Microsoft Text Analytics Project

Paul Stortz

DATA 620, Spring 2021, Assignment 12.1

University of Maryland Global Campus

Instructor: Dr. Majed AL-Ghandour

Microsoft Text Analytics Project

# Executive Summary

A text analysis was done comparing historical letters to shareholders from Microsoft to see the trend of word counts over time. This analysis can help us figure out what Microsoft sees for the future of computing. This analysis will find key market trends, product trends, and marketing strategies identified in Microsoft’s letters to shareholders.

The data used was from the Microsoft shareholders letters. The time periods included were the 1996, 2000, 2005, 2010, 2015, and 2020 shareholder letters. A custom Python script was run on each shareholder letter as a text file and had its own logic to manipulate each data set. Punctuation, bullets, “stop words”, plural words, past tense words, and adverbs were all manipulated into root words for the text analysis to find trends more accurately.

The analysis found that Microsoft sees cloud technology, data-centric solutions, Microsoft Azure, and AI as core product offerings in the future. Microsoft also sees a future with customer service being at its core with a focus on the helping the world and organizations with technological solutions.

Next steps for the organization from this analysis include:

* Focus on more data driven decisions and analysis.
* Obtain the correct employees and cloud technology for a data-driven organization.
* Maintain a customer-first organizational culture.
* Keep customer service a number 1 priority.

By following the above steps, the organization can stay on the cutting edge of technology and follow the same customer service excellence as Microsoft.

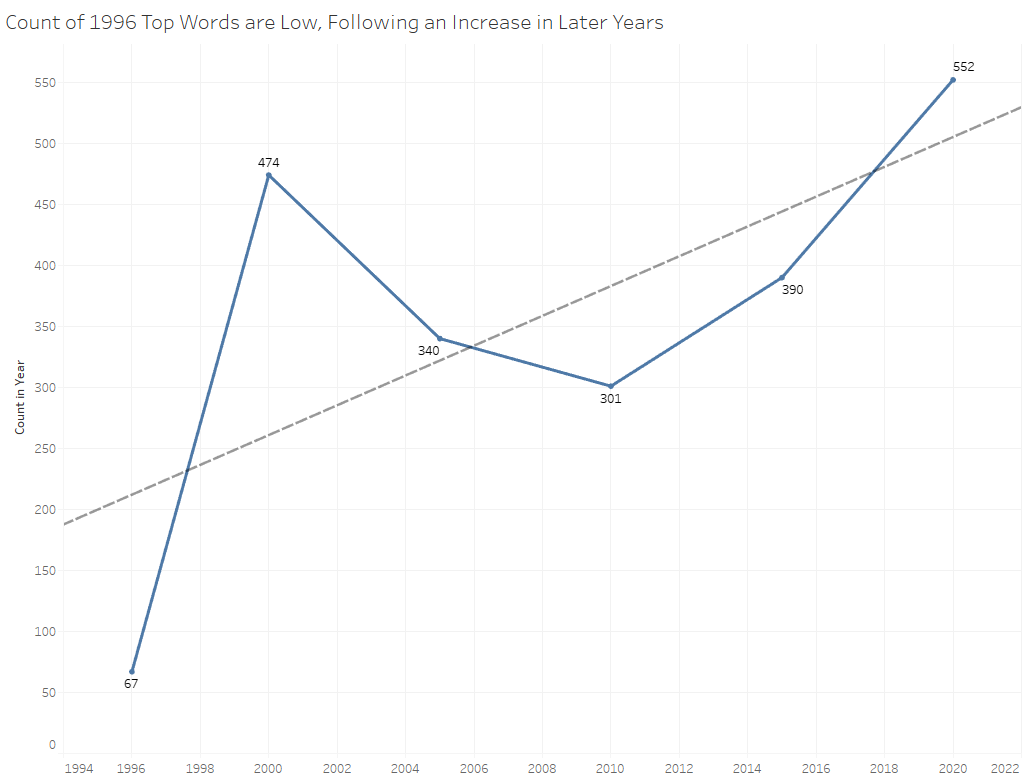
# 1. Dataset Outline, Data Preparation, and Objective

The text datasets chosen for the Microsoft Text Analytics Project are the annual Microsoft letter to shareholders sent out each year in their annual report. Six years’ worth of shareholder letters were chosen with an objective to space out each dataset by five years, although the first period used is the 1996 letter to shareholders due to availability. The years used in the analysis were the 1996, 2000, 2005, 2010, 2015, and 2020 letters to shareholders. The top fifty words by count were then exported into an Excel workbook and uploaded into Tableau for further visualization analysis. Each text dataset can be found on Microsoft’s Investor Relations webpage under the annual reports section (Microsoft Corporation, 2021).

