DATA 591, Data Science Capstone II, Project Implementation, Winter 2018, University of Washington

Findings & Recommendations for Kids on 45th

A Consignment Sales Data Analysis Project

Gary Gregg

Jahnavi Jasti

Abhishek Varma

Submitted to:

Elise and Bookis Worthy

Introduction

Our Client

Our client, *Kids on 45th* is a mixed retail and consignment shop in the Wallingford neighborhood of Seattle, Washington. The store has been a part of the Wallingford community for more than 30 years and is cherished by many in the neighborhood. In June 2017, Elise and Bookis Worthy purchased the store and are currently working on refreshing the store's policies and processes in order to increase profitability.

Kids on 45th says this about their customers:

We think of our consignors as friends and partners, working together to sell items quickly and for the best price.

The typical *Kids on 45th* customer has growing children, and children outgrow clothing. As their children grow, parents will attempt to sell clothing items by consignment. In order to maximize revenue and keep the business as healthy as possible, it is important for *Kids on 45th* to price incoming consignment items correctly. Factors affected by pricing include total revenue for the store, customer satisfaction, and consignor partner satisfaction. These three factors together ultimately affect the success of the enterprise.

The Client's Need

After acquiring *Kids on 45th*, our clients inherited several databases of data detailing ten years worth of store sales and customer information. Many of the items sold are consignment items. These are sold at low costs with small margins. Therefore, in order to better understand whether they are pricing items correctly, *Kids on 45th* required an analysis of the sales data that was left by the previous owners. Elise and Bookis Worthy are most interested in the question: *What Sells Things?* To this end, their first interest is a report with detailed analysis of the legacy sales data.

Upon item sale, *Kids on 45th* consignor partners receive 40% of the proceeds, available as a store or online credit. In this way, the partners may outfit their children with an ongoing stream of quality used clothing. *Kids on 45th* currently has more willing consignor partners than they can handle. Nevertheless, one challenge for this enterprise is the availability of quality new clothing that can be purchased from competing retail chains such as Target.

Legacy Dataset

What *Kids on 45th* Is Currently Using

When Bookis and Elise Worthy acquired *Kids on 45th*, they started a new sales tracking system that includes the following tools:

- Apache Lucene – Apache Lucene provides Java-based indexing and search technology
- Elasticsearch – Elasticsearch is an open-source, RESTful distributed search and analytics engine build on Apache Lucene
- Kibana – Kibana is an open-source data visualization and exploration tool used for log and time series analytics, application monitoring, and operational intelligence use cases

What the Previous Owners Left Behind

The legacy sales data – encompassing data from 2007 through 2017 – is contained in Microsoft Access tables. It is currently not in a form convenient for easy analysis by the enterprise.

The dataset was a collection of three disparate databases:

- Custdata (customer data)
- Product
- Sales

Each database has a number of tables that capture information about the particular field. Some of these data are replicated in other databases, and there is no foreign key that joins these databases together. For example, it is impossible to connect individual transactions in the *Sales* database with any of the product details from the *Product* database. This means that finding meaningful trends between the different databases is challenging. A diagram of the three databases and their tables is illustrated below.

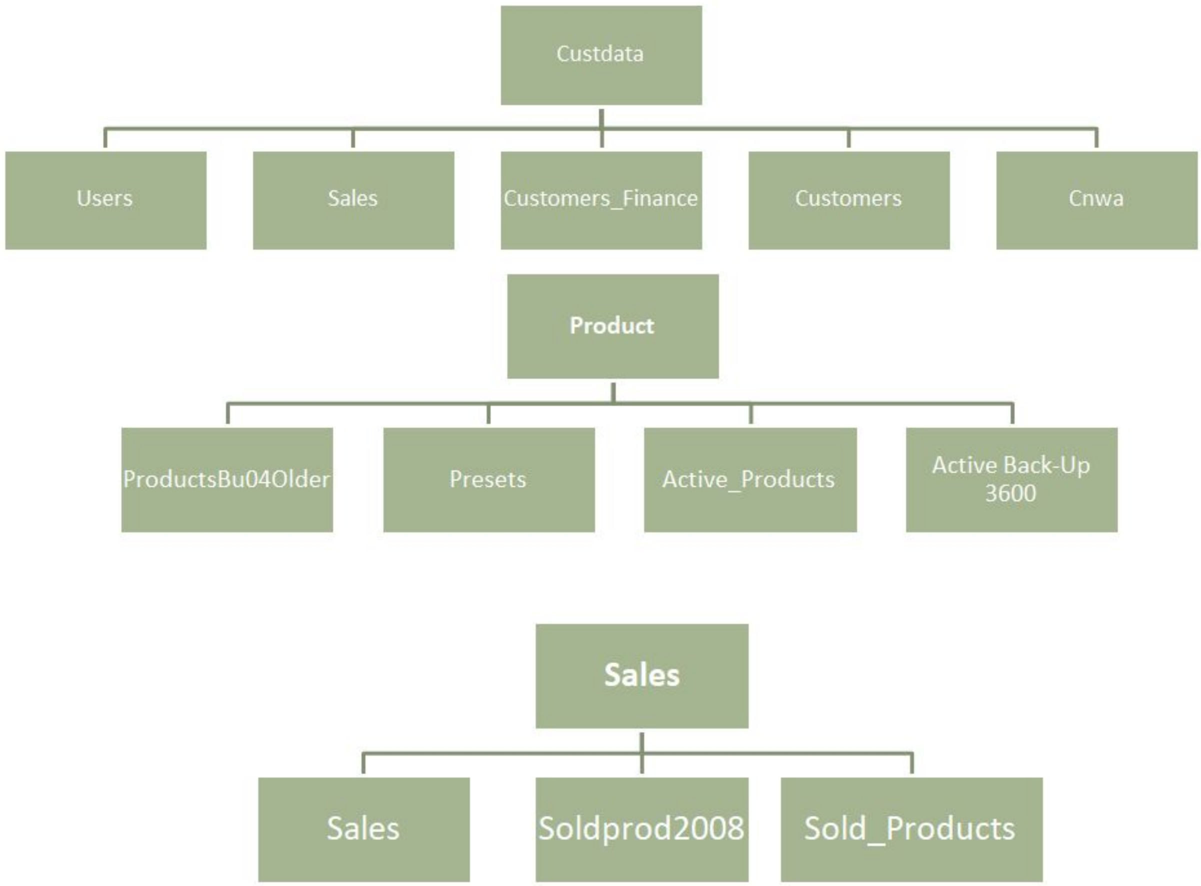

Figure 1: Table breakdown of the three large, legacy Access databases maintained by *Kids on 45th*

Questions, Objectives and Tools

Research Questions

Upon meeting with the client several times and exploring the data, we were able to define the scope of our project and understand how we would be able to provide the most value. The greatest need of *Kids on 45th* is a detailed report of our research findings. The most important questions for the enterprise to have answered were:

- *What is the mean price of consignment items per item category*
- *What is the mean time-on-shelf of consignment items per category*

These questions were identified to have the greatest impact on profitability moving forward and would enable us to explore the dataset with the right goals in mind. Due to the difficulty in tying the information in the various databases together, we decided against development of a predictive machine learning model. The output would be difficult to verify and have questionable impact.

Design Objectives

The primary objective of our team was to assist the enterprise by analyzing their legacy sales data, and produce a detailed report. In order to fully address the needs of the enterprise, we have tried to address the above mentioned research questions, and also do analysis on customer behavior. Another goal was to address the effects of inflation.

The customer behavior analysis included discovering the purchasing habits of both the consigned partners and new customers: How each of these customers are buying consigned items, new items and both types of items. This information can be used to find whether consigned or new items are being sold more often, and allows those items to be featured more prominently in the store.

The inflation analysis concentrates on whether the pricing of the items has tracked inflation. A dollars worth of item ten years ago is more expensive now (see analysis, below). Prices have therefore been adjusted to inflation.

Our Tools

Since *Kids on 45th* is an ongoing concern in the Wallingford neighborhood, it may be possible for teams from future cohorts of the *MS Data Science* program to build on the work of our project to further assist the enterprise.

The following tools have been used to analyze the legacy data:

- **RStudio**
- **Tableau**
- **Python, Jupyter Notebook**
- **SQL DBM**
- **MS Access, MS Office**

All code, notes, and visualizations produced by the authors have been shared with the enterprise for the possibility of future analysis.

Item Categories

Category Definitions

All the items in the store are classified into different categories. The most common items in each category can be seen below:

Category 1*: shirt, pants, dress, pj, nightgowns, robes, cotton one-piece outfit or set, skirt, sweatshirt, 2 or 3-piece set, vest, overalls, etc.*

Category 2*: toy, puzzle or game, Universal Toysmith, Jellycat, Small world toys, EEBOO, Janod toys, Waba fun, Rubbabu, book, Green toys, etc.*

Category 3*: New items, Children's books, sandal or shoe, sox, tights, blanket and linens, panties, boxers, briefs, bottles, cups, bowls, CD or DVD, video or audio tapes, etc.*

Category 4*: Jumper, Stroller, Basket, Bumkins, Carrier, Carseat, Highchair, Kaboost, Kleen Kanteen, Baby Monitor, etc.*

Category 5*: shorts or summer rompers, sweater, leotard or swim costume, lightweight jacket, snow pants, heavy weight jackets, snow boots and rain, swimwear, etc.*

From the above categorization, we observed that categories 1 through 4 are well classified. However, category 5 contains all the seasonal items, including summer, rainy and winter, and are fit into a single category. In the future, we recommend that these items may be classified into different categories according to their seasonal use.

There is also a category for the miscellaneous items. These are the items in the dollar rack and are all sold for a dollar. Some of the items - likes clothing or small toys - that are not sold for a certain number of days are put in the dollar rack for clearance.

Some of the items do not have a category mentioned, and we discussed ignoring them as they are only 3600 out of 530k items. This is a very small ratio.

Consigned Versus New items

From the *PP_Sold_Products* we got to know which consignor partner purchased what items. During our meetings with Elise, she had mentioned two important points that helped us during our analysis.

- *Customer_no_product* of '*999*' means that the items are new retail items. The store buys them from a wholesale dealer, and sells them at a retail price suggested by the dealer
- *Customer_no_sold* of *'600'* means that the store bought this item. This means that it was not sold to any customer. The store keeps them, or it is bought by the store itself

Using the above information, we were able to classify items as new or consigned items.

Average Days on Shelf per Category

To answer this, we only considered the consigned items. Date difference between *Date_In* and *Date_sold* of an item, in number of days, is calculated using *DATEDIFF()* method in *R*. The reason to select only the consigned items was because the new items have the same SKU number when they are sold and restocked. For example, if a new item with SKU as *123* is sold today, then the item would be restocked again on the shelf with the same SKU number. This led to the number of days on shelf for some of the new items being more than 2,000 days. This made it very difficult to actually calculate the number of days for a new item to be sold.

Figure 2 shows the average number of days on shelf for each category.

