

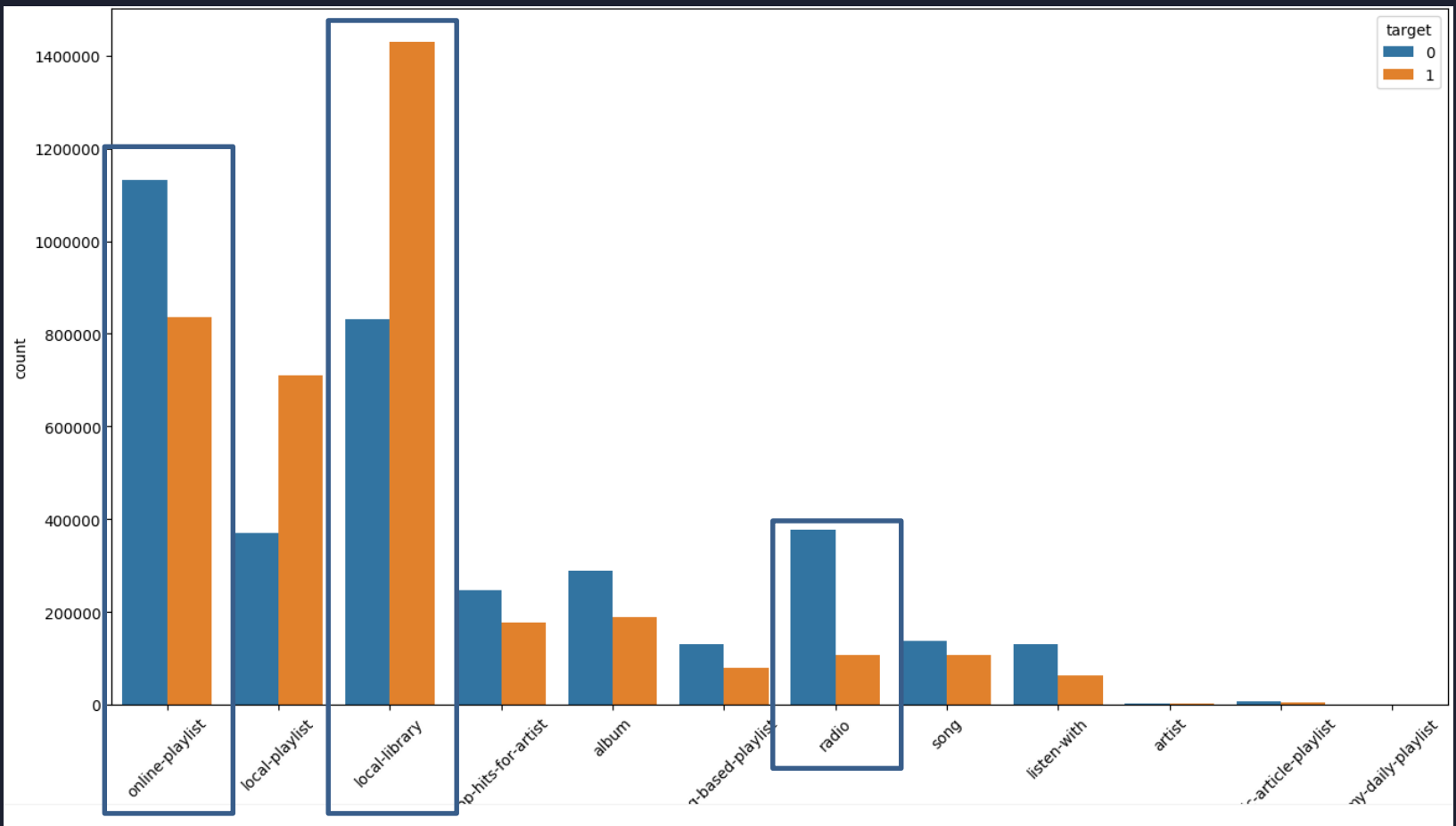
BIG DATA (X-LARGE)

