



Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

LCBB5014

Data Handling & Business Intelligence

Student Name: Mihaela Cosmina Barabas

Student ID: 2229446

Executive Summary

This report aims to apply Data Warehousing (DW) and Data Mining (DM) techniques on Tesco sales data to obtain important business inferences and provide strategic decision-making is the most common Mode of Payment. Based on the above upward trend, store Digits' predictors are predicting that the Total Number of Sales (TNMOS) will rise to more than 100K in the future. The organisation should concentrate on the digital payment, understanding the low performance of the Health and Beauty product line and utilise the BI tools for a continuous performance monitoring of the departments.

Table of Contents

Executive Summary.....	2
Introduction.....	5
Trends in BI, DW, and DM.....	6
Data Management Foundations (BI, DW, DM).....	6
Data Consolidation and Improving Decisions.....	6
Current Trends and Academic Models.....	6
Organisation data warehousing in the Organisational applications.....	7
Data Warehouse (DW).....	7
Improved data can enhance data access and efficiency.....	7
Strategic Decisions.....	7
Data Mining Techniques.....	8
Regression.....	8
Classification.....	8
Audience to Organisational Goals.....	8
Predictive Analysis.....	10
Predictive analysis tools: Power BI.....	10
Features, Usability and Integration.....	10
Ethical and Data Security Considerations.....	11
GDPR with the Data Warehouse.....	11
Data ownership, consent, and transparency.....	11
Reporting Dashboard Solutions.....	12
BI dashboards.....	12
Considering Ethics and Data Security.....	16
GDPR and Legal Compliance:.....	16
Data Ownership and Consent:.....	16
Transparency:.....	16

Access Controls:.....	16
Internal and External Threats:.....	16
Conclusion.....	17
Key findings.....	17
Reference.....	18
Appendices.....	20

List of Figures

Figure 1:Sales forecasting	8
Figure 2: Forecasting	10
Figure 3: Power BI dashboard	12
Figure 4: Average rating	13
Figure 5: Income by product line	13
Figure 6: Gender distribution	14
Figure 7: Mode of Payment	14

Introduction

Tesco is a multinational corporation based in the United Kingdom that deals in general merchandise and supermarket business. It has millions of customers who shop in its physical stores and websites in Europe and Asia day in and day out.

This report aims to use the data warehousing and data mining methods on Tesco sales data to make significant business conclusions. It can transform the uncodified information into formulated information therefore, by their ability to convert the raw transactional data into the form that aids the management in making strategic choices such customer segmentation, sales forecast, and analysis of the product performance.

Under the data mining techniques, a series of key areas is included, such as design of data warehousing, data mining findings as well as interpretation. Each section of the paper assists in additional demonstration of how advanced analytics can be used to support the activities and marketing practices of Tesco.

Trends in BI, DW, and DM

Data Management Foundations (BI, DW, DM)

Business Intelligence (BI) refers to a grouping of methodologies, processes, architectures, and technologies that are intended to provide meaning to untamed data and change it to a consumable one that can be utilised in strategic decision-making.

Data Warehouse (DW) subject-oriented subject integrated, time-variant and non-volatile collection of data helping BI activities physically. It aggregates and stores historical data from disparate sources and cleans it up, and structures it as quickly as possible for analytical querying.

Data Mining (DM): Process of identifying patterns, anomalies and correlations in large data with the help of algorithms (e.g. classification, clustering). This type of analysis, brought about by the DM, is generally concerned with data extracted from the DW.

Data Consolidation and Improving Decisions

A Data Warehouse is important for the transformation of scattered data in the operational level into strategic insight through consolidation (Dibouliya, 2023). It uses the ETL (Extract, Transforms and Loads) process to integrate data from a siloed system so that data quality and consistency are ensured throughout the organisation.

Decision Enhancement: By providing this single view of history, the DW makes decision-making faster. Managers can immediately verify that "Cash" is the least popular "Mode of Payment", and plan on strategically promoting digital wallets, supported by their decision with the solid data from the DW.

Current Trends and Academic Models

Current Trends: The way towards Cloud Warehousing (e.g. Snowflake, Google Big Query) solves for large scales in an elastic way, allowing firms to deal with the huge volumes of data in the Total Number of Sales trends. A blend of Artificial Intelligence (AI) is automating more complex tasks, like the method by which data quality is improved to use predictive analytics to anticipate sales trends, illustrated by the shaded area of the graph.

