**BIG DATA VISUALIZATION ON JOB MARKET DATA**

Bhanu Siva Kumar Kommana, Prudhvinath Reddy Katha, Vaishnavi Valluri, Vikhita Ganta, Mounika Chennuboina

Kent State University, Department of Computer Science, Kent, OH 44240

[bkomman1@kent.edu](mailto:bkomman1@kent.edu), [pkatha@kent.edu](mailto:pkatha@kent.edu), [vvalluri@kent.edu](mailto:vvalluri@kent.edu), [vganta@kent.edu](mailto:vganta@kent.edu), [mchennub@kent.edu](mailto:mchennub@kent.edu)

1. **INTRODUCTION**

In today's world, knowing about the job market is important, especially for businesses. We can find new job opportunities by looking at trends and patterns in areas like data science, analysis, and machine learning. This helps us understand what's happening in the job market for people looking for jobs, companies hiring, and others involved. Using Big Data Visualization, we can uncover hidden trends, find new job opportunities, and see what skills and certifications are in demand. This helps job seekers find jobs and helps companies use their talent better.

This project has many real-life uses. It helps people understand what skills they need for data jobs, which improves their chances of getting hired. Companies can use the data to find great candidates, make hiring better, and negotiate better job deals. And others involved can use the information to start programs that solve talent shortages, encourage new ideas, and make the data industry stronger economically. The project's earlier results have already made staff planning and development easier by showing how the job market is changing in the data industry.

1. **PROJECT DESCRIPTION**

The project aims to study and present data-related job markets using Kaggle data. This dataset contains job title, location, income, qualifications, experience levels, and type of work. Our goal is to thoroughly analyze these figures to understand the dynamics of the data industry job market.

***Scope of the Project:*** We aim to provide analytical insights on various aspects of the data-related job market, including position distribution, experience levels, hiring companies, and regional salary averages. Additionally, we'll investigate the relationship between education and compensation, as well as regional employment prospects. Once the project concludes, we'll produce a detailed study and visualization of the data industry, providing essential information for employers, job seekers, and other stakeholders.

***Challenges and Technical Contributions:*** Handling and analyzing large datasets is the primary challenge. This involves substantial data preparation, including data cleansing, resolving missing values, and formatting data for analysis. Another challenge is creating engaging graphics for analyzing complex data. To address these challenges, we've developed novel visualization techniques specifically for employment market analysis. Our technical contribution ensures clear and understandable presentation of data, facilitating the extraction of useful insights.

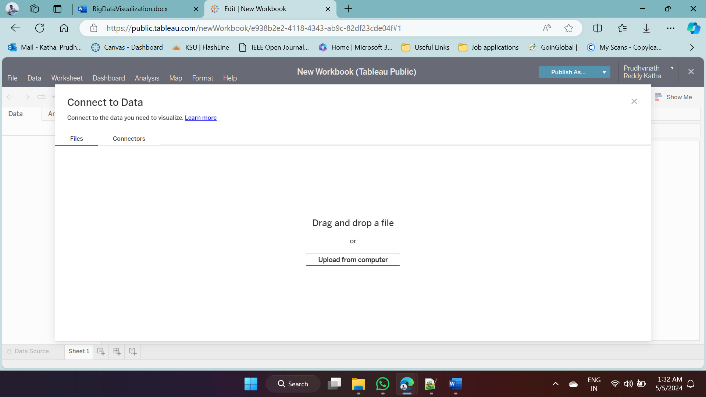
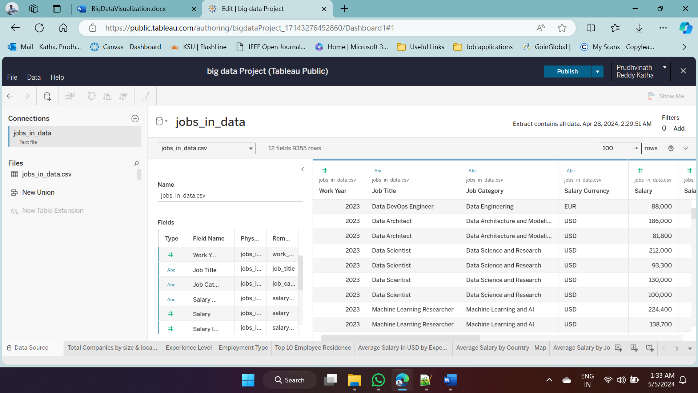
***Workload Distribution:*** The project is divided into three main stages with equal effort distribution among team members. The first stage involves designing visualizations and preparing the data. This includes cleaning the dataset, handling missing data, and creating visually appealing visualizations. The team member responsible for this stage explores the data, contributes to GUI design, and identifies trends and patterns to create a user-friendly interface. The second stage focuses on dashboard development and documentation. The team collaborates to design and build an interactive dashboard that effectively displays visualizations and conclusions. They also ensure comprehensive and well-organized project documentation. Each team member contributes equally to the data processing and in developing interaction with the visuals, ensuring successful completion of the project.

1. **BACKGROUND**

***Literature Review:*** Despite the absence of directly relevant articles or surveys, our team drew inspiration from various sources in data science, job market analysis, and data visualization during our literature review. We gained insights into effective visualization techniques from books like Kieran Healy's "Data Visualization: A Practical Introduction" and similar resources. Additionally, we explored research papers and publications on employment market trends and analysis to understand industry methodologies and best practices. This review significantly influenced the concept and execution of our project.

***Software Tools:*** We utilize Tableau for data visualization. Tableau enables users to create interactive and informative representations from various data sources, simplifying the visual expression of complex ideas. The interface is very easy, and it has features like drag-and-drop the attributes, make the data more interesting by visualizations for type of experiences levels. Tableau gives different visualization options, like bar charts, pie charts, line graphs, maps, and dashboards, helping in detailed data analysis. Also, Tableau's data joining, cleaning and visualization features allow us to merge information from multiple sources for in-depth analysis.

Python is mostly used for data preprocessing, analysis, and make ML models. Frameworks like Pandas and NumPy are used for managing and analyzing data. These libraries offer robust data structures and algorithms for cleaning, manipulating, and analyzing large datasets. Pandas provides mathematical tools for working with arrays, while NumPy offers a variety of numerical functions. For data visualization, we use Matplotlib, a Python package that allows us to create various plot types like line graphs, scatter plots, and maps, making data viewing easier.

