

**UNIVERSIDADE TÉCNICA DE LISBOA**  
**INSTITUTO SUPERIOR DE ECONOMIA E GESTÃO**

DISSERTAÇÃO DE MESTRADO EM GESTÃO

**SEGMENTAÇÃO DE INDIVÍDUOS NO *FACEBOOK* QUE  
GOSTAM DE MÚSICA: ABORDAGEM EXPLORATÓRIA,  
RECORRENDO À COMPARAÇÃO ENTRE DOIS  
ALGORITMOS, *K-MEANS* E *FUZZY C-MEANS***

**José António Teixeira Quinteiro**

Orientador: Professor Doutor Jorge Júlio Landeiro Vaz

Júri: Professor Doutor José Augusto de Jesus Felício  
Professor Doutor Maria de Nazaré Rala Esparteiro Barroso  
Professor Doutor Jorge Júlio Landeiro de Vaz

Lisboa, Setembro 2011

## Resumo

Para se poder definir os melhores planos estratégicos, as decisões de marketing que se têm que tomar, com o intuito de abordar o mercado, escolher a melhor campanha publicitária, seleccionar o segmento e o tipo de produto ou serviço a oferecer, têm que ter por base o resultado de uma boa análise técnica da informação ou dos dados disponíveis.

A escolha do método de segmentação, é de primordial importância, pois os dados que se obtêm podem alterar a estratégia de selecção do mercado alvo e a estratégia de posicionamento dos produtos ou serviços, para além dos custos inerentes á tomada da decisão.

Este estudo procura encontrar diferenças entre dois métodos de segmentação descritivos post-hoc, (*k-means* e *Fuzzy C-Means*), na obtenção dos *clusters*, tendo por base a população portuguesa que gosta de música e que tem conta activa no *Facebook*.

No âmbito deste trabalho realizou-se uma revisão da literatura conhecida tendo-se efectuado a segmentação da amostra obtida através de dois algoritmos. Complementou-se o estudo com uma análise descritiva das frequências de modo, aquisição e audição dos vários tipos de música.

Palavras-chave: segmentação, cluster, k-means, Fuzzy C-Means, Facebook, marketing

## **Abstract**

In order to define the best strategic plans, marketing decisions that have to be taken in order to tackle the market, choose the best advertising campaign, select the thread and the type of product or service to offer, they have to be based on the result of a good technical analysis of available data or information.

The choice of segmentation method is of paramount importance, since the data obtained may change the target market selection and the strategy of placement of products or services, in addition to the costs related to taking the decision.

This study seeks to find differences between two methods of descriptive post-hoc segmentation (k-means clustering and Fuzzy C-Means clustering), in obtaining of clusters, based on the Portuguese population who likes music and have an active account on Facebook.

This work there was a review of the literature known followed by the segmentation of the sample obtained through two algorithms. These were complemented with a descriptive analysis of usage situations, acquisition and hearing of various types of music.

**Keywords:** segmentation, cluster, k-means, fuzzy c-means, Facebook, marketing

## Índice

Resumo .....	2
Abstract .....	3
Agradecimentos .....	7
Lista de Figuras .....	8
Lista de Quadros .....	8
Lista de Abreviaturas e Siglas .....	9
1. Introdução.....	10
2. Revisão Bibliográfica .....	12
2.1. Segmentação .....	12
2.1.1. “ <i>Segmentação de Mercado</i> ” – Origem do Conceito.....	13
2.1.2. Níveis de Segmentação .....	14
2.1.3. Critérios de Segmentação .....	16
2.1.4. Bases de Segmentação .....	18
2.1.4.1. Bases Gerais Observáveis .....	19
2.1.4.1.1. Geográficas.....	19
2.1.4.1.2. Demográficas .....	20
2.1.4.1.3. Socioeconómicas .....	21
2.1.4.2. Bases Gerais Não Observáveis .....	23
2.1.4.3. Bases Específicas do Produto Observáveis.....	26
2.1.4.4. Bases Específicas do Produto Não Observáveis.....	28
2.1.4.5. Conclusão .....	28
2.1.5. Métodos de Segmentação.....	31
2.1.5.1. Métodos de Segmentação <i>a Priori</i> Descritivos .....	32
2.1.5.2. Métodos de Segmentação <i>A Priori</i> Preditivos .....	33
2.1.5.3. Métodos de Segmentação <i>Post Hoc</i> Descritivos.....	34
2.1.5.4. Métodos de Segmentação <i>Post Hoc</i> preditivos .....	36
2.1.5.5. Conclusão .....	39
2.2. Segmentação, <i>Targeting</i> e Posicionamento .....	39
2.2.1. <i>Targeting</i> .....	40
2.2.2. Posicionamento.....	41

2.3.	“Clustering” .....	41
2.3.1.	Os dados .....	42
2.3.2.	Os “Clusters” .....	43
2.3.3.	Similaridade e Distâncias .....	45
2.3.4.	As Partições dos “Clusters” .....	46
2.3.4.1.	Partição “Hard (crisp)” .....	46
2.3.4.2.	Partição “Fuzzy” .....	48
2.3.5.	Algoritmos de “Clustering” .....	49
2.3.5.1.	O Algoritmo <i>K-means</i> .....	49
3.	<i>Facebook</i> .....	54
4.	Metodologia de Trabalho .....	58
4.1.	Processo de Obtenção dos Dados .....	58
4.2.	Questionário .....	59
5.	Resultados Obtidos .....	60
5.1.	Recolha dos Questionários .....	60
5.2.	Pré-Tratamento dos Dados .....	61
5.3.	Tratamento dos Dados .....	61
5.3.1.	Redução de Dados .....	61
5.3.2.	<i>Outliers</i> .....	62
5.3.3.	Análise aos Resultados do PCA .....	62
5.3.5.	Processo de Segmentação .....	65
5.3.5.1.	Segmentação <i>K-means</i> .....	65
5.3.5.2.	Segmentação <i>Fuzzy</i> .....	66
6.	Análise dos Resultados .....	67
6.1.	<i>K-means</i> .....	67
6.2.	<i>FCM</i> .....	69
7.	Conclusões .....	74
7.1.	Comparação e Conclusões .....	74
7.2.	Limitações do Trabalho .....	75
7.3.	Trabalhos Futuros .....	76
8.	Anexos .....	86
	Anexo 1: Procedimento para obtenção do público-alvo no Facebook. ....	86

Anexo 2: Questionário .....	87
Anexo3: Frequências idade, Sexo, Estado civil e Escolaridade .....	90
Anexo 4: Outliers – “caixa de bigodes” .....	91
Anexo 5: Rotated Component Matrix .....	93
Anexo 6: Comunalidades .....	95
Anexo 7: Valores próprios e variância explicada.....	97
Anexo 8: Quadro resumo após PCA .....	99
Anexo 9: Determinação do número de clusters – Kmeans .....	100
Anexo 10: Análise discriminante.....	101
Anexo 11: Determinação do Número de Clusters em Fuzzy .....	102
Anexo 12: Gráfico final de Kmeans.....	103
Anexo 13: Gráfico final Fuzzy .....	104
Anexo 14: Frequência de modos, aquisição e audição de música com os clusters .....	105
Anexo 15: Caracterização dos clusters (variáveis demográficas, K-means).....	111
Anexo 16: Caracterização dos clusters (variáveis demográficas, FCM) .....	112
Anexo 17: Lista completa dos graus de pertença .....	113

## **Agradecimentos**

A realização deste trabalho contou com o apoio, aconselhamento e disponibilidade de diversas pessoas e não queria perder a oportunidade de agradecer de uma forma muito especial.

Quero agradecer em particular ao meu orientador, Prof. Doutor Jorge Landeiro Vaz, que aceitou ser meu orientador neste trabalho e que sempre me ajudou na orientação, em particular na sua disponibilidade em dispor do seu tempo para me ajudar. Quero salientar o seu carácter humano que tanto me cativou no desenrolar deste trabalho.

Ao Prof. Rui Brites por me ter ajudado no desenvolvimento nos capítulos do tratamento de dados e também em SPSS.

Ao Prof. Doutor João Dias que baseado nos seus vastos conhecimentos em estatística, me consegui guiar através deste trabalho.

À minha família, em particular à minha mulher Carla e aos meus filhos João e Francisco que ajudaram nos momentos mais difíceis, pela compreensão das ausências mais do que muitas, na realização deste trabalho, enfim por me terem acompanhado durante toda esta caminhada.

## Lista de Figuras

Figura 1: Níveis de segmentação de mercado .....	15
Figura 2: Classificação dos métodos de análise classificatória .....	34
Figura 3: Diferentes tipos de <i>clusters</i> em $\mathbb{R}^2$ .....	44

## Lista de Quadros

Quadro 1: Classificação das bases de segmentação .....	19
Quadro 2: Base de segmentação observável geográfica .....	20
Quadro 3: Base de segmentação observável demográfica .....	21
Quadro 4: <i>The National Statistics Socio-economics Classification</i> (NS-SEC) .....	22
Quadro 5: Base de segmentação, observável socioeconómica .....	23
Quadro 6: Avaliação das bases de segmentação .....	29
Quadro 7: Classificação do método usado na segmentação .....	31
Quadro 8: Avaliação dos métodos de segmentação .....	39
Quadro 9: Graus de pertença após o algoritmo FCM .....	69



## **Lista de Abreviaturas e Siglas**

ACORN – A Classification of Residential Neighborhoods

ADCLUS – Additive CLUstering

AID – Automatic Interaction Detection

AIO – Activities, Interests and Opinions

CART – Classification and Regression Trees

FCL – Fuzzy c-lines

FCM – Fuzzy c-means

FCR – Fuzzy Clusterwise Regression

GoM – Fuzzy Grade of Membership Model

LOV – List of Values

NS-SEC –National Statistics Socio-economic Classification

PCA – Principal Components Analysis

PRIZM – Potential Rating Index for ZIP Markets

RVS – Rokeach Value Survey

SPSS - Statistical Package for the Social Sciences

VALS - Values and Lifestyle Segmentation

## 1. Introdução

A internet cada vez mais está presente na vida das pessoas, tendo sido um importante factor de desenvolvimento tecnológico ao longo dos últimos anos. Tem revolucionado a forma de se obter informações, de se fazer negócios e também de se relacionar.

As redes sociais têm cada vez mais, sido um motor para o crescimento do modo de relacionamento actual através da internet. Em parte a popularidade das redes sociais, deve-se à facilidade em ultrapassar as barreiras geográficas, já que possibilita a comunicação entre as diversas partes do planeta, a qualquer hora do dia ou da noite, sendo apenas necessário para se efectuar, um computador com acesso à internet.

O acesso à informação é de tal forma actualmente facilitada que uma grande quantidade de informação é gerada e conseqüentemente partilhada entre os diversos intervenientes. No entanto, a criação e partilha da informação, é de pouca ou nenhuma utilidade, caso a mesma não resulte em conhecimento que se revele potencial gerador de vantagem para o seu interlocutor. Numa era em que a quantidade de informação é muito grande, há que criar e gerir ferramentas que potenciem a distinção entre o útil e o essencial do que é desperdício ou desnecessário para determinado fim ou respectivo interlocutor.

Conhecer os consumidores, as suas características, quais são os mais rentáveis, que padrões de comportamento assumem e que preferências apresentam, é fundamental para as empresas melhor se adaptarem a essas exigências. A tecnologia actual permite-nos recorrer a muitas ferramentas e técnicas para a análise de dados e facilita a interpretação de sinais de comportamento, de preferências, de hábitos, de desejos, que os consumidores apresentam, não inteligíveis numa análise directa dos valores de negócio obtido.

A segmentação de consumidores, de empresas, de marcas ou produtos permitem-nos compreender melhor o mercado, quer ao nível de comportamentos, quer ao nível de organização e distribuição. Segmentar significa encontrar agrupamentos de indivíduos, objectos, que partilhem, associem ou sejam entendidos como tendo algumas características comuns. Com a segmentação pretende-se encontrar diferentes grupos com características homogéneas.

## Segmentação de indivíduos no *Facebook* que gostam de música

Definir os segmentos tem por finalidade eliminar ou diminuir a arbitrariedade entre grupos de consumidores, produtos ou objectos, para assim estabelecer as melhores políticas da empresa. A segmentação pretende conseguir a definição de metas, objectivos, políticas de acção ou publicidade adequadas a esses produtos, marcas, objectos ou consumidores.

Com a análise de *clusters*, pretende-se organizar um conjunto de casos em grupos homogéneos, de tal modo que os indivíduos pertencentes a um grupo são o mais semelhante possível entre si e diferenciados dos restantes. Esta análise procura classificar um conjunto de objectos (indivíduos e produtos) em grupos ou categorias usando os valores observados das variáveis, sem que seja necessário definir critérios que classificam os dados que integram determinado grupo.

Os métodos de segmentação utilizados na análise de *clusters* podem ser classificados, em termos de operacionalização, de acordo com o momento de determinação dos segmentos e com o método estatístico utilizado. Os métodos de segmentação podem ser classificados como *a priori* e *post hoc* sendo estes podem ser descritivos ou preditivos.

A escolha do método de segmentação utilizado na análise de *clusters*, influenciam os resultados de um estudo de segmentação de mercados e bem como os seus outputs. Assim sendo a comparação entre dois métodos de segmentação *k-means* e *FCM* será efectuada neste trabalho.

A indústria da música, como grande indústria que é e que movimenta milhões de euros por todo o mundo, tem que tomar posições e atitudes na conjuntura actual.

Os especialistas prevêem que este ano, a venda mundial de música *on line* será responsável por 20% da totalidade das vendas de música em todo o mundo e que em 2020 esta percentagem seja superior a 60%.

Com o desaparecimento cada vez maior dos suportes tradicionais de audição de música, vinil e CD, os sítios de música na internet e que em particular utilizam as redes sócias, como o *Facebook* para a sua venda, têm vindo vertiginosamente a aumentar. Não é de

estranhar que actualmente o *Facebook* seja um veículo privilegiado para a venda de música e de plataforma de audição.

O objectivo fulcral do presente trabalho é a comparação entre dois métodos de segmentação, *k-means* e *FCM*, utilizando como base de estudo, as pessoas que gostam de música e que têm conta no *Facebook*.

Este trabalho começa por uma revisão bibliográfica sobre a segmentação, onde são abordados os níveis de segmentação bem como as bases e métodos de segmentação. São descritos também as partições e algoritmos de *clustering*, que este trabalho se propõe realizar, *k-means* e *Fuzzy C-Means*.

De seguida é apresentada a metodologia de trabalho e depois a análise dos resultados obtidos através da recolha dos questionários, onde primeiramente é realizada um pré-tratamento dos dados.

Para uma melhor inteligibilidade e interpretação dos resultados obtidos foi realizada uma redução dos dados utilizando o método dos *Principal Components Analysis*.

Por fim faz-se a comparação dos resultados com a utilização dos algoritmos utilizados, as limitações e é ainda apresentado trabalhos futuros referentes a este tema.

## 2. Revisão Bibliográfica

### 2.1. Segmentação

O conceito base que suporta a segmentação de mercado – subgrupos de consumidores que partilham semelhantes necessidades e desejos – não parece à primeira vista, uma problemática muito complexa para a disciplina do Marketing. No entanto, com os recentes desenvolvimentos nesta área, a segmentação de mercado tem vindo a tornar-se uma disciplina cada vez mais rica e complexa.

Neste capítulo, começa-se com uma breve descrição das origens do conceito “Segmentação de Mercado”, bem como a definição da expressão.

De seguida procede-se à identificação de mercado que são função das bases (critérios ou variáveis) e dos métodos usados para os definir. São normalmente estas dimensões que estruturam a investigação em segmentação de mercado. Assim sendo, as bases de segmentação são estruturadas e cada variável ou critério é ajustada tendo em conta as propriedades mais importantes ou desejáveis para os segmentos de mercado.

### 2.1.1. “Segmentação de Mercado” – Origem do Conceito

O conceito de segmentação de mercado é atribuído por muitos autores a *Wedel R. Smith* (1956), publicado num artigo intitulado “*Product differentiation and market segmentation as alternative marketing strategies*”.

No entanto segundo *Goller* (2002, *Cataldo e Martin Savard*, 2010), a possibilidade de segmentar um mercado homogéneo em subsegmentos foi primeiramente discutido por *Joan Robison* (1933) e por *Edward Chamberlin* (1933).

A base teórica de *Wedel R. Smith* (1956) assenta na concorrência imperfeita e o mesmo autor conclui que uma empresa que adoptasse uma estratégia de segmentação conseguiria diversificar a sua oferta ao mercado e deste modo conseguiria colher os benefícios de uma maior flexibilidade no seu *product-mix* e assim alcançar melhores resultados financeiros. Segundo *Smith* (1956), bens e/ou serviços para serem vendidos têm que ter em atenção as necessidades e logicamente o reconhecimento da heterogeneidade dos consumidores.

Para outros autores, *Bass* (1968), *Wilkie e Cohen* (1977), a segmentação de mercado é observada como a consequência natural da heterogeneidade do mercado e assim existe uma necessidade latente para o desenvolvimento de uma estratégia para a criação de subsegmentos em que os seus elementos serão cada vez mais homogéneos entre si.

Outros autores, *Punj e Stewart* (1983), definem a segmentação de mercado, indiferentemente do método utilizado, como a procura em identificar grupos de identidades em que os mesmos partilham características comuns. Essas características podem ser pessoas, empresas ou organizações.

## Segmentação de indivíduos no *Facebook* que gostam de música

Segundo *Rajeev Kamineni* (2005) a segmentação de mercado, deve ser também heterogénea em relação aos demais segmentos, ou seja, as características, necessidades e comportamentos dos consumidores de um segmento sejam, tanto quanto possível, diferentes dos consumidores de outros segmentos de mercado.

Já na opinião de *Dickson e Ginter* (1987), a segmentação de mercado é definida como um estado do mercado em que existe heterogeneidade na procura, permitindo a identificação de segmentos constituídos por diferentes funções de procura.

Uma definição bastante abrangente, encontra-se em *Wedel e Kamakura* (2000), onde se define a segmentação de mercado como “um conceito teórico da disciplina de marketing, em que se efectuam partições tendo como ponto de partida um mercado heterogéneo, sendo que se criam de sub-mercados em que a procura é homogénea. Tem como propósito o ajuste mais fino da marca, produtos ou serviços às necessidades dos consumidores para assim maximizar a rentabilidade dos recursos ao marketing”.

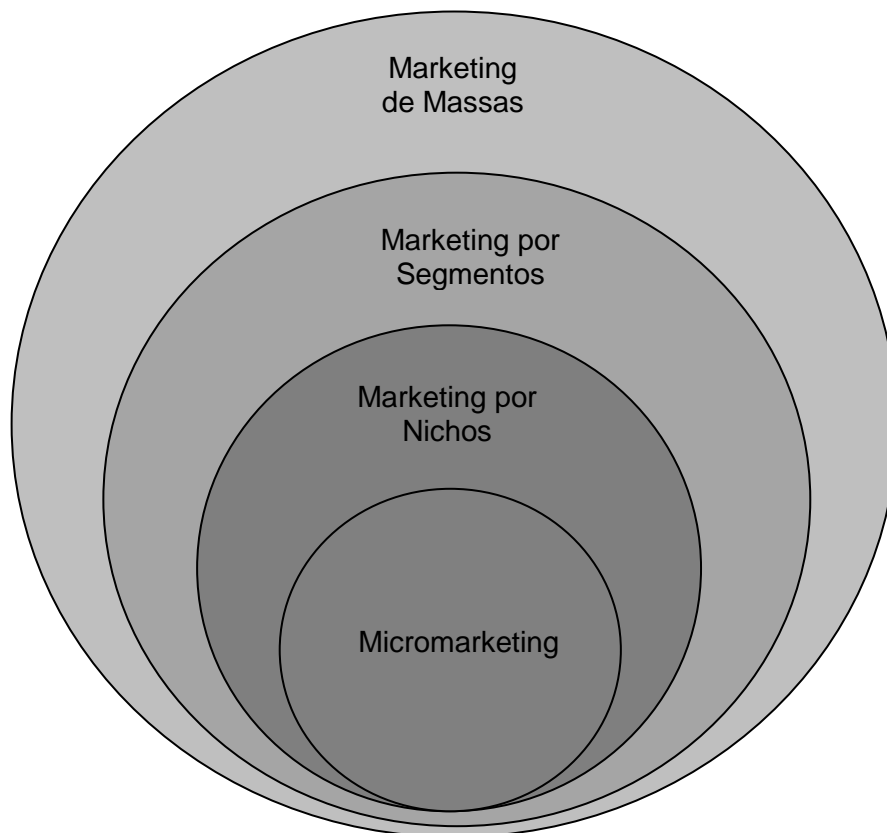
### 2.1.2. Níveis de Segmentação

Para *Kara e Kaynak* (1997) existem os seguintes níveis de segmentação no marketing:

- Marketing indiferenciado (*mass marketing*)
- Marketing diferenciado (*segment marketing*)
- Marketing concentrado (*niche marketing*)
- Marketing personalizado (*finer segmented marketing*)

Para *Kotler* (2009), a segmentação pode ser agrupada em diversos níveis, que começa na ausência da mesma, a que corresponderá o marketing de massas até ao micro marketing que inclui o marketing local e o marketing individual. Existem ainda a nível intermédio os níveis de marketing por segmentos e marketing por nichos.

**Figura 1 - Níveis de Segmentação de Mercado**



Fonte: Adaptado de Wedel e Kamakura (2000)

A ausência de segmentação foi bastante usada no princípio do século passado em que a procura de bens era bastante grande e em que os fornecedores desses produtos eram em número reduzido. Um exemplo clássico era o automóvel *Ford T*, que era idêntico em todas as versões e podia ser pintado em todas as cores, desde que fosse em preto (Kotler, 2009).

O marketing por segmentos tem vantagens claras sobre a ausência de segmentação, pois possibilita produtos ou serviços diferenciados e ajustados às necessidades díspares dos diversos tipos de consumidores.

O marketing por nichos, utiliza uma segmentação ainda mais subdividida para assim afinar ainda mais as necessidades e desejos dos consumidores. No entanto, uma maior

sintonização com as necessidades e desejos dos consumidores, a dimensão da base de cliente fica cada vez mais reduzida, correndo o risco de que a rentabilidade dos produtos ou serviços não serem justificados.

Com o desenvolvimento das tecnologias de informação e instrumentos de armazenamento, surgiu a segmentação individual (micro marketing), tornando-se a segmentação ainda mais refinada e individualizada. A este nível os dados pessoais de cada indivíduo, são tidos em conta para a elaboração dos produtos ou serviço.

No entanto esta última forma de segmentação está a causar alguma resistência. Segundo (Kotler, 2009), existem duas possíveis razões para esta resistência: a primeira é que as pessoas não querem adquirir produtos ou serviços com empresas sem conhecerem a cara de quem lhes vende e que portanto são empresas que apenas lhes pretendem vender os produtos. A outra razão está relacionada com a síndrome do “*Big Brother*”, em que as empresas colecionam vasta informação sobre os consumidores e podem assim disseminar esta informação para outros potenciais fornecedores.

### 2.1.3. Critérios de Segmentação

Após o agrupamento em conjuntos homogêneos dos consumidores, a segmentação de mercado só tem utilidade se preencherem certos requisitos. Segundo *Wedel e Kamakura*, (2000), são seis os critérios: identificabilidade, substancialidade, acessibilidade, estabilidade, capacidade de resposta e accionabilidade.

Segundo *Kotler* (2009), os segmentos devem ser passíveis de serem medidos, com dimensão suficiente para serem lucrativos e susceptíveis de serem servidos por programas específicos de marketing.

Estes critérios estabelecem condições para a obtenção de segmentos de mercado o que resulta na validação dos resultados obtidos.

Vamos apresentar uma descrição mais pormenorizada dos conteúdos dos critérios mencionados:



## Segmentação de indivíduos no *Facebook* que gostam de música

- **Identificabilidade:** quando se consegue distinguir grupos de consumidores no mercado usando bases de segmentação específicas. Tem que se poder identificar os consumidores em cada segmento com base em variáveis que são facilmente mensuráveis.
- **Substancialidade:** a dimensão de um segmento tem que possuir uma dimensão crítica, para que se possa aplicar um programa de medidas de marketing que se torne rentável. Os conceitos de micro marketing e marketing individual, podem tornar os programas de marketing mais caros e assim pouco rentáveis.
- **Acessibilidade:** mensura a capacidade para se atingir os segmentos alvos através dos esforços das políticas de distribuição e de comunicação. Não faz muito sentido segmentar um mercado, quando se sabe com antecedência que não existe um forma prática e expedita, de alcançar os membros do referido segmento.
- **Estabilidade:** Só os segmentos que permanecem estáveis durante pelo menos o tempo de identificação dos mesmos, é que podem fornecer as bases para a implementação de uma estratégia de marketing bem-sucedida e assim produzir os resultados esperados.
- **Capacidade de resposta:** os diferentes segmentos têm que responder de uma forma diferenciada a diferentes programas de marketing para eles direccionados, isto é, se não existir diferenças na resposta entre os segmentos, então não existe uma eficaz segmentação.
- **Accionabilidade:** os segmentos têm que fornecer os elementos necessários para a orientação das decisões relacionadas com a efectiva especificação dos vários instrumentos do marketing. Os consumidores de um segmento de mercado, devem ser atraídos e servidos de uma forma prática e rentável.

Os resultados de um estudo de segmentação de mercado e a presença das propriedades desejáveis já referidas dependem de uma selecção adequada das variáveis de agrupamento e de uma metodologia de análise de dados e de interpretação de resultados.

É através de uma definição das diversas variáveis de agrupamento e da análise dos dados e da respectiva interpretação dos dados que alcançamos os outputs do estudo de segmentação.

Assim sendo os dois vectores em assenta a segmentação são as bases de segmentação e os métodos de classificação, *Wedel e Kamakura, (2000)*.

#### 2.1.4. Bases de Segmentação

Tal como foi referido no ponto anterior, as bases de segmentação constituem o primeiro passo para a realização da segmentação de mercado. A definição de bases de segmentação é dada ao conjunto de características ou variáveis que estão associadas a um conjunto de consumidores a que pretende agrupar num conjunto homogéneo com as referidas características em comum.

Segundo *Frank, Massy e Wind (1972)*, na segmentação é proposta uma classificação baseada em dois vectores: natureza da variável e a natureza do processo de medida.

São consideradas *bases gerais* as características culturais, geográficas, demográficas, socioeconómicas bem como as características psicográficas, valores, padrões de personalidade e estilos de vida dos consumidores. Estas variáveis são independentes de qualquer produto ou serviço, *Wendel e Kamakura (2000)*.

As características atributos de elasticidade, frequência de uso, lealdade à marca e intenções de preferência, são alguns dos exemplos das variáveis específicas do produto.

Como foi referido anteriormente, as bases de segmentação também são classificadas em relação à natureza do processo de medida. Assim elas podem ser classificadas como *observáveis*, i.e., poder ser mensuradas directamente, sendo exemplos as características culturais, geográficas e socioeconómicas, ou podem ser classificadas como não-observáveis, i.e., inferidas, sendo exemplos as características psicográficas, personalidade e estilos de vida dos consumidores.

De uma forma resumida e segundo *Frank e tal (1972)* e por *Wedel e Kamakura (2000)*, apresenta-se o quadro 1, com a classificação das bases de segmentação.

**Quadro 1 - Classificação das Bases de Segmentação**

		Natureza da Variável	
		Gerais	Específicas do Produto
Natureza do processo de medida	Observáveis	Culturais, geográficas, demográficas socioeconómicas	Estatuto de uso, Frequência de utilização, Fidelidade á loja e marca, situações
	Não-observáveis	Psicográficas, Valores, personalidade e estilos de vida	Psicográficas, benefícios, percepções, atributos de elasticidade, preferências, intenções

Fonte: Adaptado de Wedel e Kamakura (2000, p. 7)

#### 2.1.4.1. Bases Gerais Observáveis

Numa abordagem tradicional, as bases de segmentação observáveis mais utilizadas são as geográficas, demográficas e socioeconómicas, *Wedel e Kamakura* (1998).

