Part 5: Communication and reflection

● **QUESTIONS:**

What are the questions you wanted to explore? Why are they interesting to you?

As the call for gender equality continues to gain momentum, there is a growing emphasis on empowering women through education, enabling them to achieve self-sustenance and dignity independently of spousal support.
In light of this, our investigation aims to delve into the influence of education on the economic well-being of both women and men.

According to the above, our question is:
Is there a difference in the effect of education between men and women on their contribution to household income?

By exploring the impact of education on livelihoods, we seek to comprehend how educational attainment contributes to women's self-reliance and dignity, as well as its analogous effects on men.
This analysis strives to shed light on the evolving dynamics of gender roles and economic empowerment, thereby contributing to a deeper understanding of the transformative potential of education in promoting equitable and dignified livelihoods for all.

Furthermore, as citizens of Israel, a country that focuses on protecting its residents, the question arose for us of how to reach families who need help from the state without them having to harm their dignity and turn to the authorities, so it is important for us to create a way that the government will find people who need help from the Ministry of Welfare, even if they didn't ask for it.

By creating a smart and fast system, we want to make sure essential welfare services are easy to access, demonstrating our strong commitment to helping everyone in our company get the help they need.

● **DATASET:**

Describe the dataset you use; Explain why it is appropriate for answering these questions.

Our chosen dataset, consisting of 753 observations representing married individuals in the United States at 1975, provides comprehensive insights into the complex interplay of work and household dynamics. This rich compilation uniquely positions us to uncover the intricate factors influencing gender roles, economic well-being, and family patterns.

This dataset allows us to examine the relationship between education and economic well-being for both women and men.

By analyzing variables such as education levels, income, and employment status, we can assess how educational attainment impacts the financial outcomes of married individuals.
This data can provide quantitative evidence of whether higher education is associated with higher income levels, financial stability, and overall economic well-being for both genders.

● **ANALYSIS & FINDINGS:**

What analyses did you conduct to answer your questions?
What did you find? (support with plots, but no code here).
This part should summarize everything you've done in parts 2-4.
A person reading this should be able to understand the questions you asked, the analysis you've done, and the results, without looking at the Jupiter notebook.

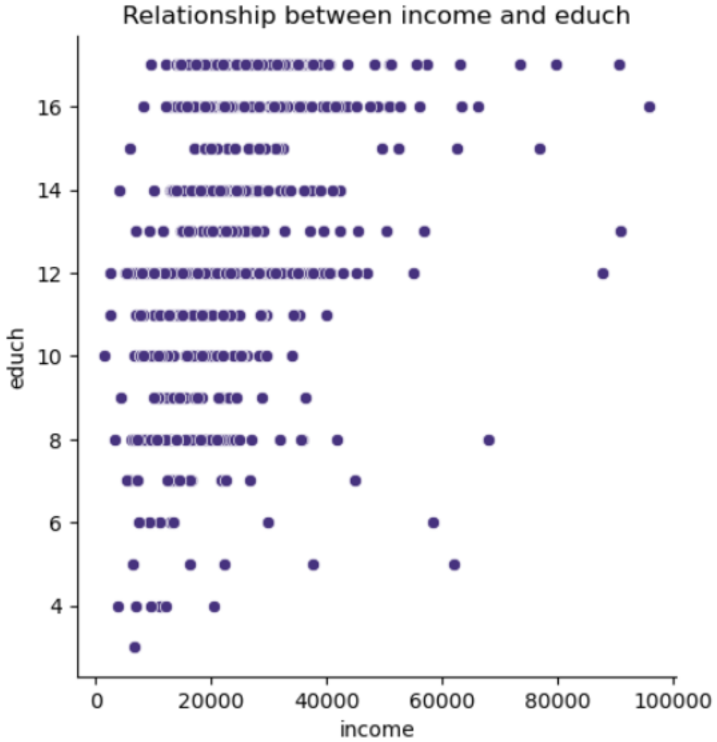
Part 2:

In the beginning, specifically in part 2, our focus aimed to identify patterns that could potentially drive our investigations in parts 3 and 4.
During this initial exploration, we noticed that the way certain variables like 'educh', 'educw', and 'income' distributed could be important for what we were trying to do later.

