

# Project

## Phase 2 : Decision Making

---

### Project Team - 2

Harsh Siddhapura

Shivkesh Madasu

Shrenika Soma

Jhansi Alugoju

04/14/2024

## Section - 1: Used Visualization Tools

**“List the visualization tools that you will use to create the dashboard. Explain why you chose these tools. This can be due to data-related issues or personal preference for certain development tools. Explain why you prefer some tools over others” (Canvas)**

We would primarily use the following tools for visualization. These tools are our preference due to their robust capabilities and user-friendly interfaces. They are:

- Tableau Prep Builder
- Tableau
- Table Calculations
- Story Creation

Tableau is a powerful data visualization tool that allows us to create various visualizations to present data and extract insights interactively. It's particularly useful for creating dashboards for data analysis, reporting, and decision-making purposes. Tableau's flexibility and customization options make it a superior choice for creating visually appealing and insightful dashboards.

Tableau Prep Builder is another tool to prepare and clean data. Data preparation is a critical step in any data analysis task. Tableau Prep Builder makes this process easier by providing an intuitive way to clean, shape, and aggregate data, crucial for accurate and meaningful data visualization. In addition to these, we would use Table Calculations to create “calculated fields” and other computations. Table Calculations provide a way to create new data from existing data.

Lastly, we would use the Creating a Story feature in Tableau for the final dashboard and story for aspiring job seekers. This feature allows us to present a sequence of visualizations that work together to convey information. It's a great way to showcase a progression, provide context, demonstrate how decisions relate to outcomes, or simply make a compelling case. In conclusion, combining these tools allows for a streamlined process from data preparation to insightful visualization, making them our preferred choice for this task. Their ease of use, flexibility, and powerful capabilities make them stand out among other data visualization tools.

As for other tools like Python, Matplotlib, Bokeh, and Excel, while they are powerful in their own right, they are not required for this specific task. Python and its libraries like Matplotlib and Bokeh are excellent for creating custom visualizations but require significant coding and data manipulation. On the other hand, Excel is excellent for fundamental data analysis and visualization, but it lacks Tableau's advanced data visualization capabilities. Therefore, to create an advanced, interactive dashboard with minimal coding, we find Tableau and its associated tools to be the most suitable.

## Section - 2: Explanation of Required Data Pre-processing

**“In case your data requires any kind of pre-processing such as computing certain attributes or removing missing values, explain how the data will be processed and prepared for visualization” (Canvas)**

To prepare the data for visualization in Tableau Prep Builder, we would first need to assess the quality and structure of the dataset. This involves identifying missing values, inconsistencies, or outliers that could affect the visualization process. Once the data quality check is complete, we can proceed with preprocessing steps tailored specifically for visualization. Here's how we would approach it:

- **Handling Missing Values:** We would check for missing values in each column and decide on the appropriate strategy to deal with them. This might involve imputing missing values, removing rows with missing data, or replacing them with placeholders, depending on the impact on visualization.
- **Data Type Conversion:** We would ensure that each column is assigned the correct data type. For example, numerical columns like 'salary' and 'remote\_ratio' should be converted to numeric types, while categorical columns like 'employment\_type' and 'job\_title' should be treated as strings or categories.
- **Currency Conversion:** Since the 'salary' column contains values in different currencies, we would convert them to a common currency, such as USD, for consistency in visualization. This would involve using exchange rates to convert salaries to a single currency.
- **Feature Engineering:** Depending on the visualization goals, we may need to create new features or derive additional insights from existing ones. For example, we could create a new feature indicating whether the employee works remotely based on the 'remote\_ratio' column.
- **Data Aggregation:** Depending on the level of detail required in the visualization, we may need to aggregate the data at different levels. For example, we could aggregate salary data by country or job title to visualize average salaries across different categories.
- **Data Cleaning and Standardization:** We would perform any necessary data cleaning and standardization steps to ensure consistency and accuracy in the dataset. This might involve removing duplicates, correcting spelling errors, or standardizing categorical values.
- **Data Joining or Blending:** If additional data sources are available or required for visualization, we may need to join or blend the dataset with other relevant data sources to enrich the analysis.

## Section - 3: List of Final Sets of Questions

**“List the final set of questions that the dashboard will be designed address. The dashboard users should be able to find answers for these questions by using your dashboard. List at least ten questions” (Canvas)**

1. What is the average salary in USD for employees categorized by their experience level and type of employment (full-time, part-time, contract, freelance)?
2. How does the average salary differ across various employment types such as contract, freelance, full-time, and part-time?
3. Which countries are among the top 10 in terms of employee residences, based on the provided data?
4. How many companies exist based on their size and location?
5. What percentage of companies fall into the categories of small, medium, and large based on their size?
6. What is the distribution of experience levels among the employees in the dataset?
7. What proportion of employees work part-time or freelance, as indicated in the dataset?
8. Considering the data provided, how does the average salary vary by country on a global scale?
9. In USD, what are the average salaries for specific job titles like 3D Computer Vision Researcher, Analytics Engineer, Applied Data Scientist, Applied Machine Learning Scientist, and BI Data Analyst across different experience levels?
10. How does the experience level of employees influence the average salary for the specified job titles in USD?
11. Which country has the highest number of employee residences according to the dataset?
12. How do salaries compare across different countries for similar roles and experience levels?
13. What is the distribution of remote work ratio among the employees represented in the data?
14. How does the size of a company affect the average salary of its employees?
15. How does the location of a company impact the average salary of its employees?

