Project Proposal

Part 1: Organization Information

Overview:

H&M Group is a global powerhouse in fashion and design, renowned for offering a diverse range of stylish, high-quality, and affordable products. Since its founding in 1947, H&M has remained dedicated to making fashion accessible to all while driving positive change. With a bold vision to become a fully circular and climate-positive business by 2040, H&M Group is transforming the future of retail through innovation, sustainability, and customer-centric strategies. Operating across 53 (spanning countries and regions) online markets and approximately 4,850 stores worldwide, the company is committed to delivering exceptional shopping experiences through cutting-edge technology and personalized services.

Employees:

As a **global leader**, H&M Group employs approximately **150,000 talented individuals** across corporate offices, logistics hubs, and retail stores.

Industry:

H&M Group operates in the **Fashion and Retail** industry, setting trends and redefining the shopping experience.

Type:

H&M Group is a **Public Company**, listed on the Nasdaq Stockholm Stock Exchange under the ticker symbol **HM-B**.

Specialties:

H&M Group excels in several key areas:

- **Fashion and Lifestyle Innovation:** Delivering an extensive array of clothing, accessories, and home products through iconic brands such as H&M, COS, Weekday, & Other Stories, and ARKET.
- **Sustainability Leadership:** Pioneering circular fashion through sustainable sourcing, eco-friendly production, and ambitious climate goals.

- **Omnichannel Excellence:** Seamlessly integrating online and in-store experiences to offer personalized shopping journeys tailored to individual customer needs.
- **Data-Driven Insights:** Leveraging advanced analytics and artificial intelligence to enhance product recommendations, optimize operations, and deliver exceptional customer satisfaction.

Part 2: Use Case/Scenario to understand Value

In the fast-paced world of fashion retail, data-driven insights are crucial for making informed decisions. For H&M, understanding sales patterns, customer preferences, and product performance can significantly enhance inventory management, marketing strategies, and customer satisfaction. This analysis focuses on key questions related to sales trends, product category performance, and distribution across departments. By leveraging these insights, H&M can optimize operations and better align with consumer demand.

1. **Which months exhibit the highest sales, and what trends can we observe in peak sales periods?**

   Identifying the periods with the highest sales and analyzing peak sales trends

- **Value:** Helps H&M identify peak sales months and plan inventory, marketing campaigns, and promotions accordingly.
- **Outcome:** This insight allows H&M to prepare for peak demand by increasing stock levels, launching targeted promotions, and optimizing marketing efforts during this period.

2. **How do different product categories perform in terms of revenue across different months?**

   Sales across different product categories

- **Value:** Helps H&M understand which product categories drive revenue during specific months and adjust inventory and marketing strategies.
- **Outcome**. This insight allows H&M to focus on promoting high-performing categories and reevaluate strategies for underperforming ones.

3. **What are the top-selling products?**

Identifying the top-selling products

- **Value:** Helps H&M identify customer preferences, optimize inventory management, and focus marketing efforts on high-demand products.
- **Outcome:** H&M can identify the top-selling products by revenue and units sold. For example, black-colored tops or organic cotton trousers might be the best-selling items. This insight allows H&M to ensure these products are always in stock, focus marketing campaigns on them, and improve product recommendations by suggesting popular items to customers.

4. **Which color and graphical appearance are most frequently used?**

Determining the most frequently used colors and graphical appearances

- **Value:** Helps H&M align product offerings with customer preferences, identify design trends, and optimize inventory by focusing on popular colors and patterns.
- **Outcome:** This insight allows H&M to focus on producing more products in these colors and patterns, ensuring they meet customer demand, and use this information for targeted marketing and improved product recommendations.

5. **How are products distributed among departments and garment groups?**

Examining the distribution of products among departments and garment groups

- **Value:** Helps H&M identify which departments and garment groups dominate the product catalog, discover opportunities for expansion, and optimize inventory management.
- **Outcome:** This insight allows H&M to focus on promoting high-performing departments and garment groups, explore opportunities for expansion in underperforming categories, and optimize inventory and marketing strategies.

Part 3: Problem Statement and Background

High-Level Problem Statement:

As a leading global fashion retailer, H&M must effectively manage its inventory, marketing strategies, and product offerings to meet evolving customer demands. With an extensive product range and diverse customer base, H&M needs to uncover key insights into sales trends, product performance, and customer preferences to maintain its competitive edge. Without a clear understanding of these factors, the company risks inefficiencies such as overstocking, understocking, missed sales opportunities, and ineffective marketing campaigns.

Specific Questions to Address:

- **Sales Trends and Peak Periods:**

    1. Which months have the highest sales, and what patterns emerge during these peak periods?
    2. How can H&M use this data to optimize inventory planning, marketing campaigns, and promotional activities?

- **Product Category Performance:**

    1. How does revenue vary across different product categories throughout the year?
    2. Which categories perform best, and which ones struggle during peak sales months?

- **Top-Selling Products:**

1. What are the best-selling products in terms of revenue and units sold?
2. How can H&M ensure these popular items remain in stock and are effectively promoted?

- **Customer Preferences in Design:**

1. Which colors and patterns are most commonly used in products?
2. How can H&M align its product designs with customer preferences to improve inventory management and stay on-trend?

