```
from google.colab import files
uploaded = files.upload()
     Choose Files adult.csv
       adult.csv(text/csv) - 5326368 bytes, last modified: 3/3/2025 - 100% done
     Saving adult csv to adult csv
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.impute import SimpleImputer
from sklearn.preprocessing import OrdinalEncoder, OneHotEncoder
from sklearn.preprocessing import StandardScaler, MinMaxScaler
from scipy import stats
import pandas as pd
# Replace 'your_file.csv' with the name of the file you just uploaded
df = pd.read_csv('adult.csv')
df.head() # Display the first few rows
₹
                                                                                                                                       hours-
                                              educational-
                                                             marital-
                                                                                                                  capital-
                                                                                                                             capital-
                                                                                                                                                native
                                                                       occupation relationship
                                                                                                   race gender
         age workclass fnlwgt education
                                                                                                                                          per-
                                                        num
                                                               status
                                                                                                                       gain
                                                                                                                                 loss
                                                                                                                                                country
                                                                                                                                          week
                                                               Never-
                                                                          Machine-
                                                                                                                                                 United
      0
          25
                  Private
                         226802
                                         11th
                                                          7
                                                                                        Own-child
                                                                                                   Black
                                                                                                            Male
                                                                                                                          O
                                                                                                                                     n
                                                                                                                                            40
                                                               married
                                                                                                                                                  States
                                                                          op-inspct
                                                              Married-
                                                                           Farming-
                                                                                                                                                 United
          38
                  Private
                           89814
                                     HS-grad
                                                          9
                                                                                         Husband White
                                                                                                                          0
                                                                                                                                     0
                                                                                                                                            50
                                                                                                            Male
                                                                  civ-
                                                                             fishing
                                                                                                                                                  States
                                                               spouse
     4
 Next steps:
              Generate code with df

    View recommended plots

                                                                    New interactive sheet
df.head(10)
₹
                                                                                                                                        hours-
                                                                                                                   capital- capital-
                                              educational-
                                                             marital-
                                                                                                                                                native
                                                                                                         gender
                         fnlwgt education
                                                                       occupation relationship
              workclass
                                                                                                   race
                                                                                                                                          per-
                                                        num
                                                               status
                                                                                                                       gain
                                                                                                                                 loss
                                                                                                                                                country
                                                                                                                                          week
                                                               Never-
                                                                          Machine-
                                                                                                                                                 United
          25
                  Private
                         226802
                                                          7
                                                                                         Own-child
                                                                                                                                     0
                                         11th
                                                                                                   Black
                                                                                                            Male
                                                                                                                          0
                                                                                                                                            40
                                                               married
                                                                          op-inspct
                                                                                                                                                  States
                                                              Married-
                                                                           Farming-
                                                                                                                                                 United
                                     HS-grad
                                                                                         Husband White
          38
                  Private
                           89814
                                                          9
                                                                  civ-
                                                                                                            Male
                                                                                                                          0
                                                                                                                                     0
                                                                                                                                            50
                                                                            fishing
                                                                                                                                                  States
                                                               spouse
                                                              Married-
                                      Assoc-
                                                                         Protective-
                                                                                                                                                 United
      2
          28
               Local-gov 336951
                                                         12
                                                                                         Husband White
                                                                                                            Male
                                                                                                                          0
                                                                                                                                     n
                                                                                                                                            40
                                                                  civ-
                                       acdm
                                                                                                                                                  States
                                                                              serv
                                                               spouse
                                                              Married-
                                      Some-
                                                                          Machine-
                                                                                                                                                 United
          44
                  Private 160323
                                                         10
                                                                  civ-
                                                                                         Husband
