IMPACTS OF MAINSTREAM-DRIVEN ALGORITHMS ON RECOMMENDATIONS FOR CHILDREN ACROSS DOMAINS

A Reproducibility Study

Robin Ungruh, Alejandro Bellogín, Dominik Kowald, Sole Pera

TUDelft

UAM Universidad Autónoma de Madrid

KNOW Center

TU Graz

REFERENCE WORK*

- Exploring **differences between children's and mainstream users'** preferences
- Measuring the effect of the mainstream on recommendations for children

REPRODUCIBILITY STUDY

MOTIVATION

1. Children ≠ Mainstream
2. Deviating recommender behavior
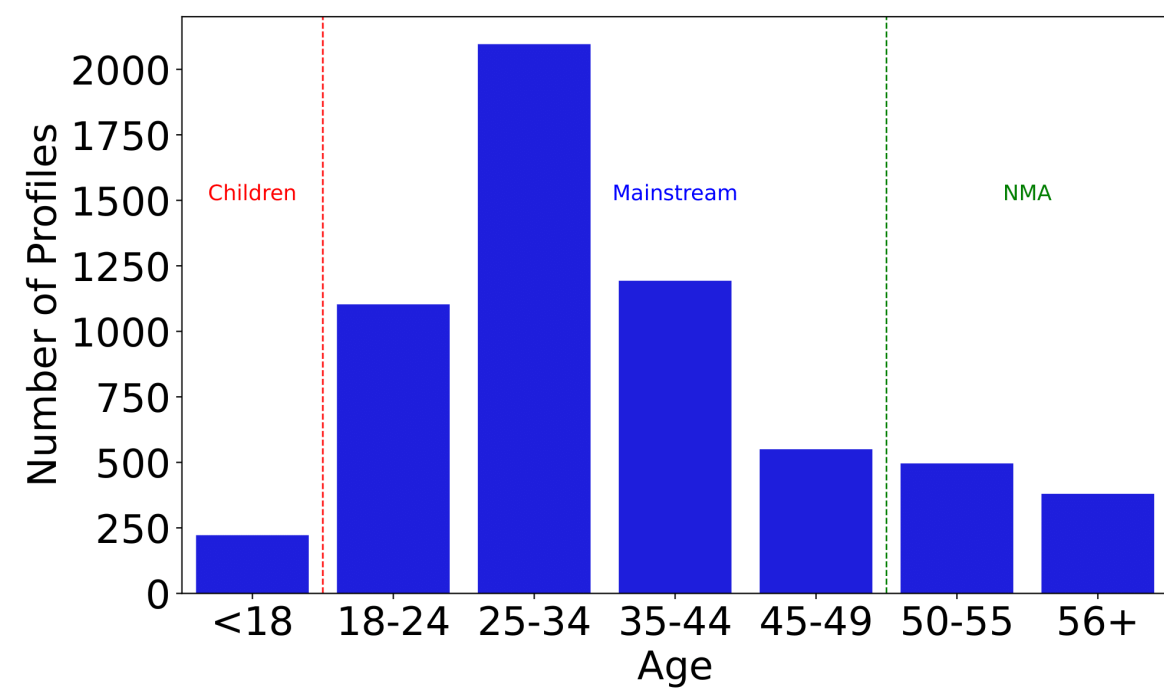3. Reproducibility Concerns

Domain	Dataset	New Domain	New Dataset	Pref. Dev. Exploration	RS Experiment
Movies	ML	✗	✗	repr + ext	repl + ext
Music	MLHD	✗	✔	repl + ext	repl + ext
Books	BX	✔	✔	repl + ext	repl + ext

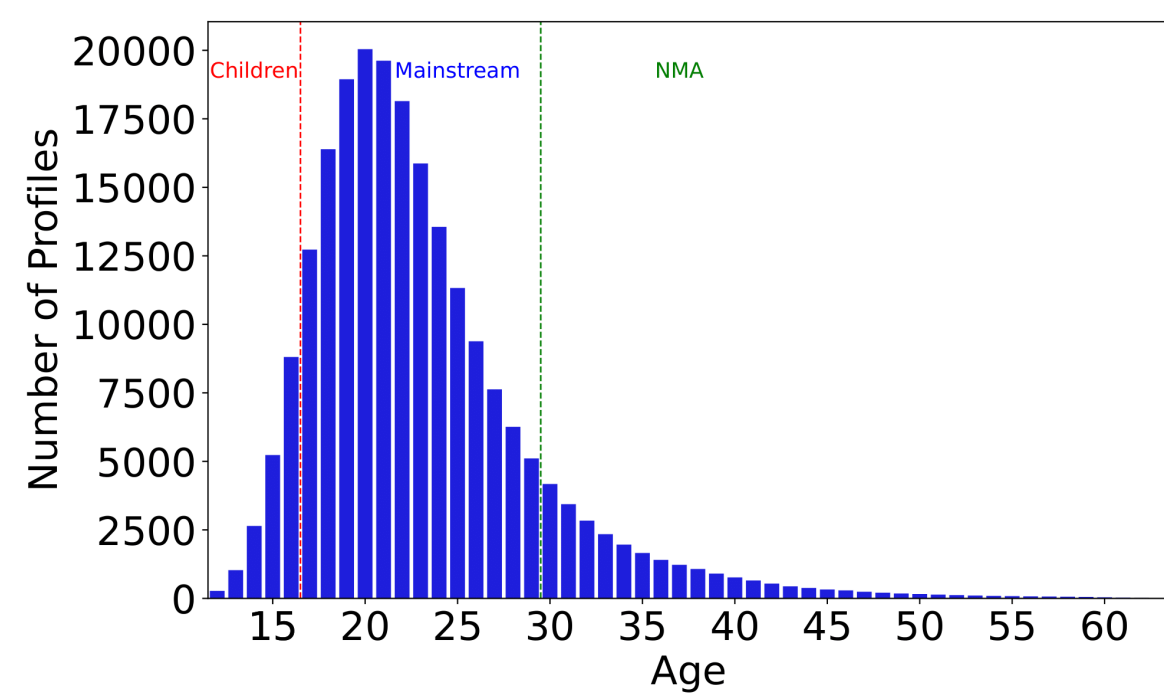**Table 1:** Overview of **reproducibility** (repr), **replicability** (repl) and **extension** (ext) efforts

