



# **DATA ANALYST INTERNSHIP**


## **TASK - 1**

By Kapil Pandey



# WHO IS DATA ANALYST?

A data analyst collects, cleans, and analyzes large datasets from sources like databases, spreadsheets, or APIs. The core responsibilities include:

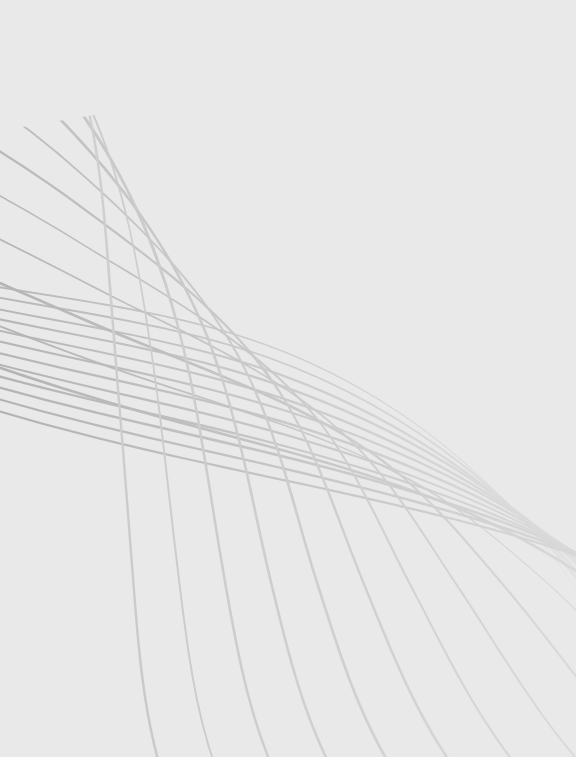
- Data cleaning and preparation: Ensuring data is accurate and ready for analysis.
  - Exploratory analysis: Using statistics to uncover patterns, trends, and anomalies.
  - Data visualization: Creating charts, graphs, and dashboards for easier interpretation.
  - Reporting results: Presenting insights in a way that supports business decision-making.
  - Collaboration: Working with different departments to solve problems using data.
- 



# Q/A SECTION

**Q1.** What skills are most important for a data analyst?

**A1.** Analytical thinking, proficiency in tools like Excel, SQL, and visualization platforms (Tableau, Power BI), and communication






**Q2. How do data analysts create value?**

**A2.** They reveal actionable insights that drive operational improvements, customer satisfaction, and strategic growth.

**Q3. What are core responsibilities of a data analyst?**

**A3.** A data analyst is responsible for gathering data, cleaning and preprocessing it, analyzing and interpreting results, visualizing insights via charts/dashboards, and communicating findings to stakeholders.



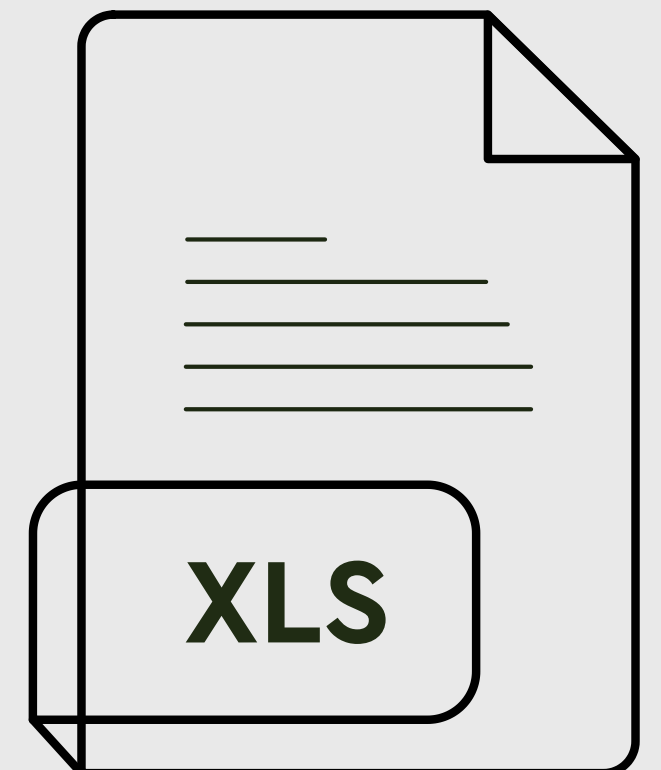
## TASK - 2

# ANALYSIS OF DATASET IN EXCEL



**NAME OF THE DATASET :**

**Amazon - Prime Insights**





# OVERVIEW:

- The dataset contains a comprehensive list of movies and TV shows available on Amazon Prime Video, detailing their attributes such as title, type, cast, genre, and more. This cleaned version appears to be structured and ready for deeper analysis, reporting, or integration into other systems.
- The dataset is a single table with **409+ records** (from s1 to at least s409), each representing a unique title on the platform. The data is organized into **12 columns**, providing a multifaceted view of each title.



