IBM Data Science Capstone Project

Schäfer, Dennis

November 01st, 2022

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

- Methodologie
  - Data Collection
  - Data Wrangling
  - Data Analysis
  - EDA with SQL
  - Predictive Analysis
- Results
  - Exploratory Data Analysis with Data Visualization
  - Interactive map with Folium
  - Predictive Analysis

Introduction

Project backround

- We are working for the company SpaceY and are analysing the data of the company SpaceX adverstised on its website
- The Falcon 9 rocket launches costs around 100 million dollars less throught SpaceX can reuse the first stage
- The goal is to predict if the Falcon 9 will land successfully

Problems

- What influences the success of the rocket landing?
- Under what conditions does the rocket launch has the best success rate?
- What is the best model to predict further rocket launches

Methodology

Data Collection

- Web Scrapping Wikipedia using BeautifulSoup
- SpaceX REST API
  - Using the following URL: https://api.spacexdata.com/v4/launches/past
  - Provides data about rocket launches including detailed information
- Normalize the data into a .csv file for further analysis

Data Wrangling

- Convert a succeed and failed landing with numeric for futher steps
- Create a new column named „Class“
- If the landing succeed assign 1 to the column „Class“, otherwise assign 0

Data Analysis

- Exploratory Data Analysis with Data Visualization
- Using scatter graphs, bar graph and line graph
- Creating an interactive map with Green and Red markers assigned from the column „Class“
- Creating an interactive dashboard to access filtered information

Predictive Analysis

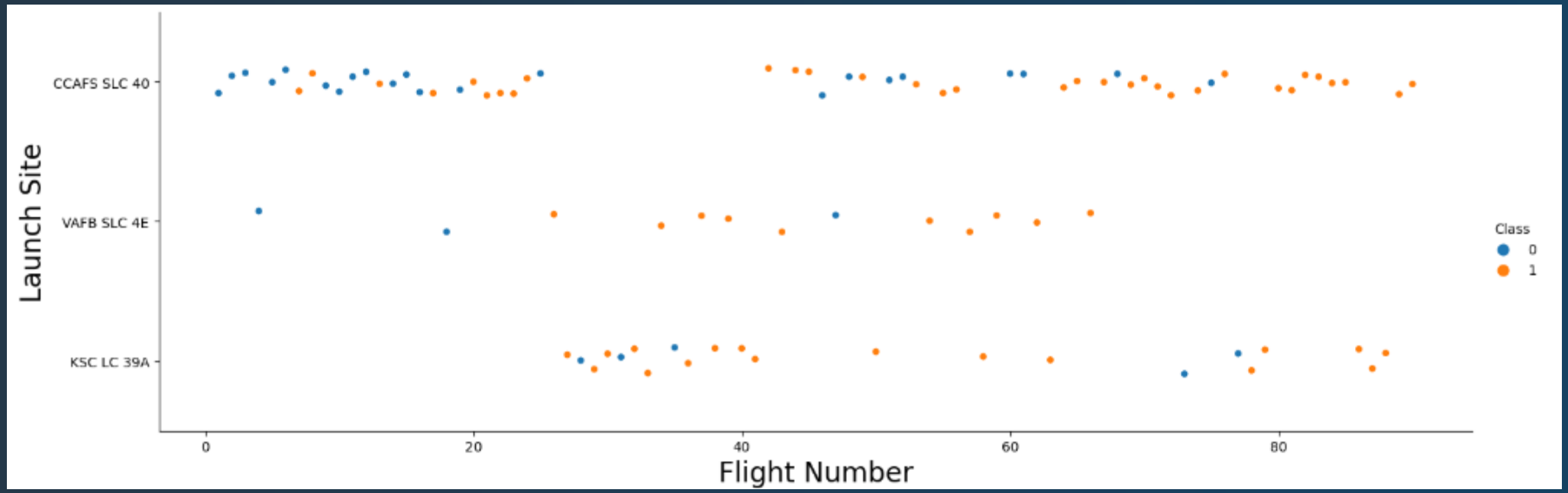
- Building the model with a 20% test data set
- Check the accuracy for each model
- Plot Confusion Matrix