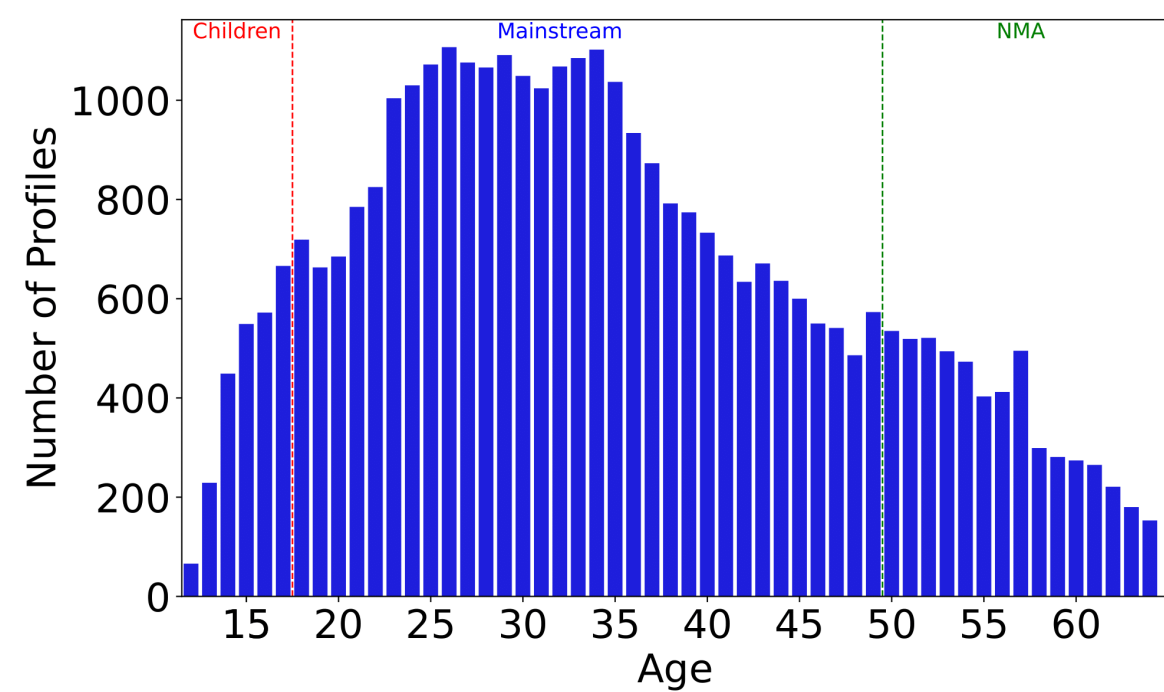## EXPERIMENTS