WSDM – KKBox's Music Recommendation Challenge

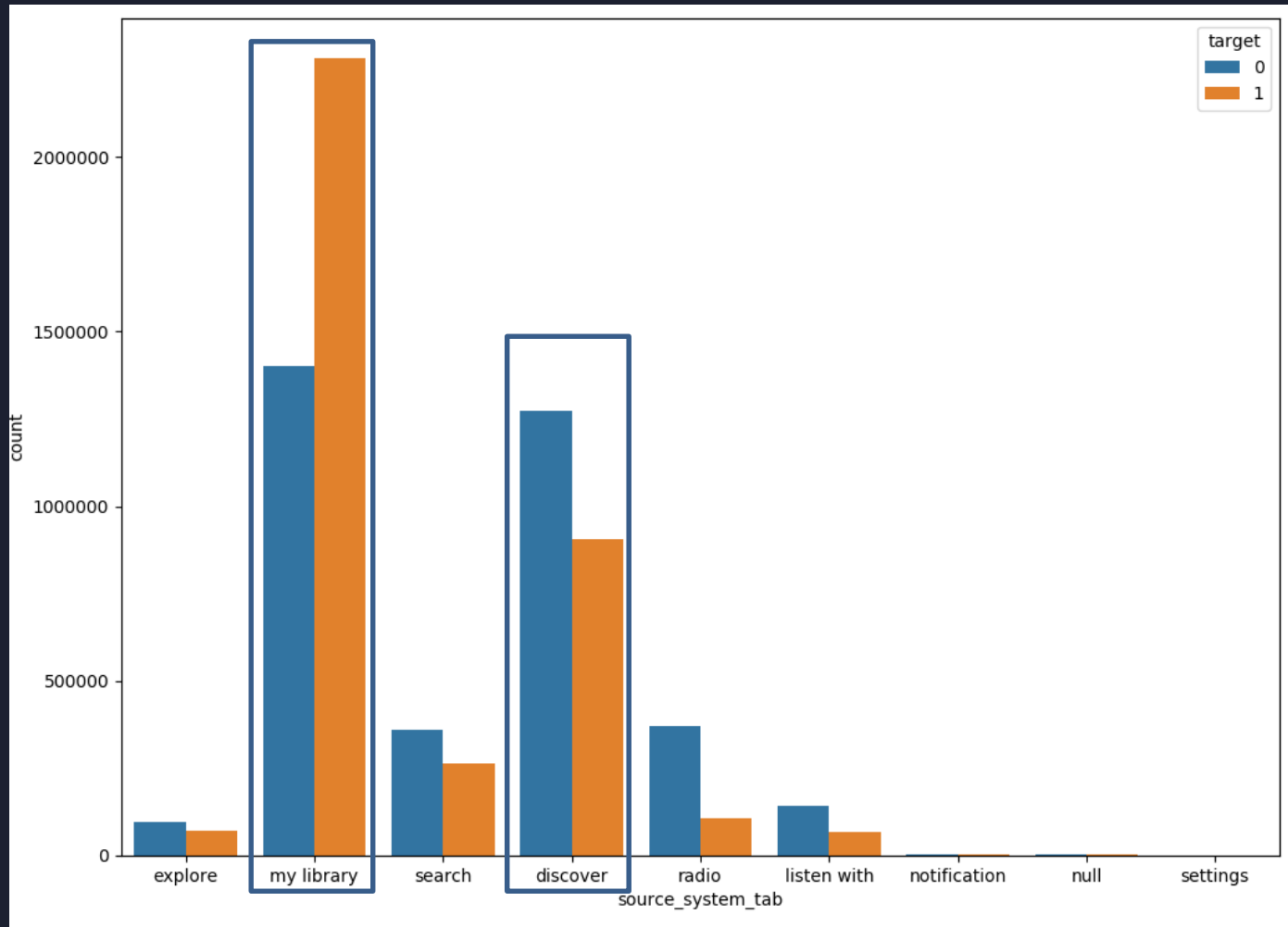