- **Product Distribution Across Departments and Garment Groups:**

1. How are products distributed across different departments and garment groups?
2. Which departments and garment groups are most prominent, and where are the opportunities for growth?

Background:

H&M operates in a fast-paced and highly competitive retail industry, where customer preferences and fashion trends shift quickly. To stay ahead, H&M must continuously analyze sales data, customer behavior, and product performance. By addressing the specific questions outlined above, H&M can gain critical insights into sales patterns, customer preferences, and product performance. These insights will help the company optimize inventory management, refine marketing strategies, and align product offerings with customer demand, ultimately driving revenue growth and improving customer satisfaction.

The analysis will focus on key datasets such as transactions_train.csv and articles.csv, which contain transaction details and product information, respectively. Using Excel tools like Pivot Tables and data merging techniques, we can extract actionable insights to address the identified challenges and support H&M's strategic decision-making processes.

Part 4: Data Sources you intend to Use

Explanation of articles.csv

file has detailed information about every product available for purchase. Here's a simple breakdown of the data and what we can learn from it:

1. Product Details:
    - It includes things like product names, types (e.g., tops, trousers), and groups (e.g., clothing, accessories).
2. Colors and Patterns:
    - It tells us the colors (e.g., black, white) and patterns (e.g., solid, stripes) of the products.
3. Categories:
    - Products are grouped into departments (e.g., Jersey Basic, Lingerie) and garment types (e.g., tops, underwear).
4. Descriptions:
    - Each product has a description that explains its features, like materials (e.g., organic cotton) or special details (e.g., elastic waist).

By analyzing this data, we can:

- Find out which products are most popular.
- See which colors and patterns customers prefer.
- Understand how products are organized in the catalog.
- Use this information to improve recommendations, plan inventory, and make better marketing decisions.

1. Dataset Overview

- Columns in the Dataset:
    - article_id: Unique identifier for each article.
    - product_code: Code associated with the product.

- prod_name: Name of the product.
- product_type_no: Numeric code for the product type.
- product_type_name: Name of the product type (e.g., Vest top, Bra, Sweater).
- product_group_name: Group to which the product belongs (e.g., Garment Upper body, Underwear).
- graphical_appearance_no: Numeric code for the graphical appearance.
- graphical_appearance_name: Description of the graphical appearance (e.g., Solid, Stripe).
- colour_group_code: Numeric code for the color group.
- colour_group_name: Name of the color group (e.g., Black, White, Light Blue).
- perceived_colour_value_id: Numeric code for the perceived color value.
- perceived_colour_value_name: Description of the perceived color value (e.g., Dark, Light, Dusty Light).
- perceived_colour_master_id: Numeric code for the perceived color master.
- perceived_colour_master_name: Description of the perceived color master (e.g., Black, White, Blue).
- department_no: Numeric code for the department.
- department_name: Name of the department (e.g., Jersey Basic, Clean Lingerie).
- index_code: Code for the index (e.g., A, B, C).
- index_name: Name of the index (e.g., Ladieswear, Lingeries/Tights).
- index_group_no: Numeric code for the index group.
- index_group_name: Name of the index group (e.g., Ladieswear, Menswear).
- section_no: Numeric code for the section.
- section_name: Name of the section (e.g., Womens Everyday Basics, Womens Lingerie).
- garment_group_no: Numeric code for the garment group.
- garment_group_name: Name of the garment group (e.g., Jersey Basic, Underwear).
- detail_desc: Detailed description of the product

2. Key Insights from the Dataset

2.1 Product Categories and Types

- The dataset contains a wide variety of product types, such as:
    - Vest tops (e.g., Strap top)
    - Bras (e.g., OP T-shirt (Idro))
    - Sweaters (e.g., SWEATSHIRT OC)
    - Tights (e.g., 20 den 1p Stockings)
    - Trousers (e.g., Jerry jogger bottoms)
    - Accessories (e.g., Hair clips, Belts)
- Analysis:
    - Identify the most common product types and groups (e.g., Garment Upper body, Underwear).
    - Determine which product types are most frequently listed, indicating their popularity or availability.

2.2 Color Analysis

- The dataset includes detailed color information, such as:
    - Color groups: Black, White, Light Blue, Dark Red, etc.
    - Perceived color values: Dark, Light, Dusty Light.
    - Perceived color masters: Black, White, Blue, etc.
- Analysis:
    - Identify the most popular colors (e.g., Black, White, Light Blue).
    - Analyze the distribution of perceived color values (e.g., are dark colors more common than light colors?).

2.3 Graphical Appearance

- The dataset provides information on the graphical appearance of products, such as:
    - Solid: Single-color products.
    - Stripe: Striped patterns.
    - All over pattern: Products with patterns covering the entire surface.
- Analysis:
    - Determine the most common graphical appearances (e.g., Solid vs. Stripe).

- Identify trends in patterns and designs.

2.4 Departments and Garment Groups

- The dataset categorizes products into departments and garment groups, such as:
  - Departments: Jersey Basic, Clean Lingerie, Tights basic.
  - Garment groups: Jersey Basic, Underwear, Socks and Tights.
- Analysis:
  - Identify the most common departments and garment groups.
  - Analyze the distribution of products across these categories.

