

## Text Mining Assignment (Maximum Possible: 50 Points)

### 1. Using QIMacros in Excel

You can perform text mining using QI Macros added to Excel. Please always remember to download the free trial. The web address to download the free trial is <https://www.qimacros.com/trial/30-day/>




[Home](#) » [Free Trial](#)

## Create Six Sigma and SPC Charts in Seconds

It's Easy Using QI Macros Affordable Add-in for Excel  
Compatible with PC and Mac, Excel 2007-2016 and Office 365

### Download QI Macros 30-day Trial

Digitally-Signed Trusted Publisher  Symantec.

**Email \***

**Name \***

**Company** (optional)

**Phone** (optional)

**Evaluating For \***  **Interest \***

**\* Required Fields**

[Download 30-Day Trial Now!](#)

Please use your UHV email account. You may use your own name. Type UHV as the organization. You do not have to provide any phone number. You may tell them you are evaluating the program for Professor and your interest is Six Sigma Tools.

**QIMacros**<sup>®</sup> Home  
Reliable Six Sigma SPC Add-in for Excel

Why QI Macros	Key Tools	Training	Services	Support	FREE Resources
---------------	-----------	----------	----------	---------	----------------

[Home](#) » [Free Trial](#)

## Download and Install QI Macros 180-Day Trial!

Soon you will be able to create control charts, histograms, pareto charts and more right in Excel with this easy SPC Software tool.

Follow this Easy 3 Step Process to Install:

### 1. Completely Exit out of Excel on Your Computer.

Don't just close a file; completely quit out of Excel.

---

### 2. Download and Run the QI Macros 180-Day Trial.

PC Users



[.exe—10 MB]

MAC Users



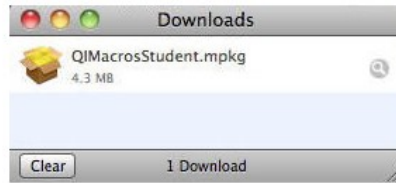
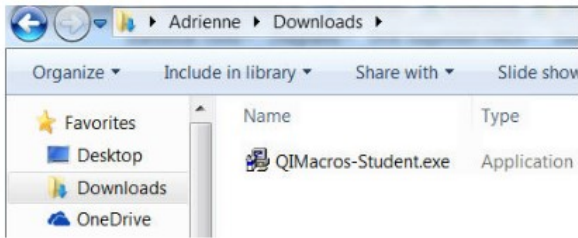
[.dmg—10 MB]

Click on **Run** if your web browser gives you the option or look in your downloads folder to run the file.

**Note:** You may have to unlock the Security & Privacy Settings and select "Anywhere" to install. [Click Here for more information.](#)

Click on **Run** if your web browser gives you the option or look in your downloads folder to run the file.

**Note:** You may have to unlock the Security & Privacy Settings and select "Anywhere" to install. [Click Here for more information.](#)



### 3. Start Excel—look for the QI Macros tab on Excel's menu:

PC



Mac 2011



Mac 2016



## [View More Detailed Installation Instructions](#)

### Look for an Email with Tips on Getting Started



If you don't see the email in a few minutes, check your junk or spam folder.

### Download the How To Guide for More Help



The [QI Macros How to Guide](#) shows you how to use our most popular charts and tools.

- [Control Charts](#)
- [Histograms](#)
- [Pareto Charts](#)
- [Statistical Tests](#)

### QI Macros User Guide



Installation and step-by-step instructions for using QI Macros

### Free Ebooks on Six Sigma

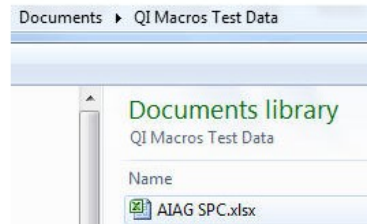


- [Lean Simplified](#)
- [Six Sigma Simplified](#)
- [Six Sigma Action Plan](#)
- [Lean Six Sigma Quick Reference Card](#)
- [SPC Quick Reference Card](#)
- [Lean Quick Reference Card](#)

Once you downloaded the QI Macro free trial, you will be asked to go through a tutorial. You may be offered to choose between manufacturing oriented exercises and healthcare oriented exercises. Simply choose the one you like and go through a few of your interest. *QI Macros Tutorial* is actually included on the QI Macros 2017 ribbon. You can go through the tutorial exercises anytime.

