SPACE X DATA ANALYSIS

Author : Neharika Joshi
22 June, 2022

Contents

- Summary
- Introduction
- Methodology used
- Results
- Conclusion

Summary

- Data Collection
- Data Wrangling
- EDA with Data Visualization
- EDA with SQL
- Building an interactive map with Folium
- Building a Dashboard with Plotly Dash
- Machine Learning (Finding the best Predictor

Introduction

- Background
  - SpaceX released Falcon rocket launches information on it's website which had a budget of 62 million dollars on contrast to the other companies that has average 165 million dollars cost. SpaceX saved more that half of the money because they reuse the first stage
- Problem
  - Predicting if the first stage of the SpaceX Falcon 9 rocket will land successfully or not

Methodology

- Data Collection

Data Collection

- SpaceX launch data which is gathered from SpaceX REST API.
- Gave us data regarding launches, rockets used, payload delivered, specifications of launches and landing, outcome of landing
- The SpaceX REST API endpoints, or URL, starts with api.spacexdata.com/v4/.
- Also used Wikipedia for Web Scrapping using BeautifulSoup.

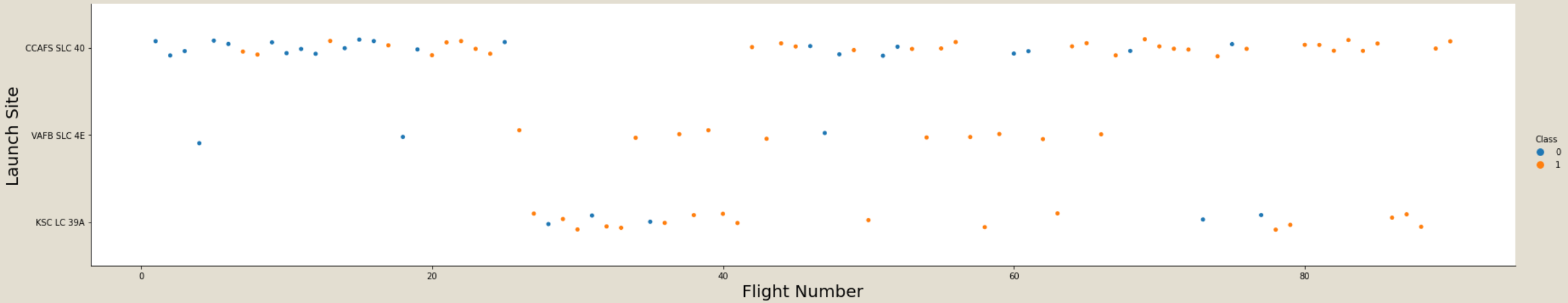
- Notebook: https://github.com/niyaryca/Applied-Data-Science-Capstone-Project/blob/main/0.%20Web%20scraping%20Falcon%209%20and%20Falcon%20Heavy%20Launches%20Records%20from%20Wikipedia.ipynb

Data Collection - Scraping

- Notebook : https://github.com/niyaryca/Applied-Data-Science-Capstone-Project/blob/main/1.%20Spacex-Data%20wrangling.ipynb

EDA WITH DATA VISUALIZATION

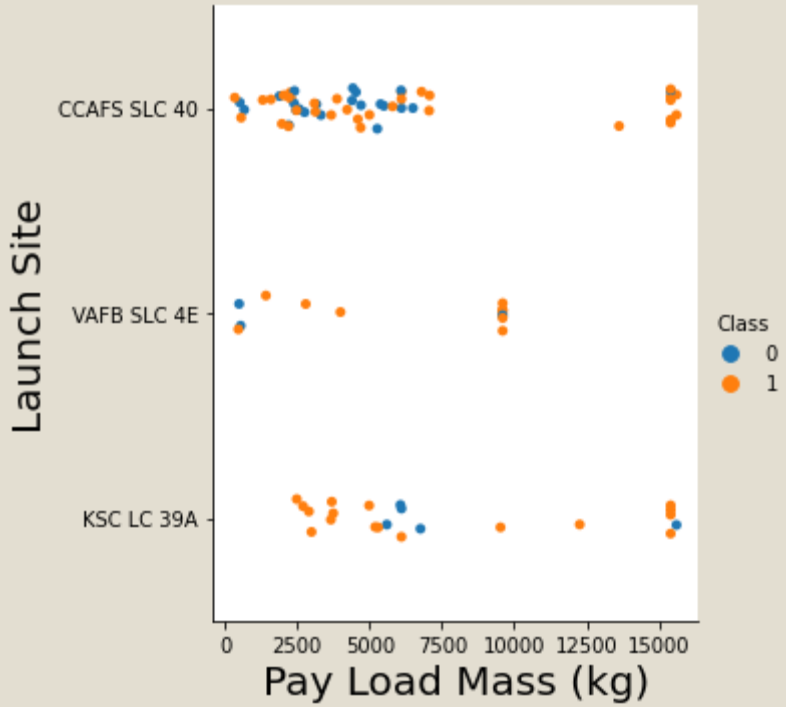
Flight Number vs. Launch Site



- The rockets launched from the site of CCAFS SLC 40 are comparatively higher than from any other sites

