Summarizing Spreadsheets for Voice Interfaces

Kerem Baktiroglu

Practical Course: Computer Science 1

Abstract

The world is facing with a pandemic situation called COVID-19. As many people suffered or lost their jobs and even their lives, the subject became/becomes very important to consider. In this Practical Course 1, a talking chatbot is designed and the chatbot answers some of the important questions regarding to people' interest.

1 Introduction

This section briefly explains the topic of chat-bots in the context of pandemic situation called Covid19 and relation to the related reader. In following sections, motivation of the project will be introduced, thematic analysis of the project will be exhibited and challenging issues are mentioned and finally technical parts are explained.

Talking chatbots are becoming more and more popular in todays world in order to ease the related person's questions. Amazon Alexa, Siri, Cortana and many others can be given as an example on this subject. People are becoming more curious in order to have an information on covid19 disease and learn some specific answers regarding to their specific questions.

In this project, regarding to people needs, a talking chatbot called "Covid19 Chatbot" is designed with Python programming language. In order to do that, firstly thematic analysis should be done on the subject and the provided interview is inspected which was given by Laura Koesten.

In the following parts, thematic analysis of the interview and information on technical part would be provided.

2 Thematic Analysis

In this section, an interview is provided so that, the corpus can be extracted and the thematic analysis can be done on it. The spreadsheet and interview cannot be provided in this report since the privacy reasons. However, the main idea behind them would be explained to the related user.

Themes	Categories	Examples
Uncertainty on Dataset	Uncertainty	No idea on Dataset
		Confused about question
		Asking questions
	Alternative Interpretations on Dataset	Numerical Interpretations
		Interpretations on dissimilar question
Improper information on Dataset	Incorrect facts	Misunderstanding of the question
		Misunderstanding of the Swine flu dataset
Distrust on the Dataset	Changing the dataset information	Changing the header of the dataset
		Changing columns of the dataset
		Reconstructing the complete dataset regarding own idea
Redundancy	Detailed Explanation	Rambling on Swine flu Dataset
Acknowledgement	True Information	Predicting true information on dataset

First of all, a code book is designed for the project and inspected how people talked about the data. There are 7 spreadsheets which were gathered from the previous study and inspected. These 7 spreadsheets are used for the thematic analysis.

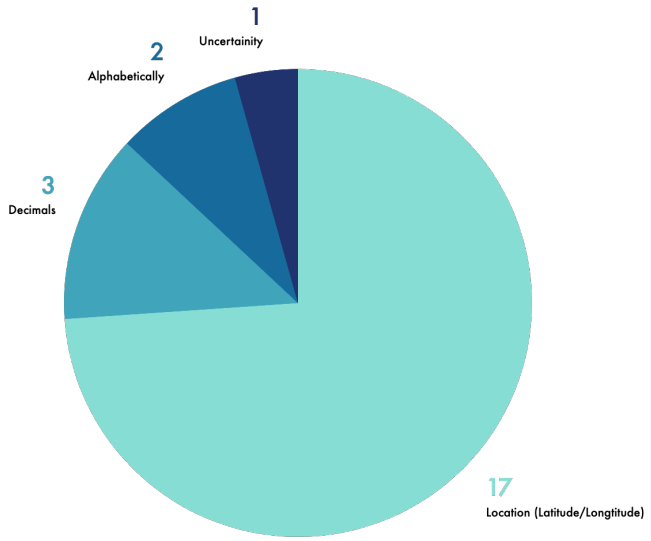
The questions which were asked to the related interviewers, which names are not provided in the project, were like below.

1. Please describe the entire column “deaths per million population” (Column D)
2. If you had to make a one-sentence summary of the headers, what would you say? (Can you see any groups?)
3. What are the three most important columns? Why?
4. Please look at columns B, K and L. Could you please describe these columns?
5. Please look at column M and describe it to me.
6. If you had to make a one-sentence summary of column E, what would you say?
7. If you had to make a one-sentence summary of row 3, what would you say?