2.5 Product Descriptions

- The detail_desc column provides detailed descriptions of products, such as:
  - Material composition (e.g., organic cotton, microfibre).
  - Features (e.g., elasticated waist, padded cups).
  - Usage (e.g., for indoor use, for sports).
- Analysis:
  - Extract keywords from product descriptions to identify common features or materials.
  - Analyze the frequency of specific terms (e.g., "organic cotton," "elasticated waist").

Explanation of transactions_train.csv

The **transactions_train.csv** file is one of the key datasets for this analysis. It contains historical purchase records, providing insights into customer behavior, product performance, and overall sales trends.

1. Content of transactions_train.csv

Column Name	Description
t_dat	Transaction date

This dataset includes the following important fields:

customer_id	Unique identifier for each customer
article_id	Unique identifier for each purchased product
price	Price at which the product was purchased
sales_channel_id	Sales channel (e.g., online store vs. physical store)

2. Purpose in Analysis

This **top-** dataset is essential for understanding **customer purchasing behavior**, **identifying selling products**, and **optimizing recommendations**. Below are the specific ways this data is useful:

A. Understanding Buying Patterns

- Track purchase frequency for individual customers to classify them into **loyal customers** and **occasional shoppers**.
- Identify **trending products** by analyzing sales volume over time.
- Determine **seasonal variations** in product demand (e.g., higher sales of winter clothing during colder months).

B. Product Performance Analysis

- Identify **best-selling items** across different time periods.
- Analyze **price trends** to see if discounts or promotions impact sales.
- Examine which products are frequently **bought together** to improve cross-selling strategies.

C. Customer Segmentation & Personalization

- Group customers based on purchasing habits to deliver **targeted marketing campaigns**.
- Identify **high-value customers** who make frequent or high-value purchases.

- Develop **personalized product recommendations** by understanding previous purchase history.

D. Online vs. In-Store Shopping Trends

- By analyzing the **sales_channel_id**, determine differences in purchasing behavior between **online and offline shoppers**.
- Understand which product categories perform **better online vs. in stores**.
- Optimize **inventory management** based on regional sales patterns.

Explanation of customer.csv

The customers.csv file comprises key customer-related information that provides insights into the demographic and behavioral characteristics of H&M's clientele. This dataset typically includes:

Customer ID: A unique identifier for each customer.

Demographic Details: Information such as age, gender, geographic location, and other relevant attributes.

Loyalty and Behavioral Data: Indicators of customer loyalty, frequency of purchases, or membership in rewards programs, if available.

Importance of the Data:

Customer Profiling and Segmentation:

The dataset allows for the creation of detailed customer profiles. By analyzing demographic and behavioral data, H&M can segment its customer base into distinct groups, such as frequent buyers, occasional shoppers, or high-value customers.

Targeted Marketing and Personalization:

With insights from customer demographics and purchase history, tailored marketing strategies can be developed. This enables the company to deliver personalized recommendations, targeted promotions, and communication that resonates with each customer segment.

Enhancing Customer Experience:

A deeper understanding of customer preferences facilitates improved product recommendations and service enhancements. This leads to a more engaging shopping experience and ultimately drives customer satisfaction and retention.

Strategic Decision-Making:

The data supports strategic initiatives by providing a clearer picture of customer behavior and trends. This information can be used to optimize product assortments, adjust inventory levels, and improve overall operational efficiency.

Integrating customers.csv, transactions_train.csv, and articles.csv for Analysis

Data Set	Key Columns	Purpose
Transaction_train.csv	Customer_id, Article_id	who bought what, when, and at what price
Customer.csv	Customer_id	Customer demographic data (age, location, loyalty program info, etc.)
Article.csv	Article_id	Product details (category, brand, color, material, etc.)

To derive meaningful insights, you need to merge the three datasets based on common identifiers. Here's a structured approach to integrating these files for analysis using Excel

Conclusion: Integrating the Datasets

By integrating the customers.csv, transactions_train.csv, and articles.csv files, you create a comprehensive dataset that offers a complete picture of H&M's business. This consolidated view enables you to:

- Understand Customer Behavior:
  Merge customer demographics with transaction records to analyze purchasing habits, segment customers, and tailor personalized recommendations.
- Evaluate Product Performance:
  Combine product details with sales data to identify top-selling items, monitor trends, and optimize inventory management.
- Enhance Decision-Making:
  Leverage the integrated data to uncover actionable insights, such as seasonal sales patterns and channel-specific performance, which inform targeted marketing strategies and overall business planning.

Integrating these datasets empowers H&M to make data-driven decisions that improve customer experience, drive sales, and ultimately enhance operational efficiency.

Part 5: Preliminary Analysis

a. Assessing the Quality of Data

Before conducting deeper analysis, it's critical to ensure the data is clean, consistent, and complete. Here's how you can assess data quality in Excel:

1.Checking the missing data.

Use functions such as COUNTBLANK to identify missing entries.

=COUNTBLANK(B2:B1000)

This formula will count the number of empty cells in the range, helping you determine if missing data is a potential issue.

