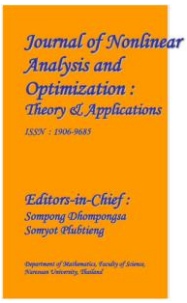
Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 15, Issue. 1, No.15 : 2024
ISSN : **1906-9685**



BIG DATA ANALYTICS

K.RAMBABU SIR[1], CHAVALI BHUVANA VENKATA DATTA[2],

[1]**Assistant professor(HOD) , MSC DEPT,** Dantuluri Narayana Raju College**, Bhimavaram, Andharapradesh**
Email:- kattarambabudnr@gmail.com
[2]**PG Student of MSC, Dantuluri** Narayana Raju College**, Bhimavaram, Andharapradesh**
Email:- Chavalidatta7@gmail.com

ABSTRACT

The project "Big Data Job Analysis" delves into the dynamics of the contemporary job market, specifically focusing on positions related to big data. Through a comprehensive analysis of job postings, skill requirements, and industry trends, this study aims to provide insights into the evolving landscape of big data jobs. The findings contribute valuable information for job seekers, employers, and educators, shedding light on the skills and qualifications in demand for successful careers in the rapidly expanding field of big data.

1 INTRODUCTION

In the introduction, the project underscores the transformative impact of big data on various industries, necessitating a closer examination of the job market. With the exponential growth in data generation, organizations seek skilled professionals capable of harnessing and analyzing this wealth of information. The study aims to identify patterns, requirements, and emerging trends in big data job postings.

Key Features:

1. **Data Collection from Job Postings:** The project involves the systematic collection of data from job postings related to big data roles. This includes positions such as data scientists, data engineers, big data analysts, and other roles crucial for managing and extracting insights from large datasets.

2. **Skill and Qualification Analysis:** Through natural language processing and data analytics, the study analyzes the skills and qualifications sought by employers in big data job postings. This includes programming languages, data manipulation tools, machine learning frameworks, and domain-specific knowledge.

2 RELEATED WORK

Literature Survey 1: Title: "Emerging Trends in Big Data Jobs: A Comprehensive Review"

Author: Sarah E. Williams

Abstract: Sarah E. Williams provides a comprehensive review of emerging trends in big data jobs. The survey covers the evolving landscape of roles, skills, and responsibilities in the big data industry, offering insights into the dynamic nature of jobs related to handling and analyzing large datasets.

Literature Survey 2: Title: "Skill Sets for Successful Careers in Big Data: State-of-the-Art Approaches"

Author: Michael J. Davis

Abstract: In this survey, Michael J. Davis explores state-of-the-art approaches to acquiring skill sets for successful careers in big data. The review covers the technical and non-technical skills required for various roles within the field, providing a foundation for understanding job requirements in big data analytics.

3 implementation study

Existing System:

The existing system for job analysis may rely on manual surveys and limited datasets, leading to incomplete or outdated information. Traditional methods may not capture the dynamic nature of the big data job market, hindering accurate trend analysis

Analysing an existing system that integrates big data with job analysis involves evaluating how organizations utilize large-scale data to optimize their workforce management and decision-making processes.

Proposed System & alogirtham

The proposed project leverages advanced data analytics techniques to systematically analyze a large volume of job postings, providing a more accurate and up-to-date representation of the big data job market. Automation ensures efficiency and the ability to capture dynamic trends in real-time.

Proposing a system that integrates big data with job analysis involves designing a framework to leverage large-scale data analytics for optimizing workforce management and decision-making processes.

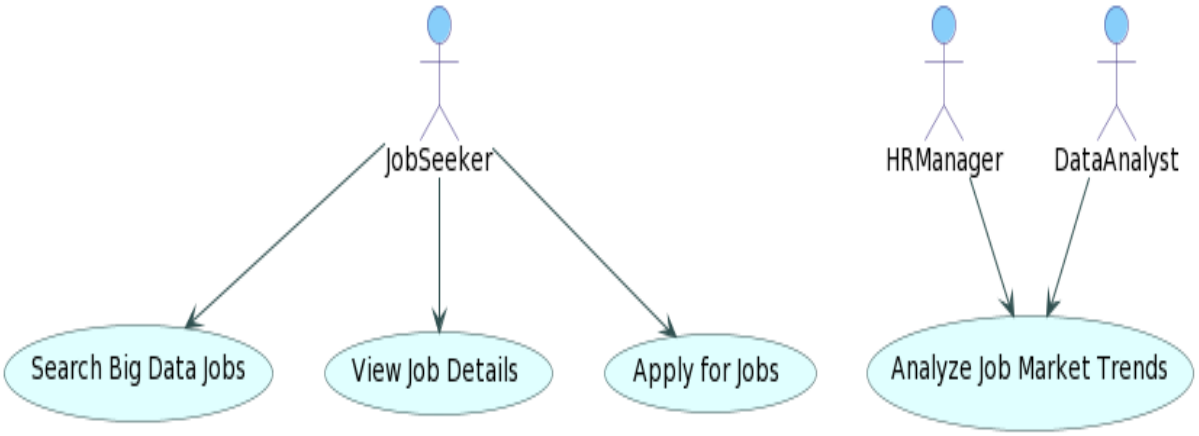


Fig1: Use Case Diagram

4. IMPLEMENTATION:

4.1 Modules Used in Project :-

Tensorflow

TensorFlow is a free and open-source software library for dataflow and differentiable programming across a range of tasks. It is a symbolic math library, and is also used for machine learning applications such as neural networks. It is used for both research and production at Google.

