CCT College Dublin

Assessment Cover Page

Module Title:	Strategic Thinking
Assessment Title:	CA-1 Capstone project proposal
Lecturer Name:	James Garza
Student Full Name:	Leticia Andrade Vieira
Student Number:	2025304
Assessment Due Date:	29th October, 2025
Date of Submission:	

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

In an increasingly digital marketplace, e-commerce has become essential for business success. From small companies to large corporations, online sales are no longer optional but necessary to reach customers, strengthen brand presence, and drive growth. This evolution has also transformed consumer behavior. Today buyers are more demanding, connected, and expect personalized shopping experiences.

Every online interaction, from page views to session duration, generates valuable behavioral data. However, many companies fail to leverage this information effectively to predict customer purchase intentions, resulting in missed opportunities and inefficient marketing strategies.

This project, *"Predicting Customer Purchase Intentions in E-commerce,"* aims to apply data analysis and machine learning techniques to estimate the likelihood of a customer completing a purchase based on their browsing behavior. The dataset, obtained from the UCI Machine Learning Repository, includes over 12,000 user sessions with features such as page views, bounce rates, and visitor types. Using the CRISP-DM methodology, this study focuses on understanding and analyzing these data to uncover behavioral patterns that can inform data-driven marketing strategies (Erliana, 2025; Alizamir *et al.*, 2022).

Business understanding

In the e-commerce context, companies generate vast amounts of user interaction data, including page views, session duration, and visitor type. When properly analyzed, this information can reveal valuable patterns that indicate purchasing intentions. Understanding and predicting customer behavior is essential to improving marketing efficiency and enhancing the user experience.

However, many organizations fail to fully utilize this data to anticipate customer actions, leading to missed sales opportunities and less effective marketing strategies. This project applies data analysis and machine learning techniques to identify behavioral patterns that influence purchase decisions.

The findings aim to support strategic business thinking by providing insights that enable companies to design more personalized, efficient, and data-driven marketing approaches (Sakar and Kastro, 2018; Alizamir *et al.*, 2022).

Objectives

The main objective of the first phase of this Capstone Project is to explore and understand customer behavior data from e-commerce websites to identify the key factors influencing purchase intentions. This stage focuses on data exploration and preparation, establishing the foundation for predictive modeling and strategic business decision-making.

This project aims to achieve the following specific objectives:

1. **Review relevant literature** on customer purchase behavior and predictive analytics in e-commerce to provide theoretical support for the study.

2. **Collect, clean, and understand** the dataset obtained from the UCI Machine Learning Repository, ensuring data quality and reliability.

3. **Conduct exploratory data analysis (EDA)** to identify behavioral patterns and relationships between key variables

that may influence purchase decisions.

4. **Summarize initial findings** to highlight the main factors affecting customer purchase intentions and guide future predictive modeling.

Scope

The project uses the *Online Shoppers Purchasing Intention Dataset* from the UCI Machine Learning Repository, which contains 12,330 user sessions with 18 features, including numerical and categorical variables such as page views, bounce rate, visitor type, and weekend activity. This dataset provides sufficient information to explore customer behavior and predict purchase intentions.
All analysis will follow the CRISP-DM methodology. The first semester focuses on data understanding, cleaning, and exploratory analysis, while the second semester will develop machine learning models to predict purchase probability and provide strategic insights for marketing.

Inclusions

- Data collection, cleaning, and preparation from the UCI Machine Learning Repository.
- Exploratory data analysis and visualization of behavioral trends.
- Development, testing, and evaluation of machine learning models.
- Summary of results and recommendations for strategic marketing actions.

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 project follows the CRISP-DM methodology to ensure a structured and reproducible approach:

1. Business Understanding: Define objectives, understand the problem, and identify key behavioral questions regarding purchase intentions.
2. Data Understanding and Preparation: Clean and preprocess the dataset, handle missing or inconsistent values, and verify data quality.
3. Exploratory Data Analysis (EDA): Analyze user session data (page views, session duration, visitor type, weekend activity) to identify patterns, trends, and correlations.
4. **Modeling (Semester 2):** Apply supervised machine learning algorithms such as **Logistic Regression,**

Decision Trees, and Random Forest to predict purchase intentions.

