IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Lena
February 23, 2024

Outline

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

Executive Summary

The project is intended to collect and analyze data from rocket launches by Space X. The aim is to predict the success of a launch in order to calculate the costs of further launches.

- Methodologies
  - Data Collection from Space X data
  - Data Analysation and Visualization
  - Machine Learning
- Results
  - Predictive Results using machine learning
  - Analysis of relationship between different parameters

Introduction

- Project background and context

The project is intended to use data from rocket launches by Space X to predict the success of further launches. These information should help the competating company Space Y to gain insights into the relevant parameters for a successful launch and therefore to reduce the costs.

- Problems you want to find answers
  - What influences the success of a rocket launch?
  - Which payload has the highest risk of failure?
  - Which locations are optimal for launch sites?
  - How can the success of rocket launch?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- 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 was collected using two different ways:
    1) Data was collected via the Space X API by making a request and collecting the dat
    2) Data was collected using wrapscaping

[hide] Flight No.	Date and time (UTC)	Version, Booster[b]	Launch site	Payload[c]	Payload mass	Orbit	Customer	Launch outcome	Booster landing
78	7 January 2020, 02:19:21[492]	F9 B5 ♺ B1049.4	CCAFS, SLC-40	Starlink 2 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)
	Third large batch and second operational flight of Starlink constellation. One of the 60 satellites included a test coating to make the satellite less reflective, and thus less likely to interfere with ground-based astronomical observations.[493]								
79	19 January 2020, 15:30[494]	F9 B5 ♺ B1046.4	KSC, LC-39A	Crew Dragon in-flight abort test[495] (Dragon C205.1)	12,050 kg (26,570 lb)	Sub-orbital[496]	NASA (CTS)[497]	Success	No attempt
	An atmospheric test of the Dragon 2 abort system after Max Q. The capsule fired its SuperDraco engines, reached an apogee of 40 km (25 mi), deployed parachutes after reentry, and splashed down in the ocean 31 km (19 mi) downrange from the launch site. The test was previously slated to be accomplished with the Crew Dragon Demo-1 capsule;[498] but that test article exploded during a ground test of SuperDraco engines on 20 April 2019.[419] The abort test used the capsule originally intended for the first crewed flight.[499] As expected, the booster was destroyed by aerodynamic forces after the capsule aborted.[500] First flight of a Falcon 9 with only one functional stage — the second stage had a mass simulator in place of its engine.								
80	29 January 2020, 14:07[501]	F9 B5 ♺ B1051.3	CCAFS, SLC-40	Starlink 3 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)
	Third operational and fourth large batch of Starlink satellites, deployed in a circular 290 km (180 mi) orbit. One of the fairing halves was caught, while the other was fished out of the ocean.[502]								
81	17 February 2020, 15:05[503]	F9 B5 ♺ B1056.4	CCAFS, SLC-40	Starlink 4 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Failure (drone ship)
	Fourth operational and fifth large batch of Starlink satellites. Used a new flight profile which deployed into a 212 km × 386 km (132 mi × 240 mi) elliptical orbit instead of launching into a circular orbit and firing the second stage engine twice. The first stage booster failed to land on the drone ship[504] due to incorrect wind data.[505] This was the first time a flight proven booster failed to land.								
82	7 March 2020, 04:50[506]	F9 B5 ♺ B1059.2	CCAFS, SLC-40	SpaceX CRS-20 (Dragon C112.3 ♺)	1,977 kg (4,359 lb)[507]	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
	Last launch of phase 1 of the CRS contract. Carries *Bartolomeo*, an ESA platform for hosting external payloads onto ISS.[508] Originally scheduled to launch on 2 March 2020, the launch date was pushed back due to a second stage engine failure. SpaceX decided to swap out the second stage instead of replacing the faulty part.[509] It was SpaceX's 50th successful landing of a first stage booster, the third flight of the Dragon C112 and the last launch of the cargo Dragon spacecraft.								
83	18 March 2020, 12:16[510]	F9 B5 ♺ B1048.5	KSC, LC-39A	Starlink 5 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Failure (drone ship)
	Fifth operational launch of Starlink satellites. It was the first time a first stage booster flew for a fifth time and the second time the fairings were reused (Starlink flight in May 2019).[511] Towards the end of the first stage burn, the booster suffered premature shut down of an engine, the first of a Merlin 1D variant and first since the CRS-1 mission in October 2012. However, the payload still reached the targeted orbit.[512] This was the second Starlink launch booster landing failure in a row, later revealed to be caused by residual cleaning fluid trapped inside a sensor.[513]								
84	22 April 2020, 19:30[514]	F9 B5 ♺ B1051.4	KSC, LC-39A	Starlink 6 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)

