AI Impact Analysis on Indian Job Domains

Strategic Workforce Transformation Report

Author: Gundeti Sai Likith – Aspiring Data Analyst

Tools Utilized: Python (Pandas, Scikit-learn, Scipy, BeautifulSoup), Power BI, Matplotlib/Seaborn

Date: December 2025

1. Executive Summary

People are worried that AI is taking over jobs, but they don't know exactly which jobs are safe, which are using AI, and which will disappear due to automation in future. This project addresses the growing concern regarding Artificial Intelligence's role in the Indian job market. By analysing 5,000 scraped job records across 11 key domains, the study classifies roles into three categories: **AI Adoption, Automation role** and **Low AI Impact**.

A predictive Machine Learning model was developed using a Decision Tree classifier, achieving **84% accuracy** in predicting a job's AI impact category. The findings indicate that while entry-level roles face the highest automation risk, human-centric roles like UI/UX and Product Management remain resilient.

The result will give people and companies clear, data-backed answers so they can make smart choices about careers, training, and hiring for the future.

2. Business Objective:

We have collected data on 11 popular jobs in India (in different cities and for different experience levels). We need to analyse this data to create a clear picture that shows:

AI Adoption roles- Jobs currently requiring AI skills like **AI Creation, Management, or Advanced Use.**

Automation role - Jobs that AI can automate in Future like **repetitive, rules-based tasks.**

Low AI Impact role - Jobs that are resistant to AI automation.

3. Data Overview:

- **Data Acquisition:** Custom web scraping pipeline using BeautifulSoup to extract 5,000 records from major Indian job portal LinkedIn and stored as a excel file

- **Schema:** A consolidated web_scraping_data excel file containing fields for Job Title, Company, Location, Experience Level, Job Description, and Application Counts.

- **Feature Engineering: Developed a heuristic classification function.**

  - **Logic:** Applied keyword-based mapping on Job Descriptions to categorize roles into *AI Adoption* (e.g., "LLM," "Neural Networks"), *Automation* (e.g., "Excel Macro" "robotic process automation"), and *Low Impact* and added this data as a column(Ai Impact Category) in the data frame for further analysis

- **Preprocessing:** Handled null values and duplicates; normalized categorical text for machine learning readiness.

4. Methodology

Tools Used:

- **Python** (via Jupyter Notebook) for web scraping, data cleaning, analysis, visualization and Hypothesis Testing, ML model building and web page design
- **Power BI** for interactive dashboard Creation

Workflow:

1. Scraped data using Beautiful Soap (Python) and saved as a excel file
2. Created final file after data cleaning, feature engineering and added new column (AI Impact Category)
3. Performed exploratory data analysis (EDA) numerical and categorical columns characteristics
4. Performed Data Analysis and Hypothesis Testing
5. Built a Machine Learning model with 84% accuracy
6. Designed a web page local host that takes certain inputs from user on and predicts output either AI Adoption role, Automation role or Low AI Impact role
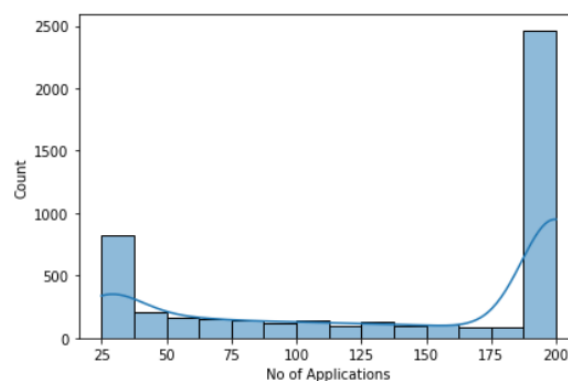
Key Techniques:

- Extracting job data from web HTML via Python
- Categorizing job descriptions into AI Impact levels.
- Using Chi-Square to prove experience-level plays a key role in AI Impact
- Predicting 84% of job outcomes using Decision Tree ML Model.

