**BIGDATA JOB ANALYSIS**

**Objective:**

Nowadays, companies are starting to realize the importance of data availability in large amounts in order to make the right decisions and support their strategies. With the development of new technologies, the Internet and social networks, the production of digital data is constantly growing. The term "Big Data" refers to the heterogeneous mass of digital data produced by companies and individuals whose characteristics (large volume, different forms, speed of processing ) require specific and increasingly sophisticated computer storage and analysis tools. Here intends to define the concept of Big Data, its concepts, challenges and applications, as well as the importance of Big Data Analytics.

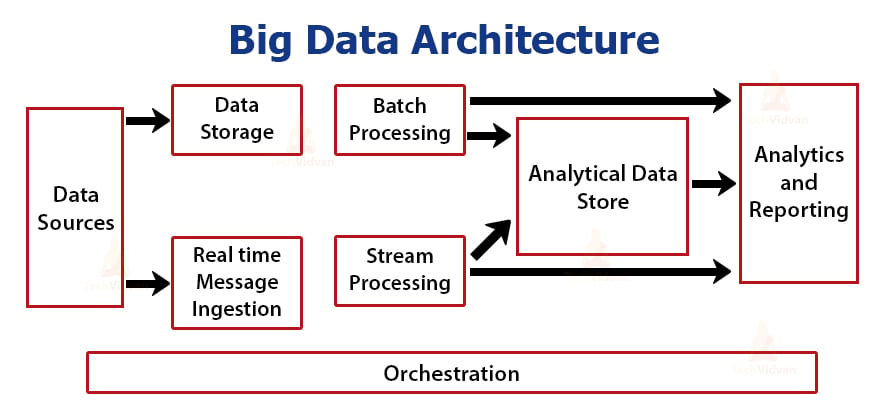
**Introduction:**

The digital data produced is partly the result of the use of devices connected to the Internet. Thus, smartphones, tablets and computers transmit data about their users. Connected smart objects convey information about consumer‟s use of everyday objects. Apart from the connected devices, data come from a wide range of sources: demographic data, climate data, scientific and medical data, energy consumption data, etc. All these data provide information about the location of users of the devices, their travel, their interests, their consumption habits, their leisure activities, and their projects and so on. But also information on how the infrastructure, machinery and apparatus are used. With the ever-increasing number of Internet and mobile phone users, the volume of digital data is growing rapidly. Today we are living in an Informational Society and we are moving towards a Knowledge Based Society. In order to extract better knowledge we need a bigger amount of data. The Society of Information is a society where information plays a major role in the economical, cultural and political stage. [1]

WHAT IS BIG DATA ?

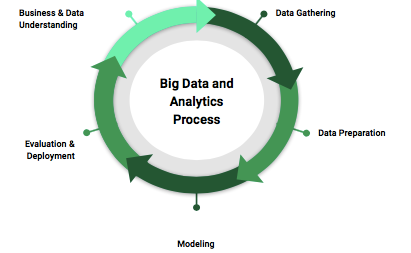
A. Definition

The term "Big Data" refers to the evolution and use of technologies that provide the right user at the right time with the right information from a mass of data that has been growing exponentially for a long time in our society. The challenge is not only to deal with rapidly increasing volumes of data but also the difficulty of managing increasingly heterogeneous formats as well as increasingly complex and interconnected data. Being a complex polymorphic object, its definition varies according to the communities that are interested in it as a user or provider of services. Invented by the giants of the web, the Big Data presents itself as a solution designed to provide everyone a real-time access to giant databases. Big Data is a very difficult concept to define precisely, since the very notion of big in terms of volume of data varies from one area to another. It is not defined by a set of technologies, on the contrary, it defines a category of techniques and technologies. This is an emerging field, and as we seek to learn how to implement this new paradigm and harness the value, the definition is changing. [2].



WHAT IS BIG DATA ANALYTICS ?

Big Data generally refers to data that exceeds the typical storage, processing, and computing capacity of conventional databases and data analysis techniques. As a resource, Big Data requires tools and methods that can be applied to analyze and extract patterns from large-scale data. [3] The analysis of structured data evolves due to the variety and velocity of the data manipulated. Therefore, it is no longer enough to analyze data and produce reports, the wide variety of data means that the systems in place must be capable of assisting in the analysis of data. The analysis consists of automatically determining, within a variety of rapidly changing data, the correlations between the data in order to help in the exploitation of it.



Big Data Anlytics refers to the process of collecting, organizing, analyzing large data sets to discover different patterns and other useful information. Big data analytics is a set of technologies and techniques that require new forms of integration to disclose large hidden values from large datasets that are different from the usual ones, more complex, and of a large enormous scale. It mainly focuses on solving new problems or old problems in better and effective ways. [4]

Types of Big Data Analytics

a) Descriptive Analytics It consists of asking the question: What is happening? It is a preliminary stage of data processing that creates a set of historical data. Data mining methods organize data and help uncover patterns that offer insight. Descriptive analytics provides future probabilities and trends and gives an idea about what might happen in the future.

b) Diagnostic Analytics It consists of asking the question: Why did it happen? Diagnostic analytics looks for the root cause of a problem. It is used to determine why something happened. This type attempts to find and understand the causes of events and behaviors.

c) Predictive Analytics It consists of asking the question: What is likely to happen? It uses past data in order to predict the future. It is all about forecasting. Predictive analytics uses many techniques like data mining and artificial intelligence to analyze current data and make scenarios of what might happen.

d) Prescriptive Analytics It consists of asking the question: What should be done? It is dedicated to finding the right action to be taken. Descriptive analytics provides a historical data, and predictive analytics helps forecast what might happen. Prescriptive analytics uses these parameters to find the best solution.