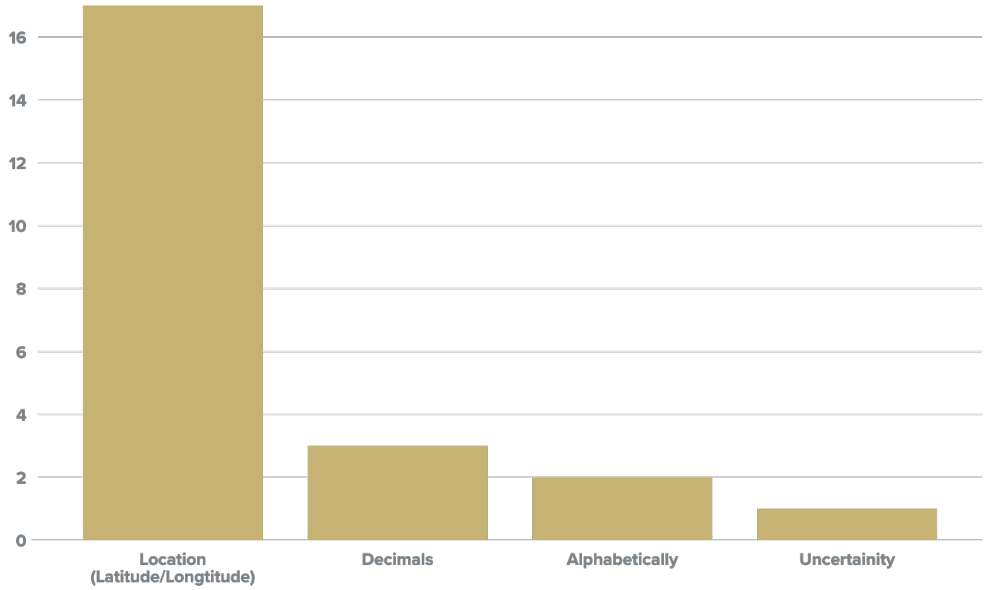
These 7 questions answers are inspected and thematic analysis was done on them.

The high approach of the thematic analysis and analysis itself is given below;

