BREAD (BIG DATA RETAIL ANALYSIS AND PRODUCT DISTRIBUTION) MODEL FOR SALES PREDICTION

Riktesh Srivastava, Skyline University College, Sharjah, UAE.

ABSTRACT

UAE retail sector is expected to grow by 5% every year and an estimated retail market to be AED 200 billion (Emirates247, 2016)*, which makes UAE ranked 7th in Global Retail Development Index. Euromonitor and ATKearney* (Hana Ben-Shabat, 2015) *study about UAE retail market, opinions that consumer confidence in UAE has not negatively wedged despite slow economic growth, rather, unexpectedly, resulted in pierce retail competition. In a state of such penetrating race, analytics can be a foremost differentiator for the companies. (Steve Lavalle, 2010) found that the top performing companies are three times more effective than those without analytics, making analytics as a sole competitive differentiator. Analytics lashes the passage from merchant-driven business models to digital models, where every decision is cognizant by data. Another study conducted by* (Brynjolfsson, 2011) *reveals that data-driven businesses have an output and productivity of 5-6% higher than similar companies who do not use data-driven decisions.* (Luckie, 2012) *state that poor data management decisions can cost up to 35% of a businesses operating revenue. In this study, a new technique called BREAD (Big Data Retail Analytics and Product Distribution) is developed for sales prediction for retail scenario. As an experiment, the model was used for sales prediction of ABC Stores (name changed, as requested), based on their 2015 sales data for 16 Item types, divided into 1559 items across 10 stores (a total of 8523 records). Based on the study, the paper gives recommendations to ABC Stores to embed analytics in its predictions.*

Keywords: *Retail Analytics, BREAD, Shopping Basket Analysis, Market Basket Analysis*

1. INTRODUCTION

Almost all retail outlets (within the same retail segment) offer alike products, incorporates similar IT tools and infrastructure and uses almost similar business models. As indicated, the sole differentiator can be analytics with consistent data, which can help retail outlets to make knowledgeable verdicts, confidently influence sales and gain competitive advantage.

Table 1 given below illustrates the types of retailers operating in UAE, market share and use of analytics by them.

Table 1: Retail Outlets (in UAE)

Types of Retail	**Market Share (Complete Retail)**	**Use of Retail Analytics**
Online Retailer(Deals, Electronics, Fashion & Organic Foods) (Euromonitor, 2016)	1%	100%
Grocery Stores (Euromonitor, 2016)	31%	5% Carrefour(Geodashboard) Spinneys (SAP HANA, SAP Business Objects)
Non-Grocery Stores (Electronics, Fashion, G&D) (Euromonitor, 2016)	63%	5%
Miscellaneous	5%	0%

Surprisingly, internet retailing in the UAE accounted for just 1% of total retailing value sales, a considerably lower share in comparison to the international average of 15-20% (Euromonitor, 2016). However, the use of analytics is much higher by internet retailers than by non-internet (traditional) retailers.

(Inna Kolyshkina, 2007) piloted a study for optimal utilization of retail analytics, by focusing on key stages of analytics process. The study also identified the factors for success or failure of analytics. (Adams, 2008), illustrates that collecting the data regarding which products constitute a customer's order is often referred to as "Shopping Basket Analysis [SBA]", used in retail and grocery environments to comprehend customer purchasing inclinations. This analysis is also called Market Basket Analysis [MBA] (Loraine Charlet Annie M.C., 2012).

The present study, assents that retail outlet has previously piloted SBA, and now planning for product wise sales in the outlets. BREAD Model uses the product distribution based on the information gathered from SBA, segments the data into Item_Identifier and Outlet_Identifier, thereby analyzing Item_Outlet_Sales.

Figure 1 illustrates the operational stages for BREAD Model.