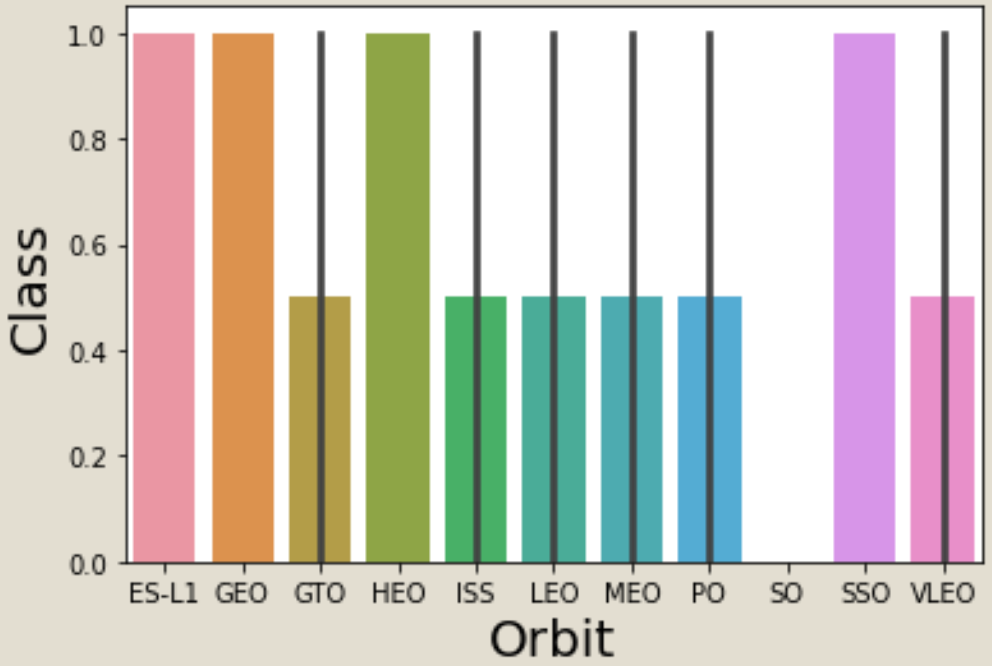
Payload vs. Launch Site

- Pay loads with lower mass have been launched from CCAFS SLC 40



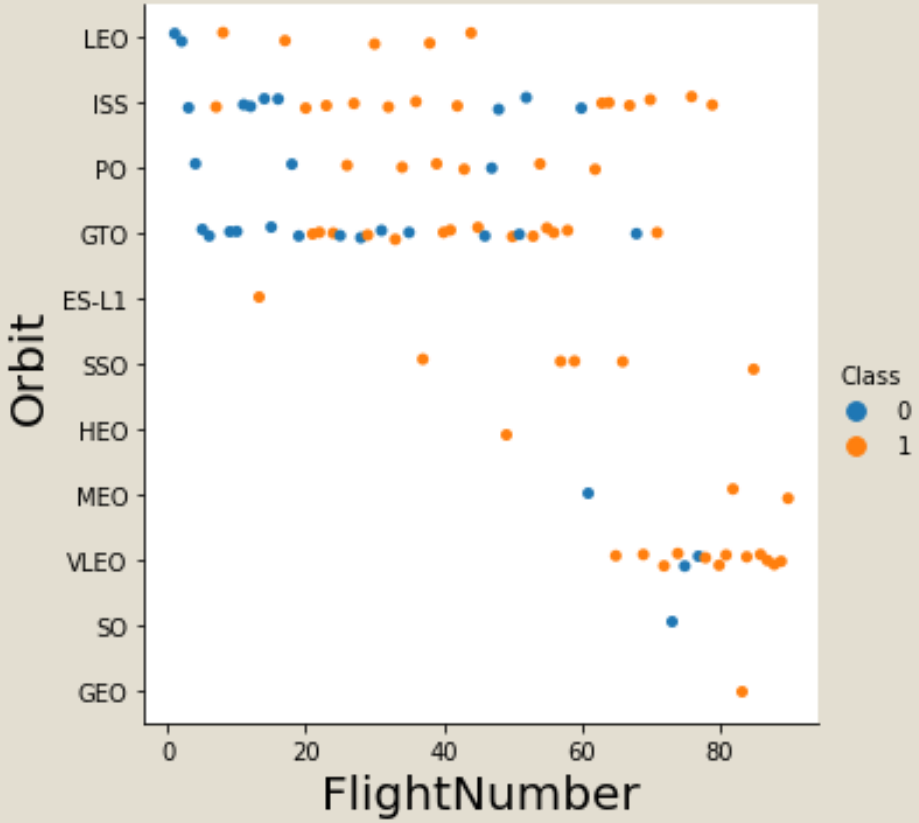
Orbit Type vs. Success Rate

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



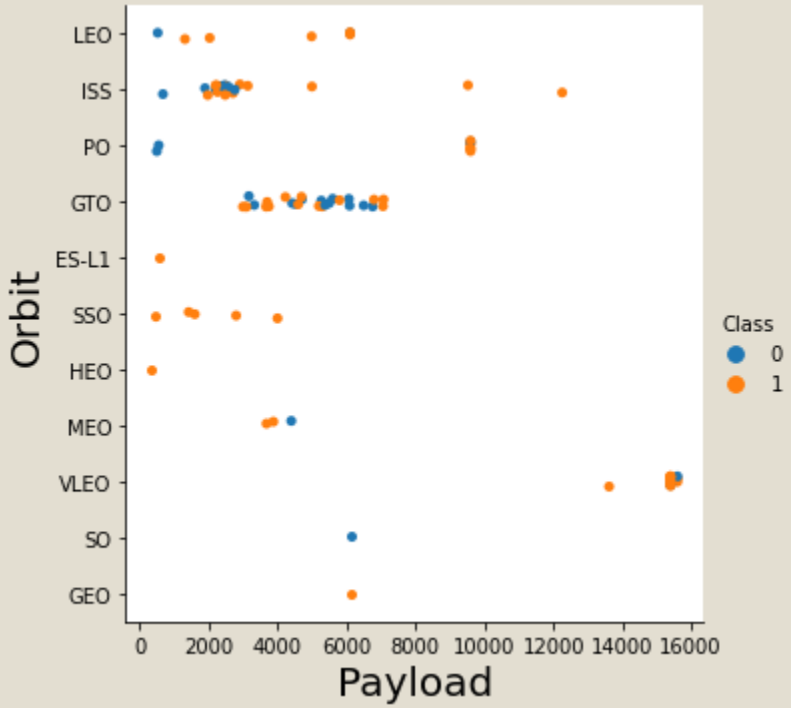
