

UCI Adult Income Dataset - Exploratory and Descriptive Analysis

This notebook is focused on the exploratory and descriptive analysis of the cleaned version of the UCI Adult Income Dataset.

```
# Import libraries
import os
import pandas as pd
import numpy as np
import plotly.express as px
```

Define and Create Paths

```
# Get working directory
current_dir = os.getcwd()
# Go one directory up to the root directory
project_root_dir = os.path.dirname(current_dir)
# Define paths to the data folders
data_dir = os.path.join(project_root_dir, 'data')
raw_dir = os.path.join(data_dir, 'raw')
processed_dir = os.path.join(data_dir, 'processed')
# Define paths to results folder
results_dir = os.path.join(project_root_dir, 'results')
# Define paths to docs folder
docs_dir = os.path.join(project_root_dir, 'docs')

# Create directories if they do not exist
os.makedirs(raw_dir, exist_ok = True)
os.makedirs(processed_dir, exist_ok = True)
os.makedirs(results_dir, exist_ok = True)
os.makedirs(docs_dir, exist_ok = True)
```

Read in the data

```
adult_data_filename = os.path.join(processed_dir, "adult_cleaned.csv")
adult_df = pd.read_csv(adult_data_filename)
adult_df.head(10)
```

| | age | workclass | fnlwgt | education_num | marital_status | relationship | race | sex | |
|---|-----|---------------|--------|---------------|-----------------------|---------------|-------|--------|---|
| 0 | 39 | government | 77516 | 13 | single | single | white | male | 2 |
| 1 | 50 | self-employed | 83311 | 13 | married | male spouse | white | male | 0 |
| 2 | 38 | private | 215646 | 9 | divorced or separated | single | white | male | 0 |
| 3 | 53 | private | 234721 | 7 | married | male spouse | black | male | 0 |
| 4 | 28 | private | 338409 | 13 | married | female spouse | black | female | 0 |
| 5 | 37 | private | 284582 | 14 | married | female spouse | white | female | 0 |
| 6 | 49 | private | 160187 | 5 | divorced or separated | single | black | female | 0 |
| 7 | 52 | self-employed | 209642 | 9 | married | male spouse | white | male | 0 |
| 8 | 31 | private | 45781 | 14 | single | single | white | female | 1 |
| 9 | 42 | private | 159449 | 13 | married | male spouse | white | male | 5 |

Check the shape of the dataset and datatypes

```
adult_df.shape
```

```
(32513, 16)
```

```
adult_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 32513 entries, 0 to 32512
Data columns (total 16 columns):
#   Column                Non-Null Count  Dtype
---  -
0   age                   32513 non-null  int64
1   workclass             32513 non-null  object
2   fnlwgt                32513 non-null  int64
3   education_num         32513 non-null  int64
4   marital_status        32513 non-null  object
5   relationship           32513 non-null  object
```

```

6   race                32513 non-null object
7   sex                 32513 non-null object
8   capital_gain        32513 non-null int64
9   capital_loss        32513 non-null int64
10  hours_per_week      32513 non-null int64
11  income              32513 non-null object
12  education_level     32513 non-null object
13  occupation_grouped  32513 non-null object
14  native_region       32513 non-null object
15  age_group           32513 non-null object
dtypes: int64(6), object(10)
memory usage: 4.0+ MB

```

Summary Statistics

Numerical Variables

```
adult_df.describe()
```

| | age | fnlwgt | education_num | capital_gain | capital_loss | hours_per_week |
|-------|--------------|--------------|---------------|--------------|--------------|----------------|
| count | 32513.000000 | 3.251300e+04 | 32513.000000 | 32513.000000 | 32513.000000 | 32513.000000 |
| mean | 38.590256 | 1.897942e+05 | 10.081629 | 1079.239812 | 87.432719 | 40.440962 |
| std | 13.638932 | 1.055788e+05 | 2.572015 | 7390.625650 | 403.243596 | 12.350184 |
| min | 17.000000 | 1.228500e+04 | 1.000000 | 0.000000 | 0.000000 | 1.000000 |
| 25% | 28.000000 | 1.178330e+05 | 9.000000 | 0.000000 | 0.000000 | 40.000000 |
| 50% | 37.000000 | 1.783560e+05 | 10.000000 | 0.000000 | 0.000000 | 40.000000 |
| 75% | 48.000000 | 2.370510e+05 | 12.000000 | 0.000000 | 0.000000 | 45.000000 |
| max | 90.000000 | 1.484705e+06 | 16.000000 | 99999.000000 | 4356.000000 | 99.000000 |

