

Winning Space Race with Data Science

Anubhav Gupta
2022.07.18



Outline

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

Executive Summary

- Data:
 - Collection was performed using API and Web Scraping
 - Analysis was performed using SQL
 - Visualization was performed with Folium and Plotly
- All the four models Logistic regression, SVM, decision tree and knn performed equally well in predicting if the Falcon 9 rocket's first stage will land successfully with an accuracy of 0.83

Introduction

- Companies such as Galactic, Rocketlab, Blue Origin and SpaceX are making space travel affordable
- SpaceX is the most successful one
 - Rocket launches by SpaceX are relatively less expensive
 - FALCON 9 rocket launches cost 62 Million USD whereas other providers cost 165 Million USD
 - Because SpaceX can reuse first stage of the rocket
- What is the price of each launch?
 - This will be determined by predicting if the first stage of the rocket will land.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
 - Using SpaceX API and making a get request
 - Using Web Scraping on HTML table of Falcon 9 from the Wikipedia
- Perform data wrangling
 - Executing SQL queries on table loaded in a Db2 database
- 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

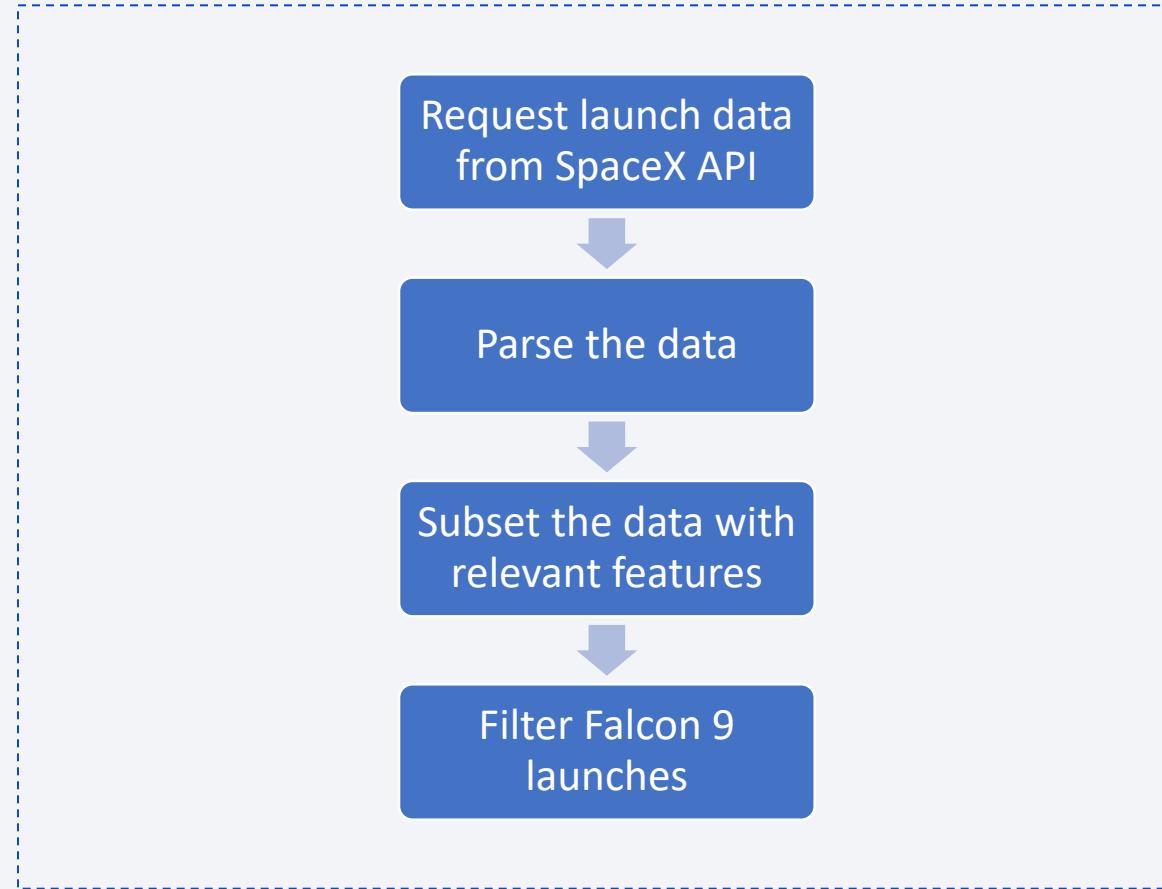
Data Collection

- Data sets were collected using:
 - SpaceX API and making a get request
 - Web Scraping on HTML table of Falcon 9 from the Wikipedia

Data Collection – SpaceX API

- GitHub URL:

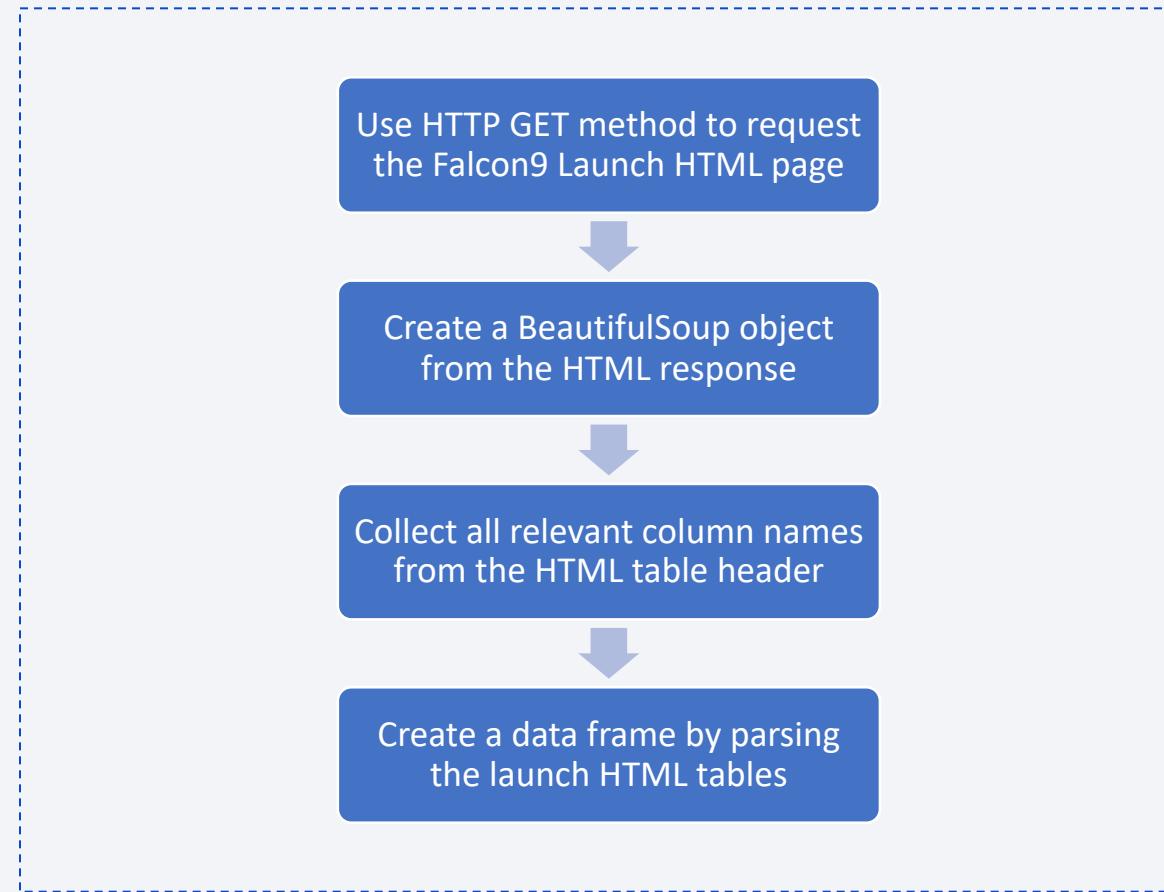
https://github.com/anubhav3/falcon_capstone/blob/main/Data_Collection_API.ipynb



