



Automated Coding for a Simple Data Science Project Using RooCode and GitHub Copilot

[HOME](#) » [BLOG](#) » [UNCATEGORIZED](#)



YONG SHENG TAN / FEBRUARY 9, 2025

Share via:

[roocode-action](#)

Download

My prompt

```
# Data Science Task: Telco Customer Churn Analysis
```



```
## **Objective**
```

You're a data scientist responsible for building a predictive model for telco

Use the **SEMMA (Sample, Explore, Modify, Model, Assess)** methodology to guid

```
## **1. Sample**
```

- Load the dataset and check for:
 - Missing values
 - Duplicate records
 - Class imbalance in the churn label

```
## **2. Explore**
```

Conduct **Exploratory Data Analysis (EDA)** with:

- **Statistical Summaries:** Analyze numerical and categorical features.
- **Visualizations:** Generate histograms, boxplots, and correlation heatmaps.
- **Feature Relationships:** Identify trends between features and customer chu

```
## **3. Modify**
```

Perform **data preprocessing** steps:

```

22 - Handle missing values, outliers, and duplicates.
23 - Encode categorical variables and scale numerical features.
24 - Engineer new features (e.g., tenure segmentation, interaction features).
25
26 ## **4. Model (Supervised Learning: Churn Prediction)**
27 Train and evaluate predictive models:
28 - **Models to Use:**
29   - Logistic Regression
30   - Decision Trees
31   - Random Forest
32   - XGBoost
33 - **Hyperparameter Tuning:** Optimize model performance.
34 - **Evaluation Metrics:**
35   - Accuracy, Precision, Recall, F1-score
36   - AUC-ROC Curve
37
38 ## **5. Clustering (Unsupervised Learning: Customer Segmentation)**
39 Perform customer segmentation using:
40 - **K-Means Clustering**
41   - Determine optimal clusters using the **Elbow method** and **Silhouette sco
42 - **DBSCAN (Density-Based Clustering)**
43   - Identify natural groupings and detect outliers.
44 - Compare clustering results and extract business insights.
45
46 ## **6. Assess**
47 Summarize key findings:
48 - Identify the most important factors influencing churn.
49 - Provide insights from **customer segmentation** for potential marketing stra
50 - Recommend actions based on **churn prediction & clustering analysis**.
51
52 ### **Deliverables:**
53 - Well-structured, reproducible analysis with code and visualizations.
54 - Interpretations and business recommendations based on the findings.

```

Result:

Telco Customer Churn Analysis Results

Key Findings

1. Sample Analysis

Dataset contains 7,043 customers with 21 features

No missing values or duplicate records

Churn rate is 26.5% (imbalanced dataset)

Clean dataset with good quality data

2. Exploratory Analysis

Contract Type Impact:

Month-to-month contracts have highest churn (42.7%)

Two-year contracts have lowest churn (2.8%)

One-year contracts show moderate churn (11.3%)

Service Usage:

Fiber optic customers have higher churn (41.9%)

DSL users show moderate churn (19%)

No internet service shows lowest churn (7.4%)

Customer Tenure:

Non-churned customers average 37.6 months

Churned customers average 18 months

Higher tenure correlates with lower churn

3. Modeling Results

Predictive Models Performance

Logistic Regression:

Accuracy: 81.6%

ROC-AUC: 0.86

Best overall performer

Random Forest:

Accuracy: 79.4%

ROC-AUC: 0.83

Good feature importance insights

Decision Tree:

Accuracy: 73.7%

ROC-AUC: 0.67

Simpler but less accurate

Most Important Churn Factors

Tenure-charge interaction

Monthly charges

Total charges

Tenure

Contract type

4. Customer Segmentation

Three distinct customer segments identified:

Loyal Low-Spenders (19.7% of customers)

Long tenure

Lowest monthly charges

Very low churn rate (3.2%)

Mostly on long-term contracts

Premium Stable Customers (29.8%)