Categorical Variables

```
adult_df.describe(include='object')
```

| | workclass | marital_status | relationship | race | sex | income | education_level | occupation |
|--------|-----------|----------------|--------------|-------|-------|--------|-----------------|------------|
| count | 32513 | 32513 | 32513 | 32513 | 32513 | 32513 | 32513 | 32513 |
| unique | 6 | 4 | 5 | 5 | 2 | 2 | 7 | 5 |

| | workclass | marital_status | relationship | race | sex | income | education_level | occ |
|------|-----------|----------------|--------------|-------|-------|--------|---------------------------|-----|
| top | private | married | male spouse | white | male | <=50k | secondary-school graduate | wh |
| freq | 22650 | 14984 | 13178 | 27771 | 21758 | 24677 | 10484 | 165 |

```
adult_df['workclass'].value_counts(normalize=True)
```

```
workclass
private      0.696644
government   0.133793
self-employed 0.112447
unknown      0.056470
voluntary    0.000431
unemployed   0.000215
Name: proportion, dtype: float64
```

```
adult_df['marital_status'].value_counts(normalize=True)
```

```
marital_status
married      0.460862
single       0.327684
divorced or separated 0.180912
widowed      0.030542
Name: proportion, dtype: float64
```

```
adult_df['relationship'].value_counts(normalize=True)
```

```
relationship
male spouse    0.405315
single         0.360686
child          0.155599
female spouse  0.048227
extended relative 0.030173
Name: proportion, dtype: float64
```

```
adult_df['race'].value_counts(normalize=True)
```

```
race
white      0.854151
```

```
black                                0.096023
asian or pacific islander           0.031926
american indian or eskimo           0.009565
other                                0.008335
Name: proportion, dtype: float64
```

Income Distribution

```
adult_df_income = adult_df.groupby('income').size().reset_index(name='total')
adult_df_income
```

| | income | total |
|---|--------|-------|
| 0 | <=50k | 24677 |
| 1 | >50k | 7836 |

```
fig = px.pie(adult_df_income, names='income', values='total', title='Overall Income Distribution')
fig.update_layout(template="presentation", legend_title=dict(text='Income Level'), paper_bgcolor='white')
fig.show()
fig.write_image(os.path.join(results_dir, 'income_distribution_pie_chart.jpg'))
fig.write_image(os.path.join(results_dir, 'income_distribution_pie_chart.png'))
fig.write_html(os.path.join(results_dir, 'income_distribution_pie_chart.html'))
```

Overall Income Distribution



Income by Age Group

```
adult_df_income_age = adult_df.groupby(['age_group', 'income']).size().reset_index(name='total_by_age')
adult_df_income_age
```

| | age_group | income | total_by_age |
|----|-----------|--------|--------------|
| 0 | 18-25 | <=50k | 5333 |
| 1 | 18-25 | >50k | 114 |
| 2 | 26-35 | <=50k | 6910 |
| 3 | 26-35 | >50k | 1591 |
| 4 | 36-45 | <=50k | 5230 |
| 5 | 36-45 | >50k | 2771 |
| 6 | 46-60 | <=50k | 4479 |
| 7 | 46-60 | >50k | 2809 |
| 8 | 61-75 | <=50k | 1580 |
| 9 | 61-75 | >50k | 511 |
| 10 | 76+ | <=50k | 200 |
| 11 | 76+ | >50k | 40 |
| 12 | <18 | <=50k | 945 |