2. Identify Duplicate Records:

Select the range (e.g., customer IDs in the customers dataset), then go to **Home > Conditional Formatting > Highlight Cells Rules > Duplicate Values** to highlight any duplicates.

3.Validate Data Consistency and Format:

Check that dates, numbers, and text entries are formatted correctly. Consistent formatting avoids errors in calculations and pivot tables.

In addition, validating data consistency is crucial. This involves ensuring that data is entered in the correct format and that there is uniformity across the dataset. For example, dates should be consistently formatted and numeric fields should contain only numbers. Using functions like ISNUMBER or the TEXT function helps standardize formats, reducing errors in subsequent calculations or pivot table analyses.

Finally, detecting outliers is another vital aspect of assessing data quality. Outliers may represent data entry errors or true variations in the data that require further investigation.

b. Providing Early Insights Using Charts/Graphs

Once data quality has been confirmed, we have generated visual insights. Below is our analysis using Excel charts and pivot tables

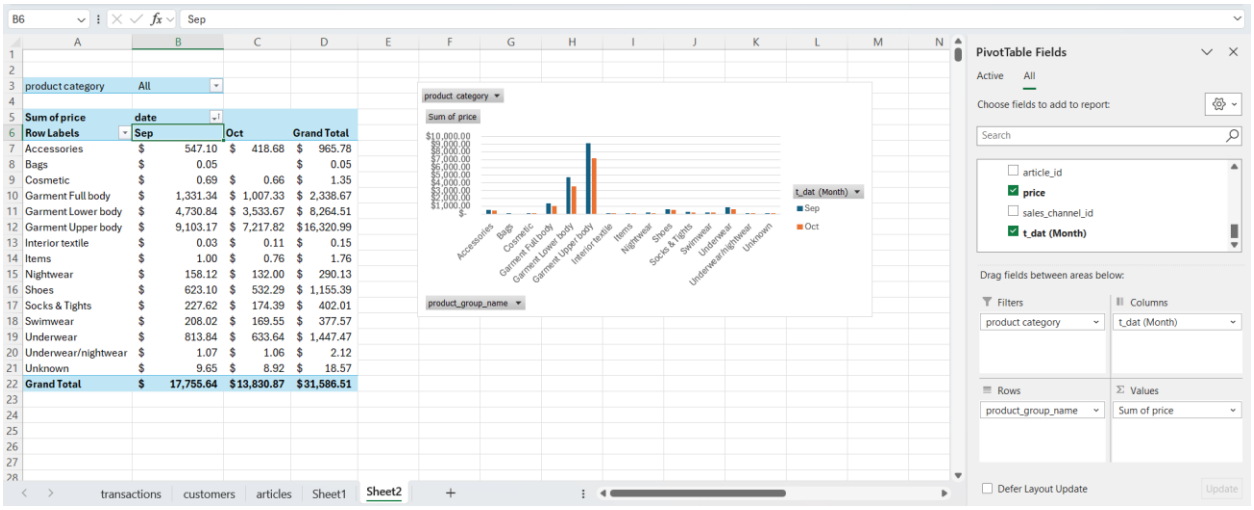
1.Which months exhibit the highest sales, and what trends can we observe in peak sales periods?

Our analysis indicates that September stands out as the peak sales month, with revenue reaching $17,755.64, substantially higher than the $13,830.87 observed in October. This suggests a strong seasonal trend where customer purchasing activity intensifies in September, potentially due to back-to-school promotions, seasonal product launches, or other marketing initiatives. Such insights are critical for optimizing inventory levels and tailoring marketing strategies to capitalize on this high-demand period.

2. How do different product categories perform in terms of revenue across different months?

Our analysis reveals that "Garment Upper body" consistently outperforms other categories, emerging as the highest-selling segment in both September and October. In September, this

category generated $9,103.17 in revenue, underscoring its robust and sustained demand. This consistent performance suggests that products in this category resonate well with customers across different periods, highlighting a strategic opportunity for focused marketing and inventory planning.



Key Findings:

- **Peak Seasonality:** September is identified as the peak sales month, generating significantly higher revenue ($17,755.64) compared to October ($13,830.87). This suggests a strong seasonal trend or promotional activities driving sales in September.
- **Top Performing Category:** "Garment Upper body" stands out as the highest-selling category in both September and October, with September revenue reaching $9,103.17. This indicates a consistent demand for this product type.
- **Low Performing Category:** "Underwear/nightwear" registers the lowest sales in September at $1.07. While it shows slightly better performance in October, it remains a low-performing category overall. This could be due to factors like lower demand, pricing strategy, or competition.

