Islamic University of Technology (IUT)

Department of Computer Science and Engineering (CSE)

Investigation of the Causal Relation of Personal Financial Activities and the Social Media Influence

Author

Rubayea Ferdows - 16041015

Supervisor

Prof. Dr. Abu Raihan Mostofa Kamal

Head, Department of CSE

A thesis submitted to the Department of CSE

in partial fulfillment of the requirements for the degree of M.Sc.

Engineering in CSE

Academic Year: 2016-17

May - 2022

Declaration of Authorship

This is to certify that the work presented in this thesis is the outcome of the analysis and experiments carried out by Rubayea Ferdows under the supervision of Prof. Dr. Abu Raihan Mostofa Kamal, Head of the Department of Computer Science and Engineering (CSE), Islamic University of Technology (IUT), Dhaka, Bangladesh. It is also declared that neither of this thesis nor any part of this thesis has been submitted anywhere else for any degree or diploma. Information derived from the published and unpublished work of others has been acknowledged in the text and a list of references is given.

Author:

Rubayea Ferdows

Student ID - 16041015

Supervisor:

Prof. Dr. Abu Raihan Mostofa Kamal

Head of the Department

Department of Computer Science and Engineering

Islamic University of Technology (IUT)

Acknowledgement

I would like to express our grateful appreciation for **Prof. Dr. Abu Raihan Mostofa Kamal**, Head of the Department, Department of Computer Science & Engineering, IUT for being our adviser and mentor. His motivation, suggestions and insights for this research have been invaluable. Without his support and proper guidance this research would never have been possible. His valuable opinion, time and input provided throughout the thesis work, from first phase of thesis topics introduction, subject selection, proposing algorithm, modification till the project implementation and finalization which helped us to do our thesis work in proper way. I am really grateful to him.

I am also grateful to **Fuad Ahmed**, M.Sc. Student, Department of Computer Science & Engineering, IUT for his valuable inspection and suggestions on my proposal of investigation of the causal relation of personal financial activities and the social media influence.

Abstract

Many social network (SN) applications such as Facebook, Twitter, Instagram, etc. provide many benefits that allow users to connect, follow each other, share various content, and influence them to engage in various activities in their personal lives. Sometimes it impacts their habits such as online buying, restaurant check-in, traveling, etc. Different existing researchers have used a variety of approaches to identify these impacts on various topics, including fitness, psychological health, and so on. However, very few research has been done on investigating individual expenditures. Thus, in this work, we aim to 1) generate an appropriate dataset based on social media usage and users financial activities, 2) investigate the correlation between social media use and personal financial activities, 3) estimate personal expenditure based on various social media aspects such as checking into restaurants, buying clothes, traveling to new locations, doing something entertaining, and so on. We collected data through an online survey using social network platforms such as Facebook. In the study we apply a causal model using propensity score-based inverse probability treatment weighting (IPTW) and a doubly robust estimator along with some other methods. We evaluate our approach by refuting the outcome. Finally, we find that social media usage has a significant impact on spending patterns.

Keywords— Causal Effect, Quasi Experiment, Causal Model, Propensity Score matching(PSM), Inverse Probability of Treatment Weighting (IPTW) Estimator, Doubly Robust Estimator.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

Social Network (SN) characteristics are essential to many computer activities today. Numerous successful websites and apps use social networking to intrigue their customers to interact, form social connections, publish updates, share and exchange content, and comment on other posts. Thus, Social networking features are almost unavoidable and has become part of our daily life.

Globally, the rapid increase in digitization and ubiquitous computing have led to a dramatic increase in active social media users [1]. People are using numerous online SN applications such as Facebook, Instagram, Twitter, Pinterest, and TikTok, capturing and sharing thousands of photographs every single day [2]. Additionally, a large population is connected with the SN platform actively. According to the existing studies, a rich collection of literature in data science and associated fields focused on various effects of such contagious usage of SN. However, it has unconsciously underestimated the need for or proper planning for the future as well as the culture of the middle class [3].

As in many other countries in the Global South [3], sharing digital pictures and videos have become a common habit across urban areas in the country because of cheap smartphones, mobile internet and Wi-Fi, as well as the popularity of social networking applications [4, 5].

Various research has been carried out to investigate the impact of social media factors on human activities. Social media influences physical activity quite strongly [6]. Another related study demonstrates that social media influence physical activity [7]. Social media can help us make better choices by exposing ourselves to more goods and services and enabling us to learn about the experiences of others. But it might also influence us to spend with or without necessity. This additional spending puts a strain on people's budgets, particularly among the middle class people of Bangladesh. This study attempts to identify possible social media-related causes of personal expenditure and analyze the effect on personal life.

1.2 Problem Statement

The primary purpose of digitizing the country through a variety of actions is to achieve success. Smartphones are becoming more widely available at a rapid rate. The virtual world has a significant impact on people in a variety of ways. Some people are using and experiencing the digital facility for their own purposes, while others are simply interested in seeing it. In addition, social network business platforms are receiving a lot of attention. Almost every type of business now has an internet platform in addition to their physical location. Despite this, many online platforms are readily available to end-users. As a developing country, many of our inhabitants are using the internet to access a variety of innovative smartphone applications.With regard to the digital facility people are now more accustomed to online work while manual pen-paper work declines in every field. Social media is used by a vast number of people. The number of people who use social media on a regular basis is continuously expanding. When adolescents spend a lot of time on social media sites like Facebook, Instagram, and Twitter, they become interested in a variety of activities like sharing random posts, a picture or video, or something else entirely. As a result, it seems social media has a significant impact on the daily lives of its users. Also, sharing various postings or videos, such as restaurant check-ins, trip to new areas, clothing purchases, doing something entertainment for themselves, and so on, may provoke the curiosity of other users. The online behaviors then influence the users in the same network, who try to have the same experience as others. They are unconsciously spending a lot of money on these. At the end of the day, this may have an impact on a middle-class household. Social media has an impact not just on teenagers, but also on housewives and the elderly.

1.3 Motivation and Scope of the Research

Social media influence the users both online and offline behavior. We want to demonstrate how social media promote people's impulsive spending in our research. We are particularly interested in learning how the impact of social media causes middle-class people to spend a substantial portion of their money unintentionally. The main things that influence the personal financial activities of this study are given below:

- **Rapid growth of internet:** Availability of smartphone and internet has rapidly increased the social media users in Bangladesh.
- **Business activities on social media:** Frequent users are attracted by social media based business activities.
- **Influential content:** Social media contents such as reviews, photos, posts highly influence the users.

1.4 Research Challenges

The main objectives of this work is to specify the influence of social media(SM) on users' daily lives in various factors. Some challenges will arise in order to achieve this goal, which are as follows:

- **Lack of appropriate dataset for the study**

  With the introduction of online social networks, we have gained a better grasp of how to characterize and affect people's behavior in a social setting [7]. However, because of lack of good quality data, it has been difficult to investigate the effects on users "offline" behavior, such as financial activities of user.

- **Development of causal influence model**

  Predicting the influence of social networks on both online and offline behavior is heavily reliant on adequate study design [2]. Existing research on social influence in social networks has primarily focused on evaluating online behaviors and outcomes such as app adoption, content downloads, suicidal tendencies, and content re-sharing. Recent research has gone beyond discovering useful correlations for creating prediction models to concentrating on causal relationships between diverse circumstances and user behavior [8]. However, Researchers have been unable to use Randomized Controlled Trials when working with chronological online trace data due to a lack of control over the balanced distribution of contents into comparable groups. However, in recent years, researchers have encouraged to focus on natural or quasi-experimental methods to determine the casual impact of multiple parameters. As a result, we can easily obtain the distributed groups for determining casualty.

- **Measure the social media influence**

  It is difficult to tell which part of activity is due to social media impact and which part is regular activity. Measuring the social media influence is also challenging. We can apply causal analysis to find users' activities based on their actions.

1.5 Thesis Contributions

In this study, the relationship between the use of social media and personal expenditure is investigated to see how much it can influence day-to-day lives. In total, 768 persons submitted data to our research, with most of the data being self-contributed. The impact has been determined based on whether they are active social media users or not. The data was acquired over five months through an online survey(google form). The main contributions of the work are given below:

- The generation of a appropriate dataset based on social media usage and users financial activities.
- Investigation on the social media aspects that have raised the cost of the users' personal life. We will also include other essential factors such as eating out, buying clothes, going to new places, personal entertainment, and so on, all of which contribute to growing spending.
- Design of a causal model that can demonstrate the correlation of a person's SN usages and personal daily spending.

1.6 Thesis Outline

Chapter 1 discuss the study in a precise and concise manner. Chapter 2 deals with the necessary literature review for our study. In Chapter 3, it stated the skeleton of the proposed method, proposed algorithm and also the flowchart to provide a detail insight of the working procedure. Chapter 4 shows the results and comparative analysis of successful implementation of our proposed method. Chapter 5, discuss about the result and validation. Chapter 6 focuses on conclusion and future work of the study. The final segment of this study contains all the references and credits used.

Chapter 2

Literature Review

With an increasing number of individuals utilizing the social media, its social impact has gotten a lot of attention. Moreover social media are regarded as crucial element in peoples daily activities these days. Social networks(SN) have been focused as a large-scale "sensor" capable of providing insights about user's activity, habits, emotions, thoughts, and health. So that, many researchers have focused on social media by analyzing user behavior pattern in order to find causal relationships and measuring user behavioural patterns in the observed data.

A subcategory of such studies looks on how people' online behavior affects their offline behaviors [6, 7, 8]. Fewer research have looked at the effects of individuals' online social media activities on their real-world behavior (offline).

2.1 Rapid growth of internet

Since the early 1990s, the internet has sparked a massive revolution that encompasses social, economic, political, technical, and cultural aspects that affect countries. Everywhere in the world, people, communities, and individuals are impacted. The information, communication, and technology convergence as well as the creation of portable multi-media service a network that allows people and organizations to connect and collaborate by social media. Social network has unquestionably become the most important instrument in business climate a more linked and flatter [9]. During the previous decade, internet usage has increased dramatically in practically every country on the planet. People use the internet to send and receive e-mails, chat, conduct research for school or work, download music or photographs, and perform a variety of other tasks at home, at work, and in other locations such as Internet cafes. By empowering people and enhancing their productivity, the Internet helps a firm compete in the market. The adoption of broadband and access to the internet has resulted in an increase in the GDP of emerging countries.

The rapid expansion of digitization and globalization has resulted in an unexpected increase in the number of active SM users in Bangladesh [1]. Like in other countries from the South, the availability of low-cost cellphones, internet access, and widespread use of social media has encouraged people to share more of their daily activities on social media [10].

2.2 Growth of social commerce

By the rapid growth of social media(SM), The field of social business has arisen as a new type of e-commerce. Use of SM, enables social interactions and influence user to buy things they did not intend to purchase or do not truly require [11]. On these social networks, consumers can discover interesting links to purchasing websites. Under these conditions, impulsive purchasing is unavoidable, especially in the case of social commerce [12].The worldwide online market was valued at USD 9.09 trillion in 2019, with a compound annual growth rate (CAGR) of 14.7 percent predicted from 2020 to 2027 [13]. In 2019, Asia Pacific dominated the online market with a 55.31 percent share. This is due to a rising propensity among enterprises to conduct business through online platforms, expanding infrastructure, and an increase in the number of internet users.

Bangladesh, like the rest of the globe, is advancing in online platforms. According to the German research company Statista, Bangladesh's GDP has surpassed one and a half billion dollars, and is predicted to reach 2 billion dollars in current year and 3 billion dollars by 2023 [13]. The spreading pandemic has confined the entire world to their houses, but that does not imply that market demand has decreased. During the COVID-19 Pandemic, it was revolutionized. When compared to the preceding period, online sales have climbed by 70 to 80 percent [14]. As it shows in figure 2.1:

Online buyers appear to be drawn to the simplicity with which they may locate things on the Internet, as well as the thorough product information and wide range of

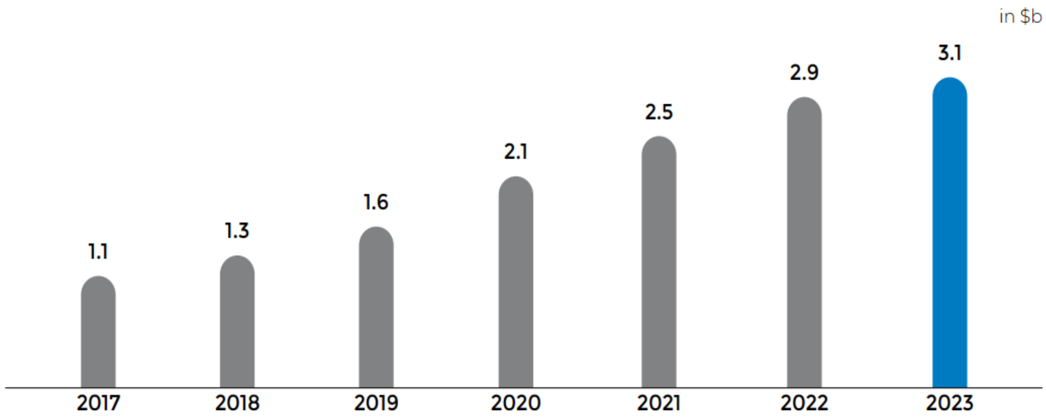


Figure 2.1: Bangladesh's e-Commerce market is projected to grow in double digits

options available. Smaller retailers have embraced the Internet due of the relative ease with which they may start up shop [15].

