IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Serge FOKWA
August 08, 2022

Outline

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

Executive Summary

- Our goal in this project is to use this data to predict whether SpaceX will attempt to land a rocket or not.
- In terms of methodology, we collected data from SPACEX API and also by Web Scraping. Then data passed through some basic wrangling where they were explored, cleaned and formatted using Pandas library. At the end of this process we obtained a Pandas dataframe ready for analysis. We perform exploratory data analysis (EDA) by using Pandas, Numpy, Matpltlib and Seaborn libraries to visualize relationship between variables. We also used SQL for supplementary analysis. Then we performed interactive visual analytics using Folium and Plotly Dash to enable users to explore and manipulate data. We performed predictive analysis using logistic regression, support vector machine, decision tree classifier and k nearest neighbors.
- In terms of results, we found that the success rate since 2013 kept increasing till 2020. Launch sites are in close proximity to coastline, railway and highway, and away from cities. For each site, success appears related to the number of flights. ES-L1, GEO, HEO, SSO orbits have the highest success rate. Predictive analysis showed that Decision tree model has the highest classification accuracy (88.9%).

Introduction

- 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.
- In this project, the problem is to predict if the Falcon 9 first stage will land successfully. So, our goal is to use the collected data from 2010 to 2020, to analyse them and to predict whether SpaceX will attempt to land a rocket or not.
- In the first section we will present the methodology. In the second section some insights drawn from EDA. In the third section Launch sites proximities analyses. In the fourth section we will build a dashboard. In the fifth section we will perform the predictive analysis.

Section 1

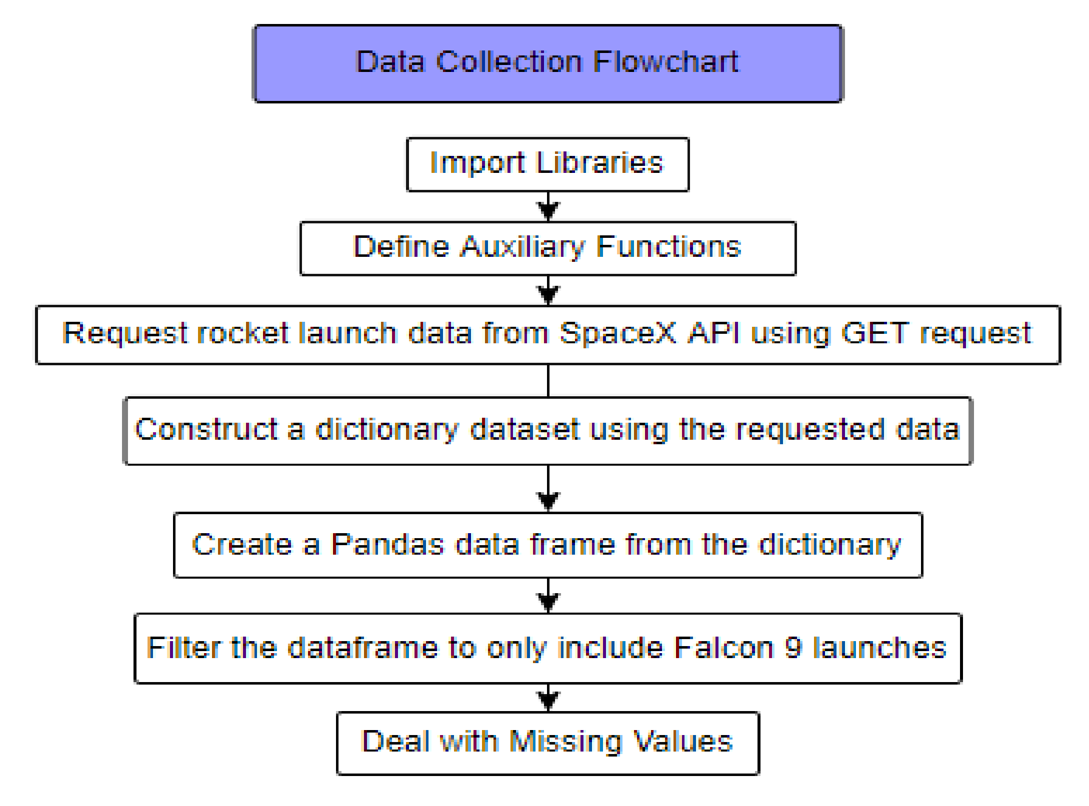
Methodology

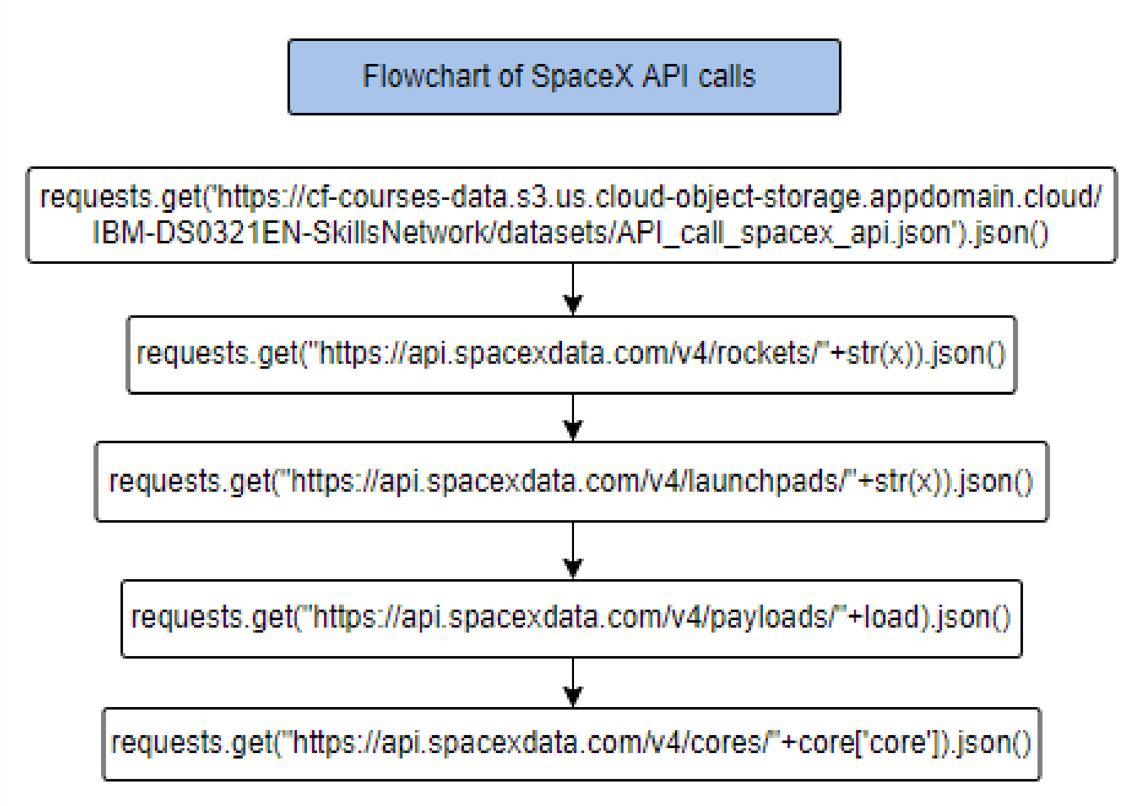
Methodology

Executive Summary

- Data collection methodology:
    - Describe how data was collected
- Perform data wrangling
    - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
    - How to build, tune, evaluate classification models

