Recommandation de musique

...

Carrard Antony, Ganguillet Anne Sophie, Maillefer Dalia, Killian Vervelle

9 juin 2023

Plan

de la présentation

- Million songs dataset
- Statistiques globales
- Genres musicaux
- Clustering musiques
- Clustering artistes
- Conclusion

Dataset

One Million Song

- Taille totale 280 Go
- Utilisation d'une sous partie
  - 1%, 1,8 Go

- Information sur les caractéristiques des chansons et des artistes

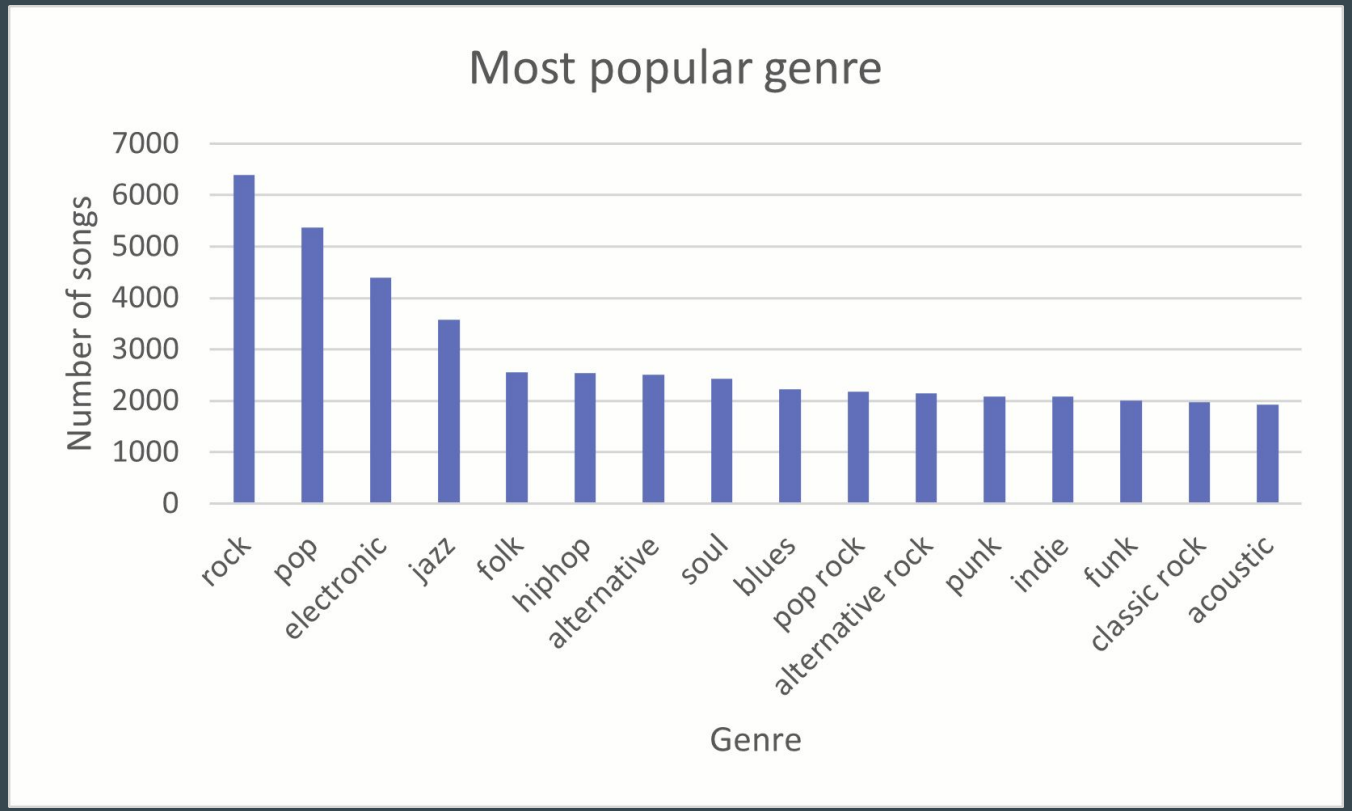
Field Name	Type	Description
artist hotttnesss	float	algorithmic estimation
artist id	string	Echo Nest ID
artist latitude	float	latitude
artist location	string	location name
artist longitude	float	longitude
artist name	string	artist name
beats confidence	array float	confidence measure
beats start	array float	result of beat tracking
duration	float	in seconds
energy	float	energy from listener point of view
key	int	key the song is in
key confidence	float	confidence measure
loudness	float	overall loudness in dB
mode	int	major or minor
mode confidence	float	confidence measure
release	string	album name
similar artists	array str	Echo Nest artist IDs (sim. unpublished)
song hotttnesss	float	algorithmic estimation
tempo	float	estimated tempo in BPM
time signature	int	estimate of number of beats per bar
time signature confidence	float	confidence measure
title	string	song title
year	int	song release year from MusicBrainz or 0

