**Objective of the project:**

To provide an interactive analysis of salary trends across industries, genders, and locations. It highlights key metrics like salary distribution, experience-based variations, and gender pay gaps in a way that users can explore the insights to make informed decisions to further the development of their respective industry and job titles.

**Dataset Description:**

The dataset had 28,103 rows in total with 16 columns denoting various parameters and attributes related to the salary data. Please go through the columns and their description provided below.

* **Age range - Defines which age range the respective employee falls in, eg: under 18, over 65**
* **Industry - Denotes which industry the employee belongs to**
* **Job title - The title of the job**
* **Job clarification – Clarification provided by the employee on their job description**
* **Annual salary - Annual salary package of each employee**
* **Annual Monetary compensation - Monetary compensation received by employees annually**
* **Currency - Currency in which they receive their salary**
* **Income clarification - Further clarification of other income through their job**
* **Country - Which country the employee belongs to or for which country they work**
* **State - State of the employee**
* **City - Current residential city of the worker**
* **Overall experience - Overall experience of the employee**
* **Field experience - Personnel’s experience in the specific field**
* **Education - the level of education an employee has completed**
* **Gender – gender of an employee**

**Additionally added columns:**

* **Annual salary in USD**
* **Additional monetary compensation in USD**

**Data Cleansing**

* **The dataset requires basic cleaning such as formatting the datatype of the measures in it, for example – currency formatting and attributes to text which were in general format. The missing values in details like salary were filled with average of that specific column whereas if the columns like education, industry or job title is empty those rows were removed.**
* **Apart from that, Country names were uniformed for example – there was lot of variations and spelling mistakes in the country names provided. USA was mentioned as United States, United states of America, US of America and USB. These names were rectified and converted to single entity – USA and so on for other countries such as Canada, UK, Australia and further.**
* **No changes were made to the job clarification and income clarification owing to the nature of the data as it provides the employees understanding of their salary and the income they receive**. We’ve had multiple empty rows where the gender was empty which was filled with common genders to make sure the number of employees is not reduced.

**An addition of new two columns were added in the name of “Annual Salary in USD” and “Additional monetary compensation in USD” to provide the insights based on a single currency and not with multiple currency. There was a formula used in order to convert all the salary and compensation values into USD which is provided below.**

**Formula for Annual Salary in USD**

**=IF(K2="USD",G2,IF(K2="GBP",G2/0.7897,IF(K2="CAD",G2/1.4345,IF(K2="EUR",G2/0.9534,IF(K2="NZD",G2/1.7559,IF(K2="CHF",G2/0.8954,IF(K2="ZAR",G2/20.482,IF(K2="SEK",G2/11.153,IF(K2="HKD",G2/7.7735,IF(K2="JPY",G2/149.45,"Currency Not Found"))))))))))**

**K2 - Currency to determine whether it is USD, CAD, JPY, HKD or other**

**G2 – The actual salary which is in a different currency received by the employee**

**Formula for Monetary compensation in USD**

**=IF(K2="USD",I2,IF(K2="GBP",I2/0.7897,IF(K2="CAD",I2/1.4345,IF(K2="EUR",I2/0.9534,IF(K2="NZD",I2/1.7559,IF(K2="CHF",I2/0.8954,IF(K2="ZAR",I2/20.482,IF(K2="SEK",I2/11.153,IF(K2="HKD",I2/7.7735,IF(K2="JPY",I2/149.45,"Currency Not Found"))))))))))**

**K2 - Denotes which currency**

**I2 – Actual compensation received by the employee**

**Using these 2 formulas all the salary details and compensation details are converted to single currency to streamline the insight extraction.**

**These two column values are now converted to USD by assigning the datatype to currency in accounting.**

**All the blank cells in the rows are converted to null instead of 0 to make sure this does not become an error while uploading the data to SQL.**

**Importing to SQL**

To import the data to SQL, the excel file was converted to CSV format and saved in a new folder called “uploads” which was created in the server folder of MySQL workbench.

The method I used to import the data is LoadIn file method. After creating a specific database and table for the salary data in SQL, I ran a query to get the data from the uploads folder which was mentioned above. The code for creating a table and exporting the csv file is provided below for reference -

**Database creation:**

create database Milestone;

use Milestone;

**Table creation:**

CREATE TABLE salary\_data (

age\_range VARCHAR(100),

industry VARCHAR(255),

job\_title VARCHAR(255),

annual\_salary INT,

annual\_salary\_usd INT,

additional\_monetary\_compensation DECIMAL(10,2) NULL,

additional\_monetary\_compensation\_usd INT,

currency VARCHAR(10),

other\_currency VARCHAR(50),

income\_clarification TEXT,

country VARCHAR(255),

state VARCHAR(255),

city VARCHAR(255),

experience\_overall VARCHAR(100),

experience\_field VARCHAR(100),

education\_level VARCHAR(100),

gender VARCHAR(100)

);

**Importing data:**

LOAD DATA LOCAL INFILE 'C:\\Users\\manik\\Desktop\\Milestone.csv'

INTO TABLE salary\_data

FIELDS TERMINATED BY ','

ENCLOSED BY '"'

LINES TERMINATED BY '\n'

IGNORE 1 ROWS;

After importing the data to SQL, few of the queries had to be executed to see whether there are any errors or warnings in the csv file that was uploaded. After a thorough inspection, the warning was rectified for example - the income clarification and job clarification strings were too long for the space assigned which was sorted by increasing the storage space for the respective columns.

