**Level 1 (Regular Data Science Questions)**

1. Calculate the average price of toys in each category. What is the category with the highest average price?

Electronic

1. Determine the proportion of toys that are in stock for each material type. Which material has the highest in-stock proportion?

All material types have a 100% in-stock proportion

1. Calculate the range (maximum - minimum) of ratings for toys in each age group. What age group has the widest range of ratings?

Ages 6-12

**Level 2 (Multiple Step Hard Data Science Questions)**

1. For each category, find the average rating of toys and compare it with the overall average rating. Which category has the highest average rating above the overall average?

Electronics

1. Create a new column "PriceCategory" based on the price of toys (Low: < $10, Medium: $10-$20, High: > $20). What is the average rating for each price category?

High Price Category (> $20): Average Rating = 4.31; Low Price Category (< $10): Average Rating = 3.85; Medium Price Category ($10 - $20): Average Rating = 4.18

1. Calculate the price elasticity of demand by age group, assuming a linear relationship between price and the number of toys in stock (use a simple regression model). Which age group shows the highest price elasticity?

Ages 3-6

1. Identify the top 5 most common toy names and calculate the average price for these toys. What is the average price for these common toys?

Top 5 most common toy names: 'Toy Car', 'Toy Boat', 'Kitchen Playset', 'Origami Kit', 'Magic Set'; Average Price: Magic Set = $23.49, Origami Kit = $12.99, Toy Car = $12.99

1. Segment the dataset based on 'Age Group' (e.g., Toddlers, Children, Teens). For each age group, calculate the average 'Price' and identify the 'Category' with the highest average sales. What is the difference in average price between the age groups with the highest and lowest values?

$11.80

1. Within each 'Category' of toys, calculate the year-over-year growth rate in 'Sales'. Identify the category with the highest average growth rate over the last 3 years. Additionally, for this category, determine the average 'Rating' for the top 10% best-selling toys.

4.6 stars

**Level 3 (Multistep Data Analysis and Machine Learning Questions)**

1. Use a Naive Bayes classifier to predict the "InStock" status based on Price, Rating, and Age. Report the F1-score of the model.

0.0

1. Employ a Linear Regression model to predict the Rating of a toy based on its Price. What is the coefficient of determination (R²) for the model?

-1.5

1. Implement a K-Nearest Neighbors algorithm to classify toys into their respective Categories based on Price and Age. Report the model's mean squared error.
2. Use a Ridge Regression model to predict the Price of toys based on their Rating and Age. What is the root mean squared error (RMSE) of this model?
3. Develop a Logistic Regression model to classify toys into "Low" and "High" price categories (split at $15). What is the recall of the model for the "High" price category?
4. Apply a Lasso Regression model to predict the Rating of toys using Price, Age, and InStock status. What is the most influential feature in this model?