

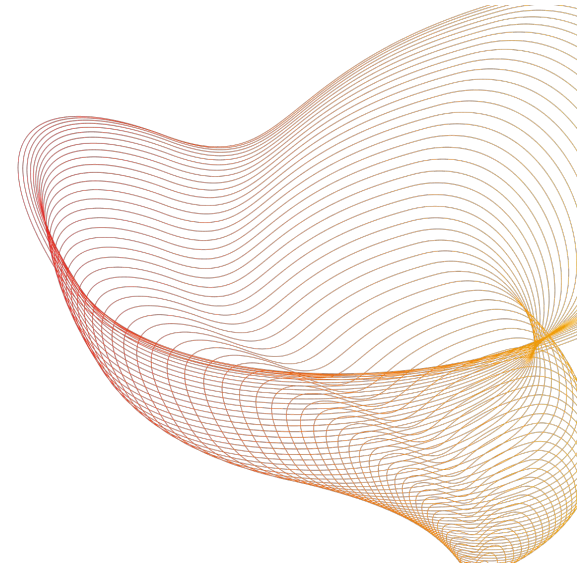


Market Basket Analysis

New Yorker Case Study

Blake Huebner

26.09.23





Technical Problem Definition

Motivation: To provide insights to optimize store layout by section, promote groups of items, etc.

Objective: Identify pairs of items that are often sold together. Items are defined as a section of items (section_name)

Metrics: Lift and Leverage





Metric Definitions

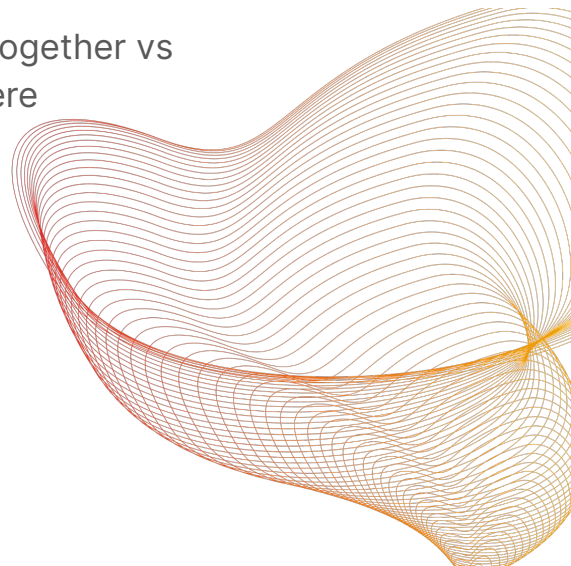
Where Support is defined as the proportion of transactions containing a certain item set:

$$\text{Lift} = \text{Support}(A \& B) / \text{Support}(A) * \text{Support}(B)$$

$$\text{Leverage} = \text{Support}(A \& B) - \text{Support}(A) * \text{Support}(B)$$

- Both measure the difference between how often items are found together vs how often they would be expected to be found together if they were independent
- Leverage favors itemsets with larger support

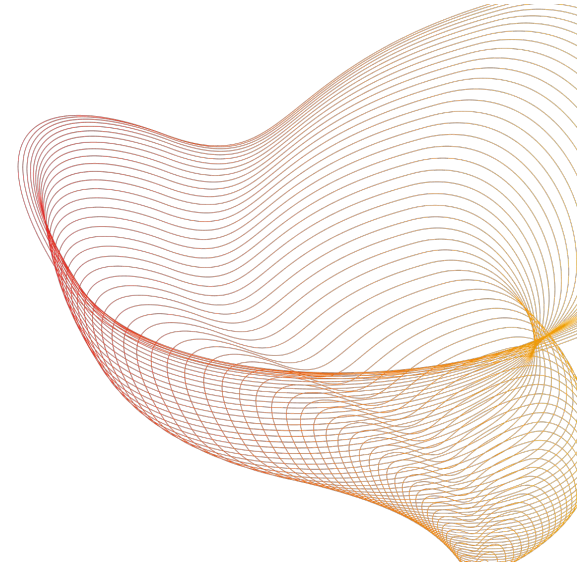
Support is calculated using the FP Growth algorithm





