Introduction:

The project for CS-513 “Knowledge discovery and Data Mining” is based on Classification and Neural Networks, which the course curriculum covers. The Black Friday dataset from Kaggle will be used towards working on this project.

The link for the dataset is given below:

https://www.kaggle.com/mehdidag/black-friday

This is a dataset of 550000 observations of the black Friday in a retail store, it contains different kinds of variables either numerical or categorical.

The dataset consists of the following columns with their description:

Columns

ID	User_ID	User ID
String	Product_ID	Product ID
String	Gender	Sex of User
String	Age	Age in bins
Integer	Occupation	Occupation
String	City_Category	Category of the City (A,B,C)
String	Stay_In_Current_C	Number of years stay in current city
Integer	Marital_Status	Marital Status
Integer	Product_Category	Product Category
Integer	Product_Category	Product may belongs to other category also
Integer	Product_Category	Product may belongs to other category also
Integer	Purchase	Purchase amount in dollars

Objective:

- Finding in which City -> Category of the product which is being sold based on the other independent variables. Keeping **City_Category** as the target variable (Using different classification models and comparing their accuracy).
- We will be working towards using Neural Nets in predicting the purchase amount of the product, having **Purchase** as the target variable.

Future Scope:

- We will try predicting the marital status of the user who bought this product, having **Marital_Status** as the target variable.
- We will also try predicting which **Age** category of people who buy expensive products during black Friday sale (Having the purchase amount more than 10000 being expensive).

There are many other ideas that we have come up with for our future predictions.