Money, Sleep, and Love:

The Effects of Income and Sleeping Arrangements on Relationships

Rabia Gondur, Kevin Fernandez, and Nicholas Wai

Department of Computer and Information Science, Fordham University

CISC-5450: Mathematics for Data Science

Prof. Ruhul Amin

May 3rd, 2022

PROJECT SUMMARY:

Relationships play an important role in human social experience, as a result, there are many studies that investigate the effects of a wide range of factors on relationships. For our project, we decided to analyze the effects of two factors, income, and sleeping arrangements, on relationships. Using the *Divorce Rates* dataset, we sought to find a correlation between household income and divorce rates in order to see just how much this external factor may affect the progression of a relationship. Based on common sense, we assumed that income plays a major role in relationships and we predicted that people in higher-income brackets would experience more success in their relationships. We also investigated the problem of whether sleeping in the same bed together helps or hinders relationship length using the *Sleeping Alone* dataset. According to previous research on sleeping patterns and relationship quality, many studies suggest that participants feel more secure and satisfied when they are sleeping with their partner versus sleeping alone as well as display improved sleep quality in polysomnography (Troxel et al., 2007; Drews et al., 2020). Thus, we also predicted that sleeping in separate beds would have a negative impact on relationships. Connecting our two predictions, we hypothesized that people who are more likely to prefer sleeping together in higher-income households are also more likely to stay together and have lower divorce rates. In order to study this, we explored how independent variables related to sleeping arrangements and income levels affect dependent variables like relationship length and divorce rates. The results from this analysis can provide an insight into the connection between sleep, income, and relationships, as well as provide information on how different sleeping patterns and income levels contribute to longer or shorter relationships, potentially offering directions for the development of successful relationships. After analyzing the data with various plots, we tested our hypothesis by training decision trees, and by examining the relationships between our variables through different plots. Finally, we used the *FINC-01* dataset to draw a picture of the current income and relationship trends.

WORKS DONE BY OTHERS:

The *Sleeping Alone* dataset was used by the website *FiveThirtyEight* in the article titled: “Dear Mona, How Many Couples Sleep in Separate Beds?”. In this article, the author sought to answer the question of how common it is for couples to sleep in separate beds using the *Sleeping Alone* dataset. *FiveThirtyEight* also used the *Divorce Rates* dataset to write an article, titled “Marriage Isn’t Dead - Yet”, in which the author tried to better understand the trend that fewer Americans are getting married by analyzing different groups of the population and their divorce and marriage rates separately. *FINC-01: Selected Characteristics of Families by Total Money Income* is a dataset created by the government to categorize data from the U.S. Census in order to help in deciding how to distribute funds and assistance to states and localities.

DATA SOURCE DETAILS:

Our main dataset is the *Sleeping Alone* dataset, and it consists of the responses collected from a survey in 2014 done by the website *FiveThirtyEight*. Before we analyzed the data, we had to pre-process it due to formatting issues. We also had to change each text/object value from every column and give it a corresponding numerical integer value with the idea that we are working with ordinal and categorical data. We checked if there were any outliers before finalizing our data frame. We then concluded that our dependent variables would be List 1 and our independent variables would be List 2. Our second dataset is called *Divorce Rates* and it was compiled by *FiveThirtyEight*. The source for all data is the Decennial Census (years 1960 to 2000) and the American Community Survey (years 2001-2012), via IPUMS USA. Values in the divorce dataset are the share of the relevant population that is currently divorced, conditional on

having ever been married. This dataset was brought in for the time series plot and rate of change plot tasks, as the *Sleeping Alone* dataset did not have the appropriate time columns for those. Our third dataset comes from the United States Census Bureau website, and it is called *FINC-01: Selected Characteristics of Families by Total Money Income*. The data contains the population of families with income salaries from $0 to $200,000. The salary brackets are distributed between different categories like age, education status, and the number of members in a family. We observed three years' worth of data (2018-2020), and each year was its own dataset, so we had to concatenate all three datasets to compare them side by side.

RESULTS AND INTERPRETATION:

Task 1 - Time Series Plots of Y for X:

Diagram A shows that although divorce rates have gone up for all groups since the 1960s, the group with the highest divorce rate has always been those in lower-income brackets, and the group with the lowest divorce rate has always been those in high-income brackets. In short, wealth is a strong predictor of whether or not a married couple will divorce, with rich couples having a divorce rate under 10% in 2010 and poor couples having almost four times the likelihood of being divorced.

Task 2 - Cumulative plots of Y for X:

In Diagram B, we can see that the cumulative frequency of the 16-20 years' response is just under 0.5. This means around half of survey respondents stated that the length of their current or most recent relationship was 16-20 years. This does not help us make any conclusions about our hypothesis, but it helps inform our conclusion by revealing the nature of our dataset. It tells us about who the survey respondents are, allowing us to account for potential shortcomings of the data. For example, as seen in Diagram C, 60% of respondents make over $100,000, this is important as this may mean our survey is not an accurate sample of the population.

Task 3 - The Rate of Change of Y for X:

Diagrams D and E show the rates of change of divorce for three groups of different income levels in two age ranges. Both plots tell a similar story: the rate of change increased for all groups until it dropped off in the 1980s-1990s, then it stabilized from 2000 onwards. The exception is that the rich demographic experienced a spike in the mid-2000s. This could have been caused by the 2008 recession, as that could have affected some assets of rich couples thus potentially causing an increase in divorce rates.

Task 4 - Scatter Plot of Y for X:

The scatter plot (Diagram F) shows the average relationship length for the different groups of people who rated the extent to which they agree with the statement that sleeping in separate beds helped them stay together. The answers ranged from 1 (strongly disagree) to 5 (strongly agree). At first, the graph appears to indicate some significant variance between the average relationship lengths. However, looking at the Y-axis, we can see that there is not even a whole category difference between the lowest average (~4.2) and the highest (~4.9). This indicates that the level of agreement with that statement has no significant impact on relationship length.

Task 5 - Bar chart for Different Xs, and Properties Y:

This bar chart (Diagram G) analyzes the sleeping arrangements of participants and how that may impact their relationship length. Each bar in this graph represents an answer to the question of how often the participants and their partners sleep in separate beds. The spectrum goes from never (1) to every night (6). The height of the bar represents on average how long each of those groups have been in their current relationships based on the categories that range

from less than a year (1) to more than 20 years (6). The bars show very little change in relationship length depending on sleeping arrangements, which suggests that there is no significant correlation between these two variables.

Task 6 - Box plots of Xs for their Properties Y:

Our box plot (Diagram H) uses the same metrics as the bar chart above, however, it reveals more information about sleep patterns and relationship length. The box plot shows that the median for relationship length goes up as the frequency of sleeping apart increases. This could suggest that couples that sleep apart more frequently are more likely to stay together for longer. However, there are many other factors that could affect the relationship length at higher sleep apart frequencies. For instance, couples that are together for long periods are also more likely to be older in age, which could explain the higher frequency of sleeping apart.

Task 7 - Draw the Histogram for all Y for all X:

Due to the categorical nature of our *Sleeping Alone* dataset, our histogram was divided into separate graphs, each corresponding to one category of an independent variable, with the columns representing the categories of the dependent variables. In the case of Diagram I, the independent variable is the frequency of sleeping in separate beds, while the dependent variable is relationship length. This histogram shows us that the highest relationship length bracket dominates all categories of the independent variable. This further supports the idea that sleeping arrangements do not affect relationship length.

