Phone Usage in India: Data Analysis Report

1. Dataset Overview

The dataset provides insights into phone usage behavior in India. It contains information on various user demographics, phone usage metrics, and spending habits. Key columns include:

- **Demographics**: Age Group, Gender, Location
- **Usage Metrics**: Screen Time (hrs/day), Data Usage (GB/month), Monthly Recharge Cost (INR)
- **Spending Behavior**: E-commerce Spend (INR/month)

Dataset Highlights:

- Total Records: [Provide count from dataset]
- Missing Values: [Summary of missing values identified]
- Data Source: Provided as CSV file
- Purpose: Analyze patterns in phone usage and spending behaviors

2. Data Cleaning Steps Taken

1. Handling Missing Data:

- **Numerical Columns**: Missing values were filled with the mean or median.
- **Categorical Columns**: Missing values were filled with the mode.

2. Duplicate Removal:

- Identified and removed duplicate rows to ensure data integrity.

3. Outlier Detection and Handling:

- Used the IQR method to identify outliers in Screen Time (hrs/day) and Monthly Recharge Cost (INR), replacing them with upper/lower bounds.

4. Data Transformation:

- Converted Screen Time (hrs/day) into a categorical column (e.g., Low, Medium, High).
- Standardized Monthly Recharge Cost (INR) for improved analysis.

5. Logical Validation:

- Verified consistency in columns like Age Group and Gender.
- Ensured numerical values (e.g., Data Usage) were within reasonable limits.

3. Key Insights from Visualizations

1. Trend Analysis:

- **Observation**: A line chart showed monthly trends in data usage, with peaks during holiday seasons.
- **Insight**: Data usage increases during specific times, likely due to seasonal events.

2. Distribution Analysis:

- **Observation**: Histogram of Screen Time (hrs/day) revealed most users spend 3-5 hours on their phones daily.
- **Insight**: Indicates moderate phone usage as a common behavior.

3. Categorical Insights:

- **Observation**: Bar chart comparing Age Group and Gender against Monthly Recharge Cost (INR) revealed higher spending among younger users.
- **Insight**: Younger demographics tend to spend more on phone-related expenses.

4. Correlation Analysis:

- **Observation**: Scatter plot showed a strong positive correlation between Screen Time (hrs/day) and Data Usage (GB/month).
- **Insight**: Higher screen time is directly associated with increased data usage.

5. Advanced Visualization:

- **Observation**: Heatmap revealed strong correlations between E-commerce Spend (INR/month) and Monthly Recharge Cost (INR).

- **Insight**: Users who spend more on e-commerce also tend to have higher recharge costs.

4. Challenges Faced and Solutions Implemented

1. Handling Missing Data:
- **Challenge**: Missing values in key columns.
- **Solution**: Applied appropriate filling methods (mean, median, mode) to retain data integrity.

2. Outlier Management:
- **Challenge**: Significant outliers in numerical columns skewed results.
- **Solution**: Used the IQR method to replace outliers with upper/lower bounds.

3. Data Transformation:
- **Challenge**: Numerical data needed conversion for better interpretability.
- **Solution**: Categorized Screen Time and normalized Recharge Cost to standardize scales.

4. Visualization Setup in Power BI:
- **Challenge**: Ensuring accurate representation of trends and correlations.
- **Solution**: Utilized appropriate chart types and added slicers for interactivity.