Report On

Online Retail Service Analysis using Big Data

Submitted in partial fulfillment of the requirements of the Course project in Semester VII of Final Year Artificial Intelligence and Data Science

by

Devashree Pawar (Roll No. 21)
Naina Roy (Roll No. 24)
Khushi Upadhyay (Roll No. 30)

Supervisor
Prof. Bhavika Gharat

University of Mumbai

Vidyavardhini's College of Engineering & Technology

Department of Artificial Intelligence and Data Science

(2023-24)

Vidyavardhini's College of Engineering & Technology
Department of Artificial Intelligence and Data Science

CERTIFICATE

This is to certify that the project entitled “Crime Detection Analysis using Big Data” is a bonafide work of "Devashree Pawar (Roll No. 21), Naina Roy (Roll No. 24), Khushi Upadhyay(Roll No. 30)" submitted to the University of Mumbai in partial fulfillment of the requirement for the Course project in semester VII of Final Year Artificial Intelligence andData Science engineering.

Supervisor

Prof. Bhavika Gharat

Dr. Tatwadarshi P. N.
Head of Department

Abstract

In today's digital age, e-commerce has become a significant driver of business growth, demanding a deep understanding of consumer behavior, trends, and sales performance. This project presents a comprehensive analysis of e-commerce sales data, harnessing the power of Big Data analytics and leveraging Microsoft Power BI as the primary tool for insights and visualization. The primary objective of this project is to improve the overall sales performance of an e-commerce platform by extracting valuable insights from the wealth of data it generates. We employ cutting-edge data collection techniques to gather information from multiple sources, including transaction records, website interactions, social media, and customer reviews. The data collected spans a wide range of metrics, including sales revenue, customer demographics, product popularity, and more.

Table of Contents

Chapter No		Title	Page No.
1		**Chapter # 1**	5
	1.1	Problem Statement	5
2		**Chapter # 2**	6
	2.1	Description and Working	6
	2.2	Software & Hardware Used	7
3		**Chapter # 3**	8
	3.1	Result	8
	3.2	Conclusion and Future Work	9
4		**Chapter # 4**	10
		References	10

Chapter # 1

1.1 Problem Statement:

E-commerce businesses are inundated with vast and diverse data, presenting several challenges. Extracting meaningful insights from this data for strategic decisions is difficult. Inefficient marketing strategies, inventory management issues, and intense competition further complicate matters. The project's goal is to utilize Big Data analytics and Power BI to tackle these challenges by providing actionable insights. This approach aims to enhance sales performance, optimize operations, and improve the customer experience, ultimately driving consistent revenue growth.

Chapter # 2

2.1 Description and Working:

The project aims to optimize the sales performance of an e-commerce platform. By collecting, integrating, and modeling data from various sources, including transaction records, website interactions, and customer reviews, the project employs Microsoft Power BI and Big Data analytics to provide actionable insights. The clean and structured data is analyzed using Data Analysis Expressions (DAX) to create custom calculations and visualizations, allowing users to interact with and explore the data dynamically. This project's goal is to improve inventory management, refine marketing strategies, and enhance customer experiences, ultimately leading to consistent revenue growth and a competitive edge in the e-commerce market.

Through the meticulous use of data-driven methodologies, e-commerce sales analytics project not only provides valuable insights but also fosters a culture of continuous improvement and adaptability within the e-commerce business landscape. By harnessing Power BI's robust capabilities, it equips e-commerce businesses to adapt swiftly to evolving customer behaviors and market dynamics, thereby positioning them to remain competitive and resilient in an ever-changing digital marketplace. The project's data-driven approach empowers organizations to make informed decisions, maximize revenue potential, and enhance customer satisfaction, solidifying their presence and profitability in the highly competitive world of online retail.

2.2 Software & Hardware used:

Software:

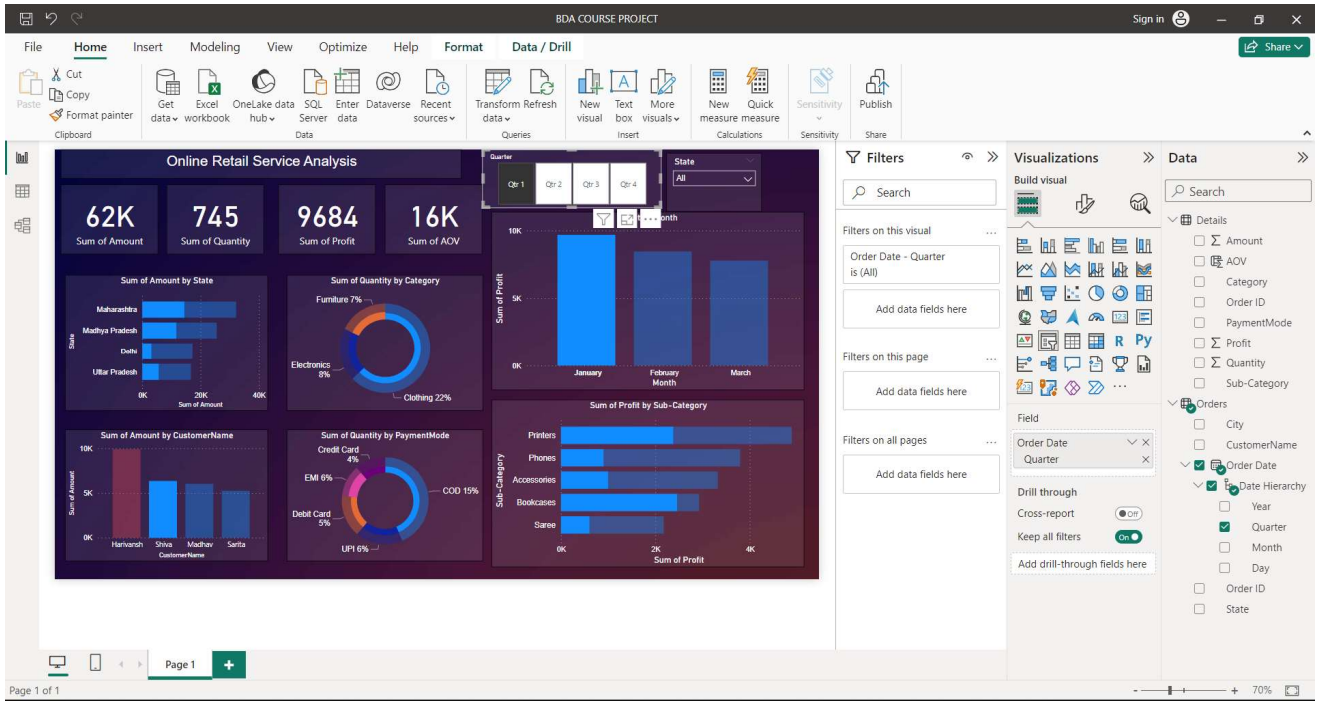
- Windows 10 OS
- Power BI

Hardware:

- 64 bit Operating System
- 6gb RAM
- Intel i5 processor

Chapter # 3

3.1 Results:



3.2 CONCLUSION AND FUTURE SCOPE:

In conclusion, this e-commerce sales analytics endeavor, driven by Power BI and Big Data analytics, is a catalyst for enhancing sales performance and bolstering e-commerce businesses in the fast-paced digital market. The project's success lies in its ability to harness diverse data sources, transform them into actionable insights, and produce dynamic, interactive visualizations. By offering real-time monitoring and data refresh capabilities, it equips organizations to remain agile and responsive to shifting market trends, thereby fostering revenue growth and reinforcing competitiveness.

Looking ahead, the project's future scope includes the potential for deeper integration with artificial intelligence and machine learning techniques to provide predictive analytics. By applying these advanced methodologies, e-commerce businesses can forecast sales trends, customer behavior, and inventory needs more accurately. Furthermore, expanding the project's capabilities to encompass additional data sources, such as social media sentiment analysis and external market data, can offer a broader perspective on consumer preferences and market dynamics. This ongoing evolution in analytics will continue to drive innovation, ensuring that e-commerce businesses maintain their edge in an ever-evolving digital landscape.

Chapter # 4

REFERENCES

[1] Charron, Mathieu. ”Neighbourhood characteristics and the distribution of police-reported crime in the city of Toronto. Statistics Canada , 2009.

[2] Kou, Gang, Yi Peng, and Guoxun Wang. ”Evaluation of clustering algorithms for financial risk analysis using MCDM methods.” Information Sciences 275 (2014): 1-12.

[3] Ding, Chris, and Xiaofeng He. ”K-means clustering via principal component analysis.” Proceedings of the twenty-first international conference on Machine learning. ACM, 2004.

[4] Pham, Duc Truong, Stefan S. Dimov, and Chi D. Nguyen. ”Selection of K in K-means clustering.”Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science 219.1 (2005): 103-119.

[5] Lloyd, Stuart P. ”Least squares quantization in PCM.” Information Theory, IEEE Transactions on 28.2 (1982): 129-137.

[6] A. Jain, R.C. Dubes,”Algorithms for clustering Data” in , Prentice Hall, 1988.

[7] Ester, Martin, et al.”A density-based algorithm for discovering clusters in large spatial databases with noise.” Kdd. Vol. 96. No. 34. 1996.

[8] Dudoit, Sandrine, and Jane Fridlyand. ”A prediction-based resampling method for estimating the number of clusters in a dataset.” Genome biology 3.7 (2002): research0036-1.

[9] Thalamuthu, Anbupalam, et al.”Evaluation and comparison of gene clustering methods in microarray analysis.” Bioinformatics 22.19 (2006): 2405-2412.

[10] Desgraupes, Bernard. ”Clustering indices.” University of Paris OuestLab ModalX 1 (2013): 34.