IBM Developer
SKILLS NETWORK

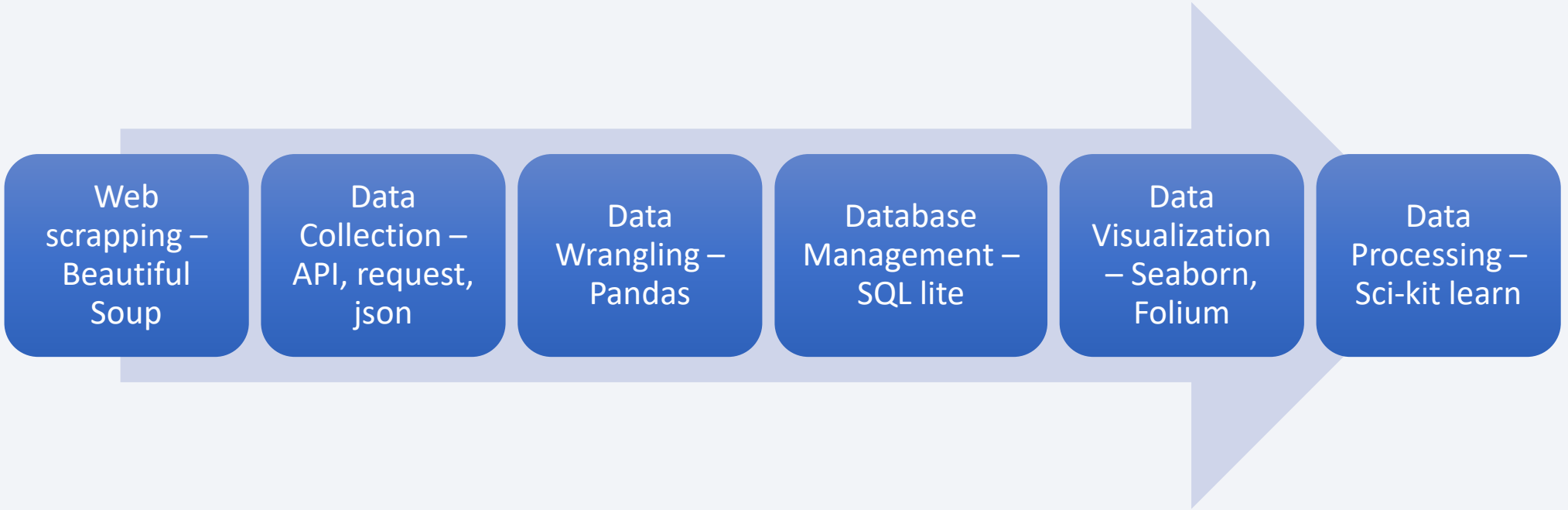
Winning Space Race with Data Science

Tomáš Vražina
30.11.2023

Outline

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

Executive Summary



Result

- Foremost features for successful landing were determined

Introduction

- SPACE Y wants compete with SPACE X
- As Data Scientist in SPACE Y company I need:
- Employ most important analytical Tools
- To maximize success rate during landing of space module

Section 1

Methodology

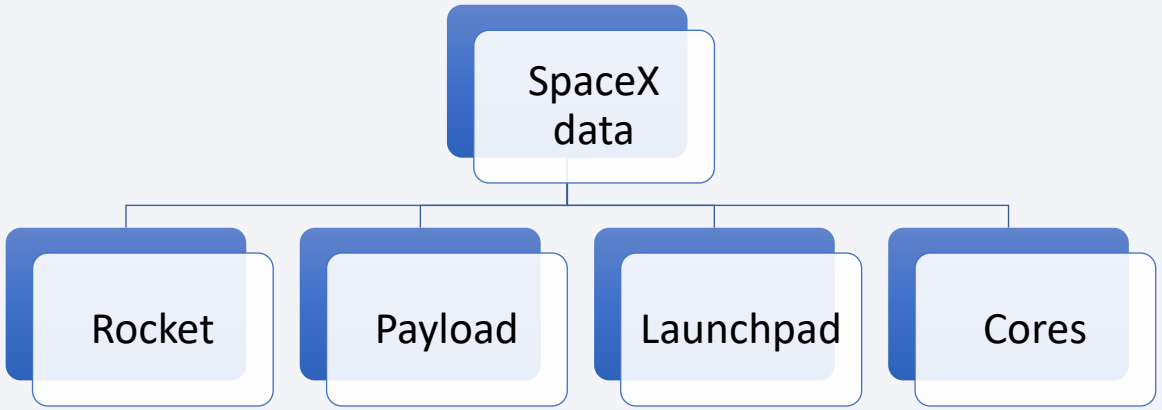
Methodology

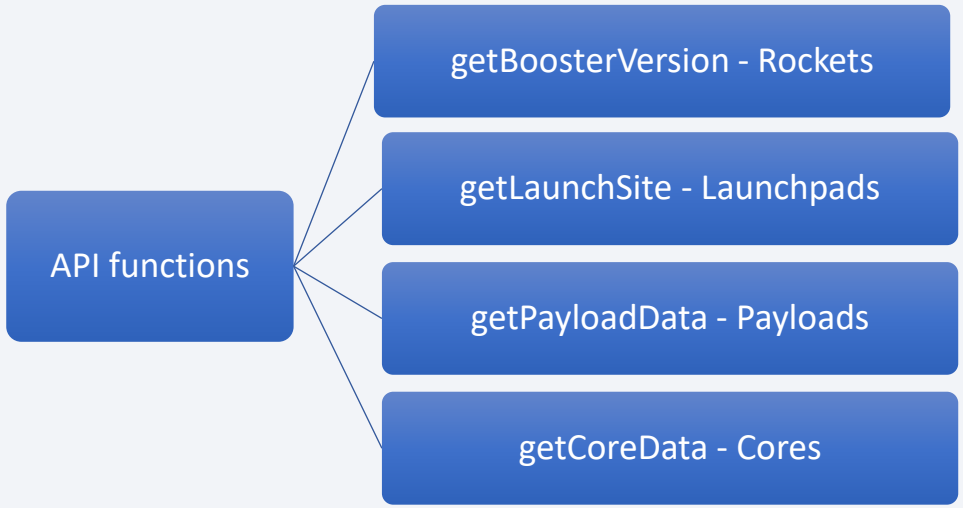
Executive Summary

- Data collection methodology:
  - Beautiful Soup for Data Scrapping
  - Request for downloading Data
- Data wrangling
  - Pandas for Data processing enables calculate: number of launches but mainly number and occurrence of each orbit with the respective outcome
- Exploratory data analysis (EDA) using visualization and SQL
- Interactive visual analytics using Folium and Dash
- Perform predictive analysis using classification models
  - Via Scikit-learn - Splitting data and using Grid Search for hyperparameter selection

Data Collection – SpaceX API

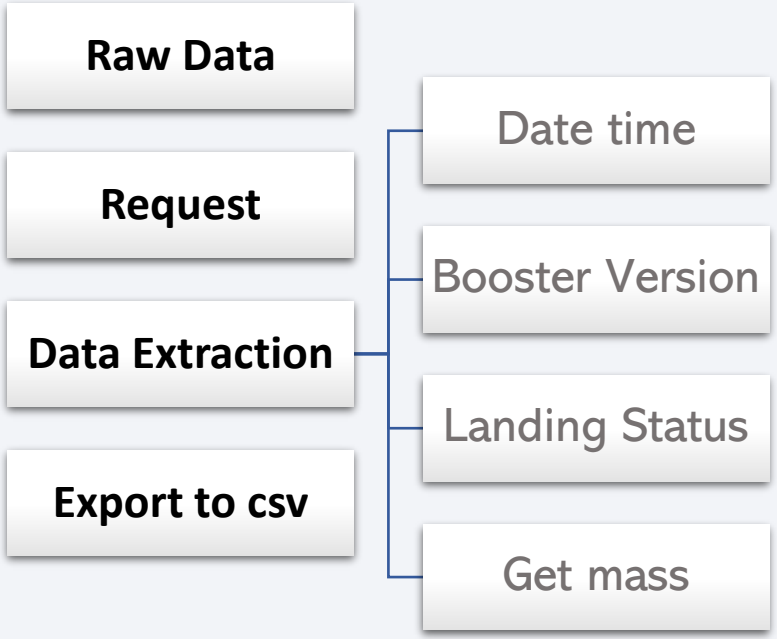
- Raw Data → API functions → JSON file → Pandas DF