2.3 Significance of social media influences

Online social media(SM) has recently become the active area of comprehensive quasi experimental design research which target to demonstrate causal relationships and making assumptions influenced by social media usage. Several studies have been conducted to determine social media's influence on human activities. For example, Althof et al. [6] shown that physical activities are highly influenced by social media where people are changing their activities by the influence of a social networking application and demonstrate that social networking causes a significant rise in users online and offline activities. Another similar study has shown that social media influences physical activities [7]. In another study, it was shown that social media had affected users' real-world affective process [2, 16]. Here, they analyzed that when people engage in or disengage from real-life activities, they are changing their behavior on social media. A current study of social media influence in domestic space showed that people are firmly moving towards western culture by the impact of social media and changing the major functionality, redecorating their home space [3].

Another similar study showed that social media(SM) has an influence on physical activities. In a study conducted by sayed amin mirlohi showed that social media have affected users' real world affective process. Here, they analyze that, when people are engaging in or disengaging from real life activities, they are changing their behaviour on social media also [8]. An existing study of social influence in domestic space showed by Nusrat Jahan Mim where people are strongly moving towards western culture by the influence of social media and changing the major functionality, redecorating their home space [3]. While preparing this article, the work has gone through a research where David Stück shows the transformation of people's physical activity by the impact of social media. They have collected the dataset from a popular fitness tracker app [7]. Moreover, Table 2.1, provides some existing research methodologies that have been used to find a causal effect.

One study focused on Twitter users who had self-disclosed having an anxiety disorder after expert validation. They analyzed their timeline data and created a model with several social interactional attributes. They then attempted to identify specific interaction patterns that were significantly affected by an individual's historical anxiety status by developing a Granger causality framework and time series forecasting models. The research actually provided novel insights into how an individual's mental health state, specifically anxiety disorders, influences online social networks and interactions. The work also demonstrated that online SM platforms have a positive impact on individuals who are vulnerable to mental health issues [17].

Wang et al. [18] showed that both theoretical and empirical studies support the positive effect of internet use on reducing overall depression levels in older adults. Online participation can have an effect on well-being outcomes. Older people can access more health information and support online, which has been proven to benefit their mental health, by using the internet. In Americans aged 80 and up, using the internet for social purposes was associated with higher levels of life satisfaction and accomplishment. Internet use and online activities can reduce loneliness, increase social participation, and strengthen social networks in older adults, lowering depression levels. They have used PSM to balance their data to find the impact of internet use in older ages.

Saha et al [19] make two significant contributions in their work. First, to detect the effects of drug use and that these modifications are sensitive to drug families, they demonstrated a proof of concept that social media is useful as an effective sensor to scalably detect behavioral changes in individuals who initiate treatment via use of psychiatric medication; and second empirical findings included the invention that people's online behavior patterns change in some unpredictable ways following drug intake, and these may varies from the named side effects.

Despite of these above mentioned statistical studies some other research has been done through machine learning(ML) approaches. Researchers trying to find the social media factors on various aspects like country's economical growth, person's behaviour prediction. In one study [20] showed the social media influence on financial services industry where They examine the roles and effects of machine learning (ML) and artificial intelligence (AI) in the country's financial services industry. They conducted a survey to assess the current AI/ML landscape in the United Kingdom. Machine learning has had a significant impact on fraud and compliance, credit scoring, financial distress prediction, robo-advising, and algorithmic trading. They look at these applications through the lens of examples from the country, as well as the role of regulation and governance in ML applications for financial sector. Another study demonstrate the relationship of a users personality and Facebook usages. They also predicted the accuracy rate in five different factors [21, 22, 23, 24]. In one work they showed the technique to measure the influence of many popular social media application like: Facebook, Instagram, Twitter etc.The impacts of social media are calculated by many Machine Learning regression approaches [25, 26, 27]. Some works have been done on the Facebook also among them one study [28] showed how machine learning approaches helps to properly infrastructure the hardware and software requirements to use Facebook. These ML models support various capabilities to the new user experiences and many services such as News Feed post ranking, speech and text translations, and photo and real-time video classification [29]. In another interesting work [30] has been done to find the causal inference estimation on rural poor people and micro-credit.They extend the work to demonstrate the future causality rate by some machine learning methods [31, 32].

At the end, we can say the work of althoff et al. [6] indicate and as well as demonstrate several interesting directions for future work including studying what causes leads to healthy behavior change in various groups of people. Although, they focused on only physical activity but other effects such as financial, psychological activities are need to explore. Then falvarjani et al. [2] focused on causal relation between users offline and online behavior. Their study limited to twitter users only. They only showed the relation of psychological processes such as anxiety and depression. However, other aspects such as financial activities can be explored. Stuck et al. [7] limits on only a fitness tracker data. They only performed between subject analysis to measure the causality but other causal model can also be explored to get better result. Some work [3] conducted on only 20 30 users data of middle class people of Dhaka. Their study also limited to few places. However, a diversified dataset with large sample size can help to achieve a better result.

Therefore, we demonstrate how social media promote people's impulsive spending in our research. We're particularly interested in learning how the impact of social media causes middle-class people to spend a substantial portion of their money unintentionally. The possible outcome of the research will be (i) To show the correlation between social media usage and personal financial spending. (ii) The generation of a suitable dataset based on social media activity. (iii) Design a suitable experiment to determine the social media influence on the users spending patterns.

These studies show that, under certain situations, an individual's online presence and activity might influence their offline activities.

Table 2.1: Few of Existing Causal Effect Estimation Studies

Reference	Objective	Approach	Limitation
Althoff et al. [6]	Use health tracking data to demonstrate how online social networks influence both online and offline user behavior.	matching based observational studies, Natural experiments, difference-in-difference models.	They focused on only physical activity but other effects such as financial, psychological activities are need to explore.
Stuck et al. [7]	Investigate the health problems that concern the trackers and quantify how these factors affect the interaction between the social network(SN) and physical activity.	Between-subject analysis, causality analysis, network science and dynamic networks.	Limits on only a fitness tracker data. They only performed between subject analysis to measure the causality but other causal model can also be explored to get better result.
N. J. Mim and S. I. Ahmed [3]	Analyze the effects of social media on urban architecture in the Global South.	Nine-month long ethnography in Dhaka, Bangladesh.	Study conducted on 20 to 30 people and location limited to few places. However, a diversified dataset with large sample size can help to achieve a better result.
Saha et al. [19]	Utilize social media data to investigate psycho pathological effects associated with self-reported psychiatric medication use.	SVM classifier to divide data. Propensity score based causal analysis.	They cannot prove actual causality even using a causal framework. They lacks proper confounding variables and also have selection bias and self report bias.
Gotanda et al. [33]	The objective of this research was to see if there was a link between the Affordable Care Act's Medicaid expansion and changes in low-income persons healthcare spending.	Quasi-experimental design and difference-in-difference analysis.	The study lacks on considering all possible covariate and only considered few covarites. It also lacks proper result validation.

Continued on next page

Table 2.1 – continued from previous page

Reference	Objective	Approach	Limitation
S. Tirunillai and G. J. Tellis. [34]	Analyzes the impact of offline television advertising on multiple metrics of online chatter or user-generated content.	Quasi-experimental design, difference in differences, synthetic control, and vector auto regression.	This study lacks appropriate data and only followed one method for quasi experiment, which not always justify the outcome.
Dutta et al. [17]	Discover different insights and practical lessons on how a person's mental health affects their online social connections.	Causal inference based time series approach, Granger causality technique, VAR.	Data gathered through tweets and that's why there is a chance of reduction in interactions. User may have fear of being judged negatively.
Falavarjani et al. [2]	Examine whether there are any causal relationships between social media users real-world activities and their online emotional expressions.	Quasi-experimental design, stratified propensity score matching.	Study limited to twitter users only. They only showed the relation of psychological processes such as anxiety and depression. However, other aspects such as financial activities can be explored.
Wang et al. [18]	Study the relation between internet use and depression in older persons.	Binary logistic regression, PSM.	This study only consider 10 covariates which may not address all of the important confounding factors.

2.4 Background

This research has a theoretical foundation that incorporates computational causal analysis. Our theoretical knowledge on social media impact factor based on the current research that addresses social media influence of various online/offline activities on people's behavior. For example, the impact of social media usage on one's physical activity has been described in a literature [6]. Based on a fitness tracker data, It claiming that active social media engagement increases peoples physical activity and disengaging from social activity, reduces physical activity. Another example is provided in the literature [2], where researchers attempted to determine the causal relationships between social media users and their real-life activities. People express their emotions on social media platforms whenever they engage in or disengage from real-world activities. Another study [18] found that older adults experience less depression when they use the internet. The work was completed by balancing various important factors such as location, profession, gender, age, urban or rural residence, pension status, educational background, physical health, life satisfaction, and intelligence level of any elderly people who use social media.

Existing research implies that it is possible to uncover causal relationships between users' online/offline activities and their social media usage, which motivates our effort. However, Our work differentiate itself from the studies that investigates how online activities influence offline activity. Existing studies explores the social media influence on physical activity, mental health, emotional change etc. But we are specifically focusing on users financial activities.

We looked into this within a causal framework to see if there are any potential causal relation between users financial activity and social media usage. Research with a quasi-experimental design are increasingly being utilized to evaluate the effect of various elements in the setting of online social media [35] and are seen as an adequate and sensible alternative to studies with complete randomly assigning of subjects [36]. As a result, we carried out a quasi experimental design study. The main elements of quasi experimental study are as follows: (1) the dataset, (2) treatment group. (3) control group, (4) causal model, and (5) the outcome. Details of these elements are briefly discussed in Chapter 4.

Chapter 3

Data Collection and Prepossessing

In this study, we gathered data through an online survey. We provided several questions and obtained information between December 2020 and April 2021. Around 768 people participated and provided information, including salary, gender, age, monthly savings, and categorized data on frequent spending patterns that influence social media. We also prepossess and clean our dataset. In the following section, we provide details of our questionnaires and how we processed them accordingly. The dataset characteristics table is followed by Table 3.2

3.1 Qualitative Research

The term "qualitative research" refers to a broad range of processes and methods for studying social life [37]. The information or data gathered and processed is primarily qualitative in nature, with text - based materials such as interviews, field observations, and records, as well as visual materials such as artifacts, images, video footage, and Web sites documenting human experiences about others and/or oneself in social action and instinctive states. [38]. According to another study [39], qualitative data is a source of well-grounded, rich descriptions and explanations of processes in discernible local contexts. With qualitative data, one can keep the chronological flow, see which events lead to which consequences, and derive useful explanations. Furthermore, good qualitative research is more likely to yield unexpected findings and new integration, as well as assist researchers in moving beyond initial conceptions to generate or revise conceptual frameworks. Finally, as a result of the qualitative studies, there are words that have a concrete sense that often provides to be far more convincing to a reader, especially when organized into stories. Leedy et al [40] said the term qualitative research contains several research approaches that are very different from one another. All qualitative approaches, however, share two characteristics. First, they concentrate on phenomena that occur in natural settings, or in the "real world." Second, they entail investigating those phenomena in all of their complexities.

The research goal is to investigate how the impact of social media causes middle-class people to spend a substantial portion of their money unintentionally. This study is qualitative in nature, based on the research objective discussed on Section 1.

3.2 Survey Design and Questionnaires

Survey questionnaires are the most standard practice of collecting data from a sample of people. A questionnaire is a set of questions with a variety of answers. It is also a format that allows for the collection of standardized, relatively structured data about each of a large number of cases. Questionnaires are now widely used in social research at all levels, from small-scale student and community projects to large-scale international surveys. The main thing that these surveys have in common is the formulation of a set of questions and answers that will help the researcher answer his research question or test his hypothesis [41, 42]. The most important things in this type of research is questionnaire design because once the questionnaire is formats, the researcher has determined the questions and answers and will not be able to go back and get more further information. The researchers must be certain that the questions they ask will allow them to gather the necessary data. [43, 44].

A questionnaire may include any or all of the following types of questions:

- Open-ended questions allow for unlimited responses. The questions are followed by plenty of blank space for responses.

- Checklists provide a list of items, and participants are asked to check the ones that apply to their situation.

- Two-way questions limit responses to a pair of alternatives (yes and no).
- Multiple-choice questions have several possible answers, and the participant must choose the one that is most relevant.
- Ranking scales necessitate that the participant rank a list of items.

A questionnaire is normally designed to gather a variety of data types, including: statistics about people or events, descriptions-descriptions of what happened to people Knowledge refers to what people know about something; opinions refer to what they have experienced or know about; attitudes/values refer to how people feel about other people, institutions, ideas, and so on; and background information about the respondent that may be relevant to the research topic [45, 46, 47].

Conducting survey research was a time-consuming process involving paper questionnaires that had to be emailed to sincerely selected samples of prospective respondents drawn from electoral rolls or customized databases. This was followed by a period of hopeful waiting for responses so that they could be manually entered into a database or spreadsheet for analysis [48]. Thanks to automated survey software on online platforms, as well as the use of email and social media to reach respondents, survey data can now be generated quickly and in massive quantities.Because of these advancements, surveys are now so simple to create, complete, and analyze that their use as a research tool has skyrocketed. Online surveys are being conducted by everyone from high school students to professional organizations. However, just because surveys can be used as a methodological tool. There are strict survey methodology conventions that should be followed in order to produce data that are valid, robust, and meaningful; many books have been written on this subject, and a selection of useful guides cited throughout this article is provided in the Reference List. Both casual and research users of online survey methodology frequently overlook two critical components of survey methodology: sample selection and question validation. As a result, data generated by online surveys can be extremely biased, and the results may be difficult to replicate or robust.