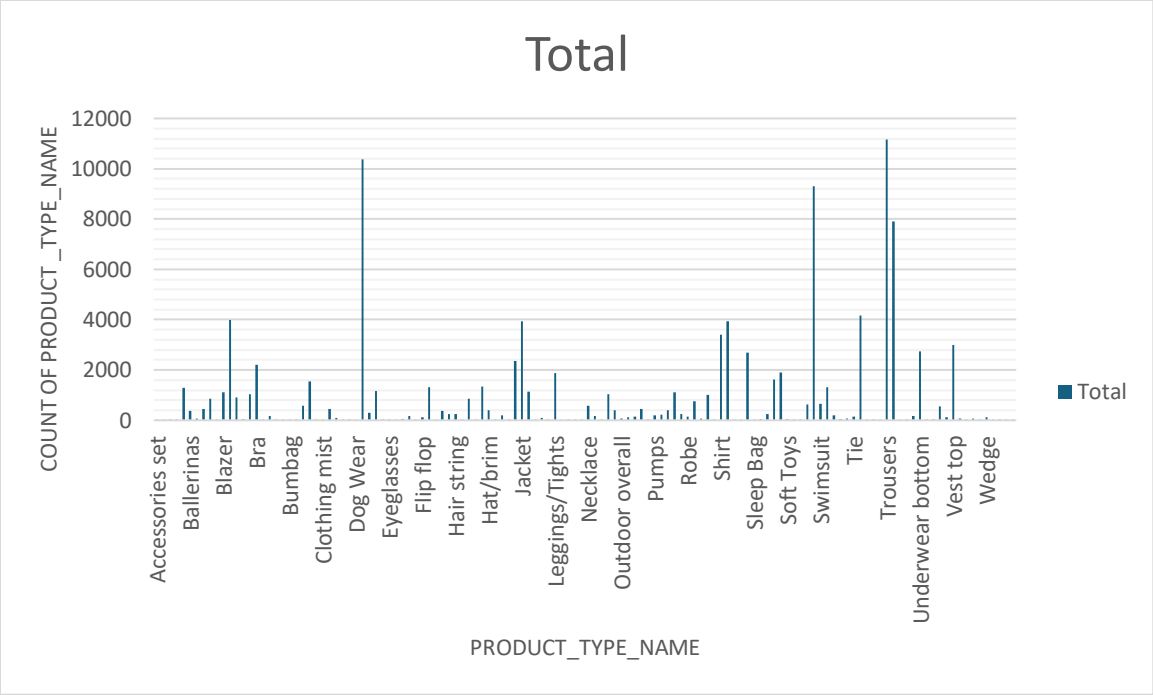
- **Other Notable Categories:** Several other categories contribute significantly to overall sales. "Garment Lower body" and "Shoes" demonstrate robust sales figures in both months. "Accessories" and "Cosmetic" also contribute substantially to revenue.
- **Data Limitations:** The analysis is limited to two months and doesn't provide insights into the entire year's performance. Understanding the broader sales trends would require a more comprehensive dataset. Additionally, the analysis doesn't explore potential factors influencing sales, such as pricing, marketing campaigns, or external market dynamics.

3.What are the most common product types and groups sold?

This analysis will help us identify which categories are the most popular or have the widest . For example, are tops, trousers, or accessories more prevalent in the catalog? Understanding this can guide inventory planning and marketing strategies.

To answer this question, we need to analyze the articles.csv dataset, specifically focusing on the column’s product_type_name and product_group_name. These columns tell us the type of product (e.g., tops, trousers, accessories) and the group it belongs to (e.g., Garment Upper body, Underwear, Accessories).

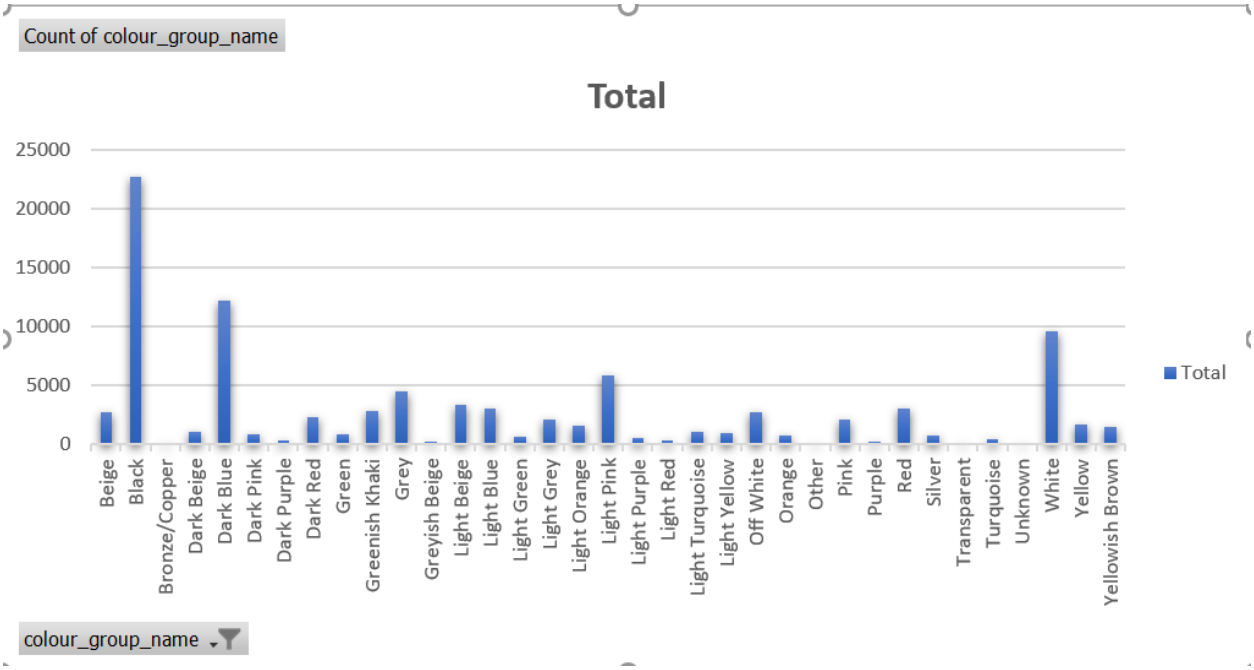
As per our analysis Trouser is the product which is widely sold followed by dresses.

