



INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

D2C INFOGRAPHIS

Yoshitha Sri H A, Dr.S.Dhanabal

Student,Professor

BE-CSE

Hindusthan College of Engineering and Technology, Coimbatore, India

Abstract: As India's digital and economic landscape evolves, a new breed of brands is taking centre stage: the Direct-to-Consumer (D2C) players. This info graphic dives deep into the heart of this dynamic ecosystem, visually mapping its explosive growth, key trends, and defining characteristics. We all delve into the challenges and opportunities these brands face, highlighting the fierce competition, logistical hurdles, and the quest for differentiation. Data has become prevalent to all, data is a variate in our life. Data analysis is the process of examining, filtering, adapting, and modeling data to help solve problems. Data analysis helps determine what is and isn't working, so you can make the changes needed to achieve your business goals. The biggest challenge for every D2C company is the shift of responsibility in distribution. Without large resellers to support your brand, you have to work harder to make your brand stand out. You are responsible for fulfillment, packaging, returns, and warehousing. This info-graphics is to know how the new engrossed brands affect our Country's economy and the number of brands that people opt to open business based on which category

Index Terms - Company, Algorithm, continuous, Visualization

I. INTRODUCTION

This project aimed to create an insightful and visually engaging info graphic exploring the burgeoning D2C brand landscape in India. Leveraging data and current trends, the info graphic seeks to inform and educate on the key players, challenges, and opportunities shaping this dynamic space. This info graphic provides a valuable resource for understanding the dynamic D2C landscape in India. By presenting data in a visually appealing and insightful way, it aims to stimulate further discussion, collaboration, and innovation within this exciting sector. The project aims to thoroughly examine the implications of these challenges and proposes innovative algorithms to navigate complexities associated with missing value generation, ultimately contributing to fairer outcomes in algorithmic decision-making. To evaluate the effectiveness of these proposed solutions, an extensive empirical analysis will be conducted using both genuine and synthetic datasets. While serving as an initial exploration, this project underscores the need for tailored strategies to address the intricacies of real-world data in algorithmic decision-making. These concepts promise to deepen our understanding of fairness in the evolving landscape of datasets in machine learning

II. D2C : DIRECT TO CONSUMER

D2C stands for direct-to-consumer. It's a business model where a company sells products directly to customers without using third-party retailers, wholesalers, or middlemen. D2C companies can offer products at lower prices than brands using traditional retail business models. The roots of D2C can be traced back to the earliest forms of commerce. Local farmers selling produce directly to consumers, artisans crafting and selling their wares, and traveling merchants all represent a direct connection between producers and consumers. With the Industrial Revolution, mass production led to the rise of wholesalers and retailers as intermediaries. This shifted the focus from local production and direct sales to large-scale manufacturing and distribution. Late 19th and 20th Century The invention of the printing press and advancements in postal services paved the way for mail-order catalogs. Companies like Sears, Roebuck and Co. offered a wider variety of goods directly to consumers through catalogs, bypassing traditional retail channels. This marked a resurgence of the D2C model. During mid 20th Century, Television commercials with direct calls to action and toll-free numbers further fueled the D2C approach. Companies like Slinky and Ginsu knives effectively used TV advertising to reach consumers directly and drive sales. During 1990, The emergence of the internet offered a major turning point. Online retailers like Amazon and Dell initially focused on a D2C strategy, leveraging the Internet's reach to bypass physical stores. However, the dot-com bubble burst in the early 2000s, leading many pure D2C companies to pivot or even disappear. The Internet's infrastructure matured, e-commerce platforms like Shopify simplified online store creation, and social media provided powerful marketing tools. This confluence of factors fueled a revival of D2C in the 21st century

III. DEVELOPMENT PHASE

During the development phase various data processing methods were used to process the data

3.1 Data Collection and Processing

Identifying the source of the dataset provides information about its origin, collection methods, and potential biases. This helps to assess the reliability and representativeness of the data. Understanding the data structure involves determining the organization and arrangement of the data elements. This includes identifying the data types, variables, and relationships between different data points. The data format refers to the way the data is stored and encoded. This information is essential for accessing and processing the data using appropriate software tools and programming languages: Data content encompasses the actual values, information, and meaning contained within the datasets. It involves understanding the range of values, distributions, and patterns within the data

