Task - 5

📊 Netflix Data Analysis – Exploratory Data Analysis (EDA)

🗓️ Date: 30-06-2025
📁 Task: Exploratory Data Analysis (EDA)

📌 Project Overview

This project performs **Exploratory Data Analysis (EDA)** on the **Netflix Movies and TV Shows** dataset. The goal is to extract business insights, understand content trends, and build familiarity with EDA using Python libraries like `pandas`, `matplotlib`, and `seaborn`.

🔧 Tools & Technologies

- Python
- Google Colab
- Pandas, NumPy
- Matplotlib, Seaborn

📂 Dataset

- Dataset Name: `netflix_titles.csv`
- Source: Netflix Movies and TV Shows on Kaggle

🚀 Steps Followed

1. Importing Libraries

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')

plt.style.use('seaborn')
sns.set_palette("pastel")
```

2. Loading the Dataset

```
from google.colab import files
uploaded = files.upload()

df = pd.read_csv('netflix_titles.csv')
df.head()
```

- Uploaded and loaded the dataset into a DataFrame to begin exploration.

3. Initial Data Exploration

```
df.info()
df.describe()
df.isnull().sum()
```

- Explored data types, null values, and descriptive statistics.

4. Handling Missing Values

```
df['director'].fillna('Not Available', inplace=True)
df['cast'].fillna('Not Available', inplace=True)
df['country'].fillna('Not Available', inplace=True)
df.dropna(subset=['date_added', 'rating'], inplace=True)
```

- Filled or dropped missing values in important columns to ensure clean analysis.

5. Feature Engineering

```
df['date_added'] = pd.to_datetime(df['date_added'])
df['year_added'] = df['date_added'].dt.year
df['month_added'] = df['date_added'].dt.month

df['duration'].fillna('Not Available', inplace=True)
df['duration_type'] = df['duration'].apply(lambda x: 'Season(s)' if 'Season' i
```

- Extracted new time-based and content-type features to enhance analysis.

📊 Exploratory Data Analysis (EDA)

6. Content Type Distribution

```
sns.countplot(data=df, x='type')
plt.title('Distribution of Content Type')
```

- Shows that Netflix has more Movies than TV Shows.

7. Rating Distribution

```
plt.figure(figsize=(12,6))
sns.countplot(data=df, y='rating', order=df['rating'].value_counts().index)
plt.title('Distribution of Ratings')
```

- Reveals most content is rated TV-MA and TV-14, targeting adult audiences.

8. Top 10 Countries by Content

```
top_countries = df['country'].value_counts().head(10)
sns.barplot(x=top_countries.values, y=top_countries.index)
plt.title('Top 10 Countries by Content')
```

- The United States, India, and the UK lead in content production.

9. Content Added Over Time

```
sns.histplot(data=df, x='year_added', bins=20, hue='type', multiple='stack')
plt.title('Content Added Over Years')
```

- Netflix's content library grew rapidly post-2015, especially in 2018–2020.

10. Duration Type by Content Type

```
sns.countplot(data=df, x='duration_type', hue='type')
plt.title('Duration Type by Content Type')
```

- Duration matches content format: Movies in minutes, Shows in seasons.

11. Correlation Heatmap

```
sns.heatmap(df[['year_added', 'month_added']].corr(), annot=True, cmap='coolwa
plt.title('Correlation Heatmap')
```

- No strong correlation among time-based engineered features, as expected.

🔍 Key Observations

- Movies dominate the catalog compared to TV Shows.
- Netflix content is mostly aimed at mature teens and adults (TV-MA, TV-14).
- USA, India, and UK are the top content contributors.
- Content growth spiked post-2015, especially around 2018–2020.
- Duration fields align well with content types (minutes for movies, seasons for shows).

- Missing values were handled using fill/drop strategies for a clean dataset.

📝 Conclusion

This project demonstrates a complete EDA workflow: data loading, cleaning, feature engineering, visualization, and insight generation — providing a clear picture of Netflix's content strategy and global expansion. These insights can further be used to build models for recommendation systems or content trend forecasting.

📁 Deliverables

- netflix_eda_AdityaKankarwal.ipynb – Jupyter Notebook

✨ Done by: **Aditya Kankarwal**