

LICENCIATURA EM INFORMÁTICA DE GESTÃO



Aluna Maria Silvina Lopes Alves nº 20081028

Orientadora DraAna Margarida Gomes Alexandre

de Humanidades e Tecnologias Humani nibil alienum

Resumo

Desde o ano de 2008, ano em que se deu o colapso da economia Mundial devido à

declaração de bancarrota da empresa norte-americana Lehman Brothers, que Portugal

tem vindo a viver uma profunda crise financeira e social, resultando num acentuado

acréscimo da taxa de desemprego no nosso país. Para colmatar este problema, isto é,

para que sejam criados novos postos de trabalho (a fim de diminuir a referida taxa de

desemprego) é necessário que a economia nacional seja estimulada. Uma das formas de

sabermos quais as áreas do nosso país onde a incidência da taxa de desemprego é mais

elevada e desta forma desenvolver acções que possam reverter esses mesmos números é

fundamental identificar e conhecer o perfil destes, ou seja, desenvolver um Modelo de

Segmentação (modelo comportamental de Data Mining).

No seguimento desta temática, este projecto consistiu no desenvolvimento e

implementação, do modelo de segmentação de beneficiários do sistema de Segurança

Social com o objectivo de criar e aplicar estratégias e campanhas de contacto sobre cada

um dos perfis dos desempregados identificados. Desta forma, utilizando este modelo,

será possível reactivar a situação profissional de cada um dos indivíduos desempregados

registados na Segurança Social.

Classificação JEL: C13, C38, C43, C44, E24

Palavras-Chave: Modelo de Segmentação, Desemprego, Análise de Clusters,

Distribuição Geográfica.

ii

UNIVERSIDADE LUSÓFONA de Humanidades e Tecnologias Humani nibil alienum

Abstract

Since 2008, year where the collapse of the worldwide economy happened due to the

declaration of bankruptcy of the U.S. firm Lehman Brothers, Portugal has been

experiencing a deep financial and social crisis, resulting in a marked increase in the

unemployment rate in our country. To overcome this problem, i.e., to create new jobs in

order to reduce this rate of unemployment, the stimulation of the national economy is

necessary. One way of knowing which areas of Portugal has the highest incidence of

unemployment and thus carry out actions that can reverse these numbers, it becomes

essential to identify and meet their profile, i.e., to develop a segmentation model so that

we can decrease these numbers in a more localized way, through the use of a clustering

model (Data Mining behavioral model).

According to this matter, this project consisted in the development and

implementation of a clustering model of beneficiaries of the Portuguese Social Security

system in order to develop and implement strategies and custom mail campaigns in

order to identify the profiles of the unemployed people. With this model, it will be

possible to reactivate the employment of each one of the registered unemployed in

Social Security.

JEL Classification: C13, C38, C43, C44, E24

Keywords: Clustering Model, Unemployment, Cluster Analysis, Geographical

Distribution

iii



Agradecimentos

Um agradecimento muito especial à Professora Ana Margarida Gomes Alexandre, por ter aceite e ter-se disponibilizado para me orientar no desenvolvimento deste projecto. Obrigada pela confiança, dedicação, disponibilidade e, acima de tudo, pela amizade demonstrada.

Aos meus irmãos, cunhada e sobrinhos pela força e apoio incondicional e por terem compreendido as minhas ausências tão frequentes.

Um agradecimento muito especial às minhas amigas, Dr.ª Aline Silva Santos e Mestre Ester Liñares, por todo o apoio, dedicação, conhecimento e confiança que me transmitiram. Obrigada por acreditarem em mim e terem estado sempre comigo.

Não posso, contudo, deixar de dar uma palavra de agradecimento ao Dr. J. Monteiro de Andrade, meu coordenador no serviço onde desempenho funções, pela disponibilidade sempre demonstrada.



Índice

Resumo	ii
Abstract	iii
Agradecimentos	iv
Índice	5
Índice de Figuras	6
Índice de Tabelas	
1. Introdução	
1.1. Objectivos e Descrição do Tema	9
2.1. Gestão do relacionamento dos beneficiários do sistema de segurança social português	11 13
3.1. Metodologia 3.2. Descrição e Análise dos Dados 3.3. Desenvolvimento do Modelo 4. Resultados	15 18 18
5. Conclusões e Trabalho Futuro	28
5.1. Conclusões	28
5.2. Trabalho Futuro	
6. Bibliografia	
I. Anexos	
Anexo 1. Código utilizado no Pré -Tratamento dos Dados	XXXII
Anexo 2. Variáveis de Entrada	
Anexo 3. Matriz de Distâncias segundo a Segmentação Hierárquica	. XXXIII
Anexo 4. Dimensão dos Segmentos obtidos	. XXXIV
Anexo 5. Semelhantes entre Segmentos	. XXXIV
Anexo 6. Variáveis Relevantes para a obtenção dos Segmentos	XXXV
Anexo 7. Source Code do Projecto	
Glossário	



Índice de Figuras

Figura 1 - Metodologia SAS de "Implementação de Plataformas Inteligentes" u	tilizada
neste projecto	16
Figura 2 - Fluxo do processo desenvolvido para o Modelo de Segmentação de	
Beneficiários	19
Figura 3 - Gráfico "Segment Plot"	21
Figura 4 - Gráfico "Segment Profile"	22
Figura 5 - Matriz de Distância segundo a Segmentação hierárquica	XXXIII
Figura 6 - Dimensão dos segmentos obtidos	XXXIV



Índice de Tabelas

Tabela 1 - Interpretação dos Segmentos obtidos	24
Tabela 2 - Variáveis de entrada	XXXIII
Tabela 3 - Semelhantes entre Segmentos	XXXIV
Tabela 4 - Variáveis relevantes para a obtenção dos Segmentos	XXXV



1. Introdução

Nos últimos tempos Portugal tem vivido uma crescente crise financeira e social, o que se traduz em diversos factores negativos, entre eles numa maior taxa de desemprego. Torna-se assim fundamental que a economia nacional seja estimulada e para tal tem de existir um maior entendimento das características e do perfil dos desempregados. Desta forma será possível efectuar acções de reinserção dos mesmos no mercado de trabalho tendo em conta cada um dos perfis.

A reparação da situação de desemprego realiza-se através de medidas activas e de medidas passivas. As medidas activas integram:

- O pagamento, por uma só vez, do montante global das prestações de desemprego, para criação do próprio emprego;
- A possibilidade de acumular subsídio de desemprego parcial com trabalho por conta de outrem a tempo parcial ou com actividade profissional independente;
- A suspensão total ou parcial das prestações de desemprego, durante a frequência de curso de formação profissional com atribuição de compensação remuneratória;
- A manutenção das prestações de desemprego, durante o período de exercício de actividade ocupacional promovida pelos Centros de Emprego;
- Outras medidas de política activa de emprego que promovam a melhoria dos níveis de empregabilidade e a reinserção no mercado de trabalho, de beneficiários das prestações de desemprego, em termos a definir em legislação própria.

Já as medidas passivas concretizam-se pela atribuição das prestações de desemprego.

A protecção no desemprego abrange, ainda, medidas excepcionais e transitórias previstas em legislação própria que, por não considerar de interesse para o presente trabalho, não enumero. As pessoas abrangidas neste universo são os beneficiários residentes em território nacional, ou seja, os trabalhadores abrangidos pelo regime geral de segurança social dos trabalhadores por conta de outrem e os pensionistas de invalidez que sejam declarados aptos para o trabalho em exame de revisão de incapacidade.



Com as dificuldades económicas e financeiras a cresceram cada vez mais o aumento do desemprego em Portugal torna-se um factor consequente pondo em causa toda a sustentabilidade social.

1.1. Objectivos e Descrição do Tema

O objectivo deste projecto consistiu no desenvolvimento e implementação de um Modelo Comportamental de Segmentação dos indivíduos desempregados em Portugal.

A partir da identificação e análise do perfil dos desempregados foi possível a criação e aplicação de estratégias e campanhas de contacto sobre cada um dos perfis. O intuito consistiu em reactivar a situação profissional de cada um dos beneficiários do sistema de segurança social português.

Para um entendimento da problemática em estudo foi feita uma breve análise às causas e consequências da crise de desemprego em Portugal na primeira década do século XXI.

1.2. Importância do Tema

A temática do desemprego tem ganho cada vez maior importância devido às grandes agitações económicas, sociais e empresariais que ocorrem na actualidade. É, então, fundamental que exista a capacidade de identificar e compreender quais as características e padrões comportamentais associados aos desempregados em Portugal.

Sendo um tema actual e preocupante, este estudo torna possível a definição de processos e métricas que permitam um modelo de gestão de relacionamento com os beneficiários e potenciais empregadores. Assim é possível melhorar as condições de vida da população portuguesa desempregada, permitindo o crescimento social e económico do país.



2. Enquadramento Teórico

As características básicas do desemprego em Portugal estão ligadas ao seu carácter predominantemente cíclico, à diferente intensidade com que atinge algumas regiões, segmentos de população e níveis etários, bem como à fragilidade social que lhe está associada. De acordo com os resultados do Inquérito ao Emprego relativos ao 1.º trimestre do corrente ano¹ a população desempregada em Portugal foi estimada em 688,9 mil indivíduos, sendo que: (1) 51,4% da população desempregada era constituída por homens e 48,6% por mulheres; (2) 18,0% dos indivíduos pertenciam ao grupo etário dos 15 aos 24 anos, 28,5% ao dos 25 aos 34 anos, 23,3% ao dos 35 aos 44 anos e 30,3% ao dos 45 e mais anos. No entanto, os dados extrapolados da aplicação mostram-nos uma realidade diferente, o que se justifica por dizerem respeito ao ano de 2010.

Existem diversos motivos para o aumento do desemprego em Portugal. Relativamente aos jovens verificamos que os obstáculos mais relevantes com que se defrontam na procura de emprego dizem respeito à falta de experiência profissional e de mecanismos eficazes que facilitem a transição da escola para a vida activa, apresentando, por isso, uma estrutura habilitacional e um nível de literacia bastante menos desfavorável que os da população adulta. O processo de integração tem vindo a tornar-se mais longo e difícil, não se limitando a um momento de passagem da escola para um emprego estável, sendo cada vez mais um itinerário descontínuo, ao longo de vários anos. Tornase então fundamental desenvolver medidas que permitam uma inserção rápida no mercado de trabalho, tais como: (1) medidas que visam a experiência em contexto de trabalho; (2) apoios ao desenvolvimento de iniciativas empresariais e (3) à contratação de jovens.

Contrariamente aos resultados obtidos na presente investigação, verifica-se através de dados do INE consultados que em Portugal, a taxa de desemprego das mulheres é sistematicamente superior à dos homens, em qualquer um dos grupos etários, apesar de o comportamento cíclico das taxas de desemprego daquelas ser próximo da dos homens, embora um pouco mais errático, agravando-se a situação no caso das jovens mães, expostas a discriminações e exclusões variadas.

