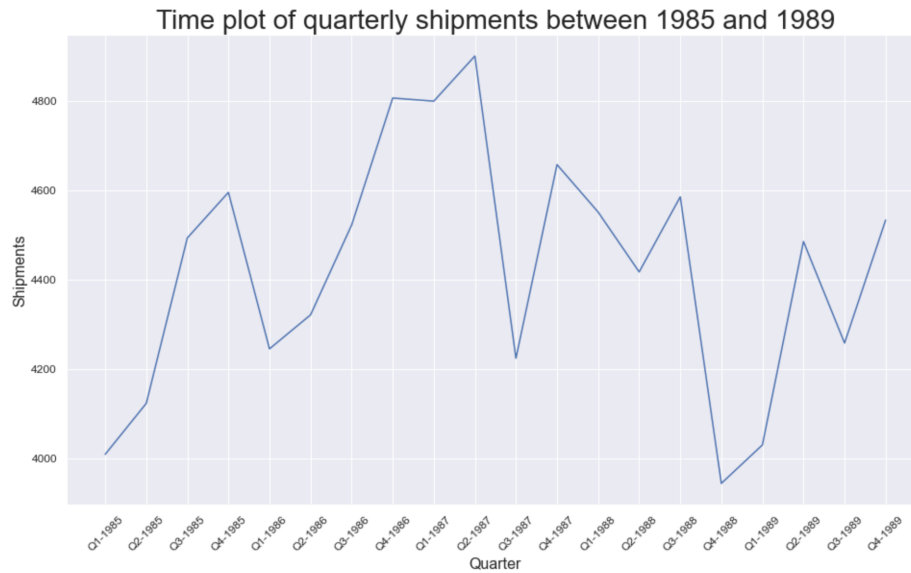
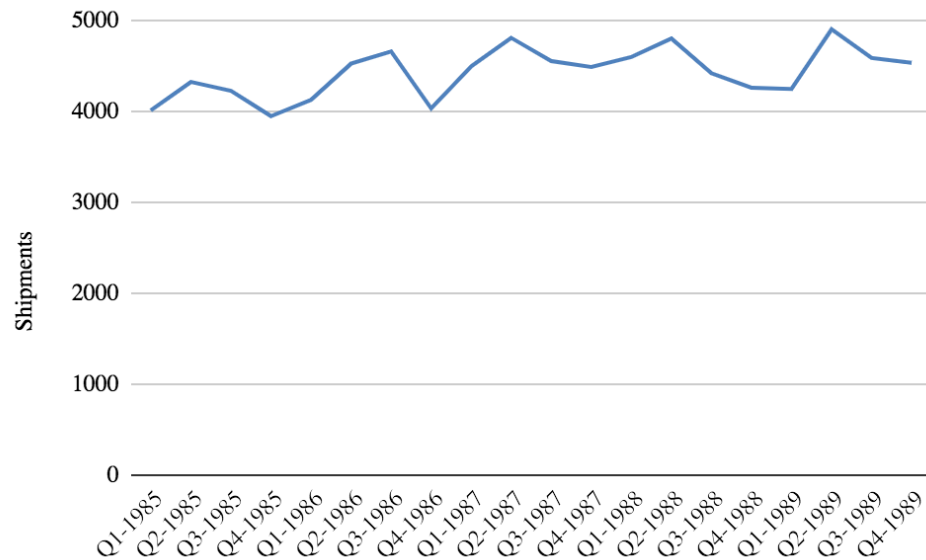


## 1. Shipments of Household Appliances: Line Graphs.

- a. Create a well-formatted time plot (line graph) of this data.

