predictive analysis of SpaceX first stage landing success rate

IBM Data Science capstone

Bin Wu

Nov. 10,2021

SPACEX

Outline

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

Executive Summary

- **Summary of methodologies**

  Data Collection
  Data Wrangling
  EDA with SQL
  EDA with data visualization

  Building an interactive map with Folium

  Building a Dashboard with plotly Dash

  Predictive analysis (Classification)

- **Summary of all results**

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

— Design by All-free-download.com —

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, much of the savings is because SpaceX can reuse the first stage. Therefore if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

- **Problems you want to find answers**

  1. What's the landing success rate in different situations?

  2. What factors affect the success rate?

  3. How to improve the success rate?

Section 1

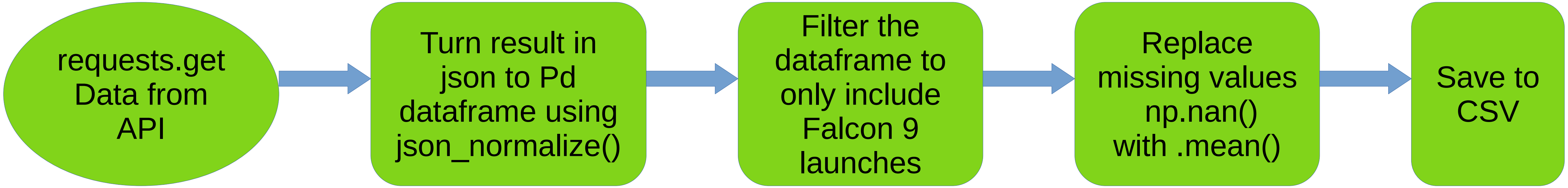
Methodology

Data Collection

There are 2 ways to collect data:

Option 1: Get data from SpaceX API using HTTP Requests:

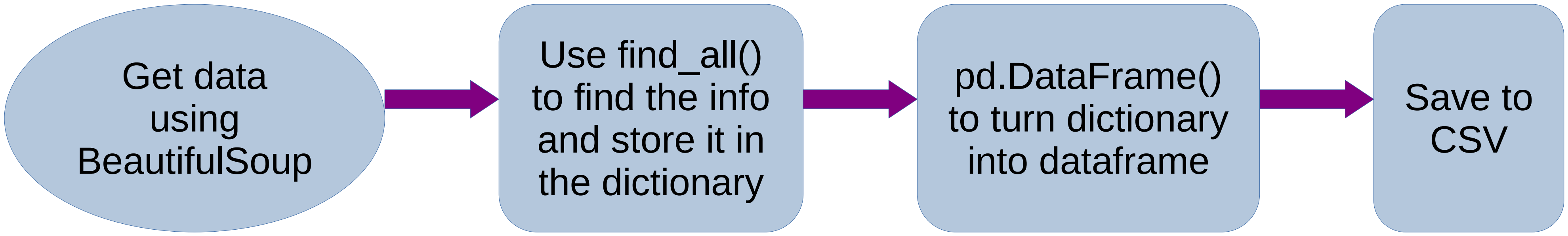
https://api.spacexdata.com/v4/launches/past



JupyterLab notebook link: click here

- **Option 2**: Get data from Wiki:

https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922



JupyterLab notebook link: click here

Data Wrangling

Load csv from data collection → value_counts() Calculate the number of launches on each site → Calculate the number and occurrence of each orbit → Calculate the number and occurrence of mission outcome per orbit type → Create a landing outcome label "class" from Outcome column

Appendix 1

1: First stage land successfully
0: Failed

True Ocean means the mission outcome was successfully landed to a specific region of the ocean while False Ocean means the mission outcome was unsuccessfully landed to a specific region of the ocean. True RTLS means the mission outcome was successfully landed to a ground pad False RTLS means the mission outcome was unsuccessfully landed to a ground pad.True ASDS means the mission outcome was successfully landed to a drone ship False ASDS means the mission outcome was unsuccessfully landed to a drone ship. None ASDS and None None these represent a failure to land.

JupyterLab notebook link: click here

EDA with SQL

- Load the dataset into the corresponding table in a Db2 database.
- Display the names of the unique launch sites in the space mission.
- Display 5 records where launch sites begin with the string 'CCA'.
- Display the total payload mass carried by boosters launched by NASA (CRS).
- Display average payload mass carried by booster version F9 v1.1.
- List the date when the first successful landing outcome in ground pad was achieved.
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000.
- List the total number of successful and failure mission outcomes.
- List the names of the booster_versions which have carried the maximum payload mass. Use a subquery.
- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015.
- 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.

JupyterLab notebook link: click here

EDA with Data Visualization

Scatter plot:

Explore the relationship between the 2 factors and their relationship with success rate, to find out what factor(s) affect the success rate,and if these 2 factors have correlation between them.

Flight Number vs. Launch Site
Payload vs. Launch Site
Flight Number vs. Orbit Type
Payload vs. Orbit Type

Bar plot:

Explore the different influences by the subgroups, to find which subgroup has the highest success rate.

- **Success Rate vs. Orbit Type**

Line plot:

Direct changes of success rate over time.

- **Launch Success Rate vs. Time**

JupyterLab notebook link: click here

SPACEX

Build an Interactive Map with Folium

1. A CSV file is imported to get the latitude and longitude for all launch sites, form a new dataframe together with column “Class”, which with 1(green) indicate success landing, 0(red) as failure.

2. Mark all 4 launch sites on a map.

3. Mark the success/failed launches for each site on the map

4. Calculate the distances between a launch site to its nearest proximities, such as coastline,railway,highway and city, and mark that on the map.

