

PROJECT - 1

(BUAN 6320: Database Foundations for Business Analytics)

Topic: Retail Data Of Turkish Sector

Abhishek Arya	axa220149@utdallas.edu
Manan Mistry	mxm220108@utdallas.edu
Tanishka Patel	txp210040@utdallas.edu
Het Shah	hrs220002@utdallas.edu
Shekhar Vashist	sxv220008@utdallas.edu

Project made under the guidance of : Farzad Kamalzadeh

SECTION - 1:

ABOUT THE DATASET:

This data is collected from a Turkish retail firm for the period starting from 2017 to 2019. In this data set we have 3 tables which are Sales table, Product hierarchy table and Stores Cities table.

The data present in these 3 tables are as follows:

Sales: Daily sales data from 2017 to 2019

Product hierarchy: Data that contains hierarchy and sizes of products.

Stores cities: Data that contains city, type and size information about the stores of this retail firm.

By using this data, we plan to achieve more insights on the sales and revenue of this retail firm and also try to understand which store and products are performing better during the mentioned time period. And to analyze how to optimize the revenue of the Turkish Retail Firm.

Data Source: <https://www.kaggle.com/datasets/berkayalan/retail-sales-data>

SECTION - 2:

1. Why has this data been gathered?

The data from the retail firm has been collected to record information about the sales and the products in stock every day. The Turkish sector has collected the data to perform further analysis as well as to maintain records and to evaluate past performance.

2. What can be done with data? What can we achieve?

Making the collected data useful and insightful is a very important process that every company should adapt to. Using this data we can :

- Build prediction models to analyze the business data to improve the store's performance.
- From this data we can create detailed report on the performance of the company in past quarter

3. What are some of the goals/targets we have regarding the business that we can achieve by investigating this data?

GOALS:

- Increase in ROI with proper data analysis
- Increase the Sales Strategy of the store
- Increase visibility of different areas of the business

4. What insightful information can this data provide us that can be used to improve the business?

Insightful information regarding the best selling products, store with highest profit, period with highest sale can be found from this data.

5. Why are we studying this data?

We are studying this data to understand how the firm has performed and what were its shortcomings during the period. Moreover, from the products and stores data we will be able to understand which store is performing better and which product is in more demand during this time. We will also try to understand the impact of promotion campaigns on the revenue of the firm.

6. Are there any problems in our business (based on the given data)?

Through exploratory data analysis we can say that not all the stores and products are performing well. We need to understand the performance of stores with respect to the products sold and hence identify which products are performing well on which store and then maintain the inventory of the stores accordingly.

7. Can we find any solutions to these problems by studying this data?

We can find the solution for the above-mentioned problem by analyzing the appropriate stores for products and hence getting to the perfect combination of stores and products which will make the business more profitable. Also, after achieving these perfect combinations we shall also find out the quantity of stocks to be maintained at different stores.

8. What are some of the things we can optimize/improve in our business by studying this data?

Following are some of the things we can optimize:

- Inventory management: After analyzing the data, we can understand the pattern of sales from each store and get some insights on seasonality i.e., at what time of the year the sales are high and which product is more in demand.

Considering all the factors we can also find out the amount of inventory needed at what time of the year and maintain the stock in stores. This will lead to optimization of stores operations and also efficient inventory management.

- Promo Campaigns: We have a few data points on promo campaigns and from this we can analyze which campaign has performed better and what marketing technique is better to promote the products. Hence this can optimize the resource allocation on promo campaigns.
- Revenue: Once we have analyzed the data we will get a picture of revenue from each stores and each product category. From this information we can also run a predictive analysis to predict the revenue trends of the firm.

SECTION - 3:

9. What information each column of the data contains and its data type?

Sales_Data_Table

Column	Datatype	Information
Product_id	Varchar	Unique product number
Store_id	Varchar	Unique store number
date	Date	Date when sales happened (YYYY-MM_DD)
Sales	Float	Quantity of Sales
Revenue	Float	Daily total sales revenue
Stock	Float	Stock Quantity at the end of the day
Price	Float	Sales Price of the product
Promo_type_1	Varchar	Type of the promotion applied on
Promo_bin_1	Varchar	Promotion rate applied to promo_type_1
Promo_type_2	Varchar	Type of the promotion applied
Promo_type_2	Varchar	Promotion rate applied to promo_type_2
Promo_discount_2	Varchar	Discount rate for applied promo_type_2
Promo_discount_type_2	Varchar	Type of the discount applied to promo_type_2