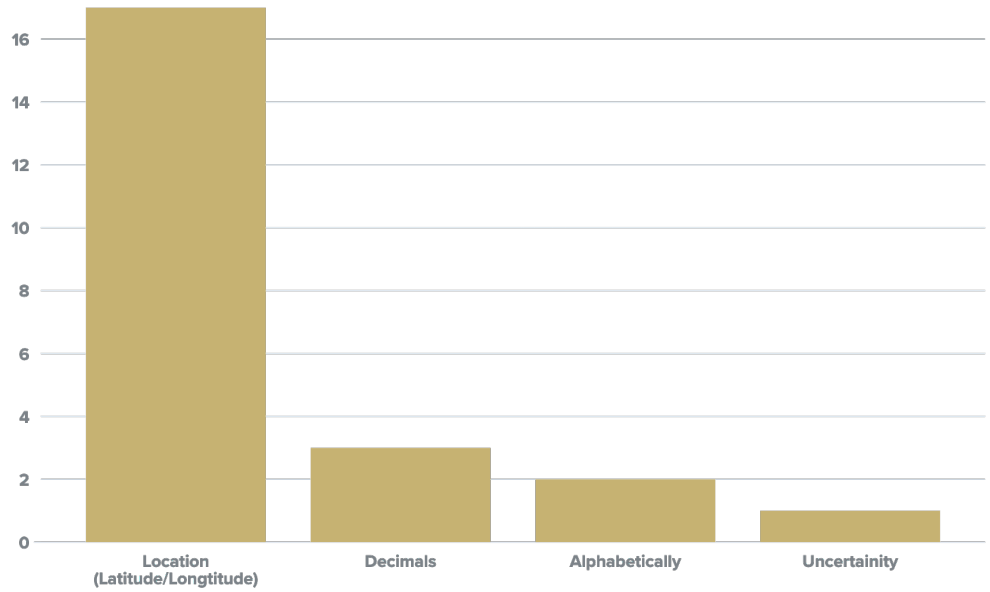
People talking on Country Information



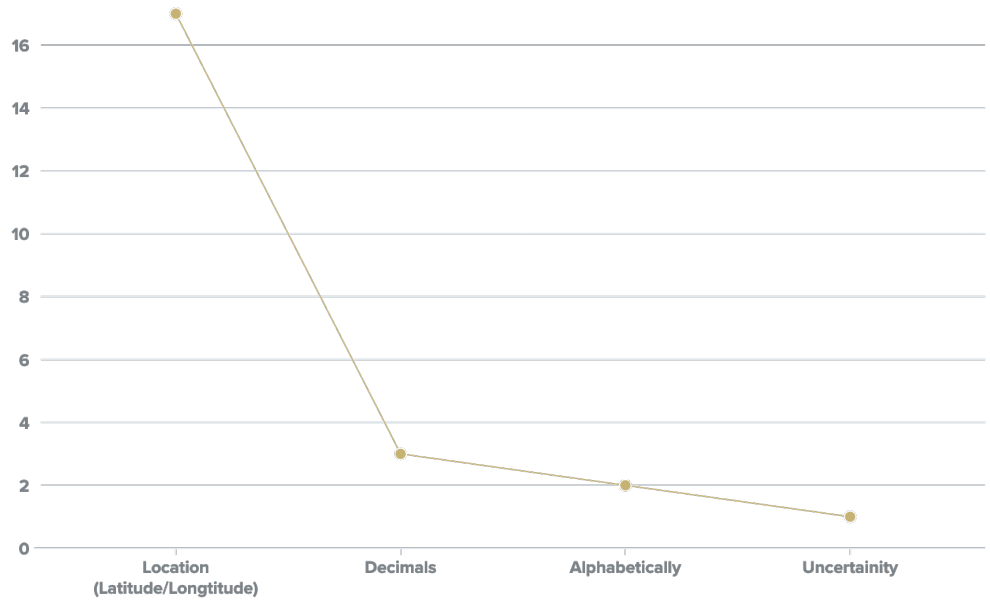
People talking on Country Information



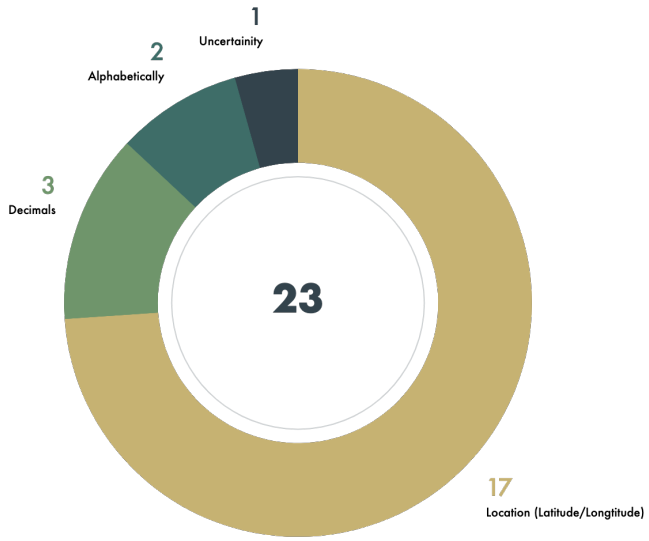
People talking on Country Information out of 21 Person



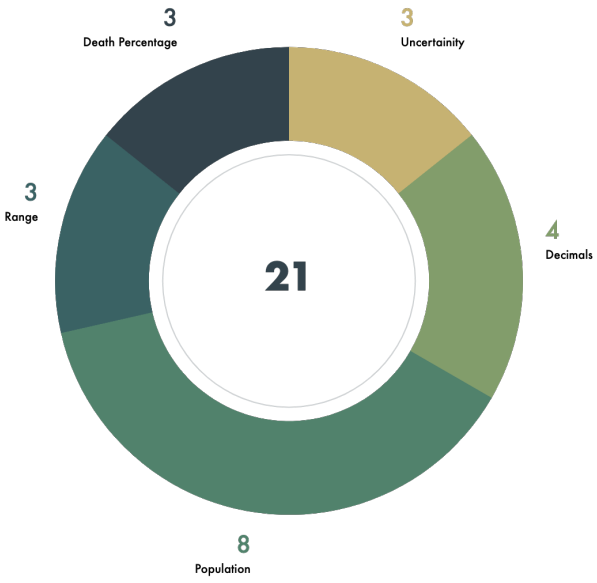
People talking on Country Information out of 21 Person



