**Project Summary**

In this project, we set out to help Google Play Store identify promising apps for promotion predicting app ratings using available metadata. We worked with the "googleplaystore.csv" dataset, which included information such as app category, number of reviews, installs, price, content rating, and more.

Key Steps

* **Data Cleaning:**  
  We removed rows with missing values, standardized data types (e.g., converting sizes to KB, parsing installs and prices), and filtered outliers in price, reviews, and installs.
* **Exploratory Data Analysis:**
* Paid apps tend to have slightly higher median ratings than free apps, but both types show a wide range of ratings.
* Most content ratings have similar distributions, with "Everyone" being the most common.
* Categories like EVENTS, EDUCATION, ART\_AND\_DESIGN, and BOOK\_AND\_REFERENCE had the highest average ratings.
* **Modelling:**  
  We trained Linear Regression, Random Forest, and Gradient Boosting models. All models achieved similar test R² scores (~0.13), indicating that the current features explain only a small portion of the variance in app ratings. Random Forest showed overfitting, performing on training data but not generalizing to new data.
* **Feature Importance:**  
  The most important predictors of app ratings were the number of reviews, installs, app size, and price.
* **Recommendations:**  
  Using our best model, we identified the top 10 apps predicted to have the highest user ratings. We also highlighted promising newcomers—apps with low install counts high predicted ratings—as strong candidates for promotion.

**Limitations & Next Steps**

* The dataset lacks potentially important features such as review text sentiment, update frequency, or user engagement metrics.
* The relationship between the provided features and app ratings is weak, limiting predictive power.
* Future work could include collecting richer data, trying natural language processing on reviews, or testing more advanced modelling techniques.

Business Recommendation

Promote apps with high predicted ratings, especially those with low install counts, to surface quality new apps and support promising developers.