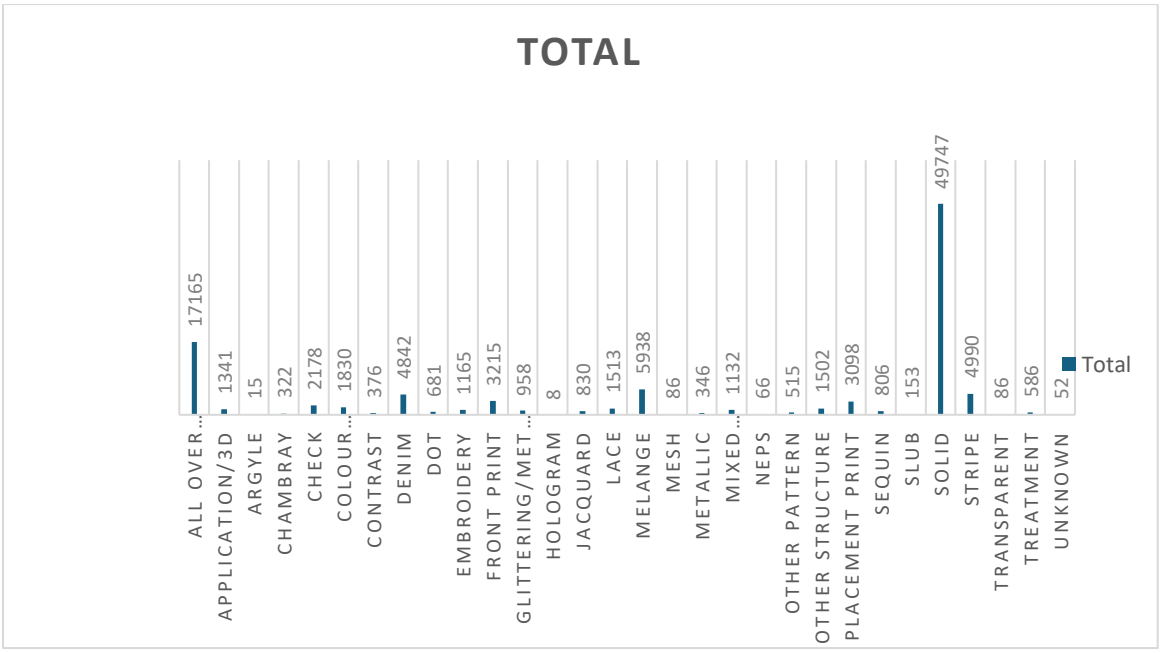


4. Which colors and graphical appearances are most frequently used?

By examining the most common colors (e.g., black, white, light blue) and patterns (e.g., solid, stripe), we can gain insights into customer preferences and current trends. This information can be used to tailor product offerings to match customer tastes.

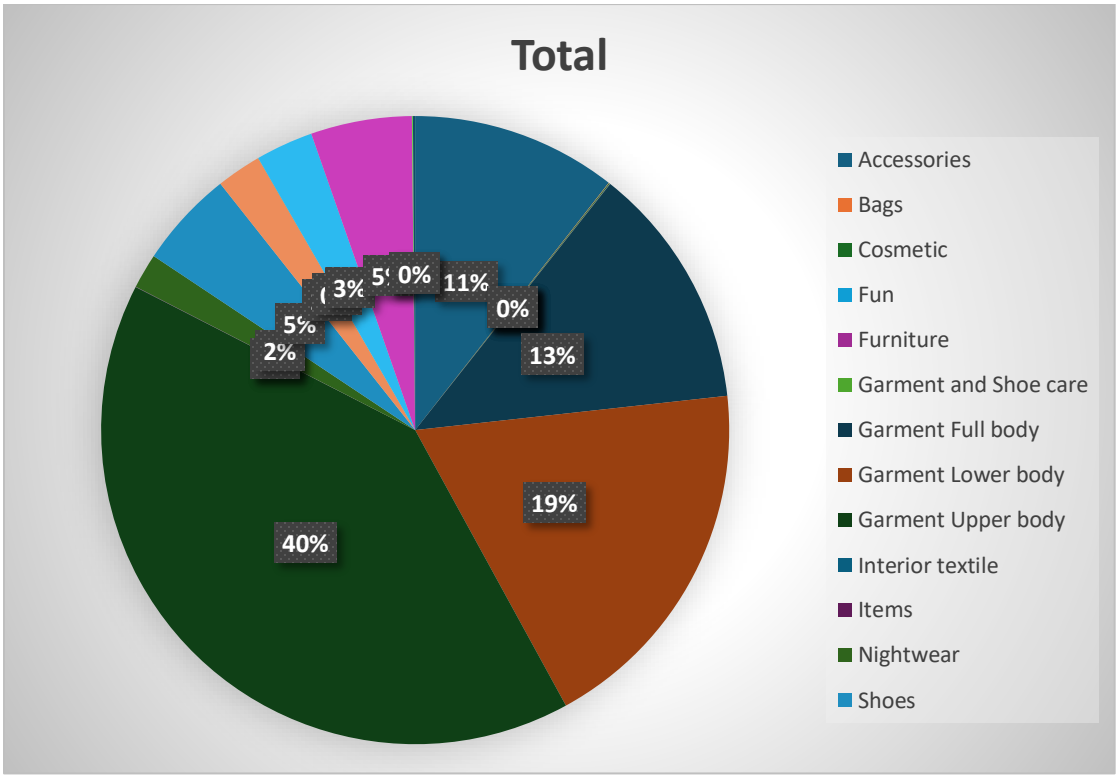
Based on an in-depth Excel pivot table analysis of customer preferences and current market trends, the most frequently used color identified is black. This finding highlights a strong customer inclination towards black as a preferred choice, which aligns with its timeless appeal and versatility. Additionally, the analysis revealed that solid patterns are the most commonly favored, further emphasizing a preference for simplicity and elegance. These insights can be leveraged to refine product offerings and align them with prevailing customer tastes, ensuring greater market relevance and customer satisfaction.