JupyterLab notebook link: click here

Build a Dashboard with Plotly Dash

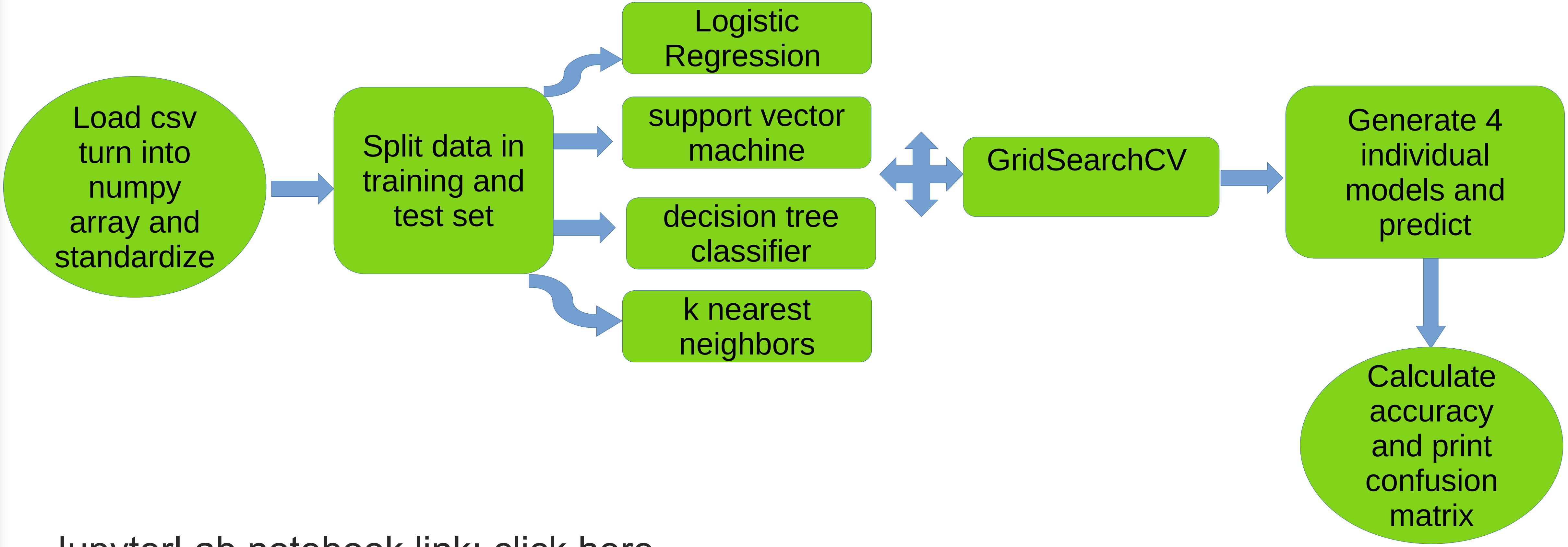
1. Build a dashboard and Pie charts, success rates of all 4 launch sites are shown interactively. It’s easier and more direct to find the launch site with highest success rate.

2. Interactive scatter plot of payload vs success rate with different booster types, it’s easier this way, to find the payload range and booster type with highest success rate.

JupyterLab notebook link: click here

Predictive Analysis (Classification)

Goal: Using the current data, find a best model to predict the success rate.



JupyterLab notebook link: click here

Results

Exploratory data analysis results

Interactive analytics demo in screenshots

Predictive analysis results

Section 2

Insights drawn from EDA

EDA with SQL

All Launch Site Names

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

- %sql SELECT UNIQUE(LAUNCH_SITE) FROM SPACEXDATA

Launch Site Names Begin with 'CCA'

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	None	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	None	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	None	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	F9 v1.0 B0006	CCAFS LC-40	None	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	None	677	LEO (ISS)	NASA (CRS)	Success	No attempt

```
%sql SELECT * FROM SPACEXDATA WHERE LAUNCH_SITE LIKE 'CCA%' limit 5
```

Total Payload Mass carried by boosters launched by NASA (CRS)

1
45596

- %sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXDATA WHERE CUSTOMER='NASA (CRS)';

Average Payload Mass by F9 v1.1

1
2534

- %sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXDATA WHERE BOOSTER_VERSION LIKE '%F9 v1.1%'

First Successful Ground Landing Date

1
2015-12-22

- %sql select min(DATE) from SPACEXDATA where landing__outcome='Success (ground pad)'

Successful Drone Ship Landing with Payload between 4000 and 6000

booster_version	payload_mass__kg_
F9 FT B1022	4696
F9 FT B1026	4600
F9 FT B1021.2	5300
F9 FT B1031.2	5200

- %sql select booster_version,payload_mass__kg_ from SPACEXDATA where landing__outcome='Success (drone ship)' and payload_mass__kg_ <6000 and payload_mass__kg_>4000

Total Number of Successful and Failure Mission Outcomes

1
100

- %sql select count(*) from SPACEXDATA where mission_outcome like '%Success%'

1
1

- %sql select count(*) from SPACEXDATA where mission_outcome like '%F%'

Boosters Carried Maximum Payload

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

- %sql select booster_version,payload_mass__kg_ from SPACEXDATA where payload_mass__kg_=(select max(payload_mass__kg_) from spacexdata)

Failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015

DATE	booster_version	launch_site
2015-01-10	F9 v1.1 B1012	CCAFS LC-40
2015-04-14	F9 v1.1 B1015	CCAFS LC-40

- %sql select DATE,booster_version,launch_site from spacexdata where landing__outcome='Failure (drone ship)' and DATE like '2015%'

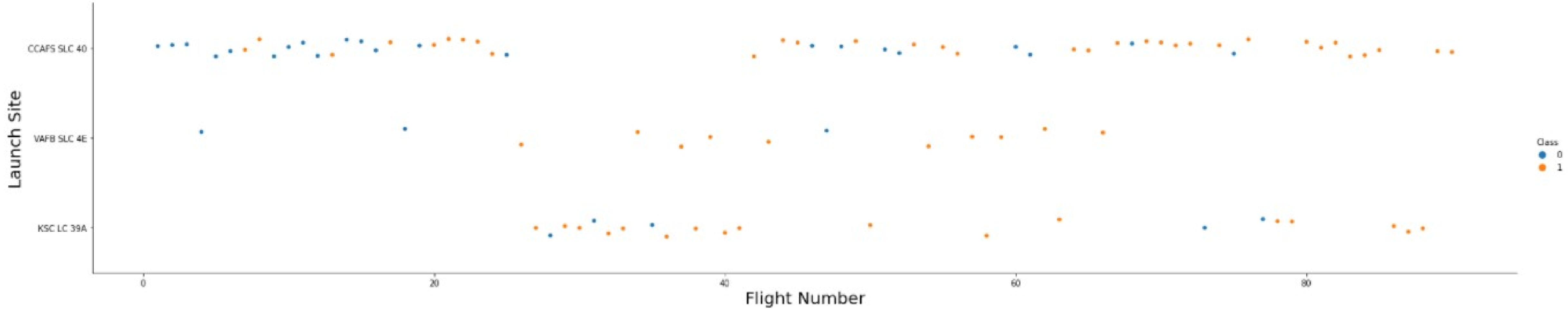
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

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

%sql select landing__outcome,count(*) as countnumber from spacexdata where DATE between '2010-06-04' and '2017-03-20' group by landing__outcome order by countnumber desc

