

OOPY1IL: Object-Oriented Programing With Python

Hannah Janout, MSc.

December 15, 2025

This exercise sheet outlines the dataset description and tasks for the final assignments in the course OOPY1IL. The tasks are to be completed collaboratively in groups of 3 to 5 students, assigned in the lecture. The results of your work will be submitted in the form of a code base, a written report, and a presentation. Your final grade will be based on:

- The quality and clarity of your presentation.
- Your ability to explain your workflows and results, and answer follow-up questions.
- The thoroughness and cleanliness of your code (incl. documentation/commenting) and report.
- The completeness of the assigned tasks.

All group members will receive the same final grade for the project. However, if a member demonstrates minimal or no participation, their individual grade will be adjusted negatively. Such cases will not affect the grades of the other group members.

The final presentation will be 25 minutes long, 15 minutes for the presentation and 10 minutes for questions and discussion.

The use of an AI assistant is prohibited. If there is any indication that your work has been copied from other students or online sources (including LLMs, CoPilot, and similar AI tools), a clarification meeting will be scheduled. During this meeting, you will be expected to demonstrate a clear understanding of your implementation and justify your approach. **If you are unable to address concerns regarding the authorship of your submission convincingly, all group members will lose credit for the entire exercise.**

1 Guidelines

Please submit your code base, report, and presentation in a single ZIP file and upload it to eLearning for grading. Your project will only be graded if all components are submitted.

Important: The submission of this zip file (in addition to holding your presentation) is mandatory to achieve your final grade.

1.1 Code Base

- Use Python 3.10 or newer.
- You may work in:
 - A Python file (.py) or a Jupyter Notebook (.ipynb).
 - Any IDE (PyCharm, VS Code, etc.).

- At the top of your code, include:
 - Python version used.
 - Group members.
 - Project name (Personality, Lung, AI Assistant, or Video Games).
 - A short description of each member's contribution.
- Comment and structure your code clearly for readability. Code quality and organization will affect your grade.
- Ensure your code runs without errors and includes:
 - Proper imports.
 - Clear variable names.
 - Logical grouping of tasks (e.g., preprocessing, analysis, visualization).

1.2 Report

- Submit as a PDF in the same folder as your code base and presentation.
- Include the same header information as in your code:
 - Python version, group members, project name, contributions.
- Your report should:
 - Be well-structured and grammatically correct.
 - Explain your work so that someone unfamiliar with the dataset can understand it.
- **Report structure:**
 - **Introduction:** Describe the dataset (origin, columns, meaning) and state the tasks. Also state possible assumptions or constraints, if present.
 - **Methodology:** Explain steps taken, packages used and why, data loading and preprocessing.
 - **Results:** Present answers with explanations, plots (with captions describing the result!), and comments.
 - **Summary:** Summarize your work:
 - * Number of tasks completed.
 - * Insights and challenges (discovered patterns, encountered difficulties).
 - * Surprising findings (unexpected results and possible reasons).
 - * Lessons learned (what you would do differently next time).

How to Write Your Report

Do **not** just copy-paste code! Instead, explain your approach and results clearly.; therefore, beginmple:

Task: What was the average temperature per month in 2024?

Solution: For this task, we need the daily temperature available in column `temp` and the corresponding month saved in `recorded_month`. Through grouping the DataFrame based on the months, we can calculate the average temperature using the `mean()` function.

Result: Using this approach, the average temperature for each month in 2024 is as follows:

...

Visualization: (Insert a plot showing monthly averages.)

This is the level of detail expected for every task:

- State the task clearly.

- Explain your approach and reasoning.
- Show the result (numbers or summary).
- Include a visualization if relevant.

1.3 Presentation

- Duration: 15 minutes presentation + 10 minutes Q&A.
- All group members must present part of the work.
- Assume your audience is unfamiliar with the dataset, so start with an introduction.
- Suggested structure:
 - **Introduction:** Dataset description and origin, tasks/questions.
 - **Methodology:** Packages used, data loading, and preprocessing steps.
 - **Results:** Show answers and visualizations clearly.
 - **Summary:**
 - * Number of tasks completed.
 - * Insights and challenges (what worked well, what was hard).
 - * Surprising findings (unexpected patterns or correlations).
 - * Lessons learned (what you would improve next time).

2 Daily AI Assistant Usage Behavior

The Daily AI Assistant Usage Behavior Dataset captures real-world interaction patterns between users and AI assistants throughout their day. It includes details such as query types, time-of-day usage, session duration, device type, user intent, and follow-up behavior.

This dataset is designed to support researchers, developers, and data enthusiasts in analyzing how people rely on AI tools for productivity, creativity, learning, and routine tasks. It is ideal for building models for user behavior prediction, recommendation systems, personalization, and conversational AI.

Size: The dataset contains 300 rows and 8 columns.

Source Link: <https://www.kaggle.com/datasets/prince7489/daily-ai-assistant-usage-behavior-dataset>

Tasks

Dataset Overview

Provide a summary of the dataset:

- Detailed description in the form of an introductory text:
 - What is the data about?
 - Where does it come from?
 - ...
- Number of rows and columns.
- List of features with their meaning and data types.
- Any initial observations about the data.

Data Quality Check

- Investigate the dataset for:
 - Missing values.
 - Outliers.
 - Unusual or inconsistent entries that may require attention later.

Data Preprocessing

- Handle missing values and outliers appropriately.
- Create a new column **timeOfDay**, which bins the **timestamp** into morning, afternoon, evening, and night.
- Create a new column **year**, which is derived from the **timestamp** column.
- Convert data types where needed (e.g., categorical columns).

Data Analysis

Answer the following questions and visualize where appropriate:

- How many different AI assistants are in the dataset? Show counts and percentages.
- What is the average session length per assistant? Are there noticeable differences?
- Are specific assistants used more often for certain tasks (e.g., education)?
 - If so, why do you think specific devices are used more often for these tasks?

- Which type of tasks has the longest average prompt size and use time?
- Is there a pattern between the task type and the time of day it was used?
- Have some AI assistants become more popular over time?
 - Analyze their occurrence for each year.
 - Plot their average number of usage per year as a scatter plot.
- Visualizations:
 - Plot distributions of key features using histograms, KDE plots, and boxplots.
 - Use color to distinguish individual assistants.
- Are there any features that clearly differentiate device types?
- Correlation Analysis:
 - Create a correlation heatmap for all features.
 - Which features are strongly correlated with each other?
 - Are any features highly correlated with assistant models and session length?
- Use pair plots to explore feature relationships, color-coded by the assistant model.
- Grouped by Assistant Model, calculate mean, median, and standard deviation for each feature. What patterns do you observe?
- Time-Based Analysis:
 - How does session length vary by day or time of day?
 - Are there peak usage times for certain assistants?
 - Create a line plot or a heatmap to show usage trends over time.

Additional Insights

Add **five** custom analyses that you believe provide valuable insights.

Presentation

Document your findings clearly and visually, and make them easily understandable to readers with no prior knowledge of the dataset or data analysis.