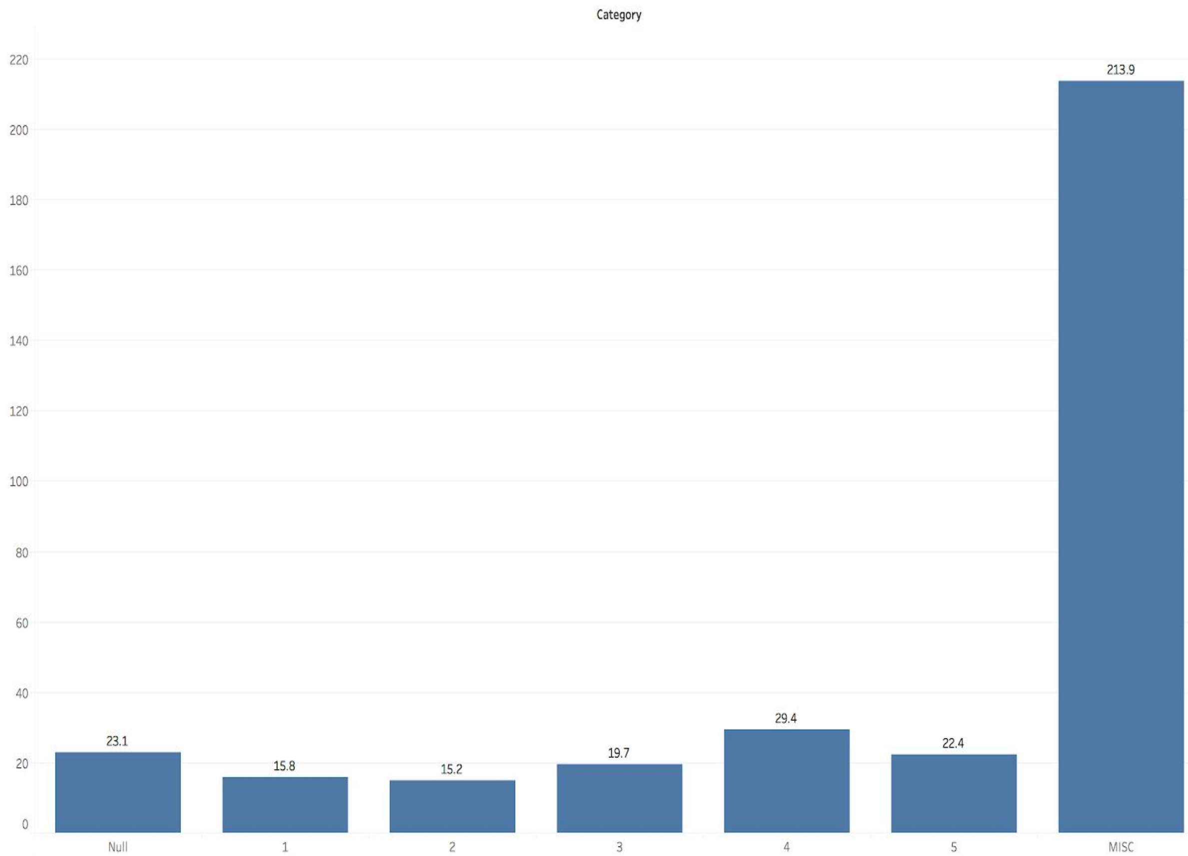

Figure 2: Average number of days on shelf of an item, per category.

From the above figure we can see that all the categories from 1 through 5 are being sold in less than a month, which is a good sign. Some of the items which are not sold and sent to the dollar rack, stay on this rack for an average of 213 days.

Average Selling Price per Category

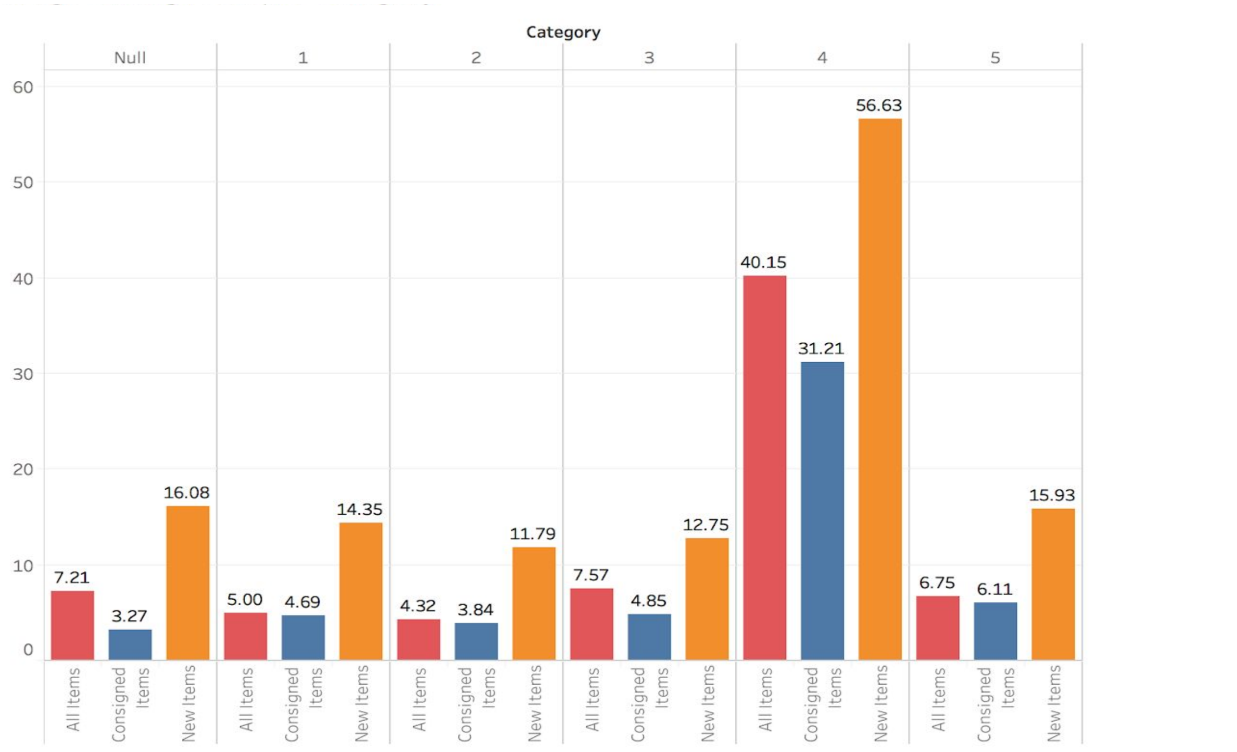

Figure 3: Average selling price per Category

It is obvious that the new items are sold for a higher price than the consigned items in all categories. From the above figure we can see that the range of average sold price of an item in any category, except 4, is between $3 and $6.

Customer Behavior

After analyzing the pricing of items with respect to category, we also thought of analyzing the purchasing behavior of customers.

There are two kinds of customers: *Consigned partners,* and *New customers.* Consigned partners are those who have sold at least one item in the store, and have an account with the store. New customers are those who do not have an account but they do come in to buy items.

From the *PP_Sold_Products* we were able to divide the customers into these two categories by assuming that the customer id - which is in both the *Customer_no_sold* field and *Customer_no_product* field - are consignor partners. Others are new customers.

The next analyses are based on these 2 customers, and their purchasing habits.

Consignor Purchasing Behavior

Figure 4: Consignors purchasing behavior

From the Figure 3, we can observe that the Consignors are buying less nowadays when compared to ten years ago. The trend line keeps decreasing from 550 to 400 since 2009. This may be because of many reasons, some of them have been discussed in the following sections.

Consignors Buying Consigned items

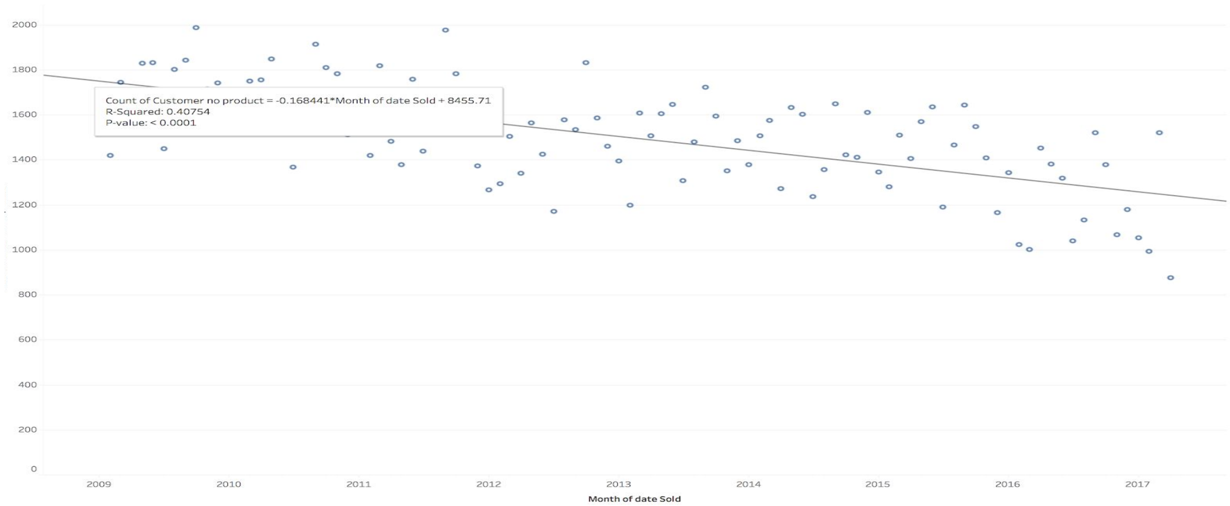

Figure 5: Consignors' behavior of buying consigned items

Consignors Buying New Items

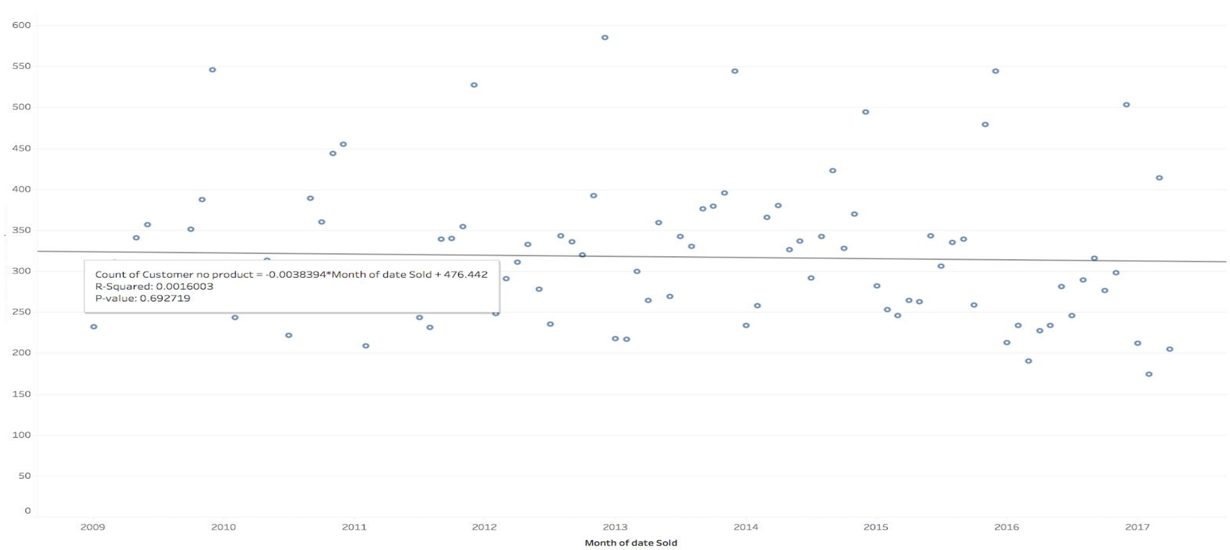

Figure 6: Consignors' behavior of buying new items

Comparing figures 5 and 6, we can see that consignors are constantly buying more new items than consigned items. Consigned items have a sloped trend line, meaning that the consignor partners have become less inclined to buy consigned items. The demand for new items, however, has remained consistent. There could be one reason for this: consignor partners get a 20% discount on the new items, even without enough credit in their account. For example, a person can consign only one item (or more) and can open a consignor account. With this account, he can build up his credit. This is what he gains from selling his items through the store. These consignors can obtain a 20% direct discount on the new items. The previous owners had a policy that the consignors needed to have at least $1 in their account balance to be able to obtain the new item discount. But this was discontinued when the new owners, Elise and Bookis, took over the store. Now, the consignors do not need to have a minimum balance in their credit account. This really encourages consignor partners to utilize their discount and buy more items in the store, as opposed to purchasing them elsewhere.

New Customers Purchasing Behavior

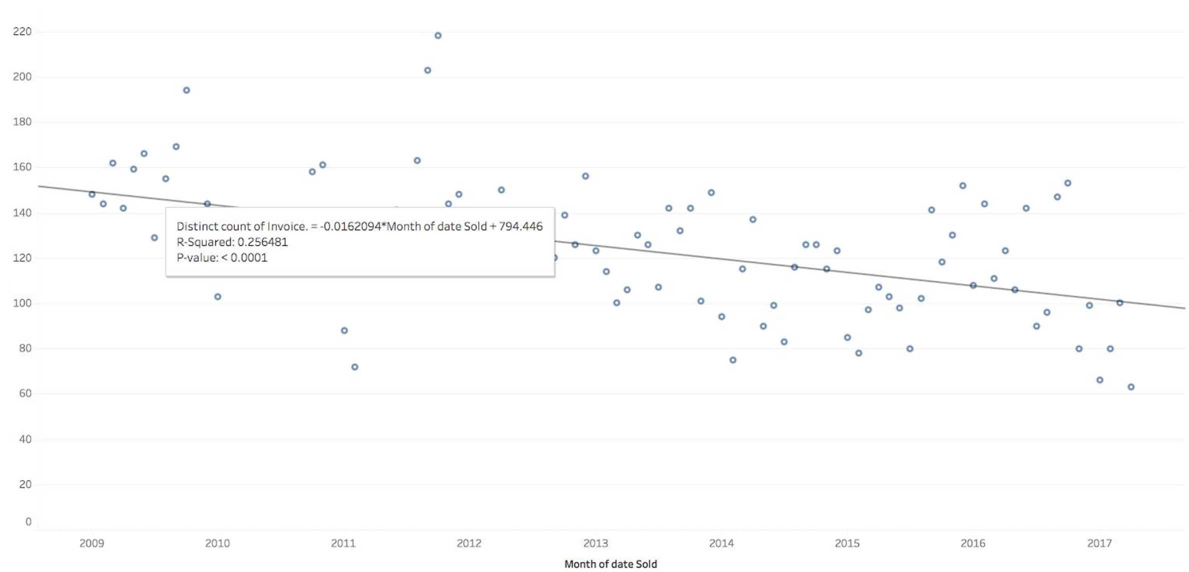

Figure 7: New customers purchasing behavior

Purchasing activity of new customers also tend to decrease over time. There are some outliers during the end of 2011, but most of their purchases tend to remain around the fitting line.