- https://github.com/TomVrazina/IBM_CapstoneProject/blob/7571bda80bdb33c41820a1510736461a14a12915/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping



RAW Data

2010 to 2013

[hide] Flight No.	Date and time (UTC)	Version, Booster [b]	Launch site	Payload[c]	Payload mass	Orbit	Customer	Launch outcome	Booster landing
1	4 June 2010, 18:45	F9 v1.0[7] B0003.1[8]	CCAFS, SLC-40	Dragon Spacecraft Qualification Unit		LEO	SpaceX	Success	Failure[9][10] (parachute)
	First flight of Falcon 9 v1.0.[11] Used a boilerplate version of Dragon capsule which was not designed to separate from the second stage.(more details below) Attempted to recover the first stage by parachuting it into the ocean, but it burned up on reentry, before the parachutes even deployed.[12]								
2	8 December 2010, 15:43[13]	F9 v1.0[7] B0004.1[8]	CCAFS, SLC-40	Dragon demo flight C1 (Dragon C101)		LEO (ISS)	NASA (COTS) NRO	Success[9]	Failure[9][14] (parachute)
	Maiden flight of Dragon capsule, consisting of over 3 hours of testing thruster maneuvering and reentry.[15] Attempted to recover the first stage by parachuting it into the ocean, but it disintegrated upon reentry, before the parachutes were deployed.[12] (more details below) It also included two CubeSats,[16] and a wheel of Brouère cheese.								

	Flight No.	Launch site	Payload	Payload mass	Orbit	Customer	Launch outcome	Version Booster	Booster landing	Date	Time
0	[4 June 2010,, 18:45]	CCAFS	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success\n	F9 v1.0B0003.1	Failure	4 June 2010	18:45
1	[8 December 2010,, 15:43]	CCAFS	Dragon	0	LEO	NASA	Success	F9 v1.0B0004.1	Failure	8 December 2010	15:43
2	[22 May 2012,, 07:44]	CCAFS	Dragon	525 kg	LEO	NASA	Success	F9 v1.0B0005.1	Not attempted\n	22 May 2012	07:44
3	[8 October 2012,, 00:35]	CCAFS	SpaceX CRS-1	4,700 kg	LEO	NASA	Success\n	F9 v1.0B0006.1	No attempt	8 October 2012	00:35
4	[1 March 2013,, 15:10]	CCAFS	SpaceX CRS-2	4,877 kg	LEO	NASA	Success\n	F9 v1.0B0007.1	Not attempted\n	1 March 2013	15:10

- https://github.com/TomVrazina/IBM_CapstoneProject/blob/7571bda80bdb33c41820a1510736461a14a12915/jupyter-labs-webscraping.ipynb

Data Wrangling

- Goal was to simplify possible landing outcomes and transformed them to boolean values

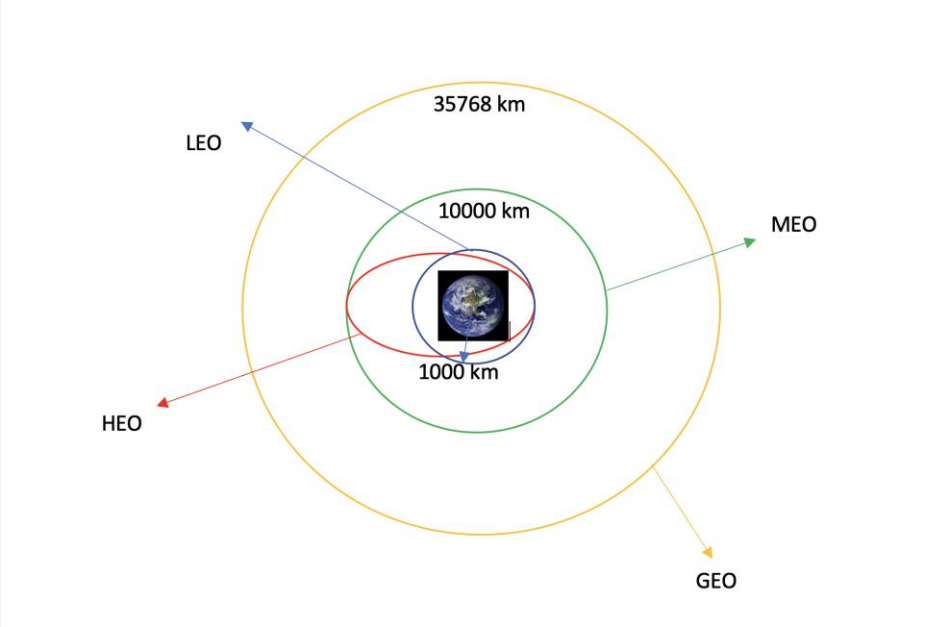
Calculation of number of each landing sites

Calculation of number and occurence of each orbit

Calculation of number and occurence of each outcome

Simplification of possible landing outcome

Export to CSV



https://github.com/TomVrazina/IBM_CapstoneProject/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

PLOT - REASON

Launch Site vs Flight number	Payload vs Launch Site	Succes rate of each orbit type	Flight number vs orbit type	Payload vs Orbit Type	Launch succes yearly
Different Launch site different succes rate	In VAFB-SLC no rocket launched for heavy payload	Simple analysis of another parameter	LEO orbit type succes related to Flight number	Connection of Payload with POLAR, LEO, ISS orbit-type	Succes rate increasing till 2017

https://github.com/TomVrazina/IBM_CapstoneProject/blob/7571bda80bdb33c41820a1510736461a14a12915/jupyter-labs-eda-dataviz.ipynb

EDA with SQL

- Selection of whole dataset
- Selection of data where launch site begin with string ‘CCA’
- Calculation of total payload mass carried by boosters
- Estimation of average payload mass carried by F9 v1.1
- Estimation of date of first succesful landing
- Determination of boosters with succes in droneship and weight range between 4000-6000
- Determinimation of number of succesful and failure mission outcomes
- Listing of defined columns in defined datetimes

https://github.com/TomVrazina/IBM_CapstoneProject/blob/7571bda80bdb33c41820a1510736461a14a12915/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Markers and circles for Launch Sites - to determine their position
- Cluster of succes/failed launches - to get better idea of succes rate with Launch sites
- Lines as a show off the distances between the launch sites and coasts

https://github.com/TomVrazina/IBM_CapstoneProject/blob/7571bda80bdb33c41820a1510736461a14a12915/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

Plots and Graphs:

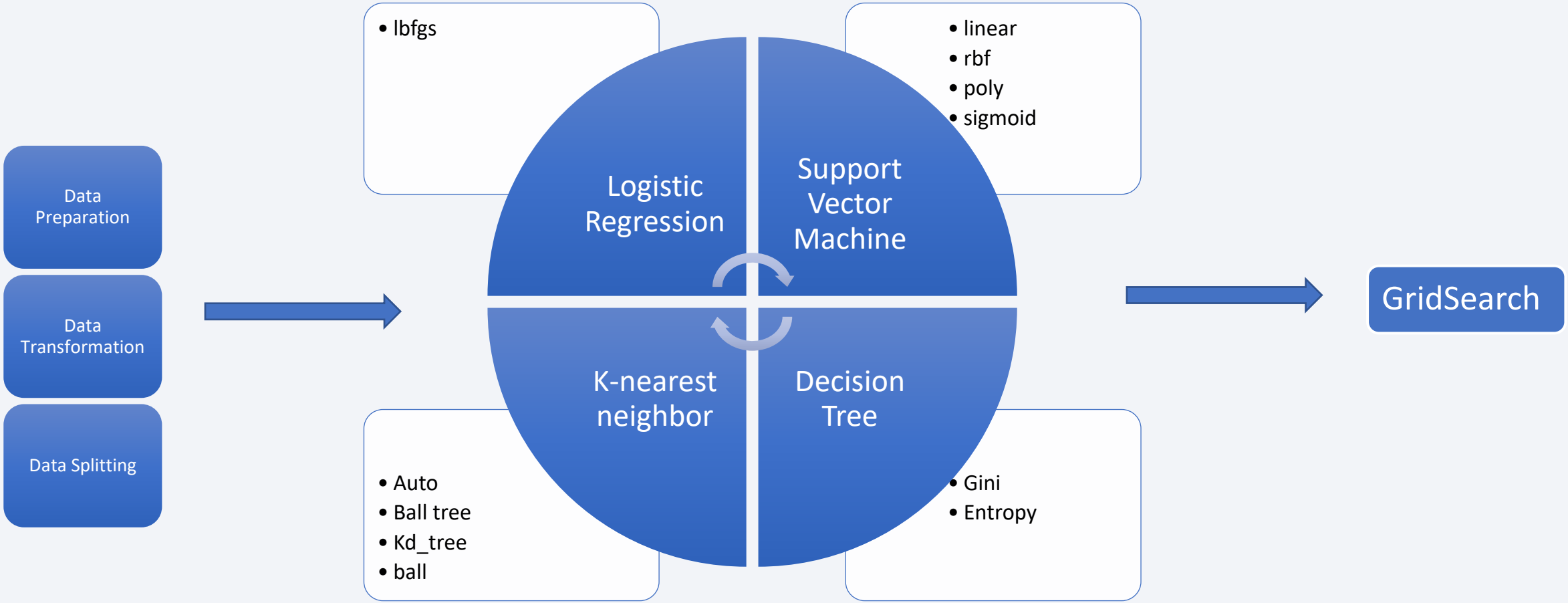
- **Pie chart** - Total launches and their success rate
- **Scatter Plot** – Payload Mass (kg) vs Mass