Questions

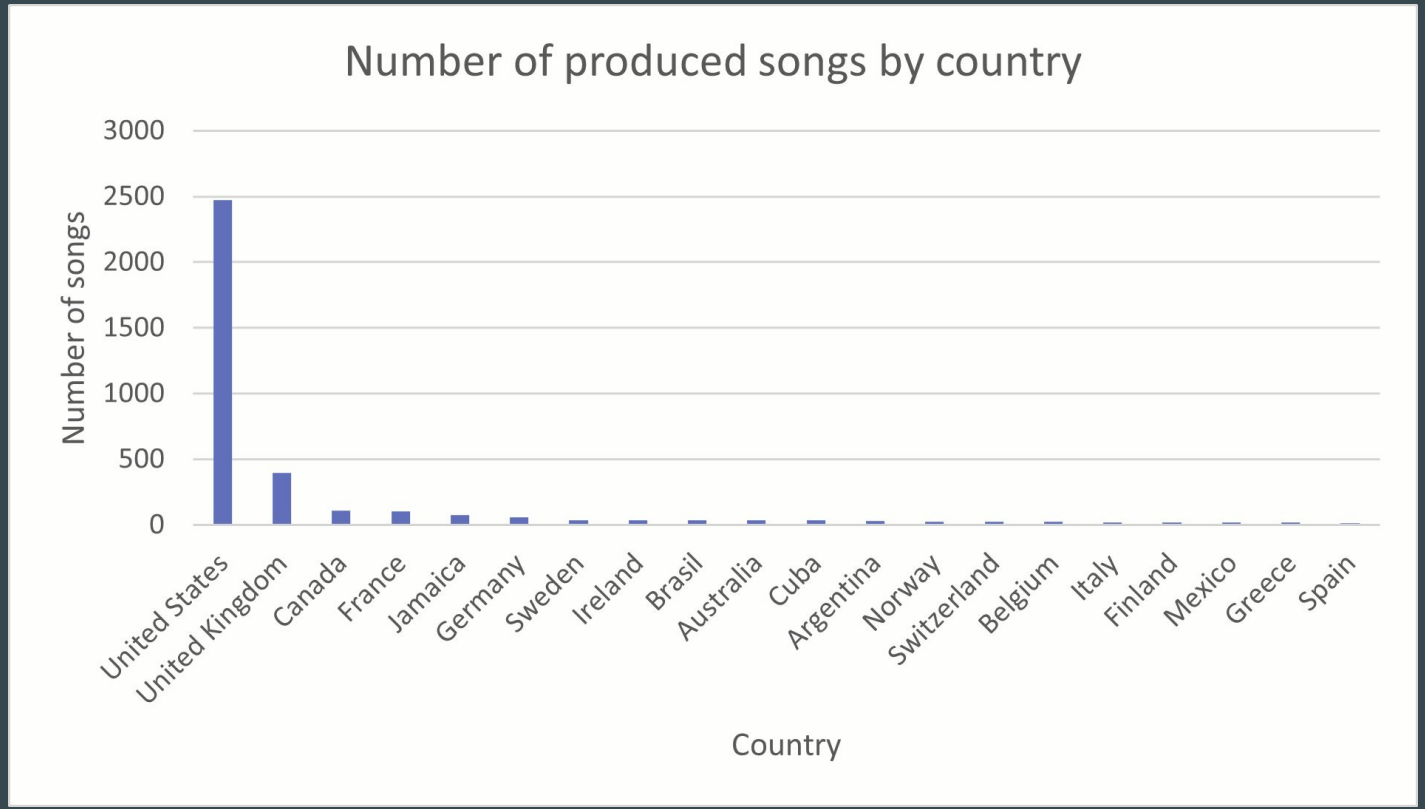
- **Question 1** : Quels sont les genres les plus populaires ? Quelle est l'année qui comptabilise le plus de chansons produites ? Quel pays détient le plus grand nombre d'artistes ?
- **Question 2** : Quel est le niveau sonore moyen et le BPM moyen (battement par minute) par genre musical ?
- **Question 3** : Comment prédire le genre musical d'une musique à partir des caractéristiques d'autres musiques (niveau sonore, tempo, gamme, durée) ?
- **Question 4** : Dans une optique de recommandation d'un artiste à un utilisateur, comment pourrait-on mesurer la similarité entre artistes ?

