EX 4- Statistical Analysis -Alyssa Varner

Dataset of Choice: Airplane Strike Data

Analysis 1: Correlation Matrix

1. Imported dataset, troubleshot some errors in the dataset (mostly polynomials appearing in an "integer" assigned column).
2. Selected the attributes of choice. Chose two numerical and two non-numerical attributes.
3. Set up the "correlation matrix" operator. Connected the "correlation matrix" output to the "results" nodule.
4. The workflow ran without an issue, and I observed the resultant matrix. Only two attributes had a numerical value, with the others showing up as a "?" in the matrix. After some research I realized that was due to the type of attributes I had chosen: non-numerical attributes cannot be assessed statistically, which meant I had only done one true correlation. Which makes sense in hindsight.
5. To get a more complete correlation matrix, I modified my attribute selection to include two more "integer" based attributes, and reran the process.
6. The new results were far more complete and interesting. Speed, altitude, and distance from airport all appear to positively correlate with higher overall costs to the airline.

Analysis 2: ANOVA (group mean comparison)

1. "reset" the design view by removing the correlation matrix operator.
2. Changed the selected variables to include the categorical variable of "damage indicated".
3. Added the "Grouped ANOVA" function to the design view.
4. Decided to compare the "damage indicated" variable to the "total cost" variable.
5. Running the process was a success- no errors were encountered.

Analysis 3: ANOVA Matrix

1. "reset" the design view by removing the ANOVA comparison function.
2. Added the "ANOVA Matrix" function to the design view.
3. Generated a matrix.
4. Played with changing variables to see what results I would get (mostly to make sure it was working correctly).
5. Success!

Overall thoughts:

This exercise went quite smoothly. I am unsure if that is due to myself becoming more familiar with the program, or if it was due to the exercise itself. Either way, how fascinating! I can see how these functions could prove quite powerful with the right dataset.