SpaceX Falcon 9 First Stage Landing Prediction

Muhammad Nadhir Fauzananto

17.11.2021

Outline

- Abstract
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

Abstract

According to the website of SpaceX, Falcon 9 rocket launches cost $62 million, compared to rocket launches from other providers, which can cost more than $165 million each. SpaceX could save much of its cost because they can reuse the first stage. In this project, we will determine if the first stage will land. This information could be useful to determine the cost of a launch.

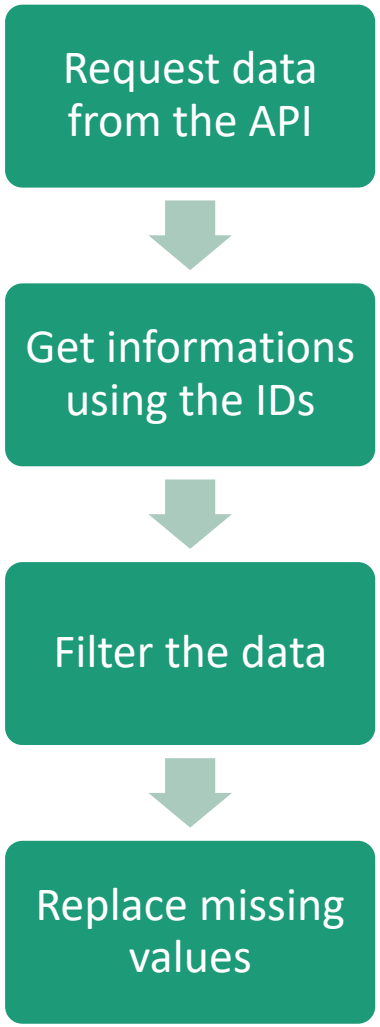
Introduction

- Background:
  - Cost of rocket launches of various providers: $165+ Mio.
  - Cost of SpaceX‘s Falcon 9 rocket launches: $62 Mio.
    - Reusable first stage rocket -> cheaper cost.
    - Not all rockets landed successfully.
- Objective:
  - Predict the likelihood of Falcon 9 first stage rocket landing successfully.
  - Predict the cost of a launch.

Methodology

Data Collection: SpaceX API

- Request rocket launch data from SpaceX API.
  - Url: https://api.spacexdata.com/v4/launches/past
- Get informations using the IDs for each launch.
  - „rocket“: Booster name.
  - „payload“: Mass of the payload and target orbit.
  - „launchpad“: Name and coordinates of the launch site.
  - „cores“: Informations about the landing (outcome, number of flights, etc.).
- Filter the data.
  - Only show informations about Falcon 9 launches.
- Replace missing values for payload mass.
  - Use the average of the payload mass.



Data Collection: Web Scraping

- Request Falcon 9 launch data from Wikipedia.
  - Url: https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches
- Get all column and variable names.
  - Extract from the HTML table header.
- Create the dataframe from the extracted data.
  - Create an empty dictionary.
  - Column names from the previously extracted column names.
  - Fill the dictionary with the extracted launch records.
  - Create a dataframe from said dictionary.

Data Wrangling

- Calculate the number of launches on each site.
  - Use .value_counts() on the column „LaunchSite“.
  - A total of 90 launches.
  - Performed mostly at Cape Canaveral Space Launch Complex 40 (CCAFS SLC-40).
- Calculate the number and occurance of each orbit.
  - Use .value_counts() on the column „Orbit“.
  - Most of the orbits are geosynchronous orbit (GTO) and ISS orbit.
- Calculate the number and occurance of mission outcome per orbit type.
  - Use .value_counts() on the column „Outcome“.
  - Most of the successful landings are on a drone ship (41 out of 90 total landings).
- Create a landing outcome label from „Outcome“ column.
  - Determine, whether a landing is successful or not.
  - 0: Landing failed 1: Landing successful
  - Average success rate: ~66,7%

Exploratory Data Analysis (EDA) with SQL

- Use „select ... from ...“ to display the requested informations.
- Use „limit“ to limit the amount of the displayed informations.
- Use „where ... like ...“ and „where ... = ...“ to display informations from the data containing the specified parameter(s)
- Use „between ... and ...“ to display numeric informations between two parameters.
- Use subquery to display informations with specific parameters.
- Use „order by“ to rank the displayed informations by the specificed parameter.
- GitHub link:

EDA with Visualization

- Scatter plot:
  - To determine the relationship between flight number and launch site
  - To determine the relationship between payload mass and launch site
  - To determine the relationship between flight number and orbit type
  - To determine the relationship between payload mass and orbit type
- Bar chart:
  - To determine the relationship between success rate of each orbit type
- Line chart:
  - To determine the launch success yearly trend

Interactive Map with Folium

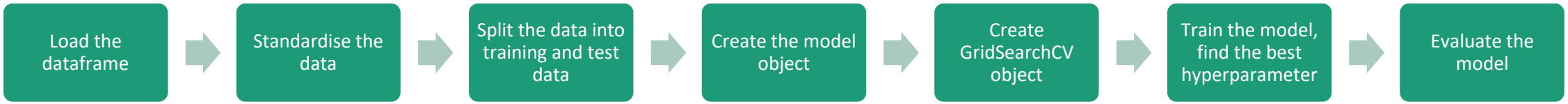
- Map Objects:
  - Circle: Location of launch site.
  - Marker: Label for the launch site.
  - Line: Distance from launch site to nearest coastline / railway / highway.
- The launch sites are located some distance away from cities
  - Closest city (Melbourne) is 51.62 km away
- The launch sites are located near highways and railways (<1 km away

Dashboard with Plotly and Dash

- Pie chart: Shows the launch result from different sites
  - To find out, which launch site has the highest success chance
- Scatter chart: Relationship between payload mass, success chance and booster version
  - To find out, which booster version has the highest chance of success if launched.
  - To find out, which payload mass range has the highest chance of success.
  - To find out, which payload masses are launched most often.

