**E-commerce Business Intelligence through Data Analytics**

Group Members

Institution of Affiliation

Date

Table of Contents

[**Introduction** 3](#_Toc195196806)

[**Methodology** 4](#_Toc195196807)

[**Results** 5](#_Toc195196808)

[**Tableau dashboards** 5](#_Toc195196809)

[1. Website Activity analysis 5](#_Toc195196810)

[2. Demographic Composition Dashboard 6](#_Toc195196811)

[3. Product Performance Dashboard 8](#_Toc195196812)

[4. Inventory Status Dashboard 9](#_Toc195196813)

[**Predictive modelling** 10](#_Toc195196814)

[**Conclusion and Recommendations** 10](#_Toc195196815)

# **Introduction**

In today’s digital economy, e-commerce companies operate in highly competitive environments where understanding user behavior, product performance, and operational efficiency is vital to maintaining a competitive edge. Decision-makers need access to reliable data and insightful analytics to respond to market trends, optimize marketing strategies, streamline inventory, and enhance customer satisfaction.

This project leverages the bigquery-public-data.thelook\_ecommerce dataset, an anonymized dataset representing a typical e-commerce business to build an end-to-end data pipeline (Google, 2022). SQL is used for data extraction from Google BigQuery, Python is employed for analytical and predictive modeling, and Tableau is used to design interactive dashboards for business storytelling. The final deliverable consists of four themed dashboards, each addressing a key business question with supporting metrics and charts.

**Research Questions**

1. Website Activity Dashboard:  
   *How are users interacting with the website, and what are the top traffic sources and user behaviors over time?*
2. Demographic Composition Dashboard:  
   *What are the demographic characteristics of our customer base, and how do they vary by gender, age, and geography?*
3. Product Performance Dashboard:  
   *Which products, categories, and brands are driving the most revenue and sales volume over time?*
4. Inventory Status Dashboard:  
   *How well is the inventory being managed, and what trends can be observed in inventory availability, aging, and distribution performance?*

By addressing these questions, this analysis equips business stakeholders with the insights necessary to make data-driven decisions for marketing, product planning, and operations.