Interactions:

- **Drop-down** menu – Encompassing all launch sites
- **Slide range** – In range of 0 – 10 000

https://github.com/TomVrazina/IBM_CapstoneProject/blob/7571bda80bdb33c41820a1510736461a14a12915/spacex_dash_app.py

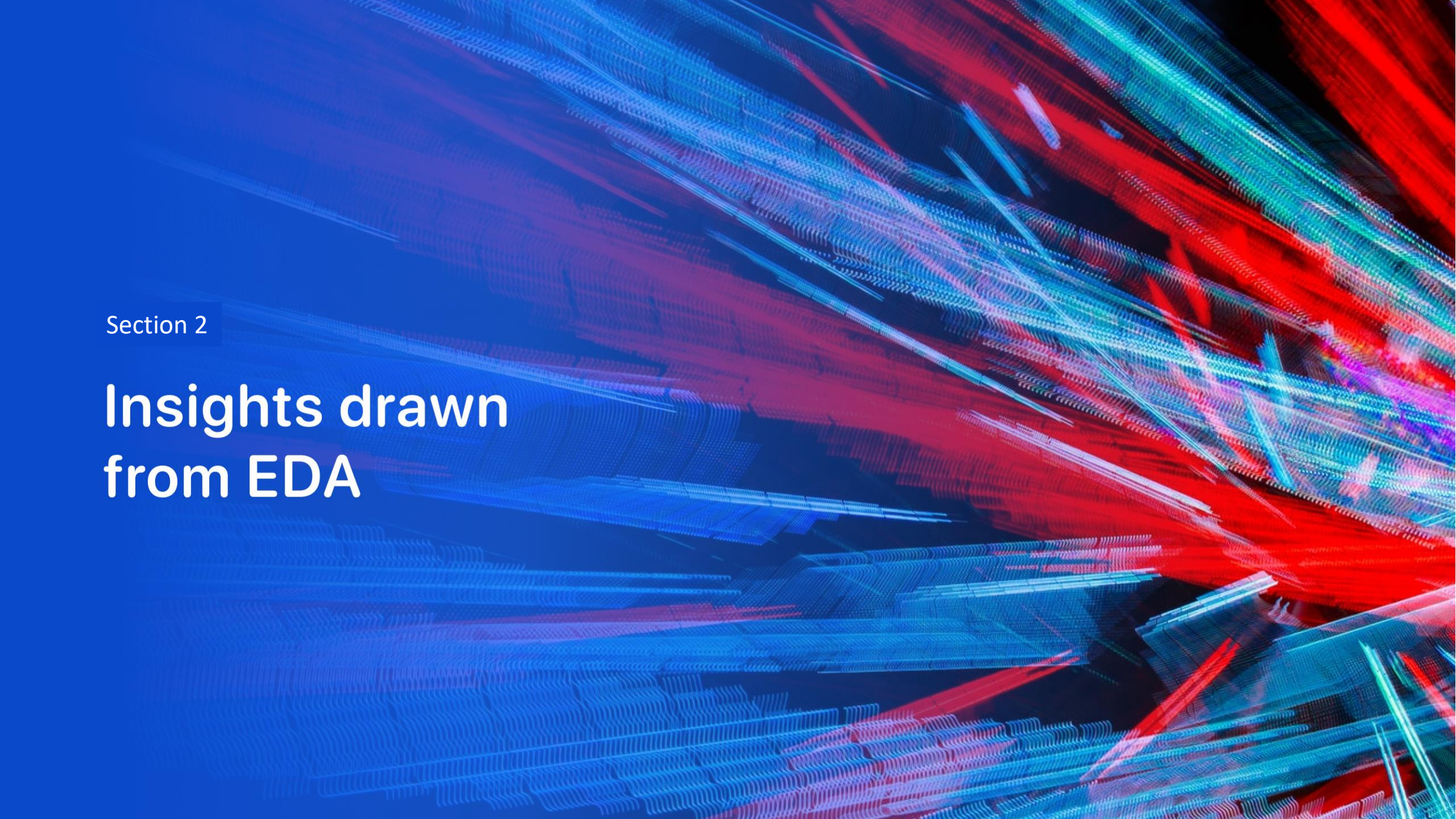
Predictive Analysis (Classification)



https://github.com/TomVrazina/IBM_CapstoneProject/blob/7571bda80bdb33c41820a1510736461a14a12915/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

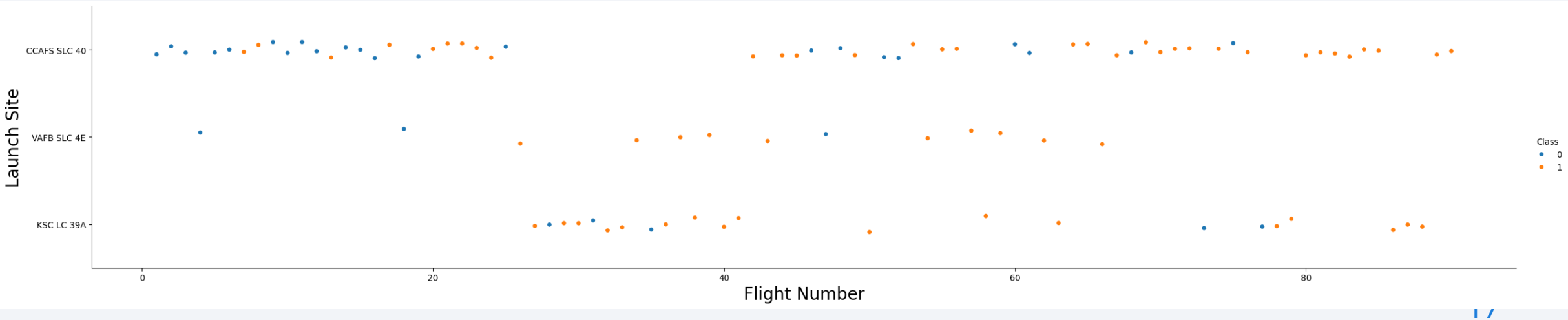
Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results