_

¹ Divulgados em 18 de Maio pelo INE.



O desemprego tem vindo a agravar-se desde 2009, ano de crise financeira e económica mundial.

Torna-se, por isso, cada vez mais importante analisar todo o percurso do desempregado em relação ao sistema de protecção social em que se encontra inserido.

2.1. Gestão do relacionamento dos beneficiários do sistema de segurança social português

Quando reflectimos sobre a gestão do relacionamento dos beneficiários do sistema de segurança social, temos subjacentes os conceitos de sistema de informação, sistemas de suporte à decisão e o conjunto de tecnologias e processos utilizados na gestão de relações (CRM), os quais se relacionam entre si.

Da boa aplicação destes conceitos surge uma boa gestão do relacionamento dos beneficiários com o sistema de segurança social português. Torna-se assim fundamental um correcto entendimento destes conceitos.

Segundo Agostinho (2000) um "Sistema de informação é um conjunto de componentes interligados, que colecta, processa, armazena e distribui as informações para suportar as tomadas de decisões e controlar a organização".

Nos anos seguintes, a mesma definição não sofreu alterações de maior, pois, em 2008 Bagad define "Sistema de Informação como um grupo de elementos organizados com o propósito de apoiar o gestor e a tomada de decisão operacional ". Por sua vez o conceito de Sistemas de Suporte à Decisão (SSD) está estritamente relacionado com o conceito de Sistemas de Informação (SI), uma vez que apoiam a decisão táctica e estratégica, orientada ao assunto, recorrendo à análise de um conjunto de dados históricos de múltiplas fontes, sendo o seu intuito modelação de dados, tomada de decisões e realização de *reporting*.

Um SSD é um sistema que proporciona aos utilizadores, não só um acesso rápido à "sua" informação, mas também a capacidade para realizar a sua análise e formatação à medida das "suas" necessidades. Assim, um SSD tem a função de criar conhecimento a partir de informação, utilizando ferramentas sofisticadas de armazenagem e análise de



base de dados, internas e externas, para permitir que os gestores garantam soluções adequadas às questões essenciais, permitindo uma fácil interacção com os utilizadores.

O que diferencia o SSD dos demais tipos de SI é que, por ser um sistema baseado em computadores, que através de informações e modelos especializados ajudam a resolver problemas organizacionais, tem como principal função apoiar o processo decisório em áreas de gestão e de planeamento estratégico. É importante que os SSD retratem a cultura organizacional, não servindo apenas para atender às necessidades específicas do utilizador, mas que seja orientado para pessoas que tomam decisões. Estes sistemas devem assim ser flexíveis na pesquisa, no acesso e na manipulação das informações, utilizando uma *interface* o mais amigável possível para satisfazer às expectativas gerais das organizações.

Existem diversos tipos de SSD, sendo o *Customer Relationship Managament* (CRM) um deles. O CRM traduz um sistema capaz de fazer a diferença na competitividade empresarial, permitindo a interacção com o beneficiário através de pontos de contacto automatizados ou semi-automatizados, e o acesso e distribuição de informações através de um sistema de *marketing* que administre as *interfaces* e forneça informações por meio de um SSD. Este tipo de sistema pode ser definido como um conjunto de tecnologias e processos utilizados na gestão das relações com os beneficiários de forma a reter informações úteis sobre os mesmos, com vista à compreensão e gestão. Os seus objectivos de implementação serão a maximização do conhecimento do beneficiário, a previsão de comportamentos, a actuação nos mais diversos canais, a integração da informação e a obtenção de condições de emprego.

Se o organismo que gere a Segurança Social em Portugal tomasse decisões no processo de relacionamento com os beneficiários do subsídio de desemprego baseadas num sistema de CRM, poderia conhecer mais profundamente o perfil desses indivíduos e consequentemente inverter a situação do desemprego em Portugal.



Esse conhecimento permitiria, não apenas a atribuição do subsídio, mas criaria novas oportunidades a partir do cruzamento do histórico de informações, do perfil do desempregado, da tendência de comportamento.

A tecnologia CRM e SSD, quando acompanhadas de toda uma infra-estrutura tecnológica e organizacional, pode ser determinante para se atingir as metas organizacionais, ou seja, não basta apenas ter informação, é necessário conhecê-la para obter vantagem.

2.2. Data Mining como processo fundamental na análise do desemprego

Com o propósito de analisar o estado actual do desemprego em Portugal, é fundamental recorrer a processos que permitam tratar uma grande quantidade de dados e retirar todo um conjunto de informação e conhecimento. Ao conjunto destes processos dá-se o nome de Data Mining (DM). Assim, o DM consiste num processo de selecção, exploração e modelação de grandes bases de dados, com o intuito de transformar dados em informação e a informação em conhecimento. Assim, DM pode ser definido como um processo de exploração e análise, de forma automática ou semi-automática, de grandes quantidades de dados para descobrir padrões e regras significativas. Existem dois tipos de modelos de DM: (1) Modelos Descritivos ou Comportamentais e (2) Modelos Preditivos. Os modelos descritivos ou comportamentais permitem identificar e descrever o comportamento dos clientes, o que na temática em estudo se traduz nos beneficiários. Já os modelos preditivos têm como objectivo prever um determinado evento e entender os motivos que o originam.

Um dos principais objectivos de Data Mining consiste em detectar grupos de observações com perfis semelhantes dentro do mesmo grupo e perfis o mais heterogéneos entre grupos. Estes grupos são coloquialmente denominados de *clusters*, e podem ser obtidos através do desenvolvimento de um Modelo Comportamental de Data Mining – Modelo de Segmentação.

É de referir a importância da identificação de perfis na análise do desemprego em Portugal. A partir da identificação e agrupamento dos beneficiários em diferentes



clusters/grupos é possível identificar e compreender as características que permitam caracterizar e diferenciar os diferentes clusters.

Utilizada pela primeira vez por Tryon (1939), a análise de clusters (ou modelo de segmentação) consiste num conjunto de métodos de estatística multivariada que engloba diferentes algoritmos de classificação e optimização para organizar informações sobre variáveis e formar grupos heterogéneos entre si e os mais homogéneos dentro de cada grupo.

Para Pestana e Gageiro (2000) "a análise de *clusters* é um procedimento multivariado para detectar grupos homogéneos nos dados, podendo os grupos ser constituídos por variáveis ou casos". Um ano mais tarde Reis (2001) defendia que "os métodos de análise de *clusters* são procedimentos de estatística multivariada que tentam organizar um conjunto de indivíduos para os quais é conhecida informação detalhada, em grupos relativamente homogéneos".

Assim, e tendo em conta o intuito deste trabalho, o objectivo básico consiste em descobrir agrupamentos naturais de beneficiários por características semelhantes e características que apresentem, simultaneamente, coesão interna e separação externa. Para tal, pode-se recorrer a dois tipos de agrupamentos: (1) agrupamento hierárquico ou (2) agrupamento não hierárquico. No primeiro agrupamento a classificação é feita com base nas similaridades (ou dissimilaridades) entre os beneficiários em estudo. Já no segundo tipo de agrupamento a classificação dos grupos é determinada pelo próprio utilizador, já que o número de grupos a formar é definida pelo mesmo.

Neste trabalho de investigação a metodologia de análise de *clusters* permitiu encontrar grupos de desempregados registados com as mesmas características, de modo a obter-se uma melhor compreensão do problema e a apresentação de possíveis soluções.



3. Caso Prático

Neste trabalho final de curso irei identificar, explicar as características e padrões comportamentais associados aos desempregados em Portugal (análise de perfil). Para tal, irá ser desenvolvido e implementado um Modelo de Segmentação (modelo comportamental de *Data Mining*). Desta forma será possível a criação e aplicação de estratégias e campanhas de contacto sobre cada um dos perfis dos desempregados identificados, com o objectivo de reactivar a situação profissional de cada um dos indivíduos registados na Segurança Social.

Os dados trabalhados são reais, referem-se ao ano de 2010 e foram fornecidos pelo Instituto de Informática, I.P., do Ministério da Solidariedade e Segurança Social.

3.1. Metodologia

A metodologia utilizada no desenvolvimento deste projecto baseou-se na metodologia que o SAS utiliza para o desenvolvimento de projectos de *Data Mining* - Metodologia "Implementação de Plataformas Inteligentes" sendo este projecto desenvolvido de acordo com 7 fases (figura 1).



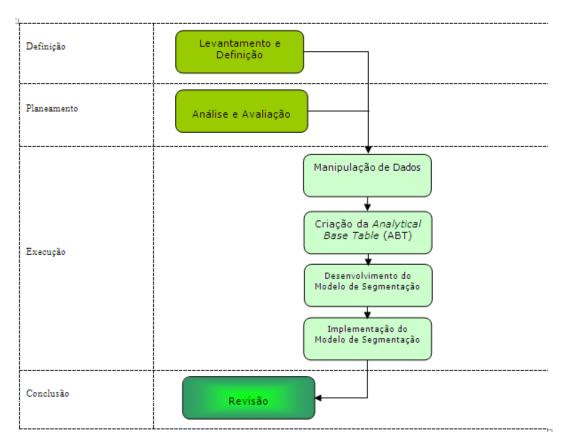


Figura 1 - Metodologia SAS de "Implementação de Plataformas Inteligentes" utilizada neste projecto

Cada uma das fases acima identificadas foi realizada tendo em conta objectivos distintos.

- Levantamento e definição: Nesta fase consultou-se vários Organismos públicos com o intuito de conhecer o universo dos desempregados em Portugal, com o intuito de poder definir a estratégia a adoptar nos processos seguintes;
- (2) Na fase de Análise e avaliação foram avaliados os dados existentes quanto à sua qualidade e histórico, variáveis explicativas e o universo temporal (ano de 2010). Nesta fase foi também escolhido o software a utilizar para desenvolver e implementar o modelo - o SAS Enterprise Miner;
- (3) Manipulação de dados, nesta fase foram implementadas as variáveis a utilizar.



