Factors determining in-person vs online purchases: a logistic regression model*

Kimlin Chin

06 April 2022

Abstract

E-commerce is booming, with more than 80% of Canadians shopping online, and they have spent over 50 billion dollars in 2018. So should certain businesses ditch their in-person services in favour of online? Using data from the 2018 Canadian Internet Use Survey, we developed a logistic regression model to investigate what factors may influence whether a person makes purchases in-person vs online. We find that young tech-savvy adults and men are more likely to make online purchases than in-person. Our findings have implications for business marketing strategies and the decision to expand to online platforms.

Keywords: e-commerce, internet, logistic regression, canadian internet use survey, canada

1 Introduction

With the Internet being ubiquitous, many businesses have entered the realm of e-commerce. From furniture to food, you can buy almost anything you could dream of online these days, with companies like Amazon making billion-dollar profits from their e-commerce services.

So when do we, as consumers, choose to shop in person over online? In this paper we take a deep dive into the data from the 2018 Canadian Internet Use Survey to investigate this question. Specifically we plot graphs of responses to the survey questions which are relevant to our research question, and construct a logistic regression model with a binary dependent variable for whether the individual purchased physical goods online or not, and independent variables for age, gender and technical skill. We find that wanting to see, hold or try on the product before purchasing is the most common reason for individuals to purchase some goods in-person instead of online. Additionally, our model predicts that young tech-savvy adults and men are more likely to make online purchases than in-person. Future work can build upon the proposed model to include other factors, such as proximity to store containing the desired item, by collecting more data, as the unavailability of detailed factors is a limitation of this study.

The rest of the paper has the following structure: Section 2 describes the data from the 2018 Canadian Internet Use Survey, Section 3 discusses the logistic regression model, Section 4 presents the results of our analysis and Section 5 discusses the findings and limitations.

2 Data

The data is from the 2018 Canadian Internet Use Survey. There are 1,426 features and 13,810 observations in the dataset.

R (R Core Team 2020) and the R package `tidyverse` which was written by Wickham et al. (2019) were used for data processing and to make the plots and tables.

---

*Code and data are available at: https://github.com/KCtt457/cius-2018-analysis.

3 Model

4 Results

5 Discussion

5.1 First discussion point

5.2 Second discussion point

5.3 Third discussion point

5.4 Weaknesses and next steps

Appendix

.1 Enhancement

A   Additional details

References

R Core Team. 2020. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.