3.2 Data Cleaning and Transformation

For this study,

1. Data Profiling: Examine the data to understand its structure, format, and distribution. This helps identify potential issues and areas for improvement. □
2. Identify and Handle Missing Values: Determine the extent of missing data and implement appropriate imputation techniques to fill in these gaps. □
3. Detect and Correct Errors: Identify and correct data entry errors, typos, outliers, and inconsistencies. This ensures the accuracy of the data. □
4. Handle Duplicates: Identify and remove duplicate records to maintain data integrity and prevent redundancy.
5. Standardization and Normalization: Standardize data formats, names, and units to ensure consistency and compatibility. Normalize data to a common scale for analysis. □
6. Data Validation: Validate the cleaned data to ensure it adheres to the expected quality standards and is ready for further analysis or modeling

3.3 Outliers And Data Analytics

Thoroughly understand the data, its characteristics, patterns, and normal behavior to identify anomalies effectively

Systematic risk is the only independent variable for the CAPM and inflation, interest rate, oil prices and exchange rate are the independent variables for APT model. Clearly define what constitutes an anomaly in the context of the data and the specific application. This definition may vary depending on the domain and the desired level of sensitivity. Assign anomaly scores to data points, indicating the degree to which they deviate from the expected behavior. This helps prioritize and focus on the most relevant anomalies.: Identify and characterize the different types of noise present in the data. This may involve analyzing data distributions, visualizing patterns, and examining outliers. Apply appropriate techniques to remove or reduce noise from the data. Common methods include data cleaning, outlier removal, smoothing algorithms, and noise filtering techniques.

Statistical Hypothesis Testing: Formulate hypotheses about the data and use statistical tests, such as t-tests, z-tests, and chi-square tests, to determine whether the hypotheses are supported by the data. Develop statistical models to represent the relationships between variables and use these models to make predictions or inferences about the data. Prepare the data for visualization by ensuring it is clean, consistent, and in a format suitable for visualization tools. Transform data into appropriate representations, such as normalized numerical values or continuous encoding

Equations

This study involves equations on identifying outliers:

$$IOP = O3 - O1$$

I. RESEARCH METHODOLOGY

The methodology section outlines the plan and method that how the study is conducted. This includes Universe of the study, sample of the study, Data and Sources of Data, study's variables and analytical framework. The details are as follows;

3.1 Sample

KSE-100 index is an index of 100 companies selected from 580 companies on the basis of sector leading and market capitalization. It represents almost 80% weight of the total market capitalization of KSE. It reflects different sector company's performance and productivity. It is the performance indicator or benchmark of all listed companies of KSE. So it can be regarded as universe of the study. Non-financial firms listed at KSE-100 Index (74 companies according to the page of KSE visited on 20.5.2015) are treated as universe of the study and the study have selected sample from these companies.

The study comprised of non-financial companies listed at KSE-100 Index and 30 actively traded companies are selected on the bases of market capitalization. And 2015 is taken as base year for KSE-100 index.

3.2 Data and Sources of Data

For this study secondary data has been collected. The data is collected from various online sources like Statistica, News and Trancxn. The data obtained are a decade info on many new brand companies till Jan, 2024.

3.3 Theoretical framework

Variables of the study contains dependent variable. The study used pre-specified method for the selection of variables. The study used the net profit as dependent variable. From the net revenue and annual abate of the company the net profit is calculated.

Programming language used would be python to do the analytical process and excel to process the data into a structured format.

Handling missing values by using various methods like Replacement or data imputation. Maintain Since the number of rows in the data set is small, replacing the null values is a viable option to handle the data. Imputing with min or max values or using replacement by statistic methods. Implementing analysis with and without outliers plays a major role in our queries, one such case analytic will be performed with and without to notice how it affects our analysis. Correlation between features helps us to use which features that better helps us in prediction, EXCEL - CORREL(),PYTHON- Use SEABORN for correlation. Multiple linear regression analysis is done by training and testing our data set, the data set is split into training and testing sets using Python ML- Scikit learn.

3.4 Machine Learning and Info-graphics

This section elaborates the proper Machine learning model which are being used to forward the study from data towards inferences. The detail of methodology is given as follows.