##### 2.1.4.1.1. Geográficas

O tipo de base geográfica é útil para a identificação, selecção e aquisição de veículos de comunicação (*Haley*, 1968).

Atendendo a que actualmente os produtos e/ou serviços são cada vez mais globais e cada vez menos locais, as empresas têm que fazer opções sobre a localização onde vão alocar mais recursos. Assim, e com o intuito de rentabilizar ao máximo os recursos da empresa, a segmentação geográfica tem uma importância cada vez maior.

A segmentação de mercado, necessita claramente de uma base geográfica porque permite à equipa de marketing dimensionar o mercado, identificar as localizações mais dinâmicas e

## Segmentação de indivíduos no *Facebook* que gostam de música

depois dos segmentos estarem definidos, poder-se então desenvolver os objectivos de marketing mais apropriados para essas mesmas localizações, (*Malcon e Ian Dunbar, 2004*).

De seguida apresenta-se um quadro 2 com as variáveis e respectivos elementos associados à base geográfica (*Weinstein, 1994*):

### **Quadro 2 – Base de Segmentação Observável Geográfica**

Variável	Elementos
Métricas geográficas do mercado	Dimensão populacional
	Densidade populacional
	Mobilidade geográfica
Abrangência do mercado	Global
	Regional
	Nacional
	Local
	Individual

Fonte: Adaptado de Weinstein (1994)

#### 2.1.4.1.2. Demográficas

Tal como a segmentação geográfica, os critérios demográficos são úteis para a identificação, selecção e aquisição de veículos de comunicação (*Haley, 1968*).

Com o aumento da área geográfica onde o produto/serviço passou a estar presente, a aplicação do tipo de base demográficos passou a ser justificada.

O tipo de base demográfica, que descreve os consumidores de uma forma mensurável, ajuda a identificar elementos de cada segmento, que por seu turno vai ajudar a determinar a forma de se chegar a esses mesmos elementos, (*Malcon e Ian Dunbar, 2004*).

Uma das razões pela qual este tipo de base é tão popular no marketing, é que normalmente associam ao consumidor necessidades e vontades, (*Kotler, 2009*).

## Segmentação de indivíduos no *Facebook* que gostam de música

Apesar da base demográfica por si só não conseguir definir um segmento, continua a desempenhar um papel importante em projecto de segmentação e de ser necessário informação adicional sobre os grupos de consumidores,

Não obstante esta forma de segmentação não conseguir, muitas vezes, discriminar de forma adequada os segmentos de mercado, as variáveis demográficas continuam a ser importantes, embora seja necessária informação adicional sobre os grupos de consumidores (*Malcon e Ian Dunbar, 2004*).

De seguida apresenta-se um quadro 3 com as variáveis e respectivos elementos associados ao tipo de base geográfica (*Weinstein, 1994*):

### **Quadro 3 – Base de Segmentação Observável Demográfica**

<b>Variável</b>	<b>Elementos</b>
Dimensão da Família	Dimensão do agregado familiar
Idade	Escalões etários
Ciclo de vida da família	Agregação de fases etárias
Género	Masculino
	Feminino
Orientação Sexual	Homossexual
	Heterossexual
Variáveis Culturais	Raça
	Nacionalidade
	Religião

Fonte: Adaptado de Weinstein (1994)

#### 2.1.4.1.3. Socioeconómicas

Este tipo de base, divide o agrupamento de consumidores tendo em conta os seus rendimentos e a sua condição social.

## Segmentação de indivíduos no *Facebook* que gostam de música

O estado socioeconómico deriva normalmente de um conjunto de medidas envolvendo o rendimento familiar bem como a educação dos seus elementos, (*Wedel e Kamakura, 2000*).

Como exemplo, a classificação segundo o “*The National Statistics Socio-economic Classification*” (NS-SEC), é a seguinte:

**Quadro 4 – *The National Statistics Socio-economic Classification* (NS-SEC)**

<b>Grupo</b>	<b>Nome</b>	<b>Descrição</b>
A	Classe Alta	Constituída por gestores de topo, administradores, profissionais liberais
B	Classe Média	Gestores intermédios, administradores e profissionais liberais
C1	Classe Média Baixa	Classe religiosa e gestores em início de carreira
C2	Técnicos Especializados	Trabalhadores qualificados em alguma área
D	Trabalhadores em geral	Trabalhadores sem qualificações e artesãos
E	Subsistência	Pensionistas, desempregados e trabalhadores ocasionais

Fonte: The National Statistics Socio-economic Classification (NS-SEC)

### Quadro 5 – Base de Segmentação Observável Socioeconómica

Variável	Elementos
Factores económicos	Nível educacional
	Ocupação
	Rendimento
Características habitacionais	Casa própria ou alugada
	Tipo de habitação
	Medidas de mobilidade
	Estabilidade familiar
Classe Social	Sistemas de classificação social
Agrupamento geodemográfico	PRIZM
	ACORN

Fonte: Adaptado de Wedel e Kamakura (2000)

Existe uma grande variedade de sistemas de segmentação geodemográficas, de entre as quais se destacam o PRIZM e ACORN, (Gunter e Furnham, 1992). A segmentação geodemográfica, é no entanto uma variável composta sendo que da sua composição, fazem parte outras variáveis do tipo gerais observáveis (Michman, 1991).

#### 2.1.4.2. Bases Gerais Não Observáveis

Este tipo de bases, teve a sua origem nos anos 60 do século passado e foi de encontro à necessidade do marketing obter respostas para assim conseguir ter uma imagem mais próxima da realidade dos consumidores e também uma melhor compreensão das suas motivações, (Wedle e Kamakura, 2000).

O desenvolvimento deste tipo de base de segmentação resulta da constatação das diversas diferenças demográficas que existem entre os vários grupos de consumidores que optam por escolherem produtos díspares (Ziff, 1971).

## Segmentação de indivíduos no *Facebook* que gostam de música

Segundo (Frank, 1972; Van Raaij e Verhallen, 1994; Wedel e Kamakura, 2000), as variáveis de segmentação destas bases, poderão ser encaixadas dentro de três grupos: personalidade, valores e estilos de vida.

A segmentação psicográfica é o resultado da necessidade em identificar causas, não se limitando à descrição de situações/comportamentos. A rápida velocidade na sua difusão constitui um evidente sinal que veio preencher necessidades não satisfeitas (Wells, 1975).

Segundo o mesmo autor Wells (1975), as questões que se colocam sobre a segmentação psicográfica podem ser agrupadas em termos de:

- Fiabilidade, dos itens e escalas, das variáveis dependentes, das relações e da estrutura;
- Validade das variáveis, das escalas, dos segmentos e da previsão;
- Aplicação aos problemas de marketing;
- Contribuição para o estudo do comportamento do consumidor.

Segundo (Kotler, 2009), a base de segmentação psicográfica, usa a psicologia e a demografia para assim melhor entender os consumidores. A segmentação psicográfica, pode ser agrupada em três grupos: psicológicos/personalidade, estilo de vida e valores.

Valores e sistemas de valores têm sido a base de segmentação de mercado segundo Kahle, Beaty e Holmer (1986), Novak (1990) e Kamakura (2000)

Os instrumentos mais usuais para mensurar os valores do comportamento humano e identificar os sistemas de valores são a escala de valores de Rokeach (RVS), Rokeach (1973) e a lista de valores de Kahle, “*List of Values*” (Kahle, 1983).

O RVS é um instrumento que inclui 18 valores terminais e também 18 valores instrumentais, que são ordenados pelos indivíduos separadamente, pela ordem de importância.



## Segmentação de indivíduos no *Facebook* que gostam de música

A lista de valores proposto por *Kahle* (1983), propõe uma escala mais pequena e mais simples, constituída por nove valores terminais directamente relacionados com os papéis e as situações diárias.

A LOV, ou seja, a lista de valores, serve como instrumento de medida do valor no estudo das tendências das similaridades e diferenças do consumidor, i.e., na natureza do consumidor (*Kahle & Kennedy*, 1989).

Alternativamente foi desenvolvida uma outra escala por *Schwartz* e *Bilsky* (1987, 1990) e *Schwartz* (1992) para aferir o sistema de valores e é constituída por 56 valores, representativos de 11 domínios motivacionais.

Em 1963, Lazer introduziu o conceito de estilo de vida no marketing que é aferido através do modelo Actividades, Interesses e Opiniões (AIO), "*Activities, Interests and Opinion*" (*Plummer*, 1974).

Este modelo estabelece compreensão mais ampla do comportamento dos indivíduos, o que enriquece o processo de definição do mercado-alvo (*Shufeldt, Oates, & Vaught*, 1998).

Os padrões de estilo de vida do consumidor são definidos por meio da inclusão do já referido AIO onde:

- (1) Actividades são o modo como o consumidor despende seu tempo e dinheiro.
- (2) Interesses são tudo aquilo que está à sua volta e que considera mais ou menos importante.
- (3) Opiniões são como os consumidores se vêem e como eles vêem o mundo à sua volta.

Dessa forma, o uso do estilo de vida apresenta uma visão geral do mercado sob forma multidimensional, tornando-se importante fonte de informação (*Shufeldt, Oates, & Vaught*, 1998).

A VALS, desenvolvida por *Mitchell* em 1983, baseou-se na hierarquia das necessidades de Maslow e no conceito da característica social (*Kahle, Beatty, & Homer*, 1986). Mais tarde,

foi desenvolvida, em 1989, a VALS2. Esta, além de definir melhor os segmentos de mercado, leva em consideração mudanças sociais e económicas. A diferença, entre a VALS e a VALS2 é que a primeira se preocupa com o que o consumidor valoriza, enquanto a segunda explica porque e como é que o consumidor decide comprar determinado produto (Gates, 1989).

Segundo Kahle e Kennedy (1989), a ideia principal empunhada pela VALS era de que os consumidores compram produtos em parte para reflectir os seus valores e estilos de vida. Assim, está intrínseco o princípio da abstracção, ou seja, associando um conceito abstracto (por exemplo, um valor) a algo específico (por exemplo, um produto), faz-se com que o algo específico do produto receba afeições positivas, associadas ao conceito abstracto do valor

A vantagem de se utilizar LOV ao invés de VALS é que esta última estuda demasiadamente, os aspectos demográficos, enquanto LOV estuda, com mais intensidade, o comportamento do consumidor. Além disso, LOV é mais interessante que RVS, pois para além de ser baseada na RVS, relaciona-se mais com a vida e com as situações diárias das pessoas, além de não ter certos problemas metodológicos que a RVS apresenta (Kahle & Kennedy, 1989).

Um dos pontos, contudo, que dificulta a utilização das bases gerais não observáveis nas organizações é a dificuldade de identificação e mensuração das variáveis que as compõem (Schiffman & Kanuk, 2000). No meio académico, o desenvolvimento de escalas específicas e a proliferação de técnicas de validação de escalas multi-itens, tais como o alfa de Cronbach e a análise factorial, tornaram bastante atractivas as variáveis psicográficas, que na sua maioria não são mensuráveis directamente por meio de uma questão (Wind, 1978).

#### 2.1.4.3. Bases Específicas do Produto Observáveis

As bases desta classe que estão relacionadas com a compra do produto/serviço bem como com a forma como os consumidores o utilizam, fazem com que este grupo de bases de segmentação seja muito utilizado.

## Segmentação de indivíduos no *Facebook* que gostam de música

As variáveis observáveis são as seguintes:

- Estado de Uso: (*Boyd e Massy, 1972; Frank, Massy e Wind, 1972*)
- Frequência de Utilização: (*Twedt, 1967*)
- Fidelidade à Marca: (*Boyd e Massy, 1972*)
- Fidelidade à Loja: (*Frank, Massy e Wind, 1972; Loudon and Della Bitta 1984*)
- Store Patronage: (*Frank, Massy e Wind, 1972*)
- Estágio de adopção: (*Frank, Massy e Wind, 1972*)
- Situação de uso: (*Belk, 1975; Dickson, 1982; Loudon e Della Bitta, 1984*)

A escala da variável observável Frequência de Utilização, segundo *Twedt, 1967*, é constituída por duas classes: *heavy users* e *light users*, enquanto a Lealdade à Marca, pode ou não ser directamente observável.

Em relação ao, Estágio de Adopção, é usual a adopção das seguintes escalas (*Rogers, 1962*):

- *Innovators* (consumidores que gostam de experimentar novas ideias com determinado risco)
- *Early adopters* (líderes de opinião, adoptam novas ideias mas com alguma reserva)
- *Early majority* (adoptam as novas ideias antes da maioria mas não são líderes)
- *Late majority* (são cépticos e só adoptam as ideias depois da maioria)
- *Laggards* (são tradicionalistas pouco propensos a mudanças)

Segundo *Kotler (2009)*, a Lealdade à Marca tem as seguintes escalas:

- *Hard-core loyal* (consumidores que só compram uma e só uma marca)
- *Split loyal* (consumidores que são leais a duas ou três marcas)
- *Shifting loyal* (consumidores que mudam de lealdade a uma marca para outra)
- *Switchers* (consumidores que não são leais a nenhuma marca)

#### 2.1.4.4. Bases Específicas do Produto Não Observáveis

As variáveis específicas do produto não observáveis são constituídas por cinco classes (*Wedel e Kamakura, 2000*): psicográficas específicas do produto, benefícios, preferências e intenções comportamentais, percepções da importância dos atributos dos produtos/marcas e elasticidades.

Segundo *Dhalla e Mathatoo (1976)*, existem três áreas chave na classe psicográfica específica do produto: orientadas para o valor, role perceptions e estilos de compra.

Para *Haley (1968)*, a segmentação por benefícios começou a ser utilizada pelas grandes empresas dos EUA a partir de 1961.

Este tipo de segmentação é interessante pois, fornece uma perspectiva algo diferente sobre o mercado. No seguimento da segmentação psicográfica, a segmentação por benefícios é apresentada como uma abordagem que permite a identificação de segmentos a partir de factores causais, por oposição aos factores descritivos (*Haley, 1968*).

O interesse da utilização deste tipo de segmentação, é que leva em consideração os benefícios dos consumidores esperam obter dos produtos/serviços e como tal, constituem as razões básicas para a existência de verdadeiros segmentos de mercado (*Loker e Perdue, 1992*).

Este critério de segmentação, segundo *Wind (1978)*, é o critério mais apropriado para a compreensão genérica do mercado sendo também o mais adequado para a tomada de decisão em termos de desenvolvimento de novos produtos, posicionamento, publicidade e distribuição. Esta ideia é partilhada por *Woodside e Jacobs (1985)*, pois segundo eles o conhecimento dos benefícios que se esperam, podem ser úteis para a definição de possíveis alterações no produto e nos conteúdos da comunicação.

#### 2.1.4.5. Conclusão

De uma forma geral não existe uma única base para a segmentação de mercado visto que uma base é só uma dimensão e não se consegue assim construir um perfil completo e robusto do consumidor.

## Segmentação de indivíduos no *Facebook* que gostam de música

Para se aferir a robustez das diversas bases de segmentação em relação aos critérios que definem as propriedades desejáveis dos segmentos de mercado, apresenta-se o seguinte quadro (Wedel e Kamakura, 2000).

**Quadro 6 – Avaliação da base de segmentação**

Bases	Critérios						
	Identificabilidad e	Sustentabilidade	Acessibilidade	Estabilidade	Accionabilidade	Resposta	
<b>Gerais</b>							
<b>Observáveis</b>	++	++	++	++	-	-	
<b>Específicas</b>							
<b>Observáveis</b>	Compra	+	++	-	+	-	+
	Uso	+	++	+	+	-	+
<b>Gerais não</b>	Personalidade	+/-	+	-	+/-	+/-	-
<b>Observáveis</b>	Estilo de Vida	+/-	+	-	+/-	+	-
	Psicográfica	+/-	+	-	+/-	+	-
<b>Específicas</b>	Psicográfica	+/-	+	-	-	++	+/-
<b>não</b>	Percepções	+/-	+	-	-	+	-
<b>Observáveis</b>	Benefícios	+	+	-	+	++	++
	Intenções	+	+	-	+/-	-	++

++ Muito Bom; + Bom; +/- Moderado; - Fraco; -- Muito Fraco

Fonte: Adaptado de Wedel e Kamakura (2000, p. 16)

Como se pode constatar, não há nenhum critério de segmentação que não revele pontos fracos.

## Segmentação de indivíduos no *Facebook* que gostam de música

As bases gerais observáveis (demográficas e geográficas) têm como ponto forte a capacidade para identificar os consumidores que pertencem a cada um dos segmentos e assim geram agrupamentos estáveis, com dimensão suficiente para a viabilização dos programas de marketing e susceptíveis de serem atingidos através de políticas de distribuição e de comunicação.

Os pontos fracos das bases demográficas e geográficas residem na (in)capacidade para gerarem segmentos de mercado que respondam de modo diferenciado e único aos programas de marketing e/ou que se revelem (in)consistentes com os objectivos das organizações.

Por seu turno, as bases específicas não observáveis (psicográficas e benefícios) apresentam boas capacidades para gerarem segmentos de mercado que respondem de modo diferenciado e único aos programas de marketing, ao invés das bases gerais observáveis (demográficas e geográficas).

Em contrapartida, os pontos fracos das bases específicas não observáveis (psicográficas e benefícios) estão na dificuldade que têm nas políticas de comunicação e de distribuição em alcançarem os consumidores em cada um dos segmentos de mercado identificados.

A capacidade para se alcançar os segmentos alvo através das políticas de distribuição e de comunicação, critério de acessibilidade, e a capacidade para produzir segmentos de mercado que respondam de modo diferenciado e único aos programas de marketing, reposta, são as situações mais críticas.

Embora se reconheça que as bases do tipo geral observável, onde se incluem as demográficas e geográficas, são as menos adequadas (efectivas) no conhecimento do comportamento de compra, continuam a ser muito utilizados, quer isoladamente quer em conjunto com as outras bases em processos de segmentação mais complexos, pois são particularmente úteis para determinar a acessibilidade aos segmentos (*Wedel e Kamakura, 2000*).

### 2.1.5. Métodos de Segmentação

A selecção das bases de segmentação bem como a escolha do método de segmentação, pela qual se agrupa os consumidores em grupos homogéneos, são os factores que influenciam os resultados de um estudo de segmentação de mercados e os seus outputs: o número e os segmentos identificados.

Os métodos de segmentação podem ser classificados, em termos de operacionalização, de acordo com o momento de determinação dos segmentos e com o método estatístico utilizado (Wedel e Kamakura, 1998):

**Quadro 7 – Classificação dos Métodos Usados na Segmentação**

	<b>A priori</b>	<b>Post Hoc</b>
<b>Descritivos</b>	Tabelas de contingência	Métodos de <i>Clustering</i> :
		Métodos não sobrepostos
		Métodos sobrepostos
	Modelos Log-lineares	Métodos <i>Fuzzy</i>
Modelos de mistura		
<b>Preditivos</b>	Tabelas cruzadas	AID
	Regressão	CART
	Modelos <i>Logit e Probit</i>	Regressão <i>Clusterwise</i>
	Análise Discriminante	Redes Neurais
		Modelos de mistura

Fonte: Adaptado de Wedel e Kamakura (1998, p. 18)

Segundo *Green* (1977) e *Wind* (1978), os métodos de segmentação podem ser classificados como *a priori* e *post hoc*. Considera-se uma segmentação *a priori* quando o tipo e o número de segmentos são determinados pelo analista, não tendo em conta a amostra do mercado. Assim a escolha da(s) variável(eis) resulta do que o investigador considerar que é importante para compreender o comportamento do consumidor (*Hsieh e tal.*, 1992).

Na abordagem *post hoc*, o número de segmentos é determinado com base na análise dos resultados da pesquisa. Começa-se pela definição de características importantes, as quais são utilizadas para a constituição de grupos (*clusters*) homogêneos. Para se obter o número de segmentos e recorrendo à análise dos dados, utilizam-se técnicas de estatística multivariada, nomeadamente a análise factorial e a análise de *clusters*.

Uma outra forma de classificação da segmentação tem por base métodos estatísticos, que podem ser descritivos ou preditivos. Nos métodos *descritivos* faz-se uma análise de associações entre um conjunto de variáveis, sem que as mesmas sejam classificadas em dependentes ou independentes, enquanto nos métodos *preditivos* a análise de associações é realizada entre dois conjuntos de variáveis, e através das variáveis independentes procura-se explicar as variáveis dependentes.

#### 2.1.5.1. Métodos de Segmentação *a Priori* Descritivos

Como anteriormente foi referido, no tipo método de segmentação *a priori* descritivo, quer o tipo quer o número de segmentos são determinados antes da recolha de dados.

Os modelos log-lineares e as tabelas de contingência parecem ter sido as ferramentas mais usadas, nos primeiros estudos de segmentação de mercado, sendo que as tabelas de contingência permitem uma representação simples de associações entre diferentes bases de segmentação e o teste de associações.

O problema em usar as tabelas de contingência é que com o aumento da dimensão pela adição de variáveis tornam-se mais difíceis de manipular, quer em termos estatísticos, quer em termos conceptuais.

*Green, Carmone e Watchpress (1976)*, sugerem o uso dos modelos Log-lineares para testar se os segmentos derivados a partir de determinadas bases exibem associações. Estas técnicas são actualmente usadas essencialmente em esquemas híbridos de segmentação que combinam métodos *a priori* com métodos *post hoc*. A segmentação híbrida pode ser vista como a combinação dos pontos forte dos métodos *a priori* e com os métodos *post hoc*,



dependendo a sua eficiência em grande parte do método utilizado post hoc, normalmente baseado em *clustering*, (Wedel e Kamakura, 1998).

#### 2.1.5.2. Métodos de Segmentação *A Priori* Preditivos

Nas abordagens de segmentação preditivas, destacam-se dois tipos de abordagens (Wilkie e Cohen, 1977; Wedel, 1990; Wedel e Kamakura, 2000), designadas por '*forward*' e '*backward*'.

Na abordagem '*forward*' são formados segmentos *a priori* com base em variáveis gerais (demográficas, psicográficas, socioeconómicas, etc.) e posteriormente relacionados entre uma variável dependente e o comportamento de compra. Na abordagem '*backward*' os segmentos são primeiramente definidos com base em variáveis relacionadas com a compra (por exemplo volume de compra, lealdade à marca) e de seguida analisa-se se existem diferenças nas características dos consumidores entre os segmentos.

Na abordagem '*backward*', o método mais popular, para se realizar a segmentação de mercado faz-se uma análise discriminante. O método é mais útil para a descrição dos segmentos de mercado do que para a sua identificação, Frank, Massy e Wind, (1972).

Para além da análise discriminante, são também utilizados os métodos estatísticos como a regressão, modelos *logit/probit*, tabelas cruzadas e *multinomial probit* (Rao e Winter, 1978; Gensch, 1985).

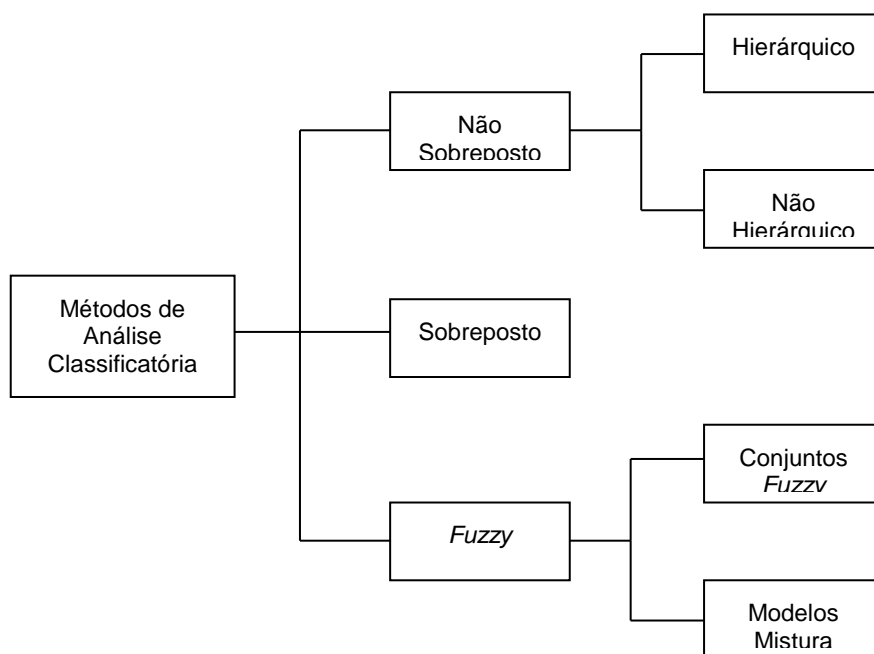
Concluindo, a implementação dos métodos de segmentação *a priori* preditivo é realizada em duas etapas: numa primeira fase são definidos segmentos *a priori* através das bases de segmentação (normalmente bases gerais) e numa segunda fase os segmentos identificados são descritos através de um conjunto de variáveis independentes, (Wedel e Kamakura, 2000). A ineficiência na definição dos segmentos *a priori* através das bases de segmentação, é a grande desvantagem na utilização destes métodos.

### 2.1.5.3. Métodos de Segmentação *Post Hoc* Descritivos

Nos métodos *post hoc* descritivos, os segmentos são identificados pela forma homogénea como os grupos de consumidores são agrupados de acordo com as características mensuradas. O número de segmentos e as suas características de cada segmento são determinados pelos dados e pela metodologia usada. A análise classificatória é a ferramenta mais usada para a segmentação *post hoc* descritiva (Punj e Stewart, 1983; Arabie e Hubert, 1994).

Na análise classificatória segundo Hruschka (1986), os métodos análise combinatória distinguem-se na natureza dos agrupamentos: não sobrepostos, sobrepostos e difusos (*fuzzy*).

**Figura 2 – Classificação dos Métodos de Análise Classificatória**



Fonte: Adaptado de Hruschka (1986)

## Segmentação de indivíduos no *Facebook* que gostam de música

O agrupamento Não Sobreposto, pressupõe que cada objecto (consumidor) só pode pertencer a um único segmento. Neste tipo de agrupamento existem dois métodos distintos: Hierárquicos e Não hierárquicos, (*Frank e Green, 1968*).

Segundo *Punj e Stewart's, (1983)*, no agrupamento Não Sobreposto, o método não hierárquico é melhor do que o método hierárquico devido à sua robustez no que diz respeito a *outliers* e atributos irrelevantes. No que diz respeito ao agrupamento hierárquico, o método de ligação de mínima variância é a mais consistente nos resultados, (*Milligan e Cooper, 1985*).

Nos métodos de agrupamento sobreposto, o primeiro modelo a ser apresentado foi o ADCLUS (*Shepard e Arabie, 1979*), seguidos de outras generalizações como o INDCLUS (*Carroll e Arabie, 1983*), GENNCLUS (*DeSarbo, 1982*) e CONCLUS (*DeSarbo e Mahajan, 1984*).

Nos agrupamentos Sobreposto e *Fuzzy*, os objectos (consumidores), podem pertencer a mais do que um segmento ao mesmo tempo, sendo que a diferença reside no valor de pertença que no caso do Sobreposto é 1 e no caso do *Fuzzy* é função do “*degree of membership*”, que varia entre zero e um.

No método de agrupamento *Fuzzy*, distinguem-se dois métodos: *Fuzzy Sets* (*Zadeh, 1965*) e Distribuição Mistura (*McLachlan e Basford, 1985*). O método *Fuzzy* tem por base a teoria dos “*fuzzy sets*” que atribui um valor de pertença a cada classe.

Existem quatro grandes métodos de agrupamento *Fuzzy*: algoritmo *fuzzy c-means*, FCM (*Bezdek, 1974*), algoritmo *fuzzy c-lines*, FCL (*Bezdek e al. 1981 a,b*), o algoritmo *fuzzy clusterwise regression*, FCR (*Wedel e Steenkamp, 1993*) e o algoritmo *fuzzy grade of membership model*, GoM (*Manton, Woodbury e Trolley, 1994*).

