Python-In-Excel

xlwings Lite - The AI Way
Getting Started

Let AI handle Python
You handle logic & outcomes

A Practitioner's Guide

Module - 02

Data Cleaning & Rule Based Transformations

Mobile Number Cleanup Example

PLUS: Validations - My Non-Negotiable Rules

Validations : The Hard Won Lessons

Across my career, I have seen analysts lose jobs, companies lose money and face lawsuits, just because the analyst skipped validations due to a deadline.
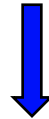
AI did the work? No- you did. AI helped. Inspect what you expect

These rules apply to everything
Machine learning models, analysis, or campaign files

1 If it's not validated, it does not go out. Period.

2 No matter the pressure. It is not worth it.

3 Deadlines can be moved. Validations are sacrosanct

4 Any validation not documented does not exist

5 Always use 2 or more methods. Never ever one

6 Use AI assisted validations and critical reviews

7 AI is not perfect - Human review is a must

The Problem: Messy Data

Common problem, right?

	A	B
1	**Name**	**Phone Num**
2	A	9845230763
3	B	919845230763
4	C	984523076
5	D	+91-919845230763
6	E	91919845230763
7	F	91 919845230763
8	G	91 919845230763,9845230764
9	H	984523076,9845230763
10	I	984523076,98452307
11	J	my phone numer is 9845230763
12	K	amar@harolikar.com, 9845230763
13	L	amar@harolikar.com, 984523076
14	M	nopes
15	N	sadf sdf
16	O	8644
17	P	llka sf(**&&) sdf

The Concept

Use AI to generate rules
for data cleaning & transformations

Validations are sacrosanct
AI or no AI

Run regular 'human' validations

Run AI assisted validations

Iterate

Real business – doesn't get done in a single instruction. Iterate

1. Download Module-02 Kit

xlwings Lite: Practice Lab

Learn xlwings Lite by practical examples

Help me build what you need   ✨ Suggest Example   🐞 Report Issue

Download AI Coder Instructions

For AI/LLM Upload

Download the complete markdown file to enhance your AI assistant's xlwings Lite knowledge

⬇ Download Markdown (1,450 lines)

Practice Workbooks and Guide

Guide is this same document

Module-02: Practical Data Cleaning

A step-by-step guide for using AI to generate, test, and validate a robust data cleaning script.

🤖 **AI-Driven Rule Generation**

A process for having the AI analyze data patterns, generate initial cleaning rules, and identify potential edge cases and exceptions.

🧪 **AI Test Case Creation**

Using AI to generate outlier scenarios and difficult edge cases to stress-test the robustness of your cleaning script before deployment.

📄 **Auditing AI**

Reviewing AI's proposed logic, identifyig what it missed and validating the results.

⬇ Practice Workbook   ⬇ Guide PDF   ⬇ Completed Workbook (WithCode)

