Predicting & Understanding the Behavioural Footprint of Older Adults in the UK

Thabsheera

30/05/2022

Introduction

Technology had succeeded in increasing the life-span of human over the years [1]. But not everyone is living a wealthy, healthy and a happy older life. It's crucial for science to advance to improve the quality of life and not just adding years to life.

Ageing Intelligence is a new term coined by Nic Palmarini (Director or UK National Innovation Centre for Ageing - **NICA**). According to NICA ageing intelligence is "*a comprehensive data-driven approach, leveraging a combination of human knowledge and experience, data analysis, and real-world applied testing and research*". Analysing the trend and amplifying the generated insights from behavioural and market data is one of the key aspects of ageing intelligence.

Behavioural data is crucial irrespective of the business domain. If it is an e-commerce market, behavioural data of customers can help in better recommendations and personalisation. When it comes to older population in the country, the behavioural data can lead to numerous opportunities benefiting both citizen and business. A person's health and lifestyle are related to their habits. If their habit/behaviour can be predicted, which can give insights about their health as well as it can prevent any unforeseen events that can happen. Most importantly, behaviour prediction can help in improve the lifestyle and thereby increase the happiness and longevity of human beings.

Project Aim

The high-level aim of this project is to predict what an older person will be doing at a given time of the day. This project will require incorporating a classification model (like decision tree or random forest) into a time series component. Because diary data has time series aspect to it and the activities are categorical variables.

Along with predicting the activities of older adults at a given hour of the day, combining two machine learning techniques adds a high value to this project.

Objective

This project will focus mainly on three objectives:

- Study the data and understand the existing trend of adult behaviours.
- Visualise the existing trend.
- Run a time series classification model on the data to predict the adult behaviour.

Optional objective if time permits:

- Recommend an alternative behaviour based on the past enjoyment level on that activity.

Progress

Initial Research

Significant number of papers have been read to understand the past work using diary data and behavioural analytics.

Theory of planned behaviour by Ajzen, 1991 [5] is the foundation of this project. According to this paper, the individual's intention to exhibit a behaviour is dependent on three main factors. First is the *attitude* towards the behaviour and this tells the depth to which a person has favourable or unfavourable judgement of the concerned behaviour. Second one is a social factor called as *subjective norm*, it refers to the perceived social pressure to perform the act. Third one is the *perceived behavioural control*, which refers to the perceived ease or difficulty for performing the behaviour and it reflects the past experience as well as the anticipated barriers.

There have been a few studies on various sectors of the industry based on theory of planned behaviour to predict the customer's or employee's behaviour, but little to no studies are done on diary data. One such study in HR analytics [3], to predict the dishonest behaviour patterns from candidate, made use of the theory of planned behaviour and used multiple classification techniques such as logistic regression, random forests and decision tree.

One study which was based on diary data [4] defined behaviour in terms of target, action, context and time(TACT) principle (Ajzen, 2006). Time series methods have been used in this study to test the dependency on daily sequential measures.

Data Exploration & Manipulation

UK time survey data for 2014-15[2] has diary and household interview data for 4,000 households in the UK. Each diary data contains what that individual is doing at a particular time of the day and around 8000 people have participated in the survey.

The data is already given with a weight for household and individual so that each entry will be comparable to the entire UK population. The granularity of the data is household, person, weekday/weekend, and 10 minute interval of the day.

- Basic EDA is conducted on the entire data to understand the distribution. Following are the major findings from the data:
  1. 50% of the individuals participated in the diary activity are from the generations *baby boomers* and *Gen X*. Gen X spends travels more than baby boomers whereas baby boomers tends to relax at home. Gen X also spends much higher time in childcare than baby boomers.
  2. When it comes to psychosocial stages among older adults, the amount of time put in the main job decrease drastically after the age of 57. Also after the age 67, people tends to go for a retirement and spends more time at home.
  3. Weekend calls for entertainment for majority of the people. Activities such as social life and TV watching happens more during weekend.
  4. Household with children tend to be more active with entertainment and travel activities compared to adults only households.
  5. The activities which people enjoy the most are self care related activities. Most people rate travelling at a medium enjoyment level.
- To visualise the above findings, a R Shiny dashboard is built. Snapshot of the first version of dashboard is given in Figure 1. More filters and plots will be added.

