IJITEB
Vol. 7
No. 2
2025

International Journal of Information Technology and Business

http://ejournal.uksw.edu/ijiteb



Sales Data Analysis and Visualization for Distribution Optimization (Case Study: US Candy Distribution)

Riska Indarwati[1*], Herti Yani[2], Beny Beny[3]

[1-3] Faculty of Computer Science, Information Systems, Universitas Dinamika Bangsa, Indonesia

Abstract : Optimization of US Candy's product distribution was conducted through sales data analysis and visualization. The main challenges identified include less strategic factory locations and high logistics costs. This research used a quantitative approach. The purpose of this analysis is to analyze the distribution pattern. Data visualization was conducted using Power BI to show the results of the analysis that New York is the highest selling city with total sales of $10,945.89, with the leading product being Wonka Bar - Scrumdiddlyumptious ($2,718). The consistent increase in sales trend in the fourth quarter of each year, triggered by celebrations such as Halloween, special promotions, and holidays can be utilized to develop future marketing and product inventory strategies. A distribution optimization strategy is recommended to minimize logistics costs and delivery time, by moving distribution to the Wicked Choccy's factory with a distance of 8237.2 km. Marketing strategies such as discounting best-selling products and bundling system for low-selling products are proposed to improve market competitiveness. The resulting interactive dashboard implementation is expected to assist the company in making more accurate and efficient data-based decisions.

Keywords :
Data Analysis, Data Visualization, Power BI, Distribution, Logistic Efficiency.

1. Introduction

1.1 Background

Efficient distribution of goods is a crucial aspect in the retail industry, especially in the consumer products sector such as candy. US Candy is a candy distribution network in the United States that covers various regions with diverse needs and demands. Effective distribution management plays an important role in ensuring products are available at strategic locations, such as retail stores, supermarkets, and seasonal events, thereby increasing customer satisfaction as well as company profitability [1].

In practice, distribution companies often face complex operational challenges. One of the main challenges is the wide geographical coverage that leads to increased delivery times and logistics costs. In addition, imbalances in the placement of distribution facilities can hamper product availability in high-demand regions [2]. To overcome these problems, delivery route optimization is an indispensable strategy to improve distribution efficiency. Route optimization involves determining the most effective delivery path based on factors such as distance, cost, traffic conditions, and vehicle capacity, with the aim of reducing travel time as well as operational costs [3].

On the other hand, advances in information technology have opened up new opportunities in supply chain management. The use of digital systems enables the integration of business flows, increases operational transparency, and supports more accurate data-based decision-making [4]. However, many companies still rely on manual methods in operational data processing, which hinders effective analysis of distribution patterns. Difficulties in identifying sales trends, evaluating product performance, and understanding consumer preferences often lead to suboptimal decision-making [5].

Therefore, this research aims to design an interactive dashboard-based analysis system that is able to process information visually and comprehensively. The implementation of this dashboard is expected to help companies recognize optimal distribution patterns, determine priority areas based on demand levels, and develop more effective strategies in managing distribution and reducing logistics costs.

1.2 Literature Review

1.2.1 Previous Research

Previous research by Qodri and Alijoyo discussed Fast Fashion Supply Chain Analysis Using Exploratory Data Analysis with Power Business

Intelligence. The challenge faced regarding the fast fashion supply chain is the need for in-depth understanding and comprehensive analysis of the fast fashion supply chain. the method used in previous research was Exploratory Data Analysis (EDA) based on Power Bi with a Knowledge Discovery in Database (KDD) approach that allows the results of analyzing information related to the fast fashion supply chain [6].

Najib and Stefany in their research produced data visualization using Power BI, including sales trends, customer consumption patterns, best-selling product categories, and purchase patterns based on time. Based on data visualization, analysis results are obtained to increase operational efficiency, improve customer service quality, and optimize marketing strategies. This research adds to the literature on the application of Business Intelligence (BI) technology in the retail sector and provides practical guidelines for supermarket management to improve competitiveness [7].

