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

→

I did a Football analysis project in python. In this project I used two datasets. One of them had information about team scores, referee, stadium where the match was played, etc. whereas the other had information about teams starting 11 players, substitution players, at what time the player passed the ball, to whom did he pass and much more.
It is my favorite because I love football and deriving insights from those datasets was pretty exciting.

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

→

Database approach is used in large systems which interrelate many files.
Database provides a crash recovery mechanism.
The file system doesn't have a crash mechanism, i.e., if the system crashes while entering some data, then the content of the file will be lost.

Changes are often a necessity to the content of the data stored in any system, and these changes are more easily done with a database approach.
A database can be easily queried for data. So it is easy to derive insights from a database having large amounts of data.

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

→

On a scale of 1-10, I rate myself at 5.
I know the basics of unix command line only but I’m willing to learn more of it.