TensorFlow was developed by the Google Brain team for internal Google use. It was released under the Apache 2.0 open-source license on November 9, 2015.

Numpy

Numpy is a general-purpose array-processing package. It provides a high-performance multidimensional array object, and tools for working with these arrays.

It is the fundamental package for scientific computing with Python. It contains various features including these important ones:

- A powerful N-dimensional array object
- Sophisticated (broadcasting) functions
- Tools for integrating C/C++ and Fortran code
- Useful linear algebra, Fourier transform, and random number capabilities

Besides its obvious scientific uses, Numpy can also be used as an efficient multi-dimensional container of generic data. Arbitrary data-types can be defined using Numpy which allows Numpy to seamlessly and speedily integrate with a wide variety of databases.

Pandas

Pandas is an open-source Python Library providing high-performance data manipulation and analysis tool using its powerful data structures. Python was majorly used for data munging and preparation. It had very little contribution towards data analysis. Pandas solved this problem. Using Pandas, we can accomplish five typical steps in the processing and analysis of data, regardless of the origin of data load, prepare, manipulate, model, and analyze. Python with Pandas is used in a wide range of fields including academic and commercial domains including finance, economics, Statistics, analytics, etc.

5 RESULTS AND DISCUSSION

SCREENSHOTS

Bigdata Job Analysis

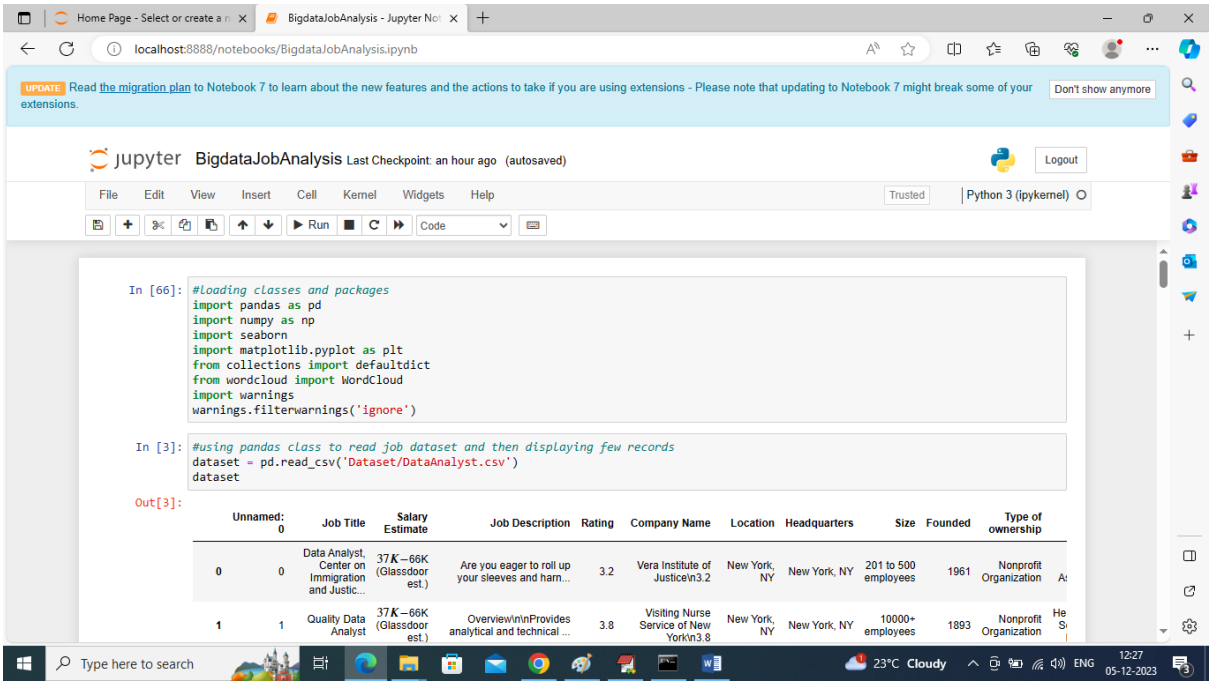
In propose work we are analyzing large amount of online Job posted dataset to find Bigdata family job skills. Since introduction of Bigdata many supporting technologies are introduced and many peoples are unfamiliar about all those technologies and their demands. Selecting suitable Bigdata job family 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

