

# Pakistan E-commerce Dashboard

**Tool Used:** Microsoft Excel

**Dataset:** Pakistan's Largest E-commerce Dataset

**Time Period Covered:** February 2016 – May 2017

**Total Records Analyzed:** 584,359 orders

**Total Revenue Processed:** PKR 4,986,068,316

---

## 1. Project Background and Purpose

This project was designed to analyze a large, real-world e-commerce dataset from Pakistan using Microsoft Excel. The dataset contains order-level data that spans from early 2016 through mid-2017 and includes product, pricing, payment, category, and customer information. The aim of the project was not only to clean and organize the raw data, but also to extract meaningful business insights and present them through a well-structured, interactive Excel dashboard.

As an aspiring data analyst, my goal was to simulate a real-world business intelligence workflow: start from raw transactional data, transform it into an analysis-ready format, and build a user-friendly dashboard that decision-makers could use to track performance and make informed decisions.

---

## 2. Understanding the Dataset Structure

The dataset included the following relevant columns:

- **item\_id:** A unique identifier for each item sold (used for internal tracking but not central to analysis)
- **status:** Indicates the outcome of an order such as 'complete', 'canceled', or 'refunded'
- **created\_at:** A timestamp column used to determine when each order was placed
- **sku:** The product SKU code; not deeply analyzed in this project, but it was available for reference
- **price:** The unit price of the item ordered
- **qty\_ordered:** The number of units sold in each transaction
- **grand\_total:** The final amount paid for the order after discounts and charges

- **increment\_id**: The order number, used internally to distinguish transactions
  - **category\_name\_1**: The product category for each item (e.g., Mobiles & Tablets, Appliances, Entertainment, etc.)
  - **sales\_commission\_completed**: This field was not used in the current dashboard, but potentially useful for future seller-level analysis
  - **discount\_amount**: Any discounts applied to the order
  - **payment\_method**: The payment type used by the customer (Cash on Delivery, Easypay, Jazzwallet, etc.)
  - **Working Date** and **BI Status**: Metadata, which I excluded from the core dashboard but could be useful for staging or quality assurance
  - **MV**: Likely short for Market Value or Merchant Value – not used in this analysis due to limited context
  - **Year, Month, Customer Since, M-Y, and FY**: Date-related columns that I refined or created for time-based analysis
  - **Customer ID**: A unique identifier used to track customer behavior and group their orders over time
- 

### 3. Data Cleaning and Preparation (in detail)

Since the raw dataset was relatively large (over half a million rows), careful cleaning and preprocessing were essential before I could begin building visuals or insights.

The most important step was to **standardize and format the date-related data**. The `created_at` column contained both date and time in a full timestamp. I extracted the year and month from this to create separate columns (Year, Month, and Month-Year or M-Y). These allowed for proper time-based grouping and comparison.

Next, I verified and adjusted the financial columns:

- I checked whether the `grand_total` matched the result of `price × qty_ordered`, factoring in the `discount_amount`. In most cases it did, but I kept `grand_total` as the main field for revenue analysis because it represented the actual money paid after adjustments.
- I removed any rows where the `grand_total` was null, zero, or clearly incorrect (e.g., very large outliers not supported by quantity or pricing).

**Order statuses** were standardized next. The status column included a mix of lowercase and inconsistently labeled values (e.g., “complete”, “completed”, “Complete”). I grouped these into three core statuses:

- **Completed:** Successful transactions
- **Canceled:** Orders placed but canceled before fulfillment
- **Refunded:** Orders that were completed but later returned

Then, I focused on the **categorical fields**. The category\_name\_1 column was cleaned by removing trailing spaces and ensuring consistent spelling. I ended up with five meaningful product categories:

1. Mobiles & Tablets
2. Appliances
3. Entertainment
4. Women’s Fashion
5. Computing

Some entries were blank or categorized as "Other", but I excluded these from the category-level breakdown for clarity.

For the **payment methods**, I mapped the values to five distinct groups:

- **Cash on Delivery**
- **Payaxis**
- **Easypay**
- **Jazzwallet**
- **Easypay Voucher**

Null or unknown values were excluded or grouped into “Other” if needed.

