Increasing Customer Loyalty with LightFMs Recommendation Algorithm

Project by Matthew Merrill

Project Goal and Results

Data from Olist was acquired from 2016 to 2018 of sales activity with customer, seller and product info for analysis.

After parsing the data, it was found that only 4% of customers in that time span purchased from Olist on more than one occasion.

Possible solution:

Recommend appropriate products to customers to increase loyal users and subsequent sales[1]**.**

Project Goal and Results

Using LightFMs recommendation algorithm we were able to achieve a 97% AUROC* for customers.

This was achieved using a hybrid recommendation system that uses collaborative and content based approaches to suggest items, coupled with clustering techniques to introduce customer features into the model.

Significant sales increases could result from implementation of this model.

*AUROC was determined by comparing items predicted with items previously purchased. The high score suggests sharpness in predicting known likes (TP), which bodes well for supplemental suggestions.

About Olist

- Considered the "Marketplace of marketplaces" due to the peculiarity of the eCommerce market [2].
- Mission is to bring together small merchants across Brazil to connect with an already existing but fragmented customer base [2].
- Founded in 2016, 95% of customer market originates from Brazil as of 2020 [3].
- Provides full stack service to merchants, from managing inventory to payments[2].

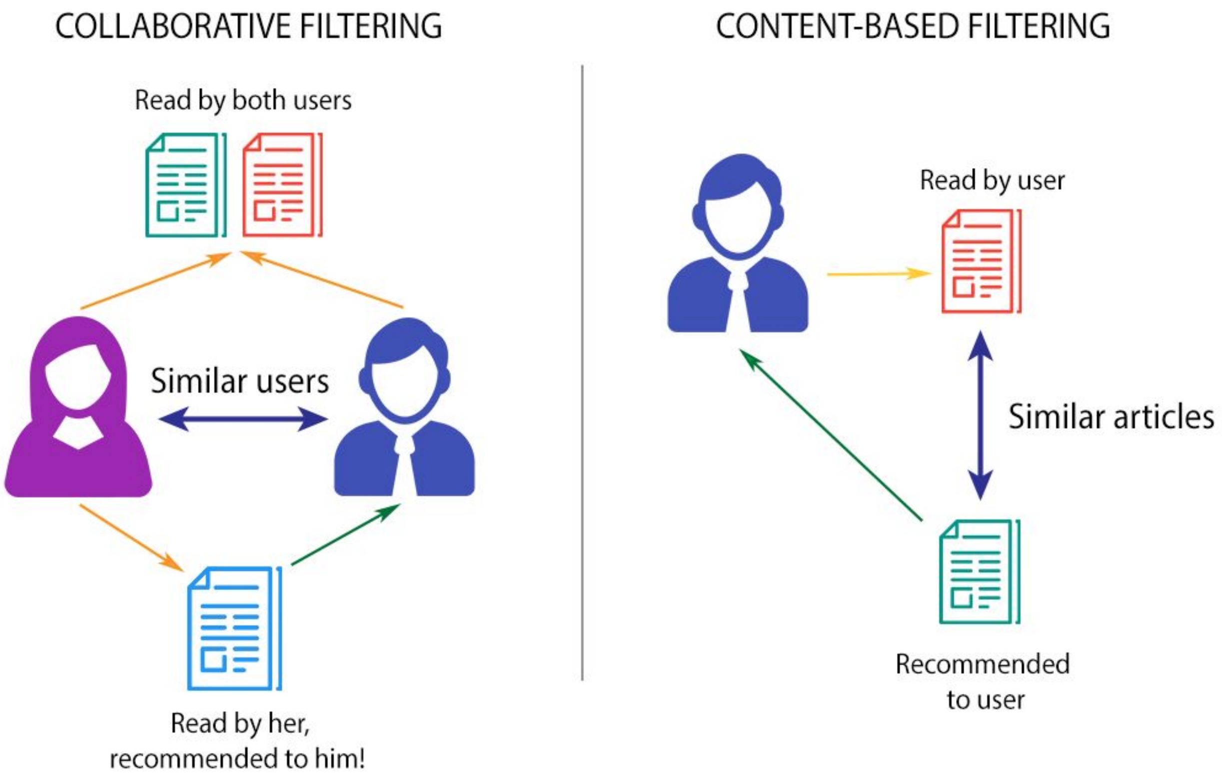
olist
empowering commerce

About Recommender Systems: Collaborative vs Content

- Collaborative recommendations:
  - Algorithm infers possible suggestions by observing similar behavior among users.
  - The strength of a prediction can be used to make the most relevant suggestion.

