FranciscarNdanu / **GROUP_9_CAPSTONE-PROJECT**

<> Code   ⊙ Issues   ⑂ Pull requests   ▷ Actions   ⊞ Projects   📖 Wiki   ⚠ Security   📈 Insights

This repository contains the final capstone project for Group 9, submitted as part of our end-of-course assessment.

☆ **0** stars   ⑂ **0** forks   ⊙ **0** watching   ⑂ Branches   ⎍ Activity   ◌ Tags

🌐 Public repository

⑂ **1 Branch**   ◌ **0 Tags**   ⑂   ◌   🔍 Go to file   t   Go to file   +   Add file ▾   Code   ⋯

◯ **FranciscarNdanu** Notebook in pdf format		c0523e4 · 7 minutes ago	◷
📁 data	Add files via upload	5 days ago	
📁 output	updated export and deleted debugg…	last week	
📄 CODE 9 - Non technical Prese…	Add files via upload	3 hours ago	
📄 CODE 9 - Non technical Prese…	Add files via upload	3 hours ago	
📄 FP_Demand_Optimization (ch…	Add files via upload	2 days ago	
📄 FP_Demand_Optimization (ch…	Add files via upload	2 days ago	
📄 FP_Demand_Optimization (ch…	Add files via upload	yesterday	
📄 FP_Demand_Optimization.ipy…	Updated notebook	yesterday	
📄 FP_Demand_Optimization.pdf	Notebook in pdf format	7 minutes ago	
📄 README.md	Create README.md	2 weeks ago	

📖 **README**   ✎ ☰

Predicting Family Planning Demand and Optimizing Service Delivery in Kenya

This project aims to leverage data science and machine learning to predict family planning (FP) demand and optimize service delivery in Kenya. By forecasting demand for specific FP methods and predicting continuation rates, this initiative seeks to enable proactive resource allocation, reduce stock-outs, and improve overall FP outcomes. This version of the project focuses specifically on a **regression-based approach**.

1. Project Structure

The project is organized into the following directories and files:

- `./`

  - `README.md` : This file.

  - `FP_Demand_Optimization.ipynb` : The main Jupyter Notebook for data analysis, modeling, and evaluation.

  - `requirements.txt` : Lists all necessary Python libraries and their versions.

  - `src/` : (Optional) Directory for reusable Python scripts (e.g., data cleaning functions, custom models).

    - `data_preparation.py`

    - `model_training.py`

  - `data/` : Contains raw and processed datasets.

    - `ke_fp_population_data.csv` : County-level demographic data.

    - `ke_fp_service_data.csv` : Provides historical family planning service delivery statistics by county and method.

    - `external_data/` : (Optional) Directory for external data sources after acquisition.

  - `models/` : (Optional) Directory to save trained machine learning models.

  - `notebooks/` : (Optional) Additional notebooks for exploration or specific analyses.

  - `reports/` : (Optional) Contains generated reports, visualizations, or presentations.

2. Setup and Installation

To set up the project environment, follow these steps:

1. **Clone the repository:**
2. **Create a virtual environment (recommended):**
3. **Install required libraries:**

3. Data

3.1 Data Sources

The project utilizes two primary datasets, located in the `data/` directory:

- `ke_fp_population_data.csv` : Contains county-level demographic information over time.

- `ke_fp_service_data.csv` : Provides historical family planning service delivery statistics by county and method.

External data sources such as DHS, Census, Guttmacher Institute data, FP Track, PMA, UN SARA, and USAID statistics will be integrated to enrich the dataset.

3.2 Data Preparation Overview

The data preparation phase involves:

- **Data Cleaning:** Handling *missing values*, addressing *outliers*, and standardizing data.

- **Feature Engineering:** Creating new features like *lagged time-series values*, *demographic ratios*, and *supply-side indicators*.

- **Target Variable Creation:** Defining the specific outcomes for prediction, focusing on regression targets:

- For **Regression**: `new_clients_method_X` (e.g., injectables, pills) and `continuation_rate_method_Y`.

- **Data Transformation:** Scaling numerical features and encoding categorical variables.

4. Methodology: Machine Learning Approach (Regression)

This project focuses on a **regression-based machine learning approach**:

4.1 Regression for Predicting Future Demand and Continuation Rates

- **Objective:** To forecast *continuous numerical values*

- **Models Used:** *Linear Regression, Random Forest Regressor, Gradient Boosting Regressor.*

- **Evaluation:** Performance will be assessed using metrics like *Mean Absolute Error (MAE)*, *Root Mean Squared Error (RMSE)*, and *R-squared (R2)*.

5. Results and Actionable Insights

The project aims to deliver:

- **Model Performance:** Quantified accuracy and reliability of predictions.

- **Feature Importance:** Identification of key drivers influencing FP demand and continuation rates.

- **Impact Quantification:** Calculation of *Couple Years of Protection (CYP)* and *averted pregnancies* based on predicted service delivery.

- **Resource Requirements:** Estimates for commodity procurement, equipment needs, and staffing allocations.

- **Policy & Programmatic Recommendations:** Data-driven insights for policymakers and program managers to optimize interventions.

6. Usage

To run the analysis and generate predictions:

1. Ensure all setup steps are completed.

2. Open the `FP_Demand_Optimization.ipynb` notebook in a Jupyter environment.

3. Execute the cells sequentially to load data, preprocess, train models, and evaluate results.

For more detailed information on specific data points, feature engineering, or model parameters, please refer to the comments within the `FP_Demand_Optimization.ipynb` notebook.

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Contributors 3

FranciscarNdanu Franciscar_Ndanu

EdgarSoudie Soudie Okwaro

RGitau-coder RGitau-RG

Languages

● **Jupyter Notebook** 100.0%