**Related work:**

**Research on the Innovation of E-business Talents Training Mode Underthe Background of Big Data (2018):**

The rapid development and continuous transformation of big data has imposed higher requirements for the e-commerce professional's data sorting and network technology applications. They provided more favorable conditions and opportunities for the development of e-commerce professionals. Universities and colleges have becoming the major exporter of talent training, and they must follow the direction of information reform and cultivate e-commerce composite innovative talents that are more in line with the requirements of the development of the times. This article combines the new requirements of e-commerce professionals and the problems in the training of e-commerce professionals in the context of big data, from cultivating applied talents, optimizing professional curriculum, strengthening cooperation between schools and enterprises, and strengthening the construction of teachers. Other aspects also propose suggestions for e-business talents training mode under the background of big data.[1].

**Practice on the Sustainable Development of Talent Cultivation Mode in the Context of BigData (2019):**

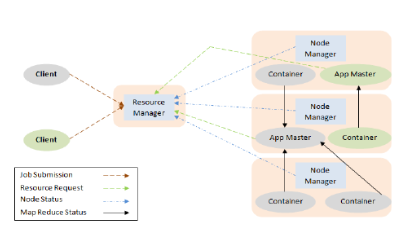
Nowadays, emerging majors including big data science and big data technology, artificial intelligence, robotics engineering, new engineering, and new media technologies are all full of digital media majors. Meanwhile, there are interdisciplinary intersections between different majors, so how to enable digital media professional to continue to balance development in new and old majors? This paper mainly analyzes and summarizes the multi-dimensional and multi-angle of the digital media technology professional talent training mode, enterprise talent demand, teachers and students in domestic universities. This laid a good foundation for the reform of the talent training of Heilongjiang International University’s digital media technology and the practical exploration of the UI direction in the digital media profession. It is necessary to implement the new policy and actively explore a high-quality, professional, practical, and applied training model that focuses on students and adapts to industries and industries.

**Big Data for Development: A Review of Promises and Challenges (2016):**

The article uses a conceptual framework to review empirical evidence and some 180 articles related to the opportunities and threats of Big Data Analytics for international development. The advent of Big Data delivers a cost-effective prospect for improved decision-making in critical development areas such as healthcare, economic productivity and security. At the same time, the well-known caveats of the Big Data debate, such as privacy concerns and human resource scarcity, are aggravated in developing countries by long-standing structural shortages in the areas of infrastructure, economic resources and institutions. The result is a new kind of digital divide: a divide in the use of data-based knowledge to inform intelligent decision-making. The article systematically reviews several available policy options in terms of fostering opportunities and minimising risks.

**Methodology:**

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.



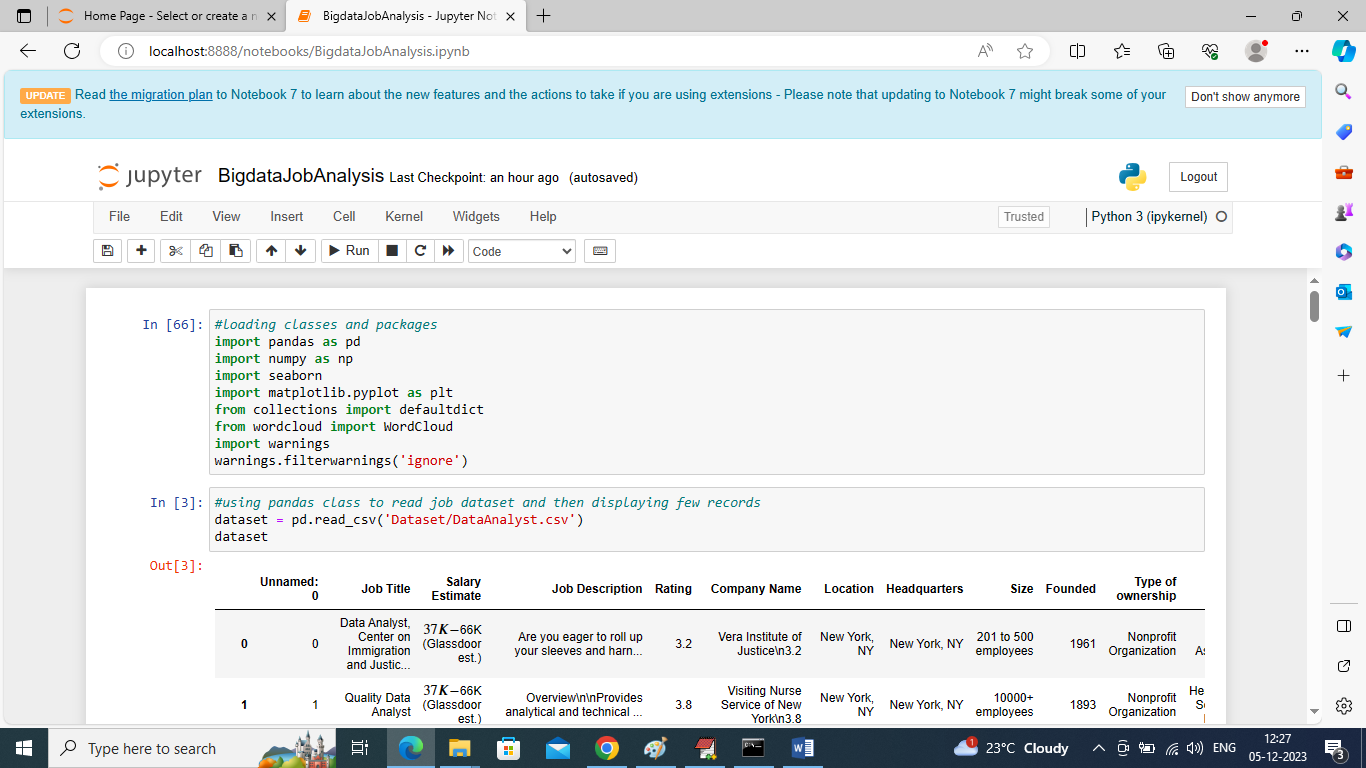
In propose work we are using JOB posting dataset from KAGGLE which can be download from below link