O tipo de agrupamento *Fuzzy* consegue uma abordagem flexível que permite assim uma identificação de grupos de consumidores de diversas formas, (*Wedel e Kamakura, 1999*).

O outro tipo de segmentação agrupado no método *Fuzzy*, é o de Distribuição Mistura, que permite uma afectação difusa dos objectos aos segmentos, no pressuposto de que os dados

provêm de uma distribuição mistura, "*mixture distribution*", sendo estimadas as probabilidades de pertença dos objectos a cada classe, (McLachlan e Basford, 1988).

Nestas duas abordagens os valores de pertença pertencem ao intervalo entre 0 e 1 o que implica que existem pontos comuns entre os elementos das duas classes. Existe no entanto uma grande diferença entre estes dois métodos: enquanto no agrupamento *fuzzy sets* os valores de pertença dos consumidores assumem que estes pertencem realmente a mais do que um segmento, no caso de agrupamentos baseados em modelos mistura "*mixture models*", assume-se que o consumidor pertence apenas a um segmento. A informação, no entanto, que é representada nos dados para esse elemento, é insuficiente para determinar o seu actual segmento, e assim sendo são estimadas probabilidades de pertença aos segmentos.

Concluindo, os métodos de segmentação *post hoc* descritivos, são procedimentos muito poderosos e muito utilizado na segmentação de mercados, sendo especialmente úteis os métodos de sobreposição e os *fuzzy* os quais têm uma maior vantagem sobre os hierárquicos não sobrepostos.

#### 2.1.5.4. Métodos de Segmentação *Post Hoc* preditivos

Este método pretende identificar segmentos tendo por base a estimação de relações entre uma variável dependente, representativa do comportamento de compra (volume de compra, escolhas de marcas, incidência e momento de compra), e um conjunto de variáveis dependentes (por exemplo, características do produto, preço, promoções, variáveis demográficas ou psicográficas) para cada classe. Assim os segmentos formados através deste método, são homogéneos na relação entre a variável dependente e as variáveis independentes.

Existem actualmente várias ferramentas para a abordagem deste método: AID, CART, Redes Neurais Artificiais, Análise Conjunta, *Clusterwise Regression* e Modelos Mistura de Regressão.

O método tradicional utilizado na segmentação *post hoc* preditivo é o AID – “*Automatic Interaction Detection*” (Assael, 1970; Assael e Roscoe, 1976), tendo sido posteriormente generalizado, permitindo múltiplas variáveis dependentes (MAID – “*Multivariate AID*” (McLachan e Johansson, 1981)), variáveis categóricas dependentes (CHAID (Kass, 1980)), variáveis dependentes binárias (THAID (Wind, 1978)), testes de hipótese (CHIAID (Magdison, 1994)).

Posteriormente Breiman et al., (1984) desenvolveram outro método, designado por *classification and regression trees*, CART.

Uma técnica interessante na abordagem deste método é a utilização de redes neuronais artificiais. Os neurónios recebem os *inputs* e estes através de uma função (normalmente não linear) de activação são transmitidos aos outputs. Uma organização típica de uma célula neuronal artificial é constituída por uma *layer* de entrada por uma *layer* intermédia e por uma *layer* de saída. Na segmentação a entrada da rede neuronal é feita com as bases de segmentação que depois são usadas para predizer as variáveis dependentes de interesse.

Apesar de alguns estudos demonstrarem superioridade desta abordagem em relação métodos estatística ou o método de agrupamento hierárquico *k-means* (Balakrishnan et al., 1995) as estimativas dos parâmetros são de difícil interpretação e em muitos casos não possuem propriedades estatísticas (Wedel e Kamakura, 2000).

A metodologia de segmentação de mercado com recurso à análise conjunta é particularmente útil na segmentação *post hoc* preditivos, porque permite o agrupamento de consumidores de acordo com as suas respostas às características dos produtos/serviços durante as decisões de escolha.

Segundo Hauser e Urban (1977); Moriarty e Venkatesan (1978) e Currim (1981), a segmentação de mercado com recurso à análise conjunta, é executada em dois passos, nos quais a identificação dos segmentos e a estimação dos modelos de análise conjunta são realizados de forma separada.

## Segmentação de indivíduos no *Facebook* que gostam de música

Uma outra abordagem é proposta por *Hargety* (1985), em que utilizando a análise conjunta, estima-se o conjunto dos pesos para cada sujeito de forma a precisão do modelo conjunto seja otimizada, utilizando para isso a análise *Q-factor*.

As desvantagens que a análise conjunta tem são minoradas pelo método "*Clusterwise Regression*", originalmente proposto por *Späth* (1981 e 1982), sendo um método que faz a previsão e simultaneamente a classificação dos elementos do segmento.

Segundo *DeSarbo et al.* (1989), é proposta uma generalização conhecida por "*Overlapping Clusterwise Regression*", que para além de permitir mais do que observações individuais, também inclui múltiplas variáveis dependentes e gera valores de pertença não sobrepostos.

A "*Fuzzy Clusterwise Regression*", foi introduzida por *Wedel e Steenkamp* (1989) sendo que a matriz de partição é difusa e mais tarde *Wedel e Steenkamp* (1991) estendem o procedimento para a identificação de segmentos de mercado e o posicionamento de marcas de forma simultânea. No entanto este método tem limitações, sendo que uma delas é a não disponibilidade das propriedades estatísticas dos segmentos, que pode ser ultrapassa através dos modelos mistura de regressão.

Os Modelos Mistura de Regressão "*Mixture Regression Models*" pertencem à classe de métodos de latência e podem ser agrupados em três categorias: "*mixture*", "*mixture regression*" e "*mixture multidimensional*", (*Wedel e Kamakura*, 2000). Estes modelos permitem simultaneamente a classificação dos elementos (consumidores) em segmentos não observáveis e estima o modelo de regressão em cada segmento, relacionando a variável dependente com o conjunto de variáveis independentes, (*Wedel e DeSarbo*, 1994).

Os modelos de "*Mixture Multidimensional Scaling*", MDS, podem ser vistos como procedimentos "*data unmixing*", por conseguem simultaneamente estimar os segmentos de mercado bem como as estruturas das preferências dos consumidores em cada segmento, (*DeSarbo, Manrai e Manrai*, 1994).

Um modelo particularmente útil na segmentação de mercado é o "*clusterwise logit*", que foi inicialmente desenvolvido por *Kamakura e Russel* (1989) e serve para analisar o comportamento da escolha do consumidor. Este método identifica os segmentos de

## Segmentação de indivíduos no *Facebook* que gostam de música

consumidores que são homogéneos na preferência da sua marca e na resposta ao preço e às promoções.

### 2.1.5.5. Conclusão

No quadro 8, faz-se uma avaliação comparativa entre os diversos métodos de segmentação e os critérios de Eficácia na segmentação, Eficácia na prevenção, Propriedades estatísticas, Aplicações conhecidas e Disponibilidades de software:

**Quadro 8 – Avaliação dos Métodos de Segmentação**

Métodos	Critérios				
	Eficácia na Segmentação	Eficácia na Previsão	Propriedades Estatísticas	Aplicações Conhecidas	Disponibilidade de software
A priori, descritivos	+/-	--	+	++	++
A priori, preditivos	-	++	++	++	++
Post hoc, descritivos	++	--	-	+/-	+
Post hoc, preditivos	++	+	+/-	+	+/-

++ Muito Bom; + Bom; +/- Moderado; - Fraco; -- Muito Fraco

Fonte: Adaptado de Wedel e Kamakura (2000, p. 28)

## 2.2. Segmentação, *Targeting* e Posicionamento

Segundo *Kotler* (1988), o processo de segmentação percorre três fases (STP): Segmentação, *Targeting* e Posicionamento. No entanto, segundo *Goller, Hogg e Kalafatis* (2002), o processo de segmentação é constituído por quatro etapas distintas: análise, avaliação, implementação e controlo.

A etapa da análise integra as actividades que contêm a identificação dos segmentos; a etapa de avaliação, verifica se são cumpridas as condições para que se possa efectuar uma segmentação eficaz do mercado; etapa da implementação é dividida em três fases: implementação estratégica, na qual é realizada a escolha dos mercados a intervir, a implementação, onde são identificados os sub-mercados e a implementação operacional onde são implementados os programas de marketing. Por fim a etapa de controlo, onde

como o próprio nome indica, é realizada a monitorização da eficácia e estabilidade dos segmentos obtidos.

### 2.2.1. *Targeting*

A segunda fase do processo de segmentação designada por *Targeting*, consiste na escolha dos segmentos anteriormente definidos pela fase de segmentação, (Kotler, 1988). Esta selecção passa pela avaliação do grau de atracção dos segmentos anteriormente identificados, sendo que serão balizados pelos objectivos e recursos da organização, (Henshall e Roberts, 1985; Mazanec, 1995; Perdue, 1996)

Segundo a (Kotler, 2009), existem muitas bases para se efectuar o *Targeting*, mas dever-se-á incluir:

- Número de segmentos e a sua rendibilidade potencial;
- Acessibilidade da organização em relação aos elementos (consumidores) dos segmentos e o custo de se realizar tal acessibilidade;
- Intensidade das organizações concorrentes nos vários segmentos;
- Nível de satisfação dos consumidores com as actuais organizações concorrentes;
- Previsão de crescimento nos segmentos;
- Barreiras à entrada.

Segundo Wind e Cardozo (1974) na identificação dos segmentos alvo, deve-se proceder à análise da rendibilidade da diferenciação dos programas de marketing para alcançar os múltiplos segmentos. Trata-se de assumir que a diferenciação dos programas de marketing implica custos, pelo que é indispensável determinar as respectivas rendibilidades. Os segmentos que não demonstrem a rendibilidade alvo, devem ser abandonados, pelo que não se procederá à aplicação de programas de marketing diferenciados.



### 2.2.2. Posicionamento

Uma vez realizadas as fases de Segmentação e *Targeting*, as organizações têm que desenvolver um plano para o produto ou serviço. É através do posicionamento que as organizações vão construir um plano, com a ajuda dos 7 P's, para se posicionarem na mente dos consumidores. É através do posicionamento que as organizações tentam conquistar na mente dos consumidores, um espaço para assim se distinguirem em relação aos concorrentes. O posicionamento escolhido deve ser influenciado pelas necessidades dos segmentos alvo a satisfazer e pelas vantagens competitivas da organização.

### 2.3. “Clustering”

A análise de “Clustering” é um método muito usado quando se pretende categorizar entidades tais como, objectos, consumidores, etc, em grupos ou *clusters*, nos quais existe uma determinada homogeneidade entre as características observadas (Wedel e Kamakura, 2000). Assim pode-se analisar de uma forma mais cuidada os grupos formados em vez de se analisar os dados no conjunto total.

A análise de “Clustering” é assim um bom método de exploração de análise de dados, sendo desta forma mais utilizado para a construção de hipóteses do que para as testar, enquadrando-se bem nos estudos exploratórios. Este tipo de análise utiliza-se muito em *Marketing Research* para a identificação e definição de segmentos de mercado para posteriormente se enquadrar na estratégia de marketing das empresas.

Segundo Gordon (1999), a análise de “Clustering” não se constitui de uma só técnica, mas sim de uma variedade de técnicas que têm por objectivo a criação de grupos com coesão interna e ao mesmo tempo isolamento externo.

Embora tenha as suas origens nas ciências biológicas, a análise de “Clustering”, disseminou-se por disciplinas, tais como a medicina, psicologia, sociologia, etc. O termo análise de “Clustering” tem outros nomes consoante as várias disciplinas onde é aplicado, assim também é conhecida por Taxonomia numérica, “*Botryology*”, *Análise Q*, reconhecimento de padrões não supervisionados.

Os métodos de “*clustering*” podem-se agrupar em dois grandes grupos através da distinção do tipo de variáveis que utilizam, dependentes e independentes. Os métodos em que se faz a distinção entre variáveis independente e dependentes são apelidados de métodos de “*clustering*” preditivos ou métodos “*clusterwise regression*”. Assim este tipo de métodos tende a criar relações entre os vários tipos de variáveis.

Por outro lado, os métodos tradicionais de “*clustering*” ou descritivos, não fazem distinção entre as variáveis dependente e as variáveis independentes nos grupos homogéneos.

Os dois grandes grupos anteriormente referidos, podem ser encaixados em três grandes categorias de análise de “*clustering*”: métodos não sobrepostos, métodos sobrepostos e métodos “*Fuzzy*”.

### 2.3.1. Os dados

As técnicas de análise de “*clustering*” podem ser aplicadas a dados, quer sejam quantitativos quer sejam qualitativos ou à mistura de ambos. Os dados são usualmente observações de quaisquer processos físicos.

Assim cada observação consiste em  $n$  variáveis mensuradas, agrupadas em  $n$ -dimensional vector linha  $\mathbf{x}_k = [x_{k1}, x_{k2}, \dots, x_{kn}]^T, \mathbf{x}_k \in \mathbf{R}^n$ . Ao conjunto de  $N$  observações é denominado por  $\mathbf{X} = \{\mathbf{x}_k | k = 1, 2, \dots, N\}$ , e é representado como a matriz  $N \times n$ :

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \dots & x_{Nn} \end{bmatrix} \quad (1.1)$$

Vamos utilizar a terminologia usualmente utilizada em análise de “*clustering*”, em que as linhas  $X$  são os objectos ou padrões e em que as colunas são denominadas por características ou atributos. A matriz  $X$  denomina-se a matriz de dados.

Neste estudo as linhas da matriz  $X$ , correspondem aos consumidores e as colunas às questões.

Atendendo ao facto de nos *clusters* não-sobrepostos, os consumidores só poderem pertencer a um e um só segmento, implica que a matriz  $X$ , só tenha nas linhas um só elemento com valor unitário, o que implica o valor zero nos outros elementos da linha.

Os métodos de Sobreposição e *Fuzzy*, têm como característica principal a não rigidez no isolamento externo entre *clusters*, e como consequência um consumidor pode pertencer a mais do que um segmento ao mesmo tempo. Deste modo a matriz  $X$ , pode ter mais do que uma entrada diferente do valor zero

Os métodos *Fuzzy* podem ser *Fuzzy Sets* ou *Mixture* e são conceptualmente bastante diferentes.

Nos métodos *Mixture*, assume-se que os segmentos são não-sobrepostos, mas devido a insuficiente informação sobre os dados dos consumidores nos segmentos, atribui-se uma probabilidade de pertença, pelo que a matriz  $X$ , tem valores pertencentes a  $\mathbf{R}^+$  na mesma linha, em que a sua soma, tem que ser obrigatoriamente igual ao valor unitário.

Em relação aos métodos Conjunto *Fuzzy*, os consumidores como pertencem a mais do que um segmento ao mesmo tempo e como têm valores parciais da *membership function*, a matriz  $X$ , tem na mesma linha valores pertencentes a  $\mathbf{R}^+$ , cuja soma tem que impreterivelmente ser igual ao valor unitário.

### 2.3.2. Os “*Clusters*”

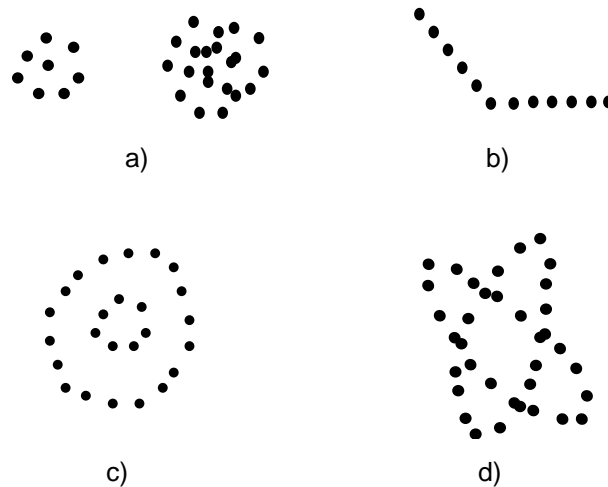
O conceito mais generalizado para “*cluster*”, define-se como um grupo de objectos que entre si são mais “*similares*” do que com outros elementos de outro grupo. O termo “*similares*” ou similaridade, deve ser interpretado como similaridade matemática, tendo como base a mensuração de alguma característica bem definida. Normalmente em espaços métricos, a similaridade é definida por meio da “*distance norm*”. Esta distância pode ser mensurada entre os próprios vectores mas também entres os vectores e algum ponto

específico do “*cluster*”. Estes pontos específicos (centroides), variam em função das várias interacções dos dados que o algoritmo executa.

Os pontos específicos podem ser vectores de igual dimensão à dos valores da matriz dos dados, mas também podem ser objectos geométricos, tais como sub-espacos ou funções lineares ou não lineares.

A matriz dos dados  $X$ , pode evidenciar *clusters* com diferentes formas geométricas, densidades ou tamanhos, como se pode constatar na figura 4.

**Figura 3 – Diferentes Tipos de Clusters em  $\mathbf{R}^2$**



Como se constata na figura 4, os *clusters* podem ser esféricos, caso a), podem ser alongados ou lineares, caso b) e finalmente podem ser “*hollow*”. Os “*clusters*” b), c) e d) são caracterizados como sub-espacos lineares ou não lineares pertencentes ao espaco em  $\mathbf{R}^2$  (nestes casos).

A performance dos algoritmos de “*clustering*” não só é influenciada pela distância entre “*clusters*”, pelas relações espaciais, como também pela densidade e forma geométrica de cada “*clusters*”.

Por fim, os “*clusters*” podem ser separados, ligados continuamente entre si ou sobrepostos uns com os outros.

### 2.3.3. Similaridade e Distâncias

Assim podemos definir “semelhança” como a quantidade que reflecte a força da relação entre dois objectos ou duas características.

Esta quantidade está geralmente compreendida no intervalo de -1 a +1 ou sendo normalizada entre o intervalo de 0 a 1.

A semelhança entre o recurso  $i$  e o recurso  $j$ , será denotada por  $s_{ij}$  e assim poderemos mensurar esta quantidade de várias formas, dependendo da escala de medição ou dos tipos de dados que tivermos.

A distância mensura a forma de dissemelhança, que não é mais do que a medida de discrepância entre os dois objectos com base em várias características, mas também pode ser vista como medida de “desordem” entre dois objectos. Essas características podem ser representadas como coordenadas do objecto no espaço de características.

Vamos definir a distância como uma variável quantitativa em geral que satisfaz pelo menos as três primeiras condições abaixo indicadas:

$d_{ij} \geq 0$ , a distância é sempre positiva ou igual a zero

$d_{ii} = 0$ , a distância é igual a zero entre o mesmo ponto

$d_{ij} = d_{ji}$ , a distância é simétrica

$d_{ij} \leq d_{ik} + d_{jk}$ , a desigualdade triangular

A distância também se chama métrica(s) se satisfizer as quatro condições acima mencionadas. Assim atendendo á condição d), desigualdade triangular, nem todas as distâncias são métricas mas todas as métricas são distâncias.

Neste estudo só será utilizada nos algoritmos o tipo de distâncias Euclidiana.

A distância euclidiana (ou distância métrica), pode ser definida como a distância entre dois quaisquer pontos sendo provada através da aplicação repetida do teorema de Pitágoras.

Matematicamente a distância euclidiana entre os ponto  $A = (a_1, a_2, \dots, a_n)$ , e  $B = (b_1, b_2, \dots, b_n)$ , num espaço euclidiano  $n - dimensional$ , é definida da seguinte forma,

$$d_{AB} = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \dots + (a_n - b_n)^2} = \sqrt{\sum_{i=1}^n (a_i - b_i)^2}$$

#### 2.3.4. As Partições dos “*Clusters*”

Como foi anteriormente referido os “*clusters*” são subconjuntos da população total de dados e assim podemos classificar os métodos de “*clustering*” de acordo com o tipo de dados, se são *fuzzy* ou “*crisp (Hard)*”.

A notação que iremos utilizar para o número de “*clusters*” é  $c$  e como denominamos que  $\mathbf{X}$  é a matriz com a totalidade dos dados, vamos considerar como  $\mathbf{U}$  a matriz partição.

Assim a matriz  $\mathbf{U} = [\mu_{ik}]$ , em que,

$$\mathbf{U} = \begin{bmatrix} \mu_{1,1} & \mu_{1,2} & \dots & \mu_{1,c} \\ \mu_{2,1} & \mu_{2,2} & \dots & \mu_{2,c} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{N,1} & \mu_{N,2} & \dots & \mu_{N,c} \end{bmatrix}$$

##### 2.3.4.1. Partição “*Hard (crisp)*”

Neste tipo de partição, assume-se que o número de “*clusters*”,  $c$ , é de antemão conhecido, devido ao conhecimento anteriormente adquirido, mas que posteriormente será validado, (*J. C. Bezdek*, 1981).

Assim a partição “*hard*” pode ser definida como uma família de subconjuntos  $\{A_i | 1 \leq i \leq c \subset P(X)\}$ , tendo as seguintes propriedades:

$$\bigcup_{i=1}^c A_i = \mathbf{X}, \quad (1.2)$$

$$A_i \cap A_j, \quad 1 \leq i \neq j \leq c, \quad (1.3)$$

$$\emptyset \subset A_i \subset \mathbf{X}, \quad 1 \leq i \leq c. \quad (1.4)$$

Estas propriedades asseguram que os subconjuntos  $A_i$  contêm todos os dados que constam da matriz de dados  $\mathbf{X}$ , são disjuntos entre si e não são nulos.

A partição “*hard*” pode ser expressa em termos da “*membership function*” da seguinte forma:

$$\bigvee_{i=1}^c \mu_{A_i} = 1, \quad (1.5)$$

$$\mu_{A_i} \vee \mu_{A_j}, \quad 1 \leq i \neq j \leq c, \quad (1.6)$$

$$0 < \mu_{A_i} < 1, \quad 1 \leq i \leq c. \quad (1.7)$$

A uma função característica do subconjunto  $A_i$  (“*membership function*”  $\mu_{A_i}$ ), só pode ter valores zero ou unitários.

Por uma questão de simplificação, iremos de futuro designar  $\mu_i$ , em vez de  $\mu_{A_i}$  e  $\mu_{ik}$  em vez de  $\mu_i(x_k)$ .

Assim a  $N \times c$  matriz partição  $\mathbf{U} = [\mu_{ik}]$ , representa uma partição “*hard*” se e só se, os seus elementos satisfizerem as seguintes condições:

$$\mu_{ij} \in 0,1, \quad 1 \leq i \leq N, \quad 1 \leq k \leq c, \quad (1.8)$$

$$\sum_{k=1}^c \mu_{ik} = 1, \quad 1 \leq i \leq N \quad (1.9)$$

$$0 < \sum_{i=1}^N \mu_{ik} < N, \quad 1 \leq k \leq c. \quad (1.10)$$

Seja  $\mathbf{X}=[\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]$ , um conjunto finito e  $2 \leq c \leq N$ , um número inteiro, a partição “hard” para  $\mathbf{X}$ , é o conjunto que tem as seguintes características:

$$M_{hc} = \left\{ \mathbf{U} \in \mathbf{R}^{N \times c} \mid \mu_{ik} \in 0,1, \forall i, k; \sum_{i=1}^N \mu_{ik} = 1, \forall i; 0 < \sum_{i=1}^N \mu_{ik} < N, \forall k \right\} \quad (1.11)$$

#### 2.3.4.2. Partição “Fuzzy”

A partição “fuzzy” pode ser vista como uma generalização da partição “hard”, sendo que a grande diferença reside no facto de que na partição “fuzzy”, se permitir que as entradas da matriz partição  $U = [\mu_{ik}]$ , possam ter valores entre 0 e 1.

Assim a  $N \times c$  matriz partição  $U = [\mu_{ik}]$ , que representa as partições “fuzzy”, tem as seguintes características:

$$\mu_{ij} \in [0,1], \quad 1 \leq i \leq N, \quad 1 \leq k \leq c, \quad (1.12)$$

$$\sum_{k=1}^c \mu_{ik} = 1, \quad 1 \leq i \leq N, \quad (1.13)$$



$$0 < \sum_{i=1}^N \mu_{ik} < N, \quad 1 \leq k \leq c. \quad (1.14)$$

Seja o espaço partição “fuzzy” definido como  $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]$  e sendo um conjunto finito em que  $2 \leq c \leq N$ , é inteiro.

O espaço de partição “fuzzy”, é o conjunto definido por:

$$M_{fc} = \left\{ \mathbf{U} \in \mathbf{R}^{N \times c} \mid \mu_{ik} \in [0,1], \forall i, k; \sum_{i=1}^N \mu_{ik} = 1, \forall i; 0 < \sum_{i=1}^N \mu_{ik} < N, \forall k \right\} \quad (1.15)$$

### 2.3.5. Algoritmos de “Clustering”

#### 2.3.5.1. O Algoritmo *K-means*

Este algoritmo é muito utilizado na formação de *clusters* apesar de nem sempre se obter resultados fiáveis. Através de uma matriz de dimensões  $N \times n$ , o algoritmo aloca cada ponto dos dados a um *cluster* de forma a minimizar a soma dos quadrados da distância ao centroide do *cluster*:

$$\sum_{i=1}^c \sum_{k \in A_i} \|\mathbf{x}_k - \mathbf{v}_i\|^2 \quad (1.16)$$

Em que  $A_i$  é o conjunto dos objectos (data points) no  $i$ -ésimo cluster e  $\mathbf{v}_i$  é a média para esses pontos sobre o cluster. Assim sendo a equação (1.16), não é mais do que a distância da norma, sendo que os centros dos clusters são obtidos através da equação:

$$\mathbf{v}_i = \frac{\sum_{k=1}^{N_i} \mathbf{x}_k}{N_i}, \quad \mathbf{x}_k \in A_i \quad (1.17)$$

## Segmentação de indivíduos no *Facebook* que gostam de música

Onde  $N_i$  é o número de objectos em  $A_i$ .

A implementação do algoritmo é realizada da seguinte forma:

Seja  $\mathbf{X}$ , a matriz dos dados e  $c$ , o número de “clusters” em que,  $1 < c < N$ , a escolha inicial dos valores dos centroides dos clusters é escolhida de uma forma aleatória.

Cria-se então um ciclo repetitivo para  $l = 1, 2, \dots$  para os seguintes passos,

**Passo 1** – Calculo das distâncias:

$$D_{ik}^2 = (\mathbf{x}_k - \mathbf{v}_i)^T (\mathbf{x}_k - \mathbf{v}_i), \quad 1 < i < c, \quad 1 < k < N \quad (1.18)$$

**Passo 2** – Seleccionar os pontos para um *cluster* em que a métrica a seguir é a distância mínima para esse *cluster*.

**Passo 3** – Calcular os centros dos *clusters*:

Repetir:

$$v_i^{(l)} = \frac{\sum_{j=1}^{N_i} x_i}{N_i} \quad (1.19)$$

Até:

$$\prod_{k=1}^n \max |v^{(l)} - v^{(l-1)}| \neq 0$$

Fim de ciclo

### 2.3.5.2. O Algoritmo “*Fuzzy c-mean*”

O algoritmo de “*clustering*” “*Fuzzy c-mean*” tem na sua fundamentação a minimização de uma função objectivo chamada “*C-means functional*”. Segundo *J.C.BEZDEK* a função é:

Segmentação de indivíduos no *Facebook* que gostam de música

$$J(\mathbf{X}; \mathbf{U}, \mathbf{V}) = \sum_{i=1}^c \sum_{k=1}^N (\mu_{ik})^m \|\mathbf{x}_k - \mathbf{v}_i\|_A^2 \quad (1.18)$$

Onde

$$\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_c], \mathbf{v}_i \in \mathbf{R}^n \quad (1.19)$$

é o vector dos centroides, um dos outputs a determinar e

$$D_{ikA}^2 = \|\mathbf{x}_k - \mathbf{v}_i\|_A^2 = (\mathbf{x}_k - \mathbf{v}_i)^T A (\mathbf{x}_k - \mathbf{v}_i) \quad (1.20)$$

é a “*squared inner-product distance norm*”.

