Predicting Falcon 9 First Stage Landings: A Cost-Saving Analysis for SpaceX Competitors

Prepared by :Abat Zhuldassov

Date: 8/21/2024

Outline

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

Executive Summary

- Summary of methodologies

In this work Python programming language was used for data collection (API, webscraping) exploratory data analysis(SQL, pandas, numpy) and visualization (matplotlib, seaborn, Folium). In addition, predictive analysis was done using machine learning algorithms (Logistic Regression, KNN, SVM and Decision Tree)

- Summary of all results

Exploratory Data Analysis

Interactive analytics in screenshots

Predictive Analytics

•How do various features interact to determine the success rate of a landing?
•What operating conditions are necessary to ensure a successful landing program?

Introduction

The aim of this project, predict the likelihood of a successful landing of the Falcon 9 first stage. SpaceX lists the cost of a Falcon 9 rocket launch at 62 million dollars on its website, while other providers charge upwards of 165 million dollars. A significant portion of SpaceX's savings comes from its ability to reuse the first stage. By predicting the success of the first stage landing, one can estimate the overall cost of a launch. This analysis could be valuable for alternative companies looking to compete with SpaceX in the rocket launch market.

The following problems should be solved:

Which the most successful launch site?

What factors influence the successful landing of a rocket?

Which model suited best for predicting successful landing?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected through the API endpoint from Space X and web scraping from Wikipedia webpage
- Data wrangling
  - Removing Nulls, unnecessary columns and assigning labels to landing outcome through the one-hot encoding
- Exploratory data analysis (EDA) using visualization and SQL
- Interactive visual analytics using Folium and Plotly Dash
- Predictive analysis using classification models

Data Collection

- Data was collected from 2 sources:

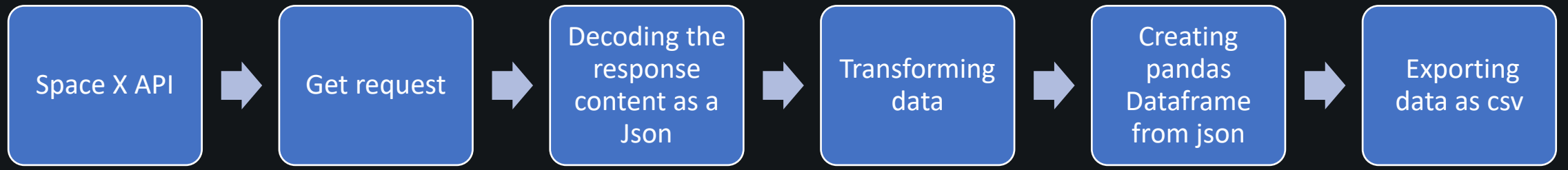
1) API endpoint from Space X https://api.spacexdata.com/v4/launches/past, containing the following columns FlightNumber, Date, BoosterVersion, PayloadMass, Orbit, LaunchSite, Outcome, Flights, GridFins, Reused, Legs, LandingPad, Block, ReusedCount, Serial, Longitude Latitude

2) Wikipedia webpage https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches containing the following columns Flight No., Date and time (UTC), Version,Booster[b], Launch site, Payload[c], Payload mass, Orbit, Customer, Launch outcome, Booster landing

Data Collection – SpaceX API

The following libraries were used for data collection from Space X API: requests, pandas, numpy, datetime.

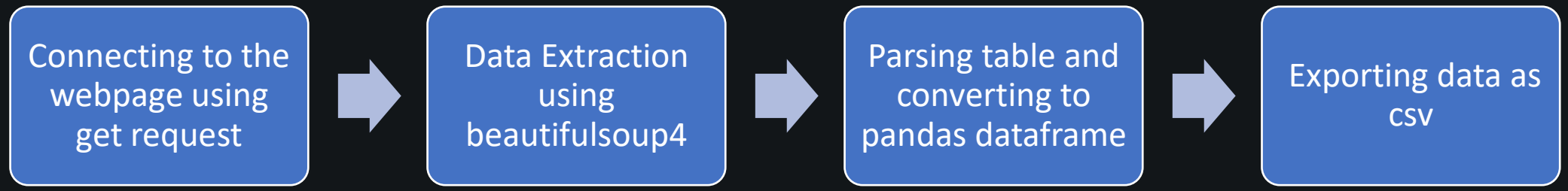
The process of data collection through API is shown below



- Link to the code

Data Collection - Scraping

- The following libraries were used for data collection from Wikipedia webpage: beautifulsoup4, sys, unicodedata, requests, pandas
- The process of data web scraping is shown below

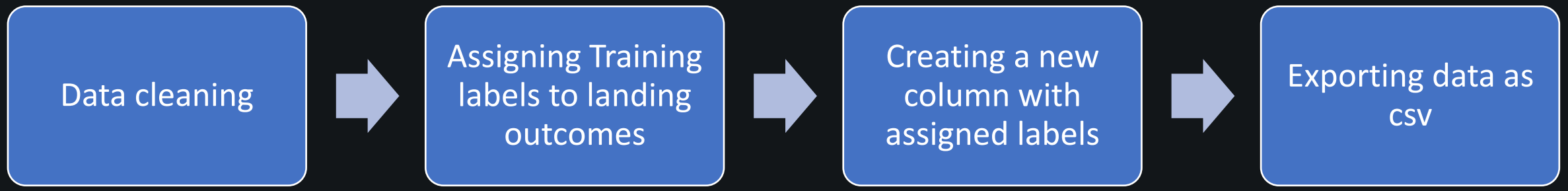


- Link to the code

Data Wrangling

After handling with null values, the following landing outcomes True ASDS, None None, True RTLS, False ASDS, True Ocean, False Ocean, None ASDS, False RTLS was assigned to 0 (unsuccessful) and 1 (successful) landing.

