

Data Visualization Project

Shara Vanessa Pineda

Southern Alberta Institute of Technology

Data Science - DATA-043-003

Table of Contents

1. The Dataset.....	3
1.1. Overview.....	3
1.2. Human Resource Dataset.....	3
2. Purpose.....	5
2.1. Overview.....	5
2.2. KPIs.....	5
3. Data Visualization.....	5
3.1. Overview.....	5
3.2. Transformed Data.....	6
3.3. Data Model.....	6
3.4. Report – Executive Dashboard.....	8
3.5. Report – Department Dashboard.....	10
4. Insights.....	13
5. Summary	14

1. The Dataset

1.1. Overview

The dataset selected for this project comes from Kaggle and is titled the "Human Resources Data Set" authored by Dr. Carla Patalano and Dr. Rich Huebner which simulates employee data for a fictional U.S. company. It was chosen for its cleanliness, granularity, and comprehensiveness, offering a rich repository of information for analysis. It's clean and well maintained and its wealth of information allows numerous insights that can be tailored to specific purposes.

As a Human Resources dataset, it holds universal applicability across industries and organizational settings. Its relatability to the any business ensures its relevance to a wide audience, regardless of company size or industry. Thus, it serves as a valuable resource for professionals seeking to gather insights into workforce dynamics and organizational management.

1.2. Human Resource Dataset

The original dataset comprises 36 columns, encompassing various aspects of employee data, including personal details, significant dates, performance ratings, and recruitment particulars. While the dataset is extensive, it underwent further cleaning and transformation specifically for the purposes of this article. Certain columns were removed, while others deemed essential were added to tailor it for the project's objectives.

To delve deeper into the dataset's features, a summary of descriptions is provided below, primarily sourced from https://rpubs.com/rhuebner/hrd_cb_v14. It's important to note that the columns listed below reflect the updated dataset that was utilized for this project, not the original dataset downloaded.

Column/ Feature	Description	Data Type
EmployeeID	Employee code is unique to each employee	Integer
EmployeeName	Employee's full name	Text
Salary	Employee's yearly salary in \$ U.S. Dollars	Integer
DateOfBirth	Employee Date of Birth	Date

Age	Calculated column based on Date of Birth and assumed present date, which is January 01, 2019	Integer
DateofHire	Date that employee was hired	Date
DateofTermination	Date that employee was terminated, if applicable	Date
YearsOf Service	Calculated column based on Date of Hire, Date of Termination and assumed present date, which is January 01, 2019	Integer
SatisfactionRate	A basic satisfaction score between 1 and 5, as reported on a recent employee satisfaction survey	Integer
GenderID	Gender code that matches the text field Gender	Integer
Gender	Gender - Male or Female	Text
RaceID	Race code that matches the text field Race	Integer
Race	Description of the race the employee identifies with - American Indian or Alaska Native, Asian, Black or African American, Hispanic, Two or more races or White	Text
EmploymentStatusID	Employment status code that matches text field Employment Status	Integer
EmploymentStatus	A description/category of the employee's employment status – Active, Terminated for Cause or Voluntarily Terminated	Text
DeptID	Department code that matches text field Department	Integer
Department	Name of the department that the employee works in	Text

2. Purpose

2.1. Overview

While the dataset contains numerous variables for analysis, the primary purpose of this article is to analyze the diversity profile of the company and provide insights for potential improvement. Specifically, attention will be directed towards examining three key dimensions of diversity: gender, age, and race. By scrutinizing these facets, the goal is to gain a comprehensive understanding of the current diversity landscape within the organization.

2.2. KPIs

The following KPIs will serve as a guide on how to present the data in terms of visualization and analysis within the HR dataset. Gender diversity, age diversity, racial (ethnicity) diversity, and salary equity will be the focal points, providing valuable insights into the workforce composition and equity practices within the organization.

