PROJECT WORK

PROJECT TITLE:-THE FUTURE OF WORK: DATA ANALYSIS OF GLASS DOOR JOBS.

FACULTY MENTOR: P. SIVA KIRAN KUMAR

TEAM ID: LTVIP2023TMID00414

TEAM SIZE:5

TEAM LEADER: POKASETTI RADHA SAI ALEKHYA

TEAM MEMBERS: RASAPILLI MADHURI

KALDI RENUKA

REVALLA GOWTAMI

GANDAMANI ANJALI

THE FUTURE OF WORK: DATA ANALYSIS OF GLASSDOOR JOBS.

1. Introduction:

1.1. Overview:

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.

1.2. Purpose:

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.

We can achieve the following using data analytics:

- Personalize the customer experience. Glassdoor jobs collect employer's data from many different channels, including physical retail, and social media.
- Inform Glassdoor job's decision-making.
- Streamline operations.
- Mitigate risk and handle setbacks.
- Enhance security.

2. 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.

Understanding the data of different jobs provided by Glassdoor can help businesses and personnel to analyse 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.

Social 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 employee value most, helping employers to create a better work environment that attracts and retains top talent.

2.1. Existing Problem:

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. Doing data analysis can solve this problem.

2.2. Proposed Solution:

To accomplish this, we have to complete all the activities listed below,

Data Collection & Extraction from Database

- Collect the dataset,
- Storing Data in DB2
- Perform SQL Operations
- Connect DB2 with Cognos

- 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 Visualizations
    - No of Unique Visualizations

- Dashboard
    - Responsive and Design of 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.

- Story
    - No of Scenes of 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.

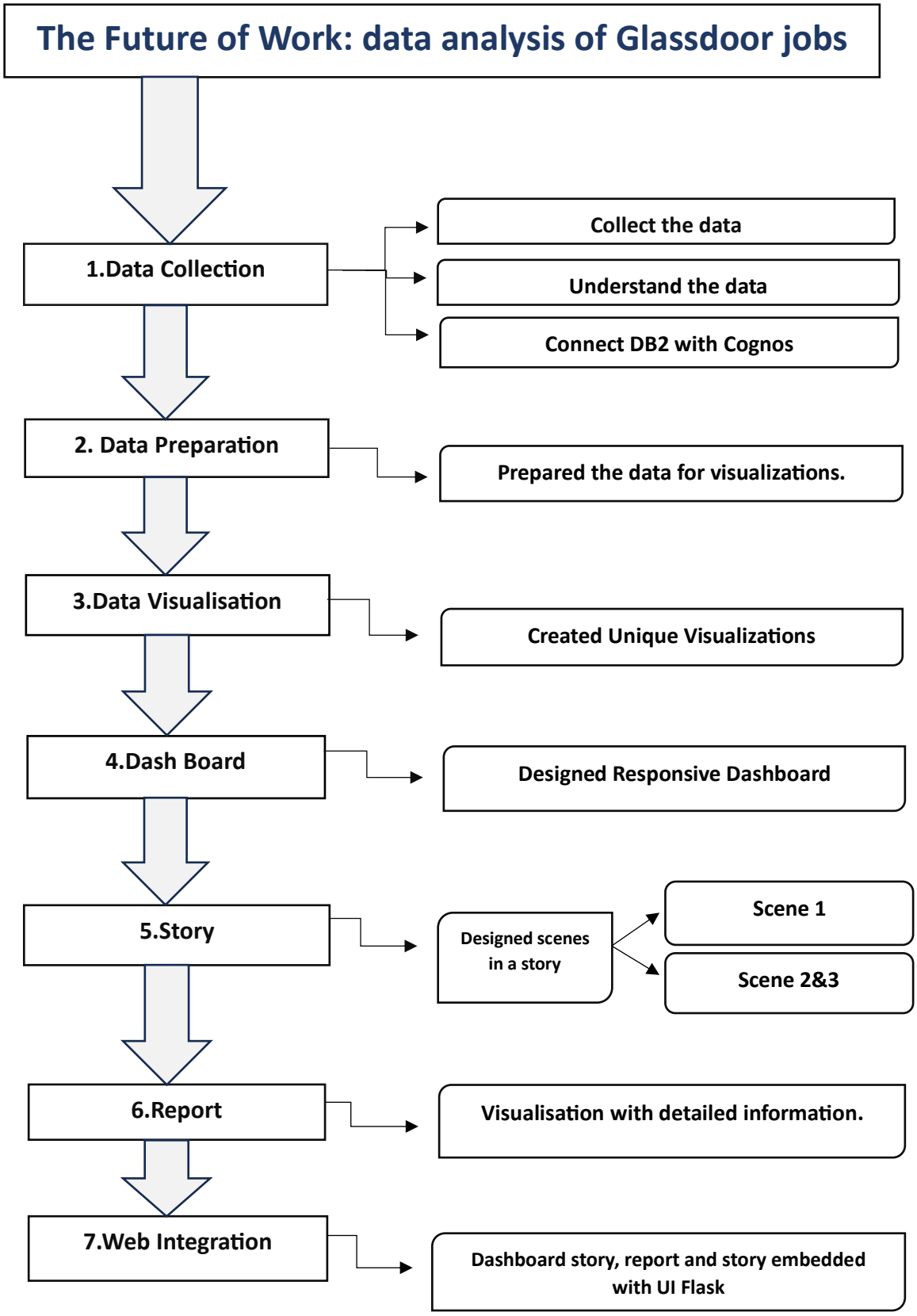
- Report
    - No of Visualization with detail information

- Performance Testing
    - Amount of Data Rendered to DB2
    - Utilization of Data Filters
    - No of Calculation Fields
    - No of Visualizations/ Graphs

- Web Integration
    - Dashboard, Report and Story embed with UI With Flask

3. Theoretical Analysis:

3.1. Block Diagram:



3.3. Hardware/Software Testing:

Understanding the data of different jobs provided by Glassdoor can help businesses and personal to analyse 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. Employers must have hardware like laptop or a desktop with latest versions of windows, mac or Linux. Also they need to login into IBM Cognos Analysis, download a dataset regarding to Glassdoor jobs, upload and creating visualizations should be done. After that we need to do web integration using Anaconda Navigator, Jupyter & Spyder.

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.

4. Result:

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.

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.

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.

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.

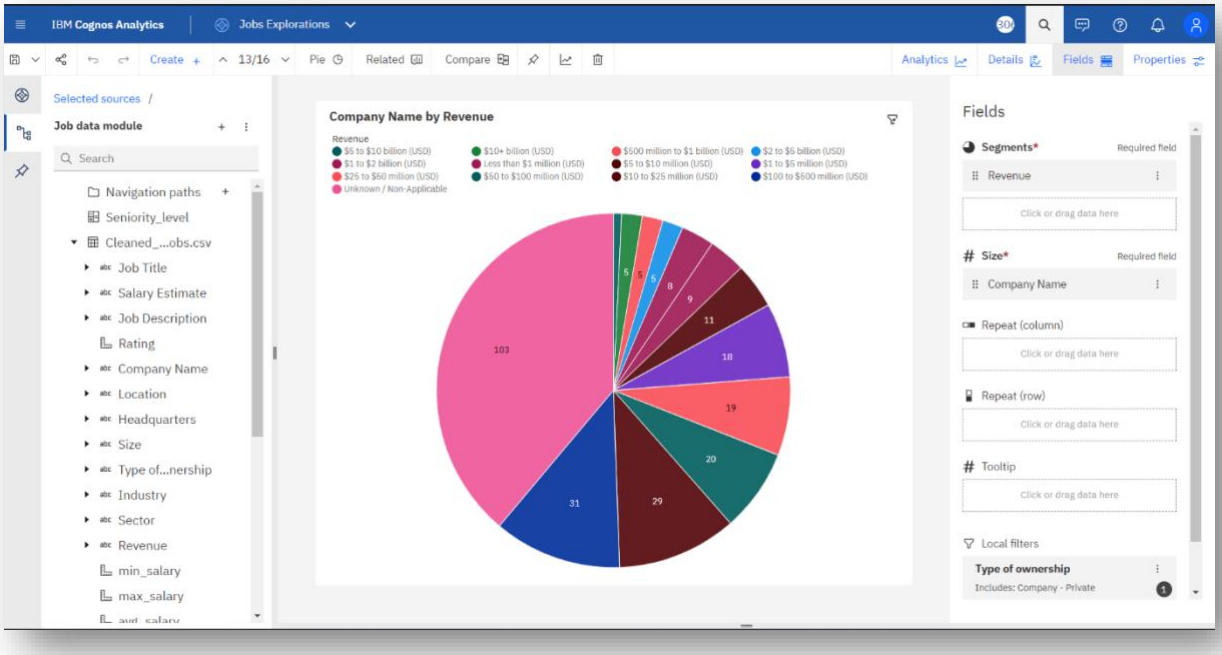
A report in data analytics typically involves analysing and interpreting data to draw insights and conclusions that can inform business decisions or address research questions. The report usually includes a summary of the data analysis process, including the methods and tools used, as well as the findings and recommendations based on the analysis. The report should begin with an executive summary, which provides a brief overview of the main findings and recommendations. The introduction should provide background information on the problem or research question being addressed and the data sources used.