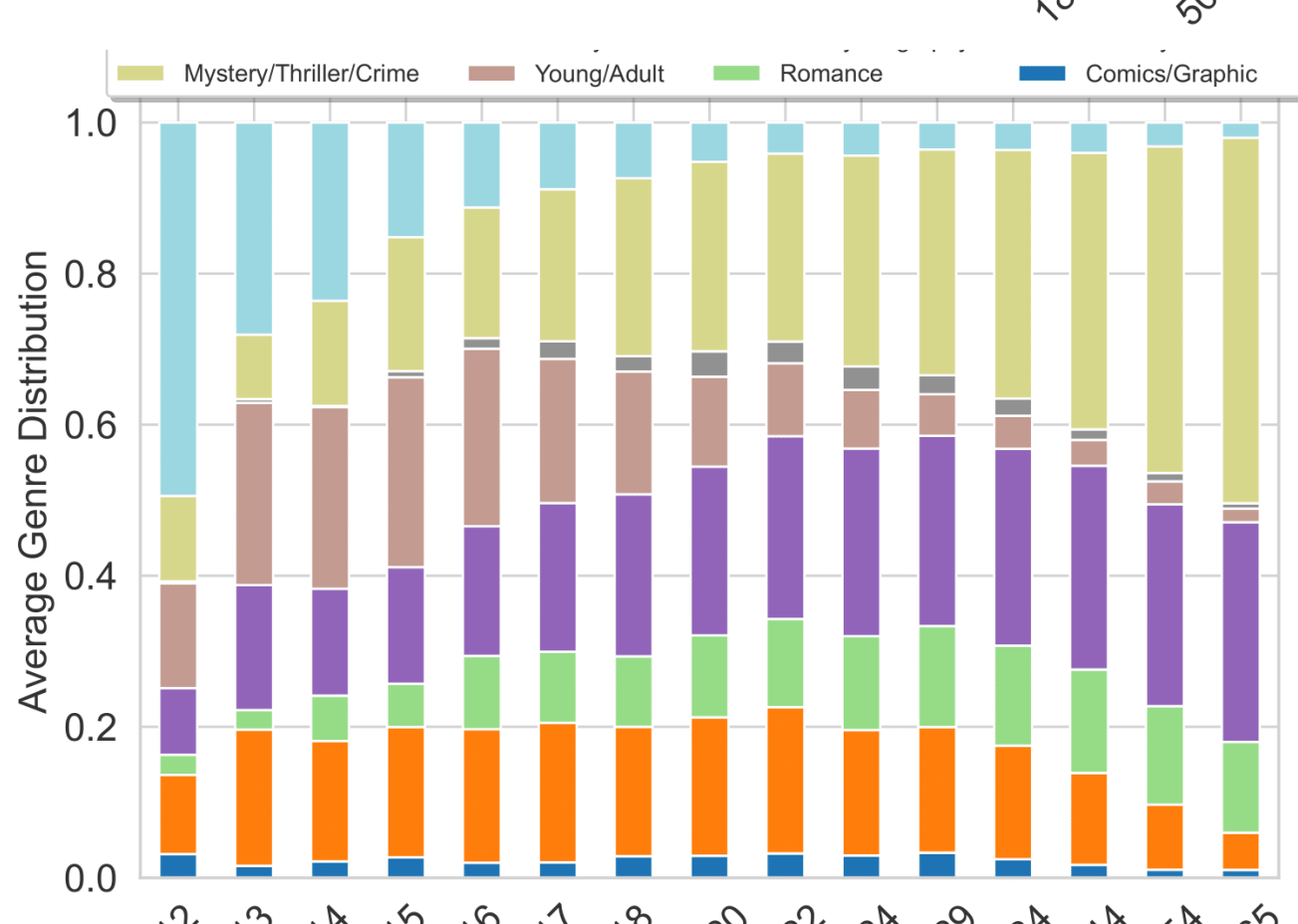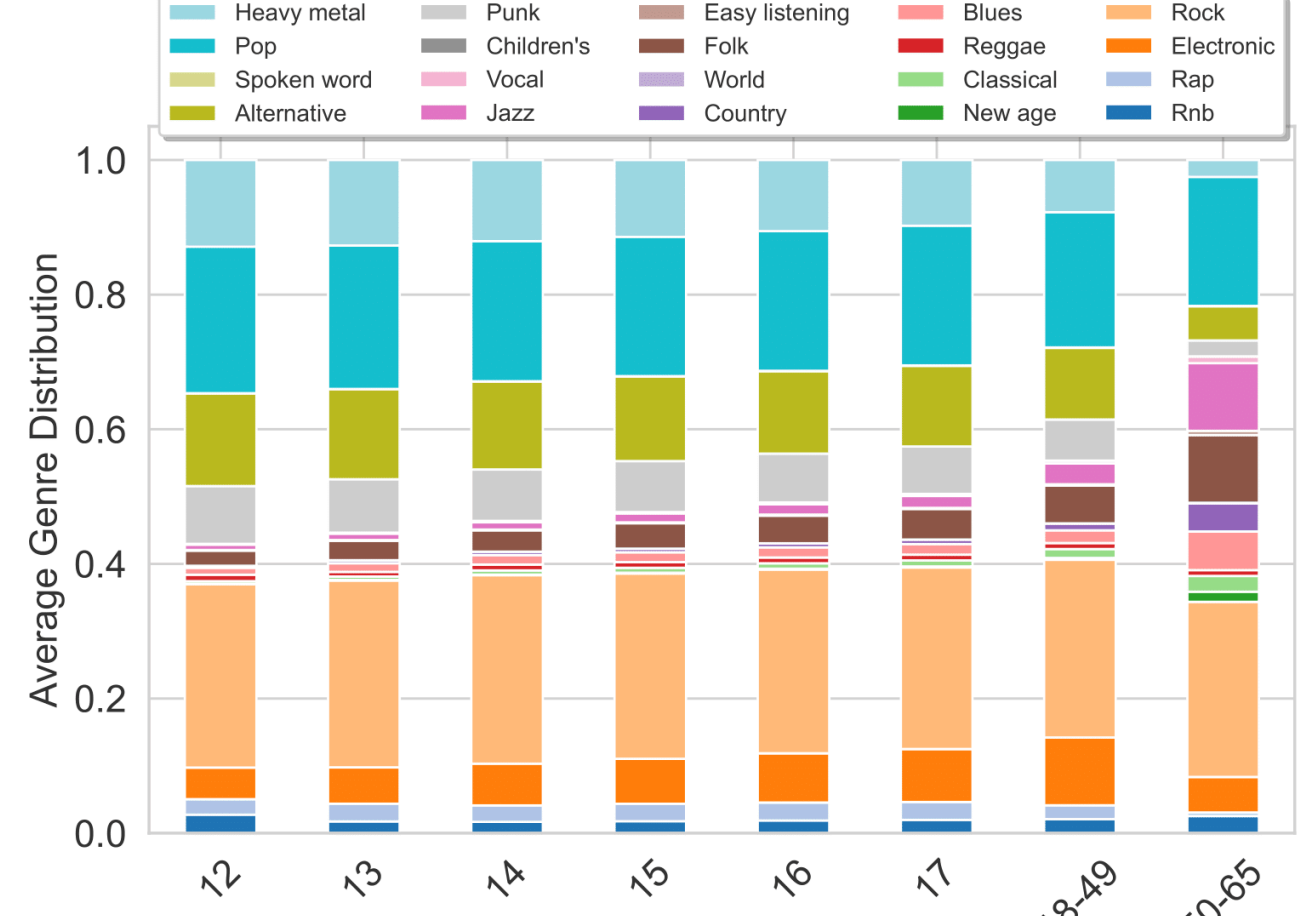