SPACEX DATA SCIENCE CAPSTONE PROJECT

BY ANKIT SHARMA
6 SEP 2021

Outline

1. EXECUTIVE SUMMARY
2. INTRODUCTION
3. METHODOLOGY
4. RESULTS
5. CONCLUSION

Executive Summary

- Spacex is the leading company in launching space missions. these missions are very costly hence predicting the success of a launch is extremely critical as it can save hundred of millions of capital

- Our priority is to predict the cost of a mission which is largely dependent on the reusability of second stage.

- The data of past launches is collected from different sources. The data is manipulated and a predictive classification model is developed to predict whether second stage will land successfully.

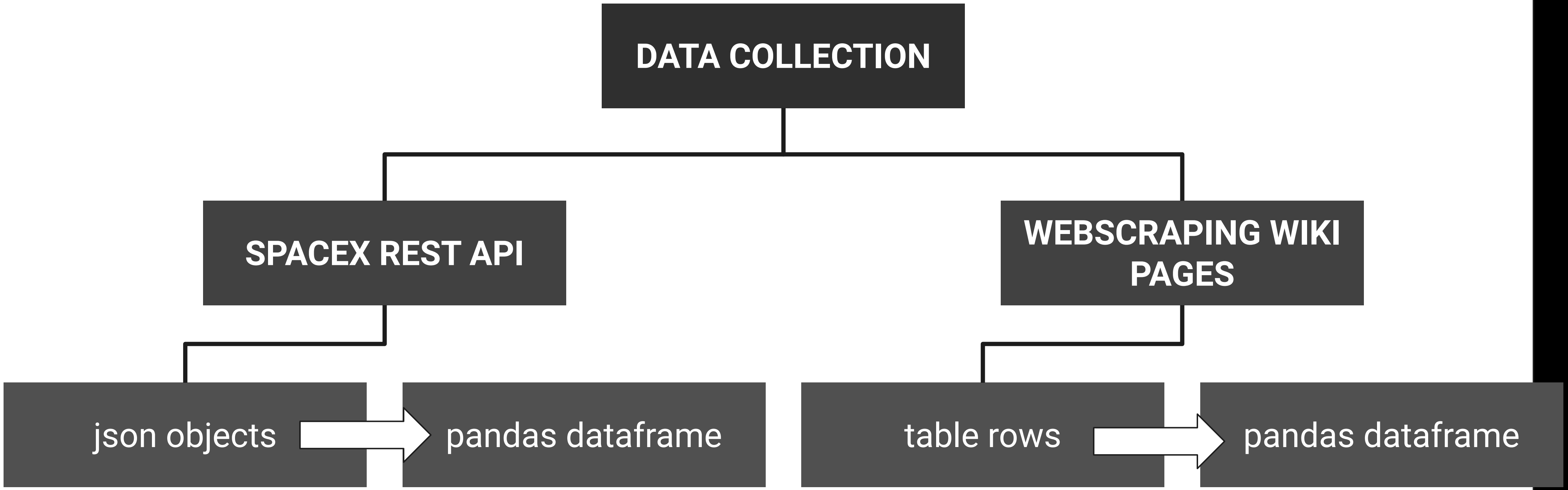
- The final outcome of the launch and hence cost of the mission is correlated to various factors including but not limited to launch site, orbit type, payload, booster version category.

Introduction

SpaceX's accomplishments include: Sending spacecraft to the International Space Station. Starlink, a satellite internet constellation providing satellite Internet access. Sending manned missions to Space. One reason SpaceX can do this is the rocket launches are relatively inexpensive. SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upwards 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. Spaces X's Falcon 9 launch like regular rockets. Our main job is to determine the price of each launch. We will do this by gathering information about Space X and creating dashboards for team. We will also determine if SpaceX will reuse the first stage. Instead of using rocket science to determine if the first stage will land successfully, We will train a machine learning model and use public information to predict if SpaceX will reuse the first stage.

Methodology

Data collection



Data collection – SpaceX API

requests.get(spacex_url)

spacex_url="https://api.spacexdata.com/v4/launches/past"

Perform a get request using the requests library to obtain the launch data, which we will use to get the data from the API

response.json()

This result can be viewed by calling the .json() method. Our response will be in the form of a JSON, specifically a list of JSON objects

pd.json_normalize()

To convert this JSON to a dataframe, we can use the json_normalize function. This function will allow us to "normalize" the structured json data into a flat table.

Data wrangling

Task 1
Wrangling Data using an API

In some of the columns, like rocket, we have an identification number, not actual data. This means we will need to use the API again targeting another endpoint to gather specific data for each ID number. These functions are already created for you, and will use the following: Booster, Launchpad, payload, and core.

Task 2
Sampling the Data

The launch data we have includes data for the Falcon 1 booster whereas we only want falcon 9. In this lab, you will need to figure out how to filter/sample the data to remove Falcon 1 launches.

Task 3
Dealing With Nulls

Our gathered data isnot perfect. We may end up with data that contains NULL values.calculate the mean of the PayloadMass data and then replace the null values in PayloadMass with the mean..

EDA with data visualization

For data visualisation following charts are plotted

- Scatter plot of Flight Number Vs Launch Site to get better insights of number of flights launched from particular sites
- Scatter plot of Payload Vs Launch Site to get insights of how payload masses distributed in different launch sites
- Bar Graph of Success Rate Vs Orbit type describing the success rates of launches for different orbits
- Scatter plot of Flight Number Vs Orbit Type showing which orbit has what number of flights.
- Scatter plot of Payload Vs Orbit Type showing which orbits have launches with particular payload masses.
- Line chart of Yearly Trend of Success rate showing the success rate in launches since year 2010.
- https://github.com/aksharma786/Data-Science-Capstone/tree/master

EDA with SQL

- 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 acheived.
- 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 records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015
- Rank the count of successful landing_outcomes between the date 2010-06-04 and 2017-03-20 in descending order.

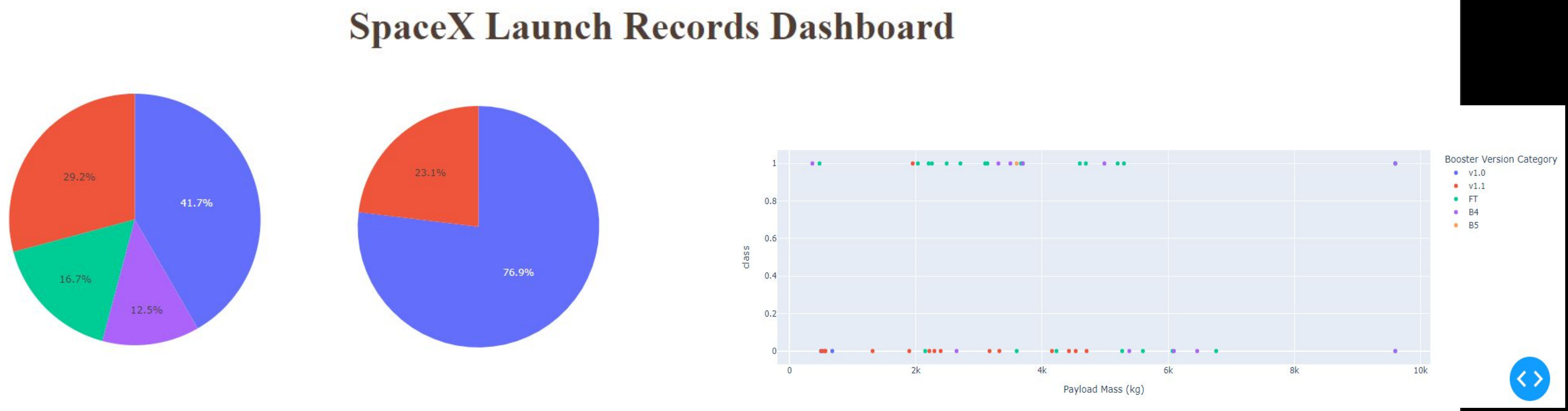
Build an interactive map with Folium

Folium is widely used to make creative and interactive maps of real world. Following map objects were used for making maps interactive.

