**Lululemon Semi and Unstructured Data Sources**

**First Data Source: Lululemon and Scraped Product Reviews**

Lululemon can leverage product reviews from their website and combine it with scraped reviews from their competitors’ sites and review forums. Lululemon’s personal reviews will provide valuable feedback on fit, quality, and customer satisfaction, offering a direct understanding of consumer preferences and beliefs about the Lululemon brand. Reviews scraped from competitor websites like Nike, Alo, Athleta, Vuori, Adidas, and other review forums will offer insights into industry trends, competitor strengths, and customer preferences for alternative products. By integrating these scraped reviews into an Amazon S3 data lake, which aligns with Lululemon's existing AWS infrastructure, the data can be preprocessed and analyzed effectively. Conversely, product reviews gathered directly from Lululemon’s website can be stored in the Orders and Sales databaseunder the customer feedback section, allowing integration with existing transactional and customer data. Sentiment analysis can be utilized to identify how customers feel about specific features, such as material quality or style, and highlight strengths and pain points in comparison to competitors. Topic modeling can uncover key themes or topics driving customer conversations, helping Lululemon identify trends and product attributes customers value most. Additionally, lift analysis can reveal unique attributes or product qualities associated with positive sentiment, distinguishing Lululemon from its competitors. This approach is valuable because these insights can be used to refine product designs and tailor marketing strategies to better align with customer preferences and beliefs, strengthening Lululemon's market position. Furthermore, by aligning its products and brand strategy with evolving customer needs and preferences, they will strengthen customer loyalty, attract new customers, and maintain its competitive edge in the activewear industry.

**Second Data Source: Customer Purchase and Return Data**

Lululemon can also incorporate customer purchase and return data to enhance personalization and improve customer experiences. Purchase data, such as reasons for buying, whether it’s a gift, a new product release, a sale, or just because, provides insights into customer motivations, enabling tailored marketing strategies. Return data can be collected both in-store through employee engagement with customers and online through questionnaires. When returning a product online, customers will be prompted to select a reason for their return with options including size issues, damaged product, or material dissatisfaction, and will also have the opportunity to discuss details in the comments section. Lululemon will be able to delve into insights that are more unique, as most stores and companies don’t ask why a customer is buying and returning. Both of these insights offer valuable feedback to refine customer service and experience. For example, if a customer is buying a gift, we know not to include this purchase into their product recommendations. To integrate these data sources effectively, Lululemon must standardize data inputs to ensure consistency across all channels, such as aligning purchase and return reasons into predefined categories for easier analysis. Then, these insights must be integrated into their Order and Sales System (purchase data) and Customer Support and Service Management System (return data) to enable personalized marketing strategies and linked to their inventory systems to inform product adjustments and demand forecasting. By creating this framework, Lululemon can enhance customer personalization, streamline operations, and drive continuous product development improvements.

# **Exploring semi-structured and unstructured data Outline**

1. Product reviews from lululemons sight and compare with reviews from scraped websites like nike, alo, athleta, vuori, adidas
   1. Conduct sentiment analysis, topic modeling, using nlp to gain insights about customer preference
   2. Data lake
      1. Amazon S3, Azure Data Lake, Google Cloud Storage data lake (The raw data does not need immediate integration with OLTP but serves as a source for preprocessing and analytics.)
      2. Lululemon uses AWS <https://aws.amazon.com/solutions/case-studies/lululemon-athletica/>
   3. Can be used for designing new products
   4. Enhance customer experience, personalized experiences
2. Why customers return/buy things - use this data to enhance customer personalization
   1. Buy (based on total purchase)
      1. Why did you buy - gift for someone, just because, needed for specific reason, new product came out, sale
   2. Return (per item basis)
      1. Why did return (in store) - have employee ask and type in answer
      2. Why did you return (online) - have questions
         1. Too big, too small, didn’t arrive in time, defective/damaged product, didn’t like the material, unexpected color, other
         2. Section for more detailed comments