What's In Your Cart?

Network Analytics with Grocery Data

Ali Khan, David Gong, Madeleine Cope, Spandan Pal, Teja Sirigina, Raghav Vaidya, Teagan Milford

TABLE OF CONTENTS

01 Introduction

02 Data

03 Cleaning

04 Visualization

05 Recommendation System

06 Conclusion

THE DATASET

KAGGLE

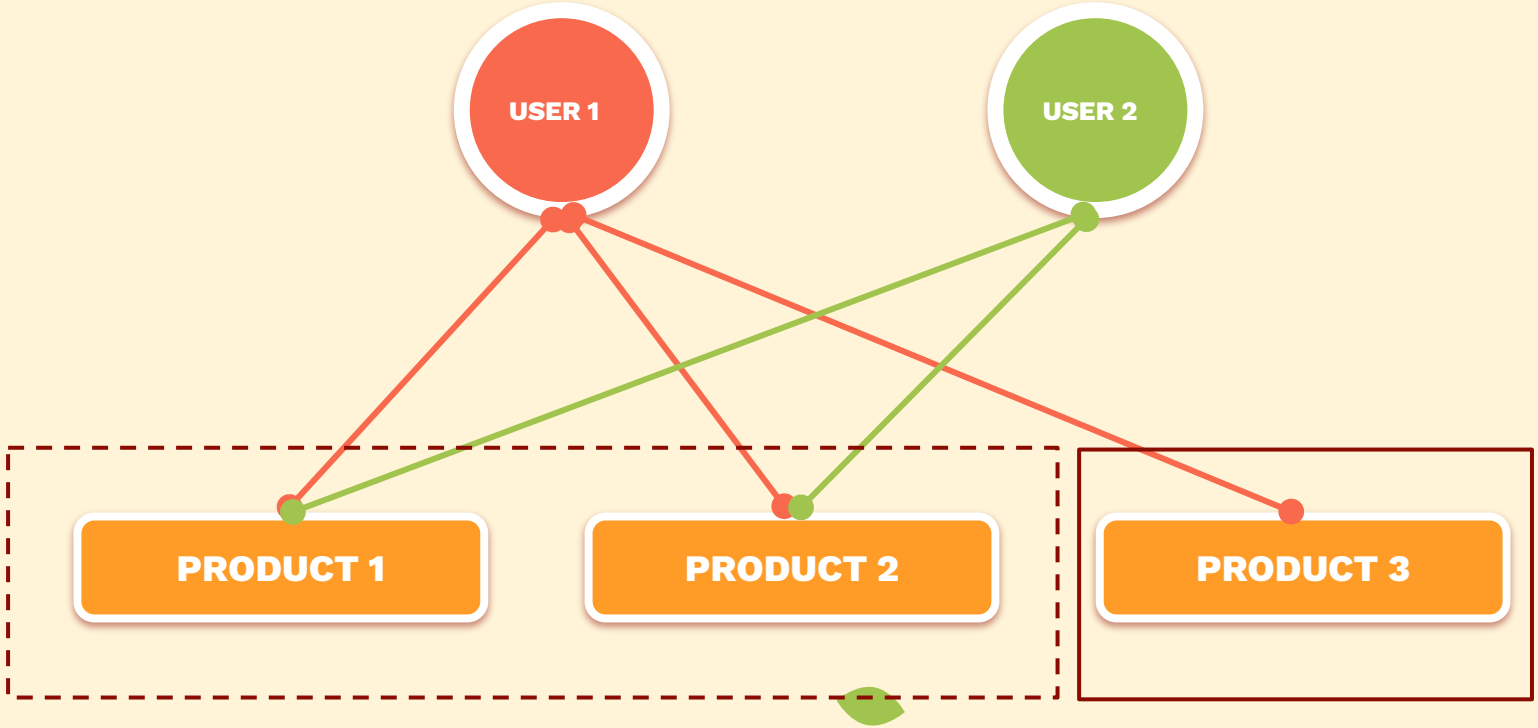
DATE

MEMBER ID

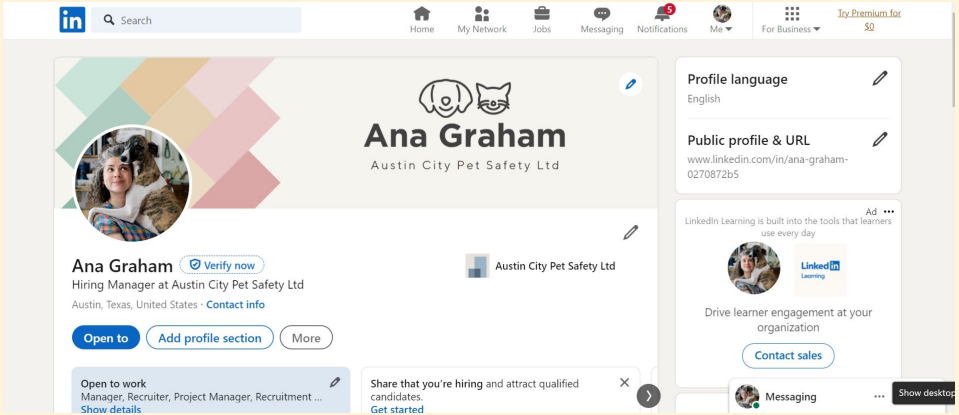
ITEM DESCRIPTION

	Member_number	Date	itemDescription
