Preparing the Data

*B. Prepare the data to be used for describing past performance.* ***List here and add to the “prepared data” tab in the spreadsheet.***

*Fix or remove the following data errors.*

1. *Remove clearly duplicated data (take care not to delete a duplicate item!).*

**Fix Data Errors:**

* + Address any inaccuracies or inconsistencies in the data.

**Remove Duplicated Data:**

* + Identify and eliminate duplicate entries while ensuring that valid duplicate items are retained.

**List Prepared Data:**

* + Document the prepared data here and add it to the "Prepared Data" tab in the spreadsheet for further analysis.

1. *Some items never made it into the auction. Look for items with no auction information e.g. 50 1/3 Pounders from McDonald's. Do not delete these items! Instead, add a “in auction” dummy variable where 0=not in auction, 1 = in auction*

***Identify Missing Auction Information:***

* 1. *Locate items without auction data, such as "50 1/3 Pounders from McDonald's."*

***Add "In Auction" Dummy Variable:***

* 1. *Introduce a binary variable where:*
     1. *0 indicates the item was not in the auction.*
     2. *1 indicates the item was included in the auction.*

1. *Look for items that were miscategorized and put them into the correct category. First sort by item description and look for the same items listed in different categories e.g. Coupon Book, Sea Life Park, 1 Maui Taco, 6 oz. Kona Coffee. Then sort by category and see if the item description reasonably matches the category. E.g. Photo Frame is in Kids Stuff. If not, look for similar items and see what category they are in Change the category to the most suitable category. (There are at least two miscategorized items)*

***Sort by Item Description:***

* 1. *Arrange items alphabetically by description.*
  2. *Identify instances of identical items listed in different categories, like "Coupon Book" or "Sea Life Park."*

***Sort by Category:***

* 1. *Organize items by category.*
  2. *Assess if item descriptions align with assigned categories.*

***Reassign Categories:***

* 1. *If mismatches are found, explore similar items to determine appropriate categories.*
  2. *Adjust categories to ensure accurate classification of items.*

1. *Look for any data that has incorrect units, coding, or description. Transform data as needed. E.g. Assagio's $51 gift card → Assagio's $50*

***Review Data for Errors:***

* 1. *Scrutinize the dataset for inconsistencies in units, coding, or descriptions.*

***Address Incorrect Data:***

* 1. *Correct errors such as incorrect units or coding discrepancies.*
  2. *Transform data to ensure accuracy, for example, changing "Assagio's $51 gift card" to "Assagio's $50."*

1. *For items with missing data, fill in the missing data for items that it was clear it was not recorded in error. E.g. 2 packs of Mountain Thunder Kona coffee 6 oz. each clearly had the same donor and initiated by person as the 3 packs of Mountain Thunder Kona coffee 6 oz. each (however, it’s not so clear that the min bid and increment were not omitted purposefully)*

***Identify Items with Missing Data:***

* 1. *Locate items where data is incomplete or omitted.*

***Evaluate Clarity of Missing Data:***

* 1. *Determine if missing data was unintentional or purposeful.*
  2. *Assess if missing data can be inferred from similar items or context.*

***Fill in Missing Data:***

* 1. *For items clearly not recorded in error, complete missing data as appropriate.*
  2. *Consider factors such as donors, initiated by persons, and similar items when filling in missing information.*

*Add needed missing variables.*

* Added Bids variable. For an item that sold, if we assume that the bidders started with the min bid and each additional bid followed the bid increment then the number of bids for that item would be 1 + (Sales - Min Bid)/Increment. However, some bids did not follow the increment (bid more or less than the increment). Some items did not start with the min bid. Therefore, this is only an approximation of the number of bids. To be consistent with the bidding policies, we assume that an item sold under the min bid had only 1 bid (not enough interest to get over the min bid). If the sales - min bid is not a multiple of increment, we will round up to account for the possibility that a bidder used a smaller increment (because the bid still shows an interest in the item, just not enough to follow the increment). Here’s the Excel formula uses to approximate bids:

*=if(Sales>0,1+CEILING.MATH(Max(0,(Sales-MinBid))/Increment),0)*

* Size (ordinal)

Use the item description and category to estimate the qualitative size of an item

1=very small, 2=small, 3=medium, 4=large, 5= very large

Examples: Earrings are very small, Kids Baskets are medium, Turtle Table is large, Dolphin Table is very large

* Solicited (dummy variable0

0 = unsolicited, 1 = solicited

If an item has no initiated by, it can be assumed it was an unsolicited donation. If the donor is an individual (even a staff member), it can be assumed it was an unsolicited donation. If it has a initiated by and the donor is a business, it was solicited.

