

Only Code If You Want To: Data Science with DataLab (Part 1)

May 28, 2024



Hello!

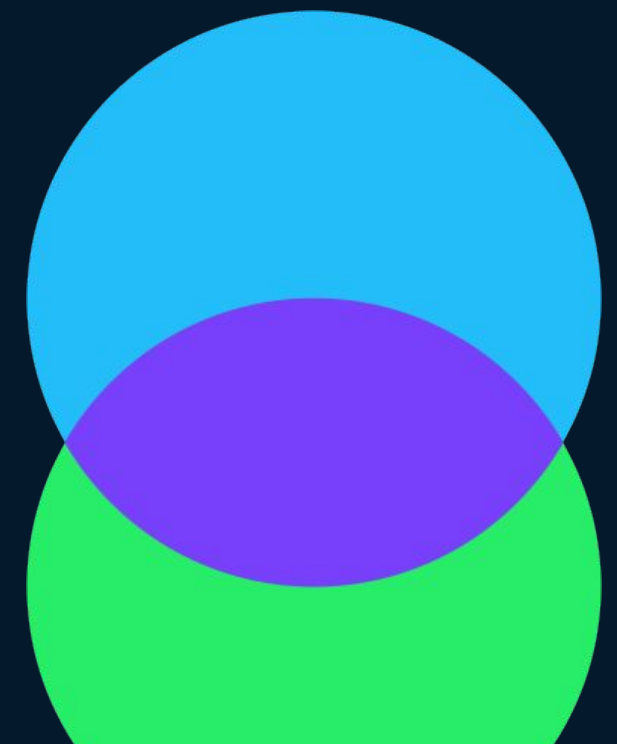
I'm Filip Schouwenaars

