

# Exploratory Data Analysis: Pakistan E-commerce Dataset

---

## Objective:

- To Understand Data preprocessing steps for Tabular data
- To do Exploratory Data Analysis
- To Understand trends in Pakistan Largest E-commerce Data

**Key Terms:** Data Preprocessing, Exploratory Data Analysis, Pakistan E-commerce

## Context:

This is the largest retail e-commerce orders dataset from Pakistan. It contains half a million transaction records from March 2016 to August 2018. The data was collected from various e-commerce merchants as part of a research study. Dataset is released as a capstone project for data science course at Alnafi (alnafi.com/zusmani). There is a dire need for such a dataset to learn about Pakistan's emerging e-commerce potential and I hope this will help many startups in many ways.

## Dataset:

The Dataset is available publicly on kaggle and can be downloaded from this [link](#).

## Assignment 1. Data Preprocessing

- To load the dataset
- Class Balance check
- Clean the Dataset (Handling structural errors, Null values, duplicates, outliers,...e.t.c.)
- Convert Ordinal or Categorical Data into Numeric
- Transform the Data (Normalization, Standardization, Scalings techniques)
- Feature Engineering (Feature generation and selection).

## Assignment 2. Exploratory Data Analysis

- What is the best-selling category?
- Visualize payment method and order status frequency
- Find a correlation between payment method and order status
- Find a correlation between order date and item category
- Find any hidden patterns that are counterintuitive for a layman