*Tableau data collection*

***Optional Hardware:*** This project can be completed on a standard laptop or desktop computer with sufficient power and memory. A multicore processor and 8GB of RAM are necessary for handling large datasets efficiently. A high-quality graphics card, such as an Intel iRISx graphics card, can help improve the performance of visualization apps like Tableau. System requirements for optimal performance include a laptop with 16GB of RAM and an Intel iRISx graphics card.

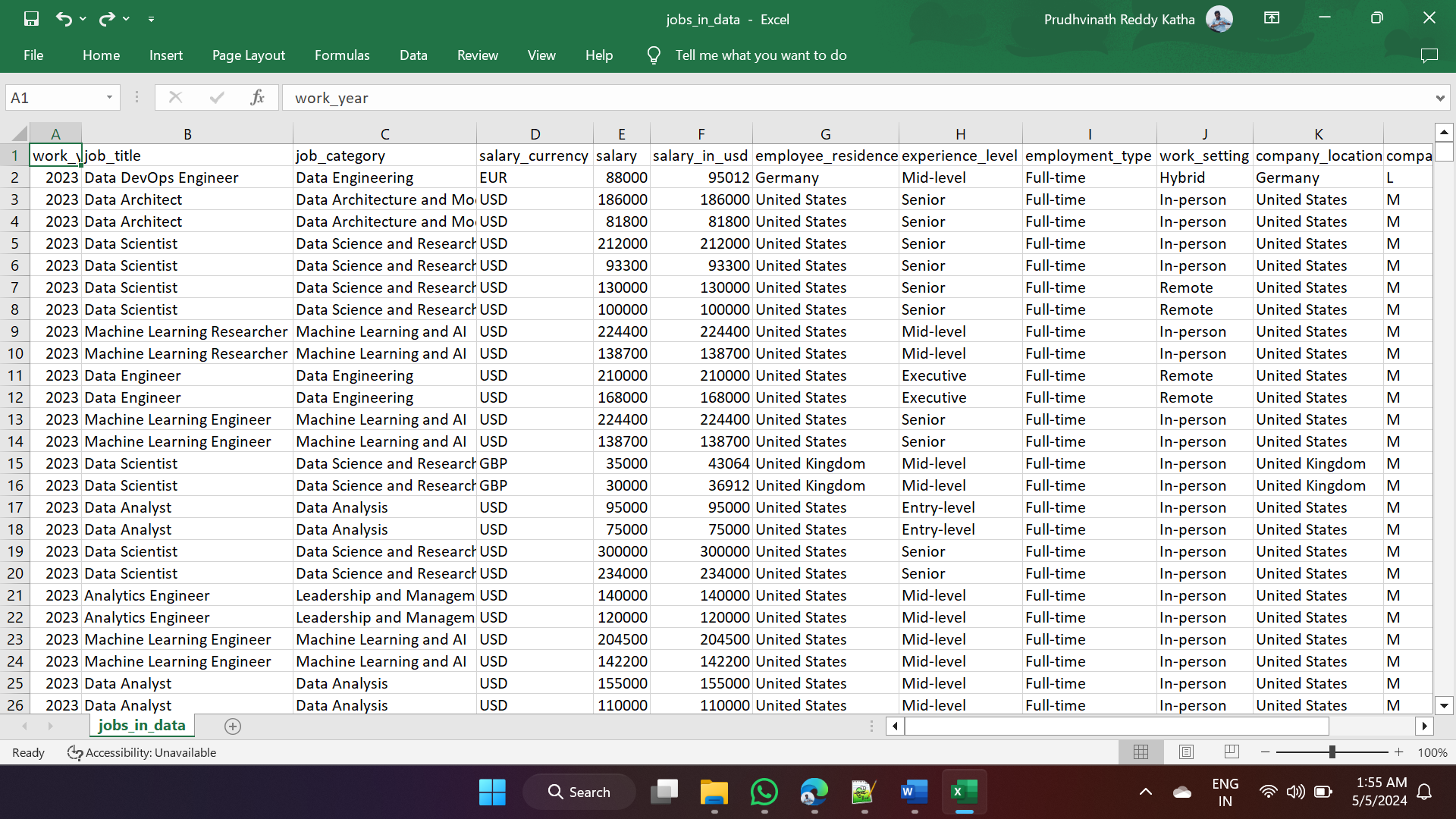
***Programming Skills:*** Python expertise is crucial, especially for data processing tasks like cleaning, modifying, and analyzing data using tools such as Pandas and NumPy. Proficiency in Matplotlib is also valuable for creating impactful visualizations and plots. Choosing appropriate chart types and color schemes is essential for effective data visualization. Additionally, knowledge of Tableau is necessary for developing interactive dashboards and visualizations with filters and calculation fields.

1. **PROBLEM DEFINITION**

This project's main objective is to do a collective and informative analysis and visualization of job market data trends. We want to analyze and visualize job market data from Kaggle. Our goal is to understand job trends like job titles, locations, salaries, qualifications, experience levels, and types of jobs. By doing this, we hope to learn how the job market operates and find important patterns and trends. This information can help both job seekers and employers make better decisions.

***Challenges:*** The processing and preprocessing of substantial amounts of data is one of the main challenges we face. Inconsistencies, missing values, or data in different formats are present in this dataset and require considerable cleaning and formatting before analysis is done. Another challenge is to create efficient visualizations that fully represent the data and give meaningful information. It is essential to ensure that the visualizations are intuitive, understandable, and visually good to make understanding and interpretation easier.