**Queries and Results:**

1. Average Salary by Industry and Gender

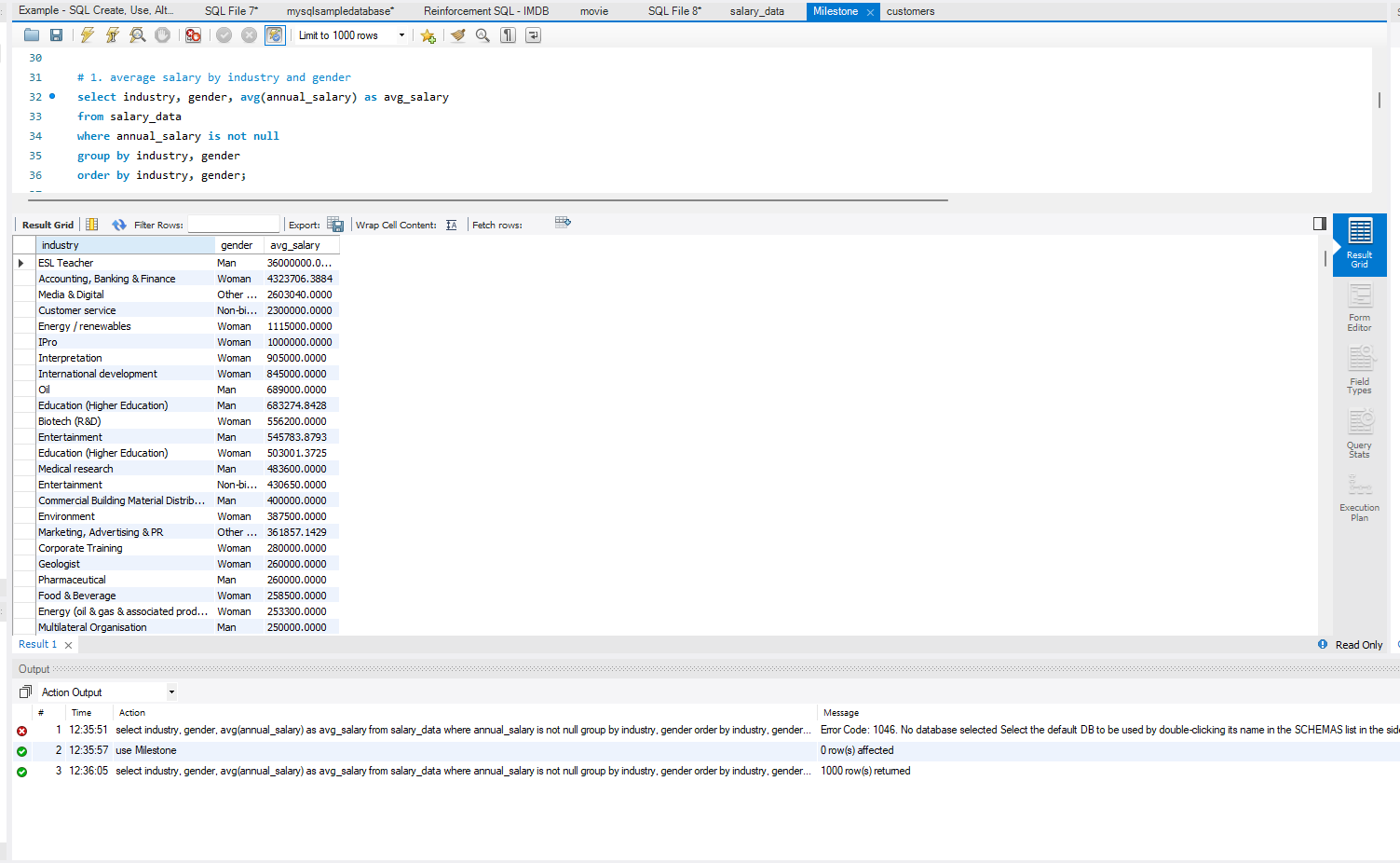
select industry, gender, avg(annual\_salary) as avg\_salary

from salary\_data

where annual\_salary is not null

group by industry, gender

order by industry, gender;



1. Total Salary Compensation by Job Title

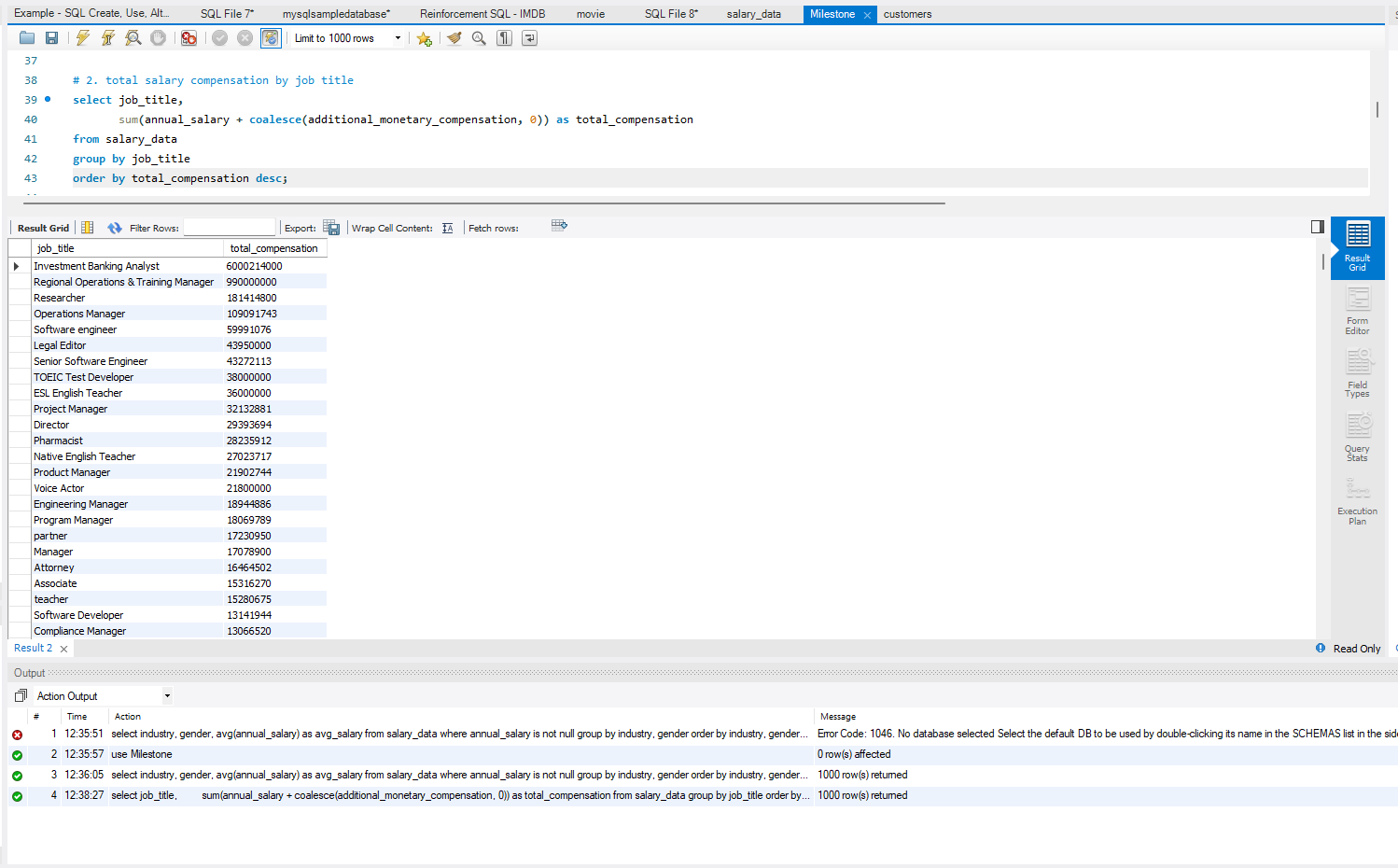
select job\_title,

sum(annual\_salary + coalesce(additional\_monetary\_compensation, 0)) as total\_compensation

from salary\_data

group by job\_title

order by total\_compensation desc;