Data Collection - Scraping

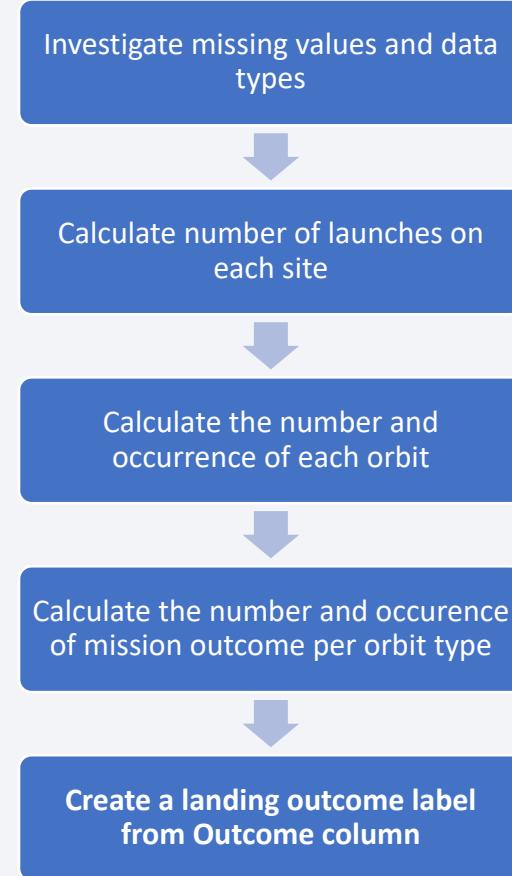
- GitHub URL:

[https://github.com/anubhav3/falcon_capstone/blob/main/
Data Collection with Web S
craping.ipynb](https://github.com/anubhav3/falcon_capstone/blob/main/Data%20Collection%20with%20Web%20Scraping.ipynb)



Data Wrangling

- GitHub:
https://github.com/anubhav3/falcon_capstone/blob/main/Data_wrangling_EDA.ipynb



EDA with Data Visualization

- Following charts were plotted:
 - Flight Number and Launch Site
 - Payload and Launch Site
 - Success rate of each orbit type
 - Flight Number and Orbit type
 - Payload and Orbit type
 - Launch success yearly trend
- URL:
https://github.com/anubhav3/falcon_capstone/blob/main/EDA_with_Data_visualization.ipynb

EDA with SQL

- Established connection with the database
- Displayed names of unique launch sites in space mission
- Displayed 5 records where launch sites begin with the string 'CCA'
- Displayed the total payload mass carried by boosters launched by NASA (CRS)
- Displayed average payload mass carried by booster version F9 v1.1
- Listed the date when the first successful landing outcome in ground pad was achieved.
- GitHub URL:
[https://github.com/anubhav3/falcon_capstone/blob/main/EDA with SQL.ipynb](https://github.com/anubhav3/falcon_capstone/blob/main/EDA%20with%20SQL.ipynb)

EDA with SQL (contd.)

- Listed the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- Listed the total number of successful and failure mission outcomes
- Listed the names of the booster versions which have carried the maximum payload mass
- Listed the failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015
- Ranked 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
- GitHub URL:
https://github.com/anubhav3/falcon_capstone/blob/main/EDA_with_SQL.ipynb

Build an Interactive Map with Folium

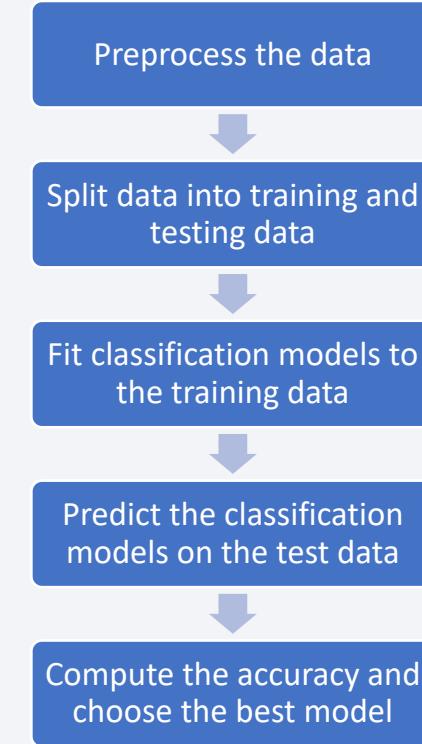
- Launch success rate may depend on the location and proximities of a launch site
- Markers, circles, and lines were added to map to calculate proximity of a launch site to closest cities, railways, highways and coastlines
- GitHub URL:
https://github.com/anubhav3/falcon_capstone/blob/main/Launch_Sites_Locations_Analysis_with_Folium.ipynb

Build a Dashboard with Plotly Dash

- Plotted:
 - Pie charts to represent success rate of launches of different sites
 - Scatter plot to identify correlation between Payload Mass and success rate of different sites
- GitHub URL:
https://github.com/anubhav3/falcon_capstone/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

- GitHub URL:
[https://github.com/anubhav3/falcon_capstone
blob/main/SpaceX%20Machine%20Learning%
20Prediction_Part_5.ipynb](https://github.com/anubhav3/falcon_capstone/blob/main/SpaceX%20Machine%20Learning%20Prediction_Part_5.ipynb)



Results

- Exploratory data analysis results
 - Different launch sites have different success rate
 - There is a period gap when sites were used of launches
 - Maximum Pay load mass is dependent on the launch sites
 - Success rate is dependent on the orbit types and payload mass
 - Success rate increased from 2013 to 2020

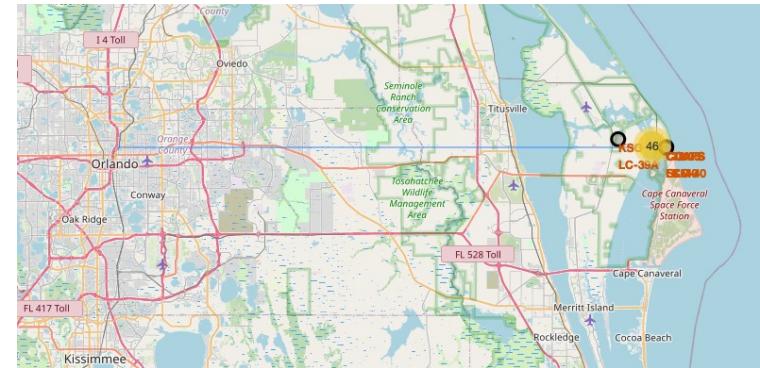
Results (contd.)