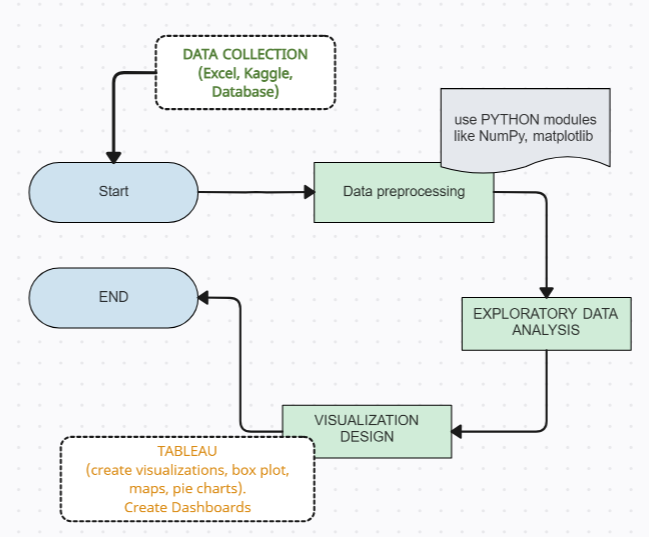
***Proposed general solution:*** We have a plan to tackle these issues using a mix of Python and Tableau. Python will help us clean up and organize the data, using tools like Pandas and NumPy. Then, we will dig into the data with exploratory analysis to spot any trends. Once the data's ready, we will switch to Tableau to make cool interactive graphs and charts. Tableau's user-friendly interface lets us create all sorts of visuals like bar graphs, pie charts, maps, and dashboards. Our goal is to analyze the data thoroughly so that everyone involved can make smart decisions."



*Raw dataset of job market*

1. **PROPOSED TECHNIQUES**

***Data Analysis Flow Diagram:*** Our framework involves several steps:



*Data Analysis Framework Flowchart*

***Methods Overview:***

***Data Preparation:*** We start by organizing our data. This includes cleaning it up, fixing mistakes or missing information, and ensuring correct formatting. Think of it like tidying up a messy room before use.

***Exploratory Data Analysis (EDA):*** Once the data is clean, we examine it closely. This step helps us understand what the data is telling us and identify any interesting patterns or trends. It's like exploring a map to understand the terrain before planning a journey.

***Visualization:*** Now that we understand our data better, we aim to make it easier to understand. We use visual tools like graphs and charts to present the information clearly and meaningfully. For instance, we might create a bar graph to show the distribution of different job titles or a map to display job opportunities' locations.

***Dashboard Creation:*** Finally, we consolidate everything into a dashboard. This acts as a control panel allowing users to interact with the data dynamically. They can explore various visualizations, filter data based on their interests, and gain deeper insights.

***Data Encoding/Indexing:*** To efficiently retrieve and analyze data, we decode and index the dataset. Classification variables like job titles, locations, and types of work are stored as numerical values. Indexing techniques may be used to improve data retrieval performance.

***Data Processing Algorithms and Query Optimization:*** We use query processing algorithms to extract relevant information. Techniques such as filtering, aggregation, and sorting are employed for efficient query processing. Query optimizations can further improve performance by reducing computational costs for complex queries.

Overall, these steps help us transform raw data into useful information for job seekers and employers to make informed decisions.

1. **VISUAL APPLICATIONS**

**Design Modules:** Here there are few steps involved in designing this project they are:

Data exploration: To provide an overview of the data, we will display the distribution of job titles, salaries, experience levels and locations.

Salary Analysis: Using bar charts and line graphs to represent salary trends, the visualisations focus on average salaries across different levels of experience and types of employment.

Geographical Analysis: Using maps, plotting company locations and examining the concentration of employment opportunities in different regions, we analyse the geographical distribution of employment opportunities.

***10 Most in Demand Job Titles:*** A bar chart displays the top ten most in demand job titles of data industry, providing insight into what skills and jobs are needed.

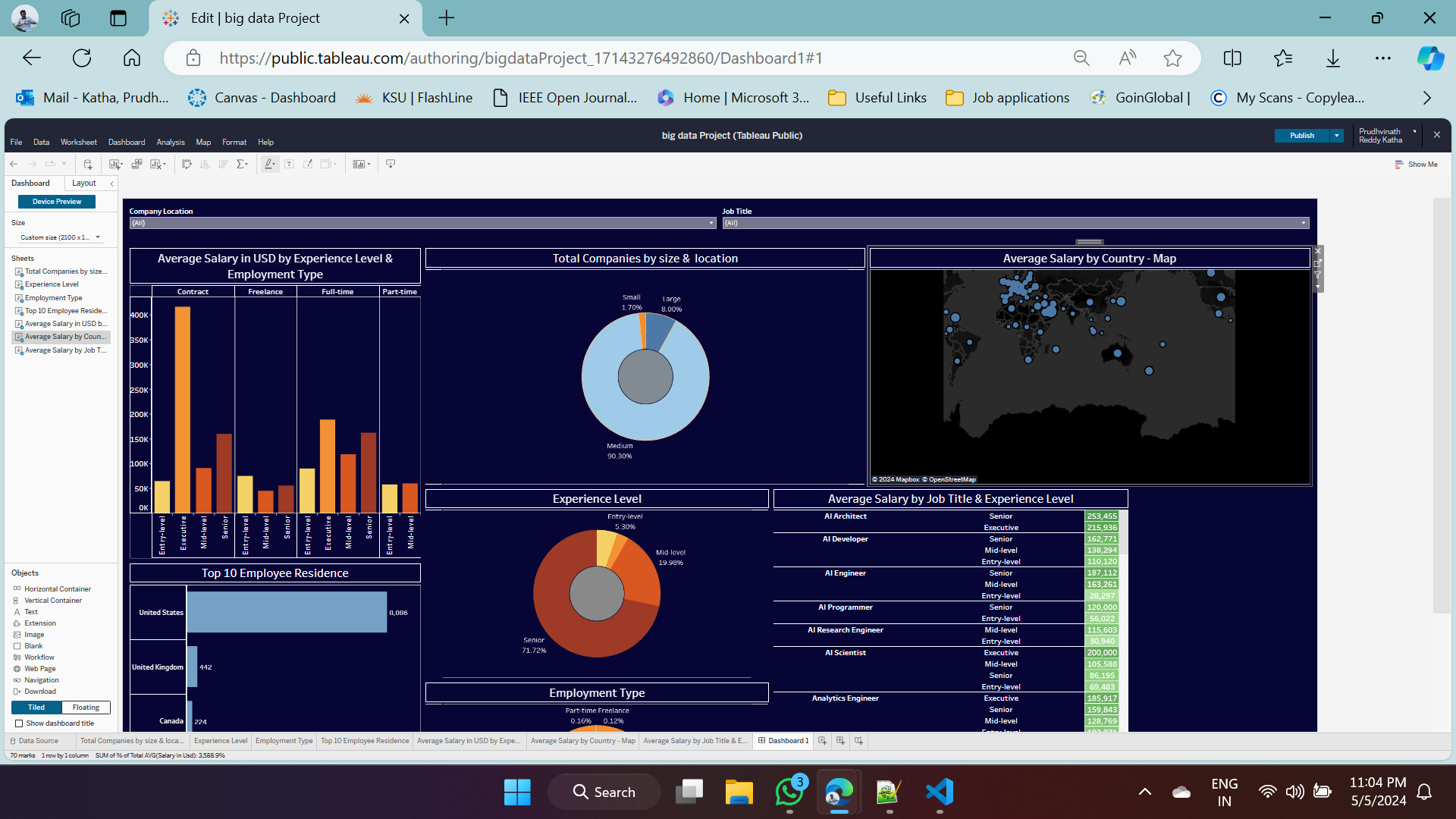
***Employee Residence Analysis:*** To show the number of employees in different areas, a visualization of the location of the employees may be used, possibly using heatmaps or choropleth maps.

