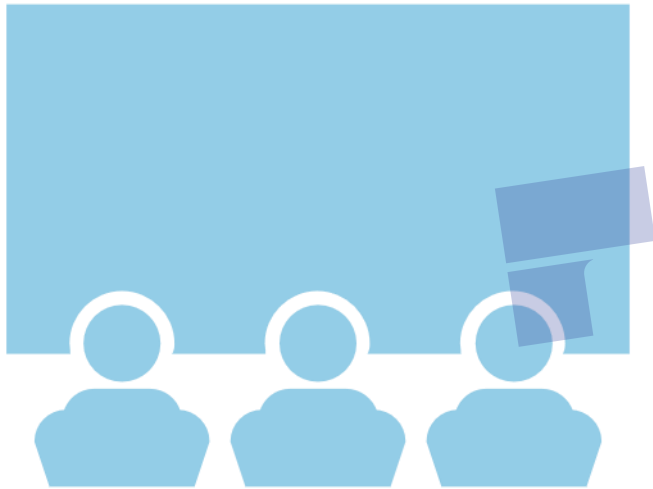


# Data Science Capstone project

---

# Outline

---



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

# Executive Summary

---



- Summary of methodologies

Data Collection

Data Wrangling

EDA with data visualization

EDA with SQL

Interactive visual analytics using Folium and Plotly Dash

Predictive analysis using classification models

- Summary of all results

All 4 machine learning models  
produced the same prediction

# Introduction

---



- Project background and context

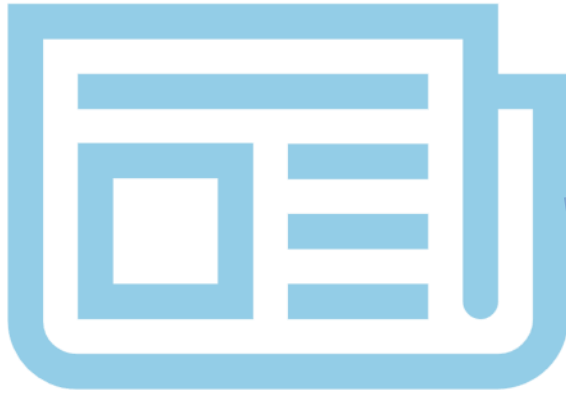
Space X advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because Space X can reuse the first stage. Therefore, if we can determine if the first stage will land, we can determine the cost of a launch

- Problems you want to find answers

What criteria determines the success of landing of the first stage of the rockets.

# Methodology

---



- Data collection methodology:
  - The data was collected through SpaceX API and the scrapped data from Wikipedia
- Perform data wrangling
  - Convert the data with for machine learning and clean(drop) irreverent columns and rows.
- Perform exploratory data analysis (EDA) using visualization and SQL  
Using various visualization tools for the data
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate using the various classification models

# Methodology

# Data collection

- The data was collected through the SpaceX API and web scraping (Wikipedia).

## API

- The SpaceX Rest API endpoints or URI, starts with `api.spacexdata.com/v4/`
- Performed a GET request using the requests library to obtain the data

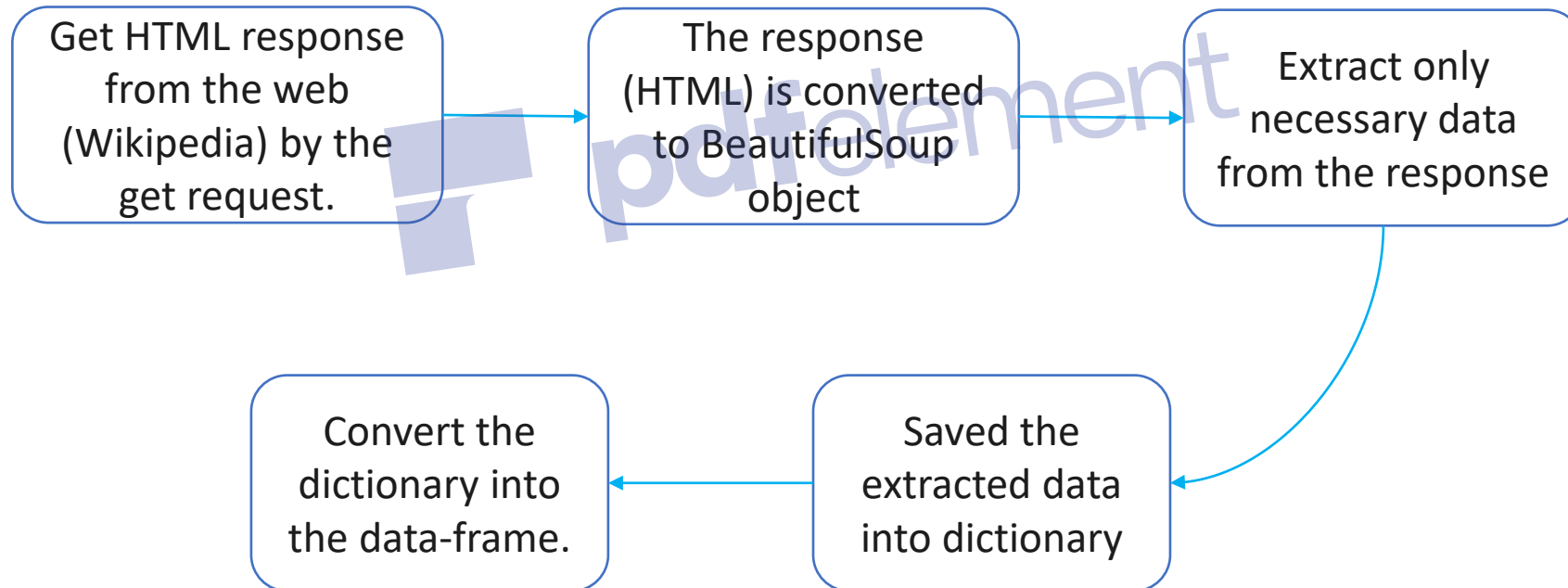
- After collecting the data, it is saved as .csv file to be processed.
- The data consists of the information about launches, component of rocket and its type, and landing/launching outcomes.

