SpaceX Falcon-9 First stage landing prediction

IBM Data Science Capstone Project

Fatih Ilhan

Outline

Executive Summary

Stages

Collection Data

Data Wrangling

Exploratory Data

EDA by Visualization

Predective Analysis by Machine Learning

Introduction

SpaceX's Goal

- Sending spacecraft to the international space station
- Providing satellite internet to the whole world with Starlink technology
- Taking people and cargo into space and contributing to space exploration.

02

FALCON 9 ROCKETS

Falcon 9 is a two-stage reusable rocket developed and manufactured by SpaceX, a private aerospace company founded by Elon Musk. Here are some key features and missions of Falcon 9 rockets

- Reusability: One of the notable features of Falcon 9 is its reusability. The first stage of the rocket is designed to return to Earth after launch, landing vertically either on land (at SpaceX's landing zones) or on an autonomous drone ship in the ocean. This reusability significantly reduces the cost of space launches.

- Payload Capacity: Falcon 9 is capable of delivering a variety of payloads to orbit, including satellites, cargo resupply missions to the International Space Station (ISS), and even crewed missions. It has a payload capacity of up to 22,800 kilograms (50,300 pounds) to low Earth orbit (LEO) and up to 8,300 kilograms (18,300 pounds) to geostationary transfer orbit (GTO).

SPACEX

- Starlink: Falcon 9 plays a crucial role in SpaceX's ambitious Starlink project, which aims to provide global broadband internet coverage from a network of thousands of small satellites in low Earth orbit. Falcon 9 launches numerous batches of Starlink satellites to gradually build up the constellation.

- NASA Missions: Falcon 9 has been selected by NASA for various missions, including resupplying the ISS as part of the Commercial Resupply Services (CRS) contract. Falcon 9 was also used to launch the Crew Dragon spacecraft, enabling NASA to resume crewed launches from U.S. soil.

- Satellite Deployment: Falcon 9 is frequently used to deploy satellites for commercial customers, government agencies, and scientific research. It offers the flexibility to deliver satellites into different orbits, such as LEO, GTO, and sun-synchronous orbit (SSO).

03

Methodology

Data Collection methodology

- Require the data from SpaceX API
- Collect data from a Wikipedia page

Data Collection methodology

- Perform EDA to find some problems
- Determine what would be the label for training supervised learning

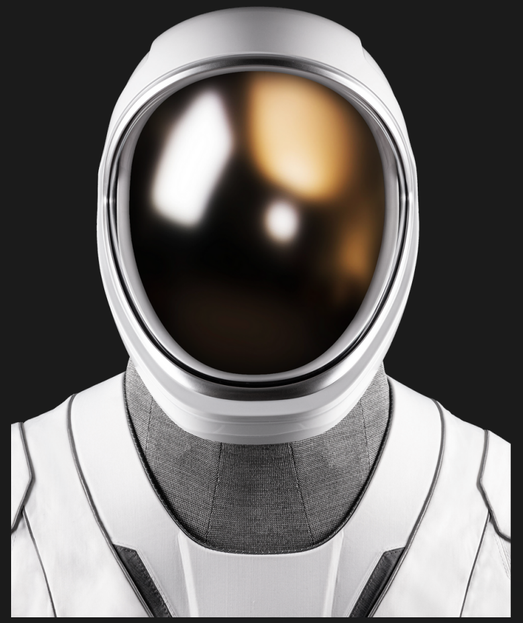
Perform exploratory data analysis (EDA) using visualization and SQL

Perform interactive visual analytics using Folium and Plotly Dash

Perform predictive analysis using classification models

03

Data Collection



API

spacex_url="https://api.spacexdata.com/v4/launches/past"

WEB PAGE

"https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922"

Data Collection - SpaceX API

- Request and parse the SpaceX launch data using the GET request
- Filter the dataframe to only include Falcon 9 launches