5. **How are products distributed across departments and garment groups?**

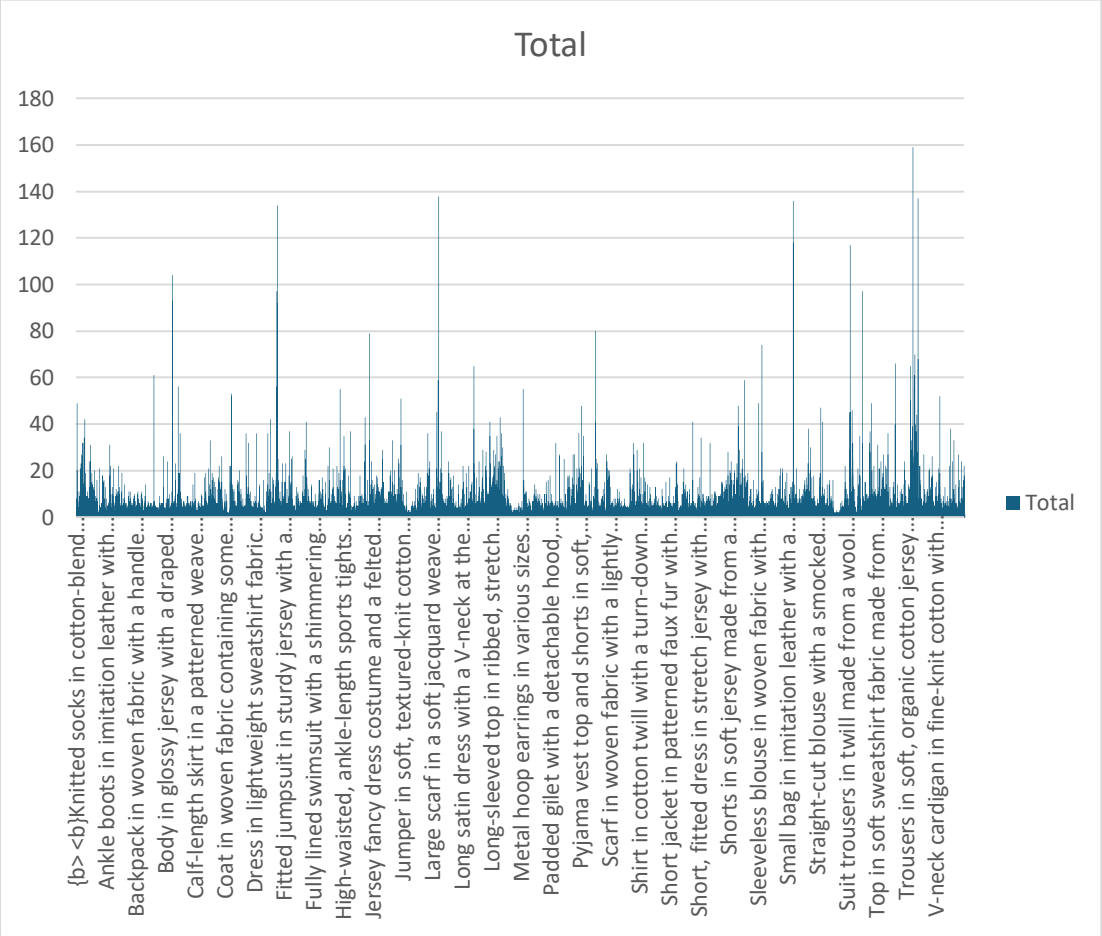
This question helps us understand the focus areas of the product catalog. For instance, are most products concentrated in the "Jersey Basic" department or the "Underwear" garment group? This analysis can reveal which departments or groups dominate the catalog and where there might be opportunities for expansion.



