**NATURAL LANGUAGE PROCESSING**

**CV ANALYSIS USING NATURAL LANGUAGE PROCESSING**

**Collaborators.**

1. **KELVIN RONO**
2. **FELIX KIPKIRUI**
3. **FELIX KOECH**
4. **MAURINE CHEROTICH**
5. **DERICK ONYONKA**

INTRODUCTION

Many are times companies need to employ or promote people in different fields for varies jobs. In order to determine who is best suited for a particular field, it becomes necessary to analyze Curriculum Vitae of people who showed interest by turning them in. Analyzing is necessary in order to rationally give the person meeting all the requirements for efficiency.

PROBLEM STATEMENT

The need to vet interested candidates many lead to several other problems. For example, in big companies there might be a lot of candidates turning in their CVs. All of which have to be analyzed. The process should also not be bias, which might be the case if the process is handled by a qualified panel.

OBJECTIVES

The main objective of this software is to utilize Natural Language Processing to analyze CVs and suggest the most suitable candidate.

By so doing, we also achieve the analysis in a timely, non-bias and convenient. The software can analyze a lot of CVs in a short period of time.

NATURAL LANGUAGE PROCESSING

Natural language processing is a subfield of AI concerned with the interactions between computers and human language.

How it works basing on our project:

1. A computer must first be taught how to extract valuable information from human input. Then it stored the important information in a table so that it can understand it better. This is **natural language understanding.**
2. At some point we will need to present data to the user. In order to do so, the computer must manipulate the data that is now stored in tables. This is a process called **natural language processing.**
3. Finally, after the understanding and processing we will have the desired result of the analysis. In order to present it in a way that humans can understand, we will utilize **natural language generation.**

In order to accomplish all this, we used IBM’s Watson studio. We first created a project then collected required data for training and some mock up data that we will analyze.

The programming language used is Python due to how fast and widely supported it is.