Results

Exploratory Data Analysis with Data Visualization

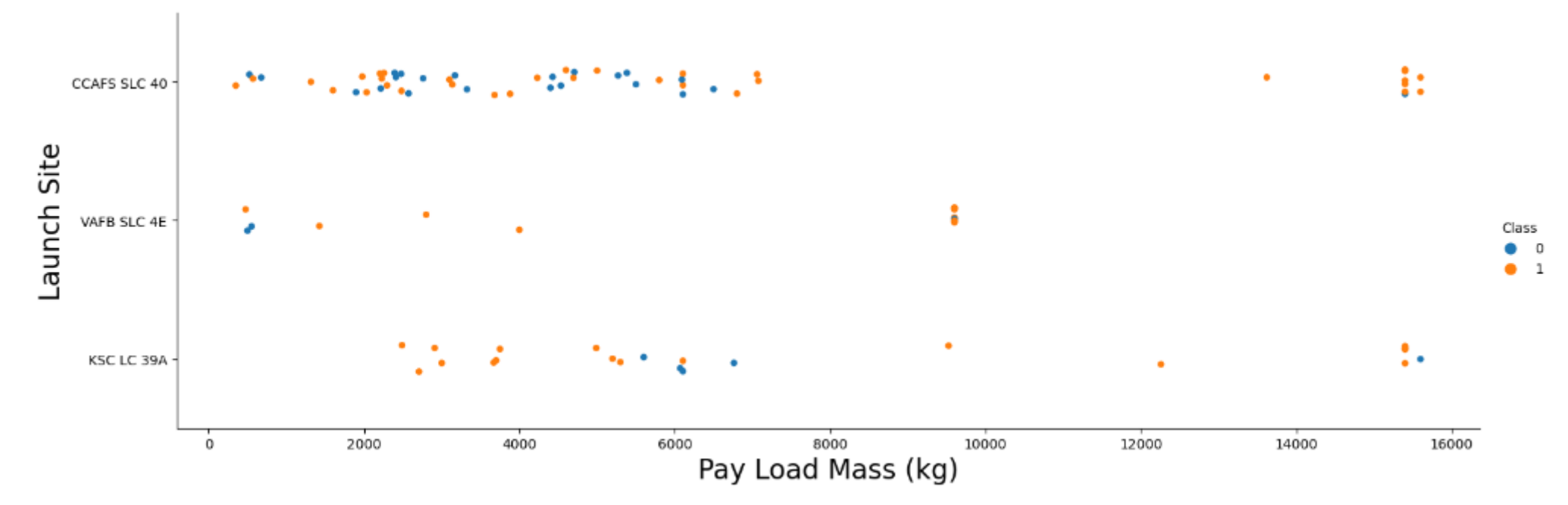
Launch Site vs Flight Number

- The the flight number scales with the success rate
- CCAFS SLC 40 has a lower success rate



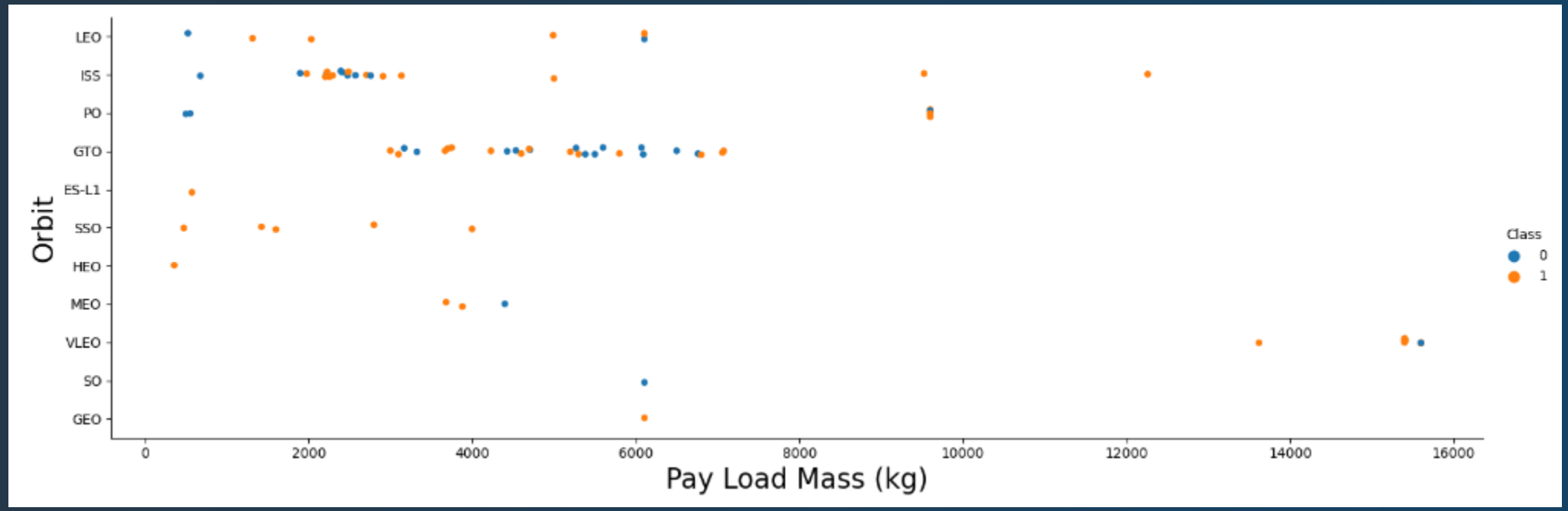
Launch Site vs Pay Load Mass (kg)

- Higher Pay Load Mass has a higher success rate
- For KSC LC 39A a very low Pay Load Mass has a higher success rate



