

Report on Time Use Survey

Tutor: Juho Heimonen

Course Name: Statistical Data Analysis

Course Code: TKO_7093-3004

Group Number: Project_013

Presented by:

Nouman Bashir (2507543)

Dominic Amoateng Sabeng (2510400)

Saif Ur Rehman (2414361)

Submission date:

27th October, 2025

University of Turku, Department of Computing

Introduction

The Time Use Survey in Finland examined how Finnish individuals allocate time to activities such as working, sleeping, reading, dining in restaurants, and visiting libraries. The survey aimed to gather insights into the habits of individuals and households across Finland. Data were collected from respondents in cities, municipalities, and rural areas using an online data collection tool. This report seeks to:

- a. Characterise respondents using demographic and activity variables
- b. Perform clustering analysis on the activity variables (working, sleeping and reading)
- c. Perform PCA on the activity variables (working, sleeping and reading)
- d. Estimate the average daily time Finnish households spend on activity variables (working, sleeping and reading)
- e. Determine differences in the average daily time spent on the activity variables based on the living environment in Finland
- f. Determine the difference in the average daily time spent on the activity variables based on the day of the week Finnish people perform the activity
- g. Determine the difference in the average daily time spent on the activity variables based on the sex of respondents.
- h. Determine the association between the demographic variables (sex, age group, living environment, day of week) and the activity variables (dining at the restaurant and visiting the library)
- i. Analyse the relationships among the activity variables (working, sleeping and reading) in Finland

Methods

The data for this report were sourced from Statistics Finland's dataset, "Teaching Use Data of the Time Use Survey," licensed under the Creative Commons Attribution 4.0 International license. The survey included 780 respondents. This secondary dataset comprised 11 variables: 6 demographic variables (household ID, member ID, day of the week, sex, age group, living environment) and 5 activity variables (working, sleeping, reading, dining at restaurants, visiting libraries). The activity variables covered activities performed by respondents over the past 12 months. Working, sleeping, and reading were measured in minutes, while dining at restaurants and

visiting libraries were recorded as binary (yes/no) responses, indicating whether respondents engaged in these activities.

Data Preparation

a. Data type

The data had both categorical and numerical variables. Table 1 describes the data type of variables in the dataset.

Table 1: Data type of variables in the dataset

Code	Variable name	Data type
khode	Household ID	Numerical variable
jasen	Member ID (within household)	Numerical variable
pvknro	Day of week	Binary (categorical variable)
sp	Sex	Binary (categorical variable)
IKAL1	Age group	Categorical variable
ASALUE	Living environment	Categorical variable
A1 (minutes)	Working	Numerical variable
A2 (minutes)	Sleeping	Numerical variable
A3 (minutes)	Reading	Numerical variable
A4	Dining at restaurant	Binary (categorical) variable
A5	Visiting library	Binary (categorical) variable

b. Data Cleaning

Initial characterization of the dataset revealed inconsistencies in the activity variables. Some data points were entered in “hours and minutes” format, as “?”, as minutes, or as zeros (e.g., 10:30, “?”, 120, 0). Data points in hours and minutes were converted to minutes for all activity variables. For the activity variables “dining at restaurants” and “visiting libraries,” “?” entries were replaced with the mode of the respective variable’s data points. Minutes entries were assigned “yes” to indicate the activity was performed, with the rationale that the activity occurred but the duration was rather recorded instead of indicating yes. For the sleeping variable, 10 “?” entries were imputed using the median of the data points, stratified by the demographic variables’ “sex” and “age group,” as sleep duration was found to vary by age group. The median imputation was done since everyone sleeps, hence “?” entries could not be treated as

missing values or discarded entirely from the dataset. For the working and reading variables, “?” entries were replaced with zeros to indicate that respondents did not engage in these activities. Zero minutes for working were assumed to represent students or retired individuals who do not work, while zeros for reading were assigned to individuals who do not read.

Descriptive Statistics

Baseline descriptive statistics were calculated for all variables in the dataset. Table 1 summarizes the baseline characteristics of respondents. Of the 780 respondents, 406 were male (52.05%), and 374 were female (47.95%). Most respondents were aged 45-54 years (21.79%), followed by 55-64 years (21.15%). Additionally, 4.35% were 24 years or younger, 13.59% were 25-34 years, 14.35% were 65-74 years, and 7.82% were 75 years or older. Furthermore, 418 respondents (53.59%) reported working from Monday to Friday. Regarding living environment, 519 respondents (66.53%) lived in cities, 124 (15.89%) in municipalities, and 137 (17.56%) in rural areas.

