BA 464

Business Analytics and Big Data
By Mehmet Gençer

Week 3B: Exploratory Analytics

Data Exploration

- Allows the business analytics professional to understand the nature of variables in the data set and their interactions, possibly leading to clues on causality.

- Good wuality exploratory analytics facilitate business discussion, hence a valuable output in itself.

Single variable: summary

- Mean, median, mode, quartiles

- Extreme values: If a data distribution is approximately normal then about 95 percent are within two standard deviations ($\mu \pm 2\sigma$). Most modeling methods are adversely effected, ie biased, by extreme values.

- Data quality: null and/or extreme values

- Shapiro test for normality:
  p-value should be more than 0.95 to be confident enough that data comes from a normal distribution.

Single variable: visualization

- Histogram
- Stem-plot
- Box-plot

Two variable

- Correlation
- Plot
- Two variable box-plot