

C&S SIG

GEOPROCESSAMENTO PARA SUPORTE A GEOMARKETING

Mapeamento de Informação Censitária ao Código Postal

Margarida da Costa Corado

Relatório de Estágio apresentado como requisito parcial
para obtenção do grau de Mestre em Ciência e Sistemas
de Informação Geográfica

C& SIG



UNIGIS PT



C&S SIG

GEOPROCESSAMENTO PARA SUPORTE A GEOMARKETING

Mapeamento de Informação Censitária ao Código Postal

Margarida da Costa Corado

Relatório de Estágio apresentado como requisito parcial
para obtenção do grau de Mestre em Ciência e Sistemas
de Informação Geográfica

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

GEOPROCESSAMENTO PARA SUPORTE A GEOMARKETING
MAPEAMENTO DE INFORMAÇÃO CENSITÁRIA AO CÓDIGO POSTAL

por

Margarida da Costa Corado

Relatório de Estágio apresentado como requisito parcial para a obtenção do grau
de Mestre em Ciência e Sistemas de Informação Geográfica

Orientador: Professor Doutor Rui Pedro Julião

Coorientador: Eng^o. José Pedro Rufino

Julho 2017

DECLARAÇÃO DE ORIGINALIDADE

Declaro que o trabalho contido neste documento é da minha autoria e não de outra pessoa. Toda a assistência recebida de outras pessoas está devidamente assinalada e é efetuada referência a todas as fontes utilizadas (publicadas ou não).

O trabalho não foi anteriormente submetido ou avaliado na NOVA Information Management School ou em qualquer outra instituição.

Lisboa, 3 de julho de 2017

(Assinatura)

AGRADECIMENTOS

Um agradecimento especial ao Professor Doutor Rui Pedro Julião e Engenheiro José Pedro Rufino por me terem proporcionado a oportunidade única de estagiar nesta empresa e colaborar num projeto tão interessante e desafiante. Todo o apoio, confiança, rigor e conselhos foram imprescindíveis para o sucesso deste projeto.

A todos os colegas dos CTT que de alguma forma me ajudaram, obrigada pelo companheirismo e amizade.

Família e amigos, obrigada!

GEOPROCESSAMENTO PARA SUPORTE A GEOMARKETING

MAPEAMENTO DE INFORMAÇÃO CENSITÁRIA AO CÓDIGO POSTAL

RESUMO

O sucesso de um negócio assenta em três pontos essenciais: localização, localização e localização (Jones, 1990). A partir desta e de modelos do século XX surgem as ideias base do conceito de *geomarketing*, a conjugação de instrumentos cartográficos e estatística espacial para valorizar e capacitar a análise socioeconómica dos mercados.

Através do projeto Geo10, o qual georreferencia e caracteriza todas as ruas, portas e domicílios de Portugal, e da informação censitária disponibilizada pelo INE, foi lançado o desafio de mapear os Censos ao código postal de 7 dígitos. Para tal foi necessário ultrapassar os constrangimentos inerentes à utilização conjunta de bases de dados de diferentes fontes, bem como informação geográfica com diferentes geometrias.

Através da integração da informação geográfica e alfanumérica do INE e códigos postais dos CTT foi possível a caracterização sociodemográfica e habitacional do código postal, permitindo alcançar resultados mais precisos nas áreas de *geomarketing* e soluções geográficas.

PALAVRAS-CHAVE

Geomarketing, Geo10, Censos, Código Postal, Geometria de dados

GEOPROCESSING AS SUPPORT FOR GEOMARKETING

MAPPING CENSUS DATA TO POSTAL CODE

ABSTRACT

The success to a business is based on three essentials: location, location, location (Jones, 1990). Based on this and models from the 20th century, the concept of geomarketing takes form as cartographic tools and spatial statistics able to value and capacitate the market socioeconomic analysis.

Through Geo10 project, which aims to georeference and characterize every Portuguese street, door and household, and census information made available by INE (Statistics Portugal), the challenge of mapping Census to postal code was launched. In order to succeed, the constraints related to the use of databases with different origins, as well as geographic information with different geometries, had to be dealt with.

Through the integration of geographic and alphanumeric information from INE and postal codes from CTT it was possible to do a sociodemographic and habitation characterization of the postal code, allowing more precise results in the geomarketing and geographic solutions domain.

