Using the Knowledge Graph in Marketing Activation

Nandin-Erdene Enkhmyagmar[1]; Enkhtuul Bukhsuren[2];
Tumen-Ochir Tumurtulga[3]; Enerlen Enkhtur[4]; Munkhtsetseg Namsraidorj[5]

[1,2,5]Department of Information and Computer Sciences School of Information Technology and Electronics,
National University of Mongolia Ulaanbaatar, Mongolia
[3]Logarithm School, Ulaanbaatar, Mongolia
[4]Absolute Elite School, Ulaanbaatar, Mongolia

Publication Date: 2025/03/31

Abstract: In this article, we present a work that uses machine learning methods to create a graph database with knowledge graphs to investigate how a wholesale business organization can monitor, improve, and manage revenue changes over time based on data from marketing activation methods used to improve sales revenue of wholesale goods.

Keywords: *Ontology; SPARQL; GraphDB; Sales; Machine Learning.*

How to Cite: Nandin-Erdene Enkhmyagmar; Enkhtuul Bukhsuren; Tumen-Ochir Tumurtulga; Enerlen Enkhtur; Munkhtsetseg Namsraidorj. (2025) Using the Knowledge Graph in Marketing Activation. *International Journal of Innovative Science and Research Technology*, 10(3), 1549-1555. https://doi.org/10.38124/ijisrt/25mar967.

I. INTRODUCTION

The primary goal of a business organization is to generate profit, and to achieve this, it is necessary to identify the factors influencing sales revenue. In today's world, businesses are leveraging advancements in information technology to enhance various aspects such as operations, revenue, and customer behavior.

By predicting sales revenue in advance, organizations can efficiently manage their resources, develop strategies to increase revenue, and optimize their operations.

When selecting a forecasting model, factors such as technical resources and costs, historical data availability, data variability and stability, accuracy, and time constraints [1] are considered. In the wholesale trade sector, time series forecasting models are advantageous for demand prediction due to their ease of implementation, better interpretability compared to competing methods, and relatively higher accuracy [2].

II. RELATED WORK

A. *Sales Revenue Prediction Model*

A multiple regression model is used to establish a linear relationship between a dependent variable (y) and two or more independent variables (xi).

The population model for multiple regression is represented as follows:

$$Y = \beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + \cdots + \beta_p * x_p + \varepsilon \tag{1}$$

Where:

Y is the dependent variable, $x_1, x_{2,\ldots,} x_p$ – independent variables, β_0 – intercept term, $\beta_1, \beta_2, \ldots, \beta_p$ are the estimated slope coefficients based on sample data, ε – the random error term, representing the deviations from the mean value Y. When the error variance follows a normal distribution, its expected value is 0.

- *The Sample Regression Model is Given as:*

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k \tag{2}$$

Where:
$\hat{y}$- predicted value, b_0 – estimated intercept term
$b_1, b_2, \ldots, b_k$ – estimated slope coefficients.

The error term determined from the regression model is:

$$e = (y - \hat{y}) \tag{3}$$

The coefficient of determination is a concept that represents how much of the variation in y is explained collectively by the considered independent variables.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25mar967

$$R^2 = \frac{SSR}{SST} = \frac{\text{Регрессийн квадрат нийлбэр}}{\text{Нийт квадрат нийлбэр}} \quad (4)$$

The statistical significance of the constructed model is tested using the F-Test.

To check whether there is a very high correlation between independent variables, the Variance Inflation Factor (VIF) is calculated as a measure of collinearity.

$$VIF_j = \frac{1}{1-R_j^2} \quad (5)$$

B. Ontology

Ontology defines the structure and relationships of concepts and entities within a specific domain, classifying the considered entities and explaining their interconnections. This enables computers and other systems to better understand and process these concepts.

Ontology is used in artificial intelligence, the semantic web, software development, and database management for knowledge sharing, integration, and reasoning [3]-[5].

By systematically organizing and linking information, ontology enhances data interactions, improves decision-making processes, and aids in the development of automated systems [6].

C. Resource Description Framework

This is a standard format used to define, organize, and share information on the internet, forming the foundation of the semantic web.

RDF represents data in a "Subject-Predicate-Object" structure [7].

It allows data to be expressed in various formats such as XML, JSON-LD, and Turtle. Additionally, RDF facilitates linking and sharing data, making it easier to integrate information from multiple sources. RDF plays a crucial role in ontology, the semantic web, and other information technologies, helping to utilize data in a more intelligent and organized manner.