EDA with data visualization

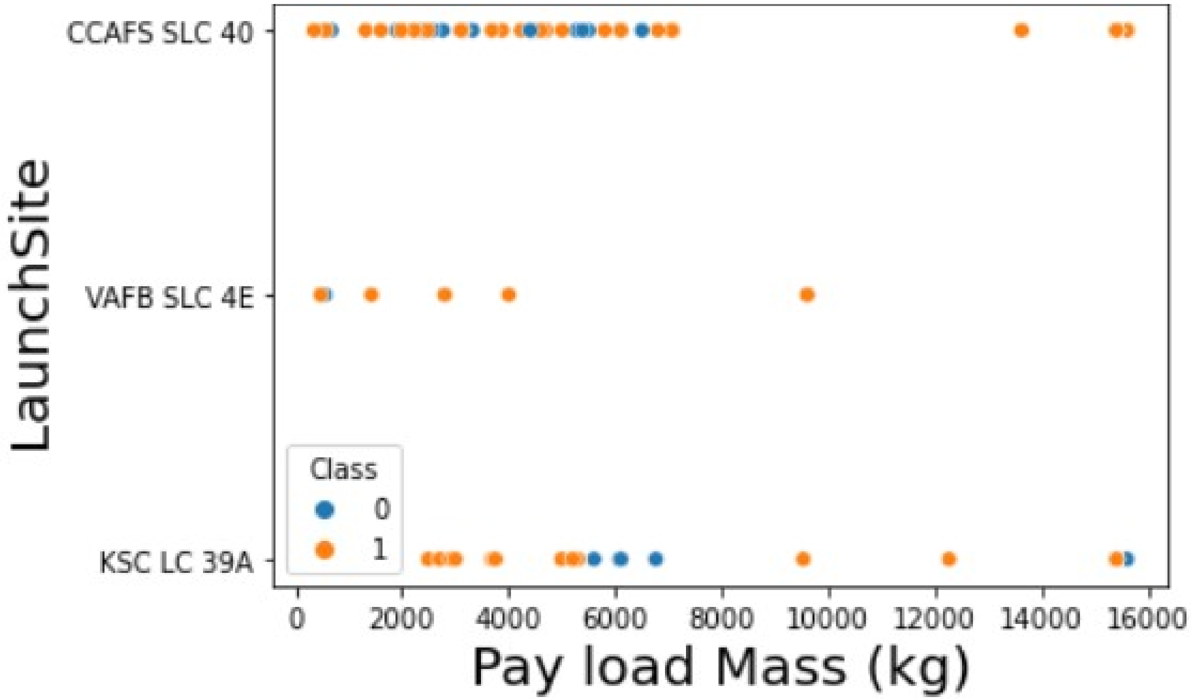
Flight Number vs. Launch Site



CCAFS SLC-40 has much more flights than the other two.

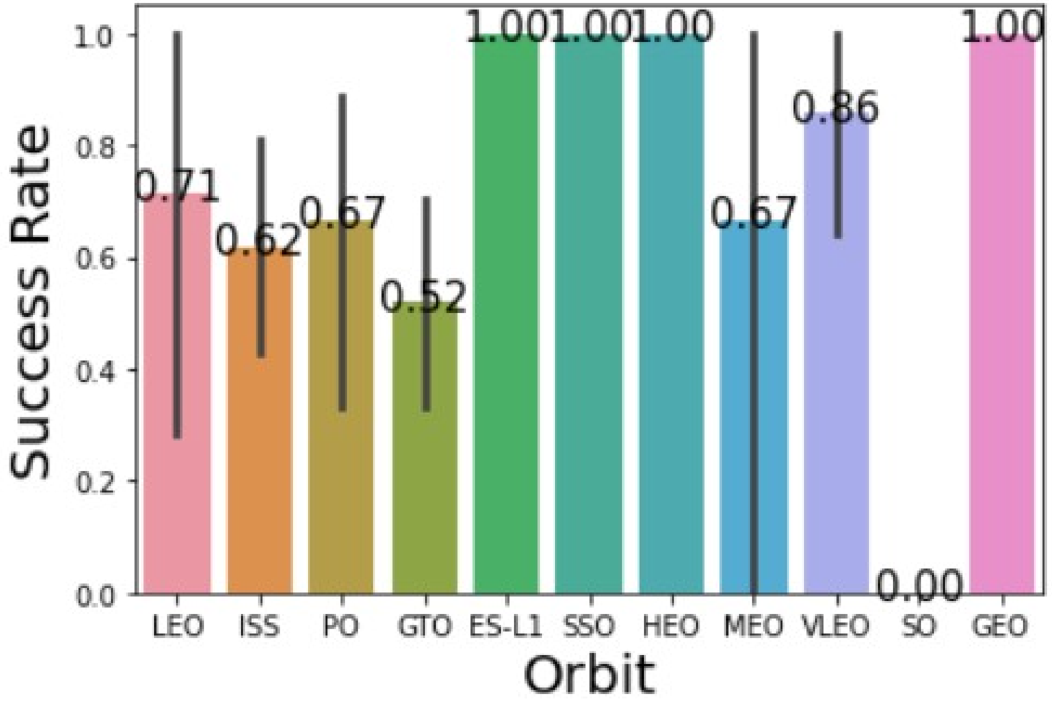
- The more flights for each launch site, the better successful rate.

Payload vs. Launch Site



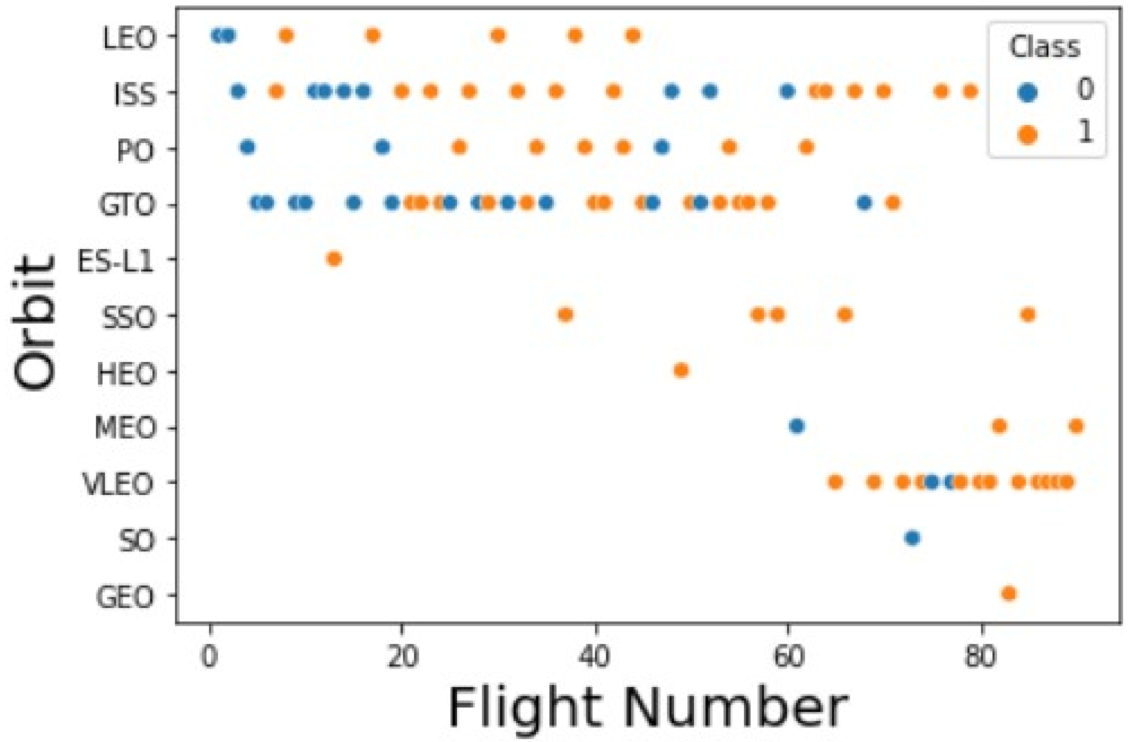
- Every payload for WAFB SLC 4E is less than 1,0000 kg, CCAFS SLC 40 has payloads concentrated below 8000 kg, KSC LC 39A has payloads distributed more evenly.
- It seems the heavier the payload, the higher success rate, but there's no enough data here to make the conclusion.