1. Salary Distribution by Education Level

select education\_level,

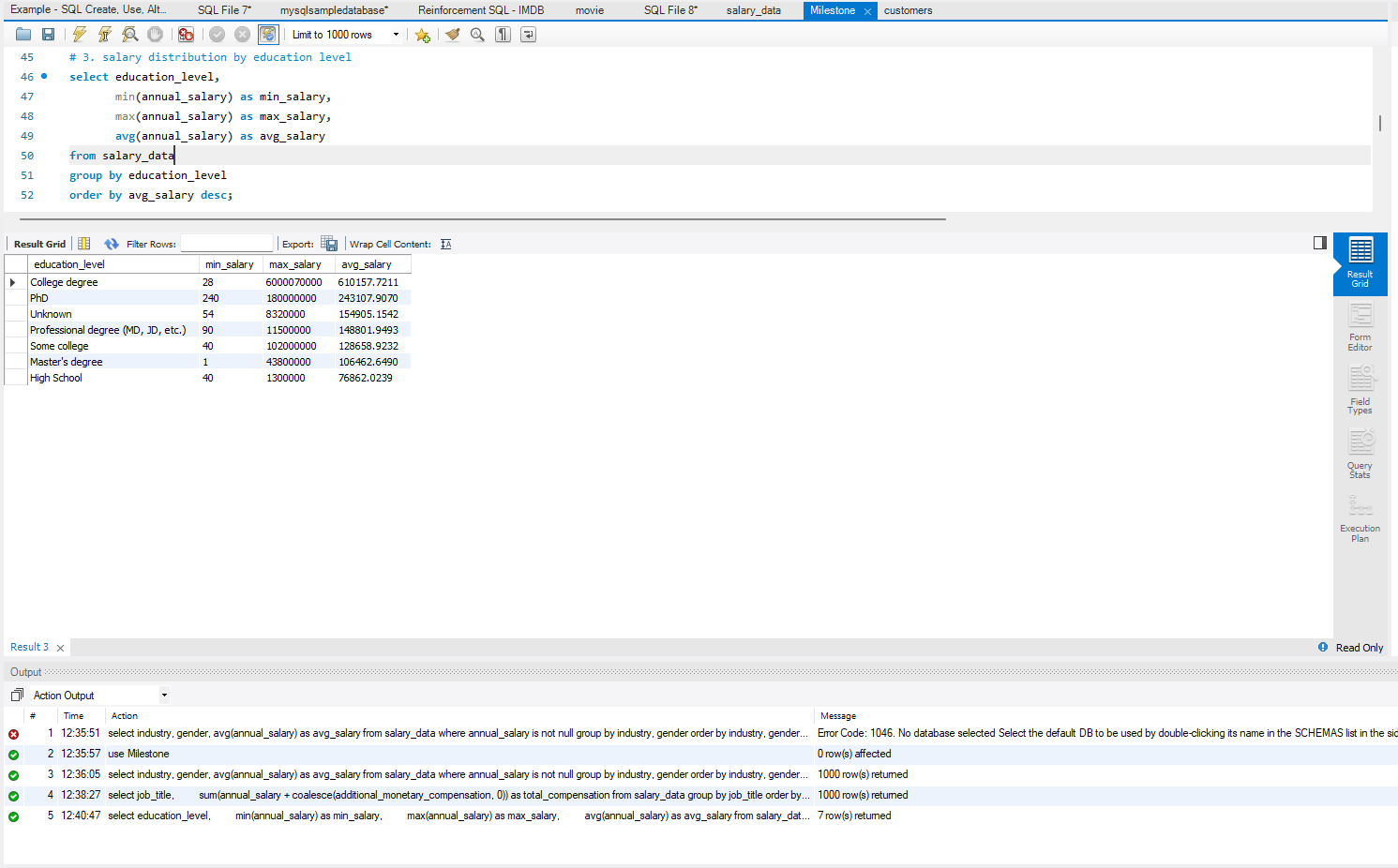
min(annual\_salary) as min\_salary,

max(annual\_salary) as max\_salary,

avg(annual\_salary) as avg\_salary

from salary\_data

group by education\_level

order by avg\_salary desc;

