

Analysing Social and Psychological Effect of the Covid-19 Pandemic in Turkey

Link to the Paper:

<https://data.mendeley.com/datasets/sv95c7ydp/9>

Team Members:

Abhishek Sharma

Tarun Jindal

Urvish Pujara

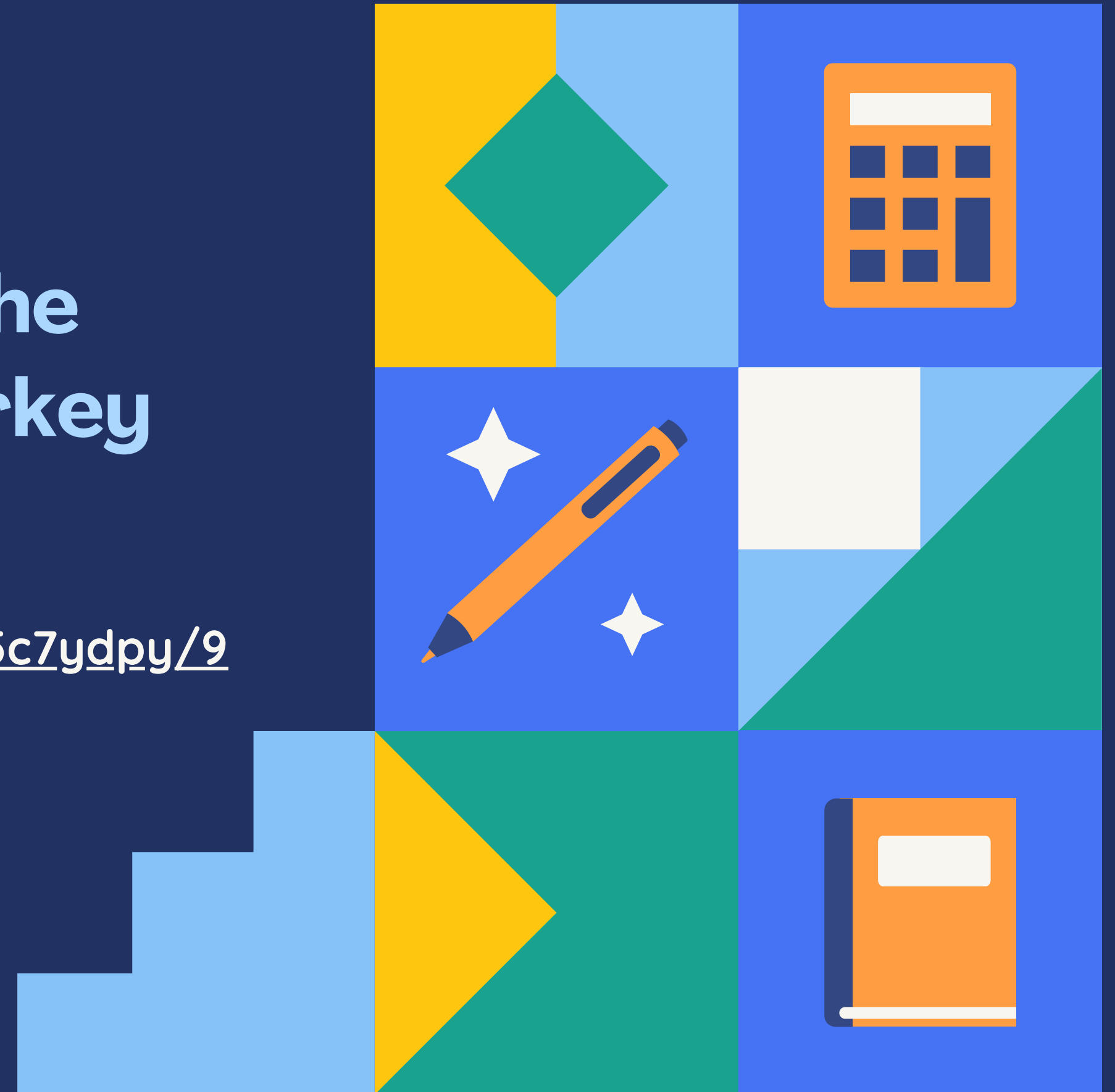
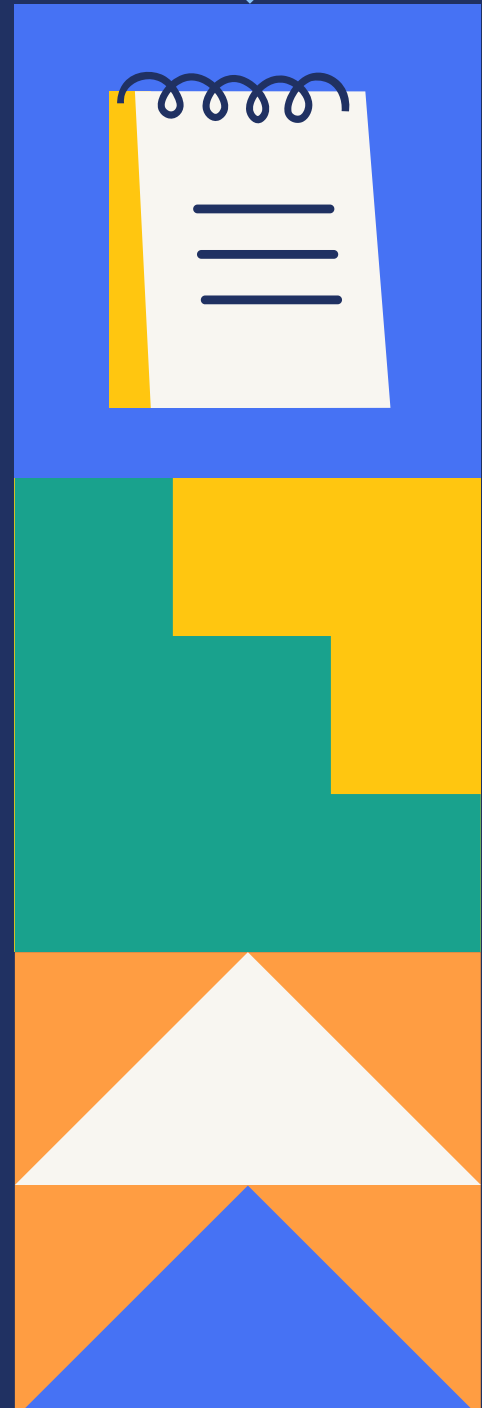
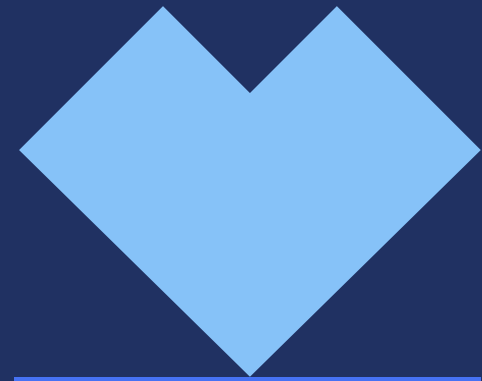


Table of Content



Introduction to the
problem statement



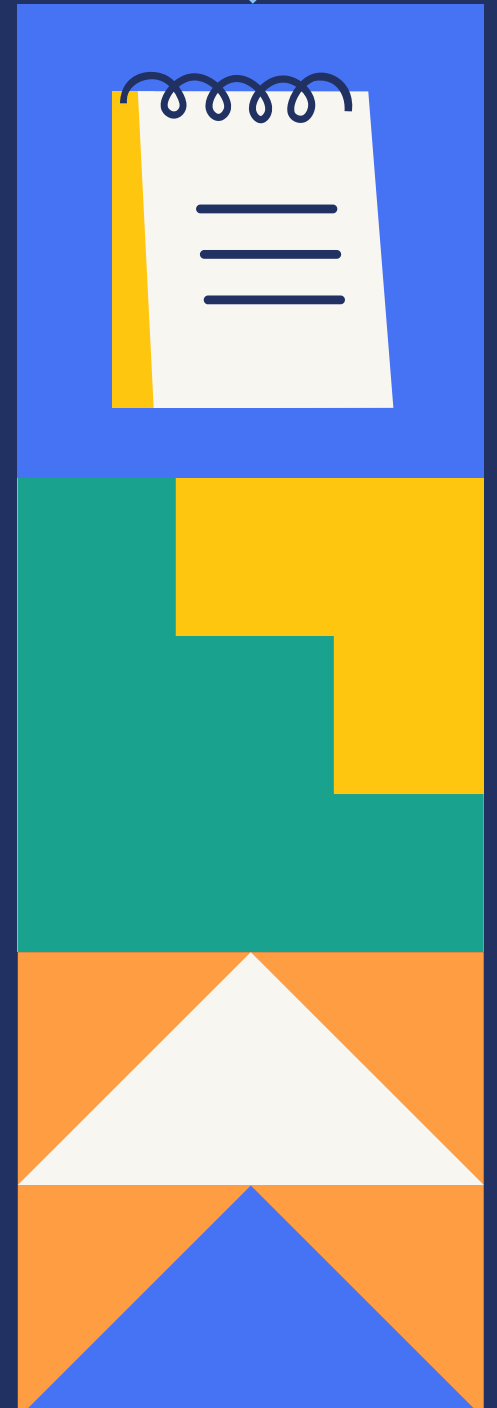
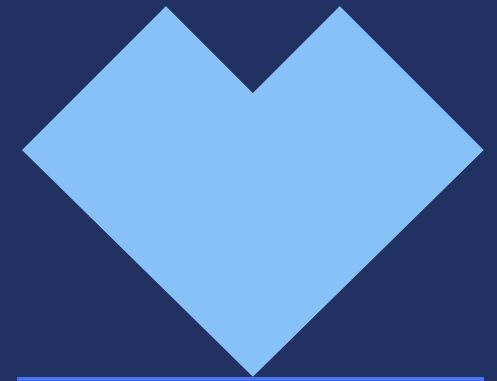
Dataset Selection



Dataset Preprocessing



Hypothesis and Results



Introduction

The COVID-19 pandemic ravaged the world began in Wuhan, China, in December 2019 and reached Turkey on March 10, 2020 . Medical research is still being conducted to determine the impact of this pandemic on humans in Turkey and around the world.

The epidemic has had a significant impact on the physical and mental well-being of many Turkish people. Despite the government's efforts to keep COVID-19 under control, the social isolation caused by the outbreak has had a significant impact on people's lives.

So we are using the survey of Turkey, presented by the paper to determine the impact of pandemic



Considering your online spending habits now compared to before COVID-19, which statement is more appropriate for you?



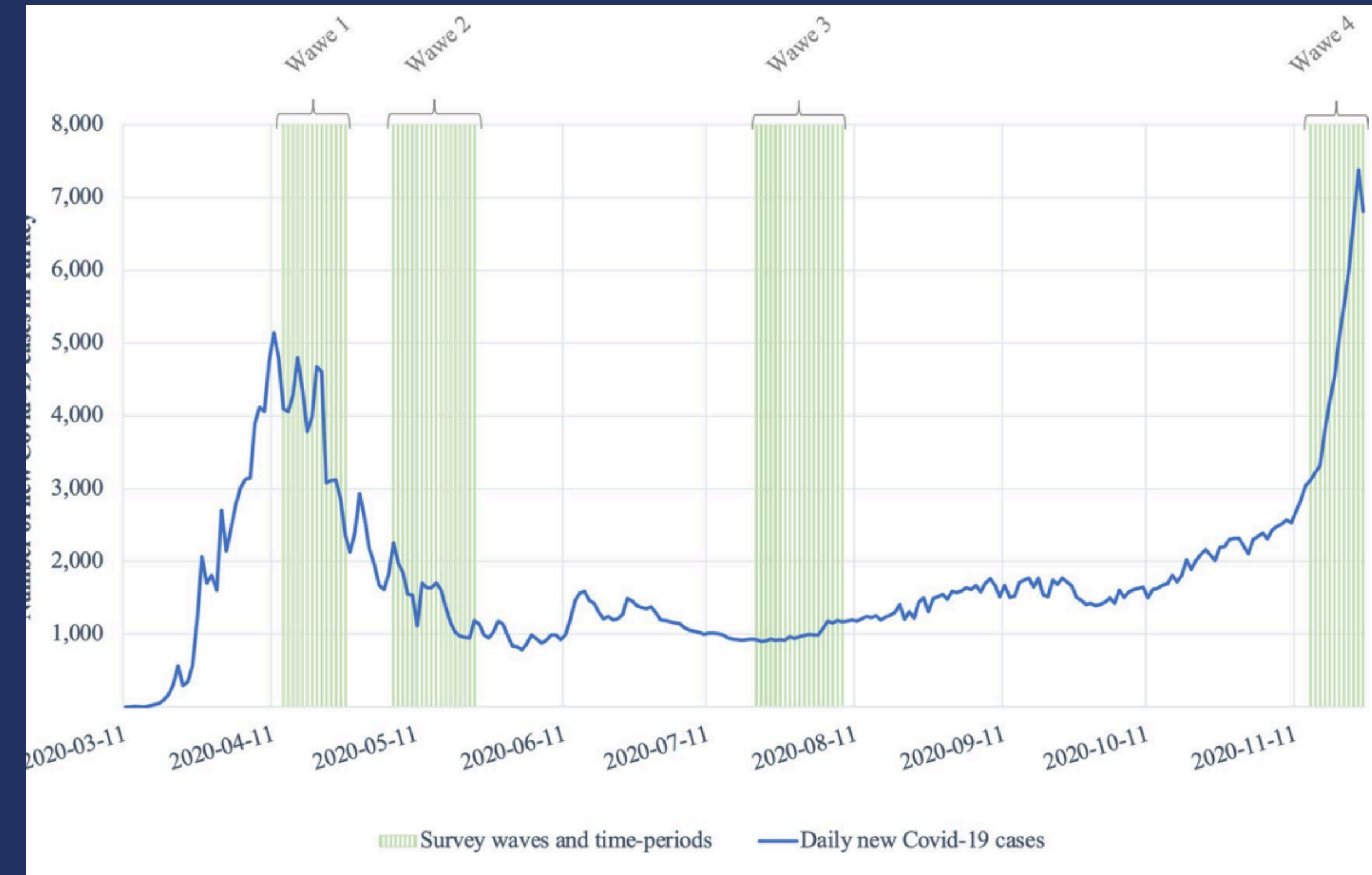
Dataset Selection

- A total of 2,817 people were surveyed online, and the results are compiled in this dataset.
- Data were collected from adults aged 18 to 65 and older, with an extensive demographic section covering location (province and rural-urban divide), income, employment status, occupation, occupational sector, family background, whether having a child, marital status, and gender identity
- The survey is conducted in four waves, where first one is in mid April daily number of newly confirmed COVID-19 cases peaked, the second is the first period in which the number of COVID-19 cases decreased, the third is the end of the summer season when the number of cases is low, the fourth is the period when the number of cases increased very rapidly