When compared to the purchasing behavior of consignor partners, new customers have a small slope. This means that the purchases done by consignors are decreasing in a faster rate than that of the new customers.

New Customers Buying Consigned Items

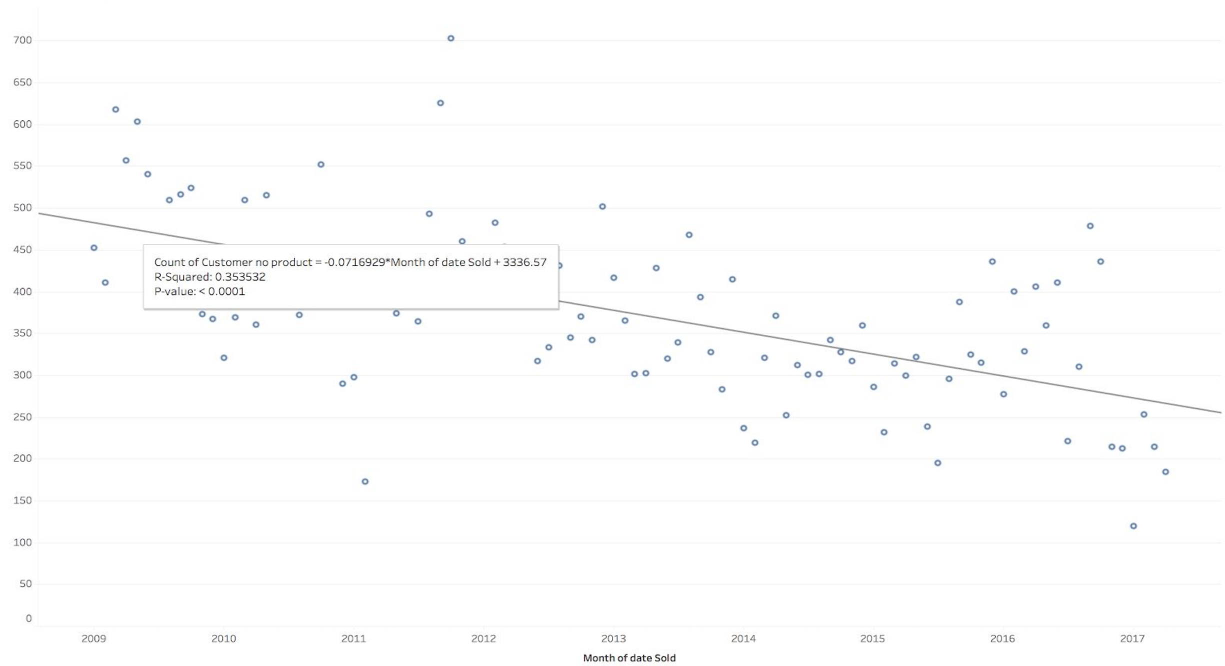

Figure 8: New customers' behavior of buying consigned items

New Customers Buying New Items

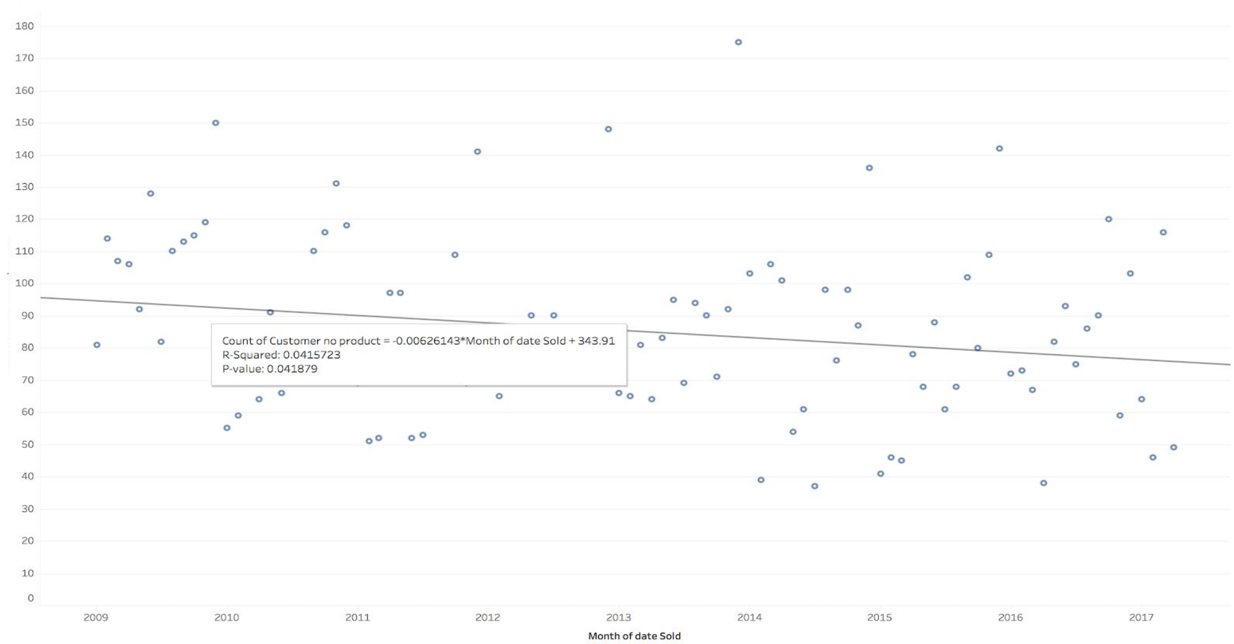

Figure 9: New customers' behavior of buying new items

New customers are buying less consigned items now than they did 10 years ago. This might be because of the competitors that have evolved in these years. Stores like *Kids on 45th* have a huge competition with stores like Target, which sells kids clothes starting from $5. So people would rather buy a new dress for $5 than buy a used for $3 - $4. This has made it really difficult for consignment stores of this type.

However, branded and seasonal items, such as winter jackets, rain boots and snow boots may have huge sales in the consignor stores as compared to stores like Target. These items are a lot more expensive when bought new than when you buy a consigned item. These seasonal items are used by kids only for 1 to 2 years, after which time the kids outgrow them. Parents prefer buying these kinds of items more often as used items. They do so more often than items such as ordinary used clothing.

Dealing with Inflation

Changes to the Consumer Price Index (CPI) 2007 - 2017

Merriam-Webster defines inflation as "*a continuing rise in the general price level usually attributed to an increase in the volume and credit relative to available goods and services*" (see Merriam-Webster). During the period when *Kids on 45th* collected their data (January 2007 to April 2017), inflation had been relatively tame in the City of Seattle, and in the country as a whole. This was principally due to the general economic malaise in the USA during the "Great Recession" (2007 - 2013), and an effort by the Federal Reserve to keep interest rates low. During these years, there were fears of *deflation* (or a general fall in the value of money). Deflation can be a far more economically damaging trend than excess inflation, and the government took great pains to avoid this condition. The government's goal for optimum economic growth is a currency inflation rate of approximately 2%. It is able to exert influence on this trend through a mixture of adjusting short-term interest rates, and also controlling the supply of money in the economy.

Over time, the impact of even modest inflation can be substantial. During the period when the previous owners of *Kids on 45th* collected sales data, the overall rate of inflation in the USA was 1.21%. That is, a dollar in January 2007 had the same buying power as $1.21 in April 2017. This works out to an annual inflation rate of only 1.92% over the period.

Because the previous owners of *Kids on 45th* recorded a sale event on the day they made a sale, the data represents a snapshot of the general value of a dollar on an almost daily basis from January 2007 through April 2017.

In order to make comparisons of sales data in most situations, we judged it necessary to normalize sales price information to a common value. As a practical matter, it made sense to try to normalize the sales data to a common dollar value as close to the present as possible. In this way, the normalized data would have the most meaning to persons analyzing it in the 2017 timeframe.

How to Normalize?

How should the sales data be normalized? Firstly, we needed an accurate source to measure the effect of inflation. There are very accurate sources for determining inflation at the national level. However, the cost of living in Seattle has risen more than in the country as a whole, and more than in the State of Washington. We were able to obtain very accurate data on the cost of living in Seattle during the required time period. These data were available from the city itself (see City of Seattle). The data supplied by the city can be downloaded as an Excel spreadsheet, or PDF file, and give a CPI index that is set relative to 100 for the 1982 to 1984 timeframe. The CPI data are given in bimonthly indices. We can see, looking at the data, that the CPI index in February 2007 was 211.704, and in April 2017 it was 261.560. Thus the rate of inflation in Seattle was (261.560 / 211.704) = 1.24%, which contrasts to the lower 1.21% rate nationwide.

But, again, CPI data are given at bimonthly intervals, and *Kids on 45th* recorded sales on a nearly daily basis. This makes it necessary to interpolate between the bimonthly CPI indices supplied by the City of Seattle. For example, a sale recorded in March 2007 would occur approximately halfway between CPI indices provided for February 2007 and April 2007. But how to interpolate? It would be quite easy to connect the CPI indices with a straight line. Such a function would be continuous, but contain sharp points of inflection at the CPI indices themselves.

We devised what we thought a better solution. Instead of interpolative lines between the CPI indices, we constructed a *spline function* with the data points. A cubic spline is a function that consists of a series of 3rd order polynomials. Using polynomials of this order, it is possible to connect the CPI data points, and also ensure that the slopes of the lines between the points remain continuous. Instead of sharp points at the provided CPI values, the curve is smooth, and still continuous. Our method seemed more logical, as the forces that control inflation in the short term - while perhaps having unknown cause - would nevertheless be applied gradually, resulting in a smooth CPI curve.

Input of the CPI Data and Creation of an Interpolative Function

Using the PDF data supplied by the City of Seattle, we created a comma-separated value file (CSV) that gave the bimonthly CPI indices supplied by the city. We used the R programming language to read this file, and the *splinefun* function (see *splinefun*) to create the continuous and smooth CPI function we needed to interpolate CPI data for any sale that occurred between January 2007 and April 2017. Figure 10 gives these results:

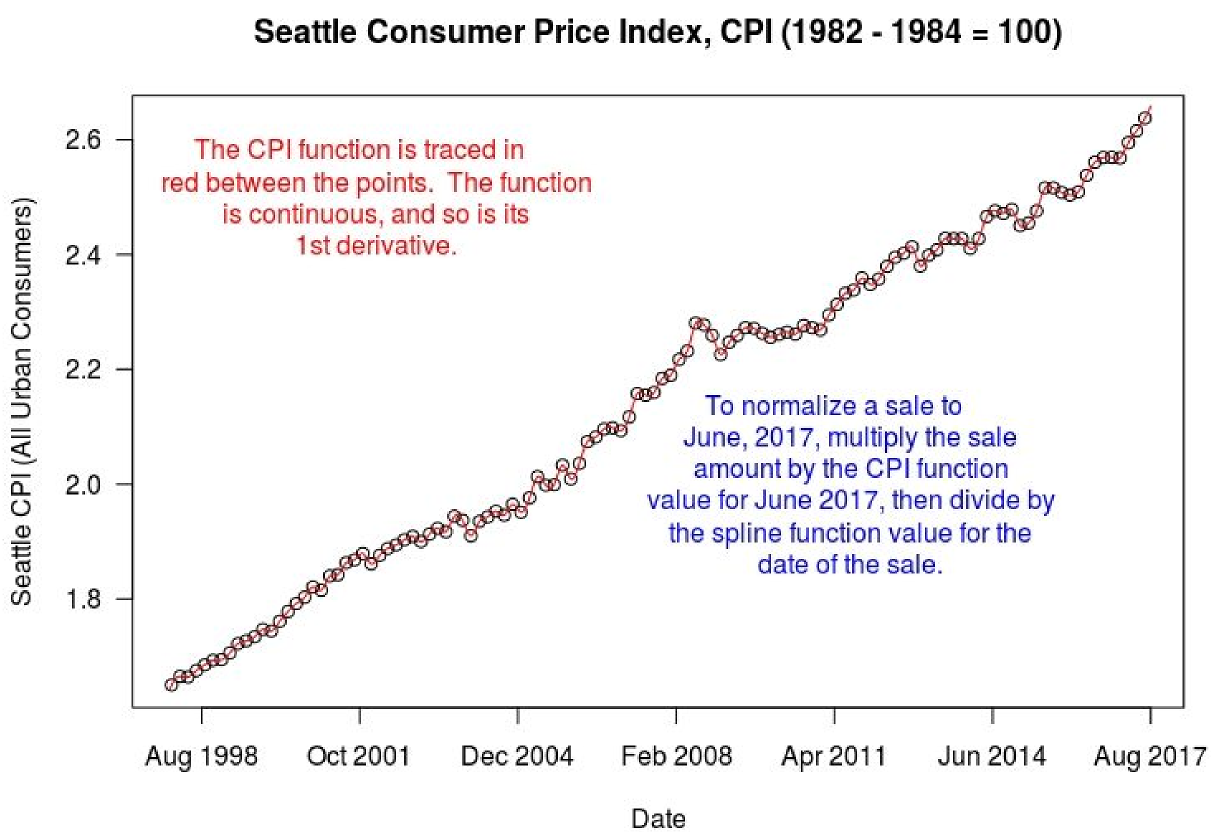

Figure 10: CPI data points supplied by the City of Seattle (circles), and interpolating function (red line)

As you can see, in the City of Seattle there has been a general rise in inflation during the dates in question. However, there are also times when there local drops, or short periods of *deflation* during this time. Our resulting spine function is valuable in this way: For a sale that occurred on any day between January 2007 and April 2017, we need only use the function to obtain a very accurate CPI index for the day in question.

Creation of Inflation (CPI) Normalized Sales Data

To make dates useful as the domain of our CPI adjustment function, we first turned the dates into seconds elapsed since the beginning of the UNIX epoch (1 January 1970). Typical values are 1,167,609,600 for 1 January 2007, and 1,491,004,800 for 1 April 2017. While the resulting integers are large, they provide very fine-grained resolution for our CPI adjustment function.

The method for CPI-modification of a currency value is thus: First, use the date of the transaction to create a UNIX epoch value. Next, use the resulting value in the CPI function (modeled in Figure 9), and arrive at a CPI index. Call that value "*x*". Use the CPI function to obtain a value, say "*y*", for UNIX-epoch value for the desired normalization date. Multiply the sale amount by "y", then divide by "x". The result is the normalized sale amount. Thus a sales value of $15.29 on 2 January 2007 is CPI normalized to $19.19 on 1 June 2017.

Decisions Made in How to CPI Normalize the Data

The legacy sales data for *Kids on 45th* comes in four Microsoft Access databases, with six or seven tables in each of the three main databases: *Custdata*, *Sales*, and *Product*. For any table that contained both a date, and a monetary column, we added two fields. First, we added the field that indicated the long-integer UNIX epoch equivalent of the date. Second, we added currency fields that adjusted the sales amount from the date in that same row, to 1 June 2017. We arbitrarily chose 1 June 2017 as our normalization date since it is the last date provided in the City of Seattle inflation data. It is the date closest to when *Kids on 45th* was acquired by its current ownership, and it is the date closest to today. Obviously, however, inflation continues to march on. Later researchers will need to account for this, but they may analyze our modified monetary fields and know that the fields represent apples-to-apples value comparisons in June 2017 dollars.

Some tables in the sales data contain "date in" and "date sold" fields. They also contain list prices for items, and their eventual sales prices. For these tables, we chose to normalize list price using the date-in value, and normalize sold price using the date-sold value. Because of this, our normalized data differ much more often between list and sold prices than do the unnormalized amounts. Consider a consignment item that comes into the store, receives a list price, but then is not sold for months. Even if it sells at that same list price, the value of money has changed slightly in the time the item sat on the shelf, waiting for a sale.

Balance Due Versus Inflation

The Sales Table

Our examination of the effect of inflation obviously raises the question of how well sales amounts, and prices for individual items have tracked inflation over time. First, we examined sales amounts. We determined that the *Custdata* and *Sales* databases have identical tables, *sales*, that have the same schema. Further analysis suggested that the previous owners of *Kids on 45th* regularly replicated the *sales* table from the *Sales* database to the *Custdata* database. We drew this conclusion because the latest *Custdata* sales table contains exactly three weeks fewer records than the equivalent table in *Sales*. The last record in *Custdata* is 1 April 2017, while the last record in *Sales* is 22 April 2017. It seems the staff performed this replication after close-of-business on Saturday, but before Monday opening.

There is an important conclusion to be drawn from the assumption of table replication across databases. It appears that the former owners of *Kids on 45th* did not apply "foreign keys" across the legacy sales databases. If they had, there would be no need to replicate tables.

An example of cross-table foreign keys would be as follows: Suppose a sales, or "products sold" database contains keys that identify the types of items sold in a transaction. They key itself might only be a number, or an index. Elsewhere, in another table, we might find expoundatory information about the key itself: A textual description of the category and other information. This design eliminates the need to duplicate the expoundatory data. For example, if included in a single table, the textual description for a category would need to be replicated again and again for each relevant occurence in a sale. To state again: This type of cross-table key information appears not to exist across databases in the legacy sales data. The other conclusion to be drawn from the activity of replicating data tables is that the database designers had limited sophistication in their database design abilities at the time. Replication is error prone, duplicative, and wasteful of space resources. For a better database design, see our recommendations later in this report.

Balance Due Versus Inflation

Using the *sales* table from the *Sales* database, we interpreted the "balance due" field to mean the total cost of all items purchased as calculated at the register, plus tax. This field seems most representative of the amount exchanged in a single transaction. We realized that a transaction might consist of a purchase of both new, and consignment items, so we made no assumptions about the types of items purchased.

First we made a cut of sold items by month, then took an average of balance due for each month. We used mean as our measure of average. Next, we plotted the average balance due per month against dollar values, and added a trendline using a linear model adjusted for exponential growth. Exponential growth is the best measure for an inflation adjustment. The resulting value for any month in such a trend can be viewed as a factor greater than one applied to the previous month. This is an exponential trend, even if the factor used is close to one. After having done this, we added a trendline for inflation starting at the first month of the graph, April 2007. Figure 11 shows our results:

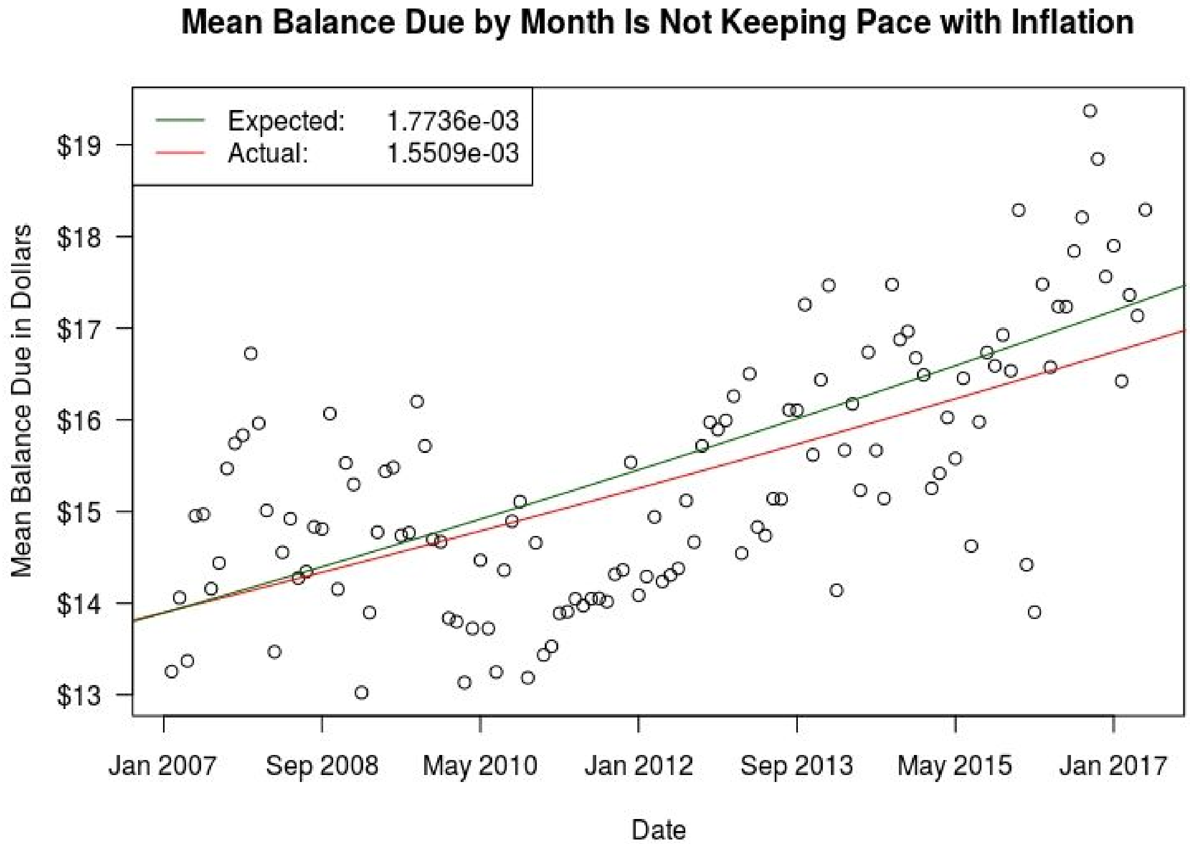

Figure 11: Mean balance due per month versus time (red line, trend; green line, inflation)

Although there is considerable variability, one can see that average balance due is not keeping pace with inflation. Our analysis identified three outliers with regard to these data. In March 2017, the previous owners of the enterprise had announced that the store would be closing. In April 2017, consignor partners were in the process of attempting to use all the store credit in their accounts. It was either that, or lose the credit. This activity resulted in outsize sales that month. Sometime later the business transitioned to new ownership, and remained open. But the sales anomaly remains in April 2017 due to the expectation of store closure. Two other outlier values are present due to the regularly back-to-school seasonality of sales in the fall of 2016. See our discussion of sales seasonality later in this report.

Interpretation of Results

Our interpretation of the results had left us with the initial conclusion that pricing of consignment items had not kept pace with inflation. We were informed by the current owners of *Kids on 45th* that the current process of pricing consignment items relies on the expertise of the staff performing the pricing. There is no right or wrong answer to this, only a general feeling about correct price from an individual who has done a lot of pricing in the past.

Question: *What is right or wrong pricing?* Were consignment items priced too high at the beginning of the data collection, and then did the staff adjust prices as market forces adjusted sales expectations? Or were values correctly priced at the beginning of the collection, and then subsequent pricing not keep pace with the CPI? It seemed possible to assume the later. With the effects of the Great Recession, it might have been possible that the staff had lowered their expectations of what customers would pay. Had there then been a lag for them to adjust their expectations as the economy recovered? This might explain an assumption of correct pricing at the beginning of the data collection, and too low pricing later on. And it might also explain the statement of one *Kids on 45th* customer, who told us that consignment items were "dirt cheap" at the store.

There is also this realization with which to contend: Our examination of balance due versus time lumped together purchase of consignment and new items. One, or the other, or a combination of both could have, and would have been purchased together in any single transaction. Since new items are somewhat more easily priced due to a suggested retail price (SRP) by the manufacturer, were our findings the result only of inefficiencies in the pricing of consignment items? Only an analysis of pricing of individual items would tell.

Item Cost Versus Inflation

The SoldProd2008 and Sold_Products Tables

Individual item sold price at the time of transaction are contained in the *pp_soldprod2008* and *pp_sold_products* tables in the *Sales* database. These tables have identical schemas, and it appears that the intent of their designers was that there would be a new table for every year of activity. Thus the sales for 2008 are all contained in *pp_soldprod2008*. However, sometime later it appears the designers began to lump all activity together in one file, *pp_sold_products*. All years from 2009 to 2017 are included here. Since data in both these tables were item specific, we were able to do separate analyses on consignment, and new items. Figure 12 gives graphic results for both consignment and new items, figure 13 gives graphic results for consignment items only, and figure 14 gives results for new items only.