# **Methodology**

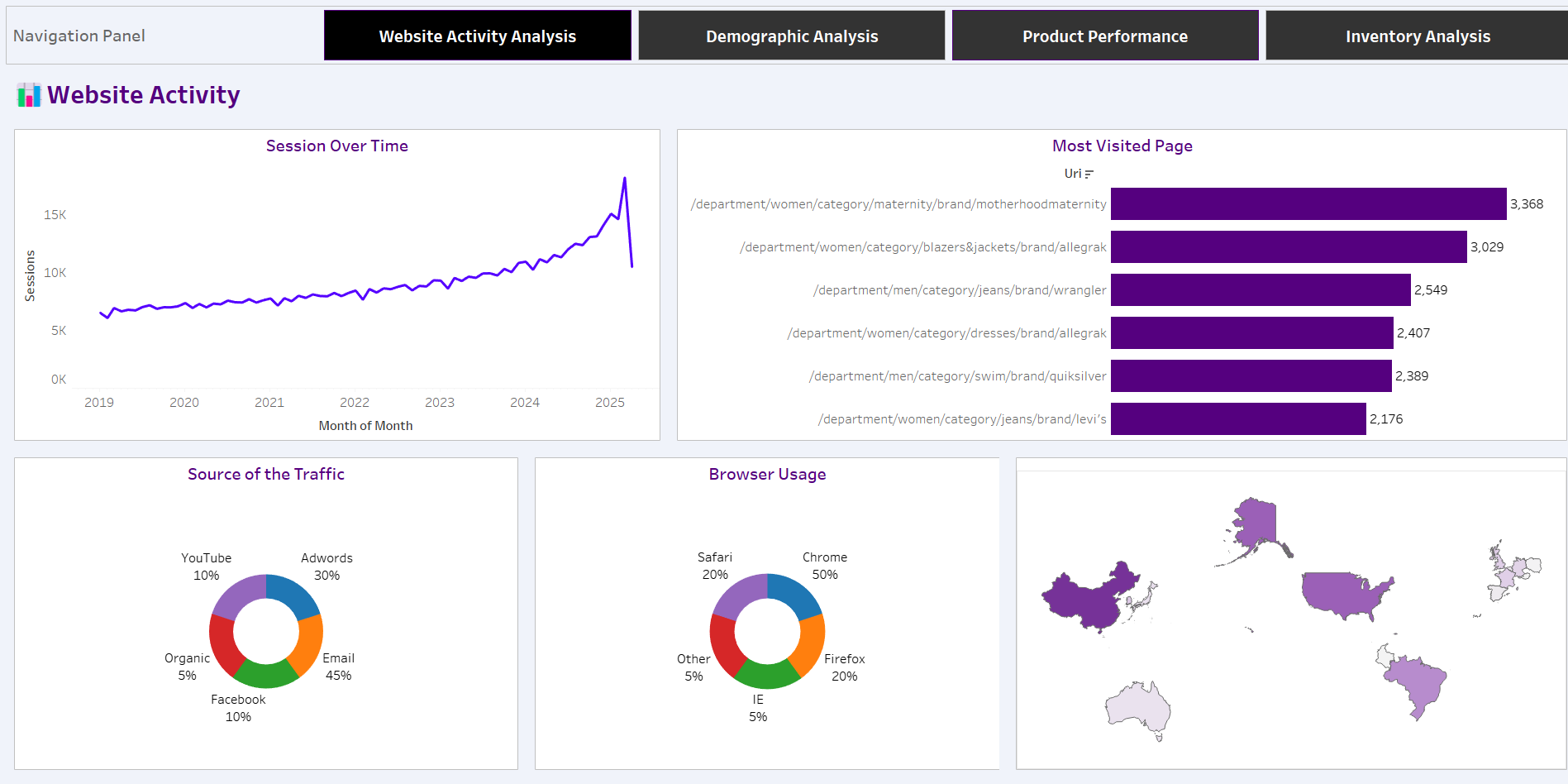
The data for this project was retrieved from Google BigQuery using SQL queries to extract relevant insights from the `bigquery-public-data.thelook\_ecommerce` dataset (Google, 2022). SQL was used to join multiple tables—such as `users`, `orders`, `order\_items`, `products`, `events`, and `inventory\_items`—to prepare datasets that reflect key business metrics like sales performance, user demographics, website activity, and inventory status. These queries were executed through a Python environment in Google Colab (Google, 2025), allowing seamless integration with BigQuery using the BigQuery client library.

After extracting and transforming the data, Python was used for further analysis and predictive modeling, including predicting user age using classification models. The cleaned and structured data was then exported to Google Drive in CSV format and imported into Tableau for visualization. Tableau served as the storytelling and dashboarding tool, where interactive charts were used to build the dashboards (Tableau Software, 2025).

# **Results**

## **Tableau dashboards**

### Website Activity analysis



The screenshot above represents a website activity dashboard. The dashboard was created to present an overview of how users interact with the website across different dimensions. The first chart in the dashboard shows the sessions over time. The chart shows an upward trend in the website traffic from 2018 to 2025. This shows that user engagement has increased over the years. The chart also indicates a spike in 2025, this suggest there was a significant event such as seasonal sale, product launch, or promotional campaign that temporarily boosted the activity which was followed by normalization or seasonal pattern.

