CS 6350.001

ASSIGNMENT 1

Names of students in your group:

Hridya Dhulipala (hxd220008)

Swetha Malaivaiyavur Elayavalli (sxm220052)

Number of free late days used: 0   
Note: You are allowed a **total** of 4 free late days for the **entire semester**. You can use at most 2 for each assignment. After that, there will be a penalty of 10% for each late day.

Please list clearly all the sources/references that you have used in this assignment.

1. <https://stackoverflow.com/questions/46758768/calculating-the-cosine-similarity-between-all-the-rows-of-a-dataframe-in-pyspark>
2. <https://stackoverflow.com/questions/63545176/filter-out-non-digit-values-in-pyspark-rdd>

Assignment Links:

1. <https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/6928907561123411/3344481849106807/1036754708689761/latest.html>
2. <https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/6928907561123411/4247023744017675/1036754708689761/latest.html>