We then went a step further and investigated how some of these variables are relates to each other.
For example, we looked at how the husband's education might be linked to the household income, and we did the

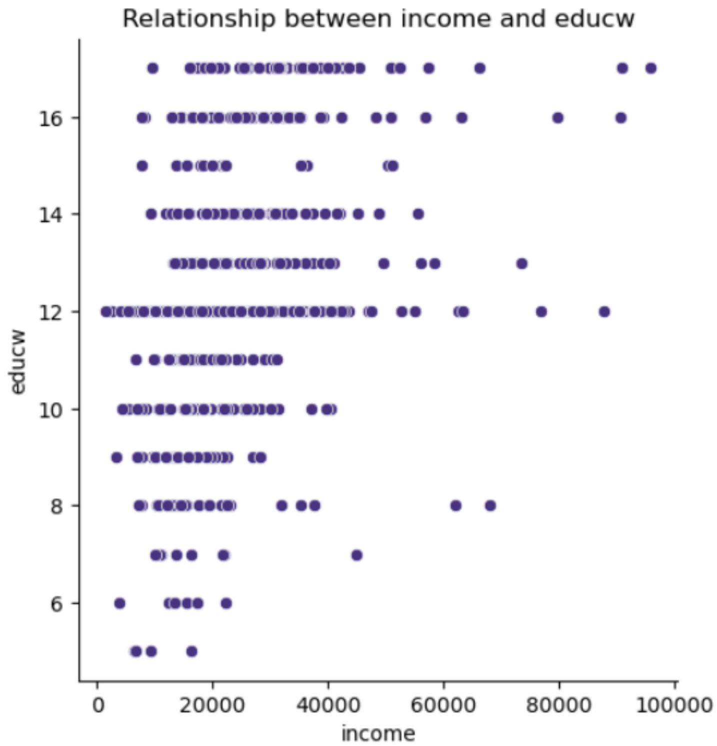
same thing for his wife's education.
We also checked if there were other connections between different variables that could help us with the decisions we made in parts 3 and 4.

The relationship between the household income to the husband's education:




```
0.3768718448508052
As we can see, there is positive correlation between income and educh.
```

The relationship between the household income to wife's education:




```
0.3612749505559233
As we can see, there is positive correlation between income and educw.
```

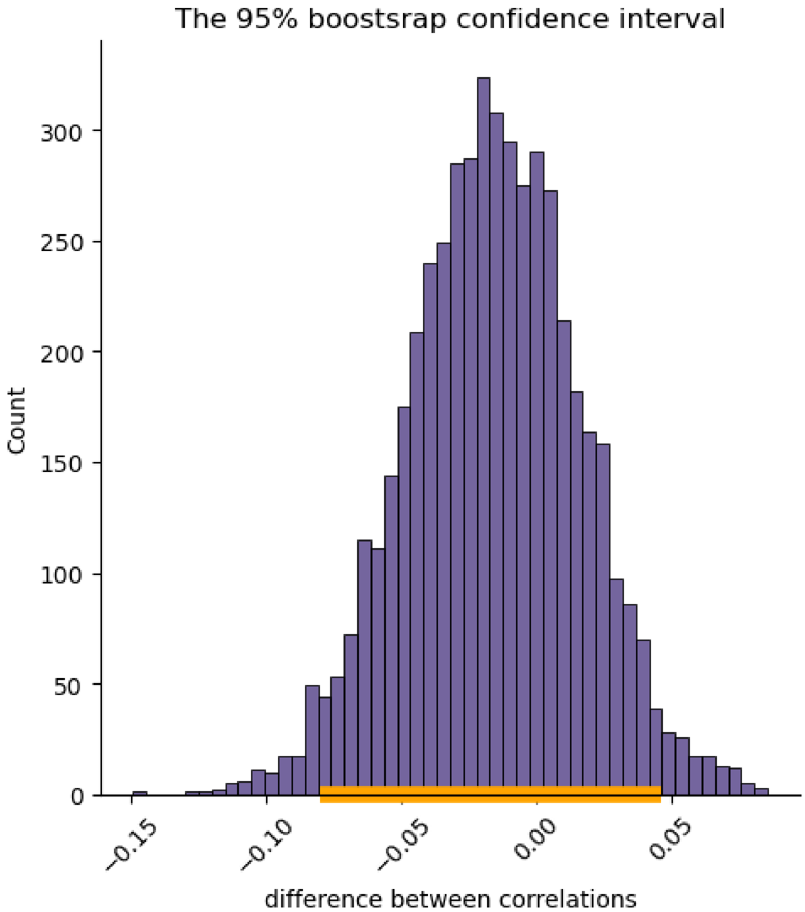
Part 3:

Our objective was to comprehend how the education levels of spouses (both wives and husbands) relate to their household income. This inquiry offers valuable insights into gender dynamics and socioeconomic influences, helping us assess gender equality and empowerment within households and whether women's education translates into increased economic prospects.

