

# **DATA-DRIVEN HR SOLUTIONS: TURNOVER AND PERFORMANCE**

**Prepared by:**

- 1-Hend Muhammed Muhammed**
- 2-Abdul Rahman Mohammed Fathy**
- 3-Mahmoud Gamal Hosny**
- 4-Menna Allah Amr Mohamed Hamed**
- 5-Bassant Mohamed Ahmed Elslawy**



# TABLE OF CONTENTS

1. cleaning
2. analyzing
3. dashboard
4. conclusions
5. suggestions
6. tools



# INTRO

**This project presents modifications to employee and performance data for analysis, focusing on turnover causes and performance trends. The data cleaning phase standardized values and addressed missing data. Key factors contributing to turnover, such as overtime and years of service, were identified, along with elements affecting job satisfaction and work-life balance. Dashboard visualizations highlighted insights on training, diversity, and fairness within the company.**

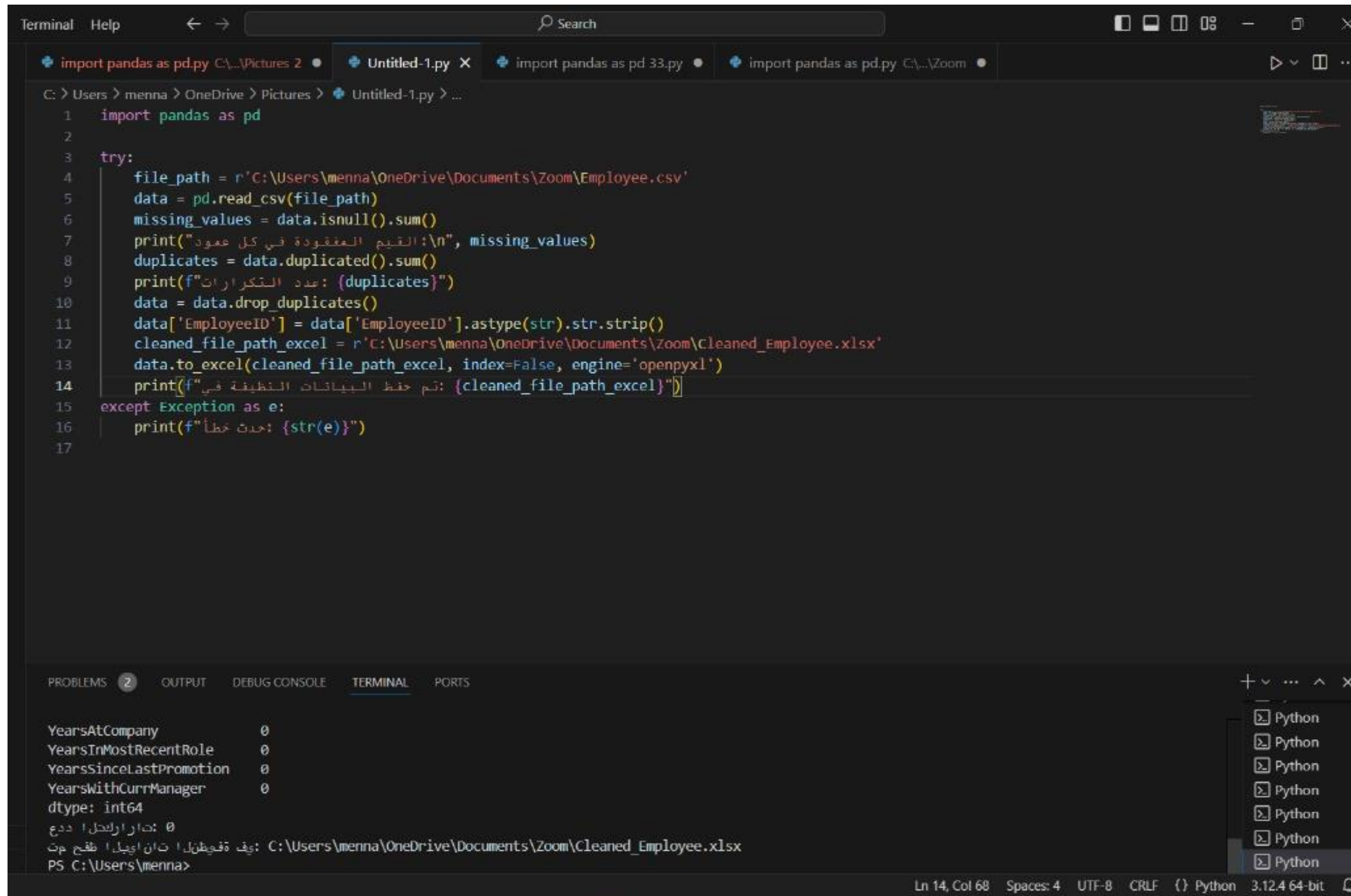
# TOOLS



SQL



# CLEANING



```
Terminal Help  Search
import pandas as pd.py C:\...Pictures 2 •  Untitled-1.py X  import pandas as pd 33.py •  import pandas as pd.py C:\...Zoom •

C: > Users > menna > OneDrive > Pictures >  Untitled-1.py > ...
1  import pandas as pd
2
3  try:
4      file_path = r'C:\Users\menna\OneDrive\Documents\Zoom\Employee.csv'
5      data = pd.read_csv(file_path)
6      missing_values = data.isnull().sum()
7      print("القيم المفقودة في كل عمود:\n", missing_values)
8      duplicates = data.duplicated().sum()
9      print(f"عدد التكرارات: {duplicates}")
10     data = data.drop_duplicates()
11     data['EmployeeID'] = data['EmployeeID'].astype(str).str.strip()
12     cleaned_file_path_excel = r'C:\Users\menna\OneDrive\Documents\Zoom\Cleaned_Employee.xlsx'
13     data.to_excel(cleaned_file_path_excel, index=False, engine='openpyxl')
14     print(f"تم حفظ البيانات النظيفة في: {cleaned_file_path_excel}")
15 except Exception as e:
16     print(f"حدث خطأ: {str(e)}")
17

PROBLEMS 2  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS
YearsAtCompany      0
YearsInMostRecentRole  0
YearsSinceLastPromotion  0
YearsWithCurrManager  0
dtype: int64
0: إدارة المحتل ا ددع
يف قفوطنلا تان اويل ا طفح م
PS C:\Users\menna>

Ln 14, Col 68  Spaces: 4  UTF-8  CRLF  {} Python  3.12.4 64-bit
```

## Check for Missing Values and Duplicates:

- A thorough review will be conducted to identify any missing values within the dataset. Additionally, duplicate entries will be removed to ensure data integrity and accuracy.



# CLEANING

The image shows a Windows IDE with a dark theme. The top bar includes 'Terminal', 'Help', and a search bar. The main editor area displays a Python script in a file named 'Untitled-1.py'. The script imports pandas, reads a CSV file, drops specific columns, and saves the cleaned data to an Excel file. The bottom panel shows the 'TERMINAL' output, which includes the execution of the script and the resulting file paths. A context menu is visible in the bottom right corner, showing options for Python.

