

A photograph of a white wire mesh basket with two wooden straps and two leather straps with silver buckles. The basket is resting on a light-colored wooden surface. In the background, there are shelves and other items, suggesting a workshop or studio environment.

Market Basket Analysis

and

Recommendation System

Apex Pattern Deployers



Today's Discussion

→ The Problem

→ The Process

→ The Solution

→ The Insights

The Problem: Business Understanding

Gift-a-Lot

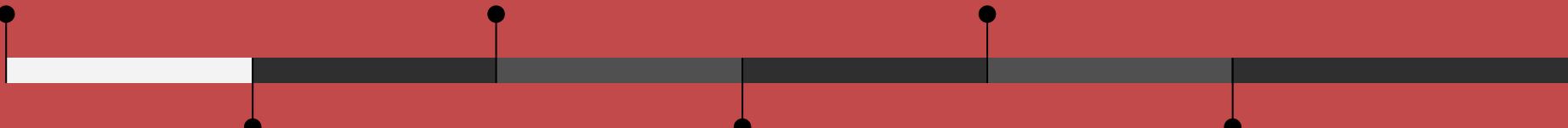
From direct mailers to online

Increased online shopping

Customer Recommendation System



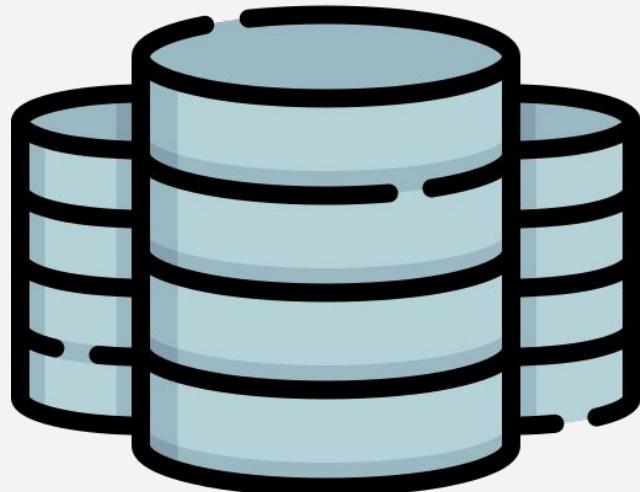
The Process: Data Understanding



A decorative horizontal bar at the bottom of the slide consists of a thick dark grey line with small black circular dots positioned along its length. A single white rectangular segment is placed on the left side of the dark grey line.

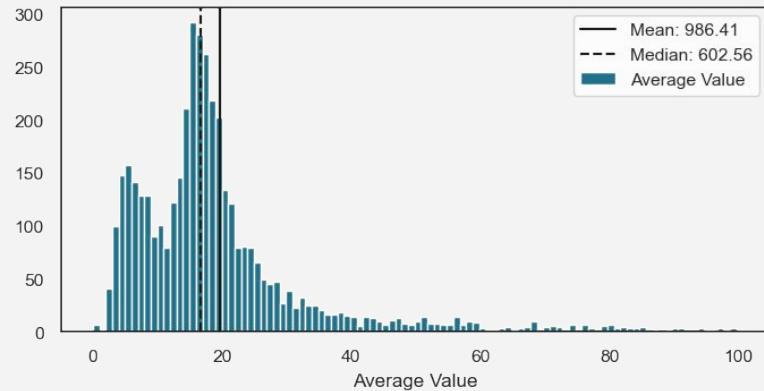
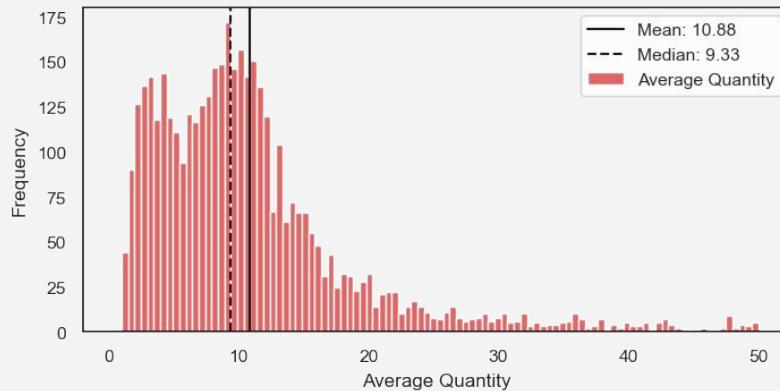
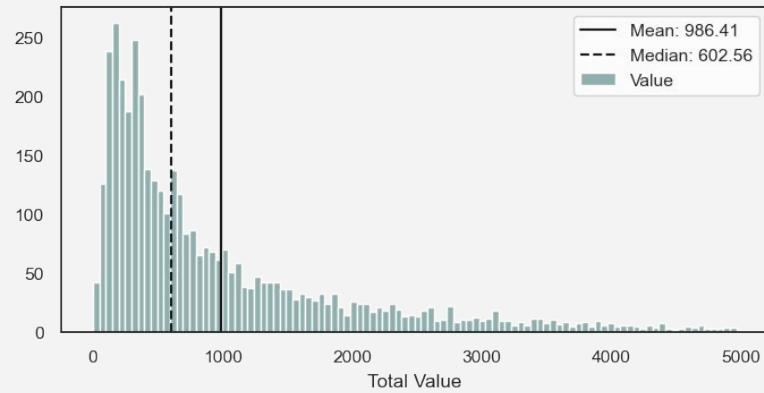
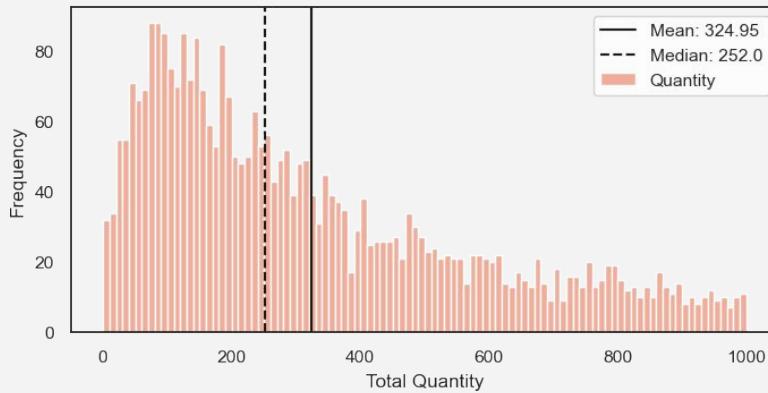
Initial Transactional Data