Once the source data was gathered, a custom Python script was run on each shareholder letter as a text file. The Python script had its own logic which excluded certain “stop words”, or words that have little meaning in the analysis such as “any” or “again”. These stop words are generally prepositions and other linking verbs. The full list of stop words was provided by Lextek.com and is a well-known general source for stop words (Lextek.com, 2021). The script then removes punctuation, special characters, and bullets so the analysis will only include actual words. Each word is then lower-cased so different capitalization methods would not count as two words. The script then removed suffixes such as “ly”, “ies”, and “ied” to make sure the root word was included. Finally, the script removed past tenses and plural words. This was done by checking the word with a dictionary of words ending in “e” to make sure words like “loved” only had the last character removed as opposed to the last two characters “ed” (Word Tool Box, 2021). The intent of all the data preparation methods was to get the root word in a present tense or present participle for the analysis to better understand each dataset.

Each dataset was then checked for significance to make sure the number of unique words that appeared could be analyzed. As shown below in Figure 1, the Microsoft letter to shareholders for 1996 had a low number of unique words, but the following years had a higher count.

*Figure 1. Count of 1996 Top Words are Low, Following an Increase in Later Years*



*From Tableau, Microsoft Text Analytics*

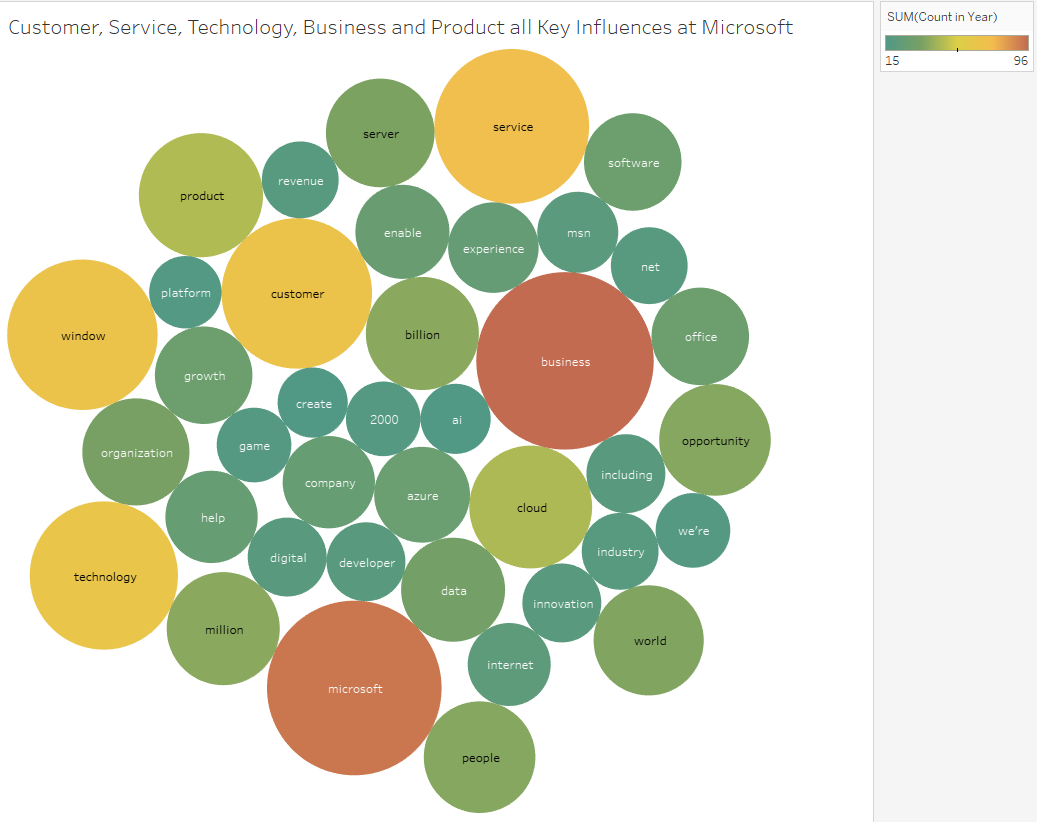
The low number of unique words in 1996 is overshadowed by higher amounts in later years which peaks in the last dataset in 2020. The trendline of the datasets are increasing, which suggests the letters to shareholders are getting larger and each one is repeating key words. This repetition of key words is key in the analysis.