It will be advisable to complete the tutorial first. Since you are in MGMT6352 class, please use the business examples. You will also be advised to watch some tutorial videos. If you feel good about learning by watching the videos, you are ready to move on with the next step:

## Find the Statistical Process Control - Second Edition Data file in the QIMacros Test Data folder:



Create these charts and diagrams in just seconds using QI Macros for Excel...

Buy It Now! ➔



Try It Now! ➔

[« Back](#)

For data mining, you need to use QI Macros Data Transformation Tools that help you analyze your data and reorganize it to run charts.

## Text Mining in Excel

### Count Words to Find Hidden Patterns in Text Data

A healthcare company's service reps entered notes about why insurance claims were denied. This text data, stored in a mainframe computer, was a rich source of information about the cause of denied claims, but extracting the data was considered too difficult because there were over 35,000 individual notes about individual cases. Every note was somewhat different depending on the rep who did the analysis. Analyzing it by hand would have taken weeks. This is where text mining comes to the rescue.

You're already familiar with various forms of Text Mining: search engines, spam filters, spelling and grammar checking and language translation. Text mining can also be useful for Six Sigma.

Companies use data mining to explore large amounts of data to discover meaningful patterns and rules. Text mining, a subset of data mining, explores the written word to find hidden patterns. By some estimate, stored text contains as much as 80 percent of the key information needed for business improvement. What are some sources of text data for mining?

- Service rep or help desk comments
- Customer complaints
- Transcriptions of recorded voice such as doctor's notes
- News stories

Text data, by its very nature, is unstructured and varies from person to person. So categorizing the text becomes difficult. In the healthcare denied claims example, reps spelled Medicare in various ways: MDCR, MedCR, Medicr, and so on. This variability makes text analysis more complex and challenging than plain numbers.

There are several types of text analysis:

- Root cause analysis
- Sentiment analysis to understand attitudes
- Clustering of documents (e.g., affinity diagram)

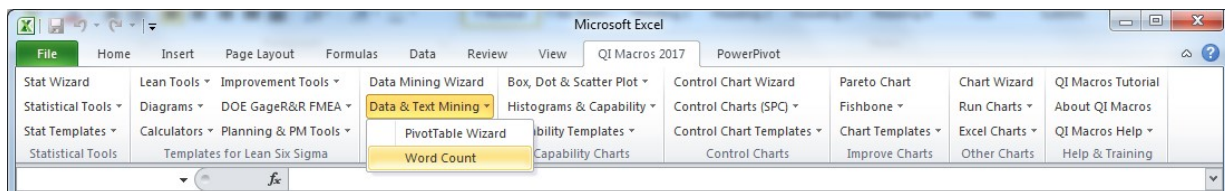
## 2. Root Cause Analysis

The first flavor of text mining attempts to look at the root cause of comments, a key method for Six Sigma. The goal is to summarize the various words and phrases to see if there's a Pareto pattern (80/20) to the text data. To do this, you will need software to parse the text into words and phrases.

*The QI Macros Word Count tool* will handle this quite easily. It may take a while if there are lots of records, but it's much faster than trying to do it by hand or using Excel's COUNTIF statement. Most help desk software allows an export to Excel, so it is easy to get the text file. Here's a subset of the 35,000 healthcare denied claims comments (Please open the attached file *Word Count\_pivottable.xlsx*):

	A
1	<b>MEMOTEXT - Wordcount</b>
	DA REJECT - DUP DOS (MULTIPLE VISITS SAME DAY). CLD AND TT SHELLY AT AETNA WAS TOLD THAT THIS SHOULD BE INCLUDED IN THE SURGERY CHARGES. E-MAILED GARY S TO SEE IF THIS IS CORRECT. THE SURGERY
2	ACCT IS IN MC STATUS. CATHY X266
	DANIEL FROM DAVIS WIRE 0 CALLED TOSAY THEY RECVD A W/C CLAIM FOR THIS PT AND THEY ARE NOT ASSOCIATED WITH DAVIS WIRE IN THAT REGARD...THE INS INFO IS INCORRECT...WILL DELETE AND REFER TO WC
3	FOLLOWUP...ADAVIS
4	DA REJECT - DUP DOS (MULTIPLE VISITS SAME DAY).
	DA REJECT - DUP DOS (MULTIPLE VISITS SAME DAY)/ RECEIVED FROM AILHANDLERS EOB THAT THEY WILL NOT PAY CLAIM/ CALLED TT KATHY/ SHE STATED CLAIM NOT PAID BECAUSE THEY RECEIVED 2CLAIMS FROM US WITH
5	SAME DOS/ NON WERE MARKED CORR

To use the Word Count tool, simply select the data then click on the QI Macros Menu and choose **Word Count** from the Data Mining sub menu:



**QI Macros Word count** function parses the words out of sentences and paragraphs and uses Pivot Tables to count the occurrences of individual words and two word phrases. Then it displays the totals in a descending order for analysis and creating charts.

