SACRAMENTO STATE
COLLEGE OF CONTINUING EDUCATION

MSBA205

Data Analytics for Business

Prepared by

Md Tarbiar Swat – 304454332
Miguel Bendezu – 219647546
Brandon Cruz – 304695313
Syed Hashim Ali Gilani - 304439369

Project:
How Generative AI Adoption is Reshaping Workforce Efficiency: A Data-Driven Study

Submission Date
Aug 06, 2025

Prepared for
Dr. Fernanda Araujo Maciel

1.0 Dataset and EDA

Our project explores how GenAI (Generative AI) adoption is shaping productivity across companies worldwide. Using a large dataset of over 100,000 records from 2022 to 2024, we looked at how different industries, countries, and AI tools like ChatGPT, Gemini, Mixtral, and Groq are making an impact. We found that most companies started using GenAI after 2022, with 2023 being the peak year. On average, productivity increased by 18.5%, showing a strong positive effect. But this came with investment, companies spent up to 25,000 hours on training and impacted large teams, often around 10,000 employees. New job roles were also created, typically around 16 per company.

Our visuals highlighted how adoption patterns varied by country, industry, and AI tool, giving us a broad, balanced picture of how GenAI is transforming work at scale.

2.0 Initial Model and Baseline Testing

To begin, a preliminary correlation analysis was conducted to understand linear relationships among numeric predictors. The correlation matrix revealed very weak correlations between variables and with the target, indicating no multicollinearity and limited linear associations.

As a baseline, a multiple linear regression model was fit using all predictors, with productivity change as the dependent variable. This initial model aimed to assess overall performance rather than refinement. The output showed no statistically significant predictors, and model fit was poor: Adjusted R^2 was –0.0009, meaning no variation in productivity change was explained. The residual standard error was 9.53, and the overall p-value was 0.516, confirming the predictors did not significantly explain the outcome.

Residuals appeared symmetrically distributed around zero, but prediction errors were high. In some cases, prediction errors exceeded 15 units. The model's MAE was 10.5, meaning predictions were on average 10.5% off in either direction, considerable given the target range of 2%–35%. The MSE was 112.55, indicating large penalties for larger deviations. While this early model lacked predictive power, it provided a valuable benchmark for improvements ahead.

3.0 Model Improvement

We first checked for multicollinearity using VIF; all values were around 1, confirming no overlap among predictors (training hours, new roles, impacted employees, industry, country, or GenAI tool). Each variable contributed unique information.

Next, we applied backward, forward, and stepwise regression (based on AIC) to find the most predictive subset. However, all approaches converged on an intercept-only model, indicating that no predictor added more value than the overall mean.

We also tested whether transforming the outcome variable (productivity change) would help. Box-Cox analysis suggested an optimal exponent of 0.75, but since 1.0 was within the 95% CI, we kept the original scale.

While this null result may seem underwhelming, it's meaningful: none of the measured factors reliably predict productivity change. For businesses, this implies that large-scale GenAI adoption doesn't automatically boost productivity, more targeted strategies or better metrics may be needed to uncover true drivers.

4.0 Model Summary and Interpretation

In our final step, we wanted to answer one core question: what really drives productivity change when companies adopt GenAI? To make the results more reliable and focused, we selected a balanced sample of 500 companies and built a model using practical factors like:

- Training hours provided to employees
- How many people were affected
- Whether new roles were created
- The industry the company operates in
- The type of GenAI tool they adopted

When we tested the model, the results were clear: no single factor stood out as a strong driver of productivity. While we expected to see strong relationships, the patterns just weren't there. We also checked how accurate the model's predictions were. The average difference between what the model predicted and what actually happened was around 8.3 percentage points, which isn't bad, but it's not precise enough to rely on.

To double-check the fairness of the model, we used a visual tool that shows whether the model is biased or off in any direction. The result? The pattern was flat and balanced, which means our model is stable, it's just that the inputs don't strongly predict productivity.

Final Thought:

Sometimes, numbers alone don't tell the full story. Productivity gains from GenAI might depend more on human factors, like leadership, mindset, or how the change was introduced which weren't captured in our dataset.