Tables for word document

Haziq Jamil

## Codebook

|  | Variable | Type | Details |
| --- | --- | --- | --- |
| 1 | id | Integer | Unique identifier for each property listing. |
| 2 | date | Date | Date when the property listing was collected. |
| 3 | quarter | Date | Quarter of the listing date in the format YYYY Qq (e.g., 2016 Q3). |
| 4 | kampong | Spatial Area | The village where the property is located. |
| 5 | mukim | Spatial Area | The sub-district administrative area where the property is located. |
| 6 | district | Spatial Area | The main district where the property is located. |
| 7 | price | Numeric | Listing price of the property in Brunei Dollars (BND). |
| 8 | type | Character | Type of property. One of "Detached", "Semi-Detached", "Terrace", "Apartment", or "Land". |
| 9 | tenure | Character | The land tenure for the property. One of "Freehold", "Leasehold", or "Strata". |
| 10 | status | Character | Current status of the listing. One of "Proposed", "Under Construction", "New", or "Resale". |
| 11 | plot\_area | Numeric | Total area of the land plot in acres. |
| 12 | floor\_area | Numeric | Built up floor area of the property in square feet. |
| 13 | storeys | Integer | Number of storeys or floors in the property. |
| 14 | beds | Integer | Number of bedrooms in the property. |
| 15 | baths | Integer | Number of bathrooms in the property. |
| 16 | agent | Character | Anonymised identifier of the real estate agent or agency handling the listing. |
| 17 | source | Character | Source of the listing. |
| 18 | method | Character | Method of data collection. |