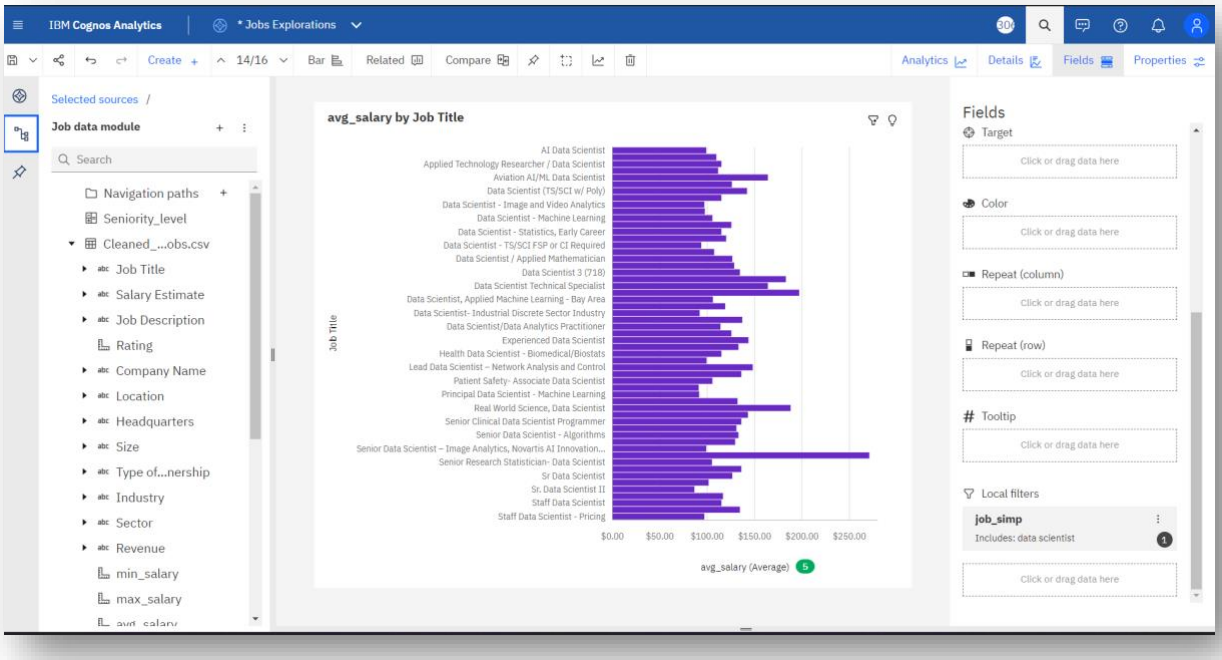
Understanding the data of different jobs provided by Glassdoor can help businesses and personnel to analyse current market trends in hiring, packages offered, etc. Businesses need to understand the Glassdoor jobs data in order to get valuable insights. Job analysis is a crucial step in validating all 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.

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.

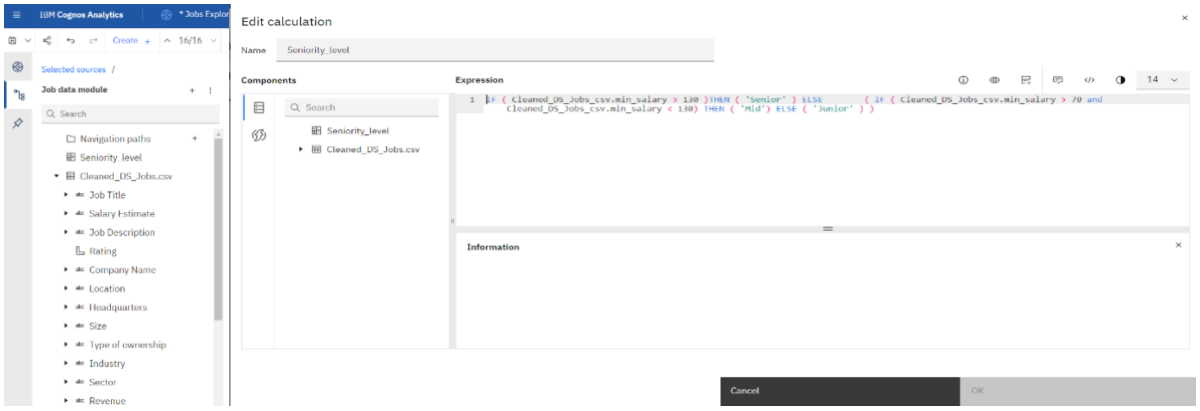
The purpose of this project is to conduct an analysis of Glassdoor job postings to gain insights into current and emerging job market trends, identify in-demand skills and experience, and understand how employers can improve their employer branding and reputation to attract and retain top talent.

Utilization Of Data Filters:

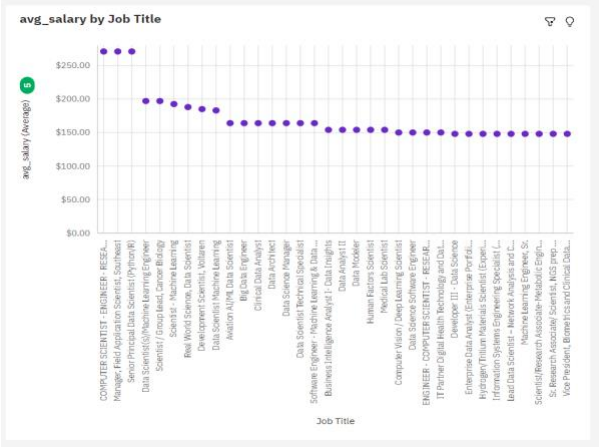
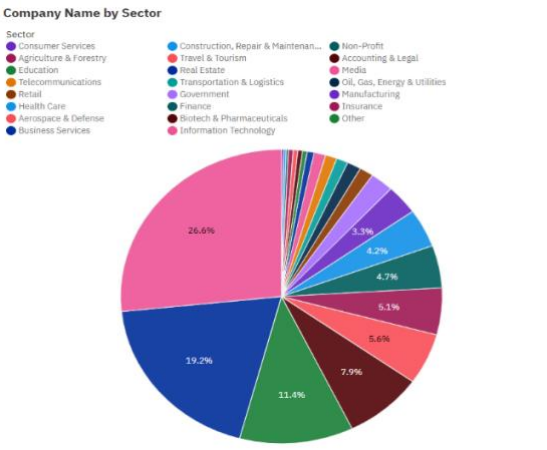
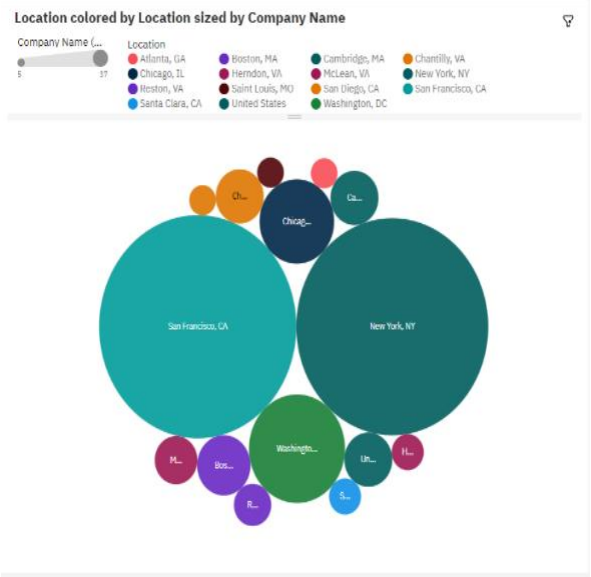
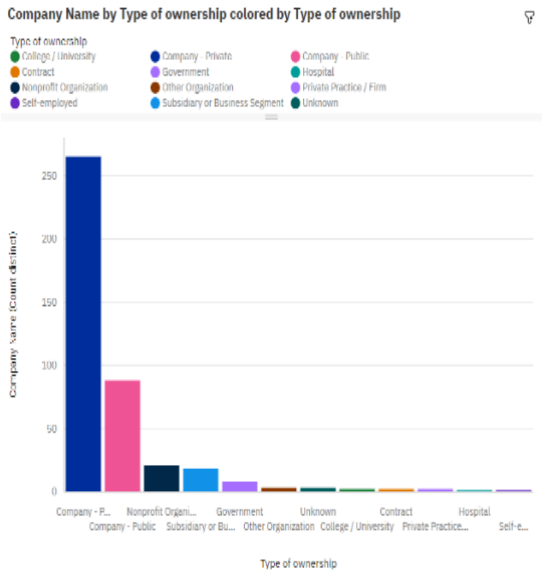


