

Tutorials 1

1. Check or install required Python packages. Anaconda is assumed.
2. Check your access to MySQL database
3. Download MovieLens datasets. Explore the data files and documentation
4. Import MovieLens datasets as Pandas dataframes. Review and augment the dataframes with column names, where needed.
5. Export the dataframes into MySQL database. Choose table names wisely. Check the database with some queries.
6. Now that we have data stored in place, import the old MovieLens user-item ratings. Which is the appropriate table?
7. User-item table is a dense COO (coordinate) representation of a sparse user-item matrix. Transform it into a proper user-item table that it represents! You can store it in the database if you wish.
8. Read data from Alice.xlsx (introductory example from lectures) into a dataframe.
9. Implement Pearson correlation and Adjusted cosine similarity measures.
10. Calculate user-user and item-item similarities.
11. Create a simple memory-based (nearest neighbor) recommender. You can use scikit-learn's NearestNeighbor.
12. Study sparse Pandas DataFrames and sparse scipy matrices (COO, CSR). Experiment with conversions sparse \leftrightarrow dense.