# 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** | **Websites Used By** |
| *Sample Number* | A numeric KEY (must be unique for each entry) | ALL |
| *PolicyStartDate* | Today’s date (=Today() ) | ALL |
| *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 | ALL |
| *Street\_number* | The street number of the house. CAN have either A or B with it | ALL |
| *Street\_name* | The street name. **MUST BE CAPITALISED** | ALL |
| *Street\_type* | The street type (E.g. STREET, ROAD, DRIVE, …). **MUST BE CAPITALISED** | ALL |
| *Suburb* | The suburb that the house is at. **MUST BE CAPITALISED** | ALL |
| *Postcode* | The postcode of the house. Is a number (the code handles cases with 0 at the front). **MUST BE CAPITALISED** | ALL |
| *City* | The city the house is in. **MUST BE CAPITALISED** | ALL |
| *BusinessUser* | Whether the house is used for business Purposes. Is one of the following:   1. Yes-Home office 2. Yes-Hobby farm 3. Yes-Bed and breakfast 4. Yes-Other 5. Yes->50% of house used for business 6. No  * Note: >50% of house used for business, means more than half the house is used for the business. For ALL OTHER Yes options, we assume that less than 50% of the house is used for the business. | ALL |
| *BuisnessStaff* | A number stating how many staff the business at the house employs. Is N/A if there is no business done at the house, else is the number of staff under employment.  If more than 3 employees, then the house in not insurable under home insurance (needs to be under business insurance) | AMI |
| *BusinessStockValue* | Does the business that is run from this house have more or less than $10,000 of stock to sell. Is either:   1. Less than 10k 2. Greater than 10k   If Greater than 10k in stock to sell, then the house in not insurable under home insurance (needs to be under business insurance) | AMI |
| *BuinessHazardousMaterials* | Do you use, store or stock flammable or hazardous materials? Is either:   1. Yes 2. No   If there are hazardous materials stored or used on the property, then the house in not insurable under home insurance (needs to be under business insurance) | AMI |
| *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 5. Let to employees 6. Unoccupied | ALL |
| *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 | AMI, Tower |
| *Tenants* | A Variable that states whether there are tenants within the house (the owner may or may not be living here). Must be EXACTLY one of the following   1. Yes 2. No | AMI, Tower |
| *NumTenanciesOnProperty* | A variable which states how many tenancies are present on the property. Must be a whole number | AMI |
| *ShortTermTenancy* | A variable that states whether there is a short-term (less than 30 days) tenancy at the property. Is only important if the property has tenants. If tenants = 0, then this variable must be “Not Applicable”. Must be one of the following   1. Not Applicable 2. Yes 3. No | AA, AMI |
| *OwnerStayInHouseWithinNextYear* | A variable that states whether the owner is going to stay at the house within the next year. This is valid for Holiday homes, thus any house that does not have Occupancy = “Holiday home” is not applicable. Must be one of the following   1. Not Applicable 2. Yes 3. No | Tower |
| *Age* | The persons Age. Is a fixed numeric variable, used to calculate their date of birth | None (but is needed to calculate DOB) |
| *DOB* | The persons Date of Birth. (Is a FORMULA, calculated based on the date today, and the persons age) | ALL |
| *Gender* | One of two options   * MALE * FEMALE | ALL |
| *CarInsurance* | Whether or not the person already has car insurance with the provider  - Note: (The 4 kinds of insurance variables are needed to be able to input whether the person already has any insurance with the given company) | AA |
| *ContentsInsurance* | Whether or not the person already has contents insurance with the provider | AA |
| *FarmInsurance* | Whether or not the person already has farm insurance with the provider | AA |
| *BoatInsurance* | Whether or not the person already has boat insurance with the provider | AA |
| *BuildingType* | A variable to say what type of building this property is. Can only currently be one of the following options:   1. Freestanding house 2. Apartment 3. Multi unit-num (where ‘num’ is the number of units in the Multi unit) 4. Semi detached / townhouse 5. Retirement unit | AA, AMI |
| *ConstructionType* | A variable which says the main material the house is made of. The options are listed in the Data Dictionary page | ALL |
| *RoofType* | A variable which says the main material the roof is made of. Many of the potential materials are not given as options to input on the websites. | ALL |
| *HazardRisk* | Has the house had previous damage, or had the council identify it as being at risk from any natural hazard. Either   1. Yes 2. No | AMI, Tower |
| *HouseStandard* | A variable stating to what standard the house was designed and built? One of the following:   1. Ordinary 2. Quality 3. Prestige | AMI, Tower |
| *LandShape* | A variable that states if the house is built on flat or sloping land. One of the following:   1. FlatAndGentle 2. Moderate 3. Severe | AMI, Tower |
| *YearBuilt* | A variable stating the year the house was built.  **NOTE:** If the year is earlier than 1935, there are a bunch of additional questions that need to be answered. I have not set up the infrastructure for this yet | ALL |
| *PurchaseYear* | A variable stating the year the person purchased the house | AMI |
| *NumberOfStories* | A number stating how many stories the house has | ALL |
| *DwellingFloorArea* | A number stating the floor area of the house (not the property) | ALL |
| *NumSelfContainedDwellings* | The number of self-contained dwellings on the property. (Self-contained means they have their own living, sleeping, bathing and cooking facilities). | ALL |
| *HasGarage* | A variable stating whether the property has a garage | AMI |
| *GarageFloorArea* | A number stating the floor area of the garage | AMI, Tower |
| *CoveredExternalCarSpaces* | The number of external covered car spaces. (Include all external covered car spaces that you have such as garages or carports. Exclude any internal garages you have). | AA |
| *SumInsured* | A number stating the likely cost to rebuild your house and domestic outbuildings at today's prices | ALL |
| *GlassExcess* | Whether to pay no excess for glass breakages. Either:   1. Yes 2. No   Note: Currently only allows ‘No’ | AA, AMI |
| *Date\_of\_incident* | A date variable in the format of dd/mm/yyyy. If an incident has occurred, is the date of the incident, if NO incident has occurred, it is 1/01/1900 | AA, Tower |