Highest tenure

High monthly charges

```
61 Moderate churn rate (15.1%)
62 Mix of contract types
63 High-Risk New Customers (50.5%)
64
65 Shortest tenure
66 Average monthly charges
67 Highest churn rate (42.4%)
68 Mostly month-to-month contracts
69 Recommendations
70 Contract Strategy:
71
72 Incentivize month-to-month customers to switch to longer contracts
73 Offer special discounts for 1-2 year commitments
74 Create loyalty rewards for long-term contracts
75 Service Improvements:
76
77 Focus on improving Fiber optic service reliability
78 Enhance technical support for new customers
79 Implement proactive customer service for high-risk segments
80 Customer Retention:
81
82 Develop targeted retention programs for customers in their first 18 months
83 Create special offers for high-ARPU customers showing churn risk
84 Implement early warning system based on identified churn factors
85 Segmented Approach:
86
87 Loyal Low-Spenders: Focus on upselling and maintaining satisfaction
88 Premium Stable: Prioritize premium service and personalized attention
    High-Risk New: Implement aggressive retention strategies and enhanced support
```

```
1 | I found that your result is not consistent with previous experiment, why? ...<
```

Pre-requisites

- Have a Github Account
- Installed Visual Studio Code

Introduction to the Tools

(i) Cline

-

(ii) Github Copilot

- Pricing



GitHub Docs

Version: Free, Pro, & Team ▾

[GitHub Copilot](#) / [About GitHub Copilot](#) / [Subscriptions](#)

Comparing Copilot subscriptions [↗](#)

	Copilot Free	Copilot Pro	Copilot Business	Copilot Enterprise
Pricing	Not applicable	\$10 USD per month, or \$100 USD per year (free for some users)	\$19 USD per granted seat per month	\$39 USD per granted seat per month
Code completion in IDEs ^[1]	2000 completions per month	✓	✓	✓
Copilot Chat in IDEs ^[2]	50 messages per month	✓	✓	✓
Copilot Chat in GitHub Mobile	×	✓	✓	✓

Source: [Github](#)

- But, if you're a student, teacher, or maintainer, you could apply for Github Education account to get free access to Copilot Pro as well as the Github Copilot subscriptions.

Learn more here: <https://docs.github.com/en/copilot/managing-copilot/managing-copilot-as-an-individual-subscriber/managing-your-github-copilot-pro-subscription/getting-free-access-to-copilot-pro-as-a-student-teacher-or-maintainer>

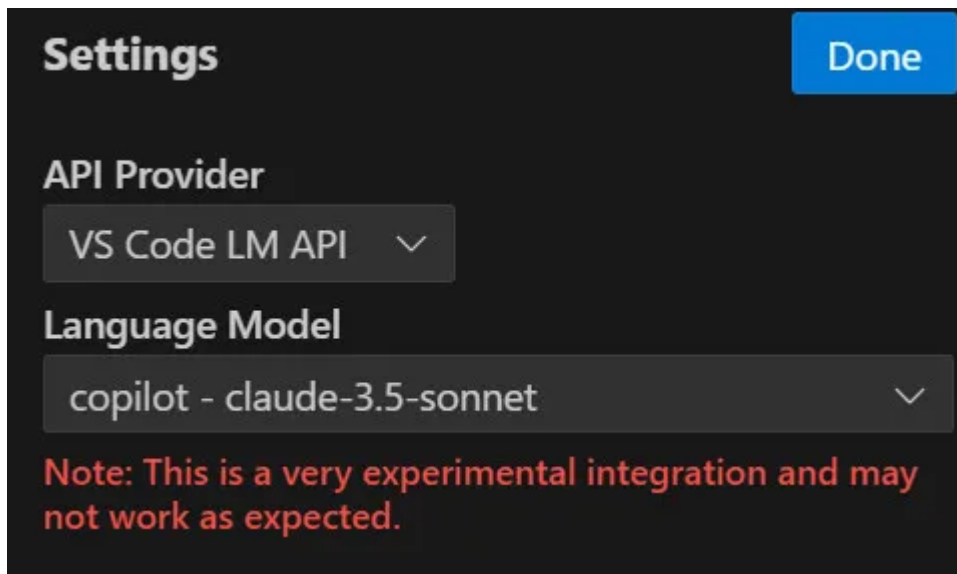
Why not just Github Copilot?

