**Task for Deloitte: Data scientist job:**

**Submitted by Or Cohen Raviv.**

**Task A –attached a pptx file and a py file-see also gitub link.**<https://github.com/OCR3785/OCR>

**Task C**

I conducted a PCA to reduce the dimensionality of the items that showed similarity but did not contribute significantly to the overall explained variance.   
I retained the first four components, named pos 1, neg 1, pos 2, and neg2, which collectively explain nearly 50% of the total variance and have eigenvalues greater than 0.5 (see to the Python file on Github for details).