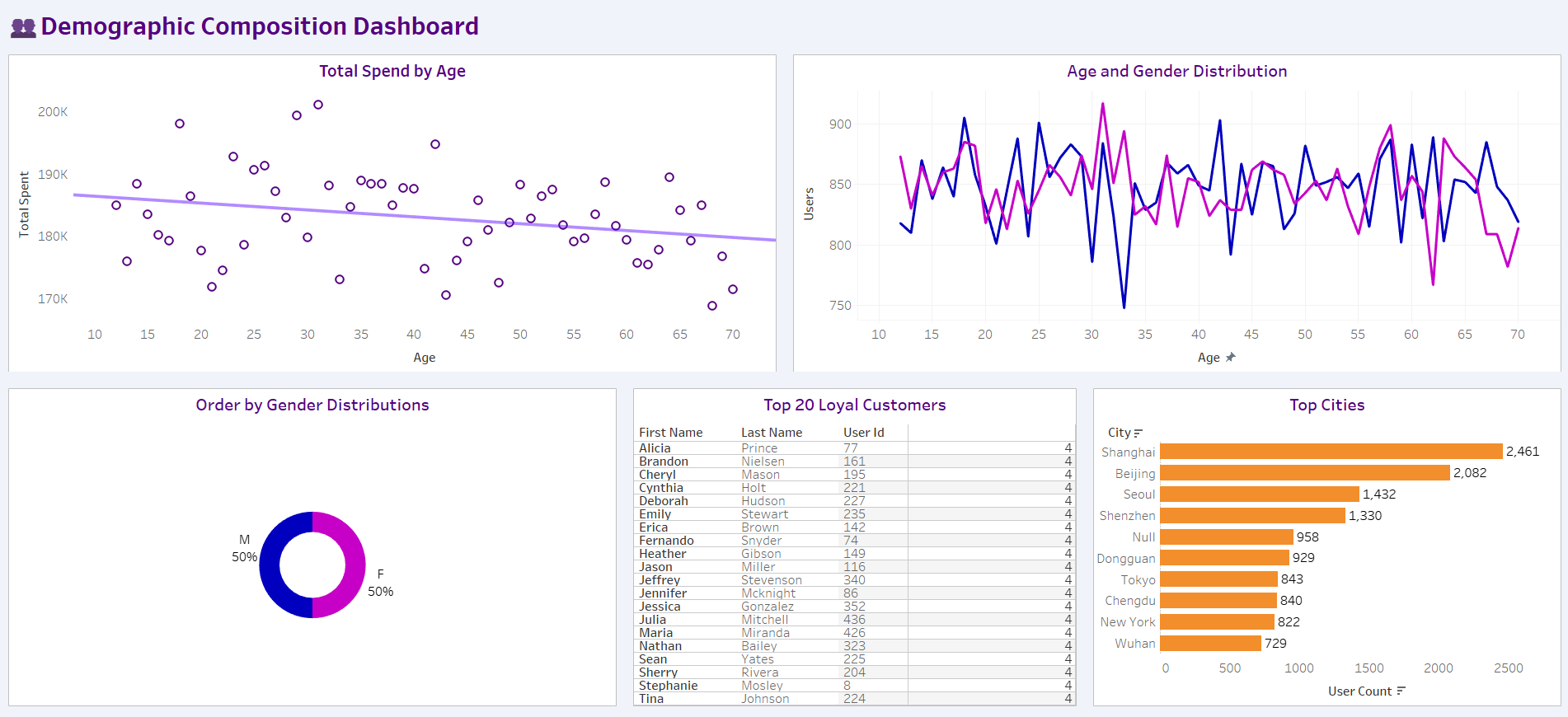
The dashboard also has a bar chart that shows the most frequently accessed URL. The chart reveals that women’s maternity wear specifically /department/women/category/maternity/brand/motherhoodmaternity, which received 3,368 visits. Other pages that followed include categories such as women blazers, men’s jeans and swimwear. From that we can conclude that there is a high level of users interest in fashion categories, with a notable emphasis on women’s clothing and branded items.

In terms of traffic sources, email emerges as the dominant channel, accounting for 45% of all website traffic. This is followed by paid search (Adwords) at 30%, with social media platforms like Facebook and YouTube each contributing 10%. Organic search and other channels make up a smaller share. From that we can conclude that the reliance on email marketing highlights the effectiveness of targeted campaigns and suggests that maintaining a strong email strategy is important for continued engagement.