0	1808	21-07-2015	tropical fruit
1	2552	05-01-2015	whole milk
2	2300	19-09-2015	pip fruit
3	1187	12-12-2015	other vegetables
4	3037	01-02-2015	whole milk

THE GOAL



DATA CLEANING



LinkedIn
r/FemaleFashionAdvice
r/Handbags
r/Legal Advice

CATEGORIES

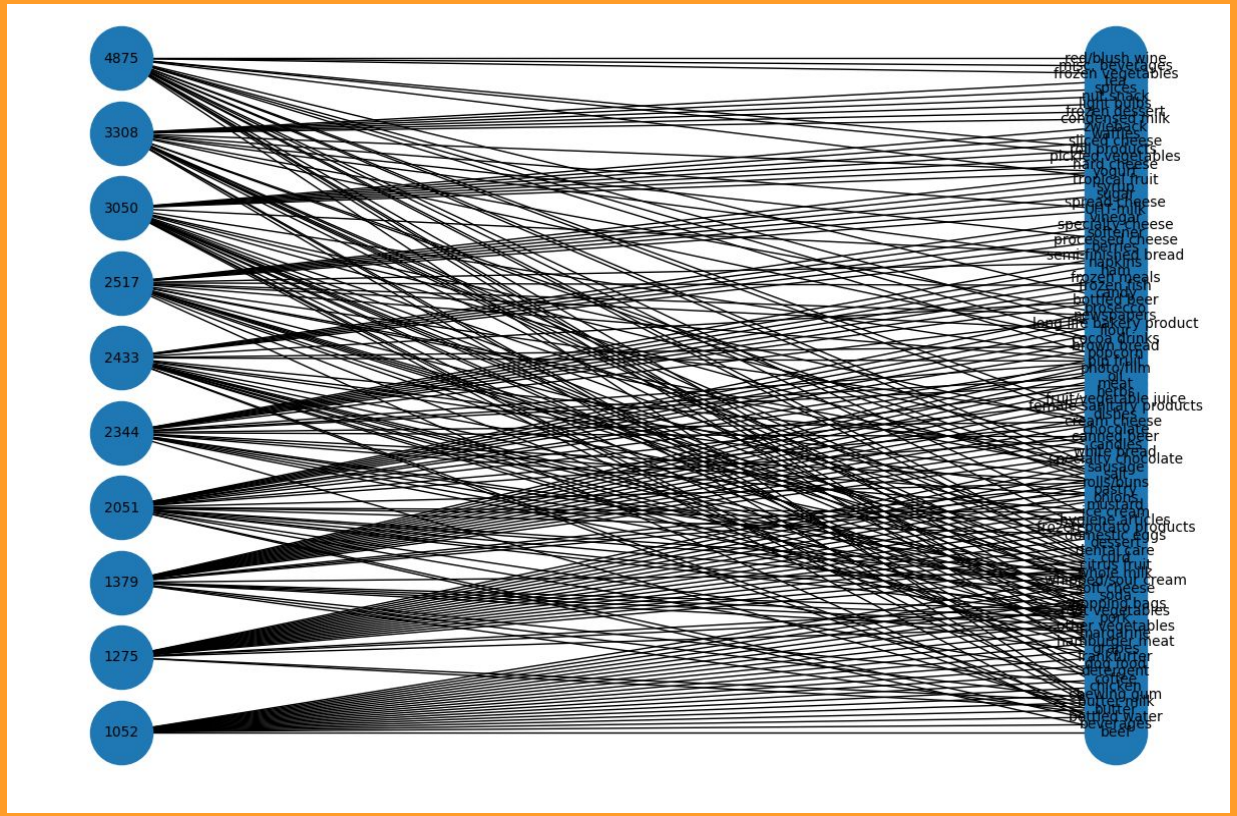
- FRESH PRODUCE
- BEVERAGES
- DAIRY & EGGS
- FROZEN & REFRIGERATED
- BAKERY & CEREALS
- HOUSEHOLD & PET
- MEATS & SEAFOODS
- MISCELLANEOUS