The flowchart of data wrangling is shown below



- Link to the code

EDA with Data Visualization

- For visualizations the following charts were used:

  1) Scatter plot to identify the relationships between Flight Number and Payload Mass, Flight Number and Launch Site, Payload and Launch Site.

  2) Bar chart to find the success rate of each orbit

  3) Line chart to find the average launch success trend.

- Link to the code

EDA with SQL

For EDA the following questions were answered using SQL queries :

- Names of unique launch sites
- Total payload mass carried by boosters launched by NASA (CRS)
- Average payload mass carried by booster version F9 v1.1
- The date when the first successful landing outcome in ground pad was achieved.
- The names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- The total number of successful and failure mission outcomes
- The names of the booster_versions which have carried the maximum payload mass

- Link to the code

Interactive Map with Folium

The following markers, circles, lines added to a folium map:

- Red circles for indication of launch cites and NASA Johnson Space Center on map.
- Markers were used for showing successful (green) and unsuccessful (red) landings.
- Lines were used for indication the distance between launch cite and nearest coastline, highway, railroad and city with marker of distance in km.

- Link to the code

Build a Dashboard with Plotly Dash

The following plots/graphs were added to a dashboard:

- Pie chart to show the percentage of success launches across different launch cites and in case of choosing one launch cite it shows the proportion successful and unsuccessful launches.
- Scatter plot to show the relationship between landing outcome and payload mass

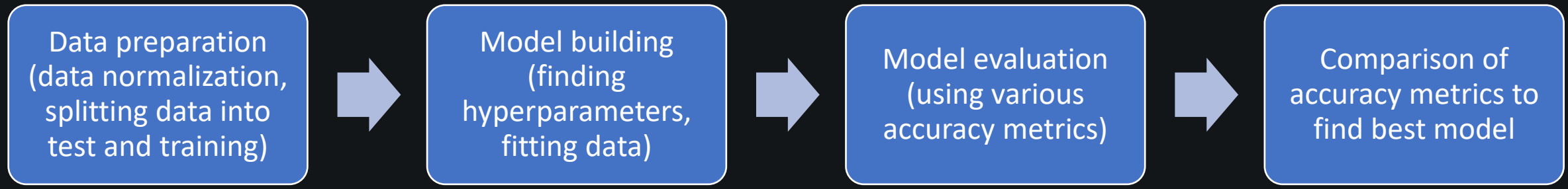
For interactivity, the following filters were added:

- Filter for choosing launch cite
- Filter for choosing pay load mass

- Link to the code

Predictive Analysis (Classification)

- Data was normalized, converted to array and splited into training and testing datasets
- Different machine learning models such as Logistic Regression, SVM, KNN and Decision Tree. Models were tuned using GridSearchCV.
- The model was evaluated using jaccard index, f1 score and accuracy score.



- Link to the code

Results

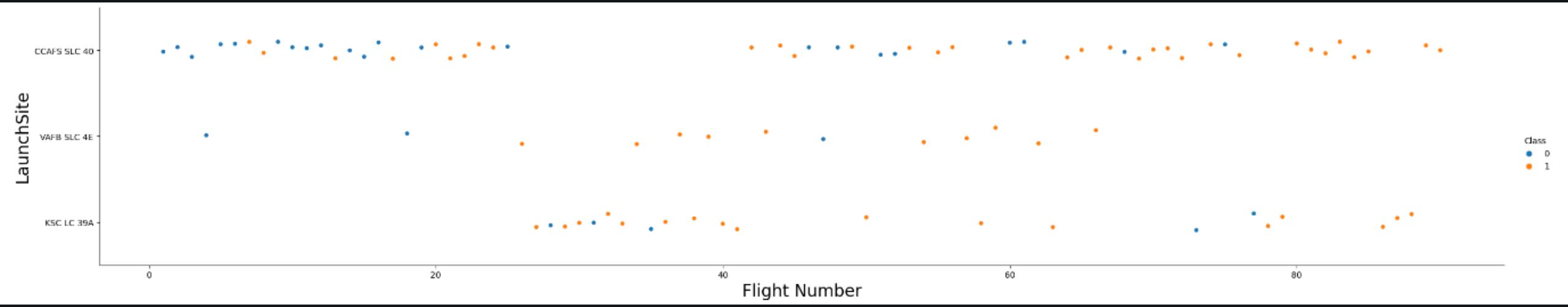
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

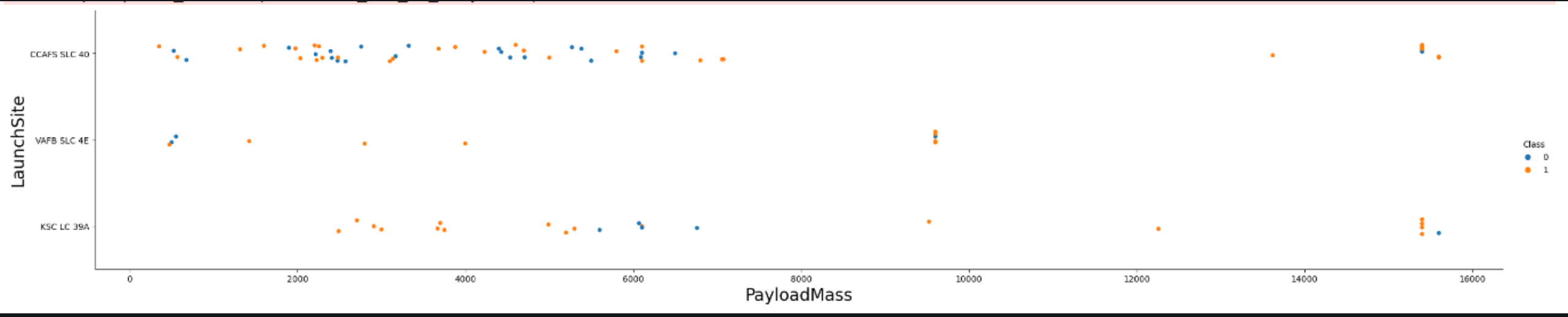
A scatter plot of Flight Number vs. Launch Site



