A study on wine drinking behaviour in correlation with age and household income in the US

The National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) is a large-scale, nationally representative survey conducted in the United States that focuses on the prevalence and correlates of alcohol use and related disorders. Data collection involves face-to-face interviews administered by trained interviewers using a structured questionnaire that covers a wide range of topics, including demographics, alcohol consumption patterns, substance use, and mental health disorders. The survey employs validated instruments to ensure accuracy and confidentiality is assured to encourage honest responses.

The sampling procedure utilizes a multistage stratified design to represent the adult population aged 18 and older. This involves stratifying the U.S. into various geographic and demographic groups, randomly selecting primary sampling units (PSUs) such as counties, and then randomly choosing households within those PSUs. Within each household, one adult is randomly selected to participate, often using methods like the Kish method to ensure randomness. To account for the complex sampling design and ensure representativeness, statistical weights are applied to the data. NESARC has been conducted in multiple waves since its inception in 2001-2002, allowing for longitudinal analysis of trends over time.

To study the wine drinking behaviour in correlation with age and household income in the US the NESARC study dataset is utilized. The data used for the research question is described in the following:

Keys "S1Q12B" for "TOTAL HOUSEHOLD INCOME IN LAST 12 MONTHS: CAT", "AGE" for age, as well as the keys for wine consumption, "S2AQ6B" for how often wine was consumed in last 12 month, as well as "S2AQ6D" for number of wine units drank when drinking in the last 12 month (refer to codebook of the NESARC study dataset).

```
206-207 S1Q12B TOTAL HOUSEHOLD INCOME IN LAST 12 MONTHS: CATEGORY

        1531 1. Less than $5,000
    2212 2. $5,000 to $7,999
    1304 3. $8,000 to $9,999
    2437 4. $10,000 to $12,999
    1288 5. $13,000 to $14,999
    3232 6. $15,000 to $19,999
    3326 7. $20,000 to $24,999
    2961 8. $25,000 to $29,999
    3050 9. $30,000 to $34,999
    2605 10. $35,000 to $39,999
    4407 11. $40,000 to $49,999
    3552 12. $50,000 to $59,999
    2729 13. $60,000 to $69,999
    2084 14. $70,000 to $79,999
    1430 15. $80,000 to $89,999
    1011 16. $90,000 to $99,999
```

```
    1171 17. $100,000 to $109,999
    451 18. $110,000 to $119,999
    939 19. $120,000 to $149,999
    745 20. $150,000 to 199,999
        628 21. $200,000 or more
```

```
68-69 AGE

    43079 18-97. Age in years
    14 98. 98 years or older
```

```
358-359 S2AQ6B HOW OFTEN DRANK WINE IN LAST 12 MONTHS

    465 1. Every day
    314 2. Nearly every day
    643 3. 3 to 4 times a week
    828 4. 2 times a week
    1193 5. Once a week
    1553 6. 2 to 3 times a month
    1819 7. Once a month
    1053 8. 7 to 11 times in the last year
    2780 9. 3 to 6 times in the last year
    3891 10. 1 or 2 times in the last year
    22 99. Unknown
    28532 BL. NA, did not drink or unknown if drank wine in last 12
months
```

```
362-363 S2AQ6D NUMBER OF GLASSES/CONTAINERS OF WINE USUALLY CONSUMED
ON DAYS WHEN DRANK
```

WINE IN LAST 12 MONTHS

```
    14530 1-12. Number of drinks of wine
    31 99. Unknown
    28532 BL. NA, did not drink or unknown if drank wine in last 12
months
```

The database was filtered, based on the mentioned keys. Further, any missing value, as well as any values which were stated as "unknown" were removed from the dataset, since they do not benefit the whole evaluation. Also, the age was categorized, to have a cleaner visualization of the

data available.

To better understand the data structure, a frequency distribution was derived for the variables, with the following results

- The frequency distribution analysis of age shows that the age ranges from 18 to 96. Further, the ages between 20 and 58 are more often represented in the data.
- Considering the distribution for the income, the income between 50,000 and 79,999 is the most common in the available data.
- There are two maximums of the frequency distribution, one for drinking wine 2 to 3 times a month and the second one for 1 or 2 times last year. Further, the amount of wine consumed per occurrence needs to be analyzed:
- The amount of wine drank in each occurrence is continuously decreasing, with a maximum of one glass/ container.