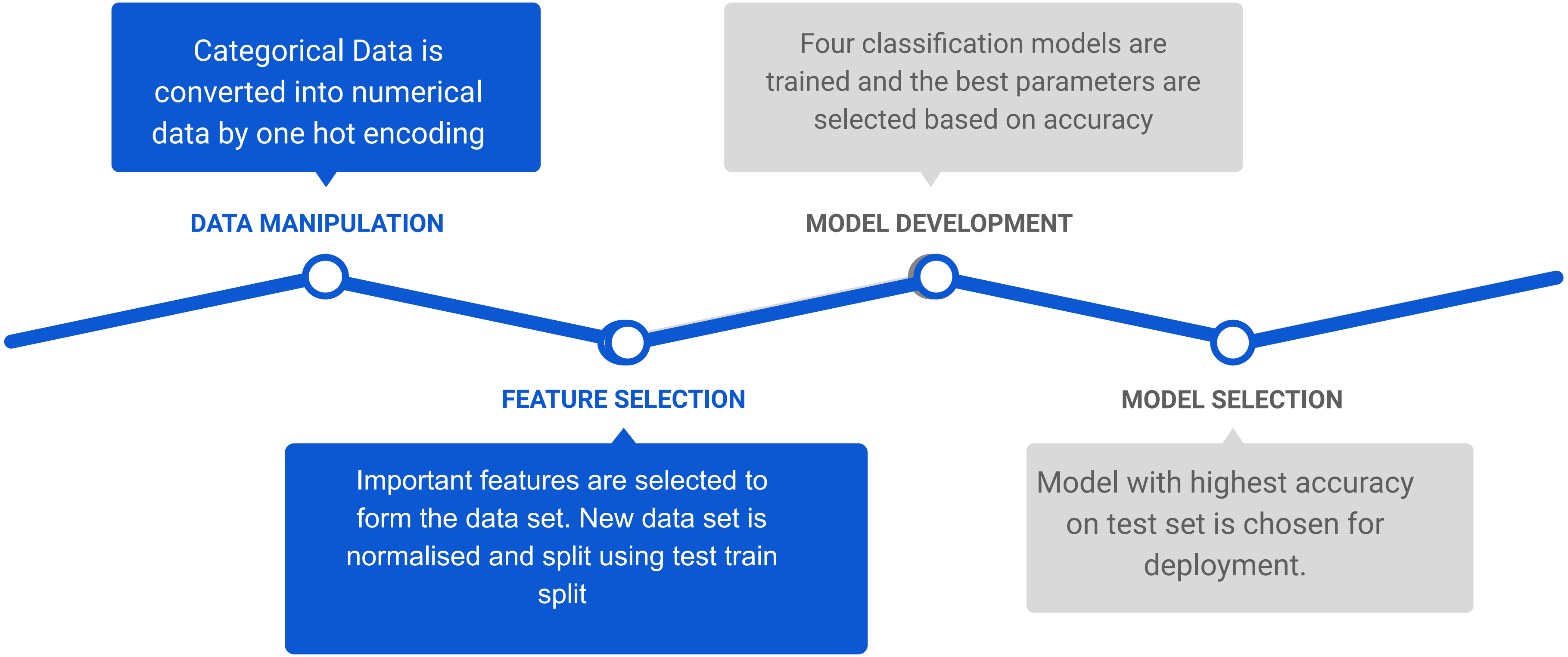
- **folium.Circle()** adds circle at the location specified
- **folium.marker()** helps adding markers with interactive icons
- **MarkerCluster()** is used in map with many markers. Every launch record is marked green or red according to landing outcome
- **folium.PolyLine()** is used to make polyline overlays on map. Nearest railway and highway points are connected to the launch site using polyline

Build a Dashboard with Plotly Dash

- Pie chart of total number of successful launches for all sites
- Pie chart of successful and fail outcomes of landing for selected launch site
- Scatter plot of Payload vs Launch Outcome for all sites with different Booster version category



Predictive analysis (Classification)



Results

Exploratory data analysis results

- All launch sites are in close vicinity of equator. All launch sites are also close to coast line.
- Developed transport infrastructure is also considered. Launch sites in close proximity with railways and highways have high success rate.
- The maximum number of successful launches have taken place at KSC LC 39A

Predictive analysis results

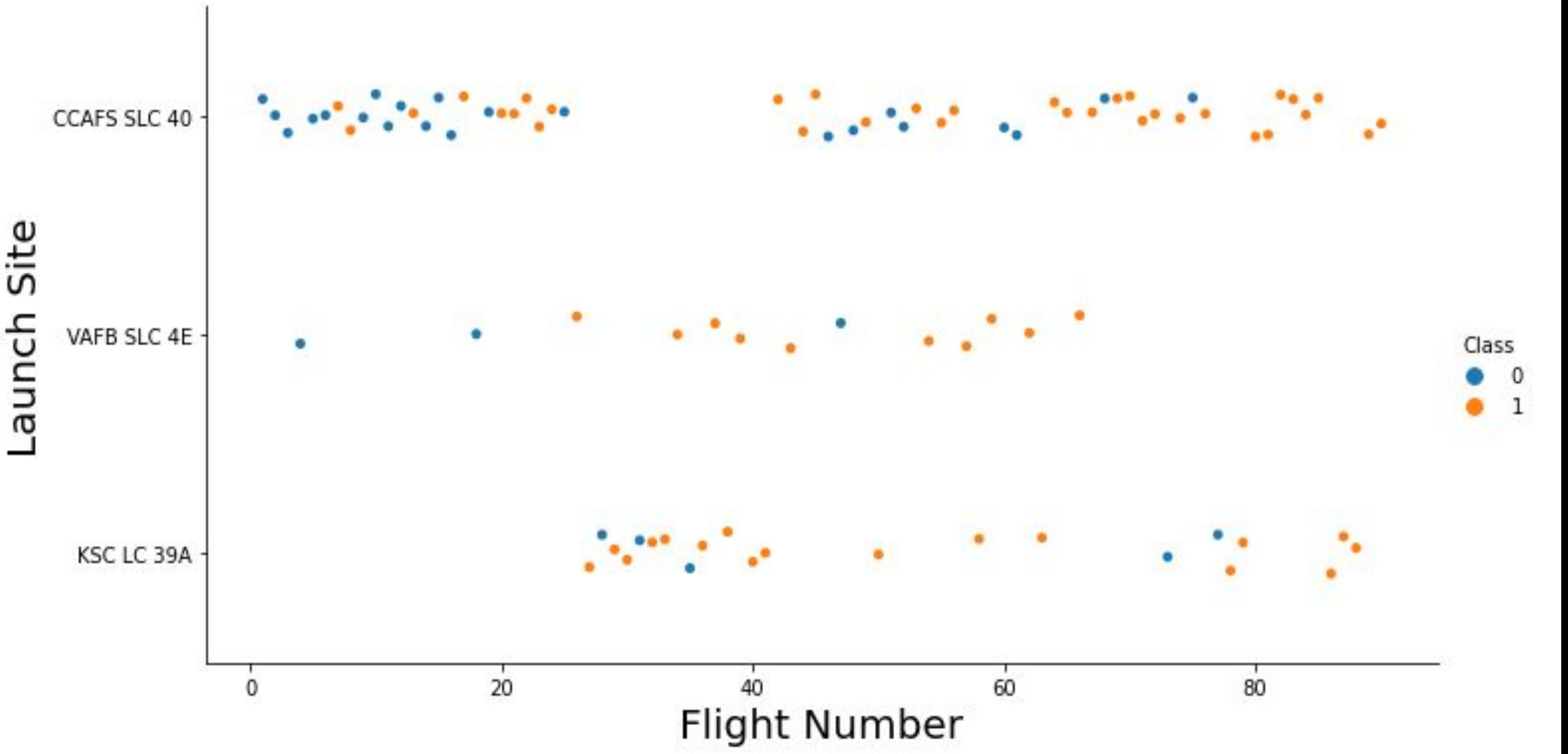
Model accuracy is 89 percent which shows that landing outcome is highly predictable.