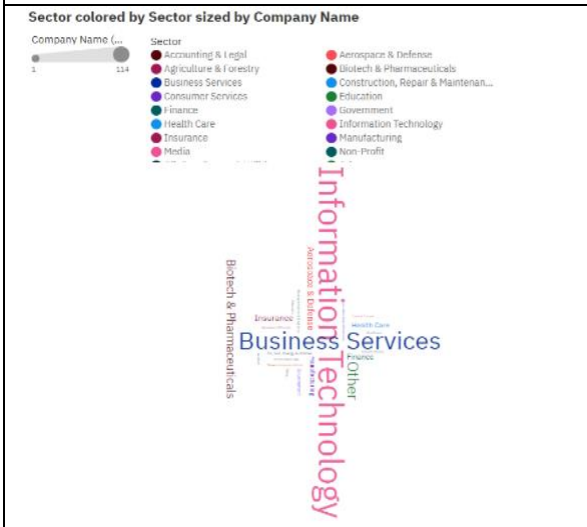
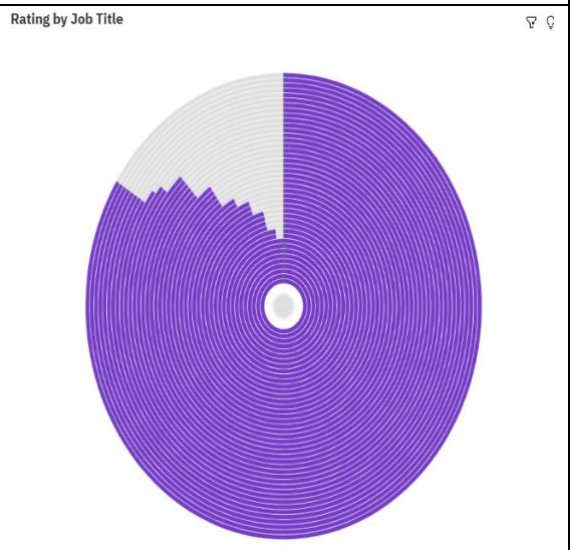
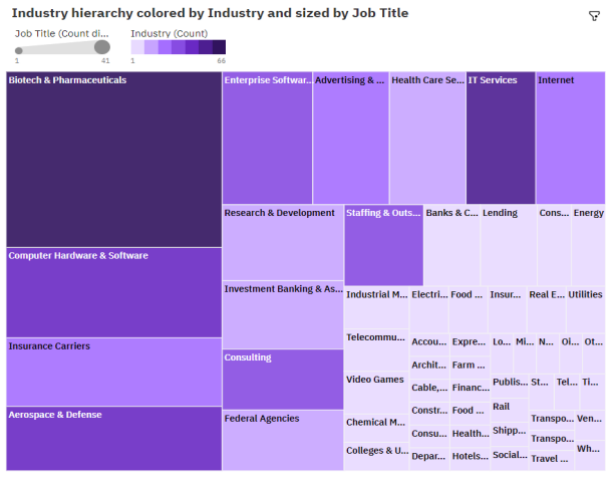
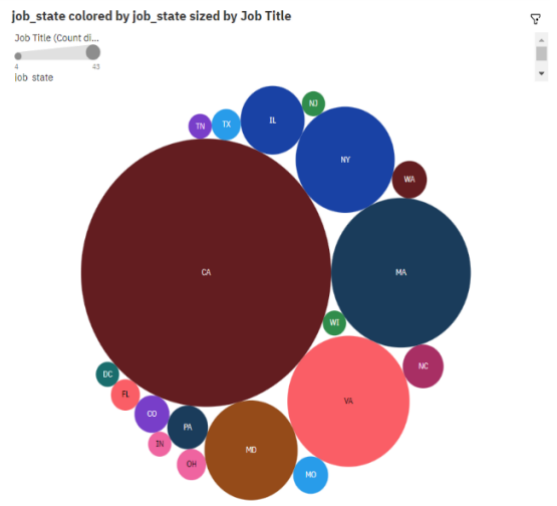


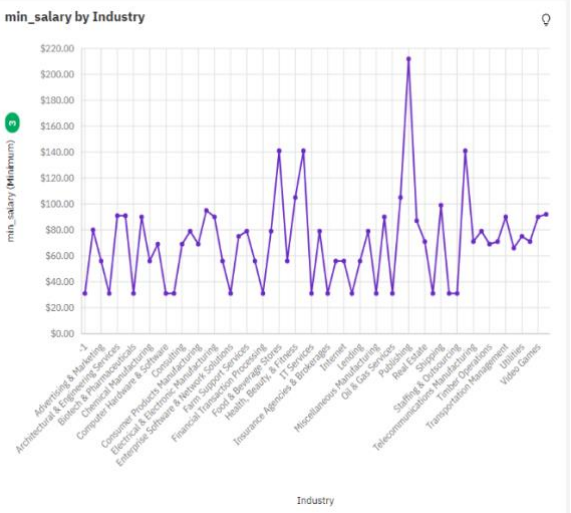
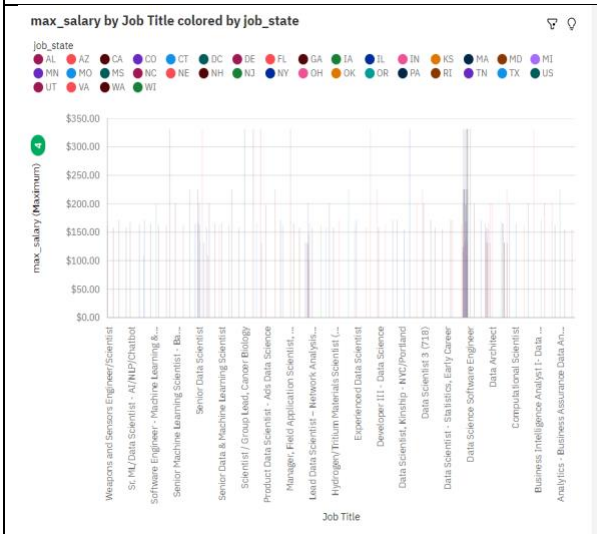
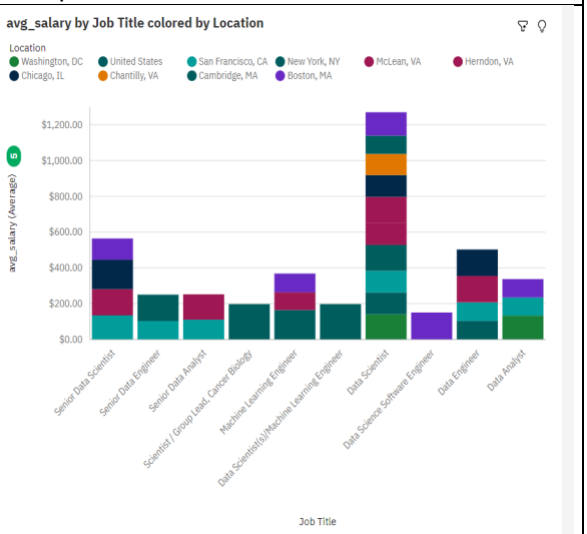
No. Of Calculation fields: Seniority_Level is the calculated column and the calculation is in below given screenshot.



