Data presented here consists of HDB flat resale transactions available for 2021 taken from Data.gov.sg and managed by the Housing and Development Board. Using information of amenities in Singapore obtained from Data.gov.sg and geospatial information acquired from OneMap.gov.sg, the direct distances and counts of nearby amenities were included. Opening dates for amenities in Singapore were cross-referenced against the resale transaction date to ensure consistency in the data.

The first HDB blocks were developed in November 1960 to address the poor housing conditions and provide safe, modern, and affordable housing on a large scale (Yueng, 1973). Initially, HDB flats were designed to be simple, standardised, and easy to build. Flats fell under one-, two- or three-room flat categories, and came in models: emergency, standard, and improved (Housing & Development Board, 1985). Since then, various other flat types and models have been introduced. For instance, the dataset contains 7 unique flat types and 15 model varieties located in 26 towns. Each town is designed to have close access to community clubs, essential education, transportation, among other amenities. Towns that were established earlier are termed by HDB as "mature" towns, while newer towns are considered to be "non-mature". Mature towns, having been around longer, are generally more developed and have better access to quality amenities. Unsurprisingly, HDB flats found in mature towns tend to be more expensive. Nevertheless, non-mature towns are increasingly becoming more developed over time with the addition of new facilities to cater to the demand of the growing population.

The hedonic pricing theory suggests that the value of a house can be viewed as a package of inherent characteristics relevant to its structure and environment. Structural characteristics refer to attributes intrinsic to the property like the size, age, and floor of the unit. In contrast, neighbourhood characteristics account for the quality and accessibility of nearby amenities. This includes proximity to both recreational and essential facilities like malls, parks, and offices. Controlling for these variables may help improve prediction accuracy.

Flat properties. In this data, we observe several structural characteristics of HDB flats such as the size, floor range, flat type, and flat model. We also observe building characteristics like the number of floors, the number of each type of flat sold, the associated property tags, and the number of each

type of rental flat. Rental flats in this case refer to HDB units set aside under the Public Rental Scheme to cater to Singapore citizen households who have no other housing options. Some HDB blocks also allocate commercial space for businesses to occupy and are demarcated by commercial property tags. Additionally, the remaining lease for each HDB flat is estimated using the date of transaction and the lease commencement.

Amenities. Neighbourhood characteristics included in this dataset fall under 5 main categories: food & recreation, transport, education, healthcare, and office facilities. We observe direct distances to the nearest hawker centres, malls, water bodies, beaches, community clubs, train stations, hospitals, schools, and offices. Where deemed relevant, variables measuring the quality and capacity of the nearest facility are included. For example, the number of cooked food and market produce stalls at the nearest hawker centre is captured by the data. Moreover, dummy variables and counts of facilities within specified proximities are added to better capture the effect of neighbourhood characteristics on resale price. In most cases, we used distances 0.5, 1, and 2km as a gauge for proximity. Schools in this dataset include primary schools, secondary schools, junior colleges, polytechnics, and autonomous universities. In this dataset, the distance to the central business district in Singapore is taken as the direct distance to Raffles station.

Existing literature has also documented the effects of superstitious beliefs on housing prices in Singapore (He et al., 2018). For example, researchers have found a preference for houses numbered “8” and a disdain for houses numbered “4” (Fortin et al. 2014; Shum et al. 2014). In Chinese culture, the number “8” is pronounced like the word “prosperity” and hence, is associated with being auspicious. In contrast, the number “4” is pronounced like the word “death” and is associated with being inauspicious. In light of this, the data contains extracts the counts of “8” and “4” numbers from the address of HDB flats (inclusive of postal code). Furthermore, distances to the nearest after-death facilities are included. After-death facilities, here, refer to cemeteries, columbaria, crematoriums, and funeral parlours.

Time controls. Other covariates that affect resale prices may also vary over time. Economic conditions, for example, can encourage the consumption of housing and spur resale prices over

time. Several studies have also noted fewer housing transactions on inauspicious months of the year (He et al., 2018). To control for potential seasonality, month dummies were created.

Geographic controls. Hedonic analysis also requires the granularity of spatial controls to be specified. In Singapore, all buildings are given a unique 6-digit postal code which comprises the sector code and the delivery point. The first two digits of a postal code represent the sector and the remaining four represents the delivery point within the sector. Moreover, the block number of each HDB block is included within the last four digits. For example, one observation located at 172 Ang Mo Kio Ave 4 has postal code 560172, where 56 represents the sector code and 172 represents the block number. This allows to vary the granularity of location controls by using 2-digit, 3-digit, or 5-digit controls, e.g., 56xxxx, 560xxx, or 5601xx. With town controls, one can control for heterogeneity across 26 locations and with 2-digit postal controls, one can control for heterogeneity at a more granular level across 51 locations.

Note: some of the variables mentioned above may have been removed from the dataset given to you due to missing observations or low relevance for the purposes of the DSE1101 project, so it is ok if you don't find some of the mentioned predictors in the data – there are plenty enough to play with!

Data description adapted from Nathanael Lam Zhao Dian's Honours Thesis (2021).