IBM Watson Analytics Capstone

by

Joshua Troup

A capstone Report submitted in partial fulfillment of the requirements for the degree

Master of Science in Applied Data Analytics

Department of Computer Information Science

Clarion University of Pennsylvania

Clarion, PA

May 2018

Capstone Supervisor: Dr. Jayakumar V Annadatha, Associate Professor

Contents

1. Abstract/Introduction

Raw data once was considered invaluable as it is unstructured with little to no insight to key decision makers. Businesses require the raw data to be transformed in a manner that an analysis can be performed to assist on that businesses success rate. Many industries rely solely on technology to improve their productivity and decrease unnecessary operational costs. Analytical tools have made a breakthrough in the way we do business today. The vast amounts of unstructured data can now be transformed into a cluster of a specific customer type, customer satisfaction, overall efficiency of a particular department within a business, fraud detection, and the list continues to grow for the use of data analytics. Analyzing marketing conditions to pin point what is effective is what is not can save companies millions if not billions of dollars a year. Introduction of new products and services can now be analyzed in more than one way to determine best location or promotion concept delivers the best result. Enhancing business day to day operations is being done by many companies and industries striving to get an edge on their competitors.

"According to a New York Times article, there are about three new McDonald's Opening every day. The main goal is not to have any person more than four minutes away from a branch." (Kannammal, 2016) According to (Kannammal, 2016) "Fast food began after the Second World War. It has come a long way and has achieved unmatchable success and popularity. The success of the fast food chains rest on its marketing strategy and selling Point." Fast food franchises use many advertising techniques to reach customers of all ages by using brand awareness, budgets, and target market with the assistance of online, radio, television, and print strategies. (Kannammal, 2016) Decision makers of these large fast food franchises are highly motivated

individuals interested in learning factual insights to better serve customers and the company as whole. Analytical software can deliver what CEO's seek in their large amounts of data.

IBM Watson Analytics is a powerful cloud based program developed to analyze and visualize large data sets using a series of technology advancements including artificial intelligence, statistics, computer science, and mathematics. IBM started Watson as a follow-on project to DeepBlue which was the computer program that defeated world chess champion Gary Kasparov.
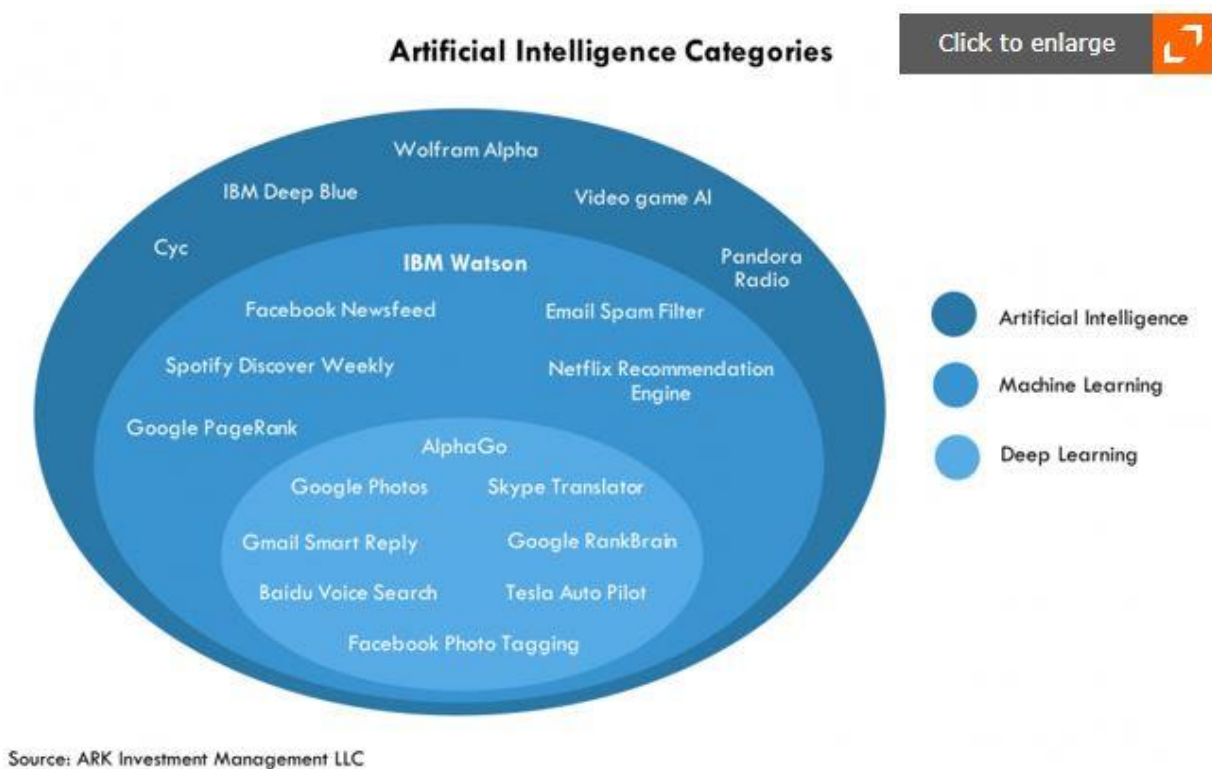