Challenges:

- Ensured reproducibility across tools by documenting Python workflows
- Designing web page using streamlit

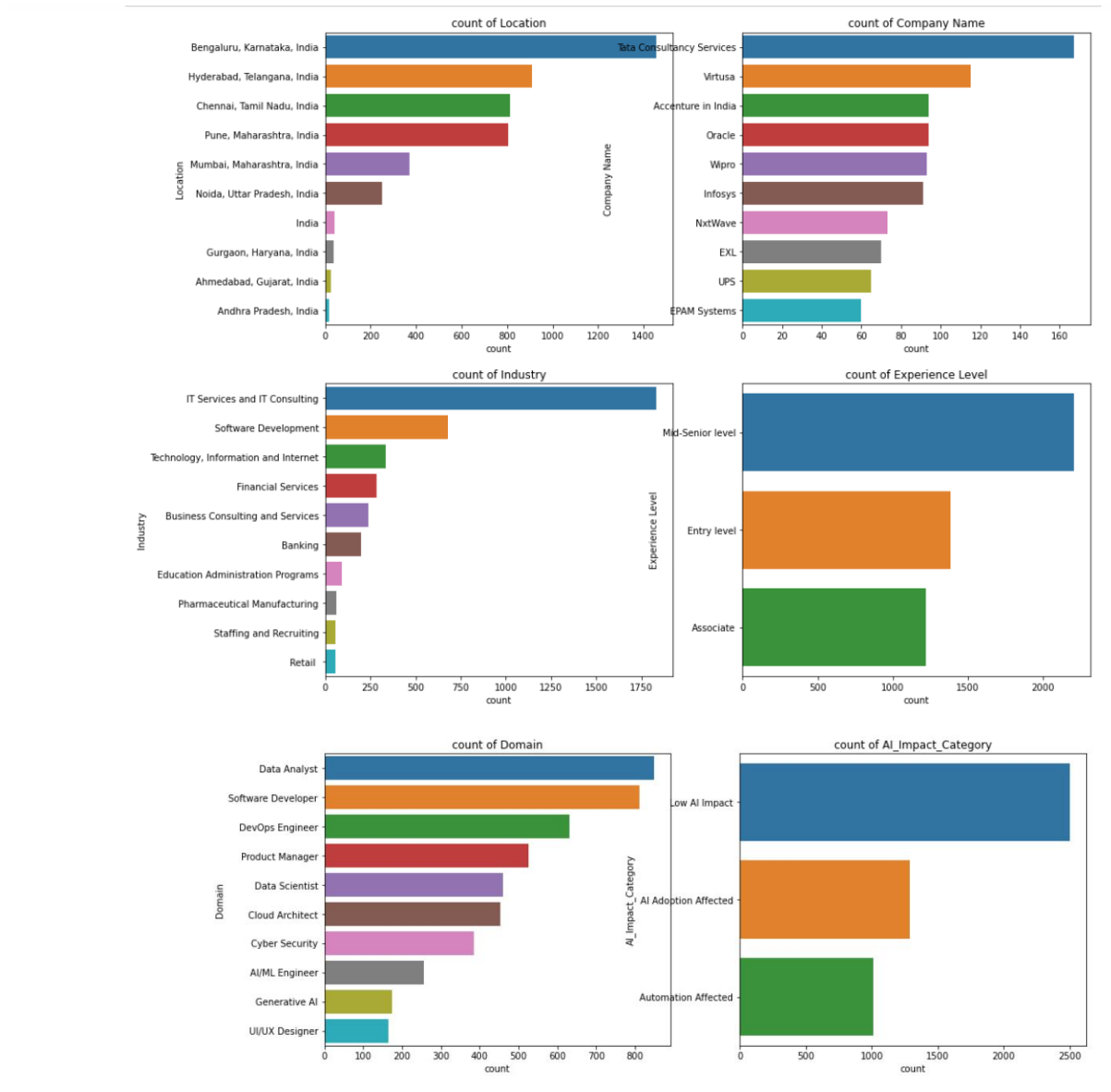**5. Exploratory Data Analysis:**

Numerical Characteristics:

Summary Statistics Insights:

- The **median (50%)** and **75th percentile** are both **200** and the **maximum** is also **200**. This means at least half of the records reached the cap of 200 applications, suggesting either a system-imposed limit or that many applicants consistently hit the maximum allowed.

