Bigdata Job Analysis

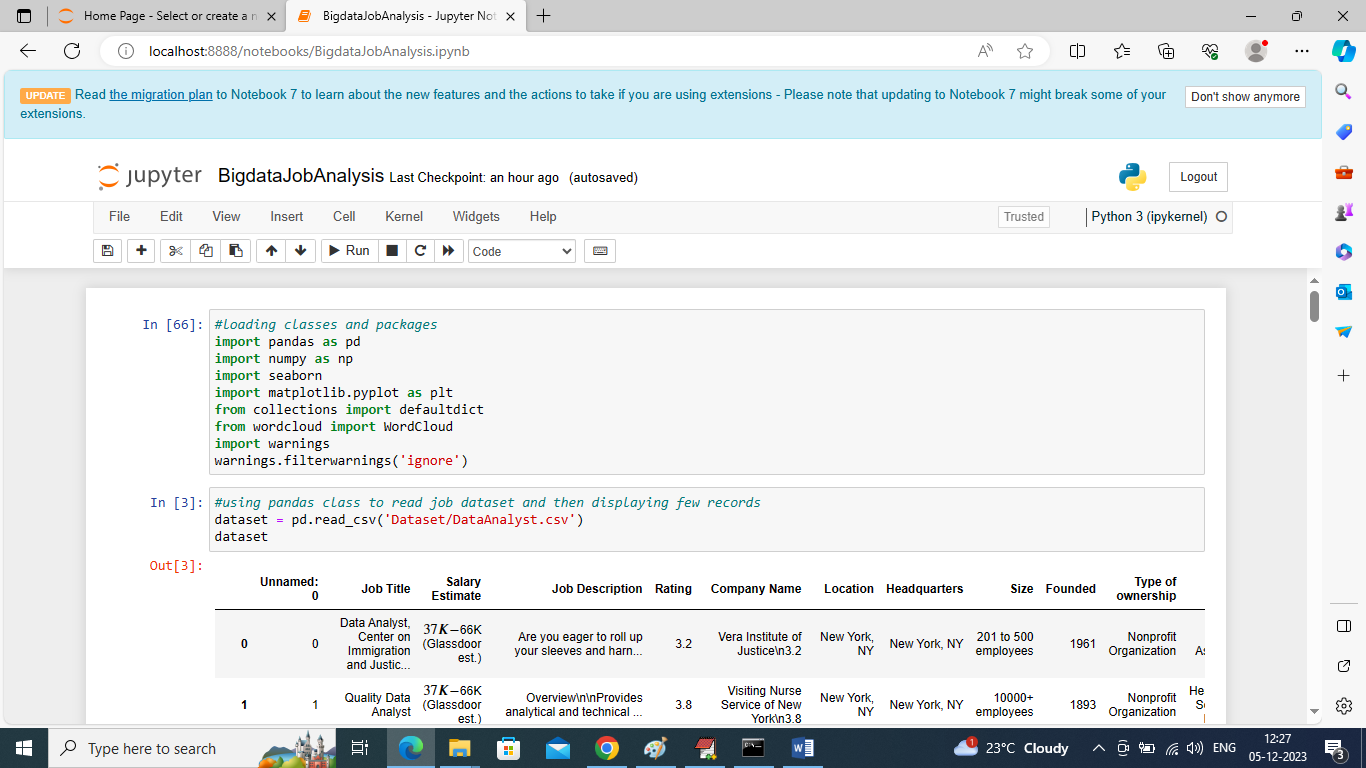
In propose work we are analyzing large amount of online Job posted dataset to find Bigdata family job skills. Since introduction of Bigdata many supporting technologies are introduced and many peoples are unfamiliar about all those technologies and their demands. Selecting suitable Bigdata job family technology can help companies in better project development. Many HR will be unaware of all Bigdata technologies and their demands.

In propose work we are using JOB posting dataset from KAGGLE which can be download from below link

