Quick introductory project.
(1) Log on to VCL following instructions.

(2) Double click Enterprise Miner icon (wait a minute or so for EM to open).

(3) Look at the properties panel at the left.
   Right click on data sources and click on “Create Data Source” pop-up
   Click “next”
   On step 2, “Select metadata source,” click “browse”
   Double click AAEM on left to open your folder then click on the Airline table
   Click “OK.”
   Click “next” until you get to step 5 “Column metadata.”
   Hold down CTRL and click on Lair, Air and Date then click Explore
   Select Actions->plot then select the line plot and next
   Give AIR the role of Y and DATE the role of X (click in role white space) then select finish.
   Discuss the plot and whether a regression approach would work well
   Make a plot of Lair versus date. Which plot looks better – why?
   Close the explore window.

(4) Mark AIR as rejected (Why?) and Lair as target
   What is the logic behind the ordering of variables mon1 – mon9?

(5) Click next until you get to the last window then click finish.
   Optional: Right click the Airline data source and select “Explore.” What happens?

(6) Click Diagrams in the properties panel then right click and create diagram. Call it Airline.
   Note the color change in the big diagram window and the label it now has.

(7) Drag and drop the Airline DATA SOURCE into the diagram window.

(8) Notice the “subtabs” above the diagram window. Click the Model subtab.

(9) Drag a regression (third from right) node into your diagram and connect the Airline data source to it with an arrow.
   To connect, move your cursor the right edge of the Airline node in the diagram window. It becomes a pencil and a little grey tab extends from the node. Hover over that tab. Press and hold the (left) mouse key and slide over to the middle of the left edge of the Regression node. Release the mouse key. The arrow should remain. If not try again.
(10) Run the Regression node (right click, select “Run”). Click Results when it completes and investigate its contents.

What is going on with the regression parameters? In the “Effects Plot” where is the coefficient for mon9 ??

(11) Close the results window.

(12) Click on the regression node to be sure its properties panel is displayed. From the properties panel, select the Ellipsis (…) next to “Exported Data.” From the popup window click on the training data (Role = Train) and click “Explore.”

(13) Plot the predicted values and actual values against date (overlaid).
   Actions -> Plot, select line plot as before
   Roles: Date=X, Lair=Y (click allow multiple role assignments) P_Lair=Y also.
   Optional: R_Lair=Tip
   Click “Finish.”

(14) Plot residuals against predicted.
   Scatter plot icon, X=P_Lair Y=R_Lair. Does it look OK?

---

Detail:
To get the graph of means in 5% chunks, SAS Enterprise Miner does this:
(1) sort by descending predicted values
(2) pick out the best (for example) 5% of predicted values and average them.
   Note: look at the output window and you will see it is an average of 8 observations then 7 for a while then back to 8. \{8 7 7 7 8 7 7 7 8 7 7 7 etc.\} Now 144(0.05) = 7.2 and 5(7.2 )=36 = 8+7+7+7+7 so they are using integer numbers of points that sum correctly at certain intervals.
(3) average the target values associated with these predicted values. For example at depth 5, these would not necessarily be the highest 8 Lair values.