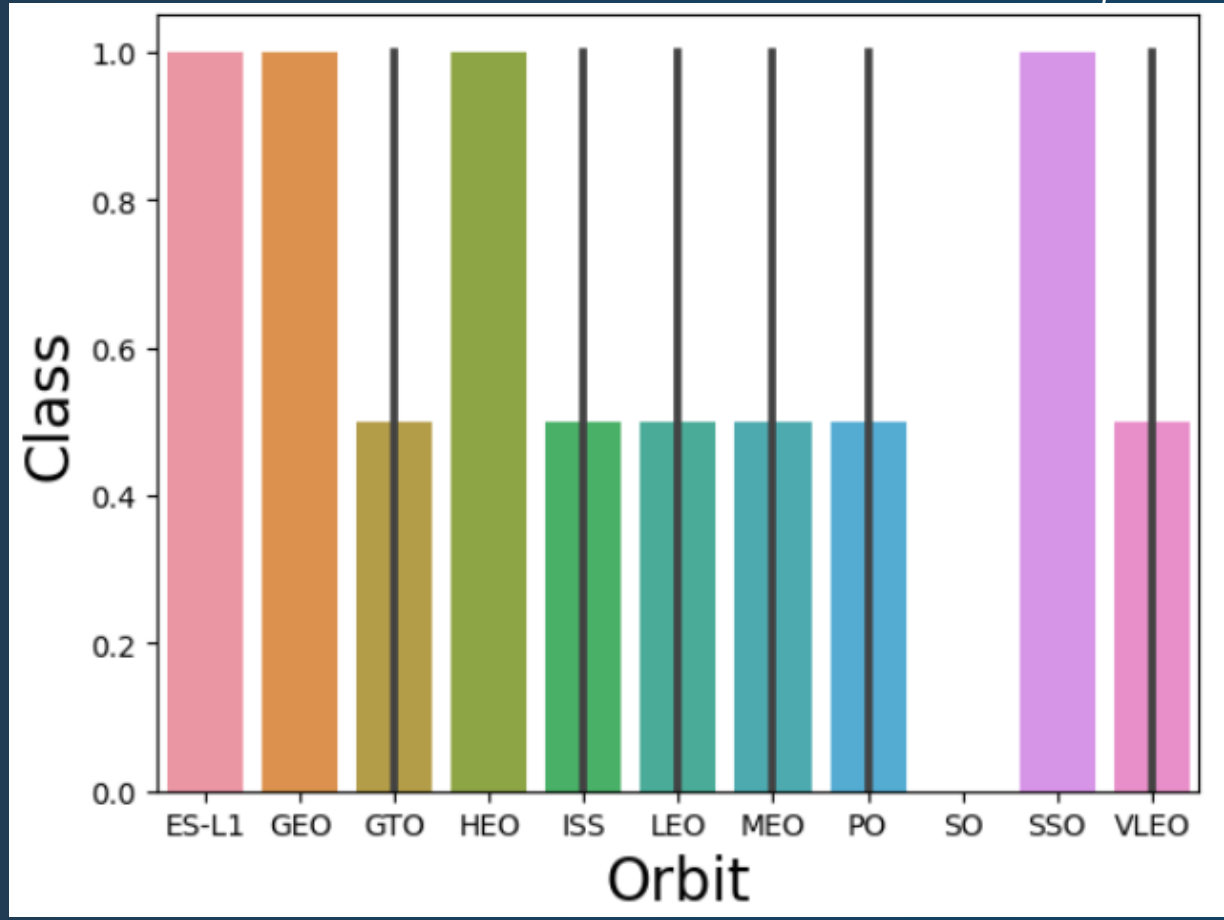
Orbit vs Pay Load Mass (kg)

The Pay Load Mass inbetween one Orbit does not have a significant difference.



