IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Ahmad Karrabi

May 2023

Github : https://github.com/karrabi/IBM-Data-Science-Capstone-Project

Outline

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

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 analysis

Introduction

- Project background and context
- SpaceX advertises Falcon 9 rocket launches on its website, with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, most of the savings is because SpaceX can reuse the first stage.

- Problems you want to find answers
- The project task is to predicting if the first stage of the SpaceX Falcon 9 rocket will land successfully

Section 1

Methodology

Methodology

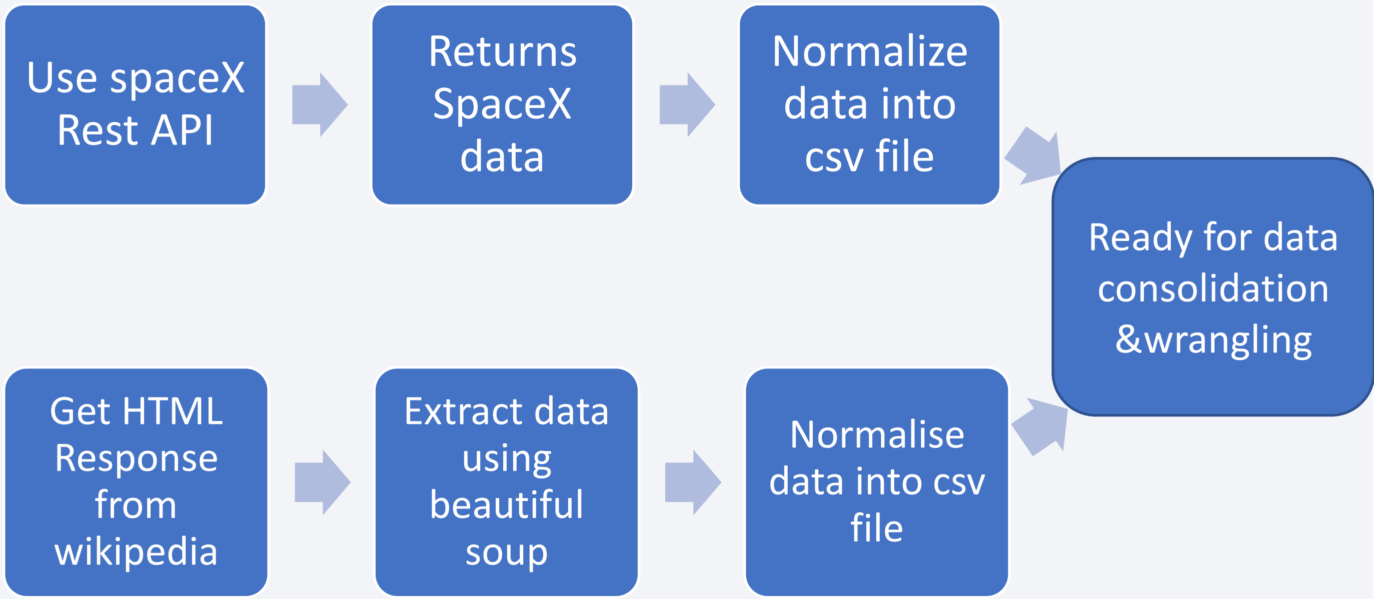
Executive Summary

- Data collection methodology:
  - SpaceX Rest API
  - Web Scrapping from SpaceX Wikipedia page
- Perform data wrangling
  - Hot Encoding data fields for Machine Learning
  - data cleaning of null values and irrelevant columns
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Logistic Regression, KNN, SVM, DT models have been built and evaluated for the best classifier

Data Collection

- Describe how data sets were collected.
- The following datasets was collected:
- SpaceX launch data that is gathered from the SpaceX REST API.
- This API will give us data about launches, including information about the rocket used, payload delivered, launch specifications, landing specifications, and landing outcome.
- The SpaceX REST API endpoints, or URL, starts with api.spacexdata.com/v4/.
- Another popular data source for obtaining

Falcon 9 Launch data is web scraping Wikipedia usingBeautifulSoup.



Data Collection – SpaceX API

- Data collection with SpaceX REST calls

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs
4	1	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False
5	2	2012-05-22	Falcon 9	525.0	LEO	CCSFS SLC 40	None None	1	False	False	False

https://github.com/karrabi/IBM-Data-Science-Capstone-Project/blob/main/Week%201/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping

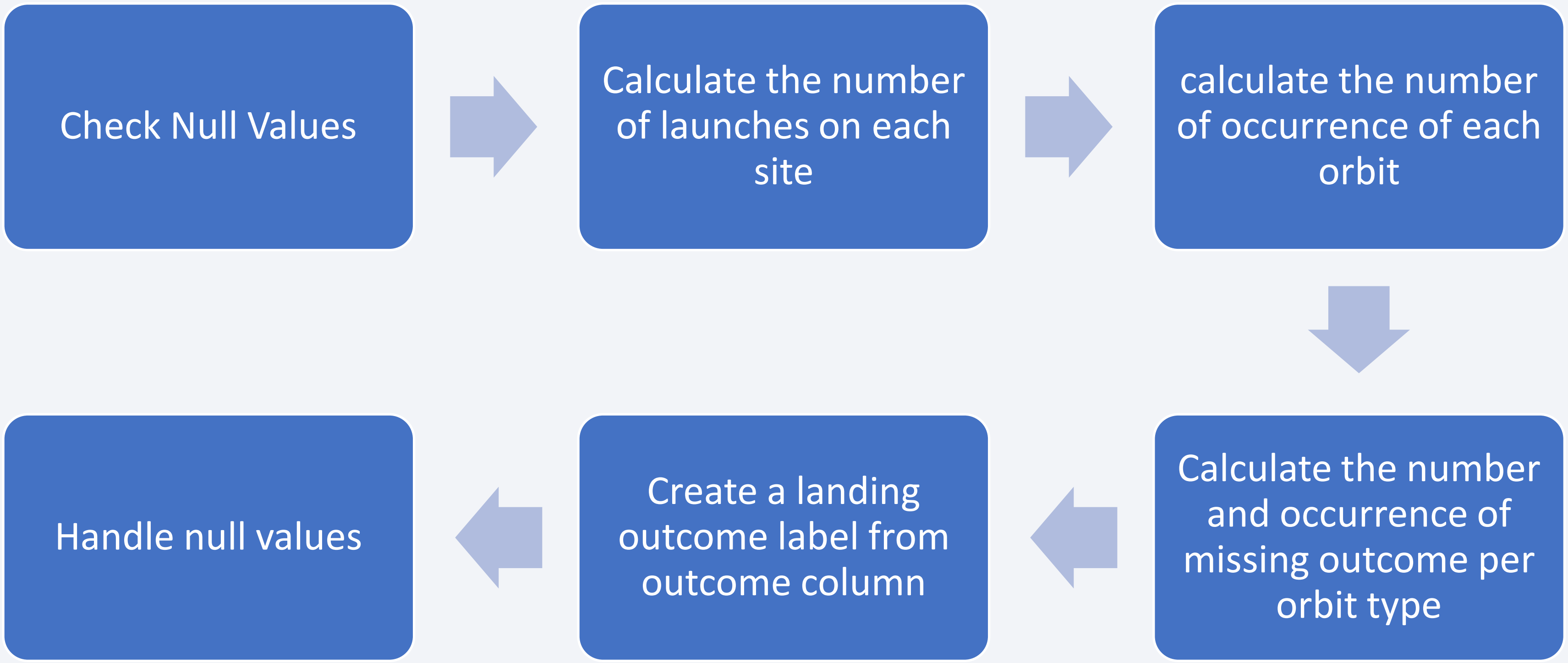
- Web scrapping from wikipedia

[hide] Flight No.	Date and time (UTC)	Version, Booster[b]	Launch site	Payload[c]	Payload mass	Orbit	Customer	Launch outcome	Booster landing
21	17 January 2016, 18:42[21]	F9 v1.1 B1017[8]	VAFB, SLC-4E	Jason-3[68][104]	553 kg (1,219 lb)	LEO	NASA (LSP) NOAA CNES	Success	Failure (drone ship)
	First launch of NASA and NOAA joint science mission under the NLS II launch contract (not related to NASA CRS or USAF OSP3 contracts) and last launch of the Falcon 9 v1.1 launch vehicle. The Jason-3 satellite was successfully deployed to target orbit.[105] SpaceX attempted for the first time to recover the first-stage booster on its new Pacific autonomous drone ship, but after a soft landing on the ship, the lockout on one of the landing legs failed to latch and the booster fell over and exploded.[106][107]								

```
#   Column
---  ------
0   Flight No.
1   Launch site
2   Payload
3   Payload mass
4   Orbit
5   Customer
6   Launch outcome
7   Version Booster
8   Booster landing
9   Date
10  Time
```

https://github.com/karrabi/IBM-Data-Science-Capstone-Project/blob/main/Week%201/jupyter-labs-webscraping.ipynb

