# Data: Transactions

* **online\_order**
  + Contains 360 NULL values
* **Brand**
  + Contains 197 NULL values
* **product\_line**
  + Contains 197 NULL values
* **product\_class**
  + Contains 197 NULL values
* **product\_size**
  + Contains 197 NULL values
* **Standard\_cost**
  + Contains 197 NULL values
* **Product\_first\_sold\_date**
  + Data does not make any sense
  + Tried converting data into date time type but it did not work

**NOTE: brand, product\_line, product\_class, product\_size, standard\_cost, product\_first\_sold\_daya have same 197 missing values which indicates that dataset is missing data of one brand.**

**Solutions:**

* Delete all rows with NULL values
* Replace NULL value with assumed value
* Replace NULL values with temporary values
* Delete Product\_first\_sold\_date column

# Data: NewCustomerList

* **last\_name**
  + Contains 29 NULL values
* **DOB**
  + Contains 17 Null values
* **Job\_title**
  + Contains 106 Null values
* **Job\_industry\_category**
  + Contains 165 NULL values

**NOTE: No correlation between job\_title and job\_industry\_category in some cases Ex – Nurse -> Property, librarian -> Entertainment**

* **Wealth\_segment**
  + Affluent customer and High net worth seem similar
* **Deceased\_indicator**
  + All values same (N)
  + Data not up to date

**Note: Following cols seem irrelevant for a company who is in a bike business**

1. **tenure**
2. **property\_valuation**
3. **Rank**

* **Value**
  + No context

**Solutions:**

* Delete rows with NULL values in last\_name and DOB columns
* Replace missing job\_title with similar job\_title to its corresponding job\_industry\_category
* Replace missing job\_industry\_category with similar job\_industry\_category to its corresponding job\_title
* Replace all Affluent customer in Welth\_segment column with High net worth or vice versa
* Delete or update Deceased\_indicator column
* Delete Value column

# Data: CustomerDemographic

* **Last\_name**
  + Contains 125 NULL values
* **Gender**
  + Redundant category e.g Male and M, Female and F
* **DOB**
  + Year 1843 seems odd and only one instance of that year available then 1931 so 1843 looks like a mistake
* **Job\_title**
  + Contains 506 missing data
* **job\_industry\_category**
  + Contains 656 NULL values
* **Wealth\_segment**
  + Affluent customer and High net worth seem similar
* **tenure**
  + Contains 87 NULL values
* **default**
  + Unclear data purpose
  + Name unclear
  + Looks like font type mismatch or garbage value

**Solutions:**

* Delete all rows with NULL values
* Replace Male with M and Female with F in Gender column
* Delete 1843-year instance or replace it with 1943
* Replace missing job\_title with similar job\_title to its corresponding job\_industry\_category
* Replace missing job\_industry\_category with similar job\_industry\_category to its corresponding job\_title
* Replace all Affluent customer in Welth\_segment column with High net worth or vice versa
* Delete default column

# Data: CustomerAddress

* **state**
  + Some state have sort form and some have not
  + For Victoria state 2 notations are used VIC and Victoria
* **property\_valuation**
  + Measure not clear

**Solutions:**

* Replace all instances of Victoria with Vic