## Section - 4: Dashboard Plot Drafts

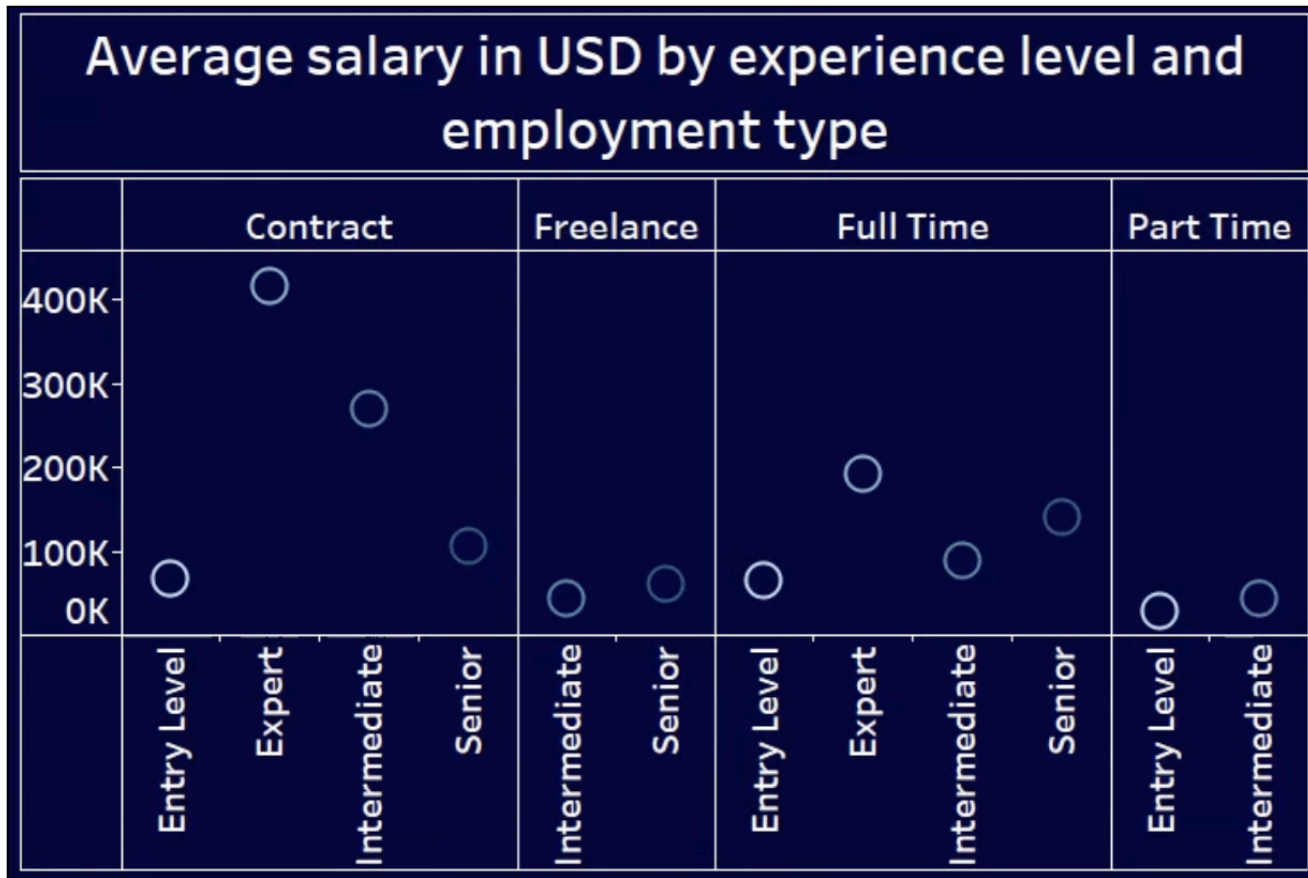
“For each of the questions listed above, think about the best plots that can be used to address it. Keep in mind that one question might require multiple plots to address. Alternatively, one plot can address multiple questions. The dashboard should contain at least ten plots. For each plot,

- Explain what it shows and how that relates to the set of questions.
- List the set of used pre-attentive attributes and colors.
- Include a rough, hand-drawn or computer-drawn, figure of the plot.” (Canvas)

The list of the plots that will be used to answer the questions is as follows:

- **Plot-1: Dot Plot - Average salary in USD by Experience Level and Employment Type**
  - **Questions:**
    - 1. What is the average salary in USD for employees categorized by their experience level and type of employment (full-time, part-time, contract, freelance)?
    - 2. How does the average salary differ across various employment types such as contract, freelance, full-time, and part-time?
    - 10. How does the experience level of employees influence the average salary for the specified job titles in USD?
    - 14. How does the size of a company affect the average salary of its employees?
    - 15. How does the location of a company impact the average salary of its employees?
  - **Explanation:** This plot will show the average salaries for different employment types (contract, freelance, full-time, part-time) at various experience levels. It directly answers questions 1 and 2. It also provides insights for questions 10, 14, and 15 about how experience level, company size, and location might affect the average salary.
  - **Pre-attentive attributes and colors:** The size of the circles represents the salary amounts, with larger circles indicating higher salaries. Different shades of blue are used to distinguish between employment types, each representing a different type.

- Graph:



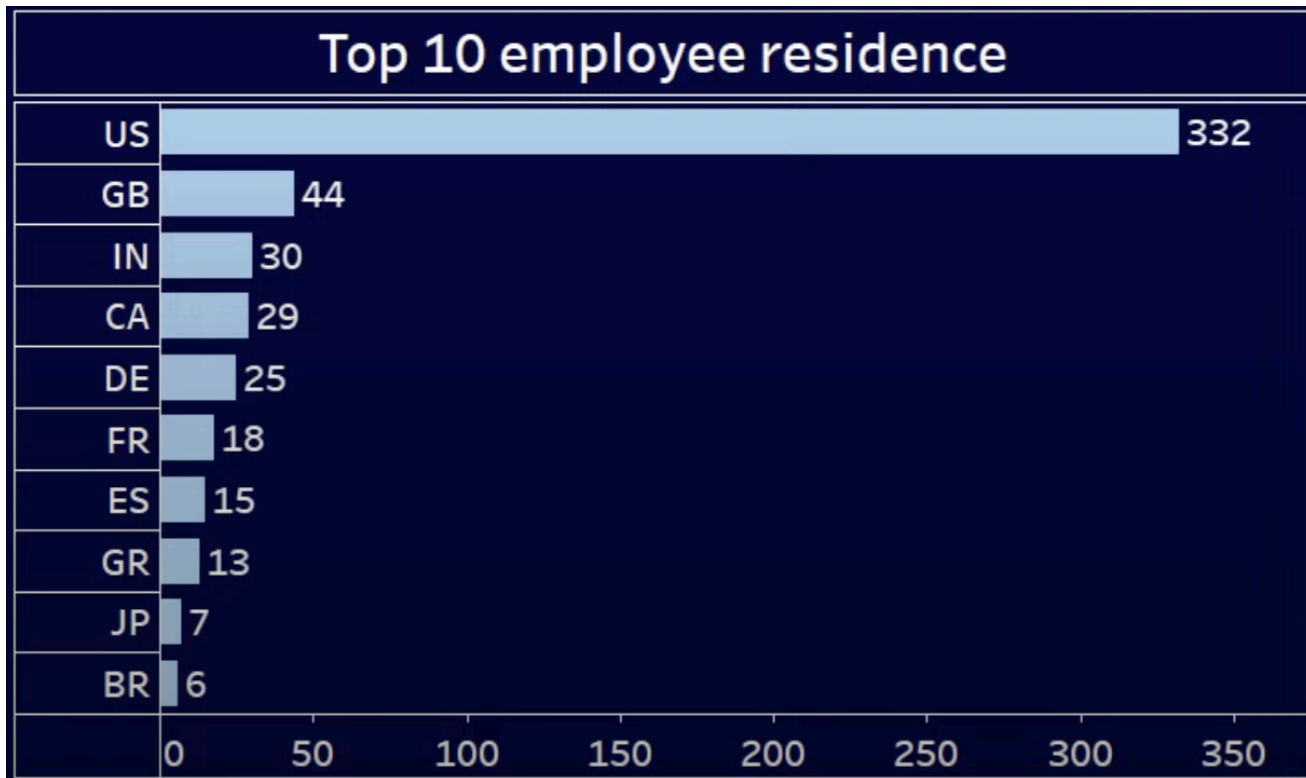
- **Plot-2: Bar Chart - Top 10 Employee Residence**