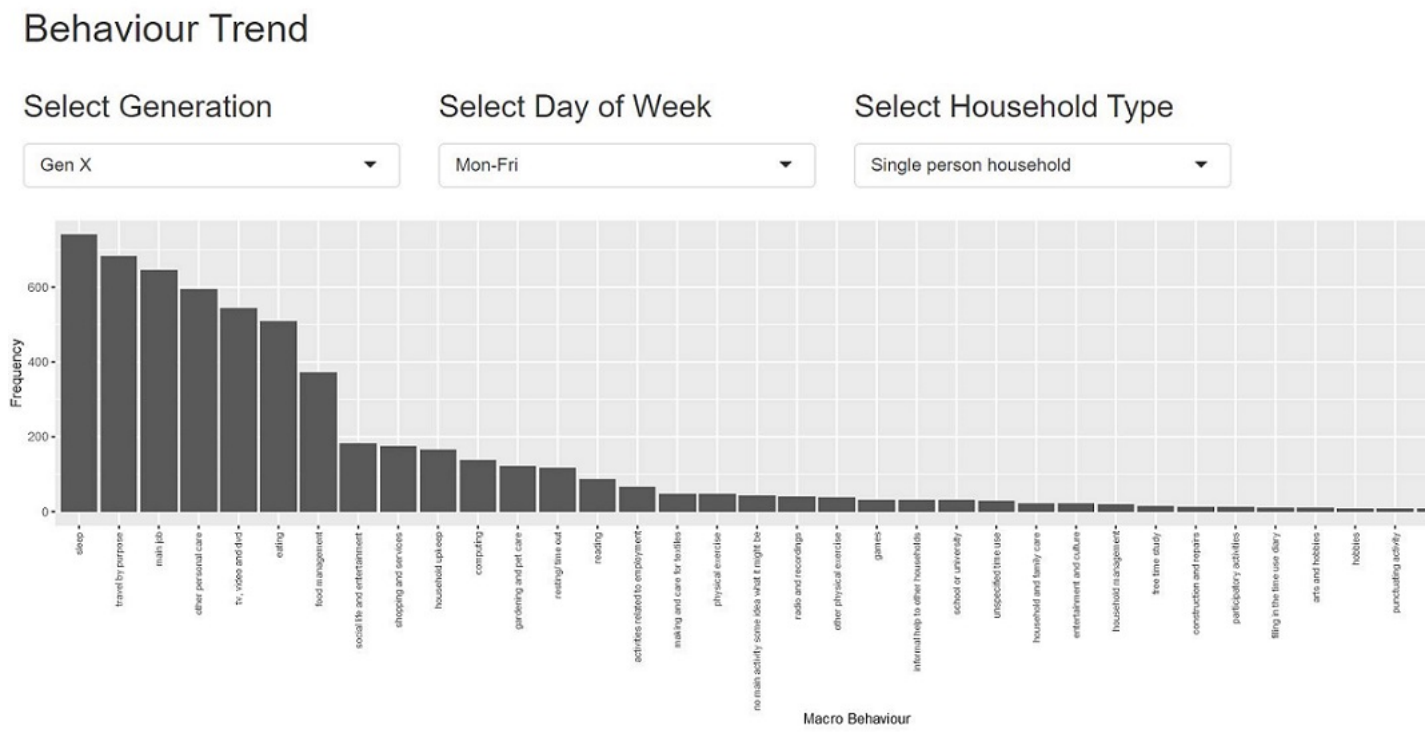

Figure 1: Shiny Snapshot

- When it comes to behaviour classification, the data needs to be prepared in a certain way. First of all, the important field like activity needed tweaking. The activities are micro behaviour like vehicle maintenance, dwelling repair, visiting historic site etc. There are 276 such behaviours which will increase the complexity of a classification model by will requiring 276 output groups. So, these micro behaviours are then grouped into 48 groups. For example, activities/ behaviours like feeding the child, teaching the child etc is now under a group called 'childcare'. This will reduce the complexity of the model.
- The activities are recorded for every 10-minute interval of the day; this interval is then converted to hour of the day. So that instead of 144-time interval, 24-time intervals can be used for predicting the activities in that interval. The 10-minute interval is avoided since most of the activities overlap the intervals. To convert the data into hour level, the activity is replaced with the most frequent activity in that hour.

Work Plan

Plan for this project is given in Figure 2.

