



IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Harsha Veena Kondeti
15-2-2022



Outline



Executive
Summary



Introduction



Methodology



Results



Conclusion



Appendix

Executive Summary

By Using logistic regression model, we predicted launch will be 83% successful

We predicted launch by using four different models, in these four models logistic regression model has highest accuracy.

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.
- We have to predict if Falcon 9 first stage will land successfully.

Section 1

Methodology

Methodology

Executive Summary

Data collection methodology:

- We collect data by sending REQUEST to SPACEX API saved into pandas dataframe, and filter data with only falcon9 launches.

Perform data wrangling

- Here we deal with missing values and saved it to dataset_part_1.csv

Perform exploratory data analysis (EDA) using visualization and SQL

Perform interactive visual analytics using Folium and Plotly Dash

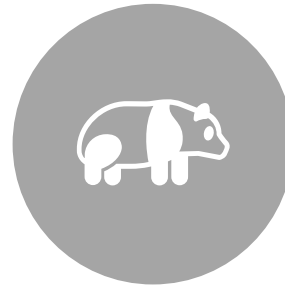
Perform predictive analysis using classification models

- Standardize data and evaluate accuracy of 4 classification models.

Data Collection



Data was collected from
SPACEX API through
REQUEST.



Store data in pandas data
frame.

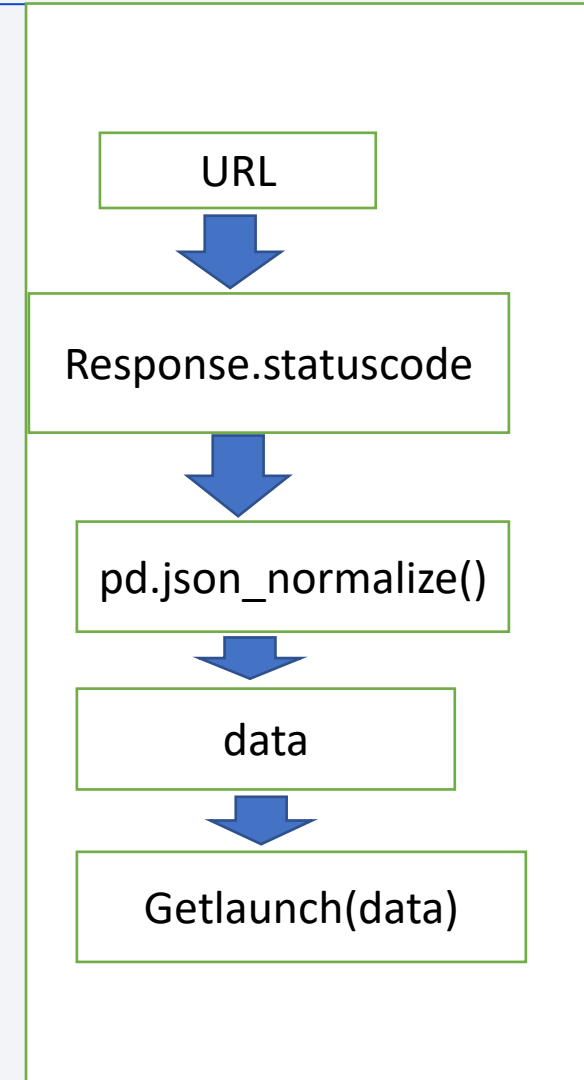
Data Collection – SpaceX API



Presenting data collection with SpaceX REST calls using key phrases and flowcharts

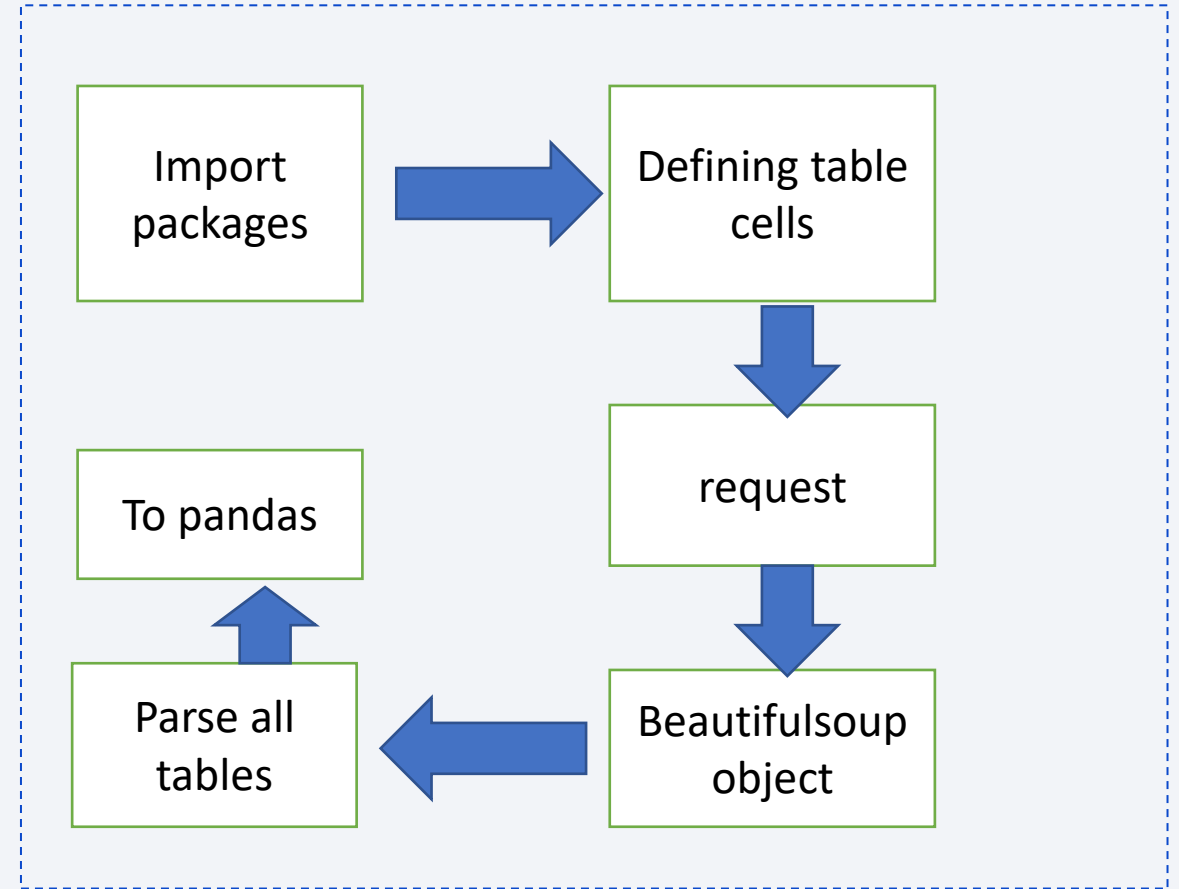


GitHub URL of the completed SpaceX API calls notebook ([spaceX data collection url](#)), as an external reference and peer-review purpose