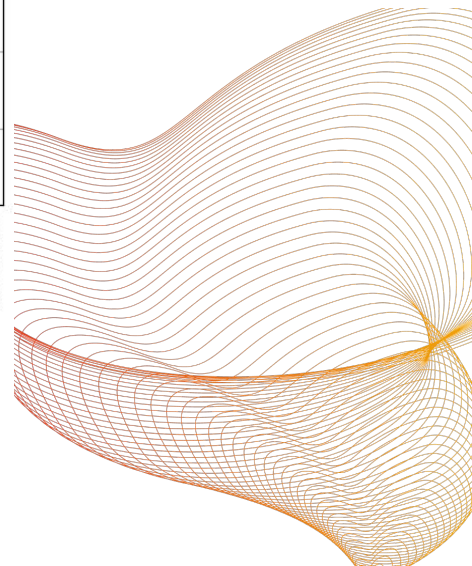
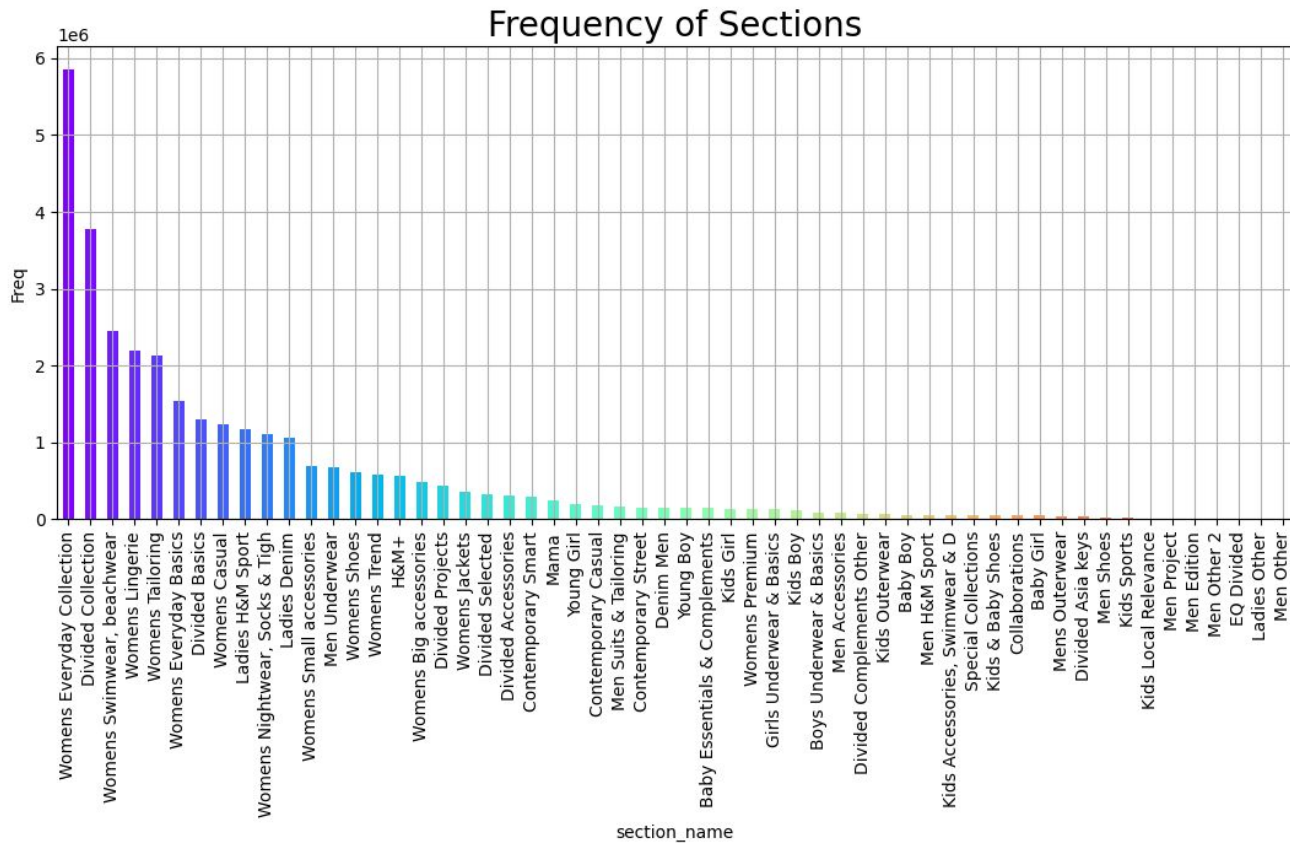
Data