The browser usage chart shows that 50% of users access the website via Chrome, making it the leading browser by a wide margin. Firefox and Safari follow, each with a 20% share, while Internet Explorer and other browsers collectively represent just 10%. These figures suggest that website performance and design should be optimized primarily for Chrome, followed by Firefox and Safari to ensure a smooth user experience for the majority of visitors.

Lastly, the geographic map and session data indicate that China is the largest source of traffic with 440,999 sessions, followed by the United States (290,920), Brazil (186,120), South Korea (70,242), and the United Kingdom (60,118). This highlights a strong international presence, particularly in Asia and the Americas. Targeted localization strategies and region-specific marketing could further enhance engagement in these key markets.

### Demographic Composition Dashboard

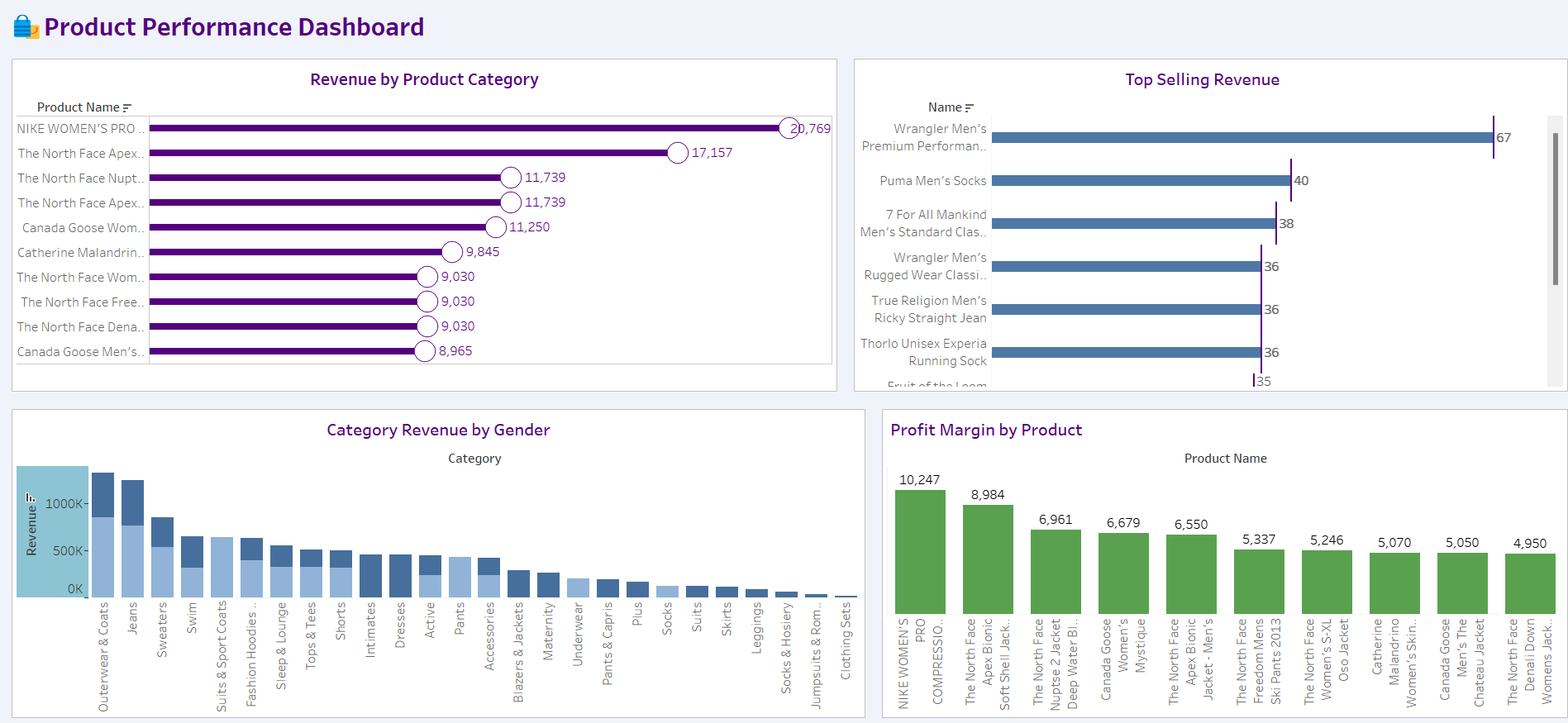


The screenshot above represents a demographic composition dashboard. The dashboard was created to present valuable insights into customer behavior based on age, gender, and geographic distribution. First, the scatterplot in the top left corner illustrates the Total Spend by Age, where each dot represents aggregated spending for users of a particular age. A fitted linear trend line reveals a slight decline in total spending as age increases, with a coefficient of -110.15 and a statistically significant p-value of 0.0419. Although the R-squared value is low (0.07), indicating a weak model fit, the negative slope suggests younger users generally spend more than older ones, albeit marginally.

The Age and Gender Distribution line chart compares user counts across age groups for males (blue) and females (pink). The two lines are closely aligned, showing a fairly even age distribution across genders, with noticeable spikes in both segments between the ages of 20 to 35. Some fluctuations exist at different ages, but no single gender appears to dominate overall. This reinforces a balanced gender presence on the platform. The Order by Gender Distributions donut chart further supports this balance, showing a 50-50 split between male and female order counts. This suggests that marketing efforts can be uniformly targeted toward both genders without needing significant customization based on gender identity.

The Top 20 Loyal Customers table highlights individuals with the highest repeat engagement, each making 4 orders. This includes a mix of male and female customers, indicating that loyalty spans across both demographics. On the bottom right, the Top Cities bar chart identifies Shanghai (2,461 users) and Beijing (2,082 users) as the most populated user bases, followed by Seoul, Shenzhen, and Dongguan. Major cities in China and East Asia dominate the top list, affirming the platform’s strong presence in these regions. Additionally, New York appears among the top cities, indicating a reach that extends into the western market.

