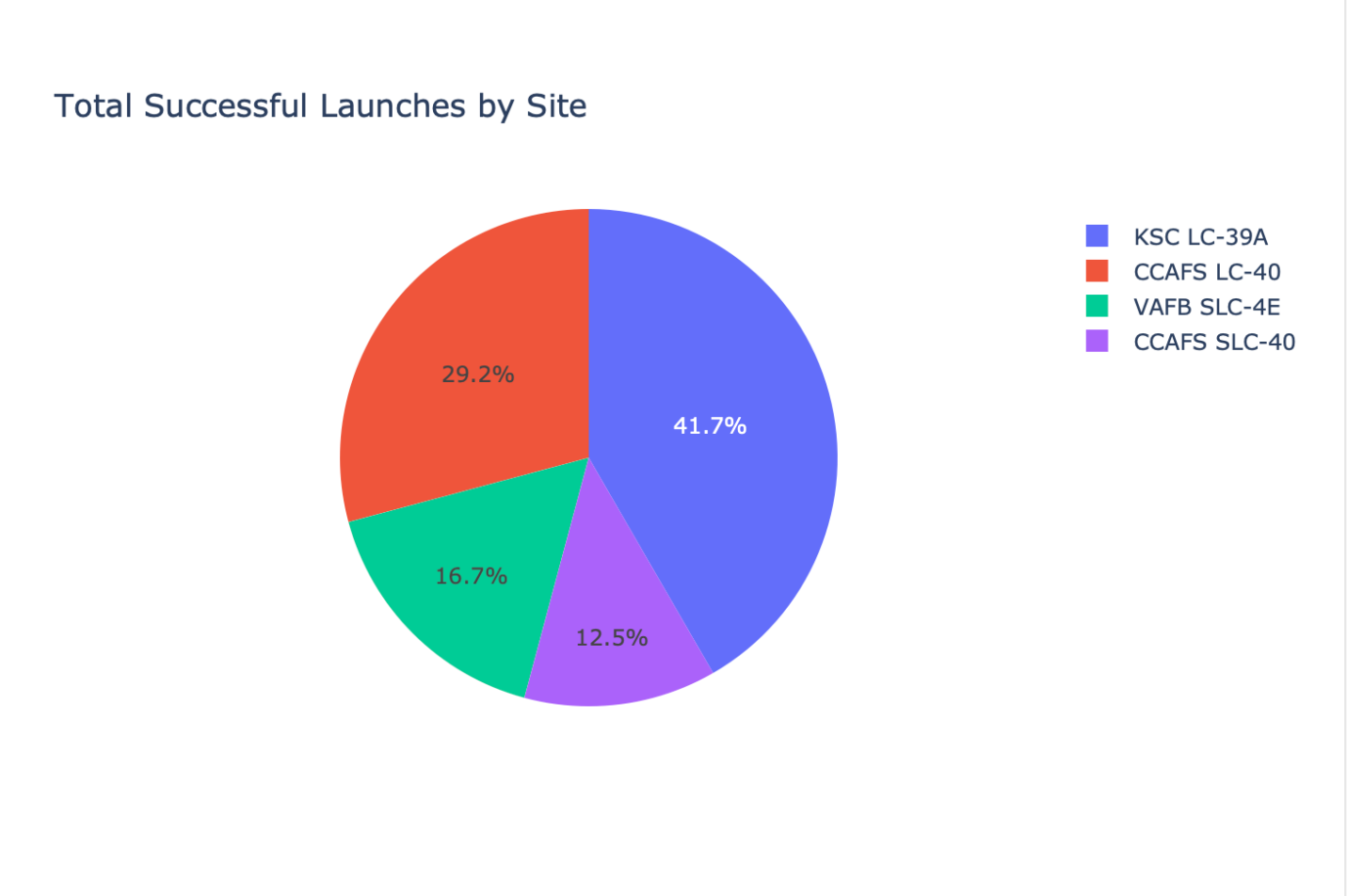
Distances between the launch site and other points