Interestingly, the data revealed that about 53.6% of the Finnish people dined out at the restaurants while 68.3% of them visited the library. However, it was revealed the number of people who visit the library but do not read was about 7.89%. The average daily time spent by Finnish was estimated to be 77.98 minutes for working, 529.17 for sleeping and 56.65 for reading. These values were believed not to be a true reflection of these activities due to some outliers in the data. Moreover, the data points from the active variables were not normally distributed hence using the mean in this context was deemed problematic and hence not trustworthy.

Table 2: Baseline descriptive statistics of the variables

Demographic variables	Statistic
Sex – no. (%)	
Male	374 (47.95)
Female	406 (52.05)
Age group (years) – no. (%)	
20 – 24	34 (04.35)
25 – 34	106 (13.59)
35 – 44	132 (16.92)
45 – 54	170 (21.79)

55 – 64	165 (21.15)
65 – 74	112 (14.35)
75 and above	61 (07.82)
Day of week – no. (%)	
Working day	418 (53.59)
Weekend	362 (46.41)
Living environment – no. (%)	
City	519 (66.53)
Municipality	124 (15.89)
Rural area	137 (17.56)
Activity Variables	
Working - mean (s.d)	77.98 (160.34)
Sleeping - mean (s.d)	529.17 (103.29)
Reading - mean (s.d)	56.65 (75.46)
Dining at a restaurant – no. (%)	
Yes	418 (53.59)
No	362 (46.41)
Visiting the library – no. (%)	
Yes	533 (62.95)
No	247 (37.05)

It can be observed from Figure 1 that most of the respondents (male and female) live in the city whereas there is an almost even distribution of male and female residents Finnish people in both municipalities and rural areas.

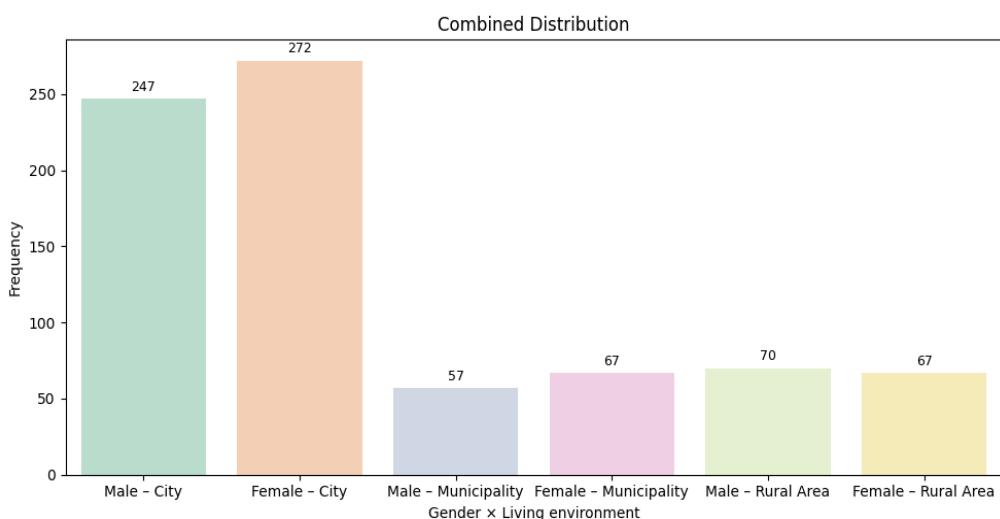


Figure 1: Distribution of Respondent by sex and living environment

Cluster Analysis of the Activity Variables:

The activity variables (working, sleeping, reading) were subjected to cluster analysis to group the Finnish population based on time allocation patterns. Table 3 and Figure 2 present the results of clustering and the average minutes spent on these activities. The analysis identified two distinct clusters, as illustrated in Figure 2. From Table 3, individuals in Cluster 1 allocate a substantial amount of time to working, whereas those in Cluster 2 spend little to no time working. On average, Cluster 2 individuals spend more time sleeping and reading compared to their Cluster 1.

Generally, Cluster 1 represents a working group characterized by high working hours and lower time spent sleeping and reading whereas, Cluster 2 represents a non-working and leisure group, characterized by minimal or no work hours, increased sleep, and greater time devoted to reading. Cluster 2 we believe might be students, the retired and old aged Finnish population not working and having more time for sleep and reading.

