



TO: Electric Utility Commission

FROM: Jawana Gutierrez, VP of Customer Care

DATE: March 18, 2011

SUBJECT: Reporting Customer Information by Zip+4 Codes

The following memo contains the results from the analysis of Austin Energy customers as it relates to aggregating consumption information by zip +4 postal codes. This memo also provides a recommended strategy and process for releasing aggregated consumption data to the public while minimizing the risk of disclosing account specific information.

Background

Certain privacy laws that require stringent security around customer account information govern Austin Energy. In order to mitigate risk Austin Energy has developed policies that prohibit the sharing of customer specific data to the general public without the customer's consent. With these constraints in mind Austin Energy has considered providing consumption data to the public for research purposes at the aggregate level only. One suggested method of aggregation is to report consumption data by zip +4 postal code.

Currently customer consumption information is stored in the eCIS Billing system. The system configuration contains premise address information with five digit zip codes. Currently there is no provision for zip +4 information in eCIS. Due to the fact that the current system will be replaced this fiscal year, it would not be prudent to modify the system to accomodate zip +4 codes at this time.

It was noted that many customers receive their utility bill with the zip +4 code on the envelope. In order to take advantage of lower postal rates, our third party bill print vendor uses postal software to append the +4 codes to customer mailing addresses prior to distribution. Austin Energy has acquired similar software for the purpose of segmenting premise addresses by zip +4 codes for this analysis.

Zip Code Analysis

Before performing an analysis of the zip +4 codes, it was important to understand the basis upon which zip +4 codes are established. The United States Postal Service reports that in 1983 it adopted the zip +4 code to further streamline postal carrier sorting processes. While the five digit zip code identified an area of the country and delivery office to which mail is directed, the sixth and seventh numbers denote a delivery sector, which may be several blocks, a group of streets, a group of Post Office boxes, several office buildings, a single high-rise building, a large apartment building, or a small geographic area. The last two numbers denote a delivery segment, which might be one floor of an office building, one side of a street, specific departments in a firm, or specific Post Office boxes. With this understanding the analysis of customer data was performed and the following assumptions were made:

- Zip +4 codes associated with PO boxes were removed from the analysis, and
- Only active premise addresses were included.

The analysis found that there are 52 five-digit zip codes with active premise addresses. For five digit zip codes there are an average of 8417 premises per zip code cell. In contrast, there are 71,523 zip +4 codes with active premises and an average of 6 accounts per code. There are 11,192 (16%) premises with unique zip +4 codes and 44,819 (63%) zip +4 codes contain 5 or less customers. The chart on page 3 graphs the frequency of zip +4 codes by number of customers.

With almost 90% of the zip +4 codes containing 10 or fewer premises, a second analysis was performed using only the first two digits of the zip +4 code. This analysis yielded a total of 1,738 zip +2 codes with an average of 252 premises per zip +2 code. Table 1 shows comparative data for the five-digit zip code, zip +4, and zip +2 codes.

TABLE 1: Summary of Zip Code Analysis

	5 Digit Zip Code	Zip +2 Code	Zip +4 Code
Total Number of Codes	52	1738	71,523
Average Num. of Premises per Cell	8417	252	6
Number of Cells with 1 Premise	0	31	11,192
Percent of Cells with 1 Premise	0%	2%	16%
Number of Cells with 3 or Fewer	1	69	29,255
Percent of Cells with 3 or Fewer	2%	4%	41%
Number of Cells with 5 or Fewer	2	97	44,819
Percent of Cells with 5 or Fewer	4%	6%	63%
Number of Cells with 10 or Less	3	144	63,383
Percent of Cells with 10 or Fewer	6%	8%	87%