## A flowchart of SpaceX API call





## flowchart of web scraping



# Data wrangling

---

- In the data set, there are different cases of whether the booster successfully landed or not. To know each case is success or failure, the outcomes are converted into training labels with `1` means the booster successfully landed `0` means it was unsuccessful and appended to the data set.
- If the column value is numeric and it has a null, then the null is filled with mean value of the whole column values.

# EDA with data visualization

---

- Bar chart
  - Success rate vs. Orbit type
  - Flight Number vs. Orbit type
  - Payload vs. Orbit type
- Scatter plot
  - Flight Number vs. Launch Site
  - Payload vs. Launch Site
- Line chart
  - Launch success yearly trend
- Each chart has its own advantage and different charts were plotted for different purpose. These were used to convey the meaning of the data in the most effective way.

# EDA with SQL

---

- Data queried by SQL

- All launch site names
- Launch site names begin with `CCA`
- The total payload carried by boosters from NASA.
- The average payload mass carried by booster version F9 v1.1
- Date of the first successful ground landing
- Successful drone ship landing with payload between 4000 and 6000
- Total number of successful and failure mission outcomes
- Boosters which carried maximum payload
- 2015 launch records
- Rank success count between 2010-06-04 and 2017-03-20

# Build an interactive map with Folium

---

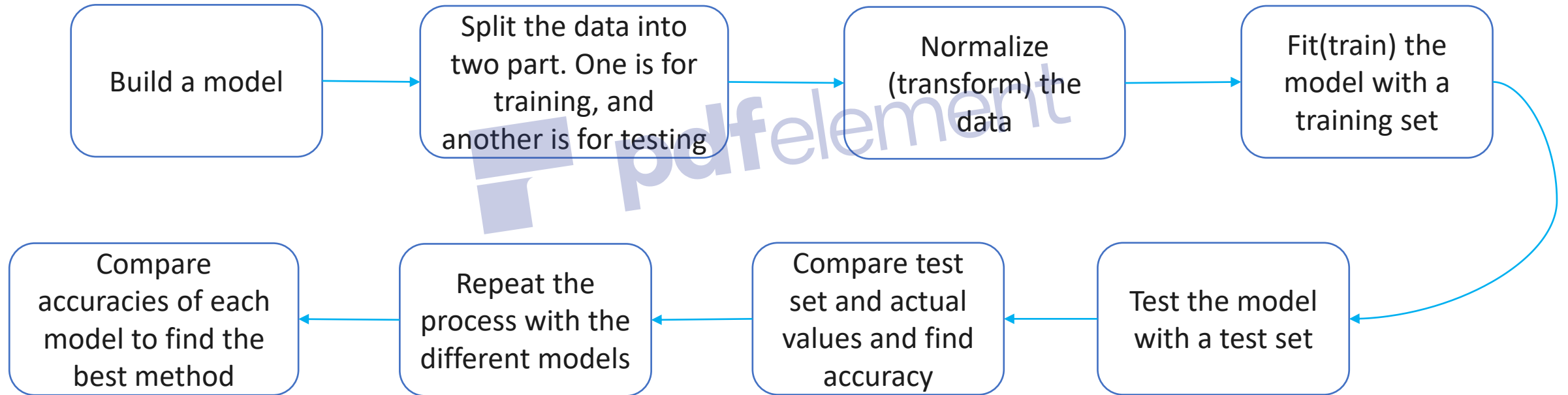
- **Markers**
  - 4 markers to indicate the name of the launch sites
  - 1 markers to indicate the distance between the launch site and the nearby railway.
- **Circles**
  - 4 circles to indicate the location of the launch sites
- **Icon**
  - Numerous icons to indicate the number of succussed and failed launch for each launch site.
- **line**
  - 1 line to indicate the linear line between the launch site and the nearby railway.

# Build a Dashboard with Plotly Dash

---

- Launch Success Count for All Sites with pie chart
- The Launch Site with The Highest Success Rate with pie chart
- Payload vs Launch Outcome with scatter plot

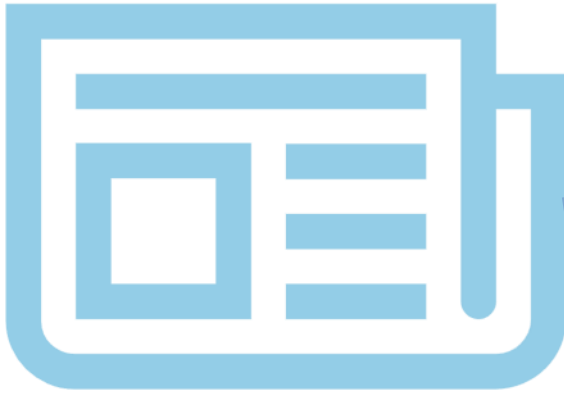
# Predictive analysis (Classification)