The questionnaire for this study will be designed in such a way that all types of questions along with multiple choice questions, will be used. For example, closed-ended questions will be asked to the all users [49]. From the frequent user and non user on social media, 768 filled the data through app and survey. Both in app and survey, questions are asked about the monthly income, monthly savings based on their age, salary and gender. Further they are asked to provide social media usage hours. In our analysis we have collected the expenses in different categories like Restaurant, clothing, tour/travel and entertainment which are strongly influenced by social media. The questionnaires are given below:

1. What is your location?
2. What is your gender?
3. What is your age?
4. What is your profession?
5. What is your monthly income (Approximate)?
6. What is your monthly savings (Approximate)?
7. How much time do you spend on social media daily?
8. Total expense in each month (Approx.) that you think has influence of social media? Submit your entry for each category below.
   - Restaurant
   - Entertainment
   - Clothing
   - Tour / travel
   - other

These questions are asked to identify the spending amount based on some category like: Restaurant; Entertainment; Clothing; Tour/travel; Others. Therefore, the dataset attributes are shown in Table 3.1.

3.3 Data Preparation

Raw data is usually ambiguous and noisy, and it is prone to containing unnecessary and redundant data as well as errors [50]. Also, bad data quality can have a significant

Table 3.1: Dataset Attributes Description

Attribute	Type	Description
Location	String	City location of Bangladesh.
Gender	String	Represent male, female or other.
Age	Number	Age in Years.
Profession	String	Various type of profession e.g. govt job, private job Student.
Monthly Income	Number	Monthly total income.
Monthly Savings	Number	Total savings per month.
Social Media Usage	Number	Time spend each day on social media.
Restaurant	Number	Money spent on restaurant each month having social media influence.
Entertainment	Number	Money spent on entertainment each month having social media influence.
Clothing	Number	Money spent on clothing each month having social media influence.
Travel	Number	Money spent on travel each month having social media influence.
Other	Number	Money spent on other purposes each month having social media influence.

impact on the effectiveness of causal method algorithms [51]. Data preprocessing, a crucial step in data mining procedures, assists in the transformation of raw data into a format that can be understood [52]. Hence, it is critical to examine and preprocess the raw data before using it into the causal models. Data preprocessing includes a variety of tasks such as data cleanup, integration, and data transformation [53].

3.3.1 Data cleaning

The survey was conducted using the Google form. After completing the study over five months, we have exported the raw data (survey entries) from Google form and saved the data in CSV format. For computation, the data has been cleaned and processed. The following steps are used to clean the data.

Dealing with missing data

In the datasets, missing data is prevalent, and it has a significant impact on the final result, making the finding inaccurate. It can also cause a large bias in the outcome model, rendering the analysis conclusion inaccurate. Many ways for dealing with missing data

Table 3.2: Demographic and Financial Characteristic of the Dataset

Factors	Overall (n=768)		Treatment (n=449)		Control (n=319)	
Location	**Freq**	**Percent**	**Freq**	**Percent**	**Freq**	**Percent**
Dhaka	382	49.7%	239	31.1%	143	18.6%
Gazipur	56	7.3%	32	4.2%	24	3.1%
Chittagong	24	3.1%	18	2.3%	6	0.8%
Rajshahi	19	2.4%	7	0.9%	12	1.6%
Bogra	19	2.4%	7	0.9%	12	1.6%
Sirajganj	17	2.2%	14	1.8%	3	0.4%
Rangpur	17	2.2%	4	0.5%	13	1.7%
Khulna	16	2%	5	0.6%	11	1.4%
Meherpur	16	2%	9	1.2%	7	0.9%
Barisal	15	1.9%	6	0.8%	9	1.2%
Jamalpur	12	1.5%	8	1.0%	4	0.5%
Mymensingh	11	1.4%	7	0.9%	4	0.5%
Sylhet	10	1.3%	8	1.0%	2	0.3%
Tangail	9	1.1%	7	0.9%	2	0.3%
Other	145	18.9%	78	51.7%	67	67.2%
Gender	**Freq**	**Percent**	**Freq**	**Percent**	**Freq**	**Percent**
Male	466	60.7%	262	58.4%	204	63.9%
Female	302	39.3%	187	41.6%	115	36.1%
Age Group	**Freq**	**Percent**	**Freq**	**Percent**	**Freq**	**Percent**
0-30	312	40.6%	269	59.9%	43	13.5%
30-40	227	29.6%	108	24.0%	119	37.3%
40-50	156	20.3%	52	11.6%	104	32.6%
50-60	67	8.7%	18	4.0%	49	15.4%
60-80	6	0.8%	2	0.4%	4	1.3%
Profession	**Freq**	**Percent**	**Freq**	**Percent**	**Freq**	**Percent**
Private job	195	25.4%	102	22.7%	93	29.2%
Student	160	20.8%	152	33.9%	8	2.5%
Government Job	159	20.7%	56	12.5%	103	32.3%
Business	143	18.6%	69	15.4%	74	23.2%
Housewife	68	8.9%	37	8.2%	31	9.7%
Self employed	29	3.8%	22	4.9%	7	2.2%
Other	14	1.8%	11	2.4%	3	0.9%
Social Media Use	**Freq**	**Percent**	**Freq**	**Percent**	**Freq**	**Percent**
Less than 1 hour	319	41.5%	0	0	319	100%
2 Hour	87	11.3%	87	19.4%	0	0%
3 Hour	66	8.6%	66	14.7%	0	0%
4 Hour	75	9.8%	75	16.7%	0	0%
5 Hour	85	11.0%	85	18.9%	0	0%
6 hour	63	8.2%	63	14.0%	0	0%
More than 6 Hour	73	9.5%	73	16.3%	0	0%

Note: The the categorical variables are shown as n(%) and the continuous variables are stated as mean value.

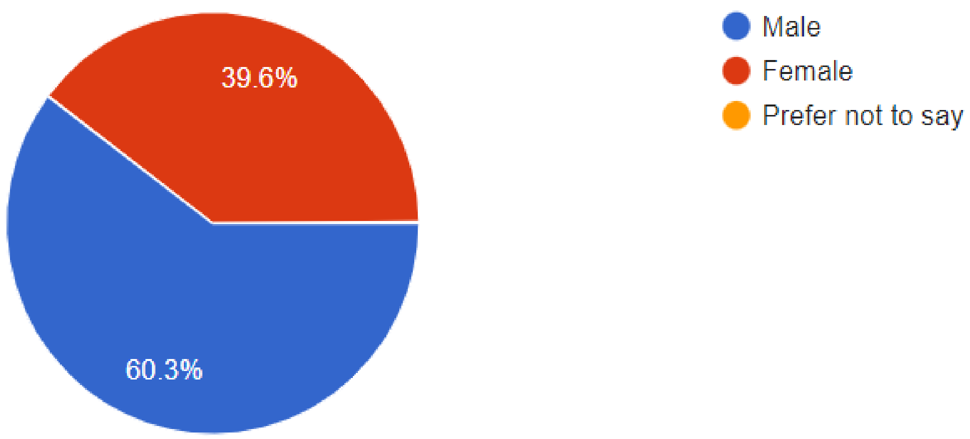


Figure 3.1: Percentage distribution of gender in survey participants

have been explored [54, 55], and they can be split into two categories. Firstly, The easiest and most straightforward option would be to remove the missing data from the dataset. And secondly, use the existing data to fill in or replace the missing data.

We have used Python Pandas [56] library to deal with missing data. In any row in dataset, if there is multiple missing values, the row is removed and if there is only a one or two missing values in a row, we have calculated the mean value and replace the missing value with it.

Dealing with outliers

An outlier is a remote observation point from other observations that can pose serious issues in statistical analysis [57]. However, identifying outliers is difficult. There are some outlier values in the dataset, those are inspected and removed from the dataset. As an example of the outlier in the dataset is some people have unemployed status in profession or student but they have income showing more than 100000 or more. Which is is possible to have but in real life it isn't acceptable in any sense. Thats the reason, those data entries are removed.

3.3.2 Data Transformation

Data preprocessing is the process of turning data into an appropriate format for analysis [58]. There are specific categorical columns in the data that was initially exported.

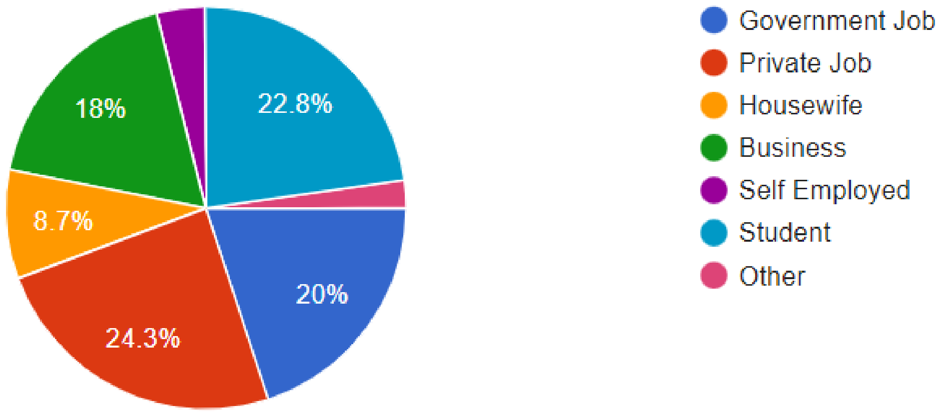


Figure 3.2: Percentage distribution of profession in survey participants

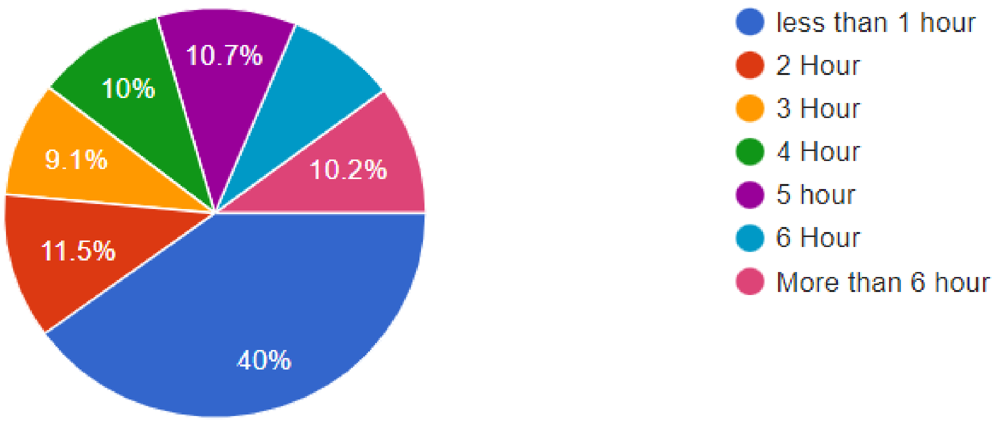


Figure 3.3: Percentage distribution of social medial usage

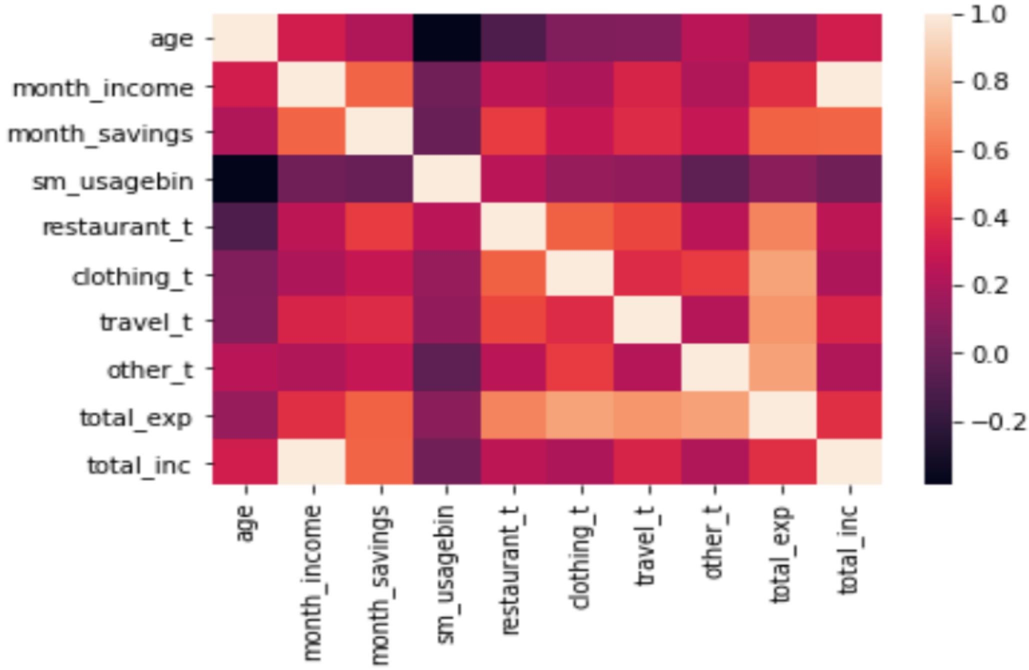


Figure 3.4: Correlation heatmap of dataset

For better results, we converted to numerical numbers. We divide the data into two groups: treatment and control. To do so, we converted the time spent on social media to a binary value. We assigned a value of 0 to people who use social media for less than one hour per day and 1 to those who use it for more than one hour per day. There are five months of observational data in the dataset. We merged the data into a single data frame. A total expense (social media influenced) column is created by combining the entire categorical expense (restaurant, entertainment, clothing, travel, and others) of 5 months of data. The same procedure is followed for income and savings. The characteristic of dataset are shown in Table 3.2.