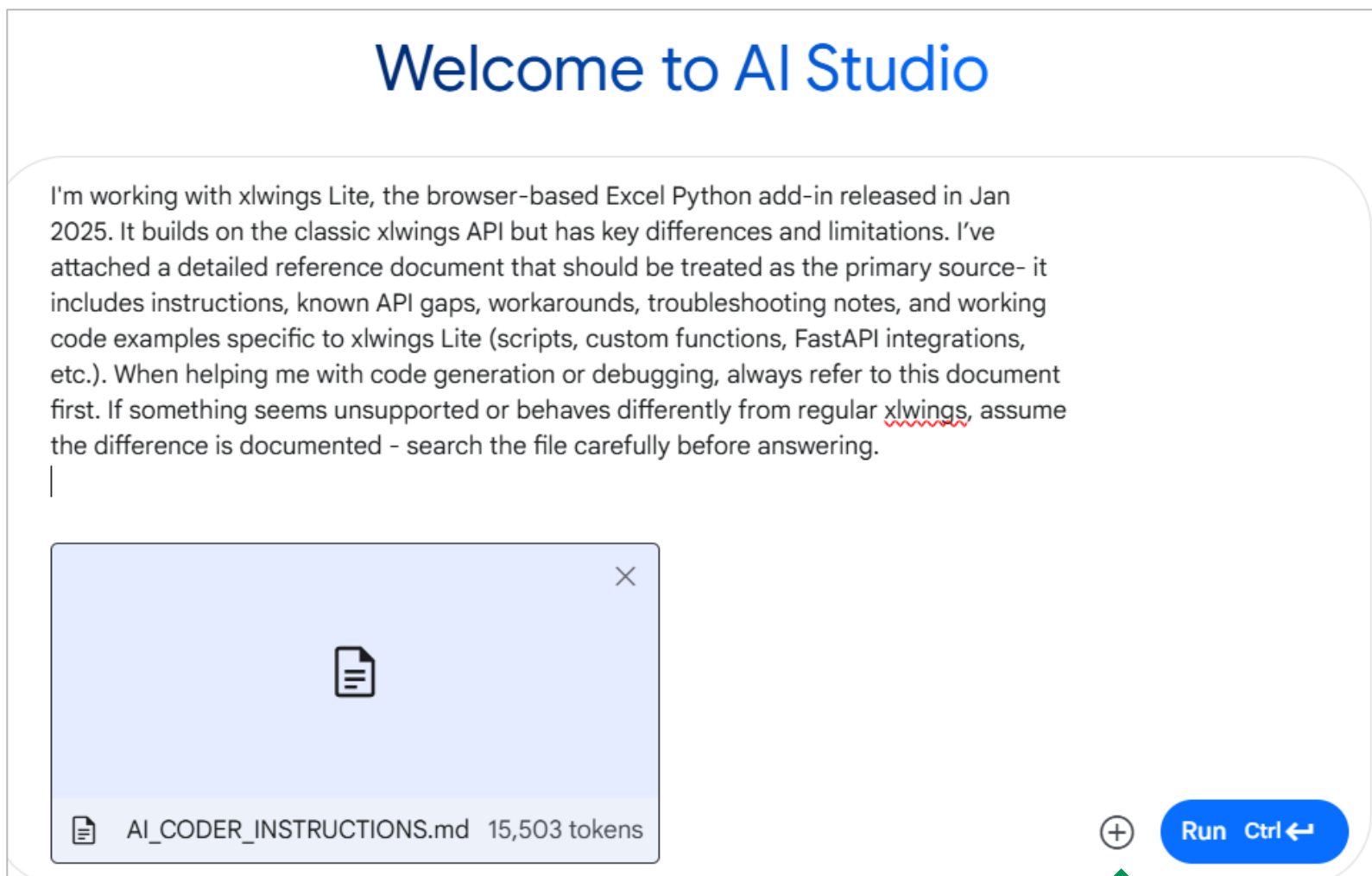# 2. Setup Google AI Studio

aistudio.google.com

Welcome to AI Studio

I'm working with xlwings Lite, the browser-based Excel Python add-in released in Jan 2025. It builds on the classic xlwings API but has key differences and limitations. I've attached a detailed reference document that should be treated as the primary source- it includes instructions, known API gaps, workarounds, troubleshooting notes, and working code examples specific to xlwings Lite (scripts, custom functions, FastAPI integrations, etc.). When helping me with code generation or debugging, always refer to this document first. If something seems unsupported or behaves differently from regular xlwings, assume the difference is documented - search the file carefully before answering.