Predictive Analysis (Classification)

- GridSearchCV is used to find, which parameters work the best for each model:
  - Logistic regression
  - Decision tree
  - Support vector machine
  - K nearest neighbours
- Model evaluation:
  - Model accuracy
  - Confusion matrix



Results

Launch Sites

All Launch Site Names

- Query: %sql select distinct launch_site from spacex
- CCAFS SLC 40: Cape Canaveral Space Launch Complex 40
  - Coordinates: 28.563197, -80.576820
- CCAFS LC-40: Cape Canaveral Launch Complex 40
  - Coordinates: 28.5623, -80.57736
- VAFB SLC 4E**:** Vandenberg Air Force Base Space Launch Complex 4E
  - Coordinates: 34.63285, -120.61079
- KSC LC 39A: Kennedy Space Center Launch Complex 39A
  - Coordinates: 28.57327, -80.6469

Launch Site Names Beginning with „CCA“

- Query: %sql select * from spacex where launch_site like '%CCA%' limit 5
- CCAFS SLC 40: Cape Canaveral Space Launch Complex 40
  - Coordinates: 28.563197, -80.576820
- CCAFS LC-40: Cape Canaveral Launch Complex 40
  - Coordinates: 28.5623, -80.57736
- Total number of successful and failed mission outcomes: 99

Landing and Mission Outcome

First Succesful Ground Landing Date

- %sql select min(date) from spacex where landing__outcome like '%Success (ground pad)%'
  - First successful ground landing date: 2015-12-22

Total number of successful and failed mission outcomes:

- %sql select count(mission_outcome) from spacex where mission_outcome = 'Success’
  - Total number: 99 missions

Booster Version – Successful Drone Ship Landing – Payload between 4000 and 6000 kg

- %sql select booster_version from spacex where landing__outcome like '%Success (drone ship)%' and payload_mass__kg_ between 4000 and 6000

Booster Version	F9 FT B1022	F9 FT B1026	F9 FT B1021.2	F9 FT B1031.2

Payload Mass

Total Payload Mass – NASA (CRS)

- Query: %sql select sum(payload_mass__kg_) from spacex where customer like 'NASA (CRS)’
- Total payload mass carried by booster launched by NASA (CRS): 45596 kg

Average Payload Mass of F9 v1.1

- Query: %sql select avg(payload_mass__kg_) from spacex where booster_version like 'F9 v1.1%’
- Average payload mass carried by booster version F9 v1.1: 2534 kg

Booster Carried Maximum Payload

- Query: %sql select booster_version from spacex where payload_mass__kg_ = (select max(payload_mass__kg_) from spacex)

Booster Version	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

- Query: %sql select landing__outcome, booster_version, launch_site, date from spacex where date like '%2015%' and landing__outcome = 'Failure (drone ship)’

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

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

%sql select landing__outcome, date from spacex where date between '2010-06-04' and '2017-03-20' order by date desc

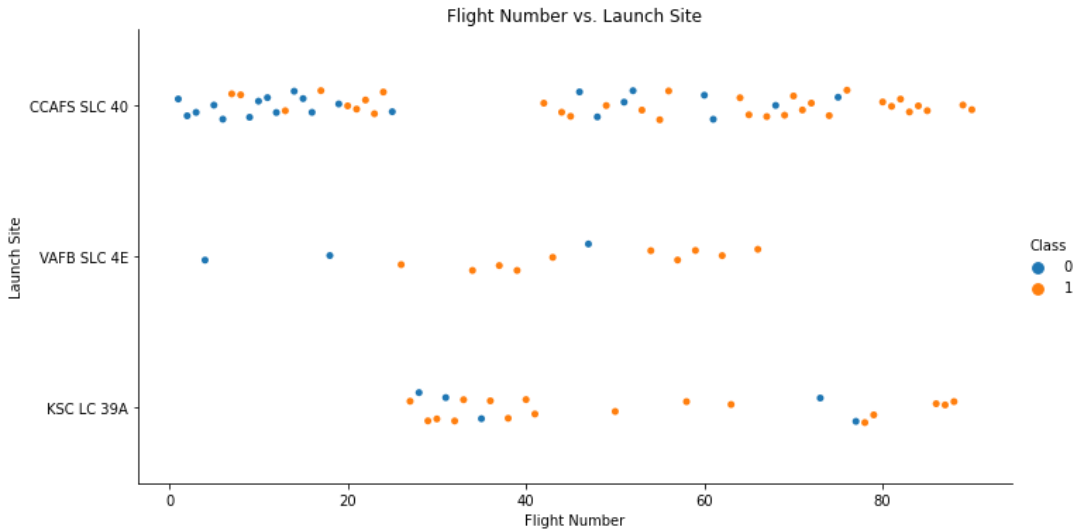
landing__outcome	DATE
No attempt	2017-03-16
Success (ground pad)	2017-02-19
Success (drone ship)	2017-01-14
Success (drone ship)	2016-08-14
Success (ground pad)	2016-07-18
Failure (drone ship)	2016-06-15
Success (drone ship)	2016-05-27
Success (drone ship)	2016-05-06
Success (drone ship)	2016-04-08
Failure (drone ship)	2016-03-04
Failure (drone ship)	2016-01-17

landing__outcome	DATE
Success (ground pad)	2015-12-22
Precluded (drone ship)	2015-06-28
No attempt	2015-04-27
Failure (drone ship)	2015-04-14
No attempt	2015-03-02
Controlled (ocean)	2015-02-11
Failure (drone ship)	2015-01-10
Uncontrolled (ocean)	2014-09-21
No attempt	2014-09-07
No attempt	2014-08-05
Controlled (ocean)	2014-07-14
Controlled (ocean)	2014-04-18
No attempt	2014-01-06

landing__outcome	DATE
No attempt	2013-12-03
Uncontrolled (ocean)	2013-09-29
No attempt	2013-03-01
No attempt	2012-10-08
No attempt	2012-05-22
Failure (parachute)	2010-12-08
Failure (parachute)	2010-06-04

