Assignment No. 2

Algorithmic Tools Analysis for Al-Beruni City

Al-Beruni City (ABC) has emerged as a regional hub of education, technology, and innovation. Over the past fifteen years, the government of ABC has made substantial investments in developing algorithmic tools and implementing Artificial Intelligence (AI) across key public sectors such as education, health, and public safety. These departments have been leveraging AI and algorithmic systems to enhance service delivery and operational efficiency.

However, during the latest annual performance review of public services, the ABC government noted a decline in the performance of several departments. A deeper investigation revealed that these departments were not effectively utilizing the available technology-based tools. Issues such as poor data quality, redundancy, and underutilization of insights generated from AI systems were identified as major contributing factors.

In response, the ABC government engaged Data Sighted (Private) Limited (DSL), a leading data and technology consulting firm, where you serve as a lead data analyst. To facilitate the analysis, the government shared two CSV files that document the use of algorithmic tools across various departments. A detailed description of the fields in these files is provided in **ANNEXURE-A**.

Your team has been tasked with analyzing the provided data to uncover actionable insights and suggest improvements to enhance the effectiveness of AI tool usage in public services.

To assess your foundational skills in using Python for data manipulation and analysis tasks, this assignment focuses on a set of predefined tasks outlined in the **Required** section. These tasks are designed to evaluate your ability to work with real-world data and derive actionable insights that can support strategic decision-making.

ANNEXURE-A

Following files have been provided to you:

- Departments.csv and Tools.csv that contains the data provided by the government of ABC to DS.
- Response.ipynb which has been provided to you for writing your answers to this assignment.

Description of each field (column name) of the CSV files is as follows:

Departments.csv

Column Name	Description
Department	Name of department
Abb	Abbreviation of the department's name

Tools.csv

Column Name	Description
Abb	Abbreviation of the department's name
Tool	Name of tool
Tool_desc	Brief description of each tool
Date	Date on which the tool was used first time by the department
Output	Brief description of the output of the tool
Analysis	Type of analysis made by the tool
Population	Type of population used by the tool
Updated	States whether the tool was ever updated

REQUIRED:
Perform the following tasks in the Jupyter Notebook file (Response.ipynb) provided to you:

(a) Load desired libraries and dataset

(i) Import necessary library packages.
(ii) Load the data from the Departments.CSV files to df1 and from Tools.CSV to df2 in Response.ipynb.
(iii) Display the first row of each data frame (df1 and df2) to verify that data from both files has been loaded successfully.

(b) Merging the two data frames and filling the missing values

(i) Run an appropriate command to merge df1 and df2 such that all rows of df2 must be selected in the merged data frame.
(ii) Run an appropriate command to display the total number of missing values in each field.
(iii) Run an appropriate command to create a mapping between Department and Abb.
(iv) Run an appropriate command to fill the missing values of Department. Run an appropriate command to confirm that there are no missing values in any field.
(v) Run an appropriate command to save the updated data frame as a CSV file.

(c) Remove duplicate values

(i) Using the 'Abb' and 'Tool' fields, run an appropriate command to remove duplicate values from the merged data frame.
(ii) Run an appropriate command to print the shape of original merged data frame and the shape of merged data frame after removing the duplicates.
(iii) Save the updated data frame (after removing the duplicates) as a CSV file. Rename this file to your CRN (CRN-Unique.CSV) and upload it with your answer files.

(d) Data Analysis
Consider the data frame obtained after removing the duplicates.

(i) Run an appropriate command to find the department that has the highest variety of 'Analysis' types.
(ii) Run an appropriate command to show the percentage of updating of each tool.

(e) Data Visualization
Consider the data frame obtained after removing the duplicates.

(i) Create a bar chart showing the count of tools per department.
(ii) Generate a pie chart for Analysis column.
(iii) Create a bar plot to show the number of tools that are marked as "Updated".

(THE END)