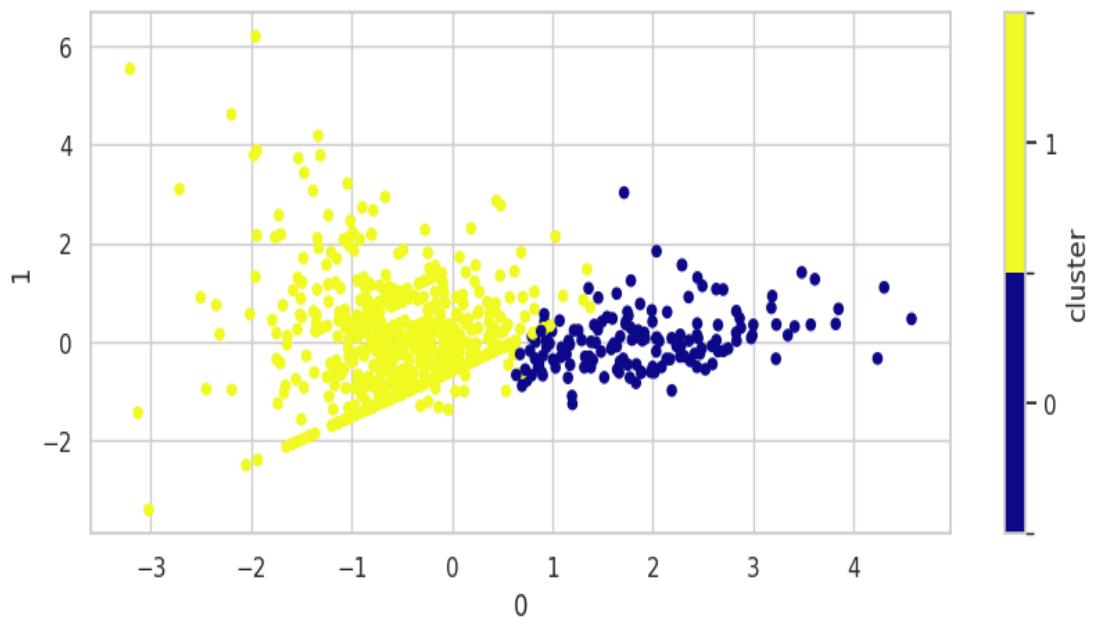


Figure 2: Clustering analysis based on the activity variables

Table 3: Average time spent on activity variables based on clusters

Activity variables	Cluster 1	Cluster 2
Working	345 Minutes (~6 hours)	0 Minutes
Sleeping	460 Minutes (~7.5 hours)	540 Minutes (9 hours)
Reading	10 Minutes	40 Minutes

Principal component analysis on Activity variables:

Principal component analysis was conducted on the activity variables: working, reading and sleeping. From the PCA, it was revealed that about 47% of the total variation in the data is explained by the first principal component, 33% variation is explained by the second component while the third component explains 21% of the variation. Table 4 presents the PCA scores. From Table 4, it can be noted that for the first component working and sleeping are quite prominent than reading though sleeping relates negatively to working. Furthermore, from the second component, reading and sleeping are prominent and also relates opposite or negatively to one another.

Table 4: PCA scores for activity variables

Components	Working	Sleeping	Reading
0	0.705129	-0.589727	-0.393720
1	0.004743	-0.551321	0.834280
2	0.709063	0.590142	0.385955

Estimation of the average daily time Finnish households spend on activity variables:

Analysis from the data cleaning showed that some Finnish were not working and not reading at all. Table 5 present the information on Finnish people that were not working and reading as well as those that were reading. From Table 5, majority of the Finnish population were not working with only a quarter of them working whereas 65.5% of them were people who read with the remaining 34.5% not reading.

Table 5: Distribution of Respondents based working and reading

Activity variables	Frequency	Percentage
Working		
Not Working	595	75.0
Working	195	25.0
Reading		
Not reading	269	34.5
Reading	511	65.5

The estimated daily average of the Finnish people who were working and reading was based on the sub population who were working and reading using the median score. The median was used due to outliers found in the data. Table 6 present the average estimated time of activity variables undertaken by the Finnish people. From Table 6, Finnish working people work on the average 225 minutes making 4.25 hours whereas the Finnish reading people also read on the average 60 minutes (1 hour). For sleeping, Finnish people sleep for 530 minutes (8.83 hours).

