

Exercise

Exercise Dataset

Data Overview

The “AI-Powered Job Market Insights” dataset provides a snapshot of the modern job market, particularly focusing on the role of artificial intelligence (AI) and automation across various industries.

This dataset includes 500 unique job listings, each characterized by different factors like industry, company size, AI adoption level, automation risk, required skills, and job growth projections.

Data Preview

```
ai_jobs |>
  gt_preview() |>
  tab_header(title = "AI-Powered Job Market Insights") |>
  tab_source_note(source_note = "Source: Kaggle https://www.kaggle.com/datasets/uom190346a/ai-powered-job-market-insights")
```

AI-Powered Job Market Insights

| | job_title | industry | company_size | location | ai_adoption_level | automation_risk | required_skills | salary_remote | job_growth_projection |
|---|------------------------|---------------|--------------|-----------|-------------------|-----------------|-----------------|---------------|-----------------------|
| 1 | Cyber-security Analyst | Entertainment | Small | Dubai | Medium | High | UX/UI Design | 111392.17 | Yes Growth |
| 2 | Marketing Specialist | Technology | Large | Singapore | Medium | High | Marketing | 73792.56 | No Decline |

Source: Kaggle <https://www.kaggle.com/datasets/uom190346a/ai-powered-job-market-insights>

AI-Powered Job Market Insights

| | job_title | industry | company_size | location | ai_adoption_level | automation_risk | required_skills | remote_friendly | job_growth_projection |
|--------|-----------------------|---------------|--------------|-----------|-------------------|-----------------|--------------------|-----------------|-----------------------|
| 3 | AI Researcher | Technology | Large | Singapore | Medium | Low | UX/UI Design | Yes | Growth |
| 4 | Sales Manager | Retail | Small | Berlin | Low | Medium | Project Management | No | Growth |
| 5 | Cybersecurity Analyst | Entertainment | Small | Tokyo | Low | Low | Java Script | Yes | Decline |
| 6..499 | | | | | | | | | |
| 500 | HR Manager | Entertainment | Medium | Berlin | Medium | High | Project Management | Yes | Decline |

Source: Kaggle <https://www.kaggle.com/datasets/uom190346a/ai-powered-job-market-insights>

To simplify our later code, I have created a separate table which is already filtered for the groups we will be looking at:

```
ai_jobs_risk <- ai_jobs |>
  filter(automation_risk %in% c("Low", "High"))

ai_jobs_high <- ai_jobs_risk |>
  filter(automation_risk == "High")
```

```
ai_jobs_low <- ai_jobs_risk |>
  filter(automation_risk == "Low")
```

Dataset Features:

Source: Kaggle <https://www.kaggle.com/datasets/uom190346a/ai-powered-job-market-insights>

1. **Job_Title:**
 - **Description:** The title of the job role.
 - **Type:** Categorical
 - **Example Values:** “Data Scientist”, “Software Engineer”, “HR Manager”
2. **Industry:**
 - **Description:** The industry in which the job is located.
 - **Type:** Categorical
 - **Example Values:** “Technology”, “Healthcare”, “Finance”
3. **Company_Size:**
 - **Description:** The size of the company offering the job.
 - **Type:** Ordinal
 - **Categories:** “Small”, “Medium”, “Large”
4. **Location:**
 - **Description:** The geographic location of the job.
 - **Type:** Categorical
 - **Example Values:** “New York”, “San Francisco”, “London”
5. **AI_Adoption_Level:**
 - **Description:** The extent to which the company has adopted AI in its operations.
 - **Type:** Ordinal
 - **Categories:** “Low”, “Medium”, “High”
6. **Automation_Risk:**
 - **Description:** The estimated risk that the job could be automated within the next 10 years.
 - **Type:** Ordinal
 - **Categories:** “Low”, “Medium”, “High”
7. **Required_Skills:**
 - **Description:** The key skills required for the job role.
 - **Type:** Categorical
 - **Example Values:** “Python”, “Data Analysis”, “Project Management”
8. **Salary_USD:**
 - **Description:** The annual salary offered for the job in USD.
 - **Type:** Numerical
 - **Value Range:** \$30,000 - \$200,000
9. **Remote_Friendly:**
 - **Description:** Indicates whether the job can be performed remotely.
 - **Type:** Categorical

- **Categories:** “Yes”, “No”

10. **Job_Growth_Projection:**

- **Description:** The projected growth or decline of the job role over the next five years.
- **Type:** Categorical
- **Categories:** “Decline”, “Stable”, “Growth”

Exercise - Mystery bags

To begin, split into three groups. Decide on the following roles, one per person:

1. **Sampler** - draws samples from the population
2. **Recording** - records the draws in R

Your group has been presented with a population model (bag containing salaries). There may be multiple populations represented in the various bags around the room.

Please DO NOT look in the bag OR empty the contents of the bag.

Exercise - Mystery bags {nonincremental}

The task

1. The **Sampler** draws 10 slips from the bag - this is your sample of size 5.
2. The **Recorder** records the values on the slips into the `sampling-exercise.R` file.
 - Create a new list for each sample (`sample_1 <- c(1, 2, 3, 2, 2)`)
3. Calculate the mean for the new sample and add it to the `sample_means` table.
4. Run the code to plot the histogram of your samples.
5. Return the slips to the bag.
6. Ensure the data is well mixed between samples.
7. Repeat

Bibliography