libaries importing

In [1]:

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings("ignore")
pd.set_option("display.max_columns",None)
pd.set_option("display.max_rows",None)
```

In [2]:

```python
data=r"C:\DsTraining\five dataset for clening\loan_defaulter.csv"
df=pd.read_csv(data)
df.head()
```

Out[2]:

	SK_ID_CURR	TARGET	NAME_CONTRACT_TYPE	CODE_GENDER	FLAG_OWN_CAR	FLAG_O
0	100002	1	Cash loans	M	N	
1	100003	0	Cash loans	F	N	
2	100004	0	Revolving loans	M	Y	
3	100006	0	Cash loans	F	N	
4	100007	0	Cash loans	M	N	

In [3]:

```python
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 307511 entries, 0 to 307510
Columns: 122 entries, SK_ID_CURR to AMT_REQ_CREDIT_BUREAU_YEAR
dtypes: float64(65), int64(41), object(16)
memory usage: 286.2+ MB
```

In [4]:

```
null_var=df.isnull().sum()/df.shape[0]*100
null_var
```

```
YEARS_BUILD_AVG                 66.497784
COMMONAREA_AVG                  69.872297
ELEVATORS_AVG                   53.295980
ENTRANCES_AVG                   50.348768
FLOORSMAX_AVG                   49.760822
FLOORSMIN_AVG                   67.848630
LANDAREA_AVG                    59.376738
LIVINGAPARTMENTS_AVG            68.354953
LIVINGAREA_AVG                  50.193326
NONLIVINGAPARTMENTS_AVG         69.432963
NONLIVINGAREA_AVG               55.179164
APARTMENTS_MODE                 50.749729
BASEMENTAREA_MODE               58.515956
YEARS_BEGINEXPLUATATION_MODE    48.781019
YEARS_BUILD_MODE                66.497784
COMMONAREA_MODE                 69.872297
ELEVATORS_MODE                  53.295980
ENTRANCES_MODE                  50.348768
FLOORSMAX_MODE                  49.760822
FLOORSMIN_MODE                  67.848630
```

In [5]:

```
null_var=df.isnull().sum()/df.shape[0]*100
null_col=null_var[null_var>17].keys()
df=df.drop(columns=null_col)
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 307511 entries, 0 to 307510
Data columns (total 71 columns):
 #   Column                       Non-Null Count   Dtype  
---  ------                       --------------   -----  
 0   SK_ID_CURR                   307511 non-null  int64  
 1   TARGET                       307511 non-null  int64  
 2   NAME_CONTRACT_TYPE           307511 non-null  object 
 3   CODE_GENDER                  307511 non-null  object 
 4   FLAG_OWN_CAR                 307511 non-null  object 
 5   FLAG_OWN_REALTY              307511 non-null  object 
 6   CNT_CHILDREN                 307511 non-null  int64  
 7   AMT_INCOME_TOTAL             307511 non-null  float64
 8   AMT_CREDIT                   307511 non-null  float64
 9   AMT_ANNUITY                  307499 non-null  float64
 10  AMT_GOODS_PRICE              307233 non-null  float64
 11  NAME_TYPE_SUITE              306219 non-null  object 
 12  NAME_INCOME_TYPE             307511 non-null  object 
 13  NAME_EDUCATION_TYPE          307511 non-null  object 
 14  NAME_FAMILY_STATUS           307511 non-null  object 
 15  NAME_HOUSING_TYPE            307511 non-null  object 
 16  REGION_POPULATION_RELATIVE   307511 non-null  float64
 17  DAYS_BIRTH                   307511 non-null  int64  
 18  DAYS_EMPLOYED                307511 non-null  int64  
 19  DAYS_REGISTRATION            307511 non-null  float64
 20  DAYS_ID_PUBLISH              307511 non-null  int64  
 21  FLAG_MOBIL                   307511 non-null  int64  
 22  FLAG_EMP_PHONE               307511 non-null  int64  
 23  FLAG_WORK_PHONE              307511 non-null  int64  
 24  FLAG_CONT_MOBILE             307511 non-null  int64  
 25  FLAG_PHONE                   307511 non-null  int64  
 26  FLAG_EMAIL                   307511 non-null  int64  
 27  CNT_FAM_MEMBERS              307509 non-null  float64
 28  REGION_RATING_CLIENT         307511 non-null  int64  
 29  REGION_RATING_CLIENT_W_CITY  307511 non-null  int64  
 30  WEEKDAY_APPR_PROCESS_START   307511 non-null  object 
 31  HOUR_APPR_PROCESS_START      307511 non-null  int64  
 32  REG_REGION_NOT_LIVE_REGION   307511 non-null  int64  
 33  REG_REGION_NOT_WORK_REGION   307511 non-null  int64  
 34  LIVE_REGION_NOT_WORK_REGION  307511 non-null  int64  
 35  REG_CITY_NOT_LIVE_CITY       307511 non-null  int64  
 36  REG_CITY_NOT_WORK_CITY       307511 non-null  int64  
 37  LIVE_CITY_NOT_WORK_CITY      307511 non-null  int64  
 38  ORGANIZATION_TYPE            307511 non-null  object 
 39  EXT_SOURCE_2                 306851 non-null  float64
 40  OBS_30_CNT_SOCIAL_CIRCLE     306490 non-null  float64
 41  DEF_30_CNT_SOCIAL_CIRCLE     306490 non-null  float64
 42  OBS_60_CNT_SOCIAL_CIRCLE     306490 non-null  float64
 43  DEF_60_CNT_SOCIAL_CIRCLE     306490 non-null  float64
 44  DAYS_LAST_PHONE_CHANGE       307510 non-null  float64
 45  FLAG_DOCUMENT_2              307511 non-null  int64  
 46  FLAG_DOCUMENT_3              307511 non-null  int64  
 47  FLAG_DOCUMENT_4              307511 non-null  int64  
 48  FLAG_DOCUMENT_5              307511 non-null  int64  
 49  FLAG_DOCUMENT_6              307511 non-null  int64  
 50  FLAG_DOCUMENT_7              307511 non-null  int64  
 51  FLAG_DOCUMENT_8              307511 non-null  int64  
 52  FLAG_DOCUMENT_9              307511 non-null  int64
 53  FLAG_DOCUMENT_10             307511 non-null  int64
 54  FLAG_DOCUMENT_11             307511 non-null  int64
 55  FLAG_DOCUMENT_12             307511 non-null  int64
 56  FLAG_DOCUMENT_13             307511 non-null  int64
 57  FLAG_DOCUMENT_14             307511 non-null  int64
 58  FLAG_DOCUMENT_15             307511 non-null  int64
 59  FLAG_DOCUMENT_16             307511 non-null  int64
 60  FLAG_DOCUMENT_17             307511 non-null  int64
 61  FLAG_DOCUMENT_18             307511 non-null  int64
 62  FLAG_DOCUMENT_19             307511 non-null  int64
 63  FLAG_DOCUMENT_20             307511 non-null  int64
 64  FLAG_DOCUMENT_21             307511 non-null  int64
 65  AMT_REQ_CREDIT_BUREAU_HOUR   265992 non-null  float64
 66  AMT_REQ_CREDIT_BUREAU_DAY    265992 non-null  float64
 67  AMT_REQ_CREDIT_BUREAU_WEEK   265992 non-null  float64
 68  AMT_REQ_CREDIT_BUREAU_MON    265992 non-null  float64
 69  AMT_REQ_CREDIT_BUREAU_QRT    265992 non-null  float64
 70  AMT_REQ_CREDIT_BUREAU_YEAR   265992 non-null  float64
dtypes: float64(19), int64(41), object(11)
memory usage: 166.6+ MB
```

In [6]:

```python
df=df.dropna()
```

In [7]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 263423 entries, 0 to 307510
Data columns (total 71 columns):
 #   Column                       Non-Null Count   Dtype
---  ------                       --------------   -----
 0   SK_ID_CURR                   263423 non-null  int64
 1   TARGET                       263423 non-null  int64
 2   NAME_CONTRACT_TYPE           263423 non-null  object
 3   CODE_GENDER                  263423 non-null  object
 4   FLAG_OWN_CAR                 263423 non-null  object
 5   FLAG_OWN_REALTY              263423 non-null  object
 6   CNT_CHILDREN                 263423 non-null  int64
 7   AMT_INCOME_TOTAL             263423 non-null  float64
 8   AMT_CREDIT                   263423 non-null  float64
 9   AMT_ANNUITY                  263423 non-null  float64
 10  AMT_GOODS_PRICE              263423 non-null  float64
 11  NAME_TYPE_SUITE              263423 non-null  object
 12  NAME_INCOME_TYPE             263423 non-null  object
 13  NAME_EDUCATION_TYPE          263423 non-null  object
 14  NAME_FAMILY_STATUS           263423 non-null  object
 15  NAME_HOUSING_TYPE            263423 non-null  object
 16  REGION_POPULATION_RELATIVE   263423 non-null  float64
 17  DAYS_BIRTH                   263423 non-null  int64
 18  DAYS_EMPLOYED                263423 non-null  int64
 19  DAYS_REGISTRATION            263423 non-null  float64
 20  DAYS_ID_PUBLISH              263423 non-null  int64
 21  FLAG_MOBIL                   263423 non-null  int64
 22  FLAG_EMP_PHONE               263423 non-null  int64
 23  FLAG_WORK_PHONE              263423 non-null  int64
 24  FLAG_CONT_MOBILE             263423 non-null  int64
 25  FLAG_PHONE                   263423 non-null  int64
 26  FLAG_EMAIL                   263423 non-null  int64
 27  CNT_FAM_MEMBERS              263423 non-null  float64
 28  REGION_RATING_CLIENT         263423 non-null  int64
 29  REGION_RATING_CLIENT_W_CITY  263423 non-null  int64
 30  WEEKDAY_APPR_PROCESS_START   263423 non-null  object
 31  HOUR_APPR_PROCESS_START      263423 non-null  int64
 32  REG_REGION_NOT_LIVE_REGION   263423 non-null  int64
 33  REG_REGION_NOT_WORK_REGION   263423 non-null  int64
 34  LIVE_REGION_NOT_WORK_REGION  263423 non-null  int64
 35  REG_CITY_NOT_LIVE_CITY       263423 non-null  int64
 36  REG_CITY_NOT_WORK_CITY       263423 non-null  int64
 37  LIVE_CITY_NOT_WORK_CITY      263423 non-null  int64
 38  ORGANIZATION_TYPE            263423 non-null  object
 39  EXT_SOURCE_2                 263423 non-null  float64
 40  OBS_30_CNT_SOCIAL_CIRCLE     263423 non-null  float64
 41  DEF_30_CNT_SOCIAL_CIRCLE     263423 non-null  float64
 42  OBS_60_CNT_SOCIAL_CIRCLE     263423 non-null  float64
 43  DEF_60_CNT_SOCIAL_CIRCLE     263423 non-null  float64
 44  DAYS_LAST_PHONE_CHANGE       263423 non-null  float64
 45  FLAG_DOCUMENT_2              263423 non-null  int64
 46  FLAG_DOCUMENT_3              263423 non-null  int64
 47  FLAG_DOCUMENT_4              263423 non-null  int64
 48  FLAG_DOCUMENT_5              263423 non-null  int64
 49  FLAG_DOCUMENT_6              263423 non-null  int64
 50  FLAG_DOCUMENT_7              263423 non-null  int64
 51  FLAG_DOCUMENT_8              263423 non-null  int64
 52  FLAG_DOCUMENT_9             263423 non-null  int64
 53  FLAG_DOCUMENT_10            263423 non-null  int64
 54  FLAG_DOCUMENT_11            263423 non-null  int64
 55  FLAG_DOCUMENT_12            263423 non-null  int64
 56  FLAG_DOCUMENT_13            263423 non-null  int64
 57  FLAG_DOCUMENT_14            263423 non-null  int64
 58  FLAG_DOCUMENT_15            263423 non-null  int64
 59  FLAG_DOCUMENT_16            263423 non-null  int64
 60  FLAG_DOCUMENT_17            263423 non-null  int64
 61  FLAG_DOCUMENT_18            263423 non-null  int64
 62  FLAG_DOCUMENT_19            263423 non-null  int64
 63  FLAG_DOCUMENT_20            263423 non-null  int64
 64  FLAG_DOCUMENT_21            263423 non-null  int64
 65  AMT_REQ_CREDIT_BUREAU_HOUR  263423 non-null  float64
 66  AMT_REQ_CREDIT_BUREAU_DAY   263423 non-null  float64
 67  AMT_REQ_CREDIT_BUREAU_WEEK  263423 non-null  float64
 68  AMT_REQ_CREDIT_BUREAU_MON   263423 non-null  float64
 69  AMT_REQ_CREDIT_BUREAU_QRT   263423 non-null  float64
 70  AMT_REQ_CREDIT_BUREAU_YEAR  263423 non-null  float64
dtypes: float64(19), int64(41), object(11)
memory usage: 144.7+ MB
```