Category	Variables	N	Missing	Distinct
Demographic	Age	2817	0	49
	Marital status	2817	0	4
	Gender	2817	0	3
	Age range of your child(ren) (if any)	2817	0	25
	City where you live (provinces)	2817	0	78
	Urban-rural divide	2817	0	2
	Education level	2817	0	7
	Mother's education level	2773	44	8
	Employment status	2817	0	9
	Occupational sector	2817	0	7
	Occupation (ISCO-08)	2506	311	10
	Parent's occupation (ISCO-08)	2817	0	10
	Home office	2411	406	3
	Current income (in month)	2817	0	6
	Last year income (average)	2817	0	6
Health	Self-rated health	2817	0	5
	Personal hygiene behaviour	2817	0	5
	Having a chronic disease	2817	0	2
	Diagnosed COVID-19	2817	0	2
Security	Safety perception	2817	0	5
	Community safety perception	2817	0	5
	Feeling uneasy in crowded places	2817	0	2
EconomicBehavior	Usage of debit/credit card	2817	0	5
	Online Shopping	2817	0	5
	Stockpiling	2817	0	2
State-Trait Anxiety Inventory (20-item sub-dimension)		2817	0	5
Measurement of Distress Intolerance		2817	0	5

Dataset Selection

The paper is just about collection of dataset through survey, it does not provide any hypothesis or tests.



	Population	Sample size
Total population of Turkey	80,810,525	
Above 15-year-old population	62,100,651	
Total survey participation		2,817
Survey waves		
Wave 1: April 13–26		1,124
Wave 2: May 6–23		975
Wave 3: July 20 - August 8		515
Wave 4: November 14–25		203

Dataset Preprocessing

- Age is changed into levels instead of **continuous** variable . <20, <=34, <=44, <=54, <=64, 65+
- Marital Status, Sex, working sector, working situation, occupation are already **categorical** variables , mapped to integer values.
- Similarly variables like education, mother-education, income , previous year income are considered as **ordinal** variables.
- There are questions like Stockpiling, diagnosed with chronic illness, diagnosed with Covid-19 which are Yes/No only (**categorical**)
- Personal Security Concern, Belief on Government for people security, change in Hygiene habits, usage of credit/debit card are **ordinal** variables. (increased a lot, increased, same,)
- There are 20 State-Trait Inventory questions which are also **ordinal** variables (low, moderate, high)
- The 10 distress tolerance questions are also **ordinal** variables (Extremely, Quite a bit, moderately.....)

Hypothesis 1

	Variable	Chi2	p-value	Dof	
0	sto_n	6.088586	0.013606	1	significant,
1	hyg_n	18.903093	0.000821	4	significant,
2	int_n	6.707053	0.152203	4	not significant, fail

	Variable	Stats	p-value	
0	sto_n	171535.5	0.010212	significant,
1	hyg_n	149924.5	0.645847	not significant, fail
2	int_n	141846.0	0.154336	not significant, fail

- There is a significant relationship between being diagnosed with COVID-19 and changes in stockpiling behavior and personal hygiene behavior.
- However, there is no significant relationship or difference in online spending habits based on COVID-19 diagnosis.

H0

There is no significant relationship between the individuals or household members being diagnosed with **COVID-19** and the changes in personal cleaning behavior, online spending habits or stockpiling behavior.

H1

There is a significant relationship between the individuals or household members being diagnosed with **COVID-19** and the changes in personal cleaning behavior, online spending habits or stockpiling behavior.

Statistical Test

We use **Chi Square Test** of Independence to check whether two variables are likely to be related or not. We can use **Man-Whitney-U** Test also for ordinal variable.

Hypothesis 2

Performing Shapiro Test for checking Normal Distribution leads to no variable to be normal distributed.

So we use Kruskal Wallis Test to check the dependency of self related health on different demographic factor variables.

	factor	p-value	Result
0	age_n	1.728047e-12	Reject H0
1	sex_n	2.280859e-17	Reject H0
2	mar_n	1.829332e-12	Reject H0
3	urb_n	7.376307e-01	Fail to reject H0
4	edu_n	1.448996e-07	Reject H0
5	med_n	1.940519e-02	Reject H0
6	iin_n	6.273135e-23	Reject H0
7	hin_n	1.999993e-14	Reject H0
8	hom_n	1.543757e-09	Reject H0

Perform Posthoc test to determine which specific group for every factor variable differ significantly from each other.