# Results

---

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





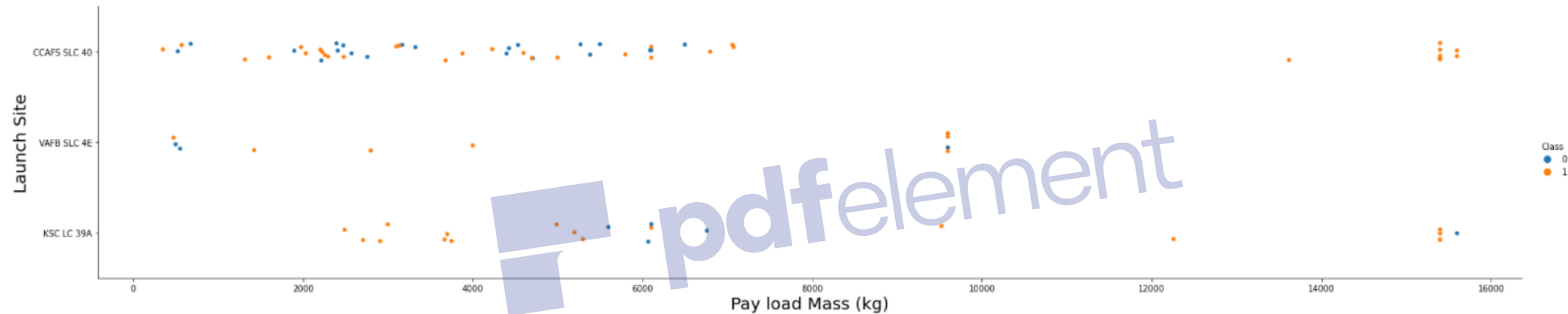
# EDA with Visualization

# Flight Number vs. Launch Site



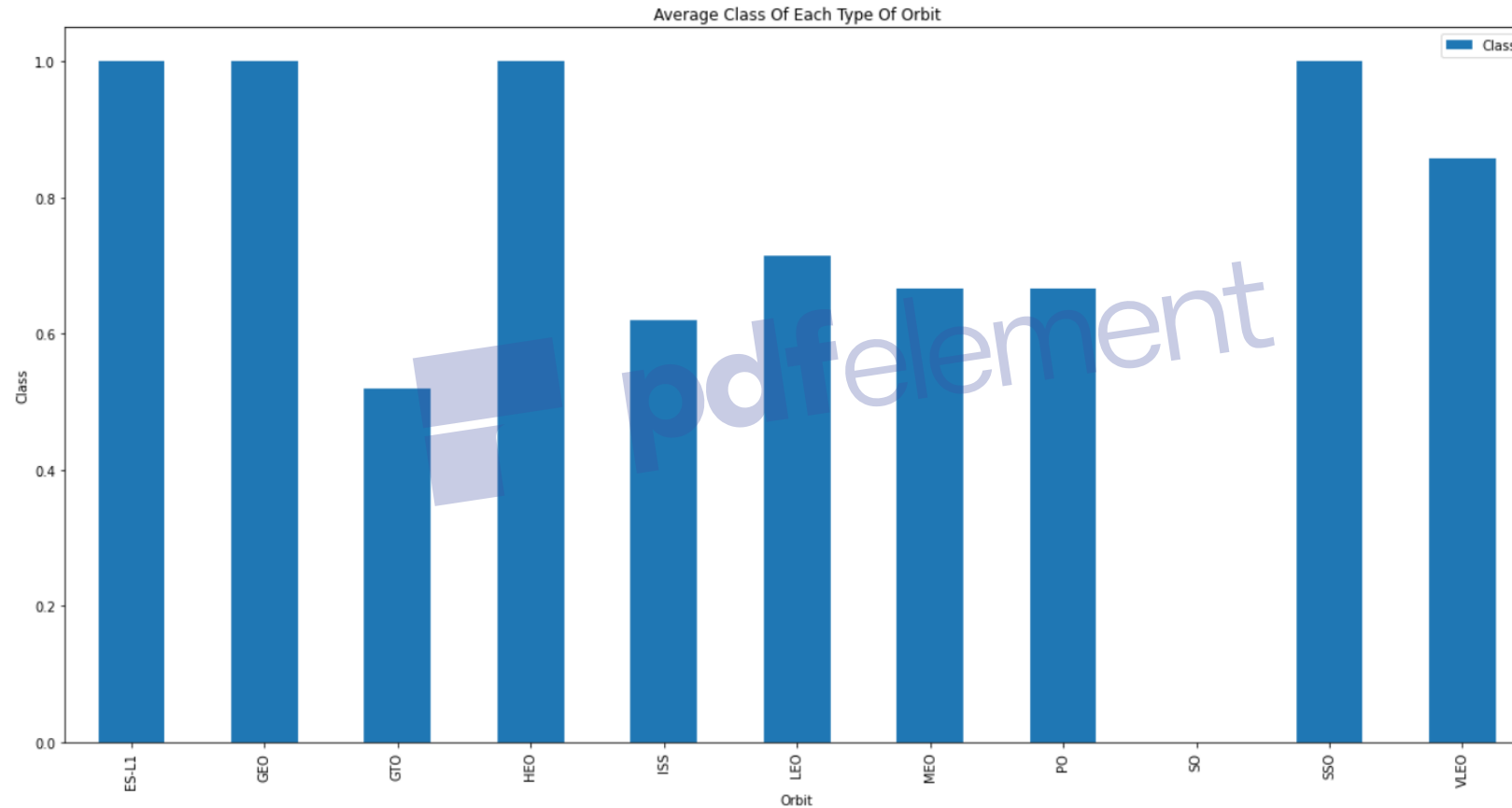
It is a scatter plot of Flight Number vs Launch Site. The blue dot indicates failure and the orange dot indicates success of landing of first stage of the Falcon 9. As flight number increases (more attempts), more orange dots are showing. On the other hand, the failure of landing is not related to the launch sites since it has no distinctive patterns.