Réponses aux questions

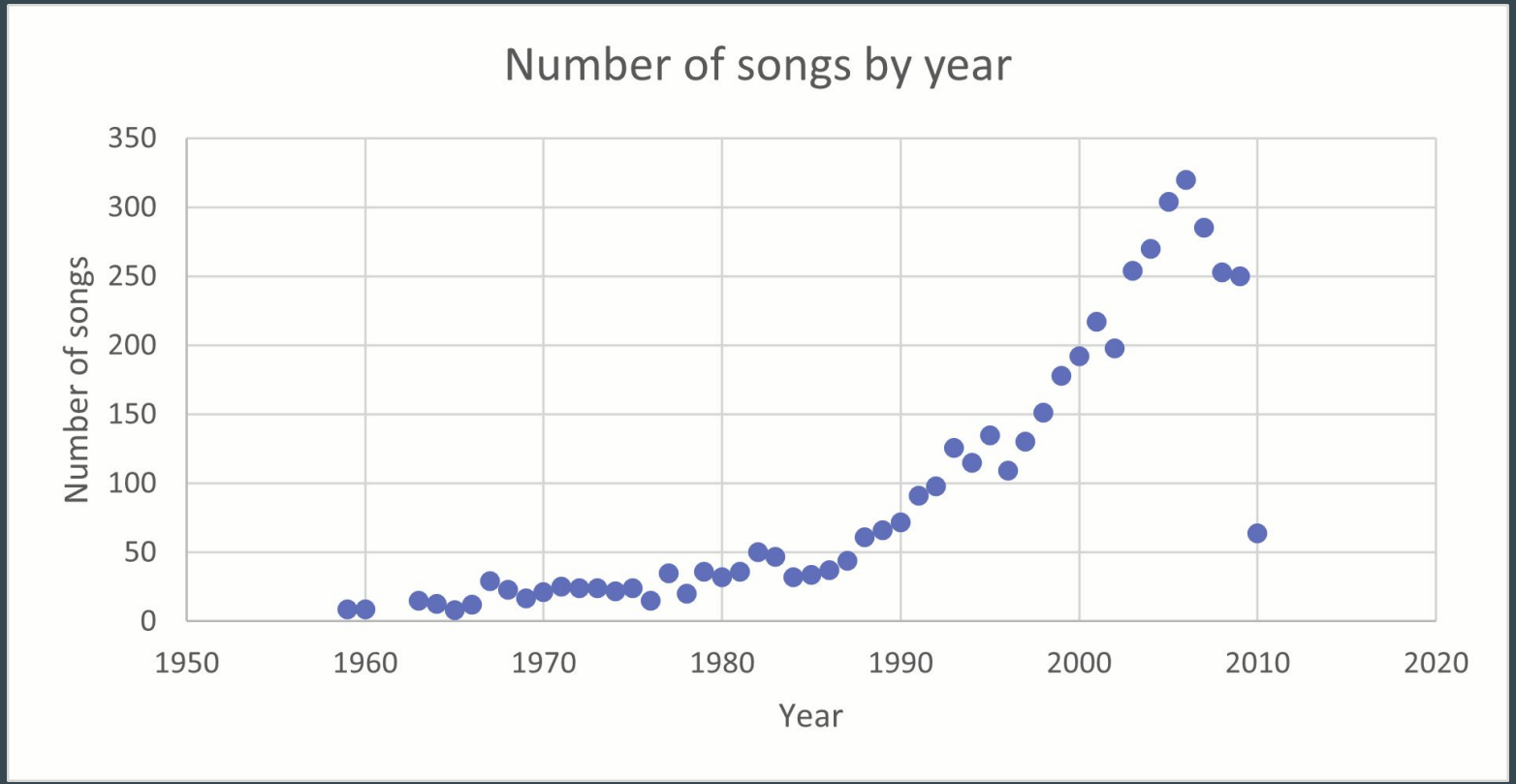
Statistiques globales



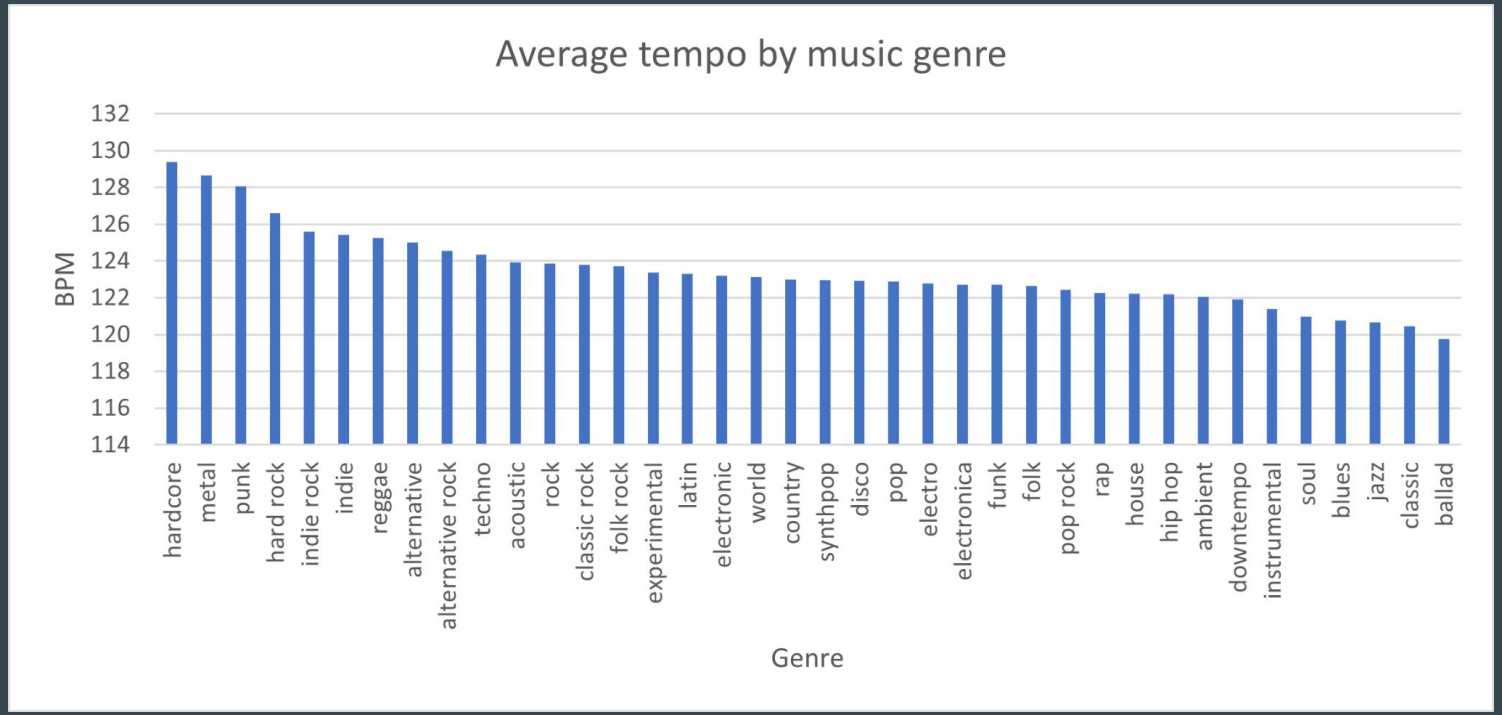
Statistiques globales



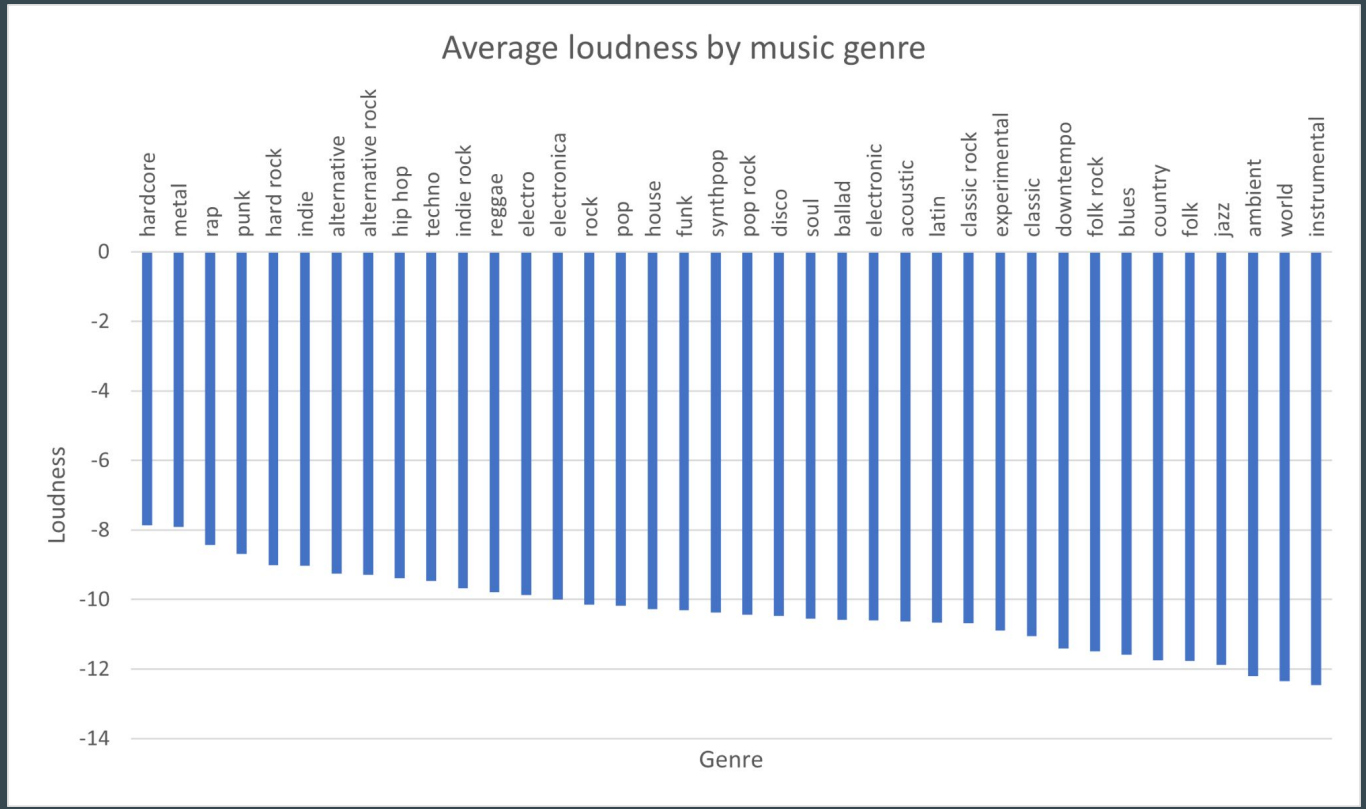
Statistiques globales



Statistiques par genres musicaux

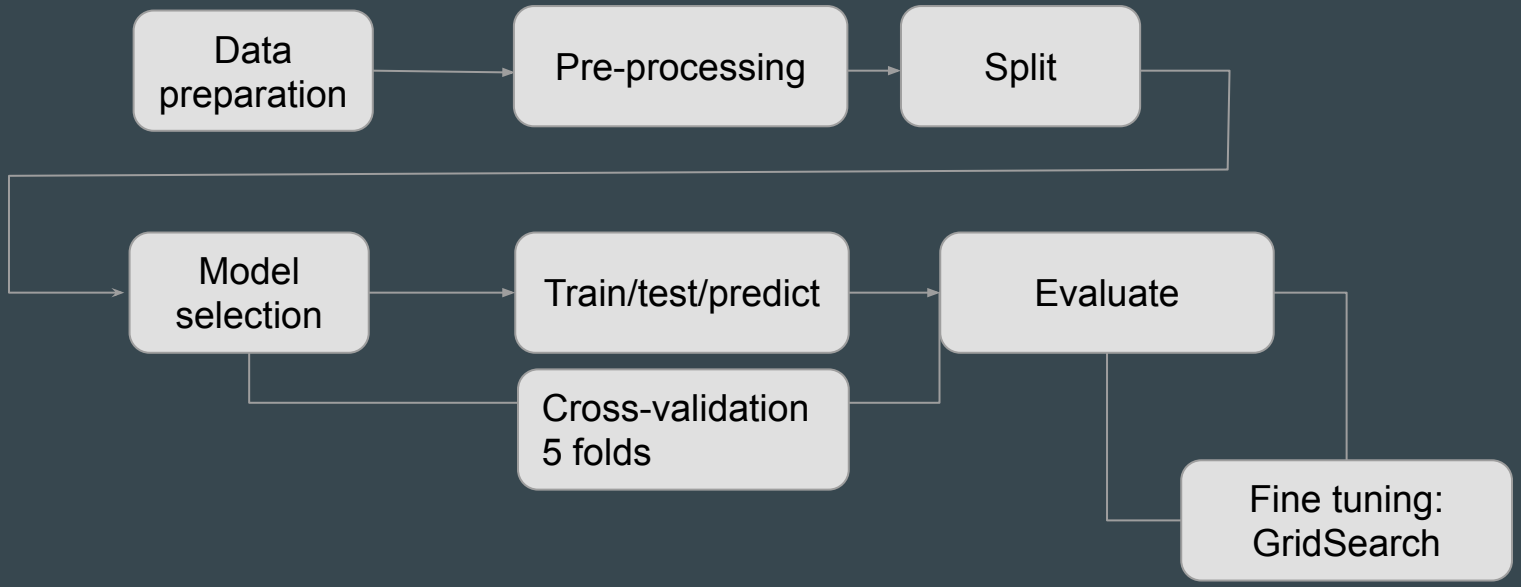


Statistiques par genres musicaux



Classification supervisée genre musicaux

Classification supervisée genre musicaux



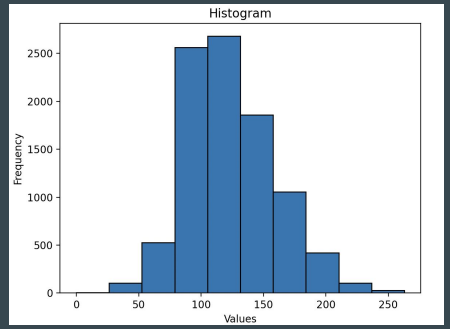
Data preparation - pre-processing

	tempo	loudness	time_signature	duration
count	10000.000000	10000.000000	10000.000000	10000.000000
mean	122.915449	-10.485668	3.564800	238.507518
std	35.184412	5.399788	1.266239	114.137514
min	0.000000	-51.643000	0.000000	1.044440
25%	96.965750	-13.163250	3.000000	176.032200
50%	120.161000	-9.380000	4.000000	223.059140
75%	144.013250	-6.532500	4.000000	276.375060
max	262.828000	0.566000	7.000000	1819.767710

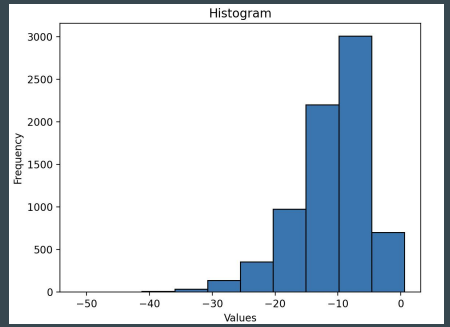
```

```

```
tempo               0
loudness            0
beats_start         0
time_signature      0
duration            0
artist_genre      155
```

