AI-Powered Data Processing & Analytics Tool

Project Proposal for Polio Eradication Initiative Data Team

1. Project Overview

Development of an in-house automation tool specifically designed for the **Polio Eradication Initiative (PEI) data team**. This specialized data processing and analytics platform aims to streamline operational data workflows, featuring AI integration for enhanced analysis and real-time data processing capabilities.

2. Problem Statement

The PEI data team currently faces challenges with:

- Manual processing of extensive vaccination campaign data
- Time-intensive data cleaning and standardization
- Inconsistent designation mappings across regions
- Limited real-time analysis capabilities
- Resource-intensive reporting processes

3. Proposed Solution

3.1 Core Features

- Automated data cleaning and standardization
- Dynamic designation mapping system
- Interactive data preview and filtering
- AI-powered data analysis
- Customized visualization tools for PEI metrics

3.2 Technical Stack

Core Framework

- Streamlit
- pandas
- plotly-express

AI Integration

- google-generativeai
- langchain

File Operations

- openpyxl
- python-dotenv

4. Implementation Details

4.1 Data Processing Module

- Campaign data processing
- Missing value handling
- Duplicate entry management
- PEI-specific data standardization

4.2 Mapping System

- Health worker designation mapping
- District and UC level mapping
- Bulk mapping capabilities
- Historical mapping retention

4.3 AI Integration

- Campaign performance analysis
- Coverage rate predictions
- Pattern recognition in vaccination data
- Automated report generation

4.4 Visualization Features

- Team performance metrics
- Campaign progress tracking
- Real-time monitoring dashboards

5. System Requirements

5.1 Development Environment

- Python 3.8+
- Git for version control
- Virtual environment setup

6. Project Timeline

Phase 1: Core Development (2 weeks)

- Basic file processing
- PEI-specific data cleaning
- Interface development

Phase 2: AI Integration (2 weeks)

- Gemini AI setup
- Analysis automation
- Report generation

Phase 3: Testing & Refinement (1 week)

- Data team testing
- Bug fixes
- Performance optimization

7. Expected Benefits

- 80% reduction in data processing time
- Improved campaign data accuracy
- Real-time analysis capabilities
- Streamlined reporting process
- Data-driven decision making

8. Future Enhancements

8.1 Google Sheets Integration

- Real-time data synchronization
- Collaborative editing capabilities
- Automated data pulling
- Version control for shared data
- Live dashboard updates

8.2 Advanced Analytics

- Predictive modeling for coverage rates
- Geographic information system (GIS) integration
- Team performance analytics
- Resource allocation optimization
- Trend analysis and forecasting

8.3 Enhanced AI Features

- Advanced natural language processing
- Dedicated PEI operations chatbot
- Automated report writing
- Anomaly detection
- Decision support system
- Historical data analysis
- Campaign strategy recommendations

9. Success Metrics

- Data processing time reduction
- Campaign analysis efficiency
- User adoption among team members
- Reporting accuracy improvement

- System reliability

10. Resource Requirements

- Development team
- Testing environment
- API access (Google Gemini)
- Documentation platform
- Version control system

This proposal outlines a specialized solution for the PEI data team's specific needs, with a clear path for future enhancements that will further optimize operational efficiency and decision-making capabilities.

The tool is designed to evolve with the team's needs, particularly through the planned integration of **Google Sheets** for real-time collaboration, **advanced analytics for deeper insights**, and **enhanced AI capabilities** for more sophisticated analysis and automation.