# Payload vs. Launch Site



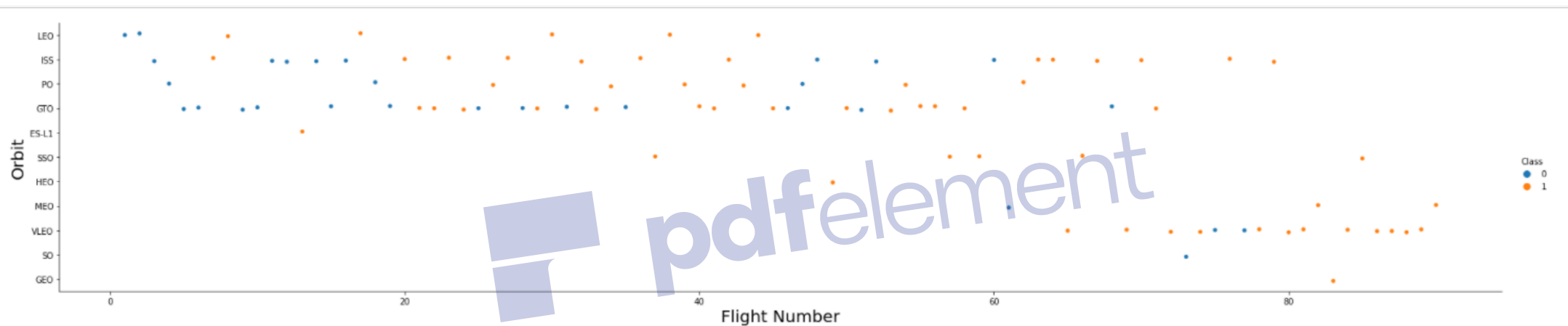
It is a scatter plot of Payload vs Launch Site. The blue dot indicates failure and the orange dot indicates success of landing of first stage of the Falcon 9. All three launch sites are showing similar patterns among different payload mass. It shows greater success rate within high pay load mass (more than 14,000kg) than low pay load mass.

# Success rate vs. Orbit type



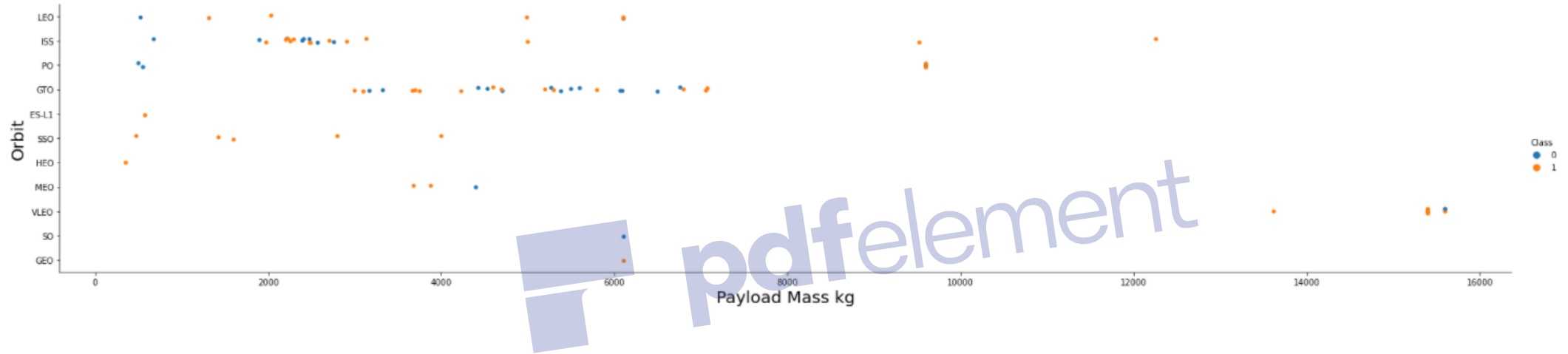
Every first stage of falcon 9 with E5-L1, GEO, HEO, and SSO orbits has landed successfully. Contrast to orbit GTO has about 50% success rate.

# Flight Number vs. Orbit type



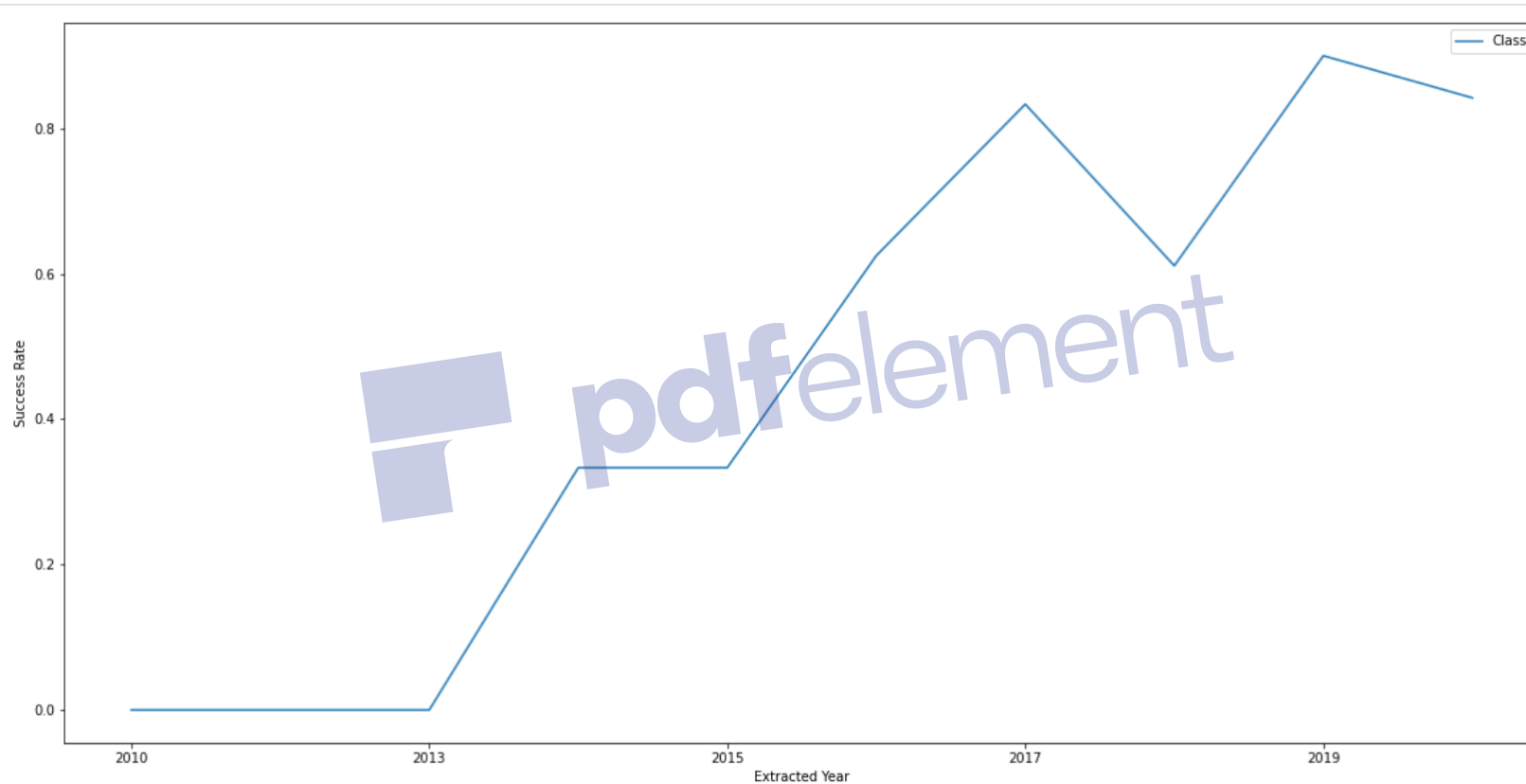
It is a scatter plot of Flight Number vs Orbit Type. The blue dot indicates failure and the orange dot indicates success of landing of first stage of the Falcon 9. Overall later attempts have higher success rate regardless orbit types. The ISS orbit shows great success rate all times except very beginning. The VLEO illustrate high success rate, however, it was used from flight number 60, therefore, hard to conclude it is the best launch site.