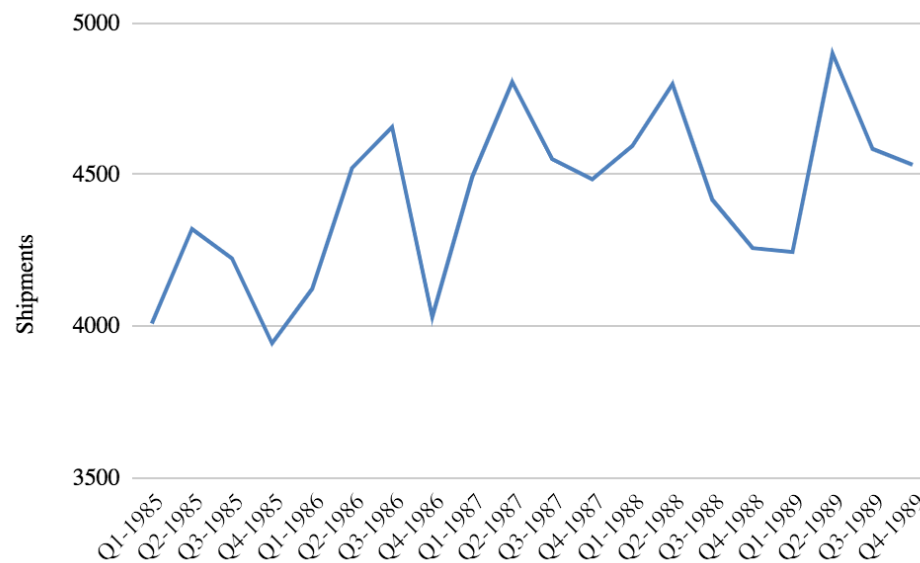
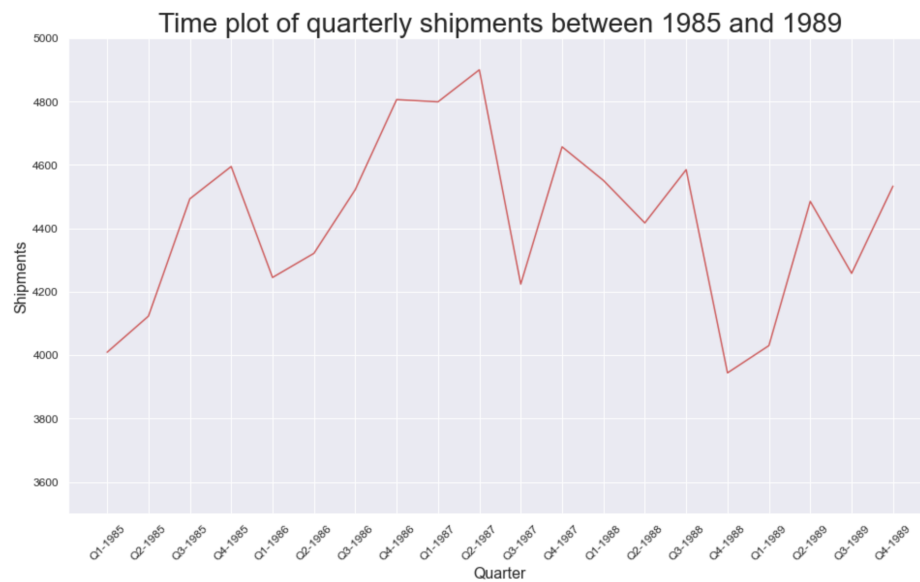