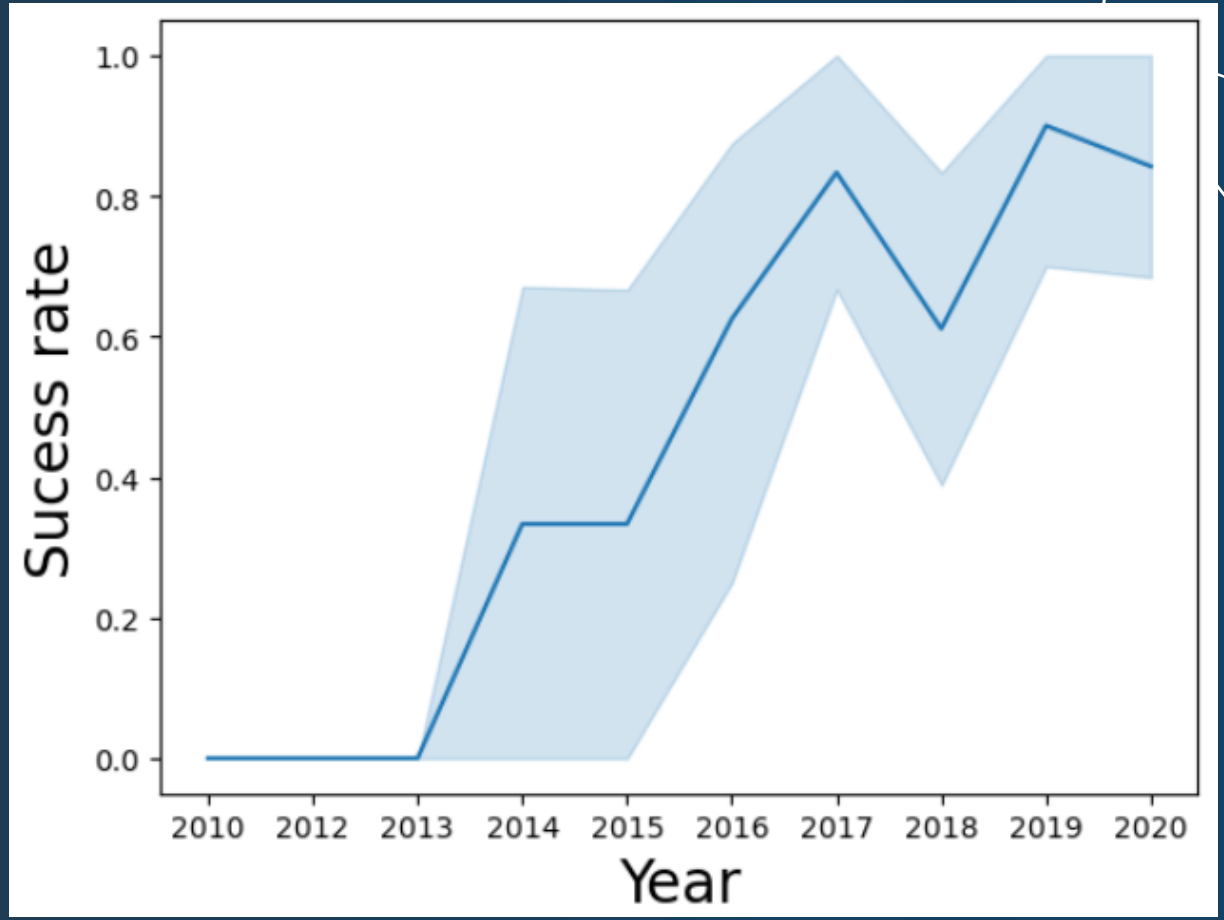
Mean of the Orbits success rate

- ES-L1, GEO, HEO and SSO has the best success rate
- SO has a very poor success rate



Success rate vs Year

The success rate is increasing nearly till 2020.



EDA with SQL

Unique launch sites

%sql SELECT

DISTIONCT(LAUNCH_SITE)

FROM SPACEXTBL

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

5 records with „CCA“ launch sites beginning

%sql SELECT * FROM SPACEXTBL WHERE LAUNCH_SITE LIKE „CCA%“ LIMIT 5

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

Total mass carried by booster launched by NASA (CRS)

```
%sql SELECT SUM(PAYLOAD_MASS_KG)
FROM SPACEXTBL
WHERE CUSTOMER = „NASA (CRS)“
```

1
45596

Average payload mass carried by boster version F9 v1.1

```
%sql SELECT AVG(PAYLOAD_MASS_KG)
FROM SPACEXTBL
WHERE BOOSTER_VERSION = „F9 v1.1“
```

1
2928.400000

Date of the first successful landing outcome in ground pad was acheived

```
%sql SELECT MIN(DATE)
FROM SPACEXTBL
WHERE LANDING__OUTCOME =
„SUCCESS (ground pad)“
```

1
2015-12-22

Booster versions with successes and a payload mass between 4000 and 6000

```
%sql SELECT BOOSTER_VERSION
FROM SPACEXTBL
WHERE LANDING__OUTCOME =
„SUCCESS (drone ship)“
AND PAYLOAD_MASS_KG > 4000
AND PAYLOAD_MASS_KG < 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

```
%sql SELECT COUNT(MISSION_OUTCOME)
FROM SPACEXTBL
WHERE MISSION_OUTCOME = „Success“
OR MISSION_OUTCOME = „Failure (in flight)“
```

1
100

Booster_versions which has carried maximum payload mass