- a. **Gender Diversity:** This KPI measures the distribution of genders within the company's workforce. It assesses the representation of male and female across different pay range and departments.
- b. **Age Diversity:** This KPI measures the distribution of employees across various age groups within the organization. It examines the proportion of younger, mid-career, and older employees, highlighting potential age-related biases or preferences in pay and career development processes.
- c. **Racial (Ethnicity) Diversity:** This KPI measures the representation of different racial and ethnic groups among the workforce. It examines whether the company's employees reflect the demographic diversity of the broader population and aims to identify any disparities or underrepresentation of certain racial or ethnic groups.
- d. **Salary Equity:** This KPI measures the fairness and parity in compensation across different demographic groups, regardless of gender, age, or race. This KPI evaluates whether employees receive equal pay for equal work and aims to identify and address any gender, age, or race-based pay gaps within the organization.

3. Data Visualization

3.1. Overview

The data visualization component of this project serves as a critical tool for presenting complex data in a clear and understandable manner. By leveraging the power of visualization, the goal is to make the dataset more accessible, allowing for easier

interpretation and analysis. The whole data visualization process will be composed of three parts: transformation of data, preparing the data model, and creating the actual data visualization report, which will consist of the executive dashboard and department dashboard.

For this article, **PowerBI** has been chosen as the primary visualization tool. PowerBI is a popular business intelligence platform that enables users to create interactive reports and dashboards, visualize data with a wide range of customizable charts and graphs, and share insights across teams seamlessly.

3.2. Transformed Data

The HR dataset underwent a transformation process to align with the objectives of the project. These transformations aimed to update the dataset and enhance its suitability for the analysis and visualization. Utilizing Power Query of PowerBI, several adjustments were made to enhance the dataset's relevance and usability including:

- a. **Multiple tables:** The dataset was evolved from a flat structure into a star schema to facilitate improved analysis and visualization. Data Model be discussed further in the subsequent section.
- b. **Years of Service:** Two different calculation methods were applied based on employment status. For employees with an Active status, the years of service were computed by finding the difference between the assumed present date (January 01, 2019) and the Date of Hire. For employees with a Terminated for Cause and Voluntarily Terminated status, the years of service were computed by finding the difference between the Date of Termination and the Date of Hire.
- c. **Age Bins:** Age was grouped into bins with a bin size set to 10 years.
- d. **Salary Bins:** Salary was grouped into bins with a bin size set to \$15,000.
- e. **Count Female:** A measure was created to count the number of female employees in the dataset.
- f. **Count Male:** A measure was created to count the number of male employees in the dataset.

3.3. Data Model

For this project, the data model is normalized by having multiple tables as opposed to having a flat structure, adhering to industry best practices to optimize performance and scalability. The following best practices are employed in the data model:

- a. **Star schema:** Utilized to organize data into fact tables surrounded by dimension tables, resembling a star shape when visualized. This approach promotes data integrity, simplifies query complexity, and enhances analytical capabilities by providing clear relationships between entities.
- b. **One-to-many relationships:** This relationship type is essential for accurately representing real-world connections within the dataset, ensuring comprehensive analysis, and reporting capabilities.
- c. **Data flow from Fact table to Dimension tables:** Implemented to streamline data retrieval and aggregation. This design enables efficient access to summarized data while maintaining granularity and detail, empowering users to derive meaningful insights from the dataset.

In the project's data model, the dataset is organized into a total of 9 tables, with the Fact Table representing the Employees, serving as the central focal point from which all other tables are connected. The Dimension Tables encompass various entities linked to each data entry, including Gender, Marital Status, State, Race, Employment Status, Department, Position, and Recruitment. This design embodies a star schema, showing clear relationships between entities, as illustrated in Figure 1.