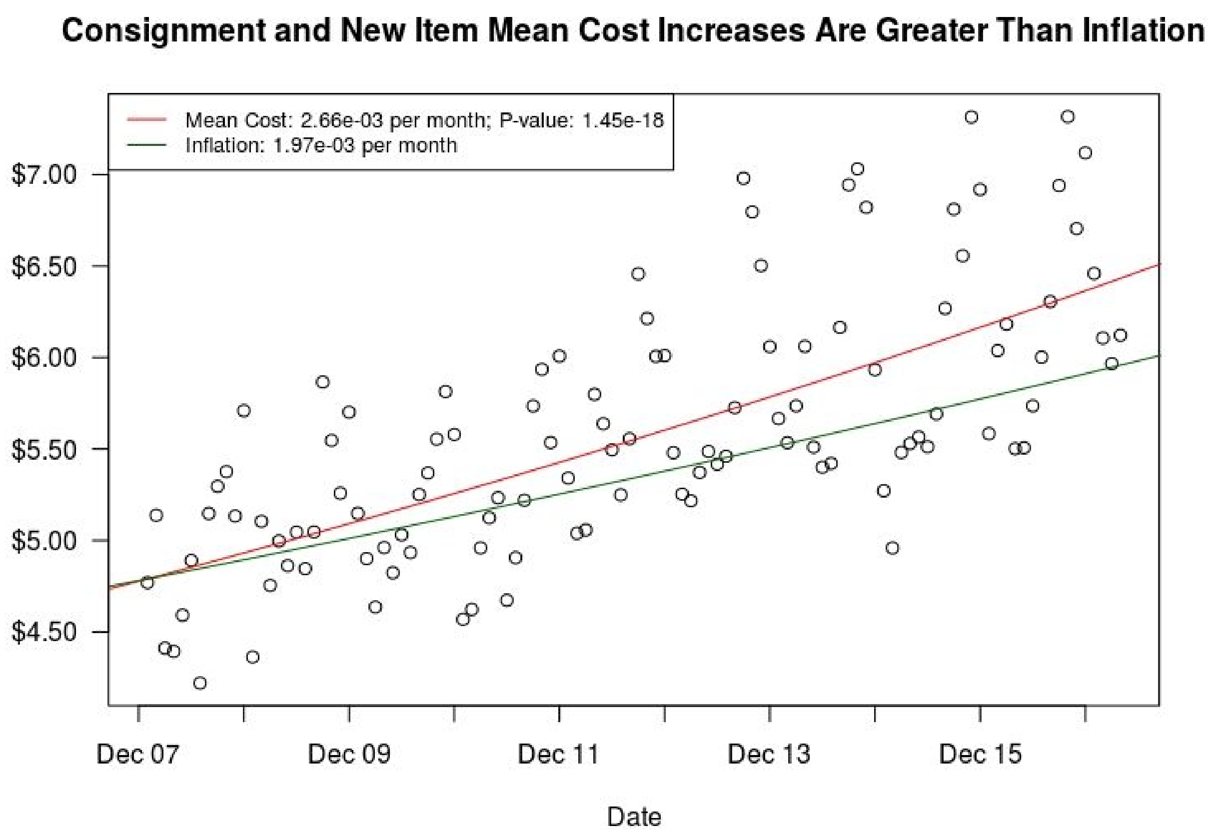

Figure 12: Consignment and new item price trend (red), and inflation (green)

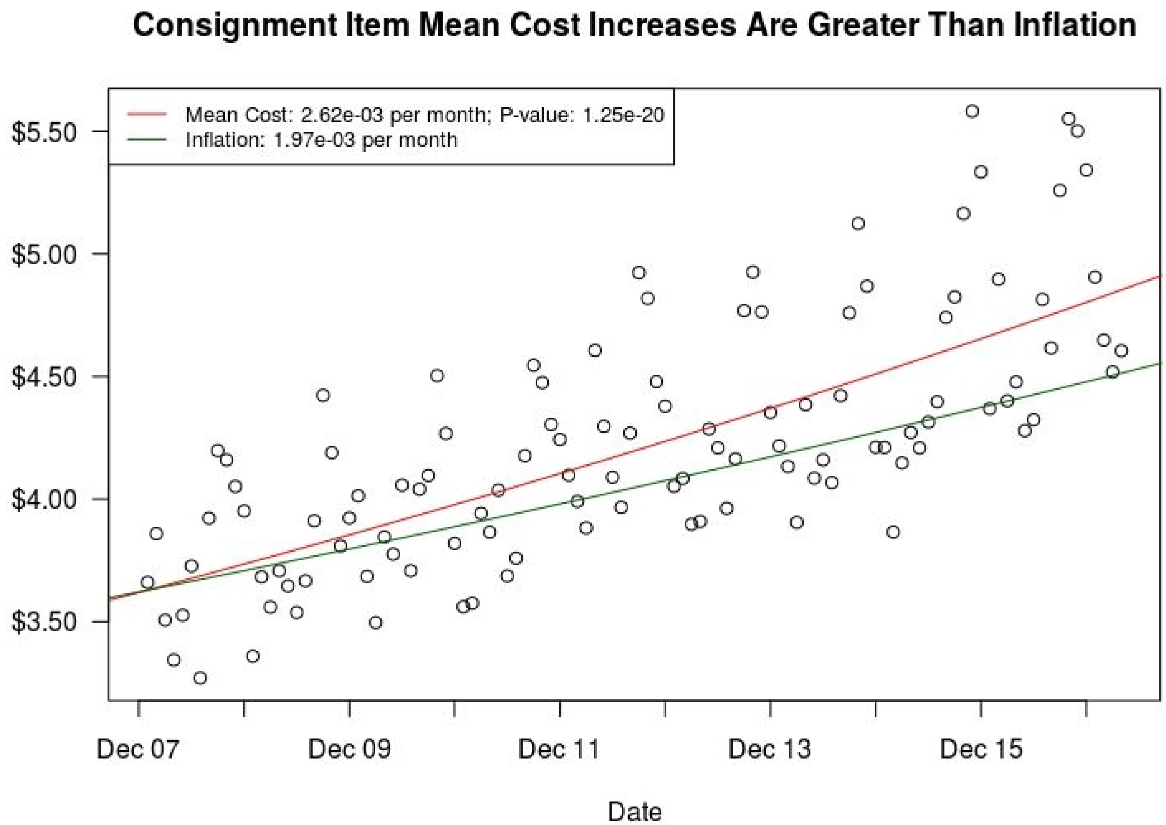

Figure 13: Consignment only item price trend (red), and inflation (green)

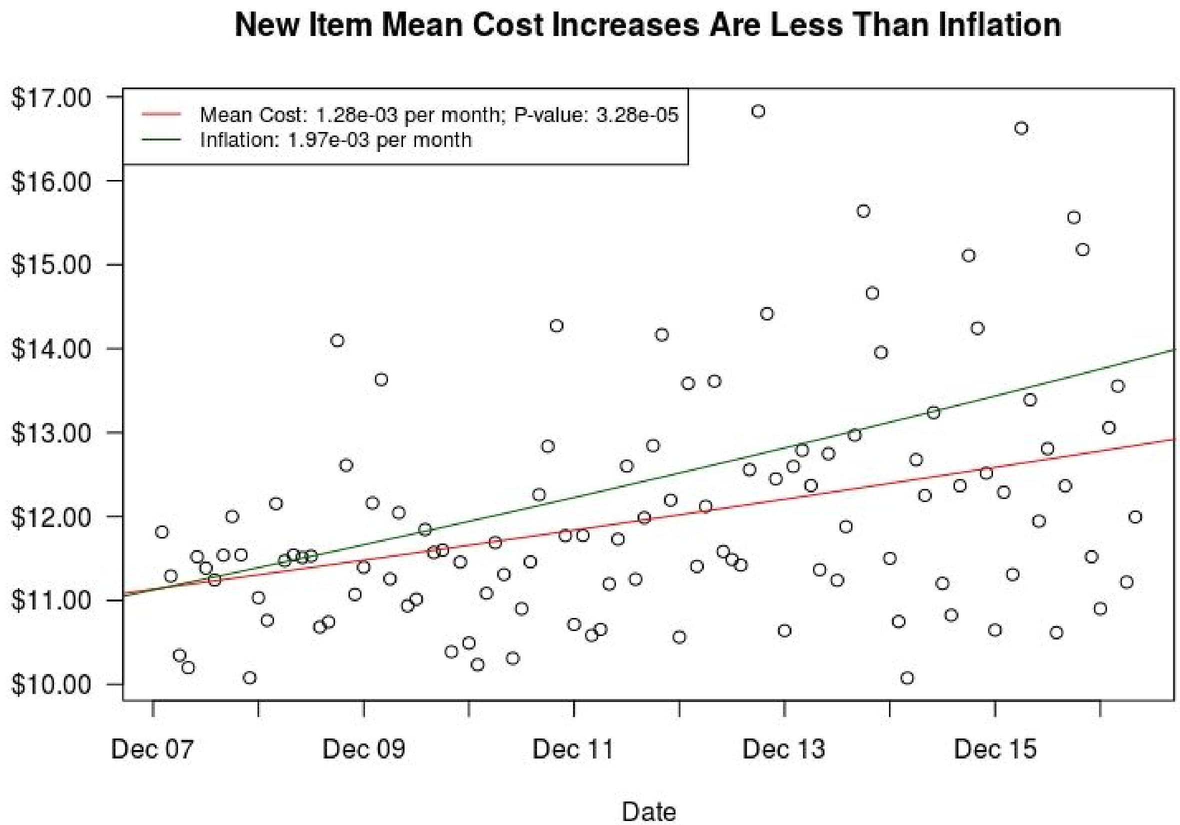

Figure 14: New only item price trend (red), and inflation (green)

Interpretation of Results

The results of this analysis were somewhat embarrassing. We had previously announced that the pricing of consignment items by the store had not kept pace with inflation, yet these results show exactly the opposite. For both new and consignment items together, we see a statistically significant trend that pricing has exceeded inflation. For consignment items only, this trend is confirmed. However, for new items only the trend is reversed: The analysis shows pricing of new items is not keeping pace with inflation. At first, we thought that the results with regard to new items were also statistically significant due to the low p-value of the trendline. This, however, is an illusion. The trendline for new items only appears significant when inflation is not taken into account. Once inflation is factored out (as in our discussion about CPI adjustment earlier in this report), then the price changes with regard to new items are not significant. *A null hypothesis that new item pricing is not* *different than the CPI adjustment cannot be rejected.* With or without CPI adjustment, the higher than inflation changes for the price of consignment items *are* statistically significant. We can state in a statistically significant way that the average, or mean price of consignment items has increased faster than inflation in the time period between January 2007 and April 2017. It appears that when considered together, the higher price of consignment items carries new items along with it. Thus the statistically significant result when the two are considered together, as in figure 12.

These findings are now at odds with the assertion of our *Kids on 45th* customer that consignment items at the store are currently "dirt cheap." Were they even more cheap back in 2007, and is the store not even now caught up? And how can we explain the results of average price of a consignment exceeding inflation while average balance due does not? There really seems to be one possible explanation, and that is that customers are, on average, *buying fewer consignment items per transaction*. We next sought to find evidence in the data to support this theory.

Item Count Purchased per Transaction over Time

Massaging the Data to Obtain Average Items Purchased per Transaction

For this analysis, we organized the *pp_soldprod2008* and *pp_sold_products* tables by customer ID and sold date. In this way, we assume that all purchases by one customer on one day count as one transaction. For the purposes of counting the number of items purchased per transaction, we discount as unimportant those occasions (if they occur) when a customer visited the store twice on the same day, and made purchases each time. We cut transactions ordered this way by month, and then averaged (or took the mean) of the result. We then graphed the result against time. Figure 15 shows the result of this exercise for both consignment and new items, figure 16 shows the result for consignment items only, and figure 17 shows the result for new items only.