Launch sites chosen: CCAFS SLC-40 and CCAFS LC-40

Distance from nearest coast = 0.85 km



Distance from nearest city = 78.05 km



Distance from nearest railway = 1.32 km

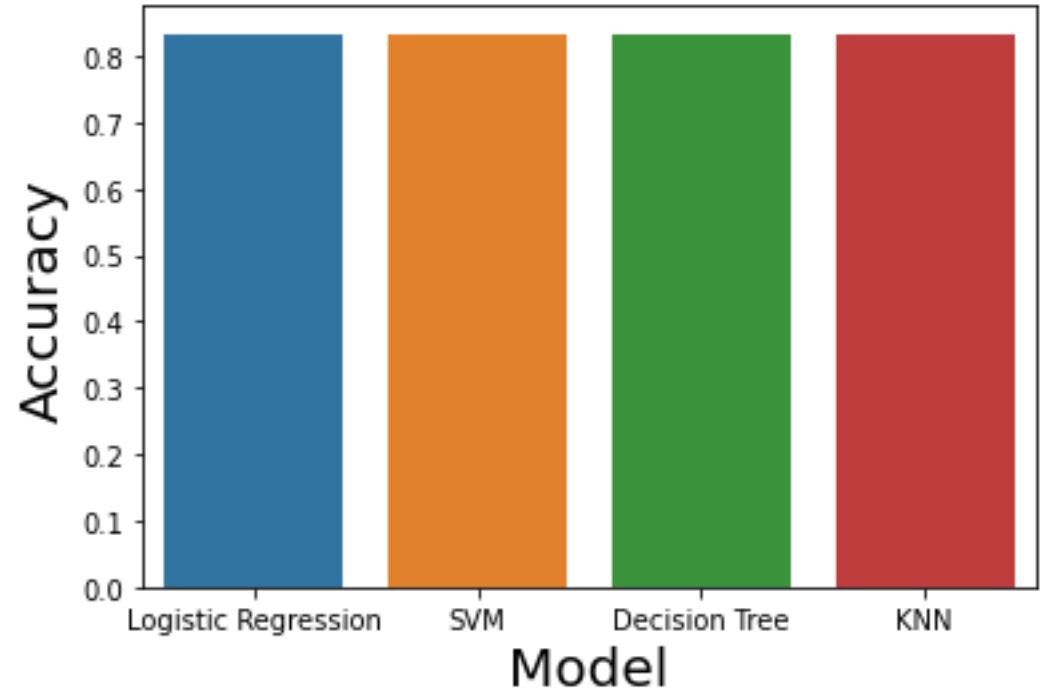


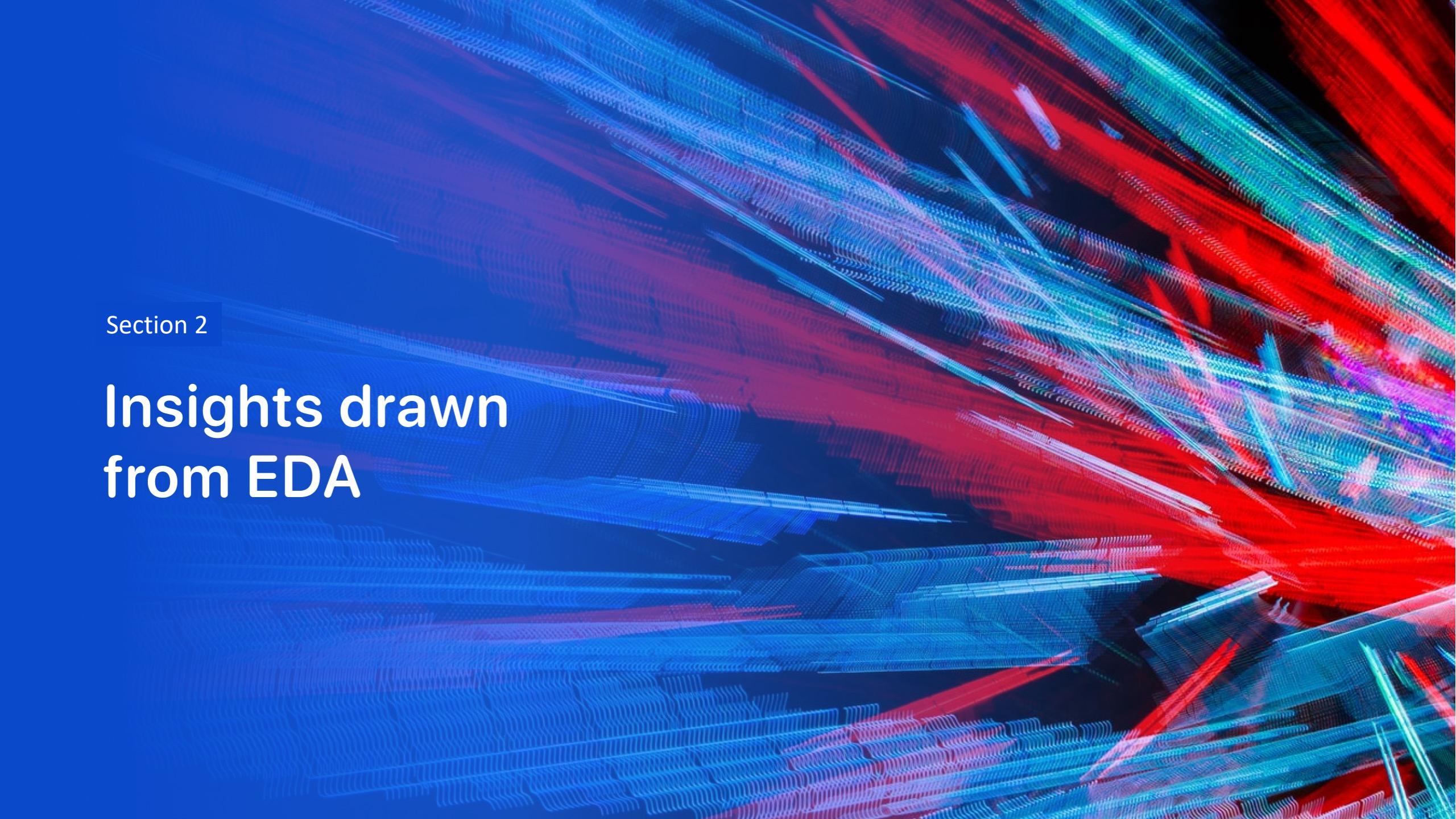
Distance from nearest highway = 0.60 km

- The launch sites are closest to highway
- The launch sites are farthest from city

Results (contd.)