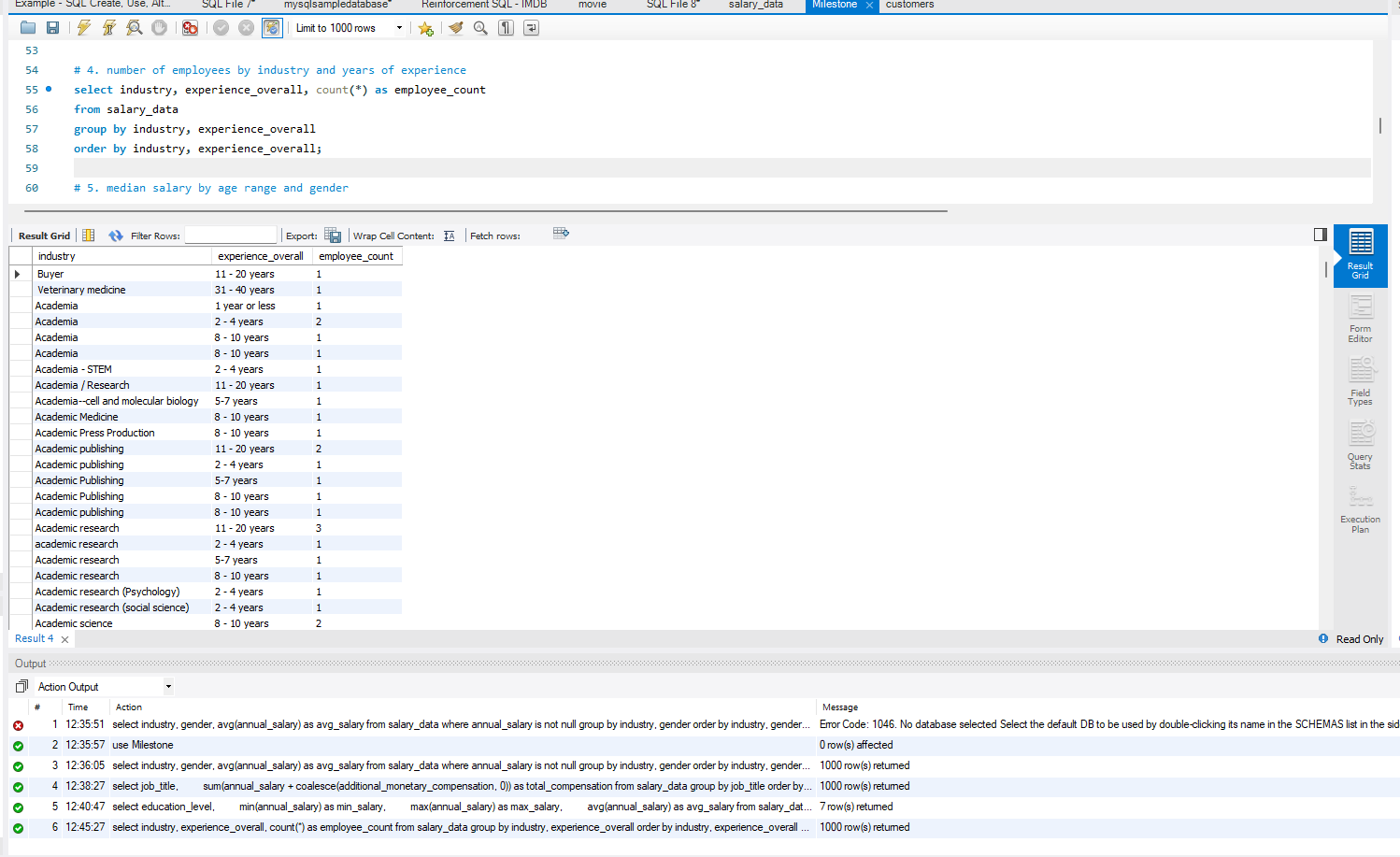
1. Number of Employees by Industry and Years of Experience

select industry, experience\_overall, count(\*) as employee\_count

from salary\_data

group by industry, experience\_overall

order by industry, experience\_overall;



1. Median Salary by Age Range and Gender

select age\_range, gender,

avg(annual\_salary) as median\_salary

from (

select age\_range, gender, annual\_salary,

row\_number() over (partition by age\_range, gender order by annual\_salary) as row\_num,

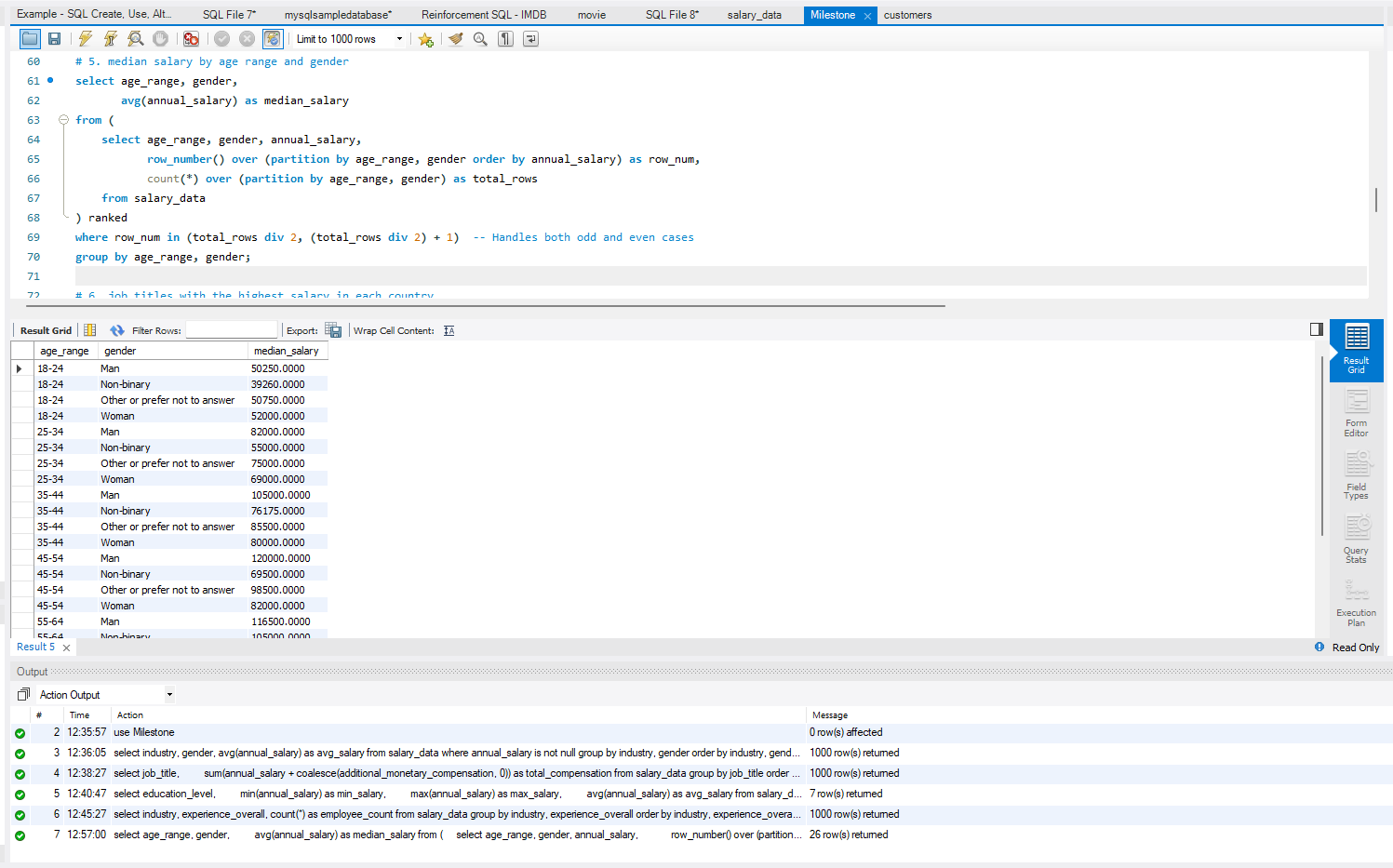
count(\*) over (partition by age\_range, gender) as total\_rows

from salary\_data

) ranked

where row\_num in (total\_rows div 2, (total\_rows div 2) + 1) -- Handles both odd and even cases

group by age\_range, gender;



