Research Proposal

Data 698

Seung Min Song

2/26/2025

**Introduction**

Annual beauty product imports to the U.S. continue to rise. According to IBISWorld, total sales in the Cosmetics & Fragrance Stores industry reached $55.4 billion, growing at a 2.6% Compound Annual Growth Rate (CAGR) from 2019 to 2024, with a projected 2.4% CAGR from 2024 to 2029. Meanwhile, total sales in the Cosmetic & Beauty Products Manufacturing industry reached $44.2 billion, declining at a 4.1% CAGR from 2019 to 2024 but expected to recover at a 2.6% CAGR from 2024 to 2029.

According to data from the U.S. International Trade Commission (USITC), the trade balance for HS Code 3304.99 (cosmetics) has steadily declined from a surplus in 2019 to a deficit in 2024. This indicates a growing reliance on imported cosmetics in the U.S. market.

The U.S. primarily imports basic skincare and color cosmetics, which remain popular categories. While all-in-one products combining multiple functions were once dominant, there is now a growing interest in functional skincare, particularly anti-aging solutions. In response to this shift, manufacturers are adjusting their marketing strategies to emphasize health and anti-aging, aligning with evolving consumer trends.

Currently, there is intense competition between legacy players and emerging brands fighting for market dominance. As a result, the barrier to entry in the beauty market is exceptionally high. Asian beauty products are recognized worldwide for their excellent quality and reasonable prices, and they are establishing a unique position among beauty enthusiasts based on ingredients that are eco-friendly and prioritize skin health. In particular, Korea is widely known as an innovative cosmetics and beauty product producer and is attracting the attention of American consumers by providing sustainable and eco-friendly products. This is creating a market position that is differentiated from the luxurious image of France, a traditional cosmetics powerhouse.

**Research Objectives**

The objective of this study is to predict and analyze U.S. skincare product imports under HS Code 3304.99 using machine learning techniques. It examines key factors in the beauty industry, including product pricing, ingredients, brand awareness, exchange rates, and tariffs, to identify the most profitable products to import, the optimal timing, and the best foreign business partners. This study focuses on HS Code 3304.99, which covers skincare products such as cleansers, creams, mask packs, moisturizers, serums, and treatments. It analyzes the impact of brand-specific ingredients and pricing on consumer preferences while investigating major skincare product import trends in the U.S. utilizing exchange rates, oil prices, and tariffs data. Additionally, by integrating Google Trends data with U.S. trade data, the study explores the correlation between consumer search behavior and actual cosmetics import volumes. This approach provides valuable insights into market demand forecasting and optimal import strategies for skincare products.

**Research Questions**

1. Do specific product ingredients significantly impact consumer preferences?
2. How do price, ingredient composition, and brand awareness influence consumer ratings?
3. Do cosmetics with high consumer ratings and strong brand awareness experience continued growth in imports, particularly from specific exporting countries?
4. How does consumer search behavior (Google Trends data) correlate with actual skincare imports?
5. What is the optimal timing for importing skincare products based on market demand and macroeconomic trends?
6. How do macroeconomic factors, such as exchange rates, oil prices, and tariffs, affect skincare imports?

**Literature study**

Traditional beauty market analyses have primarily focused on market size, product category growth, purchasing trends, and changes in distribution channels, relying on major companies' operating profits and U.S. import/export statistics. For example, McKinsey & Company (2024) examined these factors in The Beauty Boom and Beyond: Can the Industry Maintain Its Growth? However, the growth of the beauty market cannot be assessed solely based on McKinsey & Company's findings. External factors such as exchange rates, oil prices, and tariffs also play a significant role in market expansion. Additionally, consumer purchasing trends have often been analyzed through a simple trade-up vs. trade-down framework, without conducting sentiment analysis based on product-specific characteristics.

Additionally, skincare products are often classified under a single HS Code, limiting the ability to analyze detailed item-specific data, such as cleansers, creams, mask packs, moisturizers, serums, and treatments, which are among the most popular and best-selling products in the market. Moreover, existing research typically examines the most searched skincare ingredients, showing only which ingredients are globally popular at a given time. For example, Foye, L. (2024) analyzed search trends in These Are the Top 10 Skin-Care Ingredient Searches of 2024. NewBeauty. However, insights are often lacking into whether the actual import volume of products containing these ingredients increased or decreased after the surge in searches. This conventional approach struggles to uncover correlations between consumer preferences, import trends, and product characteristics. This research integrates real-time trade data, consumer search behaviors, and machine learning predictions to provide actionable insights for importers and wholesalers.

**Data Sources**

This research utilizes Kaggle’s Sephora Skincare Dataset (Nady Inky, 2023) to analyze brand-specific ingredient trends and consumer sentiment. Sephora’s skincare products reflect various ingredient combinations, consumer preferences, prices, and brand identities. By leveraging product reviews, pricing, and ingredient information, the research aims to examine the relationship between consumer sentiment and specific ingredient combinations and identify key ingredient trends by brand.

Additionally, Google Trends data is used to analyze search trends for key ingredients and brands, providing insights into how consumer interest evolves. To further assess market trends, the research incorporates U.S. Census Bureau (2024) trade data and U.S. International Trade Commission (USITC), specifically analyzing HS Code 3304.99-1000 (Other Beauty and Cosmetic Products) to evaluate the extent to which skincare market trends influence import patterns.

Furthermore, FRED (Federal Reserve Economic Data) exchange rate data is utilized to examine how exchange rate fluctuations in major supplier countries impact the scale of U.S. skincare imports, offering a deeper understanding of their effect on the cosmetics import market. In addition, oil price data from Investing.com is used to analyze how fluctuations in fuel costs affect logistics expenses and, in turn, impact the import volume of skincare products.