Product_Details_Table

Column	Datatype	Information
product_id	Varchar	Unique product number
Product_length	Float	Length of the product
Product_depth	Float	Depth of the product
Product_width	Float	Width of the product
Cluster_id	Varchar	Unique number for cluster of products
Hierarchy1_id	Varchar	Barcode Hierarchy for general level
Hierarchy_2_id	Varchar	Barcode hierarchy for second level
Hierarchy_3_id	Varchar	Barcode hierarchy for third level
Hierarchy_4_id	Varchar	Barcode hierarchy for fourth level
Hierarchy_5_id	Varchar	Barcode hierarchy for specific product

Store_Details_Table

Column	Datatype	Information
Store_id	Varchar	Unique store Id
Storetype_id	Varchar	Unique number for store type
Store_size	Float	Size of the store
City_id	Varchar	Unique ID of city

10. What are some of the values each column contains?

Sales: It has no null values and ranges from 0 to 43301.0

Revenue: Revenue ranges from 0 to 84198.0. Revenue is linearly related to sales.

Stock: The average stock is 15.9 and it ranges from 0 to 4716.

Price: The minimum price of an item is 0.01 and maximum price of an item is 1599.

Product_id: The product id is the letter P followed by a 4-digit number code.

Store_id: The store id consists of the letter S followed by a 4-digit number code.

Date: The lowest date is January 02, 2017 and highest date is December 29, 2019.

Promo_type 1: This is a variable that has 17 factors. The type 'PR14' is the most common.

Promo_bin_one: 86% of the value is null. This is a varchar set with 5 factors: very high, high, moderate, low, very low.

Product length: The maximum length of the product is 100. Some of the products don't have length and hence are null.

Product Depth: The maximum depth of the product is 165. It also contains null values for some of the products.

Product Width: The maximum width of the product is 100. It also has some missing values.

11. Do you believe there are outliers in the data?

In the data that we have gathered, there are some outliers. These outliers are few values which are very much far away from all the others in a particular column. For example, only 149 entries have the price over 200 in the sales table. Moreover, there is only one product (P0438) whose stocks are over 3000. There are only 6 values of revenue which are greater than 10,000 from almost 100k values.

12. Are there any duplicate data?

There is no duplicate data. The Product, Store, and Sales tables don't have duplicate elements.

13. Provide simple statistics of the data?

Sales: Retail sales is the sale of consumer goods, or final goods, by businesses to end consumers, and includes in-store sales as well as online sales.

Mean: 0.43619

SD: 14.66364

Mode: 0

Revenue: Retail revenue means revenue earned by the service provider from the sale of products and services directly to the end consumer. Due to the excessive number of zeros and low values, an exponential distribution can be utilized to model distribution of sales and revenue.

Mean: 2.064174

SD: 39.89569

Mode: 0

Stock: This is a threshold that inventory levels should never fall below. It is calculated as average inventory divided by average sales. This method of planning inventory levels is useful for retailers with consistent-selling items that are not subject to large fluctuations.

Mean: 15.88102

SD: 37.28009

Median: 8

Price: the price that customers pay for goods in stores.

Mean: 15.58334

SD: 7.95

Median: 32.29611

Statistics	Product_length	product_width	product_depth	Sales	Revenue	Stock	Price
Mean	7.23	13.5	18.5	0.43619	2.064174	15.88102	15.58334
SD	8.51	10.1	14.3	14.66364	39.89569	37.28009	7.95
Min	0	0	0	-	-	-	-
Q1	3	7.4	10.8	-	-	-	-
Median	5	10.8	17	-	-	8	32.29611
Q3	8	16	23.3	-	-	-	-
Max	100	100	165	-	-	-	-

14. What relationships can you find between the columns? Is there a connection between the columns? Describe.

Sales Table:

- Sales has sales_id (primary key) which makes each sales record unique and not null.
- Sales contains product_hierarchy.product_id which is not null and foreign key.
- Sales contains store_cities.store_id which is not null and foreign key.
- Sales contains sales, revenue, and stock which are correlated and influence each other.
- Sales contains promotion information.
- Overall, Sales records all the transactions and refers to Store Details and Product Details if more information about these two elements is needed.

Product Table:

- Product Details has product_id (primary key) which is not null and unique and a foreign key in Sales.
- product_length, product_depth and product_width are the dimensions of the respective product_id.
- Product Details contains cluster_id, which gives cluster information.
- Product Details contains hierarchy information, which defines the levels of hierarchy starting from largest: level 1 to specific item: level 5.

Store Details Table:

- Store_cities has store_id (primary key) which makes each sales record unique and not null and a foreign key in the Sales table.
- Store Details contains storetype_id, which gives information about store type.
- Store Details contain store information such as store_size and city_id.