The Future of Work: Data Analysis of Glassdoor Jobs

DEVAMITHRA NARESH [20BCT0156, devamithra.naresh2020@vitstudent.ac.in]

DHANANJAY K PRASAD[20BDS0182, dhananjayprasad.k@vitstudent.ac.in]

KUSHAGRA[20BIT0112,Kushagra.2020@vitstudent.ac.in]

ASWIN ANIL[20BCB0104,aswin.anil2020@vitstudent.ac.in]

The Future of Work: Data Analysis of Glassdoor Jobs

"The Future of Work: Data Analysis of Glassdoor Jobs" is a data analysis project aimed at gaining insights into the evolving landscape of work and employment trends using data from Glassdoor, a prominent online platform for job information and employee reviews. The project focuses on leveraging advanced data analysis techniques to uncover patterns, trends, and key factors influencing the job market, company performance, and employee satisfaction.

Technical Architecture:

Project Flow

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

Define Problem / Problem Understanding

- Specify the business problem
- Business requirements

- Literature Survey
- Social or Business Impact.

Data Collection & Extraction from Database

- Collect the dataset,
- Connect IBM DB2 with IBM cognos

Data Preparation

- Prepare the Data for Visualization

Data Visualizations

- No of Unique Visualizations

Dashboard

- Responsive and Design of Dashboard

Story

- No of Scenes of Story

Report

- Creating a report

Performance Testing

- Amount of Data Rendered to DB ‘
- Utilization of Data Filters
- No of Calculation Fields
- No of Visualizations/ Graphs

Web Integration

- Dashboard and Story embed with UI With Flask

Project Demonstration & Documentation

- Record explanation Video for project end to end solution
- Project Documentation-Step by step project development procedure

The Business Problem

"The Future of Work: Data Analysis of Glassdoor Jobs" project utilizes data analysis techniques to explore and understand the evolving world of work, providing actionable insights that have the potential to shape the future of employment and contribute to informed decision-making in the job market. By analyzing a vast amount of data available on Glassdoor, including job listings, salary information, company reviews, and employee ratings, this project aims to identify emerging job roles, in-demand skills, and industry-specific trends. Additionally, the project seeks to understand the impact of technological advancements, such as automation and artificial intelligence, on the job market and workforce dynamics.

Business Requirements

The business requirements for this project would likely include

Data collection:

The first requirement is to collect data from Kaggle that is relevant to the Company name, Job Title, Salary, Salaries reported, Location, Employment Status, Job roles, and rating

Data cleaning and preparation:

The collected data must be cleaned and processed to ensure it is suitable for analysis. This may involve removing irrelevant information, correcting inconsistencies and missing values, and transforming the data into a format that is compatible with the analysis tools.

Data analysis:

The data must be analyzed to uncover meaningful insights. This could involve using techniques such as descriptive statistics, regression analysis and data visualization to gain a deeper understanding of the data.

Report creation:

The insights and findings from the data analysis must be presented in a comprehensive report that includes visualizations and data tables. The report must be well organized and easy to understand, with clear and concise explanations of the results.

Literature Survey

The literature survey emphasized the significance of Glassdoor as a valuable data source for job market analysis. Researchers have utilized Glassdoor's data to study salary disparities, job satisfaction, and employer branding. The availability of rich data on job listings, company reviews, and employee ratings on Glassdoor has facilitated comprehensive analyses and provided insights into various aspects of the job market.

The literature survey analysis demonstrated the relevance and importance of the project "The Future of Work: Data Analysis of Glassdoor Jobs." It highlighted the need for data analysis in understanding job market trends, the impact of technology on employment, and employee satisfaction.

Social Or Business Impact

Social Impact:

The project provides valuable insights for job seekers, helping them make informed decisions regarding career paths, job opportunities, and skill development. By analyzing job market trends, salary dynamics, and employee satisfaction factors, individuals can align their career goals with market demands, ultimately improving their employment prospects and job satisfaction.

Business Model/Impact:

Analyzing Glassdoor data can provide valuable insights to companies regarding the factors that attract and retain top talent. Employers can leverage this information to refine their talent acquisition strategies, employer branding efforts, and employee retention programs, leading to improved recruitment outcomes and reduced turnover rates.