Success Rate vs. Orbit Type



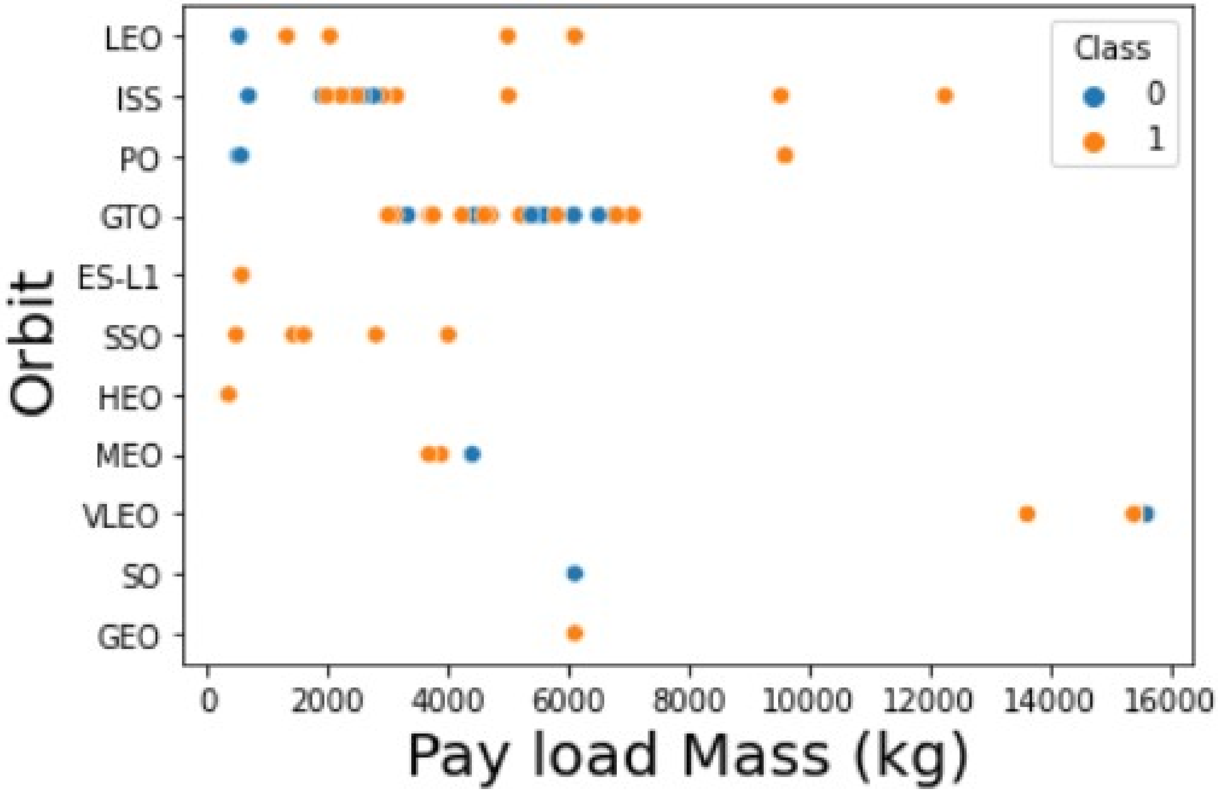
- ES-L1,SSO,HEO and GEO have 100% success rate, all of them have unique features.
- The higher the orbit, the lower success rate holds true between VLEO,LEO and GTO,but not generally, HEO and GEO are high attitude orbits.

Flight Number vs. Orbit Type



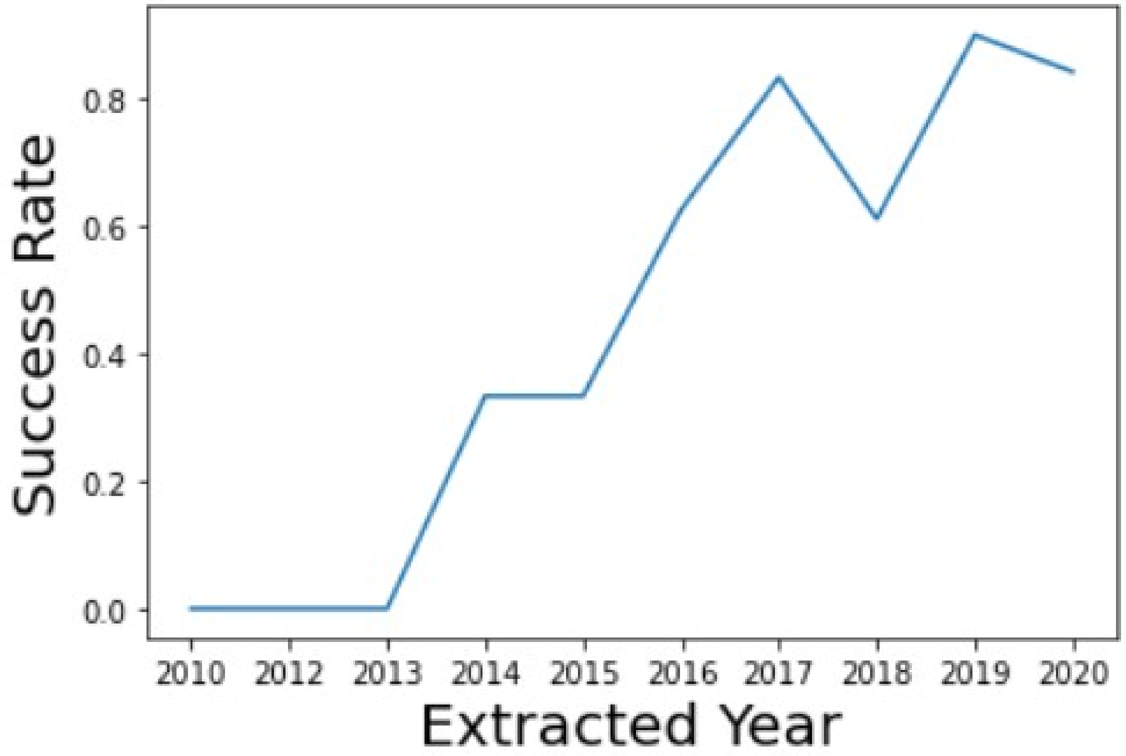
- The more flight number, the higher success rate except GTO.
- ISS,PO and GTO have more flights, doesn’t have very clear pattern.
- Although ES-L1,SSO,HEO and GEO have 100% success rate, they have few flights.