Tjahyono and Susilowati in their research included several steps starting with data collection and interviews, data transformation with the ETL method, report visualization. The output results of this study are in the form of visualization charts in Power BI such as total sales and distribution of goods in 3 years, top 5 stores with the most purchases each year, and other outputs according to the needs of UD Anugerah Sejahtera Jaya [8].

Ratna and Dewi in their research conducted by processing data using excel with the Exploratory Data Analysis (EDA) method to identify sales trends and products with the highest sales in each region. After analyzing and visualizing the data, it is found that one of the effective strategies to increase profits by 15% is to apply the right discount policy [9].

Sadewo et al. conducted research on Visual Analysis of Job Type Distribution in Bandung City Using Power BI. The method used in this research is visual analysis, which combines visualization and data analysis techniques. Data visualization shows the occupational distribution of Bandung City residents through bar charts and pie charts, with self-employed as the largest group [10].

1.2.2 Analysis

Analysis is the process of breaking data into smaller components based on certain elements and structures [11]. the purpose of analysis is to recognize a number of data obtained from a certain population in order to obtain conclusions that are used to determine policies or make decisions as a step in providing problem solving for a problem [12]. With data-based decision making, it allows organizations to make more informed and effective decisions [13].

1.2.3 Optimization

Optimization is the determination of the level of use of resources or inputs that can reduce costs as low as possible and produce maximum profit or income [14]. The purpose of optimization is to minimize the effort required or to maximize the desired results, and to define the process, it is important to imply the function of the decision variables [15]. The benefits of optimization include cost savings, increased productivity, competitive advantage, and increased customer satisfaction [16].

1.2.4 Distribution

Distribution is the distribution or distribution or availability of products in an area or sales area, where this distribution can be used as an indicator of sales potential in the area [17]. There are several types of distribution, namely direct distribution, semi-direct distribution, and indirect distribution [18]. The purpose of this distribution is to ensure that the product reaches its destination on time, in good condition, and with efficient shipping costs [19].

1.2.5 Data Visualization

Data visualization is the process of transforming data that is generally in tabular form into visual images [20]. Visualization aims to communicate information clearly and effectively through graphical elements [21]. Types of commonly used data visualizations include bar charts, line charts, pie charts, histograms, box plots, maps, scatter plots, and heatmaps [22].

1.2.6 Power BI

Power BI is a Business Intelligence (BI) tool or platform that offers tools to collect, analyze, visualize, and share data. Power BI is made to visualize data with three main characteristics, namely output (visual data) is attractive to look at, easy to understand, and easy to interpret [23]. Power BI has several advantages, namely sharing data, real time dashboards, and can process original data that exceeds the capacity of other applications [24].

2. Research Method

2.1 Research Stages

Based on the research method above, it can be described as follows:

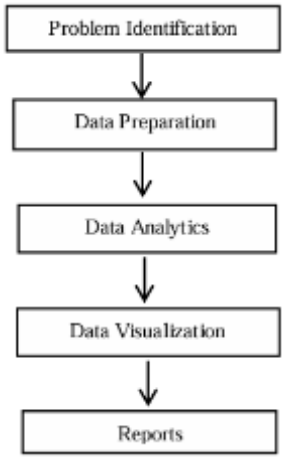


Fig. 1. Research Framework

1. Problem Identification

The first stage of this research was problem

identification, which aimed to understand the main challenges in US Candy's distribution. Analysis was conducted to identify inefficiencies, particularly in the selection of delivery routes, in order to design a more efficient and effective solution.

2. Data Preparation

Data preparation is a crucial step in analysis to ensure data quality and consistency. This stage consists of data understanding and data cleaning. Data understanding starts with understanding the content of the data starting from identifying the data type in each variable, the number of rows and columns, and the purpose of further analysis. Furthermore, the data cleaning stage starts by combining US Zips to Candy Sales data by marking unnecessary data and seeing if there is duplicate data or empty data. From the US Candy Distribution dataset, no duplicate data was found but there was some empty data only in the US Zips table which was then deleted because it was not needed in the analysis. Next we removed Canada from the country column and removed outliers based on sales. With these steps, the data can be analyzed more accurately and reliably.

