titanic

September 19, 2024

[5]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import math
import seaborn as sns
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
import warnings
warnings.filterwarnings("ignore")
```

[6]:

```
df = pd.read_csv(r"C:\Users\hp\Downloads\archive (13)\Titanic-Dataset.csv")
df
```

[6]:

```
     PassengerId  Survived  Pclass  \
0              1         0       3
1              2         1       1
2              3         1       3
3              4         1       1
4              5         0       3
..           …       …     …
886          887         0       2
887          888         1       1
888          889         0       3
889          890         1       1
890          891         0       3

                                                  Name     Sex   Age  SibSp  \
0                               Braund, Mr. Owen Harris    male  22.0      1
1    Cumings, Mrs. John Bradley (Florence Briggs Th…  female  38.0      1
2                                Heikkinen, Miss. Laina  female  26.0      0
3          Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0      1
4                              Allen, Mr. William Henry    male  35.0      0
..                                                 …     …   …    …
886                               Montvila, Rev. Juozas    male  27.0      0
887                        Graham, Miss. Margaret Edith  female  19.0      0
888            Johnston, Miss. Catherine Helen "Carrie"  female   NaN      1
889                               Behr, Mr. Karl Howell    male  26.0      0
890                                 Dooley, Mr. Patrick    male  32.0      0
```

```
     Parch            Ticket     Fare Cabin Embarked
0        0         A/5 21171   7.2500   NaN        S
1        0          PC 17599  71.2833   C85        C
2        0  STON/O2. 3101282   7.9250   NaN        S
3        0            113803  53.1000  C123        S
4        0            373450   8.0500   NaN        S
..     …              …       …   …      …
886      0            211536  13.0000   NaN        S
887      0            112053  30.0000   B42        S
888      2        W./C. 6607  23.4500   NaN        S
889      0            111369  30.0000  C148        C
890      0            370376   7.7500   NaN        Q

[891 rows x 12 columns]
```

[7]:
```
df.describe()
```

[7]:
```
       PassengerId    Survived      Pclass         Age       SibSp  \
count   891.000000  891.000000  891.000000  714.000000  891.000000
mean    446.000000    0.383838    2.308642   29.699118    0.523008
std     257.353842    0.486592    0.836071   14.526497    1.102743
min       1.000000    0.000000    1.000000    0.420000    0.000000
25%     223.500000    0.000000    2.000000   20.125000    0.000000
50%     446.000000    0.000000    3.000000   28.000000    0.000000
75%     668.500000    1.000000    3.000000   38.000000    1.000000
max     891.000000    1.000000    3.000000   80.000000    8.000000

            Parch        Fare
count  891.000000  891.000000
mean     0.381594   32.204208
std      0.806057   49.693429
min      0.000000    0.000000
25%      0.000000    7.910400
50%      0.000000   14.454200
75%      0.000000   31.000000
max      6.000000  512.329200
```

[8]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   PassengerId  891 non-null    int64
 1   Survived     891 non-null    int64
 2   Pclass       891 non-null    int64
 3   Name         891 non-null    object
 4   Sex          891 non-null    object
 5   Age          714 non-null    float64
 6   SibSp        891 non-null    int64
 7   Parch        891 non-null    int64
 8   Ticket       891 non-null    object
 9   Fare         891 non-null    float64
 10  Cabin        204 non-null    object
 11  Embarked     889 non-null    object
dtypes: float64(2), int64(5), object(5)
memory usage: 83.7+ KB
```

[9]:

```
df.isnull().sum()
```

[9]:

```
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age            177
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          687
Embarked         2
dtype: int64
```

[10]:

```
df = df.drop(columns ='Cabin',axis =1)
df
```

[10]:

```
     PassengerId  Survived  Pclass  \
0              1         0       3
1              2         1       1
2              3         1       3
3              4         1       1
4              5         0       3
..           …         …       …
886          887         0       2
887          888         1       1
888          889         0       3
889          890         1       1
890          891         0       3

                                                  Name     Sex   Age  SibSp  \
0                              Braund, Mr. Owen Harris    male  22.0      1
1    Cumings, Mrs. John Bradley (Florence Briggs Th…  female  38.0      1
2                               Heikkinen, Miss. Laina  female  26.0      0
```

```
3         Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0      1
4                             Allen, Mr. William Henry    male  35.0      0
..                                                 …     …   …    …
886                              Montvila, Rev. Juozas    male  27.0      0
887                       Graham, Miss. Margaret Edith  female  19.0      0
888           Johnston, Miss. Catherine Helen "Carrie"  female   NaN      1
889                              Behr, Mr. Karl Howell    male  26.0      0
890                                Dooley, Mr. Patrick    male  32.0      0

     Parch            Ticket     Fare Embarked
0        0         A/5 21171   7.2500        S
1        0          PC 17599  71.2833        C
2        0  STON/O2. 3101282   7.9250        S
3        0            113803  53.1000        S
4        0            373450   8.0500        S
..     …               …      …       …
886      0            211536  13.0000        S
887      0            112053  30.0000        S
888      2        W./C. 6607  23.4500        S
889      0            111369  30.0000        C
890      0            370376   7.7500        Q

[891 rows x 11 columns]
```