Data merge & analysis

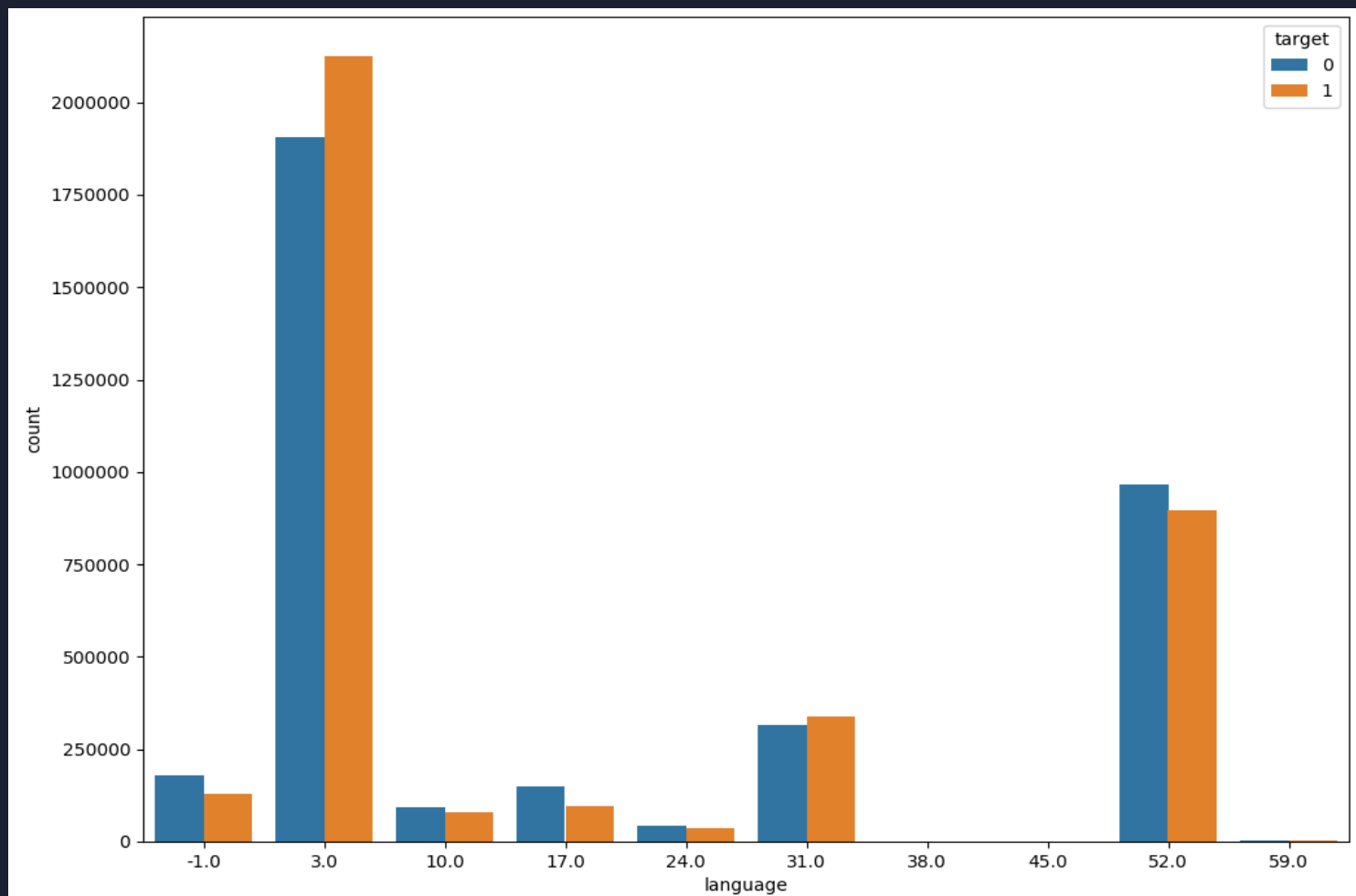




Target by source_type