3. Data Analytics

Data analysis is the process of collecting, organizing, and analyzing data to gain insight, make predictions, and support decision making. This research focuses on analyzing the improvement of distribution efficiency to cities with the highest sales with the selection of areas based on a contribution of 10% of annual sales. then calculate the distance data between the factory and the destination city using the Vincenty method using Python and the geopy library using google colab. The results of this calculation are used as the basis for distribution analysis to identify the optimal strategy to improve delivery efficiency.

4. Data Visualization

Data visualization is the process of presenting data in graphical form, such as charts, diagrams, or maps using Power BI tools, to make it easier to understand patterns, trends, and relationships in data. With visualization, complex information can be simplified so that it is easier to analyze and use in decision making.

5. Reports

The final stage of the research was the production of a report that systematically organized the findings, analysis, and recommendations. Data visualization was used to support understanding, with clear and easy-to-understand language.

3. Results and Discussion

3.1 Analysis Results

3.1.1 Sales and Products

Sales analysis was conducted to identify the number of sales in the 5 cities with the highest sales. The results of this analysis are presented in the following table:

Table 1. Highest Selling City

City	SUM of Cost	SUM of Sales	SUM of Gross Profit	SUM of Sales
New York City	3643.36	10945.89	7302.53	8.90%
Los Angeles	3101.72	9371.55	6269.83	7.62%
Philadelphia	2146.24	6369.27	4223.03	5.18%
San Francisco	2060.68	6218.24	4157.56	5.06%
Seattle	1693.78	5105.48	3411.7	4.15%

Based on available data, New York City has the largest contribution to the company's total sales, with sales value reaching $10,945.89 and gross profit of $7,302.53, which represents 8.90% of total sales. The second position is occupied by Los Angeles with total sales of $9,371.55 and gross profit of $6,269.83, or equivalent to 7.62% of total sales. Furthermore, Philadelphia, San Francisco, and Seattle also made significant contributions, each contributing more than 4% to total sales. Overall, these five cities contributed 30.91% to the company's total sales. This finding indicates that optimizing distribution in these cities has the potential to significantly improve logistics

City	Product Name	SUM of Units
New York City	Wonka Bar - Scrumdiddlyumptious	755
	Wonka Bar - Milk Chocolate	644
	Wonka Bar - Fudge Mallows	553
	Wonka Bar - Nutty Crunch Surprise	539
	Wonka Bar - Triple Dazzle Caramel	528
	Wonka Bar	49
	Kazookles	41
	Nerds	9
	SweeTARTS	5
	Hair Toffee	5
	Fun Dip	4
	Fizzy Lifting Drinks	4
	Laffy Taffy	2
	Lickable Wallpaper	1
New York City Total		3139
Los Angeles	Wonka Bar - Milk Chocolate	653
	Wonka Bar - Triple Dazzle Caramel	521
	Wonka Bar - Fudge Mallows	503
	Wonka Bar - Scrumdiddlyumptious	500

efficiency.

Next is the analysis on total sales with the number of products sold from 3 cities presented in the form of the following table:

Table 2. Number of Products Sold in Top 3 cities

	Wonka Bar - Nutty Crunch Surprise	450
	Wonka Gum	24
	Kazookles	19
	Hair Toffee	5
Los Angeles Total		2675
Philadelphia	Wonka Bar - Fudge Mallows	379
	Wonka Bar - Scrumdiddlyumptious	351
	Wonka Bar - Milk Chocolate	351
	Wonka Bar - Triple Dazzle Caramel	336
	Wonka Bar - Nutty Crunch Surprise	321
	Wonka Gum	64
	Kazookles	30
	Fizzy Lifting Drinks	5
	Laffy Taffy	2
	Lickable Wallpaper	1
Philadelphia Total		1840
Grand Total		**7.654**