Data Collection & Extraction from Database

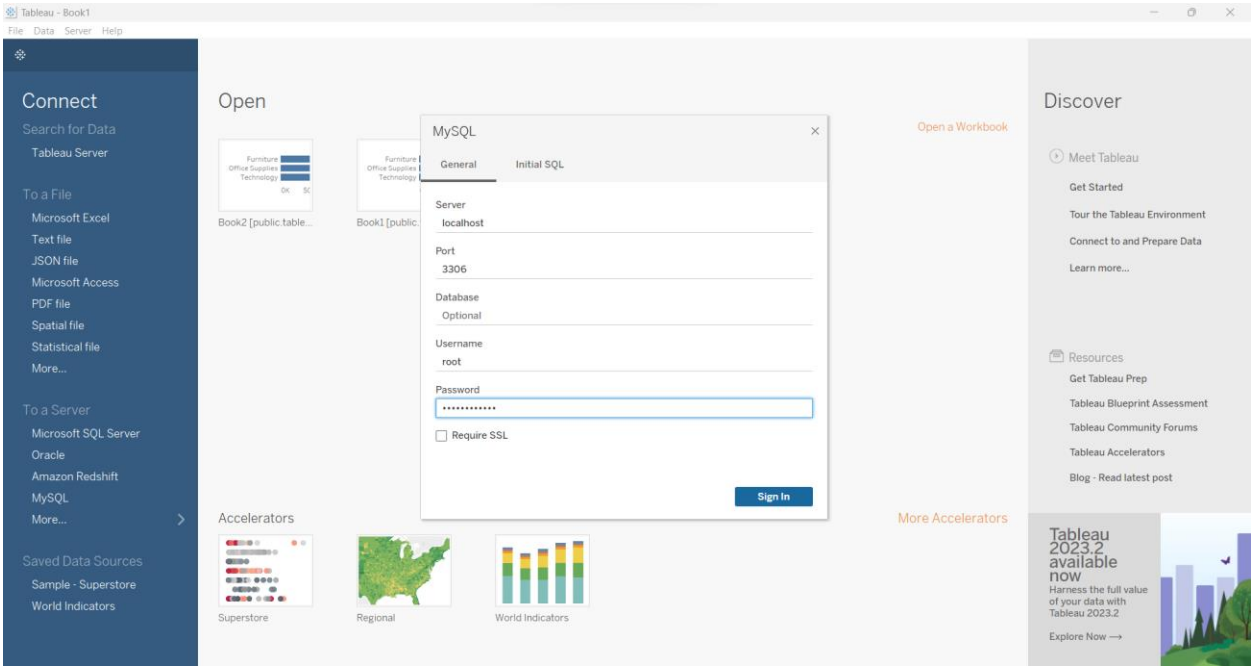
Data collection is the process of gathering and measuring information on variables of interest in an established, systematic fashion that enables one to answer stated research questions, test hypotheses, evaluate outcomes, and generate insights from the data.

Collect The Dataset

Activity 1.1: Understand the data

Check out the below link to understand the dataset in detail:
https://drive.google.com/file/d/1qBEw4_AfqozL73NyCkQW3ZvFOhok6d-_/view

Activity 2: Connect MySQL and Tableau with the dataset.



Data Preparation

Data preparation for Tableau involves the process of organizing, cleaning, and transforming raw data into a format that can be effectively visualized and analyzed within the Tableau software. This includes tasks such as data cleaning, data integration, data formatting, and data aggregation. The goal is to ensure that the data is accurate, consistent, and structured in a way that enables meaningful insights and visualizations in Tableau.

Prepare The Data For Visualization

Preparing the data for visualization involves cleaning the data to remove irrelevant or missing data, transforming the data into a format that can be easily visualized, exploring the data to identify patterns and trends, filtering the data to focus on specific subsets of data, preparing the data for visualization software, and ensuring the data is accurate and complete. This process helps to make the data easily understandable and ready for creating visualizations to gain insights into the performance and efficiency.