- To me, as at 5 Feb 2025, Cline offers superior user experience compared to Github Copilot alone...

Can refer to <https://addyo.substack.com/p/why-i-use-cline-for-ai-engineering>
(Note: Addy Osmani is a famous Googler for frontend)

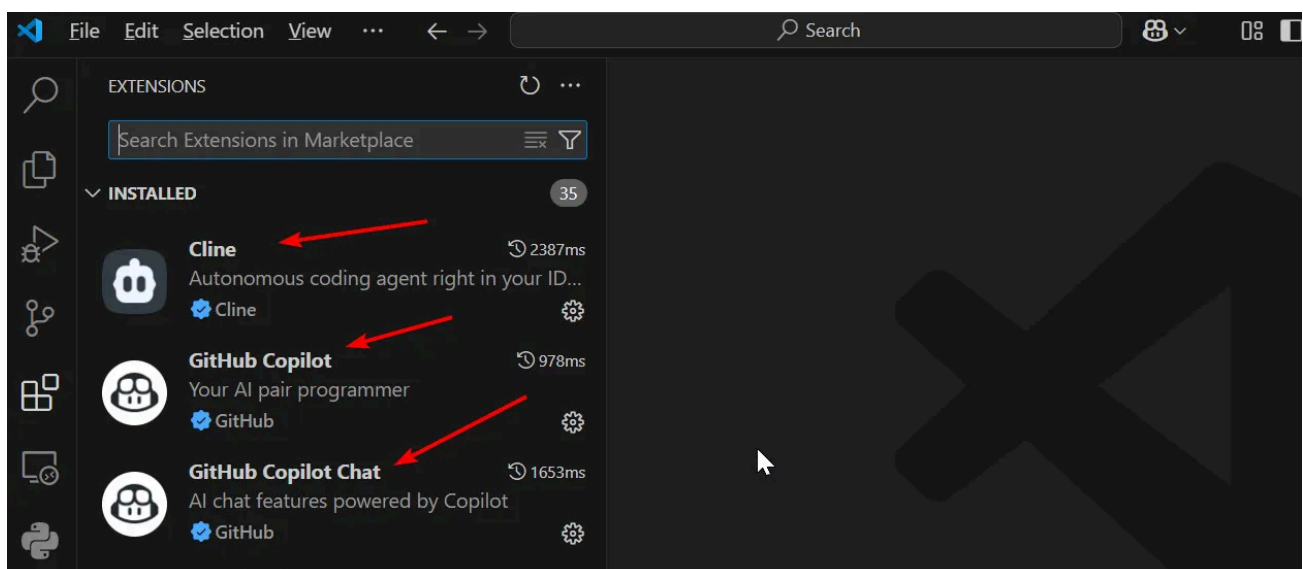
Warning first before we proceed

As at 5 Feb 2025, integrating Cline with Github Copilot LLM is a very experimental integration and may not work as expected.

