IBM DATA SCIENCE PROJECT: SPACEX LAUNCH ANALYSIS

OCHIMANA MUSTAPHA

2-9-2023

OUTLINE

- Executive Summary
- Introduction
- Methodology
- Insights from E.DA
  - Visualization – Charts
  - Dashboard
- Launch site proximity
  - Findings & Implications
- Results
- Conclusion
- Appendix

EXECUTIVE SUMMARY

- Summary of Methodologies
  - Data collection
  - - Data wrangling
  - - EDA with data visualization
  - - EDA with SQL
  - - Building an interactive map with Folium
  - - Building a Dashboard with Plotly Dash
  - - Predictive analysis (Classification)
- Summary of Results
  - Exploratory data analysis results
  - Interactive analytics demo in screenshots
  - Predictive analysis results

INTRODUCTION

- Project background and context

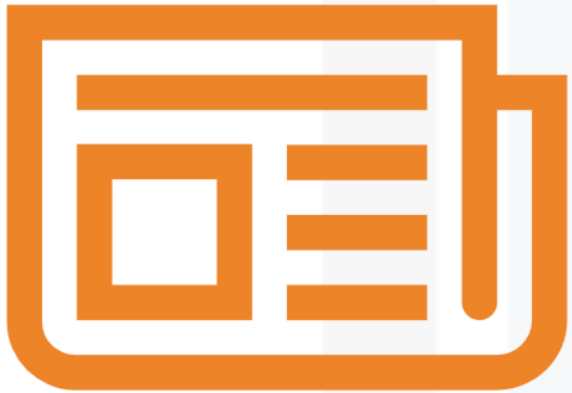
The landing outcome of the Falcon 9 first stage was predicted. SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage. Therefore, if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used by an alternate company that wants to bid against SpaceX for a rocket launch.

- Problem areas.

- What influences if the rocket will land successfully?
- The effect each relationship with certain rocket variables will impact in
- determining the success rate of a successful landing.
- What conditions does SpaceX have to achieve to get the best results

and ensure the best rocket success landing rate.

METHODOLOGY



- Data collection methodology:
  - SpaceX Rest API
  - (Web Scrapping) from Wikipedia
- Performed data wrangling (Transforming data for Machine Learning)
- One Hot Encoding data fields for Machine Learning and dropping irrelevant columns
- Performed exploratory data analysis (EDA) using visualization and SQL
  - Plotting : Scatter Graphs, Bar Graphs to show relationships between variables to show patterns of data.
- Performed interactive visual analytics using Folium and Plotly, Dash
- Performed predictive analysis using classification models

How to build, tune, evaluate classification models

EDA with Data Visualization

Data visualization helps us understand data by curating it into a form that's
easier to understand, highlighting the trends and outliers.
Several types of
charts were used in the visualization of the data:

- Cat plots and scatter plots were used to view the relationships of
categorical variables like Launch Site and Orbit.
- A bar chart was used to visualize the success rate of each orbit type.
- A line chart was used to visualize the launch success yearly trend.

EDA with SQL

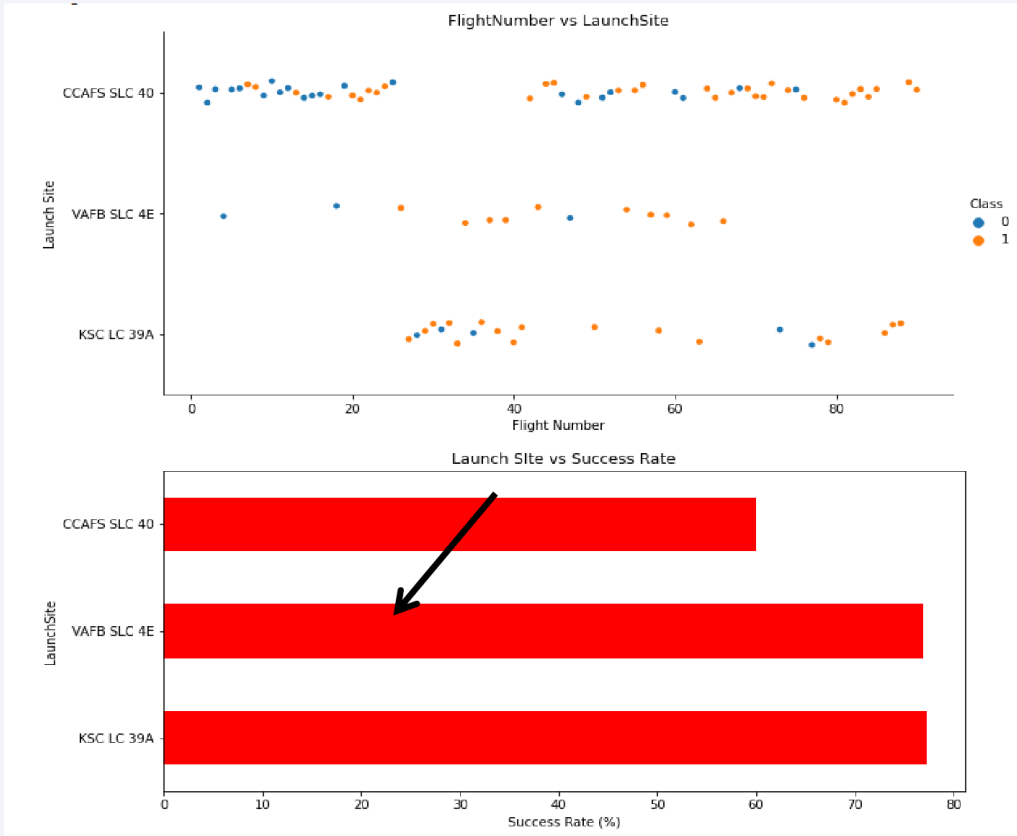
Summary of SQL queries that were used:

- Display the names of the unique launch sites in the space mission
- Compare the payload mass with boosters launched by NASA (CRS)
- Display average payload mass carried by booster version F9 v1.1
- List the total number of successful and failure mission outcomes
- Determine the dates of different landing outcomes

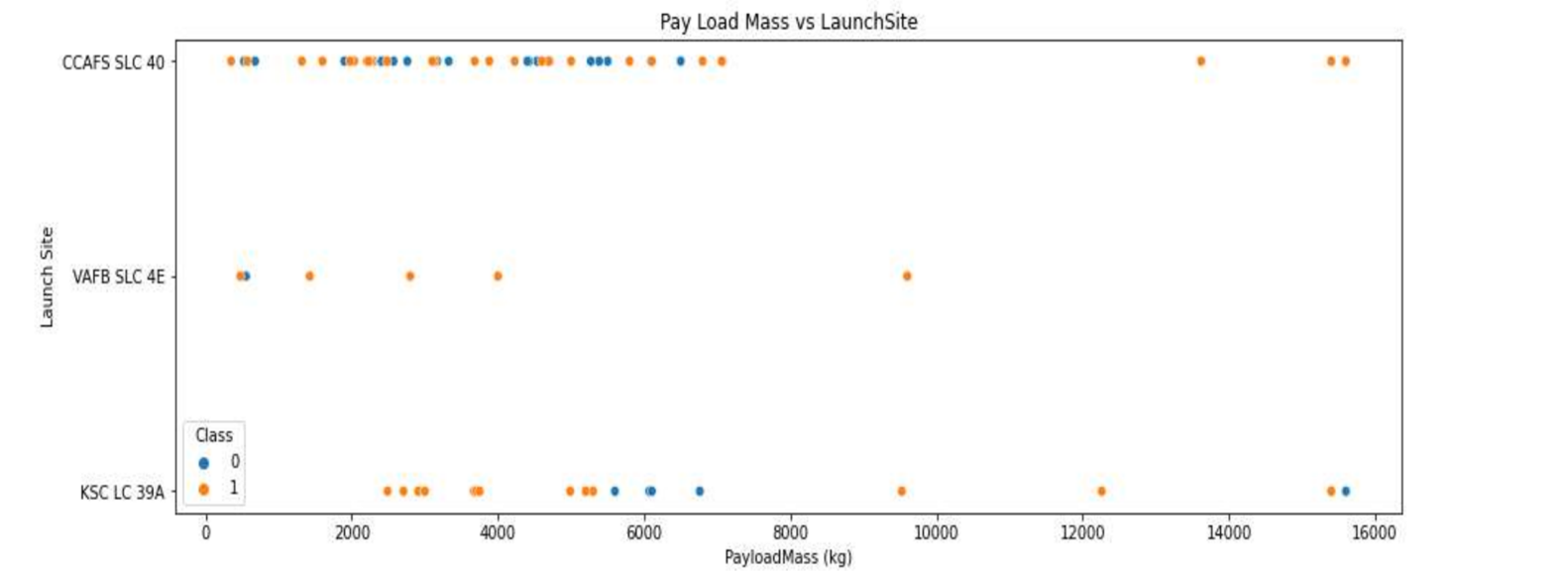
EDA with Visualization Results

Flight Number vs. Launch Site

- It appears that there were more successful landings as the flight numbers increased. It also seems that launch site **CCAFS SLC 40** had the most number of landing attempts while the site **VAFB SLC 4E** had the least number of attempts.
- Looking at the second chart, we can see that there is no Launch Site with a success rate below 60%.