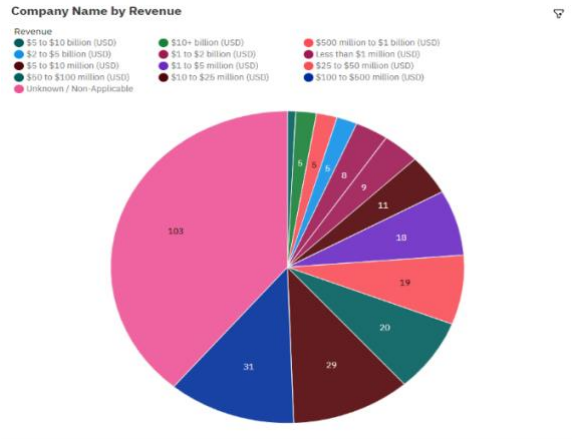
No Of Visualizations/ Graphs:

1. What is salary trend for a particular job title?	2. No of companies belonging to different Sector
	
The average values of **avg_salary** range from 148, occurring when **Job Title** is Developer III - Data Science, to 271, when **Job Title** is COMPUTER SCIENTIST - ENGINEER - RESEARCH COMPUTER SCIENTIST - SIGNAL PROCESSING. Over all **job titles**, the average of **avg_salary** is 169.1.	The total number of results for **Company Name**, across all **sectors**, is 660. Information Technology is the most frequently occurring category of **Sector** with a count of 178 items with **Company Name** values (27 % of the total).
3. How many companies belongs to the particular location?	4. what's the distribution of companies according to the type of owner?
	
The overall number of results for **Company Name** is 303. San Francisco, CA (22.8 %) and New York, NY (16.5 %) are the most frequently occurring categories of **Location** with a combined count of 119 items with **Company Name** values (39.3 % of the total).	The overall number of results for **Company Name** is 633. Company - Private is the most frequently occurring category of **Type of ownership** with a count of 386 items with **Company Name** values (61 % of the total).

5. Most popular sector on Glassdoor for data science domain	6. Top 10 rated jobs
	
The overall number of results for **Company Name** is 660. Information Technology is the most frequently occurring category of **Sector** with a count of 178 items with **Company Name** values (27 % of the total).	Over all **job titles**, the average of **Rating** is 4.6. The average values of **Rating** range from 4.2, occurring when **Job Title** is Computational Behavioral Scientist, to 5, when **Job Title** is Applied AI Scientist / Engineer. Senior Data Engineer is the most frequently occurring category of **Job Title** with a count of 5 items with **Rating** values (6.8 % of the total).
7. Which Industry is offering more job roles	8. Which state is providing more opportunities
	
The total number of results for **Job Title**, across all **industries**, is 589. Biotech & Pharmaceuticals (11.2 %), IT Services (10.2 %), and Computer Hardware & Software (9.3 %) are the most frequently occurring categories of **Industry** with a combined count of 181 items with **Job Title** values (30.7 % of the total).	The overall number of results for **Job Title** is 591. CA is the most frequently occurring category of **job_state** with a count of 165 items with **Job Title** values (27.9 % of the total).

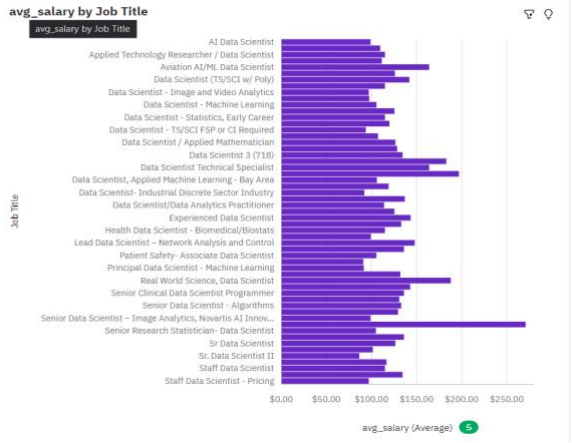
9. Show job titles from different category	10. Compare salary trend of different industry
job_simp and Job Title job_simp / Job Title data engineer: Data Science All Star Program - Data Engineer Track; Senior Data Engineer; Software Data Engineer; Sr Data Engineer (Sr BI Developer); Staff BI and Data Engineer; Tableau Data Engineer 20-0117 data scientist: AI Data Scientist; AI Ops Data Scientist; Applied Technology Researcher / Data Scientist; Associate Data Scientist; Aviation AI/ML Data Scientist; Data Scientist; Data Scientist (TS/SCI w/ Poly); Data Scientist (TS/SCI); Data Scientist - Image and Video Analytics; Data Scientist - Intermediate; Data Scientist - Machine Learning	 The total number of results for **min_salary**, across all **industries**, is 660. The smallest value of **min_salary** is 31, occurring when **Industry** is -1. -1 (10.8 %), Biotech & Pharmaceuticals (10 %), IT Services (9.1 %), and Computer Hardware & Software (8.3 %) are the most frequently occurring categories of **Industry** with a combined count of 252 items with **min_salary** values (38.2 % of the total).
11. Top 10 highest paying jobs from different states	12. what is the avg_salary of Job Title belonging from particular Location
 The largest value of **max_salary** is 331 and occurs in Senior Principal Data Scientist (Python/R) and CA. CA is the most frequently occurring category of **job_state** with a count of 139 items with **max_salary** values (25.9 % of the total). Data Scientist is the most frequently occurring category of **Job Title** with a count of 333 items with **max_salary** values (62.1 % of the total).	 Over all **job titles** and **locations**, the average of **avg_salary** is 130.1. The average values of **avg_salary** range from 99 to 197. Data Scientist is the most frequently occurring category of **Job Title** with a count of 157 items with **avg_salary** values (78.5 % of the total).

13. What is the revenue generated by companies that falls under particular category of ownership



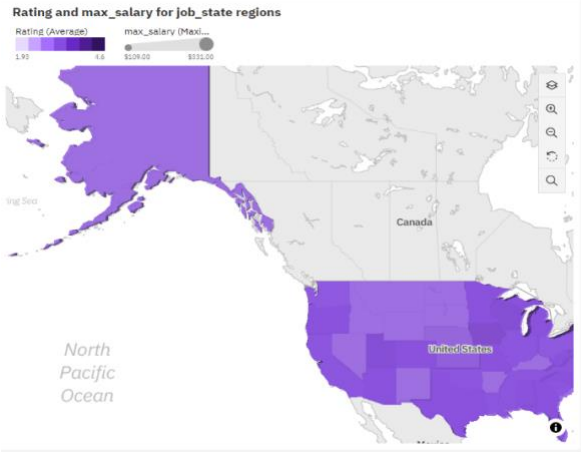
The total number of results for **Company Name**, across all **revenues**, is 386.
Unknown / Non-Applicable is the most frequently occurring category of **Revenue** with a count of 152 items with **Company Name** values (39.4 % of the total).