Data Wrangling



https://github.com/karrabi/IBM-Data-Science-Capstone-Project/blob/main/Week%201/labs-jupyter-spacex-data_wrangling_jupyterlite.jupyterlite.ipynb

EDA with Data Visualization

Exploratory Data Analysis performed on variables Flight Number, Payload Mass, Launch Site, Orbit, Class and Year.

Scatter plots, line charts, and bar plots used to compare relationships between :

- Flight Number vs. Payload Mass
- Flight Number vs. Launch Site
- Payload Mass vs. Launch Site
- Orbit vs. Success Rate
- Flight Number vs. Orbit
- Payload vs Orbit
- Success Yearly Trend

decide if a relationship exists so that they could be used in training the machine learning model

https://github.com/karrabi/IBM-Data-Science-Capstone-Project/blob/main/Week%202/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 'KSC
- 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 where the successful landing outcome in drone ship was achieved.
- Listing the names of the boosters which have success in ground pad 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 records which will display the month names, successful landing_outcomes in ground pad booster versions, launch_site for the months in year 2017
- Ranking the count of successful landing outcomes between the date 2010 06 04 and 2017 03 20 in descendingorder.

https://github.com/karrabi/IBM-Data-Science-Capstone-Project/blob/main/Week%202/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Map markers have been added to the map with aim to finding an optimal location for building a launch site

https://github.com/karrabi/IBM-Data-Science-Capstone-Project/blob/main/Week%203/lab_jupyter_launch_site_location.jupyterlite.ipynb

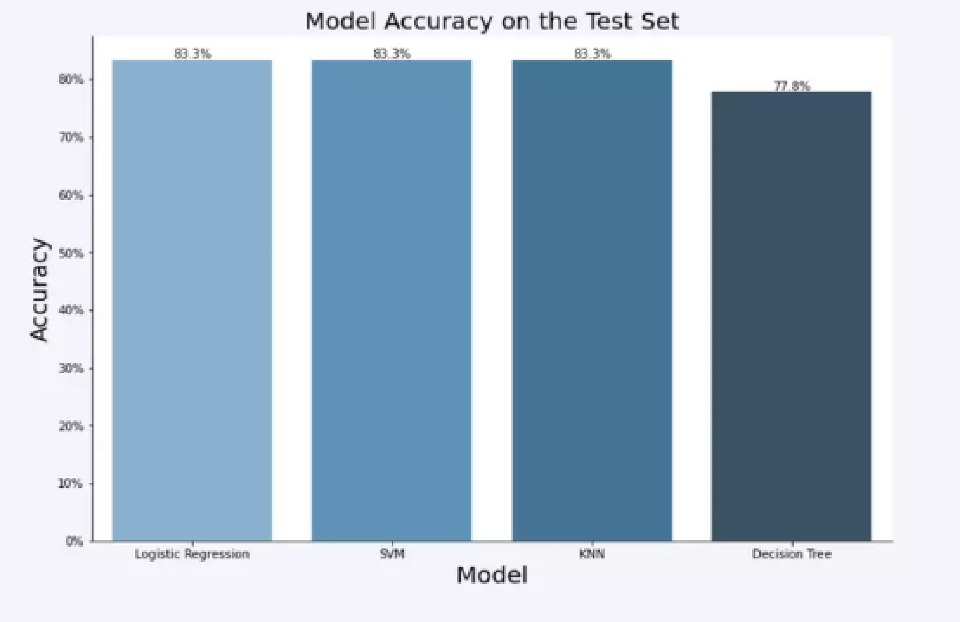
Build a Dashboard with Plotly Dash

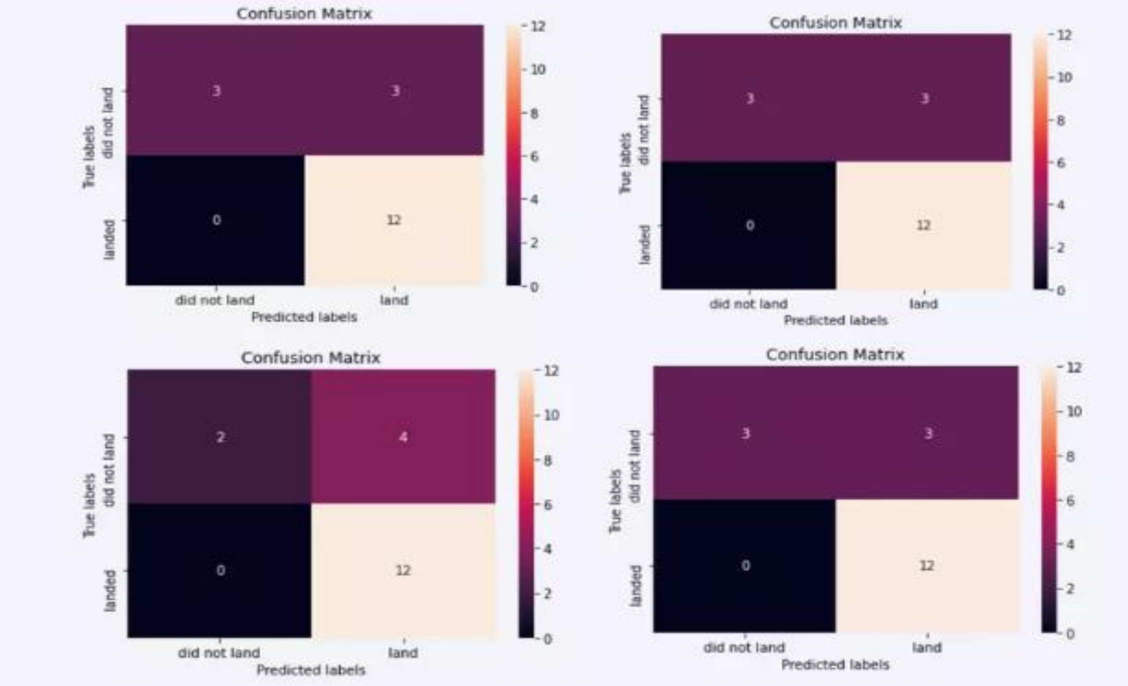
Dashboard with a selectable pie chart and a scatter plot.
Pie chart shows distribution of successful landings across all launch sites and can be selected to show individual launch site success rates.
Scatter plot takes two inputs: All sites or individual site and payload mass on a slider between 0 and 10000 kg.
The pie chart is used to visualize launch site success rate.
The scatter plot can help us see how success varies across launch sites, payload mass, and booster version category.

https://github.com/karrabi/IBM-Data-Science-Capstone-Project/blob/main/Week%203/spacex_dash_app.py

Predictive Analysis (Classification)

The SVM, KNN, and Logistic Regression model achieved the highest accuracy at 83.3%, while the SVM performs the best in terms of Area Under the Curve at 0.958.





https://github.com/karrabi/IBM-Data-Science-Capstone-Project/blob/main/Week%204/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