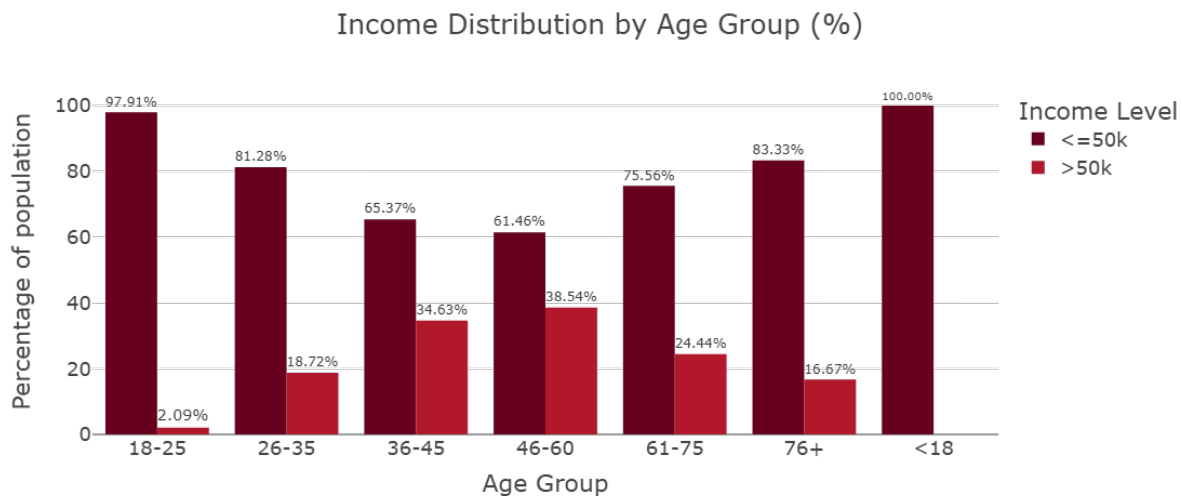
```
total_per_group = adult_df_income_age.groupby('age_group')['total_by_age'].transform('sum')
adult_df_income_age['percentage'] = (adult_df_income_age['total_by_age']/total_per_group) *100
adult_df_income_age
```

| | age_group | income | total_by_age | percentage |
|----|-----------|--------|--------------|------------|
| 0 | 18-25 | <=50k | 5333 | 97.907105 |
| 1 | 18-25 | >50k | 114 | 2.092895 |
| 2 | 26-35 | <=50k | 6910 | 81.284555 |
| 3 | 26-35 | >50k | 1591 | 18.715445 |
| 4 | 36-45 | <=50k | 5230 | 65.366829 |
| 5 | 36-45 | >50k | 2771 | 34.633171 |
| 6 | 46-60 | <=50k | 4479 | 61.457190 |
| 7 | 46-60 | >50k | 2809 | 38.542810 |
| 8 | 61-75 | <=50k | 1580 | 75.561932 |
| 9 | 61-75 | >50k | 511 | 24.438068 |
| 10 | 76+ | <=50k | 200 | 83.333333 |
| 11 | 76+ | >50k | 40 | 16.666667 |
| 12 | <18 | <=50k | 945 | 100.000000 |

```

fig = px.bar(
    adult_df_income_age,
    x = 'age_group',
    y = 'percentage',
    color = 'income',
    title = 'Income Distribution by Age Group (%)',
    barmode = 'group',
    height = 500,
    color_discrete_sequence=px.colors.sequential.RdBu,
    text= 'percentage'
)
fig.update_traces(texttemplate='%{text:.2f}%', textposition='outside')
fig.update_layout(template="presentation", xaxis_title='Age Group',
                    yaxis_title='Percentage of population', legend_title=dict(text='Income Level',
                                                                              color='darkred'),
                    paper_bgcolor = "rgba(0, 0, 0, 0)", plot_bgcolor = "rgba(0, 0, 0, 0)")
fig.show()
fig.write_image(os.path.join(results_dir, 'income_distribution_by_agegroup_bar_plot.jpg'))
fig.write_image(os.path.join(results_dir, 'income_distribution_by_agegroup_bar_plot.png'))
fig.write_html(os.path.join(results_dir, 'income_distribution_by_agegroup_bar_plot.html'))

```



Income by Native Region

```

adult_df_income_reg = adult_df.groupby(['native_region', 'income']).size().reset_index(name='count')
adult_df_income_reg

