INT 233- PROJECT REPORT

(Project Semester August-December 2024)

Data Science Salaries Dashboard

Submitted by

Donadri Naga Venkata Manibabu

Registration No: 12212606

Programme and Section: K22AU

Course Code INT233

Under the Guidance of

ASHU: 23631

Discipline of CSE/IT

Lovely School of B-Tech CSE

Lovely Professional University, Phagwara

CERTIFICATE

This is to certify that Donadri Naga Venkata Manibabu bearing Registration no. 12212606 has completed INT233 project titled, “DATA SCIENCE SALARIES DASHBOARD” under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

ASHU: 23631
Associate Professor
School of B-Tech CSE
Lovely Professional University
Phagwara, Punjab.

Date: 15-11-2024

DECLARATION

I, Aasritha Lakshmi Vemu, student of B-TECH CSE under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 15-11-2024 Signature

Registration No: 12212606 Manibabu dnv

Introduction:

This project delves into the salaries of data science employees, using Tableau to uncover actionable insights through interactive dashboards. The dataset encompasses detailed company size, categorized by country, states, job title, employee types, experience level, salary currency and employee residence. This structured data provides a foundation for understanding salaries trends and the performance of various countries across different locations.

The primary objective of this analysis is to provide a granular view of average salaries pattern to help analyser make informed decisions. The analysis investigates several critical aspects, such as:

- Identifying the total companies by size and location (e.g., the medium companies are almost 92%).
- Experience level (by entry level, middle, expert and senior).
- Evaluating employment type by part-time, freelance, contract and fulltime.
- Top 10 employee residence by job title and employee residence.
- Average salary in USD by experience level and employment type (location also)

This analysis enables the salaries and employee status to assess its performance across different geographic and job titles dimensions. For example:

- States like California, Texas, and Colorado are identified as major IT cities.
- Data science emerges as the leading sales category, with a significant share of total contribution.

By visualizing the data through Tableau dashboards, the project provides intuitive and interactive insights, facilitating better decision-making. These dashboards allow stakeholders to:

1. Quickly identify salaries hotspots and underperforming areas.
2. Assess job popularity and profitability to refine area and resolution.

.

Scope of the Analysis:

The analysis of the Data Science Job Salaries Dashboard provides a comprehensive overview of various factors that influence job salaries in the data science field. Here are the key areas covered:

1. **Salary Analysis by Experience Level and Employment Type**:
    - Evaluates the average salaries across different experience levels (entry-level, middle-level, senior, and expert).
    - Considers various employment types, including contract, freelance, full-time, and part-time positions.
2. **Company Size and Location**:
    - Analyzes the distribution of companies by size (small, medium, and large) and their respective locations.
    - Identifies patterns and trends related to company size and geographical impact on job availability and salaries.
3. **Experience Level Distribution**:
    - Provides insights into the proportion of data science professionals at different experience levels.
    - Helps understand the career progression and the demand for different experience levels within the industry.
4. **Employment Type Distribution**:
    - Breaks down the prevalence of different employment types in the data science field.
    - Highlights the dominance of full-time positions compared to part-time, freelance, and contract roles.
5. **Top Employee Residences**:
    - Identifies the most common locations where data science professionals reside.
    - Highlights the concentration of data science talent in specific countries.
6. **Salary by Job Title and Experience Level**:
    - Compares average salaries for various job titles within the data science field.
    - Evaluates how experience levels impact salaries for specific roles.
7. **Geographical Salary Analysis**:
    - Displays average salaries across different countries using a world map.
    - Provides a visual representation of salary variations by location.

Existing System:

The existing system for the Data Science Job Salaries Dashboard involves collecting and integrating data from various sources, stored in centralized databases like SQL or NoSQL systems. The data undergoes cleaning and processing using ETL (Extract, Transform, Load) processes and normalization. Analytical tools like Tableau and statistical software such as Python or R are used for analysis. Interactive dashboards in Tableau display metrics and insights, accessible to users for exploration. Regular system maintenance ensures data is up-to-date, and user feedback is incorporated to improve the system.

Drawbacks/Limitations of the Existing System

The traditional system had several limitations:

 Data **Accuracy and Consistency**: Variability in data sources can lead to inconsistencies and inaccuracies, impacting the quality of insights.

 Data **Integration Challenges**: Combining data from multiple sources might be complex and time-consuming, potentially leading to errors.

 Scalability **Issues**: As data volume increases, the system might struggle with performance and scalability, affecting efficiency.

 Maintenance **Overheads**: Regular updates and maintenance are required to keep the system and data current, which can be resource-intensive.

 User **Accessibility**: Limited access to dashboards and reports can restrict who can benefit from the insights, potentially leaving out key stakeholders.

 Technical **Expertise Requirement**: Users might need significant technical expertise to interpret complex data visualizations and analyses.

 Data **Security Concerns**: Ensuring data security and privacy can be challenging, especially with sensitive information.

 Limited **Flexibility**: The existing system might not be easily adaptable to new data sources or changing requirements without significant modifications.

Source of Dataset

The dataset used for this project was sourced from the kaggle database. It contains transactional data including:

- Sales Data: Records of each transaction, including the store location, product sold, quantity, and sales amount.
- Customer Data: Information about customer demographics (age, gender, etc.) where available.
- Promotions: Data related to promotional activities and their impact on sales.
- Inventory: Information about inventory levels to analyze sales in relation to stock availability.

The data was extracted and compiled from multiple sources within the company's IT infrastructure, ensuring accuracy and consistency.

ETL Process

The ETL (Extract, Transform, Load) process was implemented to clean and prepare the data for analysis and visualization in Tableau.

