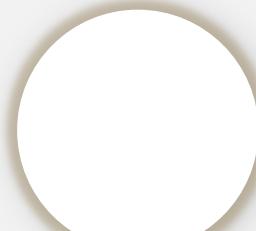


# MUSIC PLAYLIST



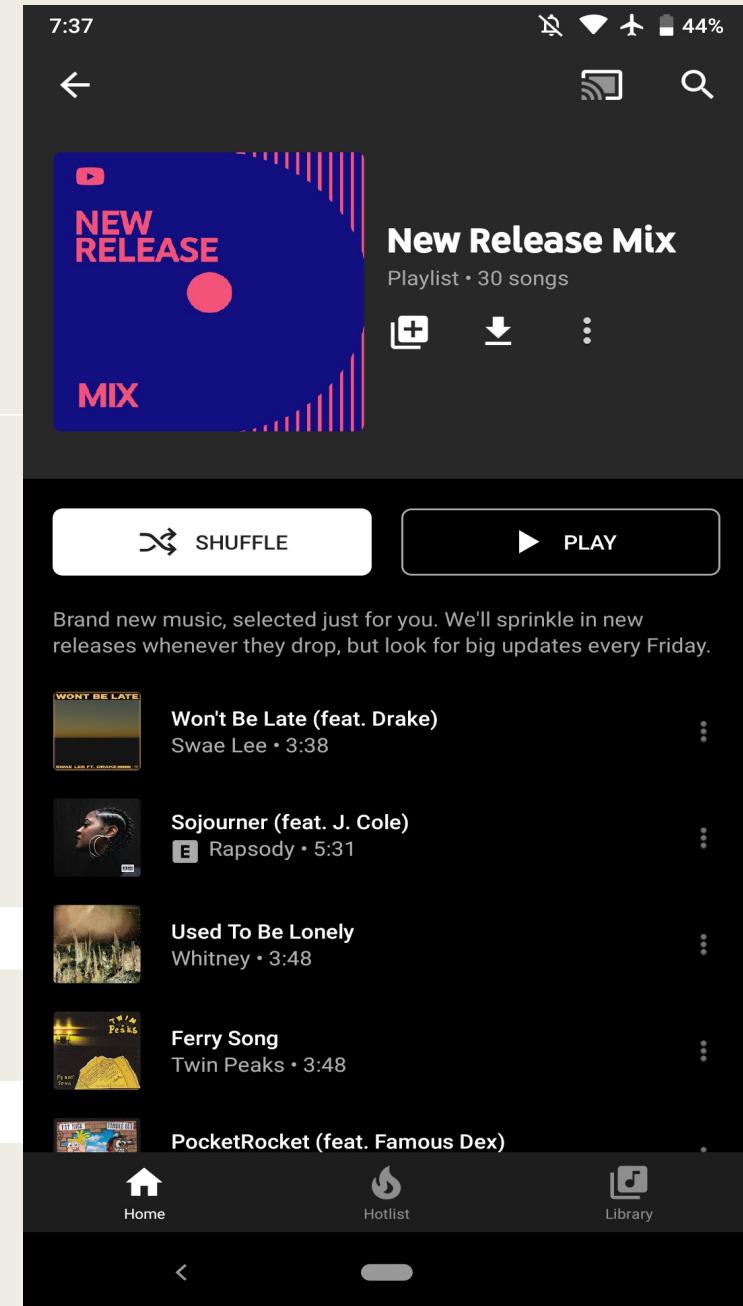
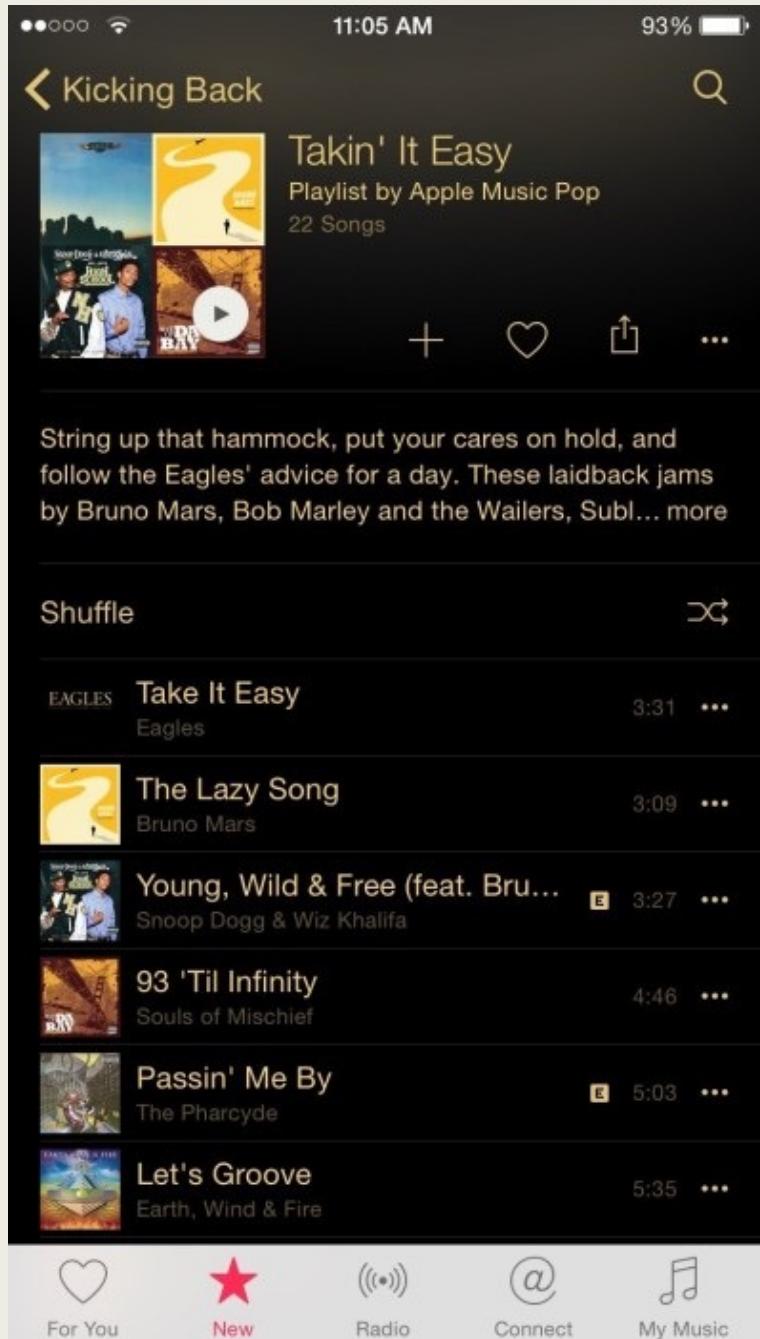
Weihao Ding  
Yiqing Fan