Data Collection – SpaceX API

- Data was collected from the Space X API via an request from a provided URL
- The data has been requested and parsed using json
- GitHub-URL: https://github.com/LenaLambers/RocketScience/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Task 1: Request and parse the SpaceX launch data using the GET request

To make the requested JSON results more consistent, we will use the following static response object for this project:

```
[10]: static_json_url='https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/API_call_spacex_api.json'
```

We should see that the request was successfull with the 200 status response code

```
[11]: response.status_code
[11]: 200
```

Now we decode the response content as a Json using `.json()` and turn it into a Pandas dataframe using `.json_normalize()`

```
[14]: # Use json_normalize meethod to convert the json result into a dataframe

respjson = response.json()
data= pd.json_normalize(respjson)
```

Using the dataframe `data` print the first 5 rows

```
[15]: # Get the head of the dataframe

data.head()
```

Data Collection - Scraping

- Data collection using webscraping from a provided URL
- Usage of beautiful soup
- Github URL: https://github.com/LenaLambers/RocketScience/blob/main/jupyter-labs-webscraping.ipynb

[10]:
```
# use requests.get() method with the provided static_url
# assign the response to a object

response = requests.get(static_url).text
```

Create a `BeautifulSoup` object from the HTML `response`

[11]:
```
# Use BeautifulSoup() to create a BeautifulSoup object from a response text content

soup = BeautifulSoup(response, 'html.parser')
```

Print the page title to verify if the `BeautifulSoup` object was created properly

[12]:
```
# Use soup.title attribute
print(soup.title)
```

```
<title>List of Falcon 9 and Falcon Heavy launches - Wikipedia</title>
```

Data Wrangling

- Data were processed using Exploratory Data Analysis (EDA)
- Analysis using Data Frames
- Github: https://github.com/LenaLambers/RocketScience/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

TASK 3: Calculate the number and occurence of mission outcome of the orbits

Use the method `.value_counts()` on the column `Outcome` to determine the number of `landing_outcomes`.Then as

```
1]: # landing_outcomes = values on Outcome column

landing_outcomes = df["Outcome"].value_counts()
landing_outcomes
```

```
1]: True ASDS      41
None None      19
True RTLS      14
False ASDS      6
True Ocean      5
False Ocean     2
None ASDS       2
False RTLS      1
Name: Outcome, dtype: int64
```

EDA with Data Visualization

- The following evaluations were done:
  - FlightNumber vs. Payload Mass
  - FlightNumber vs. LaunchSite
  - Launchsite vs. Payload
  - Orbit vs. Success Rate
  - Orbit vs. Flight Number
  - Orbit vs. Payload
  - Date vs. Success Rate
- Github: https://github.com/LenaLambers/RocketScience/blob/main/jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

EDA with SQL

- Exemplary SQL Queries:
  - Display launch sites
  - Average mass load for a specific booster version
  - date when the first succesful landing outcome in ground pad was achieved
  - names of the 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
- Github: https://github.com/LenaLambers/RocketScience/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

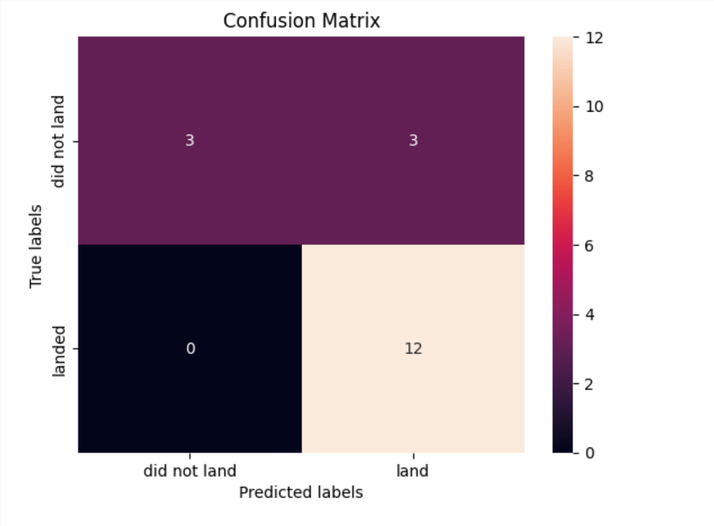
- Objects added in Folium:
  - Location of NASA Johnson Space Center
  - Location of the four launch sites
  - Marker Cluster with red and green markers for non-successful and succesfull launches
  - Mouse position
  - Distance line between launch site and coast, railway and city
- Objects were added to analysis, which parameters stand for an optimal launch site
- Github: https://github.com/LenaLambers/RocketScience/blob/main/lab_jupyter_launch_site_location.jupyterlite.ipynb