Section 4
Build a Dashboard with Plotly Dash

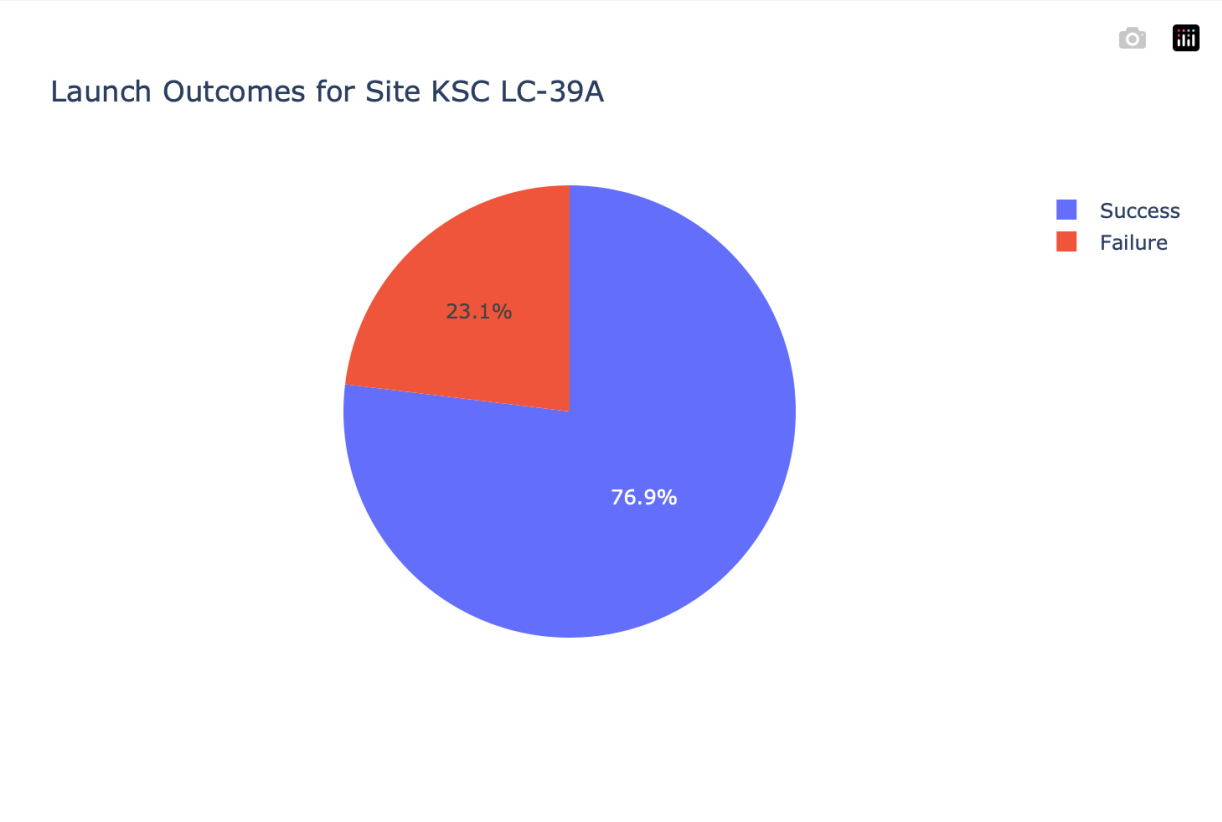
Total Successful launches

- KSC LC-39A had the most successful launches

