1. Data Description:

- The provided dataset contains information from a marketing campaign, including details about the product, campaign convergence results, and customer-specific information.

- The attributes include:

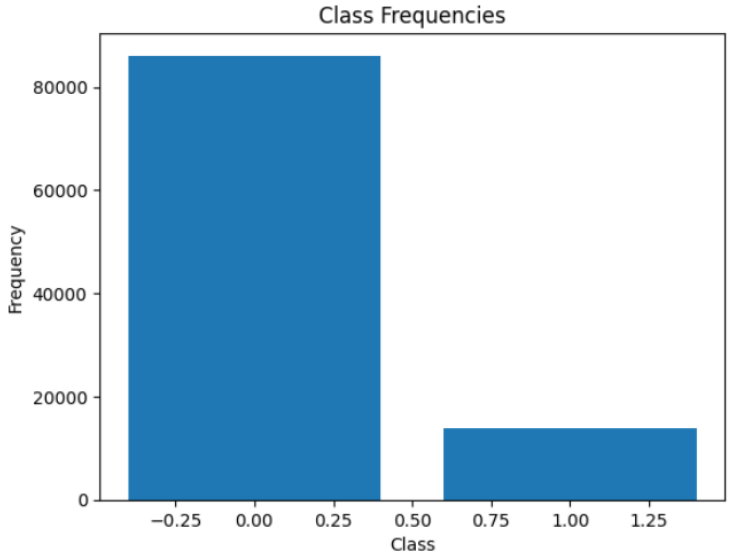
- country: Country name
- article: 6 digit article number, as unique identifier of an article
- sales: total number of units sold in respective retail week
- regular_price: recommended retail price of the article
- current_price: current selling price (weighted average over the week)
- ratio: price ratio as current_price/regular_price, such that price discount is 1-ratio
- retailweek: start date of the retail week
- promo1: indicator for media advertisement, taking 1 in weeks of activation and 0 otherwise
- promo2: indicator for store events, taking 1 in weeks with events and 0 otherwise
- customer_id: customer unique identifier, one id per customer
- article: 6 digit article number, as unique identifier of an article
- productgroup: product group the article belongs to
- category: product category the article belongs to
- cost: total costs of the article (assumed to be fixed over time)
- style: description of article design
- sizes: size range in which article is available
- gender: gender of target consumer of the article
- rgb_*_main_color: intensity of the red (r), green (g), and blue (b) primaries of the article‘s main color, taking values [0,250]
- rgb_*_sec_color: intensity of the red (r), green (g), and blue (b) primaries of the article‘s secondary, color, taking values [0,250]
- label: advertisement result after offering/sending/presenting the offer to the customer. 0 means the customer did not buy and 1 means the costumer did buy..

2. Data Preprocessing Steps:

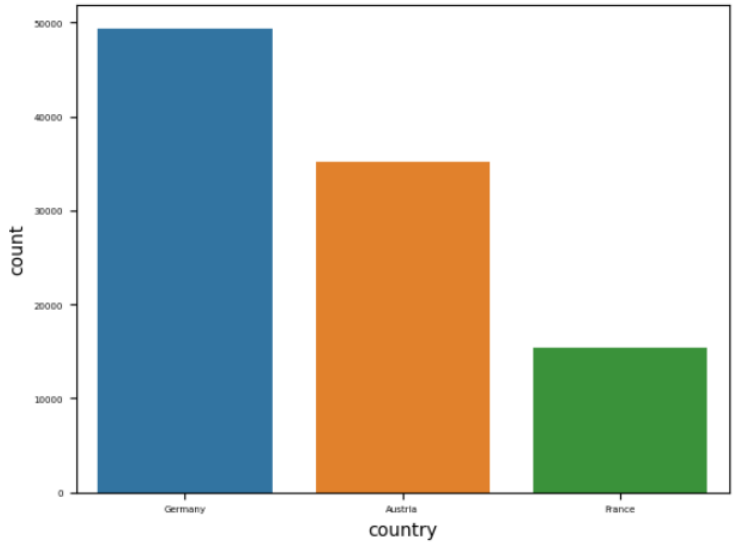
- The dataset did not contain any missing values.
- New Features Extraction:
  - Hypothesis 1: To consider the impact of the discounted amount on purchasing behavior, a new feature called price_difference was created, which represents the difference between the regular_price and the current_price.
  - Hypothesis 2: To analyze the influence of the offering time, new features were generated for each week of the month and a feature indicating the month number.
- Data Normalization:
  - The data was normalized using standard scaling and one-hot encoding techniques.