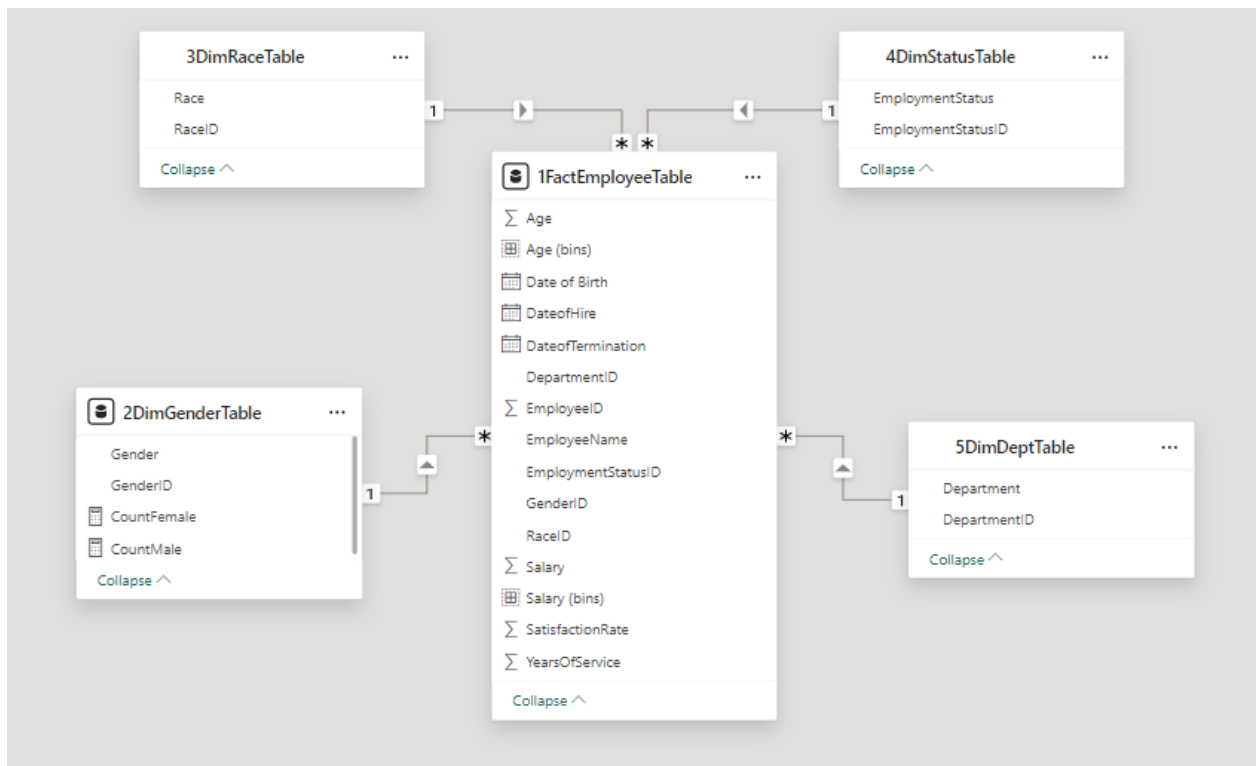


Figure 1: HR Dataset Data Model

3.4. Report – Executive Dashboard

The executive dashboard serves as a comprehensive overview of the most crucial data points, providing insights into key aspects of diversity within the organization. For this project, the executive dashboard is designed to be straightforward and informative, allowing stakeholders to quickly grasp important trends and metrics. It is divided into four main parts: slicers and cards for general reference, followed by sections dedicated to analyzing gender, age, and race diversity measures. For the diversity measures, the charts enable users to visually assess any significant disparities between employee count and salary at a glance in terms of gender, age and race. It's important to note that the data displayed in all the figures for this dashboard pertains only to employees with an Active status.

- a. **Slicer and Cards:** The slicers and cards serve as foundational elements for exploring key aspects of the dataset. The slicer allows users to filter data based on Employment Status, providing flexibility to analyze trends among Active, Terminated for Cause, and Voluntarily Terminated employees. While the focus is primarily on Active employees, the inclusion of terminated statuses enables users to recognize any patterns or trends related to gender, age, and race among departing employees. For instance, users can investigate whether there is an uneven rate of resignation among female employees or if terminations occur more frequently within specific age groups.

Additionally, the dashboard features three informative cards displaying Total Employees, Average Satisfaction Rate, and Average Salary. These metrics offer valuable insights regardless of the specific diversity measure being examined. For example, when exploring diversity within different racial groups, users can reference the count of employees within each group, assess their satisfaction levels compared to other racial demographics, and evaluate salary differences across various racial categories. These cards provide essential context for understanding the overall workforce composition and its corresponding satisfaction and compensation levels across different demographic groups.

For reference regarding the slicer and cards, please consult Figure 2.