Target by source_system_tab

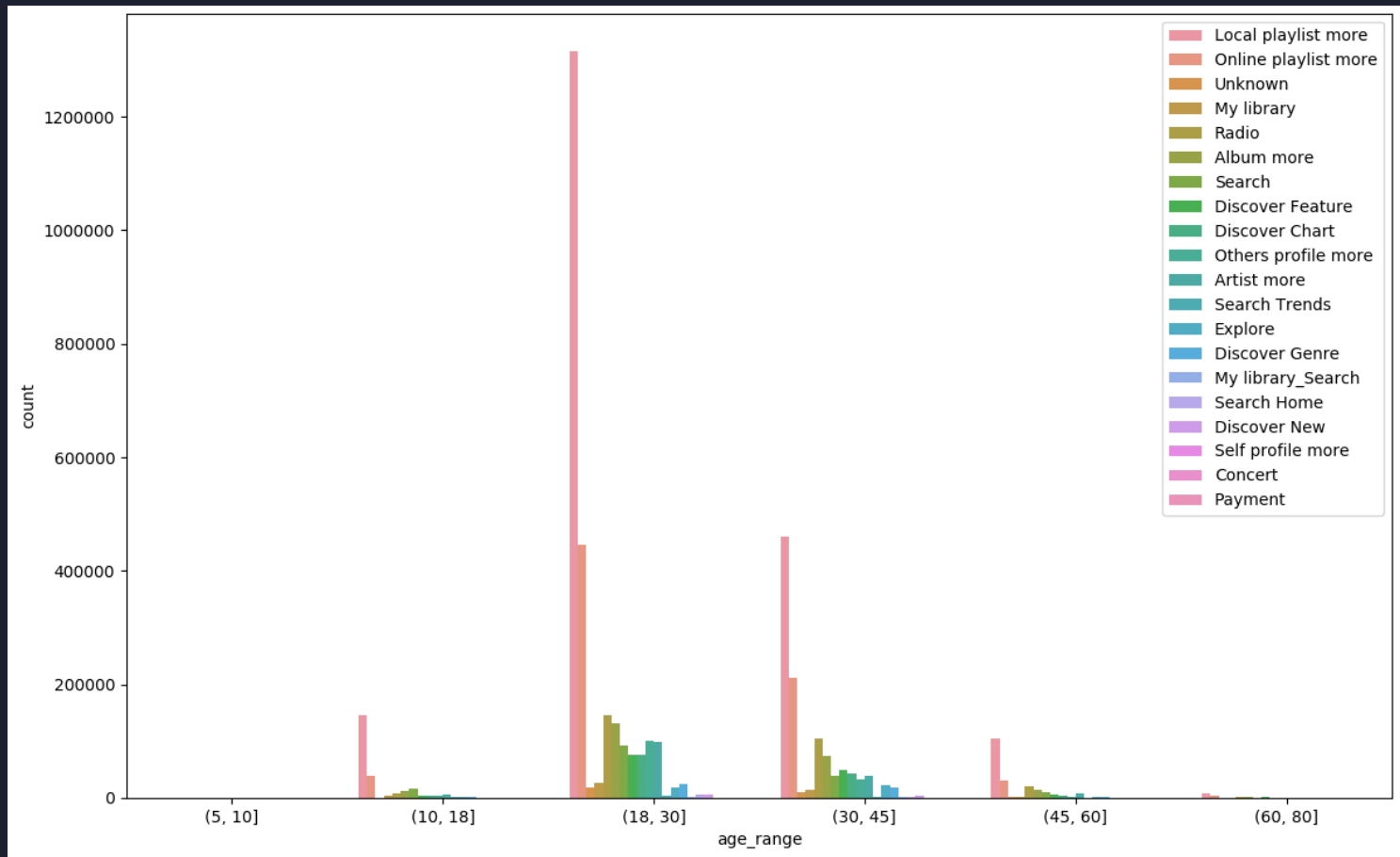


Taiwan

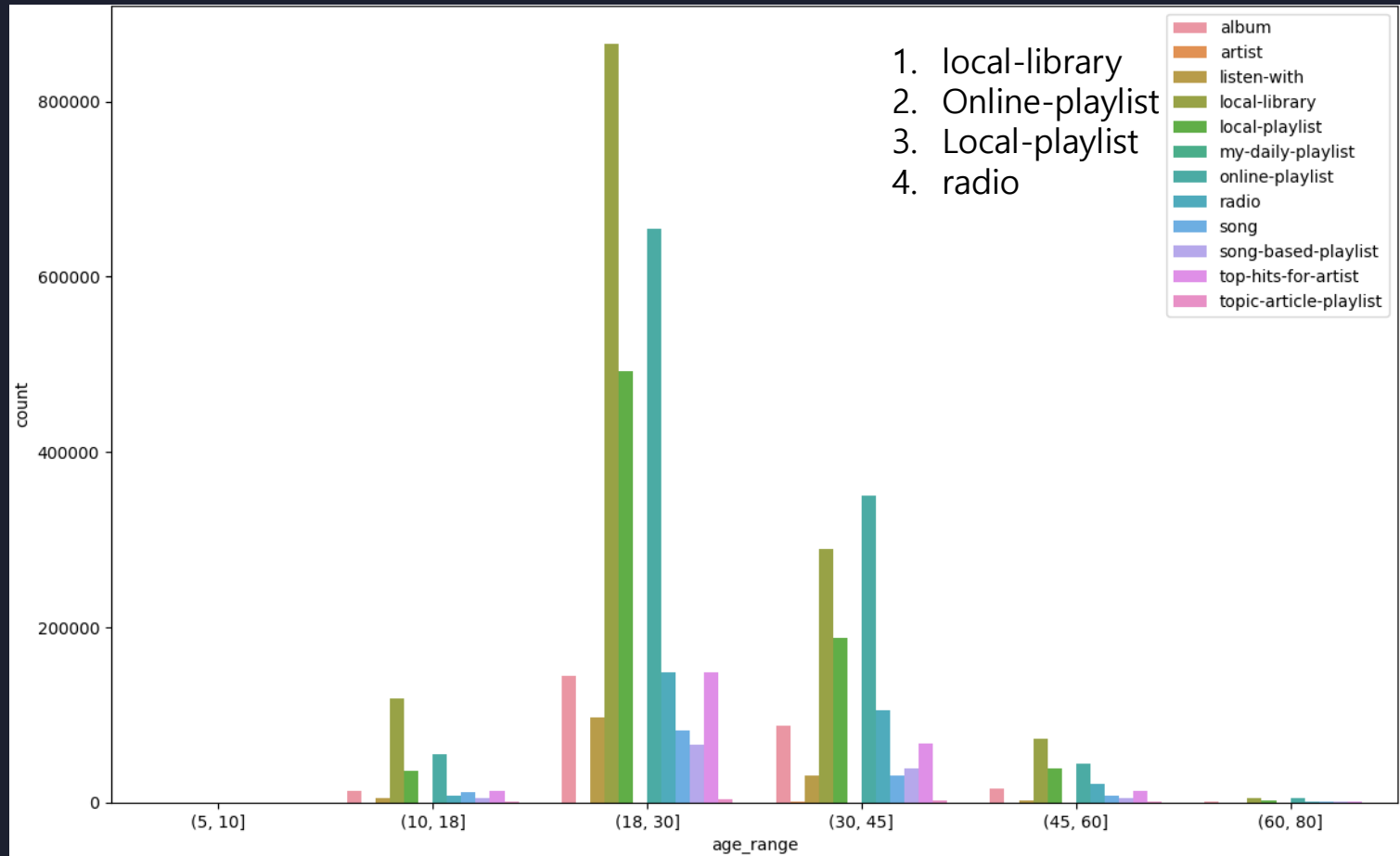
