Untitled5.ipynb

Code   git   No Kernel

[29]:
```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn import preprocessing
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn import metrics
from sklearn.metrics import classification_report, confusion_matrix
```

[2]:
```python
# Read the dataset
df = pd.read_csv("train.csv")
```

[3]:
```python
# show thw head of dataset
df.head()
```

[3]:

	Unnamed: 0	id	Gender	Customer Type	Age	Type of Travel	Class	Flight Distance	Inflight wifi service	Departure/Arrival time convenient	...	Inflight entertainment	On-board service	Leg room service	Baggage handling	Checkin service	Inflight service	Cleanliness	Departure Delay in Minutes	Arrival Delay in Minutes	satisfaction
0	0	70172	Male	Loyal Customer	13	Personal Travel	Eco Plus	460	3	4	...	5	4	3	4	4	5	5	25	18.0	neutral or dissatisfied
1	1	5047	Male	disloyal Customer	25	Business travel	Business	235	3	2	...	1	1	5	3	1	4	1	1	6.0	neutral or dissatisfied
2	2	110028	Female	Loyal Customer	26	Business travel	Business	1142	2	2	...	5	4	3	4	4	4	5	0	0.0	satisfied
3	3	24026	Female	Loyal Customer	25	Business travel	Business	562	2	5	...	2	2	5	3	1	4	2	11	9.0	neutral or dissatisfied
4	4	119299	Male	Loyal Customer	61	Business travel	Business	214	3	3	...	3	3	4	4	3	3	3	0	0.0	satisfied

5 rows × 25 columns

[5]:
```python
# Find Null values in variables
df.isnull().sum()
```

0 4 Git: idle No Kernel | Unknown   Mode: Edit Ln 1, Col 27 Untitled5.ipynb

Untitled5.ipynb

Code

No Kernel

5 rows × 25 columns

[5]:
```
# Find Null values in variables
df.isnull().sum()
```

[5]:
```
Unnamed: 0                           0
id                                   0
Gender                               0
Customer Type                        0
Age                                  0
Type of Travel                       0
Class                                0
Flight Distance                      0
Inflight wifi service                0
Departure/Arrival time convenient    0
Ease of Online booking               0
Gate location                        0
Food and drink                       0
Online boarding                      0
Seat comfort                         0
Inflight entertainment               0
On-board service                     0
Leg room service                     0
Baggage handling                     0
Checkin service                      0
Inflight service                     0
Cleanliness                          0
Departure Delay in Minutes           0
Arrival Delay in Minutes           310
satisfaction                         0
dtype: int64
```

[7]:
```
# Replace null value by most value
df["Arrival Delay in Minutes"].fillna(df['Arrival Delay in Minutes'].value_counts().idxmax(), inplace=True)
```