Task 8 - Heatmap of Correlations for Y with other Attributes:

The heatmap (Diagram J) displays that the statements “separate beds help us to stay together” and “we sleep better when we sleep in separate beds” have the highest positive correlation of 0.88 between all the other variable correlations. Diagram K is a heatmap that was constructed through further analysis of the previous heatmap. We originally intended to use a sleeping pattern question that has a correlation with income and education. However, there was too little or no correlation between those specific variables. In Diagram K, we observed that X1 and Y1 have the highest correlation of 0.52 out of the other variables. This information tells us that there is a moderate positive correlation between sleeping together and better quality of sleep.

DISCUSSION AND COMPARISONS:

Sleep, money, and relationships play an integral role in people’s lives. Although previous research has studied these variables independently from each other, to date no research looked into the link between these three variables. Therefore, our project used various data analysis tasks and algorithms to investigate the relationship between these variables and to help fill in the gap in this area of research. Based on the literature review of neurocognitive and psychosocial research, we hypothesized that people who are more likely to prefer sleeping together in higher-income households are also more likely to stay together and have lower divorce rates. To begin testing our hypothesis we completed the required tasks for the project in order to visualize the relationships between our variables. Our various plots showed that although sleeping arrangements did not play a significant role in relationship status and length, income did. To formally test our hypothesis and to check the validity of our preliminary results we used a classification algorithm: a decision tree.

A decision tree is a powerful tool that is often used to estimate the outcomes of different scenarios. Since it allows the usage of mixed-type data, this tool stands out from other algorithms as it enables us to test our hypothesis with our categorical data. In the flowchart-like structure of the decision tree, each branch represents an outcome of the test, and each leaf node holds a class label. We classified our class labels as; being divorced or non-divorced. However, since the class

distribution between the two labels is predominantly one-sided, we used the Synthetic Minority Oversampling Technique (SMOTE) to have an equal distribution of our class labels by increasing the frequency of the minority class, which is being divorced. Then, we created four decision trees: a full tree (Diagram L), a tree with a max branch level of four and five minimum nodes (Diagram M), a chi-squared tree (Diagram N), and a chi-squared tree with two fewer columns (Diagram O). Table 1 displays each tree's performance score.

The full decision tree has the highest performance with 98.90%. Thus, we used this model to predict our class label and to determine the precision score of the model, which was 0% for divorced and 99% for non-divorced. We also used other metrics to further our analysis of our hypothesis; the accuracy score for the tree is 98.90%, the recall score for the non-divorce class is one and the divorced class is zero, and the confusion matrix's array is ([[109,0],[1,0]]). These metrics demonstrate that our full decision tree model is very accurate to determine the groups of people who are non-divorced. In addition, the true positive section of the confusion matrix has the highest value, which means that we can accept our hypothesis. Overall, our decision tree accepts and supports our hypothesis that people who are more likely to prefer sleeping together in higher-income households are also more likely to stay together and have lower divorce rates.

Although the *Sleeping Alone* dataset from 2014 provided crucial information on how external factors like income and sleeping can impact our relationships, it is outdated. Therefore, we added a more recent dataset from 2018 to 2020, the *FINC-01* dataset, to investigate the current trends and to explore whether there have been any shifts since 2014. Diagram P displays which salary bracket the majority of people fall under lately, which is the $50,000 - $99,999 bracket. Diagram Q is a representation of the same data as a histogram. Both Diagrams reveal that as years go by, there has been a steady increase in the upper class ($100,000 or more) population. Thus, we can infer people from $50,000 - $99,999 have moved up to the next bracket. The trend of the upper-class population increasing from *FINC-01* coupled with our knowledge of the lower divorce rates of the upper class from *Divorce Rates* allows us to make a prediction between income and relationships. Based on the trends, we expect that the divorce rates from 2018 to 2020 would be slower since there has been a steady increase in the upper class.

The findings from all three datasets contributed to our understanding of the link between sleep, money, and relationships. However, our datasets had some limitations. In our *Sleeping Alone* dataset, only 9% of participants were divorced. This constrained our ability to apply and infer our findings into a larger divorce population. This in turn caused our decision tree to have significantly fewer amounts of divorce data and caused an overwhelmingly high score for one class over the other. Another limitation of the *Sleeping Alone* dataset is the participant pool. People who completed the survey in that dataset were a specific population, consisting of mainly older adults that were in relationships. Therefore, this could have caused us to not see any significant effects of sleeping patterns on relationship length and status, since the people who filled out the survey have been together for a long time.

Despite their limitations, the datasets produced highly valuable insights, and through the analysis of these insights, we were able to provide a new perspective and a bridge between money, sleep, and relationships. Although we were not able to find any correlation between sleep and relationship length and status, we found that income was a crucial contributing factor to people's relationship length and status. In addition to our findings, we used our newly gained knowledge to predict the outcomes of divorce based on the change in financial trends over the years.

APPENDIX

List 1: Dependent variables used in *Sleeping Alone* dataset

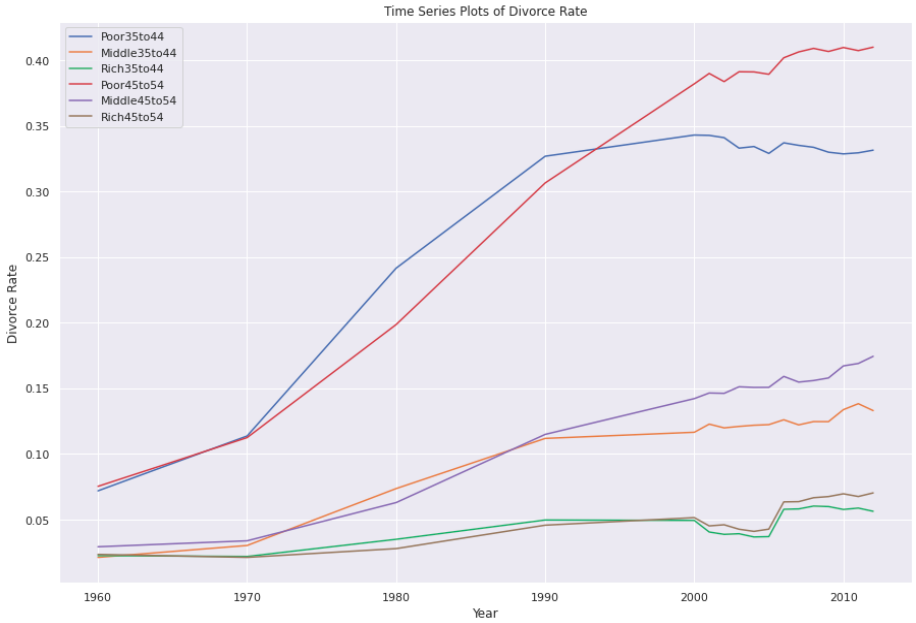
Dependent variables used in *Sleeping Alone* dataset
"Which of the following best describes your current relationship status?"
"How long have you been in your current relationship? If you are not currently in a relationship, please answer according to your last relationship."