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 datasets 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 (Filters Applied)

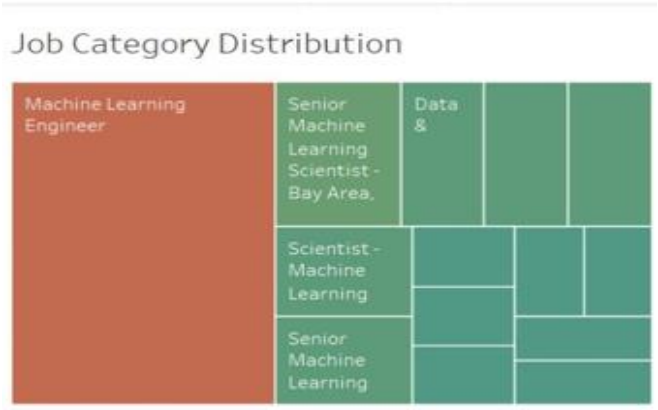
The number of unique visualizations that can be created with a given dataset. Some common types of visualizations that can be used to analyze the data include bar charts, line charts, heat maps, scatter plots, pie charts, maps, etc. These visualizations can be used to compare performance, track changes over time, show distribution, and show relationships between variables.

VISUALIZATIONS IN TABLEAU

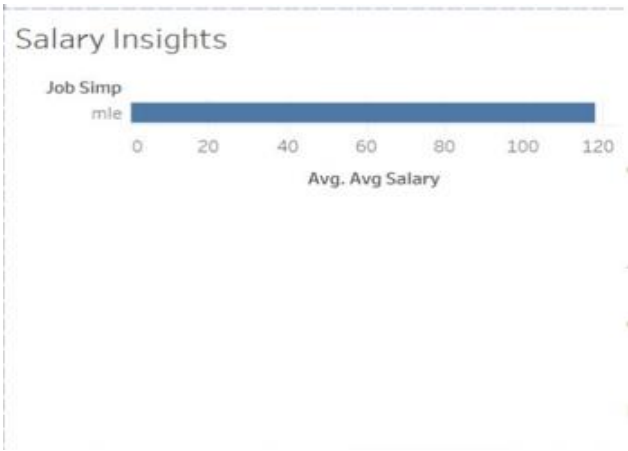
Activity 1.1: Company ratings

Activity 1.2: Company openings distribution

Activity 1.3: Job category distribution

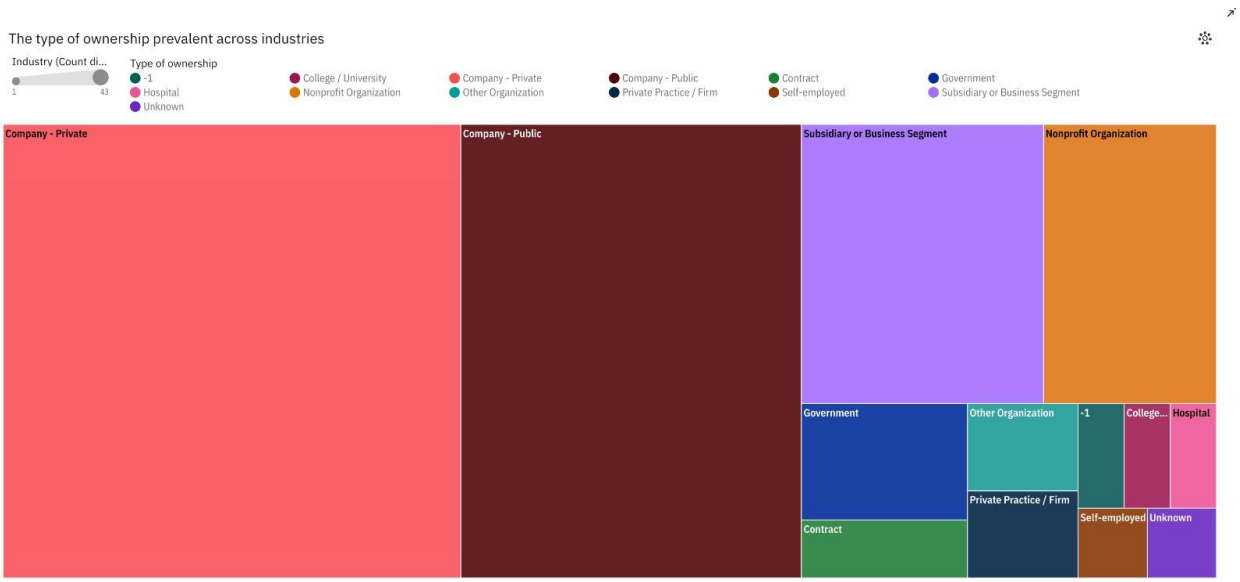


Activity 1.4: Salary insights

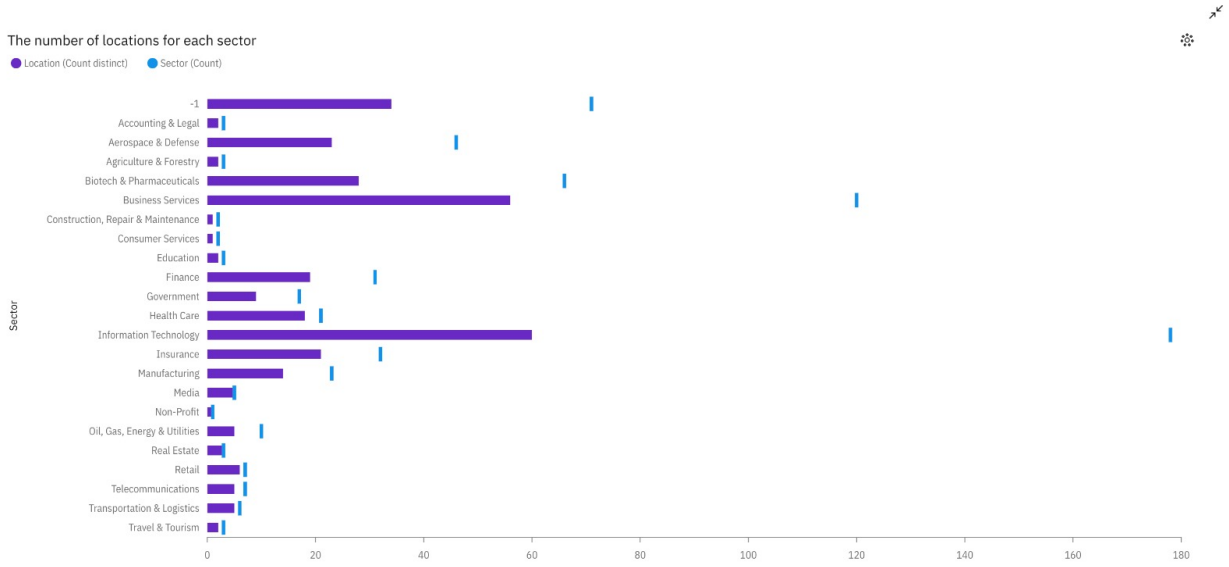


VISUALIZATIONS IN IBM COGNOS

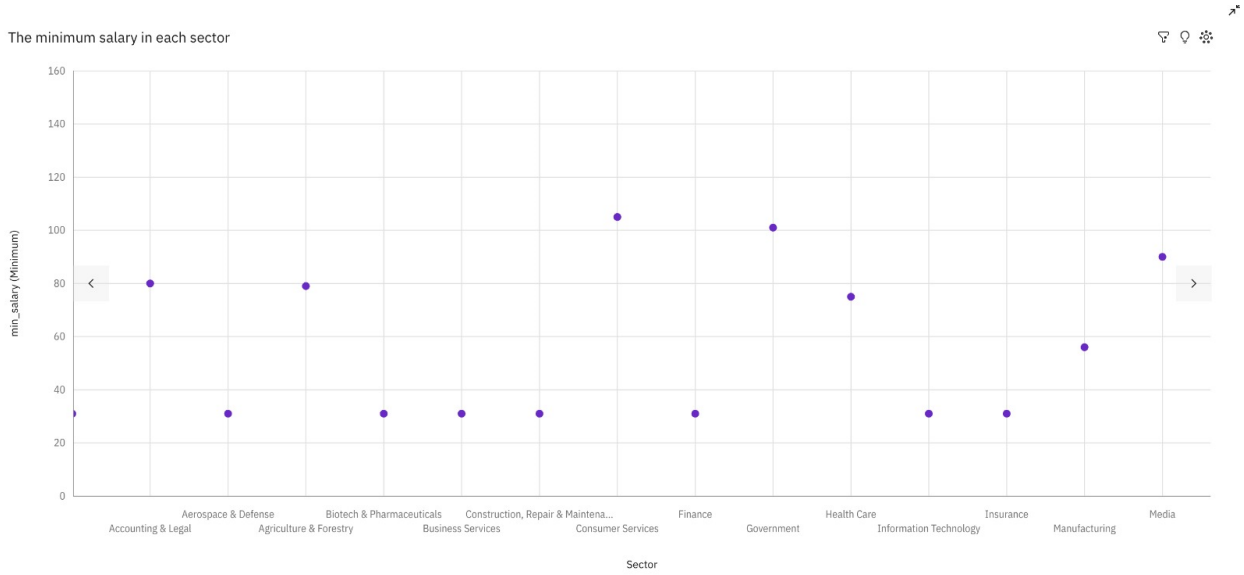
Activity 1.1:



Activity 1.2:



Activity 1.3:



Activity 1.4:

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.

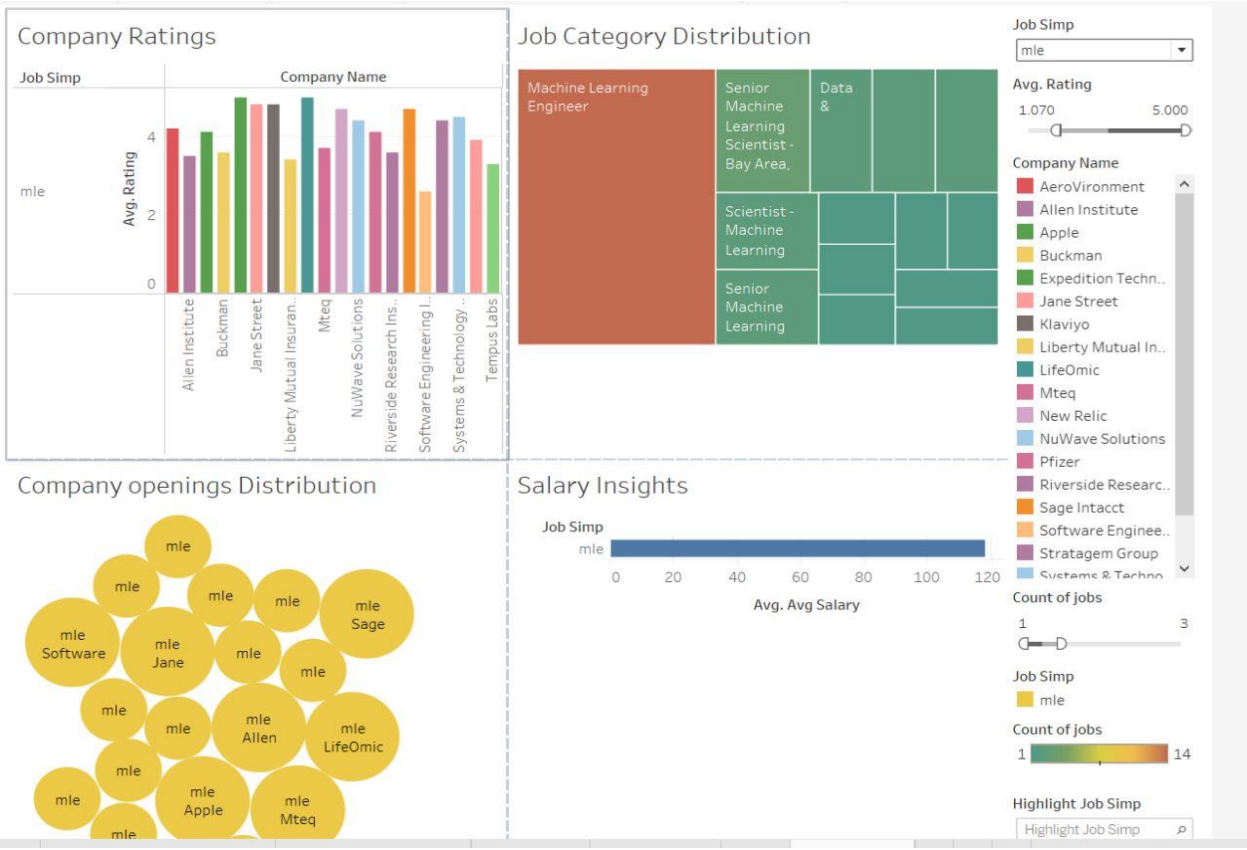
Dashboards can be used in a variety of settings, such as business, finance, manufacturing, healthcare, and many other industries. They can be used to track key performance indicators (KPIs), monitor performance metrics, and display data in the form of charts, graphs, and tables.

