

1. Introduction

- Details: Open the Excel file 'quarterly_sales.xlsx', calculate the total and average sales, then create a new PowerPoint slide name 'Sales_Report.pptx' in Documents summarizing the result in bullet points.
- Task completion conditions: A file named 'Sales_Report.pptx' with correct content is created in the right path.
- Difficulty: Medium.

2. Result

- Number of attempts until success: 2
- Time taken:

Step	Time
1	464.71
2	41.03
3	55.09
4	61.76
5	59.50
6	57.09

- Total: 739.18s (~12p19s)
- Average: 123.20s/step

- Token cost:
GPT-5:

Step	Prompt	Completion	Total	Reasoning
1	54570	15847	70417	9984
2	9717	756	10473	576
3	11863	689	12552	512
4	14013	894	14907	704
5	16170	406	16576	256
6	18282	674	18956	448

UI-Tars 7B:

Step	Prompt	Completion	Total
1	0	0	0
2	0	0	0
3	5498	24	5522
4	5498	24	5522

5	0	0	0
6	0	0	0

3. Behavior

a) Trajectory:

- The agent begins with verifying that no prior actions were taken. It analyzed the environment, noting that VS Code was open, and initiated a code agent to handle the task: locate quarterly_sales.xlsx, calculate total and average sales, and create Sales_Report.pptx in the Documents folder. The first coding sequence performed file searches, data extraction with openpyxl, and PowerPoint generation with total and average figures, verifying results by reopening the file. The second sequence repeated similar steps, but fewer steps are needed for file detection. Afterward, the model prepared a plan for GUI-based verification.
- Next, the model verified through screenshots that the PPTX file was created but not yet visually confirmed, so it planned to open File Explorer with Windows + E to check the Documents directory. It identified “Documents” in the sidebar and navigated there to verify the file’s existence. Upon confirmation, the model noted that Sales_Report.pptx was visible in Documents and planned to open it to check the content.
- Finally, after double-clicking the file, PowerPoint successfully launched. The model waited for it to open fully, then verified that the presentation contained one slide titled “Quarterly Sales Summary” with bullet points showing total and average sales, number of quarters, and a data source reference. It concluded that the task was successfully completed and the output met all specified requirements.

b) Problem:

- I tried twice but the agent always called the Code Agent two times.