Demographic Data Analyzer — Project Documentation

Project Summary

This project analyzes a demographic dataset (census data) to uncover insights about race distribution, education levels, working hours, income, and occupation trends. The goal is to use Python and Pandas to answer specific questions about the population's characteristics and earnings.

Approach / Process

1. **Data Loading**
   Loaded the dataset (CSV) into a Pandas DataFrame for easy manipulation.

2. **Data Exploration**
   Used Pandas methods like `value_counts()`, `mean()`, and filtering to explore race, education, working hours, and salary.

3. **Stepwise Analysis**
   Answered a series of questions one at a time:

   - Race count
   - Average age of men
   - Percentage with Bachelor's degree
   - Income distribution among advanced vs non-advanced education groups
   - Minimum work hours and income distribution among minimum hour workers
   - Country with highest income percentage >50K
   - Most popular occupation for high earners in India

4. **Results Interpretation**
   Calculated percentages and identified trends using Pandas aggregation functions.

Code

Python code:

```
import pandas as pd

# Load data
df = pd.read_csv('adult.data.csv')

# Step 1: How many people of each race?
race_count = df["race"].value_counts()

# Step 2: Average age of men
average_age_men = df[df["sex"] == "Male"]["age"].mean()

# Step 3.3: Percentage with Bachelor's degree
percentage_bachelors = (df["education"] == "Bachelors").mean() * 100

# Step 3.4: Percentage with advanced education making >50K
advanced_edu = ["Bachelors", "Masters", "Doctorate"]
higher_edu = df[df["education"].isin(advanced_edu)]
higher_edu_rich = higher_edu[higher_edu["salary"] == ">50K"]
percentage_higher_edu_rich = (len(higher_edu_rich) / len(higher_edu))
* 100

# Step 3.5: Percentage without advanced education making >50K
lower_edu = df[~df["education"].isin(advanced_edu)]
lower_edu_rich = lower_edu[lower_edu["salary"] == ">50K"]
percentage_lower_edu_rich = (len(lower_edu_rich) / len(lower_edu)) *
100

# Step 3.6: Minimum hours worked per week
min_work_hours = df["hours-per-week"].min()

# Step 3.7: Percentage of people working minimum hours earning >50K
```

```
min_workers = df[df["hours-per-week"] == min_work_hours]
rich_min_workers = min_workers[min_workers["salary"] == ">50K"]
percentage_rich_min_workers = (len(rich_min_workers) /
len(min_workers)) * 100

# Step 3.8: Country with highest percentage earning >50K
country_counts = df["native-country"].value_counts()
rich_country_counts = df[df["salary"] ==
">50K"]["native-country"].value_counts()
rich_percentage_by_country = (rich_country_counts / country_counts) *
100
highest_earning_country = rich_percentage_by_country.idxmax()
highest_earning_country_percentage = rich_percentage_by_country.max()

# Step 3.9: Most popular occupation for those earning >50K in India
rich_india = df[(df["native-country"] == "India") & (df["salary"] ==
">50K")]
most_popular_occupation_india = rich_india["occupation"].mode()[0]

# Printing all results
print("Race count:\n", race_count, "\n")
print(f"Average age of men: {average_age_men:.1f}")
print(f"Percentage with Bachelor's degree:
{percentage_bachelors:.1f}%")
print(f"Percentage with advanced education making >50K:
{percentage_higher_edu_rich:.1f}%")
print(f"Percentage without advanced education making >50K:
{percentage_lower_edu_rich:.1f}%")
print(f"Minimum work hours per week: {min_work_hours}")
print(f"Percentage of rich among those who work minimum hours:
{percentage_rich_min_workers:.1f}%")
print(f"Highest earning country: {highest_earning_country} with
{highest_earning_country_percentage:.1f}% earning >50K")
print(f"Most popular occupation for rich in India:
{most_popular_occupation_india}")
```

Findings

Question	Result
Number of people of each race	White: 27816, Black: 3124, Asian-Pac-Islander: 1039, Amer-Indian-Eskimo: 311, Other: 271
Average age of men	39.4 years
Percentage with Bachelor's degree	16.4%
Percentage with advanced education (Bachelors, Masters, Doctorate) earning >50K	46.5%
Percentage without advanced education earning >50K	17.4%
Minimum hours worked per week	1 hour
Percentage of people working minimum hours who earn >50K	10.0%
Country with highest % of >50K earners	Iran (41.9%)
Most popular occupation among >50K earners in India	Prof-specialty

---

Conclusion

- People with advanced education are much more likely to earn >50K than those without.
- Although some work very few hours (minimum 1 hour/week), a small percentage of them still earn high salaries.
- Iran stands out with the highest percentage of high earners relative to its population in this dataset.
- In India, the most common well-paid job is in professional specialties, indicating a skilled workforce among high earners.

This project demonstrates practical use of Pandas for demographic data exploration and income-based analysis.