# DATASET PREPROCESSING IN EXCEL

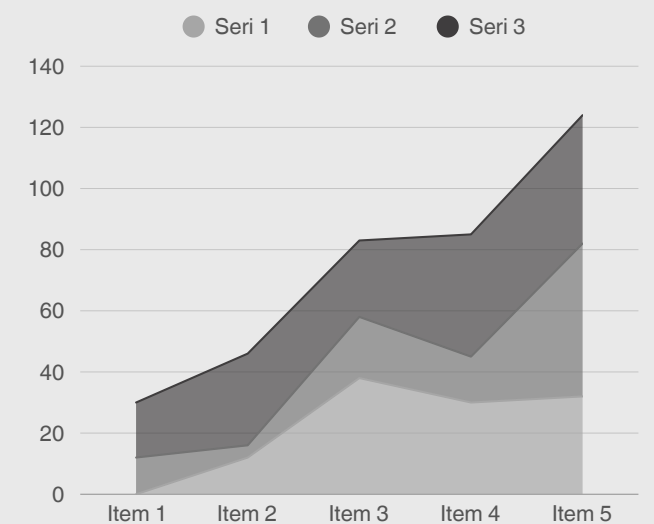
Here's what I actually did while cleaning the Amazon Prime dataset:

1. Removed Duplicates → Based on **show\_id** column.

2. Handled Missing Values:

- country → filled with "Unknown"
- director → "Unknown"
- cast → "Unknown"
- rating → "Not Rated"
- date\_added → "Unknown"

