

Questions Project 3 – Rahul Gupta

1. Which Domain is this Project in? Industrial or Manufacture

2. How would you deal with duplicate data points in an SQL query?

Using **UNIQUE** and **DISTINCT**

3. What is NumPy?

NumPy is used for efficient processing of arrays of numbers, and pandas is great for stats, which are the bread and butter of data science work

4. Can PANDAS be used for machine learning work?

Yes, Pandas is also good for preparing data for machine learning work.

5. Is it a GITHUB link for this project?

https://github.com/rahulgupta271/DSC680_Project_3_Enterprise_Asset_Maintenance

6. What Native Data Structures Can You Name in Python?

Common native data structures in Python are as follows:

Dictionaries

Lists

Sets

Strings

Tuples

7. What Python Modules are you using for this project?

- pandas
- numpy
- sklearn
- matplotlib
- seaborn
- joblib
- os

8. What is the business goal for this project?

Goal for this project is to understand the affecting factor which occurs the Maintenance cost and predicts the future cost.

9. What is model tuning?

Tuning is the method of optimizing the performance of a model without overfitting or generating a deviation that is too large. This is done in machine learning by choosing suitable "hyperparameters." Hyperparameters can be viewed as the "dials" or "knobs" of a model of machine learning.

10. What is the data availability checks?

The availability of data is about the timeliness and efficiency of data access and use. . Availability has to do with knowledge availability and longevity. Knowledge with questions about poor availability could be considered additional rather than essential.