Computer Science & Economics

Incentives for Crowd-sourced drivers

Jun Fei Cheung

Supervisors:
Yingjie Fan & Anna van der Meulen

BACHELOR THESIS

Leiden Institute of Advanced Computer Science (LIACS)
www.liacs.leidenuniv.nl 25/08/2023

Abstract

E-commerce is an ever-growing market that continues to grow at a steady pace. Crowdsourced delivery is one of the methods that would ease the demand for fast deliveries. However, crowdsourced delivery startups would struggle to retain their employees and are short on deliverer capacity. The purpose of this study is to measure the effectiveness of incentives for participating as an occasional driver. The investigated factors are financial incentives, experience and reputation by performing an empirical study in the form of a survey. The survey consists of scenario-type questions and demographic questions and was distributed on Prolific. There have been 30 responses in total. Chi-squared test, Fisher's exact test and correlation analysis have been used for our findings. Results have shown that the interaction between the factors of Financial incentives, Experience and Reputation has a statistically significant effect on participants' choices. Crowdshipping industries can leverage the knowledge gained from this study to design more effective incentive structures.

Contents

1 Introduction

E-commerce is an ever-growing market that continues to grow at a steady pace. Previous research conducted large-scale research in e-commerce in retail food stores in North America [MCK, 2022]. The survey has shown that more customers prefer to have their package delivered home than pick it up at a service point. It anticipates that amount of deliveries will continue to grow in the coming years. This pushes companies to deliver more and faster packages. Big companies such as Amazon or Uber explore new strategies to meet their customers' expectations. Crowdsourced delivery is one of the methods that would ease the demand for fast deliveries. A customer places an order on the crowdsourced delivery platform. This order information is sent to the retailer where they have tasks for a delivery driver. This driver uses his own vehicle and picks up the package from the retailer's location. The last-mile delivery is made to the customers by the driver. Afterwards, the driver gets paid for his service. Deliveries would become more efficient in urban cities where the demand for couriers is high [Intelligence, 2019]. Crowdsourced deliveries provide fast deliveries and save high costs for the companies. The drivers have their own transportation for deliveries. Despite the benefits of crowd-sourced delivery, several issues must be handled, such as human resources. Crowdsourced delivery startups would struggle to retain their employees and are short on deliverer capacity. This is due to the irregular wages and unwarranted benefits for working in a crowd-sourced delivery platform [Dolan, 2019]. The crowdsourced delivery model plays a pivotal role in the logistics industry. Crowdsourced delivery startups will need to become more efficient with their supply chain so that costs can be saved while improving customer experience [Intelligence, 2019].

Transportation mode is an important factor that could influence the environment. Having drivers using their own cars would not be as sustainable as other transportation modes for the environment. It would also result in urban cities with less area to park and more traffic [Simoni et al., 2020]. In countries like the Netherlands, the Dutch government came up with a climate policy for the upcoming years till 2030. The policy ensures that the extra carbon emissions will be reduced by 55%-60% [MinisterievanAlgemeneZaken, 2022] by 2030. The Dutch government wants to stimulate more electric vehicles in the market. There will be only electric cars sold from 2030. Its logistic sector also needs to become more sustainable using chain optimisation. Having supplies in logistics hubs outside the city would reduce CO2 emissions by 30%. More available parking spots, optimizing routes and delivering during off-peak hours could also have a positive impact on the environment and traffic [Simoni et al., 2020]. Crowd-sourced delivery is an environmentally friendly option for freight transport with the sheer number of citizens available. The environmental advantage could be even more influential by utilizing existing journeys with minimal additional routes or emission-free transportation. While there is already existing literature in crowd-sourced delivery, there are still areas that need to be further explored [Pourrahmani and Jaller, 2021].

1.1 Objective of the research

The reason for conducting this research is to explore how incentives can recruit more drivers and maintain a high retention rate for the current delivery drivers. Incentives are motivations that can be intrinsic or extrinsic, with intrinsic motivation promoting personal contentment and extrinsic motivation for financial gain. The research could assist many businesses in improving their supply chain using crowd-sourced delivery. The aim of this research is thus to face the challenge of human resources by investigating incentives to attract more delivery drivers. For this research, the following research question is constructed:

Research question

"What is the effectiveness of using incentives for attracting more occasional deliverers?"

1.2 Thesis overview

The structure of the thesis is as follows. This section includes the introduction of this thesis. Section 2 examines the existing literature related to incentives and crowdsourced delivery. Section 3 contains the methodology of how this research is developed. A broad outline of the research approach and the steps involved in the study is presented. Section 4 gives more details about how the survey was designed, which variables were used and the data collection process. Section 5 shows the basic findings in a clear and simple way. Section 6 gives a more detailed statistical analysis of the results. The results are interpreted, and their reliability and validity are discussed, using measures such as Cramer's V. Lastly, Section 7 covers a discussion of the discovered findings and a small summary. It also includes research limitations and suggestions for future work.

2 Literature review

In the domain of crowdsourced delivery, incentives play a crucial role in motivating drivers to participate and perform their tasks effectively. The section defines the concept of incentives in the context of crowdsourced delivery, presents some theoretical foundations and models related to understanding incentives, explores different incentive mechanisms commonly used in crowdsourced delivery and discusses incentives in related areas such as mobile communities or crowdsourcing in general.

2.1 Overview of Incentives in Crowdsourced Delivery

The urge for crowdsourced shipping to fight against the growing amount of freight is strong. Previous research [Miller et al., 2017] investigated the willingness of a traveller to work as an occasional driver. The experiment was an empirical study where participants have to decide between shipping jobs. The study uncovered that certain travellers would be willing to work as crowdshippers [Miller et al., 2017]. The study showed that travel time and profit earned had a diminishing effect on the utility of individuals in their decision-making process. It was implied that jobs with long working hours would require higher payment rates to stimulate the drivers. This could be overcome with cooperative shipping where multiple workers split longer jobs. Results of his study showed that certain travellers such as low-income earners would be willing to work. People who possessed sufficient leisure time, who were tolerant of extra travel time and who rather not work as a team were highly likely to work as an occasional driver. The author suggested that future studies are required to decipher what factors would lead to different travel behaviours. The limitations of his experiment were that it could not be generalized due to the sample size and the non-random sampling.

Prior study [Le and Ukkusuri, 2018] addressed this challenge to provide a sufficient supply of crowdsourced drivers in his study. His research focused on shipping behaviours and the willingness of individuals to work as crowdshippers. Le conducted his survey in the United States and Vietnam. The survey measured participants' preference and satisfaction towards pickup and time location. The article helped in understanding the behaviour of customers as well as potential crowdshippers. It also defined the people who were not willing to work and what the reason was for refusing the job. Results showed that about 80% of the participants in his survey were willing to work. The findings can be used to recruit crowd-shippers and develop business strategies that meet the expectations of both customers and potential workers. Le suggested that targeting a specific group of people could save time and money for delivery companies like Uber Eats. Past research [Le and Ukkusuri, 2018] recommended future studies to conduct a survey with current working drivers to confirm the findings in the past related works. Le suggested investigating the environmental impact of crowd-shipping so that the positive environmental benefits would lead to a successful implementation of crowd-shipping services. Additionally, Le proposed to conduct larger surveys and consider factors like public transport availability, living costs, commercial areas, and urban development. This information can help crowd-shipping companies and local governments plan and implement the service effectively. A follow-up to his research from 2018 unveiled the characteristics of potential new delivery drivers by holding a survey only in the United States this time [Le and Ukkusuri, 2019]. The findings revealed important details about the behaviour, expectations, and personal traits of requesters and potential CS driver-partners. Their findings showed that people with children, part-time jobs or full-time jobs with lower income are very willing to work as a crowdsourced delivery driver [Le and Ukkusuri, 2019]. Furthermore, the author used a mean shift clustering algorithm to have a better grasp of analysing the characteristics of people who lacked the motivation to work. Just like his previous research from 2018, the outcome can be used to improve logistics offerings, draw in potential CS driver-partners, and create business plans that satisfy both customers' and potential CS driver-partners' expectations. Le learned that when he needs to do an extended analysis on this study. He validated that behaviour variations depending on geographic location can be observed with sample size.

A related study [Dietmann et al., 2020] showed how to attract more drivers by discovering what factors were important to them to participate. Dietmann used factors that were already identified in the literature. The factors were divided into extrinsic and intrinsic motivation, customs and way of life, and shipping circumstances. Dietmann conducted an online poll to see how its audience felt about the variables. Afterwards, the author examined the data using cross-classified tables, t-tests and chi-squared tests. The following review of the literature confirmed that factors such as reputation, economic benefits, social motivation, sustainability, enjoyment, free capacity, package size, experience delivering parcels, detours taken and transportation mode had an influence on the driver's intention towards crowd-sourced delivery. However, further research could be done as the study could not draw conclusions for a specific country as his research was conducted cross-country. The article also discussed the implications that arise from not taking into account the population density of urban or rural areas in the experiment. Lastly, the results of Dietmann's research could be validated by a larger sample size.

Shifting the focus to environmental sustainability, researchers [Zalia et al., 2021] investigated the effects of crow logistics from an environmental perspective. The research was conducted in Ghana. The article focused on the behaviour of the drivers who work in crowd logistics. The author attested to the benefits of crowd logistics for the environment. Environmental sustainability was benefited by the movement of people, the form of transportation, and the lesser amount of detours taken. However, participants were required to attain more sustainability. It suggested the capability of crowd logistics for environmental sustainability depends on the crowd's behaviour. Zalio concluded that participants are both intrinsically and extrinsically driven. The implications of the findings were that policymakers could attain a more sustainable environment by creating policies that stimulate an effective engagement of transport behaviour. The survey was conducted in Ghana and failed to give a general representation of a country with high social welfare. The reason for this is that Ghana had a different economic background. The country is still under development and facing challenges like policy formulation or transportation.

Most literature conducted a survey to unwrap the motivation factors of the participant to crowd-sourced delivery. Previous work [Asdecker and Zirkelbach, 2020] used a series of interviews with American drivers from various delivery companies to seek what motivates them to work. Asdecker found that there are different types of incentives that influence their intention to deliver a package. Asdecker procured a model from his work that could be drafted in types of factors. These factors were identified from past literature for his research. He moreover came up with new factors in the model that have not been explored before in the literature. These factors were platform addictiveness and personality types. His approach helped companies with their supply management strategy in order to maintain a high supply of drivers. He further addressed some limitations of the research. For example, the author only interviewed drivers from three crowdsourced delivery platform that drives in urban cities. He also mentioned that he had not explored how drivers with different social, cultural or economic backgrounds could behave differently when motivated to deliver a package. Similar to Dietmann's study in 2020, Asdecker's paper also neglected the inclusion of drivers in rural areas but solely focused on drivers in urban areas.

2.2 Theoretical Framework of Incentives

Incentives are a motivation for you to do something. Motivation can be divided into two types: intrinsic and extrinsic. Intrinsic motivation is the motivation that can give personal contentment. Extrinsic motivation is motivation to do something to gain something, like money. There are several studies about the theory behind incentives [Libretexts, 2020]. One study stated that when the person already has an intrinsic motivation to perform a task, the extrinsic motivation can be utilized for extra stimulation. Research indicated that extrinsic motivation weakens intrinsic motivation and causes it to depend on extrinsic motivation [Eisenberger et al., 1999]. One article supported this conclusion and also proposed that extrinsic rewards improve the labour force and employer profit. It would also fit with the efficiency wage theory. The theory suggests that wages are rigid when employers pay more wages than the marginal product of labour. The motivation arises when the workers are uncertain about their job security [Kreps, 1997]. In contrast, one study even stated that extrinsic motivation would strengthen intrinsic motivation [Cameron and Pierce, 1994]. When a person receives an extrinsic reward such as a compliment for his/her performance, it appeared to have a less negative effect on intrinsic motivation. But the person must not expect any extrinsic reward. It also depends on the personality of the person to what extent extrinsic motivation can affect his/her behaviour [Libretexts, 2020]. Prior study [Coccia, 2019] investigated the differences between the two incentives and their relation in public organisations. It argued that intrinsic and extrinsic incentives can have positive or negative relationships with each other depending on the setting. Both incentives have a strong influence on the employer's performance, and motivation. It suggested reaching an equilibrium between intrinsic and extrinsic for the worker's emotional attachment to the organisation and its job involvement [Coccia, 2019].

