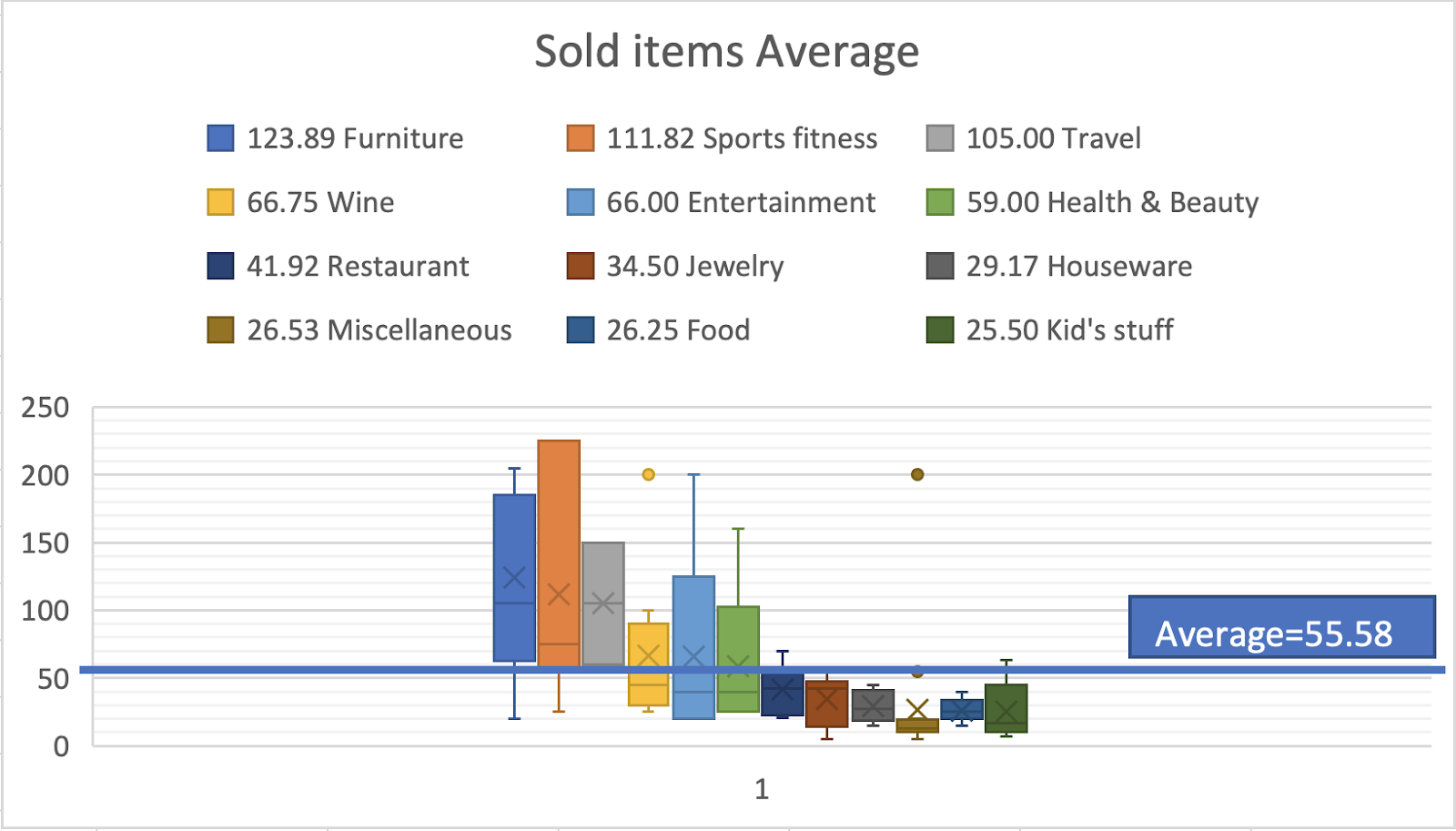
Predictive and Prescriptive Analytics

We present 8 models that may be useful predicting future auction expectations or prescribing optimal auction decisions. The models are described in green text.

M1: Predictive model for confidence that a category will have item average sale price greater than the average item sale price for all items.

For M1, we look at average sale price of sold items to figure out which categories’ items that attendees are interested in. We did not look at unsold items’ average price as unsold items’ sale price are 0.

This chart presents the distribution of sales price of sold items and the averages for each category.

By looking at the box plot above we expect:

How confident can we be that the average sale price for this category will (for a future auction) be greater than the auction average item sale price?

High confidence: Furniture, Sports fitness, Travel

Low confidence: Restaurant, Jewelry, Houseware, Miscellaneous, Food, Kid’s stuff

To determine the confidence, we test the hypothesis:

mu\_x = population mean sales for category x in an auction

mu\_all = population mean sales price for all items in an auction which we estimate as xbar\_all = $55.58

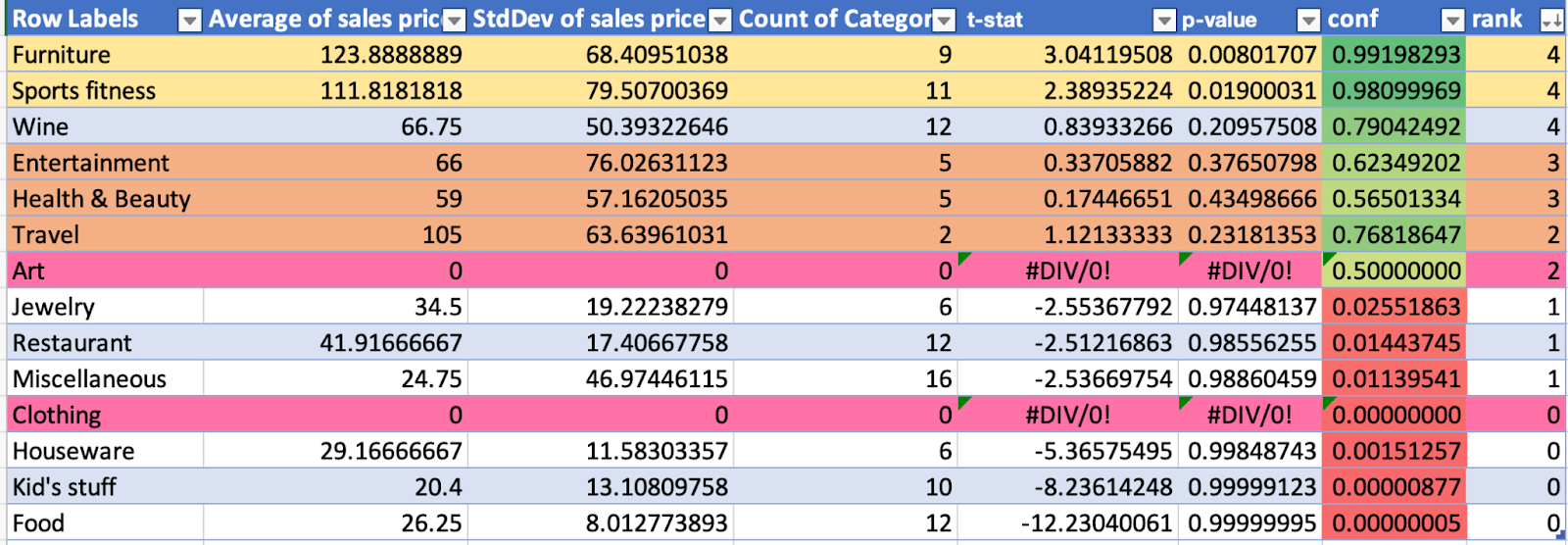
Ho: mu\_x <= xbar\_all

Ha: mu\_x > xbar\_all

Assumptions made:

Sales prices for each category are normally distributed

xbar\_all has very low variability from auction to auction (so xbar\_all is a good estimate for mu\_all)

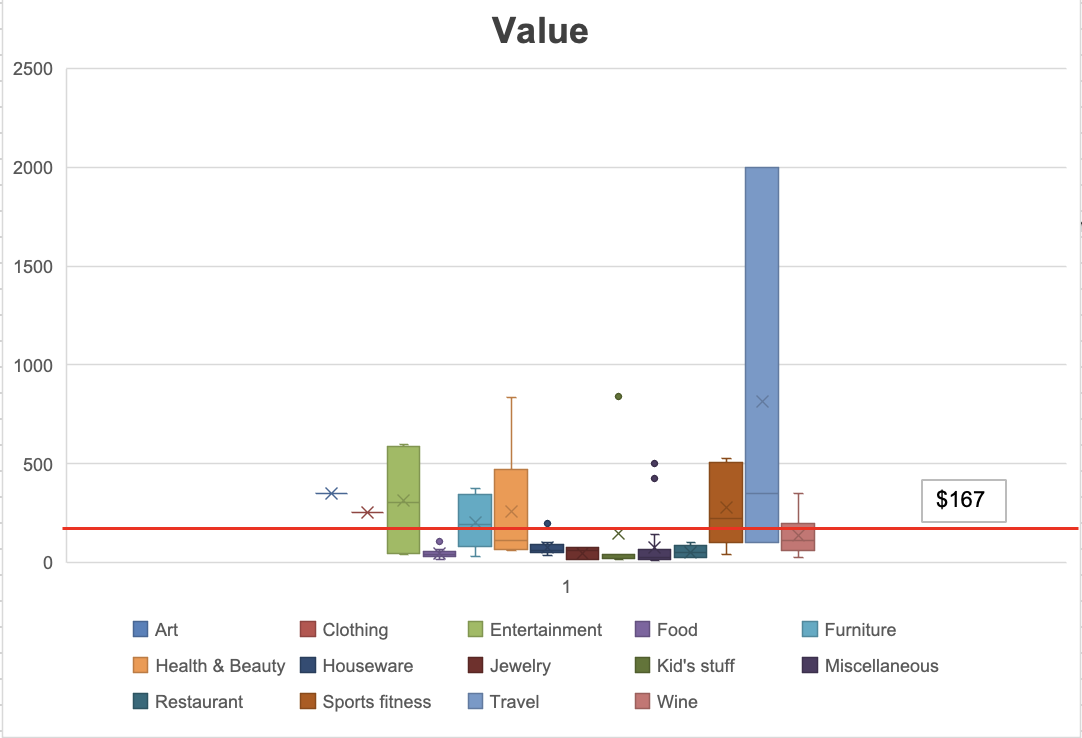
****

\*Highlighted in orange: caution, data may not be normally distributed or outliers

\*Highlighted in pink: not enough data, rank and confidence is guessed

How we define and rank for Art and Clothing categories:

Since Clothing and Art categories do not have enough data (nothing in these categories sold), we use our judgment of the sale price if something sold as follows:



Art: Items sale price tends to be relative to their value. Entertainment items have somewhat similar values (average $312) to Art (average $350) and we will assume similar variability. Since we are about 60% that Entertainment items will sell more than average, we estimate about 60% confidence that items in the art category would be sold with a sale price higher than average.

