

Project 1 : Exploratory Data Analysis

- Steffi Dorothy

Task 1

Pakistan's Largest E-Commerce Dataset - <https://www.kaggle.com/datasets/zusmani/pakistans-largest-ecommerce-dataset/data>

Introduction

This dataset represents the largest collection of retail e-commerce orders from Pakistan, comprising half a million transaction records spanning from March 2016 to August 2018. The data was gathered from various e-commerce merchants as part of a research study. It includes comprehensive details about these orders, such as item specifics, shipping methods, payment options (including credit card, Easy-Paisa, Jazz-Cash, and cash-on-delivery), product categories (such as fashion, mobile, electronics, and appliances), as well as the order date, SKU, price, quantity, total, and customer ID. This is the most detailed publicly available dataset on e-commerce in Pakistan.

Content

- Rows : 500k
- Columns : 25
- Unit Of Analysis : E-Commerce Orders
- Time Period : 03/2016 - 08/2018
- Geography : Pakistan

Motivation : Selecting an e-commerce dataset provides valuable insights into consumer behavior, market trends, and business performance in the digital retail space. Analyzing such a dataset allows for a deeper understanding of purchasing patterns, popular product

categories, payment methods, and regional preferences. This information is crucial for businesses aiming to optimize their operations, improve customer experience, and make data-driven decisions. Additionally, studying e-commerce data can reveal emerging trends, helping businesses stay competitive and adapt to the rapidly evolving market. The comprehensive nature of e-commerce datasets makes them ideal for exploring various aspects of retail economics and consumer dynamics.