- (4) Definição da *Analytical Base Table* (ABT): A ABT é uma tabela de desenvolvimento e aferição dos modelos analíticos, com todas as variáveis de entrada necessárias ao cálculo de cada componente utilizada no processo de obtenção do modelo de segmentação. Cada registo presente na ABT representa um beneficiário e em coluna existem todo um conjunto de variáveis que traduzem o comportamento dos beneficiários (sexo, idade, remuneração, antiguidade do processo, tipo de subsídio, código do distrito, nacionalidade, etc.). Assim, nesta etapa, foi preparada a tabela analítica com os dados e variáveis relevantes para a obtenção de cada componente utilizada no processo de obtenção do segmento em que cada um dos beneficiários se enquadra;
- (5) Desenvolvimento do modelo de segmentação: Nesta fase foi analisada a homogeneidade do universo considerado, verificando-se um conjunto de características comuns da população, razão pela que se procedeu ao agrupamento dos desempregados. Para tal, procedeu-se à identificação do perfil dos beneficiários, de acordo com diversas variáveis, tais como a idade, o rendimento, o sexo, etc.;
- (6) Implementação do modelo: através da disponibilização do modelo em produção é possível correr o processo periodicamente. O objectivo consiste em ficar a conhecer o perfil dos desempregados ao longo do tempo (através da aplicação do modelo desenvolvido) e em identificar o perfil de cada novo desempregado;
- (7) Revisão do processo e a conclusão: Na última fase foi efectuada toda a revisão, quer do processo de desenvolvimento quer do de implementação do modelo de Segmentação de forma a garantir o correcto desenvolvimento.



3.2. Descrição e Análise dos Dados

Os beneficiários analisados têm idades compreendidas entre os 17 e os 81 anos, sendo que 50% dos beneficiários, segundo a idade, têm até 38 anos. A idade média dos beneficiários é de 40 anos, podendo existir um desvio em termos de idade sensivelmente de 12 anos (11,75882), o que significa a existência de beneficiários em idade activa entre os 28 e os 52 anos.

Os beneficiários objecto de análise neste estudo residem maioritariamente na região do Porto.

3.3. Desenvolvimento do Modelo

O modelo de Segmentação de beneficiários foi desenvolvido de acordo com diversas etapas, que têm objectivos distintos:

- (1) Fase 1. Definição da Metadata;
- (2) Fase 2. Exploração e Modificação das Variáveis;
- (3) Fase 3. Segmentação e Análise do Perfil dos Beneficiários;
- (4) Fase 4. Implementação do Modelo.



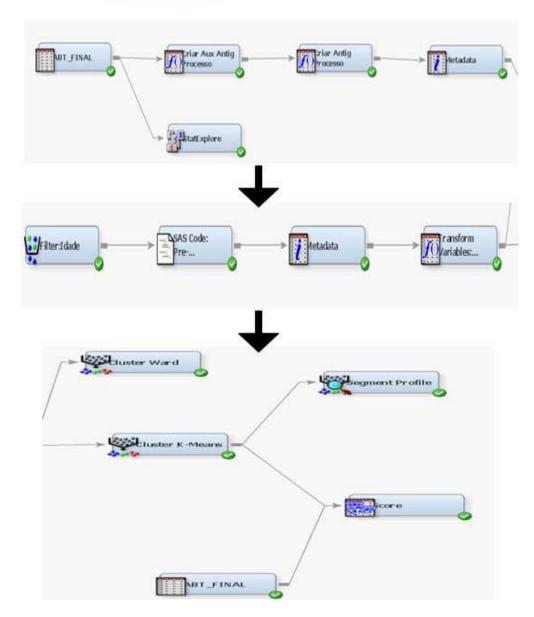


Figura 2 - Fluxo do processo desenvolvido para o Modelo de Segmentação de Beneficiários

A visão global do processo diz respeito ao estudo do perfil dos desempregados inscritos no centro de emprego na área de residência, através da aplicação de um Modelo de Segmentação.

Vejamos então com maior detalhe o que foi realizado sobre cada etapa:

- (1) Na fase 1 procedeu-se à identificação do tipo e papel que cada variável assumirá no modelo (metadata);
- (2) Na fase 2 explorou-se os dados existentes de forma descritiva, de forma a compreender os mesmos (nó de "*Explore Variables*"). Na mesma fase procedeu-se



ainda à modificação e transformação de variáveis (nó "*Transform Variable*"), com quatro grandes objectivos:

- criar uma variável que retrata-se a antiguidade de cada processo;
- suavizar a distribuição das variáveis, através de transformações feitas sobre as variáveis existentes (tipicamente as transformações são úteis para melhorar o ajustamento de um dado modelo aos dados). Para tal utilizou-se o método "Bucket", com o intuito de agrupar as variáveis intervalares em classes tendo em conta a frequência entre as classes, tais como:
 - o COD_BENEFICIO_LANC código do tipo de benefício
 - o COD_DISTRITO respectiva descrição,
 - o COD_PAIS_NACIONALIDADE respectiva descrição
 - o COD_TIPO_BENEFICIO_LANC respectiva descrição
- eliminar *outliers*, tendo-se retido apenas os beneficiários cuja idade se situava entre os 18 os 70 anos (nó de "*Filter*");
- pré-tratamento dos dados (nó de "SAS Code"), onde foi possível criar uma variável que reflicta as classes de remuneração média mensal (variável designada de Desc_Remuner), uma vez que a grande maioria dos valores de remuneração eram diferentes entre si. Desta maneira, com a criação da variável descrita acima, foi possível a síntese dos valores em 4 classes (Anexo 1)
 - o Valor = Mínima (menos de 453 €);
 - o Valor = Reduzida (maior ou igual a 453 e menor que 600 €);
 - o Valor = Intermédia (maior ou igual a 600 e menor que 900€);
 - o Valor = Elevada (maior que 900 €)
- (3) Na fase seguinte procedeu-se ao agrupamento dos beneficiários e análise do perfil dos mesmos. Para tal utilizou-se o nó de "Clusters" e "Segment Profile". Através do nó de "Clusters" foi possível desenvolver clusters através de dois métodos distintos: (1) cluster hierárquico e (2) cluster não hierárquico.



Utilizando uma abordagem hierárquica obteve-se uma solução óptima de 10 segmentos (anexo 2). Para a obtenção deste resultado foram utilizadas as variáveis que retratam o sexo, a remuneração, a idade, a nacionalidade, a antiguidade do processo, o distrito, o código de benefício lançamento e o código tipo de benefício lançamento (anexo 3). Contudo, e após uma visualização da matriz de distância verificou-se que alguns dos segmentos obtidos apresentavam distâncias próximas, o que significava que apresentavam características comuns. Deste modo, e de forma interactiva, utilizou-se uma abordagem não hierárquica em que fomos reduzindo o número de segmento a formar e visualizando a matriz de distâncias. Como resultado final obtivemos uma solução de 3 segmentos (anexos 4 e 5), onde as distâncias foram as maiores possíveis. Optou-se pela utilização deste tipo de *cluster* uma vez que é um método mais rápido e permite tratar um grande volume de dados. Para a obtenção da solução óptima as variáveis mais relevantes foram o Sexo, o Código de Beneficiário e o Valor da Remuneração de referência mensal (anexo 6). A figura abaixo ilustra o comportamento de cada segmento, o que permite identificar quais as característica dos mesmos (resultados estes que se encontram no capítulo 4).

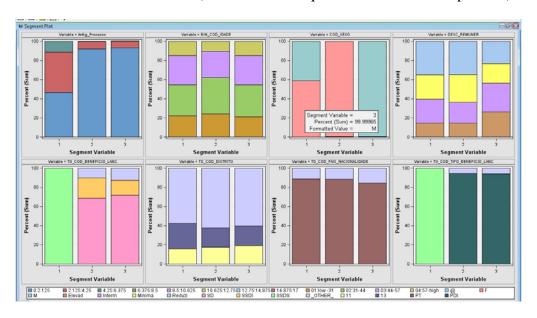


Figura 3 – Gráfico "Segment Plot"



Através do gráfico representado na Figura 3 é possível identificar e analisar como é que as variáveis se distribuem pelos respectivos segmentos, o que significa que permite identificar o comportamento de cada segmento. A título de exemplo, podemos constatar que a antiguidade do processo é menor nos segmentos 2 e 3. Já o código de distrito é semelhante nos 3 segmentos, o que à partida pode significar que a variável, por si só, não é estatisticamente relevante para distinguir os segmentos.

Um outro gráfico a analisar é o "*Profile Segment*" (gráfico 4), que nos permite identificar quais as variáveis mais relevantes para explicar cada um dos segmentos e se as características dos segmentos em causa diferem da população. Ou seja, este gráfico permite comparar os respectivos segmentos com a população.

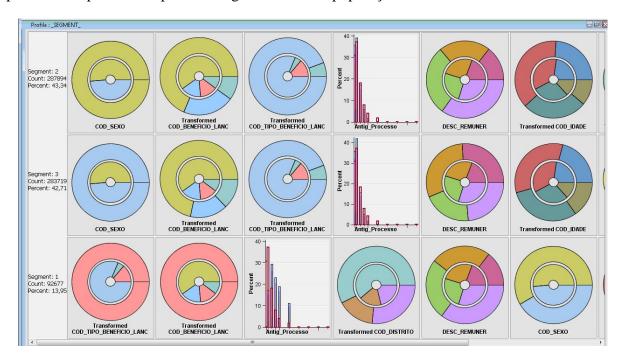


Figura 4 - Gráfico "Segment Profile"

É de realçar que a ordem pela qual os segmentos e as variáveis surgem no gráfico abaixo representado não é aleatória. Os segmentos (*clusters*) são ordenados pela proporção de indivíduos que têm e as variáveis estão ordenadas pela ordem de importância em cada segmento.



Analisando o gráfico atrás apresentado e tendo em conta as assimetrias existentes entre as características dos segmentos² e da população³ podemos referir que a solução encontrada é boa.

(4) Na fase 4, procedeu-se à implementação do modelo desenvolvido nó de "Score", que reflecte todas regras e formações de clusters. Se o modelo desenvolvido fosse corrido mensalmente numa nova base de beneficiários, permitiria obter para cada um dos beneficiários o respectivo segmento em que cada um daqueles se poderia inserir. De salientar que ao longo do tempo haverá migração entre segmentos dos beneficiários, ou seja, um dado beneficiário pode actualmente pertencer a um segmento e no próximo mês pertencer a outro. Assim, a comunicação a enviar aos beneficiários, deveria corresponder às características dos indivíduos caracterizados em cada segmento, e às estratégias determinadas para cada grupo.

² Nas variáveis categóricas (gráficos circulares) a parte de fora diz respeito à distribuição no segmento e a parte de dentro à distribuição da população.

³ Nas variáveis intervalares (histogramas) a parte a azul diz respeito à distribuição no segmento e a parte a vermelho à distribuição da população.



4. Resultados

Após o desenvolvimento do modelo de segmentação obtivemos 3 segmentos cujo comportamento é distinto entre si. A tabela 1 reflecte uma análise do comportamento inerente a cada um dos segmentos.