K-Pop

Pop song

Target by languages

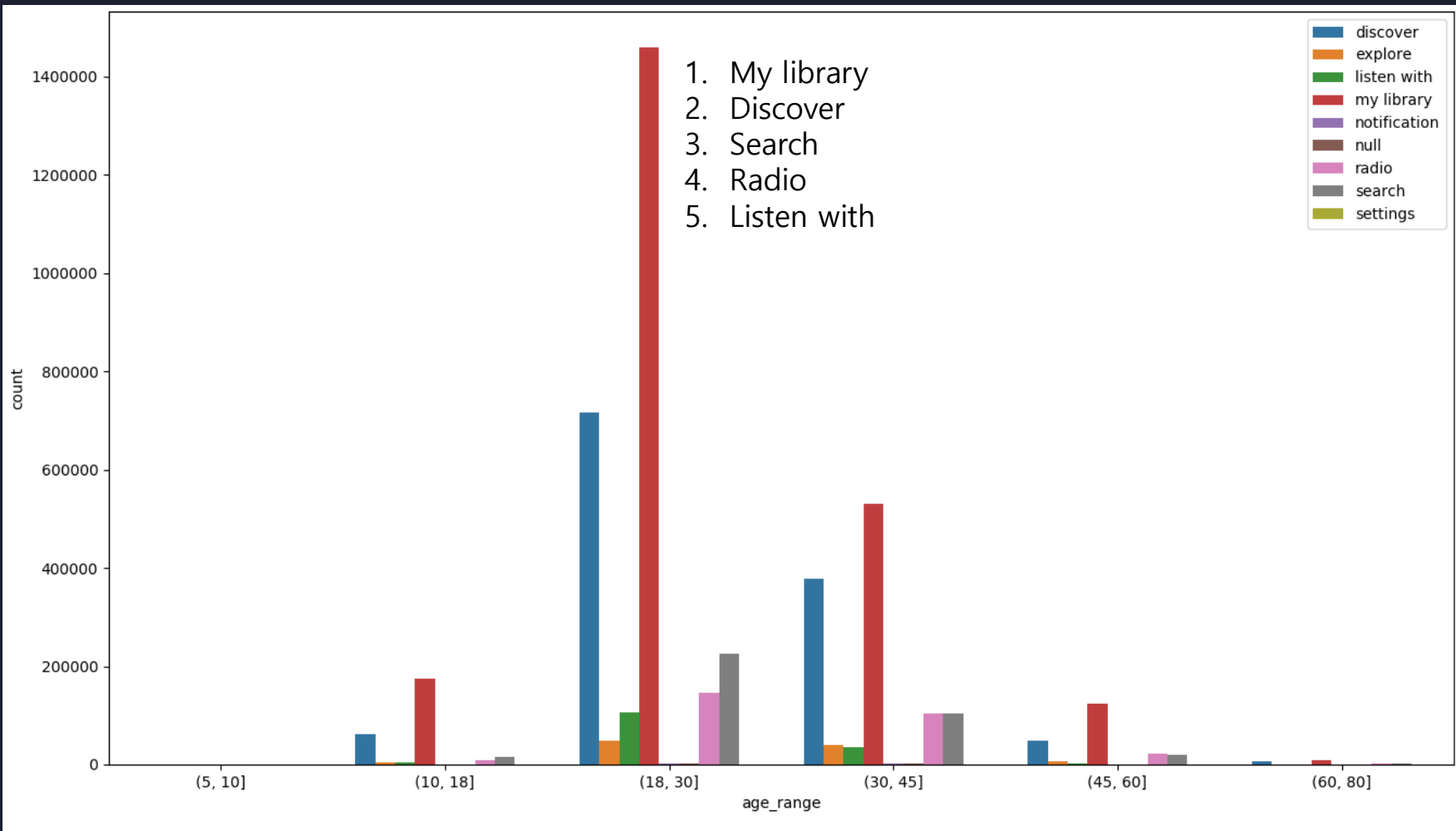