[8]:
```
df.isnull().sum()
```

0 4 Git: idle No Kernel | Unknown Mode: Command Ln 1, Col 27 Untitled5.ipynb

```
[8]: Unnamed: 0                          0
     id                                  0
     Gender                              0
     Customer Type                       0
     Age                                 0
     Type of Travel                      0
     Class                               0
     Flight Distance                     0
     Inflight wifi service               0
     Departure/Arrival time convenient   0
     Ease of Online booking              0
     Gate location                       0
     Food and drink                      0
     Online boarding                     0
     Seat comfort                        0
     Inflight entertainment              0
     On-board service                    0
     Leg room service                    0
     Baggage handling                    0
     Checkin service                     0
     Inflight service                    0
     Cleanliness                         0
     Departure Delay in Minutes          0
     Arrival Delay in Minutes            0
     satisfaction                        0
     dtype: int64
```

```
[9]: df.shape
```

```
[9]: (103904, 25)
```

```
[10]: df.info
```

```
[10]: <bound method DataFrame.info of         Unnamed: 0      id  Gender      Customer Type  Age   Type of Travel  \
      0                0   70172    Male     Loyal Customer   13  Personal Travel
      1                1    5047    Male  disloyal Customer   25  Business travel
      2                2  110028  Female     Loyal Customer   26  Business travel
      3                3   24026  Female     Loyal Customer   25  Business travel
      4                4  119299    Male     Loyal Customer   61  Business travel
      ...            ...     ...     ...                ...  ...              ...
      103899      103899   94171  Female  disloyal Customer   23  Business travel
      103900      103900   73097    Male     Loyal Customer   49  Business travel
      103901      103901   68825    Male  disloyal Customer   30  Business travel
      103902      103902   54173  Female  disloyal Customer   22  Business travel
      103903      103903   62567    Male     Loyal Customer   27  Business travel

                 Class  Flight Distance  Inflight wifi service  \
      0       Eco Plus              460                      3
      1       Business              235                      3
      2       Business             1142                      2
      3       Business              562                      2
      4       Business              214                      3
      ...          ...              ...                    ...
      103899       Eco              192                      2
      103900  Business             2347                      4
      103901  Business             1995                      1
      103902       Eco             1000                      1
      103903  Business             1723                      1
```

```
       Departure/Arrival time convenient  ...  Inflight entertainment  \
0                                      4  ...                       5
1                                      2  ...                       1
2                                      2  ...                       5
3                                      5  ...                       2
4                                      3  ...                       3
...                                  ...  ...                     ...
103899                                 1  ...                       2
103900                                 4  ...                       5
103901                                 1  ...                       4
```

```
        On-board service  Leg room service  Baggage handling  Checkin service  \
0                      4                 3                 4                4
1                      1                 5                 3                1
2                      4                 3                 4                4
3                      2                 5                 3                1
4                      3                 4                 4                3
...                  ...               ...               ...              ...
103899                 3                 1                 4                2
103900                 5                 5                 5                5
103901                 3                 2                 4                5
103902                 4                 5                 1                5
103903                 1                 1                 4                4

        Inflight service  Cleanliness  Departure Delay in Minutes  \
0                      5            5                          25
1                      4            1                           1
2                      4            5                           0
3                      4            2                          11
4                      3            3                           0
...                  ...          ...                         ...
103899                 3            2                           3
103900                 5            4                           0
103901                 5            4                           7
103902                 4            1                           0
103903                 3            1                           0

        Arrival Delay in Minutes             satisfaction
0                           18.0  neutral or dissatisfied
1                            6.0  neutral or dissatisfied
2                            0.0                satisfied
3                            9.0  neutral or dissatisfied
4                            0.0                satisfied
...                          ...                      ...
103899                       0.0  neutral or dissatisfied
103900                       0.0                satisfied
103901                      14.0  neutral or dissatisfied
103902                       0.0  neutral or dissatisfied
103903                       0.0  neutral or dissatisfied
```

[11]:
```
df.describe()
```

[11]:

	Unnamed: 0	id	Age	Flight Distance	Inflight wifi service	Departure/Arrival time convenient	Ease of Online booking	Gate location	Food and drink	Online boarding	Seat comfort	Inflight entertainment	On-board service	Leg room service	Baggage handling	Checkin service
count	103904.000000	103904.000000	103904.000000	103904.000000	103904.000000	103904.000000	103904.000000	103904.000000	103904.000000	103904.000000	103904.000000	103904.000000	103904.000000	103904.000000	103904.000000	103904.000000
mean	51951.500000	64924.210502	39.379706	1189.448375	2.729683	3.060296	2.756901	2.976883	3.202129	3.250375	3.439396	3.358158	3.382363	3.351055	3.631833	3.304290
std	29994.645522	37463.812252	15.114964	997.147281	1.327829	1.525075	1.398929	1.277621	1.329533	1.349509	1.319088	1.332991	1.288354	1.315605	1.180903	1.265396
min	0.000000	1.000000	7.000000	31.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000
25%	25975.750000	32533.750000	27.000000	414.000000	2.000000	2.000000	2.000000	2.000000	2.000000	2.000000	2.000000	2.000000	2.000000	2.000000	3.000000	3.000000
50%	51951.500000	64856.500000	40.000000	843.000000	3.000000	3.000000	3.000000	3.000000	3.000000	3.000000	4.000000	4.000000	4.000000	4.000000	4.000000	3.000000
75%	77927.250000	97368.250000	51.000000	1743.000000	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000	5.000000	4.000000	4.000000	4.000000	5.000000	4.000000
max	103903.000000	129880.000000	85.000000	4983.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000