14. Salary trend for different job titles belonging to particular job category



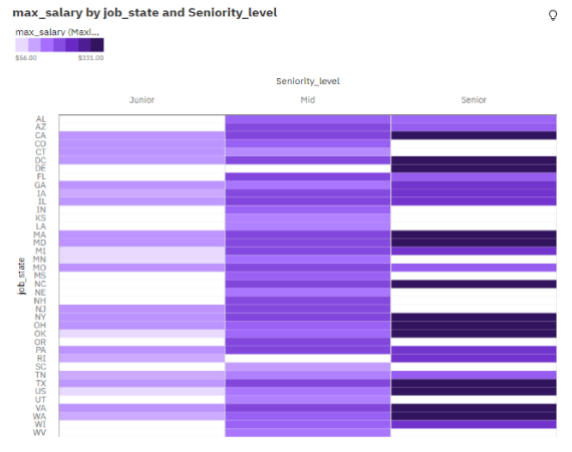
Over all **job titles**, the average of **avg_salary** is 125.3.
The average values of **avg_salary** range from 86.5, occurring when **Job Title** is Sr. Data Scientist II, to 271, when **Job Title** is Senior Principal Data Scientist (Python/R).
avg_salary is unusually high when **Job Title** is Senior Principal Data Scientist (Python/R).
Data Scientist is the most frequently occurring category of **Job Title** with a count of 333 items with **avg_salary** values (74.5 % of the total).

15. What is the rating & salary offered in particular job state



Over all values of **job_state**, the average of **Rating** is 3.587.
The average values of **Rating** range from 1.933, occurring when **job_state** is NE, to 4.6, when **job_state** is IA.
CA is the most frequently occurring category of **job_state** with a count of 165 items with **Rating** values (25 % of the total).

16. Salary trend in different states for particular seniority level



The largest value of **max_salary** is 331 and occurs in CA and Senior.

The total number of results for **max_salary**, across all **job_state**, is 660.

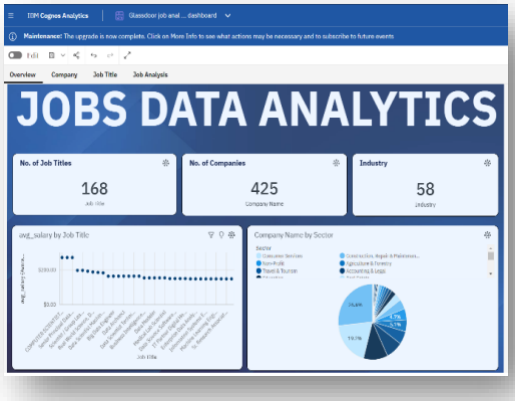
CA is the most frequently occurring category of **job_state** with a count of 165 items with **max_salary** values (25 % of the total).
Mid is the most frequently occurring category of **Seniority_level** with a count of 467 items with **max_salary** values (70.8 % of the total).

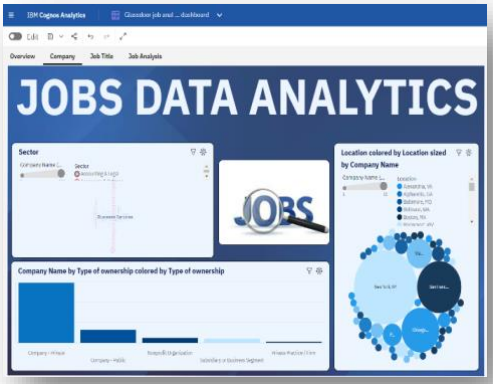
Dashboard:

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.

Jobs Data Analytics Dashboard:

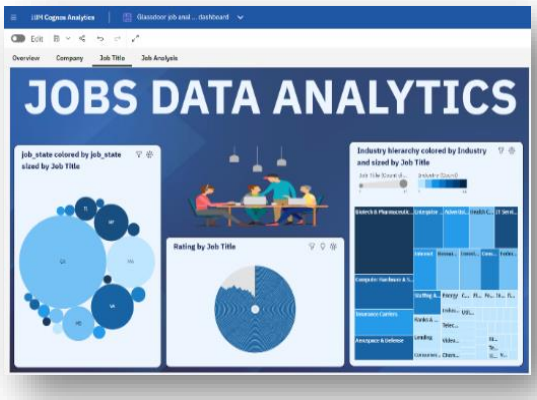
This Dashboard is created by using the visualizations from the job explorations in order to show the data in clear and analysed form.

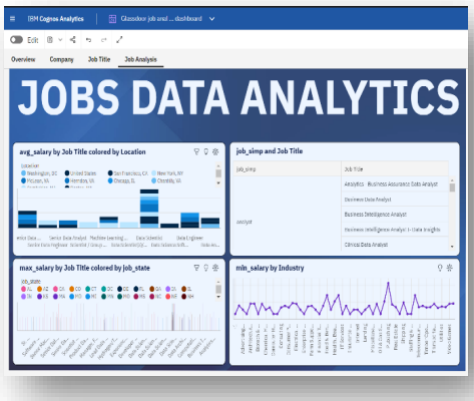




Insights:

ENGINEER - COMPUTER SCIENTIST - RESEARCH COMPUTER SCIENTIST - SIGNAL PROCESSING - SAN ANTONIO OR (**8.7** %) and **Data Science Software Engineer** (**8.7** %) are the most frequently occurring categories of **Job Title** with a combined count of **8** items with **avg_salary** values (**17.4** % of the total).





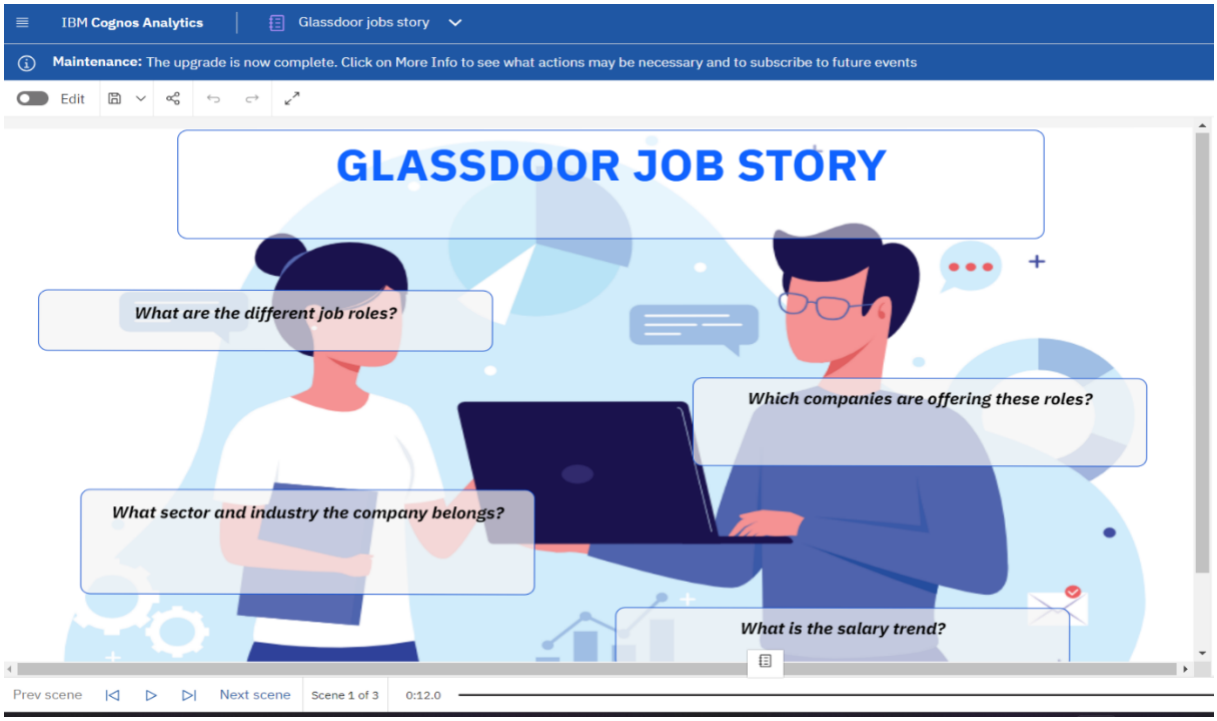
Insights:

Job Title Applied AI Scientist / Engineer has the highest **Average Rating** but is ranked #**22** in **Average avg_salary**.