Results

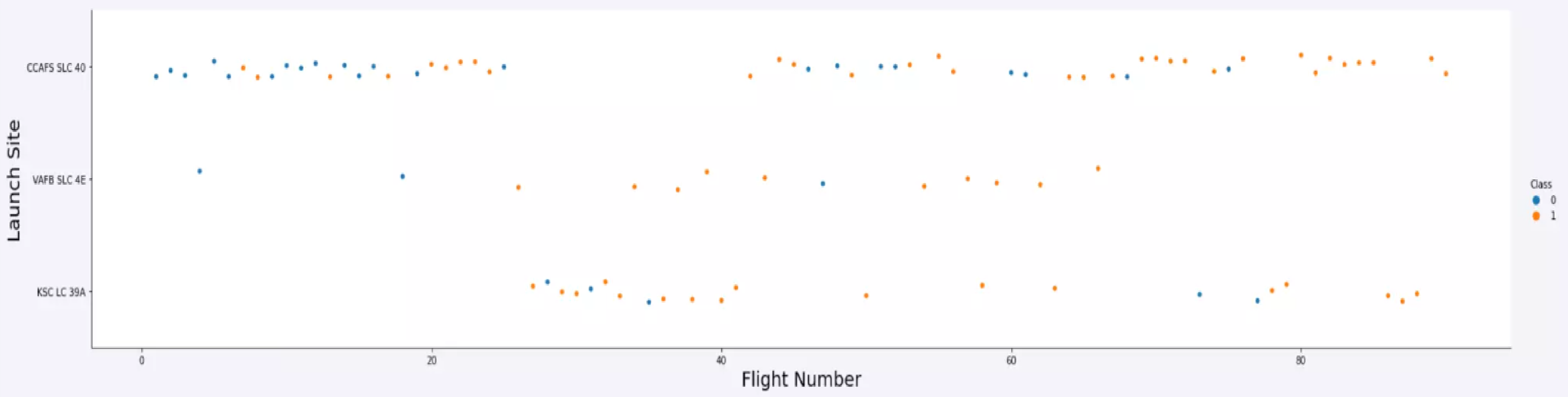
- The SVM, KNN, and Logistic Regression models are the best in terms of prediction accuracy for this dataset.
- Low weighted payloads perform better than the heavier payloads.
- The success rates for SpaceX launches is directly proportional time in years they will eventually perfect the launches.
- KSC LC 39A had the most successful launches from all the sites.
- Orbit GEO,HEO,SSO,ES L1 has the best Success Rate.

Section 2

Insights drawn from EDA

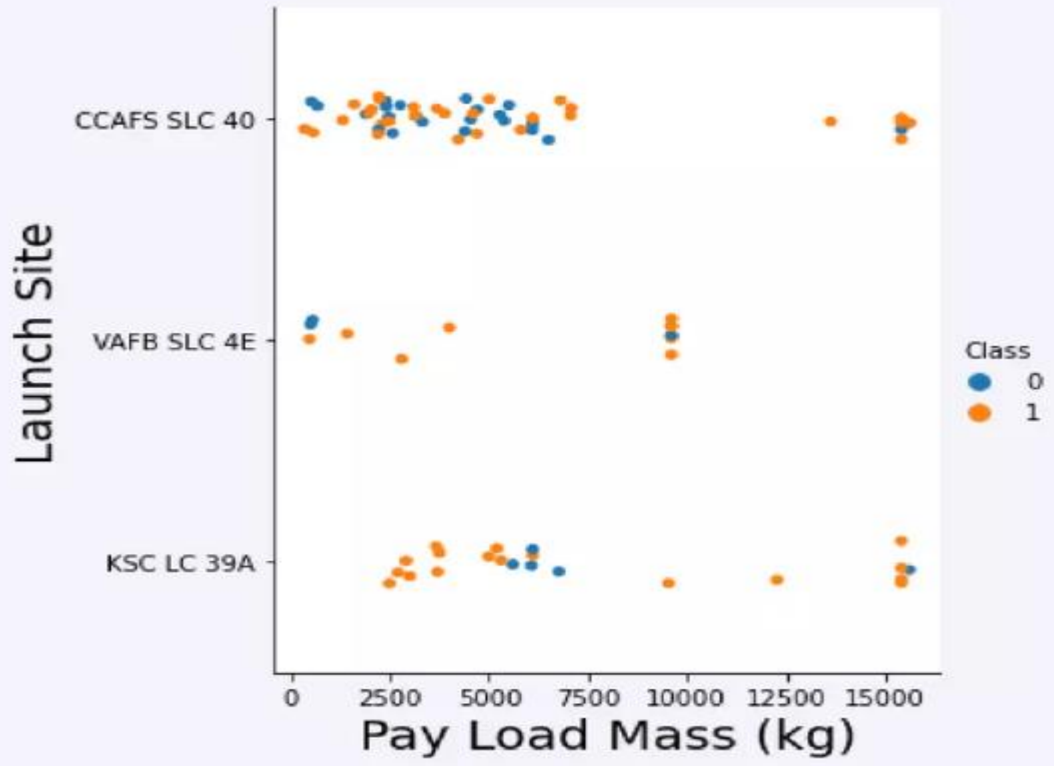
Flight Number vs. Launch Site

- Launches from the site of CCAFS SLC 40 are launches form other sites.
- significantly higher than lunches from other sites

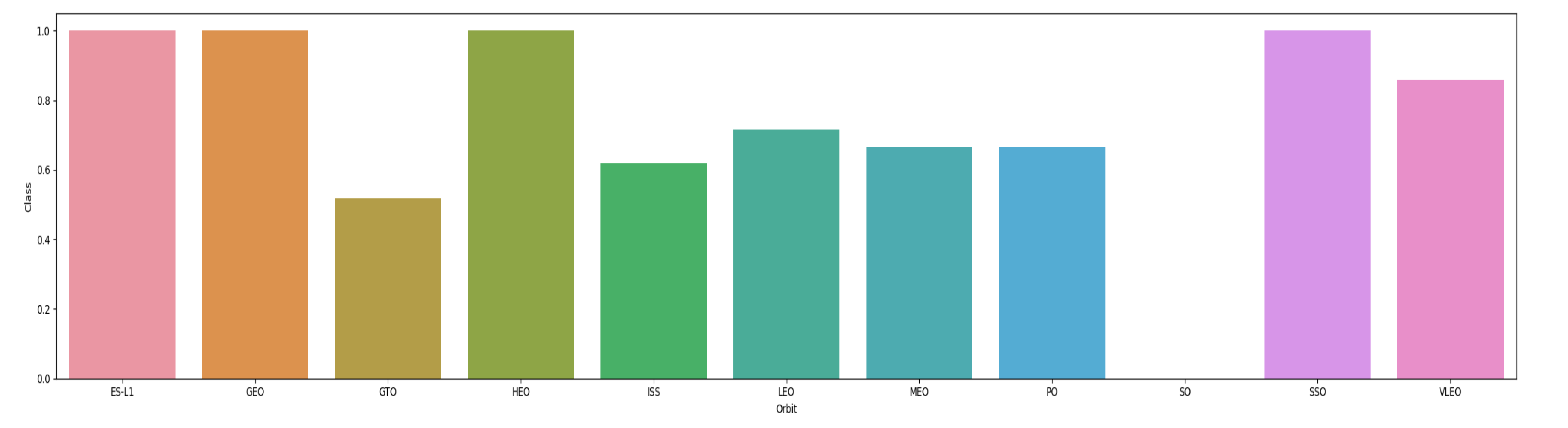


Payload vs. Launch Site

- The majority of Pay Loads with lower Mass have been launched from CCAFS SLC 40.