Table 6: Average Time of activity variable using central tendency median

Activity variables	Average time in minutes	Average time in hours
Working	255	4.25
Sleeping	530	8.83
Reading	60	1.0

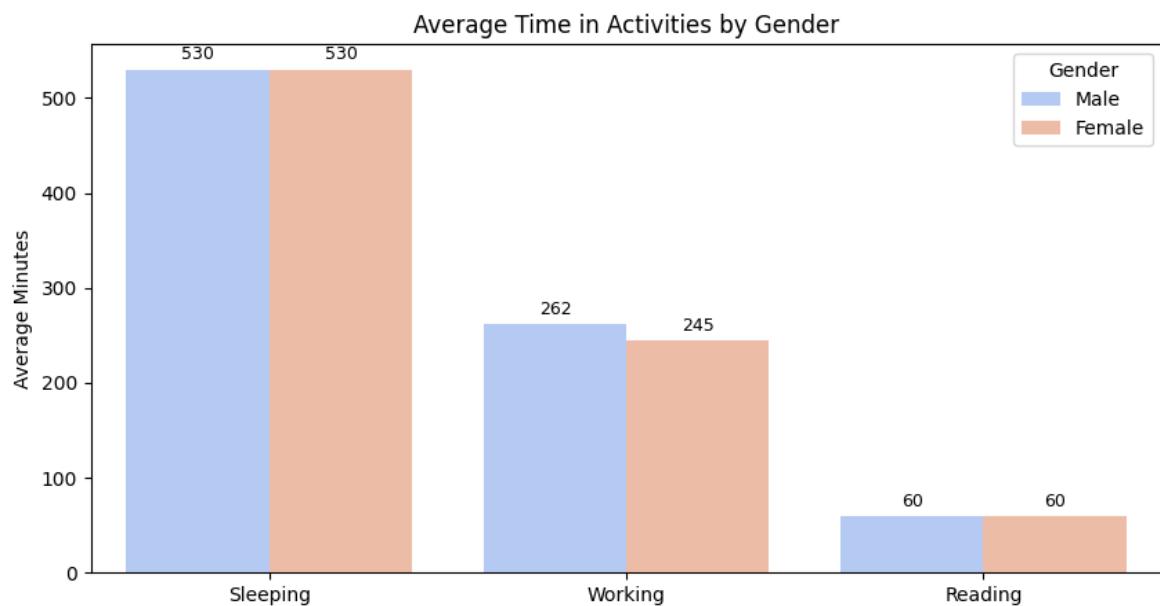


Figure 3: Distribution of activity variables based on Gender

Figure 3 illustrates that on average sleeping and reading variables have the same Average irrespective of the Gender, but we see a difference in working for gender as male tends to work more than female.

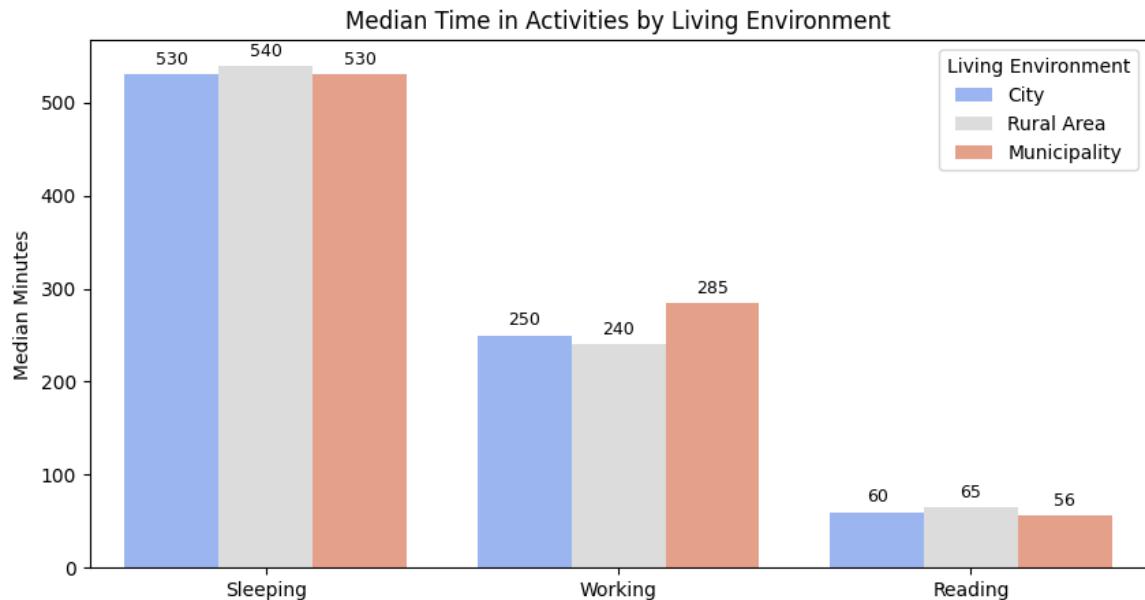


Figure 4: Distribution of activity variables based on living environment

Figure 4 illustrates that people work more in municipality and work less in rural areas. In reading and sleeping the rural area has spent more time on average.