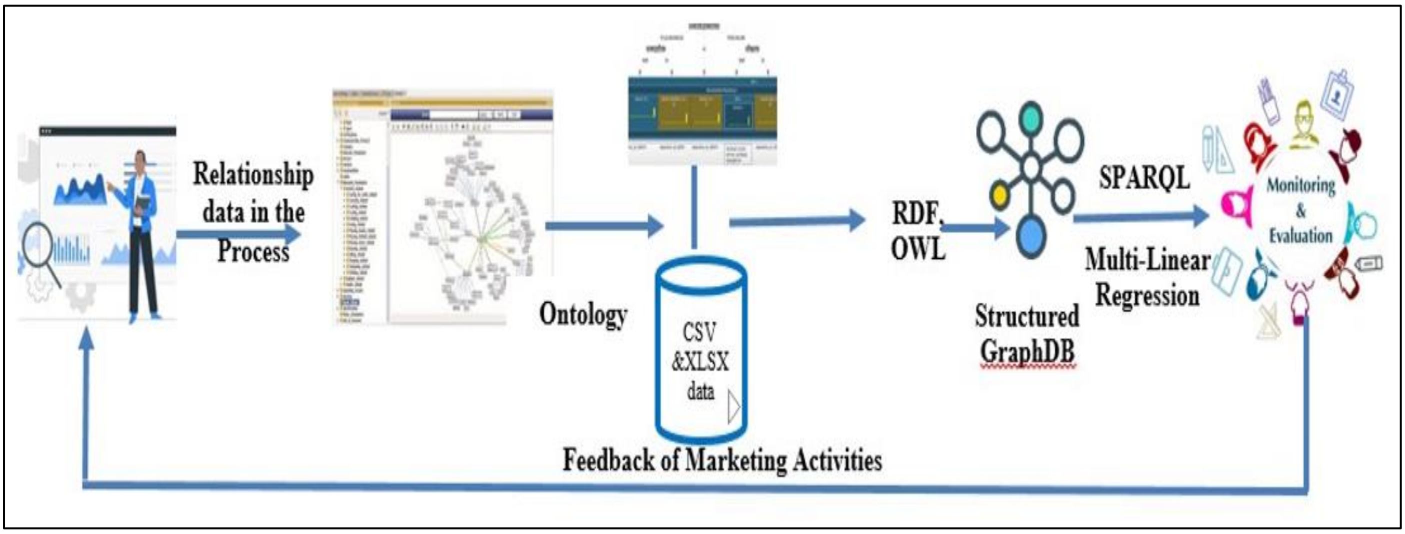


Fig 1: Ontology-Based Proposed Model for Revenue Enhancement.

D. OWL and SPARQL

Web Ontology Language (OWL) is a semantic web language designed to represent knowledge about relationships between entities. OWL files can be published on the World Wide Web. OWL is part of W3C's semantic web technologies, including RDF, RDFS, and SPARQL. SPARQL (SPARQL Protocol and RDF Query Language) is a standardized language used for querying, managing, and processing RDF (Resource Description Framework) data [8].

III. PROPOSED METHODOLOGY

To implement the monitoring, improvement, management, and prediction of wholesale sales revenue using information technology, we carried out the following steps:

- Analyzed time-based data from a distributor company and prepared the data.
- Collected and prepared data related to marketing activities aimed at improving the organization's sales revenue.
- Constructed an ontology based on the analysis of the collected data.
- Created a graph database using the developed ontology.
- Developed a SPARQL query system to extract essential data from the graph database for daily business operations and set up an infrastructure for monitoring information.
- Applied multiple regression analysis using machine learning techniques to identify the marketing activities that have the greatest impact on maximizing sales revenue.

The model we developed is illustrated in Figure 1.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25mar967

IV. EXPERIMENTAL RESULTS

A. Analysis of Existing Data

Our goal when designing our ontology was to study the impact of marketing activation on sales revenue. To achieve this, we utilized data from six files containing order information from the distributor company over a two-year period (2023, 2024), including fields such as order details, sales volume, total revenue, brand, product name, product code, product price, customer information, sales representative information, sales channel, marketing activation type and the associated costs, and activation participation levels. A key feature of our approach is that we specifically defined the more precise ontology for each type of marketing activation based on the marketing-related data extracted from these sources.

