Linear Regression Methods in Extracting Insights from Walmart Pharmacy Operations Data through Data Cleaning, Visualization, and Analysis

I. Introduction

Financialization, cultural politics, and securitization are key implementations for American retail giant Walmart (Haiven, 2013) in establishing a global presence as a provider of deeply discounted goods and services for consumers and employment with benefits for its workers. In 2012, Walmart was “the world’s largest private employer, with over 2.2 million workers, as well as its largest retailer, with global sales of $443.8 bilion” (Haiven, 2013, p. 239). In recent times, however, Walmart stock has dropped more than 30% over the course of 2015 while Amazon realized 80% gains in the same time period (Krantz, 2015). Amazon’s ecosystem of cloud computing services, digital media sales, and the “Prime” customer loyalty program was made possible through risk taking and relentless innovation. Perhaps consumers someday may see drones delivering their purchases on the same day at their doorstep. Until Walmart has a better way of delivering products and services to customers in a convenient and timely manner, they are under threat from Amazon’s technological advantage.

The use and analysis of business intelligence data through statistical and forecasting models employed in the field of Data Science may prove to be beneficial in Walmart getting the right product to consumers at the right place and time. Also, the monitoring of staffing levels and customer satisfaction levels may assist in human resources by adjusting the human capital necessary to maximize the return on investment from

employees. Hence, revenue and expense data can generate insights and provide executives decision making tools needed to gain an edge above their competition.

Walmart's division of Health and Wellness is continually a successful enterprise, distributing pharmacy, optical services, and over-the-counter medications at comparatively low cost to patients and consumers. With the passage of the Affordable Care Act of 2009, the pharmacy business has seen reduced reimbursement, cutting into the bottom line for corporate pharmacies. Chain pharmacies typically respond by increasing revenue through tighter inventory control and promotion of secondary services such as immunization and medication therapy management services while reducing labor costs through staff reduction. Currently pharmacy business literature do not quantify whether certain metrics has a greater impact on prescription sales.

Hence, this current study seeks to examine a data set provided by the Walmart Health and Wellness, North Texas division in order to quantify the significance of each available key performance indicators on prescription and over-the-counter sales.

II. Data Acquisition

Operations analysis reports are sent to managerial staff via e-mail on a weekly basis, providing a "report card" and a means for the pharmacy and optical department to assess their performance against other Walmart locations in the same district. This study acquired the operations analysis report from Week 49 (November 30, 2015 - December 6, 2015) of 2015, which may be a particular time of interest given the

transition period between Thanksgiving and the Christmas shopping season. In raw form, the provided data “TempOpsAnalysis.xls” is organized to provide an “Operational Dashboard” for management use in the pharmacy and optical department. For the purposes of this study, additional file conversion and data extraction is required.

III. Data Cleaning

The open source RStudio integrated development environment for R, a statistical computing and graphical programming language was employed to conduct data cleaning prior to using the available and powerful packages available for download for data exploration and analysis. Raw data in the XLS format was converted to comma-separated values format at the website convertio.co/xls-csv/. Next, unnecessary columns and rows were removed, leaving behind pharmacy operations data. Columns were renamed according to proper naming conventions in R (e.g. no spaces within the name or numbers as the first character of the column name). The data frame was converted from factors to numeric values given the quantitative nature of the column values. Data from Internet sources were manually added into the data frame:

(1) Name of town for which the Walmart store is located (Google Maps)

(2) Town population (Wikipedia)

(3) Average household income (Wikipedia)

(4) Average temperature (www.wunderground.com)

(5) Average temperature range (www.wunderground.com)

Additional column rearrangement was made to place the Store Number at the first position, demographic and weather information in the middle, and business data

thereafter. The result is clean data ready for analysis. For records, the write.table function was executed to save the data frame in CSV format.

IV. Data Exploration

Initial data exploration consists of comparing the Customer Satisfaction Scores of the eleven stores against the mean using the ggplot2 package. Store falling below average can benefit from the following analysis in boosting Customer Satisfaction Scores as well as Rx Sales. Prior to constructing linear regression models, the Spearman correlation (r) tests was conducted across the entire dataset, with the exception of “Store Number” and location. Key performance indicators with r-coefficients of greater than 0.5 was considered for inclusion in the linear regression model.

V. Data Analysis

Linear regression modeling takes on the assumption that a collection of independent variables provide a quantitative means of explaining the dependent variable, without “overfitting” the model from the inclusion of too many independent variables.

In fitting the regression model for the Satisfaction Scores, independent variables “Out of Stock Rate,” “Easy Pay Use,” “Discount Card Utilized,” “Partial Fill Rate,” and “OTC Sales Increase” contributed to the most significant model with R-squared as 0.9328 and Adjusted R-squared as 0.8655, p-value = 0.005918 (partial fill rate and OTC sales increase positively affected the Satisfaction Scores). This reveals that patients visiting the store have expectations of leaving with the appropriate product.

In fitting the regression model for the Rx Sales, independent variables “Average Household Income,” “Average Temperature,” “Rx Count,” “Easy Pay Use,” and “Discount Card Loss” contributed to the most significant model with R-squared of 0.932 and Adjusted R-squared of 0.8641, p-value = 0.006073 (Rx count positively affected the Rx Sales). This is common finding that high prescription volume stores benefit in amassing Rx Sales.

VI. Initial Conclusion

Easy pay use where patients have their banking card stored in their files thereby negating the need to furnish it at the point-of-sale transaction was found to be included in both the Customer Satisfaction and Rx Sales regression models. In most recent times, Walmart introduced “Walmart Pay” on December 10, 2015 where customers can pay for their merchandise using their smartphones. This is a great initial step for Walmart as a company to use information technology to not only improve convenience but also capture customer spending behaviors and habits to gain the upper edge in inventory control and staffing management.

References

Brea-Solis, H., Casadesus-Masaneli R., & Grifell-Tatije, E. (2014). Business model evaluation: Quantifying Walmart sources of advantage. *Strategic Entrepreneurship Journal, 9,* 12-33. http://dx/doi/org/10.1002/sej.1190

Haiven, M. (2013). Walmart, financialization, and the cultural politics of securitization. *Cultural Politics, 9*, 239-262. http://dx.doi.org/10.1215/17432197-2346964

Krantz, M. (2015, October 15). How Amazon is eating Walmart's lunch. USA Today. Retrieved from http://www.usatoday.com/story/money/markets/2015/10/14/walmart-amazon-dislocation-stock/73925812/