Predictive Analysis (Classification)

- Model building process:
  - Splitting the data into test and training
  - Fitting the model using different methods
  - Evaluate accuracy
  - Comparison of hyperparameters
  - Plotting of confusion matrix
- Comparison of Logistic Regression, SVM, Decision Tree and KNN
- Github: https://github.com/LenaLambers/RocketScience/blob/main/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

Results

	0
Method	Test Data Accuracy
Logistic_Reg	0.833333
SVM	0.833333
Decision Tree	0.833333
KNN	0.833333

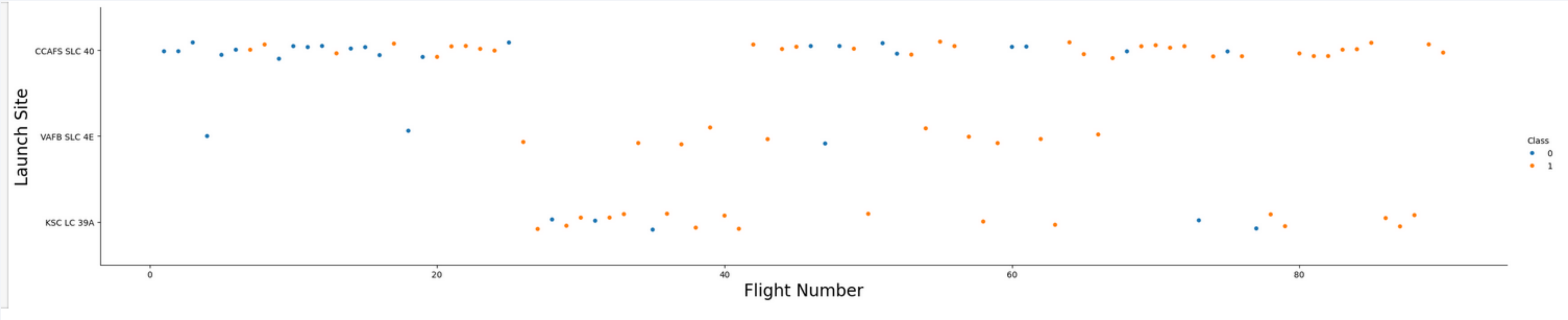


All methods have the same accuracy for predicting the data for a rocket launch.

