**Missing Values:**

1. Client Category
   1. For all missing values in client category the location was L10
   2. L10 had max count of VP as ‘Long Bergstorm’
   3. Long Bergstorm had max categories of ‘International Bank’ [110] with a highest difference of 15 records
2. Deal Cost
   1. Since the data distribution is skewed, mean is not preferable, median mode yielded same result. Hence median was chosen.

**Encoding Technique:**

Count/Frequency Encoding –

Since there are many categorical features, choosing one hot encoding would have resulted in large dimensional values. In client category itself there were 41 unique values.