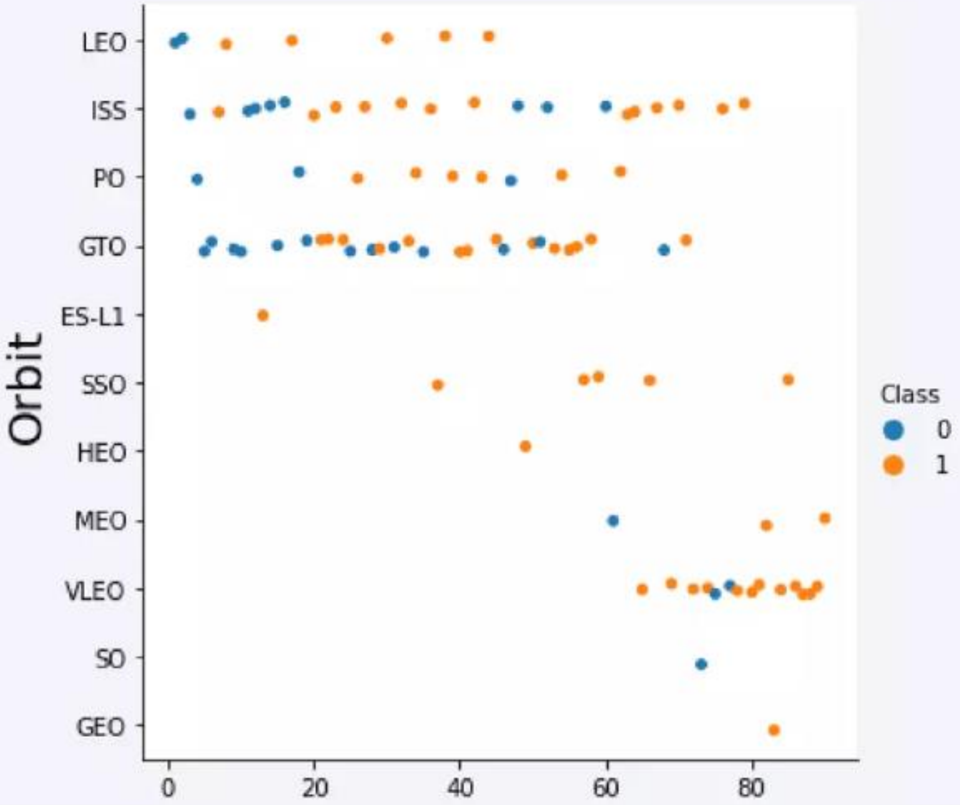
Success Rate vs. Orbit Type



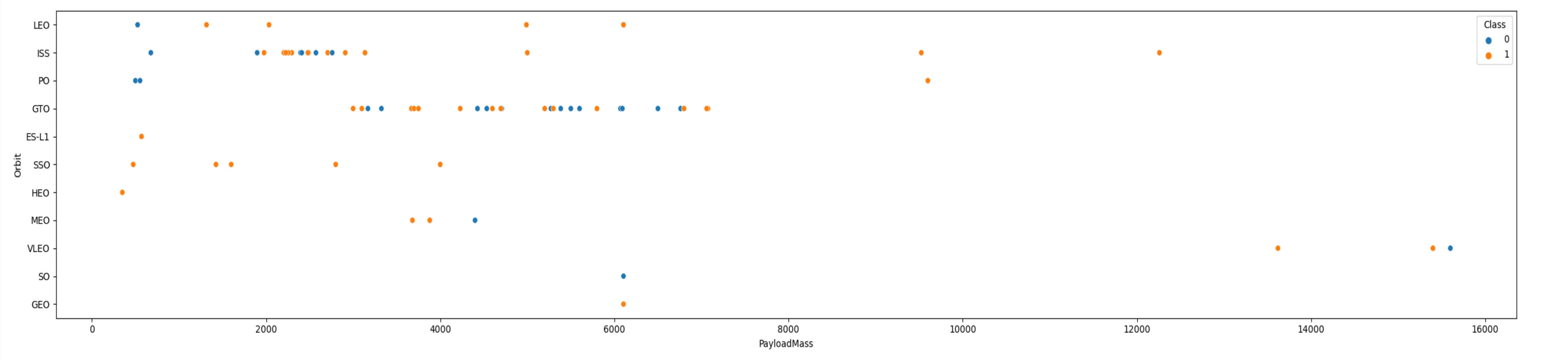
The orbit types of ES-L1, GEO, HEO, SSO are among the highest success rate.

Flight Number vs. Orbit Type

- A trend can be observed of shifting to VLEO launches in recent years.

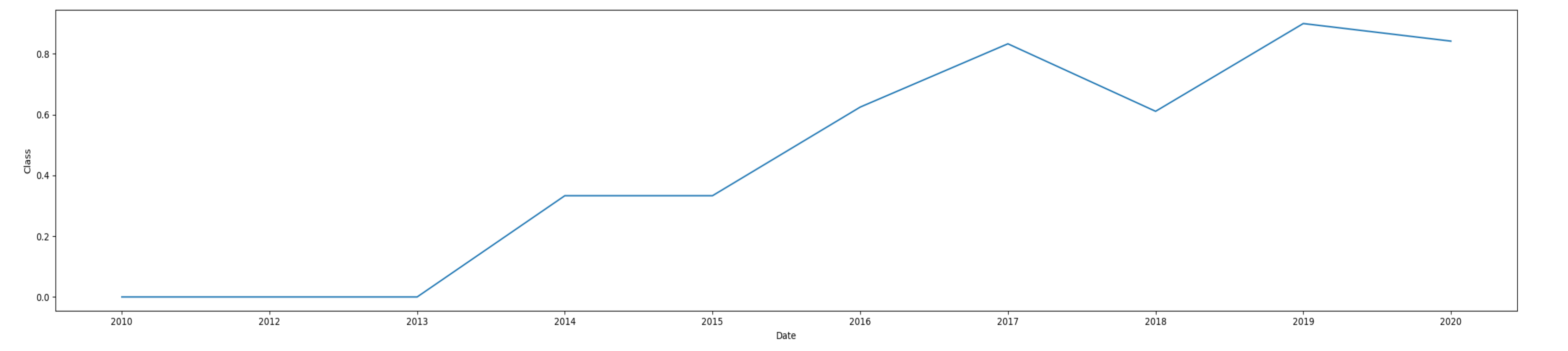


Payload vs. Orbit Type



- There are strong correlation between ISS and Payload at the range around 2000, as well as between GTO and the range of 4000-8000.

Launch Success Yearly Trend



- Launch success rate has increased significantly since 2013 and ha stablished since 2019, potentially due to advance in technology and lessons learned.

All Launch Site Names

- %sql select distinct(LAUNCH_SITE) from SPACEXTBL

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

Launch Site Names Begin with 'CCA'

- %sql select * from SPACEXTBL where LAUNCH_SITE like 'CCA%' limit 5

DATE	time_utc_	booster_version	launch_site	payload	payload_mass_kg_	orbit	customer	mission_outcome	landing_outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
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 (parachute)
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- %sql select sum(PAYLOAD_MASS_KG_) from SPACEXTBL where CUSTOMER = 'NASA (CRS)'

sum(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

- %sql select avg(PAYLOAD_MASS_KG_) from SPACEXTBL where BOOSTER VERSION = 'F9 v1.1'

avg(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

%sql select min(DATE) from SPACEXTBL where "Landing _Outcome" = 'Success (ground pad)'

min(DATE)
01-05-2017

Successful Drone Ship Landing with Payload between 4000 and 6000

%sql select BOOSTER_VERSION from SPACEXTBL where “Landing _Outcome” = 'Success (drone ship)' and PAYLOAD_MASS__KG_> 4000 and PAYLOAD_MASS_KG_< 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

%sql select Mission_Outcome, count(*) from SPACEXTBL group by Mission_Outcome

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

Boosters Carried Maximum Payload

```
%sql select distinct Booster_Version from SPACEXTBL where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_) from SPACEXTBL)
```

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

- %sql select * from SPACEXTBL where Landing Outcome like 'Success%' and (DATE between '2015-01-01' and '2015-12-31') order by date desc

time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
17:54:00	F9 FT B1029.1	VAFB SLC-4E	Iridium NEXT 1	9600	Polar LEO	Iridium Communications	Success	Success (drone ship)
05:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
04:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
21:39:00	F9 FT B1023.1	CCAFS LC-40	Thaicom 8	3100	GTO	Thaicom	Success	Success (drone ship)