2.3 Design and Implementation of Incentives in Crowdsourced Delivery

There are several literature studies that used auction and pricing-based incentive mechanisms for crowdsourced delivery. The incentive mechanisms used extrinsic incentives (monetary rewards) to stimulate workers. Researchers looked at novel incentive systems to boost the effectiveness of mobile crowdsourcing, a fresh take on the standard crowdsourcing model, in order to increase social welfare [Wang et al., 2019]. The mechanism made use of the DPSO algorithm with Gaussian white noise perturbation consists of two steps. The first step was the worker-centric task selection and the second was platform platform-centric worker. The goal of worker-centric task selection was to allocate tasks to a worker so that it was optimally utilized. The platform-centric worker selection chose appropriate workers, who participated through biddings, to do the tasks. It also took into account the unfairness of new drivers who would not get the task due to the trust the veteran drivers have gained from their work experience. The results from a research [Wang et al., 2019] showed it can stimulate the effectiveness of mobile crowd-sourcing. A similar study [Hong et al., 2019] introduced two incentive models, a platform-centric perspective and a worker-centric perspective. The platform-centric model found an optimal reward to motivate the drivers and maximized the profits of the delivery platform using the Stackelberg game theory. Afterwards, the user-centric model chose a driver based on his bidding and delivery distance. Their experiment evaluated its performance and concluded it could be used to reduce traffic congestion and air pollution [Hong et al., 2019]. Finally, one group of researchers suggested a long-term incentive mechanism for quality control in workers, named Melody [Wang et al., 2017]. Their research addressed an issue where many incentive mechanisms were only concentrated on a single set of tasks and its performance was measured by a single run [Wang et al., 2017]. Without taking the quality of information from previous runs into account, it did not accurately display the worker's quality. Whereas in actuality, there were multiple sets of tasks that needed to be continuously allocated to workers. This is where Melody came into play. For each run, there were several tasks where workers could do their bidding price on it. Melody allocated and scheduled the tasks based on several factors. It was found that Melody effectively improves worker quality [Wang et al., 2017].

It came to notice that financial reward was mostly discussed and an important incentive in crowd-sourced delivery. However, as mentioned in Section 2.3, the importance of intrinsic motivation should not be overlooked. In this section, exclusive focus is directed towards related studies that researched intrinsic incentives. A particular study [Feyisetan and Simperl, 2019] focused on making it a competitive game between the drivers through leaderboards, badges or points to see if it can enhance task performance. The investigation led to making his own platform. Workers get paid more by standing on top of the leaderboard which makes them aware of their own productivity. Or they can gain reward just from meeting the conditions of their task givers and performing a certain amount of tasks. The experiment turned out to stimulate the workers effectively and showed that gamification has some potential design improvements in crowd-sourcing delivery platforms [Feyisetan and Simperl, 2019]. A similar work [Klopfenstein and Delpriori, 2019] also took the gamification approach to keep the workers involved and amused. Its research project mostly focused on using incentive schemes in crowdsourcing while considering privacy information protection. One group of researchers investigated the impact of four incentives on consumers in e-commerce delivery, which are information, options order, social media share and social norm [Rai et al., 2021]. The research falls a bit out of the scope but provided inspiration for this research. It was also an interesting way to show that there were incentives not only affecting delivery drivers. The results from their experiment showed that all incentives except option orders have a great impact on consumer behaviour.

Previous research [Gdowska et al., 2018] studied drivers' willingness to participate depending on their compensation fee. He introduced a dynamic system that individually calculates each driver's fee which had an impact on the decision to accept or decline a delivery. The fees were based on historical data. Unfortunately, the author made these fee calculations based on computed simulations. The reason for this was that he had no access to real-world data. His work served as a foundation for developing a dynamic payment package.

Another work [Satrio Wicaksono, 2018] probed the market potential of crowdshipping using bicycles as a transportation mode in the Netherlands. He aimed to maintain a sustainable capacity for both customers and deliverers. This was far different from other literature where they only took either only drivers, the supply side, into account. The research [Satrio Wicaksono, 2018] made use of discrete choice models and presented supply, demand and market equilibrium models to help the companies understand the impact certain attributes have on the market. The article highlighted the delivery time window as an important attribute to encourage customers to choose crowdshipping. Satrio found that the couriers were not sensitive to the amount of money they were compensated for their delivery service. Customers on the other hand were strongly influenced by pricing. This means that they were more likely to choose when the prices were low.

2.4 Incentives in other related fields

Incentives were used in different fields of study such as mobile communities. While this is not directly related to crowdsourced delivery drivers, it highlights the broader spectrum of the intrinsic and extrinsic factors in online activities in China. The insights could be a useful reference point for how the participants interact on their crowdsourced delivery platform. Past research [Al Sukaini et al., 2015] used existing intrinsic and extrinsic motivational factors and explored how these factors carry out their online activities. The article explored factors such as social presence, reputation, enjoyment and creativity. It was interesting that the article confirmed limitations in a research [Asdecker and Zirkelbach, 2020] where the mobile users in China shifted towards individualism. This meant that individual pleasure stimulated users in China to perform online tasks. The author was concerned that the findings of his investigation lacked some generalization considering the study was done in China. He concluded that incentives like enjoyment and social presence are highly influential factors. However, other factors could also be considered of high importance but were not addressed in this experiment.

Another different, yet related and more general field is crowdsourcing in platforms like Amazon Mechanical Turk. People on these platforms are paid very little money for the work they do, even though there are many different types of people participating, including those with high education or working full-time jobs. Prior investigation [Kaufmann et al., 2011] addressed the gap on what motivates people to in participating platforms like Amazon Mechanical Turk and used existing studies on incentive theory. Kaufmann conducted a survey for his experiment on Mechanical Turk. Through his findings, it was uncovered that immediate rewards, delayed rewards and social motivation had a strong influence on how much time they spent on the platforms. However, many participants were also motivated by intrinsic motivations like personal enjoyment, to the extent that they seemed to overwhelm the extrinsic motivations. The findings implied that there was a difference in motivation between workers who do it occasionally and workers who like to work a lot. While the article had a different context, the insights from Kaufmann could be of relevance to this research because it served as a general ground for the motivation of crowdsourced workers.

The literature reviewed sheds light on the existing incentives that have been used before. It is clear that further research needs to be done that could define the type of workers so organizations can achieve better market penetration.

3 Research Question and Methodology

For this approach, a survey was designed and conducted that is in line with the general process of conducting survey research according to [Creswell, 2012]. Firstly, the determination of whether or not the survey is the most suitable design to use was made. For this research, it seemed to be the case, given the purpose was to evaluate the effectiveness of incentives. As mentioned in Section 1, the research question is formulated as followed.

What is the effectiveness of using incentives for attracting more occasional deliverers?

The population were identified as European citizens, who are able to work, for this survey design. The sample size must be determined using a sampling error formula. The survey study followed a cross-sectional approach. The data is collected through a questionnaire. The method was chosen so that the data can be gathered quickly. The design of the web-based questionnaire was achieved using the software program Qualtrics. Questions needed to be clear for the participants to understand so that the variables could be measured. This research had taken many inspirations from [Dietmann et al., 2020].

However, the research model in this experiment was simplified and a custom survey instrument was made for this research. Before the survey was distributed, a pilot test of the questions was performed so further improvements could be made to the survey. The pilot test is a common procedure for questionnaires to evaluate if the participants in the pilot tests are able to complete and understand the questions. These participants could write comments on the survey may be modified the survey if necessary. When the data is gathered, its analysis can be used to approach the research question with an answer.

This research is built upon the study by Dietmann, using similar factors with a different approach. His research used their survey instrument a 7-point Likert scale of questions from strongly disagree to strongly agree. This allows them to measure the level of importance of the factors [Dietmann et al., 2020]. Just like from the existing literature such as [Le and Ukkusuri, 2019] and [Dietmann et al., 2020], the term willingness to participate as a crowdsourced driver was also used as the dependent variable. For the independent variables, only the intrinsic and extrinsic motivational factors such as experience, economic benefits and reputation were looked into. A more detailed explanation of the variables was covered in Section 4.1. After the experiment, the data could be analysed and evaluated for the statistical significance of each factor. The data analysis was performed in Python on Jupyter Notebook and Excel. This enables the use of packages like NumPy for mathematical operations on the array, matplotlib and seaborn for visualizing the data and Pandas for manipulating data. To determine which statistical test is suitable, the type of predictor and outcome variables needed to be identified. Both predictor and outcome variables are categorical. This means that a nonparametric test such as the Chi-square test of independence or Fisher's exact test is suitable. Another possibility was to map the categorical values to binary values. This allows the use of other statistical tests like correlation analysis.

4 Experiment Design (and Data Collection)

4.1 Independent and dependent variables

Compared to the methodology, a more specific and thorough description of the experiment has been given in this section. The decision has been made to use an experimental design for this research. The experiment used a total of three independent variables to test their effect on one dependent variable. More than three independent variables would make the research more complex to investigate. The decision was made to have at least one intrinsic and one extrinsic motivational factor as independent variable. This decision is supported by the literature that was covered in Section 2.2. Having both intrinsic and extrinsic motivation could positively influence someone's decision.

The independent variables X1 'Financial incentives, X2 'Experience' and X3 'Reputation' are implemented in the model. The variable X1 'Financial incentives' is a common extrinsic factor used in the existing literature. This can be simply described as the economic benefit you can get for delivering a package. Related study [Hamari et al., 2016] suggested that economic benefits have a notable impact on an individual's motivation to engage in collaborative consumption. X2 'Experience' was also chosen as the second variable. The term experience could be broadly defined in this research as the enjoyment of driving to deliver a package. X2 'Experience' was categorized as intrinsic motivation. Lastly, X3 'Reputation' was included in the model. The article [Hamari et al., 2016] mentioned that enjoyment is one of the motivational factors in participating in collaborative consumption. Reputation is defined as gaining respect or recognition from the community so the participant's image is improved. It is also the second variable for extrinsic motivational factors. As mentioned before in Section 2.4, a prior study [Al Sukaini et al., 2015] also considered reputation and enjoyment as highly influential factors. The dependent variable Y is the willingness to deliver a package. The research model is visualized in Figure 1.

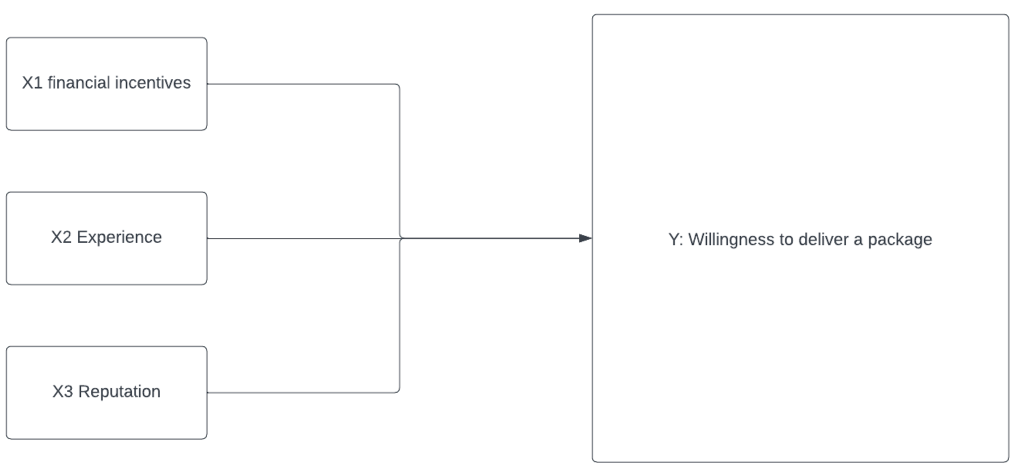

Figure 1: Research Model

4.2 Survey Design

The survey starts with a brief introduction which is then followed by two sections of questions in the survey. The introduction gives you a quick rundown of scenarios and gives you instructions on how to assess different scenario options. In this context, a scenario is a particular scenario containing the independent variables 'Financial incentives, X2 'Experience' and X3 'Reputation' that is presented to participants. The first section of the questions is the part where the independent variables were measured. It contains 14 scenario questions where the participant has to choose between two scenario options, Scenario Option 1 or Scenario Option 2, for each scenario question. Each scenario option contains a combination of the three independent variables given as factors that vary from the other option. The factors could simply be described as positive or negative. The positive is more beneficial than the negative. With three factors where each factor can be positive or negative, there are a total of $2^3 = 8$ possible combinations.