H0

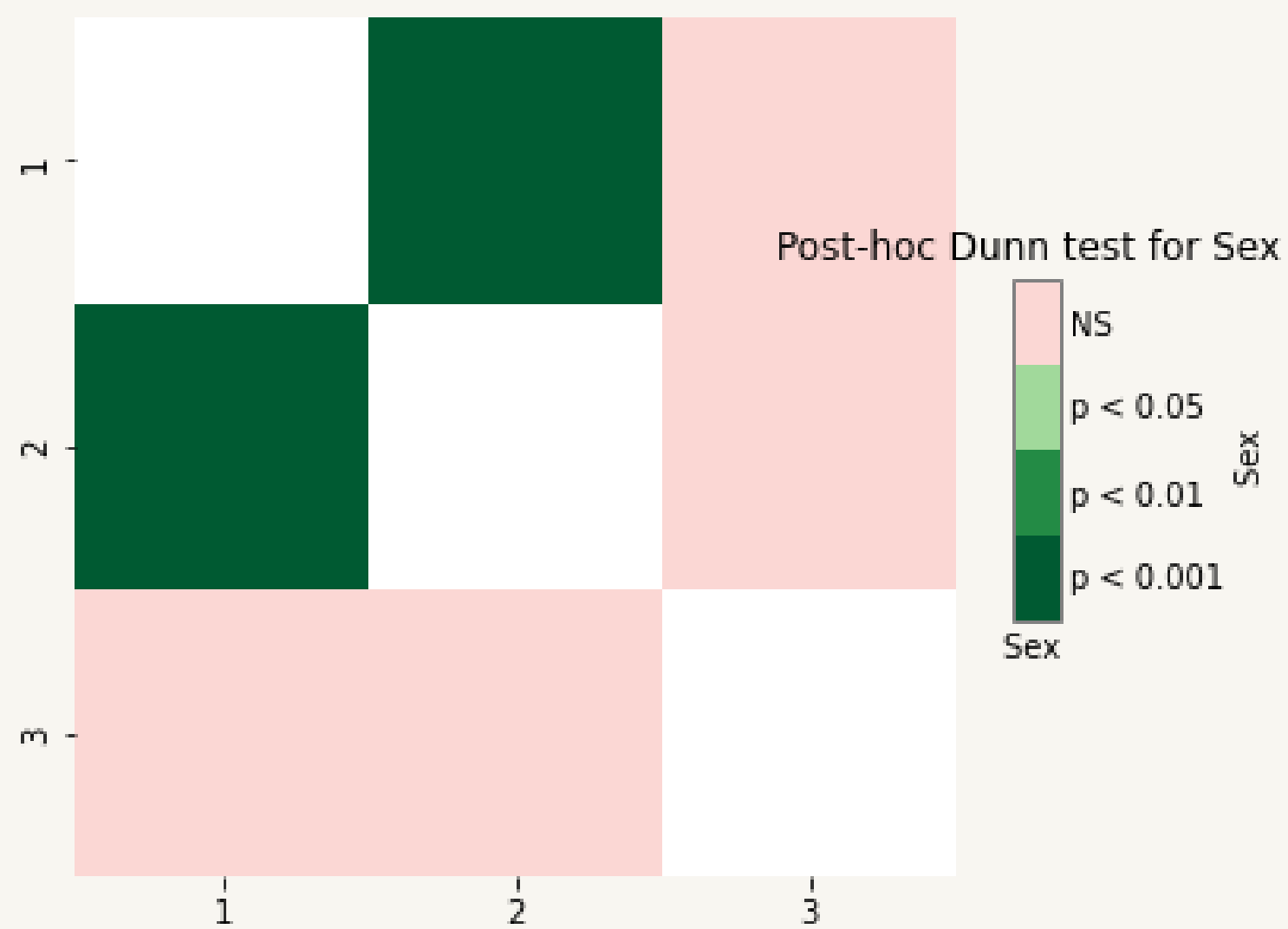
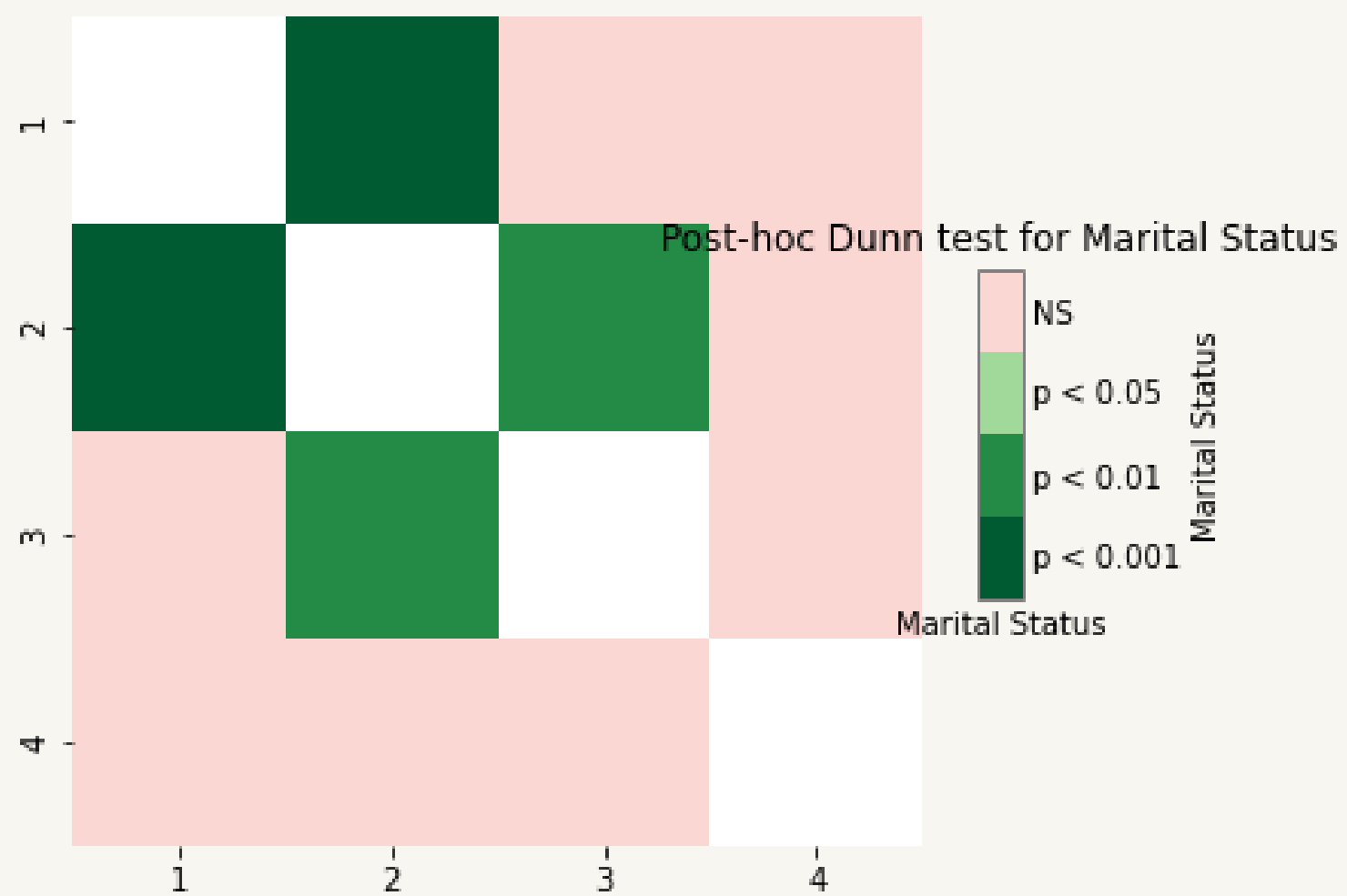
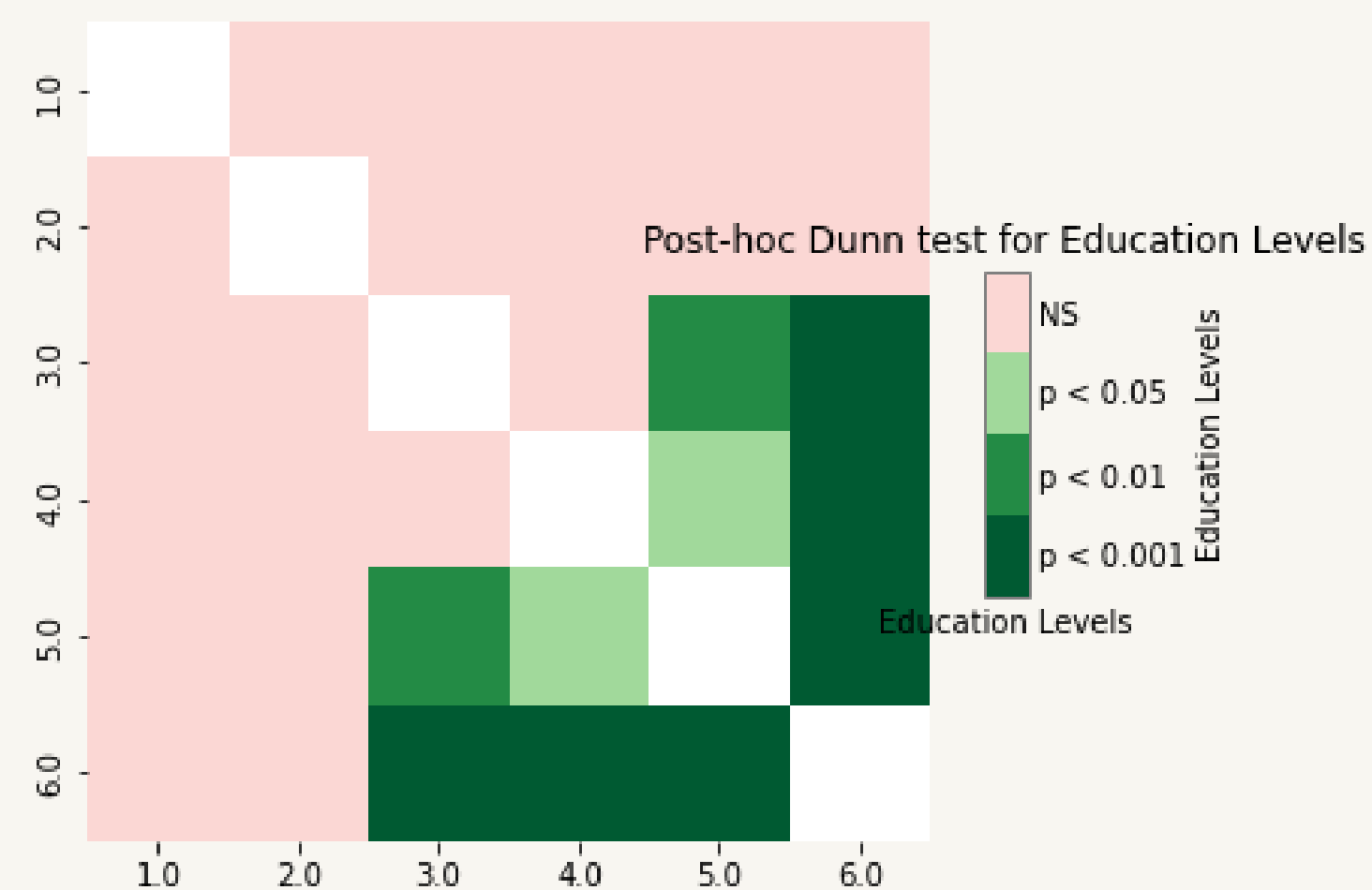
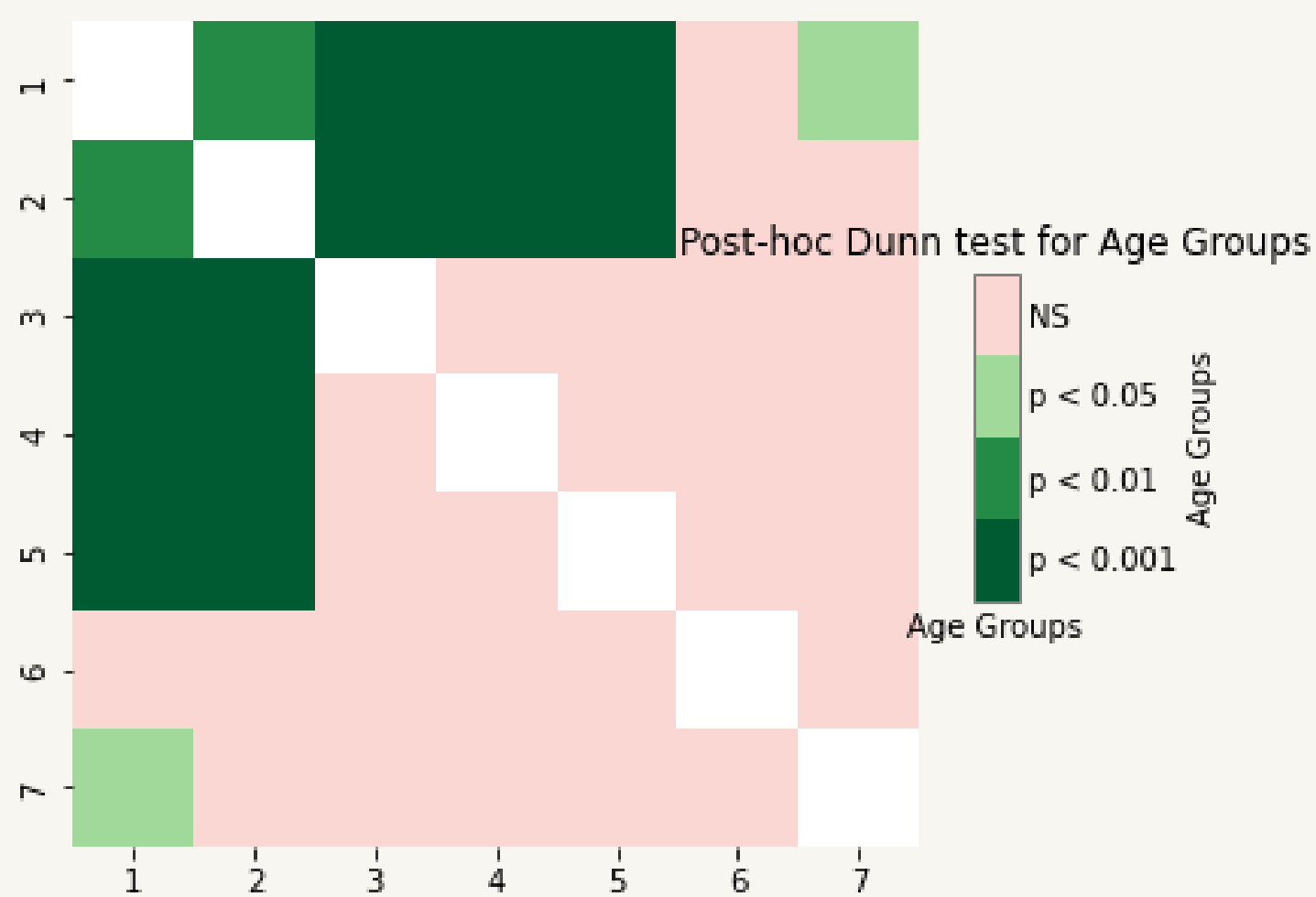
H(0) : There is no significant relationship between demographic factors(age, martial status, sex, urban/rural) and self rated health status given by the patients.

H1

There is a significant relationship between demographic factors (age, martial status, sex, urban/rural) and self rated health status given by the patients.

Statistical Test

First check the normality of variables (age, urba/rural, marital, sex, education, income, prev income etc). They do not give normal distribution, so we opt **Kruskal Wallis Test** for the hypothesis.



Hypothesis 3

Positive Emotions

- sur1_n = I am currently calm
- sur2_n = I feel safe
- sur5_n = I am at peace right now
- sur8_n = I feel physically and mentally rested
- sur10_n = I feel comfortable right now
- sur11_n = I have confidence
- sur15_n = I feel relieved
- sur16_n = I am happy with my current situation
- sur19_n = I am happy right now
- sur20_n = I am in good spirits right now

Negative Emotions

- sur3_n = I am nervous right now
- sur4_n = I feel regret
- sur6_n = I am not having fun right now
- sur7_n = I worry about what will happen to me
- sur9_n = I am concerned about my health
- sur12_n = I am in bad temper right now
- sur13_n = I am so angry
- sur14_n = I feel my nerves are so tense
- sur17_n = I am worried right now
- sur18_n = I feel stunned with excitement

We divided emotions into two parts and tried to find how they are correlated with individual's health

H0

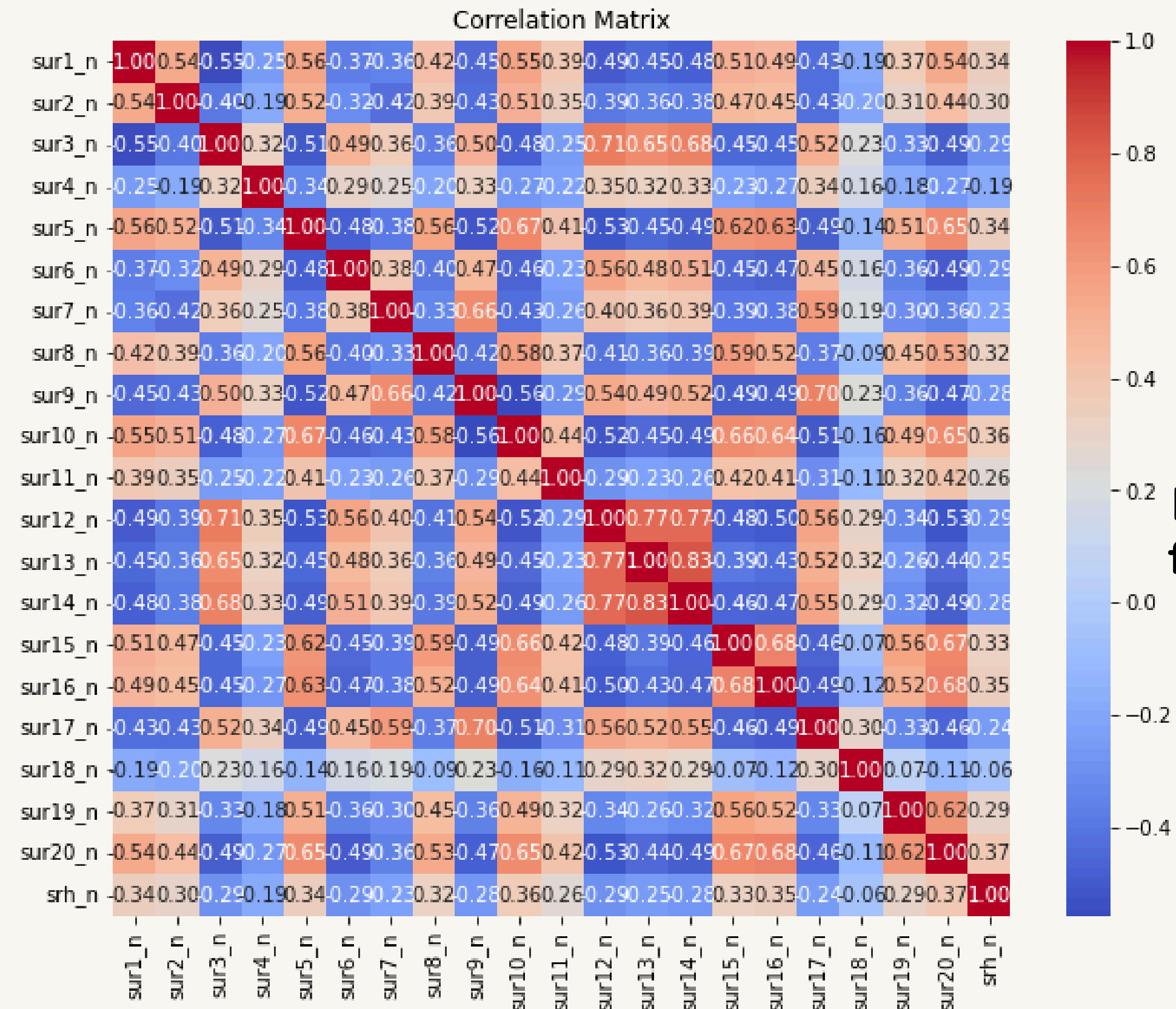
There is no significant relationship between individuals' emotional states and their overall well-being.