Flight Number vs. Orbit Type

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



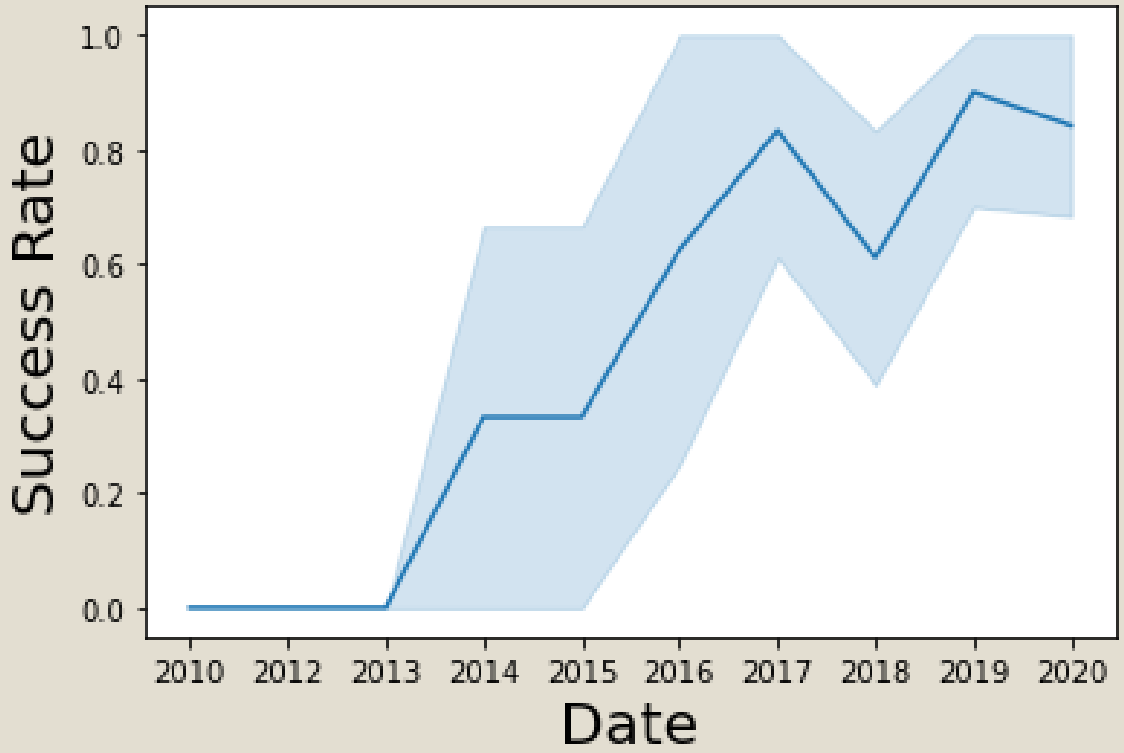
Payload vs. Orbit Type

- 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 2019 and has stabilized since 2019, potentially due to advance in technology and lessons learned.