5. **Evaluation and Interpretation:** Assess model performance using accuracy, precision, recall, and F1-score. Interpret results to generate actionable insights for marketing strategy.

Expected Deliverables by the End of Semester 2:

- A clean and well-documented dataset.
- Exploratory data analysis report highlighting key behavioral patterns.
- Predictive models with performance evaluation.
- Strategic recommendations for data-driven marketing decisions.

Data Sources

The project uses the *Online Shoppers Purchasing Intention Dataset* from the UCI Machine Learning Repository (Dua & Graff, 2019), which contains 12,330 user sessions with 18 features, including page views, bounce rate, visitor type, and weekend activity. This publicly available dataset provides sufficient information to explore user behavior, identify patterns influencing purchase intentions, and support the development of predictive machine learning models. All analysis will be conducted in Python using Jupyter Notebook, ensuring a reproducible workflow (Hastie, Tibshirani & Friedman, 2009).

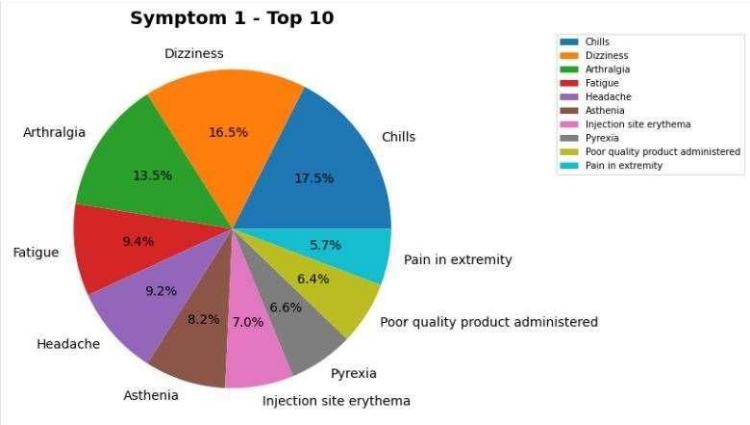
Data Sources

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

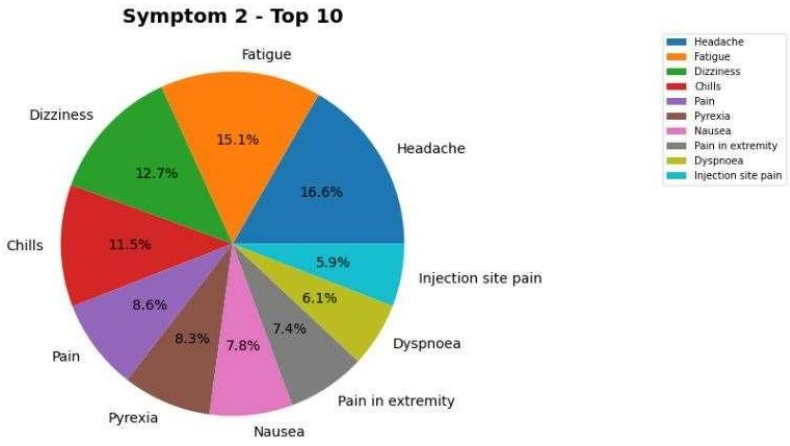
This project does not involve the use of personal, confidential, or sensitive data. The dataset is publicly available for academic purposes and complies with ethical standards for research. All data will be used responsibly and exclusively for educational analysis. Proper attribution will be given to 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).