EDA with Visualization

Flight Number vs. Launch Site

The launch site with maximum number o
flights is CCAFS SLC 40

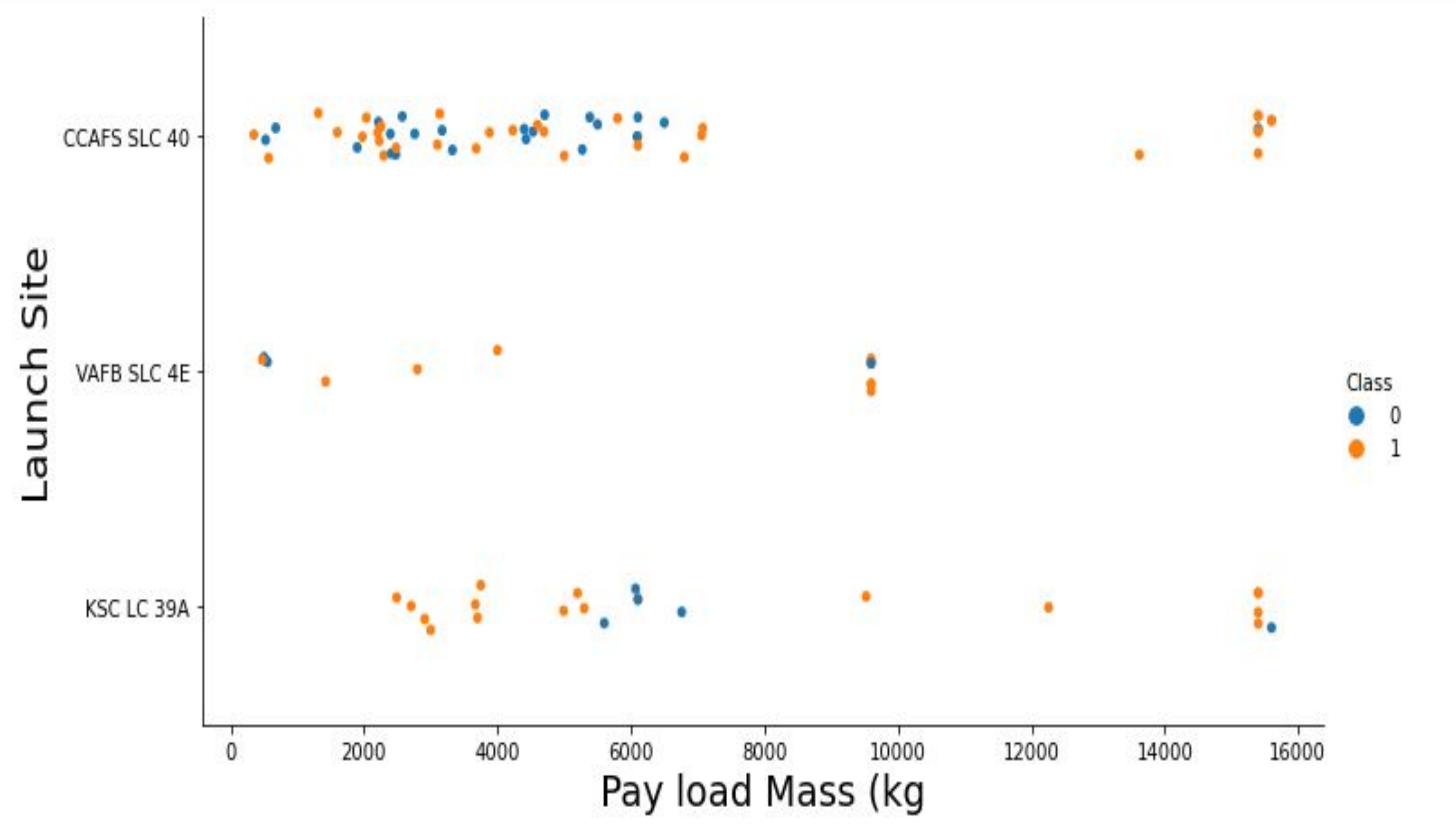
The launch site with maximum mission success rate is KSC L
39 A



Payload vs. Launch Site

Most of the light payloads were launched from CCAFS and KSC launch site

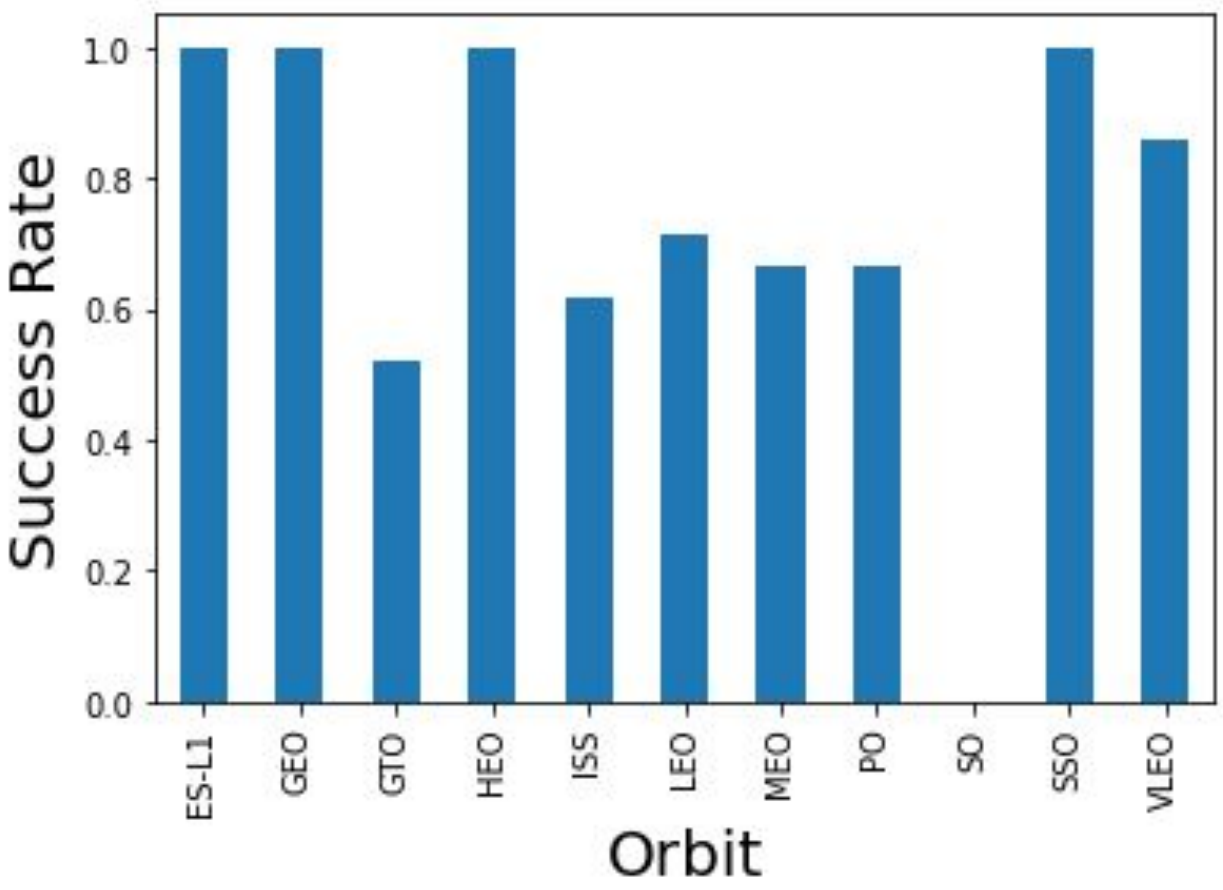
CCAFS SLC 40 has the highest success rate for launches with heavy payload.