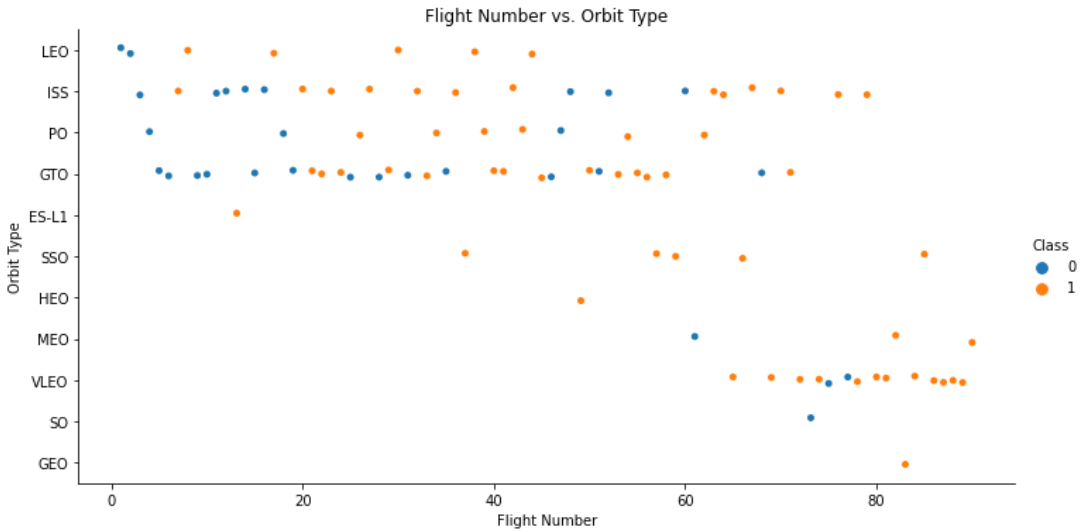
Flight Number

Flight Number vs. Launch Site



- The success rate of the launches increases as the flight number increases.
  - Later launches are more likely to succeed.

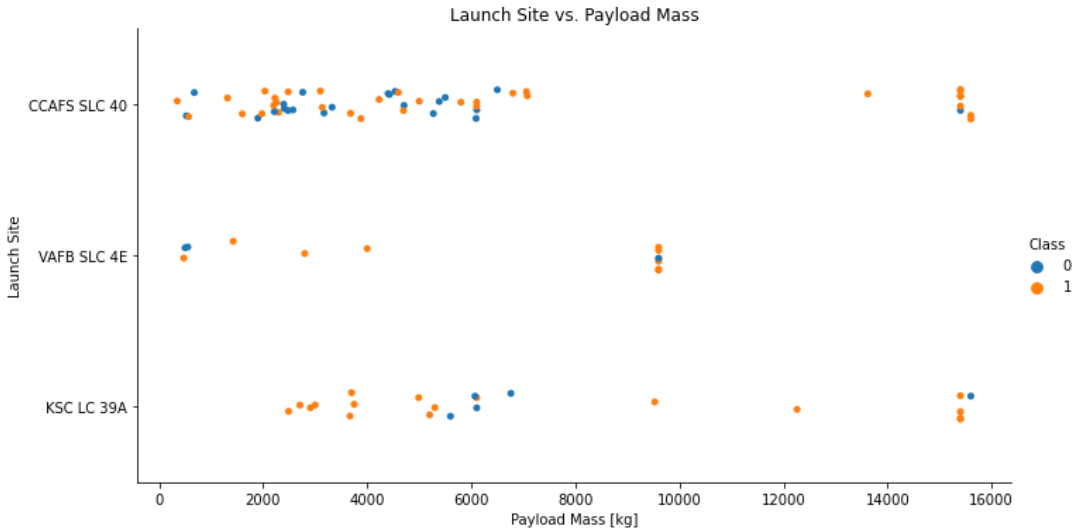
Flight Number vs. Orbit Type



- Success rate of launches to LEO orbit increases as the flight number increases.
- Most of the recent launches are launched to VLEO orbit.

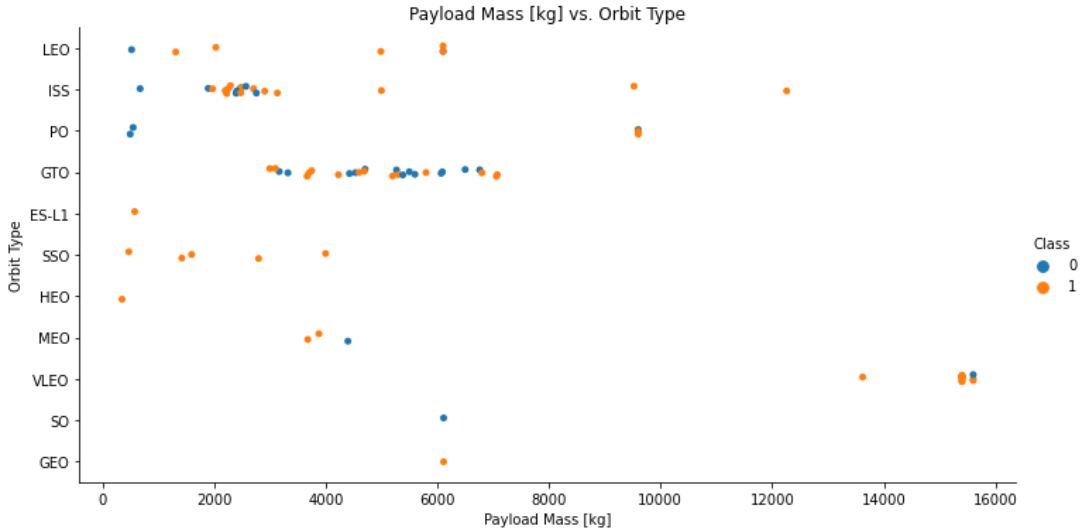
Payload Mass

Payload vs. Launch Site



- No launches for heavy payload (> 10000 kg) from VAFB SLC 4E site.