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

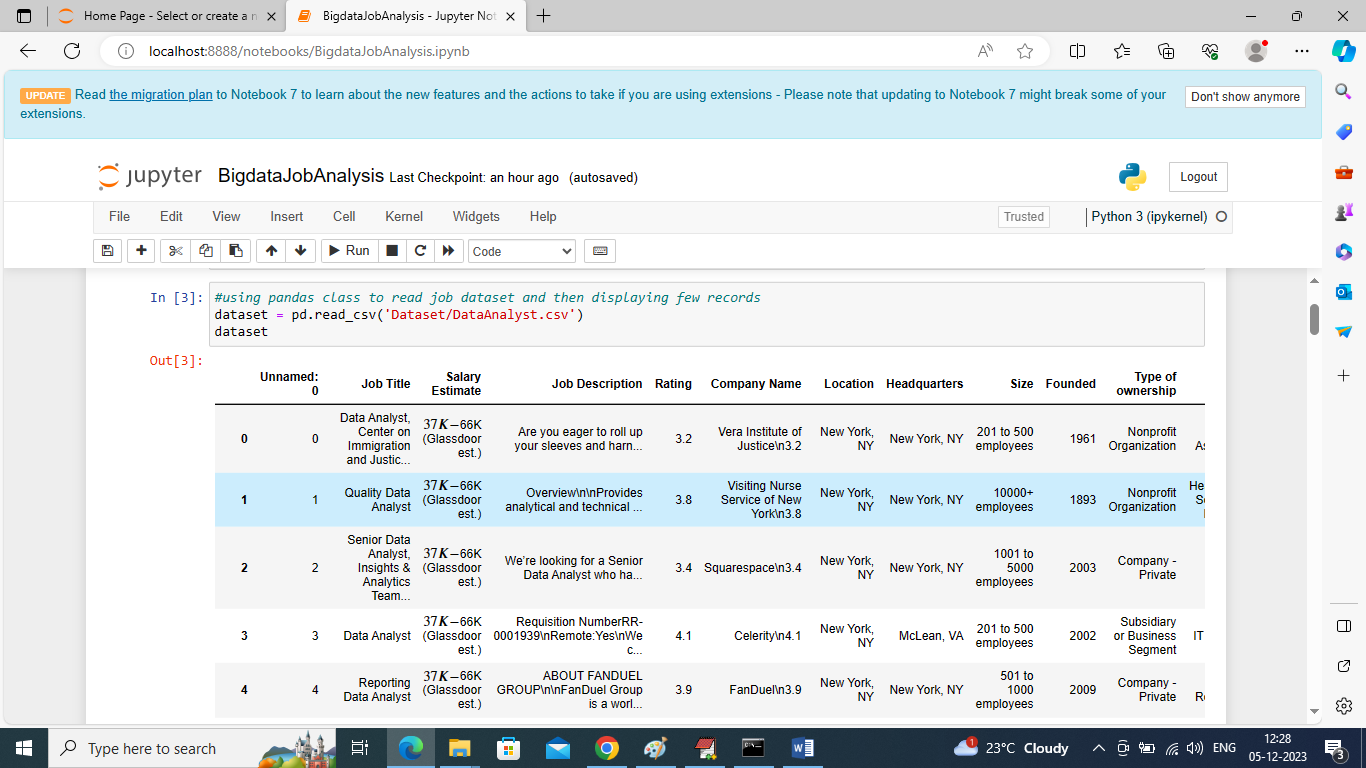
Above dataset contains job posting from various categories and more than half of the jobs are from Data Analyst. We have done extensive research on all job categories and then find all families of Bigdata technology and then plot graph of all those Bigdata technologies which are high in demand and required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

**Result discussion:**

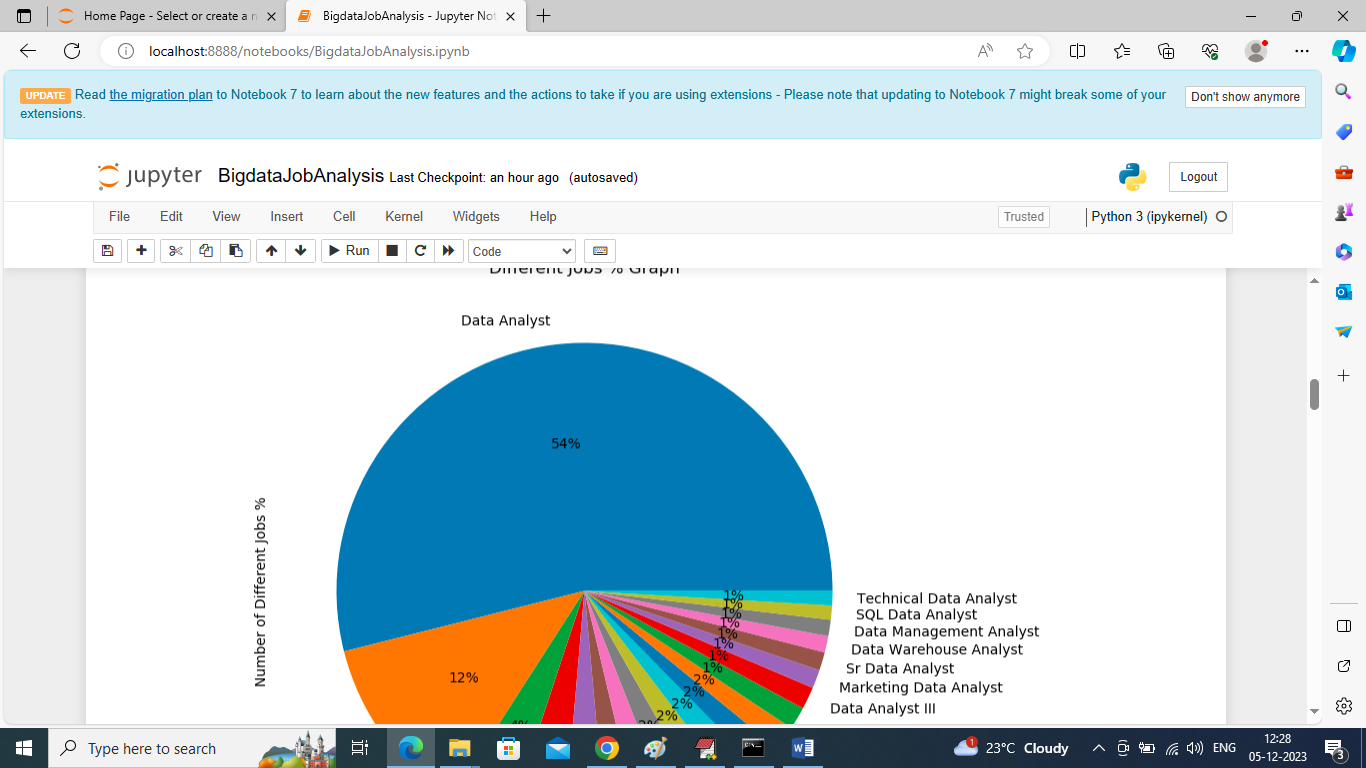
We have coded this project using JUPYTER notebook and below are the code and output screens with blue color comments