According to the graph, most launches were done in CCAFS SLC 40 site. The successful landings in this site started after 60 launches. However, in KSC LC 39A the successful landings started almost after 40 launches. VAFB SLC 4E was less popular launch site, with the lowest unsuccessful landings

Payload vs. Launch Site

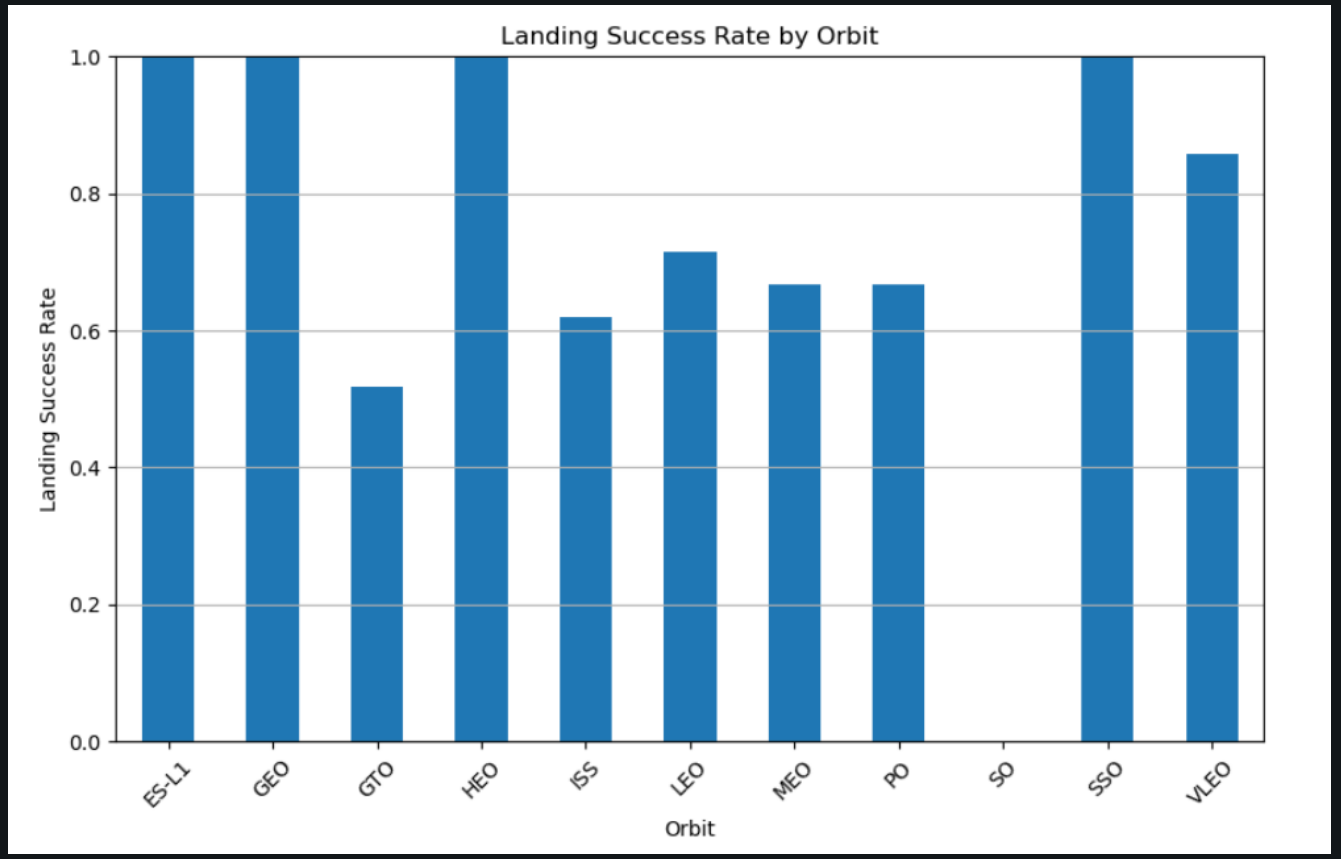
A scatter plot of Payload vs. Launch Site



It can be found from the graph that VAFB-SLC launch site never launched rockets with heavy payload mass (> 10000 kg). Most successful landings was with heavy payload mass from 12000 to 16000 kg.