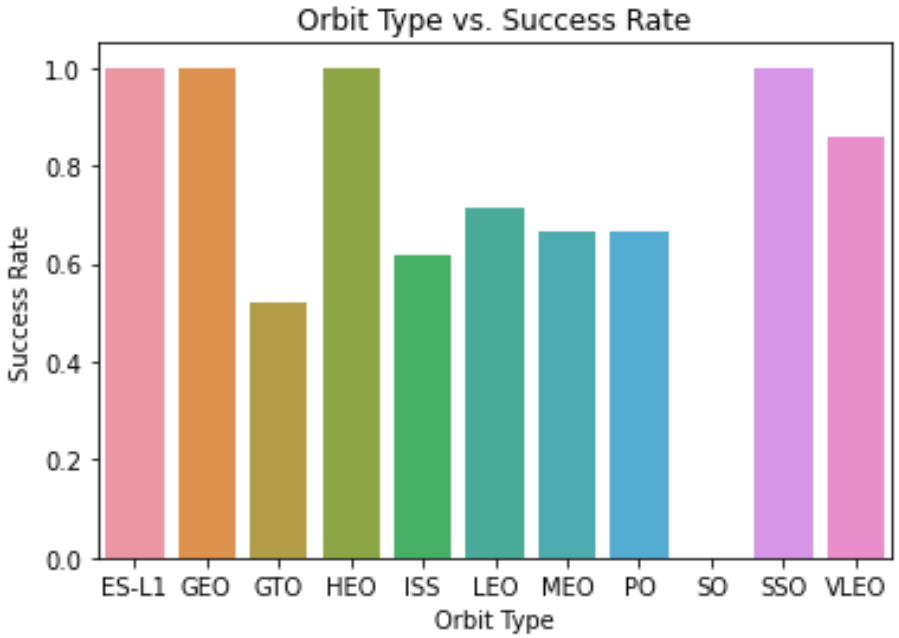
Payload vs. Orbit Type



- Heavy payloads are launched to ISS, VLEO and Polar orbit.