Clothing: We set the confidence of the clothing category (average value $127) at 55% because it has similar item values to Health & Beauty (average value $141 excluding $855 outlier ).

Based on the table above, we rank the confidence in sales performance more than average from 4 = highest confidence to 0 = lowest confidence.

What decisions might this model be useful for and how might it be used to inform the decision?[Dan]

**Decision: Priority of items to solicit for the auction**

This model allows us to figure out what category we are interested in based on the average sales price in each category. We are able to see whether a certain category brings in as many funds as possible. This will allow us to prioritize certain categories to make more money.

**Decision: Accept an item in a category with a given value or ask for a cash donation**

This model is useful for identifying and inferring whether an item will sell higher than the average sales price. Knowing the confidence level of the category can help us infer which items are likely to sell based on their ranks and their likelihood of selling.

What performance problem(s) from the performance reports is this model intended to contribute to addressing? [Dan]

Based on our ranks, we see that some categories tend to sell more than other categories. However, there was a lack of understanding in which to prioritize to list in this auction.

For example, none of the 3 items in Art and 2 items in Clothing sold. These categories have insufficient data to estimate sales likelihood for the next auction. We would have to access this information using our best guess which might be biased.

We see that Furniture and Wine should have priority as we have higher confidence that these items will be higher than the average sales price.

Controllable variables for the auction: What items to accept/reject from the auction  
Uncontrollable: Item sold for categories   
Effects: Funds raised

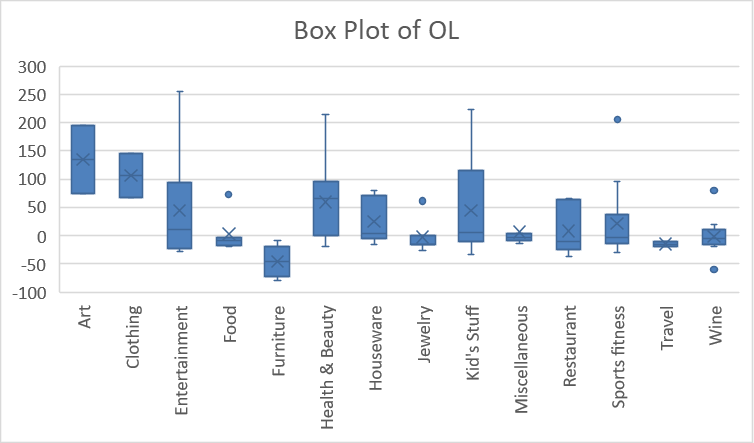
M2: Predictive model for confidence that a category will have average OL for unsold/undersold items less than 0.

We want to use data from all items because we want to also see the opportunity lost for the items that did not sell and could have been sold. Seeing the opportunity lost of the items that did not sell might give us more information on our items selection for our next auction.

Calculate the average price for sold items, Use the median (35) because we have outliers.

Compute the OL for sold/undersold items. Use the formula =0.4\*[@Value]-[@[sales price]]+IF([@[sales price]]=0,average\_item\_sold\_sale\_price,0). Keep in mind that OL < 0 means no loss from unsold/undersold.

Create box plots of sales price by categories, order from lowest to highest average OL. Add line at 0 OL. Investigate any potential outliers and categories that may not be reasonable to assume are Normally distributed. Assess which categories generally have items that have OL less than 0.



Anything below 0 is good and we add a line around 0 because that helps us differentiate between good and bad when it comes to opportunity loss.

The further up from 0 the bigger the opportunity lost. The Items that clearly have OL >0 are: Art, Clothing, health and beauty.

The Items that had OL<0 means it is good for our auctions because we do not want to lose money on items that we could have cash donation instead. Those Items come from different categories: Furniture, Food, travel, jewelry.

We noticed a few outliers: Sports and fitness We think it is because of the item -”One-month membership, two lessons, dinner for two KRC” That had a value of 375 and did not sell (0).

For food: The outlier we investigated was the Starbucks gift bag that had a value of 60$ and didn't sell. However, we noticed that the other Starbucks gift basket did sell for 40$. We were thinking maybe they had different items in the other starbucks gift basket that did not attract the audience as much.

Create a pivot table with the following for each category: average OL, std dev of OL, count of items.

Copy the table and paste as values then convert to a data table. Add columns:

t-stat for OL > 0. Use the formula

=([@[Average of sales price]]-average\_item\_sale\_price)/([@[StdDev of sales price]]/SQRT([@[Count of Category]]))

confidence level: Use the formula

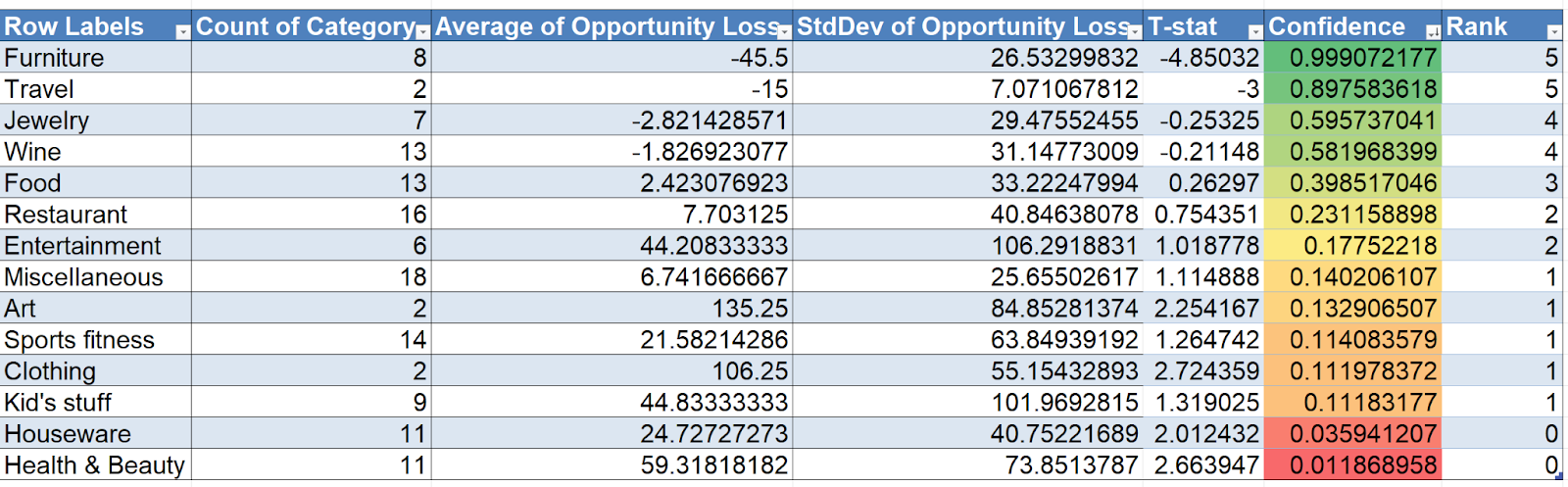
= T.DIST.RT([@[t-stat]],[@[Count of Category]]-1)

(see <https://www.tutorialspoint.com/statistics/student_t_test.htm> for details on the t-test)

Sort the table from highest to lowest confidence. Add conditional formatting to the confidence column using a 3-Color Scale where the Minimum is red, Midpoint is yellow, Maximum is green.

Add another column, Rank, and assign a rank according to the color of the confidence starting with 0 for red and the highest rank for green.

The resulting table should look something like this:



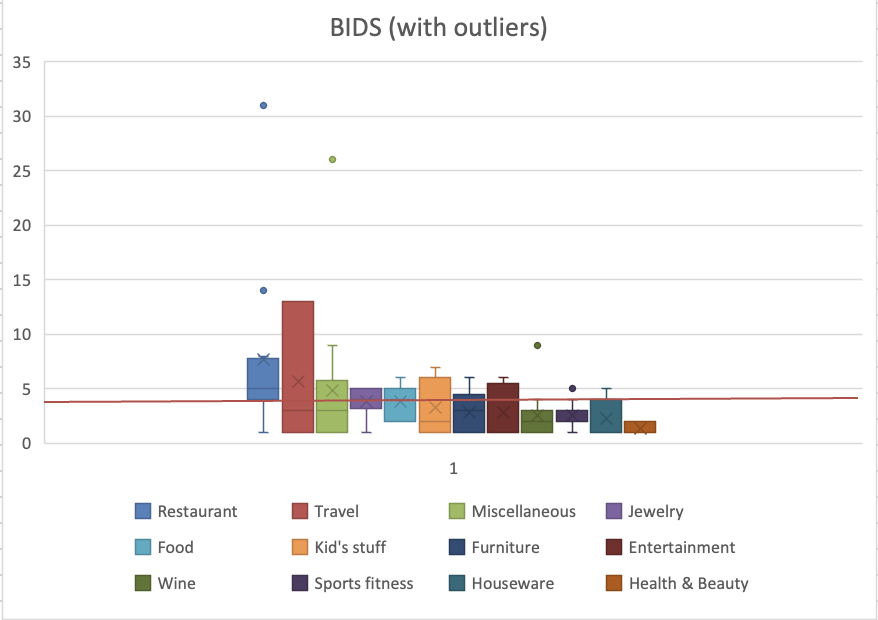
State the hypothesis(s) being tested. Explain any important assumptions that are being made for this model. Do not list technical statistical assumptions such as the data is Normally distributed. [ ]

Explain what decision(s) from the decisions listed earlier this model is useful for. It doesn’t have to address the entire decision. It can contribute information or help answer a necessary question involved in the decision. Explain how it is used to inform the decision(s). [ ]

Explain using the descriptive models from previous parts what performance problem(s) this model is intended to contribute to addressing. Also note any important considerations from the performance results that must be noted in using the predictive model. For example “We noted that 25% of the items went unsold. Our goal is 100% and this model will help prioritize selection of items that are likely to sell. Some categories of items had more unsold items than others. None of the 3 items of Art sold and only 1 of the 3 Clothing items sold. These categories have insufficient data to estimate sales likelihood for the next auction and will have to be assessed using our best guess which may biased. [ ]

M3: Predictive model for confidence that a category will have average number of bids per item greater than the average number of bids per item for all items

We use the corrected data with filled-in missing data from only sold items with outliers removed. We would want to use data from items that were sold because items that were unsold don’t have any bids (zero). It would be irrelevant to include that data for this model because we shouldn’t account for items with zero bids since it was unsold and no one bid on it. The data would be unrepresentative and could skew the data, leaving our model to be unreliable.