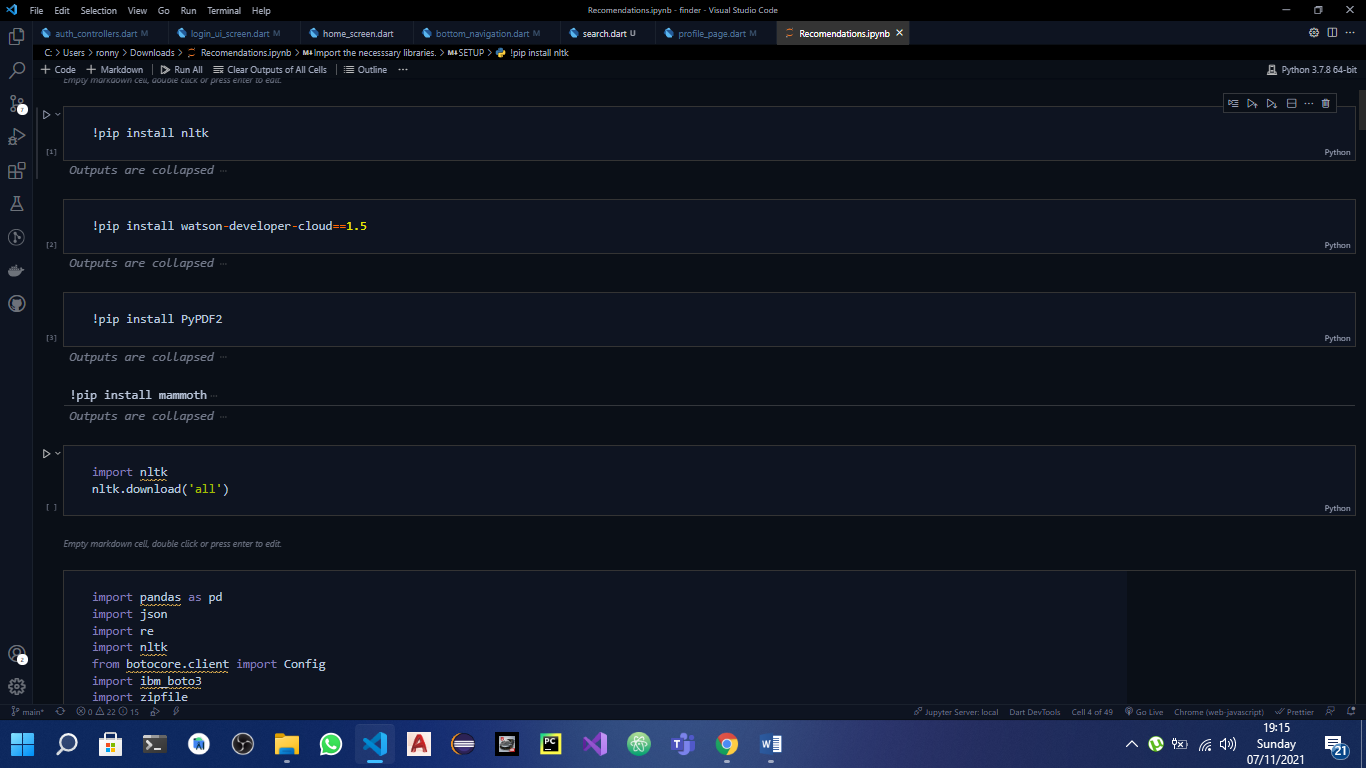
**Writing the program.**

1. Setup – Installation and Importation of libraries

First of all, we used nltk – Natural Language Toolkit, a python package for natural language processing

