CS3121 Introduction to Data Science
Class Project

Exploration of factors influencing consumers' online purchase intentions during crises in Sri Lanka

Deadline - Final Project Report and video: 14 May 2025 at 11.59 P.M

Project Overview

Imagine that you have been hired as a Data Science Engineer at Wolt, a well-known company that operates supermarkets and shopping malls in Sri Lanka. Wolt is now seeking in-depth data analysis and a comprehensive survey to examine the key factors influencing consumers' intentions to buy online during the crisis in Sri Lanka. Your task is to explore how consumer behavior changes in times of crisis and identify what drives or prevents online shopping. These insights will be vital for Wolt in shaping its plans and strengthening its marketing strategies accordingly. Suppose for this purpose that you conducted a survey and collected responses from the Sri Lankan public regarding their perceptions and experiences with making online purchases during crisis times.

The conceptual diagram you used to design the survey is shown in Figure 1. The diagram illustrates the independent variables and their subfactors, moderating variables, mediating variables, and the dependent variable, which is the consumer's intention to purchase online during crisis times.

The survey consisted of seven main sections, as described below.

- **Section 1**: Captures demographic details such as age, education, marital status, and professional background.
- **Section 2**: Evaluates participants' familiarity with commonly used online purchasing platforms in Sri Lanka.
- **Section 3**: Assesses Perceived Ease of Use (PEOU) with ten questions exploring sub-factors.
- **Section 4**: Analyzes Perceived Usefulness (PU) using ten questions to examine different aspects.
- **Section 5**: Evaluates Structural Assurance (SA) through six questions covering three sub-factors.
- **Section 6**: Focuses on Social Influence, incorporating 12 questions to evaluate six dimensions.
- **Section 7**: Focuses on intention of purchase online, incorporating four questions.

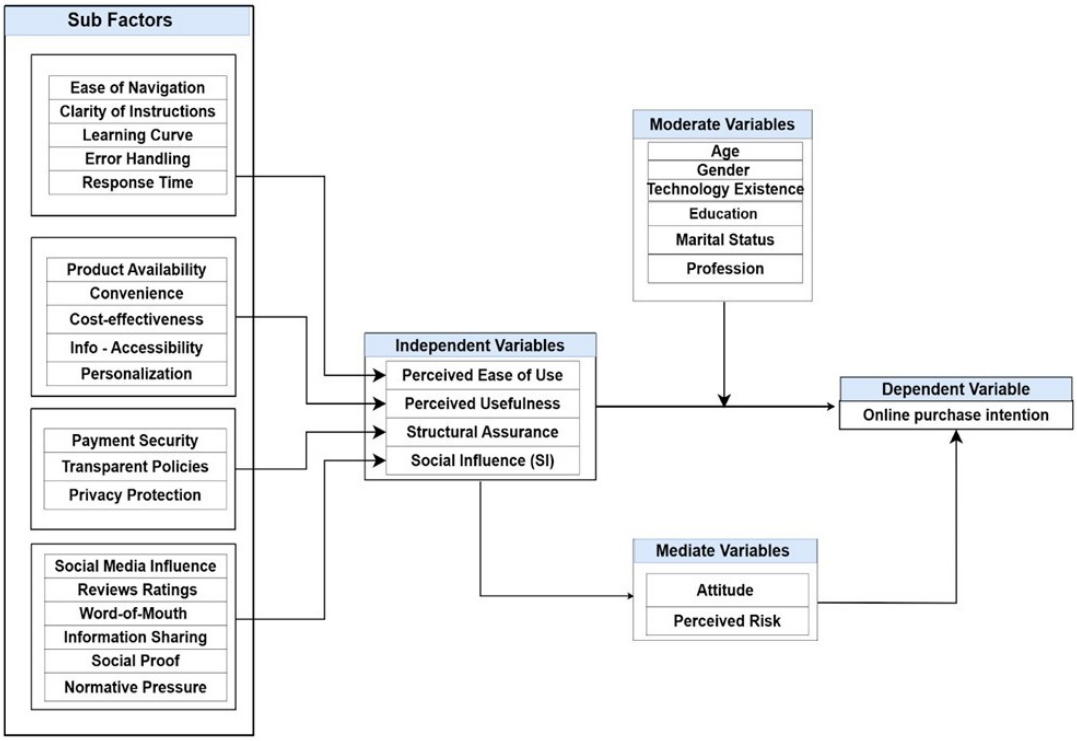

Figure 1: Conceptual diagram of the conducted survey

The mapping between each variable in the conceptual diagram and the questions in the survey is given in Table 1.