**Methodology**

1. **Data Collection:**
2. Kaggle Sephora Skincare Dataset (Nady Inky, 2023). This dataset includes brand-specific ingredient lists, consumer reviews, and pricing data. It will be used for correlation analysis and frequency analysis. Only data where primary\_category is 'skincare' and ‘face’ will be utilized.
3. Google Trends (2020–2025). This dataset captures search trends for skincare ingredients and brands and will be used to analyze **consumer interest and emerging trends.**
4. US Census Bureau & USITC Trade Data (2024). This dataset provides **import volume statistics for HS Code 3304.99,** allowing for the identification of **trends in skincare imports.**
5. FRED (Federal Reserve Economic Data). This dataset contains historical **exchange rate data** for major U.S. trading partners and will be used to assess the **impact of currency fluctuations on skincare imports.**
6. Investing.com Oil Price Data. This dataset tracks global oil price changes and will be used to evaluate how rising oil costs influence skincare import trends.
7. **Data Refinement and Processing**:
8. Brand and country of origin matching: Add the country of origin next to the brand name in the Kaggle Sephora dataset (brand-level data). This will be compared with USITC import data by each country to analyze whether product ratings correlate with each country’s import trends.
9. The collected data is refined and processed in alignment with the research objectives. This includes data cleaning, transformation, handling missing values, and integration to create a unified dataset suitable for analysis.
10. Correlation analysis will be conducted on ‘loves\_count’, ‘rating’, ‘price\_usd’, and ‘reviews’ to assess their relationships.
11. Frequency analysis will be applied to ‘ingredients’ to identify the most commonly used components in skincare products. Ingredients will be analyzed to examine whether the presence of specific components influences product ratings and consumer preferences."
12. **Analysis Execution**
13. Frequency Analysis & Correlation Analysis: Identify key ingredients commonly found in highly-rated products and analyze their correlation with consumer preferences and pricing, providing deeper insights into market dynamics.
14. XGBoost (Decision Tree Model): Evaluate the impact of specific ingredient combinations, price, and brand awareness on consumers' purchasing decisions. Feature importance analysis will be used to determine which factors have the greatest influence.
    * 1. XGBoost is chosen for its ability to model nonlinear relationships and handle high-dimensional structured data, making it suitable for analyzing complex interactions between ingredient composition, loves\_count, and price, on consumer purchasing decisions."
15. ARIMA (Time Series Forecasting): Forecast future skincare import volumes based on economic factors. Use historical import volumes (USITC) and macroeconomic indicators as input features.
    * 1. USITC data is combined with exchange rates and oil price data. All data is converted into monthly rolling averages.
      2. **ARIMA was chosen for its ability to capture temporal patterns, trends, and seasonality in time series data. It is particularly effective in predicting skincare imports by analyzing historical trade data, exchange rates, and oil prices. Unlike more complex deep learning models, ARIMA requires relatively little computational power while providing interpretable results, making it a practical choice for trade forecasting.**
16. **Model Evaluation**
17. **XGBoost’s performance** will be evaluated using Mean Squared Error (MSE) and R-squared (R²). Additionally, feature importance scores will be analyzed to understand the relative influence of each variable on purchasing decisions.
18. **ARIMA's performance will** be evaluated using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).
19. **Integration of Results:**
20. The results from frequency analysis, correlation analysis, XGBoost, and ARIMA will be synthesized to derive final insights. Findings from the frequency analysis, such as the relationship between ingredient composition and consumer reviews, will be integrated with predictive insights from XGBoost and ARIMA.
21. **Finally, graphical analysis will be used to compare product import trends during the 2018 trade war with periods unaffected by trade tensions to identify significant differences.** 
    * 1. Comparison of import volumes from specific countries (e.g. China vs. Korea)

**Conclusion**

This study provides a data-driven approach to understanding the factors influencing U.S. skincare imports under HS Code 3304.99. It integrates consumer responses, ingredient correlation and frequency analysis, machine learning, and time series forecasting to offer insights into the most profitable products, optimal timing, and key import trading partners. Correlation and frequency analysis examine key ingredient trends and their impact on consumer preferences, while XGBoost evaluates the relative importance of factors such as ingredient mix and price in shaping purchasing decisions. ARIMA captures temporal patterns, trends, and seasonality in historical trade data, exchange rates, and oil prices to forecast future import trends. Additionally, a comparative graphical analysis of skincare product import trends during the 2018 trade war and periods unaffected by trade tensions is conducted to determine whether tariffs are negatively impacting imports, as many fear. This study provides valuable insights into market demand forecasting and import strategies, equipping industry stakeholders with a tool to make informed decisions in the U.S. beauty market.

Reference:

1. Zambrano, A. M. (2024). Beauty, cosmetics & fragrance stores in the US. IBISWorld. Retrieved from <https://www.ibisworld.com>
2. Buchko, M. (2025). Cosmetic & beauty products manufacturing in the US. IBISWorld. Retrieved from <https://www.ibisworld.com>
3. Foye, L. (2024). These are the top 10 skin-care ingredient searches of 2024. NewBeauty. Retrieved from <https://www.newbeauty.com/top-10-searched-for-skincare-ingredients-2024-according-to-google/>
4. McKinsey & Company. (2024). The Beauty Boom and Beyond: Can the Industry Maintain Its Growth? Retrieved from <https://www.mckinsey.com/industries/consumer-packaged-goods/our-insights/the-beauty-boom-and-beyond-can-the-industry-maintain-its-growth>
5. Boer, L., & Rieth, M. (2024). The Macroeconomic Consequences of Import Tariffs and Trade Policy Uncertainty. Retrieved from https://www.elibrary.imf.org/view/journals/001/2024/013/article-A001-en.xml