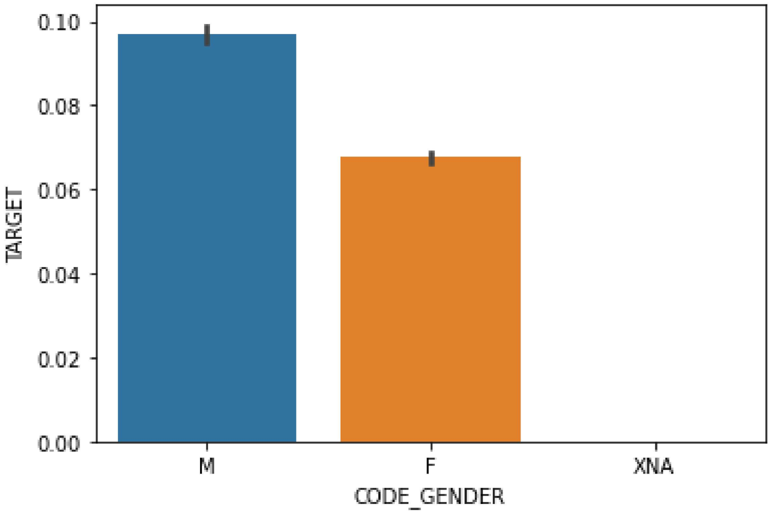
EDA

In [8]:

```
sns.barplot(y="TARGET",x="CODE_GENDER",data=df)
```

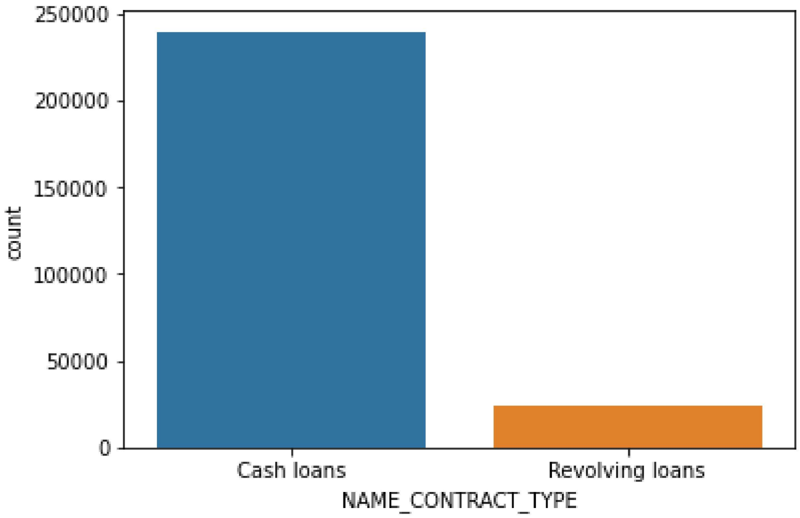
Out[8]:

```
<AxesSubplot:xlabel='CODE_GENDER', ylabel='TARGET'>
```



