**Group 1: Regular Data Science Questions**

1. Calculate the average rating of apps in the 'Social' category.

3.9

1. What is the total number of downloads for 'Fitness' category apps?

600,000

1. Find the most common version among apps in the 'Photography' category.

3.2.0

**Group 2: Multiple Step Hard Data Science Questions**

1. For each category, calculate the median number of downloads. Which category has the highest median?

Social

1. Identify the developer with the highest average app rating. Consider only developers with more than one app.

PhotoApps

1. Calculate the proportion of paid apps in each category. Which category has the highest proportion of paid apps?

Books & Reference

1. What is the average number of downloads for apps released in 2022?

2,014,706

1. Compare the average ratings of paid and free apps. Which category of app has a higher average rating?

Free

1. Identify the category with the highest increase in average ratings from version 1.0.0 to the latest version.

Music

**Group 3: Multistep Data Analysis and Machine Learning Questions**

1. Predict the category of an app based on its rating and number of downloads using a Naive Bayes classifier. What is the model's accuracy?

56.67%

1. Use a Linear Discriminant Analysis (LDA) model to classify apps into different categories based on their rating, downloads, and whether they are paid or free. What is the model's accuracy?

53.33%

1. Use cluster analysis to group apps based on their rating and downloads. How many clusters are optimal and what are their characteristics?

2 clusters

1. Apply a neural network model to classify apps into different categories based on their rating, downloads, and whether they are paid or free. What is the model's accuracy?

33.33%

1. Implement a random forest classifier to predict the app category based on rating, downloads, and version. What is the F1-score of the model?

73.87%

1. Apply a support vector machine (SVM) model to classify apps into different categories based on their rating and downloads. What is the model's recall score?

26.67%