**Issues with data**

* CustomerDemographics
  + DOB – one data value is not a datetime data type (Jephthah Bachmann)
    - This value is also not feasible as person is said to be born in 1843
  + Quite a few values have the last\_name left blank
  + Gender
    - Not named consistently
      * F, Femal and Female all for one gender
      * M, Male for the other gender
      * U – for unspecified (potentially those who are non-binary)
  + Quite a number of people without a job title, yet have indicated a job\_industry.
  + Job\_industry entered as n/a even though they have a job title
    - Shouldn’t all jobs belong to a particular industry?
  + People with tenure left blanks also have default blank as well as DOB. Unsure as to why.
* Transactions
  + Should transaction\_id be in line with transaction\_date
    - Where the first transaction of the year has id 1, instead?
  + Product Line
    - Many transactions have blanks in that area
    - It is the same for brand, product\_class, and product\_size
    - All from same transactions
  + Online\_order
    - Quite a number of cells are blank, where it should indicate whether the transaction was an online\_order or not, it is only YES or NO.
  + Product\_first\_sold\_date
    - Blank cells – standard\_cost, brand, product\_line, product\_class and product\_size are all also blank for these transactions
    - All these transactions all have product\_id 0 (correlation),
      * Is there a reason for it?
* NewCustomerList
  + 29 people who do not have a surname (last\_name) entered into the dataset
  + DOB – not the correct data type (should be a datetime datatype)
    - Additionally, 17 people without date of birth indicated.
  + Quite a number of people without a job title, yet have indicated a job\_industry.
  + Job\_industry entered as n/a even though they have a job title
    - Shouldn’t all jobs belong to a particular industry?
  + Past 3 years bike purchases
    - Deemed as a text type, not as number
  + Postcode
    - Inconsistent data types (number and text types)
  + Property\_value
    - Inconsistent data types (numbers (with decimals), and text types)
  + Value
    - Potentially many need to clean up data so that there aren’t too many decimal places for particular values.
* CustomerAddress
  + Customer\_id
    - Some numbers have been skipped (e.g. 3, 10, 22, 23). Why is this the case?
  + Address
    - People have street number 0
      * This is not possible for street addresses, street addresses start from 1.
  + State
    - Some addresses have NSW as the abbreviation, some have ti written in full (New South Wales). This is the same for addresses from Victoria (VIC and Victoria) for this column.