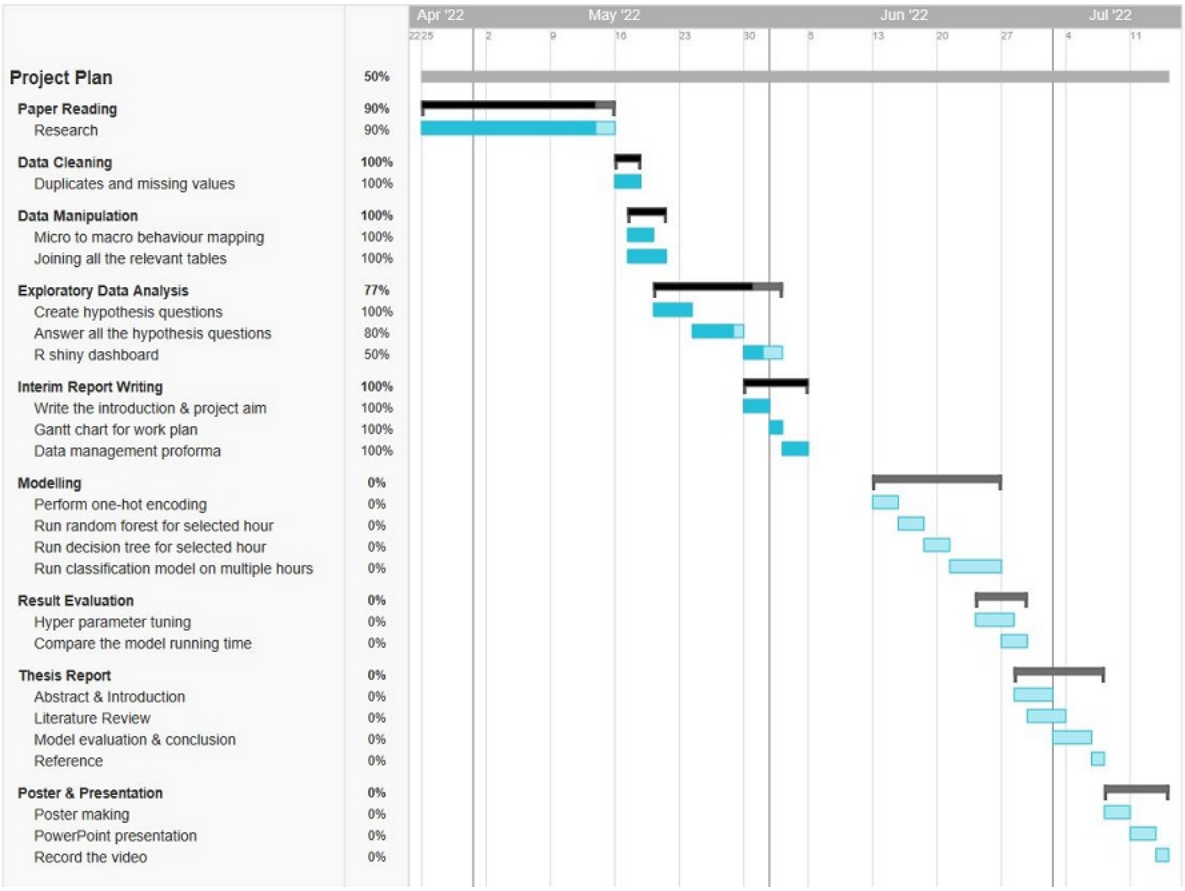

Figure 2: Project Plan

Refernces

1. Jan Vijg, Aubrey D.N.J. de Grey: Innovating Aging: Promises and Pitfalls on the Road to Life Extension. Gerontology 2014;60:373–380.
2. Sullivan, O., Gershuny, J. (2021). United Kingdom Time Use Survey, 2014-2015. [data collection]. UK Data Service. SN: 8128, DOI: 10.5255/UKDA-SN-8128-1
3. Peisl, T., Edlmann, R. (2020). Exploring Technology Acceptance and Planned Behaviour by the Adoption of Predictive HR Analytics During Recruitment. In: Yilmaz, M., Niemann, J., Clarke, P., Messnarz, R. (eds) Systems, Software and Services Process Improvement. EuroSPI 2020. Communications in Computer and Information Science, vol 1251. Springer, Cham. https://doi.org/10.1007/978-3-030-56441-4_13
4. Nicola Hobbs , Diane Dixon , Marie Johnston & Kate Howie (2013) Can the theory of planned behaviour predict the physical activity behaviour of individuals?, Psychology & Health, 28:3, 234-249, DOI: 10.1080/08870446.2012.716838
5. Ajzen, I. (1991), "The theory of planned behavior", Organizational Behavior and Human Decision Processes, Vol. 50 No. 2, pp. 179-211.