- Purchase data from H&M's kaggle competition
 - <https://www.kaggle.com/c/h-and-m-personalized-fashion-recommendations/overview>.
 - Note: Although the purchase data is from their ecommerce store, this analysis will pretend the data is from New Yorker, which has no ecommerce sales





EDA





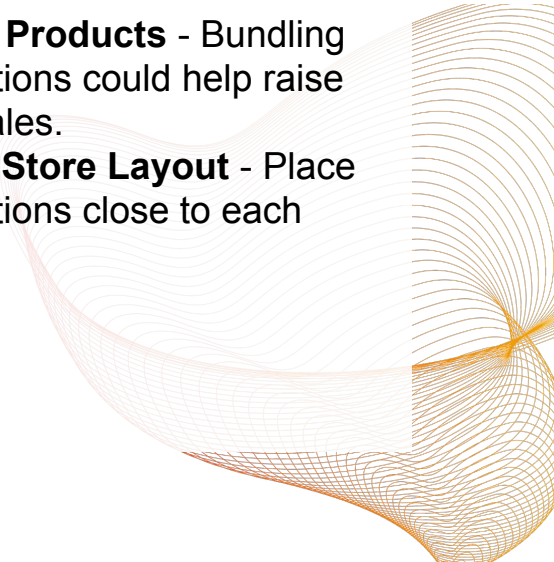
Most Frequent Pairings - Lift (Top 10)

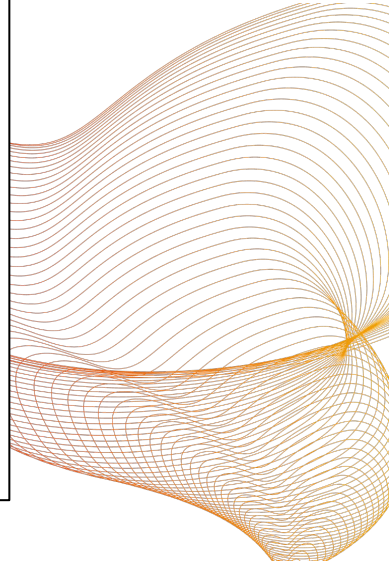
Item 1	Item 2	Support	Lift
Men Other 2	Men Edition	0.000007	69.642006
Baby Girl	Baby Essentials & Complements	0.001071	44.241461
Men Project	Men Edition	0.000009	42.300274
Baby Boy	Baby Essentials & Complements	0.001136	39.789033
Boys Underwear & Basics	Young Boy	0.001476	37.149014
Boys Underwear & Basics	Kids Boy	0.001163	36.773589
Men Project	Men Other 2	0.000006	32.604944
Girls Underwear & Basics	Kids Girl	0.001415	31.722922
Girls Underwear & Basics	Young Girl	0.002269	31.270417
Men Suits & Tailoring	Men Other 2	0.000097	30.659598

Strategy:

These sections have items that are frequently sold together, but not the most popular items.

- **Bundling Products** - Bundling these sections could help raise interest/sales.
- **Optimize Store Layout** - Place these sections close to each other.