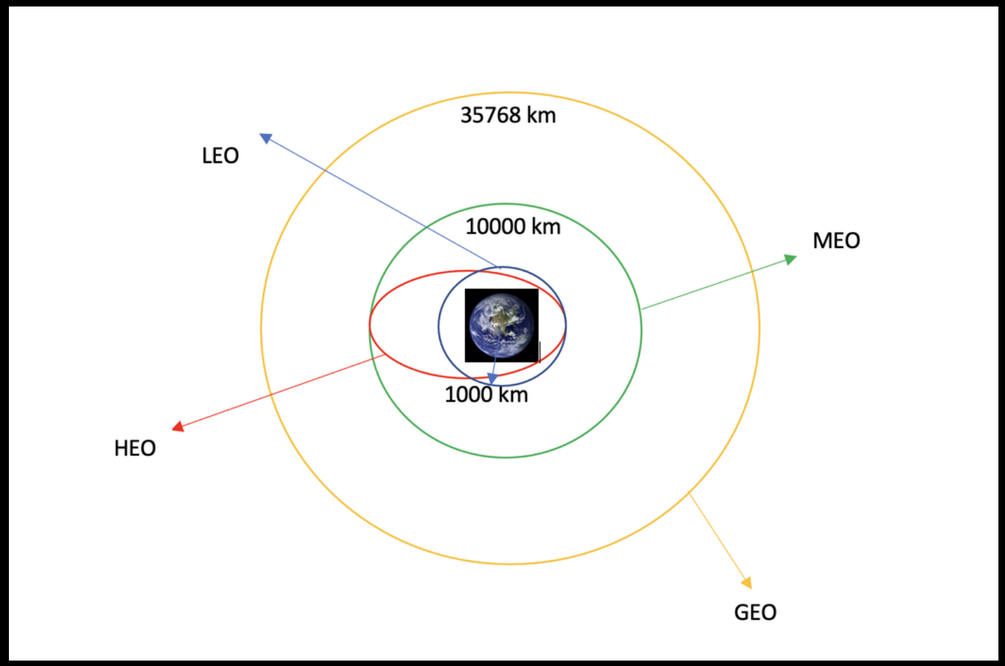
	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	
4	6	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False	
5	8	2012-05-22	Falcon 9	525.0	LEO	CCSFS SLC 40	None None	1	False	False	False	
6	10	2013-03-01	Falcon 9	677.0	ISS	CCSFS SLC 40	None None	1	False	False	False	
7	11	2013-09-29	Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean	1	False	False	False	
8	12	2013-12-03	Falcon 9	3170.0	GTO	CCSFS SLC 40	None None	1	False	False	False	
...	...	...	...	...	...	...	...	...	...	...	...	
89	102	2020-09-03	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	2	True	True	True	5e9e3032383
90	103	2020-10-06	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	3	True	True	True	5e9e3032383
91	104	2020-10-18	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	6	True	True	True	5e9e3032383
92	105	2020-10-24	Falcon 9	15600.0	VLEO	CCSFS SLC 40	True ASDS	3	True	True	True	5e9e3033383
93	106	2020-11-05	Falcon 9	3681.0	MEO	CCSFS SLC 40	True ASDS	1	True	False	True	5e9e3032383

Data Scraping - Wikipedia

- Web scraping Falcon 9 and Falcon Heavy Launches Records from Wikipedia
- Request the Falcon9 Launch Wiki page from its URL
- Extract all column/variable names from the HTML table header
- Create a data frame by parsing the launch HTML tables

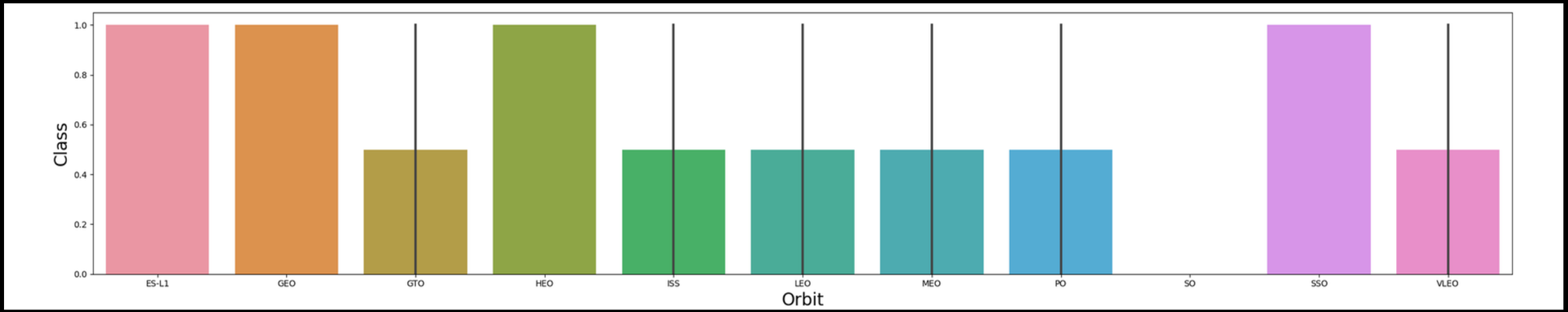
Data Wrangling

- Calculate the number of launches on each site
- Calculate the number and occurrence of each orbit
- Calculate the number and occurence of mission outcome per orbit type
- Create a landing outcome label from Outcome column

