Purpose

Superstore sales goes up and down but you usually don't know how sales item is related to each other. If we could know, we had have made more efficient sales strategy. For example, if you know tables sales is more related to furnishings sales, you would like to take an extra care about sales trend between them. Purpose of this project is to create sales data relational metric and visualize the relationship. Data driven sales strategy dashboard will be presented using a superstore sales data.

Relational Metric

All superstore data was obtained from https://www.kaggle.com/shobanama/superstore. Typical relational metric is to see the correlation of the sales volume. My specific interest is high level overview of sales. This is appropriate in monthly sales (quantity) by subcategory. Their correlations is our relational metric to see which subcategory is correlated to the others. The correlation matrix is shown below. Yellow highlighted entries are highly correlated by subcategory in terms of sales quantity. We focus on only correlated categories for our purpose.

Table 1

	Appliances	Art	Bookcases	Chairs	Copiers	Envelopes	Fasteners	Furnishings	Labels	Machines	Phones	Storage	Supplies	Tables
Appliances	1													
Art	0.756898	1												
Bookcases	0.443708	0.559544	1											
Chairs	0.702983	0.703698	0.570674	1										
Copiers	0.157113	0.220601	0.18125	0.07925	1									
Envelopes	0.441499	0.513292	0.515749	0.622797	−0.05563	1								
Fasteners	0.551608	0.519923	0.399766	0.588588	0.048852	0.379287	1							
Furnishings	0.702948	0.784488	0.594206	0.820037	0.193806	0.593478	0.630774	1						
Labels	0.594662	0.581148	0.44327	0.572674	0.173361	0.515973	0.361642	0.605371	1					
Machines	0.286328	0.2565	0.131423	0.244362	−0.27466	0.22393	0.298886	0.303378	0.335243	1				
Phones	0.713136	0.761065	0.469622	0.635	0.275201	0.6282	0.440179	0.733783	0.592123	0.133571	1			
Storage	0.709679	0.786445	0.532684	0.724828	−0.00082	0.621509	0.574822	0.784817	0.594611	0.254816	0.795486	1		
Supplies	0.456163	0.440343	0.455405	0.46802	−0.12063	0.390007	0.490083	0.537333	0.427465	−0.04649	0.515127	0.543944	1	
Tables	0.638418	0.677167	0.322099	0.594584	−0.01715	0.451359	0.467134	0.633796	0.481314	0.46105	0.710869	0.635959	0.24522	1

This correlation matrix can be expressed as a network diagram as below. The average degree is 3.714. The average path length is 1.476. This shows each subcategory has correlation with other 3 subcategories on average and it can reach out to other subcategory within 2 steps. Chart 1 below shows subcategory grouping in network by colour. Directed graph line is determined by cross correlation function (CCF) which is correlated with lagged value rather than the current one. For example, chair subcategory is correlated with furnishing but actually it is correlated with 2 period lagged value of chairs sales. This means sales of chairs subcategory is 2 period lead to furnishing sales. This indicates chair subcategory is correlated to furnishing but it is directed to furnishing sales. CCF is in fact calculated by auto correlated filter of loess decomposition - deseasonalizing and detrending but I do not go through all the details since it’s out of scope of this course.