We installed a few more packages to be utilized throughout the project.

Below is a screenshot if the installations and importation of the packages.



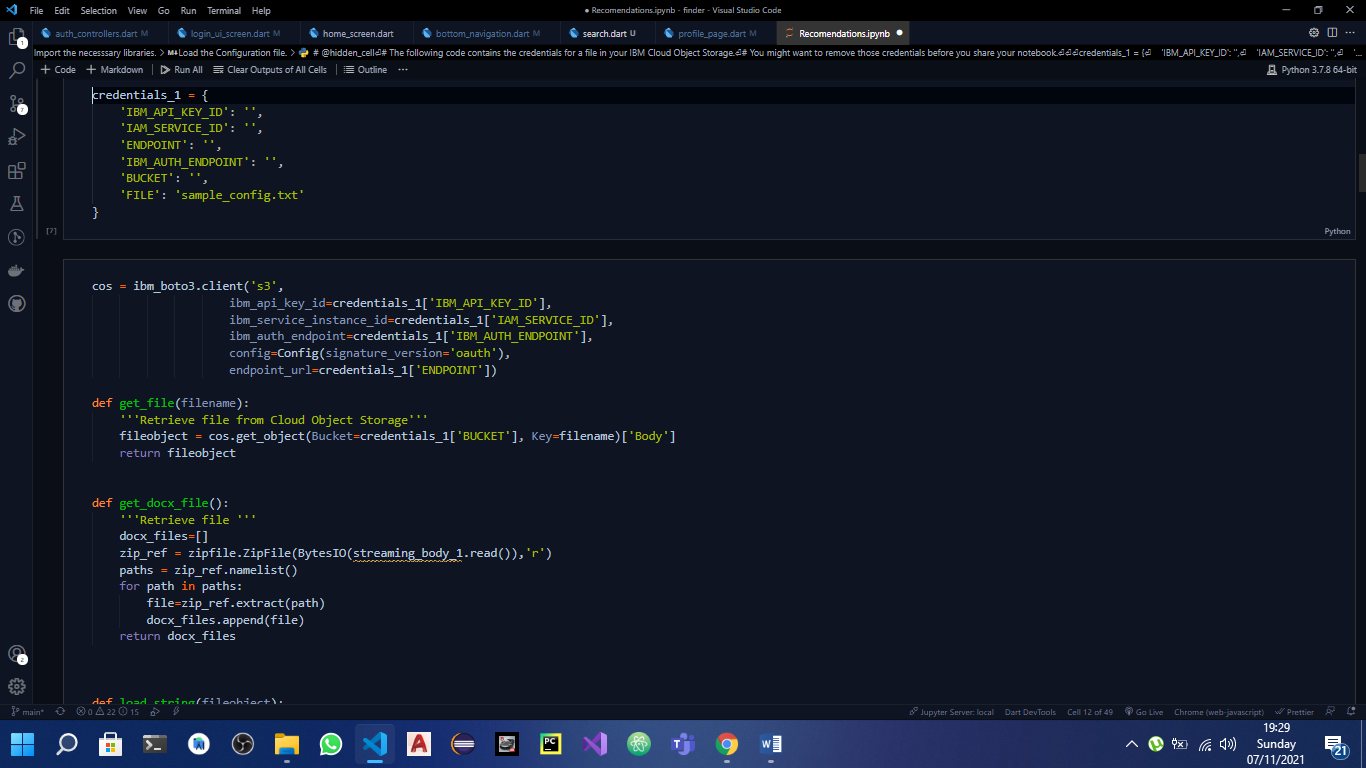
1. Natural Language Understanding.