KEYWORDS

Geomarketing, Geo10, Census, Postal Code, Data geometry

LISTA DE ACRÓNIMOS E SIGLAS

BGRI	Base Geográfica de Referenciação de Informação
BNEG	Base Nacional de Endereços dos CTT
CAE	Classificação de Atividades Económicas
CAE	Correção Assistida de Endereços
CDP	Centro de Distribuição Postal
Censos	Recenseamento da População e Habitação
CNE	Correio Não Endereçado
CP4	Código Postal de 4 dígitos
CP7	Código Postal de 7 dígitos
CTT	CTT – Correios de Portugal, S.A.
DOL	Distribuição OnLine
DPI	Soluções de Desmaterialização, Proximidade e <i>Business Intelligence</i>
INE	Instituto Nacional de Estatística
NUT	Nomenclatura de Unidade Territorial Estatística
OesteCIM	Comunidade Intermunicipal do Oeste
PD	Pontos de Distribuição
PE	Pontos de Entrega
PER	Pontos de Entrega Residenciais
QL	Quociente de Localização
S.A.	Sociedade Anónima
SIG	Sistemas de Informação Geográfica
TAE	Tratamento Automático de Endereços

ÍNDICE DO TEXTO

AGRADECIMENTOS	iv
RESUMO	v
ABSTRACT.....	vi
LISTA DE ACRÓNIMOS E SIGLAS.....	vii
ÍNDICE DE TABELAS.....	x
ÍNDICE DE FIGURAS.....	xi
1. INTRODUÇÃO.....	1
2. CTT – CORREIOS DE PORTUGAL, S.A.	3
2.1. Enquadramento	3
2.2. Missão, Visão e Valores.....	4
2.3. Contextualização	6
3. ÂMBITO DO ESTÁGIO	10
3.1. Definição do Problema e Objetivos	10
3.2. Pertinência e Relevância	11
3.3. Metodologia Adotada	15
3.3.1. Área de Estudo.....	17
4. GEOMARKETING DOS CTT	18
4.1. Dados de Suporte	18
4.1.1. Dados de Suporte Internos	18
4.1.1.1. DOL – Distribuição OnLine	18
4.1.1.2. Geoindex.....	19
4.1.1.3. BNE – Base Nacional de Endereços	20
4.1.1.4. Geo10.....	20
4.1.1.5. O Código Postal.....	22
4.1.2. Dados de Suporte Externos	23
4.1.2.1. Censos	23
4.1.2.2. Ignios.....	25
4.2. Modelo de Geoprocessamento Atual	26

4.2.1. Apresentação do Modelo Atual	26
4.2.2. Aplicação do Modelo – Estudo Geocontacto	30
4.2.3. Análise de Resultados.....	33
4.3. Modelo de Geoprocessamento Proposto.....	35
4.3.1. Tratamento de Dados.....	35
4.3.2. Apresentação do Modelo Proposto	38
4.3.3. Aplicação do Modelo – Estudo Geocontacto	42
4.3.4. Análise de Resultados.....	43
4.4. Comparação dos Modelos.....	45
4.5. Breves Conclusões.....	49
5. CONCLUSÕES	52
6. REFERÊNCIAS BIBLIOGRÁFICAS.....	55

ÍNDICE DE TABELAS

Tabela 1 - Probabilidade em atingir público-alvo em cada modelo	49
---	----

ÍNDICE DE FIGURAS

Figura 1 - Área de Estudo: OesteCIM.....	17
Figura 2 - Geoindex: Centroides de CP7 e Áreas de Influência	19
Figura 3 - Geo10: Centroides de CP7 e Ponto de Distribuição.....	21
Figura 4 - Divisão do território nacional para definição de primeiro algarismo de CP4.....	22
Figura 5 - Dados Geoindex: Relação CP7 e Subsecção Estatística, através de Centroides e Áreas de Influência	27
Figura 6 - Modelo Atual: <i>query</i> , em <i>software</i> Access, para estudo Geocontacto	32
Figura 7 - Modelo Atual: Exemplo de resultado de estudo Geocontacto	33
Figura 8 - <i>Query</i> inserida na ferramenta <i>select by attributes</i> , em <i>software</i> ArcGIS, para tratamento de dados.....	36
Figura 9 - Modelo, construído em <i>Model Builder</i> , <i>software</i> ArcGIS, para tratamento de dados.....	37
Figura 10 - Formulário para utilização do modelo anterior	37
Figura 11 - Dados Geo10: Relação CP7 e Subsecção Estatística, através de Ponto de Distribuição.....	38
Figura 12 - Modelo Proposto: <i>query</i> , em <i>software</i> Access, para estudo Geocontacto	43
Figura 13 - Modelo Proposto: Exemplo de resultado de estudo Geocontacto	44
Figura 14 - Comparação de modelos através de representação de relação de Geo10 e Geoindex com Subsecções Estatísticas.....	46
Figura 15 - Comparação de modelos – CP7 mantidos, eliminados e acrescentados pelo Modelo Proposto.....	47
Figura 16 - Comparação de modelos – CP7 mantidos e acrescentados pelo Modelo Proposto.....	48

1. INTRODUÇÃO

O presente relatório de estágio é apresentado como requisito parcial para obtenção do grau de Mestre em Ciência e Sistemas de Informação Geográfica, pela NOVA IMS (*Information Management School*), da Universidade Nova da Lisboa.

