PREDICT 420

Individual Assignment 2:
Accessing and Manipulating Relational Data

The business analysts are working on a model for customer lifetime value (CLV) analysis. They are looking at total revenues and costs associated with each customer. This initial request is for data for a prototype CLV model. The data we need are on the PostgreSQL database server, and you can access them by using your NetID and password. Hook up via the VPN and SSCC to maintain security.

Analysts need data from the **mail**, and **customer** tables under the **pilot** schema for the **xyz** database. To work on the prototype for the CLV model, analysts would like a full set of records for a sample of 100 customers—you can take the first 100 customers in the **customer** table. Any joins can be accomplished using the **acctno** columns of the tables. There should be one record per account number in the **customer** table. (There can be multiple records for each customer in the **item** table, which represents sales transactions, but we do not need to work with the item table for the CLV prototype.) What the analysts need is a single record for each customer that has the following columns/fields from the **customer** table: **acctno**, **ltd_sales**, **ltd_transactions**, **ytd_sales_2009**, and **ytd_transactions_2009**. That single customer record should also have all the fields from the **mail** table (one field from each of the sixteen mailing dates). But if there is no record in the **mail** table, we still need to get the customer record—that would be a customer who placed orders without getting any promotional mailings in 2009. You can use SQL syntax under psql to join the database tables. After you have the merged data, put them in a comma-delimited text file, something our business analysts can read into their CLV programs in Python, R, and Excel.

After you get the data on the SSCC, you can FTP the comma-delimited text file down to your personal computer before sending the work off to me. We do not care how you do the file transfer from the server to your personal computer—direct sftp may be the easiest, but some like Filezilla.

Deliverable: Submit the CSV file on canvas.