Ghulam Ishaq Khan Institute of Engineering Sciences and Technology, Topi, Pakistan.

Group Members

Muhammad Adeel	2022331
M. Hamza Mehmood Zaidi	2022379
Nauman Ali Murad	2022479
Hassan Rais	2022212

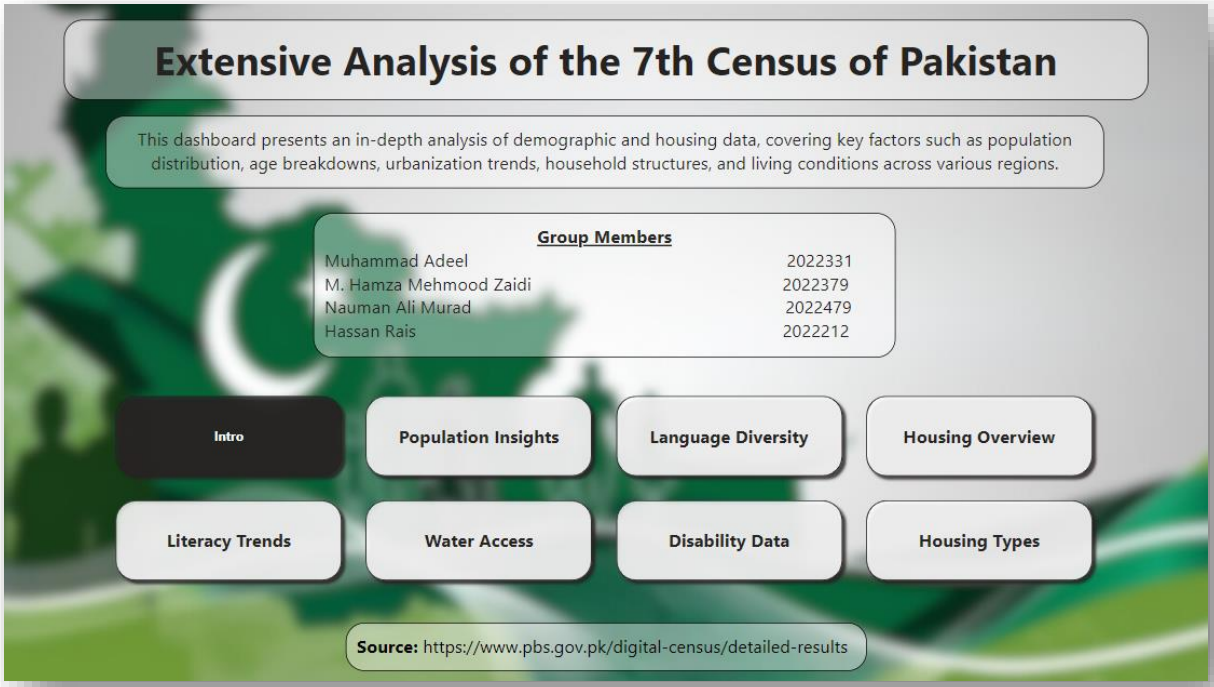
Course: DS331 – Data Warehousing and Business Intelligence (Theory & Lab)

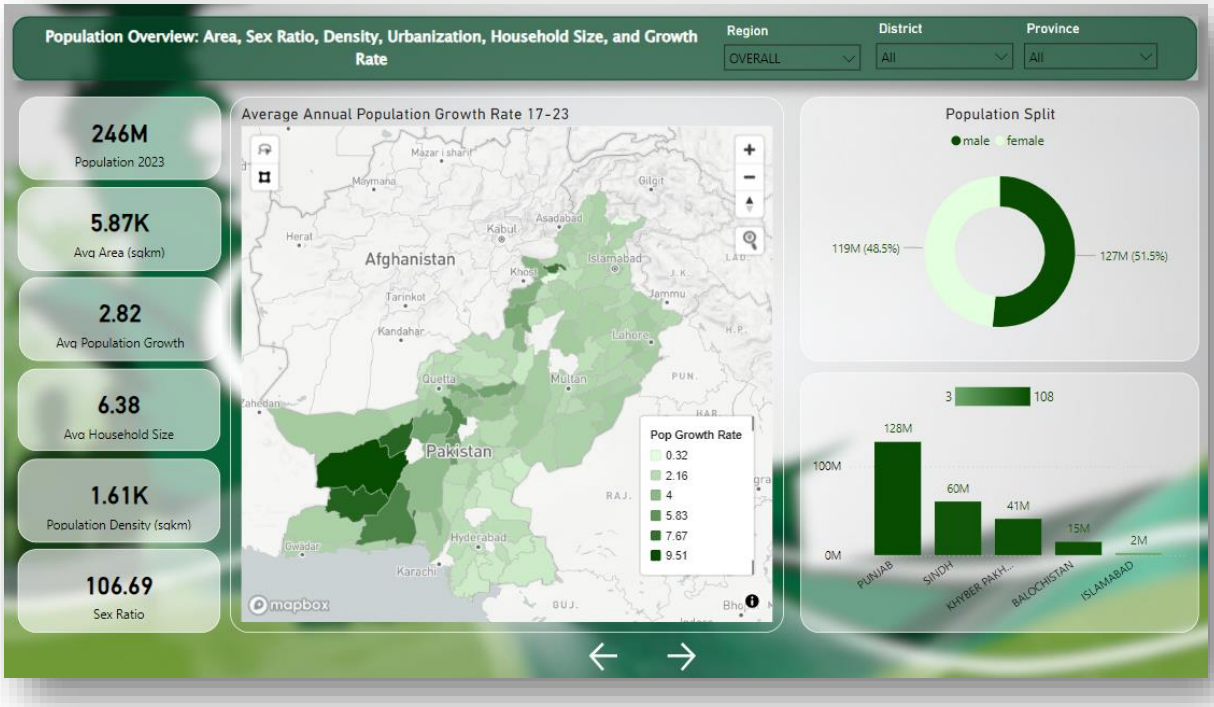
Instructors: Dr. Ayaz Umer & Engr. Hafiz Syed M. Muslim