B. Marketing Activation and Promotions

The ontology establishes the relationship between marketing activities and sales revenue, which facilitates more efficient processing of input data for multiple regression analysis. In other words, at the ontology level, sales are analyzed in relation to the corresponding month's "advertisement" and "discount" data.

The distributor company conducts marketing activities targeting both end consumers and business clients (such as supermarkets and stores). For example, television advertisements are primarily designed for end consumers, whereas order-based discounts and incentives are mainly offered to business clients. Each month, the company carries out the several marketing activities and promotions for each brand, including: Discounts and gift promotions (gifts with purchase), television advertisements, social media advertisements (Instagram, Facebook, Google Ads), collaborations with influencers (content marketing, giveaways, etc.), renting shelf space in high-traffic areas for product display, announcing stocking incentives, organizing events, hiring promotional staff. The costs of these activities vary depending on their type and level of engagement. By calculating the expenses per category each month, it becomes possible to analyze which type of activities and promotion contributes most effectively to sales growth.

C. Development of Ontology and Graph Database

By developing and utilizing the semantic web, ontology, and knowledge graphs, it becomes possible to focus on improving sales revenue, which is a key challenge in the wholesale business. We have developed an ontology that integrates the main concepts influencing sales and built a graph database using historical data. This allows for quick and efficient querying to determine what aspects should be prioritized in future sales.

➤ *To Construct a Knowledge Graph Capable of Evaluating Sales Revenue, We Followed These Steps:*

- Identifying operational entities
- Extending an existing ontology or developing a new one
- Defining classes, subclasses, their relationships, and properties
- Converting sales data into RDF format
- Building a Graph Database
- Creating a model and evaluating its performance.

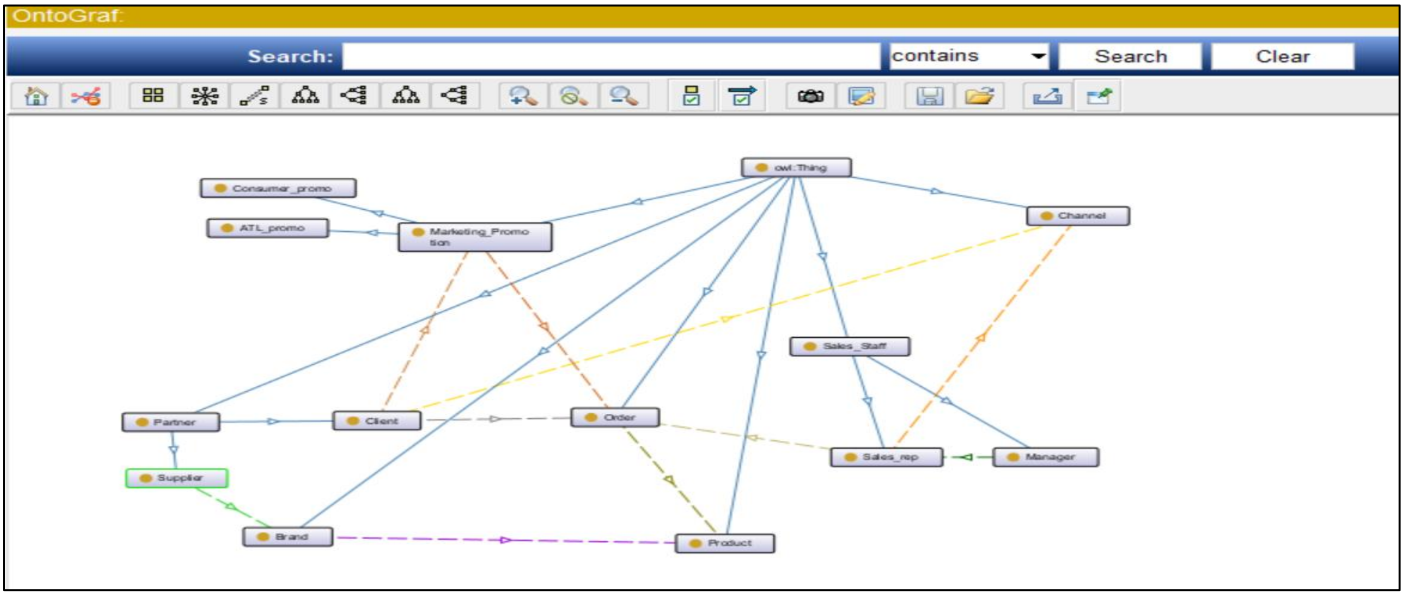


Fig 2: Ontology Graph

Our ontology is structured with a total of 13 classes, 11 object properties, and 28 data properties. Figure 2 illustrates the ontology class relationships, while Figure 3 presents the class hierarchy structure.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25mar967

