AI ASSIGNMENT-17.4

Task 1 – Employee Data Preprocessing :  
Task:  
Use AI to generate a Python script for cleaning an employee dataset.  
Instructions:  
• Handle missing values in columns (salary, department,  
joining\_date).  
• Convert the "joining\_date" column into proper datetime format.  
• Standardize department names (e.g., "HR", "hr", "Human  
Resources" → "HR").  
• Encode categorical variables (department, job\_role).  
Expected Output:  
• A cleaned Pandas DataFrame with consistent departments, proper  
dates, and encoded features.

**Prompt:**   
Write a Python script using Pandas to clean an employee dataset by:

* Handling missing values in **salary,department** and **joining\_date**.
* Converting **joining\_date** to datetime format.
* Standardizing department names (e.g., "HR", "hr", "Human Resources" → "HR").
* Encoding categorical variables (**department**, **job\_role**).  
  Output a cleaned DataFrame with proper dates, consistent departments, and encoded features.

Code:

A computer code with many letters and numbers

AI-generated content may be incorrect.

A screenshot of a computer code

AI-generated content may be incorrect.

Output:

A screenshot of a computer

AI-generated content may be incorrect.

Task 2 – Sales Transaction Data Preprocessing  
Task:  
Use AI to generate a script for preprocessing a sales transaction dataset.  
Instructions:  
• Convert transaction dates to proper datetime format.  
• Create a new column for “Month-Year” from the transaction date.  
• Remove rows with negative or zero transaction amounts.  
• Normalize the "transaction\_amount" column using Min-Max  
scaling.  
Expected Output:  
• A preprocessed DataFrame with valid dates, normalized amounts,  
and no invalid records

**Prompt:**  
Write a Python script using Pandas and scikit-learn to preprocess a sales dataset by converting **transaction\_date** to datetime, adding a **Month-Year** column, removing negative or zero **transaction\_amount** values, and normalizing **transaction\_amount** with Min-Max scaling. Output the cleaned DataFrame.

CODE:

A screenshot of a computer

AI-generated content may be incorrect.

OUTPUT:

A screenshot of a computer

AI-generated content may be incorrect.

Task 3 – Healthcare Patient Records Cleaning  
Task:  
Use AI to generate a script for cleaning healthcare patient records.  
Instructions:  
• Fill missing values in numeric columns (e.g., blood\_pressure,  
heart\_rate) with column mean.  
• Standardize units (convert height from cm to meters).  
• Correct inconsistent categorical labels (e.g., "M", "Male", "male"

→ "Male").  
• Drop irrelevant columns such as patient\_id after cleaning.  
Expected Output:  
• A cleaned healthcare dataset suitable for ML model training

PROMPT:

**Prompt :**  
Write a Python script using Pandas to clean healthcare patient records by filling missing numeric values with the mean, converting height from cm to meters, standardizing gender labels (e.g., "M", "male" → "Male"), and dropping irrelevant columns like patient\_id. Output the cleaned DataFrame.

CODE:

A computer screen shot of a computer code

AI-generated content may be incorrect.

OUTPUT:

A screenshot of a computer

AI-generated content may be incorrect.

Task 4 – Social Media Sentiment Dataset Preparation  
Task:  
Use AI to write a script to preprocess a social media text dataset.  
Instructions:  
• Remove special characters, URLs, and emojis from text.  
• Convert all text to lowercase.  
• Tokenize and remove stopwords.  
• Apply lemmatization for standardizing words.  
Expected Output:  
• A processed dataset with clean text, ready for NLP sentiment  
analysis

**Prompt :**  
Write a Python script using NLTK or spaCy to preprocess a social media text dataset by removing special characters, URLs, and emojis, converting text to lowercase, tokenizing and removing stopwords, and applying lemmatization. Output the cleaned text dataset ready for sentiment analysis.

CODE:

A computer screen shot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

OUTPUT:

A screenshot of a computer

AI-generated content may be incorrect.

Task 5 – Financial Dataset Feature Engineering  
Task:  
Use AI to create a preprocessing script for a financial dataset.  
Instructions:  
• Handle missing values in stock price and volume.  
• Create new features such as moving average (7-day, 30-day).  
• Normalize continuous variables using StandardScaler.  
• Encode categorical columns (sector, company\_name).  
Expected Output:  
• A feature-engineered DataFrame with new indicators and  
normalized values for ML tasks.  
 Deliverables (For All Tasks)  
1. AI-generated prompts for code and test case generation.  
2. At least 3 assert test cases for each task.  
3. AI-generated initial code and execution screenshots.  
4. Analysis of whether code passes all tests.  
5. Improved final version with inline comments and explanation.  
6. Compiled report (Word/PDF) with prompts, test cases, assertions,  
code, and output.

**Prompt:**  
Write a Python script to preprocess a financial dataset by handling missing values, creating 7-day and 30-day moving averages, normalizing with StandardScaler, and encoding sector and company\_name. Output the feature-engineered DataFrame.

CODE:

A screenshot of a computer

AI-generated content may be incorrect.A screenshot of a computer code

AI-generated content may be incorrect.A close up of a computer screen

AI-generated content may be incorrect.

OUTPUT:

A screenshot of a computer

AI-generated content may be incorrect.