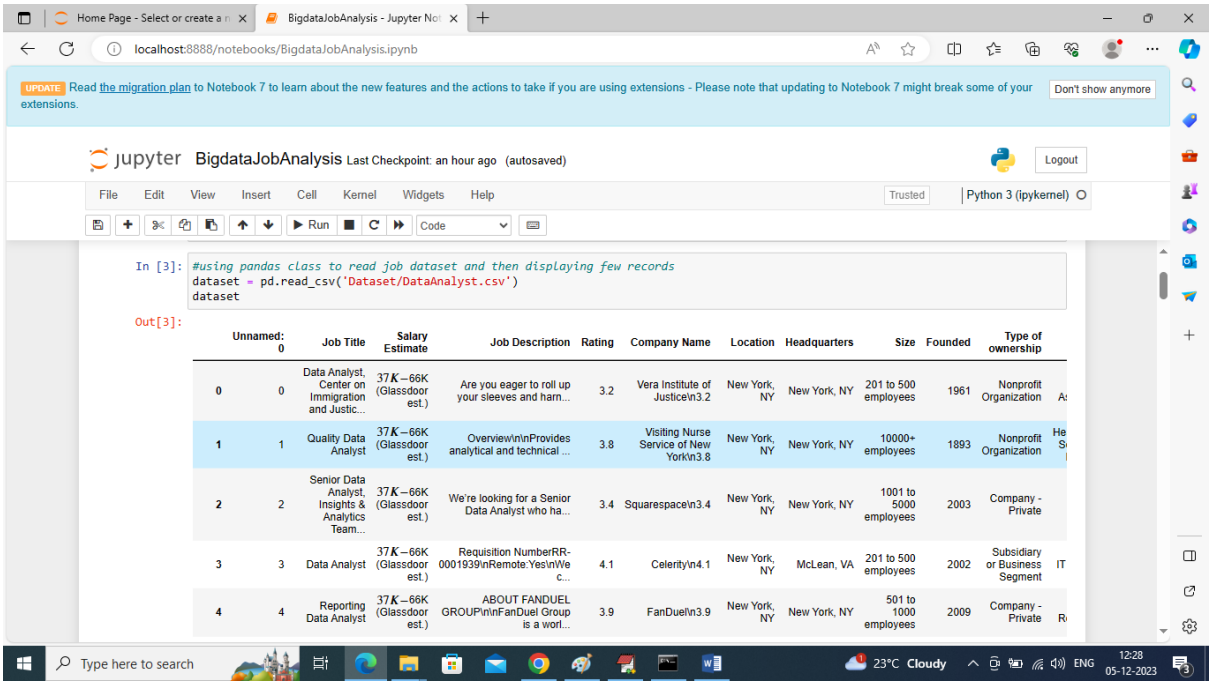
https://www.kaggle.com/code/rohitsahoo/data-analyst-job-analysis/input

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

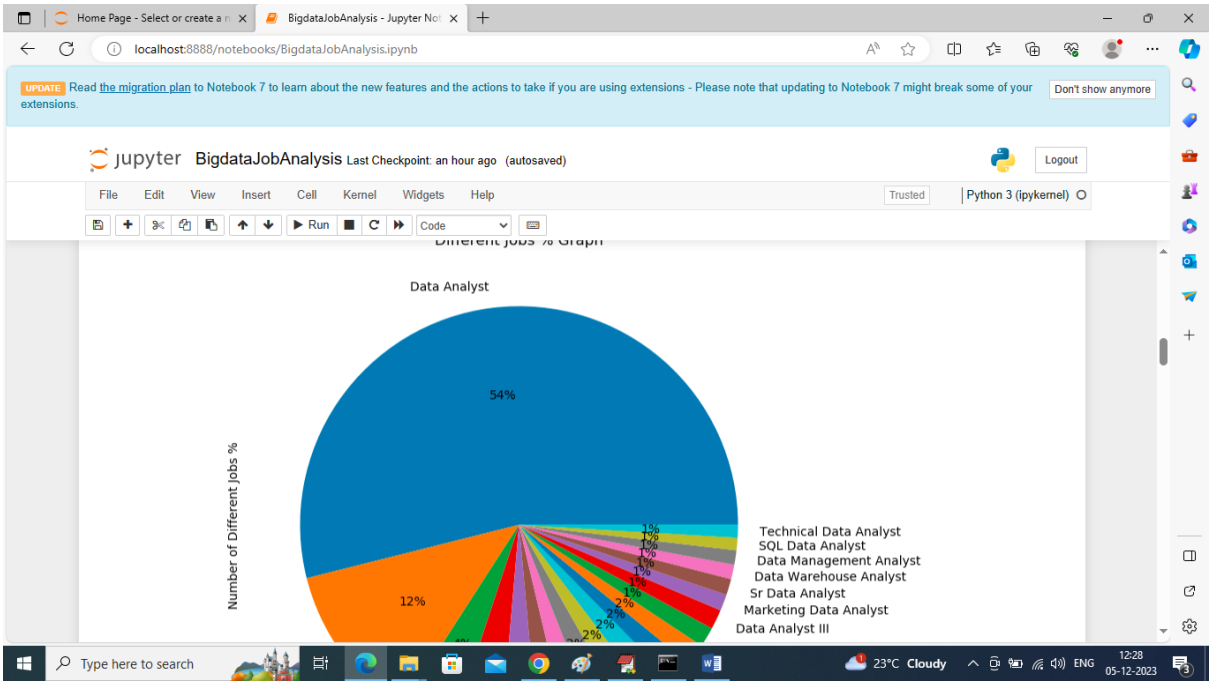
We have coded this project using JUPYTER notebook and below are the code and output screens with blue color comments



