Emily How, Elissa How, Joe Stannard
Project 1 Process Book

Food for Thought: Graphical Analysis of Grocery Sales Data

March 6th: Project Proposal/Initial Research Questions *(as submitted in Homework 2)*: From the data, we initially were curious about the following questions:

- How sensitive are the households in the dataset to pricing changes week to week?
- Is there a seasonal component present in any of the shopping patterns?
- How effective are in-store and circular features of products (i.e. do people buy more when presented with a display)?
- Do shoppers typically buy frozen pizza, for example, more often at a value store (i.e. Walmart) compared to an upscale grocery (i.e. Whole Foods)?
- Are there differences in product preferences based on geographic location/region of the store?
- When name brand items are on discount, does that affect sales of store brand/private label items?
- When consumers buy frozen pizza on discount or promotion, does this affect sales of other products included the data (i.e. pretzels)?

March 6th: Data Selection and Cleaning: The data was selected after talking with a market research professor in Farmer who has access and experience with many datasets and reliable data sources. Ultimately he pointed us to the Dunnhumby data source site which offers free datasets that are great to use for analytical projects. A retail grocery dataset about four main product categories across multiple store types and states was selected. The raw data did not require any extensive cleaning prior to loading it into Tableau, as it was consistent and high-quality.

March 10th: Meeting with Instructor: The group met with Fadel to go over the data we had selected and the plans we had for some initial visualizations. Expectations for sketches (Homework 3) for these initial visualizations were discussed. Additionally, we were told to make sure that our data was conducive to doing an analysis that could lead to useful decision-making.

March 17th: Graph Sketching *(as submitted in Homework 3):* Plots deemed most effective for representing the various metrics and patterns we aimed to show from the data were sketched on paper prior to exploring the data in Tableau. Heat maps, bar charts, line graphs and geographic maps were deemed most appropriate for visualizing the trends we wanted to emphasize and explore from the data.

- Learnings:
    - Plots that tell a meaningful, compelling story that can lead to actionable results or steps are needed

March 17th: Tableau, Session 1: Also for Homework 3, the data was loaded into Tableau and explored. The data we needed were located on three sheets that were merged based on common variables. ***Although the data did not need any preprocessing prior to being brought into Tableau, it is important to note that additional variables needed for analysis were created during this step of the project. The variables created in Tableau from the original data are as follows:***

- ***Normalized sales variable (percent of weekly sales)***

- ***Binary variable (0,1) for promotion types***
- ***Percent change in price***
- ***Price differential variable***
- ***Location variable (based on latitude and longitude)***

Once these variables were created, two initial plots were drawn up to get an idea of how things would be visualized: one that showed product seasonality and another that showed percent of sales for each product by store segment.

- Learnings:
    - Need a normalized sales variable to compare data across stores of different size and scope (average percent of weekly sales and percent change in this metric)
    - Percent increase in price tells a compelling story for the plot showing sales vs. difference between base and selling price (price sensitivity)

March 28th: Tableau, Session 2 (in class): A categorical variable consisting of the name(s) of promotions was developed. The plot showing aggregate seasonality per product category over the four years was also developed.

- Learnings:
    - How to create, manipulate and include a toggled filter for a visualization

March 29th: Tableau, Session 3: The geographic map plot was developed to show state sales by product.

- Learnings:
    - How to combine city and state into one location variable so Tableau can recognize and plot correctly on maps (avoids Tableau incorrectly assigning cities that exist in multiple states)
    - How to manually geocode cities that Tableau does not recognize (due to small size, for example)

April 2nd – 6th: Tableau videos on a majority of topics and tools were watched throughout this week.

- Learnings:
    - How to manipulate and format filters in additional ways
    - How to create table calculations to show different promotions' percent change in sales from non-promotional weeks
    - How to use LOD expressions to specify only certain data not necessarily included in the visualization
    - How to create and format dashboards in different ways
    - How to apply filters to multiple sheets included in a dashboard
    - How to create dynamic titles that change based on filters
    - How to create storyboards and use them to walk through specific findings

April 6th: Tableau, Session 4: Manually geocoded the latitude and longitude of the 16 cities Tableau did not recognize locations of to plot on a map; Created a variable for percent change in price. Created heat map of average percent of weekly sales by promotion type and product category/subcategory. Created

filters for this plot allowing using to select first a state and then drill down using a drop down to select a specific city relevant to that state. Modified scatter plot to include percent increase in price on x axis rather than raw price differential. Used table calculation to benchmark promotion types against none for product categories across a selected city and state.

April 8th : Tableau, Session 5: A few more visualizations were added and the formatting for others was fine-tuned. The focus was on creating visualizations that really helped the user discover which types of promotions were most effective, and what that effect was. We drilled down to show details at city, segment, category, and subcategory levels. Other tools included allow the user to interact with the visuals in order to explore and drill down based on specific questions.
April 10th: Tableau, Session 6: The dashboard was created; the storyboard feature was used to walk through a story told by the data for Cincinnati and a video was recorded to verbally explain this story.

Additional research questions that were discovered based on the above progress are listed below:

- What promotions are most effective in which store segment types?
- Do the same cities have highest sales for all promotional types or just some?
- How often is each promotion run and does seasonality exist in promotions run?
    - Has this changed over time?

In conclusion for part one of the project, our main takeaways/learnings were which promotions have been most effective for which cities, products, and store segments over the time period the data covers.

****

May 1st: Modifications and Final Details: Minor changes to the dashboard were made based on feedback from Fadel from part one (i.e. color schemes were tweaked and differentiated, vertical texts were realigned to be horizontal, and color was added as an additional encoding to position for the two bottom plots in the dashboard dealing with promotion types).

This process book was updated, published to html and uploaded to Github. Finally, the presentation was prepared based on content from the storyboard video submitted for part one.