

From the above side by side box plot we can see that there is slight difference between two stores. Interquartile range for the two plots is approximately equal.

Problem 3.4

3.4 Laptop Sales at a London Computer Chain: Interactive Visualization. *The next exercises are designed for using an interactive visualization tool. The file Laptop-Sales.txt is a comma-separated file with nearly 300,000 rows. ENBIS (the European 78 DATA VISUALIZATION Network for Business and Industrial Statistics) provided these data as part of a contest organized in the fall of 2009.*

Scenario: Imagine that you are a new analyst for a company called Acell (a company selling laptops). You have been provided with data about products and sales. You need to help the company with their business goal of planning a product strategy and pricing policies that will maximize Acell's projected revenues in 2009. Using an interactive visualization tool, answer the following questions.

a. Price Questions:

- i. At what price are the laptops actually selling?

[Answer to 3.4.a.i:](#)

ii. Does price change with time? (Hint: Make sure that the date column is recognized as such. The software should then enable different temporal aggregation choices, e.g., plotting the data by weekly or monthly aggregates, or even by day of week.)

A line graph (top right) shows pretty large changes over time (we show it per week, which is possible only if the date column is properly defined in date format). The histogram (lower right) shows that the highest volume of sales is for medium-priced laptops. Note that the retail price can be “auto-binned” in Spotfire (right-click on the X-Axis control to activate that function)

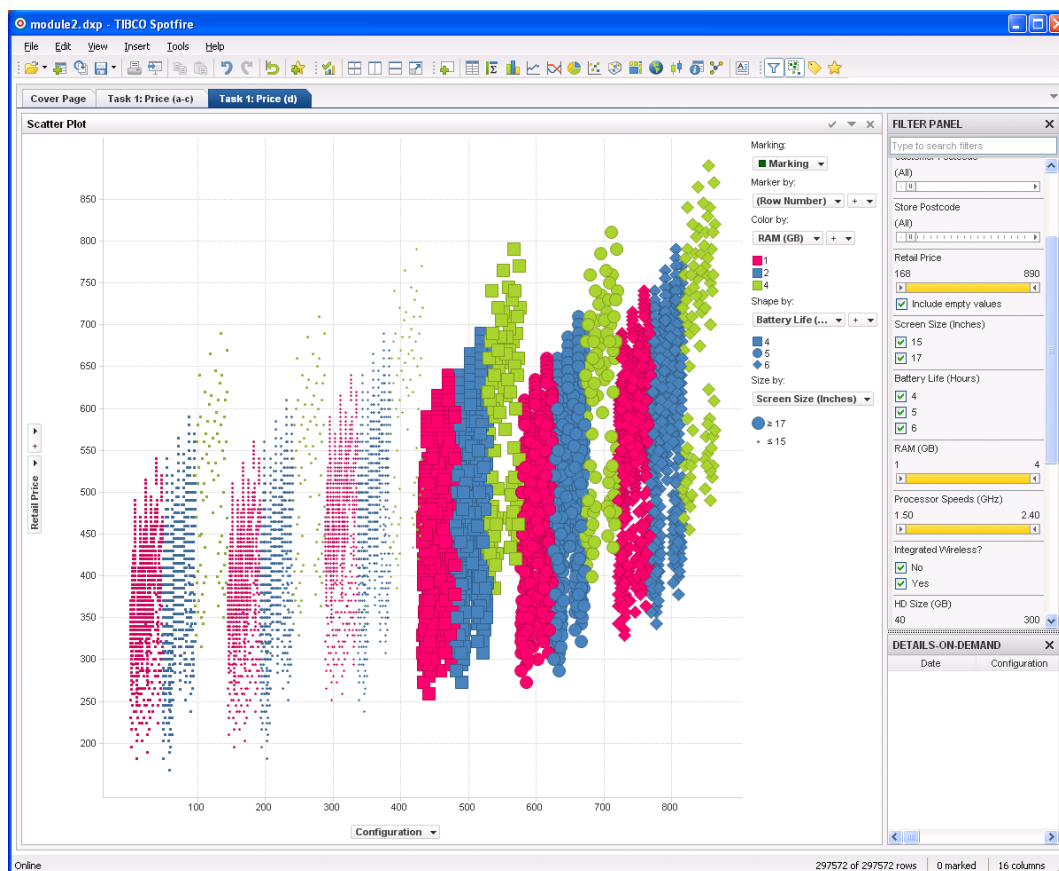
iii. Are prices consistent across retail outlets?