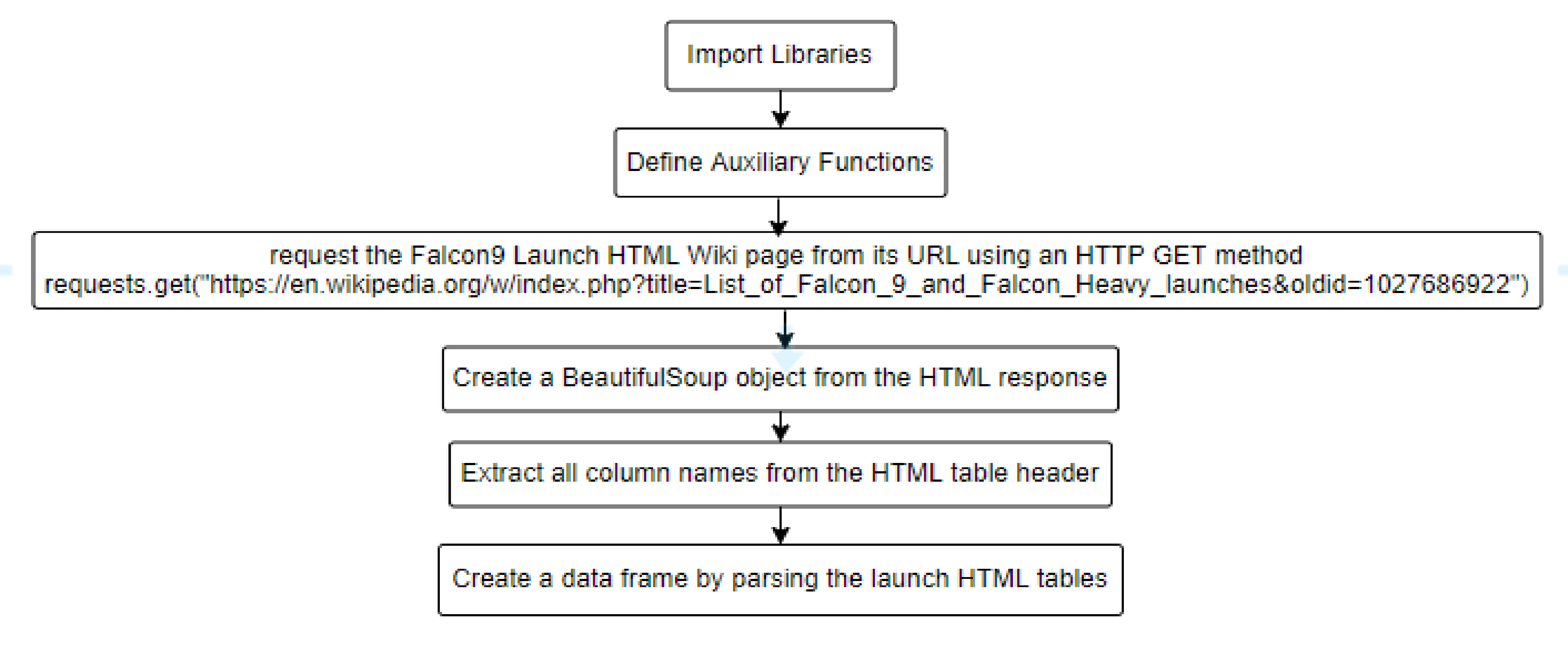
Data Collection – SpaceX API





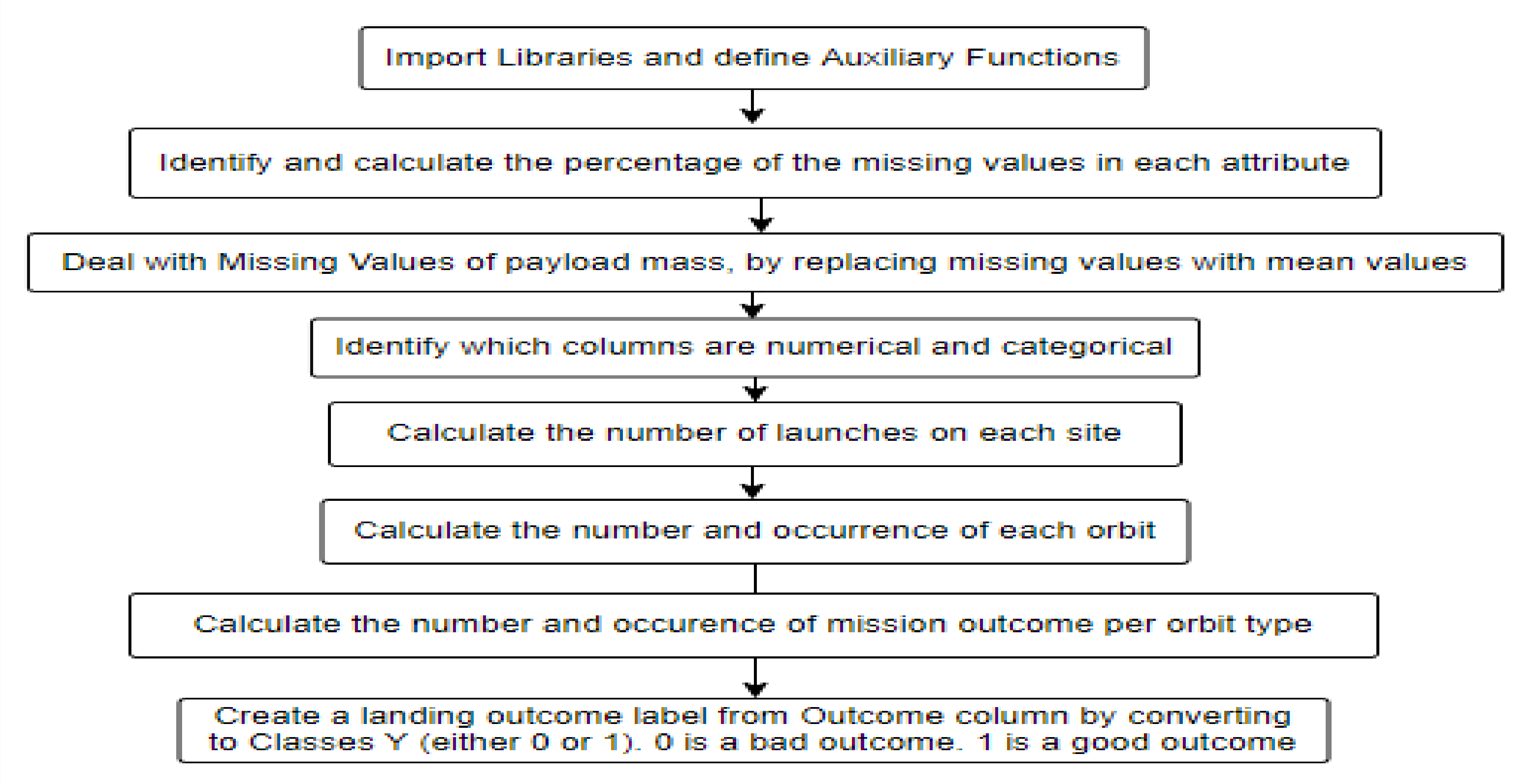
The required data for the project are: Booster Version, Payload Mass, Orbit, Launch Site, Landing outcome, Flight N°, GridFins, Reused, Legs, LandingPad, Block, ReusedCount, Serial, Longitude, Latitude, Date, Time, etc, from 2010 to 2020

Data Collection - Scraping



The required data for the project are: Booster Version, Payload Mass, Orbit, Launch Site, Landing outcome, Flight N°, GridFins, Reused, Legs, LandingPad, Block, ReusedCount, Serial, Longitude, Latitude, Date, Time, etc, from 2010 to 2020

Data Wrangling



The required data for the project are: Booster Version, Payload Mass, Orbit, Launch Site, Landing outcome, Flight N°, GridFins, Reused, Legs, LandingPad, Block, ReusedCount, Serial, Longitude, Latitude, Date, Time, etc, from 2010 to 2020

EDA with Data Visualization

We plotted out the following charts:

- FlightNumber vs. PayloadMass, to see how the Flight Number and Payload variables would affect the launch outcome.
- FlightNumber vs LaunchSite, to see how the Flight Number and Launch Site variables would affect the launch outcome.
- PayloadMass vs LaunchSite, to see how the Payload and LaunchSite variables would affect the launch outcome.
- Success rate vs Orbit type, to check if there are any relationship between success rate and orbit type.
- FlightNumber vs Orbit type, to see if there is any relationship between Flight Number and Orbit type.
- Payload vs Orbit type, to see if there is any relationship between Payload and Orbit type.
- Average success rate vs Year, to get the average launch success trend.

EDA with SQL

Using SQL queries, we displayed the following features:

- The names of the unique launch sites in the space mission
- The records where launch sites begin with the string 'CCA‘
- The total payload mass carried by boosters launched by NASA (CRS)
- The average payload mass carried by booster version F9 v1.1
- The date when the first succesful landing outcome in ground pad was acheived.
- The names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- The total number of successful and failure mission outcomes
- The names of the booster_versions which have carried the maximum payload mass.
- The month, names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.
- The count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

Build an Interactive Map with Folium

Using Folium we created and added the following map objects:

- A folium map object with an initial center location to be NASA Johnson Space Center at Houston, Texas;
- A circle area with a text label on NASA Johnson Space Center;
- For each launch site, a circle object, a marker object, and Launch site name as a popup label; So we can clearly visualize and identify the different locations.
- A marker cluster to the site map;
- For each site on the map, a Marker object to mark the success/failed launches that indicate if this launch was successful or failed. With these markers we can see which sites have high success rates.
- A distance polyline object between launch site and closest coastline, city, railway, highway; in order to analyse its proximities.

All those objects were added in order to find an optimal location for building a launch site.