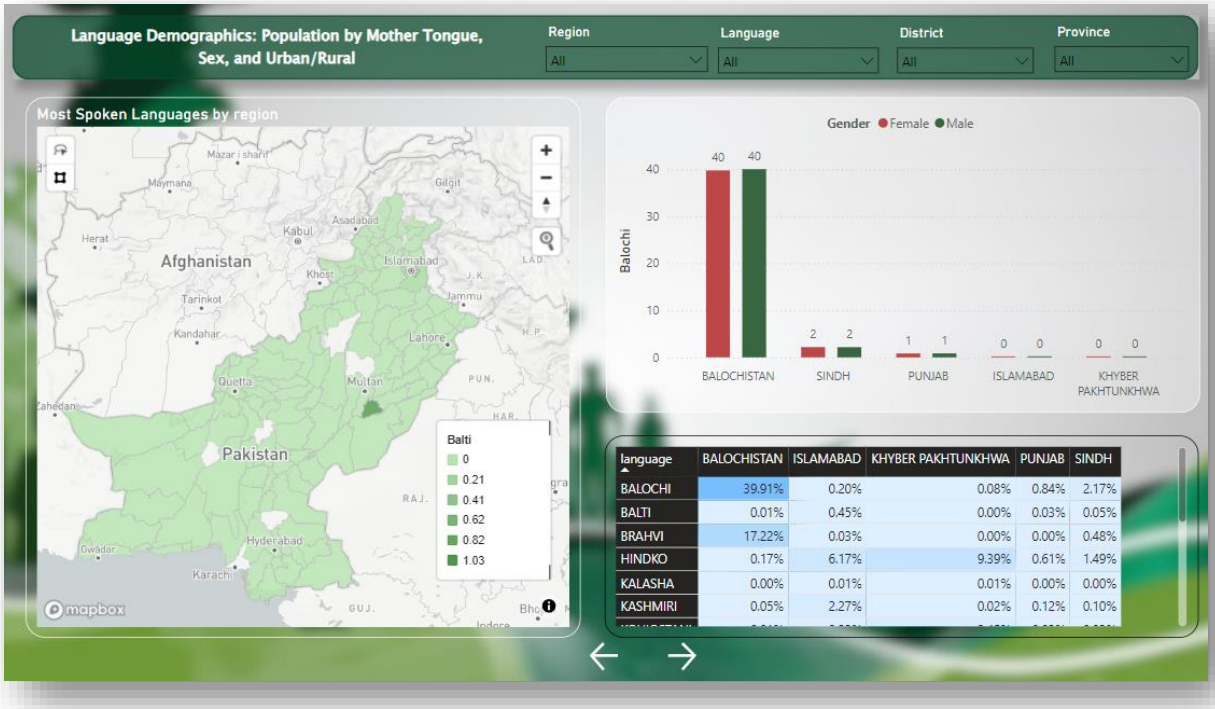
Project Title: Analysis of the 7th Population Census of Pakistan

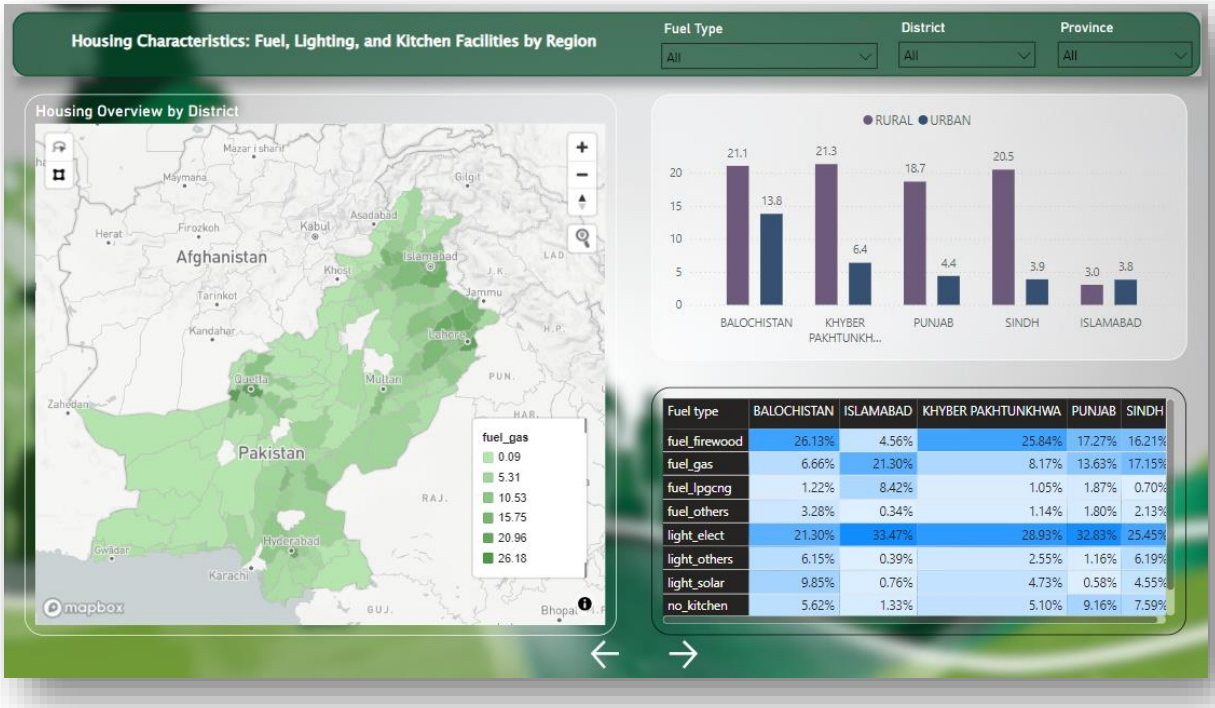
Submission Date: 9/12/2024

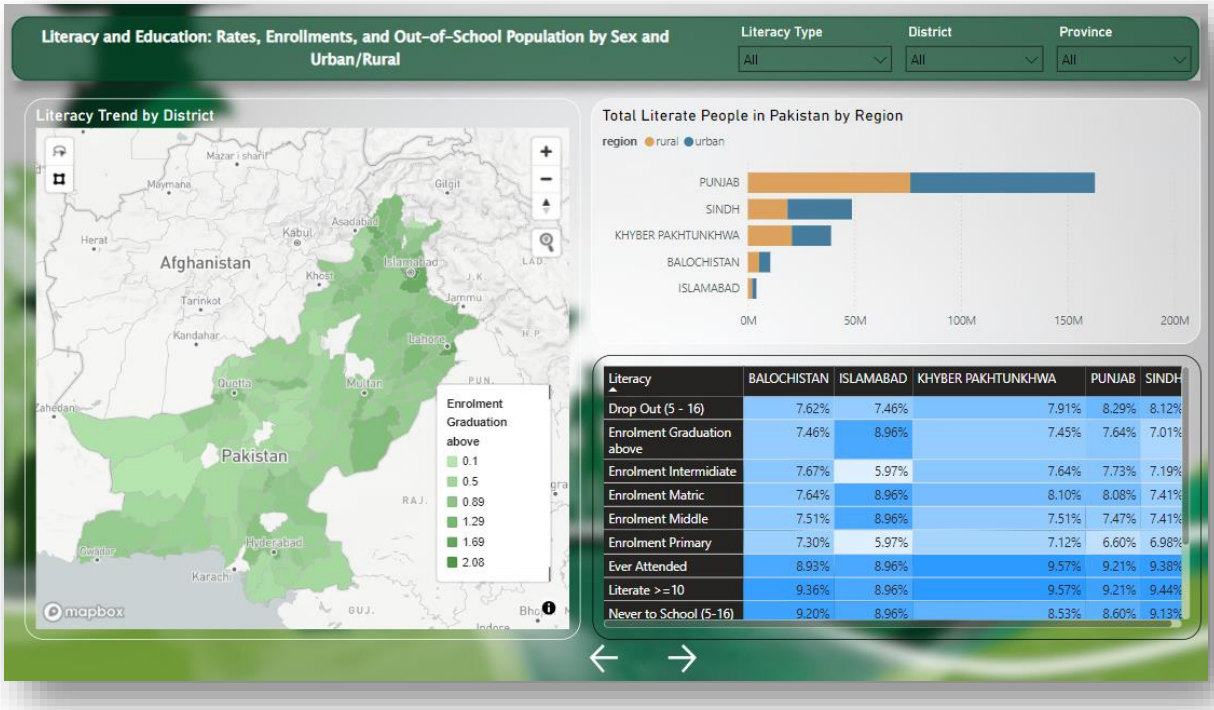
Dashboard Snapshots:

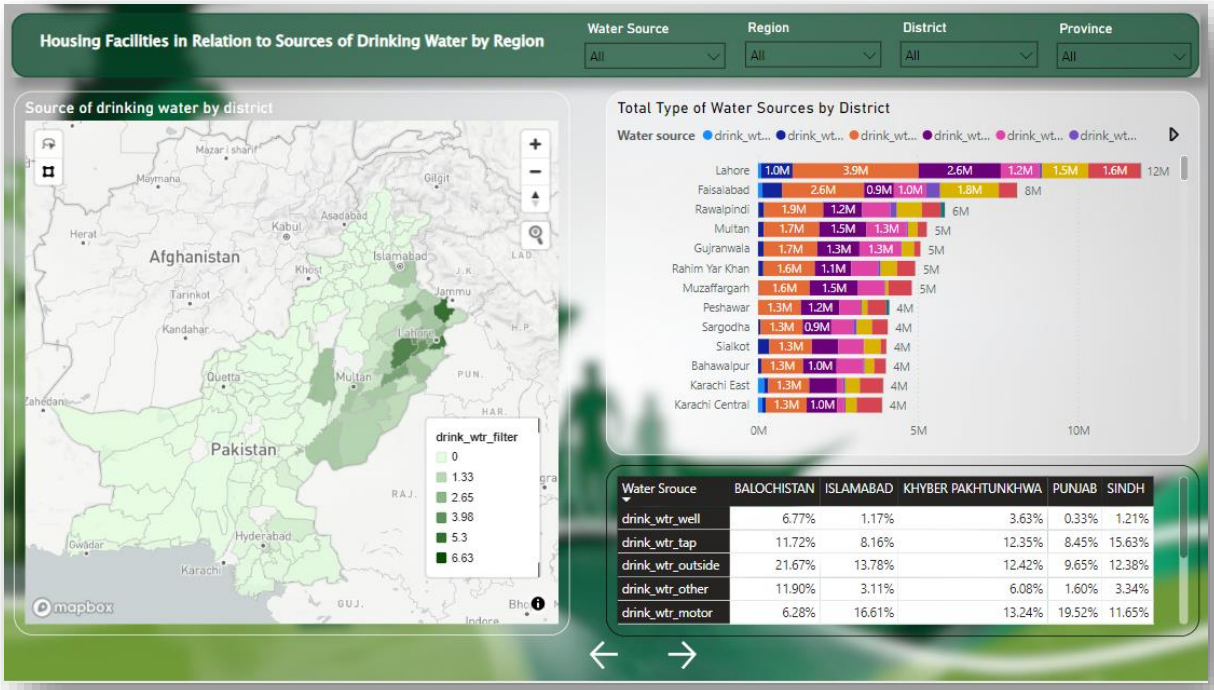


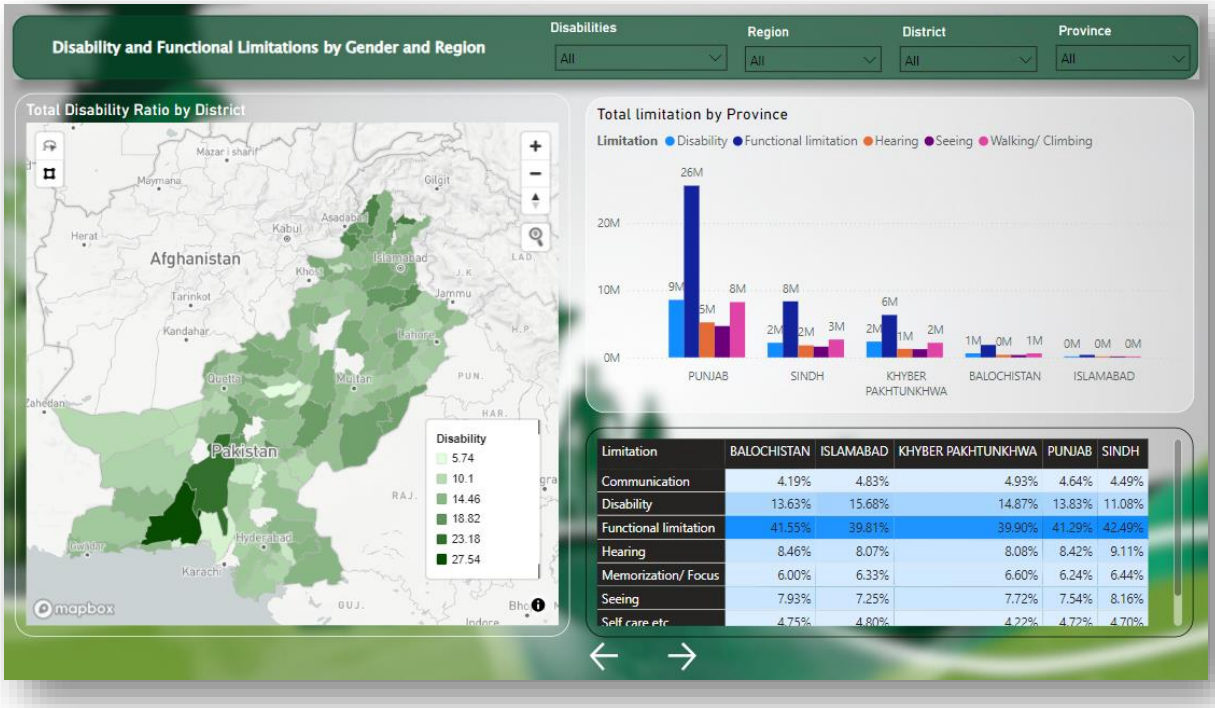


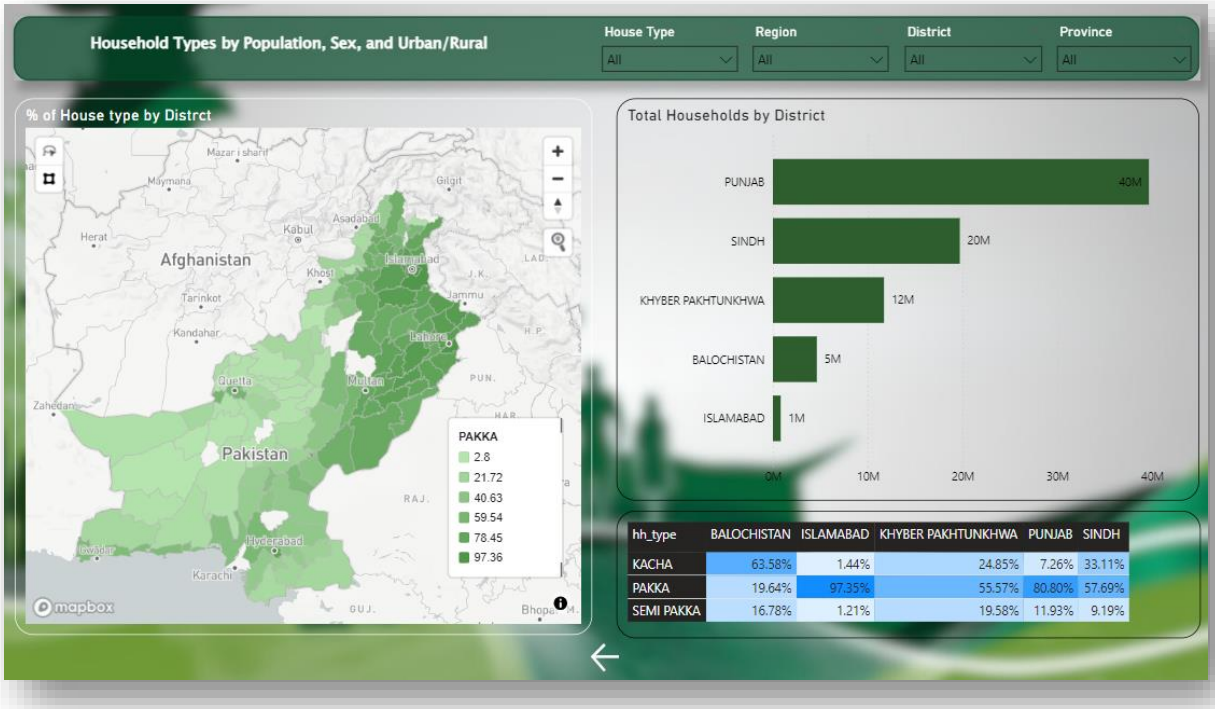

Language Demographics: Population by Mother Tongue, Sex, and Urban/Rural
Region
All
Language
All
District
All
Province
All
Most Spoken Languages by region
Afghanistan
Pakistan
Balti
0
0.21
0.41
0.62
0.82
1.03
Gender Female Male
Balochi
40
40
2
2
1
1
0
0
0
0
BALOCHISTAN
SINDH
PUNJAB
ISLAMABAD
KHYBER PAKHTUNKHWA
language | BALOCHISTAN | ISLAMABAD | KHYBER PAKHTUNKHWA | PUNJAB | SINDH
BALOCHI | 39.91% | 0.20% | 0.08% | 0.84% | 2.17%
BALTI | 0.01% | 0.45% | 0.00% | 0.03% | 0.05%
BRAHVI | 17.22% | 0.03% | 0.00% | 0.00% | 0.48%
HINDKO | 0.17% | 6.17% | 9.39% | 0.61% | 1.49%
KALASHA | 0.00% | 0.01% | 0.01% | 0.00% | 0.00%
KASHMIRI | 0.05% | 2.27% | 0.02% | 0.12% | 0.10%


Housing Characteristics: Fuel, Lighting, and Kitchen Facilities by Region
Fuel Type
All
District
All
Province
All
Housing Overview by District
Afghanistan
Pakistan
fuel_gas
0.09
5.31
10.53
15.75
20.96
26.18
RURAL URBAN
21.1
13.8
21.3
6.4
18.7
4.4
20.5
3.9
3.0
3.8
BALOCHISTAN
KHYBER PAKHTUNKH...
PUNJAB
SINDH
ISLAMABAD
Fuel type | BALOCHISTAN | ISLAMABAD | KHYBER PAKHTUNKHWA | PUNJAB | SINDH
fuel_firewood | 26.13% | 4.56% | 25.84% | 17.27% | 16.21%
fuel_gas | 6.66% | 21.30% | 8.17% | 13.63% | 17.15%