```
%sql SELECT BOOSTER_VERSION
FROM SPACEXTBL
WHERE PAYLOAD_MASS_KG =
(SELECT MAX(PAYLOAD_MASS_KG)
FROM SPACEXTBL)
```

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

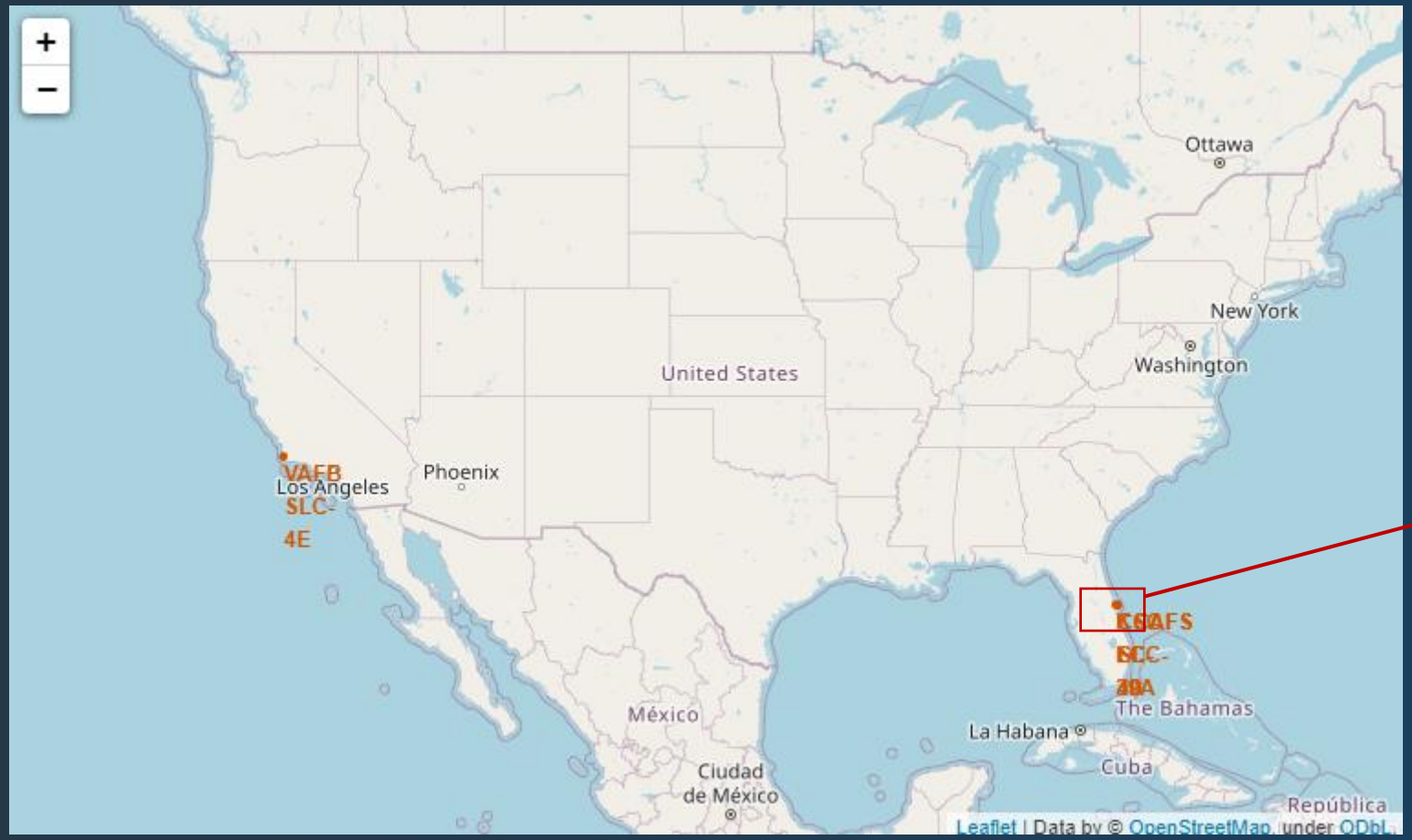
Landing_outcomes between the date 2010-06-04 and 2017-03-20 in desc order

%sql SELECT * FROM SPACEXTBL WHERE LAUNCH_OUTCOME LIKE „Success%“ AND (DATE BETWEEN „2010-06-04“ AND 2017-03-20“) ORDER BY DESC

DATE	time_utc_	booster_version	launch_site	payload	payload_mass_kg_	orbit	customer	mission_outcome	landing_outcome
2017-02-19	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
2017-01-14	17:54:00	F9 FT B1029.1	VAFB SLC-4E	Iridium NEXT 1	9600	Polar LEO	Iridium Communications	Success	Success (drone ship)