So there are 8 possibilities (n) and a selection between 2 options (k) was required for each question. To calculate the number of questions needed to cover all possible combinations, the combination formula was formulated:

$$C(n,k) = \frac{n!}{k! \cdot (n-k)!}$$

The calculation of $C(8, 2)$ using the combination formula is as follows:

$$C(8,2) = \frac{8!}{2! \cdot (8-2)!} = \frac{8!}{2! \cdot 6!} = \frac{8 \cdot 7}{2 \cdot 1} = 28$$

The combination formula suggested that 28 unique questions were needed to cover all possible combinations of three factors. However, this would require 56 questions due to the nature of the scenarios. To maintain data quality and alertness, the decision was made to not double questions for every question and excluded factors with both positive or negative values in the scenarios. This reduced the number of distinct questions to 13. However, the survey was distributed before completion, resulting in missing questions to measure the interaction between all three factors. The first part of the survey still contains 14 questions, some measuring factors twice, to ensure consistency and reliability. The questions were checked and corrected for clarity and wording. The survey is randomized and measures responses as categorical data on a nominal level. The second part is the social demographic questions to identify the demographic characteristics of the respondents. The survey can be found in Appendix A.

4.3 Demographic variables

For simplicity of the research, the extraneous variables were defined as demographic variables. In the survey, several questions such as gender, age, list of countries, current living area, educational background, mode of transportation and employment status were asked. This implied that 7 questions were required for the demographic variables. These demographic variables would also be participant variables. The variables are important and need to be analyzed so that bias can be minimized. All these demographic questions are on a categorical scale measured except for age. The variable age is measured on an interval scale.

4.4 Sample collection

In order to obtain a sufficient sample Prolific was used. Prolific is a platform where you can conduct your research and gather high-quality data in a short time. Prolific has also easy integrations with Qualtrics. The participants were randomly selected from within Europe. The sample size must be at least 30 to hold the central limit theorem. According to the central limit theorem, if the sample size is increased, then the probability distribution tends to form a bell-shaped curve [Islam, 2018]. This is useful because smaller samples can be used to make reliable predictions about the whole group. Although a larger sample size enhances the study's statistical power and strengthens the validity of the conclusions drawn from the data [Asiamah et al., 2017]. It took less than a day to gather exactly 30 responses for this research. The respondents took on average almost 5 minutes to complete the survey.

5 Results

5.1 Data Pre-processing

When the data was gathered from Qualtrics, the first thing that needed to be done was pre-processing the dataset. The dataset contained additional retrieved data from when the participant was filling in the survey. This additional recorded data frame Qualtrics such as StartDate were deleted. All these columns were omitted as they were not useful for data analysis. The first two rows were also excluded as these were not data points. The index of the dataset started with the start date which was replaced with a standard sequential numeric index. This resulted in having only the data required for the research. Since there were no missing values, there was no need to perform imputation techniques or remove instances.

5.2 Social demographic characteristics

The left-hand graph illustrated the gender distribution of respondents, showing the number of participants categorized as 'Male' and 'Female.' 16 of the 30 participants were male and 14 participants were female. The exact numbers can be found in Table 13. The right graph from Figure 2 depicts the age distribution of respondents. It shows that the most of participants (43.44%) fell in the 18-24 years old category. The next largest group was the group aged between 35 and 44 years (26.67%), followed by those aged 25-34 years (23.33%). A smaller proportion of respondents were aged 55-64 years (6.67%), see Table 14. There were no respondents with the age between 45 and 54 years, under 18 and above 65 years old.

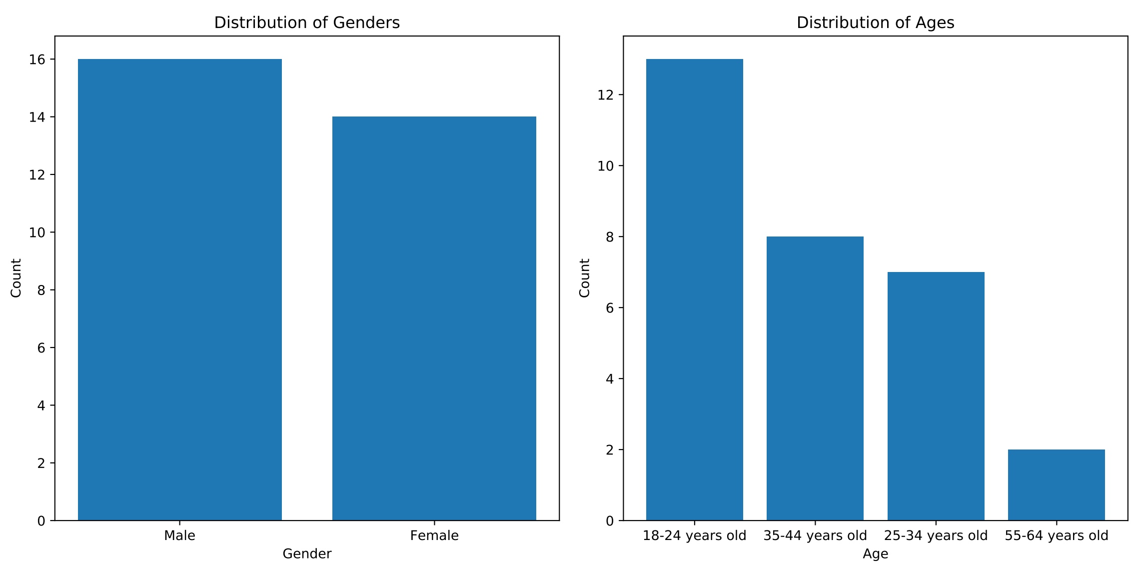

Figure 2: Gender (left) and Age (right) distribution of respondents

Country distribution is shown in the upper graph of Figure 3, where the largest number of respondents (33.66%) were from Portugal, followed by Poland (26.67%) and the UK (10.00%). The rest of the countries account for 3.33% for each of the respondents. The exact numbers can be referred back to in Table 15. The lower graph of Figure 3 presents the distribution of the living areas of the participants, with 25 respondents living in urban areas and 5 respondents in rural areas, see Table 16.

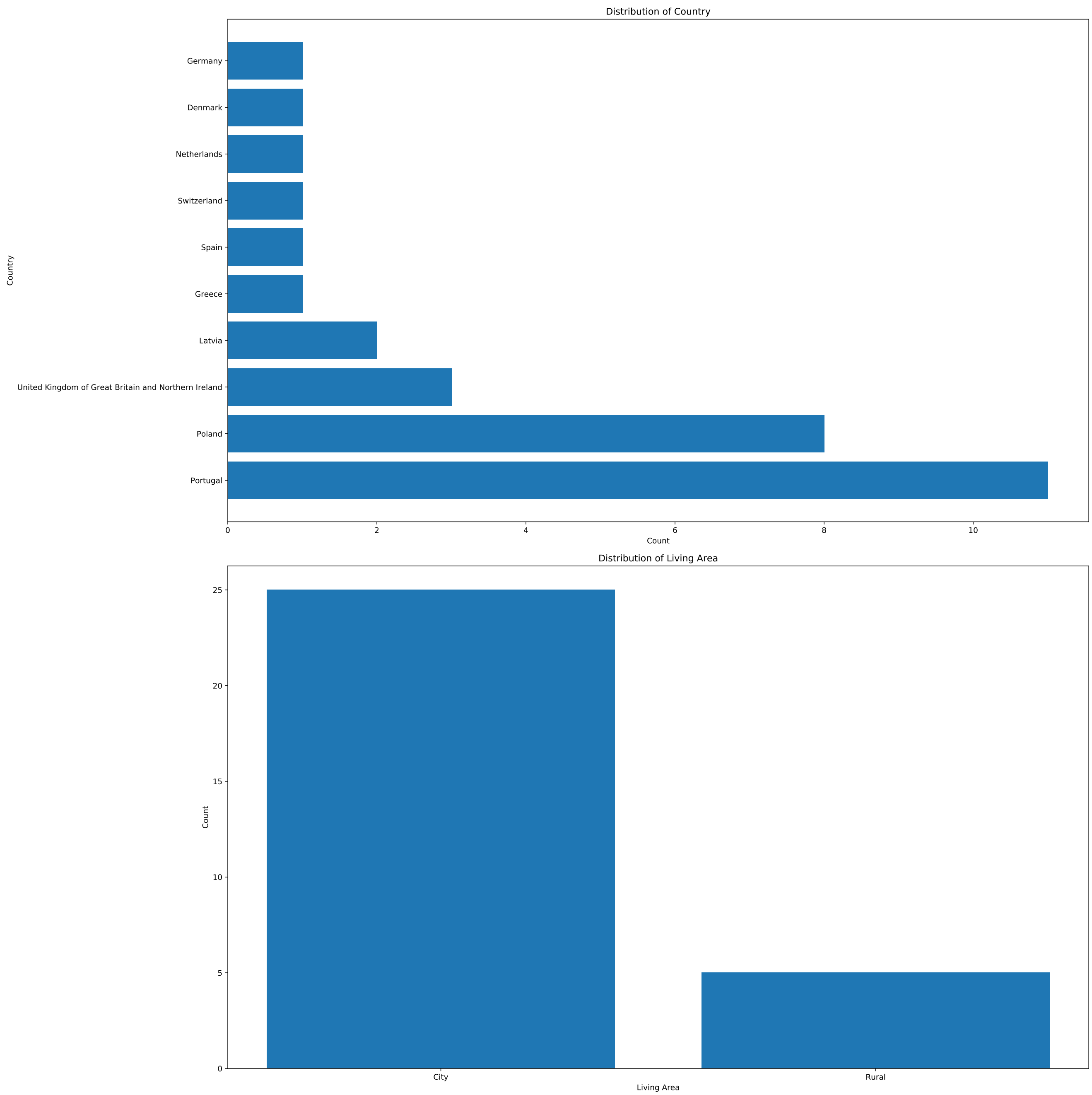

Figure 3: Distribution of respondents based on their countries (above) and living area (below)

The upper graph of Figure 4 displays the Educational background distribution. Half of the respondents hold a Bachelor's degree. A high school diploma or equivalent was the second most common educational background (23.33%). An Associate's degree or vocational certification and a Master's degree both had 4 respondents which accounts for 13.33% each of the total respondents. Specific numbers can be cross-referenced in Table 17. The lower graph of Figure 4 showcases the distribution of respondents based on their employment status, indicating that 50.00% of the respondents were working full-time, 26.67% were students, and 16.67% were working part-time. The remaining respondents represented the rest of the categories like other (6.67%), unemployed and looking for work (6.67%) and homemakers or stay-at-home parents (3.33%).Table 18 contains the exact numerical values.