Most Frequent Pairings - Leverage (Top 10)

Item 1	Item 2	Support	Lift
Womens Everyday Collection	Womens Tailoring	0.000007	69.642006
Womens Everyday Collection	Womens Casual	0.001071	44.241461
Divided Basics	Divided Collection	0.000009	42.300274
Womens Everyday Collection	Divided Collection	0.001136	39.789033
Womens Everyday Basics	Womens Everyday Collection	0.001476	37.149014
Womens Everyday Basics	Womens Casual	0.001163	36.773589
Divided Basics	Womens Everyday Basics	0.000006	32.604944
Divided Collection	Womens Casual	0.001415	31.722922
Divided Projects	Divided Collection	0.002269	31.270417
Divided Collection	Ladies Denim	0.000097	30.659598

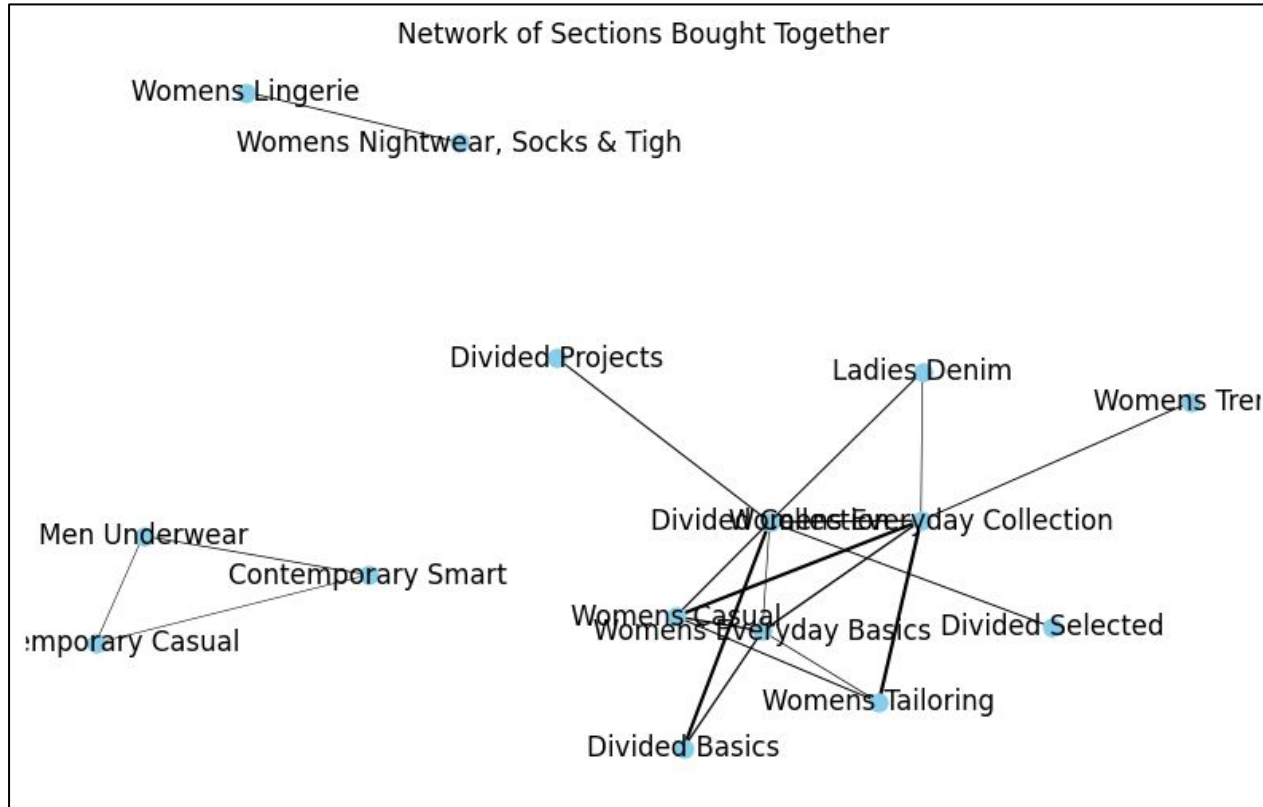
Strategy:

These sections have items are frequently sold together, and are also popular sections.