Research Data Management Proforma

0. Project title, author, version and date		
Project: Understanding and predicting the behavioural footprint of older adults in the UK		
Author: Thabsheera PK	*Version: 1*	*Date: 01/06/22*

1. Description of the data

1.1 Type of study

The given data is a diary data with activities recorded for every 10 minutes of the day. Along with the diary data, household interview and individual interview data is available. All these data are used to design and run a time series classification model to predict the behaviour of older adults

1.2 Assessment of existing data

The United Kingdom Time Use Survey, 2014-2015 (UKTUS) is a large-scale household survey that gives data on how people aged 8 years and above in the UK spend their time. The source for the mentioned data is 'uk data service' and the data is publicly available. (Sullivan, O., Gershuny, J. (2021). United Kingdom Time Use Survey, 2014-2015. [data collection]. UK Data Service. SN: 8128, DOI: 10.5255/UKDA-SN-8128-1)

The study provides the time use of individual over on a weekday and a weekend. The time diary provides information about the activities, location, presence of computer or phone, level of enjoyment etc. The questionnaire data provides the information of characteristics of the household and individual such as employment, household type, demographics information etc.

1.2 Types of data

Majority of the data fields are categorical or qualitative data. Since the data is generated from surveys, most of the fields are qualitative such as demographics, employment, activities etc.

1.3 Format and scale of the data

The data is available in spss format, Used R to read and visualise the data. The diary data is available in two data type of data sets, one is the long format where each row is for a household- person- day of the week- time interval. The other format is wide, where each time interval is transposed into columns. Both have the same data of ~4000 households and ~8000 individuals.

The questionnaire data is of the same households and individual but with their characteristics. There are of 4733 records in household data and 11421 in individual data

2. Data collection / generation

2.1 Methodologies for data collection / generation

Since the data is a publicly available UK time use survey for the period of 2014-15, the data collection was effortless. There was only a need to create an account in UK data service website and fill the form stating the purpose of the data.

2.2 Data quality and standards

The data has gone through lot of cleaning up and calibration before getting uploaded to UK data service. Weights have been given to household and individual so that it will be comparable to the entire households and population of UK. Some of the missing data has been replaced by the average values of the existing records or some of them were removed. Currently, there's no duplicate records in the data.

3. Data management, documentation, and curation

3.1 Managing, storing and curating data.

The data, which was publicly available in the UK data service, is downloaded and saved in the local system. The codes and generated small tables are getting backed up in the git hub.

3.2 Metadata standards and data documentation

The meta data and data dictionary for the data is available in the UK data service website. The data dictionary has the information about each column and what the value in that column means.

All the data manipulation that has been done as part of this project is properly commented in the codes and it is easily reproducible. The code changes are timely updated in the git log.

The entire procedure for the project has been properly documented and uploaded as a ReadME file in the github.

4. Data security and confidentiality of potentially disclosive information

4.1 Main risks to data security

The data is masked for the identity, the household number represented by a serial number. This serial number to exact household mapping is not available anywhere publicly, this is only known to the people responsible for the survey. There's no description to understand the individual as well. The field pnum(person number) is considered as individual.

5. Data sharing and access

5.1 Suitability for sharing

The data is suitable for sharing within the organisation in this case it's Newcastle University.

5.2 Discovery by potential users of the research data

The data is saved in the local system, and which is originally a publicly available data, so anyone can download from the UK data service after submitting relevant information.

5.3 Data preservation strategy and standards

The data will be preserved in the local system for the duration of this project

5.4 Restrictions or delays to sharing, with planned actions to limit such restrictions

Not Applicable

6. Responsibilities and Resources

Are there any resources (e.g. storage/ training) that you will require to fulfil the plan?

NO

7. Relevant institutional, departmental or study policies on data sharing and data security

Policy	URL or Reference
Data Management Policy & Procedures	https://www.ncl.ac.uk/media/wwwnclacuk/research/files/ResearchDataManagementPolicy.pdf
Information Security	*https://services.ncl.ac.uk/itservice/policies/InformationSecurityPolicy-v2_1.pdf*
Other	