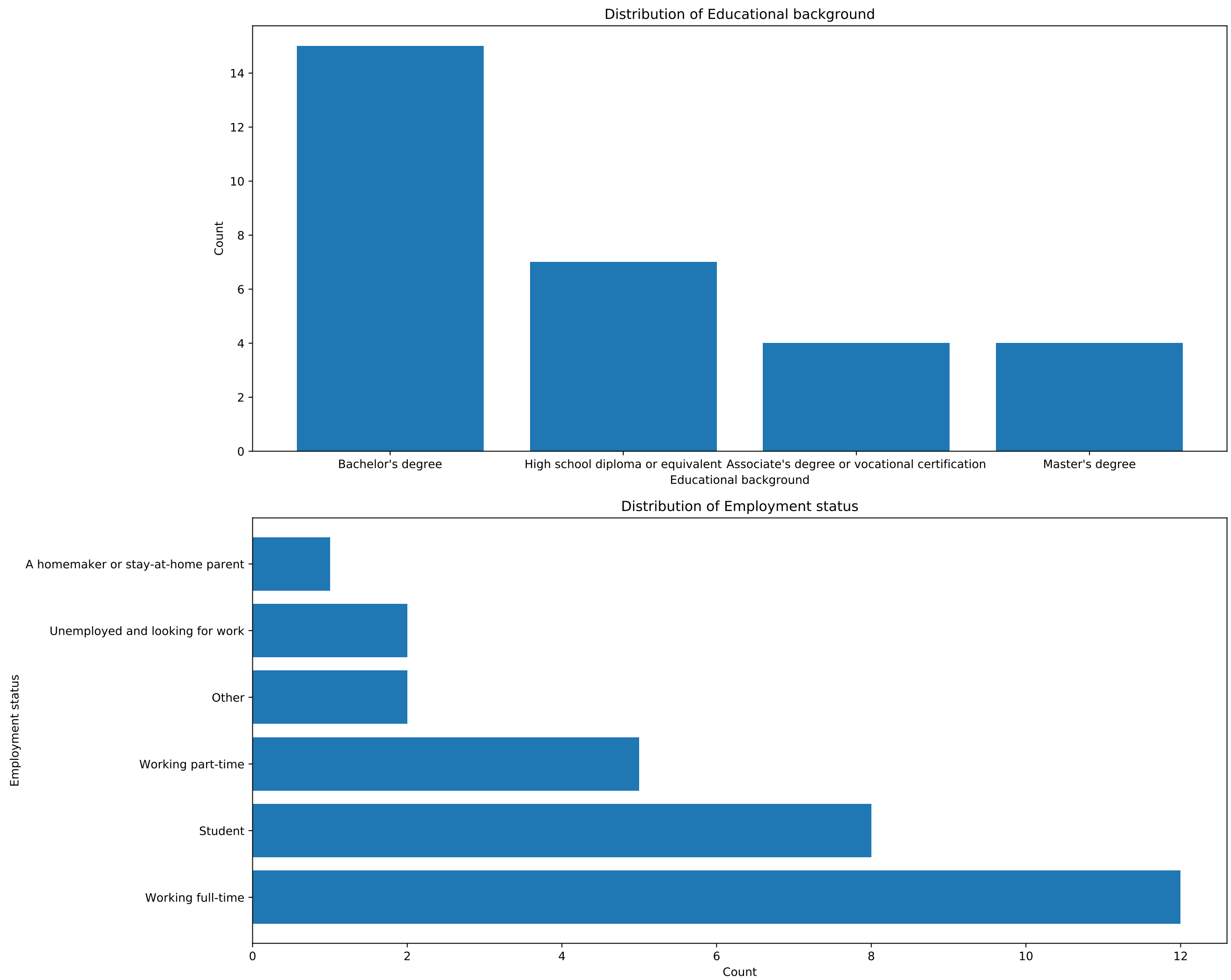

Figure 4: Distribution of respondents based on their educational background (above) and employment status (below)

In Figure 5, the frequency distribution of transportation modes from the respondents is illustrated. The majority use a car to travel (40.00%), followed by public transportation (23.33%) and walking (23.33%). Lastly, only 13.33% uses bicycle as their transportation mode which accounts for 4 participants. For the detailed numbers, please see Table 19.

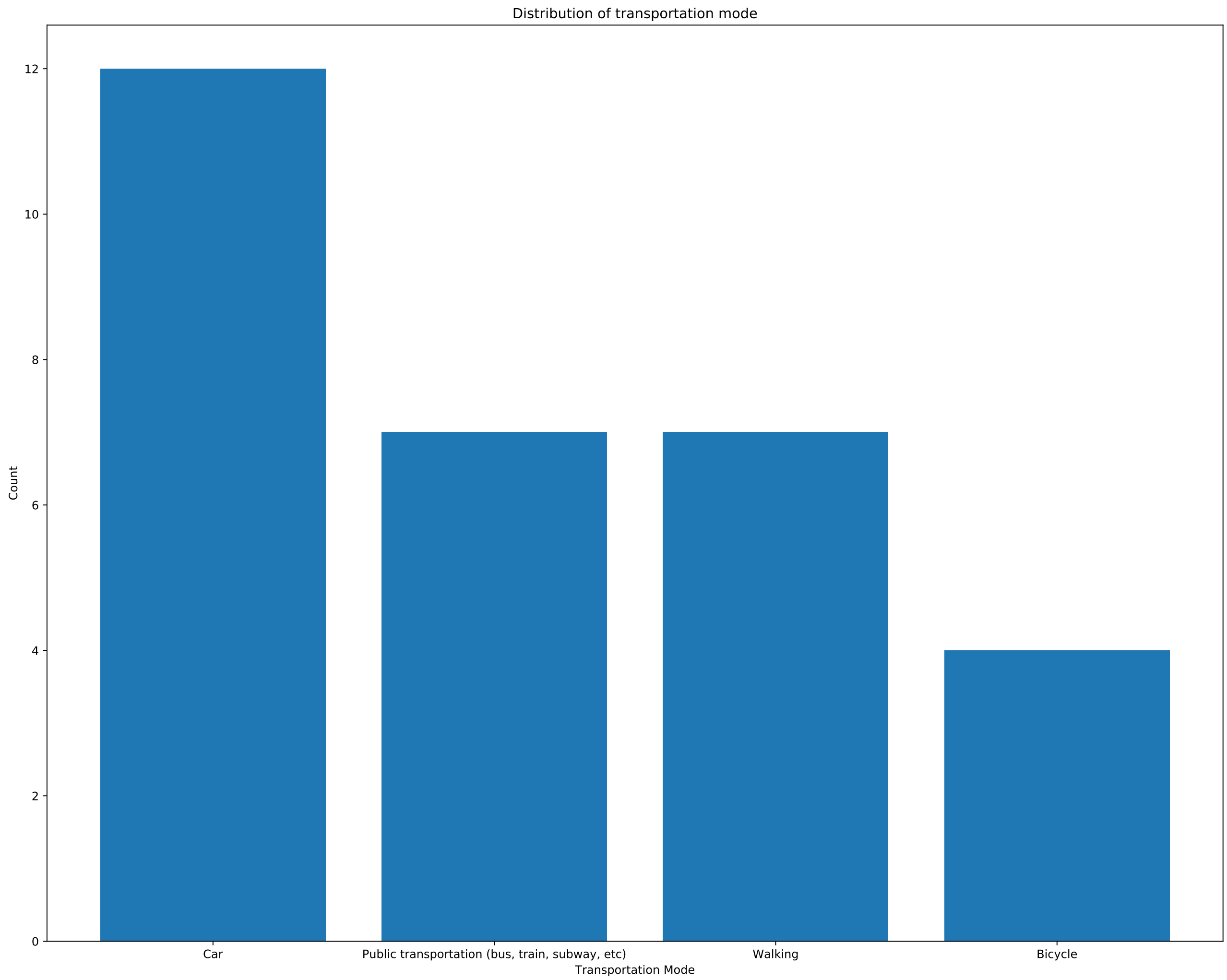

Figure 5: Transportation mode distribution of respondents

5.3 Descriptive characteristics

Most of the data consisted of categorical variables. It can be noted that the responses to the questions were mostly one-sided. There were also duplicate questions. The data have been transformed by combining those duplicate questions, grouping them into the measured factors and assigning scenario labels to the specific measured factors. This transformed data can be found in Table 1. For each group of measured factors, there was notable variation in responses, where the good and bad factors varied across different scenarios.
For each set of measured factors, a notable variation in responses was observed, with differences in positive and negative factors across diverse scenarios.
Financial incentives + Reputation varied the most in terms of responses with 21 participants choosing GoodFI + BadRep and 9 participants opting for BadFI + GoodRep. Experience + Reputation came in second with 22 respondents selecting GoodExp + BadRep and 8 respondents choosing BadExp + GoodRep. GoodFI + BadRep vs BadFI + GoodRep presents the most interesting interaction between factors for investigation.

The highest variability in responses was the choice between GoodFI + BadRep and BadFI + GoodRep. The relationship between this with the demographic variables was explored by performing some cross-tabulations. The results can be found in Table 6, 7, 8, 9, 10, 11 and 12 in Appendix C. There are lots of variations so only the most interesting points are addressed. Table 6 reveals that 10 of the male respondents chose GoodFI + BadRep and 6 chose BadFI + GoodRep. 11 of the 14 female respondents chose GoodFI + BadRep while only 3 chose BadFI + GoodRep. 7 Respondents in the age group of 18-24 years old chose GoodFI + BadRep and 6 respondents chose BadFI + GoodRep. Interestingly, every respondent between the ages of 25 and 34 years old prefers GoodFI + BadRep over BadFI + GoodRep. The majority of the respondents live in the city. Of the 25 respondents who live in the city, 17 of them preferred GoodFI + BadRep. It was observed that respondents who also chose GoodFI + BadRep have a Bachelor's degree (12 respondents), uses work full-time (10 respondents). The part-time workers were more evenly distributed. The same goes for respondents who have a High school diploma or equivalent and a Master's degree. Section 6 dives deeper into the potential relationship between GoodFI + BadRep and BadFI + GoodRep with the demographic variables using statistical tests like the Chi-Squared test and Fisher's exact test.

	Scenario 1	Scenario 2
Experience + Reputation	**GoodExp + BadRep**	**BadExp + GoodRep**
	22	8
	GoodExp	**BadExp**
	59	1
	GoodRep	**BadRep**
	60	0
	GoodExp + GoodRep	**BadExp + BadRep**
	59	1
Financial incentives + Experience	**BadFI + GoodExp**	**GoodFI + BadExp**
	27	3
	GoodFI	**BadFI**
	59	1
	GoodExp	**BadExp**
	59	1
	GoodFI + GoodExp	**GoodFI + GoodExp**
	30	0
Financial incentives + Reputation	**GoodFI + BadRep**	**BadFI + GoodRep**
	21	9
	GoodRep	**BadRep**
	60	10
	GoodFI	**BadFI**
	59	1

Table 1: Organized frequency distribution table

6 Data Analytics

This section is organized in the following order. Section 6.1 investigates the potential relationship between GoodFI + BadRep vs BadFI + GoodRep with demographic variables using the Chi-Squared test of independence. After that, a deeper look was taken into the interaction for each measured group of factors that were found in Table 1. The statistical test was used to find whether changes in factors have a significant impact on participants' choices [Turney, 2023a]. Section 6.1.1 investigated the factors of Experience and Reputation. Section 6.1.2 explores the factors of Financial incentives and Experience. Financial incentives + Reputation will be discussed in Section 6.1.3. The corresponding Cramer's V values were calculated in Section 6.1.4. In Section 6.2, Fisher's exact test has been used to ensure the robustness of findings. Lastly, an observation was done in Section 6.3 to determine whether demographic variables other than countries are related to participant choice in a given scenario.

6.1 Chi squared test

With regard to the categorical variables involved, the chi-squared test was chosen to see if there is any relationship between the variables [Bevans, 2023]. To continue with what was observed from the results in Section 5.3, the relationship between GoodFI + BadRep vs BadFI + GoodRep with the demographic variables was under investigation. Table 2 shows the Chi-squared statistic and the P-values obtained from the nonparametric test. The most common significance level used is 0.05. The obtained P-values are larger than 0.05. However, this meant that the the null hypothesis could not be rejected and that there is no significant relationship between GoodFI + BadRep vs BadFI + GoodRep with the demographic variables. This does not mean there is no relationship at all.