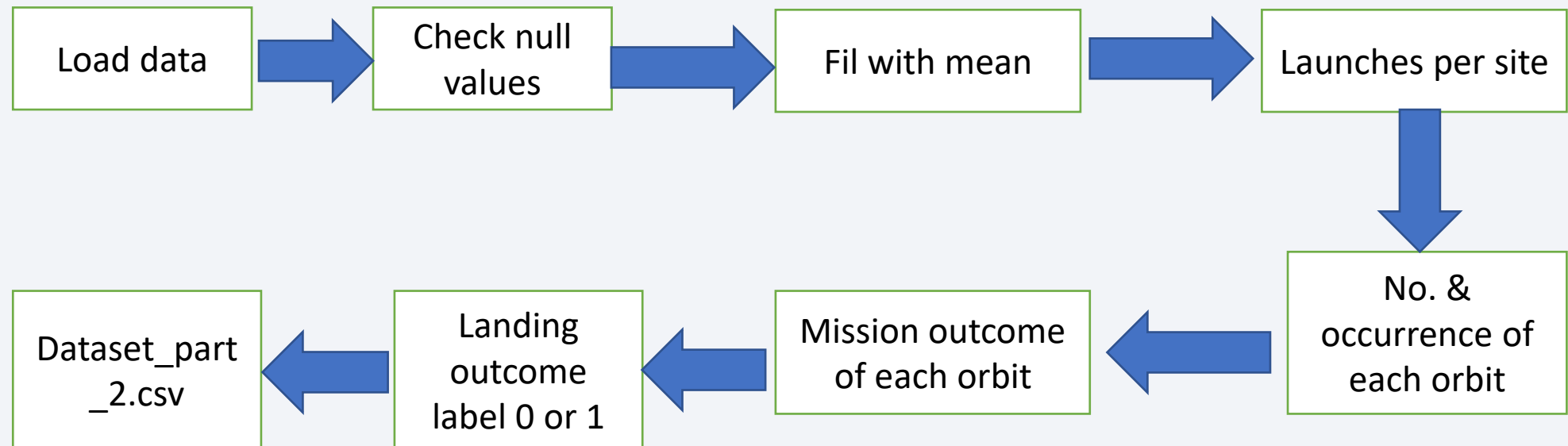
Data Collection - Scraping

- Presenting web scraping process using key phrases and flowcharts
- GitHub URL of the completed [web scraping](#) notebook, as an external reference and peer-review purpose



Data Wrangling

- Data was processed with dealing missing values
- We calculated launches per site, no. & occurrences of each orbit, mission outcome of each orbit and assign label 0 or 1 to landing outcome and saved it to csv file.
- GitHub URL of completed [data wrangling](#) notebook, as an external reference and peer-review purpose



EDA with Data Visualization

- GitHub URL of completed [EDA with data visualization](#) notebook, as an external reference and peer-review purpose
- In scatter plot of Flight number & launch site VAFB SLC4A has low success rate
- In scatter plot of payload & launch site ,no rockets launches for payload mass $> 100000\text{kg}$.
- In bar chart, success rate of each orbit ES-L1, GEO,HEO,SSO have high success.
- In yearly success trend plot success in 2013 kept increasing till 2020.
- Created dummy variables to categorical columns
- Convert numerical values to float64 and export it to dataset_part_3.csv
- These plots are plot to find success rate relate to which variables

EDA with SQL

- GitHub URL of [EDA with SQL](#) notebook, as an external reference and peer-review purpose
- Names of unique launch sites in space mission.
- 5 records where launch sites begin with string 'CCA'.
- Total payload mass carried by boosters launched by NASA.
- Average payload mass carried by booster version F9 v1.1
- Date when 1st successful landing outcome in ground pad was achieved.
- Names of boosters which have success in drone ship and payload mass between 4000 & 6000.
- Total number of successful & failure mission outcomes.
- Booster version names which have carried max payload mass.
- Failed landing outcomes in drone ship, their booster versions & launch site names in 2015.
- Ranking Landing outcomes between 2010-06-04 & 2017-03-20 in descending order.

Build an Interactive Map with Folium

- GitHub URL of [interactive map with Folium map](#), as an external reference and peer-review purpose
- I added all launch sites as markers, draw lines to near by railway line, road line, city in folium map.
- We added these objects to analyze about locations of launch sites ,whether they are near to any city or railway.

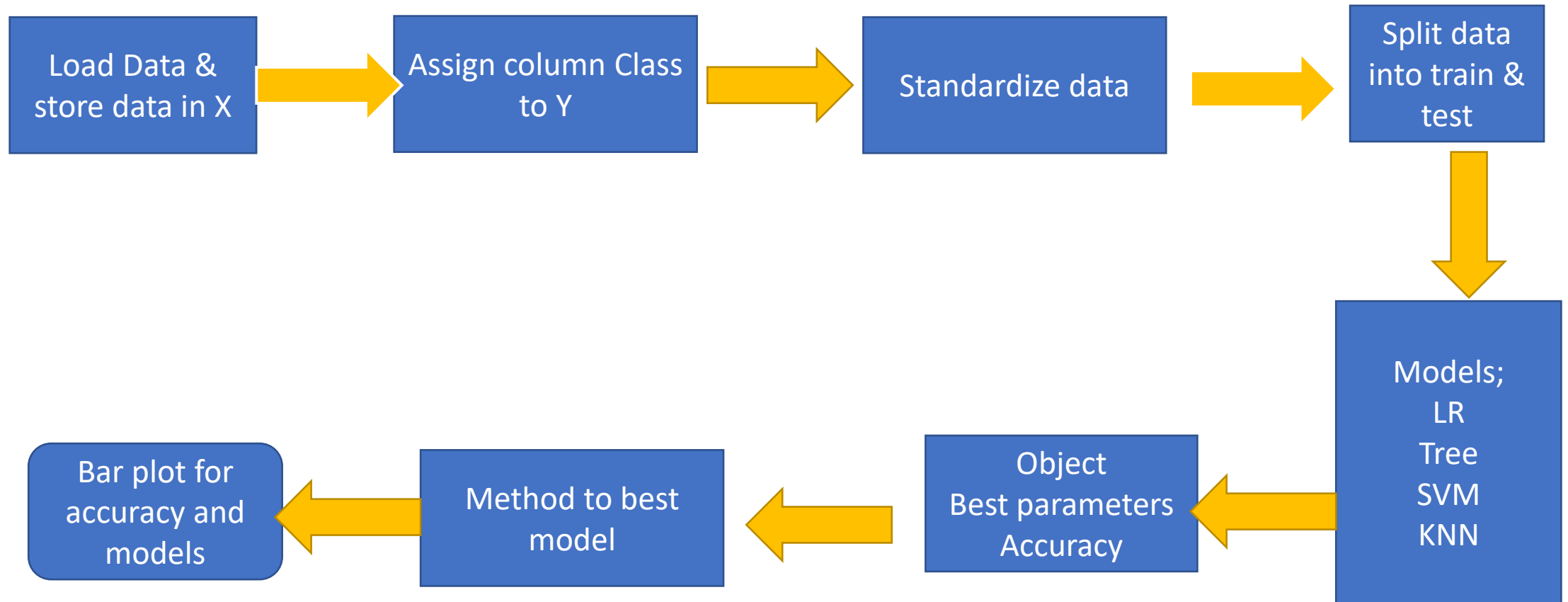
Build a Dashboard with Plotly Dash

- GitHub URL of [Plotly Dash](#) , as an external reference and peer-review purpose
- Scatter chart to show correlation between payload and launch success to dashboard.
- Added a pie chart for successful launches count of launch sites to dashboard.
- We added these charts to study which site has highest launches, in which pay range has more launches. Similarly with help of above charts we answer many questions like this

Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model
- You need present your model development process using key phrases and flowchart
- GitHub URL of [predictive analysis](#), as an external reference and peer-review purpose
- Create a column for class & standardize data, then split data into train and test sets.
- Finding best hyperparameter for SVM, KNN, Trees and Logistic Regression.
- Find the method performs best using test data.