[11]:
```python
new_age = df.Age.mean()
new_age
```

[11]:
```
29.69911764705882
```

[12]:
```python
df["Age"] = df["Age"].fillna(new_age)
df
```

[12]:
```
     PassengerId  Survived  Pclass  \
0              1         0       3
1              2         1       1
2              3         1       3
3              4         1       1
4              5         0       3
..           …       …     …
886          887         0       2
887          888         1       1
888          889         0       3
889          890         1       1
890          891         0       3

                                                  Name     Sex        Age  \
0                              Braund, Mr. Owen Harris    male  22.000000
```

```
1    Cumings, Mrs. John Bradley (Florence Briggs Th…  female  38.000000
2                               Heikkinen, Miss. Laina  female  26.000000
3         Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.000000
4                             Allen, Mr. William Henry    male  35.000000
..                                                 …     …        …
886                              Montvila, Rev. Juozas    male  27.000000
887                       Graham, Miss. Margaret Edith  female  19.000000
888           Johnston, Miss. Catherine Helen "Carrie"  female  29.699118
889                              Behr, Mr. Karl Howell    male  26.000000
890                                Dooley, Mr. Patrick    male  32.000000

     SibSp  Parch            Ticket     Fare Embarked
0        1      0         A/5 21171   7.2500        S
1        1      0          PC 17599  71.2833        C
2        0      0  STON/O2. 3101282   7.9250        S
3        1      0            113803  53.1000        S
4        0      0            373450   8.0500        S
..     …    …               …      …      …
886      0      0            211536  13.0000        S
887      0      0            112053  30.0000        S
888      1      2        W./C. 6607  23.4500        S
889      0      0            111369  30.0000        C
890      0      0            370376   7.7500        Q

[891 rows x 11 columns]
```

```
[13]: df.Age.isnull().sum()
```

```
[13]: 0
```

```
[14]: df["Embarked"].mode()
```

```
[14]: 0    S
      Name: Embarked, dtype: object
```

```
[15]: new_Embarked = (df["Embarked"].mode()[0])
      new_Embarked
```

```
[15]: 'S'
```

```
[16]: df["Embarked"] = df["Embarked"].fillna(new_Embarked)
      df
```

```
[16]:      PassengerId  Survived  Pclass  \
      0              1         0       3
      1              2         1       1
      2              3         1       3
      3              4         1       1
```

```
4              5         0       3
..           …       …     …
886          887         0       2
887          888         1       1
888          889         0       3
889          890         1       1
890          891         0       3

                                                  Name     Sex        Age  \
0                              Braund, Mr. Owen Harris    male  22.000000
1    Cumings, Mrs. John Bradley (Florence Briggs Th…  female  38.000000
2                               Heikkinen, Miss. Laina  female  26.000000
3         Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.000000
4                             Allen, Mr. William Henry    male  35.000000
..                                                 …     …        …
886                              Montvila, Rev. Juozas    male  27.000000
887                       Graham, Miss. Margaret Edith  female  19.000000
888           Johnston, Miss. Catherine Helen "Carrie"  female  29.699118
889                              Behr, Mr. Karl Howell    male  26.000000
890                                Dooley, Mr. Patrick    male  32.000000

     SibSp  Parch            Ticket     Fare Embarked
0        1      0         A/5 21171   7.2500        S
1        1      0          PC 17599  71.2833        C
2        0      0  STON/O2. 3101282   7.9250        S
3        1      0            113803  53.1000        S
4        0      0            373450   8.0500        S
..     …    …               …      …      …
886      0      0            211536  13.0000        S
887      0      0            112053  30.0000        S
888      1      2        W./C. 6607  23.4500        S
889      0      0            111369  30.0000        C
890      0      0            370376   7.7500        Q

[891 rows x 11 columns]
```