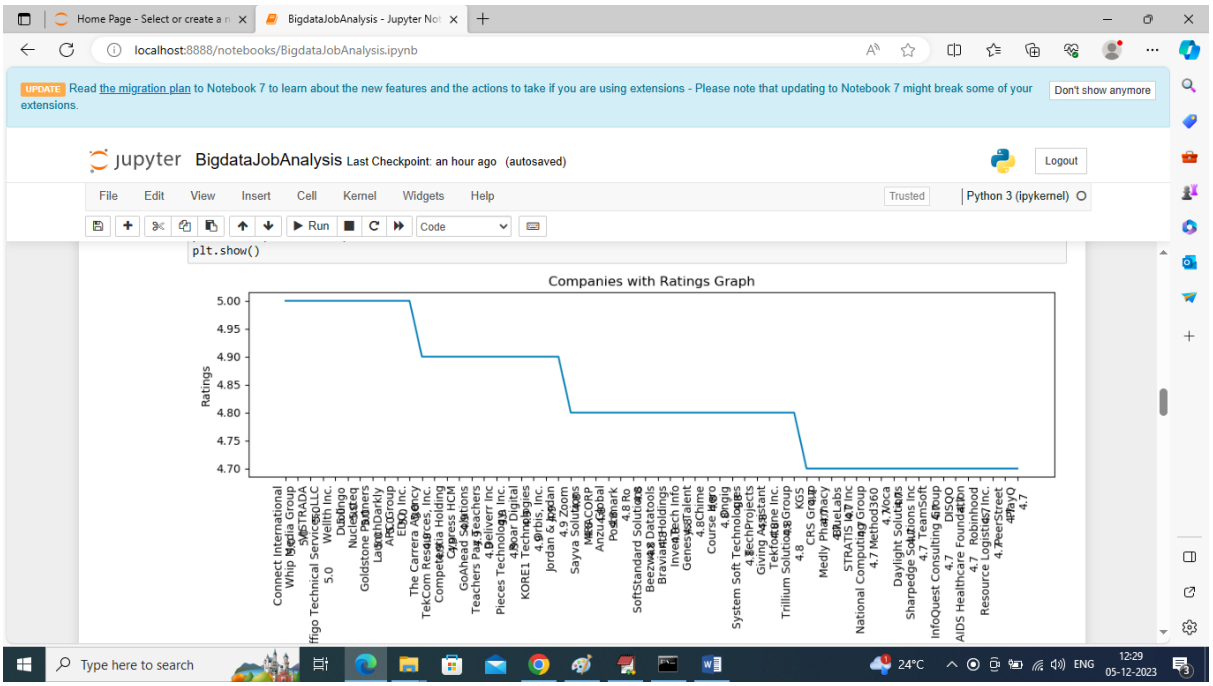
In above screen importing required python classes and packages