Examples: Baci Bistro $25 gift certificate Solicited = 1, Two white flowered vases Solicited = 0, Starbucks gift basket Solicited = 0

*Indicate outliers on the spreadsheet. Highlight these and add a comment on them as to why they are outliers. Do not delete or try to “fix” these.*

1. The 1-week stay at Thailand
2. The Baci Bistro $25 certificate had 31 bids in total and sold for $70 way over its value
3. Unique cactus and turquoise walking stick is an outlier because it is the only auction item with increment value over $25. This is considerably higher than the increment value of most other items sold.
4. Half-hour Mermaids Hawaii Show would be an outlier because this is one of the only auction items valued in the $500 that had a minimum bid of $100. The minimum bids for any other items at $500 or more were usually $200.

# Data Analysis

*C. Create descriptive statistics (use summary measures, charts, tables, graphs as you find most appropriate/useful) to answer the following questions. Explain which problems listed previously these stats may be useful for understanding. It is recommended that you download the spreadsheet and use Excel rather than work on the shared Google sheet. You should use the prepared data tab. However, note that the data preparation for this tab may not have been completed by the time you use the data. In this case, either volunteer to do the data preparation needed or use the data as is and make a note that your stats will need to be updated later when the prepared data is available.*

**1.** **How many items were in the auction? By category? The largest majority of the items came from which categories?**

*Suggested stats:*

· *Make a table of the number of items per category*

· *Pareto chart of number of items in auction by category*

· *Summary table of total number of items, average number items per category, std dev of number items per category*

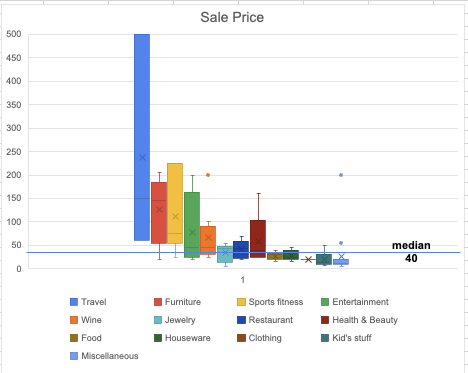
*!!! Add anything you believe is useful for describing number of items in auction*

**2.** **How much was raised by the auction? By category? Top performers of sales by category? Distribution of sales by category.**

*Suggested stats:*

· *Pareto chart of total sale price by category*

· *Box plots of sale price by category, sorted highest to lowest median*



This is helpful for understanding what we can expect in sale price for a given category and comparing categories on sale price relative to each other and sale price of all items in the auction. This information is useful for addressing the problem of accepting items for the auction (P3) because we can estimate what sale price to expect for an item. It is also useful for the problem of soliciting items for the auction (P4) because we can prioritize soliciting items from categories that bring in high sale prices.

· *Summary table of total sale price (amount raised), average sales price of item, std dev of sales price*

*!!! Add anything you believe is useful for describing sales price*

**3.** **What was the total value of items in the auction? Top 10% and bottom 10% value items? Distribution of value by category.**

*Suggested stats:*

· *Table of item values sorted highest to lowest with percentile*

· *Box plots of value by category, sorted highest to lowest median*

· *Summary table of total value, average value of item, std dev of value*

*!!! Add anything you believe is useful for describing value of items*

**4.** **Top 10% and bottom 10% bids items? Distribution of bids by category.**

*Suggested stats:*

· *Table of item bids sorted highest to lowest with percentile*

· *Box plots of bids by category, sorted highest to lowest median*

· *Summary table average item bids, std dev of bids*

*!!! Add anything you believe is useful for describing bidder interest*

**5.** **How many items were unsold? By category? What were the top and bottom performing categories in terms of sold? What was the total OL from unsold items? By category? Distribution of OL by category.**

*Suggested stats:*

· *Pareto chart of number of sold in auction by category*

· *Bar chart of % sold in category sorted highest to lowest*

· *Box plots of OL from unsold by category sorted highest to lowest median*

· *Summary table of total number items unsold, % number of items sold (from total sold), average % sold over all categories, std dev of % sold over all categories*

· *Summary table of total OL from unsold, average OL from unsold over all categories, std dev of OL from unsold over all categories*

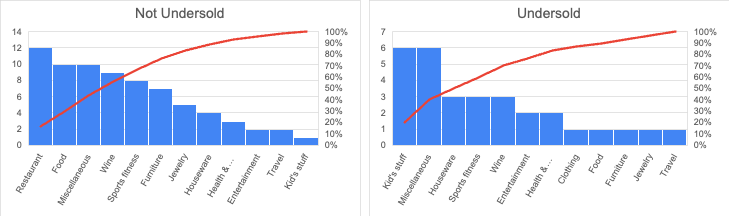
*!!! Add anything you believe is useful for describing unsold items in auction*

**6.** **How many items were undersold? By category? What were the top and bottom performing categories in terms of number of items sold? What was the total OL from unsold items? By category? Distribution of OL by category.**

**Note: do not include unsold items**

*Suggested stats:*

· *Pareto chart of number of not undersold in auction by category*

**

These graps are helpful for addressing the problem of undersold items (P2). The chart shows that about 70% of the not undersold items came from Restaurant, Food, Misc, Wine, Sports fitness. This graph also shows that 60% of the undersold items came from Misc, Health & Beauty, Kid’s stuff, Houseware, Sport fitness. Keep in mind that there are different numbers of items in each category.