Network Analysis

Chart1

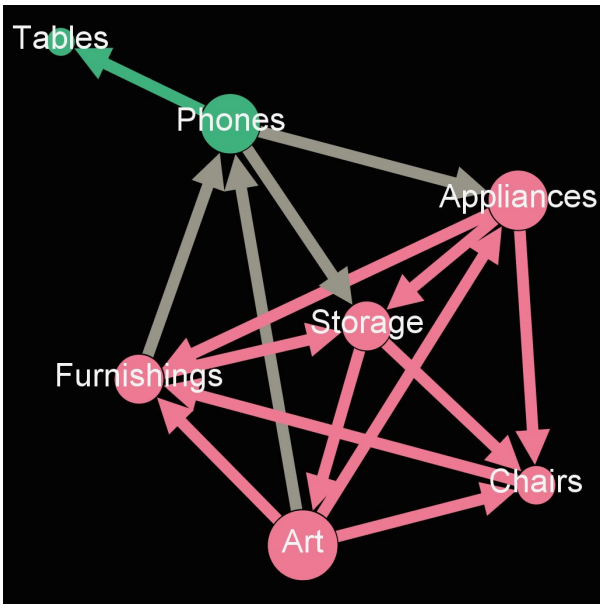
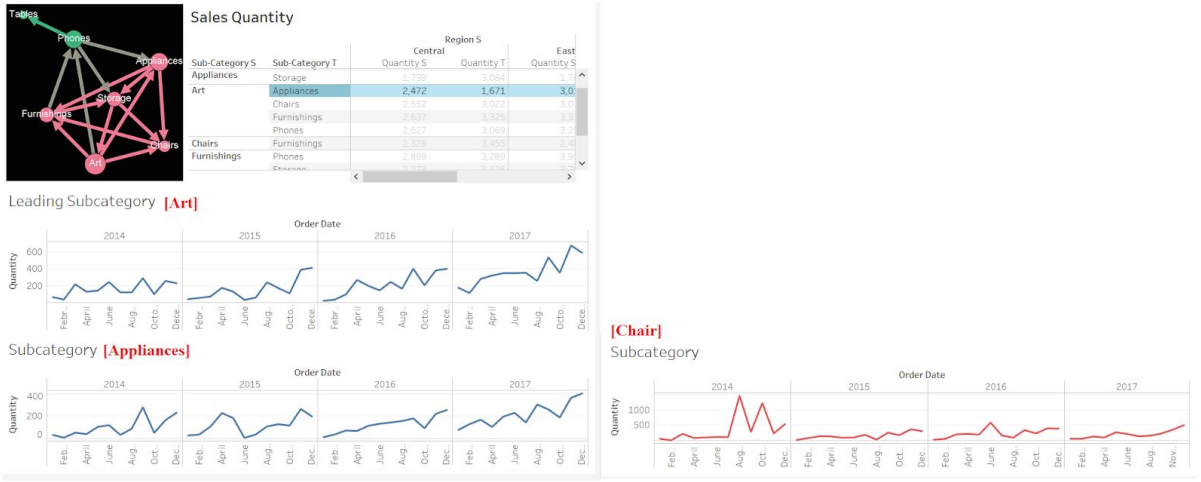


Chart1 show the most important subcategory in network that is red subgroup in which each subcategory is closely correlated. Modularity metric classifies the 2 groups. Green subgroup is a relatively isolated subgroup but “Phone” is an important subcategory in that its loss causes disconnected (or separate) network.

Chart 1 shows 2 groups (red and green). Green one is small auxiliary group. Pink one is composed of main group in network. Among the group, main players “Art” have the most connections (in terms of out-degree) to others. For sales strategy, focusing on the subcategory is expected to make the biggest ripple effect to boost entire sales since it has the most direct correlations to the other categories. Phone is a key player to keep network connected. Since network disconnection means ripple effect to be limited to one of disconnected subgroups. In this example, potential disconnected subgroup is “Tables” subcategory only. To keep “Tables” subcategory in the network (i.e if you want to include “Table” as sales strategy according to corporate management request while your sales effort also needs to see the entire sales), “Phone” subcategory is an important key player in the sales strategy.

Are subcategory sales correlated?

