Initial Data Exploration Report

Objective:
Load and perform initial exploration on the CIFAR-100 dataset to gain insights into the data's structure, distribution, and potential preprocessing needs.

Data Loading and Structure:

1. **Data Download:**
   CIFAR-100 is available through popular deep learning libraries, such as TensorFlow and PyTorch.
2. **Dataset Structure:**

   - **Train/Test Split:** 50,000 images for training and 10,000 for testing.
   - **Image Size:** 32x32 pixels.
   - **Classes:** 100 classes, labeled by category.

Exploratory Analysis:

1. **Dataset Shape and Dimensions:**
   - **Training Set:** 50,000 images of shape (32, 32, 3).
   - **Test Set:** 10,000 images of shape (32, 32, 3).
2. **Class Distribution:**
   CIFAR-100 is balanced, with each class containing 600 images across train and test sets.

Observations:

- **Image Quality:** CIFAR-100 images are low resolution (32x32 pixels).
- **Class Diversity:** Contains a wide range of categories.
- **Potential Challenges:** Image size and dataset variability might require careful tuning of deep learning models to achieve accurate classification.

Conclusions and Next Steps:

- The dataset structure and distribution are well-understood.
- Proceeding to data preprocessing, where resizing, normalization, and potentially augmentation could enhance model performance.