

# 코드북

변수명	설명	그룹핑		
id	가온차트 데이터의 id			
son_id	가온차트 내에서 song_id(반복제외)			
rank	가온차트 랭킹(1~100)			
rank_g	가온차트 랭크를 10단위로. (1~10)			
wks	그해에 몇째주			
week	몇면도_몇째주			
sex		male	female	mixed
song_type	가요인지? ost 인지 등	가요=k_pop	pop,jpop=pop	ost=ost
genre	캐롤, 블루스, 댄스 등	댄스,락,랩/힙합,일렉트 로니카,캐롤,트로트= dance_etc	발라드,블루스/포 크,r&b/소울,인디,정통 = balad_etc	드라마,애니메이션/게 임,팝,해외영화,전체 =others
runtime_g	17400데이터에서 1 quan,3quan으로 잘라서 그룹핑	short	middle	long
active		solo	group	band
active_g		solo	nonsolo	
top_freq	이 기간동안 몇주동안 top에 진입했는지?			
top_freq_g	median전까지 low, 3quan 이상 high	low	middle	high
gg_trend	2020.05.10 updated 구글 트렌드 gg trend가 2019년 6월 2일부터 있음. gg_trend있는 데이터 에서 데이터가 없으면 0처리, 2019년 6월 2일 이전의 원래 없 는 데이터에는 null값 처리함.			
nv_trend	2020.05.10 updated 네이버 트렌드 gg trend가 2019년 5월 26일부터 2019년 12월 29일까지 있음. nv_trend있는 데이터에서 데이터가 없으면 0처리, 원래 없는 데이터에는 null값 처리함.			
dc 데이터	2020.05.10 updated dc 데이터 dc 데이터는 2018년 4월 22일부터 2020년 4월 19일까지 있 으므로 basic 데이터에 부족함 없음. dc 데이터에서 데이터 가 없으면 0처리함.  dc 데이터에서 가수명변경함 ;FT아일랜드는 FTISLAND로 변경, “걸그룹 여자친구”를 여자 친구로 변경,NCT는 NCT DREAM, NCT 127로 각각 변경하 여 추가, 김청하는 청하로  모든값의 mean을 구할때는 가수의 정보가 부족하면 null처리 하였다고 보면 됨.			

변수명		
id		
son_id		
rank		
rank_g		
wks		
week		
sex		
song_type	그외장르,재즈=others	
genre		
runtime_g		
active	project	duet
active_g		
top_freq		
top_freq_g		
gg_trend		
nv_trend		
dc 데이터		