What factors predict the total alone time spent by adults in the United States from 2003 to 2023, and how do these factors relate to the amount of time spent alone?

Abstract

This study explores the factors that predict the total non-work related alone time spent by adults in the United States from 2003 to 2023, using data from the American Time Use Survey (ATUS). The aim is to understand how demographic and socioeconomic variables, such as household size, age, gender, labor force status, and time spent with family, influence the amount of time individuals spend alone. Our analysis reveals that time spent with family (TRTFAMILY) and age (TEAGE) are the most significant predictors of total alone time. Additionally, year (YEAR) and occupation code (TEIO1OCD) also contribute substantially to the model. The model's performance metrics show an R-squared value of 0.37, indicating that 37% of the variability in total alone time can be explained by these factors. These findings highlight the complex relationship between individual characteristics and time spent alone, with important implications for understanding social isolation and well-being in the context of aging and household dynamics.

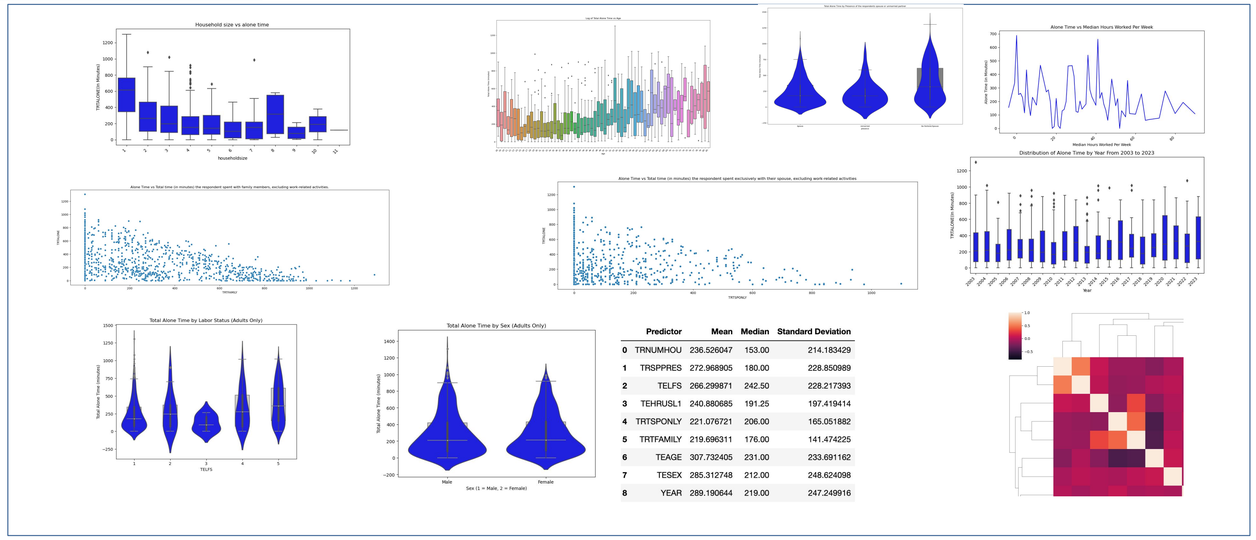
Introduction

This research investigates the factors influencing the total amount of non-work-related time adults in the United States spent alone between 2003 and 2019, using data from the American Time Use Survey (ATUS). Key demographic and socioeconomic variables, including age, gender, labor status, and household size, are examined to quantify their relationship with time spent alone. The study aims to understand how these factors contribute to social isolation, a growing concern for mental health. By analyzing relationships across various subgroups (e.g., age, gender, labor status), this research explores how different life circumstances impact social behaviors and well-being.

Methods and Materials

This study investigates the factors influencing total non-work related alone time among adults in the United States from 2003 to 2019, using data from the American Time Use Survey (ATUS). The dependent variable is the total time spent alone, excluding sleep, and key predictors include demographic and socioeconomic factors such as age, gender, labor status, household size, and work hours. A sample of 1,000 unique individuals was randomly selected from the ATUS dataset, as the full dataset was too large for analysis. The data was sourced from the respondent and roster files, merged by the unique household identifier (TUCASEID), and cleaned by filtering for adults, removing duplicates, and addressing missing data. Exploratory data analysis was performed using descriptive statistics, box plots, and scatter plots, while potential multicollinearity was assessed with a correlation matrix and visualized using a Seaborn clustermap. Due to violations of regression assumptions, Random Forest was used to assess predictor importance and model performance. The model's effectiveness was evaluated using Root Mean Squared Error (RMSE) after splitting the data into training and testing sets. All analyses were conducted using Python and relevant statistical and machine learning libraries.

Figures



	Predictor	Mean	Median	Standard Deviation
0	TRNUMHOU	236.526047	153.00	214.183429
1	TRSPPRES	272.968905	180.00	228.850989
2	TELFS	266.299871	242.50	228.217393
3	TEHRUSL1	240.880685	191.25	197.419414
4	TRTSPONLY	221.076721	206.00	165.051882
5	TRTFAMILY	219.696311	176.00	141.474225
6	TEAGE	307.732405	231.00	233.691162
7	TESEX	285.312748	212.00	248.624098
8	YEAR	289.190644	219.00	247.249916

Results

The model performance metrics reveal a Root Mean Squared Error (RMSE) of199.3, indicating a moderate level of prediction error. The R-squared value is 0.37, suggesting that approximately 37% of the variability in total alone time can be explained by the predictor variables.

In terms of feature importance, the most influential predictors are as follows: TRTFAMILY (time spent with family) with an importance score of 0.350, followed by TEAGE (age) with an importance score of 0.174. YEAR (survey year) and TEIO1OCD (occupation code) also play notable roles, with importance scores of 0.115 and 0.110, respectively. Other predictors include TRNUMHOU (household size) with an importance score of 0.092, TELFS (labor force status) with 0.091, TRTSPONLY (time spent exclusively with a spouse) with 0.032, TRSPPRES (spouse/partner presence) with 0.021, and TESEX (gender) with 0.013. These findings highlight the relative significance of each factor in predicting total alone time.

Conclusions

This study identified key factors influencing total non-work-related alone time using data from the American Time Use Survey. Significant predictors included household size, age, and family time. Larger households were associated with less alone time, supporting the idea that social support reduces isolation. Older individuals tended to spend more time alone, reflecting the increasing social isolation often seen with age. The importance of family time also underscores the role of social connections in reducing isolation.

Despite the model explaining 37% of the variation, there are limitations. The ATUS data, while comprehensive, is self-reported, which introduces potential biases due to recall or social desirability effects and unmeasured factors, such as mental health or lifestyle choices, may contribute to alone time.

WELLESLEY

Data Science Capstone. Naima Abdirahman