Vietnam National University Ho Chi Minh City

University of Science

Faculty of Information Technology

PROJECT PROPOSAL

Brand Fashion Market Seasonal Trend Analysis

Course	Introduction to Data Science
Class	22CLC
Students	**22127225 – Trần Thị Thiên Kim** **22127357 – Phạm Trần Yến Quyên** **22127374 – Lê Thanh Tâm** **22127449 – Mai Đức Vân**
Github:	Brand Fashion Market Seasonal Trend Analysis

HCMC, 2024

To have a good experience, please use a PDF reader that supports hyperlinks

Mục lục

1 About this Project:

This project aims to answer **insightful** questions regarding the Branded Fashion Market with the goal of conducting informed and predictive analysis of the recent **trend(s)** in the specify Market. Key questions include identifying the most popular brands and styles, determining the seasonal impact on sales rates, and evaluating strategies to enhance or reduce sales for specific product lines. This analysis will leverage recent data to provide actionable insights for stakeholders in the fashion industry.

2 Data Crawl Sources:

2.1 Requirements:

- Data:
  - + The data must contain features relevant to the laid-out questions.
  - + The data must be of the numerical or can be converted into numerical typing (ie. Label encoding, datetime, etc.).
  - + The data must be correct, valid.
- Sources:
  - + The sources must be of a known brand (Officially copyrighted).
  - + The sources must be crawl-able.
  - + The sources must contain the required data.

2.2 Valid sources:

NOTICE: *The sources are laid-out in order of most valid (with the first one being the most valid and the last one being the least)*

1. Fashion E-commerce Websites (e.g., ASOS, Zara, H&M) - for brand and style popularity data.
2. Social Media Platforms (e.g., Instagram, Pinterest) - for trend analysis on clothing styles and user engagement.
3. Fashion Trend Aggregators (e.g., Google Trends, Trendalytics) - for broader insights on trending keywords.
4. Online Retailer APIs (if accessible) - for direct access to seasonal sales data and other retail metrics.

3 Planning:

Following the provided Guidelines, the project plans is laid-out in steps:

1. **Data Collection:**
   (a) Crawl data from e-commerce and social media sites using web scraping tools (e.g., Selenium, BeautifulSoup) either through `Python` dedicated library or extentions.
   (b) Structure the dataset with at least five relevant attributes (e.g., brand, style, sales rate, seasonal timing, and region) and collect over 1,000 records.
2. **Data Exploration and Preprocessing:**
   (a) Conduct initial exploration using descriptive statistics to identify missing values, inconsistent data types, and outliers.
   (b) Standardize data types and handle any anomalies to prepare for analysis.
3. **Question Formulation:** Form of meaningful questions (ideally relating to the collected data)
   (a) Which brands and styles are the most popular among consumers?
   (b) Are there specific styles that outperform others during specific times of the year?
   (c) What marketing strategies are correlated with higher sales for particular brands?
   (d) How can certain products' sales be strategically boosted?
4. **Data Modeling and Evaluation:**
   (a) Use predictive modeling techniques (e.g., regression for sales predictions, classification to identify trending styles).
   (b) Validate models with cross-validation, measuring performance through metrics such as Mean Squared Error (MSE) for regression and accuracy or F1-score for classification.
   (c) Compare several models (e.g., linear regression, decision trees, or SVM) and analyze which provides the most accurate results.

   NOTICE: *For each question and model, define the method of answering through data analysis and visualization.*
5. **Reflection:** Look back at the results calculated as well as question answered to determined strong/weak points of the project.
   (a) Document challenges, lessons learned, and potential improvements.
   (b) Identify additional data that could enhance model performance or deepen insights if more time and resources were available.

3.1 Expected Result:

The analysis is expected to yield **insights** into popular brands and styles within the branded fashion market, helping to uncover correlations between seasonal changes and sales. The results will guide stakeholders in designing seasonally aligned marketing campaigns and inform strategies to boost sales for specific products. Additionally, this project will provide a basis for trend prediction in future market behavior.