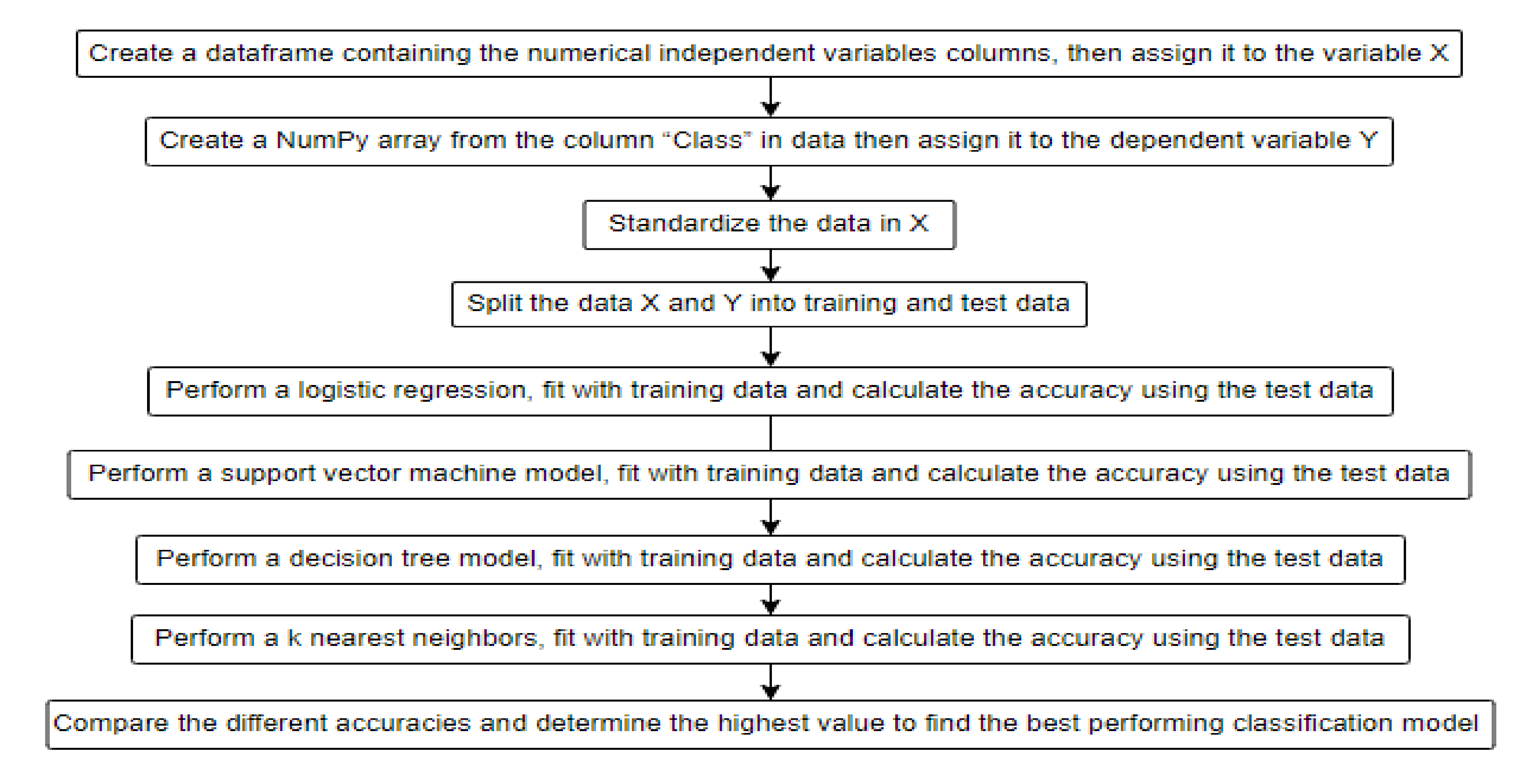
Build a Dashboard with Plotly Dash

Using Plotly Dash, we added the following to the dashboard:

- A Launch Site Drop-down Input Component
- A callback function to render success-pie-chart based on selected site dropdown
- A Range Slider to Select Payload
- A callback function to render the success-payload-scatter-chart scatter plot

All those plots and interactions were added for users to perform interactive visual analytics on SpaceX launch data in real-time and to be able to obtain some insights.

Predictive Analysis (Classification)



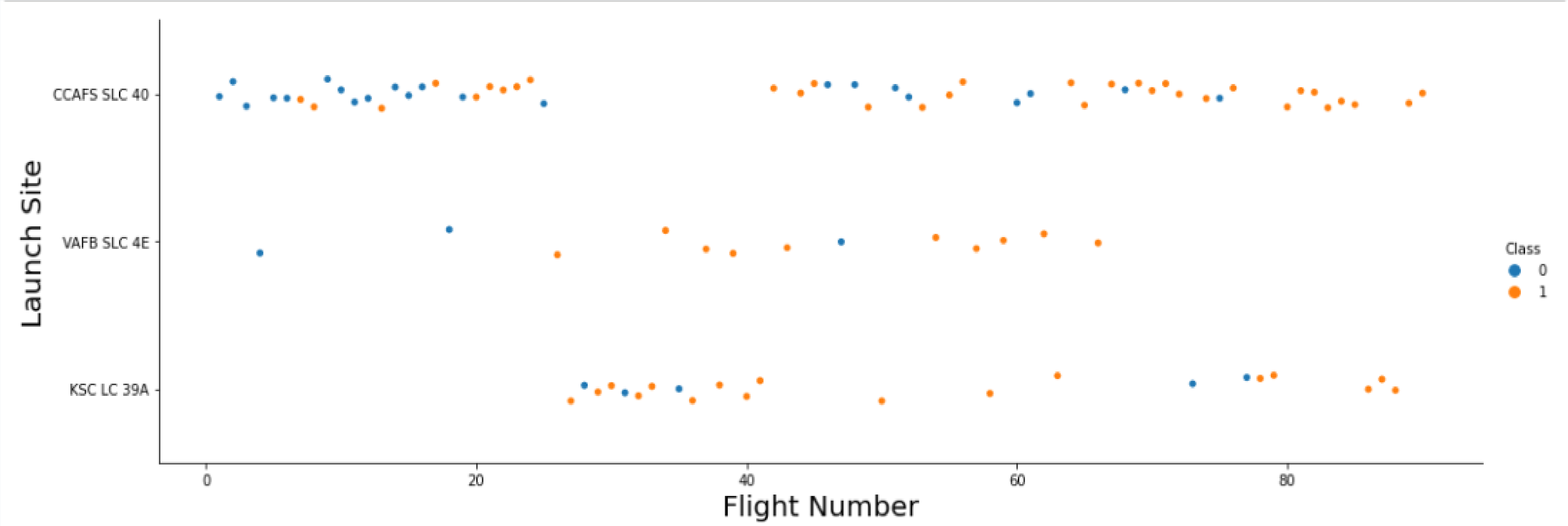
Results

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

Section 2

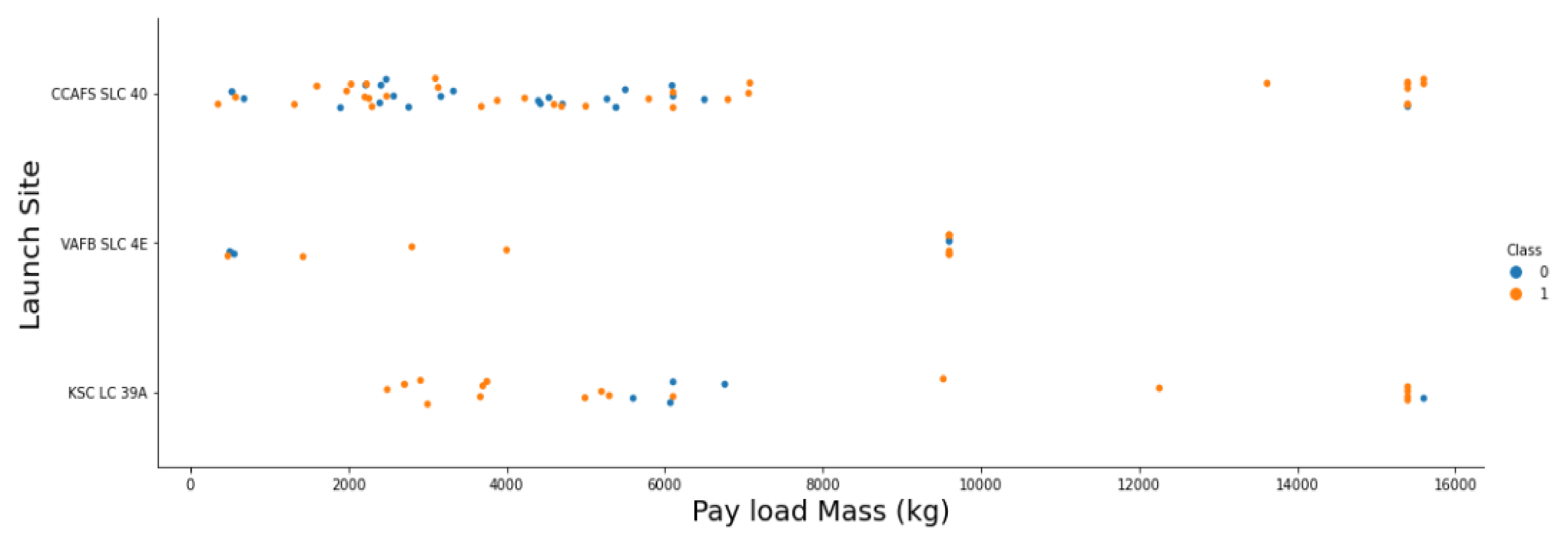
Insights drawn from EDA

Flight Number vs. Launch Site



- Launch Sites KSC LC-39A and VAFB SLC 4E have higher success rate than CCAFS LC-40.
- For each site, success also appears related to the number of flights.