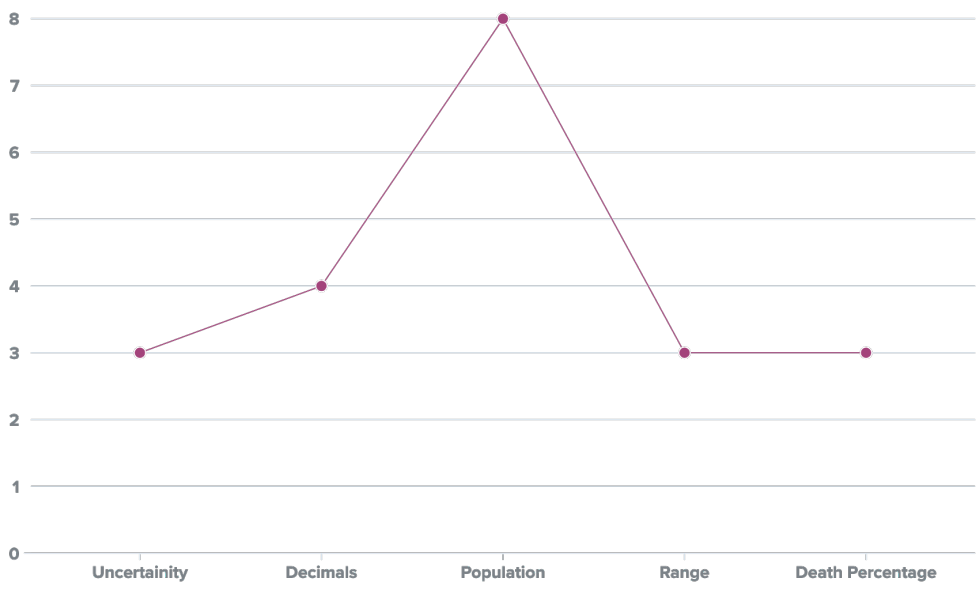
People talking on Country Information out of 21 Person



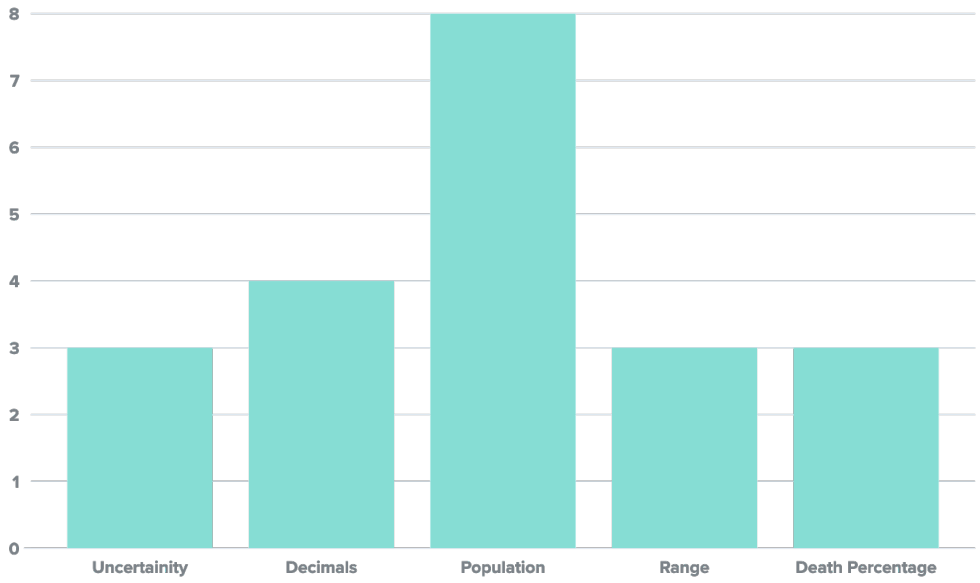
People talking on deaths per million population out of 18 Person



