

Guide to Exploratory Data Analysis

Exploratory Data Analysis (EDA) – A Practical Guide for Analysts

In my journey to becoming a better data analyst, I came across a great breakdown of how to approach Exploratory Data Analysis (EDA) effectively. I've compiled and adapted the key takeaways here as a guide for others who are building their own projects or preparing for interviews.

This framework is designed to help you **think like an analyst**, break down complex business problems, and avoid getting lost in the weeds during the analysis process.

Why EDA Is Crucial

EDA is often the **first major step in any data analysis workflow**—especially for early-career data analysts. It typically forms **30–40% of the work** in entry-level roles and is one of the most important skills you'll be assessed on during interviews.

A well-structured EDA:

- Helps you uncover meaningful insights.
- Builds trust with stakeholders.
- Makes your analysis direction clear and actionable.
- Helps communicate technical findings to non-technical audiences.

Start with the Right Questions

Before jumping into the data, it's important to understand the **business context** and clarify the problem. Here's a checklist I use:

- What's the objective of the analysis?
- Who are the stakeholders?
- What decision is being made based on this data?
- What are the important metrics and dimensions?
- Where is the data located, and how is it structured?

The EDA Framework

A helpful mental model I follow breaks down EDA into three main steps:

1. **Identify the type of question being asked**
2. **Break down the main question into smaller, more manageable sub-questions**
3. **Map each question to a suitable technique or approach**

4 Types of EDA Questions

Most business analysis requests can be grouped into one of these categories:

Type	Description	Example
Overall Trends	Analyze historical patterns	What are the overall sales trends?
Growth Rates	Track changes over time	What are the monthly and yearly growth rates?
Performance Measurement	Evaluate program or feature success	How did the loyalty program perform?
KPI Reporting	Summarize specific metrics	What are our refund rates or average order values?

4 Core Techniques for Sub-Questions

Each of the questions above can be further explored using the following techniques:

1. Seasonality Analysis

- How do metrics behave monthly/quarterly?
- Are there repeating seasonal spikes or dips?

2. Dimensional Segmentation

- How do metrics vary by region, product, or customer type?
- Which segments perform better or worse?

3. Summary Statistics

- What's the average, maximum, or minimum value?
- Which month or region had the best/worst performance?

4. Distribution Analysis

- What portion of total performance comes from each segment?
- Do a small number of products/regions contribute most to results?

These help break down larger business questions into targeted analysis goals.

Example: EDA in Excel (Sales Trends)

To explore trends in sales data, here's a basic workflow I followed using Excel:

1. **Build a Pivot Table**
 - o Metrics: Total Sales, Average Order Value (AOV), Number of Orders
 - o Rolled up to a **monthly level** for temporal analysis
2. **Calculate Monthly Growth Rates**
 - o Added a column for MoM % changes
3. **Highlight Trends Visually**
 - o Used conditional formatting to flag highs/lows
4. **Generate Summary Stats**
 - o Extracted average, min, and max values
5. **Plot Line Graphs**
 - o Visualized total sales trends with key peaks and troughs highlighted

This simple approach helped surface actionable insights—even before involving SQL or Python.



Sample Insight Format

From 2019 to 2022, sales peaked in December 2020 with 4K orders totaling \$1.2M. This likely reflects pandemic-driven consumer behavior. However, the holiday spike was missing in 2022. Investigating marketing strategies or customer behavior during that time could help explain the decline.

The goal is to turn raw data into clear, concise, and **business-relevant insights**.



How I Structure My EDA Projects

To avoid getting overwhelmed and stay productive, I follow a structured workflow:

1. **Spend 1–2 hours cleaning and understanding the data**
2. **Pick one high-level business question**
3. **Select 1–2 related sub-questions to explore**
4. **Spend 4–6 hours doing focused analysis**
5. **Spend 1–2 hours reflecting on insights and documenting them**

I use a mix of tools like:

- Excel (Pivot Tables, Charts, Conditional Formatting)
- Power BI/Tableau
- SQL (for querying data)
- Python (for advanced analysis)