Extract

- Data Sources: Data was extracted from the kaggle system, stored in relational databases (e.g., SQL).
- Extraction Methods: SQL queries were used to pull transaction data, customer demographics, and inventory details.

Transform

- Data Cleaning: The raw data was cleaned to remove any inconsistencies, such as missing values, duplicated records, or incorrect data formats.
    - Missing values in product categories were handled by assigning default values.
    - Date fields were standardized to a consistent format (YYYY-MM-DD) for easier analysis.
- Data Aggregation: Data was aggregated by store location, product category, and time periods (daily, weekly, monthly) to facilitate high-level trend analysis.
- Data Enrichment: The customer data was enriched by grouping demographic information (e.g., age groups, region) to analyze customer behavior patterns.

Load

- Data Loading into Tableau: The transformed data was loaded into Tableau using automated scripts to update the dashboard regularly. This ensures that the sales data is refreshed in real-time and reflects the most current information for decision-makers.
- Data Storage: Data is stored in Tableau’s data engine, allowing fast querying and real-time visualizations.

Analysis on Dataset

The dataset includes various details such as work year, experience level, employment type, job title, salary, salary currency, salary in USD, remote ratio, company location, and company size.

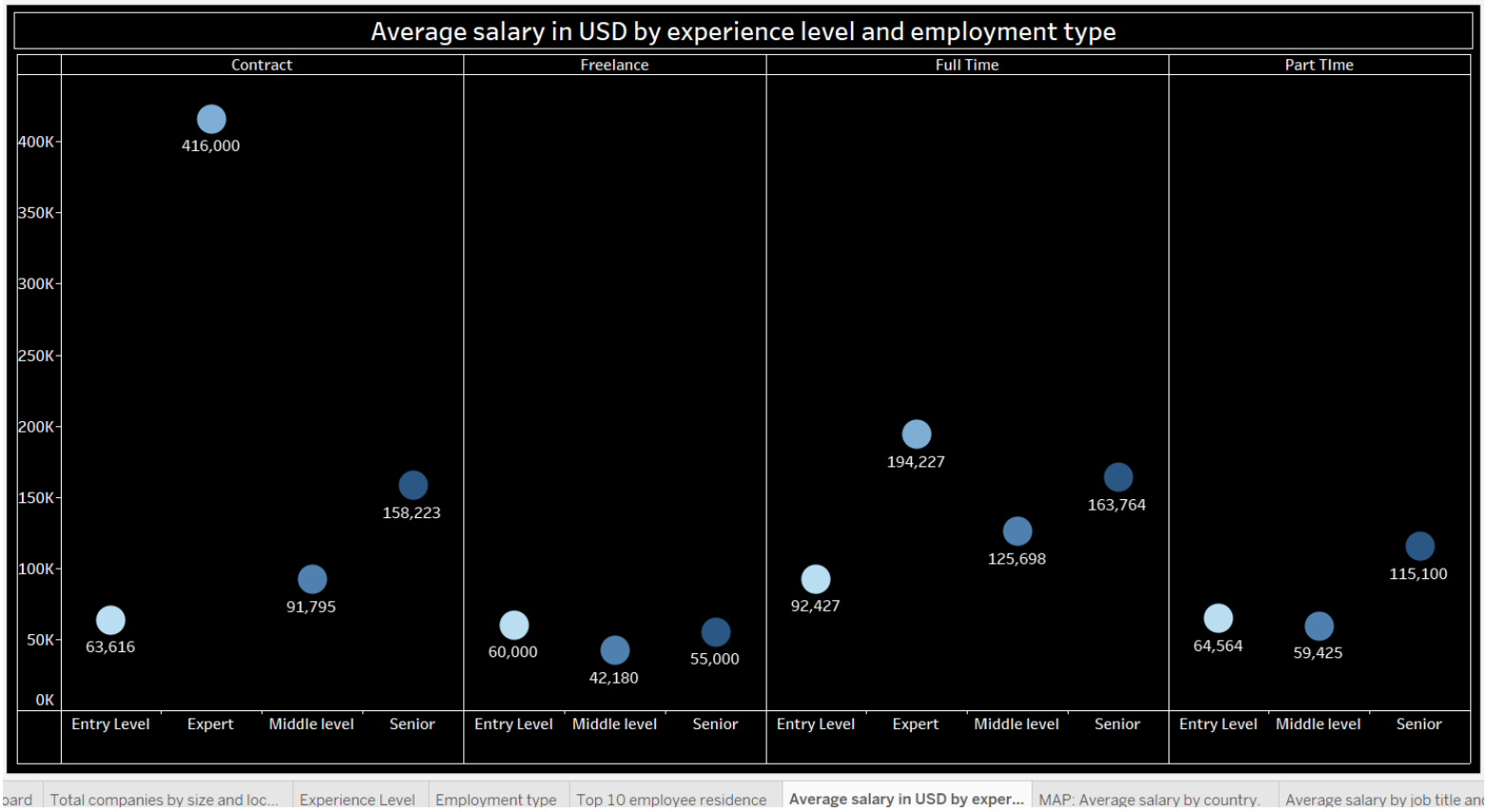
Key Insights

1. **Average Salary by Experience Level**:
    - **Entry Level**: Varied salaries based on job titles and locations.
    - **Mid-Level**: Salaries generally increase with experience.
    - **Senior Level**: Highest salaries are typically at senior levels.
    - **Expert Level**: Significant jump in salary for expert roles.
2. **Employment Type**:
    - **Full-Time Roles**: Most common employment type with a wide range of salaries.
    - **Contract and Freelance**: Less common but present, with varied salaries.
3. **Remote Work**:
    - **Remote Ratios**: Different roles offer varying levels of remote work, with some positions being fully remote.
4. **Company Size and Location**:
    - **Company Size**: Salaries vary significantly based on the size of the company (small, medium, large).
    - **Geographical Impact**: Location affects salary, with some regions offering higher or lower salaries depending on cost of living and demand for roles.
5. **Job Titles**:
    - **Diverse Roles**: Data Science, Machine Learning, AI, and related fields.
    - **Salaries by Title**: Specialized roles like AI Architect and Lead Machine Learning Engineer tend to have higher salaries.

