Wolt's Takehome Dataset Analysis

by Artur Roos

The Dataset

Why this dataset?

1. Real™ data
2. Complex structure
3. Missing data
4. Averaged everything

What is it like?

30 columns
21983 rows

From 2019 to 2021

Registration Country
Amounts of different purchases
First/last purchase dates
Most popular hour/weekday
Device/Platform
Average delivery distance

More than 50% of most columns is missing.

What processing was done?

Only Finland.
Only people who purchased.

For each metric only columns with complete data were used.

No imputations.

Total ~5000 rows.

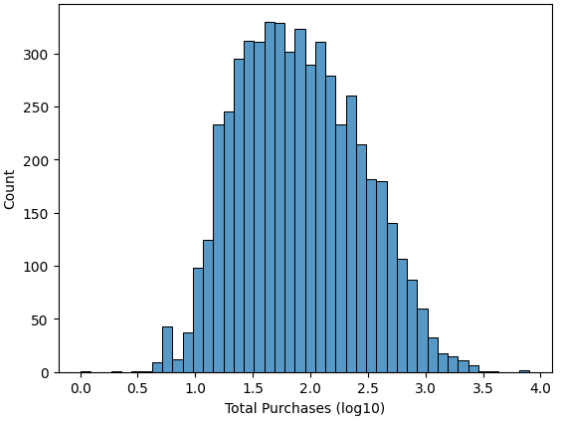
Distribution Shapes

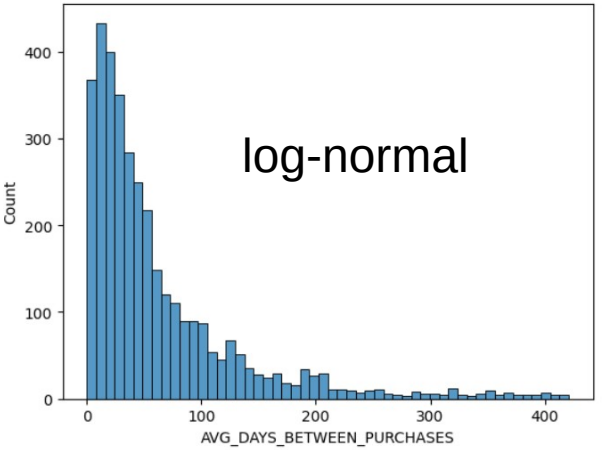
How is the data?

All data is either uniform or with a strong tail, standard deviation mostly not applicable, but averages are very valuable.

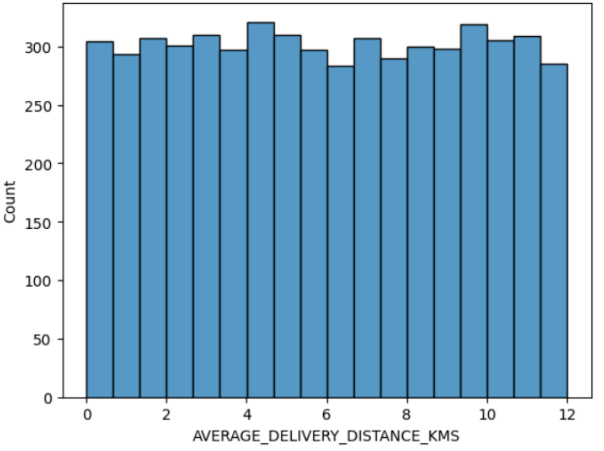
What is the average customer like?