Fig 1

Fig 1 displays the three artificial intelligence categories with real world examples inside the blue circles. "As a quick refresher, artificial intelligence can be divided into three categories, as shown above. The first category is AI itself, defined broadly and covering all possible

approaches to simulating intelligence. A subset of artificial intelligence is machine learning

(ML), which uses data and experience automatically to tune algorithms. Finally, a subset of ML

is deep learning (NYSE:DL). Deep learning uses brain inspired algorithms – neural networks –

to simulate the learning process." (Seeking Alpha, 2017) Watson is trained by data as the image
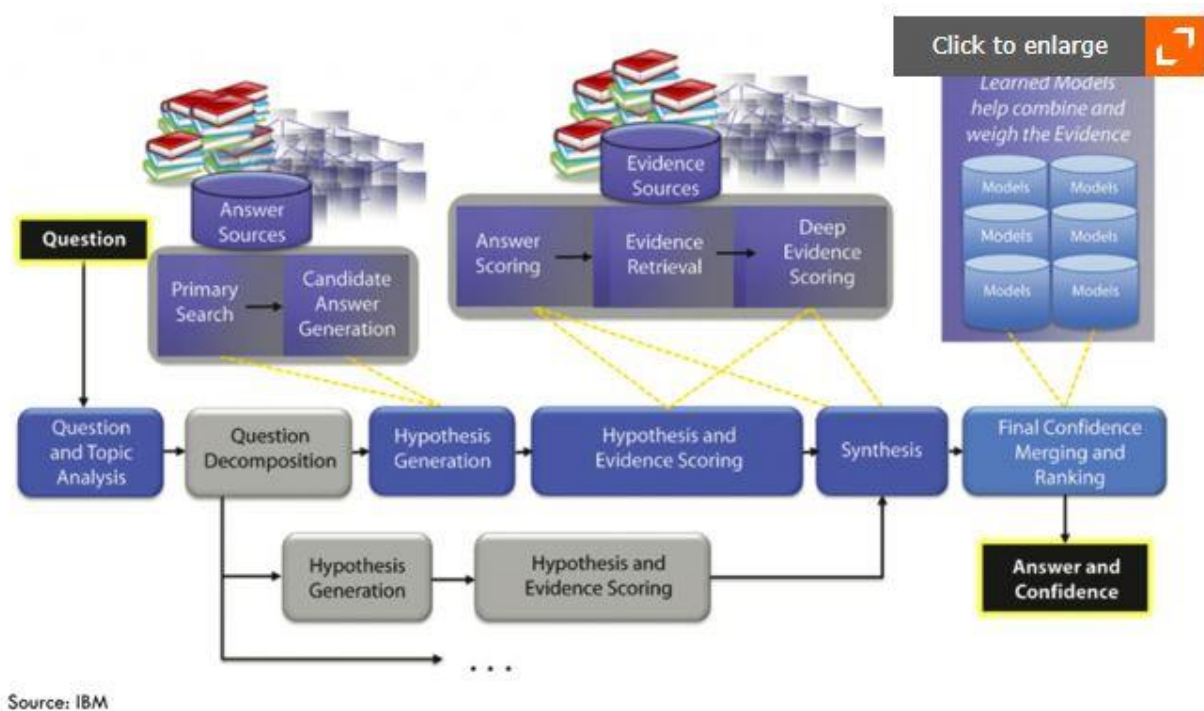
in Fig B illustrates the process.

Fig 2

The processing engine attempts to dissect the questions into words in Fig 2, seeking

correlations between words and isolating the subject of the question. Search analysis is

performed by the AI engine sifting through unstructured data in a manner similar to Google

search engine. Watson finally computes a confidence level per answer with the strongest level

selected for the user. "To summarize, IBM Watson is a system that performs open domain

question-answering. It uses multiple AI techniques such as rule based language parsing,