Demographic variable	Chi-squared statistic	P-value
Gender	0.3124999999999999	0.5761501220305788
Age	5.091575091575091	0.1652130737877296
Country	13.279220779220779	0.15037347387096417
Living area	0.0	1.0
Education	2.074829931972789	0.5570215987717599
Employment	6.825396825396825	0.23395299847694478
Transportation	0.7482993197278912	0.861788857059843

Table 2: Summary results of the association between GoodFI+BadRep vs BadFI+GoodRep with the demographic variables

6.1.1 Experience and Reputation

In this section, X2 'Experience' and X3 'Reputation' were measured. The chi-square test of independence was employed according to [Turney, 2023a]. (GoodExp+BadRep vs BadExp+GoodRep) measured the difference between X2 'Experience' and X3 'Reputation'. The following hypotheses were formulated:

Null hypothesis: The differences in factors between X2 'Experience' and X3 'Reputation' have no effect on the participant's choice.

Alternative hypothesis: The differences in factors between X2 'Experience' and X3 'Reputation' have an effect on the participant's choice.

The results are shown in Table 20, 21 and 22. In Table 20, a p-value of 0,0001939416291 was derived along with a chi-square test statistic of 13,889. In Table 21, the chi-square test statistic is 17,561 and has a p-value of 0,00002782401113. In Table 22, the same results were obtained like Table 20. The next step is to find the critical chi-square value. A significance level at α= .05 and 1 degree of freedom were present. The X^2 critical value is 3.841 and remains the same for other tests. The X^2 value in all three tests is greater than the critical value. The null hypothesis can be rejected and it can be inferred that the differences in factors between X2 'Experience' and X3 'Reputation' have an effect on the participant's choice.

6.1.2 Financial incentives and Experience

The factors X1 'Financial incentives' and X2 'Experience' were measured in this section. With regard to these factors, a comparison was made between each scenario's responses with the responses from (BadFI+GoodExp vs GoodFI+BadExp). (BadFI+GoodExp vs GoodFI+BadExp) measured the difference between X2 'Experience' and X1 'Financial incentives'. The following hypotheses were formulated:

Null hypothesis: The differences in factors between X1 'Financial incentives' and X2 'Experience' have no effect on the participant's choice.

Alternative hypothesis: The differences in factors between X1 'Financial incentives' and X2 'Experience' have an effect on the participant's choice.

The results are shown in Table 23, 24 and 25. In Table 23 and 24, a chi-square test statistic of 3,270348837 and a p-value of 0,07054262201 were obtained. In Table 25, the chi-square test statistic is 3,157894737 and has a p-value of 0,07556056753. The X^2 critical value is 3.841. The X^2 value in all three tests is lower than the critical value. The null hypothesis cannot be rejected. This means that there is not enough evidence that the differences in factors between X2 'Experience' and X3 'Reputation' have an effect on the participant's choice.

6.1.3 Financial incentives and Reputation

In this section, X1 'Financial incentives' and X3 'Reputation' were measured. With regard to these factors, a comparison was made between each scenario's responses with the responses from (GoodFI+BadRep vs. BadFI+GoodRep). (GoodFI+BadRep vs BadFI+GoodRep) measured the difference between X1 'Financial incentives' and X3 'Reputation'. The following hypotheses were formulated:

Null hypothesis: The differences in factors between X1 'Financial incentives' and X3 'Reputation' have no effect on the participant's choice.

Alternative hypothesis: The differences in factors between X1 'Financial incentives' and X3 'Reputation' have an effect on the participant's choice.

The results are shown in Table 26 and 27. A chi-square test statistic of 20 and a p-value of 0,000007744216431 were obtained in Table 26. In Table 27, the chi-square test statistic is 16,25625 and has a p-value of 0,00005532678421. The X^2 critical value is 3.841. The X^2 value in all two tests is greater than the critical value. Therefore, The null hypothesis can be rejected and it can be inferred that the differences in factors between X1 'Financial incentives' and X3 'Reputation' have an effect on the participant's choice.

6.1.4 Cramer's V

It can be inferred that there may be a statistical significance in the relationship between the factors. The next step was to consider the practical significance of a relationship. This enabled the confirmation whether the observed relationship is significant enough to be relevant in the actual world. Cramer's V was used to measure how strong the associations are between the investigated factors. The value lies between 0 and 1 and indicates the strength of the association [Merkus, 2022]. The formula to calculate Cramer's V is:

$$V = \sqrt{\frac{\chi^2}{n \cdot \min(r-1, c-1)}}$$

V stands for the Cramer's V value, and n denotes the amount of observed values. r denotes for amount of rows and c for amount of columns. Table 3 was used to interpret the results [Lee, 2016]. It can be interpreted from the results in Table 4 that there was a relatively strong association between Financial incentives and reputation. The same could be said for Experience and Reputation. There was a moderate association between Financial incentives and Experience. Unfortunately, the results cannot speak about the direction of the association. It is important to keep in mind that a strong association does not imply a causal relationship [Merkus, 2022].

Estimated values	Interpretation of association
0.00-0.10	Negligible
0.10-0.20	Weak
0.20-0.40	Moderate
0.40-0.60	Relatively strong
0.60-0.80	Strong
0.80-1.00	Very strong

Table 3: Interpretation of Cramer's V Source: [Lee, 2016]

Comparisons between factors	Cramer's V value
Experience vs (GoodExp+BadRep vs BadExp+GoodRep)	0,3928371007
Reputation vs (GoodExp+BadRep vs BadExp+GoodRep)	0,4417261043
(GoodExp+GoodRep vs BadExp+BadRep) vs (GoodExp+BadRep vs BadExp+GoodRep)	0,3928371007
Average	0.4091334352
Financial incentives vs (BadFI+GoodExp vs GoodFI+BadExp)	0,1906232129
Experience vs (BadFI+GoodExp vs GoodFI+BadExp)	0,1906232129
(GoodFI+GoodExp vs BadFI+BadExp) vs (BadFI+GoodExp vs GoodFI+BadExp)	0,2294157339
Average	0.2035540532
Financial incentives vs (GoodFI+BadRep vs BadFI+GoodRep)	0,4714045208
Reputation vs (GoodFI+BadRep vs BadFI+GoodRep)	0,425
Average	0.4482022604

Table 4: Cramer's V values obtained

6.2 Fisher's exact test

Fisher's exact test can also be an appropriate statistical test since the sample size is small. However, Fisher's exact test requires a 2 X 2 contingency table and is not suitable when variables have not 2 levels. The results from Fisher's exact test along with the results from the Chi-squared test for convenient comparisons can be found in Table 5. Fisher's exact test showed supportive results with the findings from the chi-squared test results.

Comparison	Chi-squared test	P-Values	OddsRatio	Fisher's exact test p-value
Experience vs (GoodExp+BadRep vs BadExp+GoodRep)	13,88888889	0,0001939416291	0.046610169491525424	0.0005174957047006518
Reputation vs (GoodExp+BadRep vs BadExp+GoodRep)	17,56097561	0,00002782401113	0	7.550649072144741e-05
(GoodExp+GoodRep vs BadExp+BadRep) vs (GoodExp+BadRep vs BadExp+GoodRep)	13,88888889	0,0001939416291	0.046610169491525424	0.0005174957047006518
Financial incentives vs (BadFI+GoodExp vs GoodFI+BadExp)	3,270348837	0,07054262201	0.15254237288135594	0.1060606060606063
Experience vs (BadFI+GoodExp vs GoodFI+BadExp)	3,270348837	0,07054262201	0.15254237288135594	0.1060606060606063
(GoodFI+GoodExp vs BadFI+BadExp) vs (BadFI+GoodExp vs GoodFI+BadExp)	3,157894737	0,07556056753	0	0.23728813559321998
Financial incentives vs (GoodFI+BadRep vs BadFI+GoodRep)	20	0,000007744216431	0	2.025783897404681e-05
Reputation vs (GoodFI+BadRep vs BadFI+GoodRep)	16,25625	0,00005532678421	0.03954802259887006	0.00015531009880102672

Table 5: Comparison results between Chi-squared test and Fisher's exact test

6.3 Correlation analysis

An investigation was done to decipher whether the demographic variables excluding countries are correlated with the participant choice in a certain condition. Correlation analysis works better with numerical values. Since most variables were considered categorical. Categorical values were converted into numerical ones. The Pearson correlation coefficient was employed. It was used to measure the strength and direction of the relationship between two variables. The correlation coefficient ranges from 1 to -1, with 1 being a perfect positive correlation, 0 representing no correlation, and -1 representing a perfect negative correlation. The table from a research [Turney, 2023b] was used to interpret in terms of relationship strength. An attempt was made to explore different scenarios by uncovering any potential relationship between the scenarios and demographic variables. Figure 6 illustrates that almost all cells have either a positive weak relationship or a weak negative relationship. There was a moderate negative relationship between Age and (BadExp+BadRep vs GoodExp+GoodRep). In short, it could be suggested that there could be no significant linear relationship between those variables and scenarios. Further examination can be done, by calculating the t-statistic, to see whether the relationship was significant. The correlation coefficient and sample size of 30 are known. The common significance level of 0.05 was used. The calculated critical value of t was 2.048407141795244. In Figure 7, the t-value was computed for every correlation coefficient. Only Age and (BadExp+BadRep vs GoodExp+GoodRep) showed that the t-value is greater than the critical value, meaning the relationship is statistically significant. Thus it suggests a moderate negative relationship meaning that as age increases, people are less likely to choose for 'bad' Experience and Reputation.

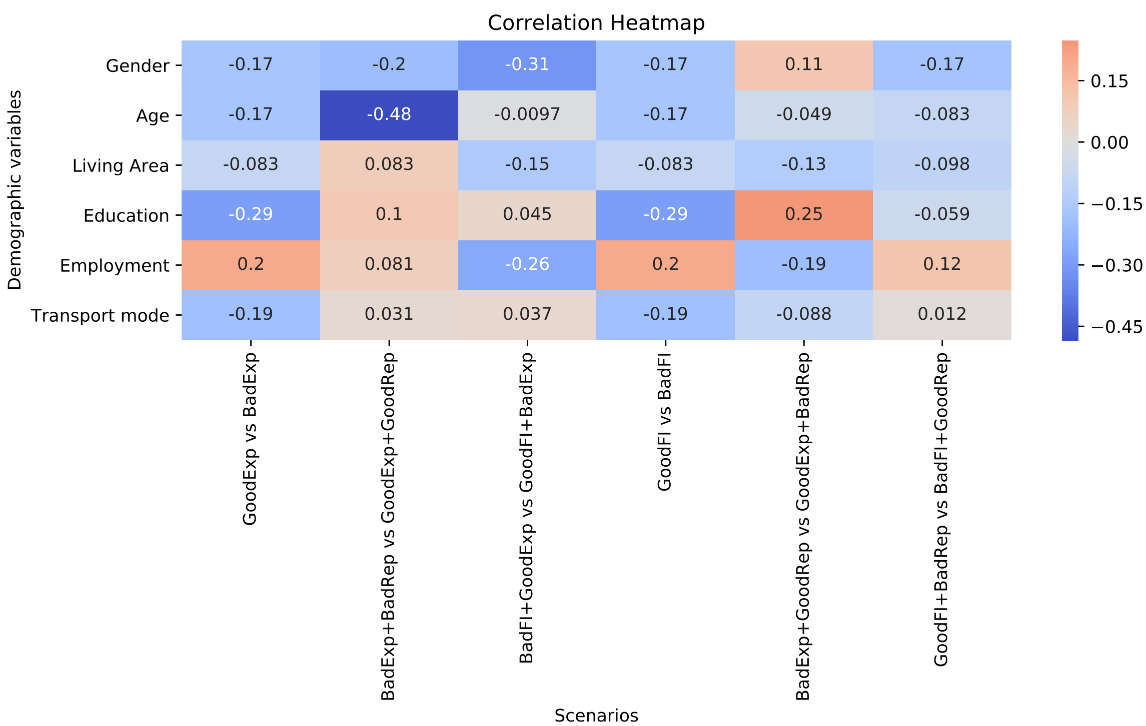

Figure 6: Correlation matrix between different scenarios and the demographic variables

7 Conclusion

7.1 Discussion

The purpose of this research is to discover the effectiveness of incentives for crowdsourced deliverers. The results of this study can be used to attract more participants to deliver a package. From what was gathered, the findings can discussed and summarized further in this section. The interaction between the factors of Financial Incentives, Experience, and Reputation showed a significant impact on participants' choices. This suggests that a combination of these factors plays a pivotal role in decision-making. There were relatively strong associations between Experience and Reputation observed, as well as between Financial Incentives and Reputation, see Table 4. This indicates that these factors tend to influence one another, possibly creating a synergistic effect. The obtained p-values in Table 5 indicated that there is statistical evidence that the differences in factors have an effect on the participant's choice. Unfortunately, the comparisons between Financial incentives and Experience showed a weaker but still moderate association. Nonetheless, it may not be statistically significant given the observed corresponding p-values. These findings provided valuable insights for the crowdshipping industry. By understanding the interaction between factors, companies can design incentive structures that align better with potential participants. Practitioners can emphasize not only Financial Incentives but also Experience and Reputation. An integrated approach that addresses these three factors could attract more occasional drivers to participate or retain their current workforce. In the correlation analysis, most of the relationships between demographic variables and factors were positively/negatively weak. Interestingly, a moderate negative correlation has been found between Experience and Reputation with age. This implied that older participants might prioritize different aspects when making their choices. It can be assumed that they are more cautious and risk-averse as they age. Other certain demographic factors could still play a role in preferences towards specific scenarios. Targeting strategies for specific participant groups can yield better engagement. For instance, focusing on cultivating a positive Experience and Reputation may be more effective for older participants.