- The mean is about **140**, but the standard deviation is **71.4**, showing wide spread. However, the **minimum** is only **25**, and the **25th** percentile is **63**, which indicates a **large gap** between low-application cases and the capped maximum. This skew suggests two groups: one with relatively few applications and another hitting the maximum.
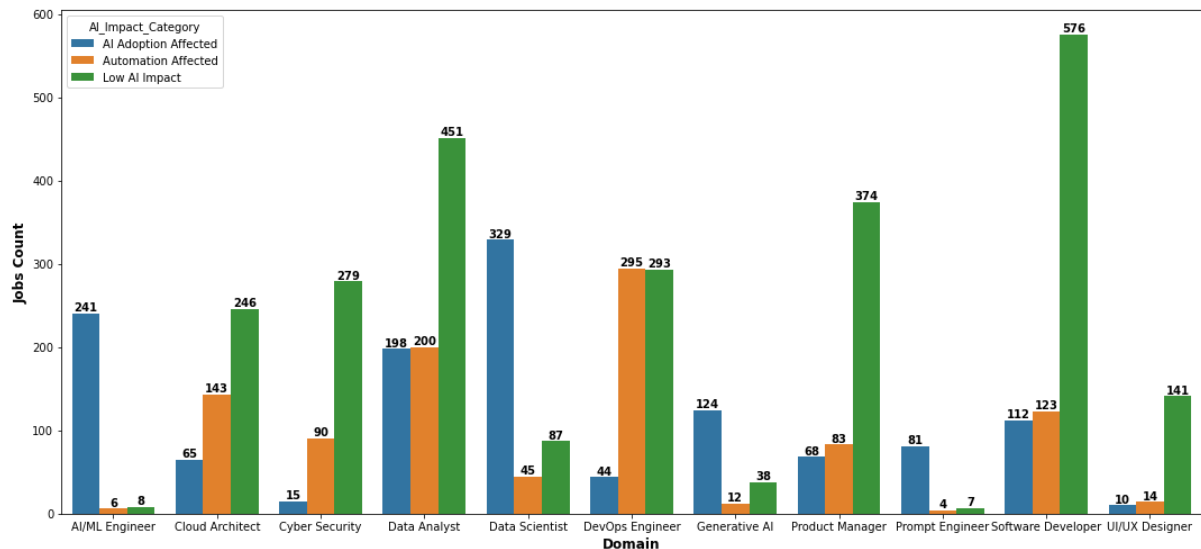
Categorical Columns Characteristics:

Summary Insights

- **Top Work Locations**: Bengaluru leads with over **1,000** job entries, followed by Hyderabad, Chennai and Pune with around **600–800** each. These cities are major tech hubs in India.

- **Popular Companies**: Tata Consultancy Services has the most listings—more than **150**. Other big companies include Virtusa, Oracle, Wipro, and Infosys, each with **100–120** roles.

- **Experience Level**: Most jobs are for **Mid-Senior level** professionals—about **2000** roles. Entry-level and Associate positions are less compared to Mid-Senior level, around **1000–1500** each.

- **AI Impact**: Around **2500** roles are marked as having **Low AI Impact**, while **1000–1500** are affected by AI adoption or automation. This shows many tech jobs are still safe from AI disruption.
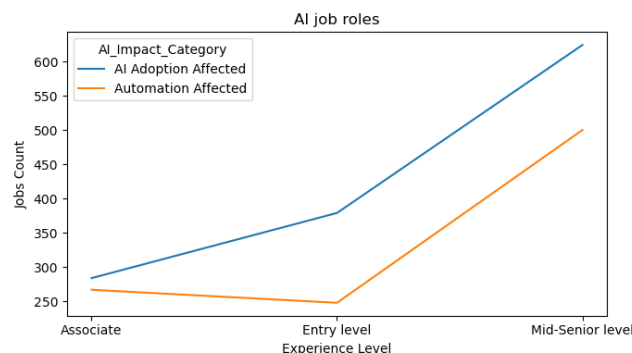
6. Data Analysis:

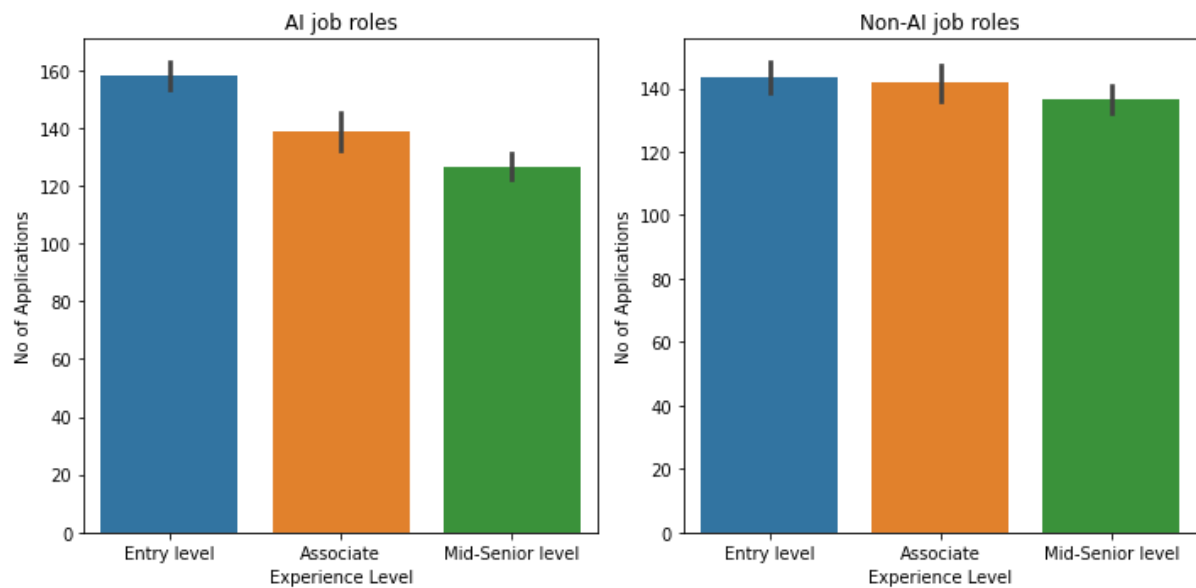**Which job domains are most associated with AI adoption, automation risk, and low AI impact?**

- Domains most associated with AI adoption are **AI/ML Engineer(94%), Prompt Engineer(88%), Data Scientist(71.3%), Gen AI(71.2%)**

- Domains most associated with automation risk are DevOps Engineer(46%) followed by Cloud Architect(31%), Data Analyst(23.5%), Cyber Security(23.4%)

    - **Data analyst and Cyber Security** were **not highly** affected by automation

- Domains associated with low AI impact UI/UX Designer, Cyber Security, Product Manager, Software Developer followed by

    - **Cloud Architect, DevOps Engineer, Data Analyst** were **not highly** safe from AI

Did AI tools make a key role of switching from junior to senior level?
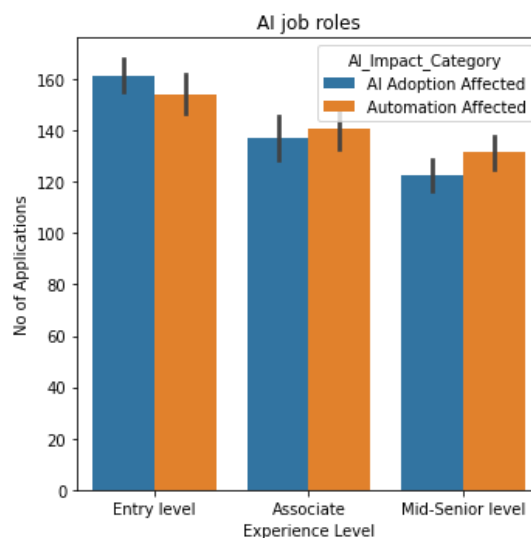

In AI adoption and automation related roles mid senior level roles are almost **twice** of junior roles - that means experience and ai tools makes a key role in switching from junior to senior level position