Stage 1: Categorize the retail outlets based on outlet features.
Stage 2: Identify the product types in each of these outlets
Stage 3: Categorize the items
Stage 4: Calculate the Item Outlet Sales

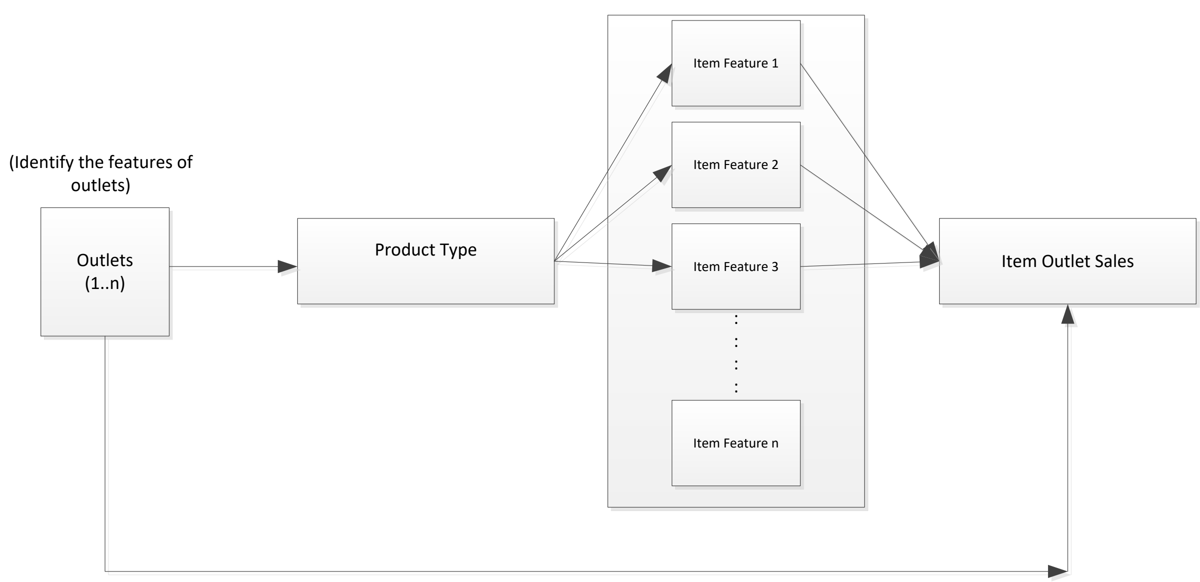

Figure 1: BREAD Model

The mathematical derivation for item *i* at Outlet *j* can thus be represented as:

$$Item\_Outlet\_Sales_{i,j} = Product\_types_i * Item\_types_{i,j}$$

Complete mapping of the outcome variable with reference to two input variables is depicted in Figure 2:

Figure 2: Stages of Mathematical derivation

METHODOLOGY

Data is obtained from a retail company, ABC Stores. Their primary focus is to:

1. Understand the sale of products in individual stores,
2. Identify the relationship between item visibility (shelf space) with sale, and
3. Evaluate the monetary value generated from different Product_types and Item_types.

Initially, the complete data set was divided, based on number of outlets. Then, various items retailed by ABC Stores were grouped into Product types. These categories are further subdivided into various item types. BREAD model was then conceded to classify the sale patterns. Appropriate recommendations were made for the company to move ahead.

1.1 PREPROCESSING OF DATA

Most of the transaction (sales) data at ABC Stores is organized based on Item_Identifier (Unique identification of each product). Also, some of the Outlet_Size and Item_Visibility data was missing, when we evaluated the summary statistics of the 8523 records. Thus a major restructuring of data was carried to identify the NA's in the data set and filling it with required inputs.

Second restructuring was carried to classify the Outlet Numbers. As R programming converts the outlet numbers labeled as OUTXXX to a factor, restructuring was required to convert and store the values as Characters in the CSV format.

Also, in order to comprehend the purchasing behavior as well as inherent purchase patterns, it was vital to abstract the items to broader categories. Inferences made at a category level would be much simpler to implement and execute. Hence classification of items into high level categories *(Baking Goods, Breads, Breakfast, Canned, Dairy, Frozen Foods, Fruits and Vegetables, Hard Drinks, Health and Hygiene, Household, Meat, Others, Snack Foods, Soft Drinks, Starchy Foods, Seafood*) was performed.

1.2 OBSERVATION CRITERIA

Based on the observation, item types of 16 product types are classified into 4 board categories:

1) Inactive
2) Cold
3) Warm
4) Active

These classifications are carried for each of the outlets in order to plan for product visibility and assortment planning.