7.2 Summary

Below is a bullet point summary of what was discussed before in Section 7.1.

Theoretical findings

- The interaction effect
- Relatively strong associations
- Age-related relationship

Practical implications

- A more efficient incentive framework can be designed
- Integrated Approach
- Targeting strategies to different demographics

7.3 Limitations and future directions

There were some limitations to this research. The chi-squared test assumes an expected cell count in the contingency table of at least 5. Unfortunately, this was not the case with the obtained data. Fisher's exact test would be more suitable when the sample size is smaller and does not require an expected cell count in the contingency table of at least 5. After the survey was distributed, 30 respondents were collected. A larger sample size would provide better data to investigate since the survey was cross-country. It would not represent the population of a specific country. The correlations and associations provide some insights. However, as previously mentioned in Section 6.1.4, causation does not imply a causal relationship. Further research could delve into the reasons behind observed patterns. It could also be suggested to improve the survey. The survey missed questions that measured the interactions between three factors. These missing questions are listed below. Some questions were responded to one-sidedly or were duplicates. This made it harder to gain meaningful insights and draw conclusions. It is important to make the questions unbiased so that the respondents would take more time to consider the options. Using a larger scale of measurements in factors would also be a good idea. This would provide a more comprehensive understanding of decision-making dynamics. More factors could be implemented into the model but need its complexity to be considered.

In conclusion, the research uncovered the complex dynamics between Financial Incentives, Experience, and Reputation in crowdshipping participation. The statistically significant associations and age-related trends provided valuable insights for both theory and practice. By aligning incentive structures with these findings, the crowdshipping industry can enhance its effectiveness and better cater to participant preferences.

Question 15:
Scenario 1: Factor 1 is bad, Factor 2 is good, and Factor 3 is bad.
Scenario 2: Factor 1 is good, Factor 2 is bad, and Factor 3 is good.

Question 16:
Scenario 1: Factor 1 is bad, Factor 2 is bad, and Factor 3 is good.
Scenario 2: Factor 1 is good, Factor 2 is good, and Factor 3 is bad.

Question 17:
Scenario 1: Factor 1 is good, Factor 2 is bad, and Factor 3 is bad.
Scenario 2: Factor 1 is bad, Factor 2 is good, and Factor 3 is good.

References

[MCK, 2022] (2022). The next horizon for grocery e-commerce: beyond the pandemic bump.

[Al Sukaini et al., 2015] Al Sukaini, A. K. M., Zhang, J., and Albazooni, A. G. Z. (2015). Mobile crowdsourcing: Intrinsic and extrinsic motivational factors influencing online communities in china. *Journal of Marketing Development and Competitiveness*, 9(1):129.

[Asdecker and Zirkelbach, 2020] Asdecker, B. and Zirkelbach, F. (2020). What drives the drivers? a qualitative perspective on what motivates the crowd delivery workforce. In *HICSS*, pages 1–10.

[Asiamah et al., 2017] Asiamah, N., Mensah, H. K., and Oteng-Abayie, E. F. (2017). Do larger samples really lead to more precise estimates? a simulation study. *American Journal of Educational Research*, 5:9–17.

[Bevans, 2023] Bevans, R. (2023). Choosing the right statistical test — types examples. *Scribbr*.

[Cameron and Pierce, 1994] Cameron, J. and Pierce, W. D. (1994). Reinforcement, reward, and intrinsic motivation: A meta-analysis. *Review of Educational research*, 64(3):363–423.

[Coccia, 2019] Coccia, M. (2019). Intrinsic and extrinsic incentives to support motivation and performance of public organizations. *Journal of Economics Bibliography*, 6(1):20–29.

[Creswell, 2012] Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Pearson Education, Inc.

[Dietmann et al., 2020] Dietmann, K. et al. (2020). Crowdshipping the last-mile delivery-an empirical investigation into the crowd's willingness to participate as crowdshipping drivers.

[Dolan, 2019] Dolan, S. (2019). Crowdsourced delivery explained: making same day shipping cheaper through local couriers. *Business Insider*.

[Eisenberger et al., 1999] Eisenberger, R., Pierce, W. D., and Cameron, J. (1999). Effects of reward on intrinsic motivation—negative, neutral, and positive: Comment on deci, koestner, and ryan (1999).

[Feyisetan and Simperl, 2019] Feyisetan, O. and Simperl, E. (2019). Beyond monetary incentives: Experiments in paid microtask contests. *ACM Transactions on Social Computing*, 2(2):1–31.

[Gdowska et al., 2018] Gdowska, K., Viana, A., and Pedroso, J. P. (2018). Stochastic last-mile delivery with crowdshipping. *Transportation research procedia*, 30:90–100.

[Hamari et al., 2016] Hamari, J., Sjöklint, M., and Ukkonen, A. (2016). The sharing economy: Why people participate in collaborative consumption. *Journal of the Association for Information Science and Technology*, 67:2047–2059.

[Hong et al., 2019] Hong, H., Li, X., He, D., Zhang, Y., and Wang, M. (2019). Crowdsourcing incentives for multi-hop urban parcel delivery network. *IEEE Access*, 7:26268–26277.

[Intelligence, 2019] Intelligence, I. (2019). The crowdsourced delivery market continues to improve last mile supply chain economics in the e-commerce era. *Business Insider*.

[Islam, 2018] Islam, M. R. (2018). Sample size and its role in central limit theorem (clt). *Computational and Applied Mathematics Journal*, 4(1):1–7.

[Kaufmann et al., 2011] Kaufmann, N., Schulze, T., and Veit, D. (2011). More than fun and money. worker motivation in crowdsourcing–a study on mechanical turk.

[Klopfenstein and Delpriori, 2019] Klopfenstein, L. C. and Delpriori, S. (2019). Incentives for crowdsourcing.

[Kreps, 1997] Kreps, D. M. (1997). Intrinsic motivation and extrinsic incentives. *The American economic review*, 87(2):359–364.

[Le and Ukkusuri, 2018] Le, T. V. and Ukkusuri, S. V. (2018). Crowd-shipping services for last mile delivery: analysis from survey data in two countries. *arXiv preprint arXiv:1810.02856*.

[Le and Ukkusuri, 2019] Le, T. V. and Ukkusuri, S. V. (2019). Crowd-shipping services for last mile delivery: Analysis from american survey data. *Transportation Research Interdisciplinary Perspectives*, 1:100008.

[Lee, 2016] Lee, D. (2016). Alternatives to p value: Confidence interval and effect size. *Korean Journal of Anesthesiology*, 69:555.

[Libretexts, 2020] Libretexts (2020). 7.1: Incentive theory of motivation and intrinsic vs. extrinsic motivation. *Social Sci LibreTexts*.

[Merkus, 2022] Merkus, J. (2022). Cramer8217;s v Begrijpen, berekenen en interpreteren (met voorbeeld). *Scribbr*.

[Miller et al., 2017] Miller, J., Nie, Y., and Stathopoulos, A. (2017). Crowdsourced urban package delivery: Modeling traveler willingness to work as crowdshippers. *Transportation Research Record*, 2610(1):67–75.

[MinisterievanAlgemeneZaken, 2022] MinisterievanAlgemeneZaken (2022). Klimaatplan 2021-2030.

[Pourrahmani and Jaller, 2021] Pourrahmani, E. and Jaller, M. (2021). Crowdshipping in last mile deliveries: Operational challenges and research opportunities. *Socio-Economic Planning Sciences*, 78:101063.

[Rai et al., 2021] Rai, H. B., Broekaert, C., Verlinde, S., and Macharis, C. (2021). Sharing is caring: How non-financial incentives drive sustainable e-commerce delivery. *Transportation Research Part D: Transport and Environment*, 93:102794.

[Satrio Wicaksono, 2018] Satrio Wicaksono, S. (2018). Exploring the market potential of bicycle crowdshipping: A bi-level acceptance perspective.

[Simoni et al., 2020] Simoni, M. D., Marcucci, E., Gatta, V., and Claudel, C. G. (2020). Potential last-mile impacts of crowdshipping services: A simulation-based evaluation. *Transportation*, 47:1933–1954.

[Turney, 2023a] Turney, S. (2023a). Chi-Square Test of Independence — Formula, Guide 038; Examples. *Scribbr*.

[Turney, 2023b] Turney, S. (2023b). Pearson Correlation Coefficient (r) — Guide 038; Examples. *Scribbr*.

[Wang et al., 2017] Wang, H., Guo, S., Cao, J., and Guo, M. (2017). Melody: A long-term dynamic quality-aware incentive mechanism for crowdsourcing. *IEEE Transactions on Parallel and Distributed Systems*, 29(4):901–914.

[Wang et al., 2019] Wang, Y., Cai, Z., Zhan, Z.-H., Gong, Y.-J., and Tong, X. (2019). An optimization and auction-based incentive mechanism to maximize social welfare for mobile crowdsourcing. *IEEE Transactions on Computational Social Systems*, 6(3):414–429.

[Zalia et al., 2021] Zalia, A.-H., Zhang, Y., Aishatu, A., Akosua, A. S., and Fathia, B. V. (2021). Crowd logistics' impact on environmental sustainability in a developing economy: An analysis in ghana.

A Survey

Crowdsourced delivery survey

Dear participant,

Thank you for taking the time to participate in this survey. Your valuable input will contribute to my thesis research for the Bachelor of Science in Computer Science & Economics at Leiden University. The purpose of this research is to measure individuals' willingness to occasionally deliver packages near their neighborhoods or workplace.

Imagine that you can make a delivery on your way home, to work, or whenever you are available. You may use your own vehicle, public transportation, or a shared vehicle for making the delivery.

In your role as an occasional delivery driver, you begin by logging into the delivery platform's app, where you receive detailed information for delivery tasks. You pick up these packages and deliver them to customers' locations, relying on the app's guidance.
Now, for the survey, I would like to ask you to review some scenario options and select the one that you find more appealing. Below you see an example:

Scenario 1	Scenario 2
Receive 10 euros in the form of compensation per delivery	Receive 5 euros in the form of compensation per delivery

Please indicate which scenario you would be more likely to make a delivery.

Thank you for your participation!

Jun Fei Cheung

→

Q1 . From the scenario options presented below, select the one you are more likely to make a delivery

Scenario 1	Scenario 2
Clear skies and mild weather (sunny or partly cloudy) with light or moderate traffic flow	Rainy or stormy weather with heavy traffic congestion.