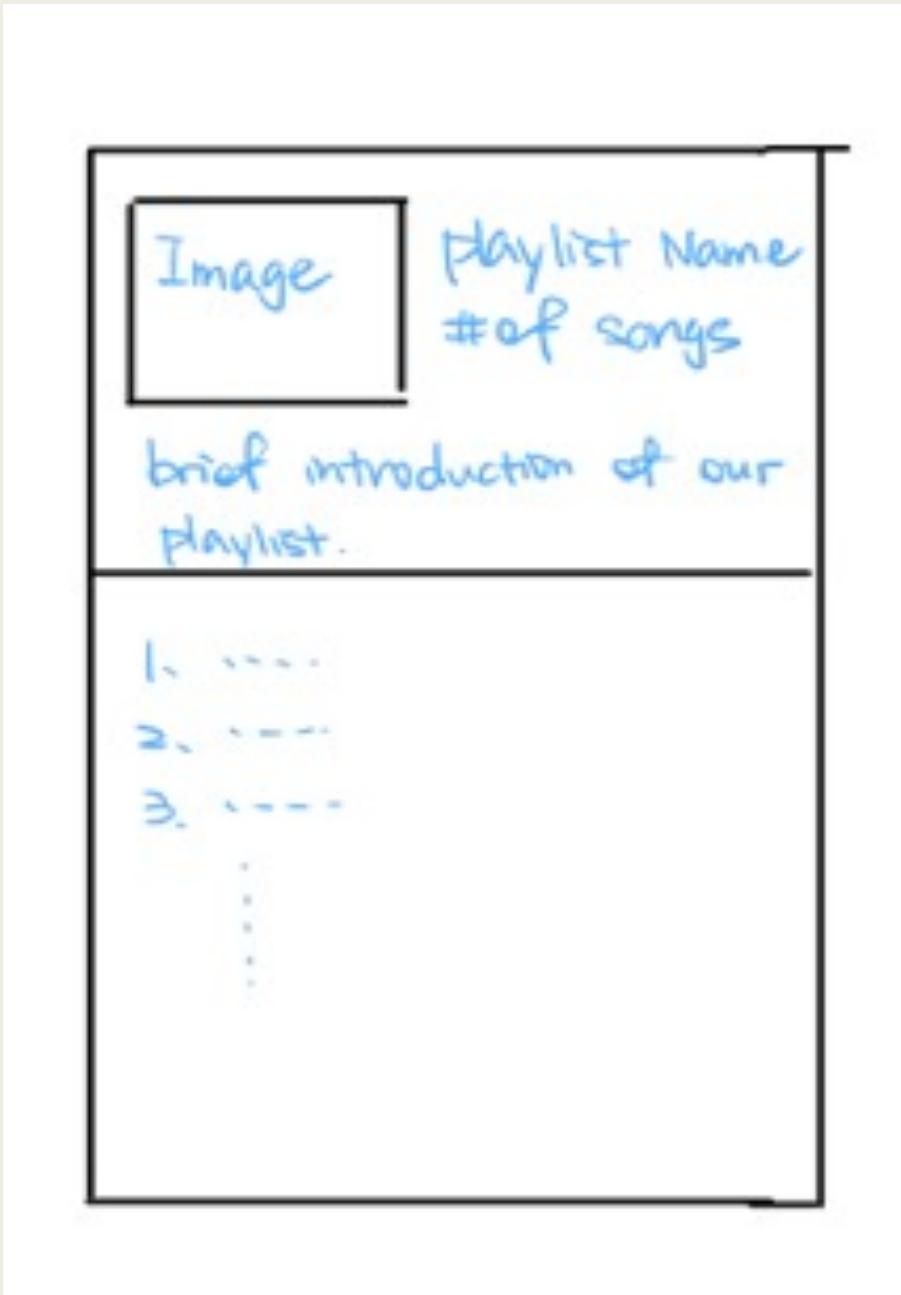


# Problem Statement

To create customized music playlists based on Amazon Product Database.

We're going to combine two datasets by their product ID and to make a few playlists based on product(music) categories and user reviews.





**Use key words to create playlist:**

Classic  
Cowboy

## **Dataset:**

### **Description:**

This dataset contains product reviews and metadata from amazon, including 142.8 million reviews spanning Mat 1996 – July 2014. This dataset includes reviews (ratings, text, helpfulness votes), product metadata (descriptions, category information, price, brand, and image features), and links (also viewed/also bought graphs).

### **Metadata:**

Metadata includes descriptions, price, sales-rank, brand info, and co-purchasing links.

# Reviews

- `reviewerID` - ID of the reviewer, e.g. [A2SUAM1J3GNN3B](#)
- `asin` - ID of the product, e.g. [0000013714](#)
- `reviewerName` - name of the reviewer
- `helpful` - helpfulness rating of the review, e.g. 2/3
- `reviewText` - text of the review
- `overall` - rating of the product
- `summary` - summary of the review
- `unixReviewTime` - time of the review (unix time)
- `reviewTime` - time of the review (raw)

# metadata

- `asin` - ID of the product, e.g. [0000031852](#)
- `title` - name of the product
- `price` - price in US dollars (at time of crawl)
- `imUrl` - url of the product image
- `related` - related products (also bought, also viewed, bought together, buy after viewing)
- `salesRank` - sales rank information
- `brand` - brand name
- `categories` - list of categories the product belongs to

# CONTENT

• 01

Pre-processing

• 02

Evaluation

• 03

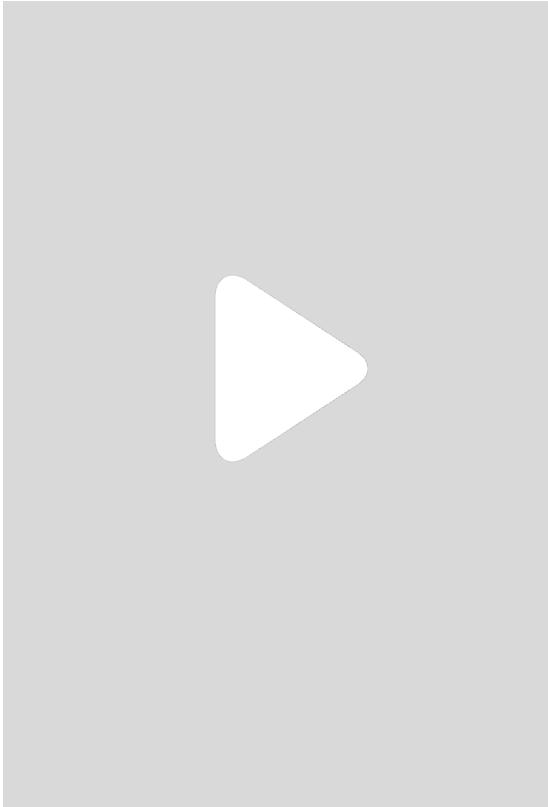
Making Playlists

• 04

Summary

# Part 01

**How to pre-processing data  
to get target output?**



## Sample review:

```
{  
    "reviewerID": "A2SUAM1J3GNN3B",  
    "asin": "0000013714",  
    "reviewerName": "J. McDonald",  
    "helpful": [2, 3],  
    "reviewText": "I bought this for my husband who plays the piano.  
He is having a wonderful time playing these old hymns. The music is  
at times hard to read because we think the book was published for  
singing from more than playing from. Great purchase though!",  
    "overall": 5.0,  
    "summary": "Heavenly Highway Hymns",  
    "unixReviewTime": 1252800000,  
    "reviewTime": "09 13, 2009"  
}
```