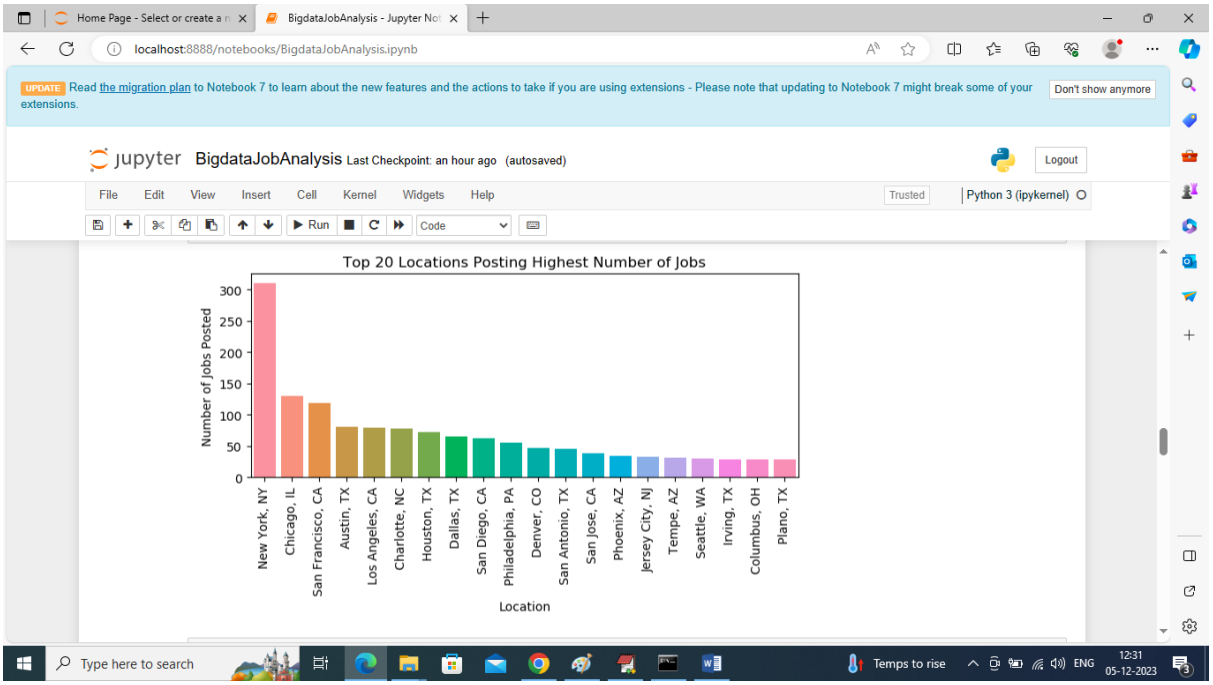
In above screen loading and displaying dataset values



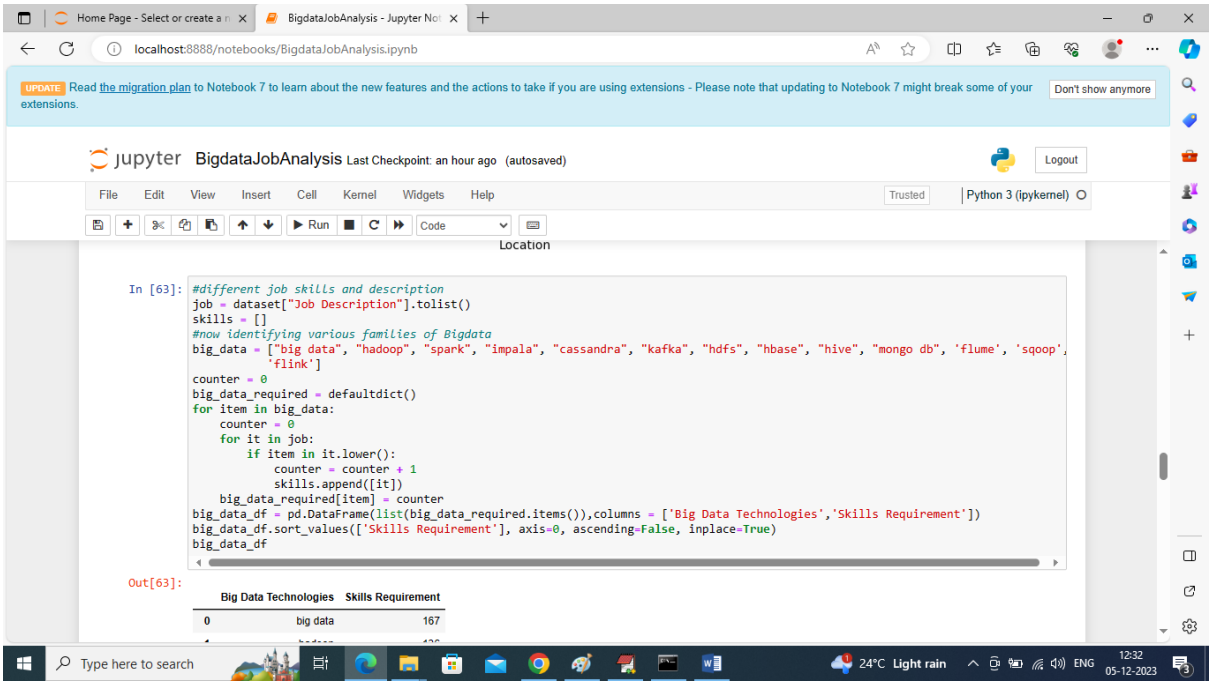
In above graph finding and displaying graph of different job categories in percentage and in all categories we can see 'Data Analyst' are more in demands. By seeing above graph HR can know easily above job which are high in demand