Flow chart for predictive analysis & model development



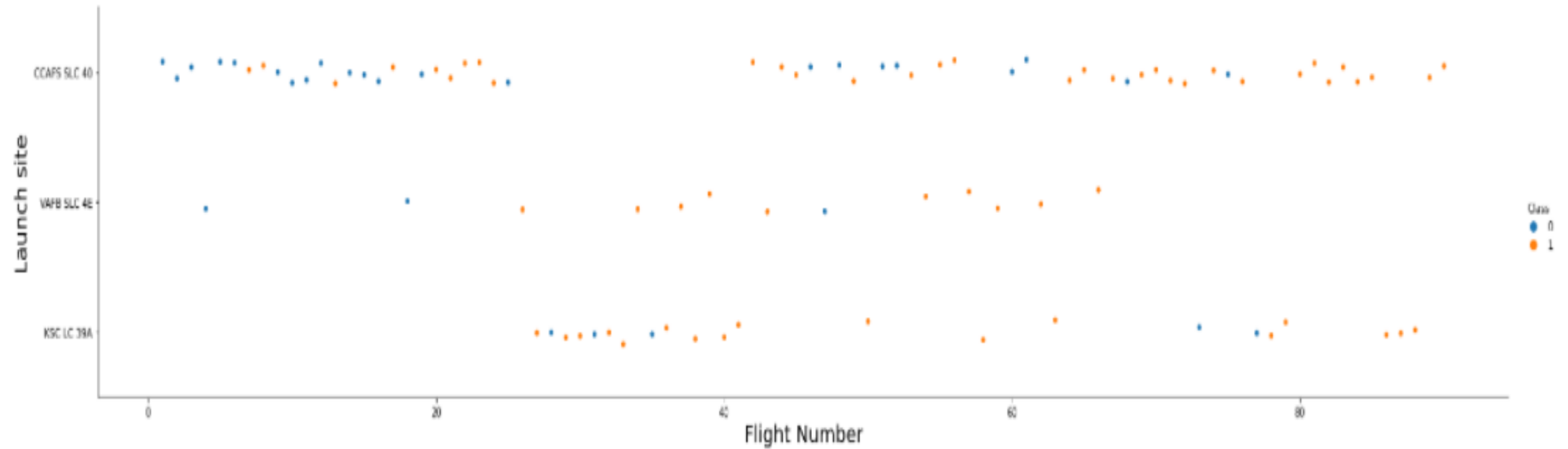
Results

- From Exploratory data analysis we can see success is increasing from 2013, for increasing payload mass makes decrease in chance of success, VAFB LC 4A have less success rate than other launch sites
- Interactive analytics demo in screenshots
- Best predicted model is Logistic regression with 83% accuracy
- So We have 83% chance in success of first stage of SpaceX launch.

The background of the slide is an abstract composition. It features a dark blue base color. Overlaid on this are numerous diagonal streaks in shades of red and cyan. A faint, light blue grid pattern is also visible, particularly in the lower half of the image. The overall effect is dynamic and technological.

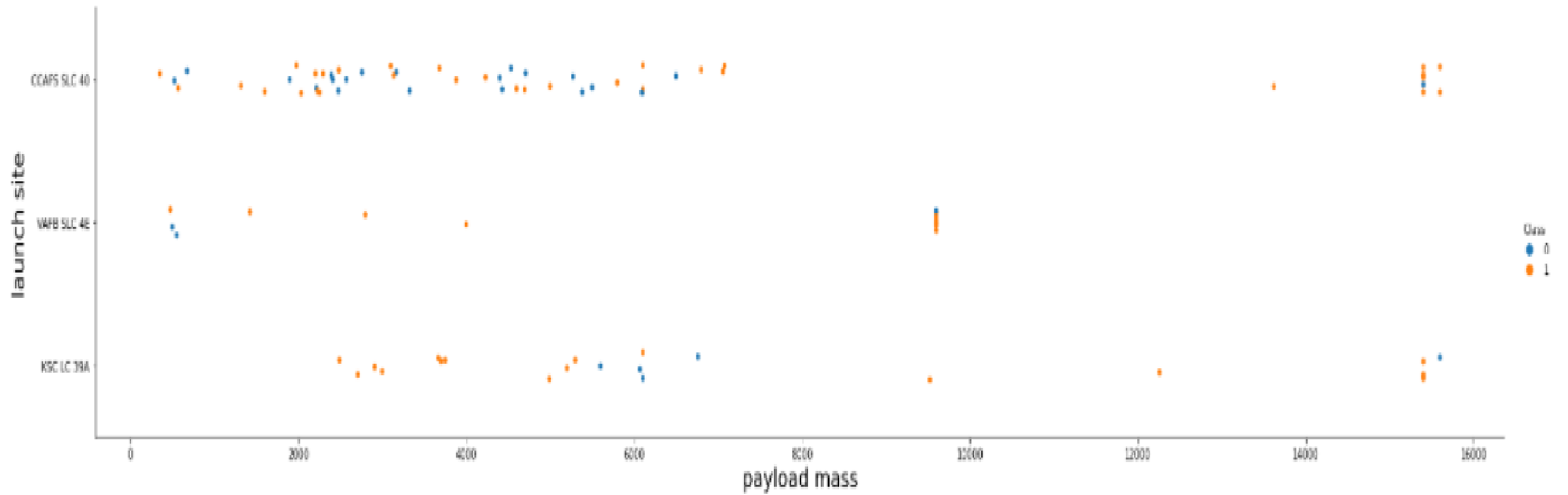
Section 2

Insights drawn from EDA



Flight Number vs. Launch Site

- Label 1 with orange color are successful outcomes
- CASFS SL 40 has more successful outcomes than other launch sites with more flight numbers.