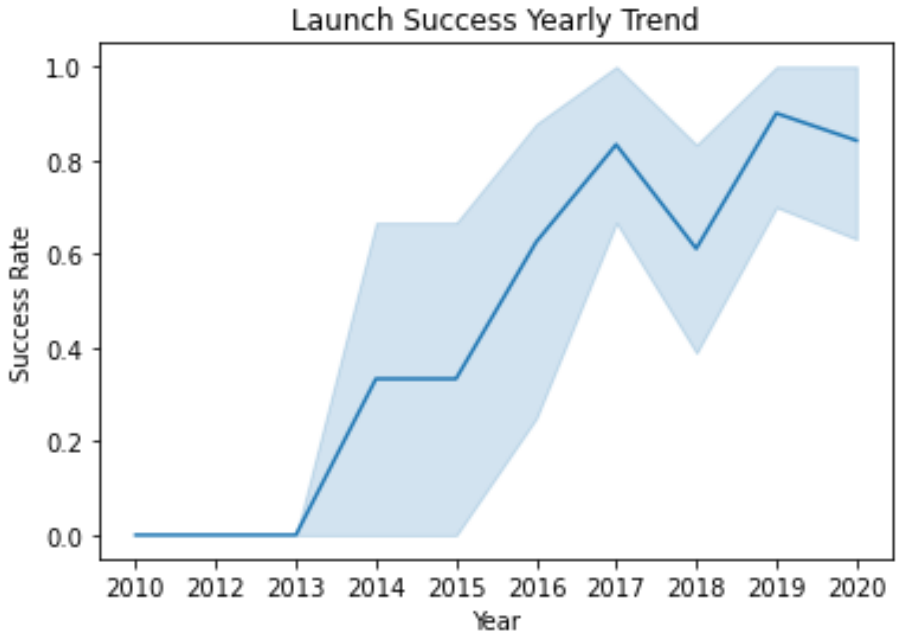
Success Rate

Success Rate vs. Orbit Type



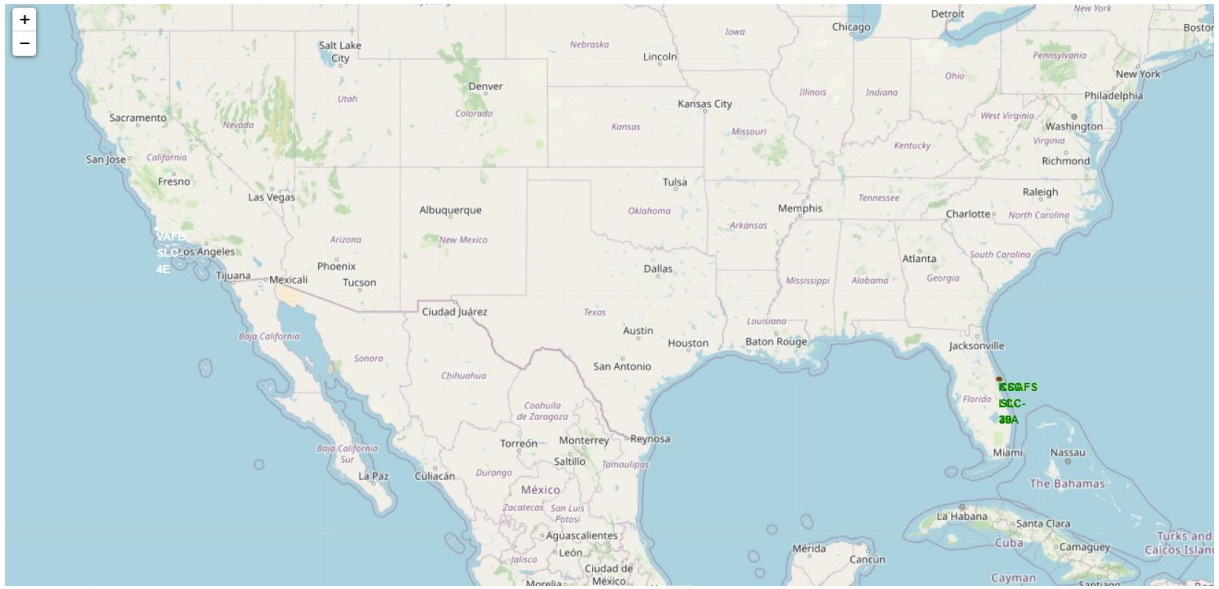
- Launches to ES-L1 orbit, GEO, HEO and SSO have 100% success rate, but:
  - There are only 1 launch to ES-L1, GEO and HEO each

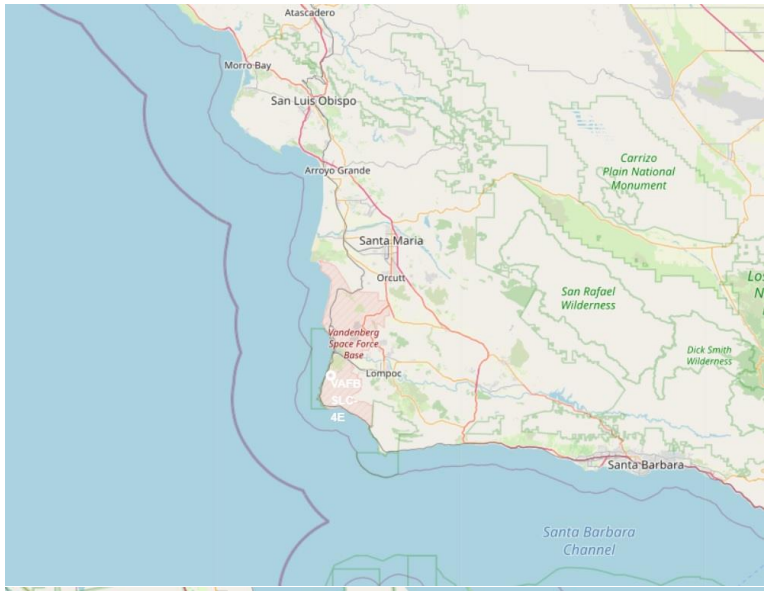
Launch Success Yearly Trend

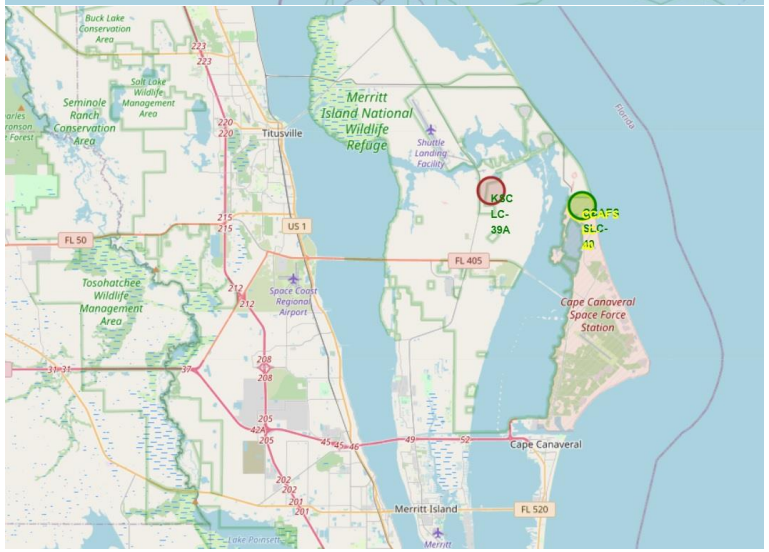


- Success rate of launches has been increasing since 2015 with a dip in 2018
- The success rate is above 50 % since 2016