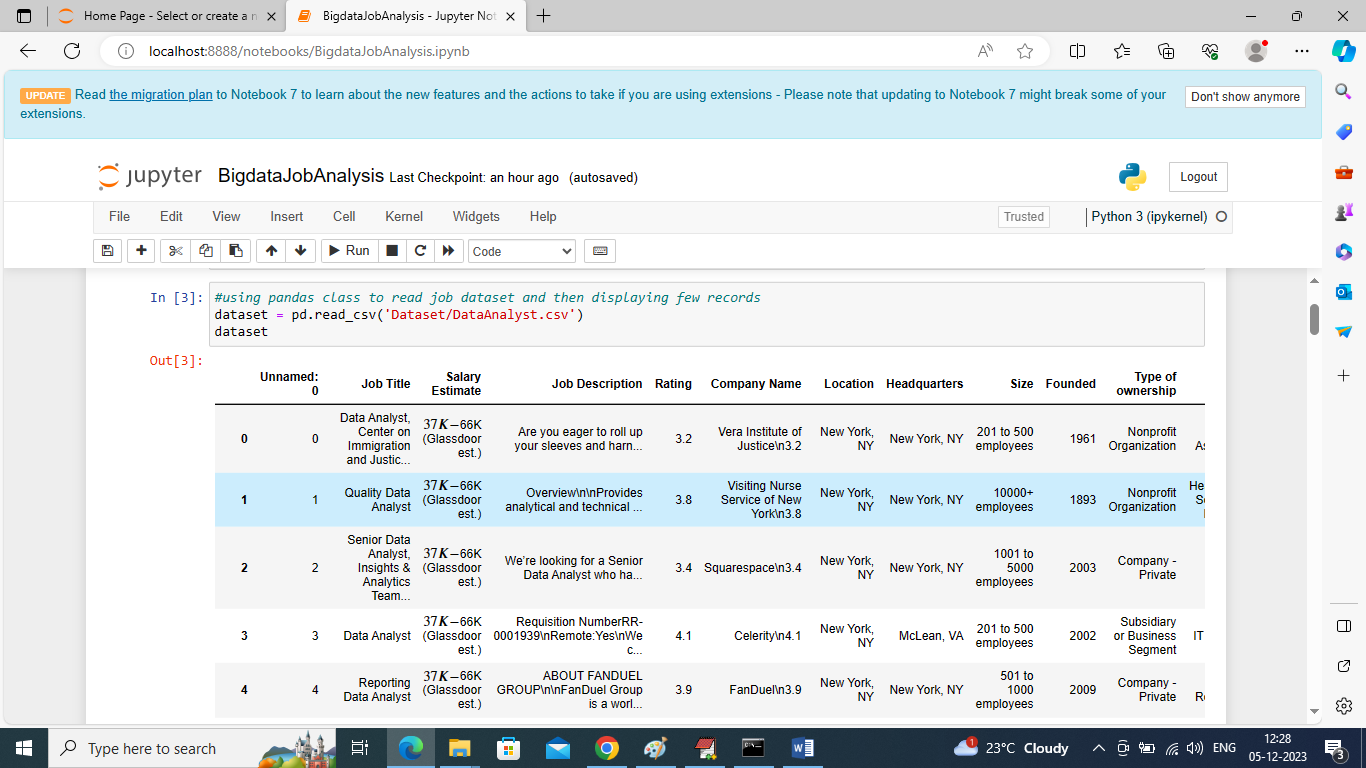
<https://www.kaggle.com/code/rohitsahoo/data-analyst-job-analysis/input>

Above dataset contains job posting from various categories and more than half of the jobs are from Data Analyst. We have done extensive research on all job categories and then find all families of Bigdata technology and then plot graph of all those Bigdata technologies which are high in demand and required most of the companies and by seeing this graph HR can easily understand which family of Bigdata is in high demand

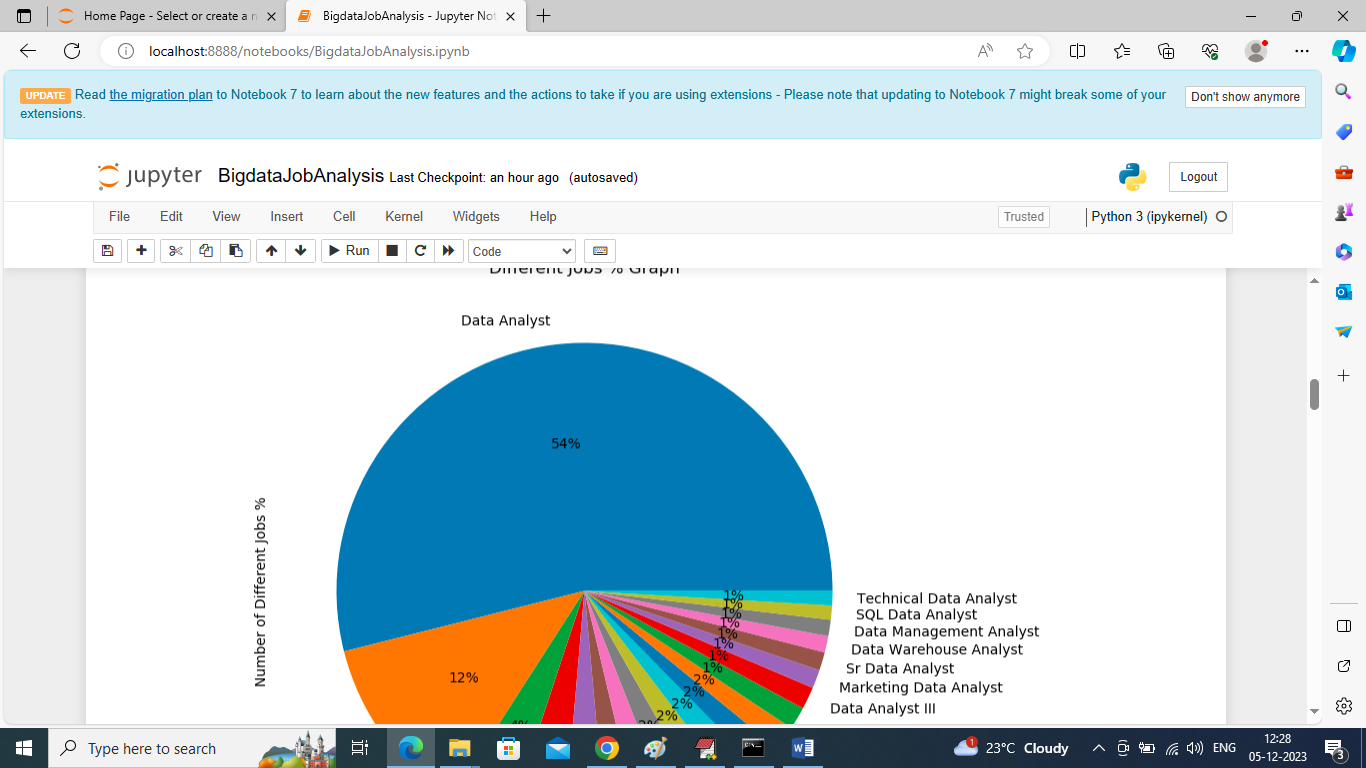
We have coded this project using JUPYTER notebook and below are the code and output screens with blue color comments



