Dataset 2: Gareth AI Job Market Insights

1. The Problem of the Dataset:

The dataset focuses on the effects of AI adoption on the job market, with an emphasis on automation risks, salaries, and job growth trends across different industries. The problem centers on how industries are responding to AI and automation, and how these technologies are reshaping employment. Companies, policymakers, and job seekers must prepare for the changes in the workforce driven by AI, including understanding which job roles are at risk of automation and which skills are most in demand.

2. Reason Behind the Selection of This Dataset:

This dataset is an ideal tool for studying the evolving job market in light of AI and automation advancements. It provides key insights into the intersection of technology and labor, showing which jobs are most susceptible to automation, which industries are embracing AI, and what skills are necessary for workers to remain competitive. Furthermore, the dataset includes information on salaries, job growth projections, and remote work trends, which makes it invaluable for businesses, government agencies, educators, and individuals looking to navigate the future of work.

3. Problems Being Solved by Analysis of the Dataset:

- **Impact of AI on Jobs:** Understand how the adoption of AI is transforming industries and job roles, helping companies and workers adapt to these changes.
- **Automation Risk:** Identify jobs that are at the highest risk of being automated and recommend appropriate measures for reskilling and upskilling.
- **Job Growth Projections**: Determine which job categories are expected to experience growth or decline, enabling students and professionals to align their skills with future market demands.
- **Salary Trends Across Roles and Locations:** Explore how salaries vary by industry, location, and job role to inform competitive salary benchmarks.
- **Remote Work Trends:** Examine how AI adoption correlates with the availability of remote work opportunities and which sectors are most open to flexible work arrangements.

4. List of Columns and Their Descriptions:

- **Job_Title (Object):** The job position being offered (e.g., Cybersecurity Analyst, Marketing Specialist)**.**
- **Industry (Object):** The industry sector that the job belongs to (e.g., Technology, Retail).
- **Company_Size (Object): Indicates the size of the company offering the position (e.g., Small, Medium, Large).**
- **Location (Object**): The physical or remote location where the job is based.

- **AI_Adoption_Level (Object):** A measure of how much the company or industry has integrated AI into its operations (Low, Medium, High).
- **Automation_Risk (Object):** The likelihood that this job will be automated in the near future (Low, Medium, High).
- **Required_Skills (Object):** A list of essential skills needed for the job (e.g., Project Management, UX/UI Design).
- **Salary_USD (Float):** The annual salary for the position, given in US dollars.
- **Remote_Friendly (Object**): Indicates whether the job can be done remotely (Yes/No).
- **Job_Growth_Projection (Object):** A forecast of whether the job category is expected to grow or decline in the future (Growth, Decline).

5. Data Cleaning Techniques:

- **Handling Missing Data:** Address missing values in key fields such as Salary_USD or Job_Growth_Projection by filling them in with estimates based on similar jobs or dropping incomplete rows if critical data is absent**.**
- **Standardizing Industry and Company Size Categories:** Ensure that similar terms (e.g., "Tech" vs. "Technology") are standardized to avoid confusion and ensure uniformity in analysis.
- **Cleaning and Formatting Job Titles:** Normalize variations in job titles (e.g., "Software Developer" vs. "Software Engineer") to reduce duplication and increase consistency in the analysis.
- **Outlier Detection in Salary Data:** Identify any extreme values in the Salary_USD field that could distort the analysis (e.g., abnormally high or low salaries) and handle them appropriately by capping or removing.
- **Verifying AI Adoption and Automation Risk Levels:** Cross-check fields like AI_Adoption_Level and Automation_Risk to ensure logical coherence; for example, a high AI adoption level should not be paired with a low automation risk without a clear explanation.
- **Location Normalization:** Standardize location fields to ensure that cities and regions are consistently represented, enabling easier location-based analysis.
- **Skill Standardization**: Group similar skills under common categories to prevent fragmentation (e.g., combining “Data Science” and “Data Analytics” into a broader category like “Data Skills”).