1. **EXPERIMENTAL EVALUATION**

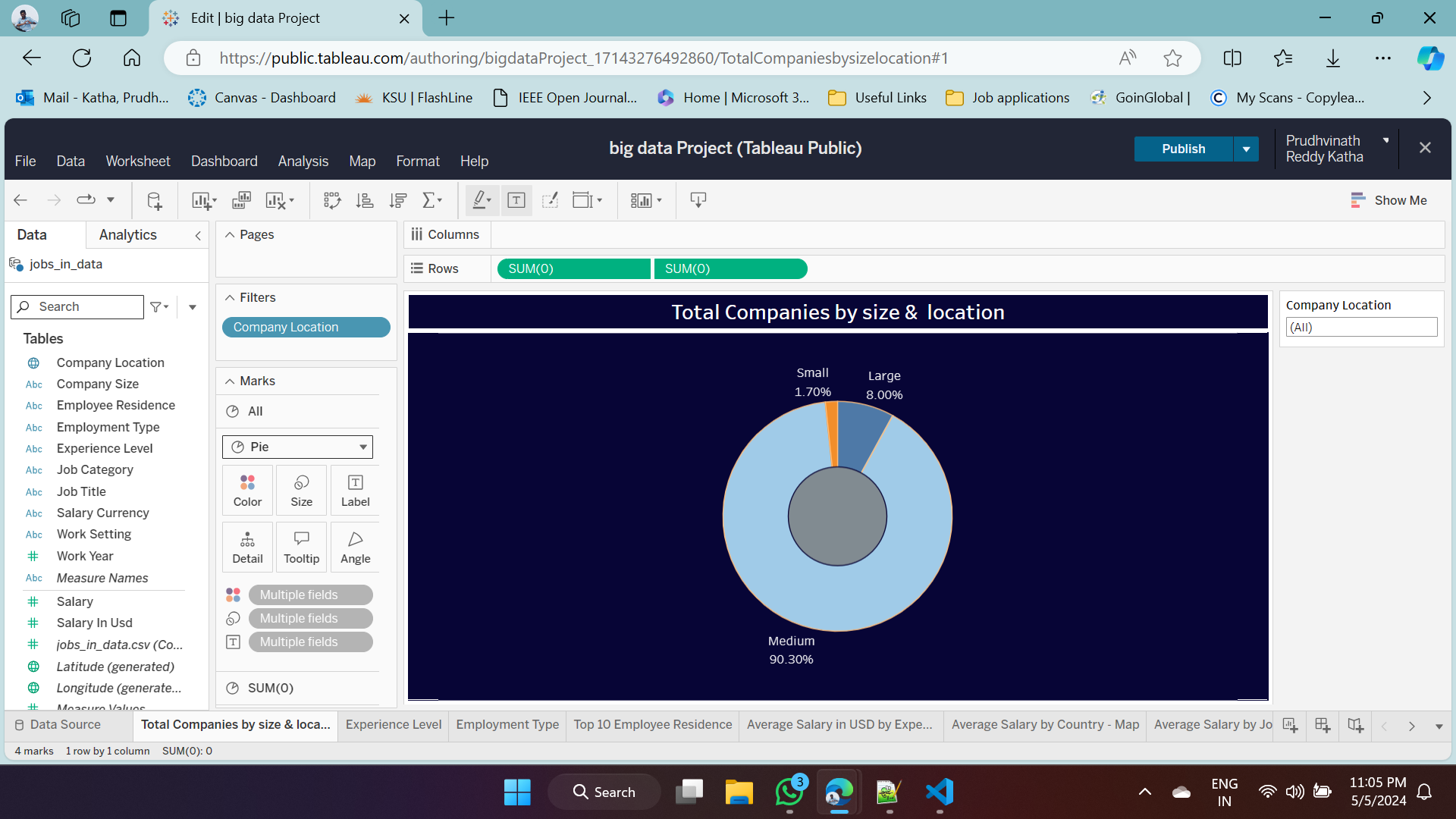
We're using the Kaggle dataset "Jobs in Data" for our analysis. In our exploratory data analysis, we compare our visualizations to industry standards. We've tweaked parameters like chart types, colors, and filters to make the visualizations more efficient.

***Evaluation Measures:*** We're evaluating our visualizations based on clarity, information, and interactivity. We invited 20 people to use Tableau visualizations during the design phase. About 80% of them were satisfied, finding the visualizations useful and easy to understand. Most respondents appreciated the clear and interactive nature of the visualizations. They liked being able to select and analyze different aspects of the data, which helped them use the information more effectively. Overall, the evaluation shows that Tableau visualizations were well received and effectively conveyed the information we wanted. These positive responses validate our approach and provide insights for further refinement of the project.