Source_screen_name by age range

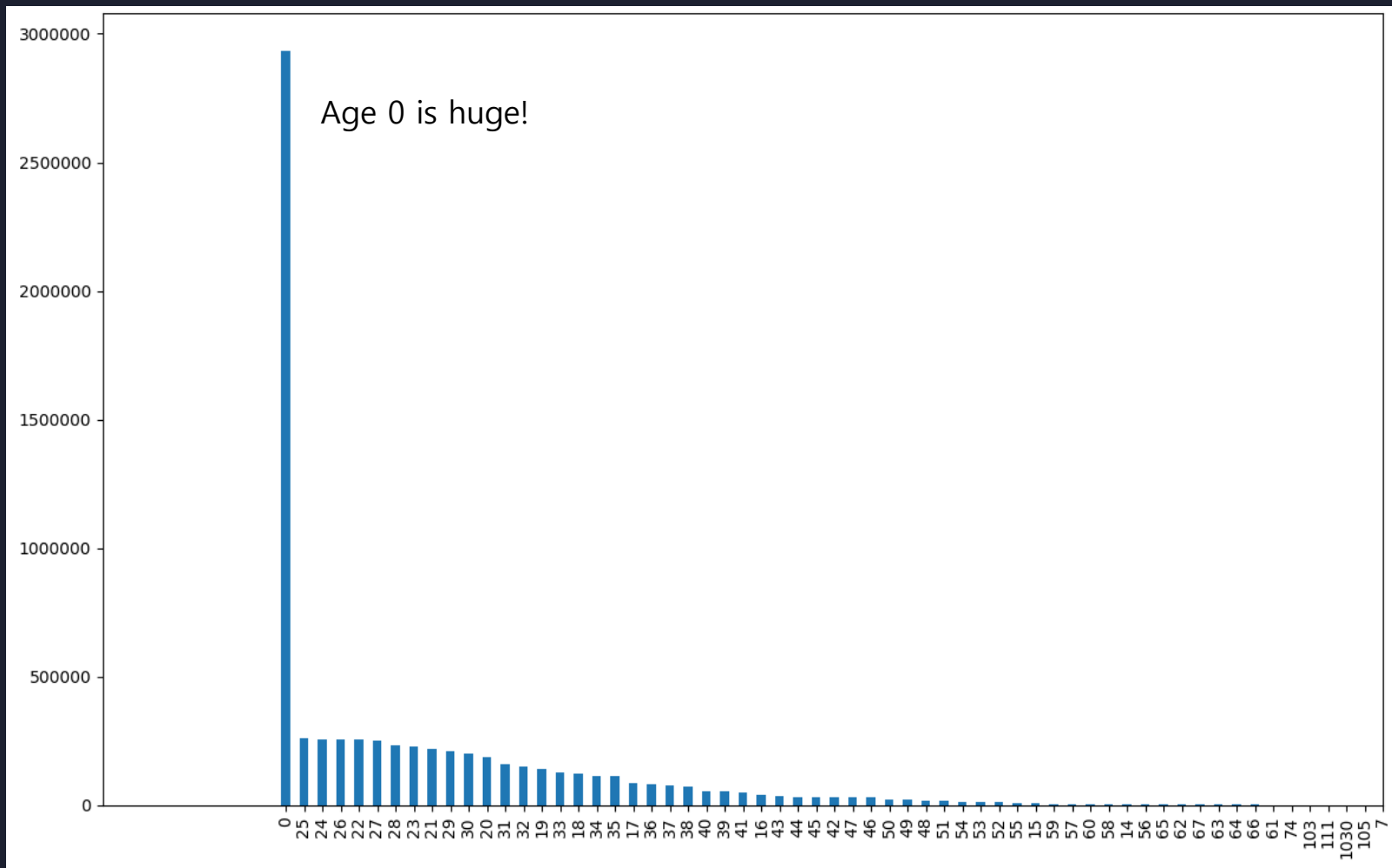


Source_type by age range

