



TASK

Exploratory Data Analysis on the Hourlywages Dataset

[Visit our website](#)

Introduction

This dataset captures the interplay between employee hourly wage, age, and position in our business. It unveils compensation trends, highlights workforce composition, and sheds light on organisational dynamics. These insights guide decisions on retention, competitive positioning, and strategic talent management.

hospital or office: 0 = "Hospital" and 1 = "Office"

age range refers to age ranges of the nurses; 1 = "18-30", 2 = "31-45" and 3 = "46-65".

years is the number of years of experience where; 1 = "5 or less", 2 = "6-10", 3 = "11-15", 4 = "16-20", 5 = "21-35" and 6 = "36 or more"

DATA CLEANING

During the data cleansing phase, several steps were taken to prepare the dataset for analysis.

1. Column Removal:

All columns were essential to the dataset and as such none of them were removed. This decision was driven by the focus on specific visualisations and the relevance of information related to travel aspects. For future projects, excluding these columns might prove advantageous.

2. Remove Duplicate Rows:

Duplicate rows were removed to maintain data integrity and ensure that each entry in the dataset is unique.

3. Remove Rows with Missing Data:

Rows with missing or zero values in crucial columns, such as profit were filtered out. This step helps in focusing the analysis on complete and meaningful data.

4. Filling missing values:

Missing data can lead to gaps in analysis and potentially distort the results. Filling in missing values helps to maintain the integrity of the dataset, ensuring that every record has complete information.

5. Change Data Types:

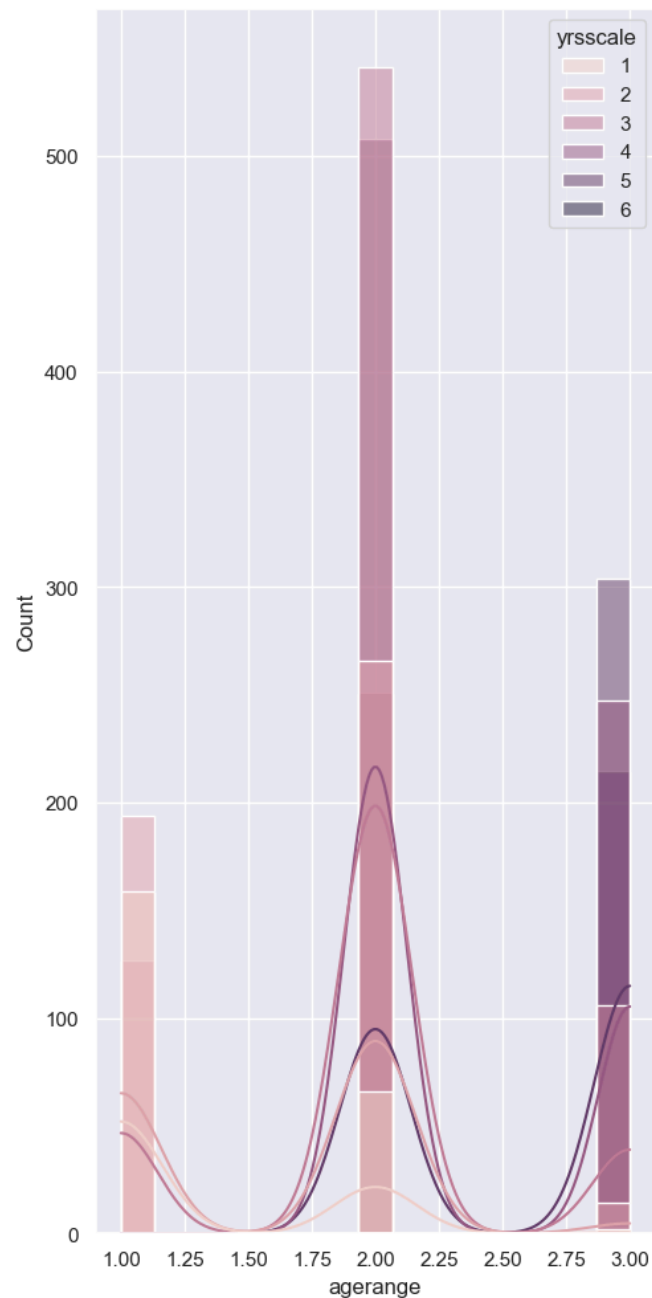
- Data types were adjusted to facilitate easier manipulation.
- Here hourly wage was rounded to 2 decimal places for better visualisation

- And also changed the information to a numerical format.