python



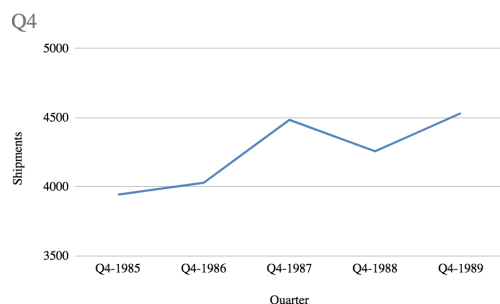
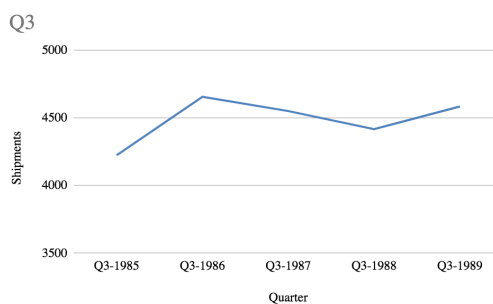
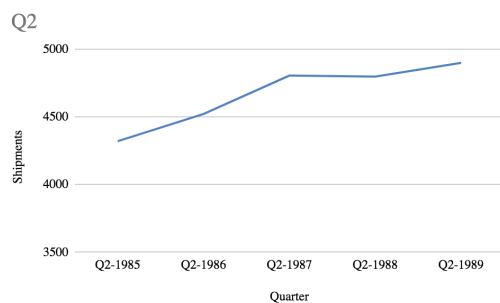
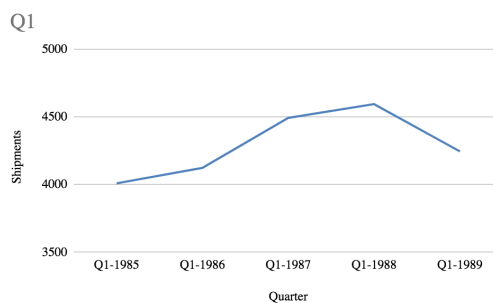
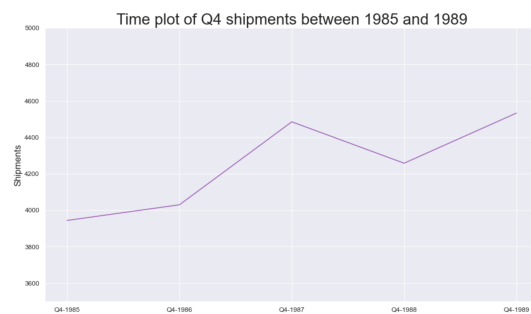
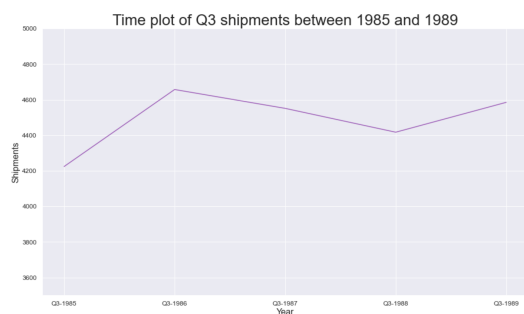
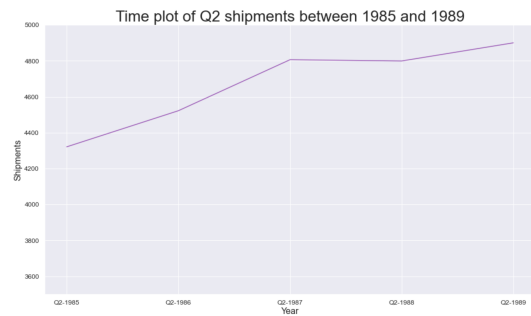
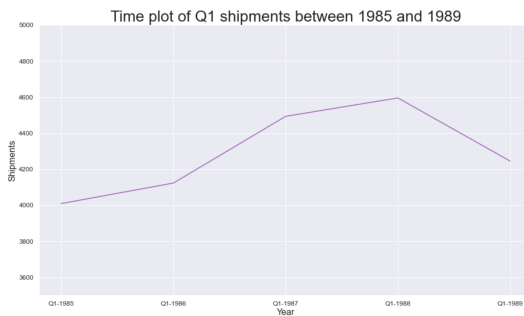
Excel

- 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 **Y** axis.



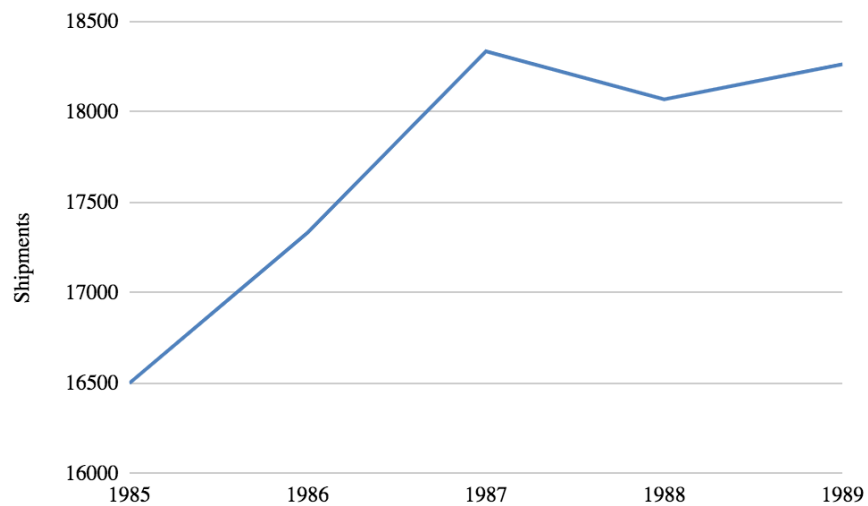
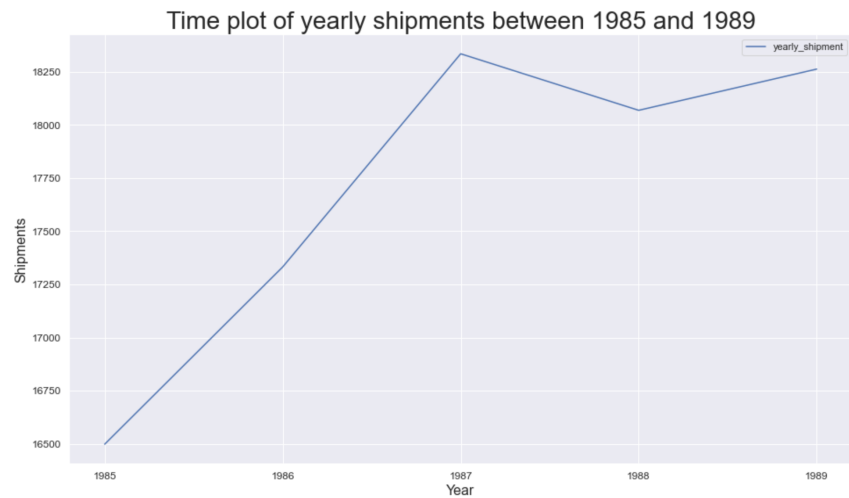
จากกราฟแสดงการขนส่งรายไตรมาสในปี 1985 ถึง 1989 แสดงให้เห็นว่าใน Q1, Q2 และ Q3 ส่วนใหญ่จะมีแนวโน้มที่เพิ่มขึ้นแต่ในทางกลับกันการขนส่งใน Q4 จะลดลง