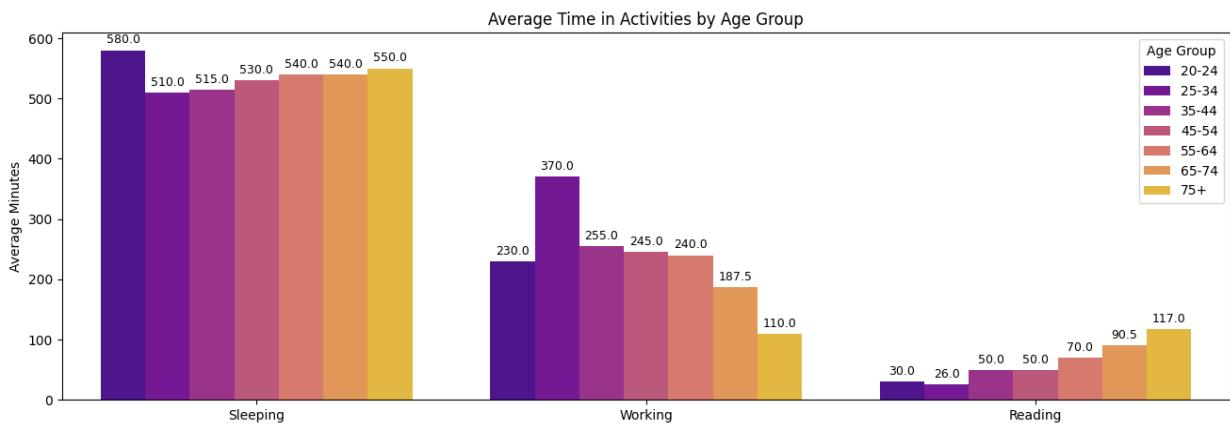


Figure 5: Distribution of activity variable based on age groups

From Figure 5, We see there is a trend in Reading as the age passes, people read more and the oldest ones read around **2** hours on average. For Working, the age group 25-34 works more than all the other groups for almost **6** hours on average and the

trend declines after that for every age group. For Sleeping, The younger ones (20-24 years) sleep for almost **10** hours on average.