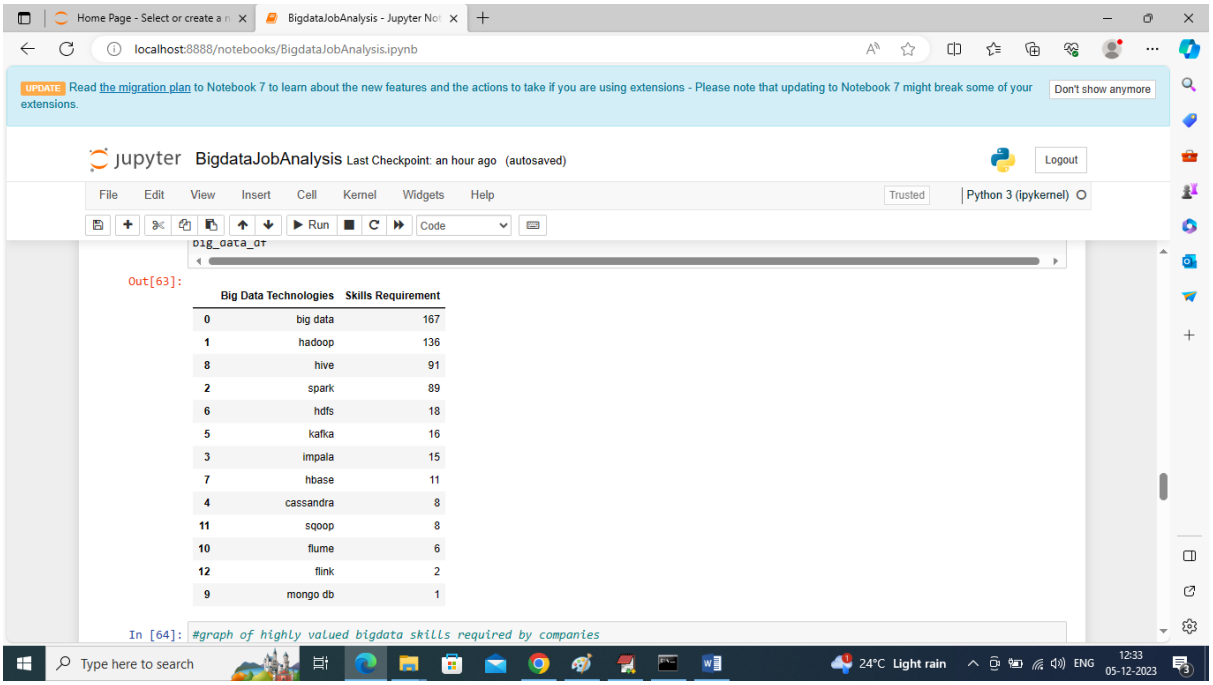
In above graph we are displaying ratings of different companies who have posted jobs and by seeing above graph HR can know this companies are genuine and posting real jobs. In above graph x-axis represents Company Names and y-axis represents Ratings



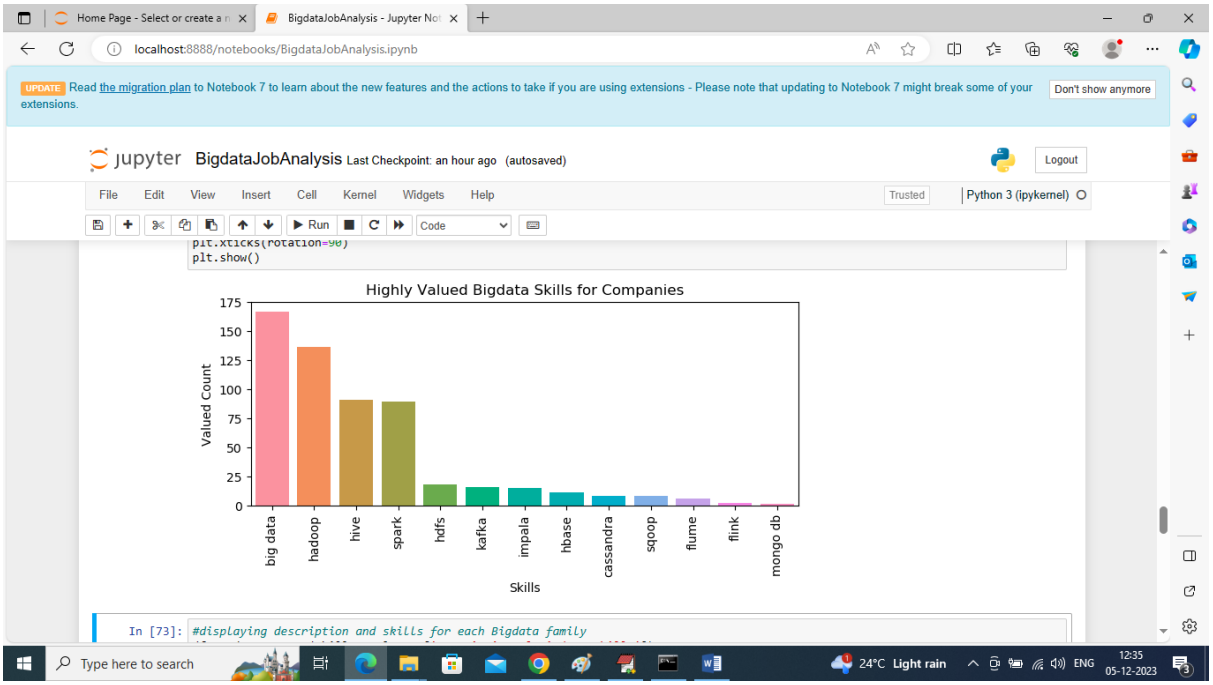
In above screen finding and displaying graph of top 20 locations who are posting more number of Jobs