# Payload vs. Orbit type



It is a scatter plot of Payload vs Orbit Type. The blue dot indicates failure, and the orange dot indicates success of landing of first stage of the Falcon 9. As the payload gets heavier, it gives negative effect on GTO orbit whereas it affects positively on ISS orbit.

# Launch success yearly trend



The line indicates success rate of landing of the first stage each year. As time goes by, it has higher success rate but there has been a drop in 2018.

# EDA with SQL



# All launch site names

---

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

- There are 4 different launch sites.

# 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	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

It is the first 5 records with launch site names begin with 'CCA'.

# Total payload mass

---

customer	total_kg
NASA (CRS)	45596

The total payload carried by boosters from NASA.

# Average payload mass by F9 v1.1

---

booster_version	average_payload
F9 v1.1	2928.400000

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

# First successful ground landing date

---

landing__outcome	first_date
Success (ground pad)	2015-12-22

The date when the first successful landing outcome in ground pad

# 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)

The names of boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000.

# Total number of successful and failure mission outcomes

---

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

- There are 100 Success and 1 failure of mission outcomes.

# Boosters carried maximum payload

booster_version	payload_mass__kg_
F9 B5 B1048.4	15600
F9 B5 B1048.5	15600
F9 B5 B1049.4	15600
F9 B5 B1049.5	15600
F9 B5 B1049.7	15600
F9 B5 B1051.3	15600
F9 B5 B1051.4	15600
F9 B5 B1051.6	15600
F9 B5 B1056.4	15600
F9 B5 B1058.3	15600
F9 B5 B1060.2	15600
F9 B5 B1060.3	15600

- The maximum payload is 15,600kg and there are several F9 B5 versions of booster carried the payload



# 2015 launch records

---

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

- There are two records where the landing outcome is “Failure (drone ship)”, and it was in 2015
- Both had booster version F9 v1.1 and launch site CCAFS LC-40

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

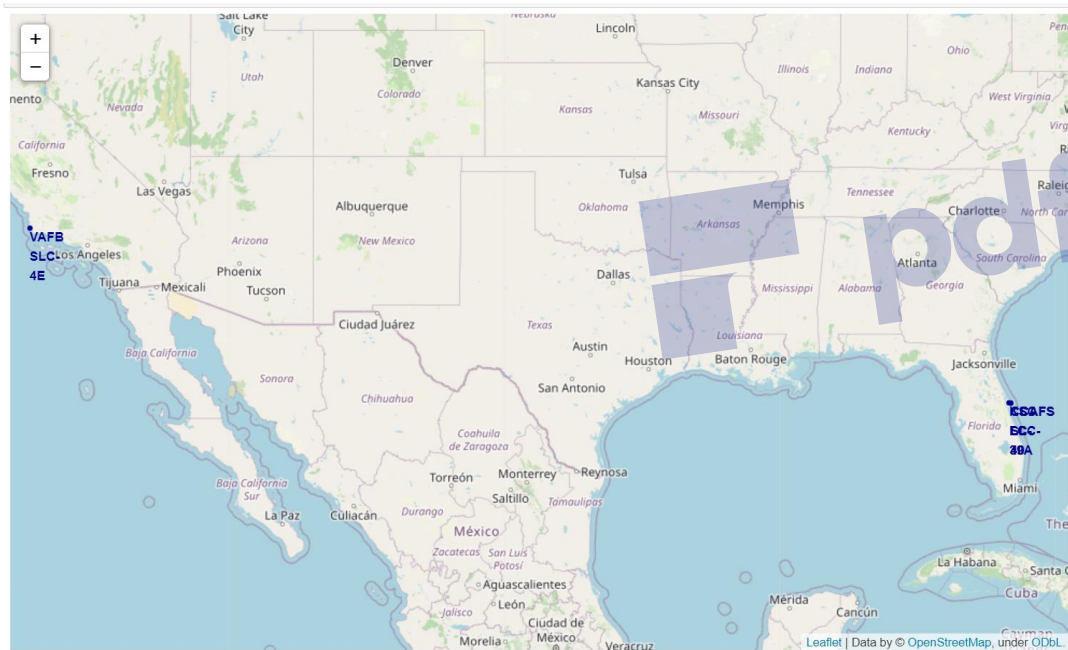
landing__outcome	COUNT
Success (drone ship)	5
Success (ground pad)	3

- There are two different success types in landing outcome from 2010-06-04 to 2017-03-20
- Success on drone ship has the most successes with 5, and on ground pad is following by 3