```

| | native_region | income | total_by_region |
|----|-----------------|--------|-----------------|
| 0 | asia | <=50k | 465 |
| 1 | asia | >50k | 206 |
| 2 | central america | <=50k | 466 |
| 3 | central america | >50k | 58 |
| 4 | europe | <=50k | 369 |
| 5 | europe | >50k | 152 |
| 6 | north america | <=50k | 22768 |
| 7 | north america | >50k | 7250 |
| 8 | other | <=50k | 435 |
| 9 | other | >50k | 146 |
| 10 | south america | <=50k | 174 |
| 11 | south america | >50k | 24 |

```
total_per_native_region = adult_df_income_reg.groupby('native_region')['total_by_region'].tr
adult_df_income_reg['percentage'] = (adult_df_income_reg['total_by_region']/total_per_native
adult_df_income_reg
```

| | native_region | income | total_by_region | percentage |
|----|-----------------|--------|-----------------|------------|
| 0 | asia | <=50k | 465 | 69.299553 |
| 1 | asia | >50k | 206 | 30.700447 |
| 2 | central america | <=50k | 466 | 88.931298 |
| 3 | central america | >50k | 58 | 11.068702 |
| 4 | europe | <=50k | 369 | 70.825336 |
| 5 | europe | >50k | 152 | 29.174664 |
| 6 | north america | <=50k | 22768 | 75.847825 |
| 7 | north america | >50k | 7250 | 24.152175 |
| 8 | other | <=50k | 435 | 74.870912 |
| 9 | other | >50k | 146 | 25.129088 |
| 10 | south america | <=50k | 174 | 87.878788 |
| 11 | south america | >50k | 24 | 12.121212 |

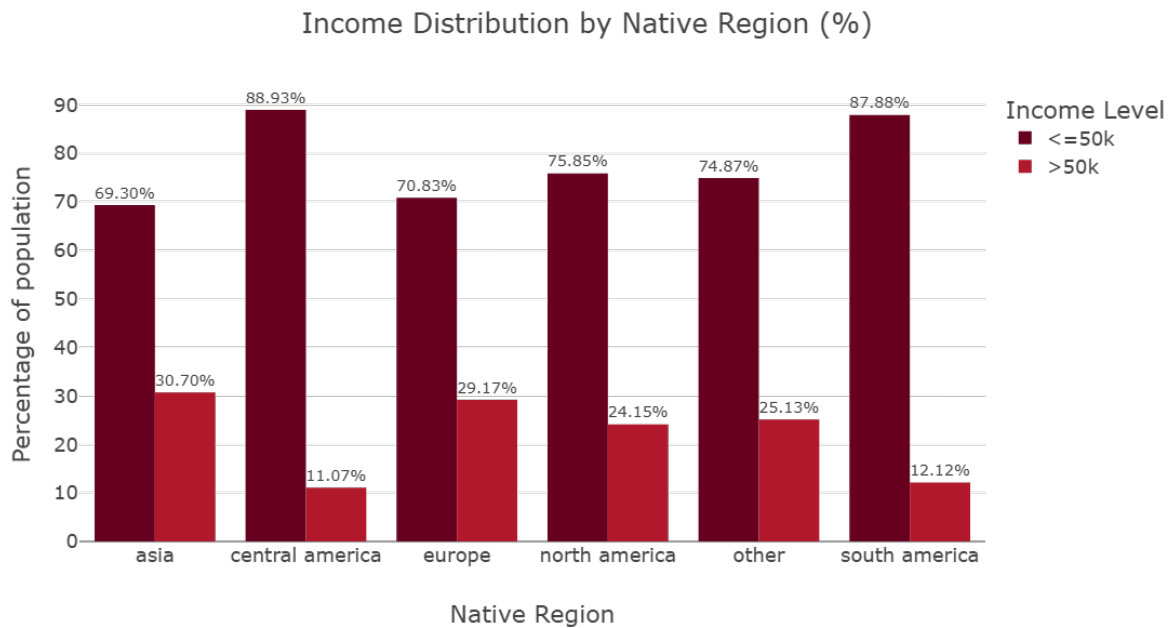
```
fig = px.bar(
    adult_df_income_reg,
    x = 'native_region',
    y = 'percentage',
    color = 'income',
    title = 'Income Distribution by Native Region (%)',
    barmode = 'group',
```



```

height = 600,
width=1000,
color_discrete_sequence=px.colors.sequential.RdBu,
text= 'percentage'
)
fig.update_traces(texttemplate='%{text:.2f}%', textposition='outside')
fig.update_layout(template="presentation", xaxis_title='Native Region', yaxis_title='Percentage of population',
                    xaxis_title_standoff=50, paper_bgcolor = "rgba(0, 0, 0, 0)", plot_bgcolor = "white")
fig.show()
fig.write_image(os.path.join(results_dir, 'income_distribution_by_nativeregion_bar_plot.jpg'))
fig.write_image(os.path.join(results_dir, 'income_distribution_by_nativeregion_bar_plot.png'))
fig.write_html(os.path.join(results_dir, 'income_distribution_by_nativeregion_bar_plot.html'))