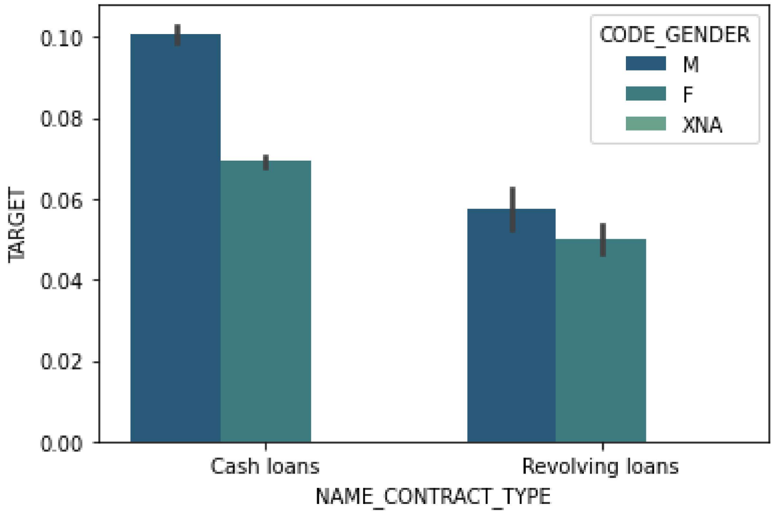
```
In [9]:  sns.countplot(x="NAME_CONTRACT_TYPE",data=df)
```

```
Out[9]: <AxesSubplot:xlabel='NAME_CONTRACT_TYPE', ylabel='count'>
```



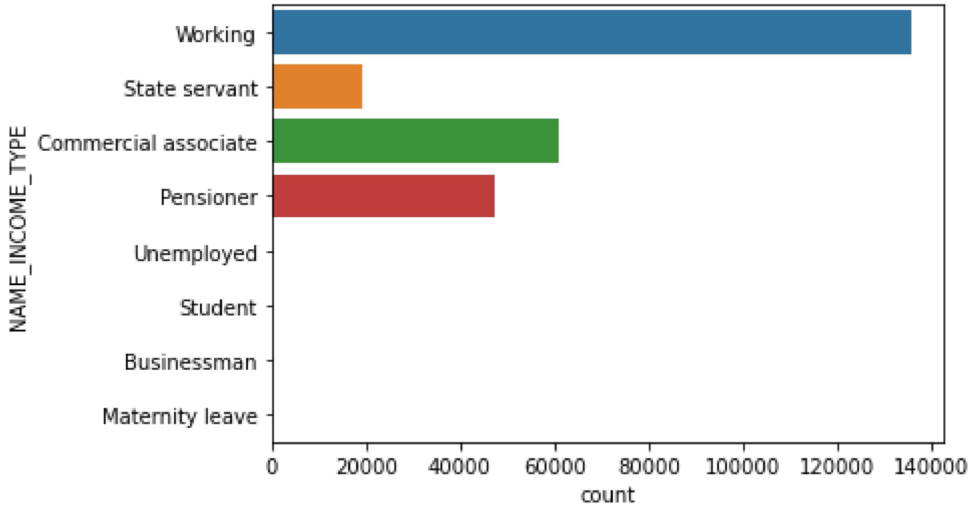
```
In [10]:  sns.barplot(x="NAME_CONTRACT_TYPE",y="TARGET",hue="CODE_GENDER",data=df,pa
```

```
Out[10]: <AxesSubplot:xlabel='NAME_CONTRACT_TYPE', ylabel='TARGET'>
```



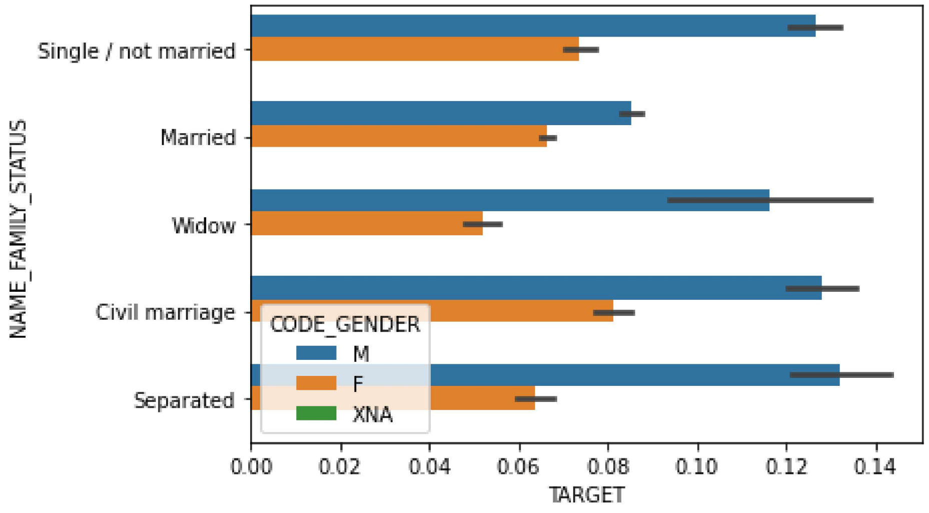
```
In [11]: sns.countplot(y="NAME_INCOME_TYPE",data=df)
```

```
Out[11]: <AxesSubplot:xlabel='count', ylabel='NAME_INCOME_TYPE'>
```



```
In [12]: sns.barplot(x="TARGET",y="NAME_FAMILY_STATUS",hue="CODE_GENDER",data=df)
```

```
Out[12]: <AxesSubplot:xlabel='TARGET', ylabel='NAME_FAMILY_STATUS'>
```