EDA WITH SQL

Site Names

- %sql select distinct(LAUNCH_SITE) from SPACEXTBL

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

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)

- 45596

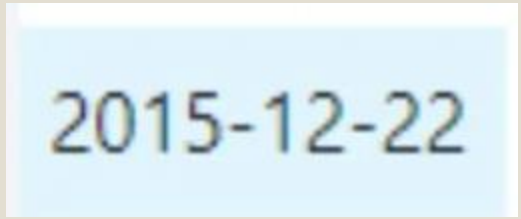
Average Payload Mass by F9 v1.1

- %sql select avg(PAYLOAD_MASS__KG_) from SPACEXTBL where BOOSTER_VERSION = 'F9 v1.1'

2928.400000

First Successful Ground Landing Date

- %sql select min(DATE) from SPACEXTBL where Landing__Outcome = 'Success (ground pad)'



Successful drone 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 count(MISSION_OUTCOME) from SPACEXTBL where MISSION_OUTCOME = 'Success' or MISSION_OUTCOME = 'Failure (in flight)'

Boosters Carried Maximum Payload

- %sql select BOOSTER_VERSION from SPACEXTBL where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_) from SPACEXTBL)

2015 Launch Records

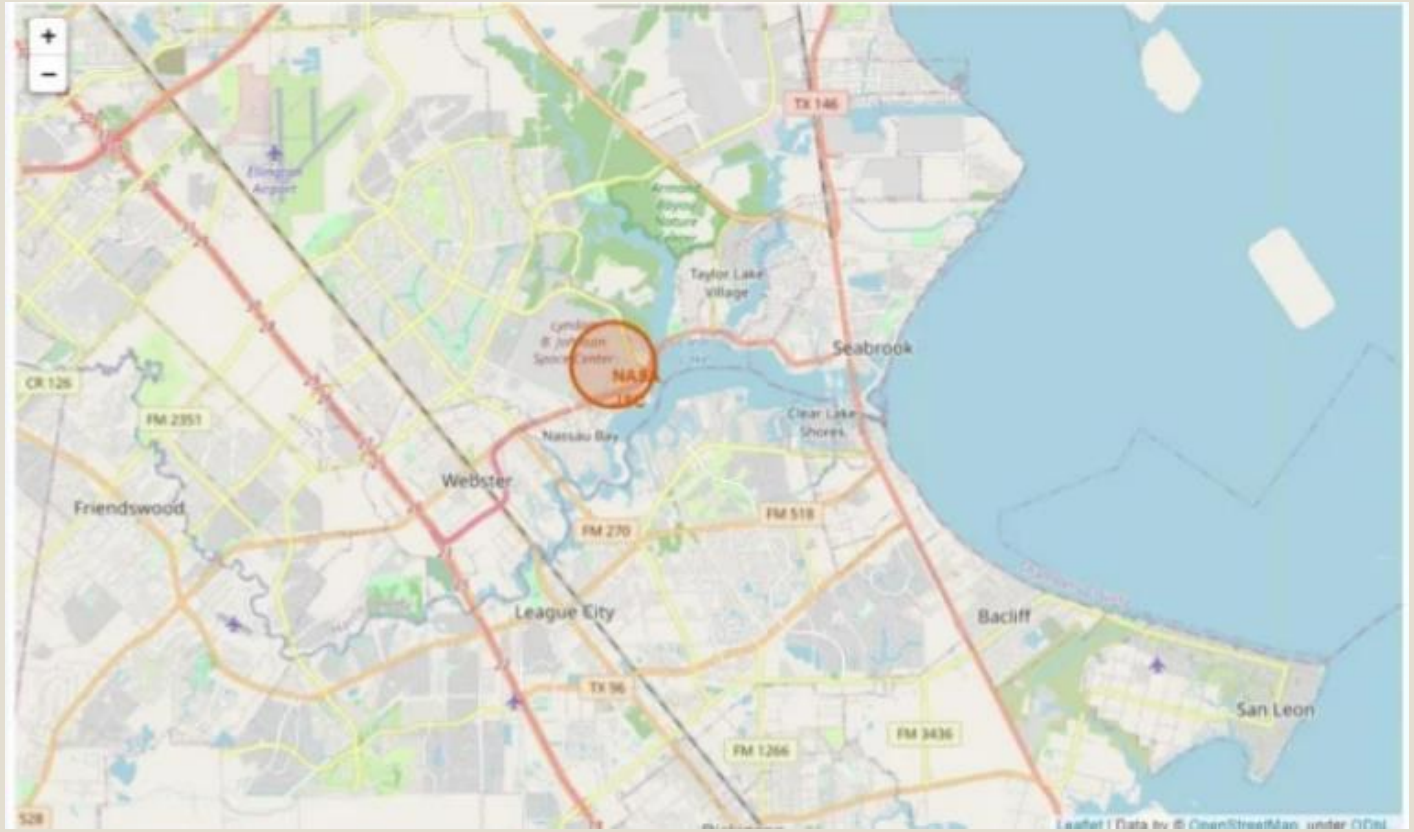
- %sql SELECT EXTRACT(MONTH, select min(DATE) from SPACEXTBL where Landing__Outcome = 'Success (ground pad)')