Visualization

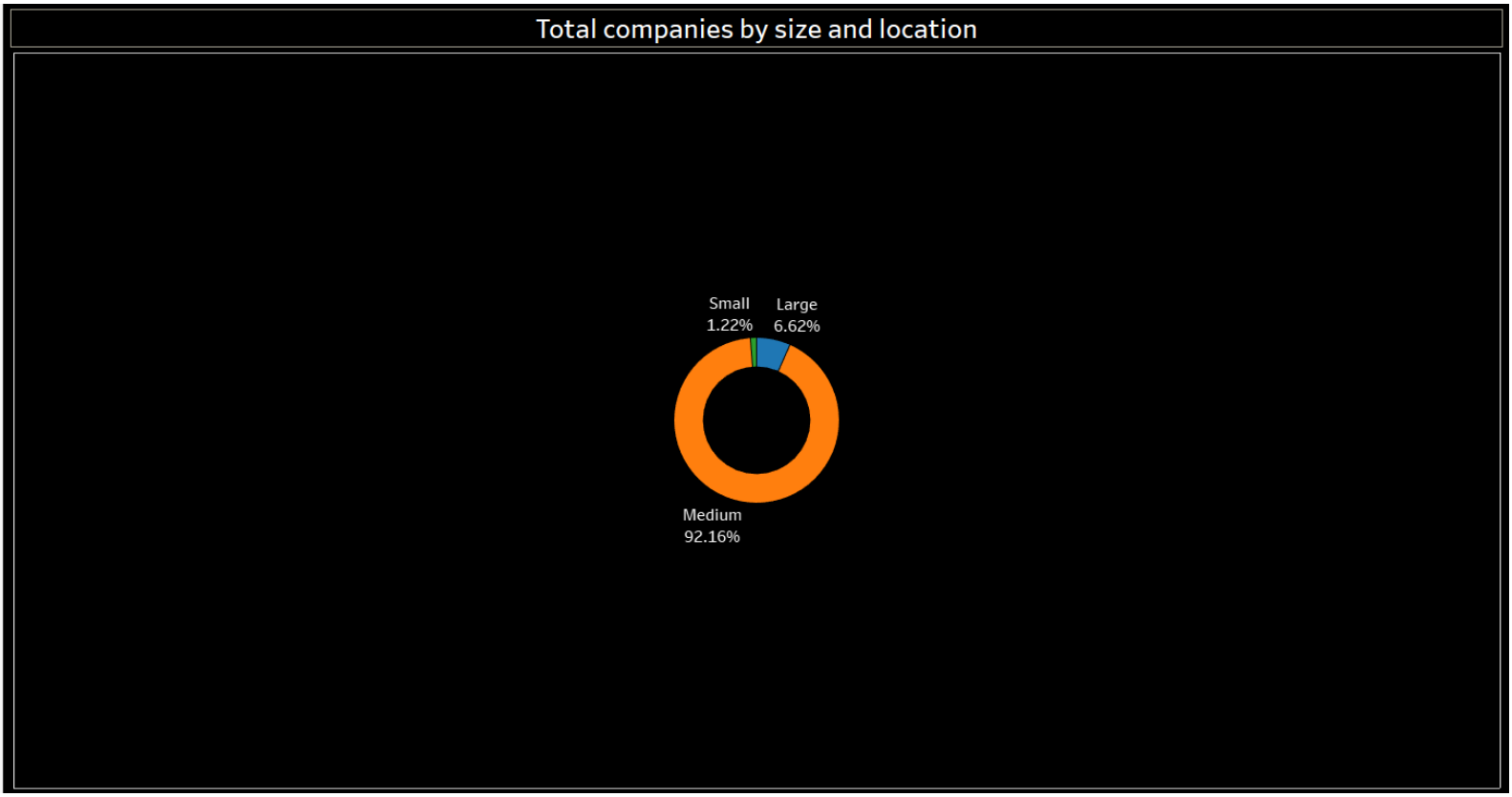
1. Average Salary in USD by Experience Level and Employment Type

- This bar chart breaks down average salaries for different experience levels (Entry, Middle, Senior, Expert) across various employment types (Contract, Freelance, Full Time, Part Time). It highlights how salary scales with experience and varies significantly based on employment type. For instance, Full Time roles often show steady salary increases with experience, while Contract roles might show higher variability.