Chapter 4

Proposal and Research Methodology

4.1 Overview of the Proposal

To analyse the relationship between users' social media activities and spending behavior, our research goals demand a causal model [19]. First, we design a quasi-experiment for our study. The quasi-experimental design includes two groups such as treatment and control [2]. Then, we apply a causal model to identify the effect on the user's spending pattern influenced by social media usage, which is becoming more common. The key components of our causal model are: (1) preparing the dataset, (2) the treatment group, (3) the control group, (4) IPTW and doubly robust estimator for causal effect estimation, and (5) the refuted outcome. Our proposed model is given in Figure 4.1.

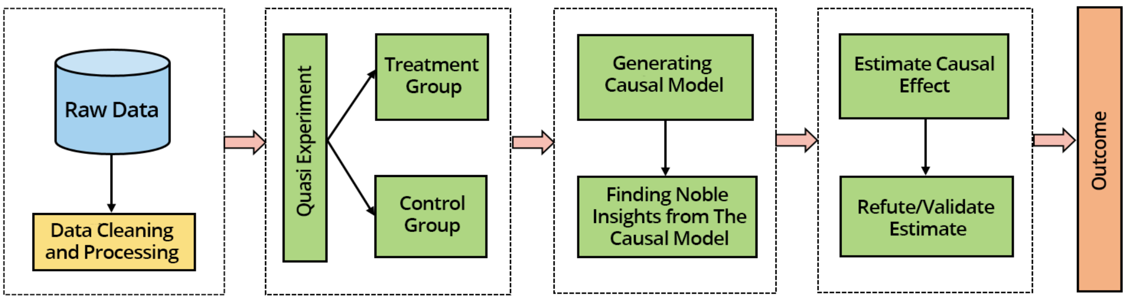


Figure 4.1: The work flow of the Proposal

4.2 Quasi Experiment

We adopted a quasi-experimental approach for this study since applying traditional natural science procedures to the social sciences will be challenging, and conventional natural science methods will be expensive as well [2]. Incorporating this design allows us to reach treatment and control groups where we may compare and contrast the results.

4.2.1 Treatment group

New items or services significantly influence the users in this experiment's treatment group through social media or web pages. We focused on users in the treatment group who spend more than an hour on social media.

4.2.2 Control group

In this study, those who do not use social media or visit web pages form the control group. As a result, social media has little impact on their spending. We focused on users in the control group who spend less than an hour on social media. We use propensity weighting to form the treatment group and control group. In section 4.4, we discuss how we form them.

4.3 Generating Causal Model

To determine the effect of both the treatment (Y_1) and control (Y_0) groups and to find the correlations between common causes, treatment, and outcome, we design a causal graph based on domain knowledge as shown in Figure 4.2. In order to establish causal graph, we focus three key points, respectively those are treatment variable (Y_1, Y_0), outcome variable (T_expense) and covariates X (professions, age, t_income, gender, t_savings).

4.4 Estimate Causal Effect

The result of causal effect depends on covariates X. Any changes in the covariates X causing changes in the outcome (T_expense). The covariates are chosen in such a way

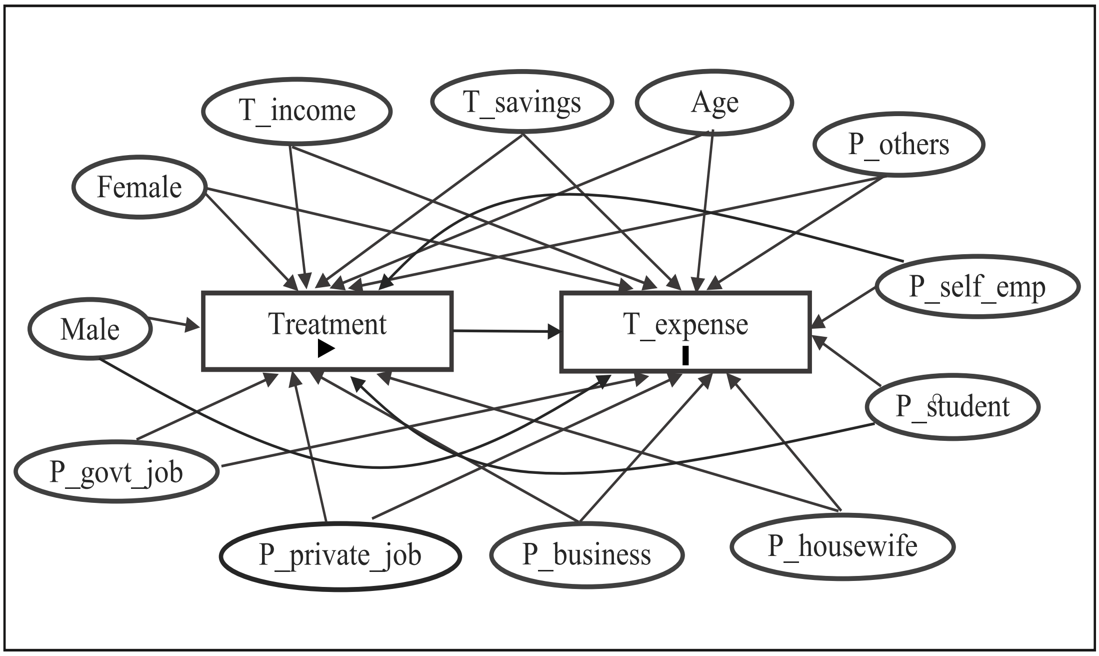


Figure 4.2: Causal Graph

that there is less biases and we get a better effect E. Therefore, to find causal effect (1) we use propensity score based Inverse Probability of Treatment Weighting (IPTW) estimator and (2) Doubly Robust estimator.

4.4.1 Propensity Score Matching (PSM)

Propensity score matching (PSM) introduced by Rosenbaum and Rubin [36] to identifying the effects of social media on the causative influence of personal expenditure. Treated and untreated data need to evaluate the causal influence [2].

Treatment and control groups are initially not comparable because of having more ambitious data. However, if we take two individuals, one from the treated and one from the control, with the same probability of receiving the treatment, they are comparable. Now, propensity score constant makes the data look as nice as random. Because it is the best process to have the balancing data for both groups.

The propensity score forms a middle ground between the covariates X and the treatment T. The equation of propensity score is:

$$(Y_1, Y_0) \perp T | P(x) \tag{4.1}$$

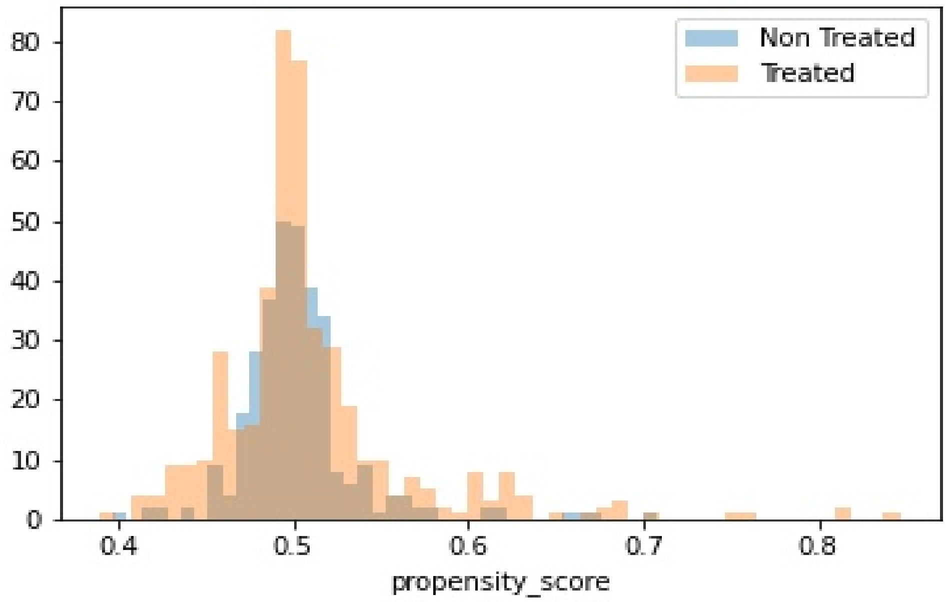


Figure 4.3: Checking balance of treatment and control group

4.4.2 Inverse Probability of Treatment Weighting (IPTW)

After selecting the propensity score, a condition needs to find a balanced treatment and control group. Here comes IPTW [59]. IPTW estimator scales each population unit by the probability of receiving some treatment other than the one it received. The equation of weighting estimator [59] follows:

$$E\Big[YT - P(x)P(x)(1 - P(x))\Big|X\Big] \tag{4.2}$$

This (4.3) becomes our IPTW weighting estimatator when integrated over X

$$E\Big[YT - P(x)P(x)(1 - P(x))\Big] \tag{4.3}$$

Here, the estimator requires that propensity score $P(x)$ and $1 - P(x)$ are larger than zero. Which means that the treated (treatment group) and untreated (control group) distributions must overlap in order for the causal inference to be demonstrated. If these groups do not overlap, it indicates that they are significantly different and that the effect of one group cannot be extrapolated to the other. In the Figure 4.3, it shows the distribution of treated and untreated population.

Here IPTW estimator reconstructs a population where everyone is treated and creates the population where everyone is untreated.

- Original Sample Size 768
- Treatment group Population Sample Size 889
- Control group Population Sample Size 644

4.4.3 Doubly Robust (DR) Estimator

We have also used Doubly Robust (DR) Estimator for calculating causal effect. It can be a more efficient method than the IPTW estimator. Because it is a combination of regression methods and IPTW estimator. It removes some drawbacks of IPTW specially in the case of small propensity score. This estimator gets its name from the fact that it only requires one of the models to be right. Even if the outcome model is incorrect, we will be able to detect the causal influence if the propensity score model is correct. If the result model is correct, we will be able to detect the causal influence even if the propensity score model(PSM) is not correct. In the Figure 4.4, the algorithm of Doubly Robust (DR) estimator are shown.

Algorithm: Doubly Robust (DR) Estimator

Input: $Z_i = \{Y_i, D_i, X_i\}_{i \in N}$

Split sample Z into K random subsets
for *k in {1, ... ,K}* **do**
  assign Sample $S_a = Z \cup S_k$ and S_k
  regress $D_i = \hat{e}\left(X_i\right) + \hat{V}_i$, with $i \in S_a$
  regress $Y_i^0 = \hat{\mu}_0\left(X_i^0\right) + \hat{U}_i^0$, with $i \in S_a|D = 0$
  regress $Y_i^1 = \hat{\mu}_1\left(X_i^1\right) + \hat{U}_i^1$, with $i \in S_a|D = 1$
    estimate $\hat{D}_i = \hat{e}\left(X_i\right)$, with $i \in S_k$
    estimate $\hat{Y}_i^0 = \hat{\mu}_0\left(X_i\right)$, with $i \in S_k$
    estimate $\hat{Y}_i^1 = \hat{\mu}_1\left(X_i\right)$, with $i \in S_k$
  create $\hat{\psi}_{DR,k} = \hat{\mu}_1(x) - \hat{\mu}_0(x) + \dfrac{D\left\{Y - \hat{\mu}_1(x)\right\}}{\hat{e}(x)} - \dfrac{(1-D)\left\{Y - \hat{\mu}_0(x)\right\}}{(1-\hat{e}(x))}$
  store $\hat{\psi}_{DR,k}$ for $i \in S_k$
end
combine $\hat{\psi}_{DR} = \{\hat{\psi}_{DR,1}, \hat{\psi}_{DR,k}, \dots, \hat{\psi}_{DR,K}\}$
Cross-fitting:
for *oob in (1:2)* **do**
  if oob $=1$: $S_{oob} = Z_i$ with $i \in \{1, ..., N/2\}$ and $S_{train} = Z_i \cup S_{oob}$
  if oob $=2$: $S_{train} = Z_i$ with $i \in \{1, ..., N/2\}$ and $S_{oob} = Z_i \cup S_{in}$
  for *l in 1:5* **do**
    split S_{train} in $\{S_1, S_2, \dots, S_5\}$
    regress $\hat{\psi}_i = \hat{t}_{DR}(X_i) + W_i$, for $i \in S_l$
      estimate $\tilde{\tau}_l(X_i) = \hat{t}_{DR}(X_i)$, with $i \in S_{oob}$
  end
  average $\hat{\tau}_{oob}(X_i) = \mathsf{E}[\tilde{\tau}(X_i)]$
end
row bind $\hat{\tau}(X_i) = \{\hat{\tau}_1, \hat{\tau}_2\}$

Figure 4.4: Doubly Robust (DR) Estimator Algorithm

4.5 Refute Estimation

The causal effect estimation is based on the data's statistical estimation, but the causality itself is not based on the data. We test our assumption validity via multiple robustness checks [60]. These are some of the methods available to test our assumptions as follows:

- **Random common cause**: Including an independent random variable in the dataset as a common cause; the estimation should not change if the assumption was true.
- **Data subset refuter**: The dataset replaced by a randomly selected subset; If the assumption was correct, the estimation should not change that much.
- **Unobserved common cause variable**: Introducing an unknown common cause variable to the dataset as a common cause variable; If the assumption was correct, the estimation should not change.