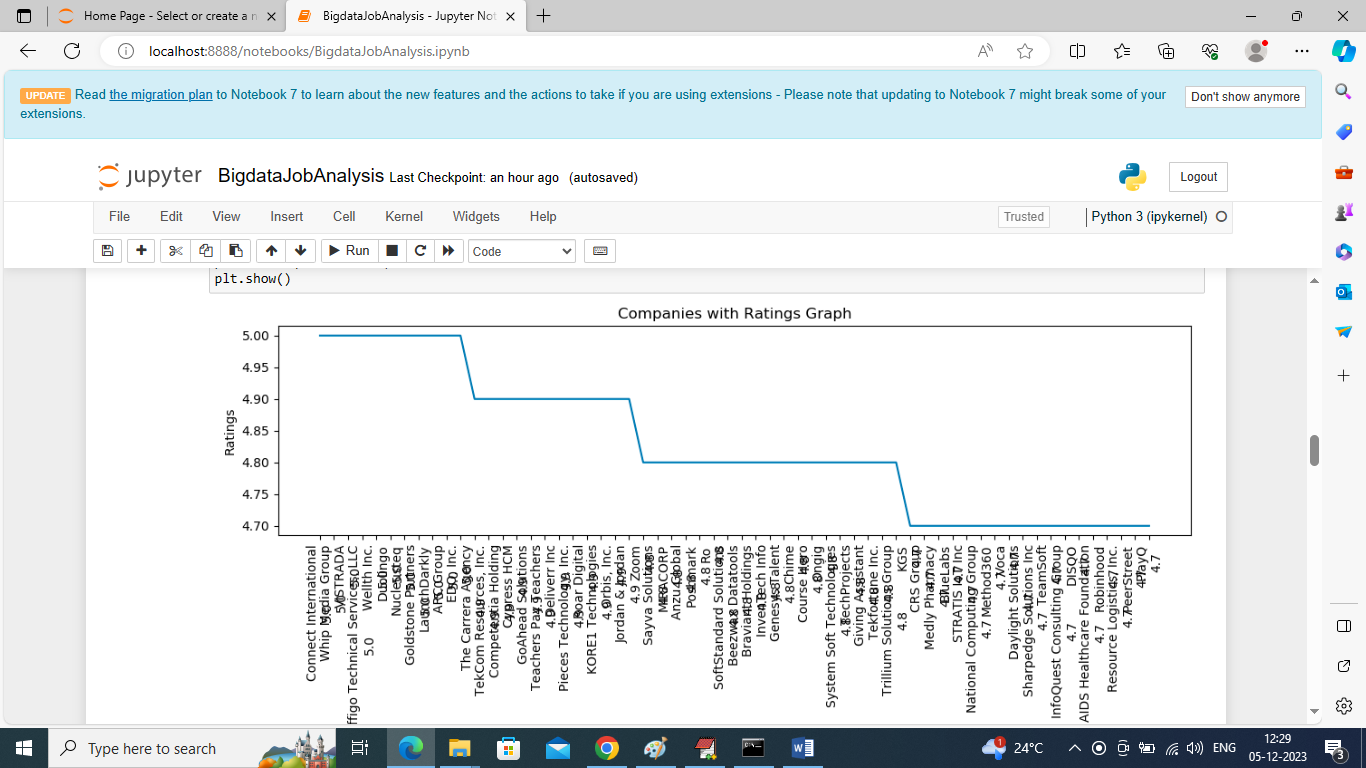
In above screen importing required python classes and packages