[17]:

```
df["Embarked"].isnull().sum()
```

[17]:

```
0
```

[18]:

```
df.isnull().sum()
```

[18]:

```
PassengerId    0
Survived       0
Pclass         0
Name           0
Sex            0
```

```
Age            0
SibSp          0
Parch          0
Ticket         0
Fare           0
Embarked       0
dtype: int64
```

[19]:

```
df.describe()
```

[19]:

```
       PassengerId    Survived      Pclass         Age       SibSp  \
count   891.000000  891.000000  891.000000  891.000000  891.000000
mean    446.000000    0.383838    2.308642   29.699118    0.523008
std     257.353842    0.486592    0.836071   13.002015    1.102743
min       1.000000    0.000000    1.000000    0.420000    0.000000
25%     223.500000    0.000000    2.000000   22.000000    0.000000
50%     446.000000    0.000000    3.000000   29.699118    0.000000
75%     668.500000    1.000000    3.000000   35.000000    1.000000
max     891.000000    1.000000    3.000000   80.000000    8.000000

            Parch        Fare
count  891.000000  891.000000
mean     0.381594   32.204208
std      0.806057   49.693429
min      0.000000    0.000000
25%      0.000000    7.910400
50%      0.000000   14.454200
75%      0.000000   31.000000
max      6.000000  512.329200
```

[20]:

```
df["Embarked"].value_counts()
```

[20]:

```
Embarked
S    646
C    168
Q     77
Name: count, dtype: int64
```

[21]:

```
df["Survived"].value_counts()
```

[21]:

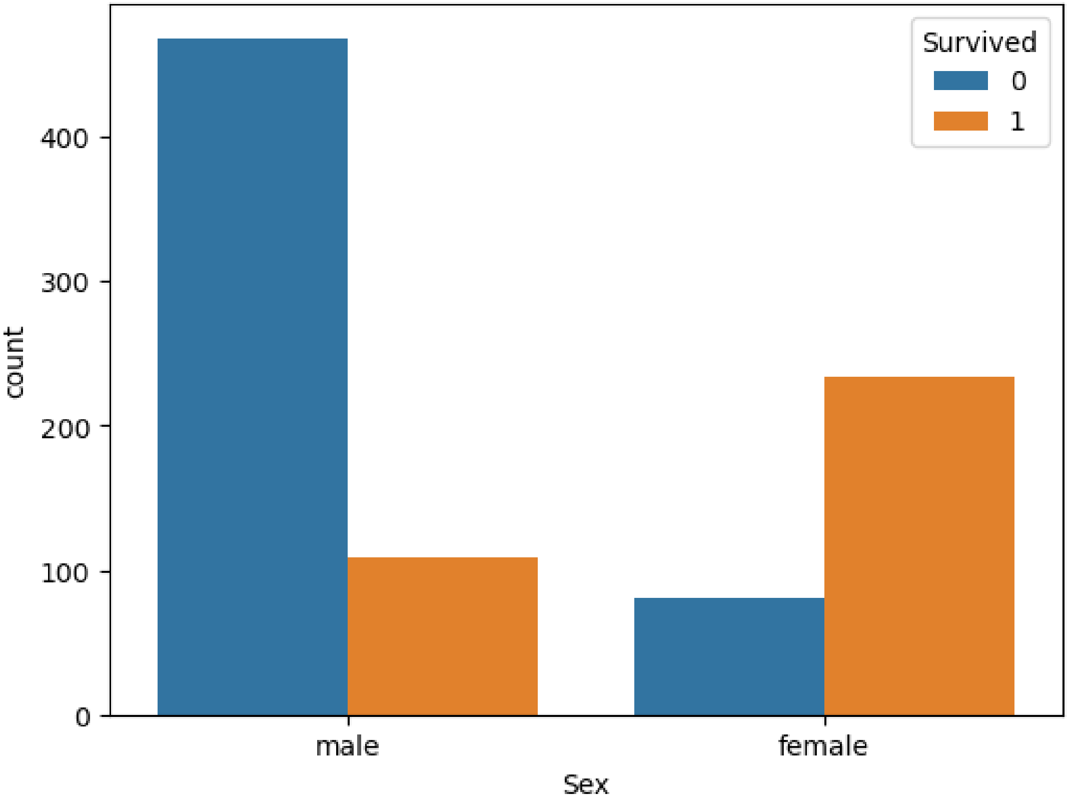
```
Survived
0    549
1    342
Name: count, dtype: int64
```