## Summary of variables

| **Variable** | **N** | **Overall** N = 31,116*1* | **Brunei-Muara** N = 28,570*1* | **Belait** N = 1,484*1* | **Tutong** N = 767*1* | **Temburong** N = 295*1* |
| --- | --- | --- | --- | --- | --- | --- |
| Price (BND 1,000) | 31,116 |  |  |  |  |  |
| Mean (SD) |  | 340 (381) | 340 (393) | 372 (208) | 259 (87) | 421 (324) |
| Min - Max |  | 70 - 13,800 | 70 - 13,800 | 98 - 2,800 | 116 - 680 | 118 - 1,800 |
| Median (Q1, Q3) |  | 288 (230, 380) | 285 (230, 380) | 320 (268, 400) | 245 (198, 310) | 390 (250, 430) |
| Property type | 27,231 |  |  |  |  |  |
| Detached |  | 17,416 (64%) | 16,307 (65%) | 520 (42%) | 509 (74%) | 80 (56%) |
| Semi-Detached |  | 3,823 (14%) | 3,591 (14%) | 97 (7.8%) | 128 (19%) | 7 (4.9%) |
| Terrace |  | 4,449 (16%) | 4,134 (16%) | 213 (17%) | 48 (7.0%) | 54 (38%) |
| Apartment |  | 1,527 (5.6%) | 1,106 (4.4%) | 414 (33%) | 4 (0.6%) | 3 (2.1%) |
| Land |  | 16 (<0.1%) | 11 (<0.1%) | 4 (0.3%) | 1 (0.1%) | 0 (0%) |
| Land tenure | 12,877 |  |  |  |  |  |
| Freehold |  | 9,296 (72%) | 8,405 (76%) | 368 (33%) | 381 (80%) | 142 (97%) |
| Leasehold |  | 2,783 (22%) | 2,221 (20%) | 467 (41%) | 91 (19%) | 4 (2.7%) |
| Strata |  | 798 (6.2%) | 504 (4.5%) | 291 (26%) | 3 (0.6%) | 0 (0%) |
| Development status | 22,481 |  |  |  |  |  |
| Proposed |  | 4,004 (18%) | 3,660 (18%) | 103 (8.8%) | 197 (33%) | 44 (31%) |
| Under Construction |  | 9,420 (42%) | 8,600 (42%) | 535 (46%) | 244 (41%) | 41 (29%) |
| New |  | 7,724 (34%) | 7,122 (35%) | 413 (35%) | 132 (22%) | 57 (40%) |
| Resale |  | 1,333 (5.9%) | 1,186 (5.8%) | 120 (10%) | 26 (4.3%) | 1 (0.7%) |
| Plot area (acres) | 23,368 |  |  |  |  |  |
| Mean (SD) |  | 0.16 (0.12) | 0.15 (0.11) | 0.19 (0.15) | 0.18 (0.17) | 0.23 (0.21) |
| Min - Max |  | 0.01 - 2.00 | 0.01 - 1.69 | 0.01 - 1.01 | 0.04 - 2.00 | 0.05 - 0.96 |
| Median (Q1, Q3) |  | 0.13 (0.08, 0.19) | 0.13 (0.08, 0.19) | 0.13 (0.06, 0.27) | 0.14 (0.10, 0.21) | 0.16 (0.13, 0.26) |
| Floor area (sq. ft.) | 16,665 |  |  |  |  |  |
| Mean (SD) |  | 2,602 (1,047) | 2,629 (1,062) | 2,423 (913) | 2,133 (651) | 2,786 (751) |
| Min - Max |  | 500 - 14,411 | 500 - 14,411 | 600 - 7,500 | 1,093 - 7,000 | 950 - 3,700 |
| Median (Q1, Q3) |  | 2,427 (2,000, 3,000) | 2,465 (2,000, 3,000) | 2,218 (1,800, 2,800) | 2,013 (1,826, 2,450) | 3,016 (2,790, 3,229) |
| Number of storeys | 13,644 |  |  |  |  |  |
| 1 |  | 1,700 (12%) | 1,462 (12%) | 160 (35%) | 71 (17%) | 7 (4.2%) |
| 2 |  | 11,266 (83%) | 10,493 (83%) | 280 (61%) | 348 (83%) | 145 (87%) |
| 3+ |  | 678 (5.0%) | 642 (5.1%) | 19 (4.1%) | 2 (0.5%) | 15 (9.0%) |
| Number of bedrooms | 26,631 |  |  |  |  |  |
| Mean (SD) |  | 4.2 (0.9) | 4.2 (0.9) | 4.0 (1.1) | 3.9 (0.7) | 4.7 (1.0) |
| Min - Max |  | 0.0 - 12.0 | 0.0 - 12.0 | 1.0 - 10.0 | 2.0 - 7.0 | 2.0 - 7.0 |
| Median (Q1, Q3) |  | 4.0 (4.0, 5.0) | 4.0 (4.0, 5.0) | 4.0 (3.0, 4.0) | 4.0 (3.0, 4.0) | 5.0 (4.0, 5.0) |
| Number of bathrooms | 19,694 |  |  |  |  |  |
| Mean (SD) |  | 3.7 (1.2) | 3.7 (1.2) | 3.3 (1.1) | 3.3 (1.0) | 3.2 (1.5) |
| Min - Max |  | 1.0 - 11.0 | 1.0 - 11.0 | 1.0 - 8.0 | 1.0 - 7.0 | 1.0 - 5.0 |
| Median (Q1, Q3) |  | 3.0 (3.0, 4.0) | 3.0 (3.0, 4.0) | 3.0 (3.0, 4.0) | 3.0 (2.0, 4.0) | 2.0 (2.0, 5.0) |
| *1*n (%) | | | | | | |

