Lab 3: More SQL --
The Zoning Variance Database & Categorization via Lookup Tables

(Same as Problem Set 2)

This Lab exercise #3 begins in lab with lecture #6 and due back before lecture #9.

Introduction

This lab has two main purposes: (1) to acquaint you with a **Zoning Variance Database**[*]recording the characteristics of some 1800 zoning variances filed during the mid 1980s building boom in Boston, and (2) to learn how SQL's 'group by' capabilities can help you interpret and analyze large, shared 'read-only' datasets by creating local lookup tables in which you can accumulate local knowledge.

The zoning variance database* contains information collected by Clark Broida (MCP '87) about all Boston zoning variance applications from early 1984 through mid 1987. Zoning variances request permission for landuses that deviate from the established zoning and landuse regulations. Much of the development and rehabilitation that occurred in Boston during the 1980s required a zoning variance. Clark encoded these data while working as an MCP intern with the Boston Redevelopment Authority (BRA) in order to help understand typical zoning variance request patterns and to inform zoning changes. His MCP thesis (C. Broida, June, 1987) analyzed these data and the zoning review process. In a few weeks, when we start linking GIS software to relational databases we will analyze and map these zoning variance data.

We will also use the zoning variance database to study a more sophisticated use of SQL's 'group by' capabilities. By creating local 'lookup' tables, we can categorize and reinterpret large, shared, 'read-only' datasets without needing 'write' access to the detailed, 'official' datasets.

A set of class notes relevant to this exercise and related lectures and homework questions are available online. They are grouped under a 'ZONING Database' section of the SQL Notes* and include:

Zoning Variances*	Schema of ZONING table (and listing of related lookup tables)

---

[*] Kindly refer to the Labs section

[1980 Census data (by Boston NSA)](#)*	Schema of 1980 Boston Census data (and related lookup tables)
[Schema of Decision, Use, NSA, Neighbrhd Lookup Tables](#)*	Schema of Lookup tables (second half of Census data web page)
[Sub-Neighborhood lookup table](#)*	The NSA and NEIGHBRHD tables (bottom of Zoning Variance web page)
[SQL examples using zoning variances](#)*	Annotated SQL queries of ZONING table
[Grouping zoning applicants via 'lookup' tables](#)*	Annotated SQL queries illustrating use of lookup tables to categorize ownership of properties seeking zoning variances
[Zoning Variance Database Evolution Chart](#)*	Stages of evolution of the ZONING variance database

PART 1: SQL Queries Exploring the Zoning Variance database

Logon to SQL Plus with your own account.

Review the [sample SQL zoning queries](#)* in the class notes. Be sure that you understand the encoding of zoning variance information -- especially existing and proposed landuse, the handling of missing values, how particular code violations are encoded, and how the Board recommendations and decisions are encoded. Review -- and rerun -- the sample queries to get a feel for the data -- which types of code violations were typical, which ones were approved, how did they spread across neighborhoods, etc.

In constructing your queries, pay attention to how you handle 'missing values'. When computing the fraction of variances approved, for example, be careful not to count variances for which the outcome is not known. *For all questions, show your SQL statements and their resulting output.*

Question I-1 (10 points). A few of the sample queries use the AVG functions (and a little arithmetic) to compute the percentage of variances that met various conditions. Which are the most common yard violations: front, side, rear yard or setback violations? What fraction of all zoning variances (with known sideyard conditions) involved sideyard violations?

Question I-2 (10 points): List the case number, sub-neighborhood, existzonin, estimated cost, board recommendation (brarecom), board decision, offstreet parking indicator (#101), floor-area-ratio indicator (far151), and sideyard indicator (syard191) for all variances proposing to convert vacant land to housing in CHARLESTOWN (this is a neighborhood, *not* a sub-neighborhood). Use the column formatting commands to adjust

---

* Kindly refer to the Labs section

column widths so the queries print on an 80-column page. Use the name 'CHARLESTOWN' in your query, not the numeric code that corresponds to this neighborhood. Sort by the case numbers in ascending order.

Question I-3 (10 points): Fix the following SQL query so that it counts the proposed uses of zoning variances where the existing use is vacant land (code 10).

```
SELECT PRPSEDUSE, LANDUSE, COUNT(*)
  FROM ZONING Z, USE U
 WHERE Z.PRPSEDUSE = U.USE_CODE
 GROUP BY PRPSEDUSE, LANDUSE
HAVING EXISTUSE = 10;
```

Question I-4 (10 points): Write a query that returns the name of the neighborhood (*neighbrhd*) along with the count of cases that propose to change the land use from vacant land to commercial. Sort by the count of cases in descending order, then the neighborhood names in ascending order. Use the names of the land uses ('COMMERCIAL', 'VACANT LAND') rather than numeric codes in your query.

