**Project Handout**

**Introduction**

The purpose of this project is to build a machine learning model that predicts the likelihood of a customer buying a product based on their demographics and purchase history. The dataset used for this project contains information on 10,000 customers and their purchase behavior. The goal of this project is to identify the key factors that influence a customer's decision to buy a product and use this information to improve sales and marketing strategies.

**Data**

The dataset used for this project contains the following features:

Customer ID

Age

Gender

Income

Occupation

Marital Status

Product Category

Purchase Amount

The dataset was preprocessed to handle missing values and remove any duplicate entries. A total of 9,200 entries were used for training the machine learning model, and the remaining 800 entries were used for testing the model's accuracy.

**Analysis**

The first step in the analysis was to explore the dataset and identify any correlations between the features. We found that the customer's age, gender, income, and occupation were all strongly correlated with their purchase behavior. We also found that customers who were married tended to spend more money than those who were single.

Next, we used a decision tree classifier to build the machine learning model. We chose this model because it is easy to interpret and can handle both categorical and numerical data. The metric we used to evaluate the model's accuracy was accuracy score, which is the percentage of correctly predicted instances. Our model achieved an accuracy score of 75%, which is a good starting point.

**Future Work**

There are several additional things that can be done in the future to improve the accuracy and usefulness of this model. Some of these include:

Feature engineering: We can create new features based on the existing ones, such as combining age and income to create a new feature that measures purchasing power.

Ensemble methods: We can use ensemble methods such as bagging and boosting to improve the accuracy of the model.

Hyperparameter tuning: We can fine-tune the model's hyperparameters to find the best combination of parameters that maximizes the accuracy score.

Online learning: We can implement an online learning algorithm that allows the model to adapt to new data in real-time, improving its accuracy and usefulness over time.

Overall, this project provides a good foundation for predicting customer behavior and can be used to inform sales and marketing strategies. With additional work, we can improve the accuracy and usefulness of the model and make it even more valuable for businesses looking to improve their sales and marketing efforts.