Success Rate vs. Orbit Type

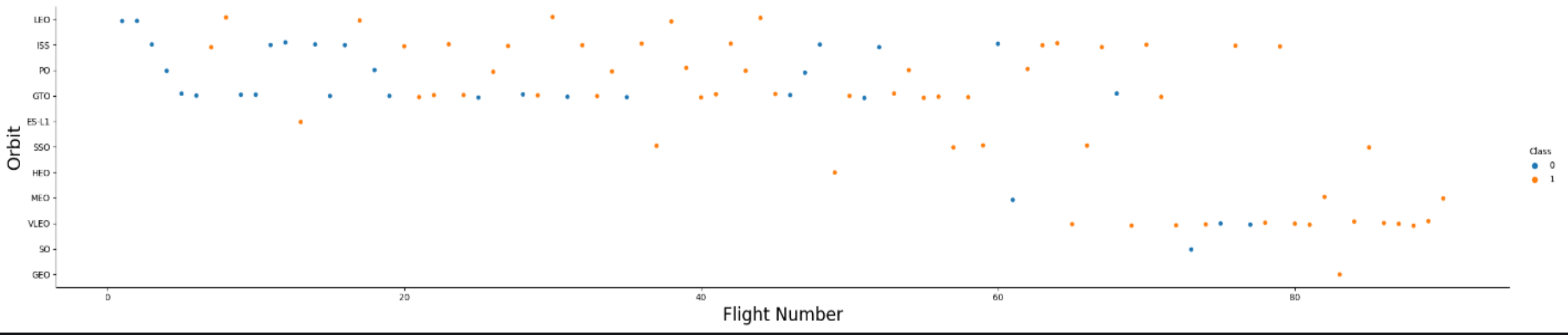
A bar chart for the success rate of each orbit type



The highest success landing was in ES-L1, GEO,HEO and SSO orbits.

Flight Number vs. Orbit Type

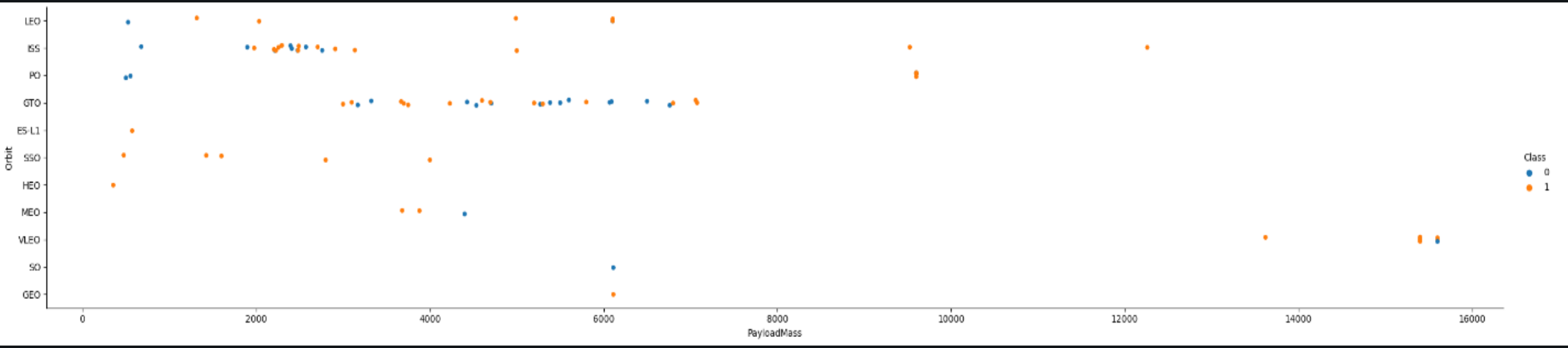
A scatter point of Flight number vs. Orbit type



Most success landings were in LEO, ISS, PO, and GTO orbits and it highly depends on the number of flights.

Payload vs. Orbit Type

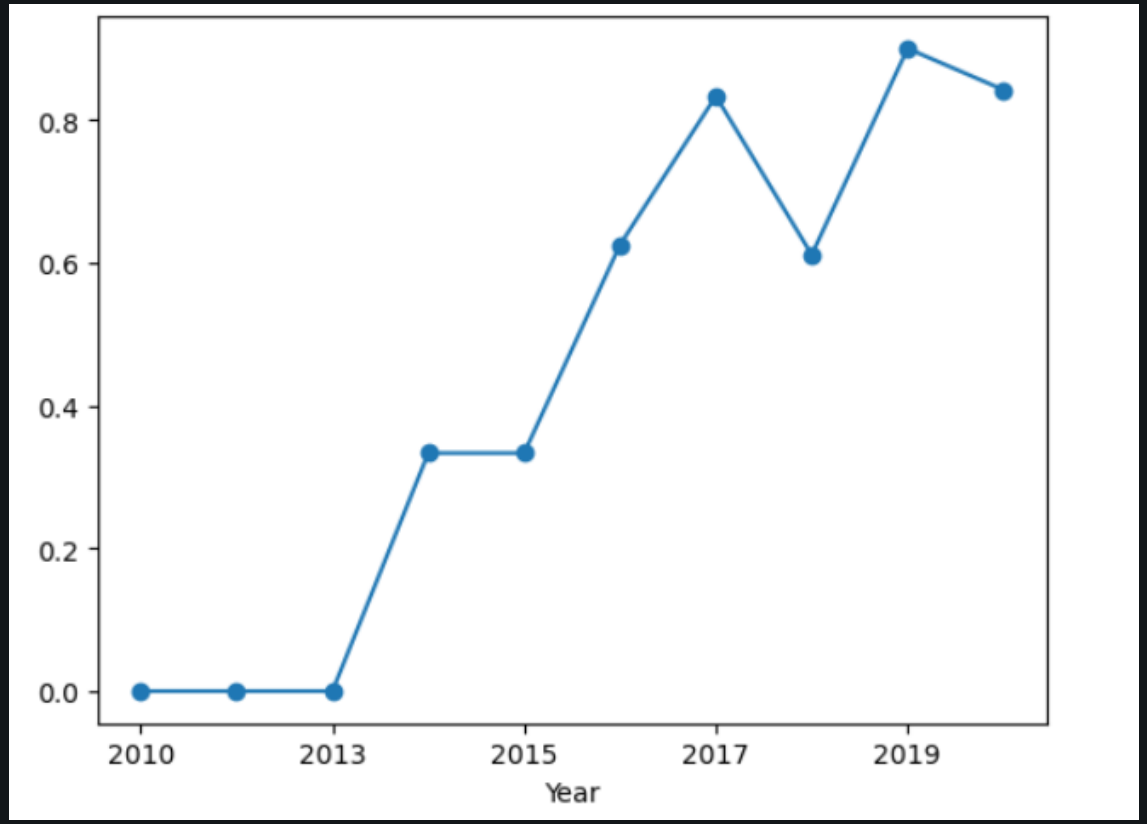
A scatter point of payload vs. orbit type