By analyzing these important annual letters, one can get inside the minds of the executives at Microsoft. Letting the market and the shareholders know the direction of such an important corporation can tell us what Microsoft sees for the future of computing. This analysis will find key market trends, product trends, and marketing strategies identified in Microsoft’s letters to shareholders.

# 2. Data Analysis and Findings

Each of the top words for all years will be shown, followed by a trend analysis, and then broken down into a trend of the top words over time. The top words that had 15 or more appearances are displayed below in Figure 2 below where one can see that the words “customer”, “service”, “technology”, “business”, and “product” are important to the Microsoft culture. Words such as “Microsoft” and “window” appear many times but provide little value since these words are the company name and the company’s flagship product. These two words will be excluded from future figures as a result.

*Figure 2. Customer, Service, Technology, Business and Product all Key Influences at Microsoft for all Time Periods*

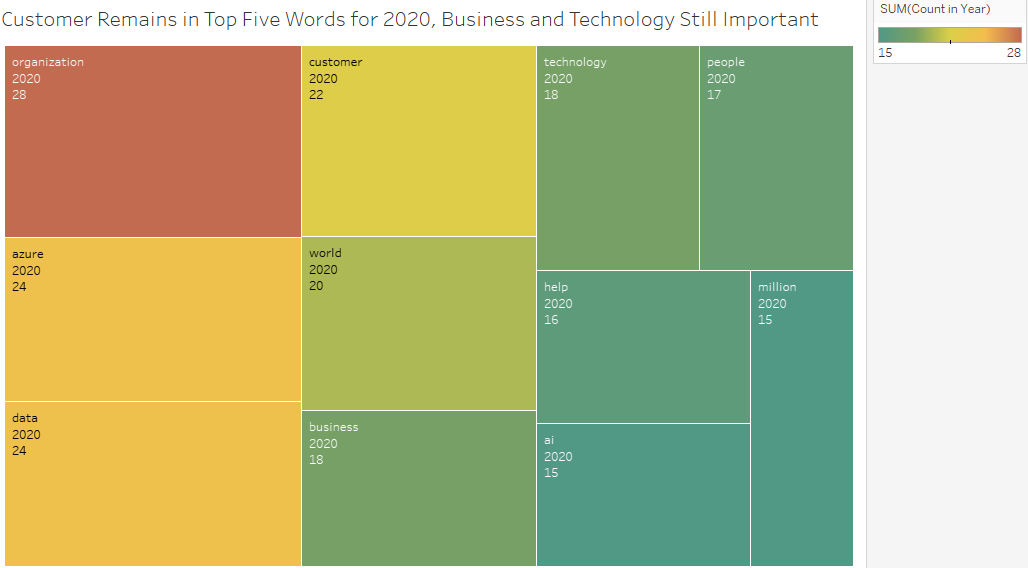


*From Tableau, Microsoft Text Analytics*

As shown above, customer service is extremely important to Microsoft in all years, which follows other tech giants such as Amazon. Microsoft also wants to convey the company and its services to be focused on the business community. The company wants others to know that their products are designed to facilitate better business. This marketing strategy can be seen in Microsoft’s product offerings, which are also displayed above, with words such as Windows, Azure, “ai”, and cloud computing. Technology is another key word that is used for all years as Microsoft’s driver is its technology services the company can provide.

Technology, more than any other industry, is rapidly changing. When looking only at the most recent time period of 2020, only the word “customer” remained in the top five words. In Figure 3 below, one can see that even though “customer” remained in the top five words used, “business” and “technology” are still in the top ten words used in the 2020 shareholder letter.

