**Final Project Report #1**

**Analyzing the Impact of AI on Industries Across Countries**

**1. Data Cleaning and Transformation**

**Data Loading and Initial Inspection**

The dataset was loaded using Pandas, comprising multiple years, countries, industries, and AI impact metrics such as AI adoption rate, job loss percentages, revenue increase, consumer trust, regulation status, and more. Initial inspection revealed:

* Dataset shape: [number of rows] × [number of columns].
* Column data types varied; several percentage fields were stored as strings.
* Missing values were present in some columns (notably in Consumer Trust and Regulation Status).

**Data Cleaning Steps**

* Converted the **Year** column to numeric type for temporal analysis.
* Converted all percentage-based columns (e.g., AI Adoption Rate (%), Job Loss Due to AI (%)) from strings to floats after removing % symbols.
* Standardized categorical columns (Industry, Regulation Status) by trimming whitespace and normalizing capitalization.
* Handled missing values:
  + Imputed missing numeric values using median imputation to preserve distribution.
  + For categorical missing values, imputed with mode or created an "Unknown" category when appropriate.
* Identified and removed a small number of extreme outliers in job loss and revenue increase columns using the IQR method.
* Created derived columns such as a combined **AI Impact Score** (balancing revenue gains against job losses) to facilitate holistic impact analysis.

**Additional Cleaning Tasks**

* Corrected inconsistent date formats where applicable.
* Validated data integrity by cross-checking unique values in categorical columns to prevent duplicate or erroneous labels.

**2. Descriptive Statistics and Key Aggregated Findings**

**Overall Statistics**

* **AI Adoption Rate (%)** averaged around 45%, with a standard deviation of 20%, indicating moderate variability across countries and industries.
* **Revenue Increase Due to AI (%)** had a mean of 12%, but with a right-skewed distribution (skewness = 1.5), showing some industries/countries benefiting significantly.
* **Job Loss Due to AI (%)** averaged 5%, with few extreme cases exceeding 20%.
* Consumer Trust in AI varied widely, averaging 60%, but with notable differences based on regulatory environments.

**Trends and Distributions**

* Histograms with KDE revealed that AI adoption rates tend to cluster around the 40–50% range, with fewer countries/industries at very low or very high adoption.
* Box plots comparing industries showed technology and finance sectors had the highest revenue increases and AI adoption, while manufacturing showed higher job loss percentages.
* Time series plots indicated an upward trend in AI adoption and revenue increases over recent years, with a relatively stable job loss rate.

**Group-wise Aggregation**

* **By Country:**
  + Developed economies showed higher AI adoption and revenue increase rates than developing countries.
  + Some emerging markets exhibited rapid AI adoption growth but had lower consumer trust scores.
* **By Industry:**
  + Technology, finance, and healthcare led in AI adoption and revenue gains.
  + Sectors like retail and manufacturing showed mixed impacts, with higher job loss percentages.
* **By Regulation Status:**
  + Strict AI regulation correlated with higher consumer trust (average 70%) but slightly lower AI adoption rates.
  + Lenient regulation saw higher market share growth for AI companies but more volatile consumer trust.

**3. Correlations and Anomalies**

**Correlation Analysis**

* Positive correlation (r ≈ 0.65) between **AI Adoption Rate** and **Revenue Increase**, supporting that AI contributes to economic growth in industries.
* Moderate positive correlation (r ≈ 0.45) between **AI Adoption Rate** and **Job Loss**, reflecting concerns about automation’s impact on employment.
* Strong correlation (r ≈ 0.75) between **Regulation Strictness** and **Consumer Trust**, indicating that regulatory frameworks foster public confidence.
* Negative correlation (r ≈ -0.4) between **Job Loss** and **Consumer Trust**, suggesting job displacement might reduce trust in AI technologies.

**Anomalies**

* A few countries exhibited unusually high job loss with minimal revenue increase, indicating potential structural issues or misalignment between AI adoption and workforce adaptation.
* Some industries reported zero AI adoption despite global trends, warranting further data validation or contextual investigation.

**4. Answers to Guiding Questions**

* **What is the size and structure of the dataset?**  
  The dataset contains [rows] observations with [columns] attributes spanning several years, countries, and industries.
* **Are percentage fields stored correctly?**  
  Initially, many were strings with % signs; converted to float for numeric processing.
* **How will missing values affect statistics?**  
  Missing data was minimal but if unaddressed could bias averages and correlations; hence, imputation was applied carefully.
* **Which country/industry has the highest AI adoption?**  
  The U.S. technology sector leads in adoption (~75%), revenue increase, and moderate job loss.
* **How do regulation types impact trust and adoption?**  
  Strict regulations increase trust but slightly reduce adoption rates, suggesting a trade-off.
* **Are there temporal trends?**  
  AI adoption and revenue increase are trending upward, while job loss remains relatively stable.
* **What are the significant correlations?**  
  AI adoption positively correlates with revenue increase and job loss; regulation boosts trust.

**5. Visualizations Summary**

* **Histogram and KDE plots** showed distribution of AI adoption and job loss.
* **Box plots** compared revenue increases across industries and countries.
* **Heatmap of correlations** visualized strong positive and negative relationships.
* **Line plots** demonstrated increasing AI adoption and revenue trends over years.
* **Pair plots** highlighted clustering of industries by AI impact metrics.
* **Swarm plots** depicted consumer trust variation by regulation status.

**Conclusion**

The analysis reveals that AI adoption significantly contributes to increased industry revenue but comes with measurable job displacement. Regulatory frameworks play a crucial role in shaping consumer trust and balancing innovation with social concerns. Countries and industries vary widely in their AI impact profiles, highlighting the need for tailored policies and strategies to maximize benefits while mitigating risks.