List 2: Independent variables used in *Sleeping Alone* dataset

Independent variables used in *Sleeping Alone* dataset
"When both you and your partner are at home, how often do you sleep in separate beds?"
"To what extent do you agree with the following statement: sleeping in separate beds helps us to stay together."
"To what extent do you agree with the following statement: we sleep better when we sleep in separate beds."
"To what extent do you agree with the following statement: our sex life has improved as a result of sleeping in separate beds."
"Household Income"

Task 1

Diagram A: *Time Series Plot - Divorce Rate*



Task 2

Diagram B: *Cumulative Plot - Relationship Length*

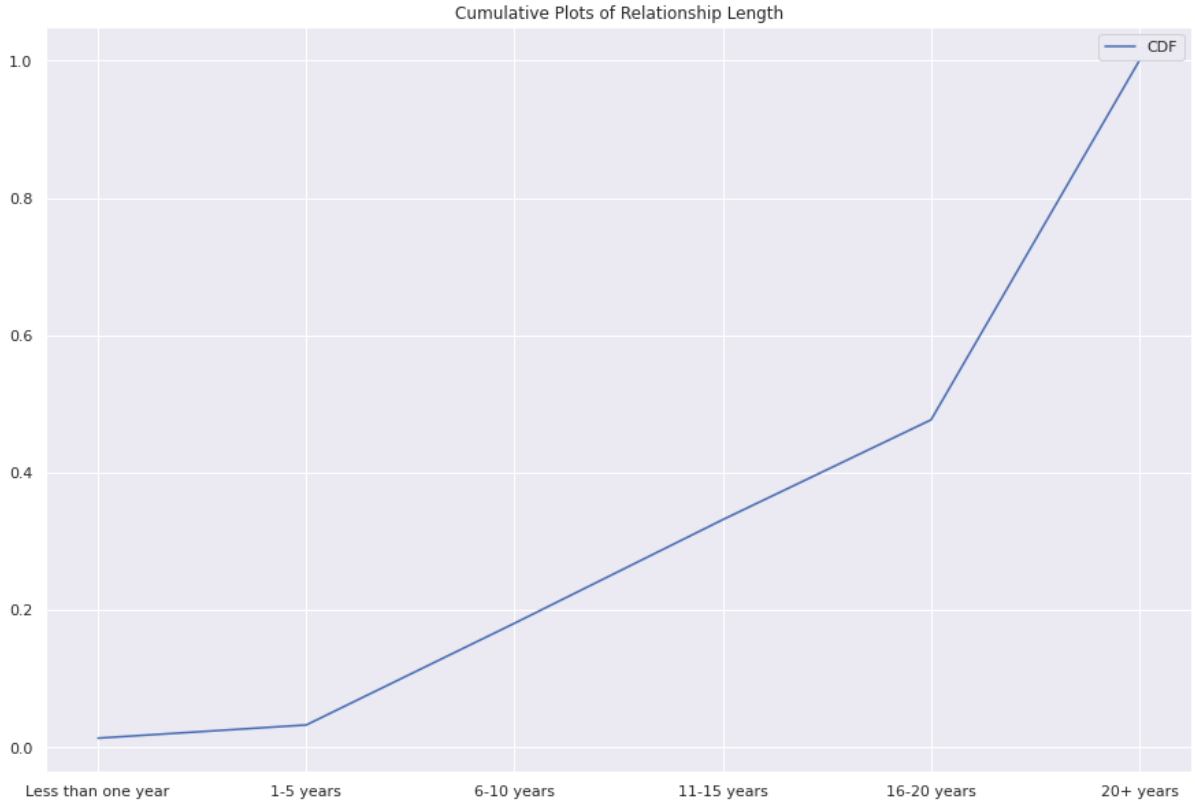
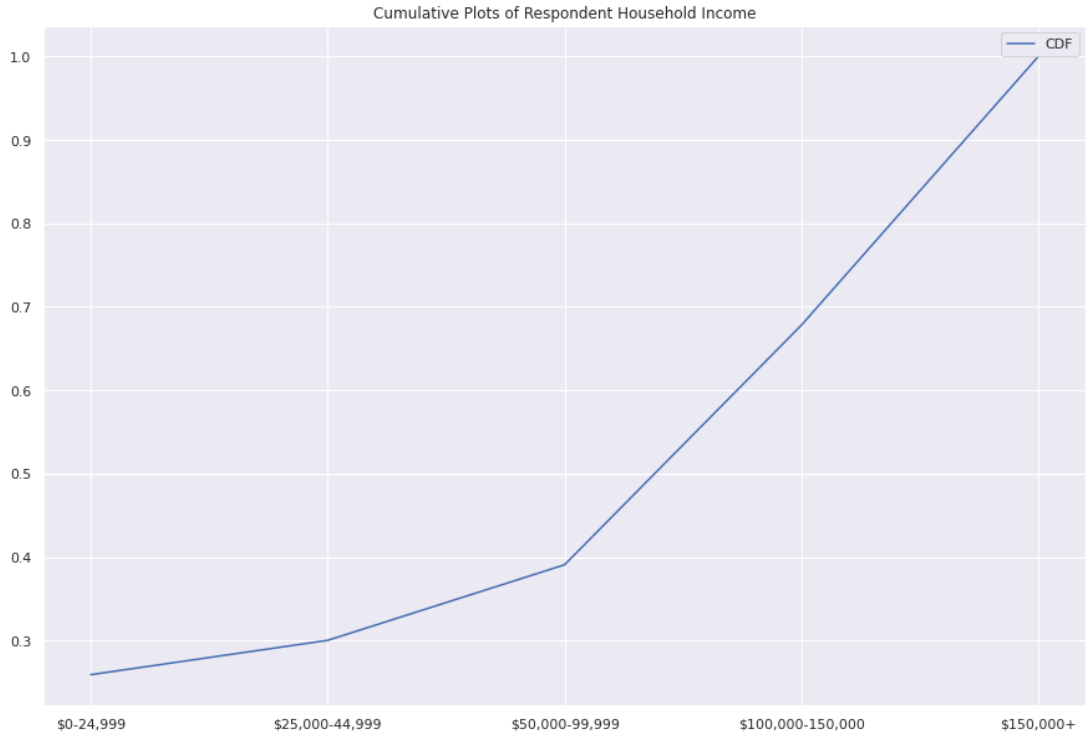


