**The Future Of Work: Data Analysis Of Glassdoor Jobs**

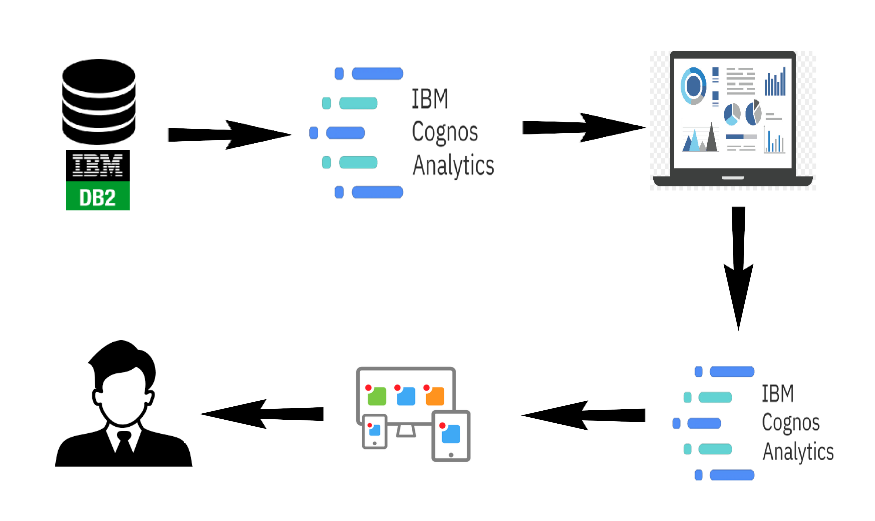
**Introduction:**

Job analysis is a systematic procedure to analyse the requirements for the job role and job profile. Glassdoor is a website and online platform that provides information about jobs, salaries, and companies. Job analysis is a systematic approach to defining the job role, description, requirements, responsibilities, evaluation, etc. It helps in finding out required level of education, skills, knowledge, training, etc for the job position. It also depicts the job worth i.e. measurable effectiveness of the job and contribution of job to the organization. Thus, it effectively contributes to setting up the compensation package for the job position.

Lack of analysis of Glassdoor jobs can result in limited understanding of job market trends, difficulty in finding relevant job opportunities, inability to attract and retain top talent, and lack of insight into company branding and reputation.

The purpose of this project is to conduct an analysis of Glassdoor job postings to gain insights into current and emerging job market trends, identify in-demand skills and experience, and understand how employers can improve their employer branding and reputation to attract and retain top talent.

**Technical Architecture:**



**Define Problem / Problem Understanding:**

A problem statement is a clear and concise description of the issue or challenge that needs to be addressed. It should define the problem in a way that is understandable to stakeholders and provide a basis for developing a solution or course of action.

### Specify The Business Problem:

Job analysis is a systematic procedure to analyse the requirements for the job role and job profile. Glassdoor is a website and online platform that provides information about jobs, salaries, and companies. Job analysis is a systematic approach to defining the job role, description, requirements, responsibilities, evaluation, etc. It helps in finding out required level of education, skills, knowledge, training, etc for the job position. It also depicts the job worth i.e. measurable effectiveness of the job and contribution of job to the organization. Thus, it effectively contributes to setting up the compensation package for the job position.

Lack of analysis of Glassdoor jobs can result in limited understanding of job market trends, difficulty in finding relevant job opportunities, inability to attract and retain top talent, and lack of insight into company branding and reputation.

The purpose of this project is to conduct an analysis of Glassdoor job postings to gain insights into current and emerging job market trends, identify in-demand skills and experience, and understand how employers can improve their employer branding and reputation to attract and retain top talent.

**Business Requirements:**

### Understanding the data of different jobs provided by glassdoor can help businesses and personnel to analyze current market trends in hiring, packages offered, etc. Businesses need to understand the glassdoor jobs data in order to get valuable insights. Job analysis is a crucial step in validating all major personnel activities. Employers must be able to show that their screening tools and appraisals are actually related to performance on the job in question. Doing this, of course, requires knowing what the job entails, which in turn requires a competent job analysis. The ultimate goal is to gain insights and improve performance through data visualization techniques.

### Literature Survey:

### A literacy survey for Data Analysis of Glassdoor Jobs involves reviewing multiple job roles in a particular domain offered by a particular organisation belonging to a given industry and sector. Job analysis defines the organization of jobs within a job family. It allows units to identify paths of job progression for employees interested in improving their opportunities for career advancement and increasing compensation.