Chapter 5

Result Analysis and Discussion

In this section, we provide the finding from causal model and then present the causal effect and refute estimation outcome accordingly. Here are the stated assumptions based on the dataset:

- *Estimand expression:* (d/d[treatment])(Expectation(t_expense|professions, gender, age, t_income, t_savings))
- *Estimand assumption:* Unconfoundedness: If U→treatment and U→t_expense then P(t_expense|treatment, professions, age, t_income, t_savings, U) = P(t_expense |treatment, professions Job, gender, age, t_income, t_savings)
- *Realized Estimand:* t_expense ∼ treatment+ professions+ gender+ age+ t_income+ t_savings

This estimand or probability expression based on our assumption that we found on the causal graph shown in Figure 4.2.

Figure 4.3, shows the empirical distribution of the propensity score on the untreated and on the treated. It also shows that no one has a propensity score of zero. Now our treatment and control group population are nicely balanced.

Using IPTW estimator discussed in Section 4.4, we have estimated the average treatment effect which is 15762.72. It says that we should expect individuals in treatment group to be 15762.72 standard deviations above than their untreated (control group) fellows.

Table 5.1: Causal effect estimation result based on multiple method

Method	Estimated causal effect
* Inverse Probability of Treatment Weighting (IPTW)	15762.72 tk
*Doubly Robust estimator	12578.65 tk
Linear regression	15196.04 tk
Self-normalized IPS weighting	12427.20 tk

Table 5.2: Refuted result

Refute Method	Original Result	Refuted result
Random Common Cause	15762.72	15150.72
Data Subset Refuter	15762.72	13682.54
Unobserved common cause variable	15762.72	14319.38

We also used doubly robust estimator to find out the causal effect [59]. doubly robust estimator is a way of combining propensity score and linear regression model into a single model. This approach helps us to better identify causal effects. If the propensity score model is correct, even if the resulting model is wrong, we will be able to determine the causal influence. On the other hand, we may also detect the causal influence when the resulting model is correct, even if the propensity score model is wrong. We're using the model on our dataset and it suggests that we should expect individuals in treatment group to be 12578.65 standard deviations above than their untreated (control group) fellows. We have shown detailed result based on multiple method in Table 5.1

5.1 Result Validation

Based on the results of the aforementioned methods, as shown in Table 5.2, It is clear from this example that even when different validation methods are used, the final result of this study remains unchanged. For example, after applying the first method, random common cause, the refuted result is 15150 taka, whereas the original result is the same. When a second method, data subset refuter, is used, the result is around 13682 taka, which is very close to the original result. The same thing happens when the third method, unobserved common cause variable, is used, and the result is around 14319 taka. So, we could agree that our assumption was correct that the social media usage had a causal effect on the individual's spending behavior.

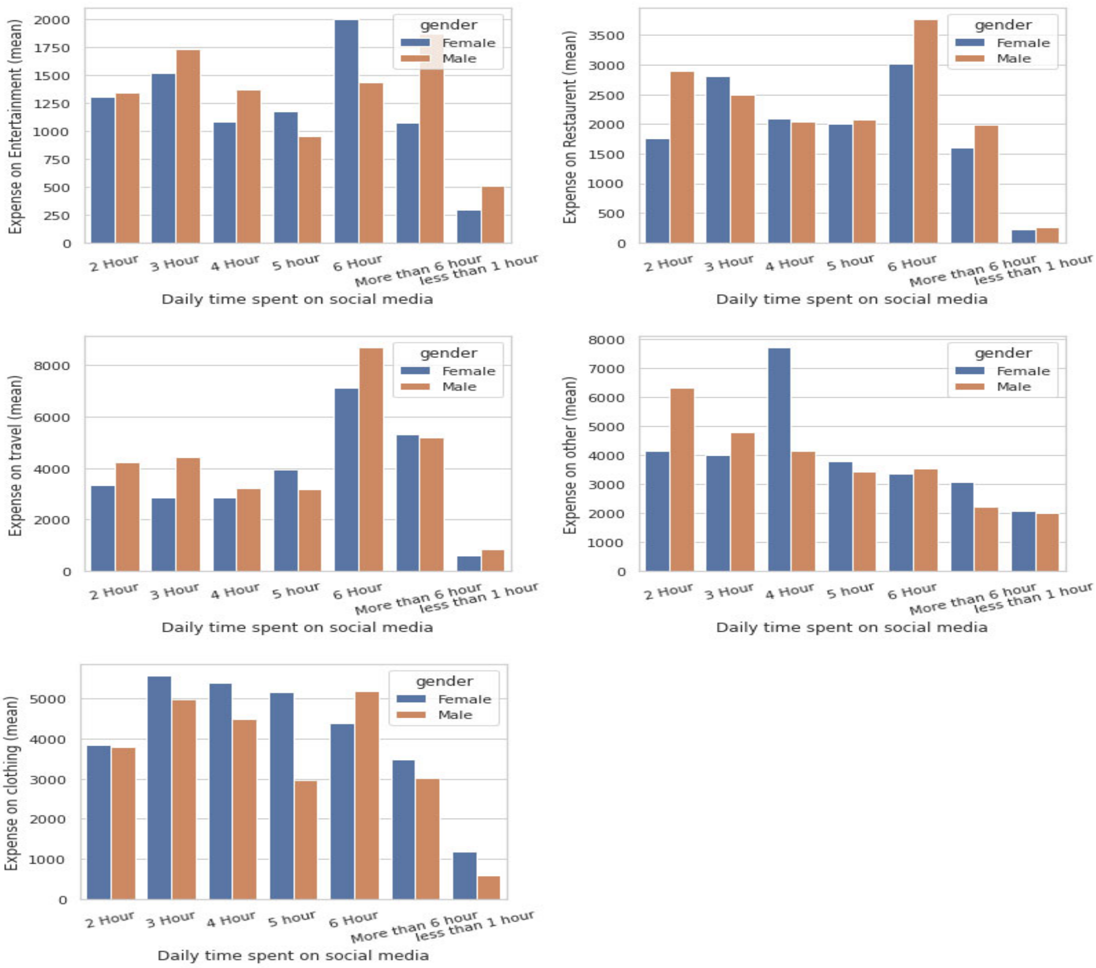


Figure 5.1: Average spending on different categories by gender influenced by social media

5.2 Spending Pattern Analysis

In Figure 5.1, we can see average user spending on different categories influenced by social media. Most of the cases for each category we can see that the user have less social media usage, have less spending on different categories. This work is tring to find the social media influence in users' daily life in different factors. The descriptions of each factors are given below:

- **Entertainment:** This graph shows the relationship between daily time spent on social media and expenditures in the entertainment category. The graph demonstrates whenever a user spends the most time on social media, his or her entertainment expenses are also the highest. However, it is notable that the expense

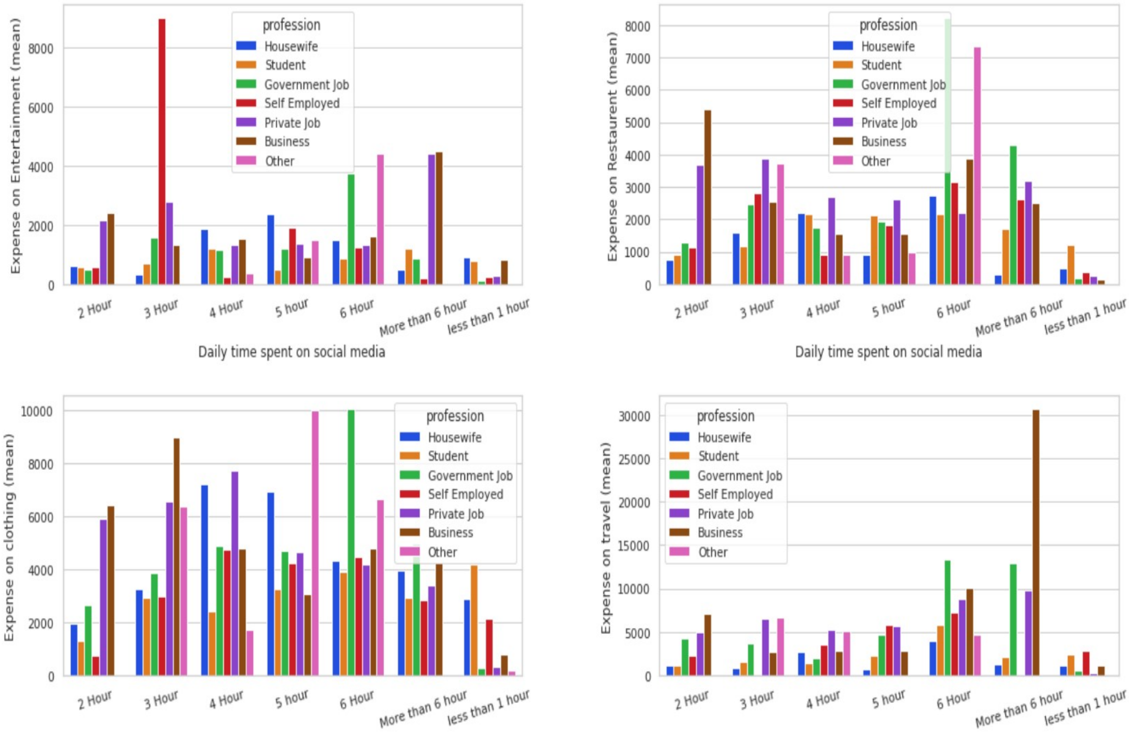


Figure 5.2: Average spending on different categories by profession influenced by social media

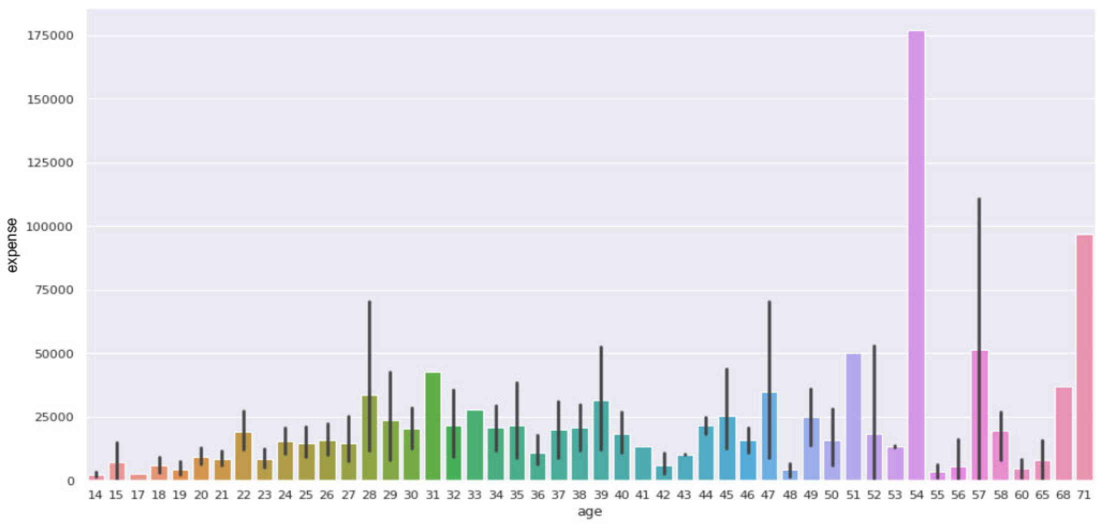


Figure 5.3: Average expense by age

rate of female users in this factor is very high.

- **Restaurant check-ins:** This graph depicts the relationship between daily social media time and restaurant check-ins. Any social media user can view posts or videos of new restaurant check-ins shared by others. People on SM frequently post pictures of food from random places. The percentage of people visiting new restaurants has increased for social media influence as well as the costs. In this graph, however, male users' expense rate is higher than female users'.

- **Travel:** Internet access is widely available in Bangladesh these days. The younger generation, in particular, is accustomed to using a variety of gadgets and taking advantage of various services provided by an internet connection. Many people enjoy traveling extensively whenever they have the opportunity. Social media plays an important role in influencing them more. People are traveling to new and exciting places more because they are connected on social media. Males spend more than females in this category.

- **Clothing:** Buying new clothes is very appealing to people of all ages, so this factor attracts a diverse range of customers. Nowadays, online platforms have captured a significant market share. People are shifting away from offline marketing and toward online shopping. On online platforms, people can buy almost any type of clothing. As a result, social media users who are frequent users benefit more from online shopping in this factor. In this case, female users spend more money than male users.

- **Others:** This graph shows the relationship between other costs that are not included in the above-mentioned factors and the amount of time spent on social media. We can see that female users have a higher cost than male users.

5.3 Discussion

The objectives of our study are to find the influence of social media on user spending behavior. The following research questions resulted from our research objective: RQ1. Whether social media impacts a user's real-life financial activity? RQ2. Whether social media impacts users' personal spending against different social media factors? To investigate these research questions, we created a treatment group and a control group for estimating causal effects between social media usages and users' spending patterns. We divided the dataset into a treatment group and a control group based on social media usage. Those who use social media for less than an hour per day are assigned to the control group, while those who use it for more than an hour per day are assigned to the treatment group. We have balanced the data from both groups using propensity weighting [17] so that they are both comparable. We estimate the average treatment effect using IPTW and a doubly robust estimator. We find that the treatment effect on both groups. The result, shows that the treatment group has a higher score than the control group. We also get a positive average treatment effect (ATE), which indicates that social media has a causal impact on users' spending. The calculation summarizes the findings that are mentioned in RQ1. In Figure 18, we also demonstrate how treatment and control groups react to other social media aspects, including restaurant check-in, clothing, traveling, entertainment, and so on, which is mentioned in RQ2.