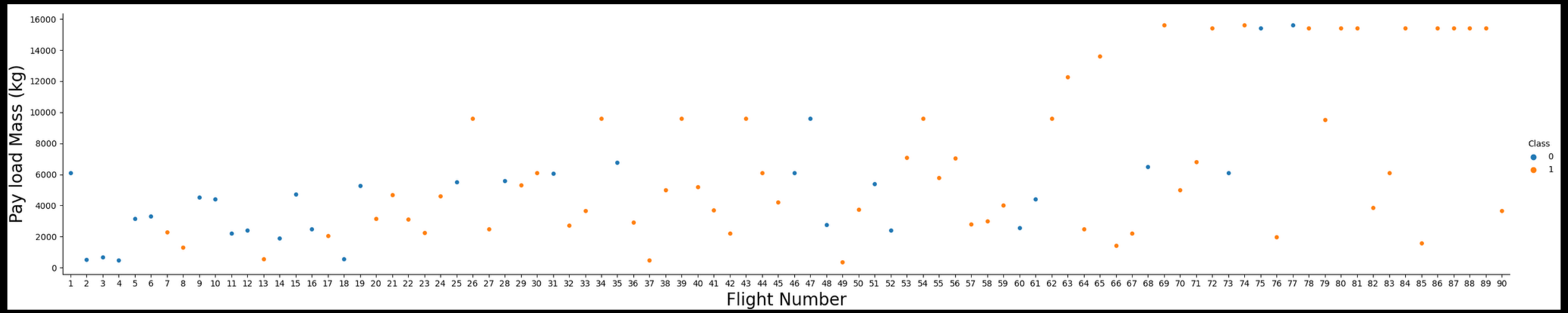


EDA with Data Visualization

- We can plot out the FlightNumber vs. PayloadMassand overlay the outcome of the launch
- Visualize the relationship between Flight Number and Launch Site
- Visualize the relationship between Payload and Launch Site
- Visualize the relationship between success rate of each orbit type
- Visualize the relationship between FlightNumber and Orbit type
- Visualize the relationship between Payload and Orbit type
- Visualize the launch success yearly trend

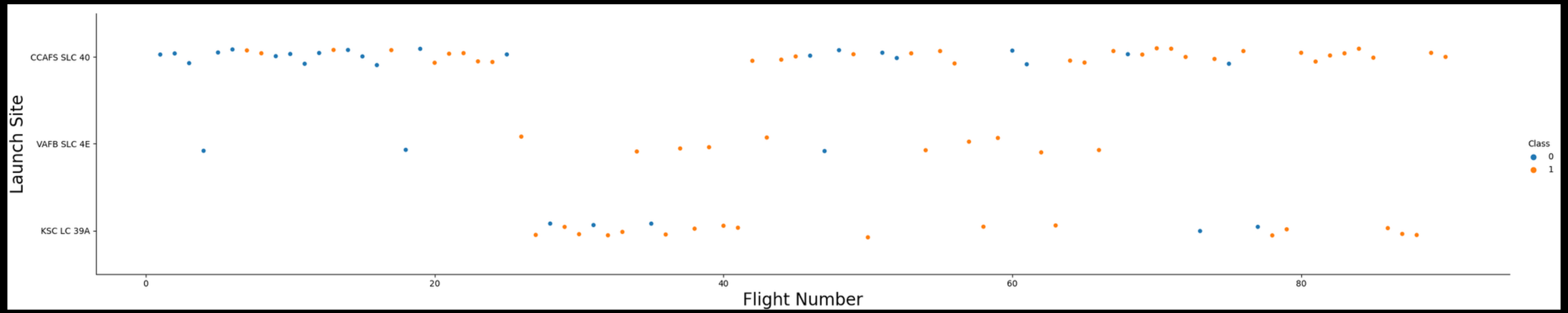


EDA with Data Visualization



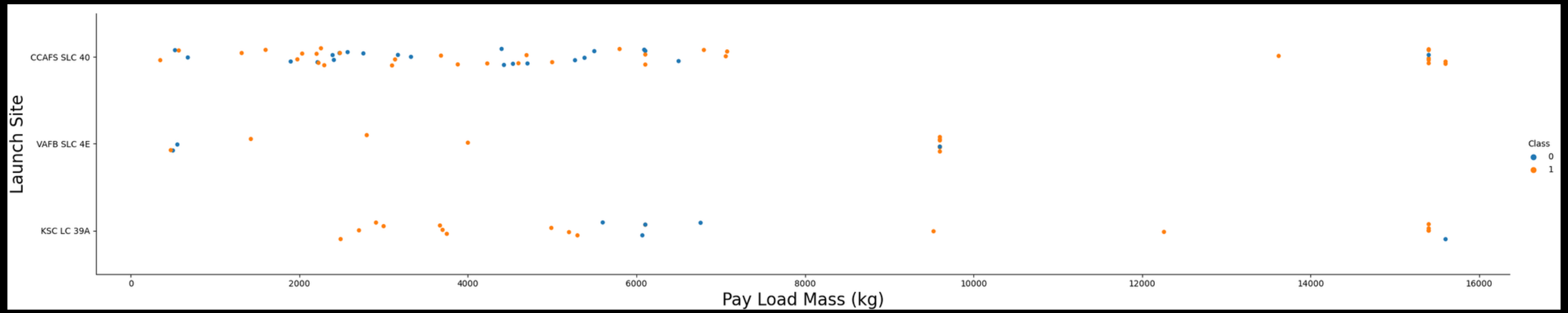
We see that different launch sites have different success rates. CCAFS LC-40, has a success rate of 60 %, while KSC LC-39A and VAFB SLC 4E has a success rate of 77%.