Based on the Excel analysis, the majority of products are concentrated in the **"Garment Upper Body"** category. This indicates that the product catalog is heavily focused on upper body clothing, such as shirts, tops, and jackets. This insight suggests that **"Garment Upper Body"** is a key area of strength and customer interest. However, it also highlights an opportunity to explore expanding into other garment categories, like lower body or full-body clothing, to create a more balanced and diverse product catalog.

6. **What are the most common materials and features mentioned in product descriptions?**

By analyzing the detail_desc column, we can identify frequently mentioned materials (e.g., organic cotton, microfiber) and features (e.g., elasticated waist, padded cups). This can highlight key selling points and customer preferences, helping to refine product recommendations.



Conclusion:

The preliminary analysis conducted using Excel charts and pivot tables has provided valuable insights into H&M's sales and product data over the two-month period. Key findings include:

Seasonal Sales Trends: September emerged as the peak sales month, generating significantly higher revenue ($17,755.64) compared to October ($13,830.87). This indicates a strong seasonal

or promotional impact, suggesting that targeted marketing efforts during this period could yield substantial benefits.

Category Performance: "Garment Upper body" has consistently been the top-performing category in both months, with September revenue reaching $9,103.17. This sustained demand highlights the category's critical role in driving overall sales and presents a strategic opportunity for focused marketing and inventory planning. Conversely, "Underwear/nightwear" remains a low-performing segment, warranting further investigation into potential causes such as market demand or pricing strategies.

Product Assortment: Analysis of product types and groups indicates that certain items, like trousers and dresses, are prevalent in the catalog, which can guide inventory planning and targeted promotional strategies. Additionally, the distribution across departments shows a concentration in the "Garment Upper Body" segment, suggesting both strength in this area and potential for diversification into other garment categories.

Design Preferences: The pivot table analysis on colors and graphical appearances reveals that black is the most frequently used color, and solid patterns dominate. These findings reflect a strong customer preference for timeless and versatile product designs, which can inform product development and merchandising strategies.

Materials and Features: An examination of the product descriptions has identified key selling points such as the frequent mention of materials like organic cotton and features like elasticated waists. These insights are valuable for refining product recommendations and enhancing the overall customer experience by aligning product offerings with consumer preferences.

The analysis of H&M's sales and product data highlights important trends in customer purchases. September had the highest sales, showing a strong seasonal pattern, while "Garment Upper Body" was the best-selling category. In contrast, "Underwear/nightwear" had the lowest sales, suggesting areas for potential improvement. The preference for black and solid patterns indicates key customer choices that can help guide marketing and inventory decisions.

These insights, gained through Excel charts and pivot tables, confirm that certain product categories drive most sales, while others may need adjustments. Although the analysis covers only

two months, it provides a useful starting point for improving product recommendations, managing stock better, and making smarter business decisions.

Part: 6 Summary & Recommendations

The analysis focuses on H&M Group, a leading global fashion retailer, to identify key sales trends, product performance, and customer preferences using datasets such as articles.csv, transactions_train.csv, and customers.csv.

Key Findings:

- Sales Trends & Peak Periods
    - September is the peak sales month, generating $17,755.64, significantly higher than October ($13,830.87).
    - Seasonal trends indicate higher purchasing activity, possibly due to back-to-school promotions or new seasonal launches.
- Product Category Performance
    - "Garment Upper Body" is the best-performing category in both September and October ($9,103.17 revenue in September).
    - "Underwear/Nightwear" is the worst-performing category, indicating a need for different marketing strategies.
- Best-Selling Products
    - Trousers and Dresses are the most sold products.
    - Ensuring popular items are well-stocked can help maximize revenue.
- Customer Preferences
    - Black is the most preferred color, followed by White and Light Blue.
    - Solid patterns are the most common, indicating a preference for simple and elegant designs.
- Product Distribution
    - The "Garment Upper Body" category dominates the product catalog.
    - Opportunity to expand into lower-body or full-body clothing to diversify the product range.
- Material & Features

    - Organic Cotton and Elasticated Waistbands are commonly mentioned in product descriptions.
    - Indicates customer preference for comfortable and sustainable materials.

Recommendations for H&M

- Optimize Inventory & Marketing for Peak Periods
    - Focus on September for major promotions and increase stock levels in high-demand categories.
- Improve Performance of Underperforming Categories
    - Analyze why Underwear/Nightwear underperforms (pricing, demand, marketing).
    - Consider discounts, better visibility, or bundling strategies.
- Stock & Promote Best-Selling Items
    - Ensure trousers and dresses are well-stocked.
    - Promote these items on social media and e-commerce platforms.
- Align Product Designs with Customer Preferences
    - Continue focusing on black and solid-patterned products.
    - Experiment with limited edition colors and patterns based on customer insights.
- Expand Product Offerings
    - Increase variety in lower-body and full-body garments to diversify the catalog.
- Enhance Sustainability Messaging
    - Highlight the use of organic cotton and sustainable materials in marketing campaigns.
    - Educate customers on eco-friendly fashion choices.

Conclusion

H&M's sales and product analysis reveals strong seasonal trends, product demand variations, and customer preferences. The company can enhance revenue by optimizing inventory, focusing on top-selling categories, adjusting marketing strategies, and expanding product variety while maintaining a customer-centric and sustainability-focused approach.