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

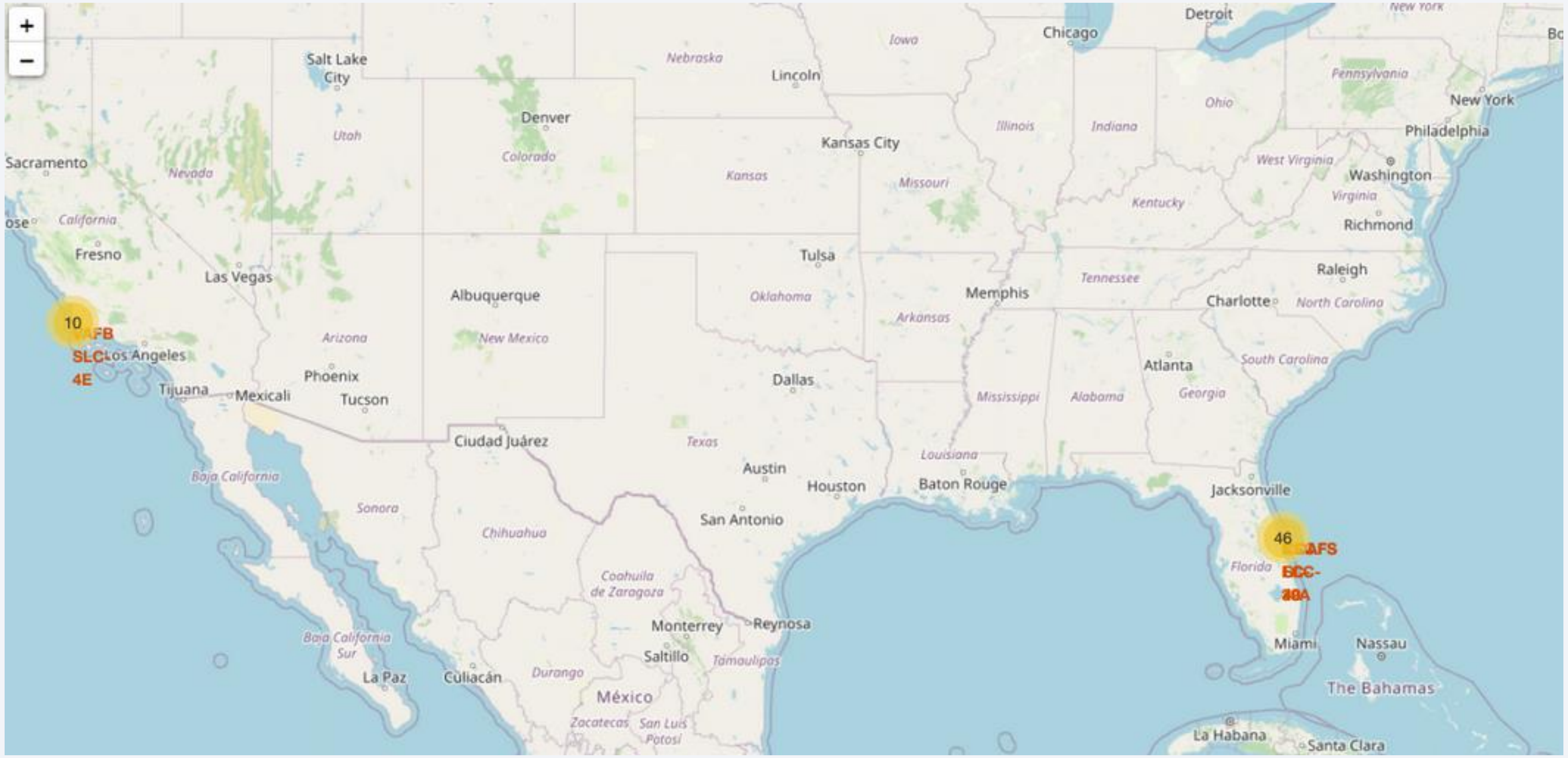
%sql SELECT "Landing _Outcome", count(*) AS count, RANK() OVER (ORDER BY count(*) DESC) AS rank FROM SPACEXTBL WHERE "Landing _Outcome" like 'Success%' and Date between '04-06-2010' and '20-03-2017' GROUP BY "Landing _Outcome" ORDER BY count DESC

Landing _Outcome	count	rank
Success	20	1
Success (drone ship)	8	2
Success (ground pad)	6	3