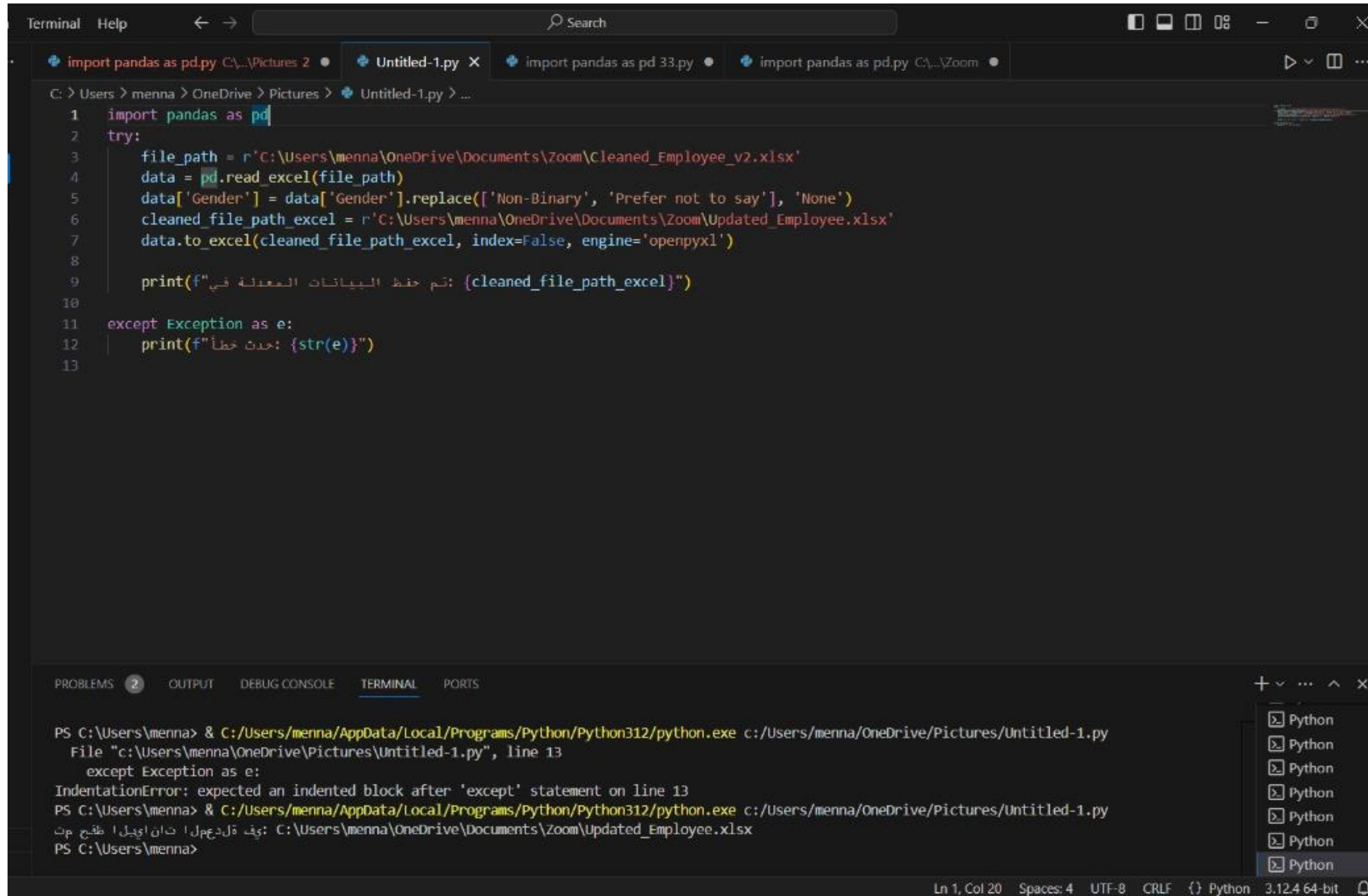
```
import pandas as pd
file_path = r'C:\Users\menna\OneDrive\Documents\Zoom\Employee.csv'
data = pd.read_csv(file_path)
columns_to_drop = [
    'FirstName',
    'LastName',
    'State',
    'HireDate',
    'BusinessTravel',
    'YearsInMostRecentRole',
    'StockOptionLevel',
    'MaritalStatus'
]
data_cleaned = data.drop(columns=columns_to_drop)
cleaned_file_path_excel = r'C:\Users\menna\OneDrive\Documents\Zoom\Cleaned_Employee_v2.xlsx'
data_cleaned.to_excel(cleaned_file_path_excel, index=False, engine='openpyxl')
print(f"Cleand data saved at: {cleaned_file_path_excel}")
except Exception as e:
    print(f"Error during processing: {e}")
```

YearsSinceLastPromotion 0  
YearsWithCurrManager 0  
dtype: int64  
تدار لفل 0  
وقف قفيلنل تان ايل طفح م  
PS C:\Users\menna> & C:/Users/menna/AppData/Local/Programs/Python/Python312/python.exe c:/Users/menna/OneDrive/Pictures/Untitled-1.py  
Cleaned data saved at: C:\Users\menna\OneDrive\Documents\Zoom\Cleaned\_Employee\_v2.xlsx  
PS C:\Users\menna>

## Remove Unused Columns:

- **Unnecessary columns that are not relevant to our analysis will be eliminated. Each column name will be listed to specify which ones will be discarded, streamlining the dataset for further processing.**

