**Title- Exploratory Data Analysis (EDA) Report**

**Date:** 4/3/2025  
**Author(s):** Samuel Jones

**Introduction**

* **Dataset Name**: PlantGrowth
* **Source**: Built-in R dataset
* **Objective**: The purpose of this analysis is to explore the effect of different treatments on plant growth by analyzing the weight of plants across different treatment groups.

**2. Data Description**

* **Number of Observations**: 30
* **Number of Variables**: 2
* **Brief Description of Key Variables**:

| **Variable Name** | **Description** | **Data Type** |
| --- | --- | --- |
| weight | Weight of the plants | Numeric |
| group | Treatment group (ctrl, trt1, trt2) | Factor |

* **Missing Values**: There are no missing values in this dataset.
* **Data Cleaning Steps Taken**: No cleaning was necessary as the dataset is already clean.

**3. Data Transformation**

* **New Variables Created (mutate() used)**:
  + weight\_kg – Weight of the plants converted to kilograms.
  + weight\_squared – Square of the weight of the plants.
* **Grouping & Aggregation (group\_by() used)**:
  + **Grouped by**: Treatment group (group)
  + **Summary statistics computed**: Total weight and average weight for each group.

**4. Exploratory Data Analysis (EDA)**

**4.1 Summary Statistics**

* **Mean, Median, Standard Deviation for Key Variables**:
  + Mean weight: 5.073
  + Median weight: 5.155
  + Standard Deviation of weight: 0.701
* **Outliers detected?**: No significant outliers detected.

**4.2 Visualizations**

1. **Histogram**: Distribution of plant weight across all samples.
2. **Box plot**: Displays weight distribution by treatment groups.
3. **Scatter plot**: Relationship between scaled weight and original weight.
4. **Bar chart**: Mean weight comparison among groups.
5. **Density plot**: Kernel density estimates for weight distribution per group.
6. **Line plot**: Line graph showing changes in mean weight.
7. **Correlation heatmap**: Correlation matrix of numeric variables (e.g., weight and scaled weight).
8. **Violin Plot** – Visualizes weight distribution per group with density and box plot elements.
9. **Custom Visualization** – A dot plot to examine weight variation across groups.
10. **Facet Grid Plot** – Shows density distribution broken down by group in a grid layout.

**5. Findings & Insights**

* **Trends Observed**: The average weight of plants varies across different treatment groups, with some groups showing higher average weights than others.
* **Patterns & Anomalies**: No significant anomalies were detected in the data.
* **Surprising Observations**: The treatment groups have a noticeable impact on the plant weights, indicating the effectiveness of certain treatments.

**6. Conclusions & Recommendations**

* **Final Summary of Findings**: The analysis shows that different treatments have a significant effect on plant growth, with some treatments leading to higher average weights.
* **Possible Further Analysis**: Further analysis could include a more detailed examination of the variance within each treatment group and the application of statistical tests to determine the significance of the differences observed.
* **Potential Business/Research Recommendations**: The insights from this analysis can be used to optimize plant growth treatments in agricultural practices, potentially leading to improved crop yields.

**7. References**

* Built into the R Studio system