IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Francisco Stigliano
26-April-2023

Outline

Executive Summary

- Summary of methodologies
  - Data Collection
  - Data Wrangling
  - EDA with data visualization
  - Building Folium powered interactive maps
  - Building a Plotly Dash powered Dashboard
  - Classification predictive analysis
- Summary of all results
  - EDA
  - Interactive Analysis
  - Predictive Analysis

Introduction

- Project background and context:
  - Space tourism will be a big thing. Companies must find the way to create the most sustainable way to take people to space. Leading this race is SpaceX with the Falcon 9 Rockets. Each launch costs SpaceX about 62 million dollars a much smaller value when compared with the 165 million dollars that other providers must spend. The reason for the difference? Reusability! The first stage of the rocket can be reused as long as it lands.
- Problems you want to find answers
  - The goal of the project is to predict if the first stage of the SpaceX Falcon9 rocket will land successfully.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - SpaceX REST API
  - Web Scrapping from Wikipedia
- Perform data wrangling
  - OHE for Machine Learning and data cleaning of null, nan values and irrelevant collumns
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - LR, KNN, SVM and DT models were built, trained and evaluated

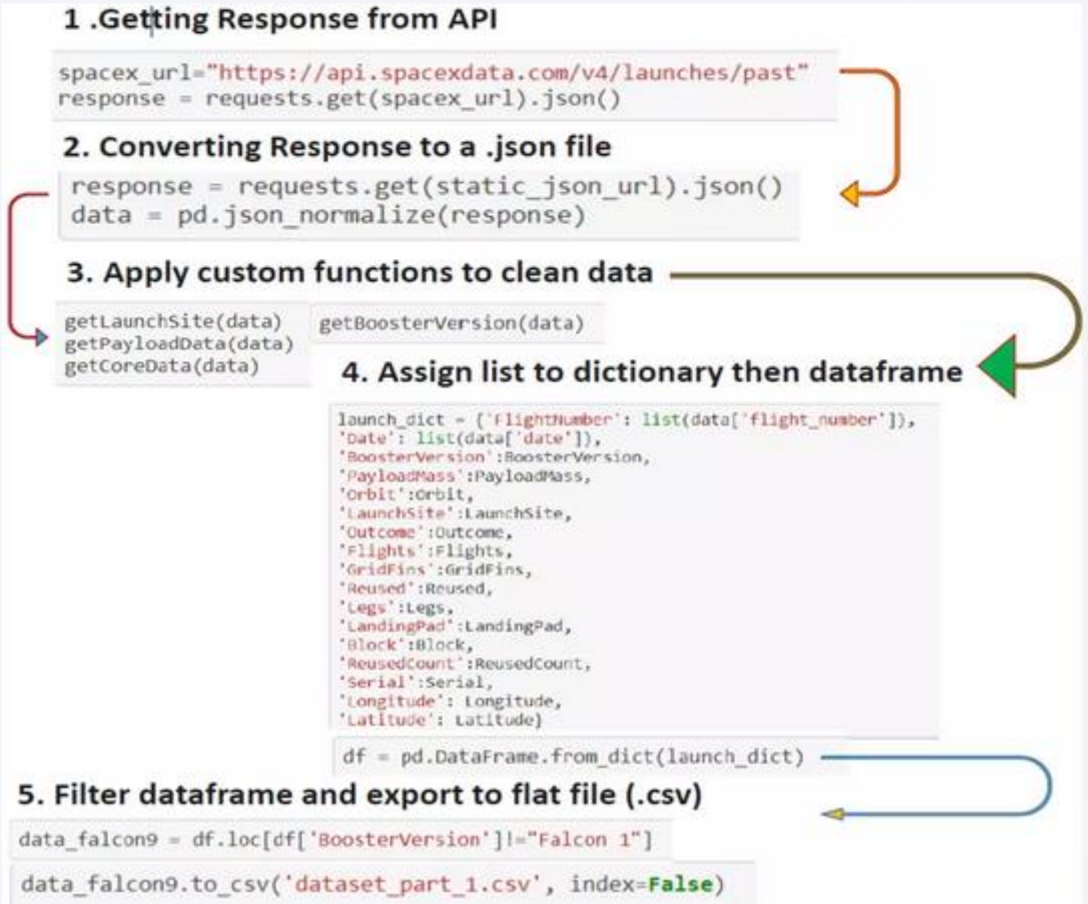
Data Collection

- Sources
  - Wikipedia Web Scraping Using BeautifulSoup
  - Information provided by the SpaceX REST API
    - Launches
    - Rocket Used
    - Payload Delivered
    - Launch and Landing info
    - Landing Outcome (if it landed or if an unscheduled rapid disassembly happened)

Data Collection – SpaceX API

- SpaceX REST API for Data Collection

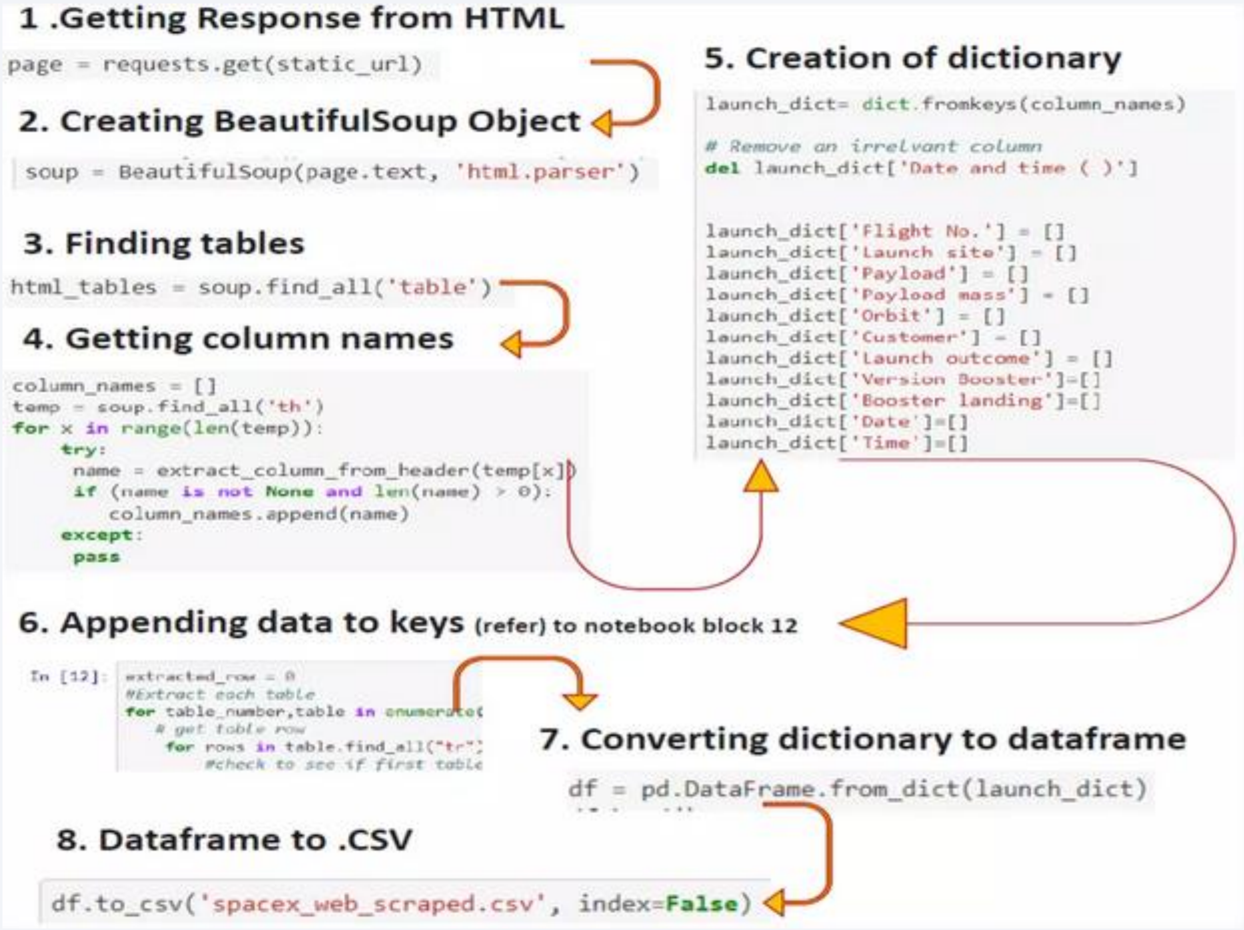
- [GitHub - SpaceX API requests]()