1. ~9 purchases totalling 260€.
2. Each purchase is ~30€.
3. Active for 8 months.
4. Orders takeaway every ~2 months.
5. Has tried around 4 places.





uniform



Recency, Frequency, Monetary

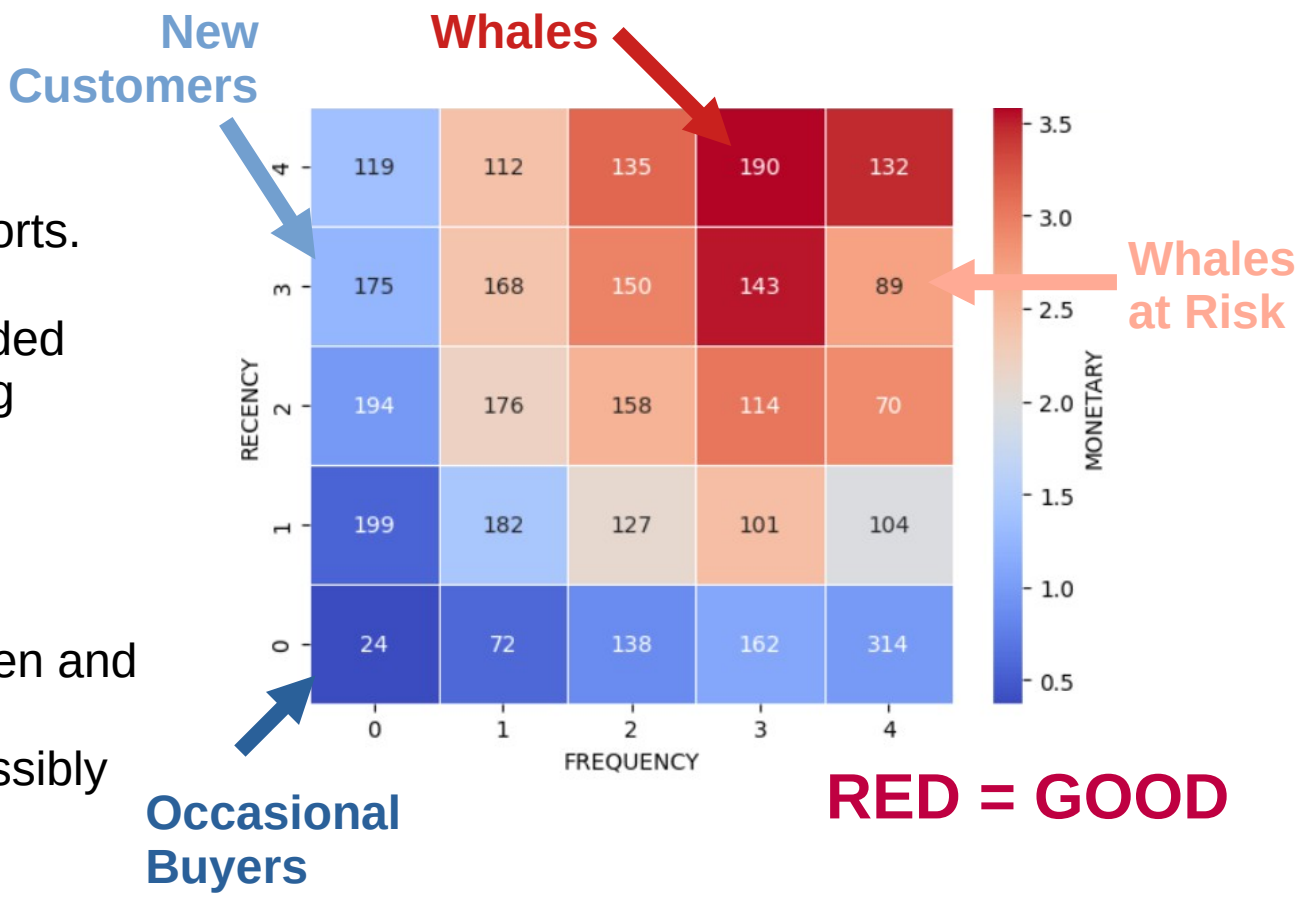
What is RFM?

Customer's activity is scored on 3 axes and split into 4 cohorts.

Customer data over time needed to estimate possible marketing strategies for each group.

What does this mean?

1. Bulk of the revenue buy often and bought recently.
2. Frequency is important, possibly because of habit formation.



Relationships

What correlations exist?