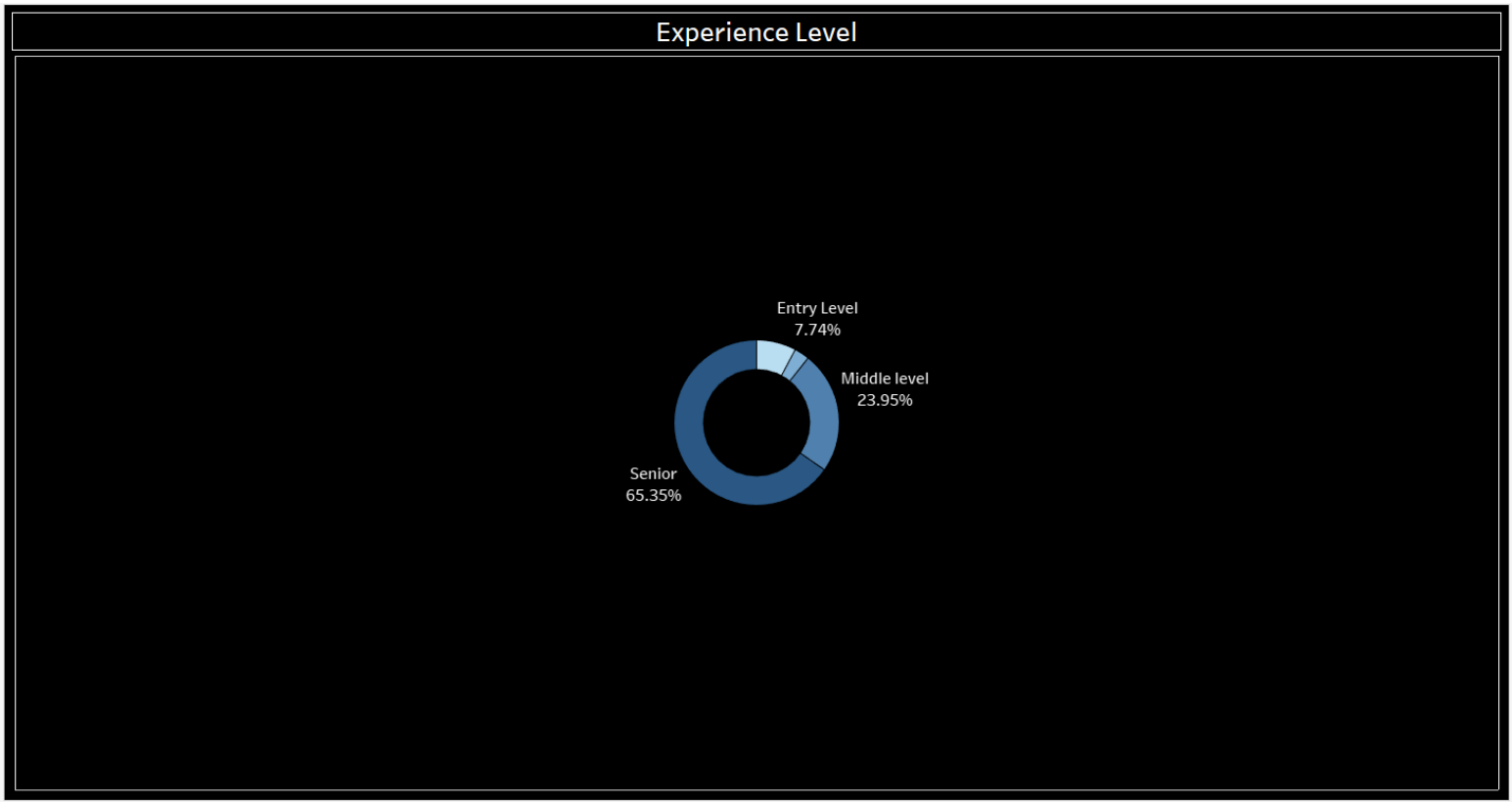
2. Total Companies by Size and Location

- This pie chart illustrates the distribution of companies by size (Small, Medium, Large). A map overlay can show where these companies are located globally, indicating how company size correlates with geographic regions. This helps identify regions with a higher concentration of medium or large companies, which might offer more job opportunities or higher salaries.