3.4.1 Machine Learning Model

Machine Learning helps to attain a proper result on the testing data. Linear regression is used to handle the business data. Correlation between features helps us to use which features that better helps us in prediction, PYTHON- Use SEABORN for correlation. After finding correlation between features the regression model is used where training and testing is done to predict the economic growth of the country

3.4.2 Info-graphics

Prepare the data for visualization by ensuring it is clean, consistent, and in a format suitable for visualization tools values or categorical encoding. Design visualizations that are clear, consistent, and visually appealing, adhering to design principles like clarity, hierarchy, and color theory. Transform data into appropriate representations, such as normalized numerical Annotate visualizations with clear labels and explanations to guide the viewer's understanding.

To make the model more effective and efficient the selection criteria for the shares in the period are: Data with no missing values in the period, Revenues with replacement on outliers or omit on it is necessary while working with machine learning algorithm. Info-graphics presents information in the visualized manner to understand. Mostly, the graph depicted are visualized using ML model.

IV. RESULTS AND DISCUSSION

4.1 Results of Info-graphic obtained after analysis using Machine Learning Model

Output 4.1 inform us about the category of the business most people opt for. These brands include: Personal Care, Brand Food, Based Company, Clothing Brand. The funding provided on the clothing based company is higher than any other brand. The Net Revenue and their GDP where the GDP amounts to how much a company can affect the economy of a country. The number of D2C startups in India has likely seen an exponential rise since 2014. This surge reflects the increasing internet penetration, growing disposable income, and a shift towards online shopping among Indian consumers. Fashion, beauty, and personal care are likely to be among the leading sectors in the Indian D2C space, driven by a young and fashion-conscious population. Other prominent sectors might include food and beverages, consumer electronics, and home ware. New-age sectors like organic products, sustainable fashion, and personalized wellness solutions could be gaining traction, catering to evolving consumer preferences

Output 4.2 inform us about the employee count and the increased net revenue of each brand. The number of categories used more is displayed in pie format. Overall, Price sensitivity remains a key factor for Indian consumers. D2C brands offering high-quality products at competitive prices are likely to see success. Social media platforms play a crucial role in brand discovery and purchase decisions. D2C brands leveraging social media effectively for marketing and customer engagement are likely to have an edge. We have explored different data and insight over their growth. The funding provided on the clothing based company is higher than any other brand. Through the decade, there has been a surmount aggrandize in D2C Sector.

This study has investigated the impact of missing data, noisy data, and selection bias on the fairness of machine learning models.