knowledge bases, search, and statistical machine learning. Its strength is that it can interpret complex queries expressed in natural language, consult many data sources, generate many possible answers, score them based on evidence, and select answers with good accuracy." (Seeking Alpha, 2017)
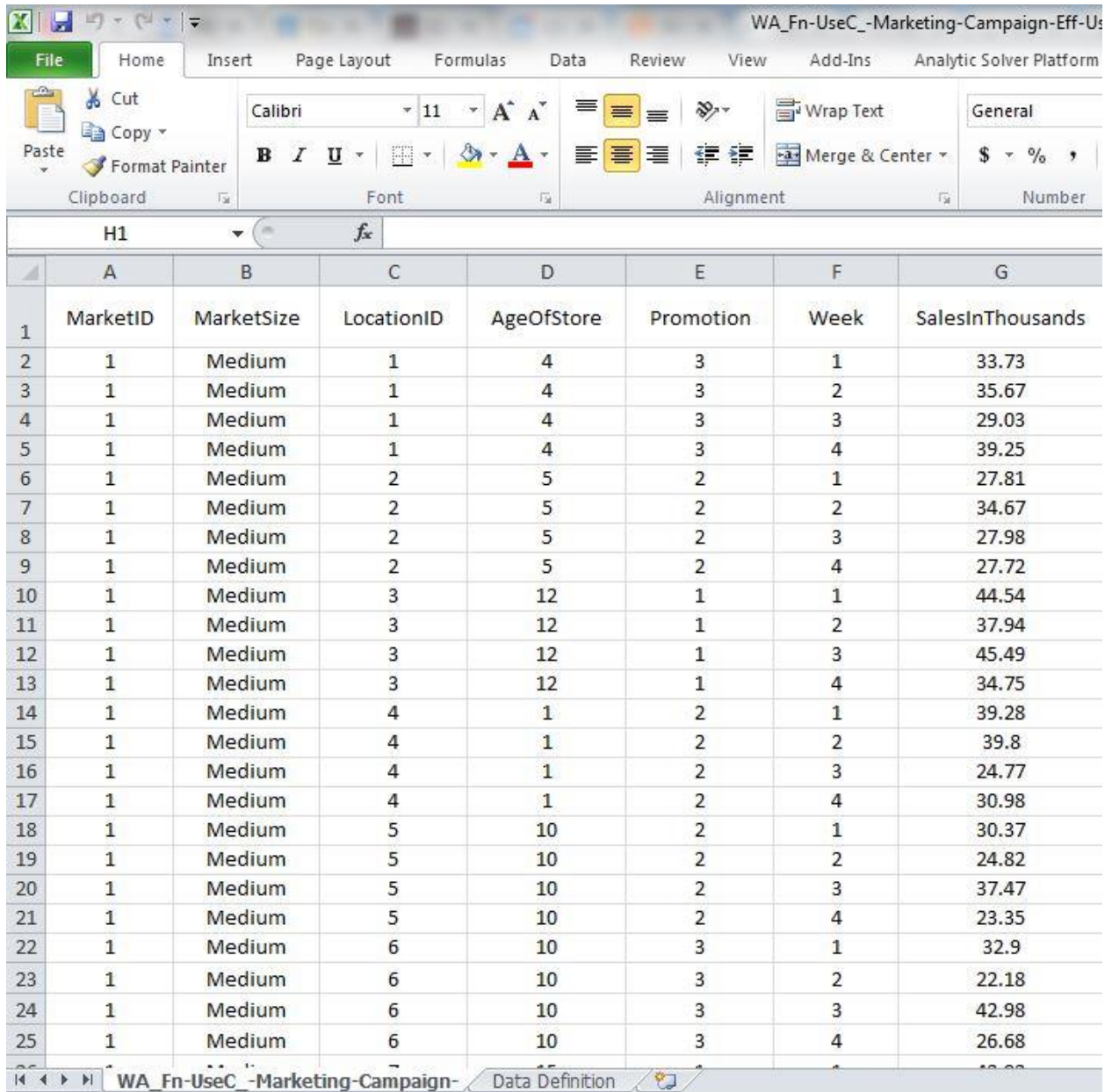
2. Project and Data Selection

Marketing campaigns are a must for every profitable business interested increasing revenue whether it is a brick and mortar establishment or online based company. A large portion of a company's revenue is constantly reinvested back into marketing and R&D. Data analytics can assist in cost savings for a marketing team with hard facts from the data on which campaigns work and which ones do not. I have chosen to analyze a fast food marketing data set attempting to determine the best marketing campaign for a newly introduced product in a variety of locations.

Variable Glossary:

MarketID	Unique identifier for market (1 – 10).
MarketSize	Size of market area by sales (Small, Medium, Large).
LocationID	Unique identifier for store location.
AgeOfStore	Age of store in years (1 – 28).

Promotion	One of three promotions that was tested (1, 2, 3).
Week	One of four weeks when the promotions were run (1 – 4).
SalesInThousands	Sales amount for a specific LocationID, Promotion and week.

3. Exploring the data

Fig 3

The fast food marketing campaign dataset is displayed above in Fig 3 showcasing the seven variables and values.

Fig 3.1

Dataset shown in Fig 3.1 imported into IBM Watson Analytics.

