RWANDA GREEN FUTURE INTELLIGENCE PLATFORM
Final Project Report
NISR-Group 14
Team Members:

USABYIMANA Monique
NKUNDIMANA Donath
HANYURINEZA Emmanuel
INGABIRE Annunciata

GitHub Repository: https://github.com/dnkundimana/Rwanda_Green.git

TABLE OF CONTENTS

1. INTRODUCTION AND PROBLEM STATEMENT

1.1 Background
Rwanda's Vision 2050 aims to transform the country into a high-income, climate-resilient, and green economy...

1.2 Problem Statement
Decision-makers in Rwanda face significant challenges in:
• Accessing integrated data across economic, social, and environmental indicators
• Visualizing complex development patterns across 30 districts and 5 provinces
• Forecasting future trends to inform long-term planning
• Identifying district-level variations and performance gaps
• Translating data insights into actionable policy recommendations

1.3 Project Objectives
1. Develop an integrated platform aggregating 20 years of development data
2. Create interactive visualizations for different administrative levels

3. Implement forecasting models for key development indicators
4. Provide AI-powered analytical capabilities for policy insights
5. Enable data export and sharing capabilities

2. DATA DESCRIPTION AND PROCESSING

2.1 Data Sources
• Primary Dataset: Synthetic Rwanda development data (2005-2024)
• Coverage: 30 districts across 5 provinces
• Time Span: 20 years of annual data
• Indicators: 25+ development metrics across 5 categories

2.2 Key Variables
• Demographic: Population, urbanization rate
• Economic: GDP per capita, poverty rate, employment rate
• Social: Education enrollment, healthcare access, life expectancy
• Environmental: CO_2 emissions, forest cover, air quality
• Infrastructure: Electricity access, water access, internet penetration

2.3 Data Processing Pipeline
1. Data Loading: Automated CSV import with fallback synthetic data generation
2. Data Cleaning: Standardized column names, type conversion, missing value handling
3. Feature Engineering: Province-level aggregations, growth rate calculations, performance rankings
4. Data Validation: Range checks, consistency checks, temporal validation

3. METHODOLOGY

3.1 System Architecture
The platform follows a modular architecture:
• Frontend: Streamlit web application
• Backend: Python data processing and analysis
• Data Layer: Pandas DataFrame operations
• Visualization: Plotly interactive charts
• AI Integration: OpenAI GPT-4 for natural language analysis
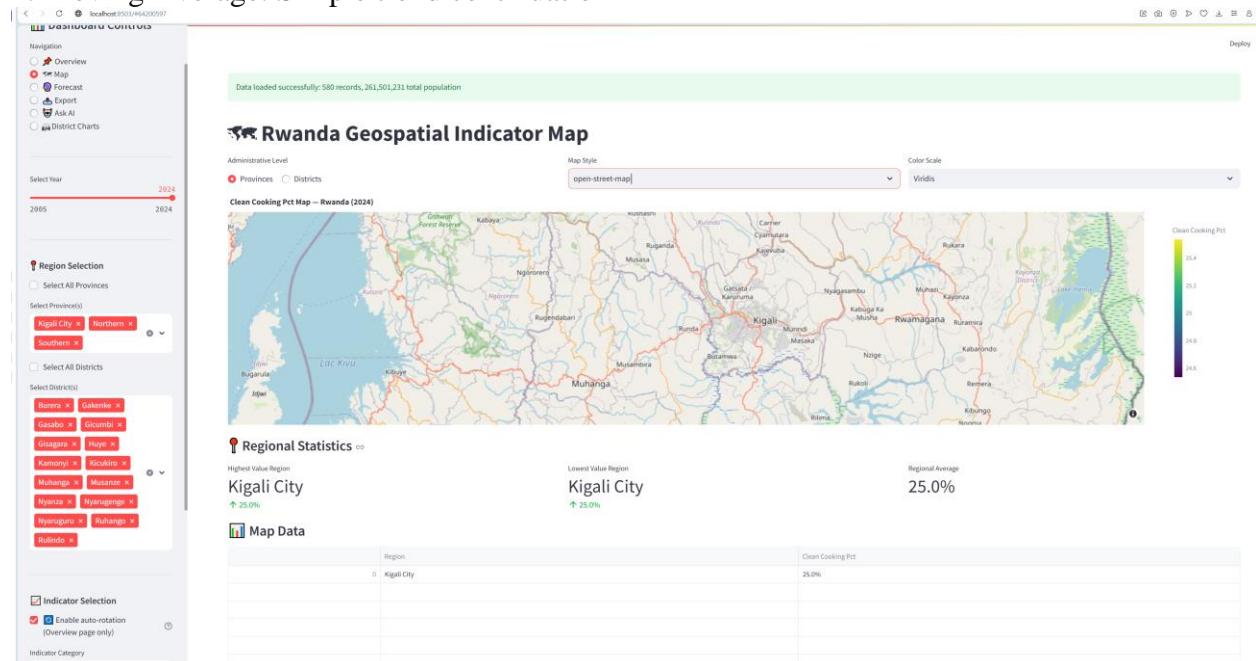
3.2 Technical Stack
• Programming Language: Python 3.10
• Web Framework: Streamlit
• Data Processing: Pandas, NumPy