CHART 1: Frequency Chart for Zip +4 Codes

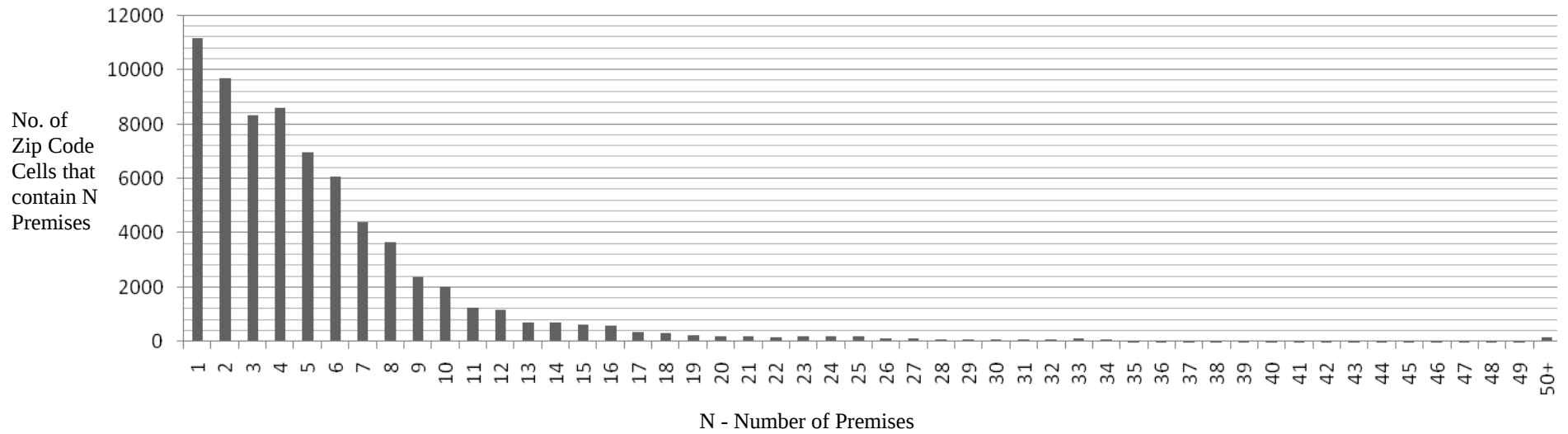
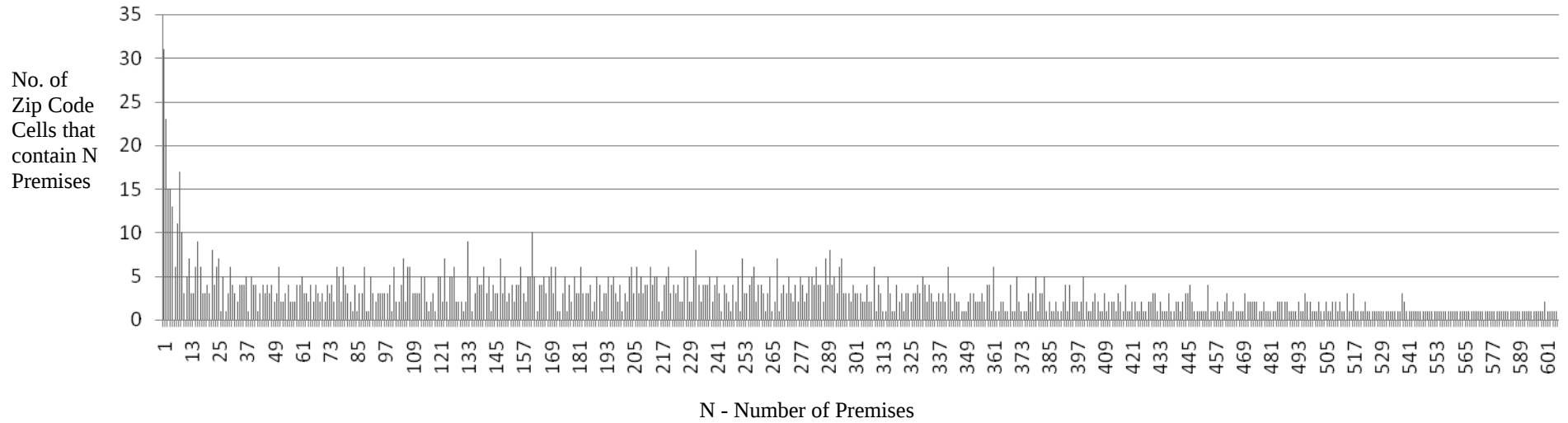