Diagram C: *Cumulative Plot - Household Income*



Task 3

Diagram D: *Rate of Change Plot - Divorce Rate (34-44-year-old age range)*

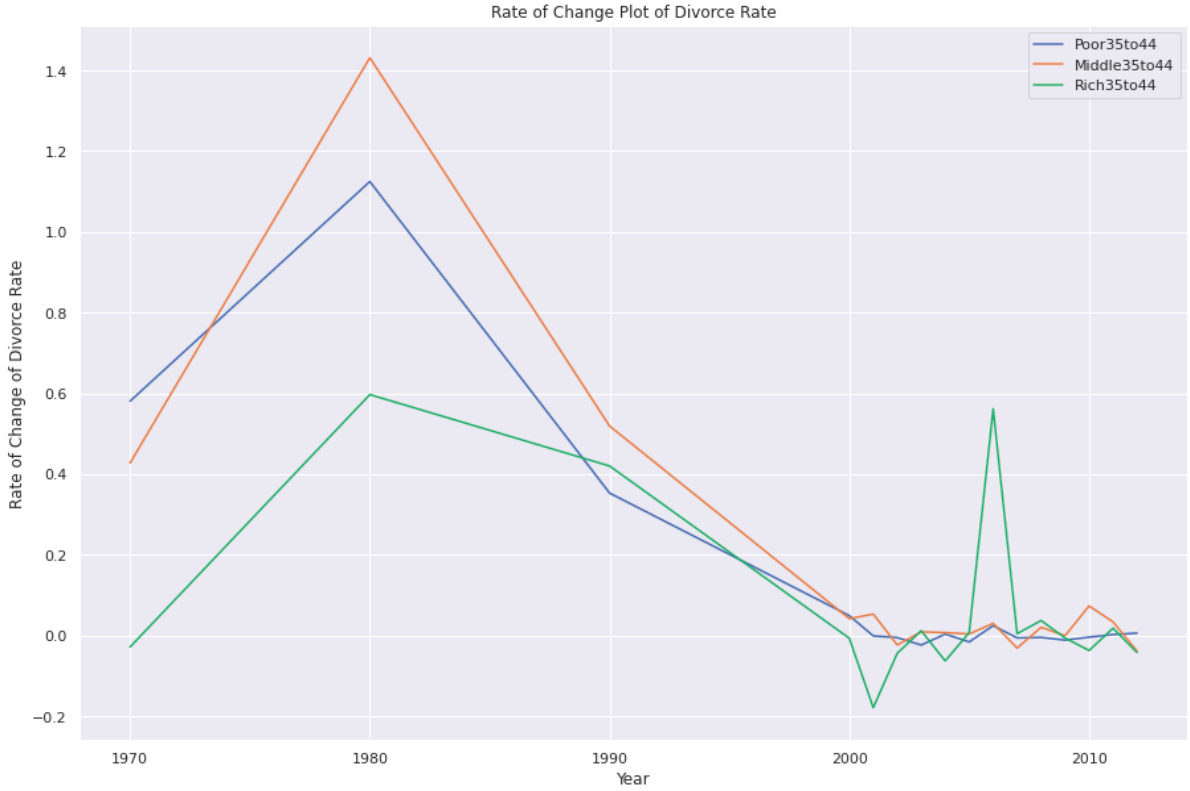
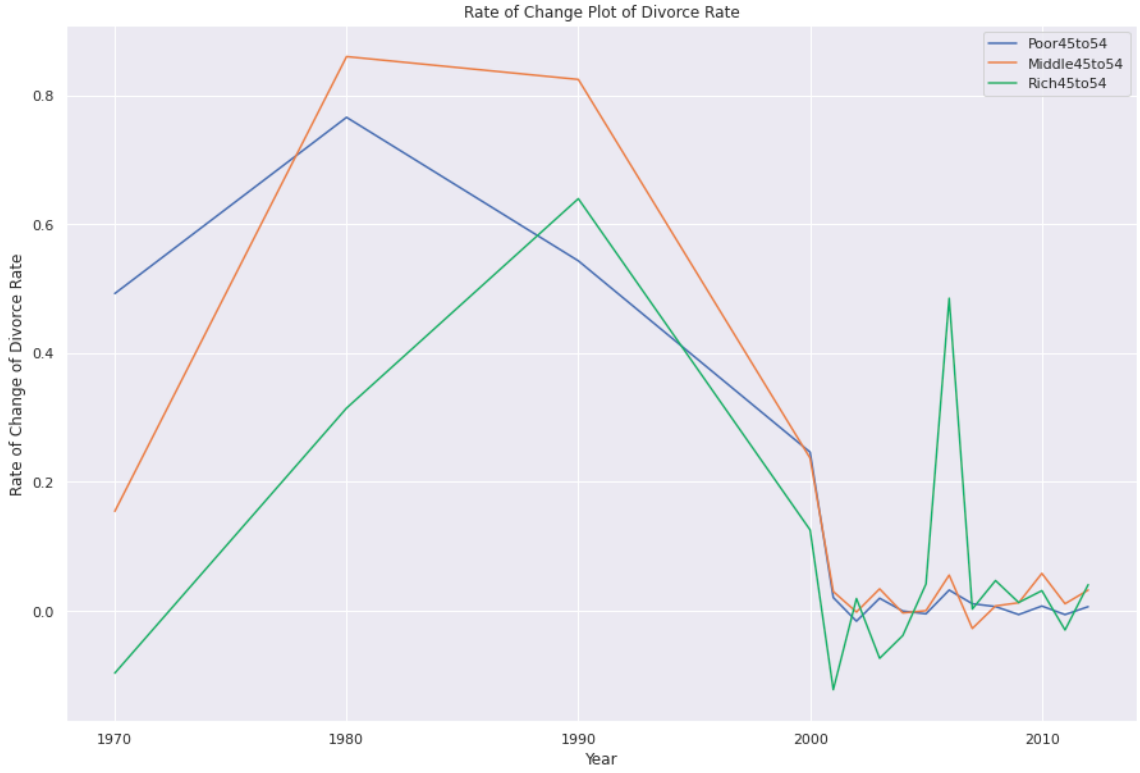
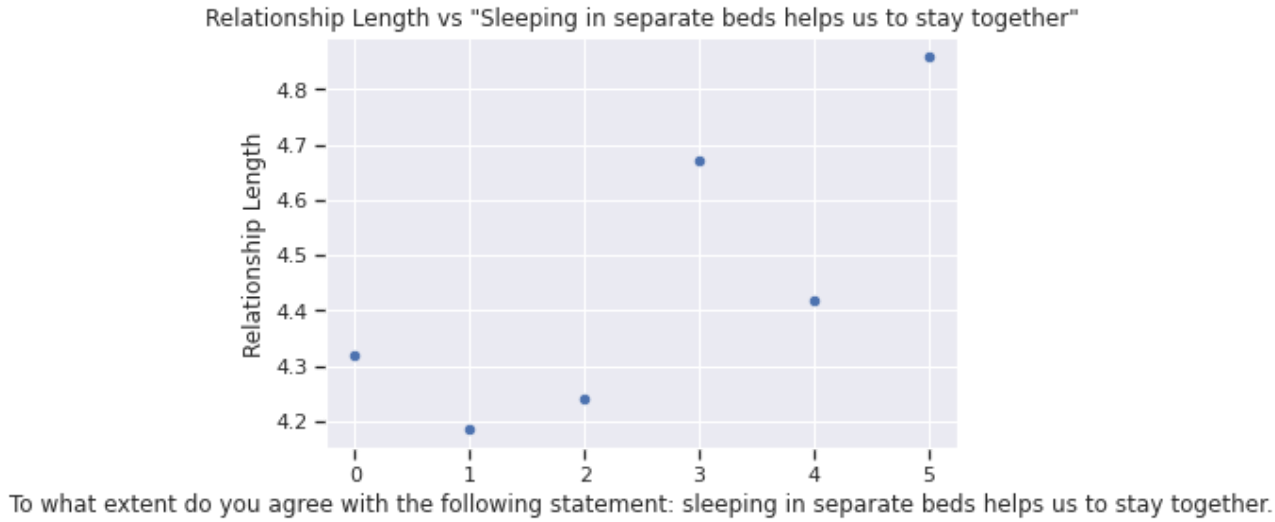


Diagram E: *Rate of Change Plot - Divorce Rate (45-54-year-old age range)*



Task 4

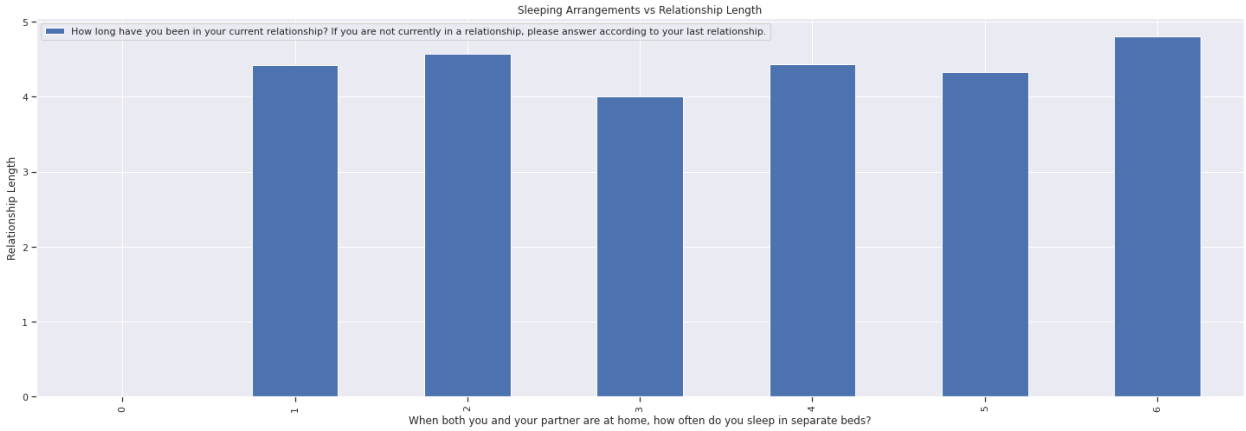
Diagram F: *Scatter plot - Relationship Length vs "Sleeping in separate beds helps us to stay together"*