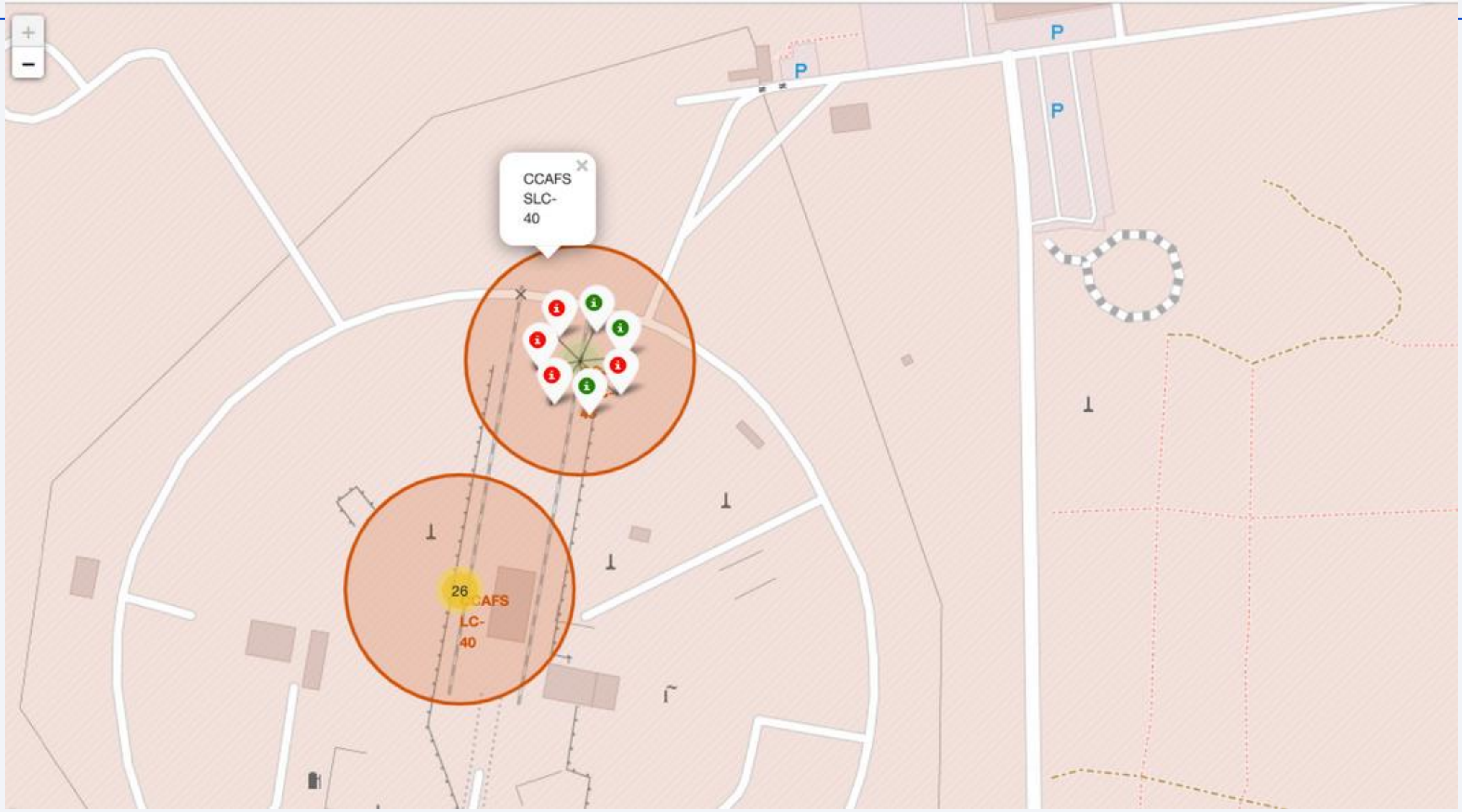
Section 3

Launch Sites Proximities Analysis

<Folium Map Screenshot 1>



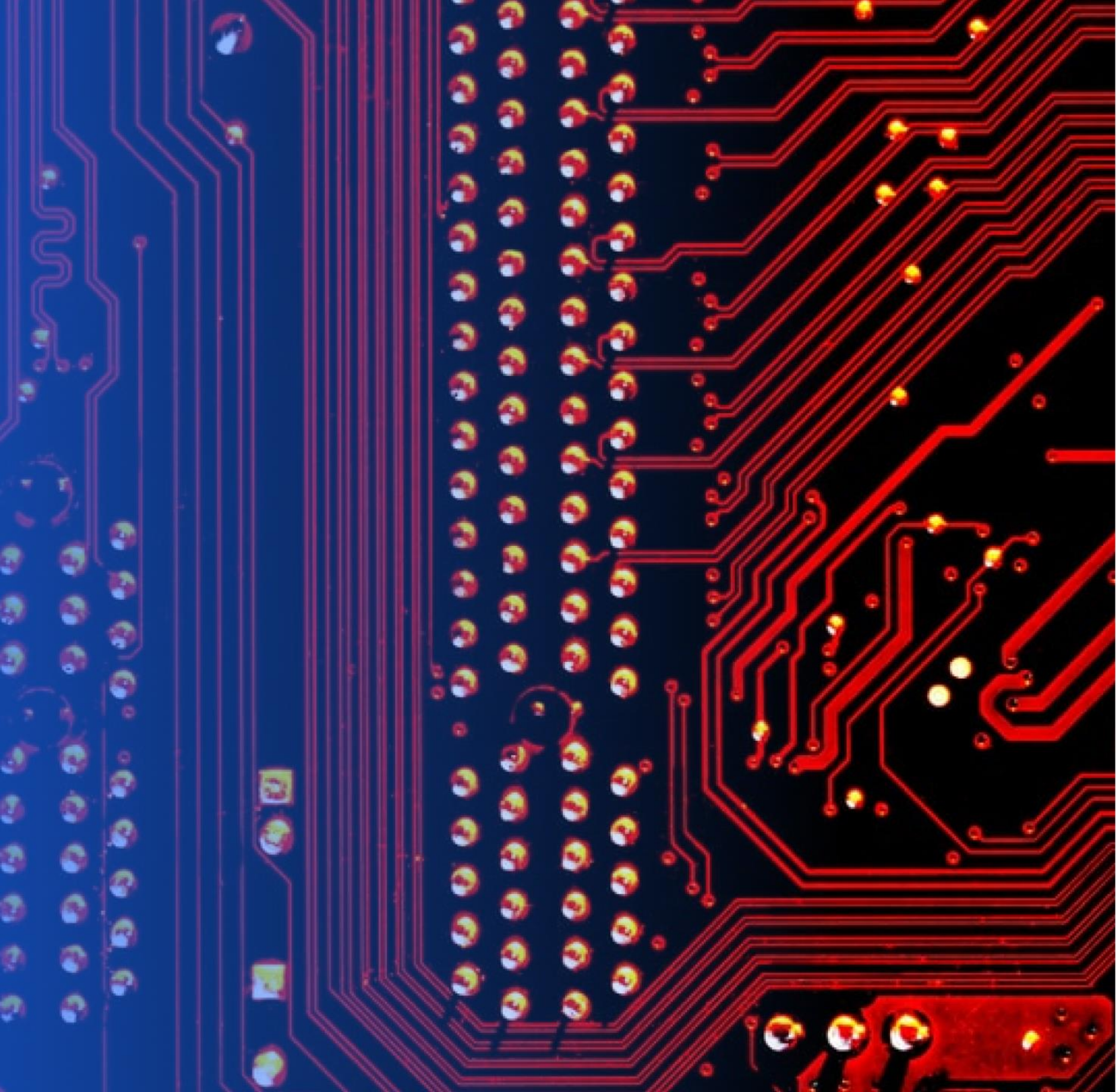
<Folium Map Screenshot 2>



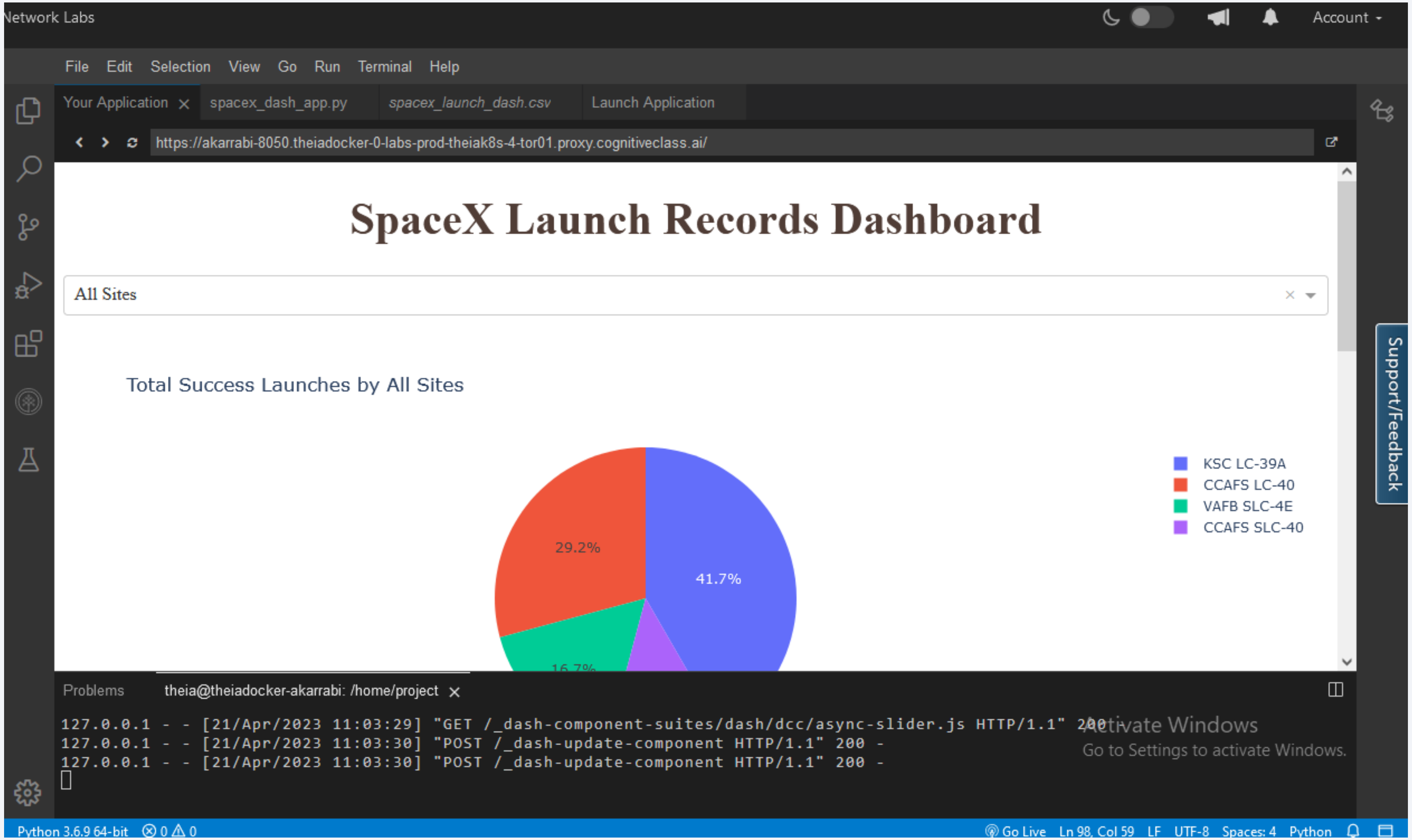
<Folium Map Screenshot 3>

Section 4

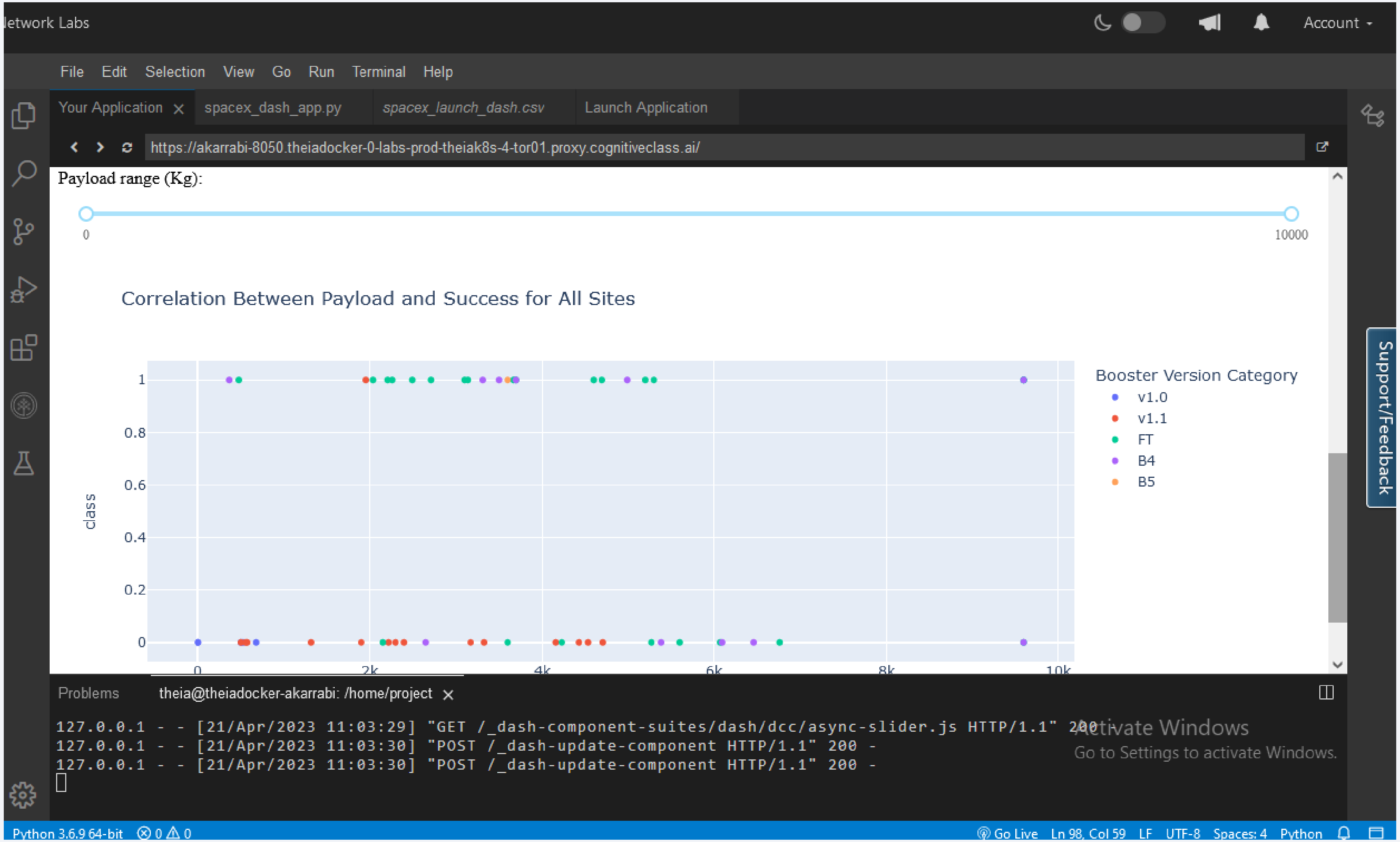
Build a Dashboard with Plotly Dash