The **customer behavior columns** Customer ID and Customer Since were especially valuable for customer-level analysis. I used these to identify:

- First-time vs returning customers
- Repeat buyers
- Top customers by number of orders

## 4. Source of Data

This dataset was obtained from Kaggle, where it was published by a contributor named **zusmani**. It's titled "**Pakistan's Largest E-commerce Dataset**" and is based on real transactional data from one of the country's leading online retailers. The dataset contains over **584,000 records**, representing detailed sales transactions including pricing, customer behavior, product categories, and payment methods, spanning across **15 months** from early 2016 to mid-2017.

Link to dataset: [Kaggle – Pakistan's Largest E-commerce Dataset](#)

Using this dataset allowed me to explore the structure and scale of real e-commerce operations, simulate business analytics processes, and build a functional dashboard using just Excel.

---

## 5. Key Insights and Business Analysis

After cleaning and preparing the data, I moved into the **exploration and visualization phase**. My focus was on identifying **meaningful business patterns** rather than just building charts. I wanted the dashboard to answer real business questions like:

- Where is our revenue coming from?
- When are sales highest?
- What products and categories perform best?
- What payment methods do customers prefer?
- Who are our most loyal or valuable customers?
- How often are orders canceled or refunded?

Here's a detailed breakdown of the analysis and findings:

---

### 5.1 Revenue and Order Overview

The platform generated **PKR 4.98 billion** in sales from **584,359** total orders. The average transaction value was roughly **PKR 8,528**, though this varied widely by category and time of year. Orders ranged from low-ticket items in fashion to high-value electronics in computing and mobile devices.

Out of the total records:

- **Completed orders** accounted for the largest share, around **70,000+**
- **Canceled orders** represented a surprisingly high percentage, signaling possible fulfillment or customer service challenges
- **Refunded orders** were less common but still measurable

I used this information to build a visual breakdown of order status by volume, helping identify how many transactions were successful versus lost revenue opportunities.

---

## 5.2 Time-Based Sales Trends

By using the created\_at, Year, Month, and M-Y columns, I visualized **monthly revenue trends** across the 15-month timeline. A clear pattern emerged:

- **Sales peaked in November**, particularly in **November 2016**, where there were over **190,000 orders**, coinciding with likely marketing events like **Black Friday** and end-of-year promotions
- The months of **July and August** also showed strong performance, potentially tied to Eid or summer shopping behaviors
- **Sales dropped significantly in early months like February and March**, possibly due to seasonality or limited campaigns

This analysis allowed me to highlight not just the raw performance numbers, but also their strategic context — for example, helping a business plan promotions more effectively.

---

## 5.3 Product Category Performance

One of the most important sections in the dashboard analyzed the **category\_name\_1** field.

The five main product categories with clean and usable labels were:

- Mobiles & Tablets
- Appliances
- Entertainment
- Women's Fashion
- Computing

Among these, **Mobiles & Tablets alone generated more than 59% of total revenue**, far outpacing all others. This makes sense considering the higher average order value and consistent demand for smartphones and accessories.

Appliances came second, contributing around 16%, followed by other categories that were smaller in scale. This insight helped illustrate how dependent the business was on electronics, and how diversification might be needed for long-term stability.

---

## 5.4 Payment Method Preferences

Using the payment\_method column, I created a breakdown of transaction volume by payment type. This gave a clear view of customer preferences and operational dependencies.

Key findings:

- **Cash on Delivery (COD)** was the dominant method, used in over **250,000 orders**
- Digital wallets and online payments (like **Easypay, Jazzwallet, and Payaxis**) were used far less frequently
- COD preference suggests potential logistical overhead and higher return risk

This analysis helped underline a business risk: heavy dependence on COD could mean lower pre-paid revenue, higher cancellation rates, and more delivery failures.

---

## 5.5 Top Customer Behavior

Using the customer\_id and customer\_since columns, I tracked repeat purchases and identified **the most active and valuable customers**. These were customers who placed multiple orders, across multiple months, often in high-value categories.