4. Modify Data

4.1 Working with Missing Values/Data Partitioning/Managing Outliers/Transformation of Variables

Fig 4

The Fast Food Marketing dataset does not contain any missing values therefore no modifying to values will be conducted. Fig 4 shows the data having an 84% quality rate. Data partitioning data into training, testing, and validation sizes is not supported by IBM Watson Analytics therefore the dataset including variables will be closely inspected before analysis process. The dataset does not contain outliers or noise as the data is considered clean. Transformation of variables is not necessary in the dataset.
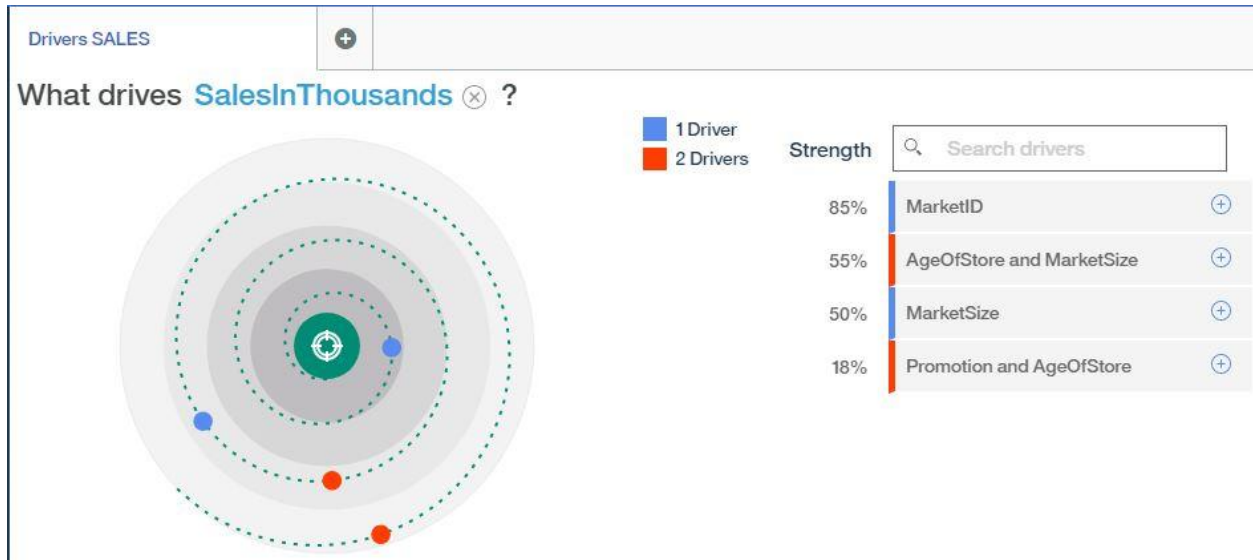
5. Development of Models