                                                                                                   Black
                                                                                                            Male
                                                                                                                       7688
                                                                                                                                     0
                                                                                                                                            40
                                      college
                                                                           op-inspct
                                                                                                                                                  States
                                                               spouse
                                                                                                                                                 United
                                      Some-
                                                               Never-
                                                                                                                                            30
          10
                          103/07
                                                         10
                                                                                         Own-child White Female
 Next steps: ( Generate code with df
                                      View recommended plots
                                                                    New interactive sheet
df.shape
→ (48842, 15)
print(df.describe())
₹
                                          educational-num
                                                             capital-gain \
                                  fnlwgt
                      age
     count 48842.000000 4.884200e+04
                                              48842.000000
                                                             48842,000000
                38.643585
                           1.896641e+05
                                                 10.078089
                                                              1079.067626
     mean
                13.710510
                                                  2.570973
     std
                            1.056040e+05
                                                              7452.019058
                17,000000
                           1.228500e+04
                                                  1.000000
                                                                 0.000000
     min
     25%
                28.000000 1.175505e+05
                                                  9,000000
                                                                 0.000000
                37.000000 1.781445e+05
                                                 10.000000
                                                                 0.000000
```

```
75%
                                             12.000000
                                                            0.000000
               48.000000 2.376420e+05
               90.000000 1.490400e+06
                                             16.000000 99999.000000
     max
            capital-loss hours-per-week
                           48842.000000
     count 48842.000000
               87.502314
                              40.422382
             403.004552
                               12.391444
     std
               0.000000
                               1.000000
     min
                               40.000000
     25%
                0.000000
     50%
                0.000000
                               40.000000
     75%
                0.000000
                              45.000000
             4356.000000
                              99.000000
     max
#Code to Find Missing Values
# Check for missing values in each column
missing_values = df.isnull().sum()
# Display columns with missing values
print(missing_values[missing_values > 0])
→ Series([], dtype: int64)
import numpy as np
# Introduce missing values at specific locations
df.loc[5, 'educational-num'] = np.nan # Set missing value for 'AGE' at row index 5
df.loc[7, 'age'] = np.nan # Set missing value for 'BMI' at row index 10
# Display the first 10 rows to check the changes
print(df.head(10))
df
```

```
₹
                        rrivate
                                                   TTLU
                                                                      1.0
        25.0
                                  220802
                                   89814
        38.0
                        Private
                                                HS-grad
                                                                      9.0
     1
     2
        28.0
                      Local-gov
                                  336951
                                             Assoc-acdm
                                                                     12.0
     3
        44.0
                                  160323
                                           Some-college
                                                                     10.0
                        Private
     4
        18.0
                                  103497
                                                                     10.0
                                          Some-college
     5
        34.0
                        Private
                                  198693
                                                   10th
                                                                      NaN
     6
        29.0
                                  227026
                                                HS-grad
                                                                      9.0
         NaN
              Self-emp-not-inc
                                  104626
                                            Prof-school
                                                                     15.0
     8
        24.0
                        Private
                                  369667
                                          Some-college
                                                                     10.0
     9
        55.0
                        Private
                                  104996
                                                7th-8th
                                                                      4.0
                                                   relationship
            marital-status
                                     occupation
                                                                   race
                                                                          gender
     0
             Never-married
                             Machine-op-inspct
                                                      Own-child
                                                                  Black
                                                                            Male
        Married-civ-spouse
                                Farming-fishing
                                                         Husband
                                                                  White
                                                                            Male
     1
     2
        Married-civ-spouse
                                Protective-serv
                                                         Husband
                                                                  White
                                                                            Male
     3
        Married-civ-spouse
                             Machine-op-inspct
                                                        Husband
                                                                  Black 

                                                                            Male
     4
             Never-married
                                                      Own-child
                                                                  White
                                                                          Female
             Never-married
                                  Other-service
                                                  Not-in-family
                                                                  White
                                                                            Male
     6
             Never-married
                                                      Unmarried
                                                                  Black
                                                                            Male
     7
        Married-civ-spouse
                                 Prof-specialty
                                                        Husband
                                                                  White
                                                                            Male
                                  Other-service
             Never-married
                                                      Unmarried
                                                                  White
                                                                          Female
     9
        Married-civ-spouse
                                   Craft-repair
                                                        Husband
                                                                  White
                                                                            Male
        capital-gain capital-loss
                                      hours-per-week native-country income
     0
                                                                        <=50K
                    0
                                   0
                                                   40
                                                       United-States
                    0
                                   0
                                                       United-States
                                                                        <=50K
     1
                                                   50
     2
                    0
                                   0
                                                   40
                                                       United-States
                                                                         >50K
     3
                 7688
                                   0
                                                   40
                                                       United-States
                                                                         >50K
     4
                    0
                                   0
                                                       United-States
                                                                        <=50K
                                                   30
     5
                    a
                                   a
                                                                        <=50K
                                                   30
                                                       United-States
                                                                        <=50K
     6
                    0
                                   0
                                                   40
                                                       United-States
     7
                 3103
                                   0
                                                       United-States
                                                                        >50K
                                                   32
                                   0
                                                       United-States
     8
                    0