	A	B	C	D	E
1	Count of Word			Count of Two-Word Phrases	
2	Word	Total		Two-Word Phrases	Total
3	DUP	11		REJECT DUP	8
4	DOS	9		DUP DOS	8
5	REJECT	8		DA REJECT	8
6	DA	8		MULTIPLE VISITS	6
7	SAME	7		DOS MULTIPLE	6
8	FOR	7		VISITS SAME	6
9	MULTIPLE	6		SAME DAY	5

Word Count strips out common English words such as "the" or "and". What's left is a Pareto pattern.

**Most common words:** DUP and DOS (which stands for Date Of Service).

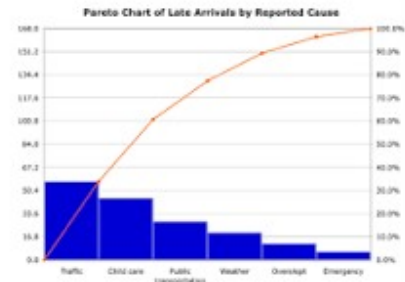
**Most common phrases:** DUP DOS, multiple visits, visits same day.

**Ah Ha!** Insurance companies were rejecting claims because they had an overlapping (duplicate) date of service. This led the healthcare company to examine these rejects and they discovered that the patients were admitted from the Emergency Department resulting in one claim for the ED and one for the nursing unit, rather than one claim for both. Procedures were changed and this type of denied claim vanished saving millions of dollars and endless hours of rework.

Text mining can use more complex natural language methods, but Word Count's "bag of words" approach will serve most improvement teams. Text mining is a powerful tool for Six Sigma and the QI Macros Word Count tool helps make it simple.

Another useful tool for text mining is **Pareto Chart**. Pareto Chart is a handy QI Macro tool to provide a graphical display of the results. Here is the definition:

A **Pareto chart**, named after Vilfredo **Pareto**, is a type of **chart** that contains both bars and a line graph, where individual values are represented in descending order by bars, and the cumulative total is represented by the line.



[Pareto chart - Wikipedia](https://en.wikipedia.org/wiki/Pareto_chart)  
[https://en.wikipedia.org/wiki/Pareto\\_chart](https://en.wikipedia.org/wiki/Pareto_chart)

Now it is your turn to do text mining in QI Macro.

1. Please download and open the Excel file *Pareto*.
2. Choose "Text" sheet tab.
3. Select the text column.
4. Select QI Macro/Data & Text Mining/Word Count.
5. You should see a pivot table created. Please tell me the top three defects according to the tabled totals.
6. Choose "Text" sheet tab again.
7. Click the heading and click the Pareto chart as instructed on the sheet.
8. By this time you should have added three sheets: Word Count, Words, Defect.
9. Please briefly compare and explain what you've found from your text mining experience.

### 3. Sentiment Analysis

This exercise uses purely Excel for text mining. This type of data analysis is almost the same as open-ended question data analysis in a survey study. Please watch the video clip in YouTube at <https://www.youtube.com/watch?v=yWBXV651yd4> to understand the common procedures. Here is the exercise for you:

1. Please download and open the file *Feedback on Shoes.xlsx*
2. Some feedback data has already loaded into the spreadsheet "Shoes" for you.
3. Please follow the procedures from the video clip to perform your sentiment analysis.
4. Please eye ball the review column and create a few theme categories as you see fit.
5. Complete the numbering in Column A and complete the codes based on the themes.

6. Complete coding in Column D for every review feedback.
  7. Use the formula Countlf() to calculate frequency for each feedback theme in Frequency column and calculate the percentage over the total number of reviews for each in % column.
  8. Please also create a chart based on the result.
  9. If you can perform such text mining over EBay data set on the next spreadsheet (in the same workbook), you can win 10 bonus points on this assignment.
  10. Hint: when some feedback contain several themes, you may have to copy and paste to split these theme based reviews into several new rows before coding. Nowadays, there are many Excel add-in tools to help ease the work on text mining. However, text mining using simply an Excel worksheet helps you to see the work logic of how text mining is handled.
-