AYTS 5001 – Week 6 Lab Sheet

Title: Exploring and Analysing Data with Pandas

Duration: 2 hours

Marks: Not graded

Tools: Python, Pandas

Learning Outcomes

- Load a CSV file into a Pandas DataFrame.
- Explore, clean, and summarise data.
- Perform basic filtering and grouping operations.
- Create calculated columns.
- Save processed data back to CSV.

Scenario

The Bridgeburners have been tracking their squad members' performance across recent campaigns. Your task is to load their data, explore patterns, calculate performance metrics, and export a cleaned dataset for further analysis.

Dataset: malazan_bridgeburners.csv

Lab Tasks

Task 1 – Load and Inspect the Data

• Import pandas as pd.
• Load the CSV file into a DataFrame.
• Display the first 5 rows.
• Show column names and data types using .info().
• Print the number of rows and columns using .shape.

Hint: Use pd.read_csv().

Task 2 – Summary Statistics

• Display summary statistics using .describe().
• Find the mean of Battles_Fought.
• Find the median of Avg_Damage.
• Find the maximum Healing_Spells.

Hint: Use df['column'].mean() and similar methods.

Task 3 – Filtering and Selection

- Filter to find warriors with Battles_Fought > 7 and Avg_Damage > 85.
- Create a new DataFrame of these elite fighters and display it.
- Sort the result by Avg_Damage in descending order.

Hint: Use boolean conditions and .sort_values().

Task 4 – Create a Calculated Column

- Add a new column called Battle_Efficiency defined as:
  (Avg_Damage × Battles_Fought) / (Injuries_Sustained + 1)
- Display the top 5 characters by Battle_Efficiency.

Hint: Use vectorised operations (df['new'] = ...).

Task 5 – Grouping and Saving Results

- Group data by Role and calculate the average Battle_Efficiency.
- Save the grouped summary to a new file bridgeburners_summary.csv.

Hint: Use df.groupby('Role')['Battle_Efficiency'].mean() and .to_csv().

Deliverables

Submit:

1. Lab6_pandas.py (or .ipynb)
2. bridgeburners_summary.csv
3. Screenshot of output tables