A logo for college computing

Description automatically generated

**Assessment Cover Page**

|  |  |
| --- | --- |
| *Student Full Name* Leticia Andrade Vieira |  |
| *Student Number*  2025304 |  |
| *Module Title*  Strategic Thinking |  |
| *Assessment Title*  CA 1- Casptone Project Proposal |  |
| *Lecturer/Supervisor*  Taufique Ahmed |  |
| *Assessment Due Date*  29th October, 2025 at 23:59 |  |
| *Date of Submission*  29th October, 2025 |  |

**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.

**Use of AI Tools**

I acknowledge the use of ChatGPT (GPT-5) to assist in proofreading the report. The tool was used to improve academic language and clarity, although all ideas, analyses, and final conclusions presented in this report are my own.

# **Table of Contents**

1. **Introduction ....................................................................................** 2
2. **Business Understanding .................................................................** 2
3. **Problem Definition .........................................................................** 2
4. **Objectives .......................................................................................** 3
5. **Scope ...............................................................................................** 3
6. **Methodology ...................................................................................** 4
7. **Data Sources ...................................................................................** 4
8. **Ethical Considerations .................................................................**5
9. **Conclusion .......................................................................................** 5
10. **References .....................................................................................** 5
11. Appendix **.....................................................................................** 5

# Assignment Title: CA 1 – Capstone Project Proposal

# Project Title:Predicting Customer Purchase Intentions

# Introduction

No digital marketplace, e-commerce is essential for business success. Online sales are necessary to reach customers, strengthen brand presence, and drive growth. Consumer behavior has evolved: buyers are more demanding and expect personalized experiences.

Every online interaction, from page views to session duration, generates behavioral data. Many companies fail to use this data effectively to predict purchase intentions, resulting in missed opportunities.

This project applies data analysis and machine learning to estimate the likelihood of a purchase based on browsing behavior. The dataset, from the UCI Machine Learning Repository, includes over 12,000 sessions with features such as page views, bounce rates, and visitor types. Using CRISP-DM, the study analyzes these data to uncover behavioral patterns for data-driven marketing (Erliana, 2025; Alizamir et al., 2022).

# Business understanding

In e-commerce, companies generate large volumes of user interaction data such as page views, session duration, and visitor type. Properly analyzed, this information reveals patterns indicating purchase intentions. Understanding and predicting customer behavior is crucial to improving marketing efficiency and user experience.

However, many organizations underuse this data, missing opportunities to anticipate customer actions and optimize marketing. This project applies data analysis and machine learning to identify behavioral patterns that influence purchase decisions, supporting data-driven marketing.

The findings aim to provide insights that enable companies to design more personalized and effective strategies (Sakar & Kastro, 2018; Alizamir et al., 2022).

# Problem Definition

E-commerce businesses collect extensive behavioral data but often struggle to extract insights that enhance marketing effectiveness or predict customer decisions. The key problem is the insufficient conversion of these data into reliable indicators of purchase intention, leading to missed sales and suboptimal marketing investments. Addressing this gap with advanced analytics and machine learning can improve customer experience, targeting, and sales performance.

# Objectives

# The main goal is to analyze e-commerce user behavior data and identify the factors that most influence purchase intentions, providing a basis for business-driven strategies. Specifically, the objectives are:

# Review literature on customer purchase behavior and data-driven analytics in e-commerce for a theoretical foundation.

# Prepare and examine data from the UCI repository to ensure analytical relevance and quality.

# Conduct exploratory data analysis (EDA) to reveal patterns between variables and purchasing decisions.

# Present initial findings on behavioral drivers and guide subsequent predictive modeling for marketing strategies.

# The central business objective is to support increased e-commerce conversion and marketing efficiency through robust behavioral insights.

# Scope

This two-semester project uses the Online Shoppers Purchasing Intention Dataset (over 12,000 sessions, 18 features: page views, bounce rate, visitor type, weekend activity). The first semester is devoted to data understanding and exploration; the second to predictive modeling and business recommendations, all following the CRISP-DM methodology.

### Inclusions

* Data collection and preparation
* EDA and visualization of behaviors
* Development and evaluation of predictive models
* Reporting results and marketing recommendations

### Exclusions

* Use of personal or sensitive data from real customers.
* Real-time deployment of predictive models.
* Financial forecasting or return-on-investment (ROI) analysis.

### Boundaries

* The project is limited to the dataset available in the UCI Machine Learning Repository.
* The analysis focuses only on user behavioral and session data.

# Methodology

The CRISP-DM model structures the analytical process. The project starts with understanding business context and key factors, followed by data cleaning and EDA to detect patterns. Supervised machine learning techniques (Logistic Regression, Decision Trees, Random Forest) will predict purchase intent, balancing interpretability and accuracy. Model evaluation will use accuracy, precision, recall, and F1-score, with results communicated clearly for both technical and business users.

### Expected Deliverables by the End of Semester 2:

* A cleaned, documented dataset.
* EDA report with key patterns.
* Predictive models with performance metrics.
* Actionable recommendations for marketing decisions.

# Data Sources

The dataset used for this project is the *Online Shoppers Purchasing Intention Dataset* from the UCI Machine Learning Repository (Dua & Graff, 2019). It contains 12,330 user sessions and 18 features, including numerical and categorical variables such as page views, bounce rate, visitor type, and weekend activity. This publicly available dataset does not contain any personal or sensitive information and therefore does not require special permissions for use. It provides sufficient data for conducting exploratory and predictive analyses to identify behavioral patterns and purchasing intentions. All analysis will be conducted in Python using Jupyter Notebook.

# Ethical Considerations

This project does not involve the use of personal, confidential, or sensitive data. The dataset is publicly available for academic purposes and complies with ethical standards for research. All data will be used responsibly and exclusively for educational analysis. Proper attribution will be given to the data source, following the Harvard Referencing style, to ensure academic integrity and avoid plagiarism. Additionally, the project will prioritize accuracy, transparency, and ethical handling of data in all analytical stages (Zook et al., 2017).

# Conclusion

This project provides a structured foundation to model and interpret e-commerce customer behavior, offering data-driven insights for better marketing decisions. By identifying core factors behind purchase intentions, it directly supports improved user targeting and conversion strategies. The next phases will develop and apply predictive models to operationalize these findings for real-world business value.

# References

Alizamir, S., Li, M. and Zhan, Y. (2022) ‘Predictive analytics in e-commerce: Understanding customer behavior through data mining and machine learning’, Journal of Business Analytics, 5(2), pp. 134–150.

Erliana, T. (2025) ‘Artificial intelligence applications in online retail: Enhancing customer engagement and sales prediction’, International Journal of Digital Business, 12(1), pp. 45–59.

Khanal, P., Shrestha, R. and Gautam, B. (2018) ‘Understanding predictive variables in supervised machine learning models’, Journal of Data Science and Analytics, 4(3), pp. 77–89.

Sakar, C.O. and Kastro, Y. (2018) ‘A real-time customer purchase prediction model for e-commerce’, Expert Systems with Applications, 97, pp. 89–100.

UCI Machine Learning Repository (2018) ‘Online Shoppers Purchasing Intention Dataset’. Available at: <https://archive.ics.uci.edu/ml/datasets/online+shoppers+purchasing+intention+dataset> (Accessed: 28 October 2025).

# 

# Appendix

All project-related materials will be available on the GitHub repository at: <https://github.com/leticiavieira03/-CA1-Strategic-Thinking-student2025304>