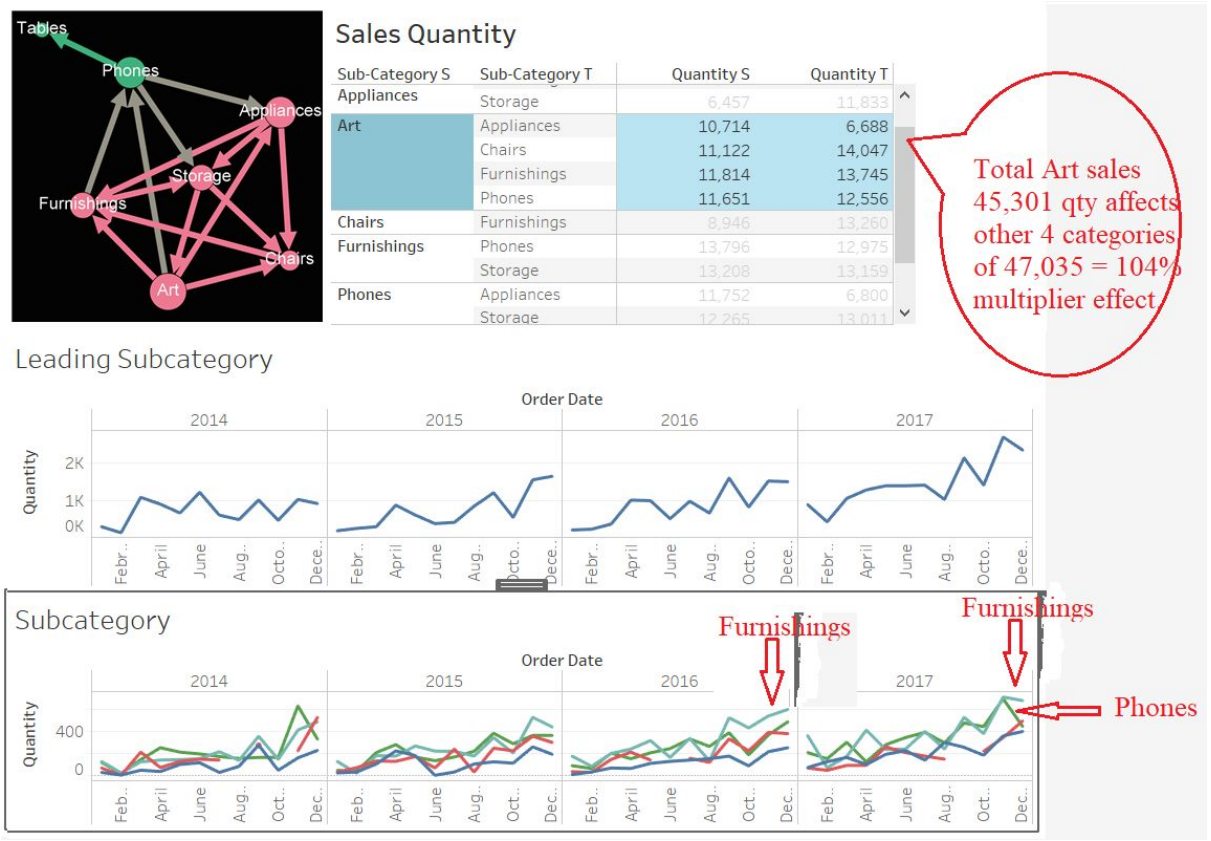
To see if they are correlated, for example mentioned above, “Art” and “Appliance” looks more correlated than “Chair”.



Which subcategory make biggest impact on entire sales?