Do ponto de vista estatístico, podemos considerar a equação (1.18), como a medida da variância total de  $\mathbf{x}_k$  de  $\mathbf{v}_i$ .

Pode-se utilizar vários métodos (desde “*grouped coordinate minimization* até a algoritmos genéticos) para a minimização da função *c-means* (1.18), que representa a optimização de um problema não linear. No entanto o método utilizado neste algoritmo, “*Fuzzy C-Means*” (*FCM*), é a interacção de *Picard* através das condições de primeira ordem, para os pontos estacionários da equação (1.18).

Os pontos estacionários da função objectivo (1.18), podem ser encontrados, adicionando a condição descrita em (1.13) á função objectivo  $J$ , através dos múltiplos de *Lagrange*:

$$\tilde{J}(\mathbf{X}; \mathbf{U}, \mathbf{V}, \lambda) = \sum_{i=1}^c \sum_{k=1}^N (\mu_{ik})^m \|\mathbf{x}_k - \mathbf{v}_i\|_A^2 + \sum_{k=1}^N \lambda_k \left( \sum_{i=1}^c \mu_{ik} - 1 \right) \quad (1.21)$$

e através dos “*setting*” dos gradientes de  $\tilde{J}$ , em relação a  $\mathbf{U}$ ,  $\mathbf{V}$  e  $\lambda$  to zero. Se as condições  $D_{ikA}^2 > 0, \forall i, k$  e  $m > 1$ , então  $(\mathbf{U}, \mathbf{V}) \in M_{fc} \times \mathbf{R}^{n \times c}$ , pode minimizar (1.18) se e só se

$$\mu_{ik} = \frac{1}{\sum_{j=1}^c \left( \frac{D_{ikA}}{D_{jkA}} \right)^{\frac{2}{m-1}}}, \quad 1 \leq i \leq c, \quad 1 \leq k \leq N \quad (1.22)$$

e

$$\mathbf{v}_i = \frac{\sum_{k=1}^N \mu_{ik}^m \mathbf{x}_k}{\sum_{k=1}^N \mu_{i,k}^m}, \quad 1 \leq i \leq c \quad (1.23)$$

Com esta solução ficam também satisfeitos os constrangimentos que estão patentes nas equações (1.12) e (1.14). É de salientar que na equação (1.23), os  $\mathbf{v}_i$ , são os pesos médios dos dados que pertencem ao “cluster”, onde os referidos pesos não são mais do que os graus de pertença. É por esta razão que o algoritmo se intitula “*c-means*” e como se verá mais adiante, o algoritmo interage entre as equações (1.22) e (1.23).

O algoritmo FCM, recorre à distância normalizada Euclidiana, que como já anteriormente foi mencionado, induz “*clusters*” hiperesféricos. Assim como consequência da escolha desta medida de similaridade, este algoritmo só detecta “*clusters*” com a mesma forma e orientação, devido ao facto da matriz da norma induz  $\mathbf{A} = \mathbf{I}$ , ou pode ser escolhida como uma matriz diagonal de  $n \times n$  que contem as diferentes variâncias segundo a direcção das coordenadas do eixo  $X$ .

Assim a matriz  $\mathbf{A}_D$ ,

$$\mathbf{A}_D = \begin{bmatrix} \left(\frac{1}{\sigma_1}\right)^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \left(\frac{1}{\sigma_n}\right)^2 \end{bmatrix} \quad (1.24)$$

ou  $\mathbf{A}$ , também pode ser definida como a matriz inversa de dimensão  $n \times n$  das covariâncias.

Assim, a matriz  $\mathbf{A} = \mathbf{F}^{-1}$ , em que,

$$\mathbf{F} = \frac{1}{N} \sum_{k=1}^N (\mathbf{x}_k - \bar{\mathbf{x}})(\mathbf{x}_k - \bar{\mathbf{x}})^T \quad (1.25)$$

É de salientar que  $\bar{\mathbf{x}}$ , representa a média simples dos dados. Neste caso, a matriz  $\mathbf{A}$ , induz a norma de Mahalanobis no espaço  $\mathbf{R}^n$ .

A implementação do algoritmo é realizada da seguinte forma:

Seja  $\mathbf{X}$ , a matriz dos dados e  $c$ , o numero de “clusters” com a condição,  $1 < c < N$ ,  $m$ , o peso do expoente em que  $m > 1$ , a tolerância de erro  $\varepsilon > 0$  e a matriz  $\mathbf{A}$ . Inicia-se a matriz de partição de forma aleatória para que  $\mathbf{U}^{(0)} \in M_{fc}$ .

Cria-se um ciclo repetitivo para  $l = 1, 2, \dots$  para os seguintes passos,

**Passo 1** – Calcular os “clusters” protótipos (“means”):

$$\mathbf{v}_i^{(l)} = \frac{\sum_{k=1}^N (\mu_{ik}^{(l-1)})^m \mathbf{x}_k}{\sum_{k=1}^N (\mu_{ik}^{(l-1)})^m}, \quad 1 \leq i \leq c$$

**Passo 2** – Cálculo das distâncias:

$$D_{ikA}^2 = (\mathbf{x}_k - \mathbf{v}_i)^T \mathbf{A} (\mathbf{x}_k - \mathbf{v}_i), \quad 1 \leq i \leq c, \quad 1 \leq k \leq N$$

**Passo 3** – Recalcular a matriz de partição:

$$\mu_{i,k}^{(l)} = \frac{1}{\sum_j^c \left( \frac{D_{ikA}}{D_{jkA}} \right)^{\frac{2}{m-1}}}$$

Fim do ciclo:

$$\|\mathbf{U}^{(l)} - \mathbf{U}^{(l-1)}\| < \varepsilon$$

### 3. *Facebook*

O *Facebook* é um sítio da internet onde existe um serviço de relacionamento social que foi criado por *Mark Zuckerberg*, em conjunto com três colegas da universidade de *Harvard*, a 4 de Fevereiro de 2004.

Inicialmente designado por “*Thefacebook*” este sítio estava limitado ao âmbito populacional de *Harvard*, sendo posteriormente alargado o universo de utilizadores, começando pelo Instituto de Tecnologia de Massachusetts (MIT), a Universidade de *Boston*, o *Boston College* e todas as escolas *Ivy League*. Só em Setembro de 2005 é que foi lançada uma versão para o universo das escolas secundárias e em Setembro de 2006 o sítio, foi disponibilizado para todos os utilizadores com idade igual ou superior a 13 anos que tenham um “email” válido, condições esta que se mantêm até hoje.

O *Facebook* tornou-se num sítio de grande popularidade que rapidamente passou a ser uma ferramenta de rentabilidade financeira, sendo que em 2011 prevê-se uma facturação de 4.300 mil milhões de US\$, com um resultado líquido de cerca de 1.000 milhões, e um número de utilizadores activos de cerca de 800 milhões. Número este, que segundo a empresa *Nielsen*, e se a tendência de crescimento se mantiver, atingirá em 2012, o número de mil milhões de utilizadores. Isto significa que cerca de 14% da população mundial entrará no *Facebook* pelo menos uma vês por mês.

Segundo uma empresa de estudos de mercado - *Hitwise Pty* - no dia 16 de Março de 2010, o *Facebook* teve mais visitas ao seu sítio do que o próprio *Google*, e em Agosto de 2011, o

## Segmentação de indivíduos no *Facebook* que gostam de música

*Facebook* ultrapassou a *Yahoo!* . Tornou-se assim na empresa número um, em termos de receitas pela exibição de anúncios *online*, atingindo cerca de 16,3% da quota de mercado, tal como demonstram os estudos que a empresa *EMarketer* realizou.

Existem alguns marcos importantes na história do *Facebook* que o tornaram naquilo que é hoje.

O primeiro foi passar da exclusividade do ecossistema universitário para a população em geral, o que permitiu a massificação e portanto o crescimento fulminante desta rede social.

Outro facto importante foi a colocação do botão gosto (*like*). Até 2009, o *Facebook* possuía apenas um botão para partilhar os conteúdos (*Share*). Esta função para além de confundir os utilizadores, não era exibida no ” *news feed*” e assim perdia-se a informação perante os elementos do grupo de amigos. Em 2010, o *Facebook*, inseriu o botão gosto (*Like*), que para além de ser mais bem compreendida pelos utilizadores, facilita a partilha ao exibir o conteúdo no perfil e no mural. Hoje em dia esta funcionalidade é amplamente explorada por empresas e portais de notícias, sendo que é uma das marcas registrada do *Facebook*.

Também o lançamento do filme *Facebook* , em 2010, foi um marco importante na história e divulgação desta rede social, passando mesmo a ser a rede social mais utilizada no mundo. O filme causou alguma polémica ao mostrar um Mark Zuckerberg sem escrúpulos e obstinado com o sucesso, sem que o próprio tenha sido consultado.

No entanto a característica mais importante, foi o aparecimento das “*news feeds*”, pois foi a forma mais eficiente de os utilizadores ficarem fidelizados com o *Facebook*, já que de uma forma constante alimentam “*news feeds*” e porque estando acessível na página inicial, torna-se uma forma de cativar a utilização constante do *Facebook*, como instrumento de comunicação.

Outro dos momentos mais marcantes na história do *Facebook*, foi o facto de ter tido um papel muito importante na chamada “Primavera Árabe”. No final de 2010, o *Facebook* foi considerado um dos principais pilares na chamada Primavera Árabe. Nesta ocasião

## Segmentação de indivíduos no *Facebook* que gostam de música

milhares de cidadãos de países do Oriente Médio, que vivem sob regimes autoritários, realizaram revoluções com a ajuda da rede social para comunicar os seus ideais, organizar e convocar a todos para os protestos.

Mas existem também diversas razões que justificam que o *Facebook* continue a crescer da forma a que está a crescer neste momento.

Uma das razões é que o *Facebook* acumula informação sobre os seus utilizadores ao contrário dos seus rivais como a *Google*. O *Facebook* está continuamente a coleccionar todo o tipo de informação sobre os seus utilizadores, fazendo com que os analistas consigam retirar cada vez mais e melhor informação sobre o perfil de hábitos e costumes dos seus utilizadores. Assim a tarefa de conseguir segmentar os utilizadores está cada vez mais facilitada e precisa, o que faz com que os anúncios colocados pelos fabricantes de produtos e serviços sejam cada vez mais orientados. As empresas vêem o *Facebook* como um canal de promoção das suas marcas e em Agosto de 2011 o *Facebook* destronou a Yahoo! e tornou-se a empresa líder em colocação de anúncios online nos Estados Unidos da América.

Outra razão é que o *Facebook* é actualmente a plataforma mais importantes no que se refere ao *marketing* em redes sociais. Através da colocação da função “gosto” o *Facebook* conseguiu que a persuasão de uma compra através de recomendação de um amigo fosse mais poderosa de que qualquer outra forma convencional. Assim existem hoje milhões de interacções no *Facebook* de muitas pessoas em quase tudo, desde música até a influências de tendências de voto.

Mas não é o facto de o *Facebook* ter um tão grande número de utilizadores que faz com seja apetecível pelas empresas de marketing. São todas as informações sobre as ligações sociais e afinidades das pessoas, gostos e preferências que fazem com que o *Facebook* tenha um conhecimento tão vasto das pessoas que constituem a sua base de dados. Segundo *Nicolas Mendensohn*, do *Institute of Practitioners in Advertising*, o *Facebook* dispõe de uma base de dados de interesses e acções, pensamentos e opiniões das pessoas que constituem o universo do *Facebook*, que faz com que o mesmo seja muito apetecível para as empresas de *marketing*.



## Segmentação de indivíduos no *Facebook* que gostam de música

No entanto, o rápido crescimento do *Facebook* só foi possível devido ao também rápido crescimento dos pontos de conectividade da *internet*, o que faz com que cada vez mais pessoas se consigam conectar hoje em dia. Segundo um estudo de 2011, da *Boston Consulting Group*, em 2010 existiam cerca de 1.6 mil milhões de pessoas com acesso à *internet* no mundo inteiro, e prevê-se que este número seja em 2016 de cerca de 3 mil milhões de utilizadores.

Um outro factor de crescimento, foi o facto de nos últimos anos ter havido um desenvolvimento muito grande na área das ligações móveis, através nomeadamente do telemóvel, bem como das aplicações ligada aos mesmos. Assim sendo, e segundo um estudo que *Booz&Company* efectuou em 2011, o número de utilizadores que actualmente utilizam comunicações móveis (telemóvel, *tablet*, etc) é de cerca de 450 milhões de pessoas, sendo que se estima que em 2015 sejam perto de 1.000 milhões de utilizadores.

O *Facebook* tem um impacto na vida social e nas actividades das pessoas em diversas formas. Através da mobilidade tecnológica actual (telemóveis e outros dispositivos móveis), os utilizadores conseguem de uma forma continua estar em contacto com os amigos, família e outros utilizadores. Assim o impacto social que o *Facebook* provocou é enorme pois criou uma plataforma onde diversas pessoas se podem juntar de uma forma virtual em torno de assunto e interesses diversificados.

No entanto existem opiniões que dizem que o *Facebook* está a originar uma tendência anti-social devido ao facto de as pessoas não comunicarem directamente, devido à falta de comunicação visual directa.

A informação armazenada no *Facebook* por alguns utilizadores tem periodicamente chamado o tema da privacidade para a primeira página dos jornais, bem como as guerras com os reguladores e com os utilizadores.

O *Facebook* está bloqueado em países como a China, Irão Uzbequistão, Paquistão e Síria devido a motivos religiosos ou por alegadamente conterem valores discriminativos.

Atendendo a que o modelo de negócio do *Facebook*, está predominantemente assente na exibição de anúncios, existe a tentação de utilizar cada vez mais a informação existente na

## Segmentação de indivíduos no *Facebook* que gostam de música

base de dados. O maior activo que o *Facebook* tem são os seus utilizadores e em particular a informação dos mesmo que está depositada na sua base de dados.

Um dos problemas que o *Facebook* actualmente enfrenta é até que ponto pode utilizar a informação que possui na base de dados sobre os seus utilizadores para assim conseguir disponibilizar informação mais precisa aos seus clientes.

A comissão de comércio americana (*America's Federal Trade Commission*), obrigou o *Facebook* a submeter-se a uma auditoria externa duas vezes por ano, para analisar as políticas e procedimentos da privacidade dos dados dos utilizadores constantes na base de dados.

Uma das ameaças que o *Facebook* enfrenta actualmente é a crescente desconfiança que os utilizadores têm sobre a privacidade dos seus dados na base de dados. Se os utilizadores começarem a sentir que os seus dados estão a ser fornecidos de uma forma generalizados às empresas, por certo que o maior activo do *Facebook* desvaloriza muito.

Do universo das redes sociais que estão activas actualmente (*Facebook, Myspace, Orkut, etc.*), o *Facebook* atendendo à sua dimensão e por estar tão disseminada na população portuguesa, foi a escolha lógica para a execução deste trabalho.

## 4. Metodologia de Trabalho

### 4.1. Processo de Obtenção dos Dados

Para obtenção dos dados para a realização dos questionários foi utilizado a plataforma do *Facebook*. Inicialmente foram seleccionados os grupos de pessoas que gostam de música com maiores utilizadores.

No processo de selecção desses grupos, utilizou-se uma ferramenta disponível no *Facebook*, em que a mesma permite saber qual é a dimensão da população que se enquadra na descrição que é realizada.

## Segmentação de indivíduos no *Facebook* que gostam de música

Os parâmetros a seleccionar podem ser demográficos, idade, país, nível de escolaridade, entre outros e nos interesses, o *Facebook* possibilita a introdução directa dos nomes dos grupos sobre os quais se pretende fazer o estudo.

Assim foram escolhidos os seguintes grupos de pessoas que têm mais de 12 anos (idade mínima para se ter conta no *Facebook*), que vivem em Portugal e que gostam de ouvir música, seguindo a metodologia atrás referida (anexo 1).

O resultado da pesquisa foi de um universo de 51.760 indivíduos, à data de 16 de Maio de 2011.

O passo seguinte, foi o envio do questionário às pessoas que faziam parte dos grupos. Após a escolha dos grupos com o maior número de elementos que se encontraram no *Facebook*, foi-lhes enviado um convite, em que se incluía o endereço da recolha do questionário.

No convite era salientado que se tratava da realização de um estudo de mercado de um estudante de uma universidade portuguesa (ISEG), o tempo estimado para a realização da mesma (5 minutos) e que a recolha dos dados seria estritamente confidencial.

No fim era dado um endereço de email para estabelecimento de contacto se os utilizadores assim o entendessem para colocar questões ou comentários.

### 4.2. Questionário

Vários factores podem influenciar o uso de serviços ou meios para a audição de música bem como as suas preferências. Para as pessoas é expectável que um serviço ou meio seja funcional, social e que dos mesmos se possam recolher benefícios, (*Sheth e Newman, 1991*).

Adicionalmente as pessoas têm em consideração, para além das suas preferências, outros factores tais como o preço, a percepção do risco (*Holak, 1990*), os custos de mudança

(*Klempere*, 1987), o medo da complexidade da tecnologia (*Higgins*, 1992), e o custo da aprendizagem de uma nova tecnologia (*Shugan*, 1980).

As pessoas tendem a tomar a decisão de usar um serviço ou meio de ouvir música, influenciadas pelas suas atitudes, interesses, e opiniões sobre as plataformas de uso (*Florence*, 1990 e *Teo*, 1999), tais como computadores, meios portáteis de audição, situações de uso (*Ratneshwar*, 1991) e também tendo em conta os seus dados demográficos.

Baseado na literatura anterior e em sugestões “*lead-users*” (*von Hippel*, 1986), o questionário foi desenhado, para captar os interesses nos diversos tipos de música, na frequência de audição dos referidos tipos de música, opiniões sobre a própria música, frequência na aquisição de música, opiniões sobre as plataformas de audição, as actividades relacionadas com a audição de música e também os dados demográficos.

No anexo 2, é apresentado o questionário que foi proposto.

A duração da recolha dos questionários foi de aproximadamente um mês (16 de Maio de 2011 a 20 de Junho de 2011).

Para a colocação dos questionários *online*, foi utilizada a plataforma do Google, o *Google docs*, e foi sobre esta plataforma que se desenhou todo o questionário. Utilizou-se esta plataforma pois para além de ser gratuita incorpora todo o ambiente do Google, tais como as contas do *gmail*, de grande capacidade e outras características úteis.

## 5. Resultados Obtidos

### 5.1. Recolha dos Questionários

Os questionários foram recolhidos na plataforma do Google, o *Google docs*, tal como foi referido anteriormente e esteve na fase de recolha até ao dia 20 de Junho de 2011. Neste período de tempo, foram validados 251 questionários.

Quando foi construído o questionário, foi seleccionado a hipótese de resposta obrigatória pelo que não houve respostas incompletas.

## 5.2. Pré-Tratamento dos Dados

Neste trabalho pressupõe-se que as pessoas que responderam ao questionário, aceitaram as condições de acesso ao *Facebook*, em particular a idade, que tem que se ter obrigatoriamente maior ou igual a 12.

Assim, no caso da Idade, reagrupou-se as classes de idades de 12 anos a 20 anos e as de 21 anos a 30 anos numa só classe que se chamou menores ou igual a 30 anos, (anexo 3).

Também se reagrupou as classes de idades de 41 anos a 50 anos, de 51 anos a 60 anos e a classe de maiores de 60 anos, em uma única classe que se chamou de maiores ou iguais a 41 anos.

Em relação ao Estado Civil, agrupou-se a classe Divorciado com Outro a que se chamou Divorciado.

Quanto à Escolaridade, atendo às respostas que se obtiveram, agrupou-se as classes do 9º Ano com a 12º Ano e ainda a classe Outro, a que se intitulou  $\leq 12^\circ$  Ano. Pela mesma razão, agrupou-se a classe de Mestrado e Doutoramento a que se chamou de Mestrado/DR.

## 5.3. Tratamento dos Dados

### 5.3.1. Redução de Dados

Depois da recolha dos 251 questionários, atendo à quantidade de questões a analisar, foi necessário recorrer a uma redução dos dados, antes de se iniciar o processo de segmentação quer utilizando a técnica *k-means* quer a técnica *Fuzzy C-Means*.

A redução de dados identifica o subconjunto óptimo das características de indivíduos mas com a garantia que não se perde a informação essencial. Este processo elimina a redundância e características menos importantes, sendo que assim se reduz bastante o “esforço” computacional ao reduzir a dimensão dos dados.

Adicionalmente à redução do esforço computacional, e porque agrupam características semelhantes, torna o processo de legibilidade e interpretação dos resultados mais fácil.

O objectivo do processo de redução de dados é identificar novas variáveis, em menor número que o conjunto inicial, mas sem perda significativa da informação contida neste conjunto, (Reis, 1990).

Dos vários métodos que existem para a redução de dados, e após consulta de vários *papers*, o método “*Principal Components Analysis*”, foi considerado o melhor em relação a outras técnicas lineares e não lineares de mapeamento, (Gelseme, 1980).

O PCA, é uma técnica de análise exploratória multivariada que transforma um conjunto de variáveis correlacionadas entre si, num conjunto menor de variáveis independentes, combinações lineares das variáveis originais, designadas por componentes principais, (Maroco, 2007).

### 5.3.2. *Outliers*

Antes da aplicação da técnica PCA, foi verificada a existência de “*outliers*” através da técnica de “caixa de bigodes”. Utilizou-se o SPSS para a detecção dos “*outliers*”.

Como se pode observar no Anexo 4, não existem *outliers* severos sendo que todas as variáveis apresentam variabilidade.

No entanto existem *outliers* moderados como por exemplo na questão de Música Brasileira e Música Clássica, mas não afectam o processo de redução de dados PCA.

### 5.3.3. Análise aos Resultados do PCA

Para a utilização da técnica PCA, utilizou o SPSS e efectuou-se a análise sobre o grupos de questões 1.1 a 1.9 (Interesses em tipos de música), 3.1 a 3.8 (Opiniões sobre a música) e 5.1 a 5.8 (Opiniões sobre as plataformas de audição de música).

## Segmentação de indivíduos no *Facebook* que gostam de música

Na produção da técnica foi parametrizada para a rotação *varimax* (ortogonal), e foram inicialmente extraídas componentes com valores próprios maiores que a unidade (Churchill, 1995).

Foram também utilizadas as métricas de *Kaiser-Meyer-Olkin* e o teste de esfericidade de *Bartlett*.

A validação da extracção dos factores foi medida através da fiabilidade, validade discriminante e convergência. A fiabilidade é mensurada através do alfa de *Cronbach*, sendo que a validação da convergência requer que todas as perguntas se apoiem fortemente numa componente, isto é, os valores das perguntas referentes à componente têm que ser elevados.

A validação discriminante pressupõe que as perguntas que não fazem parte da componente tenham, valores baixos.

Como se podem observar no anexo 5, para todos os grupos de questões, todas as perguntas apoiam-se fortemente nas componentes (validação convergência) e apoiam-se de uma forma fraca nos outros (validação discriminante). Os valores acima de 0,5 estão assinalados a negrito.

As comunalidades (anexo 6), representam a proporção de variância das variáveis explicadas pelas componentes, sendo que a média das mesmas está acima de 0,7 para o grupo de Interesses em tipos de música, acima de 0,8 para os grupos de Opiniões sobre a Música e Opiniões sobre as plataformas de audição de música.

No grupo de questões Opiniões sobre a música, com os pressupostos iniciais, componentes só com valores próprios acima da unidade, verificou-se que havia comunalidades abaixo do valor de 0,5 bem como os valores da matriz de rotação não respeitavam os princípios de validação (convergência e discriminação). Assim sendo e porque a componente mais próxima tinha um valor muito próximo de 0,9, forçou-se mais uma componente, (anexo 7).

Segmentação de indivíduos no *Facebook* que gostam de música

O mesmo se passou para os grupos de questões Opiniões sobre as plataformas de audição de música.

De salientar que o KMO, com excepção do grupo de questões Interesses em tipos de música, se encontra ao nível de médio.

#### 5.3.4. Descrição dos Subconjuntos após PCA

Foram apurados 12 subconjuntos, após a aplicação da técnica PCA aos dados do questionário, como se pode observar no anexo 8.

Os subgrupos são os seguintes:

Música "Difícil": Clássica + *Jazz*

Música Pesada: *Hard, Havy, Punk* e Metal + Alternativa, *Indie*

Música Popular: Música Portuguesa + Música Brasileira + *Pop Rock* Internacional

Música Dança: Dança, *House e Electro* + *Soul, Funk, R&B e Hip-Hop*

Acesso à música: O acesso fácil através da internet à música é muito importante para mim  
+ O acesso gratuito a *sítios* da internet de música é muito importante para mim  
+ O acesso rápido através da internet à música é muito importante para mim

Valor de aquisição: A música deve ter um preço baixo de venda + Os *sítios* de partilha de música na internet deveriam ser permitidos por lei

Baixa de valor: As músicas deveriam ter o valor do IVA reduzido + Eu tenho receio em comprar música via internet

Direitos de autor: Os direitos de autor deveriam ser preservados

Acesso à Informação: Eu gosto mais de ouvir música no carro do que nos aparelhos pessoais + Costumo procurar informação sobre os autores e músicas



Segmentação de indivíduos no *Facebook* que gostam de música

que oiço na imprensa + Costumo procurar informação sobre os autores e músicas que oiço na internet

Dificuldades com a Tecnologia: Eu tenho dificuldade na utilização dos aparelhos portáteis: *iPod*, mp3 + Eu acho muito complicado o carregamento dos ficheiros para aparelhos portáteis

Partilha: Gosto de partilhar nas redes sociais os meus gostos sobre as músicas que oiço + Quando gosto da música, clico no item "eu gosto" no *Facebook*

Individualismo: Quando ouço música, não partilho com mais ninguém

### 5.3.5. Processo de Segmentação

O propósito nesta fase do processo é o de identificar segmentos homogéneos baseados na extracção de factores utilizando a técnica PCA.

Neste estudo exploratório usou-se dois tipos de segmentação, ambos do tipo não hierárquico, sendo um não sobreposto (*k-means*) e outro sobreposto (*Fuzzy*) que posteriormente se irá comparar os resultados.

#### 5.3.5.1. Segmentação *K-means*

Para a realização da segmentação do tipo não hierárquico, não sobreposto, utilizou-se o SPSS e começou-se por determinar o número de *clusters*. A técnica mais popular é a de realizar a segmentação utilizando a análise hierárquica de clusters, utilizando o método de segmentação de variância mínima (*Ward's*). A utilização deste método embora afectada negativamente por outliers é pouco sensível pelo ruído dos dados (*Kamakura, 2000*). Após estudos efectuados, verificou-se que o método da variância mínima é o melhor dos métodos tendo em conta uma grande variedade de factores que afectam o propósito da segmentação, (*Punj e Stewart, 1983*).

Utilizando o Excel constrói-se um gráfico (anexo 9) com os coeficientes de aglomeração e o número de clusters é determinado quando a diferença entre coeficientes começa a estabilizar.

Como se pode constatar na análise ao gráfico, não é clara a determinação do número de *clusters*, sendo que o valor poderia ser três ou seis. Para se esclarecer o número de clusters, utilizou-se a análise discriminante e chegou-se à conclusão que 96,8 % dos casos são validados contra os 94,4 % dos casos validados, de três e seis clusters respectivamente, como se pode constatar no anexo 10.

#### 5.3.5.2. Segmentação *Fuzzy*

Neste tipo de segmentação o número de *clusters* bem como o peso do expoente ( $m$ ) têm que ser determinados *a priori* e são obrigatórios antes do início do algoritmo. A determinação do número de *clusters*, sempre foi um dos parâmetros deste algoritmo que mais influência o algoritmo, (Bezdek, 1995). Assim sendo, uma das técnicas mais usuais e sugerida na literatura, (Steenkamp, 1993), para a determinação *a priori* do número de *clusters* é a adição do número de segmentos e verificar se a função objectivo não se altera substancialmente, (Krieger, 1996).