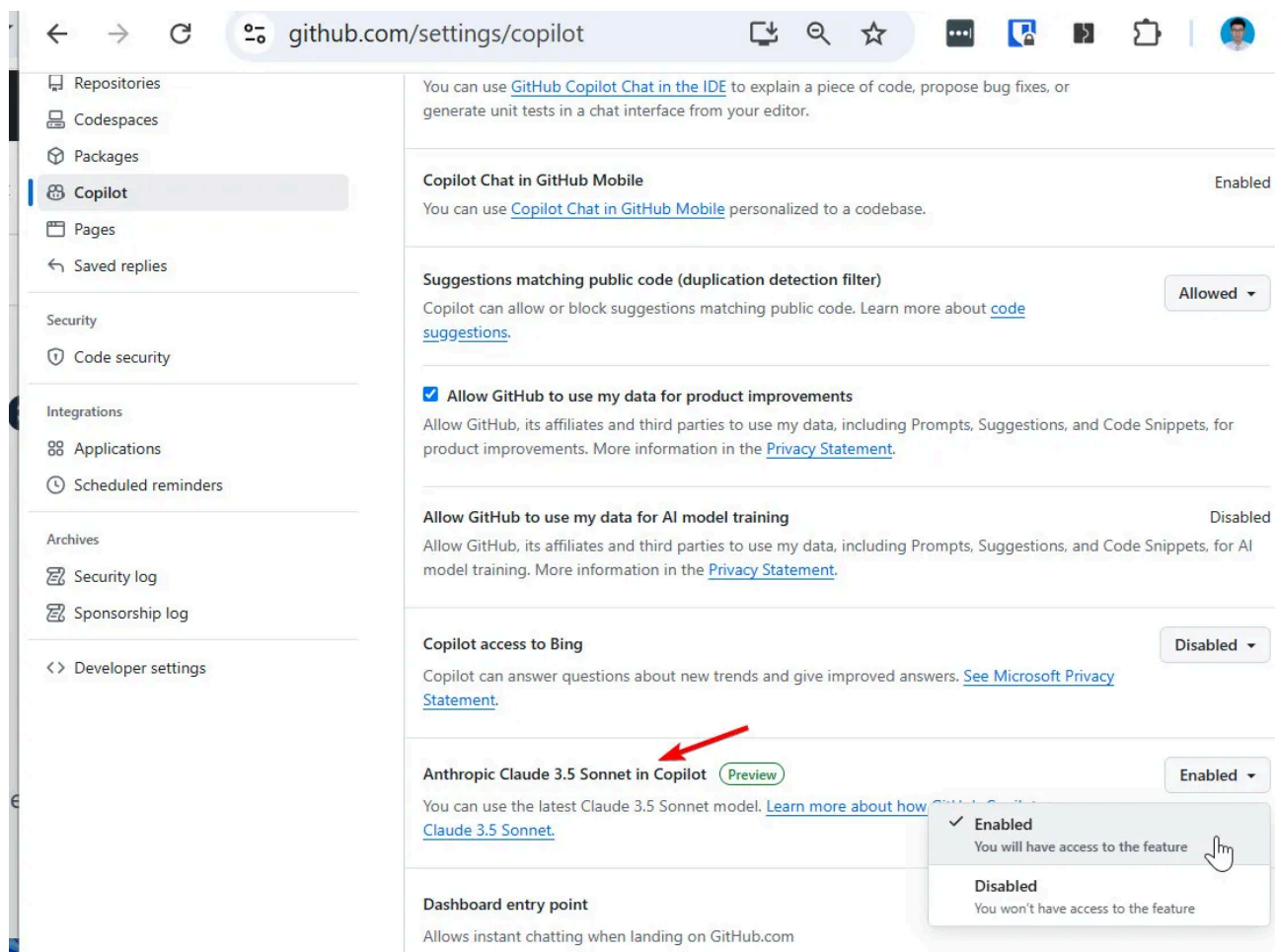


Steps

- Plugin to install
 - RooCode: <https://marketplace.visualstudio.com/items?itemName=RooVeterinaryInc.roo-cline>
 - Github Copilot: <https://marketplace.visualstudio.com/items?itemName=GitHub.copilot>
 - Github Copilot Chat: <https://marketplace.visualstudio.com/items?itemName=GitHub.copilot-chat>