Figure 4.1

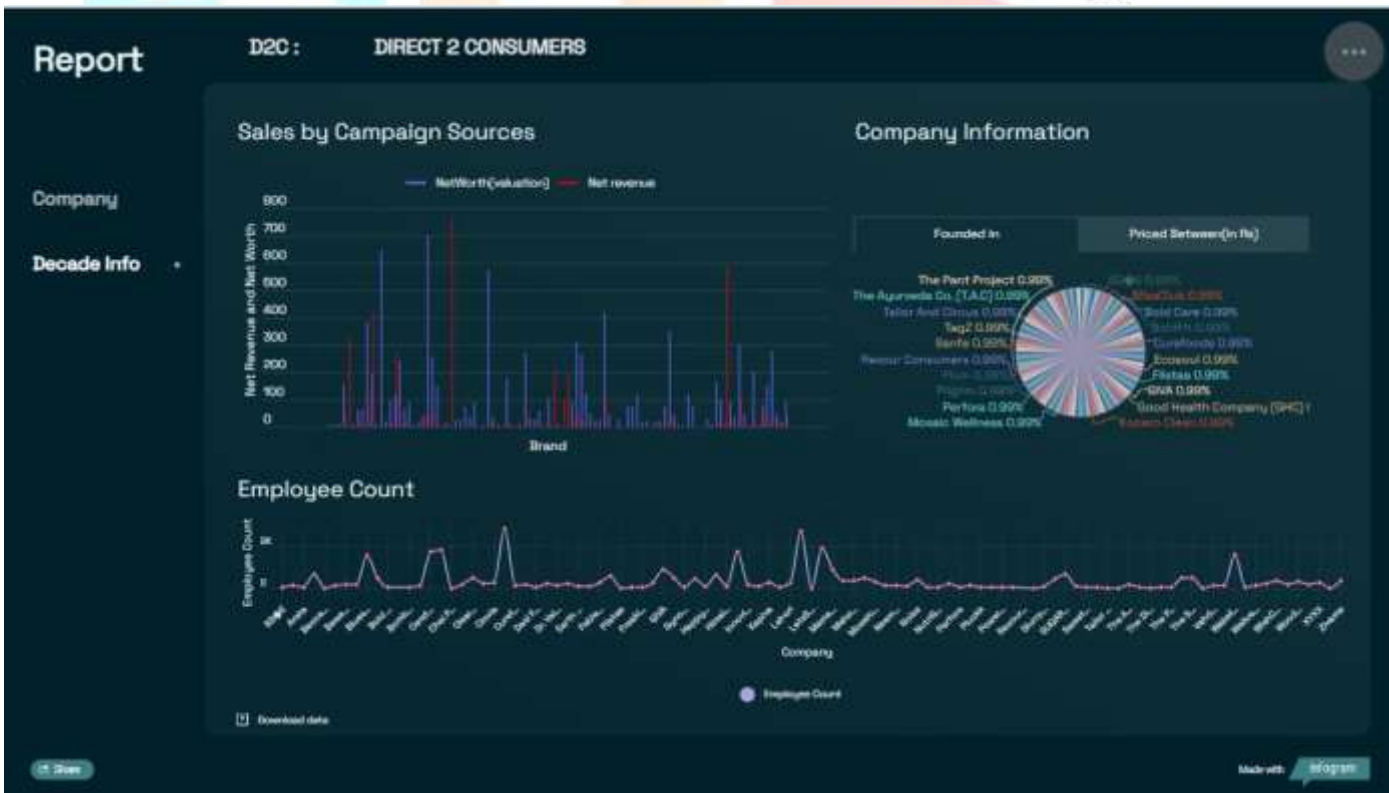


Figure 4.2

IV. ACKNOWLEDGMENT

We take this opportunity to express our wholehearted thanks and our profound respect to all those who guided and inspired us in the completion of this phase I project work. We extend our sincere thanks to the Founder and Chairman of Hindusthan Educational and Charitable Trust Shri. T. S. R. Khannaiyann and the Managing Trustee Smt. Sarasuwathi Khannaiyann and Executive Trustee Secretary Mrs Priya Satish Prabhu for providing essential infrastructure. We would like to reveal our profound thanks to our respected CEO Dr. K. Karunakaran and Principal Dr. J. Jaya who happens to be a striving force in all endeavours. We would like to express our gratitude to the Head of the Department Dr. Shankar S for bringing out the project successfully and for strengthening the ray of hope. We would like to express our sincere thanks and a deep sense of gratitude to our Guide Mr. Dr. S Dhanabal, M.E., Ph.D., Professor of Computer Science and Engineering Department.

REFERENCES

- [1] W. Deng, Y. Guo, J. Liu, Y. Li, D. Liu and L. Zhu, "A missing power data filling method based on improved random forest algorithm," in Chinese Journal of Electrical Engineering, vol. 5, no. 4, pp. 33-39, Dec. 2019, doi: 10.23919/CJEE.2019.000025.
- [2] A. Pereira Barata, F. W. Takes, H. J. van den Herik and C. J. Veenman, "Imputation Methods Outperform Missing-Indicator for Data Missing Completely at Random," 2019 International Conference on Data Mining Workshops (ICDMW), Beijing, China, 2019, pp. 407-414, doi: 10.1109/ICDMW.2019.00066.
- [3] Bhatti, U. and Hanif. M. 2010. Validity of Capital Assets Pricing Model.Evidence from KSE-Pakistan.European Journal of Economics, Finance and Administrative Science, 3 (20).
- [4] Q Liu and M. Hauswirth, "A Provenance Meta Learning Framework for Missing Data Handling Methods Selection," 2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), New York, NY, USA, 2020, pp. 0349-0358, doi: 10.1109/UEMCON51285.2020.9298089.
- [5] V. K. Tripathi, P. Parashar and M. Rao K, "Missing Data: Its Emergence in the Real-world-A Practical Review on Google Play Apps dataset using Python," 2022 IEEE IAS Global Conference on Emerging Technologies (GlobConET), Arad, Romania, 2022, pp. 209-214, doi: 10.1109/GlobConET53749.2022.9872426.