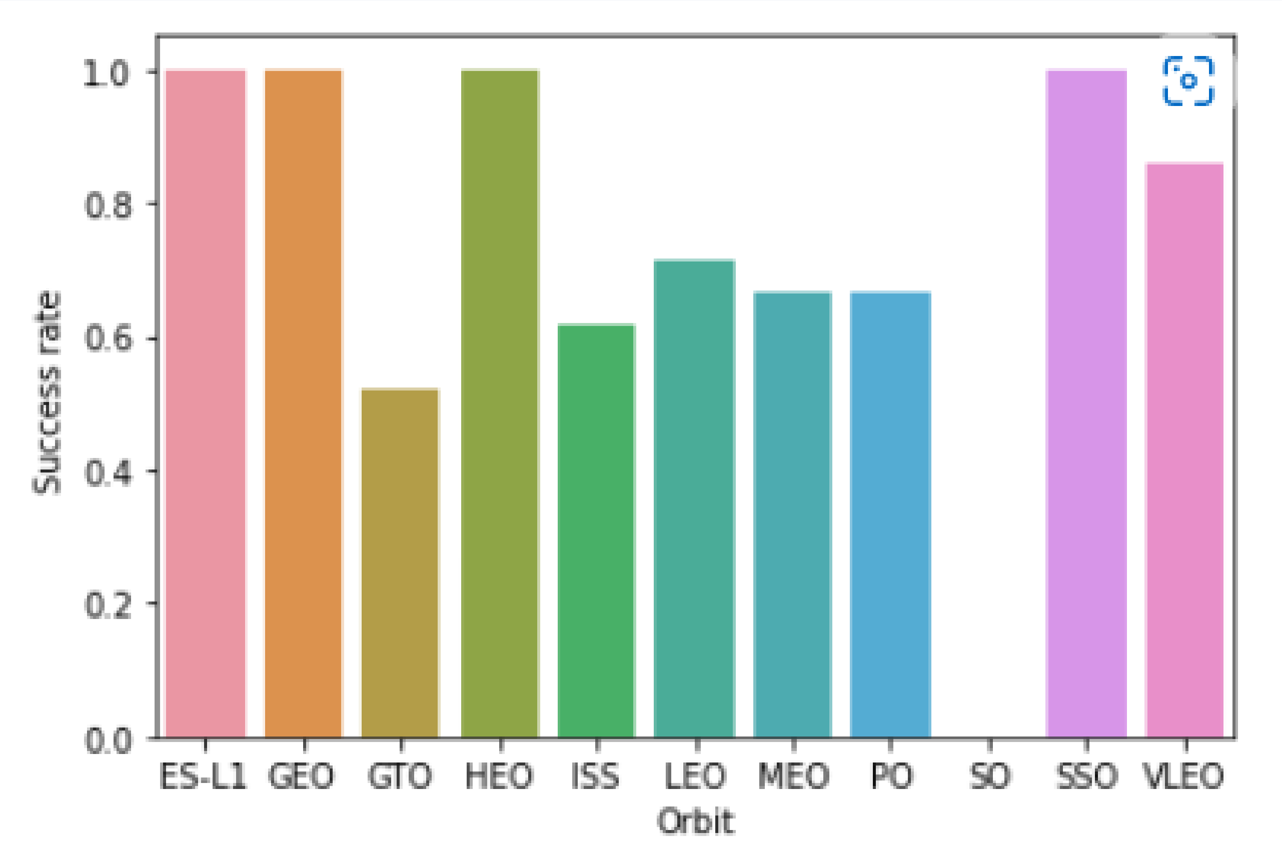
Payload vs. Launch Site



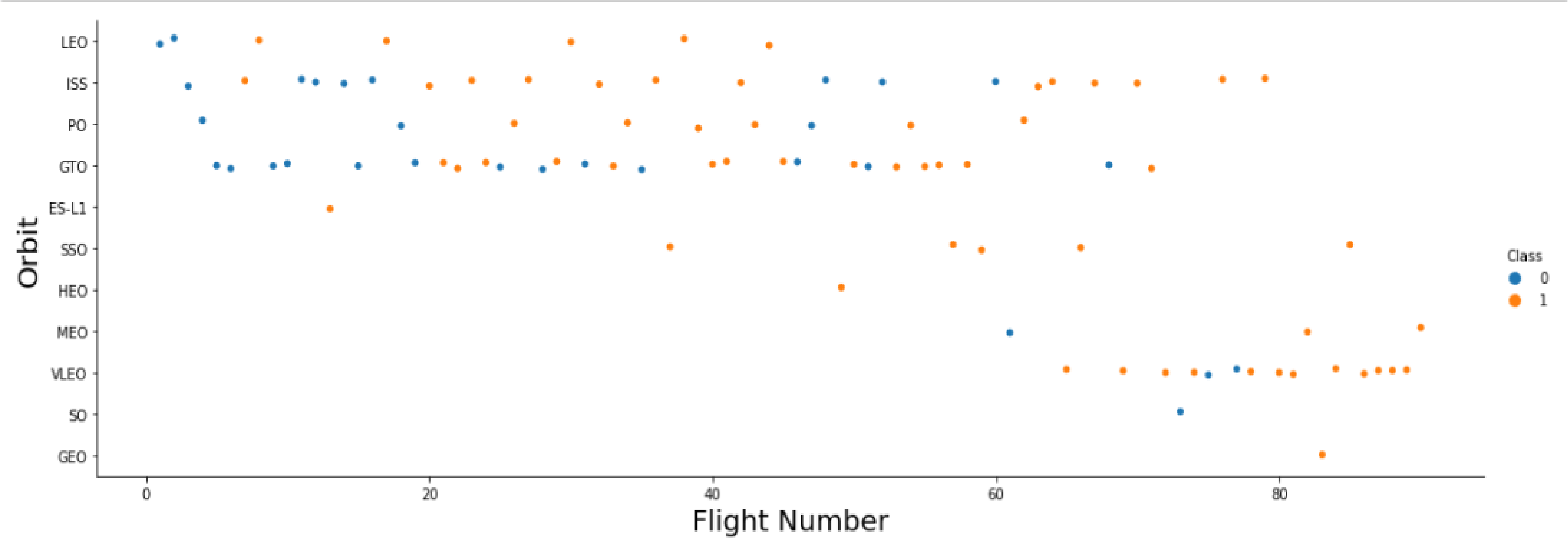
- For the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).
- For the CCAFS LC-40 launchsite has a poor success rate for small payload mass (less than 6000)
- For small payload mass (less than 6000), the KSC LC-39A launchsite has the highest success rate.

Success Rate vs. Orbit Type

- ES-L1, GEO, HEO, SSO and VLEO orbits have high success rate

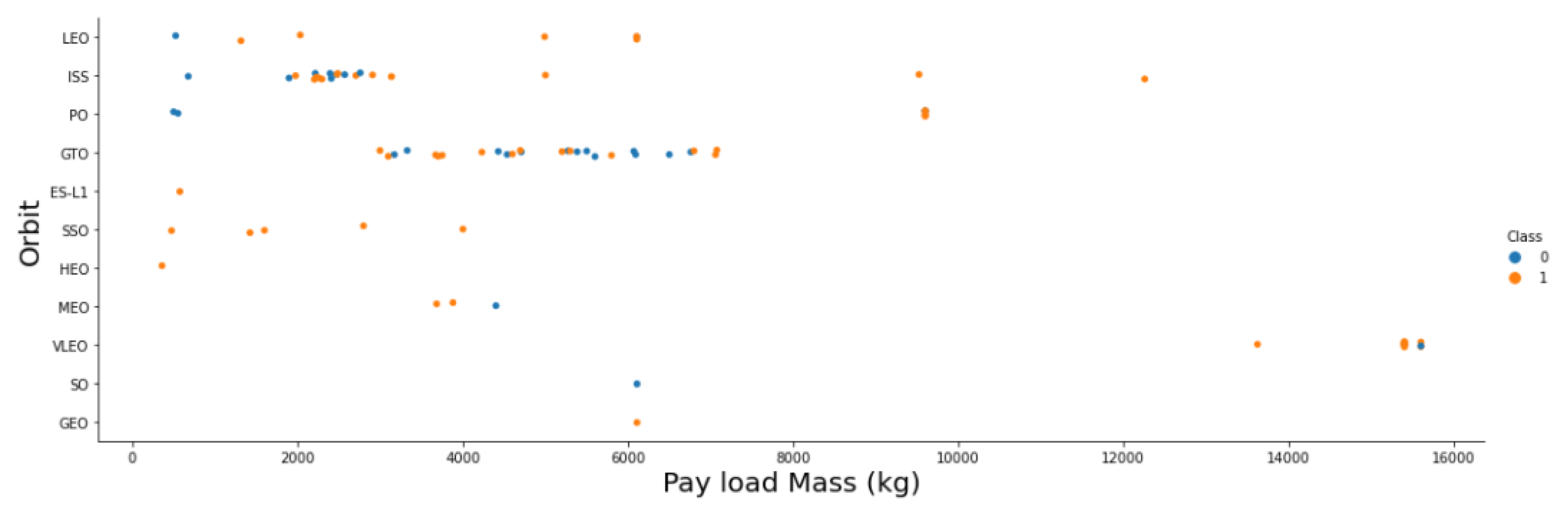


Flight Number vs. Orbit Type



- In the LEO orbit the Success appears related to the number of flights;
- In GTO orbit there seems to be no relationship with flight number.