- 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 **Y** axis. Does there appear to be a difference between quarters?



จากกราฟแสดงให้เห็นว่าใน Q2 มียอดการขนส่งที่สูงมากที่สุดอย่างเห็นได้ชัด ในขณะที่ Q1, Q2 และ Q3 มีปริมาณการขนส่งค่อนข้างใกล้เคียงกัน

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

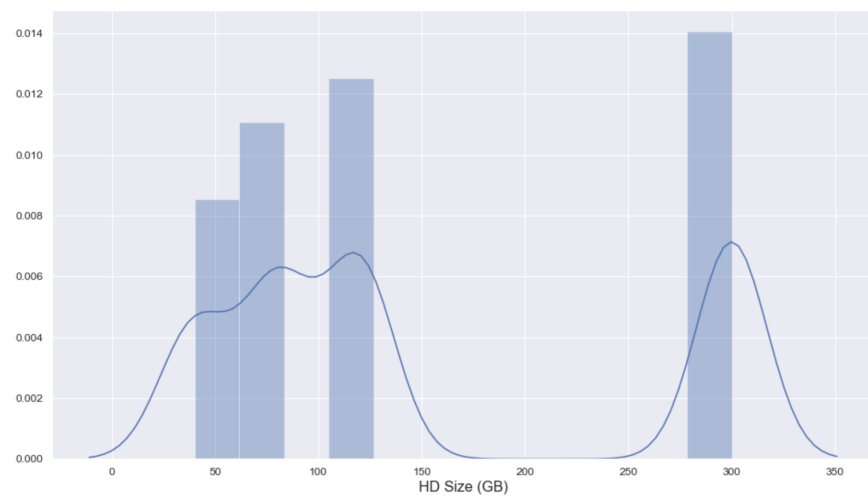


จากกราฟแสดงให้เห็นว่าปริมาณการขนส่งมีการเพิ่มขึ้นสูงมากในช่วงปี 1985 ถึง 1987 และค่อนข้างคงที่ไปจนถึงปี 1989

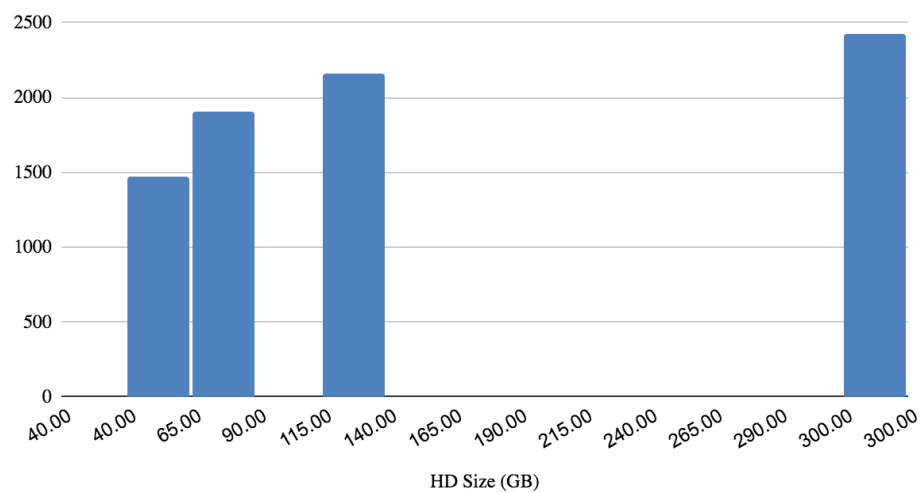
## 2. Laptop Sales at a London Computer Chain: Histogram, Bar Charts and Box plots (Quartile Chart).