EDA with Data Visualization



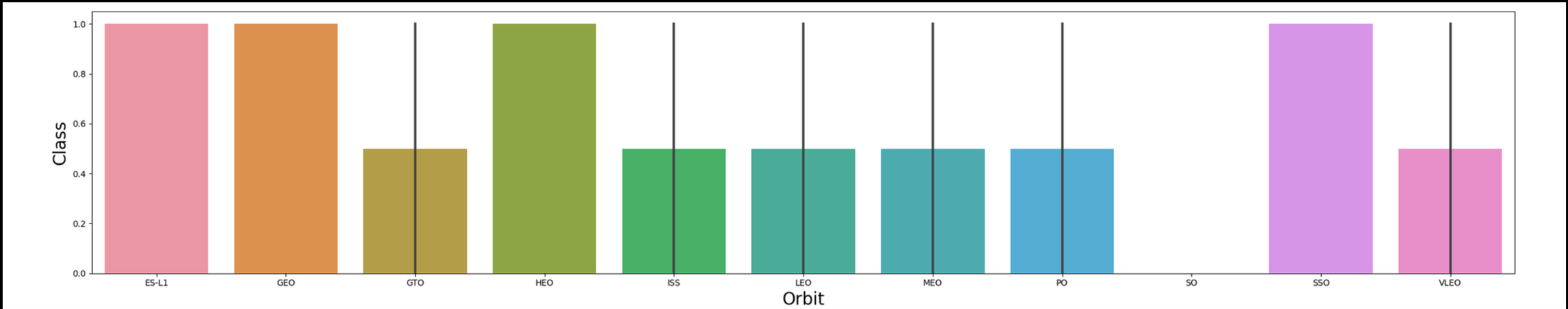
Use the function catplot to plot FlightNumber vs LaunchSite, set the parameter x parameter to FlightNumber, set the y to Launch Site and set the parameter hue to 'class'

EDA with Data Visualization



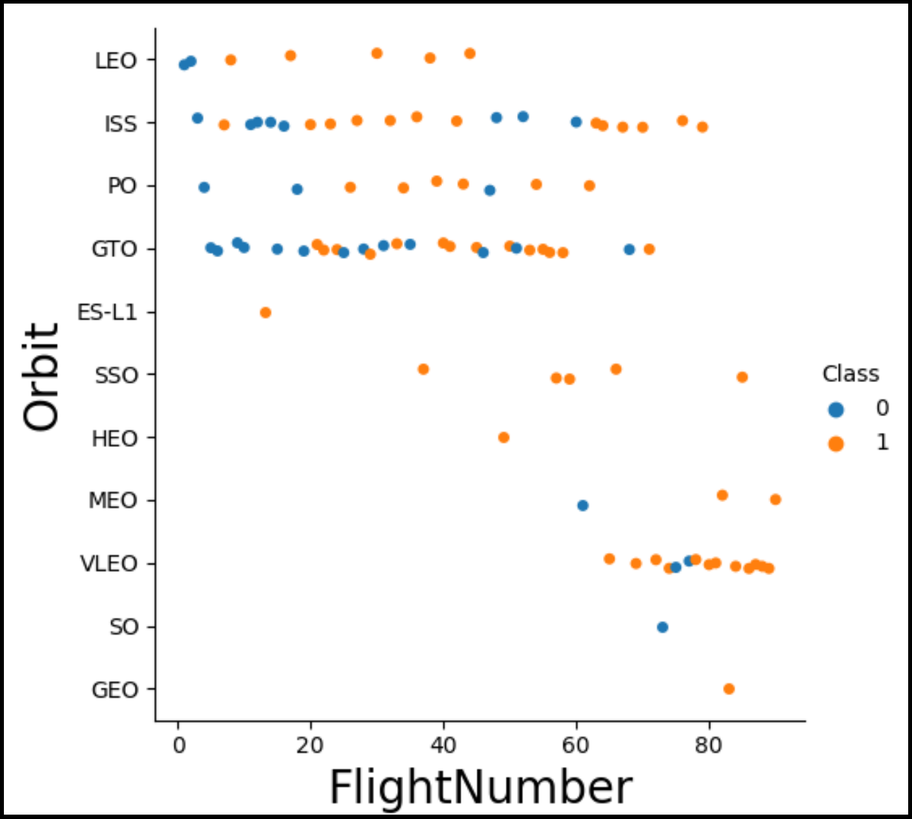
We also want to observe if there is any relationship between launch sites and their payload mass.