## Data availability by year

|  |  |  | Missing data severity | | Data source | | |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Count | Spatial coverage (mukim)*1* | Property Type*2* | Property Characteristics*3* | National Archive | Online Archive | Web Scraping | LLM post-processing |
| 1993 | 400 | 33.3% | 0.0% | 19.0% | ✔ |  |  |  |
| 1994 | 653 | 51.9% | 65.8% | 27.9% | ✔ |  |  |  |
| 1995 | 668 | 70.4% | 66.8% | 21.3% | ✔ |  |  |  |
| 1996 | 561 | 51.9% | 69.7% | 12.1% | ✔ |  |  |  |
| 1997 | 385 | 51.9% | 38.4% | 26.8% | ✔ |  |  |  |
| 1998 | 345 | 48.1% | 36.8% | 28.7% | ✔ |  |  |  |
| 1999 | 317 | 51.9% | 31.9% | 26.2% | ✔ |  |  |  |
| 2000 | 378 | 63.0% | 0.8% | 4.2% | ✔ |  |  |  |
| 2001 | 342 | 63.0% | 0.3% | 2.3% | ✔ |  |  |  |
| 2002 | 437 | 63.0% | 0.0% | 20.4% | ✔ |  |  |  |
| 2003 | 449 | 66.7% | 0.0% | 13.4% | ✔ |  |  |  |
| 2004 | 440 | 63.0% | 0.0% | 19.1% | ✔ |  |  |  |
| 2005 | 493 | 66.7% | 0.0% | 13.2% | ✔ |  |  |  |
| 2006 | 653 | 59.3% | 0.2% | 11.3% | ✔ |  |  |  |
| 2007 | 638 | 55.6% | 0.0% | 12.9% | ✔ |  |  |  |
| 2008 | 687 | 59.3% | 0.3% | 5.8% | ✔ |  |  |  |
| 2009 | 531 | 51.9% | 0.2% | 4.0% | ✔ |  |  |  |
| 2010 | 571 | 55.6% | 0.0% | 2.1% | ✔ |  |  |  |
| 2011 | 594 | 55.6% | 0.2% | 10.1% | ✔ |  |  |  |
| 2012 | 934 | 63.0% | 8.0% | 4.3% | ✔ |  | ✔ |  |
| 2013 | 882 | 59.3% | 2.9% | 26.4% | ✔ | ✔ | ✔ |  |
| 2014 | 709 | 66.7% | 10.6% | 8.0% |  | ✔ | ✔ |  |
| 2015 | 868 | 66.7% | 12.6% | 6.3% |  | ✔ | ✔ |  |
| 2016 | 1461 | 70.4% | 13.6% | 4.6% |  | ✔ | ✔ |  |
| 2017 | 1638 | 70.4% | 14.5% | 4.5% |  | ✔ | ✔ |  |
| 2018 | 2646 | 66.7% | 17.6% | 0.0% |  |  | ✔ |  |
| 2019 | 3586 | 63.0% | 15.9% | 0.0% |  |  | ✔ |  |
| 2020 | 1363 | 66.7% | 10.4% | 0.0% |  |  | ✔ | ✔ |
| 2021 | 1115 | 77.8% | 2.0% | 0.0% |  |  | ✔ | ✔ |
| 2022 | 1235 | 77.8% | 3.6% | 0.0% |  |  | ✔ | ✔ |
| 2023 | 1593 | 77.8% | 2.8% | 0.0% |  |  | ✔ | ✔ |
| 2024 | 2972 | 77.8% | 4.6% | 0.1% |  |  | ✔ | ✔ |
| 2025 | 572 | 59.3% | 14.7% | 0.0% |  |  | ✔ | ✔ |
| mean | 943 | 62.0% | 13.5% | 10.1% | — | — | — | — |
| *1*Of Brunei’s 39 mukims, only 27 are considered transactable—excluding water villages and remote, non-developable areas. | | | | | | | | |
| *2*Unknown property type. | | | | | | | | |
| *3*Missing all of plot area, floor area, beds, and baths variables. | | | | | | | | |

## LLM Test

|  | Time (seconds) | | | |
| --- | --- | --- | --- | --- |
| Model | Minimum | Mean | Median | Maximum |
| llama3.2 | 1.08 | 6.40 | 9.60 | 10.16 |
| mistral | 2.66 | 10.44 | 2.91 | 22.21 |
| phi4 | 3.72 | 14.11 | 4.27 | 38.01 |
| deepseek-r1:8b | 36.83 | 65.97 | 54.61 | 112.78 |
| deepseek-r1:14b | 40.89 | 117.67 | 81.71 | 151.20 |
| gpt-4o | 1.28 | 2.46 | 1.73 | 5.38 |
| o1-mini | 5.97 | 7.95 | 7.51 | 9.94 |