How does the number of applications vary between ai-affected and ai-resistant jobs across experience levels?
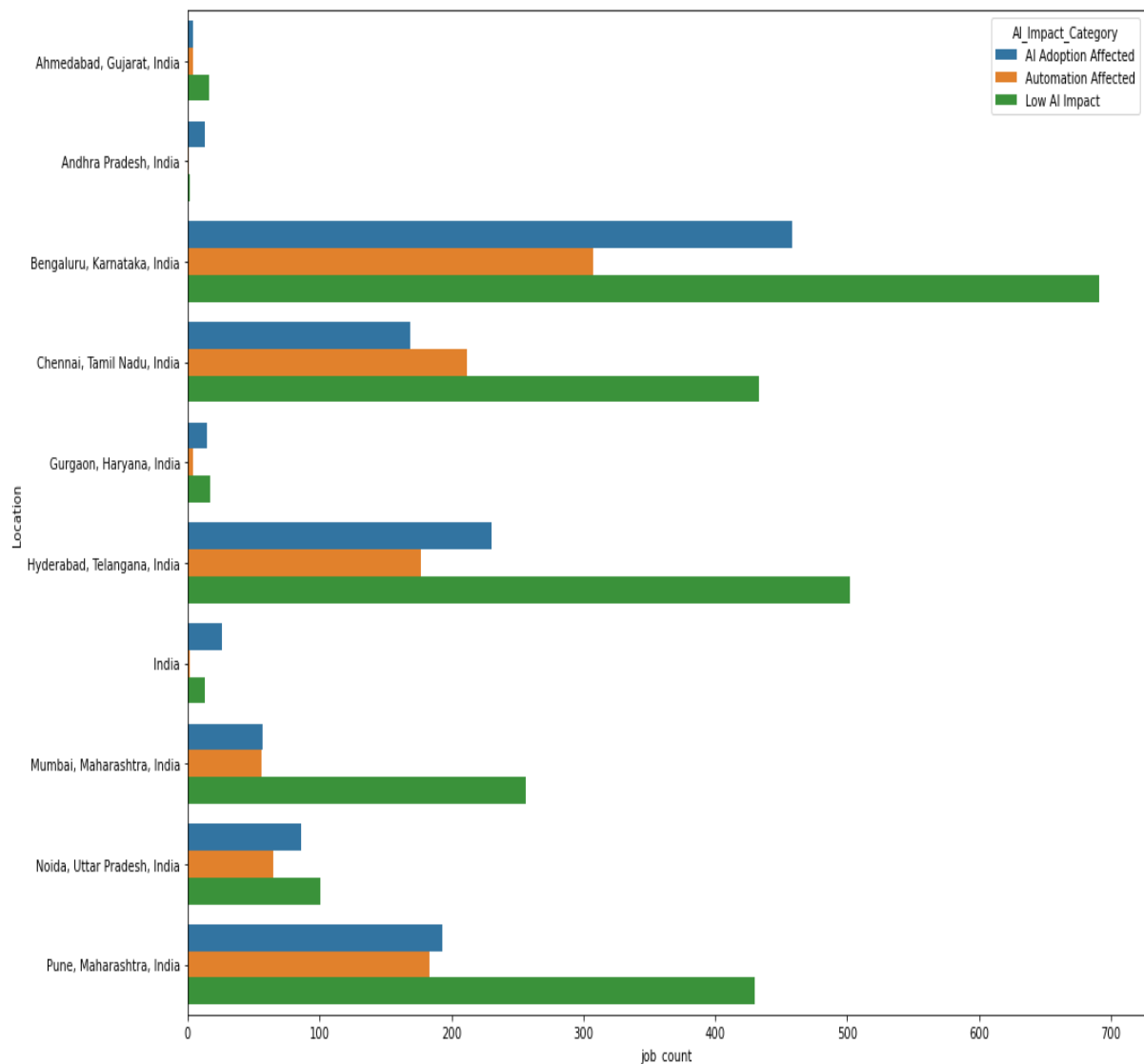