- **Questions:**

- 3. Which countries are among the top 10 in terms of employee residences, based on the provided data?
- 11. Which country has the highest number of employee residences according to the dataset?

- **Explanation:** This bar graph displays the top 10 countries where employees reside. It directly answers question 3 about the countries with the most employee residences. It also provides partial information for question 11 about the country with the highest number of employee residences.
- **Pre-attentive attributes and colors:** The length of the bars represents the number of employees, with longer bars indicating more employees. Shades of blue are used for different countries, each representing a different country.

- Graph:



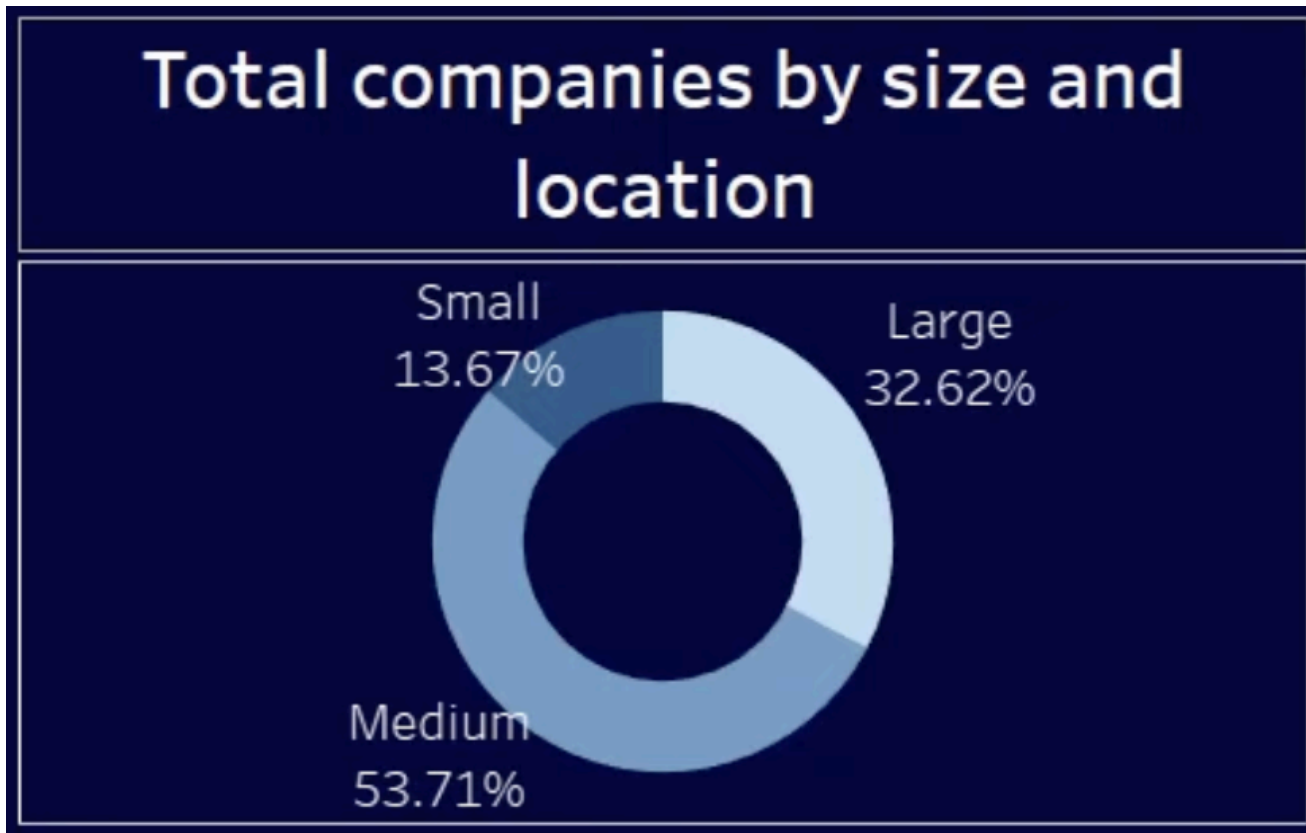
- **Plot-3: Donut Chart - Total Companies by Size and Location**

- Questions:

- 4. How many companies exist based on their size and location?
- 5. What percentage of companies fall into the categories of small, medium, and large based on their size?

- **Explanation:** This donut chart provides insights into company sizes and locations. They can be used to answer questions 4 and 5 about the number of companies by size and location, and the percentage of companies that are small, medium, or large.
- **Pre-attentive attributes and colors:** The segments' sizes in the donut charts represent the number of companies, with larger segments indicating more companies or employees. Shades of color are used for different categories within each chart, with each shade representing a different category.