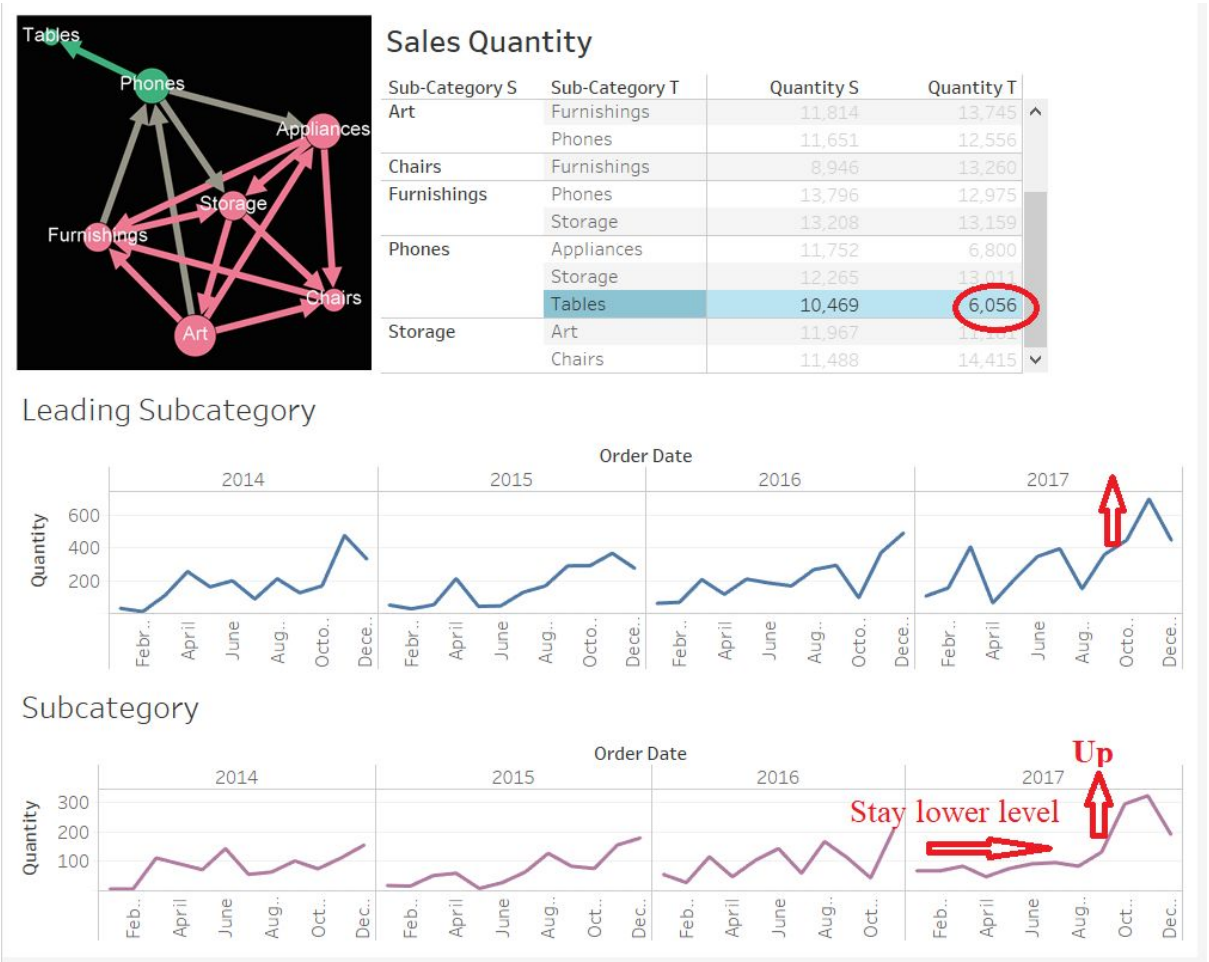
“Art” is considered as the most important subcategory. First it has 4 out-degrees. It means “Art” sales makes an impact on other 4 subcategories - “Chairs”, “Appliances”, “Phones” and “Furnishings”. Focusing on “Art” sales boost is expected to make the biggest ripple effect to thee 4 subcategories. Total “Art” sales in the last 4 years is 45,301 quantity and 47,035 for the other 4 subcategories. It’s roughly extra 4% multiplier effect from “Art” sales. “Phones” especially is an important key player to connect “Tables” sales to impact on “Tables”. Moreover “Art” can be connected to other categories within 2 steps v.s 3 steps required for some categories by “Appliances”. All together, focusing on “Art” sales strategy is the most efficient or beneficial way in considering entire sales strategy.

Recent data (2017) shows “Phones” and “Furnishings” sales is bigger than the others. Sales team has to pay special attention to these relationships because they are the biggest contributor to sales compared to others.



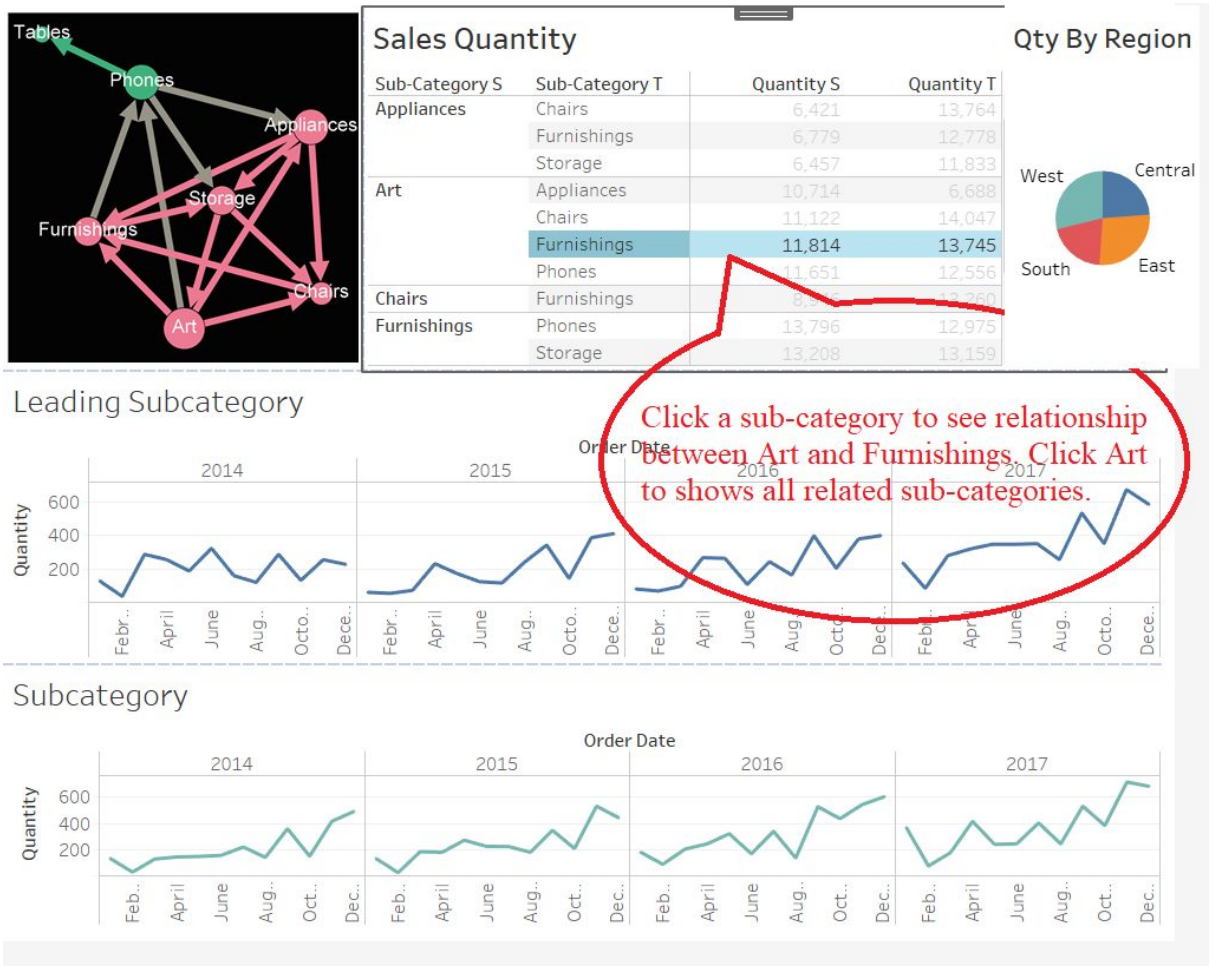
How does the loss of “Phones” make an impact on sales?

“Phones” is the only one subcategory correlated to “Tables”, the loss of which affect “Tables” sales. The last 4 years sums up 6,056 quantity sales and latest sales in Dec, 2017 is 192. This is small but current year data shows strong upward trend for both “Tables” and “Phones”. Especially Aug, 2017 - Dec, 2017 exhibits strong correlation between them. We note that “Tables” sales has been in lower level before Aug, 2017 but it goes up along with “Phones” sales up after that point of month. This is an eminent phenomena in 2017. Analytic team needs to investigate further why this happens. If “Tables” is strategic category from sales team request, we cannot ignore “Phones” sales trend.



User Manual

Sales Quantity table besides network graph image is subcategory relational list to map the network graph. This table is clickable to filter subcategory relation. All corresponding quantity data and its historical line chart shall be displayed based on your selection. Try to click some category and area. “Sub-Category S” is leading subcategory which will affect other category(s). For example below, “Art” leading subcategory is supposed to affect other 4 categories - “Appliances”, “Chairs”, “Furnishings” and “Phones”, which is corresponding to network graph image on the side. Qey by Region shows total sales quantity of leading and its related subcategory(s).



Click leading Sub-Category as below case “Art” to see all related 4 subcategories’ historical data. In some cases, outlier point hiders from viewing clear trend line. You can remove the point simply by clicking the outlier point and “Exclude”.