Fig 5

Fig 5 displays the top drivers of SalesInThousands. MarketID has the strongest predictor of 85% in correlation with driver 2 AgeOfStore and MarketSize with a predictive strength of 55%. MarketSize, Promotion, and AgeOfStore are the second set of predictive drivers.
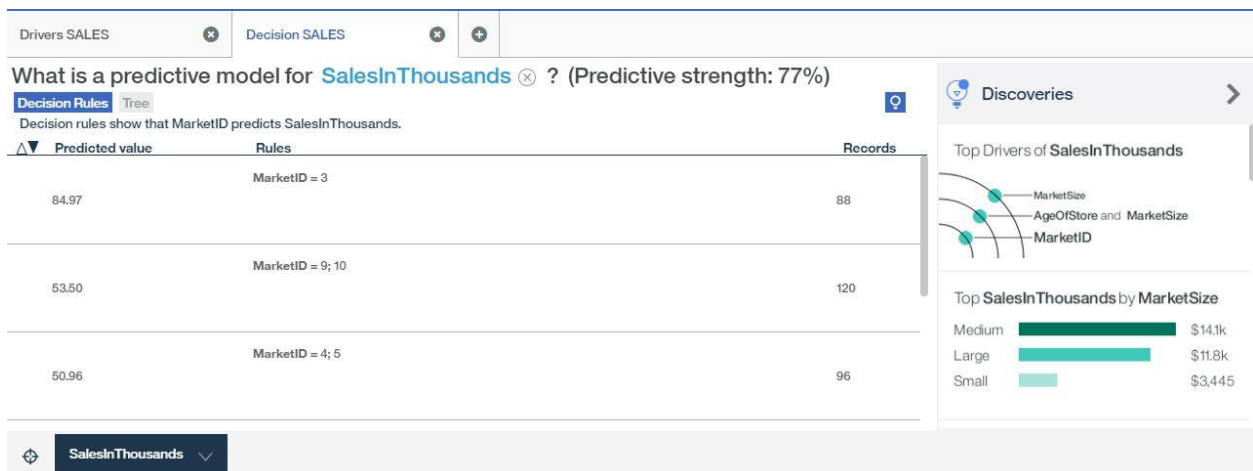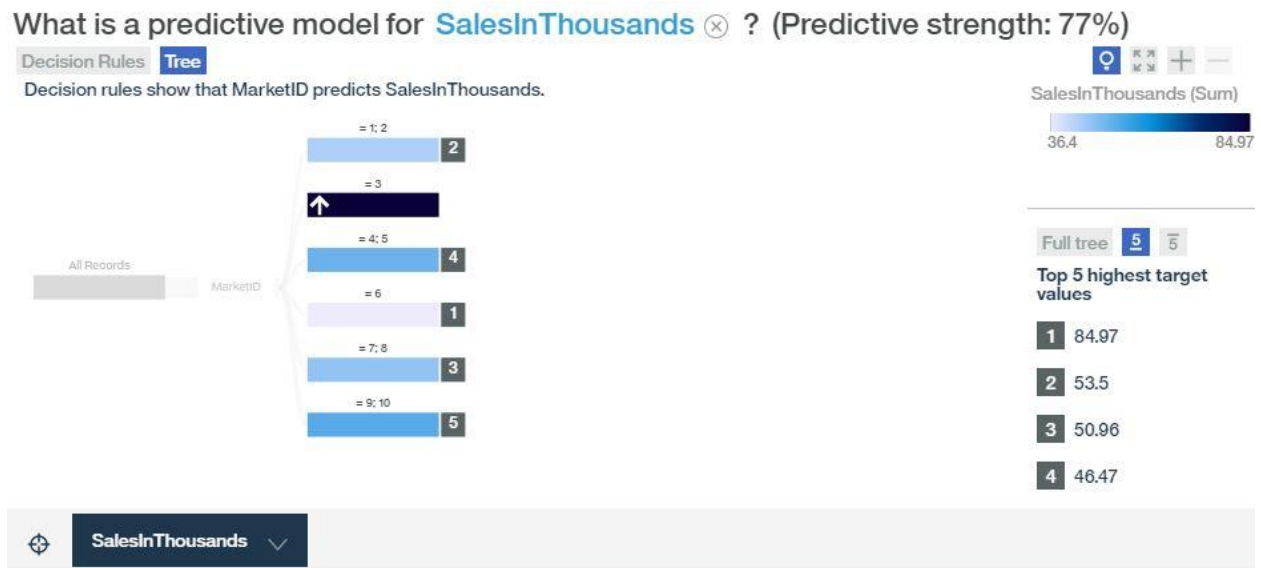

Fig 5.1

Fig 5.2

Fig 5.1 and 5.2 display the decision tree and rules based on SalesInThousands. MarketID

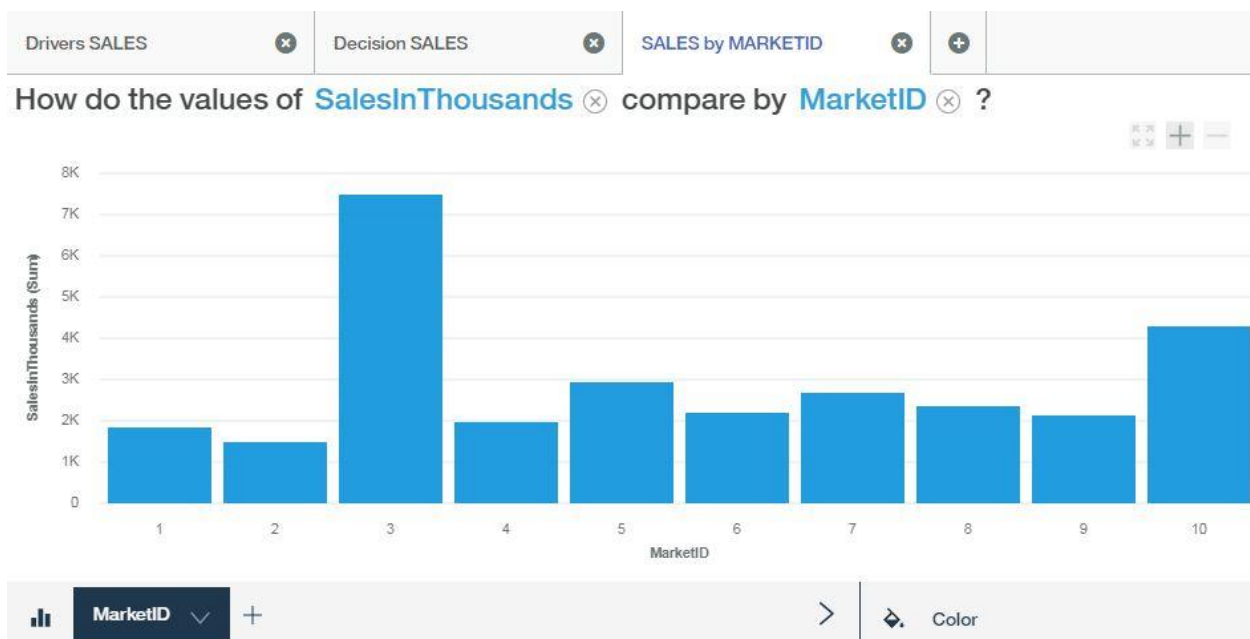three shows to have the greatest impact on SalesInThousands with a target value of 84.97%.