[12]:
```
df.corr()
```

[12]:

	Unnamed: 0	id	Age	Flight Distance	Inflight wifi service	Departure/Arrival time convenient	Ease of Online booking	Gate location	Food and drink	Online boarding	Seat comfort	Inflight entertainment	On-board service	Leg room service	Baggage handling	Checkin service	Inflight service	Cleanliness	Departure Delay in Minutes	Arrival Delay in Minutes
Unnamed: 0	1.000000	0.002991	0.004786	0.002794	-0.002490	0.000739	0.001913	0.005073	-0.002162	0.001002	0.000044	0.001363	0.000813	0.004052	-0.000526	-0.004321	-0.000134	-0.001117	-0.000045	-0.000063
id	0.002991	1.000000	0.022857	0.095544	-0.021276	-0.002110	0.014163	-0.000606	0.001063	0.055477	0.052903	0.002300	0.055241	0.044634	0.074940	0.079273	0.079346	0.024965	-0.019546	-0.036997
Age	0.004786	0.022857	1.000000	0.099461	0.017859	0.038125	0.024842	-0.001330	0.023000	0.208939	0.160277	0.076444	0.057594	0.040583	-0.047529	0.035482	-0.049427	0.053611	-0.010152	-0.012105
Flight Distance	0.002794	0.095544	0.099461	1.000000	0.007131	-0.020043	0.065717	0.004793	0.056994	0.214869	0.157333	0.128740	0.109526	0.133916	0.063184	0.073072	0.057540	0.093149	0.002158	-0.002470
Inflight wifi service	-0.002490	-0.021276	0.017859	0.007131	1.000000	0.343845	0.715856	0.336248	0.134718	0.456970	0.122658	0.209321	0.121500	0.160473	0.120923	0.043193	0.110441	0.132698	-0.017402	-0.019042
Departure/Arrival time convenient	0.000739	-0.002110	0.038125	-0.020043	0.343845	1.000000	0.436961	0.444757	0.004906	0.070119	0.011344	-0.004861	0.068882	0.012441	0.072126	0.093333	0.073318	0.014292	0.001005	-0.000918
Ease of Online booking	0.001913	0.014163	0.024842	0.065717	0.715856	0.436961	1.000000	0.458655	0.031873	0.404074	0.030014	0.047032	0.038833	0.107601	0.038762	0.011081	0.035272	0.016179	-0.006371	-0.007947
Gate location	0.005073	-0.000606	-0.001330	0.004793	0.336248	0.444757	0.458655	1.000000	-0.001159	0.001688	0.003669	0.003517	-0.028373	-0.005873	0.002313	-0.035427	0.001681	-0.003830	0.005467	0.005178
Food and drink	-0.002162	0.001063	0.023000	0.056994	0.134718	0.004906	0.031873	-0.001159	1.000000	0.234468	0.574556	0.622512	0.059073	0.032498	0.034746	0.087299	0.033993	0.657760	-0.029926	-0.032466