Rank Landing Outcomes Between 2010/06/04 & 2017-03-20

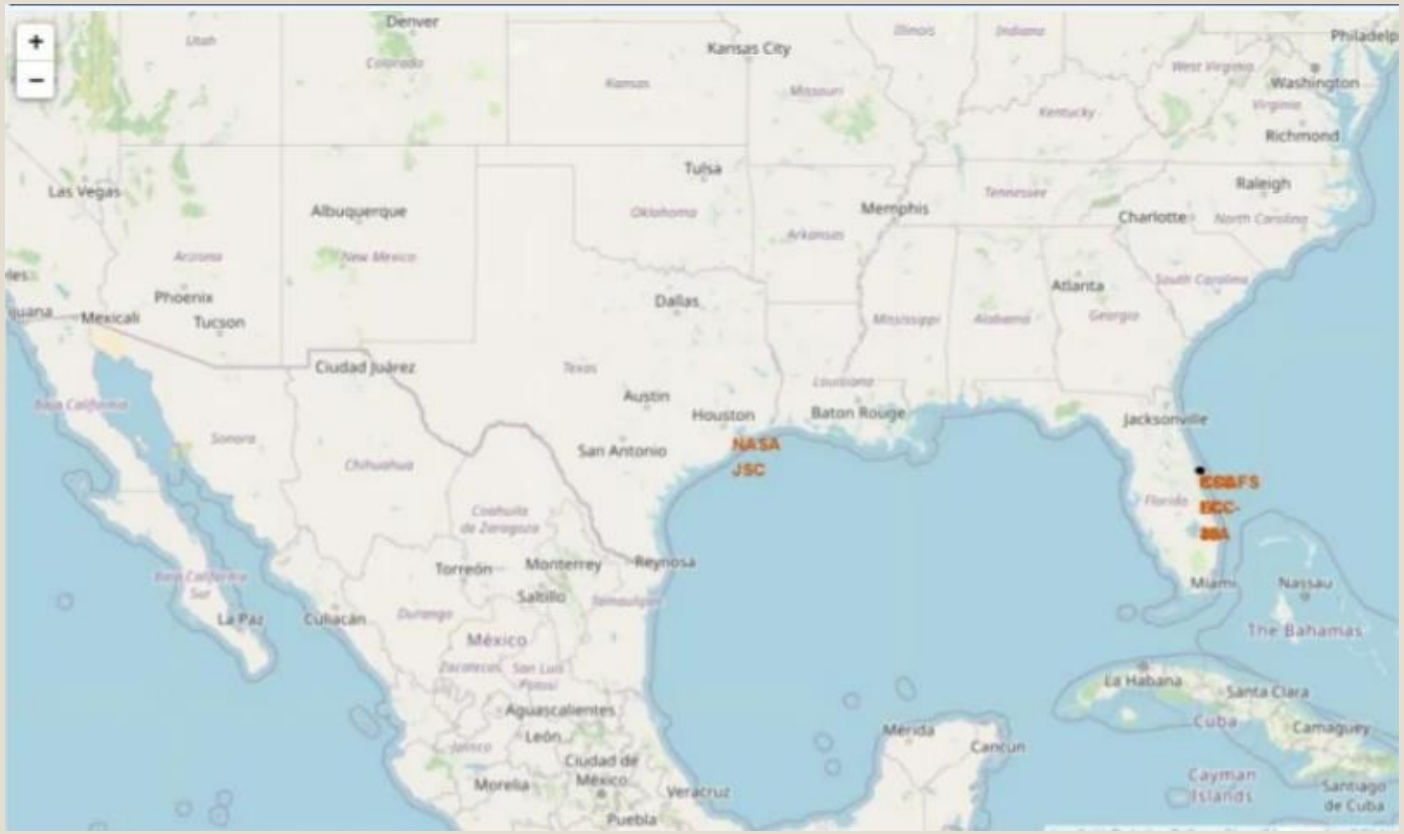
- %sql select * from SPACEXTBL where Landing__Outcome like 'Success%' and (DATE between '2010-06-04' and '2017-03-20') order by date desc