Section 2
Insights drawn from EDA

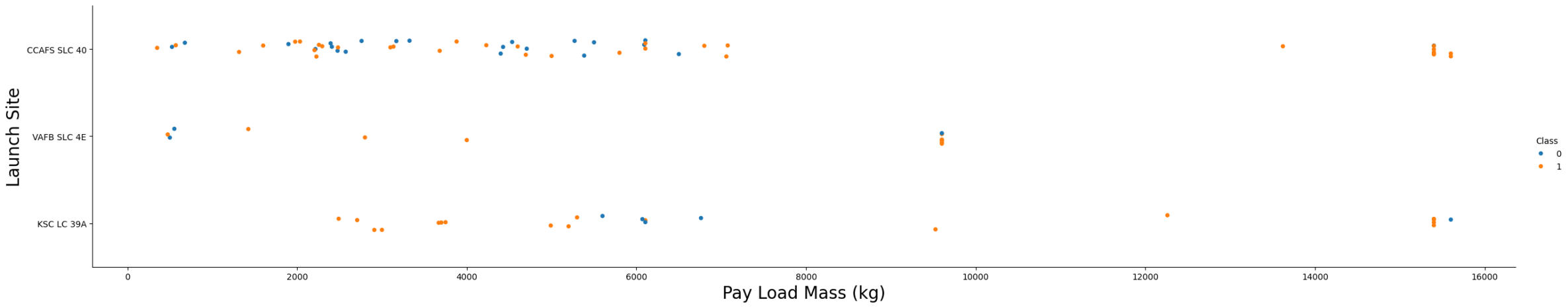
Flight Number vs. Launch Site

- Succes rate increased with increasing of Flight number
- Succes rate of CCAFS SLC-40 is lower then other two Launch sites



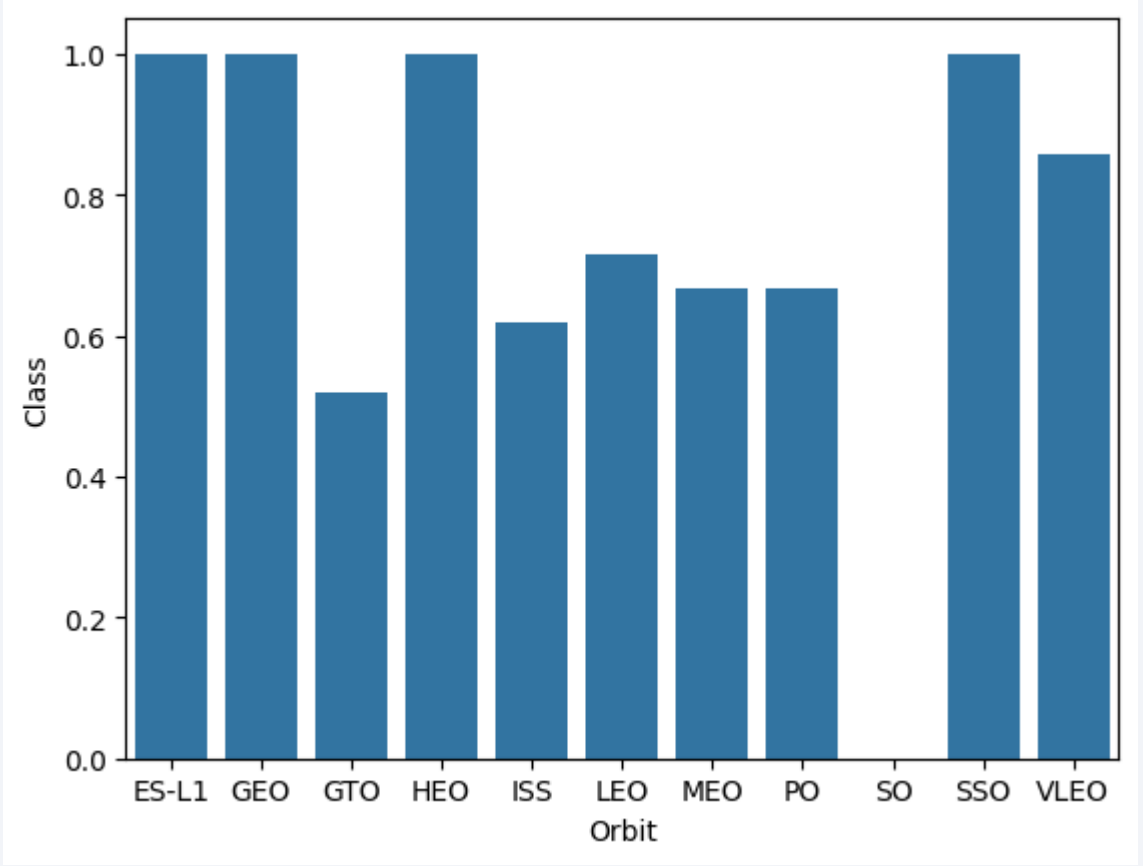
Payload vs. Launch Site

- For VAFB SLC 40 launch site there are no rockets launched for heavy payloads
- For CCAFS SLC 40 launch site there are no rockets launched for payloads in range of 8000-12000



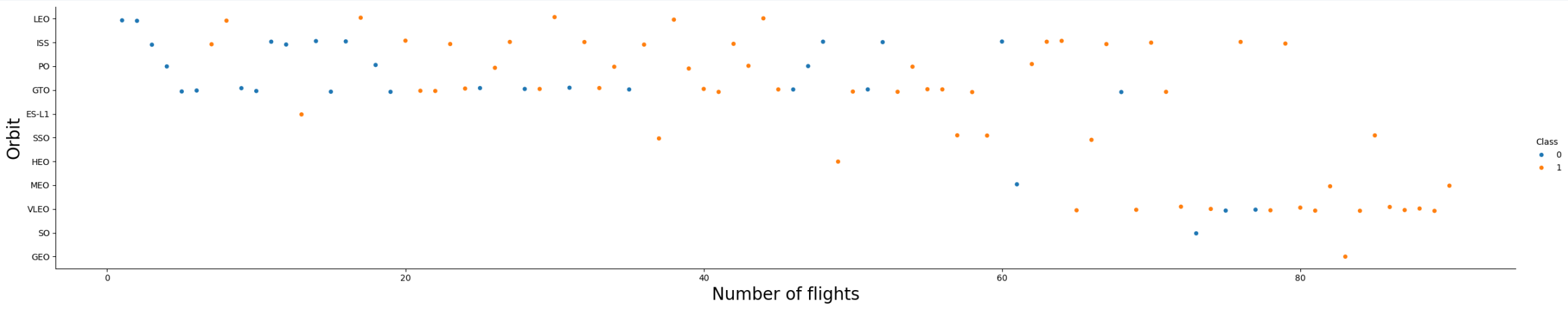
Success Rate vs. Orbit Type

- Highest succes rate was determined for ES-L1, GEO, HEO, SSO orbit type



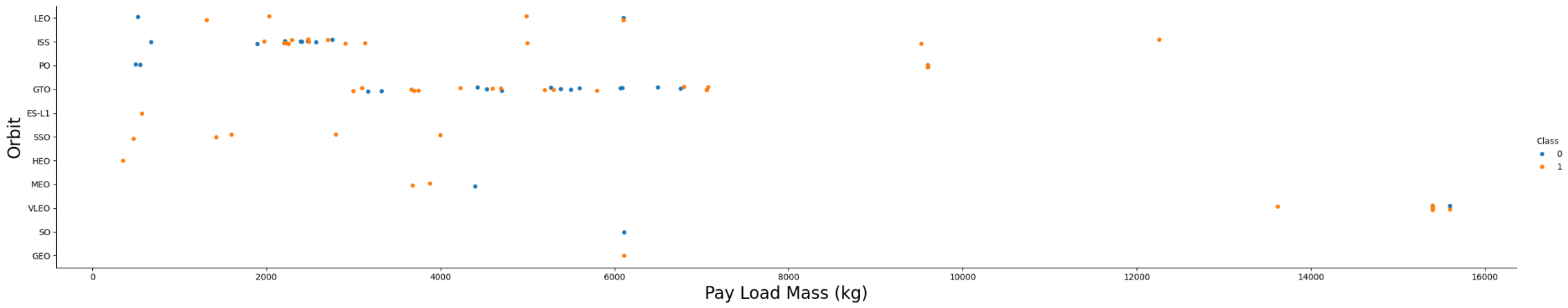
Flight Number vs. Orbit Type

- LEO orbits appears to be related to number of flights
- There is no relationship between between flight number and GTO orbit