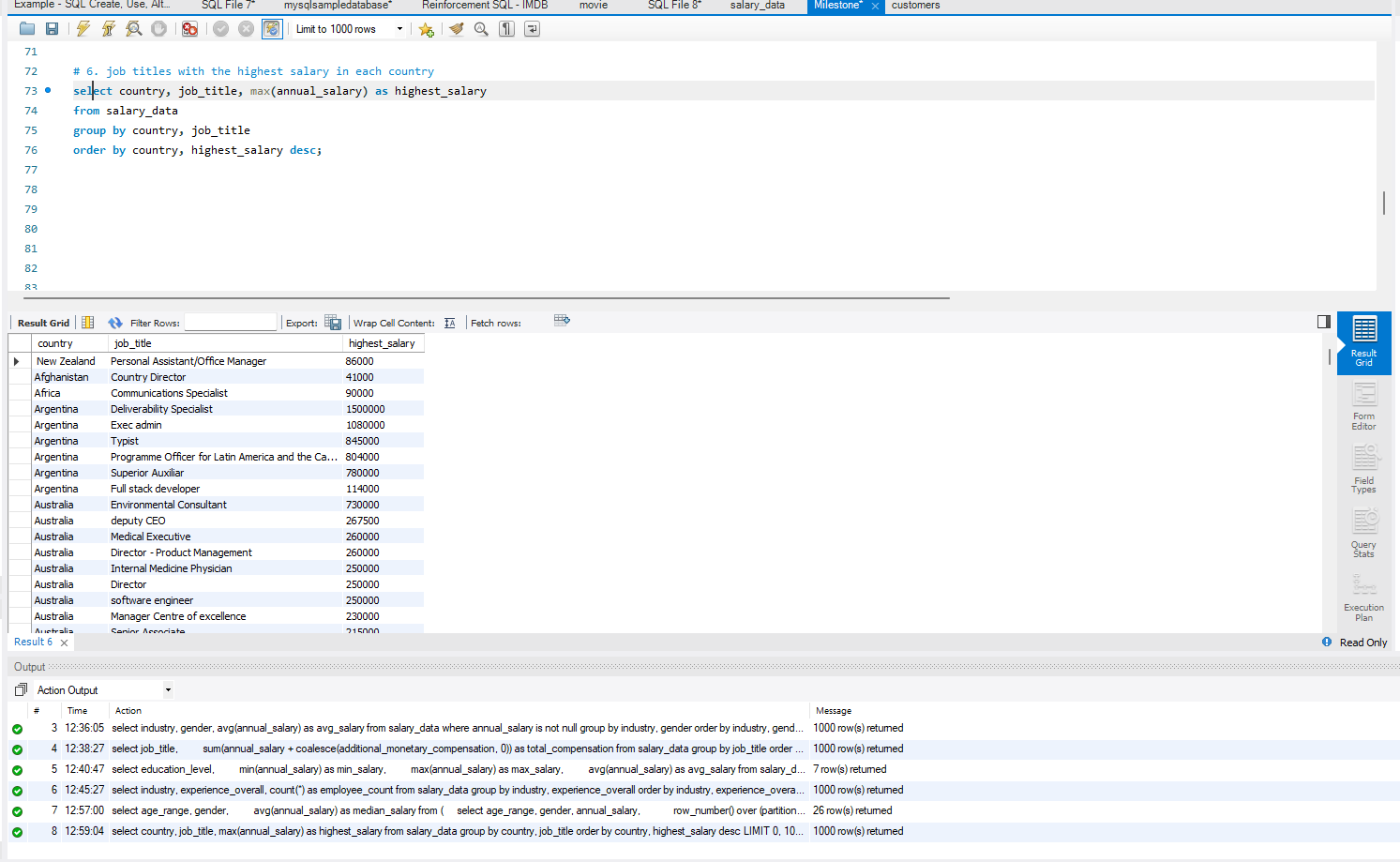
1. Job Titles with the Highest Salary in Each Country

select country, job\_title, max(annual\_salary) as highest\_salary

from salary\_data

group by country, job\_title

order by country, highest\_salary desc;