Chapter 6

Conclusion

In this study, it used personal spending data from users to examine the impact of social media on several expenditure determinants. The information comes from the online survey. We used a quasi-experimental method to conduct an observational study with two groups: control and treatment. Control groups in the study have no influence from social media, whereas treatment groups have a significant influence from social media. We identified that social networking does have an impact on users' financial activities. Thus we achieved the followings:

- Demonstrated the link between social activities and personal spending.
- Gathered relevant data from social media interactions.
- Conducted a suitable experiment to evaluate the impact of social media on consumer financial habits.

Future Work

Overall, the findings suggest that users spending behavior is influenced by social media, whether consciously or unknowingly. I hope that the research will enable those who are unintentionally influenced by social media to better control their spending and savings. A large-scale financial activity tracker that uses social media data to better run tests and achieve the best outcomes. In future, research should be focused on how social media affects various user aspects as well as long-term social media involvement and its effects. Also, the work can be enhanced to find the pattern of different attribute like:

age and location. The study's findings are limited to Bangladeshi middle-class families. Our study may be carefully extended to other contexts.

References

[1] M. R. Minar and J. Naher, "Violence originated from facebook: A case study in bangladesh," *arXiv preprint arXiv:1804.11241*, 2018.

[2] S. A. M. Falavarjani, F. Zarrinkalam, J. Jovanovic, E. Bagheri, and A. A. Ghorbani, "The reflection of offline activities on users' online social behavior: an observational study," *Information Processing & Management*, vol. 56, no. 6, p. 102070, 2019.

[3] N. J. Mim and S. I. Ahmed, "Others' images: Online social media, architectural improvisations, and spatial marginalization in bangladesh," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–14.

[4] A. Nazir, S. Raza, and C.-N. Chuah, "Unveiling facebook: a measurement study of social network based applications," in *Proceedings of the 8th ACM SIGCOMM conference on Internet measurement*, 2008, pp. 43–56.

[5] D. Nemer, "Online favela: The use of social media by the marginalized in brazil," *Information technology for development*, vol. 22, no. 3, pp. 364–379, 2016.

[6] T. Althoff, P. Jindal, and J. Leskovec, "Online actions with offline impact: How online social networks influence online and offline user behavior," in *Proceedings of the tenth ACM international conference on web search and data mining*, 2017, pp. 537–546.

[7] D. Stück, H. T. Hallgrímsson, G. Ver Steeg, A. Epasto, and L. Foschini, "The spread of physical activity through social networks," in *Proceedings of the 26th International Conference on World Wide Web*, 2017, pp. 519–528.

[8] S. A. M. Falavarjani, J. Jovanovic, H. Fani, A. A. Ghorbani, Z. Noorian, E. Bagheri *et al.*, "On the causal relation between real world activities and emotional expres-

sions of social media users," *Journal of the Association for Information Science & Technology*, vol. 72, no. 6, pp. 723–743, 2021.

[9] H. O. Quarshie and J. Ami-Narh, "The growth and usage of internet in ghana," *Journal of Emerging trends in computing and information sciences*, vol. 3, no. 9, pp. 1302–1308, 2012.

[10] N. Kumar, "Facebook for self-empowerment? a study of facebook adoption in urban india," *New media & society*, vol. 16, no. 7, pp. 1122–1137, 2014.

[11] C.-Y. Li, "How social commerce constructs influence customers' social shopping intention? an empirical study of a social commerce website," *Technological Forecasting and Social Change*, vol. 144, pp. 282–294, 2019.

[12] L.-T. Huang, "Flow and social capital theory in online impulse buying," *Journal of Business Research*, vol. 69, no. 6, pp. 2277–2283, 2016.

[13] S. F. W. Lim, X. Jin, and J. S. Srai, "Consumer-driven e-commerce: A literature review, design framework, and research agenda on last-mile logistics models," *International Journal of Physical Distribution & Logistics Management*, 2018.

[14] Statista. ecommerce bangladesh - market forecast. [Online]. Available: https://www.statista.com/outlook/dmo/ecommerce/bangladesh

[15] M. R. Ward and M. J. Lee, "Internet shopping, consumer search and product branding," *Journal of product & brand management*, 2000.

[16] M. R. Islam, S. Liu, X. Wang, and G. Xu, "Deep learning for misinformation detection on online social networks: a survey and new perspectives," *Social Network Analysis and Mining*, vol. 10, no. 1, pp. 1–20, 2020.

[17] S. Dutta, J. Ma, and M. De Choudhury, "Measuring the impact of anxiety on online social interactions," in *Twelfth international AAAI conference on web and social media*, 2018.

[18] Y. Wang, H. Zhang, T. Feng, and H. Wang, "Does internet use affect levels of depression among older adults in china? a propensity score matching approach," *BMC public health*, vol. 19, no. 1, pp. 1–10, 2019.

[19] K. Saha, B. Sugar, J. Torous, B. Abrahao, E. Kıcıman, and M. De Choudhury, "A social media study on the effects of psychiatric medication use," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 13, 2019, pp. 440–451.

[20] B. G. Buchanan and D. Wright, "The impact of machine learning on uk financial services," *Oxford Review of Economic Policy*, vol. 37, no. 3, pp. 537–563, 2021.

[21] Y. Bachrach, M. Kosinski, T. Graepel, P. Kohli, and D. Stillwell, "Personality and patterns of facebook usage," in *Proceedings of the 4th annual ACM web science conference*, 2012, pp. 24–32.

[22] P. T. Costa and R. R. McCrae, *Neo Pi-R*. Psychological assessment resources Odessa, FL, 1992.

[23] L. R. Goldberg, "The structure of phenotypic personality traits." *American psychologist*, vol. 48, no. 1, p. 26, 1993.

[24] M. T. Russell, R. B. Cattell, A. Cattell, H. E. Cattell, and D. L. Karol, *16PF fifth edition administrator's manual*. Institute for Personality and Ability Testing, Incorporated, 1994.

[25] A. Arora, S. Bansal, C. Kandpal, R. Aswani, and Y. Dwivedi, "Measuring social media influencer index-insights from facebook, twitter and instagram," *Journal of Retailing and Consumer Services*, vol. 49, pp. 86–101, 2019.

[26] G. Razis, I. Anagnostopoulos, and S. Zeadally, "Modeling influence with semantics in social networks: A survey," *ACM Computing Surveys (CSUR)*, vol. 53, no. 1, pp. 1–38, 2020.

[27] P. S. Popescu, M. C. Mihaescu, E. Popescu, and M. Mocanu, "Using ranking and multiple linear regression to explore the impact of social media engagement on student performance," in *2016 IEEE 16th International Conference on Advanced Learning Technologies (ICALT)*. IEEE, 2016, pp. 250–254.

[28] K. Hazelwood, S. Bird, D. Brooks, S. Chintala, U. Diril, D. Dzhulgakov, M. Fawzy, B. Jia, Y. Jia, A. Kalro *et al.*, "Applied machine learning at facebook: A data-

center infrastructure perspective," in *2018 IEEE International Symposium on High Performance Computer Architecture (HPCA).* IEEE, 2018, pp. 620–629.

[29] M. Kabiljo and A. Ilic, "Recommending items to more than a billion people," *Retrieved May*, vol. 2, p. 2018, 2015.

[30] D. Jacob, "Cate meets ml–the conditional average treatment effect and machine learning," *arXiv preprint arXiv:2104.09935*, 2021.

[31] ——, "Cross-fitting and averaging for machine learning estimation of heterogeneous treatment effects," *arXiv preprint arXiv:2007.02852*, 2020.

[32] E. H. Kennedy, "Optimal doubly robust estimation of heterogeneous causal effects," *arXiv preprint arXiv:2004.14497*, 2020.

[33] H. Gotanda, A. K. Jha, G. F. Kominski, and Y. Tsugawa, "Out-of-pocket spending and financial burden among low income adults after medicaid expansions in the united states: quasi-experimental difference-in-difference study," *bmj*, vol. 368, 2020.

[34] S. Tirunillai and G. J. Tellis, "Does offline tv advertising affect online chatter? quasi-experimental analysis using synthetic control," *Marketing Science*, vol. 36, no. 6, pp. 862–878, 2017.

[35] P. C. Austin, "An introduction to propensity score methods for reducing the effects of confounding in observational studies," *Multivariate behavioral research*, vol. 46, no. 3, pp. 399–424, 2011.

[36] P. R. Rosenbaum and D. B. Rubin, "The central role of the propensity score in observational studies for causal effects," *Biometrika*, vol. 70, no. 1, pp. 41–55, 1983.

[37] E. Hargittai and A. Shaw, "Digitally savvy citizenship: The role of internet skills and engagement in young adults' political participation around the 2008 presidential election," *Journal of Broadcasting & Electronic Media*, vol. 57, no. 2, pp. 115–134, 2013.

[38] J. Saldana, *Fundamentals of qualitative research.* Oxford university press, 2011.

[39] M. B. Miles and A. M. Huberman, *Qualitative data analysis: An expanded sourcebook.* sage, 1994.

[40] P. D. Leedy and J. E. Ormrod, *Practical research.* Pearson Custom Saddle River, NJ, USA, 2005, vol. 108.

[41] T. I. Farkhad, "Socio-economic problems of sustainable development," *Economic and Social Development: Book of Proceedings*, pp. 887–893, 2019.

[42] H. Taherdoost, "What is the best response scale for survey and questionnaire design; review of different lengths of rating scale/attitude scale/likert scale," *Hamed Taherdoost*, pp. 1–10, 2019.

[43] R. Matthews and E. Ross, *Research methods: A practical guide for the social sciences.* Pearson Education Ltd, 2010.

[44] J. A. Krosnick, "Survey research," *Annual review of psychology*, vol. 50, no. 1, pp. 537–567, 1999.

[45] M. E. Sanchez, "Effects of questionnaire design on the quality of survey data," *Public opinion quarterly*, vol. 56, no. 2, pp. 206–217, 1992.

[46] K. N. Ross, "Quantitative research methods in educational planning," *International Institute for Educational Planning/Unesco. Retrieved from www. unesco. org/iiep.[Accessed: 8th may 2014]*, 2005.

[47] J. Kahne and B. Bowyer, "The political significance of social media activity and social networks," *Political Communication*, vol. 35, no. 3, pp. 470–493, 2018.

[48] H. L. Ball, "Conducting online surveys," *Journal of human lactation*, vol. 35, no. 3, pp. 413–417, 2019.

[49] D. Plankina, "Suitability of user authentication solutions on mobile devices used by people with dementia," 2011.

[50] M. Kuhn, K. Johnson *et al.*, *Applied predictive modeling.* Springer, 2013, vol. 26.

[51] S. F. Crone, S. Lessmann, and R. Stahlbock, "The impact of preprocessing on data mining: An evaluation of classifier sensitivity in direct marketing," *European Journal of Operational Research*, vol. 173, no. 3, pp. 781–800, 2006.

[52] F. L. Quilumba, W.-J. Lee, H. Huang, D. Y. Wang, and R. Szabados, “An overview of ami data preprocessing to enhance the performance of load forecasting,” in *2014 IEEE Industry Application Society Annual Meeting.* IEEE, 2014, pp. 1–7.

[53] J. Han, J. Pei, and M. Kamber, *Data mining: concepts and techniques.* Elsevier, 2011.

[54] P. Li, E. A. Stuart, and D. B. Allison, “Multiple imputation: a flexible tool for handling missing data,” *Jama*, vol. 314, no. 18, pp. 1966–1967, 2015.

[55] J. W. Graham, P. E. Cumsille, and A. E. Shevock, “Methods for handling missing data.” 2013.

[56] T. pandas development team, “pandas-dev/pandas: Pandas,” Feb. 2020. [Online]. Available: https://doi.org/10.5281/zenodo.3509134

[57] F. E. Grubbs, “Procedures for detecting outlying observations in samples,” *Technometrics*, vol. 11, no. 1, pp. 1–21, 1969.

[58] R. Bro and A. K. Smilde, “Centering and scaling in component analysis,” *Journal of Chemometrics*, vol. 17, no. 1, pp. 16–33, 2003.