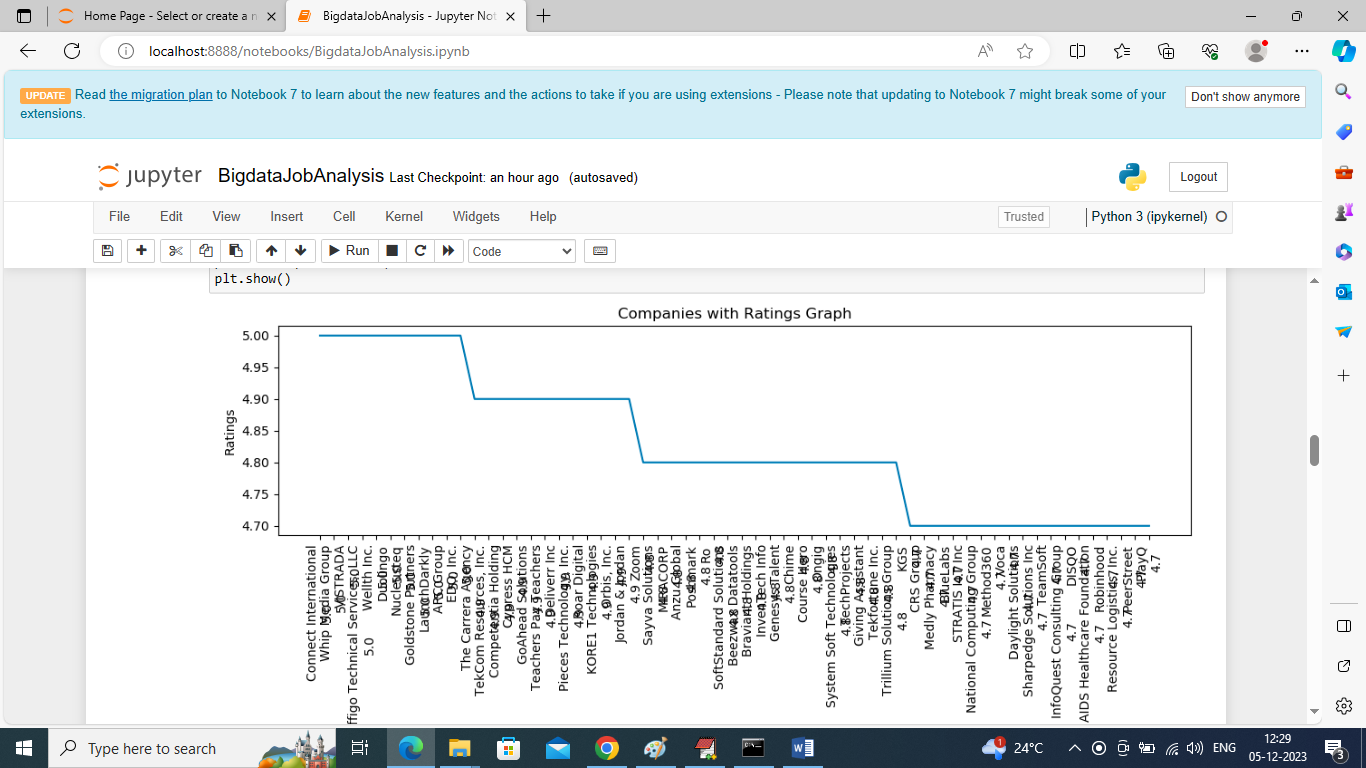
In above screen importing required python classes and packages