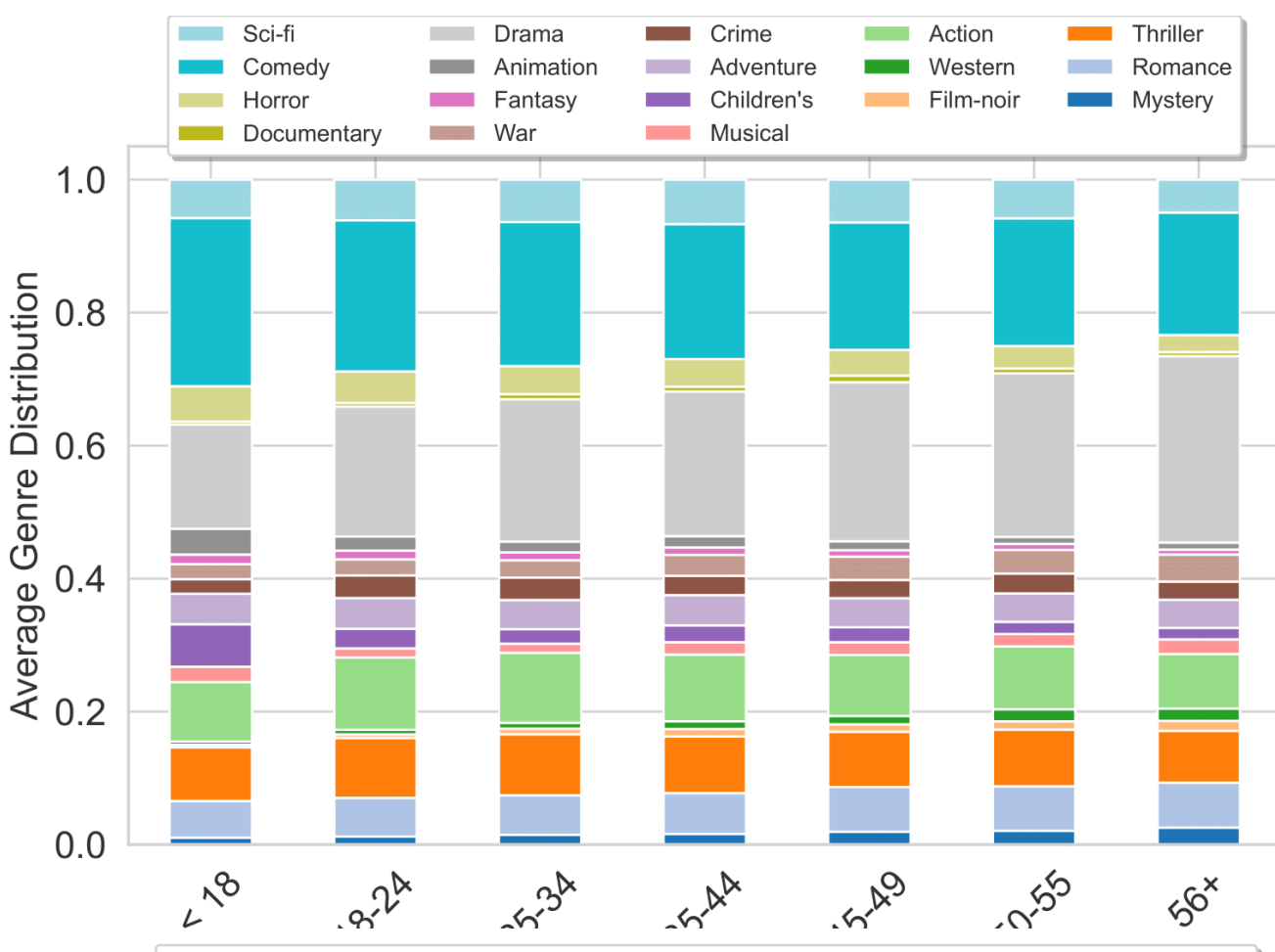### Number of **user profiles** per age in dataset