# CLEANING



```
Terminal Help Search
import pandas as pd.py C:\Users\menna\OneDrive\Pictures 2 • Untitled-1.py X import pandas as pd 33.py • import pandas as pd.py C:\Users\Zoom •
C: > Users > menna > OneDrive > Pictures > Untitled-1.py > ...
1 import pandas as pd
2 try:
3     file_path = r'C:\Users\menna\OneDrive\Documents\Zoom\Cleaned_Employee_v2.xlsx'
4     data = pd.read_excel(file_path)
5     data['Gender'] = data['Gender'].replace(['Non-Binary', 'Prefer not to say'], 'None')
6     cleaned_file_path_excel = r'C:\Users\menna\OneDrive\Documents\Zoom\Updated_Employee.xlsx'
7     data.to_excel(cleaned_file_path_excel, index=False, engine='openpyxl')
8
9     print(f"تم حفظ البيانات المعدلة في: {cleaned_file_path_excel}")
10
11 except Exception as e:
12     print(f"حدث خطأ: {str(e)}")
13

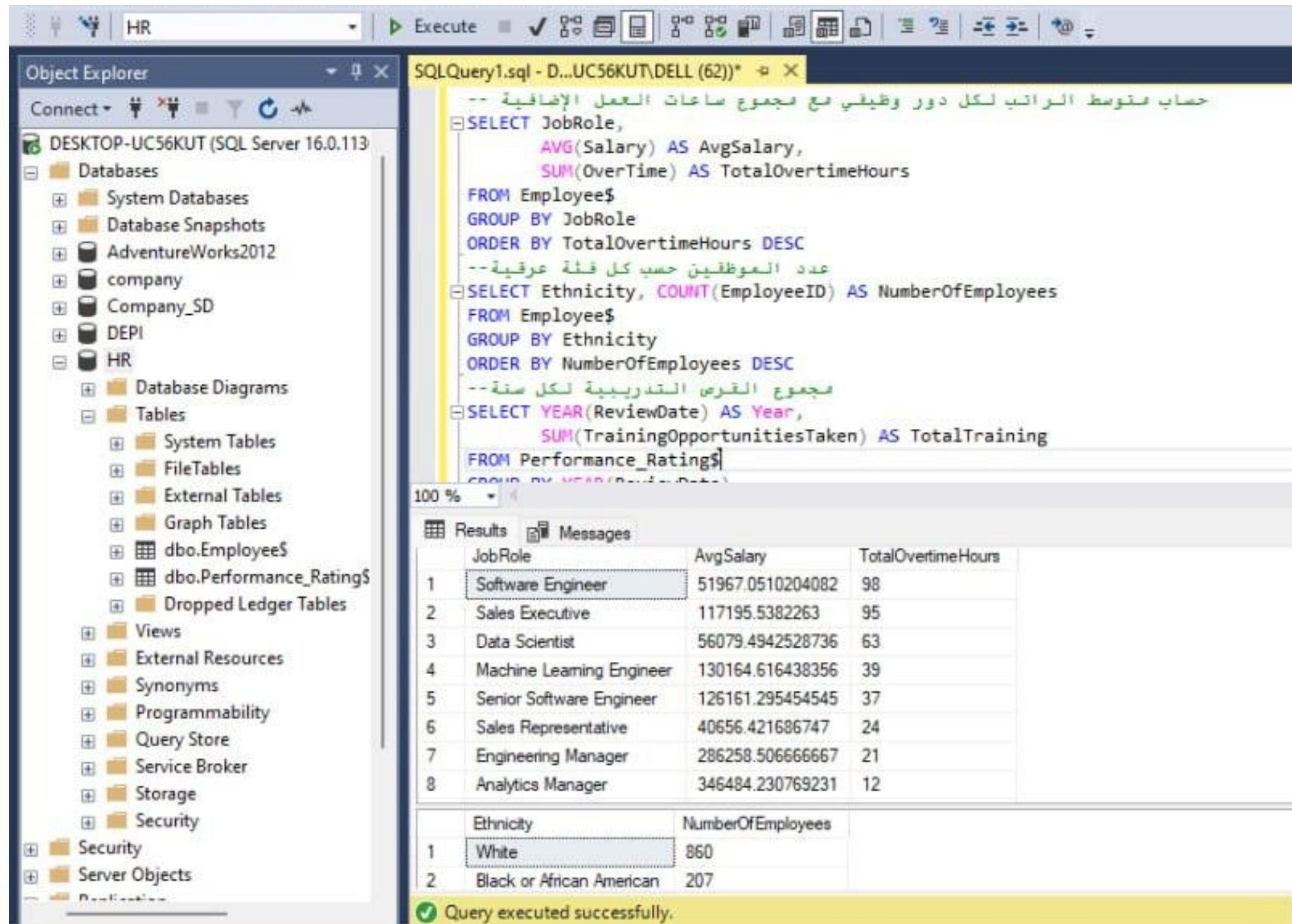
PROBLEMS 2 OUTPUT DEBUG CONSOLE TERMINAL PORTS
PS C:\Users\menna> & C:/Users/menna/AppData/Local/Programs/Python/Python312/python.exe c:/Users/menna/OneDrive/Pictures/Untitled-1.py
File "c:\Users\menna\OneDrive\Pictures\Untitled-1.py", line 13
    except Exception as e:
IndentationError: expected an indented block after 'except' statement on line 13
PS C:\Users\menna> & C:/Users/menna/AppData/Local/Programs/Python/Python312/python.exe c:/Users/menna/OneDrive/Pictures/Untitled-1.py
يف قد يعمل! ث ان ايدل! طبع م
PS C:\Users\menna>
Ln 1, Col 20 Spaces: 4 UTF-8 CRLF {} Python 3.12.4 64-bit
```

## Transform Non-Binary Values:

- To ensure inclusivity, all cells containing "non-binary" or "not preferred to say" will be replaced with "NULL" This standardization helps maintain consistency in the dataset.



# ANALIZING



The screenshot displays the SQL Server Enterprise Manager interface. On the left, the Object Explorer shows the database structure for 'DESKTOP-UC56KUT (SQL Server 16.0.113)', including 'Databases', 'System Databases', 'Database Snapshots', 'AdventureWorks2012', 'company', 'Company\_SD', 'DEPI', and 'HR'. The 'HR' database is expanded, showing 'Database Diagrams', 'Tables', 'Views', 'External Resources', 'Synonyms', 'Programmability', 'Query Store', 'Service Broker', 'Storage', and 'Security'. The 'Tables' folder is expanded, showing 'dbo.Employee\$' and 'dbo.Performance\_Rating\$'. The 'Results' pane shows the output of a query executed successfully. The query is a multi-part SQL statement with three SELECT clauses. The first clause calculates average salary and total overtime hours by job role. The second clause calculates the number of employees by ethnicity. The third clause calculates the total training opportunities taken by year. The results are displayed in two tables. The first table shows job roles and their corresponding average salaries and total overtime hours. The second table shows the number of employees by ethnicity.

```
-- حساب متوسط الراتب لكل دور وظيفي مع مجموع ساعات العمل الإضافية
SELECT JobRole,
       AVG(Salary) AS AvgSalary,
       SUM(OverTime) AS TotalOvertimeHours
FROM Employee$
GROUP BY JobRole
ORDER BY TotalOvertimeHours DESC

-- عدد الموظفين حسب كل فئة عرقية
SELECT Ethnicity, COUNT(EmployeeID) AS NumberOfEmployees
FROM Employee$
GROUP BY Ethnicity
ORDER BY NumberOfEmployees DESC

-- مجموع القرض التدريبية لكل سنة
SELECT YEAR(ReviewDate) AS Year,
       SUM(TrainingOpportunitiesTaken) AS TotalTraining
FROM Performance_Rating$
GROUP BY YEAR(ReviewDate)
```

JobRole	AvgSalary	TotalOvertimeHours
1 Software Engineer	51967.0510204082	98
2 Sales Executive	117195.5382263	95
3 Data Scientist	56079.4942528736	63
4 Machine Learning Engineer	130164.616438356	39
5 Senior Software Engineer	126161.295454545	37
6 Sales Representative	40656.421686747	24
7 Engineering Manager	286258.506666667	21
8 Analytics Manager	346484.230769231	12