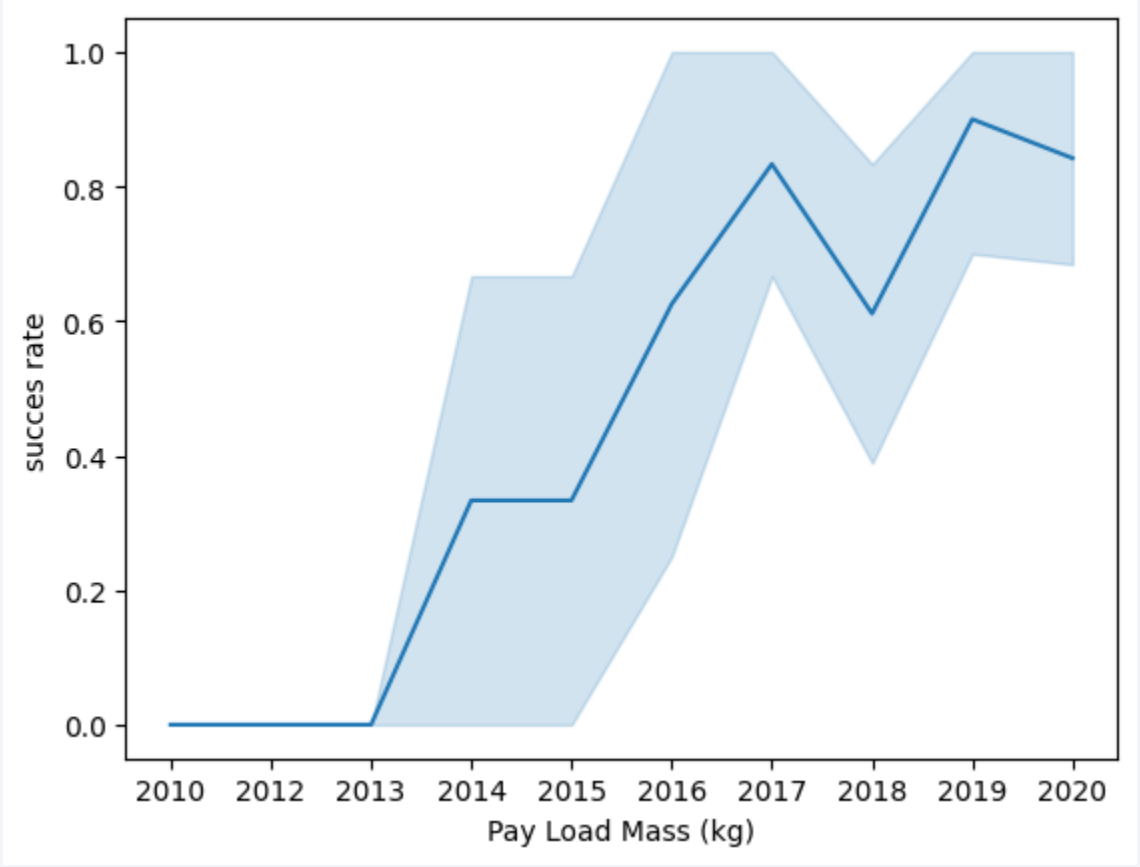
Payload vs. Orbit Type

- Most Succesful or positive landing happens with heavy payloads happend for POLAR, LEO, ISS
- But GTO has either positive and negative landigs based on Pay Load Mass (kg)



Launch Success Yearly Trend

- Succes rate kept increasing toll 2017
- In 2018 there is slight decrese
- But till that the value again increase



All Launch Site Names

- Only 4 unique launch sites exists:
- ,CCAFS LC-40
- VAFB SLC-4E
- KSC LC-39A
- CCAFS SLC-40

```
%sql select DISTINCT(Launch_Site) from SPACEXTBL
✓ 0.0s

 * sqlite:///my_data1.db
Done.
```

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

Launch Site Names Begin with 'CCA'

- 5 records of Launch site Name starting with letters “CCA”
- Like and sign % needs to be used
- Each Mission_outcome was succes in this query

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
6/4/2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
12/8/2010	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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
10/8/2012	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
3/1/2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- Total payload carried by boosters launched by NASA (CRS) is 45596 kg

SUM_of_Payload
45596

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 is 2928.4 kg
- Function AVG and command where was used

Booster_Version	Average_payload_mass_kg
F9 v1.1	2928.4

First Successful Ground Landing Date

- First succesful landing in ground pad was achieved na 2015-12-22
- MIN function was not very reliable so transformation based on date was used

date_format	Landing_Outcome
2015-12-22	Success (ground pad)

Successful Drone Ship Landing with Payload between 4000 and 6000

- There were only 4 Boosters where Payload was between 4000 and 6000 kg and they were succesful

Booster_Version	PAYLOAD_MASS__KG_	Landing_Outcome
F9 FT B1022	4696	Success (drone ship)
F9 FT B1026	4600	Success (drone ship)
F9 FT B1021.2	5300	Success (drone ship)
F9 FT B1031.2	5200	Success (drone ship)

Total Number of Successful and Failure Mission Outcomes

- There were only 1 Failure (in flight) as a Mission_outcome
- And 100 Succesful Mission Outcomes
- COUNT function was used
- Succes has various instances, they need to be transformed

Mission_Outcome	COUNT(Mission_Outcome)
Failure (in flight)	1
Success	100

Boosters Carried Maximum Payload

- Maximum payload mass was 15600 kg
- Subquery was used to get whole list

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

2015 Launch Records

- There were only 2 occurences of Failures in 2015
- For this purpose new columns „MonthName“ and „YEAR“ needed to be defined
- Subquery was employed

Date	MonthName	YEAR	Landing_Outcome	Booster_Version	Launch_Site
1/10/2015	October	2015	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
14/04/2015	April	2015	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

- Count of Landing outcomes with their respective outcome within dates 2010-06-04 and 2017-03-20

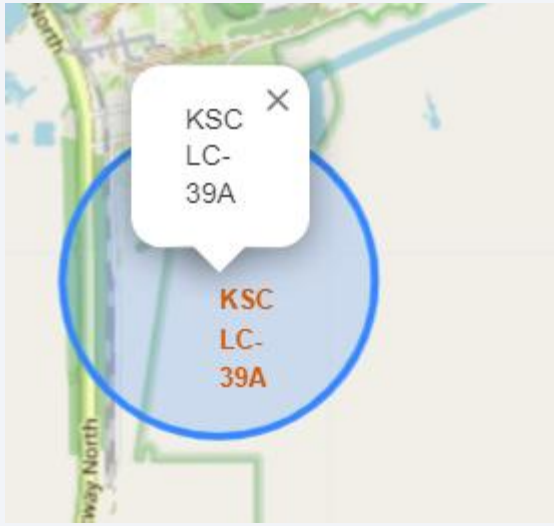
date_format	Landing_Outcome	COUNT(Landing_Outcome)
2017-01-14	Success (drone ship)	5
2016-07-18	Success (ground pad)	5
2015-10-01	Failure (drone ship)	5
2015-06-28	Precluded (drone ship)	1
2014-09-21	Uncontrolled (ocean)	2
2014-07-14	Controlled (ocean)	3
2014-06-01	No attempt	10
2010-08-12	Failure (parachute)	1