- Graph:



- **Plot-4: Map - Average Salary by Country**

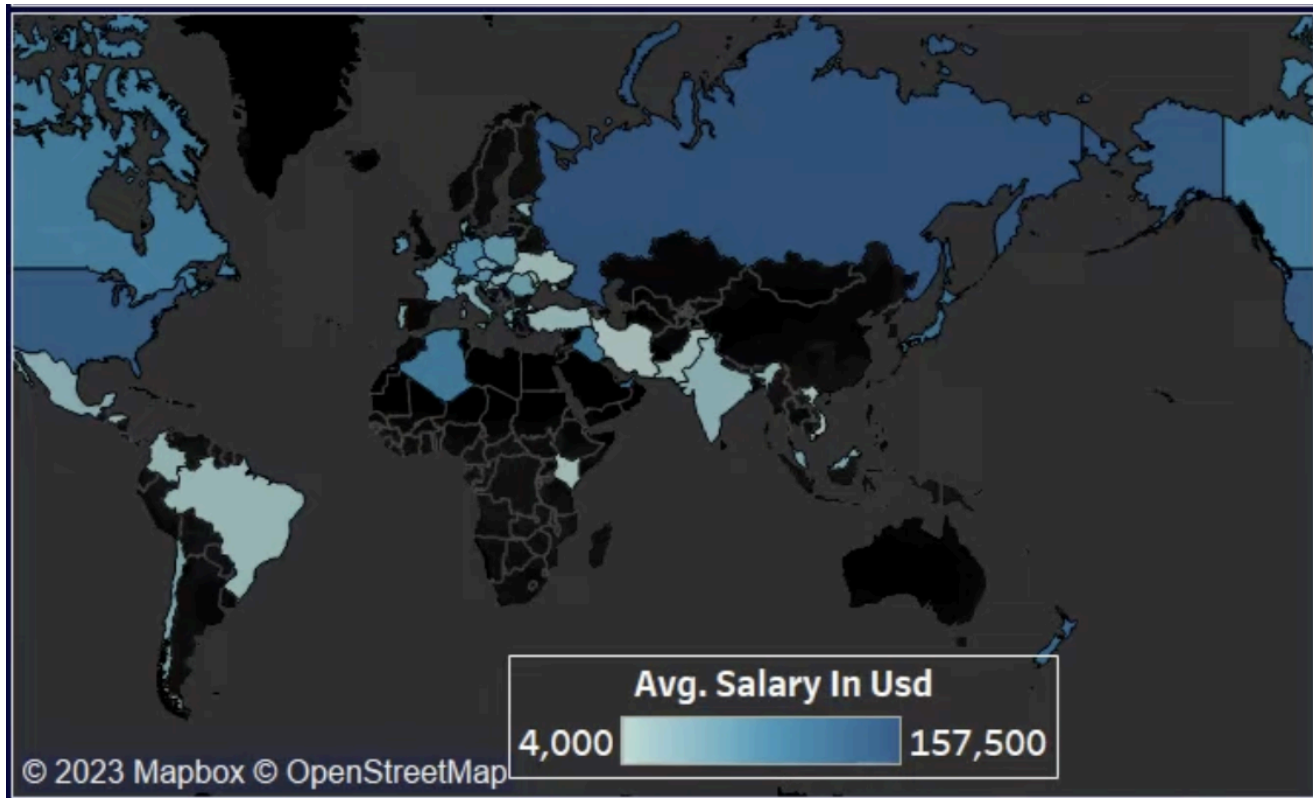
- **Questions:**

- 8. Considering the data provided, how does the average salary vary by country on a global scale?
- 12. How do salaries compare across different countries for similar roles and experience levels?
- 15. How does the location of a company impact the average salary of its employees?

- **Explanation:** This map visualization will show average salaries in various countries worldwide. It is color-coded to represent different salary ranges. This plot addresses questions 8, 12, and 15 as it allows users to visually compare salaries across countries for similar roles or experience levels.



- **Pre-attentive attributes and colors:** The color intensity on the map represents average salary amounts, with darker shades indicating higher salaries. Geographical locations are easily identifiable.
- **Graph:**



- **Plot-5: Table Heat Map - Average Salary by Job Title & Experience Level**

- **Questions:**
  - 9. In USD, what are the average salaries for specific job titles like 3D Computer Vision Researcher, Analytics Engineer, Applied Data Scientist, Applied Machine Learning Scientist, and BI Data Analyst across different experience levels?
  - 10. How does the experience level of employees influence the average salary for the specified job titles in USD?

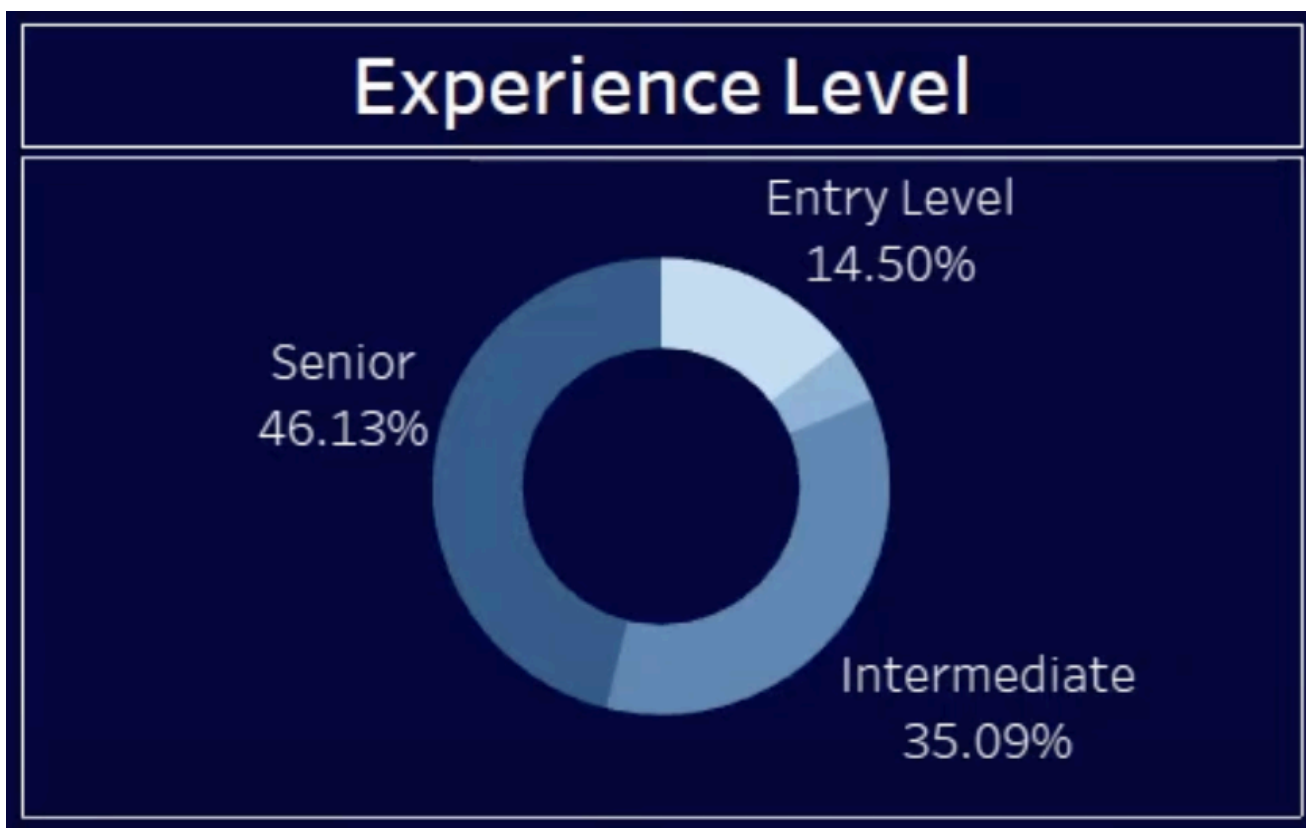
- **Explanation:** This table heat map lists specific job titles and their respective average salaries at various experience levels. It directly answers questions 9 and 10 as users can quickly identify how much each role earns on average at each level of expertise.
- **Pre-attentive attributes and colors:** The text information is organized into rows/columns, and distinct rows make it easy to read individual data points. The specific intensity of colors is used in this plot to see higher and lower salaries.
- **Graph:**