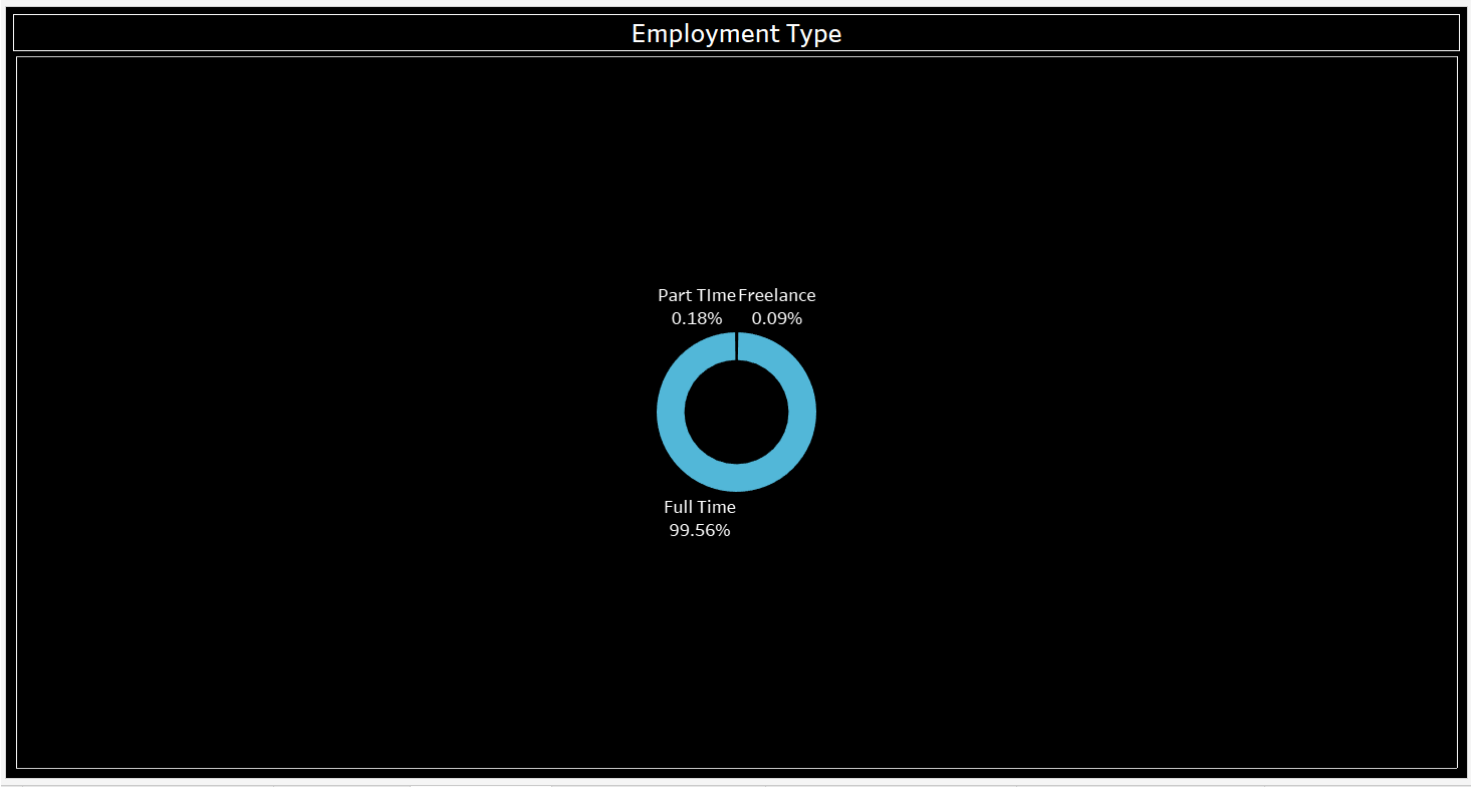
3. Experience Level Distribution

- This pie chart displays the proportion of employees at various experience levels (Entry Level, Middle Level, Senior). It provides a visual representation of the workforce composition, helping to understand the spread of experience within the field. A higher proportion of senior-level professionals could indicate a mature industry with a strong demand for experienced talent.



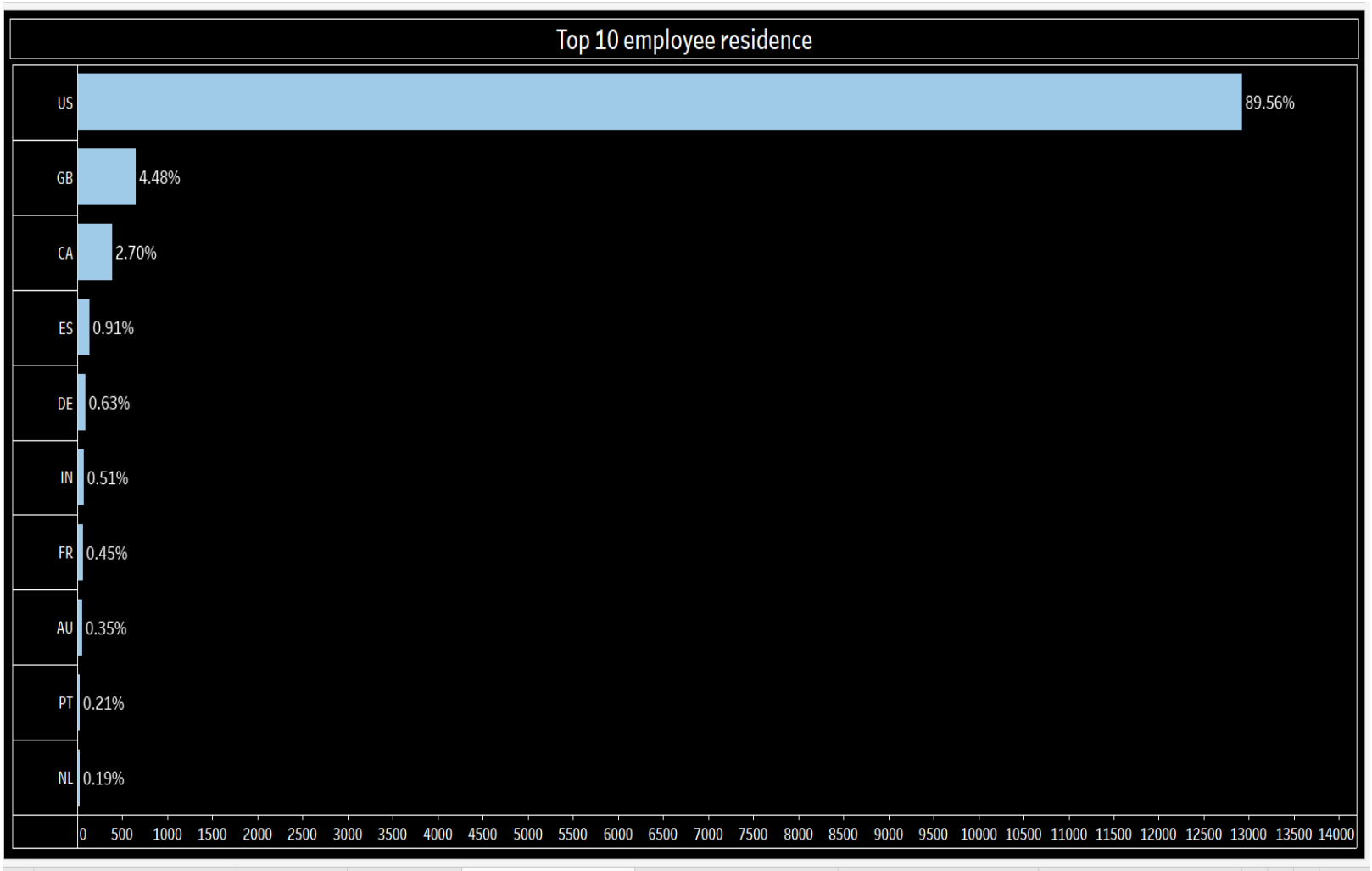
4. Employment Type Distribution

- This bar chart shows the distribution of employees across different employment types (Full Time, Part Time, Freelance). It highlights the dominance of Full-Time roles in the industry and gives an idea of how prevalent other types of employment are. This can be useful for understanding the stability and flexibility of the job market in data science.



