**Task Report**

**1. Preprocessing**

* There were **no null** values in the dataset.
* **Outlier Detection** and optimization using inter quartile technique with help of box plot.
* Tried approaches like **dropping** outliers , **log transformation** on outliers.

**2. Visualization**

* Plotted **Average Spectral Reflectance vs Wavelength**
* Plotted heatmap for Spectral reflectance of random 50 corn samples.
* Plotted box plots for all columns for outlier detection.

**3. Dimensionality Reduction**

* Applied **PCA** for dimensionality reduction.
* Got 4 principal components for 95 % variance.

**4. Feature Extraction**

* Since models were not performing well with principal component variables, I extracted features which were contributing these principal components.
* Putting principal component variables to models gave negative R2 score , RMSE and MAE.
* Tried diverse no .of features , finally model was working fine for 423 unique features which taken out by unique top 150 contributors of each principal component variable.

**5. Model Selection**

* **Artificial Neural Network:**

MAE: 3416.7812770462037

RMSE: 109932615.86233217

R² Score: 0.6067277515621837

* **RandomForest(with Log Transformation):**

MAE: 1.9759457541084504

RMSE: 2.665965428341797

R² Score: 0.12453103373042618

* **RandomForest(without Log Transformation):**

MAE: 3836.4763345870683

RMSE: 11095.237259573893

R² Score: 0.559607487910612

* **XGBoost(with Log Transformation):**

MAE: 1.9526275463211038

RMSE: 2.634929838060914

R² Score: 0.14479577030870883

**6. Improvements**

**Increasing the number of dataset should give better results.**