The next step is to train our system to understand what we need from a CV. This step involves passing in our credentials from Watson Studio using API key in order to access its services and loading all the necessary files for the analysis.

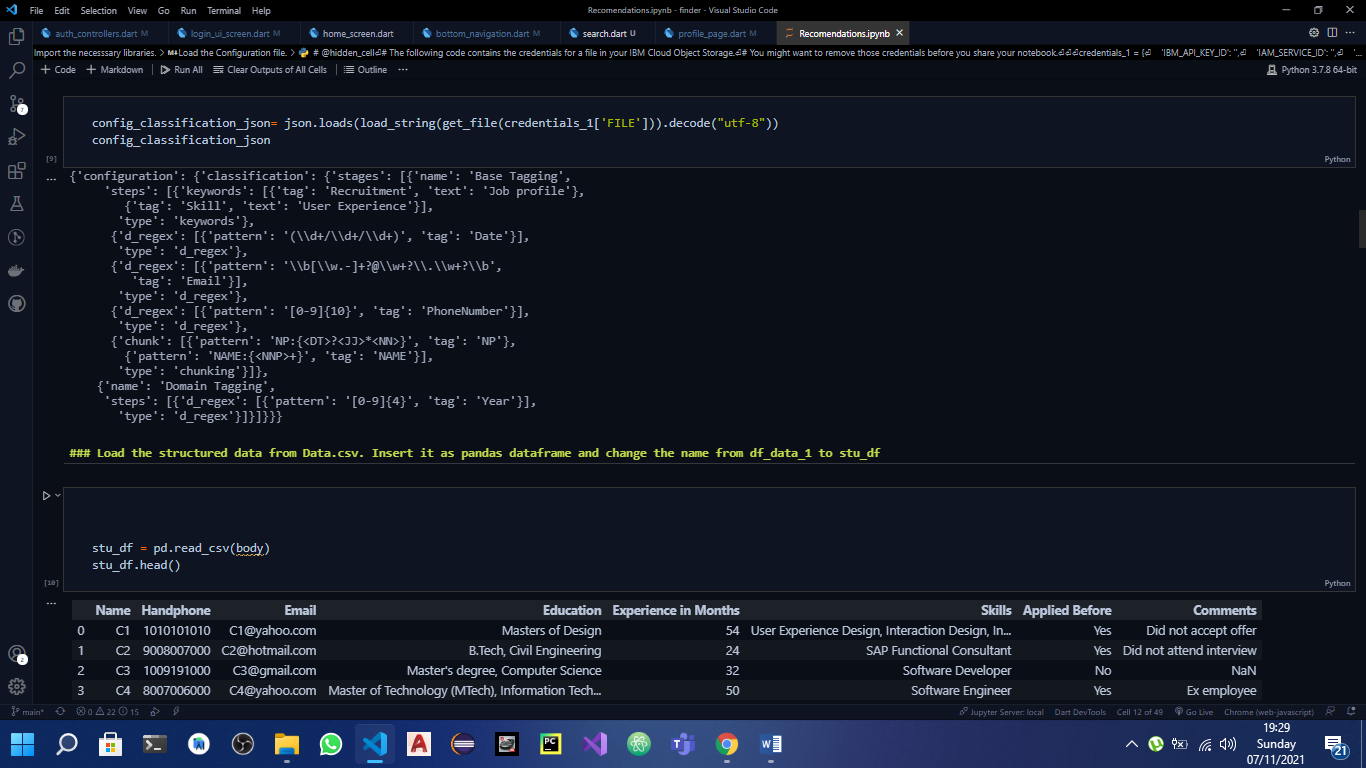
The files include:

1. A configuration file
2. A zip including sample CVs

Passing in credentials, loading files and using functions to read content



Decoding file contents into a json format then passing results to a pandas dataframe

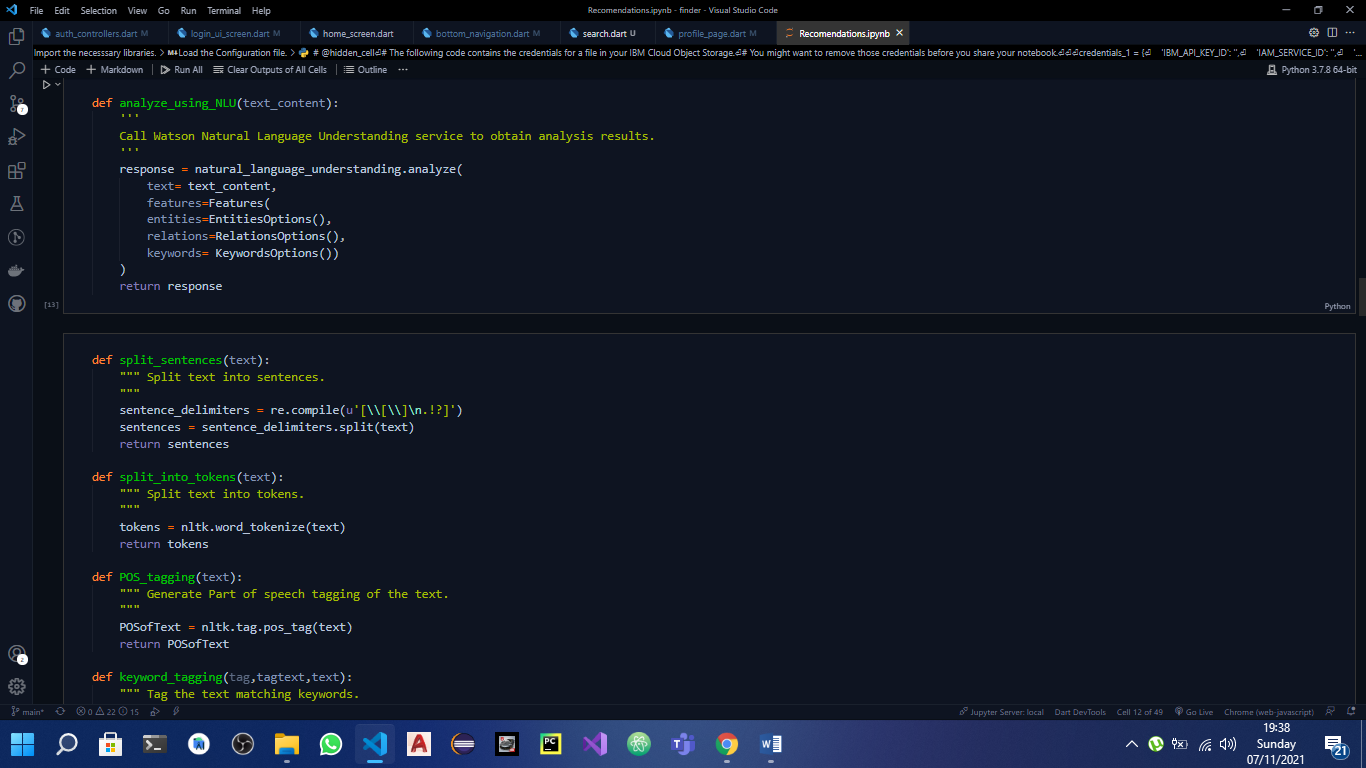


The process’ final result is a table of structured data that can easily be accessed and understood by the system.

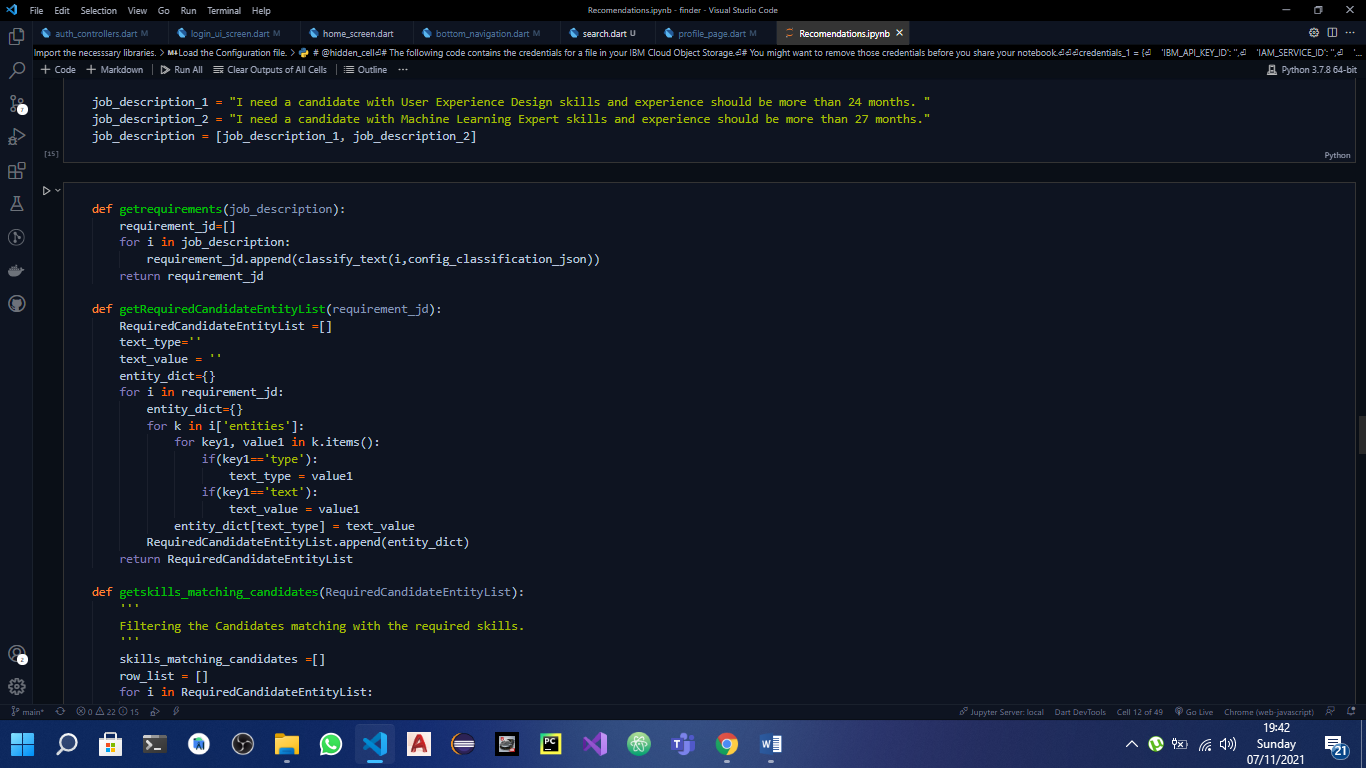
1. Natural Language Processing.