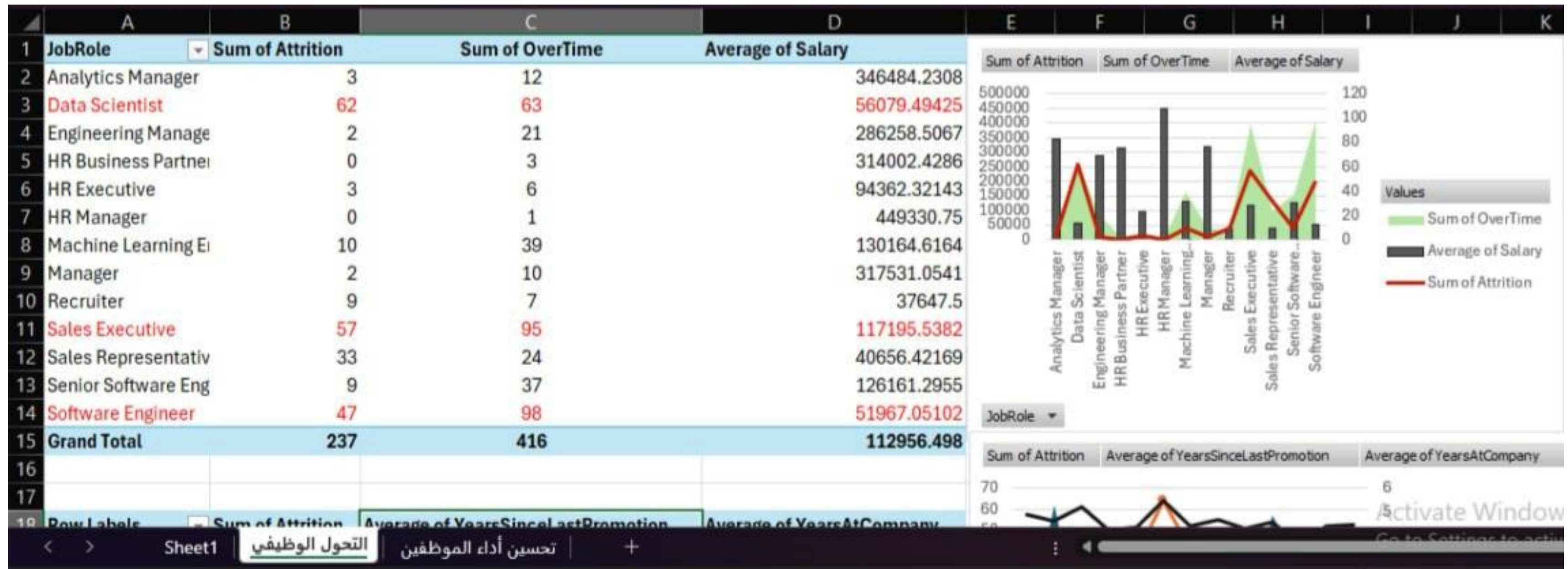
Ethnicity	NumberOfEmployees
1 White	860
2 Black or African American	207

Query executed successfully.

A database was created using an SQL server to analyze relationships such as number of employees for each racial group to look for discrimination in employment opportunities or racism, also the relation between "average salaries" and "total overtime hours" for each job, as shown.

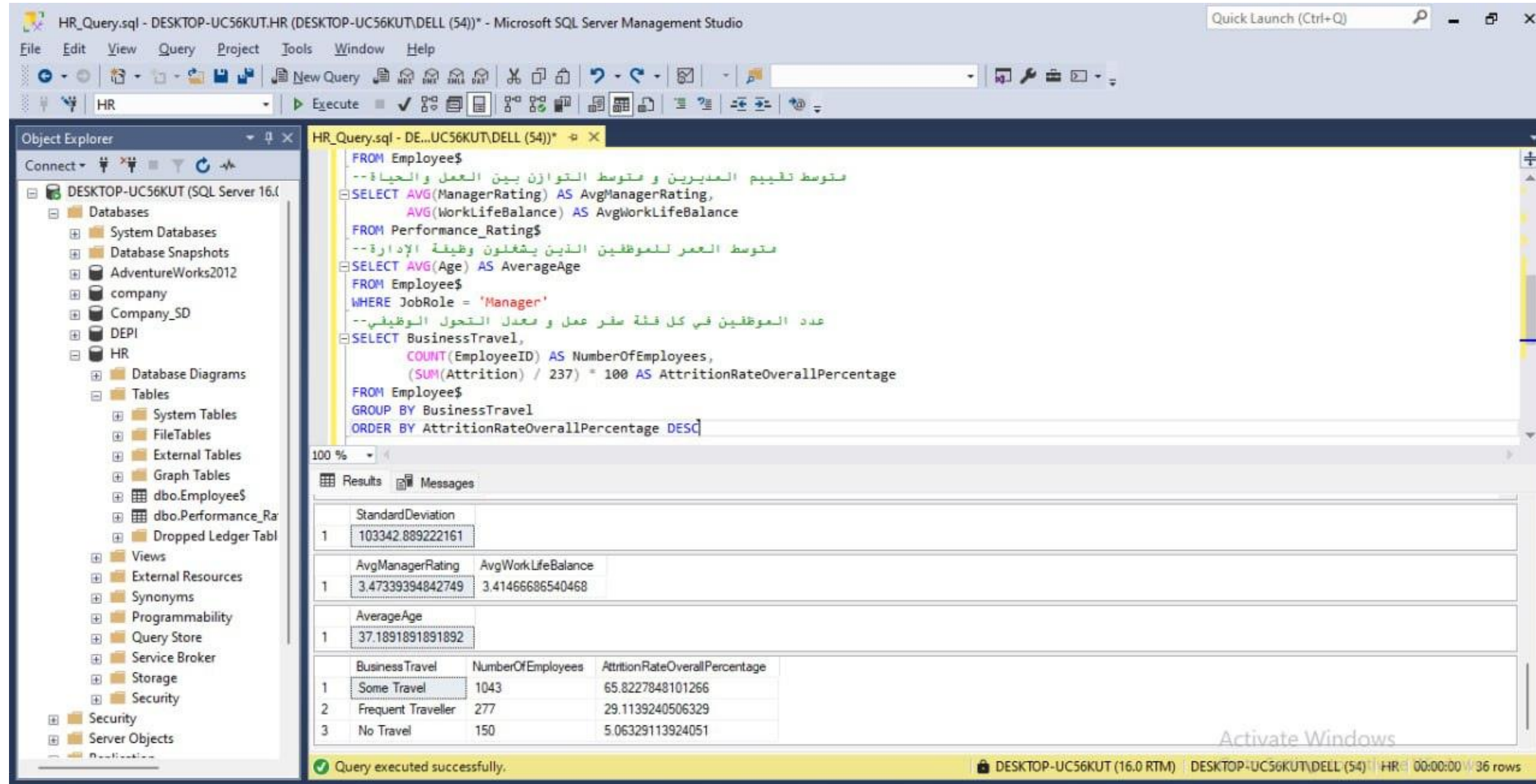


# ANALIZING



The same data was visualized in Excel, incorporating other factors analysis to explore the effect on Attrition, illustrated in the Figure.