- Data Resampling:
  - Due to class imbalance, with class 0 occurring approximately six times more frequently than class 1, upsampling was performed to increase the number of samples in the minority class.
- PCA:
  - After extracting relevant features, PCA (Principal Component Analysis) was applied to reduce the dimensionality of the problem, aiming to decrease runtime while preserving important information.

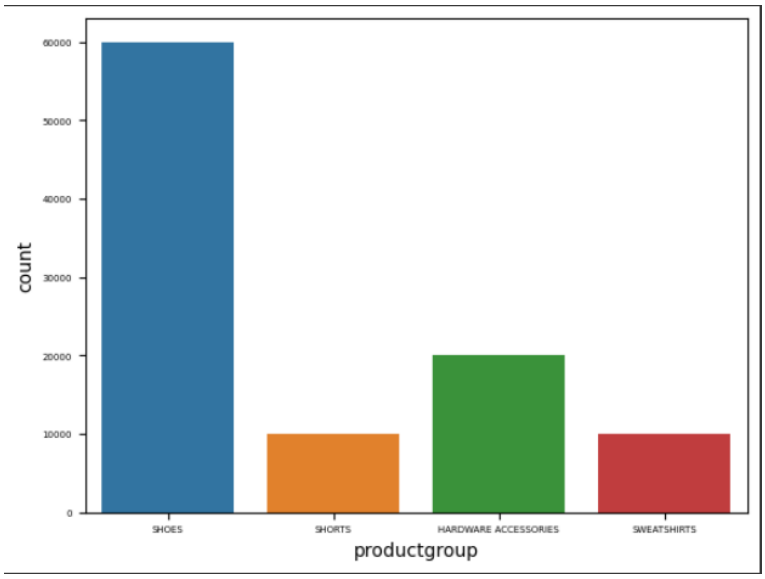
3. Data Visualization:



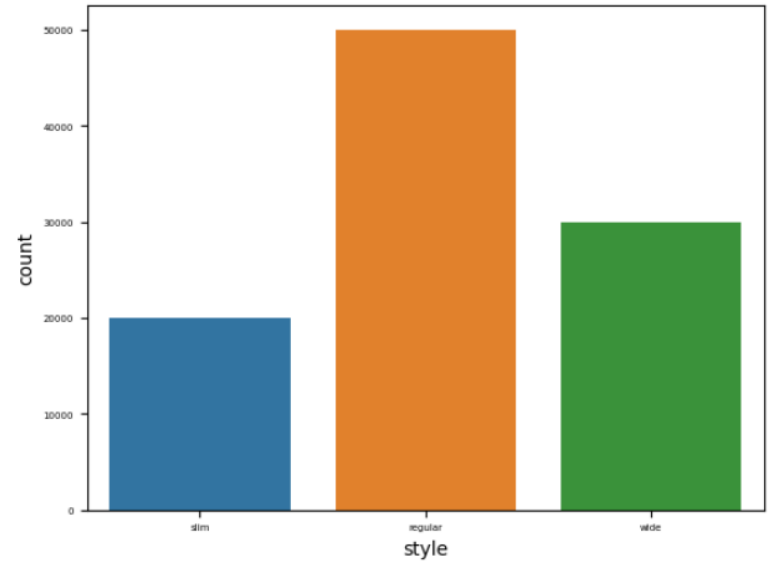
Major class : 0



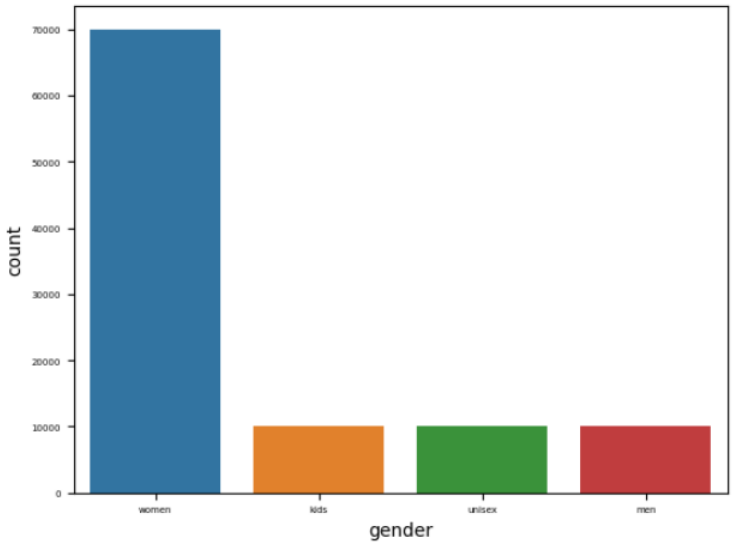
Major country : Germany



Major Product Group: Shoes

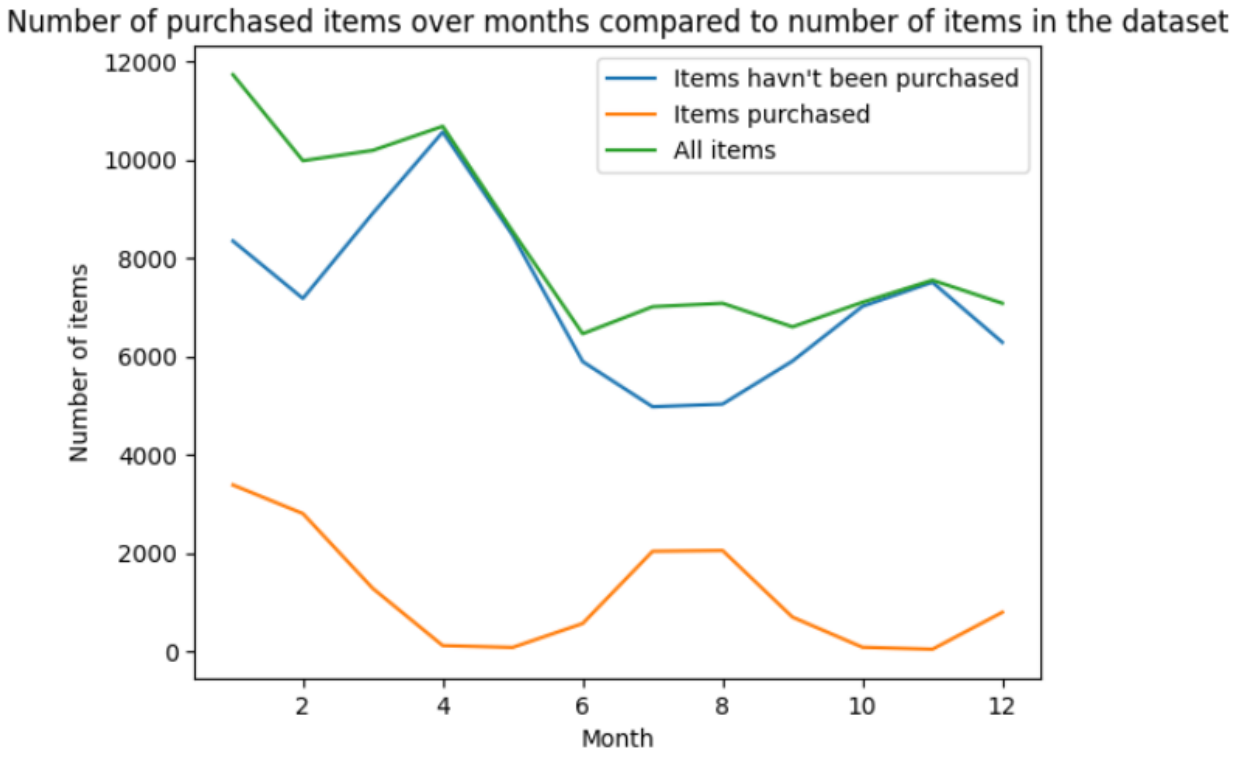


Major Style: Regular

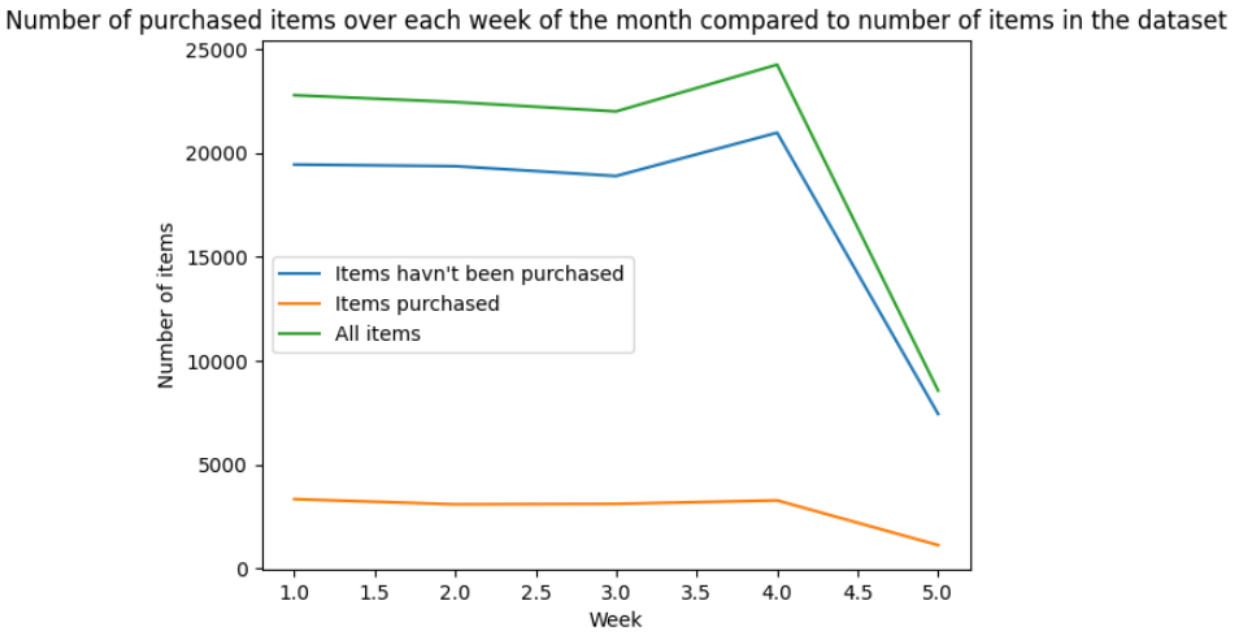


Major Gender : Woman

4. Key Findings:



Discovered that the 6 -> 8 and 10 -> months have the fewer number of items purchased.



Discovered that people's purchasing power decreases by the end of every month.

4. Modeling:

Using a Random Forest model on the dataset can help generate predictions and gain further insights. Here's the approach applied the Random Forest algorithm and extracting insights:

Data Preparation:

Splited the dataset into training and testing sets. Typically, a common split is 80% for training and 20% for testing.

Separated the target variable (label) from the feature variables.

Performed any necessary feature encoding or scaling, such as one-hot encoding for categorical variables and scaling numerical variables.

Random Forest Model Training:

Trained a Random Forest classifier on the training data using the scikit-learn library in Python.

Model Evaluation:

Evaluated the performance of the trained Random Forest model on the testing data.

	precision	recall	f1-score	support
class 0	1.00	0.93	0.96	17215
class 1	0.93	1.00	0.96	17200
accuracy			0.96	34415
macro avg	0.96	0.96	0.96	34415
weighted avg	0.97	0.96	0.96	34415

5. Recommendations & Conclusion:

Pricing Strategy:

Consider the impact of the discounted amount on purchasing behavior. The newly created feature, "price_difference," can help analyze the relationship between the discount amount and sales. Evaluate whether larger discounts lead to increased sales and adjust pricing strategies accordingly.

Timing of Offers:

Take into account the influence of the offering time on customers' purchasing behavior. The generated features for each week of the month and the month number can provide insights into when customers are more likely to make purchases. Align marketing campaigns and offers with periods when customers are more receptive and have higher purchasing power.

Country-Specific Marketing:

Focus marketing efforts on the major country identified in the data analysis (Germany). Tailor campaigns to the preferences and behaviors of customers in that specific country. Consider cultural, economic, and social factors that may influence purchasing decisions.

Product Group and Style Recommendations:

Pay attention to the major product group (Shoes) and style (Regular) identified in the data analysis. Allocate resources to promote and highlight products within this category. Explore trends and customer preferences within the identified product group and style to align product offerings with market demand.

Gender-Specific Targeting:

Utilize the gender information provided in the dataset to customize marketing messages and promotions to specific target audiences. Consider the different preferences, needs, and purchasing patterns between male and female customers.

Monthly Variation in Purchasing Power:

Recognize the trend of decreasing purchasing power towards the end of every month. Plan marketing campaigns and offers that align with customers' financial cycles, focusing on periods when customers have higher disposable income.