					Remi	ıneração			Tipo	Des_Dist.	Nacion.	
	Dimensão	Sexo	Idade	Mínima	Reduzida	Intermédia	Elevada	Ant_Proc	de Subs		Port	Est
Segmento 1 (Beneficiário Médio)	13,95%	НеF	Intermédia	25%	35%	25%	15%	Maior	SSDS	Porto 27% Lisboa 23% Outros 50%	89%	11%
Segmento 2 (Beneficiário Comum)	42,71%	F	Menor	29%	35%	22%	14%	Intermédia	PDI	Porto 21% Lisboa 17% Outros 62%	88%	12%
Segmento 3 (Beneficiário TOP)	43,34%	M	Superior	24%	20%	30%	26%	Menor	PDI	Porto 32% Lisboa 25% Outros 43%	85%	15%

Tabela 1 - Interpretação dos Segmentos obtidos

Analisemos, então, cada um dos segmentos com maior detalhe:

- O Segmento 1 representa 13,95% da população em estudo e é caracterizado pelas seguintes características:
 - Homogeneidade relativamente ao sexo, uma vez que a dimensão de Homens e Mulheres é similar;
 - A remuneração é maioritariamente intermédia (intervalo de 453 €a 600 € tem um peso de 35%);
 - o A antiguidade do processo é de longa duração;
 - O tipo de subsídio é classificado como sendo o subsídio social de desemprego subsequente;
 - A cidade do Porto é aquela que apresenta um maior peso (27%), seguindo-se a cidade de Lisboa (25%) e os restantes concelhos de Portugal com um somatório de 50%, predominando a nacionalidade Portuguesa em todos eles.



- O segmento 2 tem a dimensão de 42,71% dos beneficiários. Este segmento é caracterizado pelas seguintes características:
 - o O sexo feminino é aquele que sobressai;
 - A remuneração é maioritariamente reduzida (entre os 453 € a 600 € -35% dos casos).
 - O registo da antiguidade do processo é inferior ao segmento 1, considerando-se de menor duração;
 - O subsídio que apresenta maior evidência é a prestação de desemprego inicial;
 - A cidade do Porto representa 21% dos casos e Lisboa 17%. Os restantes concelhos de Portugal somam 62 %;
 - o 12% dos utentes tem nacionalidade estrangeira e os restantes 88% nacionalidade Portuguesa.
- O segmento 3 representa 43,34% dos beneficiários e caracteriza-se por:
 - o Maior predominância do sexo masculino;
 - A remuneração ser maioritariamente de classe intermédia (600€a 900€-30% dos casos);
 - o A antiguidade do processo é diminuta;
 - O subsídio, é maioritariamente classificado como sendo a prestação de desemprego inicial;
 - o Predomina a nacionalidade portuguesa.

Após a identificação e análise das características dos segmentos é fundamental a criação de oportunidades/campanhas a efectuar com os beneficiários com o intuito de reverter a situação. Assim, a classificação dos subsídios e o conhecimento das médias remuneratórias, existentes em cada segmento, dá a possibilidade de se projectar uma carta personalizada a cada utente. Nessa comunicação individual a Segurança Social poderá orientar o beneficiário para uma formação profissional próxima da sua realidade, ou sugerir um outro tipo de acção. Assim sendo, o objectivo será uma aproximação do beneficiário à Segurança Social, tendo este organismo a capacidade de mostrar e orientar como gerir o trabalho.



Deveriam ser iniciadas um maior número de acções de forma a incentivar os indivíduos nesta situação a procurarem novos postos de trabalho, numa perspectiva de inserção social, estabelecendo acções e estratégias específicas, potenciando uma relação mais eficaz entre o sistema de segurança social e o próprio beneficiário.

Vejamos alguns exemplos, por cada um dos segmentos, ou seja, para cada um dos perfis encontrados:

- No segmento 1 (designado por Beneficiário Médio) os beneficiários de remuneração de classe mínima atingem os 25%, os quais deveriam ser encaminhados para o sector terciário, a agricultura e pequenas industrias. Isto porque estes sectores voltaram a ser objecto de procura, além de que também poderá ser proposta a mudança de área funcional, aproveitando as ofertas nas respectivas áreas de residência. Já os beneficiários de remuneração de classe reduzida e intermédia, 35% e 25% respectivamente, deveriam ser encaminhados para cursos de formação profissional, deveriam frequentar estágios remunerados e incentivados a criar a sua pequena empresa (o chamado empresário em nome individual). Para os beneficiários de remuneração de classe elevada, seria proposto o recebimento do subsídio de uma só vez, com o objectivo do beneficiário poder criar o seu próprio negócio (média empresa).
- Relativamente ao segmento 2 (Beneficiário Comum) os beneficiários de remuneração de classe mínima atingem os 29%, os quais deveriam ser encaminhados para o sector terciário, ser-lhes ia dada a possibilidade de integração por exemplo na agricultura numa pequena quinta perto das suas residências, ou outros trabalhos semelhantes. Os beneficiários de remuneração de classe reduzida atingiram 35%. Estes deveriam ter a possibilidade de frequentar cursos de formação profissional, e obterem estágios remunerados, deveriam ser incentivados a criar a sua pequena empresa. (o chamado empresário em nome individual).
- Quanto ao segmento 3 (Beneficiário TOP) as acções a desenvolver deveriam ser as mesmas, tendo em conta o subsídio mensal encontrado.



A caracterização dos segmentos passou necessariamente pela identificação das variáveis mais importantes em cada segmento.

Uma vez que todos os segmentos apresentam beneficiários de remuneração elevada, dever-se-ia fazer um tipo de "comunicação individualizada" carta ou e-mail, visando a experiência em contexto de trabalho e referindo a criação de apoios ao desenvolvimento de iniciativas empresariais e à contratação de jovens. Com todas as acções acima descritas seriam criados mais postos de trabalho e consequentemente mais entradas positivas para a Segurança Social, assegurando com isso as pensões sociais e os respectivos subsídios aos mais carenciados ou idosos.

_

⁴ "Comunicação individualizada" entenda-se, a título de exemplo, carta com a maior referência possível a dados do beneficiário durante o texto e com assinatura, por exemplo do director responsável.



5. Conclusões e Trabalho Futuro

5.1. Conclusões

Assiste-se quase diariamente à discussão sobre a situação dos milhares de desempregados em Portugal, matéria cada vez mais preocupante se tivermos em atenção o momento que o País atravessa e que envolve não apenas números mas pessoas cujos perfis importa conhecer para que a sua qualidade de vida seja melhorada. É neste sentido que se desenvolve a presente investigação, uma vez que se desenvolveu um modelo de segmentação comportamental.

A análise dos resultados obtidos permite concluir que a distribuição de várias características, algumas das quais não contempladas neste trabalho, por dificuldades inerentes à recolha dos respectivos dados, altera-se entre características demográficas e laborais verificando-se o crescimento do desemprego registado na população masculina tornando a distribuição equitativa entre os géneros. A distribuição das características do desemprego registado não é, no entanto, similar ao longo do território nacional. De facto, observam-se diferenças significativas, relativamente a tal distribuição entre os vários distritos do país, tanto em termos demográficos como laborais. Pode observar-se, por exemplo, que os distritos de Porto e Lisboa são os que apresentam uma maior taxa de desemprego, sendo compreensível atendendo ao número de habitantes destas zonas.

A esta situação não serão alheias situações relacionadas com o desenvolvimento do sector da economia informal, os baixos níveis de produtividade das micro e pequenas empresas, a precarização e vulnerabilização das condições de emprego, as características sócio-demográficas da população, entre muitas outras situações.

É um território de mobilidade difusa e com fortes assimetrias no seu tecido social, tendo observado nos últimos anos, o impacto de fluxos migratórios de entrada, por parte de famílias sem laços de proximidade entre si e ao concelho.



5.2. Trabalho Futuro

Tendo em conta o trabalho desenvolvido existem alguns tópicos que fazem sentido rever e investigar num trabalho futuro. Sugere-se os seguintes aspectos:

- Avaliar o impacto da implementação deste modelo de segmentação.
- Melhorar o Modelo através da adição de variáveis, tais como: habilitações literárias, áreas de especialização, número de anos de educação, número de anos de trabalho, etc.;
- Acompanhar a divulgação das metas alcançadas pela Segurança Social.
- Acompanhar as acções a implementar pelos Centros de Emprego, recriando cursos de formação adequados ao perfil do desempregado.
- Criar de forma atempada a resposta à procura de emprego.
- Divulgação do site da Segurança Social, http://www.seg-social.pt.

Atendendo a que exerço funções num organismo do âmbito da segurança social, no caso de as próximas leis orgânicas a serem publicadas, não alterarem este facto, penso que se tornará fácil tentar, mesmo num pequeno grupo de indivíduos desempregados, aplicar o modelo de segmentação aqui proposto, acompanhando não só a divulgação das metas alcançadas pelo sector bem como as acções a implementar pelos Centros de Emprego.

Sabendo, por experiência própria, a dificuldade que o cidadão comum tem em consultar o actual site da segurança social, seria de toda a conveniência acções de divulgação junto dos beneficiários de modo a que pudessem obter todos os esclarecimentos possíveis sobre as variadas situações. Aguarda-se o próximo lançamento no novo portal da segurança social, bastante mais amigável que o actual.

Tenho consciência do grande desafio a que me proponho, que me permitirá posteriormente aprofundar conhecimentos na matéria.



6. Bibliografia

INE (2011). Estatísticas do Emprego 1.º Trim, Lisboa.

JISTEM - Journal of Information Systems and Technology Management Online version ISSN 1807-1775, URL acedido em 18-07-2011, às 16h

PESTANA, Maria H.; GAGEIRO, João N. (2000). *Análise de dados para Ciências Sociais: A complementaridade do SPSS*, 2ª edição, Lisboa, Edições Sílabo.

REIS, E. (2001). Estatística multivariada aplicada, 2ª edição, Lisboa, Edições Sílabo.

TRYON, R. C. (1939). Cluster Analysis, New York, McGraw-Hill,

VAREJÃO, José Manuel (2005). Emprego e Desemprego – Documento de Enquadramento Preliminar – NORTE 2015, Grupo de Prospectiva: As Pessoas, Atelier Temático: Emprego e Desemprego, (S. 1.), CCDR-N.

http://www.seg-social.pt, acedido em 16-03-2011, às 23h e em 19-09-2011, às 16h

http://www.dre.pt, acedido em 16-03-2011, às 10h

http://www.ine.pt, acedido em 18-07-2011, às 15h



I. Anexos



Anexo 1. Código utilizado no Pré -Tratamento dos Dados

```
data &EM_EXPORT_TRAIN;
```

set &EM_IMPORT_DATA;

IF COD_PAIS_NACIONALIDADE='@' Then COD_PAIS_NACIONALIDADE='DE';

IF VAL_REMUNERACAO_MEDIA_MENSAL<453 Then DESC_REMUNER='Minima';

IF VAL_REMUNERACAO_MEDIA_MENSAL>=453 and VAL_REMUNERACAO_MEDIA_MENSAL<600 Then DESC_REMUNER='Reduzida';

IF VAL_REMUNERACAO_MEDIA_MENSAL>=600 and VAL_REMUNERACAO_MEDIA_MENSAL<900 Then DESC_REMUNER='Intermedia';

IF VAL_REMUNERACAO_MEDIA_MENSAL>=900 Then DESC_REMUNER='Elevada';

run;



Anexo 2. Variáveis de Entrada

Tabela 2 - Variáveis de entrada

NISS_ENCRIPTADO	NISS do Beneficiário
DAT_ENTRADA	Data entrada do Processo
COD_ANO	Ano de lançamento do processo
COD_BENEFICIO_LANC	Código do benefício
COD_DISTRITO	Código do distrito de residência do beneficiário no final de cada ano de lançamento (cod_ano)
COD_IDADE	Idade do beneficiário no final de cada ano de lançamento (cod_ano)
COD_SEXO	Sexo do beneficiário
VAL_REMUNERACAO_MEDIA_MENSAL	Valor da remuneração de Referência média mensal.