### 

### Social Or Business Impact:

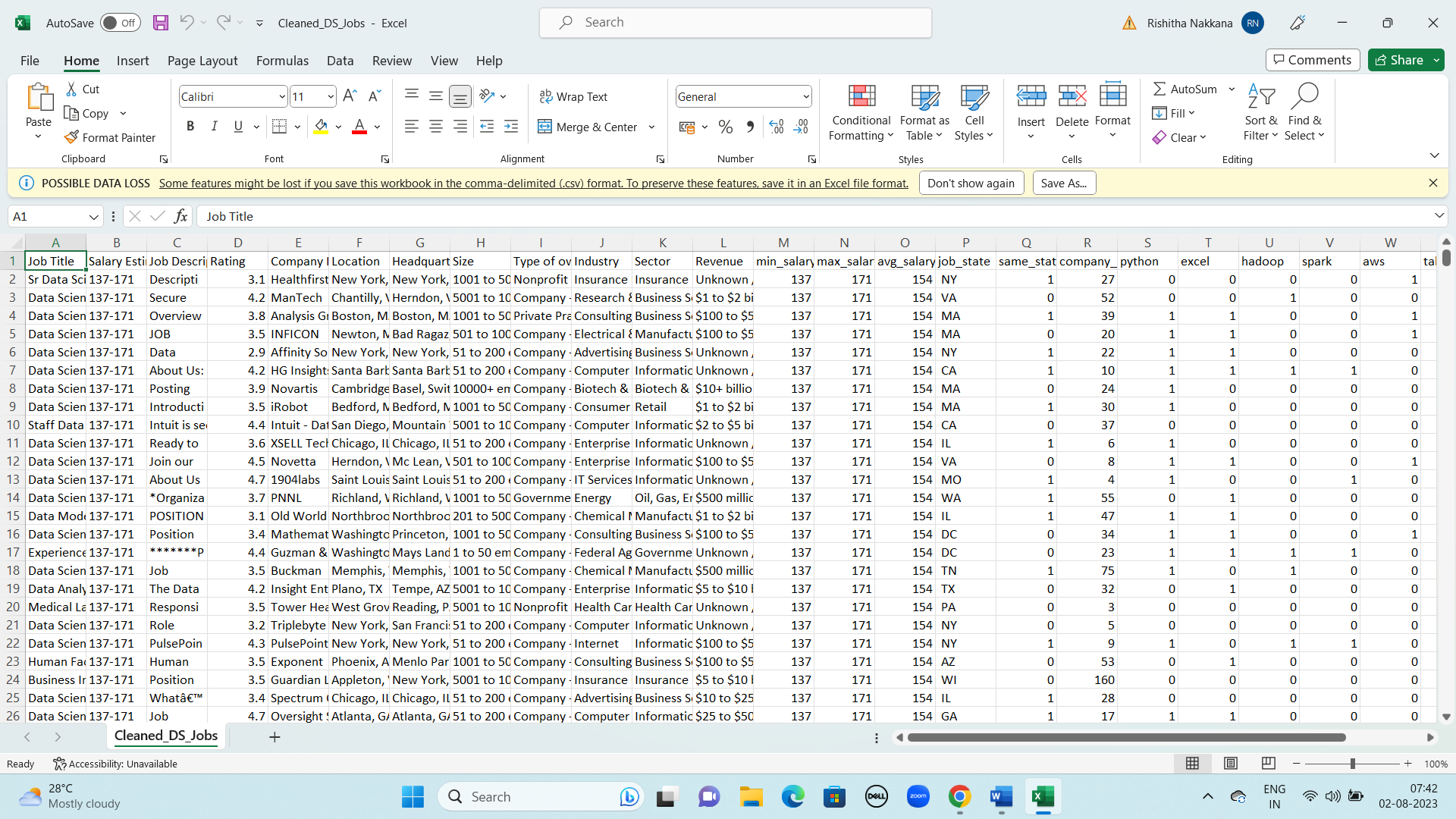
This project can help job seekers make more informed decisions about their careers and negotiate for better compensation and working conditions. This can ultimately contribute to greater economic mobility and reduce income inequality.

**Business Model/Impact:**

It can help to improve retention rates, reduce turnover costs, and increase productivity. An analysis of Glassdoor jobs can provide insights into what employees value most, helping employers to create a better work environment that attracts and retains top talent.