- 25 900 Transactions associated with unique items;
- 541 909 instances for each item;
- 4 372 different customers;



Customer Purchases

Distribution of Customer Purchases



Additional Information that we found



Unique Invoices



Unique Invoices with
NULL Customer IDs



Unique Cancelled
Invoices



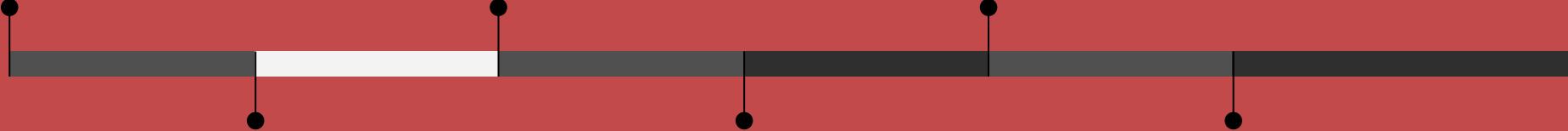
Unique Customer IDs

19790 Invoices x 3995 Products

Sparsity



The Process: Data Preparation

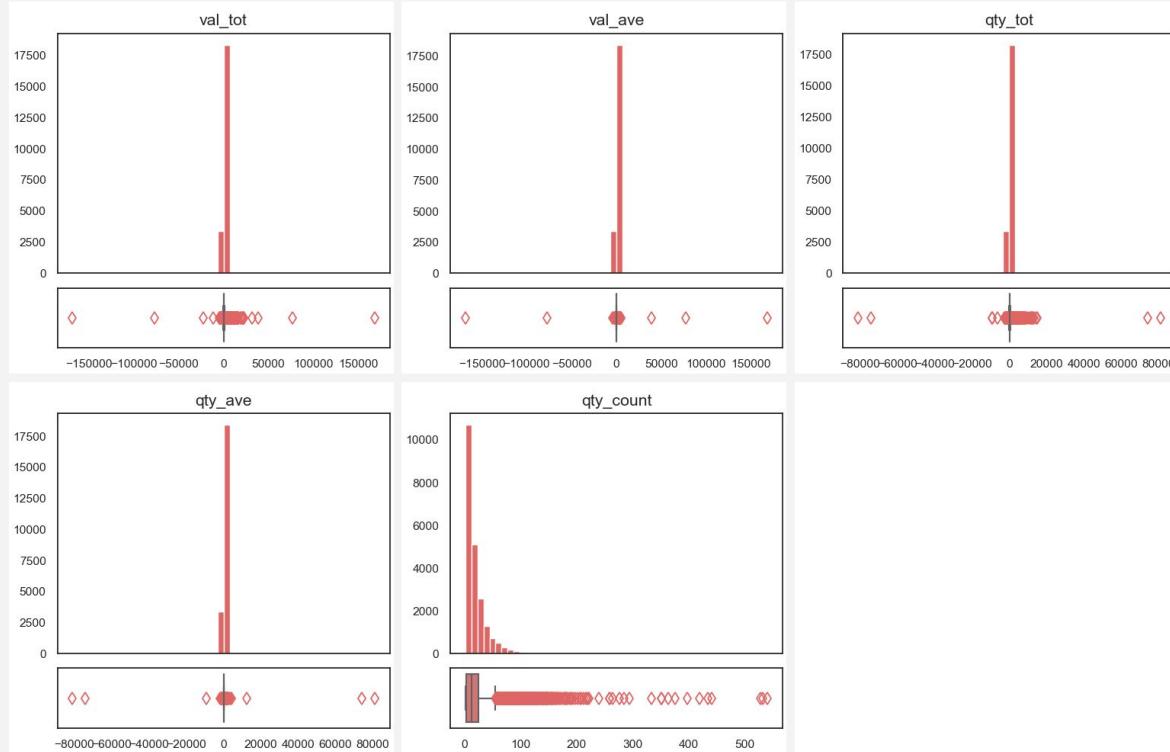
A decorative horizontal bar at the bottom of the slide consists of a thick dark grey line with several small black circular dots of varying sizes attached to it by thin black lines, creating a pattern reminiscent of a barcode or a stylized representation of data points.