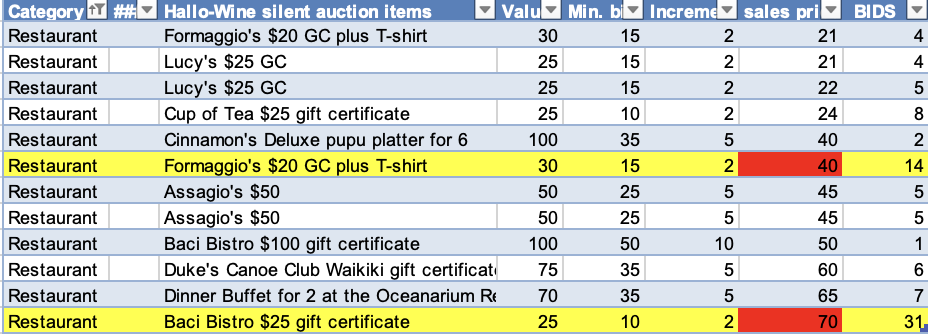
Looking at the box plot, we see that there are obvious points that are worth investigating. These points on the box and whisker plot above gave us reason to doubt the confidence in our model if it were to be included. For restaurant, miscellaneous, wine, and sports fitness, these categories had obvious outliers with points far from their box and whisker plot. However, for miscellaneous, we found that because this category is pretty random, it was best to leave the point in because it could help with decision making on what kinds of items to accept in the category for the auction. Also, when looking into sports fitness, there wasnʻt any indication of errors or unusual data, so we decided to keep it in our data set. For travel & kid stuff, their box plots have notability larger variability. For travel, it is probably due to the high value item of the 1-week stay at the Resort in Thailand. So, might want to leave travel in because it can give us a clue on how to make better decisions for that category.

To have our data representative and to make our model more reliable, we had the following outliers removed:

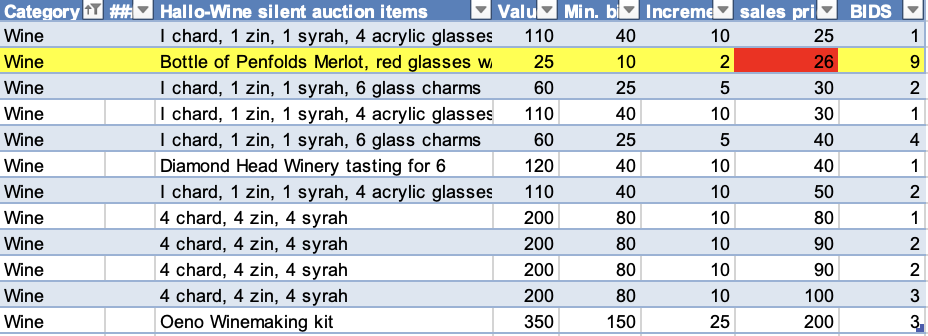
Restaurant - Baci Bistro $25 gift certificate

Restaurant - Formaggioʻs $20 GC plus T-Shirt

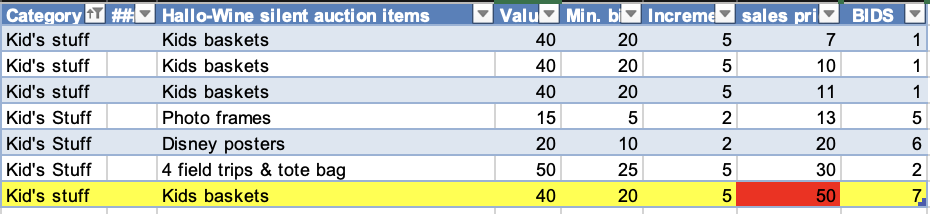
Point 14 & 31 obvious outliers and incorrect data



Wine - Bottle of Penfolds Merlot, red glasses w/charms

Point 9 obvious outlier and incorrect data

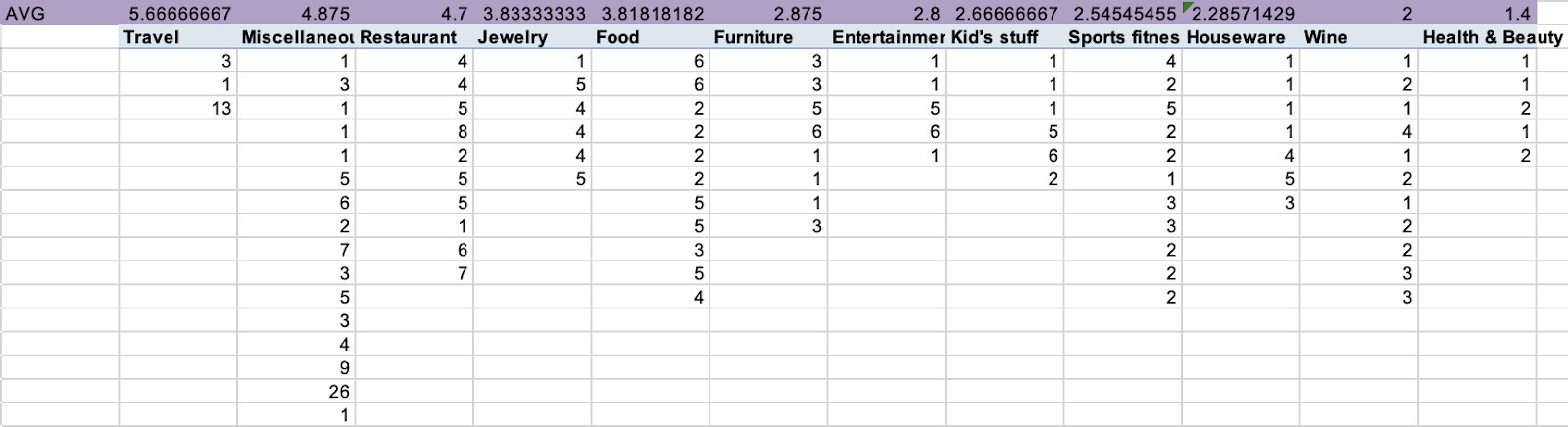
Kid Stuff - Kid Baskets

Point 7 is part of incorrect data

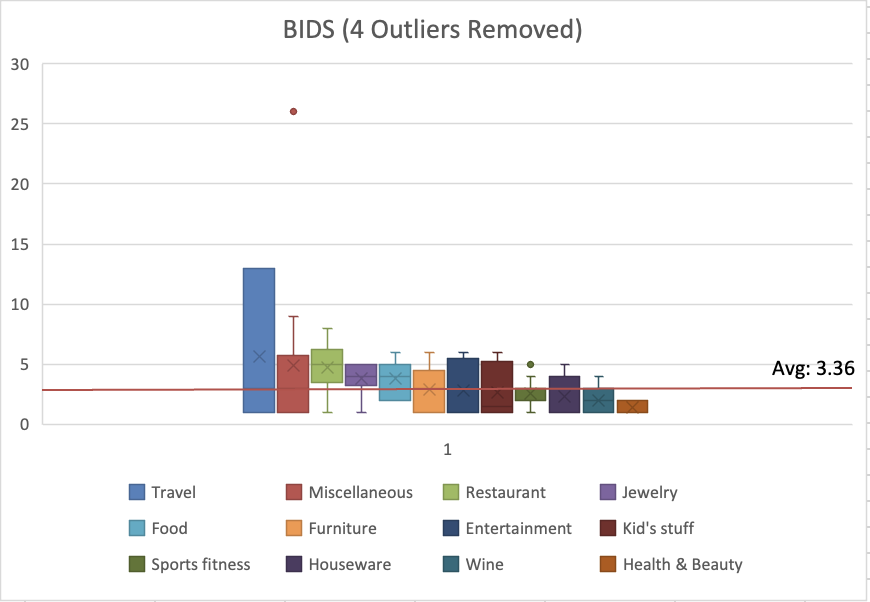
Upon investigating these outliers, we found that the sales price is larger than the value for these items, which affects the amount in BIDS. We can’t be too sure if the actual items were sold for more than their value or if it is an error. So, we removed these points from our data set to be representative for our model.

We compute the average item bid. This will be your estimate of the population mean item bids for all auctions.

Average bids of only sold data with outliers removed = 3.36 (average\_item\_bids)

Average bids for each category: 

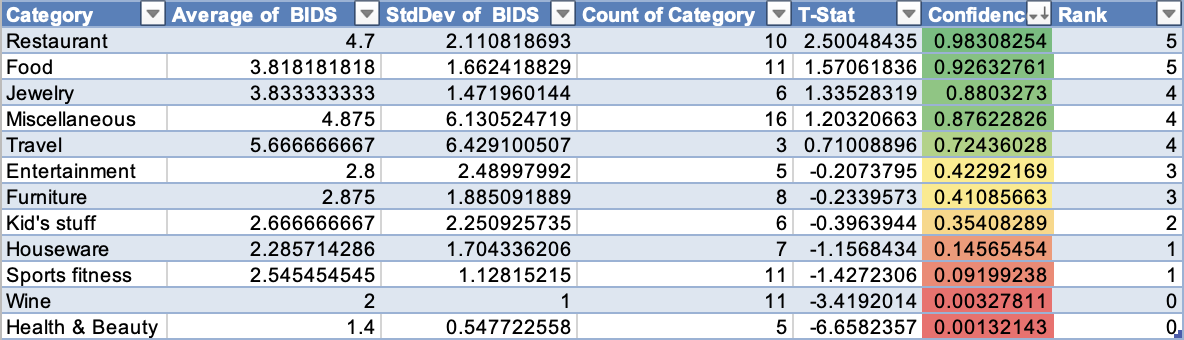
Now box plots of bids by categories, ordered from highest to lowest average with a line with the average item sale price. We investigate any potential outliers and categories that may not be reasonable to assume are Normally distributed. Assess which categories generally have items that sell more than average bids, about the average bids, and less than the average bids



As seen in the box plot above the Travel, Miscellaneous, Restaurant, and Jewelry category have bids above the average item bid. The Food, Furniture, Entertainment, and Kid Stuff category have bids near the average item bid. It could be assumed that people are more interested in Items in these categories since their bids are near or above average.

On the other hand, Sports fitness, Houseware, Wine, and Health & Beauty have less than average item bids. It could be assumed that people are less interested in items in these categories as there a less bids in these categories.

The predictive model is a table with the following for each category: average bid, std dev of bid, count of items T-Stat, confidence and rank. Explain the ranks and if the results are as expected from the box plots.



