

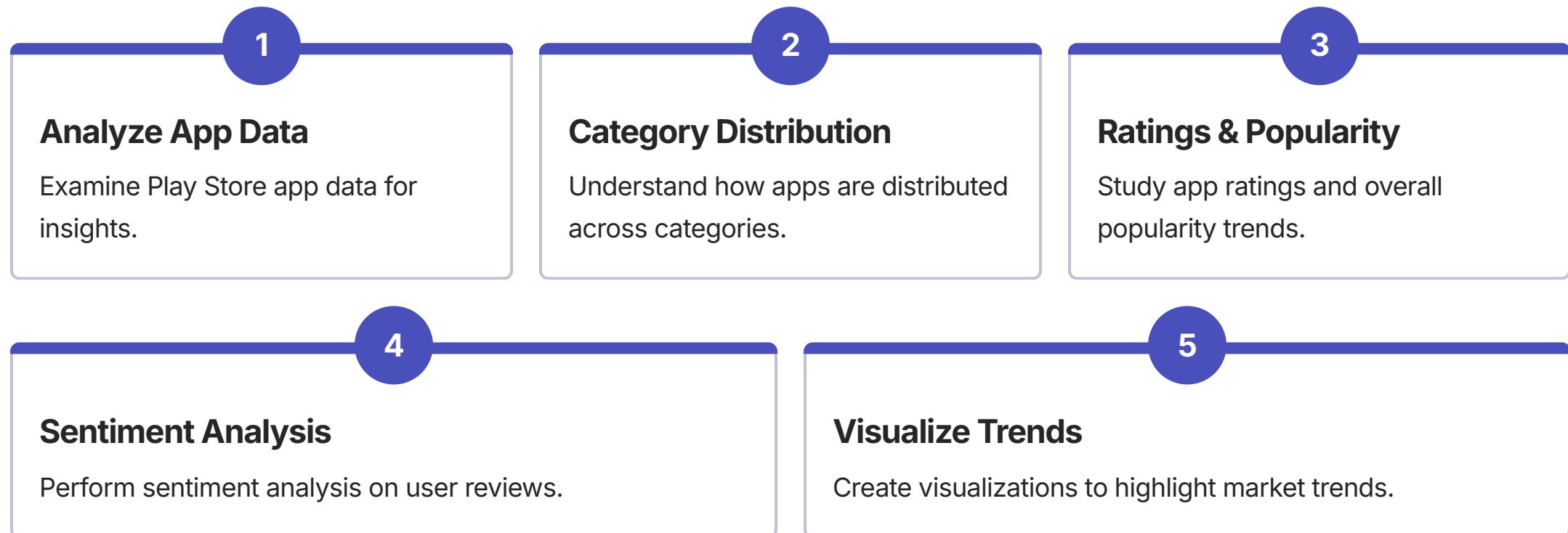
Unveiling the Android App Market

Darshana Chaudhari- Data Analytics Internship



Problem Statement & Objectives

Our goal was to analyze Google Play Store data to understand market dynamics.



Dataset Overview & Cleaning

10,841

Total Apps

Extensive dataset for analysis.

13

Features

Key columns like Category, Rating, Reviews, Size, Installs, Price.

An additional reviews dataset was utilized for deeper insights.

Data Cleaning Steps

- Removed duplicate apps.
- Handled missing ratings.
- Converted 'Installs' to numeric format.
- Cleaned 'Price' column for consistency.
- Corrected various data types.

Ratings & Popularity Insights



→ High Ratings

Majority of apps rated above 4.0, indicating overall user satisfaction.

→ Free App Dominance

Free applications hold a significant market share.

→ Top Installs

Social and Communication categories lead in terms of installations.



Category Exploration: Distribution & Trends

Most Apps in FAMILY

The 'FAMILY' category has the highest number of applications.

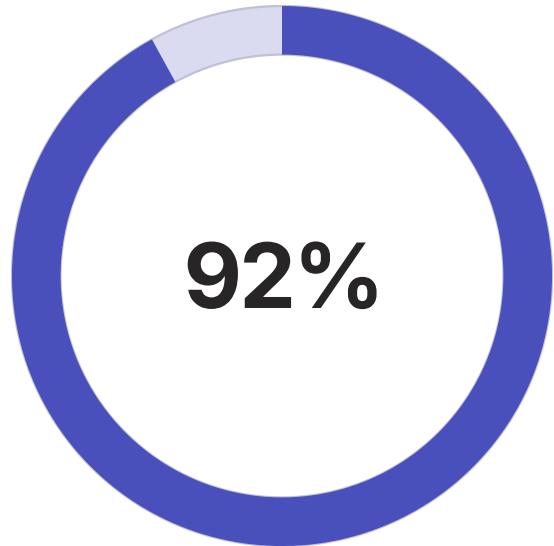
GAME & TOOLS Follow

Gaming and Utility tools are the next most populated categories.

Uneven Distribution

Some categories are underrepresented, indicating potential niche markets.

Pricing & Size Analysis



Apps are free, highlighting a strong freemium model.

Paid apps constitute a small percentage of the market, suggesting a preference for free content.

App sizes vary widely, with larger applications often found in the gaming sector.



Sentiment Analysis on Reviews



Reviews Dataset Analyzed

In-depth analysis of user comments and feedback.

Dominant Positive Sentiment

Overall sentiment in reviews leans heavily towards positive feedback.

Word Frequency Analysis

Identified key terms and phrases used by users.

NLP Techniques Utilized

Applied Natural Language Processing for context and meaning.

Key Visualization Techniques Used



Bar Charts

Illustrated category distribution effectively.



Scatter Plots

Analyzed relationships between ratings and reviews.



Heatmaps

Revealed correlation patterns in the data.



Pie Charts

Showcased the proportion of free versus paid applications.