MOVIELENS-1M

MLHD

BOOK-CROSSING

Average **genre distributions** of users per age group

Dissimilarity of genre distributions
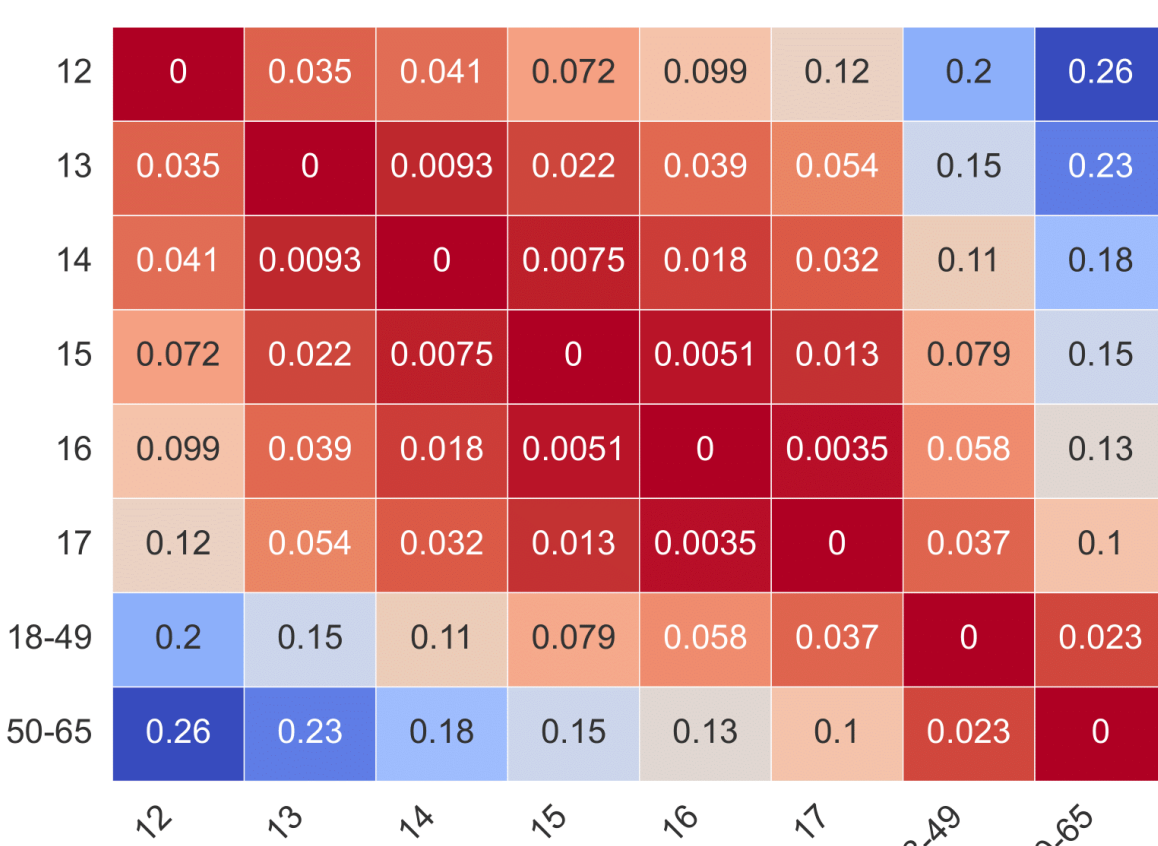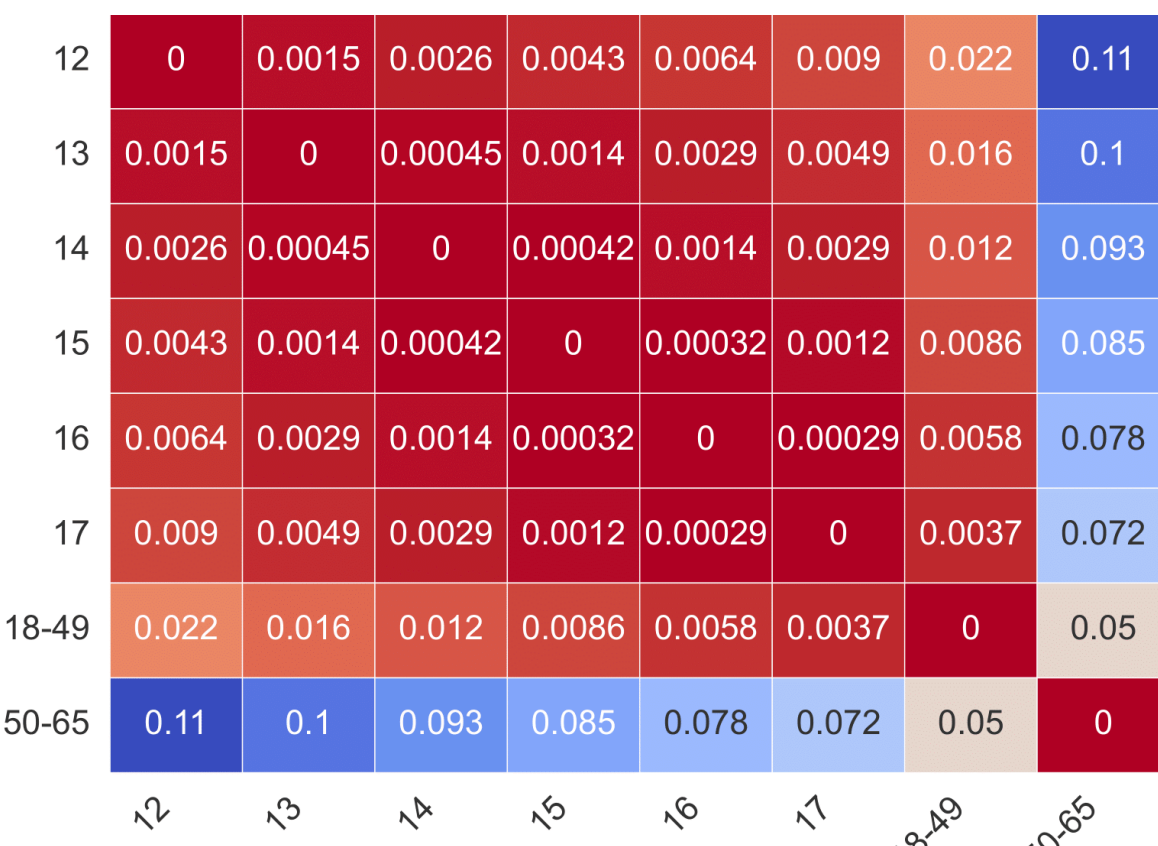
Popularity of consumed items

RS EXPERIMENT

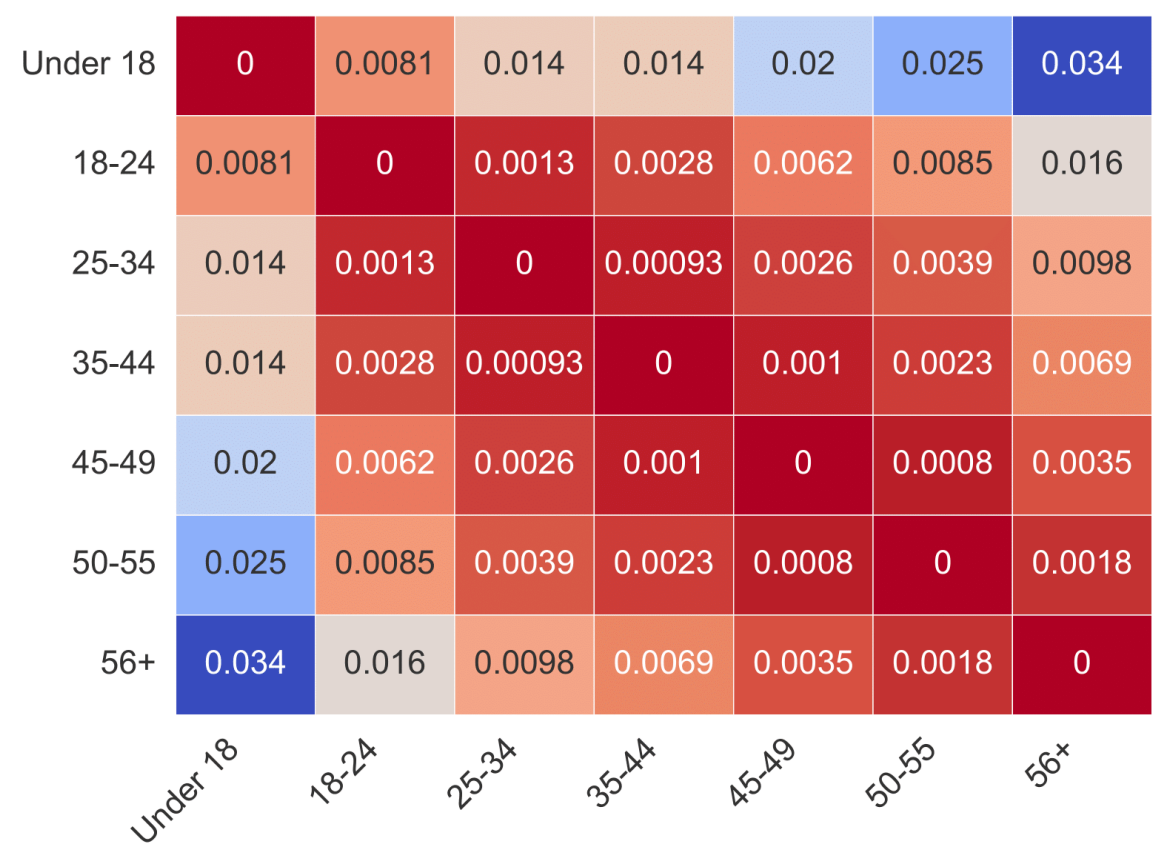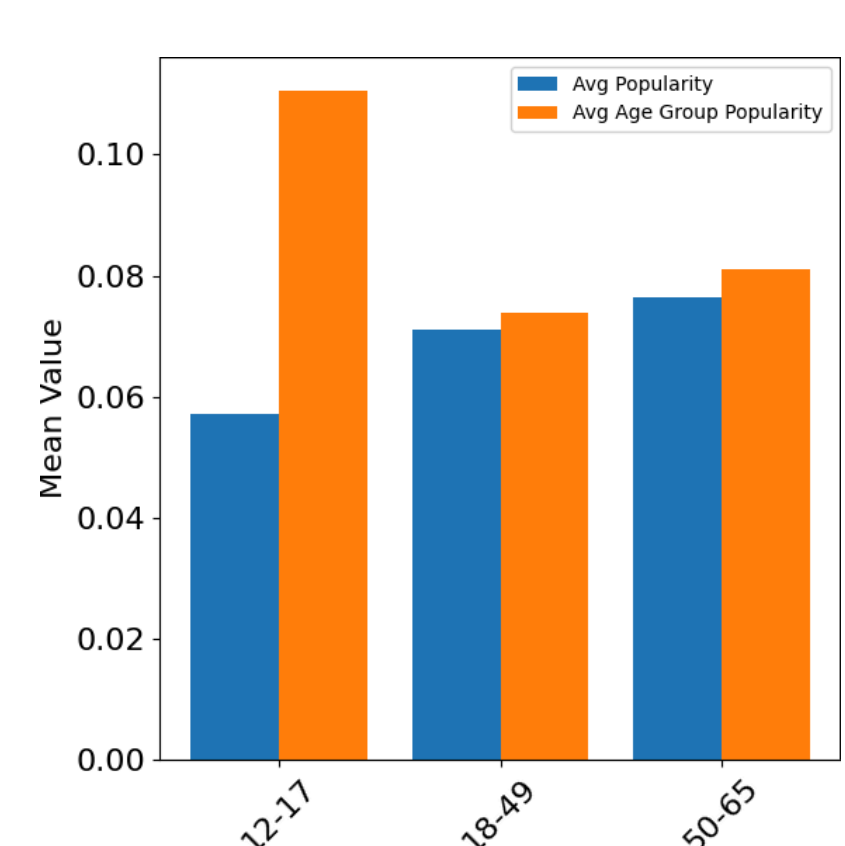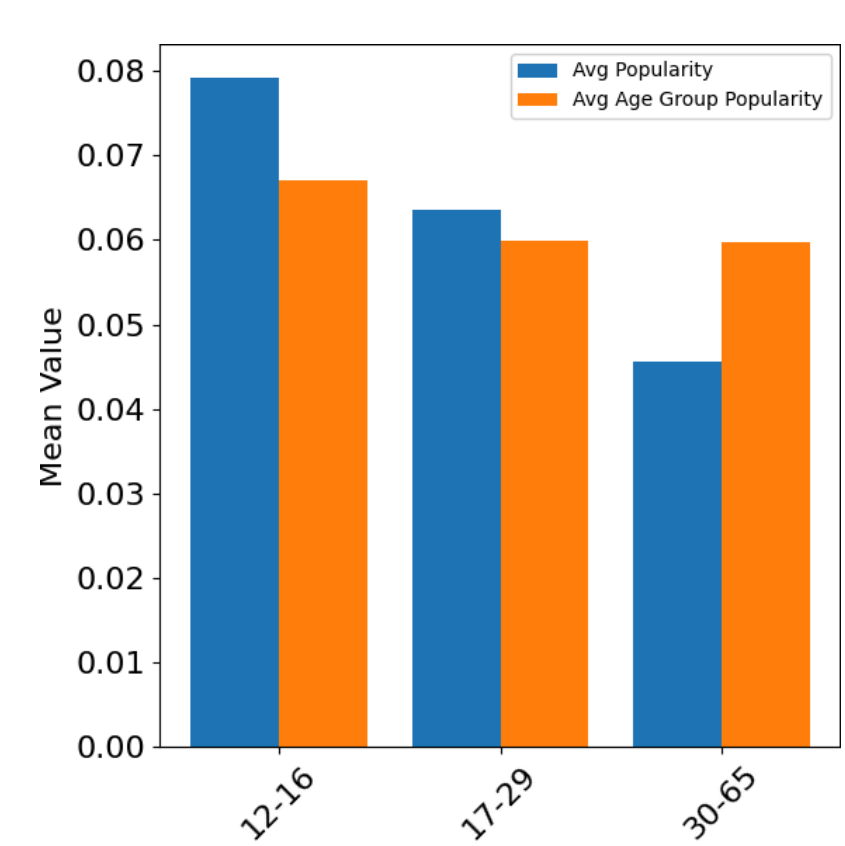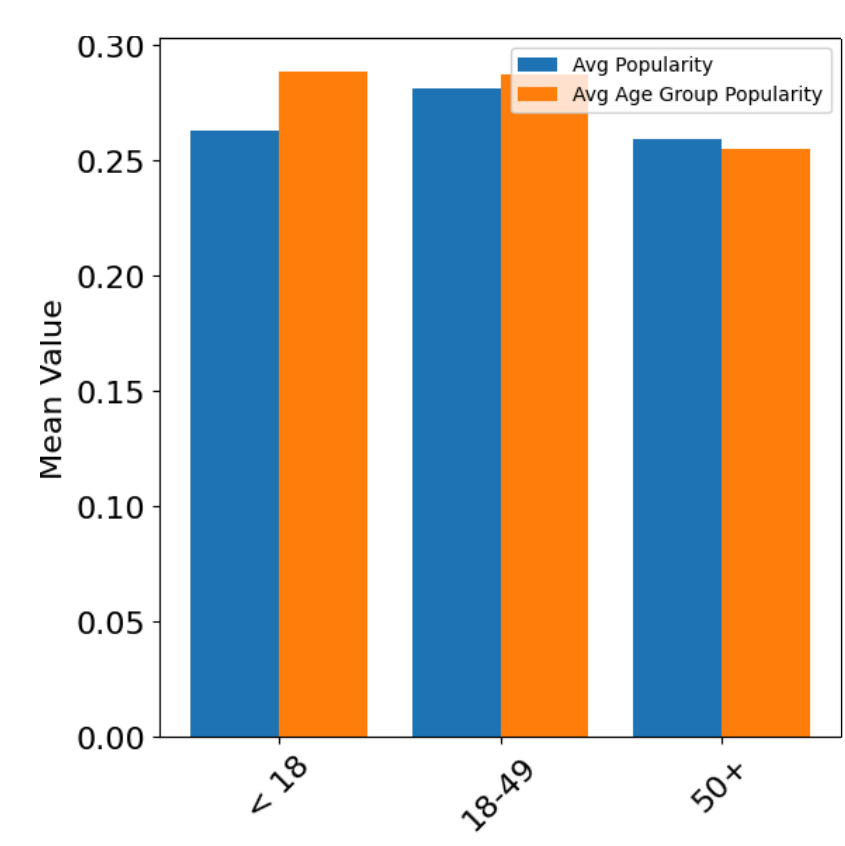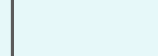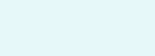Exploring the **impact of the mainstream** on recommendations for children.