*Figure 3. Customer Remains in Top Five Words for 2020, Business and Technology Still Important*



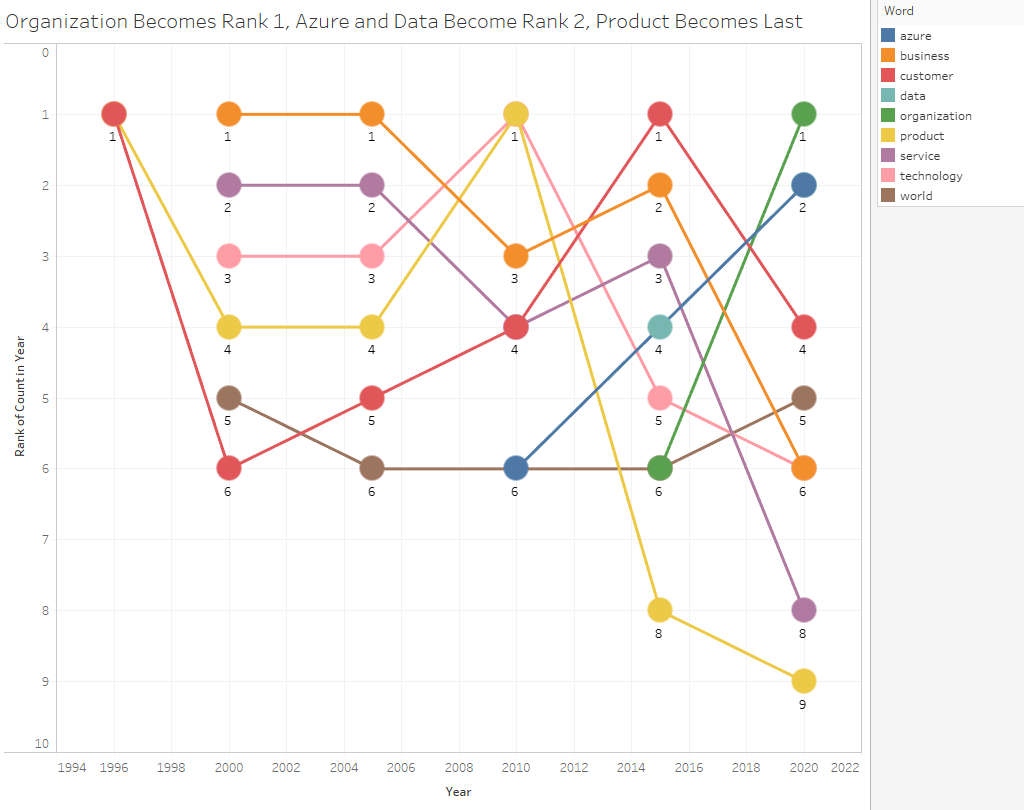
*From Tableau, Microsoft Text Analytics*

The word “organization” appears most frequently, perhaps as a more inclusive and friendly synonym for “business”. The other words which appeared were Microsoft’s fairly new cloud computing platform, Azure, which Microsoft has been heavily marketing as a competitor to Amazon Web Services. The word “data” is tied for the second most used word, as a clear trend in the marketplace is around understanding data and analyzing data to turn into actionable decisions. The word “world” appears as the fifth most used word, highlighting Microsoft’s global focus to be the number one business software platform in the world. The words “technology” and “ai” also appear which shows Microsoft’s commitment to invest in emerging technologies.

# 3. Top Word Analysis and Trend

The top five words for all time periods, excluding “window” and “Microsoft”, are “customer”, “service”, “technology”, “business”, and “product”. These words may have been important to Microsoft in the past, but the technology sector changed as shown in the aforementioned Figure 3 and in the 2020 shareholder letter. In the 2020 letter, the top five words were “organization”, “azure”, “data”, “customer”, and “world”. In Figure 4 below, one can see the trend of the top five words over time and the top 2020 words over time to display a trend analysis. Each word is displayed in a ranking within that specific time period.

*Figure # 4. Organization Becomes Rank 1, Azure and Data Become Rank 2, Product Becomes Last*