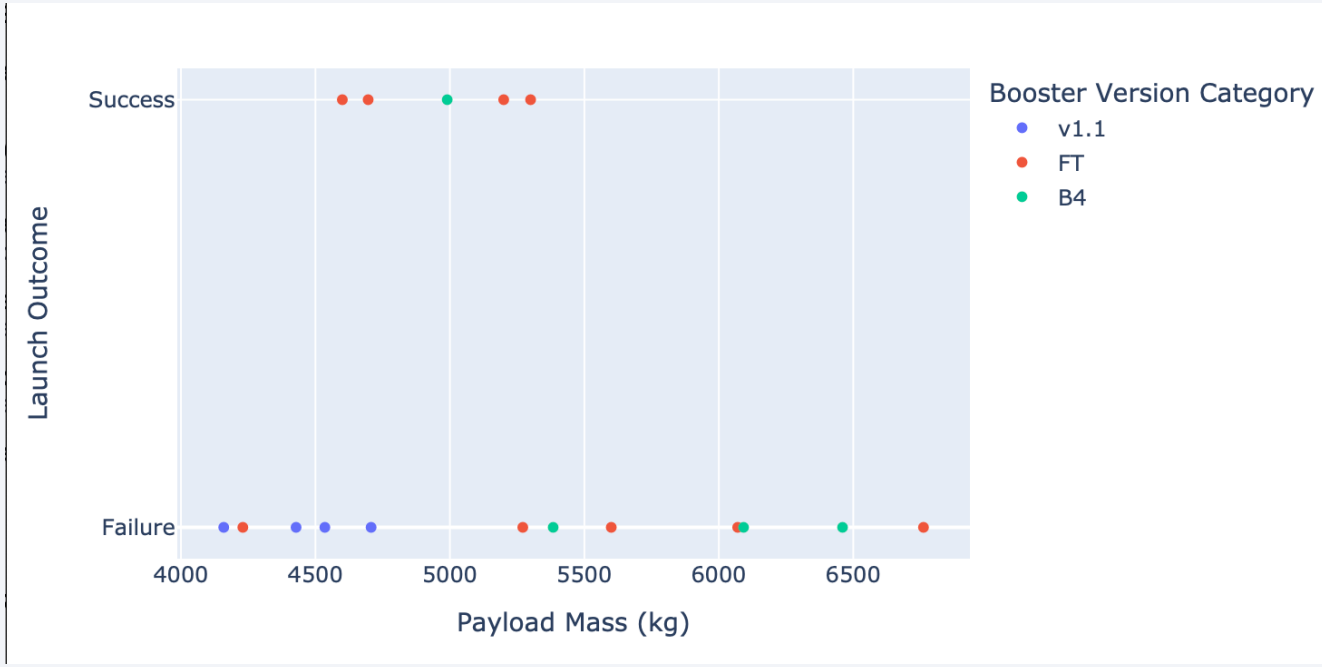


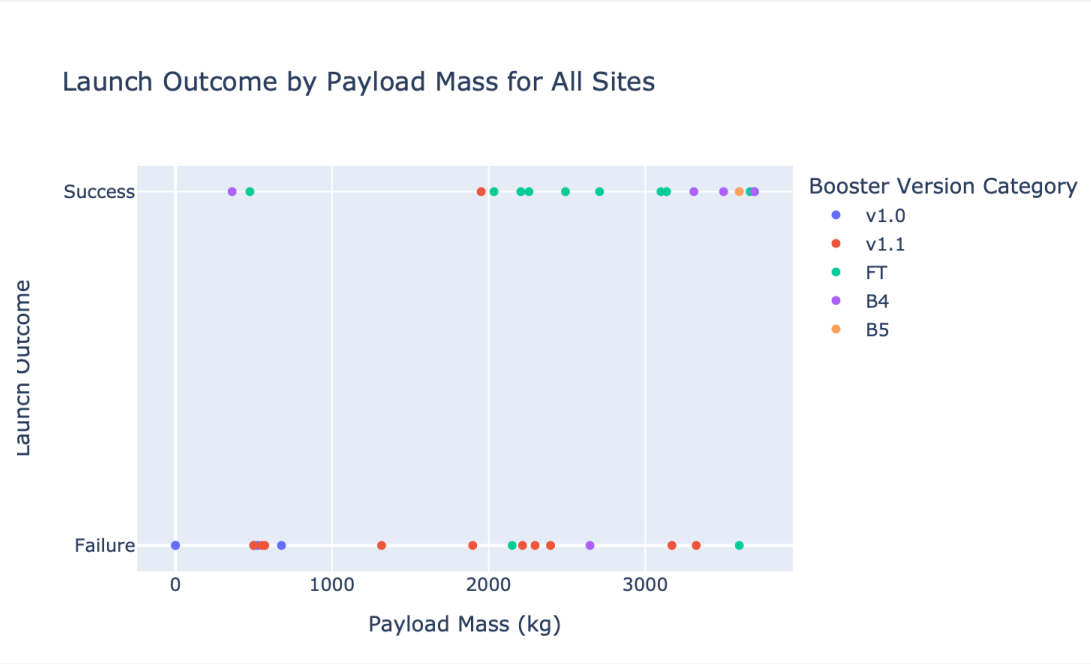
Launches for KSC LC-39A

- KSC-IC-39A has 76.9% success and 23.1% failure rate



Payload vs launch outcome