1. Purchase count is basically equivalent to delivery count, most order take out.

2. Having a lot of purchases also correlates with having a lot of lunch dinner purchases, supporting the earlier conclusions.

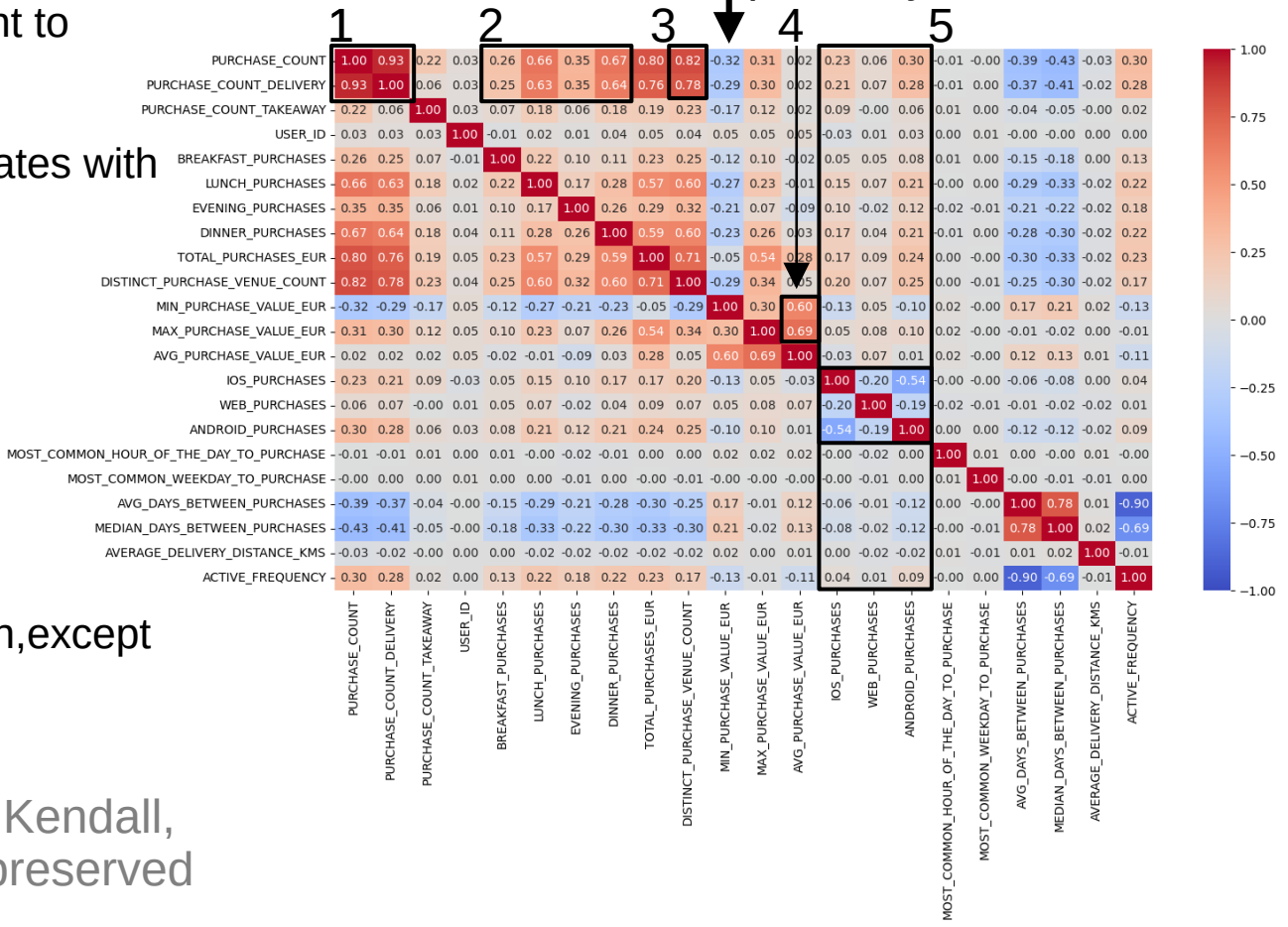
3. More purchases = More venues

4. Averages correlate with minimum and maximum, so customers have a comfortable spending range

5. Platforms don't give much information,except generally excluding each other

NB: Correlation is calculated using Kendall, so all monotonic relationships are preserved

Doesn't make logical sense, so probably a fluke



Temporal Patterns

What are the patterns?