fuel_lpgcng | 1.22% | 8.42% | 1.05% | 1.87% | 0.70%
fuel_others | 3.28% | 0.34% | 1.14% | 1.80% | 2.13%
light_elect | 21.30% | 33.47% | 28.93% | 32.83% | 25.45%
light_others | 6.15% | 0.39% | 2.55% | 1.16% | 6.19%
light_solar | 9.85% | 0.76% | 4.73% | 0.58% | 4.55%
no_kitchen | 5.62% | 1.33% | 5.10% | 9.16% | 7.59%


Literacy and Education: Rates, Enrollments, and Out-of-School Population by Sex and Urban/Rural
Literacy Type
District
Province
Literacy Trend by District
Total Literate People in Pakistan by Region
region rural urban
PUNJAB
SINDH
KHYBER PAKHTUNKHWA
BALOCHISTAN
ISLAMABAD
0M
50M
100M
150M
200M
Enrolment Graduation above
0.1
0.5
0.89
1.29
1.69
2.08
Literacy | BALOCHISTAN | ISLAMABAD | KHYBER PAKHTUNKHWA | PUNJAB | SINDH
Drop Out (5 - 16) | 7.62% | 7.46% | 7.91% | 8.29% | 8.12%
Enrolment Graduation above | 7.46% | 8.96% | 7.45% | 7.64% | 7.01%
Enrolment Intermidiate | 7.67% | 5.97% | 7.64% | 7.73% | 7.19%
Enrolment Matric | 7.64% | 8.96% | 8.10% | 8.08% | 7.41%
Enrolment Middle | 7.51% | 8.96% | 7.51% | 7.47% | 7.41%
Enrolment Primary | 7.30% | 5.97% | 7.12% | 6.60% | 6.98%
Ever Attended | 8.93% | 8.96% | 9.57% | 9.21% | 9.38%
Literate >=10 | 9.36% | 8.96% | 9.57% | 9.21% | 9.44%
Never to School (5-16) | 9.20% | 8.96% | 8.53% | 8.60% | 9.13%