Average salary by job title and experience level		
3D Computer Vision Researcher	Intermediate	5,409
AI Scientist	Intermediate	160,000
	Senior	55,000
	Entry Level	21,987
Analytics Engineer	Senior	195,000
	Expert	155,000
Applied Data Scientist	Senior	278,500
	Entry Level	110,037
	Intermediate	105,619
Applied Machine Learning Scientist	Intermediate	178,800
	Entry Level	31,875
BI Data Analyst	Expert	150,000
	Intermediate	78,086

● **Plot-6: Donut Chart - Experience Level vs Location**

- **Questions:**
  - 6. What is the distribution of experience levels among the employees in the dataset?

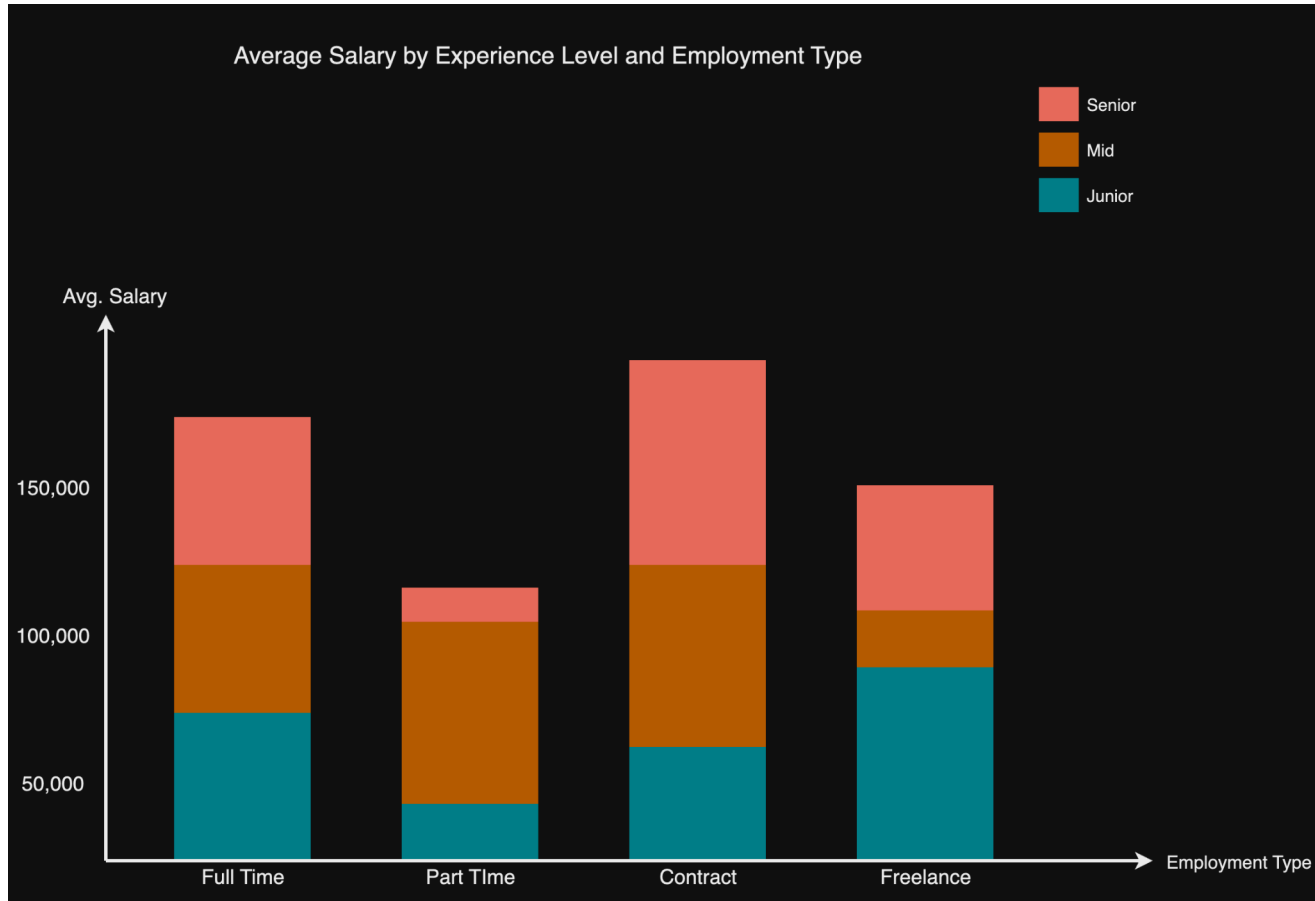
- **Explanation:** This donut chart provides insights into experience levels and locations. It can be used to answer question 6 about the percentage of experience level and location.
- **Pre-attentive attributes and colors:** The segments' sizes in the donut charts represent the percentage of experience level, with larger segments indicating more employees. Shades of color are used for different categories within each chart, with each shade representing a different category.
- **Graph:**



- **Plot-7: Bar Chart - Average Salary by Experience Level and Employment Type**

- **Questions:**
  - 1. What is the average salary in USD for employees categorized by their experience level and type of employment (full-time, part-time, contract, freelance)?
  - 2. How does the average salary differ across various employment types such as contract, freelance, full-time, and part-time?