We considered reasons for potential differences in correlations between education and income. Elements like family planning, parental leave, childcare, and household dynamics can affect income differently for men and women. Gender-based job market bias also impacts income, even with comparable education levels.

Analyzing 1975 US data, we found no significant difference in how education influenced household income between spouses. This result aligns with the period's context, marked by evolving gender roles, increased female workforce participation and education, and shifting household responsibilities. This suggests societal changes were promoting balance in education's income impact.

The graph we got according to the bootstrap process we have made:



According to the graph above, Our confidence interval analysis indicated that the null hypothesis—stating equal correlations—cannot be rejected. This underscores the consistency in education-income correlations between spouses. This aligns with the era's shifting norms and dynamics—a time of changing gender roles, rising female workforce presence, and evolving household responsibilities. Educational attainment's rising importance for both genders helped level the income impact.

In summary, our investigation illuminates the connections between education, gender dynamics, and income, reflecting a period of transformative societal shifts.

Part 4

In part 4, we decided to construct a classifier aimed at identifying households that may require assistance from the state.
Using domain knowledge, we defined families with multiple children and below-average earnings as potential candidates for assistance.
This classifier was chosen to leverage the state's capability to locate households in need and efficiently direct resources to those who need it most.
By targeting assistance to these households, we aimed to optimize the impact of welfare support and ensure it reaches those most in need.

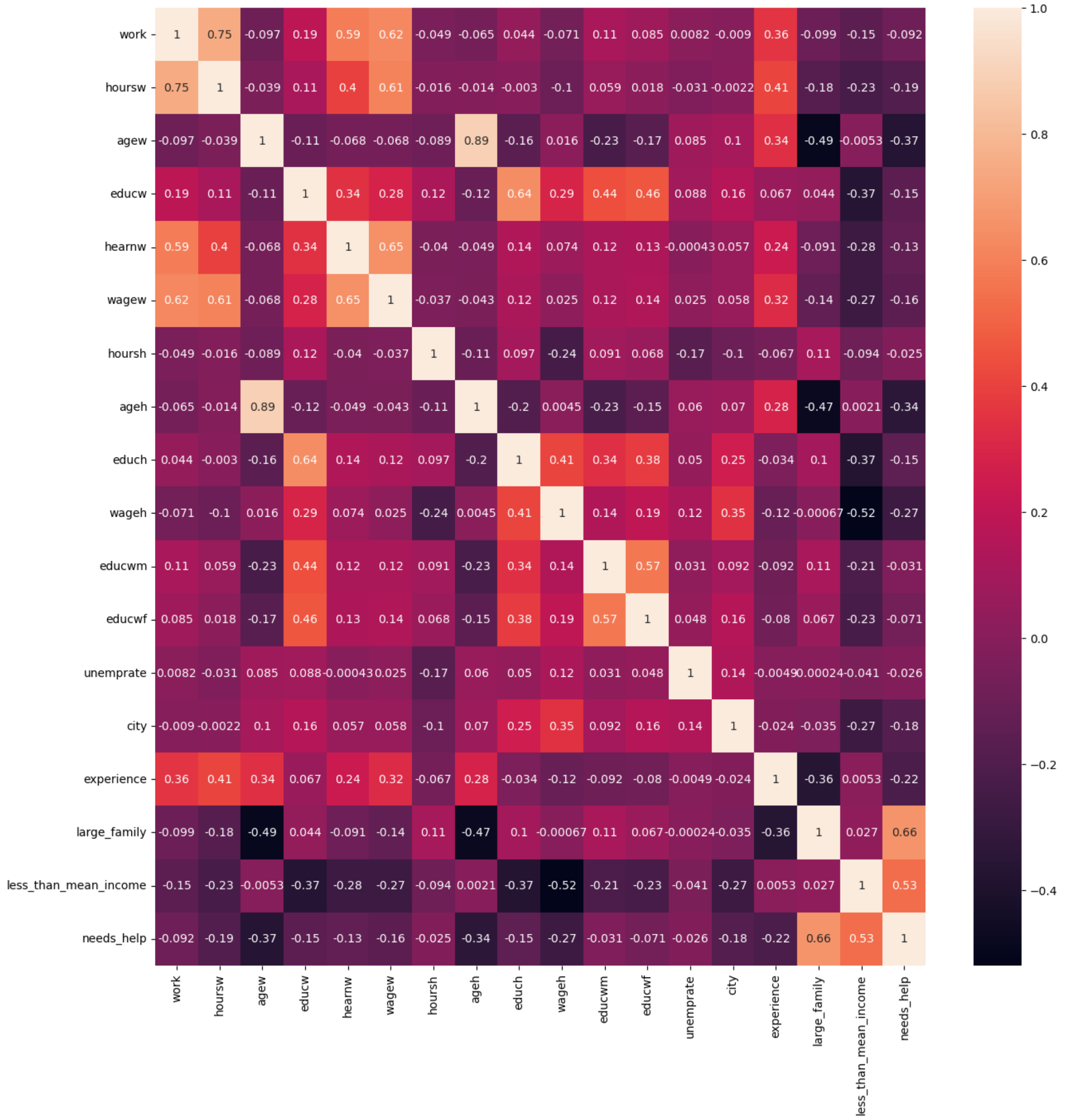
1. Pre-Processing Stage:

Conversion of Categorical Features: In this stage, we transformed the categorical features ('work' and 'city') into a numerical scale. This conversion allows the algorithms to work with these variables effectively, as they typically require numerical inputs.

Creation of the Target Variable: We established a target variable that indicates households potentially needing assistance from the welfare department. This target variable serves as the basis for our classification task.

2. Feature Selection Stage:

Heat-Map Correlation Analysis: In this step, we generated a heat-map of correlations between every pair of variables in the dataset. A heat-map visually represents the strength and direction of correlations using colors.



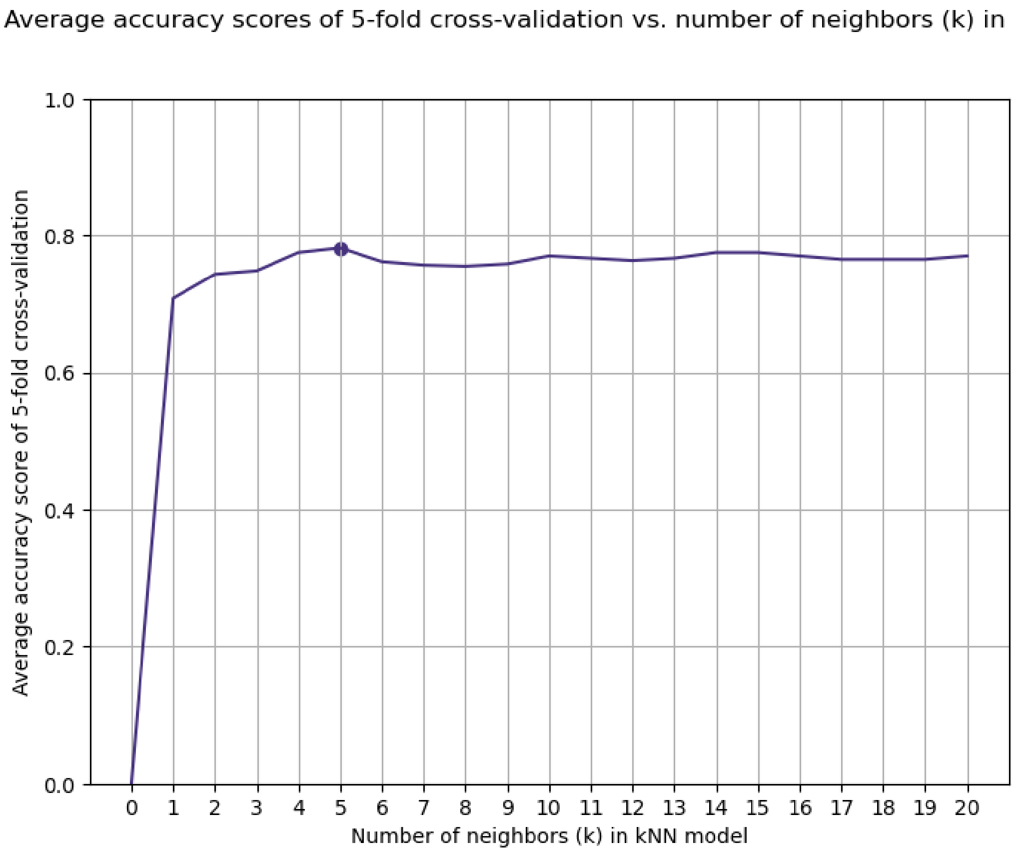
This analysis helped us understand how variables relate to each other, in addition to the graphs from part 2.