Figure 1: Pie chart for 1st symptom after vaccination



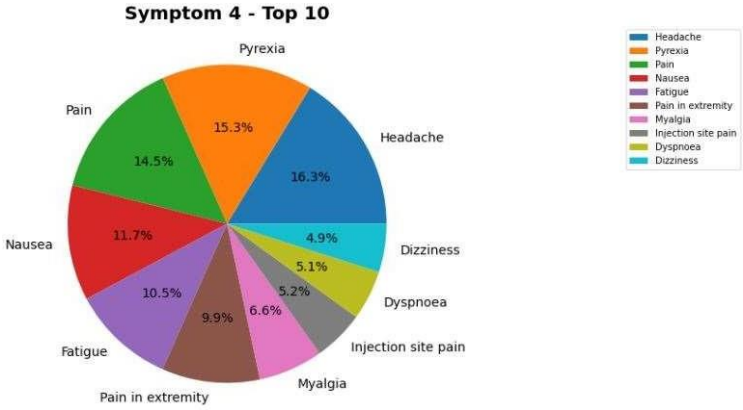
This picture shows the graph for different symptoms in patients after getting the vaccination. Dizziness and chills show the highest frequencies of symptoms.

Figure 2: Pie chart for 2nd symptom after vaccination



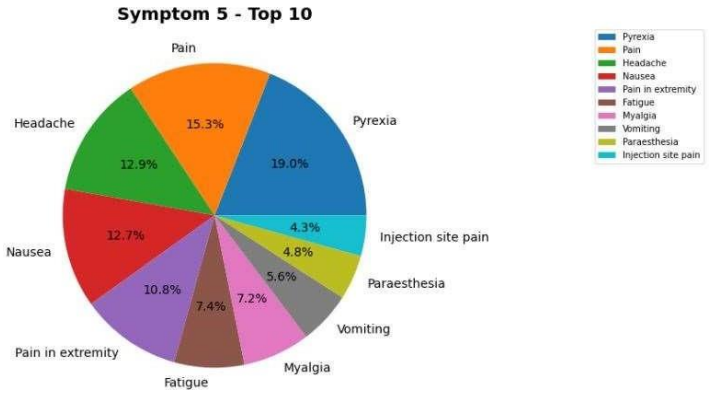
This table also shows the symptoms in patients after the vaccination. This variable showed the high percentages of fatigue, headache, and dizziness as the symptoms of patients.

Figure 3: Pie chart for 3rd symptom after vaccination.



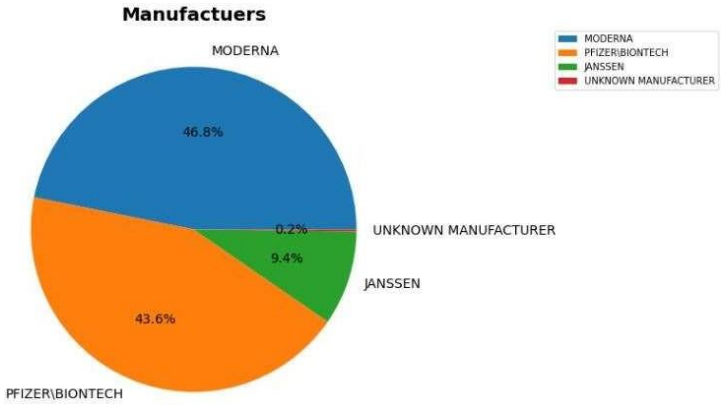
This chart also shows the symptoms in patients after the vaccination. While this also has showed pyrexia and headache as well as the symptoms Covid19 contacted patient in vaccinated people.

