**Data Engineering Exercise**

1. Write a python script to simulate the following three Tables:

* Table 1: High frequency timeseries accelerometer (3-axis) and gyroscope (3-axis) data (50 Hz, duration of 5 min) for 200 subjects
* Table 2: Demographic information for 200 subjects (Any 5 demographic variables)
* Table 3: Programming information including program name, amplitude, frequency and pulse width for 200 subjects

Use random variable generators to create above tables and make sure the range of values are reasonable.

1. Write the three Tables (from step 1) to Azure SQL database (Secured with firewall)
2. Create an Azure Blob Storage and integrate it with VNet.
3. Write an Event trigger function app (both python and .Net) to parse the Json files (attached example file) from the Blob Storage (step 3) and write it to the Azure SQL database.
4. The attached csv file contains IOT data from several sensors. “SubjectID” column is the unique identifier for each subject.

Write a python script to clean the data (missing and duplicate values), find the most important sensor and model the output variable. Try several techniques to handle the missing data and propose the optimized method.

You could either create a test environment in Microsoft Azure for this exercise and share the credentials or prepare a presentation with detailed illustrations and explanations for each of above items. Please provide all the source codes.