H1

There is a significant relationship between individuals' emotional states and their overall well-being.

**Statistical
Test**

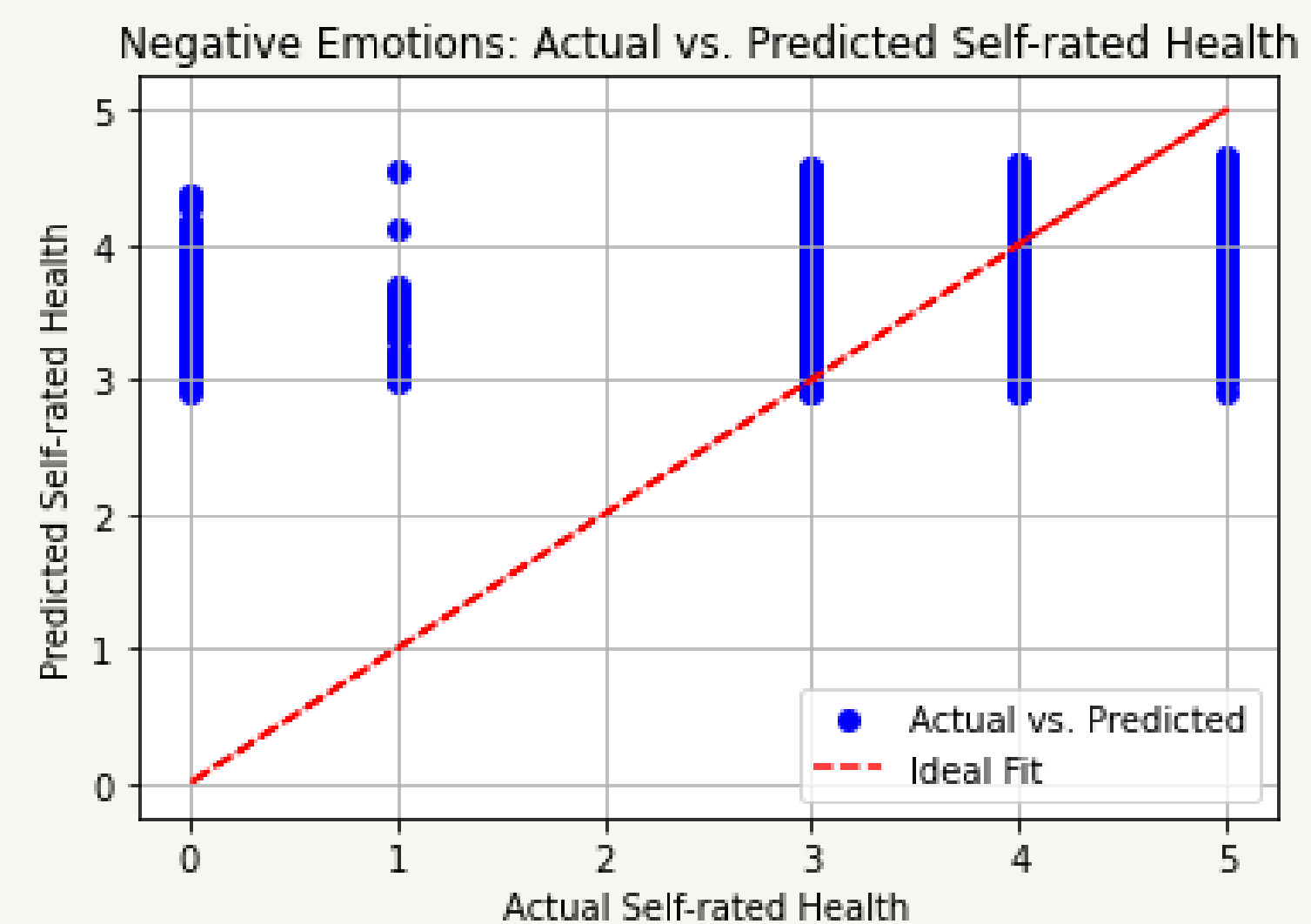
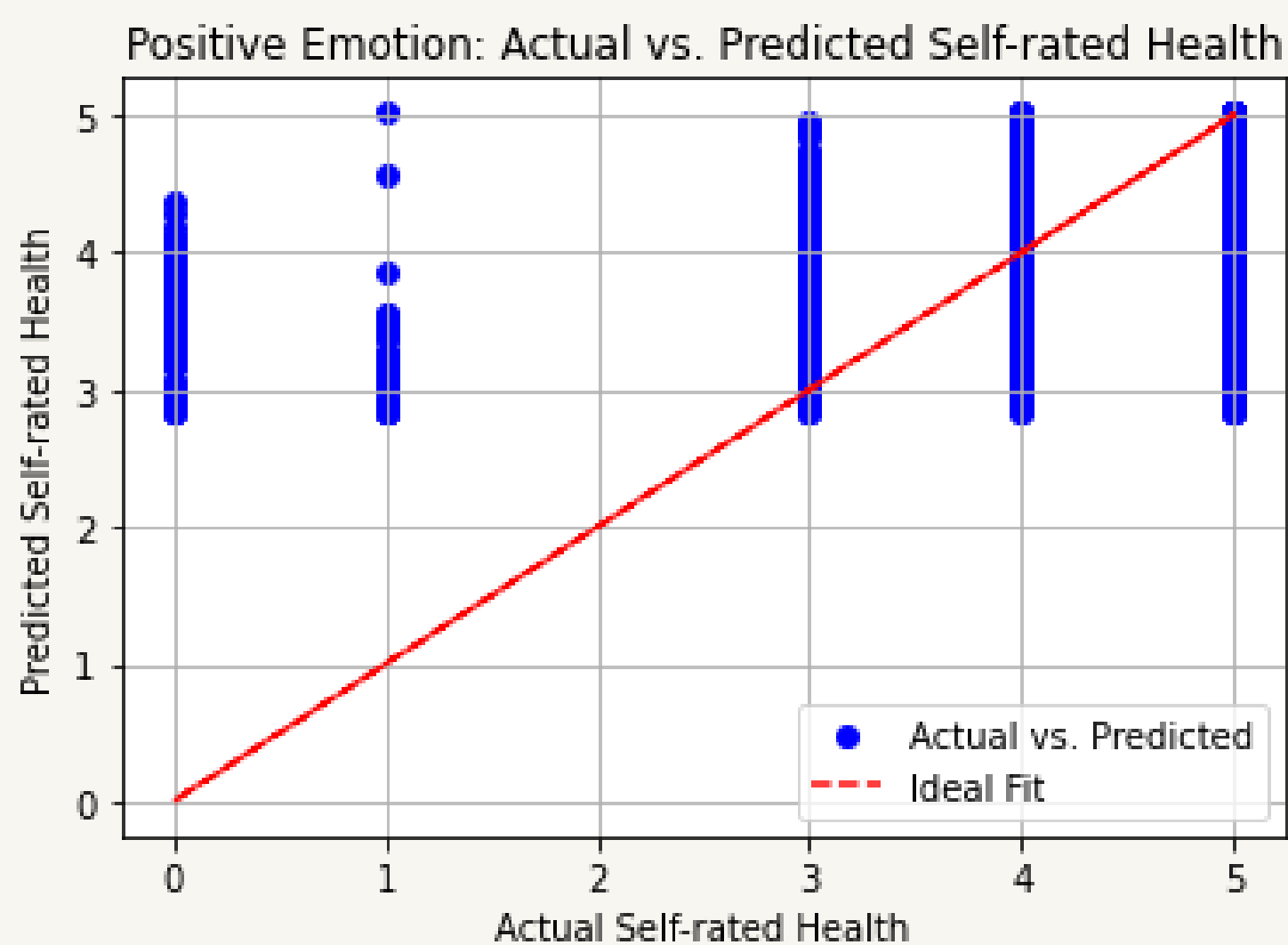
Spearman Correlation to find correlation between emotions of people with self rated health. We tried to find any linear relationship exists with help of **Multiple Linear Regression**.



Among positive emotions, the highest positive correlation with SRH is observed for sur20_n (feeling happy right now), followed by sur10_n (feeling comfortable right now) and sur16_n (being happy with the current situation).

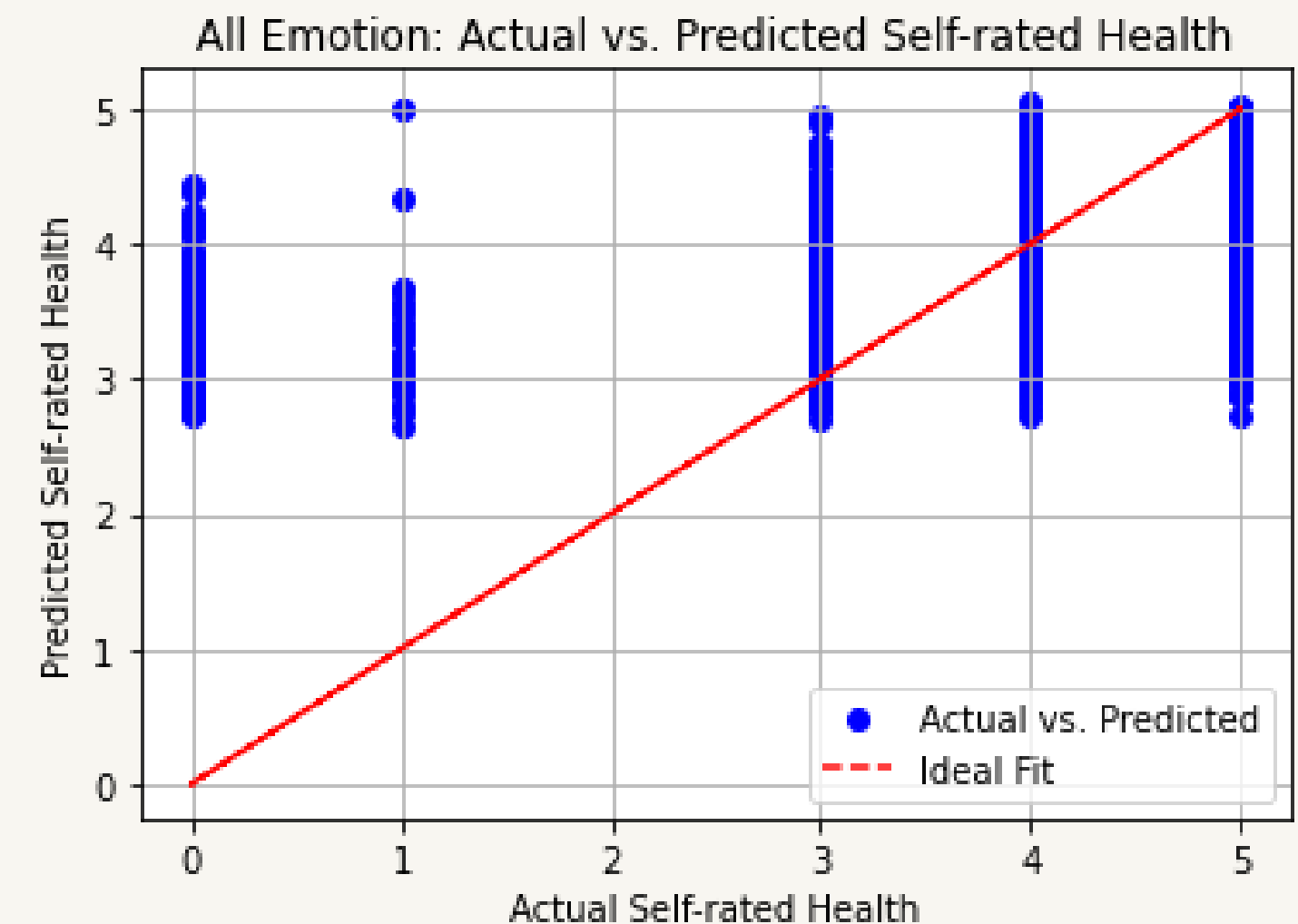
Conversely, among negative emotions, the highest negative correlation with SRH is observed for sur6_n (not having fun right now), followed by sur12_n (being in a bad temper right now) and sur3_n (feeling nervous right now).