Answer to 3.4.a.ii:

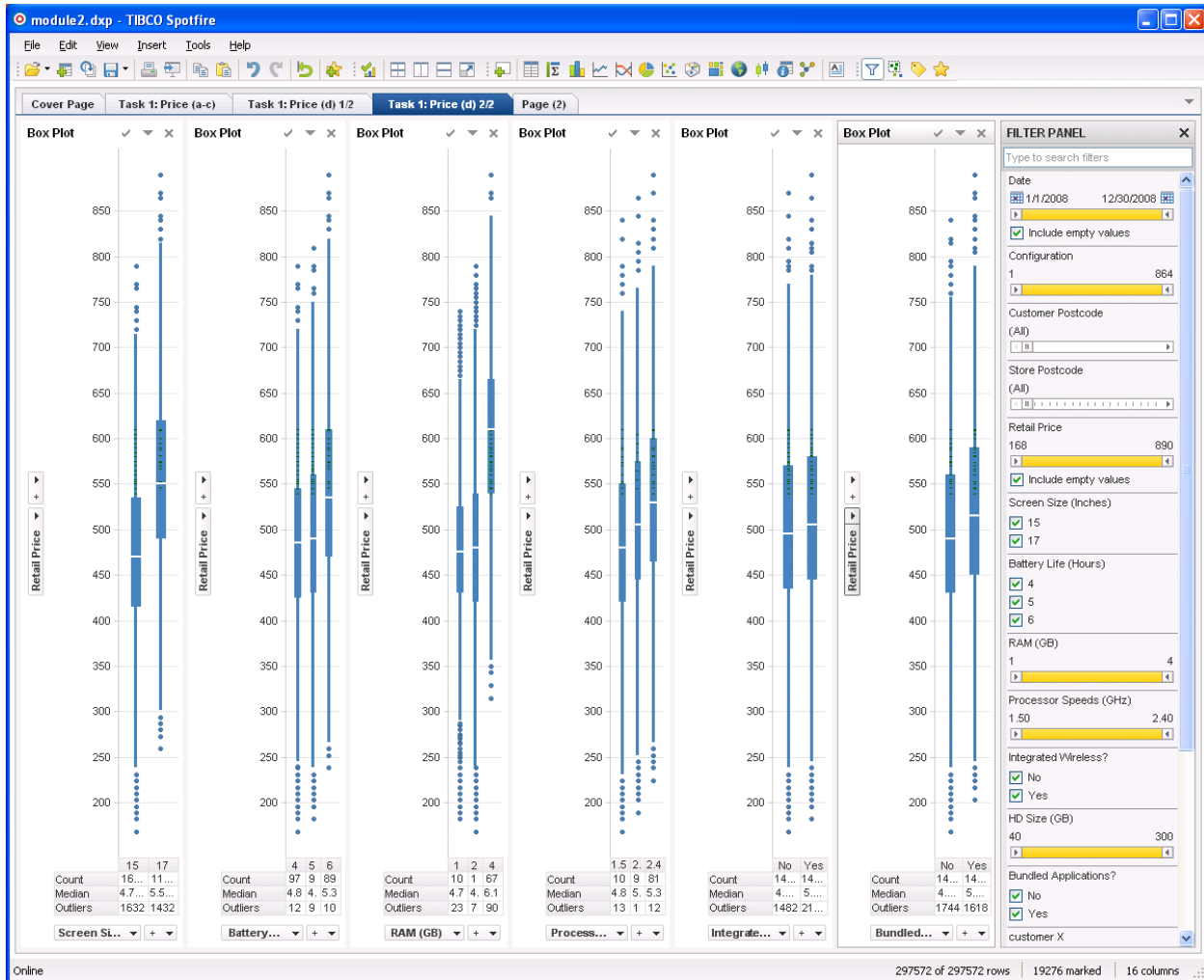
The bottom left visualization shows boxplots of price by store. We can see that overall the median price is similar across stores (around \$500, plus minus \$50). However, it is easy to see that there are two types of stores: those with larger price ranges and those with smaller price ranges.

iv. How does price change with configuration?

Answer to 3.4.a.iv:



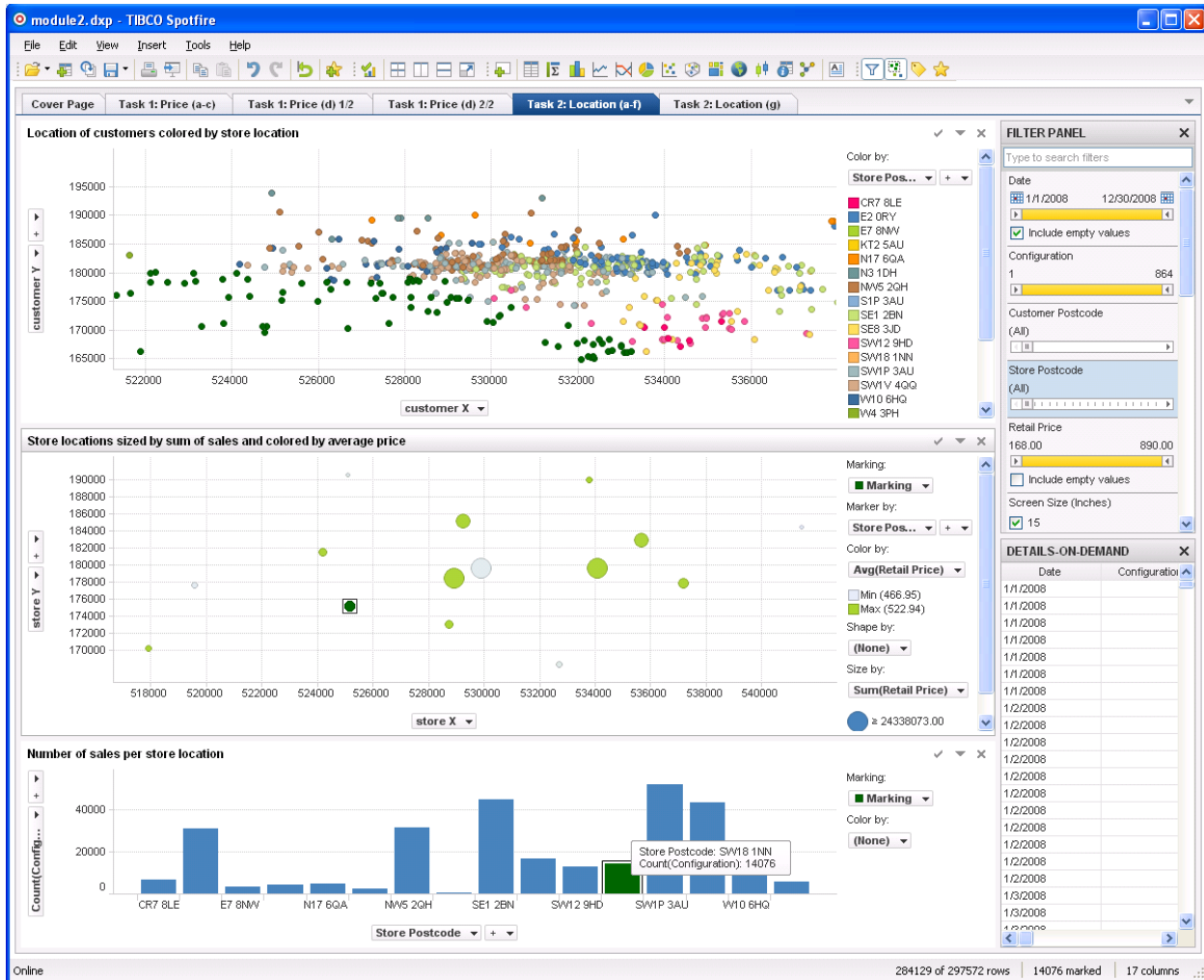
Here we tried to show all configurations at once. We mapped different attributes to size, color and shape, making the display richer, but also more complex to interpret. Depending on your audience, you should decide how much is “too much encoding”.



Here we compare the different configurations one by one. Some configuration options make a bigger price difference than others (e.g., screen size has a clear impact, while bandwidth does not, RAM size makes a difference mostly if you choose the largest size.)

Note that it is helpful to hide the legend which appears by default in Spotfire in each boxplot, to maximize the space for the boxplot (use the visualization properties control – or right-click - to hide the legend).

b. Location Questions:



i. Where are the stores and customers located?

Answer to 3.4.b.i:

The top and middle scatter plots correspond to the customers' and stores' locations. Note that the x-axis and y-axis scales are identical in both, for easy comparison). Most of the customers are located in the center (downtown) and in the south-east. Most of the stores are located downtown, with a few stores scattered in the suburbs. We used circle size in the stores scatterplot to denote average price. We can immediately see that the downtown stores sell the most expensive laptops, on average. Also, using color on the customers' scatterplot shows us that many customers (dark green) shop at a particular store (also colored in dark green).

ii. Which stores are selling the most?