Factory	*City*	AVERAGE of Distance Vincenty	SUM of Units	SUM of Cost
Wicked Choccy's	Los Angeles	8035.997799	1174	1421.72
	New York City	8237.209143	1172	1420.56
	Philadelphia	8230.658528	687	836.94
	San Francisco	7800.399981	763	932.38
	Seattle	7837.511196	613	746.98
Wicked Choccy's Total		8056.161156	4409	5358.58

The analysis shows that there are different product preferences in each city. Wonka Bar - Scrumdiddlyumptious has the highest demand in New York City, while Wonka Bar - Milk Chocolate is more dominant in Los Angeles. Philadelphia showed higher interest in Wonka Bar - Fudge Mallows. In addition, some products such as Nerds, Fun Dip and Lickable Wallpaper recorded very low sales figures, selling less than 10 units each. This suggests the need for special marketing strategies, such as bundling with popular products or providing discounts to increase the appeal of these products.

3.1.2 Factory Distance to Destination City

Calculating the distance between factories and destination cities is an important factor in analyzing distribution efficiency and logistics costs. The results of the analysis are shown in table 3.3, which shows the distance between the factory location and the five cities with the highest sales contribution.

Table 3. Factory Distance to Destination City

Sugar Shack	New York City	259.4654936	24	22.3
	Philadelphia	249.5333883	7	9
	San Francisco	312.9800158	4	3.2
Sugar Shack Total		262.8305248	35	34.5
Lot's O' Nuts	Los Angeles	10904.19852	1453	1603.6
	New York City	11102.10675	1847	2033.1
	Philadelphia	11096.08061	1051	1161.9
	San Francisco	10673.90784	974	1065.7
	Seattle	10703.27928	815	896.4
Lot's O' Nuts Total		10932.98313	6140	6760.7
Grand Total		18,989.144286	10584	12.153,78

Based on the results of the distance analysis between factories and destination cities, it shows that Lot's O' Nuts is the factory with the largest distribution volume (6,140 units), despite having the farthest average distance of 10,932 km. In contrast, Sugar Shack is the factory with the closest distance (263 km on average), but only distributes 35 units. This suggests that factories with large production capacity such as Lot's O' Nuts remain a mainstay despite their high distribution costs, while closer factories need to be optimized to support cost efficiency.

3.1.3 Shipping Mode

To determine which shipment is the most efficient, an analysis was conducted based on the data presented in the following table:

Table 4. Shipping Mode Analysis

City	*Ship Mode*	COUNTA of Ship Mode	SUM of Cost	AVERAGE of Cost
New York City	Standard Class	533	2208.84	4.144165103
	Second Class	162	645.44	3.984197531
	First Class	139	553.82	3.984316547
	Same Day	50	235.26	4.7052
New York City Total		884	3643.36	4.121447964
Los Angeles	Standard Class	431	1931.34	4.481067285

	Second Class	155	605.78	3.908258065
	First Class	109	466.52	4.28
	Same Day	28	98.08	3.502857143
Los Angeles Total		723	3101.72	4.290069156
Philadelphia	Standard Class	334	1351.02	4.044970006
	First Class	92	407.739	4.431521739
	Second Class	79	328.12	4.153417722
	Same Day	14	59.4	4.242857143
Philadelphia Total		519	2146.24	4.135337187
San Francisco	Standard Class	278	1163.42	4.184964029
	First Class	88	388.92	4.419545455
	Second Class	86	346.28	4.026511628
	Same Day	40	162.06	4.0515
San Francisco Total		492	2060.68	4.188373984
Seattle	Standard Class	237	1050.56	4.432742616
	Second Class	78	304.62	3.905384615
	First Class	58	206.7	3.563793103
	Same Day	35	131.9	3.768571429
Seattle Total		408	1693.78	4.151421569

The analysis of shipping mode data indicates that Standard Class is the most efficient method for the majority of shipments, with a low average shipping cost ($4.14–$4.5). Meanwhile, the Same Day mode has the highest average cost ($4.71) but is used selectively for shipments requiring high time precision. Optimizing the use of Standard Class can help the company reduce total operational costs.