Organisation data warehousing in the Organisational applications.

Data Warehouse (DW)

A Data Warehouse (DW) is a repository of historical and current data from the different operational systems (such as point-of-sale terminals and inventory databases) of an organisation, centralised and with subject-oriented data. This data is carefully cleansed, integrated, and transformed through the ETL/ELT process before being loaded. Its only goal is to be one single consistent source of truth optimised to support reporting, analysis and Business Intelligence (BI), as opposed to transactional systems whose purpose is to manage day-to-day operations (Nambiar and Mundra2022).

Improved data can enhance data access and efficiency.

Data Warehousing led to increased data access and increased operational efficiency. First, it provides a nature by bringing data from such siloed systems throughout the company. Instead of manually pulling the data, combining it, and viewing the "Average Rating" across all "Branches" (Giza, Alex, Cairo), analysts will be able to go to one central location and query the information, which will save a lot of time. Second is the optimisation of DW structure for high-performance query analytics on large historical data sets. This means performing a complex calculation for the Sum of gross income (15.38K), Nebula example across years does not affect the overall performance of live checkout systems (Andreadis, 2014).

Strategic Decisions

The DW is the strategic direction engine, giving executives the capacity to make decisions based on the data they have. It can perform historical and trend analysis by keeping data for large periods of time that leaders can use for measuring their long-term performance, and make predictions, such as forecasting the increase in the "Total Number of Sales" trend.

Sales Reporting

Tesco's DW includes all of the data from hundreds of thousands of stores and from the online channels of approximately seven countries, united in one single place and analytical source. This is fed into the dashboards, which have an in-depth analysis and are run daily, such as using a batch of important metrics, such as "Sum of gross income", for example, and segmented analyses of the customers, such as delayed payments per customer in the following example ", Customer Gender Distribution".

Data Mining Techniques

Regression

The regressed line that is shown by the Total Number of Sales in Tesco charts shows a positive linear relationship between the quantity sold and total sales. The higher the amount, the higher the total number of sales, which indicates good sales performance with an increase in product flow. This form of regression analysis is useful for determining trends in sales and can be used to predict sales in the future (Tyagi, Rane and Manry, 2022).



Figure 1: Sales forecasting

Classification

The classifier in this dashboard can be used to classify sales as well as customer data into meaningful categories for analysis. This helps to identify the behaviour patterns - e.g. which gender or branch is contributing most to sales, or which payment mode is preferred. Overall, it assists in strategically making decisions by converting to ruthless crafter categories and inspiration.

Audience to Organisational Goals.

Regression is important for the supermarket to predict sales in the future according to quantities sold so that they can accurately plan, forecast sales, manage their inventory control and try to improve their profitability. Classification helps segment customers based on

gender, branch, or payment mode to help the organisations target different marketing strategies and enhance customer satisfaction.

Power BI is used as both a platform and a tool in the above regression and classification process, which helped us to predict future sales with the help of forecasting.

Predictive Analysis

Predictive analysis tools: Power BI

The predictive analysis tool available in Power BI allows companies to predict trends, forecast the future, and make decisions using data. In the case of such sales data, tools like the Power BI Forecasting, the Azure ML integration, and Python/ R scripting are very appropriate.

Power BI Forecasting, in its turn, is an inbuilt tool based on exponential smoothing algorithms to predict future values based on past trends (Mahmud and Ikbal 2024). The Graph below shows that the total sales will surpass 100K in the future with 95% accuracy. The orange line in the graph is the line of forecast (Palanisamy, 2022).



Figure 2: Forecasting

Features, Usability and Integration

Report Automation and Forecasting are driven by the promotion of Business Intelligence (BI) tools as their main contributor. The BI layer (on top of the DW) can be used for self-service analytics and dashboard creation that automates the distribution of insights without the need for the stakeholders to manually compile them. For forecasting, DM algorithms are used to exploit the data that is saved in the DW for previous cycles. The future trend graph in "Total Number of Sales" is a direct result of DM's predictive analytics models - statistical analytics

analysis applied to the non-volatile time-series DW database (Eden, Chisom and Adeniyi, 2024).