Fig 5.3

The bar chart is Fig 5.3 visualizes the difference in revenue market three produced a value of $7,477.51 compared to the other nine markets.
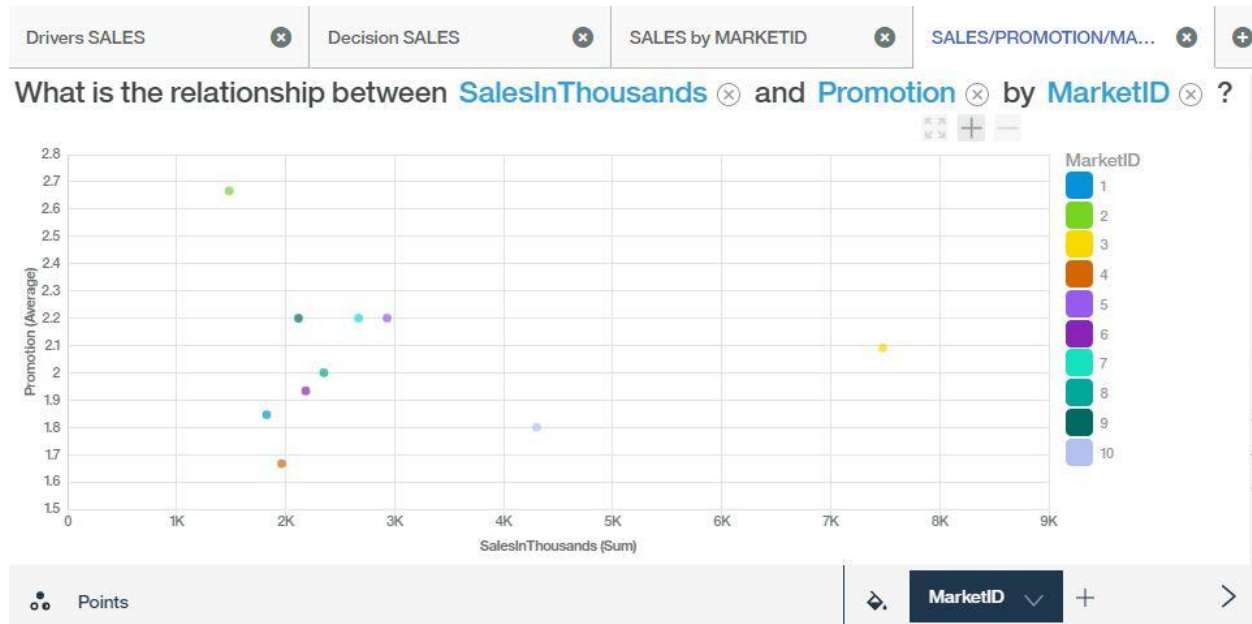

Fig 5.4

The relationship between SalesInThousands, Promotion, and MarketID is shown above in Fig 5.4. Market three with a sales value of $7,477.45 used an average promotion of 2.09 based out of the one, two, or three promotion types used.
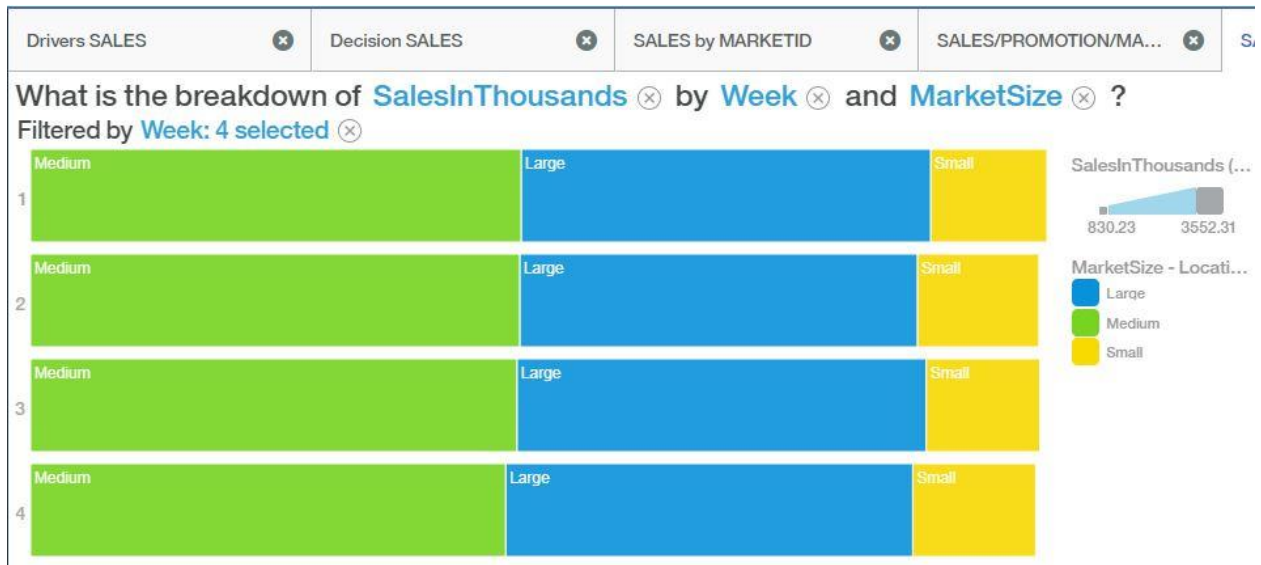
Fig 5.5