Section 2

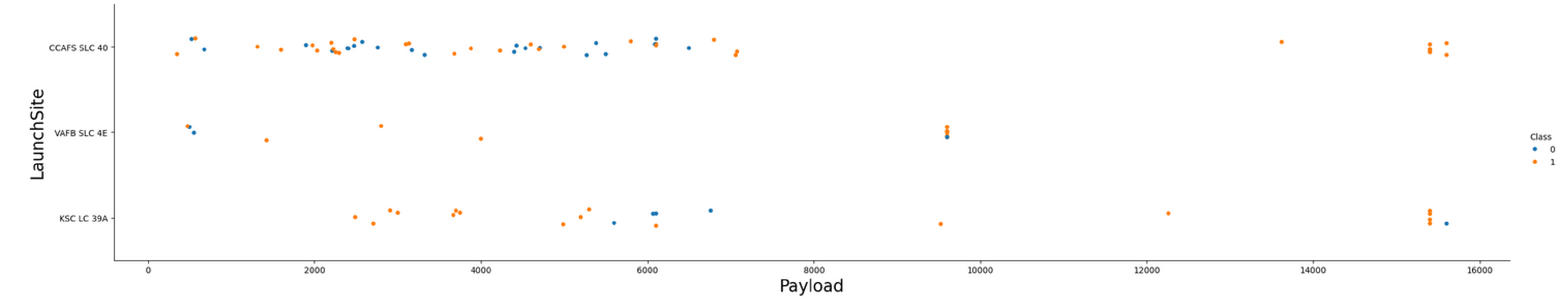
Insights drawn from EDA

Flight Number vs. Launch Site



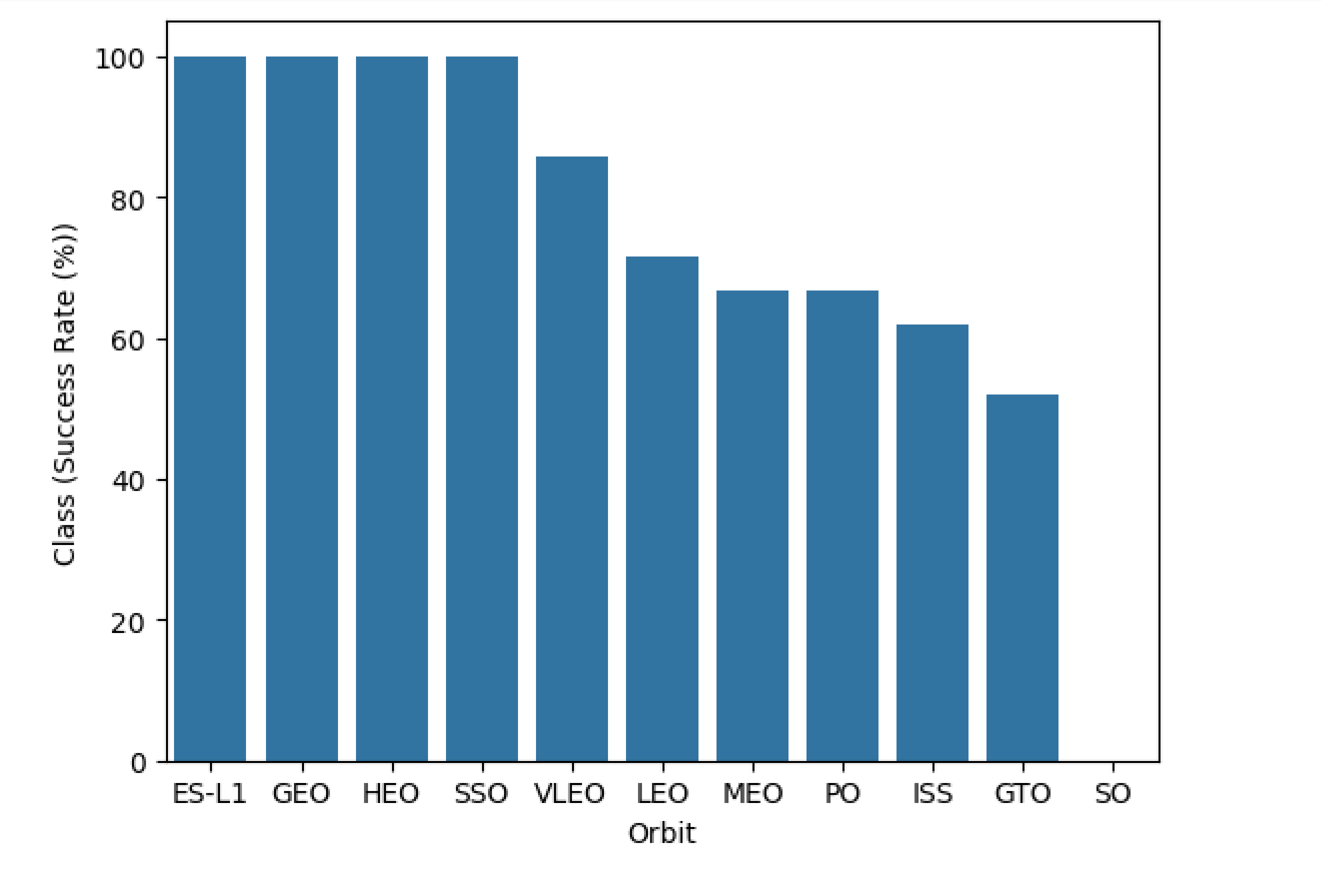
It becomes obvious, that the success rate increases with a higher number of flights. This can be explained, by the better training and planning due to longer experience of the team.
There is no significant difference between the success at different launch sites