- All the four models logistic regression, SVM, decision tree and knn have the same accuracy of 0.83 on the test data

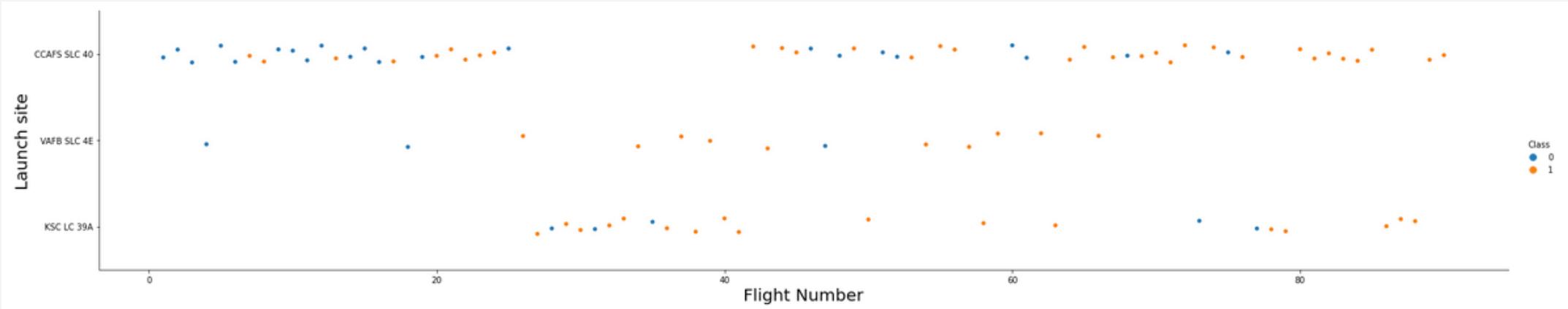


The background of the slide features a complex, abstract digital visualization. It consists of numerous thin, glowing lines that create a sense of depth and motion. The lines are primarily blue and red, with some green and purple highlights. They form a grid-like structure that curves and twists across the frame, resembling a three-dimensional space or a network of data points. The overall effect is futuristic and dynamic.

Section 2

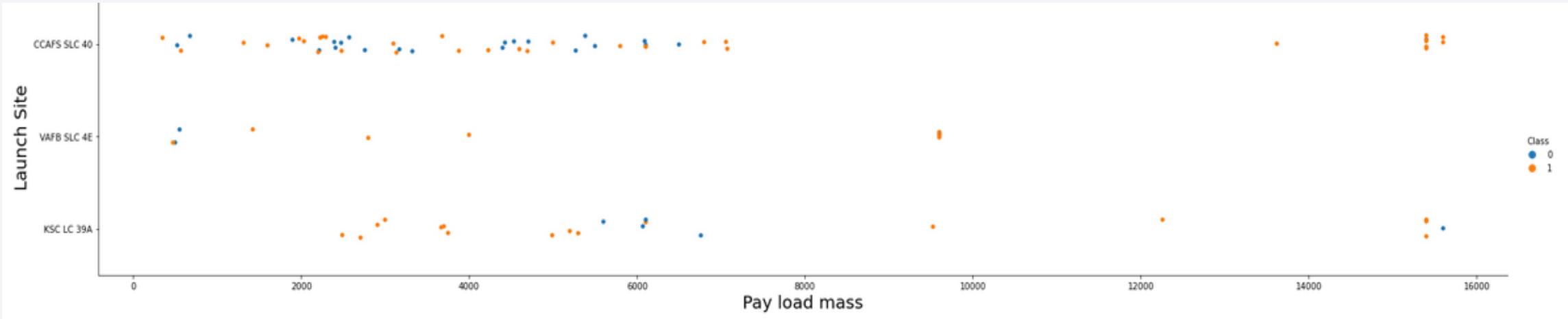
Insights drawn from EDA

Flight Number vs. Launch Site



- Launch sites are used in time intervals
 - KSC LC 39A was used after ~20 flight number
 - VAFB SLC 4E was used before ~70 flight number
 - CCAFS SLC 40 was used before ~23 and after ~40 flight number; it had more successes after ~40 flight number as compared to before ~23 flight number