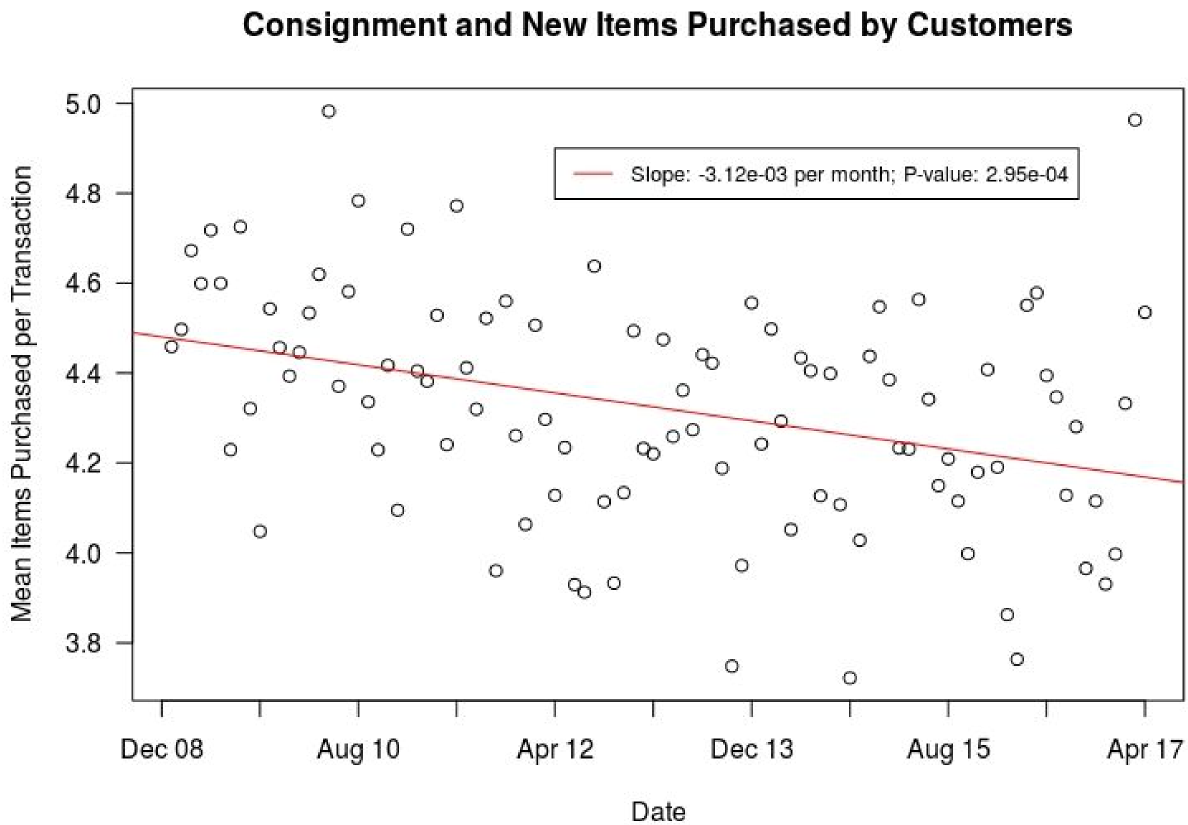

Figure 15: Consignment and new items - mean items purchased per transaction

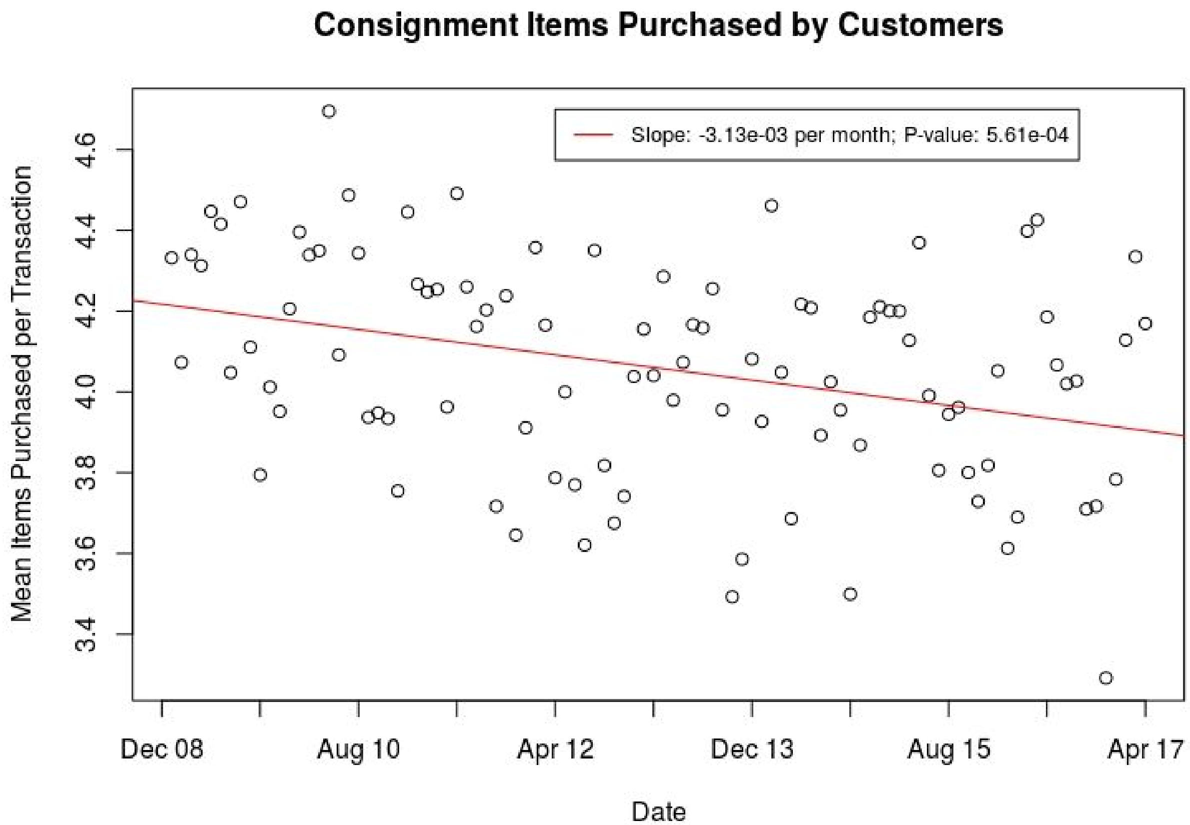

Figure 16: Consignment items only - mean items purchased per transaction

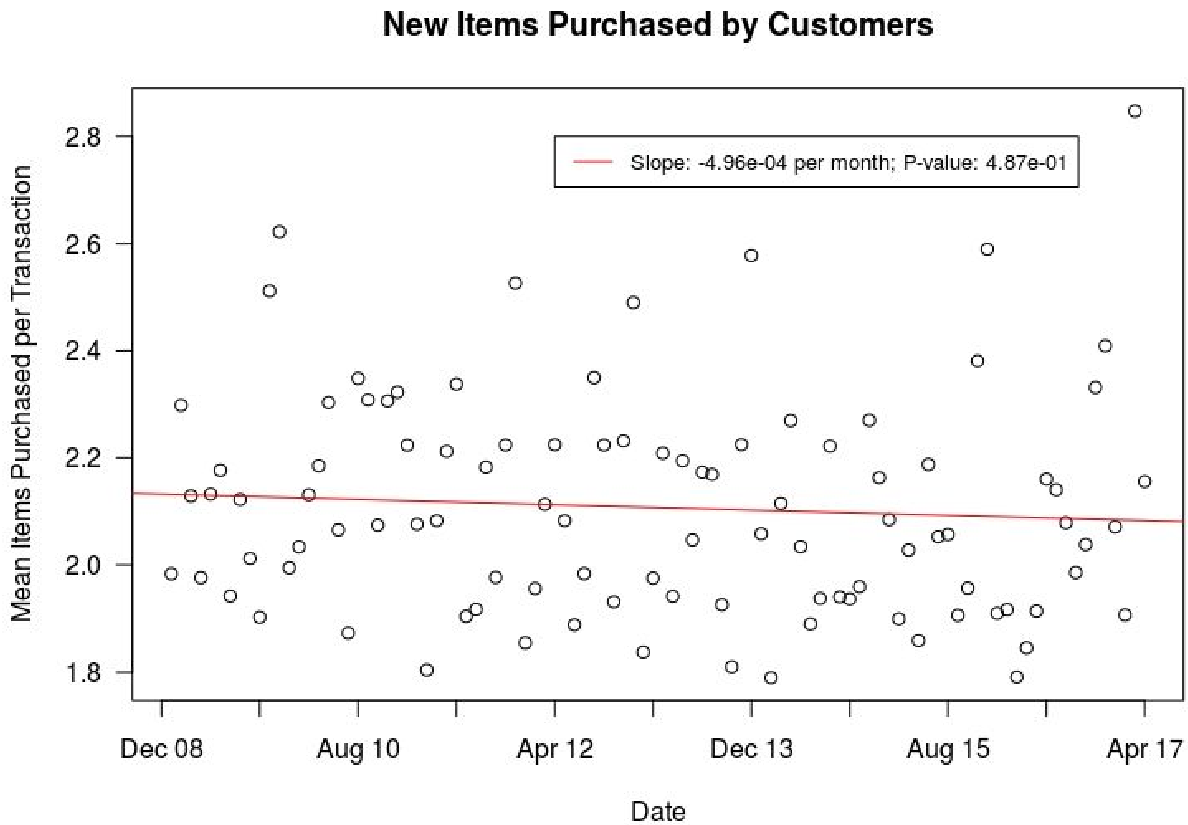

Figure 17: New items only - mean items purchased per transaction

Interpretation of Results

Considering our previous results, this was not really a shock. The number of all items purchased per transaction declined from about 4.5 to less than 4.2 from January 2007 to April 2017. The number of consignment items purchased per transaction declined from about 4.2 per transaction to 3.8. These results are statistically significant. However, considering only new items, the result is not significant. The average number of new items purchased per transaction has remained steady at about 2.1 during the 2007 to 2017 time period. Again, as in our previous results, it appears a decline in the number of consignment items sold has “carried along” transactions in which both consignment and new items were purchased together.

The decline in the number of consignment items purchased per transaction may be a somewhat depressing result for a business that specializes in these types of transactions. However, this fact fully accounts for previous finding that average balance due has not kept pace with inflation, while average consignment item price has exceeded inflation.

We were left hypothesizing what might account for the declining interest in the purchase of consignment items from *Kids on 45th*. These are the theories:

- Customers are buying fewer consignment items because they cost too much. This may be a result of our findings that pricing of these items by the store has exceeded inflation, but conflicts with the idea that items - today, at least - are “dirt cheap”
- Customers are valuing used items less now than they did more than ten years ago. Do they now prefer new items due to the greater affluence afforded by an improving economy? If so, where are they buying these items if not *Kids on 45th*?
- Demographic changes in the Wallingford neighborhood where the store is located have resulted in different customer behavior

We have not been able - as of this writing - to find anything in the legacy sales data to support any of these conclusions. We doubt there is anything in the data that can support or refute these theories. The trend of declining sales, though - although statistically significant - is not so striking that a salesperson employed for the whole period might be aware of it. Perhaps he or she might have a vague feeling that consignment items were not selling as well as they were before. The question of the reason for the trend remains open for further analysis, perhaps using data not available to us now.

Seasonality of Sales

How Much Do Prices Vary Seasonally?

A glance at any of the legacy sales data from *Kids on 45th* will show the tell-tale signs of seasonality in the data. Both the previous, and current owners of the enterprise are well aware of seasonal trends in demand for consignment, and new items. The store's ability to price items more, or less aggressively is heavily tied to this seasonality. For example, during back to school months, and the holiday run-up, the demand for merchandise from the store - for both consignment and new - is noticeably greater. In the winter and spring - without a driver such as new school year, or gift-giving season - the demand for consignment merchandise is noticeably less.

But how much more or less? Our research sought to quantify the ability of the store to adjust pricing monthly. For this we used the *time series creation* (see *ts*), and the *time series decomposition* (see *decomp*) functions of the R programming language to break sales data into its constituent components. For example, with this functionality we were able to categorize separate quantifiable components of sales data into seasonal, trend and random components.

Seasonality of Consignment Item Sales

We performed this work for consignment and new items together, consignment items alone, and new items alone. For the purposes of this report, we regard only the results for consignment, and new items considered alone to be significant. This was our first use of the inflation normalized currency fields that we described earlier in this report. For example, when an inflation component is removed from pricing, the sale pricing of new items in particular revert to what is known as a *stationary time series*. A stationary time series has important properties beyond the scope of this report. Our research produced a considerable number of visualizations. In figure 18, we show the time-series decomposition of consignment items only. In figure 19, we show in tabular format the pricing adjustments for these items. In figure 20, we show the pricing trend for consignment items with seasonality and randomness removed.

