



Google Play Store App Analysis

This presentation details the data cleaning process. It also covers the dashboard creation for analyzing Google Play Store apps. We aim to uncover trends and insights within the app ecosystem. The process covers data handling from raw dataset to actionable visualizations.

Dataset Overview

The dataset contains over 9658 app records. It also includes 13 columns such as App, Category, Rating, and Reviews. Further columns include Size, Installs, and Price. Expect to find missing values and inconsistencies. Data cleaning is therefore a necessary step.

13 Columns

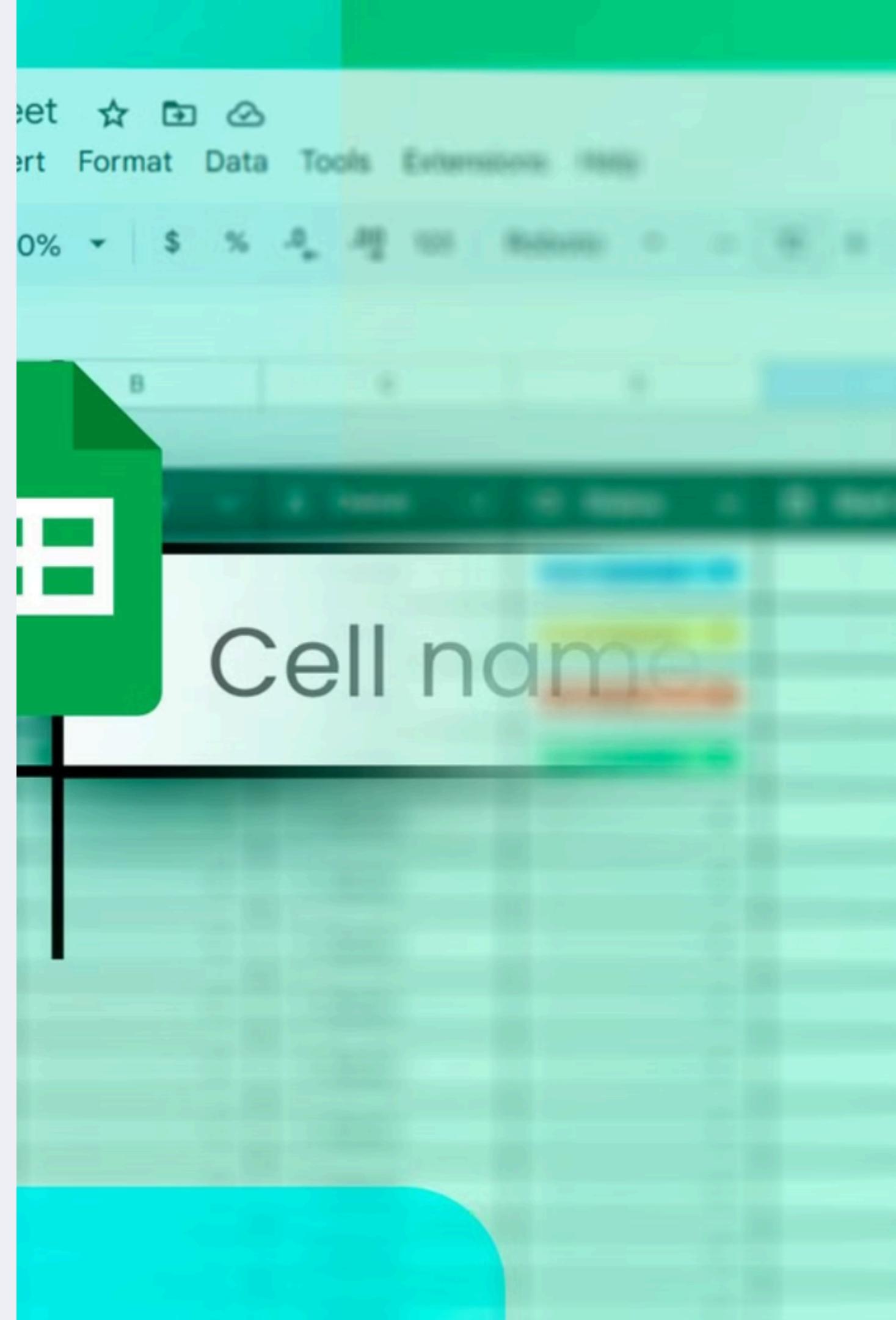
Key app attributes

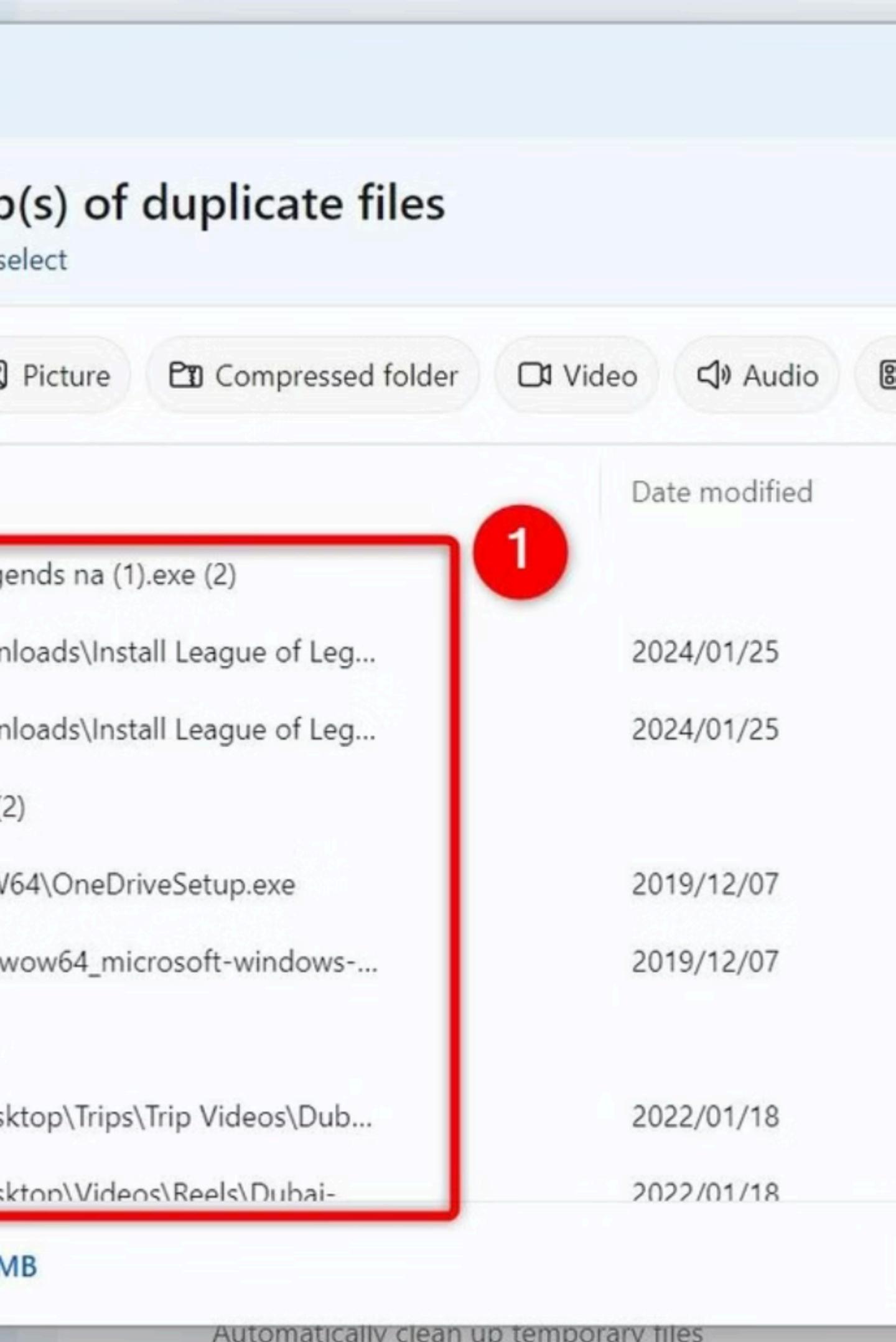
9658 Records

Wide app coverage

Inconsistencies

Requires cleaning





Removing Duplicate Entries

Duplicate entries can skew analysis results. Removing duplicates ensures accurate data interpretation. We identified and removed exact duplicates. The process used 'App' and other key fields for identification. Only unique app entries were preserved.

1 Identify Duplicates

Based on key fields

2 Remove Duplicates

Ensure unique entries

3 Accurate Analysis

Prevents skewed results

Handling Missing Values

Missing and invalid values can compromise data integrity. Apps with ratings exceeding 5 were removed. Also, apps with missing ratings were removed. Missing entries in 'Type' and 'Content Rating' were filled or removed. We focused on maintaining data accuracy.

