

**CS 5433: Bigdata Management**  
**Programming Assignment 2**  
**ReadMe for Task5**

**GROUP 4**

---

**Task5 – Identify the principal components**

**Steps to run code:**

1. After logging into Hadoop cluster, create a python file.

File Name: Assign3\_Group4\_Task5.py

Refer to “Group\_4\_Task\_5\_code.pdf”

2. Execute the below command to run the file. After this, it takes some time to get executed.

```
sdarapu@hadoop-nn001:~$ spark-submit /home/sdarapu/Assign3_Group4_Task5.py
```

3. Once, the file gets executed. The output gets displayed on the console.

**Note:** We have given permissions to username “aanmol” for all our input and output file under username “sdarapu”. So that, you can access all my files. If there is any issue with permissions then please let us know.