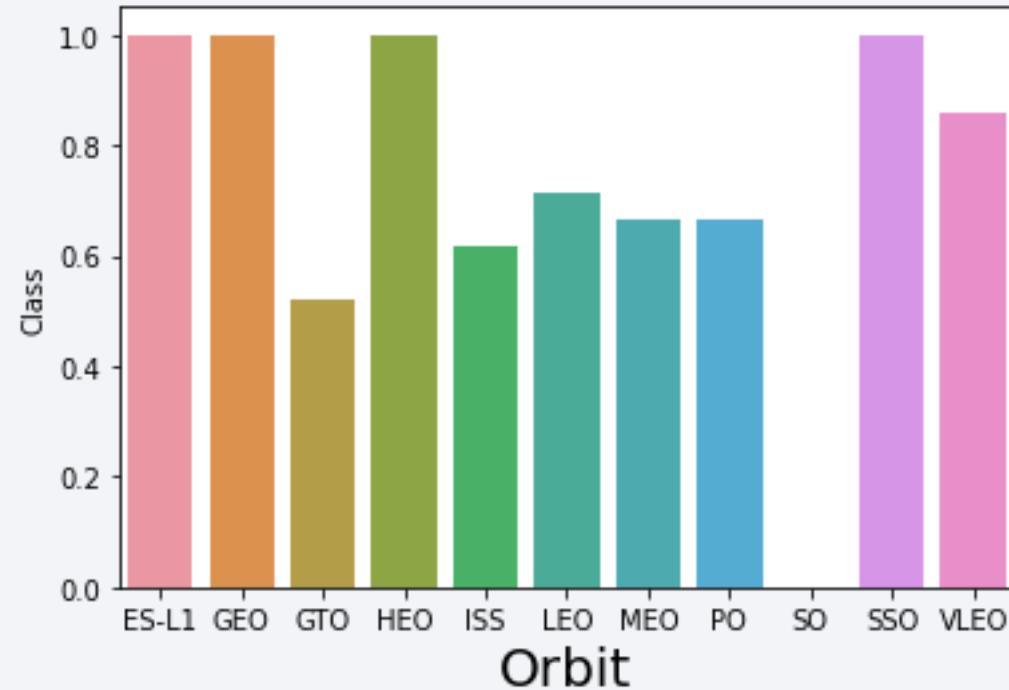
Payload vs. Launch Site



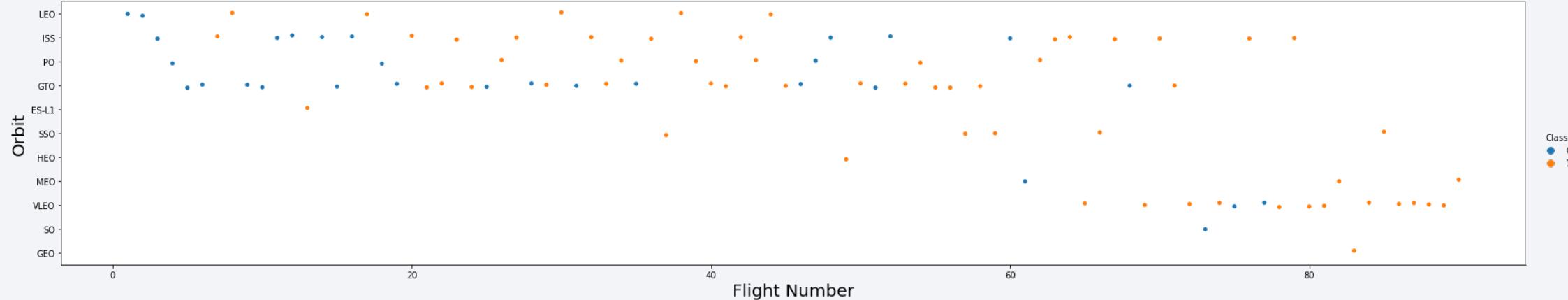
- For VAFB SLC 4E launch site, there are no rockets launched for heavy pay load mass i.e greater than ~ 10000
- For KDC LC 39A, there are no rockets launched for light pay load mass i.e. lighter than ~ 2000
- For CCAFS SLC 40, there are no rockets launched between pay load mass ~ 8000 and ~ 13000

Success Rate vs. Orbit Type

- Success rate is dependent on the orbit type
 - GTO orbit type has the lowest launch success rate (=0.5)
 - ES-L1, GEO, HEO and SSO have the highest launch success rate (=1)

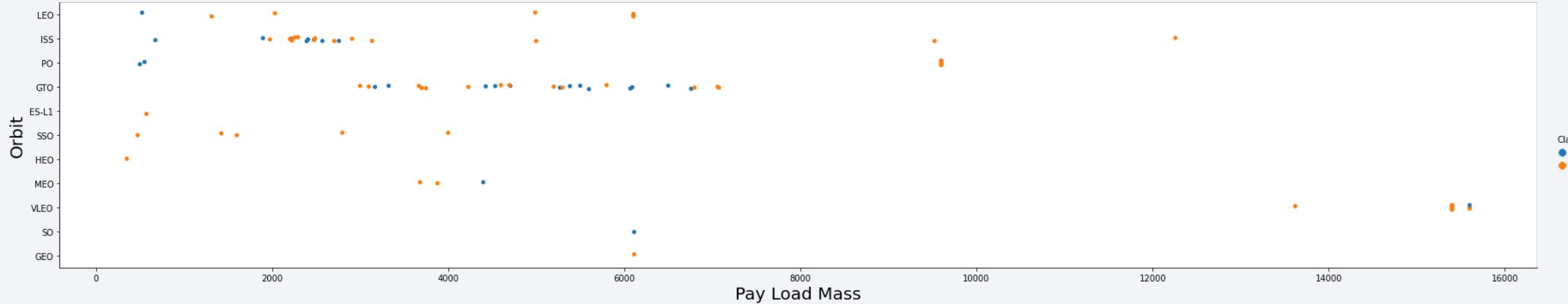


Flight Number vs. Orbit Type



- Orbit LEO always had success after ~ 5 flight number
- Orbit ISS had success between ~ 20 and ~ 45 flight number and after ~ 60 flight number
- Orbit VLEO had success for most of flight number except for two consecutive flight numbers before ~ 80
- Orbits MEO, VLEO, SO and GEO were used after flight number ~ 60

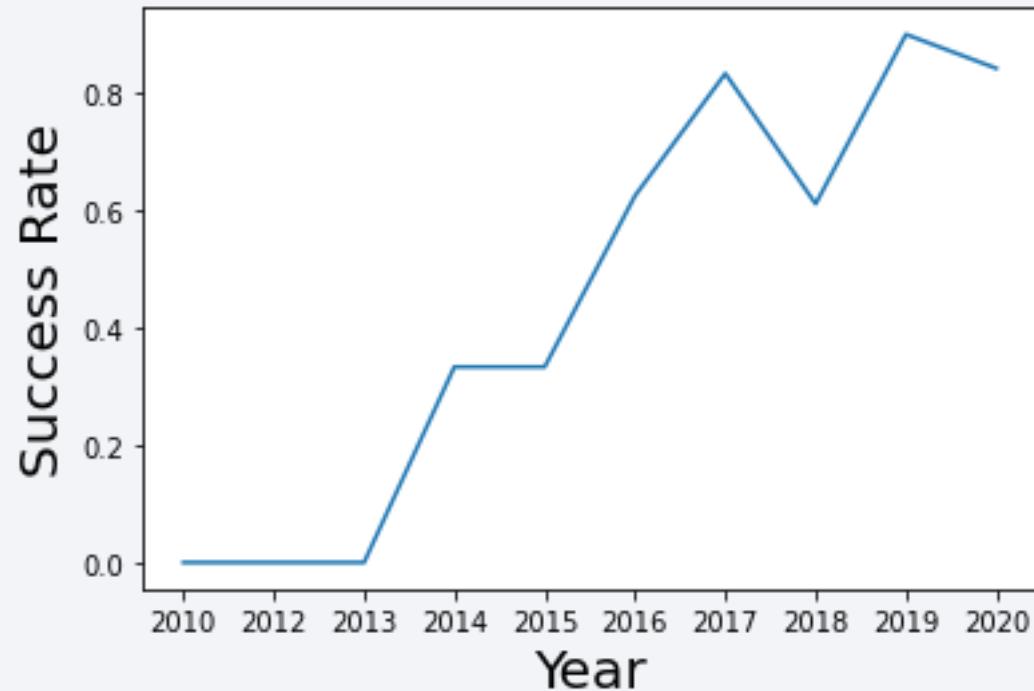
Payload vs. Orbit Type



- Success rate is higher for heavier pay load mass for orbits LEO, PO and SS
- SSO always success independent of the pay load mass
- For GTO, there is no relation between the launch success rate and pay load mass

Launch Success Yearly Trend

- The launch success rate was 0 between 2010 and 2013
- From 2013 to 2020, the launch success increased with some drop in between



All Launch Site Names

- Names of launch sites:
 - CCAFS LC-40, CCAFS SLC-40, KSC LC-39A, VAFB SLC-4E
- There are four 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

- There was no pay load for first two launches

Total Payload Mass

- The total payload mass for boosters launches by NASA (CRS) is:
 - 45596 kg

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1
 - 2534 kg

First Successful Ground Landing Date

- Date of the first successful landing outcome on ground pad:
 - 2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000:

F9 B4 B1040.2

F9 B4 B1040.1

F9 B4 B1043.1

F9 B5 B1046.2

F9 B5 B1047.2

F9 B5 B1048.3

F9 B5 B1051.2

F9 B5 B1058.2

F9 B5B1054

F9 B5B1060.1

F9 B5B1062.1

F9 FT B1021.2

F9 FT B1031.2

F9 FT B1032.2

F9 FT B1020

F9 FT B1022

F9 FT B1026

F9 FT B1030

F9 FT B1032.1

F9 v1.1

F9 v1.1 B1011

F9 v1.1 B1014

F9 v1.1 B1016

Total Number of Successful and Failure Mission Outcomes

- Number of successful outcomes = 100
- Number of unsuccessful outcomes = 1
- The dataset is unbalanced.

