

Internship Project Report

1. Dataset Used

Primary Dataset

- **Name:** Google Play Store Apps Dataset
- **Files:**
 - Play Store Data.csv
 - User Reviews.csv

The same dataset was used for Python (Plotly) analysis and Google Looker Studio dashboards, in line with internship instructions (no external datasets).

2. Data Transformations Applied

Because the raw dataset contains **text-based and inconsistent formats**, several transformations were required before analysis.

2.1 Installs Normalization

- Original format: 1,000,000+
- Transformation:
 - Removed + and commas
 - Converted to numeric

2.2 App Size Conversion (Task-Critical)

- Original format:
 - 19M, 512k, Varies with device
- Transformation:
 - Converted all sizes into **megabytes (MB)**
 - Ignored non-numeric values

2.3 Date Handling (Last Updated)

- Original values internally stored as serial/date-time
- Issue:
 - Looker Studio does not consistently support date functions
- Solution:

- Month extraction performed using formatted datetime logic

2.4 Category Translation (Multi-language Requirement)

For visualization clarity, categories were translated **only at the display level**:

3. KPIs Measured

The following **Key Performance Indicators** were defined and used across dashboards:

Core KPIs

- **Average Rating**
- **Total Reviews**
- **Total Installs**
- **Average Installs**
- **Average Revenue**
- **Cumulative Installs (Over Time)**

Supporting Metrics

- App Size (MB)
- Android Version
- Content Rating