1 Remove Invalid Ratings

Ratings > 5

2 Handle Missing Entries

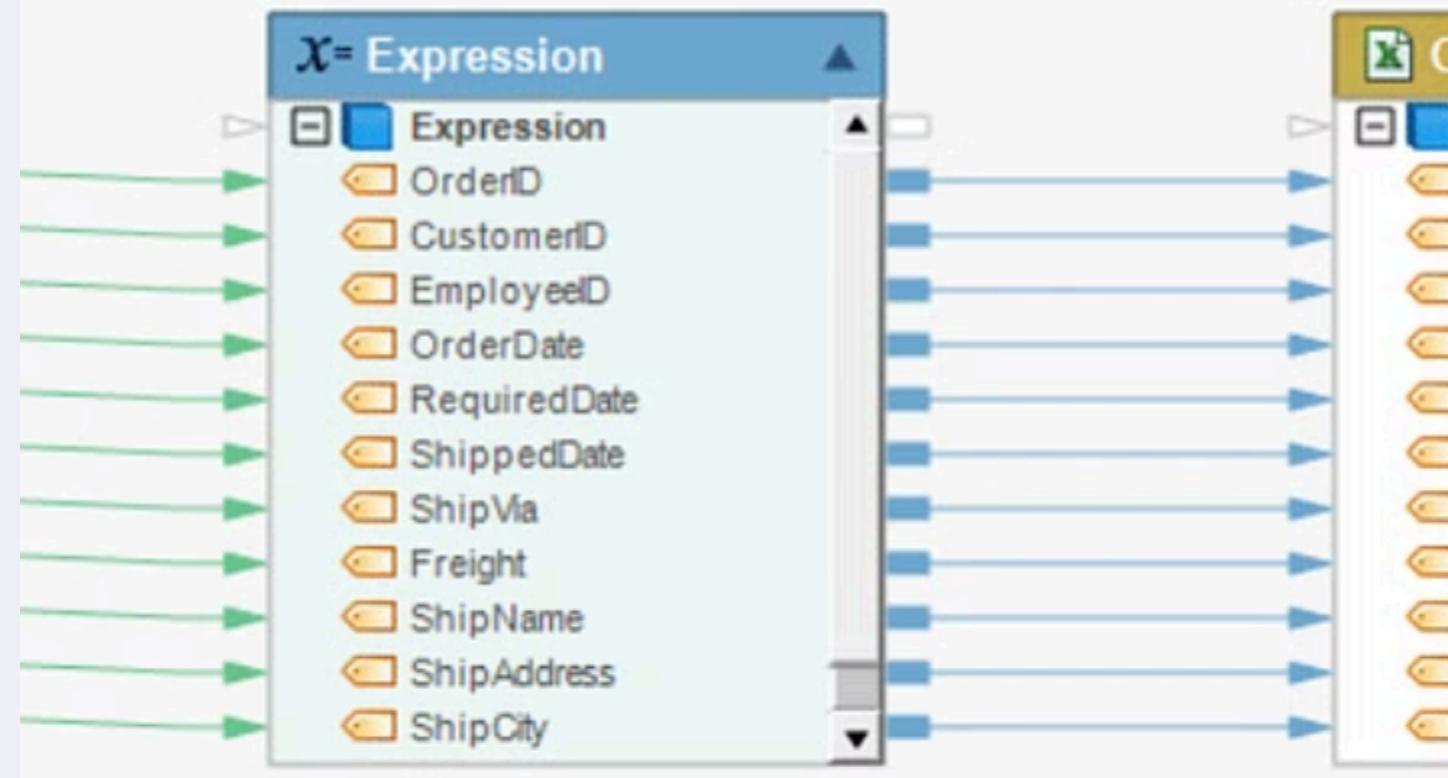
Type, Content Rating

3 Ensure Integrity

Maintain data accuracy

Data Type Conversion

Data type conversion is vital for calculations. 'Reviews', 'Installs', and 'Price' were converted to numeric types. Symbols like '+', ',', and '\$' were removed. This ensures consistency across all columns. Cleaned data now supports robust analysis.



Numeric Conversion

Reviews, Installs, Price



Symbol Removal

+, comma, dollar



Consistency

Across columns

Records With Errors 0. Duration 00:00:00.679.