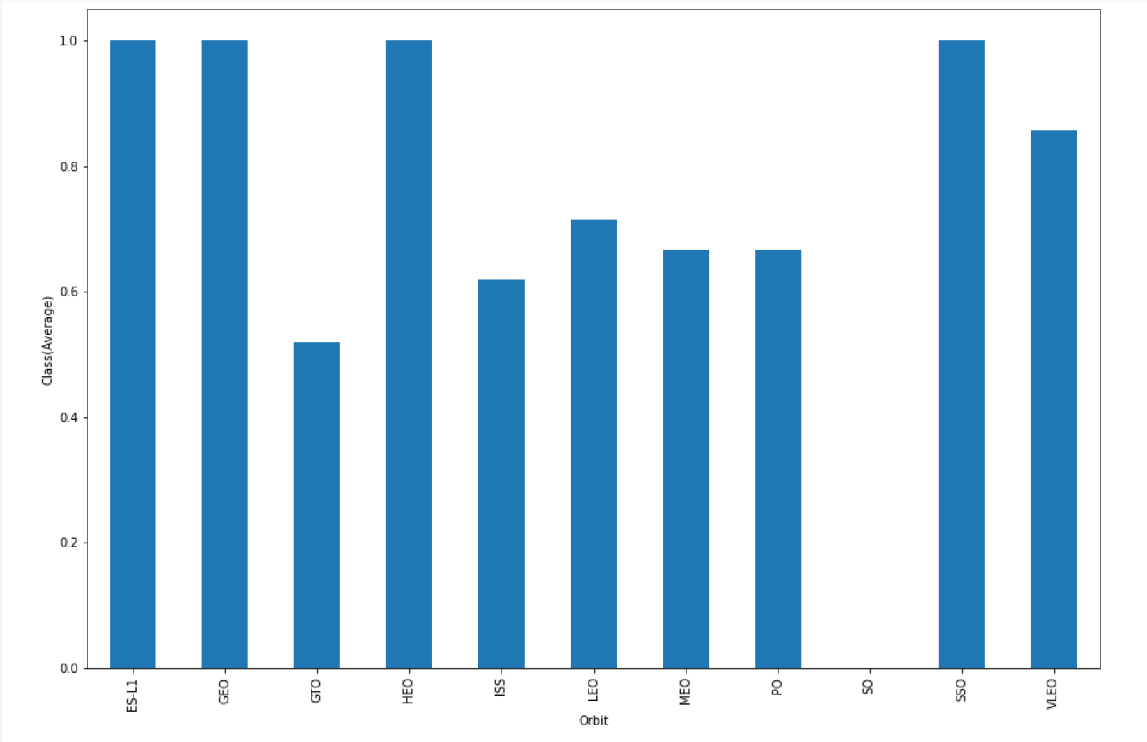
Payload vs. Launch Site



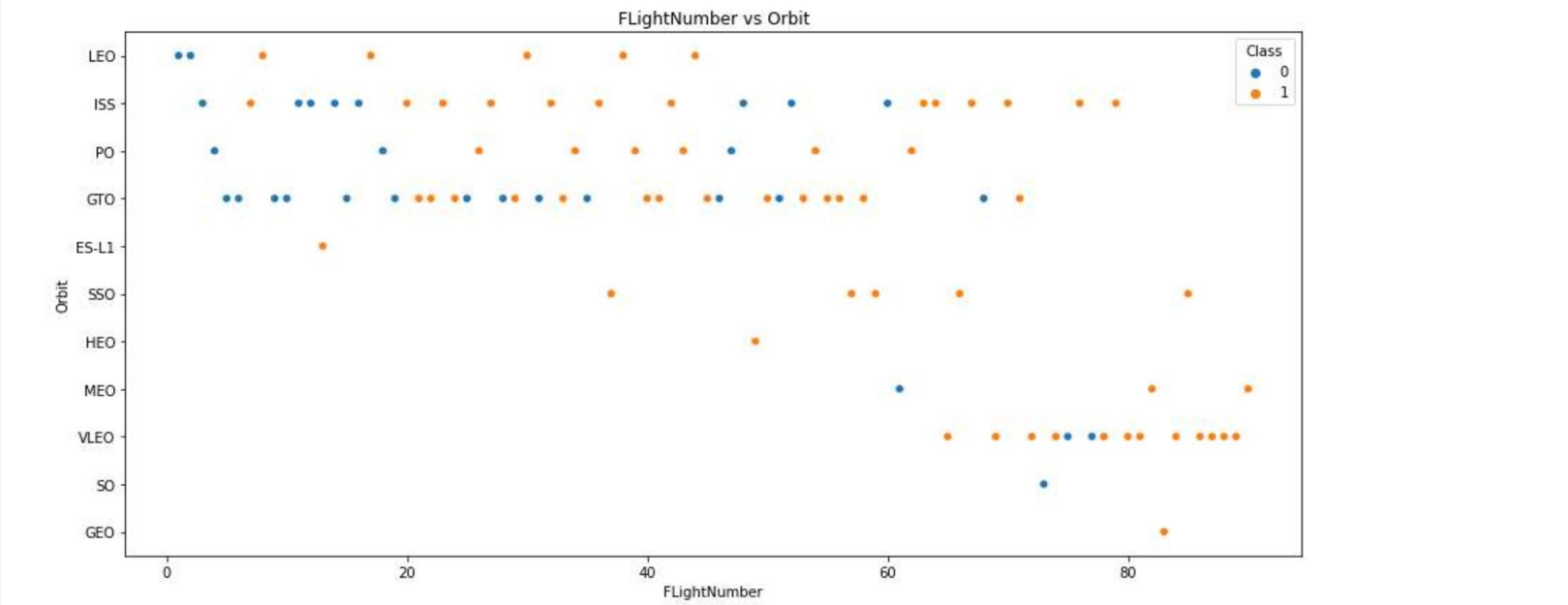
- Now if you observe the scatter point chart, you will find for the VAFB-SLC launch site there are no rockets launched for heavy payload mass(greater than 10000).

Success Rate vs. Orbit Type

The orbit types **SSO**, **HEO**, **GEO** and **ES-L1** had the highest success rate.