SalesInThousands breakdown by Week and MarketSize displays a significant pattern in Fig 5.5. Medium size market is shown to be the largest revenue producing market consistently for four weeks compared to the other two market sizes.
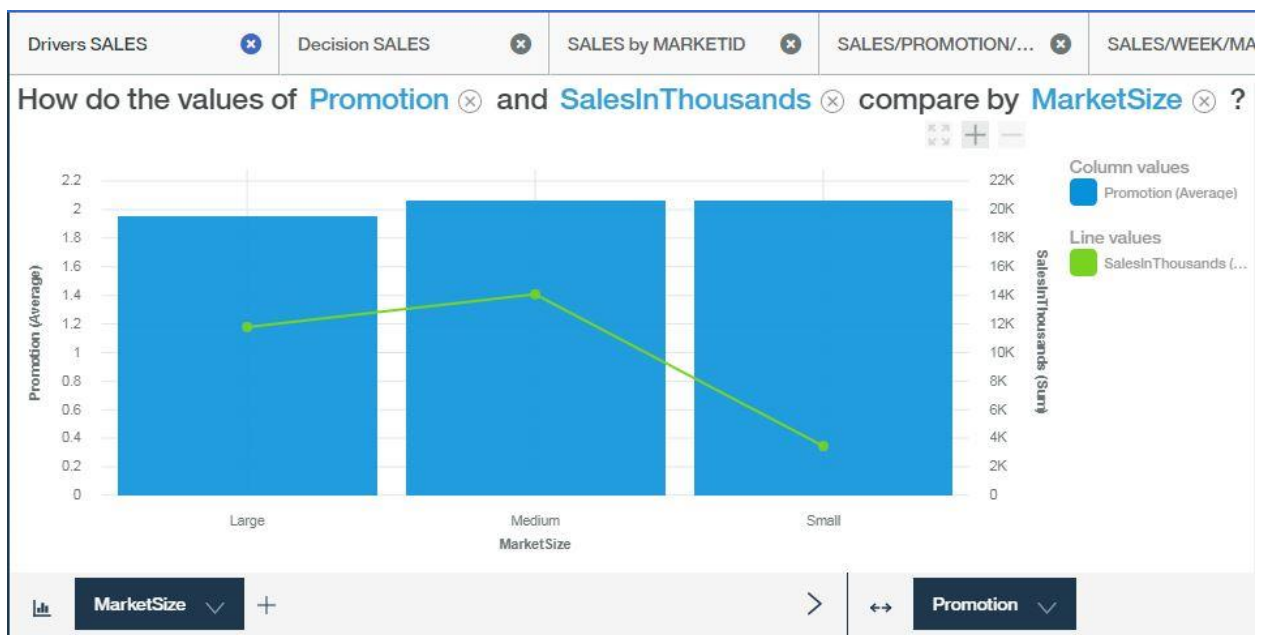

Fig 5.6

Fig 5.6 displays the correlation between Promotion, SalesInThousands, and MarketSize. The bar chart easily showcases the most common type of promotion is based on promotion type two.