Housing Facilities in Relation to Sources of Drinking Water by Region
Water Source
Region
District
Province
Source of drinking water by district
Total Type of Water Sources by District
Water source drink_wt... drink_wt... drink_wt... drink_wt... drink_wt... drink_wt...
Lahore 1.0M 3.9M 2.6M 1.2M 1.5M 1.6M 12M
Faisalabad 2.6M 0.9M 1.0M 1.8M 8M
Rawalpindi 1.9M 1.2M 6M
Multan 1.7M 1.5M 1.3M 5M
Gujranwala 1.7M 1.3M 1.3M 5M
Rahim Yar Khan 1.6M 1.1M 5M
Muzaffargarh 1.6M 1.5M 5M
Peshawar 1.3M 1.2M 4M
Sargodha 1.3M 0.9M 4M
Sialkot 1.3M 4M
Bahawalpur 1.3M 1.0M 4M
Karachi East 1.3M 4M
Karachi Central 1.3M 1.0M 4M
0M
5M
10M
drink_wtr_filter
0
1.33
2.65
3.98
5.3
6.63
Water Srouce | BALOCHISTAN | ISLAMABAD | KHYBER PAKHTUNKHWA | PUNJAB | SINDH
drink_wtr_well | 6.77% | 1.17% | 3.63% | 0.33% | 1.21%
drink_wtr_tap | 11.72% | 8.16% | 12.35% | 8.45% | 15.63%
drink_wtr_outside | 21.67% | 13.78% | 12.42% | 9.65% | 12.38%
drink_wtr_other | 11.90% | 3.11% | 6.08% | 1.60% | 3.34%
drink_wtr_motor | 6.28% | 16.61% | 13.24% | 19.52% | 11.65%


Disability and Functional Limitations by Gender and Region
Disabilities
All
Region
All
District
All
Province
All
Total Disability Ratio by District
Afghanistan
Pakistan
Disability
5.74
10.1
14.46
18.82
23.18
27.54
Total limitation by Province
Limitation Disability Functional limitation Hearing Seeing Walking/ Climbing
26M
20M
10M
0M
PUNJAB
SINDH
KHYBER PAKHTUNKHWA
BALOCHISTAN
ISLAMABAD
Limitation | BALOCHISTAN | ISLAMABAD | KHYBER PAKHTUNKHWA | PUNJAB | SINDH
Communication | 4.19% | 4.83% | 4.93% | 4.64% | 4.49%
Disability | 13.63% | 15.68% | 14.87% | 13.83% | 11.08%
Functional limitation | 41.55% | 39.81% | 39.90% | 41.29% | 42.49%
Hearing | 8.46% | 8.07% | 8.08% | 8.42% | 9.11%
Memorization/ Focus | 6.00% | 6.33% | 6.60% | 6.24% | 6.44%
Seeing | 7.93% | 7.25% | 7.72% | 7.54% | 8.16%
Self care etc | 4.75% | 4.80% | 4.22% | 4.72% | 4.70%


Household Types by Population, Sex, and Urban/Rural
House Type
All
Region
All
District
All
Province
All
% of House type by Distrct
Afghanistan
Pakistan
PAKKA
2.8
21.72
40.63
59.54
78.45
97.36
Total Households by District
PUNJAB 40M
SINDH 20M
KHYBER PAKHTUNKHWA 12M
BALOCHISTAN 5M
ISLAMABAD 1M
0M 10M 20M 30M 40M
hh_type | BALOCHISTAN | ISLAMABAD | KHYBER PAKHTUNKHWA | PUNJAB | SINDH
KACHA | 63.58% | 1.44% | 24.85% | 7.26% | 33.11%
PAKKA | 19.64% | 97.35% | 55.57% | 80.80% | 57.69%
SEMI PAKKA | 16.78% | 1.21% | 19.58% | 11.93% | 9.19%

Executive Summary

The **2023 Digital Census of Pakistan** provides an unprecedented look into the nation's demographic, linguistic, and socio-economic dynamics. This report explores key trends, disparities, and opportunities identified through an extensive analysis performed using **Power BI**, incorporating **DAX queries**, **Mapbox visualizations**, and parameterized metrics.

Key findings highlight stark inequalities in education, energy access, water infrastructure, and housing conditions, as well as the linguistic and cultural diversity across provinces. Insights emphasize critical areas for intervention, including equitable resource distribution, sustainable energy solutions, and education reforms.

Dataset Overview

The dataset, sourced from the **Pakistan Bureau of Statistics Digital Census Results**, provides granular insights into:

- **Population demographics:** Including total population (246M), gender distribution (male: 127M, female: 119M), population growth (2.82%), and sex ratio (106.69 males per 100 females).
- **Housing and infrastructure:** Types of housing (Pakka, Kacha, Semi-Pakka), household size (average 6.38), and access to utilities.
- **Water sources:** Usage of filtered, piped, and motorized pumps across provinces.
- **Languages:** Regional linguistic patterns highlighting cultural diversity.
- **Geographical details:** Land area (5.87K sq km) and population density (1.61K/sq km).

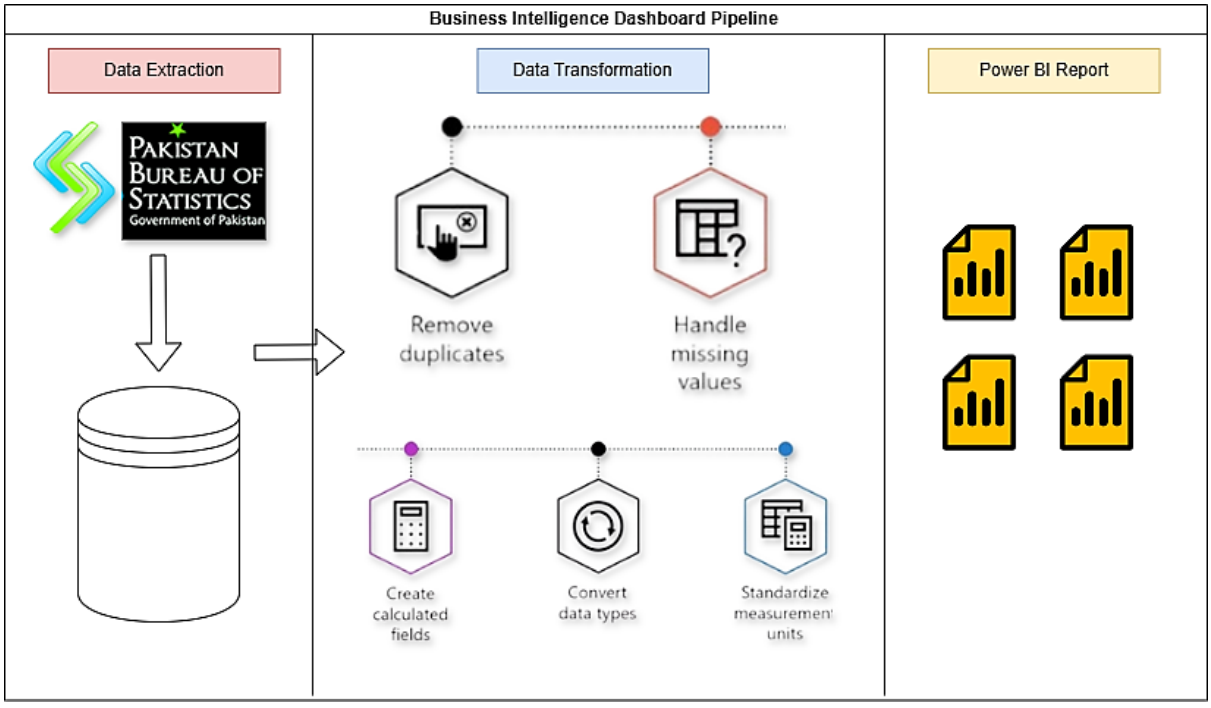
The dataset has the following attributes:

1. **Building Structures**
2. **Disabilities**
3. **House Structure**
4. **Housing Characteristics: Fuel, Lighting, etc.**
5. **Language**
6. **Literacy**
7. **Population Insights**
8. **Source of Drinking Water**

The dataset offers a balanced view of urban and rural dynamics, providing a foundation for targeted policymaking.

Business Intelligence Pipeline for Census Data Analysis

- **Purpose:** The pipeline is designed to analyse data from the 7th Population Census of Pakistan.
- **Step 1: Data Extraction**
  - Raw census data is sourced from the Pakistan Bureau of Statistics.
  - The extracted data is stored in a centralized database.
- **Step 2: Data Transformation**
  - Ensures data quality through:
    - Removing duplicates.
    - Handling missing values.
    - Creating calculated fields.
    - Converting data types.
    - Standardizing measurement units.
- **Step 3: Data Visualization**
  - Transformed data is visualized using Power BI.
  - Interactive reports are generated to provide meaningful insights into the census data.
- **Outcome**: This pipeline effectively streamlines the process from raw census data to actionable insights, enhancing data analysis and decision-making.



Data Preprocessing

The dataset underwent rigorous preprocessing to ensure reliability and accuracy:

1. **Cleaning:** Addressed missing values, standardized province and language names, and removed outliers.
2. **Transformation:** Aggregated linguistic and housing data for provincial comparisons.
3. **Categorization:** Segmented population data into urban and rural categories.
4. **Mapping:** Geo-referenced water infrastructure and housing data for enhanced visualization.

Feature Selection

Selected features reflect critical socio-economic and infrastructural parameters:

1. **Housing Type:** Distinguishing Pakka (55% in KP), Kacha (24.85%), and Semi-Pakka (19.58%) housing for regional development insights.
2. **Languages:** Highlighting linguistic diversity, e.g., Punjabi dominance in North Punjab vs. Saraiki in South Punjab, and Balochi's prevalence in Balochistan (39.91%).
3. **Water Access:** Examining reliance on improved sources (e.g., Sindh: 32.57%, Balochistan: 21.67% outside sources).
4. **Energy Access:** Insights into solar panel usage (7% nationally, high in Balochistan rural areas).

These features were selected for their relevance to understanding disparities and resource utilization patterns.

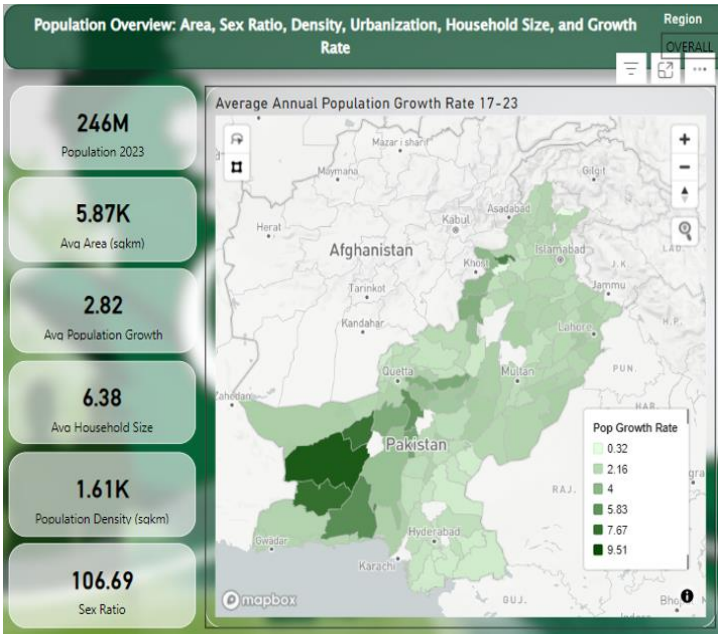
Analysis

Descriptive Statistics

- **Population:**
    - **2023 Population:** 246M (127M males, 119M females).
    - **Population Density:** 1.61K/sq km.
    - **Growth Rate:** 2.82% (fig.1).
- **Housing:**
    - In KP, **55%** of houses are Pakka, followed by Kacha (24.85%) and Semi-Pakka (19.58%) (fig.2).
    - Urban areas lead in solar adoption (e.g., 33% in rural Baluchistan) (fig.3).

- **Water Access:**
  - Sindh: High use of filtered water (3.54%) and improved sources (32.57%).
  - Islamabad: High reliance on motorized pumps (16.61%).
  - Balochistan: Heavy dependence on outside sources (21.67%) (fig.4).
- **Languages:**
  - North Punjab: Predominantly Punjabi (fig.5).
  - South Punjab: Saraiki dominance (fig.6).
  - Balochistan: Balochi (39.91%), Brahvi (17.22%), and Pashto (2.10%) (fig.7).