This step includes working on the table of results above to eliminate candidates basing on requirements specified.

Using nltk and Watson studio’s functions to carry analyze the tables and get keywords , entities and relations.

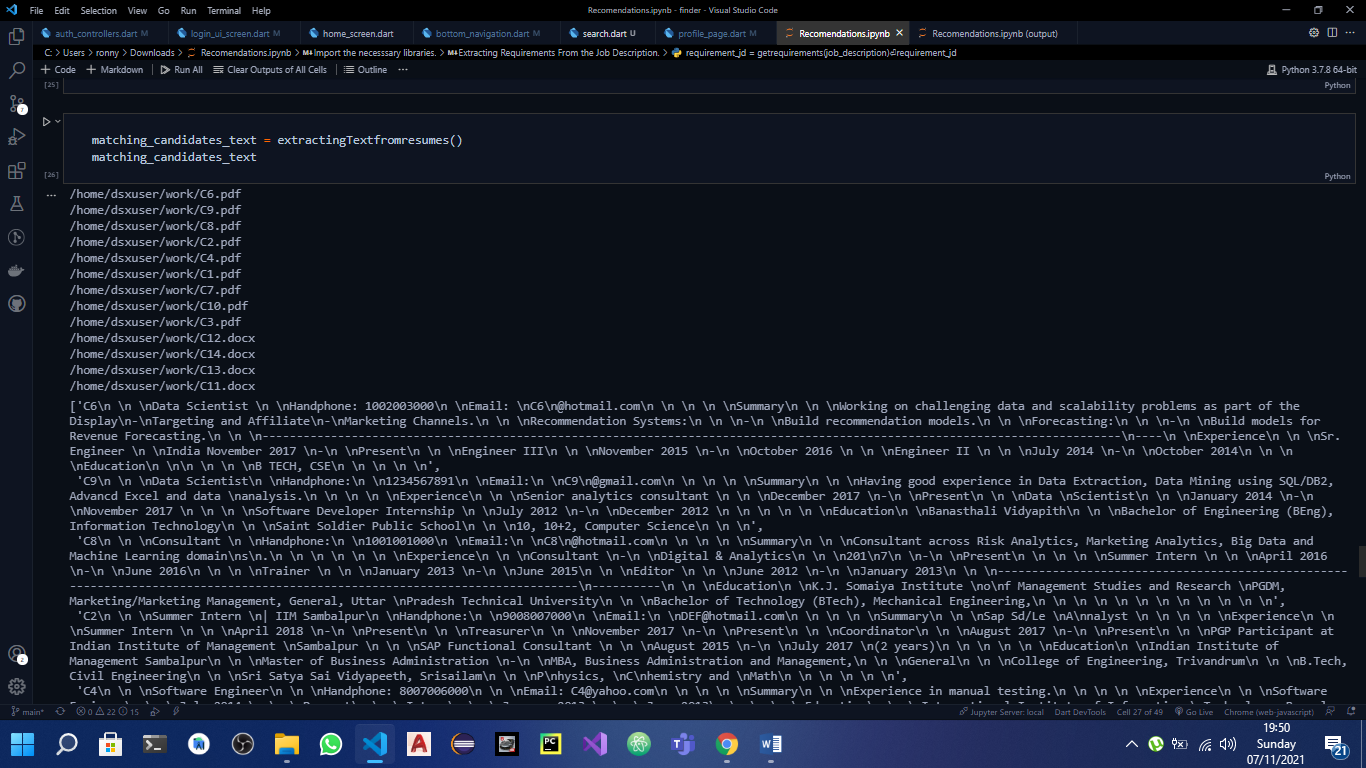


We now had to pass in a job description and extract important words for the comparison