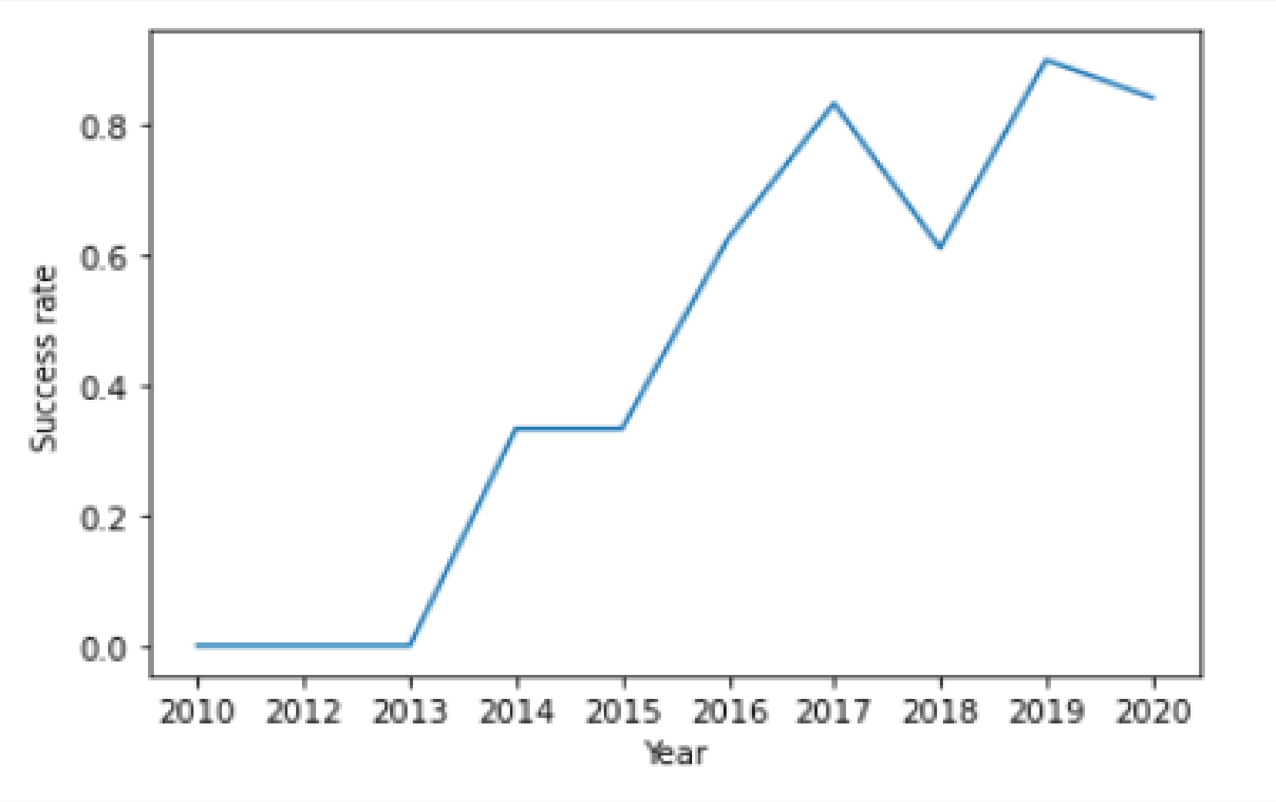
Payload vs. Orbit Type



- With heavy payloads the successful landing are higher for PO,LEO and ISS orbits.
- However for GTO we there seem to be no relationship between payloads and success rate.

Launch Success Yearly Trend

- The sucess rate since 2013 kept increasing till 2020



All Launch Site Names

```
%sql select distinct Launch_Site from SPACEXTBL;
```

There are four different launch sites in the space mission:

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

Launch Site Names Begin with 'CCA'

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

- Here we have 5 records where launch sites names begin with `CCA`
- %sql select * from SPACEXTBL where Launch Site like 'CCA%' limit 5;

Total Payload Mass

sum(PAYLOAD_MASS__KG_)
45596

- This is the total payload mass carried by boosters launched by NASA (CRS).
- %sql select sum(PAYLOAD_MASS__KG_) from SPACEXTBL where Customer = 'NASA (CRS)';

Average Payload Mass by F9 v1.1

avg(PAYLOAD_MASS__KG_)
2928.4

- This is the average payload mass carried by booster version F9 v1.1;
- %sql select avg(PAYLOAD_MASS__KG_) from SPACEXTBL where Booster_Version = 'F9 v1.1';

First Successful Ground Landing Date

- The date when the first succesful landing outcome in ground pad was acheived.
- %sql select min(Date) from SPACEXTBL where "Landing_Outcome" = "Success";

Successful Drone Ship Landing with Payload between 4000 and 6000

- There are 24 boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

Booster_Version	PAYLOAD_MASS__KG_	Mission_Outcome	"Success (drone ship)"
F9 v1.1	4535	Success	Success (drone ship)
F9 v1.1 B1011	4428	Success	Success (drone ship)
F9 v1.1 B1014	4159	Success	Success (drone ship)
F9 v1.1 B1016	4707	Success	Success (drone ship)
F9 FT B1020	5271	Success	Success (drone ship)
F9 FT B1022	4696	Success	Success (drone ship)
F9 FT B1026	4600	Success	Success (drone ship)
F9 FT B1030	5600	Success	Success (drone ship)
F9 FT B1021.2	5300	Success	Success (drone ship)
F9 FT B1032.1	5300	Success	Success (drone ship)
F9 B4 B1040.1	4990	Success	Success (drone ship)
F9 FT B1031.2	5200	Success	Success (drone ship)
F9 B4 B1043.1	5000	Success (payload status unclear)	Success (drone ship)
F9 FT B1032.2	4230	Success	Success (drone ship)
F9 B4 B1040.2	5384	Success	Success (drone ship)
F9 B5 B1046.2	5800	Success	Success (drone ship)
F9 B5 B1047.2	5300	Success	Success (drone ship)
F9 B5 B1046.3	4000	Success	Success (drone ship)
F9 B5B1054	4400	Success	Success (drone ship)
F9 B5 B1048.3	4850	Success	Success (drone ship)
F9 B5 B1051.2	4200	Success	Success (drone ship)
F9 B5B1060.1	4311	Success	Success (drone ship)
F9 B5 B1058.2	5500	Success	Success (drone ship)
F9 B5B1062.1	4311	Success	Success (drone ship)

Total Number of Successful and Failure Mission Outcomes

- %sql select count(*) as Number_of_mission from SPACEXTBL where (Mission_Outcome like "Success%") or (Mission_Outcome like "Failure%");

Number_of_mission
101

Boosters Carried Maximum Payload

- The names of the booster which have carried the maximum payload mass
- %sql select Booster_Version,PAYLOAD_MASS__KG_ from SPACEXTBL where PAYLOAD_MASS__KG_=(select max(PAYLOAD_MASS__KG_) from SPACEXTBL);

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 are 07 failed landing_outcomes in drone ship, their booster versions, and launch site names for year 2015

Month	Year	Booster_Version	Launch_Site	"Failure (drone ship)"
01	2015	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
02	2015	F9 v1.1 B1013	CCAFS LC-40	Failure (drone ship)
03	2015	F9 v1.1 B1014	CCAFS LC-40	Failure (drone ship)
04	2015	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)
04	2015	F9 v1.1 B1016	CCAFS LC-40	Failure (drone ship)
06	2015	F9 v1.1 B1018	CCAFS LC-40	Failure (drone ship)
12	2015	F9 FT B1019	CCAFS LC-40	Failure (drone ship)