                                                   40
                                                                        <=50K
     9
                    a
                                   0
                                                   10
                                                       United-States
                                                                        <=50K
                                                                                                                                             hours-
                                                                                                                       capital-
                                                                                                                                 capital-
                                                   educational-
                                                                 marital-
                  workclass fnlwgt education
                                                                            occupation relationship
                                                                                                       race gender
                                                                                                                                               per-
                                                                                                                            gain
                                                                    status
                                                                                                                                      loss
                                                            num
                                                                                                                                               week
                                                                    Never-
                                                                               Machine-
        0
             25.0
                      Private 226802
                                                             7.0
                                                                                             Own-child
                                                                                                        Black
                                                                                                                               0
                                                                                                                                          0
                                             11th
                                                                                                                 Male
                                                                                                                                                 40
                                                                    married
                                                                               op-inspct
                                                                   Married-
                                                                               Farming-
             38.0
                                89814
                                                                                              Husband White
                                                                                                                                          0
        1
                      Private
                                          HS-grad
                                                             9.0
                                                                       civ-
                                                                                                                 Male
                                                                                                                               0
                                                                                                                                                 50
                                                                                 fishing
                                                                    spouse
                                                                   Married-
                                                                              Protective-
                                           Assoc-
        2
             28.0
                    Local-gov
                               336951
                                                            12.0
                                                                       civ-
                                                                                              Husband
                                                                                                        White
                                                                                                                 Male
                                                                                                                               0
                                                                                                                                          0
                                                                                                                                                 40
                                            acdm
                                                                                   serv
                                                                    spouse
                                                                   Married-
                                           Some-
                                                                               Machine-
                                                            10.0
        3
             44.0
                       Private
                               160323
                                                                       civ-
                                                                                              Husband
                                                                                                        Black
                                                                                                                 Male
                                                                                                                            7688
                                                                                                                                          0
                                                                                                                                                 40
                                           college
                                                                               op-inspct
                                                                    spouse
                                           Some-
                                                                    Never-
             18.0
                               103497
                                                            10.0
                                                                                      ?
                                                                                             Own-child
                                                                                                        White
                                                                                                               Female
                                                                                                                               0
                                                                                                                                          0
                                                                                                                                                 30
                                           college
                                                                    married
                                                              ...
                                                                                                                                                 ...
                                                                   Married-
                                           Assoc-
                                                                                  Tech-
                                                                                                                                                 38
      48837
              27 N
                      Private
                               257302
                                                            12.0
                                                                                                  Wife
                                                                                                        White Female
                                                                                                                               0
                                                                                                                                          0
              Generate code with df
                                      View recommended plots
                                                                    New interactive sheet
print(df.describe())
₹
                                           educational-num
                                                             capital-gain
                      age
                                  fnlwgt
            48841.000000
                                              48841.000000
     count
                           4.884200e+04
                                                             48842.000000
                                                 10.078172
     mean
               38.643087
                           1.896641e+05
                                                              1079.067626
     std
               13.710207
                           1.056040e+05
                                                  2.570933
                                                              7452.019058
               17.000000
                           1.228500e+04
                                                  1.000000
                                                                 0.000000
               28.000000
                                                  9.000000
                                                                 0.000000
     25%
                           1.175505e+05
     50%
               37.000000
                           1.781445e+05
                                                 10.000000
                                                                 0.000000
     75%
               48.000000
                           2.376420e+05
                                                 12.000000
                                                                 0.000000
               90.000000
                           1.490400e+06
                                                 16.000000
                                                             99999.000000
     max
             capital-loss
                           hours-per-week
     count
            48842.000000
                              48842,000000
               87.502314
                                 40.422382
     mean
     std
               403.004552
                                 12.391444
     min
                 0.000000
                                  1.000000
     25%
                 0.000000
                                 40.000000
     50%
                 0.000000
                                 40,000000
     75%
                 0.000000
                                 45.000000
```