The complete paper is divided into 3 sections. Section 2 explains the BREAD model in detail. Section 3 performs the implementation of BREAD model on the data set. Section 4 provides the recommendations and conclusion.

2. BREAD MODEL

Figure 3 illustrates the generic activity diagram of proposed BREAD model for Item Demand Estimation and Item Visibility for ABC Stores. The comprehensive model comprises three steps as:

Step 1: Assess the Overall Sales and Outlet sales

The step implicates analysis of the overall sales w.r.t store sales, which customs the base of the Observation criteria.

Step 2: Assessment of Sales w.r.t. Item Types [Observation Criteria]

Algorithm in this step assesses the Status of Items sold and includes three stages.

Stage 1: Indentify the Status Range of Sales w.r.t Overall Sales based on criteria mentioned in Table 2.

Table 2: Status Range of Overall Sales

STATUS	Range 1	Range 2
Inactive	5000	10000
Cold	10001	25000
Warm	25001	50000
Active	50001	

Stage 2: Indentify the Status Range of Sales w.r.t Store Sales based on Table 3. It is also used to categorize the status range of Sales w.r.t Store Tier (I, II and III) Sales.

Table 3: Status Range of Store Sales

STATUS	Range 1	Range 2
Inactive	0	5000
Cold	5001	10000
Warm	10001	15000
Active	15001	

Stage 3: Indentify the Range of Sales w.r.t Item_Type (LF, R) based on Table 3.

Step 3: Estimate Optimum Item Visibility (OIV)

Evaluation of OIV depends on Item Visibility Chart (IVC), which is a status range for item visibility for diverse items at ABC stores. ABC stores has positioned the items based on demand from customers or on seasonal requirements. This can be called as Demand function and is demonstrated as:

$$D = \alpha_i * f(S_i, S_{ji})$$

where,

$\alpha_i = Latent\ demand\ for\ product\ i$

$S_i = Space\ allocated\ to\ product\ i\ (as\ mentioned\ in\ Item\ Visibility\ Chart)$

$S_{ji} = Space\ allocated\ to\ product\ j$

Table 4 provides the IVC status range.

Table 4: IVC Status Range

IVC		
Value	**Low**	**High**
Invisible	0	20
Slightly Visible	21	40
Moderately Visible	41	60
Visible	61	80
Highly Visible	81	100
Demanding	101	

There is a display space for the brands in different product categories. Item visibility is the space which is perceptible to the consumer. The optimal item visibility (OIV) is assumed to be based on the gross profit margin.

$$OIV = \left({}^{Annual\ Average\ Sales\ of\ item}\!/_{IV\ of\ item\ in\ all\ outlets} \right)$$

Based on three steps, the flow of BREAD model is as under:

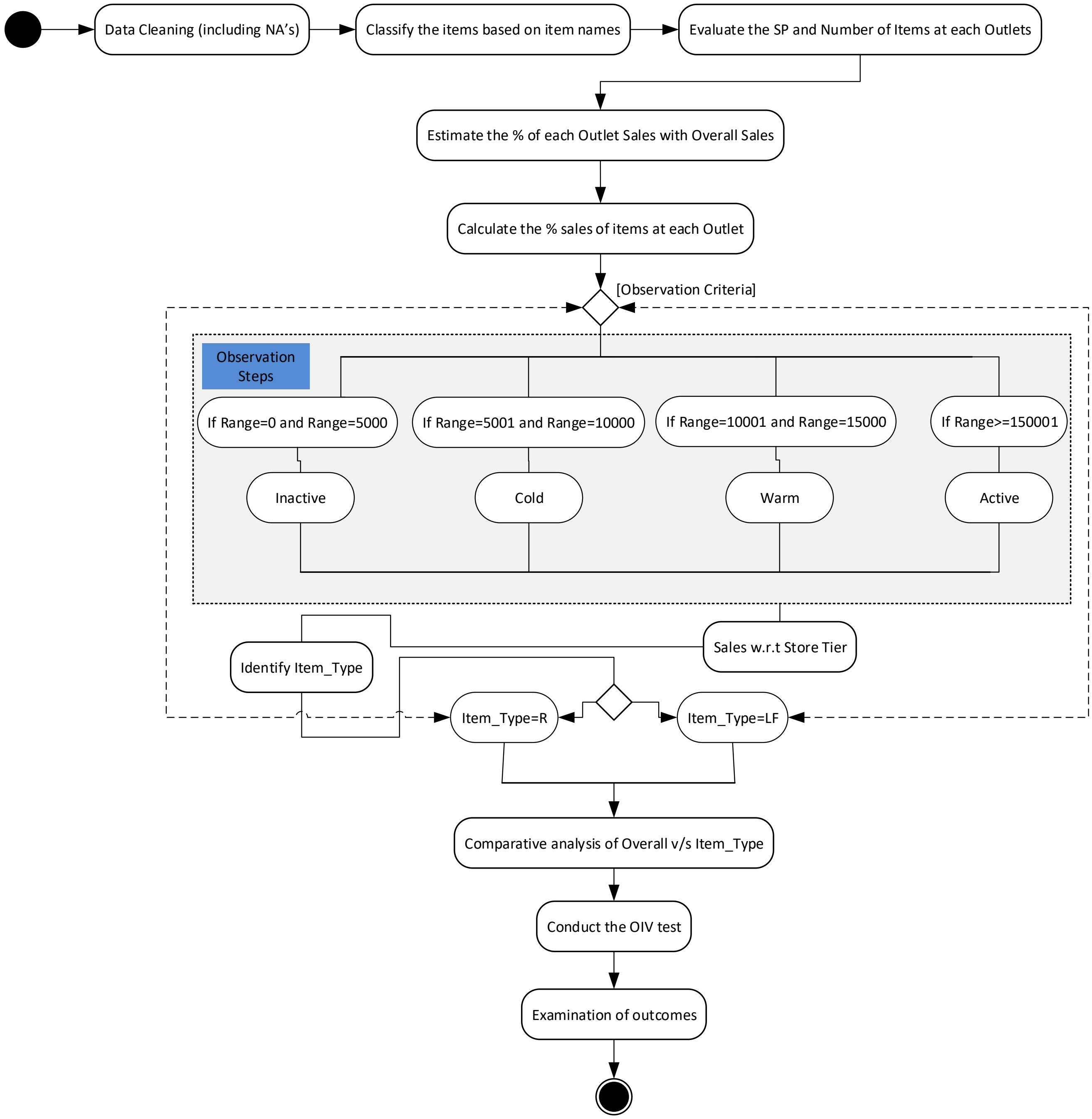

Figure 3: BREAD Model – Generic Activity Diagram

3. IMPLEMENTATION OF BREAD MODEL

BREAD Model is developed using R to convey 4 results. These outcomes are represented in Tables 5,6,7 and 8 distinctly. The implementation of BREAD model comprises the following phases:

1) Data Extraction from ABC.csv (Data Set) and use of sqldf package for Data Evaluation.
2) Data Transfer to ABC.xlsx for Data Visualization.

The situation is displayed in Figure 4 below:

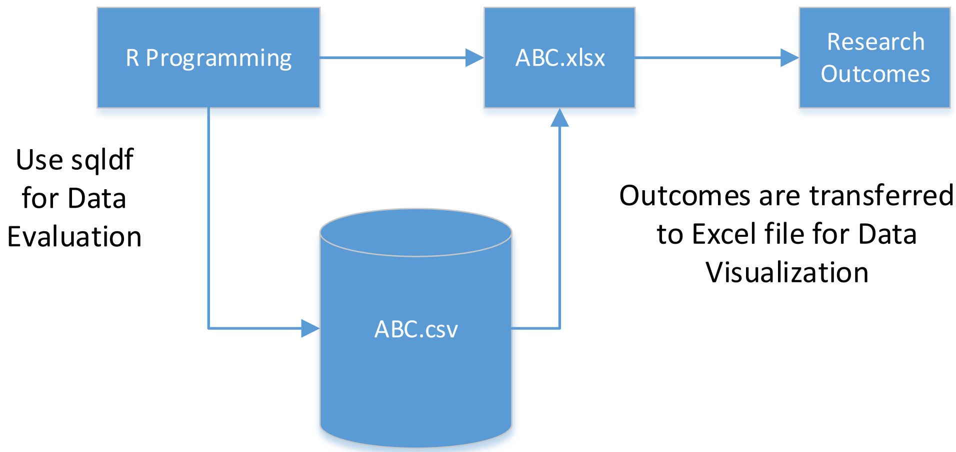

Figure 4: BREAD Model Implmentation steps

Brief Description of Tables are:

Table 5	*Depicts the count and Monetary Value generated from each store of all 16 Item_Type*
Table 6	*Analysis of sales w.r.t. Outlet Tiers (1,2 and 3) of all 16 Item_Type*
Table 7	*Analysis of sales w.r.t. Item_Type=LF/R*
Table 8	*Sales v/s OIV Comparative Analysis*

Item	Outlet 1 (OUT010)		Outlet 2 (OUT013)		Outlet 3 (OUT017)		Outlet 4 (OUT018)		Outlet 5 (OUT019)		Outlet 6 (OUT027)		Outlet 7 (OUT035)		Outlet 8 (OUT045)		Outlet 9 (OUT046)		Outlet 10 (OUT049)	
	C	**SP**	**C**	**SP**	**C**	**SP**	**C**	**SP**	**C**	**SP**	**C**	**SP**	**C**	**SP**	**C**	**SP**	**C**	**SP**	**C**	**SP**
Baking Goods	60	8,263.9	92	11,397.7	88	12,180.3	102	13,095	48	6,316.3	100	12,497.7	100	12,962.8	94	11,020.8	86	11,338.6	106	13,976.4
Breads	26	3,696.5	40	5,956.9	44	6,258.7	36	5,601.6	24	3,636.4	28	3,651.5	32	4,258	24	4,023.5	36	4,414	40	5,380.6
Breakfast	6	824.9	14	1,621.5	16	1,763.7	16	2,340.6	8	1,370.9	18	2,196.8	20	2,566.3	20	2,302.8	20	3,057	14	2,092.2
Canned	62	8,840.4	106	15,305.4	98	13,647.6	82	11,008.5	56	7,923.5	96	12,947.6	78	10,063.1	92	11,273	96	13,129.7	104	14,977.5
Dairy	56	8,132.5	88	11,763.8	102	14,759.5	102	15,256.1	42	7,444.2	116	16,807.9	106	14,665.5	112	16,492	92	13,669.8	92	12,964.3
Frozen Foods	78	10,635.0	126	17,241.7	98	13,971.9	124	15,593.6	94	12,827.1	132	17,925.8	126	15,915.1	146	19,037.3	92	12,623.5	124	16,315.8
Fruits and Vegetables	116	16,489.6	156	22,346.1	176	25,341.3	168	25,470.3	106	15,316.5	160	23,035.7	180	26,574.9	150	22,093.2	188	24,154.6	162	23,898.3
Hard Drinks	18	2,892.6	34	4,446	36	4,590.2	34	3,457.7	22	3,817.9	34	4,349.1	36	5,032.3	24	3,619.6	38	5,174.4	20	3,352.1
Health and Hygiene	36	4,844.4	66	8,568.6	62	8,653.5	68	9,945.3	46	5,537.5	70	9,979.6	86	11,733.3	78	10,966.7	78	10,555.7	86	11,576.8
Household	74	10,720.2	130	19,932.5	148	22,099.7	146	21,374.1	82	10,940.4	142	21,718.2	134	18,808.8	142	21,446.6	132	19,383.2	146	22,111.1
Meat	36	4,930.6	78	11,278.7	72	10,705.4	64	9,143.1	34	4,484.8	48	6,849	74	9,633.1	82	11,102.1	72	9,674.6	62	9,807