EmployeeID	OrderDate	RequiredDate	ShippedDate
5	7/4/1996 12:00:00	8/1/1996 12:00:00	7/16/1996 12:00:00
6	7/5/1996 12:00:00	8/16/1996 12:00:00	7/10/1996 12:00:00
4	7/8/1996 12:00:00	8/5/1996 12:00:00	7/12/1996 12:00:00
3	7/8/1996 12:00:00	8/5/1996 12:00:00	7/15/1996 12:00:00
4	7/9/1996 12:00:00	8/6/1996 12:00:00	7/11/1996 12:00:00
3	7/10/1996 12:00:00	7/24/1996 12:00:00	7/16/1996 12:00:00
5	7/11/1996 12:00:00	8/8/1996 12:00:00	7/23/1996 12:00:00
9	7/12/1996 12:00:00	8/9/1996 12:00:00	7/15/1996 12:00:00
3	7/15/1996 12:00:00	8/12/1996 12:00:00	7/17/1996 12:00:00
4	7/16/1996 12:00:00	8/13/1996 12:00:00	7/22/1996 12:00:00
1	7/17/1996 12:00:00	8/14/1996 12:00:00	7/23/1996 12:00:00
4	7/18/1996 12:00:00	8/15/1996 12:00:00	7/25/1996 12:00:00

Creating Additional Columns

Creating new columns enhances analysis capabilities. We extracted the year from 'Last Updated' for trend analysis. 'Size' was categorized into Small, Medium, and Large. Reviews were categorized for grouping purposes. New columns enable deeper insights.

Year Extraction

From 'Last Updated'

Size Categorization

Small, Medium, Large

Review Categories

For grouping

Dashboard Objective & KPIs

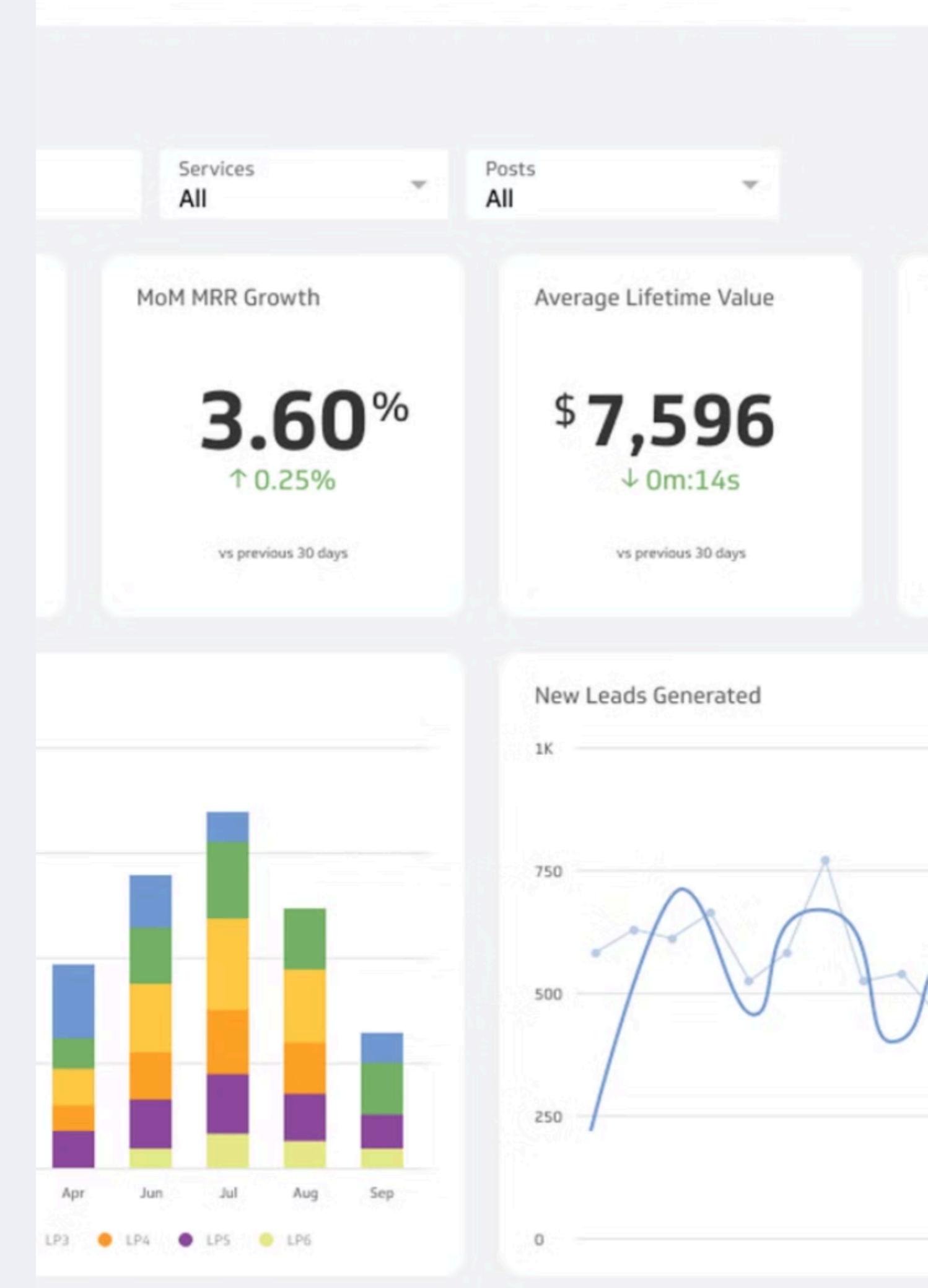
The objective is to understand app performance and user behavior. Key KPIs include Total Apps and Average Rating. Further KPIs are Total Installs and Distribution by Category. These metrics drive the dashboard's focus.