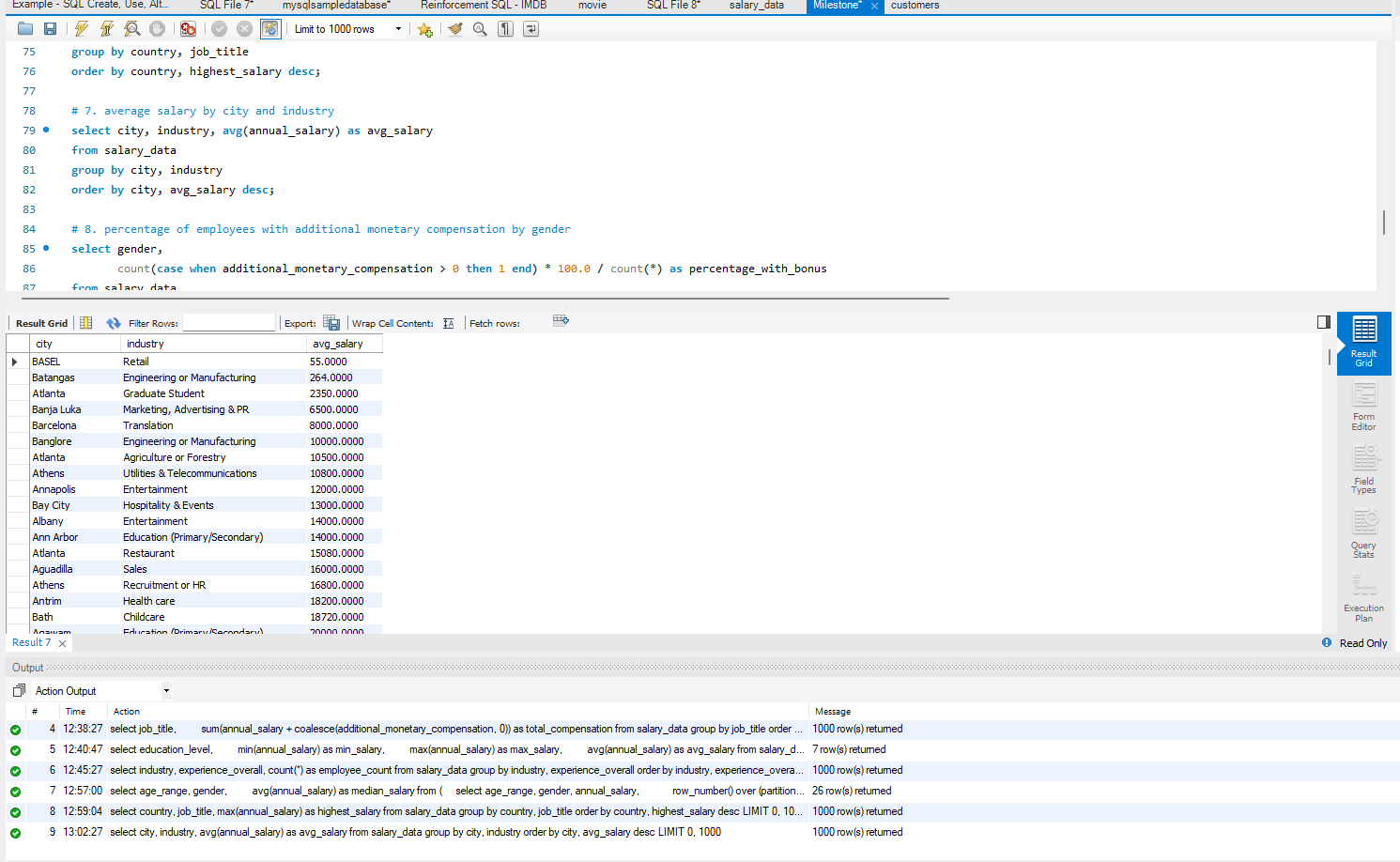
1. Average Salary by City and Industry

select city, industry, avg(annual\_salary) as avg\_salary

from salary\_data

group by city, industry

order by city, avg\_salary desc;



1. Percentage of Employees with Additional Monetary Compensation by Gender

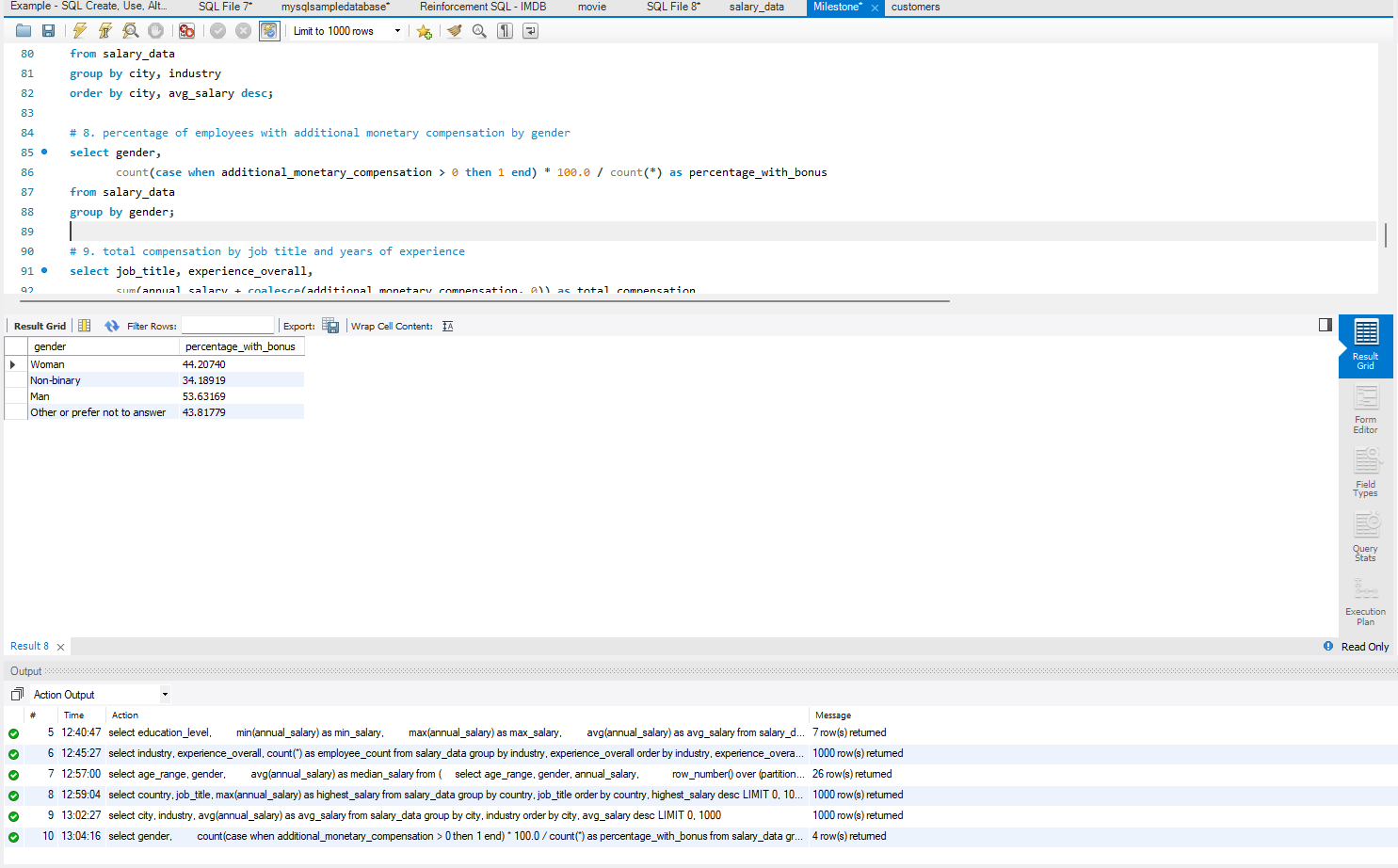
select gender,

count(case when additional\_monetary\_compensation > 0 then 1 end) \* 100.0 / count(\*) as percentage\_with\_bonus

from salary\_data

group by gender;