3.2 Data Visualization

The processed and analyzed data is presented in an interactive dashboard that facilitates monitoring, evaluation, and decision-making regarding product distribution. The visualization results can be seen in the following figure:



Fig 2. US Candy Sales Dashboard

The US Candy Sales Dashboard was created using Power BI**,** providing visual and comprehensive information that enables users to identify sales trends, total sales by city, total sales of the top five best-selling products, total sales by division, as well as the average distance and distribution factories**.** This allows for easier understanding and analysis.

4. Conclusion

Based on the results of the analysis, New York City has the highest sales rate of $10,945.89, which indicates the high interest of customers in the product. Therefore, the focus of increasing sales needs to be directed to the other five cities with lower sales. Then there are Wonka Bar products – Scrumdiddlyumptious is the highest selling product, so a more intensive promotion strategy can further increase overall sales. In addition, the consistent trend of increasing sales in the fourth quarter of each year, driven by celebrations such as Halloween, special promotions, and holidays, can be leveraged for future marketing strategy planning and product inventory. As a recommendation, optimization of distribution and marketing strategies plays an important role in increasing sales of Wonka Bar – Scrumdiddlyumptious products. More efficient distribution, such as allocating production to Wicked choccy's plant closer to New York City, could maximize the market potential in the region. In addition, moving distribution to a plant closer to Seattle could also boost sales in the city with the lowest current sales rate. Marketing strategies through the provision of discounts on the best-selling products in each city as well as bundling systems on products with low sales rates proved effective in increasing consumer awareness and expanding market reach.

5. References

[1] R. Jamal, A. Ikhval, N. Nisa, S. Qulbi, and M. Arifin, "Penggunaan teknologi informasi dalam mengoptimalisasi supply chain management," *Jurnal Inovasi Global*, vol. 2, pp. 737–750, Jul. 2024, doi: 10.58344/jig.v2i7.117. [Online]. Available: https://doi.org/10.58344/jig.v2i7.117

[2] M. M. Abdurrahman and Y. Handayati, "Rute distribusi yang efisien dengan menggunakan metode saving matrix pada PT Asahimas Flat Glass Pabrik Cikampek."

[3] R. V. Martono, *Manajemen Logistik* (Edisi Revisi). Jakarta: Gramedia Pustaka Utama, 2023. [Online]. Available: https://books.google.co.id/books?id=P_L4EAAAQBAJ

[4] K. F. Waringga, F. D. Riana, and A. Aprilia, "Pengaruh penerapan teknologi informasi terhadap peningkatan keunggulan kompetitif pada usaha kedai kopi di Kota Bandung," *SEPA: Jurnal Sosial Ekonomi Pertanian dan Agribisnis*, vol. 19, no. 1, p. 31, Jul. 2022, doi: 10.20961/sepa.v19i1.51628. [Online]. Available: https://doi.org/10.20961/sepa.v19i1.51628

[5] E. R. Damanik and M. I. P. Nasution, "Evaluasi penggunaan teknologi big data untuk analisis data bisnis dan pengambilan keputusan."

[6] M. A. Qodri and F. A. Alijoyo, "Analisis rantai pasokan fast fashion menggunakan exploratory data analysis dengan Power Business Intelligence," *JATI: Jurnal Mahasiswa Teknik Informatika*, vol. 9, no. 1, pp. 1075–1082, Dec. 2024, doi: 10.36040/jati.v9i1.12629. [Online]. Available: https://doi.org/10.36040/jati.v9i1.12629

[7] M. K. Najib and E. M. Stefany, "Visualisasi data penjualan supermarket dengan Microsoft Power BI untuk menghasilkan insight dan rekomendasi," *Jurnal Sistem Informasi*, vol. 2, no. 12, 2024.

[8] H. V. Tjahyono, "Business intelligence data penjualan distributor kembang api di area Jawa Timur menggunakan Power BI," vol. 4, 2024.