- Content recommendations:
  - Product details of previous purchases are used to recommend items with similar features.
  - For music streaming services, if song is of similar genre, artist or tempo to one previously listened to, a user might benefit from the suggestion.

About Recommender Systems: Collaborative vs Content

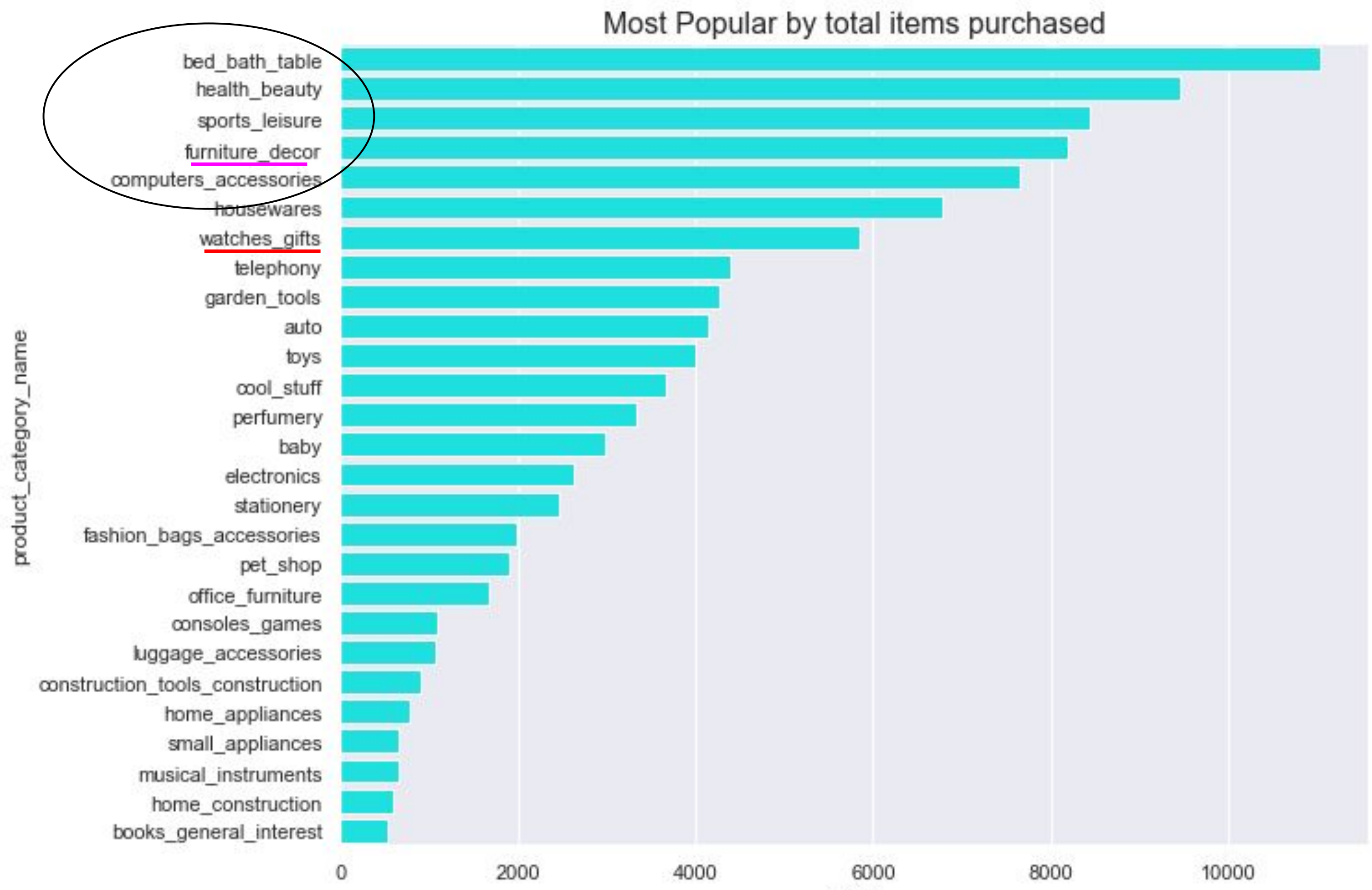


About LightFM

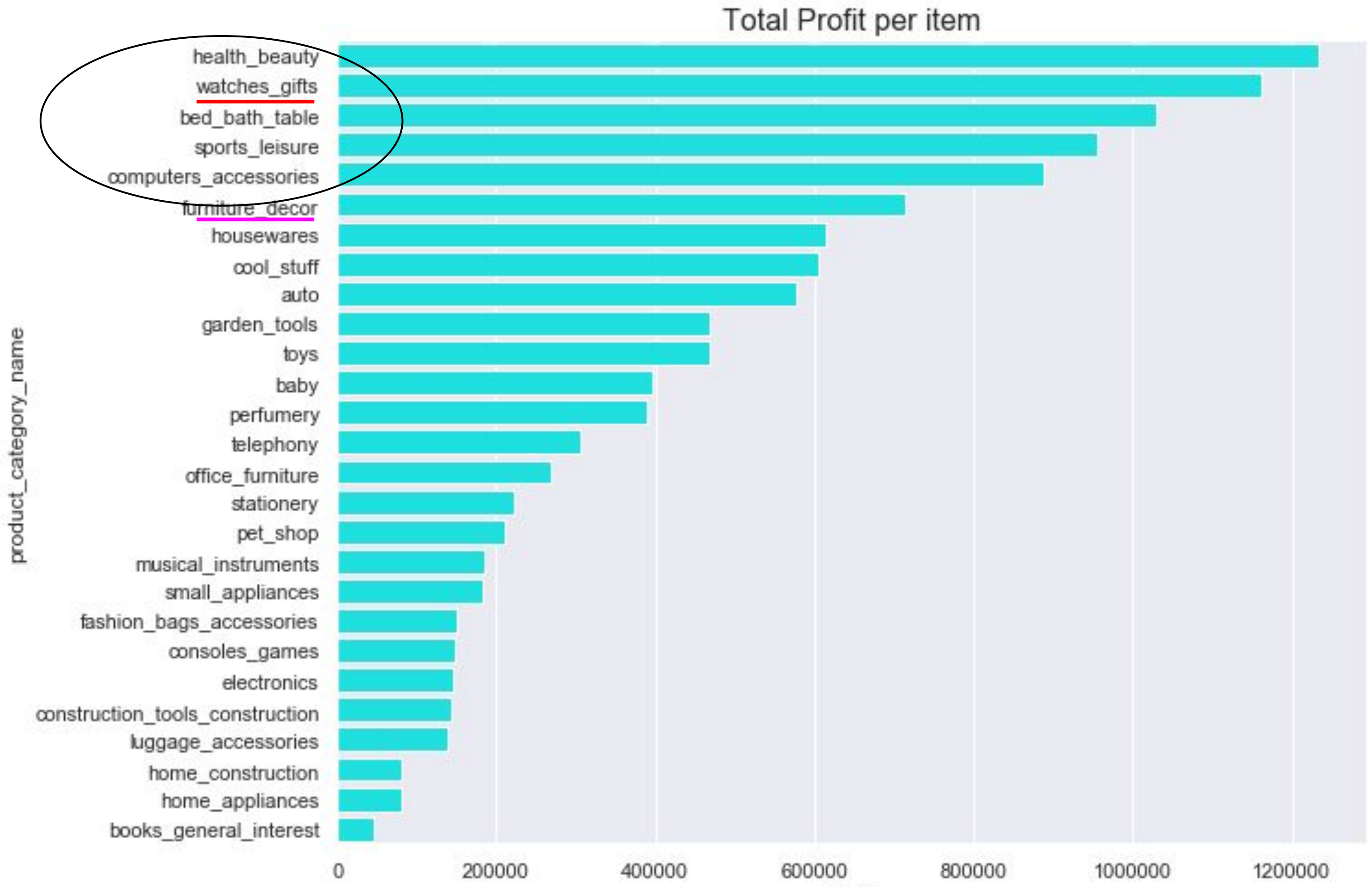
- ❑ Methodology allows us to use item and user metadata to our recommender algorithm[4].
- ❑ With user data we can predict without any prior history for a user.
- ❑ This solves the “cold start” problem that collaborative and content based methods don’t handle on their own.
- ❑ This “Hybrid” method will be used to create a recommendation engine for Olist with data from the machine learning and data science community, kaggle