2016-08-14	05:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2016-07-18	04:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
2016-05-27	21:39:00	F9 FT B1023.1	CCAFS LC-40	Thaicom 8	3100	GTO	Thaicom	Success	Success (drone ship)
2016-05-06	05:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2016-04-08	20:43:00	F9 FT B1021.1	CCAFS LC-40	SpaceX CRS-8	3136	LEO (ISS)	NASA (CRS)	Success	Success (drone ship)
2015-12-22	01:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)

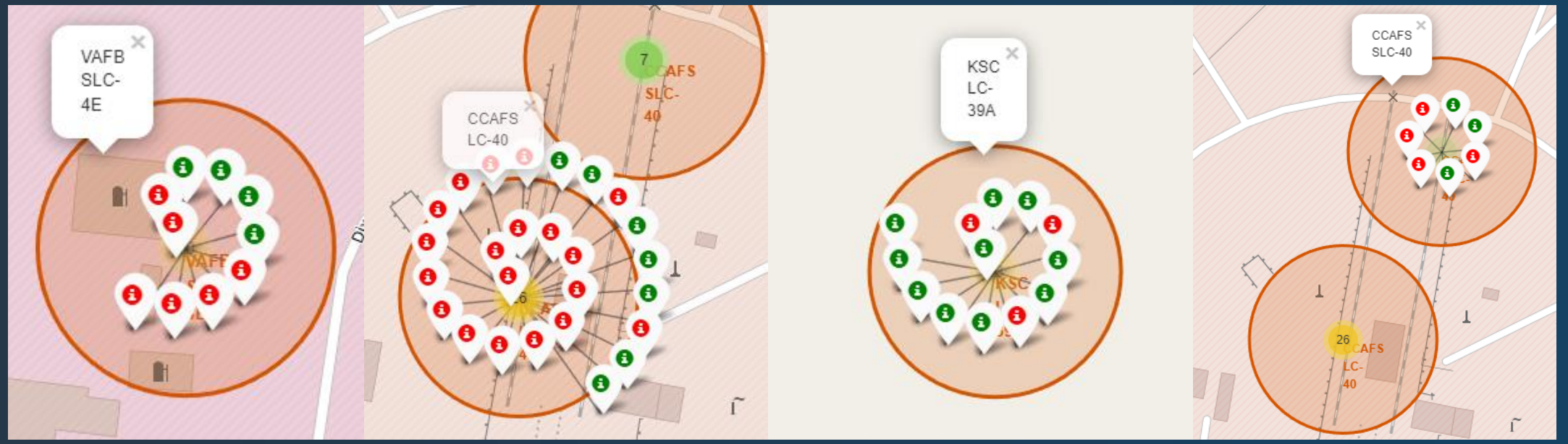
Interactive map with Folium

All launch sites



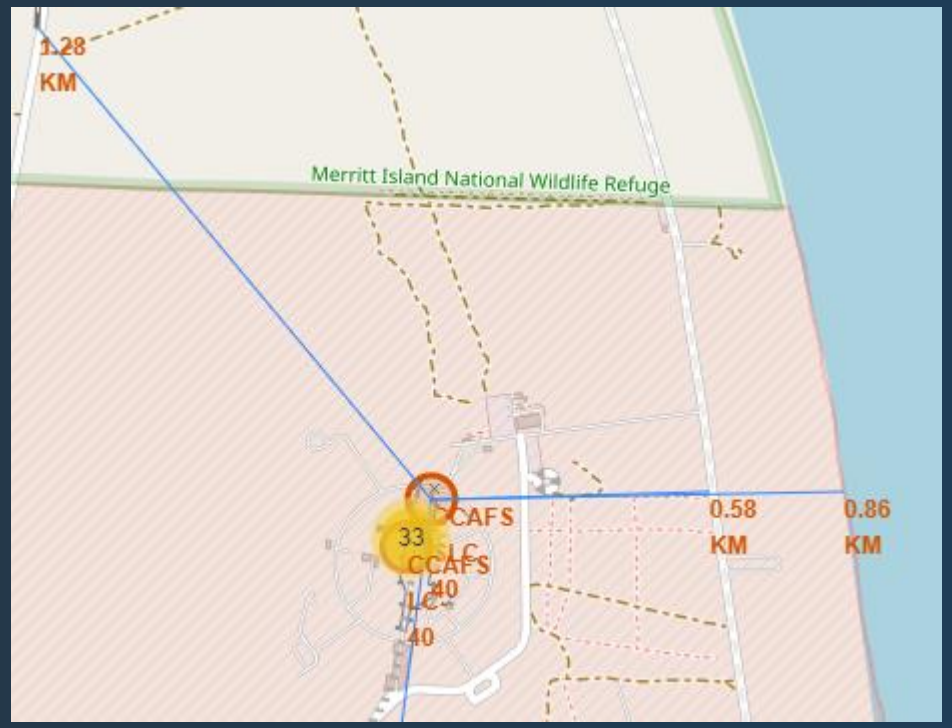