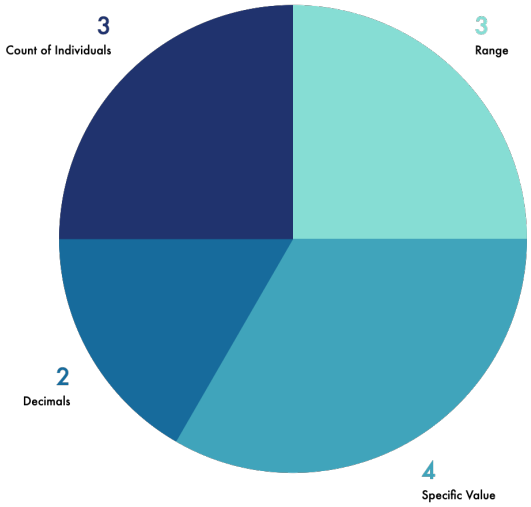
People talking on deaths per million population out of 18 Person



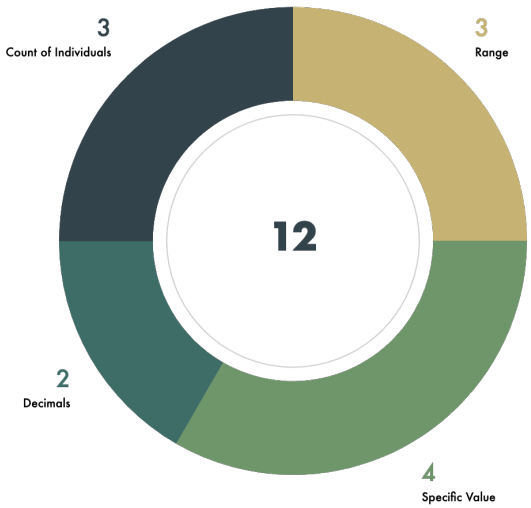
People talking on deaths per million population out of 18 Person



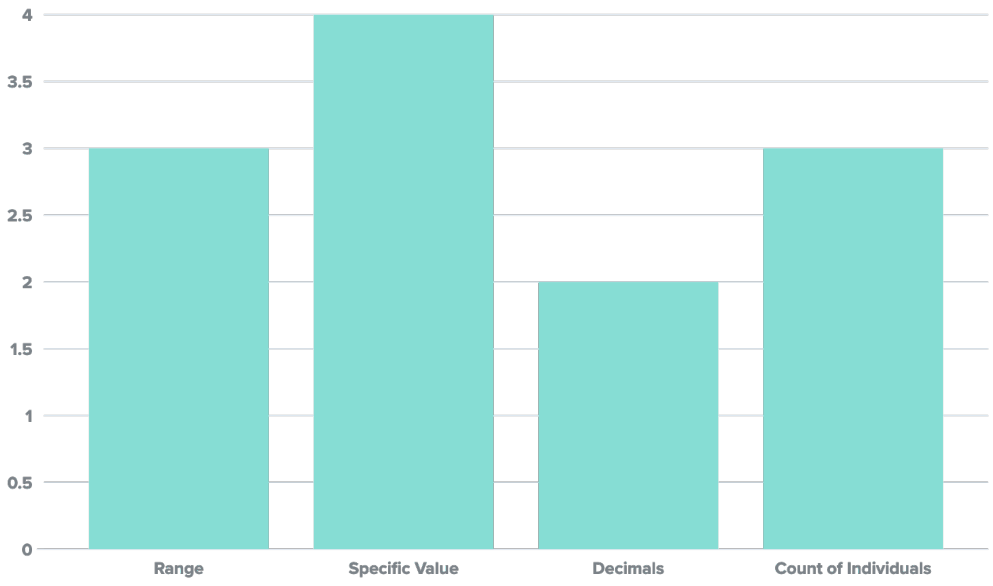
People talking on confirmed cases out of 17 Person