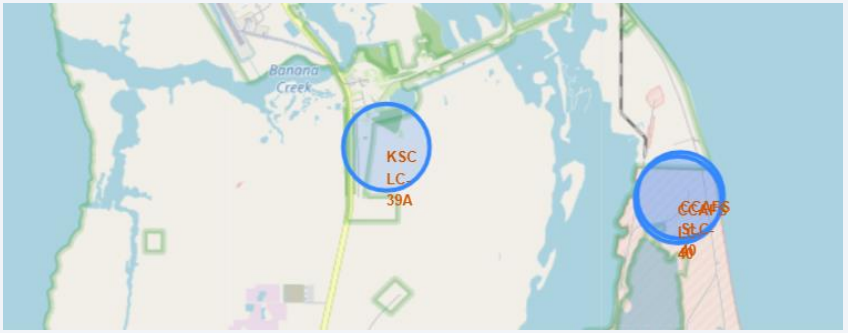
Section 3

Launch Sites Proximities Analysis

Folium map with launch sites

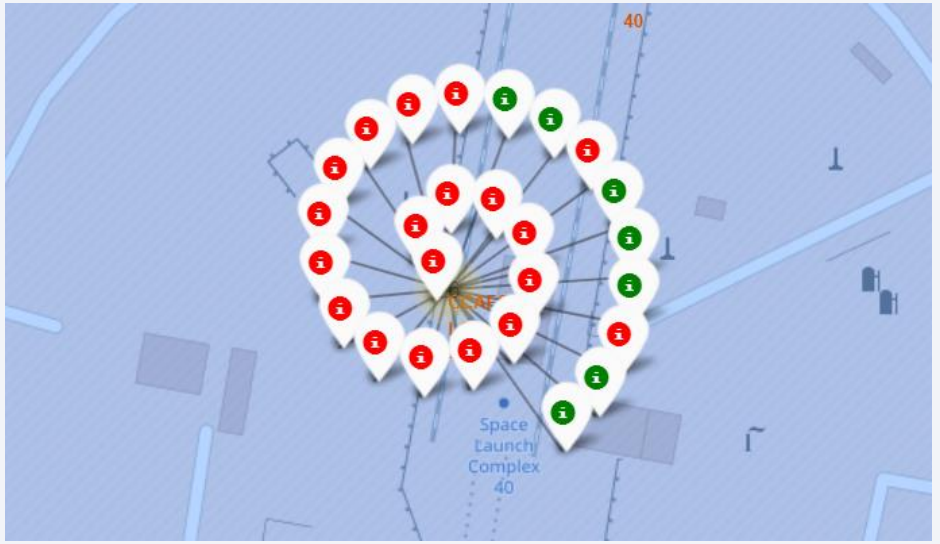
- 3 Launch sites determined by circle are located in basically same position in the east side of US
- 1 Launch site VAFB SLC 4E is located onto west side
- Each circle is labeled





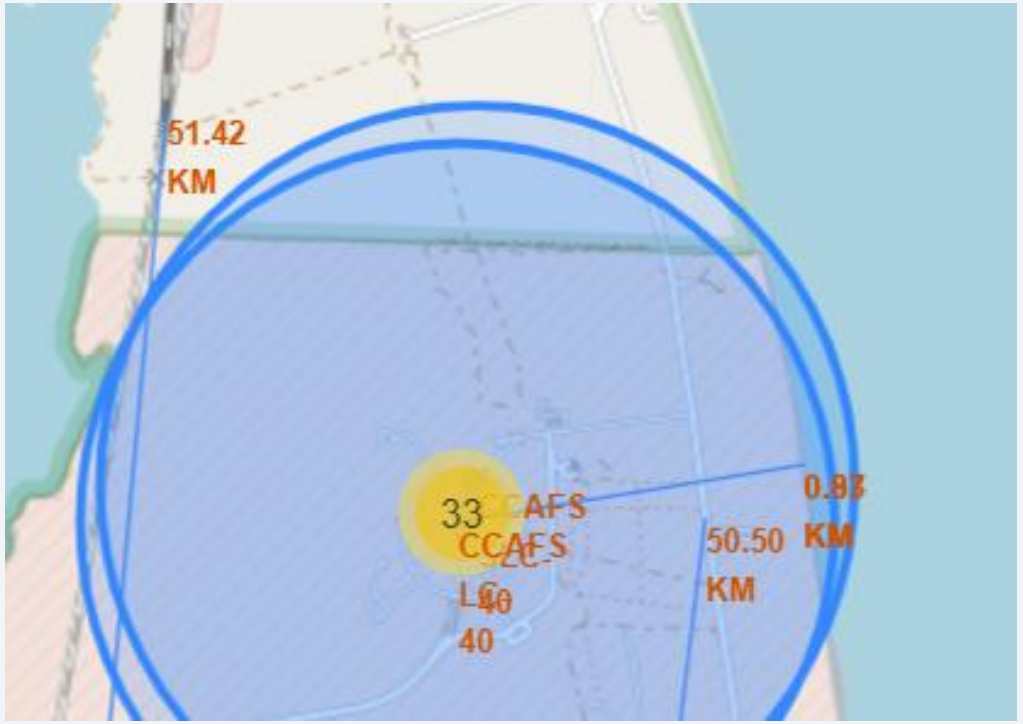
Map with Succes/failed launches

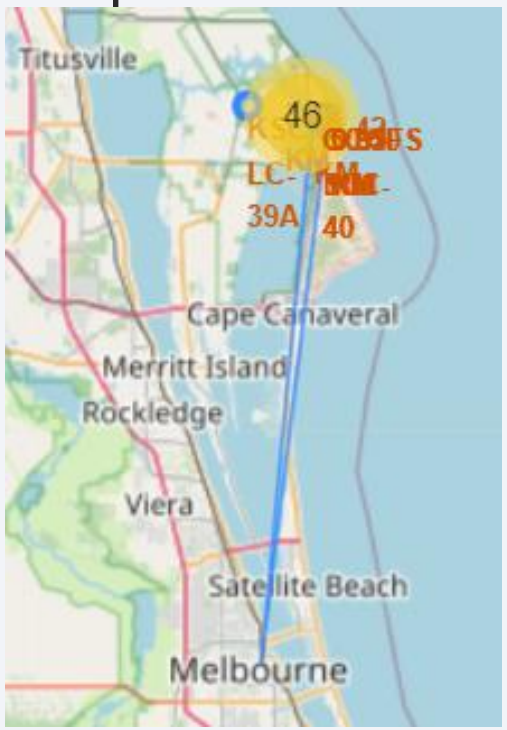
- Colour labeled outcomes are set to show up when they are clicked
- Onto First map is it possible to see how many of this are placed in map

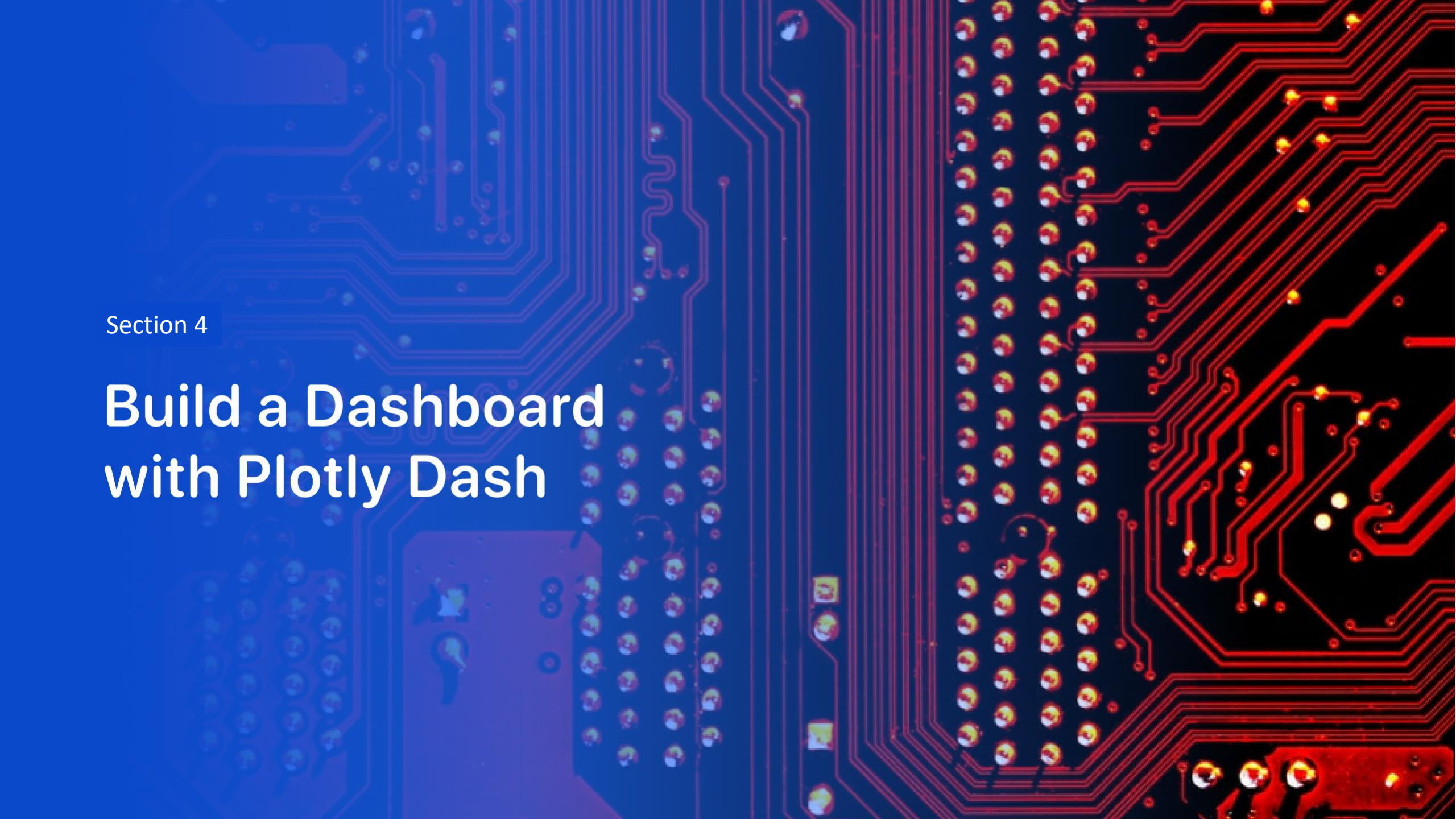


Folium map distances from launch site

- Lines from Launch site, highway, railway to the closest city (Melbourne) is shown
- Also Distances from launch site to coast is incorporated

