In [6]: **import** pandas **as** pd import matplotlib.pyplot as plt In [7]: | titanic_train = pd.read_csv('titanic_train.csv') titanic_test = pd.read_csv('titanic_test.csv') titanic_train.head() In [8]: Out[8]: Passengerld Survived Pclass Name Sex Age SibSp Parch Ticket Fare Cabin Embarked 0 0 3 22.0 A/5 21171 7.2500 1 Braund, Mr. Owen Harris male NaN 2 0 PC 17599 71.2833 C85 С 1 Cumings, Mrs. John Bradley (Florence Briggs Th... female 38.0 1 3 3 1 Heikkinen, Miss. Laina female 26.0 0 0 STON/O2. 3101282 7.9250 NaN S 4 1 1 Futrelle, Mrs. Jacques Heath (Lily May Peel) female 35.0 1 113803 53.1000 C123 S 5 0 3 Allen, Mr. William Henry male 35.0 0 373450 8.0500 S 4 0 NaN titanic_train.shape (891, 12)Out[9]: titanic_train['Survived'].value_counts() In [10]: Out[10]: Name: Survived, dtype: int64 In [11]: plt.figure(figsize=(5,5)) plt.bar(list(titanic_train['Survived'].value_counts().keys()),list(titanic_train['Survived'].value_counts()),color=["r","g"]) plt.show() 500 400 300 200 100 -0.25 0.00 0.25 0.50 0.75 1.00 1.25 titanic_train['Pclass'].value_counts() In [12]: 491 Out[12]: 216 184 Name: Pclass, dtype: int64 plt.figure(figsize=(5,5)) plt.bar(list(titanic_train['Pclass'].value_counts().keys()),list(titanic_train['Pclass'].value_counts()),color= ["Yellow","orange","blue"]) plt.show() 500 400 300 200 100 2.0 2.5 3.0 3.5 0.5 1.0 1.5 titanic_train['Sex'].value_counts() male Out[14]: female 314 Name: Sex, dtype: int64 plt.figure(figsize=(5,5)) In [15]: plt.bar(list(titanic_train['Sex'].value_counts().keys()), list(titanic_train['Sex'].value_counts()), color="Green") plt.show() 600 500 400 300 200 100 0 male female import matplotlib.pyplot as plt In [16]: plt.figure(figsize=(5, 7)) plt.hist(titanic_train['Age']) plt.title("Distribution of Age") plt.xlabel("Age") # Corrected method name plt.show() Distribution of Age 175 -150 -125 100 -75 50 25 0 10 20 30 40 50 60 70 80 Age titanic_train['Survived'].isnull() False Out[17]: False 2 False False False 886 False 887 False 888 False 889 False 890 False Name: Survived, Length: 891, dtype: bool In [18]: sum(titanic_train['Survived'].isnull()) Out[18]: titanic_train['Age'].isnull() In [19]: False Out[19]: False 2 False 3 False False 886 False 887 False 888 True 889 False 890 Name: Age, Length: 891, dtype: bool In [20]: sum(titanic_train['Age'].isnull()) Out[20]: **177** In [21]: titanic_train = titanic_train.dropna() In [22]: sum(titanic_train['Survived'].isnull()) Out[22]: sum(titanic_train['Age'].isnull()) In [23]: Out[23]: In [28]: x_train = titanic_train [['Age']] y_train = titanic_train [['Survived']] $\textbf{from} \ \text{sklearn.tree} \ \textbf{import} \ \text{DecisionTreeClassifier}$ In [29]: dtc = DecisionTreeClassifier() In [30]: dtc.fit(x_train, y_train) In [31]: DecisionTreeClassifier() Out[31]: sum(titanic_test['Age'].isnull()) In [32]: Out[32]: titanic_test = titanic_test.dropna() In [33]: In [34]: sum(titanic_test['Age'].isnull()) Out[34]: x_test = titanic_test[['Age']] In [35]: In [36]: y_pred = dtc.predict(x_test) In [37]: y_pred array([1, 0, 1, 1, 1, 1, 0, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 0, 0, 1, Out[37]: 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 1, 1, 0, 1, 0, 0, 1, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 0, 0, 0, 1, 1, 1, 1, 1, 0, 0, 1, 0, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 0, 0, 0, 1], dtype=int64)