Anexo 3. Matriz de Distâncias segundo a Segmentação Hierárquica

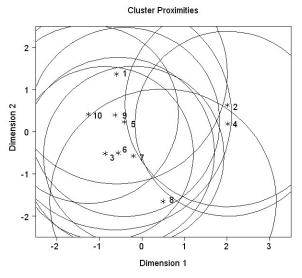


Figura 5 - Matriz de Distância segundo a Segmentação hierárquica



Anexo 4. Dimensão dos Segmentos obtidos

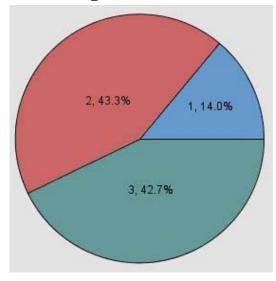


Figura 6 - Dimensão dos segmentos obtidos

Anexo 5. Semelhantes entre Segmentos

Tabela 3 - Semelhantes entre Segmentos

Segment Id	Nearest Cluster	Distance to Nearest Cluster
1		1.946985
2		3 1.433414
3		2 1.433414



Anexo 6. Variáveis Relevantes para a obtenção dos Segmentos

Tabela 4 - Variáveis relevantes para a obtenção dos Segmentos

NAME	LABEL	NRULES	NSURROGATES	IMPORTANCE
COD_SEXO		1	0	1.00000
TG_COD_BENEFICIO_LANC	Transformed COD_BENEFICIO_LANC	0	2	0.97276
DESC_REMUNER		0	1	0.77624
BIN_COD_IDADE	Transformed COD_IDADE	0	1	0.73402
TG_COD_PAIS_NACIONALIDADE	Transformed COD_PAIS_NACIONALIDADE	0	1	0.71872
TG_COD_DISTRITO	Transformed COD_DISTRITO	0	1	0.71696
TG_COD_TIPO_BENEFICIO_LANC	Transformed COD_TIPO_BENEFICIO_LANC	1	0	0.64697
Antig_Processo		0	1	0.60909