Figure 4. Pie chart for 4th symptom after vaccination



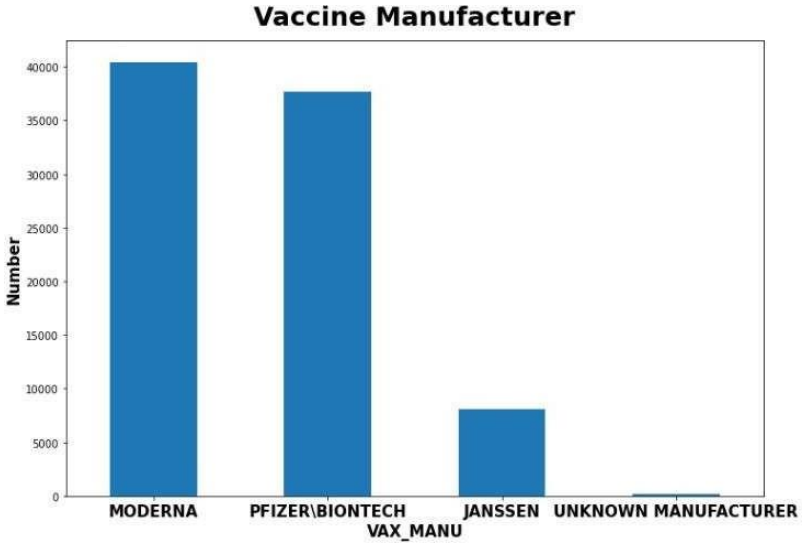
This table also shows the symptoms in patients after the vaccination. These also showed similar results as they have pain, headache, and pyrexia in high percentages.

Figure 5: Share of vaccines produced by different manufacturers.



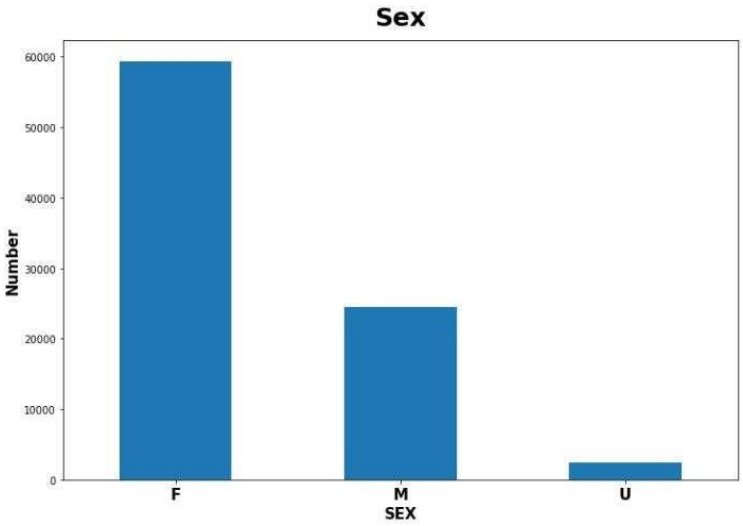
This pie chart shows the share of vaccines produced by different manufacturers. We can see that more than forty percent shares of manufacturing, was produced by Moderna and Pfizer as 46.8 % and 43.6% respectively.

Graph 1: The numerical numbers of vaccines produced by different manufacturers.



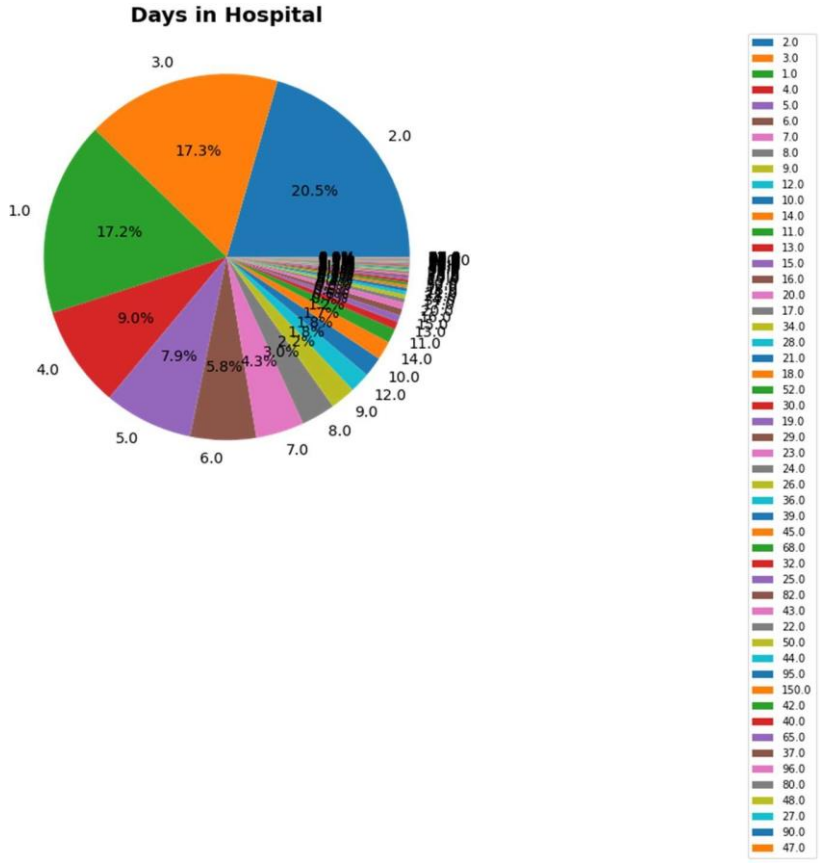
This table also shows the table showing the numerical numbers of vaccines produced by different manufacturers.