Payload vs. Launch Site



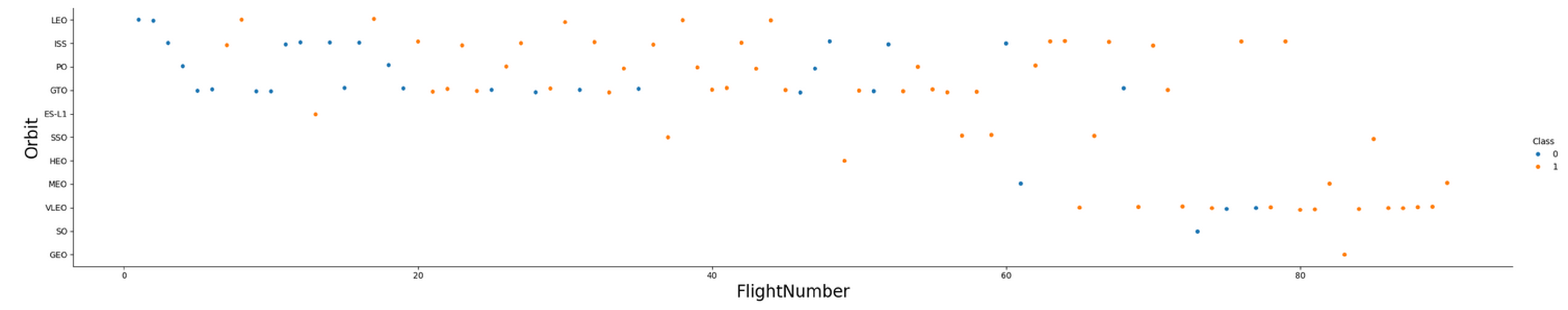
No Rockets with payload more than 10000 were launched at launch site VAFB-SLC
The success rate is higher for launches with higher payload

Success Rate vs. Orbit Type



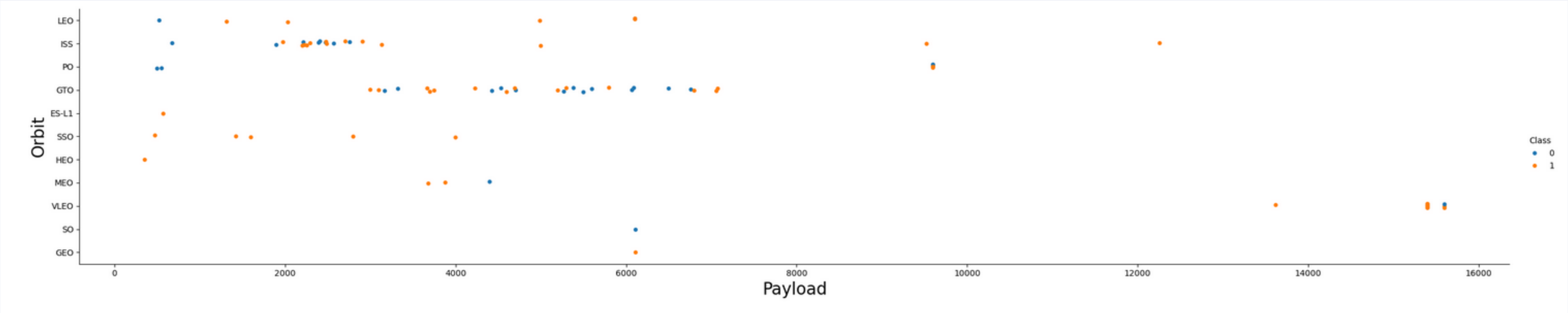
The success rate was higher at orbits near the earth. There was no successfull launch for the orbit SO.
Except for this orbit, the success rate is over 50% for all orbits.

