# Data Quality Copilot – Requirements Document (POC Phase)

1. Overview  
The Data Quality Copilot project focuses on building a minimal, well-structured AI agent system for performing basic data quality checks on structured datasets, specifically CSV and Excel files. The primary purpose of this initiative is to understand the design and development of multi-agent systems with a focus on clean architecture, agent coordination, and simple reasoning flows. This phase is strictly a Proof-of-Concept (POC), deliberately scoped to avoid unnecessary complexity.  
  
2. Goals and Objectives  
- Develop a basic functional AI Agent system with a focus on learning agent architecture.  
- Demonstrate Planner–Executor–Critic agent roles with clean separation of concerns.  
- Focus on data parsing, schema inference, and simple anomaly detection.  
- Deliver a lightweight, well-documented codebase suitable for educational and reference purposes.  
  
3. Scope Definition  
In Scope:  
- Processing CSV/Excel files with small sample datasets.  
- Implementing foundational agent roles (Planner, Executor, Critic).  
- Basic schema detection and common anomaly checks (e.g., missing values, duplicates).  
- Generating simple console-based reports summarizing results.  
  
Out of Scope:  
- User Interface (UI) development.  
- Integration with external systems (e.g., SQL Server, APIs).  
- Handling unstructured data formats (e.g., PDF).  
- Production-grade scalability and deployment.  
- Complex business logic or domain-specific rules.

4. Deliverables  
- Minimal viable code repository with clear modularization.  
- Agent orchestration framework with defined Planner, Executor, Critic flows.  
- Sample datasets and example validation flows.  
- Architecture diagram and technical documentation explaining design decisions.  
  
5. Success Criteria  
- A working AI agent prototype demonstrating end-to-end flow from input to output.  
- Clean, easy-to-understand code focusing on agent coordination.  
- Ability to easily extend the architecture in subsequent phases.  
- High-level documentation suitable for technical onboarding.

A diagram of a data quality service

AI-generated content may be incorrect.