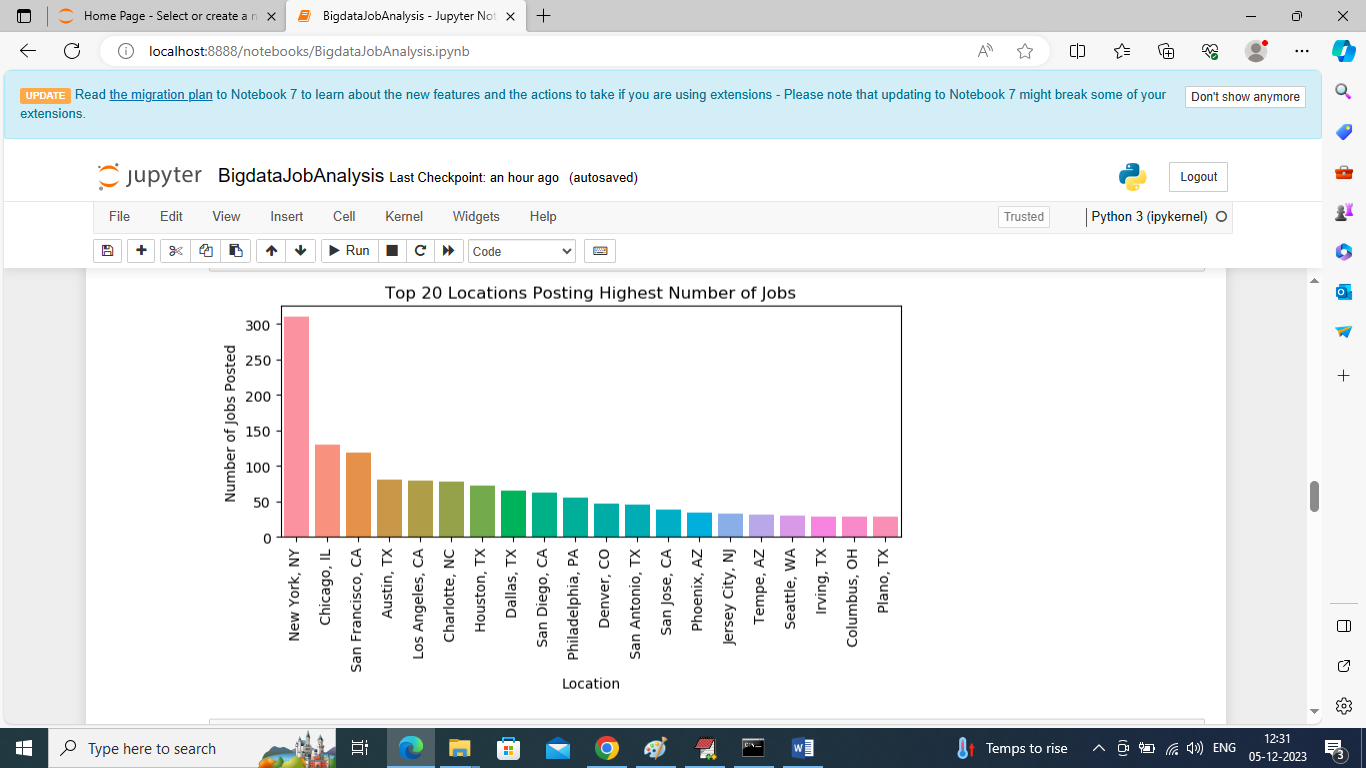
In above screen loading and displaying dataset values



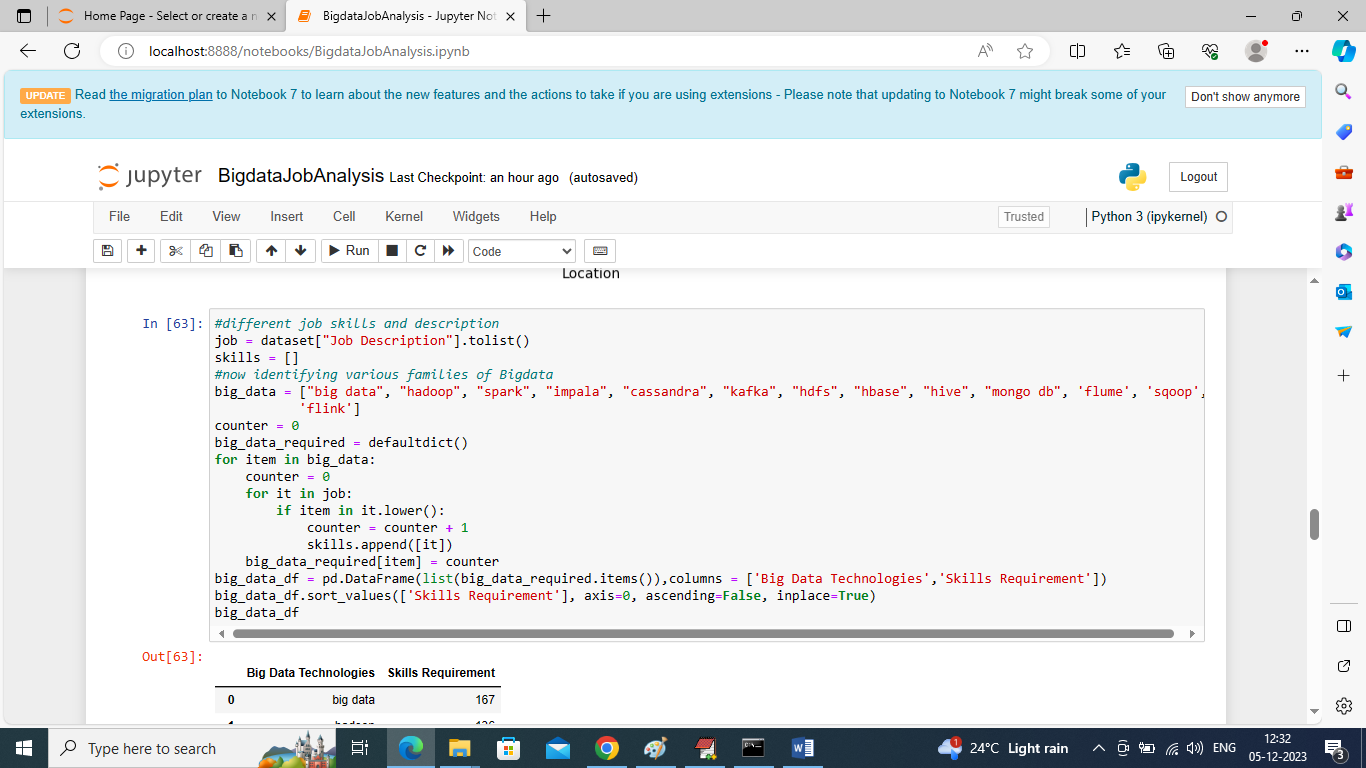
In above graph finding and displaying graph of different job categories in percentage and in all categories we can see ‘Data Analyst’ are more in demands. By seeing above graph HR can know easily above job which are high in demand