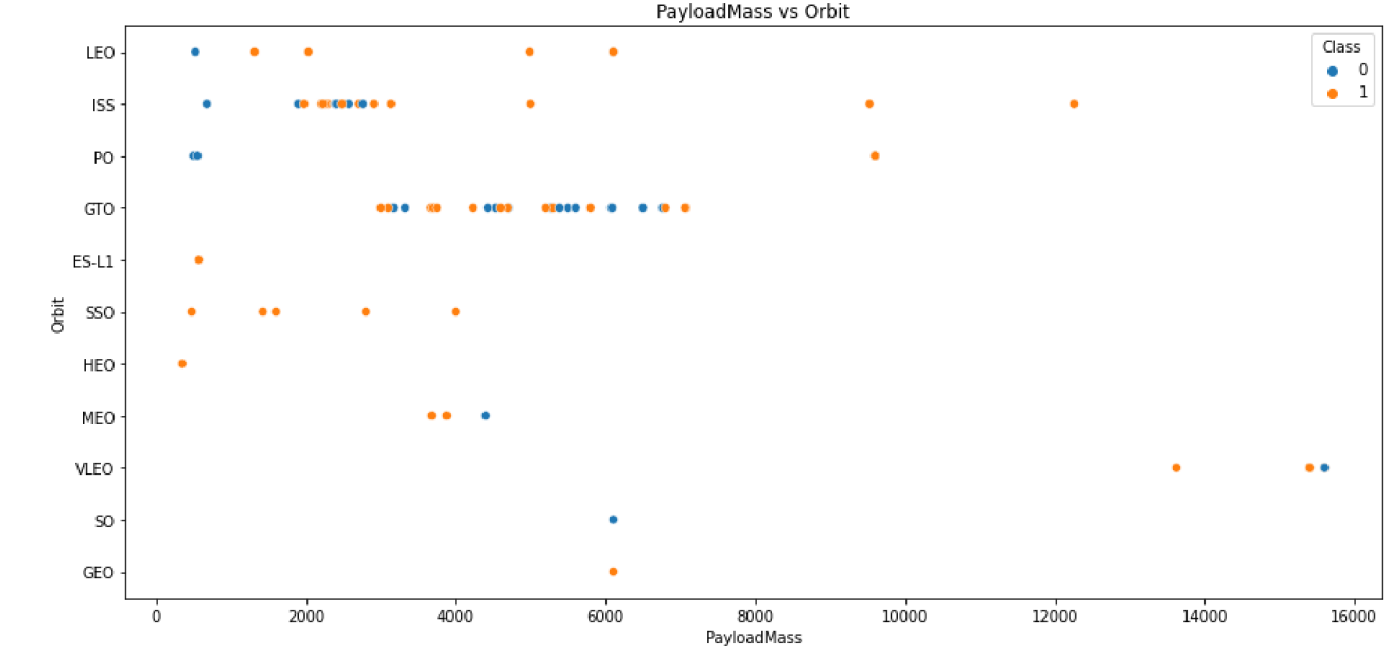


Flight Number vs. Orbit Type



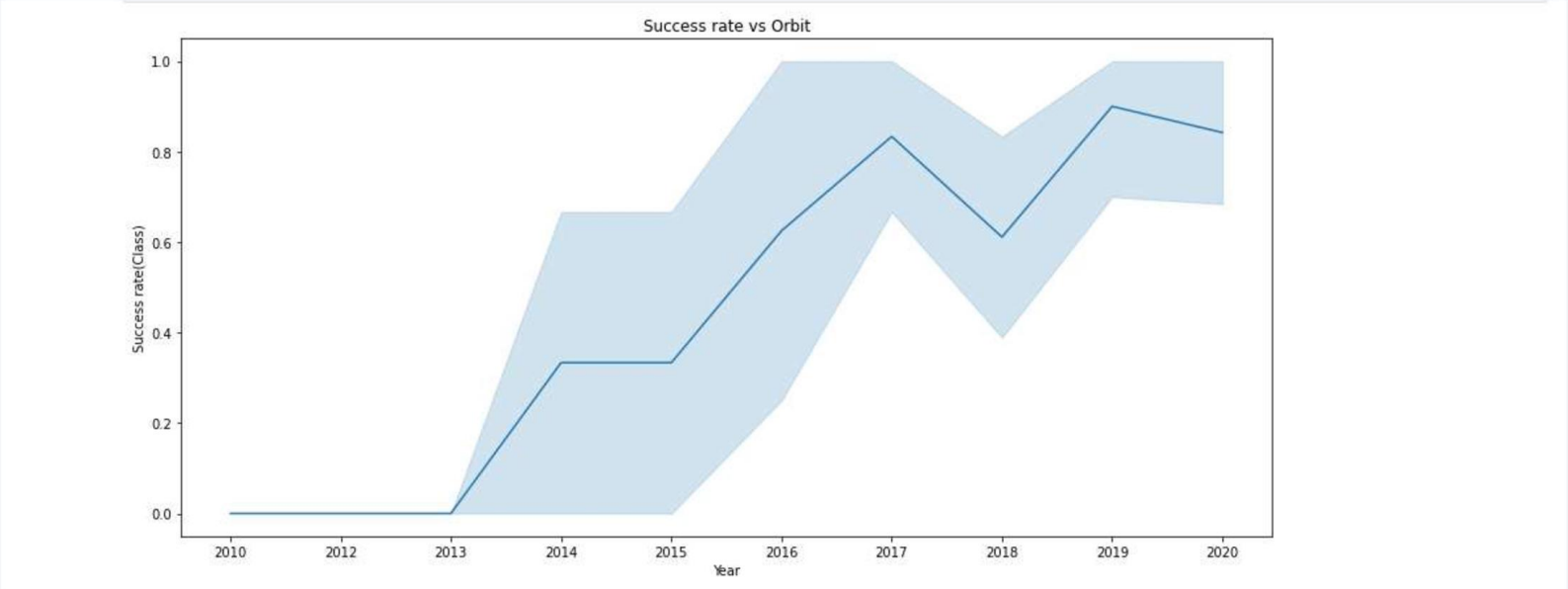
You can see that in the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

Payload vs. Orbit Type



- With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS.
- However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there.

Launch Success Yearly Trend



It is apparent that the success rate has significantly increased from 2013 to 2020.

EDA with SQL Results

All Launch Site Names

Given the data, these are the names of the launch sites where different rocket landings where attempted:

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

Launch Site Names Beginning with 'CCA'

Date	Launch_Site	Orbit	Customer	Mission_Outcome	Landing_Outcome
04-06-2010	CCAFS LC-40	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	CCAFS LC-40	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
22-05-2012	CCAFS LC-40	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	CCAFS LC-40	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	CCAFS LC-40	LEO (ISS)	NASA (CRS)	Success	No attempt

These are 5 records where launch sites begin with the letters 'CCA'. As we can see, there are other organizations besides SpaceX that were testing their rockets.

Total Payload Mass

- The information in the table displays the total payload mass carried by boosters launched by NASA .
- It seems that *NASA (CRS)* had a significantly higher total payload mass compared to the rest.

Customer	Total_Payload_Mass
NASA (CRS)	45596
NASA (CCDev)	12530
NASA (CCP)	12500
NASA (CCD)	12055
NASA (CTS)	12050
NASA (CRS), Kacific 1	2617
NASA / NOAA / ESA / EUMETSAT	1192
NASA (LSP) NOAA CNES	553
NASA (COTS)	525
NASA (LSP)	362
NASA (COTS) NRO	0

Average Payload Mass by F9 v1.1

Average_Payload_Mass (kg)	Booster_Version
2928.4	F9 v1.1

- The average payload mass carried by F9 v1.1 was 2928.4 kg.

First Successful Ground Landing Date

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

- The first successful ground pad landing took place in December 2015. This was a historic reusable-rocket milestone for both SpaceX and the world.
- Prior to this, no one had ever brought an orbital class booster back intact.

Successful Drone Ship Landing with Payload between 4000 and 6000

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)

- It appears that there only 4 Boosters with a payload mass between 4000 and 6000.
- It is interesting to see that they all had successful landing outcomes.

Total Number of Successful and Failure Mission Outcomes

Mission_Outcome	Outcomes
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

- It appears that missions generally tend to be successful with the exception of one failure.

Boosters That Carried the Maximum Payload Mass

- 12 boosters have carried the maximum payload mass of 15600 kg.
- Since the version names are similar, they might be from the same manufactures.

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 - Failed Landing Outcomes

Date	Launch_Site	Booster_Version	Landing_Outcome
10-01-2015	CCAFS LC-40	F9 v1.1 B1012	Failure (drone ship)
14-04-2015	CCAFS LC-40	F9 v1.1 B1015	Failure (drone ship)

- It appears that 2 boosters failed to land at the beginning of the year..
- The first successful landing took place later that year in December as we saw earlier.

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

- If we observe the table, it is apparent that the number of successful landings have increased since 2015.
- Before 2013, it seems that there were no attempts to land the boosters.

date	Landing_Outcome	Outcomes
2016-04-08	Success (drone ship)	14
2015-12-22	Success (ground pad)	9
2015-06-28	Precluded (drone ship)	1
2015-01-10	Failure (drone ship)	5
2014-04-18	Controlled (ocean)	5
2013-09-29	Uncontrolled (ocean)	2
2012-05-22	No attempt	22

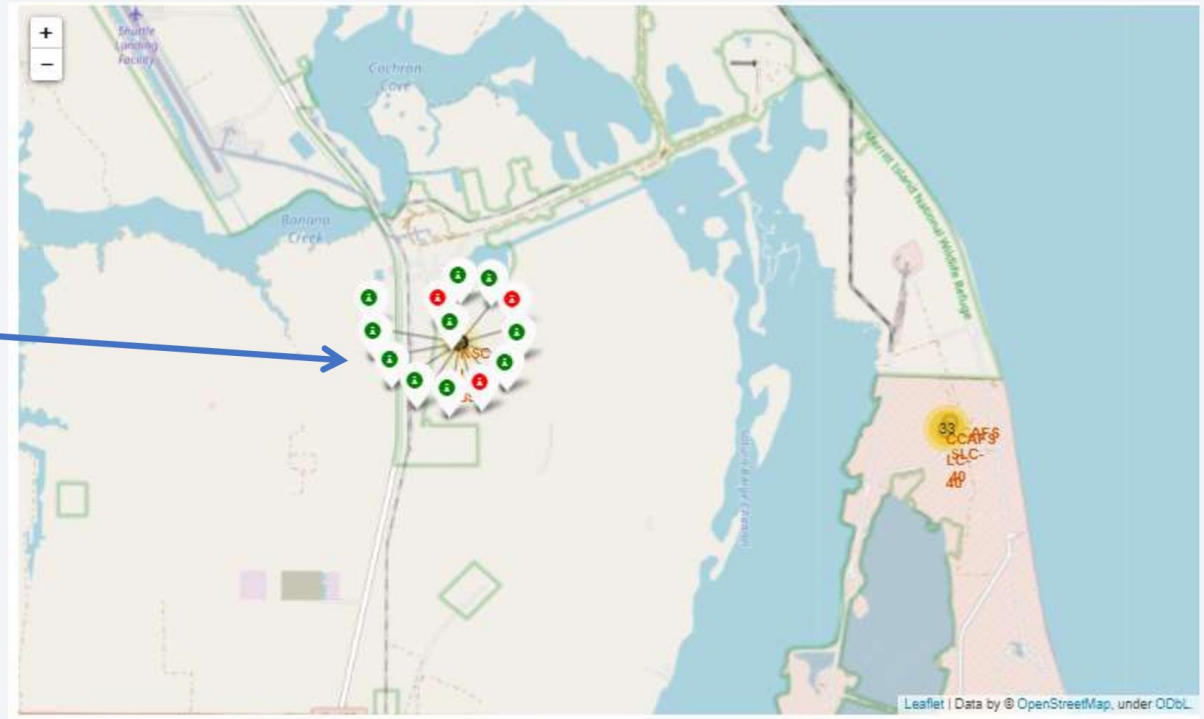
Interactive map with Folium

Launch Site Locations

- We can see that all launch sites are in very close proximity to the coast and they are also a couple thousand kilometers away from the equator line.
- It is interesting to see that most launch sites are concentrated near Miami.