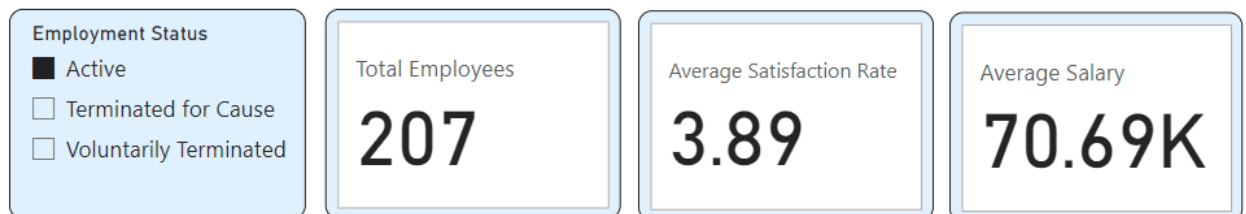


Figure 2: Executive Dashboard – Slicer and Cards

- b. **Gender Measures:** The dashboard includes two visuals (card and column chart) dedicated to Gender, focusing on the number of employees and average salary. For reference, please consult Figure 3.

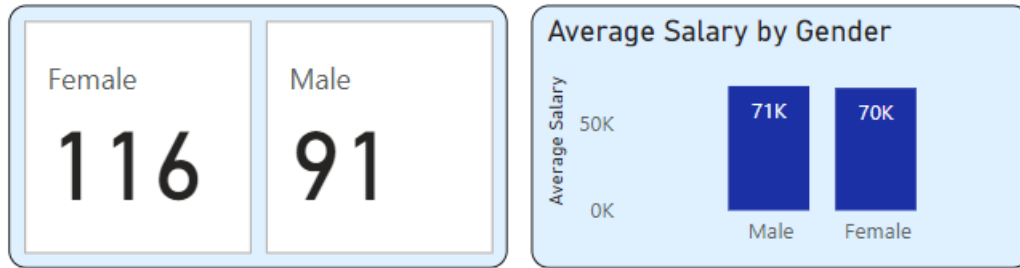


Figure 3: Executive Dashboard – Gender Measures

- c. **Age Measures:** Similarly, two charts, column and clustered bar chart, are dedicated to Age, displaying the percentage of employees within each age group (with a 10-year coverage) and average salary for each group. For reference, please consult Figure 4.

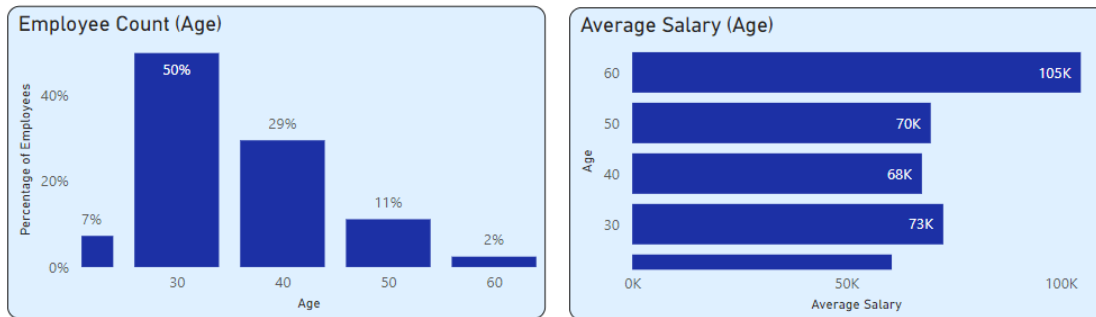


Figure 4: Executive Dashboard – Age Measures

- d. **Race Measures:** Like Gender and Age, two visuals, tree map and clustered bar chart, are dedicated to Race, showcasing the percentage of employees per race and their corresponding average salary. For reference, please consult Figure 5.