Data cleaning

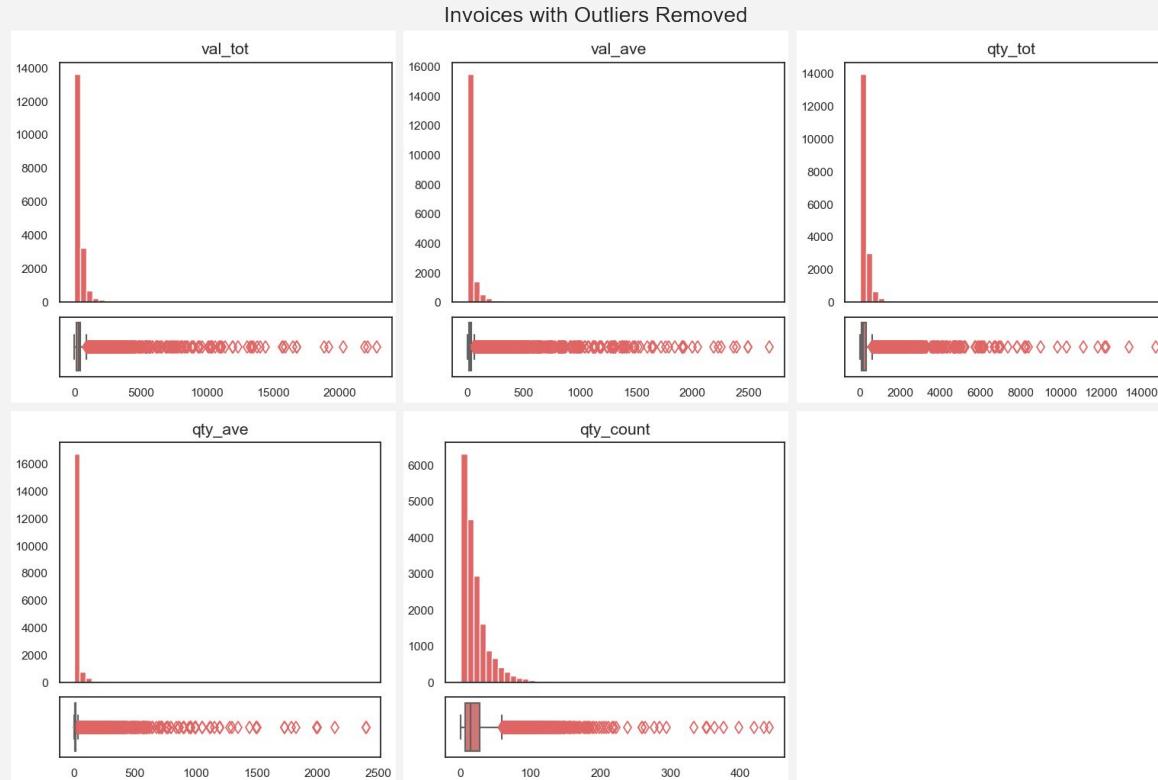
- Filtered out 9 291 invoices that contained adjustments or cancellations;
- Removed 3 641 rows with non-product descriptions;
- Removed 1 454 rows with outliers just for the clustering;
- Created a new variable Value (Quantity * Unit Price);

Invoices

Distribution of Invoice Information



Invoices with outliers removed



Clean up product descriptions

	Unique product descriptions	"set of 4 rose botanical candles"	4224
215	Removed non-products	Postage, Damage, Error	4009
191	Removed non words + symbols	→ rose botanical candles	3818
11	Used lemmatizer	→ rose botanical candle	3807
476	Removed colors	→ botanical candle	3331
173	Removed adjectives	→ candle	3158

Product variants

3 158

487

with variants

“drop earring”

purple crystal drop earrings
black crystal drop earrings
amber crystal drop earrings
green crystal drop earrings

2 671

without variants

“paisley cake cases”