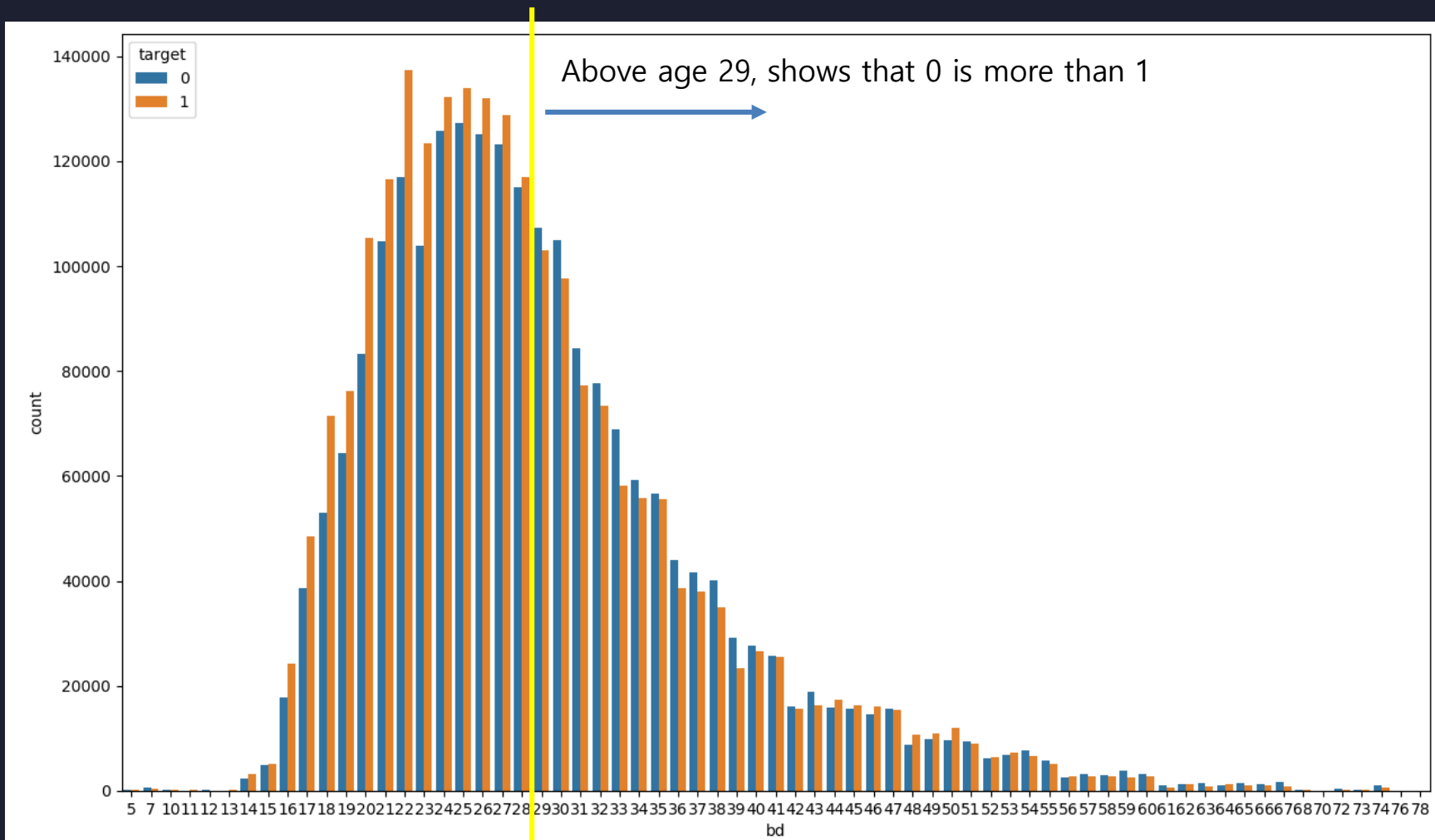
1. My library
2. Discover
3. Search
4. Radio
5. Listen with