Ethical and Data Security Considerations

GDPR with the Data Warehouse

The whole of the information will be required to be done in a strict, disciplined way to comply with the provisions of GDPR. The owner of the data is the individual (data subject), and his/her rights are transferred to the individual. Maximum transparency is indispensable - the view of the consent to process data and the intention of the entire processed data, databases behind the analytical dashboard (Allahrakha, 2023).

Data ownership, consent, and transparency

Endless efforts are needed to carry this out. Access Controls - Role-Based Access Control (RBAC) must take steps to safeguard any emanating protruding sensitive PII inside the BI system from internal threats (such as employee misuse, malpractice, and employees with designated saved inclusion to sensitive PII) other than due to their contracting to their role. Access controls, encryption, and other safeguards (Damgård, Haagh and Orlandi, 2016).

These technical security countermeasures, which support organisational processes including continuous auditing and penetration testing, are key to the battle against external threats.

Reporting Dashboard Solutions.

BI dashboards

Power BI dashboards offer real-time monitoring and decision-making decision making by way of providing a centralised platform, which is visual and interactive. They enable managers to monitor the business key performance indicators (KPIs) in real-time so that appropriate decisions made using the data can be implemented immediately.

The Sum of Gross Income (15.38K) and the Total Number of Sales are used to assess the international finance performance and profitability in general.



Figure 3: Power BI dashboard

Metrics/KPIs

The chart depicts average ratings by customers in each branch. Giza has the highest satisfaction, followed by Cairo and Alex.



Figure 4: Average rating

The graph/networks products based on their income, and Food and Beverages are the products with the highest income, while Health and Beauty are the lowest ones.

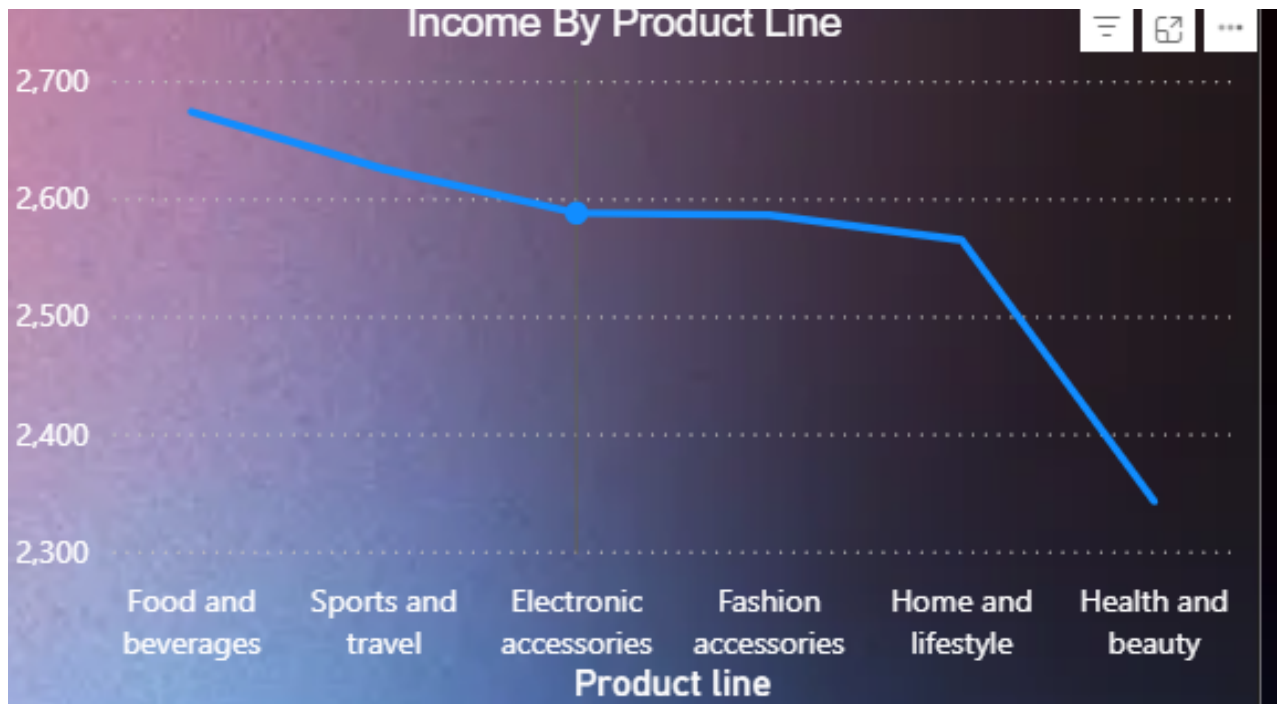


Figure 5: Income by product line

Female customers were more numerous than the total number of customers targeting all occupational groups, and especially those in Giza; male customers were less numerous but equally distributed between all three branch locations.