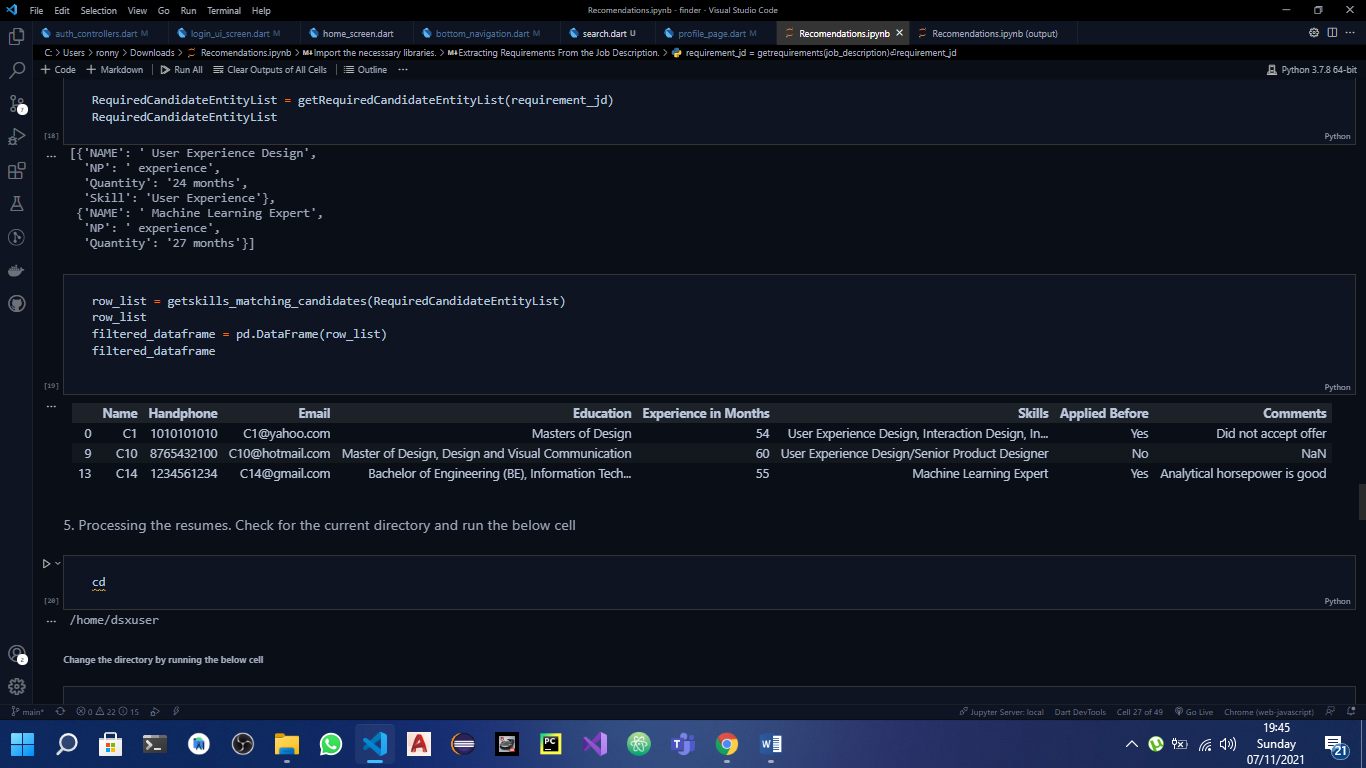
After the processes, we had a json of keywords about a job description.

All CVs read and printed in plaintext from pdf.



Using a created function we compared the job description and the dataframe we had and filtered people who did not have the requirements from the description.