```



Income by Race

```

adult_df_income_race = adult_df.groupby(['race', 'income']).size().reset_index(name='total_by_race_income')
adult_df_income_race

```

| | race | income | total_by_race |
|---|---------------------------|--------|---------------|
| 0 | american indian or eskimo | <=50k | 275 |
| 1 | american indian or eskimo | >50k | 36 |
| 2 | asian or pacific islander | <=50k | 762 |
| 3 | asian or pacific islander | >50k | 276 |
| 4 | black | <=50k | 2735 |
| 5 | black | >50k | 387 |
| 6 | other | <=50k | 246 |
| 7 | other | >50k | 25 |
| 8 | white | <=50k | 20659 |
| 9 | white | >50k | 7112 |

```
total_per_race = adult_df_income_race.groupby('race')['total_by_race'].transform('sum')
adult_df_income_race['percentage'] = (adult_df_income_race['total_by_race']/total_per_race)
adult_df_income_race
```

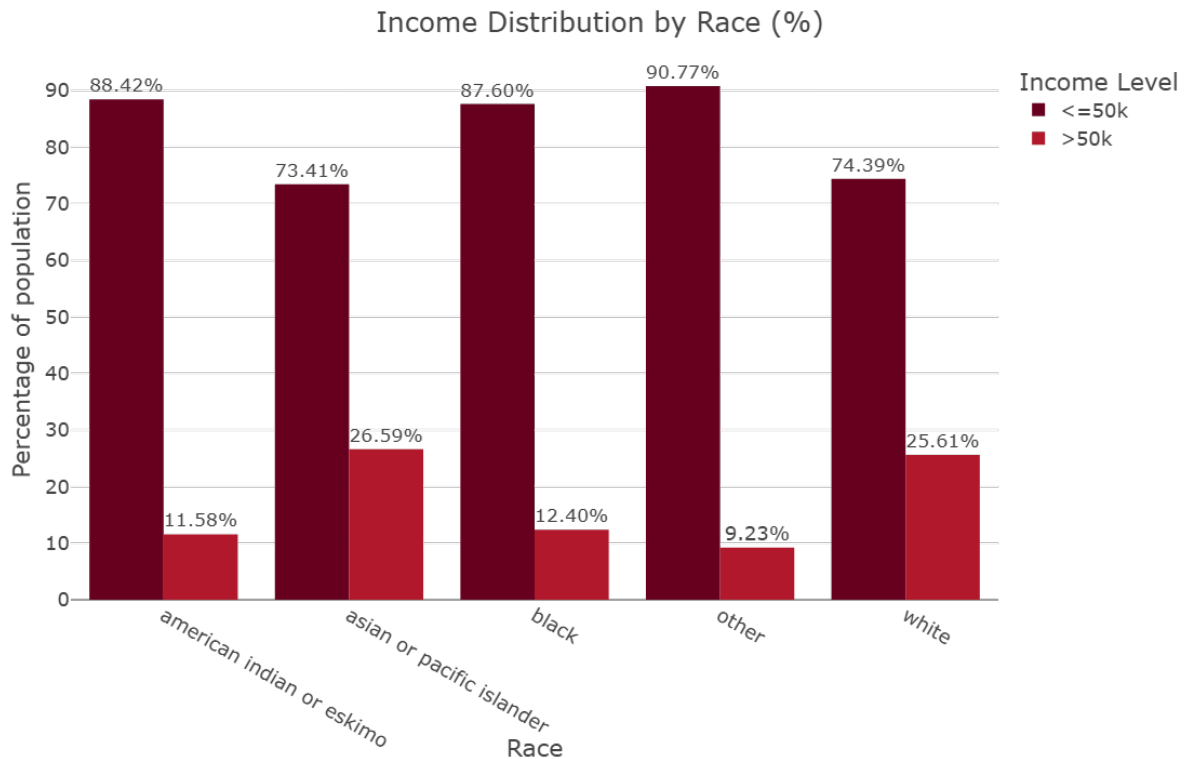
| | race | income | total_by_race | percentage |
|---|---------------------------|--------|---------------|------------|
| 0 | american indian or eskimo | <=50k | 275 | 88.424437 |
| 1 | american indian or eskimo | >50k | 36 | 11.575563 |
| 2 | asian or pacific islander | <=50k | 762 | 73.410405 |
| 3 | asian or pacific islander | >50k | 276 | 26.589595 |
| 4 | black | <=50k | 2735 | 87.604100 |
| 5 | black | >50k | 387 | 12.395900 |
| 6 | other | <=50k | 246 | 90.774908 |
| 7 | other | >50k | 25 | 9.225092 |
| 8 | white | <=50k | 20659 | 74.390551 |
| 9 | white | >50k | 7112 | 25.609449 |

```
fig = px.bar(
    adult_df_income_race,
    x = 'race',
    y = 'percentage',
    color = 'income',
    title = 'Income Distribution by Race (%)',
    barmode = 'group',
    height = 700,
    width=1000,
    color_discrete_sequence=px.colors.sequential.RdBu,
    text= 'percentage'
```

```

)
fig.update_traces(texttemplate='%{text:.2f}%', textposition='outside')
fig.update_layout(template="presentation", xaxis_title='Race', yaxis_title='Percentage of popo
                    xaxis_title_standoff=30, margin=dict(l=60, r=50, t=50, b=150), paper_bgcolor
fig.show()
fig.write_image(os.path.join(results_dir, 'income_distribution_by_race_bar_plot.jpg'))
fig.write_image(os.path.join(results_dir, 'income_distribution_by_race_bar_plot.png'))
fig.write_html(os.path.join(results_dir, 'income_distribution_by_race_bar_plot.html'))

```



Income by Education Level and Occupation Group

```