| *Type\_incident* | The type of incident that occurred. Is one of the following:   1. Accidental Glass Breakage 2. Accidental Loss/Damage at the home 3. Accidental Loss/Damage away from home 4. Burglary (with break in) 5. Damage by an animal 6. Escaped water or other liquid 7. Explosion 8. Fire 9. Flood at current address 10. Flood at previous address 11. Impact (e.g. falling tree, space debris or vehicle) 12. Legal Liability 13. Loss of Rent 14. Malicious damage/Vandalism 15. Motor burnout, fusion or food spoilage 16. Natural Event 17. Other 18. Storm, cyclone or rainwater runoff 19. Theft (without break in) 20. No Incident | AA, Tower |
| *Excess* | A number stating the amount of excess you must pay. Is currently always 500. | ALL |
| *Balcony* | Whether the property contains a Balcony/Deck/Verandah/Pergola. Is either:   1. Yes 2. No | AA |
| *SwimmingPool* | Whether the property contains a fixed swimming pool. Is either:   1. Yes 2. No | AA |
| *SportCourt* | Whether the property contains a sports court. Is either:   1. Yes 2. No | AA |
| *AdditionalFeaturesAboveLimit* | The policy covers retaining walls up to $20,000 and recreational features (e.g. fixed spa pool, in ground swimming pool or tennis court) up to $40,000.  This policy currently does not cover any:   * Private utility plant and associated equipment, other than solar power or heating (e.g. Wind or water mills or diesel generators) * Cable car and its associated equipment * Bridge or culvert, permanent ford or dam * Wharf, pier, landing or jetty   Do you need cover for any items worth more than these limits or currently not covered? Either:   1. Yes 2. No | AMI |
| *GardenShed* | Whether the property contains a garden shed (small). Is either:   1. Yes 2. No | AA |
| *Shed* | Whether the property contains a shed (large). Is either:   1. Yes 2. No | AA |
| *WaterTanks* | Whether the property contains any water tanks. Is either:   1. Yes 2. No | AA |
| *AAMember* | States whether this person is an AA member. Must be EXACTLY one of the following   1. Yes 2. No | AA |
| *CurrentInsurer* | States who this person’s current insurer is. Currently MUST BE *“None”*, as the code is not fit to handle anything else | AA |
| *AMICurrentRenovations* | A variable which states what type of current renovations the house is undergoing. Must be one of the following:   1. No (for the case where no current renovations are ongoing) 2. Non-structural>75k 3. Non-structural<75k 4. Structural | AMI |
| *LostRentCover* | State whether you wish to add cover for lost rent if your rental house is so damaged that it is unfit to live in. Must be one of the following:   1. Yes 2. No | AMI |
| *UnexpectedVacancyCover* | State whether you wish to add cover for the actual net rent you lose if your tenant leaves without giving the required notice. Must be one of the following:   1. Yes 2. No | AMI |
| *TheftOrDamageByTenantsCover* | State whether you wish to add cover for theft or deliberate damage caused by your tenant. Must be one of the following:   1. Yes 2. No | AMI |
| *ChattelsCover* | State whether you wish to add cover for the chattels, furniture and appliances in your rental property. Must be one of the following:   1. Yes 2. No | AMI |
| *BodyCorporate* | States whether the house is a part of a body corporate: Either:   1. Yes 2. No | AA |
| *HouseSecurity* | A variable to state the level of security protection that house has. Is one of the following:   1. No Security 2. Burglar alarm, not monitored by a security provider 3. Burglar alarm, monitored by a security provider | AA |
| *HouseWellMaintained* | This means that your house, property or unit and contents do not have any faults or defects that might cause loss or damage to your house/property and contents, or loss or damage to property of others or injury to people. Is either   1. Yes 2. No | AA, Tower |
| *HeritageProperty* | Have Heritage New Zealand or your local council placed any restrictions or preservation orders on this house? Is either:   1. Yes-Heritage 2. Yes-Local Council 3. No | AA |
| *Insurance Refused In Last 7 Years* | In the last seven years have you or anyone covered by this policy had any insurance refused, cancelled, special terms imposed, renewal not offered or a claim declined? Is one of the following:   1. Yes-Fraud 2. Yes-Non-disclosure 3. Yes-Misrepresentation 4. Yes-Breach of policy conditions 5. Yes-Other 6. No | AA, Tower |
| *Crime in Last 7 Years* | In the last seven years have you or anyone covered by this policy had any criminal convictions? Is one of the following:   1. Yes-Fraud 2. Yes-Arson 3. Yes-Burglary 4. Yes-Theft 5. Yes-Drugs 6. Yes-Other 7. No | AA, Tower |
| *PolicyHeldByCompany/Trust* | State whether the policy is going to be held by a company or trust. Must be:   1. No   (if Yes then you need to provide a whole bunch of information about the company/trust) | AA, Tower |
| *Mortgage* | State whether there is a mortgage on the house, and if there is, who is it with. Is either:   1. Yes- 2. No   Following the dash must be one of the following institutions that may have lent the money for the mortgage the brackets just says what the acronym means:   1. ABL (ASB Bank Limited) 2. ABNZ (ANZ Bank New Zealand Limited) 3. BNZ (Bank of New Zealand) 4. NZCU (NZCU) 5. GEMO (GE Money) 6. HNZL (The Hongkong and Shanghai Banking Corporation Limited) 7. KWBL (Kiwibank Limited) 8. SCBS (Southern Cross Building Society) 9. PSIS (The Co-Operative Bank Limited) 10. TSBL (TSB Bank Limited) 11. AAF (AA Finance) 12. AAFL (AA Finance Limited (AA Money)) 13. MHTC (Mortgage Holding Trust Company Limited) 14. NZLL (New Zealand Home Lending Limited) 15. NZHL (NZ Home Loans) 16. OTHR (Other (Not Listed)) 17. SBS (Southland Building Society (SBS Bank)) 18. SOVE (Sovereign) 19. WNZL (Westpac New Zealand Limited)  * NOTE: Code does not currently allow for 2nd mortgage | AA |
| *MortgageeSale* | Did the current owner purchase the house in a mortgagee sale in the current year? If they did then need to call in to AMI to get insurance (so no online quote).  As we only care about mortgagee sale houses that were purchased in the current year, we put ‘Not Relevant’ for any other purchase year.  Options are the following:   1. Yes 2. No 3. Not Relevant | AMI |
| *Re-roofedRe-linedAndRe-wired* | Has the house ever been re-roofed, re-lined and re-wired? Older houses were often built with materials or methods that don't meet current safety standards, and can increase the risk of loss or damage. We can only offer cover if the original roofing, lining, and wiring have been updated. Answer is either:   1. Yes 2. No | Tower |