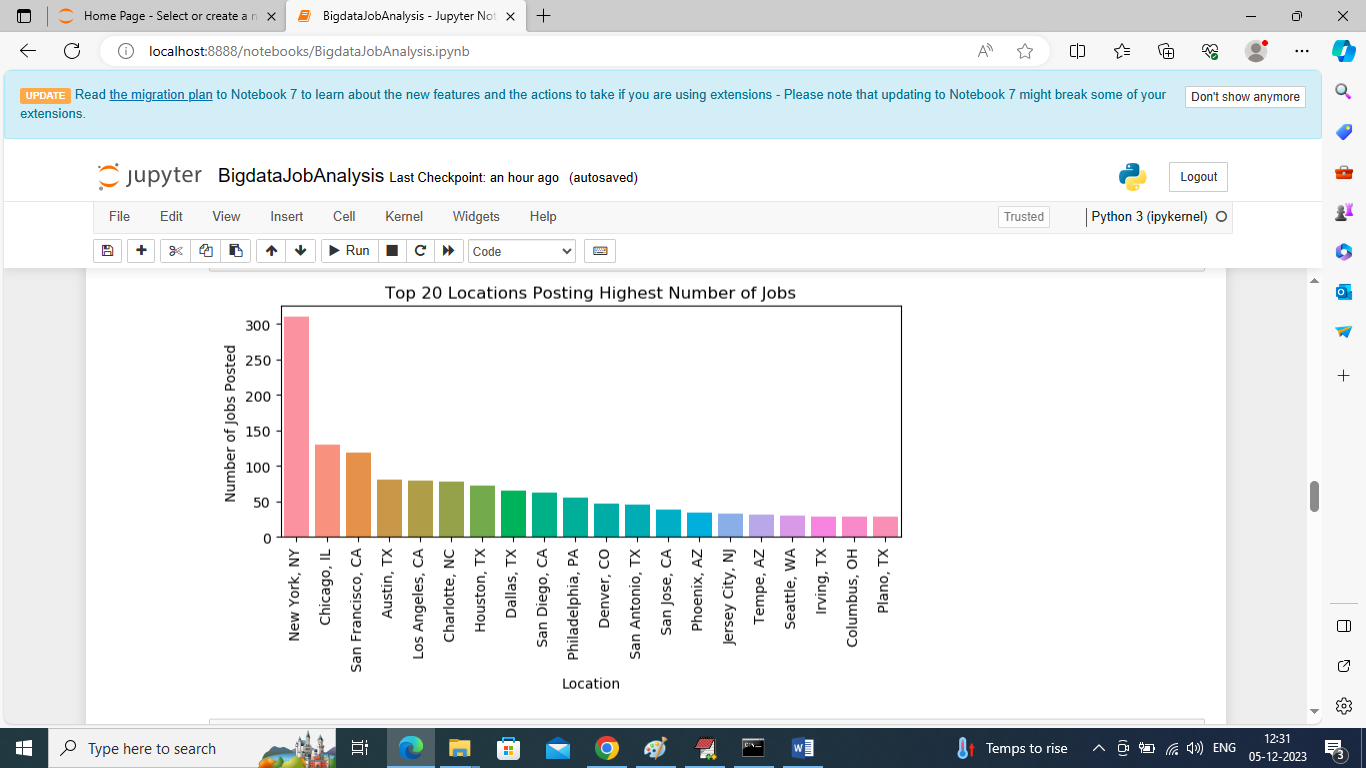
In above screen loading and displaying dataset values



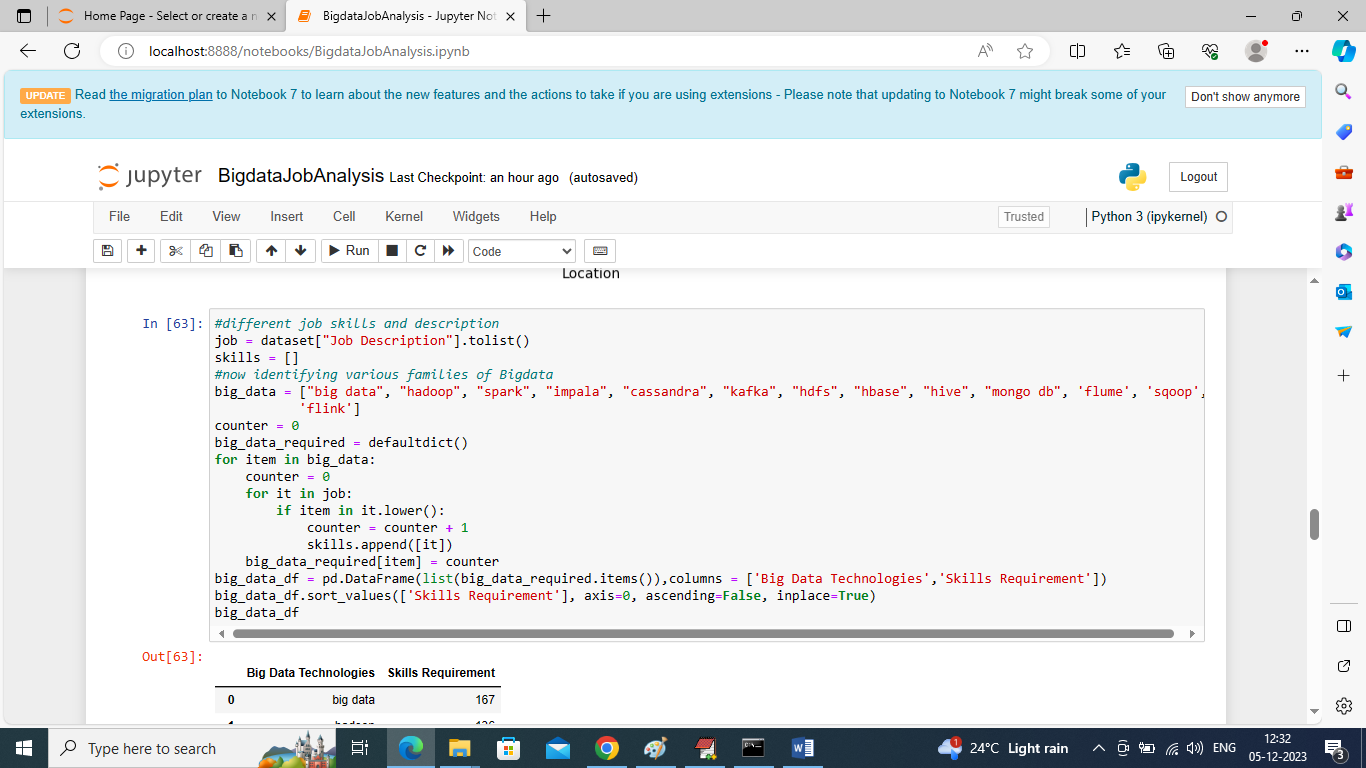
In above graph finding and displaying graph of different job categories in percentage and in all categories we can see ‘Data Analyst’ are more in demands. By seeing above graph HR can know easily above job which are high in demand