State the hypothesis(s) to tested. Explain any important **assumptions** that are being made for this model. Do not list technical statistical assumptions such as the data is Normally distributed. [ ]

Explain what decision(s) from the decisions listed earlier this model is useful for and interpret what the model says for these decisions.. It doesn’t have to address the entire decision. It can contribute information or help answer a necessary question involved in the decision. Explain how it is used to inform the decision(s). [ ]

Explain using the descriptive models in Step2 what performance problem(s) this model is intended to contribute to addressing. Also note any important considerations from the performance results that must be noted in using the predictive model. For example “We noted that 25% of the items went unsold. Our goal is 100% and this model will help prioritize selection of items that are likely to sell. Some categories of items had more unsold items than others. None of the 3 items of Art sold and only 1 of the 3 Clothing items sold. These categories have insufficient data to estimate sales likelihood for the next auction and will have to be assessed using our best guess which may be biased. [ ]

M4: Predictive model for **confidence that a category will have average sales/value greater than the average sales/value for all items**

We want to use the data from items that were sold because we are looking at the overall problem of undersold items (these are sold items that sold but at no more than 40% of its value) and how we can increase the average sales price. For items that undersold we could have just asked for a cash donation of 40% of the item’s value. It is important to only look at sold items because we can look at its sales value and calculate confidence intervals from there. There was no value gain from the unsold items and so, because we intend to develop a predictive model for future auctions, we should look at the sold items so we can consider the value gained from those items.

Estimate of the population mean item sales/value price for all auctions

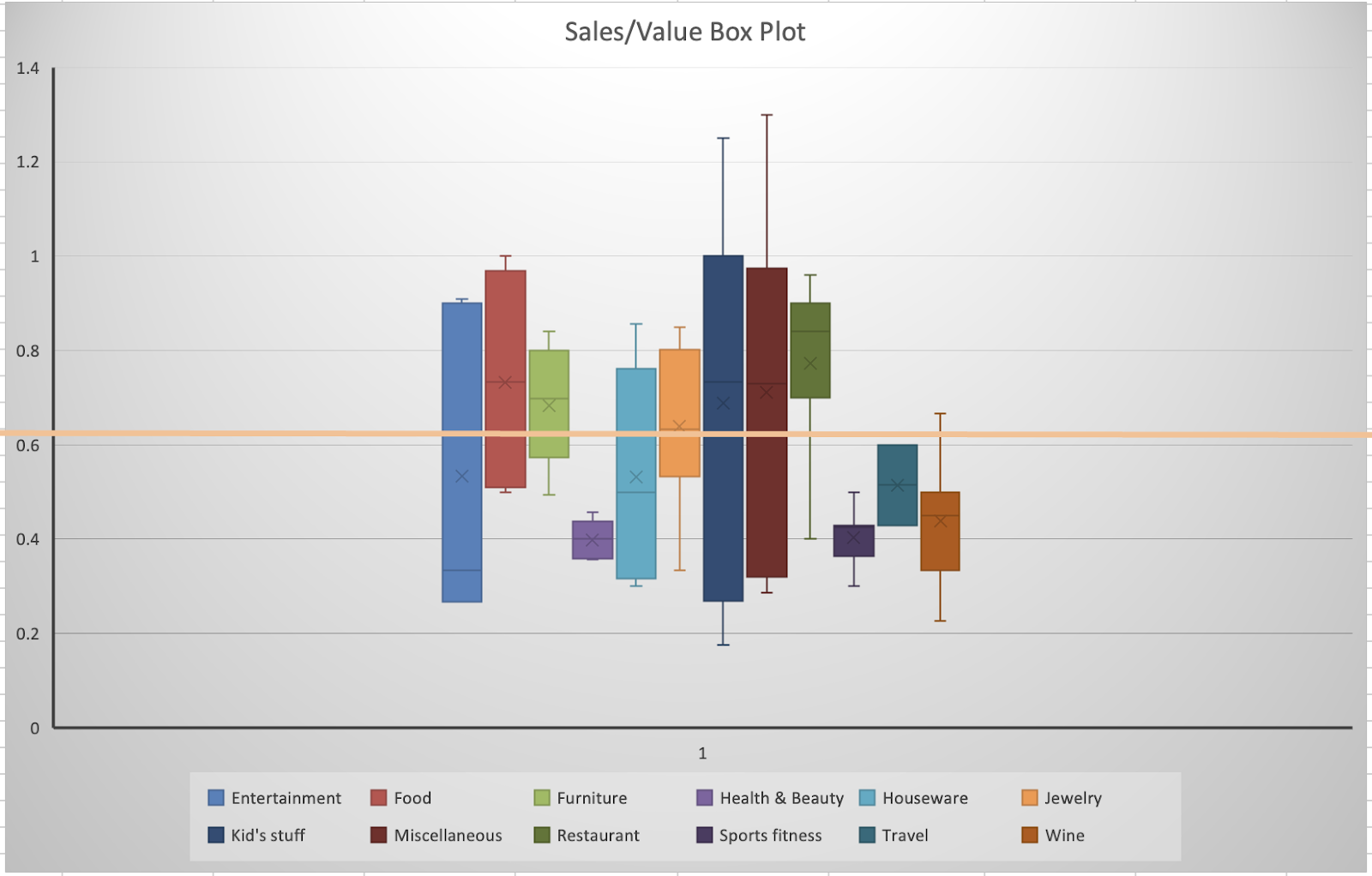
Our original population mean item sales/value price for all items was 0.629314847 which is good because it is greater than 0.40, the cash value that could have been asked as a donation instead of the items.

|  |
| --- |
| 0.629314847 |

However, after noting that there were several outliers outside of the box and whisker plot, we removed several outlier items from the calculations. The new population mean item sales/value price was 0.617291013. This is the number we are running with for all future auctions.

|  |
| --- |
| 0.617291013 |

Box plots of sales price by categories, order from highest to lowest average. Add line with average item sale price and explain results



Outliers We Deleted

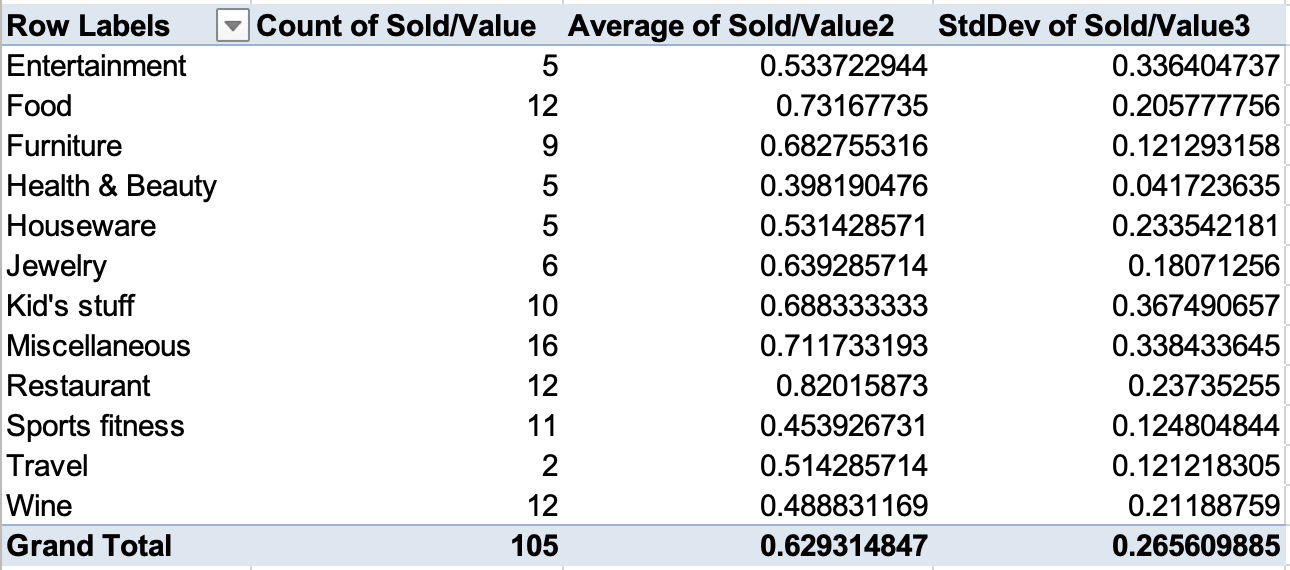
* Restaurants 1.333 (outlier in box and whisker) - Formaggio’s $20 GC plus T-shirt
* Sports fitness .73333 (outlier in box and whisker) - Fishing pole
* Sports fitness .625 (outlier in box and whisker) - Swimming pool boat
* Wine 1.04 (outlier in box and whisker) - Bottle of Penfolds Merlot, red glasses with charms
* Houseware .28 (outlier in box and whisker)

\*\*Thought about taking out the kids basket but we believe that it is important to keep this auction item in our data because yes this could be an error but we believe this was just a competitive auction item and there was a bidding war\*\*\*\*

We chose to delete these outliers because they were outside of the initial box and whisker plots and they do not represent the typical sales/value for the items. Removing these outliers likely made the data more representative of future YMCA HalloWine auctions.

Assessment of Categories

* All items above the orange line (the population mean item sales/value price) typically have a higher than average sales value (food, furniture, kids stuff, miscellaneous, and restaurants).
* About the same average sales value (jewelry)
* Below the average sales value (entertainment, health & beauty, houseware, sports fitness, travel and wine).



Hypotheses being tested and Assumptions:

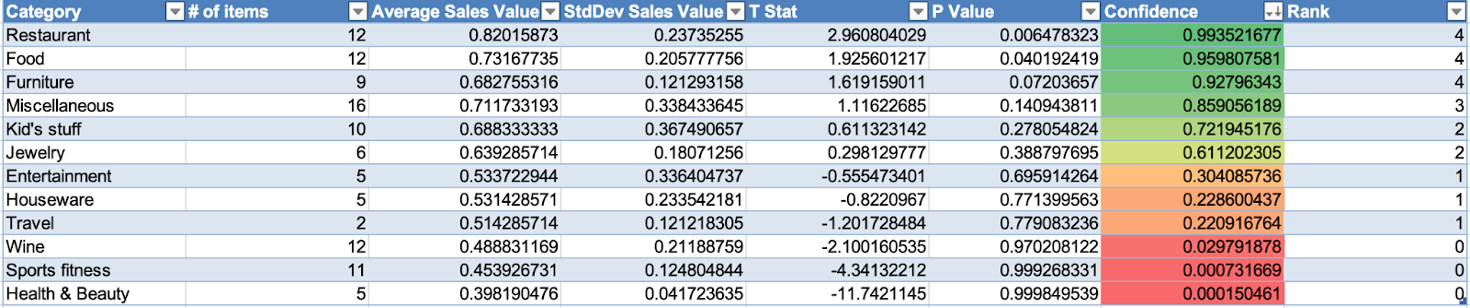
mu\_all = population mean sales/value for all items

mu = population mean sales/value for category x

Ho: mu <= mu\_all

Ha: mu > mu\_all

The predictive model is a table with the following for each category: average bid, std dev of bid, count of items T-Stat, confidence and rank. Explain the ranks and if the results are as expected from the box plots.