## Read json to CSV : got formatted CSV file

```
def main():
    fin = open("reviews_Digital_Music.json", "r")
    line = fin.readline()
    fout=open("reviews_Digital_Music.csv","a")
    #, reviewerName
    fout.write("reviewerID, asin, helpful, reviewText, overall, summary, unixReviewTime, reviewTime\n")
```

## Preprocessing data

Delete quotation

Unify string

Remove Space

Consistent Casing

```
str_=str_.replace("&quot;","");
str_=re.sub("[^a-zA-Z\s]+","",str_">#1. unify string
str_=str_.replace(" "," ")#2. remove unnecessary spacing
str_=str_.lower()#3. consistent casing
```

### Json file:

```
{"reviewerID": "A1WR23ER5HMAA9", "asin": "5555991584", "reviewerName": "AKB", "helpful": [1, 1], "reviewText": "Just when I thought Enya couldn't possibly get any better, I found \"The Memory of Trees\"; \"Trees\"; is Enya's best album, and like \"Watermark\"; and \"Shepherd Moons\";, each song is a gem. No other music conveys beauty, love, joy and peace the way Enya's does, and \"Trees\"; is no exception. Enya is a world treasure, and \"The Memory of Trees\"; is a must-have for anyone who loves her music.", "overall": 5.0, "summary": "How to improve upon perfection.", "unixReviewTime": 953424000, "reviewTime": "03 19, 2000"}
```

### Formatted CSV file:

reviewerID	asin	helpful
A1WR23ER5HMAA9	5555991584	[1 1]

reviewText
just when i thought enya couldnt possibly get any better i found the memory of trees trees is enyas best album and like watermark and shepherd moons each song is a gem i

overall	summary	unixReviewTime	reviewTime
5	How to improve upon perfection.	953424000	03 19 2000

## Calculate TF to select features:

Use TF: More weight is given to documents where the query terms appear many times

$$w_{d,t} = \begin{cases} 1 + \log_2 f_{d,t} & \text{if } f_{d,t} > 0 \\ 0 & \text{otherwise} \end{cases}$$
$$w_{q,t} = \begin{cases} \log(1 + \frac{N}{f_t}) & \text{if } f_{q,t} > 0 \\ 0 & \text{otherwise} \end{cases}$$

```
for i in range(0,maxi):
    strout=""
    for x in L:
        score=0
        if x in tok[i]:
            score=v[i][tok[i].index(x)]
        score=math.log(1+score,2)
        strout=strout+str(score)+","
```

# Outputs:

Cancel	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
	werID	asin	beautiful	singer	cd	the	ole	ahhh	days	good	memories	classic	coolio	rap	gangsta
3	A30T65W4IL	B0000004VV	0	0	1	0	0	0	0	0	0	0	0	0	0
4	A1I7110J3IM	B00000053X	0	0	0	1	1	1	1	1	0	0	0	0	0
5	A1WKTN9NJ	B000000D1V	0	0	0	0	0	0	0	0	1	0	0	0	0
6	A3RWOQAM	B000000HKV	0	0	0	0	0	0	0	0	0	1	1	1	1
7	A3EW0YLSU	B000000OM	0	0	0	0	0	0	0	0	0	0	0	0	0
8	A2OON1160	B0000000Q5	0	0	0	0	0	0	0	0	0	0	0	0	0
9	A1W63J18FC	B000000WC	0	0	0	1.5849625	0	0	0	0	0	0	0	0	0
10	A2ZF18JZGF	B000001A5X	0	0	0	1	0	0	0	0	0	0	0	0	0
11	ALNODPMRF	B000001A5X	0	0	0	1	0	0	0	0	0	0	0	0	0
12	A8Z8EJ7NLQ	B000001FAC	0	0	0	0	0	0	0	0	0	0	0	0	0
13	A39L5ZODCY	B00000255B	0	0	0	0	0	0	0	0	0	0	0	0	0
14	A3CYSZTOC1	B000002564	0	0	1	0	0	0	0	0	0	0	0	0	0
15	A35ZJDKR3Z	B00000259D	0	0	1	0	0	0	0	0	0	0	0	0	0
16	A25QXWTEA	B0000025CO	0	0	0	0	0	0	0	0	0	0	0	0	0
17	A1I72OXASR	B0000025LB	0	0	0	0	0	0	0	0	0	0	0	0	0
18	AHQ4BHT03	B0000025NL	0	0	0	1	0	0	0	0	0	0	0	0	0
19	A2RFKH1GZ1	B0000025RI	0	0	0	0	0	0	0	1	0	0	0	0	0
20	A2E0DRGIOX	B0000028RG	0	0	0	0	0	0	0	0	0	0	0	0	0
21	A15PG37TAI	B0000029J5	0	0	0	0	0	0	0	0	0	0	0	0	0
22	A201QRZNU	B000002G7C	0	0	0	0	0	0	0	0	0	0	0	0	0
23	A1640AMVN	B000002GYL	0	0	0	1	0	0	0	0	0	0	0	0	0
24	A3RT17EKJ1	B000002IJG	0	0	0	0	0	0	0	0	0	0	0	0	0

## Sort by the weight of Classic

reviewerID	asin	beautiful	singer	cd	the	ole	ahhh	days	good	memories	classic
A3DH84VE9AOM77	B0026LWUCC	0	0	0	0	0	0	0	0	0	1.58496
APM2F4D36644R	B000V66H3M	0	0	0	0	0	0	0	0	0	1.58496
A13WKCAR30ETEC	B00137RHT6	0	0	0	0	0	0	0	0	0	1.58496
A3NODD5SETTKE5	B001EW8BAO	0	0	0	0	0	0	0	0	0	1.58496
A3IHOMGAK8FIC8	B005LPG4O0	0	0	0	0	0	0	0	0	0	1.58496
AVC40H1AVADPA	B00171I5YO	1	0	0	0	0	0	0	0	0	1
A2K7TWCH1GZ3WX	B000W239UY	1	0	0	0	0	0	0	0	0	1
AR7N1ITXXCKAE	B00137YR9Y	1	0	0	0	0	0	0	0	0	1
A1KD6MXSOLFDVL	B004IAYI0G	1	0	0	0	0	0	0	0	0	1
ARZQ3K5LUIE99	B00937K4RA	1	0	0	0	0	0	0	0	0	1
A2KDJJP5283E7B	B00123FJUU	1	0	0	0	0	0	0	0	0	1
A19DU0YV4K2MRY	B001KVXJPG	1	0	0	1	0	0	0	0	0	1
A21B9HIVZVZK63	B001Q1OG1Q	1	0	0	0	0	0	0	0	0	1
A3RWOQAM6DXJ1K	B000000HKV	0	0	0	0	0	0	0	0	0	1
AC3P8IKGQ70IH	B000TDSRVM	0	0	0	0	0	0	0	0	0	1