- 7. What proportion of employees work part-time or freelance, as indicated in the dataset?
- 10. How does the experience level of employees influence the average salary for the specified job titles in USD?
- **Explanation:** This chart can address Questions 1, 2, 7, and 10. It shows the average salary in USD for different experience levels (MI, SE) and employment types (FT, PT, FL, CO). The x-axis represents the experience level and employment type, and the y-axis represents the average salary in USD.
- **Pre-attentive attributes and colors:** Positions along a common scale and length are the pre-attentive attributes used in this chart. The position along the x-axis represents the combination of experience level and employment type, while the length of the bars represents the average salary in USD. Different colors are used for different employment types. This helps to distinguish between full-time, part-time, contract, and freelance employees.
- **Graph:**



- **Plot-8: Donut Chart - Employment Type vs Location**

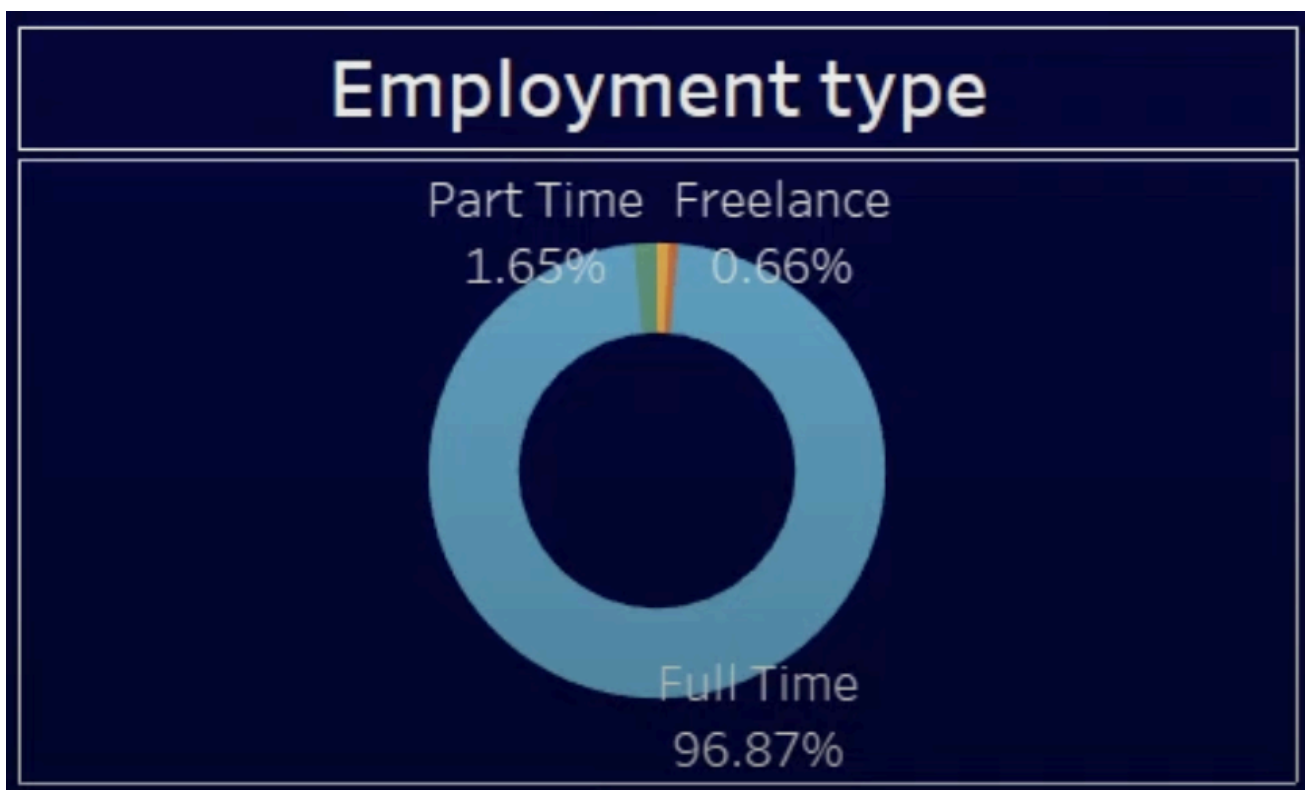
- **Questions:**

- 7. What proportion of employees work part-time or freelance, as indicated in the dataset?
- 13. What is the distribution of remote work ratio among the employees represented in the data?

- **Explanation:** This donut chart provides insights into the percentage of employment type and locations They can be used to answer questions 7 and 13 about the percentage of employment type and location.

- **Pre-attentive attributes and colors:** The segments' sizes in the donut charts represent the percentage of employment type, with larger segments indicating more employees. Shades of color are used for different categories within each chart, with each shade representing a different category.

- **Graph:**



• **Plot-9: Line Graph - Average Salary at Different Experience Levels in USD vs Work Year**