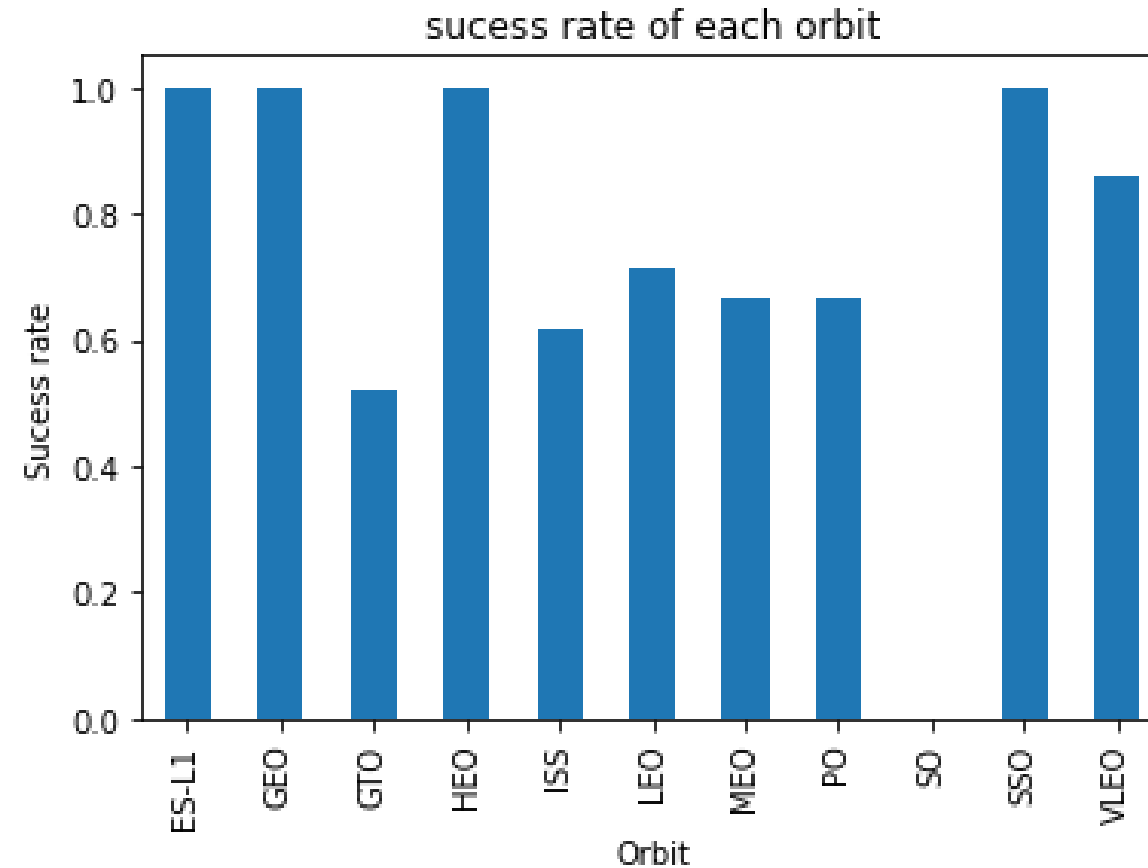


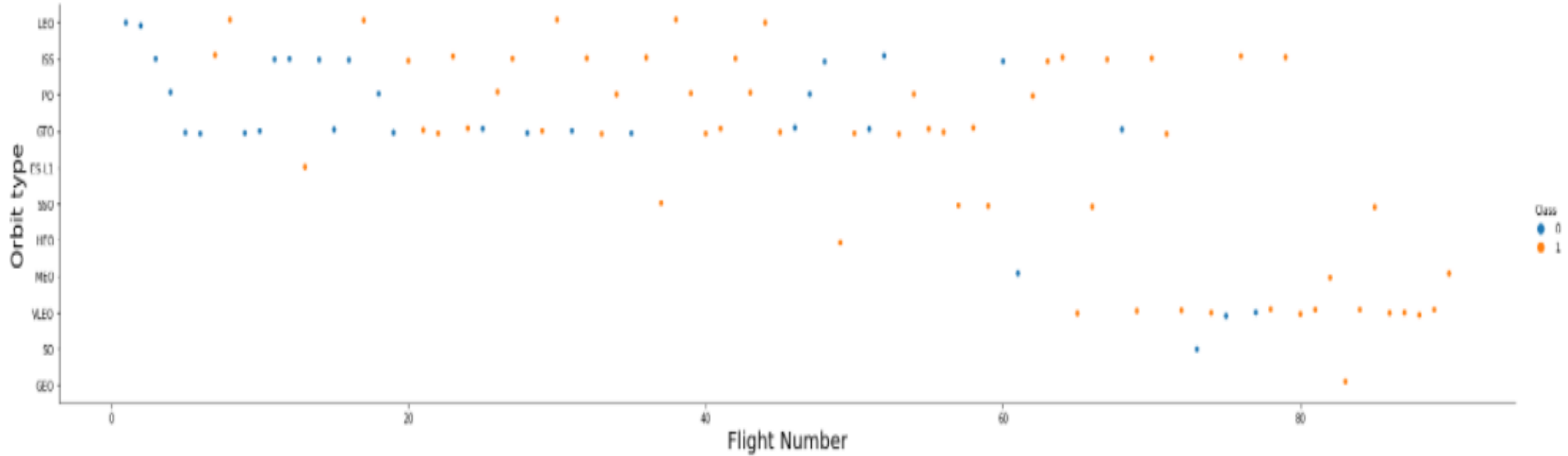
Payload vs. Launch Site

- Label 1 is successful outcomes.
- With increasing payload mass, we can notice less successful mission outcomes.

Success Rate vs. Orbit Type

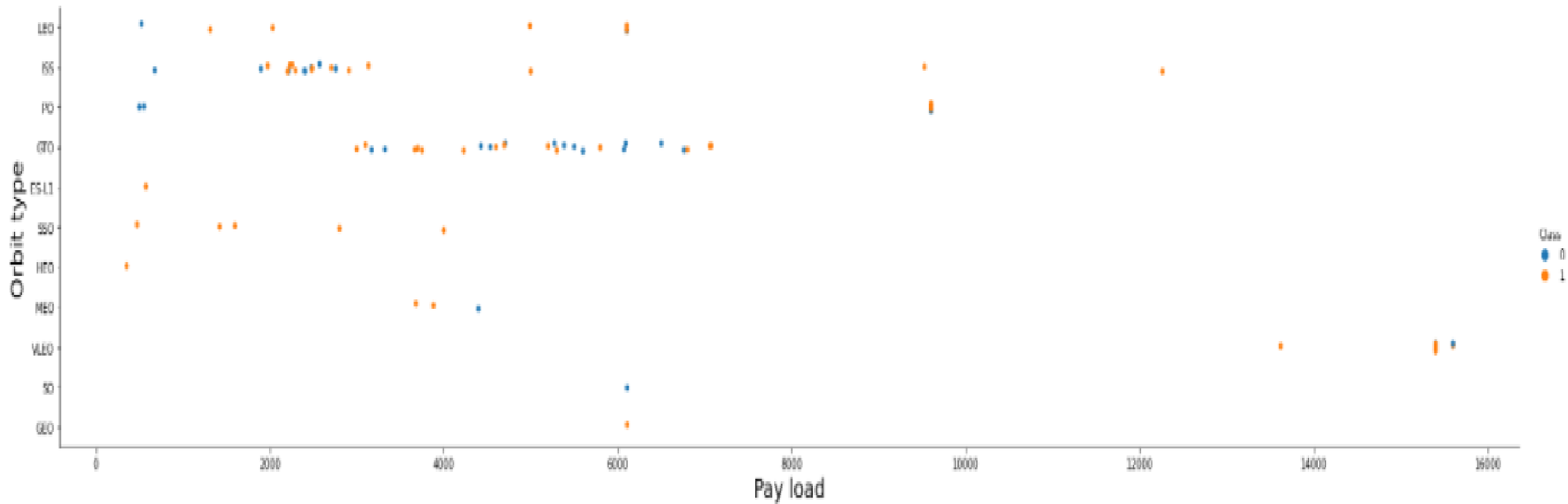
- ES-L1, GEO, HEO, SSO has highest success rate
- GTO has lowest success rate.