5. Top 10 Employee Residences

- This bar chart or map displays the top 10 countries where data science professionals reside, based on the number of employees. It helps identify geographic hotspots for data science talent, showing where the industry is most concentrated. This can be useful for companies looking to hire or professionals considering relocation for better job prospects.



6. Average Salary by Job Title and Experience Level

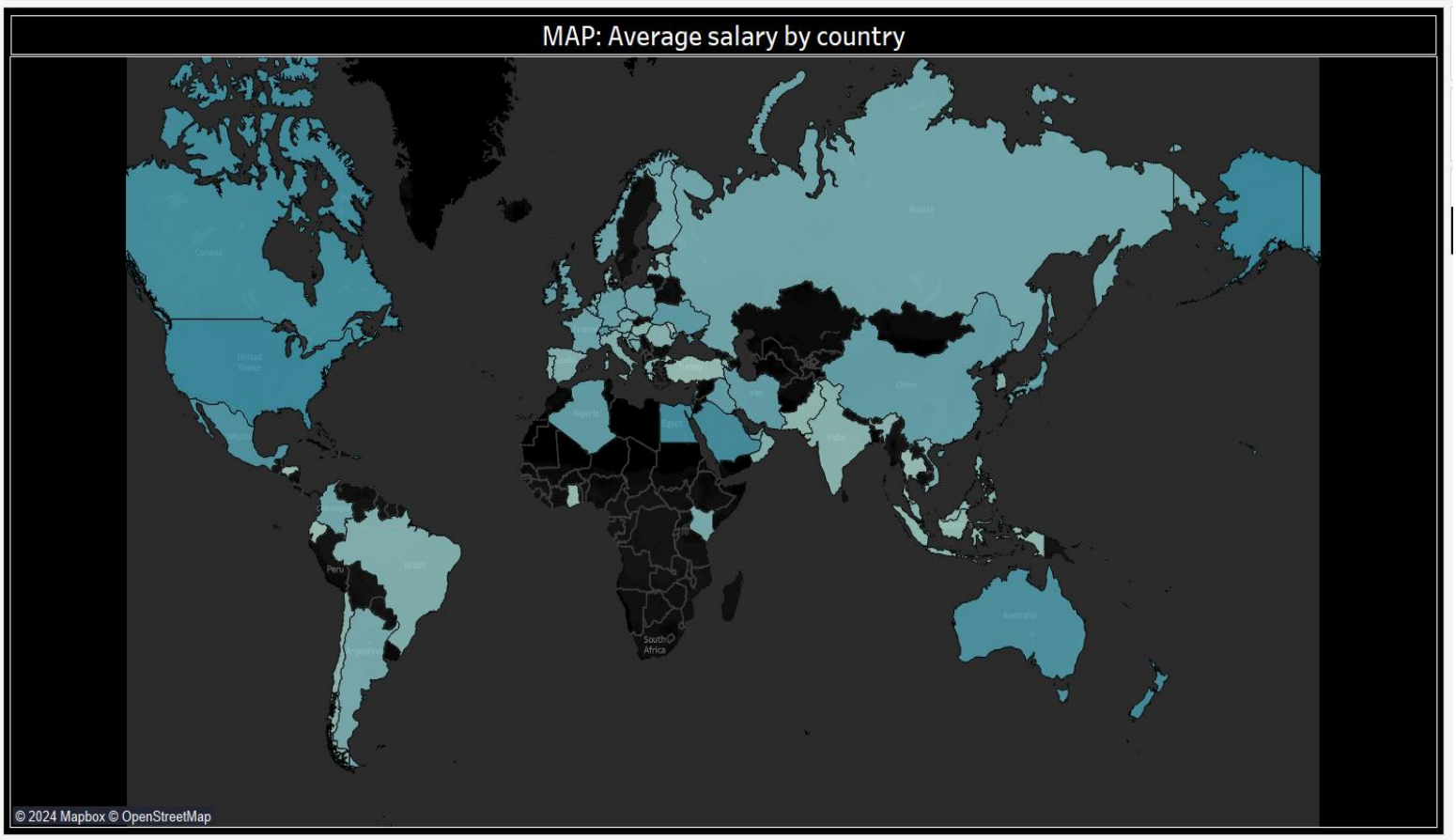
- This clustered bar chart compares average salaries for various job titles within the data science field, segmented by experience levels (Entry, Middle, Senior, Expert). It provides detailed insights into how different roles and levels of experience impact salary. For example, it can show how salaries for specialized roles like AI Architect or Lead Machine Learning Engineer compare to more general roles like Data Analyst.