99.000000

4356.000000

```
#Code to Find Missing Values
# Check for missing values in each column
missing_values = df.isnull().sum()
# Display columns with missing values
print(missing_values[missing_values > 0])
→ age
     educational-num
     dtype: int64
#Set the values to some value (zero, the mean, the median, etc.).
# Step 1: Create an instance of SimpleImputer with the median strategy for Age and mean stratergy for Salary
imputer1 = SimpleImputer(strategy="median")
imputer2 = SimpleImputer(strategy="mean")
df_copy=df
# Step 2: Fit the imputer on the "Age" and "Salary"column
# Note: SimpleImputer expects a 2D array, so we reshape the column
imputer1.fit(df_copy[["educational-num"]])
imputer2.fit(df_copy[["age"]])
# Step 3: Transform (fill) the missing values in the "Age" and "Salary"c column
df_copy["educational-num"] = imputer1.transform(df[["educational-num"]])
df_copy["age"] = imputer2.transform(df[["age"]])
# Verify that there are no missing values left
print(df_copy["educational-num"].isnull().sum())
print(df_copy["age"].isnull().sum())
→ 0
import pandas as pd
from sklearn.preprocessing import OrdinalEncoder, OneHotEncoder
# Normalize the Gender column to be consistent (uppercase in this case)
df['gender'] = df['gender'].str.upper() # Convert to uppercase
df['race'] = df['race'].str.upper()
# Initialize OrdinalEncoder for 'Gender' column
ordinal_encoder = OrdinalEncoder(categories=[["FEMALE", "MALE"]]) # Encoding 'F' as 0, 'M' as 1
# Fit and transform the data in the 'Gender' column
df["gender_Encoded"] = ordinal_encoder.fit_transform(df[["gender"]])
# Initialize OneHotEncoder for the 'City' column (if the column exists)
# You should replace "City" with the actual column name in your dataset
onehot_encoder = OneHotEncoder()
# Fit and transform the "City" column (replace 'City' with the actual name of the column)
if 'race' in df.columns:
    encoded_data = onehot_encoder.fit_transform(df[["race"]])
    # Convert the sparse matrix to a dense array
    encoded_array = encoded_data.toarray()
    # Convert to DataFrame for better visualization
    encoded_df = pd.DataFrame(encoded_array, columns=onehot_encoder.get_feature_names_out(["race"]))
    # Concatenate the one-hot encoded columns with the original DataFrame
    df_encoded = pd.concat([df, encoded_df], axis=1)
    # Drop the original 'City' column as it is now encoded
    df_encoded.drop("race", axis=1, inplace=True)
# If there is no 'City' column, proceed with just encoding 'Gender'
    df_encoded = df.copy()
# Drop the original 'Gender' column
df_encoded.drop("gender", axis=1, inplace=True)
# Display the first few rows of the encoded dataframe
```

```
print(df_encoded.head())
df_encoded
```

```
educational-num
₹
         age
               workclass
                           fnlwgt
                                       education
                                                                          marital-status
     0
        25.0
                 Private
                           226802
                                            11th
                                                                7.0
                                                                           Never-married
     1
        38.0
                 Private
                            89814
                                         HS-grad
                                                                9.0
                                                                     Married-civ-spouse
     2
        28.0
                           336951
                                      Assoc-acdm
                                                                     Married-civ-spouse
               Local-gov
                                                               12.0
     3
        44.0
                 Private
                           160323
                                   Some-college
                                                               10.0
                                                                     Married-civ-spouse
     4
        18.0
                       ?
                           103497
                                   Some-college
                                                               10.0
                                                                           Never-married
                occupation relationship capital-gain
                                                          capital-loss
                                                                          hours-per-week
     0
        Machine-op-inspct
                               Own-child
                                                       0
                                                                      0
                                                                                       40
     1
           Farming-fishing
                                 Husband
                                                       0
                                                                      a
                                                                                       50
     2
                                                       0
                                                                      0
                                                                                       40
           Protective-serv
                                 Husband
     3
        Machine-op-inspct
                                                    7688
                                                                      0
                                                                                       40
                                 Husband
                                                                      0
                                                                                       30
     4
                               Own-child
                                                       a
                                gender_Encoded
       native-country income
                                                  race_AMER-INDIAN-ESKIMO
     0
                         <=50K
       United-States
                                            1.0
                                                                       0.0
                         <=50K
        United-States
                                            1.0
                                                                       0.0
        United-States
                          >50K
                                            1.0
                                                                       0.0
        United-States
                          >50K
                                            1.0
                                                                       0.0
     3
     4
        United-States
                         <=50K
                                            0.0
                                                                       0.0
        race_ASIAN-PAC-ISLANDER
                                                race_OTHER
                                                             race_WHITE
                                  race_BLACK
     a
                              0.0
                                           1.0
                                                        0.0
                                                                     0.0
     1
                              0.0
                                           0.0
                                                        0.0
                                                                     1.0
     2
                              0.0
                                           0.0
                                                        0.0
                                                                     1.0
     3
                              0.0
                                           1.0
                                                        0.0
                                                                     0.0
     4
                              0.0
                                           0.0
                                                        0.0
                                                                     1.0
                                                                                                                               hours-
                                                    educational-
                                                                  marital-
                                                                                                         capital-
                                                                                                                    capital-
                                                                                                                                       native-
              age workclass fnlwgt education
                                                                              occupation relationship
                                                                                                                                 per-
                                                                                                                                                 incom
                                                                     status
                                                                                                              gain
                                                                                                                        loss
                                                                                                                                       country
                                                             num
                                                                                                                                 week
                                                                     Never-
                                                                                Machine-
                                                                                                                                         United-
                                                                                                                            0
        0
              25.0
                       Private
                               226802
                                              11th
                                                              7.0
                                                                                               Own-child
                                                                                                                 0
                                                                                                                                   40
                                                                                                                                                  <=501
                                                                                                                                         States
                                                                    married
                                                                                op-inspct
                                                                    Married-
                                                                                Farming-
                                                                                                                                         United-
                                          HS-grad
                                                                                                                 0
