**Apple Quality Classification**

**Problem Statement**: This project aims to build a classification model to predict the quality of apples based on various features such as size, weight, sweetness, crunchiness, juiciness, ripeness, and acidity. The dataset used for this analysis contains 4000 entries with two classes of quality: 'good' and 'bad'. The data was loaded, explored, and pre-processed to ensure it was suitable for model training.

**Libraries Used:**

* Pandas: Data manipulation and analysis
* Matplotlib and seaborn: Data visualization
* Numpy: Mathematical operations
* Sklearn: Machine learning library
* Scipy: Scientific computing library
* Pygwalker: To analyse the data

**Data Collection**: I have collected this data from Kaggle containing 4000 datapoints with 8 features and a class label in the form of a csv file.

**Analysing the raw data**: In the process of analysing the data, I have observed that the data is almost balanced. And in this dataset it contains one null row in the data. And also the column called “Acidity” is in object datatype.

**Data Cleaning:** Handled missing values by dropping the last row and converting the 'Acidity' column to a float. Removed unnecessary columns like 'A\_ id' for model training.

**Exploratory Data Analysis (EDA):** Utilized the Python Graphic-Walker library to visualize the distribution of features after cleaning. Examined the distribution of each feature using probability plots. After this analysis, I observed that the data is following Gaussian distribution.

**Model Training and Evaluation**:

**Splitting Data:** Divided the dataset into training and testing sets using the **train\_ test\_ split** method.

**Gaussian Naive Bayes Model:** Created an instance of the Gaussian Naive Bayes classifier. Trained the model on the training data.

**Pipeline:** Here, I have created a pipeline for all this function to make things easy to transform and share the code to any where.

**Prediction and Evaluation:** Made predictions on the test set and calculated the accuracy of the model. Achieved an accuracy score of 75%.

**Conclusion:**After completing all the preprocessing and training i deployed my model into streamlit so that i can share with some other people.In conclusion, the project successfully explored and cleaned the apple quality dataset, performed exploratory data analysis, and built a Gaussian Naive Bayes classification model. The model achieved a 75% accuracy in predicting apple quality based on the provided features. Further improvements can be explored, such as trying different models, hyperparameter tuning, and feature engineering, to enhance the predictive performance.