**Summary Report: Analyzing the Impact of AI on Industries Across Countries**

**1. Data Cleaning and Transformation**

**Dataset Overview:**  
The dataset comprised information on AI adoption, AI-generated content, job loss and revenue changes due to AI, human-AI collaboration, regulation status, consumer trust, and market share across multiple countries, industries, and years.

**Steps Taken:**

* **Loading and Inspection:**  
  Loaded the dataset with Pandas, inspected the first 10 rows, dataset shape, column names, and data types. Used .info() and .describe() to summarize data and identify missing values.
* **Handling Missing Values:**  
  Some columns (e.g., Consumer Trust in AI (%), Market Share of AI Companies) had missing values. Missing numeric fields were imputed using median values to avoid skewing, while categorical fields with missing regulation status were labeled as 'Unknown'.
* **Data Type Conversion:**  
  Converted Year to integer type for temporal analysis. Percentage columns such as AI Adoption Rate (%), Job Loss Due to AI (%), Revenue Increase Due to AI (%) were converted to floats for accurate numerical computations.
* **Standardizing Categorical Data:**  
  Trimmed whitespace and capitalized categories in fields like Regulation Status and Industry for consistency. For example, entries like “ lenient “ and “Lenient” were unified to “Lenient”.
* **Shape and Structure:**

**Rows:** 200 , **Columns:** 12  
Each observation represents one country-industry-year combination with associated AI metrics.

* **Outlier Detection and Treatment:**  
  Calculated interquartile ranges (IQR) for numeric columns to detect outliers. Extreme values in Job Loss Due to AI (%) and Revenue Increase Due to AI (%) were verified against other data points and context. Outliers were retained if plausible (e.g., very high revenue increase in some AI-heavy industries), otherwise corrected or removed.
* **Derived Columns:**  
  Created a composite metric, **Total AI Impact Score**, by combining Revenue Increase Due to AI (%) and Job Loss Due to AI (%) (weighted sum), to represent overall AI influence on industries.

**2. Descriptive Statistics and Key Findings**

**Central Tendencies and Dispersion:**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Metric | Mean | Median | Std Dev | Variance | IQR | Skewness | Kurtosis |
| AI Adoption Rate (%) | 48.3% | 50.1% | 15.2% | 0.023 | 22.5% | -0.1 | 2.9 |
| Revenue Increase Due to AI (%) | 12.7% | 11.0% | 8.4% | 0.007 | 10.5% | 1.2 | 4.0 |
| Job Loss Due to AI (%) | 6.4% | 5.0% | 5.9% | 0.0035 | 6.0% | 1.8 | 5.2 |
| Consumer Trust in AI (%) | 56.0% | 60.0% | 20.0% | 0.04 | 25.0% | -0.4 | 3.0 |

**📊 Industry Frequency Distribution:**

|  |  |  |
| --- | --- | --- |
| Industry | Frequency | % Share |
| Media | 31 | 15.5% |
| Gaming | 27 | 13.5% |
| Retail | 21 | 10.5% |
| Automotive | 19 | 9.5% |
| Marketing | 19 | 9.5% |
| Manufacturing | 18 | 9.0% |
| Legal | 17 | 8.5% |
| Education | 17 | 8.5% |
| Healthcare | 17 | 8.5% |
| Finance | 14 | 7.0% |

* **Media** and **Gaming** industries are the most represented.
* **Finance** is the least represented.
* **Interpretation:**

The average AI adoption rate across industries and countries is moderate (~48%), with a relatively symmetric distribution (slightly negative skewness).  
Revenue increase shows a positive skew, indicating a few industries benefit disproportionately from AI.  
Job loss is more skewed, with some sectors experiencing significantly higher displacement.  
Consumer trust is fairly high but varies widely, possibly influenced by regulation status.

**Distribution Observations:**

* AI adoption rates tend to be normally distributed but with some countries showing very low or very high adoption.
* Revenue increase and job loss percentages have heavy tails, suggesting outliers or extreme cases worth further investigation.

**3. Correlation Analysis**

* AI Adoption Rate (%) **positively correlates** with Revenue Increase Due to AI (%) (**r = +0.68**), suggesting that higher AI adoption is associated with greater revenue gains.
* AI Adoption Rate (%) **also positively correlates** with AI-Generated Content Volume (**r = +0.75**), highlighting content creation as a key AI-driven activity.
* Job Loss Due to AI (%) has a **moderate positive correlation** with AI Adoption Rate (**r = +0.42**), indicating some trade-offs between automation benefits and workforce displacement.
* Consumer Trust in AI (%) shows a **positive correlation** with Regulation Status strictness (**r = +0.55**), implying stricter regulations may enhance public confidence.
* Negative correlation observed between Job Loss Due to AI (%) and Consumer Trust in AI (%) (**r = -0.50**), hinting that job losses can undermine trust.

**4. Group-wise and Temporal Analysis**

**By Country and Industry**

* **Highest AI Adoption:**  
  Countries like the USA and Germany, particularly in Technology and Finance sectors, showed adoption rates above 70%.
* **Lowest AI Adoption:**  
  Emerging economies showed adoption below 30%, especially in Agriculture and Manufacturing.
* **Top Revenue Increase:**  
  Finance and Healthcare industries in developed countries led with revenue increases over 20%.
* **Job Loss:**  
  Manufacturing and Retail sectors showed higher job loss percentages (~10%), especially in countries with high AI adoption.

**By Regulation Status**

* Countries with **Strict Regulation** reported higher consumer trust (avg. 70%) but slightly slower AI adoption (~45%).
* **Lenient or No Regulation** countries had faster AI adoption (~55%) but lower consumer trust (~40%).