The last column shows correlation between self rated helath and emotion of people



Positive Emotion:
R-squared value : 0.175 => 17.5% of variance in self rated health can be explained
p-value is very small, shows at least one of the predictor variable has non-zero effect

Negative Emotion
R-squared value : 0.13 => 13% of variance in self rated health can be explained
p-value is very small, shows at least one of the predictor variable has non-zero effect
All Emotion leads to 18.6% of variance in self rated health



Hypothesis 4

ANOVA Test Results:

F-statistic: 18.827939730472224

P-value: 2.722236623909921e-15

The post-hoc Tukey HSD test results provide further insights into the specific differences between groups

Post-hoc Tukey HSD Test Results:
Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
1	2	-0.1963	0.0349	-0.3838	-0.0088	True
1	3	-0.3448	0.0	-0.5328	-0.1568	True
1	4	-0.3897	0.0	-0.5775	-0.2019	True
1	5	-0.5507	0.0	-0.7501	-0.3512	True
2	3	-0.1485	0.0227	-0.2836	-0.0135	True
2	4	-0.1934	0.0009	-0.3282	-0.0586	True
2	5	-0.3544	0.0	-0.505	-0.2038	True
3	4	-0.0449	0.8954	-0.1803	0.0906	False
3	5	-0.2059	0.0019	-0.357	-0.0547	True
4	5	-0.161	0.0298	-0.3119	-0.0101	True

H0

There is no significant difference in the mean perception scores of the consequences of COVID-19 on personal security (dan) among individuals with different perceptions of the adequacy of state protection of individual and community security during the COVID-19 pandemic (gov).

H1

There is a significant difference in the mean perception scores of the consequences of COVID-19 on personal security (dan) among individuals with different perceptions of the adequacy of state protection of individual and community security during the COVID-19 pandemic (gov).

Statistical Test

One Way ANOVA is performed for comparing mean perception scores across multiple levels of the independent variable (gov) and determining if there are significant differences in perceptions of personal security among individuals with different perceptions of state protection adequacy during the COVID-19 pandemic.

Hypothesis 5

The data is filtered on the basis of the dates of wave 3 and wave 4 surveys.

Man-Whitney-U Results

Question	Result	Question	Result
0 sur1_n	Reject Null Hypothesis	0 sur3_n	Reject Null Hypothesis
1 sur2_n	Reject Null Hypothesis	1 sur4_n	Fail to Reject
2 sur5_n	Reject Null Hypothesis	2 sur6_n	Reject Null Hypothesis
3 sur8_n	Reject Null Hypothesis	3 sur7_n	Reject Null Hypothesis
4 sur10_n	Reject Null Hypothesis	4 sur9_n	Reject Null Hypothesis
5 sur11_n	Fail to Reject	5 sur12_n	Fail to Reject
6 sur15_n	Reject Null Hypothesis	6 sur13_n	Fail to Reject
7 sur16_n	Reject Null Hypothesis	7 sur14_n	Fail to Reject
8 sur19_n	Reject Null Hypothesis	8 sur17_n	Reject Null Hypothesis
9 sur20_n	Reject Null Hypothesis	9 sur18_n	Fail to Reject

There is significant differences in the distribution of emotions between wave 3 and wave 4 for most of the survey questions.

We can see the bar plots for mean of all the survey questions on next slide.

H0

There is no difference in the distribution of emotions between wave3 and wave4 of COVID-19

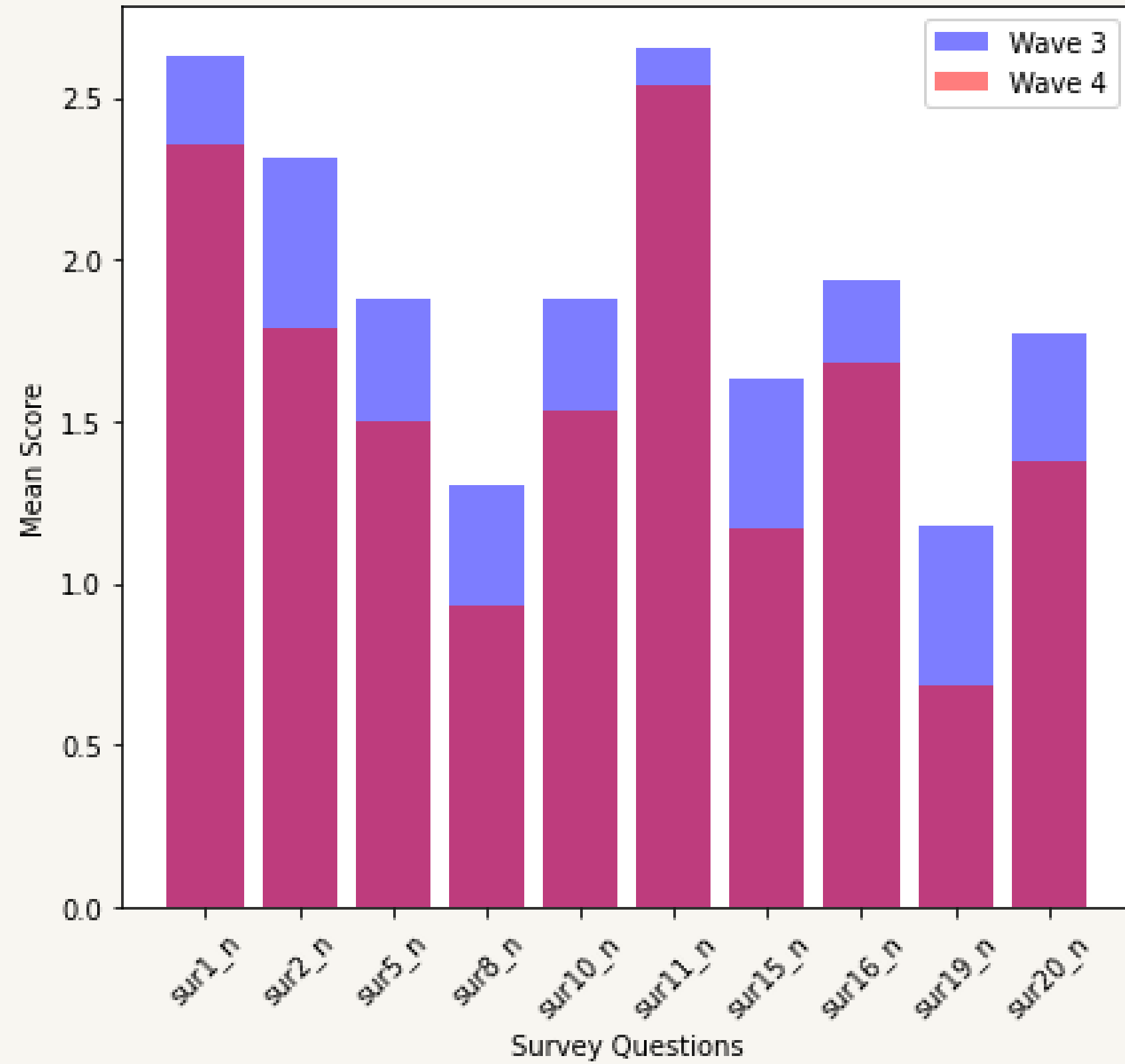
H1

There is a difference in the distribution of emotions between wave3 and wave4 of COVID-19

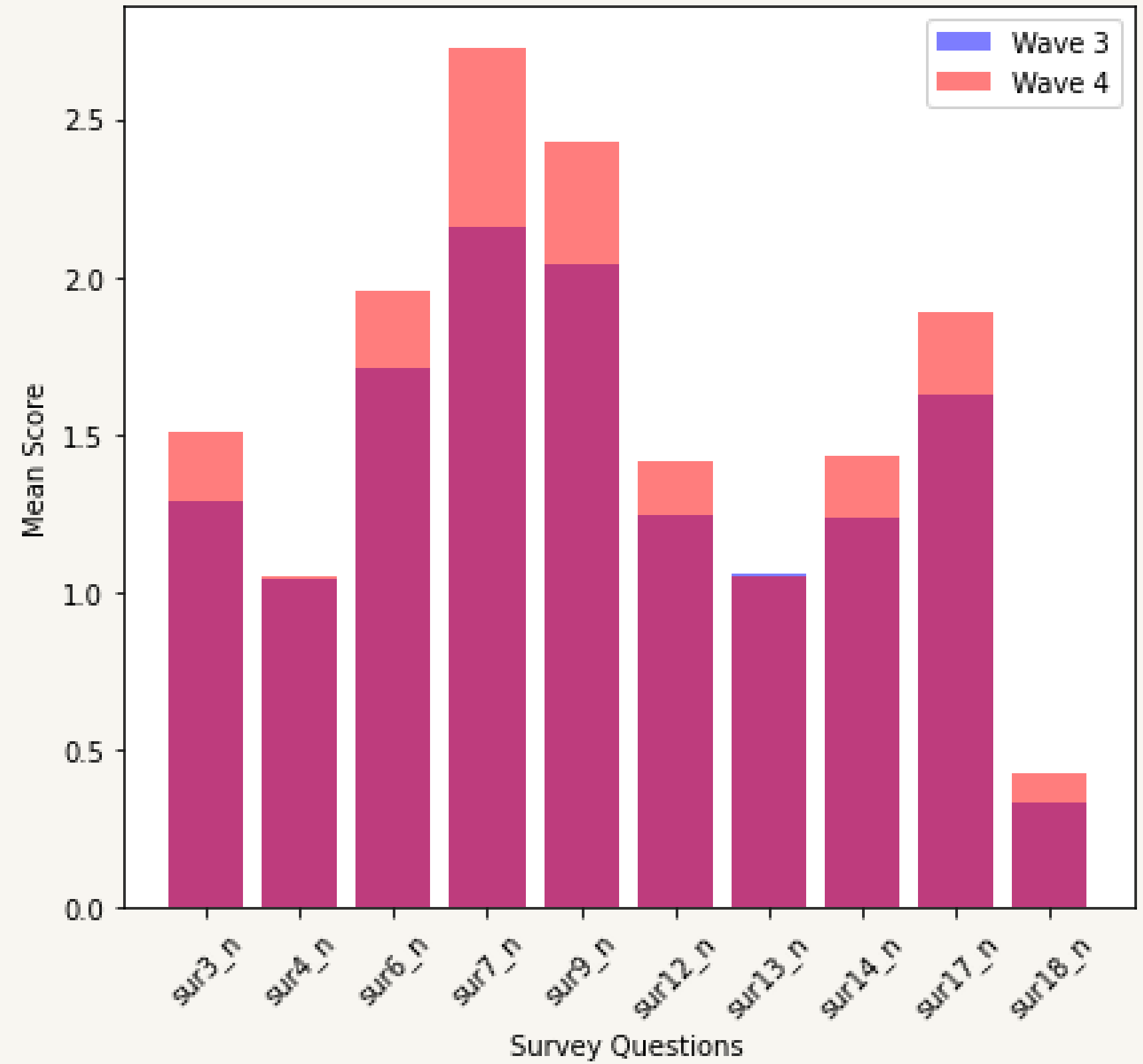
Statistical Test

First check the normality of all survey question using **Shapiro Normality Test**. The result of test is that no one is normally distributed. We perform **Man-Whitney-U Test** to compare between two waves.

Positive Emotions



Negative Emotions



Hypothesis 6

- sur21_n = I cannot cope with feeling distressed or sad.
- sur22_n = Other people can withstand feelings of distress or sadness better than I can
- sur23_n = Being distressed or sad is always a great ordeal for me.
- sur24_n = Feelings of being distressed or sad scare me
- sur25_n = I can do anything to not feel distressed or sad
- sur26_n = When I'm feeling distressed or sad, I can't help but focus on how bad the distress really feels.
- * sur27_n = I have to get rid of the uncomfortable feelings very quickly, otherwise I cannot bear them
- sur28_n = I cannot tolerate situations where I might feel sad
- sur29_n = I cannot stand uncomfortable feelings
- sur30_n = Being nervous scares me

	Distress Question	Statistic	P-value	
0	sur21_n	83.519971	5.392791e-18	Reject null hypothesis: distrib
1	sur22_n	2.566004	4.634806e-01	Fail to reject null hypothesis
2	sur23_n	10.607633	1.404826e-02	Reject null hypothesis: distrib
3	sur24_n	5.902730	1.164397e-01	Fail to reject null hypothesis
4	sur25_n	5.690808	1.276613e-01	Fail to reject null hypothesis
5	sur26_n	16.968178	7.174728e-04	Reject null hypothesis: distrib
6	sur27_n	7.697036	5.270614e-02	Fail to reject null hypothesis
7	sur28_n	3.426894	3.303682e-01	Fail to reject null hypothesis
8	sur29_n	64.615173	6.062791e-14	Reject null hypothesis: distrib
9	sur30_n	1.656040	6.467509e-01	Fail to reject null hypothesis

H0

There is no significant difference in distress tolerance levels across the four survey waves.

