

Project - Due by December 8

Choose groups of one, two, or three people.

Find and procure a substantial (at least thousands of records, but tens-to-hundreds of thousands would be best) transactional dataset. The dataset can come from any mode of activity, but it should center around some form of fine-grained business transactions, rather than summary data, user profiles, or strictly raw content such as text or images. Describe the data, its source, and the context in which it was generated. Model the data using a relational model; implement and populate that model, and demonstrate how it may be queried in this form (note: this is assignment #5). Transform it into a dimensional model suitable for analysis; implement and populate that model, and demonstrate how that model may be queried (note: this is assignment #7).

Prepare a notebook describing your entire process from selection, through both models, and exploring the transformed data, providing some descriptive statistics and visualizations. Describe modeling decisions you faced and any interesting patterns or anomalies you find in the data. Use the tools you prefer, or with which you wish to gain experience. A five-minute presentation summarizing your work for class and a 15-25 page notebook writeup from each group are due in December; turn these in through Blackboard.

Data sources:

- Professor Kanungo's Programming course syllabus suggests several public data sources
- Follow up with Shmuel Ben-Gad (Gelman Library) regarding subscribed sources available only through GWU - many excellent Business and Finance transaction sets are available