Hello spyrocket,

Thank you for putting your trust in KPMG AU and supplying us with the datasets, we promise to leave you with a great experience working with us.

I was performing a data check on the dataset and I did notice some inconsistency and faults in the dataset. I have outlined the various issues with the dataset below and I made sure to include the approach we would be taking to clean the data and tackle those inconsistencies.

The data quality challenges faced include the following:

- **Problems of accuracy(eg: job title, job industry and property value)**: the dataset set has some inaccurate values, for example, the inconsistent representation of values, wrong values and also misspelled values.
- Problems of completeness: the dataset has a lot of blank cells(values), some are represented with n/a while the rest are left just as blanks
- **Problems of consistency(eg: gender ,state)**: the dataset has some columns having contradicting values(represented in different ways), some values are represented wrongly and cannot be assumed to belong to any category
- **Problems of currency(eg: DOB)** : the dataset has some missing dates, incorrect date formats, dates missing the days and month and there is not an indication of the latest or earliest dates.
- **Problems of relevance**: the absence of the meta data posses a threat to our analysis as we do not have comprehensive information of the data and what all columns represent.
- **Problems of validity(eg:first sold date)**: there are some values that are not supposed to be in the data set and conflict with the allowed values

We have decided to take the following steps toward data cleaning and validation:

- Request a recheck from the data collection team at spyrocket
- Change the categorical values into one acceptable format
- Replace missing continuous values with the mean value
- Remove duplicate values and fix in the mode date for missing date values
- Replace blanks with a null indicating value for the name columns
- Decide on keys for all tables
- And use the customer value as a criteria to pick the first 1000 customers to focus on

I hope to hear from you soon enough to get a response on the method and approach which we have chosen to follow at KPMG.