Which scenario would you be more likely to make a delivery?

Scenario 1

Scenario 2

Q2. From the scenario options presented below, select the one you are more likely to make a delivery

Scenario 1	Scenario 2
Rainy or stormy weather with heavy traffic congestion.	Clear skies and mild weather (sunny or partly cloudy) with light or moderate traffic flow
No feedback	Positive feedback*

* Positive feedback: Successfully delivery of this package earns you a great reputation among your neighbors/colleagues and recognition for your contribution to reducing CO2 emissions

Which scenario would you be more likely to make a delivery?

Scenario 1

Scenario 2

Q3. From the scenario options presented below, select the one you are more likely to make a delivery

Scenario 1	Scenario 2
Receive 5 euros for each delivery	Receive 7,50 euros for each delivery
Clear skies and mild weather (sunny or partly cloudy) with light or moderate traffic flow	Rainy or stormy weather with heavy traffic congestion.

Which scenario would you be more likely to make a delivery?

Scenario 1

Scenario 2

Q4. From the scenario options presented below, select the one you are more likely to make a delivery

Scenario 1	Scenario 2
Clear skies and mild weather (sunny or partly cloudy) with light or moderate traffic flow	Rainy or stormy weather with heavy traffic congestion.
Positive feedback*	No feedbback

* Positive feedback: Successfully delivery of this package earns you a great reputation among your neighbors/colleagues and recognition for your contribution to reducing CO2 emissions

Which scenario would you be more likely to make a delivery?

Scenario 1

Scenario 2

Q5. From the scenario options presented below, select the one you are more likely to make a delivery

Scenario 1	Scenario 2
Receive 7,50 euros for each delivery	Receive 5 euros for each delivery

Which scenario would you be more likely to make a delivery?

Scenario 1

Scenario 2

Q6. From the scenario options presented below, select the one you are more likely to make a delivery

Scenario 1	Scenario 2
Positive feedback*	No feedback

* Positive feedback: Successfully delivery of this package earns you a great reputation among your neighbors/colleagues and recognition for your contribution to reducing CO2 emissions

Which scenario would you be more likely to make a delivery?

Scenario 1

Scenario 2

Q7. From the scenario options presented below, select the one you are more likely to make a delivery

Scenario 1	Scenario 2
Receive 7,50 euros for each delivery.	Receive 5 euros for each delivery
Clear skies and mild weather (sunny or partly cloudy) with light or moderate traffic flow	Rainy or stormy weather with heavy traffic congestion.
Positive feedback*	No feedback

* Positive feedback: Successfully delivery of this package earns you a great reputation among your neighbors/colleagues and recognition for your contribution to reducing CO2 emissions

Which scenario would you be more likely to make a delivery?

Scenario 1

Scenario 2

Q8. From the scenario options presented below, select the one you are more likely to make a delivery

Scenario 1	Scenario 2
Rainy or stormy weather with heavy traffic congestion.	Clear skies and mild weather (sunny or partly cloudy) with light or moderate traffic flow
Positive feedback*	No feedback

* Positive feedback: Successfully delivery of this package earns you a great reputation among your neighbors/colleagues and recognition for your contribution to reducing CO2 emissions.

Which scenario would you be more likely to make a delivery?

Scenario 1

Scenario 2

Q9. From the scenario options presented below, select the one you are more likely to make a delivery

Scenario 1	Scenario 2
No feedback	Positive feedback*

* Positive feedback: Successfully delivery of this package earns you a great reputation among your neighbors/colleagues and recognition for your contribution to reducing CO2 emissions.

Which scenario would you be more likely to make a delivery?

Scenario 1

Scenario 2

Q10. From the scenario options presented below, select the one you are more likely to make a delivery

Scenario 1	Scenario 2
Receive 7,50 euros for each delivery	Receive 5 euros for each delivery
Clear skies and mild weather (sunny or partly cloudy) with light or moderate traffic flow	Rainy or stormy weather with heavy traffic congestion.

Which scenario would you be more likely to make a delivery?

Scenario 1

Scenario 2

Q11. From the scenario options presented below, select the one you are more likely to make a delivery

Scenario 1	Scenario 2
Receive 7,50 euros for each delivery	Receive 5 euros for each delivery
No feedback	Positive feedback*

* Positive feedback: Successfully delivery of this package earns you a great reputation among your neighbors/colleagues and recognition for your contribution to reducing CO2 emissions.

Which scenario would you be more likely to make a delivery?

Scenario 1

Scenario 2

Q12. From the scenario options presented below, select the one you are more likely to make a delivery

Scenario 1	Scenario 2
Receive 5 euros for each delivery	Receive 7,50 euros for each delivery

Which scenario would you be more likely to make a delivery?

Scenario 1

Scenario 2

Q13. From the scenario options presented below, select the one you are more likely to make a delivery

Scenario 1	Scenario 2
Clear skies and mild weather (sunny or partly cloudy) with light or moderate traffic flow	Rainy or stormy weather with heavy traffic congestion.

Which scenario would you be more likely to make a delivery?

Scenario 1

Scenario 2

Q14. From the scenario options presented below, select the one you are more likely to make a delivery

Scenario 1	Scenario 2
.Receive 7,50 euros for each delivery	Receive 5 euros for each delivery
Clear skies and mild weather (sunny or partly cloudy) with light or moderate traffic flow	Rainy or stormy weather with heavy traffic congestion.
Positive feedback*	No feedback

* Positive feedback: Successfully delivery of this package earns you a great reputation among your neighbors/colleagues and recognition for your contribution to reducing CO2 emissions.

Which scenario would you be more likely to make a delivery?

Scenario 1

Scenario 2

Q15. What is your gender?

- Male
- Female
- Other

Q16. How old are you?

- Under 18
- 18-24 years old
- 25-34 years old
- 35-44 years old
- 45-54 years old
- 55-64 years old
- 65+ years old

Q17. In which country/region do you currently reside?

Q18. Which of the following best describes your current living area?

- City
- Rural

Q19. What is your highest level of education?

- High school diploma or equivalent
- Associate's degree or vocational certification
- Bachelor's degree
- Master's degree
- Doctorate or professional degree

Q20. What best describes your employment status over the last three months?

- Working full-time
- Working part-time
- Unemployed and looking for work
- A homemaker or stay-at-home parent
- Student
- Retired

Other

Q21. What is your most frequent mode of transportation?

Car

Public transportation (bus, train, subway, etc)

Bicycle

Walking

Other

Q22. This is the end of the survey. Thanks for your support!
What are the reasons for you to make those answers? Please feel free to leave any comments or suggestions to us:

Q23. What is your Prolific ID?

Q24. **Your Prolific Code is: CDXNCT8S**
Please copy the code above to the Prolific page to receive the payment.

← →

B Types of questions asked in the survey

The number behind the question number denotes which Question it was from Appendix A
Questions for factor 1:
Question 1 (12):
Scenario 1: Factor 1 is good, Factor 2 is irrelevant, and Factor 3 is irrelevant.
Scenario 2: Factor 1 is bad, Factor 2 is irrelevant, Factor 3 is irrelevant.

Question 2 (5):
Scenario 1: Factor 1 is bad, Factor 2 is irrelevant, Factor 3 is irrelevant.
Scenario 2: Factor 1 is good, Factor 2 is irrelevant, and Factor 3 is irrelevant.

Question for Factor 2:
Question 3 (13):
Scenario 1: Factor 1 is irrelevant, Factor 2 is good, and Factor 3 is irrelevant.
Scenario 2: Factor 1 is irrelevant, Factor 2 is bad, and Factor 3 is irrelevant.

Question 4 (1):
Scenario 1: Factor 1 is irrelevant, Factor 2 is bad, Factor 3 is irrelevant.
Scenario 2: Factor 1 is irrelevant, Factor 2 is good, Factor 3 is irrelevant.

Question for Factor 3:
Question 5 (6):
Scenario 1: Factor 1 is irrelevant, Factor 2 is irrelevant, and Factor 3 is good.
Scenario 2: Factor 1 is irrelevant, Factor 2 is irrelevant, Factor 3 is bad.

Question 6 (9):
Scenario 1: Factor 1 is irrelevant, Factor 2 is irrelevant, Factor 3 is bad.
Scenario 2: Factor 1 is irrelevant, Factor 2 is irrelevant, and Factor 3 is good.

Questions for Factors 1 and 2:
Question 7 (10):
Scenario 1: Factor 1 is good, Factor 2 is good, and Factor 3 is irrelevant.
Scenario 2: Factor 1 is bad, Factor 2 is bad, and Factor 3 is irrelevant.

Question 8 (3):
Scenario 1: Factor 1 is good, Factor 2 is bad, and Factor 3 is irrelevant.
Scenario 2: Factor 1 is bad, Factor 2 is good, and Factor 3 is irrelevant.

Questions for Factors 2 and 3:
Question 9 (4):
Scenario 1: Factor 1 is irrelevant, Factor 2 is good, and Factor 3 is good.
Scenario 2: Factor 1 is irrelevant, Factor 2 is bad, and Factor 3 is bad.

Question 10 (8):

Scenario 1: Factor 1 is irrelevant, Factor 2 is good, and Factor 3 is bad.
Scenario 2: Factor 1 is irrelevant, Factor 2 is bad, and Factor 3 is good.

Questions for Factors 1 and 3:
Question 11 (2):
Scenario 1: Factor 1 is good, Factor 2 is irrelevant, and Factor 3 is good.
Scenario 2: Factor 1 is bad, Factor 2 is irrelevant, and Factor 3 is bad.

Question 12 (11):
Scenario 1: Factor 1 is good, Factor 2 is irrelevant, and Factor 3 is bad.
Scenario 2: Factor 1 is bad, Factor 2 is irrelevant, and Factor 3 is good.

Question for Factors 1, 2, and 3:
Question 13 (14):
Scenario 1: Factor 1 is good, Factor 2 is good, and Factor 3 is good.
Scenario 2: Factor 1 is bad, Factor 2 is bad, and Factor 3 is bad.

Question for Factors 1, 2, and 3:
Question 14 (7):
Scenario 1: Factor 1 is bad, Factor 2 is bad, and Factor 3 is bad.
Scenario 2: Factor 1 is good, Factor 2 is good, and Factor 3 is good.