adult_df_income_edu_occ = (adult_df.groupby(['education_level', 'occupation_grouped', 'income_level']
                                             .size().reset_index(name='total').sort_values('total', ascending = False)
adult_df_income_edu_occ

```

| | education_level | occupation_grouped | income | total |
|----|---------------------------|--------------------|--------|-------|
| 29 | secondary-school graduate | blue collar | <=50k | 3976 |
| 56 | tertiary | white collar | >50k | 3545 |
| 55 | tertiary | white collar | <=50k | 3369 |
| 45 | some college | white collar | <=50k | 3003 |
| 36 | secondary-school graduate | white collar | <=50k | 2900 |
| 38 | some college | blue collar | <=50k | 1503 |
| 32 | secondary-school graduate | service | <=50k | 1444 |
| 20 | secondary | blue collar | <=50k | 1349 |
| 6 | associate | white collar | <=50k | 1015 |
| 41 | some college | service | <=50k | 902 |
| 46 | some college | white collar | >50k | 858 |
| 30 | secondary-school graduate | blue collar | >50k | 796 |
| 37 | secondary-school graduate | white collar | >50k | 731 |
| 23 | secondary | service | <=50k | 663 |
| 12 | primary | blue collar | <=50k | 634 |
| 27 | secondary | white collar | <=50k | 552 |
| 34 | secondary-school graduate | unknown | <=50k | 487 |
| 0 | associate | blue collar | <=50k | 482 |
| 43 | some college | unknown | <=50k | 481 |
| 39 | some college | blue collar | >50k | 397 |
| 7 | associate | white collar | >50k | 397 |
| 47 | tertiary | blue collar | <=50k | 375 |
| 25 | secondary | unknown | <=50k | 307 |
| 14 | primary | service | <=50k | 243 |
| 2 | associate | service | <=50k | 237 |
| 51 | tertiary | service | <=50k | 232 |
| 48 | tertiary | blue collar | >50k | 183 |
| 53 | tertiary | unknown | <=50k | 172 |
| 1 | associate | blue collar | >50k | 166 |
| 21 | secondary | blue collar | >50k | 116 |
| 16 | primary | unknown | <=50k | 111 |
| 33 | secondary-school graduate | service | >50k | 100 |
| 52 | tertiary | service | >50k | 97 |
| 42 | some college | service | >50k | 95 |
| 18 | primary | white collar | <=50k | 93 |
| 4 | associate | unknown | <=50k | 89 |
| 54 | tertiary | unknown | >50k | 82 |
| 28 | secondary | white collar | >50k | 49 |
| 35 | secondary-school graduate | unknown | >50k | 46 |
| 3 | associate | service | >50k | 44 |
| 13 | primary | blue collar | >50k | 40 |