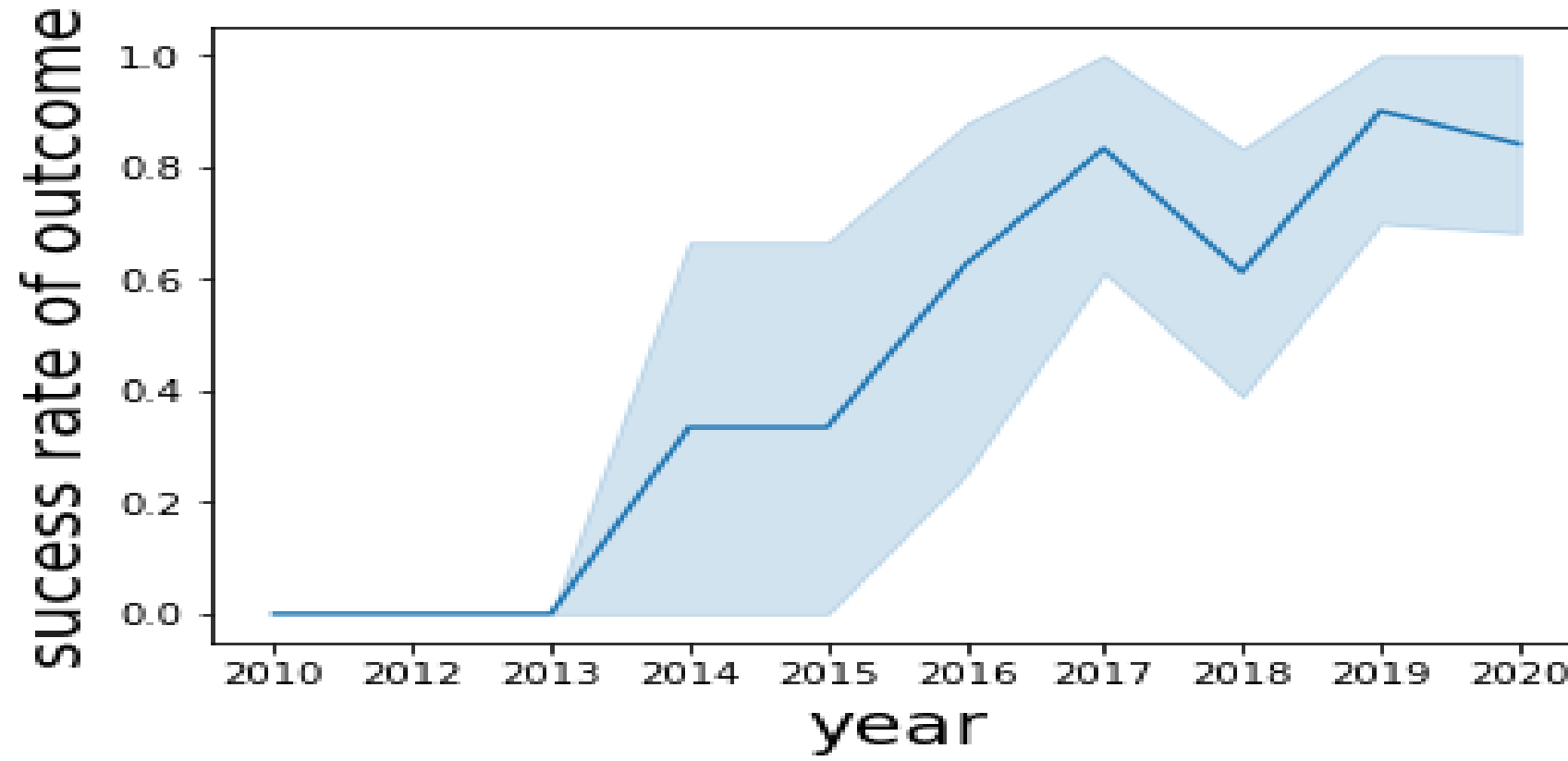
Flight Number vs. Orbit Type

- Label 1 is successful outcomes.
- VLEO, ISS, PTO orbit types has highest success with high flight numbers.
- LEO, ISS, PO, GTO has more success in medium flight numbers.



Payload vs. Orbit Type

- Label is for success outcomes.
- With increasing payload, orbits have less success.
- From 8k to 30k payload range only ISS,PO,VLEO, has success.
- GEO has most failures in 3k to 8k payload range.



Launch Success Yearly Trend

- From 2013 yearly success outcome rate kept increasing.

All Launch Site Names

- names of the unique launch sites
- Here are the launch site names with success & failure outcomes.

launch_site

CCAFS LC-40

CCAFS SLC-40

KSC LC-39A

VAFB SLC-4E

Launch Site Names Begin with 'CCA'

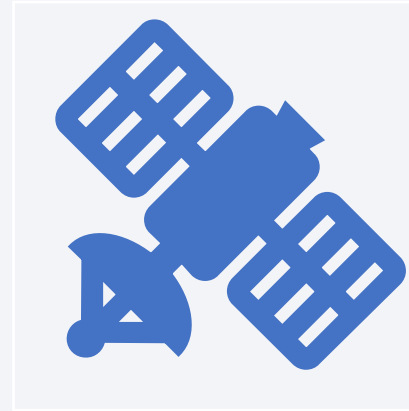
- Find 5 records where launch sites begin with `CCA`
- These are the records of launch sites with launch site name starts with CCA

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



Calculate the total payload carried
by boosters from NASA?



Total payload mass carried by
boosters from NASA is 45596

Average Payload Mass by F9 v1.1



Calculate the average
payload mass carried
by booster version F9
v1.1?

Average payload
mass carried by
booster version F9
v1.1 is 2928



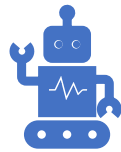
- The date of the first successful landing outcome on ground pad 2015-12-22

First Successful Ground
Landing Date

Successful Drone Ship Landing with Payload between 4000 and 6000



List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000?



These are the booster versions which have success on landing on drone ships with payload between 4000 & 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



Calculate the total number of successful and failure mission outcomes?



We have 99 success outcomes and 1 failure in flight and 1 success but not clear of payload.

COUNT	MISSION_OUTCOME
1	Failure (in flight)
99	Success
1	Success (payload status unclear)

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass?
- Here maximum payload is 15600 kg, and boosters carried this payload are F9 B5 B10 type versions

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



landing__outcome	booster_version	launch_site	DATE
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40	2015-01-10
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40	2015-04-14

2015 Launch Records

- List the failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015?
- F9 v1.1 b1012 and 1015 are the 2 versions failed in landing on drone ship from CCAFS LC-40 launch site in 2015.

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
- We have most outcomes are from no attempt after that drone ship for both failure & success.