2- 6 1.584962501

7-610 1

611-27302 0

## Classic Music Playlist

REVIEWER ID	PRODUCT ID
A3DH84VE9AOM77	B0026LWUCC
APM2F4D36644R	B000V66H3M
A13WKCAR30ETEC	B00137RHT6
A3NODD5SETTKE5	B001EW8BAO
A3IHOMGAK8FIC8	B005LPG400
AVC40H1AVADPA	B00171I5YO
A2K7TWCH1GZ3WX	B000W239UY
AR7N1ITXXCKAE	B00137YR9Y

## Sort by the weight of Cowboy

reviewerID	asin	singer	cd	the	beautiful	ole	ahhh	days	cowboy		
A2RARKYHW0YP6H	B000V66UV6	0	0	0	0	0	0	0	1	2 - 8	1
A3NQDOB129UC14	B000SZDI42	1	0	0	0	0	0	0	1	9 - 27302	0
A22G7FWP4SNOEC	B000QNU0RY	0	0	1	0	0	0	0	1		
A1TL721YECDIM8	B002IEW504	0	0	0	0	0	0	0	1		
AHE1GGVZMTSUW	B00123HIPY	0	0	1	0	0	0	0	1		
A1M4J5T1PQLQ4H	B00B13OPB0	0	0	0	0	0	0	0	1		
A1BHYIEB7TC7ZZ	B003NJXWKA	0	0	1	0	0	0	0	1		
A37R1RNA0PMLDG	B000Q3NYS6	0	0	0	0	0	0	0	0		
A1MJWRYG60QPQR	B004X97NYU	0	0	0	0	0	0	0	0		
A20RJXQWN251LX	B000EHQ80A	0	0	0	0	0	0	0	0		
AGI34T5QZWBBC	B000VMYM5Q	0	0	0	0	0	0	0	0		
A37FOMQPPIVLAV	B00124HEFM	0	0	0	0	0	0	0	0		
A3P991L6XH9G2J	B007XEHNUQ	0	0	0	0	0	0	0	0		
A35B1IURYB9EVM	B00E9OGMPA	0	0	0	0	0	0	0	0		
AH25EEU5QYRDW	B00GHJ6VME	0	0	0	0	0	0	0	0		
A3A4QHHS49KJWV	B00122C2QA	0	0	0	0	0	0	0	0		
A1L7JM06814O0B	B009Z8G390	0	0	0	0	0	0	0	0		
AJ2HAF14FZNKF	B00B9TPXE4	0	0	0	0	0	0	0	0		
A1T1FIGFIGXHZZ	B00GZ962A2	0	0	0	0	0	0	0	0		

## Cowboy Music Playlist

REVIEWER ID	PRODUCT ID
A2RARKYHW0YP6H	B000V66UV6
A3NQDOB129UC14	B000SZDI42
A22G7FWP4SNOEC	B000QNUORY
A1TL721YECDIM8	B002IEW504
AHE1GGVZMTSUW	B00123HIPY
A1M4J5T1PQLQ4H	B00B13OPB0

## Part 02

Evaluate our  
resulting playlist



# Classic Music Playlist

PRODUCT ID	AVERAGE RATING	TITLE
B0026LWUCC	5.0	Velvet Martini
B00137RHT6	5.0	A long time Comin
B001EW8BA0	5.0	Rodeo Waltz
B00171I5YO	4.875	Lift Up Your Hearts
B000W239UY	4.865	It's Christmas Time Again
B00137YR9Y	4.0	Slow Train Coming

```
In [21]: r6 = data[(data.iloc[:,1]=='B00171I5YO')]
ratings = r6.iloc[:,4]
ratings
```

```
Out[21]: 362912    5.0
362913    5.0
362914    5.0
362915    5.0
362916    5.0
362917    4.0
362918    5.0
362919    5.0
Name: overall, dtype: float64
```

```
In [22]: np.average(ratings)
```

```
Out[22]: 4.875
```

## Cowboy Music Playlist

PRODUCT ID	AVERAGE RATING	TITLE
B000V66UV6	4.0	Slippery When Wet
B000SZDI42	5.0	Deeper Still
B000QNUORY	5.0	American Cowboy
B002IEW504	4.596	Statesboro Blues
B00123HIPY	5.0	How Time Flies
B00B130PBO	4.909	

