

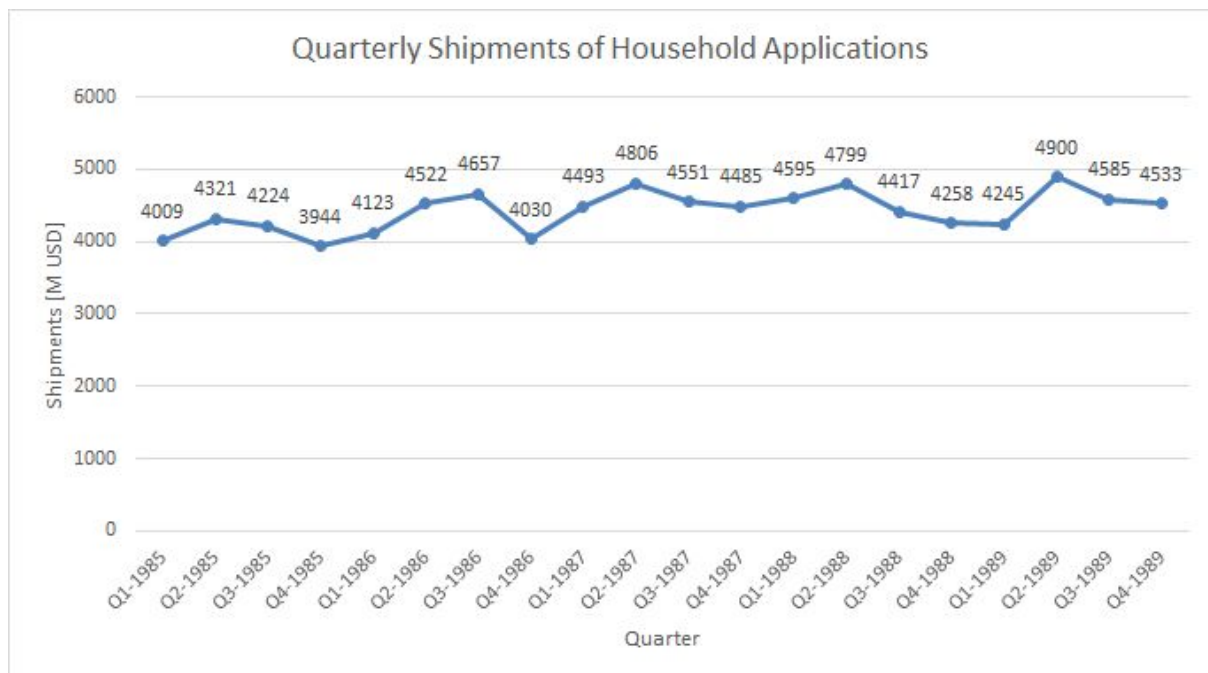
Data Mining - Homework 1

Sirakorn Lamyai, 5910500023

Exercise 1

(This part of the exercise is done on Excel)