# Interactive map with Folium

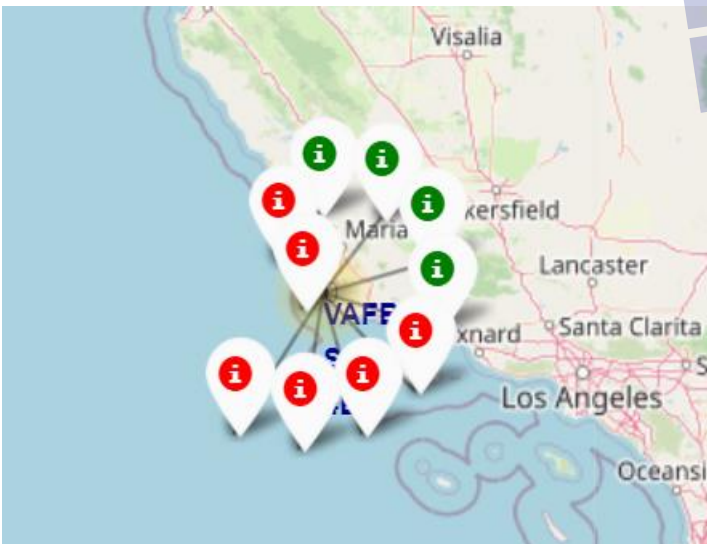
# <All Launch Sites>



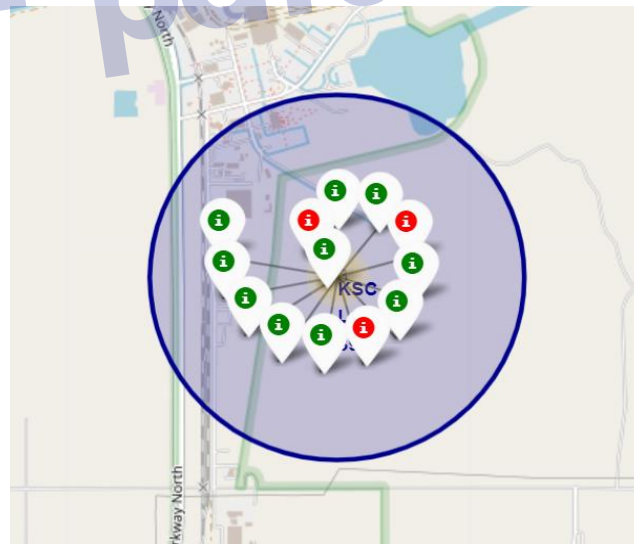
- There are three launch sites on the East coast, and one site on the West coast.
- All of the launch sites are located near the coast.

## <successed/failed launches>

- CCAFS LC-40 has 26 launches and only 7 of them were success.
- KSCLC-39A has the highest success rate.