# ANALIZING



The average age of managers was calculated to assess the fairness of hiring criteria, as depicted in the Figure



# ANALYZING



while gender inequality in salaries and job opportunities was assessed, as shown in the Figure

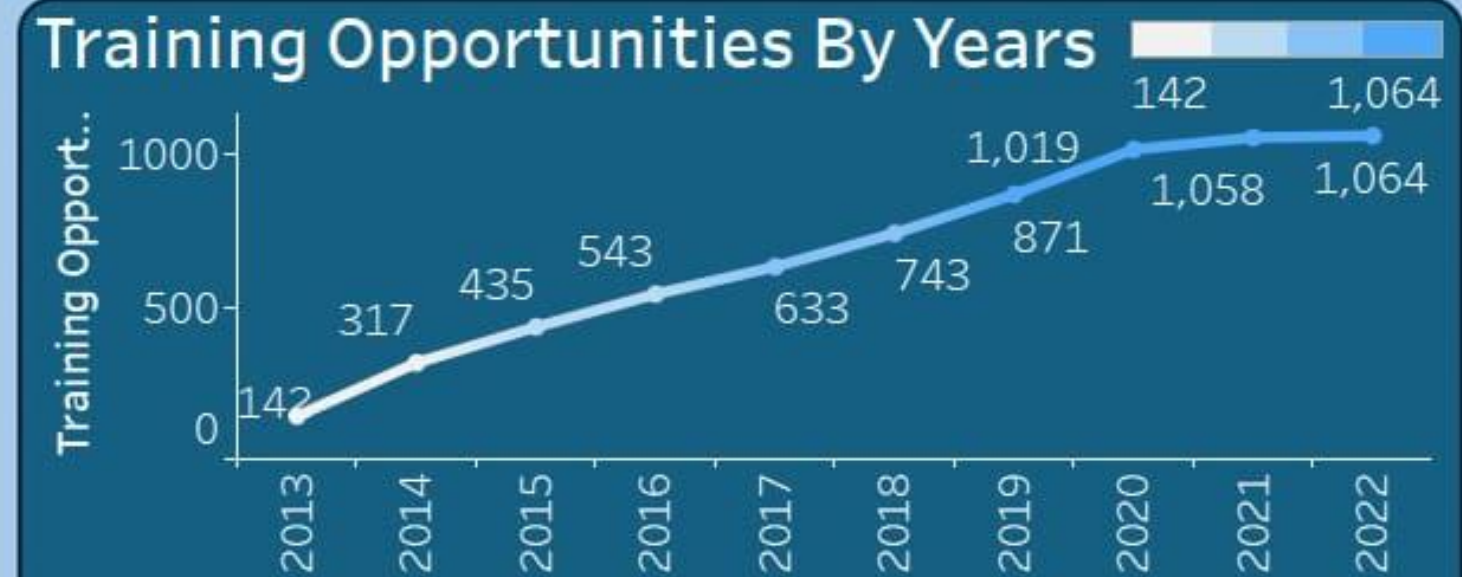
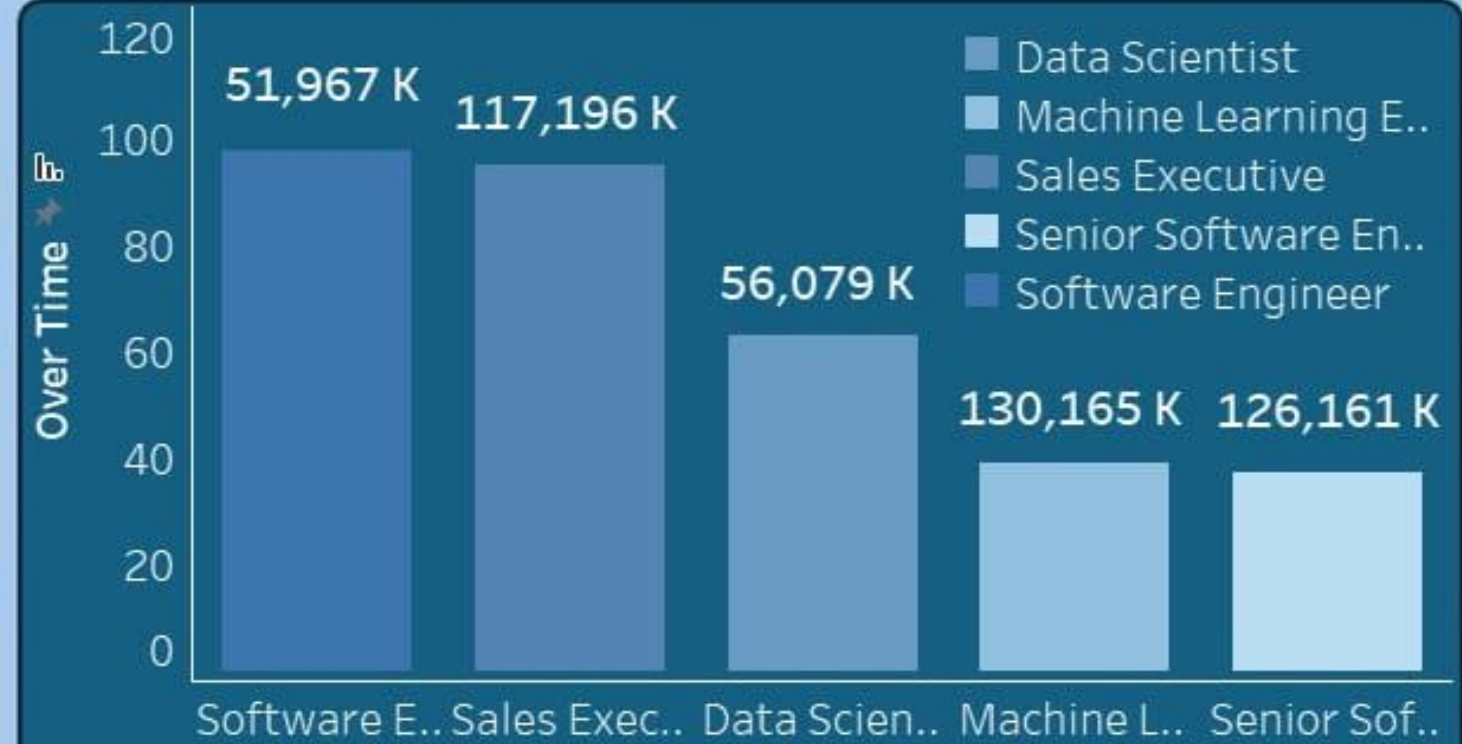
# DASHBOARD