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.

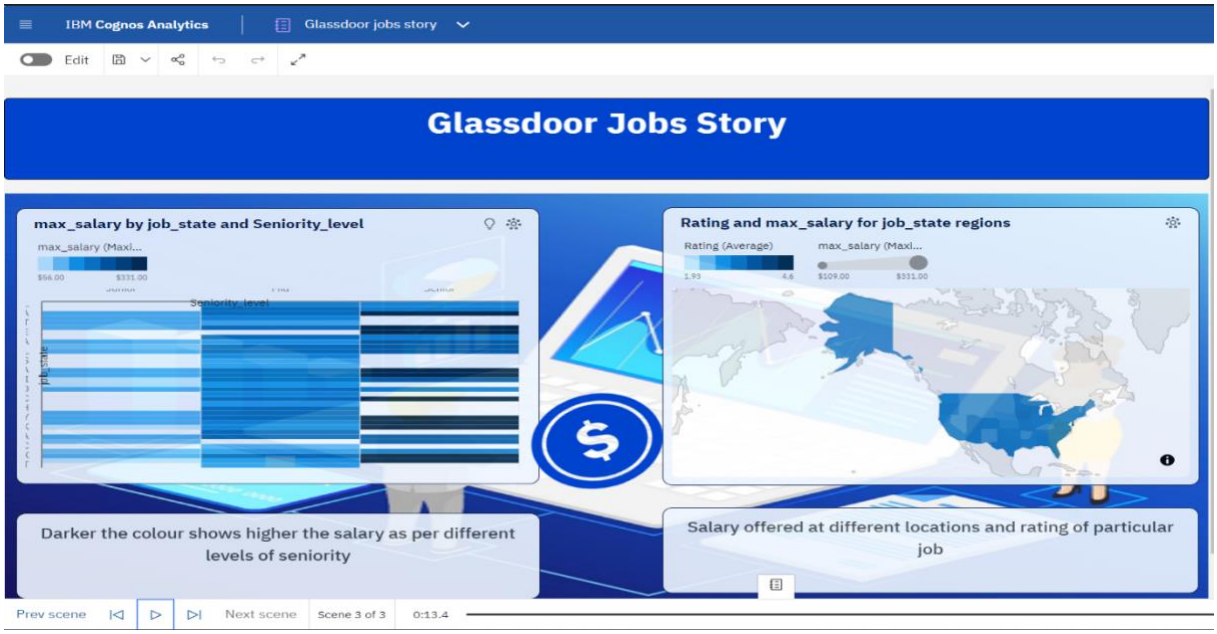
Scene-1:



This story is created without any visualizations to analyse the data in Glassdoor jobs.

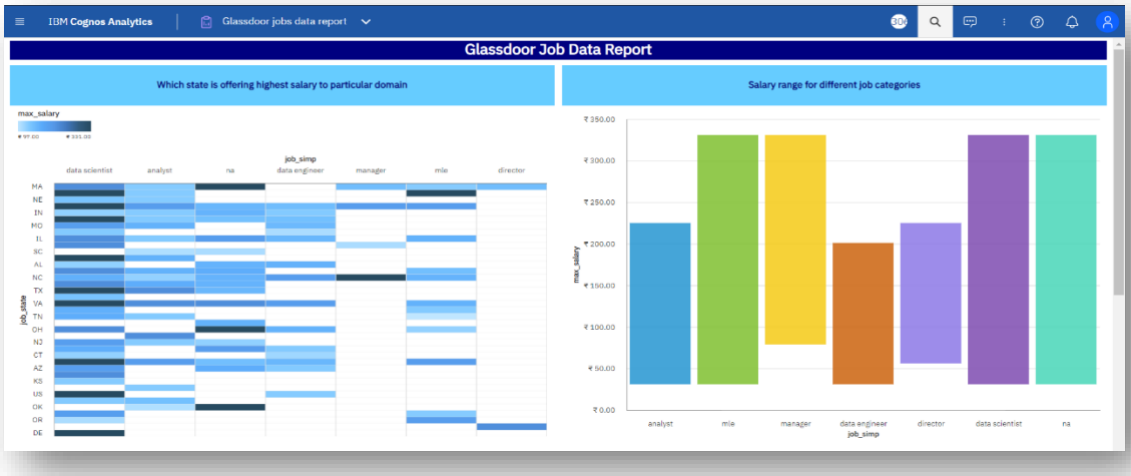
Scene-2: Salary trends per market standards.

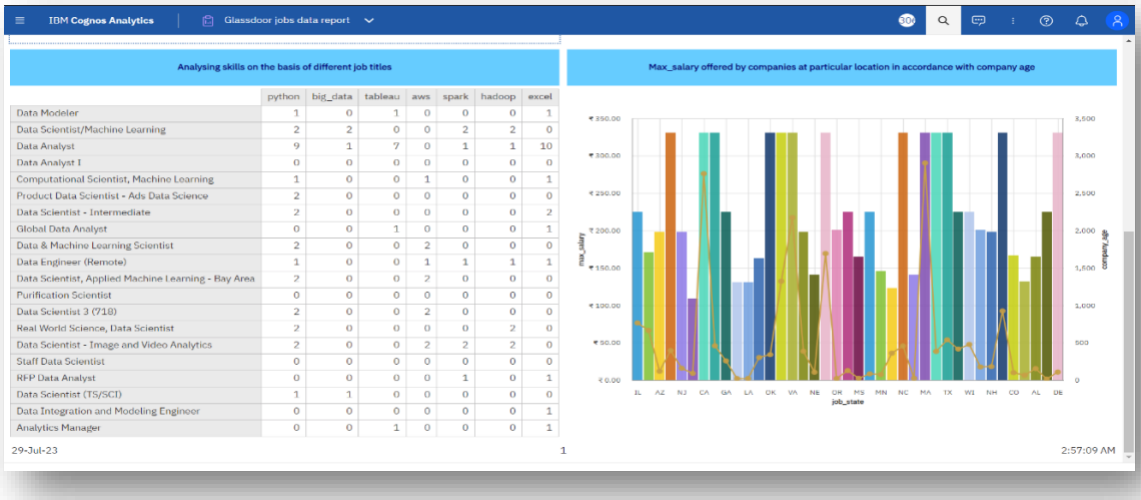
SCENE-3: SALARY BY SENIORITY LEVEL.



REPORT:

When creating a report in Cognos, it is often helpful to include visualizations to help communicate the findings of the analysis.





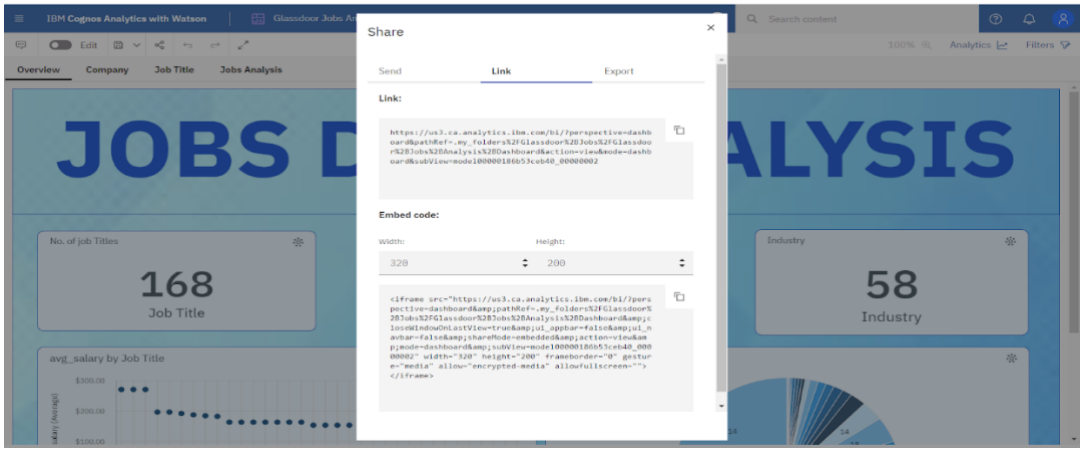
Web Integration:

Publishing helps us to track and monitor key performance metrics, to communicate results and progress. help a publisher stay informed, make better decisions, and communicate their performance to others.