Anexo 7. Source Code do Projecto

*	*·
* EM SCORE CODE;	,
* VERSION: 6.2;	
* GENERATED BY: Acer-Acer;	
* CREATED: 18JUL2011:20:27:28;	
*	*•
*	, *;
* TOOL: Input Data Source;	
* TYPE: SAMPLE;	
* NODE: Ids;	
*	*•
*	
* TOOL: Transform;	,
* TYPE: MODIFY;	
* NODE: Trans;	
*	*•
<pre>aux_Antig_Processo = year(DAT_ENTRADA);</pre>	,
*	*;
* TOOL: Transform;	
* TYPE: MODIFY;	
* NODE: Trans2;	
*	*•
Antig_Processo = COD_ANO_LANC-aux_Antig	, _Processo;
	· ,
* TOOL: Metadata Node;	
Τ Ι Υ PH· I I I I I I V ·	



```
* NODE: Meta:
*_____*:
*-----*:
* TOOL: Filtering;
* TYPE: MODIFY;
* NODE: Filter;
*-----*:
*_____*.
* TOOL: SASHELP.EMCORE.EMCODETOOL.CLASS;
* TYPE: UTILITY:
* NODE: EMCODE;
*-----*:
*-----*:
* TOOL: Metadata Node;
* TYPE: UTILITY;
* NODE: Meta2;
*_____*:
*-----*:
* TOOL: Transform;
* TYPE: MODIFY;
* NODE: Trans3:
length BIN_COD_IDADE $10;
if (COD_IDADE ne.) and (COD_IDADE < 31) then
BIN COD IDADE = "01:low -31";
else
if (COD IDADE \geq 31 and COD IDADE < 44) then
BIN_COD_IDADE = "02:31-44";
if (COD IDADE \geq 44 and COD IDADE < 57) then
BIN_COD_IDADE = "03:44-57";
else
if (COD\_IDADE >= 57) then
BIN_COD_IDADE = "04:57-high";
length _DC_Format $200;
length TG_COD_BENEFICIO_LANC $%DMNORLEN;
_DC_Format = put(COD_BENEFICIO_LANC, $CHAR7.0);
%dmnormip(_DC_Format);
select(_DC_Format);
when('CRAD') TG COD BENEFICIO LANC = " OTHER ";
when('CRADSCR') TG_COD_BENEFICIO_LANC = "_OTHER_";
when('LOR') TG_COD_BENEFICIO_LANC = "_OTHER_";
when('LOS') TG_COD_BENEFICIO_LANC = "_OTHER_";
when('MU') TG_COD_BENEFICIO_LANC = "_OTHER_";
when('PO') TG_COD_BENEFICIO_LANC = "_OTHER_";
when('PSSD') TG_COD_BENEFICIO_LANC = "_OTHER_";
when('SDAD') TG_COD_BENEFICIO_LANC = "_OTHER_";
when('SDBO') TG_COD_BENEFICIO_LANC = "_OTHER_";
when('SDCO') TG COD BENEFICIO LANC = "OTHER";
```



```
when('SDDO') TG_COD_BENEFICIO_LANC = "_OTHER_";
when('SDMG') TG_COD_BENEFICIO_LANC = "_OTHER_";
when('SDMI') TG COD BENEFICIO LANC = " OTHER ";
when('SDNIAD') TG_COD_BENEFICIO_LANC = "_OTHER_";
when('SDP') TG_COD_BENEFICIO_LANC = "_OTHER_";
when('SDPI') TG_COD_BENEFICIO_LANC = "_OTHER_";
when('SDSA') TG_COD_BENEFICIO_LANC = "_OTHER_";
when('SDST') TG_COD_BENEFICIO_LANC = "_OTHER_";
when('SSDIDO') TG_COD_BENEFICIO_LANC = "_OTHER_
when('SSDIMG') TG COD BENEFICIO LANC = " OTHER ";
when('SSDIMI') TG_COD_BENEFICIO_LANC = "_OTHER_
when('SSDISA') TG_COD_BENEFICIO_LANC = "_OTHER_";
when('SSDSEQ') TG_COD_BENEFICIO_LANC = "_OTHER_";
when('SSDSI') TG COD BENEFICIO LANC = " OTHER ";
otherwise TG_COD_BENEFICIO_LANC = DC Format:
end;
drop DC Format;
length _DC_Format $200;
length TG COD DISTRITO $% DMNORLEN;
_DC_Format = put(COD_DISTRITO, $CHAR3.0);
%dmnormip(_DC_Format);
select(_DC_Format);
when('01') TG_COD_DISTRITO = "_OTHER_";
when('02') TG COD DISTRITO = " OTHER
when('03') TG_COD_DISTRITO = "_OTHER_";
when('04') TG_COD_DISTRITO = "_OTHER_
when('05') TG COD DISTRITO = " OTHER
when('06') TG COD DISTRITO = " OTHER
when('07') TG COD DISTRITO = " OTHER
when('08') TG COD DISTRITO = " OTHER
when('09') TG_COD_DISTRITO = "_OTHER_
when('10') TG COD DISTRITO = " OTHER
when('12') TG_COD_DISTRITO = "_OTHER_
when('14') TG COD DISTRITO = " OTHER
when('15') TG_COD_DISTRITO = "_OTHER_
when('16') TG COD DISTRITO = " OTHER
when('17') TG_COD_DISTRITO = "
                             OTHER
when('18') TG_COD_DISTRITO = " OTHER
when('31') TG COD DISTRITO = " OTHER
when('32') TG COD DISTRITO = "OTHER
when('41') TG_COD_DISTRITO = "_OTHER
when('42') TG COD DISTRITO = "OTHER
when('43') TG_COD_DISTRITO = "_OTHER_
when('44') TG COD DISTRITO = " OTHER
when('45') TG COD DISTRITO = " OTHER
when('46') TG_COD_DISTRITO = "_OTHER_
when('47') TG COD DISTRITO = " OTHER
when('48') TG_COD_DISTRITO = "_OTHER_";
when('49') TG COD DISTRITO = " OTHER ";
```



```
when('XXX') TG_COD_DISTRITO = "_OTHER_";
when('ZZZ') TG COD_DISTRITO = "_OTHER_";
otherwise TG_COD_DISTRITO = _DC_Format;
end;
drop DC Format;
length DC Format $200;
length TG_COD_PAIS_NACIONALIDADE $% DMNORLEN;
_DC_Format = put(COD_PAIS_NACIONALIDADE, $CHAR2.0);
%dmnormip(_DC_Format);
select( DC Format);
when('AD') TG_COD_PAIS_NACIONALIDADE = "_OTHER_";
when('AF') TG COD PAIS NACIONALIDADE = " OTHER ";
when('AI') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('AL') TG COD PAIS NACIONALIDADE = " OTHER "
when('AM') TG COD PAIS NACIONALIDADE = " OTHER
when('AO') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('AR') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('AT') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('AU') TG COD PAIS NACIONALIDADE = " OTHER
when('BA') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('BD') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('BE') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('BG') TG_COD_PAIS_NACIONALIDADE = "_OTHER_":
when('BI') TG COD PAIS NACIONALIDADE = " OTHER
when('BJ') TG COD PAIS NACIONALIDADE = " OTHER "
when('BM') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('BN') TG COD PAIS NACIONALIDADE = " OTHER
when('BO') TG COD PAIS NACIONALIDADE = " OTHER
when('BR') TG COD PAIS NACIONALIDADE = " OTHER
when('BS') TG_COD_PAIS_NACIONALIDADE = "_OTHER_"
when('BT') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('BW') TG COD PAIS NACIONALIDADE = " OTHER "
when('BY') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('BZ') TG COD PAIS NACIONALIDADE = " OTHER
when('CA') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('CD') TG COD PAIS NACIONALIDADE = "OTHER":
when('CG') TG COD PAIS NACIONALIDADE = " OTHER
when('CH') TG_COD_PAIS_NACIONALIDADE = "_OTHER_":
when('CI') TG_COD_PAIS_NACIONALIDADE = "_OTHER_"
when('CL') TG COD PAIS NACIONALIDADE = " OTHER "
when('CM') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('CN') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('CO') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('CS') TG COD PAIS NACIONALIDADE = " OTHER
when('CU') TG_COD_PAIS_NACIONALIDADE = "_OTHER_"
when('CV') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('CZ') TG COD PAIS NACIONALIDADE = " OTHER ";
when('DE') TG_COD_PAIS_NACIONALIDADE = "_OTHER_":
when('DK') TG COD PAIS NACIONALIDADE = " OTHER ";
```



```
when('DO') TG_COD_PAIS_NACIONALIDADE = "_OTHER_";
when('DZ') TG COD PAIS NACIONALIDADE = " OTHER
when('EC') TG COD PAIS NACIONALIDADE = " OTHER
when('EE') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('EG') TG COD PAIS NACIONALIDADE = " OTHER
when('ES') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('ET') TG COD PAIS NACIONALIDADE = " OTHER
when('FI') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('FO') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('FR') TG COD PAIS NACIONALIDADE = " OTHER
when('GA') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('GB') TG_COD_PAIS_NACIONALIDADE = "_OTHER_":
when('GE') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('GF') TG COD PAIS NACIONALIDADE = " OTHER
when('GH') TG COD PAIS NACIONALIDADE = " OTHER
when('GM') TG_COD_PAIS_NACIONALIDADE = "_OTHER_";
when('GN') TG COD PAIS NACIONALIDADE = " OTHER
when('GQ') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('GR') TG COD PAIS NACIONALIDADE = " OTHER
when('GT') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('GW') TG_COD_PAIS_NACIONALIDADE = "_OTHER_";
when('GY') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('HR') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('HU') TG COD PAIS NACIONALIDADE = " OTHER
when('ID') TG COD PAIS NACIONALIDADE = " OTHER "
when('IE') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('IL') TG COD PAIS NACIONALIDADE = " OTHER
when('IN') TG COD PAIS NACIONALIDADE = " OTHER
when('IQ') TG COD PAIS NACIONALIDADE = " OTHER
when('IR') TG COD PAIS NACIONALIDADE = " OTHER
when('IS') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('IT') TG COD PAIS NACIONALIDADE = " OTHER
when('JM') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('JP') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('KE') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('KG') TG COD PAIS NACIONALIDADE = " OTHER
when('KR') TG COD PAIS NACIONALIDADE = " OTHER
when('KZ') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('LB') TG COD PAIS NACIONALIDADE = " OTHER
when('LR') TG COD PAIS NACIONALIDADE = " OTHER
when('LT') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('LU') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('LV') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('MA') TG COD PAIS NACIONALIDADE = " OTHER
when('MD') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('MG') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('ML') TG COD PAIS NACIONALIDADE = "OTHER":
when('MN') TG_COD_PAIS_NACIONALIDADE = "_OTHER_";
when('MR') TG COD PAIS NACIONALIDADE = " OTHER ";
```



```
when('MT') TG_COD_PAIS_NACIONALIDADE = "_OTHER_";
when('MW') TG COD PAIS NACIONALIDADE = " OTHER ":
when('MX') TG COD PAIS NACIONALIDADE = " OTHER
when('MZ') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('NE') TG COD PAIS NACIONALIDADE = " OTHER
when('NG') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('NI') TG COD PAIS NACIONALIDADE = " OTHER
when('NL') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('NO') TG_COD_PAIS_NACIONALIDADE = "_OTHER]
when('NP') TG COD PAIS NACIONALIDADE = " OTHER
when('PA') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('PE') TG COD PAIS NACIONALIDADE = " OTHER
when('PH') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('PK') TG COD PAIS NACIONALIDADE = " OTHER
when('PL') TG COD PAIS NACIONALIDADE = " OTHER
when('PR') TG_COD_PAIS_NACIONALIDADE = "_OTHER_"
when('PW') TG COD PAIS NACIONALIDADE = " OTHER
when('RO') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('RS') TG COD PAIS NACIONALIDADE = " OTHER
when('RU') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('RW') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('SA') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('SE') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('SI') TG COD PAIS NACIONALIDADE = "
                                       OTHER
when('SK') TG_COD_PAIS_NACIONALIDADE = "_OTHER_";
when('SL') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('SN') TG COD PAIS NACIONALIDADE = " OTHER
when('ST') TG COD PAIS NACIONALIDADE = " OTHER
when('SV') TG COD PAIS NACIONALIDADE = " OTHER
when('SY') TG COD PAIS NACIONALIDADE = " OTHER
when('SZ') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('TD') TG COD PAIS NACIONALIDADE = " OTHER
when('TG') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when("TH') TG COD PAIS NACIONALIDADE = " OTHER
when('TL') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when("TN") TG COD PAIS NACIONALIDADE = " OTHER
when("TO") TG COD PAIS NACIONALIDADE = " OTHER
when('TR') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('TV') TG COD PAIS NACIONALIDADE = " OTHER
when("TW") TG COD PAIS NACIONALIDADE = " OTHER
when('TZ') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('UA') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('UG') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('US') TG COD PAIS NACIONALIDADE = " OTHER
when('UY') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('UZ') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('VC') TG COD PAIS NACIONALIDADE = "OTHER":
when('VE') TG_COD_PAIS_NACIONALIDADE = "_OTHER_
when('YU') TG COD PAIS NACIONALIDADE = " OTHER ";
```



```
when('ZA') TG_COD_PAIS_NACIONALIDADE = "_OTHER_";
when('ZM') TG_COD_PAIS_NACIONALIDADE = "_OTHER_";
when('ZW') TG_COD_PAIS_NACIONALIDADE = "_OTHER_";
otherwise TG_COD_PAIS_NACIONALIDADE = _DC_Format;
end;
drop _DC_Format;
length _DC_Format $200;
length TG_COD_TIPO_BENEFICIO_LANC $%DMNORLEN;
_DC_Format = put(COD_TIPO_BENEFICIO_LANC, $CHAR5.0);
%dmnormip(_DC_Format);
select(_DC_Format);
when('CR') TG_COD_TIPO_BENEFICIO_LANC = "_OTHER_";
when('LO') TG_COD_TIPO_BENEFICIO_LANC = "_OTHER_";
when('MU') TG COD TIPO BENEFICIO LANC = "OTHER";
when('PDNIA') TG COD TIPO BENEFICIO LANC = " OTHER ";
when('PO') TG_COD_TIPO_BENEFICIO_LANC = "_OTHER_";
when('PSSD') TG_COD_TIPO_BENEFICIO_LANC = "_OTHER_";
when('SDEQ') TG_COD_TIPO_BENEFICIO_LANC = "_OTHER_";
when('SDP') TG COD TIPO BENEFICIO LANC = " OTHER ";
otherwise TG_COD_TIPO_BENEFICIO_LANC = _DC_Format;
end;
drop _DC_Format;
*_____*:
* TOOL: Clustering;
* TYPE: EXPLORE:
* NODE: Clus;
*****************
*** Begin Scoring Code from PROC DMVQ ***;
****************
*** Begin Class Look-up, Standardization, Replacement;
drop \_dm\_bad; \_dm\_bad = 0;
*** Standardize Antig_Processo;
drop T Antig Processo;
if missing(Antig_Processo) then T_Antig_Processo = .;
else T_Antig_Processo = (Antig_Processo - 0) * 0.05882352941176;
*** Generate dummy variables for BIN_COD_IDADE;
drop BIN_COD_IDADE01_low__31 BIN_COD_IDADE02_31_44
BIN_COD_IDADE03_44_57
   BIN COD IDADE04 57 high;
*** encoding is sparse, initialize to zero;
BIN_COD_IDADE01_low_31 = 0;
BIN COD IDADE02 31 44 = 0;
BIN_COD_IDADE03_44_57 = 0;
BIN COD IDADE04 57 high = 0;
```



```
if missing(BIN_COD_IDADE) then do;
 BIN COD IDADE01 low 31 = ...
 BIN_COD_IDADE02_31_44 = .;
 BIN_COD_IDADE03_44_57 = .;
 BIN COD IDADE04 57 high = .;
end;
else do;
 length _dm10 $ 10; drop _dm10;
 %DMNORMCP(BIN_COD_IDADE, _dm10)
 if dm10 = '02:31-44' then do;
   BIN_COD_IDADE02_31_44 = 1;
 end;
 else if _{dm10} = '03:44-57' then do;
   BIN COD IDADE03 44 57 = 1;
 end;
 else if _{dm10} = '01:LOW - 31' then do;
   BIN COD IDADE01 low 31 = 1;
 end;
 else if _dm10 = '04:57-HIGH' then do;
   BIN_COD_IDADE04_57_high = 1;
 end;
 else do;
   BIN_COD_IDADE01_low__31 = .;
   BIN_COD_IDADE02_31_44 = .;
   BIN COD IDADE03 44 57 = ...
   BIN_COD_IDADE04_57_high = .;
   DM BAD = 1;
 end;
end;
*** Generate dummy variables for COD_SEXO;
drop COD SEXO COD SEXOF COD SEXOM;
*** encoding is sparse, initialize to zero;
COD\_SEXO\_ = 0;
COD\_SEXOF = 0;
COD SEXOM = 0;
if missing(COD_SEXO) then do;
 COD\_SEXO\_ = .;
 COD SEXOF = .;
 COD SEXOM = .;
end;
else do;
 length _dm1 $ 1; drop _dm1 ;
 dm1 = put( COD SEXO, $CHAR1.);
 %DMNORMIP(_dm1)
 if _{dm1} = 'F' then do;
   COD SEXOF = 1;
 end;
 else if _{dm1} = 'M' then do;
```



```
COD\_SEXOM = 1;
 end:
 else if _{dm1} = '@' then do;
   COD\_SEXO\_ = 1;
 end;
 else do:
   COD_SEXO_=:
   COD\_SEXOF = .;
  COD\_SEXOM = .;
   DM BAD = 1;
 end;
end;
*** Generate dummy variables for DESC REMUNER;
drop DESC_REMUNERElevad DESC_REMUNERInterm DESC_REMUNERMinima
    DESC_REMUNERReduzi;
*** encoding is sparse, initialize to zero;
DESC_REMUNERElevad = 0;
DESC_REMUNERInterm = 0;
DESC_REMUNERMinima = 0;
DESC_REMUNERReduzi = 0;
if missing( DESC_REMUNER ) then do;
 DESC_REMUNERElevad = .;
 DESC REMUNERInterm = .;
 DESC REMUNERMinima = .;
 DESC_REMUNERReduzi = .;
end;
else do:
 length dm6 $ 6; drop dm6;
 %DMNORMCP(DESC_REMUNER, _dm6)
 if _{dm6} = 'REDUZI' then do;
  DESC_REMUNERReduzi = 1;
 end;
 else if _{dm6} = 'INTERM' then do;
   DESC_REMUNERInterm = 1;
 end;
 else if _{dm6} = 'MINIMA' then do;
   DESC_REMUNERMinima = 1;
 end;
 else if dm6 = 'ELEVAD' then do;
   DESC_REMUNERElevad = 1;
 end;
 else do;
   DESC_REMUNERElevad = .;
   DESC_REMUNERInterm = .;
   DESC_REMUNERMinima = .;
   DESC REMUNERReduzi = .;
   _DM_BAD = 1;
 end;
```