Most positive landings with heavy payloads were in Po, LEO and ISS orbits.

Launch Success Yearly Trend

A line chart of yearly average success rate



The success landing started from 2013 kept increasing till 2017. However, there was a sharp drop in 2018.

All Launch Site Names

There are 4 unique launch cites of Space X rockets

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

Launch Site Names Begin with 'CCA'

5 records where 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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	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

SUM(PAYLOAD_MASS__KG_)
48213

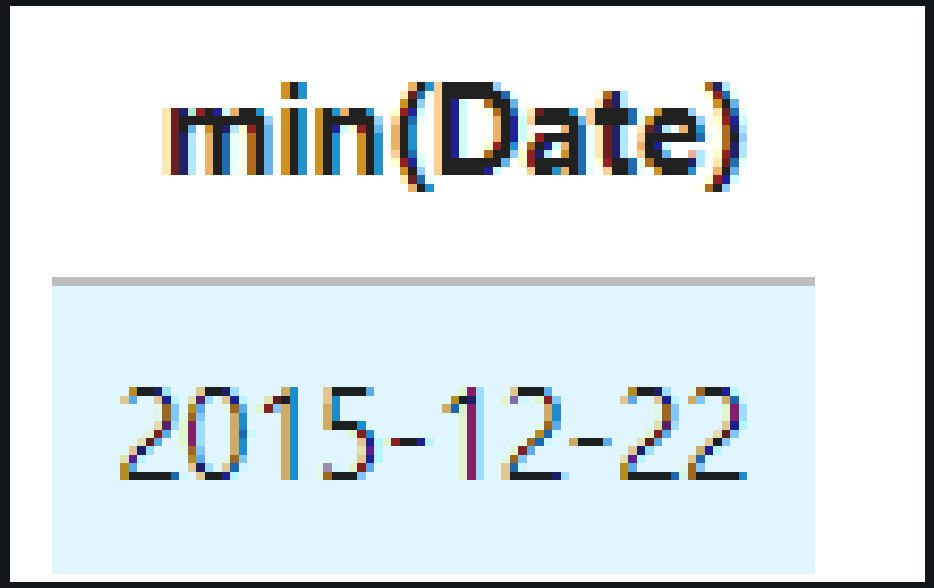
Average Payload Mass by F9 v1.1

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

AVG(PAYLOAD_MASS__KG_)
2534.6666666666665

First Successful Ground Landing Date

The date of the first successful landing outcome on ground pad



Successful Drone Ship Landing with Payload between 4000 and 6000

The names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 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

The total number of successful and failure mission outcomes

Categorized_Outcome	Outcome_Count
Failure	1
Success	100

Boosters Carried Maximum Payload

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

Booster_Version
F9 v1.0 B0003
F9 v1.0 B0004
F9 v1.0 B0005
F9 v1.0 B0006
F9 v1.0 B0007
F9 v1.1 B1003
F9 v1.1
F9 v1.1 B1011
F9 v1.1 B1010
F9 v1.1 B1012
F9 v1.1 B1013
F9 v1.1 B1014
F9 v1.1 B1015
F9 v1.1 B1016
F9 v1.1 B1018
F9 FT B1019
F9 v1.1 B1017
F9 FT B1020
F9 FT B1021.1
F9 FT B1022
F9 FT B1023.1
F9 FT B1024
F9 FT B1025.1
F9 FT B1026
F9 FT B1029.1
F9 FT B1031.1
F9 FT B1030
F9 FT B1021.2
F9 FT B1032.1
F9 FT B1034
F9 FT B1035.1
F9 FT B1029.2
F9 FT B1036.1
F9 FT B1037
F9 B4 B1039.1
F9 FT B1038.1
F9 B4 B1040.1
F9 B4 B1041.1
F9 FT B1031.2
F9 B4 B1042.1
F9 FT B1035.2
F9 FT B1036.2
F9 B4 B1043.1
F9 FT B1032.2
F9 FT B1038.2
F9 B4 B1044
F9 B4 B1041.2
F9 B4 B1039.2
F9 B4 B1045.1
F9 B5 B1046.1
F9 B4 B1043.2
F9 B4 B1040.2
F9 B4 B1045.2
F9 B5B1047.1
F9 B5B1048.1
F9 B5 B1046.2
F9 B5B1049.1
F9 B5 B1048.2
F9 B5 B1047.2
F9 B5 B1046.3
F9 B5B1050
F9 B5B1054
F9 B5 B1049.2
F9 B5 B1048.3
F9 B5B1051.1
F9 B5B1056.1
F9 B5 B1049.3
F9 B5 B1051.3
F9 B5 B1056.4
F9 B5 B1059.2
F9 B5 B1048.5
F9 B5 B1051.4
F9 B5B1058.1
F9 B5 B1049.5
F9 B5 B1059.3
F9 B5B1060.1
F9 B5 B1058.2
F9 B5 B1051.5
F9 B5 B1049.6
F9 B5 B1059.4
F9 B5 B1060.2
F9 B5 B1058.3
F9 B5 B1051.6
F9 B5 B1060.3
F9 B5B1062.1
F9 B5B1061.1
F9 B5B1063.1
F9 B5 B1049.7
F9 B5 B1058.4