Responsiveness And Design of Dashboard

The responsiveness and design of a dashboard for analyzing the factors important for A comprehensive analysis of the IT sector's salaries and roles analyzes various engagement metrics such as Company name, Job Title, Salary, Salaries reported, Location, Employment Status, Job roles, and Rating.

TABLEAU DASHBOARD

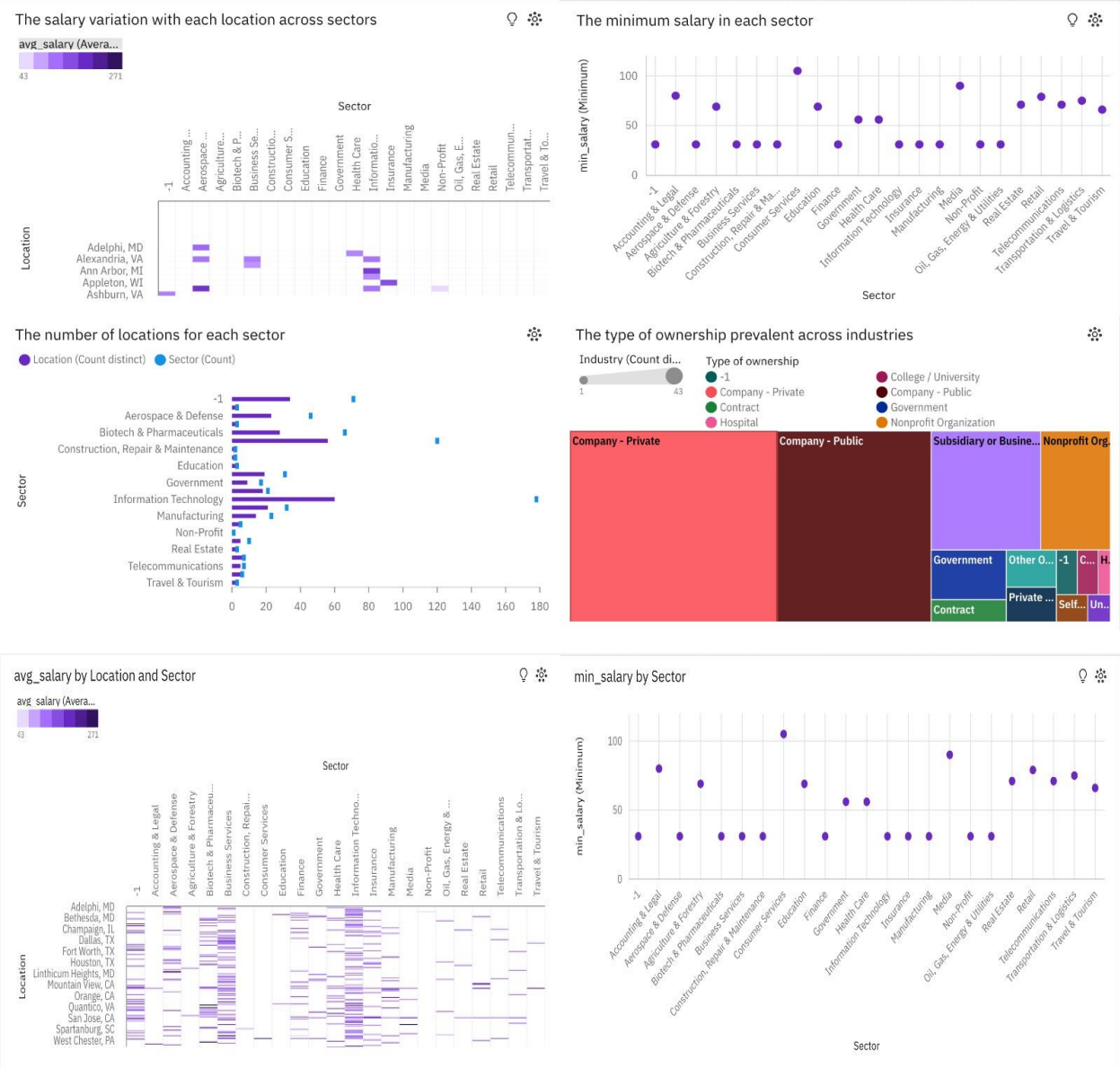
Explanatory Video link:
https://www.loom.com/share/e1181694546f461bac4c14780498e726?sid=02e51936-9f76-4c88-836f-65157bb67390



IBM COGNOS DASHBOARD

Explanation video link:

https://www.loom.com/share/48d5895026db4430849e376a114d2427?sid=63ac105c-07a4-436a-904d-88455e73dae1



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.