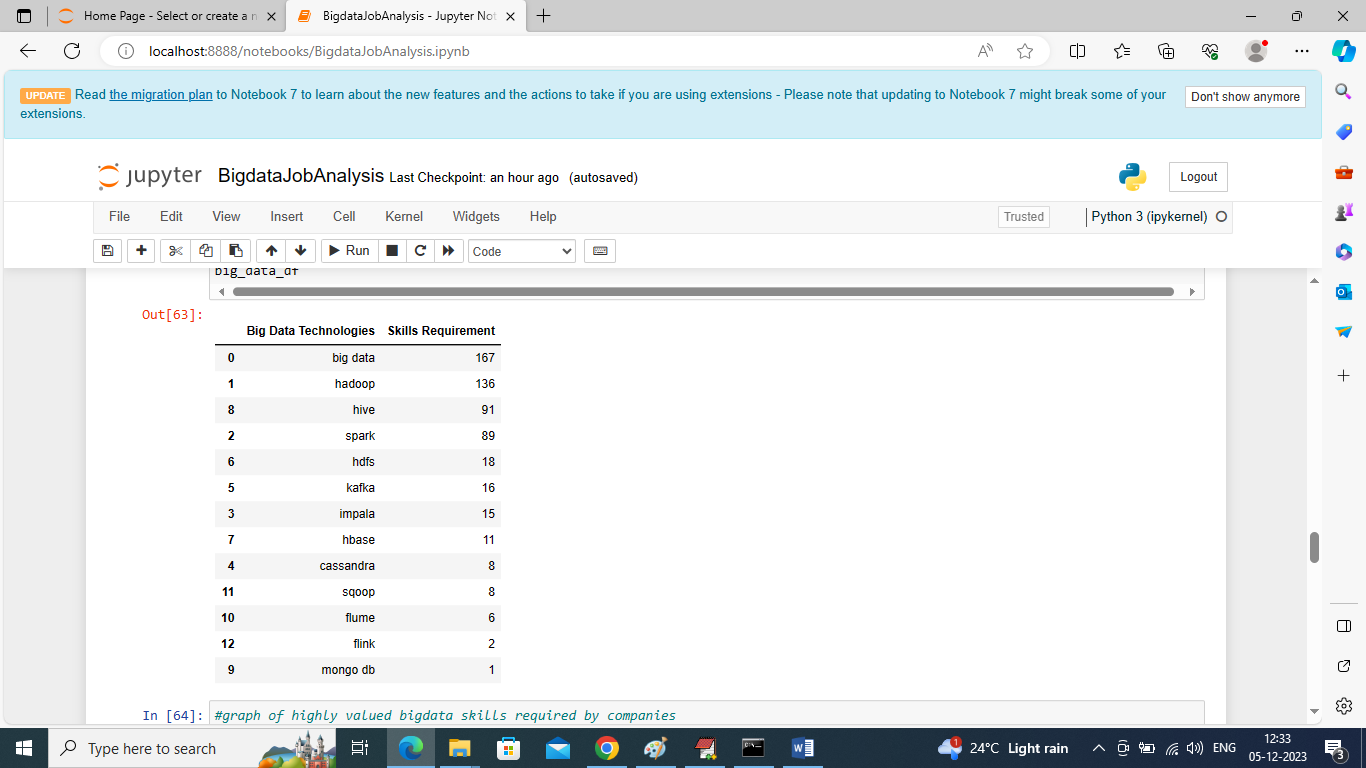
In above graph we are displaying ratings of different companies who have posted jobs and by seeing above graph HR can know this companies are genuine and posting real jobs. In above graph x-axis represents Company Names and y-axis represents Ratings



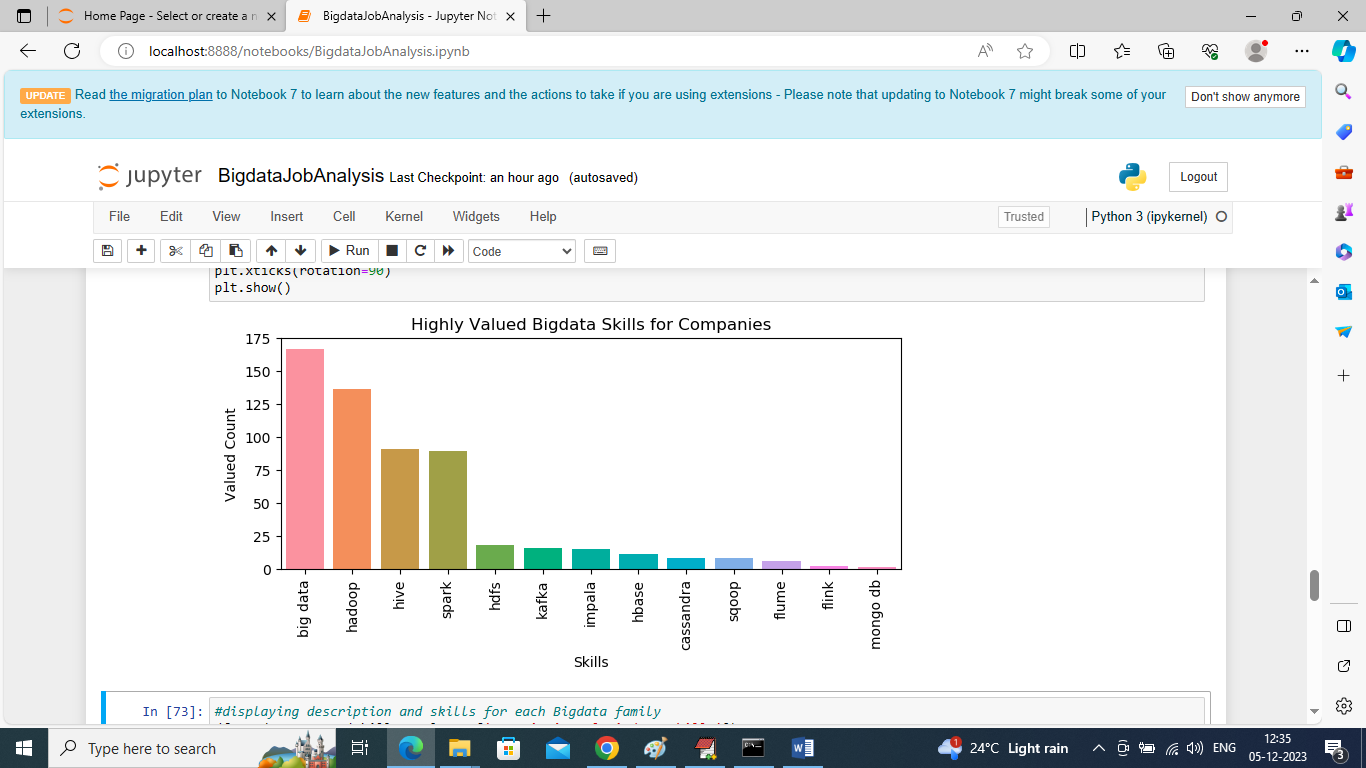
In above screen finding and displaying graph of top 20 locations who are posting more number of Jobs