EDA with Data Visualization



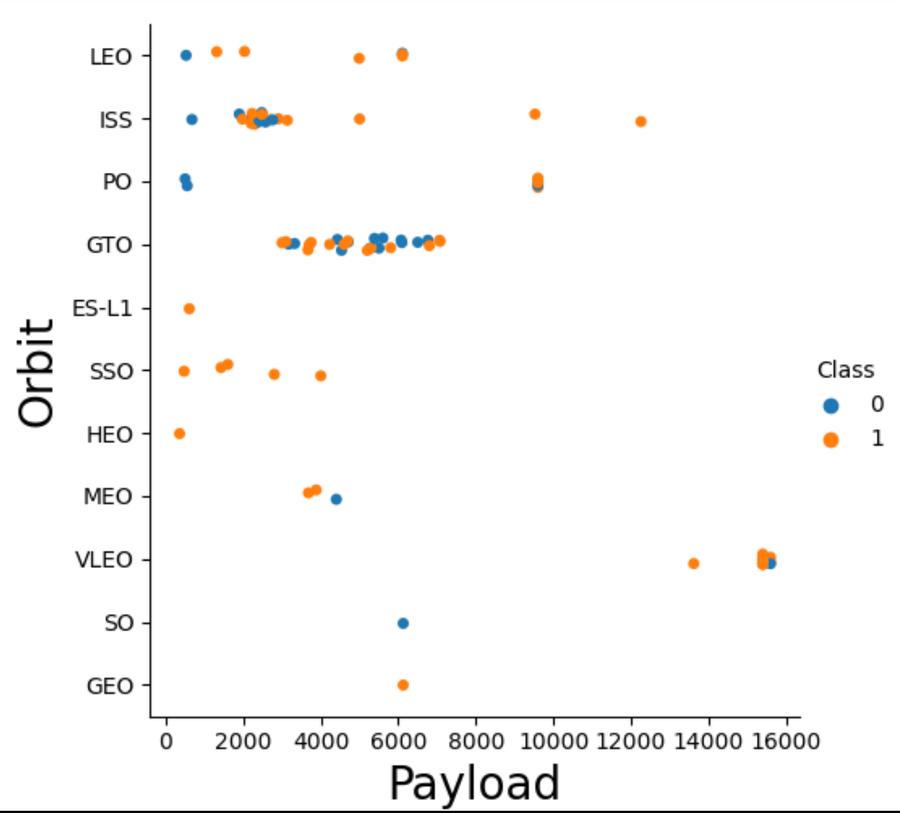
Visualize the relationship between success rate of each orbit type

EDA with Data Visualization



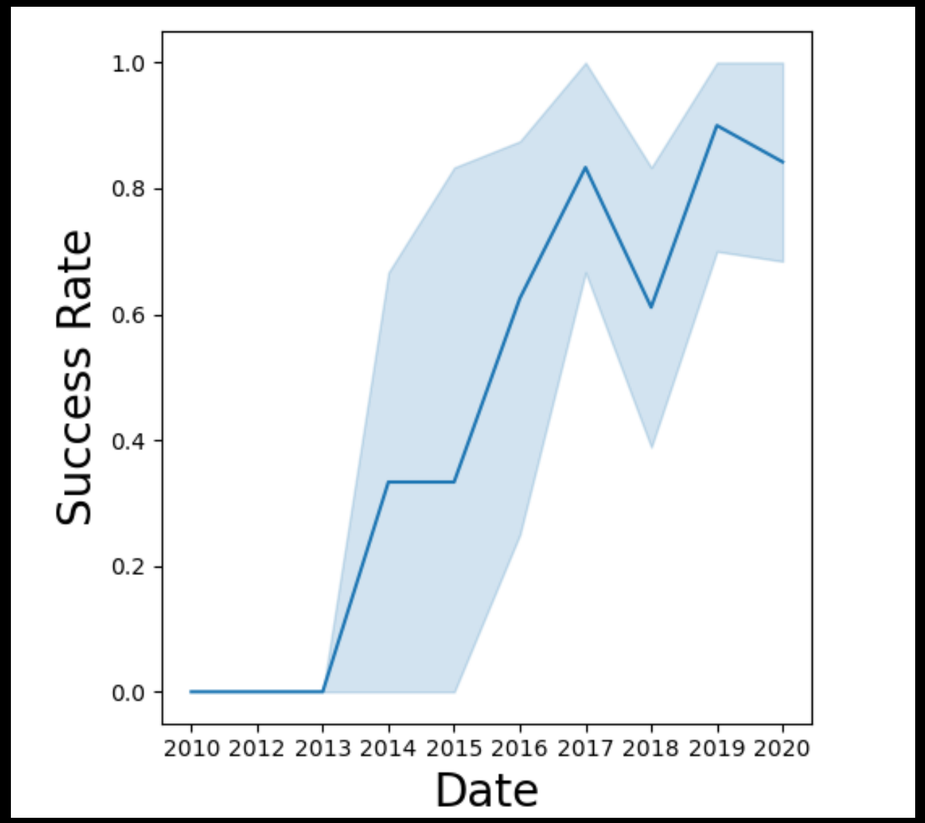
Visualize the relationship between FlightNumber and Orbit type