                                                                                                                            0
                                                                                                                                                  <=501
        1
              38.0
                       Private
                                89814
                                                              9.0
                                                                        civ-
                                                                                                Husband
                                                                                                                                   50
                                                                                  fishing
                                                                                                                                         States
                                                                     spouse
                                                                    Married-
                                                                               Protective-
                                                                                                                                         United-
                                            Assoc-
                               336951
                                                             12.0
                                                                                                Husband
                                                                                                                 0
                                                                                                                            0
                                                                                                                                   40
                                                                                                                                                   >501
        2
              28.0
                     Local-gov
                                                                        civ-
                                             acdm
                                                                                     serv
                                                                                                                                         States
                                                                     spouse
                                                                    Married-
                                            Some-
                                                                                Machine-
                                                                                                                                         United-
                                                             10.0
                                                                                                                                   40
              44.0
                               160323
                                                                                                Husband
                                                                                                              7688
                                                                                                                            0
                                                                                                                                                   >501
        3
                       Private
                                                                        civ-
                                           college
                                                                                                                                         States
                                                                                op-inspct
                                                                     spouse
                                            Some-
                                                                     Never-
                                                                                                                                         United-
        4
              18.0
                            ?
                               103497
                                                             10.0
                                                                                       ?
                                                                                               Own-child
                                                                                                                 0
                                                                                                                            0
                                                                                                                                   30
                                                                                                                                                  <=501
                                           college
                                                                     married
                                                                                                                                         States
                                                                    Married-
                                                                                                                                         United-
                                            Assoc-
                                                                                    Tech-
      48837
             27.0
                       Private
                               257302
                                                             12.0
                                                                                                   Wife
                                                                                                                 0
                                                                                                                            0
                                                                                                                                   38
                                                                                                                                                  <=501
                                                                        civ-
                                             acdm
                                                                                  support
                                                                                                                                         States
                                                                     spouse
                                                                    Married-
                                                                                Machine-
                                                                                                                                         United-
      48838
             40.0
                       Private
                               154374
                                          HS-grad
                                                              9.0
                                                                        civ-
                                                                                                Husband
                                                                                                                 0
                                                                                                                            0
                                                                                                                                   40
                                                                                                                                                   >501
                                                                                op-inspct
                                                                                                                                         States
                                                                     spouse
                                                                                    Adm-
                                                                                                                                         United-
                                                                                                                            0
      48839
             58.0
                       Private 151910
                                          HS-grad
                                                              9.0
                                                                   Widowed
                                                                                              Unmarried
                                                                                                                 0
                                                                                                                                   40
                                                                                                                                                  <=501
                                                                                                                                         States
                                                                                  clerical
                                                                     Never-
                                                                                    Adm-
                                                                                                                                         United-
      48840
             22.0
                       Private 201490
                                          HS-grad
                                                              9.0
                                                                                               Own-child
                                                                                                                 0
                                                                                                                            0
                                                                                                                                   20
                                                                                                                                                  <=501 ▼
                                                                                                                                         States
              Generate code with df_encoded
                                               View recommended plots
 Next steps:
                                                                              New interactive sheet
import pandas as pd
from sklearn.preprocessing import MinMaxScaler
# Initialize the MinMaxScaler
normalizer = MinMaxScaler()
# Apply the MinMaxScaler to the 'AGE' column only
df_encoded[['hours-per-week']] = normalizer.fit_transform(df_encoded[['hours-per-week']])
# Display the first few rows to verify the transformation
print(df_encoded.head())
df_encoded
```