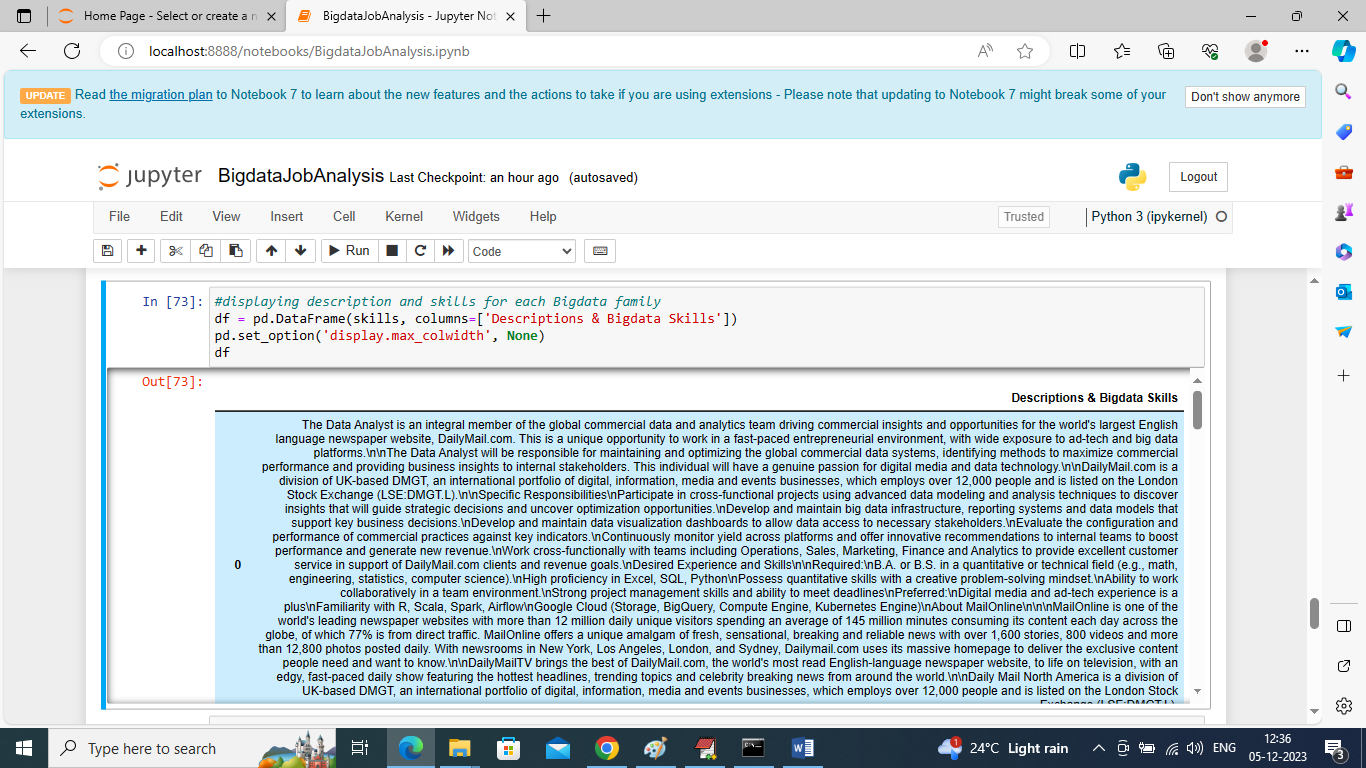
In above screen writing code to find number of jobs posted in each Bigdata family to identify its demand and valued for company



