Instructions for the Exam:

1. This is an open book, take home exam
2. This is a group activity. Each group has 4 members who have to work together to solve the problem. The group allocations will be announced in class
3. Please spend sufficient amount of time in solving the problem instead of a last-minute rush.
4. Final day would be a presentation by the team followed by viva for individual team members and the conceptual understanding of each student is thoroughly tested. So, prepare accordingly.

Problem Statement:

There are two variants of the same product that are being manufactured by a company. The characteristics of the products are given in the two data files, each for one variant respectively. The attributes of the variants are masked (this is not uncommon when the client due to certain restrictions may not want to give away the details of it, he would mask the attributes and in some cases the values as well), so by looking at the names of the columns you may not be able to identify what it means. You need to analyse this data and come up with insights based on your initial exploratory data analysis and visualizations

Here are some of the questions to guide you to solve. ***This is not limited to the hints but you need to come up with additional analysis as well.***

Data Reading and Basic Descriptions

1. Read the two datasets
2. You need to figure out a way to combine both the datasets. After combining, you should also know which records belong to variant A and which to B

   Hint: you may create a new column with variant names
3. Get the dimensions of the data and the data types. Do you find the need to change the data type of any attribute?
4. Get the summary statistics of the data
   a. Are there any missing values, any imputation needed etc
   b. Are there any outliers or extreme values (you can compute z scores to identify this). How would you deal with this given the size of the data

Data Exploration and Visualizations and Statistical Analysis

5. Observe the distribution of values of certain attributes for each variant and check if the distributions are different

   Hint: You can plot histograms of a certain attribute for two variants and check it. Also you can perform statistical tests.

6. Does the PH values of the variants significantly different from one another? Justify

   Hint: Frame the hypothesis and use appropriate statistical test

7. How is Q distributed over the product variants? What are your observations.
    a. Does a variant of one product has higher values of Q compared to the other? How would you answer this question?
8. Do you find any correlation between PH and AF. Perform both qualitative and quantitative analysis.

    Hint: Compute the correlation value and draw an appropriate plot
9. Which of the variables have negative correlation among themselves?