

Dear Sir/Madam,

**DATA QUALITY ISSUES IDENTIFIED FROM THE SKYROCKET**  
**Central Pty LTD DATA SET**

The following where the data quality issue that were identified in the data set and should be carefully looked into in order not to influence any data analysis negatively and to avoid data integrity issues.

1. For proper representation of the data type the following fields contained in the dataset should be represented as *string* data type instead of *numerical* data type; 'transaction\_id', 'product\_id' and 'customer\_id'.
2. In the field 'product\_first\_sold\_date' in the '**Transactions**' worksheet has a *numerical* data type instead of it been a *date* data type.
3. **Transactions** worksheet consist of 197 missing values which cut across the following fields: 'brand', 'product\_line', 'product class', 'product\_size', 'standard\_cost' and 'product\_first\_sold' and all these missing values has 'product\_id' of 0, also the field 'online\_order' has 360 missing values
4. The gender field in the **CustomerDemographic** worksheet does not contain a consistence value at such it should be readjusted to female and male
5. The state field in the **CustomerAddress** worksheet does not contain a consistence value
6. The DOB field in the **CustomerDemographic** worksheet has a wrongly inputted data 1843-12-21