Success rate vs. Orbit type

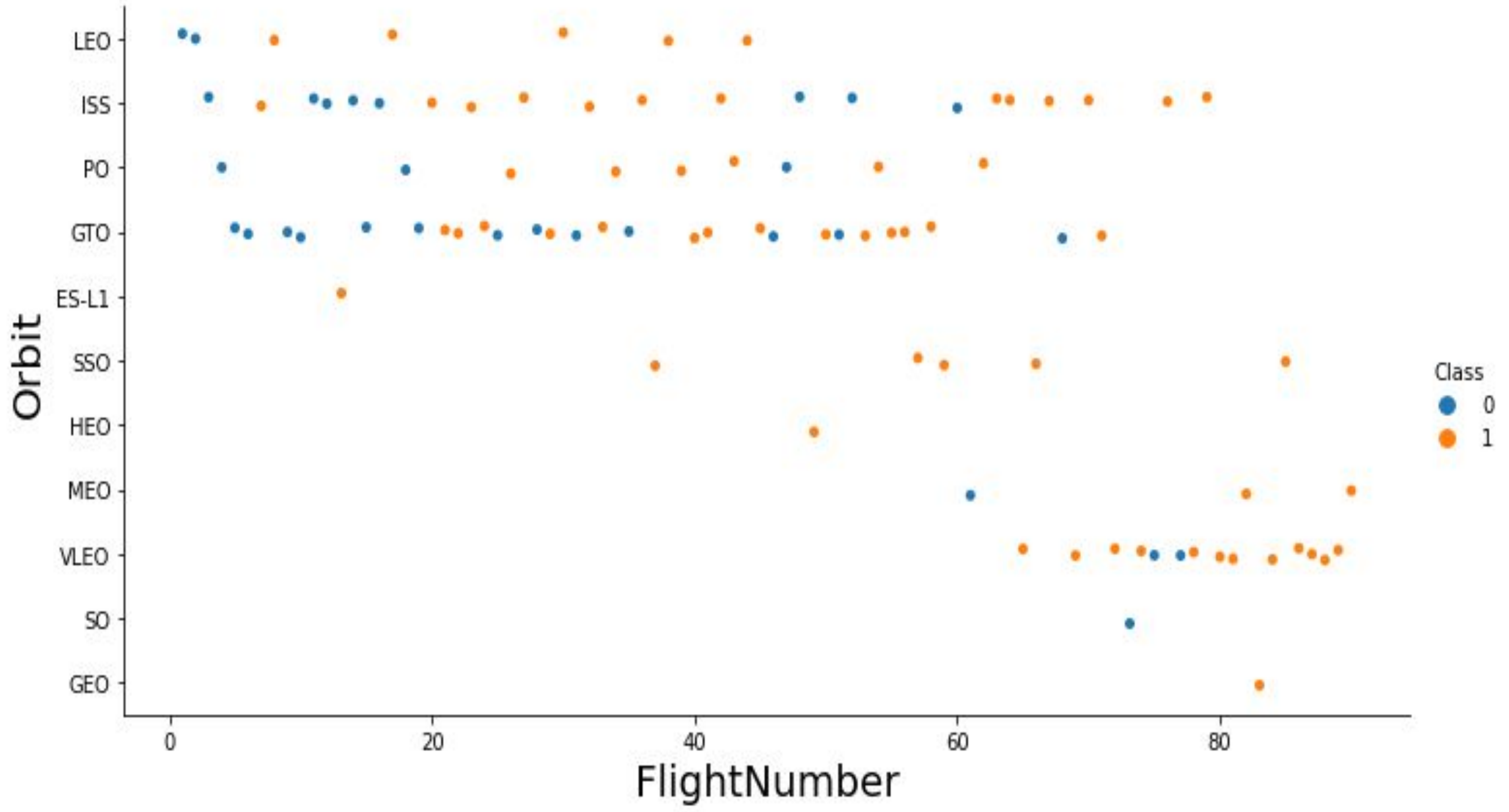
ES-L1,GEO,HEO,SSO orbit types have the highest success rate of 100 percent.

GTO orbit type has the lowest success rate and **SO** orbit type has no successful launch.

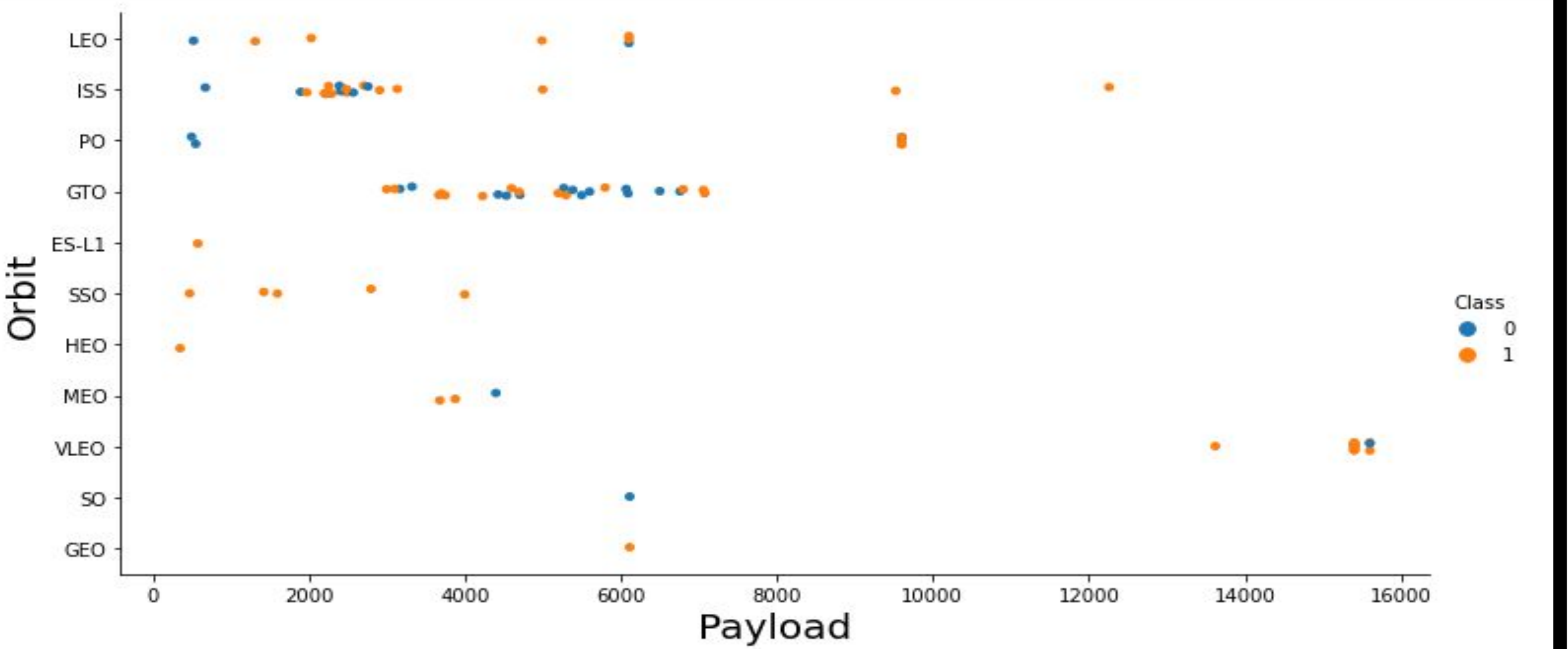


Flight Number vs. Orbit type

Landing success rate is low for low flight number and increases as the flight number is increased.