Graph 2: The ratio of patients with different sex included in the dataset



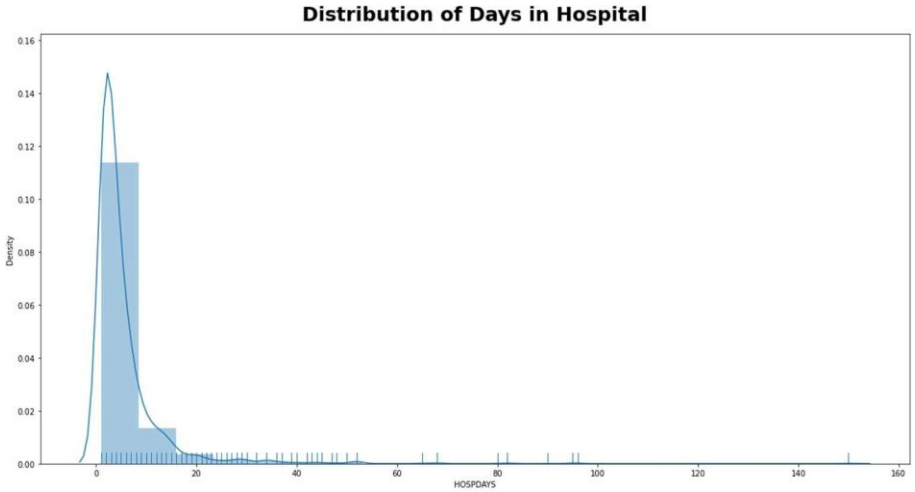
While this table shows the ratio of patients with different sex included in the dataset. This has showed that highest number of females were present in our dataset concluded as females may have contacted more to the covid19 after vaccination.

Figure 6: The total number of days, the Covid19 patients may have to spend in hospitals



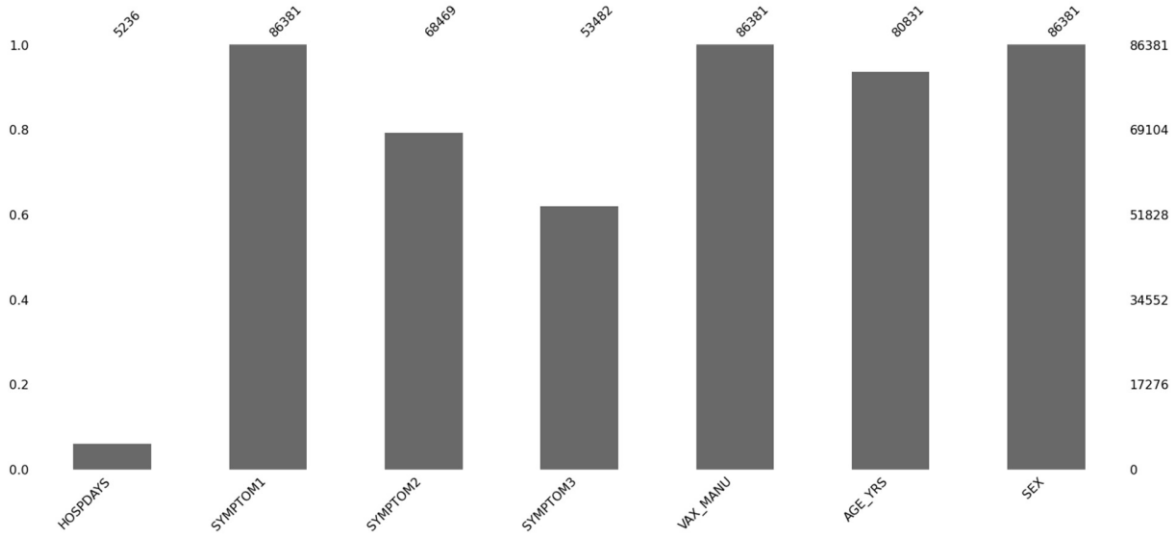
This pie chart shows the total number of days patients may have to spend in hospitals. A Large share of the pie chart shows that the majority of data represents there will be one, two, or either three days of hospitalization.