2015 Launch Records

List of the failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015

substr(Date, 6,2)	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

Landing outcomes between the date 2010-06-04 and 2017-03-20, in descending order

Landing_Outcome	Outcome_Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

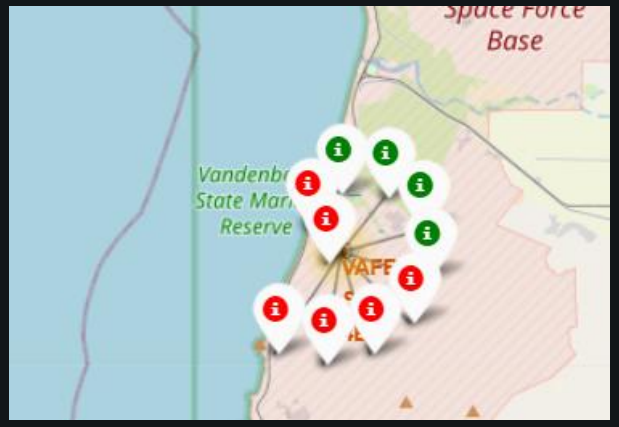
Section 3

Launch sites proximities analysis

Space X launch sites

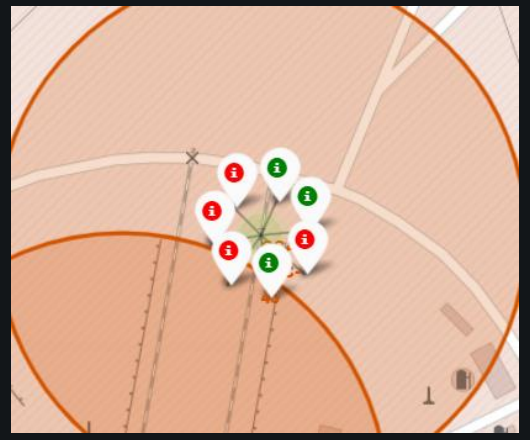
All launch sites located near coasts

Successful landings on launch sites

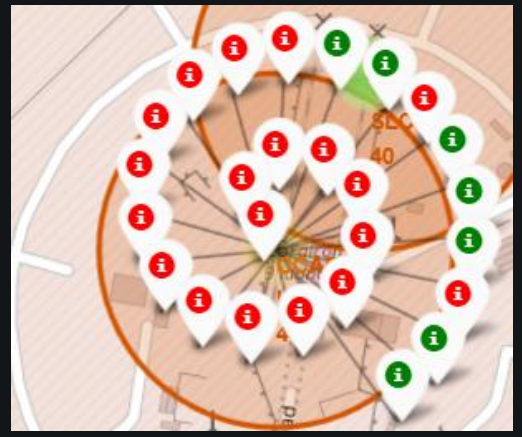


VAFB SLC-4E launch site

KSC LC-39A launch site



CAAFC SLC-40 launch site



CAAFC LC-40 launch site

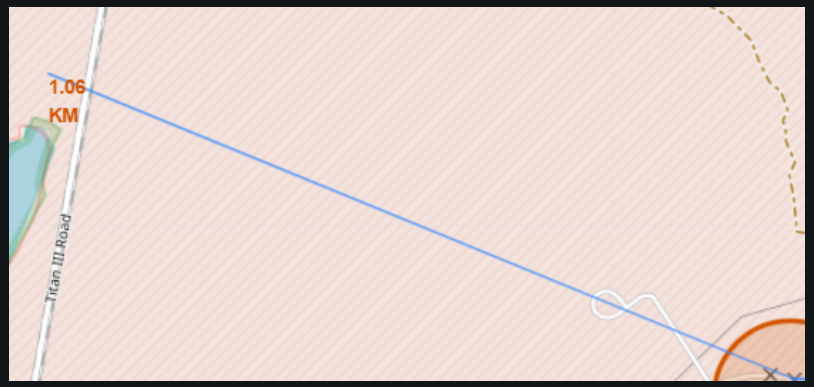
Green – positive landings

Red – unsuccessful landings

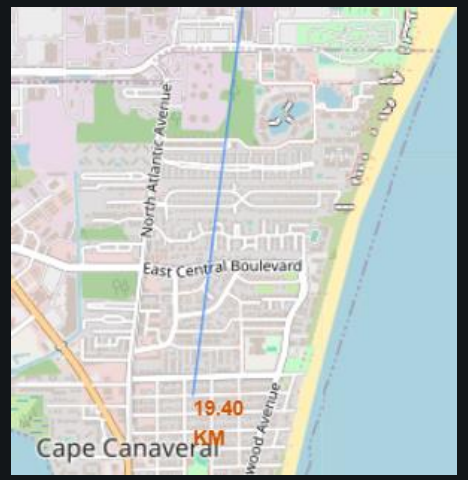
According to viz, most successful landings were in KSC LC-39A, the least in CAAFC LC-40