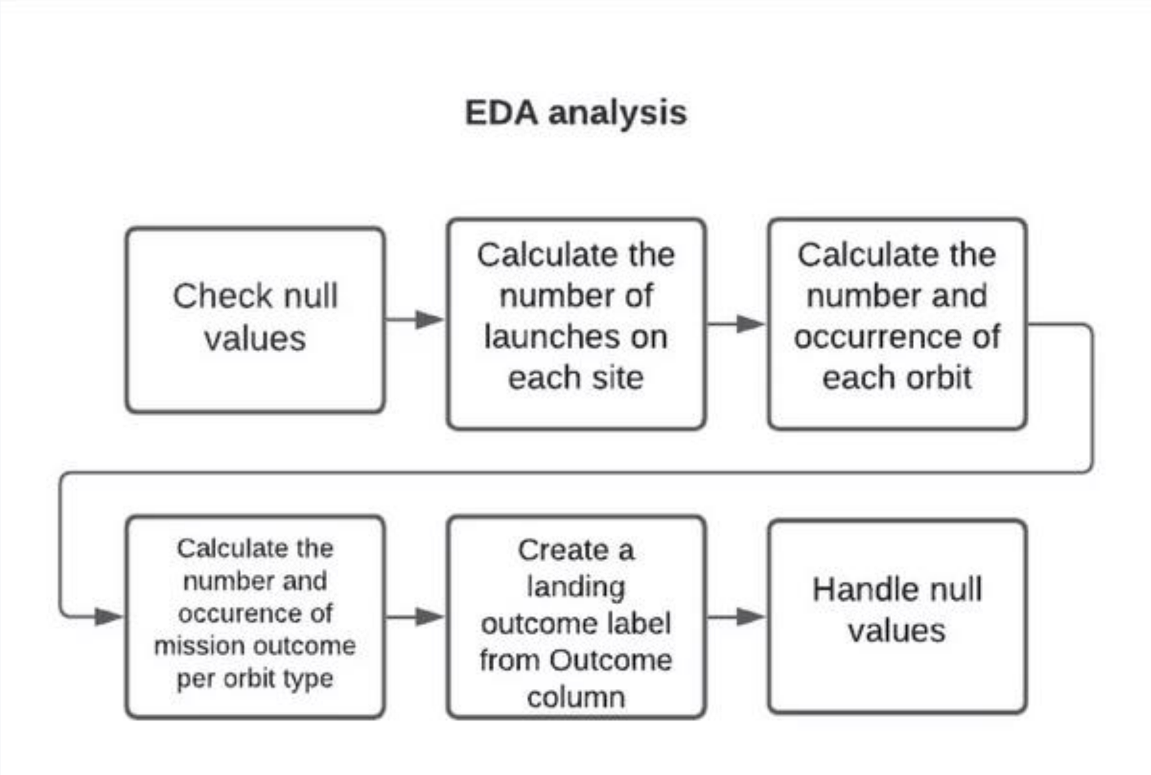
Data Collection - Scraping

- Web Scraping Wikipedia pages

- [Github - WebScraping]()

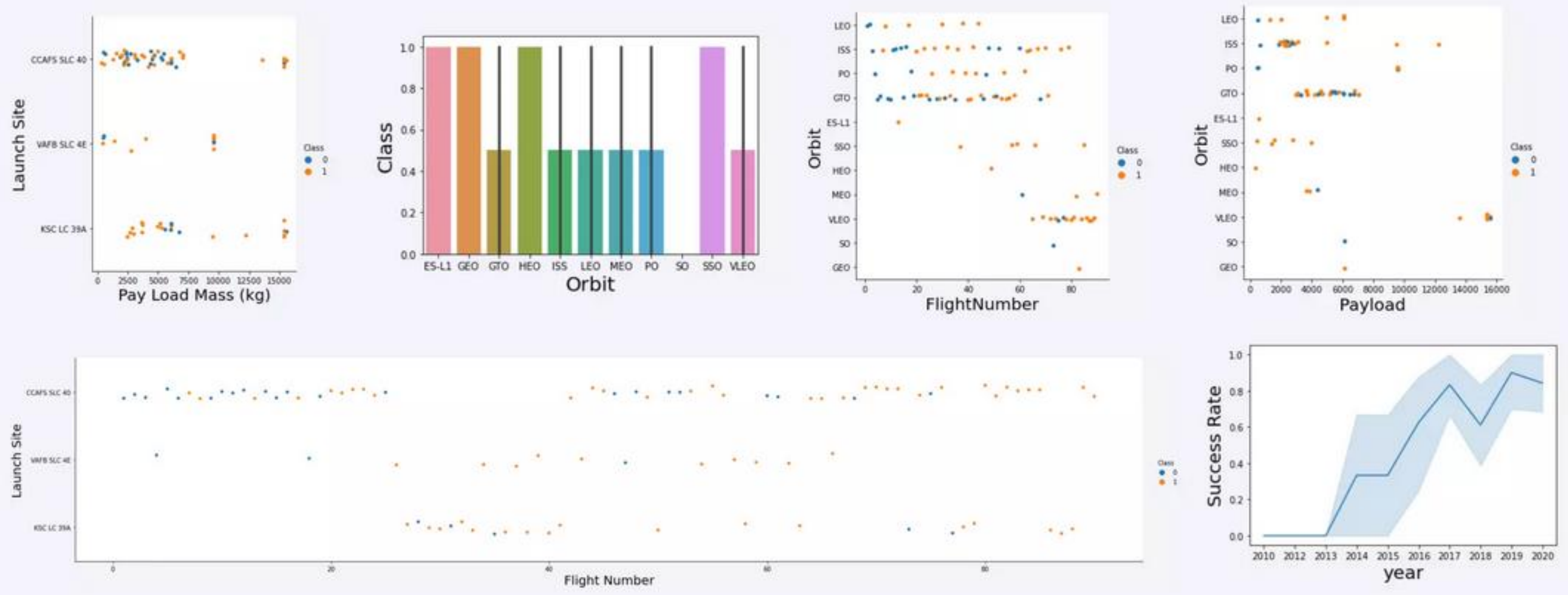


Data Wrangling



- [Github - Data Wrangling]()

EDA with Data Visualization



- [Github- Viz]

EDA with SQL

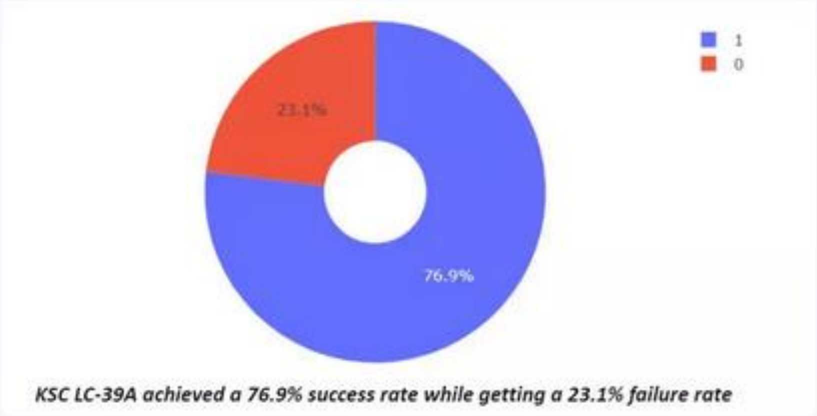
- Some queries performed:
  - Displaying singular and combined categories
  - Listing the successful events
  - Ranking the outcomes on a given interval