Visual Representations



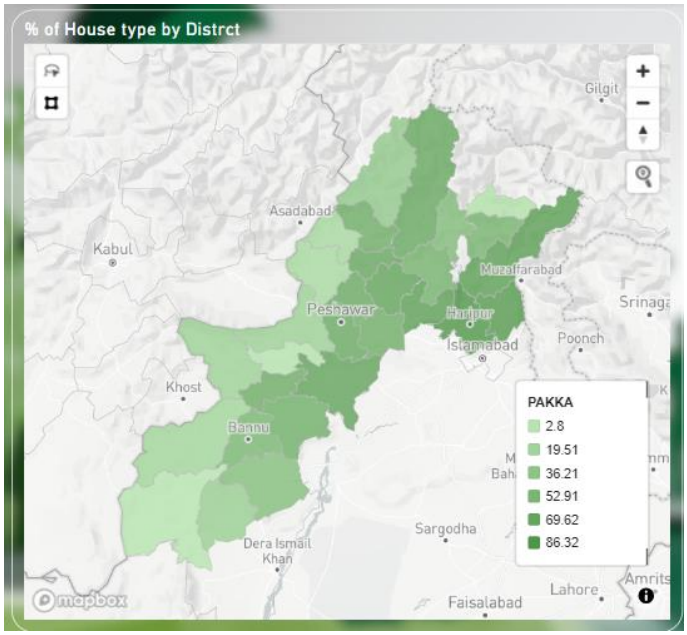
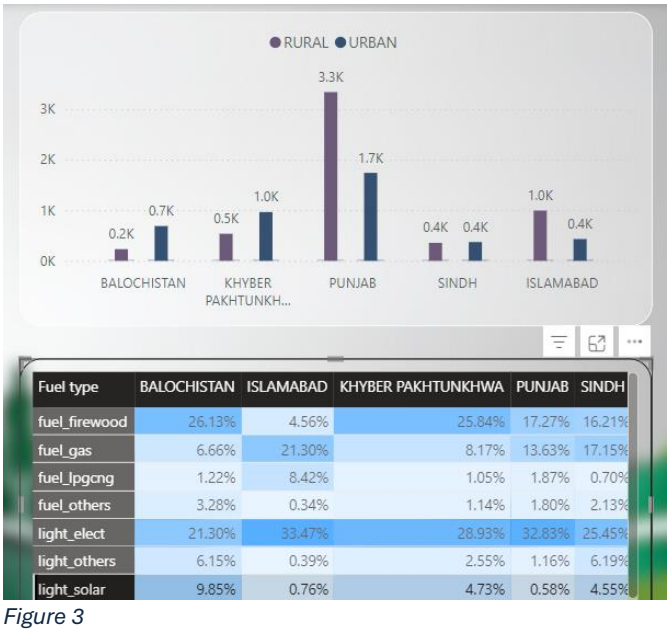


Figure 2



Fuel type	BALOCHISTAN	ISLAMABAD	KHYBER PAKHTUNKHWA	PUNJAB	SINDH
fuel_firewood	26.13%	4.56%	25.84%	17.27%	16.21%
fuel_gas	6.66%	21.30%	8.17%	13.63%	17.15%
fuel_lpgcng	1.22%	8.42%	1.05%	1.87%	0.70%
fuel_others	3.28%	0.34%	1.14%	1.80%	2.13%
light_elect	21.30%	33.47%	28.93%	32.83%	25.45%
light_others	6.15%	0.39%	2.55%	1.16%	6.19%
light_solar	9.85%	0.76%	4.73%	0.58%	4.55%

Figure 3

Water Srouce	BALOCHISTAN	ISLAMABAD	KHYBER PAKHTUNKHWA	PUNJAB	SINDH
drink_wtr_well	6.77%	1.17%	3.63%	0.33%	1.21%
drink_wtr_tap	11.72%	8.16%	12.35%	8.45%	15.63%
drink_wtr_outside	21.67%	13.78%	12.42%	9.65%	12.38%
drink_wtr_other	11.90%	3.11%	6.08%	1.60%	3.34%
drink_wtr_motor	6.28%	16.61%	13.24%	19.52%	11.65%
drink_wtr_inside	15.17%	19.93%	23.01%	24.06%	21.76%
drink_wtr_improve	26.32%	32.57%	29.14%	32.57%	31.72%
drink_wtr_filter	0.05%	3.77%	0.11%	3.54%	1.02%
drink_wtr_bottle	0.12%	0.89%	0.03%	0.28%	1.29%