4	חיפד		; T0:	349/ SOM6	e-correde		10.0	never-marriea						•
0	Mach		cupation re p-inspct	lationshi Own-chil		l-gain capit 0	al-loss 0	hours-per-week 0.397959						
1	Fa	rming	-fishing	Husbar	ıd	0	0	0.500000						
2	Pr	otect	ive-serv	Husbar		0	0	0.397959						
3	Mach	ine-o	p-inspct	Husbar		7688	0	0.397959						
4			}	Own-chil	.d	0	0	0.295918						
0			ntry income ates <=50k	_	Encoded r	race_AMER-IND	DIAN-ESKIM 0.							
1		ed-St			1.0		0.							
2		ed-St			1.0		0.							
3	Unit	ed-St			1.0		0.							
4	Unit	ed-St	ates <=50k	(0.0		0.	0						
	race	_ASIA	N-PAC-ISLAN		_	ace_OTHER ra								
0				0.0	1.0	0.0	0.0							
1 2				0.0 0.0	0.0 0.0	0.0 0.0	1.0 1.0							
3				0.0	1.0	0.0	0.0							
4				0.0	0.0	0.0	1.0							
		age	workclass	fnlwgt	education	educational nu	marita	occupation	relationship	capital- gain	capital- loss	hours- per- week	native- country	inco
	0	25.0	Private	226802	11th	7.	.0 Neve		Own-child	0	0	0.397959	United- States	<=5
	1	38.0	Private	89814	HS-grad	9.	Marrie 0 ci spous	iv- Farming-	Husband	0	0	0.500000	United- States	<= <u></u> 5
	2	28.0	Local-gov	336951	Assoc- acdm	12.	Marrie 0 ci spous	iv- Protective-	Husband	0	0	0.397959	United- States	>5
	3	44.0	Private	160323	Some- college	10.	Marrie 0 ci spous	iv- on-inspct	Husband	7688	0	0.397959	United- States	>5
	4	18.0	?	103497	Some- college	10.	.0 Neve marrie	,	Own-child	0	0	0.295918	United- States	<=5
48	3837	27.0	Private	257302	Assoc- acdm	12.	Marrie 0 ci spous	iv- support	Wife	0	0	0.377551	United- States	<= 5
48	3838	40.0	Private	154374	HS-grad	9.	Marrie 0 ci spous	iv- Machine- op-inspct	Husband	0	0	0.397959	United- States	>5
48	3839	58.0	Private	151910	HS-grad	9.	.0 Widowe	ed Adm- clerical	Unmarried	0	0	0.397959	United- States	<= <u>5</u>
48	3840	22.0	Private	201490	HS-grad	9.	.0 Neve marrie		Own-child	0	0	0.193878	United- States	<= <u></u> 5
lext 48	3841 ⁽	Gene	era Sedfeden pr i inc	th df_enco 287927	ded HS-grad	View recomme	Marrie ended plots Ci spous		ive sheet Wife	15024	0	0.397959	United- States	>5

[#] Standardization (mean=0, variance=1)

[#]Pros: Works well for normally distributed data; suitable for many models.