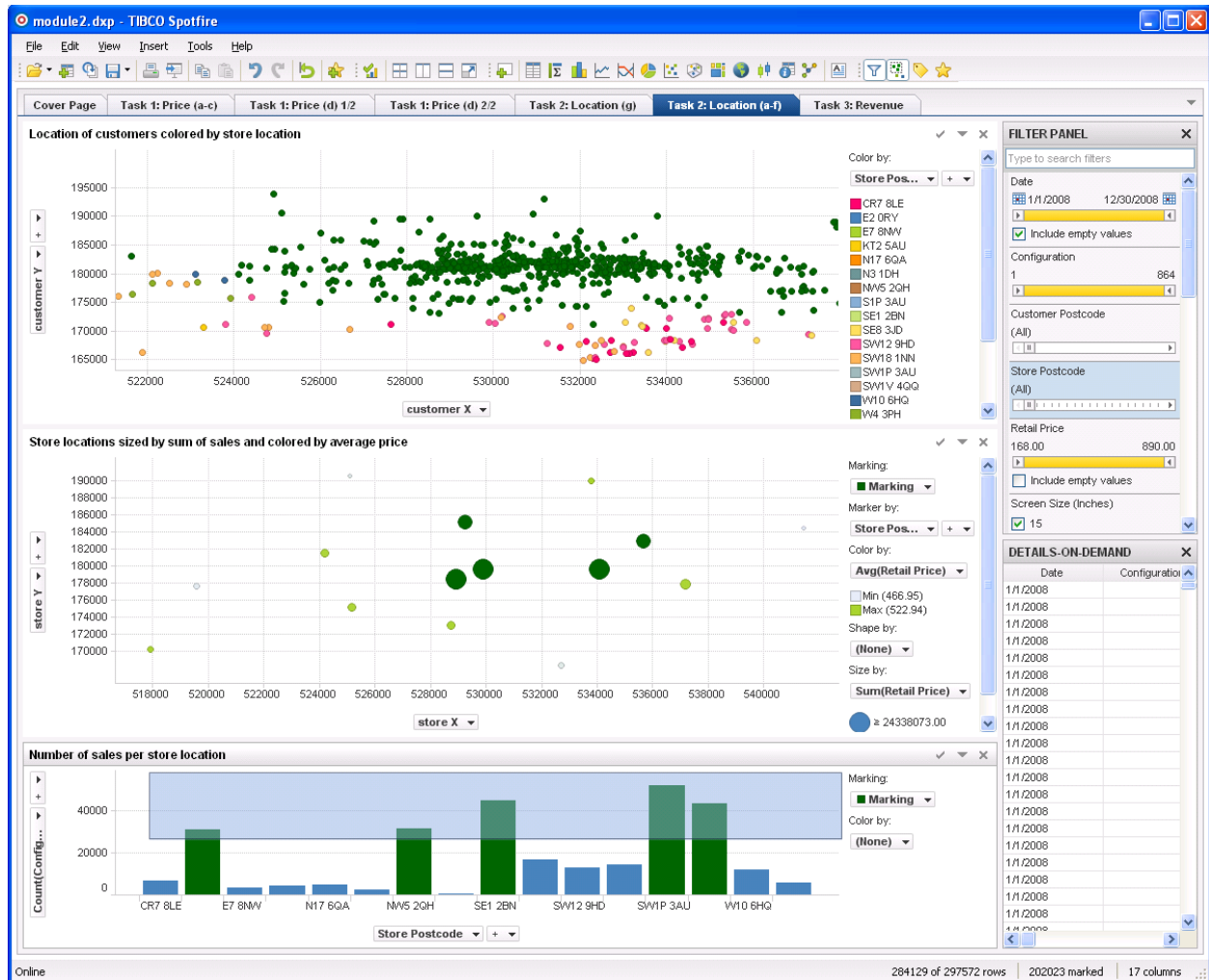
Answer to 3.4.b.ii:

In general, stores downtown seem to sell more than stores in the suburbs.

One interesting outlier in the scatter plot is the store that brings a fair amount of money but sells a smaller number of configurations (large but lightly colored circle in the center).

The bar chart reveals one store that has very few sales. This might be a new store -- Plotting a time line could help to confirm this hypothesis.

If you select the stores with the highest revenue, they are all downtown and customers come mostly from the north of the town (see below).



iii. How far would customers travel to buy a laptop?

Hint 1: you should be able to aggregate the data, e.g., plot the sum or average of the prices.

Hint 2: Use the coordinated highlighting between multiple visualizations in the same page, for example, select a store in one view to see the matching customers in another visualization.

Hint 3: Explore the use of filters to see differences. Make sure to filter in the zoomed out view. For example, try to use a “store location” slider as an alternative way to dynamically compare store locations. This might be more useful to spot outlier patterns if there were 50 store locations to compare.

Answer to 3.4.b.iii:

To study the distance travelled by customers you can visually estimate it on the visualizations, or use create a new data column with the distance between the store and the customer (in Spotfire: *Insert > Calculated column*). The formula would be $\sqrt{([OS\ X\ Customer] - [OS\ X\ Store])^2 + ([OS\ Y\ Customer] - [OS\ Y\ Store])^2}$. The snapshot below shows how such a column can be created in Spotfire.

The screenshot shows the 'Column Properties' dialog box in Spotfire. The 'Columns and hierarchies' tab is active, displaying a list of columns. The 'Distance travelled...' column is highlighted, showing it is a 'Calculated' column of 'Real' data type. Below the list, the 'General' tab is selected, showing the column's name, description, and data type. The formula for the calculated column is entered in the 'Formula' field.

Name	ColumnType	Data Type	ExternalName	ExternalId	IsValid
Integrated Wirele...	Imported	String	Integrated Wirele...		True
HD Size (GB)	Imported	Integer	HD Size (GB)		True
Bundled Applicati...	Imported	String	Bundled Applicati...		True
customer X	Imported	Integer	OS X		True
customer Y	Imported	Integer	OS Y		True
store X	Imported	Integer	OS X		True
store Y	Imported	Integer	OS Y		True
Distance travelle...	Calculated	Real			True