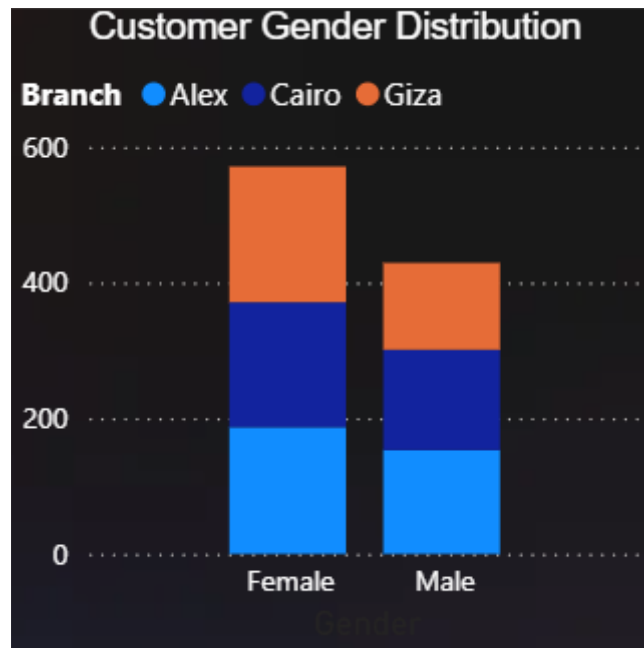


Figure 6: Gender distribution

The Pie Chart shows that Cash was the mode, which was mostly used by the customers, 34.74 % and Ewallet was used by 34.06% and Credit Card was the least used among all, which was 31.2%

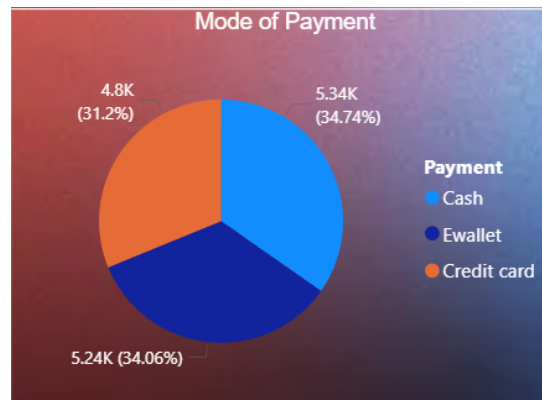


Figure 7: Mode of Payment

The Time index graph shows the Total Number of Sales as a function of an unknown time or quantity index. It is modelled with a general upward trend, with the future pathway indicated by the shaded area.

This performance process data element represents the total gross income, which amounts to \$15,380. This value is a total of pre-Dorfman-Likely profits before subtracting the cost of goods sold or outlays.

This interactive feature provides the capability for the user to filter the data being displayed on the dashboard. The current view covers the period Jan 1, 2019, to Dec 3, 2019.

Considering Ethics and Data Security

GDPR and Legal Compliance:

Users involved in predictive analysis with Power BI must respect such data protection regulations as the local acts on data privacy. This guarantees that individual or confidential customer information is handled legally, reasonably, and openly (Lukic, Miller, and Skiera, B. 2024).

Data Ownership and Consent:

Information gathered should strictly be utilised through the express permission of customers or owners of data. The organisation should be very clear on who owns the data and there should be policies that will limit its usage for other purposes (Dhirani et al., 2023).

Transparency:

Organisations are required to inform the users of the way their data is being collected, analysed, and stored (TINGHE, 2023). Clear reporting establishes confidence and agrees with the principle of ethical AI and analytics.

Access Controls:

The role-based access control (RBAC) is supported by Power BI, and only authorised personnel can access or manipulate sensitive data.

Encryption and Safeguards:

End-to-end encryption should be used in the protection of data, during transit and at rest. The security gateways and compliance certifications make it more secure.

Internal and External Threats:

Firewalls, multi-factor authentication, and regular security updates should be used to mitigate external cybersecurity threats like hacking or phishing.