landing__outcome	count
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

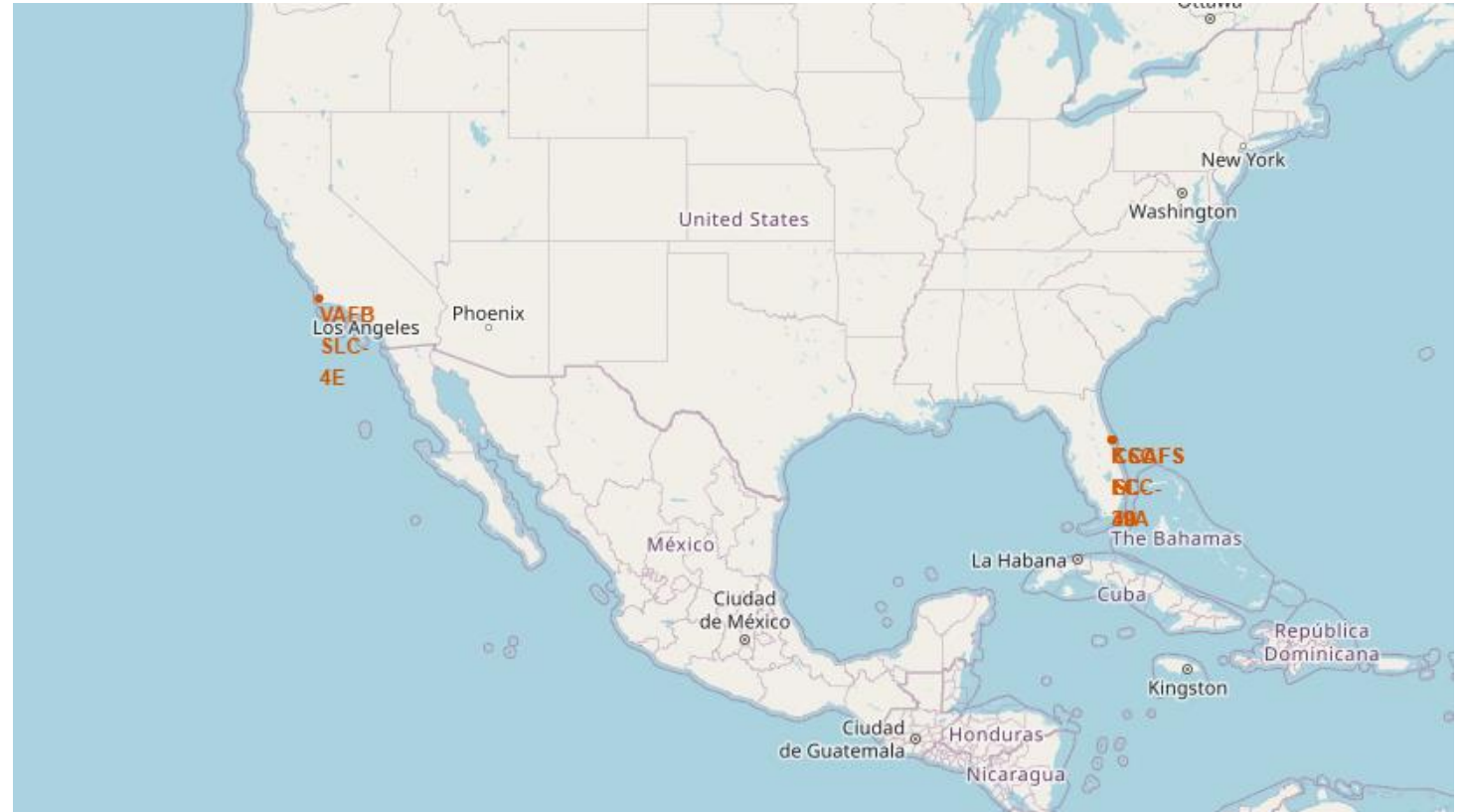
A satellite view of Earth from space, showing the curvature of the planet and city lights at night. The background is a deep blue gradient.