Fig 5.7

The correlation between SalesInThousands and AgeOfStore is fairly significant shown in Fig 5.7. The stores with an age of one year to nine years prove to produce more sales with the exception of the second year store. The older the store, the less sales are made.
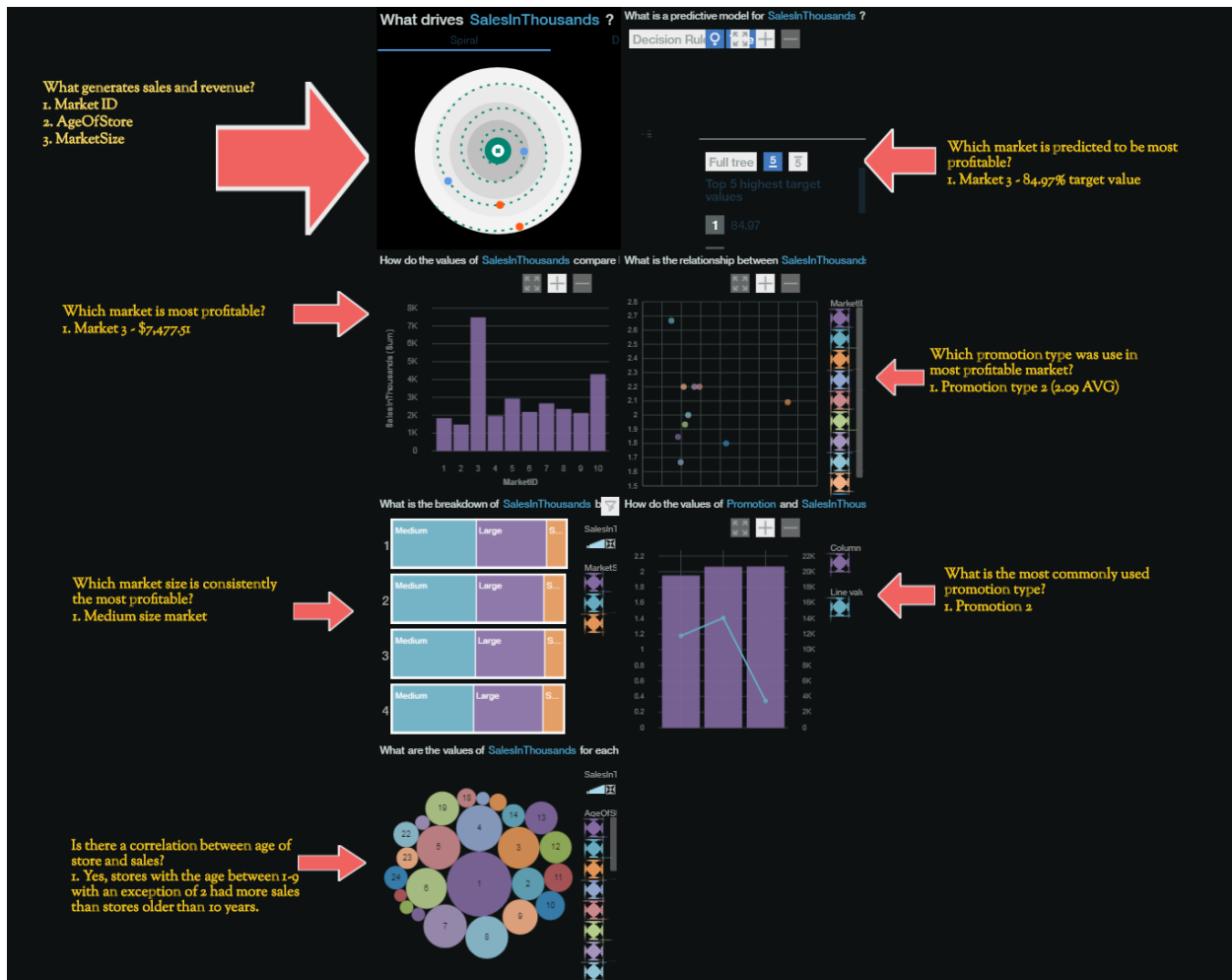
6. Analytical Findings/Insights

Fig 6

Fig 6 displays the dashboard display view of the analysis performed with insights for each model.

7. Conclusions

After discovering many valuable insights within the fast food marketing data set, promotion type two is shown to be the most widely used campaign delivering excellent results in medium size markets. The age of store proves to be ironic in my findings as it does have an impact on sales and marketing. The decision to build more brick and mortar stores in a medium size market similar to marketID three location using a promotion type two would be beneficial to the company. According to (Kannammal, 2016) "Fast food contains several ingredients that are very harmful to health of the consumers. Therefore, as a long term business strategy, fast food companies have to utilize their research and development to make their recipes healthier and filled with nutritional values. This would ensure not only the growth of fast food businesses but also safeguard its consumers from health hazards and lifestyle diseases."

8. References

ARK Investment Management Portfolio strategy, Growth, research analyst, infrastructureARK Invest.cls-1{fill:#024999;}. (2017, July 13). How Much Artificial Intelligence Does IBM Watson Have? Retrieved March 03, 2018, from https://seekingalpha.com/article/4087604-much-artificial-intelligence-ibm-watson

SAMPLE DATA: Marketing Campaign, Promotion Effectiveness – Fast Food Chain. (2015, April 11). Retrieved February 24, 2018, from https://www.ibm.com/communities/analytics/watson-analytics-blog/marketing-campaign-eff-usec_-fastf/

Kannammal, G., & Suvakkin, M. (2016, December). Emerging Marketing Strategies of Fast Food Industry in India. Retrieved April 14, 2018, from

http://www.internationaljournalssrg.org/IJEMS/2016/Special-Issues/NCEBSED/IJEMS-

NCEBSED-P110.pdf