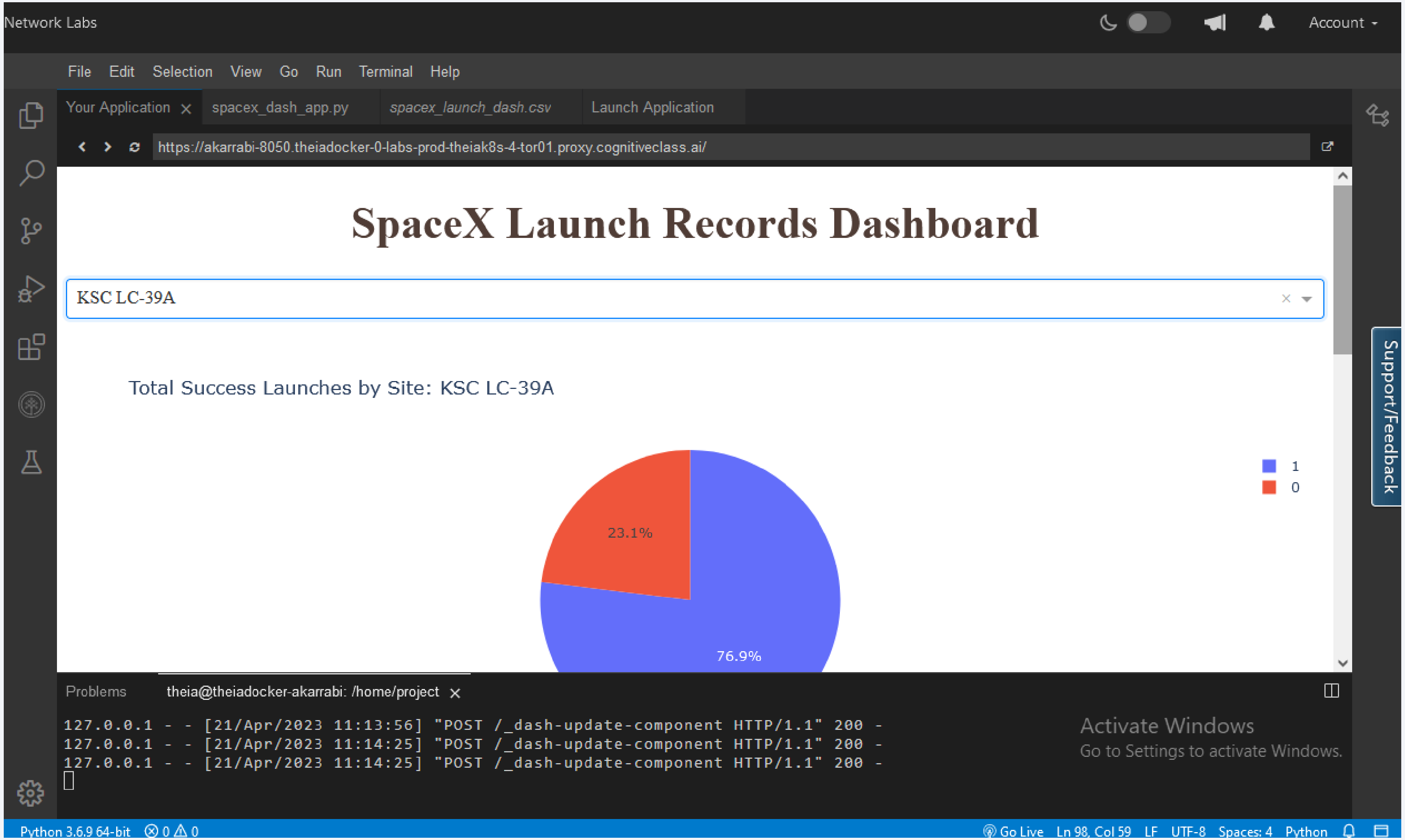
<Dashboard Screenshot 1>



<Dashboard Screenshot 2>



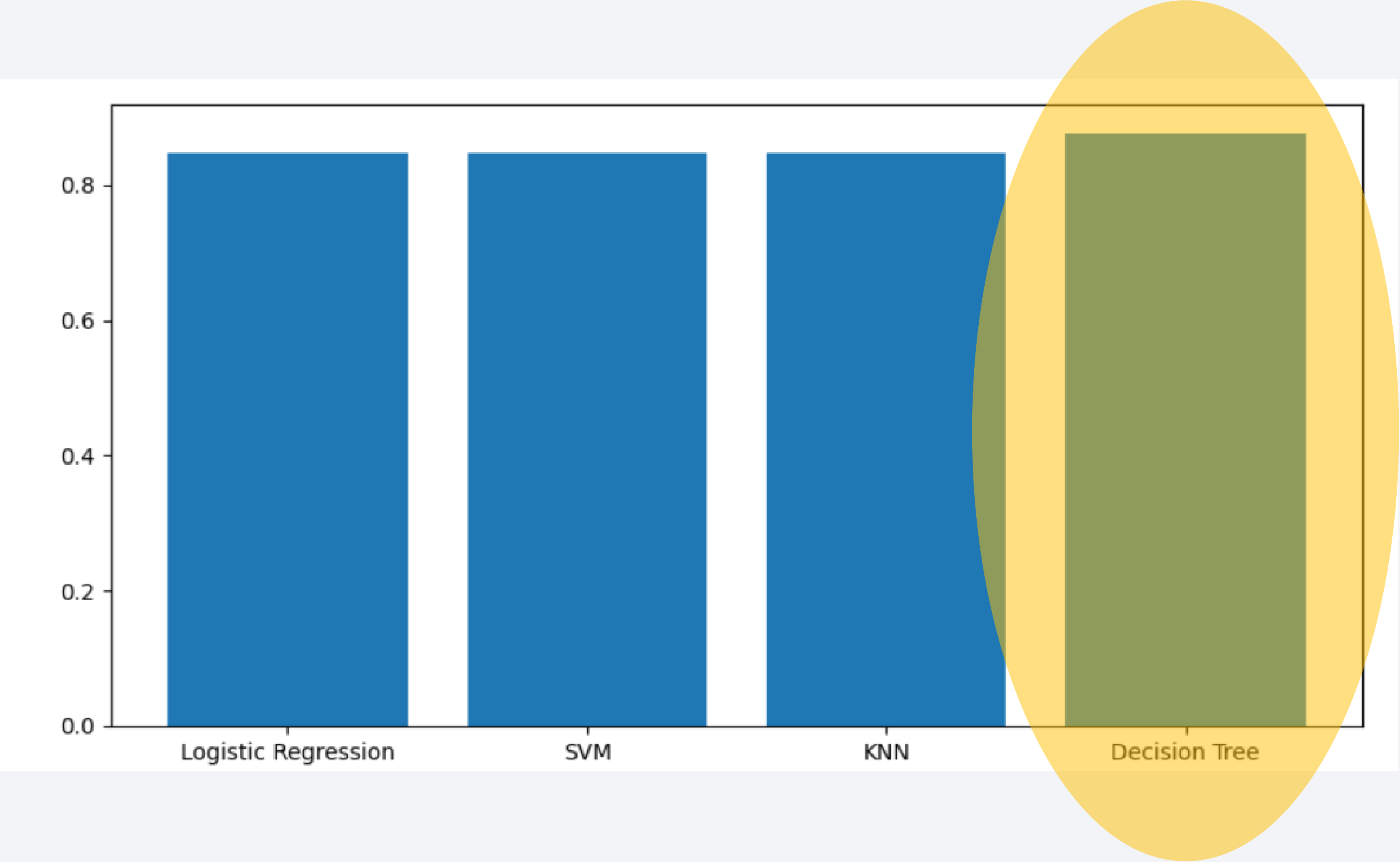
<Dashboard Screenshot 3>



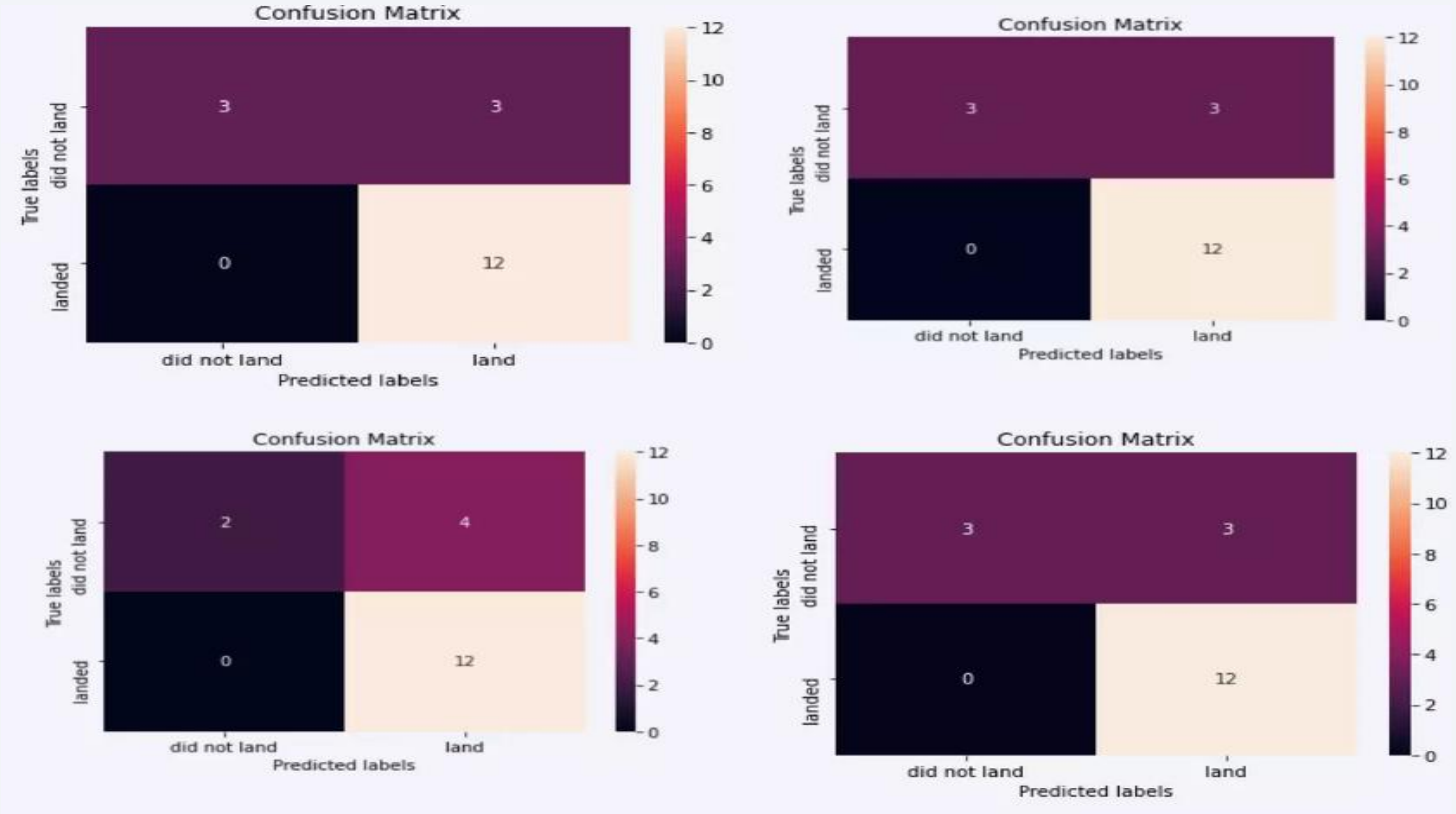
Section 5

Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix



Conclusions

- The Decision Tree model is the best in terms of prediction accuracy for this dataset.
- Low weighted payloads perform better than the heavier payloads.
- The success rates for SpaceX launches is directly proportional time in years they will eventually perfect the launches.
- KSC LC 39A had the most successful launches from all the sites.
- Orbit GEO,HEO,SSO,ES L1 has the best Success Rate.

Thank you!