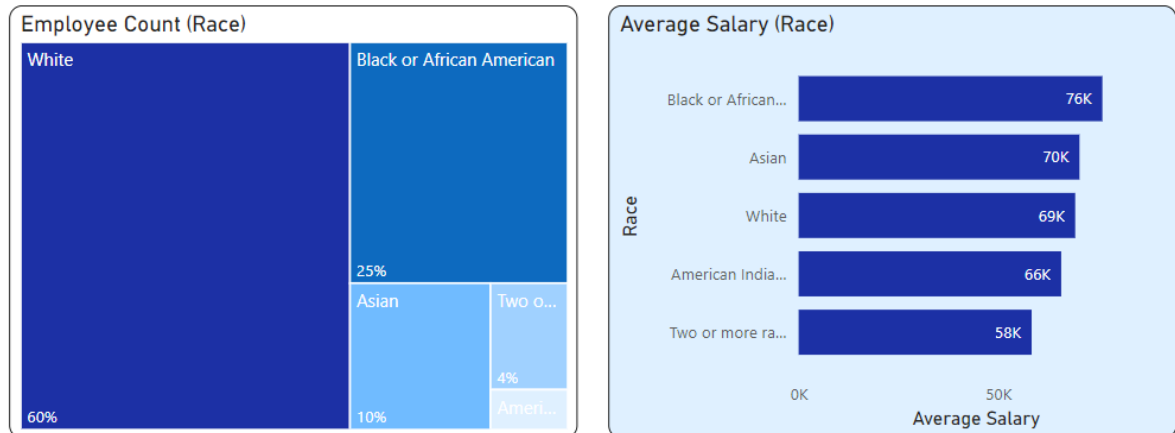


Figure 5: Executive Dashboard – Gender Measures

In summary, the executive dashboard provides a comprehensive overview of key diversity measures, including gender, age, and race, along with important HR metrics such as employee count, average salary, and satisfaction rate. For a consolidated view of all elements discussed, please refer to Figure 6.



Figure 6: Executive Dashboard

3.5. Report – Department Dashboard

The Department Dashboard serves as a detailed analysis of various diversity measures within each department of the company. While the Executive Dashboard provides a high-level overview, this dashboard delves deeper into the data by connecting each measure to the different departments. Like the Executive Dashboard though, it is also composed of four main parts. The first part consists of a slicer, allowing users to filter the data based on employee status, and the subsequent parts focus on departmental analysis based on Gender, Age, and Race. For all the department analysis charts, the Department will be used as the x-axis and the Average Salary as the y-axis.

It's important to note that the Executive Office was removed from the Department axis for all visuals. This decision was made because the Executive Office comprises only one person, which skews the data as it represents the highest position and therefore the highest salary in the company. It's also worth mentioning that the data displayed in all the figures for this dashboard pertains only to employees with an Active status.

- a. **Slicer:** The Employee Status slicer, as used in the Executive Dashboard, is retained here to facilitate filtering of data based on employment status.
- b. **Department Analysis by Gender:** This section utilizes a line and clustered column chart to visualize gender diversity within each department. The chart displays male and female salaries side by side within each department, allowing for easy comparison. The line represents employee count per department. For visual reference, please consult Figure 7.