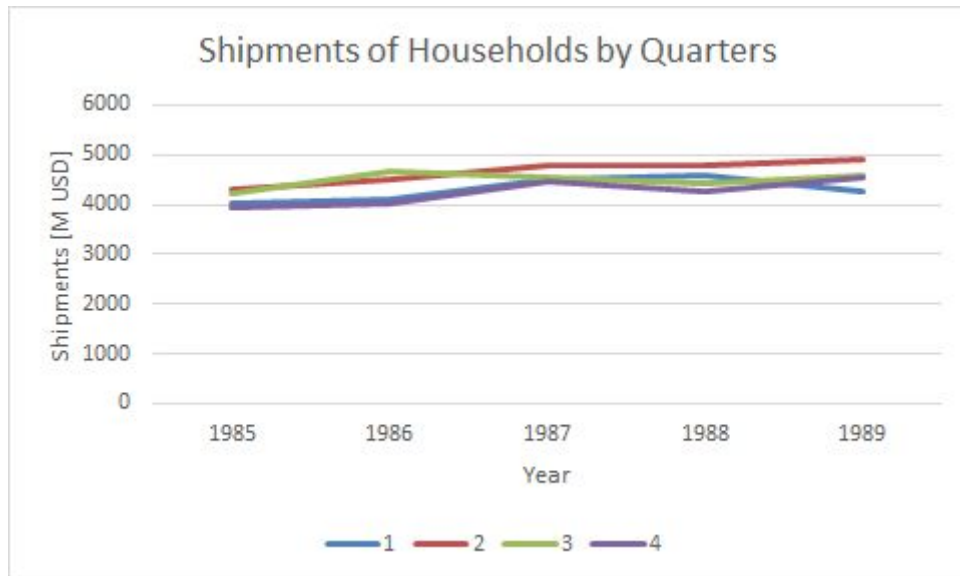
1.1



1.2

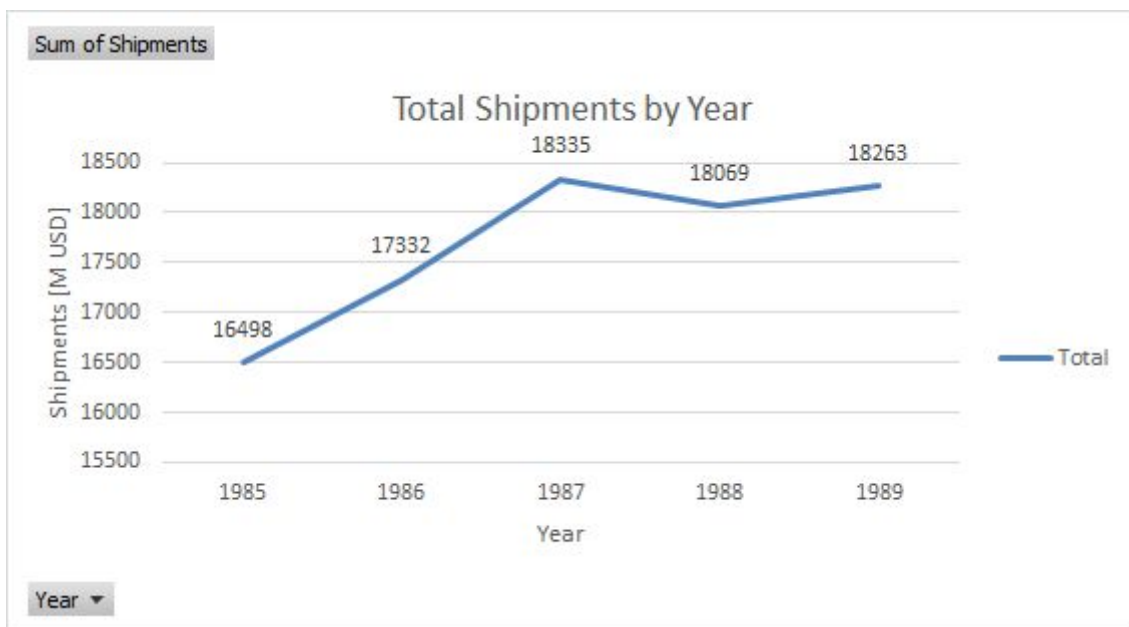
Yes, there exists a pattern based on a period of one year. It could be observed that the sale in one single year peaks up in the second quarter, before decreasing down on the third and fourth quarter.

1.3



Yes, there exist gaps between quarters. However, all of them show the same trend of increasing shipments.