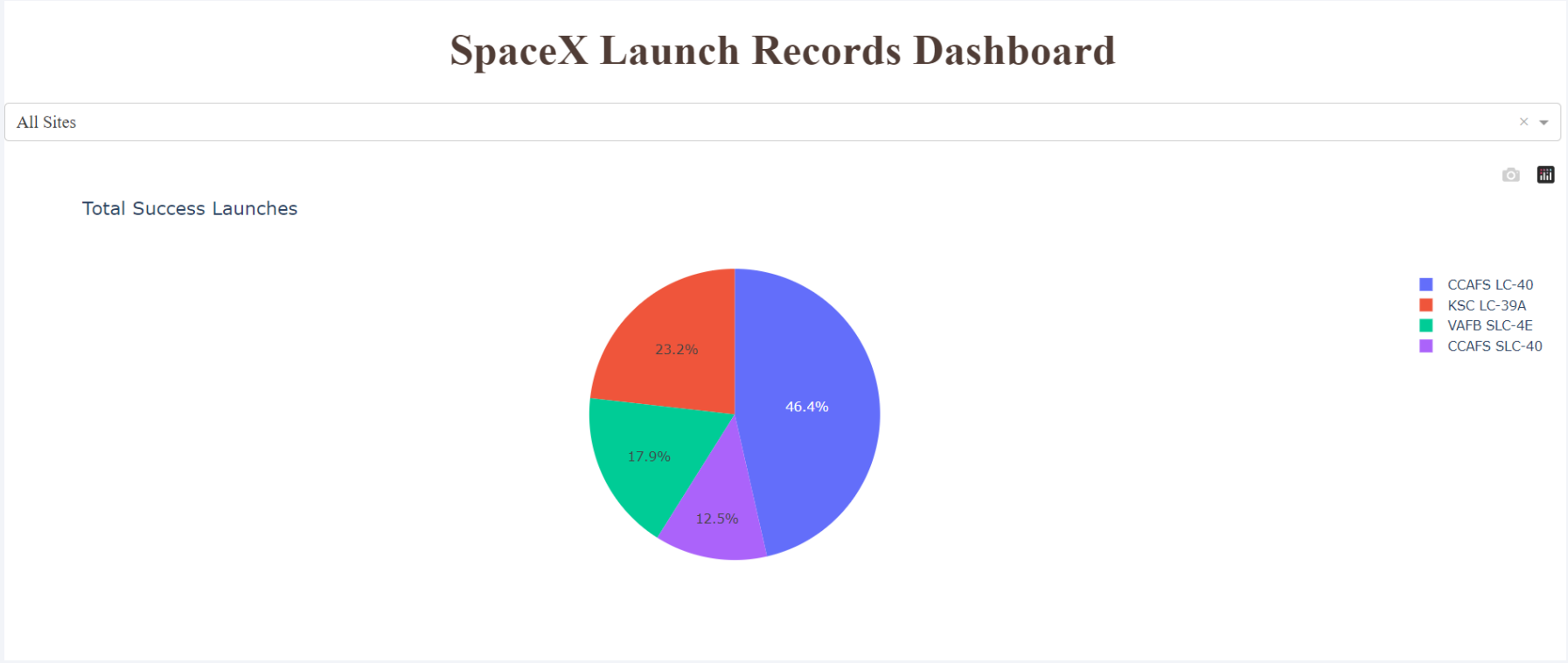





Section 4
Build a Dashboard with Plotly Dash

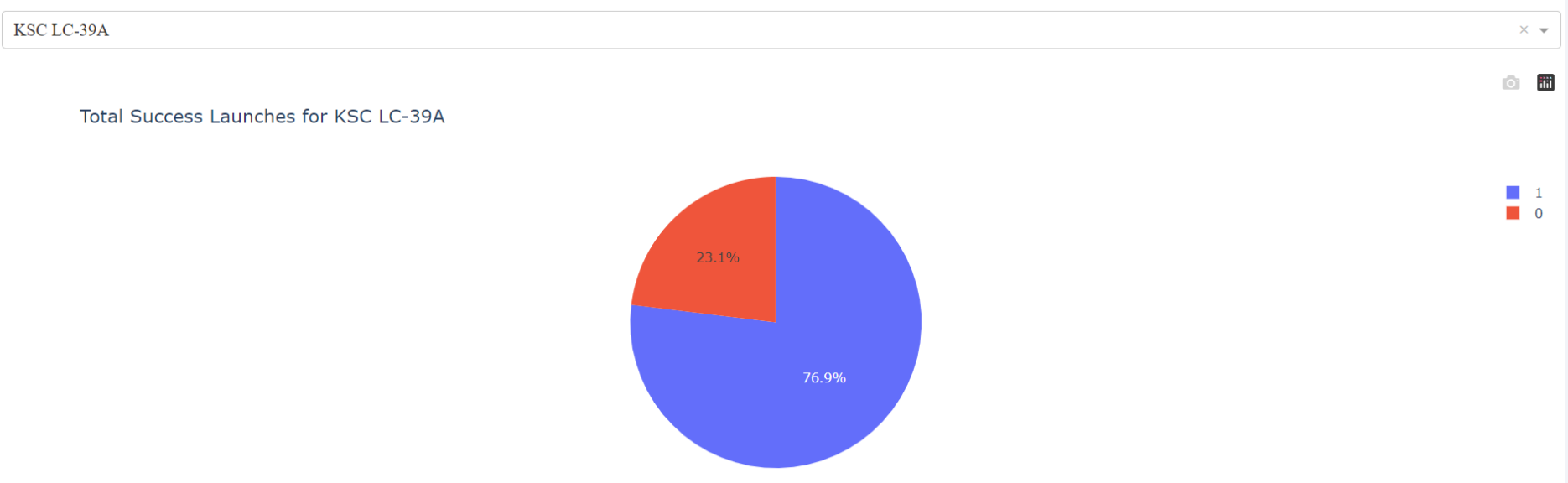
Total succes launches PieChart

- The number of launch sites is not consistent so piechart can cause little bit of mystifacation