Success und fail landing marks

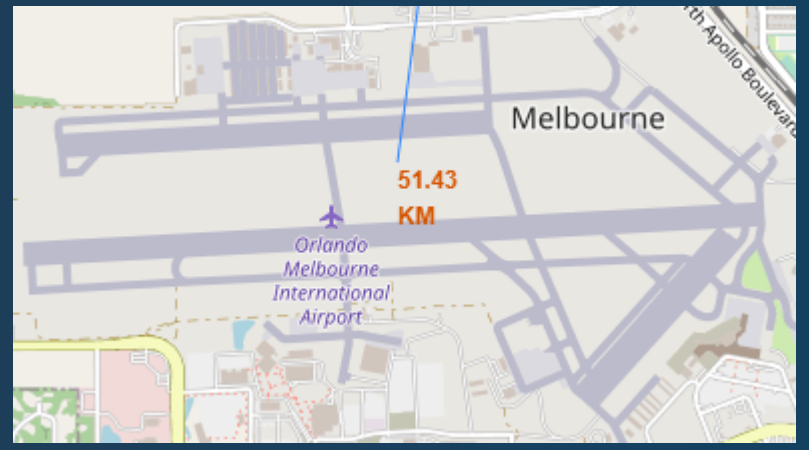
Green marks a succeed landing and Red marks a failed landing.



Success und fail landing marks

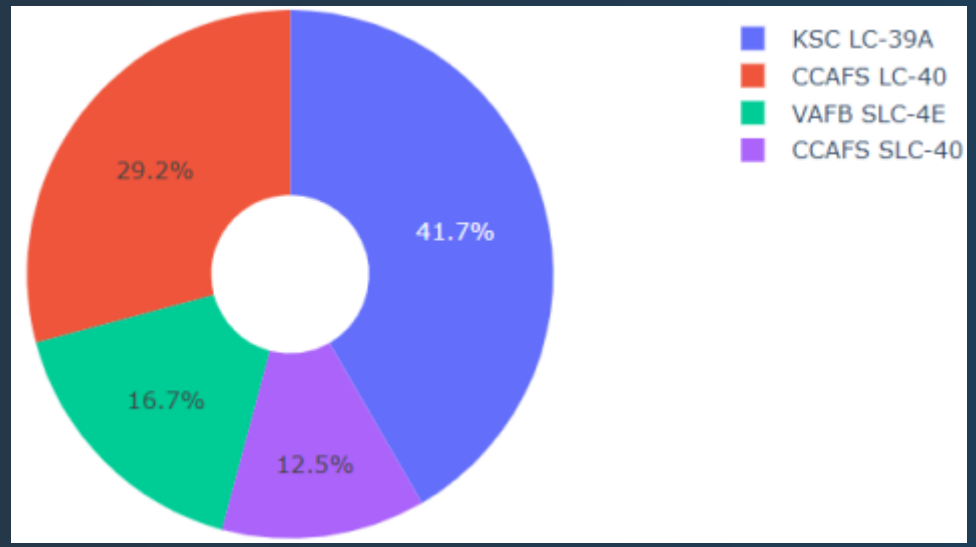
Distance to the closest Highway, coast, rails and airport



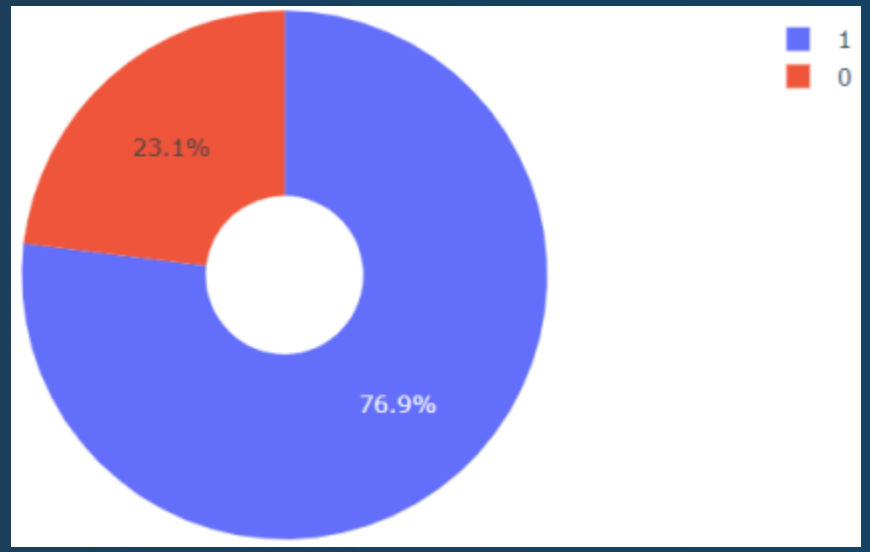


Dashboard – Pie Chart

Success rate all Launches

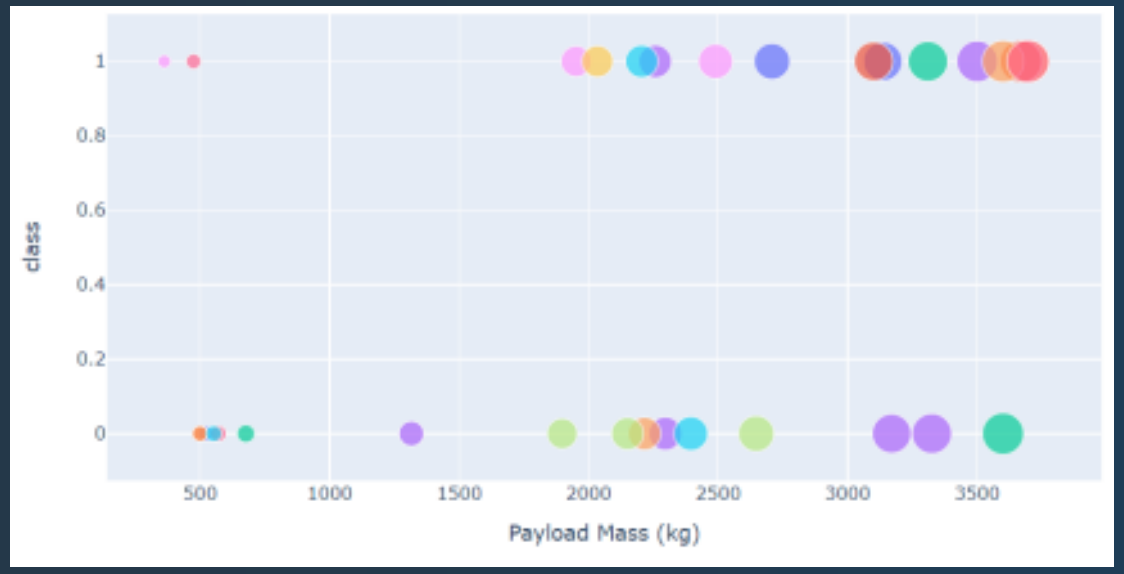


KSC LC-39A success rate

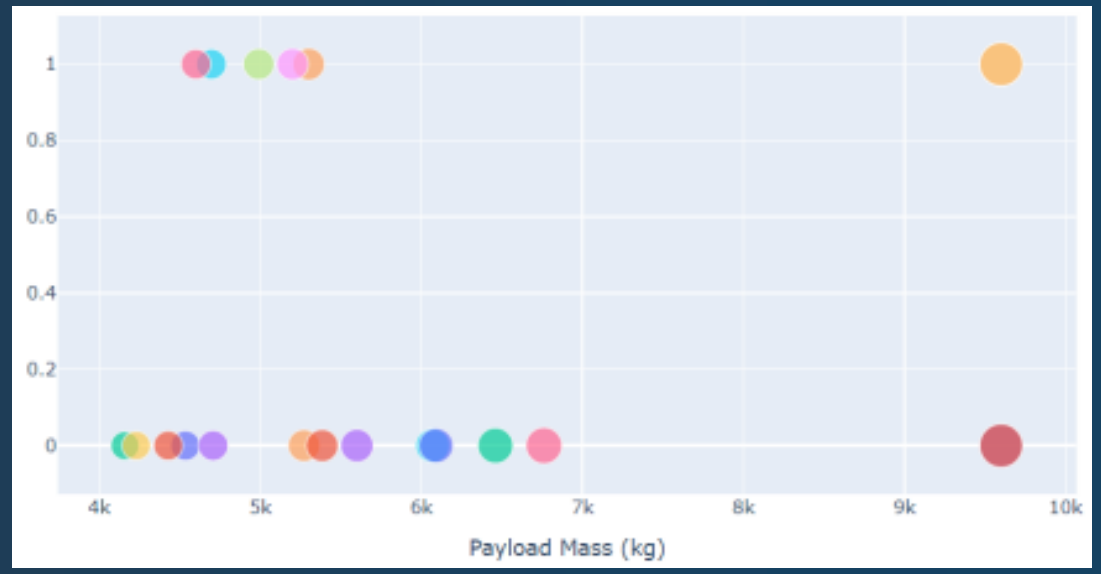


Dashboard – Scatter Plot

Light Payload max 4000 kg



Heavy Payload min 4000 kg



Predictive analysis

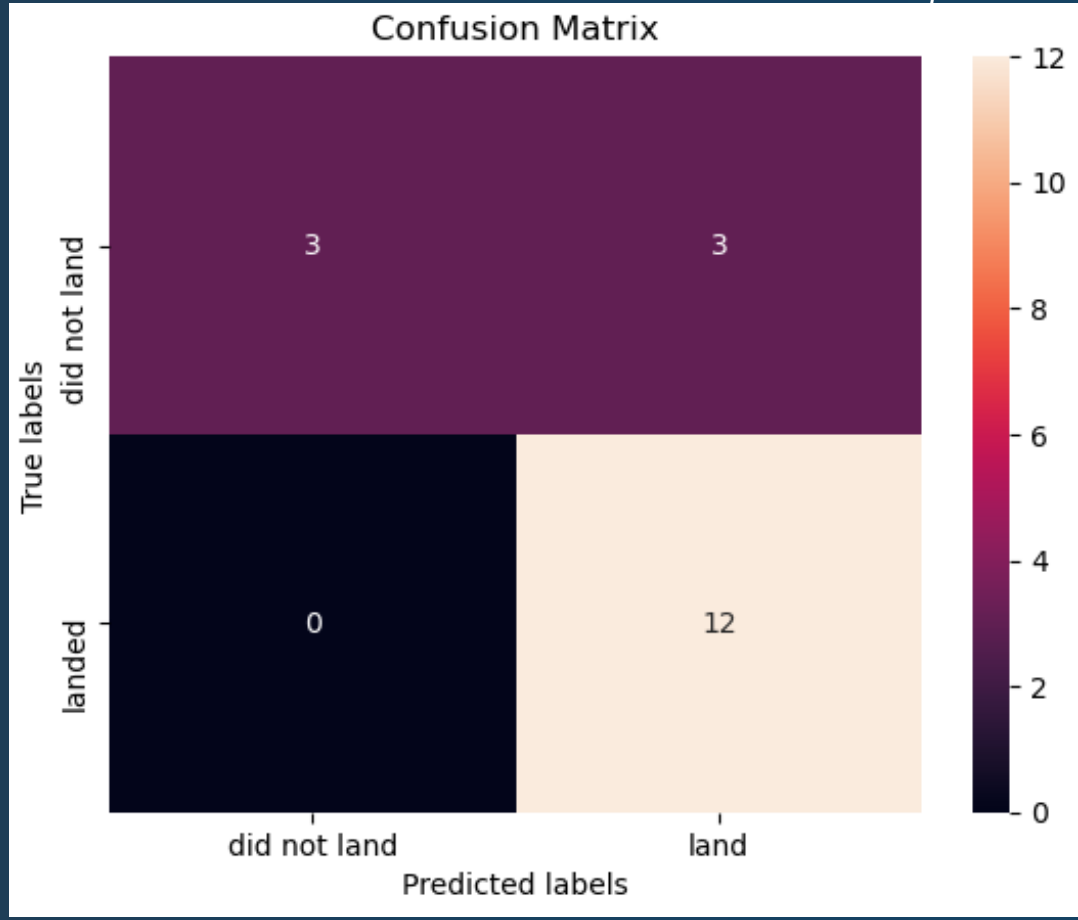
Classification models

Classification accuracy using models

- Performing accuracy tests on the following models
  - K-nearest Neighbors (KNN)
  - Support Vector Machine
  - Linear Model (Logistic Regression classifier)
  - Decision Tree Model
- The Decision Tree has the highest accuracy

Confusion Matrix for the tree

After Examination we see problems in the row „did not land“ for the actual values.



Conclusion

Conclusion

- Higher Pay Load Mass performs better than lighter Pay Load Mass, except for KSC LC 39A
- Orbits ES-L1, GEO, HEO and SSO has the best success rate
- Success rate of SpaceX increases with time in years
- The Decision Tree Model works the best for this dataset

Thank You!