[9] N. Ratna and S. Dewi, "Analisis dan visualisasi data dalam meningkatkan rata-rata profit penjualan produk rumah tangga di Amerika Serikat," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 1, Jan. 2025, doi: 10.23960/jitet.v13i1.5679. [Online]. Available: https://doi.org/10.23960/jitet.v13i1.5679

[10] A. D. Sadewo and F. Sulianta, "Analisis visual distribusi jenis pekerjaan di Kota Bandung menggunakan Power BI."

[11] Z. E. Hasibuan, *Metodologi Penelitian Pendidikan*. AE Publishing, 2024. [Online]. Available: https://books.google.co.id/books?id=sTIwEQAAQBAJ

[12] M. Taufiq and A. Vidya, *Analisis Sistem Informasi (Konsep, Metodologi, dan Pendekatan)*. Ananta Vidya. [Online]. Available: https://books.google.co.id/books?id=aXDQEAAAQBAJ

[13] N. B. Sukamdani and T. Sukwika, *Manajemen Talenta dan Technopreneur*. MEGA PRESS NUSANTARA, 2024. [Online]. Available: https://books.google.co.id/books?id=krISEQAAQBAJ

[14] B. Hartono and U. Press, *Prinsip Analisis Ekonomi: Teori dan Aplikasi di Bidang Peternakan*. Universitas Brawijaya Press, 2016. [Online]. Available: https://books.google.co.id/books?id=elxKDwAAQBAJ

[15] J. A. B. Sinaga and H. Cipta, *Optimasi Pengelolaan Hutan Berkelanjutan dan Terpadu*. CV. Gita Lentera, 2023. [Online]. Available: https://books.google.co.id/books?id=W83bEAAAQBAJ

[16] R. Jusdijachlan, A. Apriyanto, N. Nurfatma, E. Efitra, and I. Uzma, *Strategi Manajemen Efektif: Meningkatkan Kinerja Perusahaan di Era Digital*. PT. Sonpedia Publishing Indonesia, 2024. [Online]. Available: https://books.google.co.id/books?id=I-Y5EQAAQBAJ

[17] R. Alfisyahr and L. D. R. Devita, *Manajemen Penjualan*. Universitas Brawijaya Press, 2021. [Online]. Available: https://books.google.co.id/books?id=0UV_EAAAQBAJ

[18] M. S. I. Widaningsih, *Be Smart Ilmu Pengetahuan Sosial*. PT Grafindo Media Pratama. [Online]. Available: https://books.google.co.id/books?id=d9YOj8Ze-2AC

[19] K. H. Widodo, *Manajemen Rantai Pasok Pangan Strategis Disertai Implementasinya*. CV. Oxy Consultant. [Online]. Available: https://books.google.co.id/books?id=ako4EQAAQBAJ

[20] J. Enterprise, *Analisis dan Visualisasi Data dengan Power BI dan Tableau*. Elex Media Komputindo, 2024. [Online]. Available: https://books.google.co.id/books?id=Y4f3EAAAQBAJ

[21] D. Pratiwi, *Visualisasi Data: Teori dan Penerapannya*. Uwais Inspirasi Indonesia, 2024. [Online]. Available: https://books.google.co.id/books?id=asj-EAAAQBAJ

[22] N. R. Silaen *et al.*, *Manajemen Sumber Daya Manusia (Manusia, Data dan Analisis)*. Penerbit Widina, 2021. [Online]. Available: https://books.google.co.id/books?id=YbhVEAAAQBAJ

[23] S. Santoso, *Cara Praktis dan Sistematis Menguasai Power BI*. Elex Media Komputindo, 2024. [Online]. Available: https://books.google.co.id/books?id=pOMeEQAAQBAJ

[24] N. A. Novianti, B. H. Irawan, N. P. Ariyanto, H. Widiastuti, F. Restu, and N. L. Arifin, "Penggunaan Power BI untuk pengolahan data non-conformance material," *Jurnal Teknologi dan Riset Terapan (JATRA)*, vol. 4, no. 2, pp. 52–57, Dec. 2022, doi: 10.30871/jatra.v4i2.3885. [Online]. Available: https://doi.org/10.30871/jatra.v4i2.3885