Conclusion

Key findings

The analysis shows critical insights into sales performance, customer behaviour and revenue in different branches with different product lines. The average rating of customer ratings is the highest in Cairo, while the flavour of food and beverages generates the highest income. Strategic importance

Business Intelligence (BI) tools such as Power BI have improved the performance of businesses by leaps and bounds, changing raw data into actionable insights (Selvarajan, G.P, 2023).

Regularly update datasets, blend predictions to forecast demand and apply interactive dashboards for departmental performance reviews for continuous improvement in business.

Reference

- Allahrakha, N., (2023) 'Balancing cyber-security and privacy: legal and ethical considerations in the digital age', *Legal Issues in the digital Age*, (2), pp.78-121.
<https://cyberleninka.ru/article/n/balancing-cyber-security-and-privacy-legal-and-ethical-considerations-in-the-digital-age>
- Dhirani, L.L., Mukhtiar, N., Chowdhry, B.S. and Newe, T., (2023) 'Ethical dilemmas and privacy issues in emerging technologies: A review', *Sensors*, 23(3), p.1151.
<https://www.mdpi.com/1424-8220/23/3/1151>
- Dibouliya, A. (2023) 'Review on: Modern Data Warehouse & how is it accelerating digital transformation', *International Journal of Advance Research, Ideas and Innovations in Technology*, 9(2), pp. 912-1224.
https://www.researchgate.net/profile/Ashish-Dibouliya/publication/377399166_Review_on_Modern_Data_Warehouse_how_is_it_accelerating_digital_transformation/links/65a49eddc77ed940477852ff/Review-on-Modern-Data-Warehouse-how-is-it-accelerating-digital-transformation.pdf
- Eden, C.A., Chisom, O.N. and Adeniyi, I.S., (2024) 'Integrating AI in education: Opportunities, challenges, and ethical considerations', *Magna Scientia Advanced Research and Reviews*, 10(2), pp.006-013.
https://www.researchgate.net/profile/Nneamaka-Onyebuchi-4/publication/379428577_Integrating_AI_in_education_Opportunities_challenges_and_ethical_considerations/links/67e2c62fe0f5a760f90195a/Integrating-AI-in-education-Opportunities-challenges-and-ethical-considerations.pdf
- Lukic, K., Miller, K.M. and Skiera, B. (2024) 'The impact of the general data protection regulation (gdpr) on online tracking'. *Available at SSRN 4399388*.
<https://www.tandfonline.com/doi/pdf/10.1080/19460171.2021.1927776>
- Mahmud, D. and Ikbali, M.Z. (2024) 'Power BI and data analytics in financial reporting: a review of real-time dashboarding and predictive business intelligence tools', *International Journal of Scientific Interdisciplinary Research*, 5(2), pp.125-157.
<https://ijsir.org/index.php/IJSIR/article/download/18/18>

Nambiar, A. and Mundra, D. (2022) 'An overview of data warehouse and data lake in modern enterprise data management', *Big data and cognitive computing*, 6(4), p.132.

<https://www.mdpi.com/2504-2289/6/4/132>

Palanisamy, S. (2022) 'Predictive analytics with data visualization', *Journal of Ubiquitous Computing and Communication Technologies*, 4(2), pp.75-96.

<https://irojournals.com/jucct/V4/I2/03.pdf>

Selvarajan, G.P. (2023) 'Augmenting Business Intelligence with AI: A Comprehensive Approach to Data-Driven Strategy and Predictive Analytics', *International Journal of All Research Education and Scientific Methods*, 11(10), pp.2121-2132.

https://www.researchgate.net/profile/Guru-Selvarajan/publication/385782587_Augmenting_Business_Intelligence_with_AI_A_Comprehensive_Approach_to_Data-Driven_Strategy_and_Predictive_Analytics/links/67351de7f255d5728669eab3/Augmenting-Business-Intelligence-with-AI-A-Comprehensive-Approach-to-Data-Driven-Strategy-and-Predictive-Analytics.pdf