1.4

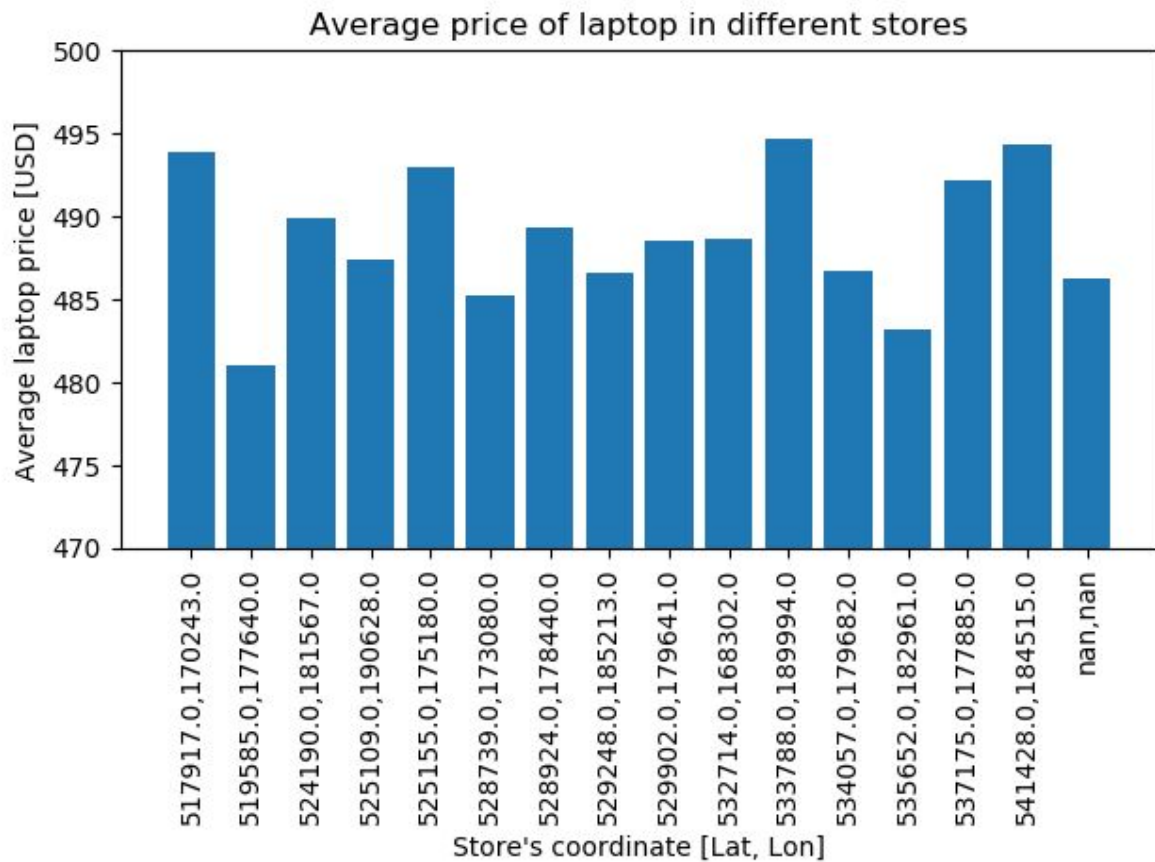


This chart illustrates the total shipments by year, summing all sales in the quarter in such a year. It could be observed that the shipments in a year start saturating in 1987, which before that year the sale is linearly increasing.

Exercise 2

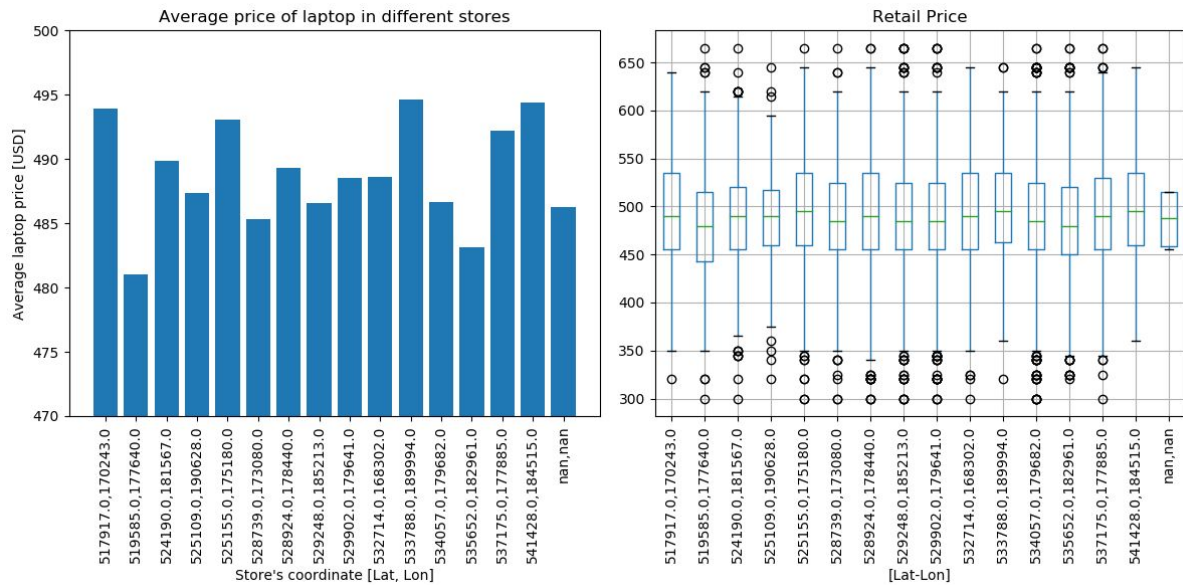
(This part of the exercise is done in Python. The Python code for the plots can be)

2.1



The store at (533788.0, 189994.0) and (519585.0, 177640.0) have the highest and lowest average laptop price sold respectively.