- a. List all the variables that are quantitative
  - Configuration
  - Retail Price
  - Screen Size (Inches)
  - Battery Life (Hours)
  - RAM (GB)
  - Processor Speeds (GHz)
  - HD Size (GB)
  - OS X Customer
  - OS Y Customer
  - OS X Store
  - OS Y Store
  - CustomerStoreDistance

- b. Show the distribution of HDSize. What type of its distribution? Normal or Not?

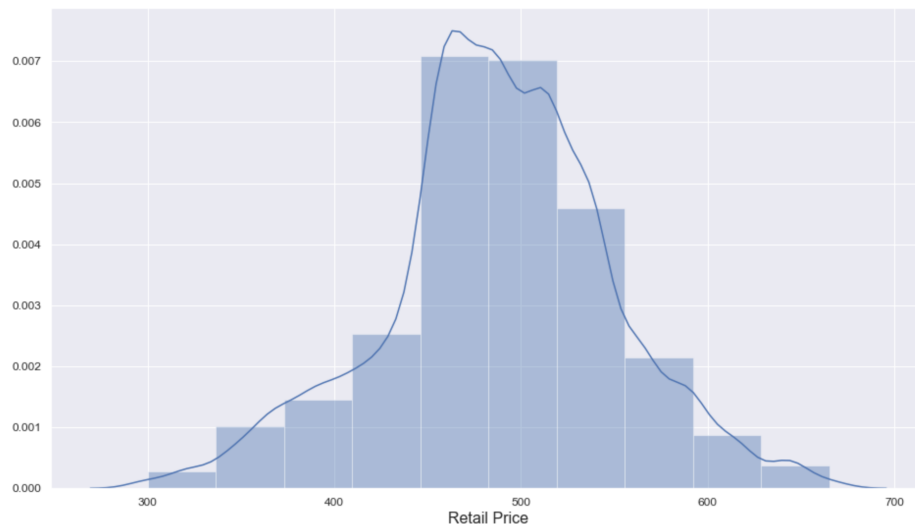


Histogram of HD Size (GB)



จากกราฟแสดงให้เห็นว่ามีจุด peak ของกราฟมากกว่า 1 จุด ทำให้กราฟแสดงการกระจายตัวของ HD Size ไม่ใช่ Normal distribution

- c. Show the distribution of RetailPrice. What type of its distribution? Normal or Not?

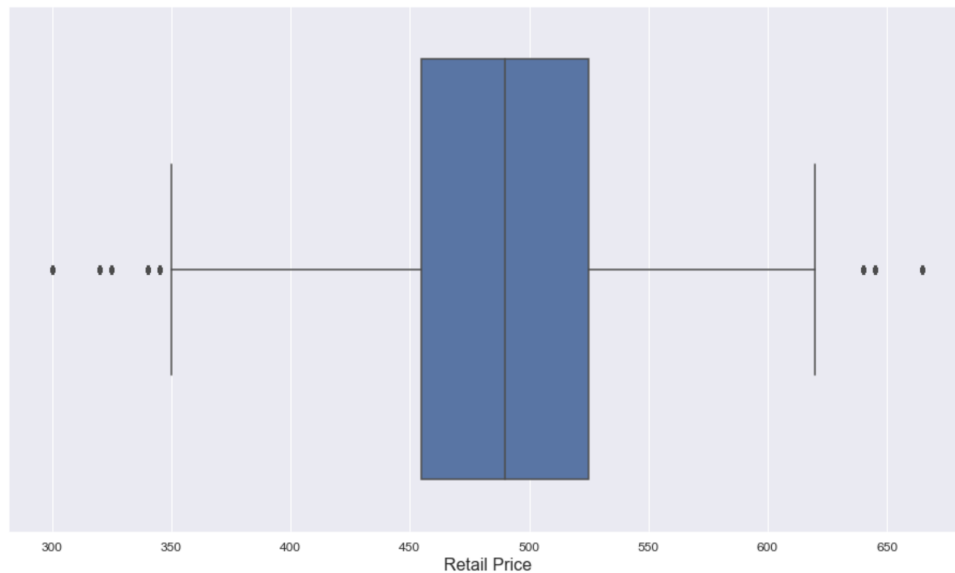


Histogram of Retail Price



จากกราฟแสดงให้เห็นว่ามีจุด peak ของกราฟเพียง 1 จุด รวมถึงจากการกระจายตัวของ HD Size ทำให้สรุปได้ว่าเป็น Normal distribution