Flight Number vs. Orbit Type



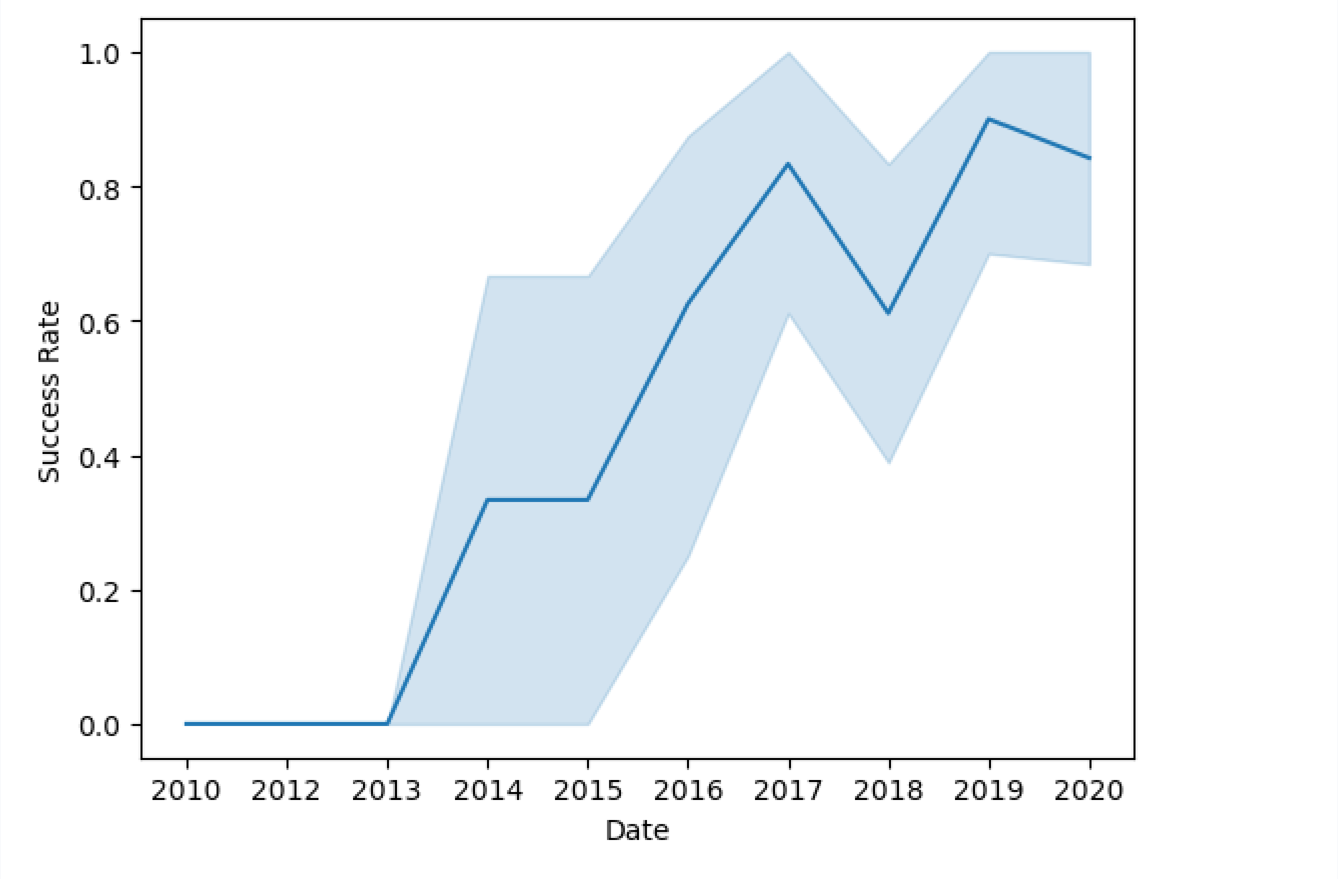
In the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

Payload vs. Orbit Type



With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS. However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.

Launch Success Yearly Trend



The success rtae significantly increases over the years, with only one exception in 2018.

All Launch Site Names

Space X uses different launch sites to launch their rockets.

The different launches site can be determined using a SQL query:

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

Launch Site Names Begin with 'CCA'

[10]:
```
%sql SELECT * FROM 'SPACEXTBL' WHERE Launch_Site LIKE 'CCA%' LIMIT 5
```

```
 * sqlite:///my_data1.db
Done.
```

[10]:

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

Task 3

The results show a list of launch site, that start with the letters CCA.

Total Payload Mass

Task 3

Display the total payload mass carried by boosters launched by NASA (CRS)

[16]:
```
%sql SELECT SUM(PAYLOAD_MASS__KG_) as "Total Payload Mass (kg)", Customer FROM 'SPACEXTBL' WHERE Customer='NASA (CRS)'
```

```
 * sqlite:///my_data1.db
Done.
```

[16]:

Total Payload Mass (kg)	Customer
45596	NASA (CRS)

In total, the customer NASA carried a payload of 45596 kg into space using Space X rockets.

Average Payload Mass by F9 v1.1

Task 4

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

[20]:
```
%sql SELECT AVG(PAYLOAD_MASS__KG_) as "AVERAGE Payload Mass (kg)", Booster_Version FROM 'SPACEXTBL' WHERE Booster_Version LIKE 'F9 v1.1%'
```

```
 * sqlite:///my_data1.db
Done.
```

[20]:

AVERAGE Payload Mass (kg)	Booster_Version
2534.6666666666665	F9 v1.1 B1003

The booster version F9 v1.1 carried an average payload of 2534.67 kg per launch.

First Successful Ground Landing Date

The first sucessful landing outcome in ground pad was performed on December 22, 2015

1
2015-12-22

```
%sql SELECT min(date) from SPACEXTBL where landing__outcome = 'Success (ground pad)'
```

Successful Drone Ship Landing with Payload between 4000 and 6000

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

```
%sql SELECT DISTINCT Booster_Version, Payload FROM 'SPACEXTBL' WHERE 'Landing _Outcome' = "Success (drone ship)" AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ <6000
```

```
 * sqlite:///my_data1.db
Done.
```

Booster_Version	Payload

- The following results show a list of 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

List the total number of successful and failure mission outcomes

```
%sql SELECT "Mission_Outcome", COUNT("Mission_Outcome") as Total FROM SPACEXTBL GROUP BY "Mission_Outcome";
```

```
 * sqlite:///my_data1.db
Done.
```

Mission_Outcome	Total
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

The result list the successful and failed outcomes of all missions.