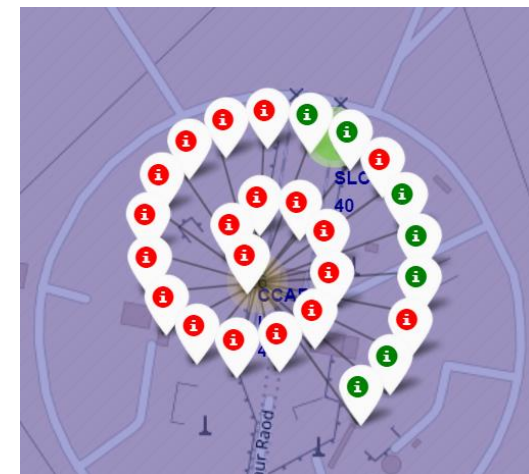
VAFB SLC-4E



KSCLC-39A



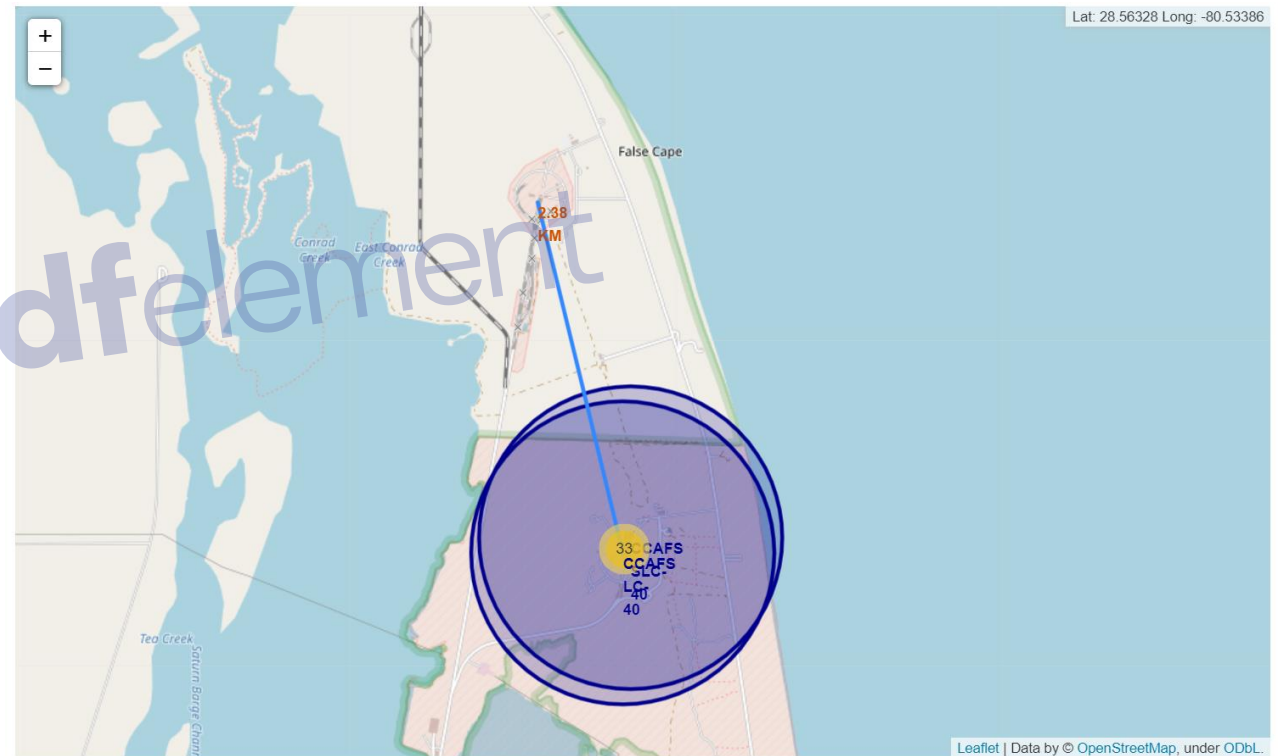
CCAFS SLC-40



CCAFS LC-40

## <Launch site and nearby railway>

- The distance between CCAFSLC-40 and nearby railway is about 2.38km



# Build a Dashboard with Plotly Dash

# < Launch Success Count for All Sites >

All Sites

% of successful launch by launch sites



- KSC LC-39A site has the highest success counts whereas CCAFS LC-40 has the lowest