2.2

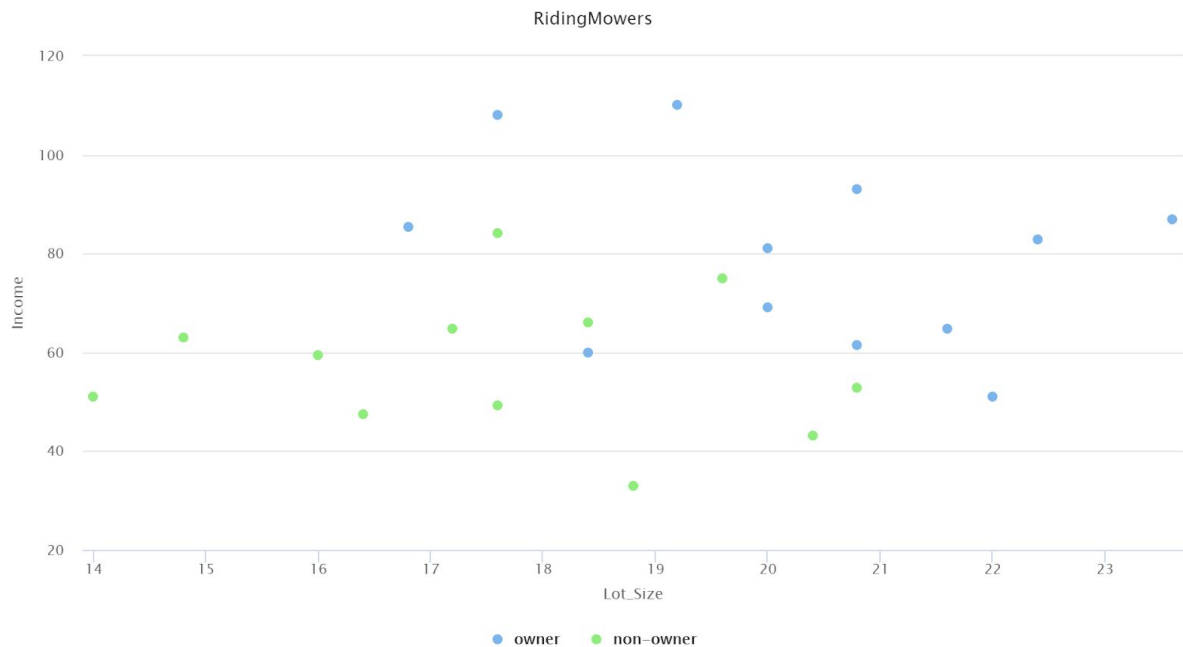


No, the distribution appears to be the same (normally distributed). However, it could be observed that the deviation of the store with the highest average laptop price sold is less than the store with the least average laptop price.

Exercise 3

(This part of the exercise is done on RapidMiner Studio)

3.1



3.2

By the intuition, it could be observed that the people who tend to own riding mowers have a higher income and lot size. This observation could lead to creating a simple model to predict the likeliness of a potential customer becoming the owner of riding mowers.

Exercise 4

It could be asserted the following likeliness and differences between Excel, Python, and RapidMiner:

- Excel offers outstanding data analytics tools. This includes PivotTable for data aggregation and grouping. The Graphical User Interface (GUI) makes it easy for non-programmers to use Excel.
- Python, with the power of additional libraries like Pandas and Matplotlib, is a very powerful tool for data analysis. Python offers a strongly flexible and customisable data analysis tools. When familiar with, ones can create complex plots and charts simply.
- RapidMiner is not an ideal tool for visualisation. Although it offers a solid data analytics tool. However, when comes to visualisation, it doesn't perform outstandingly.