

CloudComp - TP 1

César de Paula Morais - 2021031521

Philipe Dutra Cunha - 2021031734

Tarefa 1:

Duração das Músicas - Sem Tratamento (em ms)

Minimum	0
Mean	234408.54976216817
Maximum	10435467

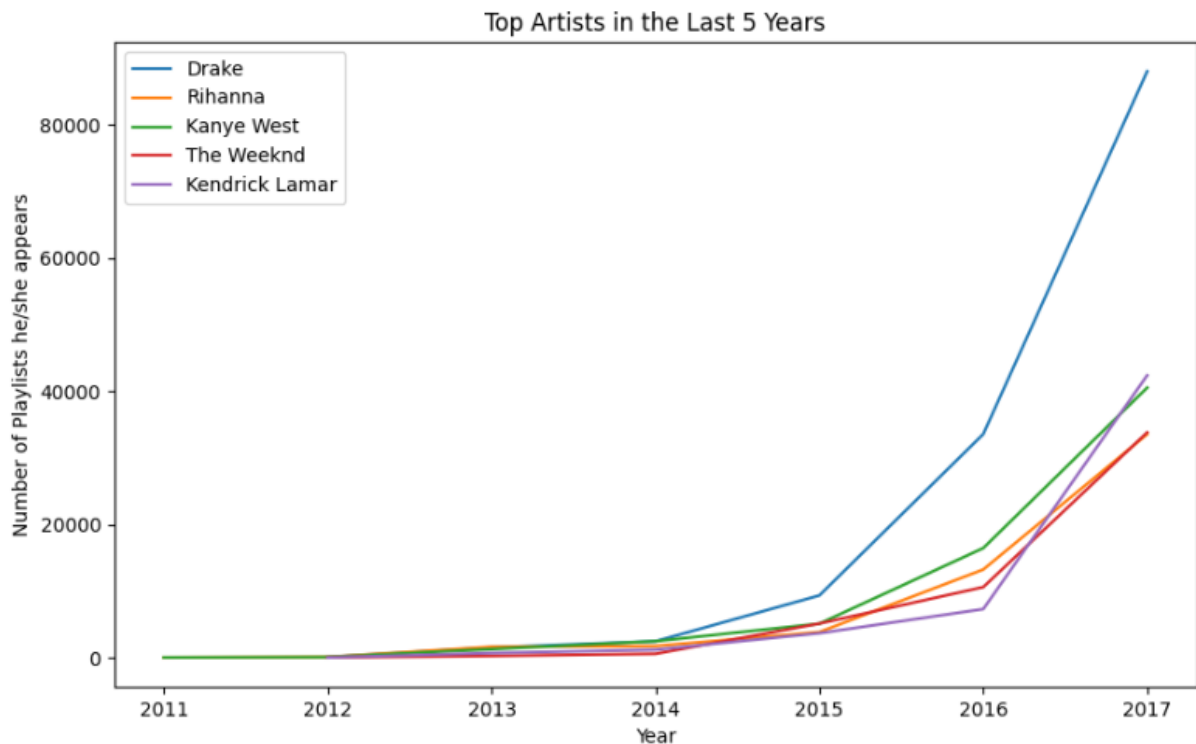
Duração das Músicas - Com Tratamento (em ms)

Minimum	107582
Mean	226899.35353939075
Maximum	349583

Foram agrupadas metadados sobre o dataset sem qualquer tratamento no primeiro gráfico e, no segundo gráfico, essas informações foram submetidas ao IQR (Interquartile Range Rule) que é basicamente um algoritmo para remover pontos fora da curva (outliers).

