



Chatting with your data A hands on Tutorial on Generative AI for Interactive Data Analysis

Presenter: Muhammad Umair Aslam, Asad Ullah Khan

Data Science Lab, AITeC

Workshop Agenda



1.



Introduction

What is Generative AI?
The Challenge of Using AI in Data Analysis
Introducing PandasAI and Ollama

2.



Setting Up Your Environment

Installing Required Libraries
Brief Overview of Ollama Setup

3.



Hands-On with PandasAI and Ollama

Loading and Exploring a Sample Dataset
Live Demo: Chatting with Your Data

4.



Beyond PandasAI

Additional Tools and Platforms
Q&A and Next Steps

What is Generative AI



Artificial intelligence that creates new content (text, images, code, etc.)



It learns patterns from massive datasets



Examples

Text: GPT4 ,Google Gemini for NLG and NLU and multimodal tasks

Images: Dall-E,Stable Diffusion and SORA

Code: GitHub Copilot for suggesting and completing code snippets



Relevance to Data Analysis:

Automate the generation of reports and summaries

Uncover hidden patterns and insights

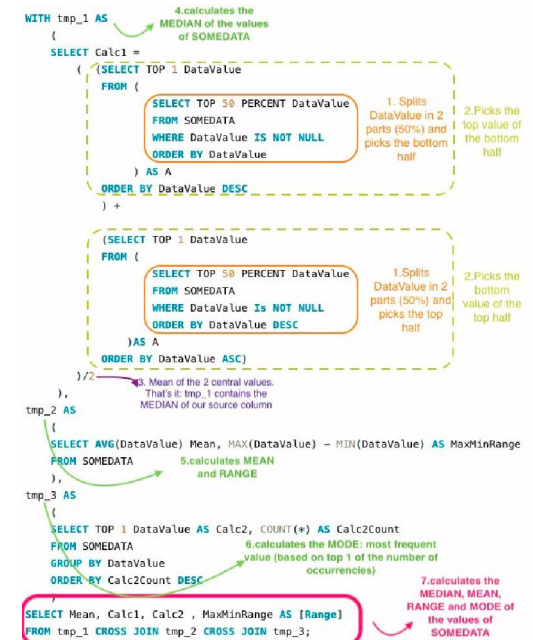
Simplify complex analyses through natural language interaction

The Challenge of Using AI in Data Analysis



Barriers to Entry:

- Requires specialized skills (coding, statistics, machine learning)
- Complex tools and APIs
- Difficult to translate data questions into code
- Privacy Concerns

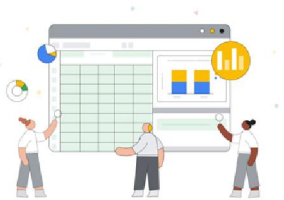


Case Study AI in Data Analysis :DataChat.ai



Enterprise Solution:The DataChat.ai Solution

- Raised \$30M in funding
- Generates \$3M in revenue
- No-code platform: Simplifies AI-powered data analysis for non-experts
- Conversational interface: Ask questions in plain English
- Automated insights: Generates reports, visualizations, and predictions
- Democratization of AI: Makes AI accessible to a wider audience



[Datachat](#)

Introducing PandasAI & Ollama



PandasAI:

- An open-source library that lets you interact with your Pandas DataFrames using natural language.
- Translates your questions into Pandas code, making data analysis more intuitive.
- Works seamlessly with popular large language models (LLMs), including OpenAI's GPT models and local models like Ollama.

Ollama:

- An open-source framework for running LLMs locally.
- Provides privacy, cost savings, and customization compared to cloud-based APIs.
- Privacy Preserving

Applications of GenAI in Data Domain



- Rapid Prototyping & Exploration:** Quickly test hypotheses and uncover insights.
- Self-Service Analytics:** Empower non-technical users to explore data independently.
- Automated Reporting:** Generate reports and summaries
- Anomaly Detection:** Proactively identify unusual patterns or outliers.
- Data Cleaning & Transformation:** Streamline data preparation and feature engineering.

Environment Setup



Required Libraries:

- PandasAI :Enables natural language interaction with Pandas DataFrames
- Ollama : Running LLMs locally

Live Demo



- Goal is to replicate the functionality of datachat.ai
- We will use llama3

[Colab](#)

Startups



[Cooloop](#)

[Harmonize AI](#)

[Maxqda](#)



THANK YOU