Table 5: Sales and Count of Items

Others	16	2,275.6	28	3,842	28	3,449	20	2,975.9	10	903.8	30	4,893.4	28	3,625.2	20	3,034	16	1,862.2	26	3,132
Seafood	0	-	10	1,470.2	8	1,087.4	6	649.6	0	-	6	1,085.8	4	782	2	397.5	8	709.4	6	811.4
Snack Foods	104	15,568.6	190	27,831.8	182	27,620.6	174	24,263.4	82	12,420.6	166	24,784.5	158	23,293	170	24,810.6	192	28,770.8	160	22,387.1
Soft Drinks	36	4,266.5	62	8,750.03	52	7,018.8	68	9,036	34	4,808.1	70	10,539.3	56	7,973.6	54	7,985.8	64	9,227.3	66	9,972.1
Starchy Foods	16	2,501.0	22	2,973.40	24	3,766.4	26	3,807.6	16	2,360.7	32	4,984.5	22	3,337	28	4,602.7	30	4,576	26	3,896.8

Table 6: Sale w.r.t Outlet Tiers

Item Sales (Type/Tier)	Outlets 5,9,10	Outlets 3,7,8	Outlets 1,2,4,6	OUTCOMES				
Row Labels	Tier 1	Tier 2	Tier 3	Grand Total	Status Tier 1	Status Tier 2	Status Tier 3	Overall Status
Baking Goods	15815.68	18082.02	22627.20	56524.90	Active	Active	Active	Active
Breads	6715.59	7270.16	9453.29	23439.05	Cold	Cold	Cold	Cold
Breakfast	3260.09	3316.47	3492.02	10068.59	Inactive	Inactive	Inactive	Cold
Canned	18015.42	17491.90	24051.03	59558.36	Active	Active	Active	Active
Dairy	17039.16	22958.57	25980.21	65977.95	Active	Active	Active	Active
Frozen Foods	20883.23	24462.22	30698.15	76043.60	Active	Active	Active	Active
Fruits and Vegetables	31684.74	37004.71	43670.88	112360.34	Active	Active	Active	Active
Hard Drinks	6172.26	6621.11	7572.74	20366.12	Cold	Cold	Cold	Cold
Health and Hygiene	13835.06	15676.81	16669.01	46180.89	Warm	Active	Active	Warm
Household	26217.37	31177.61	36872.55	94267.57	Active	Active	Active	Active
Meat	11983.26	15720.39	16100.80	43804.41	Warm	Active	Active	Warm
Others	2949.03	5054.16	6993.51	14996.72	Inactive	Cold	Cold	Cold
Seafood	760.44	1133.49	1602.88	3496.82	Inactive	Inactive	Inactive	Inactive
Snack Foods	31789.29	37862.13	46224.20	115875.63	Active	Active	Active	Active
Soft Drinks	12003.82	11489.13	16295.98	39788.94	Warm	Warm	Active	Warm
Starchy Foods	5416.81	5853.11	7133.31	18403.23	Cold	Cold	Cold	Cold

Table 7: Sales w.r.t Item_Type=LF/R

Item	Sales (Item Type=LF)	Sales (Item Type=R)	Status (Item Type=LF)	Status (Item Type=R)
Baking Goods	54201.23	58848.59	Active	Active
Breads	21829.17	25048.94	Active	Active
Breakfast	7698.3	12438.9	Cold	Warm
Canned	58688.51	60428.22	Active	Active
Dairy	78801.37	53154.53	Active	Active
Frozen Foods	72159.62	79927.6	Active	Active
Fruits and Vegetables	126749.35	97971.34	Active	Active
Hard Drinks	40732.25	0	Active	Inactive
Health and Hygiene	92361.79	0	Active	Inactive
Household	188535.16	0	Active	Inactive
Meat	30041	57567.83	Active	Active
Others	29993.44	0	Active	Inactive
Seafood	3354.58	3639.07	Inactive	Inactive
Snack Foods	141141.67	90609.6	Active	Active
Soft Drinks	70296.44	9281.44	Active	Cold
Starchy Foods	24426.48	12380	Active	Warm