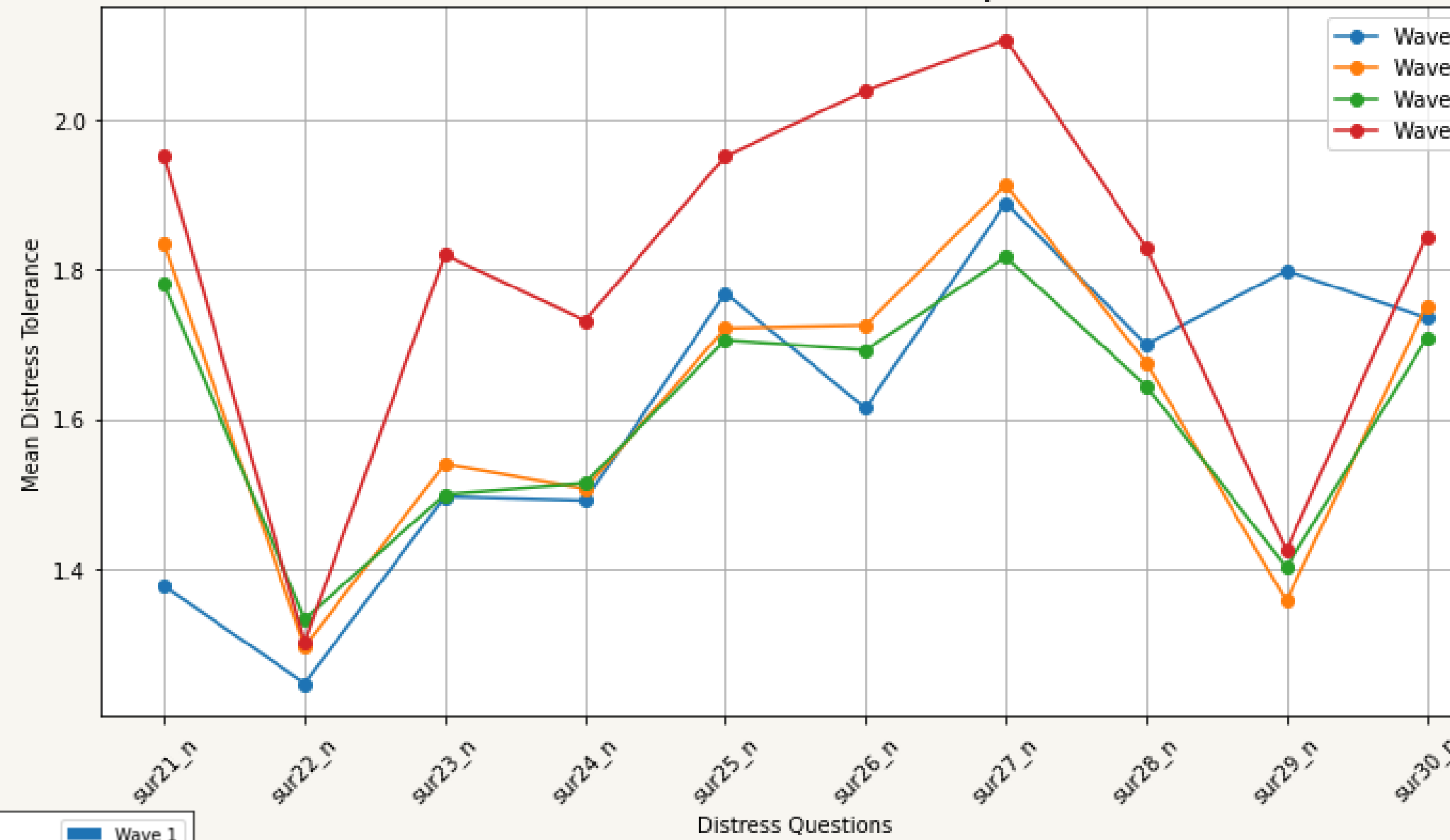
H1

There is a significant difference in distress tolerance levels across the four survey waves.

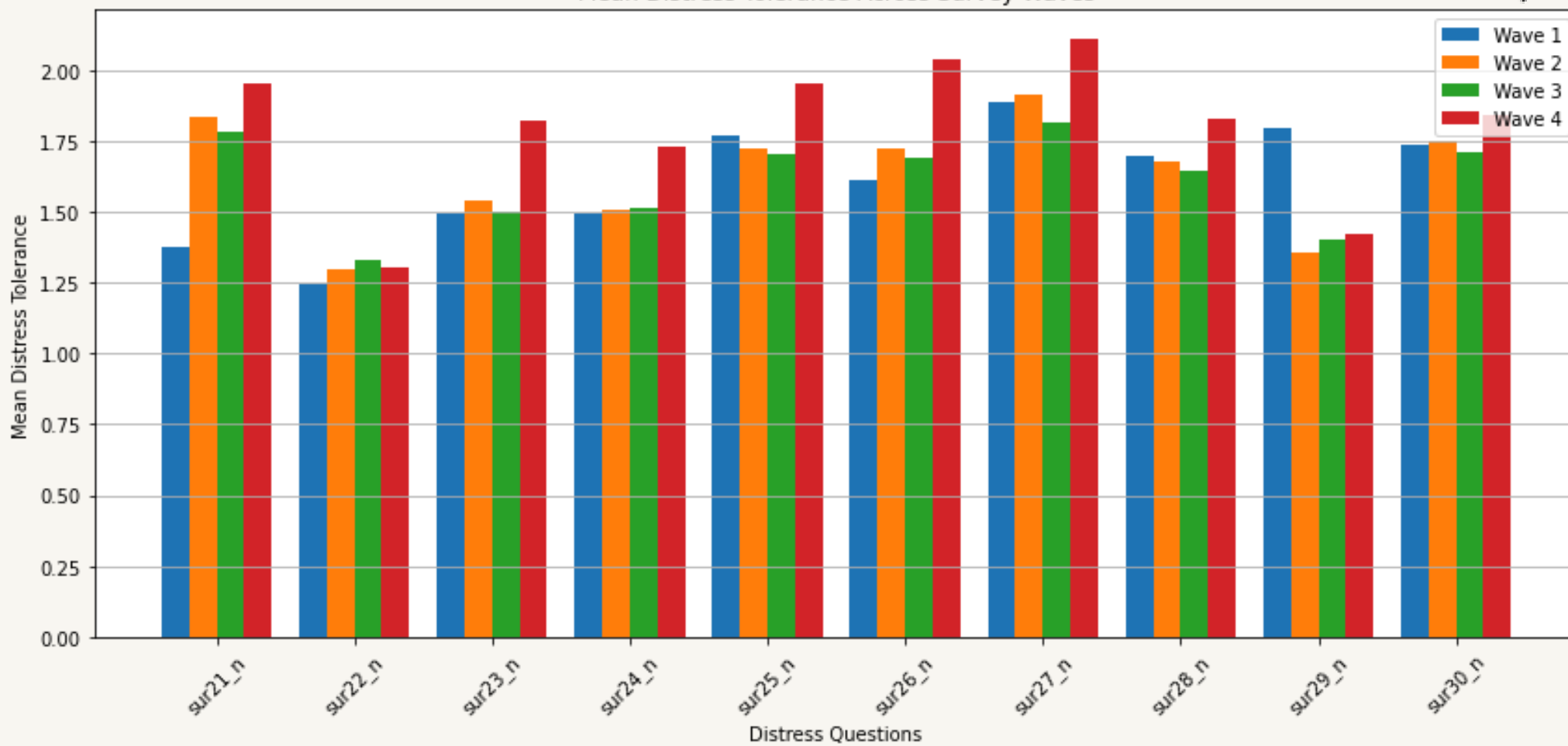
Statistical Test

The **shapiro normality test** results in no normal distribution. We perform **Kruskal-Wallis test** for all questions across all waves.

Mean Distress Tolerance Across Survey Waves

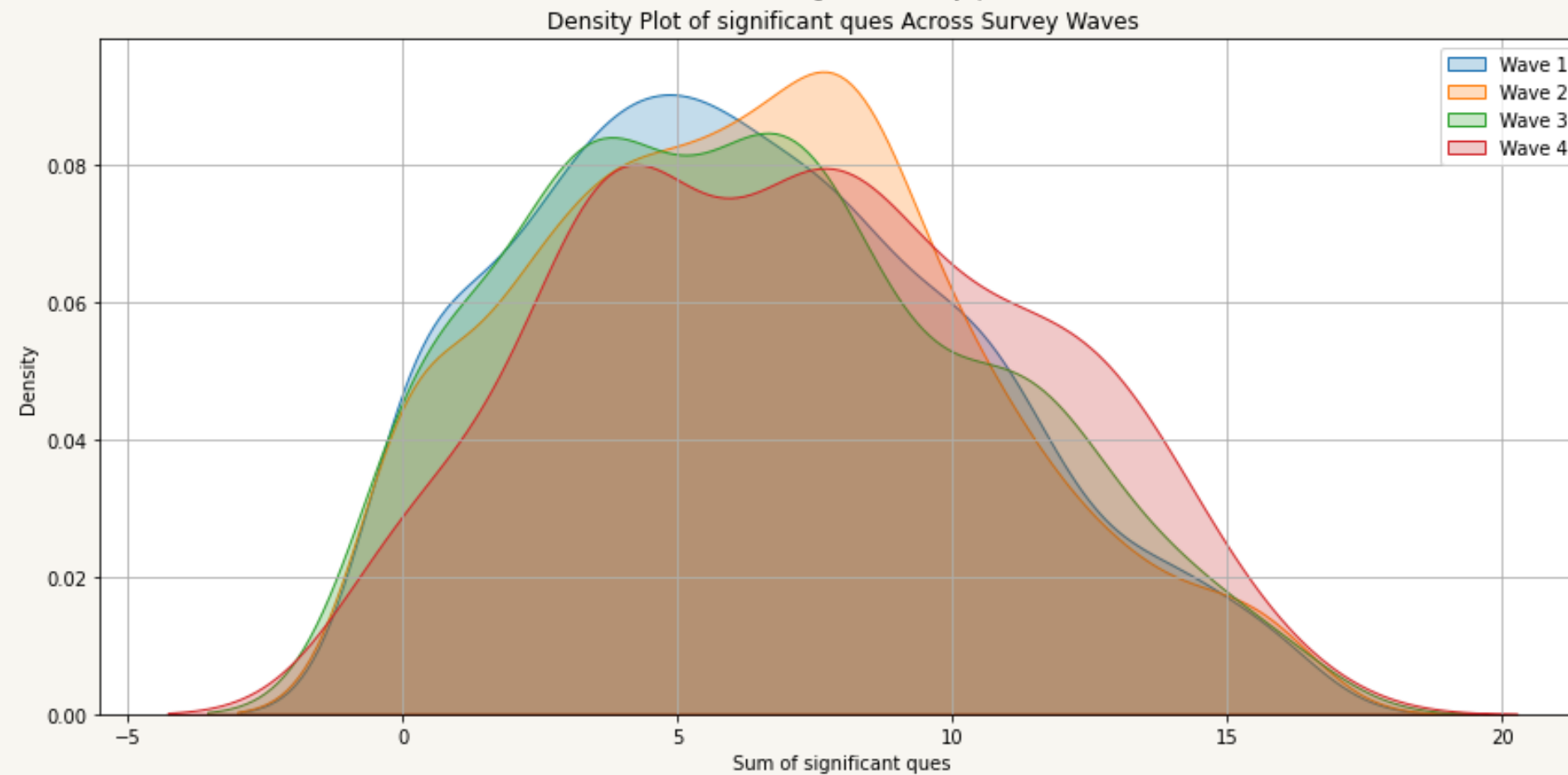
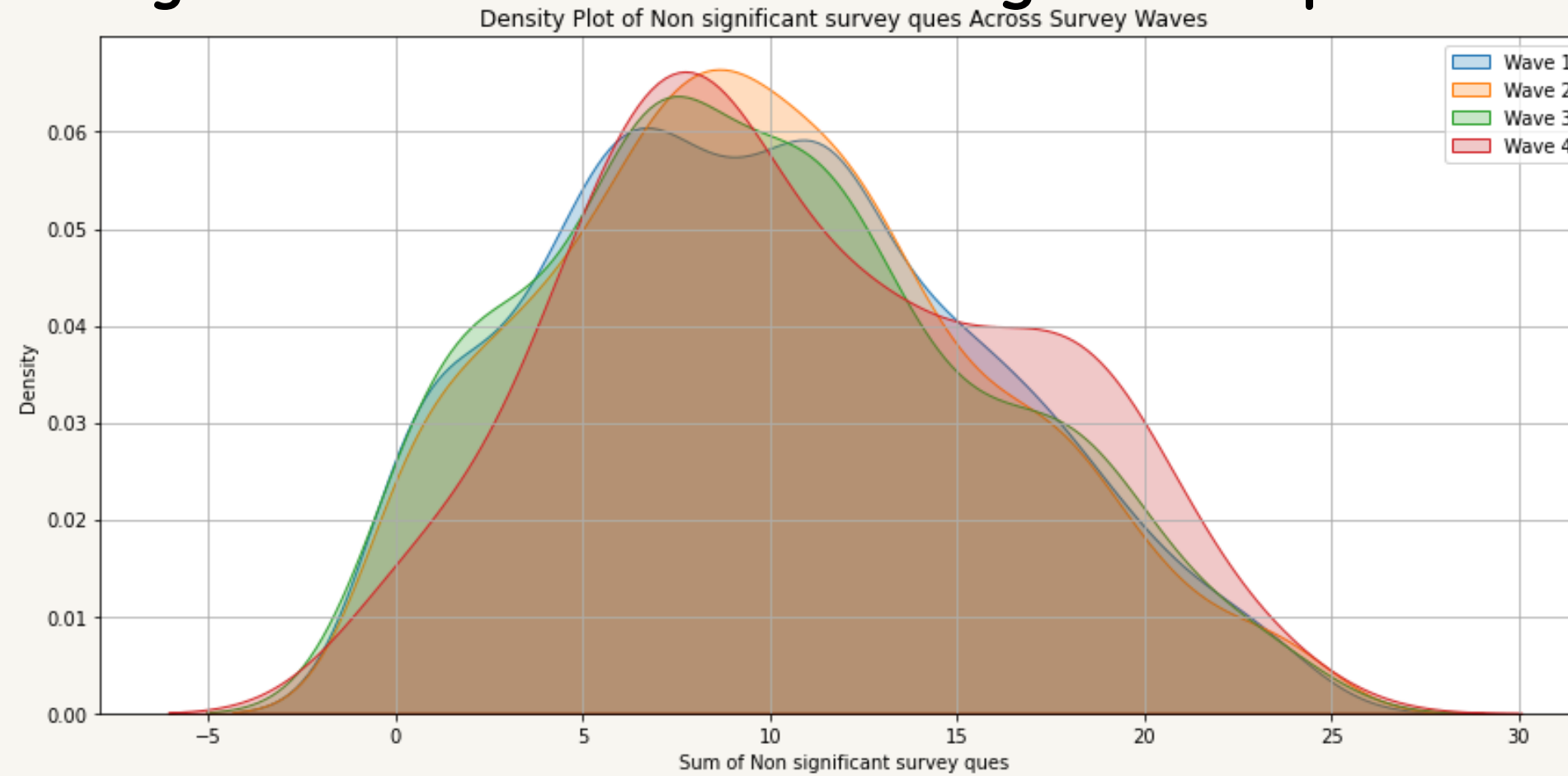