×

📄

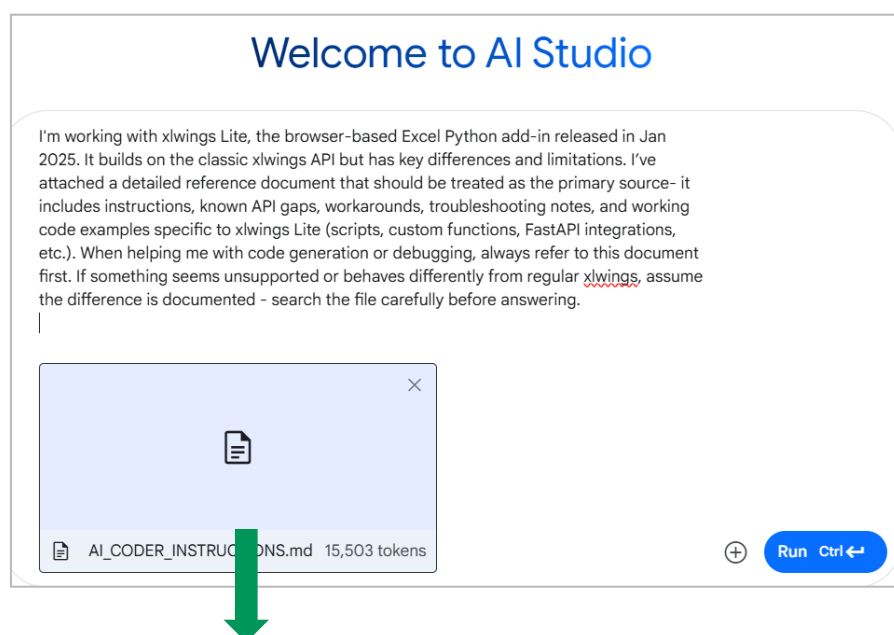📄 AI_CODER_INSTRUCTIONS.md 15,503 tokens   ⊕   **Run Ctrl ↵**

① Upload AI Coder Instructions File

② Enter Instruction -01 : Base Instruction

③ Hit Run

What's that about?

It's about providing AI the context to do its job

When AI has context = you get results

When AI lacks context = you get fluff

Welcome to AI Studio

I'm working with xlwings Lite, the browser-based Excel Python add-in released in Jan 2025. It builds on the classic xlwings API but has key differences and limitations. I've attached a detailed reference document that should be treated as the primary source- it includes instructions, known API gaps, workarounds, troubleshooting notes, and working code examples specific to xlwings Lite (scripts, custom functions, FastAPI integrations, etc.). When helping me with code generation or debugging, always refer to this document first. If something seems unsupported or behaves differently from regular xlwings, assume the difference is documented - search the file carefully before answering.

AI_CODER_INSTRUCTIONS.md   15,503 tokens   Run Ctrl↵

AI Coder Instructions File

- This is my personal, 1,450-line rulebook for xlwings Lite that I feed to the AI right at the start
- Covers tool limitations, non-negotiable rules, error patterns, and best practices from my client projects
- Helps in delivering clean, reliable xlwings Lite code - reducing debugging and iterations.

3. Input Instruction-02 & Hit Run

❌ **NO**

Give me code to clean phone number data

✅ **YES**

Context is all AI needs

See attached few sample rows. The data is formatted as a table named PHONES.

Give me code to clean up the phone numbers.
- I want a single 10-digit Indian mobile number as output.

- In case of invalid input, return 0.

- In case of multiple numbers, use the first valid one.

- This is for Indian mobile numbers — make sure the number starts with 6, 7, 8, or 9.

Also:
- Share the pseudo logic you used for the cleanup.

- Convert the field to string first.

- Use Regex where appropriate.

Name Phone Num A 9845230763 B 919845230763 C 984523076 D +91-919845230763 E 91919845230763 F 91 919845230763 G 91 919845230763,9845230764 H 984523076,9845230763 I

➡ **Where is the data?**

➡ **What is the outcome you want**

➡ **What to do in different scenarios . e.g. invalids or multiple entries**

➡ **Provide background plus any helpful tips**

➡ **Ask for pseudo logic**

➡ **Tell how you want it done**

➡ **Show what you see Share sample rows / images**

4. Validate Pseudo Logic Shared by AI

- Don't take AI output at face value
- It talks with confidence - but misses occasionally
- Human review necessary

What did AI miss?

