DATA SCIENCE JOBS PROJECT.

BUSINESS PROBLEM.

KenyaData Insights, a cutting-edge data firm in Kenya, is about to debut with a unique focus on learning about the wage expectations of people aspiring to venture into the world of Data Science. From upper management to lower-level employees, every tier of the organisational structure will be covered by our investigation. Acknowledging the critical role that pay considerations play in determining job satisfaction, our startup is committed to thoroughly investigating the effects of these expectations on employee satisfaction and engagement. KenyaData Insights makes it a priority to develop a predictive model to analyse and forecast salaries for data science job positions based on factors such as job titles, experience level, company size, and other relevant variables. The goal is to provide actionable insights and guidance to our stakeholders, enabling them to make informed decisions about job opportunities and negotiate competitive salaries.

OVERVIEW.

The goal of the study is to better understand the complex relationships that exist within the data and machine learning business. Specifically, the study will look at how job features, compensation structures, and the dynamics of remote work interact with one another. The main goals are to investigate the differences in pay against various variables including employee residency, job titles, remote work ratios, and experience levels. This is approached by creating visuals/graphs and statistical metrics like correlation coefficients. The study also attempts to provide light on how distant work patterns are distributed within the sector by examining variables such as job titles, work experience, and their effects on work-life balance and overall job satisfaction.

BUSINESS STAKEHOLDERS.

The primary business stakeholders for this project are future data scientists, data science students, academic institutions, employers in the data science industry, career advisors and even recruitment agencies. The objective is to provide valuable insights that will aid in the strategic determination of salaries and facilitate important decision-making processes. Our aim is to empower the people to make decisions

that not only enhance the overall working environment but also contribute to the success of their businesses. By addressing the needs of these stakeholders, the project aims to contribute to a more efficient and informed data science job market and to foster an environment that supports both employee well-being and business prosperity.

BUSINESS OBJECTIVE.

Our business objectives include developing a strong salary prediction model to better understand the factors that influence industry salaries, optimising placement strategies for future data scientists based on identified success factors, improving recruitment processes with data-driven insights, and optimising career advisory services by providing personalised guidance. The project also intends to build ethical data governance standards and cultivate engagement with educational institutions and business partners in order to ensure responsible data usage and improve alignment between educational and industrial demands. Overall, the initiative aims to benefit the industry by delivering actionable data and supporting better decisions in personnel management and career development.

HYPOTHESIS.

i.) Hypothesis on Job Titles and Compensation:

- Null Hypothesis (H0): There is no significant difference in compensation across various job titles.
- Alternative Hypothesis (H1): Compensation levels vary significantly based on different job titles.

ii) Hypothesis on Work Experience and Salary:

- Null Hypothesis (H0): Work experience does not have a significant impact on salary levels.
- Alternative Hypothesis (H1): Employees with longer work experience receive higher salaries.

iii.) Hypothesis on Regional Salary Disparities:

- Null Hypothesis (H0): There are no significant differences in salary distributions across various nations and regions.
- Alternative Hypothesis (H1): Salary distributions vary significantly across different nations and regions.

DATA.

Here are the datasets:
https://www.kaggle.com/datasets/georgejnr/advertised-data-science-jobs-dataset/discussion/393915
https://www.kaggle.com/code/hasibalmuzdadid/data-science-jobs-salary-analysis-retro-vibe/input

- Business understanding
- Data understanding
- Modeling
- Regression results
- Conclusions
- Recommendations

Multiple linear regression models

A simple linear regression model that can

Predict salary based on:

1. Experience level
2. Company Size
3. Employment type
4. Company location
5. Country of Residence