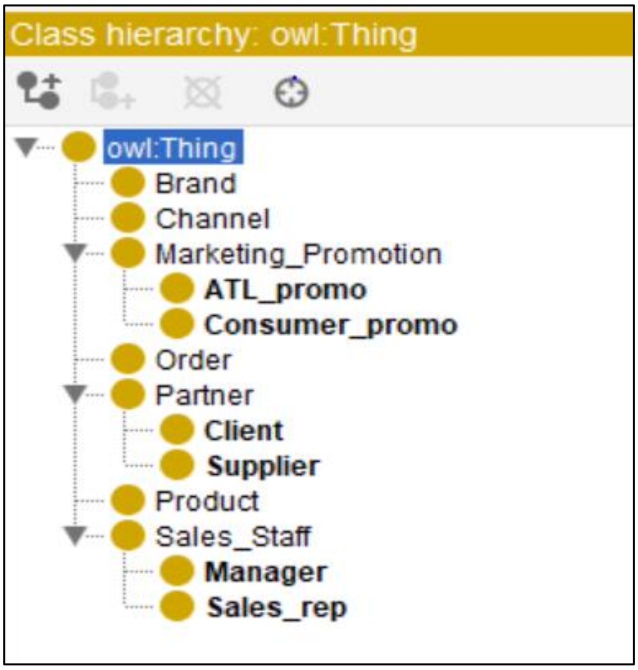


Fig 3: Ontology Classes

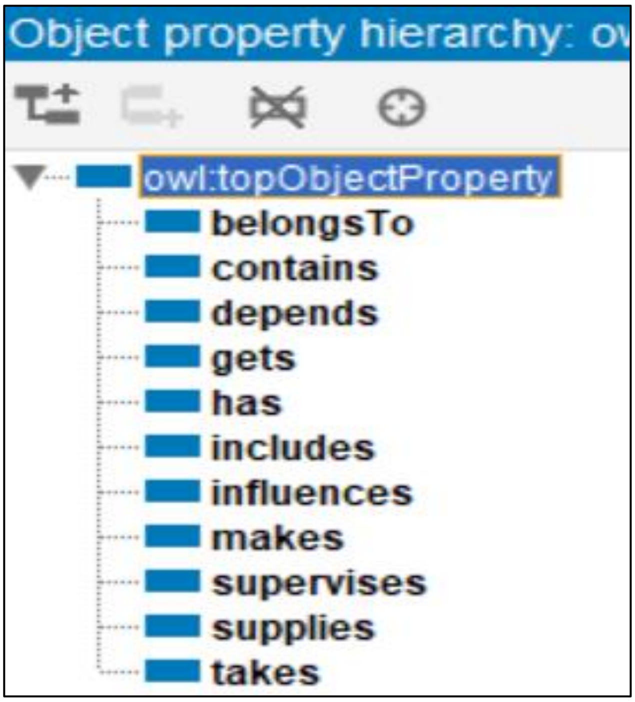


Fig 4: Object properties.

The relationships between classes are represented by the properties shown in Figure 4. For example, the "contains" property indicates the relationship where a Brand contains products, as illustrated in Figure 5.