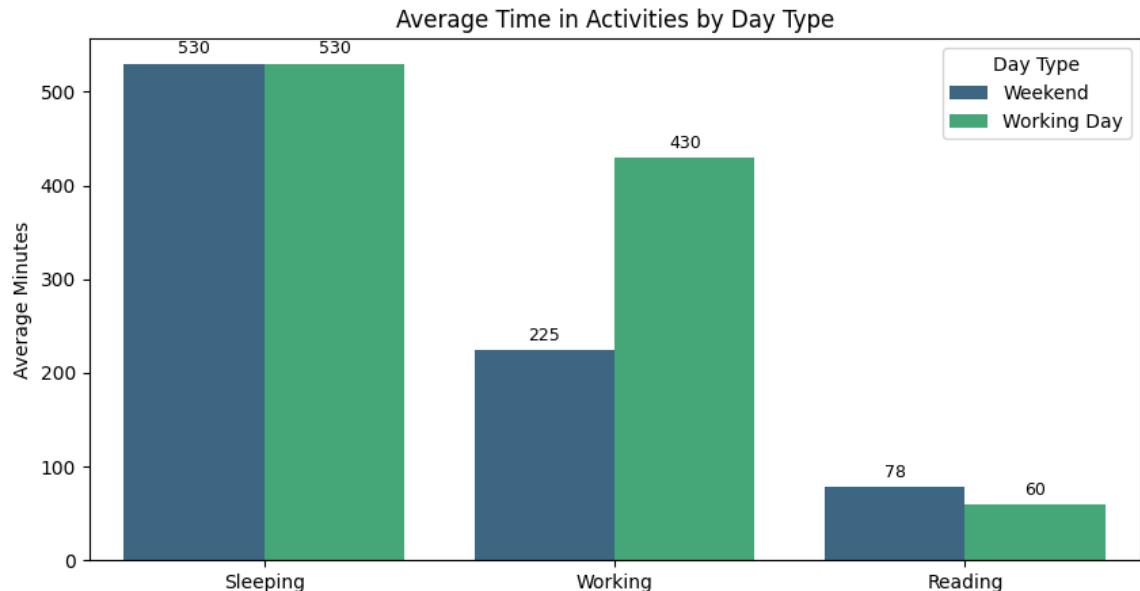


Figure 6: Distribution of activity variables based on Day of week

It is an interesting fact that Finns work on weekends for **225** minutes (almost 4 hours) on average, and **430** minutes (more than 7 hours) on working days. There is no difference found between the sleeping in this group as illustrated in Figure 6. Finns tend to read more on weekends instead of working days.

Differences in Activity Variables based on Living Environment in Finland:

To test the difference in the activity variables, a normality test was conducted on the variables based on the living environment. The null hypothesis of normally distributed data was tested against the alternative hypothesis of non-normally distributed data using the Shapiro-Wilk test. Table 7 displays the results of the analysis, with p-values less than 0.05 indicating significance of the test, hence rejecting the null hypothesis of normally distributed data. From Table 7 we can conclude that the activity variables working, sleeping and reading based on the living environment of respondents are not normally distributed. The non-parametric test, the Kruskal-Wallis test, for more than two independent samples, will be conducted to test the difference in the activity variables.

Table 7: Normality test of activity variables based on living environment

Source of variation	Groups	W-statistic	p-value	Decision
Working	City	0.565	< .001	Not normal
	Municipality	0.553	< .001	Not normal
	Rural	0.537	< .001	Not normal
Sleeping	City	0.17423	< .001	Not normal
	Municipality	0.17423	0.0077	Not normal
	Rural	0.17423	0.6851	Normal
Reading	City	9902.0	< .001	Not normal
	Municipality	9902.0	< .001	Not normal
	Rural area	9902.0	< .001	Not normal

The Kruskal-Wallis test was conducted with the null hypothesis of no significant difference in the minutes of activity variables performed by respondents based on their living environment against the alternative hypothesis of a significant difference in the minutes of activity variables performed. The results of the test for the activity variables are displayed in Table 8. The results in Table 8 indicate that there is no significant difference in the working and reading minutes of respondents in the city, municipality and rural areas. However, there was a significant difference in the sleeping time of respondents in the city, municipality and rural areas. The pairwise comparisons showed that there was a significant difference in the sleeping minutes between rural area respondents and both city and municipality respondents, whereas there was no significant difference between city and municipality respondents.

Table 8: Difference in activity variables based on the living environment of respondents

Test	Source of variation	Groups	Statistic	df	P-value	Decision
Kruskal-Wallis	Working	Living environment	0.1742	2	0.9166	Not significant
Kruskal-Wallis	Reading	Living environment	3.6951	2	0.1576	Not significant

Kruskal-Wallis	Sleeping	Living environment	10.8915	2	0.0043	Significant
Post-hoc test						
Mann-Whitney U	Sleeping	City vs rural areas	29113.5	1	0.001096	Significant
Mann-Whitney U	Sleeping	City vs municipality	31439.5	1	0.691099	Not significant
Mann-Whitney U	Sleeping	Rural area vs municipality	9902.0	1	0.020705	Significant

Differences in Activity Variables based on Day of Week the activity is performed:

To test the difference in the activity variables, a normality test was conducted on the variables based on the day of the week activities are performed. The null hypothesis of normally distributed data was tested against the alternative hypothesis of non-normally distributed data using the Shapiro-Wilk test. Table 9 displays the results of the analysis, with p-values less than 0.05 indicating significance of the test, hence rejecting the null hypothesis of normally distributed data. The results from Table 9, indicate that the activity variables working, sleeping and reading based on the day of the week that respondents performed these activities are not normally distributed. The non-parametric test, Mann-Whitney U test, for two independent samples, would be conducted to test the difference in the activity variables.