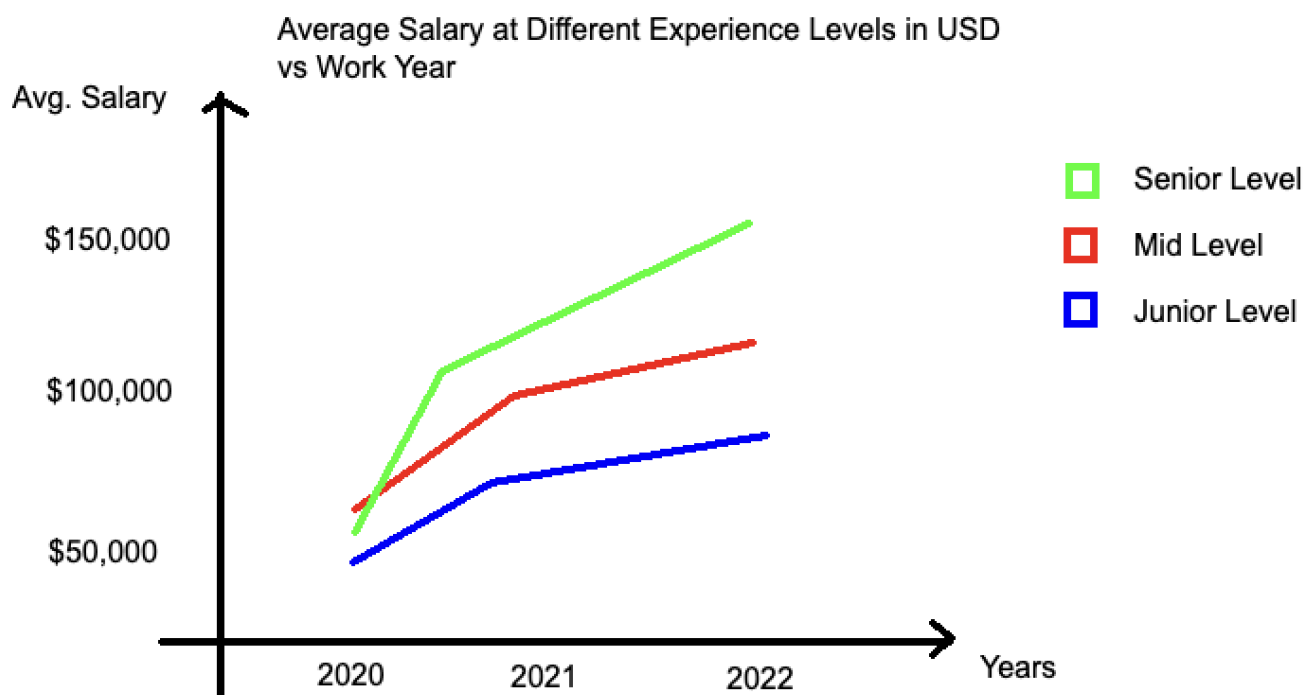
○ **Questions:**

- 6. What is the distribution of experience levels among the employees in the dataset?
- 10. How does the experience level of employees influence the average salary for the specified job titles in USD?

○ **Explanation:** This graph can address questions 6 and 10 about how the average salary in USD has changed over the years at various experience levels. The x-axis represents the work year, and the y-axis represents the average salary in USD. Each line in the graph represents a different experience level.

○ **Pre-attentive attributes and colors:** Positions along a common scale and length are the pre-attentive attributes used in this chart. The position along the x-axis represents the work year, while the position along the y-axis represents the average salary in USD. The length between points on a line represents the change in average salary from one year to the next. Different colors are used for different experience levels. This helps to distinguish between mid-level and senior-level employees.

○ **Graph:**



- **Plot-10: Bar Chart - Distribution of Experience Levels Among Employees**

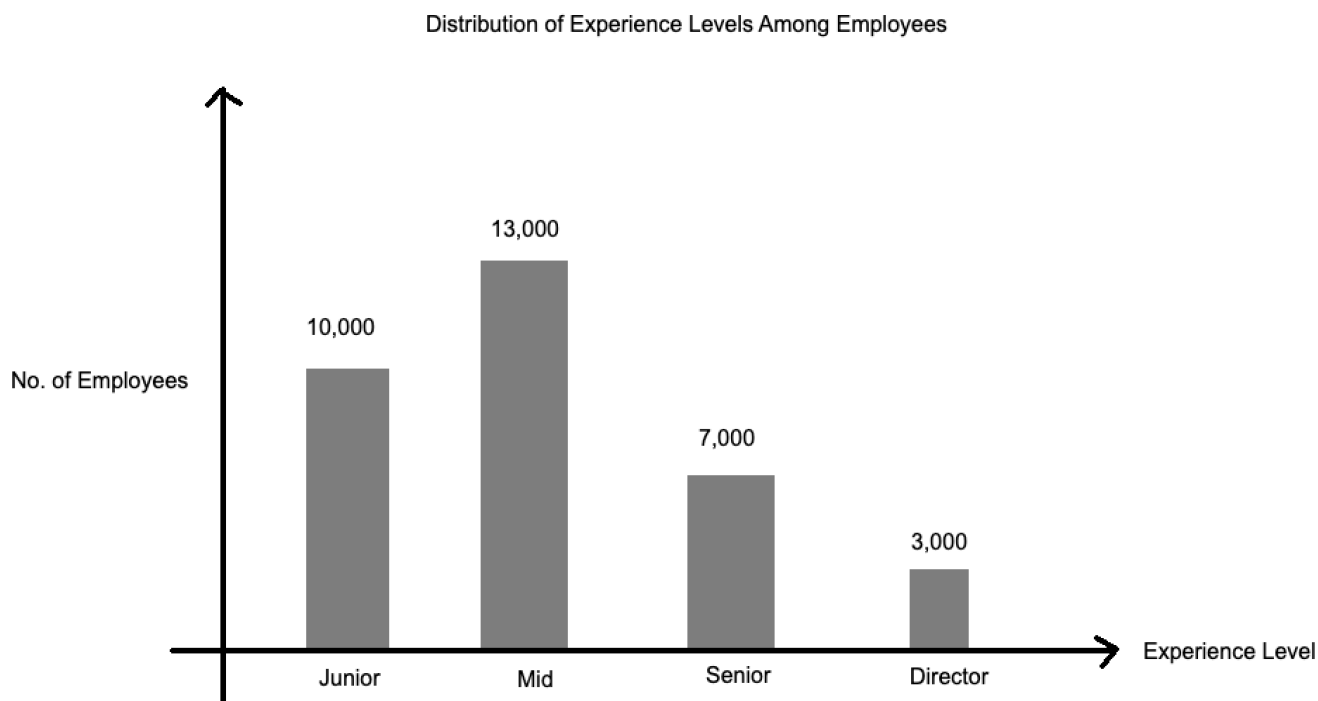
- **Questions:**

- 6. What is the distribution of experience levels among the employees in the dataset?

- **Explanation:** This chart can address Question 6. It shows the number of employees at each experience level. The x-axis represents the experience level, and the y-axis represents the number of employees.

- **Pre-attentive attributes and colors:** Positions along a common scale and length are the pre-attentive attributes used in this chart. The position along the x-axis represents the experience level, while the length of the bars represents the number of employees. Different colors are used for different experience levels. This helps to distinguish between mid-level and senior-level employees.

- **Graph:**



## Section - 5: Dashboard Interactivity

“The dashboard user can change the visualizations via interactive controls. If your dashboard contains any controls,

- List what they will be used for
- Which plots are connected to each one
- The value range for each control and whether or not it is loaded from a certain attribute in the data.” (Canvas)

The dashboard user can change the visualizations via interactive controls such as:

- **Dropdown Filter: Company Locations**
  - **Usage:** The "Company Locations" dropdown filter on the dashboard is a powerful tool that allows users to focus on data related to companies in a specific location. When a location is selected, all plots on the dashboard are updated to reflect this choice. For instance, the "Average salary in USD by Experience Level and Employment Type" plot will show data for the chosen location, as will the "Top 10 Employee Residence" bar chart.

