

Title: Fashion Trend Visualization (Group H)

Names of all project participants:

Eunji Kim ek3223@columbia.edu
Dustin Nguyen dn2519@columbia.edu
Xintong Tang xt2249@columbia.edu
Frances Yang zy2479@columbia.edu

Abstract: The project seeks to understand characteristics of clothing trends based on apparel offered on popular e-commerce sites. The data collected provides site categorizations of products up for sale such as main categories (Apparel, Accessories, etc.), subcategories (Top, Shoes, etc.), article type (T-shirts, Handbags, etc.), and other descriptives such as color, usage, year, and season. The visualizations that are planned to be produced are seasonality of color, yearly trends in categories, and clothing usage in various categories. We would also like to explore how clothing usage and article type may trend differently between male and female consumers.

How does baseColour differ from season to season and year to year?

How has the trend in certain subCategory changed over time?

Frequently purchased brand over years for male and female?

Number of item change in a certain subCategory over the years

Techniques: ggplot2, NLP text analysis, interaction

Data Description:

Dataset: <https://www.kaggle.com/paramagarwal/fashion-product-images-dataset>

Smaller version: <https://www.kaggle.com/paramagarwal/fashion-product-images-small>

The dataset is from 2010-2018, with 31136 unique values of various apparel, accessories, footwear, and in some cases personal care - all products that are sold by big name fashion brands. Cataloged products were scraped from e-commerce websites and organized into columns and categories as described below. This dataset can give us a unique insight into fashion trends from 2010-2018.

The dataset contains the following columns:

- id: a unique identifier for the product
- gender: the gender a product was marketed towards (Men, Women, Other)
- masterCategory: overall category for the product (Apparel, Accessories, Other)
- subCategory: more specifically, what the product is (Topwear, Shoes, Other)
- articleType: specifically, what the product is (Tshirt, Shirts, Handbags, etc)

- baseColour: color of the product
- season: season that the product was released (Fall, Winter, Spring, Summer)
- year: year that the product was released
- usage: type of style of the product
- productDisplayName: a description of the product itself, including brand name

Twitter API data: we can web scrape sentiment data from twitter with subCategory keywords on individual's opinions by region (country).

Visualizations:

Word cloud: to see which articles of clothing were the most popular during each year

Line chart: number of item change in a certain subCategory over the years

Bar chart: How does baseColour differ from season to season and year to year?

Point chart: size of the number of items in subcategories/article type

Map: using ggmap to show clothing trends differences across the world