6. Flatten Columns:

- were flattened, making the data more accessible for analysis.
- Inspecting flattened data involves reviewing the structure, content, and quality of the data in its flattened form.

DATA STORIES AND VISUALISATIONS



{Total Earnings} = \$15.75/hour times 40 \ hours

The perception of whether \$630 for 40 hours of work is a lot or not can vary depending on factors such as location, industry standards, and individual financial circumstances. However, we can discuss some general considerations:

1. Hospital Work vs. Office Work:

- Hospital Work: In healthcare settings, especially for roles like nursing, the pay might be influenced by factors like the level of expertise, responsibilities, and the specific healthcare facility. Healthcare professionals often work in shifts, including nights and weekends, and may receive additional compensation for overtime or working during less desirable hours.

- Office Work: Office-based roles might follow a standard 9-to-5 schedule with no or limited overtime. The compensation structure may be different, and the nature of the work may not involve the same level of urgency or shift differentials as hospital work.

2. Overtime and On-Call Considerations:

- Hospital Work: Overtime and on-call work in hospitals may result in higher earnings due to shift differentials and overtime pay.

- **Office Work:** Overtime might be less common in traditional office settings, and compensation structures may differ.

It's essential for individuals to consider their own financial needs, the demands of the job, and industry standards when evaluating the adequacy of their earnings. Additionally, factors like benefits, job satisfaction, and work-life balance play a crucial role in determining overall job satisfaction. The first histogram suggests that individuals in position 0 generally earn higher hourly wages, particularly for at 40 hours (5 days) . However, working such long hours may not align with everyone's preferences as position 0 is the only chart that continues to increase. For those seeking a stable income without extended work hours, position 1 could be a more suitable choice. Nevertheless, in the long run, position 0 offers the potential for higher earnings.

In the second histogram, a clear correlation emerges between the age range of employees and their experience at the company . This pattern implies either a higher job satisfaction among existing employees or a continual influx of new hires (people still happy in this job industry both 0 and 1). This dynamic contributes to a positive work environment, with both newer and more experienced employees playing a role in the company's success

Number of positions for positions 0 and 1

0	1
1945	966

A lower turnover in position 0 could imply various factors:

Job Stability: Employees in position 0 may consider their roles more stable, having invested years in acquiring specialised skills. This investment makes transitioning to new industries a highly stressful decision.

Job Satisfaction: The lower turnover rate in position 0 could be attributed to higher job satisfaction or better working conditions. Employees in this position may find their work environment more fulfilling, reducing the inclination to seek alternative opportunities.

Skills Retention: Position 0 might demand specialised skills or experience, making employees less likely to explore opportunities outside the organisation. The retention of these skills contributes to the overall stability of the workforce.

While lower turnover is positive for stability, it could also mean fewer opportunities for new hires or career advancement in that specific position. Regularly monitoring turnover rates and understanding the underlying reasons can offer valuable insights into workforce dynamics.

A higher turnover in position 1 compared to position 0 could imply various factors, and one potential indication is:

Job Scarcity:

This may suggest that there is a higher demand for the type of work associated with position 1, leading to more opportunities for employees to explore alternative roles or industries.

This thorough data cleansing and exploration process establishes a robust foundation for subsequent in-depth analysis and storytelling. The refined dataset now serves as a solid basis for extracting meaningful insights within the context of the automotive industry.

THIS REPORT WAS WRITTEN BY : Elsy Theledi