Table 8: OIV of Item_Type

Item	Optimum Item Visibility (OIV)	Visibility Status
Baking Goods	57.83	Moderately Visible
Breads	24.61	Slightly Visible
Breakfast	11.82	Invisible
Canned	58.64	Moderately Visible
Dairy	63.73	Visible
Frozen Foods	79.63	Visible
Fruits and Vegetables	107.25	Demanding
Hard Drinks	19.78	Invisible
Health and Hygiene	39.24	Slightly Visible
Household	73.28	Visible
Meat	34.64	Slightly Visible
Others	11.57	Invisible
Seafood	3.05	Invisible
Snack Foods	107.18	Demanding
Soft Drinks	37.86	Slightly Visible
Starchy Foods	16.20	Invisible

4. OBSERVATIONS AND RECOMMENDATIONS

Chart 1 presents the comprehensive analysis of the study piloted. The key observations and recommendations are specified in Table 9:

Chart 1: Observations and Recommendations Chart

Item	Observations	Recommendations
Baking Soda	Status is ACTIVE (Both Item_Type=LF/R, Tier 1, 2 and 3) No much difference between Sales in both LF and R	OIV Status to be upgraded to VISIBLE.
Breads	Status = ACTIVE (Item_Type=LF/R) Status = COLD (Overall,Tier 1, 2 and 3) No much difference between Sales in both LF and R	OIV Status to be upgraded to DEMANDING.
Breakfast	Status = COLD (Item_Type=LF and Overall) Status = INACTIVE (Tier 1, 2 and 3) Status = WARM (Item_Type=R) Difference in Sales of LF and R (R is approx. $5000 more than LF)	OIV Status of Breakfast (Item_Type=R) to be upgraded to VISIBLE. OIV Status of Breakfast (Item_Type=LF) to be upgraded to MODERATELY VISIBLE.
Canned	Status=ACTIVE (LF/R, Tier 1,2 and 3, Overall)	No Recommendations
Dairy	Status=ACTIVE (LF/R, Tier 1,2 and 3, Overall)	No Recommendations
Frozen Foods	Status=ACTIVE (LF/R, Tier 1,2 and 3, Overall)	No Recommendations
Fruits and Vegetables	Status=ACTIVE (LF/R, Tier 1,2 and 3, Overall)	No Recommendations
Hard Drinks	Status=ACTIVE (Item_Type=LF) Status = INACTIVE (Item_Type=R) Status = COLD (Overall,Tier 1, 2 and 3)	Concentrate on Item_Type=LF. OIV Status to be upgraded to VISIBLE
Health and Hygiene	Status=ACTIVE (Item_Type=LF) Status = INACTIVE (Item_Type=R) Status = WARM (Overall,Tier 1) Status = COLD (Tier 2 and 3)	Concentrate on Item_Type=LF. OIV Status to be upgraded to VISIBLE
Household	Status=ACTIVE (Item_Type=LF, Overall,Tier 1, 2 and 3) Status = INACTIVE (Item_Type=R)	No Recommendations
Meat	Status = ACTIVE (Item_Type=LF/R, Tier 2 and 3) Status = WARM (Overall,Tier 1)	OIV Status to be upgraded to VISIBLE
Others	Status=ACTIVE (Item_Type=LF) Status = INACTIVE (Item_Type=R, Tier 1) Status = COLD (Overall,Tier 2 and 3)	OIV Status to be upgraded to MODERATELY VISIBLE.
Seafood	Status = INACTIVE (LF/R, Tier 1,2 and 3, Overall)	**Discontinue the Item**
Snack Foods	Status=ACTIVE (LF/R, Tier 1,2 and 3, Overall) Difference in Sales of LF and R (LF is approx. $50000 more than R)	No Recommendations
Soft Drinks	Status = ACTIVE (Item_Type=LF) Status = COLD (Item_Type=R) Status = WARM (Overall,Tier 1, 2 and 3) Difference in Sales of LF and R (LF is approx. $60000 more than R)	OIV Status to be upgraded to MODERATELY VISIBLE.
Starchy Foods	Status = ACTIVE (Item_Type=LF) Status = WARM (Item_Type=R) Status=ACTIVE (Overall,Tier 1, 2 and 3) Difference in Sales of LF and R (LF is double the sales of R)	OIV Status to be upgraded to SLIGHTLY VISIBLE.

