

H&M data deep dive / chap.1 Understand article

1. article 자료를 이용한 변수 분석.

set_gender_freq(X) 함수를 설정

⇒ rs_for_male / rs_for_female / rs_for_mama 변수 추가.

Menswear	Ladies wear / divided	season/male = mama
Baby / children, sport (boy / men)	Baby / children, sport (girl, ladies)	

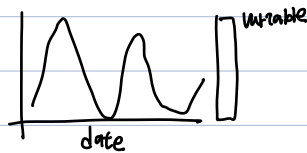
2. transactions_train 자료를 이용 (product sales seasonality)

→ 데이터의 크기가 매우 커서 sample(Arc=0.1)을 이용하여 downsampling을 진행

날짜에 해당하는 변수만을 가짐 → dt_trans

① 위에서 이용한 자료에서 dt_article pd.merge (article_id를 기준으로)
→ dt에 저장

dt를 더 가격, 날짜로 group전행,



3. Classify seasonal products by GMM clustering.

Customer Segmentation: modelling & visuals

Gender 별 plot을 그려서 확인.

- distribution of Age
- distribution of Annual Income
- " spending Score

kmeans / silhouette analysis