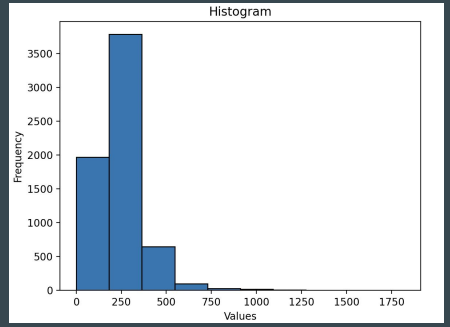
Tempo



Loudness



Duration



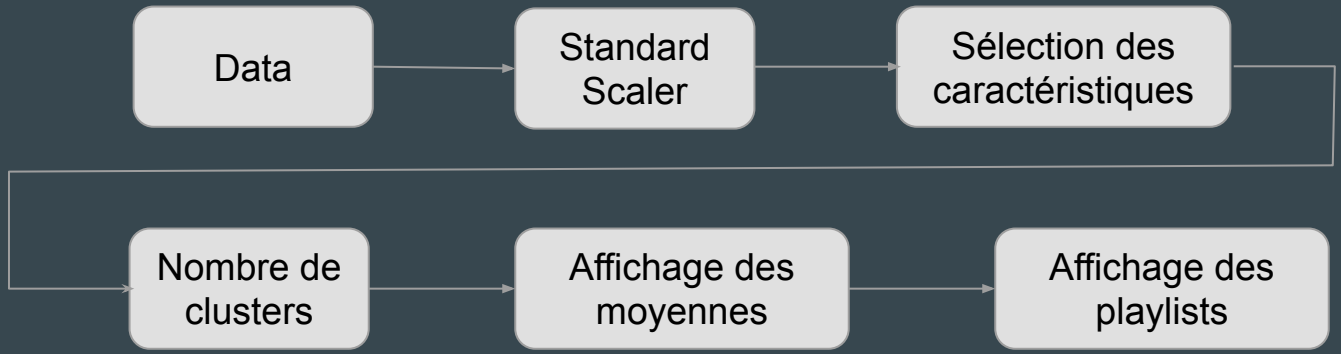
Model evaluation

	Random forest				Decision tree			MLP			
num trees	100	200	300	300				num layers	4, 10, 10, 5	4, 5, 4, 5	4, 10, 10, 5
max depth	5	10	20	20	max depth	5	20	max iter	50	100	50
max bins	32	32	32	32	max bins	32	32	solver	l-bfgs	adam	l-bfgs
sub sampled ra	0.8	0.8	0.8	0.8				step size	0.01	0.3	0.01
data split	80/20	80/20	80/20	80/20	data split	80/20	80/20	data split	80/20	80/20	80/20
preprocessing	oui	oui	non	oui	preprocessing	oui	oui	preprocessing	oui	oui	non
accuracy	**43.00%**	**44.00%**	**43,8%**	**44.70%**	**accuracy**	**42.40%**	**41.60%**	**accuracy**	**42.80%**	**41,1%**	**38,6%**

```
Best Model Parameters:
maxBins: 50
impurity: gini
maxDepth: 5
subsamplingRate: 0.8
numTrees: 200
```

Clustering genre musicaux

Clustering genres musicaux



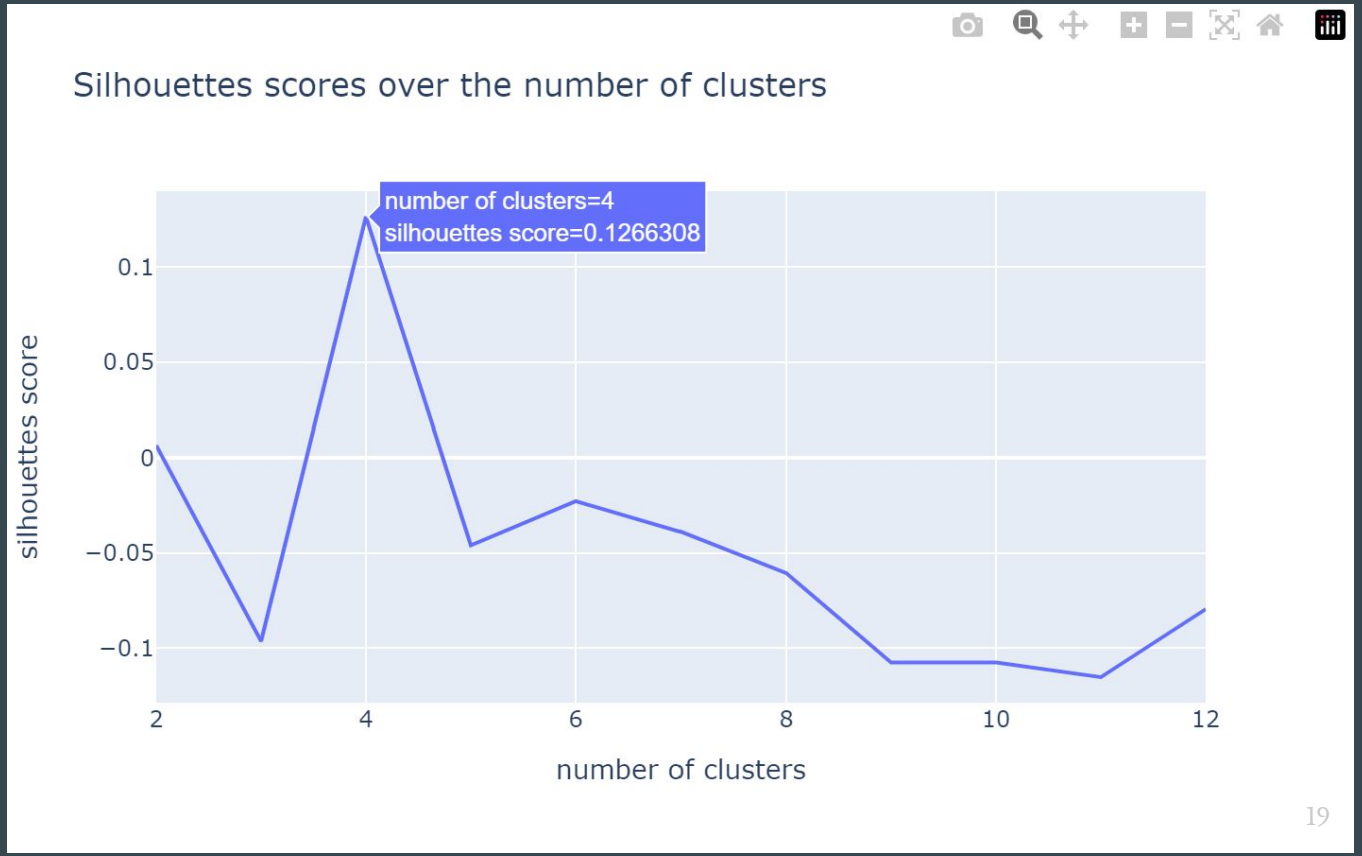
Caractéristiques sélectionnées

Selected data for k-means:

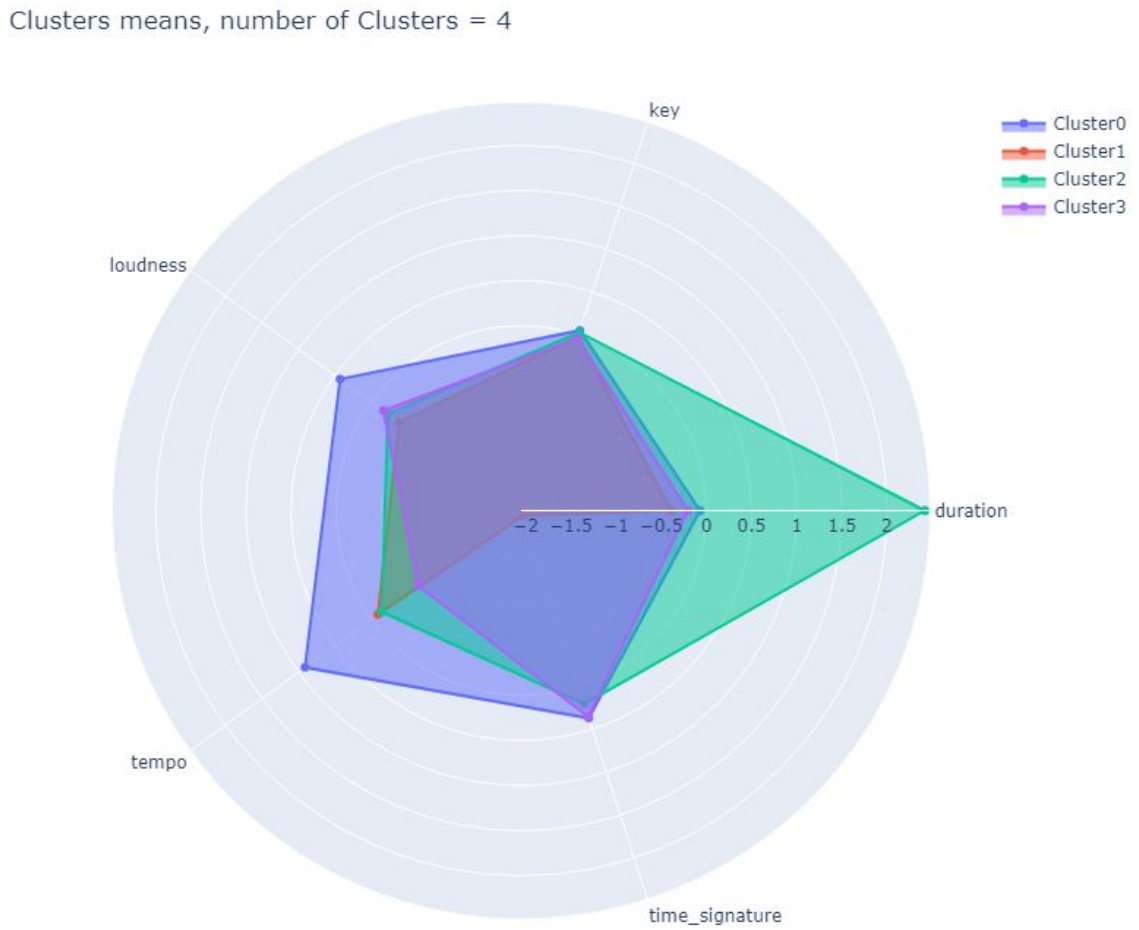
duration	key	loudness	tempo	time_signature
148.74077	0	-9.636	124.059	4
252.99546	1	-11.061	80.084	4
78.0273	3	-24.14	54.874	4
163.63057	7	-5.795	77.15	3
199.99302	10	-16.477	120.382	4
279.35302	9	-12.474	99.024	4
255.03302	9	-4.393	175.673	4
259.3171	1	-5.05	87.999	4
216.842	10	-4.264	92.897	4
312.99873	4	-13.885	86.981	5

Nombre de clusters

`(b-a)/max(a,b)`



Moyennes des clusters



Affichage des playlists

```
First 20 musics of cluster 0 :
+------------------------+-----------------------------------------------+---------+-------+-----------------------------------------------------------------------------------------------------+
|artist_name             |title                                          |duration |tempo  |artist_genre                                                                                         |
+------------------------+-----------------------------------------------+---------+-------+-----------------------------------------------------------------------------------------------------+
|Amorphis                |Misery Path (From the Privilege of Evil)       |255.03302|175.673|['Progressive metal', 'death metal', 'Melodic Death Metal', 'doom metal', 'seen live']|
|Atreyu                  |You Eclipsed By Me (Album Version)             |218.90567|157.715|['metalcore', 'hardcore', 'metal', 'screamo', 'rock']                                                |
|Spoonie Gee             |Spoonie Is Back                                |393.63873|135.503|['Hip-Hop', 'rap', '80s', 'hip hop', 'old school']                                                   |
|UFO                     |Out In The Street (Live) (2008 Digital Remaster)|314.17424|131.5  |['hard rock', 'classic rock', 'rock', 'heavy metal', 'Progressive rock']                              |
|Dave Hollister          |Calm Da Seas                                   |258.16771|117.936|['rnb', 'soul', 'Dave Hollister', 'r and b', 'gospel']                                               |
|Bob Marley & The Wailers|Rainbow Country (Red Gold And Green Version)   |258.29832|152.99 |['reggae', 'roots reggae', 'ska', 'roots', 'classic rock']                                           |
|Naseebo Lal             |Dholna Dholna                                  |376.16281|151.983|['Punjabi', 'Bhangra', 'Indian', 'folk', 'World Music']                                              |
|Jimmy Riley             |Amaze                                          |216.39791|159.943|['reggae', 'roots reggae', 'seen live', 'jamaica', 'jamaican']                                       |
|The Cortinas            |Radio Rape                                     |219.01016|134.985|['punk', 'punk rock', 'Punk 77', 'british', 'bristol']                                               |
|Jongo Trio              |Cavaleiro De Aruanda                           |157.72689|144.581|['Bossa Nova', 'brasil', 'jazz', 'mpb', 'easy listening']                                            |
|George Nooks            |TELL ME WHY                                    |221.90975|152.172|['reggae', 'lovers rock', 'dancehall', 'jamaica', 'George Nooks']                                    |
|HA-ASH                  |Amor a Medias                                  |253.90975|136.945|['pop', 'latin pop', 'latin', 'mexico', 'Espanol']                                                   |
|Radiohead               |15 Step                                        |237.21751|188.91 |['alternative', 'alternative rock', 'rock', 'indie', 'electronic']                                   |
|Bon Jovi                |Raise Your Hands                               |311.27465|139.95 |['rock', 'hard rock', 'classic rock', '80s', 'hair metal']                                           |
|John Holt               |I Need a Veggie                                |228.30975|131.297|['reggae', 'roots reggae', 'rocksteady', 'jamaica', 'roots']                                         |
|Capleton                |Cry For Love                                   |216.81587|137.425|['reggae', 'dancehall', 'ragga', 'jamaica', 'roots']                                                 |
|Kisha                   |Wohär dr Wind wäiht                            |203.04934|152.792|['swiss', 'seen live', 'Mundart', 'pop', 'switzerland']                                              |
|DJ Vix                  |Putt Jhatt Da Gulabi Phull Varga               |98.76853 |195.755|['Bhangra', 'Punjabi', 'Indian', 'Desi Artist', 'desi']                                              |
|Crematorium             |Unlearn                                        |210.1024 |122.186|['death metal', 'deathcore', 'black metal', 'metal', 'hardcore']                                     |
|Månegarm                |Vargbrodern Talar                              |92.76036 |163.086|['viking metal', 'folk metal', 'black metal', 'pagan metal', 'swedish']                              |
+------------------------+-----------------------------------------------+---------+-------+-----------------------------------------------------------------------------------------------------+
```

Comparaison des valeurs

Cluster 0

duration	tempo
255.03302	175.673
218.90567	157.715
393.63873	135.503
314.17424	131.5
258.16771	117.936
258.29832	152.99
376.16281	151.983
216.39791	159.943
219.01016	134.985
157.72689	144.581
221.90975	152.172
253.90975	136.945
237.21751	188.91
311.27465	139.95
228.30975	131.297
216.81587	137.425
203.04934	152.792
98.76853	195.755
210.1024	122.186
92.76036	163.086