UPDATED DATASET

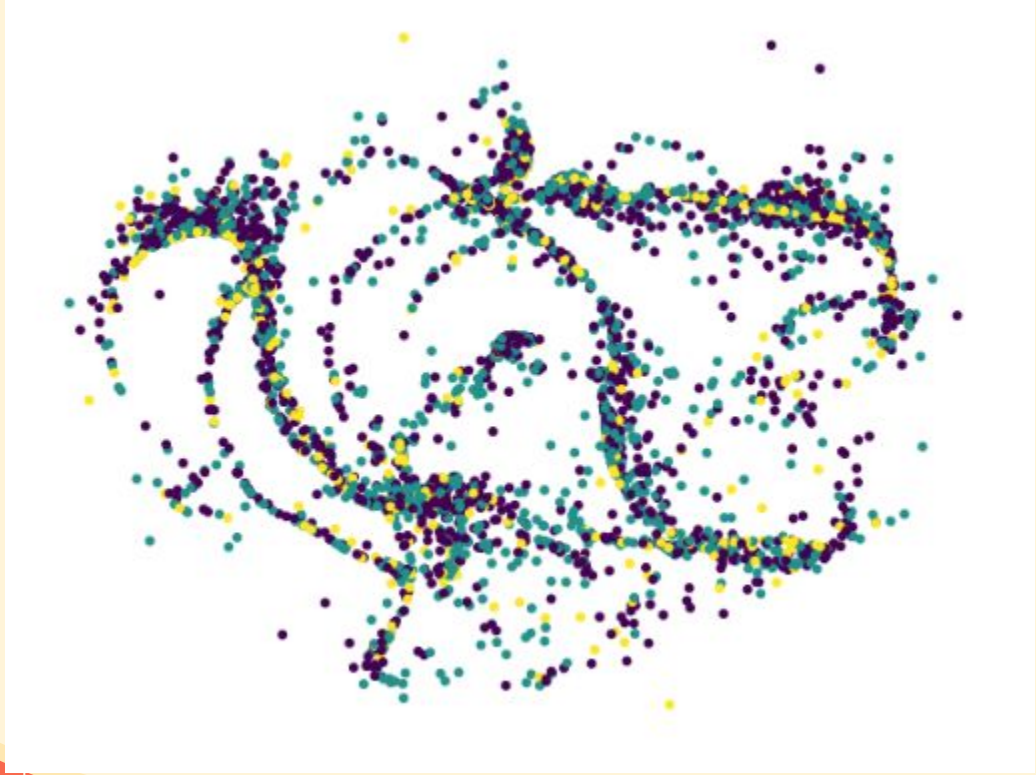
- Member ID
- ITEM
- COUNT
- ITEM CATEGORY

	member_id	item	count	item_category
0	1000	canned beer	1	Beverages
1	1000	hygiene articles	1	Household and Pet
2	1000	misc. beverages	1	Beverages
3	1000	pastry	1	Bakery and Cereals
4	1000	pickled vegetables	1	Miscellaneous