Explain what decision(s) from the decisions listed earlier this model is useful for. It doesn’t have to address the entire decision. It can contribute information or help answer a necessary question involved in the decision. Explain how it is used to inform the decision(s). [ ]

Explain using the descriptive models in previous parts what performance problem(s) this model is intended to contribute to addressing. Also note any important considerations from the performance results that must be noted in using the predictive model. [ ]

M5: Predictive model for confidence that a category will have proportion of sold items greater than proportion of sold for all items [Names here (no more than 4):Sydney, Stacy, Mitch, and Katelyn]

Compute p\_all, the proportion of items sold (# sold / # items) =76%

Pivot table showing count of items sold and count of items

|  |  |  |
| --- | --- | --- |
| **Row Labels** | **Count of sold** | **Count of Category** |
| Art | 0 | 2 |
| Clothing | 0 | 2 |
| Entertainment | 5 | 6 |
| Food | 11 | 13 |
| Furniture | 8 | 8 |
| Health & Beauty | 5 | 11 |
| Houseware | 7 | 11 |
| Jewelry | 6 | 7 |
| Kid's stuff | 7 | 9 |
| Miscellaneous | 15 | 18 |
| Restaurant | 11 | 16 |
| Sports fitness | 11 | 14 |
| Travel | 2 | 2 |
| Wine | 12 | 13 |
| **Grand Total** | **100** | **132** |

Copy the table and paste as values then convert to a data table. Add columns:

p-value for proportion items sold > p\_all. Use the formula

=1-BINOM.DIST([@[count if items sold]],[@[Count of Category]],p\_all,TRUE)

confidence level: Use the formula

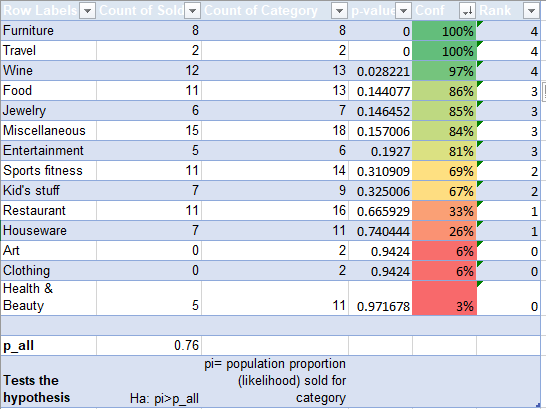
=1 - p-value

(see[https://www.tutorialspoint.com/statistics/one\_proportion\_z\_test.htm#:~:text=The%20test%20statistic%20is%20a,deviation%20of%20the%20 sampling%20 distribution](https://www.tutorialspoint.com/statistics/one_proportion_z_test.htm#:~:text=The%20test%20statistic%20is%20a,deviation%20of%20the%20sampling%20distribution) for details on the binomial test for proportion)

Sort the table from highest to lowest confidence. Add conditional formatting to the confidence column using a 3-Color Scale where the Minimum is red, Midpoint is yellow, Maximum is green.

Add another column, Rank, and assign a rank according to the color of the confidence starting with 0 for red and the highest rank for green.

Your resulting table should look something like this:



1. Hypotheses being tested and Assumptions: [ ]

Explain what decision(s) from the decisions listed earlier this model is useful for. It doesn’t have to address the entire decision. It can contribute information or help answer a necessary question involved in the decision. Explain how it is used to inform the decision(s). [ ]

Explain using the descriptive models in Step2 what performance problem(s) this model is intended to contribute to addressing. Also note any important considerations from the performance results that must be noted in using the predictive model. [ ]

**M6: Prescriptive model for the economically optimal number of items in each category to solicit and the total number of items in the auction to have.**

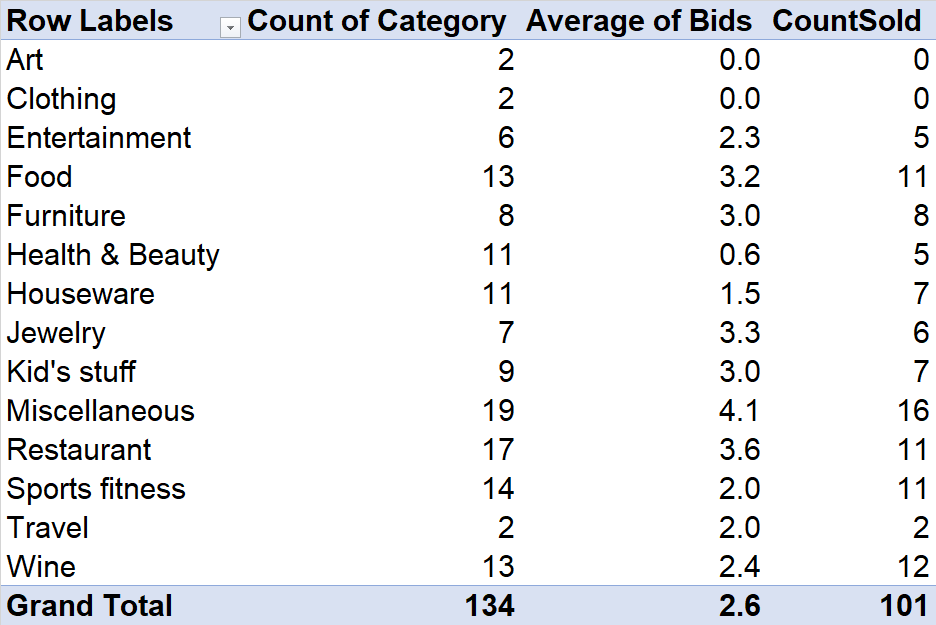
We are going to use the Newsvendor model. Here are some references:

https://medium.com/@pdemarle/the-basics-of-the-newsvendor-model-ef756f203433

https://smartsupplychains.ai/2019/08/09/a-gentle-introduction-to-newsvendor-model/

Use the corrected data with filled in missing data from **all** items with outliers removed.

Create a pivot table with the following for each category: count of items, # Sold, Average Bids



Copy the table and paste to another area, convert this to a data table, then add columns: 95% demand lower, 95% demand upper, Average Demand, Std Demand. ]

For the 95% Demand low use the formula =FLOOR.MATH([@sold]-T.INV.2T(1-0.95,[@N]-1)\*SQRT([@sold]\*(1-[@[proportion sold]])),1)

In the case where this value is negative set the value manually to 0.

For the 95% Demand high use the formula:

=CEILING.MATH([@sold]+T.INV.2T(1-0.95,[@N]-1)\*SQRT([@sold]\*(1-[@[proportion sold]])),1)

In the case where zero items sold (sold = 0), use the formula =3/[@N] to account for the possibility that the demand is not zero in future auctions

In the case where 100% of the items sold, use the formula =[@[Average of Bids]]+[@N]. The theory here is that these items were popular and the demand is higher than the number of items available. We will guess the demand was more by the average number of bids for these items assuming that if there were this many more items, bids would have been shifted to these items. This of course is a highly unfounded assumption!

For Average Demand use the formula:

=([@[95% Demand low]]+[@[95% Demand high]])/2

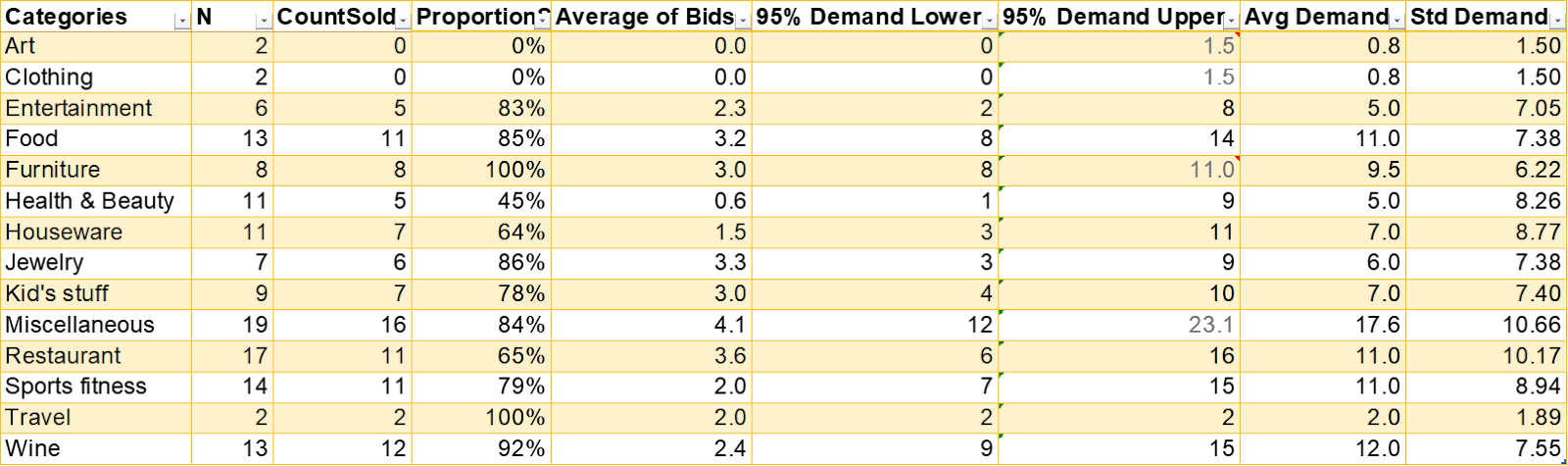
This will generally be the same as # sold.

For Std Demand use the formula:

=([@[95% Demand high]]-[@[95% Demand low]])\*SQRT([@N])/(2\*T.INV.2T(1-0.95,[@N]-1))

This accounts for the additional variability in estimating the standard deviation from a small number of data points.

The table looks something like this:



In cases where 0 items sold, we assumed that demand is not actually zero, but is less than the number of items available.

In cases where 100% of the items sold, we assume the demand is higher than the number available. We assumed that the demand is higher by the avg number of bids per item in the category.Demand was calculated as a function of count, count sold, and proportion sold.

Now create a pivot table with the following for each category: count of items, average value

Copy the table and paste to another area then go back to the pivot table and add a filter to view only **sold** items (sales price > 0). Use this to get the **count of items sold, average sales price of items sold**. Copy these new columns to your table. If there were no items sold, the average sale price will be 0 and you should make a guess at what this might be if there were items sold. A good starting point might be 40% of the average value.

Convert table to a data table. Add columns Average demand, Std demand. Copy the cell references for these from your other table above (don’t just copy the values just in case you adjust the other table)

Add a column Co to the table. For each category, compute the average loss from having one item that did not sell (overage cost Co). This should be 40% of the average value for that category.