Folium Map

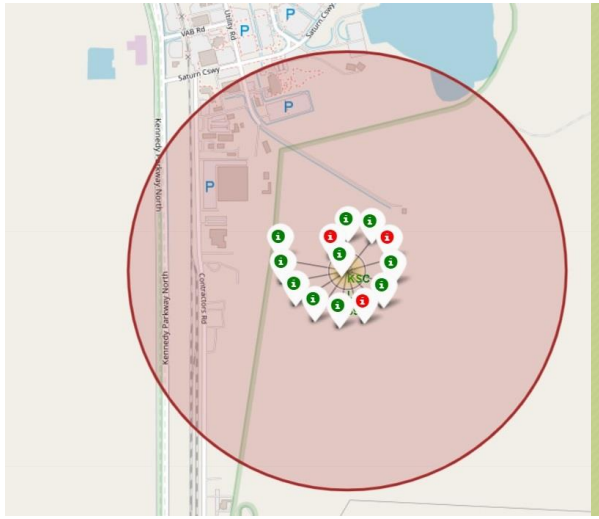






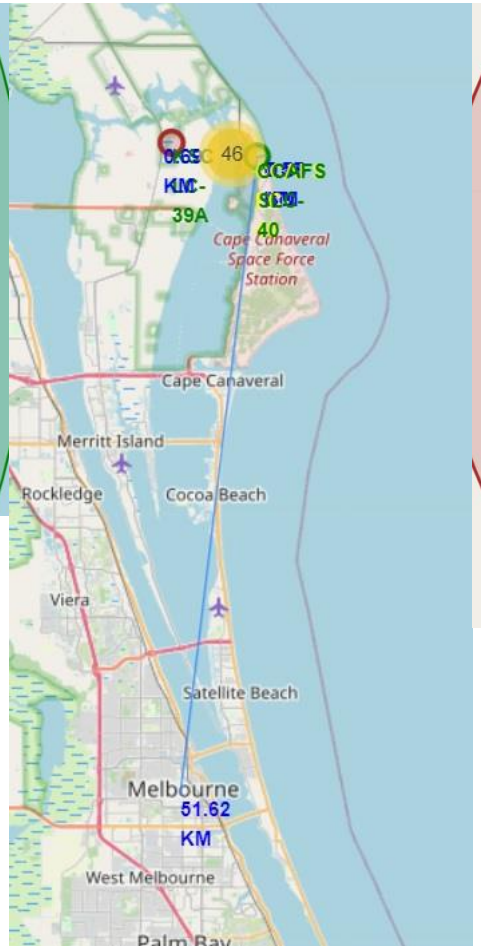
- All launch sites are located:
  - near the coast
  - as south as possible
  - as far away from the city as possible

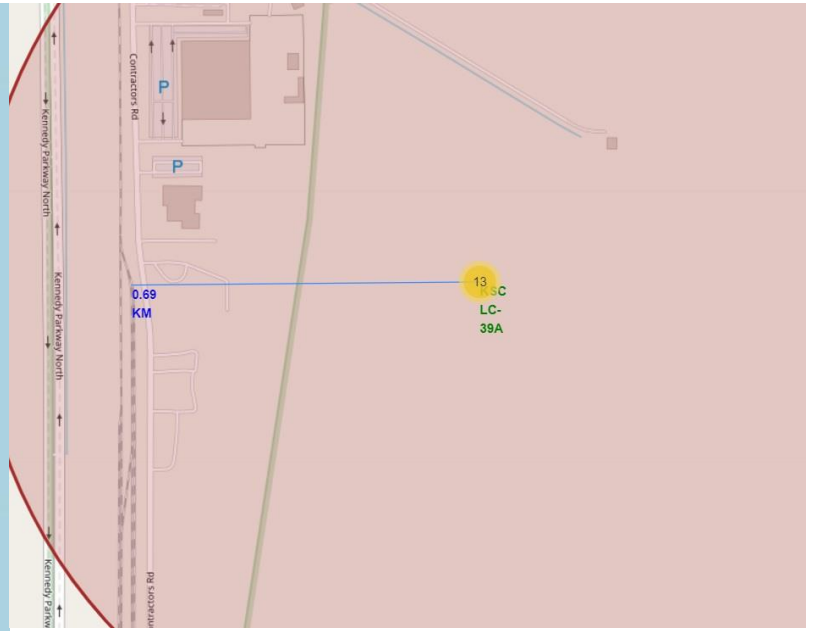
Folium Map



- Launch site at KSC LC 39A (Kennedy Space Center Launch Complex 39A) has the highest success rate compared to others.

