Q2. Which is an interesting data set you discovered recently? Why is it your favourite? No datasets on Kaggle, please.

Recently, I worked on Capstone Project in which I cluster the areas based on similarity of the different venues present in the city. For that what I did was to collect the pincode, latitude and longitude data from data.gov.in and then passing the latitude and longitude data to the foursquare API to get the different venues in 2KM range of the given location. Now the dataset is collected and then I applied the k-means clustering to cluster different venues of the city. This was the first dataset that I made by myself so that's why it was my favourite.

Q3. Why do we need a database? We can store everything in a file, no?

If we store everything in a file then there are very chances of data redundancy where as in database management system there is no redundant data.

If the data is lost there is no recovery mechanism in file system but database management system also provides us with recovery mechanism.

There is no efficient query processing in file system but dbms provide us with efficient query processing.

In database the data is stored in different tables and these tables are logically linked whereas in file system everything is stored in a single file, so for some records there is so much data suplicity. In DBMS we have a concept of normalization which helps split data into different logically linked tables eliminating data redundancy.

Q4. How well versed are you on the Unix command line?

During 1st year of my B.Tech I learnt C language from Coursera Course and in that course I need to submit my assignments through their portal which was entirely Unix based ie. We need to write unix commands to access the assignments and submit solutions. So, I have a clear understanding of the Unix command line.