StatBot Pro

StatBot Pro – Autonomous CSV Data Analyst

Created & Architected by: Sarvesh Ghotekar
Role: AI / Backend Engineer
Core Focus: Autonomous Agents, Secure Code Execution, Data Intelligence
Built as: Production-grade AI system (not a demo)

StatBot Pro demonstrates how AI agents can safely reason, write code, execute analysis, and return verifiable results.

Project Overview

StatBot Pro is an autonomous data analysis agent that allows users to upload CSV files and ask natural language questions. The system intelligently generates Python analysis code, executes it in a sandboxed environment, and returns insights and visualizations through a modern web interface.

Key Capabilities

- Natural language → data analysis
- Autonomous Python code generation
- Secure sandboxed execution
- Automatic visualization generation
- Streaming, step-by-step reasoning

Technology Stack

▼ Backend

- Python 3.11+

- FastAPI
- Pandas
- Matplotlib
- Pydantic
- AsyncIO

▼ Frontend

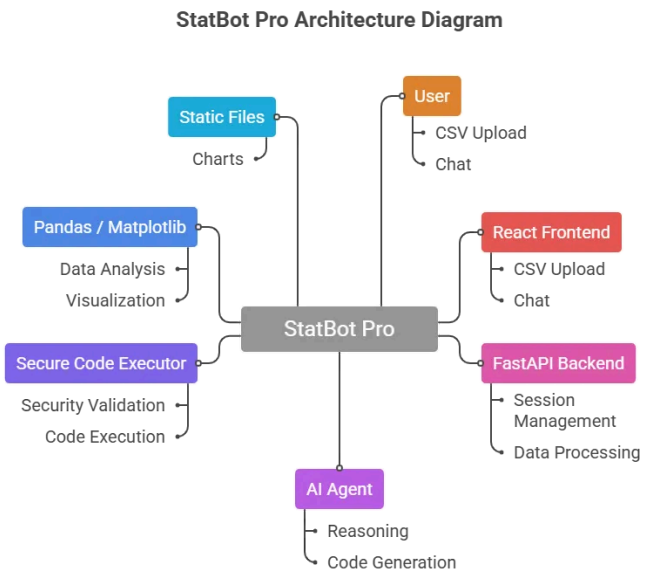
- React 18
- TypeScript
- Vite
- Tailwind CSS
- Radix UI / Lucide

▼ Infrastructure

- Docker & Docker Compose
- Nginx (production)
- GitHub Actions (CI/CD)

System Architecture

Architecture Diagram



Architecture Flow

1. User uploads CSV
2. Session initialized
3. Question parsed by agent
4. Python code generated
5. Code validated (AST checks)
6. Secure execution
7. Chart / result generated
8. Response streamed to UI

Core Components

StatBot Agent (`agent.py`)

- Interprets user intent
- Generates Pandas / Matplotlib code
- Self-corrects on execution failure

- Supports filtering, aggregation, visualization

Secure Executor

- AST-based code validation
- Restricted imports
- Blocked system calls
- CPU & memory limits
- Execution timeout

> 🚫 The agent can write code, but **cannot escape the sandbox**.

FastAPI Backend

- Session management
- Rate limiting
- Streaming responses
- API logging & monitoring

React Frontend

- Dataset inspector
- Chat-based interaction
- Results & charts panel
- Streaming reasoning display

Security Model

Threats handled

- Malicious code execution
- Prompt injection

- Resource exhaustion
- File system access

Safeguards

- AST inspection
- Import whitelisting
- Timeout enforcement
- Isolated workspace per session

> 🔒 Autonomy without safety is dangerous. StatBot Pro enforces both.

API Documentation

API Endpoints

Method	Endpoint	Description
POST	/upload_csv	Upload dataset
POST	/ask_question	Ask analysis question
GET	/health	System health
GET	/static/{file}	Charts & outputs

Features Deep Dive

Natural Language Analysis

- Intent detection
- Context awareness
- Follow-up support

Autonomous Code Generation

- Template-based + dynamic code
- Self-repair on errors
- Pandas + Matplotlib optimized

Visualization Engine

- Histograms
- Correlations
- Time-series plots
- Exportable PNGs

Deployment

Local

```
pip install -r requirements.txt
python main.py
```

Docker

```
docker-compose up --build
```

Production Setup

- Nginx reverse proxy
- Static file serving
- Health monitoring
- Log rotation

Monitoring & Performance

Observability

- Structured logs
- CPU / memory metrics
- Execution time tracking
- Session activity monitoring

> 📈 Every analysis is measurable, traceable, and auditable.

Roadmap

Short-term

- Excel / JSON support
- Improved visualizations

Mid-term

- Interactive charts (D3.js)
- Multi-user collaboration

Long-term

- ML model generation
- Scheduled analysis
- Report export (PDF / PPT)

Conclusion

StatBot Pro demonstrates how autonomous AI agents can be safely deployed in real-world systems. By combining intelligent code generation with strict sandboxing and

observability, the platform enables powerful data analysis without compromising security or reliability.

The project serves as a practical foundation for secure, agent-driven analytics at scale.

 Autonomy, with guardrails.