The "Total Companies by Size and Location" donut chart will display the distribution of companies by size in the selected location, while the "Map - Average Salary by Country" will highlight the chosen location. The "Table Heat Map - Average Salary by Job Title & Experience Level" will show salary information for the selected location, and both the "Experience Level vs Location" and "Employment Type vs Location" donut charts will show distributions for the chosen location.

This interactive control enhances the user's ability to explore and understand the data in a more detailed and customized manner. The actual functionality of this control may vary depending on the data visualization tool or library used. The description provided is a general guideline and may need to be adjusted based on the specific characteristics of the data.

- **Plots connected:**
  - Plot-1: Dot Plot - Average salary in USD by Experience Level and Employment Type



- Plot-2: Bar Chart - Top 10 Employee Residence
- Plot-3: Donut Chart - Total Companies by Size and Location
- Plot-4: Map - Average Salary by Country:
- Plot-5: Table Heat Map - Average Salary by Job Title & Experience Level
- Plot-6: Donut Chart - Experience Level vs Location
- Plot-8: Donut Chart - Employment Type vs Location
- **Range:** 'DE', 'JP', 'GB', 'HN', 'US', 'HU', 'NZ', 'FR', 'IN', 'PK', 'CN', 'GR', 'AE', 'NL', 'MX', 'CA', 'AT', 'NG', 'ES', 'PT', 'DK', 'IT', 'HR', 'LU', 'PL', 'SG', 'RO', 'IQ', 'BR', 'BE', 'UA', 'IL', 'RU', 'MT', 'CL', 'IR', 'CO', 'MD', 'KE', 'SI', 'CH', 'VN', 'AS', 'TR', 'CZ', 'DZ', 'EE', 'MY', 'AU', 'IE'
- **Attribute:** company\_location

- **Slider Filter: Years**

- **Usage:** The “Years” slider filter on the dashboard allows users to focus on data from specific years. The range for this slider is from 2020 to 2022. When a year or a range of years is selected, the “Average Salary at Different Experience Levels in USD vs Work Year” line graph updates to reflect this choice.

Specifically, the line graph will show the average salary in USD for different experience levels for the selected year(s). This can provide insights into how salaries have changed over time and how these changes vary between different experience levels.

- **Plots connected:**
  - Plot-9: Line Graph - Average Salary at Different Experience Levels in USD vs Work Year
- **Range:** 2020 to 2022
- **Attribute:** work\_year

- **Dropdown Filter: Job Titles**

- **Usage:** The "Job Titles" dropdown filter on the dashboard allows users to focus on data related to specific job titles. When a job title is selected, all plots on the dashboard are updated to reflect this choice.

For instance, the plot of the "Average salary in USD by Experience Level and Employment Type" will show data for the chosen job title. The "Top 10 Employee Residence" bar chart will display the top 10 residences for employees with the selected job title. The "Total Companies by Size and Location" donut chart will show the distribution of companies that employ individuals with the selected job title.

The "Map - Average Salary by Country" will highlight the average salary for the chosen job title in different countries. The "Table Heat Map - Average Salary by Job Title & Experience Level" will show salary information for the selected job title at different experience levels.

The "Experience Level vs Location" and "Employment Type vs Location" donut charts will show distributions for the chosen job title in different locations and employment types.

This interactive control enhances the user's ability to explore and understand the data in a more detailed and customized manner. The actual functionality of this control may vary depending on the data visualization tool or library used. The description provided is a general guideline and may need to be adjusted based on the specific characteristics of the data.

- **Plots connected:**
  - Plot-1: Dot Plot - Average salary in USD by Experience Level and Employment Type
  - Plot-2: Bar Chart - Top 10 Employee Residence
  - Plot-3: Donut Chart - Total Companies by Size and Location
  - Plot-4: Map - Average Salary by Country:
  - Plot-5: Table Heat Map - Average Salary by Job Title & Experience Level
  - Plot-6: Donut Chart - Experience Level vs Location

- Plot-8: Donut Chart - Employment Type vs Location
  - **Range:** 'Data Scientist', 'Machine Learning Scientist', 'Big Data Engineer', 'Product Data Analyst', 'Machine Learning Engineer', 'Data Analyst', 'Lead Data Scientist', 'Business Data Analyst', 'Lead Data Engineer', 'Lead Data Analyst', 'Data Engineer', 'Data Science Consultant', 'BI Data Analyst', 'Director of Data Science', 'Research Scientist', 'Machine Learning Manager', 'Data Engineering Manager', 'Machine Learning Infrastructure Engineer', 'ML Engineer', 'AI Scientist', 'Computer Vision Engineer', 'Principal Data Scientist', 'Data Science Manager', 'Head of Data', '3D Computer Vision Researcher', 'Data Analytics Engineer', 'Applied Data Scientist', 'Marketing Data Analyst', 'Cloud Data Engineer', 'Financial Data Analyst', 'Computer Vision Software Engineer', 'Director of Data Engineering', 'Data Science Engineer', 'Principal Data Engineer', 'Machine Learning Developer', 'Applied Machine Learning Scientist', 'Data Analytics Manager', 'Head of Data Science', 'Data Specialist', 'Data Architect', 'Finance Data Analyst', 'Principal Data Analyst', 'Big Data Architect', 'Staff Data Scientist', 'Analytics Engineer', 'ETL Developer', 'Head of Machine Learning', 'NLP Engineer', 'Lead Machine Learning Engineer', 'Data Analytics Lead'
  - **Attribute:** job\_title

## Section - 6: References

1. Dataset: Data Science Job Salaries

<https://www.kaggle.com/datasets/ruchi798/data-science-job-salaries>

2. Sample Mural Board used for Reference:

<https://app.mural.co/t/dvproject7266/m/dvproject7266/1680555603113/665bc60debfd6186673e06d0766c5f6101e556ab?sender=uab3d8f6d05efd4b7a75a4728>

3. Team's Mural Board:

<https://app.mural.co/t/datavisualization9024/m/datavisualization9024/1712091965326/958f0526d92ec8f0d396974a6a49c06289e5ca2e?sender=uab3d8f6d05efd4b7a75a4728>