EDA with Data Visualization



Visualize the relationship between Payload and Orbit type

EDA with Data Visualization



Visualize the launch success yearly trend

EDA with SQL

- Display the names of the unique launch sites in the space mission
- Display 5 records where launch sites begin with the string 'CCA'
- Display the total payload mass carried by boosters launched by NASA (CRS)
- Display average payload mass carried by booster version F9 v1.1
- List the date when the first succesful landing outcome in ground pad was acheived.
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- List the total number of successful and failure mission outcomes
- List the names of the booster_versions which have carried the maximum payload mass. Use a subquery
- List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015
- Rank the count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

EDA with SQL

Display the names of the unique launch sites in the space mission

```
%sql select distinct(LAUNCH_SITE) from SPACEXTBL
```

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

Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40
None

Display the names of the unique launch sites in the space mission

EDA with SQL


```
%sql select * from SPACEXTBL where LAUNCH_SITE like 'CCA%' limit 5
```

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_
06/04/2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0.0	LEO	SpaceX	Success	Failure (
12/08/2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0.0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (
22/05/2012	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525.0	LEO (ISS)	NASA (COTS)	Success	N
10/08/2012	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500.0	LEO (ISS)	NASA (CRS)	Success	N
03/01/2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677.0	LEO (ISS)	NASA (CRS)	Success	N

Display 5 records where launch sites begin with the string 'CCA'

03

EDA with SQL

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

```
%sql select sum(PAYLOAD_MASS__KG_) from SPACEXTBL where CUSTOMER = 'NASA (CRS)'
```

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

sum(PAYLOAD_MASS__KG_)
45596.0

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

03

EDA with SQL

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

```
%sql select avg(PAYLOAD_MASS__KG_) from SPACEXTBL where BOOSTER_VERSION = 'F9 v1.1'
```

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

avg(PAYLOAD_MASS__KG_)
2928.4

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

03

EDA with SQL

List the date when the first succesful landing outcome in ground pad was acheived.

Hint:Use min function

```
%sql select min(DATE) from SPACEXTBL where Landing_Outcome = 'Success (ground pad)'
```

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

min(DATE)
01/08/2018

List the date when the first succesful landing outcome in ground pad was acheived.

EDA with SQL

Task 6

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 Booster_Version from SPACEXTBL WHERE Landing_Outcome = 'Success (drone ship)' and PAYLOAD_MASS__KG_ > 4000
```

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

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

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

03

EDA with SQL

Task 7

List the total number of successful and failure mission outcomes

```
%sql select count(Mission_Outcome) from SPACEXTBL WHERE Mission_Outcome = 'Success' or Mission_Outcome = 'Failure (in f
```

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

count(Mission_Outcome)
99

List the total number of successful and failure mission outcomes

03

EDA with SQL

Task 8

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

```
%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

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

EDA with SQL

Task 9

List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.

Note: SQLLite does not support monthnames. So you need to use substr(Date, 4, 2) as month to get the months and substr(Date,7,4)='2015' for year.

```
%sql SELECT SUBSTR(Date,4,2) AS Month, Booster_Version, Launch_site FROM SPACEXTBL WHERE Landing_Outcome LIKE 'Failure%
```

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

Month	Booster_Version	Launch_Site
10	F9 v1.1 B1012	CCAFS LC-40
04	F9 v1.1 B1015	CCAFS LC-40

List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.

03

EDA with SQL

Task 10

Rank the count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

```
%sql SELECT Landing_Outcome, COUNT(*) AS Numbers FROM SPACEXTBL WHERE Landing_Outcome LIKE 'Success%' AND Date BETWEEN
```

