**DATA ANALYSIS CAPSTONE PROJECT**

**Postpartum Depression Among Mothers in Kenya/Africa**

**Problem**: Postpartum depression affects 13-18.7% of mothers in Kenya, with Sub-Saharan Africa showing pooled estimates of 22.1% and some regions reaching as high as 39.96% - yet it remains largely undiagnosed and untreated.

**Postpartum depression (PPD)** affects up to 1 in 7 women globally, but it's often undiagnosed in low-resource settings like Kenya due to stigma, poor mental health literacy, and limited access to maternal care. There's a need for better data-driven insights into risk factors, trends, and awareness levels.

**Data Sources:**

**1. Twitter API**

* **Use Case**: Extract tweets from Kenyan users using keywords like *"postpartum"*, *"baby blues"*, *"new mother"*, *"depression after giving birth"*.
* **Tool**: Use Tweepy to extract tweet content, location, date, sentiment.
* **Goal**: Measure awareness, stigma, and experience sharing across time.

**2. Reddit**

* **Subreddits**: /r/postpartum, /r/depression, /r/Kenya
* **Method**: Web scraping (with PRAW or BeautifulSoup) or Reddit API.
* **Goal**: Qualitative and sentiment-rich discussions from real mothers.

**3. Kenya DHS & KNBS Data**

* **Source**: [Kenya Demographic and Health Survey (DHS)](https://dhsprogram.com/)
* **What to Use**: Questions on maternal health, birth outcomes, mental health if included.
* **Method**: Download CSV datasets → Clean and analyze with Pandas.
* **Goal**: Find correlations between health access, income, and depression risk.

**4. Google Trends API**

* **Search Terms**: "Postpartum depression Kenya", "new mother sad", "baby blues"
* **Tool**: PyTrends
* **Goal**: Track public interest over time, e.g. around Mother’s Day, Mental Health Month.

**5. WHO/UNICEF Global Maternal Health Indicators**

* **API/CSV Access**: https://ghoapi.azureedge.net/api/
* **Use**: Compare PPD prevalence or mental health funding across Africa.

**APIs:**

* **WHO Global Health Observatory API** (https://ghoapi.azureedge.net/)
  + Mental health indicators by country
  + Maternal mortality and morbidity data
* **UNICEF Data API** (https://sdmx.data.unicef.org/)
  + Child health outcomes, breastfeeding rates
* **World Bank Health Data API**
  + Healthcare expenditure, facility density
* **Kenya Open Data Portal API** (https://www.opendata.go.ke/)
  + Birth registration data, healthcare facility statistics

**Web Scraping Sources:**

* **Ministry of Health Kenya** (https://www.health.go.ke/)
  + Policy documents, health facility reports
* **Kenya Medical Research Institute (KEMRI)** publications
* **Daily Nation/The Standard Health sections**
  + Public awareness articles, case studies
* **Nairobi County health department reports**
* **Private hospitals websites** (Nairobi Hospital, Aga Khan, etc.)
  + Service availability, maternal health programs

**Database Sources:**

* **Kenya Demographic Health Survey (KDHS)** data
* **NHIF maternal health claims data**
* **Birth registry data from civil registration**
* **Health facility surveys from MOH**

**Specific Research Angles:**

**1. Early Detection System**

* Analyze patterns in prenatal care attendance vs. PPD risk
* Build predictive model using maternal health indicators

**2. Treatment Access Gap Analysis**

* Map mental health facilities vs. maternal health centers
* Analyze geographical disparities in care availability

**3. Economic Impact Assessment**

* Calculate costs of untreated PPD (healthcare, lost productivity)
* ROI analysis of screening and treatment programs

**4. Cultural Barriers Investigation**

* Analyze social media sentiment around maternal mental health
* Study traditional vs. modern healthcare seeking behaviors

**Multi-Source Analysis Framework:**

**Primary Analysis (Real-Time Data):**

├── API data (WHO, Kenya Open Data)

├── Web scraped data (news, reports)

**Secondary Analysis (Supplementary Datasets):**

├── Kaggle PPD datasets → Model training/validation

├── UCI maternal health → Risk factor correlation

├── GitHub mental health collections → Comparative analysis

├── Research datasets → Academic benchmarking

**Tertiary Analysis (Context):**

├── Social media data → Public sentiment

├── Government surveys → Policy context

└── International data → Global comparison

**Analysis Ideas**

* Use DialogFlow
* Regional mapping of maternal mental health discussion volume
* Time-series of awareness (Google Trends)
* Correlation between healthcare access and risk of PPD (KNBS/DHS)
* Keyword frequency trends (e.g. “sad”, “crying”, “alone”)

**Expected Impact**

* Identify gaps in maternal mental health support.
* Provide actionable insights to NGOs and public health ministries.
* Promote awareness campaigns targeting new mothers.
* Support development of chatbots or helplines for anonymous support.