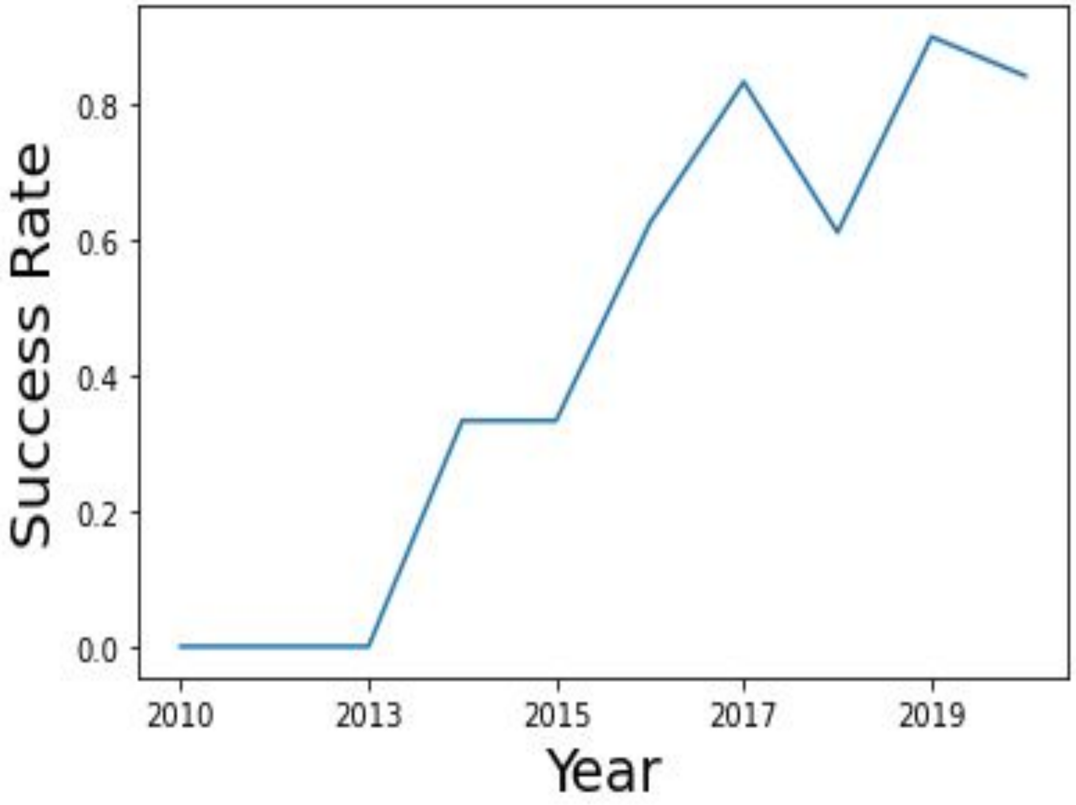
Payload vs. Orbit type



Launch success yearly trend

There have been substantial growth in success rate after 2013.

There is a dip in between years 2017 to 2019 which shows that there have been many fail launches in the company.



EDA with SQL

All launch site names

Knowing what launch sites are being used by SPACEX have critical importance in predicting the launch outcome.

SQL Query Result

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

Launch site names begin with `CCA`

- Find all launch sites 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	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

- The total payload carried by boosters from NASA

45596 kg.

Average payload mass by F9 v1.1

The average payload mass carried by booster version F9 v1.1 is 2534 kg

First successful ground landing date

The date when the first successful landing outcome in ground pad **2015-12-22**

Successful drone ship landing with payload between 4000 and 6000

- F9 FT B1022
- F9 FT B1026
- F9 FT B1021.2
- F9 FT B1031.2

These are the booster Versions with successful Drone Ship landing with payload mass between 4000 and 6000 kg.

Total number of successful and failure mission outcomes

- Calculate the total number of successful and failure mission outcomes

mission_outcome	occurences
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters carried maximum payload

- List the names of the booster which have carried the maximum payload mass
-

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

- The launch record from the year 2015 for failure in drone ship landing with the month name,

YEAR	month_name	landing__outcome	booster_version	launch_site
2015	January	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
2015	April	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

Rank success count between 2010-06-04 and 2017-03-20

- Rank the count of successful landing_outcomes between the date 2010-06-04 and 2017-03-20 in descending order.
- The highest number of succesful landing outcones are with drone ship followed by ground pad landing.

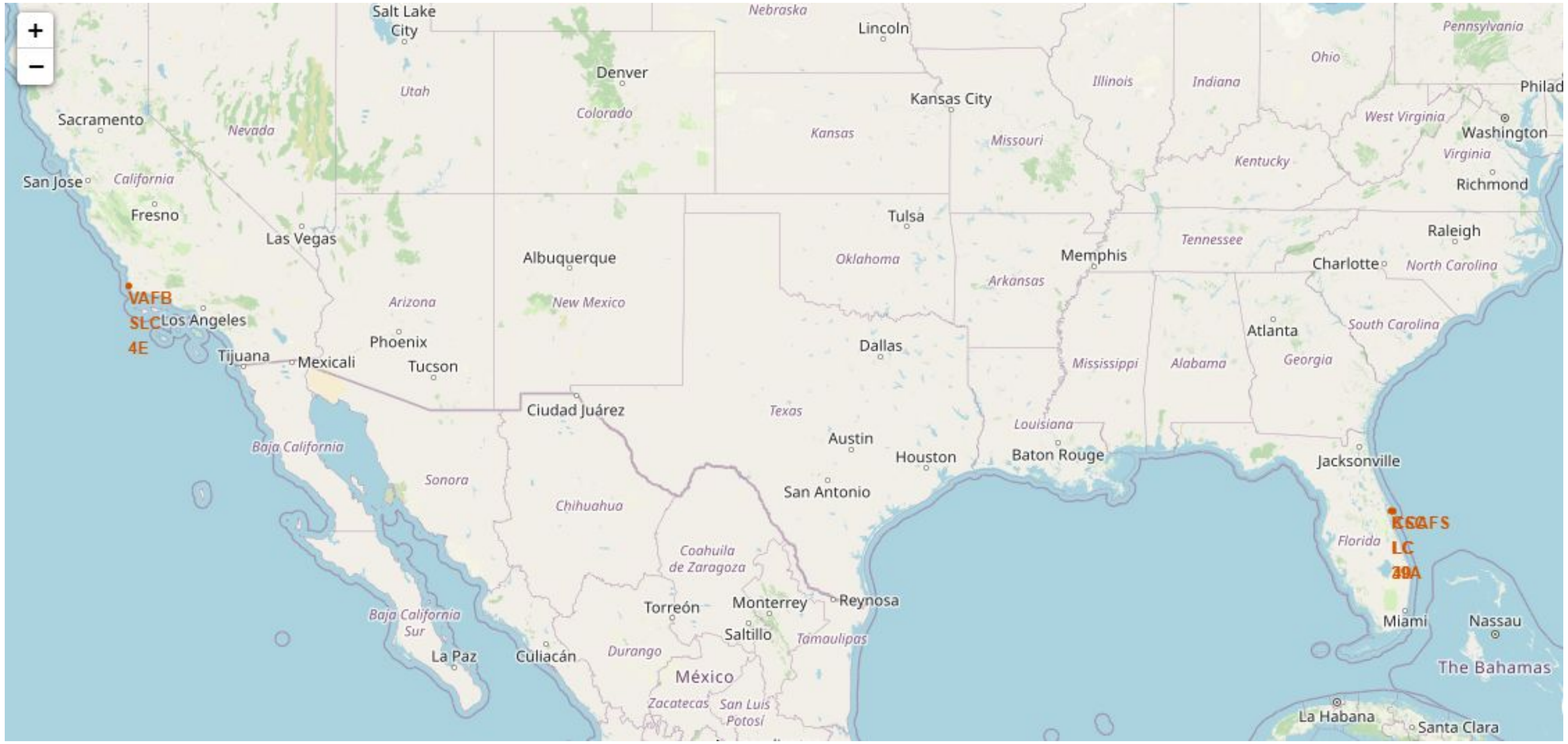
landing__outcome	count_
Success (drone ship)	5
Success (ground pad)	3

