TASK:

After reading the assigned articles and textbook chapters, how can you use exploratory data analysis techniques to uncover patterns or relationships in data that are not immediately apparent? Give an example of a dataset and explain how you would approach the task of finding hidden patterns or relationships. How can Python be used to develop a useful tool in this context?

What is Exploratory Data Analysis?

Exploratory Data Analysis (EDA) is like exploring a new place. You look around, observe things, and try to understand what's going on. Similarly, in EDA, you look at a dataset, check out the different parts, and try to figure out what's happening in the data.

About Dataset

Overview:

This dataset provides a comprehensive overview of customer interactions with an online retail store, aiming to predict customer churn based on various behavioral and demographic features. It includes data on customer demographics, spending behavior, satisfaction levels, and engagement with marketing campaigns. The dataset is designed for analysis and development of predictive models to identify customers at risk of churn, enabling targeted customer retention strategies.

Description of Columns:

- Customer_ID: A unique identifier for each customer.
- Age: The customer's age.
- Gender: The customer's gender (Male, Female, Other).
- Annual_Income: The annual income of the customer in thousands of dollars.
- Total_Spend: The total amount spent by the customer in the last year.
- Years_as_Customer: The number of years the individual has been a customer of the store.
- Num_of_Purchases: The number of purchases the customer made in the last year.
- Average_Transaction_Amount: The average amount spent per transaction.
- Num_of_Returns: The number of items the customer returned in the last year.
- Num_of_Support_Contacts: The number of times the customer contacted support in the last year.
- Satisfaction_Score: A score from 1 to 5 indicating the customer's satisfaction with the store.
- Last_Purchase_Days_Ago: The number of days since the customer's last purchase.

- Email_Opt_In: Whether the customer has opted in to receive marketing emails.
- Promotion_Response: The customer's response to the last promotional campaign (Responded, Ignored, Unsubscribed).
- Target_Churn: Indicates whether the customer churned (True or False).

Approche:

- Statistiques descriptives : nous calculerons des statistiques récapitulatives (moyenne, médiane, écart type) pour les variables numériques et des décomptes de fréquence pour les variables catégorielles afin de mieux comprendre la distribution des données.
- Exploration visuelle : créez des visualisations telles que des histogrammes, des diagrammes en boîte et des nuages de points pour explorer les relations entre les variables.
- Analyse de corrélation : calculez les corrélations entre les variables pour identifier toute relation significative.

Dataset: https://www.kaggle.com/datasets/hassaneskikri/online-retail-customer-churn-dataset