BIPARTIDE NETWORK OF TOP 10 USERS & PURCHASE FREQUENCY



COMMUNITY CLUSTERING

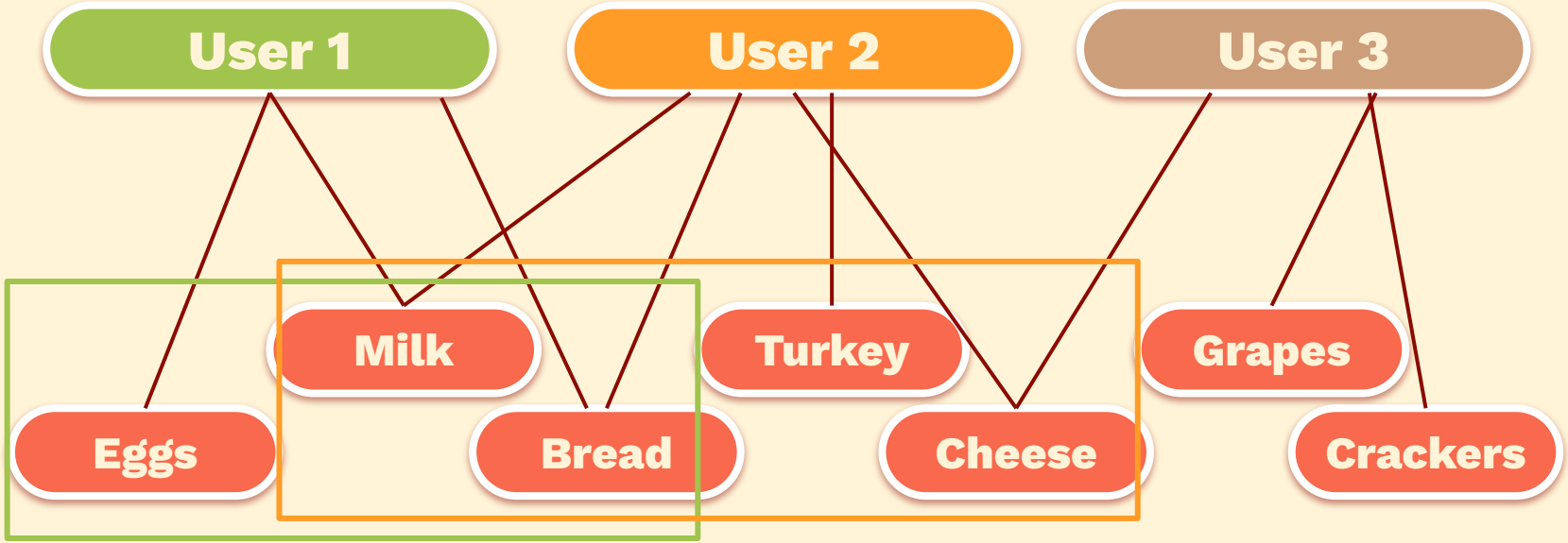


There are no strong clusters

Overall, we cannot confidently interpret the clusters

This led to us approaching the data from another angle

Recommender System



How to find similar users?

Maximizing Intersection Size

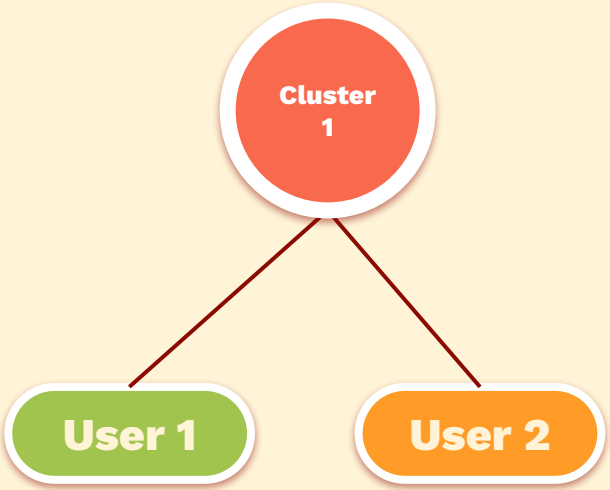
Hit Rate: 22%

Cosine Similarity

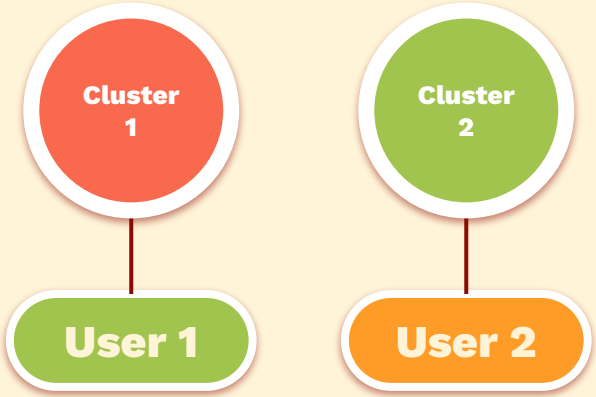
Hit Rate: 24%

Clustering of similar users

50% within cluster



50% out of cluster



Future Uses

Grocery online storefronts

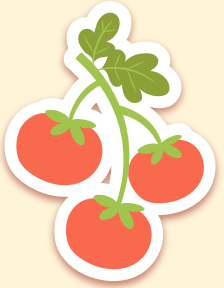
Can offer recommendations of items not typically bought by the user

Trials and Samples

Companies can use this logic to offer samples to customers that they actually may want to buy later

E-commerce across industries

Because the e-commerce area is growing, this could be a way for many industries to have a competitive advantage



Thank You!