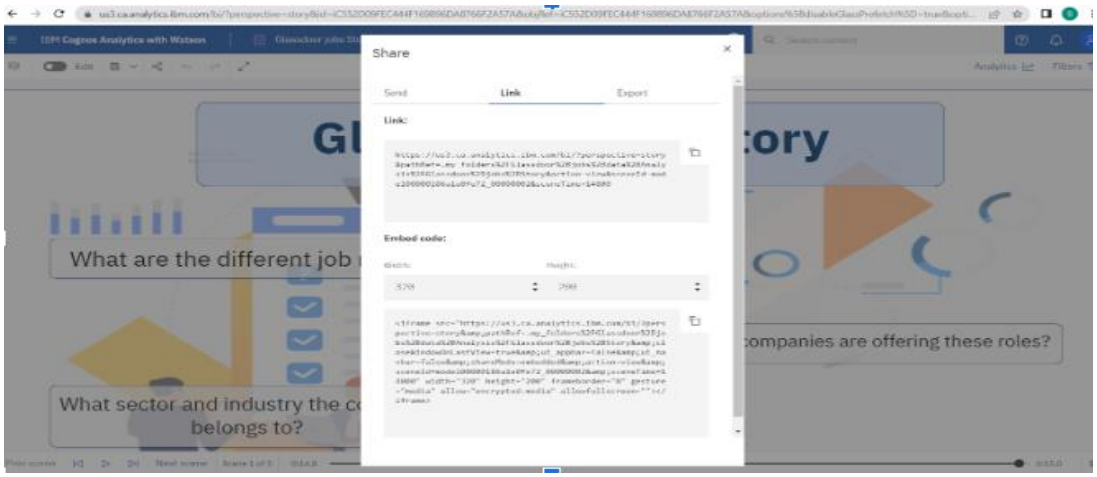
Publishing dashboard, report & story.

Go to Dashboard, report & /story, click on share button on the top.