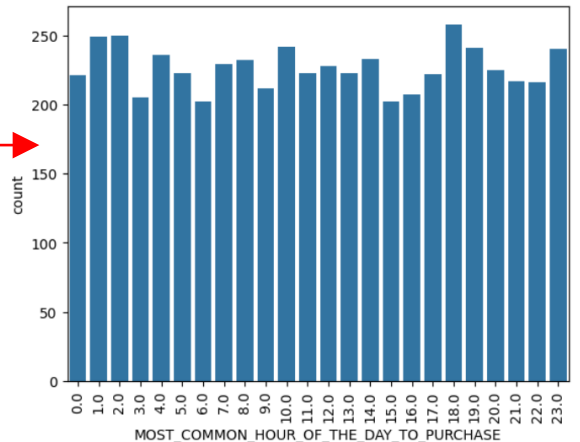
Strong preference towards lunch and dinner order. 0 orders late at night.

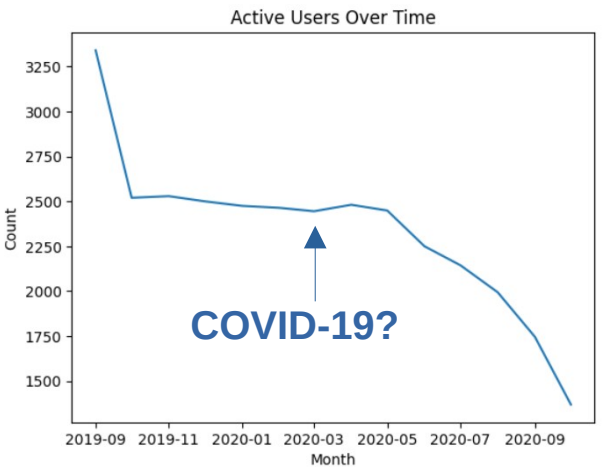
The user count is continuously falling at an accelerating rate.

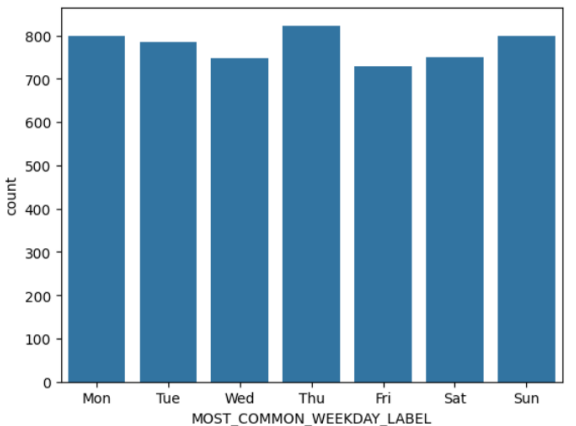
Most preferred hour of purchase is inconsistent across two different columns

Most common weekday looks similar, so it is also invalidated.

Inconsistent







Conclusions

What are the key takeaways?

1. Frequency (habit formation) is very important for generating revenue.

2. Customers have a comfortable spending range the bounds of which are proportional to their average spending.

3. The amount of active users is declining, measures to increase conversion are needed.

4. Optimize for load during lunch and dinner times.

5. Different platforms are in use, all need to be supported.

References

Where were the images taken from?
1. https://explore.wolt.com/en/fin/about
2. https://wolt.com/en/fin/ruka/article/helsinki-welcome-to-wolt

Where is the code for this?
https://www.github.com/ktnlvr/wolt-takehome