TG COD DISTRITO11 = .;

end: *** Generate dummy variables for TG_COD_BENEFICIO_LANC; drop TG_COD_BENEFICIO_LANCSD TG_COD_BENEFICIO_LANCSSDI TG COD BENEFICIO LANCSSDS TG COD BENEFICIO LANC OTHER ; *** encoding is sparse, initialize to zero; TG COD BENEFICIO LANCSD = 0; TG_COD_BENEFICIO_LANCSSDI = 0; TG COD_BENEFICIO_LANCSSDS = 0; TG_COD_BENEFICIO_LANC_OTHER_ = 0; if missing(TG_COD_BENEFICIO_LANC) then do; TG_COD_BENEFICIO_LANCSD = .; TG_COD_BENEFICIO_LANCSSDI = .; TG COD BENEFICIO LANCSSDS = .; TG_COD_BENEFICIO_LANC_OTHER_ = .; end; else do: length _dm32 \$ 32; drop _dm32; %DMNORMCP(TG_COD_BENEFICIO_LANC, _dm32) if $_{dm32} = 'SD'$ then do; TG_COD_BENEFICIO_LANCSD = 1; end: else if $_{dm32} = 'SSDI'$ then do; TG COD BENEFICIO LANCSSDI = 1; else if $_{dm32} = 'SSDS'$ then do; TG COD BENEFICIO LANCSSDS = 1; end; else if dm32 = 'OTHER' then do; TG_COD_BENEFICIO_LANC_OTHER_ = 1; end; else do: TG_COD_BENEFICIO_LANCSD = .; TG COD BENEFICIO LANCSSDI = .; TG_COD_BENEFICIO_LANCSSDS = .; TG_COD_BENEFICIO_LANC_OTHER_ = .; DM BAD = 1; end; end; *** Generate dummy variables for TG_COD_DISTRITO; drop TG_COD_DISTRITO11 TG_COD_DISTRITO13 TG_COD_DISTRITO_OTHER_; *** encoding is sparse, initialize to zero; TG COD DISTRITO11 = 0; $TG_COD_DISTRITO13 = 0;$ TG COD DISTRITO OTHER = 0; if missing(TG_COD_DISTRITO) then do;



```
TG_COD_DISTRITO13 = .;
 TG COD DISTRITO OTHER = .;
else do:
 length dm32 $ 32; drop dm32;
 %DMNORMCP(TG_COD_DISTRITO, _dm32)
 if _{dm32} = '_{OTHER_{'}} then do;
  TG_COD_DISTRITO_OTHER_ = 1;
 end;
 else if dm32 = '13' then do;
   TG\_COD\_DISTRITO13 = 1;
 end;
 else if _{dm32} = '11' then do;
   TG COD DISTRITO11 = 1;
 end;
 else do;
  TG COD DISTRITO11 = .;
  TG_COD_DISTRITO13 = .;
  TG_COD_DISTRITO_OTHER_ = .;
   _DM_BAD = 1;
 end;
end;
*** Generate dummy variables for TG_COD_PAIS_NACIONALIDADE;
drop TG COD PAIS NACIONALIPT TG COD PAIS NACIONALI OTHER;
if missing(TG_COD_PAIS_NACIONALIDADE) then do;
 TG COD PAIS NACIONALIPT = .;
 TG_COD_PAIS_NACIONALI_OTHER_ = .;
end:
else do;
 length _dm32 $ 32; drop _dm32;
 %DMNORMCP(TG COD PAIS NACIONALIDADE, dm32)
 if _{dm32} = 'PT' then do;
   TG_COD_PAIS_NACIONALIPT = 1;
   TG_COD_PAIS_NACIONALI_OTHER_ = 0;
 end:
 else if _dm32 = '_OTHER_' then do;
  TG_COD_PAIS_NACIONALIPT = 0;
  TG_COD_PAIS_NACIONALI_OTHER_ = 1;
 end;
 else do:
  TG_COD_PAIS_NACIONALIPT = .;
  TG_COD_PAIS_NACIONALI_OTHER_ = .;
   DM BAD = 1;
 end;
end;
*** Generate dummy variables for TG_COD_TIPO_BENEFICIO_LANC;
drop TG_COD_TIPO_BENEFICIOPDI TG_COD_TIPO_BENEFICIOSSDS
```



```
TG_COD_TIPO_BENEFICIO_OTHER_;
*** encoding is sparse, initialize to zero;
TG COD TIPO BENEFICIOPDI = 0;
TG_COD_TIPO_BENEFICIOSSDS = 0;
TG COD TIPO BENEFICIO OTHER = 0;
if missing(TG_COD_TIPO_BENEFICIO_LANC) then do;
 TG_COD_TIPO_BENEFICIOPDI = .;
 TG_COD_TIPO_BENEFICIOSSDS = .;
 TG_COD_TIPO_BENEFICIO_OTHER_ = .;
end;
else do:
 length _dm32 $ 32; drop _dm32;
 %DMNORMCP(TG_COD_TIPO_BENEFICIO_LANC, _dm32)
 if dm32 = 'PDI' then do:
   TG COD TIPO BENEFICIOPDI = 1;
 end;
 else if dm32 = 'SSDS' then do;
   TG_COD_TIPO_BENEFICIOSSDS = 1;
 end;
 else if _{dm32} = '_{OTHER_{'}} then do;
   TG_COD_TIPO_BENEFICIO_OTHER_ = 1;
 end;
 else do;
   TG COD TIPO BENEFICIOPDI = .;
   TG COD TIPO BENEFICIOSSDS = .;
   TG_COD_TIPO_BENEFICIO_OTHER_ = .;
   DM BAD = 1;
 end;
end;
*** End Class Look-up, Standardization, Replacement;
*** Omitted Cases;
if dm bad then do;
 _SEGMENT_ = .; Distance = .;
 goto CLUSvlex;
end; *** omitted;
*** Compute Distances and Cluster Membership;
label _SEGMENT_ = 'Segment Id' ;
label Distance = 'Distance';
array CLUSvads [3] _temporary_;
drop _vqclus _vqmvar _vqnvar;
_vqmvar = 0;
do _vqclus = 1 to 3; CLUSvads [_vqclus] = 0; end;
if not missing( T_Antig_Processo ) then do;
 CLUSvads [1] + ( T_Antig_Processo - 0.16438877848373 )**2;
 CLUSvads [2] + ( T_Antig_Processo - 0.05935538366057 )**2;
```



```
CLUSvads [3] + ( T_Antig_Processo - 0.0563961069187 )**2;
end:
else vgmvar + 0.00514279329997;
if not missing(BIN_COD_IDADE01_low__31) then do;
 CLUSvads [1] + (BIN COD IDADE01 low 31 - 0.21759444090766)**2;
 CLUSvads [2] + (BIN_COD_IDADE01_low__31 - 0.23948397674143)**2;
 CLUSvads [3] + (BIN_COD_IDADE01_low__31 - 0.21094110722228)**2;
end;
else _vqmvar + 0.17395636080632;
if not missing(BIN_COD_IDADE02_31_44) then do;
 CLUSvads [1] + (BIN_COD_IDADE02_31_44 - 0.32604637612352)**2;
 CLUSvads [2] + (BIN_COD_IDADE02_31_44 - 0.37961194050588)**2;
 CLUSvads [3] + (BIN_COD_IDADE02_31_44 - 0.33299849498976)**2;
else vgmvar + 0.22816442987572;
if not missing(BIN_COD_IDADE03_44_57) then do;
 CLUSvads [1] + (BIN COD IDADE03 44 57 - 0.30592272084767)**2;
 CLUSvads [2] + (BIN_COD_IDADE03_44_57 - 0.27378132229223)**2;
 CLUSvads [3] + (BIN_COD_IDADE03_44_57 - 0.30362788533725)**2;
end;
else _vqmvar + 0.20632472821076;
if not missing(BIN_COD_IDADE04_57_high) then do;
 CLUSvads [1] + (BIN_COD_IDADE04_57_high - 0.15043646212113)**2;
 CLUSvads [2] + (BIN COD IDADE04 57 high - 0.10712276046044)**2;
 CLUSvads [3] + (BIN COD IDADE04 57 high - 0.15243251245069)**2;
end;
else vgmvar + 0.1149567298432;
if not missing (COD SEXO ) then do;
 CLUSvads [1] + (COD SEXO -0)**2;
 CLUSvads [2] + ( COD_SEXO_ - 3.4735006634386E-6 )**2;
 CLUSvads [3] + (COD_SEXO_ - 3.5246141428667E-6)**2;
end:
else _vqmvar + 3.0107287318222E-6;
if not missing( COD_SEXOF ) then do;
 CLUSvads [1] + ( COD_SEXOF - 0.58381259643708 )**2;
 CLUSvads [2] + (COD SEXOF - 0.99999652649933)**2;
 CLUSvads [3] + (COD SEXOF - 0)**2;
end;
else vgmvar + 0.24978033193228;
if not missing (COD SEXOM) then do;
 CLUSvads [1] + (COD_SEXOM - 0.41618740356291)**2;
 CLUSvads [2] + ( COD_SEXOM - 0 )**2;
 CLUSvads [3] + ( COD_SEXOM - 0.99999647538585 )**2;
end:
else vgmvar + 0.24978024260135;
if not missing( DESC_REMUNERElevad ) then do;
 CLUSvads [1] + ( DESC REMUNERElevad - 0.14427527865597 )**2;
 CLUSvads [2] + ( DESC_REMUNERElevad - 0.14456362411165 )**2;
 CLUSvads [3] + ( DESC REMUNERElevad - 0.26204801229385 )**2;
```



```
end;
else vgmvar + 0.1567928234628;
if not missing (DESC REMUNERInterm ) then do:
 CLUSvads [1] + ( DESC_REMUNERInterm - 0.25009441393225 )**2;
 CLUSvads [2] + ( DESC_REMUNERInterm - 0.21511042258609 )**2;
 CLUSvads [3] + ( DESC_REMUNERInterm - 0.29850309637352 )**2;
end;
else _vqmvar + 0.1902729557305;
if not missing( DESC_REMUNERMinima ) then do;
 CLUSvads [1] + ( DESC REMUNERMinima - 0.25256536141653 )**2;
 CLUSvads [2] + ( DESC_REMUNERMinima - 0.29057222449929 )**2;
 CLUSvads [3] + ( DESC_REMUNERMinima - 0.20408573271441 )**2;
end:
else vgmvar + 0.1866631469846;
if not missing( DESC_REMUNERReduzi ) then do;
 CLUSvads [1] + ( DESC_REMUNERReduzi - 0.35306494599523 )**2;
 CLUSvads [2] + ( DESC REMUNERReduzi - 0.34975372880296 )**2;
 CLUSvads [3] + ( DESC_REMUNERReduzi - 0.23536315861821 )**2;
end;
else _vqmvar + 0.21054220754897;
if not missing( TG_COD_BENEFICIO_LANCSD ) then do;
 CLUSvads [1] + (TG_COD_BENEFICIO_LANCSD - 0)**2;
 CLUSvads [2] + (TG_COD_BENEFICIO_LANCSD - 0.68667981965584)**2;
 CLUSvads [3] + (TG COD BENEFICIO LANCSD - 0.71270517660079)**2;
end:
else vgmvar + 0.23959746005233;
if not missing (TG COD BENEFICIO LANCSSDI) then do;
 CLUSvads [1] + (TG COD BENEFICIO LANCSSDI - 0)**2;
 CLUSvads [2] + (TG COD BENEFICIO LANCSSDI - 0.2120120599943)**2;
 CLUSvads [3] + (TG_COD_BENEFICIO_LANCSSDI - 0.15906231165343)**2;
end;
else vqmvar + 0.13427692190893;
if not missing( TG_COD_BENEFICIO_LANCSSDS ) then do;
 CLUSvads [1] + (TG_COD_BENEFICIO_LANCSSDS - 0.99983814754469)**2;
 CLUSvads [2] + (TG_COD_BENEFICIO_LANCSSDS - 0)**2;
 CLUSvads [3] + (TG_COD_BENEFICIO_LANCSSDS - 0)**2;
else _vqmvar + 0.12003292454003;
if not missing (TG COD BENEFICIO LANC OTHER ) then do;
 CLUSvads [1] + (TG COD BENEFICIO LANC OTHER - 0.0001618524553
)**2;
 CLUSvads [2] + (TG_COD_BENEFICIO_LANC_OTHER_ - 0.10130812034985
 CLUSvads [3] + (TG COD BENEFICIO LANC OTHER - 0.12823251174577
)**2;
end;
else vqmvar + 0.08895551641117;
if not missing (TG COD DISTRITO11) then do:
 CLUSvads [1] + (TG COD DISTRITO11 - 0.15562653085447)**2;
```



```
CLUSvads [2] + (TG_COD_DISTRITO11 - 0.17100043766108)**2;
 CLUSvads [3] + (TG COD DISTRITO11 - 0.18873603812222)**2;
else _vqmvar + 0.14530298099392;
if not missing (TG COD DISTRITO13) then do;
 CLUSvads [1] + (TG_COD_DISTRITO13 - 0.2676068495959)**2;
 CLUSvads [2] + (TG COD DISTRITO13 - 0.20226889063335)**2;
 CLUSvads [3] + (TG_COD_DISTRITO13 - 0.20436065261755)**2;
end;
else vgmvar + 0.16721617022987;
if not missing (TG_COD_DISTRITO_OTHER_) then do;
 CLUSvads [1] + (TG_COD_DISTRITO_OTHER_ - 0.57676661954961)**2;
 CLUSvads [2] + (TG_COD_DISTRITO_OTHER_ - 0.62673067170555)**2;
 CLUSvads [3] + (TG COD DISTRITO OTHER - 0.60690330926021)**2;
end;
else _vqmvar + 0.23761450294215;
if not missing (TG COD PAIS NACIONALIPT) then do;
 CLUSvads [1] + (TG_COD_PAIS_NACIONALIPT - 0.88743701241947)**2;
 CLUSvads [2] + (TG COD PAIS NACIONALIPT - 0.87935142795612)**2;
 CLUSvads [3] + (TG_COD_PAIS_NACIONALIPT - 0.84394770882457)**2;
end:
else vgmvar + 0.11651333963454;
if not missing( TG_COD_PAIS_NACIONALI_OTHER_ ) then do;
 CLUSvads [1] + (TG COD PAIS NACIONALI OTHER - 0.11256298758052
)**2:
 CLUSvads [2] + ( TG_COD_PAIS_NACIONALI_OTHER_ - 0.12064857204387
 CLUSvads [3] + (TG COD PAIS NACIONALI OTHER - 0.15605229117542
)**2:
end;
else _vqmvar + 0.11651333963454;
if not missing (TG COD TIPO BENEFICIOPDI) then do;
 CLUSvads [1] + (TG_COD_TIPO_BENEFICIOPDI - 0)**2;
 CLUSvads [2] + (TG COD TIPO BENEFICIOPDI - 0.9420654824345)**2;
 CLUSvads [3] + (TG_COD_TIPO_BENEFICIOPDI - 0.94099091002012)**2;
end:
else _vqmvar + 0.15379090780939;
if not missing( TG_COD_TIPO_BENEFICIOSSDS ) then do;
 CLUSvads [1] + (TG COD TIPO BENEFICIOSSDS - 1)**2;
 CLUSvads [2] + (TG COD TIPO BENEFICIOSSDS - 0)**2;
 CLUSvads [3] + (TG_COD_TIPO_BENEFICIOSSDS - 0)**2;
end:
else _vqmvar + 0.12004920503362;
if not missing( TG COD TIPO BENEFICIO OTHER ) then do;
 CLUSvads [1] + (TG_COD_TIPO_BENEFICIO_OTHER_ - 0)**2;
 CLUSvads [2] + (TG_COD_TIPO_BENEFICIO_OTHER_ - 0.05793451756549
 CLUSvads [3] + (TG_COD_TIPO_BENEFICIO_OTHER_ - 0.05900908997987
)**2:
```



```
end;
else _vqmvar + 0.04777974768177;
_{\text{vqnvar}} = 3.66002277789758 - _{\text{vqmvar}};
if _vqnvar <= 9.9863139745982E-12 then do;
 SEGMENT = .; Distance = .;
end;
else do;
 _SEGMENT_ = 1; Distance = CLUSvads [1];
 _vqfzdst = Distance * 0.9999999999988; drop _vqfzdst;
 do _{\text{vqclus}} = 2 \text{ to } 3;
  if CLUSvads [_vqclus] < _vqfzdst then do;
    _SEGMENT_ = _vqclus; Distance = CLUSvads [_vqclus];
    _vqfzdst = Distance * 0.9999999999988;
  end:
 end;
 Distance = sqrt(Distance * (3.66002277789758 / _vqnvar));
CLUSvlex:;
*****************
*** End Scoring Code from PROC DMVQ ***;
************
*-----*;
* Clus: Creating Segment Label;
*_____*:
length _SEGMENT_LABEL_ $80;
label _SEGMENT_LABEL_='Segment Description';
if _SEGMENT_ = 1 then _SEGMENT_LABEL_="Cluster1";
if _SEGMENT_ = 2 then _SEGMENT_LABEL_="Cluster2";
else
if SEGMENT = 3 then SEGMENT LABEL ="Cluster3";
*-----*:
* TOOL: Score Node;
* TYPE: ASSESS;
* NODE: Score:
*_____*:
*-----*:
* Score: Creating Fixed Names;
*-----*:
LABEL EM_SEGMENT = 'Segment Variable';
EM_SEGMENT = _SEGMENT_;
```