Mean Distress Tolerance Across Survey Waves



The survey questions , sur21, sur23, sur26, sur29 shows significant difference between different waves.

The below plot is density distribution of sum of all significant question and non significant question.



Hypothesis 7

From wave1 to wave2

Chi-square statistic: 84.56675485639165
p-value: 0.04357732210788379

From wave2 to wave3

Chi-square statistic: 80.21565452113265
p-value: 0.08298243945319692

From wave3 to wave4

Chi-square statistic: 39.43515939802765
p-value: 0.9933149021192091

There is a significant difference in the distribution of working situations between wave1 and wave2 but not from wave2 to wave3 and wave3 to wave4.

There might be issues in later waves as the data is not sufficient for waves3 and wave4.

H0

There is no significant difference in the distribution of working situations and sectors between Wave 1 to Wave 4.

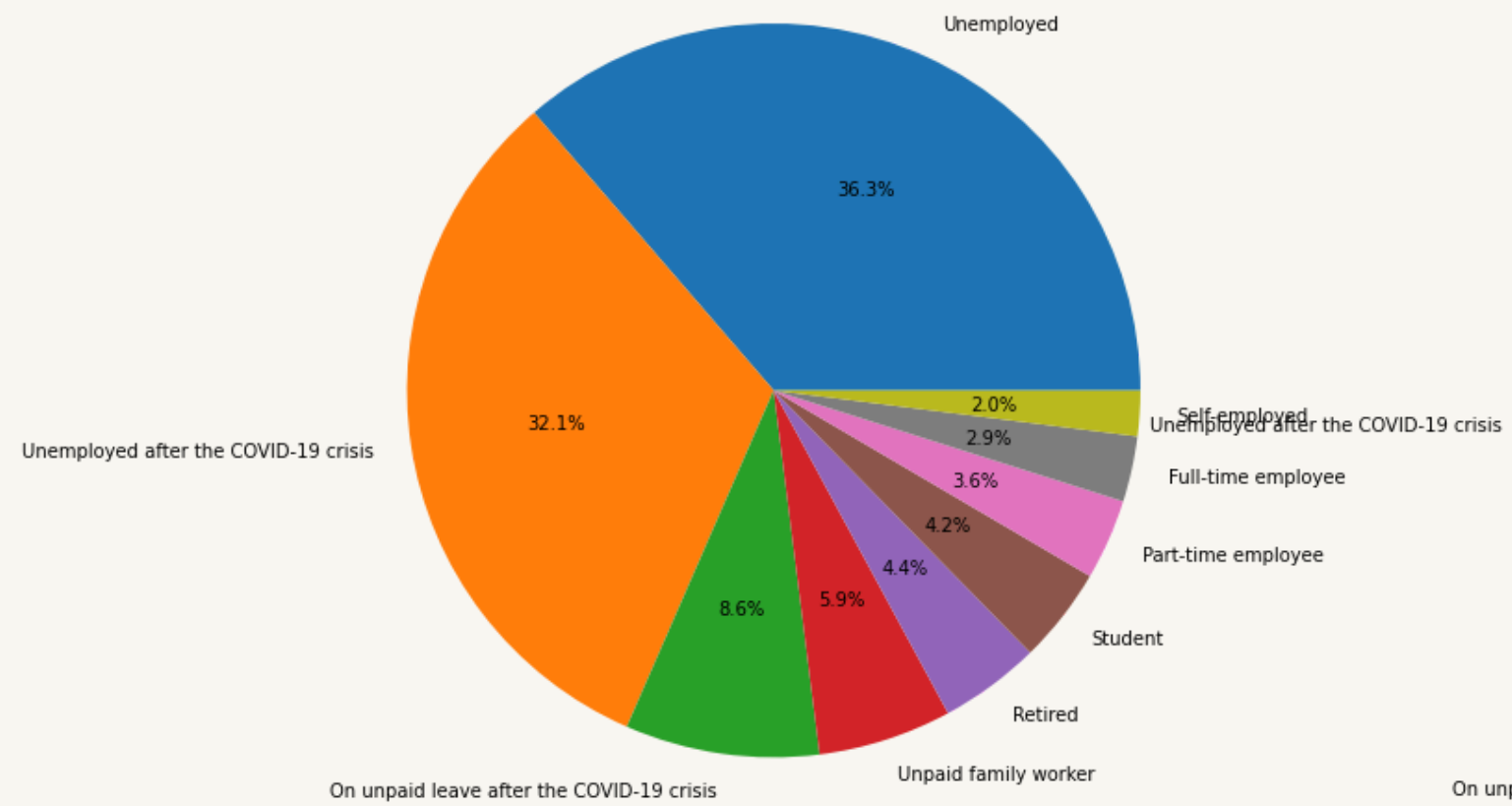
H1

There is a significant difference in the distribution of working situations and sectors between Wave 1 to Wave 4.

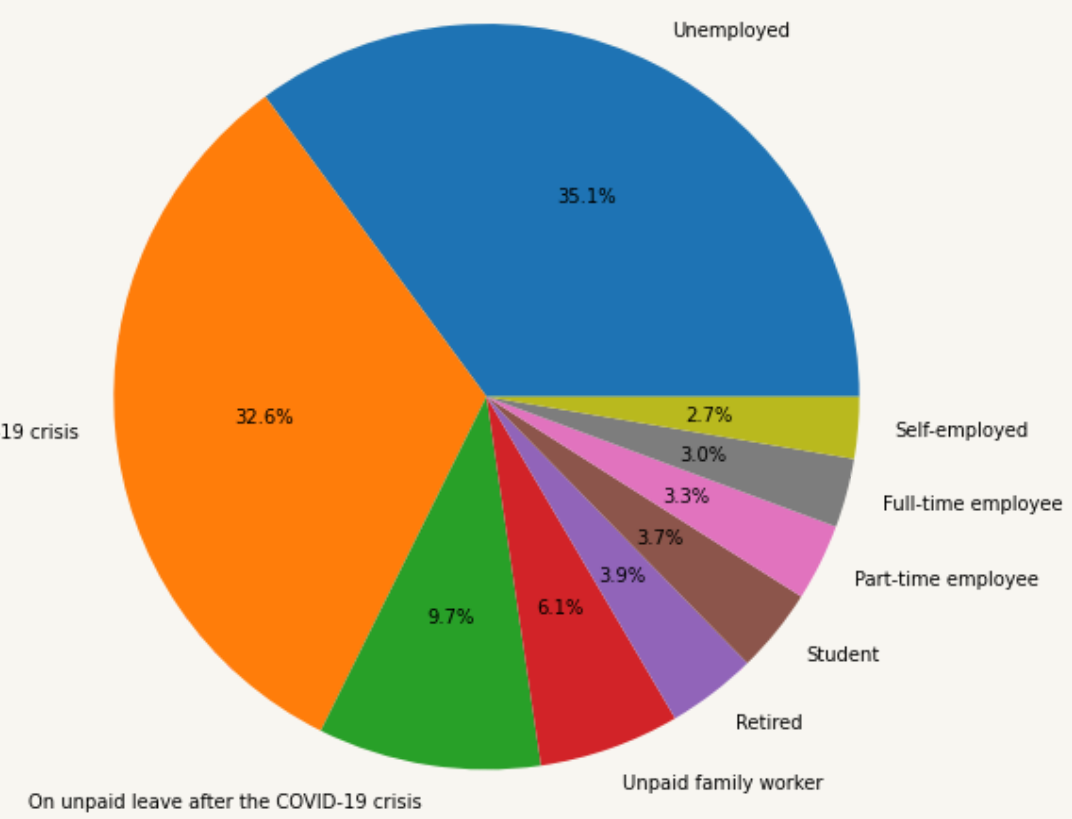
**Statistical
Test**

As the variable working status is a categorical variable, we can use **Chi Square Test**. We took wave1 & wave2, wave2 & wave3 and wave3 & wave4.