Average salary by job title and experience level

Admin & Data Analyst	Middle level	42,222
	Entry Level	50,000
	Senior	60,000
AI Architect	Expert	215,936
	Senior	235,694
	Middle level	800,000
AI Developer	Entry Level	110,120
	Middle level	112,054
	Senior	162,771
AI Engineer	Entry Level	33,679
	Middle level	151,910
	Senior	176,148
	Expert	178,800
AI Product Manager	Senior	120,000
	Middle level	152,650
AI Programmer	Middle level	30,000
	Entry Level	56,859
	Senior	120,000
AI Research Engineer	Entry Level	28,734
	Middle level	123,635
AI Research Scientist	Entry Level	88,888
	Senior	150,000
AI Scientist	Entry Level	60,912
	Middle level	105,588
	Senior	164,858
	Expert	200,000
AI Software Engineer	Senior	174,100
Analytics Engineer	Entry Level	100,935
	Middle level	146,141
	Senior	165,200

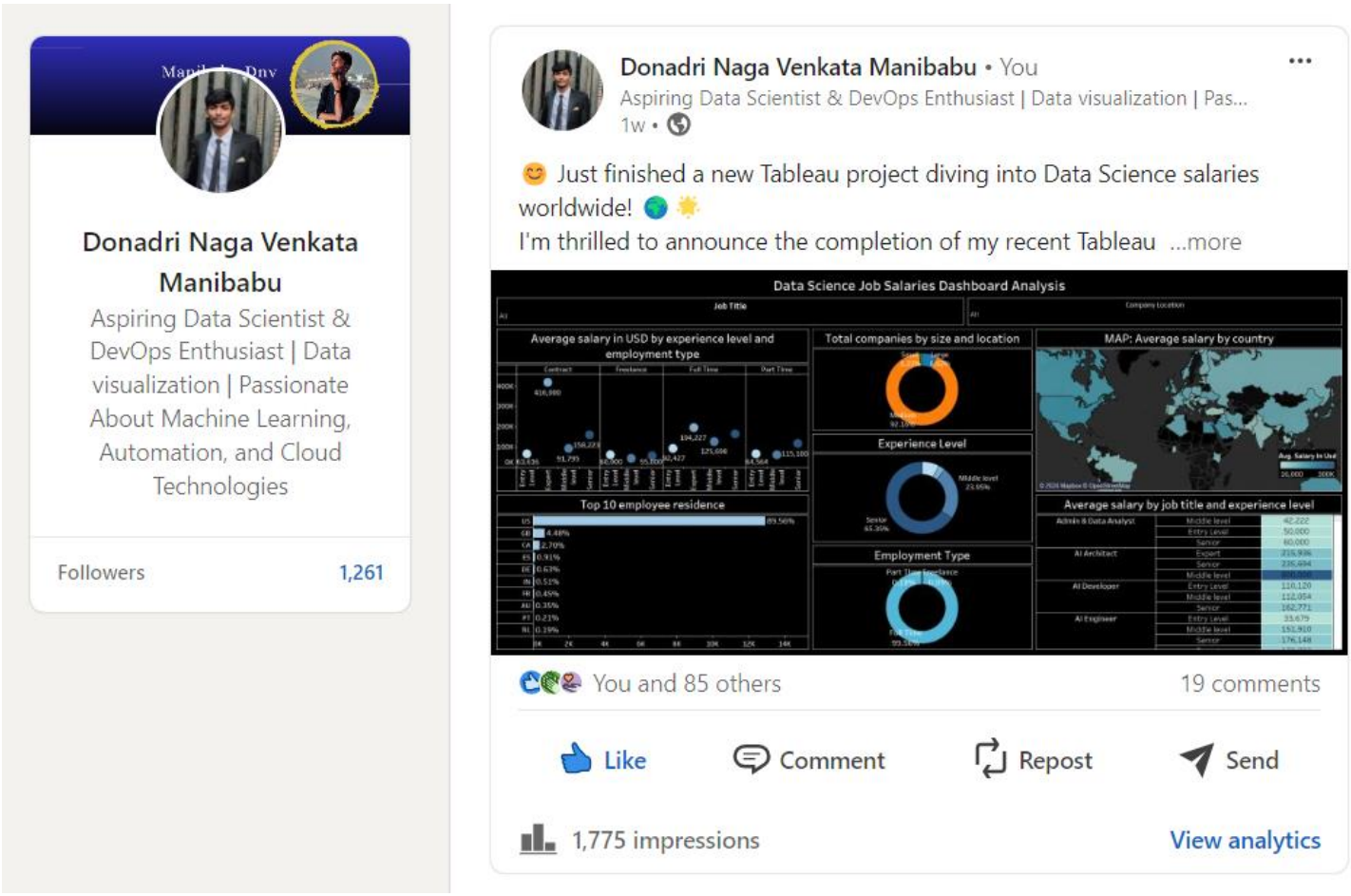
7. **Map: Average Salary by Country**

- This choropleth map visually represents average salaries across different countries. Darker shades indicate higher average salaries, while lighter shades indicate lower salaries. This map helps understand geographic disparities in salary, potentially highlighting regions with higher demand for data science skills and better compensation.



These detailed descriptions provide a deeper understanding of what each chart conveys and how it can be used to derive insights about the data science job market.

Linkedln ScreenShort:



List of Analysis with Results:

1. Average Salary in USD by Experience Level and Employment Type

- **Results**:
    - **Contract**: Entry Level: $63,616, Middle Level: $91,795, Senior: $158,223, Expert: $416,000
    - **Freelance**: Entry Level: $60,000, Middle Level: $55,000
    - **Full Time**: Entry Level: $92,427, Middle Level: $125,698, Senior: $194,227
    - **Part Time**: Entry Level: $64,564, Middle Level: $115,100

2. Total Companies by Size and Location

- **Results**:
    - **Small**: 1.22%
    - **Medium**: 92.16%
    - **Large**: 6.62%

3. Experience Level Distribution

- **Results**:
    - **Entry Level**: 10.49%
    - **Middle Level**: 23.95%
    - **Senior**: 65.35%

4. Employment Type Distribution

- **Results**:
    - **Full Time**: 99.56%
    - **Part Time**: 0.18%
    - **Freelance**: 0.09%