Pseudo-Logic for Phone Number Cleaning

This logic is designed to be robust and handle the varied formats in your sample data.

1. **Input Standardization:** First, convert the content of each cell in the 'Phone Num' column into a plain text string to ensure consistent processing.
2. **Extraction:** Use a Regular Expression (regex) to find all separate sequences of digits within the text. This correctly handles cases where numbers are mixed with text, symbols, or spaces (e.g., "my phone is 9845230763" or "+91-9845...").
3. **Iteration & Validation:**
   - Go through the extracted digit sequences one by one, from left to right.
   - For each sequence, check if it is at least 10 digits long. This immediately filters out invalid short codes.
   - If it is long enough, take the **last 10 digits** of that sequence. This effectively strips any leading country codes like 91 .
   - Check if this 10-digit number is a valid Indian mobile number by verifying if its first digit is 6 , 7 , 8 , or 9 .
4. **First Match Wins:** The very first number that passes this validation is selected as the clean phone number for that row, and the process stops for that cell.
5. **Default Value:** If, after checking all extracted digit sequences in a cell, no valid 10-digit mobile number is found, the script assigns a default value of '0' .

Key Principle: Have AI modify logic based on your 'human' review and real world experiences
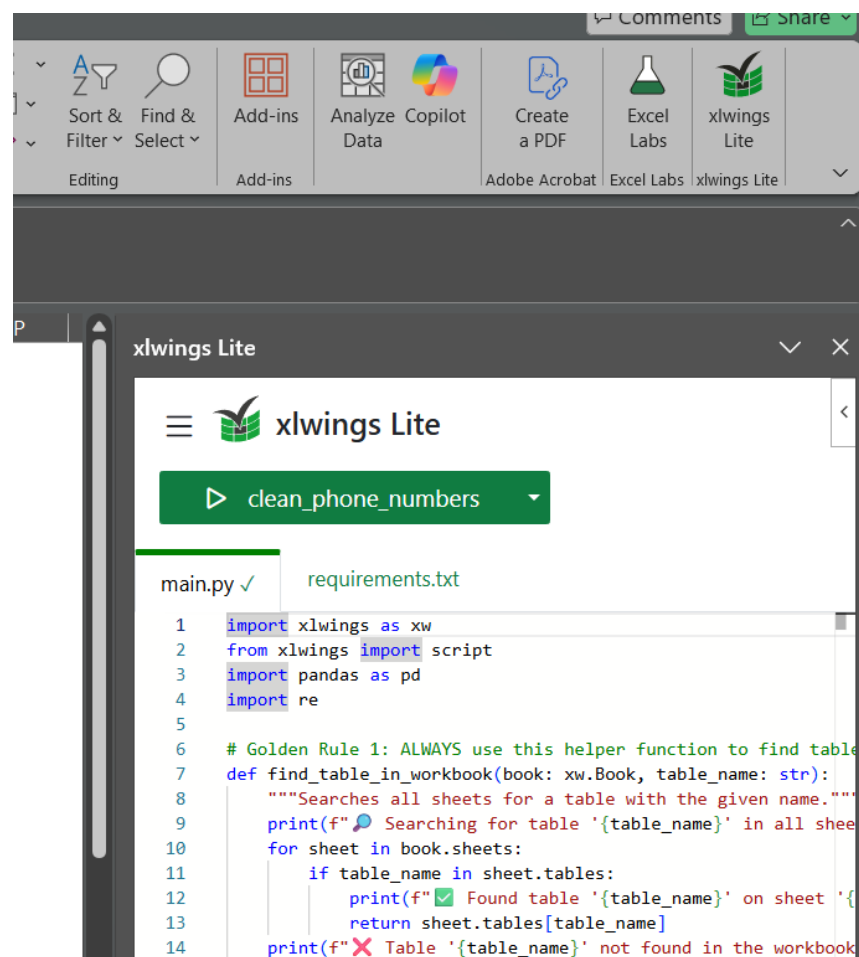
5. Run Code

xlwings Lite Python Script

You can paste this code directly into the `main.py` tab of the xlwings Lite task pane. It follows all the required coding guidelines.