Payload vs. Orbit Type



- The heavier the payload, the higher success rate,except for GTO or VLEO.
- ES-L1,SSO,HEO and GEO have 100% success rate, but their payload is less than 6,200 kg.
- GTO has concentrated payloads between 2,200~6,200 kg.

Launch Success Yearly Trend



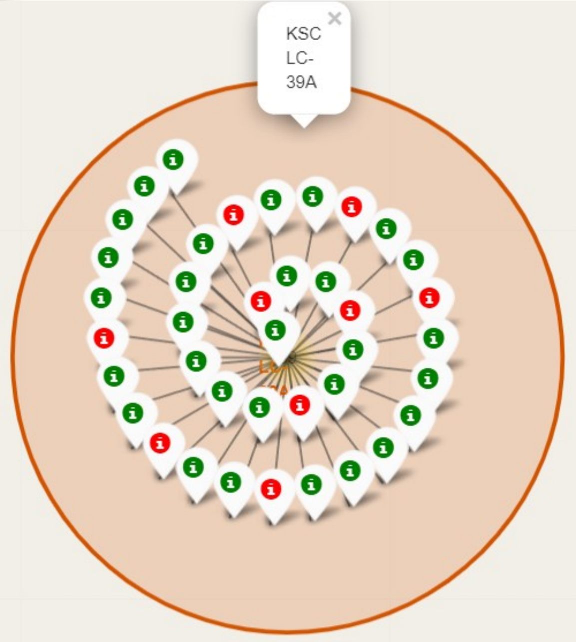
- From 2010~2013, the success rate is terrible.
- From 2013, it started climbing up, dipped around 2018.