### Product Performance Dashboard



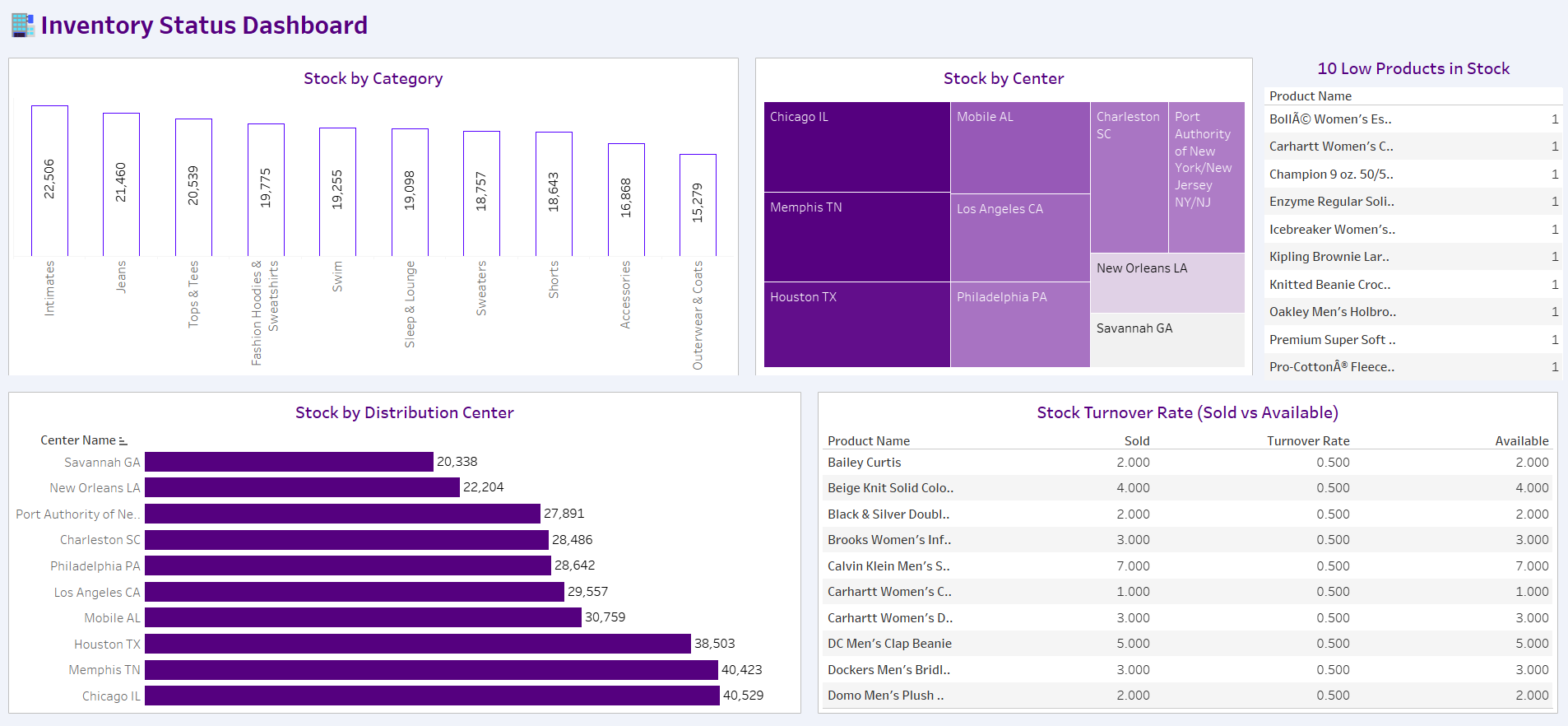
The screenshot above represents a product performance dashboard which provides a comprehensive overview of sales, revenue, and profitability across product lines, helping identify top-performing items and categories. In the top left, the Revenue by Product Category chart reveals that NIKE WOMEN'S PRO COMPRESSION SHORTS is the top revenue-generating product with 20,769 in revenue, far surpassing the next highest, The North Face Apex Bionic Jacket, which generated 17,157. This indicates strong demand for activewear, especially among women. Several North Face and Canada Goose products also perform well, highlighting the popularity of premium outdoor and cold-weather apparel.

To the right, the Top Selling Revenue bar chart showcases units sold rather than revenue. Wrangler Men’s Premium Performance Jean leads with 67 units sold, followed by items like Puma Men’s Socks and 7 For All Mankind Men’s Standard Classic Straight Jean. While these items have high sales volumes, they may not necessarily top the revenue list, suggesting they are more affordable but widely popular, especially among male customers.

In the bottom left, the Category Revenue by Gender bar chart reveals that Outerwear & Coats, Jeans, and Sweaters dominate revenue across both genders. Male-associated categories are shaded darker, and female-associated ones lighter. While both genders contribute to high-revenue categories, there is a clear difference in preference—Sleep & Lounge, Dresses, and Intimates perform better among female consumers, while Jeans and Activewear show a more balanced gender split.

Lastly, the Profit Margin by Product bar chart on the bottom right highlights the most profitable items. Once again, NIKE WOMEN’S PRO COMPRESSION SHORTS tops the list with a margin of 10,247, followed by The North Face Apex Bionic Jacket and North Face Soft Shell Jacket. Profit margins are generally higher for branded outerwear and performance gear, suggesting that while these items may have a higher cost, their markup allows for strong profitability.