1. Total Compensation by Job Title and Years of Experience

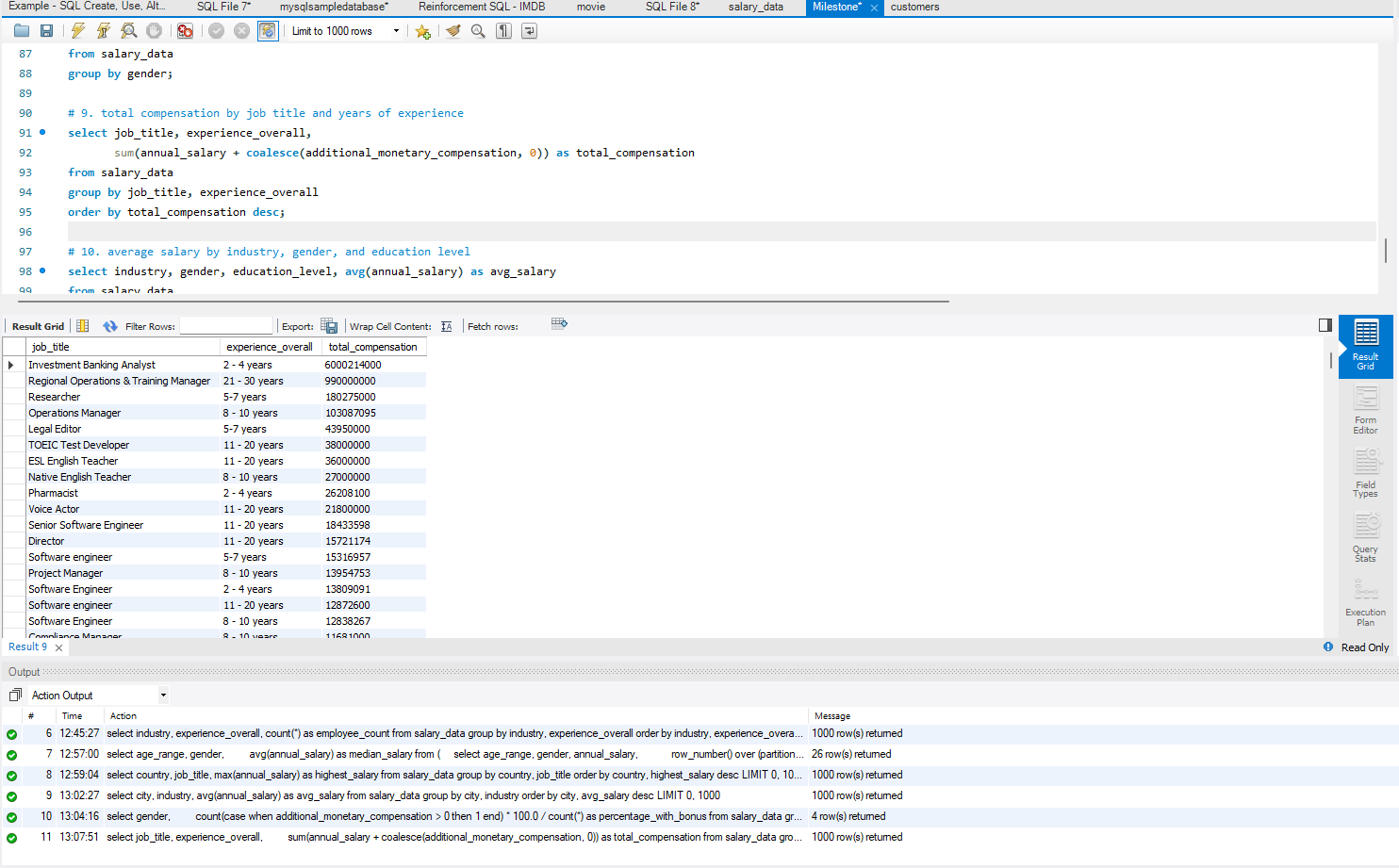
select job\_title, experience\_overall,

sum(annual\_salary + coalesce(additional\_monetary\_compensation, 0)) as total\_compensation

from salary\_data

group by job\_title, experience\_overall

order by total\_compensation desc;