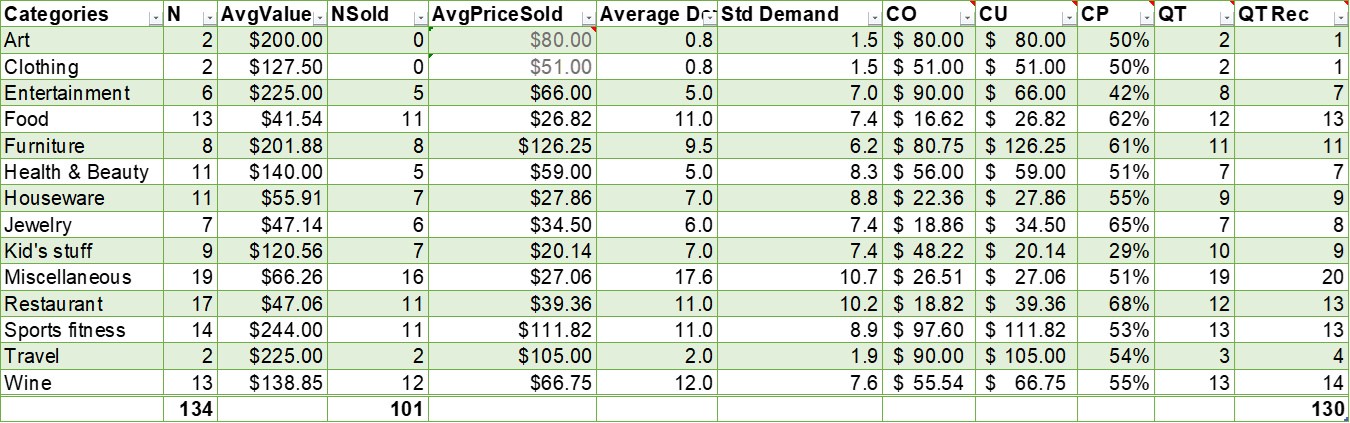
Add a column Cu to the table. For each category, compute the average loss from having one item that could have sold but was not available for sale (underage cost Cu). This should be the average sale price of sold items in the category. If there are no items sold in the category (like Art), you can make your best guess for this.

Add the following columns to the table

CP to the table using the formula Cu/(Co+Cu)

QT\* using the formula =average demand + std dev deman d \* T.INV.2T(CP,count of items - 1)/SQRT(count of items)

Recommended # using the formula =IF(Co < Cu,CEILING.MATH(QT\*),FLOOR.MATH(QT\*))

When you have all the recommended computed, add the total. This will be the recommended total number of items to have in the auction.

CO = Cost of overage

CU = Cost of underage

CP = Critical Percentile, equals the profit maximizing point on the demand distribution

QT = profit maximizing quantity

QTRec= profit maximizing quantity adjusted for round-up and round-down rule.

Explain what the objective of this model is and any important assumptions that are being made for this model. Do not list technical statistical assumptions such as the data is Normally distributed. [ ]

Explain what decision(s) from the decisions listed earlier this model is useful for. It doesn’t have to address the entire decision. It can contribute information or help answer a necessary question involved in the decision. Explain how it is used to inform the decision(s). [ ]

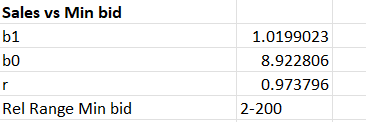
Explain using the descriptive models from previous parts what performance problem(s) this model is intended to contribute to addressing. Also note any important considerations from the performance results that must be noted in using the predictive model. [ ]

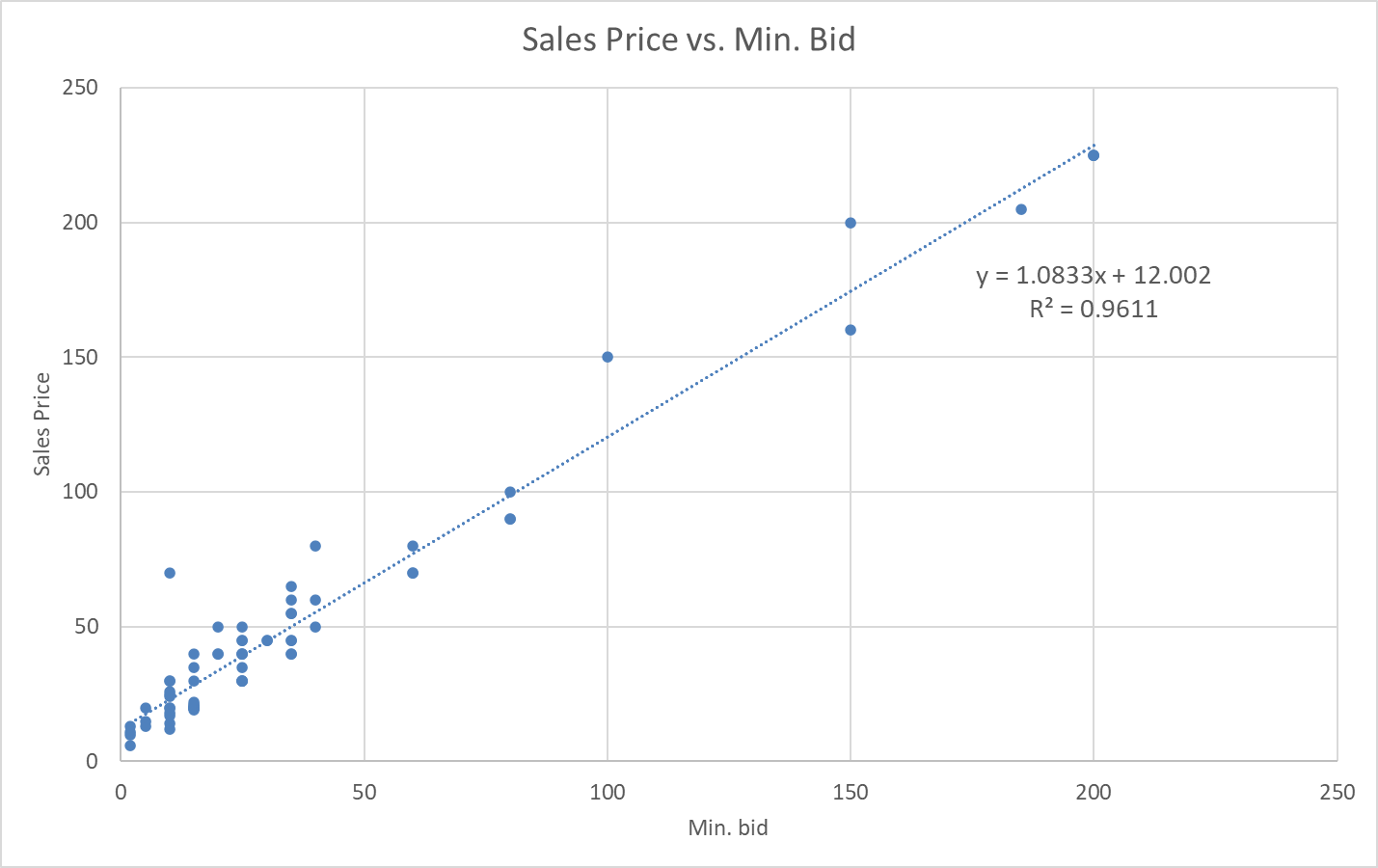
**M7: Prescriptive model for setting the optimal min bid and increment**

Assumptions:

* We imputed the min bid and increment for the following items that did not have min bid and increment based on other items with similar value so that we could have more data to use in our analysis:
  + Food
    - 2 packs of Mountain Thunder Kona coffee 6 oz. each
  + Kid's Stuff
    - Photo frames
    - Disney posters
    - 4 field trips & tote bag
  + Misc
    - Scrapbook ribbon thingy
* Removed items that we were not able to analyze or we considered outliers for our analysis. We considered outliers as items that had a large value to min bid ratio since they would be unrepresentative of items how to set the min bid in general for future items
  + Furniture
    - Carved wooden bowl - no item that is $125 or similar item to make a confident imputation
  + Miscellaneous
    - Real estate licensing class - the value ($425) to min bid ($5) is vastly different from other Misc items and items with higher values
  + Restaurant
    - 20 $5 gift cards from Maui Tacos - no data
    - 50 1/3 Pounders from McDonald's - no data
    - 50 Small hot or medium iced McCafe - no data
    - 50 Local breakfast platters - no data
  + Travel
    - 1-week stay at Resort in Thailand (air fare not included) - the value ($2000) to min bid ($200) is vastly different from other items that had a min bid of $200

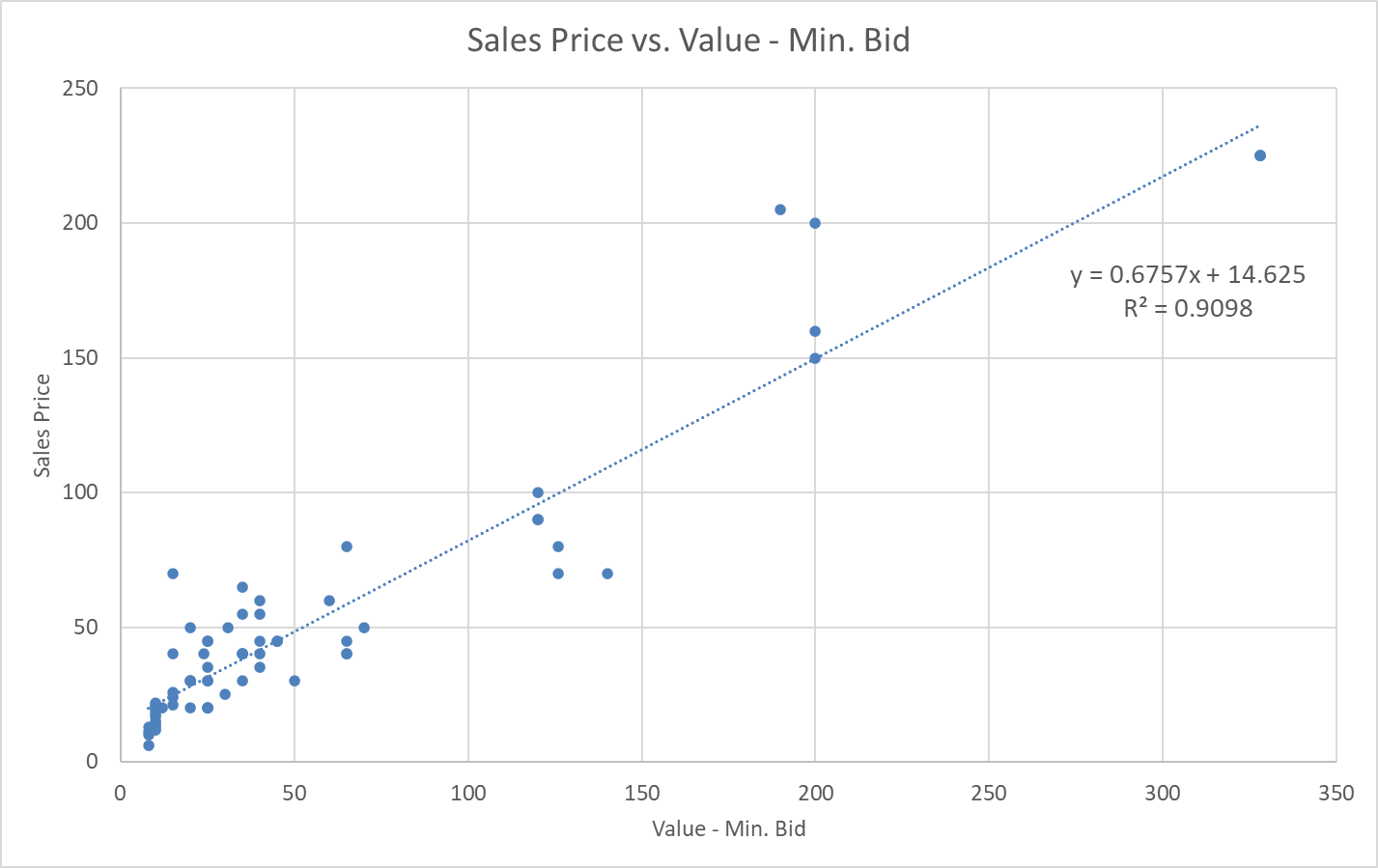
Using the **not underbid** (sales price >= min bid) data with outliers removed, look at a scatterplot of the sales price versus min bid. Explain why you want to use only data from items that were not underbid. Add a linear trend line. Consider if there are any regression outliers. Remove any outliers that appear unrepresentative of this relationship. Get the formula for the trend line Sales = b0 + 11\*Min\_Bid. In theory, when min bid increases, we expect sales price to increase. Verify that b1 > 0. Add the regression information to your spreadsheet in a table like this:





Our theory is that min bid and increment affect sales price somehow. We want to maximize sales and enforce the behavior of people bidding at or over the min bid. Items underbid indicate behavior that we do not want and will not maximize sales. Looking at a model of data using only non-underbid items will help us to see the relationship between min bid and sales that we want to repeat and optimize. Looking at the Sales Price vs. Min Bid Graph, we can affirm the relationship of when the min bid increases, so does the sales price with an R^2 of 0.96 and a positive slope of our regression line.

Now add a column to your data table Value - Min Bid. Look at a scatterplot of the sales price versus Value - Min bid. Add a linear trend line. Consider if there are any regression outliers. Remove any outliers that appear unrepresentative of this relationship. Get the formula for the trend line Sales = a0 + a1\*(Value - Min\_Bid). Also, in theory, when the difference between the value of an item and the mid bid increases, we expect the sales price to increase. Verify that a1 > 0. Add the regression information to your spreadsheet as you did above.



After creating a scatter plot to display the relationship of Sales Price and Value - Min. Bid, we affirmed the theory that when the difference between value and min. Bid increases, so does the sales price. We also noticed there were a lot of items where the difference between the value and min. bid was small, hence the sales price was lower. This implies that value and min. bid are related because in order to get the maximum sales price, the min. bid needs to be low enough to have a significant difference when compared to the value.

So the advice for setting Min bid to maximize Sales is that we look for something that tends to increase Sales when Min Bid increases and something that decreases Sales when Min bid increases in general (i.e. not for a given category). We have two such relationships:

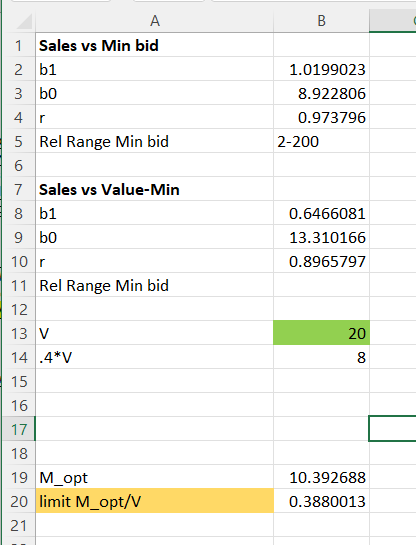
Sales = b0 + b1\*Min bid (since b1 >0, sales increase when Min Bid increases)

Sales = a0 + a1\*(Value - Min bid) = a0 + a1\*Value - a1\*Min Bid (since a1 >0, sales decreases when Min Bid increases)

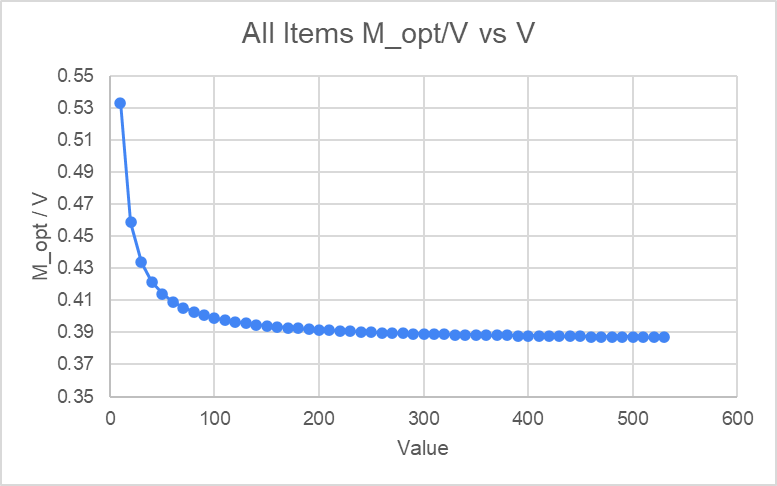
Setting (1) = (2) will and solving for Min Bid will give is the optimal Min Bid (M\_opt):

M\_Opt = (a0 + a1\*Value - b0)/(a1 + b1)

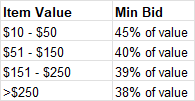
Add cells to your spreadsheet for Value (V) as an input, M\_Opt using the above formula, and the limit of M\_opt/V using the formula a1/(a1+b1). Your spreadsheet should look like this:



Now create a what-if data table using V as the input cell and M\_opt as the output column. Add another column M\_opt/V. Make a graph of M\_opt/V versus V and interpret what this chart suggests for setting the Min Bid for an item. Your graph should looks something like this:



What this chart is saying is that the optimal min bid to set for any item in the $10 to $500 should be from 38%-53% of its value. We recommend this policy for setting the min bid based on the item’s value



The above suggests that lower value items should have higher min bid (% wise). This may make low value items seem less of a “deal” and inhibit interest and reduce sellability. This is due to the model we used to counter the increase in sales price with increase min bid

Sales price = a0 + a1(Value - Min bid)

This is a “perceived deal” model. This is a bit counter to our goal of maximizing sales price in that it relies on the assumption the better the deal (what the item may obtain relative to its value), the higher an item will sell for. This is a pretty dubious assumption, especially given that people are looking for deals at an auction more than their desire for a particular item. This is because the majority of the items are not terribly unique or unobtainable outside the auction and there’s only so much extra people are willing to pay as a donation.

Let’s look at a model where we consider “actual deal” which doesn’t rely on a perceived deal assumption. Consider

Value - Sales = a0 + a1\*Min bid → Sales = Value - a0 - a1Min bid

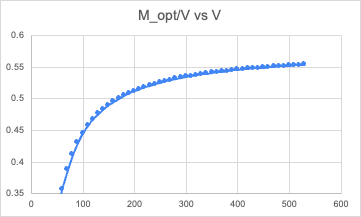
To explain the above model, the larger Value - Sales, the “better deal” was achieved at the auction. We would like to set the min bid to curb “good deals” as they encourage underselling (sales/value less than 40%) and thus lower sales price than what we would hope to obtain. So in using this model to balance the increase in sales price we get from increasing the min bid:

b0+b1\*min bid = Value - a0 - a1Min bid

Solving for Min bid gives

M\_opt = (Value - a0 - b0)/(a1 + b1)

This gives an M\_opt/V graph that looks like this:



The above model suggests lower value items be given lower min bids (encouraging deals) and higher value items be given higher bids (discouraging deals). My thoughts are this model makes more sense for setting the min bid than the previous model.

To determine the optimal increment, assume we have already determined the optimal min bid M\_opt (as above). An Increase in Increment (Inc) tends to increase Min bid\*Sales

Min bid \* Sales = a0 + a1\*Inc → Sales = (a0+a1\*Inc)/Min bid

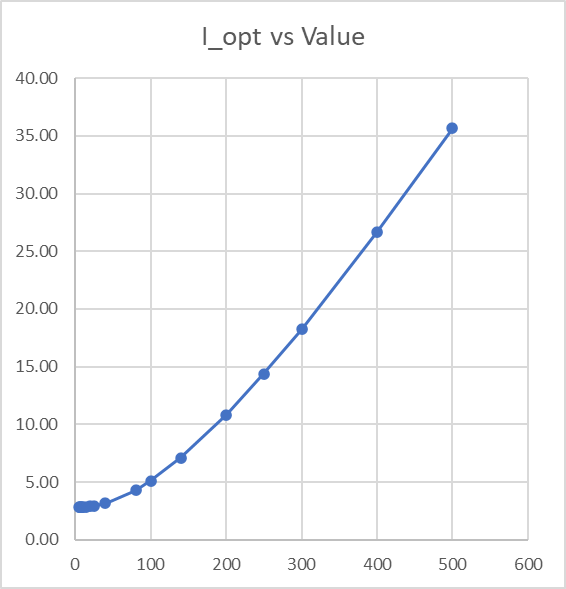
Inc is dependent on Value which we cannot set, we see an increase in Inc would increase Value - Sales which implies a decrease in sales