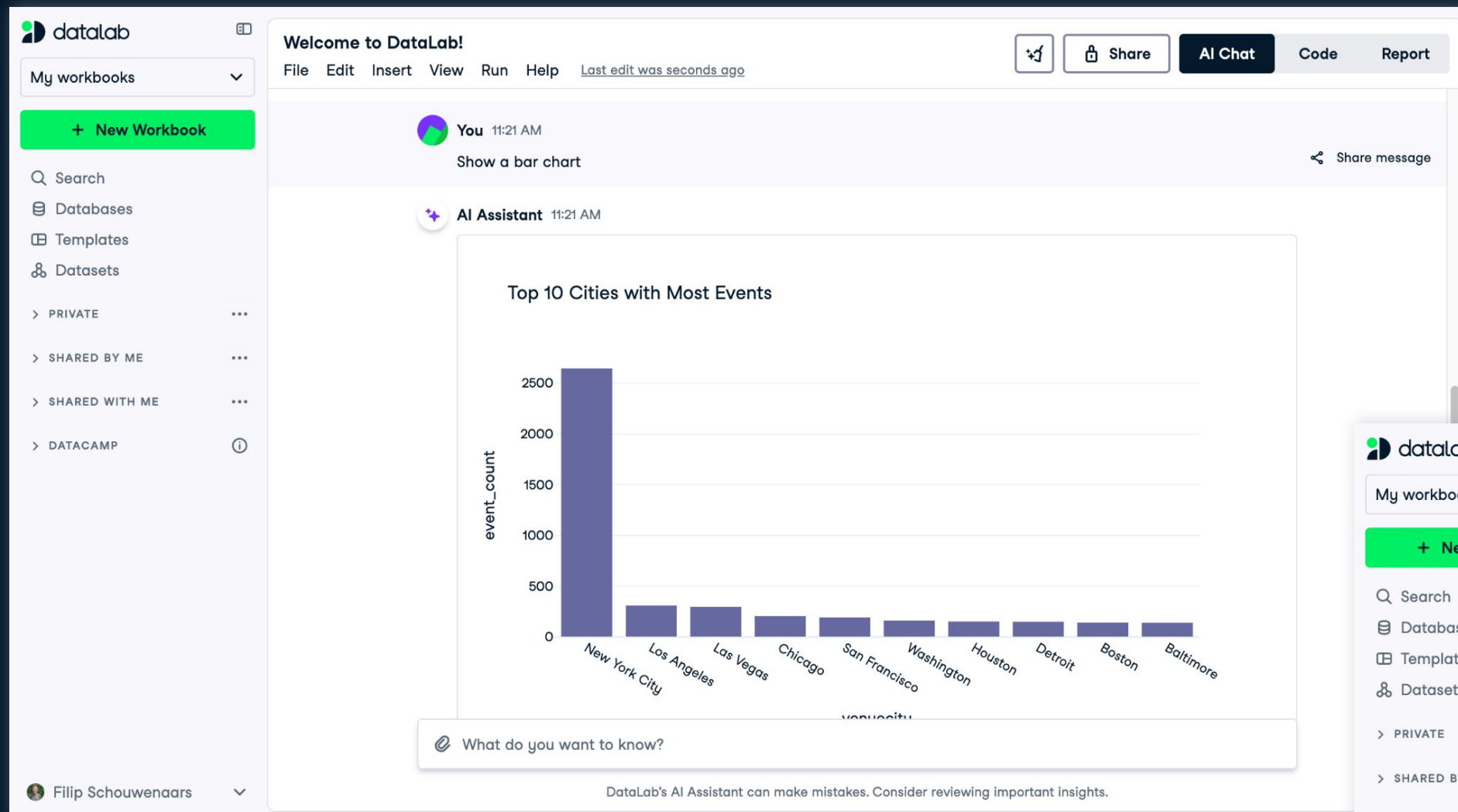


Filip Schouwenaars

VP of Product & Engineering

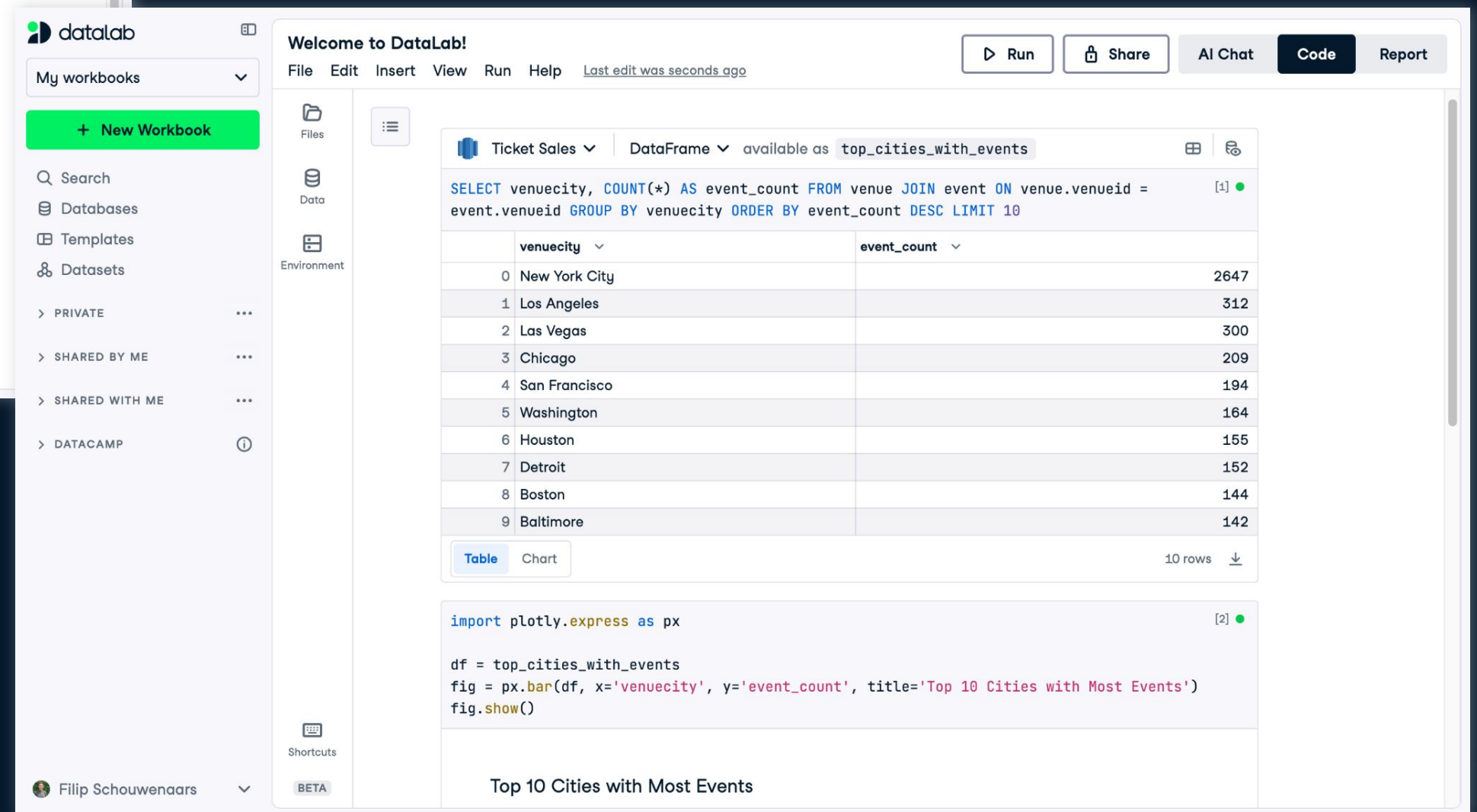
 [fschouw](#)