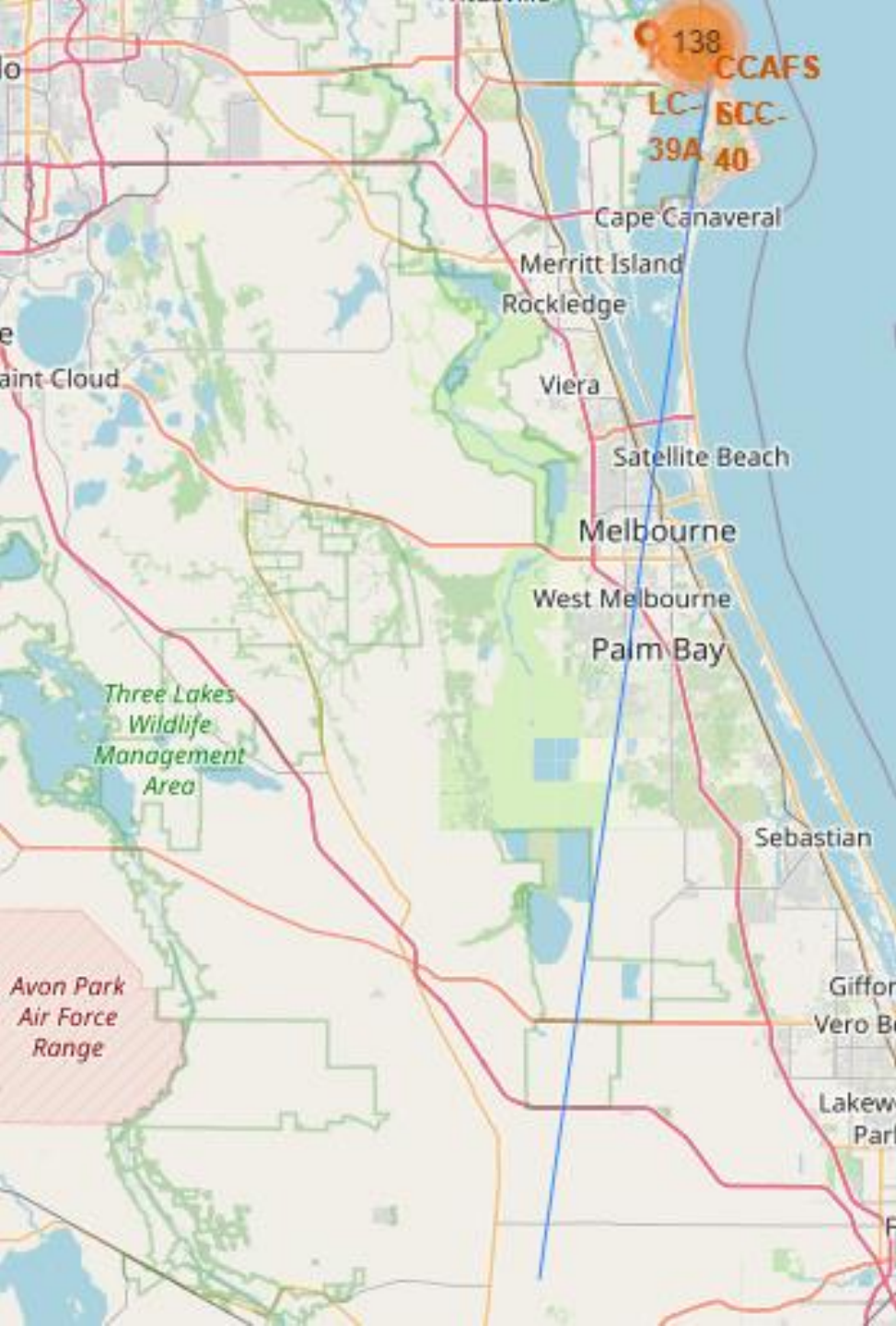
Section 3

Launch Sites Proximities Analysis

Marking of Launch sites on global map

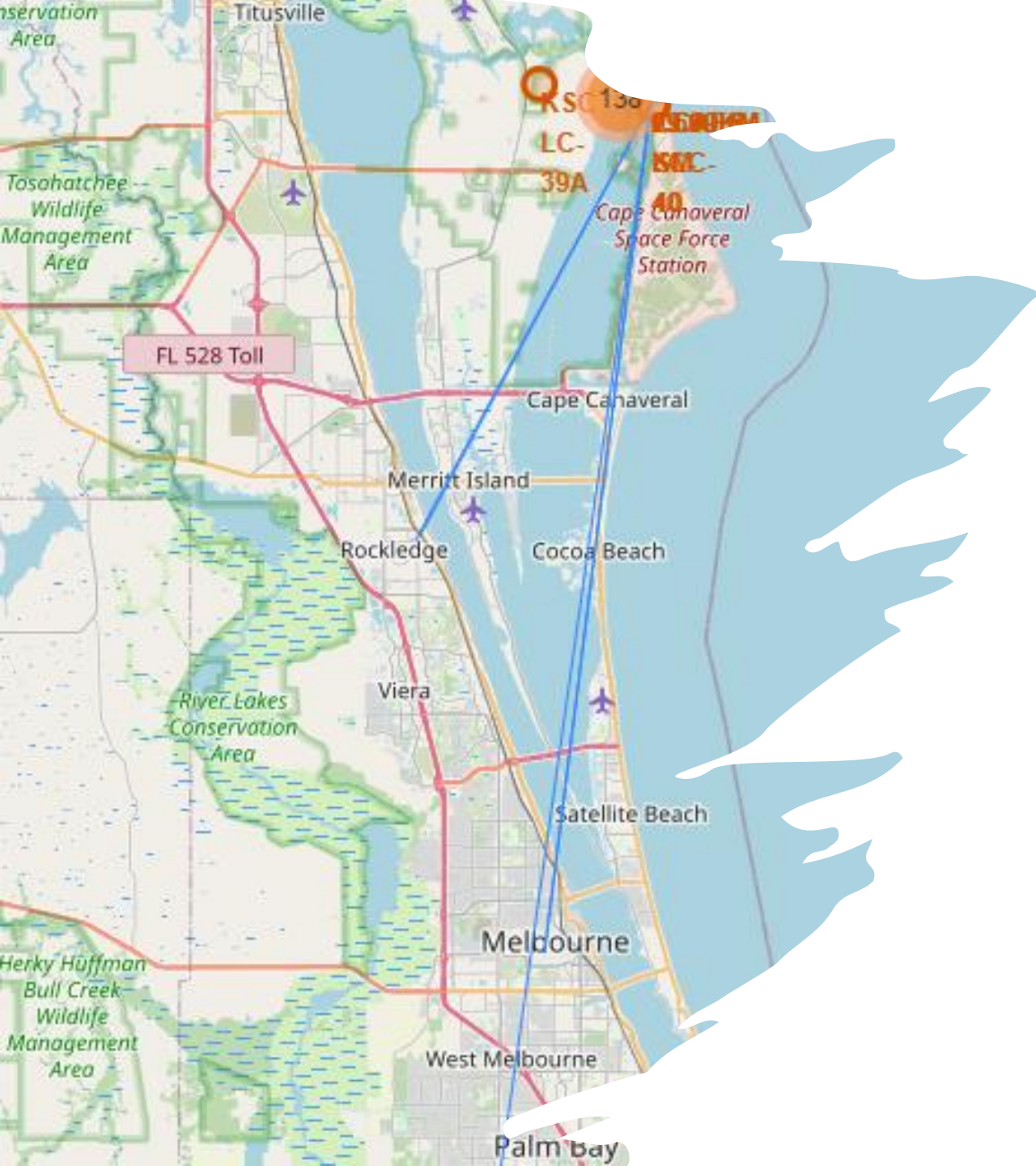
- Launch sites are located near coast regions
- Two launch sites are near to each other





Folium map with line

- We can draw line from launch sites to any location and can measure distance.



Folium map with launch site to its proximities

- This launch site is near to Melbourne city and near to railway and highway but very close to coastal line



Section 4

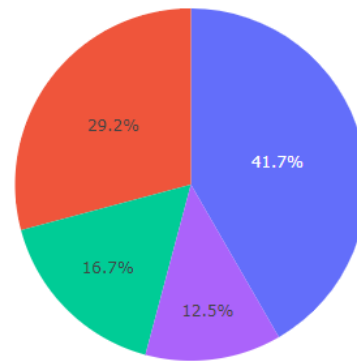
Build a Dashboard with Plotly Dash

SpaceX Launch Records Dashboard

ALL

x ▼

Success count for all launch site

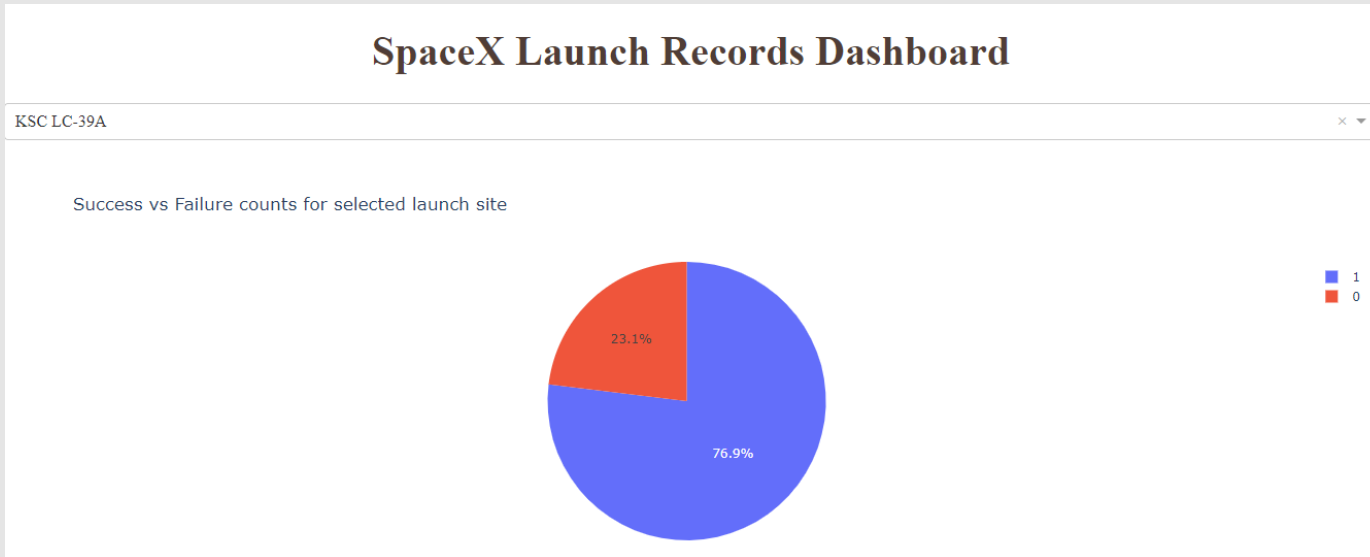


■ KSC LC-39A
■ CCAFS LC-40
■ VAFB SLC-4E
■ CCAFS SLC-40

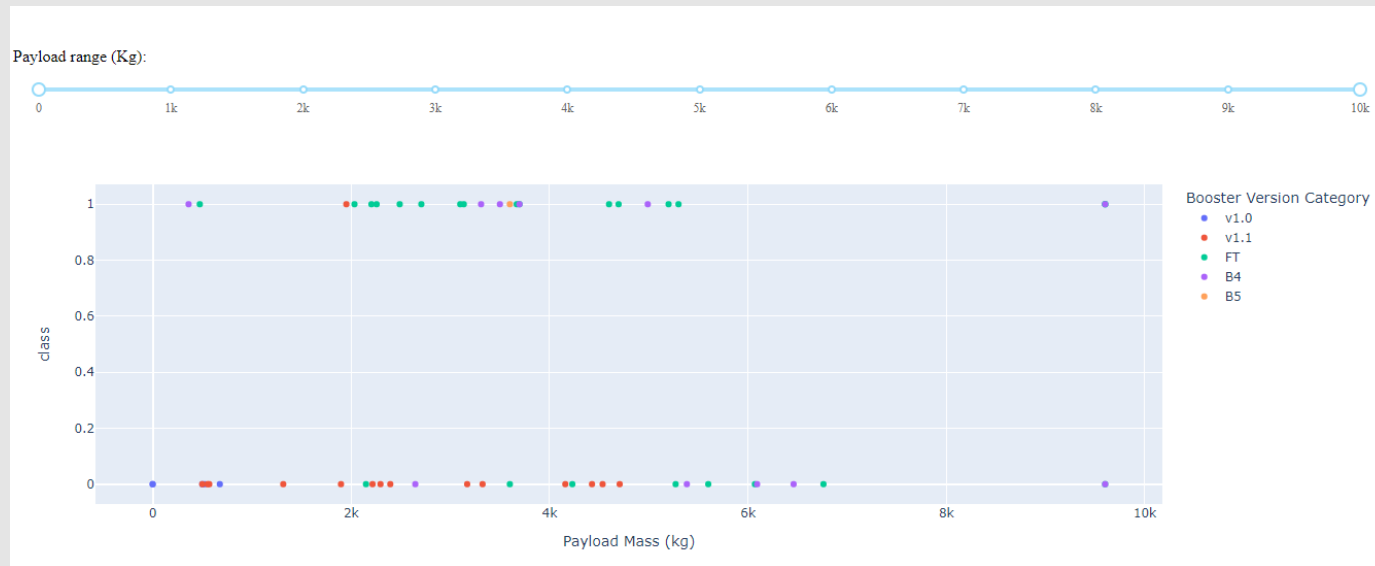
Launch success
count for all sites

- From pie chart we can see KSC LC-39A have high success count.
- CCAFS SLC-40 has low success count.