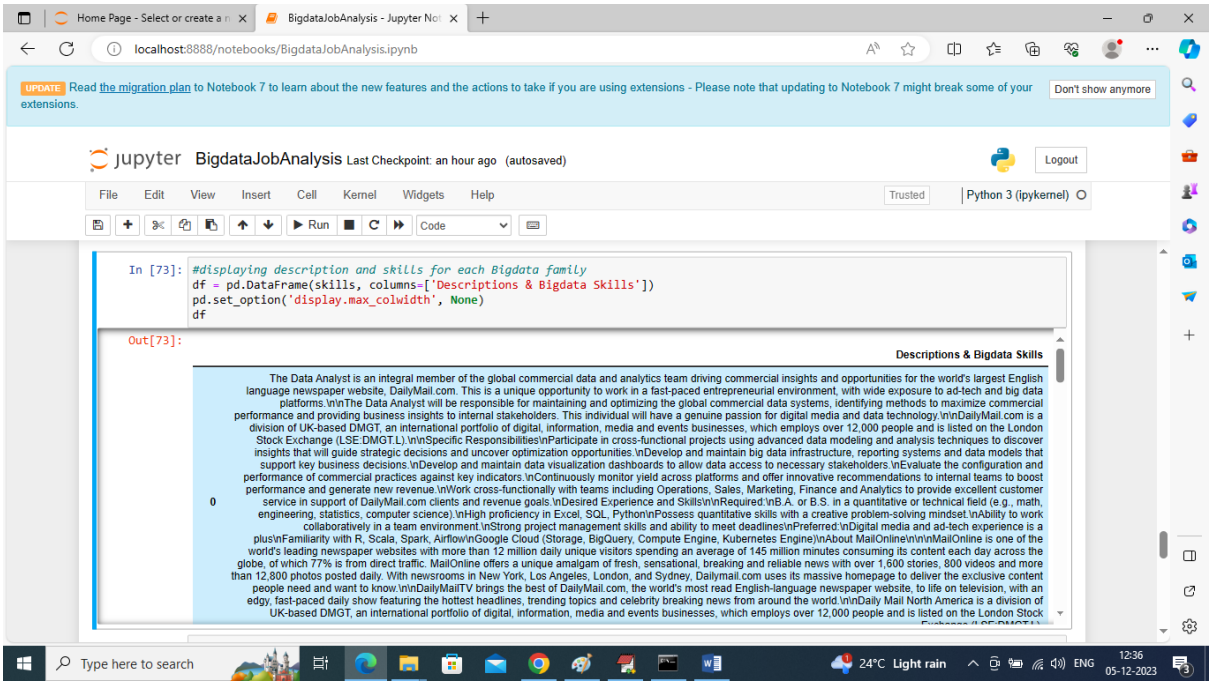
In above screen writing code to find number of jobs posted in each Bigdata family to identify its demand and valued for company