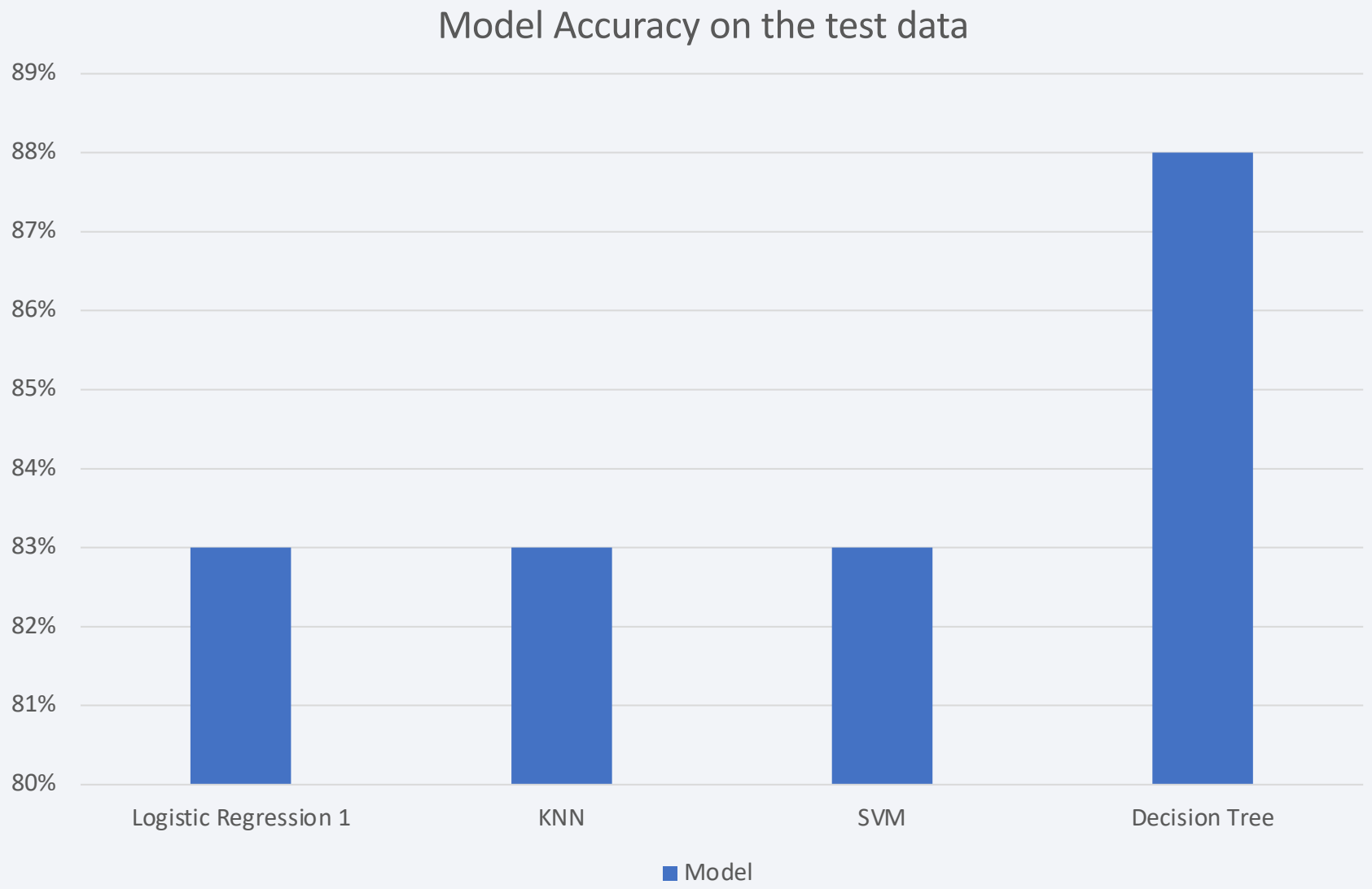


- We can see that the success rate for the payloads with lower weights is much higher that that for the heavy payloads

