Overview of Data Files

# Modelling Data Files

|  |  |  |
| --- | --- | --- |
|  | **File Name** | **Description** |
| Crime Cases | police centre\_dataset.xlsx | List of neighbourhood police centres and their respective locations |
| crime cases\_dataset.csv | Records of reported crime cases in respective neighbourhood police centres from 2011 to 2018 |
| Train Stations | mrt.sg.csv | Each row documents an individual MRT or LRT station. Interchanges have multiple rows, each representing the train lines that an interchange carries. Details include station name, station number, location, and the associated color of the train (eg. Yellow for Circle Line). |
| mrt\_openingDates.xlsx | Records the opening dates of each MRT station. |
| lrt\_openingDates.xslx | Records the opening dates of each LRT station. |
| train\_stations\_data.csv | Preprocessed and com dataset to include train station information from the other files above. |
| Primary Schools | general-information-of-schools.csv | Information like postal code, street addresses of schools from Primary to Tertiary level. We only extracted data of Primary schools. |
| priSch\_openingDates.csv | Opening dates and closing dates (if available) of primary schools. This dataset was filled manually as there was no available dataset online on opening and closing dates. |
| Property Price Index | property\_price\_index.csv | Records the property price index for each quarter from 2010 to 2020. |
| Property Specifications | Completion\_Date\_Manual\_Fill.csv | Includes the year of completion for properties with missing completion date. |
|  | preliminary\_dataset.csv | Each row consists of information about a resale transaction, including the project specifications and our engineered features. All features compiled into a single dataset for further preprocessing  It excludes duplicated resale transactions and those that are out of the scope of our project. |
| final\_dataset.csv | Subset of preliminary\_data.csv with a smaller number of features. |
| modelling\_dataset.csv | Consists of all the variables that will be used for model experimentation process |
|  |  |

# Interface-related Datasets

|  |  |  |
| --- | --- | --- |
|  | **File Name** | **Description** |
| Interface-related datasets | average\_cases\_by\_npc.csv | Get the average number of crime cases for each police centre.  Dataset is used to display the average yearly number of crime cases for the nearest police centre to the user’s inputted listing in Interface. |
| police\_centre\_gdf.csv | Get geo data frame for police centres.  Dataset is used to get nearest police centre to the user’s inputted listing in Interface. |
| train\_gdf.csv | Get geo data frame for train stations.  Dataset is used to get nearest stations and lines to the user’s inputted listing in Interface. |
| primary\_sch\_gdf.csv | Get geo data frame for primary schools.  Dataset is used to get nearest stations and lines to the user’s inputted listing in Interface. |