- **Optimize Store Layout** - Place these sections close to each other.
- **Discounting One Item** - Discount one of these sections, but not the other. Customers are probably willing to pay full price to 'complete' the pair



Most Frequent Pairings - Leverage (Top 20)





Least Frequent Pairings

Out of 9,080,179 transactions, 13 item pairs were never purchased together

Item 1	Item 2
Kids Boy	Men Other
Ladies Other	Men Project
Baby Girl	EQ Divided
EQ Divided	Mens Outerwear
Men Other 2	Men Other
Kids Sports	Men Other
Ladies Other	Men Other 2
Men Project	Men Other
Ladies Other	Men Other
Men Project	EQ Divided
Kids & Baby Shoes	Men Other
EQ Divided	Men Other
Kids Local Relevance	Men Other

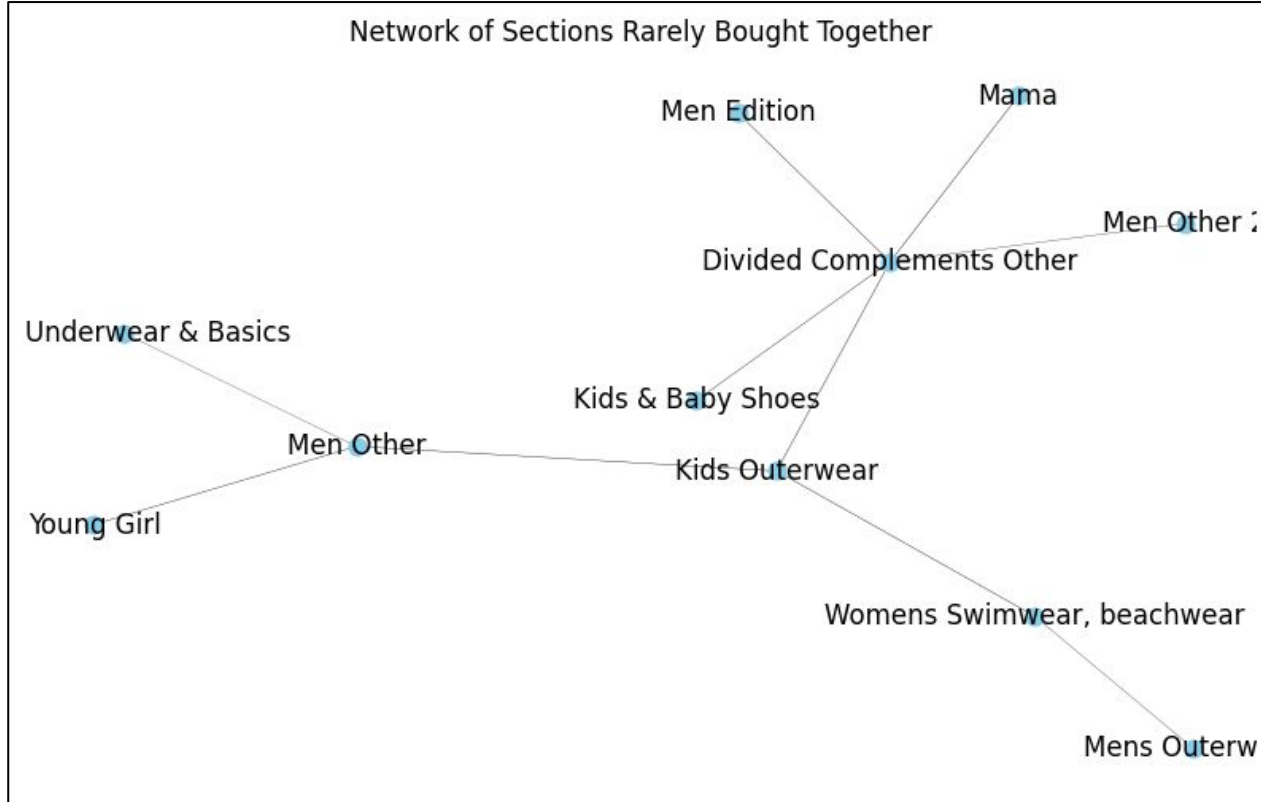
Strategy:

Items least frequently purchased together may be placed far away in the store.