Boosters Carried Maximum Payload

- Names of the booster which have carried the maximum payload mass

"F9 B5 B1048.4"

"F9 B5 B1049.4"

"F9 B5 B1051.3"

"F9 B5 B1056.4"

"F9 B5 B1048.5"

"F9 B5 B1051.4"

"F9 B5 B1049.5"

"F9 B5 B1060.2 "

"F9 B5 B1058.3 "

"F9 B5 B1051.6"

"F9 B5 B1060.3"

"F9 B5 B1049.7 "

2015 Launch Records

- Failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015:

Booster Version	Launch Site	Date
"F9 v1.1 B1012"	"CCAFS LC-40"	"10-01-2015"
"F9 v1.1 B1015"	"CCAFS LC-40"	"14-04-2015"

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

Number of failure = 4

Number of success = 6

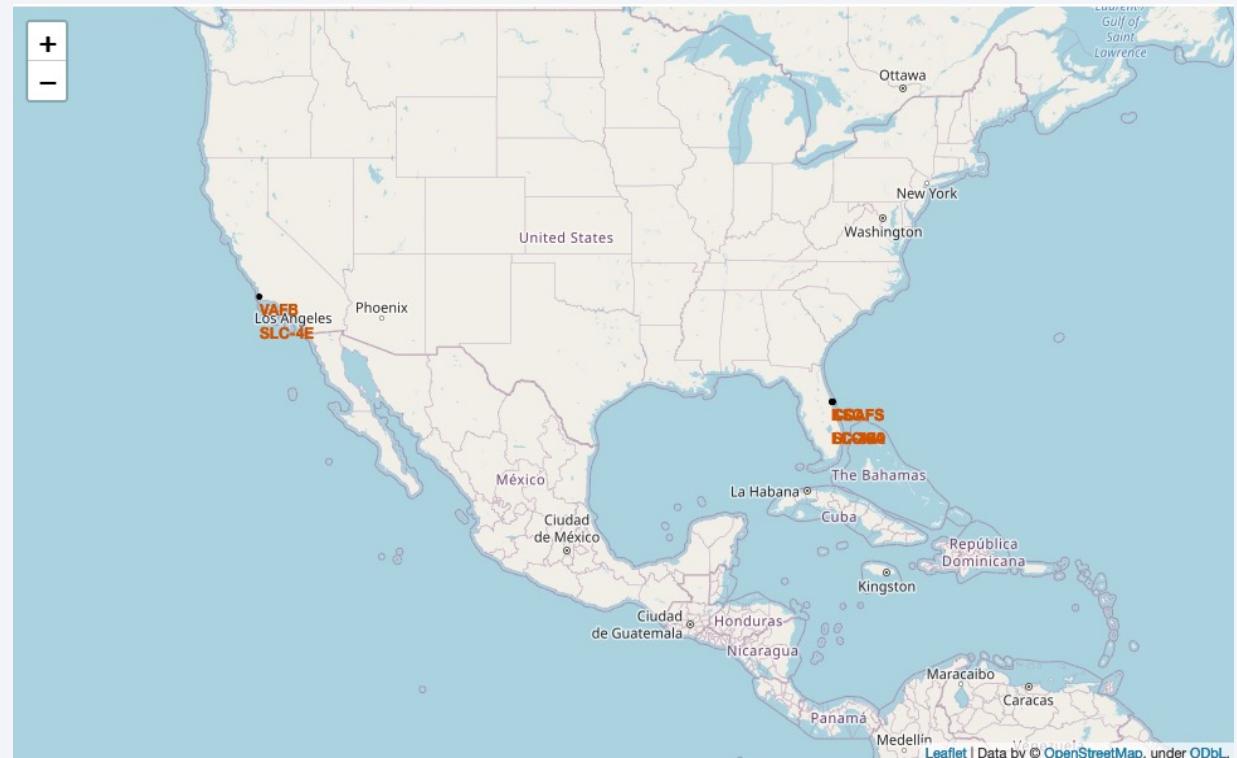
The background of the slide is a photograph taken from space at night. It shows the curvature of the Earth against a dark blue-black void of space. City lights are visible as numerous small white and yellow dots, primarily concentrated in the lower right quadrant where the United States appears. In the upper right, the green and yellow glow of the aurora borealis is visible. The atmosphere of the Earth is thin and hazy, appearing as a light blue band near the horizon.