Generated python

```python
import xlwings as xw
from xlwings import script
import pandas as pd
import re


# Golden Rule 1: ALWAYS use this helper functi    to find tables.
def find table in workbook(book: xw Book  tabl  name: str):
```


Want to understand the code? Have AI walk you through each code element, step-by-step.

No Black Boxes Allowed

Ask. Interrogate. Challenge AI

Why this way, and not the other way?

If you can't explain the logic, you can't trust the output

AI is a Tool. Not Your Friend

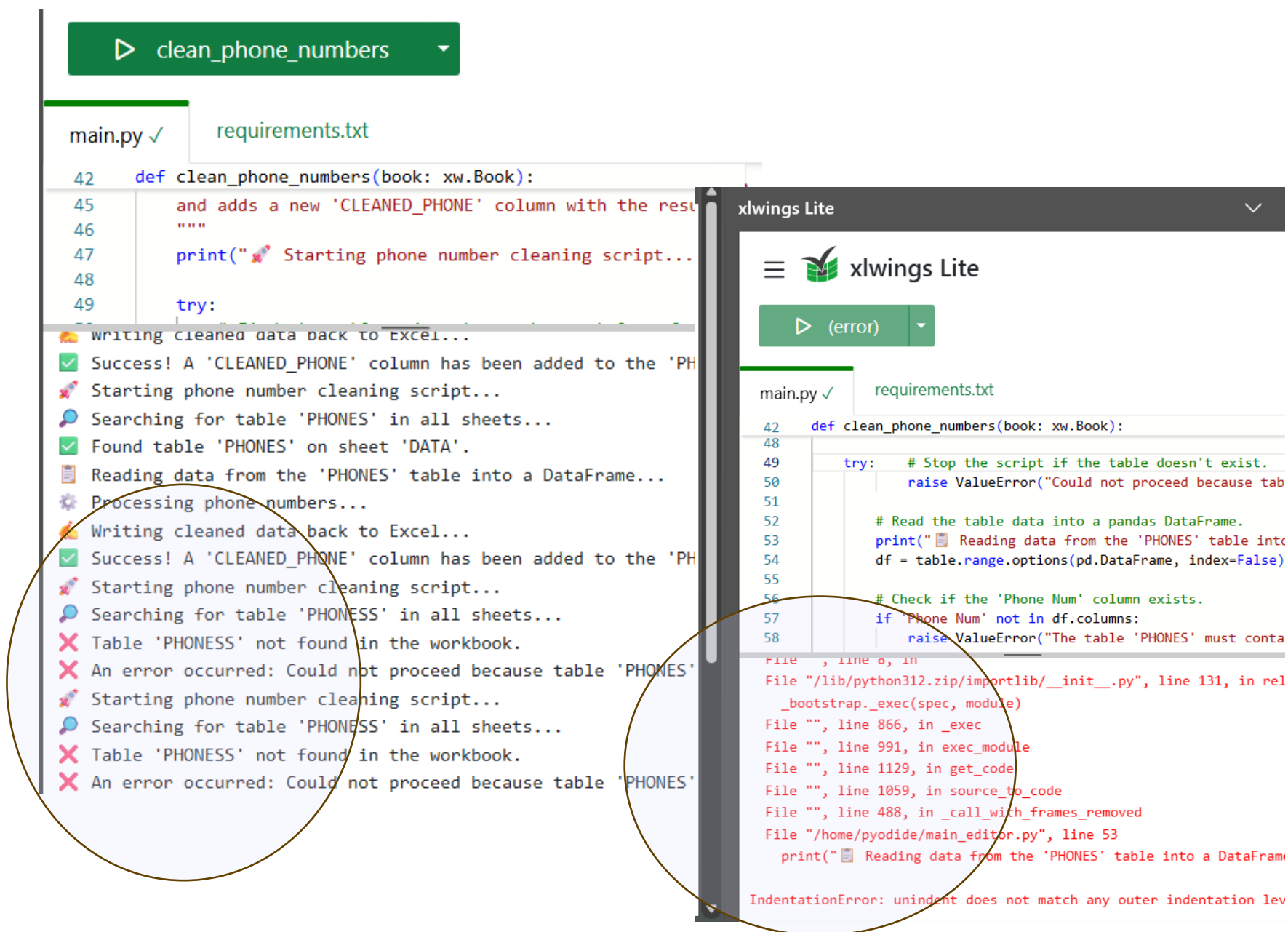It Works For You.

