**Business Problem:**

Executives from Jiffyshirts.com are interested in learning how their dataset might be used to create a model that could have commercial value. What products, for instance, should we suggest to their most loyal clients? Why? Data manipulation, exploratory data analysis (EDA), and model development are carried out to address this business issue.

**Data Preparation**

A good representative sample of the company's sales data was included in the datasets for analysis. An online store with a concentration on "soft goods" including t-shirts, sweaters, and bags, is JiffyShirts.com, which is situated in North America. Data were presented in three sets: 1) The orders dataset, which contains eight columns and 46433 records; 2) The customers’ dataset, which contains five columns and 10,000 records; and 3) The line items dataset, which contains eight columns and 373878 records. Data-wrangling was done as part of the data cleaning and preparation (outlier detection, handling missing values, altering variables, etc.).

**EDA (Exploratory Data Analysis) and Feature Selection**

The EDA helps in addressing some of the business problems. According to the analysis results, customers pay only 35.6% of shipping fees. The EDA results also revealed that Customers that have indicated they are a business used the organic search and paid social acquisition channels while customers who have created user accounts with JiffyShirts.com do so through referral networks, direct, organic search, sponsored search, coupon aggregators, and direct marketing acquisition channels. The analysis further shows that the state where consumers pay the highest delivery costs is Texas (TX). Similarly, Texas (TX) is the state where Jiffyshirts.com pays the highest delivery expenses as revealed in the EDA analysis. The analysis also answers other questions like knowing the top-selling products, customers that bought the most items by category, and so on.

**Model**

Product recommendation and clustering models were built to answer some of the business questions further. Questions like “What product categories should we recommend to our top customers?” The results from the product recommendations model indicated that Basic tee, fashion tee, and other product categories can be recommended to Jiffyshirts.com’s top customers.

**Model Evaluation**

AUC, Recall, and Precision are three classification accuracy metrics that are used to assess how well recommendation system models work. The metrics Silhouette Score, Calinski Harabasz Score, and LGBMClassifier are frequently employed in cluster analysis.

**Conclusion and Recommendations**

The analysis's results demonstrate that data analytics can solve Jiffyshirts' business issues, enabling the enterprise to operate more productively and efficiently. It is important to compare several models as they are being built and trained in order to improve model optimizations and performance. For instance, to determine which model performs the best, we can compare the effectiveness of our KPrototypes (Clustering with Mixed Data Types) and KMeans with other methods, such as CLTree (Clustering Through Decision Tree Construction). While KMeans performs well with continuous (numerical) Data Types, KPrototypes performs well with mixed data types. CLTree clustering model uses the supervised machine-learning approach.