Boosters Carried Maximum Payload

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

```
 * sqlite:///my_data1.db
Done.
```

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

The listed booster versions have carried the maximum payload mass

2015 Launch Records

DATE	booster_version	launch_site	landing__outcome
2015-01-10	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

There were only two failed landings in 2015, namely on October 01. and on April 14.

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

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

- The results show a ranking of 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

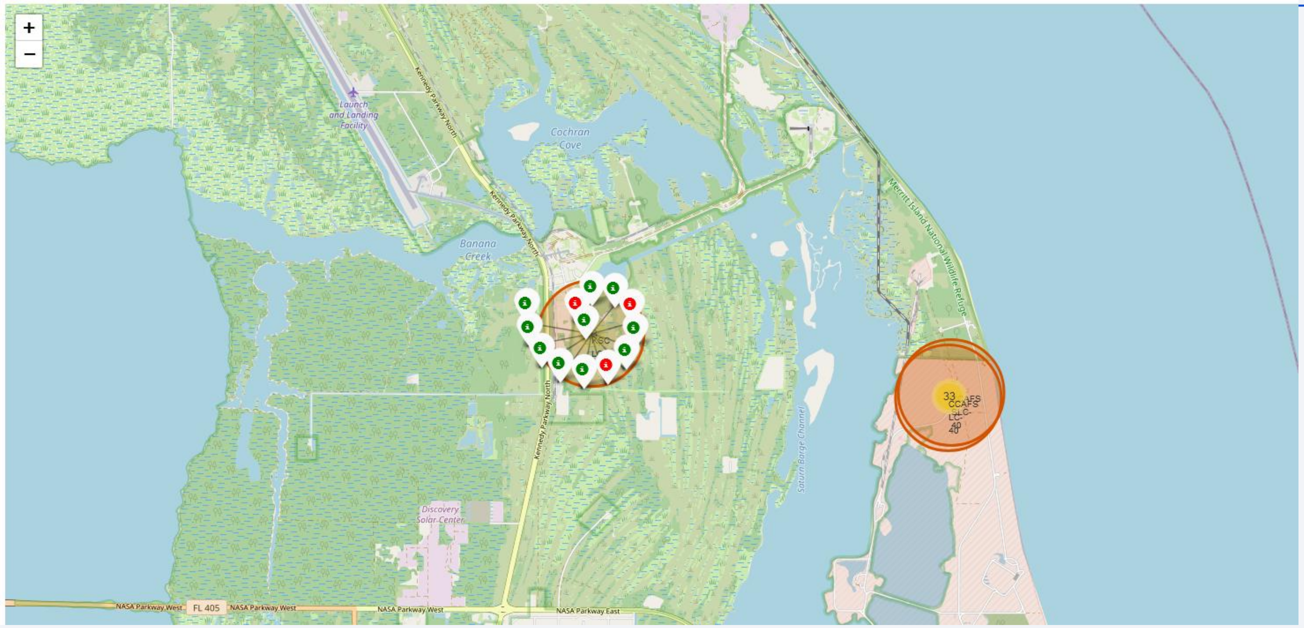
Section 3

Launch Sites Proximities Analysis

Spatial distribution of launch sites

The map shows the locations of Space X launch sites in California and Florida. We added the location as well as information on the successfusll launches from these locations. Both launch sites are located next to the coast, as well as in close distance to a railroad and are therefore optimal for rocket launches.

Launch outcomes at locations



- The map shows the two launch sites located near the Kennedy Space Center in Florida. Green markers represent successful launches from this launch sites, while red markers stand for failed launches.