Table 9: Normality test of activity variables based on living environment

Source of variation	Groups	W-statistic	P-value	Decision
Working	Weekend	3.79207	< .001	Not normal
	Working day	3.79207	< .001	Not normal
Sleeping	Weekend	1700.0	< .001	Not normal
	Working day	1700.0	< .001	Not normal
Reading	Weekend	73498.5	< .001	Not normal
	Working day	73498.5	< .001	Not normal

The Mann-Whitney U test was conducted with the null hypothesis of no significant difference in the minutes of activity variables performed by respondents based on day of week against the alternative hypothesis of a significant difference in the minutes of activity variables performed. The results of the test for the activity variables are displayed in Table 10. The results in Table 10 indicate that there is a significant difference in the minutes of working and reading by respondents on the working day and the weekend. However, there was no significant difference in the sleeping time of respondents performing this activity on a working day and on the weekend.

Table 10: Difference in activity variables based on the day of the week activity was performed by respondents.

Test	Source of variation	Groups	Statistic	df	P-value	Decision
Mann-Whitney U	Working	Day of week	1700.0	1	< .001	Significant
Mann-Whitney U	Sleeping	Day of week	73498.5	1	0.4912	Not significant
Mann-Whitney U	Reading	Day of week	38785.50	1	< .001	Significant

Differences in Activity Variables based on the Sex of Respondent in Finland:

To test the difference in the activity variables, a normality test was conducted on the variables based on the sex of respondents. The null hypothesis of normally distributed data was tested against the alternative hypothesis of non-normally distributed data using the Shapiro-Wilk test. Table 11 displays the results of the analysis, with p-values less than 0.05 indicating significance of the test, hence rejecting the null hypothesis of normally distributed data. The results from Table 11, indicate that the activity variables working, sleeping and reading based on the sex of respondents are not normally distributed. The non-parametric test, Mann-Whitney U test, for two independent samples, would be conducted to test the difference in the activity variables.

Table 11: Normality test of activity variables based on living environment

Source of variation	Groups	W-statistic	p-value	Decision
---------------------	--------	-------------	---------	----------

Working	Female	0.195762	0.018689	Not normal
	Male	0.195762	0.010305	Not normal
Sleeping	Female	4341.0	< .001	Not normal
	Male	4341.0	< .001	Not normal
Reading	Female	75697.0	< .001	Not normal
	Male	75697.0	< .001	Not normal

The Mann-Whitney U test was conducted with the null hypothesis of no significant difference in the minutes of activity variables performed by respondents based on their sex against the alternative hypothesis of a significant difference in the minutes of activity variables performed. The results of the test for the activity variables are displayed in Table 12. The results in Table 12 indicate that there is no significant difference in the minutes of the performance of the activity variables based on the respondent's sex. This shows that there is no difference in the working, sleeping and reading time of males and females in Finland.

Table 12: Difference in activity variables based on the day of the week the activity was performed by respondents.

Test	Source of variation	Groups	Statistic	df	P-value	Decision
Mann-Whitney U	Working	Sex	4341.0	1	0.4588	Not significant
Mann-Whitney U	Sleeping	Sex	75697.0	1	0.9430	Not significant
Mann-Whitney U	Reading	Sex	32726.5	1	0.7922	Not significant

Association of activity variables and Demographic variables:

The test of association was conducted on categorical variables Dine-out and Visiting-library. **Table 13** illustrates that going to the library is dependent on age-group, so there is a significant association between them $X^2(6) = 58.5917, p=0.0000$. The age-group and living-environment have significant association with dine-out as well $X^2(6) = 90.5619, p=0.0000$ and $X^2(2) = 16.4043, p = 0.0003$.

Table 13: Association Results

	Statistic	df	p-value	Association
Dine out				
Gender	1.2974	1	0.2547	Not Significant
Age Group	90.5619	6	0.0000	Significant
Living Environment	16.4043	2	0.0003	Significant
Day Type	1.4956	1	0.2213	Not Significant
Visiting Library				
Gender	2.9073	1	0.0882	Not Significant
Age Group	58.5917	6	0.0000	Significant
Living Environment	3.7921	2	0.1502	Not Significant
Day Type	0.1958	1	0.6582	Not Significant

Analysis of the Relationship between Activity Variables:

To determine the relationship between the activity variables (working, sleeping, reading) we adopted the spearman correlation because it first ranks the values and repeated zeros will not affect the relationship, also the main reason to use this method is for the outliers. A Spearman correlation analysis showed a significant negative relationship between working and sleeping time ($\rho = -0.34$, $p < 0.001$), indicating that individuals who work more tend to sleep less. A weaker but significant negative relationship was also found between working and reading time ($\rho = -0.21$, $p < 0.001$), so, people who work more tend to read slightly less, though not as strongly as the work-sleep link. No significant relationship was observed between sleeping and reading ($\rho = 0.02$, $p = 0.52$).

