

# E-commerce Customer Data

## Business Statistics Assignment

António Fidalgo

Miguel Salema

### Assignment Overview

This is a group assignment where you will analyze a dataset from an online e-commerce company. The company aims to understand which customer characteristics and behaviors lead to higher revenue. Your task is to perform an exploratory analysis using plots and summary statistics to provide insights.

### Instructions

- This is group work. Each group should submit two files: an R script (.R) or Quarto file (.qmd) containing your code and a PDF report (.pdf) with your findings.
- The pdf must be have a maximum of 5 pages, tables, figures and bibliography included.
- All relevant R code must be in an .R or .qmd file and must run on the grader's computer
- Upload the full directory of your assignment to Moodle by the indicated deadline as a .zip file. The .zip file should include the R file, the pdf and the data.
- The detection of any form of plagiarism in your work will result in a “fail” grade.
- You must indicate explicitly the questions you are answering.
- Deadline: 10th of December 2025 at 23:59.

## Criteria for Evaluation

Each question is worth 50% of the grade, and will both be evaluated based on the following criteria:

- **Technical Accuracy (40%)**: Correct use of plots, statistics, and code in tidyverse.
- **Interpretation (25%)**: Clear, logical, and data-driven explanations of your findings.
- **Originality & Relevance (25%)**: Insights that go beyond the obvious or provide additional value to the company.
- **Presentation (10%)**: Clean, well-formatted report and code.

## Dataset Description

Imagine yourselves as data analysts working for a company, ready to tackle a practical business case. This case involves delving into customer data to gain insights and solve business challenges.

We have a thriving e-commerce company that exclusively operates online. Our aim is to customize our marketing strategy by understanding the factors that drive varying levels of customer spending on our website. By analyzing customer behavior and characteristics, we hope to enhance our marketing approach and improve customer engagement.

To this purpose, we will analyze a data set including customer demographics and purchase history. The dataset covers a sample 1000 registered customers and their browsing activity on the company's website during the previous month (30 days). You must download the file `ecommerce.csv` from Moodle.

**Assume that you know nothing else about this firm and draw no conclusions that cannot be drawn from this data alone.**

Table 1: Description of the Variables

Variable	Description
Age	Customer age, categorized into <code>young</code> , <code>middle-aged</code> , and <code>old</code>
Gender	Customer gender, categorized into <code>male</code> or <code>female</code>
OwnHome	Home ownership status ( <code>own</code> or <code>rented</code> )
Married	Marital status ( <code>single</code> or <code>married</code> )
Location	Customer location ( <code>close</code> or <code>far</code> from brick-and-mortar stores)
Children	Number of children
History	Past purchase history ( <code>low</code> , <code>medium</code> , <code>high</code> , or <code>NA</code> )
Visits	Total website visits in the past month
AmountSpent	Total money spent in the past month (in USD)
TimeSpent	Average time spent per website visit

## Questions

1. **Customer Profiles.** Use the customer demographic variables (**Age**, **Gender**, **Location**, **Married**, **OwnHome** and **Children**) to describe two or three distinct customer profiles. Based on your analysis, how can the company use these insights to improve its marketing strategy?
2. **Maintaining Interest.** Explore the relationship between a customer's past purchase history (**History**), time spent on the website (**TimeSpent**), and total spending (**AmountSpent**). What patterns can you observe?