LAUNCH SITES PROXIMITIES ANALYSIS

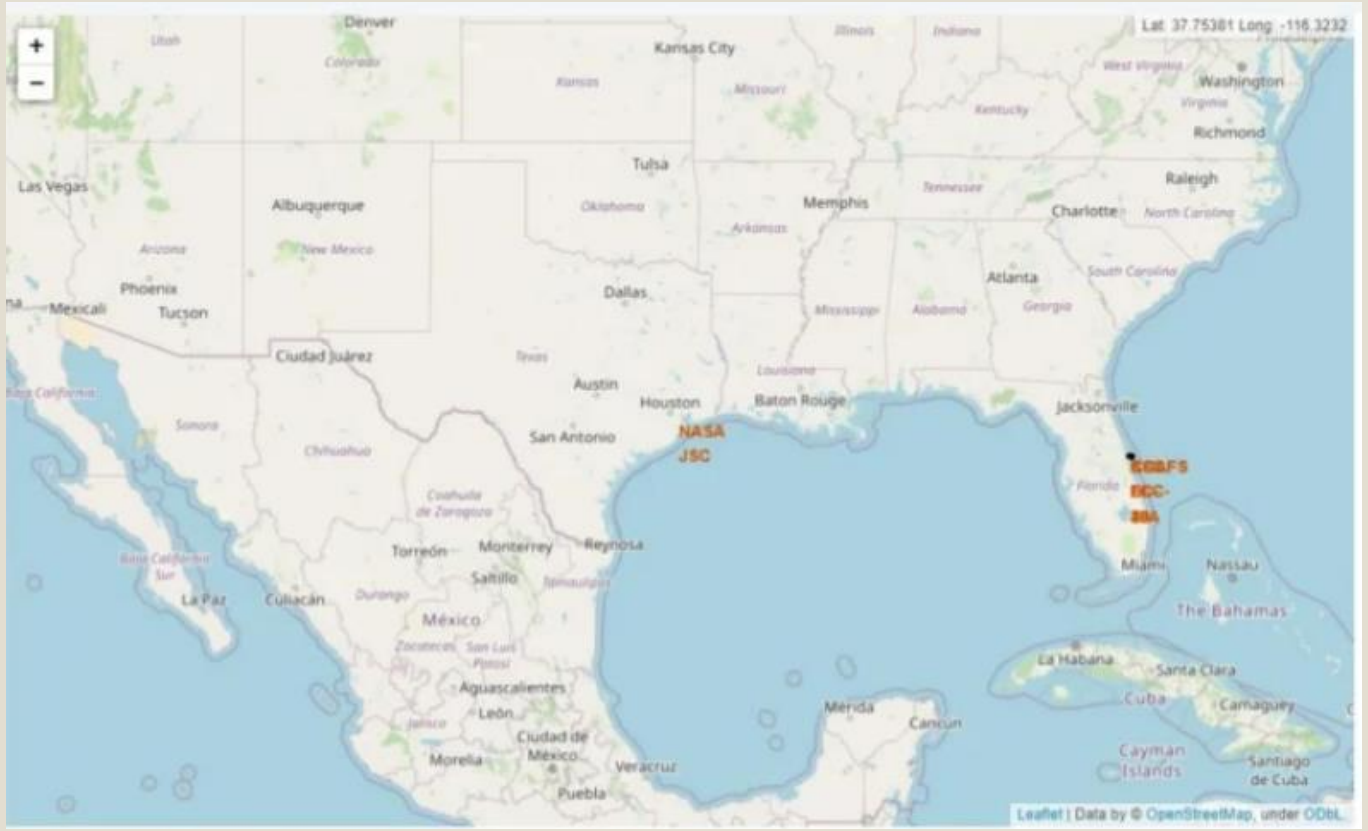
All Launch sites marked on a map



Success/failed launches marked on the map

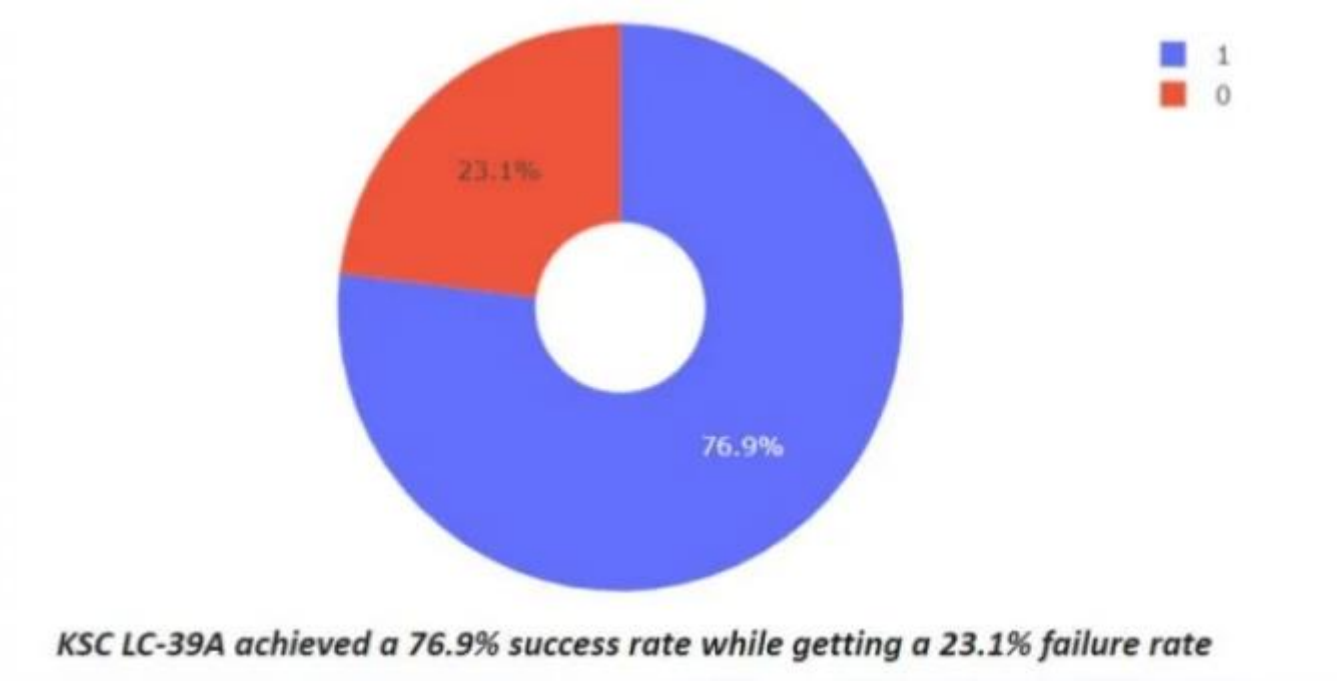