Table 9: Complete Analysis

Items	Total Sales	Sales (Item_Type=LF)	Sales (Item_Type=R)	Status (Item_Type=LF)	Status (Item_Type=R)	Overall Status	Status Tier 1	Status Tier 2	Status Tier 3	OIV
Baking Goods	113049.82	54201.23	58848.59	Active	Active	Active	Active	Active	Active	Moderately Visible
Breads	46878.11	21829.17	25048.94	Active	Active	Cold	Cold	Cold	Cold	Slighthly Visible
Breakfast	20137.19	7698.30	12438.90	Cold	Warm	Cold	Inactive	Inactive	Inactive	Invisible
Canned	119116.74	58688.51	60428.22	Active	Active	Active	Active	Active	Active	Moderately Visible
Dairy	131955.91	78801.37	53154.53	Active	Active	Active	Active	Active	Active	Visible
Frozen Foods	152087.22	72159.62	79927.60	Active	Active	Active	Active	Active	Active	Visible
Fruits and Vegetables	224720.69	126749.35	97971.34	Active	Active	Active	Active	Active	Active	Demanding
Hard Drinks	40732.25	40732.25	0.00	Active	Inactive	Cold	Cold	Cold	Cold	Invisible
Health and Hygiene	92361.79	92361.79	0.00	Active	Inactive	Warm	Warm	Active	Active	Slighthly Visible
Household	188535.16	188535.16	0.00	Active	Inactive	Active	Active	Active	Active	Visible
Meat	87608.83	30041.00	57567.83	Active	Active	Warm	Warm	Active	Active	Slighthly Visible
Others	29993.44	29993.44	0.00	Active	Inactive	Cold	Inactive	Cold	Cold	Invisible
Seafood	6993.66	3354.58	3639.07	Inactive	Inactive	Inactive	Inactive	Inactive	Inactive	Invisible
Snack Foods	231751.27	141141.67	90609.60	Active	Active	Active	Active	Active	Active	Demanding
Soft Drinks	79577.88	70296.44	9281.44	Active	Cold	Warm	Warm	Warm	Warm	Slighthly Visible
Starchy Foods	36806.48	24426.48	12380.00	Active	Warm	Cold	Cold	Cold	Cold	Invisible

References

Adams, M. (2008). *Predictive Analytics for the Retail Industry (MS SQL Server Technical Article).* Microsoft.

Brynjolfsson, E. H. (2011). Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance? *ebusiness.mit.edu*.

Emirates247. (2016, April 16). *UAE retail sector to grow 5% each year through 2017: Dubai Chamber*. Retrieved from Emirates 247: http://www.emirates247.com/business/economy-finance/uae-retail-sector-to-grow-5-each-year-through-2017-dubai-chamber-2016-04-11-1.626809

Euromonitor. (2016, May). *Retailing in United Arab Emirates.* Retrieved from http://www.euromonitor.com/retailing-in-the-united-arab-emirates/report

Hana Ben-Shabat, M. M. (2015). *Global Retail Expansion: An Unstoppable Force*. Retrieved from ATKearney: https://www.atkearney.com/consumer-products-retail/global-retail-development-index/2015

Inna Kolyshkina, S. S. (2007). Proceedings, AusDM '07 Proceedings of the sixth Australasian conference on Data mining and analytics - Volume 70. *Australian Computer Society* (pp. 13-19). ACM Digital Library.

Loraine Charlet Annie M.C., A. K. (2012). Market Basket Analysis for a Supermarket based on Frequent Itemset Mining. *International Journal of Computer Science Issues*, 257-264.

Luckie, C. (2012). *"Big Data" Facts and Statistics That Will Shock You.* Fathom.

Steve Lavalle, M. S. (2010). *Analytics: The New Path to Value.* MIT Sloan Management Review.

AUTHOR PROFILE

Riktesh Srivastava, is PhD (CS) and EGMP (IIMA), Certification in Marketing, Customer Analytics from Wharton School (University of Pennsylvania) and Electronic Commerce from NTU, Singapore. Currently is Associate Professor (Information Systems) at Skyline

University College, Sharjah, UAE. His area of interest is Business Analytics, Queuing Theory and Electronic Commerce.