pack of 60 pink paisley cake cases

Sparsity

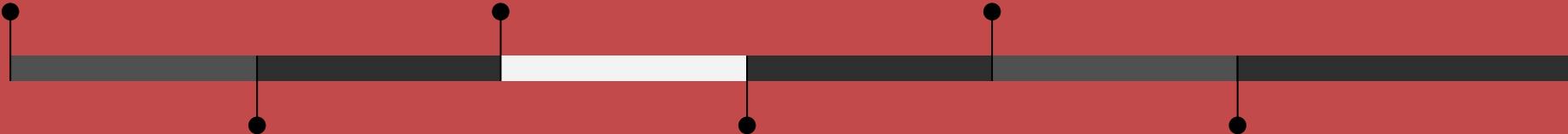
19790 Invoices x 3146 Products



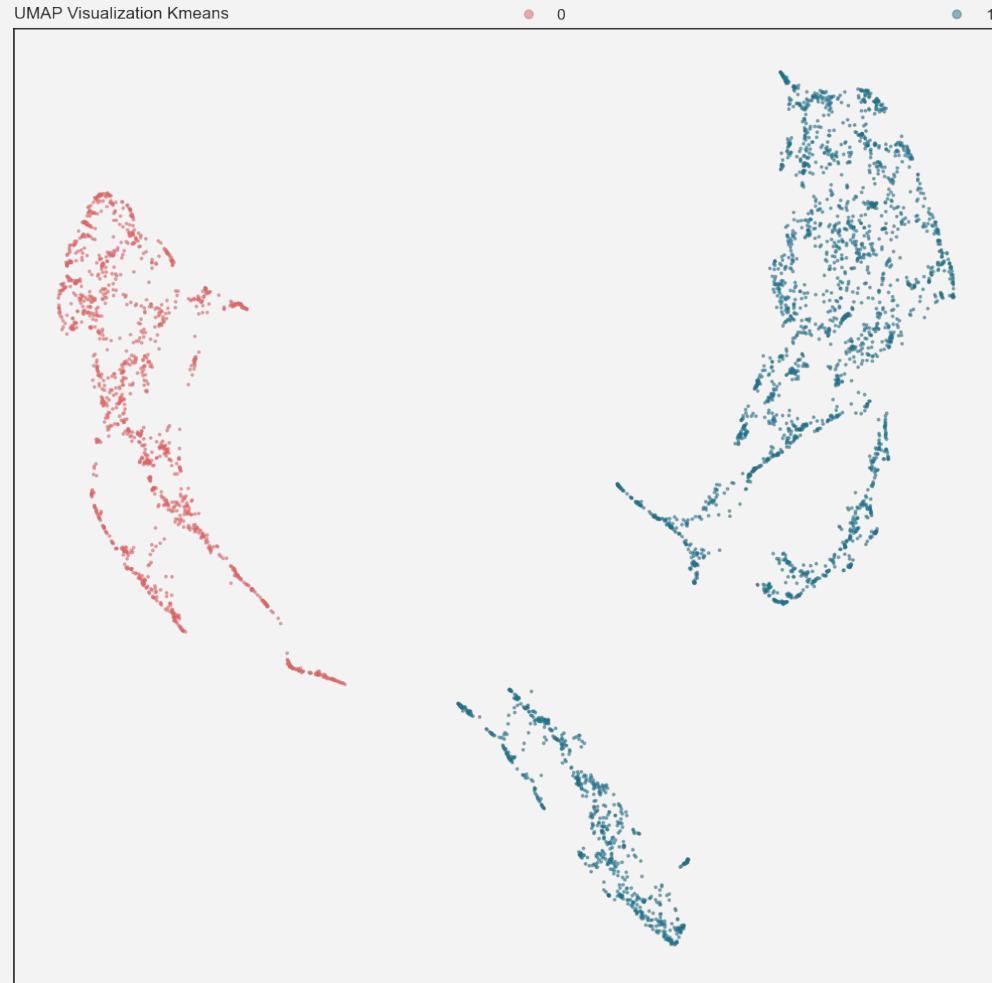
Sparsity (zoomed in)