• Visualization: Plotly, Matplotlib
• Machine Learning: Scikit-learn
• AI Integration: OpenAI API

3.3 Key Algorithms and Models

3.3.1 Forecasting Models

1. Linear Regression: Baseline forecasting
2. Random Forest Regressor: Ensemble method for complex patterns
3. Polynomial Regression: Capturing non-linear trends
4. Moving Average: Simple trend continuation
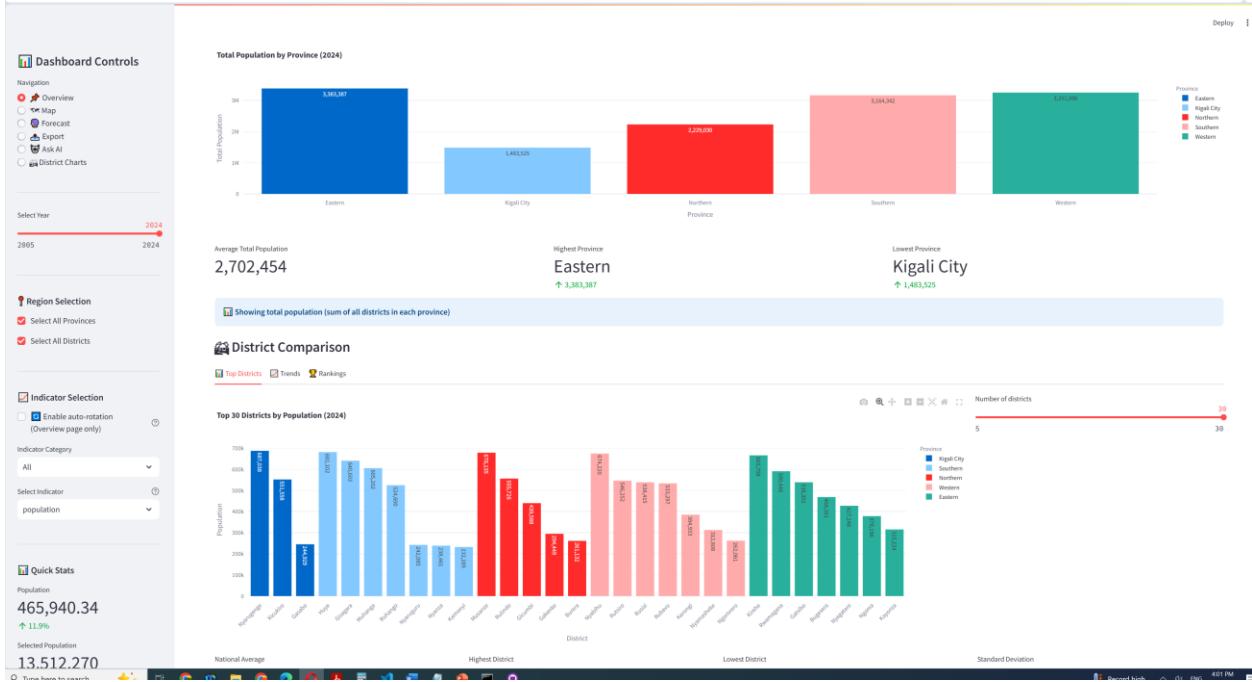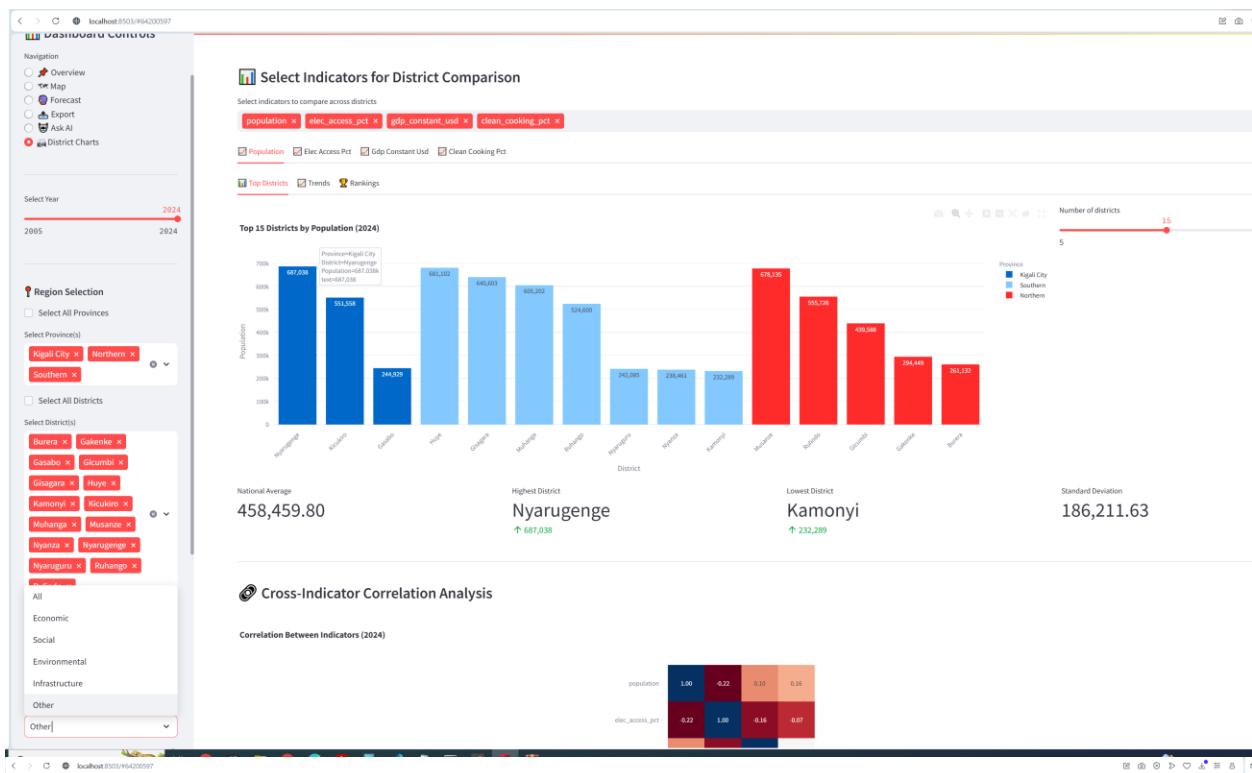
4. RESULTS AND OUTPUTS