Interactive map with Folium

Launch Sites on Global map of US

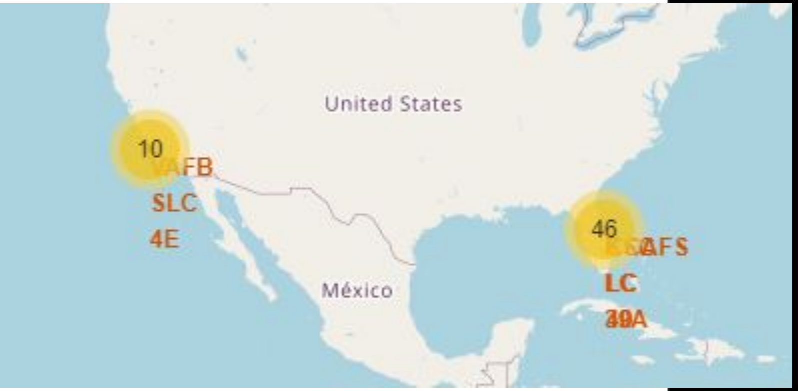
The location of four launch sites is marked on the map.

All launch sites are near equator and near coastline also.

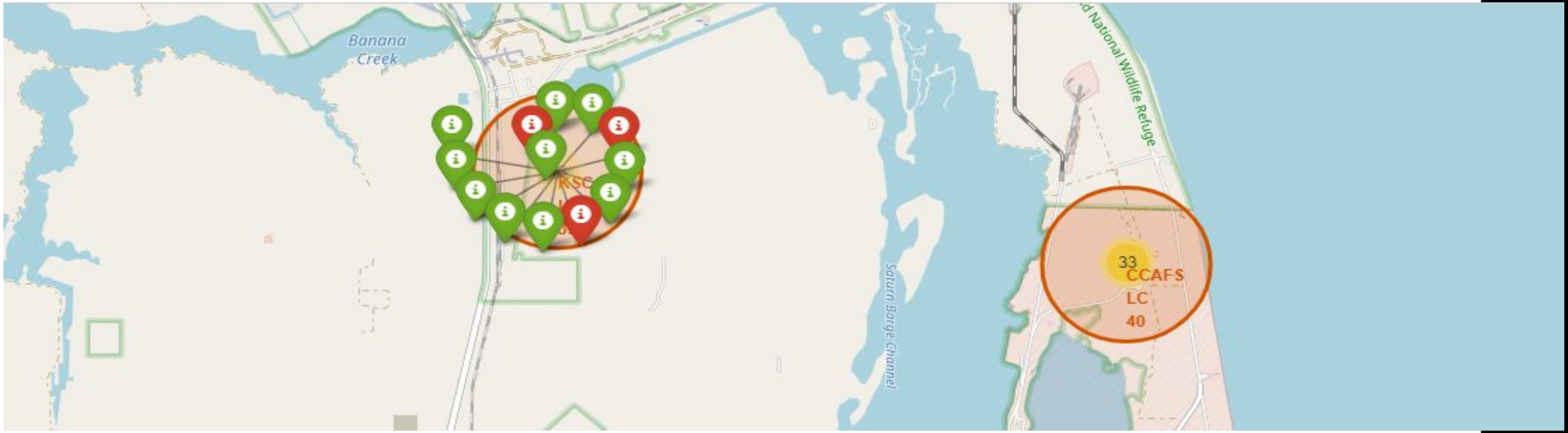


Location wise Launch record

First map shows the number of launches from different launch sites



Second map shows the color labeled launch record for the KSC LC 39 A launch site with 13 outcomes 10 success and 3 failures.

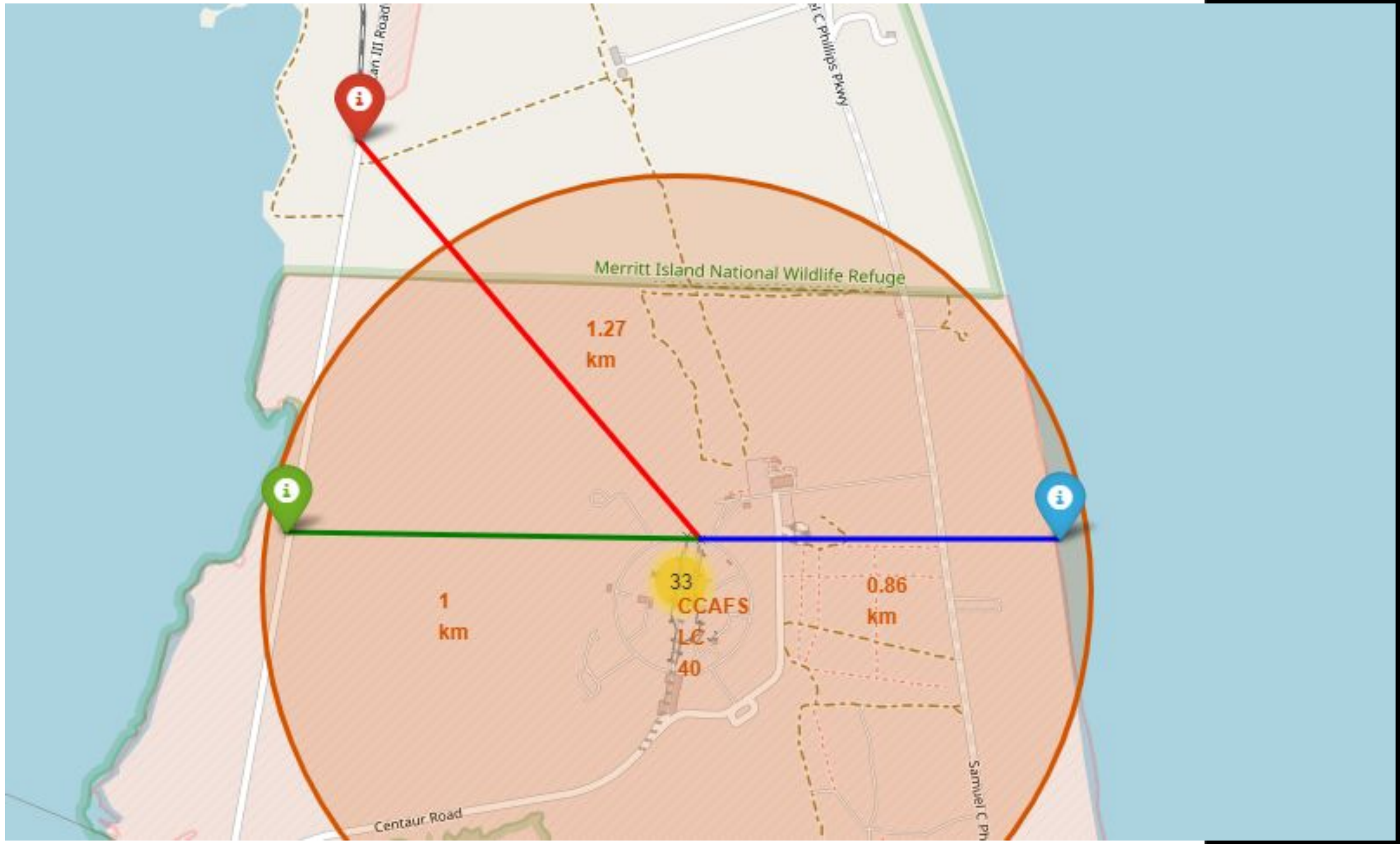


Analysis of close vicinity of Launch Sites

In this screenshot distance of a launch site from nearest Railway point, Highway and coastline is shown.

