Yelp dataset

* Using only reviews data, build a simple recommendation system first
  + Matrix factorization – derive embedding vectors for users and restaurants
* Next stage ideas
  + Clustering food preferences.
    - E.g. people who like Indian food have taste preferences similar to people who like Sri Lankan food
    - …..and maybe interesting insights like “people who prefer Swedish food have tastes very different from all other preference-based groups of people”

#####

**Notes on EMR setup**

AWS blog on EMR setup and a few other useful things:

<https://aws.amazon.com/blogs/big-data/building-a-recommendation-engine-with-spark-ml-on-amazon-emr-using-zeppelin/>

Adding new users to your AWS instance/EMR:

<https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/managing-users.html>

Make new user sudo user (increase permissions)

<https://www.helicaltech.com/create-multiple-sudo-users-to-ec2-amazon-linux/>

* Edit visudo file

<https://www.garron.me/en/linux/visudo-command-sudoers-file-sudo-default-editor.html>

* Note: the new user can now do ‘sudo su’ and assume the role of root user. This allows