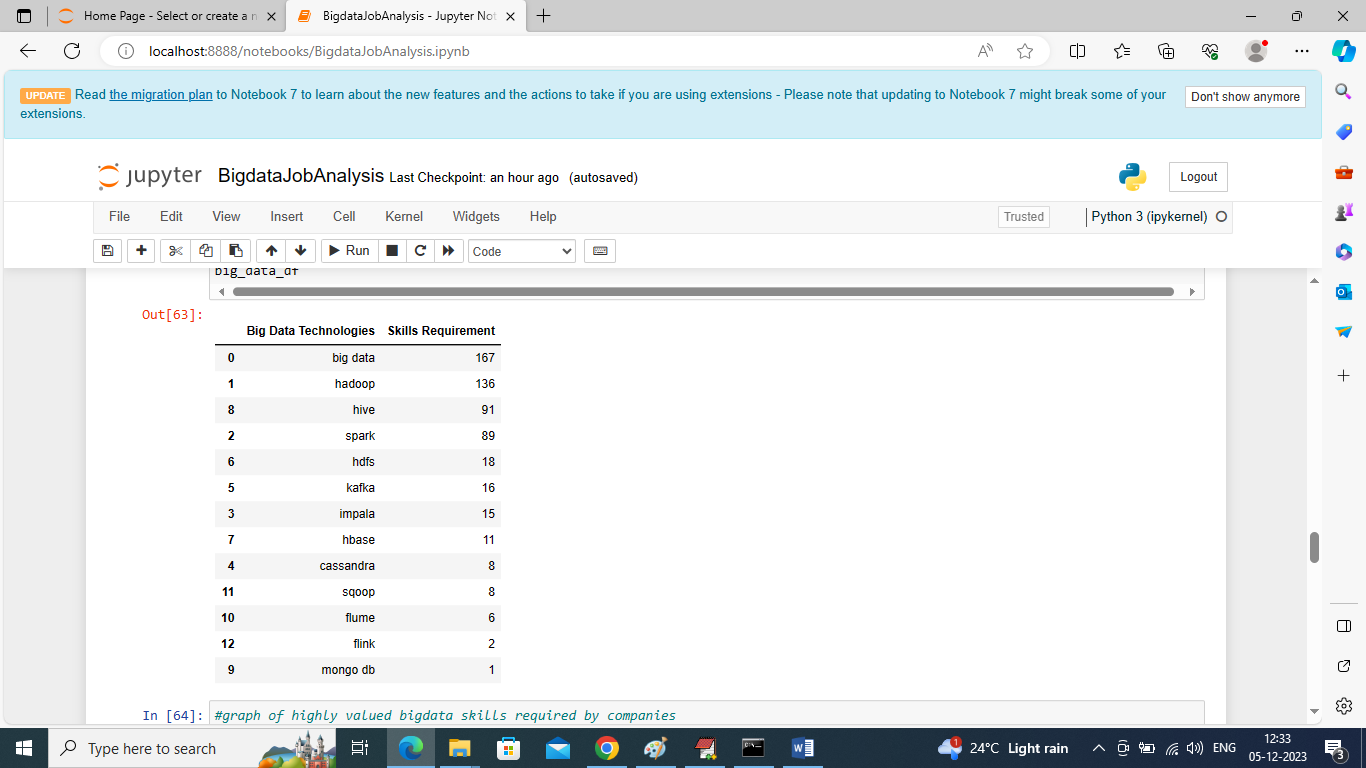
In above graph we are displaying ratings of different companies who have posted jobs and by seeing above graph HR can know this companies are genuine and posting real jobs. In above graph x-axis represents Company Names and y-axis represents Ratings