**Data Collection & Extraction From Database:**

Data collection is the process of gathering and measuring information on variables of interest, in an established systematic fashion that enables one to answer stated research questions, test hypotheses, and evaluate outcomes and generate insights from the data.



### Understand The Data:

Data contains all the meta information regarding the columns described in the CSV files. The name of file is Cleaned\_DS\_Jobs.csv

Description for Cleaned\_DS\_Jobs.csv:

The file Cleaned\_DS\_Jobs.csv contains 660 rows.Each row corresponds to record of state with details and marks in respective subjects. The columns are:

* Job Title: Title of the job posting
* Salary Estimation: Salary range for that particular job
* Job Description: This contains the full description of that job
* Rating: Rating of that post
* Company: Name of company
* Location: Location of the company
* Headquarter: Location of the headquater
* Size: Total employee in that company
* Type of ownership: Describes the company type i.e non-profit/public/private farm etc
* Industry, Sector: Field applicant will work in
* Revenue: Total revenue of the company
* min\_salary,max\_salary,avg\_salary: Refers to the minimum, maximum and average salary for that post
* job\_state: State where the applicant will work
* same\_state: Same state as headquarter or not(Boolean)
* company\_age: Age of company

### Storing Data In DB2 & Connect DB2 With Cognos:

To connect to the database from your Cognos server, you need the db2jcc.jar file. The db2jcc.jar file is part of the [Db2 driver package](https://www.ibm.com/docs/en/SS6NHC/com.ibm.swg.im.dashdb.doc/connecting/connect_driver_package.html), which you can download from the web console. The same file is available in the Windows or the Linux® download.

1. Download the Db2 driver package from the web console.
   * The Windows package is an executable file that installs multiple drivers on your computer.
   * The Linux package contains the db2jcc.jar file within compressed files. You can use the Linux package to access the db2jcc.jar file on a Windows computer too. You will need a decompress utility, such as 7-Zip.

2. Locate the db2jcc.jar file:

* + Windows: After you run the ibm\_data\_server\_driver\_package\_win64\_v10.5.exe file, the db2jcc.jar file is installed in the C:\Program Files\IBM\IBM DATA SERVER DRIVER\java directory.
  + Linux: Decompress the ibm\_data\_server\_driver\_package\_linuxx64\_v10.5.tar.gz file, and then decompress the ibm\_data\_server\_driver\_package\_linuxx64\_v10.5.tar file. Go to the dsdriver\jdbc\_sqlj\_driver\linuxamd64 subdirectory. Decompress the db2\_db2driver\_for\_jdbc\_sqlj.zip file to access the db2jcc.jar file.

1. Copy the db2jcc.jar file to the c10\_location\webapps\p2pd\WEB-INF\lib directory on your Cognos Content Manager computer.

Follow these steps to connect the Cognos server to the Db2 database, BLUDB:

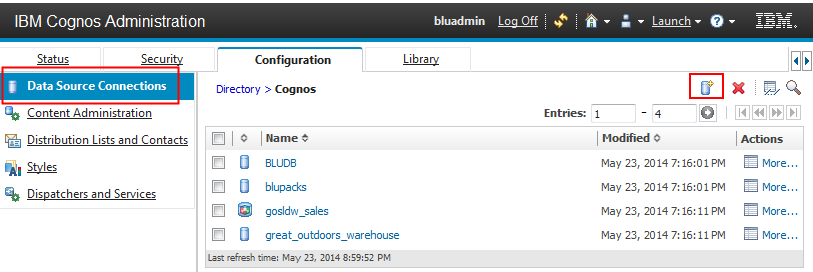
1. Collect [connect information and credentials](https://www.ibm.com/docs/en/SS6NHC/com.ibm.swg.im.dashdb.doc/connecting/connect_credentials.html).

5. On your Cognos server, start IBM Cognos Administration. Use one of the following methods:

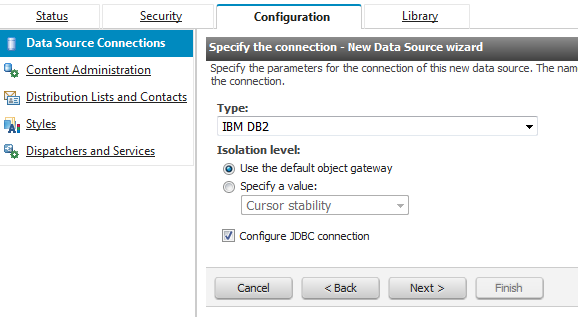
* + In the Welcome page, click **Administer IBM Cognos Content**.
  + In IBM Cognos Connection, from the toolbar, click **Launch**> **IBM Cognos Administration**.