Master It.

6. Validate Results

- Are the results as you expected?

- If not, then share the offending row or screenshot with AI and ask it to fix it

Name	Phone Num	CLEANED_PHONE
A	9845230763	9845230763
B	919845230763	9845230763
C	984523076	0
D	+91-919845230763	9845230763
E	91919845230763	9845230763
F	91 919845230763	9845230763
G	91 919845230763,9845230764	9845230763
H	984523076,9845230763	9845230763
I	984523076,98452307	0
J	my phone numer is 9845230763	9845230763
K	amar@harolikar.com, 9845230763	9845230763
L	amar@harolikar.com, 984523076	0
M	nopes	0
N	sadf sdf	0
O	8644	0
P	llka sf(**&&) sdf	0
Q	9845230760	9845230760

Key Principle: Do not compromise on validations just because AI did it

7. Review Logs

Review logs for any errors - sometimes they might not be apparent. Copy paste the error log to AI and ask it to fix it

Tip: If you are not sure, just dump the whole log to AI and let it review.

Power User Tip 1 : AI Audit

1. Ask AI to run an Audit

2. Always in a new session

3. Provide all context

   What you are trying to do, pseudo code, actual code, and any other relevant background

Instruction 03 : AI Audit

Id like you to conduct and audit of the following code and logic against the business requirement. Please review it critically and let me know if there are any issues, gaps, or potential edge cases that may have been overlooked. Be honest, objective, and professional in your assessment.

I ran the Audit in a new session. AI found some potential gaps. What did you find?

Power User Tip 2 : AI Test Cases

1. Ask AI to create test cases

2. Focusing on outliers and edge cases

3. Always in a new session

4. Provide all context

Instruction 4 : AI Test Cases

> Based on the business requirement and sample records I'll provide, generate edge cases, outlier scenarios, and unusual data combinations. Focus on Including invalid, extreme, or rare cases. As well as both functional and data-related anomalies. Objective is to test my script to see if it misses any edge cases. Use your creativity and business understanding. Share cases not covered in my sample

I had AI pull out edge cases, that the code misses.
Did you find any?

What's possible with xlwings Lite?

A few of the xlwings Lite apps I've built, from web scrapers to database tools.

⬇

🗇 Explore Live Apps

🌐 Web Scraper ↗

Transform Excel into a powerful web scraping platform using Jina AI and Google Gemini

• Uses Jina AI API for webpage rendering to markdown
• Google Gemini analyzes content for intelligent data extraction
• Process batches of URLs with custom column specifications

📈 Technical Analysis ↗

Create professional technical analysis reports with AI-powered chart pattern recognition

• Pulls live data from Yahoo Finance via custom FastAPI server
• Gemini Vision API analyzes charts for technical patterns
• Generates PDF/HTML reports with embedded charts and tables

🗄 Database & ML ↗

Connect to any database with custom FastAPI layer for data analysis and machine learning

• Connects to PostgreSQL/MySQL via web API layer
• Pull random records, run custom queries, explore data
• Build XGBoost models with evaluation and decile tables