Como se pode observar no gráfico do anexo 11, o “cotovelo” tem a curva no número de clusters igual a três e atendo ao que foi descrito atrás o número de clusters escolhido foi três.

Se o valor de  $m$  for próximo de 1, como se pode verificar segundo as equações 1.22 e 1.23, os coeficientes da função *membership*, ou são um ou são nulos, o que significa que os elementos dos segmentos são distintos entre segmentos. Por outro lado, se o valor de  $m$  for muito alto, os coeficientes da função *membership* têm todos o mesmo valor de  $(1/c)$ , o que significa que se sobrepõem na totalidade e conseqüentemente não existe variação entre segmentos.

Como o valor de  $m$ , depende do número de *clusters*, foi realizada uma análise empírica para a determinação do valor óptimo de  $m$ , para vários números de clusters, utilizando o erro e a função objectivo. Os resultados demonstram que o valor  $m$ , deve estar entre o valor 1.25 e 1.15, sendo que o valor adoptado foi 1.2.

## 6. Análise dos Resultados

Neste estudo foram utilizadas duas técnicas de segmentação, ambas não hierárquicas, sendo que uma não é *overlapping* (*Kmeans*) e a outra é *fuzzy* (FCM).

Ambos os processos de segmentação concluíram que o número de *clusters* óptimo era de três.

### 6.1. *K-means*

Como se pode observar no gráfico (Anexo 12) as pessoas diferem em relação aos interesses por tipos de música, opiniões e tipos de plataformas de audição de música.

Na caracterização dos *clusters* em função das variáveis de input, podemos observar que no *cluster* 1 existe uma menor importância na música “pesada” e música de dança ao invés da música popular que tem uma importância maior em relação à média. Já relativamente à música “difícil”, embora a importância seja abaixo da média, esse valor é insignificante pelo que será considerado na média.

No que diz respeito ao acesso à música, o *cluster* 1 apesar de dar uma importância abaixo da média, o valor é reduzido pelo que se considera dentro da média. O mesmo não se passa em relação ao valor de aquisição onde a importância é significativamente abaixo da média e o inverso se passa em relação à baixa de valor onde aqui a importância é significativamente acima da média.

*Relativamente* aos direitos de autor a importância está dentro da média.

## Segmentação de indivíduos no *Facebook* que gostam de música

O *cluster* 1 tem uma importância acima da média no diz respeito ao acesso à informação e em particular á dificuldade com a tecnologia onde tem uma importância manifestamente acima da média.

No que diz respeito à partilha e individualismo, releva uma importância abaixo da média.

No *cluster* 2 a música “difícil” é a única que tem uma importância acima de média, pois todas as outras têm uma importância abaixo da média em particular a música de dança.

No que se refere ao *cluster* 3 a música “Pesada” e a música de dança têm uma importância acima da média sendo que a música popular está dentro da média. A música “difícil” tem uma importância abaixo da média mas o valor não é muito baixo, pelo que se considera dentro da média.

O acesso à música no *cluster* 3 tem uma importância acima da média assim como no valor de aquisição, sendo que na baixa de valor a importância é abaixo da média. No que diz respeito aos direitos de autor o valor está dentro da média.

Em relação ao acesso à informação, a importância que o *cluster* 3 tem, está acima da média embora não tão significativamente quanto á partilha. Nas dificuldades com a tecnologia a importância está abaixo da média sendo que no individualismo está dentro da média.

Segundo a caracterização dos clusters em função das variáveis demográficas (anexo 15), o *cluster* 1 é composto maioritariamente por pessoas jovens, abaixo dos 30 anos, predominantemente do sexo feminino, solteiras e com uma escolaridade predominantemente ao nível da licenciatura. A este grupo de indivíduos chamam-se “*Mulheres*”.

O *cluster* 2 há uma predominância de indivíduos acima dos 31 anos, predominantemente do sexo masculino, não solteiros e com elevado grau de escolaridade ao nível de licenciatura, mestrados e doutoramentos. A este grupo de indivíduos chamam-se “*Maduros*”.

Em relação ao *cluster* 3 podemos caracterizá-lo do ponto de vista das variáveis demográficas como maioritariamente jovens, abaixo dos 30 anos, predominantemente do

## Segmentação de indivíduos no *Facebook* que gostam de música

sexo masculino e solteiro, com elevado grau de escolaridade. Estes grupos de indivíduos chamam-se “*Jovens*”.

### 6.2. FCM

Após a utilização do algoritmo *Fuzzy*, os graus de pertença de cada indivíduo aos *clusters* variam entre zero e um como se pode constatar no quadro 10. A lista completa está no anexo 17.

**Quadro 9 – Graus de Pertença de Cada Indivíduo nos *Clusters***

Indivíduo	Cluster 1	Cluster 2	Cluster 3
1	0,1056	0,1236	<b>0,7708</b>
2	0,2050	<b>0,6777</b>	0,1172
3	0,1177	<b>0,3411</b>	<b>0,5413</b>
4	0,2402	<b>0,5677</b>	0,1921
5	0,0257	0,0593	<b>0,9150</b>
6	0,1281	<b>0,6005</b>	0,2714
7	<b>0,7340</b>	0,1324	0,1337
8	0,1682	<b>0,4934</b>	<b>0,3384</b>
9	0,1085	<b>0,6621</b>	0,2294
10	0,3116	0,0228	<b>0,6656</b>
11	0,2262	<b>0,3532</b>	<b>0,4206</b>
12	0,0061	<b>0,9886</b>	0,0053
13	0,0878	0,2833	<b>0,6289</b>
14	<b>0,3488</b>	<b>0,3828</b>	0,2684
15	0,0227	0,0173	<b>0,9599</b>
16	<b>0,8599</b>	0,0575	0,0825
17	0,0473	0,2087	<b>0,7440</b>
18	<b>0,4746</b>	0,1040	<b>0,4213</b>
19	<b>0,6338</b>	0,0702	0,2960
20	0,1510	<b>0,7198</b>	0,1292
...	...	...	...
235	<b>0,4223</b>	<b>0,3511</b>	0,2266
236	0,0769	<b>0,7432</b>	0,1799
237	0,0699	0,0607	<b>0,8694</b>
238	<b>0,4303</b>	0,2986	0,2711
239	0,1383	<b>0,7810</b>	0,0807
240	<b>0,4351</b>	0,0959	<b>0,4689</b>
241	0,2390	0,0896	<b>0,6714</b>

## Segmentação de indivíduos no *Facebook* que gostam de música

242	0,0698	0,2266	<b>0,7037</b>
243	0,3190	<b>0,5531</b>	0,1280
244	0,1319	0,2145	<b>0,6535</b>
245	0,2032	<b>0,5868</b>	0,2100
246	<b>0,8503</b>	0,0136	0,1362
247	0,1601	<b>0,7712</b>	0,0687
248	0,1621	<b>0,7090</b>	0,1289
249	0,0493	0,1387	<b>0,8120</b>
250	0,0080	<b>0,9716</b>	0,0204
251	0,0434	<b>0,8786</b>	0,0779

Esta tabela sugere que existem diferentes tipos de graus de pertença e pode-se constatar o seguinte:

- a) Existem indivíduos que se posicionam claramente mais perto de um *cluster* do que outros *clusters*. Por exemplo o indivíduo 5 tem um grau de pertença de 0,9150 no *cluster* 3, enquanto o indivíduo 12 tem um grau de pertença de 0,9886 no *cluster* 2.
- b) Alguns indivíduos partilham características em mais do que um *cluster* com aproximadamente o mesmo grau de pertença. Se reparáramos no indivíduo 14, este tem um grau de pertença de 0,3488 no *cluster* 1 e 0,3828 no *cluster* 2.
- c) Pode haver no entanto, indivíduos que têm grau de pertença diferentes e que pertencem a *cluster* diferentes. Por exemplo o individuo 3, tem um grau de pertença de 0,3411 no *cluster* 2 e um grau de pertença de 0,5411 no *cluster* 3.

Observando o gráfico (Anexo 13), também se constata que as pessoas diferem em relação aos interesses por tipos de música, opiniões e tipos de plataformas de audição de música.

Fazendo a caracterização dos *clusters* em função das variáveis de input, podemos observar que no *cluster* 1, os indivíduos dão uma preferência acima da média, e é a maior de todos os outros, à música “difícil” e à música pesada, sendo que nesta última a diferença em relação à média é significativa. As preferências dos indivíduos deste *cluster*, em relação à música de dança é inferior à media e fica dentro da média no diz respeito à musica popular.

O *cluster* 1 dá uma importância acima da média em relação ao valor de aquisição, direitos de autor e acesso à música, sendo que é mais significativa neste último. No entanto, a

importância que este *cluster* dá relativamente à baixa de valor, encontra-se dentro da média. De salientar que os indivíduos do *cluster* 1 dão uma importância muito acima da média ao acesso à informação e estão abaixo da média no que diz respeito às dificuldades com a tecnologia. Do mesmo modo, relativamente à partilha e individualismo o *cluster* 1 dá uma importância acima da média.

Em relação ao *cluster* 2, os indivíduos que nele estão agrupados não têm uma grande preferência por música de dança, música pesada ou música difícil sendo que nesta última esteja próximo do valor médio. A música popular, embora revele uma preferência acima da média, não é significativa, e portanto considera-se dentro da média. Denota-se abaixo da média, o acesso à música e ao valor de aquisição, sendo que é significativamente maior no primeiro caso.

O *cluster* 2 dá uma preferência acima da média em relação à baixa de valor e em relação aos direitos de autor considera-se dentro da média.

Quer a partilha quer o acesso à informação têm uma importância menor em relação à média sendo maior nesta última.

Os indivíduos agrupados no *cluster* 2 têm manifestas dificuldades com a tecnologia ficando muito acima da média e são mais individualistas do que a média.

Em relação ao *cluster* 3 as preferências pelos tipos de música vão indiscutivelmente para a música de dança, onde estão muito acima da média. O mesmo já não se pode dizer da música pesada pois estão abaixo da média.

A música “difícil” tem uma ligeira preferência neste *cluster*, assim como a música popular está ligeiramente abaixo das preferências dos indivíduos do *cluster* 3.

O acesso música tem uma preferência acima da média neste cluster e a baixa de valor assim como os direitos de autor têm uma preferência abaixo da média neste cluster, sendo que este último é significativo. Assim sendo para os indivíduos deste cluster, a importância da preservação dos direitos de autor não é muito relevante.

## Segmentação de indivíduos no *Facebook* que gostam de música

Apesar do valor de aquisição ser algo que está abaixo da média neste *cluster*, como o valor é reduzido, vamos considerar que está dentro da média.

Para os indivíduos que pertencem ao *cluster* 3, o acesso à informação não é muito importante, quando comparado com a média dos outros *clusters* assim como não são individualistas e não têm dificuldades com a tecnologia. Em relação à partilha de preferência dos seus gostos musicais, são adeptos da mesma.

O *cluster* 1, caracteriza-se em função das variáveis demográficas (anexo 16), como composto por indivíduos com idade maioritariamente acima dos 30 anos, predominantemente do sexo masculino e solteiros com elevado grau de escolaridade quer ao nível da licenciatura quer ao nível de mestrados e doutoramentos. Este grupo de indivíduos denomina-se “Solteirões”.

Já o *cluster* 2 caracteriza-se com um grupo de indivíduos com idade maioritariamente acima dos 30 anos, maioritariamente feminino, não casado e com elevado grau de escolaridade ao nível de mestrados e doutoramentos. Este grupo de indivíduos chama-se “Mulheres Maduras”.

O *cluster* 3 é composto maioritariamente por jovens, abaixo dos 30 anos, maioritariamente do sexo masculino, predominantemente solteiros e com maior grau de escolaridade comparativamente aos outros *clusters*. Estes grupos de indivíduos chamam-se “Jovens”.

### 6.3. Análise às frequências de modos, aquisição e audição de música

No anexo 14 estão representadas as frequências de modos, aquisição e audição de música, utilizado o algoritmo *K-means* e o algoritmo FCM.

Como se pode observar no processo em foi utilizado o algoritmo *K-means*, é no *cluster* 3 que se ouve maioritariamente, acima de 60%, mais música portuguesa quando o período de audição é uma semana. Nos tipos de música Brasileira e Pop maioritariamente é neste *cluster* que se ouve mais mas com o período de tempo a ser de um mês ou raramente. De salientar que o *cluster* 3 dá uma importância ao tipo de música popular igual á média.



## Segmentação de indivíduos no *Facebook* que gostam de música

No *cluster 1*, em relação à frequência de audição no tipo de música pesada, a mesma é muito baixa nos períodos mais curtos (diariamente e 2 a 3 x semana) o que se enquadra na análise de *cluster* que se fez anteriormente, pois este tipo de cluster tem uma aversão a este tipo de música.

Em relação à frequência de aquisição de música em lojas *online*, é no *cluster 3*, que se evidencia o maior valor sendo que é total no período de tempo diariamente.

De salientar que o *cluster 2* é predominantemente o grupo de indivíduos que compra com mais frequência nas lojas especializadas de música e em sítios de específicos de autor. Esta tendência vai de encontro á análise de *clusters* que se fez, pois será nestes dois lugares onde se pode encontrar mais facilmente música difícil.

No que diz respeito à frequência nos modos de audição de música, é predominantemente o *cluster 3* que mais frequentemente ouve rádio, sendo que o *cluster 1* é que menos frequentemente vai a concertos de música ao vivo.

É de salientar que o *cluster 3*, que tem uma faixa etária mais baixa de todos os outros *clusters* é no entanto o que mais regularmente ouve música através de discos em vinil.

Fazendo agora a análise às frequências de audição tendo como base a segmentação do tipo *Fuzzy*, observa-se que é no *cluster 1*, que se ouve mais frequentemente a tipo de música pesada ao contrário do *cluster 3*. Este facto fica bem patente neste tipo de música na análise de *clusters* que se fez.

Quem adquire música em lojas especializadas de música com mais regularidade são os *cluster 2* e *3* sendo no entanto este último que também adquire com mais regularidade música nas lojas online e em sítios específicos, tais como o *iTunes* ou outros.

O *cluster 1*, onde predominam os indivíduos de sexo masculino e com uma faixa etária mais elevada, verifica-se uma maior frequência aquisição de música via sítios de partilha do tipo *emule*, *torrents* ou similares.

Em relação á audição de música, por via dos discos de vinil, o grupo de indivíduos que mais frequentemente ouve neste formato é o *cluster 1* sendo que através do computador a maior frequência está dividida entre o *cluster 1* e *cluster 3*.

## 7. Conclusões

Este capítulo tem por objectivo apresentar as principais conclusões do trabalho realizado e apresentado nos capítulos anteriores.

Para o efeito, começa-se por descrever as comparações entre os dois métodos e principais conclusões encontradas. De seguida faz-se referência a algumas limitações do presente trabalho e por último enumeram-se as linhas orientadoras de potencial trabalho futuro a realizar.

### 7.1. Comparação e Conclusões

O principal objectivo deste trabalho é o de comparar dois métodos de segmentação, *K-means* e *Fuzzy C-Means*, utilizando como base de estudo, as pessoas que gostam de música e que têm conta no *Facebook*.

Nesta abordagem exploratória aos interesses, opiniões sobre música e diversidade de plataformas para audição, conclui-se que não é indiferente a utilização do método *k-means* ou do método *FCM*. Apesar dos segmentos gerados serem bastante similares, a informação que se consegue extrair é diferente.

Ambos os métodos utilizados, originaram três segmentos, no entanto, no caso do método *FCM*, foi necessário recorrer à análise discriminante, para o apuramento do número de segmento.

Outra conclusão que se pode retirar deste estudo, é que existe uma sobreposição em cerca de 20% de elementos pertencente a mais do que um segmento, quando se utiliza o método *FCM*, enquanto no método *k-means* só se pode pertencer a um e um só segmento.

Neste estudo, conclui-se que utilizando o método *FCM*, ao contrário do método *k-means*, consegue-se quantificar quais são os elementos que pertencem a mais do que um segmento.

Outra conclusão que se obtém deste estudo, é o facto de quando se utiliza o método *FCM*, consegue-se retirar mais informação acerca do grau de pertença dos elementos, do que quando se utiliza o método *k-means*. Assim quando se utiliza o método *fuzzy sets*, consegue-se saber qual é o grau de pertença a cada um dos segmentos a que o elemento pertence.

Como os graus de pertença são quantificáveis, consegue-se hierarquizar no mesmo indivíduo os segmentos. Isto pode ser importante na definição de estratégia de venda de produto ou serviços aos consumidores por parte das empresas.

Quando se utiliza o método *k-means*, como método não sobreposto que é, só se consegue atribuir um segmento a cada indivíduo. Como se pode verificar no caso da utilização do método *FCM*, existem cerca de 20 % de casos em que existem indivíduos que pertencem a segmentos múltiplos. O não considerar esta informação, faz com que se desperdice 20% de vendas de produtos ou serviços a potenciais clientes.

A título de exemplo, quando se utiliza o método *k-means*, o segmento C2 (maduros), este trabalho constata que só existe um interesse acima da média no tipo de música “difícil”. Quando se utiliza o método *FCM*, o segmento equivalente C1 (Solteirões), não só gosta de música difícil como também gosta acima da média de música pesada, sendo que o interesse neste último tipo de música é muito acima da média.

Assim sendo, com este trabalho conclui-se que se retira mais informação dos elementos que constituem os segmentos quando se utiliza o método *FCM* do que quando se utiliza o método *k-means*.

## 7.2. Limitações do Trabalho

Uma das limitações deste trabalho, foi o tempo em que esteve o questionário “*on-line*”, que foi de aproximadamente de um mês. Seria desejável que o tempo de recolha das respostas dos questionários fosse maior.

Outra limitação foi que a recolha dos questionários teve como base indivíduos que têm conta no *facebook* pelo que não se considerou as pessoas que não têm, ainda conta nesta rede social.

### 7.3. Trabalhos Futuros

Durante o desenvolvimento deste trabalho e tendo por base as limitações do mesmo, foram identificadas várias possibilidades para futuras reflexões, no sentido de explorar e desenvolver outros trabalhos com o intuito de promover melhorias nas técnicas utilizadas. Em forma de sistematizar possíveis futuros trabalhos sugere-se:

- a) A utilização do método *Fuzzy* mas com novas medidas de semelhança. Assim podia-se realizar o mesmo estudo exploratório mas utilizando o algoritmo *Fuzzy Gustafson-Kessel* e/ou o algoritmo *Gath-Geva*;
- b) Fazer a comparação e analisar as diferenças entre o algoritmo *Fuzzy* e um algoritmo utilizando redes neuronais;
- c) Utilizar este tipo de estudo exploratório mas com variáveis de input de outro género, tais como a política ou os gostos futebolísticos.

## 1. Bibliografia

- Arabie, P. e L. Hubert (1994), “Cluster Analysis in Marketing Research”, *Advanced Methods of Marketing*, R.P. Bagozzi (ed.), Cambridge: Blackwell, 160-189
- Assael, H. (1970), “Segmenting Markets by Group Purchasing Behavior: An Application of the AID Technique”, *Journal of Marketing Research*, 7, 153-158
- Assael, H. e A. M. Roscoe, Jr. (1976), “Approach to Market Segmentation Analysis”, *Journal of Marketing*, 40, 67-76
- Balakrishnan, P. V., M. C. Cooper, V. S. Jacob and P. A. Lewis (1995), “A Study of the Classification Capabilities of Neural Networks Using Unsupervised Learning: A Comparison with K-Means Clustering”, *Psychometrika*, 59, 509-524
- Bass, F. M., D. J. Tigert e R. T. Lonsdale (1968) “Market Segmentation: Group versus Individual Behavior “, *Journal of Marketing Research*, 5, 264-270.
- Belk, R. W. (1975),”Situational Variables and Consumer Behavior”, *Journal of Consumer Research*, 2, 157-164
- Bezdek JC, Chiou ER. Core zone scatterplots: a new approach to feature extration for visual display. *Computer Vision Graphics and Image Processing* 1988;41:186-209
- Bezdek, J. C. (1974), “Numerical Taxonomy with Fuzzy Sets”, *Journal of Mathematical Biology*, 1, 57-71
- Bezdek, J. C., C. Coray, R. Gunderson e J. Watson (1981a), “Detection and Characterization of Cluster Substructure. I. Linear Structure: Fuzzy c-Lines”, *SIAM Journal on Applied Mathematics*, 40, 339-357
- Bezdek, J. C., C. Coray, R. Gunderson e J. Watson (1981b), “Detection and Characterization of Cluster Substructure. II. Fuzzy c-Varieties and Convex Combinations Thereof”, *SIAM Journal on Applied Mathematics*, 40, 358-372

- Breiman, L., J. H. Friedman, R. A. Olshen and C. J. Stone (1974), "Classification and Regression Trees", Belmont, CA: Wadsworth International Group.
- Carrol, J. D. e P. Arabie (1983), "INDLUS: An Individual Differences Generalization of the ADCLUS Model and MAPCLUS Algorithm", *Psychometrika*, 48, 157-169
- Churchill Jr GA. *Marketing Research: methodological foundations*, 6 th ed. Orlando, FL: The Dryden Press, 1995
- Currim, I. S., (1981), "using Segmentation Approaches for Better Prediction and Understanding from Consumer Mode Choice Models", *Journal of Marketing Research*, 18, 301-309
- DeSarbo, W. S. (1982), "GENNCLUS: New Models for General Nonhierarchical Clustering Analysis", *Psychometrika*, 47, 449-476
- DeSarbo, W. S. e V. Mahjan (1984), "Constrained Classification: The Use of A Priori Information in Cluster Analysis", *Psychometrika*, 49, 57-78
- DeSarbo, W. S., A. K. Manrai e L. A. Manrai (1984), "Latent Class Multidimensional Scaling: A Riview of Recent Development in the Marketing and Psychometrics Literature", *Advance Methods of Marketing Research*, Richard P. Bagozzi (ed.), Cambridge: Blacwell, 190-222.
- Dhalla, N. K. e W. H. Mahatoo (1976), "Expanding teh Scope of Segmentation Research", *Journal of Marketing*, 40, 34-41.
- Dickson, P. R. (1982), "Pearson-Situation: Segmentation's Missing Link", *Journal of Marketing*, 46, 56-64
- Dickson, P. R. e J. L. Ginter (1987), "Market Segmentation, Product Differentiation, and Marketing Strategy", *Journal of Marketing*, 51, 1-11
- Florence PV, Jolibert A. Social values, AIO, and consumption patterns. *Journal of Business Reseach* 1990;20:109-22

## Segmentação de indivíduos no *Facebook* que gostam de música

- Frank, R. E., W. F. Massy e Y. Wind (1972), “Market Segmentation”, Englewood Cliffs, NJ: Prentice Hall
- Gates, Michael, “VALS Changes With The Times”, *Incentive*. V. 163, n 3, pag. 27-31, 1989
- Gelsem ES, Eden G. Mapping algorithms in ISPAHAN. *Pattern Recognition* 1980;12:127-36
- Gensch, D. H. (1985), “Empirically Testing a Disaggregate Choice Model for Segments”, *Journal of Marketing Research*, 22, 462-467
- Goller, S., A. Hogg e S. Kalafatis, (2002), “A new research agenda for business segmenation”, *European Journal of Marketing* 36 (1/2), 252-271
- Gordon, A. D. (1999), “Classification”, Chapman and Hall
- Green, P. E. e F. J. Carmone (1977), “Segment Congruence Analysis: A Method for Analyzing Association Among Alternative Bases for Market Segmentation”, *Journal of Consumer Research*, 3, 217-222
- Green, P. E. e F. J. Carmone e D. P. Wachspress (1976), “Consumer Segmentation via Latent Class Analysis”; *Journal of Consumer Research*, 3, 170-174.
- Gunter, B. and A. Furnham (1992), “Consumer Profiles. A Introduction to Psychographics”, London: Routledge.
- Hagerty, M. R. (1985), “Improving the Predictive Power of Conjoint Analysis: The Use of Factor Analysis and Cluster Analysis”, *Journal of Marketing Research*, 22, 168-184
- Haley, R. I. (1968), “Benefit Segmentation: A Decision-Oriented Research Tool”, *Journal of Marketing*, 32, 30-35
- Hauser, J. R. e G. L. Urban (1977), “A Normative Methodology for Modeling Consumer Response to Innovation”, *Operations Research*, 25, 579-617

- Henshall, B., e R. Roberts, (1985), “Comparative Assessment of Tourist Generating Markets for New Zealand”, *Annals of Tourism Research* 12 (2), 219-238
- Higgins SH, Shanklin WL. Seeking mass market acceptance for high technology consumer products. *The Journal of Consumer Marketing* 1992;9(1): 5-14
- Holak SL, Lemann DR. Purchase intentions and the dimensions of innovation: an exploratory model. *Journal of Product Innovation Management* 1990; 7:59-73.
- Hruschka, H. (1986), Marketing Definition and Segmentation Using Fuzzy Clustering Methods”, *International Journal of Research Marketing*, 3, 117-134
- Hsieh, S.,J. O’Leary, e A. Morrison (1992), “Segmenting the international travel market by activity”, *Tourism Management* 13(2), 209-223
- Kahle, L. R. (1983), “Social Values and Social Change: Adaptation to life in America”, New York: Praeger.
- Kahle, L. R., S. E. Beatty e P. Holmer (1986), “Alternative Measurement Approaches to Consumer Values: The List of Values (LOV) and Values and Life Style (VAL)”, *Journal Of Consumer Research*, 13, 405-409.
- Kamakura, W. A. e G. J. Russel (1986), “A Probabilistic Choice Model for Market Segmentation and Elasticity Structure”, *Journal of Marketing Reserarch*, 26, 379-390
- Kamineni, R. (2005), “The next stage of psychographic segmentation: usage of enneagram”, *Journal of American Academy of Business*, 6(1), 315-320.
- Kara, A., e E. Kaynak (1997), “Markets of a single customer: exploiting conceptual development in market segmentation”, *European Journal of Marketing* 31(11/12), 873-895
- Kass, G. (1980),” An Exploratory Technique for Investigating Large Quantities of Categorical Data”, *Applied Statistics*, 29, 119-127



Klemperer P. Market with consumer switching costs. *The Quarterly Journal of Economics* 1987; 102(2): 375-94

Kotler, P., (1988), “Marketing Mangement: analysis, palnning, implementation and control, 6 th ed, Prentice-Hall, Englewood Cliffs – NJ – USA

Loker, L. e R. Perdue (1992), “A Benefit-Based Segmentation of a Nonresident Summer Travel Market”, *Journal of Travel Research* 31(1), 30-35

Loudon, D. e A. J. Della Bitta (1984), “Consumer Behavior Concepts and Applications”, London: McGraw-Hill International Editions

Magdison, J. (1994), “The CAID Approach to Segmentation Modelling: Chi-squares Automatic Interaction Detection”, *Advance Methods of Marketing Research*, R.P.BAgozzi (ed), Cambridge, MA: Blackwell, 118-119

Malcoln M, Ian Dunbar (2004), “Market segmentation: How to do it, how to profit from it”, Elsevier

Mark Sauders, Adrian Thornhill (2009), “Research Methods for Business Students”, Prentice Hall

Mazanec, J. (1995), “Análise da Posição Estratégica de Cidades Turísticas: Um estudo para Lisboa”, *revista Portuguesa de Gestão II/III*, 83-94

McLachlan, G. J. e K. E. Basford (1988), “Mixture Models: Inference and Applications to Clustering”, New York: Marcel Dekker

Michman, R. D. (1991), “Lifestyle Market Segmentation”, Praeger

Milligan, G. W. e M. C. Cooper (1985), “An Examination of Procedures for Determining the Number of Clusters in a Dataset”, *Psychometrika*, 50, 159-179