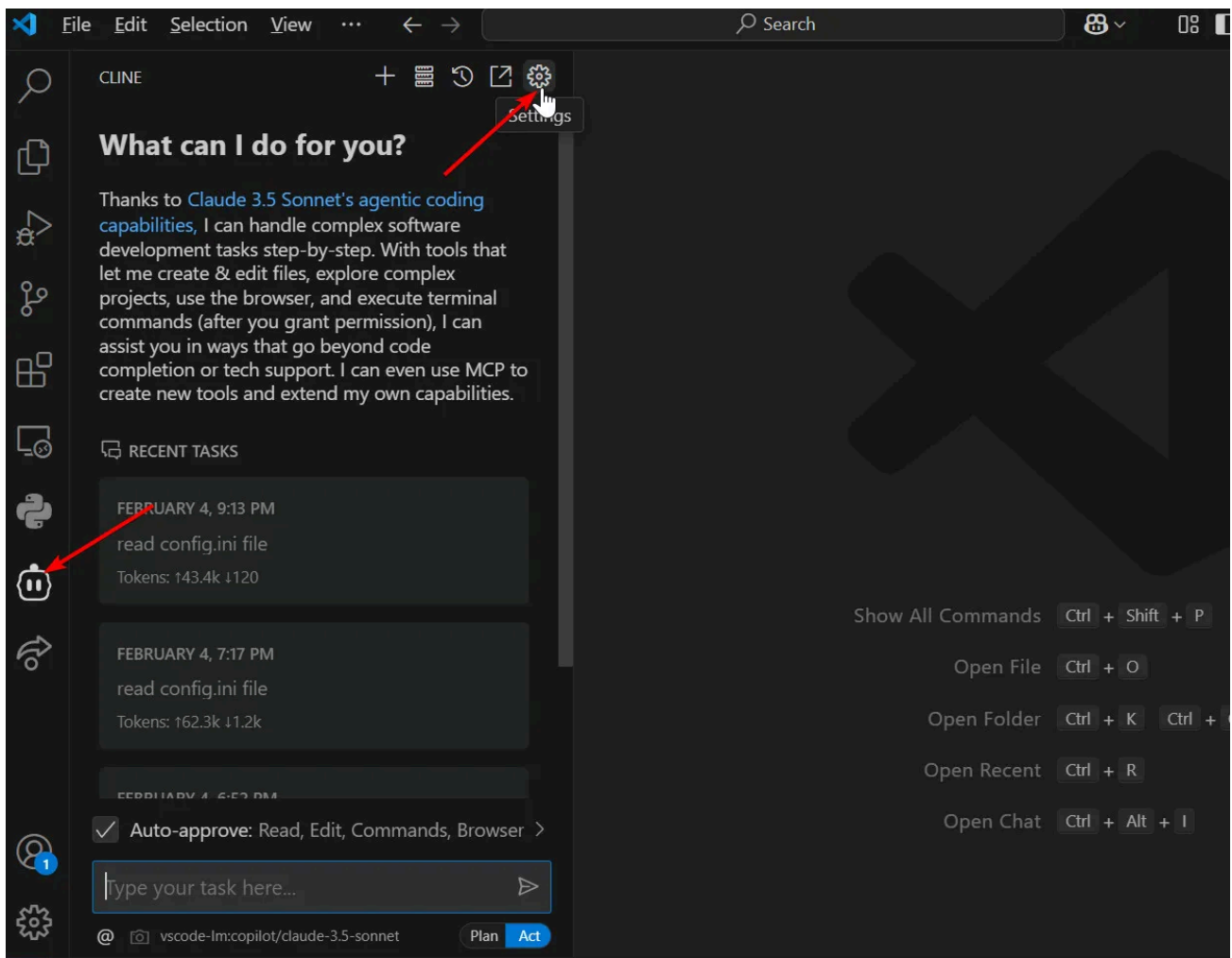
- Anthropic access in Github Copilot is disabled by default. Go to <https://github.com/settings/copilot> to enable it.



