**##Libraries to install**

pip install Flask

pip install neo4j

pip install pandas

pip install sklearn

**##Add output\_university.csv to import folder**

**##Create a neo4j database and create nodes**

create database with:

username : "neo4j"

password : "root@12345"

**##Create nodes for university:**

LOAD CSV WITH HEADERS FROM 'file:///output\_university.csv' AS row

CREATE (u:University {

name: row.displayName,

actAvg: toFloat(row['act-avg']),

satAvg: toFloat(row['sat-avg']),

hsGpaAvg: toFloat(row['hs-gpa-avg']),

enrollment: toInteger(row.enrollment),

city: row.city,

zip: row.zip,

acceptanceRate: toFloat(row['acceptance-rate']),

rankingDisplayScore: toFloat(row['rankingDisplayScore']),

percentReceivingAid: toFloat(row['percent-receiving-aid']),

costAfterAid: toFloat(row['cost-after-aid']),

state: row.state,

rankingDisplayRank: toInteger(row['rankingDisplayRank']),

businessRepScore: toFloat(row['businessRepScore']),

tuition: toFloat(row.tuition),

engineeringRepScore: toFloat(row['engineeringRepScore']),

institutionalControl: row.institutionalControl

})

RETURN u;

**## Run Similarity and Graph Creation Query in Neo4j Browser, so it stays in the memory:**

**# Similarity Query:**

MATCH (u1:University), (u2:University)

WHERE id(u1) < id(u2)

WITH u1, u2,

gds.similarity.euclidean(

[u1.actAvg, u1.satAvg, u1.acceptanceRate, u1.hsGpaAvg, u1.rankingDisplayRank, u1.businessRepScore, u1.engineeringRepScore],

[u2.actAvg, u2.satAvg, u2.acceptanceRate, u2.hsGpaAvg, u2.rankingDisplayRank, u2.businessRepScore, u2.engineeringRepScore]

) AS euclideanDistance

MERGE (u1)-[similarity:SIMILARITY\_EDGE]->(u2)

ON CREATE SET similarity.euclideanDistance = euclideanDistance;

**# MyGraph Query:**

CALL gds.graph.exists('myGraph')

YIELD exists AS graphExists

// If the graph does not exist, create it

WITH 'myGraph' AS graphToCreate, graphExists

WHERE NOT graphExists

CALL gds.graph.project(graphToCreate, 'University', 'SIMILARITY\_EDGE', {

nodeProperties: ['actAvg', 'satAvg', 'acceptanceRate', 'hsGpaAvg', 'rankingDisplayRank', 'businessRepScore', 'engineeringRepScore']

}) YIELD graphName, nodeCount, relationshipCount

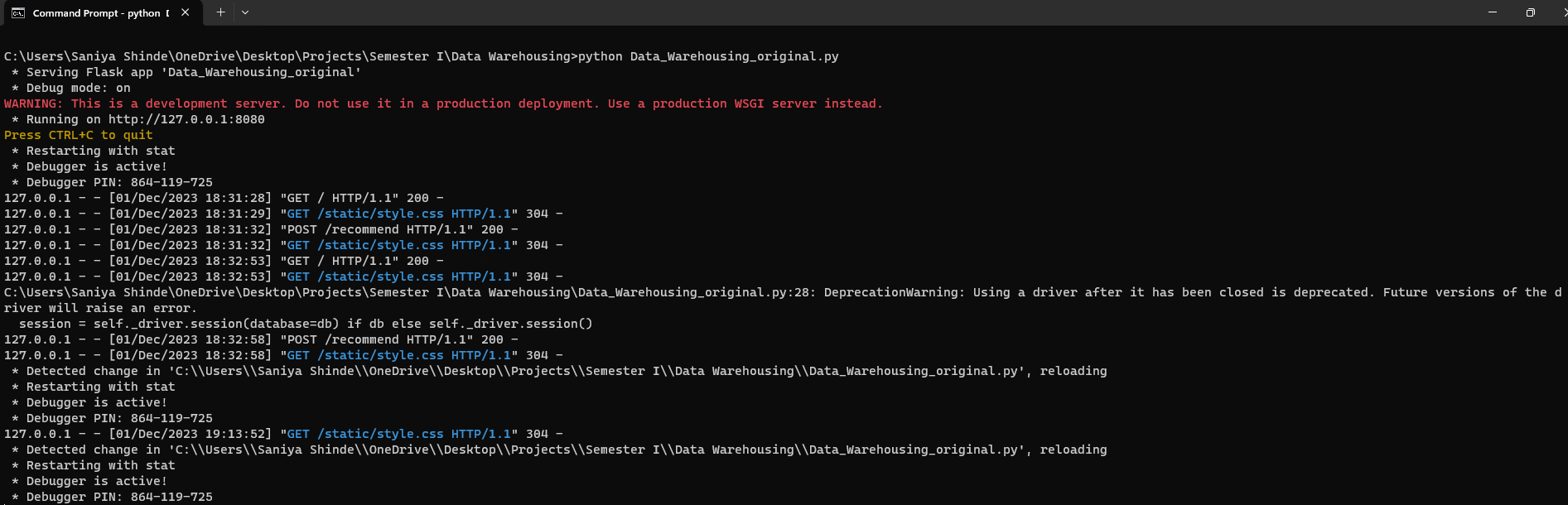
RETURN graphName, nodeCount, relationshipCount;

**## Change the .csv file path in the python file:**

uni = pd.read\_csv('C:/Users/DEVARSH SHETH/Desktop/Data\_Hazim/Project Files/output\_university.csv')

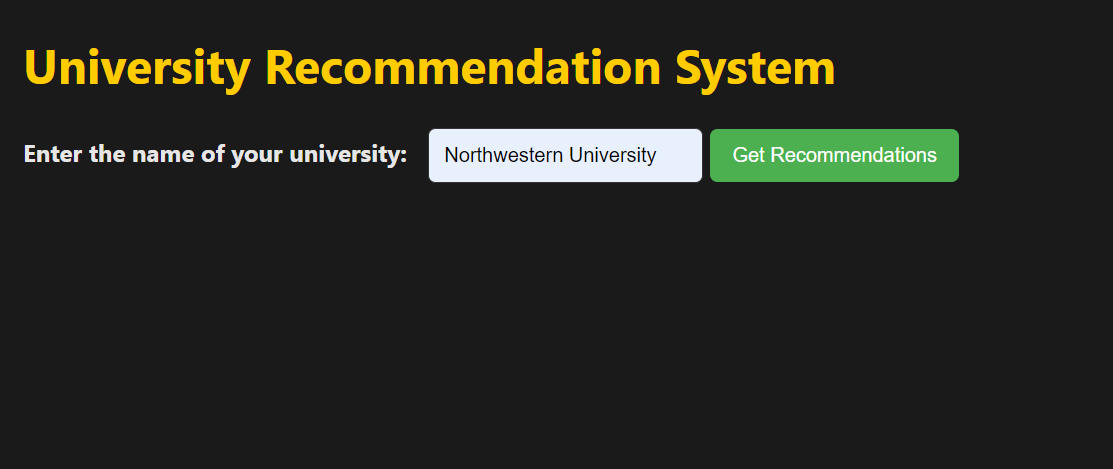
Add your path here, wherever you have stored the file in your local.

**## Run the Python file from the command prompt**



**## Navigate to the URL:** [**http://127.0.0.1:8080**](http://127.0.0.1:8080)

Enter the university name and click on Get Recommendations



The top 5 recommended universities will be displayed:

