Department of Computer Engineering
Faculty of Engineering

Kasetsart University

Data Exploration & Visualization using Excel & Python (import Pandas, Math, …)

1. **Shipments of Household Appliances: Line Graphs.**

The file ApplianceShipments.xls contains the series of quarterly shipments (in million $) of U.S. household appliances between 1985 and 1989 (data courtesy of Ken Black).

- **a.** Create a well-formatted time plot (line graph) of this data.
- **b.** Does there appear to be a quarterly pattern? For a closer view of the patterns, zoom in to the range of 3500-5000 on the γ axis.

c. Create four separate line graph for Q1, Q2, Q3, and Q4. In each, plot a line graph. Recommend to order the data by Q1, Q2, Q3, Q4 (alphabetical sorting will work), and plot them as separate series on the line graph. Zoom in to the range of 3500-5000 on the γ axis. Does there appear to be a difference between quarters?

d. Create a line graph of the series at a yearly aggregated level (i.e., the total shipments in each year). Comment your graph

2. **Laptop Sales at a London Computer Chain: Histogram, Bar Charts and Boxplots (Quartile Chart).**

The file LaptopSalesJanuary2008.xls contains data for all sales of laptops at a computer chain in London in January 2008. This is a subset of the full dataset that includes data for the entire year.

- **a.** List all the variables that are quantitative
- **b.** Show the distribution of HDSize. What type of its distribution? Normal or Not?
- **c.** Show the distribution of RetailPrice. What type of its distribution? Normal or Not?
- **d.** Find whether there ate outliers in RetailPrice. Show your method to determine outliers
- **e.** What are mean, median, mode of HDSize and RetailPrice
- **f.** What are SD, variance, and IQR of HDSize and RetailPrice. Explain their meaning.
- **g.** Create a bar chart, showing the average RetailPrice by store. Which store has the highest average? Which has the lowest?
- **h.** To better compare retail-prices across stores, create side-by-side boxplots of retail price by store. Now compare the prices in the two stores above. Do you see a difference between their price distributions? Explain.

3. **Based on your point(s) of view, compare Excel & Python in terms of Exploration & Visualization Ability**