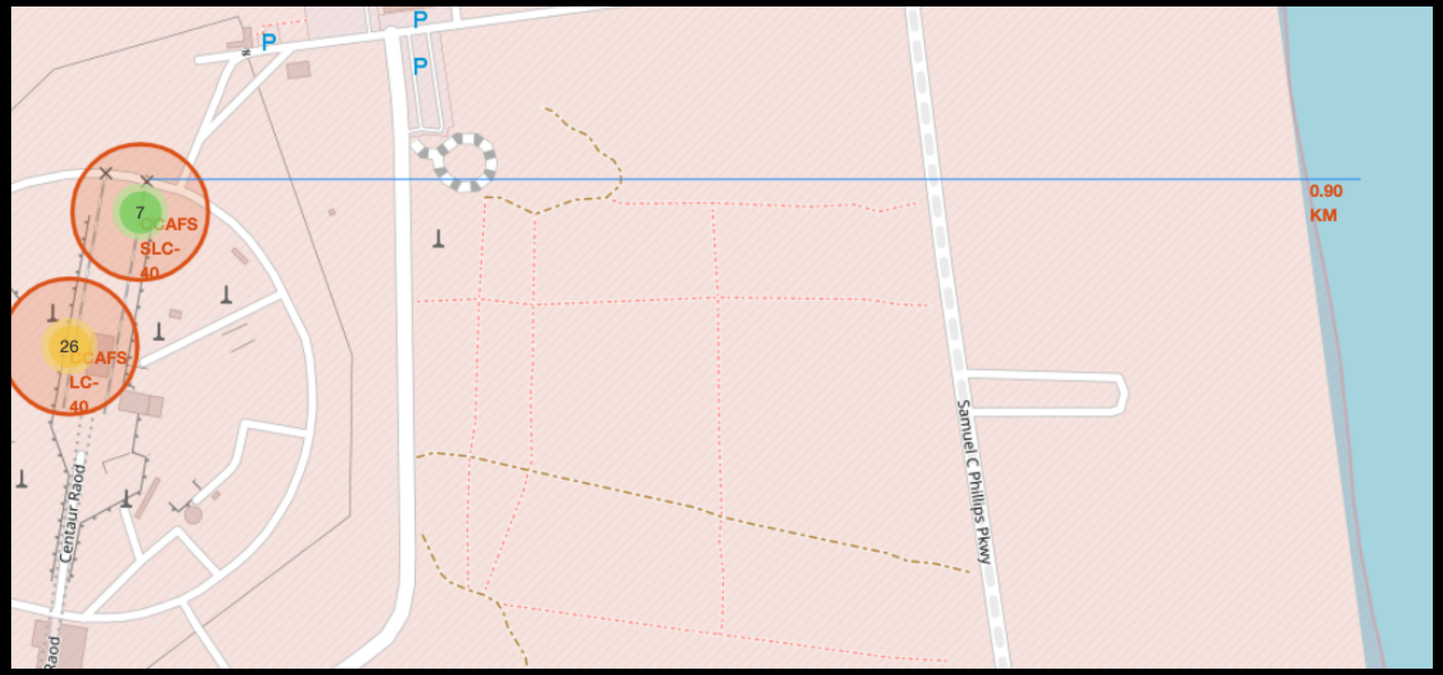
```
 * sqlite:///my_data1.db
Done.
```

Landing_Outcome	Numbers
Success	20
Success (drone ship)	8
Success (ground pad)	7

Rank the count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

Interactive Visual Analytics with Folium

- Mark all launch sites on a map
- Mark the success/failed launches for each site on the map
- Calculate the distances between a launch site to its proximities



Machine Learning Prediction

- Create a NumPy array from the column Class in data, by applying the method to_numpy() then assign it to the variable Y,make sure the output is a Pandas series (only one bracket df['name of column']).

- Standardize the data in X then reassign it to the variable X using the transform provided below.

- Use the function train_test_split to split the data X and Y into training and test data. Set the parameter test_size to 0.2 and random_state to 2. The training data and test data should be assigned to the following labels.

- Create a logistic regression object then create a GridSearchCV object logreg_cv with cv = 10. Fit the object to find the best parameters from the dictionary parameters.

