CIFAR-10 Dataset Analysis Report - Task 2

Course: CSE449 | **Project:** Event Recognition via PageRank

Dataset Summary: The CIFAR-10 dataset contains 60,000 color images (50,000 training, 10,000 test) across 10 classes: airplane, automobile, bird, cat, deer, dog, frog, horse, ship, truck. These classes represent different event types: transportation (airplane, automobile, ship, truck), wildlife (bird, deer, frog), domestic animals (cat, dog, horse). Each image is 32×32 pixels with 3 RGB channels. Dataset is perfectly balanced with 6,000 images per class.

Basic Analysis

- **Data Points:** 60,000 total images **Range of values:** Pixel values 0-255 (uint8) **Mean RGB:** [125.31, 122.95, 113.87] **Median RGB:** [124.0, 121.0, 112.0] **Standard Deviation:** [63.02, 62.09, 66.70] **Mode:** Most frequent pixel values around [120, 115, 95]
- **Common images:** Ships and airplanes (bright backgrounds, consistent patterns) **Uncommon images:** Cats and dogs (varied poses, indoor/outdoor settings, high variance)
- **Data Quality:** No missing values, no corrupted images, consistent 32×32×3 dimensions across all images.

EDA Results

- **Class Distribution:** Perfectly balanced - each class has exactly 6,000 images (16.67%)
- **Brightness Analysis:** Ships highest (130.2), Frogs lowest (108.3). Animals generally darker than vehicles/aircraft.
- **Color Patterns:** Blue-dominant (ships, airplanes), Green-dominant (frogs, deer), Balanced RGB (cars, trucks)
- **Visualizations Generated:** 9 charts including bar chart, pie chart, histogram, sample images, box plots, color analysis, correlation heatmap, scatter plot, and summary statistics table.

Conclusion: CIFAR-10 dataset is suitable for PageRank-based event recognition. The balanced distribution, diverse visual characteristics, and clean data quality make it ideal for algorithm development. Color and brightness features effectively distinguish between event types. Dataset prepared for PageRank implementation with 1,500-image subset created for testing.