Evaluation of optimal location

- To evaluate the optimal parameters for a launch site, the distances to important landmarks like coast lines and railroad were calculated. All launch sites are located at the coast, as shown in the result plot.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

	0
Method	Test Data Accuracy
Logistic_Reg	0.833333
SVM	0.833333
Decision Tree	0.833333
KNN	0.833333

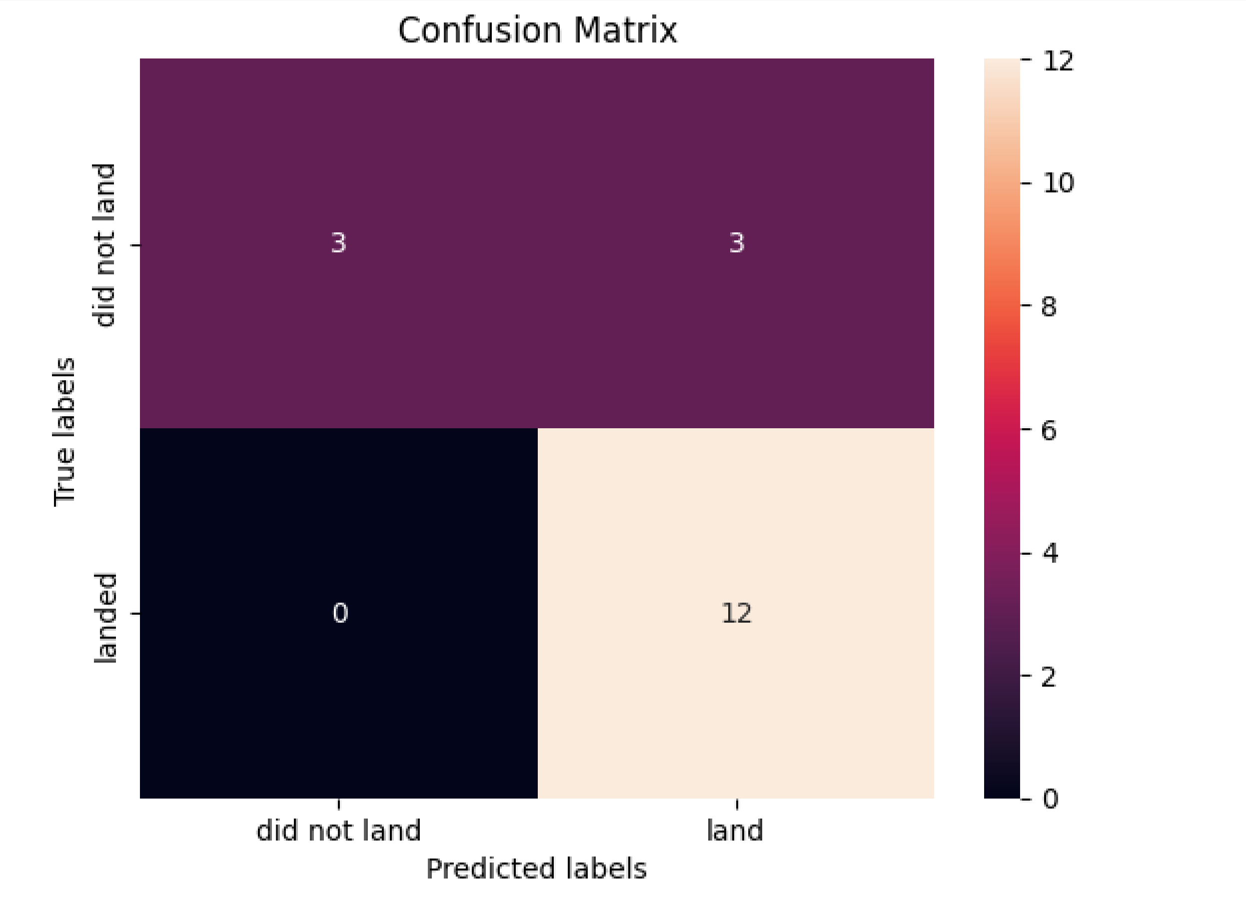
The accuracy for all four models is the same.

TASK 12

Find the method performs best:

```
Report = pd.DataFrame({'Method' : ['Test Data Accuracy']})

knn_accuracy=knn_cv.score(X_test, Y_test)
Decision_tree_accuracy=tree_cv.score(X_test, Y_test)
SVM_accuracy=svm_cv.score(X_test, Y_test)
Logistic_Regression=logreg_cv.score(X_test, Y_test)

Report['Logistic_Reg'] = [Logistic_Regression]
Report['SVM'] = [SVM_accuracy]
Report['Decision Tree'] = [Decision_tree_accuracy]
Report['KNN'] = [knn_accuracy]




Report.transpose()
```

Confusion Matrix



The confusion matrix shows a good prediciton for the failed launches, but has slight problems with too much false positive predictions.

Conclusions

- Data Analysis can help to predict rocket launches from Space X
- Data visualization and analysis showed an increasing successrate over time and with increasing flights
- Machine learning approaches show a good accuracy over 80% for the prediction
- Spatial analysis of location sites help to find the optimal location parameters for possible future launch sites.

Thank you!