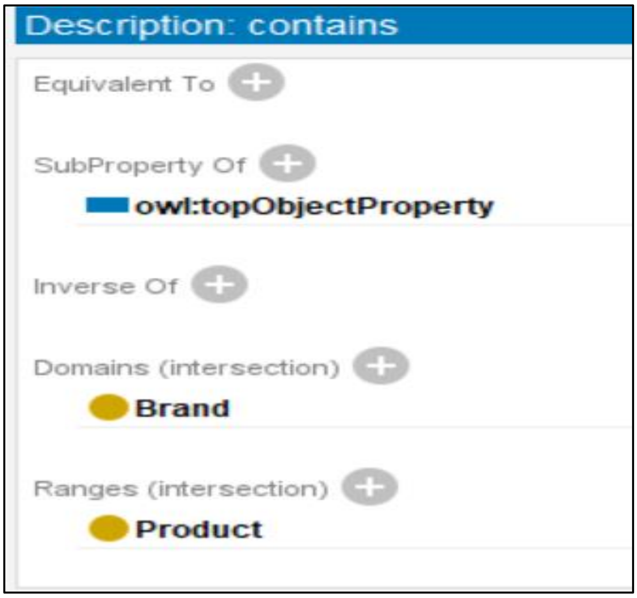


Fig 5: "Contains" Object Property

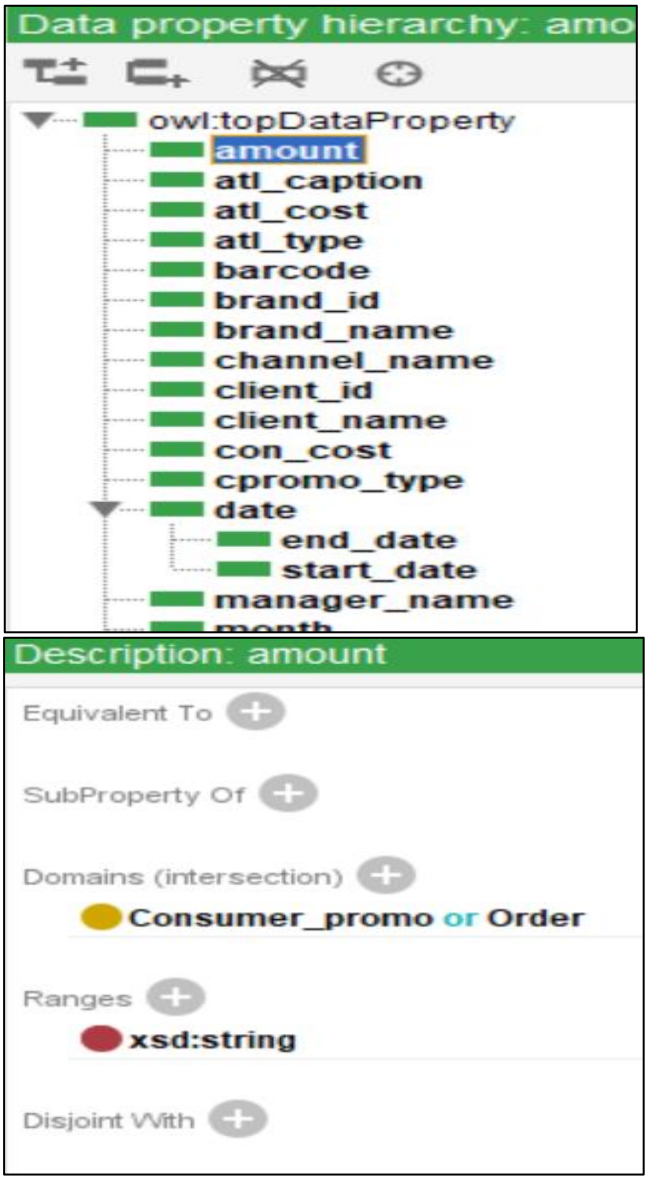


Fig 6: Data Properties.