Online boarding	0.001002	0.055477	0.208939	0.214869	0.456970	0.070119	0.404074	0.001688	0.234468	1.000000	0.420211	0.285066	0.155443	0.123950	0.083280	0.204462	0.074573	0.331517	-0.018982	-0.021874
Seat comfort	0.000044	0.052903	0.160277	0.157333	0.122658	0.011344	0.030014	0.003669	0.574556	0.420211	1.000000	0.610590	0.131971	0.105559	0.074542	0.191854	0.069218	0.678534	-0.027898	-0.029735
Inflight entertainment	0.001363	0.002300	0.076444	0.128740	0.209321	-0.004861	0.047032	0.003517	0.622512	0.285066	0.610590	1.000000	0.420153	0.299692	0.378210	0.120867	0.404855	0.691815	-0.027489	-0.030597
On-board service	0.000813	0.055241	0.057594	0.109526	0.121500	0.068882	0.038833	-0.028373	0.059073	0.155443	0.131971	0.420153	1.000000	0.355495	0.519134	0.243914	0.550782	0.123220	-0.031569	-0.035089
Leg room service	0.004052	0.044634	0.040583	0.133916	0.160473	0.012441	0.107601	-0.005873	0.032498	0.123950	0.105559	0.299692	0.355495	1.000000	0.369544	0.153137	0.368656	0.096370	0.014363	0.011924
Baggage handling	-0.000526	0.074940	-0.047529	0.063184	0.120923	0.072126	0.038762	0.002313	0.034746	0.083280	0.074542	0.378210	0.519134	0.369544	1.000000	0.233122	0.628561	0.095793	-0.005573	-0.008576
Checkin service	-0.004321	0.079273	0.035482	0.073072	0.043193	0.093333	0.011081	-0.035427	0.087299	0.204462	0.191854	0.120867	0.243914	0.153137	0.233122	1.000000	0.237197	0.179583	-0.018453	-0.020324
Inflight service	-0.000134	0.079346	-0.049427	0.057540	0.110441	0.073318	0.035272	0.001681	0.033993	0.074573	0.069218	0.404855	0.550782	0.368656	0.628561	0.237197	1.000000	0.088779	-0.054813	-0.058980
Cleanliness	-0.001117	0.024965	0.053611	0.093149	0.132698	0.014292	0.016179	-0.003830	0.657760	0.331517	0.678534	0.691815	0.123220	0.096370	0.095793	0.179583	0.088779	1.000000	-0.014093	-0.015732
Departure Delay in Minutes	-0.000045	-0.019546	-0.010152	0.002158	-0.017402	0.001005	-0.006371	0.005467	-0.029926	-0.018982	-0.027898	-0.027489	-0.031569	0.014363	-0.005573	-0.018453	-0.054813	-0.014093	1.000000	0.960247
Arrival Delay in Minutes	-0.000063	-0.036997	-0.012105	-0.002470	-0.019042	-0.000918	-0.007947	0.005178	-0.032466	-0.021874	-0.029735	-0.030597	-0.035089	0.011924	-0.008576	-0.020324	-0.058980	-0.015732	0.960247	1.000000

