

CS 6350

ASSIGNMENT 1

Names of students in your group:

Sanjeev Penupala (sxp170022)

Gautam Sapre (gss170001)

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.

- Class PySpark Labs
- NLTK docs
- <https://towardsdatascience.com/tf-idf-calculation-using-map-reduce-algorithm-in-pyspark-e89b5758e64c>
- <https://janav.wordpress.com/2013/10/27/tf-idf-and-cosine-similarity/>