- d. Find whether there ate outliers in RetailPrice. Show your method to determine outliers



จาก boxplot แสดงให้เห็นว่ามี Outlier ที่เกิน Upper Bound อยู่จำนวน 3 ตัว และ Outlier ที่เกิน Lower Bound อยู่ 5 ตัว

และทำการหา Outlier ได้ด้วยวิธีการคำนวณ

IQR (Interquartile Range) ด้วยสูตร

$Q3 - Q1$  ( $Q3$  = third quartile,  $Q1$  = first quartile)

ดังนั้นข้อมูลที่สูงกว่า 630 หรือต่ำกว่า 350 จะถือเป็น Outlier

```
Q1 = sales_data['Retail Price'].quantile(0.25)
Q3 = sales_data['Retail Price'].quantile(0.75)

# interquartile range (IQR)
IQR = Q3 - Q1
upper = Q3 + 1.5*IQR
lower = Q1 - 1.5*IQR
```

```
print('Upper: '+str(upper))
print('Lower: '+str(lower))
```

```
Upper: 630.0
Lower: 350.0
```

	Q1	Q3	IQR	Upper bound	Lower bound
Retail Price	455	525	70	630	350

Excel



- e. What are mean, median, mode of HDSize and RetailPrice

```
mean_HD_size = sales_data['HD Size (GB)'].mean()
mean_retail_price = sales_data['Retail Price'].mean()

median_HD_size = sales_data['HD Size (GB)'].median()
median_retail_price = sales_data['Retail Price'].median()

mode_HD_size = sales_data['HD Size (GB)'].mode()[0]
mode_retail_price = sales_data['Retail Price'].mode()[0]
```

```
mean HD Size: 150.3795877325289
mean Retail Price: 487.93489190548013
median HD Size: 120.0
median Retail Price: 490.0
mode HD Size: 300
mode Retail Price: 490
```

	Mean	Median	Mode
HD Size (GB)	150.3795877	120	300
Retail Price	487.9348919	490	490

- f. What are SD, variance, and IQR of HDSize and RetailPrice. Explain their meaning.

```
std_HD_size = sales_data['HD Size (GB)'].std()
std_retail_price = sales_data['Retail Price'].std()

var_HD_size = sales_data['HD Size (GB)'].var()
var_retail_price = sales_data['Retail Price'].var()

Q1_HD_size = sales_data['HD Size (GB)'].quantile(0.25)
Q3_HD_size = sales_data['HD Size (GB)'].quantile(0.75)
IQR_HD_size = Q3_HD_size - Q1_HD_size

Q1_retail_price = sales_data['Retail Price'].quantile(0.25)
Q3_retail_price = sales_data['Retail Price'].quantile(0.75)
IQR_retail_price = Q3_retail_price - Q1_retail_price
```

```
SD HD Size: 102.49830440874854
SD Retail Price: 61.51851757214998
variance HD Size: 10505.90240666848
variance Retail Price: 3784.528004274926
IQR HD Size: 220.0
IQR Retail Price: 70.0
```