Moriarty, M e M. Venkatesan (1978), “Concept Evaluation and Market segmentation”, *Journal of Marketing*, 42, 82-86

- Novak, T. P. e B. MAcEvoy (1990), “On Comparing Alternative Segmentation Schemes: The List of Values and Values and Life-Styles”, *Journal of Consumer Research*, 17, 105-109
- Perdue, R., (1996), “Target Market Selection and Marketing Strategy: The Colorado Downhill Skiing Industry”, *Journal of Travel Research* 34 (4), 39-46
- Philip Kotler, Kevin L. Keller (2009), “Marketing Management”, 13th Edition, Prentice Hall
- Plummer, J. T. (1974), “The Concept of Style Segmentation”, *Journal of Marketing*, 38, 33-37
- Punj, G. e D. W. Stewart (1983), “Cluster Analysis in Marketing Research: Review and Suggestions for Application”, *Journal of Marketing Research*, 20, 134-148
- Rao, V. R e F. W. Winter (1978), “An Application of the Mutivariate Probit Model to Market Segmentation and Product Design”, *Journal of Marketing Research*, 15, 361-368
- Ratneshwar S, Shocker AD. Substitution-in-use and the role of usage context in product category structures. *Journal of Marketing Research* 1991;28:281-95
- Robinson, J. (1938), “The Economics of the Imperfect Competition”, London: McMillan
- Rogers, M. (1962), “Diffusion of Innovation”, New York: The Free Press
- Rokeach, M. (1973), “The Nature of Human Values”; New York: The Free Press
- Schiffman, Leon G.; Kanuk, Leslie Lazar, “Comportamento do consumidor”, 6. Ed Rio de Janeiro
- Schwartz, S. H. (1992), “Universals in the Content and Structure of Values: Theoretical Advances and Empirical Tests in 20 Countries”, *Advances in Experimental Social Psychology*, Mark P. Zanna (ed), New York: Academic Press

## Segmentação de indivíduos no *Facebook* que gostam de música

- Schwartz, S. H. e W. Bilsky (1987), “Towards Universal Psychological Structure of Human Values”, *Journal of Personality and Social Psychology*, 53, 550-562
- Schwartz, S. H. e W. Bilsky (1990), “Towards a Theory of the Universal Content and Structure of Values: Extensions and Cross-cultural Replications”, *Journal of Personality and Social Psychology*, 58, 878-891
- Shepard, R. N. e Arabie (1979), “Additive Clustering: Representation of Similarities as Combinations of Discrete Overlapping Properties”, *Psychological Review*, 86, 87-123
- Sheth JN, Newman BI, Gross BL. Consumption values and market choices. Cincinnati: South Western Publishing Co., 1991.
- Shufeldt, L., Oates, B., & Vaught, B. (1998), “Is lifestyle an important factor in the purchase of OTC drugs by the elderly?”, *The Journal of Customer Marketing*, 15(2), 111
- Smith. W. (1956), “Product Differentiation and Market Segmentation as Alternative Marketing Strategies”, *Journal of Marketing*, 21, 3-8
- Spath, H. (1979), “Clusterwise Linear Regression”, *Computing*, 22, 367-373
- Spath, H. (1981), “Clusterwise Linear Regression”, *Computing*, 26, 275
- Spath, H. (1982), “A fast Algorithm for Clusterwise Linear Regression”, *Computing*, 29, 175-181
- Steen kamp, J. E. B. M. e M. Wedel (1993), “Fuzzy Clusterwise Regression in Benefit Segmentation: Application and Investigation into its Validity”, *Journal of Business Research*, 26, 237-249
- Sugan SM. The cost of thinking. *Journal of Consumer Research* 1980; 7:99-111
- Teo TSH, Lim VKG, Lai RYC. Intrinsic and extrinsic motivation in internet usage. *Omega* 1999; 27(1):25-37

- Tewdt, D. W. (1967), “How Does Awareness-Attitude affect Marketing Strategy?”, *Journal of Marketing*, 31, 64-66
- Van Raaij e T. M. M. Verhallen (1994), “Domain-Specific market segmentation”, *European Journal of Marketing*, 28, 49-66
- Von Hippel E. Lead users: a source of novel product concepts. *Management Science* 1986:32:791-805
- Wedel, M. e J. B. E. M. Steenkamp (1989), “Fuzzy Cluster wise Regression Approach to Benefit Segmentation”, *International Journal of Research in Marketing*, 6, 241-258
- Wedel, M. e J. B. E. M. Steenkamp (1991), “A Clusterwise Regression Method for Simultaneous Fuzzy Market Structuring and Benefit Segmentation!”, *Journal of Marketing Research*, 28, 385-396
- Wedel, M. e W. S. DeSarbo (1994), “A Review of Latent Class Regression Models and their Applications”, *Advanced Methods for Marketing Research*, Richard P. Bagozzi (ed.), 353-388
- Weinstein, A. (1994), “Market Segmentation”, IRWIN
- Wells, W. D. (1975), “Psychographics: A Critical Review”, *Journal of Marketing Research*, 12, 196-213
- Wilkie, W. L. e J. B. Choen (1977), “An Overview of Market Segmentation: Behavioral Concepts and Research Approaches”, *Marketing Science Institute Working Paper*
- Wind, Y. (1978), “Issues and Advances in Segmentation Research”, *Journal of Marketing Research*, 15, 317-337
- Wind, Y., e R. Cardozo (1974), “Industrial Market Segmentation”, *Industrial Marketing Management* 3, 155-165
- Woodside, A., e L. Jacobs, (1985), “Step Two in Benefit Segmentation: Learning the Benefits Realized by Major Travel Markets”, *Journal of Travel Research* 24 (1), 7-13

Segmentação de indivíduos no *Facebook* que gostam de música

Zadeh, L. A. (1965), "Fuzzy Sets", *Information and Control*, 8, 338-353

Ziff, R. (1971), "Psychographics for Market Segmentation", *Journal of Advertising Research*, 11, 3-9

## 8. Anexos

### Anexo 1: Procedimento para obtenção do público-alvo no Facebook.

2. Definição do público-alvo FAQ sobre definição do público-alvo de anúncios

---

**Local**

País:  ?

Em todo o lado ?  
 Por cidade ?

---

**Dados demográficos**

Idade:  -  ?

Exigir correspondência de idade exacta ?

Sexo:  Todos  Homens  Mulheres

---

**Interesses**

Interesses Precisos:  ? +

Gostos & Interesses sugeridos

<input type="checkbox"/> Dormir Escuchando Música	<input type="checkbox"/> Escuchar Musica Mientras Chateo
<input type="checkbox"/> #Tree	<input type="checkbox"/> #La Cama
<input type="checkbox"/> #Traer, Iowa	<input type="checkbox"/> Ares

[Mudar para público-alvo Categoria Alargada](#) ?

---

**Ligações no Facebook**

Ligações:  Qualquer pessoa ?  
 Público-alvo de ligações avançadas

---

**Dados demográficos avançados**

Interessado/a em:  Todos  Homens  Mulheres

Relação:  Todas  Solteiro(s)  Noivo/a ?  
 Num(a) relação  Casado(s)

Idiomas:

---

**Formação & Trabalho**

Alcance estimado ?

**51.760** pessoas

- que vivem no país: Portugal
- que gostam de musica

**Anexo 2: Questionário**

<b>1. Interesses em tipos de música</b>					
Por favor indique qual é o seu interesse nos seguintes tipos de música					
	Pouco interesse			Muito interesse	
Música Portuguesa (P101)	1	2	3	4	5
Música Brasileira (P102)	1	2	3	4	5
Pop Rock Internacional (P103)	1	2	3	4	5
Alternativa, Indie (P104)	1	2	3	4	5
Hard, Havy, Punk e Metal (P105)	1	2	3	4	5
Dança, House e Electro (P106)	1	2	3	4	5
Soul, Funk, R&B e Hip-Hop (P107)	1	2	3	4	5
Clássica (P108)	1	2	3	4	5
Jazz (P109)	1	2	3	4	5
Outro tipo de música (P110)	1	2	3	4	5

<b>2. Frequência de audição</b>					
Por favor indique qual é a frequência que ouve os seguintes tipos de música					
	Todos os dias	2 - 3 x semana	1 x semana	1 x mês	Raramente
Música Portuguesa (P201)	1	2	3	4	5
Música Brasileira (P202)	1	2	3	4	5
Pop Rock Internacional (P203)	1	2	3	4	5
Alternativa, Indie (P204)	1	2	3	4	5
Hard, Havy, Punk e Metal (P205)	1	2	3	4	5
Dança, House e Electro (P206)	1	2	3	4	5
Soul, Funk, R&B e Hip-Hop (P207)	1	2	3	4	5
Clássica (P208)	1	2	3	4	5
Jazz (P209)	1	2	3	4	5
Outro tipo de música (P210)	1	2	3	4	5

<b>3. Atitudes e opiniões perante a música</b>					
Das seguintes afirmações, quais são as que mais concorda e as que mais discorda					
	Discordo muito			Concordo muito	
O acesso fácil através da internet á música é muito importante para mim (P301)	1	2	3	4	5
O acesso gratuito a sítios da internet de música é muito importante para mim (P302)	1	2	3	4	5

## Segmentação de indivíduos no *Facebook* que gostam de música

O acesso rápido através da internet á música é muito importante para mim (P303)	1	2	3	4	5
A música deve ter um preço baixo de venda (P304)	1	2	3	4	5
Os direitos de autor deveriam ser preservados (P305)	1	2	3	4	5
As músicas deveriam ter o valor do IVA reduzido (P306)	1	2	3	4	5
Os sítios de partilha de música na internet deveriam ser permitidos por lei (P307)	1	2	3	4	5
Eu tenho receio em comprar música via internet (P308)	1	2	3	4	5

### 4. Frequência na aquisição da música

Quais são os locais e a frequência na aquisição de música	Todos os dias	2 - 3 x semana	1 x semana	1 x mês	Raramente
Lojas especializadas em música (P401)	1	2	3	4	5
Supermercados (P402)	1	2	3	4	5
Lojas generalistas (P403)	1	2	3	4	5
Através de catálogos (P404)	1	2	3	4	5
Ofertas de jornais ou revistas (P405)	1	2	3	4	5
Lojas online de música: amazon, emusica, etc. (P406)	1	2	3	4	5
Sítios específicos de fabricantes: itunes ou outros (P407)	1	2	3	4	5
Sítios específicos de autores (P408)	1	2	3	4	5
Troca com amigos em formato digital (P409)	1	2	3	4	5
Sítios de partilha, tipo emule, torrents ou outros (P410)	1	2	3	4	5

### 5. Atitudes e opiniões sobre as plataformas de audição de música

Das seguintes afirmações, quais são as que mais concorda e as que mais discorda	Discordo muito				Concordo muito
Eu tenho dificuldade na utilização dos aparelhos portáteis: ipod, mp3 (P501)	1	2	3	4	5
Eu acho muito complicado o carregamento dos ficheiros para aparelhos portáteis (P502)	1	2	3	4	5
Eu gosto mais de ouvir música no carro do que nos aparelhos pessoais (P503)	1	2	3	4	5
Quando ouço música, não partilho com mais ninguém (P504)	1	2	3	4	5



## Segmentação de indivíduos no *Facebook* que gostam de música

Gosto de partilhar nas redes sociais os meus gostos sobre as músicas que oiço (P505)	1	2	3	4	5
Quando gosto da música, clico no item "eu gosto" no Facebook (P506)	1	2	3	4	5
Costumo procurar informação sobre os autores e músicas que oiço na imprensa (P507)	1	2	3	4	5
Costumo procurar informação sobre os autores e músicas que oiço na internet (P508)	1	2	3	4	5

### 6. Actividades relacionadas com a audição de música

Quais são os modos e a frequência na audição de música

	Todos os dias	2 - 3 x semana	1 x semana	1 x mês	Raramente
Discos em vinil (P601)	1	2	3	4	5
Rádio (P602)	1	2	3	4	5
Televisão (P603)	1	2	3	4	5
Concertos de música ao vivo (P604)	1	2	3	4	5
CD (P605)	1	2	3	4	5
Leitor portátil de mp3 (P606)	1	2	3	4	5
Computador (P607)	1	2	3	4	5
Predominantemente durante o dia (P608)	1	2	3	4	5
Predominantemente durante á noite (P609)	1	2	3	4	5

### 7. Dados demográficos - Idade

12 - 20  
21 - 30  
31 - 40  
41 - 50  
51 - 60  
> 60

### 8. Dados demográficos - Sexo

Masculino  
Feminino

### 9. Dados demográficos - Estado civil

Solteiro  
Casado  
Divorciado  
Outro

### 10. Dados demográficos - Escolaridade

9º ano  
12º Ano  
Licenciatura  
Mestrado  
Doutoramento  
Outro

**Anexo3: Frequências idade, Sexo, Estado civil e Escolaridade**

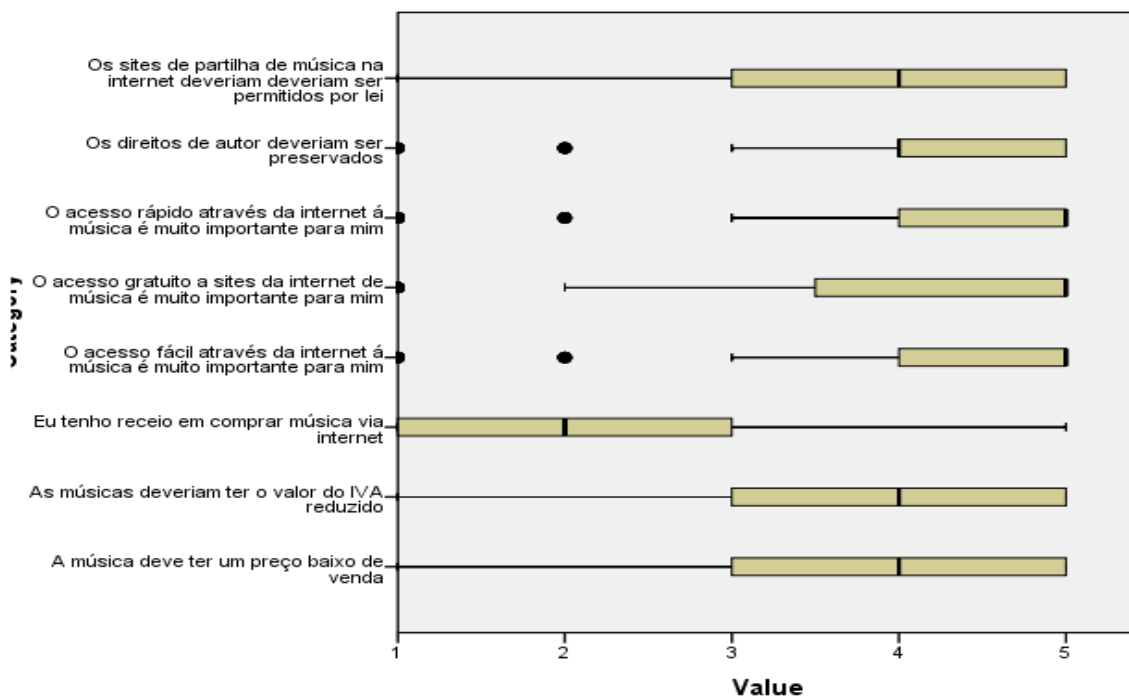
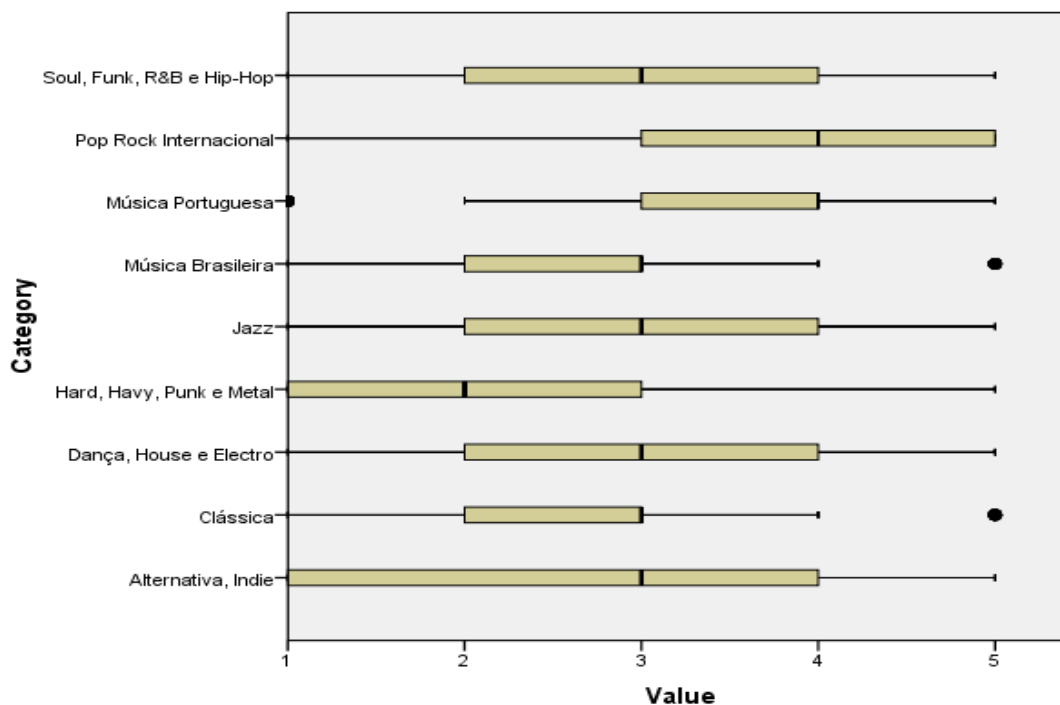
<b>Idade</b>				
	Frequency	Percent	Valid Percent	Cumulative Percent
21 a 30 anos	122	48,6	48,6	48,6
31 a 40 anos	102	40,6	40,6	89,2
41 a 50 anos	25	10,0	10,0	99,2
> 60 anos	2	,8	,8	100,0
Total	251	100,0	100,0	

<b>Sexo</b>				
	Frequency	Percent	Valid Percent	Cumulative Percent
Masculino	132	52,6	52,6	52,6
Feminino	119	47,4	47,4	100,0
Total	251	100,0	100,0	

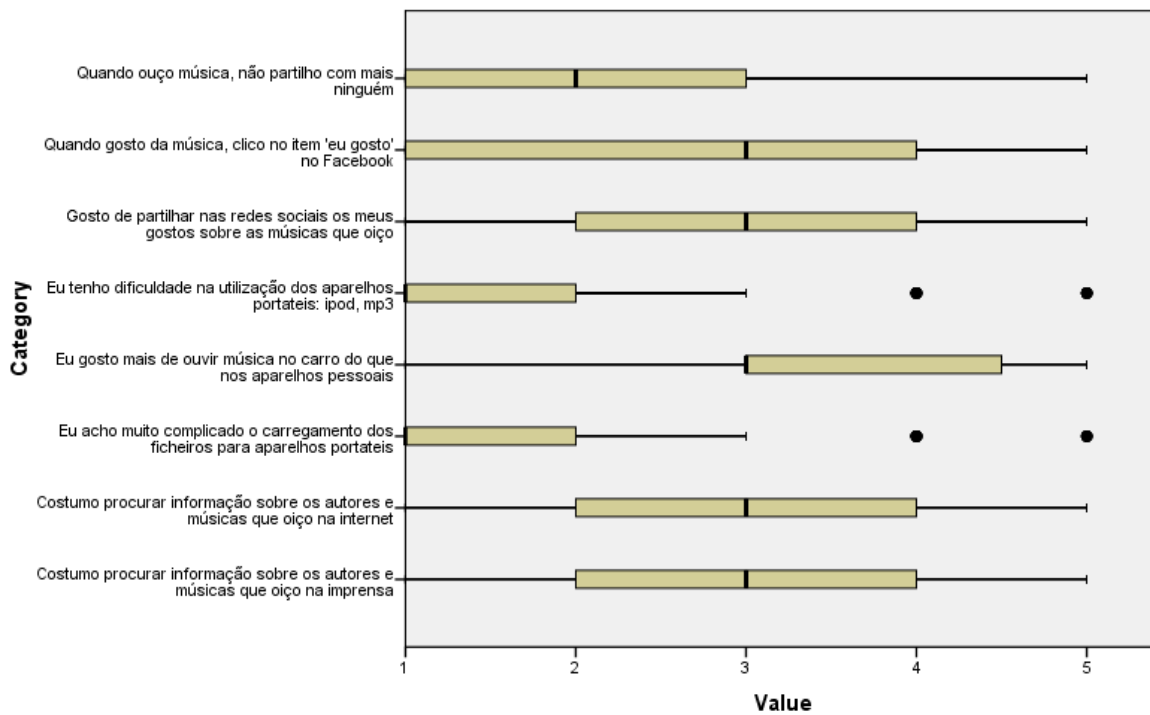
<b>Estado Civil</b>				
	Frequency	Percent	Valid Percent	Cumulative Percent
Solteiro	143	57,0	57,0	57,0
Casado	83	33,1	33,1	90,0
Divorciado	15	6,0	6,0	96,0
Outro	10	4,0	4,0	100,0
Total	251	100,0	100,0	

<b>Escolaridade</b>				
	Frequency	Percent	Valid Percent	Cumulative Percent
12° Ano	26	10,4	10,4	10,4
Licenciatura	140	55,8	55,8	66,1
Mestrado	70	27,9	27,9	94,0
Doutoramento	5	2,0	2,0	96,0
Outro	10	4,0	4,0	100,0
Total	251	100,0	100,0	

**Anexo 4: Outliers – “caixa de bigodes”**



## Segmentação de indivíduos no *Facebook* que gostam de música



**Anexo 5: Rotated Component Matrix**

Interesses em tipos de música

**Rotated Component Matrix<sup>a</sup>**

	Component			
	1	2	3	4
Jazz	<b>,885</b>	,080	-,016	,035
Clássica	<b>,790</b>	,057	,068	-,068
Hard, Havy, Punk e Metal	-,013	<b>,809</b>	-,052	,019
Alternativa, Indie	,198	<b>,761</b>	,075	,057
Música Portuguesa	,087	-,016	<b>,770</b>	-,118
Pop Rock Internacional	-,192	,220	<b>,650</b>	,236
Música Brasileira	,231	-,434	<b>,612</b>	-,018
Soul, Funk, R&B e Hip-Hop	,182	-,058	,148	<b>,840</b>
Dança, House e Electro	-,231	,140	-,130	<b>,750</b>

Opiniões sobre música

**Rotated Component Matrix<sup>a</sup>**

	Component			
	1	2	3	4
O acesso rápido através da internet á música é muito importante para mim	<b>,949</b>	,108	,017	-,019
O acesso fácil através da internet á música é muito importante para mim	<b>,912</b>	,180	-,070	-,015
O acesso gratuito a sítios da internet de música é muito importante para mim	<b>,881</b>	,223	-,040	-,093
A música deve ter um preço baixo de venda	,170	<b>,807</b>	-,187	,106
Os sítios de partilha de música na internet deveriam ser permitidos por lei	,264	<b>,662</b>	,206	-,393
Eu tenho receio em comprar música via internet	-	-,144	<b>,865</b>	-,058
As músicas deveriam ter o valor do IVA reduzido	,147	,202	,457	<b>,569</b>
Os direitos de autor deveriam ser preservados	-,064	-,008	,042	<b>,928</b>

Segmentação de indivíduos no *Facebook* que gostam de música

Opiniões sobre as plataformas de audição de música

**Rotated Component Matrix<sup>a</sup>**

	Component			
	1	2	3	4
Costumo procurar informação sobre os autores e músicas que oiço na internet	<b>,857</b>	-,028	,301	,001
Costumo procurar informação sobre os autores e músicas que oiço na imprensa	<b>,824</b>	,040	,369	,083
Eu gosto mais de ouvir música no carro do que nos aparelhos pessoais	<b>,555</b>	,389	,294	,296
Eu tenho dificuldade na utilização dos aparelhos portáteis: iPod, mp3	-	<b>,916</b>	-,058	-,019
Eu acho muito complicado o carregamento dos ficheiros para aparelhos portáteis	-	<b>,870</b>	-,153	,129
Gosto de partilhar nas redes sociais os meus gostos sobre as músicas que oiço	,187	-,097	<b>,836</b>	-,147
Quando gosto da música, clico no item "eu gosto" no <i>Facebook</i>	,242	-,110	<b>,836</b>	-,062
Quando ouço música, não partilho com mais ninguém	,010	,076	-,171	<b>,948</b>

**Anexo 6: Comunalidades**

Interesses em tipos de música

<b>Communalities</b>		
	Initial	Extraction
Música Portuguesa	1,000	,615
Música Brasileira	1,000	,616
Pop Rock Internacional	1,000	,563
Alternativa, Indie	1,000	,627
Hard, Havy, Punk e Metal	1,000	,658
Dança, House e Electro	1,000	,652
Soul, Funk, R&B e Hip-Hop	1,000	,765
Clássica	1,000	,636
Jazz	1,000	,790

Opiniões sobre música

<b>Communalities</b>		
	Initial	Extraction
O acesso fácil através da internet á música é muito importante para mim	1,000	,869
O acesso gratuito a sítios da internet de música é muito importante para mim	1,000	,837
O acesso rápido através da internet á música é muito importante para mim	1,000	,912
A música deve ter um preço baixo de venda	1,000	,726
Os direitos de autor deveriam ser preservados	1,000	,866
As músicas deveriam ter o valor do IVA reduzido	1,000	,677
Os sítios de partilha de música na internet deveriam ser permitidos por lei	1,000	,705
Eu tenho receio em comprar música via internet	1,000	,793

## Segmentação de indivíduos no *Facebook* que gostam de música

### Opiniões sobre as plataformas de audição de música

<b>Communalities</b>		
	<b>Initial</b>	<b>Extraction</b>
Eu tenho dificuldade na utilização dos aparelhos portateis: ipod, mp3	1,000	,844
Eu acho muito complicado o carregamento dos ficheiros para aparelhos portateis	1,000	,798
Eu gosto mais de ouvir música no carro do que nos aparelhos pessoais	1,000	,634
Quando ouço música, não partilho com mais ninguém	1,000	,934
Gosto de partilhar nas redes sociais os meus gostos sobre as músicas que oiço	1,000	,766
Quando gosto da música, clico no item "eu gosto" no Facebook	1,000	,773
Costumo procurar informação sobre os autores e músicas que oiço na imprensa	1,000	,823
Costumo procurar informação sobre os autores e músicas que oiço na internet	1,000	,825



**Anexo 7: Valores próprios e variância explicada**

Interesses em tipos de música

<b>Total Variance Explained</b>									
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	1,764	19,595	19,595	1,764	19,595	19,595	1,630	18,114	18,114
2	1,631	18,120	37,714	1,631	18,120	37,714	1,504	16,708	34,823
3	1,444	16,047	53,761	1,444	16,047	53,761	1,442	16,020	50,842
4	1,085	12,060	65,821	1,085	12,060	65,821	1,348	14,979	65,821
5	,813	9,032	74,853						
6	,736	8,183	83,036						
7	,598	6,649	89,684						
8	,555	6,166	95,850						
9	,374	4,150	100,000						

Opiniões sobre música

<b>Total Variance Explained</b>									
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3,151	39,387	39,387	3,151	39,387	39,387	2,674	33,420	33,420
2	1,265	15,818	55,206	1,265	15,818	55,206	1,412	17,655	51,075
3	1,076	13,447	68,652	1,076	13,447	68,652	1,157	14,463	65,538
4	,894	11,173	79,826	,894	11,173	79,826	1,143	14,288	79,826
5	,695	8,691	88,517						
6	,547	6,841	95,357						
7	,238	2,975	98,332						
8	,133	1,668	100,000						