3. Standardized Text → Trimmed spaces, fixed column names, and cleaned **title**.

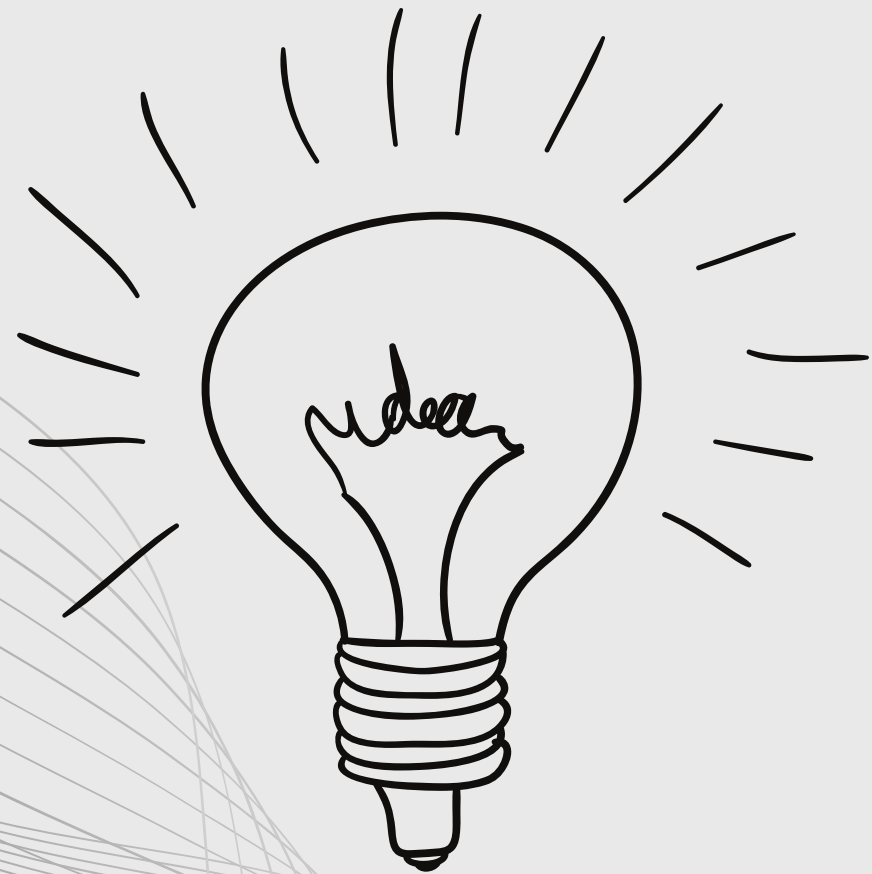


## TASK - 3

# DATA VISUALIZATION

**MAKE AN INTERACTIVE  
DASHBOARD USING:**

**Power BI**








# OBJECTIVE

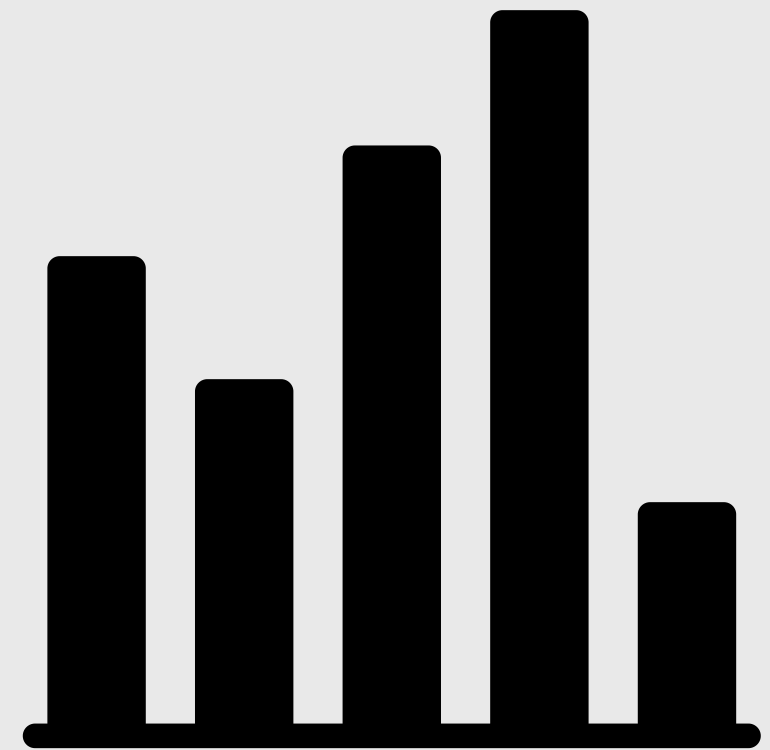
The objective of this dashboard is to analyze the Amazon Prime Titles dataset and extract meaningful insights about the type of content available, the most common genres, distribution of ratings, contribution by countries, and trends in release years.

## Key Metrics (KPIs)

- Total Titles: 9,655
  - Total Ratings: 25
  - Total Genres: 519
  - Total Directors: 5,771
  - Release Years Covered:  
1920 – 2021
- 

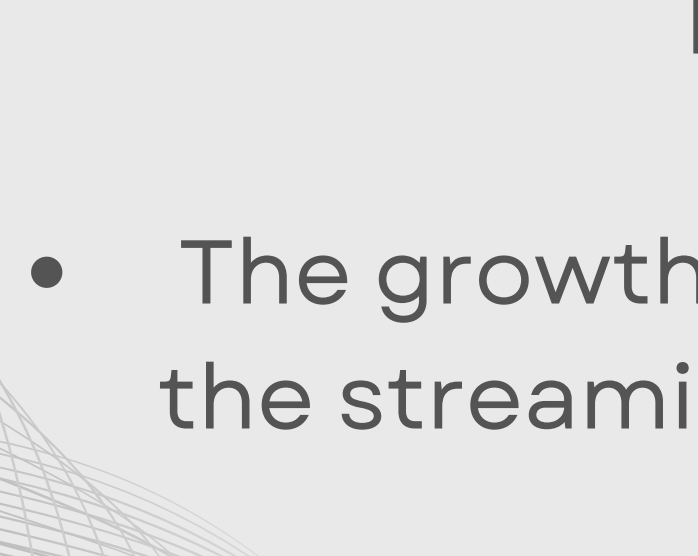
# Visualizations

- **Ratings Distribution:** Shows that 13+ is the most common rating.
- **Genre Analysis:** Drama, Comedy, and Action-Drama dominate the platform.
- **Country-wise Distribution:** Most titles are from the US, India, UK, and Canada.
- **Content Type Split:** About 81% Movies vs 19% TV Shows.
- **Trend Over Time:** Significant growth in content releases after 2000, peaking around 2020.





# INSIGHTS & STORYTELLING

- Amazon Prime has positioned itself as a movie-first platform, which differentiates it from competitors like Netflix.
  - The genre dominance of Drama and Comedy shows Prime caters to mass entertainment and family audiences.
  - India's strong contribution highlights the importance of regional and Bollywood content in Prime's global strategy.
  - The growth trend post-2010 proves Amazon's aggressive push into the streaming industry, especially after 2015 when OTT competition intensified.
- 

The image features a minimalist design with the words "THANK YOU" centered in a bold, dark grey, sans-serif font. The background is a light grey. In the top-left and bottom-left corners, there are decorative elements consisting of numerous thin, dark grey lines that curve and overlap, creating a sense of movement and depth.

**THANK YOU**