Terminal command: python test.py (press enter) ,

type the name of the document(pdf,doc,txt) with the extension. The required json file is obtained in the same directory. (as out.json)

Our code takes .doc, .pdf and .txt. We convert the .doc and .pdf into .txt file.

We take the basic details - name(2nd line), phone number(3rd line) and email address(4th line). Then we take whatever heading comes our way and put the content below the heading. The only constraint is the fact that the content below the heading will be numbered(to differentiate between heading and content).

So our sample Resume looks like:

Basic Details

\*Name\*

\*Phone no\*

\*Email\*

Heading 1

1. Point 1
2. Point 2
3. Point 3

Heading 2

1. Point 4
2. Point 5
3. Point 6

Heading 3

1. Point 7
2. Point 8
3. Point 9

Another way to approach this would be to have a dictionary of key words to recognize the headings. This will remove the necessity of having bullets for the content under the heading. But this dictionary would have to have all possible headings. Our code is good for a general resume. If the question is to parse a specific topic resume- a dictionary of words from that topic would be the approach we would go for.