The Solution: Market Basket Analysis



Cluster the Customers

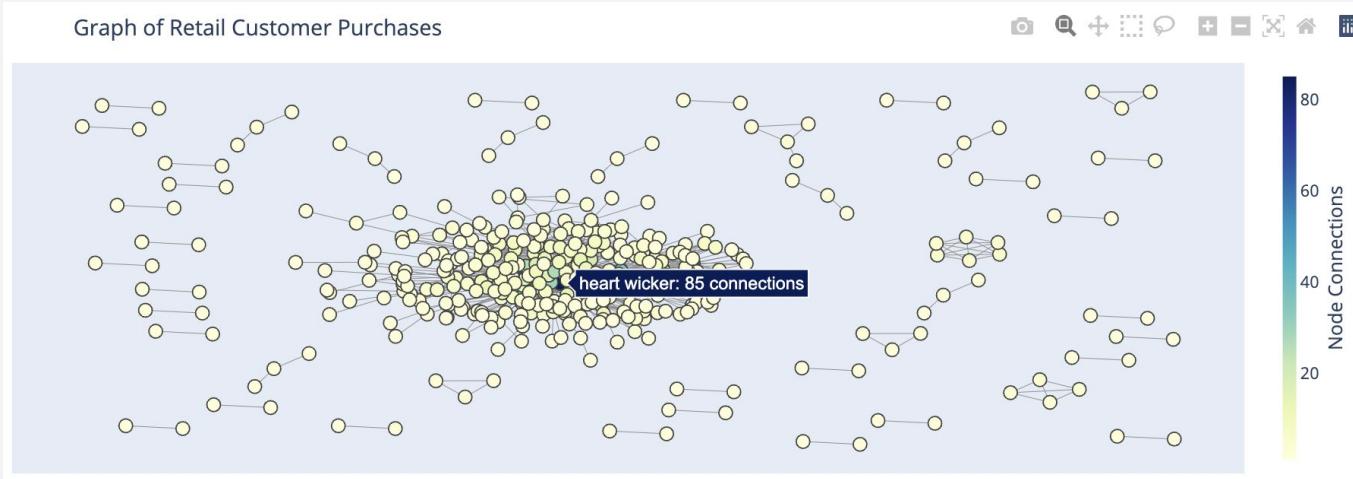


Association Rules

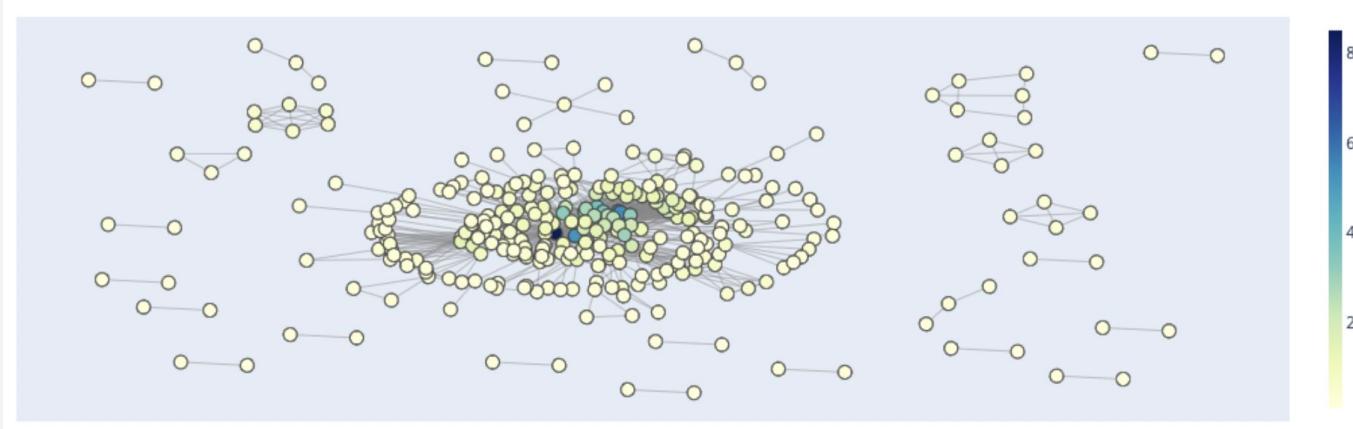


- Minimum Support: 10%
- Wholesalers: 672 rules
- Retail: 278 rules
- Few substitutes, lots of complementary products