I ranked customer IDs by:

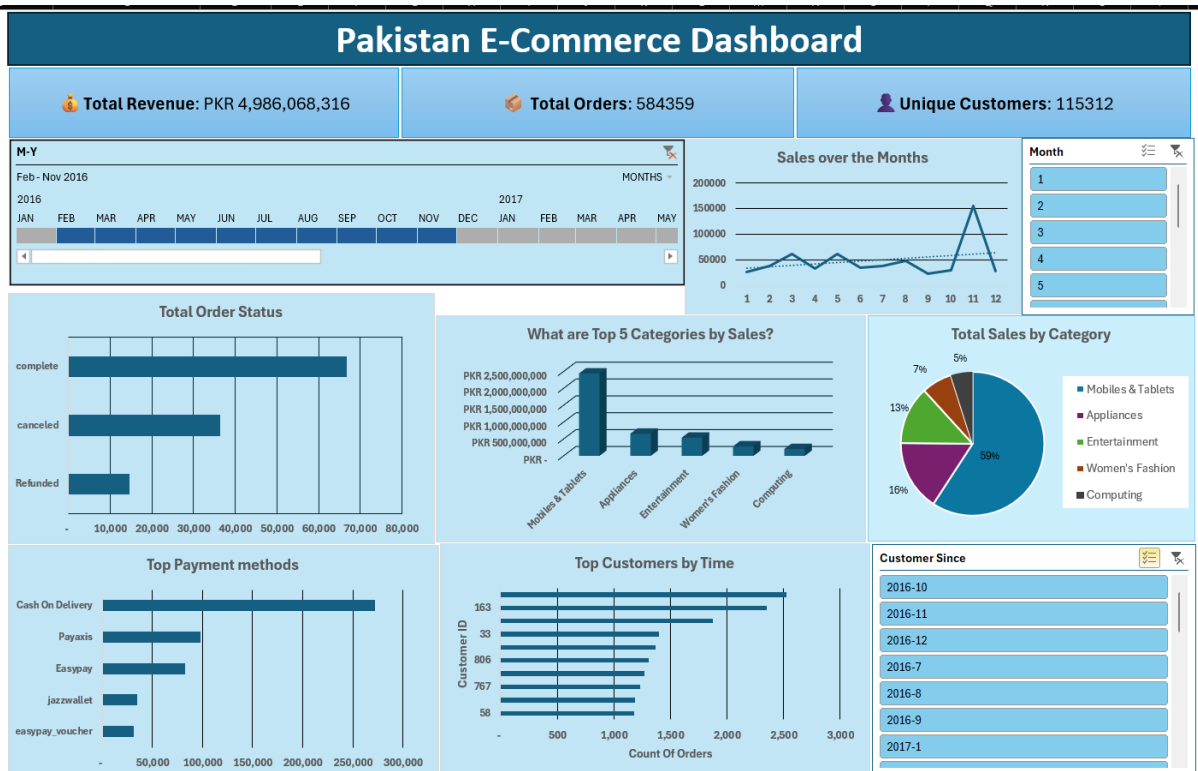
- Number of orders
- Total revenue generated
- Length of time as a customer (based on customer\_since and M-Y)

This helped me simulate **customer cohort analysis** grouping buyers by when they first joined the platform and how their behavior evolved over time.

## 6. Dashboard Design and Structure

The dashboard was created entirely in **Microsoft Excel**, using a combination of **pivot tables, charts, formulas, named ranges**, and **interactive slicers**. The goal was to design a dashboard that could be used by business stakeholders — such as sales managers or operations leads to **explore performance trends independently**, without needing to modify or filter raw data manually.

I built the dashboard on a **dedicated sheet**, separating it from the data and pivot sources, to maintain clarity and avoid accidental edits.



## Layout Overview

The layout of the dashboard follows a top-to-bottom flow:

1. **KPI Summary Section (Top Row)**
2. **Time-Based Sales Trends (Line Chart)**
3. **Product Category Performance (Bar and Pie Charts)**
4. **Order Status Breakdown (Column Chart)**
5. **Payment Method Preference (Donut Chart)**
6. **Top Customers by Order Volume (Horizontal Bar Chart)**
7. **Interactive Slicers** (Right side or top for filters)

Each section was designed to tell a specific part of the business story, and all visuals were **dynamically connected** through shared pivot caches and slicers.