Performance Testing

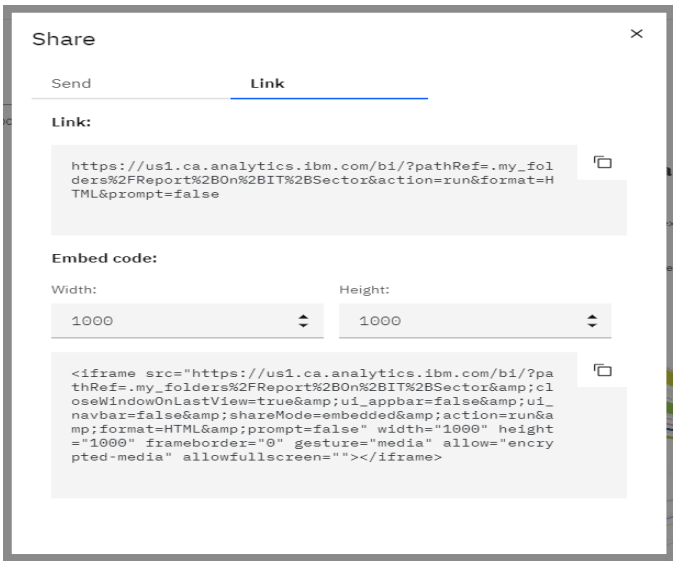
Performance testing for Tableau focuses on evaluating the software's speed, responsiveness, and scalability under various conditions and workloads. It involves measuring and analyzing key performance indicators such as query response time, data loading speed, dashboard rendering time, and concurrent user handling capacity. The testing process helps identify any performance bottlenecks, optimize system configurations, and ensure that Tableau can handle the expected workload efficiently, providing users with a smooth and responsive experience while working with large datasets and complex visualizations.

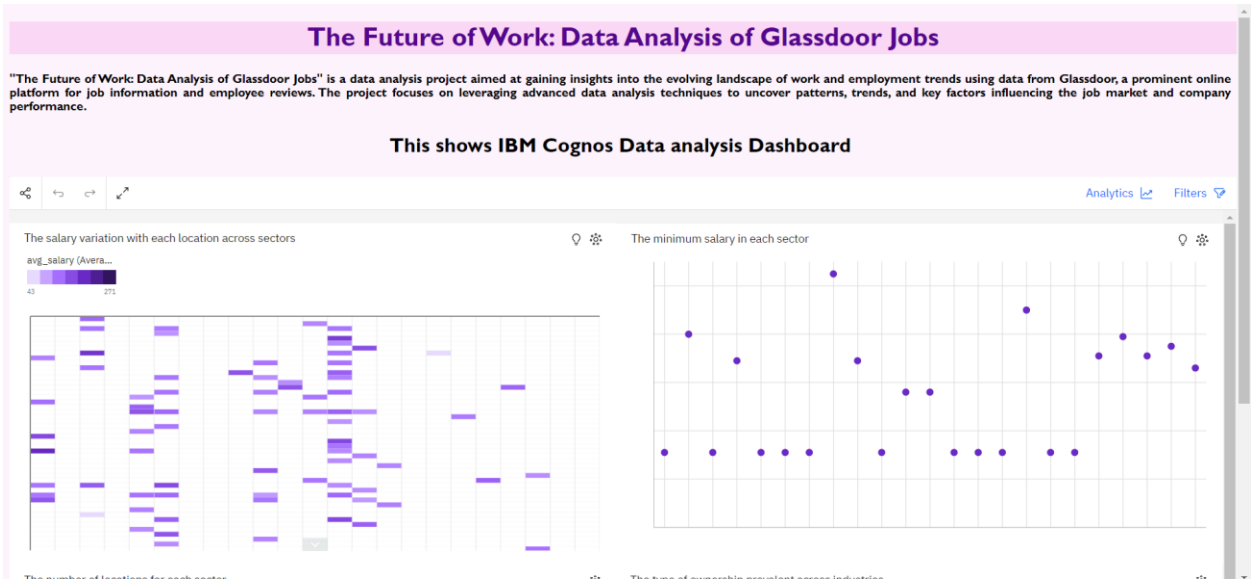
Web Integration

Publishing helps us 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.