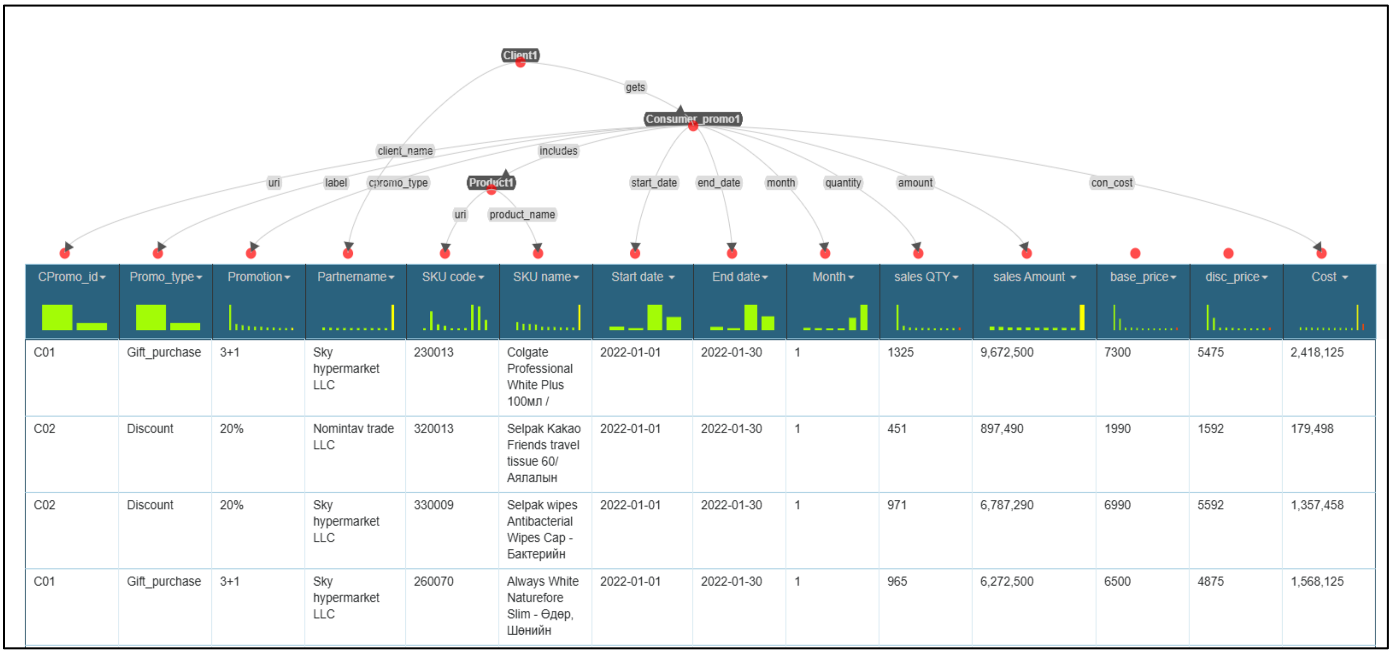
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25mar967

Figure 6 illustrates the data properties associated with each Class. For example, the "amount" property can be used in both the Consumer_promo and Order classes.

D. Building a Graph DB

➢ *Data Integration*

Since it was necessary to integrate data according to the designed ontology, we used the "Karma" data integration tool.



CPromo_id	Promo_type	Promotion	Partnername	SKU code	SKU name	Start date	End date	Month	sales QTY	sales Amount	base_price	disc_price	Cost
C01	Gift_purchase	3+1	Sky hypermarket LLC	230013	Colgate Professional White Plus 100мл /	2022-01-01	2022-01-30	1	1325	9,672,500	7300	5475	2,418,125
C02	Discount	20%	Nomintav trade LLC	320013	Selpak Kakao Friends travel tissue 60/ Аяллын	2022-01-01	2022-01-30	1	451	897,490	1990	1592	179,498
C02	Discount	20%	Sky hypermarket LLC	330009	Selpak wipes Antibacterial Wipes Cap - Бактерийн	2022-01-01	2022-01-30	1	971	6,787,290	6990	5592	1,357,458
C01	Gift_purchase	3+1	Sky hypermarket LLC	260070	Always White Naturefore Slim - Өдөр, Шөнийн	2022-01-01	2022-01-30	1	965	6,272,500	6500	4875	1,568,125

Fig 7: Data Integration

By using the Karma tool, it becomes easy to convert .csv and other file formats into RDF based on the ontology, generate an R2RML model for reuse, and perform various data integration tasks efficiently.

The integrated for of marketing activation and promotion data targeted at end consumers is shown in Figure 7.

➢ *Graph Database*

The Graph Database was created using the generated RDF, allowing SPARQL queries to easily retrieve all types of semantically structured information required for daily business operations.