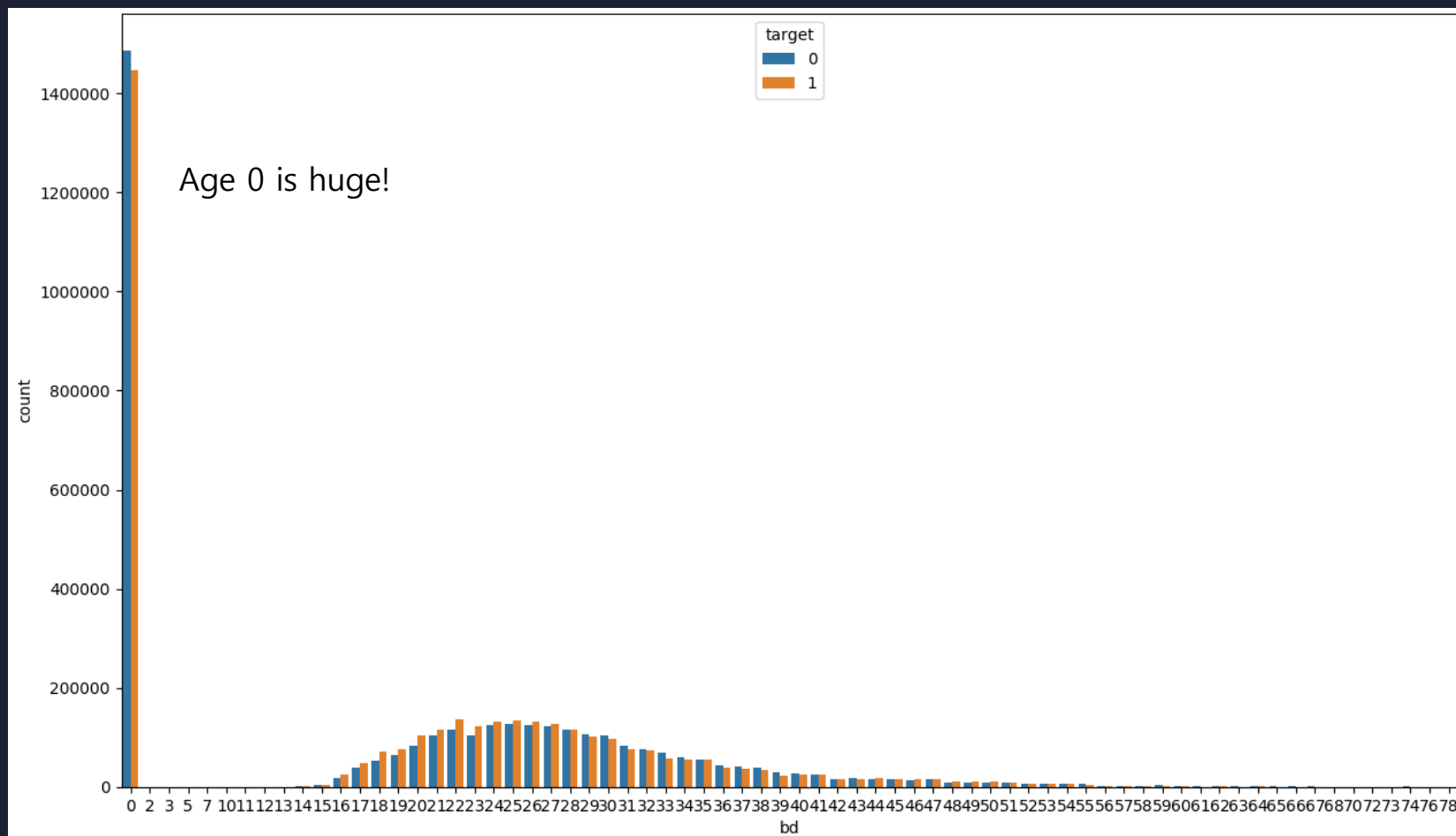
Source_syetem_tab by age



Age count



Target by ages



Target by ages including age 0

Random Forest in a nutshell



Random Forest



Tensorflow-GPU

Feature Range: 3 source columns (3 nominal, 0 numeric)

Parameters

Records	Used for train	Used for validation	# Trees	# Nodes
7377418	80%	20%	3	1000

Results

Accuracy	Elapsed Time
62.624%	4h 50m



THANKYOU

WSDM – KKBox's Music Recommendation Challenge