General | Formatting | Properties | Sort Order

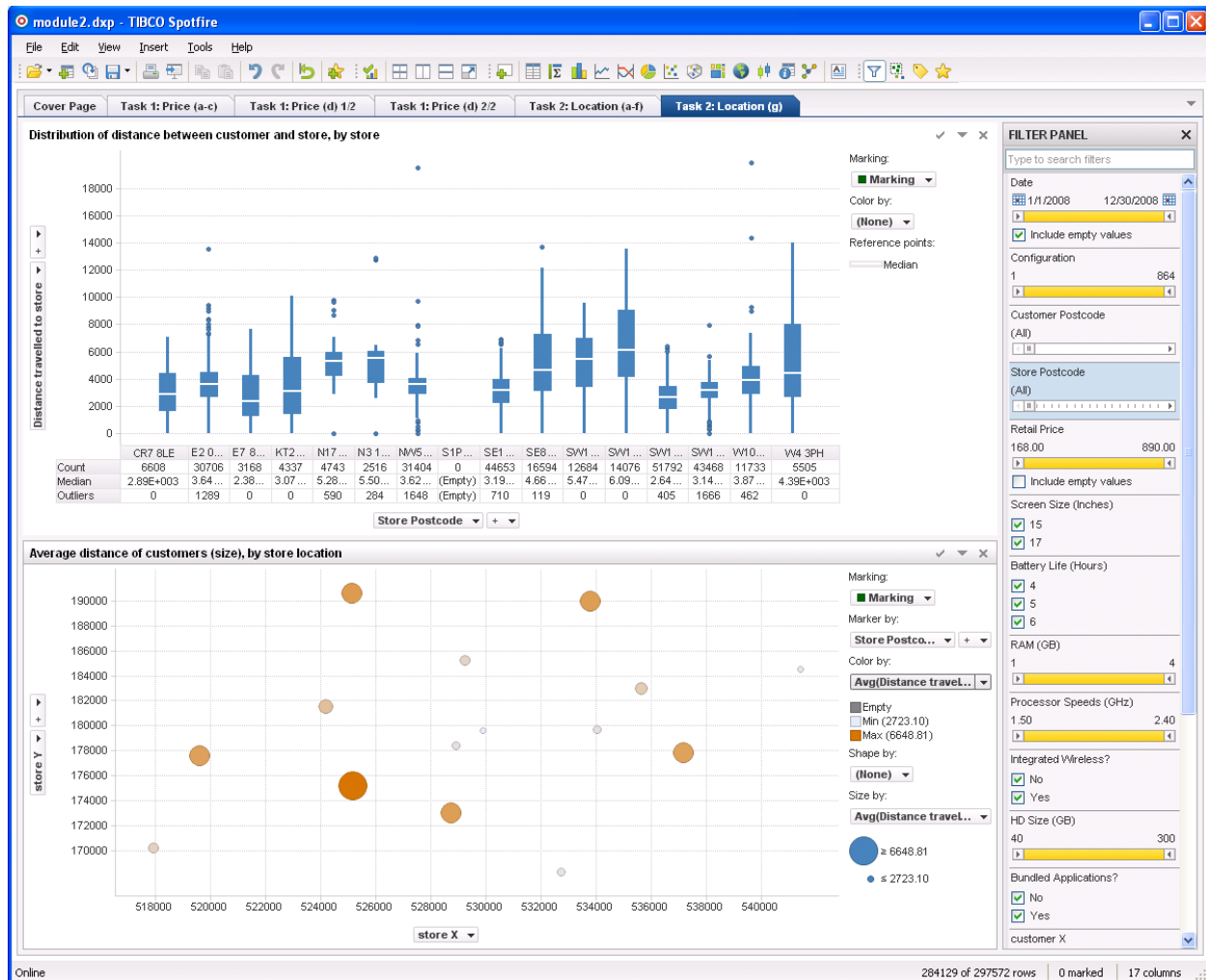
Name: Distance travelled to store

Description:

Data type: Real

Formula: $\sqrt{([customer\ X] - [store\ X])^2 + ([customer\ Y] - [store\ Y])^2}$

Buttons: Insert, Delete, Edit..., Freeze Column, OK, Cancel, Apply, Help



Stores in the suburbs require customers to travel longer distances.

