Thursday, October 10, 2024

SALARY DISTRIBUTION ANALYSIS

AI/ML SALARIES

Objective : Analyze the distribution of salaries across different job roles experience levels , and , Locations

SQL Queries use GROUP BY and ORDER to calculate average, minimum and maximum salaries by job role.

To create a **Salary Distribution Analysis** using **Microsoft Excel** and **SQL** in **Microsoft SQL Server Express**, we will break the process into two main tasks:

1. **SQL Analysis**: Use SQL to analyze the salary distribution, focusing on calculating the **average**, **minimum**, and **maximum salaries** by **job role**, **experience level**, and **location**.
2. **Excel Visualization**: Use Excel to create visual representations (charts, histograms) to better understand the salary distribution.

Part 1: SQL Queries for Salary Distribution Analysis

In SQL, you will perform an analysis to group data by **job role**, **experience level**, and **location**. You will use aggregate functions like `AVG()`, `MIN()`, and `MAX()` to calculate salary statistics.

Assumptions:

- **Table Name**: `salaries`
- **Columns**:
  - `job_role` (job title or role of the employee)
  - `experience_level` (e.g., "Junior", "Mid", "Senior")
  - `location` (the location of the employee)
  - `salary` (the salary of the employee)

```
-- Analyzing salary distribution by job role, experience level, and location
SELECT
    job_role,
    experience_level,
    location,
    AVG(salary) AS average_salary,
    MIN(salary) AS min_salary,
    MAX(salary) AS max_salary
FROM salaries
GROUP BY
    job_role,
    experience_level,
    location
ORDER BY
    job_role,
    experience_level,
    location;
```

Explanation:

- **`AVG(salary)`**: This calculates the average salary for each combination of `job_role`, `experience_level`, and `location`.
- **`MIN(salary)`**: This returns the minimum salary for each combination of `job_role`, `experience_level`, and `location`.
- **`MAX(salary)`**: This returns the maximum salary for each combination of `job_role`, `experience_level`, and `location`.
- **`GROUP BY job_role, experience_level, location`**: This groups the data by the three categories to perform aggregate calculations for each unique combination.
- **`ORDER BY job_role, experience_level, location`**: This orders the results in a clear and structured way.

Part 2: Visualizing the Salary Distribution in Microsoft Excel

Once you have the data from SQL, you can use **Microsoft Excel** to visualize the salary distribution across the different segments.

Steps to Visualize the Data in Excel:

1. **Export SQL Query Results to Excel**:
    - After running the SQL query, export the query results to an Excel file.
    - In SQL Server Management Studio (SSMS), right-click on the results grid and choose **Save Results As...**. Choose Excel as the file format.
2. **Create a Pivot Table in Excel**:
    - Open the exported file in Excel.
    - Go to the **Insert** tab and click **PivotTable**.
    - In the PivotTable Field List:
        - Drag `job_role`, `experience_level`, and `location` to the **Rows** section.
        - Drag `average_salary`, `min_salary`, and `max_salary` to the **Values** section.
3. **Create Visualizations**:
    - **Bar Chart** for **Salary Distribution by Job Role**:
        - Select the data for `job_role` and the corresponding salary columns (`average_salary`, `min_salary`, and `max_salary`).
        - Go to the **Insert** tab and choose **Bar Chart** or **Column Chart**.

        - Customize the chart to show salary distributions for each `job_role`.
    - **Pie Chart** for **Salary Distribution by Location**:
        - Select the data for `location` and `average_salary`.
        - Go to the **Insert** tab and choose **Pie Chart** to show how salaries vary across locations.
    - **Histogram** for **Salary Ranges**:
        - Select the `salary` column to visualize how salaries are distributed (frequency distribution).
        - Go to **Insert > Histogram** to create a histogram that shows salary ranges.
4. **Formatting and Analysis**:
    - Format the numbers as currency to ensure clarity.
    - Add titles and axis labels to charts for better interpretation.
    - Filter or use slicers in Excel to drill down into specific job roles, experience levels, or locations.

Example Output:

1. **Pivot Table Example**:

Job Role	Experience Level	Location	Average Salary	Min Salary	Max Salary
Data Scientist	Junior	NY	$80,000	$65,000	$95,000
Data Scientist	Senior	CA	$120,000	$100,000	$150,000
Software Eng.	Mid	TX	$95,000	$75,000	$120,000
Software Eng.	Senior	FL	$110,000	$90,000	$130,000

5. **Bar Chart Example** (Salary Distribution by Job Role):
    - A bar chart will show each job role along the x-axis and the corresponding average salary on the y-axis, giving a clear visual comparison of salaries across roles.
3. **Histogram Example**:
    - A histogram will show the salary ranges (e.g., $50,000-$60,000, $60,000-$70,000) and the frequency of employees in each range.

Summary:

By using **SQL** for data aggregation and **Excel** for visualization, you can analyze and present the salary distribution for different job roles, experience levels, and locations in a clear, actionable

way. This combination of SQL queries for data extraction and Excel for reporting and visualization is a powerful approach to performing salary distribution analysis.