Question I-4extra: Optional (just for fun) (3 extra credit points): Repeat the query above, only show **all** the neighborhoods, regardless of whether they had a qualifying case or not (i.e., the count should be zero for any neighborhoods that did not show up in the last query). *Note: This query involves multiple SQL queries or a complex SQL query using techniques beyond what we have demonstrated in class; that's why it's an extra credit question. Don't spend a lot of time on it -- just understand why it isn't so easy to get the zero-case rows added to the table in one easy step.*

Question I-5 (15 points): Next, we want to compute the percentage of variances, by neighborhood, that proposed to convert vacant lantdto housing. This sounds easy, but isn't because it's hard to compute the numerator and denominator of this fraction in the same query. It is less elegant but more easily understood to do this in three steps. First, generate a view for the denominator -- i.e., a table that counts (by zoning.neighbrhd) all zoning variances (with known existing/proposed use).

Next, compute the numerator -- i.e., a table showing the counts (by neighborhood) of those vacant-to-housing variances.

Now combine these two temporary tables into a third table showing the total and filtered counts (for each neighborhood) in side-by-side columns, along with their ratio (the desired fraction, expressed as a percent).

Include in your lab assignment all three SQL statements and the final result showing the total and filtered counts (and percents). Sort your results by the neighborhoods *by name.* Include both the name and number of the neighborhood in your results. (It is possible but hard to use SQL to generate such a table all in one step. We'll show you a way later on. Think about why this is the case.)

Question I-5extra: Optional (just for fun) (2 extra credit points): Here's another question involving percentages. Compute by neighborhood (***neighbrhd***), the percent of variances for which the existing land use is 'housing' *and* the proposed use is 'housing'. Include the neighborhood name (via the lookup table) in your results, and sort them by the neighborhood name.

Question I-6 (15 points): Now let's look at the percentage of variances (with a known outcome) that were approved. For all variances with a known BRARECOM and known BOARDDECIS, determine the number of variances and the percentage approved for each neighborhood (`neighbrhd`). Consider 'approved' to mean codes 1 and 2 (i.e., approved as submitted or approved with provisions).

Question I-6extra: Optional (just for fun) (2 extra credit points): Use the 1980 census table (aggregated by neighborhood statistical area, nsa) to determine which of the NSAs are above average in income. Which sub-neighborhoods (the 64 NSAa) had above average incomes? (Think about and explain how you developed the estimate of the 'average' income that you choose to use here. This is not a trick question with a precise answer. There are several plausible answers and you should think about which one you choose.) Now determine the fraction of housing-to-housing variances (with a known decision) that were approved for these two groups of sub-neighborhoods: (a) those with above average income, and (b) those with below average income.

PART 2: Using Lookup Tables to Update and Categorize Read-Only Databases

In this part, you will be categorizing the zoning variances based on the type of land ownership of the parcels. We began discussing good and bad ways of doing this in class, and we have extensive online notes about the concepts and queries needed for the suggested strategy: "Grouping Zoning Applicants via Lookup Tables"[*]. Be sure you understand the concept behind using lookup tables and 'group by' queries to create and use your customized categorizations of variance without ever needing to change anything in the 'official' zoning table. Please also note that the class reading that develops these ideas further using, as a context, the problem of correcting 'spelling errors' in owner names recorded in the official Boston parcel database: *Ferreira, Joseph Jr., "Information Technologies that Change Relationships between Low-Income Communities and the Public and Non-profit Agencies that Serve Them,"* Chapter 7 in MIT Press Book from Spring 1996 DUSP colloquium on High Technology and Low-Income Communities. In addition, the PowerPoint slides that were used to discuss this paper[*] are also available online.

Question II-1 (15 points): Using the lookup table technique, develop a new table that categorizes property owners in the zoning database into groups called 'individual', 'public', 'corporate', 'other', or 'unknown'. Provide an annotated list of all the SQL

---

[*] Kindly refer to the Labs section

statements (including 'update' commands) that you use to generate this table. For example, one such 'update' command from the notes is:

```
UPDATE apptype
   SET newapp = 'BOSTON, PFD'
 WHERE applicant LIKE '%FACILIT%';
```

Don't spend a lot of time building 'updates' to catch every last corporation or trust. Once you've developed half a dozen update 'rules' (in addition to those in the notes), you should stop and focus your time on the rest of the questions below that use these results.

Question II-2 (15 points): For each of your 5 ownership categories in III-5 above, prepare a summary table showing: the total number of variances with a known proposed use, the total number of variances with 'housing' as the proposed use, and the percentage of variances that involved a proposed use of housing. Sort by the name of your categories. Briefly indicate whether the results suggest a pattern that might be worth looking into.

Question II-2extra (2 extra credit points): *If you have extra time, try using 'group by' commands as illustrated in latter sections of the "Grouping Zoning Applicants via Lookup Tables"[*] notes to produce frequency distributions of a few other zoning violations* -- broken down by the ownership categories you create in above. See if you can find categories of land ownership for which the rate of zoning variance approval is significantly different.

What to Turnin

We only need your answers (including SQL statements) in digital format. The problem set is due before the start of **lecture #9.**

---

[*] Kindly refer back to the Labs section