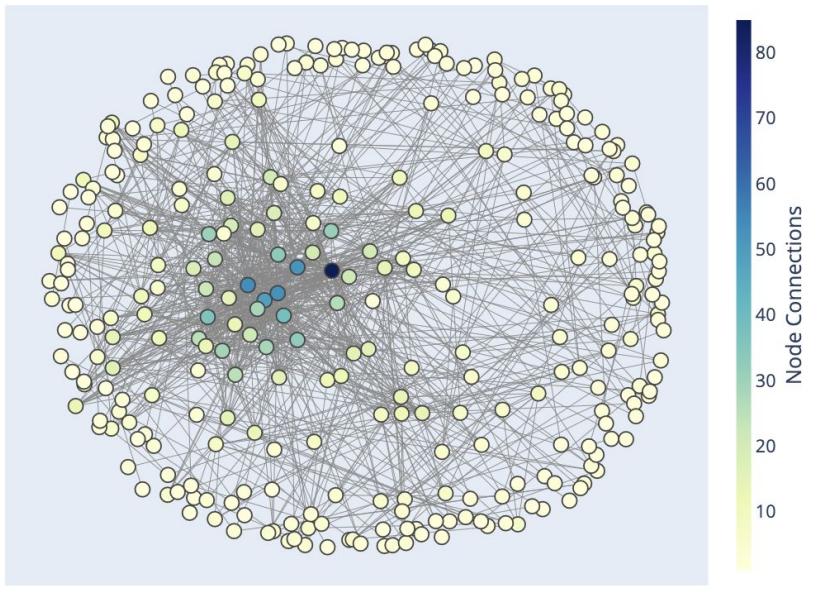
Graph of Retail Customer Purchases



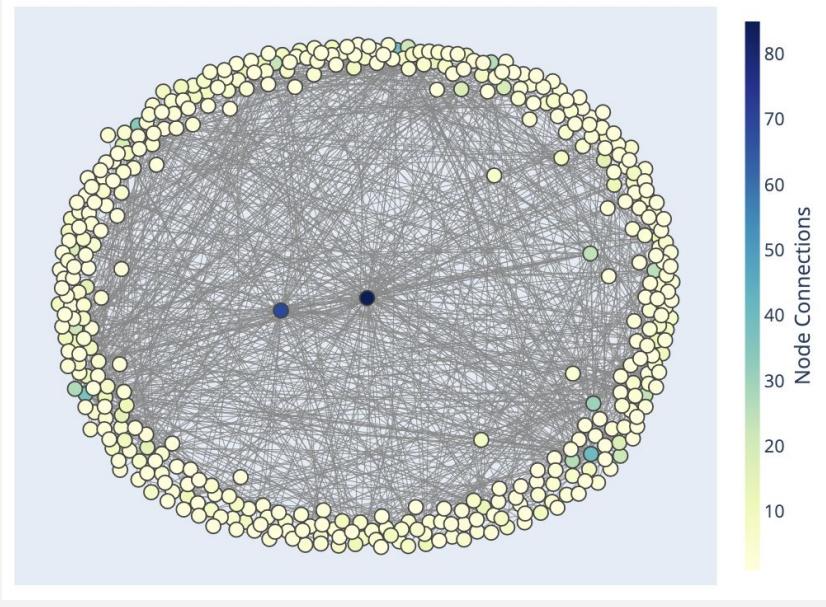
Graph of Wholesale Customer Purchases



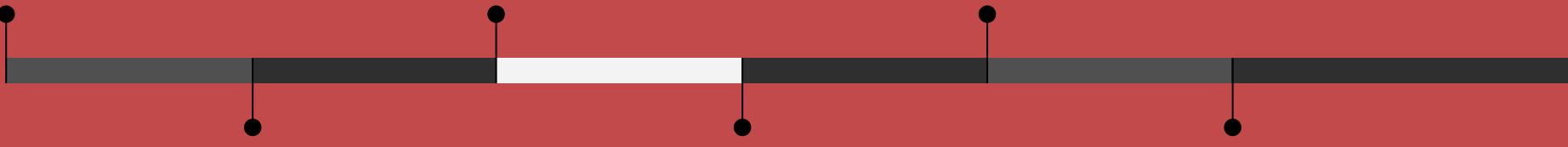
Graph of Most Frequent Wholesale Customer Purchases



Graph of Most Frequent Retail Customer Purchases



The Solution: Recommender System

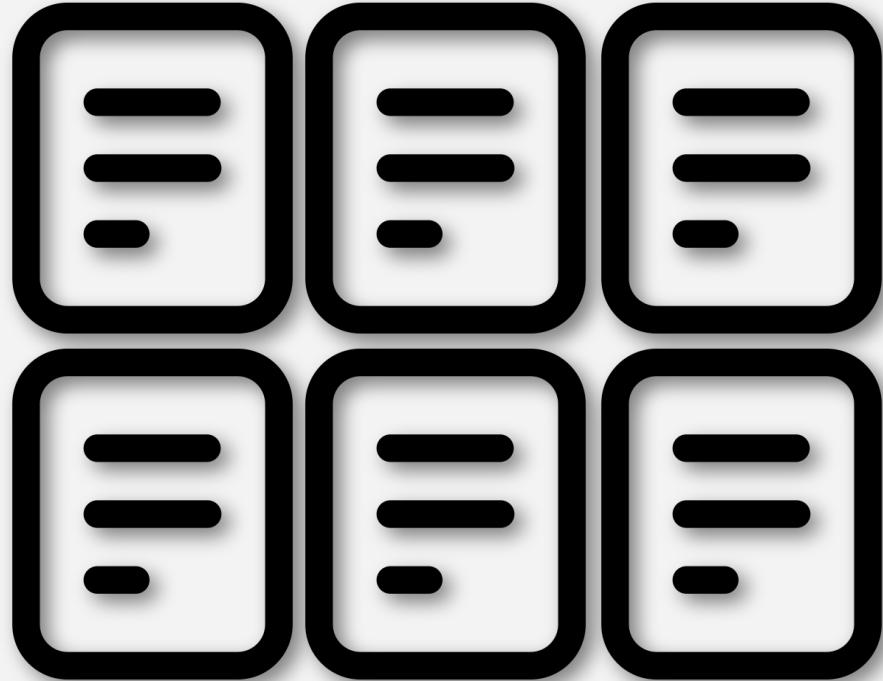


A decorative horizontal bar at the bottom of the slide consists of a thick dark grey line with small black circular dots attached by thin vertical lines. There are ten dots in total, with two on the left, one in the center, and seven on the right.

