Ali Powell - DS 2023

Dataset - AI and the Job Market

Are AI Generated Datasets Reflective of Real Trends?

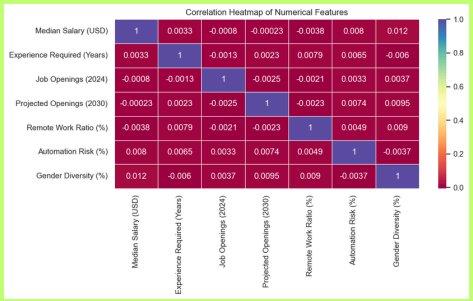
About the Main Dataset

This dataset was created by user Sahillslam007 on Kaggle, and was last updated in June of 2025. This AI generated dataset contains 30,000 rows and 13 columns, and reflect's AI's potential influence over the job market, from 2024 to 2030. According to the author, ""This is a synthetic dataset generated using realistic modeling, public job data patterns (U.S. BLS, OECD, McKinsey, WEF reports), and AI simulation to reflect plausible scenarios from 2024 to 2030. Ideal for educational, research, and AI project purposes".

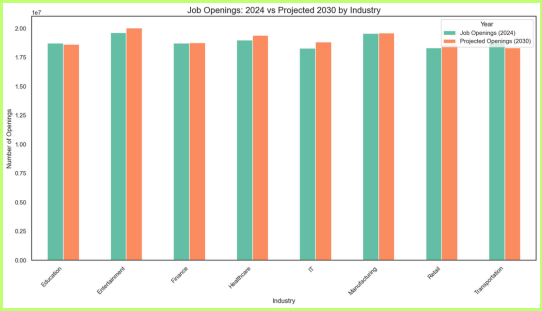
NOTE - For the best experience and easiest reading, zoom into a column below and scroll down through.

AI Dataset Outcome vs Real-World Patterns

Numeric Features



Job Openings



Salary v Experience

About Numeric Features Figure

This figure is a heat map displaying the correlations between all numerical variables in the dataset. As shown by the scale, most features exhibit very weak correlations with one another, indicated by low negative values. This means the variables largely behave independently and have minimal influence on each other. While some correlations are negative, the values are so small that they do not indicate any meaningful inverse relationships, just noise rather than a true trend. Overall, the heat map suggests that the dataset contains dispersed values without strong linear patterns, and no numeric feature meaningfully drives or predicts another (aside from self-correlation).

About Job Openings Figure

This figure is a bar chart comparing job openings in 2024 to projected openings in 2030 across various industries. The x-axis lists each industry, and the y-axis represents the number of openings in tens of millions. According to the chart, the dataset suggests only minimal change in job openings over time. A few industries, such as entertainment, healthcare, IT, and retail, show slight growth, while others, including education, finance, manufacturing, and transportation, remain relatively flat or show small declines. Notably, every industry falls within a narrow range of approximately 17.5 to 20 million openings in both years. This uniformity is unusual and raises questions about the realism of the dataset's projections.

About Salary v Experience Figure

This figure presents scatterplots illustrating the relationship between required experience and median salary for each industry in the dataset. Despite representing different sectors, all graphs display an almost identical pattern: years of experience appear to have no correlation with salary. The data points are evenly scattered in every plot, indicating uniformly low correlation across industries. Additionally, each industry shares the exact same salary range (approximately $30,000 to $160,000) with no outliers. All industries also cap required experience at 20 years, again with no variation. This uniformity suggests that the dataset lacks realistic differentiation between industries and does not reflect typical real world trends.

Is this realistic?

The table below examines several key features from the AI dataset and compares them with real-world evidence. As expected, some variables that appeared uncorrelated in the heatmap, such as median salary and experience, median salary and job openings, and experience and gender diversity, are in fact related in real labor-market research. However, certain features that showed no correlation in the heatmap, like automation risk and projected job openings, also show little to no real-world relationship. Overall, this analysis shows that many of the "no-correlation" patterns in the AI-generated dataset do not reflect real labor-market dynamics.

Is this realistic?

The graph below uses data from the US Bureau of Labor Statistics and mirrors the structure of the AI-generated employment graph. It compares projected employment levels across major US sectors between 2024 and 2030. Unlike the AI dataset, which showed minimal or stagnant global employment growth, the real BLS projections show a strong, consistent upward trend across nearly all sectors. This highlights that the AI-generated dataset does not accurately reflect real-world labor market patterns.

Is this realistic?

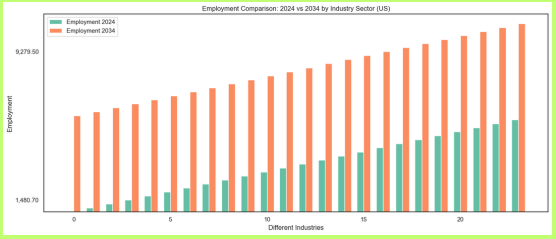
The graphs below, sourced from the US Bureau of Labor Statistics, show the relationship between salary and work experience across various US industries. In every scatterplot, higher experience levels consistently correspond to higher wages, with virtually no outliers. In contrast, the AI generated dataset shows no relationship between experience and salary, an outcome that does not align with real-world labor market trends. If the AI dataset were reflective of reality, this clear upward relationship would be present.

Real World Data

Compared Variables	Correlation	Source
Median Salary and Experience Required	**RELATED** - *generally experience positively affects salary*	US Bureau of Labor Statistics - The Employment Projections (EP)
Median Salary and Job Openings	**RELATED** - *large number of job openings usually signals a "tight" labor market, meaning that labor demand exceeds supply at the offered wage*	US Bureau of Labor Statistics - "Making Sense of Job Openings and Other Labor Market Measures"
Current Openings and Projected Openings	**UNCLEAR** - *no publicly documented direct relationship between current openings and projected openings*	US Bureau of Labor Statistics - - The Employment Projections (EP) Methodology
Experience Required and Gender Diversity	**RELATED** - Gender diversity and experience requirements are linked because female dominated jobs tend to be undervalued even when they require similar experience or skill levels.	Acker, Joan. "Gender Inequalities in the Workplace" (2015). - England, Paula et al. (2020). The Gender Wage Gap, Between-Firm Inequality, and Job Segregation by Gender
Automation Risk and Projected Openings	**NOT RELATED** - *automation/AI exposure and projected job openings are not strictly negatively correlated — many "high-risk" occupations continue to see projected employment growth*	Upjohn Institute — "AI exposure and the future of work: Linking task-based measures to U.S. occupational employment projections" - BLS — "Growth trends for selected occupations considered at risk from automation" (2022, Monthly Labor Review)

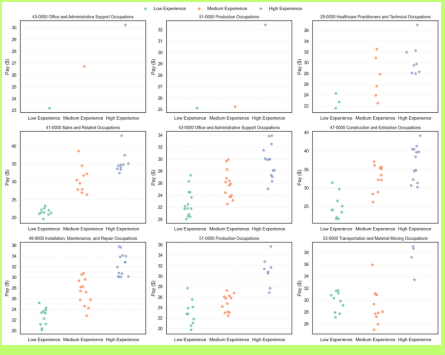
Various Cited Sources

Real-World Data



US Bureau of Labor Statistics

Real-World Data



US Bureau of Labor Statistics