Segmentação de indivíduos no *Facebook* que gostam de música

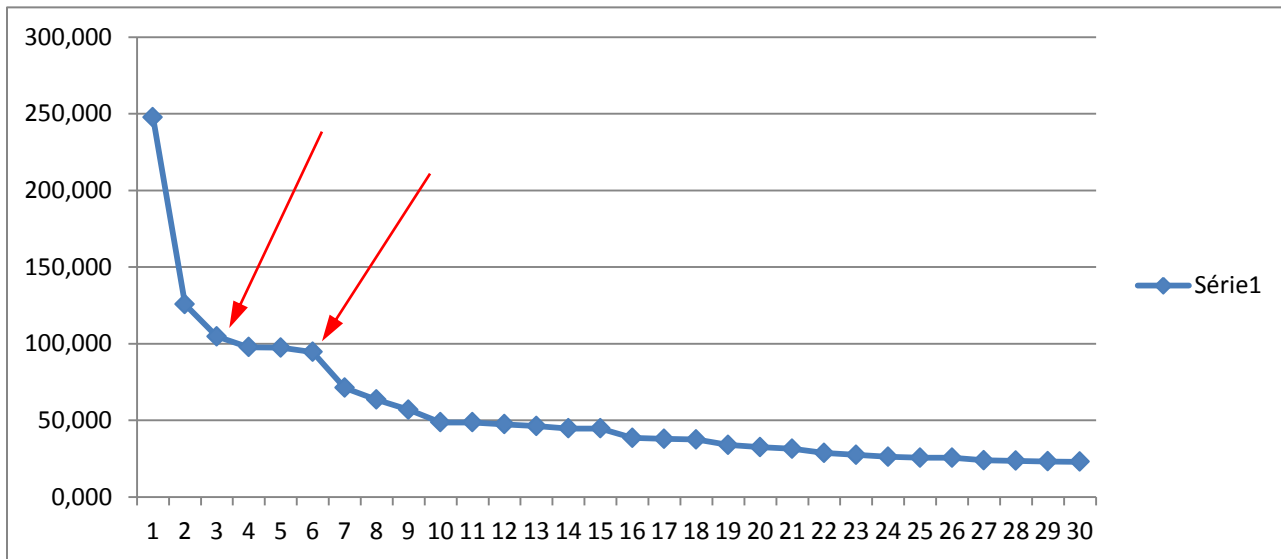
Opiniões sobre as plataformas de audição de música

<b>Total Variance Explained</b>									
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2,799	34,982	34,982	2,799	34,982	34,982	1,816	22,696	22,696
2	1,652	20,644	55,625	1,652	20,644	55,625	1,777	22,210	44,906
3	1,076	13,447	69,073	1,076	13,447	69,073	1,768	22,094	67,000
4	,870	10,879	79,952	,870	10,879	79,952	1,036	12,952	79,952
5	,747	9,335	89,288						
6	,330	4,127	93,415						
7	,310	3,872	97,287						
8	,217	2,713	100,000						

**Anexo 8: Quadro resumo após PCA**

Interesses em tipos de música													
Factor Loading									Estatísticas				
Factor	P101	P102	P103	P104	P105	P106	P107	P108	P109	Média	Desvio Padrão	Nº Itens	Alpha Cronbach
F11								0,790	0,885	2,737	1,192	2	0,664
F12				0,761	0,809					2,392	1,345	2	0,620
F13	0,770	0,612	0,650							3,308	1,236	3	0,723
F14						0,750	0,840			2,954	1,235	2	0,648
Determinante		KMO	Significância	Média Comunalidades									
0,335		0,501	0,000	0,658									
Opiniões sobre música													
Factor Loading									Estatísticas				
Factor	P301	P302	P303	P304	P305	P306	P307	P308	Média	Desvio Padrão	Nº Itens	Alpha Cronbach	
F31	0,912	0,881	0,949						4,155	1,142	3	0,924	
F32				0,807			0,662		3,912	1,167	2	0,650	
F33						0,569		0,865	3,088	1,546	2	0,773	
F34					0,928				4,127	1,031	1	N.A.	
Determinante		KMO	Significância	Média Comunalidades									
0,046		0,751	0,000	0,798									
Opiniões sobre plataformas de audição de música													
Factor Loading									Estatísticas				
Factor	P501	P502	P503	P504	P505	P506	P507	P508	Média	Desvio Padrão	Nº Itens	Alpha Cronbach	
F51			0,555				0,824	0,857	3,214	1,304	3	0,622	
F52	0,916	0,870							1,524	0,970	2	0,804	
F53					0,836	0,836			2,771	1,417	2	0,799	
F54				0,948					2,084	1,175	1	N.A.	
Determinante		KMO	Significância	Média Comunalidades									
0,072		0,646	0,000	0,800									

**Anexo 9: Determinação do número de clusters – *Kmeans***



**Anexo 10: Análise discriminante**

Três *clusters*

Classification Results <sup>a</sup>						
	Cluster Number of Case	Predicted Group Membership			Total	
		Cluster 1	Cluster 2	Cluster 3		
Original	Count	Cluster 1	49	2	2	53
		Cluster 2	0	80	3	83
		Cluster 3	0	1	114	115
	%	Cluster 1	92,5	3,8	3,8	100,0
		Cluster 2	,0	96,4	3,6	100,0
		Cluster 3	,0	,9	99,1	100,0

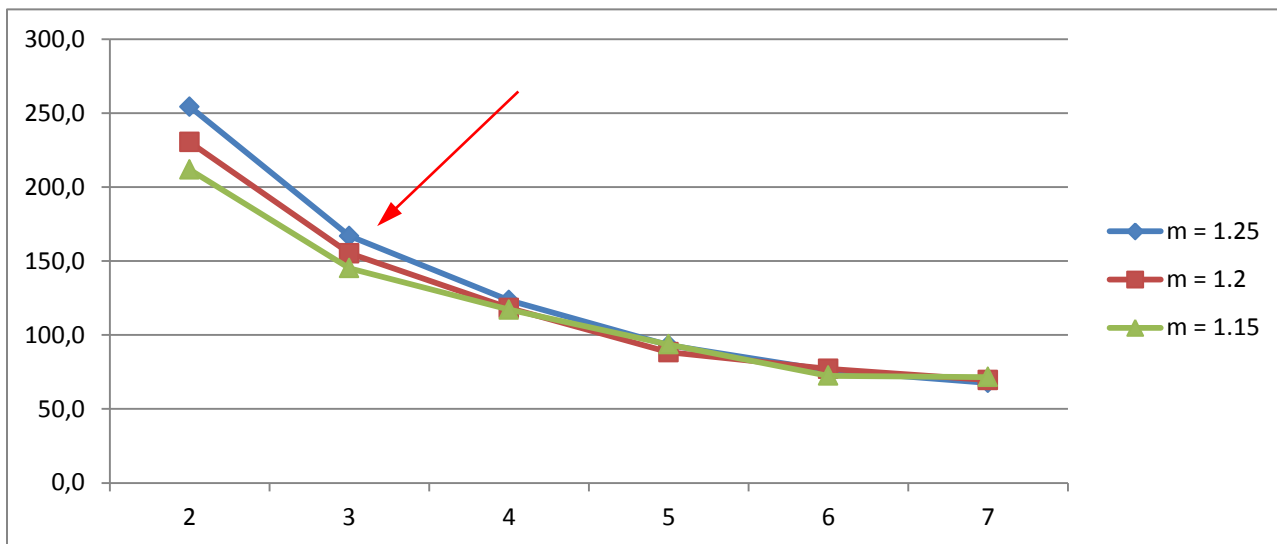
a. 96,8% of original grouped cases correctly classified.

Seis *clusters*

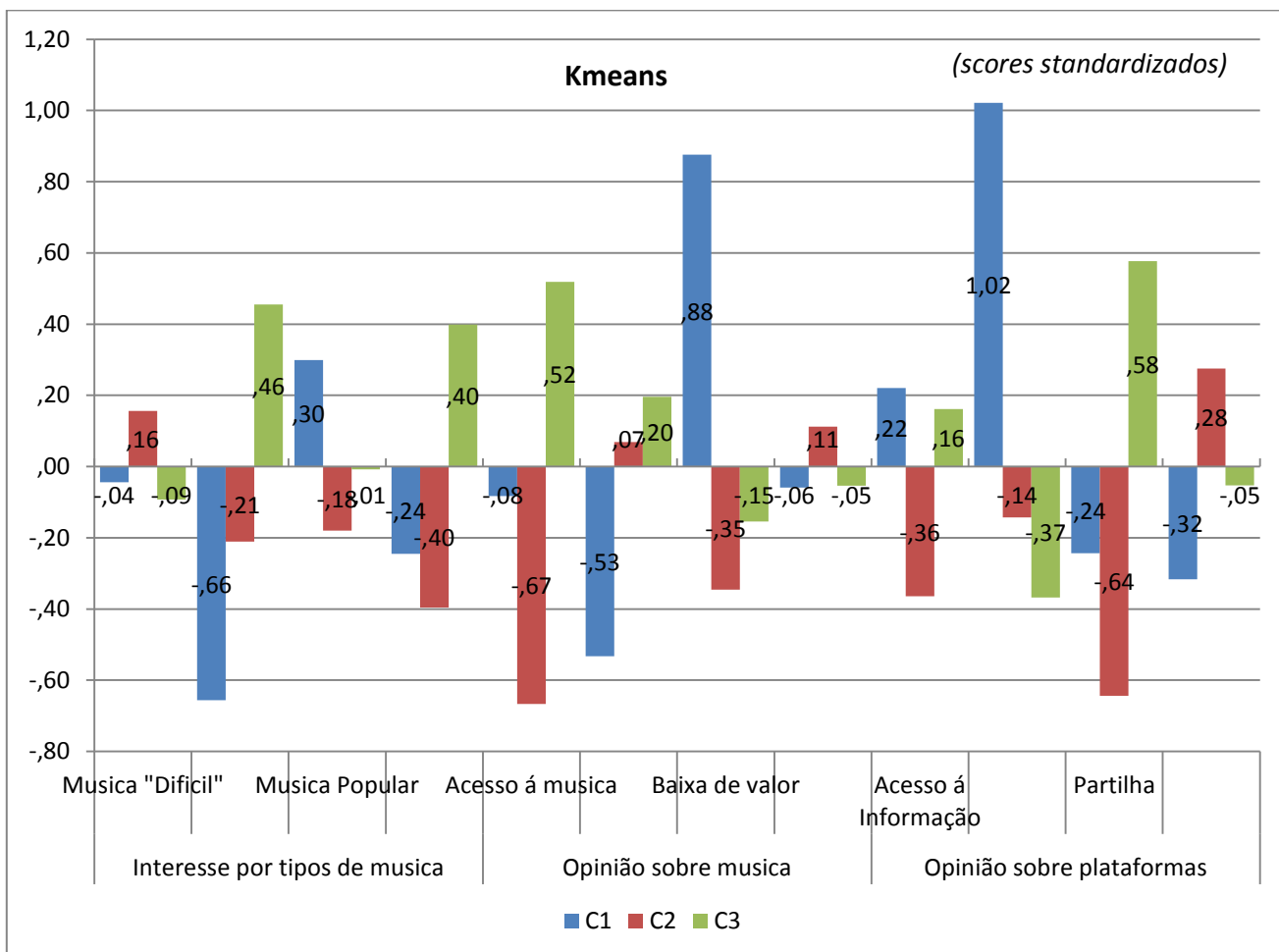
Classification Results <sup>a</sup>									
	Cluster Number of Case	Predicted Group Membership						Total	
		Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6		
Original	Count	Cluster 1	44	0	2	0	1	1	48
		Cluster 2	1	23	0	2	0	0	26
		Cluster 3	0	1	22	0	0	0	23
		Cluster 4	1	0	0	36	0	2	39
		Cluster 5	1	0	0	0	54	0	55
		Cluster 6	0	2	0	0	0	58	60
	%	Cluster 1	91,7	,0	4,2	,0	2,1	2,1	100,0
		Cluster 2	3,8	88,5	,0	7,7	,0	,0	100,0
		Cluster 3	,0	4,3	95,7	,0	,0	,0	100,0
		Cluster 4	2,6	,0	,0	92,3	,0	5,1	100,0
		Cluster 5	1,8	,0	,0	,0	98,2	,0	100,0
		Cluster 6	,0	3,3	,0	,0	,0	96,7	100,0

a. 94,4% of original grouped cases correctly classified.

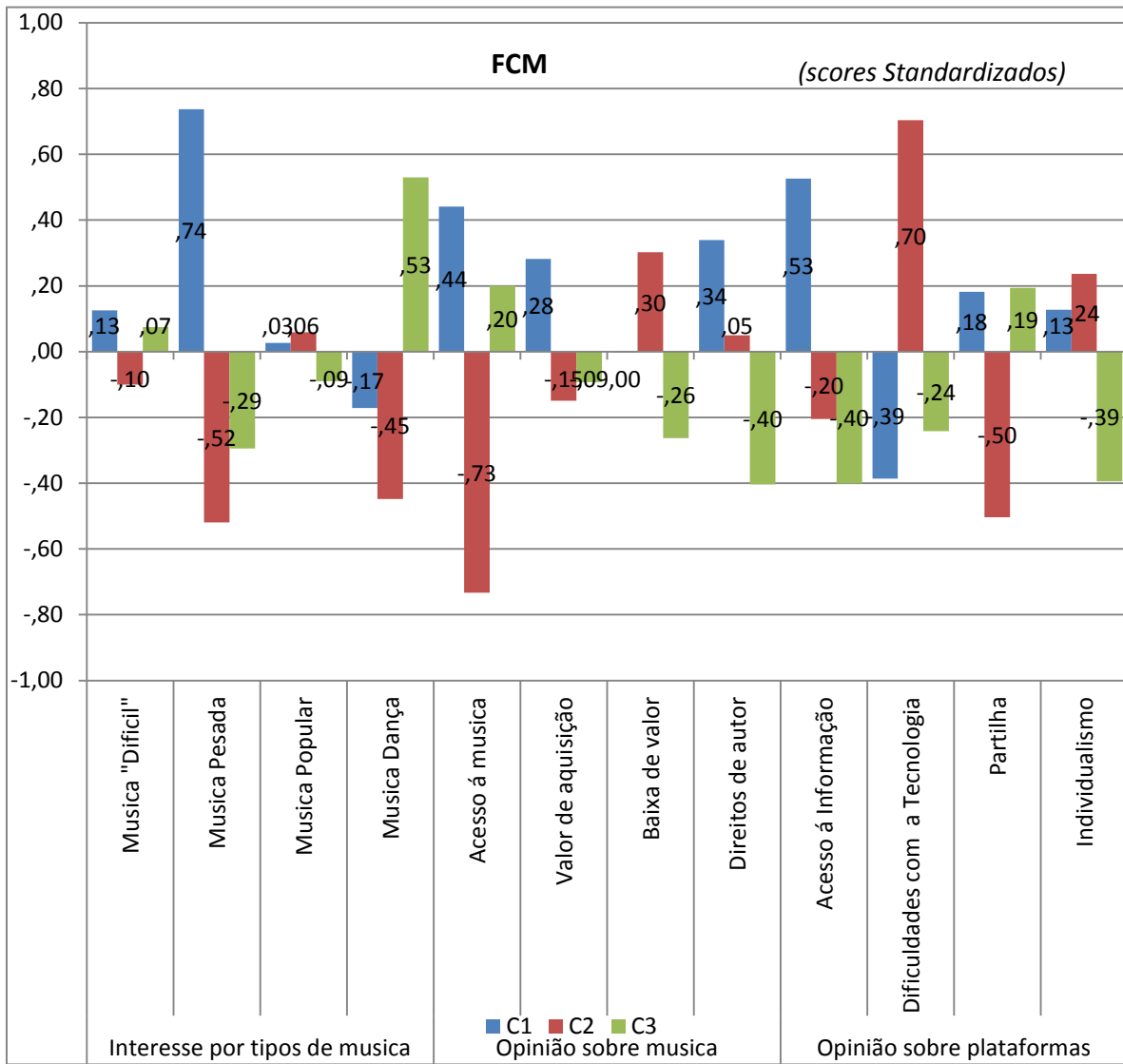
*Anexo 11: Determinação do Número de Clusters em Fuzzy*



**Anexo 12: Gráfico final de Kmeans**



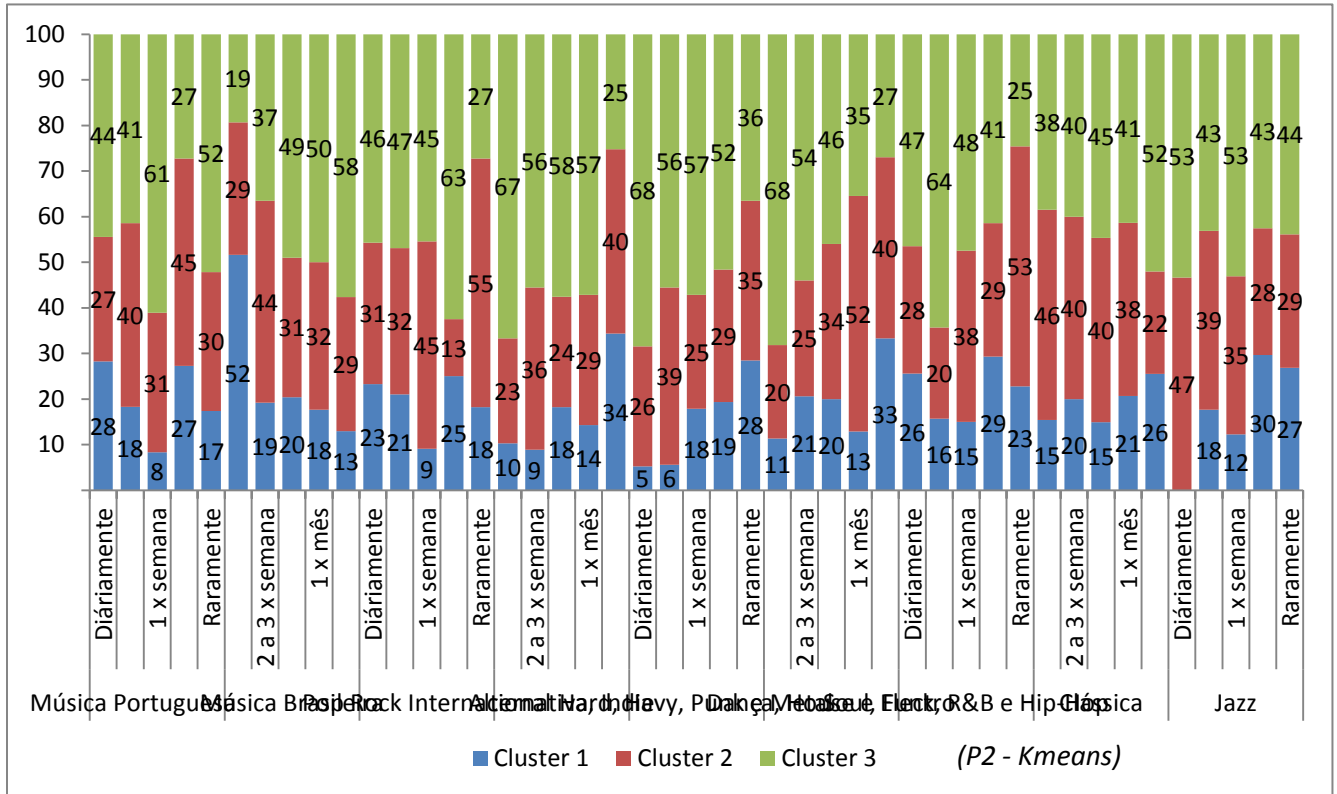
Anexo 13: Gráfico final Fuzzy





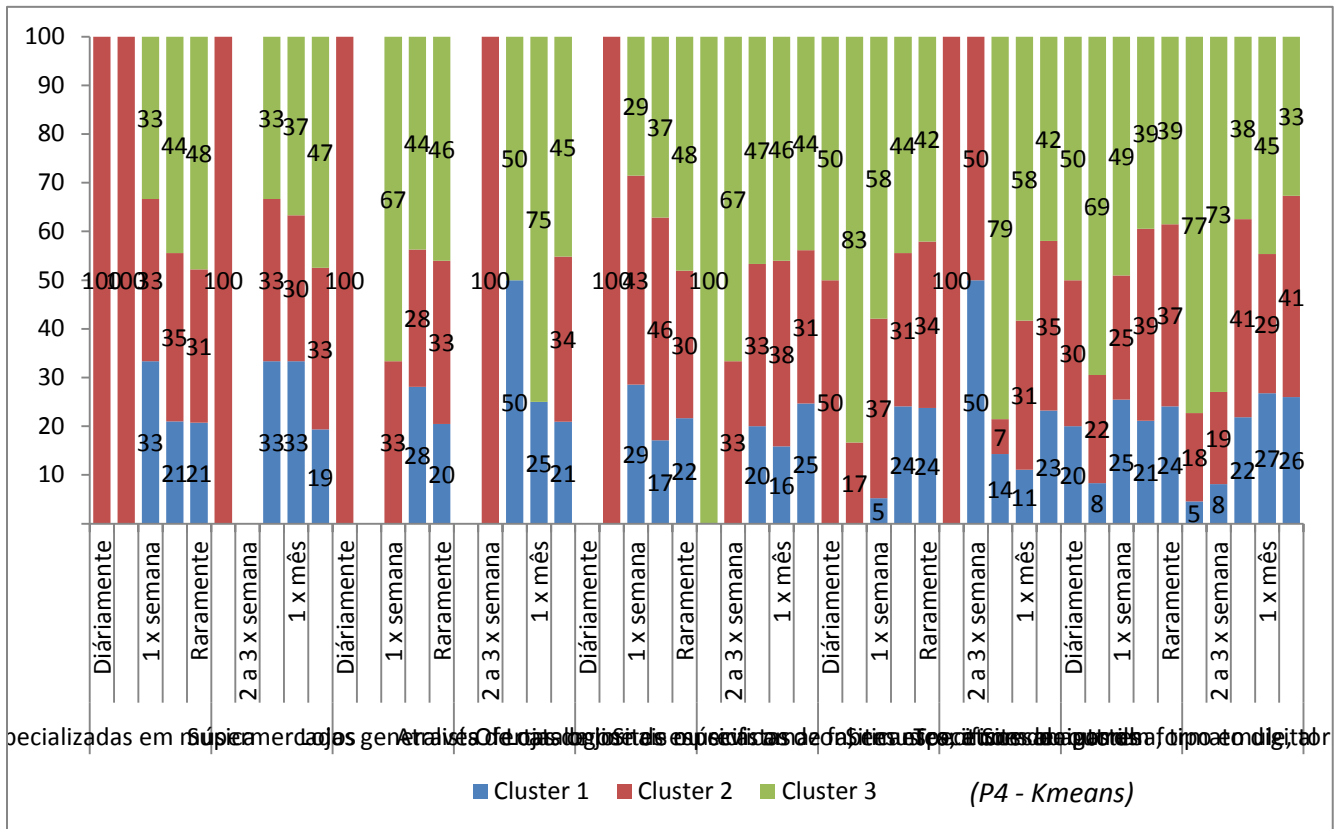
**Anexo 14: Frequência de modos, aquisição e audição de música com os clusters**

Frequência na audição de tipos de música (Kmeans)



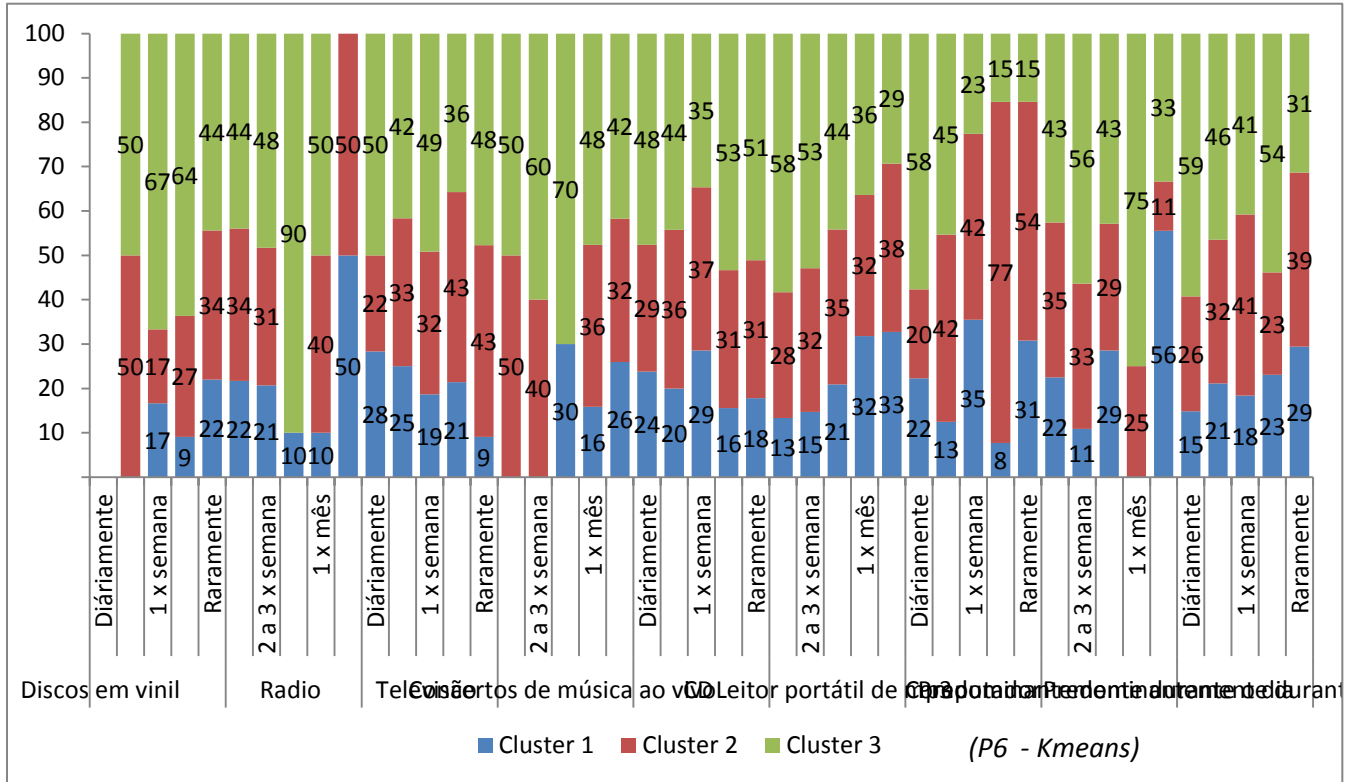
# Segmentação de indivíduos no Facebook que gostam de música