5. Top 10 Employee Residences

- **Results**:
    - **US**: 89.56%
    - **GB**: 4.48%
    - **CA**: 2.70%
    - **ES**: 0.91%
    - **DE**: 0.63%
    - **IN**: 0.51%
    - **FR**: 0.45%
    - **AU**: 0.35%
    - **PT**: 0.21%
    - **NL**: 0.19%

6. Average Salary by Job Title and Experience Level

- **Results**:
    - **Admin & Data Analyst**: Middle Level: $42,222, Entry Level: $50,000, Senior: $60,000
    - **AI Architect**: Expert: $215,936, Senior: $235,694
    - **AI Developer**: Middle Level: $800,000, Entry Level: $110,120, Senior: $162,771
    - **AI Engineer**: Entry Level: $33,679, Middle Level: $151,910, Senior: $176,148

7. Map: Average Salary by Country

- **Results**:
    - Salaries range from $16,000 to $300,000 across different countries, with higher salaries in regions like the US and lower salaries in other parts of the world.

8. Future Scope

1. Advanced Predictive Analytics

- **Incorporate Machine Learning Models**: Implement predictive models to forecast salary trends based on historical data, economic factors, and industry developments.
- **Career Path Analysis**: Use predictive analytics to suggest potential career paths and salary projections for professionals based on their current role and experience.

2. Real-time Data Integration

- **Live Data Feeds**: Integrate real-time data sources to keep the dashboard updated with the latest job market trends and salary information.
- **Dynamic Adjustments**: Allow for dynamic updates and adjustments based on real-time changes in the job market, such as new job postings and industry shifts.

3. Geographical Expansion

- **Global Salary Trends**: Expand the geographical scope to include more countries and regions, providing a more comprehensive global view of the data science job market.
- **Localized Insights**: Offer localized insights and comparisons to help professionals understand regional differences in salaries and job opportunities.

4. Enhanced Visualization

- **Interactive Features**: Incorporate more interactive elements, such as drill-down capabilities, filters, and customizable views, to allow users to explore data in more detail.
- **Advanced Visualizations**: Use advanced visualization techniques like heat maps, bubble charts, and network diagrams to present complex data more intuitively.

5. User Personalization

- **Custom Dashboards**: Allow users to create personalized dashboards tailored to their specific needs and preferences, such as focusing on specific job roles, regions, or experience levels.
- **Recommendation Systems**: Implement recommendation systems to suggest career opportunities, training programs, or job roles based on user profiles and data.

6. Integration with Other Platforms

- **Job Portals**: Integrate with job portals and professional networking sites to provide seamless access to job listings, networking opportunities, and industry news.
- **Educational Resources**: Link to educational platforms offering courses and certifications relevant to the data science field, helping users enhance their skills and career prospects.

7. Feedback and Continuous Improvement

- **User Feedback Mechanism**: Implement feedback mechanisms to gather user insights and continuously improve the dashboard's features and usability.
- **Regular Updates**: Ensure regular updates to the underlying data and analytical models to reflect the latest trends and maintain accuracy.

By incorporating these future enhancements, the Data Science Job Salaries Dashboard can evolve into a more powerful and versatile tool, providing deeper insights and greater value to users in the data science community.

9. References

1. Tableau Public Dashboard (2024). Data science salaries analysis - Visualizations designed for salaries trends and insights of data science.
2. Data Analysis Best Practices (2024). Guide to Effective Data Visualization in Tableau.
3. Regional Marketing Strategies (2024). Boosting Sales Performance in Competitive Markets.
4. Seasonal Sales Trends Report (2024). Impact of Seasonal Fluctuations on Coffee Sales.

10. Bibliography

- Chen, Y., & Gupta, R. (2023). Data Analytics for Business Growth. McGraw Hill Education.
  Smith, J. (2022). Visual Storytelling with Tableau: Strategies for Business Analysts. Wiley Publications.
- Coffey, M., & Kim, S. (2023). Consumer Preferences and Product Strategies in the Beverage Industry.
  Pearson Education.
- Tableau Documentation (2024). Official User Guide for Tableau Desktop.

Conclusion:

The Data Science Job Salaries Dashboard offers a comprehensive analysis of salary trends and employment dynamics within the data science field. By examining various factors such as experience level, employment type, job title, and company size, the dashboard provides valuable insights that are essential for professionals, employers, and industry analysts.

The analysis reveals that experience levels significantly impact salaries, with senior and expert roles commanding the highest compensation. Employment types also play a crucial role, with full-time positions being the most prevalent and typically offering the most stable and competitive salaries. The distribution of companies by size shows a dominance of medium-sized enterprises, highlighting their significant presence in the industry. Moreover, the geographical distribution of data science professionals indicates a strong concentration in the United States, followed by other key regions such as the United Kingdom, Canada, and Germany.

Key visualizations in the dashboard, including bar charts, pie charts, and choropleth maps, provide a clear and intuitive understanding of the data. These visual tools enable users to quickly identify patterns and trends, making the information accessible and actionable. For instance, the average salary by job title and experience level chart helps professionals understand potential career growth and salary expectations, while the map of average salaries by country offers a global perspective on compensation trends.

The future scope for the dashboard includes integrating real-time data feeds, enhancing predictive analytics, and expanding geographical coverage to provide a more holistic and dynamic view of the job market. User personalization features, such as custom dashboards and recommendation systems, can further enhance the user experience, making the tool even more valuable for career planning and decision-making.

In conclusion, the Data Science Job Salaries Dashboard is a powerful tool that sheds light on the intricate landscape of the data science job market. By leveraging detailed analysis and visualization, it helps stakeholders navigate the complexities of salary trends, employment types, and industry dynamics. Future enhancements will further solidify its role as an indispensable resource for understanding and optimizing career opportunities in data science.