CHART 2: Frequency Chart for Zip +2 Codes



Recommended Minimum Number of Customers

Austin Energy reviewed the practices of some of the larger Federal Statistical Agencies. These agencies conduct a large number of either household or establishment surveys and have well-established policies and review bodies focused on disclosure issues and protocols.

Statistical methods used to limit disclosure vary depending on the program. In the case of tabular data, the most commonly used procedure has two steps – the threshold rule, followed by a concentration rule. Under the threshold rule, a cell in a table of frequencies is defined to be sensitive if the number of respondents is less than some specified number. Some agencies require at least 5 respondents in a cell, others require 3. The threshold rule may be applied to an entire table (e.g. a minimum size may be needed to publish values in all cells of a table); therefore, the agency may restructure tables and combine categories or use cell suppression, random rounding, or controlled rounding to meet the establish threshold.

The Bureau of Labor Statistics (BLS) programs and the US Census Bureau (Census) use what is called the p percent rule or the (n, k) rule to assess concentration depending upon program. The (n, k) rule is described as follows:

“Regardless of the number of respondents in a cell, if a small number (n or fewer) of these respondents contribute a large percentage (k percent or more) of the total cell value, then the so-called **n respondent, k percent rule** of cell dominance defines this cell as sensitive.”

The value of the parameters used for thresholds and various concentration rules used by the BLS is not released to the public. Establishment of the p -percent rule is generally more complicated than other methods and requires statistical exercises to implement.

Given the similar nature of the data released by Energy Information Administration (EIA) and Austin Energy, it is illustrative to examine the EIA disclosure procedures used. EIA has established statistical standards including standards for data protection, accessibility, and nondisclosure. EIA provides micro-level data through the auspices of the Residential Energy Consumption Survey (RECS) and the Commercial Building Energy Consumption Survey (CBECS). Cells are suppressed in many cases due to reasons related to accuracy rather than disclosure. For the RECS data, cell values are suppressed if there are fewer than 10 respondents or the Relative Standard Errors (RSE's) are 50% or greater. For the CBECS data, cell values are suppressed if there are fewer than 20 respondents or the RSE's are 50% or greater.

Austin Energy recommends the following cell minimums:

1. In the case of residential data and small commercial (non-demand) data, individual cells should not be disclosed in cases where there are less than 10 accounts;
2. In the case of large commercial (demand) or industrial data, individual cells should not be disclosed in cases where there are less than 5 accounts or 1 account contributes 60% or more to total, additive peak demand (kW).

Recommended Data Release Process

Austin Energy recommends publishing customer consumption that is aggregated by zip +2 postal codes. Zip +2 codes that require suppression as outlined in the section above, will be grouped and reported at a less granular level. This approach will minimize the amount technical resources allocated to determinations of disclosure and minimize the likelihood that any individual customer can be identified.

This data would be published on-line quarterly, showing monthly totals for the prior quarter. Reports would be presented in table format and could be uploaded by the user.

An initial report will be provided at the end of the year containing FY11 monthly aggregated consumption data gathered from eCIS merged with postal zip codes. Quarterly reports will be provided on on-going basis once the new billing system is implemented. The original requirements for the new billing system do not include zip +4, however the implementation project team is currently evaluating the impacts of adding this requirement to the project costs and timelines. Once the impacts are estimated, the project team will determine whether this new requirement can be implemented along with the initial go-live.

This memo outlines the analysis of Austin Energy zip code data, which led to the recommendation to provide monthly customer consumption data aggregated by zip +2 postal codes. This approach will allow for a best practices method for the release of customer consumption data while protecting the individual nature of that data.