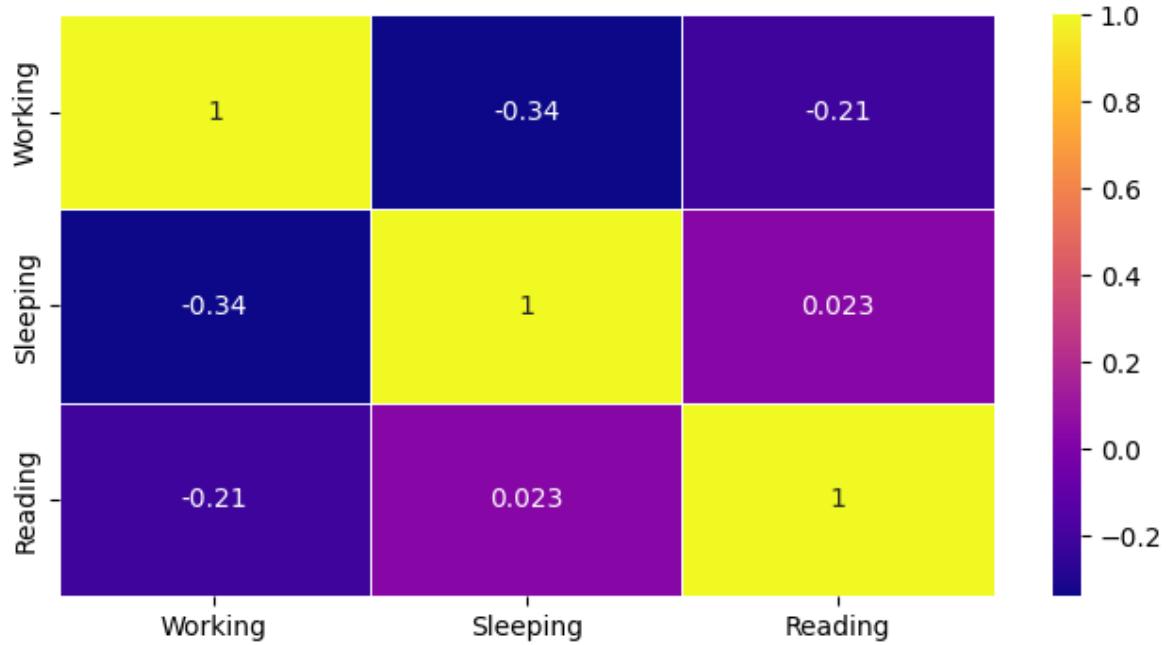


Figure 7: Correlation Matrix of Activities

Appendix:

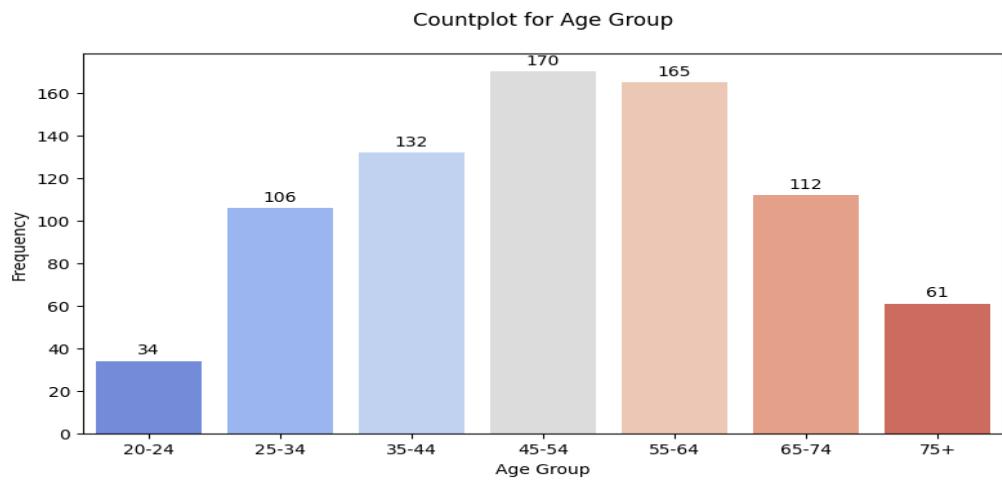


Figure 1: Distribution of respondents by Age groups