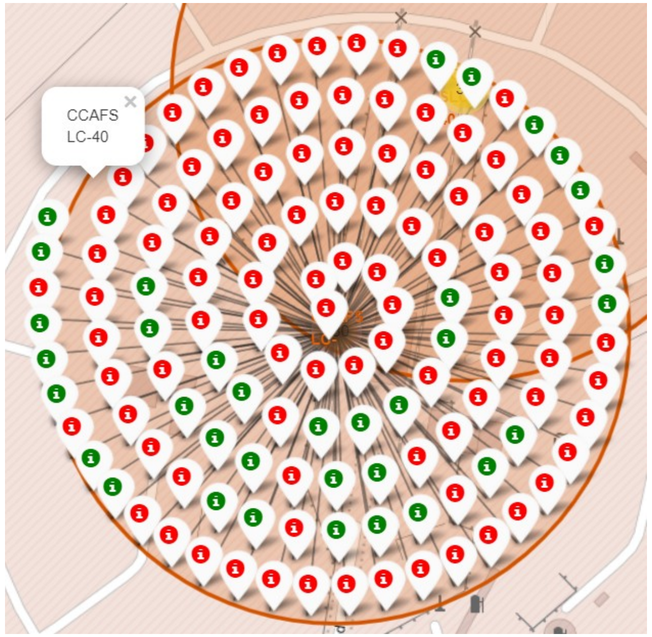
Section 4

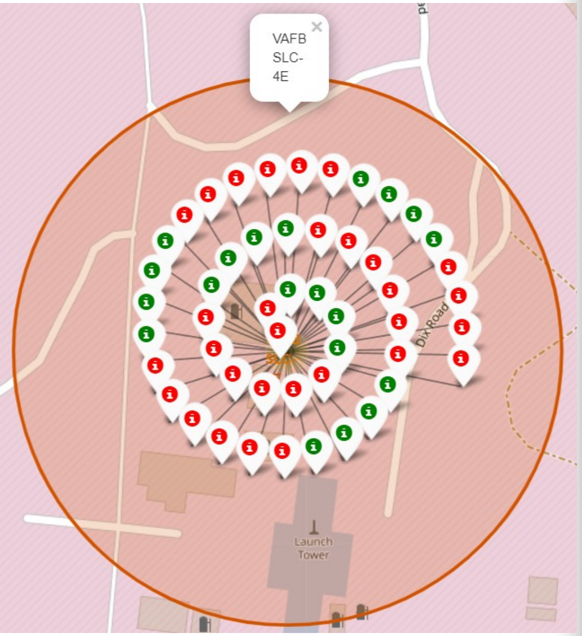
Launch Sites Proximities Analysis

<4 launch sites on global map>

<Launch sites with success ratio>

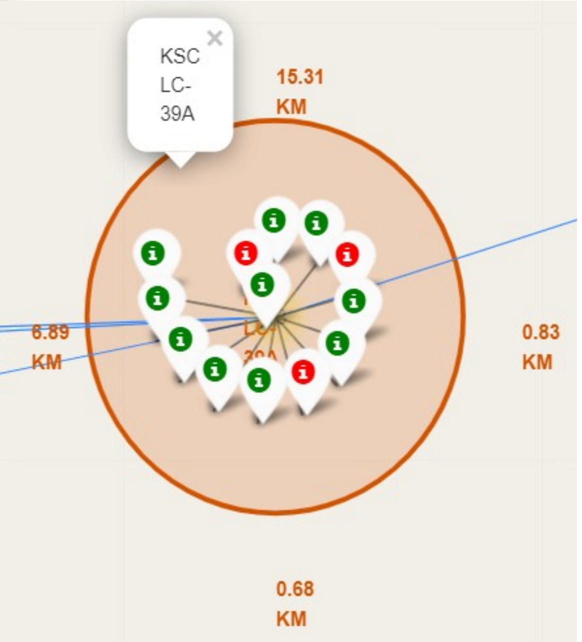


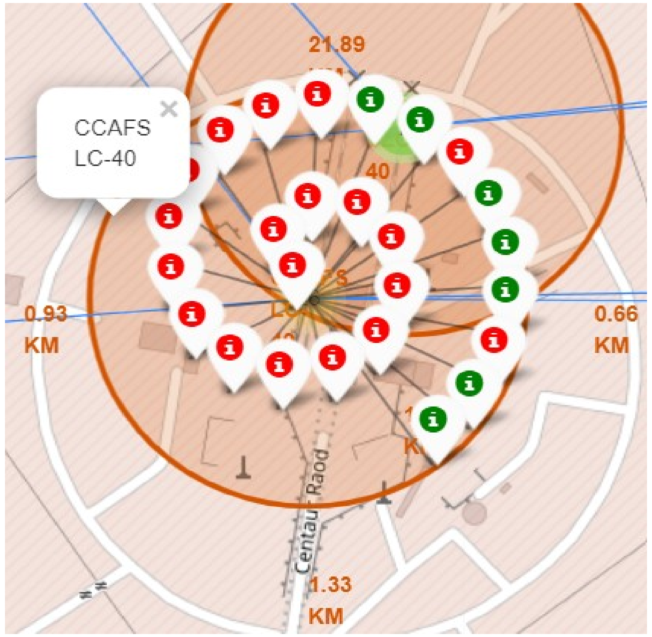


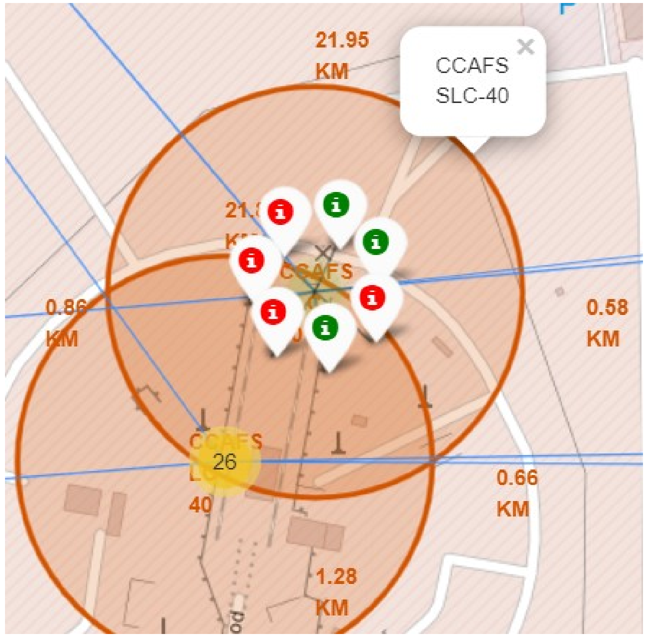


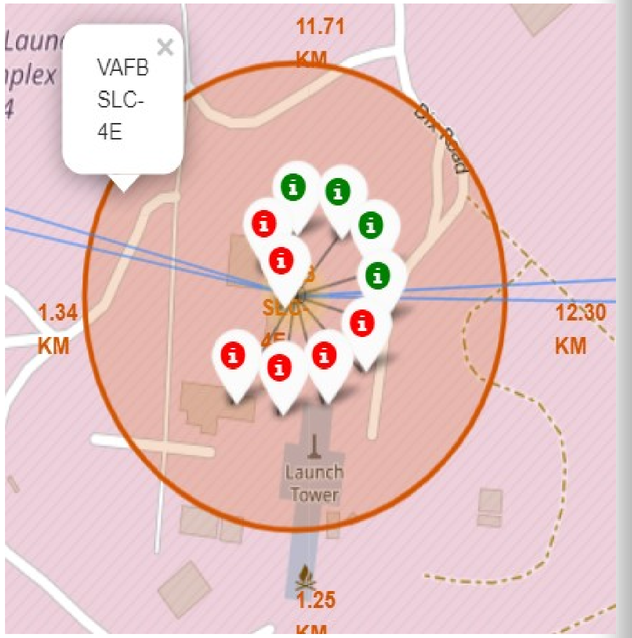
- KSC LC-39A has the highest success ratio.

<Launch outcomes w/distance to coastline,railway,highway and city>

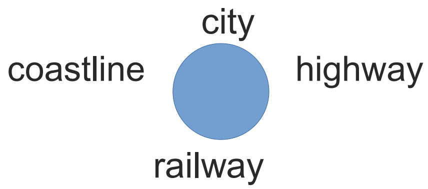








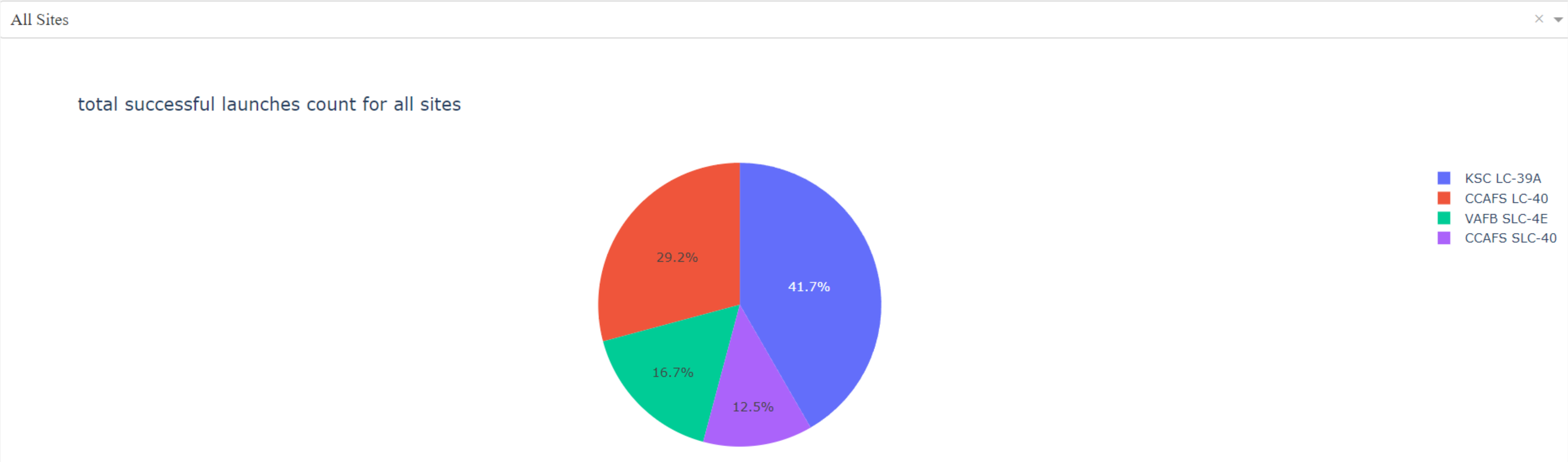
- It seems maintaining certain distance to coastline helps success rate.