Distances between a launch site to its proximities



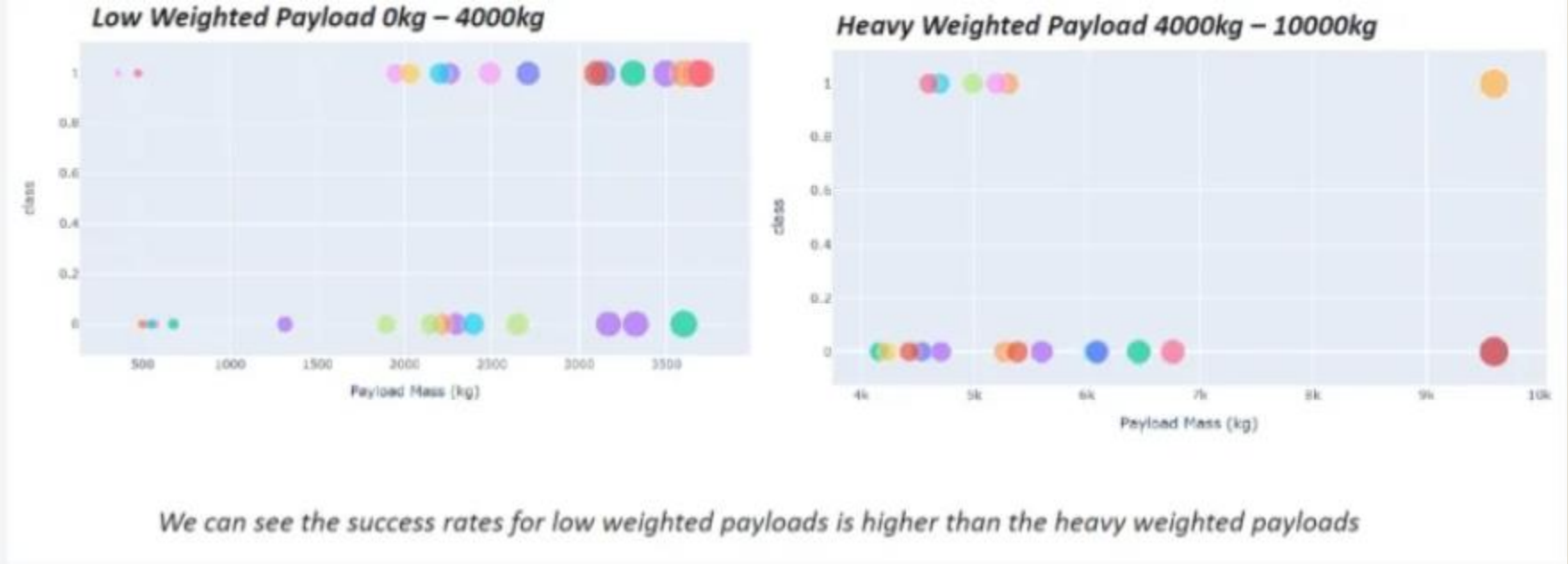
Total success launches by all sites

Success rate by site

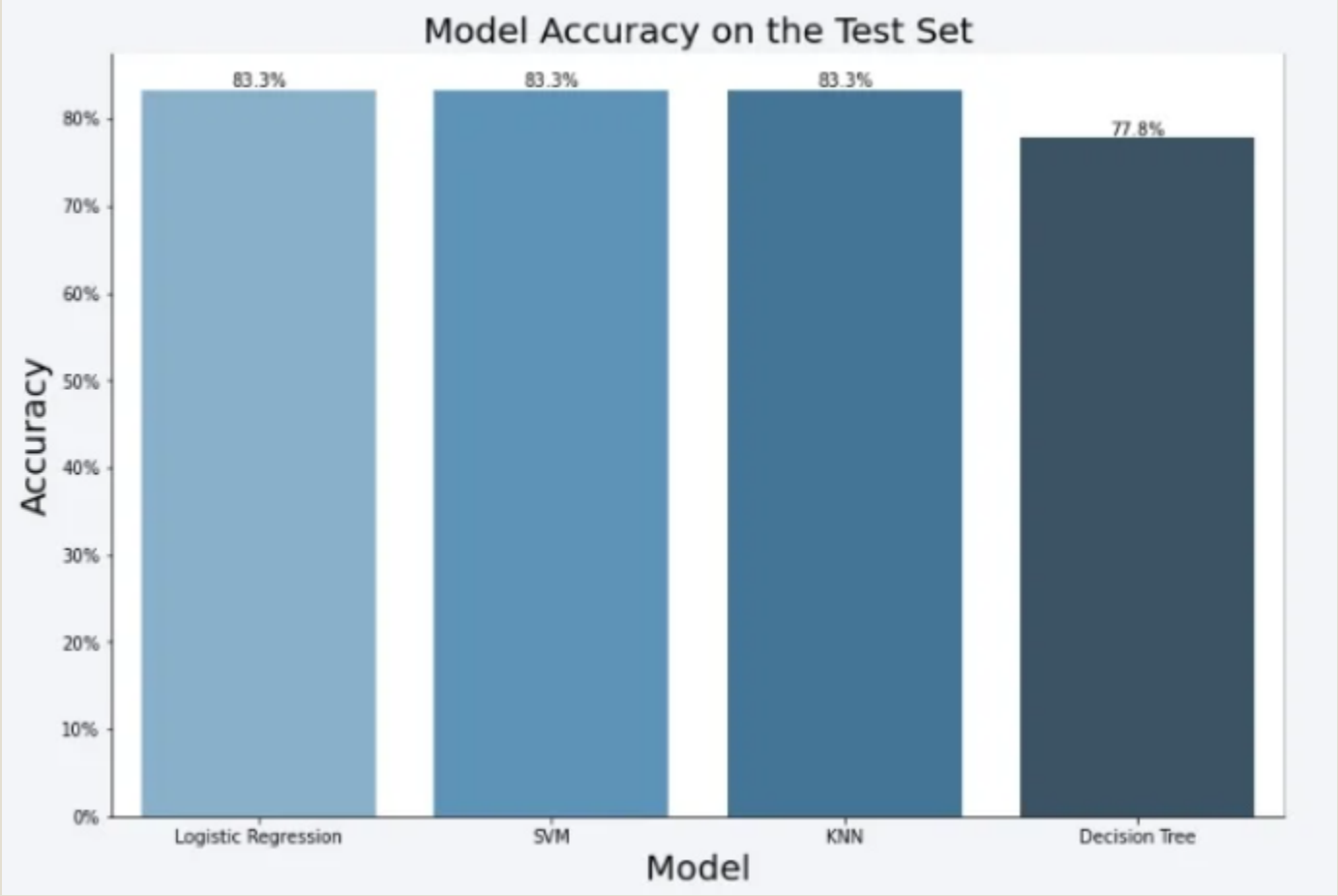


KSC LC-39A achieved a 76.9% success rate while getting a 23.1% failure rate

Payload vs Launch outcome



Accuracy of Classification algorithms



Conclusions

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