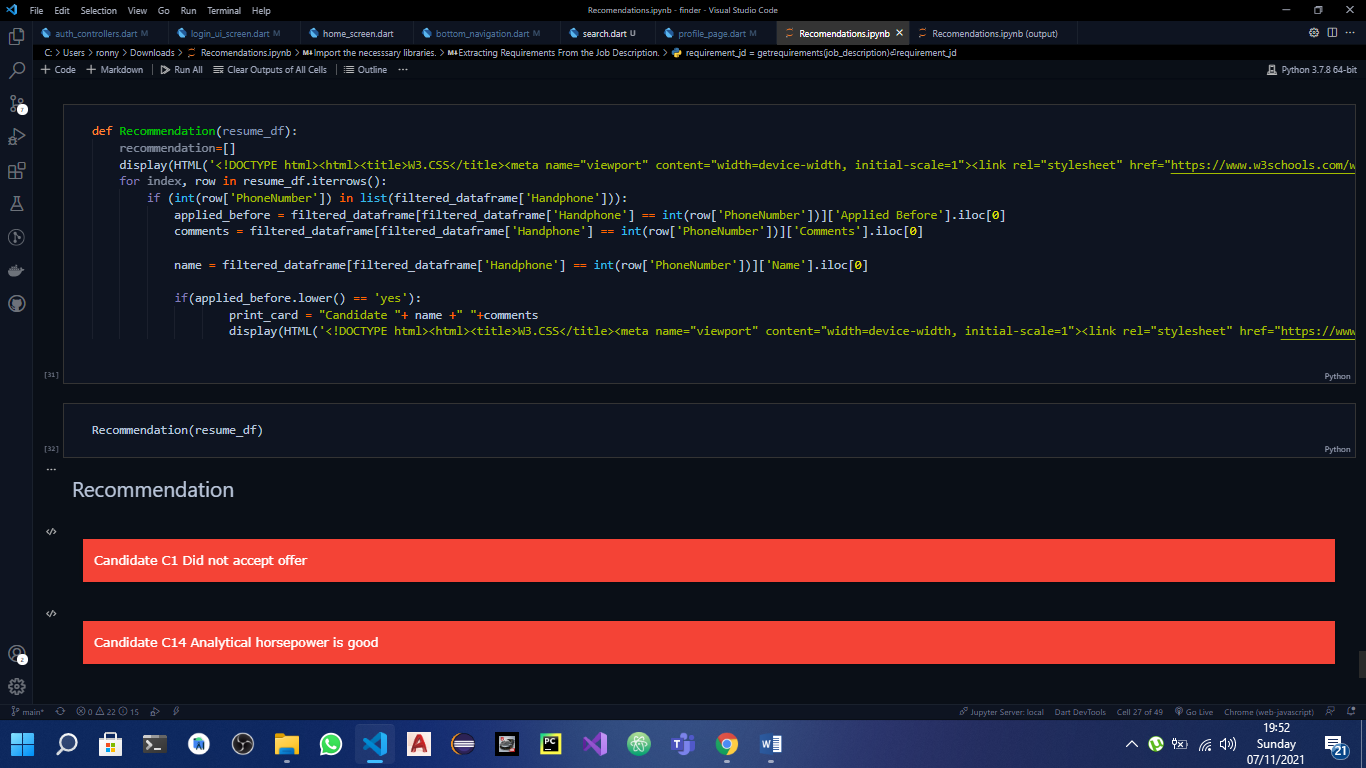
Keywords from job description and filtered candidates from dataframe.



1. Natural Language Generation.

This being the last step dealt only with presenting our final output to the analyst.

Below is the NLG step and how it helped generate user-friendly result.



CONCLUSION

The use of natural language processing helped a lot to analyze a lot of CVs faster and without any bias. IBM’s Watson studio is so helpful in providing storage services and a place to work with colleagues to break down bigger problems to smaller bits of manageable tasks.