C Additional tables and graphs from Results and Data analysis

	GoodFI+BadRep	BadFI+GoodRep
Female	11	3
Male	10	6

Table 6: Contingency table between GoodFI+BadRep vs BadFI+GoodRep with Gender

	GoodFI+BadRep	badFI+GoodRep
18-24 years old	7	6
25-34 years old	7	0
35-44 years old	6	2
55-64 years old	1	1

Table 7: Contingency table between GoodFI+BadRep vs BadFI+GoodRep with Age

	GoodFI+BadRep	BadFI+GoodRep
Denmark	0	1
Germany	1	0
Greece	1	0
Latvia	2	0
Netherlands	1	0
Poland	3	5
Portugal	9	2
Spain	1	0
Switzerland	0	1
United Kingdom of Great Britain and Northern Ir...	3	0

Table 8: Contingency table between
GoodFI+BadRep vs BadFI+GoodRep
with Country

	GoodFI+BadRep	badFI+GoodRep
City	17	8
Rural	4	1

Table 9: Contingency table between
GoodFI+BadRep vs BadFI+GoodRep
with Living Area

	GoodFI+BadRep	badFI+GoodRep
Bicycle	3	1
Car	9	3
Public transportation	4	3
Walking	5	2

Table 10: Contingency table between GoodFI+BadRep vs BadFI+GoodRep
with Transportation mode

	GoodFI+BadRep	BadFI+GoodRep
Associate's degree or vocational certification	3	1
Bachelor's degree	12	3
High school diploma or equivalent	4	3
Master's degree	2	2

Table 11: Contingency table between GoodFI+BadRep vs BadFI+GoodRep
with Educational background

	GoodFI+BadRep	BadFI+GoodRep
A homemaker or stay-at-home parent	0	1
Other	1	1
Student	6	2
Unemployed and looking for work	2	0
Working full-time	10	2
Working part-time	2	3

Table 12: Contingency table between GoodFI+BadRep vs BadFI+GoodRep with Employment status

Gender	Number of Respondents	Percentage
Male	16	53.33%
Female	14	46.67%
Total	30	100

Table 13: Gender frequency table

Age	Number of Respondents	Percentage
18-24 years old	13	43.44%
25-34 years old	7	23.33%
35-44 years old	8	26.67%
55-64 years old	2	6.67%

Table 14: Age frequency table

Country	Number of Respondents	Percentage
Portugal	11	33.657%
Poland	8	26.67%
UK	3	10.00%
Latvia	2	6.67%
Greece	1	3.33%
Spain	1	3.33%
Switzerland	1	3.33%
Netherlands	1	3.33%
Denmark	1	3.33%
Germany	1	3.33%

Table 15: Country frequency table

Living Area	Number of Respondents	Percentage
City	25	83.33%
Rural	5	16.67%

Table 16: Living area frequency table

Educational background	Number of Respondents	Percentage
Bachelor's degree	15	50.00%
High school diploma or equivalent	7	23.33%
Associate's degree or vocational certification	4	13.33%
Master's degree	4	13.33%

Table 17: Educational background frequency table

Employment status	Number of Respondents	Percentage
Working full-time	12	50.00%
Student	8	26.67%
Working part-time	5	16.67%
Other	2	6.67%
Unemployed and looking for work	2	6.67%
A homemaker or stay-at-home parent	1	3.33%

Table 18: Employment status frequency table

Transportation	Number of Respondents	Percentage
Car	12	40.00%
Public transportation (bus, train, subway, etc)	7	23.33%
Walking	7	23.33%
Bicycle	4	13.33%

Table 19: Transportation mode frequency table

Observed frequencies (O)			
Intervention	Scenario 1	Scenario 2	Row totals
GoodExp + BadRep vs BadExp + GoodRep	22	8	30
GoodExp vs BadExp	59	1	60
Column totals	81	9	90
Expected frequencies (E)			
Intervention	Scenario 1	Scenario 2	Row totals
GoodExp + BadRep vs BadExp + GoodRep	27	3	30
GoodExp vs BadExp	54	6	60
Column totals	81	9	90
(O − E)ˆ2/ E			
	Scenario 1	Scenario 2	
GoodExp + BadRep vs BadExp + GoodRep	0,9259259259	8,333333333	
GoodExp vs BadExp	0,462962963	4,166666667	
Xˆ2	13,88888889		
df	1		
p-value	0,0001939416291		

Table 20: Calculated test statistic of Experience vs (GoodExp + BadRep vs BadExp + GoodRep)

Observed frequencies (O)			
Intervention	Scenario 1	Scenario 2	Row totals
GoodExp + BadRep vs BadExp + GoodRep	22	8	30
GoodRep vs BadRep	60	0	60
Column totals	82	8	90
Expected frequencies (E)			
Intervention	Scenario 1	Scenario 2	Row totals
GoodExp + BadRep vs BadExp + GoodRep	27,33333333	2,666666667	30
GoodRep vs BadRep	54,66666667	5,333333333	60
Column totals	82	8	90
(O − E)ˆ2/ E			
	Scenario 1	Scenario 2	
GoodExp + BadRep vs BadExp + GoodRep	1,040650407	10,66666667	
GoodRep vs BadRep	0,5203252033	5,333333333	
Xˆ2	17,56097561		
df	1		
p-value	0,00002782401113		

Table 21: Calculated test statistic of Reputation vs (GoodExp + BadRep vs BadExp + GoodRep)

Observed frequencies (O)			
Intervention	Scenario 1	Scenario 2	Row totals
GoodExp + BadRep vs BadExp + GoodRep	22	8	30
GoodExp + GoodRep vs BadExp + Bad Rep	59	1	60
Column totals	81	9	90
Expected frequencies (E)			
Intervention	Scenario 1	Scenario 2	Row totals
GoodExp + BadRep vs BadExp + GoodRep	27	3	30
GoodExp + GoodRep vs BadExp + BadRep	54	6	60
Column totals	81	9	90
(O − E)ˆ2/ E			
	Scenario 1	Scenario 2	
GoodExp + BadRep vs BadExp + GoodRep	0,9259259259	8,333333333	
GoodExp + GoodRep vs BadExp + BadRep	0,462962963	4,166666667	
Xˆ2	13,88888889		
df	1		
p-value	0,0001939416291		

Table 22: Calculated test statistic of (GoodExp + GoodRep vs BadExp + BadRep) vs (GoodExp + BadRep vs BadExp + GoodRep)

Observed frequencies (O)			
Intervention	Scenario 1	Scenario 2	Row totals
BadFI + GoodExp vs GoodFI + BadExp	27	3	30
GoodExp vs BadExp	59	1	60
Column totals	86	4	90
Expected frequencies (E)			
Intervention	Scenario 1	Scenario 2	Row totals
BadFI + GoodExp vs GoodFI + BadExp	28,66666667	1,333333333	30
GoodExp vs BadExp	57,33333333	2,666666667	60
Column totals	86	4	90
(O − E)ˆ2 / E			
	Scenario 1	Scenario 2	
BadFI + GoodExp vs GoodFI + BadExp	0,09689922481	2,083333333	
GoodExp vs BadExp	0,0484496124	1,041666667	
Xˆ2	3,270348837		
df	1		
p-value	0,07054262201		

Table 23: Calculated test statistic of Experience vs (BadFI + GoodExp vs GoodFI + BadExp)

Observed frequencies (O)			
Intervention	Scenario 1	Scenario 2	Row totals
BadFI + GoodExp vs GoodFI + BadExp	27	3	30
GoodFI vs BadFI	59	1	60
Column totals	86	4	90
Expected frequencies (E)			
Intervention	Scenario 1	Scenario 2	Row totals
BadFI + GoodExp vs GoodFI + BadExp	28,66666667	1,333333333	30
GoodFI vs BadFI	57,33333333	2,666666667	60
Column totals	86	4	90
(O − E)ˆ2 / E			
	Scenario 1	Scenario 2	
BadFI + GoodExp vs GoodFI + BadExp	0,09689922481	2,083333333	
GoodFI vs BadFI	0,0484496124	1,041666667	
Xˆ2	3,270348837		
df	1		
p-value	0,07054262201		

Table 24: Calculated test statistic of Financial incentives vs (BadFI + GoodExp vs GoodFI + BadExp)

Observed frequencies (O)			
Intervention	Scenario 1	Scenario 2	Row totals
BadFI + GoodExp vs GoodFI + BadExp	27	3	30
GoodFI + GoodExp vs BadFI + BadFI	30	0	30
Column totals	57	3	60
Expected frequencies (E)			
Intervention	Scenario 1	Scenario 2	Row totals
BadFI + GoodExp vs GoodFI + BadExp	28,5	1,5	30
GoodFI + GoodExp vs BadFI + BadFI	28,5	1,5	30
Column totals	57	3	60
(O − E)ˆ2 / E			
	Scenario 1	Scenario 2	
BadFI + GoodExp vs GoodFI + BadExp	0,07894736842	1,5	
GoodFI + GoodExp vs BadFI + BadFI	0,07894736842	1,5	
Xˆ2	3,157894737		
df	1		
p-value	0,07556056753		

Table 25: Calculated test statistic of (GoodFI + GoodExp vs BadFI + BadFI) vs (BadFI + GoodExp vs GoodFI + BadExp)

Observed frequencies (O)			
Intervention	Scenario 1	Scenario 2	Row totals
GoodFI + BadRep vs BadFI + GoodRep	21	9	30
GoodRep vs BadRep	60	0	60
Column totals	81	9	90
Expected frequencies (E)			
Intervention	Scenario 1	Scenario 2	Row totals
GoodFI + BadRep vs BadFI + GoodRep	27	3	30
GoodRep vs BadRep	54	6	60
Column totals	81	9	90
(O − E)ˆ2 / E			
	Scenario 1	Scenario 2	
GoodFI + BadRep vs BadFI + GoodRep	1,333333333	12	
GoodRep vs BadRep	0,6666666667	6	
Xˆ2	20		
df	1		
p-value	0,000007744216431		

Table 26: Calculated test statistic of Reputation vs (GoodFI + BadRep vs BadFI + GoodRep)

Observed frequencies (O)			
Intervention	Scenario 1	Scenario 2	Row totals
GoodFI + BadRep vs BadFI + GoodRep	21	9	30
GoodFI vs BadFI	59	1	60
Column totals	80	10	90
Expected frequencies (E)			
Intervention	Scenario 1	Scenario 2	Row totals
GoodFI + BadRep vs BadFI + GoodRep	26,66666667	3,333333333	30
GoodFI vs BadFI	53,33333333	6,666666667	60
Column totals	80	10	90
(O − E)ˆ2 / E			
	Scenario 1	Scenario 2	
GoodFI + BadRep vs BadFI + GoodRep	1,204166667	9,633333333	
GoodFI vs BadFI	0,6020833333	4,816666667	
Xˆ2	16,25625		
df	1		
p-value	0,00005532678421		

Table 27: Calculated test statistic of Financial incentives vs (GoodFI + BadRep vs BadFI + GoodRep)

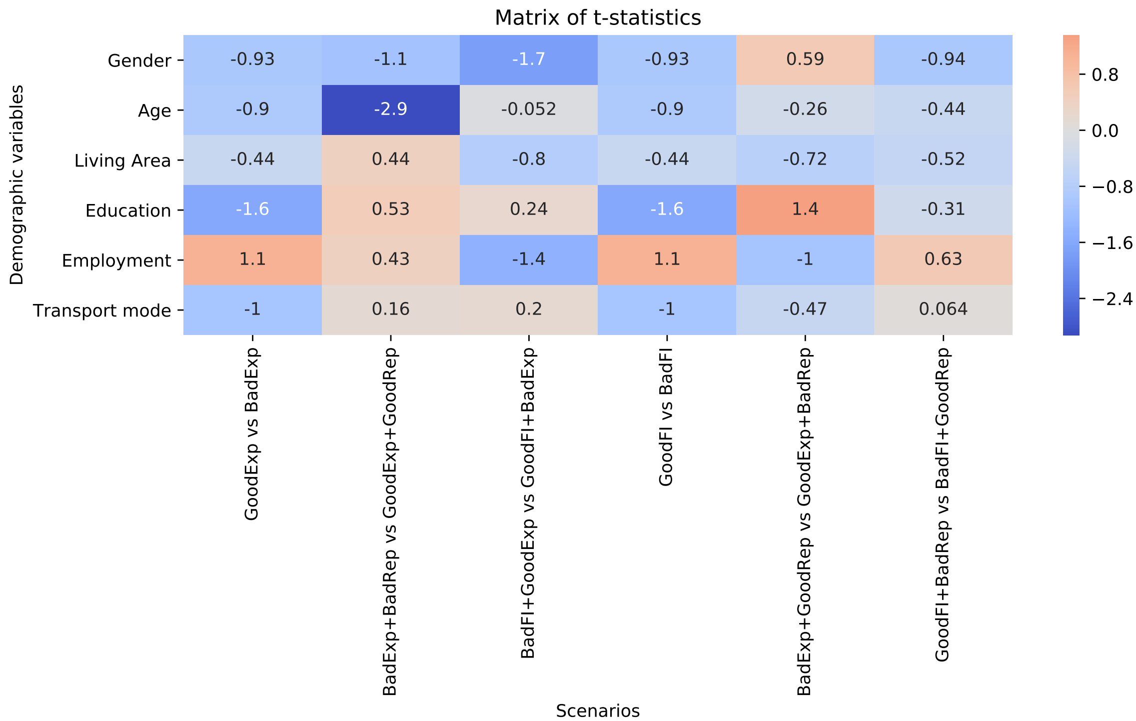

Figure 7: Calculated t-statistics for every correlation coefficient