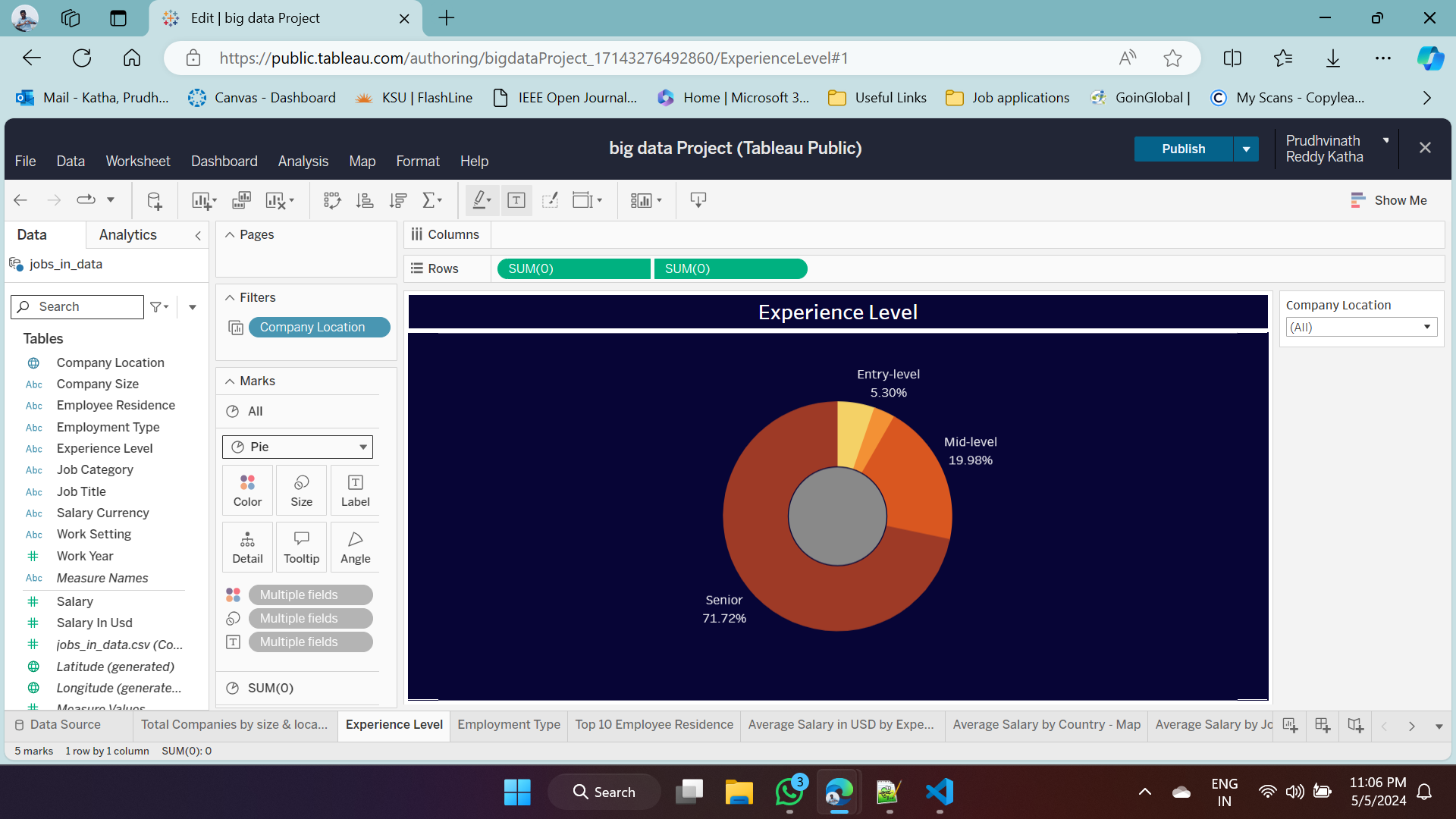
**Output screenshots:**



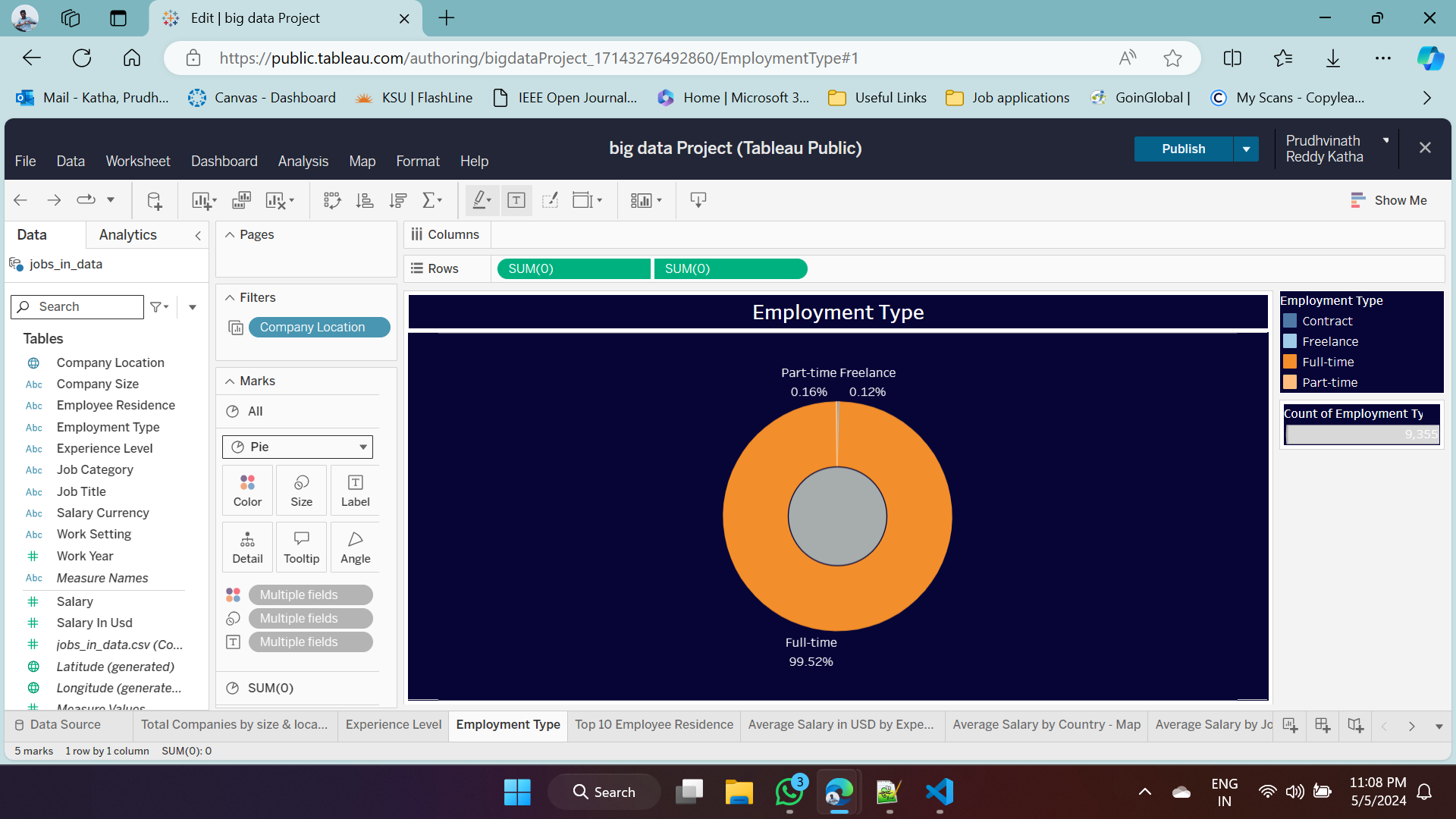
*An Interactive dashboard*



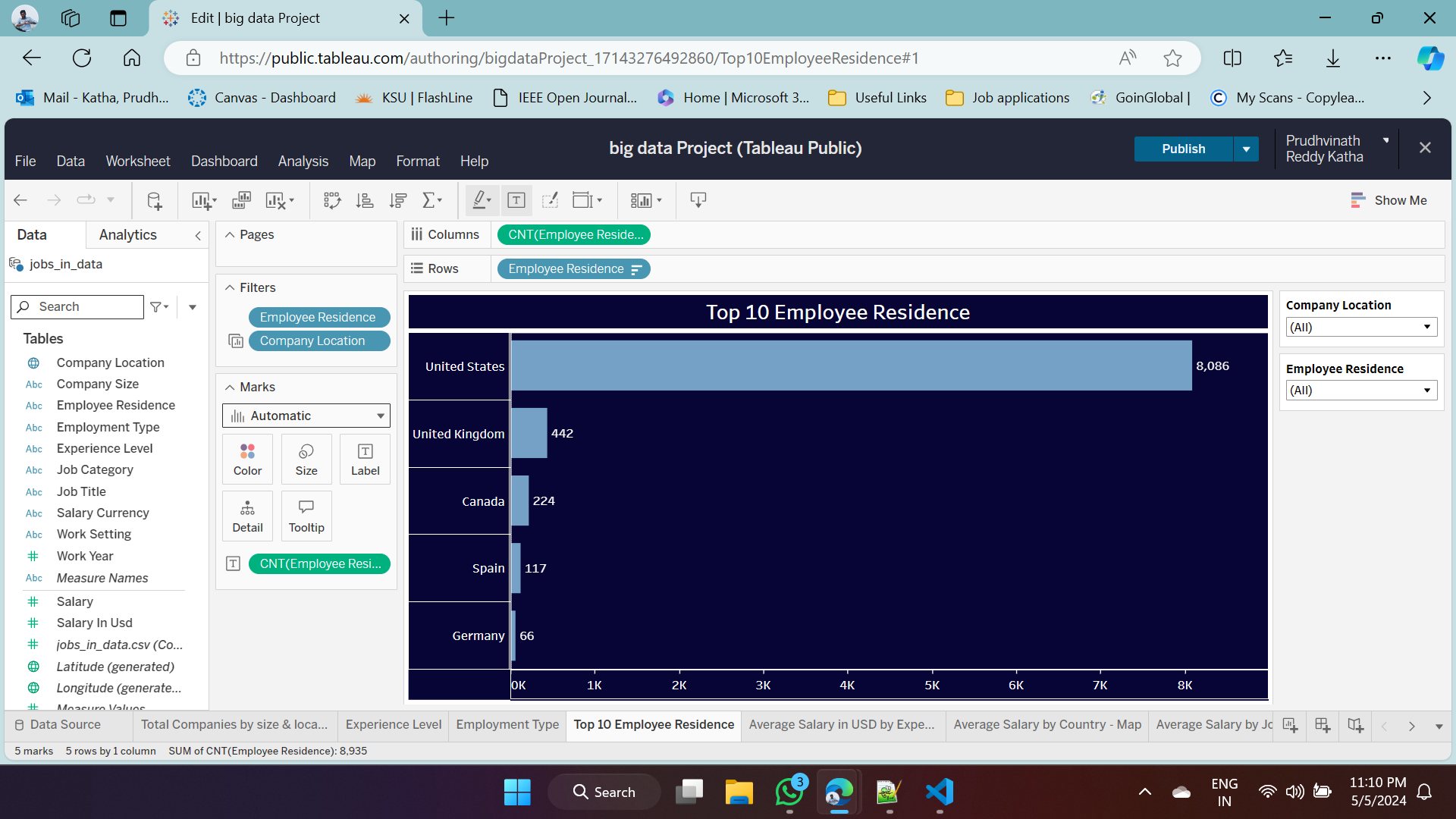
*Total Companies by size & location*



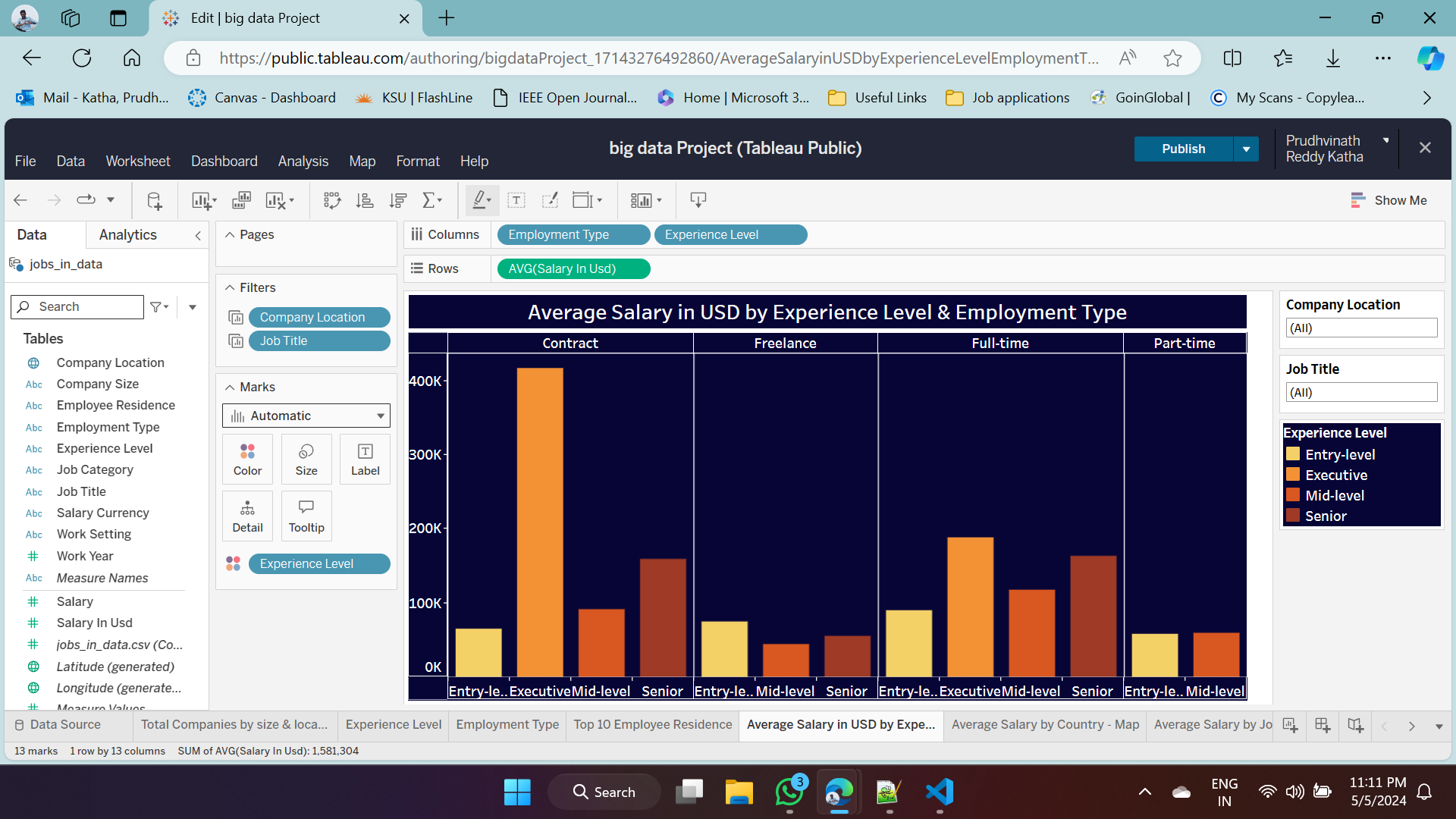
*Different Experience levels*



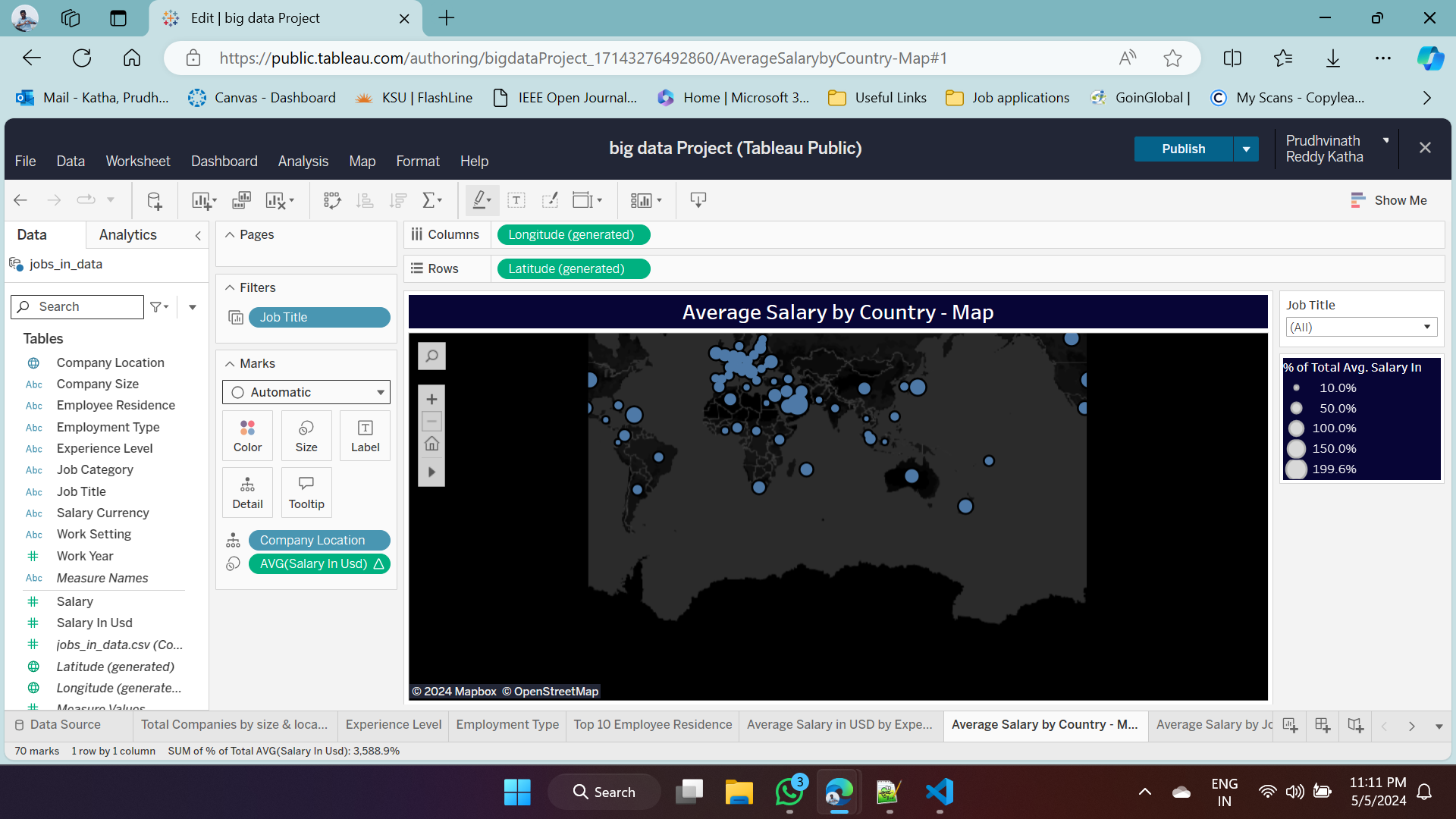
Employment types in each company



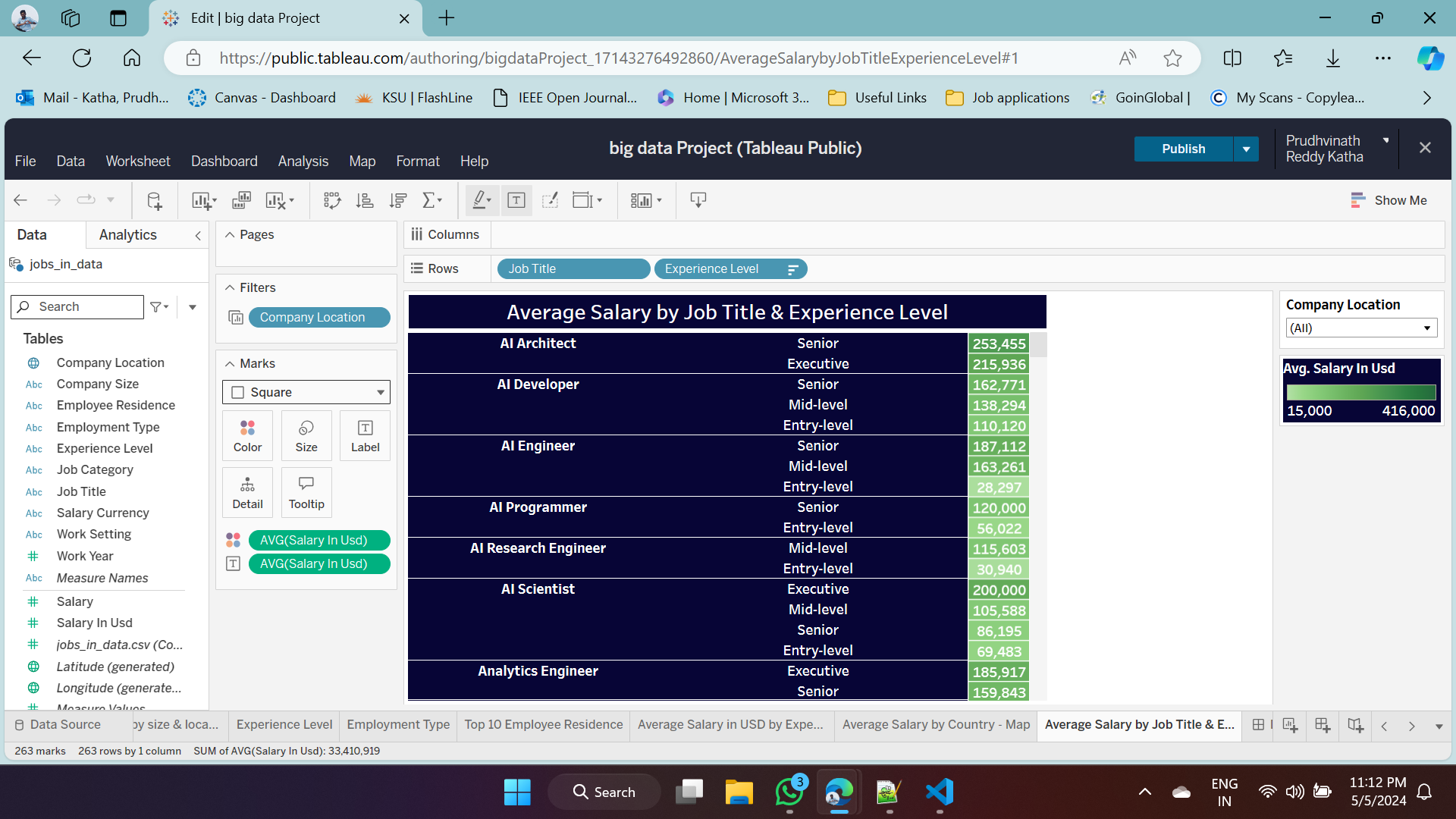
*Top 10 employees Residents in bar chat*



*Average Salary in USD by Experience & Employment type*



*Average salary by country - Map*



*Average Salary by job title & experience level in highlight table*

1. **FUTURE WORK**