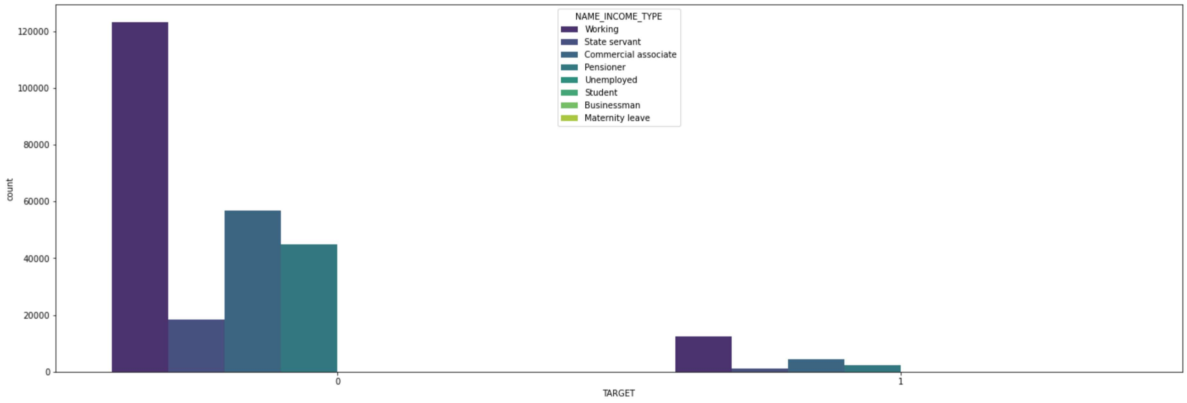


In [13]:

```
plt.figure(figsize=(24,8))
sns.countplot(data=df,x="TARGET",hue="NAME_INCOME_TYPE",palette="viridis"
```

Out[13]:

```
<AxesSubplot:xlabel='TARGET', ylabel='count'>
```



In [ ]:

```

```

In [14]:

```
#sns.displot(data=df,x="AMT_INCOME_TOTAL",hue="TARGET",kde=True)
```

PreProcessing

In [15]:

```
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
object_list=df.select_dtypes(include=['object']).columns
for i in object_list:
    df[i]=le.fit_transform(df[i])
```

Slicing

In [25]:

```
x=df.iloc[:,2:].values
y=df.iloc[:,1].values
```

Split data into train and test

In [17]:

```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,random_state=42,test_s
```

Scaling

In [18]:

```
from sklearn.preprocessing import StandardScaler
sc=StandardScaler()
x_train=sc.fit_transform(x_train)
x_test=sc.transform(x_test)
```

check accuracy score

In [19]:

```
from sklearn.linear_model import LogisticRegression
classifier=LogisticRegression(random_state=0)
classifier.fit(x_train,y_train)
y_pred=classifier.predict(x_test)
```

In [27]:

```
from sklearn.metrics import accuracy_score
accuracy_score(y_test,y_pred)*100
```

Out[27]:

```
92.17044699629876
```

In [31]:

```
import xgboost as xgb
xgb.XGBClassifier().get_params()
xg_classifier = xgb.XGBClassifier()
xg_classifier.fit(x_train,y_train)
xgb_preds = xg_classifier.predict(x_test)
print("The score of XGBoost classifier is",xg_classifier.score(x_test, y_
```

```
The score of XGBoost classifier is 92.14577204137801
```

In [ ]:

```

```