PCET's

PIMPRI CHINCHWAD UNIVERSITY

Department of CSE - Artificial Intelligence & Data Science

Mini Project Report

Base Algorithm in R - Manual Vs Library

Project Title: Exploratory Data Analysis of Skincare Products: Understanding Product Preferences Based on Skin Type, Skin Concern, Climate, and Ingredients

Academic Year: 2025–26

Submitted By

Siddhi R. Ibatti

Roll No: A-64

PRN : SOE23201030058

Department of AI & DS

Pimpri Chinchwad University

Guided By

Prof. Tushar R. Mahore

Contents

Abstract

This project explores patterns and trends in skincare product data to understand how factors such as climate, skin type, skin concerns, and ingredients influence product ratings and preferences. Using Exploratory Data Analysis (EDA) in the R programming language, a dataset of 200 skincare products is analyzed using visualizations, statistical summaries, and pattern identification.

Problem Statement

Selecting skincare products can be difficult because performance varies with climate, skin type, and product formulation. Consumers often face dissatisfaction due to mismatch between product and requirement.

This project answers:

"How do skincare features such as ingredients, climate suitability, and skin concerns influence product ratings?"

Objectives

- Perform exploratory data analysis on skincare product features.
- Investigate how climate and skin type affect product ratings.
- Analyze the effect of ingredients on performance.
- Study relationships such as price vs rating.
- Present insights using meaningful visualizations.

Dataset Description

The dataset contains 200 rows and the following columns:

- Product_ID
- Product_Name
- Brand
- Product_Type
- Price ()
- Rating
- Skin_Type
- Skin_Concern
- Climate
- City
- Avg_Temperature
- Humidity
- Packaging_Type
- Gender_Target
- Main_Ingredients

Methodology

4.1 Data Loading

The dataset was imported into R using `read.csv()`.

4.2 Data Cleaning

- Cleaned column names using `janitor`.
- Converted numeric and categorical variables correctly.
- Checked and handled missing values.

4.3 Exploratory Data Analysis

Performed:

- Univariate Analysis
- Bivariate Analysis
- Multivariate Analysis

Visualizations created:

- Histograms, Scatter plots
- Bar charts
- Boxplots

Exploratory Data Analysis

5.1 Univariate Findings

- Most product ratings range from 4.0–4.8.
- Skincare prices generally fall between 300–1200.
- Products for oily and combination skin types are most common.

5.2 Bivariate Findings

- Products in humid climates show slightly higher average ratings.
- Dry skin products tend to have higher pricing.
- Weak relationship between price and rating.

5.3 Multivariate Findings

- Hyaluronic Acid and Niacinamide appear frequently in high-rated products.
- Dry climate + Dry skin products have the highest mean rating.
- Ingredient profiles strongly influence ratings.

Results and Insights

- Hyaluronic Acid + Ceramides work best for dry climate users.
- Niacinamide-based products perform best for oily skin.
- Climate significantly influences product performance.
- Ingredients matter more than price or brand in determining satisfaction.

Conclusion

This exploratory data analysis concludes that skincare product performance depends largely on ingredient composition and climate suitability. The study shows that climate-based product formulation plays a more significant role than price or brand identity in user satisfaction.

Future improvements include:

- Adding machine learning prediction.
- Creating a skincare recommendation system.
- Expanding dataset with real consumer reviews.