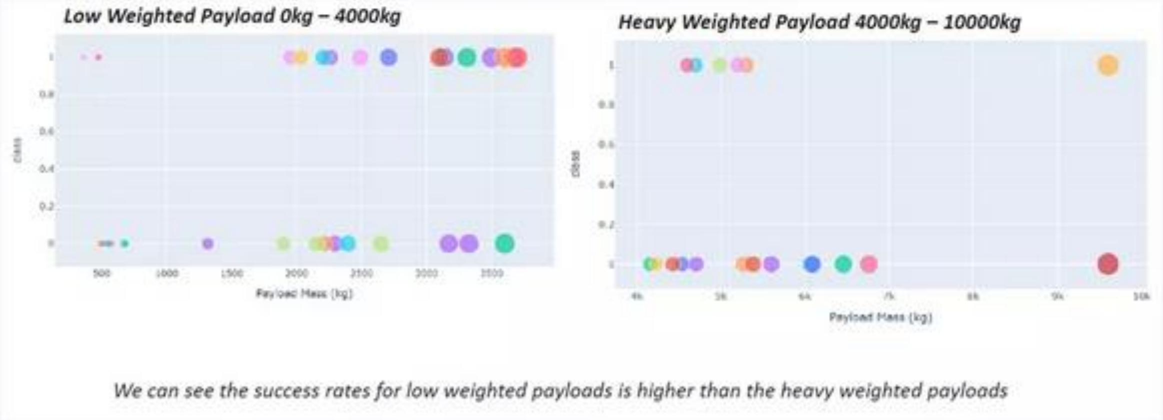
[Github - SQL]

Build an Interactive Map with Folium

- [Github - Folium]()

Build a Dashboard with Plotly Dash





- [Github - Plotly Dash]

Predictive Analysis (Classification)

- Both the SVM and the KNN got an accuracy score of 83.3%, the highest amongst the 4 models trained.
- The SVM also has the best under the curve performance at 0.958

Github - Machine Learning

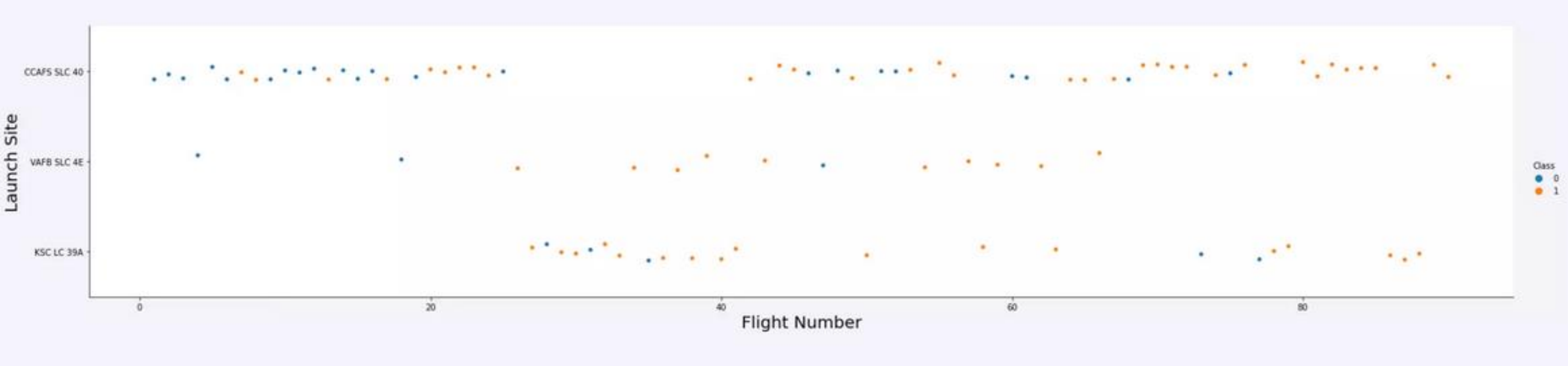
Results

- SVM, KNN and LR work well in terms of accuracy
- Heavy payloads generally landing to fail
- The success rate of launches and landings as risen throughout the years
- Orbit GEO,HEO,SSO,ES L1 has provides the best rate for sucess

