Analytics Foundations: Lab 1

1. Use the *NormTemp* dataset to answer the following:
   a. Determine the following statistics for the variable **temperature:**

Minimum	96.3
Maximum	100.8
Mean	98.25
Standard Deviation	0.733

   b. Does **temperature** appear to be normally distributed?

YES, THE QQ PLOT LOOKS NORMAL

   c. Create box plots for **temperature**. Display a reference line at 98.6. Does the median body temperature seem to be 98.6 degrees?

NO. THE REFERENCE LINE LOOKS CLOSER TO THE 75 PERCENTILE THAN THE MEDIAN.

2. Using the *Ameshousing* dataset from our in-class examples, run some distributional analysis on **SalePrice**, **Log_Price**, and **Gr_Liv_Area**.
   a. Create histograms of these three variables.
      - Overlay with a kernel density estimator of the variables.

   b. Create a QQ Plot for both SalePrice and Log_Price. Based on these exploratory procedures, which version of the price information (SalePrice or log_price) would you say is closer to being normally distributed?

Log(SALE_PRICE) LOOKS CLOSER TO A NORMAL DISTRIBUTION THAN SALE_PRICE

3. Using the *Ameshousing* dataset from our in-class examples, determine the following:

   a. What type of variables are each of these columns (Nominal, Ordinal, or Continuous)? Keep in mind that the way they are represented in the SAS dataset may not be appropriate, so you should make this determination using your *own judgement based on the data you are looking at*.

   - **Overall_Qual** **ORDINAL**
   - **Lot_Shape** **ORDINAL**
   - **Heating_QC** **ORDINAL**
   - **Lot_Area** **CONTINUOUS/QUANTITATIVE**