		Movielens-1M			MLHD			Book-Crossing		
		nDCG↑	GMC↓	PL→0	nDCG↑	GMC↓	PL→0	nDCG↑	GMC↓	PL→0
MostPop	Children (CS)	**.152**	**.140**	**1.410**	**.013**	**.136**	**5.690**	.025	**.261**	**3.509**
	Children (GS)	.129	.151	1.768	.011	.139	6.429	.020	.308	7.961
	Mainstream	.174	.120	1.263	.008	.135	7.751	.030	.199	6.220
	NMA	.126	.147	1.594	.005	.146	9.229	.021	.181	5.864
iALS	Children(CS)	**.197**	**.073**	**0.214**	**.033**	**.060**	**1.811**	**.034**	**.210**	**0.508**
	Children (GS)	.292	.054	0.377	.038	.042	0.826	.106	.160	1.242
	Mainstream	.322	.047	.311	.030	.050	1.083	.080	.115	0.926
	NMA	.302	.055	0.363	.026	.052	1.278	.060	.096	1.006
	Trends reproduced?	👎			👍			👎		

Table 2: nDCG, Genre-Miscalibration (GMC), and Popularity Lift (PL) per age group.
Color shows significant difference between children and mainstream (red = better, green = worse).
Impact of mainstream assessed by comparing outcomes of training on the entire set (GS) vs training solely on child data (CS).
Boldened score marks significant difference between GS and CS.

IMPLICATIONS

- Some trends reproduced (MLHD aligns with LFM-2b); salient differences between domains.
- In domains where children prefer popular items (i.e., music), classic recommenders work well; in others they **fail** to capture children's preferences.
- Most recommenders require mainstream data to create fitting recommendations for children.

REFERENCES

* Ungruh, R., Bellogín, A. and Pera, M.S., 2025. The Impact of Mainstream-Driven Algorithms on Recommendations for Children. In *ECIR 2026*

Vigliensoni, G. and Fujinaga, I., 2017. The music listening histories dataset. In *ISMIR 2017*

Ekstrand, M.D., Tian, M., Azpiazu, I.M., Ekstrand, J.D., Anuyah, O., McNeill, D. and Pera, M.S., 2018. All the cool kids, how do they fit in?: Popularity and demographic biases in recommender evaluation and effectiveness. In *FAccT 2018*