TF-IDF



Frequency within invoice



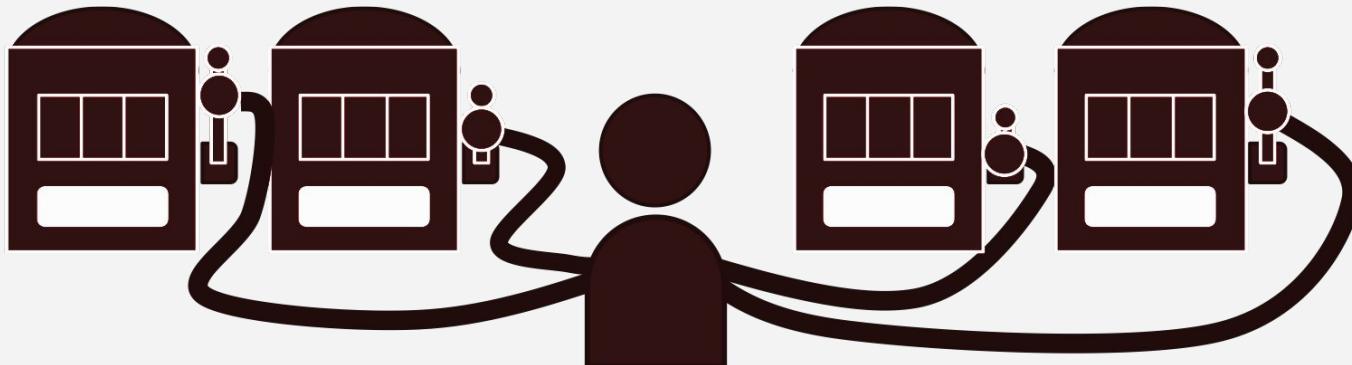
Frequency across all invoices

Cosine Similarity Matrix

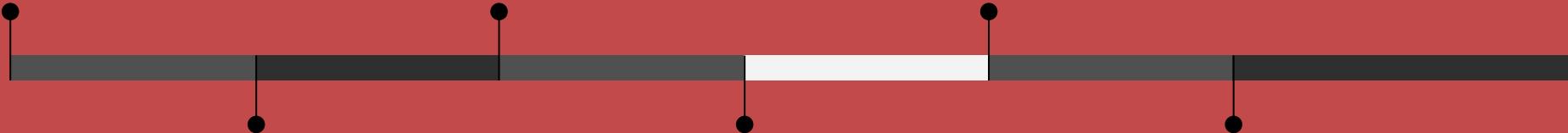


Cold Start

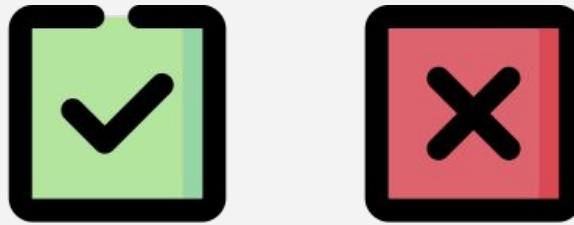
- Overstocked items
- Popular items from the current week
- Random items
- Anchor items
- For future: multi-armed bandit algorithm



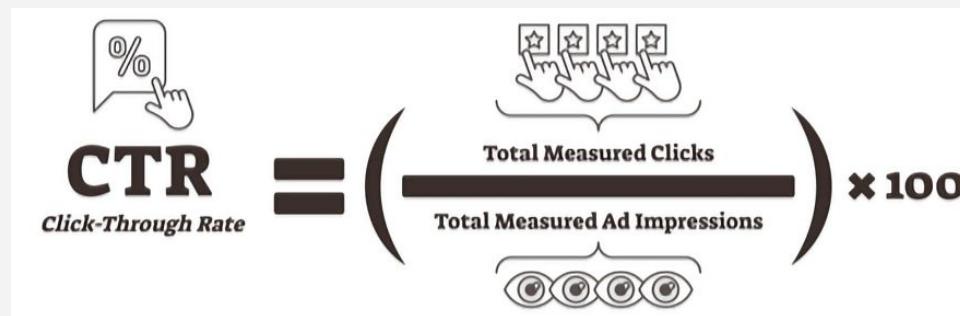
The Solution: Evaluation



Accuracy & Forecasting

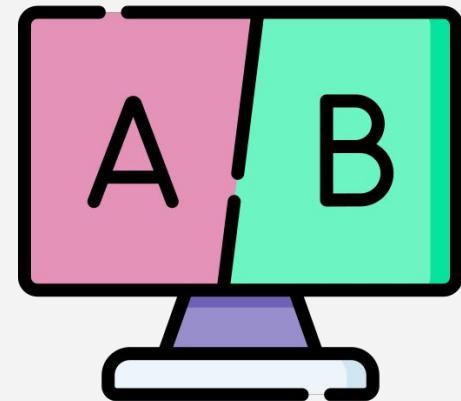


Utility Metrics & Value Test

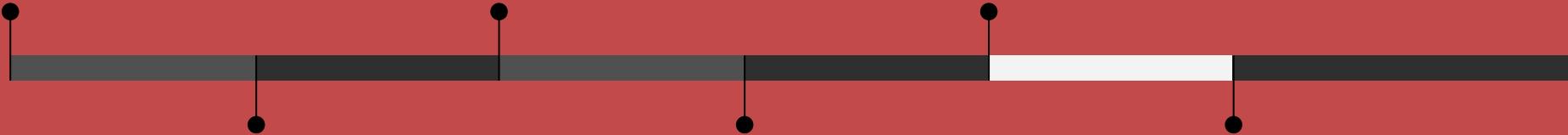


A diagram illustrating the Click-Through Rate (CTR) formula. It features a hand icon pointing at a percentage sign (%) inside a speech bubble icon. Below this is the bold text "CTR". Underneath "CTR" is the italicized text "Click-Through Rate". To the right of "CTR" is an equals sign (=). To the right of the equals sign is a large parentheses containing a fraction. The numerator of the fraction is a hand icon pointing at a star icon, with the text "Total Measured Clicks" written below it. The denominator of the fraction is a hand icon pointing at an eye icon, with the text "Total Measured Ad Impressions" written below it. A multiplication sign (×) followed by the number 100 is positioned to the right of the fraction.

$$\text{CTR} = \left(\frac{\text{Total Measured Clicks}}{\text{Total Measured Ad Impressions}} \right) \times 100$$



The Solution: Deployment



Proof of Concept: Recommender

```
new_items = ['lunch bag retrospot', 'towel design']
similar_baskets = get_similar_baskets(new_items, df_inv, )
```

```
similar_baskets
```

	InvoiceNo	Description	CustomerID
0	538049	glass dressing table pot,towel design,towel de...	12748
1	552456	retrospot oven glove,party bunting,child garde...	13999
2	581583	lunch bag retrospot,love heart tlights	13777
3	565154	lunch bag design,lunch bag skull,lunch bag ret...	13470
4	562466	towel design,mirrored disco ball,towel design,...	13435

```
get_product_recs(similar_baskets, new_items)
```

```
['polkadot bowl',
 'retrospot cup',
 'wood garden plant ladder',
 'love heart tlights',
 'lunch bag design',
 'lunch bag skull',
 'feltcraft shoulder bag',
 'vintage paisley',
 'lunch bag vintage design',
 'polkadot cup',
 'flag',
 'spring paper cup',
```