Value - Sales = b0 + b1\*Inc → Sales = Value - b0 -b1\*Inc

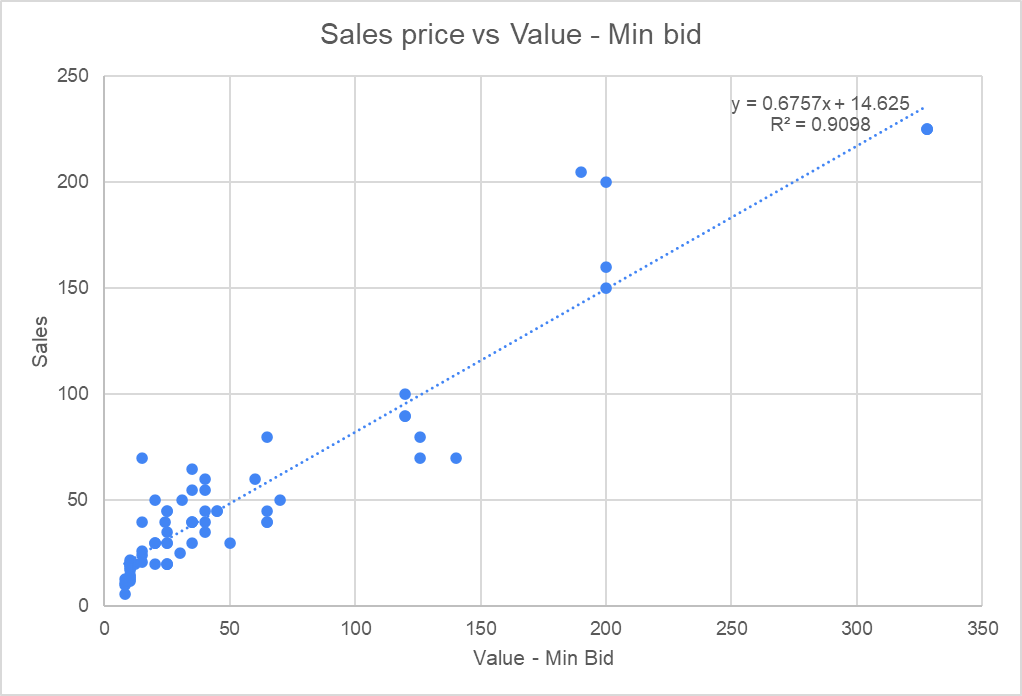
So to maximize Sales given a Value and optimal Min bid M\_opt, we look at where these are equal and solve for Inc (we’ll call it Inc\_opt) giving

Inc\_opt = (M\_opt \* (Value - b0) - a0) / (a1 + b1\*M\_opt)

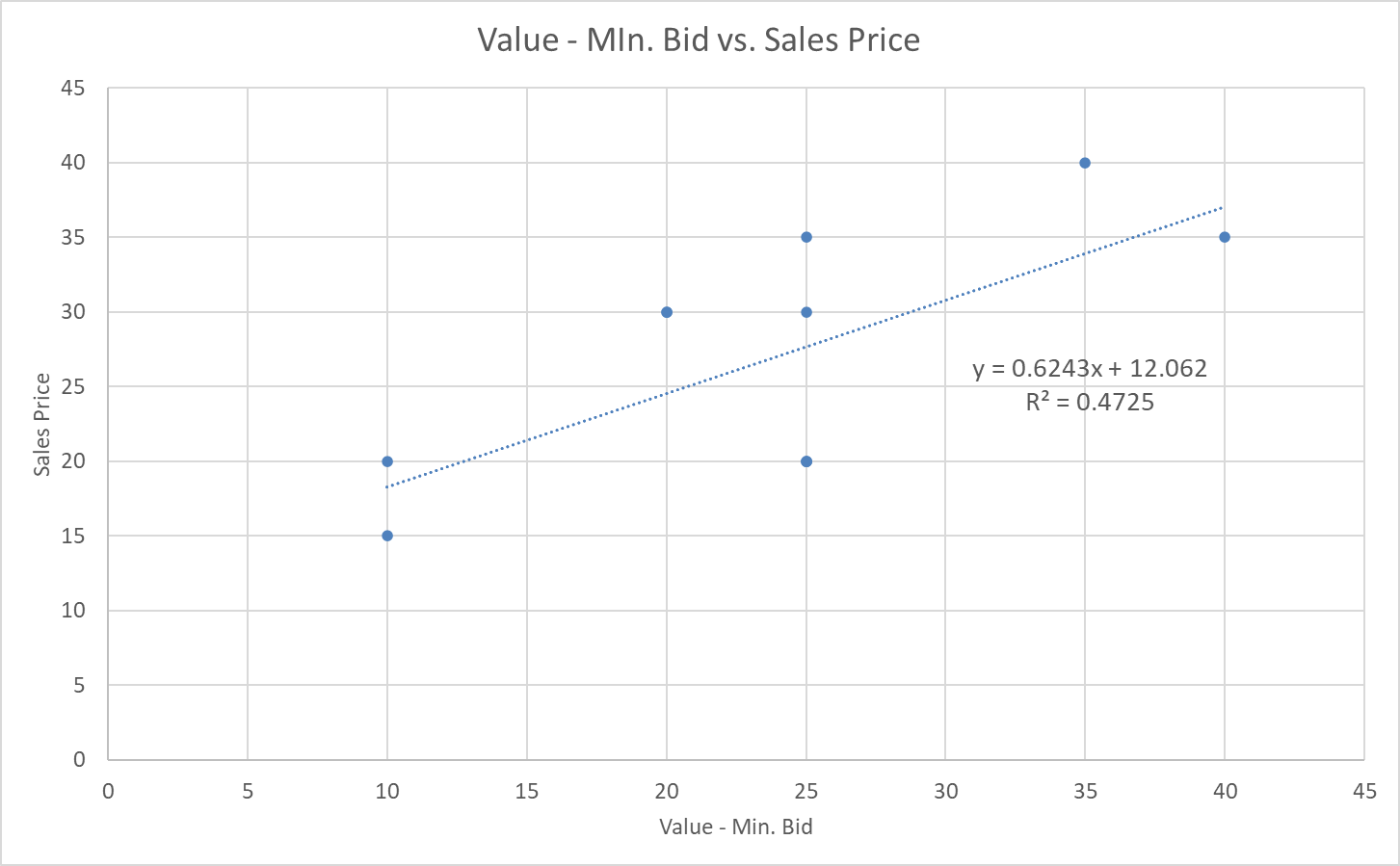
Get the coefficients a0, a1, b0, b1 by adding Min Bid \* Sales and Value - Sales to your data table, look at the scatter plots, add trend lines etc. Then use the above formula as you did above and create a graph I\_opt/Value. Interpret how this can be used to set the increment for an item given its value and M\_Opt. Your graph should look like this:



Investigate if the category affects M\_opt and I\_opt. If so, discuss how you might handle this? Hint: you can look at what the relationships discussed above are for each category and note if they appear notably different for some categories.



Food Category



Looking at our initial graph of Sales Price vs Value-Min Bid we see a cluster around the lower difference of value-min bid and lower sales indicating that there must be a possibility certain items within categories are affecting the Min Bid. We now see within the food category there are discrepancies that led us to believe that we must have the optimal bid in certain categories.

After calculating the optimal min. bid for the 5 categories with the most amount of items (using not underbid items), we determined that category does play a role in affecting optimal min. Bid and optimal increment. Using Food, Miscellaneous, Restaurant, Sports Fitness, and Wine, we discovered these categories do not have the same relationship as the original optimal min bid and increment. For example, the optimal min. Bid for Sports Fitness and Wine items are almost statistically perfect with r-square values being close to 1. This means the min bid is set perfectly for Sports Fitness and Wine items. However, the same could not be said for the optimal increment where Wine items seem to have the optimal increment set. Overall, we should set the Min. Bid differently for each category.

Discuss how these models could be used in a practical way for setting Min Bid and Increment for an item at the auction [ ]

Explain what decision(s) from the decisions listed earlier this model is useful for. It doesn’t have to address the entire decision. It can contribute information or help answer a necessary question involved in the decision. Explain how it is used to inform the decision(s). [ ]

|  |
| --- |
| Explain using the descriptive models in previous parts what performance problem(s) this model is intended to contribute to addressing. Also note any important considerations from the performance results that must be noted in using the predictive model. [ ] |

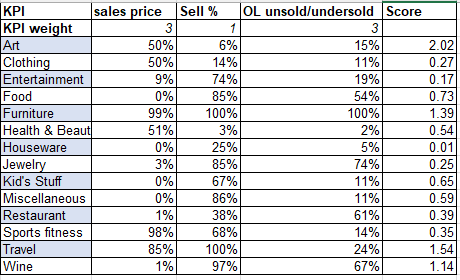
*M8: Prescriptive model for goal driven prioritization of item solicitation by category and value*

*List the goals of the auction (e.g. maximize funds raised) then assign a weight (1-5) to the importance of each of these, 1 = not so important, 2 = somewhat important, 3 = moderate importance, 4 = important, 5 = most important*

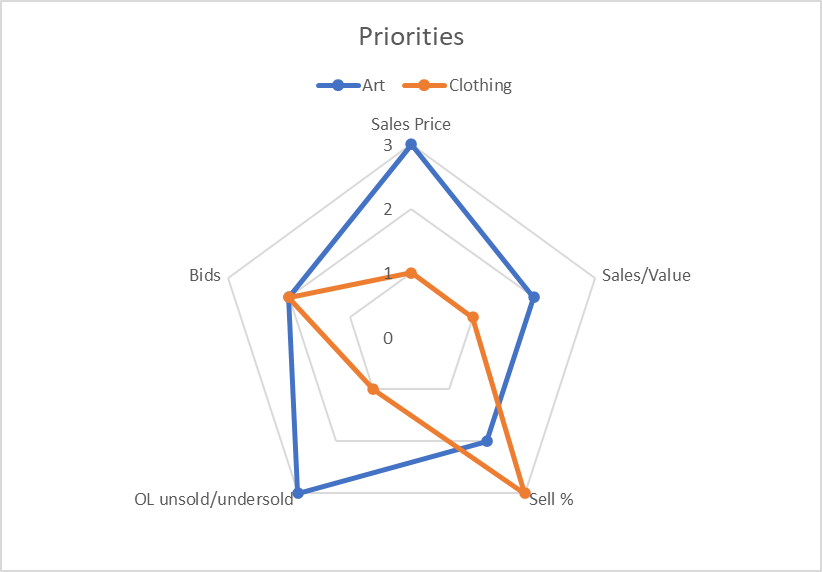
*Use the rankings from models M1-M5 and the weights from above to determine a weighted rank for each category. Use a table like this:*

**

*Alternatively, you can use the confidence levels instead of the ranks:*

**

*For each category make a spider diagram that profiles the rankings such as:*

**

Discuss how these models could be used in a practical way for setting Min Bid and Increment for an item at the auction [ ]

Explain what decision(s) from the decisions listed earlier this model is useful for. It doesn’t have to address the entire decision. It can contribute information or help answer a necessary question involved in the decision. Explain how it is used to inform the decision(s) [ ]

*M9: Prescriptive model for accepting an item of category X and value V or asking for cash donation*

*Which of the models M1-M7 are useful for this? Do we need other models? Do we need to determine some acceptance criteria? Can you think of this in terms of a decision payoff matrix? Would a dashboard tool be useful here?* [ ]

*Do we need anything for the size of items? Why or why not?* [ ]