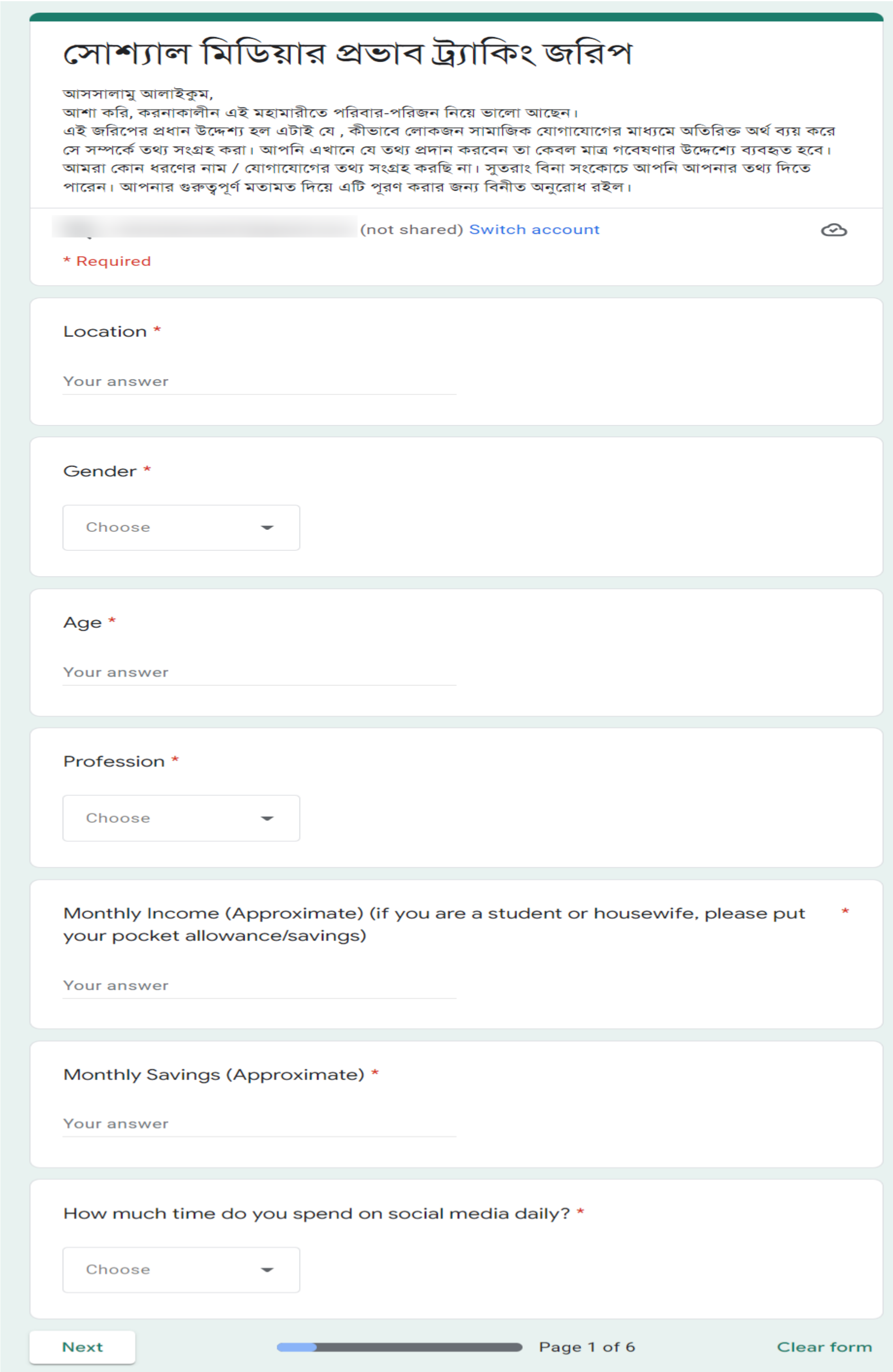
[59] M. Facure and M. Germano, “matheusfacure/python-causality-handbook: First Edition,” Jan. 2021. [Online]. Available: https://doi.org/10.5281/zenodo.4445778

[60] A. Sharma, E. Kiciman *et al.*, “DoWhy: A Python package for causal inference,” https://github.com/microsoft/dowhy, 2019.

Appendix A

Survey Form

This chapter shows the survey form that used to collect data. Later that raw data used to develop the dataset which is used in the causal analysis.



সোশ্যাল মিডিয়ার প্রভাব ট্র্যাকিং জরিপ

আসসালামু আলাইকুম,
আশা করি, করনাকালীন এই মহামারীতে পরিবার-পরিজন নিয়ে ভালো আছেন।
এই জরিপের প্রধান উদ্দেশ্য হল এটাই যে , কীভাবে লোকজন সামাজিক যোগাযোগের মাধ্যমে অতিরিক্ত অর্থ ব্যয় করে সে সম্পর্কে তথ্য সংগ্রহ করা। আপনি এখানে যে তথ্য প্রদান করবেন তা কেবল মাত্র গবেষণার উদ্দেশ্যে ব্যবহৃত হবে। আমরা কোন ধরণের নাম / যোগাযোগের তথ্য সংগ্রহ করছি না। সুতরাং বিনা সংকোচে আপনি আপনার তথ্য দিতে পারেন। আপনার গুরুত্বপূর্ণ মতামত দিয়ে এটি পূরণ করার জন্য বিনীত অনুরোধ রইল।

(not shared) Switch account

* Required

Location *

Your answer

Gender *

Choose

Age *

Your answer

Profession *

Choose

Monthly Income (Approximate) (if you are a student or housewife, please put your pocket allowance/savings) *

Your answer

Monthly Savings (Approximate) *

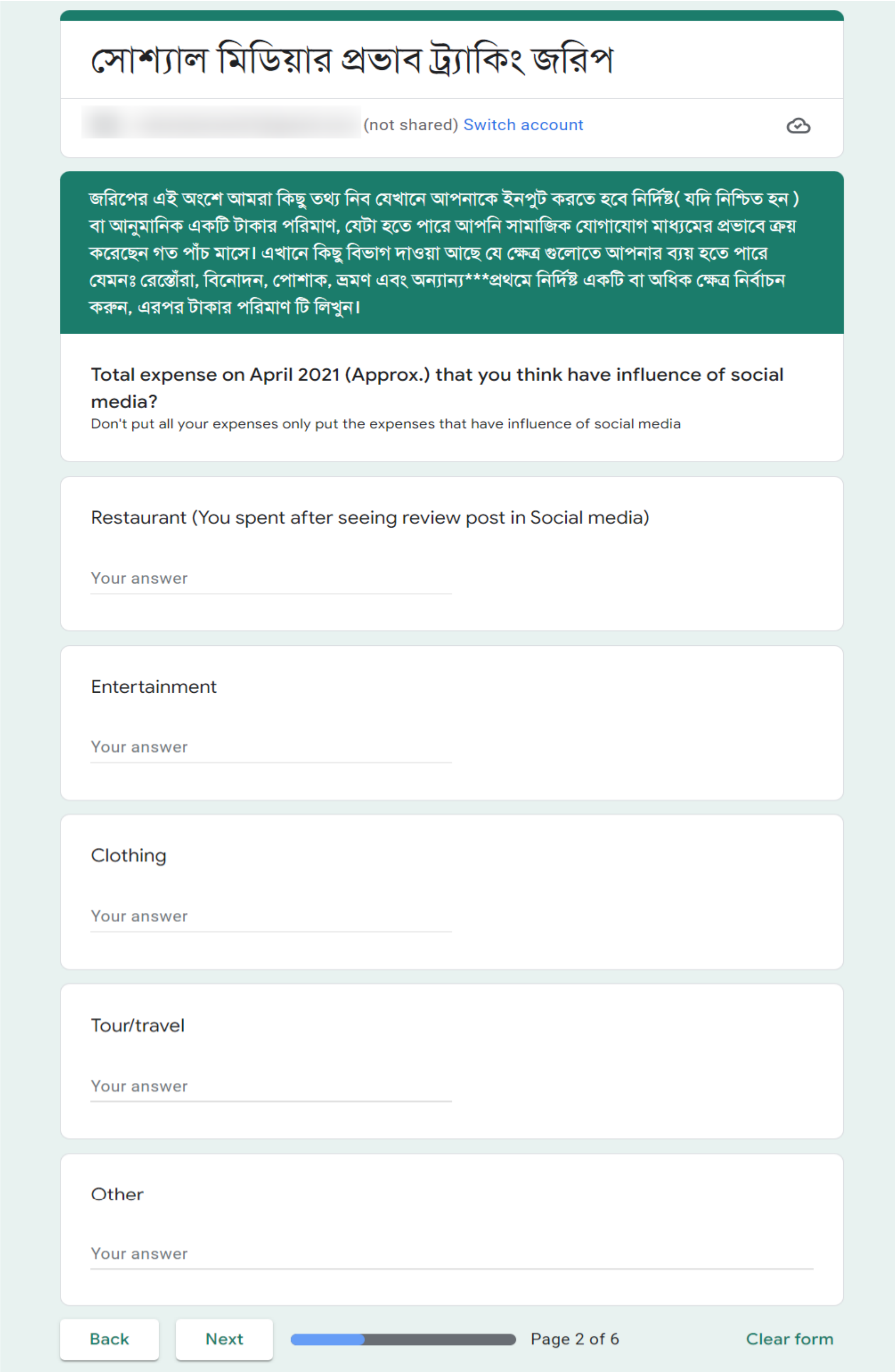
Your answer

How much time do you spend on social media daily? *

Choose

Next | Page 1 of 6 | Clear form

Figure A.1: Survey form first page


সোশ্যাল মিডিয়ার প্রভাব ট্র্যাকিং জরিপ

(not shared) Switch account

জরিপের এই অংশে আমরা কিছু তথ্য নিব যেখানে আপনাকে ইনপুট করতে হবে নির্দিষ্ট(যদি নিশ্চিত হন) বা আনুমানিক একটি টাকার পরিমাণ, যেটা হতে পারে আপনি সামাজিক যোগাযোগ মাধ্যমের প্রভাবে ক্রয় করেছেন গত পাঁচ মাসে। এখানে কিছু বিভাগ দাওয়া আছে যে ক্ষেত্র গুলোতে আপনার ব্যয় হতে পারে যেমনঃ রেস্তোঁরা, বিনোদন, পোশাক, ভ্রমণ এবং অন্যান্য***প্রথমে নির্দিষ্ট একটি বা অধিক ক্ষেত্র নির্বাচন করুন, এরপর টাকার পরিমাণ টি লিখুন।

Total expense on April 2021 (Approx.) that you think have influence of social media?

Don't put all your expenses only put the expenses that have influence of social media

Restaurant (You spent after seeing review post in Social media)

Your answer

Entertainment

Your answer

Clothing

Your answer

Tour/travel

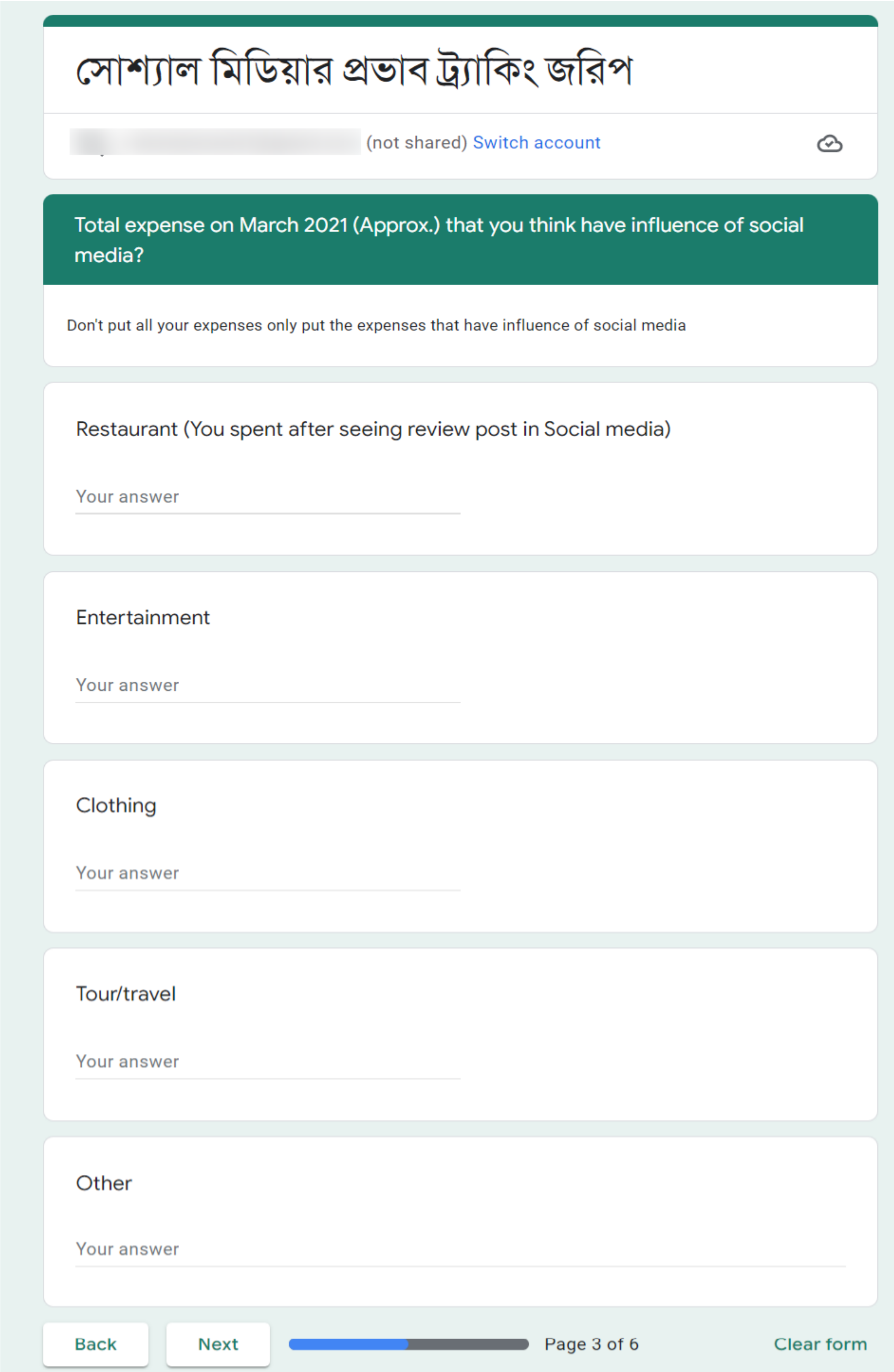
Your answer

Other

Your answer

Back Next Page 2 of 6 Clear form

Figure A.2: Survey form second page


সোশ্যাল মিডিয়ার প্রভাব ট্র্যাকিং জরিপ

(not shared) Switch account

Total expense on March 2021 (Approx.) that you think have influence of social media?