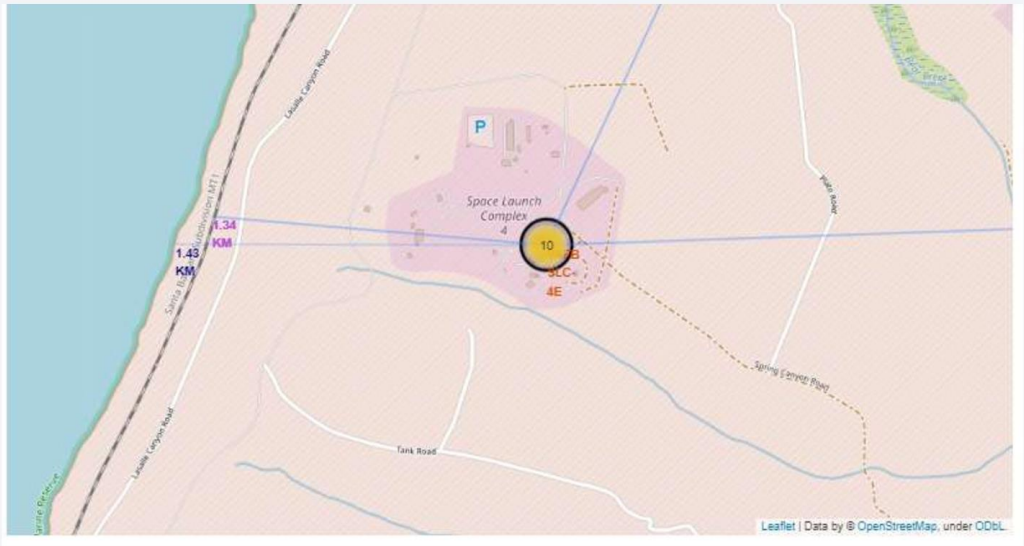
Success Rate of Rocket Launches

- The successful launches are represented by a green marker while the red marker represents failed rocket launches.
- It appears that **KSC LC-39A** had the highest success rate of rocket launches compared to other launch sites.



Surrounding Landmarks

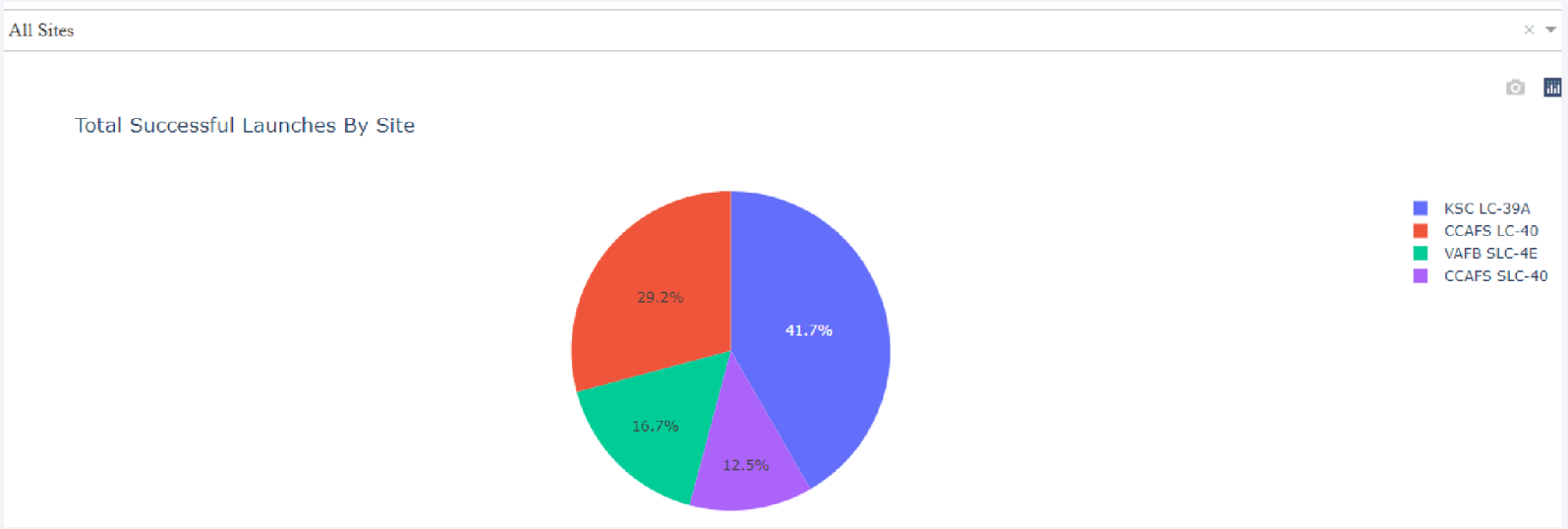
- It appears that launch sites are usually set up at least 18 km away from cities. This may be because of the desire to prevent any crashes near populated areas.
- It is also apparent that launch sites are in very close proximity to railways and highways. Perhaps, due to the necessary transportation requirements for rocket parts.
- The sites are close the coast line. This is evident with the many rocket landing tests on water bodies like the ocean.



Map Object	Colour
Nearest Highway	Green
Nearest Railway	Purple
Nearest City	Crimson
Nearest Coastline	Dark Blue