Figure 4

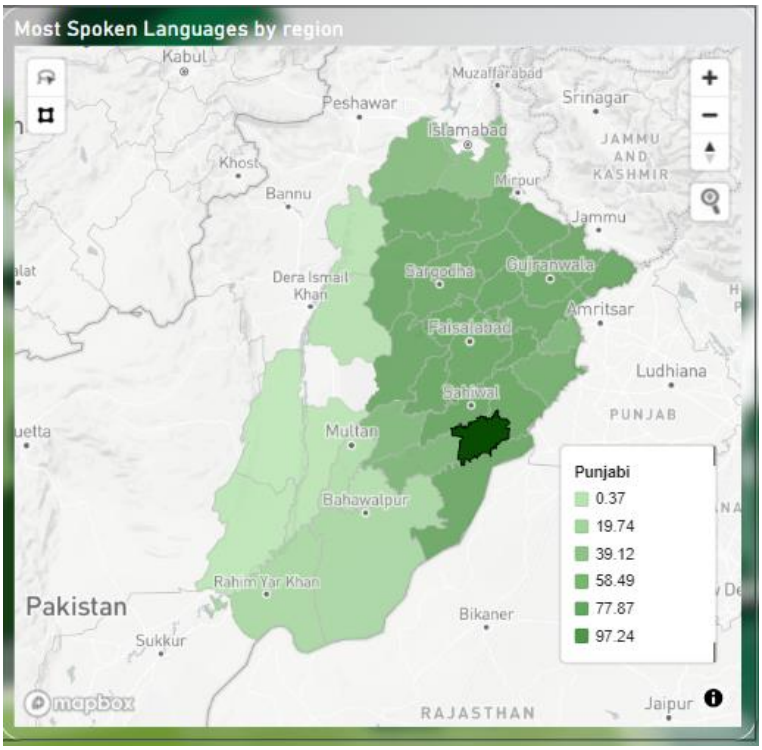


Figure 5

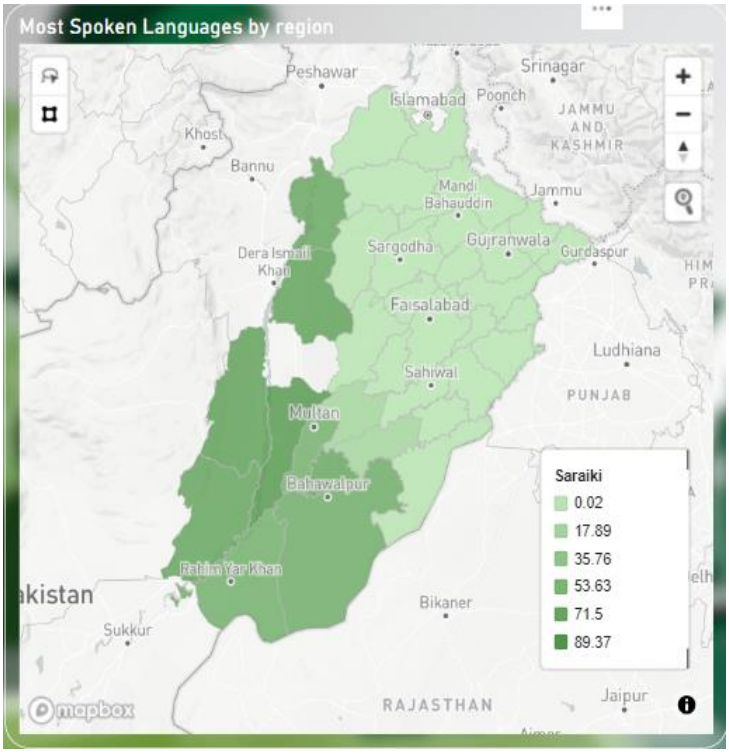


Figure 6

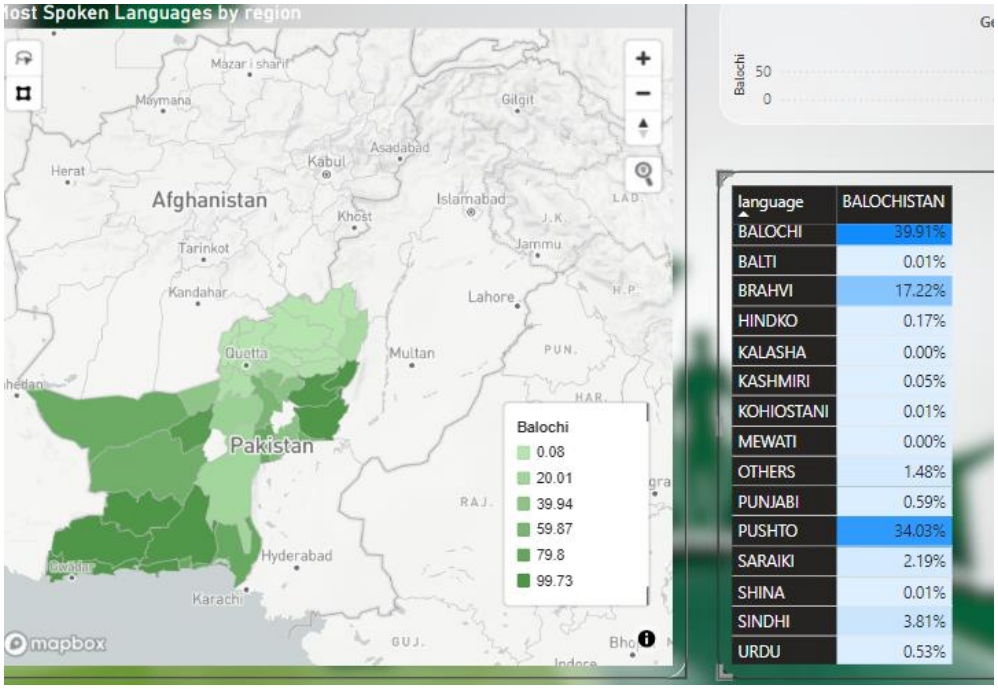


Figure 7

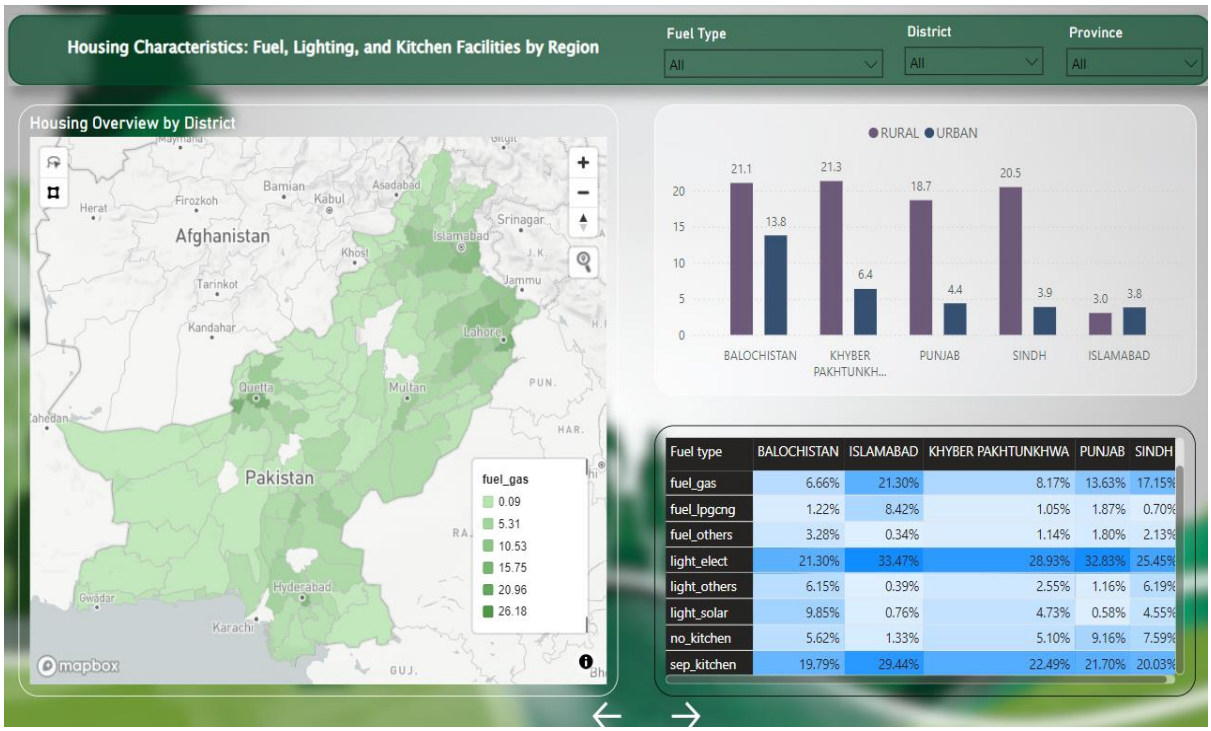


Figure 8

Insights and Recommendations

Key Insights

1. **Housing and Energy:**
    - The reliance on wood for cooking (over 50%) and low solar adoption indicate a need for sustainable energy solutions.
2. **Water Access:**
    - Sindh's high use of filtered water reflects water scarcity, while Balochistan's reliance on external sources underscores infrastructure gaps.
3. **Linguistic Diversity:**
    - Distinct linguistic zones in Punjab and Balochistan highlight the need for culturally tailored education and communication policies.
4. **Population and Growth:**
    - Pakistan's population growth (2.82%) and density (1.61K/sq km) demand urgent urban planning and resource optimization.

Recommendations

1. **Sustainable Energy Policies:**
    - Expand solar energy programs in rural areas, leveraging Balochistan's high adoption rates as a model.
2. **Water Infrastructure Development:**
    - Invest in piped water systems and filtration facilities in underserved regions like Balochistan and Sindh.
3. **Targeted Education Programs:**
    - Address educational disparities through province-specific programs, prioritizing rural Balochistan and South Punjab.
4. **Urban Housing Policies:**
    - Enhance rental housing markets in urban centers like Islamabad to meet growing demand.

Conclusion

The **2023 Digital Census** provides a detailed lens into Pakistan's evolving socio-economic landscape. By addressing critical disparities in housing, energy, water, and education, policymakers can drive meaningful change for sustainable development.