AMPBA
Data Collection Individual Assignment – 50 points

PART I - Goal of the assignment: Identify Key Tech Hires in the Hyderabad area – 25 Points
Companies (both international and local) competing for local talent pool.
Our aim is to Identify the key Companies in the market who are hiring in Hyderabad area.
When you have those, we want to know how many positions they have and what are the Key skills they are hiring for.

Your output file may look like this

Name of the company, Number of Positions , Segmentation, Skills they are looking for
===
Amazon, 2300, "Information technology" , "Tableau, R, Python, Java, Salesforce"
Flipkart, 1300, "Information technology" , "Java, API, REST API, Integration, C++, Python"
Microsoft, 8600, "information technology", ".net, Machine Learning, AI, Tableau"

- **You are free to choose any jobs website for crawling purpose (for example https://www.naukri.com)**

- **Select Information technology jobs**

- **Crawl for number of positions in last 6 months**

- **Save the Code – Hyderabad_FirstName_LastName.ipynb**

- **Save the output file in csv format("Hyderabad_FirstName_LastName.csv")**

PART II - Extracting data from API - 25 points

Task:

- Create an API key for NewsApi.org (or find another site where they are serving News)

- Use that API to create queries.

- Understand the API – **Pick a topic**, for example: All articles published by the Wall Street Journal in the last 6 months, sorted by recent first

- Fetch the recent articles published about topic you have selected.

- For each article extract the followings– [source-id , source-name, author, title, description, content]

- Convert the JSON output to data frame with each parameter as a column and each article as a row.

- **Save the code in NewsApi_FirstName_LastName.ipynb**

- **Save the output file in csv format("News_api_FirstName_LastName.csv")**

Deliverables for the assignments

- We need you to submit **two** files on LMS.

- **A word/pdf file which has all the codes/explanations for Turnitin check.**

- **Zipped folder with all notebooks (rmd / ipynb) and their html versions describing code workflows and step-by-step explanations, Documentation for each task detailing your code (in R or Py or ipynb) and results.**

- **Zip ALL files.** Name of the Zip file will be **FULLNAME.zip**

- **Kindly upload both files (A word/pdf file and zipped file) on LMS.**

- Due time is **11.55 PM, August 16th**

- Please ensure you write the **name** and **PGID** clearly at the beginning.

- Honor code for this assignment is **4N**.

- **We need to Execute your code to grade you.**