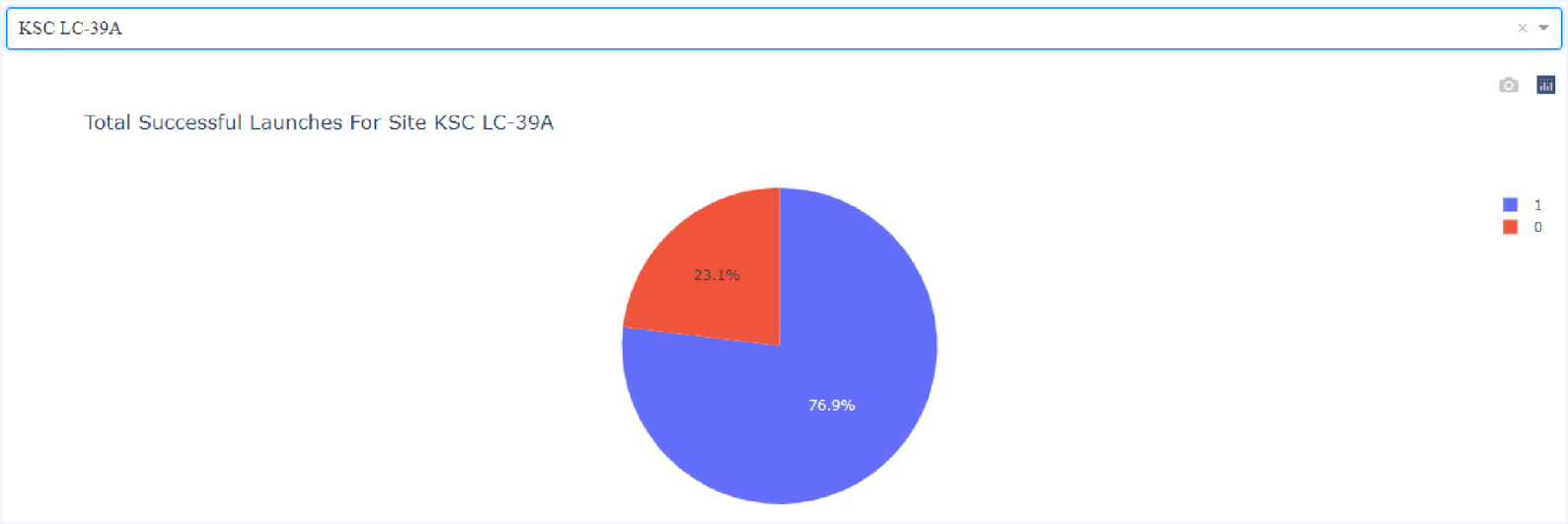
Plotly,Dash dashboard results

Successful Launches by Site



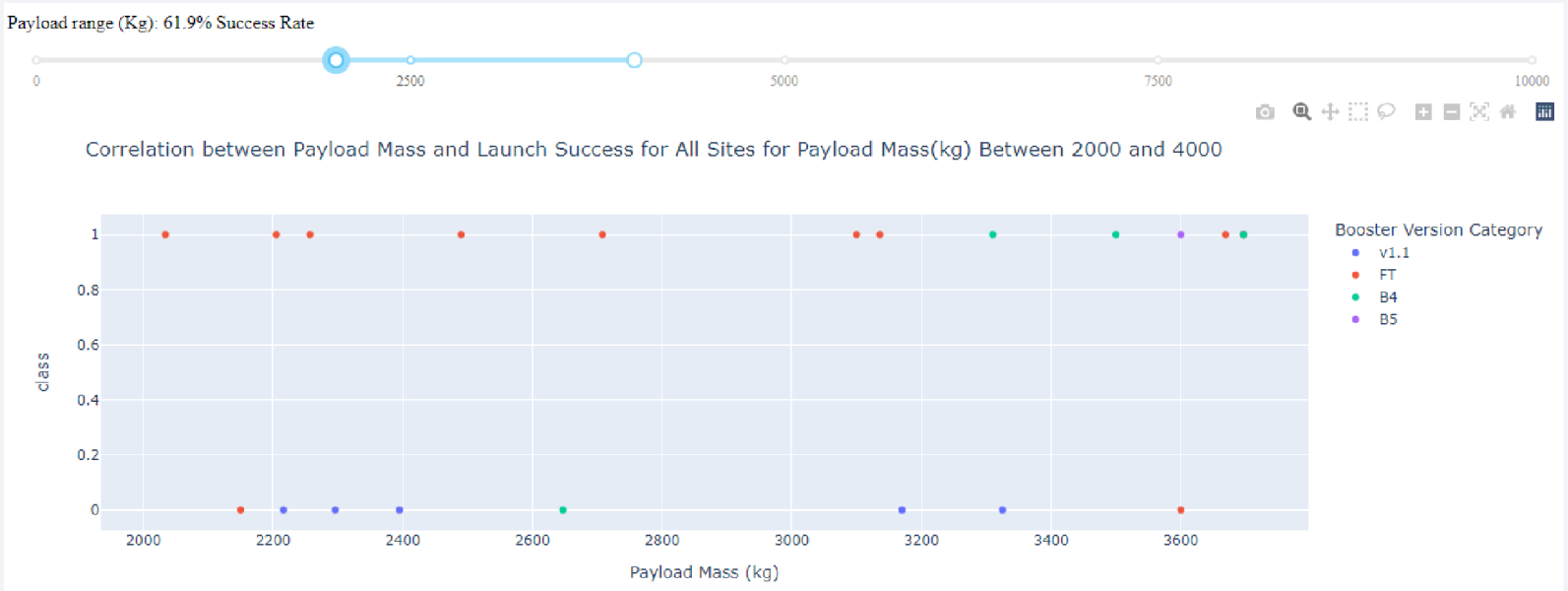
- Site KSC LC-39A has the largest successful launches as well the highest launch success rate.
- More investigation may be needed to determine why KSC LC-39A is the preferred launch site.

Total Successful Launches for Site KSC LC-39A



- As we can see, 76.9% of the total launches at site KSC LC-39A were successful. This is a the highest success rate of all the different launch sites.
- However, this success rate was only around 3% higher than the runner up; site CCAFS LC-40.

Payload Mass vs. Launch Success for All Sites

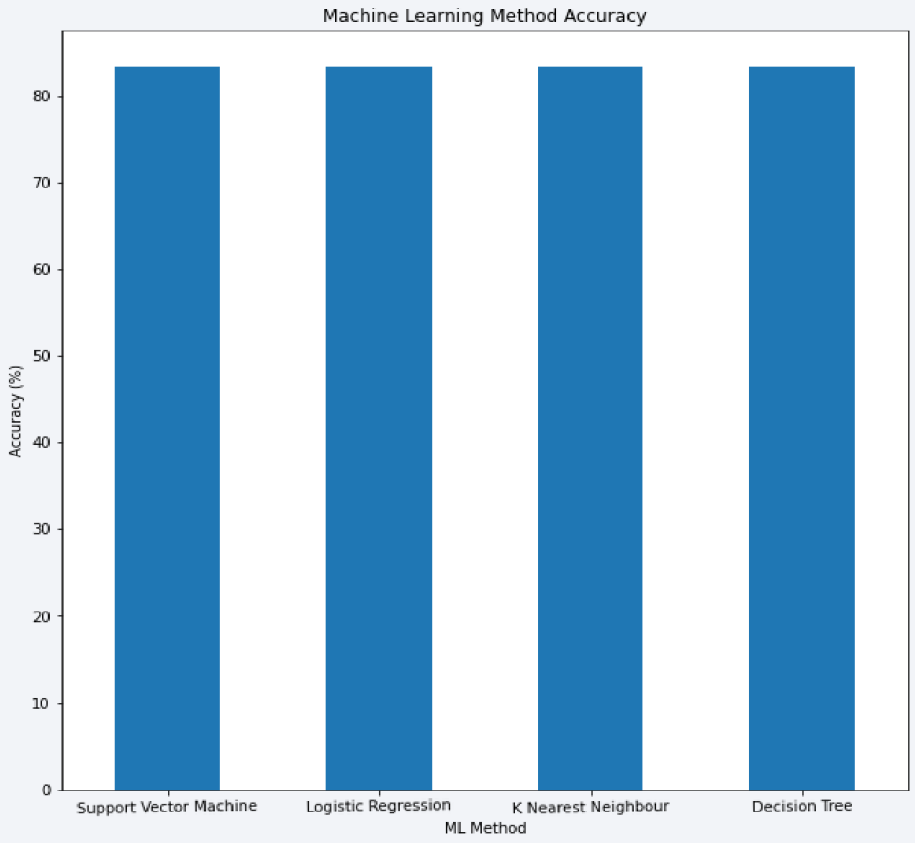


- It appears that the payload range between 2000 kg and 4000 kg has the highest success rate.
- The launch success rate was also dramatically low between the payload range of 0kg and 2500kg. Perhaps very low masses decrease launch success.
- The booster version **FT**, seems to have a higher success rate than other booster versions

Predictive Analysis (Classification)

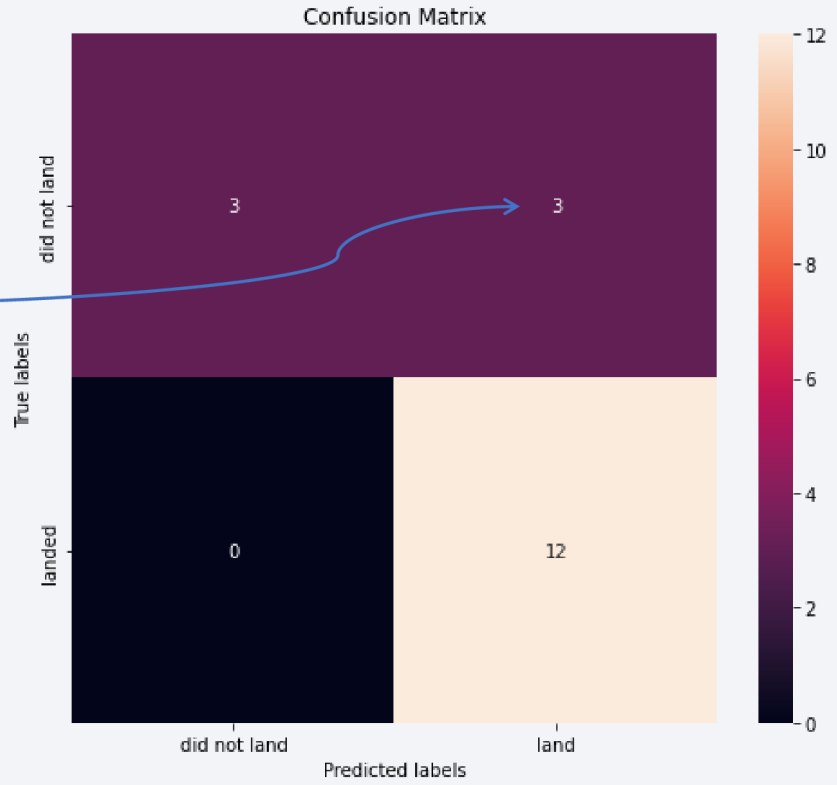
Classification Accuracy

- Since all the methods have an identical accuracy score of 83.33%, we decided to use Logistic Regression for the classification



Confusion Matrix

- The chart shows the confusion matrix of the Logistic Regression model that was chosen.
- The model only failed to accurately predict 3 labels.



Conclusions

In order to compete with SpaceX, it was crucial to analyze their data. Through this process, a general picture of their success methods was produced.

- All their launch sites are located near the coast, away from nearby cities. This enabled to them to test their rocket landings without much interference.
- Site KSC LC-39A had the highest launch success rate out of all the launch sites.
- From 2015 onwards, the success rate of rocket landings significantly increased. It was also apparent that landing success increased with flight number

All this data was used to train a machine learning model that is able to predict the landing outcome of rocket launches with 83.33% accuracy.

This will allow our company to make more attractive offers than SpaceX and increase the confidence of our investors and customers. Can anyone say “No” to a company that can predict the success of their product?

Thank you