Integrating dashboard/reports/stories to web

Step 1: Go to Dashboard/story/report, click on share button on the top ribbon





Activity 1: Integrating with Tableau Public

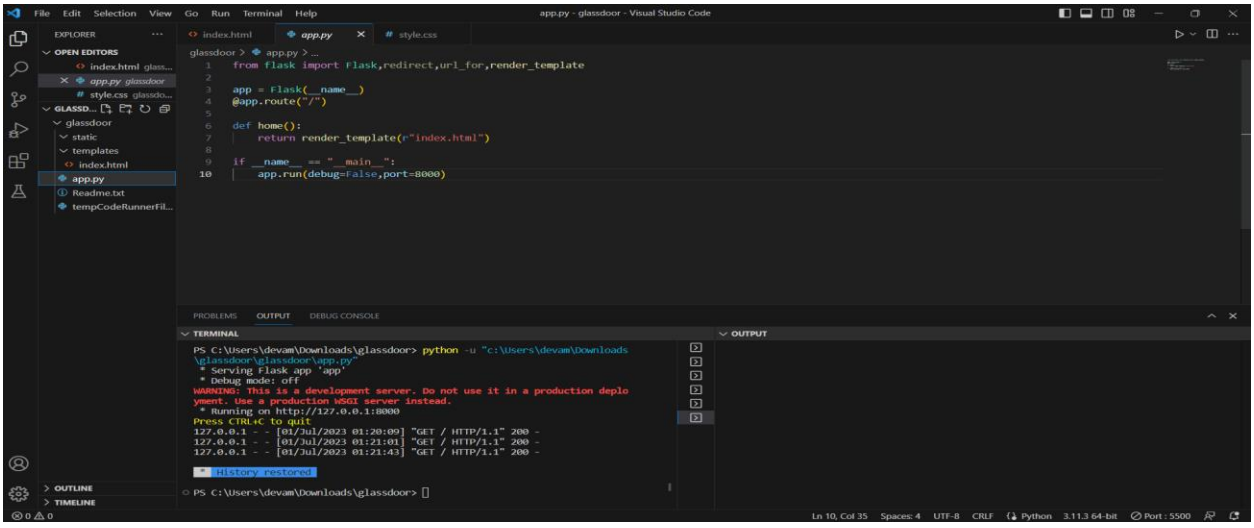
Explanatory video:

https://www.loom.com/share/e1181694546f461bac4c14780498e726?sid=02e51936-9f76-4c88-836f-65157bb67390

Activity 2: Integrating with bootstrap website and Implementing Flask

Explanatory video:

https://www.loom.com/share/fbb5915aa4ad4f6d9c8ac305ca34bc5a?sid=b8afb6bd-ec41-4991-9513-cea341d45c17



CONCLUSION:

In conclusion, the data analysis of jobs in the Glassdoor dataset holds significant potential to unlock valuable insights and contribute to various aspects of the job market, employment landscape, and business decision-making. By leveraging advanced data analysis techniques and utilizing the rich data available on Glassdoor, this project can yield numerous benefits.

Information Technology is the most frequently occurring category of **Sector** with a count of **178** items with **average salary** values (**27** % of the total).

Over all **locations** and **sectors**, the average of **average salary** is **123.6**.

Biotech & Pharmaceuticals has a **minimum salary** of **212** for **Location Pleasanton, CA**.

Location New York, NY has the highest **minimum salary** at **1021**, out of which **Sector Retail** contributed the most at **141.**

Information Technology is the most frequently occurring category of **Sector** with a count of **178** items with **Location** values (**27** % of the total).

Company - Private is the most frequently occurring category of **Type of ownership** with a count of **386** items with **Industry** values (**58.5** % of the total).