A logo for college computing

Description automatically generated

**Assessment Cover Page**

|  |  |
| --- | --- |
| *Student Full Name* | Isabella Gubitoso |
| *Student Number* | sba25155 |
| *Module Title* | Strategic Thinking |
| *Assessment Title* | Predicting Purchasing Intention and conversion optimization in e-commerce |
| *Lecturer/Supervisor* | Taufique Ahmed |
| *Assessment Due Date* | 7th November |
| *Date of Submission* | 6th November |

**Use of AI Tools**

**Choose the statement that fits your work and delete the other.**

I acknowledge the use of Chat GPT for the purpose of **[provide a brief explanation of how you used the tool].**

I have not used any AI tools or technologies in the preparation of this assessment.

**Declaration**

By submitting this assessment, I confirm that I have read the CCT policy on academic misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source.

I declare it to be my own work and that all material from third parties has been appropriately referenced.

I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution.

Contents

[Introduction 1](#_heading=h.gjdgxs)

[Objectives](#_heading=h.30j0zll) 2

Problem Definition3

Scope4

Methodology5

Data Sources6

Ethical Considerations7

[References 8](#_heading=h.2et92p0)

# Introduction

E-commerce has experienced massive growth over the past decades, becoming a leading retail channel worldwide. With over five billion internet users worldwide, consumers benefit from the advantages of online transactions (Statista, 2025). However, despite this evolution, most e-commerce businesses struggle with customer retention, particularly in maintaining client loyalty. According to recent data, only 2-3% of visitors to e-commerce platforms actually make purchases (Oberlo, 2024), indicating significant conversion challenges.

Predicting purchase intention is crucial for e-commerce businesses because most website visitors leave without purchasing. This creates major challenges: companies cannot identify which visitors are likely to buy, resulting in wasted marketing spend on uninterested browsers and missed opportunities to convert high-intent visitors into long-term customers. Without predictive capabilities, businesses apply generic strategies to all visitors rather than targeted approaches based on purchase likelihood.

This capstone project addresses these challenges by developing machine learning models to predict online shopping purchase intention. Using the Online Shoppers Purchasing Intention Dataset from the UCI Machine Learning Repository, this research will analyze behavioral patterns, temporal factors, and user characteristics to identify key predictors of purchase behavior.

The significance of this project lies in its practical business impact. By accurately predicting purchase intention, e-commerce businesses can implement targeted retention strategies, optimize marketing budgets, and personalize customer experiences. This data-driven approach has the potential to increase conversion rates substantially, directly improving revenue and profitability while enhancing overall customer satisfaction (Content Square, 2025).

# Objectives

This project is guided by the following business hypothesis: by predicting purchasing intention and conversion patterns, business can use machine learning to identify high-potential customers, create better strategies, and significantly improve their conversion rates.

This capstone has four clear objectives:

1. To build and test different machine learning models (like Logistic Regression and Random Forest) that can predict which visitors will make a purchase. The goal is to create models that are accurate enough for businesses to trust and use in their decision-making.

2. To find out what factors make people more likely to buy. This includes analyzing visitor behavior (how they browse the website), timing factors (like special days or weekends), and user characteristics (new vs. returning customers). Understanding these patterns will help businesses know what drives purchases.

3. To create practical strategies that businesses can use to increase conversions. These strategies will be customized for different types of customers, with specific recommendations for engaging each group effectively.

4. To present the findings and recommendations through clear visualizations that communicate insights effectively to business stakeholders. This will help managers understand the results and make data-driven decisions without needing technical expertise.

# Problem Definition

A large number of people who visit e-commerce sites do not make a purchase. In fact, only 2-3% of visitors purchase products, leaving a whopping 97-98% of people empty-handed (Oberlo, 2024). This is extremely frustrating for businesses as they spend a considerable amount of money on advertisements to attract visitors to sites, yet barely any of them turn out to be paying customers. Companies are essentially throwing money at people who wouldn't purchase in the first place.

The critical problem here is that businesses cannot distinguish between a visitor who is considering a purchase and one who is simply browsing. Without this predictive ability, businesses end up treating visitors in the same way, unable to identify who is likely to make a purchase. This means that they will waste money on marketing campaigns targeting visitors who are not interested in purchasing.

This is a problem that is costing businesses substantial money. It is impacting marketing budgets as they are spent on the wrong people, and potential sales are lost. A small increase in conversion rates will result in a large increase in the bottom line. This makes developing an effective solution essential for survival in today’s competitive marketplace.

# Scope

This project uses the Online Shoppers Purchasing Intention Dataset from UCI, containing 12,330 sessions over one year. The work involves building and comparing machine learning models—Logistic Regression and Random Forest—to predict purchase behavior. The analysis will identify key factors influencing purchases, create customer segments based on behavior patterns, and develop business recommendations. Deliverables include trained models, factor analysis, customer insights, strategic recommendations, and visualizations.

The project will not involve implementing models on live websites, collecting new data, or building e-commerce platforms. It also excludes specific product analysis and detailed pricing strategies. Instead, the focus is on demonstrating how predictive modeling can help businesses understand and improve conversion rates using existing data.

# Methodology

This project follows the CRISP-DM methodology, a widely-used framework for data analytics projects. The six phases are:

1. Business Understanding - Define the conversion optimization problem and project objectives

2. Data Understanding - Explore the dataset's 12,330 sessions and understand patterns in the data

3. Data Preparation - Clean and prepare data for analysis, selecting relevant features

4. Modeling - Build machine learning models (Logistic Regression, Random Forest) to predict purchases

5. Evaluation - Compare models using accuracy, precision, and recall to select the best performer

6. Deployment - Create clear visualizations that communicate insights effectively to business stakeholders.

The CRISP-DM framework ensures the analysis remains focused on business needs while maintaining analytical rigor throughout the project.

# Data Source

This project uses the Online Shoppers Purchasing Intention Dataset from UCI Machine Learning Repository (Sakar and Kastro, 2018), containing 12,330 sessions from one year. The dataset includes 17 features: behavioral metrics (page visits, bounce rates, time spent), temporal factors (month, weekends, special days), and user characteristics (new/returning visitors, browser, region). The target variable shows whether each session resulted in a purchase. The dataset is publicly available, licensed for academic use, and contains no missing values.

Ethical Consideration

This project uses anonymized, publicly available data with no personal identifiers, ensuring participant privacy. The dataset complies with data protection regulations as it contains no names, email addresses, or payment information. The research focuses on behavioral patterns rather than individual tracking. All findings and recommendations will emphasize improving customer experience rather than manipulative practices. The project will be conducted transparently, acknowledging any limitations in the data or methodology.

# References

[https://www.statista.com/topics/871/online-shopping/#topicOverview](https://www.statista.com/topics/871/online-shopping/" \l "topicOverview)

https://www.oberlo.com/statistics/average-ecommerce-conversion-rate

https://contentsquare.com/guides/customer-retention/rate/

https://archive.ics.uci.edu/dataset/468/online+shoppers+purchasing+intention+dataset

ContentSquare (2025) Customer retention rate: what it is and why it matters. Available at: https://contentsquare.com/guides/customer-retention/rate/ (Accessed: 30 October 2025). Oberlo (2024) Average ecommerce conversion rate. Available at: https://www.oberlo.com/statistics/average-ecommerce-conversion-rate (Accessed: 30 October 2025). Statista (2025) E-commerce worldwide - statistics & facts. Available at: https://www.statista.com/topics/871/online-shopping/ (Accessed: 30 October 2025)

Appendix