Total Apps

Average Rating

Total Installs

Distribution by Category



Visualizations Created

Various visualizations were created for the dashboard. These include bar charts and pie charts. Line charts and treemaps were also used. Filters for Category, Content Rating, and Last Update Year were added. This enables dynamic exploration of the data.

Bar Chart

Top 10 Installed Apps

1

Pie Chart

App distribution by Category

2

Treemap

Installs by Category

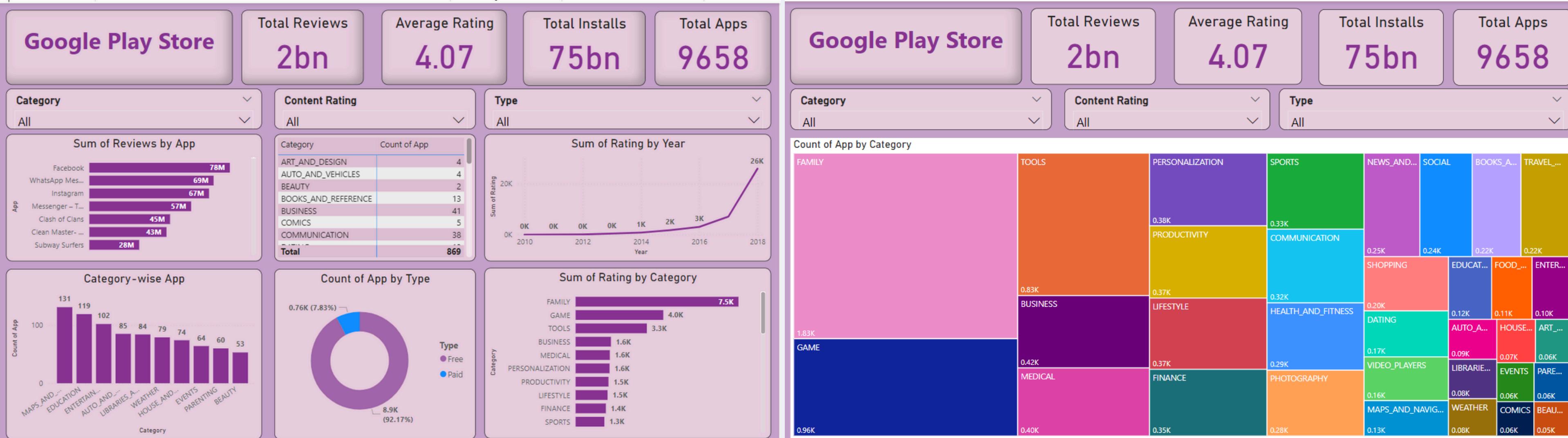
4

3

Line Chart

Ratings trend over years

Visualization



Insights from Dashboard

Most apps are free in the FAMILY or GAME categories. Apps with high installs generally have high average ratings. Most apps target all age groups (Everyone). Most updates happened in 2017 and 2018. The dashboard reveals key trends.

1 **Free FAMILY/GAME Apps**

2 **High Installs = High Ratings**

3 **Updates in 2017/2018**

Conclusion

Data cleaning was critical for accurate insights. The dashboard highlights trends in installs, ratings, and categories. Clean, interactive visuals aid in understanding Google Play Store dynamics. This comprehensive analysis delivers valuable perspectives.



Data Cleaning

Crucial for accuracy



Key Trends

Installs, Ratings,
Categories



Interactive Visuals

Easy understanding