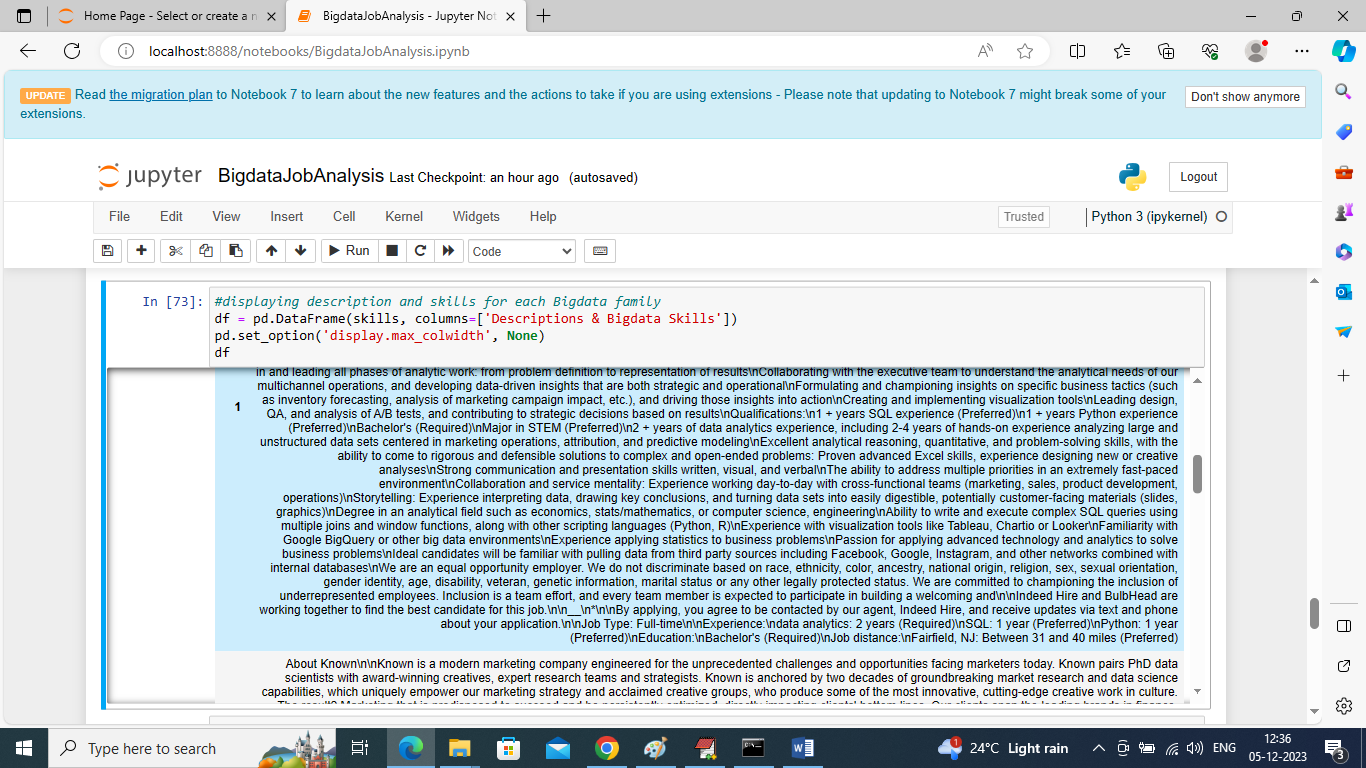
In above screen fetching and categorizing only Bigdata family technologies and their skills demands and in above table Bigdata, Hadoop, Spark, Hive libraries are in more demand



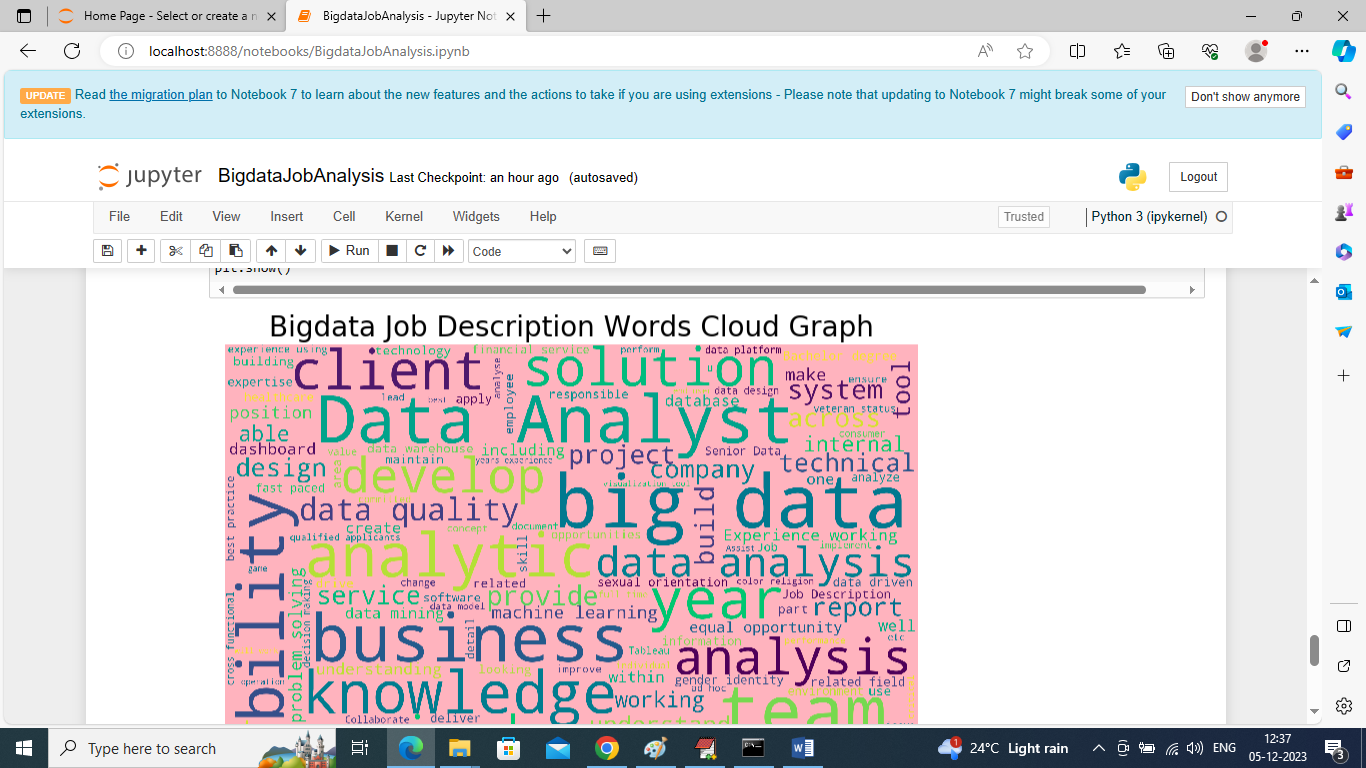
In above screen plotting graph of each Bigdata category where x-axis represents Bigdata technology names and y-axis represents requirements of Jobs for that technology



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