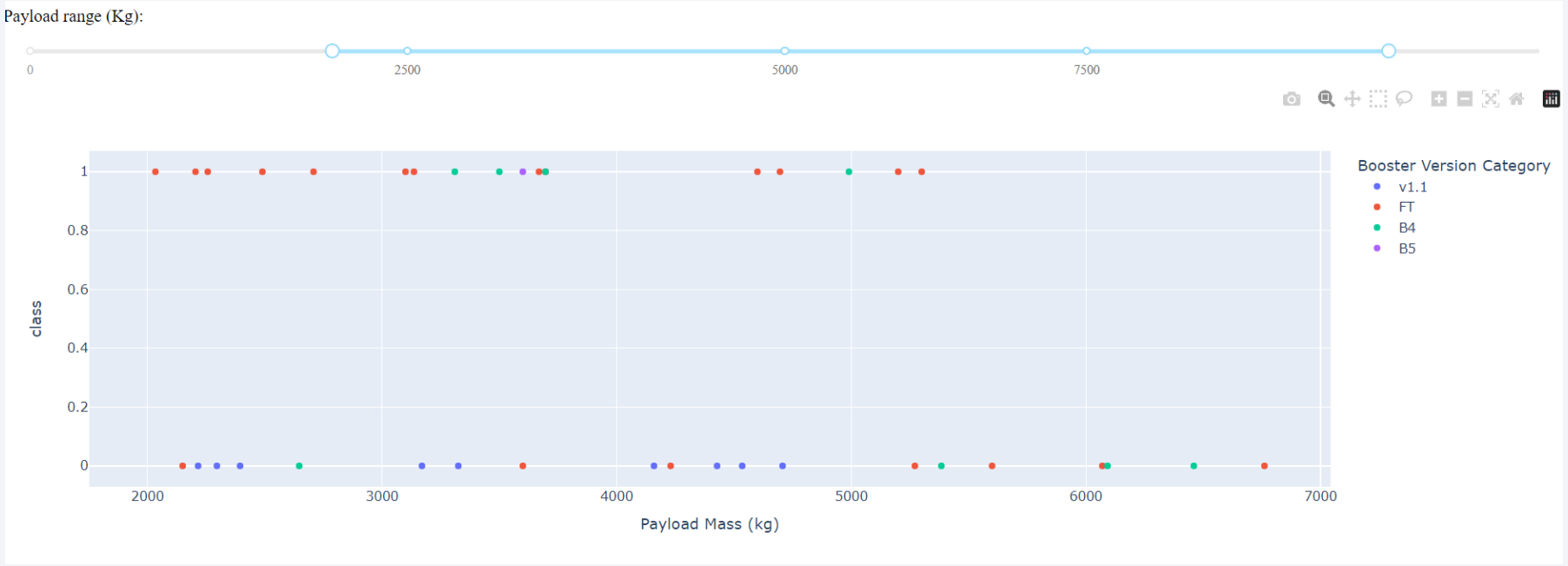
Total Succes Launches for KSC-LC-39A

- The highest ratio for succes/failed launches is for KSC-LC-39A Launch Site



Payload vs Launch Outcome

- Highest succes rate has FT booster version in Payload range 2300-9000
- And Lowest succes rate in the Payload range is for v1.1

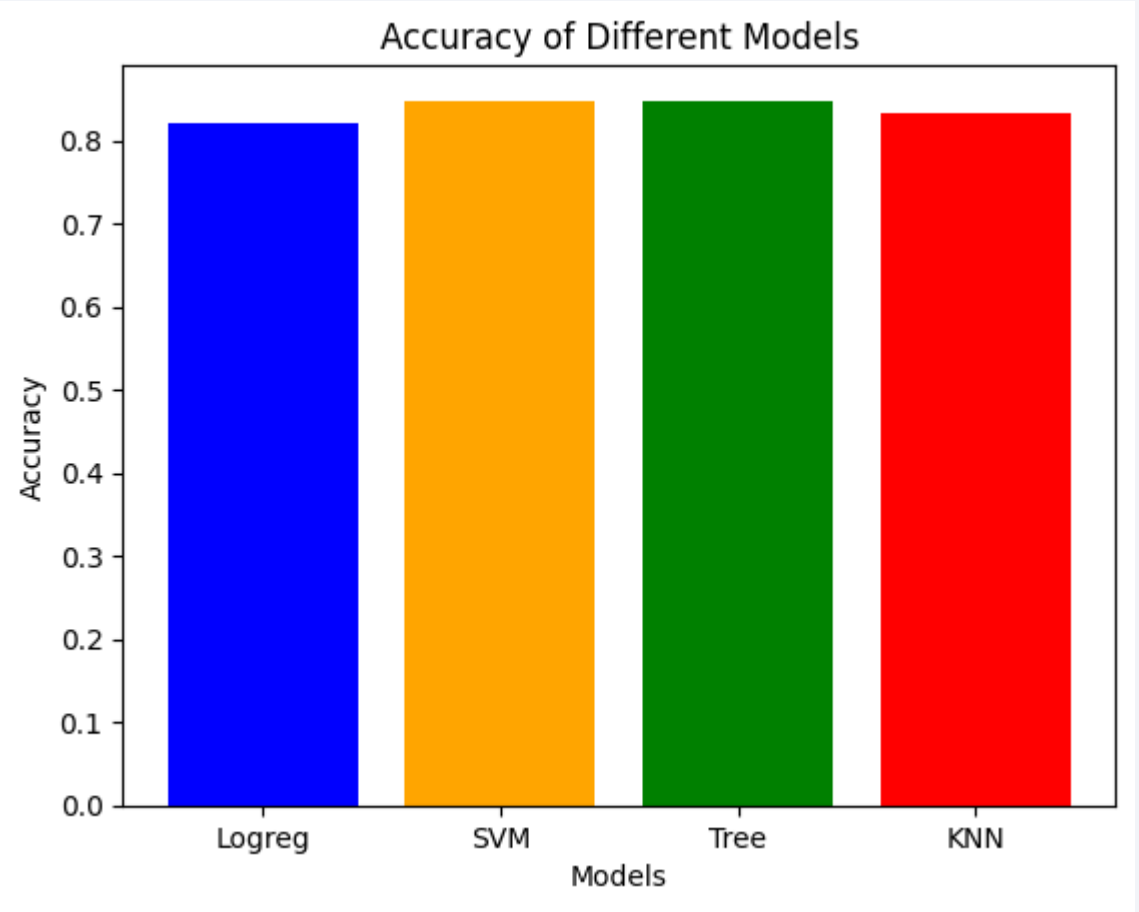


Section 5

Predictive Analysis (Classification)

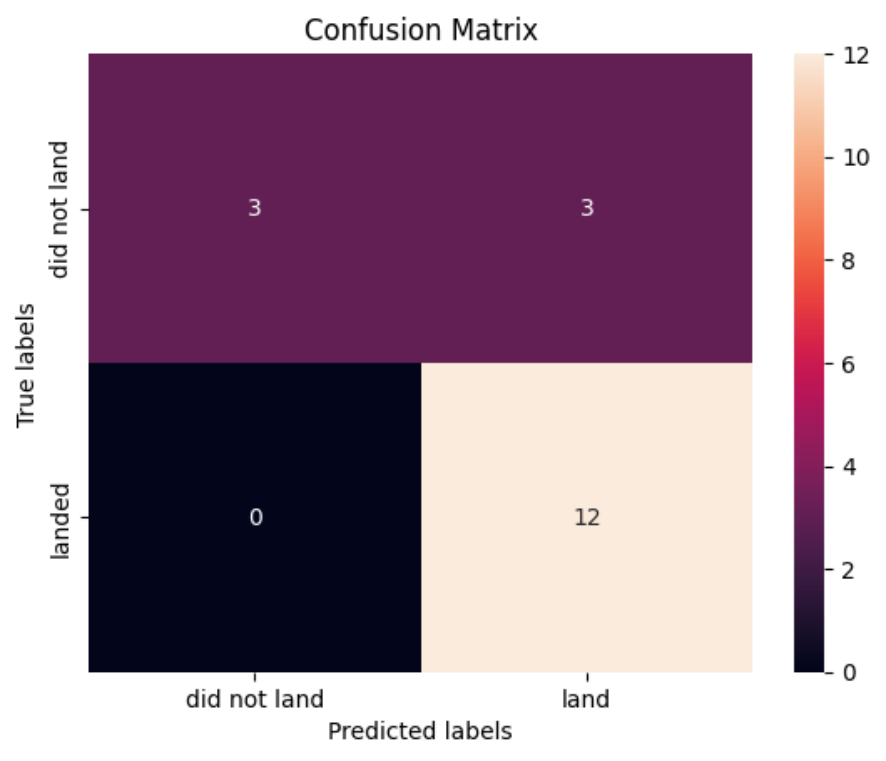
Classification Accuracy

- Highest classification accuracy has the SVM model – train data
- Score – train to test data is same for all of them



Confusion Matrix

- The SVM model has more of Predicted labels in column land when the did not land



Conclusions

- Lauch succes rate through years is increasing
- The highest ratio for succes/failed launches is for KSC-LC-39A Launch Site
- Most Succesful or positive landing happens with heavy payloads happend for POLAR, LEO, ISS
- Best model for training of this data is SVM

Appendix

- Data_set_part1.csv
- Data_set_part2.csv
- Data_set_part3.csv
- Space_launch_dash.csv
- Space_launch_geo.csv
- Space_web_scraped.csv

Thank you!