· *Bar chart of % not undersold in category sorted highest to lowest*

· *Box plots of OL from undersold by category sorted highest to lowest median*

· *Summary table of total number items undersold, % number of items undersold (from total sold), average % undersold over all categories, std dev of % undersold over all categories*

· *Summary table of total OL from undersold, average OL from undersold over all categories, std dev of OL from undersold over all categories*

*!!! Add anything you believe is useful for describing undersold items in auction*

**7.** **What is the distribution of size of the items? What is the typical size of an item per category? What was the number of unsold and OL from unsold by size? What was the number of undersold and OL from undersold by size?**

*Suggested stats:*

· *Pie chart of item sizes*

· *Table of modes of size by category*

· *Pareto of number unsold by size*

· *Pareto of OL unsold by size*

· *Pareto of number undersold by size*

· *Pareto of OL undersold by size*

· *Summary table of total OL unsold+undersold, average OL, std dev OL by size*

*!!! Add anything you believe is useful for describing sizes of items in auction*

**8.** **How percentage of items were solicited vs unsolicited? What categories were the most and least solicited? What is the sales % for solicited vs. unsolicited items? What is the distribution of OL for solicited vs. unsolicited?**

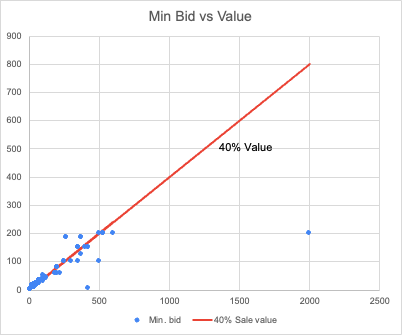
***Suggested Stats:***

1. ***Pareto Chart of Solicited Items by Category:***
   * *Visualize the distribution of solicited items across different categories.*
2. ***Box Plots of OL for Solicited vs. Unsolicited Items:***
   * *Compare the distribution of OL between solicited and unsolicited items to understand their performance.*
3. ***Box Plots of OL for Solicited and Unsolicited by Category:***
   * *Examine how the OL varies between solicited and unsolicited items within each category.*
4. ***Summary Table of Solicitation Statistics:***
   * *Provide summary statistics including the percentage of solicited and unsolicited items, total OL for solicited and unsolicited items, and any other relevant metrics.*

**9.** **Did the auction perform as expected in terms of sales vs value (higher value items sell for more)? Did min bid perform as expected (frequency of sales/value for sold items over 40% for given min bid/value, frequency of sold for given min bid/value)? Did increment perform as expected (frequency of sales/value for sold items over 40% for given increment/min bid, number of bids for min bid/value)? Does it appear that min bid was set based on value? Does it appear that the increment was set based on min bid? Did category matter in setting the min bid?***Suggested stats:*

· *Scatter plot of sale price vs value with linear trend*

· *Scatter plot of min bid vs value*

**

*We see a clear association between min bid and value providing evidence that the mon bid was predominantly set based on item value. For low-value items, it is always nearly 40% of the value. For higher-value items this varies notably but is still typically around 40%.*

· *Scatter plot of sale price vs value by categories with linear trend for each category*

· *Scatter plot of increment vs min bid*

· *Histogram of not undersold items for ranges of min bid/value*

· *Histogram of number of sold items for ranges of min bid/value*

· *Histogram of not undersold items for ranges of increment/min bid*

· *Histogram of bids for ranges of increment/min bid*

*\* Scatterplots may be more useful than histograms for the above*

*!!! Add anything you believe is useful for describing expected relationships*

**10.** **How many items were missing min bid? Missing increment? What was the total OL from missing min bids? What was the total OL from missing increments? What categories of items had the most missing min bids or increments? Did it matter if the item was solicited or unsolicited?**

*Suggested stats:*

· *Table of total number of items missing min bid, total number of items missing increment, total OL form missing min bid, total OL form missing increment*

· *Same as above by category*

· *Same as above by solicited and unsolicited*

· *Pareto chart of number of missing min bids by category*

· *Pareto chart of number of missing increments by category*

*!!! Add anything you believe is useful for describing missing min bids and increments*

**11.**  **How many items was the min bid not enforced? OL for this? How many items was the increment not enforced? OL for this?**

*Suggested stats:*

· *Table of total number of items min bid not enforced, total number of items increment not enforced, total OL form not enforcing min bid, total OL form not enforcing increment*

· *Same table as above by category*

· *Same as above by solicited and unsolicited*

· *Pareto chart of number of unenforced min bids by category*

· *Pareto chart of the number of unenforced increments by category*

*!!! Add anything you believe is useful for describing enforcement of min bids and increments*

**12.**  **[open]**

*Add useful questions and descriptive stats to address them not covered by any of the above.*