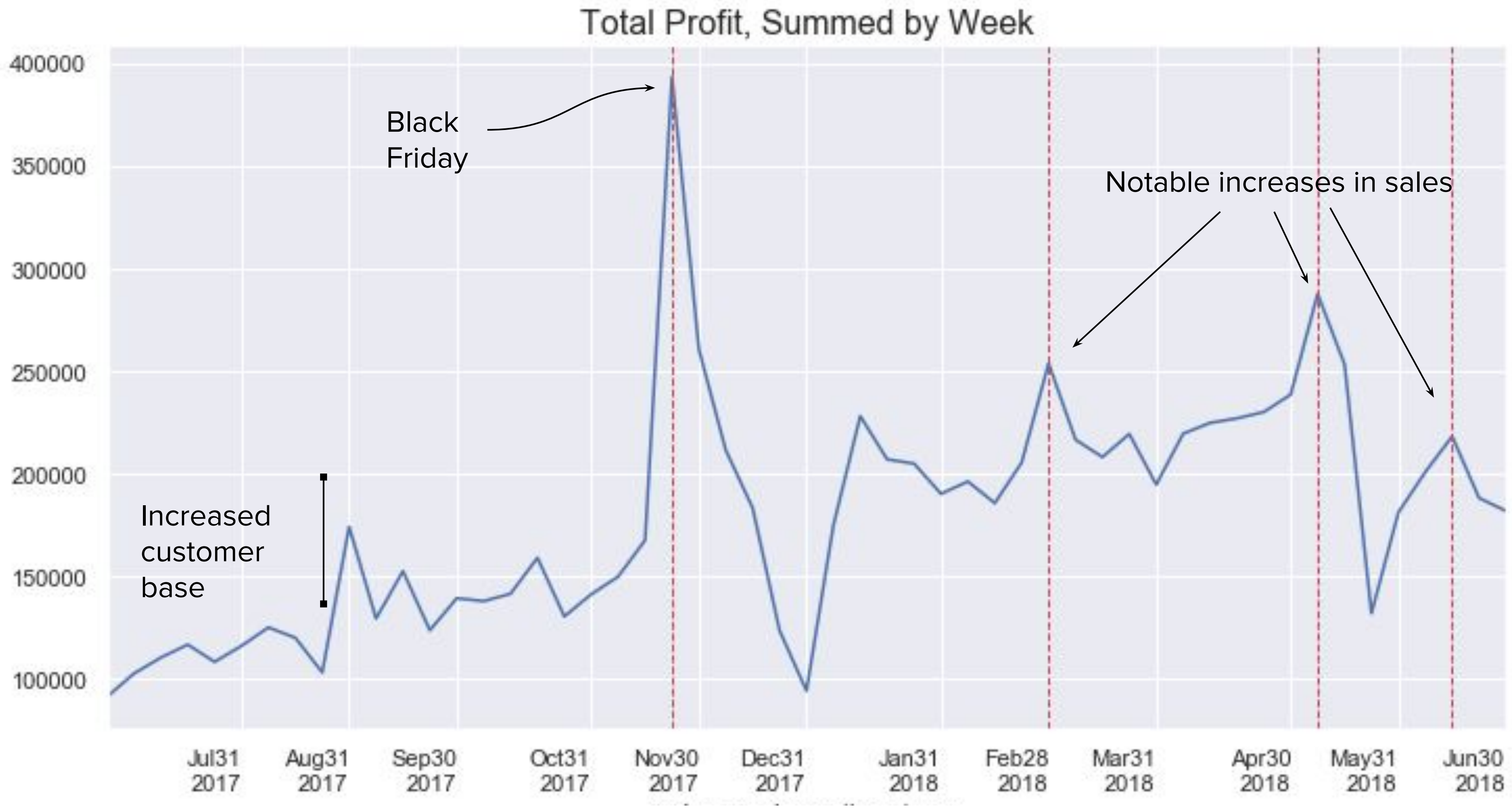
Exploratory Analysis

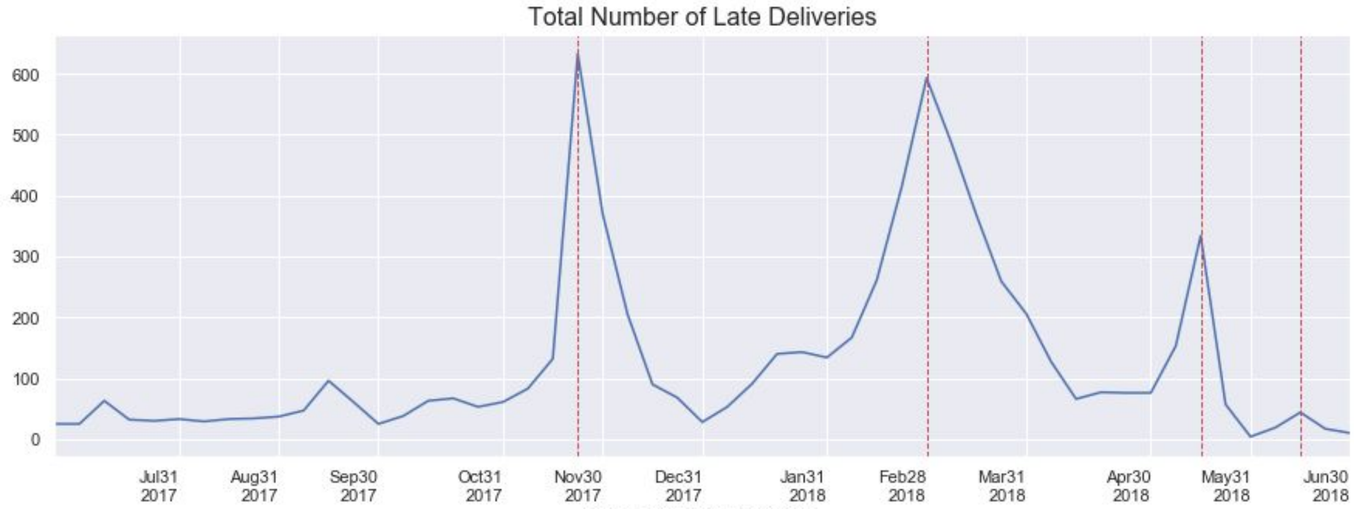
- ❏ Analysis was conducted to reveal customer seasonal trends in the data.
- ❏ Most popular items were found to be **bed and bath**, **health and beauty**, **sports leisure**, **furniture decor** and **computer accessories**.
- ❏ Most profitable items were the same except **watches and accessories**, came in 2nd for profit, although being 7th in popularity..
- ❏ Olist managed to increase sales by **600%** in a single day in 2017, Black Friday.
- ❏ Following a seasonal dip after christmas, the **customer base increased by 50%**.
- ❏ Despite this success, **only 4%** of the users in this time span purchased on more than one occasion.
- ❏ Analysis found a strong correlation between **late deliveries** and **customer satisfaction**.



Most Popular by total items purchased
product_category_name
bed_bath_table
health_beauty
sports_leisure
furniture_decor
computers_accessories
housewares
watches_gifts
telephony
garden_tools
auto
toys
cool_stuff
perfumery
baby
electronics
stationery
fashion_bags_accessories
pet_shop
office_furniture
consoles_games
luggage_accessories
construction_tools_construction
home_appliances
small_appliances
musical_instruments
home_construction
books_general_interest
0
2000
4000
6000
8000
10000


Total Profit per item
product_category_name
health_beauty
watches_gifts
bed_bath_table
sports_leisure
computers_accessories
furniture_decor
housewares
cool_stuff
auto
garden_tools
toys
baby
perfumery
telephony
office_furniture
stationery
pet_shop
musical_instruments
small_appliances
fashion_bags_accessories
consoles_games
electronics
construction_tools_construction
luggage_accessories
home_construction
home_appliances
books_general_interest
0
200000
400000
600000
800000
1000000
1200000


Total Profit, Summed by Week
400000
350000
300000
250000
200000
150000
100000
Jul31 2017
Aug31 2017
Sep30 2017
Oct31 2017
Nov30 2017
Dec31 2017
Jan31 2018
Feb28 2018
Mar31 2018
Apr30 2018
May31 2018
Jun30 2018
Black Friday
Notable increases in sales
Increased customer base



Late deliveries and satisfaction over time.