Fully-featured data notebook to build analyses

AI Chat to analyze data without writing code

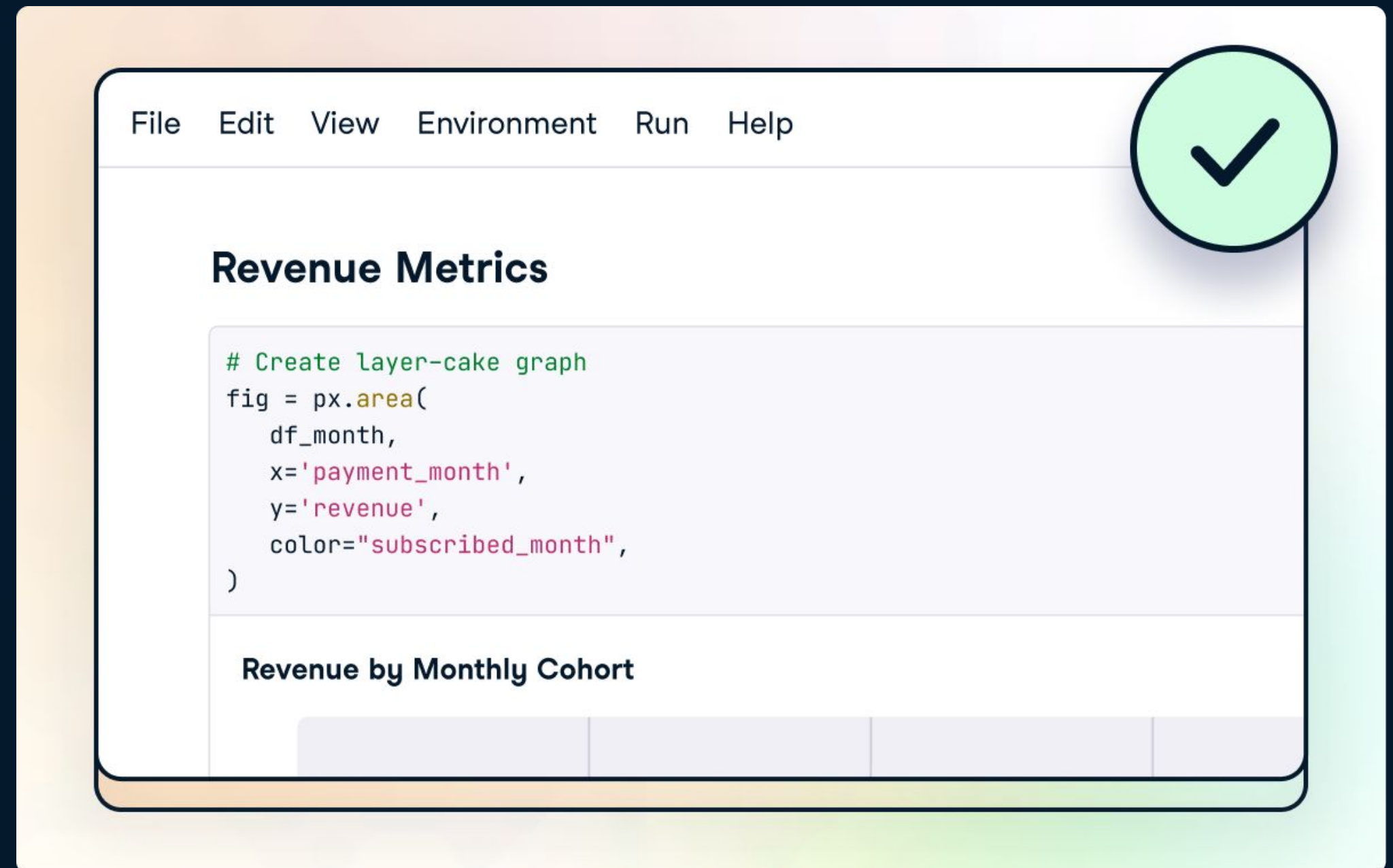


Ready to go

You can participate in this code along by just clicking a link. Everything happens in the web browser. No installation is required.