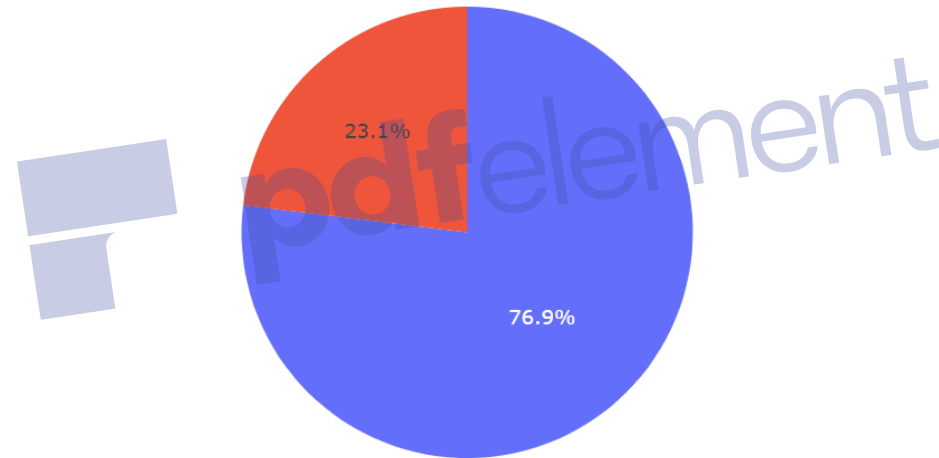


# <The Launch Site with The Highest Success Rate >

KSC LC-39A



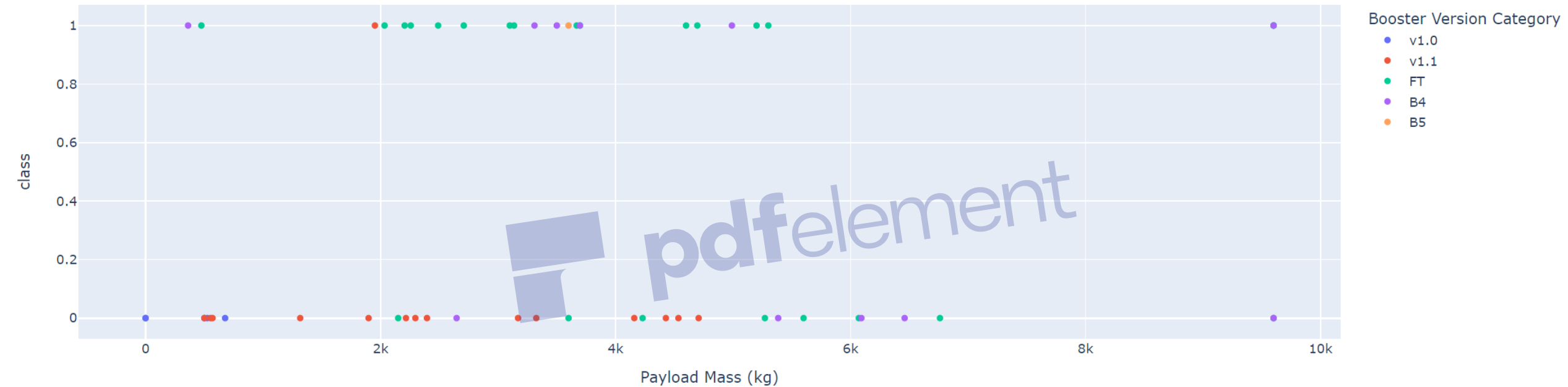
% of successful launch by KSC LC-39A



0  
1

- KSC LC-39A has the highest success rate of 76.9%

# < Payload vs Launch Outcome >

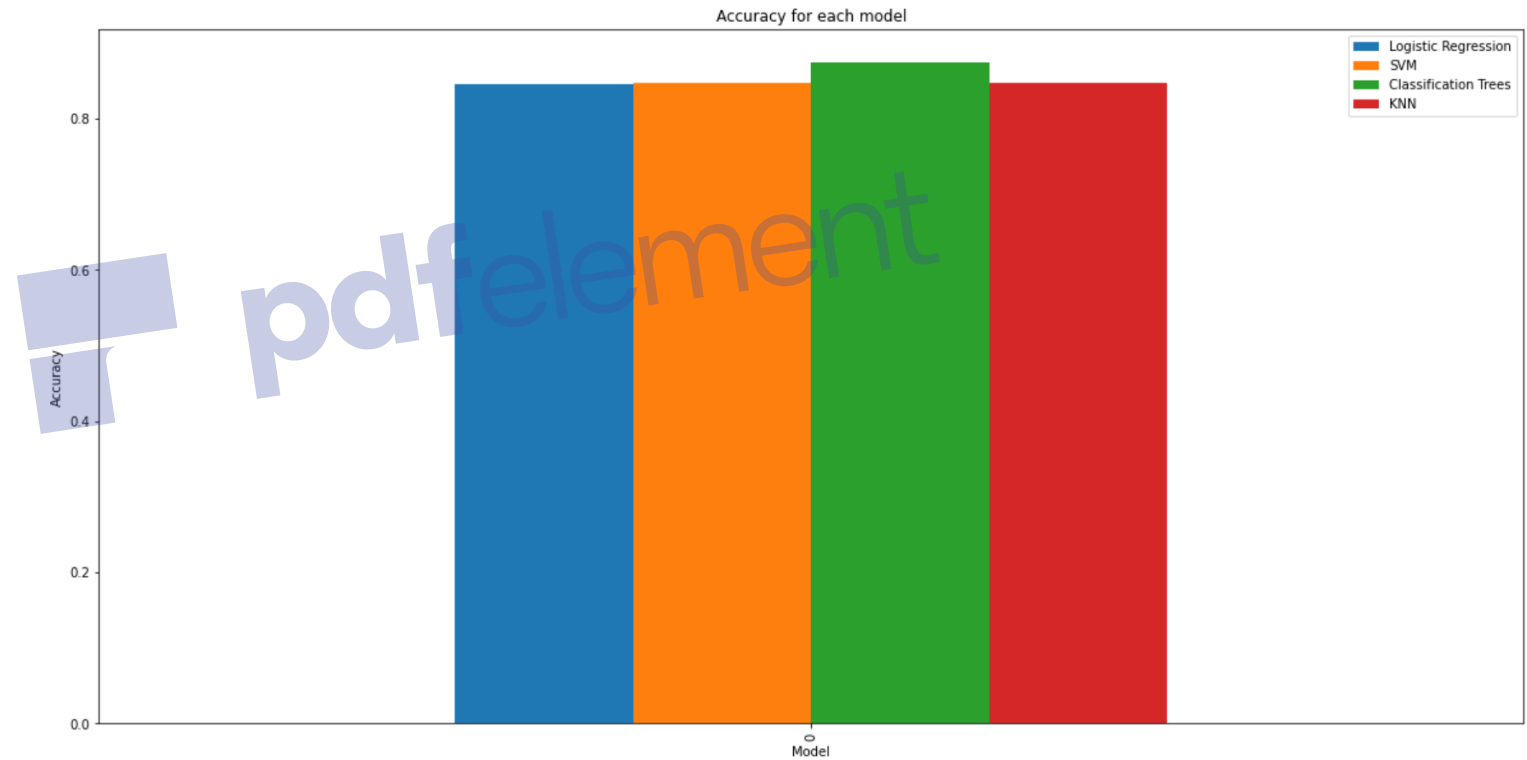


The payload between 2,000kg to 4,000kg has the highest success rate among all different payload mass. Also, when the booster version is FT, it shows great outcome whereas booster version v1.1 does not perform very well.

# Predictive analysis (Classification)

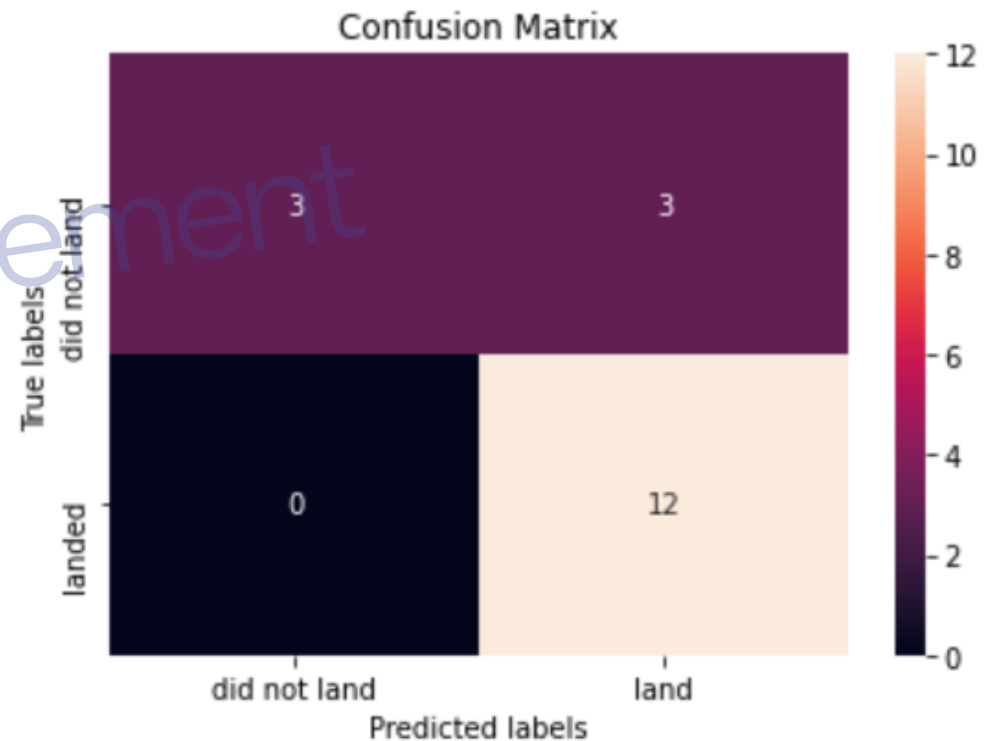
# Classification Accuracy

- Four different algorithms were used
  - Logistic regression
  - SVM
  - Classification tree
  - KNN
- Among four different models, the classification tree has the highest accuracy.



# Confusion Matrix

- This is the confusion matrix of classification tree model
- There are three prediction which the model thought it “land”, but it was “did not land” which is a false positive.



# CONCLUSION

---



- Success rate is increased with more attempts regardless other factors.
- Success rate is increased with payload mass between 2,000 and 5,000kg or over 14,000kg.
- Success rate is increased with orbits E5-L1, GEO, HEO, and SSO.
- Success rate is increased with KSCLC-39A launch site.
- Success rate is increased with booster version FT.
- Success rate has been steadily increased from 2013.