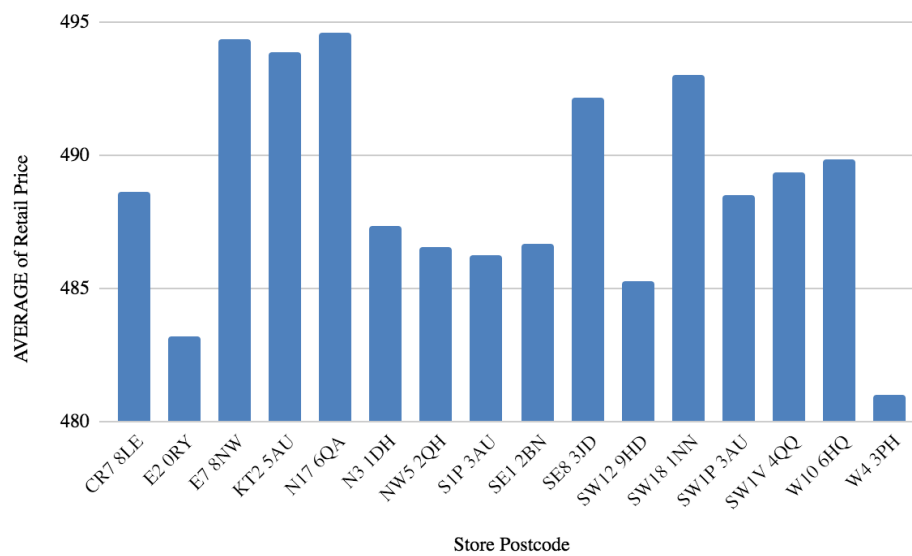
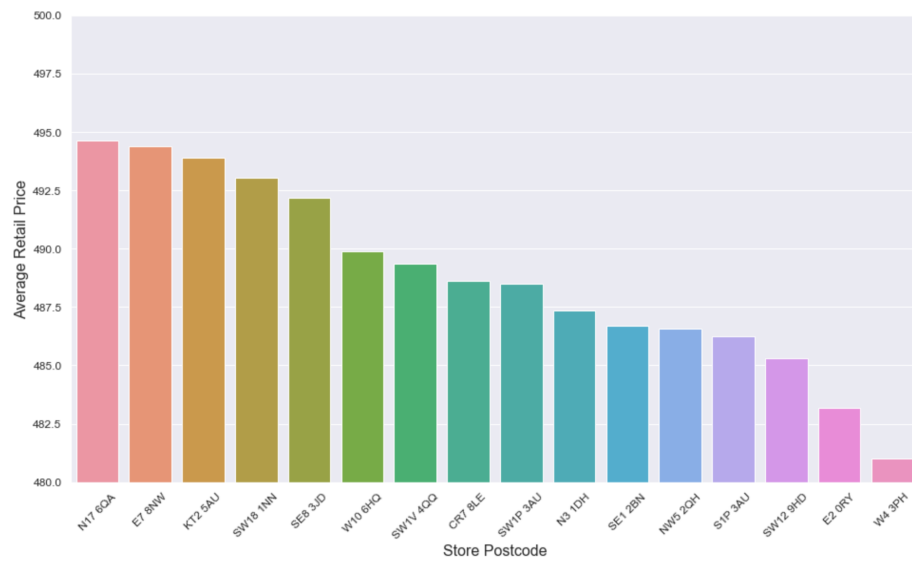
	SD	variance	IQR
HD Size (GB)	102.4983044	10505.90241	220
Retail Price	61.51851757	3784.528004	70

SD หรือ Standard deviation คือ ค่าเบี่ยงเบนมาตรฐาน ใช้เพื่อดูการกระจายข้อมูลจากค่าเฉลี่ย

Variance คือ ค่าความแปรปรวน ใช้เพื่อวัดการกระจายของข้อมูลคิดจากค่าเฉลี่ยของความต่างจากค่าเฉลี่ยยกกำลัง 2

IQR หรือ Interquartile range คือ ความแตกต่างระหว่างควอไทล์แรกและควอไทล์ที่สาม :  $IQR = Q3 - Q1$