Selecting High Correlation Variables: Building on the heat-map results, we identified variables with significant correlations to the target variable. These variables were chosen based on their individual correlation with the target variable, ensuring that they provide valuable information for the classification task.

<u>3. Building the Intended Classifier (using KNN Algorithm):</u>

Scaling Features: Before applying the K-Nearest Neighbors (KNN) algorithm, we standardized the features to the same scale. Scaling is important for KNN, as it's a distance-based algorithm that requires uniform scales to effectively measure distances between data points.

Cross-Validation for Evaluation: To assess the performance of our classifier and ensure its generalizability, we emplo
yed cross-validation. Cross-validation involves dividing our dataset into subsets (folds) for training and testing the mo
del. This approach helps prevent overfitting and provides a more accurate estimation of the classifier's performance.
We chose the "accuracy" score as our evaluation metric to measure how well our classifier performs.
This choice ensures that we focus on correctly identifying households in need of assistance from the welfare depart
ment. By doing this, we aim to improve the classifier's overall accuracy in correctly classifying households that requir
e support.



Best average accuracy score is: 0.782, for number of neighbors (k) = 5

Computation of Classifier Accuracy: Following the training of the KNN classifier and parameter tuning, we assessed i
ts accuracy on an independent test set. This step allowed us to measure the classifier's performance in correctly
predicting households that might require assistance.

Test accuracy score is: 0.762

In summary, our data preprocessing involved converting categorical features and defining the target variable.
The feature selection stage utilized correlation analysis to identify relevant variables.

The construction of the KNN classifier involved feature scaling, cross-validation, and the calculation of accuracy score to assess its predictive capability.

These stages collectively form a comprehensive process to create and assess your intended classifier for identifying households that could require welfare assistance.

● **LIMITATIONS:**
What are some limitations of your analyses and potential biases of the data you used?
How might these biases affect your findings?

Temporal Bias: The data is from 1975, and societal, economic, and gender dynamics have evolved since then. Findings might not accurately reflect current gender roles, workforce dynamics, and household structures, impacting the relevance of conclusions to the present day.

Changing Definitions and Measures: Over time, definitions of terms like education, income, or family dynamics might have evolved or varied across different regions. This could impact the consistency and comparability of data, potentially affecting the reliability of findings.

Cultural and Contextual Bias: Societal norms, cultural values, and economic conditions of 1975 might not fully align with present-day perspectives. This bias could influence the interpretation of results and their applicability to current social contexts.

Sample Representation Bias: The dataset might not fully represent all demographics and socioeconomic groups within the United States. This bias could affect the generalizability of our findings to the entire population, particularly if certain groups are underrepresented or overrepresented.

These biases and limitations could collectively impact the robustness and relevance of our findings.
It's important to acknowledge and account for these potential sources of bias when interpreting and generalizing our results. While our dataset provides valuable insights into historical trends and dynamics, recognizing its limitations ensures a balanced and cautious interpretation of our analyses.

● **FUTURE DIRECTIONS:**
What new questions came up following your exploration of this data?
Describe at least one question that could not be answered using your data alone, and specify what additional data you would collect to address it.

Following the exploration of our data, some new questions that we faced are:

1. Long-Term Trends and Changes:
How have the relationships between education, gender dynamics, and household income evolved over subsequent decades beyond 1975?

2. Cultural and Societal Influences:
How do cultural norms and societal values impact the correlation between education and income for different genders in various regions or countries?

3. Intersectionality and Diversity:
How do other factors like race, ethnicity, and socioeconomic background intersect with gender and education in influencing income levels and opportunities?

4. Effects of war on the household:
In times of war, do women take a greater part in the labor market compared to times when there is no war?

A question that might be challenging to answer using our current dataset alone is question number 4:
In our small research for Exercise 2, Question 10, Section B, we tried to explain an unexpected graph from our data. We realized that the Vietnam War happened in the USA in 1975, the same year as our data. This made us wonder if the war affected our data when we were working with it.

But we faced a problem.
Our data only covered 1975, so we couldn't do a proper test to check if the war influenced the data. To answer this question better in the future, we need data from different years, including ones without wars, for comparison. This way, we can see if war or other events actually impact the patterns we observe.