- Vertical dotted lines represent the same surges in profit seen in previous slide.
- It's clear that Olist had trouble dealing with surges, resulting in late deliveries and significant drops in customer satisfaction.

Recommendation System with LightFM

light fm

LightFM Recommender System

Two models were created for evaluation, both were hybrid recommendation algorithms but only one included user features, including clustered groupings for customers.

```
final_model1.fit(user_to_product_interaction,
          user_features=None,
          item_features=product_to_feature_interaction,
          sample_weight=None,
          epochs=5,
          num_threads=4,
          verbose=False)
```

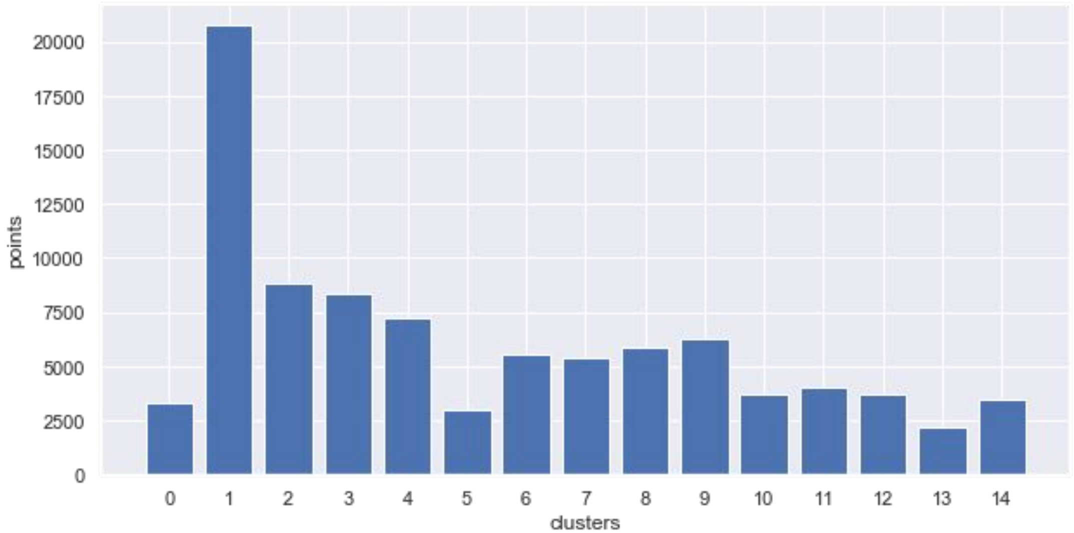
```
final_model2.fit(user_to_product_interaction,
          user_features=user_to_feature_interaction,
          item_features=product_to_feature_interaction,
          sample_weight=None,
          epochs=hyperparams2.pop('num_epochs'),
          num_threads=4,
          verbose=False)
```

Model1 includes information about the items and of what users purchased in the past.

Model2 has the same information, but also factors in user features, where we have clusters users into groups.

Clustering user tendencies for LightFM model use

In order to understand more about users, KMeans clustering techniques were used to create groupings. Each user was assigned a cluster based on similarities in purchase behavior. The groups below represent the number of users in each group, in all there were 15 groups created.



Sample Predictions (User ’37b5ae93bb8e35c742a4bcd701395daa’)

Model1: Without user clusters

Known positives:

c6dd917a0be2a704582055949915ab32
cool_stuff

Recommended:

389d119b48cf3043d311335e499d9c6b
garden_tools
368c6c730842d78016ad823897a372db
garden_tools
422879e10f46682990de24d770e7f83d
garden_tools

Model2: With user clusters

Known positives:

c6dd917a0be2a704582055949915ab32
cool_stuff

Recommended:

f35927953ed82e19d06ad3aac2f06353
books_general_interest
5d66715cc928aadd0074f61332698593
electronics
6a8631b72a2f8729b91514db87e771c0
electronics

I like to think that books and electronics are cool! Although that’s just my opinion, nothing against garden tools. It’s difficult to say who wins this comparison.