## Part 03

### Making our playlists



# Classic One:

```
今天:  
  
{'asin': 'B0026LWUCC', 'imUrl': 'http://ecx.images-amazon.com/images/I/51co1KUgeUL._SL500_AA280_.jpg', 'categories': [['Digital Music']]}  
{'asin': 'B000V66H3M', 'imUrl': 'http://ecx.images-amazon.com/images/I/51wfCwLFT1L._SL500_AA280_.jpg', 'related': {'also_bought': ['B000V619US', 'B000V62TES', 'B000VHKFZM', 'B002O4H6OI', 'B000V62STY', 'B00136PSXY', 'B000V64MRK', 'B000VZZ252', 'B000V62RXQ', 'B000V698DS', 'B00A1U2SVS', 'B001BKPV6G', 'B0065CCPYG', 'B000VRPGVA', 'B000V61A6G', 'B000V68LJ0', 'B000W25JUW', 'B000V698OC', 'B000W022JA', 'B001GNERHW', 'B0047B9GQM']}, 'categories': [['Digital Music']]}  
{'asin': 'B00137RHT6', 'imUrl': 'http://ecx.images-amazon.com/images/I/51Eu7nCc4gL._SL500_AA280_.jpg', 'related': {'also_bought': ['B001F67WUO', 'B000QO5DAC', 'B001VFYJPK', 'B002A715KK', 'B002FTD4YI']}, 'categories': [['Digital Music']]}  
{'asin': 'B001EW8BAO', 'imUrl': 'http://ecx.images-amazon.com/images/I/51AkXhfF2OL._SL500_AA280_.jpg', 'related': {'also_bought': ['B00897RX2O', 'B008BN7XA8']}, 'categories': [['Digital Music']]}  
{'asin': 'B005LPG4O0', 'price': 1.29, 'imUrl': 'http://ecx.images-amazon.com/images/I/417F-2gcc5L._SL500_AA280_.jpg', 'related': {'also_viewed': ['B000VZPJCI', 'B004RF4FP0', 'B000S57KVO', 'B00F5L159Y', 'B001EST2BA'], 'buy_after_viewing': ['B000VZO366', 'artist-red', 'B001O39AYY', 'artist-red', 'B001NCM8HM', 'artist-red', 'B000VZPJCI', 'artist-red']}, 'categories': [['Digital Music']]}  
{'asin': 'B00171I5YO', 'price': 0.99, 'imUrl': 'http://ecx.images-amazon.com/images/I/513jh1eyBkL._SL500_AA280_.jpg', 'related': {'also_bought': ['B005NWUXY8', 'B0060GB6FG', 'B000QVO7S4', 'B0011Z310Y', 'B005P37HQW', 'B0013G42KY', 'B0012ADS1U', 'B0081EA1JC', 'B002XSWNES', 'B002U4NEW0', 'B005MVHHFI', 'B001C155JW', 'B001GCIRUQ', 'B0019GS53S', 'B009P13UNY', 'B0011W1W5S', 'B001A818WK', 'B00136J7ZE', 'B0032PGJFU', 'B003B8N0DW', 'B002VOPASA', 'B000S4DDXY', 'B000S4K71A', 'B001225078', 'B003UBYWCI', 'B001MG4TJY', 'B000S4INAW', 'B002FN1HR0', 'B001451VK8', 'B00CDU9VWS', 'B0063JFMHS', 'B004ZLSBQU', 'B0097BO4SW', 'B000S4DDWK', 'B00122HHP6', 'B000W29AR0', 'B003AUEBN4', 'B0058ZGDL2', 'B000Z9KCRW', 'B0018GYD8K', 'B005P37AOQ', 'B00A95VWBW', 'B0029S38C8', 'B0049SEVPO', 'B005A0N04I', 'B006G8W6JS', 'B0013CPL1M', 'B001BGU412', 'B000VNXR2Y', 'B00GILXNA4', 'B003BS7PL0', 'B007HID19K', 'B0043VPW28', 'B0081EA0Q6', 'B00GANK53I', 'B00137ZN0Q', 'B001386Z6Q', 'B0011XDO4O', 'B002TZNPI', 'B0017RTE7A', 'B000QVL5PW', 'B001EUTW3G', 'B000V6NQ8Q', 'B005LYQBQ2', 'B007Z558NW', 'B004KEO1I4', 'B000QVOF9U', 'B002XSSVHQ', 'B00A0A6HSE', 'B003IR3SVK', 'B00123N6VO', 'B001AKI70E', 'B004UE3784', 'B000W1M500', 'B000Z9MSWY', 'B000ZOCEBO', 'B001SJH0VY', 'B0014O6CYW', 'B000W0AFDA', 'B001BKTVQM', 'B000TD7E4S'], 'buy_after_viewing': ['B0060GB6FG', 'artist-red', 'B005NWUXY8', 'artist-red', 'B0011Z310Y', 'artist-red', 'B004B2RSGM', 'artist-red']}, 'categories': [['Digital Music']]}  
{'asin': 'B000W239UY', 'price': 0.99, 'imUrl': 'http://ecx.images-amazon.com/images/I/51CpGswGtbL._SL500_AA280_.jpg', 'related': {'also_bought': ['B0013CPL1M', 'B000W22QSA', 'B0013G42KY', 'B000W20G6Y', 'B0011XDO4O', 'B0018PZ4UC', 'B000W238I2', 'B000W23IAU', 'B00136J7ZE', 'B001VW2ZS6', 'B00BNX8YSI', 'B00137TA14', 'B00137GB0M', 'B001H5OYXG', 'B000S52DRK', 'B003ER8SBY', 'B000W2388W', 'B000W271WQ', 'B0011Z5KBC', 'B001451VK8', 'B00137XHN6', 'B0011Z310Y', 'B007HID3JS', 'B00137KM3Y', 'B000W298WW', 'B003LG58GQ', 'B001O3K17C', 'B00122D6RE', 'B0047B4O50', 'B001U9UQ9U', 'B0013G1JZ0', 'B0011W1W5S', 'B003TUH0KQ', 'B002YSLBWM', 'B00BNX90RC', 'B001E54TMO', 'B00137OG54', 'B00EGIL7M2', 'B001GCIRUQ', 'B007F5JSJ2', 'B001JPFUFU', 'B000VK5H5C', 'B00AZ7TG94', 'B0013GL9E6', 'B001DDULQG', 'B00136NOVM', 'B001O3UHY6', 'B007V5XV30', 'B00136JO6G', 'B001NT4HW4', 'B00137OA0K', 'B00137Z19Y', 'B000V617PA', 'B001O03NRM', 'B000W0CNOO', 'B001BKTVQM', 'B001KRVX6C', 'B00137V0S0', 'B002RJ8VA8', 'B00157PBLU', 'B00137KRY5', 'B000S51XMQ', 'B001EC5N2S', 'B0011W3WRE', 'B0013GDXZO', 'B0063JUONK', 'B004P1Q8U6', 'B000TEBHU', 'B00656Q4NU', 'B00E91QA9Q', 'B0056D7EDW', 'B0014LSX3A', 'B004B2VUMO', 'B001VG0G0Q', 'B00B4N1V8C', 'B008Y8I6CY', 'B000W27428', 'B002TU1REW', 'B00DYMMDIS', 'B001KRO0US', 'B001O3R9P6', 'B000WLLOE8', 'B001DQH7OM', 'B000THFUHM', 'B000ZKFTRE'], 'buy_after_viewing': ['B003TUH0KQ', 'artist-red', 'B000W20FO2', 'artist-red', 'B0013G42KY', 'artist-red', 'B0013CPL1M', 'artist-red']}, 'categories': [['Digital Music']]}  
{'asin': 'B00137YR9Y', 'imUrl': 'http://ecx.images-amazon.com/images/I/51zCazEvVGL._SL500_AA280_PJStripe-Robin,TopLeft,0,0_.jpg', 'related': {'also_bought': ['B00137WUXE', 'B00137STG6', 'B00136PMV2', 'B00136JN94', 'B00137ML5Q', 'B00136NUUW', 'B00137YQVI', 'B00136JIIC', 'B00136PR6C', 'B00137V6LQ', 'B00GPFYQIQ', 'B00136RJS6', 'B00137XRHC', 'B00137YV2W', 'B00136Q4EQ', 'B001YXVOWA', 'B00137YV9A', 'B00FITJ70M', 'B00137X85I', 'B001YXLU2E', 'B002GO2U8S', 'B00137Z1QM', 'B00138CXJY', 'B00FITJQY4', 'B00136Q3S8', 'B00137TL3G', 'B00137TFZU', 'B00137IBQY', 'B00136NYR6', 'B005NNV0X0', 'B00138H1PK', 'B001YXP3CC', 'B00137QV1Q', 'B006645H1Q', 'B00137KT3M', 'B003LWJCGM', 'B00137R81S', 'B001UEZ816', 'B000W272FC', 'B00137URKW', 'B00137MTGC', 'B0016CTSQ8']}, 'categories': [['Digital Music']]}
```