[#]Cons: Sensitive to outliers.

scaler = StandardScaler()

df_encoded[['age']] = scaler.fit_transform(df_encoded[['age']])

df_encoded.head()

```
₹
                                                                                                                            hours-
                                                  educational-
                                                                 marital-
                                                                                                      capital- capital-
              age workclass fnlwgt education
                                                                          occupation relationship
                                                            num
                                                                   status
                                                                                                          gain
                                                                                                                    loss
                                                                   Never-
                                                                              Machine-
      0 -0.995125
                                                            7.0
                                                                                                             0
                                                                                                                       0 0.397959
                       Private
                              226802
                                             11th
                                                                                           Own-child
                                                                   married
                                                                              op-inspct
                                                                  Married-
                                                                              Farming-
         -0.046907
                                89814
                                                            9.0
                                                                                            Husband
                                                                                                             0
                                                                                                                       0
                                                                                                                          0.500000
                       Private
                                         HS-grad
                                                                      civ-
                                                                                fishing
                                                                   spouse
                                                                  Married-
                                          Assoc-
                                                                            Protective-
      2 -0.776305
                    Local-gov
                              336951
                                                           12.0
                                                                                            Husband
                                                                                                             0
                                                                                                                       0 0.397959
                                                                      civ-
                                            acdm
                                                                                  serv
                                                                   spouse
                                                                  Married-
                                           Some-
                                                                              Machine-
         0.390732
                             160323
                                                           10.0
                                                                                            Husband
                                                                                                          7688
                                                                                                                       0 0.397959
                       Private
                                                                      civ-
                                          college
                                                                              op-inspct
                                                                   spouse
 Next steps: ( Generate code with df_encoded
                                             View recommended plots
                                                                          New interactive sheet
#Removing Outliers
# Outlier Detection and Treatment using IQR
#Pros: Simple and effective for mild outliers.
#Cons: May overly reduce variation if there are many extreme outliers.
df_encoded_copy1=df_encoded
df_encoded_copy2=df_encoded
df_encoded_copy3=df_encoded
Q1 = df_encoded_copy1['fnlwgt'].quantile(0.25)
Q3 = df_encoded_copy1['fnlwgt'].quantile(0.75)
IQR = Q3 - Q1
lower\_bound = Q1 - 1.5 * IQR
upper_bound = Q3 + 1.5 * IQR
df_encoded_copy1['fnlwgt'] = np.where(df_encoded_copy1['fnlwgt'] > upper_bound, upper_bound,
                         np.where(df_encoded_copy1['fnlwgt'] < lower_bound, lower_bound, df_encoded_copy1['fnlwgt']))</pre>
print(df_encoded_copy1.head())
₹
                                fnlwgt
                                            education
                                                       educational-num
                  workclass
             age
     0 -0.995125
                              226802.0
                                                                   7.0
                    Private
                                                11th
     1 -0.046907
                    Private
                               89814.0
                                              HS-grad
                                                                   9.0
                                                                  12.0
     2 -0.776305
                  Local-gov
                              336951.0
                                          Assoc-acdm
     3 0.390732
                              160323.0 Some-college
                                                                  10.0
                    Private
     4 -1.505704
                              103497.0 Some-college
                                                                  10.0
                                    occupation relationship capital-gain
            marital-status
     a
             Never-married
                            Machine-op-inspct
                                                   Own-child
                                                                          a
     1
        Married-civ-spouse
                               Farming-fishing
                                                     Husband
                                                                          0
        Married-civ-spouse
                               Protective-serv
                                                     Husband
                                                                          0
     2
        Married-civ-spouse
                            Machine-op-inspct
                                                     Husband
                                                                       7688
     3
     4
             Never-married
                                                   Own-child
                                                                          0
        capital-loss
                      hours-per-week native-country income
                                                              gender Encoded
     a
                             0.397959
                   a
                                       United-States
                                                      <=50K
                                                                          1.0
                                                       <=50K
     1
                   0
                             0.500000
                                       United-States
                                                                          1.0
                   0
     2
                             0.397959
                                       United-States
                                                        >50K
                                                                          1.0
                   0
                             0.397959
                                                                          1.0
     3
                                       United-States
                                                        >50K
     4
                             0.295918
                                       United-States
                                                       <=50K
                                                                          0.0
                                                                        race_OTHER \
        race_AMER-INDIAN-ESKIMO
                                 race_ASIAN-PAC-ISLANDER race_BLACK
     0
                             0.0
                                                       0.0
                                                                   1.0
                                                                                a a
     1
                             0.0
                                                       0.0
                                                                   0.0
                                                                                0.0
     2
                             0.0
                                                       0.0
                                                                   0.0
                                                                                0.0
     3
                             0.0
                                                       0.0
                                                                   1.0
                                                                                0.0
     4
                             0.0
                                                       0.0
                                                                   0.0
                                                                                0.0
        race WHITE
     a
               9.9
     1
               1.0
     2
               1.0
     3
               0.0
     4
               1.0
#Removing Outliers
# Z-score method
#Pros: Good for normally distributed data.
#Cons: Not suitable for non-normal data; may miss outliers in skewed distributions.
df_encoded_copy2['fnlwgt_zscore'] = stats.zscore(df_encoded_copy2['fnlwgt'])
```