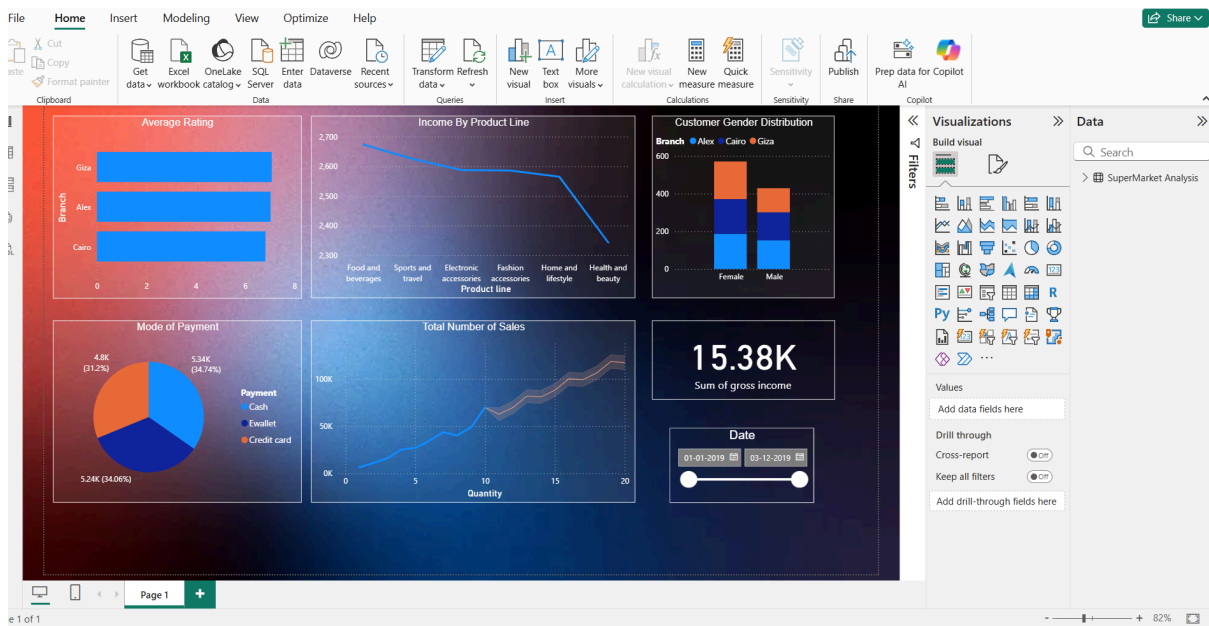
TINGHE, W. (2023) 'STUDY ON COST CONTROL IN THE CONSTRUCTION INDUSTRY BASED ON VALUE CHAIN THEORY---TAKE CHANG'AN CONSTRUCTION COMPANY AS AN EXAMPLE' (Doctoral dissertation, SIAM UNIVERSITY).

<https://e-research.siam.edu/wp-content/uploads/2023/11/MBA-2023-IS-Wei-Tinghe-6417195034-Study-on-Cost-Control.pdf>

Tyagi, K., Rane, C. and Manry, M. (2022) 'Regression analysis'. In *Artificial intelligence and machine learning for EDGE computing* (pp. 53-63). Academic Press.

https://www.researchgate.net/profile/Kanishka-Tyagi/publication/367711473_Regression_analysis/links/63f89ab40d98a97717b51761/Regression-analysis.pdf

Appendices



FileHomeHelpTable tools

NameSuperMarket Analy...