Suppose you collected 836 responses, which are given as a CSV file with the project description.

Project Tasks

Your task is to analyze the provided dataset and identify valuable insights into the factors influencing consumers' intentions to purchase online during crises in Sri Lanka. To accomplish this, please follow the tasks listed below.

1. Data Pre-processing

   (a) Load the provided CSV files into a DataFrame. Use Python 3.6 for this project, and you can utilize the following libraries for data cleaning.

   - Pandas
   - Numpy
   - Scikit-learn

   (b) Verify the reliability and validity of the survey questions.

   i. Cronbatch's alpha coefficient value

Variable	Related Questions in the Questionnaire
Ease of Navigation	Section 3 [13,14]
Clarity of Instructions	Section 3 [15,16]
Learning Curve	Section 3 [17,18]
Error Handling	Section 3 [19,20]
Response Time	Section 3 [21,22]
Product Availability	Section 4 [23,24]
Convenience	Section 4 [25,26]
Cost-effectiveness	Section 4 [27,28]
Information Accessibility	Section 4 [29,30]
Personalization	Section 4 [31,32]
Payment Security	Section 5 [33,34]
Transparent Policies	Section 5 [35,36]
Privacy Protection	Section 5 [37,38]
Social Media Posts/Influence	Section 6 [39,40]
Online Reviews and Ratings	Section 6 [41,42]
Word of Mouth and Recommendations	Section 6 [43,44]
Information Sharing	Section 6 [45,46]
Social Proof	Section 6 [47,48]
Normative Pressure	Section 6 [49,50]
Intention to Purchase online	Section 7 [51,52]

Table 1: Sub Factors and Related Questions

ii. Inter-item correlation

(c) Identify and resolve data quality issues, such as missing values, duplicates, etc.

(d) Apply necessary data transformations.

2. Data Analysis Conduct descriptive and exploratory data analysis to derive insights.

3. Hypothesis Testing

(a) Validate the following hypotheses using suitable statistical tests.

i. Perceived ease of use has no impact on intention to purchase.

ii. Perceived usefulness has no impact on the intention to purchase online.

iii. Having structural assurance has no impact on the intention to make an online purchase.

iv. Social influence has no impact on the intention to purchase online.

v. The attitude towards online shopping does not mediate the relationship between perceived ease of use and the intention to purchase online.

vi. Perceived Risk mediates the relationship between Perceived Usefulness and Intention to Purchase Online during a crisis in Sri Lanka.

vii. Ease of navigation has no impact on the intention to purchase online.

viii. Clarity of instructions has no impact on the intention to purchase online.

ix. Response Time has no impact on intention to purchase online.

x. Error handling has no impact on intention to purchase online

(b) In addition to the above hypotheses, create your own 10 hypotheses based on the given conceptual diagram and validate them using suitable statistical tests.

4. Rule Mining - Using the Apriori algorithm as your rule-mining technique, discover five interesting rules about consumer perception of online purchasing during a crisis.

This is a group project. Each group will include five members.

Deliverables

As the deliverables, you must submit a report and a video. Each should be renamed as the project group name and also include the index numbers of all the team members.

1. Final Report - Submit a report (PDF format) including:

   - Problem Overview
   - Data Pre-processing Steps
   - Insights from Data Analysis
   - Hypothesis Testing
   - Rule mining
   - Discussion
   - Recommendations
   - Conclusion

   The report should be limited to a maximum of 10 pages. Reports longer than that will not be considered for grading.

2. Video Presentation - Create a video explaining your main findings and recommendations. All the group members should be present in the video. **The maximum duration of the video is two minutes. Videos longer than that will not be considered for grading.**