Exploratory Analysis Plan

Research Question	EDA Approach	Expected Insights
Which app categories dominate downloads, and how has this changed over the past decade?	Group by `category × release_year`; compute total `installs`. Visualise trend using line or area plots.	Identify rising vs. declining app domains (e.g., Social Media vs. Utilities).
Do app ratings correlate strongly with install counts, or do marketing/network effects play a bigger role?	Correlation matrix (`rating`, `installs`, `reviews`); regression analysis of installs on ratings and reviews.	Distinguish whether installs are driven by satisfaction (ratings) or visibility (reviews).
Are free apps rated higher or lower than paid apps? Does monetisation affect user satisfaction?	Boxplots or violin plots of `rating` by `type`; T-test or Mann-Whitney test for mean difference.	Test if paid apps yield better quality or if free apps dominate ratings.
How do reviews reveal evolving concerns (privacy, ads, etc.)?	NLP on review text (topic modelling via LDA); time-series of topic frequency.	Extract themes from user feedback over time.
Which countries/regions show the fastest growth in app adoption, and are preferences culturally specific?	Group by `region × category`; compute CAGR (compound annual growth rate) of installs.	Identify regional app preferences (e.g., India → Social/Entertainment; Japan → Games).