Manage relationships

New

Quick

New

New

Mark as date table

measure

measure column

Calculations

Calendars

Share

Structure

×

✓

Invoice ID

Branch

City

Customer type

Gender

Product line

Unit price

Quantity

Tax 5%

Sales

Date

Time

Payment

costs

gross margin

Search

SuperMarket Analysis

299-46-1805	Cairo	Mandalay	Member	Female	Sports and travel	93.72	6	28.116	590.436	01 January 2019	31-12-1899 16:19:00	Cash		562.32
145-94-9061	Cairo	Mandalay	Member	Female	Food and beverages	88.26	5	22.09	463.89	01 January 2019	31-12-1899 19:48:00	Cash		441.8
149-71-6266	Cairo	Mandalay	Member	Female	Sports and travel	78.07	9	35.1315	737.7615	01 January 2019	31-12-1899 12:43:00	Cash		702.63
817-48-8732	Alex	Yongon	Member	Female	Home and lifestyle	72.35	10	36.175	759.675	01 January 2019	31-12-1899 15:55:00	Cash		723.5
232-11-3025	Alex	Yongon	Member	Female	Sports and travel	78.77	10	39.385	827.085	01 January 2019	31-12-1899 10:04:00	Cash		787.7
756-01-7507	Giza	Naypyitaw	Member	Female	Health and beauty	20.38	5	5.095	106.995	01 January 2019	31-12-1899 18:56:00	Cash		101.9
565-80-5980	Giza	Naypyitaw	Member	Female	Home and lifestyle	47.38	4	9.476	196.996	01 January 2019	31-12-1899 10:25:00	Cash		186.52
870-54-3162	Alex	Yongon	Member	Female	Sports and travel	32.25	5	8.0625	169.3125	01 January 2019	31-12-1899 13:26:00	Cash		161.25
191-10-6171	Cairo	Mandalay	Normal	Female	Fashion accessories	39.62	7	13.867	291.207	01 January 2019	31-12-1899 13:18:00	Cash		277.34
280-17-4359	Giza	Naypyitaw	Member	Male	Health and beauty	90.5	10	45.25	950.25	01 January 2019	31-12-1899 13:48:00	Cash		905
400-89-4171	Giza	Naypyitaw	Normal	Female	Sports and travel	80.97	8	32.388	680.148	01 January 2019	31-12-1899 13:05:00	Cash		647.76
234-65-2137	Giza	Naypyitaw	Normal	Male	Home and lifestyle	95.58	10	47.79	1003.59	01 January 2019	31-12-1899 13:32:00	Cash		955.8
796-32-9050	Alex	Yongon	Normal	Male	Food and beverages	51.28	6	15.384	323.064	01 January 2019	31-12-1899 16:31:00	Cash		307.68
434-83-9547	Giza	Naypyitaw	Member	Male	Food and beverages	38.47	8	15.388	323.148	01 January 2019	31-12-1899 11:51:00	Cash		307.76
556-86-3144	Giza	Naypyitaw	Member	Female	Fashion accessories	74.29	1	3.7145	78.0045	01 January 2019	31-12-1899 19:30:00	Cash		74.29
426-39-2418	Giza	Naypyitaw	Normal	Male	Electronic accessories	61.41	7	21.4935	451.3635	01 January 2019	31-12-1899 10:02:00	Cash		429.87
677-11-0152	Giza	Naypyitaw	Normal	Female	Food and beverages	93.26	9	41.967	881.307	01 January 2019	31-12-1899 18:08:00	Cash		839.34
289-65-5721	Cairo	Mandalay	Normal	Female	Fashion accessories	81.37	2	8.137	170.877	01 January 2019	31-12-1899 19:28:00	Cash		162.74
530-90-9855	Alex	Yongon	Member	Male	Home and lifestyle	47.59	8	19.036	399.756	01 January 2019	31-12-1899 14:47:00	Cash		380.72
110-48-7033	Cairo	Mandalay	Member	Male	Fashion accessories	32.62	4	6.524	137.004	01 January 2019	31-12-1899 14:12:00	Cash		130.48
633-91-1052	Alex	Yongon	Normal	Female	Home and lifestyle	12.03	2	1.203	25.263	01 January 2019	31-12-1899 15:51:00	Cash		24.06
226-71-3580	Giza	Naypyitaw	Normal	Female	Sports and travel	23.75	9	10.6875	224.4375	01 January 2019	31-12-1899 12:02:00	Cash		213.75
316-68-6352	Alex	Yongon	Member	Female	Food and beverages	36.36	2	3.636	76.356	01 January 2019	31-12-1899 10:00:00	Cash		72.72
307-83-9164	Alex	Yongon	Member	Male	Home and lifestyle	60.01	4	12.002	252.042	01 January 2019	31-12-1899 15:54:00	Cash		240.04
779-06-0012	Giza	Naypyitaw	Member	Female	Home and lifestyle	88.61	1	4.4305	93.0405	01 January 2019	31-12-1899 10:21:00	Cash		88.61
532-59-7201	Cairo	Mandalay	Member	Male	Sports and travel	79.93	6	23.979	503.559	01 January 2019	31-12-1899 14:04:00	Cash		479.58
623-20-1945	Giza	Naypyitaw	Normal	Female	Health and beauty	39.42	1	1.971	41.391	01 January 2019	31-12-1899 15:08:00	Cash		39.42

Table: SuperMarket Analysis (1,000 rows)