Cluster 2

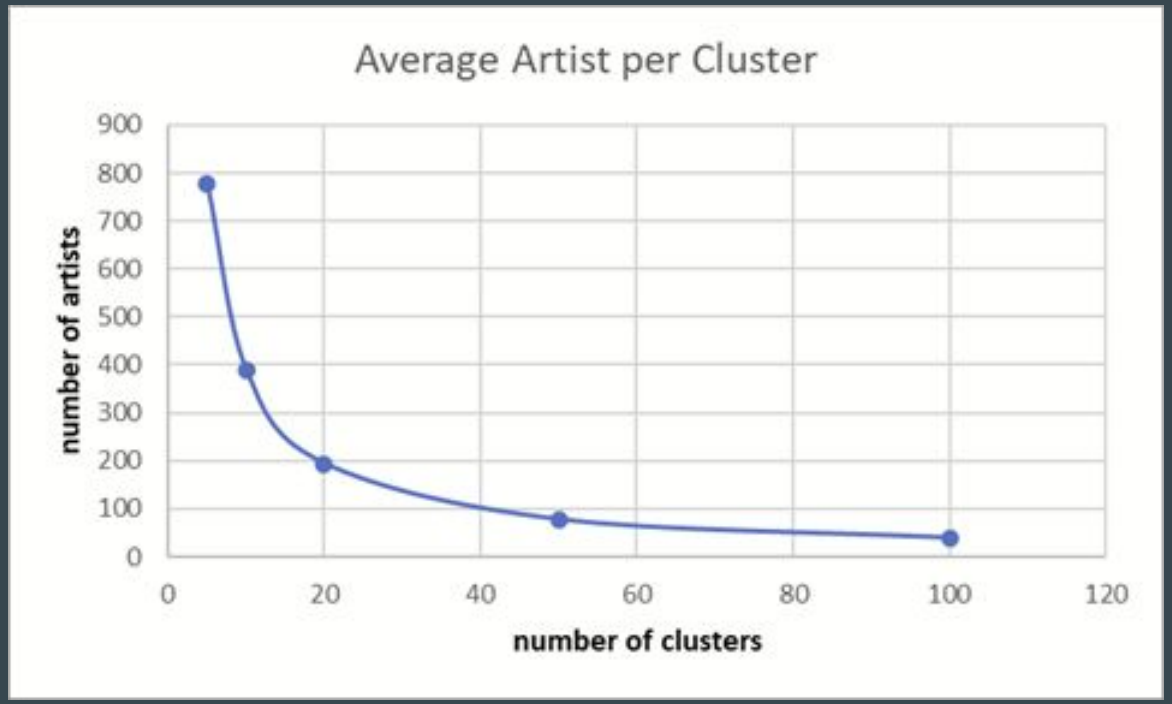
duration	tempo
580.70159	146.331
625.78893	89.572
528.22159	137.658
461.71383	138.512
424.82893	124.984
465.47546	131.999
381.23057	83.991
483.39546	70.256
600.11057	144.252
565.96853	0.0
532.27057	131.991
380.08118	131.999
482.21995	112.964
472.39791	121.518
770.35057	92.731
425.16853	120.006
485.14567	126.914
371.33016	110.909
486.97424	86.308
608.23465	127.996

Clustering artistes similaires

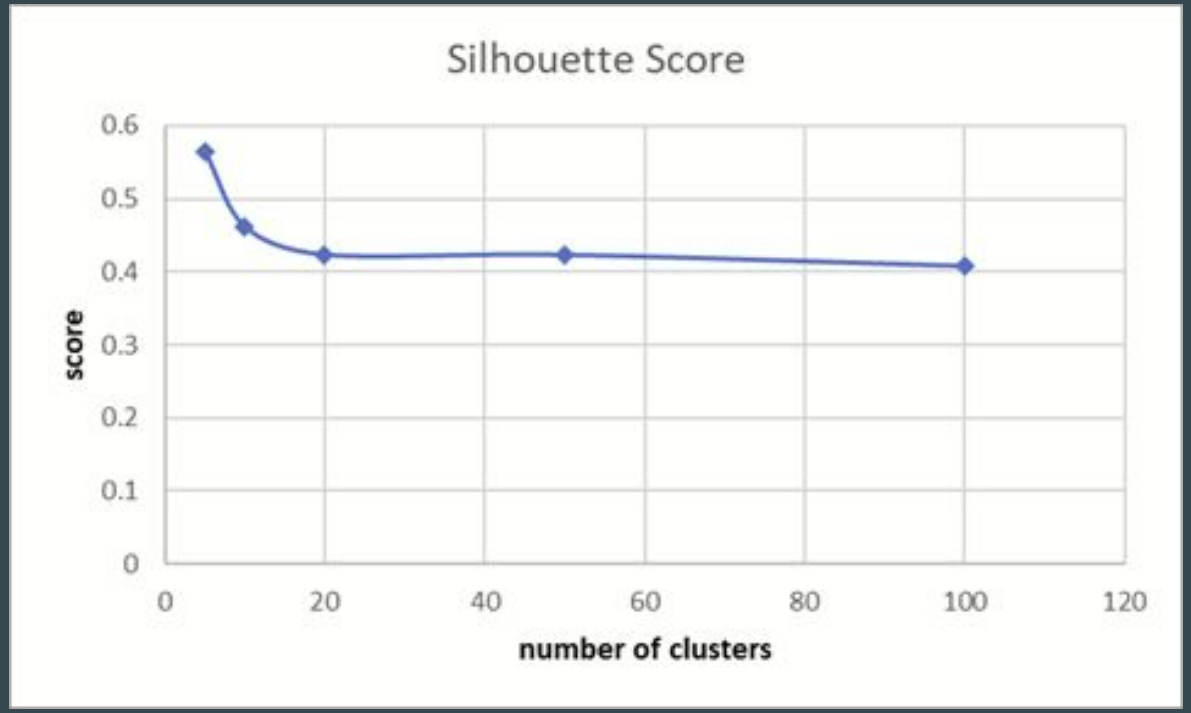
Clustering Artistes

1. Création features agrégées
2. Extraction artistes similaires
3. Preprocessing
4. K-Means pour différents nombres de clusters
5. Évaluations à l'aide de différentes métriques
    a. Nombres d'artistes par clusters
    b. Silhouette score
    c. Accuracy artistes similaires