Installing MCP server for data exploration in Cline

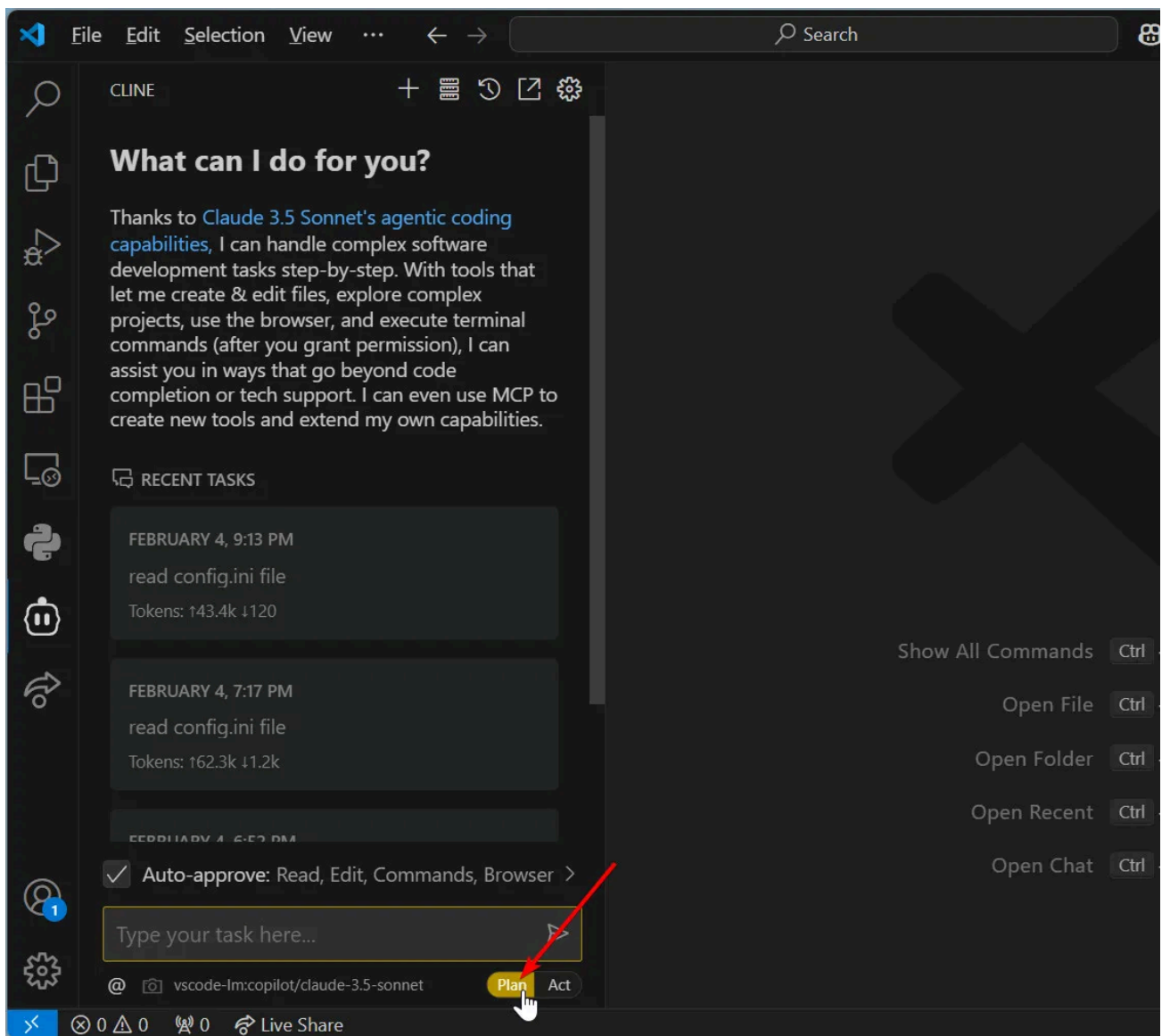
- Dataset: <https://www.kaggle.com/datasets/blastchar/telco-customer-churn>
- MCP server for data exploration: https://github.com/reading-plus-ai/mcp-server-data-exploration/tree/main/src/mcp_server_ds

