# Insurance Premium Web Scraping Readme

### Input Data

### Test\_auto\_data1 Data Dictionary (Explaining the formats that variables must be in)

#### Clean Data, Data Dictionary

**Note: Should NOT edit the clean data table. All new data/ changes to the data should be done on the Raw data sheet (will flow through to the Clean data sheet)**

|  |  |
| --- | --- |
| **Variable** | **Required Data and Format** |
| *Sample Number* | A numeric KEY (must be unique for each entry) |
| *Unit\_number* | The unit number of the address that the car is usually kept at overnight. Can be a letter, a number or a blank |
| *Street\_number* | The street number of the house. CAN have either A or B with it |
| *Street\_name* | The street name. **MUST BE CAPITALISED** |
| *Street\_type* | The street type (E.g. STREET, ROAD, DRIVE, …). **MUST BE CAPITALISED** |
| *Suburb* | The suburb that the house is at. **MUST BE CAPITALISED** |
| *Postcode* | The postcode of the house. Is a number (the code handles cases with 0 at the front). **MUST BE CAPITALISED** |
| *City* | The city the house is in. **MUST BE CAPITALISED** |
| *BusinessUser* |  |
| *Occupancy* | The type of occupancy the house has. Must be EXACTLY one of the following   1. Owner occupied 2. Holiday home 3. Rented 4. Let to family/employees 5. Unoccupied |
| *Boarders* | A Variable that states whether there are boarders within the house (the owner may or may not also be living here). Must be EXACTLY one of the following   1. Yes 2. No |
| *Tenants* | A Variable that states whether there are tenants within the house (the owner may or may not be also be living here). Must be EXACTLY one of the following   1. Yes 2. No |
| *Age* |  |
| *DOB* | The persons Date of Birth. (Is a FORMULA, calculated based on the date today, and the persons age) |
| *Gender* | One of two options   * MALE * FEMALE |
| CarInsurance |  |
| ContentsInsurance |  |
| FarmInsurance |  |
| BoatInsurance |  |
| HomeCombInsurance |  |
| BuildingType |  |
| ConstructionType |  |
| RoofType |  |
| HouseStandard |  |
| LandShape |  |
| YearBuilt |  |
| PurchaseYear |  |
| NumberOfStories  DwellingFloorArea |  |
| HasGarage |  |
| GarageFloorArea |  |
| SumInsured |  |
| GlassExcess |  |
| PreviousClaim |  |
| Date\_of\_incident |  |
| Type\_incident |  |
| HaveAAPolicies |  |
| CarSpaces |  |
| Excess |  |
| Balcony |  |
| SwimmingPool |  |
| SportCourt |  |
| GardenShed |  |
| Shed |  |
| WaterTanks |  |
| TownWaterSupply |  |
| HaveTowerPolicies |  |
| *AAMember* | States whether this person is an AA member. Must be EXACTLY one of the following   1. Yes 2. No |
| *CurrentInsurer* | States who this person’s current insurer is. Currently MUST BE *“None”*, as the code is not fit to handle anything else |