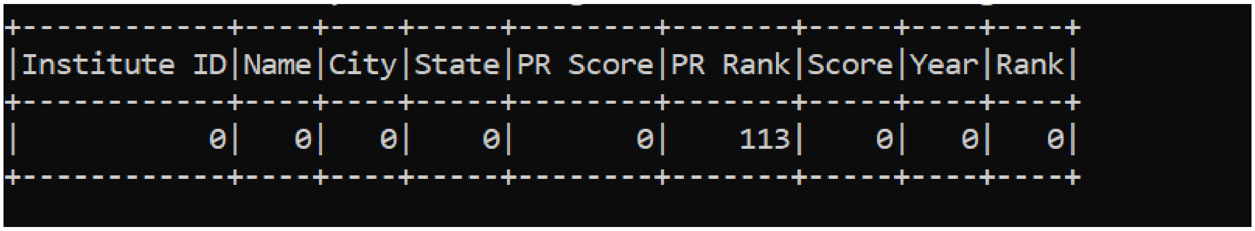
CS 5433: Bigdata Management
Programming Assignment 3
Task 4– Output

Group 4

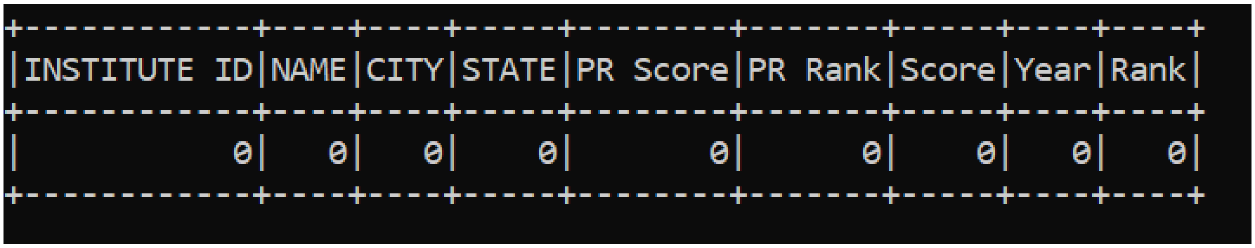
Output results for Task1:

Before performing data correction using cosine similarity using pipeline



```
+------------+----+----+-----+--------+-------+-----+----+----+
|Institute ID|Name|City|State|PR Score|PR Rank|Score|Year|Rank|
+------------+----+----+-----+--------+-------+-----+----+----+
|           0|   0|   0|    0|       0|    113|    0|   0|   0|
+------------+----+----+-----+--------+-------+-----+----+----+
```

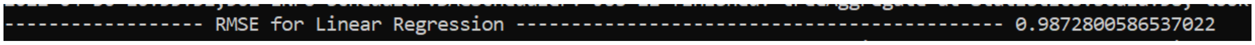
No null values present after performing data correction using cosine similarity



```
+------------+----+----+-----+--------+-------+-----+----+----+
|INSTITUTE ID|NAME|CITY|STATE|PR Score|PR Rank|Score|Year|Rank|
+------------+----+----+-----+--------+-------+-----+----+----+
|           0|   0|   0|    0|       0|      0|    0|   0|   0|
+------------+----+----+-----+--------+-------+-----+----+----+
```

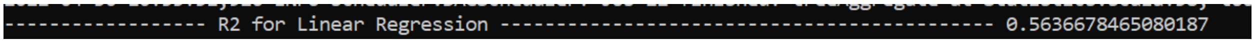
Output results for Task2 & Task3 (Linear Regression):

RMSE Value after using pipeline:



```
----------------- RMSE for Linear Regression ------------------------------------------- 0.9872800586537022
```

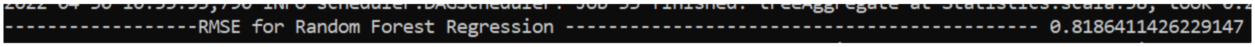
R2 Value after using pipeline:



```
----------------- R2 for Linear Regression ------------------------------------------- 0.5636678465080187
```

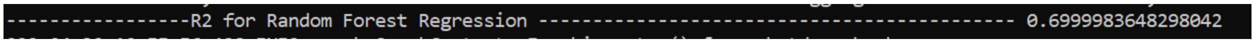
Output results for Task2 & Task3 (Random Forest):

RMSE Value after using pipeline:



```
-----------------RMSE for Random Forest Regression ------------------------------------------- 0.8186411426229147
```

R2 Value after using pipeline:



```
----------------R2 for Random Forest Regression ------------------------------------------- 0.6999983648298042
```