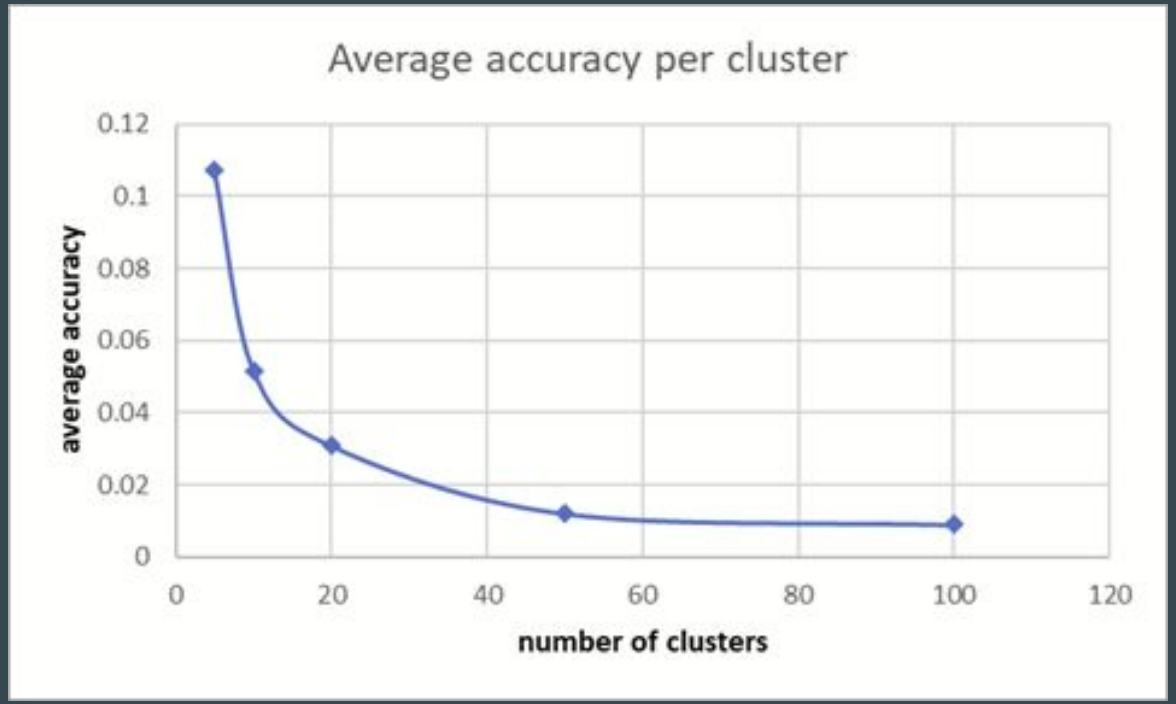
Clustering Artistes



Clustering Artistes



Clustering Artistes



Conclusion

Conclusion

- Clustering pas très accurate
- Possible problème dans la labellisation initiale

Améliorations :

- Utilisation d'un notebook Zeppelin
- Utilisation du dataset entier (280 Go)
- Ajout de features
- Meilleure sélection de features

Questions ?

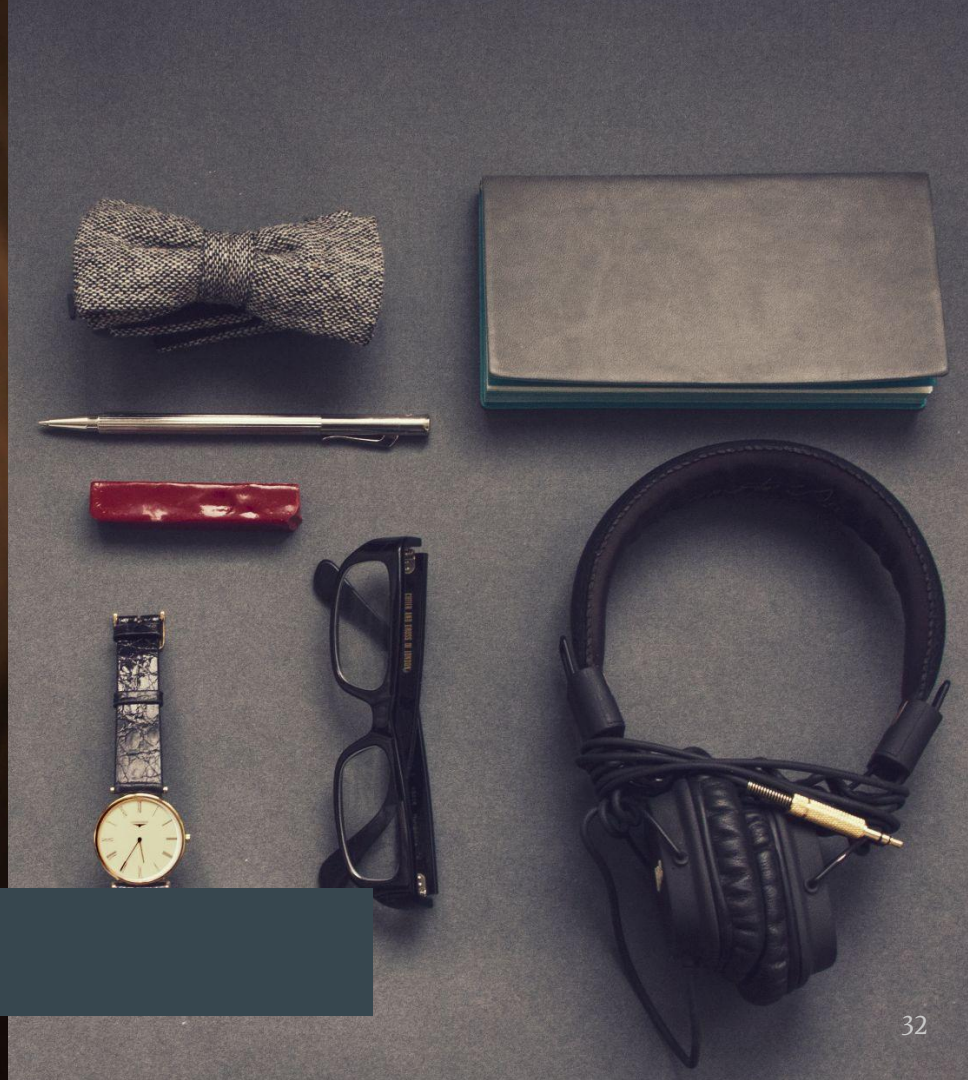
Hypothesis support

I think this is what's going to happen because...

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip.

Variables that may affect the outcome...

- Lorem ipsum dolor sit amet, consectetur adipiscing elit
- Sed do eiusmod tempor incididunt ut labore et dolore magna aliqua



The experiment

The Experiment

Tell the audience what you expect to happen...