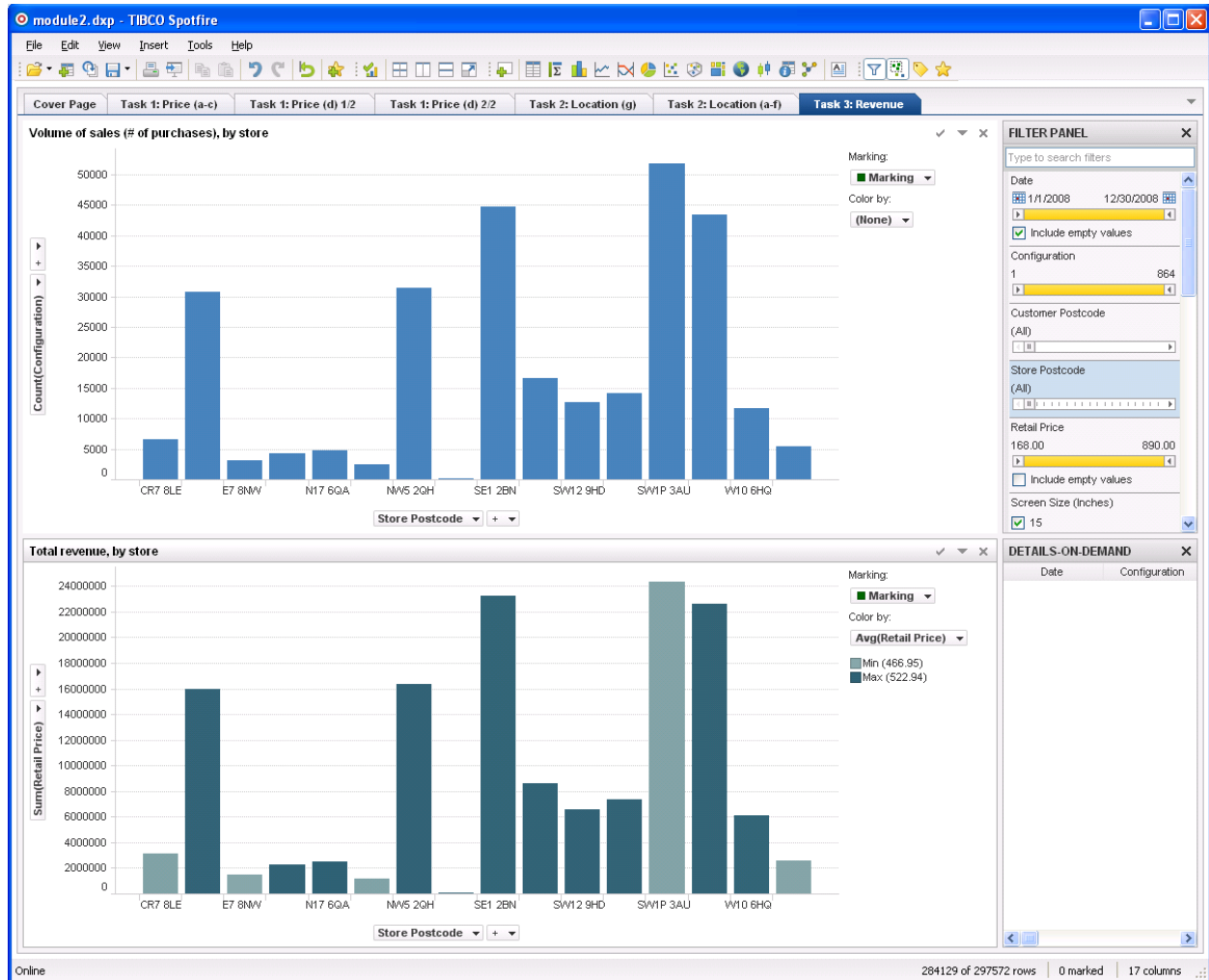
The store corresponding to the large circle in the lower left area of the scatter plot requires the longest average travel. If you go back to the preview page showing where the customers come from, and click on that store location to select it (shown in dark by default) you can see that indeed customers come from far away.

iv. Try an alternative way of looking at how far customers traveled. Do this by creating a new data column that computes the distance between customer and store.

c. Revenue Questions:

i. How do the sales volume in each store relate to Acell's revenues?

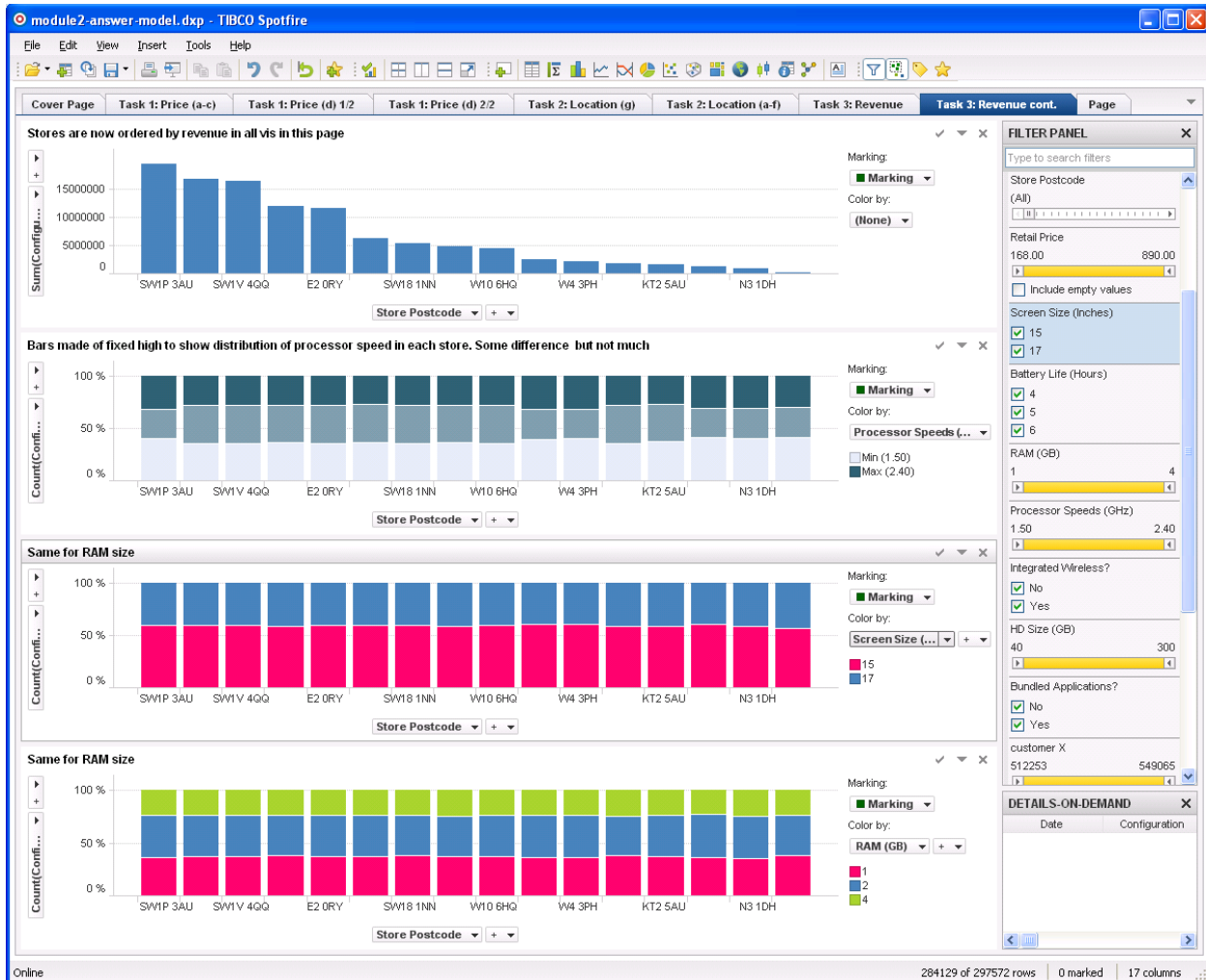
Answer to 3.4.c.i:



There are big differences between the stores. The number of sales and total revenue show similar patterns (see above). We reinforced the slight differences in the bottom chart by color coding by the average retail price, showing which stores tend to sell cheaper laptops. One store with high revenue sells lower price laptops.

ii. How does this relationship depend on the configuration?

Answer to 3.4.c.ii:



There does not appear to be much difference in the distribution of configuration options between the stores (see above).

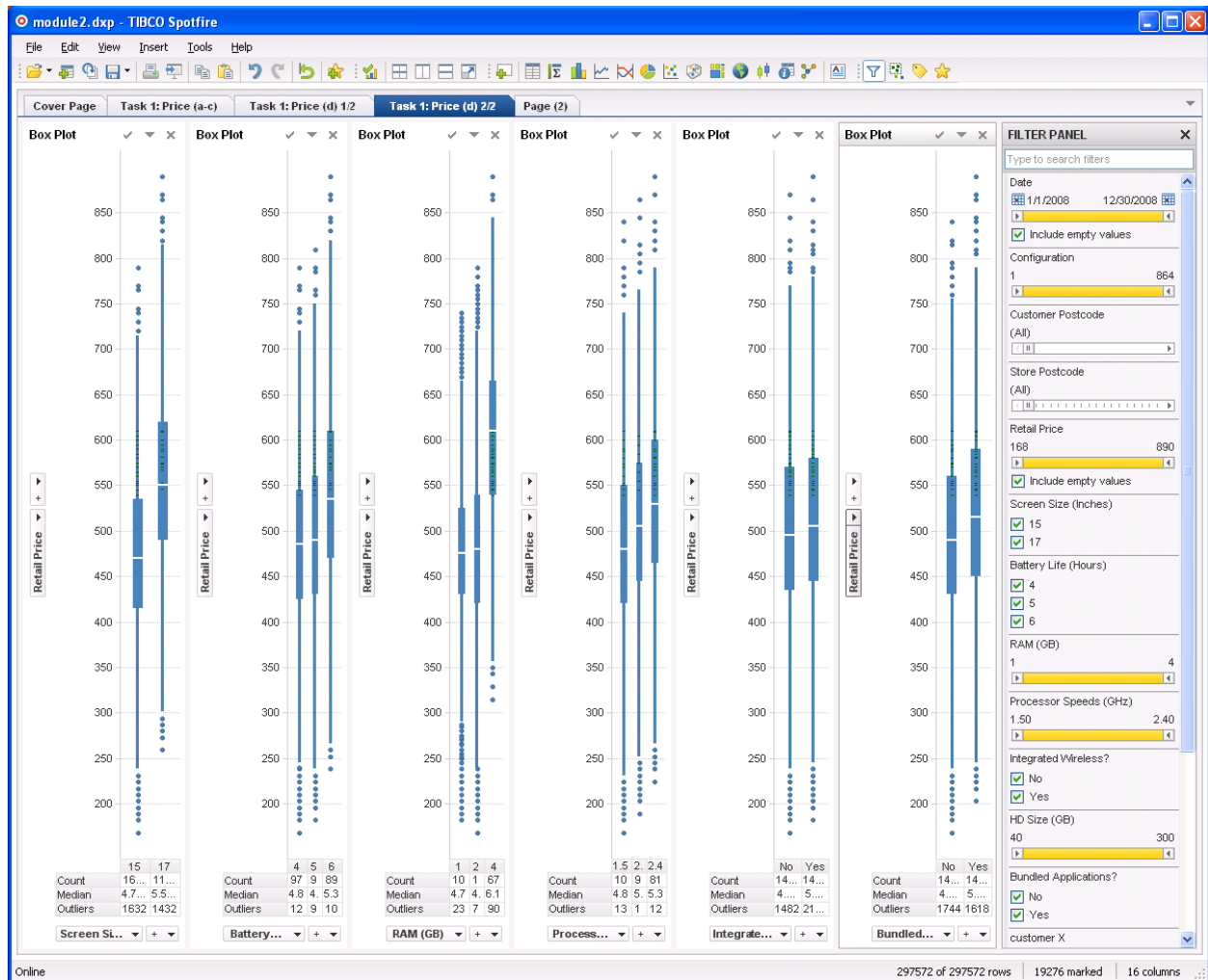
Here, we first sorted all the bars by size (i.e., we sorted the stores by revenue), and then colored by different options (e.g., speed, screen size, and Ram)

