**[Weekly Report]**

* Report No.2 (updated week1)
* Date: Nov. 08, 2023~ Nov. 14, 2023
* Team: Yang Junseob (2019034639), Ryu Seung Gwon (2019087147)

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Highlights | Self  Evaluation |
| Last  Week | Summary | * Conducting Exploratory Data Analysis to examine key features of each file | Medium |
| Current  Week | Baseline Goals (Given) | * Reconduct data cleansing and manipulation each file. * Determine the reference area(s) to be compared in ACS | Medium |
| Additional Goals (O.Y.O) | * checking the logical relationships between variables | Medium |
| Key Issues to Be Resolved | * Several duplicate data included in geo1 and geo3 datasets. * Some missing values included in geo3’s columns. * Outliers included in geo3. * Detecting Logical Violation between indicator and property type | Medium |
| Strategies | * Step 1: Reconduct EDA to examine key characteristic of variables and to remove outliers, duplicate and Missing values. * Step 2: Identify logical violations and logical relationship. * Step 3: Select Reference area to create a list of queries in ACS | Medium |
| Results | * In 2017 and 2018, there were logical violations of indicator in 566 and 437 obs each. This is presumed to be due to a change in housing type. On the contrary, in 2019, all rows are matched, suggesting that the measurement point for geo2 is in 2019. * Remove Outlier “Unknown” values in geo3. * Merge duplicate values and delete missing values (geo1: 153021 obs -> 77101 obs, geo3: 1210596 obs -> 638270 obs) * Set Reference area to Chicago city, Cook County, Illinois * Create list of filtered data in ACS * comparing housing characteristics of 5 communities in Chicago and the reference area, The proportion of multi-unit houses in the reference area is relatively higher. | Medium |
| Implications | * It helps us prepare the data for further analysis or modeling. | Medium |
| Next  Week | Things to Do | * Based upon those queries, we will extract data from the ACS. * Select Model for Analysis | Medium |
| Remarks | Core Libraries & Packages | * Data Cleansing: dplyr * Visualization: ggplot2 | Medium |
|  | Additional  Remarks |  |  |

* Remarks
* Visualizations and Results are given in following Appendix.
* Work Progress

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Week 1** | **Week 2** | **Week 3** | **Week 4** | **Week 5** |
| Yang Junseob | EDA | Reconduct EDA | - | - | - |
| Ryu Seung Gwon | EDA | Reconduct EDA | - | - | - |
|  | **Week 6** | **Week 7** | **Week 8** | **Week 9** | **Week 10** |
| Yang Junseob | - | - | - | - | Final Report |
| Ryu Seung Gwon | - | - | - | - | Final Report |

|  |  |
| --- | --- |
| **Appendix1.**  **Frequency of Community ID** | **Appendix2.**  **Total House unit by Property type of house** |
| 텍스트, 스크린샷, 직사각형, 도표이(가) 표시된 사진  자동 생성된 설명 | 텍스트, 스크린샷, 도표, 직사각형이(가) 표시된 사진  자동 생성된 설명 |
|  |  |
| **Appendix3.**  **Frequency of Property type** | **Appendix4.**  **Average log units by property type of house** |
| 텍스트, 스크린샷, 도표, 그래프이(가) 표시된 사진  자동 생성된 설명 | 텍스트, 스크린샷, 직사각형, 도표이(가) 표시된 사진  자동 생성된 설명 |

|  |  |
| --- | --- |
| **Appendix5.**  **Relation between Property type and Residential Id by years** | **Appendix6.**  **Distribution of assessor city in “geo3”** |
| **텍스트, 스크린샷, 도표, 라인이(가) 표시된 사진  자동 생성된 설명** |  |

**[Appendix7. Examine Logical Violations between Indicator and property type]**

|  |  |  |
| --- | --- | --- |
|  | **Pass** | **Fail (Logical Violation)** |
| **2017** | **211792** | **566** |
| **2018** | **212112** | **437** |
| **2019** | **213363** | **0** |
| **Total** | **637267** | **1003** |

* In 2017 and 2018, there were logical violations of indicator in 566 and 437 obs each. This is presumed to be due to a change in housing type. On the contrary, in 2019, all rows are matched, suggesting that the measurement point for geo2 is in 2019.

**[Appendix8. Create the list of queries(filters) for further analysis]**

**텍스트, 번호, 폰트, 스크린샷이(가) 표시된 사진

자동 생성된 설명**

**Reference Area:**

**Chicago city, Cook County, Illinois,**

units and stories in structure -> B25024 Units in Structure

2019: ACS 1-Year Estimates Detailed Tables

**[Appendix9. Closely examine relevant data from ACS to compare housing characteristics of 5 communities in Chicago and the reference area]**

|  |  |
| --- | --- |
|  |  |

We compared the community area data with the data for Chicago County based on the criteria of the number of house units: single house units, house units (2 to 4), and house units( five or more).

The results showed that in geo2, 34% of households had single house unit structure, 33% had two to four and 32% had five or more.

In contrast, for Chicago City, the percentages were 29%, 26%, and 44% respectively.

So, the proportion of multi-unit houses in the reference area is relatively higher.