Sample Predictions (User ’c8ed31310fc440a3f8031b177f9842c3’)

Model1: Without user clusters

Known positives:

4a5c3967bfd3629fe07ef4d0cc8c3818
construction_tools_construction

Recommended:

18fa9cc25ea8b54f32d029f261673c0f
construction_tools_construction
97d94ffa4936cbc2555e83aefc1f427b
construction_tools_construction
cd46a885543f0e169a49f1eb25c04e43
computers_accessories

Model2: With user clusters

Known positives:

4a5c3967bfd3629fe07ef4d0cc8c3818
construction_tools_construction

Recommended:

0aabfb375647d9738ad0f7b4ea3653b1
consoles_games
d017a2151d543a9885604dc62a3d9dcc
fashion_bags_accessories
4fe644d766c7566dbc46fb851363cb3b
art

I don’t know if someone who purchased construction tools would immediately be interested in video games, fashion bags or art. This suggestion is a bit off for model2.

Sample Predictions (User ‘0848ef3901afaa99199cbe1bbbd71e1a’)

Model1: Without user clusters

Known positives:

cba54528d3adcef3c12d8e8b9a48bc17
bed_bath_table

Recommended:

a0fe1efb855f3e786f0650268cd77f44
agro_industry_and_commerce
1a300f482e35d7eac74b229be067aefd
computers_accessories
466d263ce8b7bd275003ee2104428127
telephony

Model2: With user clusters

Known positives:

cba54528d3adcef3c12d8e8b9a48bc17
bed_bath_table

Recommended:

99a4788cb24856965c36a24e339b6058
bed_bath_table
06edb72f1e0c64b14c5b79353f7abea3
bed_bath_table
ec2d43cc59763ec91694573b31f1c29a
bed_bath_table

This customer only bought items in bed and bath category. I don’t think that’s enough to determine interest in such a broad range of items for model1. Model2 seems more appropriate here.

Sample Predictions (User ’02ce431b0797023384d47edad5dd7284’)

Model1: Without user clusters

Known positives:
893b9464c6ab7f700148fe9db838b6b4
stationery

Recommended:
43423cdffde7fda63d0414ed38c11a73
watches_gifts
2ffdf10e724b958c0f7ea69e97d32f64
watches_gifts
e0d64dcfaa3b6db5c54ca298ae101d05
watches_gifts

Model2: With user clusters

Known positives:
893b9464c6ab7f700148fe9db838b6b4
stationery

Recommended:
fb55982be901439613a95940feefd9ee
stationery
e03102efbc2229024c89be731f0aedcb
stationery
c706d50b57c9e83293c2586d01f32445
stationery

Stationary items are like office supplies. Often when buying office supplies you can be lured into buying more. I think model2 has the advantage here.

Summary of Results

It’s impossible to determine which model is better by looking at every customers recommendations so we have to rely on metrics here, in that case model2 is the better choice. Using clustering may have given our final model (model2) the ability to recognize latent factors among users and therefore the ability to connect relevant products in a more consistent manner.

While this did produce what seemed to be a better performing recommendation engine, there is still some healthy scepticism. It is reasonable to assume that our model is fitting too closely to the cluster ids and not enough to other traits. As could be inferred from some of the example recommendations in previous slides. This will be an ongoing project, check for updates at my public repository to see what developments have been made!

Personal Profile links

Repository of results.
Check it out!

References

1. http://rejoiner.com/resources/amazon-recommendations-secret-selling-online/
2. https://valorcapitalgroup.com/case-studies/olist-redesigned-the-marketplace-business-model-to-fit-the-realities-of-ecommerce-in-brazil/
3. https://www.similarweb.com/website/olist.com/#referrals
4. https://www.themarketingtechnologist.co/building-a-recommendation-engine-for-geek-setting-up-the-prerequisites-13/
5. https://making.lyst.com/lightfm/docs/home.html#:~:text=LightFM%20is%20a%20Python%20implementation,the%20traditional%20matrix%20factorization%20algorithms.