Every DataLab workbook runs in a fully-managed, cloud-hosted environment with all common data science packages pre-installed.

DataLab supports Python, R and SQL.

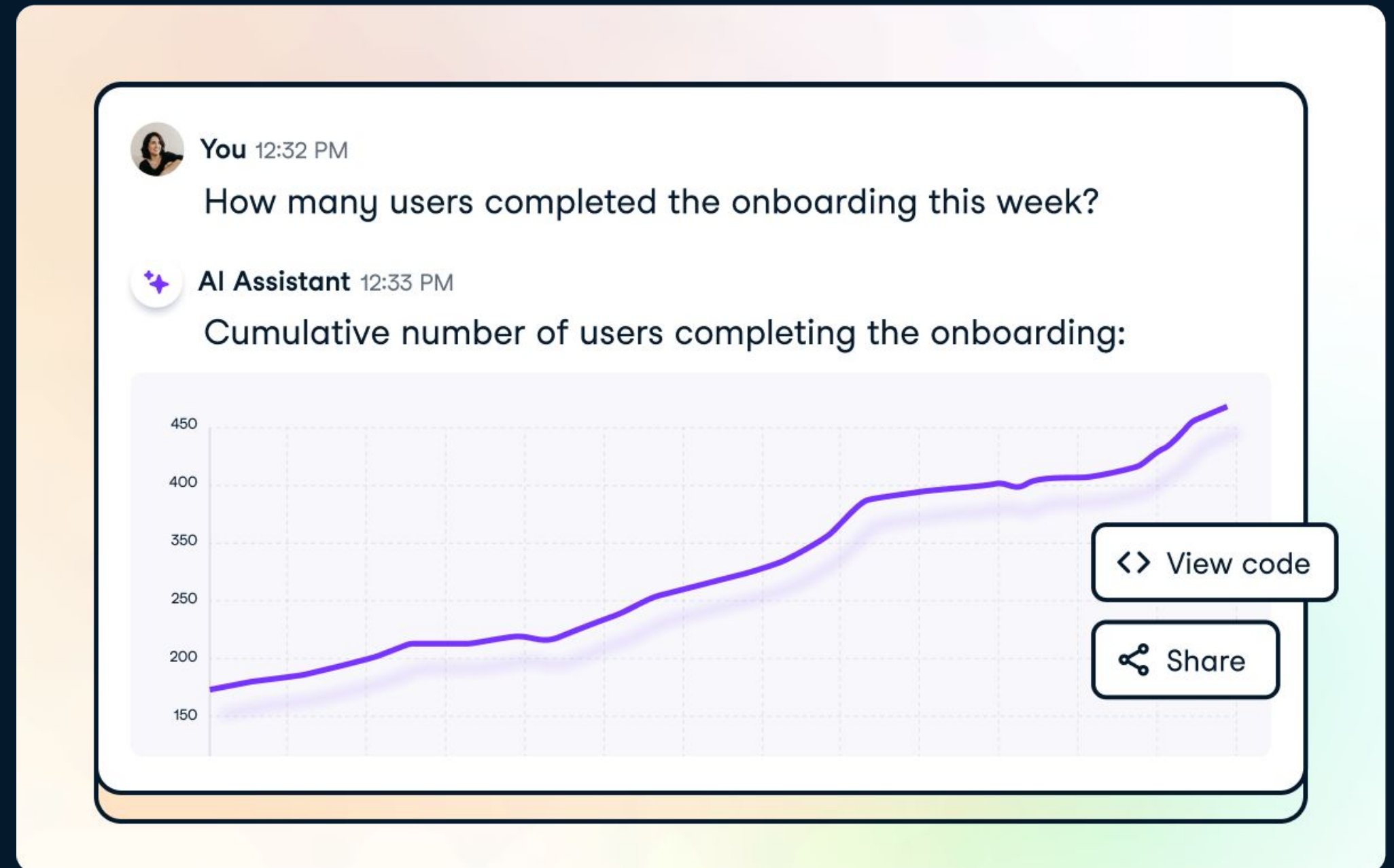


AI built-in throughout

AI Chat: explore your data and dig for insight *without* writing code.

