

Course : Data Analytics

Year : 2024

Group Project Report Professionalism

**Complied By :**

|  |  |  |
| --- | --- | --- |
| **Surname** | **Full Names** | **Student Number** |
| ALEXANDRE | EDNA DA SILVA | 223031461 |
| AMAKUTSI | HERBERT DATI-TANGI | 223067962 |
| ANTONIO | ROBIN PANDULENI | 220112312 |
| AZUWA | KWENACHUKWU | 223039659 |
| BASSON | NICOLA YVONNE | 222115823 |
| BERNARD | FRANK NDATALIISHUNA | 223012807 |
| BIRECH | SHALOM SASHA JERUTO | 222129433 |
| BVOCHO | TATENDA | 221140263 |
| CHAVES | LEANDRO ISAIAS | 223082279 |
| CHIWARA | GARETH NEIL MUNYARADZI | 223003956 |

**Introduction to AI in the Workforce: A Job Market Analysis**

This analysis delves into the "AI Powered Job Market Insights" dataset, offering a comprehensive exploration of how artificial intelligence (AI) and automation are influencing the modern job market. The dataset provides a snapshot of job listings across various industries, highlighting the following factors AI adoption levels, automation risk and salary ranges. By examining these elements, we aim to uncover key trends and insights into how AI is reshaping employment opportunities and job market dynamics.

**Objective**The primary objective of this analysis is to identify and understand key patterns related to salaries, required skills, remote work and job market trends influenced by AI. By employing various analytical techniques, we seek to reveal how AI adoption impacts compensation, skill requirements and job growth, as well as how it correlates with the likelihood of remote work.

**Data Overview The dataset includes the following key features:**

* Job\_Title: The title of the job role.
* Industry: The industry in which the job is located.
* Company\_Size: The size of the company offering the job.
* Location: The geographic location of the job.
* AI\_Adoption\_Level: The extent to which the company has adopted AI in its operations.
* Automation\_Risk: The estimated risk of the job being automated within the next 10 years.
* Required\_Skills: The key skills required for the job role.
* Salary\_USD: The annual salary offered for the job in USD.
* Remote\_Friendly: Indicates whether the job can be performed remotely.
* Job\_Growth\_Projection: The projected growth or decline of the job role over the next five years.

**Columns :**

**1.Job Title:**

* Description: The title of the job role.
* Type: Categorical

**2.Industry:**

* Description: The industry in which the job is located.
* Type: Categorical

**3.Company Size:**

* Description: The size of the company offering the job.
* Type: Categorical

**4.Location:**

* Description: The geographic location of the job.
* Type: Categorical

**5.AI\_Adoption Level:**

* Description: The extent to which the company has adopted AI in its operations.
* Type: Categorical

**6.Automation Risk:**

* Description: The estimated risk that the job could be automated within the next 10 years.
* Type: Categorical

**7.Required Skills:**

* Description: The key skills required for the job role.
* Type: Categorical

**8.Salary USD:**

* Description: The annual salary offered for the job in USD.
* Type: Numerical

**9.Remote Friendly**:

* Description: Indicates whether the job can be performed remotely.
* Type: Categorical

**10.Job Growth Projection:**

* Description: The projected growth or decline of the job role over the next five years.
* Type: Categorical

**Reason behind the selection of this dataset (why the dataset is ideal)**

The dataset is ideal due to its comprehensive coverage of AI adoption trends across diverse industries and job roles, offering valuable insights into the changing skill demands and their implications. Moreover, its inclusion of salary data allows for a deep dive into the correlation between AI integration and income dynamics across various job titles and locations, making it an ideal resource for in-depth analysis and trend identification in the context of AI impact on employment landscapes.

**The problem being solved by analysis of the dataset.**

The analysis of this dataset aims to address the critical gaps in understanding the impact of AI adoption on different industries and job roles. By exploring the demand for specific skills across sectors and how AI influences these requirements, the analysis seeks to provide actionable insights for individuals and organizations looking to navigate the evolving job market shaped by AI technologies. Additionally, by conducting a salary analysis and examining the correlation between AI adoption and salary ranges across diverse job titles and locations, the analysis aims to uncover patterns and trends that shed light on the relationship between AI implementation and compensation structures within the workforce. Ultimately, the goal is to provide a comprehensive understanding of how AI adoption is reshaping industries, job roles, and earning potentials, facilitating informed decision-making and strategic planning in the context of the AI-driven economy. The following are the problems being addressed :

* AI and Job Market Research: Analyzing the impact of AI adoption on different industries and job roles.
* Skill Gap Analysis: Understanding which skills are in demand across industries and how AI influences this demand.
* Policy Making: Assisting policymakers in identifying job roles at high risk of automation and strategizing for workforce transitions.
* Salary Analysis: Exploring the correlation between AI adoption and salary ranges across different job titles and locations.

**Data Cleaning Techniques**

1.Handling Outliers : Detection, Identifying outliers using statistical methods like Z-Score, IQR (Interquartile Range) and visualization techniques.

2. Data Transformation: Feature Engineering creating new features from existing ones to improve model performance and Log Transformation Addressing skewed data distributions.

3. Error Correction : Consistency Checks Ensuring consistency in data formats and values

4. Handling Inconsistent Data: Normalization Ensuring consistency in naming conventions and units.

5. Data Validation:Conducting sanity checks to ensure data accuracy and integrity.

6.Handling Missing Values

7.Standardization and Normalization: Standardize numerical data like "Salary USD" to ensure all values are on a similar scale.

8.Encoding Categorical Variables: Converting categorical variables like "Job Title," "Industry," "Company Size," using the following techniques one-hot encoding and label encoding.

9.Handling Duplicates: Check for and remove any duplicate entries in the dataset to maintain data integrity.

10.Data Formatting: Ensure consistency in data formats, especially for categorical columns e.g "Company Size" and "AI Adoption"

**Analysis Highlights**

* **Distribution of Salaries Across Industries and AI Adoption Levels: A box plot will illustrate how salaries differ across various industries, segmented by AI adoption levels. This analysis will help identify high-paying industries and assess the impact of AI on compensation.**
* **Skill Frequency Analysis: A count plot will show the frequency of required skills, segmented by AI adoption levels. This will reveal which skills are more common in roles with high AI adoption.**
* **Correlation Matrix: A heatmap will explore the relationships between numerical features such as salary, AI adoption, automation risk, and job growth. This will uncover trends and correlations, such as whether higher AI adoption is linked to increased salaries or reduced automation risk.**
* **AI Adoption Level Impact on Remote Work: A heatmap will examine the relationship between AI adoption levels and the likelihood of a job being remote-friendly, exploring if AI adoption promotes remote work opportunities.**
* **Industry-Specific Salary Distribution: A focused box plot will analyze salary distributions within the Technology industry, segmented by AI adoption level and automation risk, providing deeper insights into compensation within this sector.**
* **Clustering Analysis: KMeans clustering will group similar job roles based on AI adoption level, automation risk, and salary, helping to identify patterns and categorize jobs with similar characteristics.**
* **Point-Biserial Correlation: An analysis of the point-biserial correlation will explore the relationship between AI adoption levels and the correlation between remote work and salary, revealing how AI adoption influences this dynamic.**

**By combining these analyses, this report aims to provide a thorough understanding of how AI is transforming the job market, offering valuable insights into compensation trends, skill demands, and employment opportunities in an AI-driven landscape.**