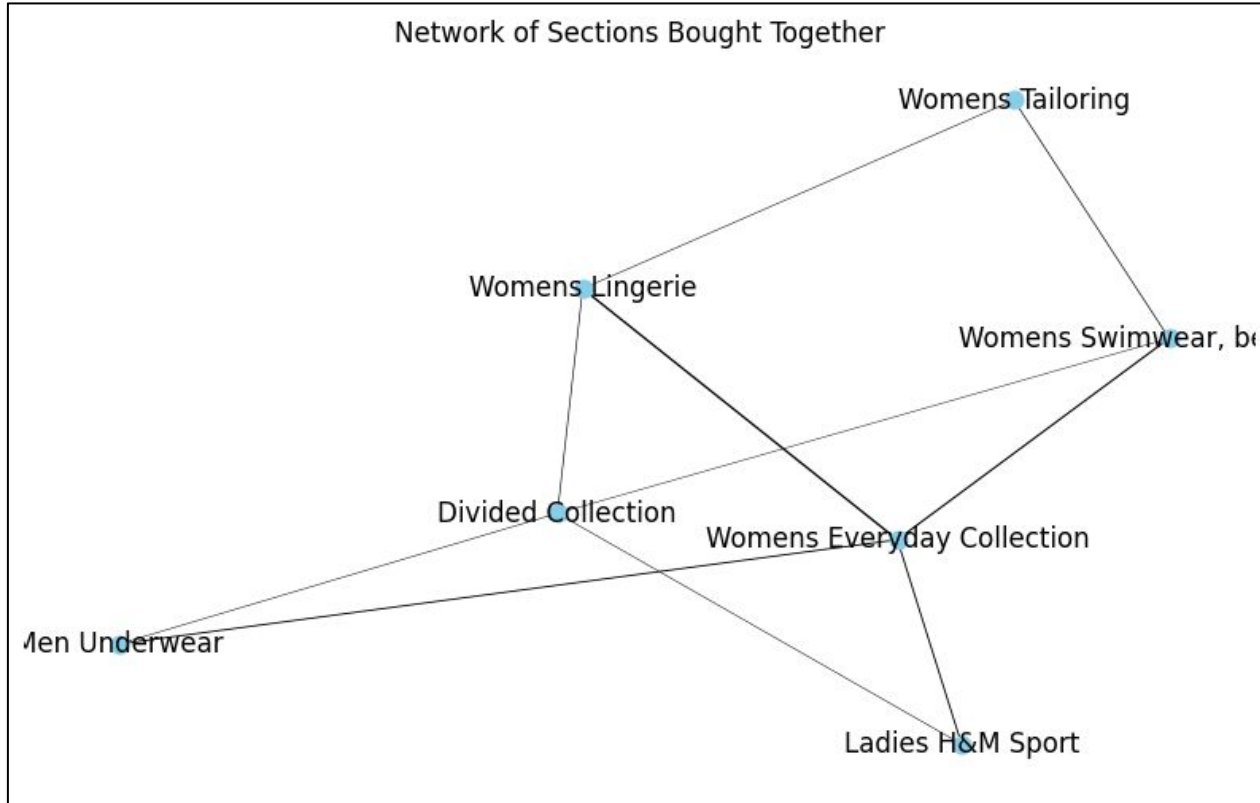


Least Frequent Pairings - Lift (Bottom 10)





Least Frequent Pairings - Leverage





Future Work

- (On Jupyter Report) Find which sections have highest lift with the most sold section, Womens Everyday Collection, and locate them together in stores. These are sections are undersold, so it may help to boost sales.
- Expand analysis to itemsets of 3 or more
- Use more specific definition of item (SKU)
 - Useful for promotions, recommendations, etc.