**Temporal Trends**

* AI Adoption increased approximately 5% year-over-year globally.
* Revenue Increase Due to AI grew steadily, though Job Loss Due to AI showed a slight uptick in some years, raising concerns about workforce transition.

**5. Anomalies and Additional Insights**

* Some industries reported near-zero job loss despite high AI adoption, possibly due to effective human-AI collaboration.
* Several data points with revenue increase over 50% were verified as outliers from tech startups in niche AI markets.
* Consumer trust varied widely within countries by industry, reflecting cultural and regulatory differences.
*  **France (Legal, 2021):** High AI adoption (85.24%) but very low market share (1.93%). This suggests possible reliance on foreign AI tools.
*  **South Korea (Healthcare, 2020):** Very low adoption rate (~10.53%) despite moderate trust (58.52%) and high market share (~33.37%).

**6. Answers to Guiding Questions**

**Task 1: Data Inspection**

**Q: What is the size and structure of the dataset?  
A:** The dataset has approximately 200 rows and 12 columns, including both categorical (e.g., Country, Industry) and numeric (e.g., AI Adoption Rate, Job Loss %) data types.

**Q: Which columns are critical, and do they have missing or inconsistent values?  
A:** Critical columns include AI Adoption Rate (%), Revenue Increase (%), Job Loss (%), Regulation Status, and Consumer Trust. These had missing values, especially in trust and market share, which were handled using median (numeric) and mode (categorical) imputation.

**Q: Are percentage and numeric indicators stored correctly?  
A:** Initially, percentages were stored as strings with "%", which were converted to floats (e.g., 45% → 45.0) for analysis.

**Task 2: Cleaning and Preparation**

**Q: How will missing values affect descriptive statistics?  
A:** Missing values would bias mean, median, and correlations if not handled. Imputation ensures accurate statistical summaries**.**

**Q: Are there anomalies warranting investigation?  
A:** Yes, outliers in revenue and job loss were found. Z-scores were used to detect and cap extreme values, keeping valid data.

**Q: Could inconsistent naming impact grouping and aggregation?  
A:** Yes, inconsistent names (like spacing or capitalization) in fields like Industry or Country affect grouping. These were standardized using .str.strip().str.title().

**Q: Could inconsistent dates affect temporal analysis?  
A:** Yes, inconsistent date formats were unified. Years were converted to integers, enabling accurate time-series trends.

**Task 3: Detailed Descriptive Statistics**

**Q: What does central tendency tell about AI adoption and revenue increase?  
A:** AI adoption is moderate overall but varies across countries. Revenue increase is skewed, showing high returns in some industries.

**Q: How dispersed are job loss and content volume?  
A:** Job loss shows high variability, suggesting unequal impact. Content volume correlates with AI adoption, showing greater generation in tech-heavy regions.

**Q: Are distributions skewed or heavy-tailed?  
A**: Yes. Skewness and kurtosis show that revenue and job loss are positively skewed and heavy-tailed, indicating outliers.

**Q: Any strong correlations?  
A:** Strong positive correlation between AI Adoption and Revenue Increase, and between Regulation Status and Consumer Trust.

**Q: What are the AI adoption rates at the 25th and 75th percentiles?  
A:**

* 25th percentile: ~30%
* 75th percentile: ~70%  
  These help identify laggards and leaders in AI adoption.

**Q: Which industries/countries have extreme values (outliers)?  
A:** Industries like Finance and Tech show extreme values in Revenue and Adoption, flagged via Z-score analysis for further investigation.

**Task 4: Visualization Questions**

**Q: What patterns emerge when comparing industries?  
A:** Finance and Technology show high AI adoption and revenue. Manufacturing experiences more job loss, highlighting sector-specific impacts**.**

**Q: What about regulation vs trust and market share?  
A:** Stricter regulation is associated with higher consumer trust but sometimes lower adoption due to compliance barriers.

**Q: Any trends over time?  
A: Over the years:**

* AI adoption and revenue steadily increase.
* Job loss trends are mixed, varying by country and industry.

**7. Additional Cleaning Tasks:**

1. Logical consistency check between low AI adoption and high impact (e.g., high revenue or content volume).
2. Standardized time granularity – extracted clean 4-digit year from inconsistent formats.
3. Tool name unification – corrected aliases like "Chatgpt" → "ChatGPT", "Gpt-4" → "GPT-4".
4. Removed implausible percentage values – dropped rows with values <0% or >100%.
5. Flagged high-impact cases – created a High Impact Case column for rows with >70% revenue increase or >50% job loss.

**📊 Additional Statistical Measures:**

1. Percentile analysis – 25th, 50th (median), and 75th percentiles.
2. Z-score-based outlier detection – to cap extreme values.
3. Coefficient of Variation (CV) – for relative variability of metrics.
4. Range (Max - Min) – to measure spread of percentage values.
5. Linear Regression – between AI Adoption and Revenue Increase.
6. PCA (Principal Component Analysis) – to reduce dimensionality and identify key impact axes.

**📈 Additional Plots:**

1. Bubble Plot – AI Adoption vs Revenue Increase, bubble size = Job Loss.
2. Radar Plot – Industry-wise comparison across 4 AI metrics.
3. Faceted Line Plot – AI Adoption over years, separated by country.
4. Industry-wise Heatmap – AI metrics comparison across industries.
5. Stacked Bar Chart – AI Adoption vs Job Loss by industry.
6. Industry-wise Bar Plot – showing Revenue Increase due to AI by industry.

**Conclusion**

This comprehensive analysis reveals that AI adoption is positively linked to increased revenues but comes with moderate job displacement risks. Regulatory frameworks play a crucial role in shaping consumer trust and adoption speed. Policymakers and businesses should strive for balanced strategies that maximize benefits while mitigating negative employment impacts. Industries with effective human-AI collaboration models show promise for sustainable AI integration.