- No of applications for each job role is almost same for non-ai job roles

- No of applications for ai job roles are more from entry level followed by associate and mid-senior level
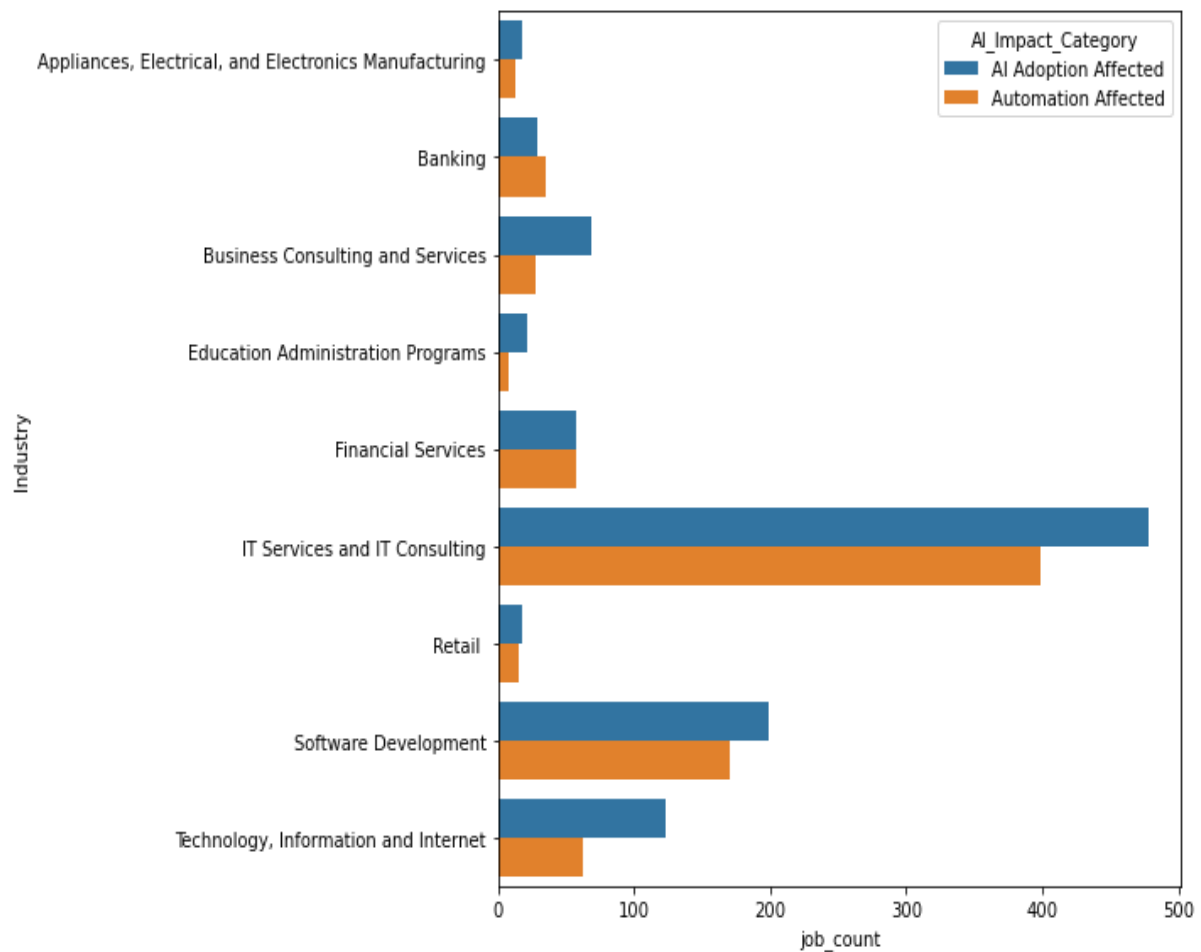

- But if we deeper for ai job roles ai adoption entry level job roles are little more than automation entry level job roles

- In mid-senior level automation jobs were more than ai adoption job roles

- Associate job roles both categories perform nearly same

Are certain locations in India recruiting more AI adoption or automation than low AI impact jobs?

- Andhra Pradesh, Gurugraon and Noida are emerging as **AI-first hiring hubs**, where opportunities in AI adoption and automation significantly **outpace** traditional non-AI roles.

- Bengaluru showcases a dual surge: **more than half** of openings lean toward AI adoption, while nearly **half** are tied to automation compared to non ai roles**, underscoring** the city's role as a national leader in AI-driven transformation.

- In Chennai, the hiring narrative tilts toward automation-heavy roles, which **clearly outnumber** AI adoption positions, signaling a market preference for operational efficiency over strategic AI integration.

- Hyderabad, Mumbai, and Pune stand apart with a **dominance** of non-AI roles, reflecting a comparatively **slower** pivot toward AI adoption and automation compared to other major metros.

