**STAT112 FINAL PROJECT**

**THE ANALYSIS OF DEPRESSION DATA**

Members:

1. Berrin Kolcu
2. Ceren Ak
3. Emir Çakır
4. Enes Ege Bozkır
5. Nuray Taghiyeva

**ABSTRACT**

This report presents the findings of an analysis of a study conducted in Australia in 2022, which collected information on various factors, including the number of children, gender, marital status, and some economic factors. The dataset was cleaned step by step and prepared for analysis. Then eight research questions were written to explore the relationship between these factors and the depression status of Australian people**.**

**INTRODUCTİON**

The report will focus on the eight research questions that were formulated to explore the relationship between various demographic and lifestyle factors and the economic status of individuals in Australia. This paper mainly investigates the association between the given variables and depression, such as whether the education level of people and depression are related or not, whether non-depressed individuals have more savings than the ones who are depressed, and if depressed people have plans for the far future and etc. The purpose of this report is to present the findings of the analysis and provide insights into the factors that may be related to the depression status of individuals in Australia. The results of this study can be used to inform future research and contribute to the development of policies aimed at improving the mental health of people.

**DATA TİDYİNG AND CLEANİNG STEPS**

1)**Renamed columns**: We renamed the columns to make them more meaningful, correct, and easier to understand. For example, there was a quotation mark in most column names (for ex:durable\_asset,save\_asset), and we removed those marks. T was missing in the column name No\_Lasting\_Investmen; we added t by the function rename. Then we made column names as titles.

2) **Duplicated rows:** We checked whether there were any duplicated observations; since the answer was 0, there was no need to drop rows in the dataset.

3) **Getting rid of the irrelevant strings in the values:** The best way to examine the values of strings is to create their frequency tables. The frequency table showed that Sex, Married, and Depressed columns had irrelevant punctuations or strings to be removed or replaced. Under the depressed column, we changed depressed/not\_depressed strings to yes/no, because it is easier to read in that case.

4) **Examining the descriptive statistics of the numerical variables:** The descriptive statistics showed that we had an unlogical value in the ‘Number Children’ column (99). We replaced this value with a median.

5) **Filling null values:** We created a missing percentage table for each column. Since no percentage was above 60, there was no need to drop any column. We observed that there was missing data in each variable. We filled them with the mode in an object variable column. Then we filled other columns with median because they had outliers.

**EXPLORATORY DATA ANALYSİS**

1. **What is the relationship between the saved asset and the depression condition of married/non-married individuals, and how does this relationship differ between the two groups in Australia in 2022**?

One possible graph to answer the question is a stacked bar chart with the x-axis depressed and y-axis saved asset because it allows for the comparison of asset and depression based on the marriage condition within the same graph. This allows us to see patterns in the data quickly.

Chart, box and whisker chart

Description automatically generated

This question examines the association between the saved asset and the depression condition of married/ not married individuals, based on the data provided, which shows the average saved assets of people. According to the data, on average, not-depressed but married individuals have 23 million saved assets, not-depressed not-married individuals have 25 million saved assets, depressed and married individuals have 21 million saved assets, and depressed not-married individuals have 24 million saved assets in Australia in 2022.

This data suggests that there may be a correlation between having more saved assets and not being depressed; it would indicate that having more savings could prevent financial stress, which also helps to prevent depression. Additionally, the fact that not-depressed and not-married individuals have more saved assets than not depressed married individuals, it can be inferred that being not married could be positively associated with having more savings.

However, the data also shows that depressed married individuals have fewer saved assets than not depressed married individuals; it means that marriage could be negatively related to having more savings. This could be a result of financial stressors that come with marriage, such as sharing expenses.

Additionally, the fact that the average saved assets of depressed individuals are lower than the average saved assets of not depressed individuals, regardless of their marital status, indicates that having fewer saved assets could be a risk factor for depression. This could be a result of financial insecurity, which can contribute to depression.

Furthermore, it is possible that the relationship between saved assets and depression could be affected by cultural and societal expectations of financial stability. For example, some cultures may place a high value on saving and financial stability, while others may place a greater emphasis on spending.

In conclusion, the data given suggests that there may be a correlation between the saved assets and the depression condition of married and non-married individuals, with not-depressed individuals having more savings on average than depressed individuals and not-depressed, not-married individuals having more savings than not depressed married individuals.

1. **Create a scatter plot matrix of numerical variables and interpret it.**

Graphical user interface, calendar

Description automatically generated with medium confidence. The scatter plots of Age, Number of Children, Education Level, and Saved assets vs. all other variables show that there is a stable relationship rather than being linear, where the points do not have a trend. This suggests that the increasing/decreasing of the given variable does not have any impact on the other variable.

The scatter plots of Gained Assets, Living Expenses, Other Expenses, Farm Expenses, and Lasting Investment mostly tend to show a weak positive linear relationship, where the points trend upward to the right. This suggests that as those variables of individuals increases, the other variables slightly increase.

The histograms on the diagonal show the distribution of each variable, which appears to be right-skewed for all variables except Education Level(left-skewed), with a few outliers.

1. **Is there any relationship between education level and depression in females?**

Chart, box and whisker chart

Description automatically generated

In general, it is thought that people with lower education levels may be more depressed because having a life with good education requires mental health. However, when we look at the plot prepared for this research topic, we see that depressive mood is unrelated to education. The average education level of depressed females is 8,7 and of non-depressed females is 8,5, which indicates that they are very close to each other. Outlier values are more common in females who are not depressed. Looking at this plot, we can conclude that low or high education does not affect depression much.

1. **It can be predicted that individuals with good salaries tend to be happy more, and the level of depression is lower. Is this statement true based on our dataset?**

Chart, box and whisker chart

Description automatically generated

When we analyzed this data, we thought that low-income people might be more likely to have depression because life might be more difficult for low-income people, and they might not be able to afford it. This can drive them into depression. But when we look at the graph, people with depression have higher average gained assets. The average gained assets of people who are not depressed are approximately 30 million Australian dollars, while those who are depressed are close to 35 million Australian dollars. The fact that people may have given up on things to be happy while trying to increase their assets may be effective in the result, or maybe depression is just a mental disorder not associated with assets.

1. **Do depressed people have plans for the far future? Interpret the question by looking at the relationship between lasting investment and depression.**

Chart, box and whisker chart

Description automatically generated

We can see those medians are exactly the same for both groups and equal to 2.8 million Australian dollars. People who are not depressed have many outliers, which are quietly high/low. The data of people who are depressed have more variability, approximately from 0 to 8 million Australian dollars, when compared to lasting investments of people who are not depressed. While most people who are not depressed prefer to make lasting investments of around 3 million Australian dollars, some people may prefer to make lasting investments of less than 1 million Austrian dollars. We can see that the median line is very close to the first quartile, so we may infer that approximately half of the people are making a comparatively small amount of lasting investment which ranges from approximately 0.2 to 2.5. While the lasting investments of people who are not depressed have more consistent values, the lasting investments of those who are depressed may vary and have very high or very low lasting investments, so depression may lead people to invest more or less than normal. In conclusion, by looking at these two boxplots, even if we see that some people's investments are less than normal, we can say that generally depressed people have plans for the distant future.

1. **Whose living expense is higher: the women in depression or the women who are not in depression?**

Chart, box and whisker chart

Description automatically generated

We can say that the medians are the same in both plots, around 2.5 Australian million dollars. The living expenses of women who are depressed have less variety when compared to the living expenses of people who are not depressed. Half of the women who are depressed and who are not depressed expend around 1 and 3 million Australian dollars. The median line is very close to the third quartile in both plots; we can say that %25 of the women on both plots expend money around 3 million Australian dollars. There are more outliers on the boxplot, which shows the living expenses of women who are not depressed than other plots. Both plots have a negatively skewed distribution. The amount of money between Q3 and the maximum living expenses of women who are depressed generally has lower amounts relative to the another. In conclusion, we can say that by looking at these boxplots, depression causes some females to have relatively fewer living expenses. This may be due to the fact that people who are depressed do not want to do anything.

1. **Is there a difference between the relationship between living expenses and gained assets for depressed and non-depressed people?**

**Chart, scatter chart

Description automatically generated**

We looked at the relationship between living expenses and gained assets for depressed and not-depressed people. As you can see, there is no significant relationship between these two variables for people who are not depressed. However, there is a positive correlation for depressed people shown as the trend line. We can say that depressed people with more gained assets tend to have higher living expenses. This leads us to deduce that depressed people are more likely to spend the money they gained during the year.

**CONCLUSION**

The results of the analysis provided insights into the factors that may be correlated/not correlated with the depression status of individuals in Australia. For example, the analysis revealed that non-depressed individuals have more savings on average compared to depressed individuals. Moreover, depression causes some females to have relatively less living expenses. Additionally, the study found that there is almost no relationship between education level and depression. Furthermore, the analysis revealed that investments tend to be positively associated with the depression status of individuals, with those who reported having more investments being in depression.

It is important to note that the study is limited to one year of data collection in Australia; thus, the findings may not be generalizable to other populations or time periods, and further research is needed to confirm these findings.