Section 5

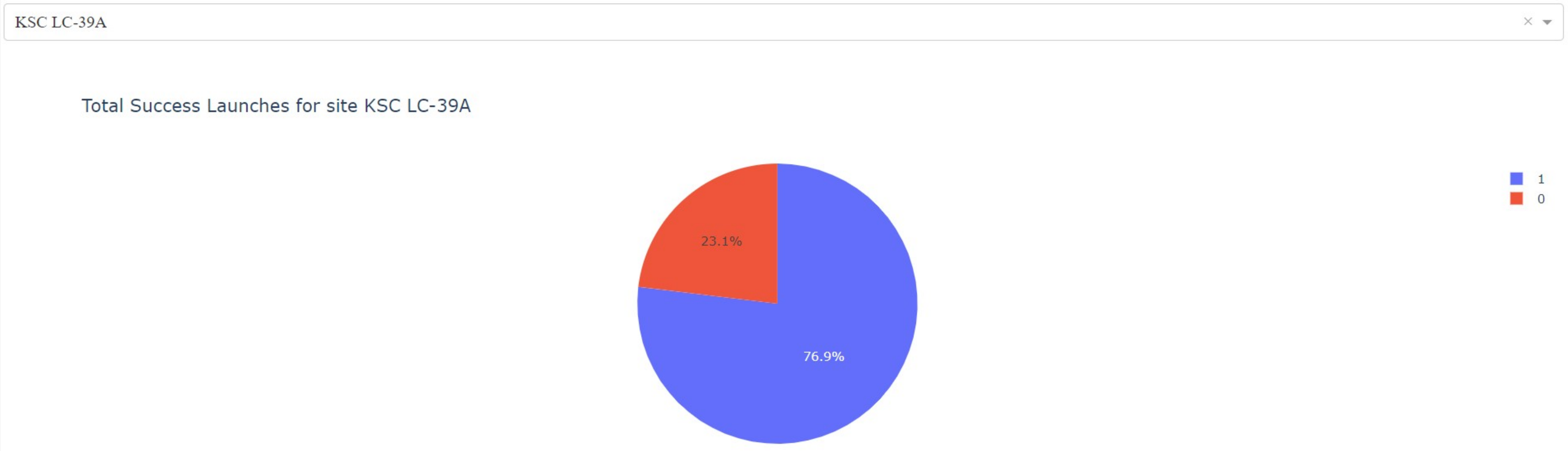
Build a Dashboard with Plotly Dash

<launch success count for all sites>

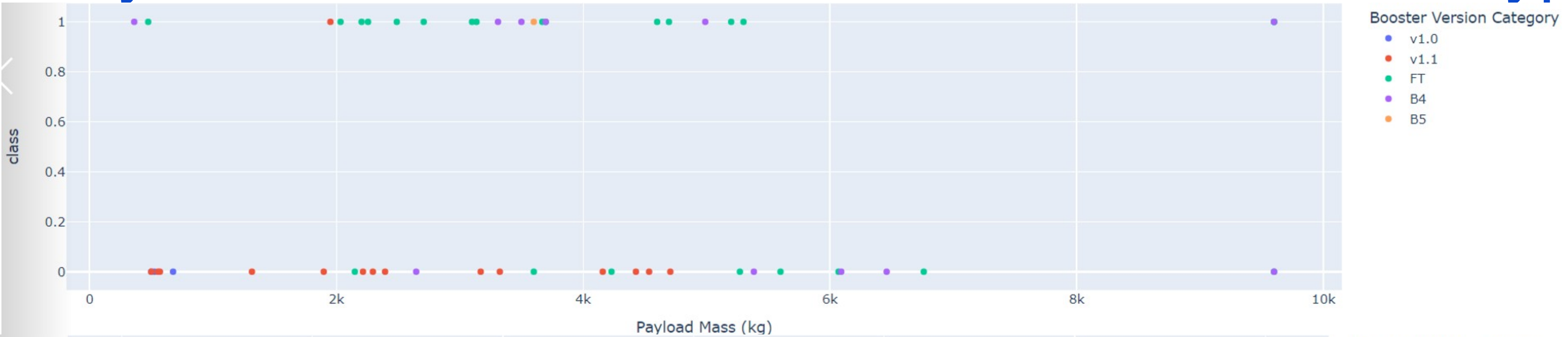


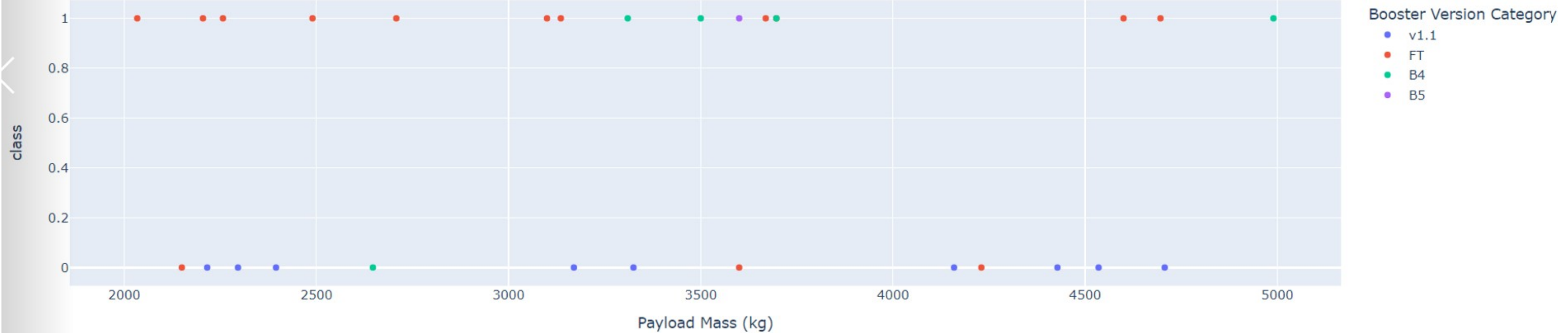
- KSC LC-39A has the highest success ratio.

<Launch sites with highest success ratio>



<Payload vs Success rate with different booster type>



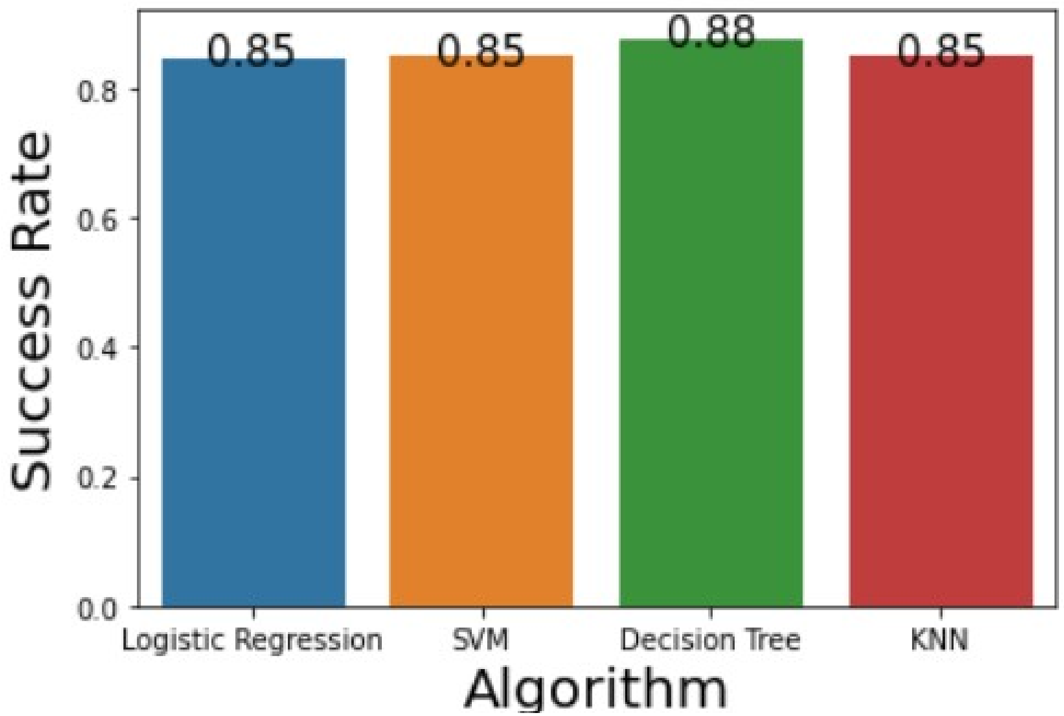


- Payload 1,900~5,300 kg has higher success rate, Booster FT has higher success rate.