Which industries show the highest concentration of AI adoption vs automation risk?

- Most ai roles are from IT Services and IT Consulting followed by Software Development, Technology, Information and Internet
- In retail and Financial services Automation related jobs were almost equal to ai adoption jobs
- In others industries Ai adoption dominated automation related jobs

Does the experience level have no effect on the likelihood of a job being AI-affected?

Null Hypothesis (H₀):

Experience level has no relationship with AI impact; the probability of a job being AI-affected is the same for entry, associate, and senior roles.

Alternative Hypothesis (H₁):

Experience level affects AI impact; entry-level jobs are more likely to be AI-affected than associate or senior roles.

```python
df1=df.groupby(['Experience Level','AI_Impact_Category']).agg(job_counts=('AI_Impact_Category','count')).reset_index()
df1
```

	Experience Level	AI_Impact_Category	job_counts
0	Associate	AI Adoption Affected	284
1	Associate	Automation Affected	267
2	Associate	Low AI Impact	666
3	Entry level	AI Adoption Affected	379
4	Entry level	Automation Affected	248
5	Entry level	Low AI Impact	756
6	Mid-Senior level	AI Adoption Affected	624
7	Mid-Senior level	Automation Affected	500
8	Mid-Senior level	Low AI Impact	1078

```python
data=pd.pivot_table(
    df1,
    values='job_counts',
    index='Experience Level',
    columns='AI_Impact_Category',
    aggfunc='sum'
)
data
```

AI_Impact_Category Experience Level	AI Adoption Affected	Automation Affected	Low AI Impact
Associate	284	267	666
Entry level	379	248	756
Mid-Senior level	624	500	1078

```python
from scipy.stats import chi2_contingency

# Perform chi-square test of independence
chi2, p, dof, expected = chi2_contingency(data)

print("\nChi-square Statistic:", chi2)
print("Degrees of Freedom:", dof)
print("P-value:", p)
print("\nExpected Frequencies:\n", expected)
```

```
Chi-square Statistic: 24.775336732646252
Degrees of Freedom: 4
P-value: 5.5822454772549265e-05

Expected Frequencies:
 [[ 326.17221991  257.23760933  633.59017076]
 [ 370.66243232  292.32507289  720.01249479]
 [ 590.16534777  465.43731778 1146.39733444]]
```

```python
alpha = 0.05
if p < alpha:
    print("\nResult: Reject the null hypothesis.")
    print("Conclusion: Experience level affects AI impact; entry-level jobs are more likely to be AI-affected than associate or m
else:
    print("\nResult: Fail to reject the null hypothesis.")
    print("Conclusion: Experience level has no relationship with AI impact; the probability of a job being AI-affected is the sam
```

```
Result: Reject the null hypothesis.
Conclusion: Experience level affects AI impact; entry-level jobs are more likely to be AI-affected than associate or mid-senior
level roles.
```

7. Recommendations:

Early-Career Professionals & Job Seekers

- Focus on AI tool proficiency and AI-assisted workflows, as entry-level roles face the highest automation risk.

- Build careers around AI-augmented roles (e.g., Data Analytics, Cyber Security) or low AI-impact roles like UI/UX and Product Management.

- Use hands-on projects and real-world datasets to improve competitiveness in AI-related entry-level roles.

Regional Hiring & Talent Strategy

- Leverage Andhra Pradesh, Gurgaon, and Noida as AI-first hiring hubs due to strong AI adoption and automation demand.

- Position Bengaluru as an AI innovation and transformation center, balancing both adoption and automation roles.

- Use Chennai for automation-driven efficiency roles, while maintaining Hyderabad, Mumbai, and Pune for stable non-AI and legacy systems.

Recruiters & Organizations

- Implement Human-in-the-loop training to protect junior roles from automation risks.

- Reskill DevOps and Cloud teams toward AIOps to align with AI-driven infrastructure trends.

- Continuous learning programs should be promoted to help professionals reskill and adapt as AI impact evolves.

Education System & Training Institutions

- Integrate AI, data analytics, and automation concepts into core technical curricula.

- Emphasize project-based and applied learning over theory-heavy instruction.

- Align courses with industry-relevant AI tools and workflows to improve employability.