### Error Code system

***No Error***: If the code ran on that website for the given example and scraped out the premium

***Website Does Not Quote For This House/ Person***: Occurs when the website in question does not automatically quote for the given car. To get quotes for these cars the websites will prompt the user to call into their help centre to get a quote. Sometimes it takes issue with the person who is insuring the car, for example on some cars tower refuses to insure under 25 drivers, or if a person has had too many incidents recently. Ultimately, this error code occurring mean that it is not possible for the web scraper to get a premium from that website

***Unable to find car variant***: Occurs when there are no options for car variant that align with the data in the spreadsheet

***Invalid Input Data Error:*** An error code to notify that the data that was input was invalid

***Unknown Error***: A general message for any other error encountered while scraping. Often these errors are caused by slowness on the browser preventing the scraping from executing correctly. The code attempts to remedy this by going back over all the unknown error once, at the end of the code, to attempt to ‘fix’ the situation scraping error. However, some of these errors are unavoidable, however.

**Code:**

Data: {… , … }

Wait 1 or 3 or

Wait10.until(EC.element\_to\_be\_clickable((By.ID, 'xxxxxxx'))).click() – right click to inspect

driver.find\_elements(By.XPATH, "xxxxxx") ---- right click to copy Xpath

driver.find\_element(By.XPATH, "xxxxxx") ---- right click to copy Xpath(one element)

*note: AI tend to use “driver.find\_elelemetn\_by\_path” but it is same to “driver.find\_element(by.Xpath..)*