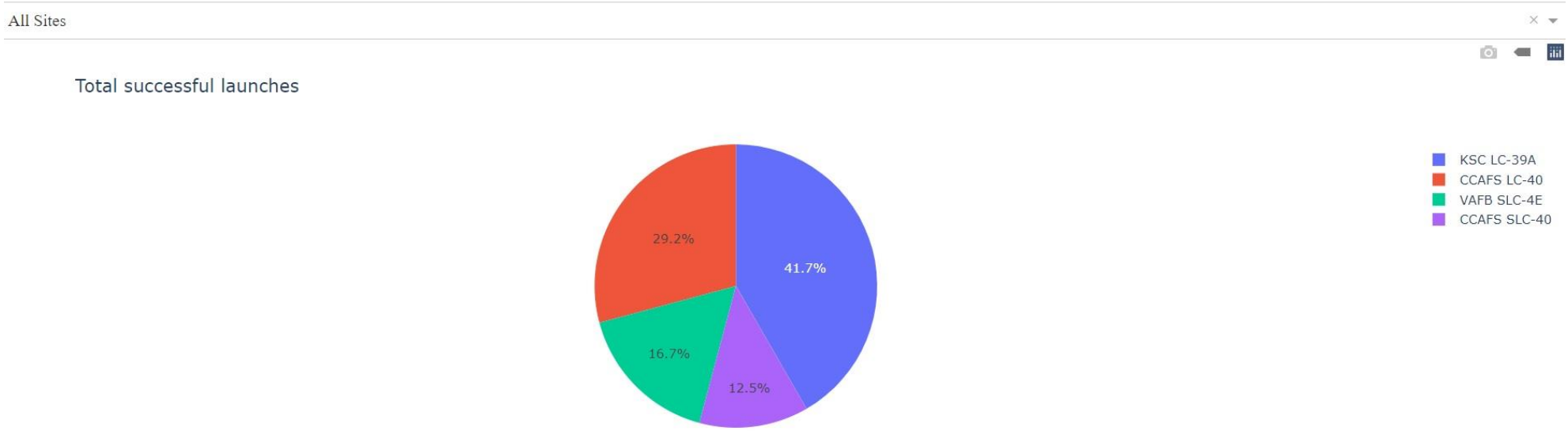
Folium Map





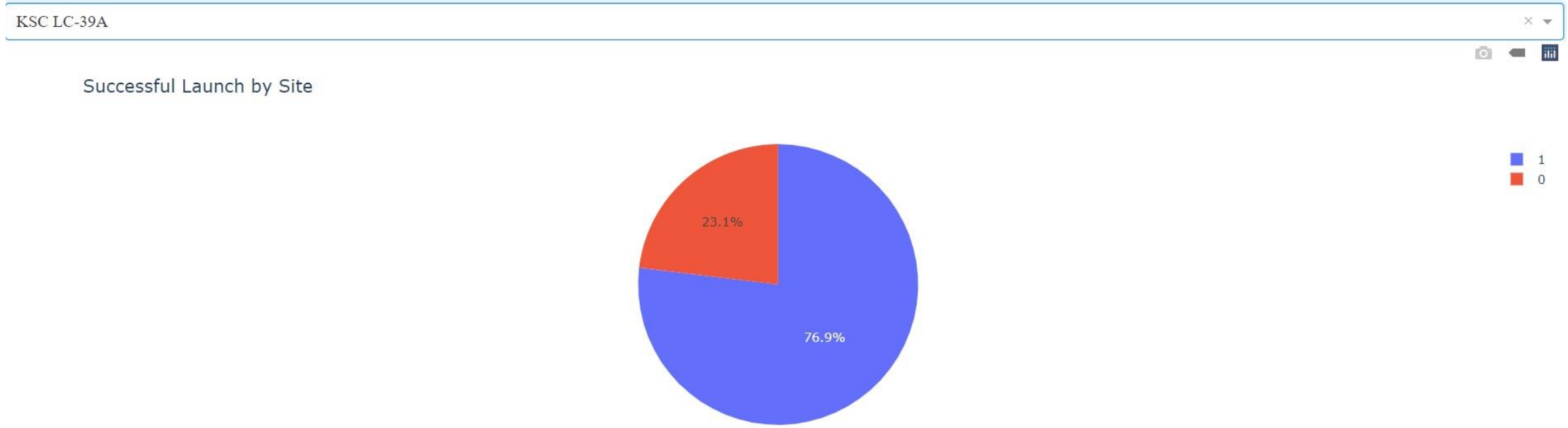
- The launch sites are located:
  - at least 50 km away from the nearest city
  - near a railway
  - near a highway
  - near the coast

Dashboard with Plotly and Dash



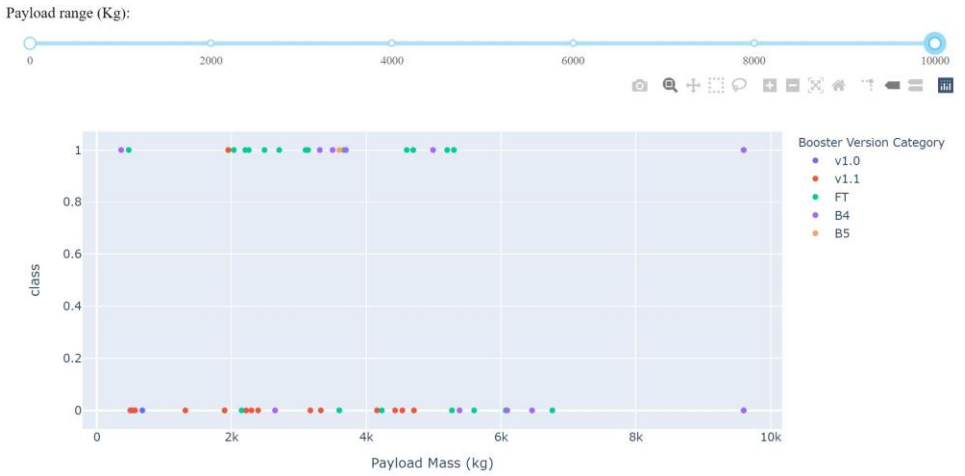
- Launches from KSC LC-39A (Kennedy Space Center Launch Complex 39A) have the highest success rate
  - 41.7 % of all successful launches are from there.

