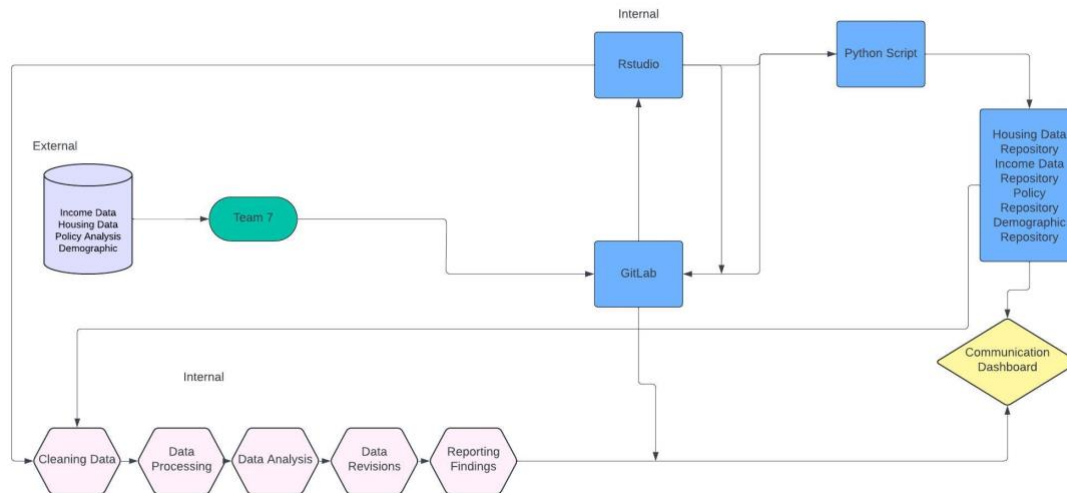# Seattle Housing Crisis Analysis

Part 1: **Fatima & Thao**
Data Flow Diagram (DFD)

Markdown Summary

1. **Housing Characteristics Data**
   - **Source**: Seattle Housing Characteristics Data from Seattle.gov
   - **Description**: Contains information on housing units, including average costs and unit counts by neighborhood.
2. **Rent Burden Data**
   - **Source**: Rent Burden Data from Seattle.gov
   - **Description**: Includes records on income levels and the percentage of income spent on rent.
3. **Policy Analysis Data**
   - **Source**: MHA Overview PDF from Seattle.gov
   - **Description**: Contains information about housing policies and their effectiveness scores.
4. **Demographic Data**
   - **Source**: Seattle GeoData
   - **Description**: Provides demographic statistics for neighborhoods, including median income and population counts.

Special Properties

- **Housing Characteristics**: Focuses on average housing costs and availability across different neighborhoods.

- **Rent Burden**: Highlights the economic stress experienced by residents in terms of housing costs relative to income.
- **Policy Analysis**: Evaluates the impact of housing policies on affordability and accessibility.
- **Demographic Data**: Offers insights into income distributions and population characteristics.

Proposed Wrangling Techniques

- **Data Cleaning**:
    - o   Handle missing values and standardize formats (e.g., currency formats).
- **Data Transformation**:
    - o   Convert categorical data (e.g., neighborhoods, income levels) into numerical formats for analysis.
    - **Join Techniques**:
        - ▪   Connect **Housing Characteristics** with **Demographic Data** using the Neighborhood field to analyze how demographic factors influence housing characteristics.
        - ▪   Merge **Rent Burden** with **Demographic Data** on Income_Level to assess rent burdens across different income groups.
- **Reason for Joins**: These joins will enable a multifaceted analysis of how housing costs, income levels, and policies interact and impact housing insecurity.

Storage

- Data will be stored in a **relational database** (e.g., PostgreSQL, R studio, Python Script) to facilitate efficient querying and analysis.
- Each dataset will be stored in separate tables, ensuring normalization and ease of access for analysis.

Data Sources Using:

- https://data.seattle.gov/dataset/Affordable-and-Available-Rental-Units-per-100-Hous/q9hb-qmcp/about_data
- https://data.seattle.gov/dataset/Incomes-Occupations-and-Earnings-Seattle-Neighborh/cm78-ajfy/about_data
- https://data.seattle.gov/dataset/Housing-Tenure-and-Costs-Seattle-Neighborhoods/r4hu-gxdi/about_data