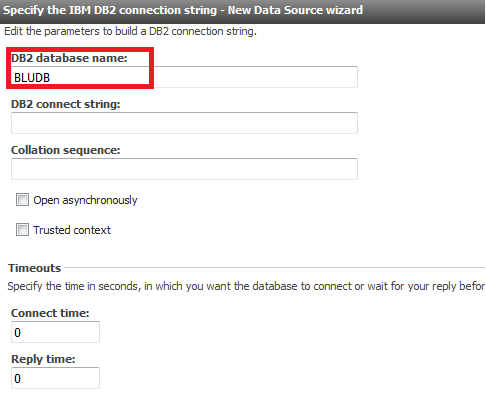
1. From the **Configuration** tab, select **Data Source Connections**.



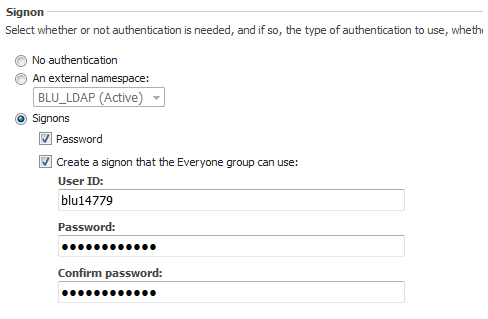
1. Click the new data source icon New data source icon.. The New Data Source wizard opens.
2. In the "Specify a name and description" page, enter a unique name for the BLUDB data source and an optional description and screen tip. Click **Next**.
3. In the "Specify the connection" page, select **IBM DB2** for the type and specify the isolation level. Ensure that **Configure JDBC** connection is selected, and then click **Next**.



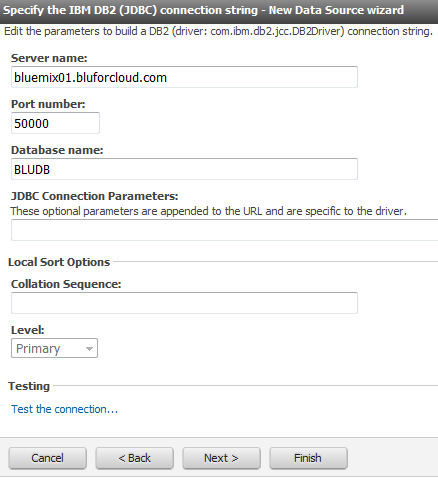
1. In the "Specify the IBM DB2 connection string" page, use the following parameters:
   * Enter BLUDB in the **DB2 database name** field.



* + Ensure that the **Signons** radio button is selected, and then select the two check boxes for **Password** and **Create a signon that the Everyone group can use**. Use the user ID and password that you obtained in step [4](https://www.ibm.com/docs/en/db2woc?topic=bi-cognos#connect_connecting_cognos__param). Click **Finish**.



1. On the "Specify the IBM DB2 (JDBC) connection string" page, specify the server name and port number that you obtained in step [4](https://www.ibm.com/docs/en/db2woc?topic=bi-cognos#connect_connecting_cognos__param). The database name is BLUDB.



1. Click **Test the connection**, then click **Test**. On the "View the results" page, the status of the connection tests for the dynamic query mode should be **Succeed**. Click **Finish**.

## **Results:**

Your Cognos server is now connected to the Db2 database. Use your Cognos tools to create dynamic cubes and start generating reports.

**Data Preparation:**

Data modules are containers that describe data and rules for combining and shaping data to prepare it for analysis and visualization in IBM Cognos Analytics. Data module sources. Data modules can be based on data servers, packages, uploaded files, data sets, and other data modules.

### Prepare The Data For Visualization:

Data preparation is a pre-analysis step that is used by most data analytic algorithms to ensure that the data is suitable for analytic use.

## **Overview:**

Data preparation is critical in IBM® Cognos Analytics. Only prepared data is entered into analysis for key drivers, decision trees, and relationships that are displayed in the advanced analytics visualizations: Spiral, Driver analysis, Decision tree, Sunburst, and Explore relationships. Data is not automatically prepared for other visualizations and their corresponding insights.

## **Algorithms:**

All applied algorithms are based on values of a single field at a time. Missing values are removed or handled for each field, all numeric predictor driver fields are binned. All categorical fields are adjusted for large number of categories and outliers are handled in the target field. While all data preparation influences the analysis results, corresponding data preparation summaries are not currently reported to you.

