Google Play Store Apps –

Q1. Data Loading & Inspection
Load the Google Play Store dataset and display the first few rows. Examine the shape of the dataset, column names, data types, and check for missing values.

Q2. Summary Statistics
Generate descriptive statistics for all numerical columns and identify which features may influence app ratings and installs.

Q3. Duplicate Records Handling
Check whether duplicate records are present in the dataset. If found, remove them and report how many rows were deleted.

Q4. Category Exploration
List all unique values in the *Category* column and identify any inconsistent or invalid categories. Remove such records from the dataset.

Q5. Missing Value Treatment & Rating Categorization
Handle missing values in the *Rating* column. Create a new column called Rating_category where:
Rating > 3.5 → High
Rating ≤ 3.5 → Low

Q6. Distribution of Rating Category
Plot the count distribution of the Rating_category variable and interpret the results.

Q7. Reviews Column Cleaning & Outlier Treatment
Convert the *Reviews* column into numeric format. Detect extreme values and apply a logarithmic transformation to reduce skewness.

Q8. Cleaning the Size Column
Process the *Size* column by:
Converting "M" to millions
Converting "K" to thousands
Dropping rows with "Varies with device"
Converting the column to numeric type.

Q9. Cleaning the Installs Column
Remove special characters such as "+" and commas from the *Installs* column and convert it into numeric format.

Q10. Cleaning the Price Column
Remove the "$" symbol from the *Price* column and convert it to numeric values.

Q11. Dropping Irrelevant Columns
Drop columns such as *App*, *Rating*, *Genres*, *Last Updated*, *Current Ver*, and *Android Ver*. Explain why these variables are excluded from modeling.

Q12. Encoding Categorical Variables
Apply Label Encoding to convert categorical columns (*Category*, *Content Rating*, *Type*, and *Rating_category*) into numeric form.

Q13. Feature–Target Separation
Separate the dataset into:
Target variable → Rating_category
Predictor variables → remaining columns.

Q14. Train–Test Split
Split the dataset into training and testing sets in an appropriate ratio.

Q15. Feature Scaling
Standardize the feature values so that all numerical variables are on the same scale.