Railways and highways are suitable for careful transportation of launch vehicles hence launch sites are in close vicinity.

Launch sites are also close to coastal boundary in order to minimise the loss due to debris from fail launching scenarios



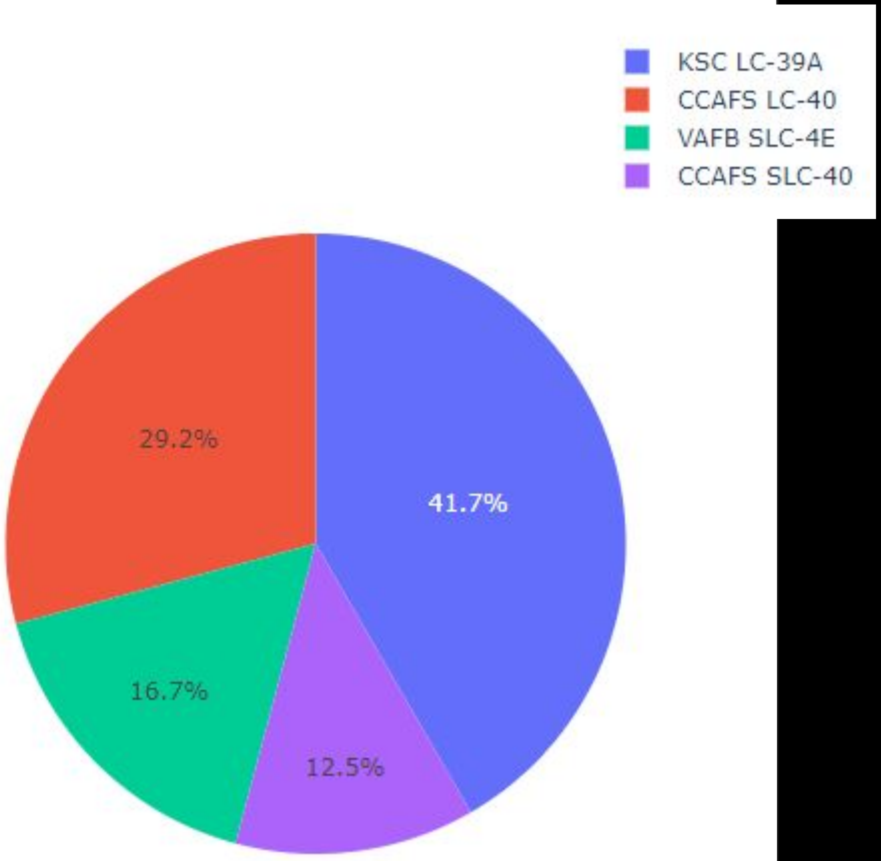
Build a Dashboard with Plotly Dash

Launch Success rate for all sites

- KSC LC 39A launch site has the highest number successful launches .

- CCAFS SLC 40 launch site has the lowest successful launches

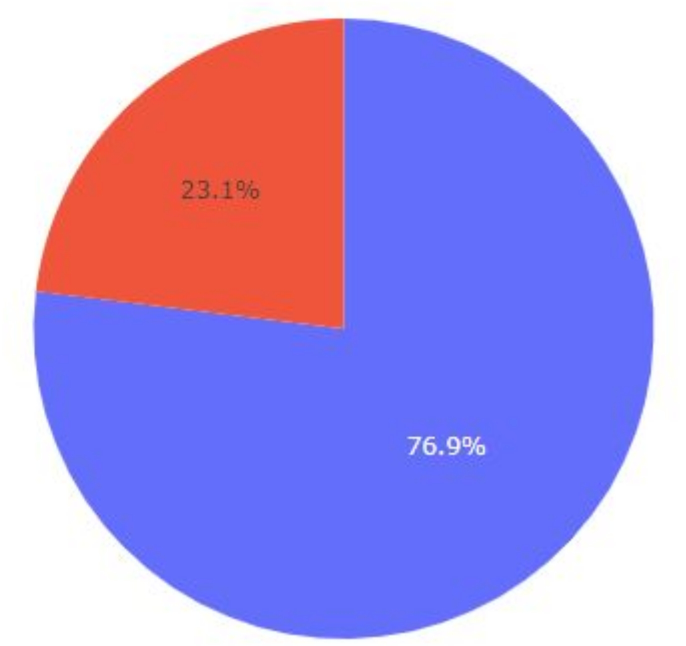
- The success launches from CCAFS LC 40 are more than double of CCAFS SLC 40



Launch record for KSC LC 39A

KSC LC 39A has the success rate of 76.9 percent.

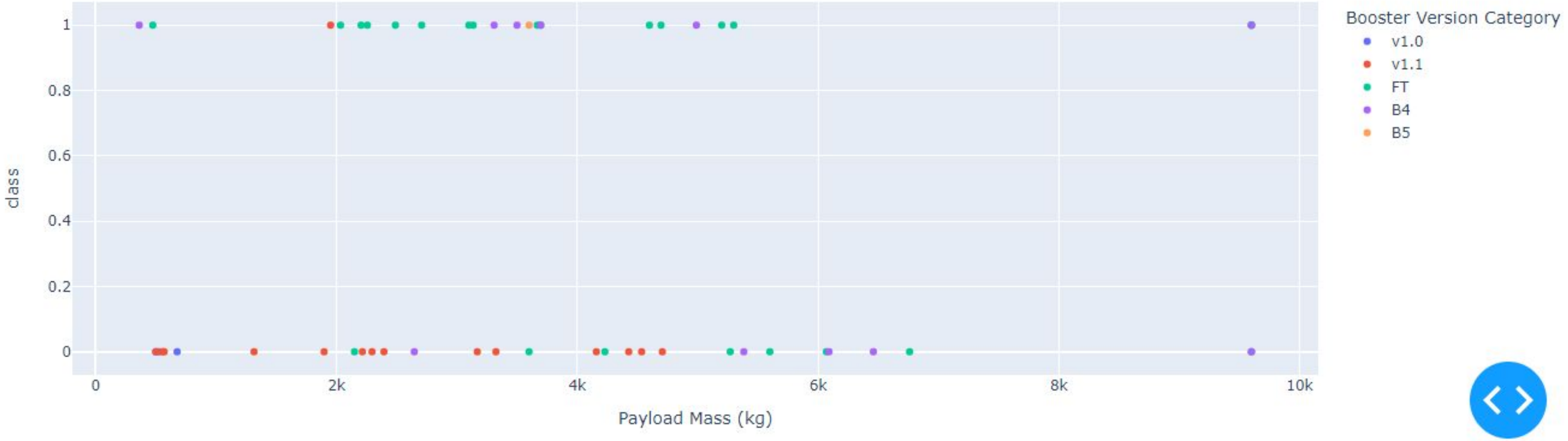
10 out of 13 launches were successfully landed.