## **Details:**

Data preparation and subsequent key drivers, decision trees and relationships are based on a data sample with approximately 10,000 rows when the original data is larger. Bernoulli random sampling, equal probability without replacement random sampling, is applied to uploaded data and any connected data sources that are supporting random sampling. Otherwise, systematic sampling is used.

**Data Visualization:**

Data visualization is the process of creating graphical representations of data in order to help people understand and explore the information. The goal of data visualization is to make complex data sets more accessible, intuitive, and easier to interpret. By using visual elements such as charts, graphs, and maps, data visualizations can help people quickly identify patterns, trends, and outliers in the data.

### No.Of Unique Visualizations:

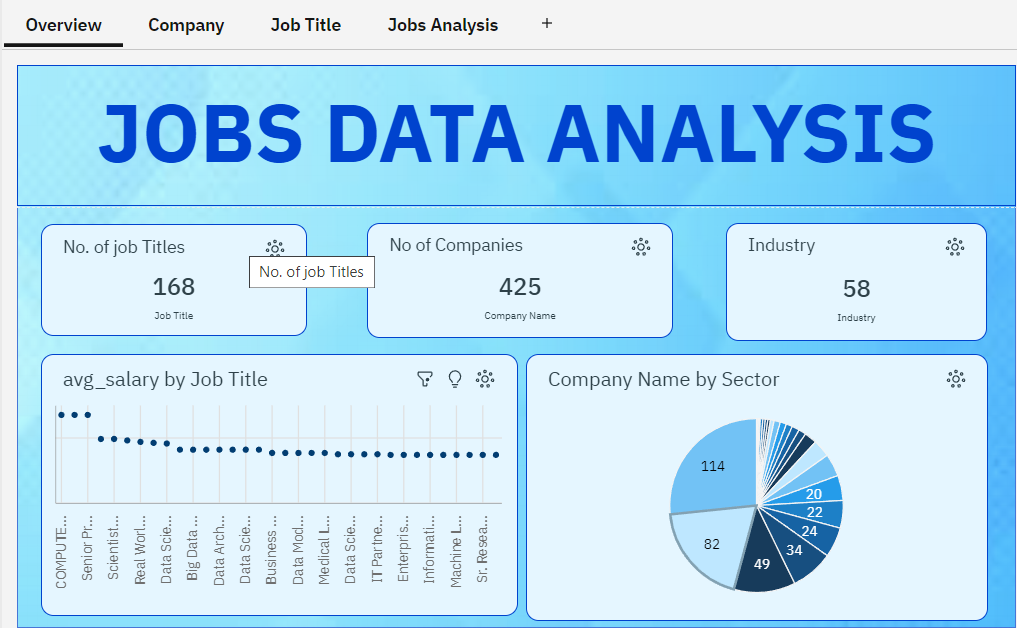
### The number of unique visualizations that can be created with a given dataset. Some common types of visualizations that can be used to analyse the Literacy include bar charts, line charts, heat maps, scatter plots, pie charts, Maps etc. These visualizations can be used to compare performance, track changes over time, show distribution, and relationships between variables, breakdown of revenue and customer demographics, workload, resource allocation and location of different job roles.

### Dashboard:

A dashboard is a graphical user interface (GUI) that displays information and data in an organized, easy-to-read format. Dashboards are often used to provide real-time monitoring and analysis of data, and are typically designed for a specific purpose or use case. Dashboards can be used in a variety of settings, such as business, finance, manufacturing, healthcare, and many other industries. They can be used to track key performance indicators (KPIs), monitor performance metrics, and display data in the form of charts, graphs, and tables.

**Responsive And Design Of Dashboard:**

The responsiveness and design of a dashboard for Data-Driven insights on Student Performance is crucial to ensure that the information is easily understandable and actionable. Key considerations for designing a responsive and effective dashboard include user-cantered design, clear and concise information, interactivity, data-driven approach, accessibility, customization, and security. The goal is to create a dashboard that is user-friendly, interactive, and data-driven, providing actionable insights.



### Story:

A data story is a way of presenting data and analysis in a narrative format, with the goal of making the information more engaging and easier to understand. A data story typically includes a clear introduction that sets the stage and explains the context for the data, a body that presents the data and analysis in a logical and systematic way, and a conclusion that summarizes the key findings and highlights their implications. Data stories can be told using a variety of mediums, such as reports, presentations, interactive visualizations, and videos.