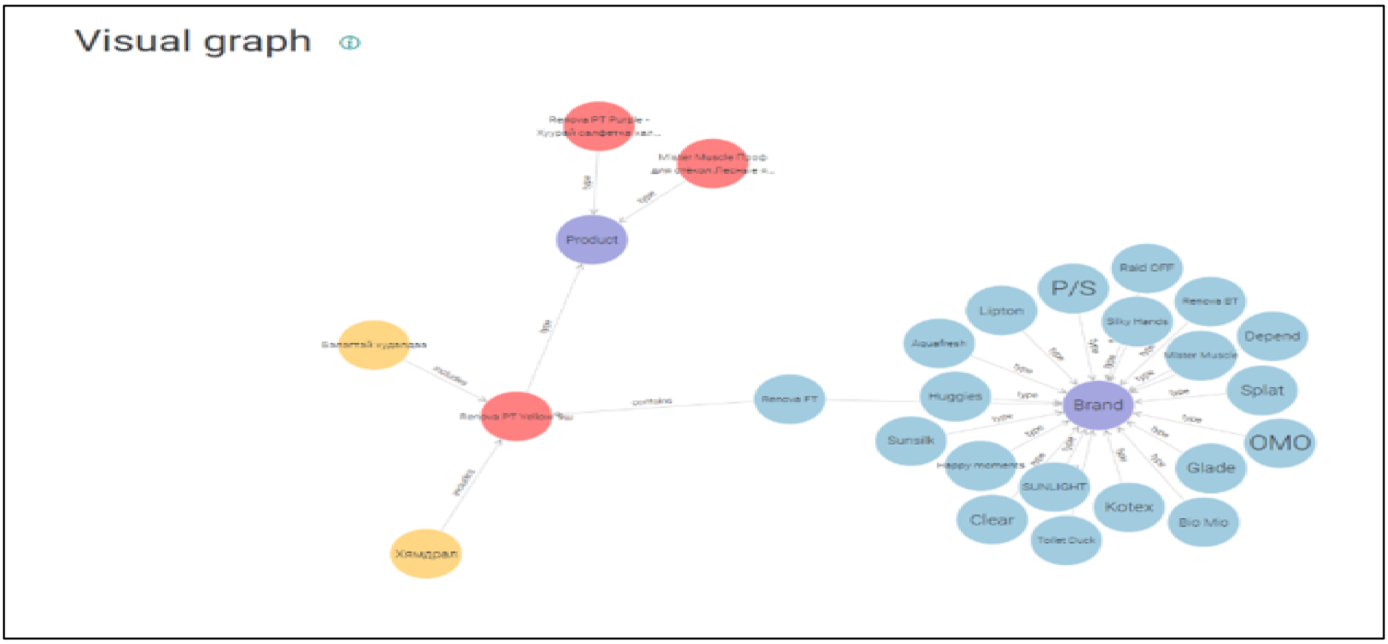


Fig 8: Visual Graph in GraphDB

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25mar967

The data in the Graph Database is visualized, as shown in Figure 8, providing a unified view of the marketing activation and promotions applied to each product. This is crucial for the company in planning future marketing strategies.

➢ *Model Evaluation*

To determine which marketing promotion method has the most significant impact on sales revenue, data from the Graph was used to perform multiple regression analysis.

Machine learning techniques were applied to estimate the parameters of the best-fitting linear model in the multiple regression analysis.

First, we checked for high correlation between independent variables. To do this, we calculated the Variance Inflation Factor (VIF_j) as a measure of collinearity, and the results were as follows.

Table 1: Variance Inflation Factor

	Variable	VIF
0	const	26.767809
1	Gift_purchase	1.470370
2	Discount	1.675928
3	Giveaway	1.672399
4	Gondola	2.394477
5	Influencer	1.630362
6	Merchandise	1.373275
7	Promoter	1.426720
8	Social_media	1.190947
9	TV_commercial	1.382248

Since the examined variables do not exhibit high correlation with each other, they are suitable for analysis.

Multiple regression analysis was computed using Python.

```
X = df[['Gift_purchase','Discount', 'Giveaway', 'Gondola', 'Influencer', 'Merchandise', 'Promoter', 'Social_media', 'TV_commercial']]

y = df['Sales']
print(df.dtypes)
print(df.isnull().sum
X = sm.add_constant(X)
model = sm.OLS(y, X).fit()
print(model.summary())
...
```

Figure 9 shows the results of the multiple regression analysis.

	coef	std err	t	P>|t|
const	5.746e+09	9.18e+08	6.261	0.000
Gift_purchase	2.0480	0.841	2.434	0.029
Discount	15.9784	5.207	3.069	0.008
Giveaway	109.5655	107.840	1.016	0.327
Gondola	-140.7755	141.610	-0.994	0.337
Influencer	5.7321	81.448	0.070	0.945
Merchandise	-744.2826	627.490	-1.186	0.255
Promoter	10.3284	289.206	0.036	0.972
Social_media	20.5052	28.369	0.723	0.482
TV_commercial	42.0760	30.424	1.383	0.188

Fig 9: Variance Inflation Factor

The coefficient indicates how sales revenue changes with a one-unit change in the corresponding variable.

