Text Analysis of Proposal(Work order) requests

Step 1: Creating the text files

In the current Data model, The PM table is on the many(*) end of several of it's relationships with other tables, making it a sort of "transaction" table. This table contains several proposal(work order) requests with the information about the nature of the request. In Excel 2016, Powerquery was used to import the PM_Active table from an oracle database, by providing the server name and SQL statement that had been used to put together the columns that constitute the PM_Active table. I was then able to preview the data and do some editing.

In the category column , I was interested in five specific categories : service call, repair, corrective, preventive and regular. I needed only the category and c_description columns in the PM_Active table, I therefore deleted all the other columns and kept just those two columns. While previewing each table, I added a filter to the category column and loaded the work order requests that fell in each category into separate worksheets in excel. I repeated this process till I had 5 different worksheets for each category of work order. I renamed each worksheet to match the kind of word order requests it contained and I deleted the category column from each worksheet.

Next, because the number of rows in the regular and preventive tables would pose a problem later for the analysis in R, I reduced their number of rows to 40000 and 50000 rows respectively. I moved all the worksheets into separate workbooks and then saved each of them as text(tab delimited) files. Now that all the files were in text format, I could then load them into R studio and begin the analysis.

Step 2: R Analysis

Analysis can be performed on any of the text files by running the R code. I was able to report the data in a word cloud, bar plot and pie chart which showed the most frequent words/terms associated with each category of work order.

R contains several packages that can be used for analysis and visualizations. Those used in this project are : tm package for text mining, SnowballC package for text stemming, the wordcloud package for word cloud generation and the RColorBrewer for the color palette.

Text stemming was done to reduce words to their root form e.g flies, flying both originate from fly. The text has also been cleaned to:

- remove numbers, punctuations, special characters like "[", "*", "@", whitespace;
- convert everything to lower case letters;
- remove common words or stop words in English(at , the, we, from,….) as well as words that are expected to be common in these requests (contact, thank, you, fix, email, illinois,…..) because these words do not provide much information value.

Word clouds are read from the inner to outer with the words at the center having the highest frequencies. Words with the same frequency have the same color.

Below is a side by side comparison of the word clouds of the preventive and regular work order requests.
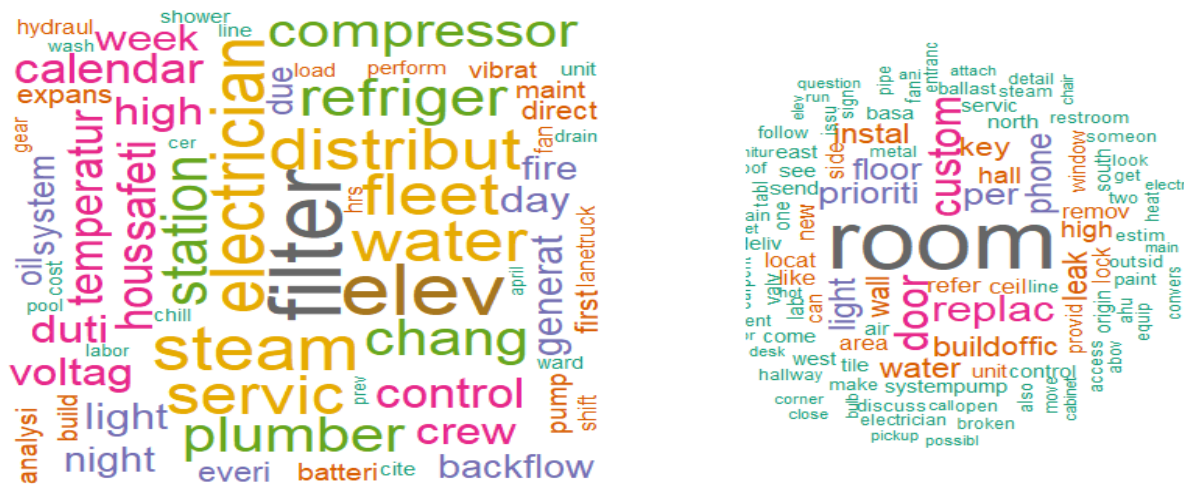
Here is a side by side comparison of the pie charts which show the 20 most frequent words in preventive and regular work order requests

20 most frequent words in preventive reque   **20 most frequent words in regular request**
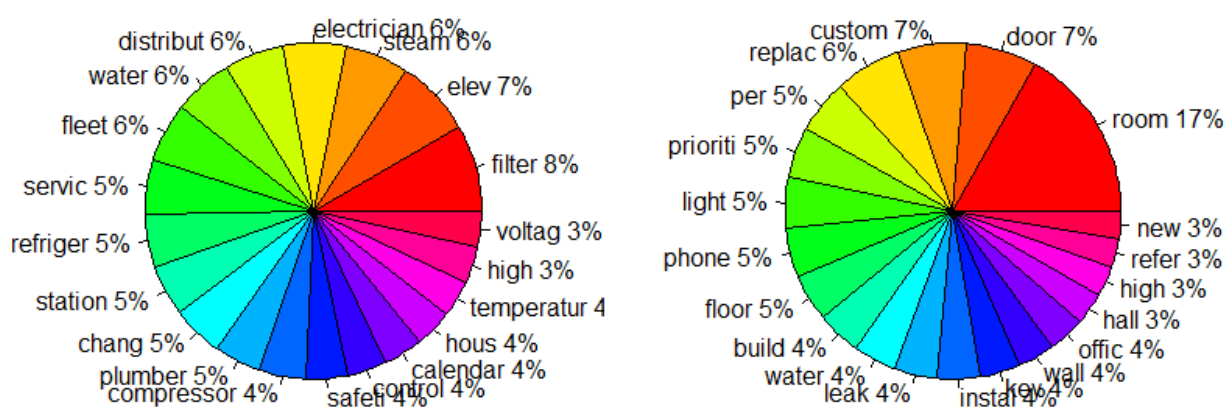
We can find the association/correlation between words in R. For instance, the word "filter" in our preventive work orders is a frequent word and has an association with

```
$filter
maintenancesid        stream        hous        chang        week        station        water
          0.63          0.63        0.60         0.53        0.43          0.40         0.35
```

These visualizations are easily understandable, visually engaging and can be shared with others. Analysis and Visualizations such as this are useful because they help us see at a glance some of the recurring terms related to the categories of work order requests received and give us a better understanding of the nature of these requests.