Section 3

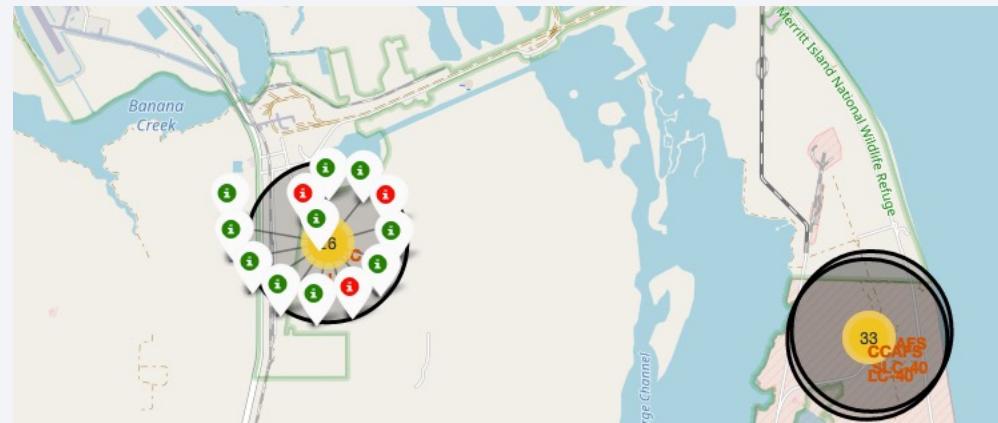
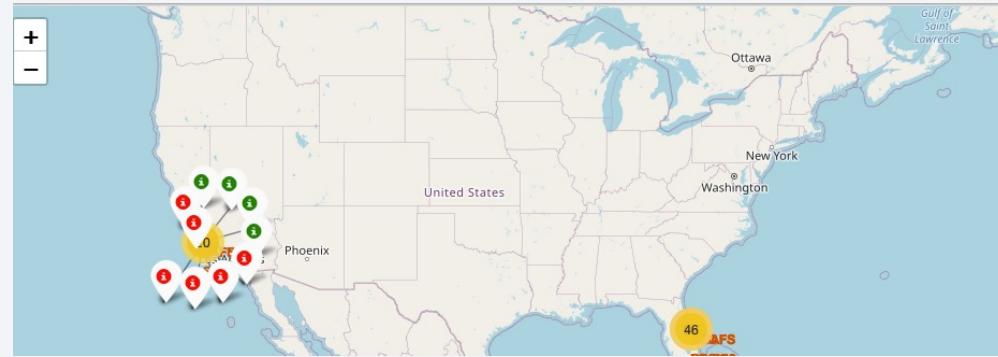
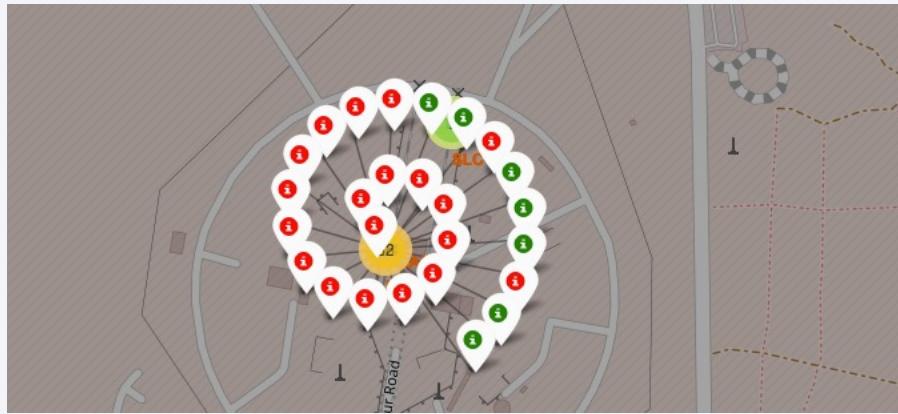
Launch Sites Proximities Analysis

Launch sites' locations

- Launch sites are very close to the coast
- Two of the launch sites (CCAFS LC-40 and CCAFS SLC-40) are at same position and close to KSC LC-39A
- VAFB SLC-4E is on the west coast of USA as compared to others which are on the east coast



Launch success for each sites



- KSC LC-39A has the highest proportion of successes
- CCAFS LC-40 has the highest proportion of failure

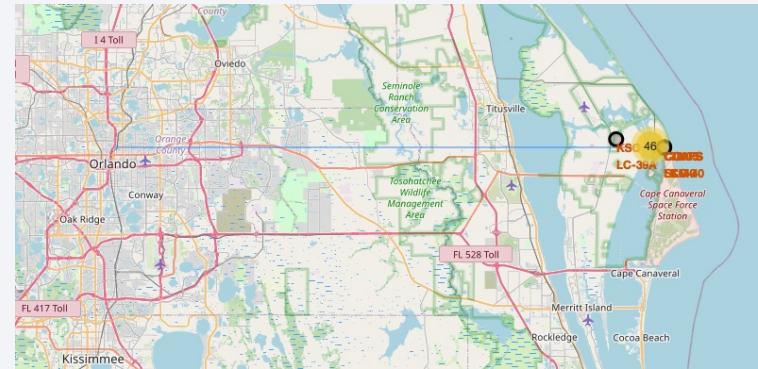
Proximity of launch site to artificial structures

Launch sites chosen: CCAFS SLC-40 and CCAFS LC-40

Distance from nearest coast = 0.85 km



Distance from nearest city = 78.05 km



Distance from nearest railway = 1.32 km



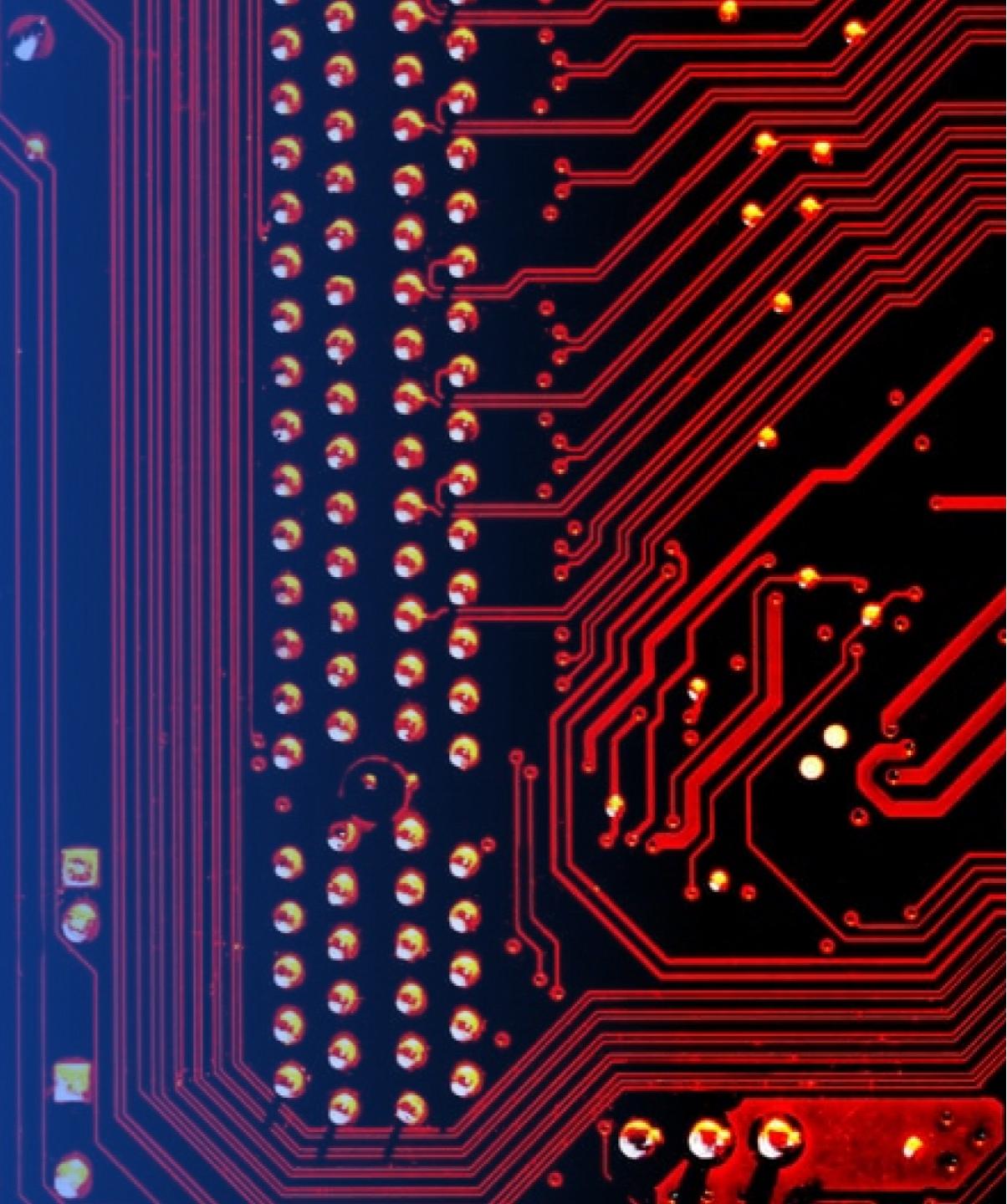
Distance from nearest highway = 0.60 km



- The launch sites are closest to highway
- The launch sites are farthest from city