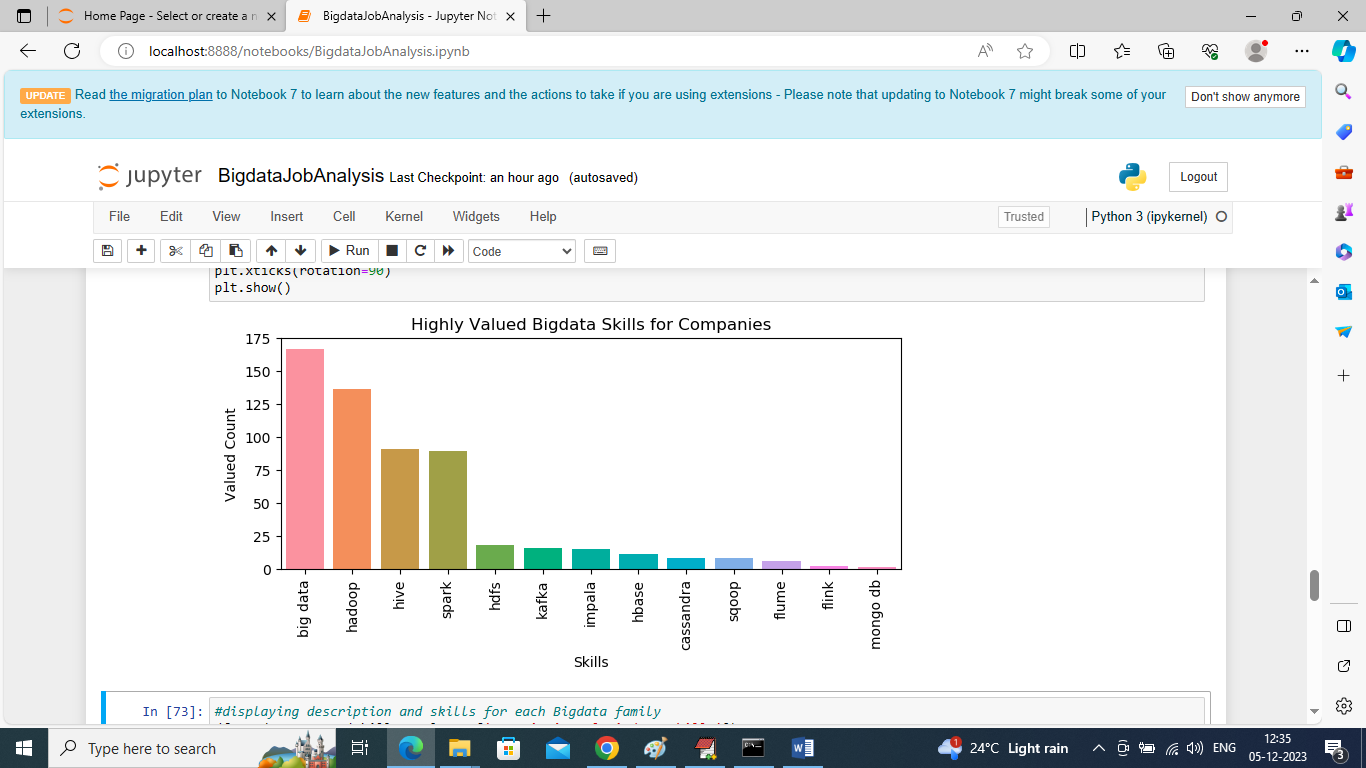
In above screen finding and displaying graph of top 20 locations who are posting more number of Jobs