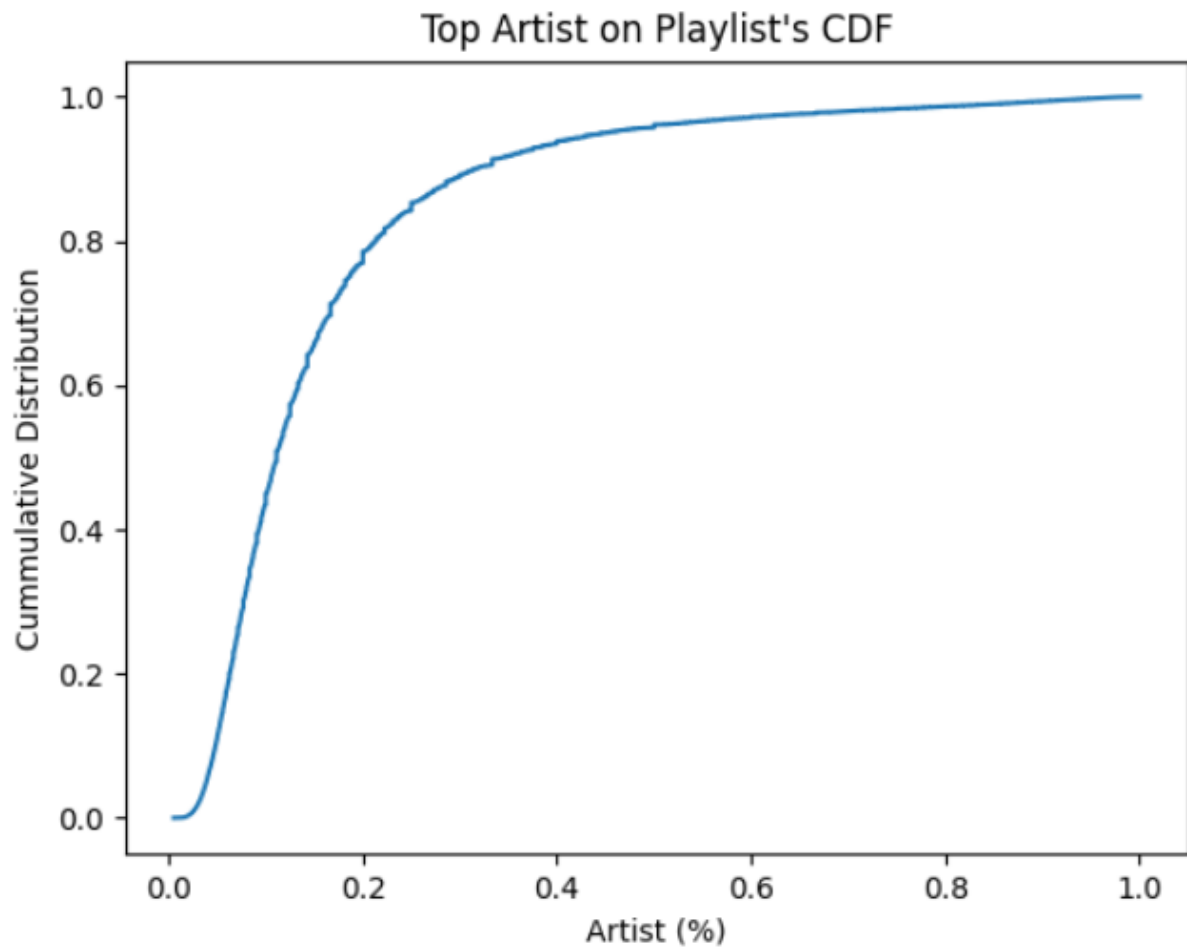
Para exemplificar os outliers, podemos pegar a duração mínima do primeiro gráfico que é de 0 ms, enquanto no gráfico 2 é de aproximadamente 1,8 minutos. É completamente descabido pensar em uma música que dure 0 ms, por outro lado, uma música durar 1,8 minutos é totalmente aceitável. Agora analisemos os máximos, no primeiro gráfico a duração convertida para horas resulta em aproximadamente 2,9 horas, e o segundo é aproximado em 5,8 minutos. Novamente, o outlier do primeiro gráfico é completamente absurdo, provavelmente se trata de um concerto, apesar disso, o valor tratado também não corresponde exatamente à realidade, já que existem músicas bem mais longas. Por fim, o valor da média do primeiro gráfico resulta em 3,9 minutos e o do segundo 3,8, ambos aproximados, ou seja, os outliers praticamente se anulam, nesse caso.

Tarefa 2:



De acordo com o gráfico, é possível notar que houve um crescimento considerável da popularidade dos top 5 artistas ao longo dos anos. Podemos, então, considerar duas hipóteses desse motivo: ou a popularidade da plataforma começou a crescer consideravelmente a partir de 2014 (ou será que a função de criar playlists não era tão usada?), ou tal crescimento diz respeito somente a esses artistas, ou seja, pode ser que houvessem outros artistas com números mais expressivos que foram eventualmente ultrapassados. Provavelmente tais resultados são consequência das duas hipóteses que apresentamos.

Tarefa 3:



Como visto no gráfico de CDF, temos um resultado que confirma uma hipótese que pode corroborar o senso comum: playlists geralmente têm grande variedade de artistas. Pelo gráfico, é evidente que, em aproximadamente 80% das playlists, temos 20% ou menos de músicas do artista mais popular nela. Novamente, isso deixa claro a preferência dos usuários de variar artistas, provavelmente em gêneros/estilos semelhantes.