Machine Learning Prediction

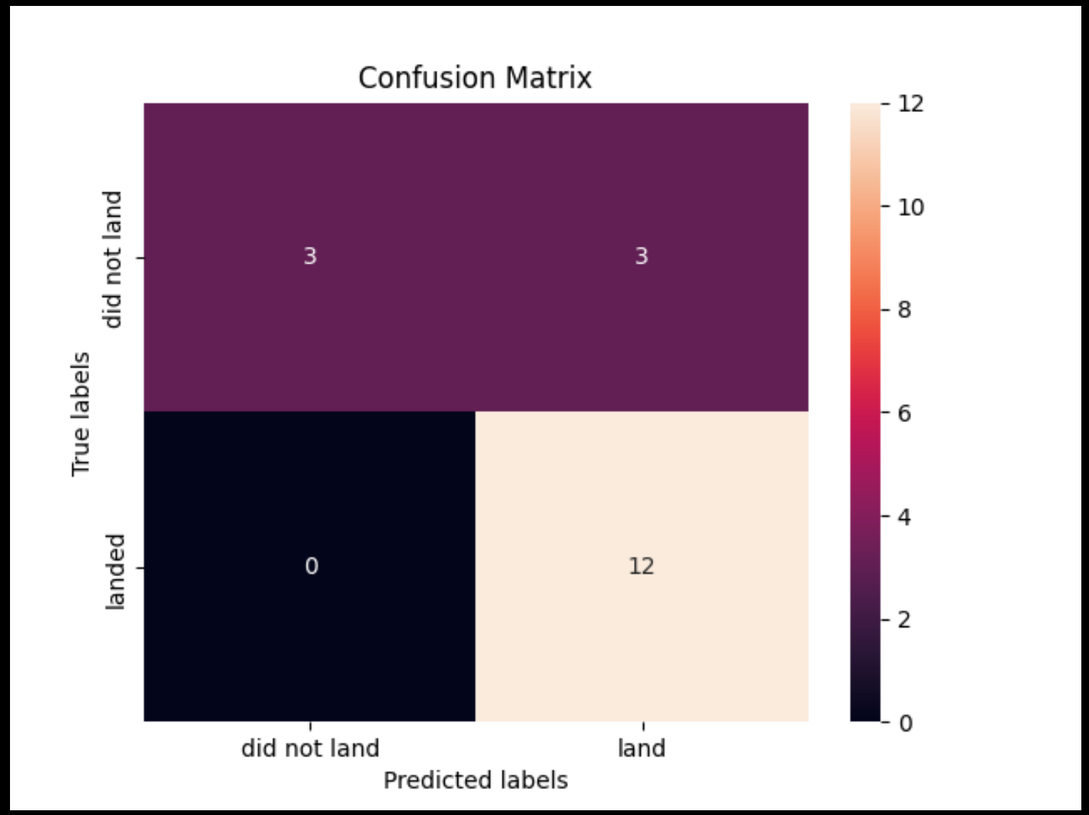
- We output the GridSearchCV object for logistic regression. We display the best parameters using the data attribute best_params_ and the accuracy on the validation data using the data attribute best_score_.

```
tuned hpyerparameters :(best parameters)  {'C': 0.01, 'penalty': 'l2', 'solver': 'lbfgs'}
accuracy : 0.8464285714285713
```

- Calculate the accuracy on the test data using the method score:

0.8333333333333334

Machine Learning Prediction



Machine Learning Prediction

- Create a support vector machine object then create a GridSearchCV object svm_cv with cv - 10. Fit the object to find the best parameters from the dictionary parameters.

accuracy : 0.8482142857142856

- Calculate the accuracy on the test data using the method score:

0.8333333333333334

Machine Learning Prediction

- Create a decision tree classifier object then create a GridSearchCV object tree_cv with cv = 10. Fit the object to find the best parameters from the dictionary parameters

accuracy : 0.871428571428571 3

- Calculate the accuracy of tree_cv on the test data using the method score:

0.8333333333333334

Machine Learning Prediction

- Create a k nearest neighbors object then create a GridSearchCV object knn_cv with cv = 10. Fit the object to find the best parameters from the dictionary parameters

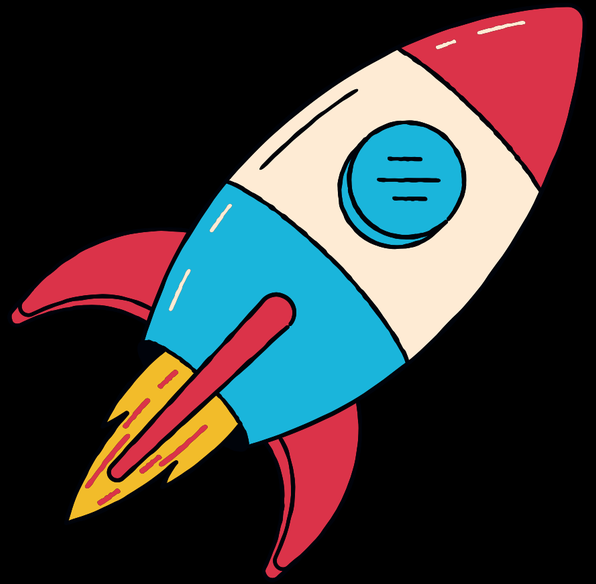
accuracy : 0.8482142857142858

- Calculate the accuracy of knn_cv on the test data using the method score:

0.8333333333333334

Machine Learning Prediction

"Practically all these algorithms give the same result"



Konu 05

Conclusion

- There is a correlation between launch site and success rate Payload mass is also associated with the success rate.: the more massive the payload, the less likely the first stage will return

- For orbit type, SO has the least success rate while ES-L1, GEO, HEO and SSO have the highest success rate According to the yearly trend

• There has been an increase in the success rate since 201 3 kept increasing till 2020 • With best parameter provided, decision tree classifier used in prediction yielded the highest accuracy of 84% .

Thank you!

Fatih Ilhan

SPACEX

Appendix
Resources

- https://github.com/fatihilhan42/SpaceX_Falcon-9_First_stage_landing_prediction/tree/main