d. Configuration Questions:

i. What are the details of each configuration? How does this relate to price?

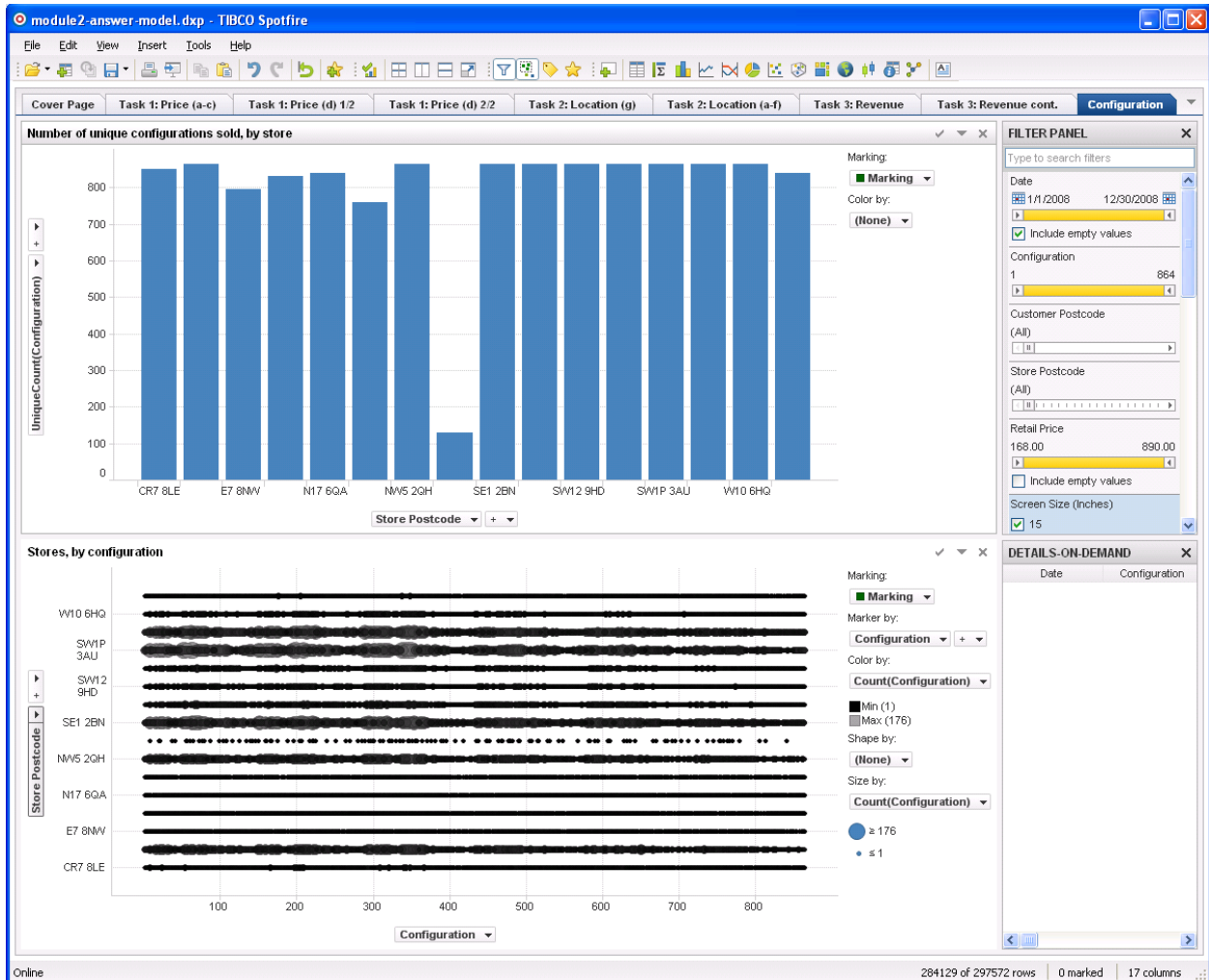
Answer to 3.4.d.i:

One of the screenshots shown in (a) already partially answered the revenue question (we reproduce it below). This set of boxplots shows the effect of various configuration options on the retail price.



ii. Do all stores sell all configurations?

Answer to 3.4.d.ii:



The bar chart in the top panel shows that all stores except one sell most configurations.

The bottom scatterplot display shows the same information in a different way. Sizing by the number (count) of configurations gives an idea of which configurations are more successful.