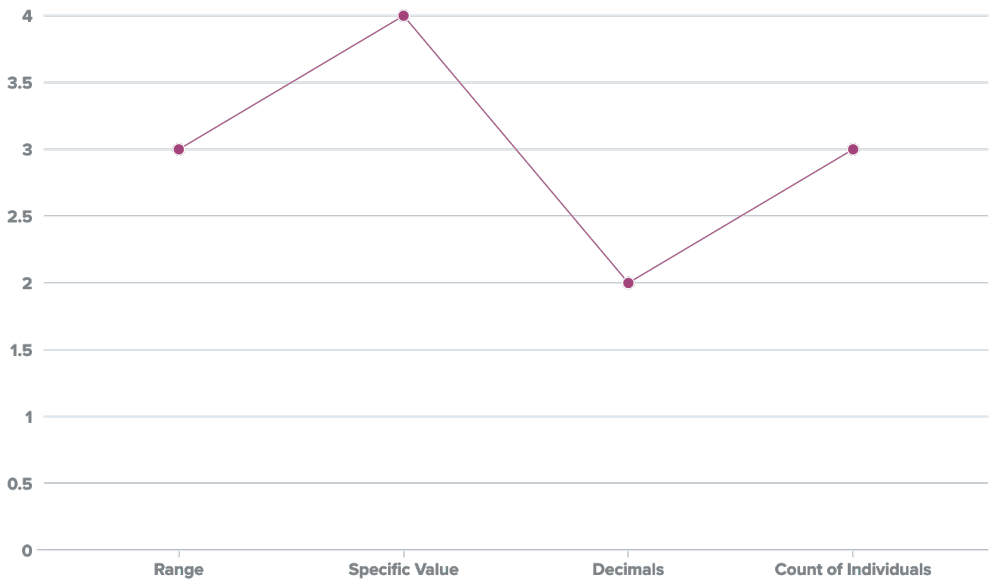
People talking on confirmed cases out of 17 Person



People talking on confirmed cases out of 17 Person



People talking on confirmed cases out of 17 Person



3 Questions on Chatbot

Regarding the thematic analysis on the spreadsheets, the questions are gathered from different kind of sources such as Quora, Yahoo, CDC and CNN.

These questions are listed above;

· Which country has had the most deaths due to COVID-19? - Quora

· How many COVID deaths occurred in (country) as of (date) ? - Yahoo

· Which country has the most death number? - CDC

· Which country has the least death number? - CDC

· What is the current pandemic situation in (Country Name)? – CNN

These are the question set which the talking chatbot should answer.

4 Datasets regarding Questions on Chatbot

While dealing with answers on talking chatbot, mainly 3 datasets are used. Mainly each question will be answered by 2 datasets.

These datasets are;

1. WHO (World Health Organization) Dataset
2. ECDC (European Centre for Disease Prevention and Control) Dataset
3. John Hopkins Coronavirus Resource Center Dataset on Covid19

5 Designing the talking Chatbot

Talking chatbot is designed with the Python programming language tool. The language of the talking chatbot is English.

While designing and coding the chatbot, google text to speech, csv, os, playsound, pycountry, iso3166, pandas, numpy and speech recognition libraries are implemented.

Anaconda was used during the compilation of the python code.

On the other hand, there needs to be implemented such as PyobJC, pyaudio iso3166 and gTTs in order to compile the python code to run the code.

6 Open Issues

Since the project time is limited and very serious problems occured in developer side, the project is not completed as it planned to be so that there are plenty of open issues such as, the talking chatbot,

Covid19 Chatbot does not provide any other languages other than English language. Also, questions needs to be asked grammatically correct and with a clear speech in order to be answered. Finally, an application and GUI can be designed later on.

7 Conclusion and Outlook

Because of pandemic situation called Covid19 disease in whole world, a useful solutions needs to be done in order to ease the process of getting information. That is the main reason, why a talking chatbot is designed to contribute it.

Previously, a research was given and regarding to that research, thematic analysis done and also regarding to thematic analysis, an answer dataset such as WHO, ECDC and John Hopkins datasets which are reliable, were found and answers were provided.

To sum up, this talking chatbot can be improved more and useful functions can be implemented regarding to people questions with the help of Data Science approaches.

During this project, I would like to thank a lot Ms. Laura Koesten, MSc Ph.D., for her help and patience.