***Enhancing visual techniques:*** We are planning to introduce more advanced visual techniques in subsequent versions of this project so that we can get a better picture of the data. Also using 3D views or animations to represent a continuous trend could be part of it.

***Predictive modeling:*** It might be helpful to expand the project and add prediction models. We can predict trends in the job market and anticipate future demand for specific skills or job roles by applying machine learning algorithms.

***Integration of visualizations with real-time data sources***: the project's relevance and usefulness would be enhanced by integrating visualizers with realtime data sources. To provide timely information on trends in the job market, this could entail connecting with APIs and streaming data sources.

1. **Conclusion**

Our project mainly focussed on visualizing a large and complex data in a better way to understand its attributes, information and insights using a visualization tool called tableau. We used a data set on job market around the world from kaggle and used data processing techniques and visualizing tools like tableau to present more valuable insights to users. Using tableau to generate dynamic and interactive visulaizations to clearly understand the dataset. We used bar plots, line graphs and geo maps to generate static experience to understand the attributes and data clearly and know the present insights on them.

1. **REFERENCES**

[1] Kieran Healy. (2018). Data Visualization: A Practical Introduction. Princeton University Press.

[2] McKinney, W., & others. (2010). Data structures for statistical computing in python. Proceedings of the 9th Python in Science Conference, 445, 51-56.

[3] Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. Computing in Science & Engineering, 9(3), 90-95.

[4] Tableau Software. (2024). Tableau. Retrieved from <https://www.tableau.com>.

[5] Kaggle. (2024). Jobs in Data Dataset. Retrieved from https://www.kaggle.com/datasets/hummaamqaasim/jobs-in-data.