[16]:
```
# Find null values in variables
df.isnull().sum()
```

[16]:
```
Unnamed: 0                           0
id                                   0
Gender                               0
Customer Type                        0
Age                                  0
Type of Travel                       0
Class                                0
Flight Distance                      0
Inflight wifi service                0
Departure/Arrival time convenient    0
Ease of Online booking               0
Gate location                        0
Food and drink                       0
Online boarding                      0
Seat comfort                         0
Inflight entertainment               0
On-board service                     0
Leg room service                     0
Baggage handling                     0
Checkin service                      0
Inflight service                     0
Cleanliness                          0
Departure Delay in Minutes           0
Arrival Delay in Minutes             0
satisfaction                         0
dtype: int64
```

[18]:
```
#label_encoder
label_encoder = preprocessing.LabelEncoder()
df['Age']= label_encoder.fit_transform(df['Age'])
df['Gender']= label_encoder.fit_transform(df['Gender'])
df['Customer Type']= label_encoder.fit_transform(df['Customer Type'])
df['Type of Travel']= label_encoder.fit_transform(df['Type of Travel'])
df['Class']= label_encoder.fit_transform(df['Class'])
df['Flight Distance']= label_encoder.fit_transform(df['Flight Distance'])
df['Inflight wifi service']= label_encoder.fit_transform(df['Inflight wifi service'])
df['Departure/Arrival time convenient']= label_encoder.fit_transform(df['Departure/Arrival time convenient'])
df['Ease of Online booking']= label_encoder.fit_transform(df['Ease of Online booking'])
df['Gate location']= label_encoder.fit_transform(df['Gate location'])
df['Food and drink']= label_encoder.fit_transform(df['Food and drink'])
df['Online boarding']= label_encoder.fit_transform(df['Online boarding'])
df['Seat comfort']= label_encoder.fit_transform(df['Seat comfort'])
df['Inflight entertainment']= label_encoder.fit_transform(df['Inflight entertainment'])
df['On-board service']= label_encoder.fit_transform(df['On-board service'])
df['Leg room service']= label_encoder.fit_transform(df['Leg room service'])
df['Baggage handling']= label_encoder.fit_transform(df['Baggage handling'])
df['Checkin service']= label_encoder.fit_transform(df['Checkin service'])
df['Inflight service']= label_encoder.fit_transform(df['Inflight service'])
df['Cleanliness']= label_encoder.fit_transform(df['Cleanliness'])
df['Departure Delay in Minutes']= label_encoder.fit_transform(df['Departure Delay in Minutes'])
df['Arrival Delay in Minutes']= label_encoder.fit_transform(df['Arrival Delay in Minutes'])
df['satisfaction']= label_encoder.fit_transform(df['satisfaction'])
```

[19]:
```
df.head()
```

[19]:

	Unnamed: 0	id	Gender	Customer Type	Age	Type of Travel	Class	Flight Distance	Inflight wifi service	Departure/Arrival time convenient	...	Inflight entertainment	On-board service	Leg room service	Baggage handling	Checkin service	Inflight service	Cleanliness	Departure Delay in Minutes	Arrival Delay in Minutes	satisfaction
0	0	70172	1	0	6	1	2	358	3	4	...	5	4	3	3	4	5	5	25	18	0
1	1	5047	1	1	18	0	0	144	3	2	...	1	1	5	2	1	4	1	1	6	0
2	2	110028	0	0	19	0	0	994	2	2	...	5	4	3	3	4	4	5	0	0	1
3	3	24026	0	0	18	0	0	446	2	5	...	2	2	5	2	1	4	2	11	9	0
4	4	119299	1	0	54	0	0	124	3	3	...	3	3	4	3	3	3	3	0	0	1

5 rows × 25 columns

[22]:
```
X = df.drop('satisfaction', 1)
y = df.satisfaction
```

[23]:
```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=2)
```

[25]:
```
logicRegression_model = LogisticRegression()
logicRegression_model.fit(X_train, y_train)
```

[25]:
```
LogisticRegression()
```

[26]:
```
pred = logicRegression_model.predict(X_test)
```

[27]:
```
pred
```

[27]:
```
array([1, 0, 0, ..., 0, 0, 0])
```

[28]:
```
#comparsion
a = pd.DataFrame({'Actual value': y_test, 'Predicted value':pred})
a.head()
```

[28]:

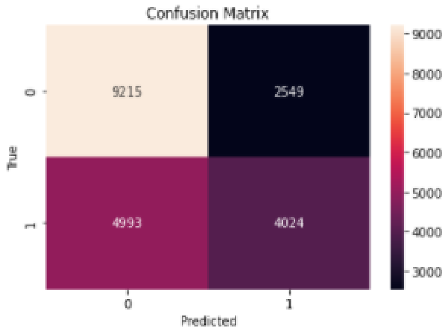
	Actual value	Predicted value
18981	1	1
4555	0	0
44022	0	0
84411	0	0
91989	1	0

[30]:
```
# Evaluation model
EvalMatrix = confusion_matrix(y_test, pred)
sns.heatmap(EvalMatrix, annot=True, fmt="d")
plt.title('Confusion Matrix')
plt.xlabel('Predicted')
plt.ylabel('True')
print(classification_report(y_test, pred))
```

```
              precision    recall  f1-score   support

           0       0.65      0.78      0.71     11764
           1       0.61      0.45      0.52      9017

    accuracy                           0.64     20781
   macro avg       0.63      0.61      0.61     20781
weighted avg       0.63      0.64      0.63     20781
```



Confusion Matrix
9215
2549
4993
4024
True
Predicted
0
1
0
1
9000
8000
7000
6000
5000
4000
3000