The distance between launch site and its proximities

The distance between CCAFS SLC-40 and nearest coastline (0.86 km) and highway (0.59 km)



The distance between CCAFS SLC-40 and nearest railroad (1.06 km)



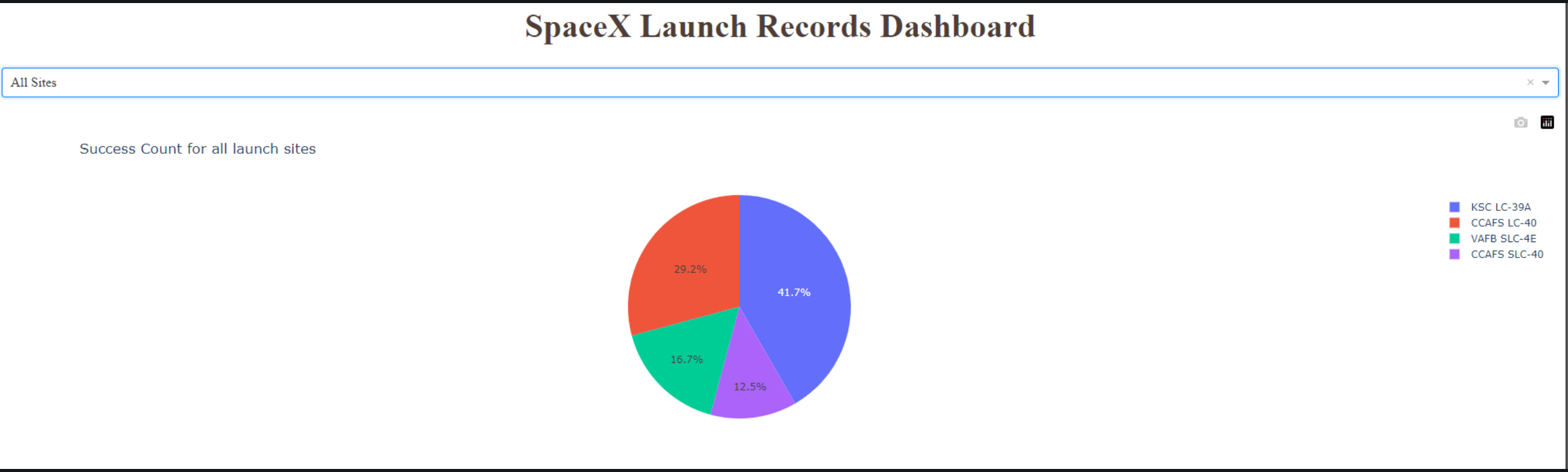
The distance between CCAFS SLC-40 and closest city (19.4 km)

CCAFS SLC-40 is near Samuel highway and coastline.
The nearest railroad called Titan III is within 1 km.
However, the closest city Cape Canaveral locates 20 km from launch site.

Section 4

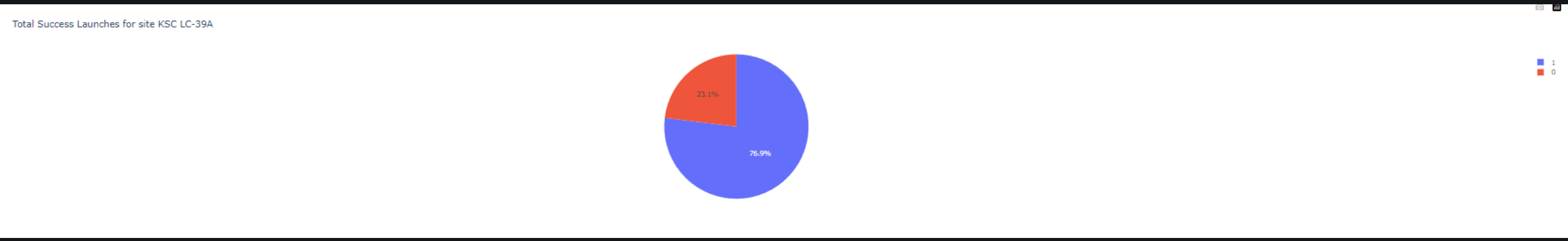
Dashboard

The proportion of positive landings in different launch sites



According to the graph, most success landings were in KSC site. While CCAFS SLC-40 was the least.

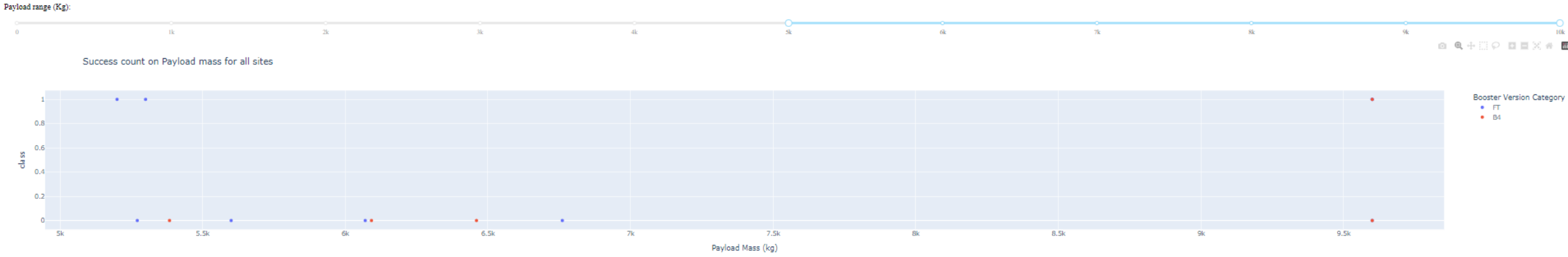
Launch site with the highest launch success rate



In KSC site around 77 % of landings were positive, and 23 % unsuccessful.