---

### 6.1 KPI Summary Section

At the top of the dashboard, I created a **summary of key performance indicators**, including:

- **Total Revenue** (sum of grand\_total)
- **Total Orders** (count of increment\_id)
- **Unique Customers** (count of distinct customer\_id)
- **Average Order Value** (Total Revenue ÷ Total Orders)

These metrics are shown in **visually distinct boxes**, often using conditional formatting (green for growth, red for decline) and bold fonts. They immediately give the viewer a sense of business scale and health.

---

### 6.2 Time-Based Sales Trends

Using the extracted Month-Year (M-Y) field, I created a **line chart** to visualize revenue over time.

- This chart shows **monthly revenue fluctuations** over the 15-month period



- **November 2016** shows a clear spike in orders, likely from **Black Friday** or seasonal campaigns
- The time trend helps spot **seasonality**, **growth**, or **sales dips** that may require business action

I also made this chart responsive to **slicers**, so a stakeholder could filter by product category, payment method, or order status and immediately see how those segments performed over time.

---

### 6.3 Category-wise Sales Breakdown

The category\_name\_1 column was grouped and cleaned, then used to create:

- A **bar chart** showing total revenue by category
- A **pie chart** showing the **percentage share** of each category

**Mobiles & Tablets** dominates the chart contributing 59% of all revenue which is immediately obvious to any viewer. The pie chart helps visualize **relative performance**, while the bar chart shows **absolute sales figures**.

This visual lets business leaders easily see which categories are high performers and which are under-leveraged.

---

### 6.4 Order Status Analysis

Using the cleaned status column, I visualized how many orders were:

- Completed
- Canceled
- Refunded

This section uses a **stacked column chart** or a **segmented bar chart**, allowing viewers to quickly assess how many transactions were successful vs lost. This is essential for understanding **operational efficiency** and identifying problems in fulfillment or customer experience.

In a future version, this could be linked to customer cohorts or product categories to see if certain types of orders get canceled more often.

---

## 6.5 Payment Method Breakdown

A **donut chart** was created to visualize order volume by payment\_method. This chart makes it clear that:

- **Cash on Delivery** was by far the most used method
- Digital methods like **Easypay**, **Jazzwallet**, and **Payaxis** accounted for a small fraction of transactions

This visualization has strategic importance. It can inform business decisions such as promoting digital payments, offering incentives for prepaid orders, or understanding where COD might be causing delivery failures.

---

## 6.6 Top Customers

This section ranks customers by the number of orders placed. I used a **horizontal bar chart** to display:

- Top 10 customers by count of increment\_id
- Optionally, top customers by total grand\_total

This helps simulate **customer segmentation**, identifying power users or loyal buyers who may be eligible for loyalty programs, upsells, or retention campaigns.

---

## 6.7 Interactive Slicers

To make the dashboard dynamic, I added Excel **slicers** connected to the pivot tables. These included filters for:

- Month-Year
- Product Category
- Order Status
- Payment Method

Slicers allow users to **drill down** into the data with a single click. For example, they can instantly see:

- How November 2016 performed compared to March 2017
  - What product categories dominated during Eid or holiday seasons
  - What happens to revenue when only “Completed” orders are included
- 

## 7. Technical Skills Demonstrated

This project showcases a strong command of Excel, including:

- **Pivot tables** for grouped analysis
- **Slicers** for interactivity
- **Formulas** like AVERAGEIFS, COUNTIF, VLOOKUP, IFERROR, TEXT, and CONCATENATE
- **Named ranges** and hidden helper columns for clean visuals
- **Conditional formatting** for KPI indicators
- **Chart customization** for clear, professional visuals

Everything was done within Excel no VBA, Power Query, or external tools to simulate what a data analyst might build in a company using standard Microsoft Office tools.

## 8. Reflection & Key Learnings

This project deepened my understanding of how raw transactional data can be transformed into actionable insights. I learned how to:

- Clean and structure large datasets in Excel for business analysis
- Build dynamic dashboards using pivot tables, charts, and slicers
- Extract KPIs and interpret performance trends across time, product, and customer segments
- Simulate a real business scenario and deliver insights in a visual, user-friendly format

It also helped me improve my storytelling and taught me how to connect data back to business goals — a critical skill for any data analyst.

Project Done By: Muhammad Annas @ Air University