The const coefficient represents the baseline sales revenue when all variables are set to zero. For example, if no marketing activities are conducted, the sales revenue is 5.746 billion.

The analysis shows that the variables Gift_purchase, Discount, and TV_commercial are statistically significant, whereas Giveaway, Gondola, Influencer, Merchandise, Promoter, and Social_media have a p-value > 0.05, indicating that they are not statistically significant and do not impact sales revenue.

The multiple regression model for predicting sales revenue was established as follows.

$$Sales=5.746e+2.04*Gift\_purchase+15.97*Discount+42.07*TV\_commercial \quad (6)$$

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25mar967

The parameters of the established model can be interpreted as follows: When the costs of discount sales and TV advertisements remain constant, increasing the budget for gift promotions by one unit increases sales revenue by 2.04 units, whereas when the costs of gift promotions and TV advertisements remain constant, increasing the budget for discount sales by one unit increases sales revenue by 15.97 units, and when the costs of gift promotions and discount sales remain constant, increasing the budget for TV advertisements by one unit increases sales revenue by 42.07 units.

Our model's F-statistic = 4.71 and Prob (F-statistic) = 0.005, indicating that the model is statistically significant.

The R-squared value of 0.752 shows that the selected independent variables explain 75.2% of the total sales revenue variation.

The Adjusted R-squared value, which accounts for the number of independent variables, is 59.2%, suggesting that some variables may not have a significant effect on the sales revenue.

Based on these findings, the distributor company can use past month's data to predict the most effective marketing activities for the upcoming month, helping in decision-making to maximize sales revenue.

V. CONCLUSIONS

In terms of results, high-cost marketing activities such as Gift Promotions, Discounts, and TV Advertisements were found to have a statistically significant impact on sales. However, this only reflects the direct quantitative effect on total sales revenue. Other marketing activities may have an indirect impact on sales by influencing brand reputation, recognition, recall, consumer behavior, habits, and differentiation from competitors, which are qualitative factors that are not immediately visible in sales figures. In this study, we explored and proposed a method for integrating data in a more semantic manner using ontology to evaluate the impact of marketing promotions on product sales. By utilizing ontology, we built a graph database that links sales and marketing activities, allowing for a unified view of all marketing promotions conducted for a product at a given time and analyze their impact.

Moving forward, we aim to develop a model for assessing the effectiveness of marketing activities on sales.

REFERENCES

[1]. Herbig, P. a., Milewicz, J., & Golden, J. E. The do's and don'ts of sales forecasting. Industrial Marketing Management, 22(1), 1993, pp. 49–57.

[2]. Liu, L.-M., Bhattacharyya, S., Sclove, S. L., Chen, R., & Lattyak, W. J. Data mining on time series: an illustration using fast-food restaurant franchise data. Computational Statistics & Data Analysis, 2001, 37(4), pp.455–476.

[3]. Chaccour, C., Saad, W., Debbah, M., Han, Z., & Poor, H. V. Less Data, More Knowledge: Building Next Generation Semantic Communication Networks. IEEE Communications Surveys & Tutorials, 2024. doi:https://doi.org/10.1109/COMST.2024.3412852

[4]. Fu, C., Jiang, H., & Chen, X. RETRACTED ARTICLE: Modeling of an Enterprise Knowledge Management System Based on Artificial Intelligence. Knowledge Management Research & Practice, 2022, pp.1–13. doi:https://doi.org/10.1080/14778238.2020.1854632

[5]. Olan, F., Arakpogun, E. O., Suklan, J., Nakpodia, F., Damij, N., & Jayawickrama, U. Artificial intelligence and knowledge sharing: Contributing factors to organizational performance. Journal of Business Research, 145, 2022, pp. 605-615. doi:https://doi.org/10.1016/j.jbusres.

[6]. Staab, S.; Studer, R. Handbook on Ontologies; Springer Science & Business Media: Berlin, Germany, 2013.

[7]. De Nicola, A., & Villani, M. L. Smart City Ontologies and Their Applications: A Systematic Literature Review. Sustainability, 2021, 13(10). doi:https://doi.org/10.3390/su13105578

[8]. Gruber, T.R. A translation approach to portable ontology specifications. Knowl. Acquis. 1993, 5, pp.199–220.

[9]. "SPARQL Query Language for RDF." Available: https://www.w3.org/TR/rdf-sparql-query/.