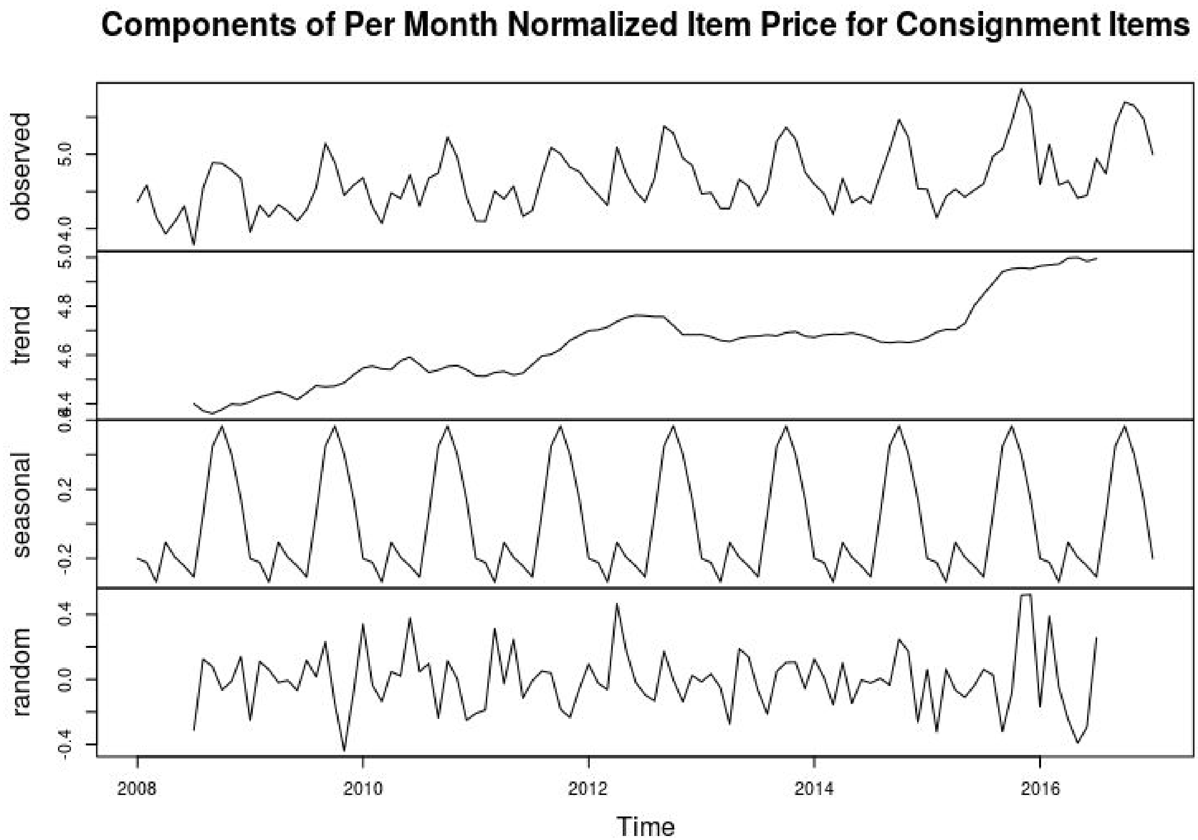

Figure 18: Time-series decomposition of consignment item sale prices

Table of Seasonal Variances for Consignment Item Sales

	Month	Adjustment
1	January	$-0.20
2	February	$-0.22
3	March	$-0.34
4	April	$-0.11
5	May	$-0.19
6	June	$-0.24
7	July	$-0.31
8	August	$0.05
9	September	$0.45
10	October	$0.57
11	November	$0.40
12	December	$0.14

Figure 19: Table of seasonal adjustment for consignment items, by month

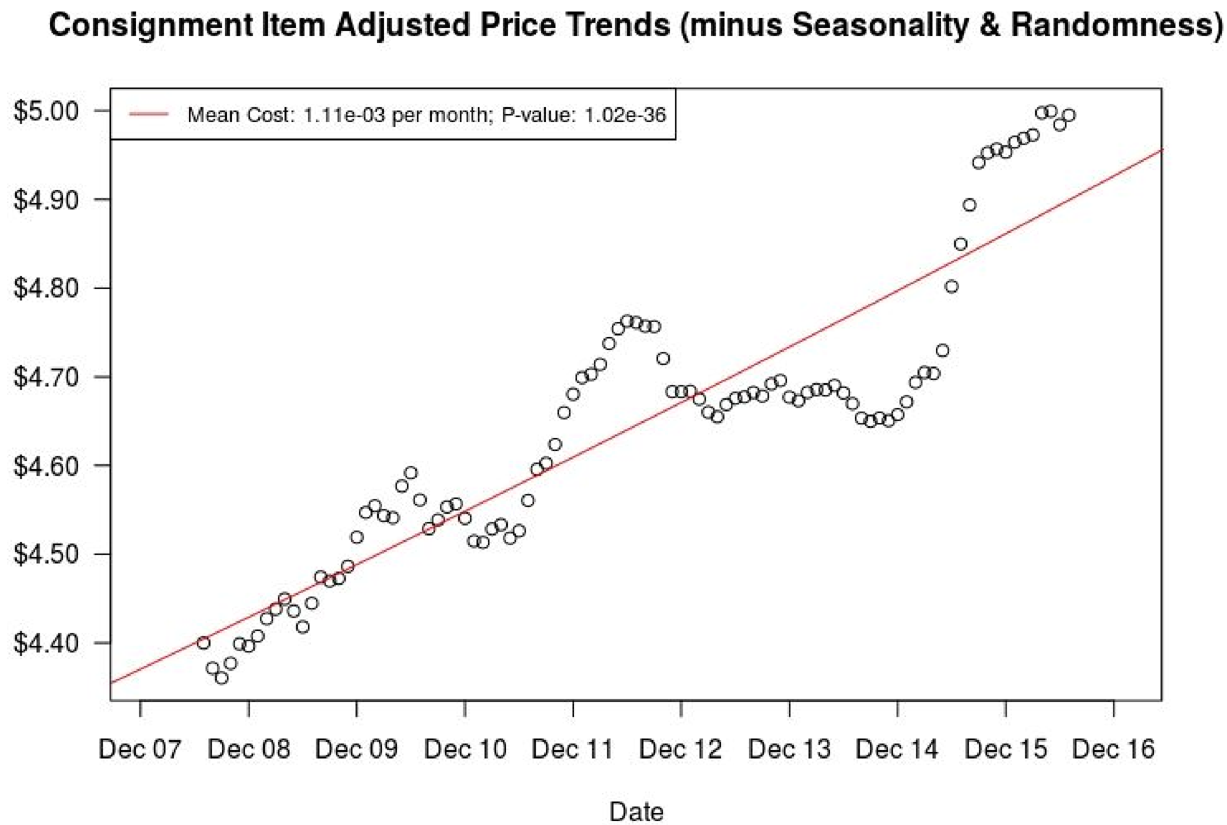

Figure 20: Trend of the price for consignment items, by month, with seasonality and randomness removed

Interpretation of Consignment Item Seasonality

We can definitely see the effect of annual seasonality. In figure 18, as expected, higher sales prices occur in the late summer and early fall. The random pricing subpanel gives the impression of exactly what it is: random. However, in the trend panel - with the seasonal and random components removed - we do see the statistically significant uptrend in consignment item prices noted earlier. To emphasize, these trends are from the inflation-normalized transaction values. The pricing of consignment items has indeed climbed faster than the CPI for the period January 2007 to April 2017.

Figure 19 is a powerful result, as it gives a table of the exact amounts over the more than ten years that can be attributed to seasonality, given by month. We see positive premiums for the sale of consignment items for the last five months of the year, with the significantly higher premiums coming in September, October and November. In the first seven months of the year, we see negative pricing seasonality. April seems to have a slightly lower discount, and that may be due to spring shopping. Still, budget minded consumers might consider winter and spring months a good time to obtain value for consignment items at the store. The enterprise might well be advised to apply more aggressive pricing for these items at back-to-school time, and the holidays. Demand is greatest then.

In figure 20, we see the time series decomposition of consignment item prices, graphed against time, by-month, with seasonality and randomness removed. Again, we emphasize that these are prices that have been normalized to June 2017. We can see a multi-year periodicity to the data when compared to the trendline that time series decomposition has not accounted for. There seems to be a multi-year cycle to consignment item prices that may have a genesis in economic factors. Or it may be due to changes in personnel at the store, and how those persons priced the consignment items placed on offer. We believe, however, that staff changes would account for rather more abrupt changes to consignment item pricing trends. In any event, an analogous multi-year cyclicity is present in the pricing of new items (see figure 23). It is harder then, to explain those by changes in store salespersons, and how items were priced. This is due to new item pricing being more driven by the SRP of manufacturers. It may be that the multi-year cyclicity of consignment item pricing is being driven by some function of new item pricing. At this time, we do not know.

Seasonality of New Item Sales

In figure 21, we show the time-series decomposition of new items only. In figure 22, we show in tabular format the pricing adjustments for these items. In figure 23, we show the pricing trend for new items with seasonality and randomness removed.

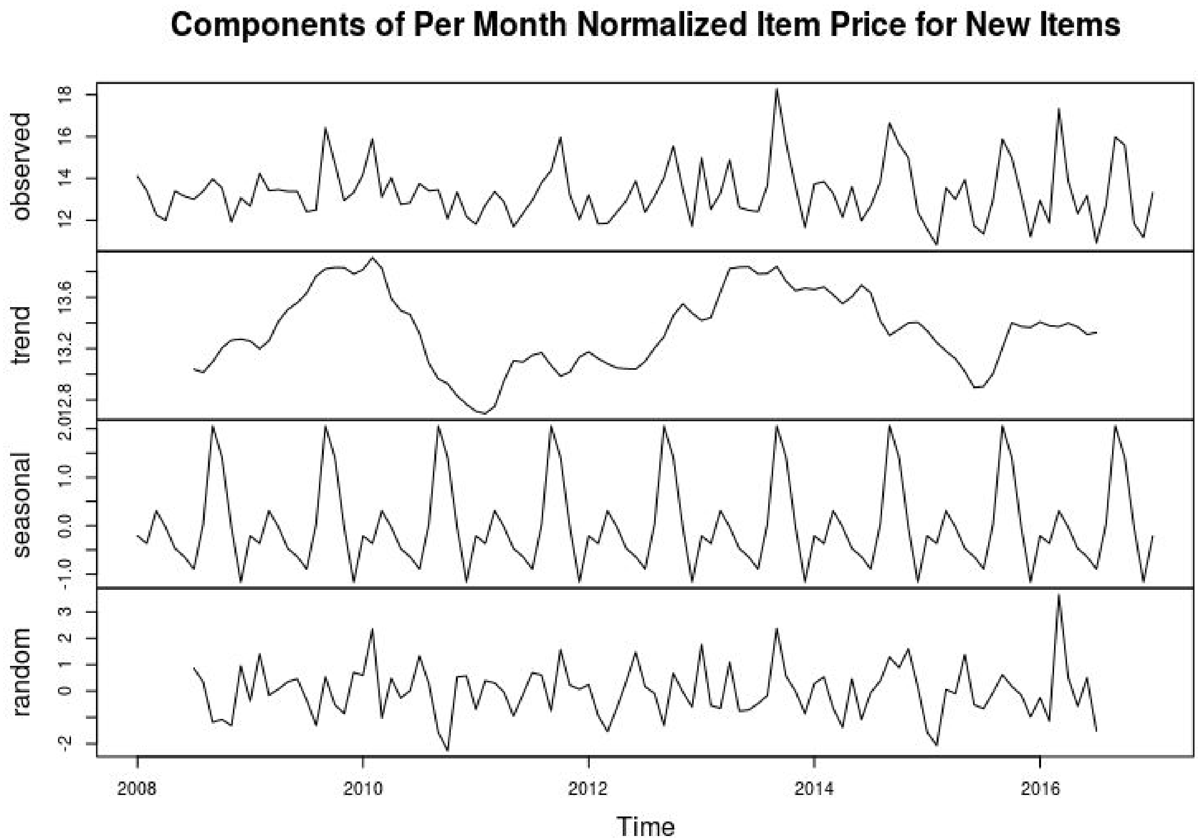

Figure 21: Time-series decomposition of new item sale prices

Table of
Seasonal Variances
for New Item Sales

	Month	Adjustment
1	January	$-0.21
2	February	$-0.37
3	March	$0.31
4	April	$-0.03
5	May	$-0.47
6	June	$-0.64
7	July	$-0.90
8	August	$0.03
9	September	$2.06
10	October	$1.41
11	November	$-0.03
12	December	$-1.16