Section 5

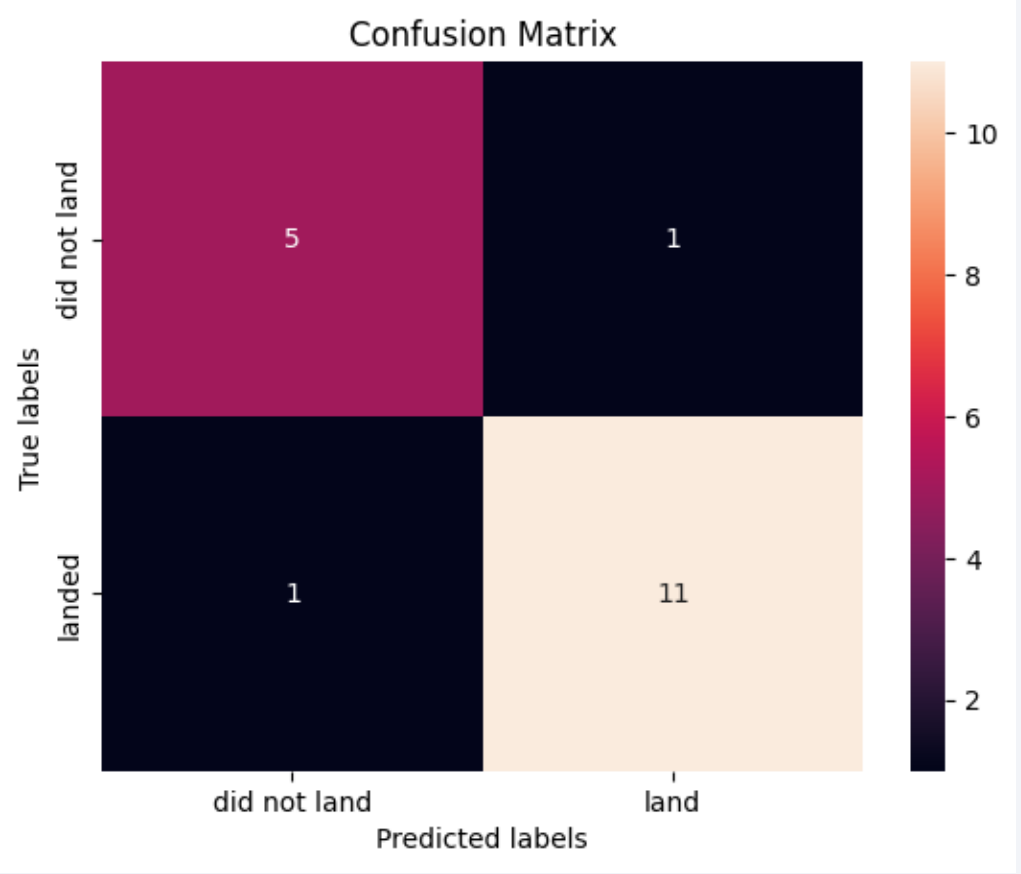
Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix

- Decision Tree model had the highest accuracy
- Confusion Matrix shows that it had only two predictions wrong, one false positive and one false negative



Conclusions

- Payloads with lower masses had higher performance rates than heavier payloads
- Average launch success rate is directly proportional to time, becoming higher each year
- KSC LC 39A had largest amount of successful launches
- Orbit GEO,HEO.SSO.ES L1 has the largest success rate
- Decision tree model had highest classification accuracy

Thank you!