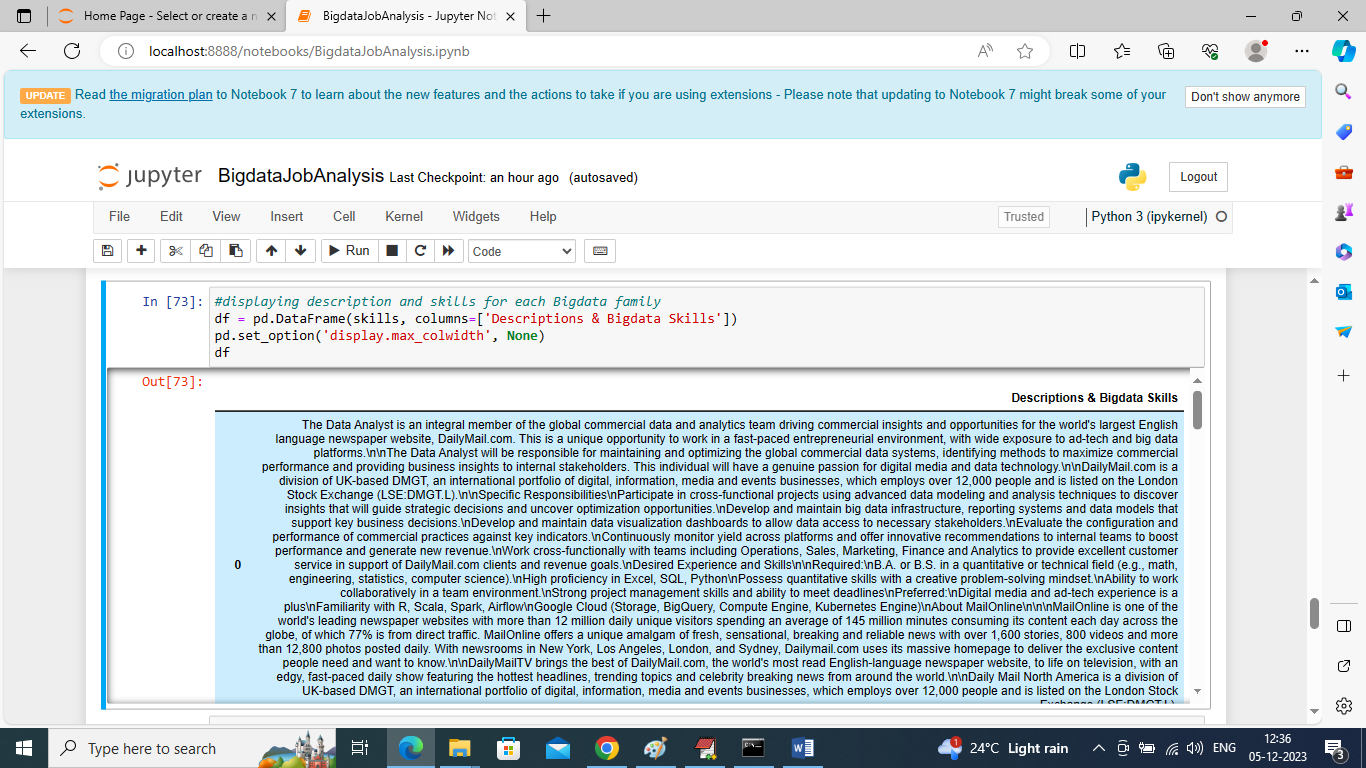
In above screen writing code to find number of jobs posted in each Bigdata family to identify its demand and valued for company



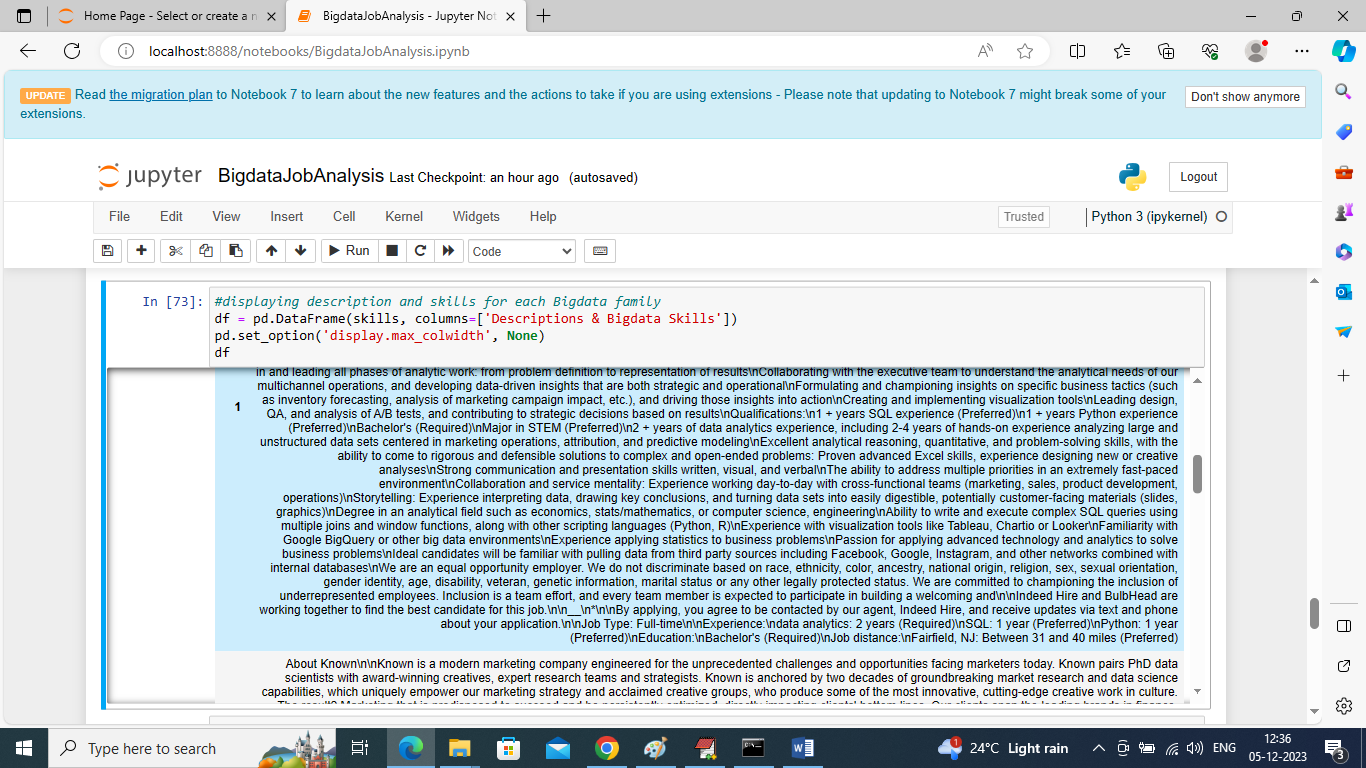
In above screen fetching and categorizing only Bigdata family technologies and their skills demands and in above table Bigdata, Hadoop, Spark, Hive libraries are in more demand