Figure 22: Table of seasonal adjustment for new items, by month

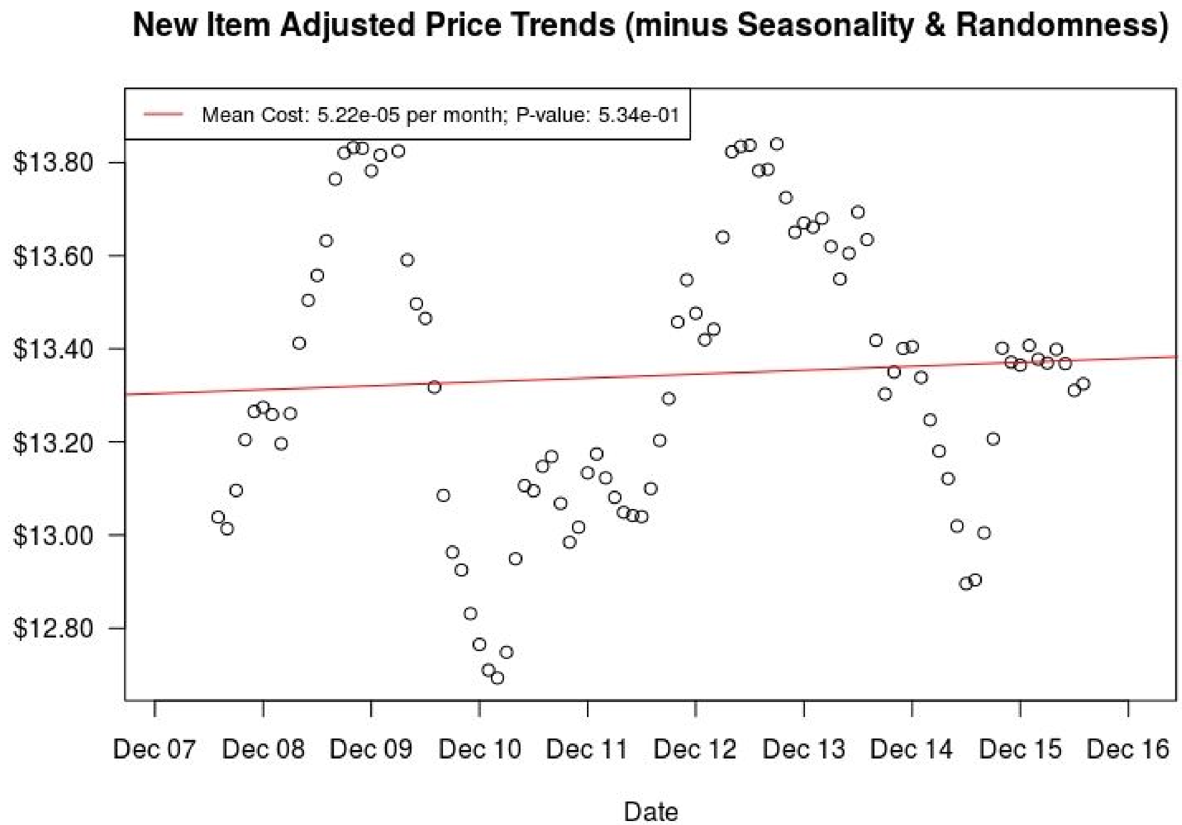

Figure 23: Trend of the price for new items, by month, with seasonality and randomness removed

Interpretation of New Item Seasonality

Again, for new items we see that the effects of annual seasonality are readily apparent. In the analogous panel for that shown for consignment items, we see variability and multi-year cyclicity for the price of new items, but no sustained trend that is greater, or less than the CPI.

Figure 22 shows the same monthly breakdown of new item prices by month, showing premiums and discounts. However, there are important differences. The back-to-school months of September and October show a considerable premium. Thereafter, though, into the holidays, prices show a significant discount. In December, the discount is greater than any other month, and we suspect that this is due to efforts by the store to match other stores' discounting efforts during the holidays. No other month shows a premium but March. In fact, with a premium of greater than $2 in September, and $1 in October, it really becomes a matter of trying to figure out which month with a discount is the least significant. Outside of a small premium in March and August, it seems that July has the highest new item discount, outside of December.

Figure 23 shows what we had mentioned earlier with regard to a multi-year cyclicity in the pricing of new items, once components of seasonality and randomness are removed. An eyeball glance at this multi-year cyclicity seems to show that it is uncorrelated with the same effect seen for consignment item sales. This would seem to indicate that one effect is not a causative of the other. Considering new items only, what causes this multi-year effect? We see local highs and lows usually near the end of the year, with highs occuring in 2009 and 2013, and a low occuring in 2010. However, there seems to be a local low mid-year 2014. The cause of these effects are not currently known, but may be tied to economic forces.

Recommendations

The manner in which the legacy database was organized made making a sophisticated pricing model challenging. However, through a variety of data analysis techniques, we were able to analyze trends over time and come up with the following recommendations.

Higher Number of Items per Transaction

As noted in the analysis above, the average number of items per transactions have gone down through the years. With margins being so low per unit, the high Customer Acquisition Costs (CAC) associated with a retail store mean that the simplest way to increase profitability is to encourage more people to purchase more items whenever they make a purchase. We recommend experimenting with a 'Buy 2 get 20% discount on next purchase' strategy as this has an added benefit of incentivizing customers to create an account. In this manner, reaching out to past customers through email marketing will be able to decrease CAC.

Seasonal Items Have Higher Margins

Our analysis showed that season items, like winter coats and rain boots have the highest margins in the store. As kids grow so fast, these items generally only fit them for one season. However, high quality products are built to last more than one season, therefore many parents are likely to buy these high quality seasonal items from a consignment store like *Kids on 45th.* By increasing the percentage of seasonal items being offered, specifically during back to school season, we anticipate an increase in margins.

Database Design

Under new management, *Kids on 45th* are utilizing new technologies in order to track inventory and sales data. We were advised that they are now using Square as their Point of Sale (PoS) system. Using this knowledge, we were able to design a new database schema for the client. The Entity Relationship Diagram is illustrated below and each table is expanded upon to illustrate its function.

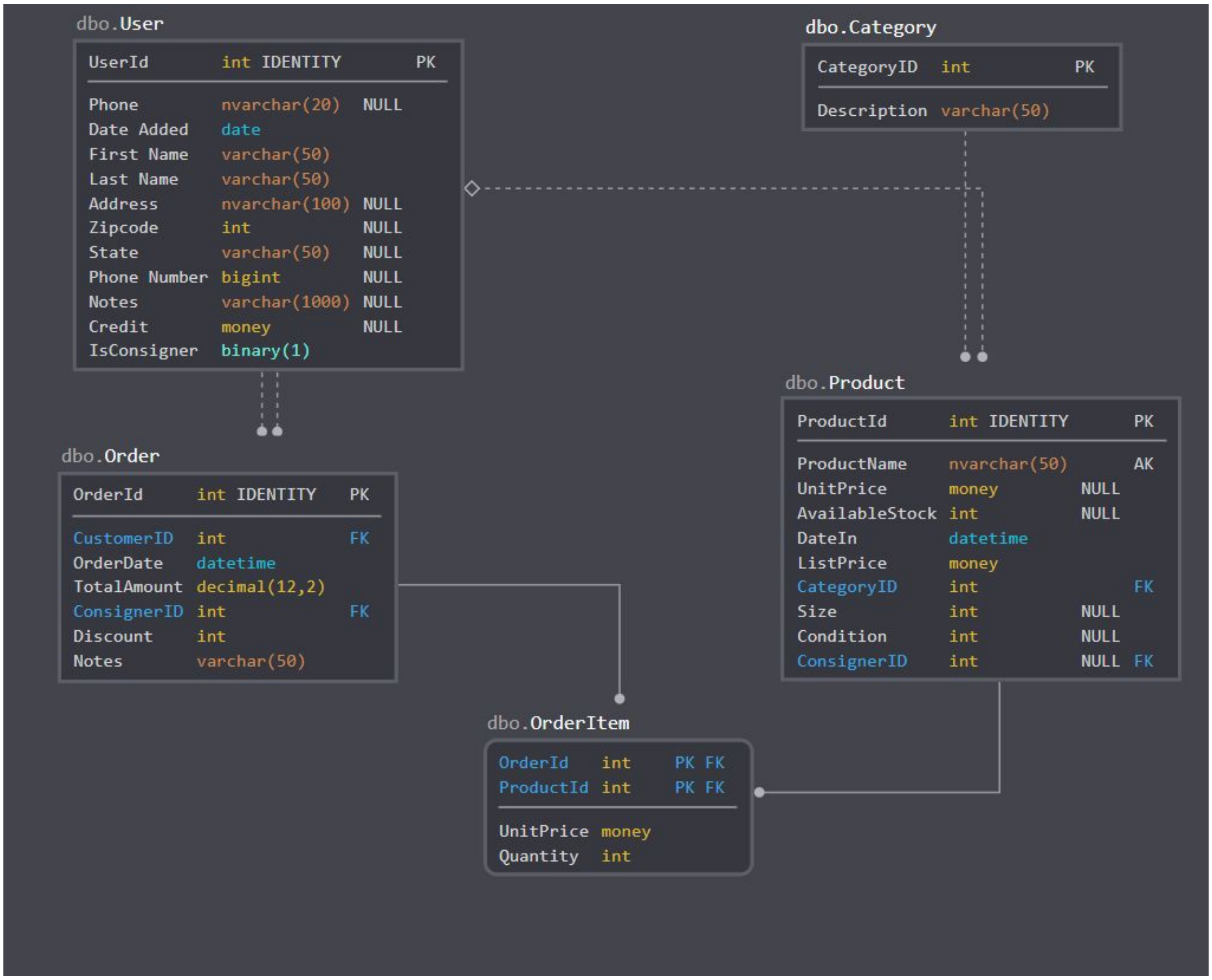

Figure 24: Suggested database design for *Kids on 45th*

Tables

Users

This table stores the information about both customers and consigners. One of the concerns the client had communicated with us was that they were concerned by the amount of new items being bought by consigners, as they enjoyed a 20% discount. In the legacy database, it was tricky to query how many products and what kind of products are being purchased by consigners. By having one database of both customers and consigners, it will be easier to investigate the behavior consigners have as consumers. The rest of the fields are taken from the legacy database and serve as demographic data.

Orders

The Orders table stores information about the orders being placed. It is this order that has two Foreign Keys from the Users table to closely link the relationship between consignors and customers. It is also handy to note any discounts here in order to quickly measure revenue.

Products

This table holds all the information about the present and past inventory and details about the products. There are fields that track details like unit price (Cost) and list price in order to easily measure the Cost of Goods Sold (COGS) which in turn makes profitability calculations easier. The condition field can also be used to designate new items and, if the data capturing process allows, further details, e.g. 'Like New' . The table also connects to a consignorID that allows all consigned products to have an associated consignor. This data is redundant when the product is ordered, as the consignerID is recorded in the orders table as well but this relationship will be useful when measuring goods that are not yet ordered.

Order Item

This is a bridge table to join the Orders and Product tables together and measure the total of every transaction.

Category

Kids on 45th already has a list of categories that they follow. By inputting it into a table and creating a relationship to the product table, it will be easy to query and measure the sales of products by categories.

Conclusions

Due to the difficulties of obtaining meaningful relationships between product details, customer behavior and sales, we were unable to build a sophisticated pricing model for our client. However, by analyzing historical trends and patterns, we believe that there we have identified certain strategies that will enable *Kids on 45th* to increase profitability. In addition, by implementing the new database design that we have proposed, *Kids on 45th* will be able to capture germane data in a more meaningful manner. This will enable our client to seek more detailed insights in the future.

References

R, https://www.r-project.org/

RStudio, https://www.rstudio.com/

Tableau, https://www.tableau.com/

Python, https://www.python.org/

Jupyter, http://jupyter.org/

SQL Database Modeler (SQL DBM), https://sqldbm.com/en/Home/

Microsoft Access, https://products.office.com/en-us/access

Microsoft Office, https://www.office.com/

RDocumentation, datediff,
https://www.rdocumentation.org/packages/SparkR/versions/2.1.2/topics/datediff

Merriam-Webster, Inflation, https://www.merriam-webster.com/dictionary/inflation

City of Seattle, City Budget Office, Historical Consumer Price Index Data,
http://www.seattle.gov/financedepartment/cpi/historical.htm

RDocumentation,splinefun,
https://www.rdocumentation.org/packages/stats/versions/3.4.3/topics/splinefun

RDocumentation,ts,

https://www.rdocumentation.org/packages/stats/versions/3.4.3/topics/ts

RDocumentation,decomp,

https://www.rdocumentation.org/packages/numOSL/versions/1.5/topics/decomp