Section 2

Insights drawn from EDA

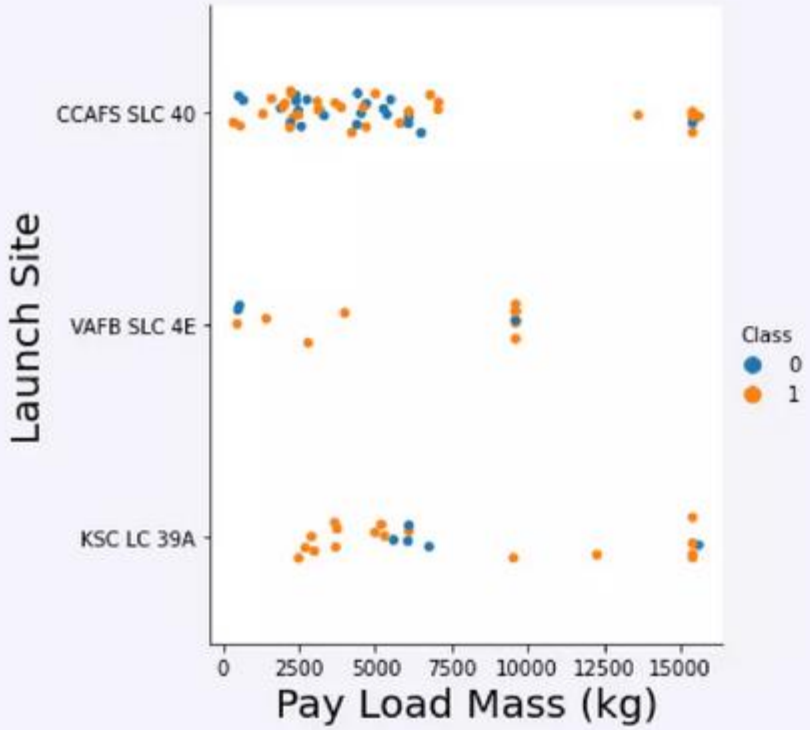
Flight Number vs. Launch Site



- CCAFS SLC 40 Launches have the best performance

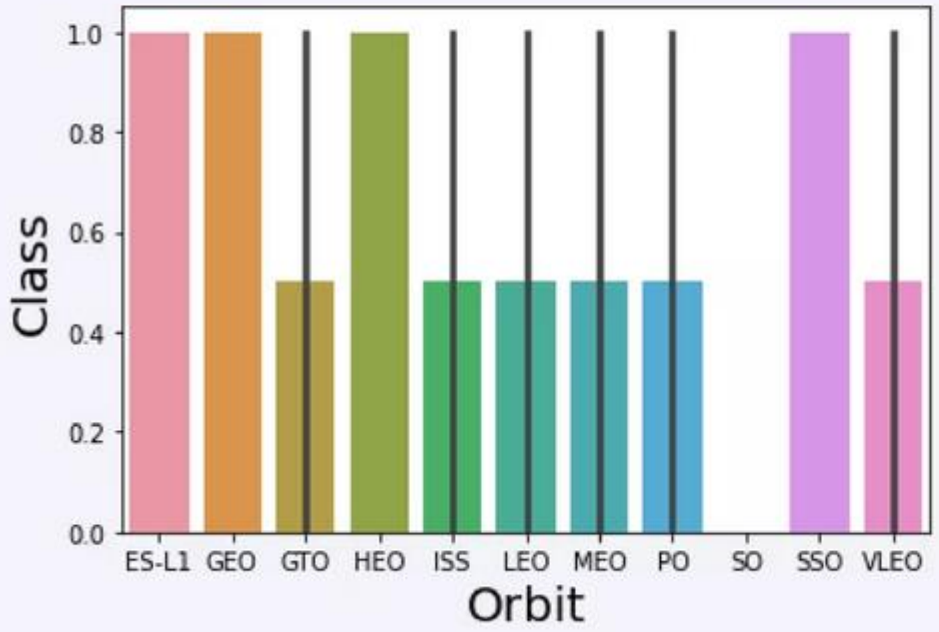
Payload vs. Launch Site

- Showing the correlation between launch Site and the payload mass



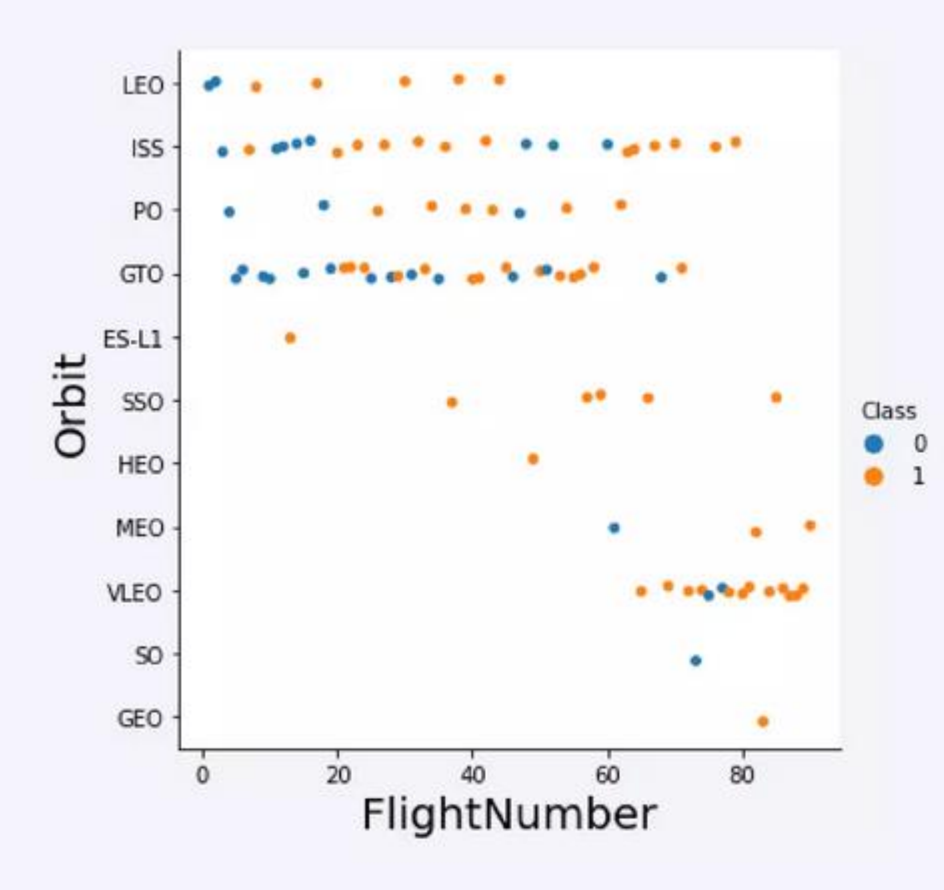
Success Rate vs. Orbit Type

- Success rate of each orbit



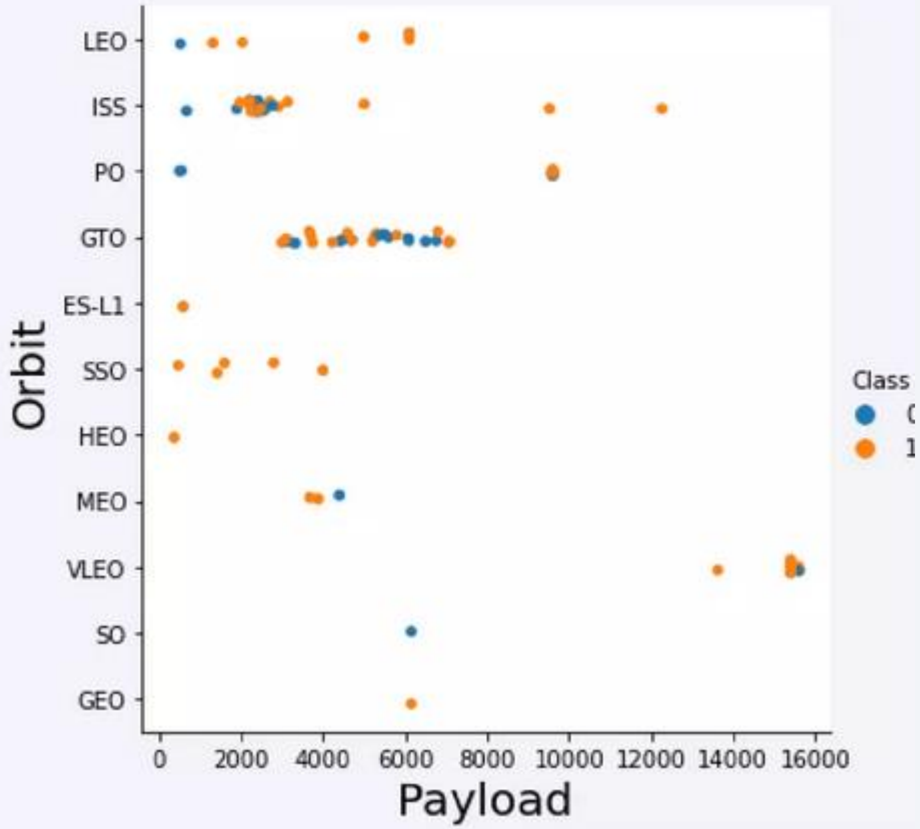
Flight Number vs. Orbit Type

- Types of orbit for each flight number.



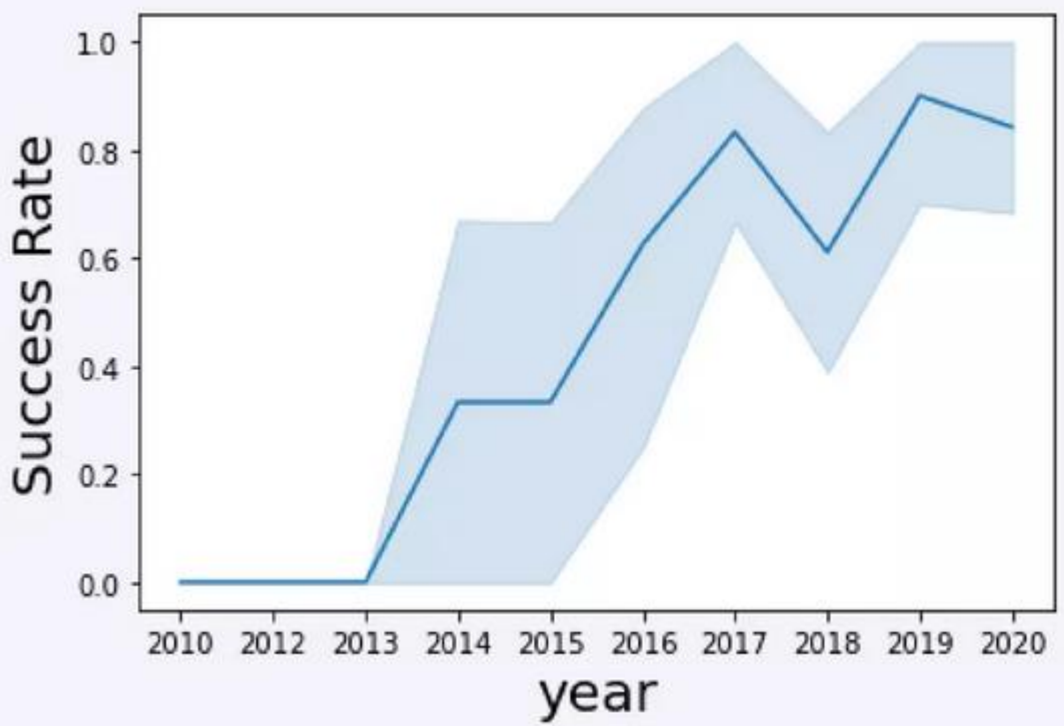
Payload vs. Orbit Type

- A scatter plot that explain the relationship between the payload mass and the orbit chosen