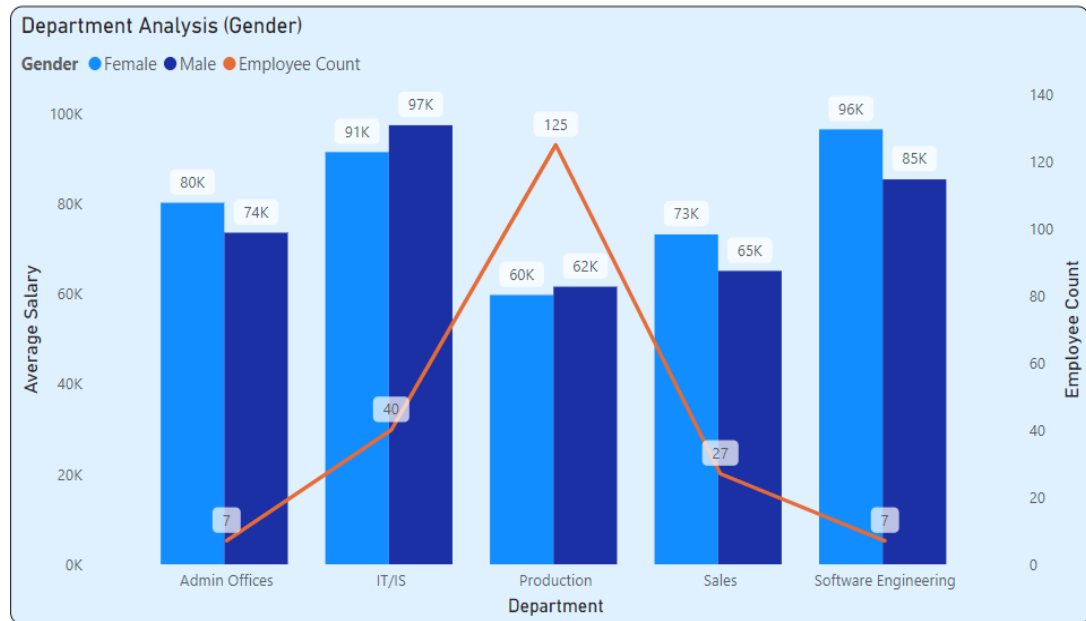


Figure 7: Department Analysis by Gender

- c. **Department Analysis by Age:** This section utilizes a line and stacked column chart to analyze age diversity across departments. The chart displays average salary per department, with the line representing the average age per department. For visual reference, please consult Figure 8.

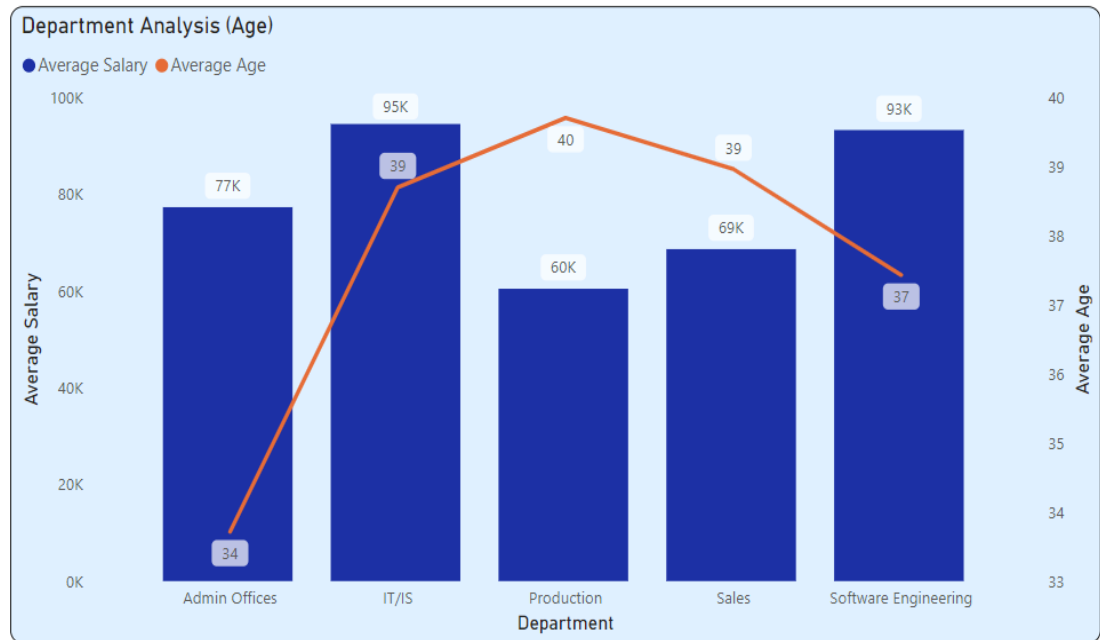


Figure 8: Department Analysis by Age

- d. **Department Analysis by Race:** This section utilizes a line and stacked column chart to examine racial diversity within each department. The chart stacks salary data for each race within departments, facilitating the identification of any significant disparities. Like the Gender analysis, the line represents employee count per department. For visual reference, please consult Figure 9.