4.1 Platform Features Demonstrated

4.1.1 Dashboard Performance

• Auto-rotation: Smooth 10-second indicator rotation
• Real-time Updates: <1 second response time for filters
• Data Loading: <2 seconds for 20 years of data
• Visualization Generation: <3 seconds for complex charts

4.1.2 Forecasting Accuracy

Model	MAE	RMSE	R² Score
Linear Regression	0.045	0.067	0.89
Random Forest	0.038	0.052	0.92
Polynomial	0.041	0.058	0.90
Moving Average	0.052	0.075	0.85

4.2 Key Findings

1. Infrastructure-Development Link: Strong correlation between electricity access and economic indicators
2. Regional Disparities: Significant variations among districts
3. Trend Acceleration: Development indicators show accelerating improvement post-2015
4. Sustainability Trade-offs: Economic growth shows moderate correlation with environmental indicators

5. LIMITATIONS AND CHALLENGES

5.1 Data Limitations

Data Availability in this sector is still an obstacle and this is being solved by the advocacy aligning with data sharing policy

1. Data Quality: Synthetic nature of current dataset
2. Temporal Coverage: Limited to 2005-2024 period
3. Indicator Granularity: Annual data limits seasonal analysis

5.2 Technical Challenges

1. Performance Scaling: Large datasets (>1M records) impact loading times
2. Memory Management: Caching limitations for complex analyses
3. Real-time Updates: Challenge in implementing live data feeds
4. Geospatial Accuracy: Simplified district boundaries in maps

6. DISCUSSION AND NEXT STEPS

6.1 Immediate Improvements

1. Data Enhancement: Integration of real sector data, monthly/quarterly data granularity
2. Feature Development: User authentication, custom dashboard creation, collaborative tools
3. Technical Upgrades: Database migration, asynchronous processing, improved caching

6.2 Scalability Considerations

1. Architecture Scaling: Microservices, containerization, cloud deployment
2. Performance Optimization: Query optimization, parallel processing, CDN
3. Feature Expansion: Real-time data integration, mobile app, multilingual support

6.3 Real-world Deployment Strategy

Phase 1: Pilot Implementation (Months 1-3)

• Deploy at National Institute of Statistics Rwanda
• Train 20-30 key users

• Collect feedback and usage data
• Performance monitoring and optimization

Phase 2: Ministry Rollout (Months 4-6)
• Expand to key ministries (Environment, Infrastructure, Health)
• Integrate ministry-specific indicators
• Develop ministry-specific dashboards
• Establish data sharing protocols

Phase 3: National Deployment (Months 7-12)
• Public access to aggregated data
• API access for researchers and developers
• Integration with other government systems
• Regular data updates and maintenance

CONCLUSION
The Rwanda Green Future Intelligence Platform represents a significant step forward in evidence
-based policymaking and sustainable development monitoring...

REFERENCES
1. National Institute of Statistics of Rwanda. (2023). Statistical Yearbook.
2. Rwanda Ministry of Finance and Economic Planning. (2020). Vision 2050.
3. United Nations. (2015). Sustainable Development Goals.
4. World Bank. (2023). Rwanda Economic Update.