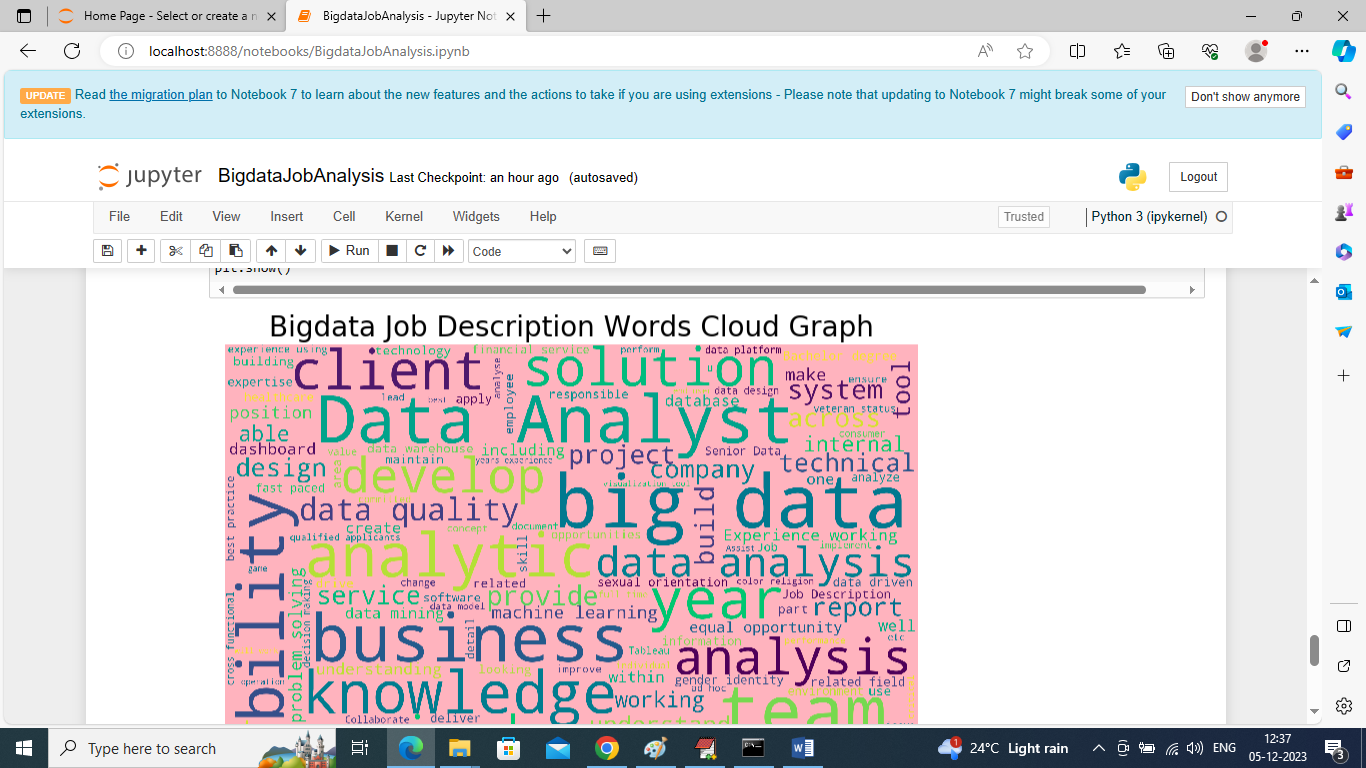
In above screen plotting graph of each Bigdata category where x-axis represents Bigdata technology names and y-axis represents requirements of Jobs for that technology



In above screen displaying JOB description for each Bigdata family job requirements



From above description HR can easily understand about candidates to select or he can write his own company job requirement by seeing above description



In above screen displaying graph of technologies word cloud and this graph will display “all words” in “bold” format which used many number of times in all Job description and from above graph we can see ‘Data Analyst, Big data, analysis’ are used many times. So above technologies are more in demands