Assignment: Predictive Maintenance Analysis using Pandas

You are a Data Scientist at an advanced manufacturing plant. The company has deployed sensors to monitor various machine parameters like temperature, speed, torque, and tool wear. The operations team wants to reduce unplanned downtimes by analyzing machine behavior and predicting failures in advance. Your role is to conduct data exploration and feature analysis using Pandas to extract actionable insights that can feed into future classification models.

Dataset Summary:

- **Type**: Synthetic real-world-like industrial data
- **Instances(Records)**: 10,000
- **Features**: Air temperature, Process temperature, Rotational speed, Torque, Tool wear, Machine failure, etc., [refer to CSV]
- **Domain**: Predictive Maintenance, Manufacturing Analytics

---

Questions:

1. **Display the first 10 rows of the dataset**
   Hint: Use `.head()`
2. **What are the unique values in the `Type` column? How many machines of each type exist?**
   Hint: Use `.unique()` and `.value_counts()`
3. **Select all records where `Tool wear` is greater than 50.**
   Hint: Filtering
4. **Rename the column `Torque [Nm]` to `Torque_Nm` and `Air temperature [K]` to `AirTemp_K`.**
   Hint: Use `.rename()`

5. **Check for missing values in each column and report the count.**
   Hint: Use `.isnull().sum()`

6. **Filter all rows where the `Rotational speed` is greater than 1600 and `Torque [Nm]` is less than 30.**

7. **Create a new column called `Temp_Diff` which is the difference between `Process temperature [K]` and `Air temperature [K]`.**

8. **Group the dataset by `Type` and compute the average `Tool wear` and `Torque [Nm]` for each type.**
   Hint: Use `.groupby()` and `.agg()`

9. **Sort the dataset in descending order of `Tool wear` and display the top 5 rows.**

10. **Create a new column `Failure_Flag` that has value 1 if any of the failure indicators (`TWF`, `HDF`, `PWF`, `OSF`, `RNF`) are 1, else 0.**
    Hint: Use row-wise `.sum(axis=1)` and compare

11. **For each `Product ID`, compute the total number of failures (`Machine failure`) it encountered. Display only those products with more than one failure.**

12. **Use multi-level groupby on `Type` and `Machine failure` and compute average `Torque [Nm]` and `Tool wear [min]`.**

13. **Replace any zero `Rotational speed` values with the median of that column.**
    Hint: Data imputation

14. **Find out which machine type had the highest average difference between process and air temperature (`Temp_Diff`).**

15. **Filter rows where `Tool wear` is in the top 10% of the dataset. Then, group by `Type` and calculate the average failure rate.**
    Hint: Use `.quantile()` for threshold

Variable Information:

1. ***UID*** *– Unique identifier for each record (1 to 10,000).*
2. ***Product ID*** *– Encodes product quality as L (Low), M (Medium), or H (High) with a unique serial.*
3. ***Air temperature [K]*** *– Ambient air temperature around 300K with slight random variation.*
4. ***Process temperature [K]*** *– Process temperature, typically ~10K higher than air temperature.*
5. ***Rotational speed [rpm]*** *– Machine's rotational speed derived from power with noise added.*
6. ***Torque [Nm]*** *– Force applied in rotation, normally around 40 Nm, no negative values.*
7. ***Tool wear [min]*** *– Minutes of tool usage, higher for better product quality.*
8. ***Machine failure*** *– Indicates if any type of machine failure occurred (1 = failure).*
9. ***TWF*** *– Tool wear failure: happens when tool wear exceeds a threshold.*
10. ***HDF*** *– Heat dissipation failure: occurs when temp difference < 8.6K & speed < 1380 rpm.*
11. ***PWF*** *– Power failure: power outside 3500–9000 W range causes failure.*
12. ***OSF*** *– Overstrain failure: excessive tool wear × torque for each product type.*
13. ***RNF*** *– Random failure: small chance (0.1%) of failure unrelated to parameters.*