Demonstration

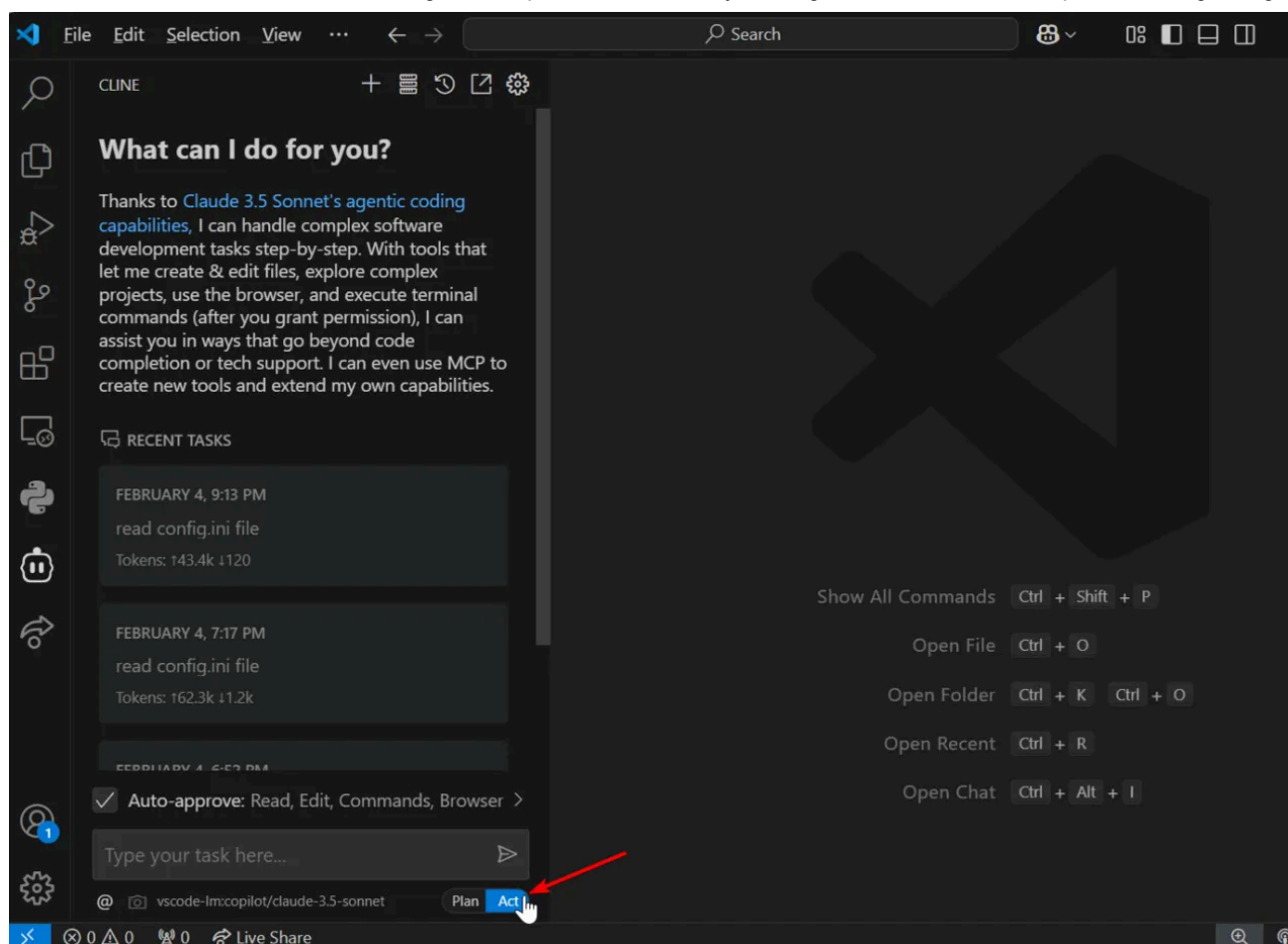


- Switch plan and act

Plan (note: can switch to use smarter model like o1-ga or o3-mini?)