⚡ AI Workflows ↗

Automate Excel workflows with AI schema detection using Gemini 2.0-Flash or GPT-4o

• Automatic schema detection with structured JSON output
• Identify categorical and numerical variables intelligently
• Generate SQL-compatible data types for database integration
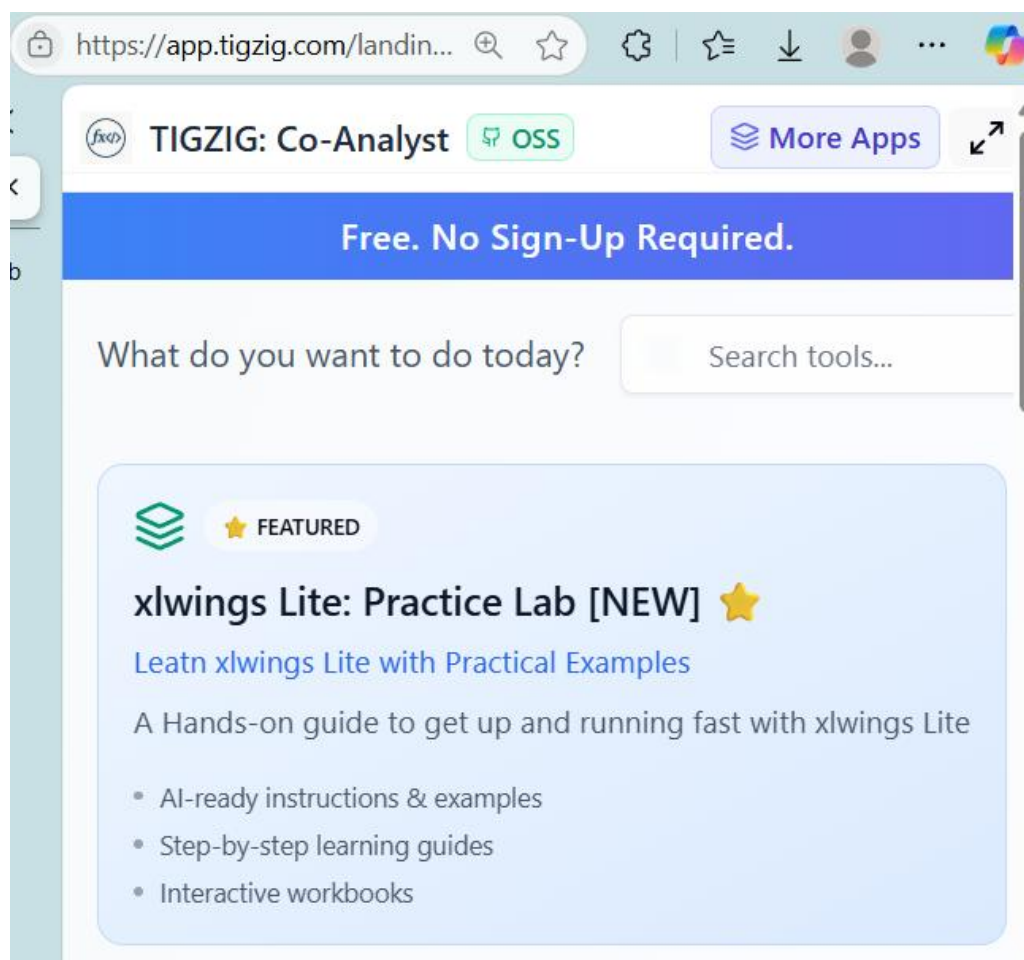
TIGZIG: Micro-Apps for Analytics

25+ apps: Text-to-SQL / Database AI / Automation / Quants

Get the full module and workbooks:

app.tigzig.com

Path: xlwings Lite: Practice Lab

Amar Harolikar

Specialist – Decision Sciences & Applied Gen AI

Builder of app.tigzig.com