## Work Life Balance & Training Insights Dashboard

Total Employee  
1,470

Sum Of Attrition  
237

### Relationship Between Manager Rating And Work Life Balance

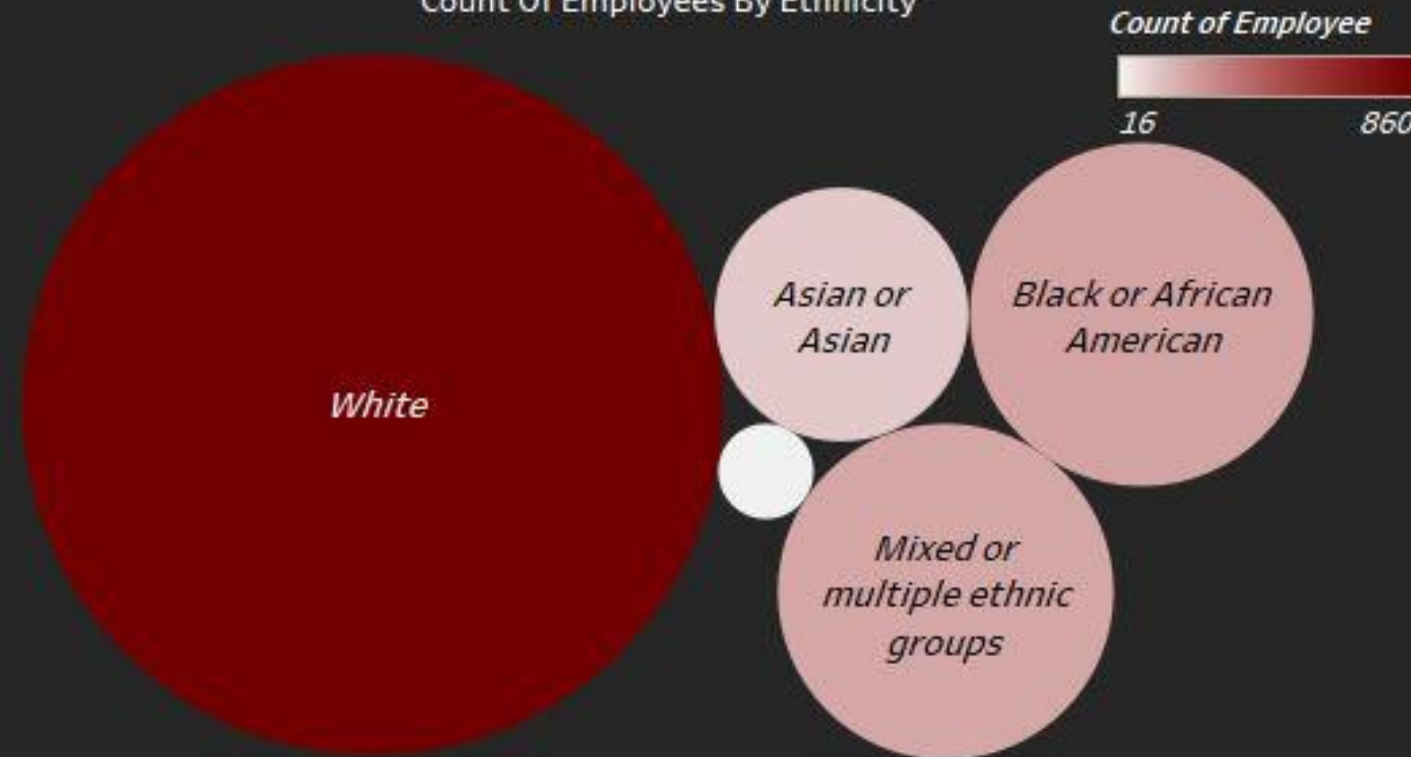




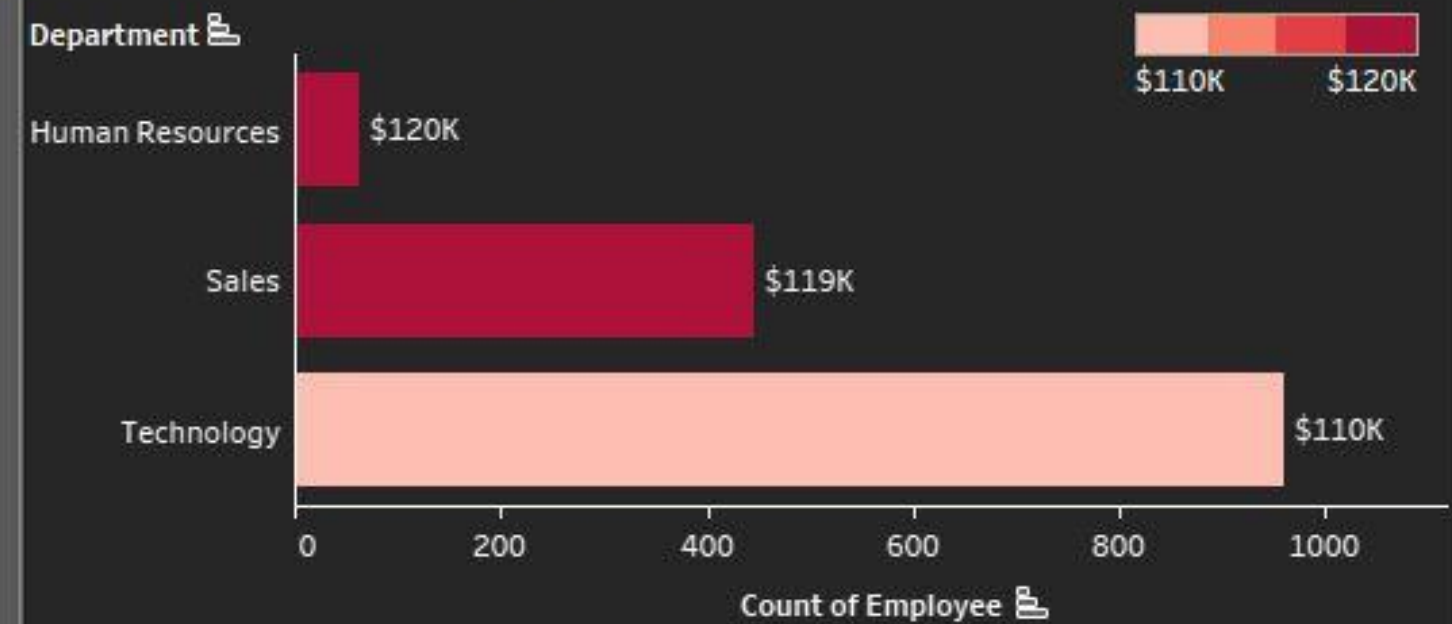
# DASHBOARD

## Workplace Diversity & Salary Fairness Dashboard

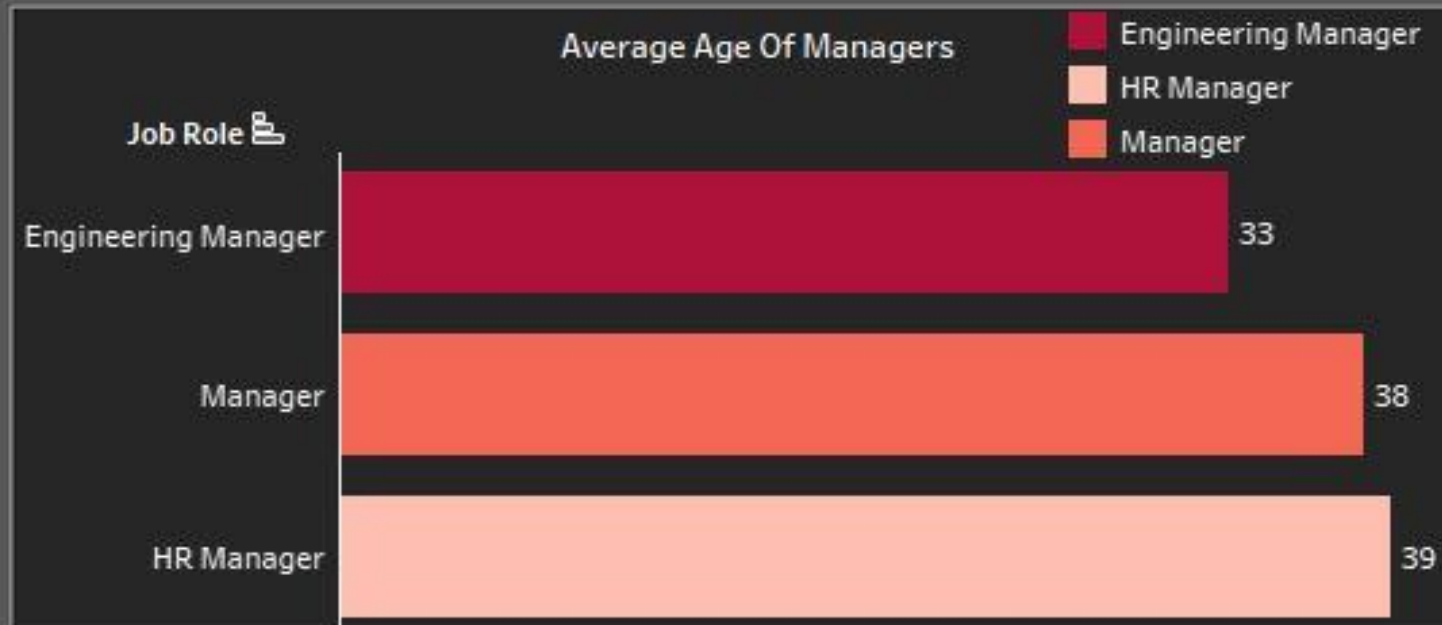
Count Of Employees By Ethnicity



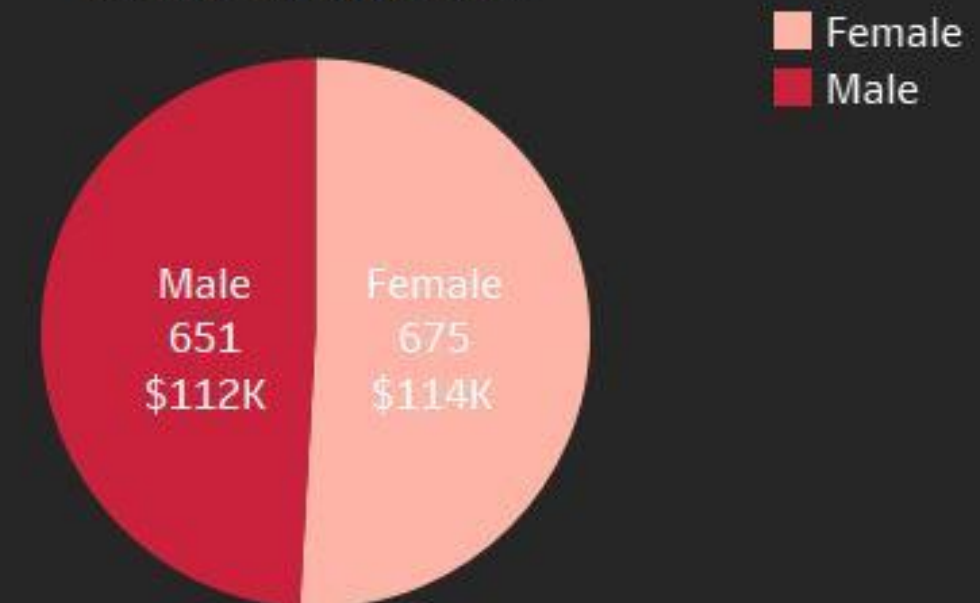
Average Salary Between Departments



Average Age Of Managers

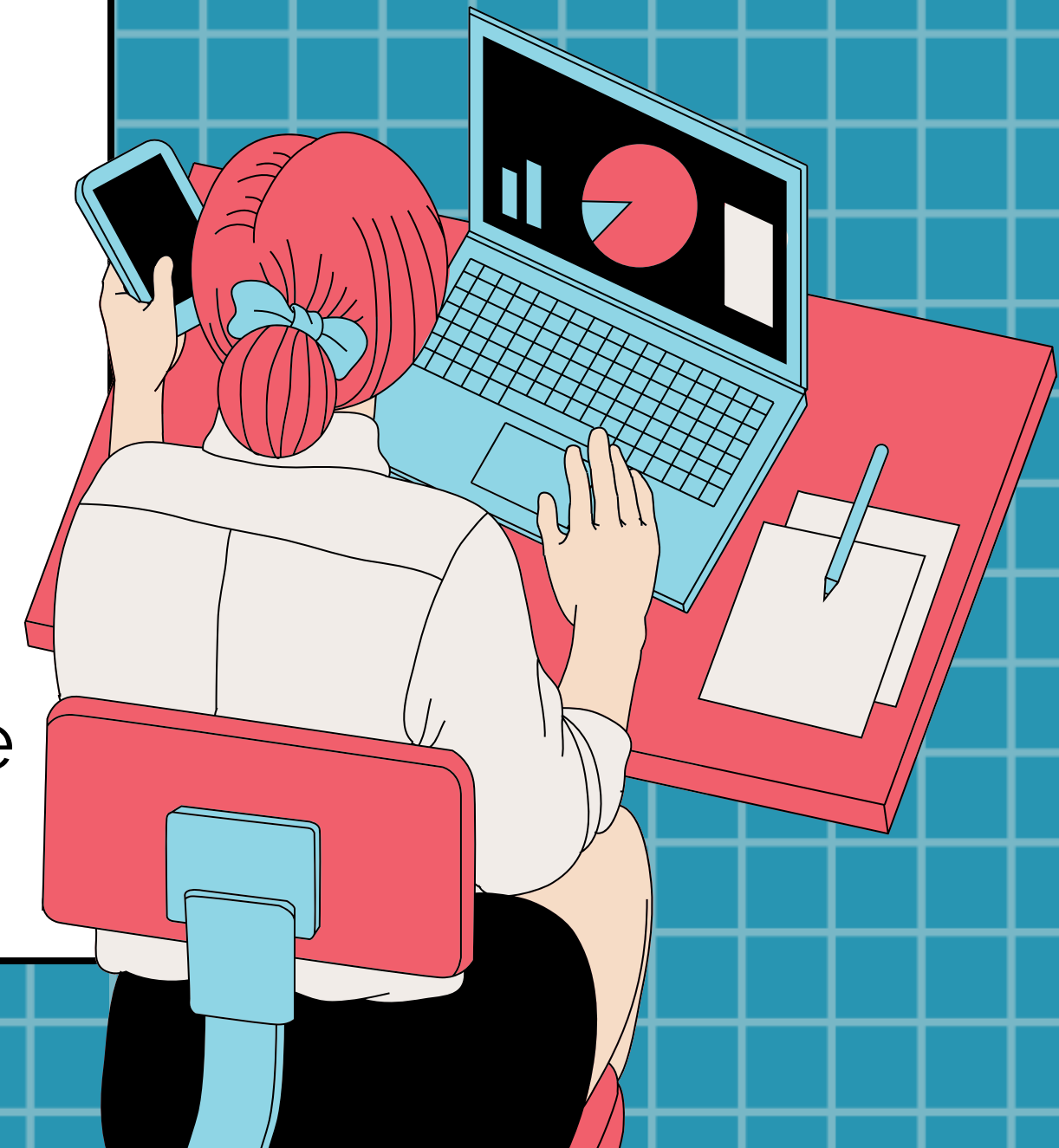