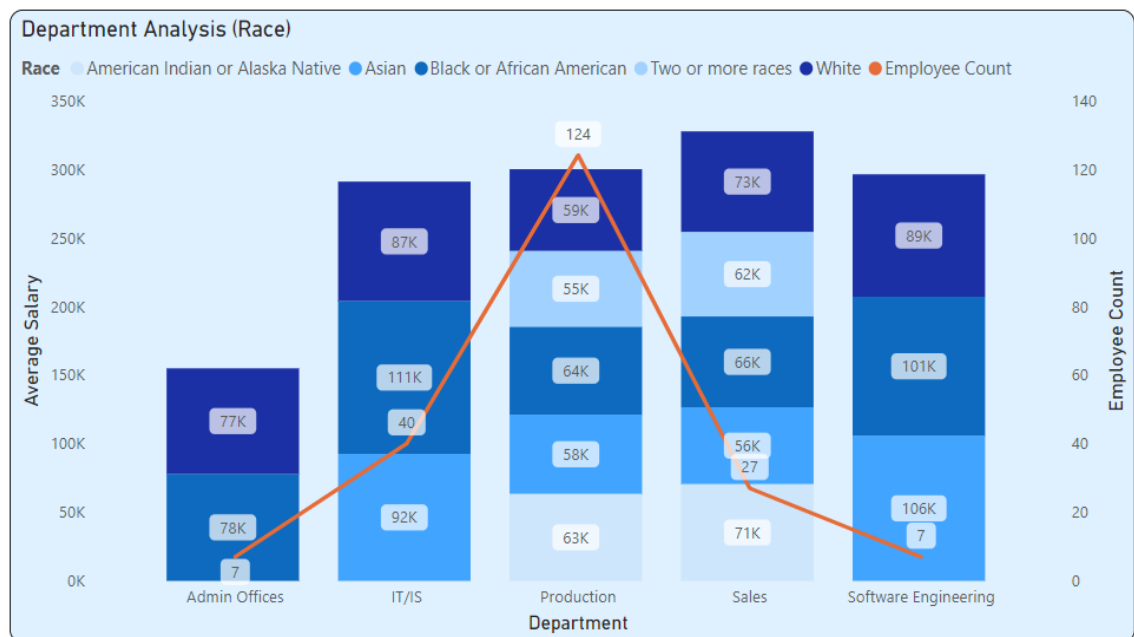


Figure 9: Department Analysis by Race

In summary, the department dashboard offers an insightful analysis of diversity measures within each department of the company, allowing stakeholders to gain deeper insights into diversity dynamics at the departmental level. For a consolidated view of all elements discussed, please refer to Figure 10.