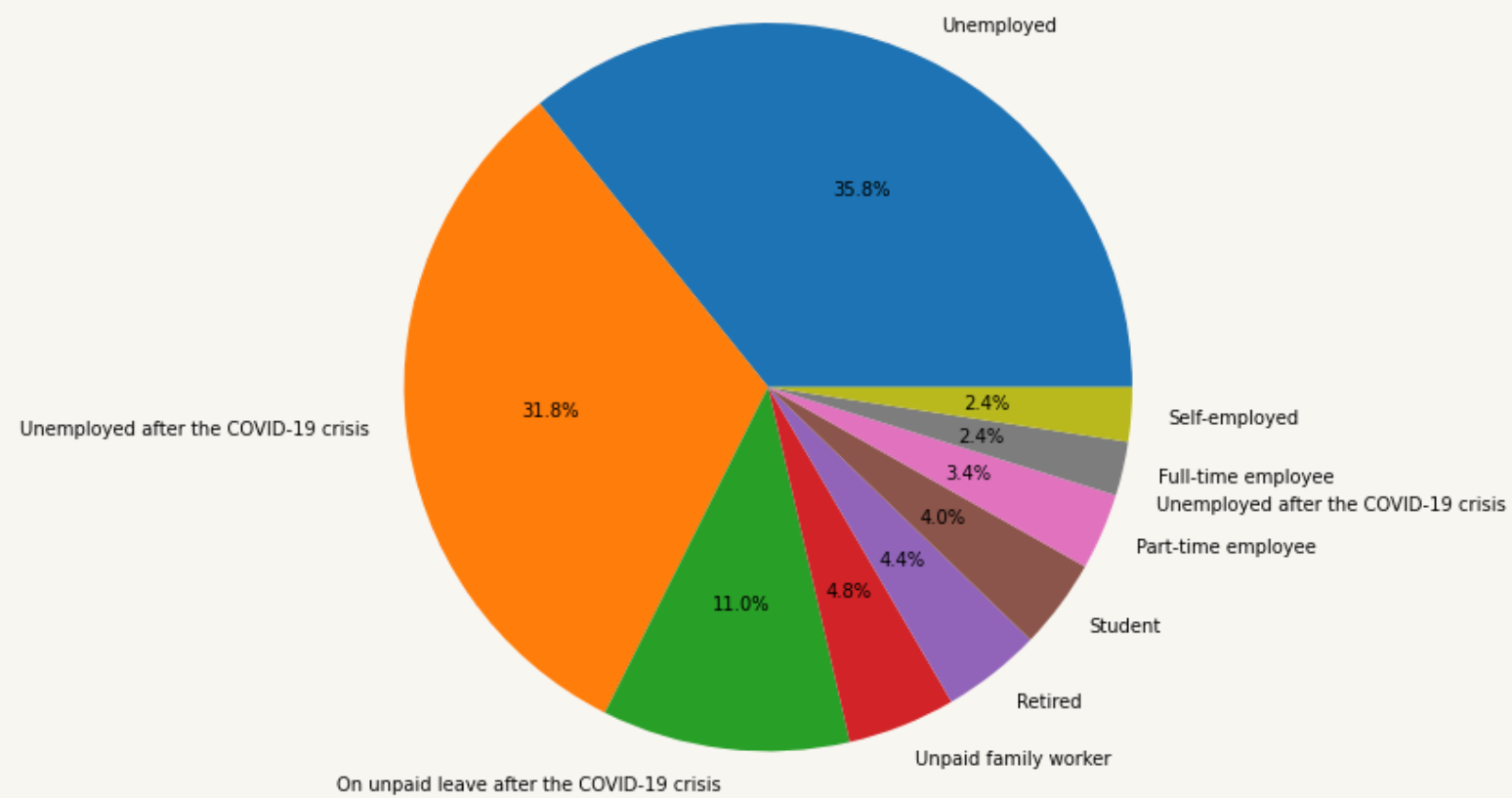
Wave1: Distribution of wor_n



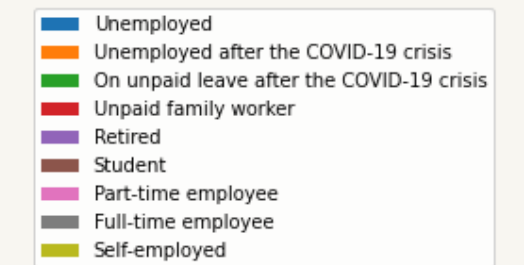
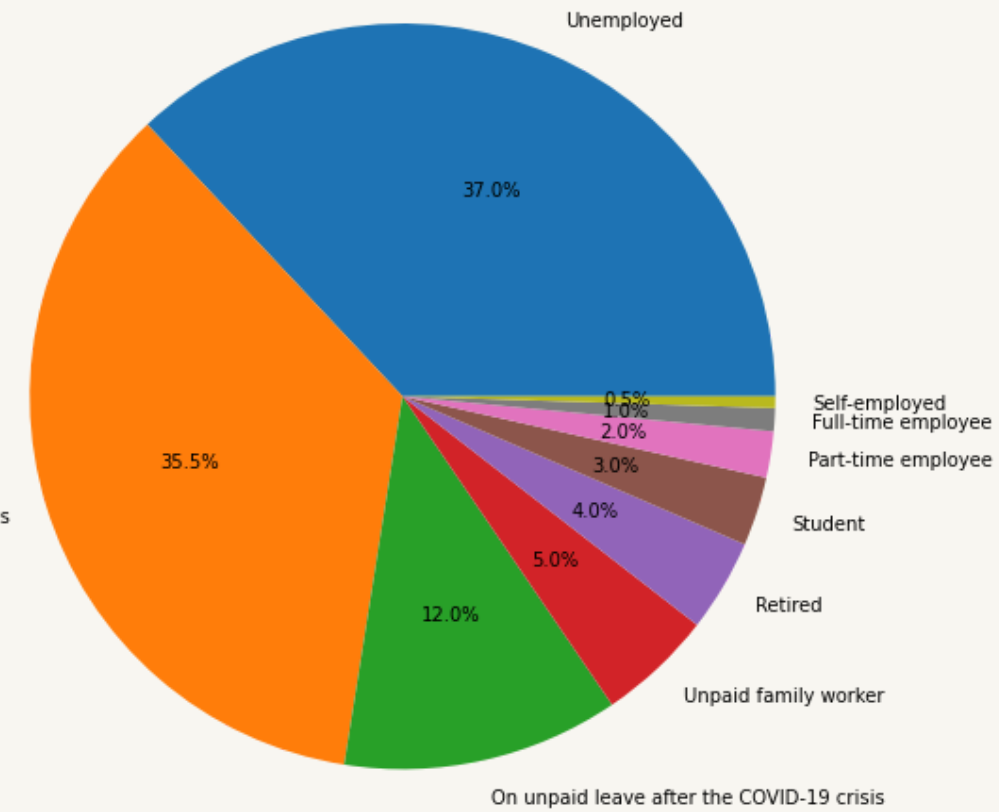
Wave2: Distribution of wor_n



Wave3: Distribution of wor_n



Wave4: Distribution of wor_n



Hypothesis 8

Mann-Whitney U Test Results:

Credit or Debit Card Usage:

Mann-Whitney U statistic: 27813.0

p-value: 0.6070684557853891

Online Spending Habits:

Mann-Whitney U statistic: 28824.5

p-value: 0.8543931134666313

The p-values are greater than the typical significance level of 0.05, indicating that there is not enough evidence to reject the null hypothesis. Therefore, based on these results, we cannot conclude that there is a significant difference in usage between urban and rural populations for either credit/debit card usage or online spending habits.

H0

There is no significant difference in the usage of credit/debit cards and online spending habits between urban and rural people.

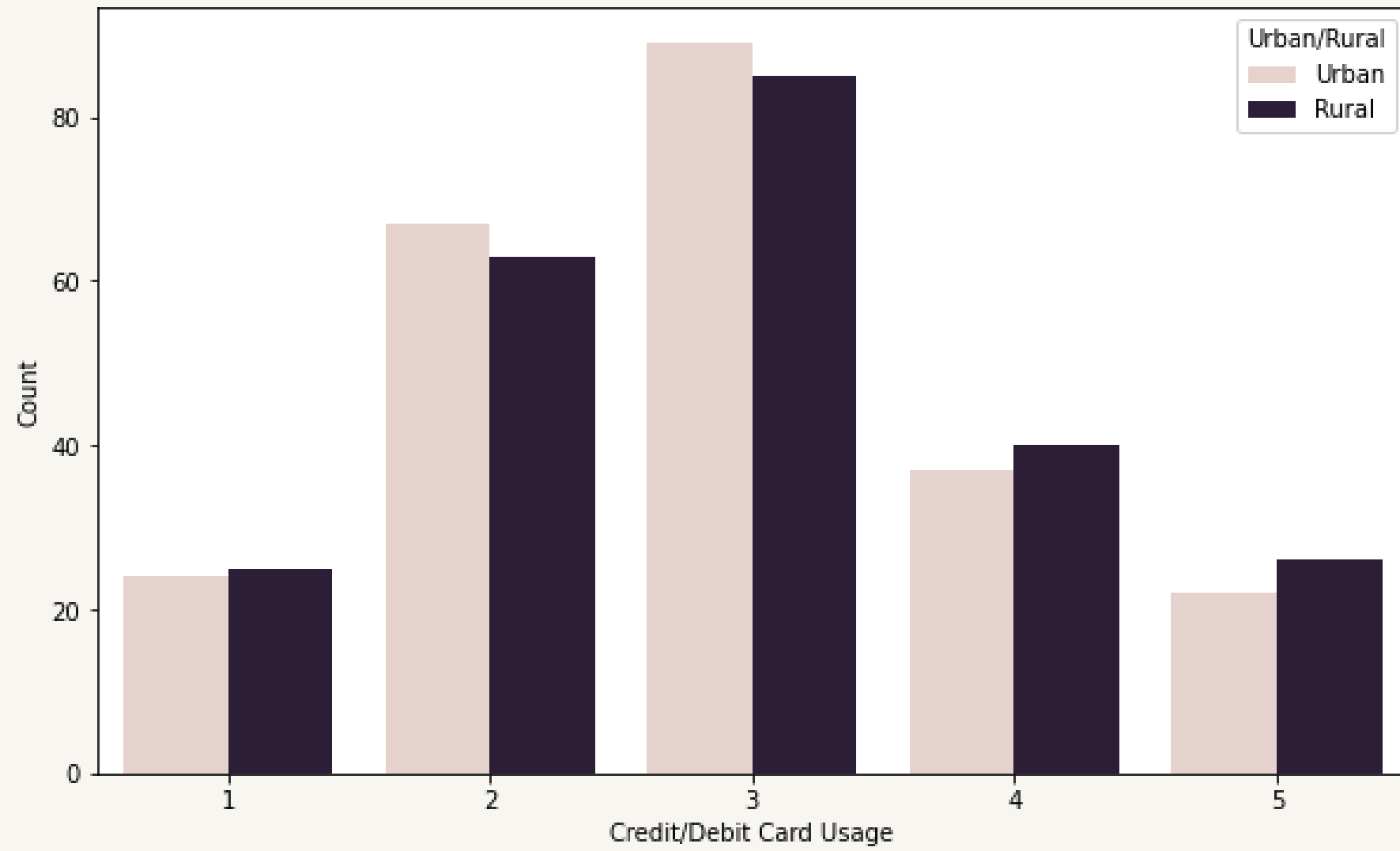
H1

The usage of credit/debit cards and online spending habits has increased more for urban people than rural people.

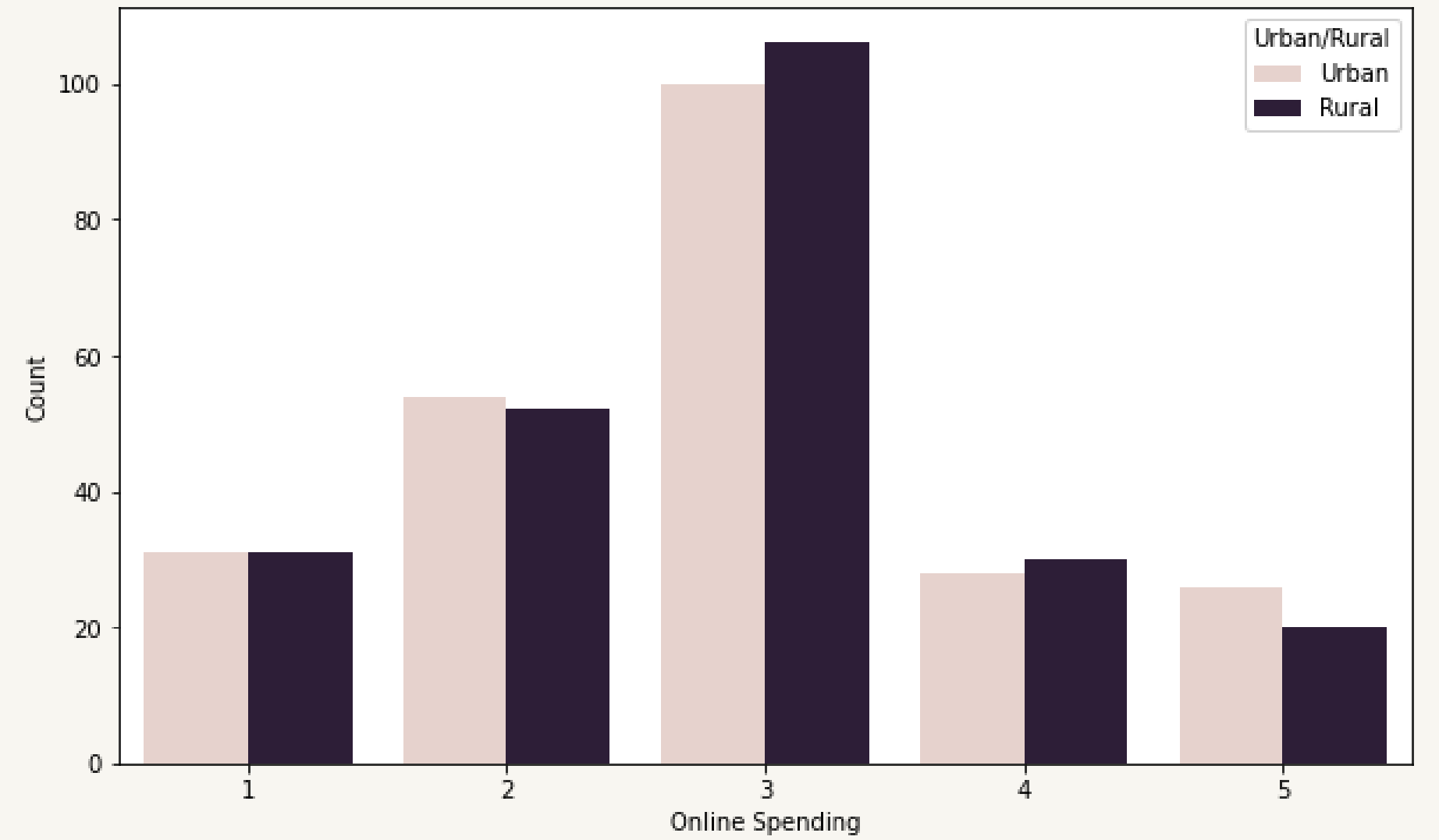
**Statistical
Test**

Since we are comparing two independent groups (urban vs. rural) and the variables are ordinal, we can use the **Mann-Whitney U test** to determine if there is a significant difference in the median usage between urban and rural groups.

Credit/Debit Card Usage by Urban/Rural Areas



Online Spending by Urban/Rural Areas





Reference

Sari E, Kağan G, Karakuş BŞ, Özdemir Ö. Dataset on social and psychological effects of COVID-19 pandemic in Turkey. Sci Data. 2022 Jul 23;9(1):441. doi: 10.1038/s41597-022-01563-4. PMID: 35871225; PMCID: PMC9308400.