AI Coding Assistant: write, adjust and refactor code with the help of AI. Have the AI fix and explain coding mistakes, helping you build more skill in the process.

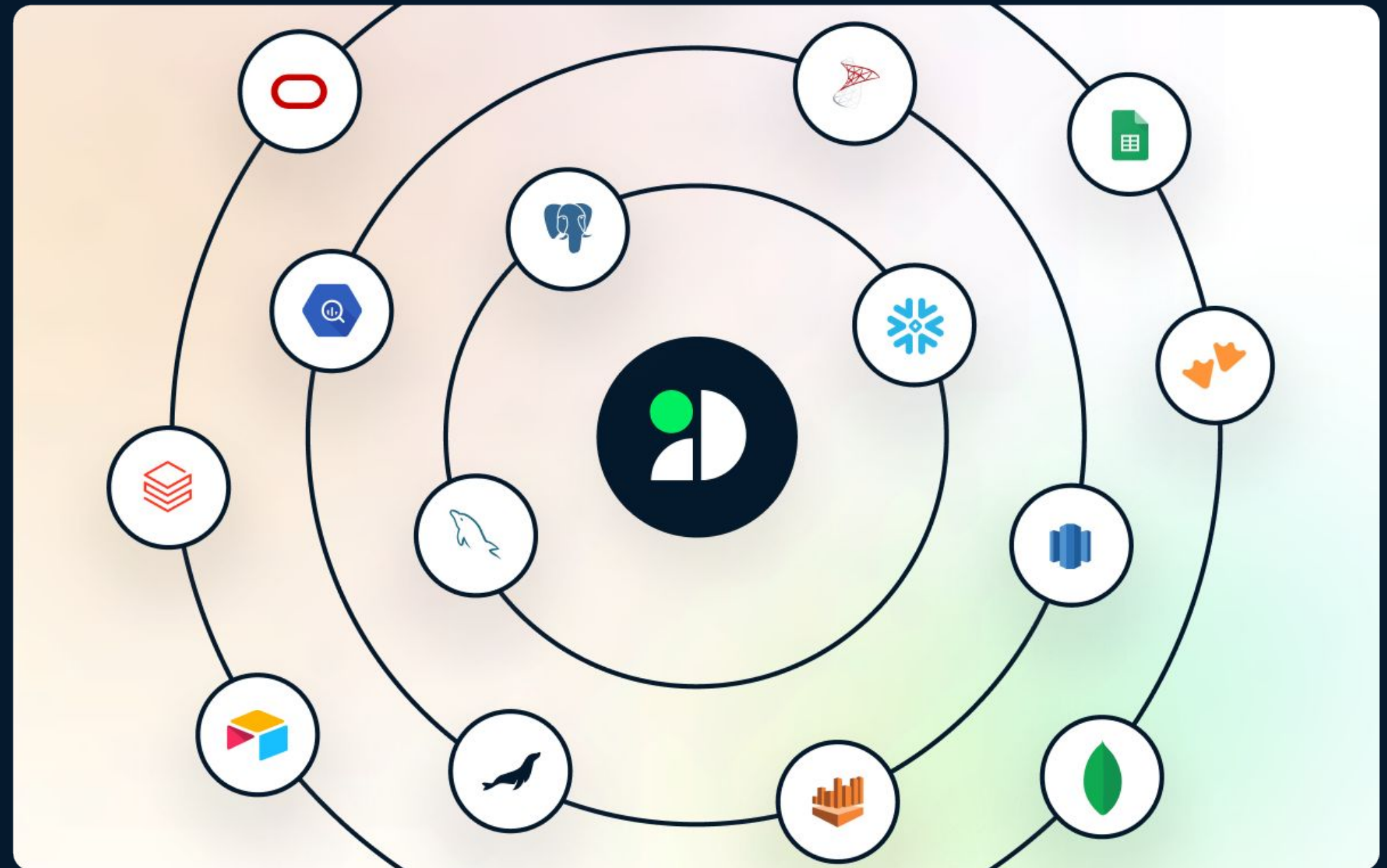
DataLab smartly enhances your question with your unique context to get great results fast.