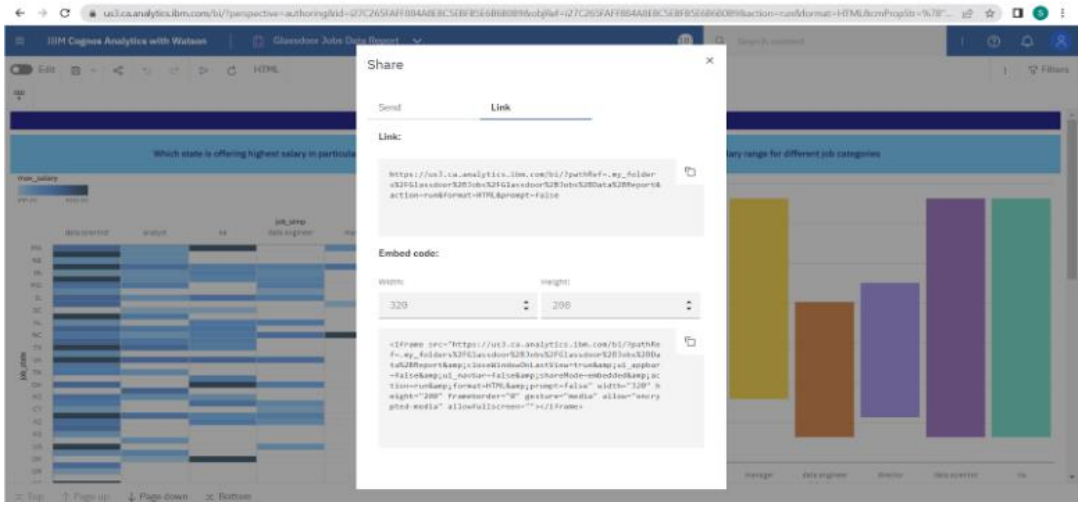
Dashboard



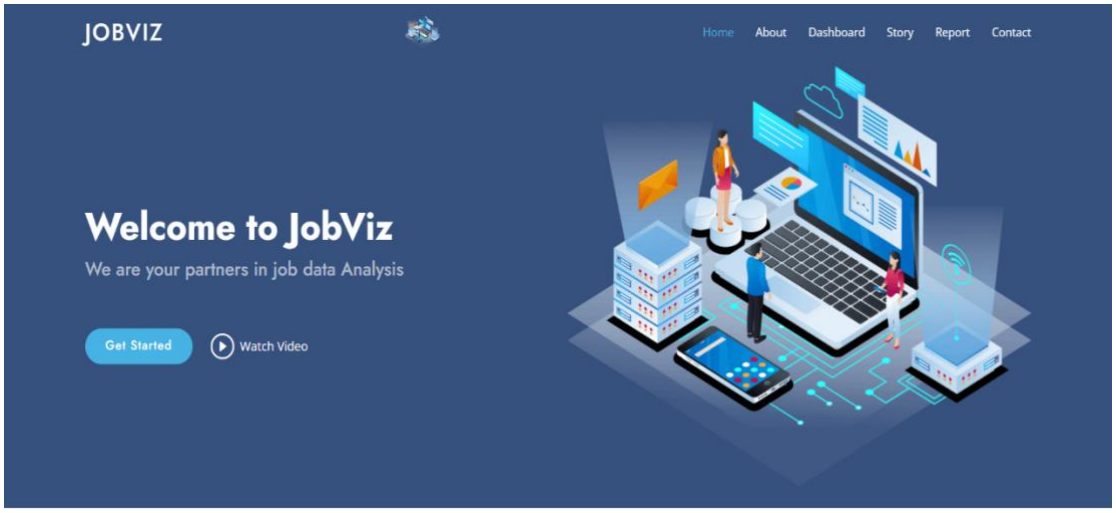
Story

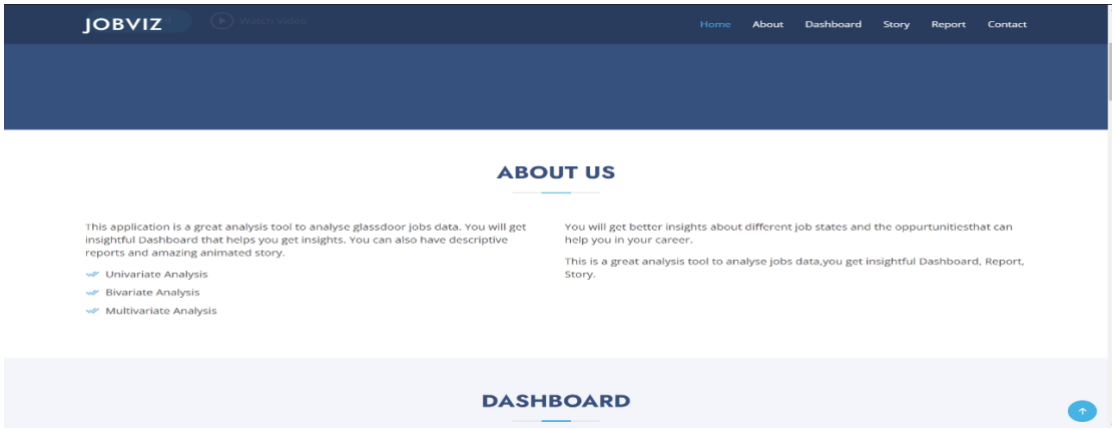


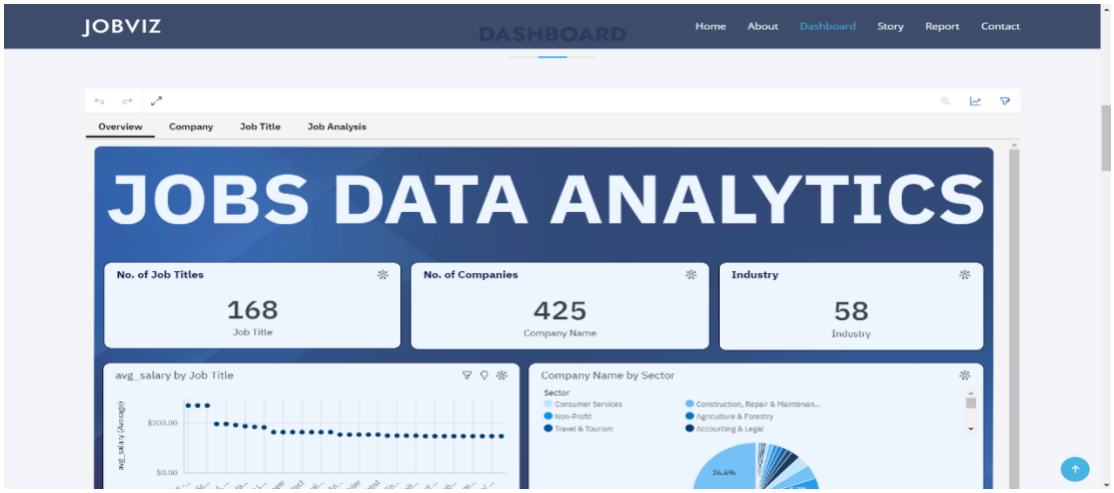
Report

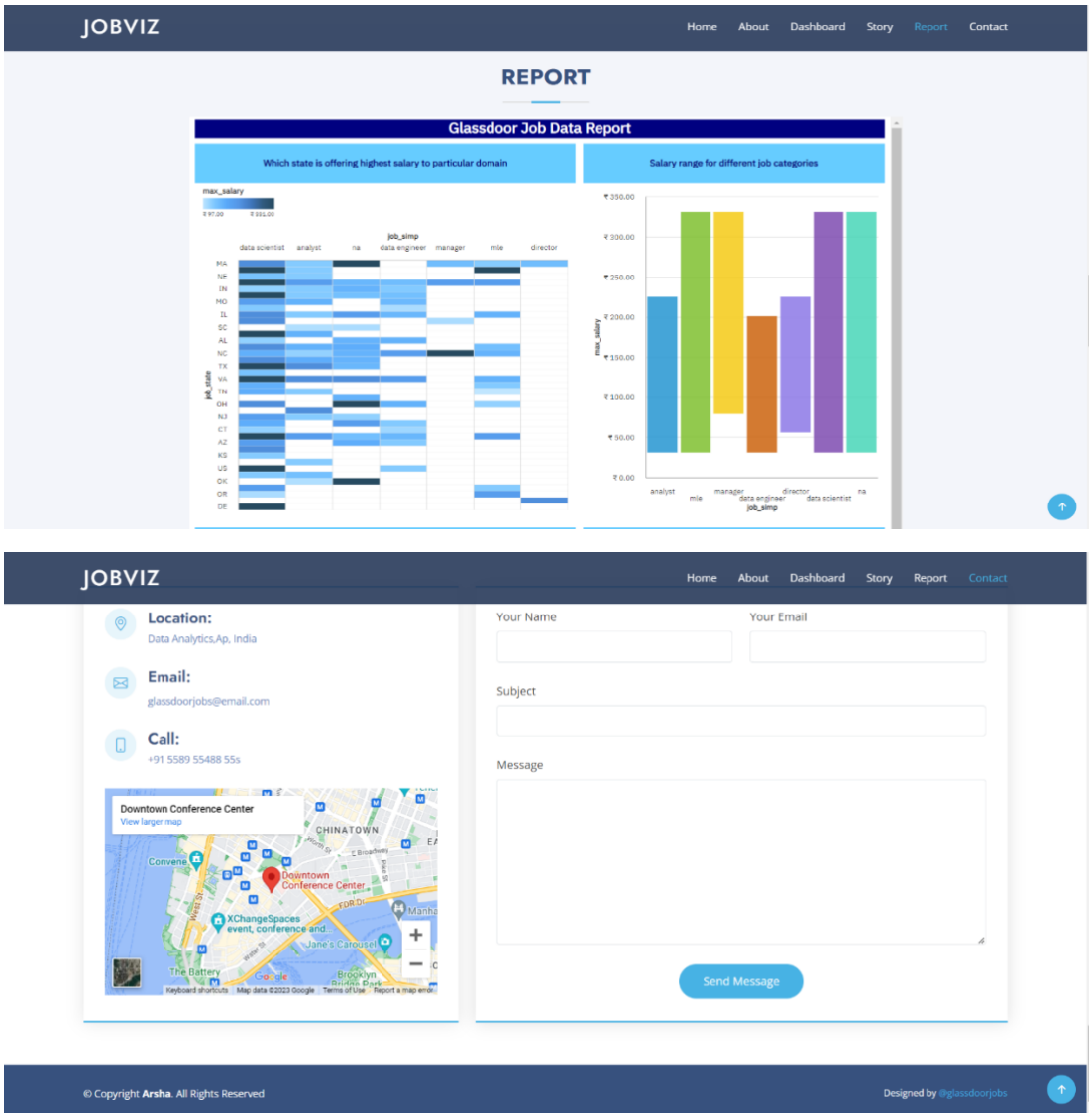


Dashboard, Report, Story Embed With UI Using Flask:









5.Advantages and Disadvantages:

Advantages:

➡It detects and correct the errors from data sets with the help of data cleansing.

➡It removes duplicate information from data sets and hence saves large amount of memory space. This decrease cost to the company.

➡It helps in displaying relevant advertisements on the online shopping websites based on historic data and purchase behaviour of the users. Machine learning algorithms are applied for the same. This helps in increasing revenue and productivity of the companies.

➡It reduces banking risks by identifying probable fraudulent customers based on historic data analysis. This helps institutes in deciding whether to issue loan or credit cards to the applicants or not.

➡It is used by security agencies for surveillance and monitoring purpose based on information collected by huge number of sensors. This helps in preventing any wrongdoings and/or calamities.

Disadvantages:

➡This may breach privacy of the customers as their information such as purchases, online transactions, subscriptions are visible to their parent companies. The companies may exchange these useful customer databases for their mutual benefits.

➡The cost of data analytics tools vary based on applications and features supported. More over some of the data analytics tools are complex to use and require training. This increases cost to the company willing to adopt data analytics tools or software.

➡The information obtained using data analytics can also be misused against group of people of certain country or community or caste.

➡It is very difficult to select the right data analytics tools. This is due to the fact that it requires knowledge of the tools and their accuracy in analysing the relevant data as per applications. This increases time and cost to the company.

6.Applications:

➡ By using Data analysis, the number of jobs that are giving opportunities to employees can be classified and can be displayed.

➡Can determine which country giving more jobs and more salaries to the employees.

➡Employees can study which job is the perfect match for their skill set and can land in their dream job.

➡ Can determine which country is best in hiring new employees every year.

➡Can determine which company started in which year.

➡ Can determine the number of companies increasing or decreasing over the years.

7.Conclusion:

The main objective of this study was to analyse and visualise the various factors which have contributed to the Evolution of the Glassdoor jobs over the years. This type of analysis is very helpful as this type of analysis can be performed by any employer which can help them in analysing their performance so that they can find their jobs by challenging their strategies.

We have used a technique named Exploratory Data Analysis which enables you to encapsulate the primary factors of a dataset into a visual format. We selected Python language to implement our work because it is one of the best languages suitable for Data Analysis and is the platform where we have performed this Analysis. As a result of the Analysis, we can conclude that it is true that Glassdoor jobs have creating considerable platform to the employees to land into their dream job.

8.Future Scope:

We all know that any Analysis is not perfect and it consists of some limitations which define the Future scope of the Research Work. This project work also contains some limitations which we are considering as the Future Scope of the Project. These are ASCI-2020 IOP Conf. Series: Materials Science and Engineering We have visualised our data only in Graphical format. We can also describe the data in other formats like Geographical format where we can depict the countries on the World map.

Till now we have only performed Data Analysis using Exploratory Data Analysis. We can also apply various Machine Learning Algorithms to the data set after Analysis and can create a Predictive Model which can predict the statistics of the Future jobs in Glassdoor jobs.