Launch Success Yearly Trend

- Launch success has experienced an uprising trend throughout the years



All Launch Site Names

- Find the names of the unique launch sites

- %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'

- Find 5 records where launch sites begin with `CCA`
- %sql select * 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

- Total payload carried by boosters from NASA
- % select sum(PAYLOAD_MASS_KG_) from PACEXTBL where CUSTOMER = NASA (CRS)

45596

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1
- %sql select avg(PAYLOAD_MASS_KG_) from SPACEXTBL where BOOSTER_VERSION = ‘F9 v1.1’

2928.400000

First Successful Ground Landing Date

- Dates of the first successful landing outcome on ground pad
- %sql select min(DATE) from SPACEXTBL where Landing_Outcome 0 'Success (ground pad)'

2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 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

- 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)’

100

Boosters Carried Maximum Payload

- Names of the booster which have carried the maximum payload mass
- %sql select 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

- Failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
- %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

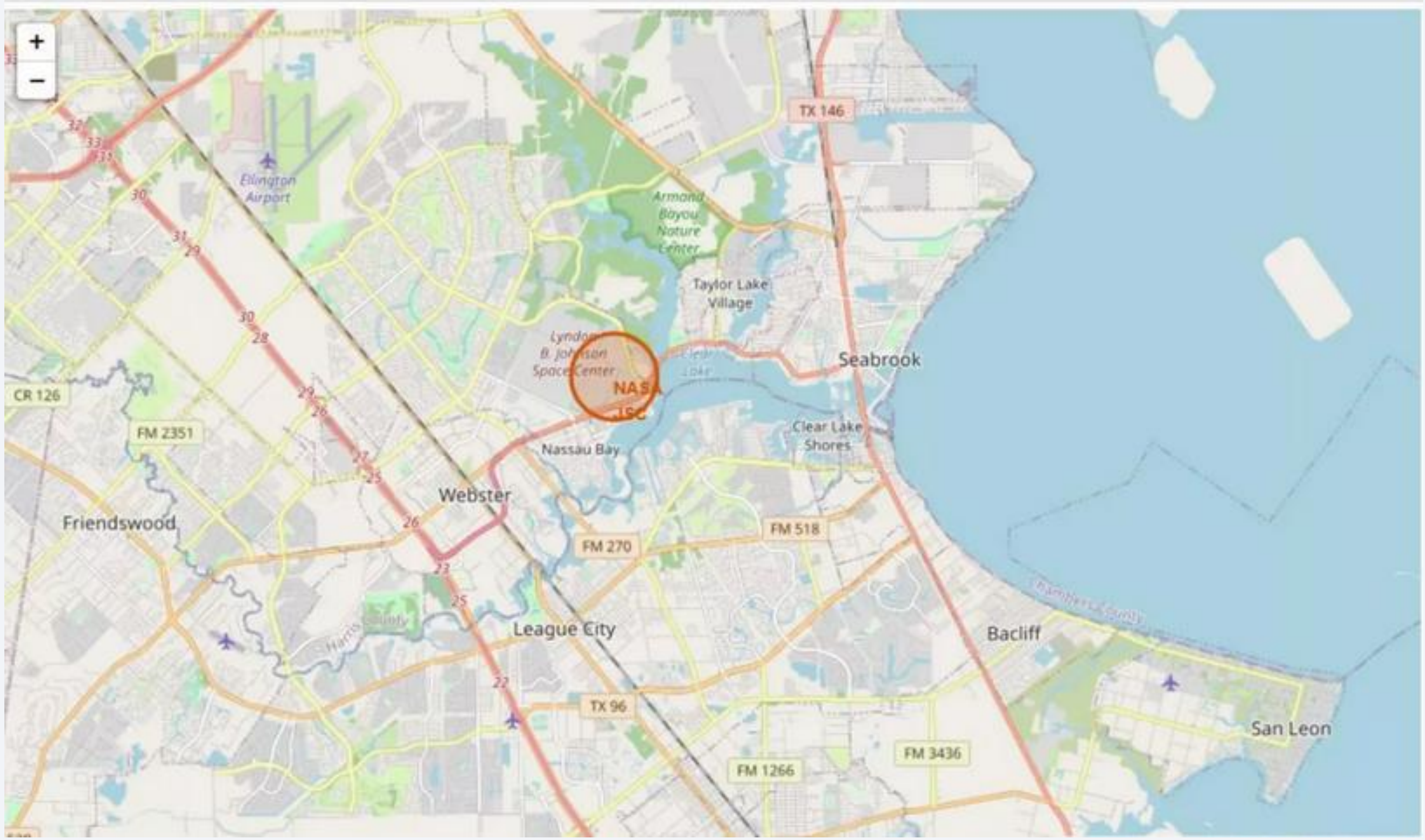
- Ranking the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order
- %sql select* from SPACEXTBL where Landing_Outcome like 'Success%' and (DATE between '2010-06-04' and '2017-03-20') order by date desc

2016-05-27	21:39:00	F9 FT B1023.1	CCAFS LC-40	Thaicom 8	3100	GTO	Thaicom	Success	Success (drone ship)
2016-05-06	05:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2016-04-08	20:43:00	F9 FT B1021.1	CCAFS LC-40	SpaceX CRS-8	3136	LEO (ISS)	NASA (CRS)	Success	Success (drone ship)
2015-12-22	01:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)