Launch site with highest launch success



- KSC LC-39A has highest success ratio with 76.9%
- In pie chart label 1 is success outcomes
- Label 0 is failed outcomes.



- Payload vs launch outcome scatter plot categorized by booster versions
- From 2k to 6k payload range have higher launch outcome.
- FT booster version have higher success rate
- 6k-9k has lowest success launch outcome.

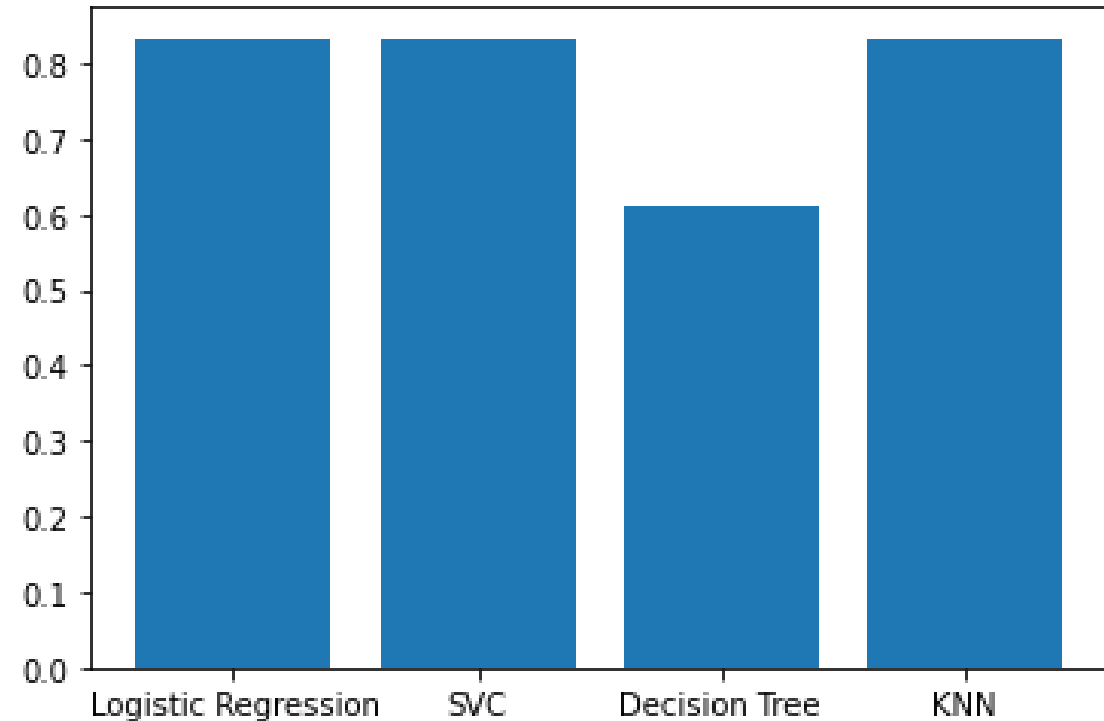
Scatter plot for all sites

Section 5

Predictive Analysis (Classification)

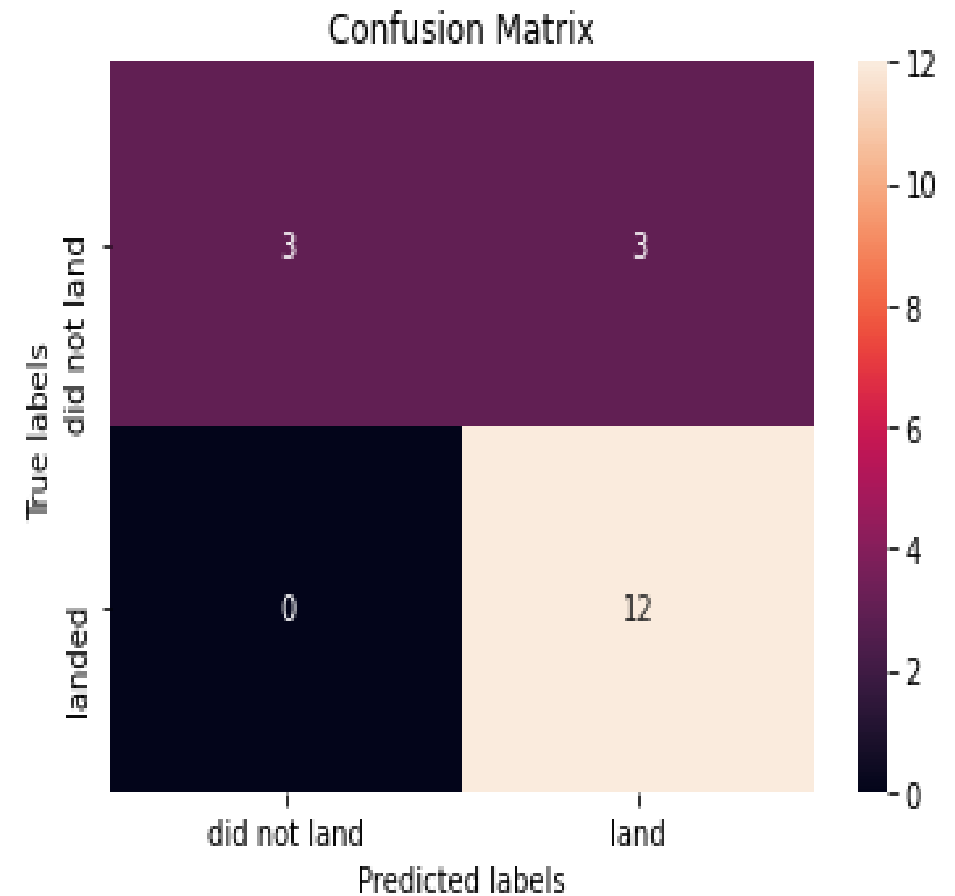
Classification Accuracy

- Logistic regression , SVC ,KNN shows highest accuracy with 83%.
- Decision tree has lowest accuracy with 60%



Confusion Matrix

- This is confusion matrix of logistic regression.
- Here false positives are not sure



Innovative insights



Payload mass inversely proportional to launch outcome after 8k.



Launch sites are near to coastal sides and far away from city and highways.



With increasing flight numbers only few orbit types have success launch outcomes.



Conclusions

By evaluating 4 models we got predicted accuracy as 83%.

We can predict launch of first stage will be 83% successful.

We got relation between payload and launch outcome

Here launch outcome is highly dependent on orbit type, launch site, payload mass.

With the help of these attributes , we can estimate cost of launch to bid against SpaceX.

Appendix



Relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that have created during this project in GitHub.



GitHub repository of [SpaceX](#) capstone project.



[SpaceX](#) dataset.

Thank you!