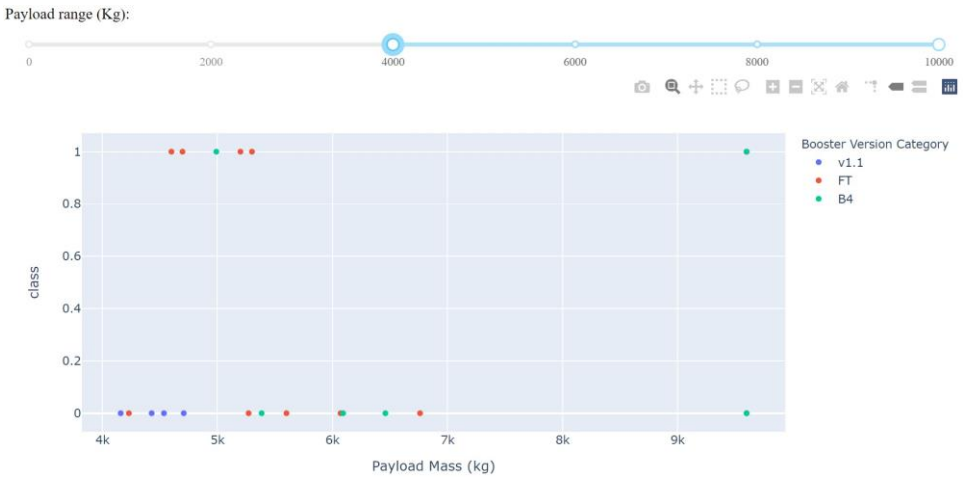
Dashboard with Plotly and Dash

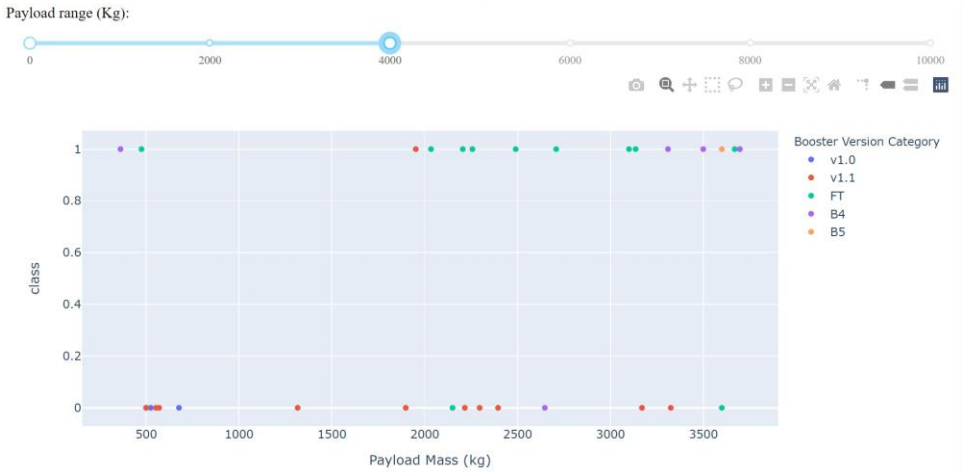


- 76.9 % of launches from KSC LC-39A are successful.
  - Highest success rate compared to other sites.

Dashboard with Plotly and Dash

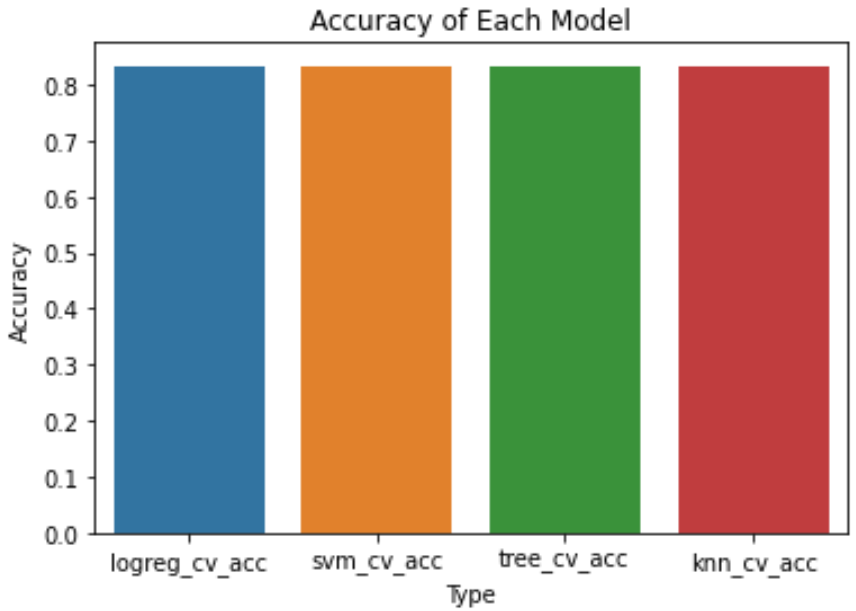




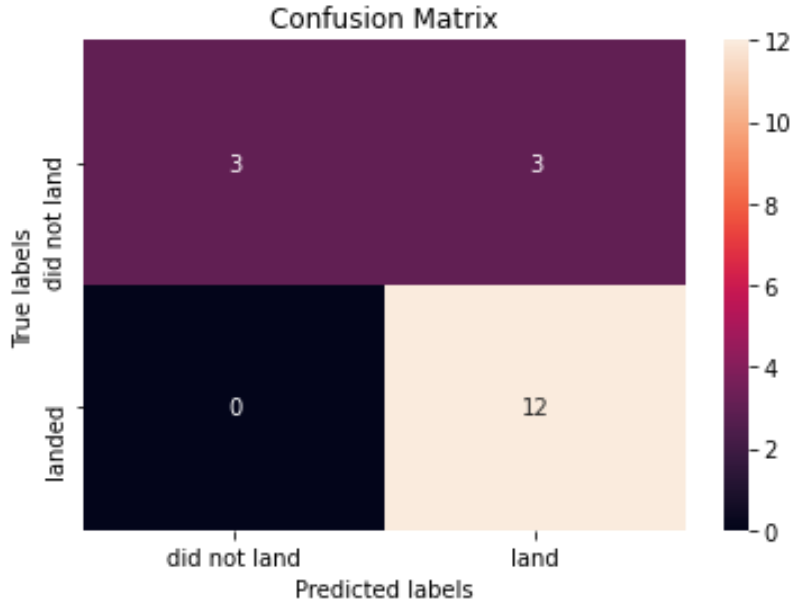


- Most successful booster version: FT (15 success out of 23 total).
- The success rate decreases as the payload mass increases.
- Payload between 2000 and 4000 kg have the highest success rate.

Predictive Analysis



- All models have the same accuracy (83 %)



- There are 12 succesful landing attempts.
  - The model correctly predicted, that all 12 will land
- There are 6 unsuccesful landing attempts.
  - However, the model falsely predicted, that 3 of them will land.
- There are 3 false positives and 0 false negative

Conclusions and Appendix

Conclusions

- 76.9% of launches from KSC LC-39A (Kennedy Space Center Launch Complex 39A) are successful.
  - Highest success rate compared to the other sites.
- All launch sites are located near to the coast, railway and highway while located a good distance away from the nearest city (> 50 km).
- The heavier payloads (> 10000 kg) are launched to ISS, VLEO and Polar orbit.
- Most successful booster version: FT (15 success out of 23 total).
- Payloads between 2000 and 4000 kg have the highest success rate.
  - The success rate decreases as the payload mass increases.

Appendix

- Github link: https://github.com/chessf/SpaceX-Falcon-9-First-Stage-Landing-Prediction.git