# Cowboy

{'asin': 'B000V66UV6', 'price': 1.29, 'imUrl': 'http://ecx.images-amazon.com/images/I/61pAMpBJ9tL.\_SL500\_AA280\_.jpg', 'related': {'also\_bought': ['B000V62V9G', 'B000V66Z3Y', 'B000VZZ252', 'B000V66V5G', 'B000SXKWXY', 'B000V66ZPW', 'B008593MLC', 'B00IKNBLP2', 'B00FAXF4S0', 'B004A2QHWO', 'B001O03NKY', 'B000V62YY8', 'B000YMQIYG', 'B001AL64A8', 'B000V66VYC', 'B001BZOUQI', 'B0035MJERK', 'B00137GHF6', 'B001O8069M', 'B004A2U4XM', 'B000VRWW26', 'B004A2S4H0', 'B002DQLDPU', 'B001O4Q3N4', 'B00136PKKU', 'B000V62YSO', 'B004M4AKSW', 'B001OGROSA', 'B001NU4AA2', 'B002FVKL5G', 'B000V64X42', 'B004A2V3YQ', 'B002O4H6O1', 'B008593C94', 'B002JBBEV2', 'B000VRWVYK', 'B001B9DD8U', 'B000VC8VJO', 'B001NZZNOY', 'B001BZFEZY', 'B000V634TM', 'B000VRSXNI', 'B000W25C3Q', 'B00137MFXY', 'B000TDB4R6', 'B002FVMMJE', 'B00137IEJS', 'B000V62YT8', 'B000YMOQZO', 'B001B91KEO', 'B00136JL46', 'B00137X1FU', 'B001NZX8I2', 'B001NU2E50', 'B000W01746', 'B001NZ8ILO', 'B00124AJPO', 'B002CNZVJS', 'B001NZ76OE', 'B00DAPK7BY', 'B00136NVXI', 'B001NTVDUS', 'B001NYBVDW', 'B002JB9F7W', 'B00137IDT4', 'B00137KFLS', 'B001OB4TAG', 'B00L4RQZFC', 'B00123JYR4', 'B0011Z3G8G', 'B002FVTDHS', 'B001FEEYRK'], 'buy\_after\_viewing': ['B004A2X6DC', 'artist-red', 'B000V68ZM8', 'artist-red']}, 'categories': [['Digital Music']]}

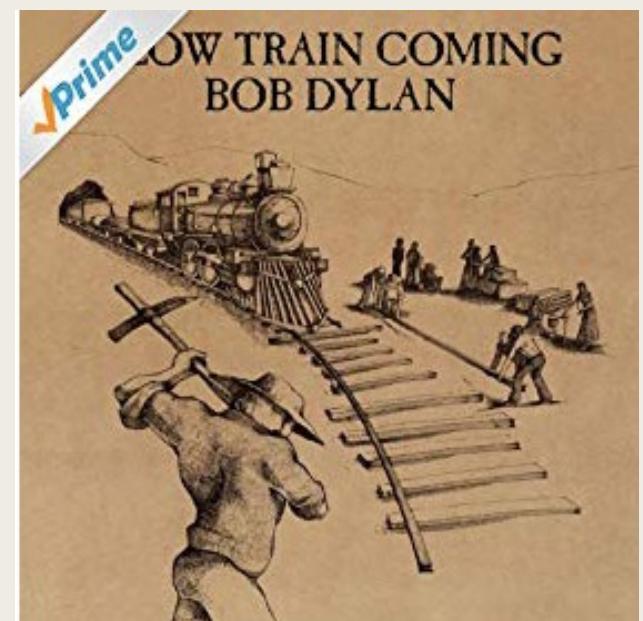
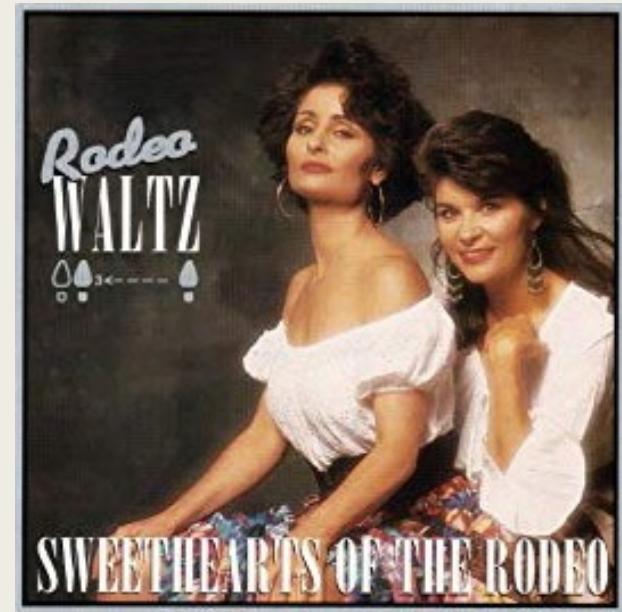
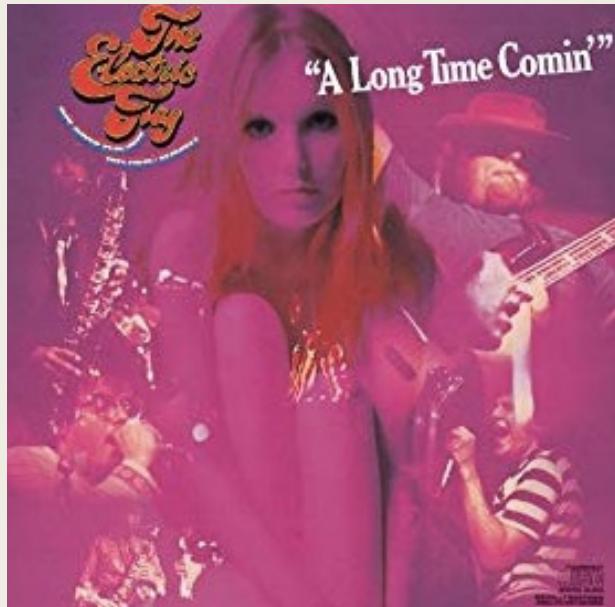
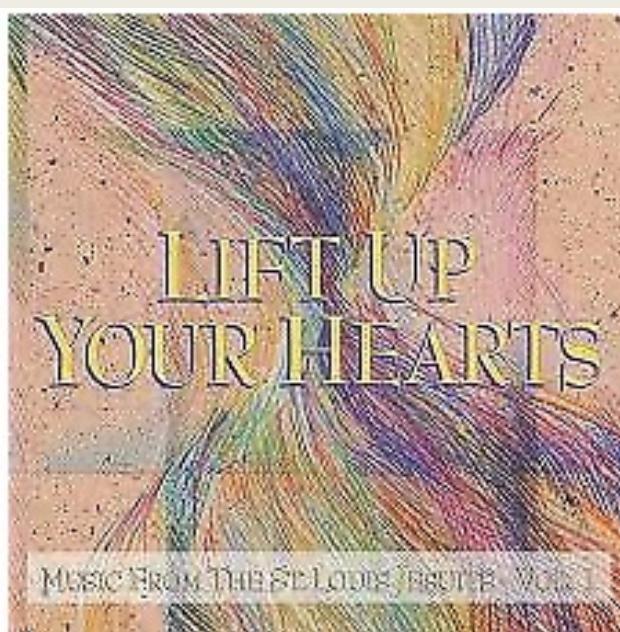
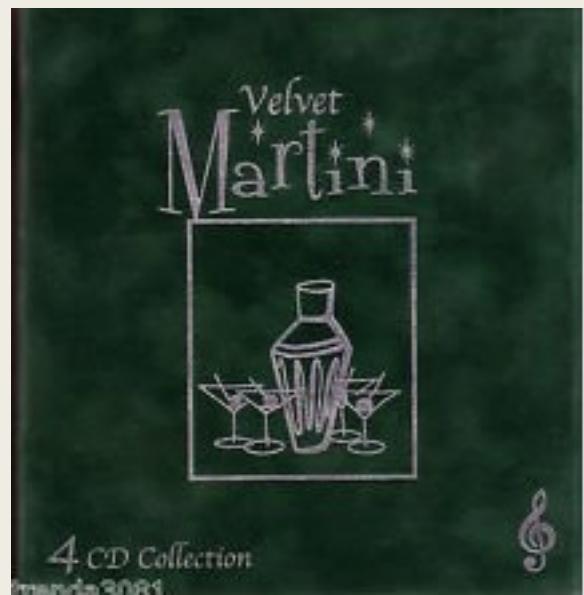
{'asin': 'B000SZDI42', 'imUrl': 'http://ecx.images-amazon.com/images/I/41qb10SBSwL.\_SL500\_AA280\_.jpg', 'related': {'also\_bought': ['B000TDCR5E']}, 'categories': [['Digital Music']]}