| | education_level | occupation_grouped | income | total |
|----|---------------------------|--------------------|--------|-------|
| 44 | some college | unknown | >50k | 35 |
| 8 | preschool | blue collar | <=50k | 25 |
| 5 | associate | unknown | >50k | 19 |
| 9 | preschool | service | <=50k | 17 |
| 19 | primary | white collar | >50k | 17 |
| 24 | secondary | service | >50k | 12 |
| 10 | preschool | unknown | <=50k | 5 |
| 26 | secondary | unknown | >50k | 5 |
| 17 | primary | unknown | >50k | 4 |
| 31 | secondary-school graduate | military | <=50k | 4 |
| 11 | preschool | white collar | <=50k | 3 |
| 40 | some college | military | <=50k | 2 |
| 49 | tertiary | military | <=50k | 1 |
| 50 | tertiary | military | >50k | 1 |
| 15 | primary | service | >50k | 1 |
| 22 | secondary | military | <=50k | 1 |

```
adult_df_income_edu_occ['edu_occ'] = (adult_df_income_edu_occ['education_level'] + " | "
                                     + adult_df_income_edu_occ['occupation_grouped'])
adult_df_income_edu_occ
```

| | education_level | occupation_grouped | income | total | edu_occ |
|----|---------------------------|--------------------|--------|-------|--|
| 29 | secondary-school graduate | blue collar | <=50k | 3976 | secondary-school graduate blue collar |
| 56 | tertiary | white collar | >50k | 3545 | tertiary white collar |
| 55 | tertiary | white collar | <=50k | 3369 | tertiary white collar |
| 45 | some college | white collar | <=50k | 3003 | some college white collar |
| 36 | secondary-school graduate | white collar | <=50k | 2900 | secondary-school graduate white collar |
| 38 | some college | blue collar | <=50k | 1503 | some college blue collar |
| 32 | secondary-school graduate | service | <=50k | 1444 | secondary-school graduate service |
| 20 | secondary | blue collar | <=50k | 1349 | secondary blue collar |
| 6 | associate | white collar | <=50k | 1015 | associate white collar |
| 41 | some college | service | <=50k | 902 | some college service |
| 46 | some college | white collar | >50k | 858 | some college white collar |
| 30 | secondary-school graduate | blue collar | >50k | 796 | secondary-school graduate blue collar |
| 37 | secondary-school graduate | white collar | >50k | 731 | secondary-school graduate white collar |
| 23 | secondary | service | <=50k | 663 | secondary service |
| 12 | primary | blue collar | <=50k | 634 | primary blue collar |
| 27 | secondary | white collar | <=50k | 552 | secondary white collar |
| 34 | secondary-school graduate | unknown | <=50k | 487 | secondary-school graduate unknown |