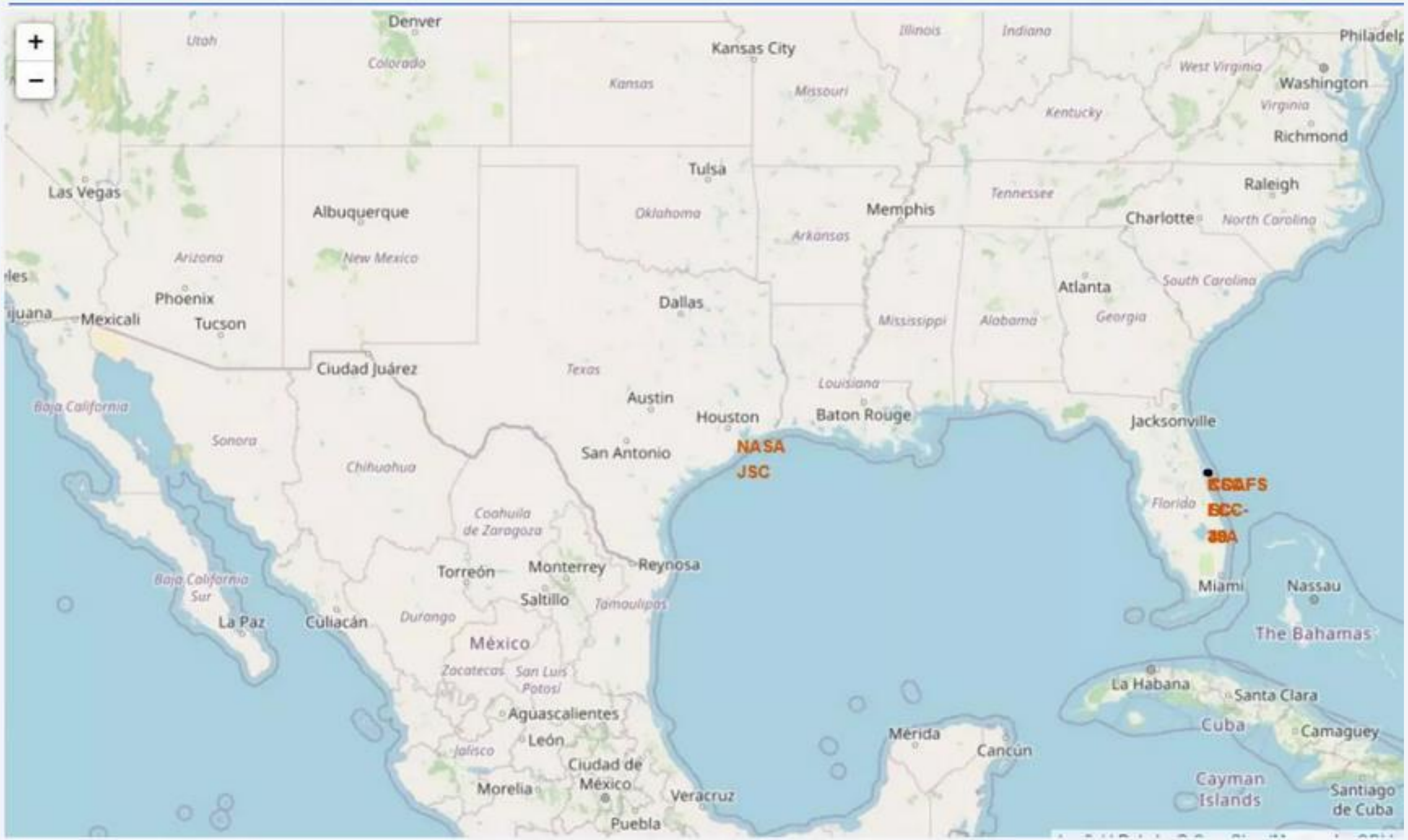
Section 3

Launch Sites Proximities Analysis

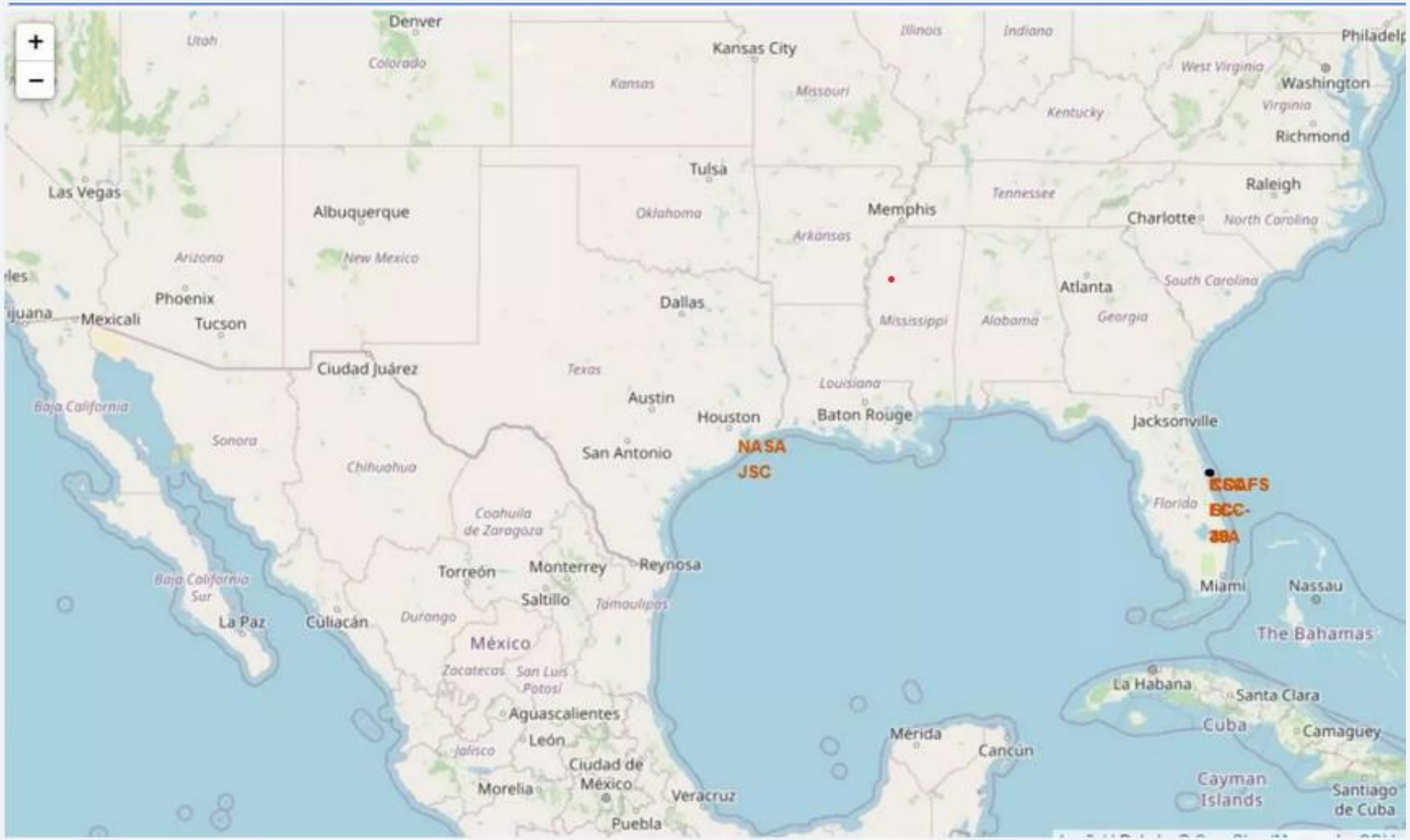
Launch Site Location



Successful vs Unsuccessful launches on a map



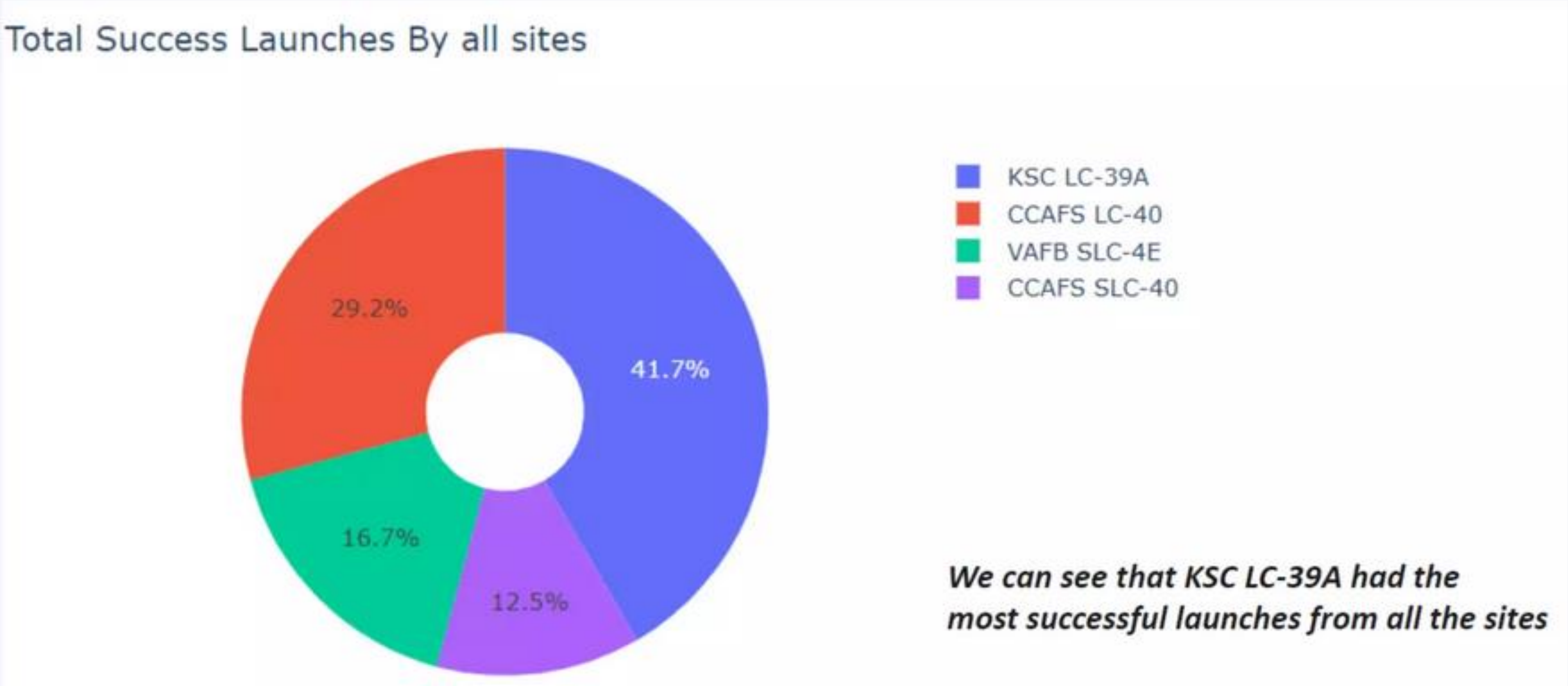
Distances from launch site to proximities