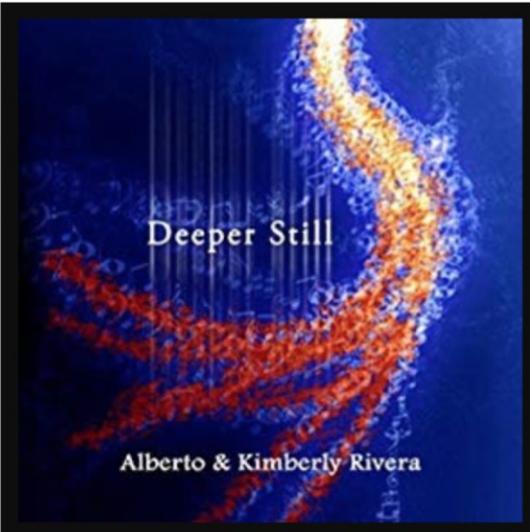
{'asin': 'B000QNU0RY', 'imUrl': 'http://ecx.images-amazon.com/images/I/51uaLkQjR-L.\_SL500\_AA280\_PJStripe-Robin,TopLeft,0,0\_.jpg', 'related': {'also\_bought': ['B00123JYR4', 'B001MCW6BG']}, 'categories': [['Digital Music']]}

{'asin': 'B002IEW504', 'price': 1.29, 'imUrl': 'http://ecx.images-amazon.com/images/I/51xpUUGQ8GL.\_SL500\_AA280\_.jpg', 'related': {'also\_bought': ['B00BKRLJG', 'B0013ASJP4', 'B000S51XMQ', 'B00CF0WSIU', 'B008KY78KI', 'B00920DUI8', 'B006W9T31K', 'B00B113E0U', 'B004SBW1LS', 'B002IV59P0', 'B00KDE46HQ', 'B00B9ZI5PC', 'B001D5FC7W', 'B006KYH2G0', 'B001CDNJWU', 'B00K5S089U', 'B004QRP53U', 'B0048W3SQ4', 'B0084O4CK8', 'B00D80EBFU', 'B005UWEKZE', 'B00658KCEA', 'B00F878NQI', 'B009Y6V44W', 'B002NQTG5Y', 'B002IERYH8', 'B0053ZFOYE', 'B001EUTW3G', 'B007QNBRBY', 'B009ZKNATY', 'B000S54MOC', 'B005IDAN38', 'B00D69AGAG', 'B00ISACZN4', 'B001YW9X8I', 'B006W9WGN2', 'B00BLICN3W', 'B0038AOXD4', 'B00FPZRJSG', 'B006OISTVY', 'B00J6155L8', 'B006L5O9GY', 'B00JAA56QY', 'B00136J7ZE', 'B008TMR676', 'B007QNRC3W', 'B006KYJLAK', 'B00658KCOA', 'B005UWEK8G', 'B0013G42KY', 'B00ELD37AC', 'B00HSSZW64', 'B001R5WZ28', 'B003UHO9PM', 'B007JCPELM', 'B002LBSIJG', 'B00EOHQE4Q', 'B00GM5Y5P8', 'B00GM5Y4YK', 'B00C5ZMMVI', 'B00JXPBE94', 'B00L4LENW0', 'B00AL5OQH2', 'B0048W3SRS', 'B001KW100W', 'B00ELD36SK', 'B007Y1ANKK', 'B00JS00AMQ', 'B00L4LEJY2', 'B00B9ZI3YA', 'B00C6MPIP2', 'B00FBCIYZU', 'B009ZAXVGQ', 'B00CEIGQLI', 'B00K4GRDGE', 'B00AIVEUJS', 'B006T2U1V6', 'B005Q7E0D0', 'B005UYZ5CO', 'B00ISAEAM8', 'B001UQPYTK', 'B00GIGYKFG', 'B009JDUXWY', 'B00L4LEHX0', 'B00B9TPXE4', 'B002Y4ADTI', 'B00C6MPHBC', 'B00I2VLJYA', 'B00DVV8KG6', 'B002PD7A7Q', 'B00JJD2LSI', 'B00K67U11A', 'B001AAAT7I', 'B00420HEF8', 'B0058U21N6', 'B007SQBS96', 'B00C6MPZE6', 'B00H2EBJSE', 'B004ZIH70E', 'B00713VD82'], 'buy\_after\_viewing': ['B000S51XMQ', 'artist-red', 'B00B9ZI5PC', 'artist-red', 'B00IOVH8AW', 'artist-red', 'B00CTTEKJW'], 'categories': [['Digital Music']]}