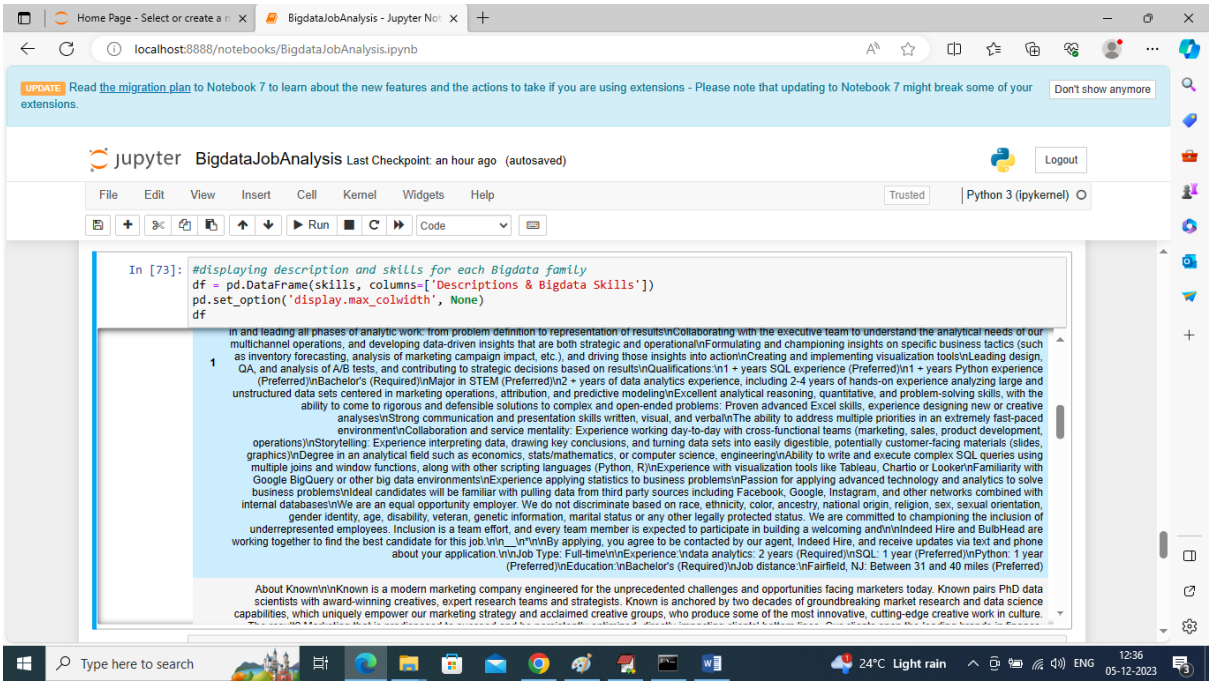
In above screen fetching and categorizing only Bigdata family technologies and their skills demands and in above table Bigdata, Hadoop, Spark, Hive libraries are in more demand



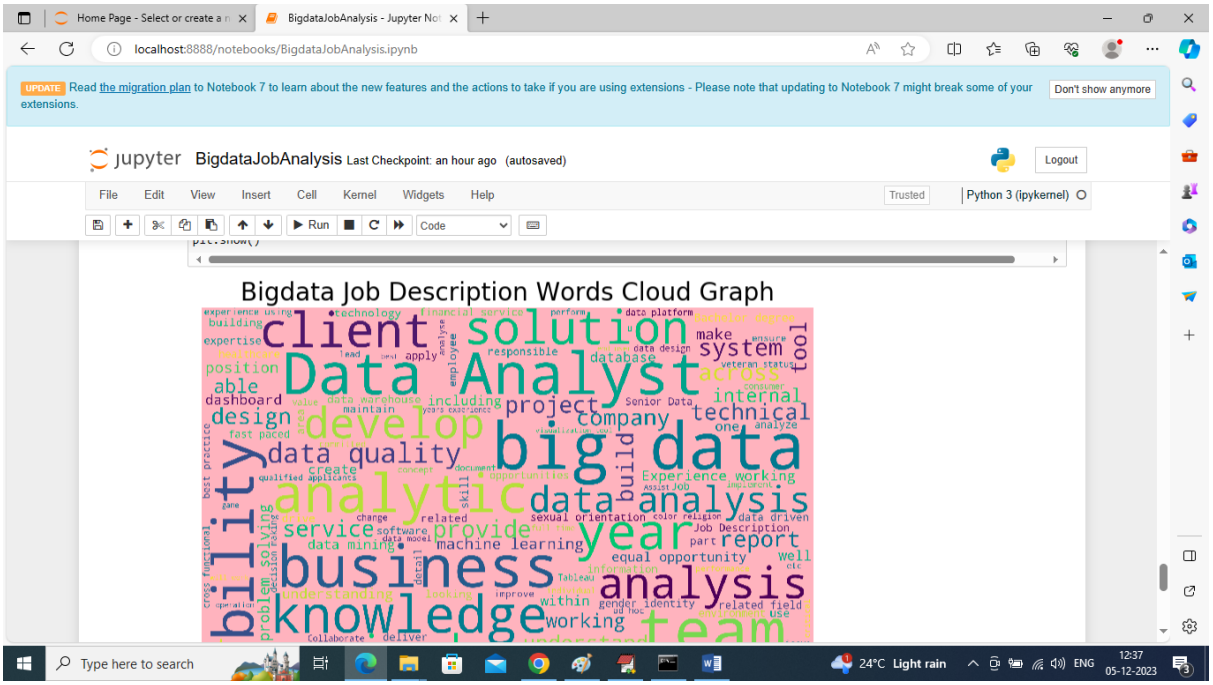
In above screen plotting graph of each Bigdata category where x-axis represents Bigdata technology names and y-axis represents requirements of Jobs for that technology



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

6. CONCLUSION AND FUTURE WORK

CONCLUSION

In conclusion, "Big Data Job Analysis" serves as a valuable resource for stakeholders in the big data ecosystem. The insights gained from this study empower job seekers to align their skillsets with industry demands, assist employers in refining recruitment strategies, and guide educators in shaping curricula to meet the evolving needs of the big data job market.