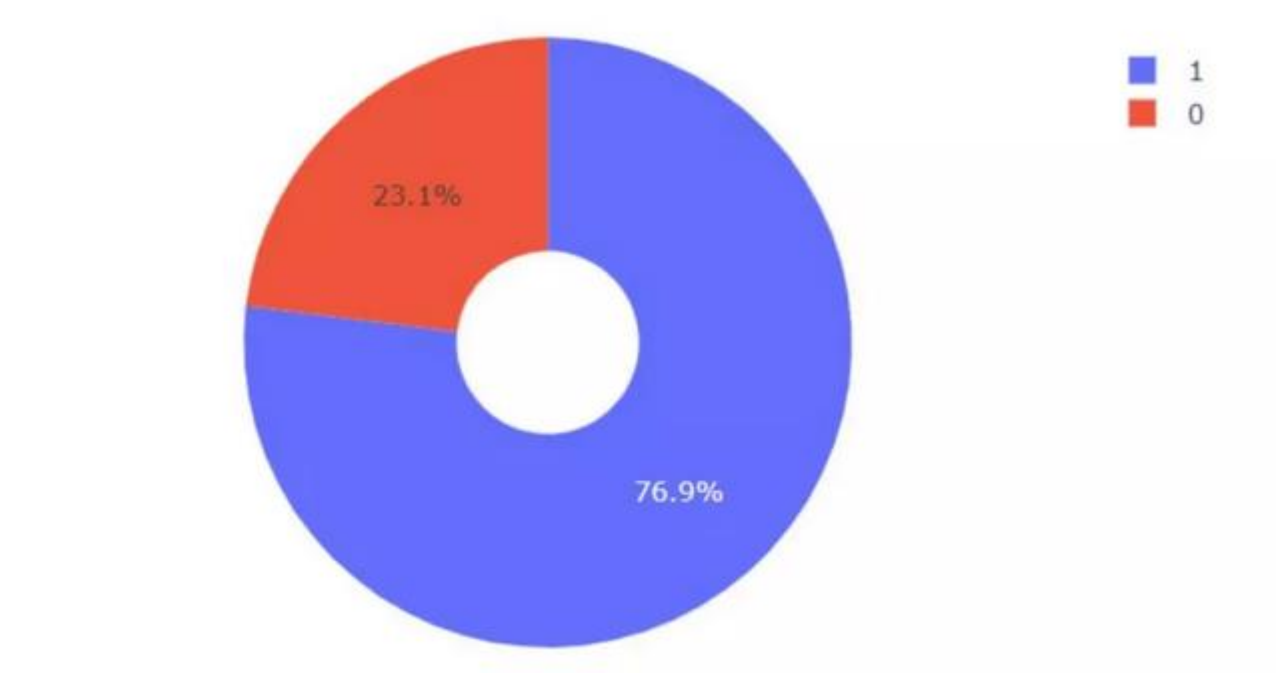
Section 4

Build a Dashboard with Plotly Dash

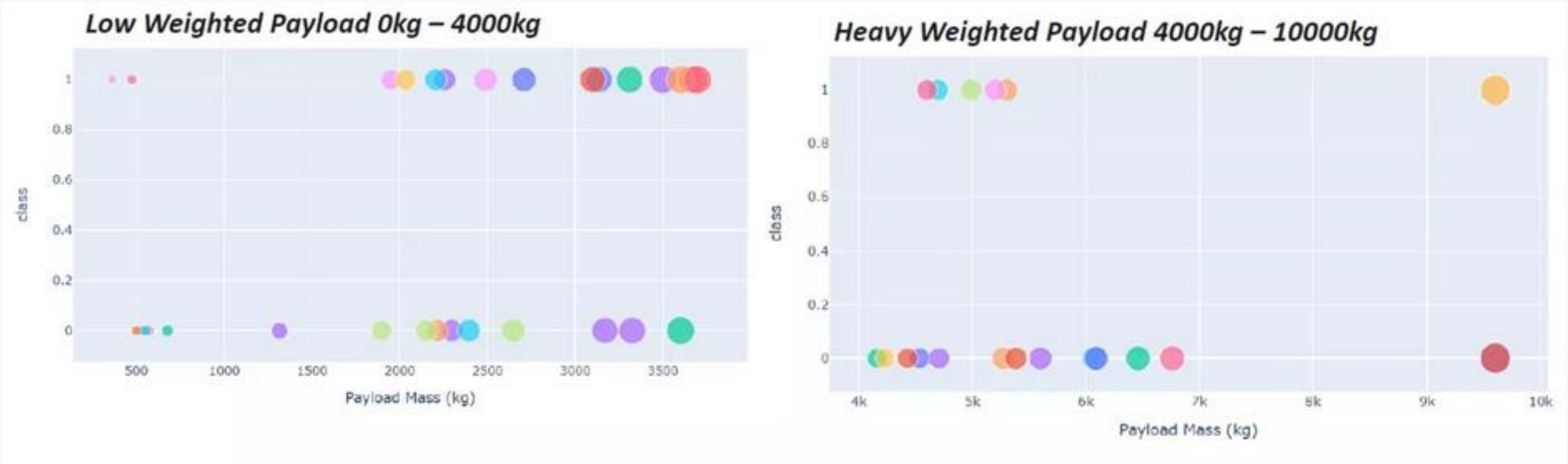
Success Percentages by Launch Site



Success Rate by Site



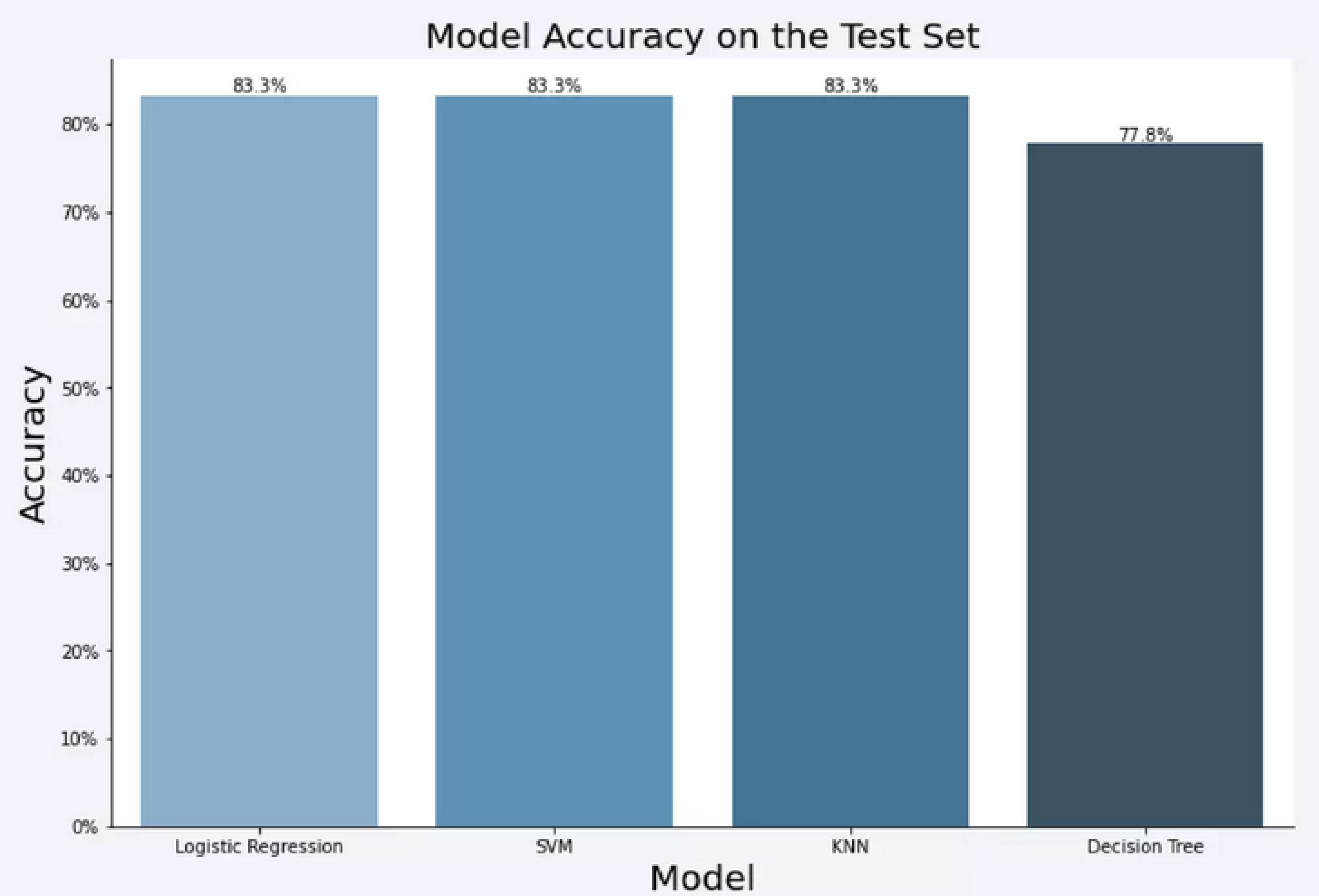
Payload Mass vs Outcome



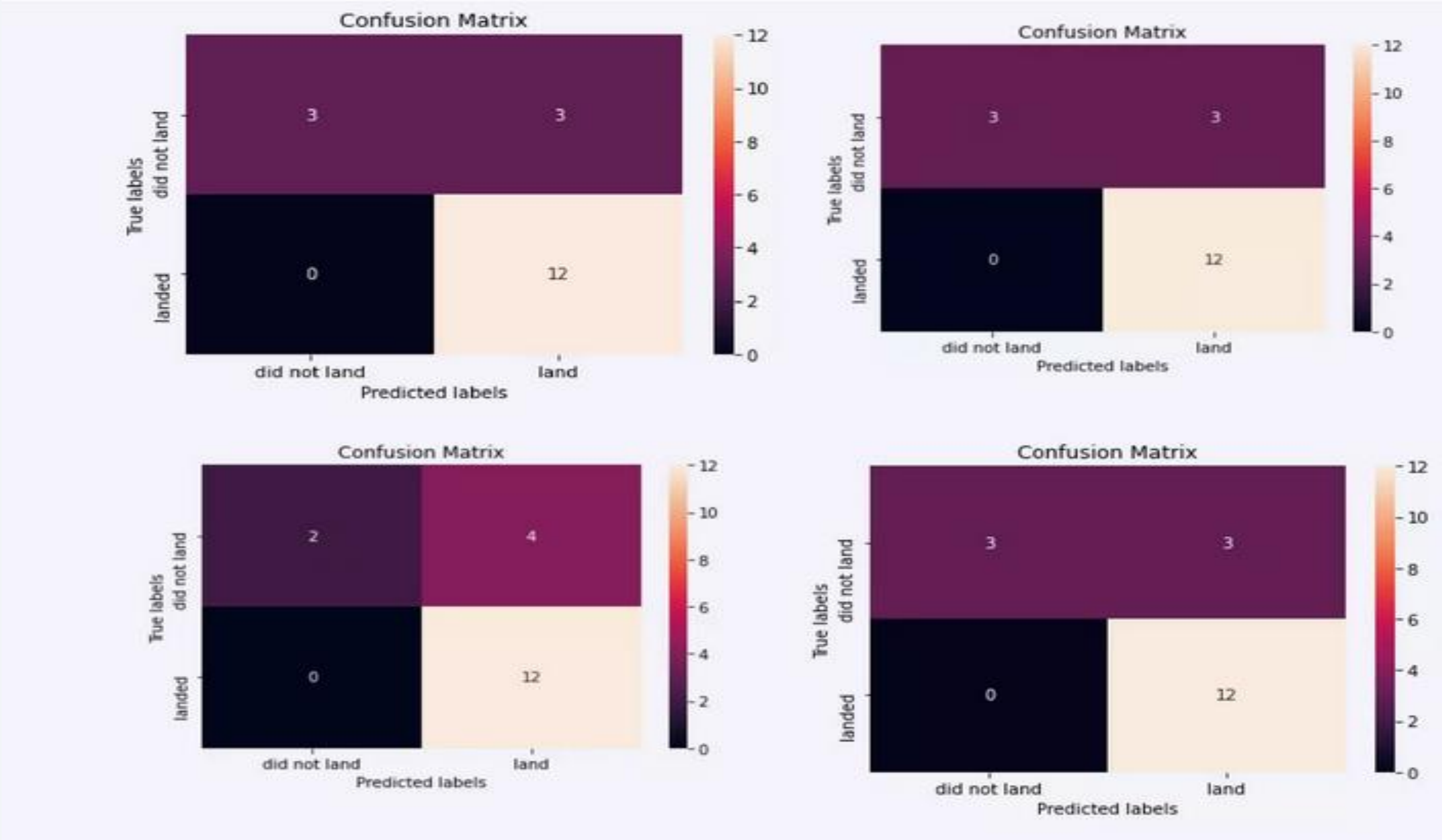
Section 5

Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix



Conclusions

- SVM, KNN and LR can all be used to predict the landing outcome with the same accuracy
- Heavier Payloads generally perform worse than lighter ones
- The success of SpaceX Landings has been increasing throughout the years
- Some launch sites have better performance, such as the KSC LS 39A. This might be because sites like this generally are used for launches with lighter payloads
- Some orbits like the GEO,HEO,SSO,ES L1 have better rates than other, the reasons for these orbital differences were not studied.

Thank you!