Integrate with your stack

From CSV files and Google Sheets data to Snowflake and BigQuery: DataLab instantly and securely connects to all your data sources.

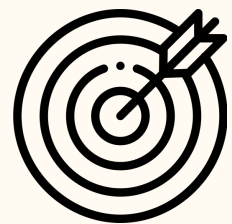
Don't have a dataset lying around? Pick one of DataLab's many sample data sources available in the app.



About this ~~code~~ *prompt*-along



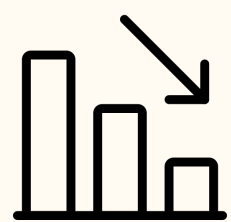
Data: Agricultural raw material prices ([source](#)): wool, copra, cotton, fine wool, hardlog, etc.



Primary Goal: Understand trends over time and correlations in raw material prices.



Secondary Goal: Understand the power and limitations of conversational AI-powered exploration.



Why? Generative AI is rapidly changing how we analyze data; it pays to keep up with this evolution.



Important: AI systems are stochastic; the same prompt may lead to different results!

If you click the link that we'll share in chat, you'll create your own DataLab workbook to follow along!





You May 27

I want to analyze time series data on wood alone, can you zoom in on those columns?



Prompt generator

You are an AI assistant for DataLab, a data notebook. [...]
This **dataset** is connected: [...]
These **variables** are defined in the session: [...]
The **previous chats and code** in the workspace: [...]
My question or goal: **I want to analyze time [...]** wood alone.



OpenAI



output parser

```
[  
  createCell('python', 'wood_columns = ...')  
  runCell('1ja8e390n')  
]
```



AI Assistant May 27

	Month ▾	Hard log Price ▾	Hard log price % Change ▾	Hard sawnwood Price ▾	Hard sawnwood price % C... ▾	So
0	Apr-90	161.2	-	549.91	-	
1	May-90	172.86	7.23%	491.88	-10.55%	
2	Jun-90	181.67	5.10%	495.39	0.71%	
3	Jul-90	187.96	3.46%	485.86	-1.92%	
4	Aug-90	186.13	-0.97%	487.52	0.34%	

5 rows



Prompt generator

You are an AI assistant for DataLab, a data notebook. [...]
My question or goal: **Describe the results.**

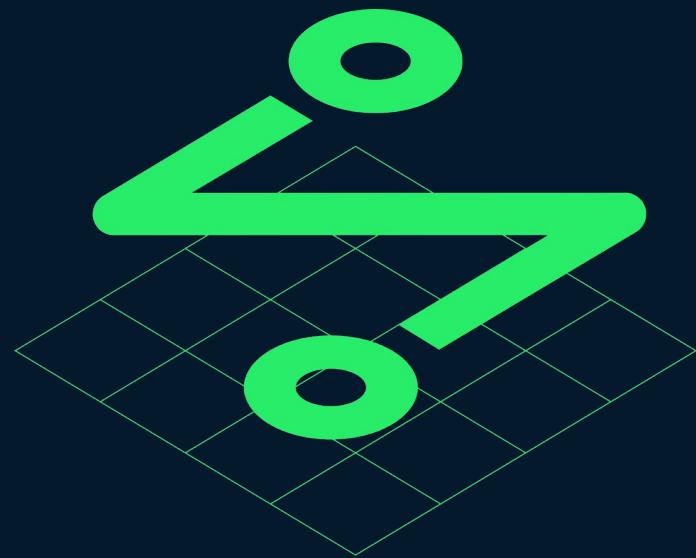


OpenAI



output parser

The filtered dataset focuses on wood-related materials and includes the following columns: Month, Hard log Price, Hard log price % Change, Hard sawnwood Price, Hard sawnwood price % Change, Softlog Price, Softlog price % Change, Soft sawnwood Price, Soft sawnwood price % Change, Wood pulp Price, and Wood pulp price % Change. The data spans multiple months, starting from April 1990, and provides both the prices and percentage changes for each wood-related material.



**Want to learn
more?**

- ✓ Play around with more [sample datasets](#)
- ✓ Relive [past code alongs](#)
- ✓ Read the [DataaLab Docs](#)





Thank you