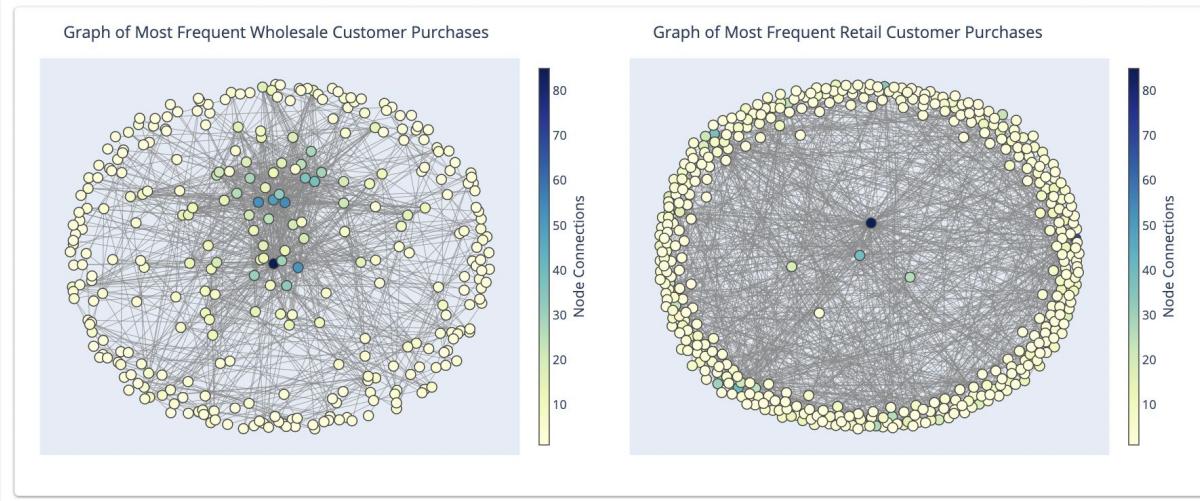
Proof of Concept: Recommender + Network Explorer

Select Products

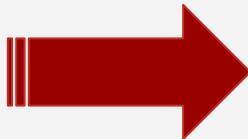
x alarm clock bakelite x heart coat hanger x babushka nesting box x
x bird ornament

Other Customers Also Bought:

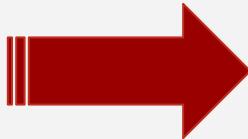
vintage billboard drink mug feather penhot
retrospot pudding bowl retrospot teapot
vintage billboard lovehate mug retro mug
decoration hen nest hanging heart trellis



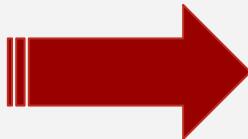
Challenges to Deployment



The company website and servers must be able to handle the increased processing, since the recommender system must give results in real time;



Deciding on how many products to recommend, as research has shown that the right number of recommended items depends on both product category and customer type;



New invoices will need to be fed back into the algorithm at regular intervals;

Future Recommendations

General Recommendations

Collaborative Filtering Algorithm instead of only content-based filtering;

Adding product categories and keeping product descriptions free of extraneous information

Improving the cold start problem by adding a multi-armed bandit algorithm;

Database Recommendations

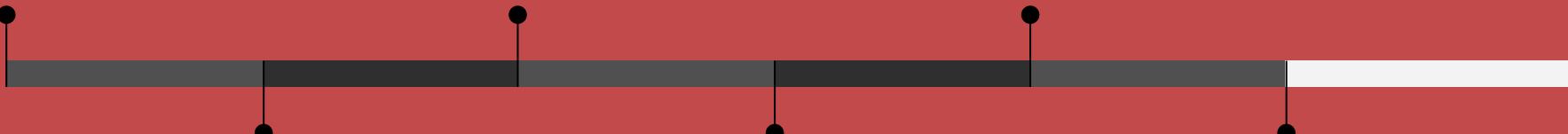
Keep Data Clean;

Add product categories to the database;

Track retail/wholesale as well as customer demographic data (e.g. google analytics);



Insights



A decorative horizontal bar at the bottom of the slide consists of a thick dark grey line with small black circular dots positioned along its length, some near the top edge and others near the bottom edge.

Retail Customers

```
{ bird ornament } => { hanging heart tlight holder }
{ hanging heart tlight holder } => { bird ornament }
{ bird ornament } => { heart wicker }
{ heart wicker } => { bird ornament }
{ hanging heart tlight holder } => { party bunting }
{ party bunting } => { hanging heart tlight holder }
{ hanging heart tlight holder } => { rex cashcarry shopper }
{ rex cashcarry shopper } => { hanging heart tlight holder }
{ heart water bottle } => { hanging heart tlight holder }
{ hanging heart tlight holder } => { heart water bottle }
```

Wholesale Customers

```
{ bag } => { hanging heart tlight holder }
{ hanging heart tlight holder } => { bag }
{ hanging heart tlight holder } => { regency cakestand tier }
{ regency cakestand tier } => { hanging heart tlight holder }
{ bag retrospot } => { hanging heart tlight holder }
{ hanging heart tlight holder } => { bag retrospot }
{ hanging heart tlight holder } => { lunch bag retrospot }
{ lunch bag retrospot } => { hanging heart tlight holder }
{ water bottle } => { hanging heart tlight holder }
{ hanging heart tlight holder } => { water bottle }
```

Thank You!

Apex Pattern Deployers