Average Salary By Gender



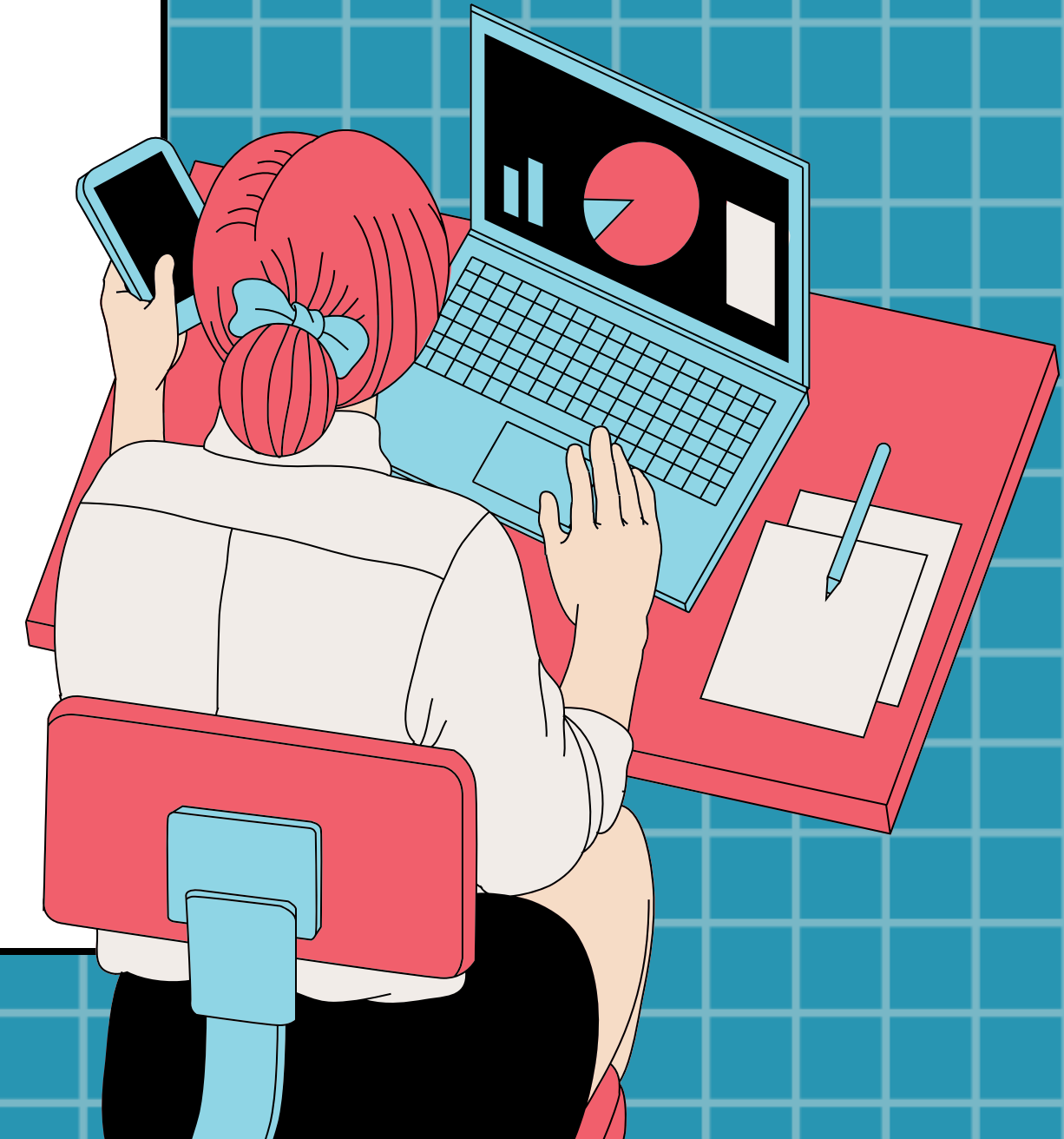
# CONCLUSIONS

- Jobs requiring more overtime tend to offer lower average salaries.
- Better work-life balance improves employee productivity and manager evaluations.
- The company has increasingly focused on providing regular training opportunities.
- The company shows a hiring preference for white employees.



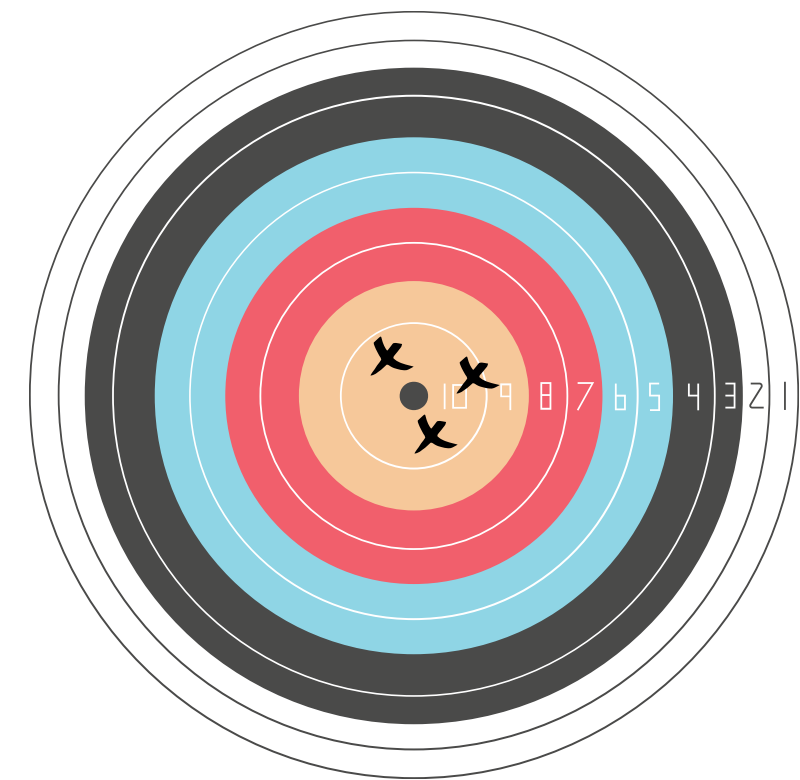
# CONCLUSIONS

- No significant gender differences in job opportunities or salaries.
- Desk work combined with business travel influences career shifts.
- Younger managers are selected based on competence rather than years of service.



# SUGGESTIONS

- Offer incentives for overtime work and provide targeted training for skill gaps.
- Increase flexible work policies and allow remote work when suitable.
- Align hiring and training departments to focus on relevant skill development.
- Implement anti-racism awareness programs and monitor recruitment practices.





**THANK YOU!**