[22]:

```
sns.countplot(x="Sex",hue ="Survived",data=df)
```

[22]:

```
<Axes: xlabel='Sex', ylabel='count'>
```

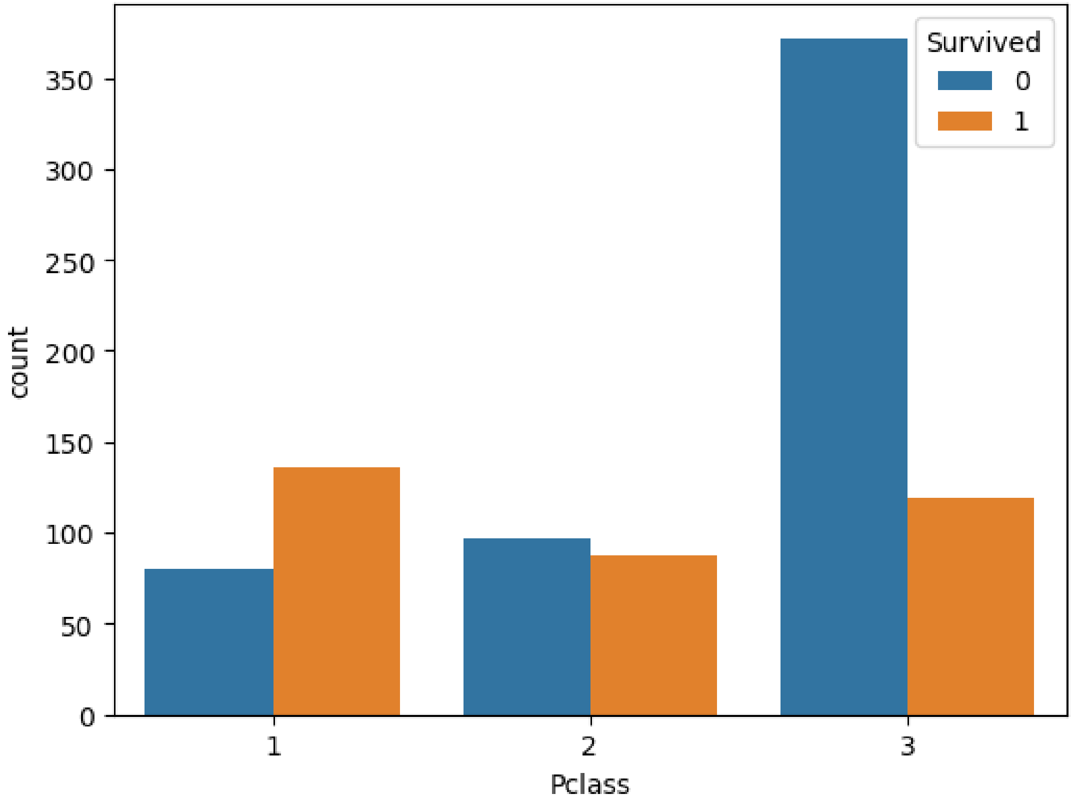


```
[23]: df["Pclass"].value_counts()
```

```
[23]: Pclass
      3    491
      1    216
      2    184
      Name: count, dtype: int64
```

```
[24]: sns.countplot(x="Pclass",hue ="Survived",data=df)
```

```
[24]: <Axes: xlabel='Pclass', ylabel='count'>
```



[25]:
```
df = df.replace({'Sex':{'male': 0 , 'female':1}, 'Embarked':{'S':0,'C':1,'Q':
 ↪2}})
df
```

[25]:
```
     PassengerId  Survived  Pclass  \
0              1         0       3
1              2         1       1
2              3         1       3
3              4         1       1
4              5         0       3
..           …         …       …
886          887         0       2
887          888         1       1
888          889         0       3
889          890         1       1
890          891         0       3

                                                  Name  Sex        Age  SibSp  \
0                              Braund, Mr. Owen Harris    0  22.000000      1
1    Cumings, Mrs. John Bradley (Florence Briggs Th…    1  38.000000      1
2                               Heikkinen, Miss. Laina    1  26.000000      0
```

```
3         Futrelle, Mrs. Jacques Heath (Lily May Peel)    1  35.000000       1
4                             Allen, Mr. William Henry    0  35.000000       0
..                                                  …  …         …    …
886                              Montvila, Rev. Juozas    0  27.000000       0
887                       Graham, Miss. Margaret Edith    1  19.000000       0
888           Johnston, Miss. Catherine Helen "Carrie"    1  29.699118       1
889                              Behr, Mr. Karl Howell    0  26.000000       0
890                                Dooley, Mr. Patrick    0  32.000000       0

     Parch            Ticket     Fare  Embarked
0        0         A/5 21171   7.2500         0
1        0          PC 17599  71.2833         1
2        0  STON/O2. 3101282   7.9250         0
3        0            113803  53.1000         0
4        0            373450   8.0500         0
..     …               …       …       …
886      0            211536  13.0000         0
887      0            112053  30.0000         0
888      2        W./C. 6607  23.4500         0
889      0            111369  30.0000         1
890      0            370376   7.7500         2

[891 rows x 11 columns]
```