Section 4

Build a Dashboard with Plotly Dash



Number of successes per launch sites

Total Success Launches



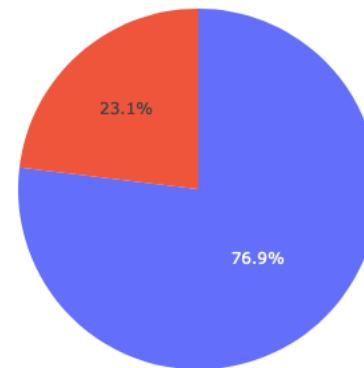
- KSC LC-39A had the highest proportion of successful launches (41.2%)
- CCAFS LC-40 had the lowest proportion of successful launches (14.4%)

Success rate of launch site KSC LC-39A

KSC LC-39A

Align to Right
Success probability

x ▾



1
0

- KSC LC-39 A had the success rate of 76.9%

Launch success rate with booster version and pay load mass



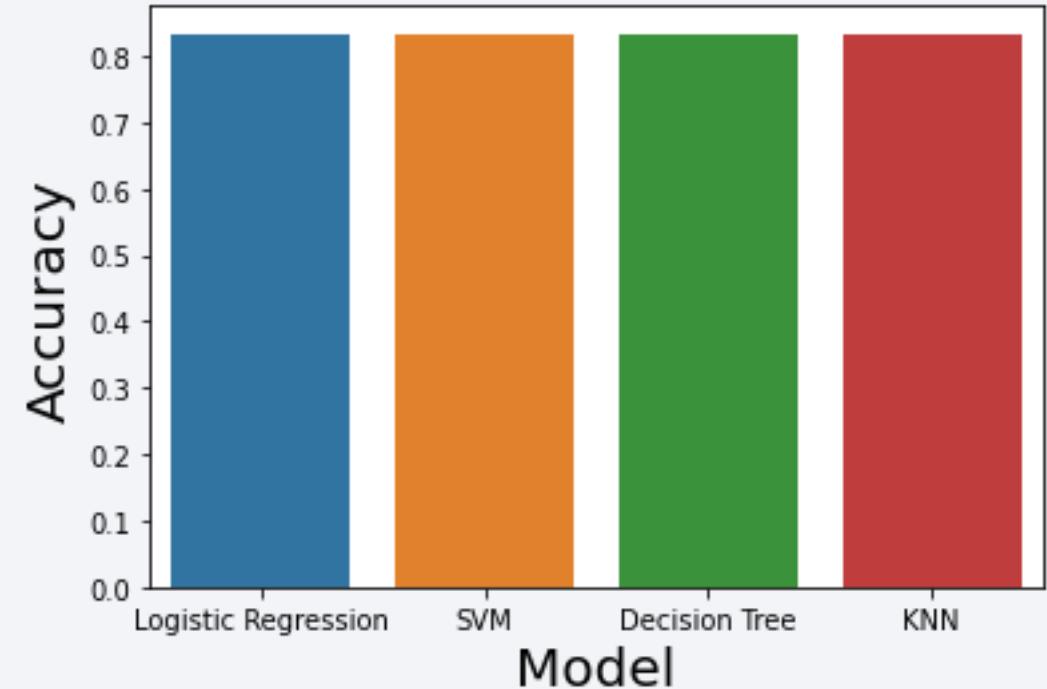
- Booster FT had a higher success rate till pay load mass of around 5500 kg
- Booster version v1.1. had the lowest success rate irrespective of the pay load mass

Section 5

Predictive Analysis (Classification)

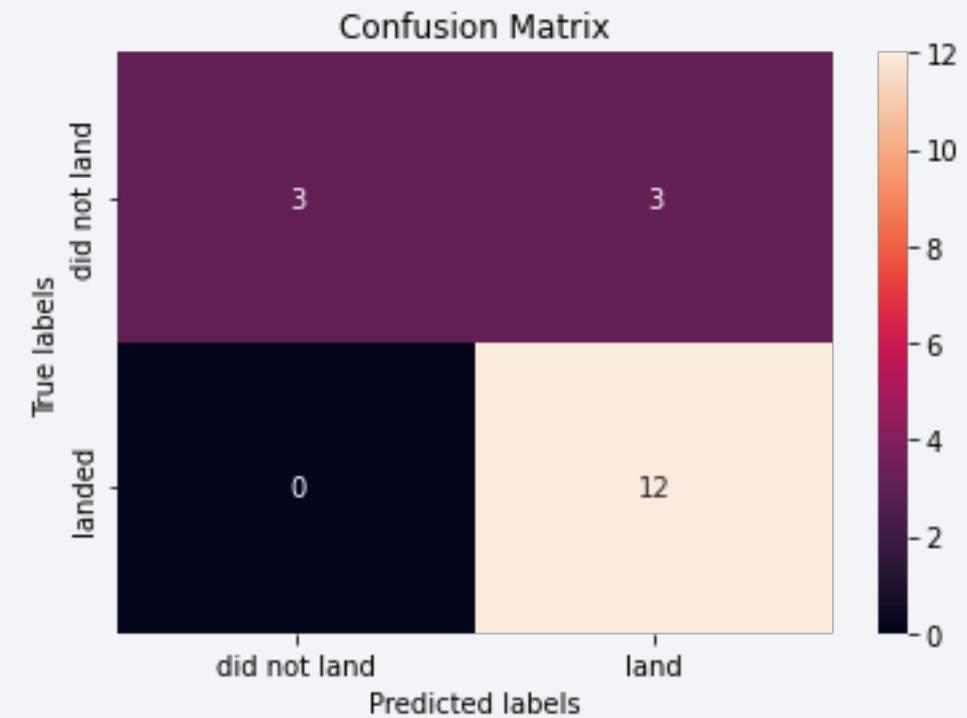
Classification Accuracy

- All the models have the same accuracy of 0.83 on the test data



Confusion Matrix

- All the four models have the same confusion matrix
- All the four models have some false positives



Conclusions

- The launch success increased from 2013 to 2020
- All the four models Logistic regression, SVM, decision tree and KNN performed equally well in predicting launch of Falcon 9 rocket's first stage with an accuracy of 0.83
- Any of the above four models can be used for predicting launch success rate

Thank you!