	0	**1**	**2**	**3**	**4**	**5**	**6**
X: *"Sleeping in separate beds helps us stay together."*	Null	Strongly disagree	Somewh at disagree	Neither agree nor disagree	Somewh at agree	Strongly agree	N/A
Y: *Relationship Length*	Null	Less than 1 year	1-5 years	6-10 years	11-15 years	16-20 years	More than 20 years

Task 5

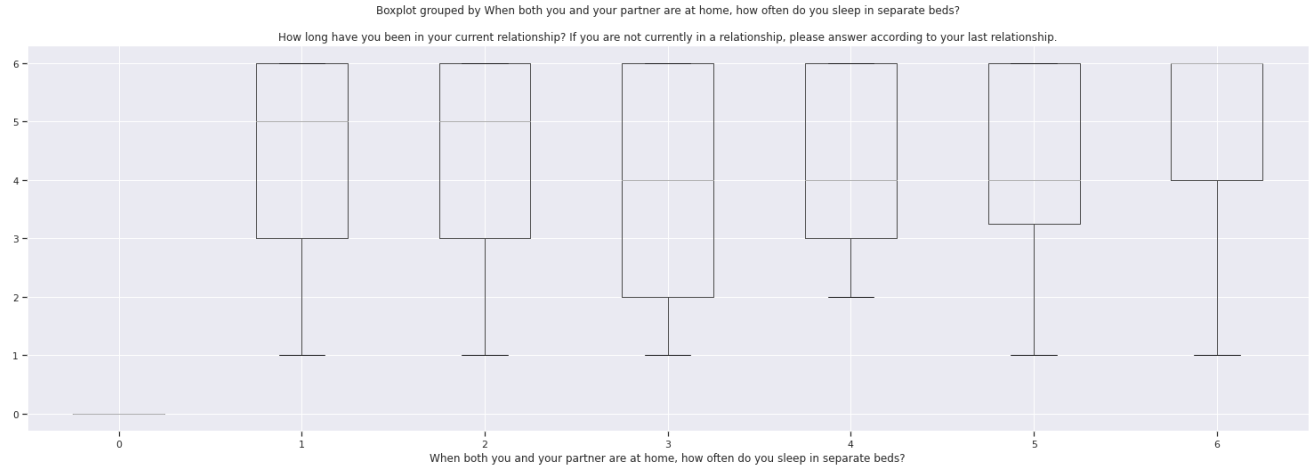
Diagram G: *Bar chart - Relationship Length vs Sleeping Apart Frequency*



	0	**1**	**2**	**3**	**4**	**5**	**6**
X: *"When both you and your partner are at home, how often do you sleep in separate beds?"*	Null	Never	Once a year or less	Once a month or less	A few times per month	A few times per week	Every night
Y: *Relationship Length*	Null	Less than 1 year	1-5 years	6-10 years	11-15 years	16-20 years	More than 20 years

Task 6

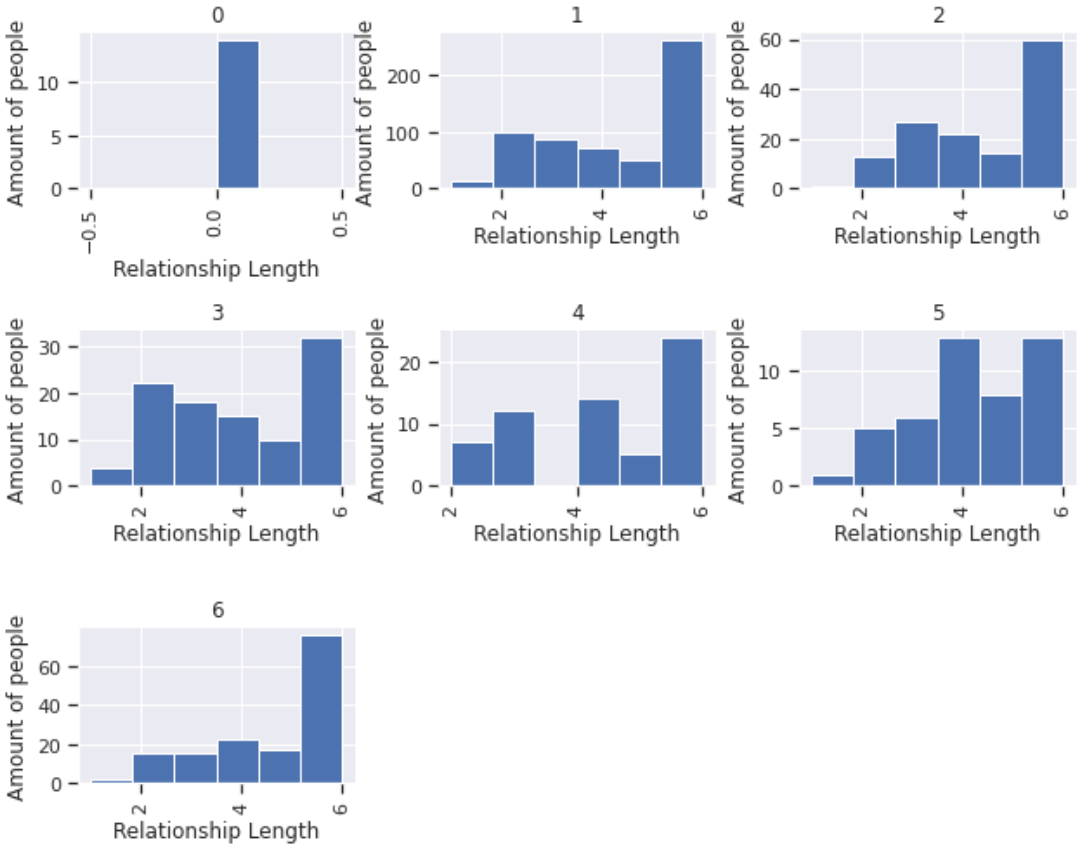
Diagram H: *Boxplot - Relationship Length vs Sleeping Apart Frequency*



	0	**1**	**2**	**3**	**4**	**5**	**6**
X: *"When both you and your partner are at home, how often do you sleep in separate beds?"*	Null	Never	Once a year or less	Once a month or less	A few times per month	A few times per week	Every night
Y: *Relationship Length*	Null	Less than 1 year	1-5 years	6-10 years	11-15 years	16-20 years	More than 20 years

Task 7

Diagram I: *Histogram - Sleeping Apart Frequency vs Relationship Length*



Each graph corresponds to one answer to the question: “When both you and your partner are at home, how often do you sleep in separate beds?”

