EMPLOYEE PRODUCTIVITY MODELING FOR A MID-SIZED IT SERVICES FIRM USING PYTHON REGRESSION

1. Introduction

An IT services company with over 200 employees across development, design, and testing departments struggled to understand the factors influencing team productivity. The HR team needed help determining how variables such as training hours, working hours, and employee role affected individual output. We were brought in to design a regression-based analytical solution using Python.

2. Objective

- To identify which work-related and demographic factors influence productivity
- To create a multiple linear regression model using Python, including dummy and interaction variables
- To deliver interpretable, actionable results for the HR and operations teams

3. Data Provided by Client

The client shared anonymized data for 210 employees over one fiscal quarter. Each record included:

- Productivity_Score (measured using internal KPI scale, 0–100)
- Avg_Working_Hours_Per_Week
- Monthly_Training_Hours
- Experience_Level (Junior, Mid-Level, Senior)
- Department (Development, Design, Testing)

Dummy variables were created for Department and Experience_Level. An interaction term was created between Training_Hours and Experience_Level.

4. Methodology

4.1 Data Preparation

- Cleaned inconsistent values and normalized productivity scores
- Encoded categorical variables using dummy variables
- Created interaction term between Monthly_Training_Hours × Experience_Level
- Identified and removed 3 high-leverage outlier records using Cook's Distance

4.2 Model Development

- Built the regression model using statsmodels.api.OLS
- Checked regression assumptions using residual plots and multicollinearity checks (VIF)
- Used backward elimination to retain only significant predictors ($p < 0.05$)

4.3 Deliverables

- Python notebook with documented code and results
- Visual summary (PDF) with coefficient table and charts
- Slide deck summary prepared for HR review meeting

5. Key Results

- Model R-squared: **0.67**, Adjusted R-squared: **0.65**
- **Training Hours** significantly improved productivity only for **Junior** employees
- **Avg_Working_Hours** had a positive effect up to 45 hours/week; beyond that, the effect plateaued
- **Department** differences were notable: Development had the highest baseline productivity
- The **interaction between training and experience** showed that training investment had minimal effect on seniors

6. Business Impact

- HR shifted its training budget toward junior employees, improving ROI
- Mid-level managers received department-specific productivity insights to manage workload
- Helped optimize resource allocation in the design team, which had the widest performance spread

- Estimated 10–12% improvement in productivity across Q3 after applying model-driven changes

7. Future Scope Suggested

- Add qualitative feedback data (employee surveys) for enriched analysis
- Introduce time-series tracking for productivity to observe trends across quarters
- Create department-specific models for deeper focus
- Build an interactive HR dashboard using Plotly or Power BI for ongoing tracking