## Frequência na aquisição (Kmeans)



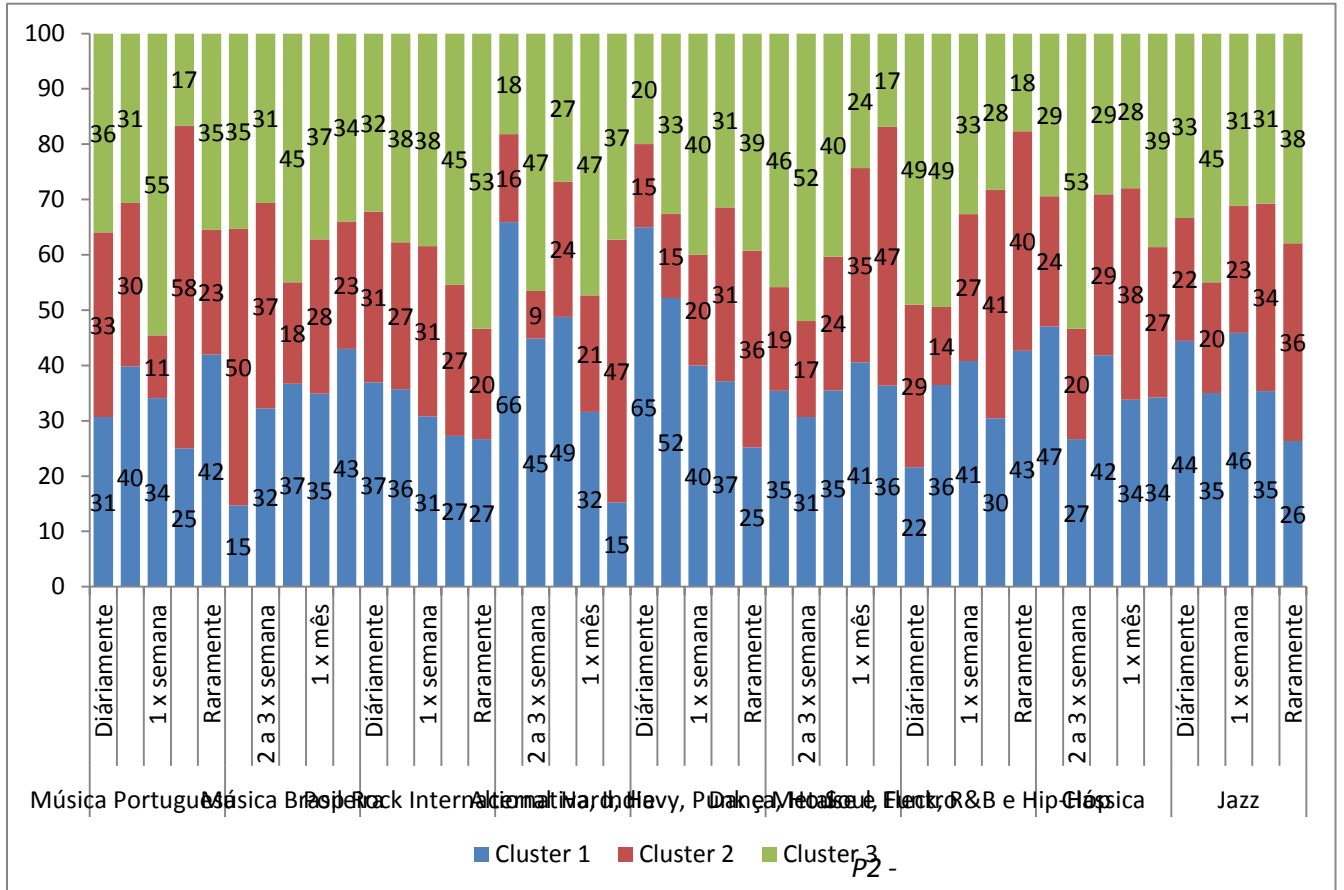
# Segmentação de indivíduos no Facebook que gostam de música

## Frequência nos modos de audição (Kmeans)



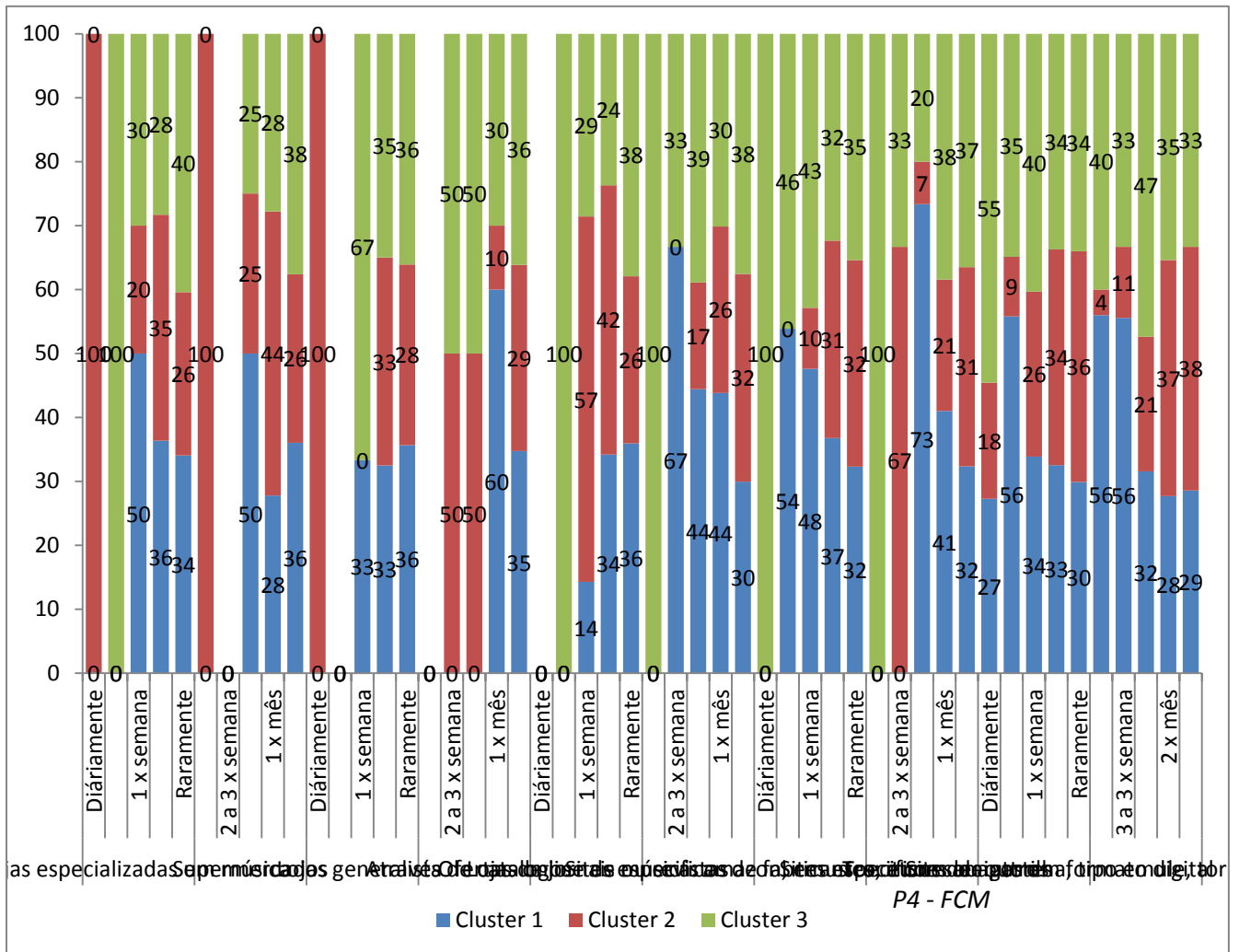
# Segmentação de indivíduos no Facebook que gostam de música

## Frequência na audição de tipos de música (FCM)



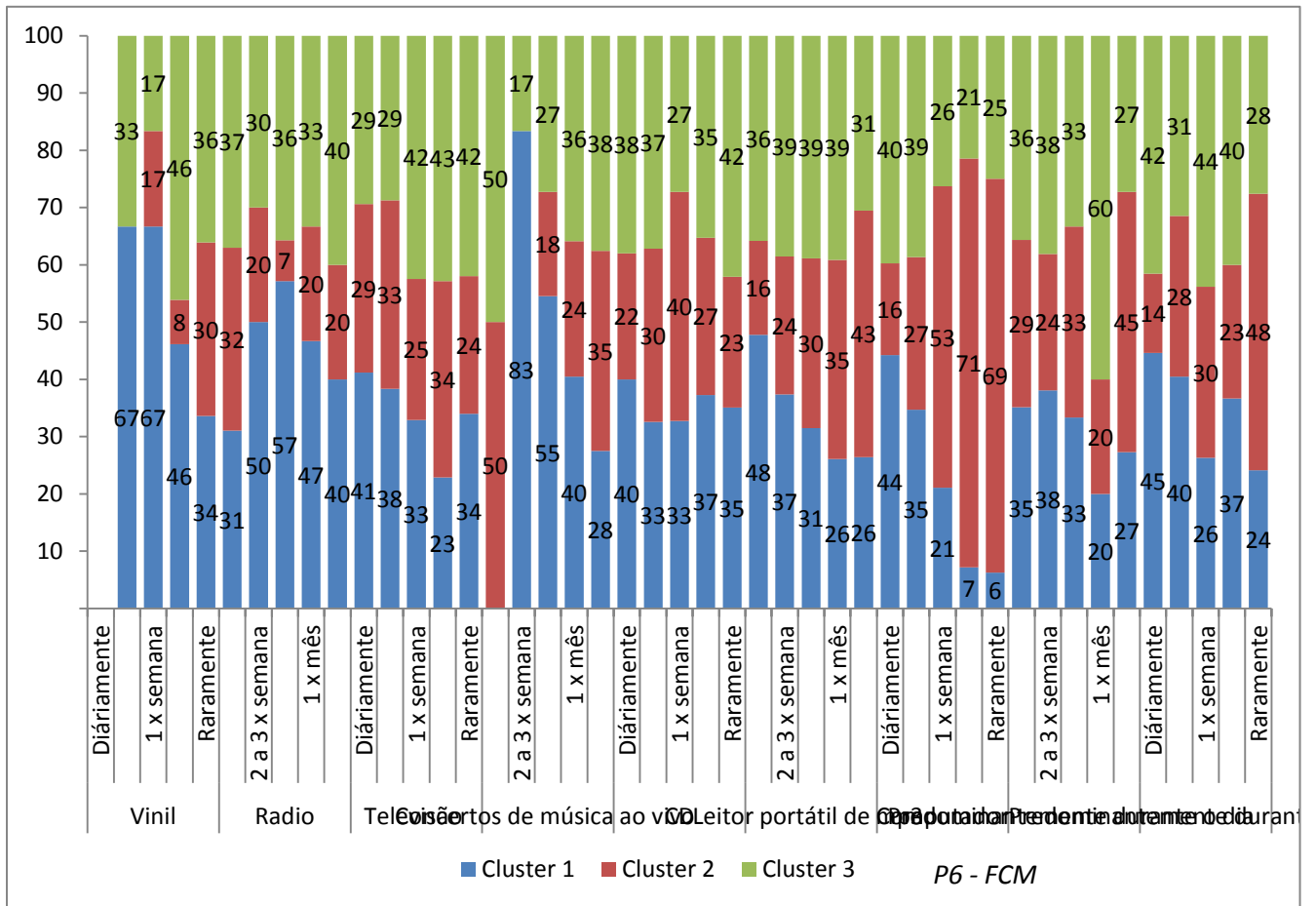
# Segmentação de indivíduos no Facebook que gostam de música

## Frequência na aquisição (FCM)



# Segmentação de indivíduos no Facebook que gostam de música

## Frequência nos modos de audição (FCM)



**Anexo 15: Caracterização dos clusters (variáveis demográficas, K-means)**

		Cluster Number of Case			
		C 1	C 2	C 3	Total
Idade	≤ 30 anos	50,9%	43,4%	51,3%	48,6%
	31 a 40 anos	34,0%	42,2%	42,6%	40,6%
	> 40 anos	15,1%	14,5%	6,1%	10,8%
	Subtotal	100,0%	100,0%	100,0%	100,0%
Sexo	Masculino	30,2%	60,2%	57,4%	52,6%
	Feminino	69,8%	39,8%	42,6%	47,4%
	Subtotal	100,0%	100,0%	100,0%	100,0%
Estado Civil	Solteiro	47,2%	48,2%	67,8%	57,0%
	Casado	45,3%	38,6%	23,5%	33,1%
	Divorciado	7,5%	13,3%	8,7%	10,0%
	Subtotal	100,0%	100,0%	100,0%	100,0%
Escolaridade	≤ 12º Ano	20,8%	8,4%	15,7%	14,3%
	Licenciatura	54,7%	51,8%	59,1%	55,8%
	Mestrado/DR	24,5%	39,8%	25,2%	29,9%
	Subtotal	100,0%	100,0%	100,0%	100,0%

**Anexo 16: Caracterização dos clusters (variáveis demográficas, FCM)**

		Cluster 1	Cluster 2	Cluster 3	Total
Idade	≤ 30 anos	48,6%	38,4%	54,6%	47,8%
	31 a 40 anos	45,7%	43,0%	36,1%	41,5%
	> 40 anos	5,7%	18,6%	9,3%	10,7%
	Total	100,0%	100,0%	100,0%	100,0%
Sexo	Masculino	61,0%	44,2%	55,6%	54,2%
	Feminino	39,0%	55,8%	44,4%	45,8%
	Total	100,0%	100,0%	100,0%	100,0%
Estado Civil	Solteiro	61,0%	43,0%	62,0%	56,2%
	Casado	25,7%	47,7%	32,4%	34,4%
	Divorciado	13,3%	9,3%	5,6%	9,4%
	Total	100,0%	100,0%	100,0%	100,0%
Escolaridade	≤ 12º Ano	14,3%	15,1%	13,0%	14,0%
	Licenciatura	58,1%	54,7%	54,6%	55,9%
	Mestrado/DR	27,6%	30,2%	32,4%	30,1%
	Total	100,0%	100,0%	100,0%	100,0%



*Anexo 17: Lista completa dos graus de pertença*

Indivíduo	Cluster 1	Cluster 2	Cluster 3
1	0,1056	0,1236	<b>0,7708</b>
2	0,2050	<b>0,6777</b>	0,1172
3	0,1177	<b>0,3411</b>	<b>0,5413</b>
4	0,2402	<b>0,5677</b>	0,1921
5	0,0257	0,0593	<b>0,9150</b>
6	0,1281	<b>0,6005</b>	0,2714
7	<b>0,7340</b>	0,1324	0,1337
8	0,1682	<b>0,4934</b>	<b>0,3384</b>
9	0,1085	<b>0,6621</b>	0,2294
10	0,3116	0,0228	<b>0,6656</b>
11	0,2262	<b>0,3532</b>	<b>0,4206</b>
12	0,0061	<b>0,9886</b>	0,0053
13	0,0878	0,2833	<b>0,6289</b>
14	<b>0,3488</b>	<b>0,3828</b>	0,2684
15	0,0227	0,0173	<b>0,9599</b>
16	<b>0,8599</b>	0,0575	0,0825
17	0,0473	0,2087	<b>0,7440</b>
18	<b>0,4746</b>	0,1040	<b>0,4213</b>
19	<b>0,6338</b>	0,0702	0,2960
20	0,1510	<b>0,7198</b>	0,1292
21	<b>0,8551</b>	0,0405	0,1044
22	<b>0,6357</b>	0,2766	0,0877
23	0,2223	0,0372	<b>0,7405</b>
24	<b>0,4941</b>	0,0144	<b>0,4915</b>
25	<b>0,9137</b>	0,0346	0,0517
26	0,0696	0,0491	<b>0,8813</b>
27	0,0597	0,1963	<b>0,7440</b>
28	0,0275	0,0934	<b>0,8791</b>
29	<b>0,8932</b>	0,0535	0,0533
30	0,0234	0,1690	<b>0,8076</b>
31	0,1987	0,1982	<b>0,6031</b>
32	0,3324	0,0407	<b>0,6269</b>
33	0,0421	0,0785	<b>0,8794</b>
34	0,0457	<b>0,6970</b>	0,2573
35	<b>0,7297</b>	0,0118	0,2586
36	0,0309	0,0453	0,9237
37	0,0484	<b>0,7408</b>	0,2108

Segmentação de indivíduos no *Facebook* que gostam de música

38	0,2362	0,2354	0,5284
39	0,2181	<b>0,6439</b>	0,1380
40	<b>0,9449</b>	0,0083	0,0468
41	0,1247	<b>0,5594</b>	0,3159
42	<b>0,8540</b>	0,0226	0,1235
43	<b>0,9498</b>	0,0035	0,0467
44	<b>0,5791</b>	0,0518	<b>0,3691</b>
45	0,1217	0,0344	<b>0,8439</b>
46	<b>0,9404</b>	0,0261	0,0335
47	<b>0,9174</b>	0,0249	0,0577
48	<b>0,7591</b>	0,1005	0,1405
49	0,2899	0,2484	<b>0,4618</b>
50	<b>0,4142</b>	0,0247	<b>0,5611</b>
51	<b>0,9071</b>	0,0223	0,0706
52	0,0984	0,0654	<b>0,8362</b>
53	0,0038	<b>0,9933</b>	0,0029
54	<b>0,7843</b>	0,1684	0,0472
55	<b>0,9506</b>	0,0079	0,0415
56	0,2317	<b>0,5082</b>	0,2601
57	0,0254	<b>0,9249</b>	0,0497
58	<b>0,9550</b>	0,0217	0,0233
59	<b>0,4261</b>	0,2825	0,2914
60	<b>0,9437</b>	0,0367	0,0196
61	<b>0,9437</b>	0,0367	0,0196
62	0,1213	<b>0,8336</b>	0,0451
63	<b>0,4047</b>	0,0215	<b>0,5738</b>
64	<b>0,4126</b>	0,0388	<b>0,5486</b>
65	<b>0,8381</b>	0,0372	0,1247
66	<b>0,7396</b>	0,1311	0,1293
67	0,0187	<b>0,9260</b>	0,0554
68	0,0504	<b>0,9019</b>	0,0476
69	<b>0,3380</b>	<b>0,4392</b>	0,2228
70	<b>0,6579</b>	0,0235	0,3186
71	<b>0,5255</b>	0,1205	<b>0,3540</b>
72	<b>0,4057</b>	<b>0,4081</b>	0,1862
73	0,1825	<b>0,4048</b>	<b>0,4127</b>
74	<b>0,7237</b>	0,0287	0,2476
75	0,0383	<b>0,8412</b>	0,1205
76	0,0316	<b>0,5563</b>	<b>0,4121</b>
77	0,1016	0,0581	<b>0,8403</b>

Segmentação de indivíduos no *Facebook* que gostam de música

78	<b>0,6233</b>	0,1335	0,2433
79	0,0389	0,2498	<b>0,7113</b>
80	0,0806	<b>0,7920</b>	0,1274
81	<b>0,4697</b>	0,0228	<b>0,5075</b>
82	0,2650	<b>0,3359</b>	<b>0,3991</b>
83	<b>0,7843</b>	0,0391	0,1765
84	0,0759	0,1593	<b>0,7648</b>
85	0,0117	<b>0,9349</b>	0,0534
86	<b>0,9836</b>	0,0038	0,0126
87	0,0415	0,1929	<b>0,7655</b>
88	<b>0,8649</b>	0,0195	0,1156
89	0,0319	0,0930	<b>0,8751</b>
90	0,1067	<b>0,4078</b>	<b>0,4855</b>
91	0,0263	<b>0,8261</b>	0,1477
92	0,2740	0,0747	<b>0,6514</b>
93	<b>0,4316</b>	<b>0,4140</b>	0,1544
94	<b>0,3862</b>	0,0432	<b>0,5706</b>
95	0,1047	<b>0,7464</b>	0,1489
96	<b>0,7318</b>	0,1163	0,1519
97	0,1260	<b>0,6545</b>	0,2195
98	0,0245	<b>0,8935</b>	0,0821
99	<b>0,5836</b>	0,3226	0,0938
100	0,1568	<b>0,7901</b>	0,0531
101	<b>0,3921</b>	0,0082	<b>0,5997</b>
102	<b>0,7654</b>	0,0841	0,1505
103	0,1163	<b>0,4640</b>	<b>0,4197</b>
104	<b>0,3450</b>	<b>0,4443</b>	0,2108
105	0,0519	<b>0,9080</b>	0,0402
106	<b>0,6028</b>	0,0137	<b>0,3835</b>
107	0,1524	0,0317	<b>0,8159</b>
108	<b>0,5738</b>	0,1877	0,2386
109	0,0904	<b>0,8545</b>	0,0552
110	0,2395	<b>0,5072</b>	0,2533
111	0,3318	0,3196	<b>0,3486</b>
112	0,0338	0,1101	<b>0,8561</b>
113	0,2910	0,2418	<b>0,4672</b>
114	0,0271	<b>0,9203</b>	0,0526
115	0,3206	<b>0,4251</b>	0,2543
116	0,0051	<b>0,9688</b>	0,0261
117	<b>0,5162</b>	0,1431	<b>0,3407</b>

Segmentação de indivíduos no *Facebook* que gostam de música

118	0,1652	0,1435	<b>0,6912</b>
119	<b>0,5897</b>	0,2283	0,1820
120	<b>0,5181</b>	0,0928	<b>0,3891</b>
121	0,2635	<b>0,5471</b>	0,1894
122	0,0610	<b>0,8098</b>	0,1292
123	<b>0,9476</b>	0,0265	0,0259
124	<b>0,6681</b>	0,0492	0,2828
125	<b>0,7288</b>	0,2065	0,0647
126	<b>0,8974</b>	0,0346	0,0680
127	0,1986	<b>0,3416</b>	<b>0,4597</b>
128	<b>0,8135</b>	0,0363	0,1502
129	<b>0,9654</b>	0,0058	0,0288
130	<b>0,7576</b>	0,0939	0,1484
131	<b>0,8941</b>	0,0547	0,0512
132	<b>0,9521</b>	0,0033	0,0446
133	0,0610	<b>0,9131</b>	0,0260
134	<b>0,7474</b>	0,1223	0,1303
135	<b>0,9068</b>	0,0256	0,0676
136	0,1144	0,1080	<b>0,7776</b>
137	<b>0,8559</b>	0,0151	0,1289
138	0,0761	0,0972	<b>0,8267</b>
139	<b>0,9992</b>	0,0001	0,0007
140	<b>0,6972</b>	0,2107	0,0921
141	0,1472	0,0226	<b>0,8302</b>
142	0,2912	<b>0,3552</b>	<b>0,3536</b>
143	0,3044	0,1145	<b>0,5810</b>
144	0,3040	0,1200	<b>0,5760</b>
145	0,2188	0,0185	<b>0,7627</b>
146	0,0010	<b>0,9909</b>	0,0081
147	0,0429	<b>0,7934</b>	0,1637
148	0,0843	<b>0,6347</b>	0,2810
149	<b>0,9325</b>	0,0379	0,0296
150	0,0364	0,0392	<b>0,9244</b>
151	0,1113	<b>0,4323</b>	<b>0,4564</b>
152	<b>0,4548</b>	0,1029	<b>0,4423</b>
153	0,2278	<b>0,3815</b>	<b>0,3907</b>
154	0,0196	<b>0,9673</b>	0,0130
155	0,0931	0,0353	<b>0,8716</b>
156	<b>0,8587</b>	0,0460	0,0953
157	0,0447	0,1788	<b>0,7765</b>

Segmentação de indivíduos no *Facebook* que gostam de música

158	0,0931	0,0768	<b>0,8301</b>
159	<b>0,6530</b>	0,2220	0,1249
160	0,2292	<b>0,6073</b>	0,1634
161	0,0269	<b>0,8736</b>	0,0995
162	<b>0,6636</b>	0,1387	0,1977
163	0,1287	0,0416	<b>0,8297</b>
164	0,1235	<b>0,5464</b>	0,3301
165	0,0084	<b>0,9356</b>	0,0560
166	0,0005	0,0047	<b>0,9948</b>
167	0,0233	<b>0,9481</b>	0,0286
168	<b>0,9684</b>	0,0013	0,0303
169	<b>0,5090</b>	<b>0,3864</b>	0,1046
170	<b>0,3360</b>	<b>0,5673</b>	0,0967
171	0,0627	0,0038	<b>0,9335</b>
172	0,0325	<b>0,9142</b>	0,0533
173	<b>0,3830</b>	<b>0,4443</b>	0,1727
174	0,0490	<b>0,8928</b>	0,0583
175	0,0461	0,0241	<b>0,9297</b>
176	0,2358	<b>0,3802</b>	<b>0,3840</b>
177	<b>0,7621</b>	0,0595	0,1784
178	<b>0,7936</b>	0,0450	0,1614
179	<b>0,9489</b>	0,0242	0,0270
180	0,1533	<b>0,5095</b>	<b>0,3373</b>
181	<b>0,4579</b>	0,0399	<b>0,5022</b>
182	0,2205	<b>0,3500</b>	<b>0,4295</b>
183	<b>0,3922</b>	0,2885	0,3193
184	0,0323	0,0364	<b>0,9314</b>
185	0,2380	0,1332	<b>0,6288</b>
186	<b>0,4837</b>	0,0110	<b>0,5053</b>
187	<b>0,5949</b>	0,0115	<b>0,3936</b>
188	0,0959	0,0069	<b>0,8972</b>
189	0,3094	0,0350	<b>0,6556</b>
190	0,2184	0,1766	<b>0,6050</b>
191	0,2080	<b>0,5533</b>	0,2388
192	<b>0,3465</b>	0,1449	<b>0,5086</b>
193	<b>0,9816</b>	0,0047	0,0137
194	<b>0,4508</b>	0,0351	<b>0,5141</b>
195	0,0048	<b>0,9567</b>	0,0385
196	0,1705	<b>0,4337</b>	<b>0,3958</b>
197	0,1673	0,0149	<b>0,8178</b>

Segmentação de indivíduos no *Facebook* que gostam de música

198	<b>0,9040</b>	0,0350	0,0610
199	<b>0,8265</b>	0,1504	0,0232
200	<b>0,7942</b>	0,0199	0,1859
201	0,1178	0,0361	<b>0,8461</b>
202	<b>0,4209</b>	0,1173	<b>0,4618</b>
203	0,0678	0,1698	<b>0,7624</b>
204	<b>0,9882</b>	0,0049	0,0069
205	0,0737	<b>0,8384</b>	0,0879
206	0,1825	0,1476	<b>0,6699</b>
207	0,1822	<b>0,5675</b>	0,2503
208	0,1099	0,0900	<b>0,8001</b>
209	0,0461	<b>0,8240</b>	0,1300
210	0,3221	0,0835	<b>0,5943</b>
211	0,1955	0,0433	<b>0,7612</b>
212	0,0706	0,0704	<b>0,8590</b>
213	<b>0,9447</b>	0,0121	0,0432
214	<b>0,3717</b>	<b>0,4087</b>	0,2197
215	0,0327	0,2353	<b>0,7320</b>
216	0,0348	<b>0,4843</b>	<b>0,4809</b>
217	0,0888	<b>0,8150</b>	0,0962
218	0,1256	0,0687	<b>0,8058</b>
219	<b>0,5592</b>	0,1289	0,3119
220	<b>0,7477</b>	0,1008	0,1515
221	0,0551	<b>0,9155</b>	0,0294
222	0,3045	0,0746	<b>0,6209</b>
223	0,2277	0,1598	<b>0,6125</b>
224	0,0150	0,0215	<b>0,9635</b>
225	<b>0,4371</b>	0,2480	0,3149
226	<b>0,5841</b>	0,3241	0,0918
227	<b>0,9322</b>	0,0399	0,0279
228	<b>0,6024</b>	0,0776	0,3201
229	0,1112	0,2162	<b>0,6726</b>
230	<b>0,6979</b>	0,0130	0,2891
231	<b>0,9940</b>	0,0010	0,0050
232	<b>0,6959</b>	0,0406	0,2636
233	0,0506	0,0292	<b>0,9201</b>
234	0,0435	0,2330	<b>0,7235</b>
235	<b>0,4223</b>	<b>0,3511</b>	0,2266
236	0,0769	<b>0,7432</b>	0,1799
237	0,0699	0,0607	<b>0,8694</b>

Segmentação de indivíduos no *Facebook* que gostam de música

238	<b>0,4303</b>	0,2986	0,2711
239	0,1383	<b>0,7810</b>	0,0807
240	<b>0,4351</b>	0,0959	<b>0,4689</b>
241	0,2390	0,0896	<b>0,6714</b>
242	0,0698	0,2266	<b>0,7037</b>
243	0,3190	<b>0,5531</b>	0,1280
244	0,1319	0,2145	<b>0,6535</b>
245	0,2032	<b>0,5868</b>	0,2100
246	<b>0,8503</b>	0,0136	0,1362
247	0,1601	<b>0,7712</b>	0,0687
248	0,1621	<b>0,7090</b>	0,1289
249	0,0493	0,1387	<b>0,8120</b>
250	0,0080	<b>0,9716</b>	0,0204
251	0,0434	<b>0,8786</b>	0,0779