native-

country

United-

States

United-

States

United-

States

United-

States

incom

<=501

<=501

>501

>501

per-

week

df_encoded_copy2['fnlwgt'] = np.where(df_encoded_copy2['fnlwgt_zscore'].abs() > 3, np.nan, df_encoded_copy2['fnlwgt']) # Replace outliers wiprint(df_encoded_copy2.head())

```
education educational-num \
                          fnlwgt
       age
            workclass
0 -0.995125
              Private
                       226802.0
                                         11th
                                                           7.0
1 -0.046907
               Private
                         89814.0
                                                           9.0
                                       HS-grad
2 -0.776305 Local-gov
                       336951.0
                                                          12.0
                                    Assoc-acdm
3 0.390732
              Private 160323.0 Some-college
                                                          10.0
4 -1.505704
                       103497.0 Some-college
                                                          10.0
                             occupation relationship capital-gain \
       marital-status
0
       Never-married Machine-op-inspct
                                           Own-child
                                                                  0
   Married-civ-spouse
                         Farming-fishing
                                              Husband
                                                                  0
1
                                              Husband
                                                                  0
   Married-civ-spouse
                        Protective-serv
3
  Married-civ-spouse Machine-op-inspct
                                              Husband
                                                               7688
                                           Own-child
        Never-married
   capital-loss hours-per-week native-country income
                                                      gender_Encoded
0
              0
                       0.397959 United-States <=50K
              0
1
                       0.500000
                                United-States
                                                <=50K
                                                                 1.0
2
              0
                      0.397959 United-States
                                                >50K
                                                                 1.0
3
              0
                      0.397959
                                United-States
                                                >50K
                                                                 1.0
4
              0
                      0.295918
                                United-States <=50K
                                                                  0.0
   race_AMER-INDIAN-ESKIMO race_ASIAN-PAC-ISLANDER race_BLACK race_OTHER \
0
                      0.0
                                                0.0
                                                           1.0
                                                                        0.0
1
                       0.0
                                                0.0
                                                           0.0
                                                                        0.0
                                                0.0
                                                                        0.0
2
                      0.0
                                                           0.0
3
                      0.0
                                                0.0
                                                           1.0
                                                                        0.0
4
                      0.0
                                                0.0
                                                           0.0
                                                                        0.0
   race_WHITE fnlwgt_zscore
0
          0.0
                    0.419934
1
          1.0
                   -1.017089
                   1.575412
2
          1.0
3
          0.0
                   -0.277440
4
                   -0.873553
          1.0
```

#Removing Outliers

Median replacement for outliers

#Pros: Keeps distribution shape intact, useful when capping isn't feasible.

#Cons: May distort data if outliers represent real phenomena.

df_encoded_copy3['fnlwgt_zscore'] = stats.zscore(df_encoded_copy3['fnlwgt'])