Don't put all your expenses only put the expenses that have influence of social media

Restaurant (You spent after seeing review post in Social media)

Your answer

Entertainment

Your answer

Clothing

Your answer

Tour/travel

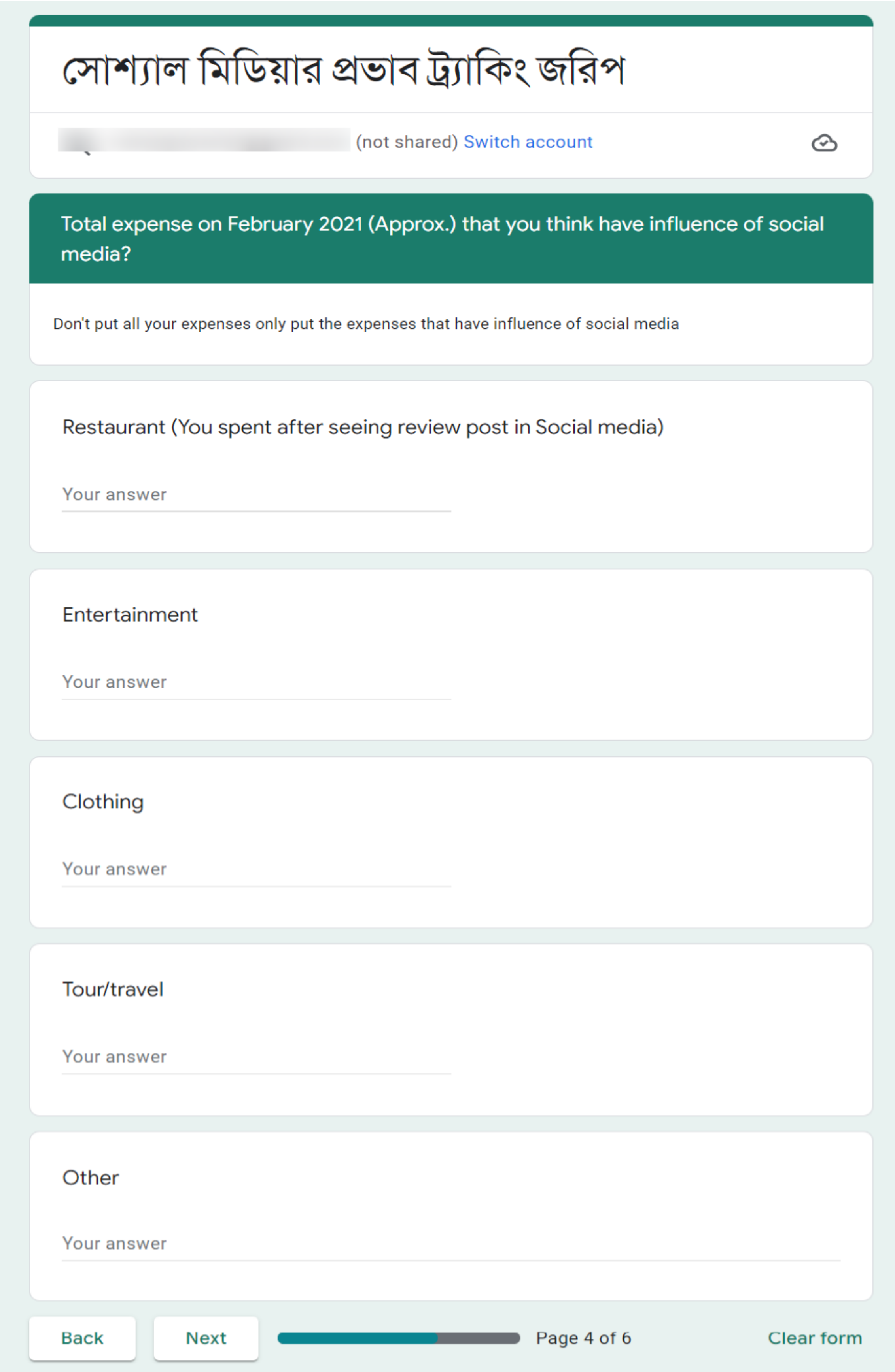
Your answer

Other

Your answer

Back Next Page 3 of 6 Clear form

Figure A.3: Survey form third page


সোশ্যাল মিডিয়ার প্রভাব ট্র্যাকিং জরিপ

(not shared) Switch account

Total expense on February 2021 (Approx.) that you think have influence of social media?

Don't put all your expenses only put the expenses that have influence of social media

Restaurant (You spent after seeing review post in Social media)

Your answer

Entertainment

Your answer

Clothing

Your answer

Tour/travel

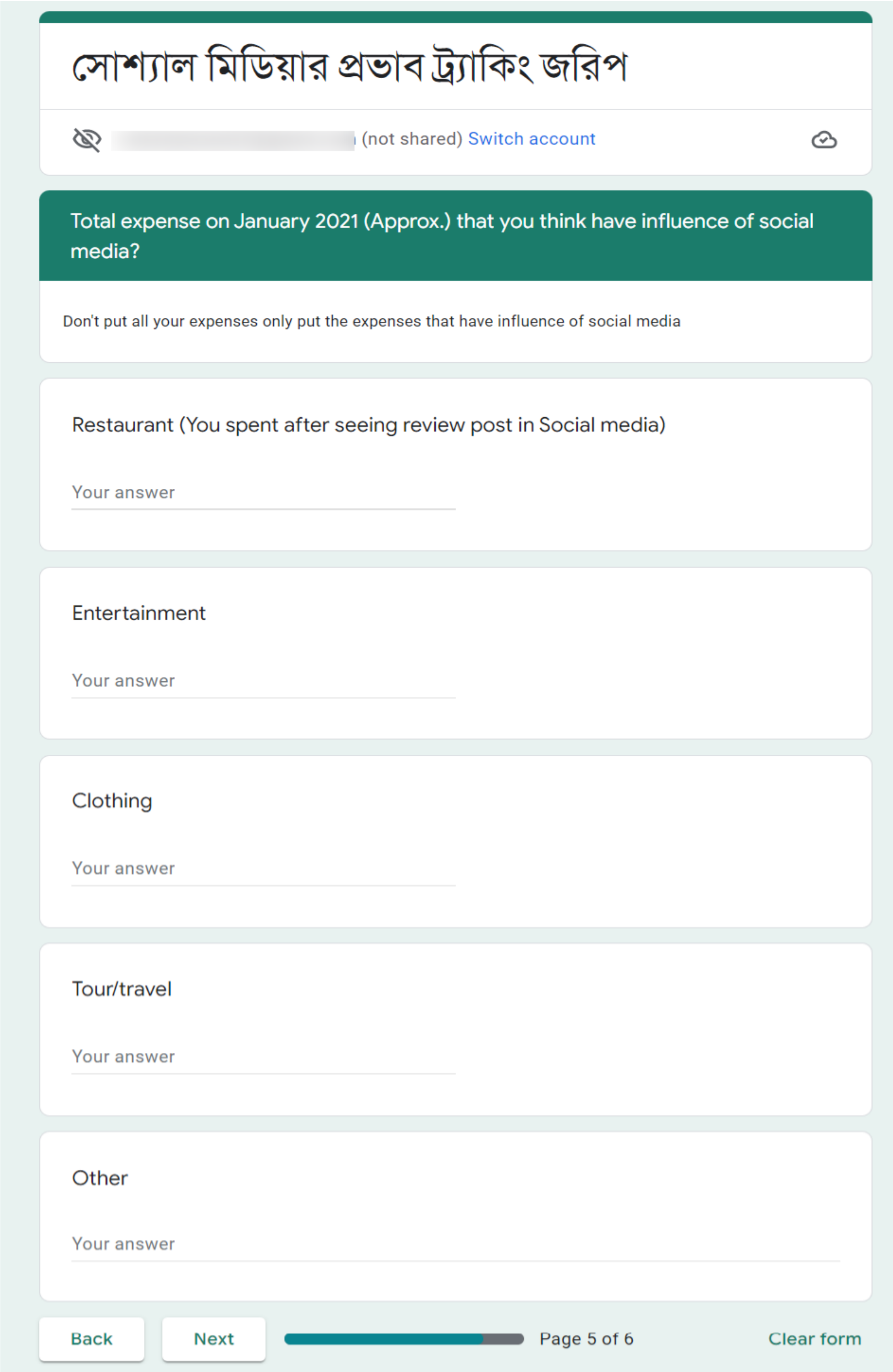
Your answer

Other

Your answer

Back Next Page 4 of 6 Clear form

Figure A.4: Survey form fourth page



সোশ্যাল মিডিয়ার প্রভাব ট্র্যাকিং জরিপ

(not shared) Switch account

Total expense on January 2021 (Approx.) that you think have influence of social media?

Don't put all your expenses only put the expenses that have influence of social media

Restaurant (You spent after seeing review post in Social media)

Your answer

Entertainment

Your answer

Clothing

Your answer

Tour/travel

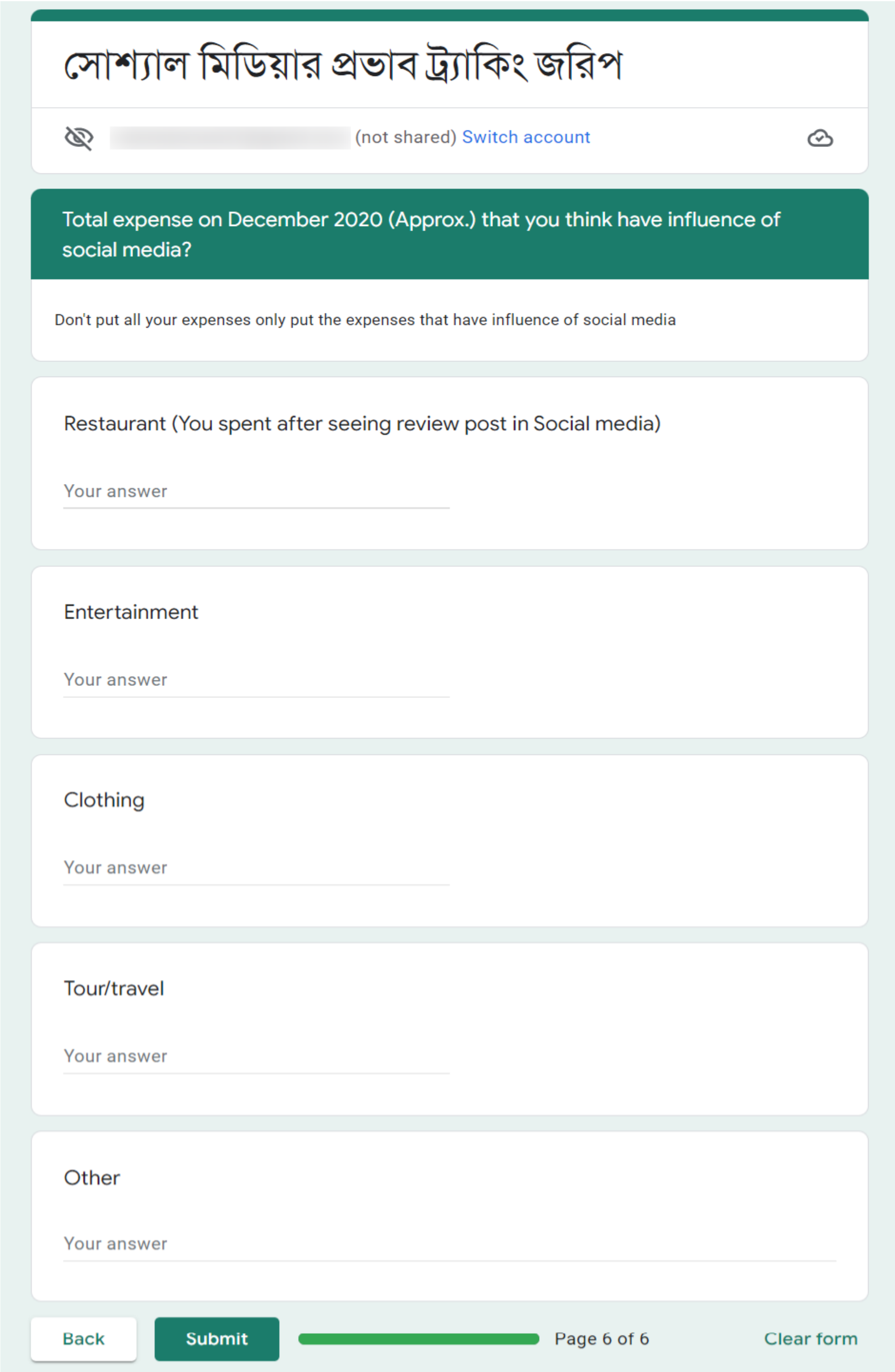
Your answer

Other

Your answer

Back Next Page 5 of 6 Clear form

Figure A.5: Survey form fifth page


সোশ্যাল মিডিয়ার প্রভাব ট্র্যাকিং জরিপ

(not shared) Switch account

Total expense on December 2020 (Approx.) that you think have influence of social media?

Don't put all your expenses only put the expenses that have influence of social media

Restaurant (You spent after seeing review post in Social media)

Your answer

Entertainment

Your answer

Clothing

Your answer

Tour/travel

Your answer

Other

Your answer

Back | Submit | Page 6 of 6 | Clear form

Figure A.6: Survey form last page