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

- 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

- Present your query result with a short explanation here

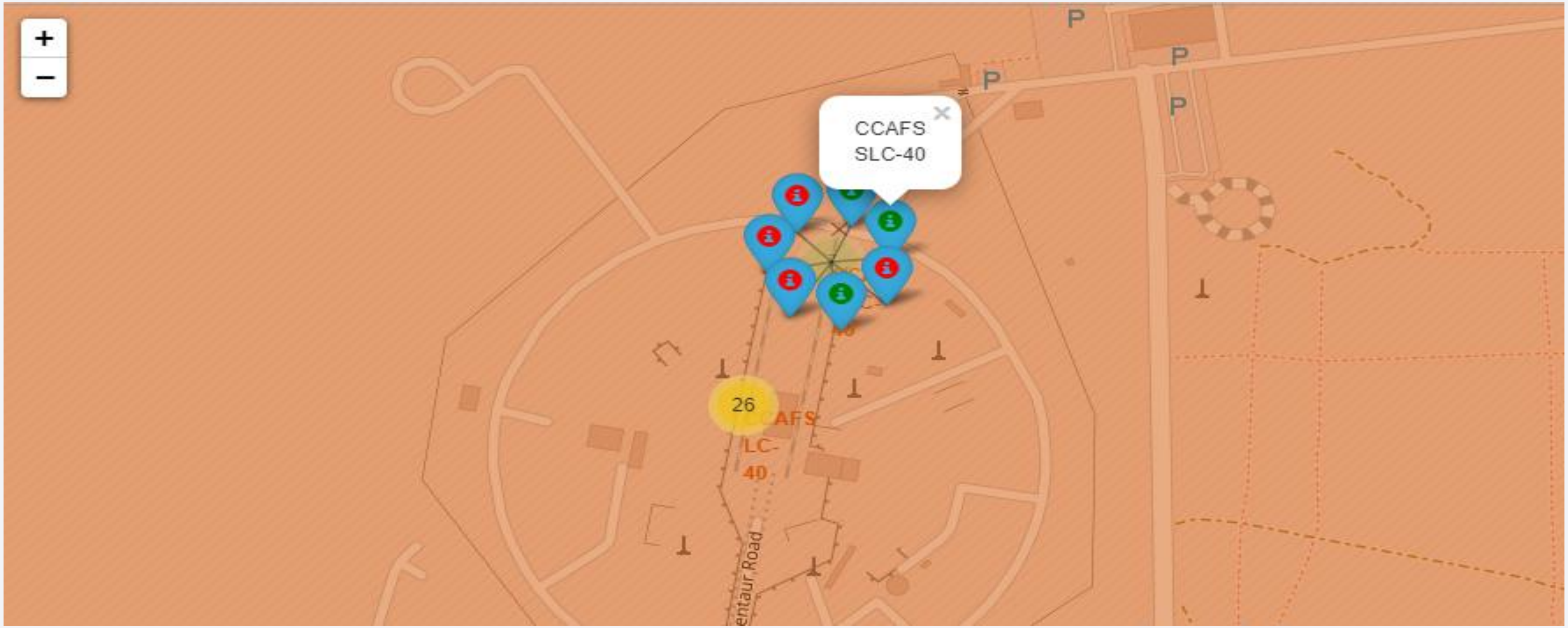
Section 3

Launch Sites Proximities Analysis

Launch sites’ location markers

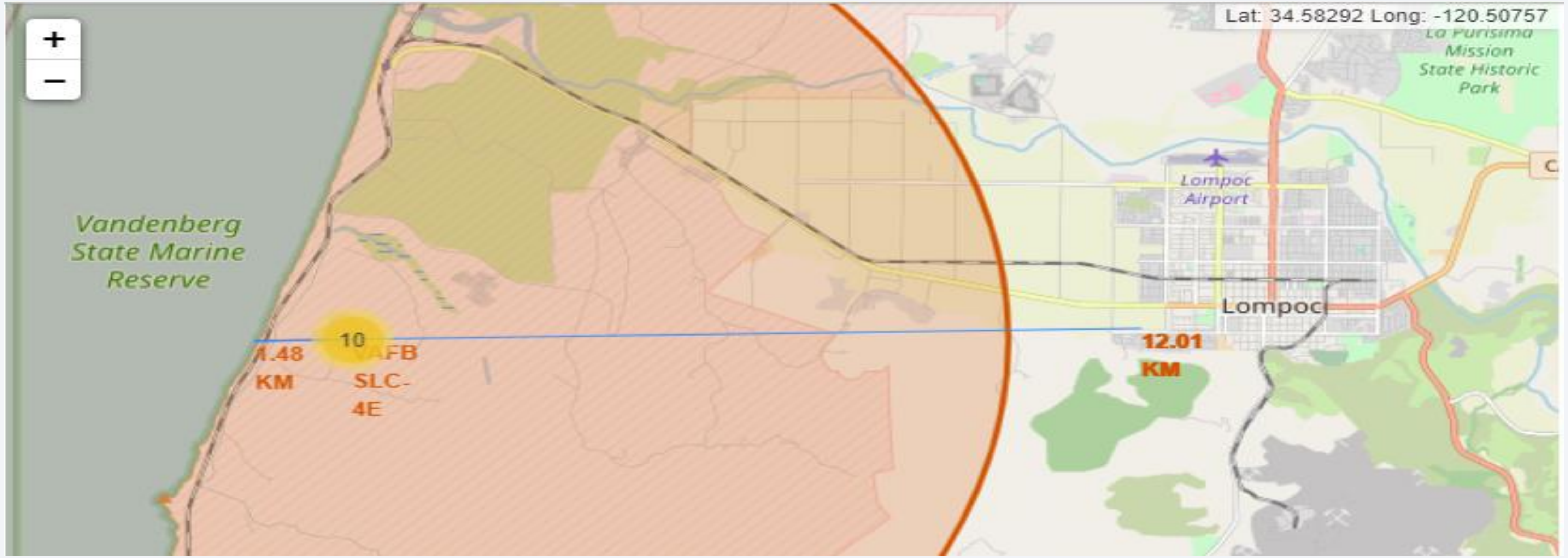
- All launch sites are in very close proximity to the coast

Launch outcomes on the map



- From the color-labeled markers in marker clusters, it is easy to identify which launch sites have relatively high success rates (green color).

Launch site proximities

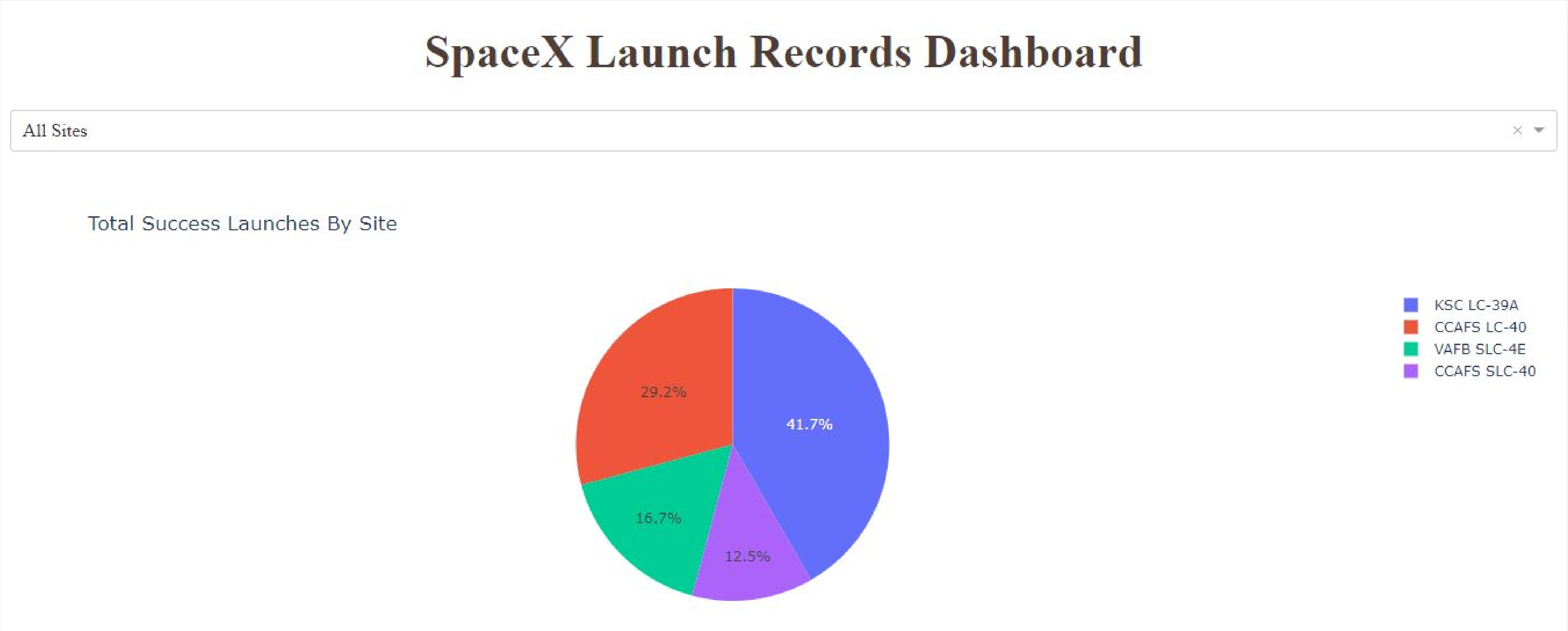


- VAFB SLC-4E launch site is in close proximity to coastline, railway and highway (1.48KM)
- VAFB SLC-4E launch site keep certain distance away from cities (12.01KM)

Section 4

Build a Dashboard with Plotly Dash

launch success count for all sites



- KSC LC-39A launch site has the highest success rate

<Dashboard Screenshot 2>

- Replace <Dashboard screenshot 2> title with an appropriate title

- Show the screenshot of the piechart for the launch site with highest launch success ratio

- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 3>

- Replace <Dashboard screenshot 3> title with an appropriate title

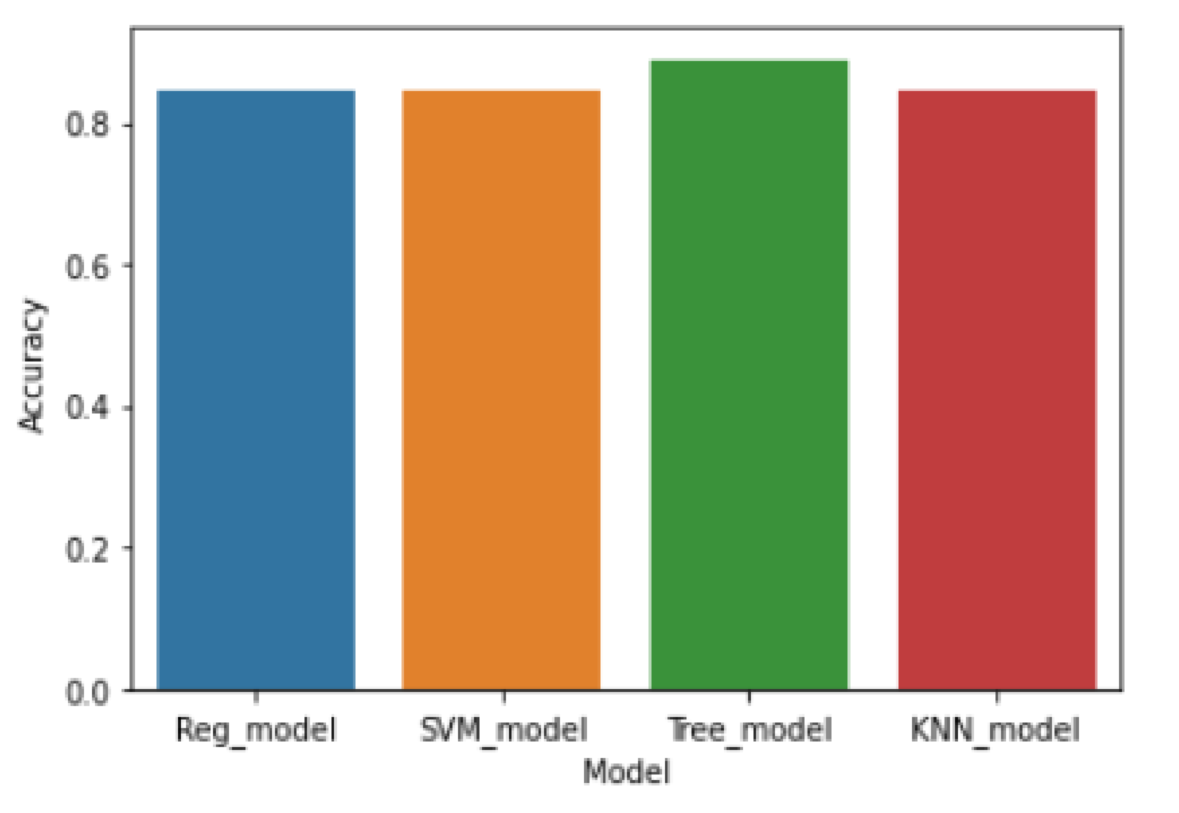
- Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider

- Explain the important elements and findings on the screenshot, such as which payload range or booster version have the largest success rate, etc.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

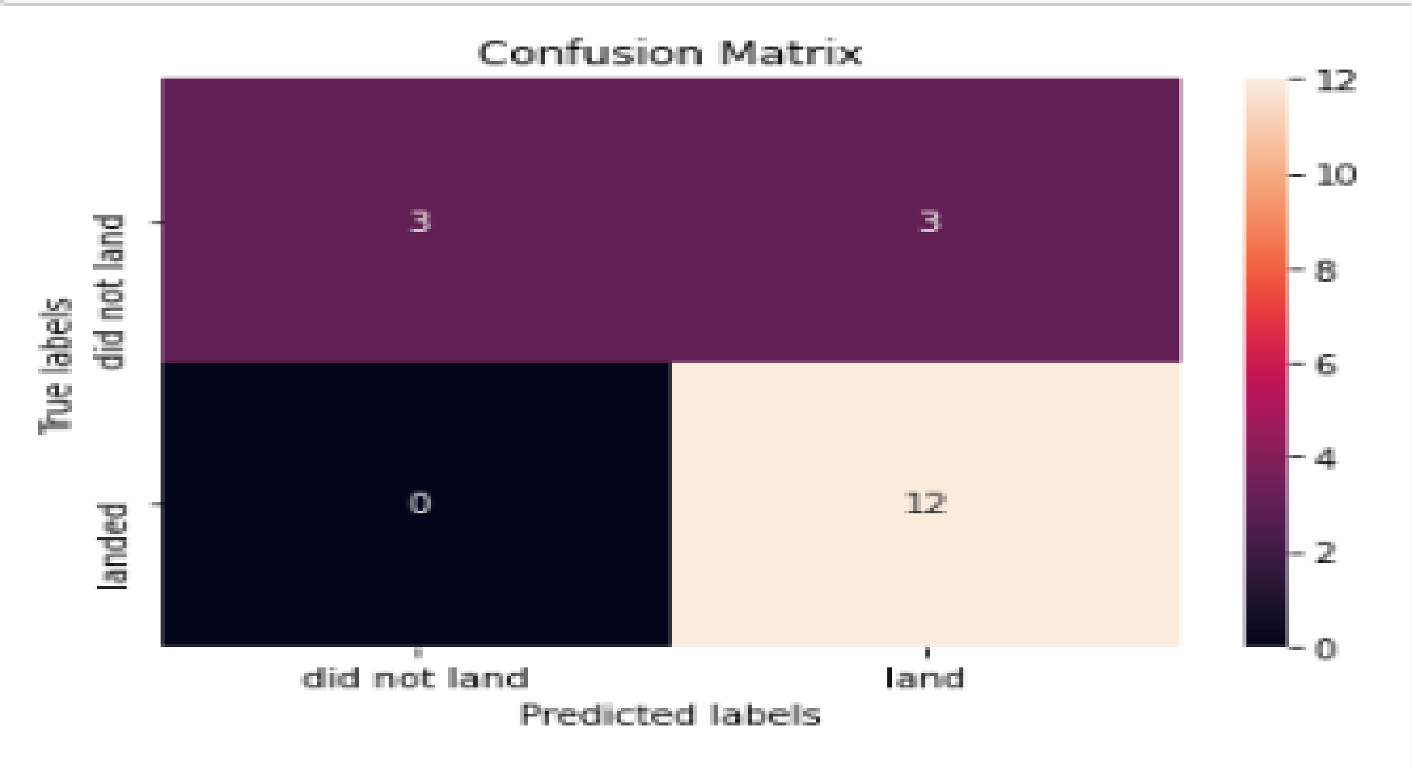


	Model	Accuracy
0	Reg_model	0.847222
1	SVM_model	0.847222
2	Tree_model	0.888889
3	KNN_model	0.847222

- Decision tree model has the highest classification accuracy

Confusion Matrix

- The confusion matrix of the best performing model: Decision tree classifier



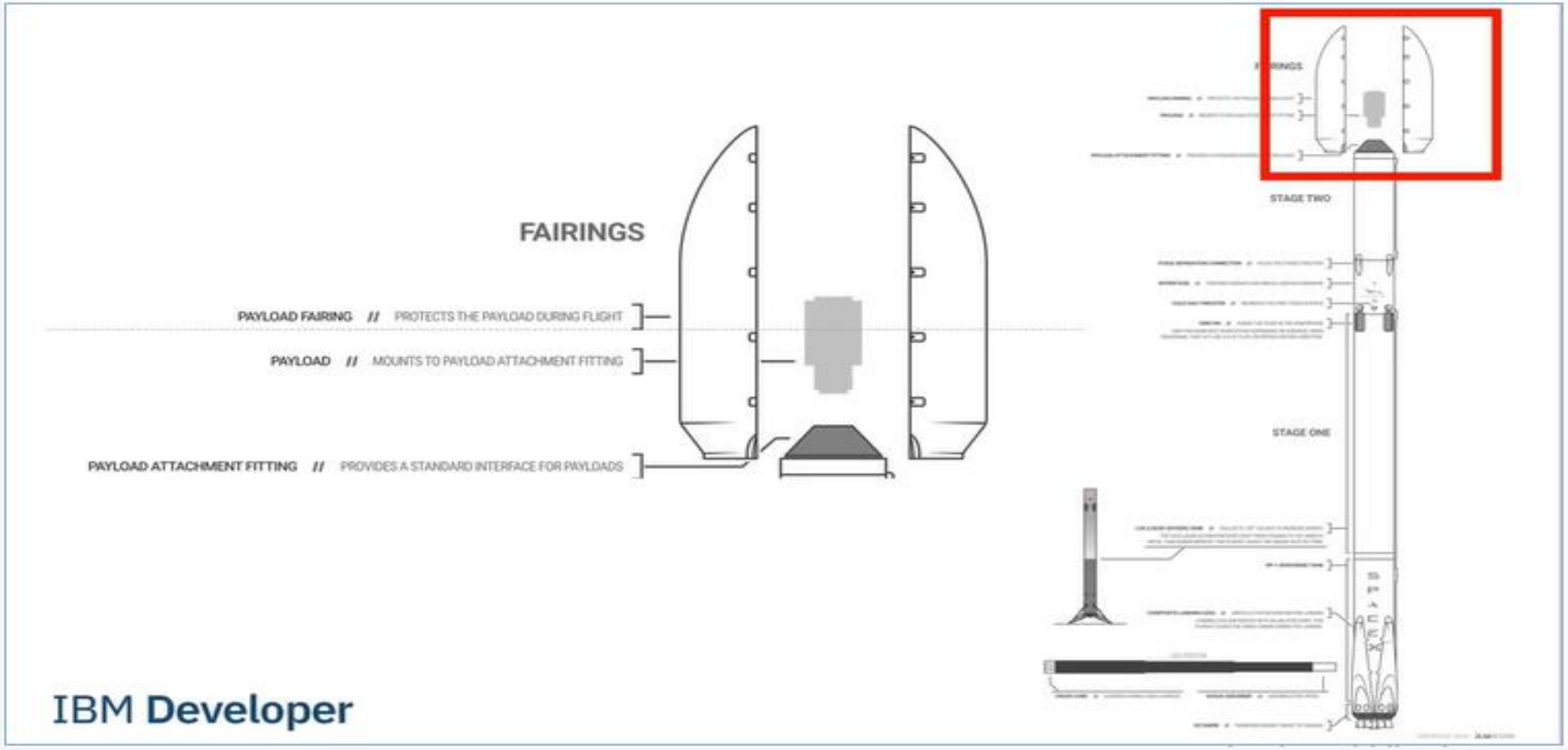
- In the test dataset there are 12 True/Positive, 03 False/Positive, 00 False/Negative and 03 True/Negative