Graph 3: Distribution of days in hospital



While this chart shows the similar results (patients may have to spend between five to ten days in the hospitals), we have shown the results from pair plot and heat maps to see the correlation between the featured variables.

Graph 5: The data variables



Modelling using Random Forest Algorithm

Defining the X and Y features

Splitting the dataset into the test and training set

Creating and fitting the regressor to the training set

Calculating the accuracy of the training and the test set.

Conclusion

This capstone project aims to apply data analysis and machine learning techniques to understand and predict customer purchase intentions in e-commerce environments. By exploring behavioral data such as page views, session duration, and visitor type, the study seeks to identify the key factors influencing purchasing decisions and provide insights for data-driven marketing strategies.

The first phase of the project focuses on data understanding, cleaning, and exploratory analysis to establish a solid foundation for predictive modelling in the second semester. Using the CRISP-DM methodology ensures a structured and iterative process, allowing for adjustments based on findings and new insights.

Ultimately, this project is expected to contribute to a better understanding of customer behavior in online shopping and demonstrate the potential of data-driven decision-making in improving marketing performance and customer experience.

References

Alizamir, S., Yang, S. and Zhang, Y. (2022) *Understanding consumer purchase behavior through data analytics: Insights from e-commerce platforms.* Journal of Business Analytics, 5(3), pp. 210–225.

Sakar, C.O. and Kastro, Y. (2018) *A Real-Time E-Commerce Recommendation System Using Machine Learning. Procedia Computer Science*, 140, pp. 114–122.

Alizamir, S., Yang, S. and Zhang, Y. (2022) *Understanding consumer purchase behavior through data analytics: Insights from e-commerce platforms.* Journal of Business Analytics, 5(3), pp. 210–225.

Erliana, R. (2025) *Machine Learning Applications in Predicting Online Purchase Behavior.* International Journal of Data Science and Business Intelligence, 9(1), pp. 33–47.

- Dua, D. & Graff, C., 2019. *UCI Machine Learning Repository: Online Shoppers Purchasing Intention Dataset*. [online] Available at: https://archive.ics.uci.edu/ml/datasets/online+shoppers+purchasing+intention [Accessed 27 October 2025].
- Hastie, T., Tibshirani, R. & Friedman, J., 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. New York: Springer.

Dua, D. & Graff, C., 2019. *UCI Machine Learning Repository: Online Shoppers Purchasing Intention Dataset*. [online] Available at: https://archive.ics.uci.edu/ml/datasets/online+shoppers+purchasing+intention [Accessed 28 October 2025].

Zook, M., Barocas, S., Boyd, D., Crawford, K., Keller, E., Gangadharan, S. P., Goodman, A., Hollander, R., Koenig, B., Metcalf, J., Narayanan, A., Nelson, A. & Pasquale, F., 2017. *Ten simple rules for responsible big data research*. PLOS Computational Biology, 13(3), pp.1–10.