[26]:
```
X = df.drop(columns =["Name", "PassengerId","Survived","Ticket"],axis =1)
X
```

[26]:
```
     Pclass  Sex        Age  SibSp  Parch     Fare  Embarked
0         3    0  22.000000      1      0   7.2500         0
1         1    1  38.000000      1      0  71.2833         1
2         3    1  26.000000      0      0   7.9250         0
3         1    1  35.000000      1      0  53.1000         0
4         3    0  35.000000      0      0   8.0500         0
..      …  …        …    …    …      …       …
886       2    0  27.000000      0      0  13.0000         0
887       1    1  19.000000      0      0  30.0000         0
888       3    1  29.699118      1      2  23.4500         0
889       1    0  26.000000      0      0  30.0000         1
890       3    0  32.000000      0      0   7.7500         2

[891 rows x 7 columns]
```

[27]:
```
Y = df.Survived
Y
```

[27]:
```
0      0
1      1
```

```
2      1
3      1
4      0
      ..
886    0
887    1
888    0
889    1
890    0
Name: Survived, Length: 891, dtype: int64
```

```
[28]: X_train,X_test,Y_train,Y_test = train_test_split(X,Y,test_size=0.2)
```

```
[29]: from sklearn.linear_model import LogisticRegression
      from sklearn.metrics import accuracy_score, confusion_matrix,␣
       ↪classification_report
```

```
[30]: model = LogisticRegression()
      model
```

```
[30]: LogisticRegression()
```

```
[31]: model.fit(X_train,Y_train)
```

```
[31]: LogisticRegression()
```

```
[32]: model_predict= model.predict(X_test)
      model_predict
```

```
[32]: array([0, 1, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 1, 0,
             0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0,
             1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 1, 1, 1, 1, 0, 0, 0, 0,
             0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0,
             1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 1, 1, 0, 0, 1, 0, 0, 0,
             0, 0, 1, 0, 1, 0, 0, 1, 1, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0, 1, 0, 1,
             1, 0, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 1, 0, 1, 1, 1, 1, 0, 0, 1,
             1, 0, 0, 0, 1, 0, 1, 0, 1, 1, 1, 0, 1, 0, 0, 1, 0, 1, 0, 1, 1, 0,
             1, 1, 0], dtype=int64)
```

```
[33]: accuracy =accuracy_score(model_predict, Y_test)
      accuracy
```

```
[33]: 0.8156424581005587
```

```
[34]: confusion_matrix(Y_test, model_predict)
```

```
[34]: array([[93, 12],
             [21, 53]], dtype=int64)
```

```
[35]: model_predict
```

```
[35]: array([0, 1, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 1, 0,
             0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0,
             1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 1, 1, 1, 1, 0, 0, 0, 0,
             0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0,
             1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 1, 1, 0, 0, 1, 0, 0, 0,
             0, 0, 1, 0, 1, 0, 0, 1, 1, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0, 1, 0, 1,
             1, 0, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 1, 0, 1, 1, 1, 1, 0, 0, 1,
             1, 0, 0, 0, 1, 0, 1, 0, 1, 1, 1, 0, 1, 0, 0, 1, 0, 1, 0, 1, 1, 0,
             1, 1, 0], dtype=int64)
```

```
[40]: import joblib
```

```
[41]: joblib.dump(model,'model.joblib')
```

```
[41]: ['model.joblib']
```

```
[44]: u = joblib.load('model.joblib')  # Ensure you are loading the actual model file

      u
```

```
[44]: LogisticRegression()
```

```
[45]: u.predict(X_test)
```

```
[45]: array([0, 1, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 1, 0,
             0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0,
             1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 1, 1, 1, 1, 0, 0, 0, 0,
             0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0,
             1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 1, 1, 0, 0, 1, 0, 0, 0,
             0, 0, 1, 0, 1, 0, 0, 1, 1, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0, 1, 0, 1,
             1, 0, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 1, 0, 1, 1, 1, 1, 0, 0, 1,
             1, 0, 0, 0, 1, 0, 1, 0, 1, 1, 1, 0, 1, 0, 0, 1, 0, 1, 0, 1, 1, 0,
             1, 1, 0], dtype=int64)
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```