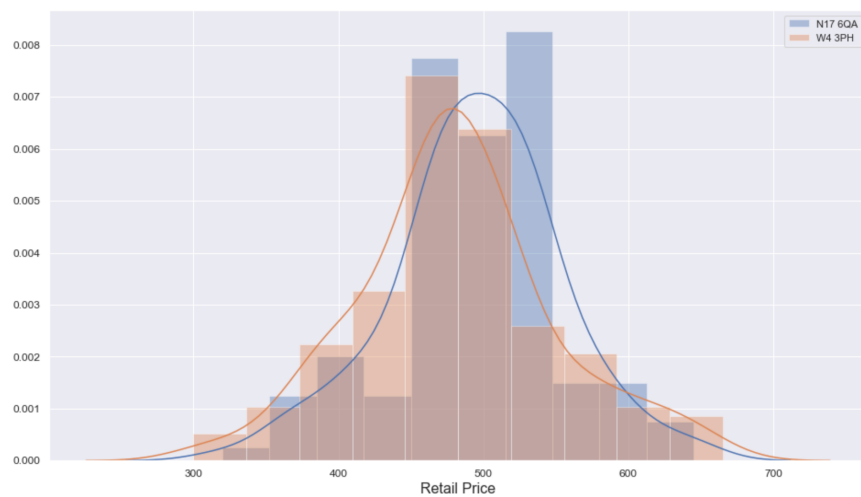
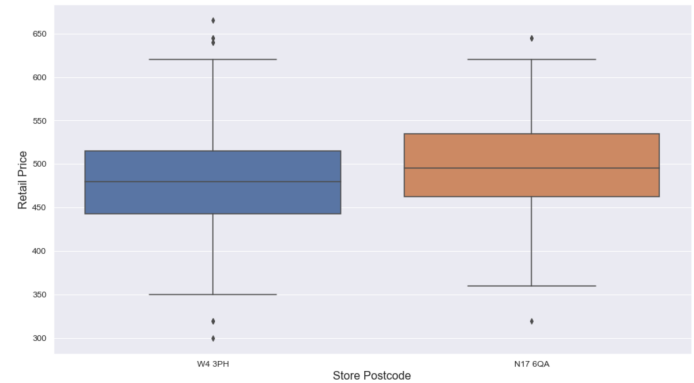
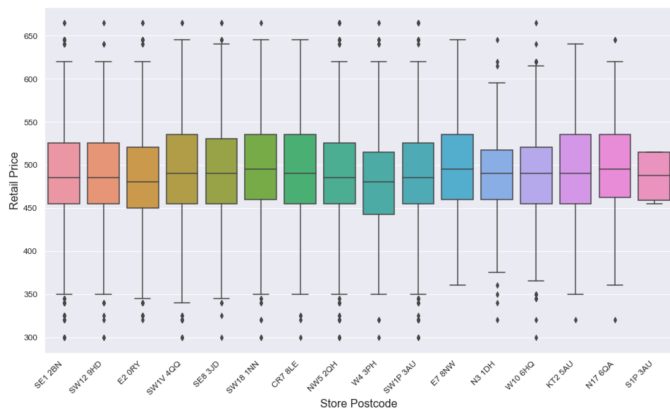
- g. Create a bar chart, showing the average RetailPrice by store. Which store has the highest average? Which has the lowest?



ร้าน N17 6QA มีราคาขายปลีกเฉลี่ยสูงสุดอยู่ที่ 494.63

ร้าน W4 3PH มีราคาขายปลีกต่ำสุดอยู่ที่ 481.01

- h. To better compare retail-prices across stores, create side-by-side box plots of retail price by store. Now compare the prices in the two stores above. Do you see a difference between their price distributions? Explain.



จากกราฟแสดงให้เห็นว่าร้าน N17 6QA ซึ่งเป็นสีฟ้ามีจุด peak รวมทั้งข้อมูลส่วนใหญ่ที่กระจายตัวอยู่ทางขวาของร้าน W4 3PH ซึ่งเป็นสีส้มอย่างเห็นได้ชัด ทำให้สามารถวิเคราะห์ได้ว่าราคาขายปลีกเฉลี่ยของร้าน N17 6QA มากกว่า W4 3PH

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

โดยส่วนตัวคิดว่าประสิทธิภาพในการ Exploration & Visualization ของ Python มีมากกว่า Excel อยู่พอสมควร เริ่มจากการจัดการกับข้อมูลจำนวนมาก Excel จะมีความหน่วงในการทำงานที่มากกว่าการใช้ python อย่างเห็นได้ชัด เนื่องจาก Python มี library อย่าง pandas ที่ใช้สำหรับจัดการและวิเคราะห์ข้อมูลขนาดใหญ่ และในส่วนของสร้างกราฟต่างๆ Excel ทำได้ค่อนข้างง่ายโดยแค่เลือก column และกดสร้างกราฟ ต่างจาก Python ที่จำเป็นต้องเขียน code เพื่อสร้างกราฟ แต่เนื่องจากความง่ายในการใช้งานของ Excel ทำให้เกิดข้อจำกัดในการใช้งานขึ้น ในขณะที่ Python ถึงแม้จะใช้งานได้ยากกว่า แต่ก็สามารถจัดการกับกราฟได้มากกว่า สรุปได้ว่าหากต้องการวิเคราะห์ข้อมูลอย่างรวดเร็วและไม่ได้ต้องการปรับแก้อะไรมากสามารถใช้ Excel ได้เลย แต่หากข้อมูลมีปริมาณมากและต้องการจัดการเกี่ยวกับการวิเคราะห์ที่มีความละเอียดมากขึ้นควรเลือกใช้ python