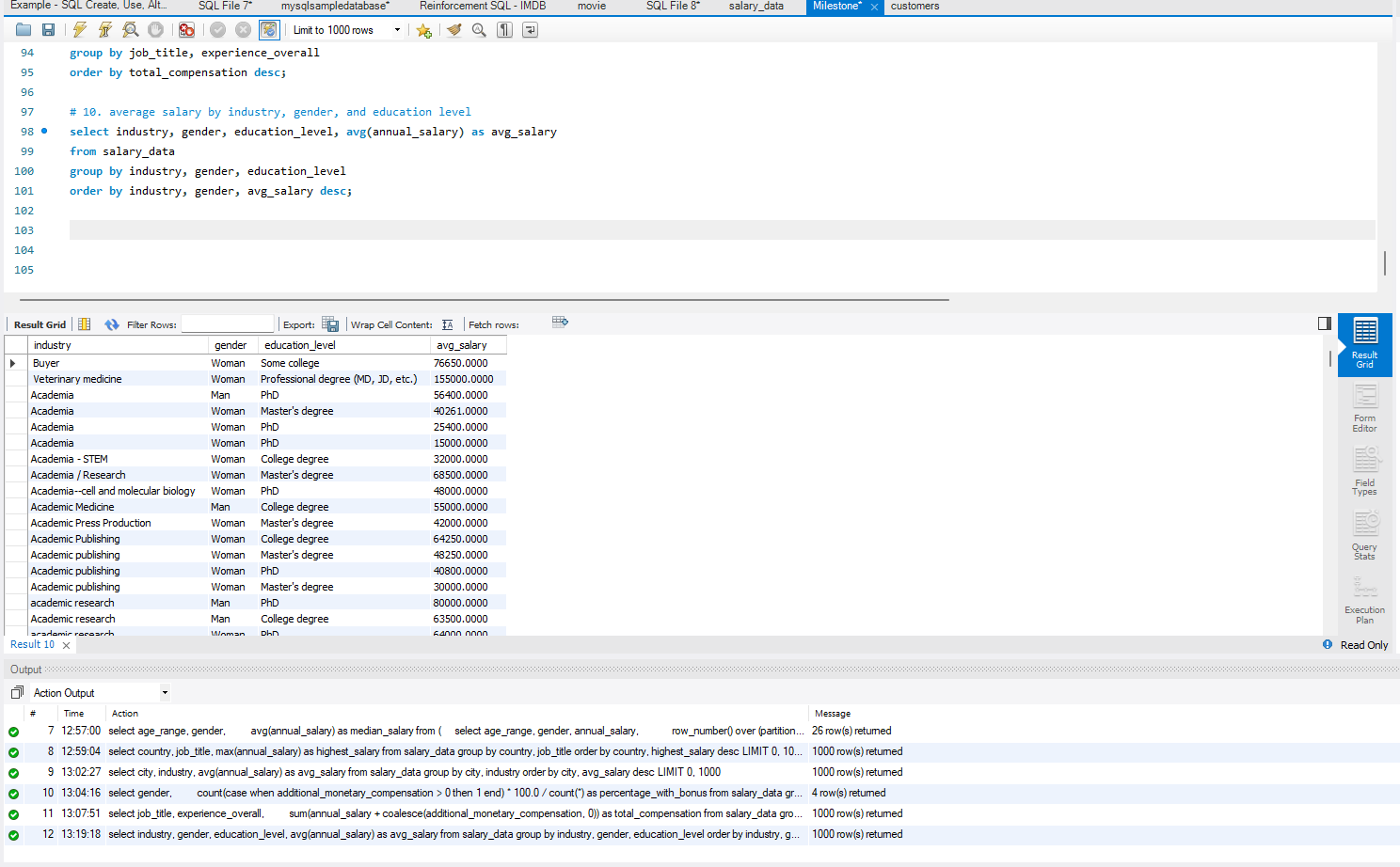


1. Average Salary by Industry, Gender, and Education Level

select industry, gender, education\_level, avg(annual\_salary) as avg\_salary

from salary\_data

group by industry, gender, education\_level

 order by industry, gender, avg\_salary desc;

**Insights:**

**Based on Age range:**

* Personnel within the age range of 25-34 has the highest workforce
* Employee count within the age range of 55-64 and above 65 is low
* 18-24 has the highest average salary
* People under 18 has the lowest average salary and the least count
* Female Employees lead all the age range with their count compared to other genders
* Average monetary compensation of employees within age range 55-64 is the highest
* Tech, Higher Education and NGOs are the industries with employees from all the age-range
* The highest and lowest age range employees count contributes 3.90% of the total workforce.

**Based on Gender:**

* More than 75% of the complete workforce comprises of female employees
* Average Annual salary of female employees is higher than any other
* People who chose not to disclose their gender and non-binary adds up-to not more than 5% of the total work group
* More than 30 Industries has employees from all the genders working
* Only 13 countries have employees from all the gender categories
* Majority of countries follows only the idea of Male/Female in the aspect of gender
* Male employees receive the highest annual average Monetary compensation plan.

**Based on Industry:**

* There are 1,089 industries total
* Tech companies, Higher education institutions and Non-profit organizations are the industries with the biggest labour force
* 9 Industries in total has more than 1000 employees whereas majority of the industries (1033) has employees in single digits
* Teaching Industry has the highest Annual average salary whereas Plumbing industry has the lowest average annual salary
* Finance, Food & Beverages and Life Sciences industry offers the most beneficial average monetary compensation plan
* Almost half of the industries do not offer monetary plans to their employees.
* Majority of the Industries are from first world countries.

**Based on Job Roles:**

* There are 12,792 Jobs overall
* Most of the employees are from the role – Software engineer
* Investment banking analyst has the highest Average annual salary
* Media executive and Game specialist earn the least annual salary compared to other job roles
* Regional Operations & Training Manager are benefited the most from their monetary compensation plan whereas a creative Co-Ordinator on the other hand are offered the least monetary compensation plan
* More than 6000 Job titles does not provide additional monetary compensation plan
* Only 7220 Employees has provided a clarification about their job roles whereas majority of the employees (almost 75%) has not provided a clarification.

Based on Education:

* Majority of the workforce – almost 50% has completed their college degree and the number of people who completed just the high school is less than 3%
* Doctorate holders earns the highest salary compared to people with lesser education level
* Less than 1% of the labour force has not disclosed or have not completed their education, most of them are in the tech field as software engineers
* Less than 5% of the workforce has completed their PhD who are currently in the govt and public administration and education industries in the role of professors
* Almost every employee under the age of 18 has completed their high school
* First world countries stand top with producing employees that have completed their basic education and high school

**Based on Experience:**

|  |  |
| --- | --- |
| Overall Experience | Field Experience |
| People with less than 1 year experience gets the lowest salary package. | People with the respective field experience gets the highest salary package. |
| Industries requiring employees with overall experience is higher. | Employees with the respective field experience are hired in less quantity compared to the overall experience. |

* People with more than 41 years’ experience in both their field and overall is very few
* Employees with the most experience are in managerial or senior positions
* Workers with experience of 1 year or less than that are in the positions of assistant
* First world countries do have more employees working with 41 years’ experience in both category

**Recommendations:**

1. Industries with the Highest Salaries:
   * Top-Paying Sectors: Energy (Oil & Gas, Renewables), Pharmaceuticals, and Consulting dominate the high-paying industries.
   * Industries with the Highest Bonuses: Pharmaceuticals, Strategy Consulting, and Petroleum provide the highest additional compensation.
   * Actionable Insight: If maximizing salary is a priority, professionals should consider careers in these industries.
2. Job Titles with the Highest Salaries:
   * Top Earning Roles: Vice Presidents, Senior Directors, and Chief-level executives (CTO, CEO) earn the most.
   * Job Titles with the Highest Bonuses: IT Directors and Senior M&A Tax Managers receive significant additional compensation.
   * Actionable Insight: Employees seeking career growth should target leadership roles and industries with structured bonus systems.
3. Experience & Salary Growth:
   * Peak Salary Period: Salaries steadily increase with experience and peak between 21-30 years of professional work.
   * Decline after 30+ Years: Salaries stabilize or slightly decrease after 30 years, likely due to retirements or career transitions.
   * Actionable Insight: Professionals should strategically upskill and negotiate salaries between 10-20 years of experience to maximize earnings.
4. Education & Salary Impact:
   * Highest Earnings: Professionals with MD, JD, or PhD degrees earn significantly higher salaries.
   * Master’s Degrees Provide a Salary Boost: Master’s degree holders earn ~$4,500 more annually than those with only a college degree.
   * Actionable Insight: Higher education (especially professional degrees) provides significant salary advantages. Employees should consider investing in relevant advanced degrees.
5. Salary vs. Bonuses Correlation:
   * Moderate Correlation (0.457): Higher salaries often come with bonuses, but this is not always the case.
   * Actionable Insight: Employees should negotiate both base salary and bonus structures, especially in industries where bonuses form a significant portion of compensation.

**Conclusion:**

Salaries peak between 21-30 years of experience, with executive roles (VP, Directors, CTOs) earning the most, especially in Energy, Pharma, and Consulting industries. Professional degrees (MD, JD) and PhDs significantly boost earnings, while bonuses are moderate but industry-dependent. Strategy Consulting and Pharma offer the highest additional compensation, making negotiation crucial. To maximize earnings, strategic career moves, industry selection, and advanced education are key factors.