Payload vs. Launch Outcome

Scatter plot for Payload mass Vs Launch outcome for different Booster version category

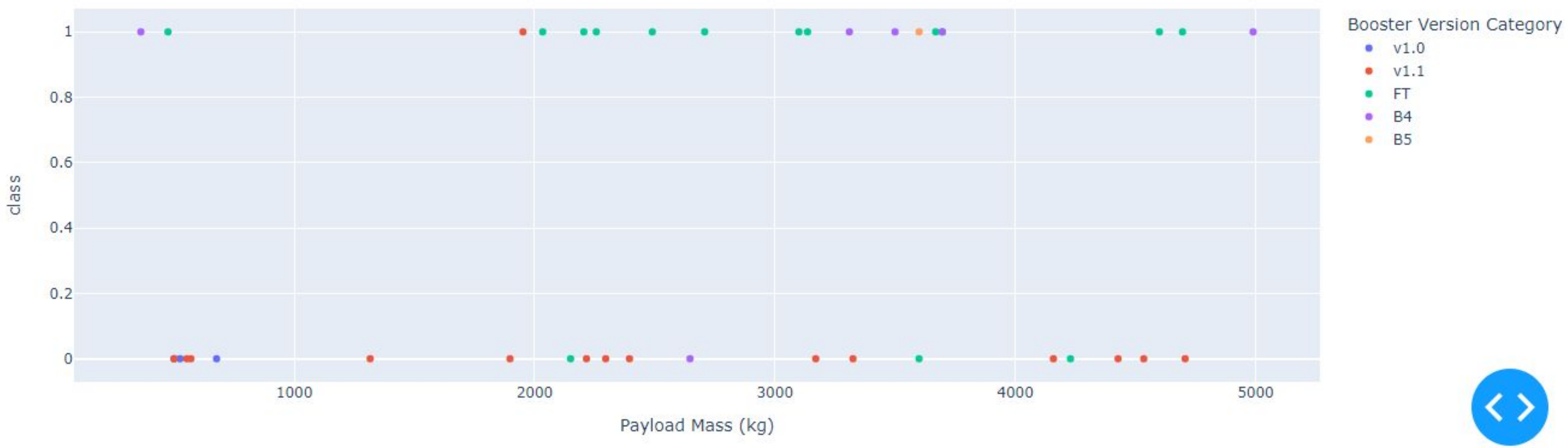
Payload Range 0 kg to 10000 kg



Payload vs. Launch Outcome

Scatter plot for Payload mass Vs Launch outcome for different Booster version category

Payload Range 0 kg to 5000 kg

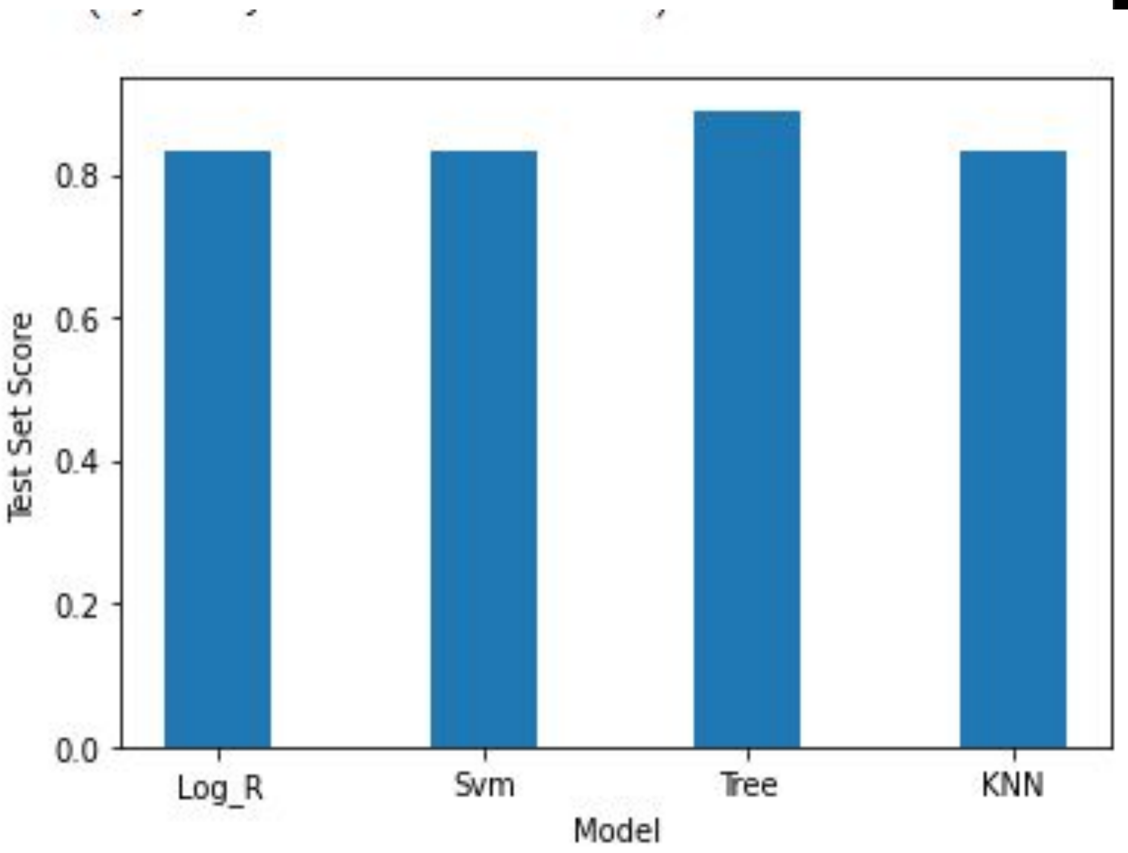


Predictive analysis (Classification)

Classification Accuracy

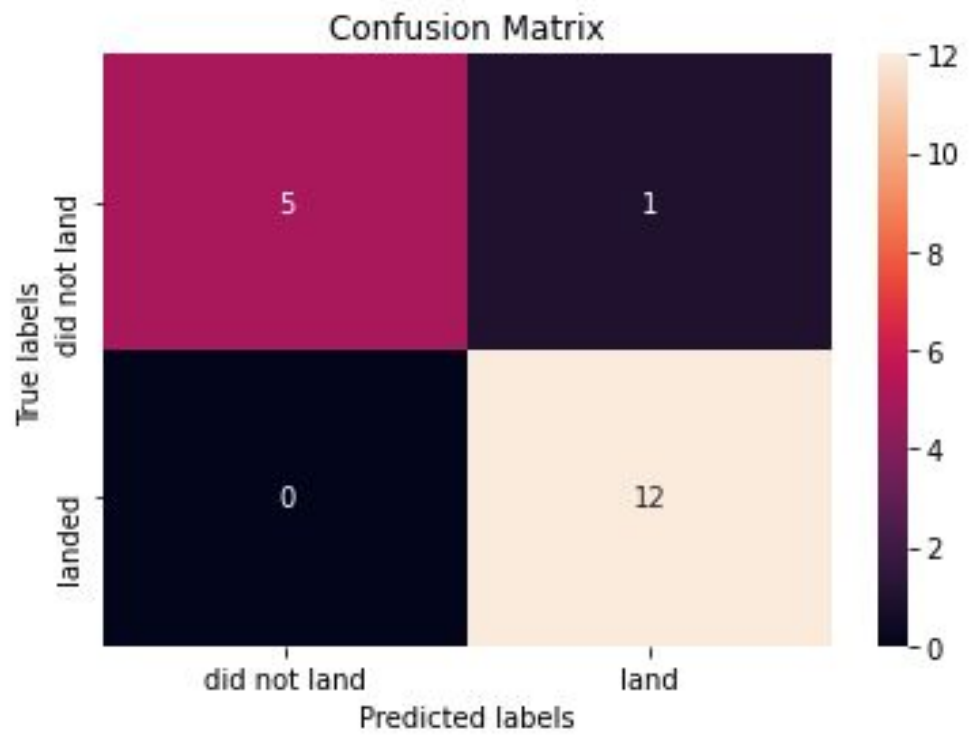
Decision Tree model has the highest classification accuracy

Model	Accuracy
Decision tree	0.89
SVM	0.83
KNN	0.83
Logistic Reg	0.83



Confusion Matrix

The predictive model developed for the data has considerably high accuracy. The confusion matrix shows that the model has only misclassified one launch record from the entire data set.



CONCLUSION

- Landing outcome is correlated to many factors and difficult to predict by human observation.
- Predictive model is fairly accurate and cost can be estimated on the basis of the landing outcome.
- Selection of appropriate launch site, orbit type, payload and booster version is critical for the success of mission.

APPENDIX

- The github repository where all the completed files are uploaded at

  https://github.com/aksharma786/Data-Science-Capstone/tree/master