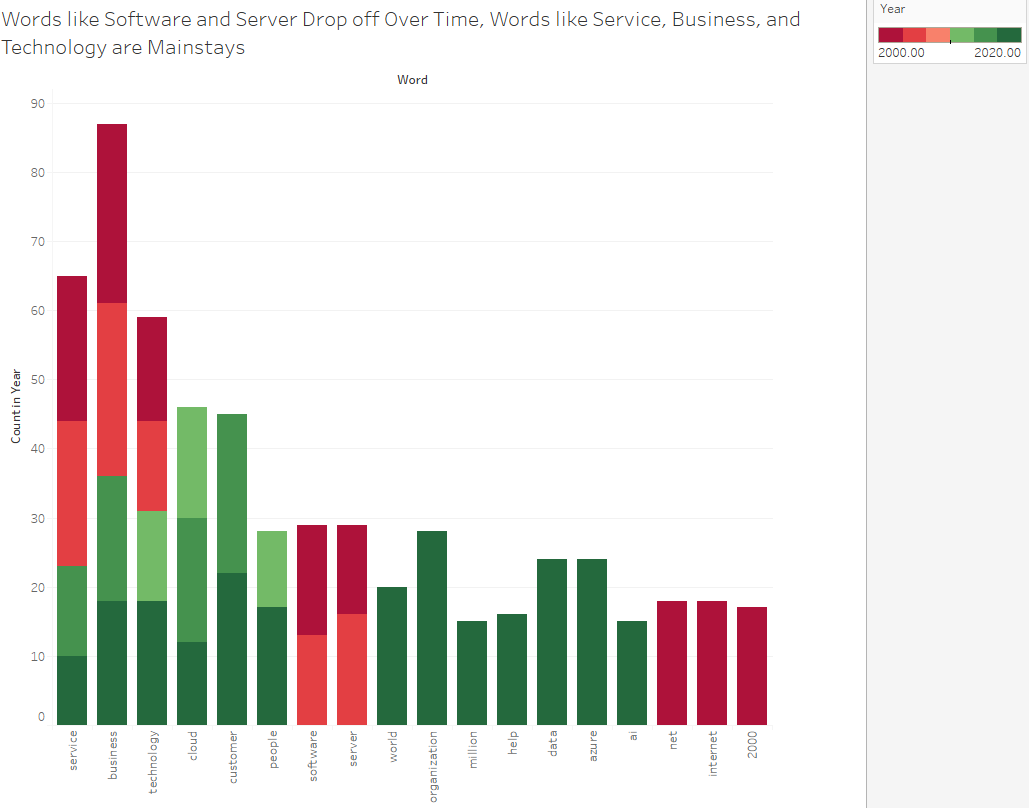


*From Tableau, Microsoft Text Analytics*

The word “product” was an historically highly ranking word between 1996 and 2010 but becomes the poorest ranking in the selected dataset. The word “organization” becomes the mostly used word in the 2020 letter, which seems to be replacing the word “business” as the latter word falls in use over the same time periods. The Microsoft cloud product Azure skyrockets from rank 6 in 2010 to rank 2 in 2020 with a renewed focus on cloud software and rapid growth in that area. The word “data” also ties for rank 2 in the 2020 letter from rank 4 in 2015, showing that data analytics is more important than ever to Microsoft’s future.

When comparing the total usage of words from the past to the present, some words drop off massively in frequency. In Figure 5 below, words that appear more than 10 times in a letter are shown. Words such as “service”, “business”, and “technology” are repeatedly used but words such as “software”, “server”, “net”, “internet” and “2000” are not used with a count of 10 occurrences since the 1996-2005 letters.

*Figure # 5. Words like Software and Server Drop off Over Time, Words like Service, Business, and Technology are Mainstays*



*From Tableau, Microsoft Text Analytics*

Figure 5 shows that some words do not change overtime, even in a rapidly changing technology business. When the internet was new and being rapidly utilized around the year 2000, the shareholder letters reflected the prevalence of popular terms from that period, such as “net”, “internet”, and “2000” in Microsoft’s shareholder letters. In 2020, it is clear that Microsoft is focused on the cloud, data, Microsoft Azure, and AI, along with other mainstays shown above.

# 4. Next Steps and Conclusion

The next steps in evaluating this analysis would be to focus the company down a more data driven path. Microsoft sees the future with easy access to cloud technology and data. In order to capture business value from data, employing the right human capital to extract that value and making sure the infrastructure at the organization is suitable becomes more critical. Having a customer-first attitude and having a strong customer service organizational culture has been serving Microsoft well in the past and will continue to serve all organizations well into the future.

This analysis could be used in a variety of different marketing strategies and shows where Microsoft sees the future of their business. Microsoft sees cloud technology, data-centric solutions, Microsoft Azure, and AI as core product offerings in the future. Microsoft often sees a future with customer service being at its core with a focus on the helping the world and organizations with technological solutions.

# References

Lextek.com. (2021, April 4). *Stop Word List*. Retrieved from Lextek.com: http://www.lextek.com/manuals/onix/stopwords1.html

Microsoft Corporation. (2021, April 4). *Microsoft Annual Reports*. Retrieved from https://www.microsoft.com/en-us/Investor/annual-reports.aspx

Word Tool Box. (2021, April 6). *Verbs Ending with E*. Retrieved from Word Tool Box: https://wordtoolbox.com/verbs-ending-with/e