### Inventory Status Dashboard



The screenshot above represents inventory status dashboard that gives a clear picture of how products are distributed across categories and fulfillment centers, and it helps us understand which items might need attention. Looking at the Stock by Category chart, we can see that items like Intimates, Jeans, and Tops & Tees are stocked in high numbers, with over 20,000 units each. On the other hand, Outerwear & Coats, which were top performers in revenue on the Product Performance Dashboard, are sitting at the bottom here with just over 15,000 units.

On the distribution centers, it's clear that Chicago, Memphis, and Houston are the heavy lifters, each holding upwards of 38,000 units. In contrast, Savannah GA and New Orleans LA are operating with much smaller stock levels, which might limit how quickly orders can be fulfilled in those regions. This is visualized in both the treemap and the bar chart.

On the 10 Low Products in Stock table, several items, including women’s outerwear and popular casual wear like Champion and Oakley products, are down to just one unit left. These could very easily go out of stock completely if no action is taken. Finally, the Stock Turnover Rate section gives us a peek into product movement. Every product listed has a turnover rate of 0.5, meaning they’ve only sold half of what’s available. While that’s not terrible, it does suggest these items are moving slowly. Some may benefit from promotions or markdowns to clear out space and drive sales.

## **Predictive modelling**

A linear regression was used to predict users’ exact age based on their shopping behavior—things like total orders, number of unique products purchased, how much they spent, their average order value, the variety of product categories and departments they bought from, along with their gender, country, traffic source, and even the day and time they placed orders. However, the model didn’t perform well at all. With an R-squared value of -0.00%, it essentially meant the model couldn’t explain any variation in age from the available data, and the predictions were almost as good as guessing.

Given these poor results, we took a different approach: instead of trying to predict a specific age, we grouped ages into broader categories (like "<25", "25–34", "35–44", and so on) and used a classification model—specifically, a Random Forest classifier. Using the same features as before, we trained the model to classify users into age groups rather than predict an exact number. This approach gave us slightly better results, with an overall accuracy of 18.15%. While this is still relatively low, it suggests that age is not strongly reflected in transaction-level behavior alone.

# **Conclusion and Recommendations**

The analysis reveal several key conclusions that call for strategic actions. First, while products like NIKE Women’s Pro Compression and The North Face outerwear are driving significant revenue and profit, there is a mismatch between top-selling items and those with the highest profit margins—highlighting the need to evaluate pricing and promotional strategies to better capitalize on high-volume products. Secondly, gender and category-based revenue differences suggest that marketing and product placement can be more targeted to match consumer demand. On the inventory side, categories like Intimates and Jeans are overstocked, while high-turnover items such as Outerwear & Coats have relatively low stock, signaling a need for rebalancing inventory levels. Additionally, distribution centers like Savannah and New Orleans carry less inventory compared to Chicago or Memphis, which may be contributing to the stock shortages seen in high-demand products. Finally, low stock turnover rates across many products indicate potential overstocking or slow-moving items, which ties up capital and warehouse space.

To address these issues, the business should optimize inventory by increasing stock for fast-moving, profitable products, reduce inventory in underperforming categories, and adjust marketing and pricing strategies to boost profitability on top-selling items. Regular stock reviews and improved demand forecasting will also help align supply with customer demand, ultimately improving operational efficiency and profitability.

References

Google. (2022). *The Look eCommerce Dataset* [Data set]. Google BigQuery Public Datasets. [https://console.cloud.google.com/bigquery?authuser=1&project=big-query- data-projects&ws=!1m4!1m3!3m2!1sbigquery-public data!2sthelook\_ecommerce&pli=1&invt=AbuYvw](https://console.cloud.google.com/bigquery?authuser=1&project=big-query-%09data-projects&ws=!1m4!1m3!3m2!1sbigquery-public%20%09data!2sthelook_ecommerce&pli=1&invt=AbuYvw) *(Accessed April 9, 2025)*

Tableau Software. (2025). *Tableau Desktop (Version 2025.x)* [Computer software]. <https://www.tableau.com/>

Google. (2025). Google Colaboratory. Retrieved April 09, 2025, from <https://colab.research.google.com/>