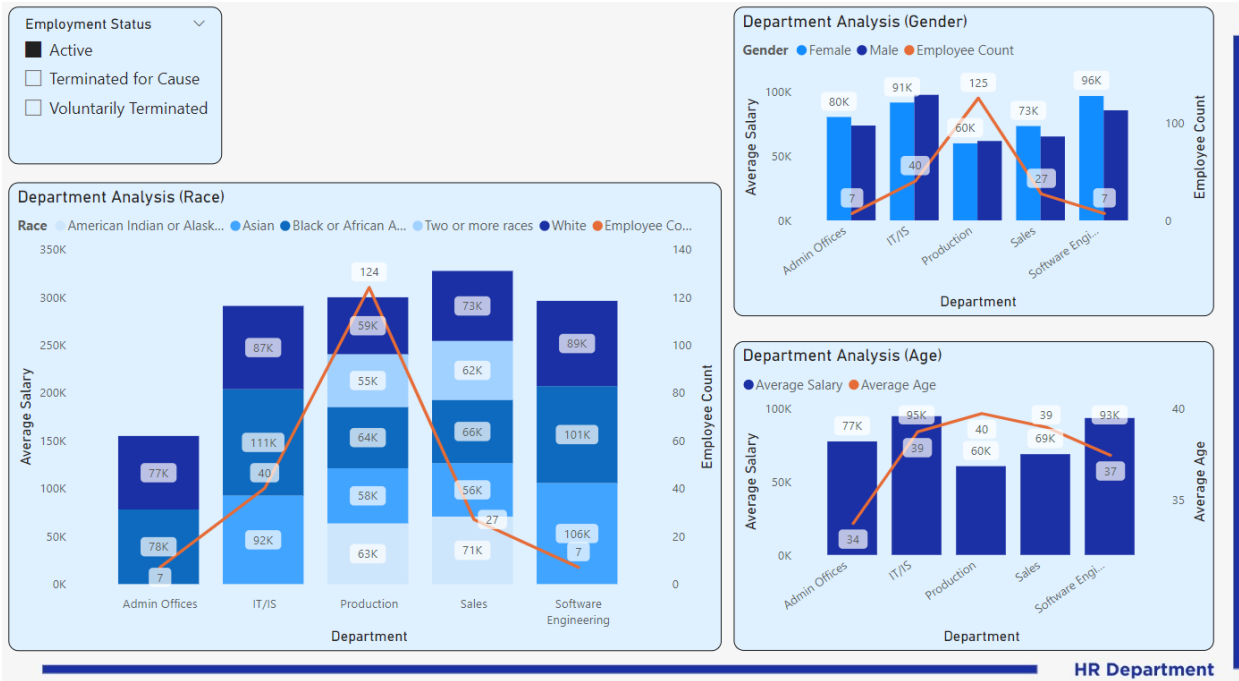


Figure 10: Department Dashboard

4. Insights

At this point, a deeper understanding of the dataset is gained through thorough observation and analysis. By examining factors such as gender, age and race, stakeholders can identify trends, patterns, and potential areas for improvement, ultimately guiding strategic decision-making and fostering a more inclusive workplace culture. In this section, observations and analyses are presented based on the diversity categories: Gender, Age, and Race.

- Gender:** The employee count in terms of gender shows no significant disparity, indicating a fair hiring process regardless of gender even in a labor-intensive company. The average salary for both genders is almost similar, with only the production department showing a slightly higher average for males, which may be understandable given that labor-intensive jobs require more physical strength. Satisfaction ratings, although slightly higher for females, fall within a 1% gap of the overall average, suggesting overall contentment among employees regardless of gender.

- b. **Age:** Employee counts across age groups follow typical trends, with a lower percentage of the workforce in the extremes (less than 25 and above 60) and the majority falling within the 30 to 50 age range. Average salaries across age groups are fairly consistent, with slight variations within a \$10,000 range from the overall average. The higher average salary for the 60s age group is influenced by the salary of the President, an outlier in this analysis. Excluding this outlier, the average salary of the 60s drops down to \$74,000 which aligns with other age groups. However, satisfaction ratings show a notable difference among the 20s age group, with a rating of 3.53.
- c. **Race:** The company's racial diversity profile is in line with the average for US companies, with approximately 60% of employees comprising white employees. Salary gaps among racial groups are minimal and not alarming. Similarly, average satisfaction rates among racial groups are within a good range of the overall average, indicating overall satisfaction among employees.

Upon comprehensive analysis, the company demonstrates a commendable level of diversity across gender, age, and race. The examination of gender diversity reveals a balanced distribution of male and female employees. Age-wise, the workforce distribution aligns with typical trends and overall salary ranges remain consistent across age groups. Regarding racial diversity, the company mirrors the demographic composition of the broader population, showcasing a diverse workforce. Moreover, salary parity across racial groups underscores the company's commitment to equality. Collectively, these insights affirm the company's proactive approach to fostering an inclusive environment, with no significant disparities detected across diversity dimensions.

5. Summary

The dataset selected for this project, sourced from Kaggle, revolves around Human Resources, chosen for its cleanliness, comprehensiveness, and universal relevance to businesses. Despite its initial quality, further transformation was undertaken to ensure alignment with the project's KPIs, which predominantly center on analyzing the company's diversity profile and identifying areas for enhancement. These KPIs incorporate gender, age, racial diversity, and salary equity, representing crucial aspects of organizational inclusivity and equity. Utilizing PowerBI for data visualization, the report is structured into two dashboards, each offering a visual narrative of key data points pertinent to the defined KPIs. The Executive Dashboard provides a high-level overview, allowing stakeholders to quickly grasp important trends and metrics related to diversity measures. Meanwhile, the

Department Dashboard delves deeper, connecting each diversity measure to the different departments within the company.

Upon thorough analysis and interpretation of the visuals presented in both dashboards, it is evident that the company demonstrates a commendable level of diversity across gender, age, and race. Notably, the data reveals no glaring disparities in diversity metrics, suggesting a healthy organizational culture characterized by fairness and equity. This confirms the company's dedication to promoting diversity and inclusion, serving as a testament to its efforts in creating a workplace environment that celebrates and embraces differences among its workforce.