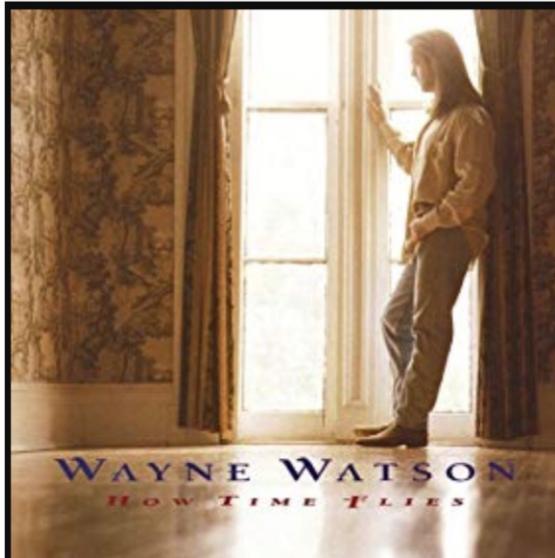
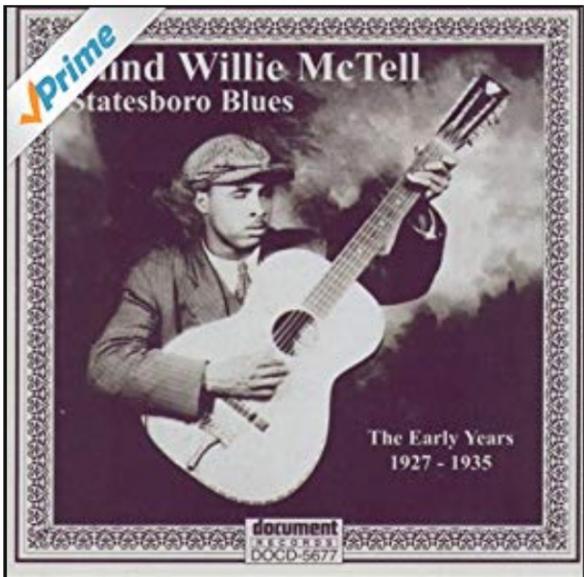
{'asin': 'B00123HIPY', 'price': 11.49, 'imUrl': 'http://ecx.images-amazon.com/images/I/51jLJ5iuE%2BL.\_SL500\_AA280\_.jpg', 'related': {'also\_bought': ['B00123FM84', 'B00123FBZI', 'B00K5S089U', 'B00123NLA0', 'B00123FQG2', 'B001I8YKKY', 'B00KDE46HQ', 'B00123HOTE', 'B004CA9406', 'B000TH94RE', 'B001Q1OIDW', 'B00123I23G', 'B000WD645A', 'B007QNRDD6', 'B00123F8Y2', 'B00123FQLM', 'B00137KG6W', 'B000TEC09Q', 'B001237CUK', 'B00137T0BE', 'B001INGJ3K', 'B00123HN9U', 'B000SZZBOC', 'B001INK3VY', 'B000VIP5TM', 'B00137O7EE', 'B005SZTF8A', 'B00123HFNY', 'B00123NHMW', 'B00124SGRC', 'B005F9VTJM', 'B00123N61Y', 'B0043VPW28', 'B001239EK6', 'B000TE738O', 'B002CJT2R4', 'B002KWBHFI', 'B001JTW0CM', 'B00137KL9Y', 'B007QNBRBY', 'B00CY56UE4', 'B00138061G', 'B002VHOISA', 'B0013805W6', 'B006W9WGN2', 'B00HSSZW64', 'B001ISZUTY', 'B001387A74', 'B0013855VW', 'B002FU1IAO', 'B00ARLEUHG', 'B009ZKNATY', 'B001OPPV7M', 'B000X6U6IM', 'B00123JT6U', 'B003IS3ZDK', 'B00122FVOU', 'B00136NXGS', 'B001KSS3YG', 'B00137PTPG', 'B00C5ZMMVI', 'B005UWEKZE', 'B000WY7WLY', 'B00123HNJA', 'B000S53WH0', 'B00137KJCI', 'B000VIR9NW', 'B0065CCPYG', 'B000TDSTOW', 'B001JTZB3M', 'B00BQ0KLNE', 'B001RJAHW', 'B00123NUO2', 'B0083EXE76', 'B00137KM3Y', 'B00137MHQ4', 'B00370CCGA', 'B00136J7ZE', 'B000TE38VA', 'B00136RG8Y', 'B000TDD5BE'], 'buy\_after\_viewing': ['B00123N9SY', 'artist-red', 'B002T00TSW', 'B00123FM84', 'artist-red'], 'categories': [['Digital Music']]}

{'asin': 'B00B13OPB0', 'price': 1.29, 'imUrl': 'http://ecx.images-amazon.com/images/I/41YOMa6Pj8L.\_SL500\_AA280\_.jpg', 'related': {'also\_bought': ['B001NZI54Y', 'B001NSAJCW', 'B00B6HWOSM', 'B00ENFIXGQ', 'B001NY9YGI', 'B00IL8182W', 'B00DH72Y7U', 'B0052SC6NO', 'B0045KYGG0', 'B00A5057EE', 'B001N37XXI', 'B004TMHJP4', 'B006G21UUA', 'B00I164HK4', 'B00KLKTZPK', 'B00E0FPF06', 'B00CMKRAVS', 'B00DE3QDKG', 'B002YSN364', 'B00K67U11A', 'B0047FOEFQ', 'B001NTEYJA', 'B001NSG25K', 'B009TZA6UE', 'B00D3R4RAG', 'B00BKJ5OD8', 'B00FGIHX9C', 'B00EQO0GB4', 'B00FENV9QW', 'B00FWFL2AU', 'B00IHWSM76', 'B001NY9YHW', 'B00KU5XPZM', 'B00K2RD9QI', 'B001NZ178K', 'B00IL4HYXS', 'B00KU5XRA0', 'B00B2LVYKQ', 'B00IQ42HQ2', 'B00JH1W452', 'B00137IJXO', 'B00E80JIH4', 'B001LJTCF8', 'B001NZGTRY', 'B0083WMA4Q', 'B000VZKR0W', 'B005W3ED3A', 'B00GBKIAQJ', 'B00L2FGLP0', 'B00CCUAKG0', 'B00DLPIGWA', 'B00LNP2KZY', 'B000W11HT0', 'B0052SC6DY', 'B00LAP6MAQ', 'B00K5BP5LI', 'B004XU084G', 'B000W177K8', 'B000W168H6', 'B0040IBFU2', 'B009E2PDGG', 'B00K67TXUK', 'B00K5RXRNU', 'B00367068G', 'B00KQPFQLW', 'B000VZTGMW', 'B00D80I1PQ', 'B00KGITY0I', 'B00CCUAHCM', 'B001NYGHTU', 'B001MYOE5K', 'B00JVOZDYO', 'B00D80I2PU']}





AmericanCowboy





# **Part 04**

**Future plan & what  
we have learned.**



- The importance of data cleaning
- The importance of baseline for evaluation

Challenging:

The evaluation part of our project is ambiguous. ....

Since our dataset is very large, we may need some optimization.