	0	**1**	**2**	**3**	**4**	**5**	**6**
X: *Relationship Length*	Null	Less than 1 year	1-5 years	6-10 years	11-15 years	16-20 years	More than 20 years
Graph #: *“When both you and your partner are at home, how often do you sleep in separate beds?”*	Null	Never	Once a year or less	Once a month or less	A few times per month	A few times per week	Every night

Task 8

Diagram J: *Heat Map - All Variables*

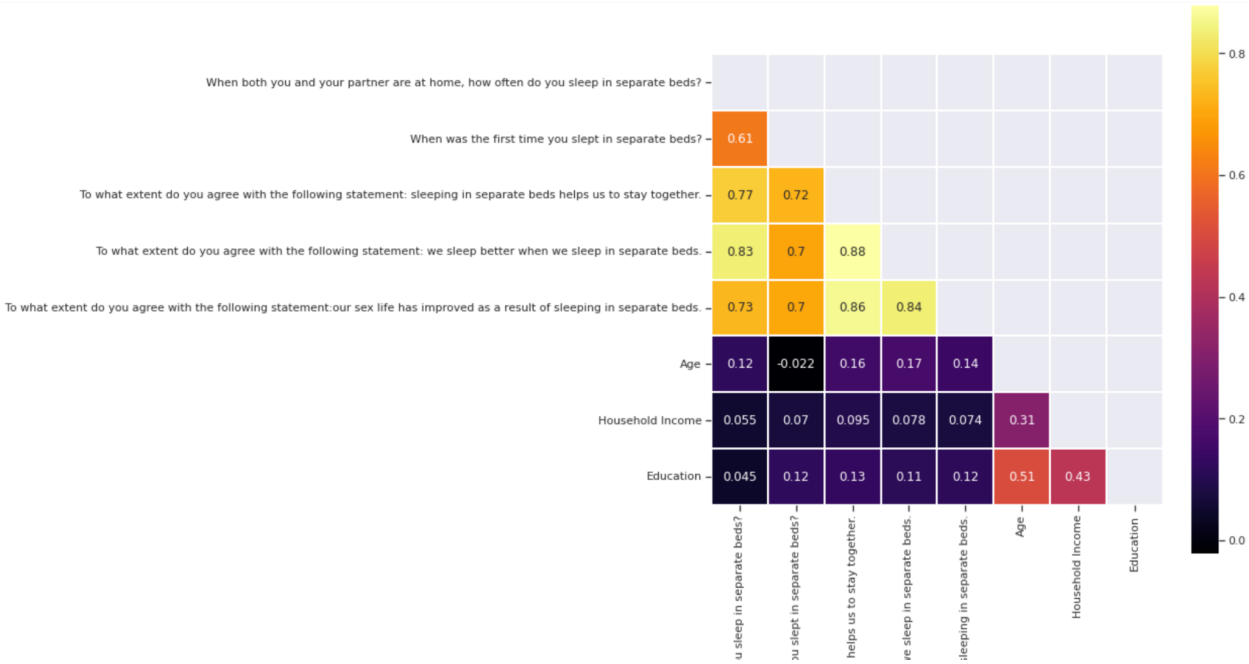
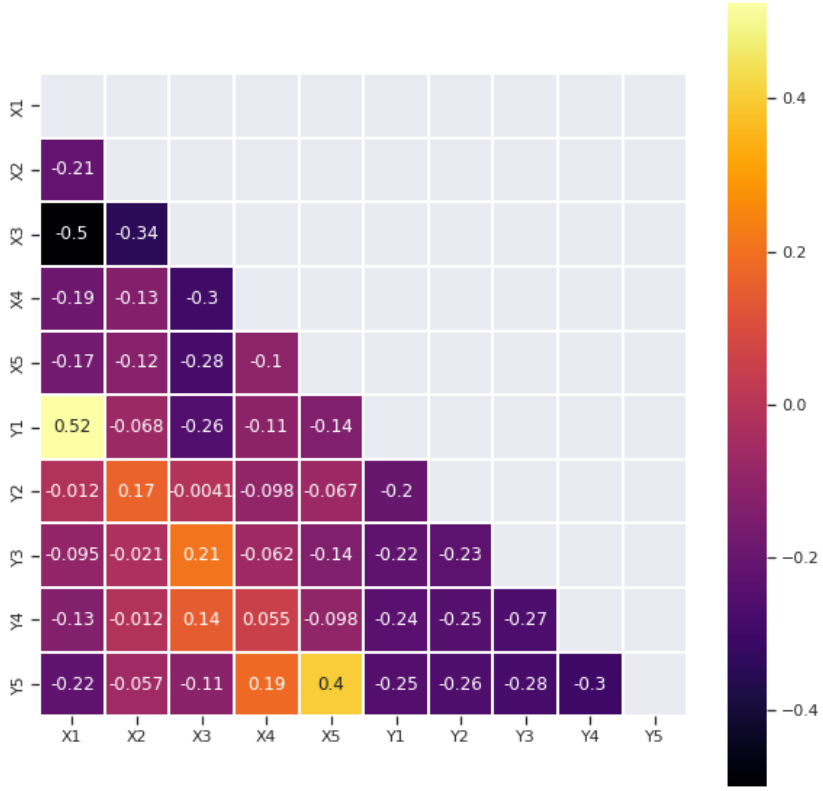


Diagram K: *Heat Map - "We Sleep Better When We Sleep in Separate Bed" Vs "Sleeping in Beds Helps Us Stay Together"*



	1	2	3	4	5
X: *We Sleep Better When We Sleep in Separate Bed*	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
Y: *Sleeping in Beds Helps Us Stay Together*	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree

Task 9

Diagram L: *Decision Tree - Full Tree*

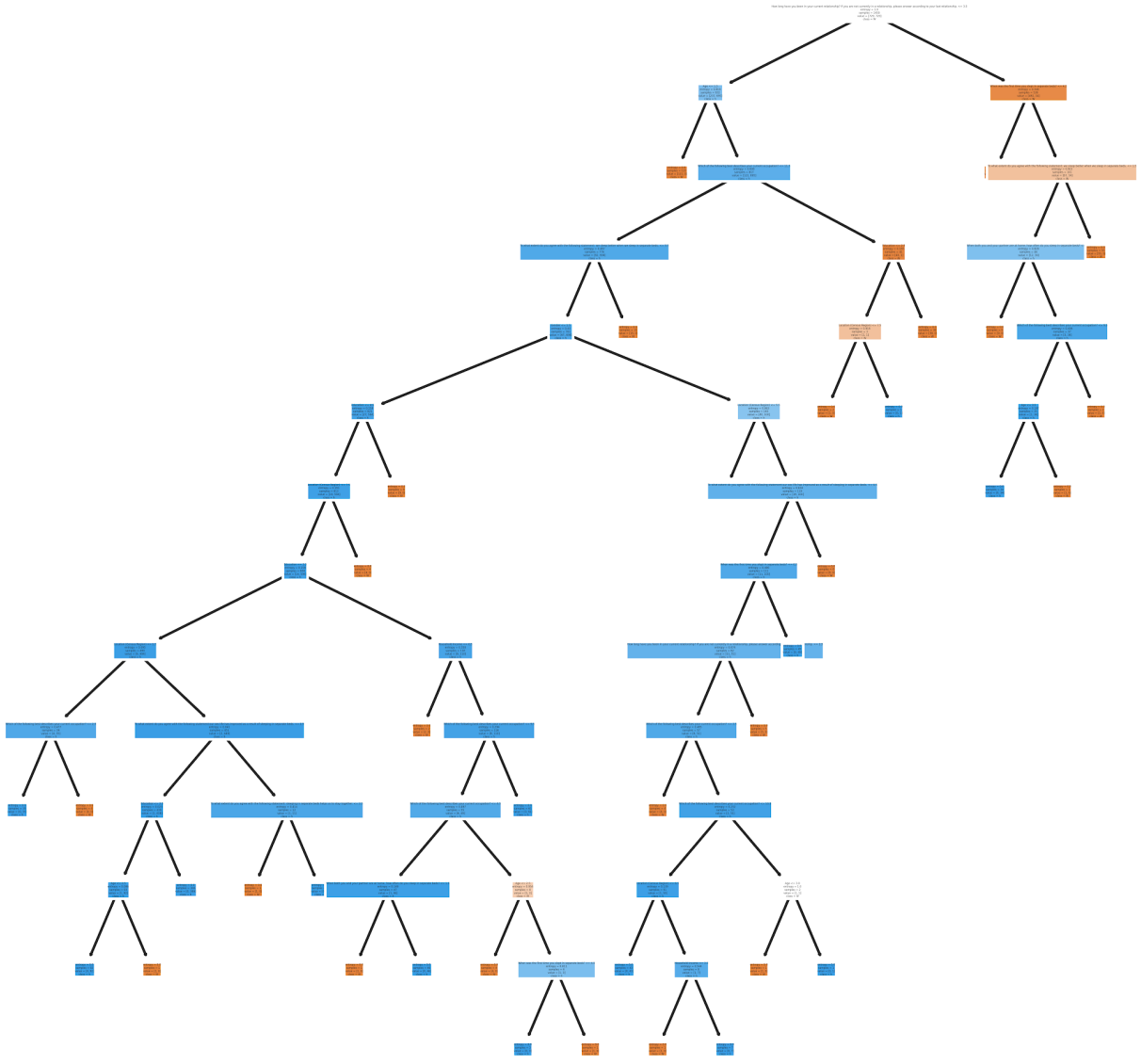


Diagram M: *Decision Tree - Max Depth of Four and Minimum Samples Nodes of Five*

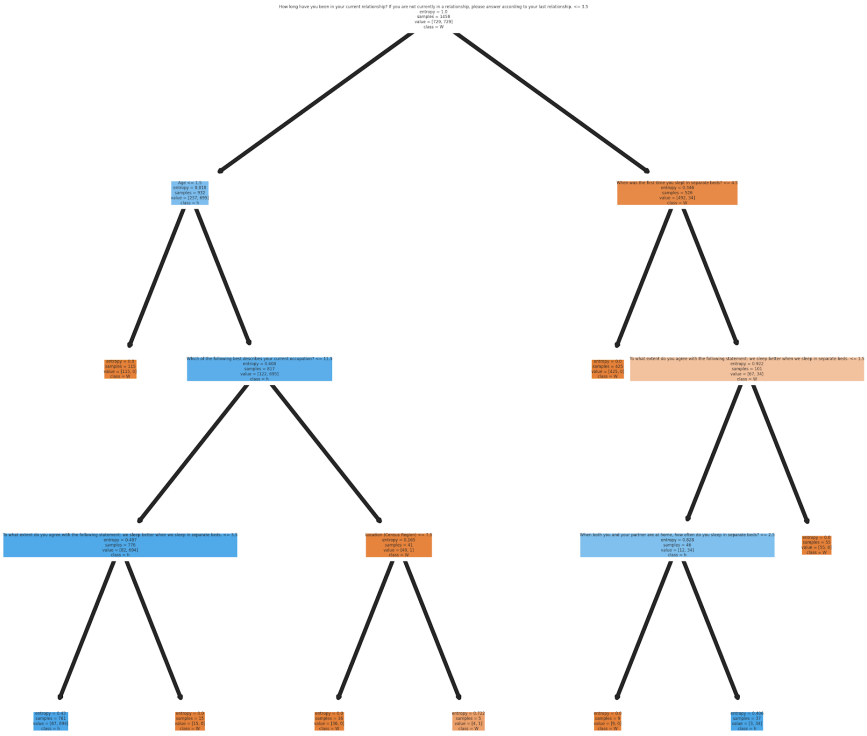


Diagram N: *Decision Tree - Chi-Squared Test*

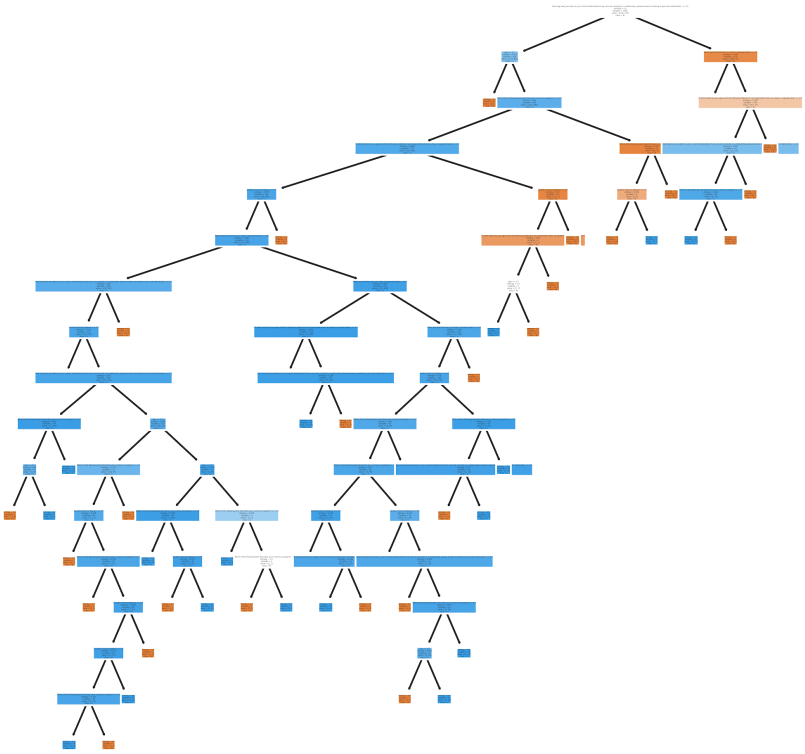


Diagram O: *Decision Tree - Chi-Squared Test With 2 Fewer Features*

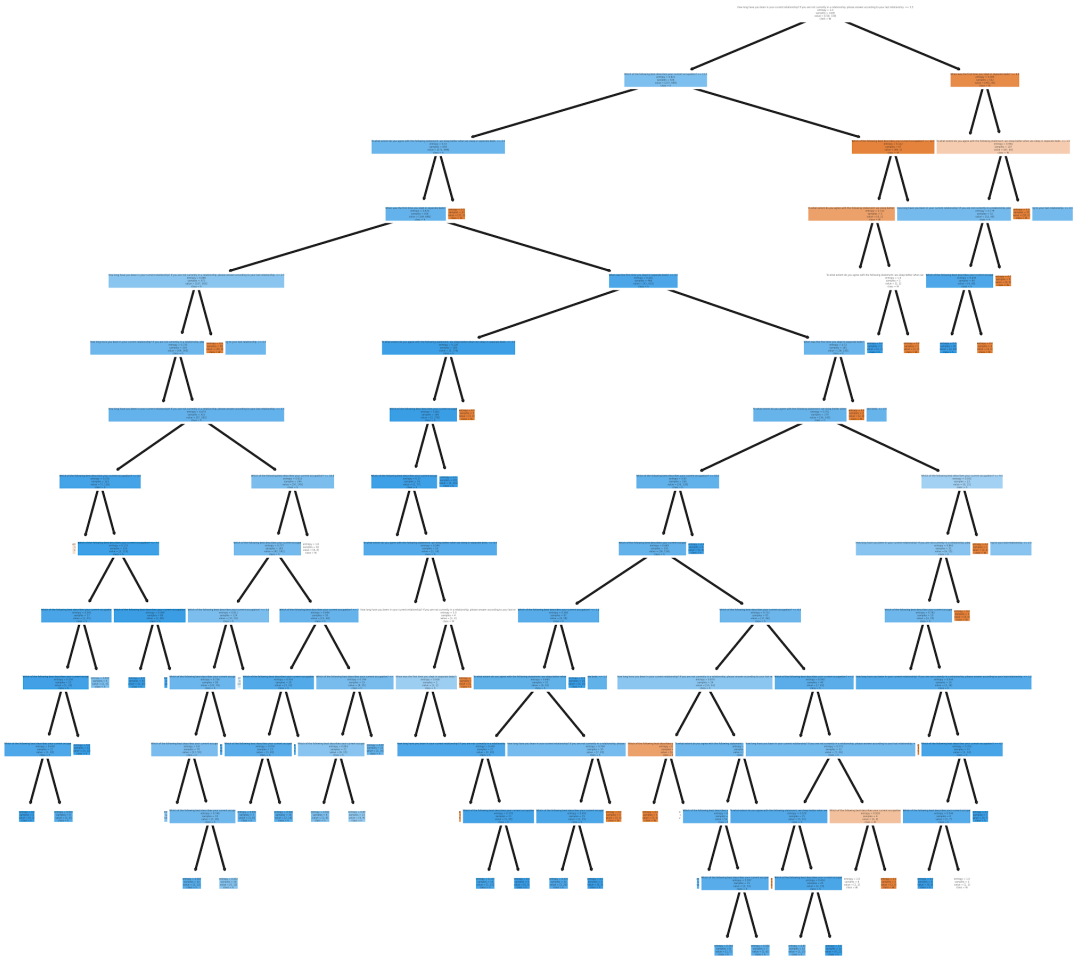


Table 1:

Decision Tree Performance Scores			
Full Tree	Max Depth of Four & Sample Nodes of Five	Chi-Squared Test	Chi-Squared Test With Two Fewer Features
0.989090	0.890909	0.981818	0.954545

Task 10

Diagram P: *Plotline - Population (in the Thousand) in Each Salary Range (2018-2020)*

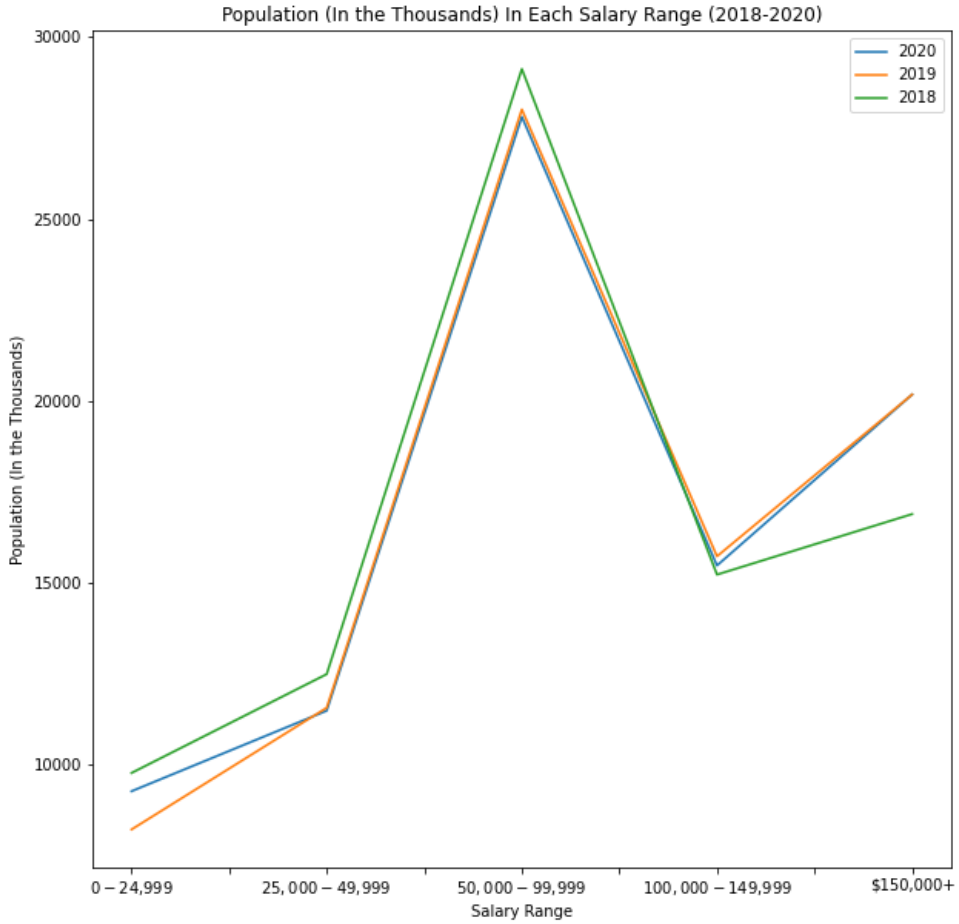
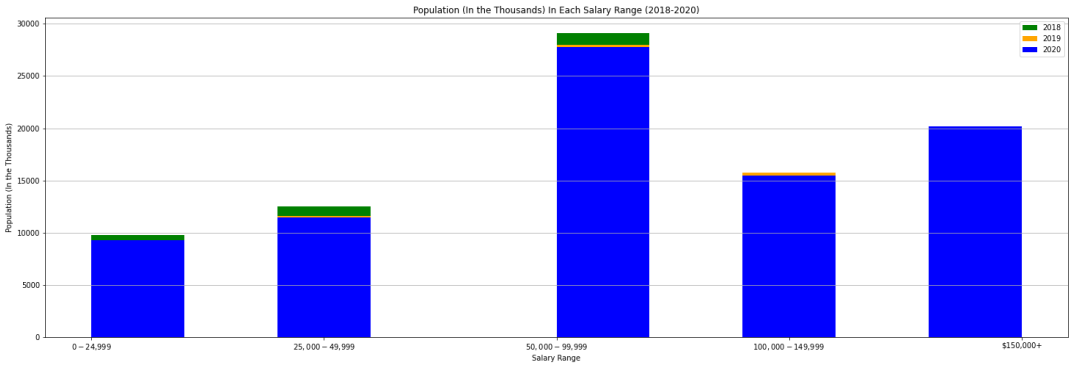


Diagram Q: *Histogram - Population (in the Thousand) in Each Salary Range (2018-2020)*



All of the code for our project are in the following links:
https://tinyurl.com/MainCode5450
https://tinyurl.com/FINC01Code

References

Drews, H. J., Wallot, S., Brysch, P., Berger-Johannsen, H., Weinhold, S. L., Mitkidis, P., Baier, P. C., Lechinger, J., Roepstorff, A., & Göder, R. (2020). Bed-Sharing in Couples Is Associated With Increased and Stabilized REM Sleep and Sleep-Stage Synchronization. *Frontiers in Psychiatry*, *11*. https://doi.org/10.3389/fpsyt.2020.00583

Drews, H. J., & Drews, A. (2021). Couple Relationships Are Associated With Increased REM Sleep—A Proof-of-Concept Analysis of a Large Dataset Using Ambulatory Polysomnography. *Frontiers in Psychiatry*, *12*. https://doi.org/10.3389/fpsyt.2021.641102

Troxel, W. M., Robles, T. F., Hall, M., & Buysse, D. J. (2007). Marital quality and the marital bed: Examining the covariation between relationship quality and sleep. *Sleep Medicine Reviews*, *11*(5), 389–404. https://doi.org/10.1016/j.smrv.2007.05.002

Troxel, W. M. (2010). It's More than Sex: Exploring the Dyadic Nature of Sleep and Implications for Health. *Psychosomatic Medicine*, *72*(6), 578–586. https://doi.org/10.1097/psy.0b013e3181de7ff8

Sleeping Alone Data. (2014, July 25). [Dataset]. FiveThirtyEight. https://github.com/fivethirtyeight/data/blob/master/sleeping-alone-data/sleeping-alone-data.csv

US Census Bureau. (2021, October 8). *FINC-01. Selected Characteristics of Families by Total Money Income.* [Dataset]. https://www.census.gov/data/tables/time-series/demo/income-poverty/cps-finc/finc-01.2020.html

Divorce. (2014, September 29). [Dataset]. FiveThirtyEight. https://github.com/fivethirtyeight/data/blob/master/marriage/divorce.csv