Conclusions

- In this project, the objective is to predict if the Falcon 9 first stage will land successfully.
- The success rate since 2013 kept increasing till 2020.
- Launch sites are in close proximity to coastline, railway and highway, and away from cities. For each site, success appears related to the number of flights. But for the VAFB-SLC launchsite there are no rockets launched for heavy payload mass.
- ES-L1, GEO, HEO, SSO orbits have the highest success rate. But with heavy payloads the successful landing are higher for PO,LEO and ISS orbits.
- Predictive analysis showed that Decision tree model has the highest classification accuracy (88.9%)

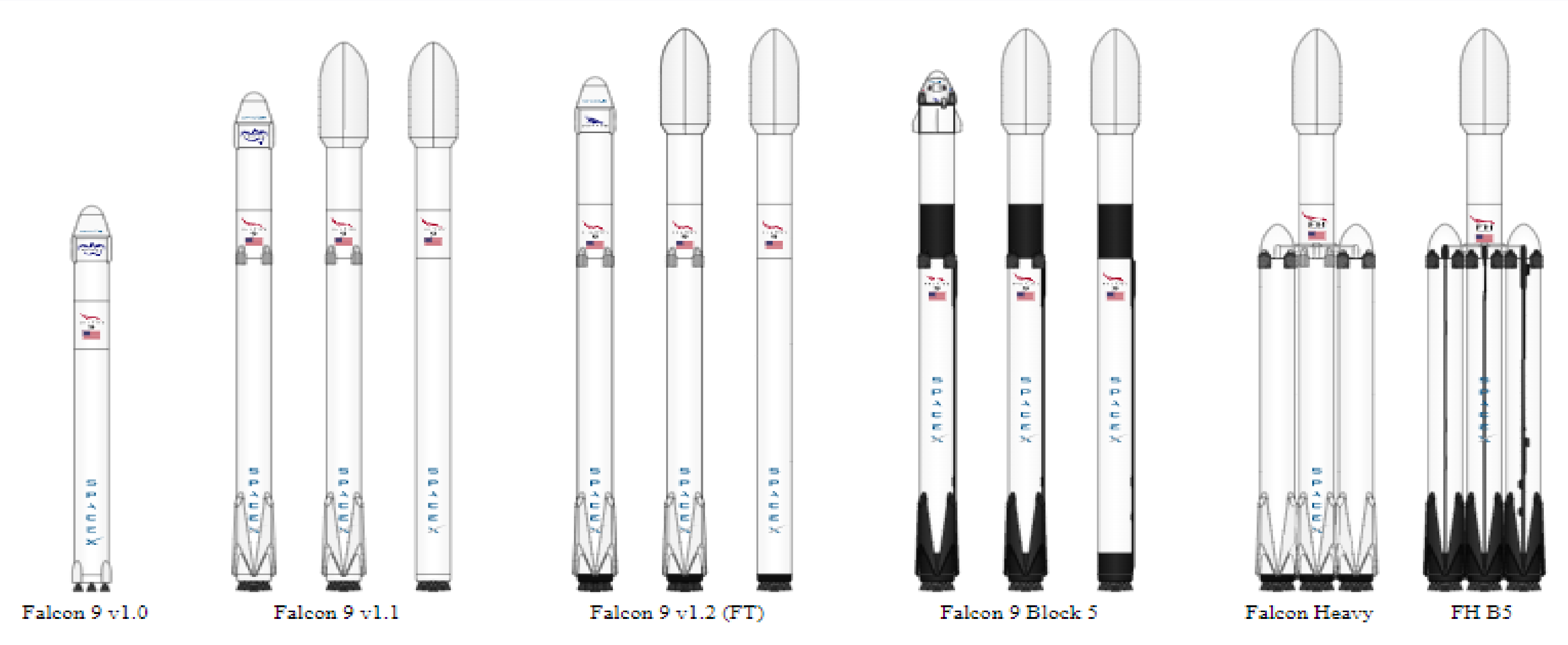
Appendix

- Spaces X’s Falcon 9 Rocket



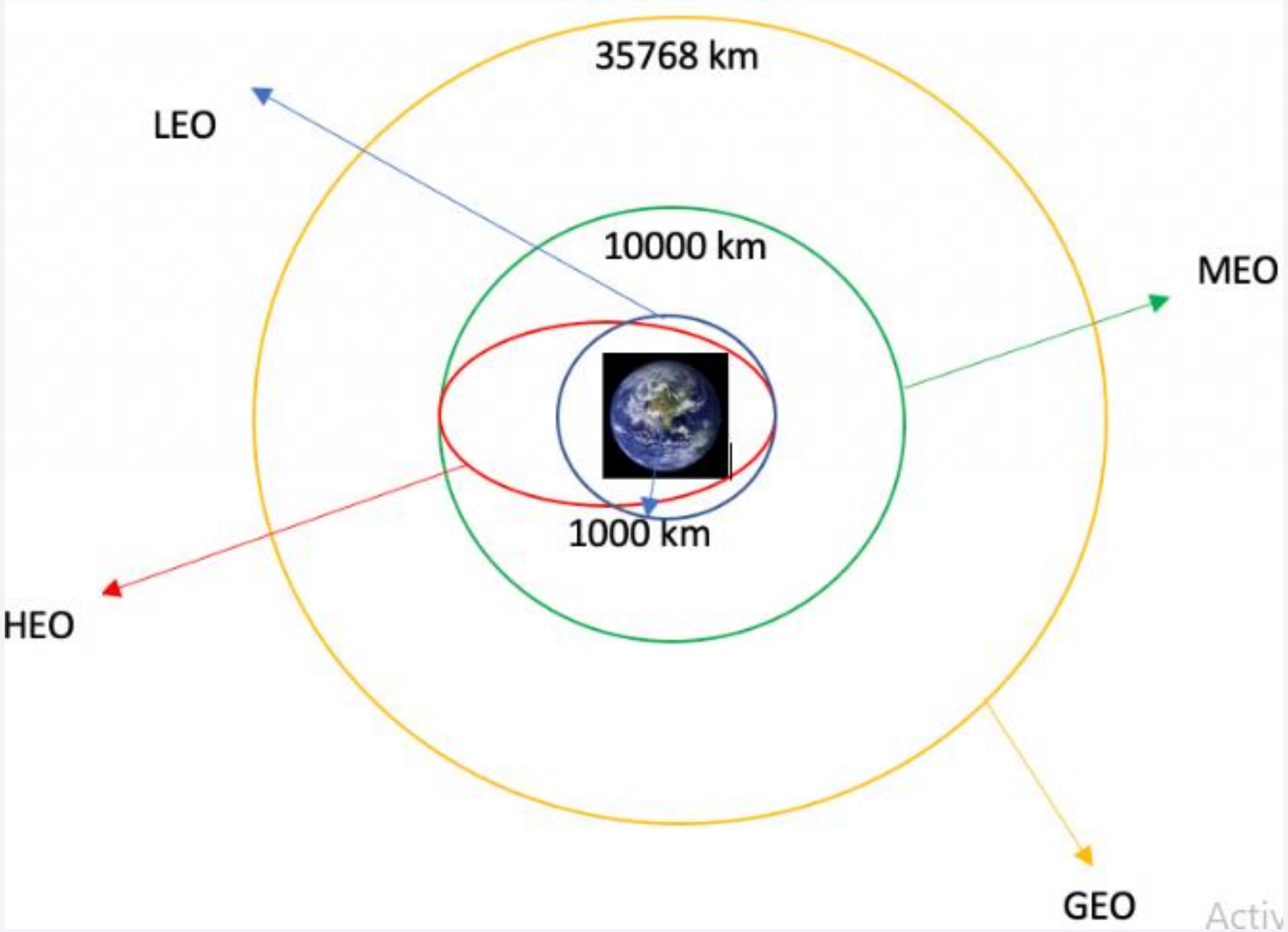
Appendix

- Versions of Spaces X's Falcon 9 Rocket



Appendix

- Some Orbit types



Appendix

- The columns of the Pandas dataset

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude	Class
0	1	2010-06-04	Falcon 9	6104.959412	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0003	-80.577366	28.561857	0
1	2	2012-05-22	Falcon 9	525.000000	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0005	-80.577366	28.561857	0
2	3	2013-03-01	Falcon 9	677.000000	ISS	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0007	-80.577366	28.561857	0
3	4	2013-09-29	Falcon 9	500.000000	PO	VAFB SLC 4E	False Ocean	1	False	False	False	NaN	1.0	0	B1003	-120.610829	34.632093	0
4	5	2013-12-03	Falcon 9	3170.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1004	-80.577366	28.561857	0

Thank you!