Section 6

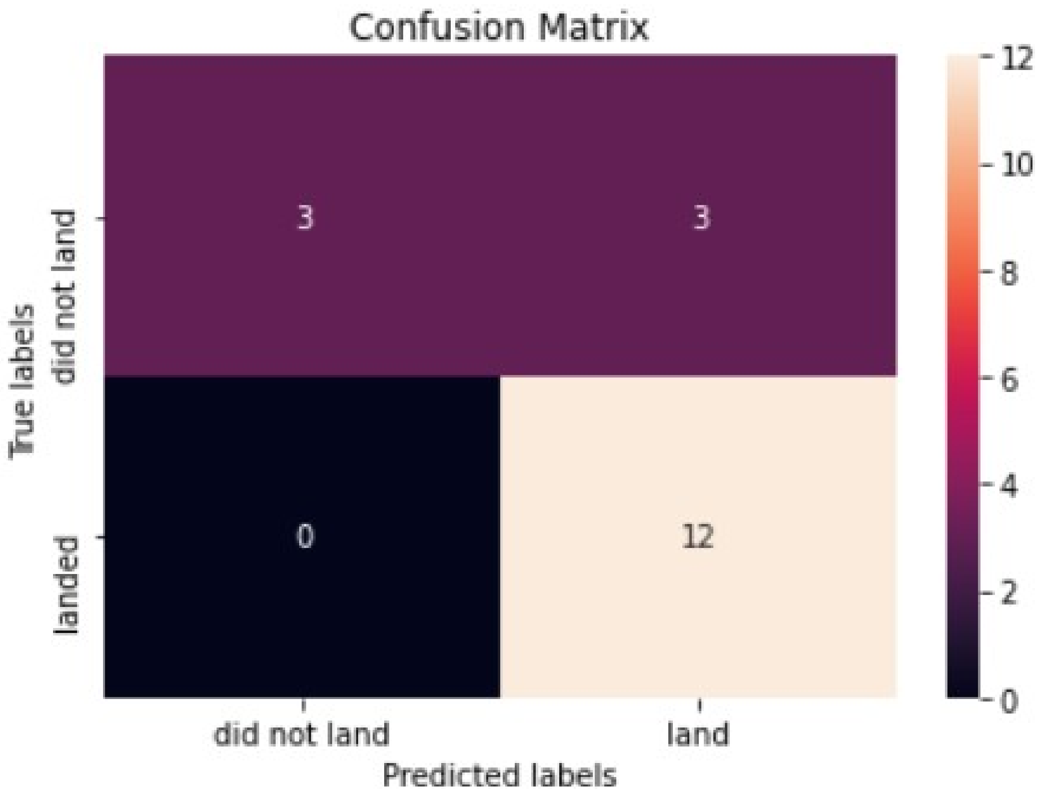
Predictive Analysis (Classification)

Classification Accuracy



- Decision tree method has the highest accuracy.

Confusion Matrix



- Inaccuracy comes from false positive, in which it's predicted successfully landed but actually didn't.

Conclusions

- ES-L1,SSO,HEO and GEO have 100% success rate, all of them have unique physical features. They have less than average flight numbers though.
- The more flight number, the higher success rate except GTO.
- The most successful payloads are from 1,900 kg to 5,300 kg.
- From 2013, the success rate started climbing up, dipped around 2018.
- KSC LC-39A has the highest success ratio.
- Decision tree model predicts the success rate the best.
- It seems maintaining certain distance to coastline helps success rate.

Appendix 1: Common orbit types:

- LEO: Low Earth orbit (LEO)is an Earth-centred orbit with an altitude of 2,000 km (1,200 mi) or less (approximately one-third of the radius of Earth),[1] or with at least 11.25 periods per day (an orbital period of 128 minutes or less) and an eccentricity less than 0.25.[2] Most of the manmade objects in outer space are in LEO [1].
- VLEO: Very Low Earth Orbits (VLEO) can be defined as the orbits with a mean altitude below 450 km. Operating in these orbits can provide a number of benefits to Earth observation spacecraft as the spacecraft operates closer to the observation[2].
- GTO A geosynchronous orbit is a high Earth orbit that allows satellites to match Earth's rotation. Located at 22,236 miles (35,786 kilometers) above Earth's equator, this position is a valuable spot for monitoring weather, communications and surveillance. Because the satellite orbits at the same speed that the Earth is turning, the satellite seems to stay in place over a single longitude, though it may drift north to south," NASA wrote on its Earth Observatory website [3] .
- SSO (or SO): It is a Sun-synchronous orbit also called a heliosynchronous orbit is a nearly polar orbit around a planet, in which the satellite passes over any given point of the planet's surface at the same local mean solar time [4] .
- ES-L1 :At the Lagrange points the gravitational forces of the two large bodies cancel out in such a way that a small object placed in orbit there is in equilibrium relative to the center of mass of the large bodies. L1 is one such point between the sun and the earth [5] .

- HEO A highly elliptical orbit, is an elliptic orbit with high eccentricity, usually referring to one around Earth [6].

- ISS A modular space station (habitable artificial satellite) in low Earth orbit. It is a multinational collaborative project between five participating space agencies: NASA (United States), Roscosmos (Russia), JAXA (Japan), ESA (Europe), and CSA (Canada) [7]

- MEO Geocentric orbits ranging in altitude from 2,000 km (1,200 mi) to just below geosynchronous orbit at 35,786 kilometers (22,236 mi). Also known as an intermediate circular orbit. These are "most commonly at 20,200 kilometers (12,600 mi), or 20,650 kilometers (12,830 mi), with an orbital period of 12 hours [8]

- HEO Geocentric orbits above the altitude of geosynchronous orbit (35,786 km or 22,236 mi) [9]

- GEO It is a circular geosynchronous orbit 35,786 kilometres (22,236 miles) above Earth's equator and following the direction of Earth's rotation [10]

- PO It is one type of satellites in which a satellite passes above or nearly above both poles of the body being orbited (usually a planet such as the Earth [11]

Thank you!