| | education_level | occupation_grouped | income | total | edu_occ |
|----|---------------------------|--------------------|--------|-------|--------------------------------------|
| 0 | associate | blue collar | <=50k | 482 | associate blue collar |
| 43 | some college | unknown | <=50k | 481 | some college unknown |
| 39 | some college | blue collar | >50k | 397 | some college blue collar |
| 7 | associate | white collar | >50k | 397 | associate white collar |
| 47 | tertiary | blue collar | <=50k | 375 | tertiary blue collar |
| 25 | secondary | unknown | <=50k | 307 | secondary unknown |
| 14 | primary | service | <=50k | 243 | primary service |
| 2 | associate | service | <=50k | 237 | associate service |
| 51 | tertiary | service | <=50k | 232 | tertiary service |
| 48 | tertiary | blue collar | >50k | 183 | tertiary blue collar |
| 53 | tertiary | unknown | <=50k | 172 | tertiary unknown |
| 1 | associate | blue collar | >50k | 166 | associate blue collar |
| 21 | secondary | blue collar | >50k | 116 | secondary blue collar |
| 16 | primary | unknown | <=50k | 111 | primary unknown |
| 33 | secondary-school graduate | service | >50k | 100 | secondary-school graduate service |
| 52 | tertiary | service | >50k | 97 | tertiary service |
| 42 | some college | service | >50k | 95 | some college service |
| 18 | primary | white collar | <=50k | 93 | primary white collar |
| 4 | associate | unknown | <=50k | 89 | associate unknown |
| 54 | tertiary | unknown | >50k | 82 | tertiary unknown |
| 28 | secondary | white collar | >50k | 49 | secondary white collar |
| 35 | secondary-school graduate | unknown | >50k | 46 | secondary-school graduate unknown |
| 3 | associate | service | >50k | 44 | associate service |
| 13 | primary | blue collar | >50k | 40 | primary blue collar |
| 44 | some college | unknown | >50k | 35 | some college unknown |
| 8 | preschool | blue collar | <=50k | 25 | preschool blue collar |
| 5 | associate | unknown | >50k | 19 | associate unknown |
| 9 | preschool | service | <=50k | 17 | preschool service |
| 19 | primary | white collar | >50k | 17 | primary white collar |
| 24 | secondary | service | >50k | 12 | secondary service |
| 10 | preschool | unknown | <=50k | 5 | preschool unknown |
| 26 | secondary | unknown | >50k | 5 | secondary unknown |
| 17 | primary | unknown | >50k | 4 | primary unknown |
| 31 | secondary-school graduate | military | <=50k | 4 | secondary-school graduate military |
| 11 | preschool | white collar | <=50k | 3 | preschool white collar |
| 40 | some college | military | <=50k | 2 | some college military |
| 49 | tertiary | military | <=50k | 1 | tertiary military |
| 50 | tertiary | military | >50k | 1 | tertiary military |
| 15 | primary | service | >50k | 1 | primary service |
| 22 | secondary | military | <=50k | 1 | secondary military |

```

num = 15
adult_df_combos = adult_df_income_edu_occ.head(num)
fig = px.bar(
    adult_df_combos,
    x = 'total',
    y = 'edu_occ',
    color = 'income',
    orientation = 'h',
    title = f'Top {num} Education and Occupation Groups Combinations by Income Group',
    # barmode = 'group',
    height = 500,
    width=1100,
    color_discrete_sequence=px.colors.sequential.RdBu,
    text = 'total'
)
fig.update_layout(template="presentation",
                    xaxis_title='Number of Individuals',
                    yaxis_title='Education | Occupation Group',
                    legend_title=dict(text='Income Level'),
                    margin=dict(l=450, r=50, t=50, b=50),
                    paper_bgcolor = "rgba(0, 0, 0, 0)",
                    plot_bgcolor = "rgba(0, 0, 0, 0)")
fig.update_traces(textposition='inside')
fig.show()
fig.write_image(os.path.join(results_dir, 'income_distribution_by_eduandocc_bar_plot.jpg'))
fig.write_image(os.path.join(results_dir, 'income_distribution_by_eduandocc_bar_plot.png'))
fig.write_html(os.path.join(results_dir, 'income_distribution_by_eduandocc_bar_plot.html'))

```

Top 15 Education and Occupation Groups Combinations by Income Group