Relationship between landing outcomes and payload mass

Most success landing occurred with FT booster version in the payload mass range between 0 – 5ooo kg. While v 1.1 was most negative.



In payload mass range between 5000 – 10ooo kg, only 3 landings were successful.

Section 5

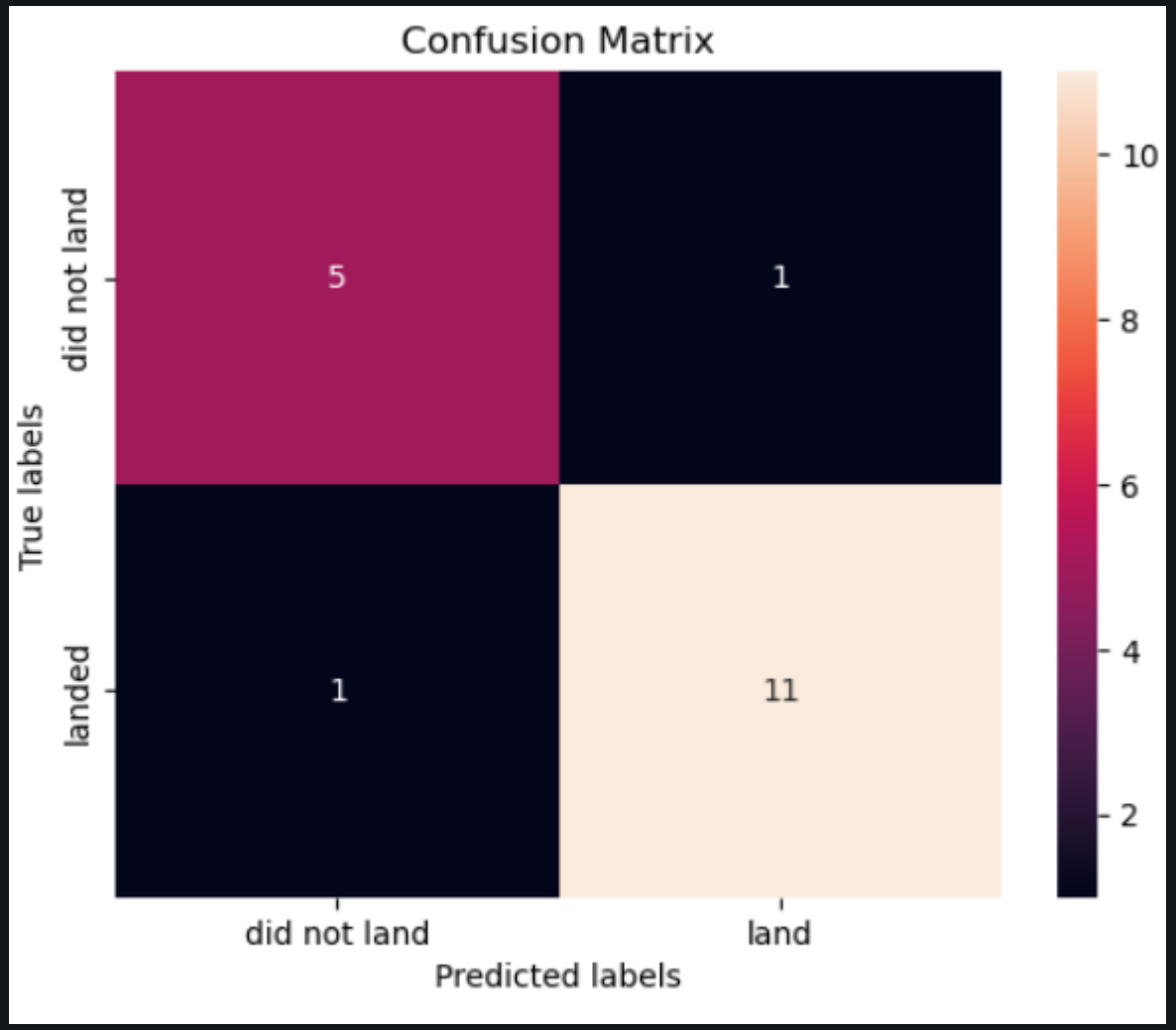
Predictive analysis

Classification Accuracy

According to evaluation metrics Decision Tree model best suited for predicting landing outcome.

	Model	Accuracy Score	Jaccard Index	F1 Score
0	Logistic Regression	0.833333	0.65000	0.777778
1	SVM	0.833333	0.65000	0.777778
2	Decision Tree	0.888889	0.78022	0.875000
3	KNN	0.833333	0.65000	0.777778

Confusion Matrix



- According confusion matrix Decision Tree can clearly distinguish between the different classes. The problem related with false positive value, where out 6 landings 5 were unsuccessful, while actually it was 1.
- True Positive – 11
- False positive- 1
- True Negative – 5
- False Negative – 1

Conclusions

- There are 4 unique launch sites and positive landings highly depends on the attempts of launching, type of orbit and payload mass. According to analysis KSC launch site most successful in terms of landing.
- Generally, every year starting from 2013 landings launch success rates increases
- In order to predict landing outcome for reducing risks of unsuccessful landings Decision Tree model can be used.

Appendix

- All code snippets can be found here: https://github.com/zhuldassov/capstone_ibm_project_space_x

Thank you!