Act (note: can switch to use Antropic Claude instead)



Why Using RooCode (the fork) over Cline

- I just want a quick prototype
- I want to be quick and fully autonomous

Current Challenges of AI Coding

- AI Recommendations of known malicious or deprecated libraries during coding task
- AI suggestions might not follow secure coding practices
- Sensitive information (e.g., credentials) could be accidentally exposed to AI models

Future Consideration

Try to evaluate our LLM's data analysis task via some public benchmark such as [InfiAgent-DABench: Evaluating Agents on Data Analysis Tasks](https://github.com/InfraAgent/InfraAgent-DABench).

References

- **Using AI for Coding: My Journey with Cline and Large Language Models** <https://pgaleone.eu/ai/coding/2025/01/26/using-ai-for-coding-my-experience/>
- **Why I use Cline for AI Engineering** <https://addyo.substack.com/p/why-i-use-cline-for-ai-engineering>
- **CodeGate: Open Source Tool Secures AI Coding Assistants** <https://thenewstack.io/codegate-open-source-tool-secures-ai-coding-assistants/>
- 🍷 **Cline + MCP: VS Code's AI Assistant Gets Agentic Superpowers** <https://medium.com/@elio.verhoef/cline-mcp-vs-codes-ai-assistant-gets-agentic-superpowers-3008aefba178>
- **InfiAgent-DABench: Evaluating Agents on Data Analysis Tasks** <https://github.com/InfiAgent/InfiAgent>, <https://arxiv.org/html/2401.05507v1>
- **Automating Data Analytics with Large Language Models** <https://medium.com/%40felixbastian.hans/automating-data-analytics-with-large-language-models-3e83bcdbd51d>
- **Can Large Language Models Replace Data Scientists in Clinical Research?** <https://arxiv.org/html/2410.21591v1>
- **Roo Code vs Cline** https://www.reddit.com/r/ClaudeAI/comments/1i7vd0i/a_comprehensive_look_roo_code_vs_cline/?tl=fil

 Edit

PREVIOUS POST

How to Use Claude's MCP Server for Large CSV Data Exploration: A Step-by-Step Guide



Yong Sheng Tan

[View Profile](#)

Hi, I'm Yong Sheng. I am currently a marketing executive at ShareInvestor. I am good at utilizing tools to fulfil and achieve my work such as using ChatGPT for marketing copy, Mailchimp for email marketing, and some coding works (mostly in python) to automate some of task flows.



Subscribe to our newsletter to receive the latest news and offers from us.



Subscribe

[Edit Form](#)

Latest Post

How to Use Claude's MCP Server for Large CSV Data Exploration: A Step-by-Step Guide

How to Self-host your E-Book Library on S3 storage with Kavita Reader?

Web Scraping Tutorial with Python (2024) – Using Python requests, BeautifulSoup, Selenium, anti-CAPTCHA and Proxy

How to Self-host your E-Book Library on your VPS server?

How to Create a Free Business Email? (2024)

Leave a Reply

Logged in as Yong Sheng Tan. [Edit your profile.](#) [Log out?](#) Required fields are marked *

Add Comment *



Subscribe to our newsletter to receive the latest news and offers from us.

☐ I accept the [Privacy Policy](#).

Post Comment



Copyright © 2025 Tan Yong Sheng

[Privacy Policy](#) | [Cookie Policy](#)