Glossário

Beneficiário: Pessoa inscrita como titular do direito a protecção social no âmbito dos Regimes da Segurança Social, contributivos e não contributivos. (Lei n.º 28/84, de 14/8).

Prestações de desemprego: Pessoa inscrita como titular do direito a protecção social no âmbito dos Regimes da Segurança Social, contributivos e não contributivos. (Lei n.º 28/84, de 14/8).

Desemprego: Situação decorrente da inexistência total e involuntária de emprego do beneficiário com capacidade e disponibilidade para o trabalho, inscrito para emprego no centro de emprego. O requisito de inexistência total de emprego considera-se ainda preenchido nas situações em que, cumulativamente com o trabalho por conta de outrem, cujo contrato de trabalho cessou, o beneficiário exerce uma actividade independente cujos rendimentos não ultrapassem mensalmente 50% do SMN.

Subsídio de desemprego: Prestação pecuniária concedida aos trabalhadores que à data do desemprego reúnam as seguintes condições gerais: tenham trabalhado 540 dias por conta de outrem com o correspondente registo de remunerações nos 24 meses imediatamente anteriores à ocorrência da eventualidade (prazo de garantia); estejam em situação de desemprego involuntário; tenham capacidade e disponibilidade para o trabalho; estejam inscritos no centro de emprego da área de residência. O decreto-lei n.º 84/2003 de 24 de Abril reduz o prazo de garantia para a atribuição do subsídio de desemprego para 270 dias de trabalho por conta de outrem, com o correspondente registo de remunerações no período de 12 meses imediatamente anterior à data do desemprego.

Subsídio de desemprego parcial: É um valor em dinheiro que é pago a quem (1) seja requerente do subsídio de desemprego e à data em que cessou o contrato de trabalho, que determina a concessão do subsídio de desemprego, tinha outro emprego por conta de outrem a tempo parcial ou exercia uma actividade independente e desde que a



retribuição do trabalho por conta de outrem ou o rendimento relevante da actividade independente (70% no caso de profissionais livres e 20% no caso de empresários em nome individual) seja inferior ao valor do subsídio de desemprego (2) esteja a receber subsídio de desemprego e comece a trabalhar como trabalhador por conta de outrem (TCO) a tempo parcial ou como independente (TI) e desde que a retribuição do trabalho por conta de outrem ou o rendimento relevante da actividade independente (70% no caso de profissionais livres e 20% no caso de empresários em nome individual) seja inferior ao valor do subsídio de desemprego.

Subsídio social de desemprego inicial: Prestação pecuniária concedida aos trabalhadores que à data do desemprego não preencham o prazo de garantia para atribuição do subsídio de desemprego e reúnam as seguintes condições gerais: tenham trabalhado 180 dias por conta de outrem com o correspondente registo de remunerações nos 12 meses imediatamente anteriores à ocorrência da eventualidade (prazo de garantia); os rendimentos mensais, per capita, do agregado familiar do beneficiário não sejam superiores a 80% do SMN (condição de recursos); estejam em situação de desemprego involuntário; tenham capacidade e disponibilidade para o trabalho; estejam inscritos no centro de emprego da área de residência.

Subsídio social de desemprego subsequente: Prestação pecuniária concedida aos trabalhadores que tenham esgotado os prazos de concessão do subsídio de desemprego e preencham a condição de recursos exigida para atribuição do subsídio social de desemprego inicial.