O estágio desenvolvido teve a duração de 800 horas e decorreu na empresa CTT – Correios de Portugal, S.A., na área comercial de Soluções de Desmaterialização, Proximidade e *Business Intelligence*, a qual integra a Direção de Publicidade e Soluções Empresariais.

O estágio permitiu o contacto com o mercado de trabalho, neste caso uma empresa de renome em Portugal, adquirir *soft skills* que apenas a prática num meio empresarial permite e conhecer de perto as soluções oferecidas pelos CTT, que vão muito além da distribuição postal.

O relatório, que visa descrever a principal atividade desenvolvida no decorrer do estágio, encontra-se estruturado em 4 capítulos distintos.

Primeiramente são abordados os CTT, apresentando resumidamente a sua história de quase 500 anos, citando a sua missão, visão e valores e contextualizando o âmbito do estágio na área comercial a que fui alocada.

De seguida é explicitado o âmbito do estágio, onde se define o problema e objetivos a alcançar, a pertinência e relevância dos trabalhos desenvolvidos para o estado da arte, bem como a metodologia adotada.

No capítulo seguinte é apresentado o *geomarketing* atual dos CTT, os dados de suporte utilizados bem como os postos à disposição para este fim, o modelo de geoprocessamento atualmente utilizado e o modelo de geoprocessamento proposto. No final deste capítulo são analisados os dois modelos, comparados e avaliadas as melhorias alcançadas.

O último capítulo visa resumir os trabalhos desenvolvidos, os objetivos alcançados, os constrangimentos sentidos e apresenta as perspectivas futuras do *geomarketing* e soluções geográficas dos CTT.

Através do estágio curricular e redação deste relatório foi possível perceber a importância do *geomarketing* nos mercados atuais, o seu potencial evolutivo e de que forma os sistemas de informação geográfica (SIG) permitem o processamento, manipulação e análise de dados, tanto alfanuméricos como geográficos, potenciando a sua integração e harmonização para fins conjuntos.

2. CTT – CORREIOS DE PORTUGAL, S.A.

2.1. Enquadramento

Os CTT - Correios de Portugal, S.A.¹ percorrem já quase 500 anos de história tendo sempre presente o objetivo de aproximar os portugueses e o mundo.

Os primeiros serviços de correio portugueses são fruto da necessidade do Rei e nobreza em comunicar com regularidade os atos políticos, diplomáticos e económicos (FPC, 2012). Com os Descobrimentos, a necessidade da Coroa, nobreza e burguesia contactarem outros Estados e mercadores facilitou o acesso ao serviço de correio, potenciando a sua democratização, passando este a ser acessível a qualquer cidadão, mediante o pagamento de uma quantia de subscrição do serviço (FPC, 2012).

Até 1800, os destinatários de correspondência eram divulgados numa lista pública e teriam de levantar o seu correio em local designado. Nesta data é aprovado o diploma para distribuição domiciliária de correio em Lisboa, a qual viu o primeiro protótipo de cadastro onde foram criados dezassete distritos postais, com identificação de ruas e números de casa. A distribuição de correspondência passou assim a ser constituída pela possibilidade de distribuição porta a porta, dependente da subscrição do serviço, e pelas caixas postais públicas, que dariam origem aos marcos de correio, apenas implementados em 1882 (FPC, 2012). Por dificuldades relacionadas com o processo de cadastro, a distribuição domiciliária só viu o seu pleno funcionamento em 1821 (FPC, 2012).

Até 1853 o pagamento da correspondência estaria a cargo do destinatário, situação invertida nesse ano com a introdução da obrigatoriedade do selo postal adesivo, tendo o primeiro a efígie da Rainha D. Maria II (FPC, 2012).

¹ Doravante denominado apenas como CTT.

A evolução dos modos de transporte, tecnologia, informatização e necessidades dos cidadãos foram sempre sendo acompanhadas pelos CTT, que procuraram estar um passo à frente e foram consolidando a sua organização e estrutura. Uma das medidas organizacionais mais importantes no setor foi a implementação do código postal de 4 dígitos em 1978, e a posterior introdução do código postal de 7 dígitos em 1998 (CTT - Correios de Portugal, S.A., 2012). O código postal facilitou a identificação dos concelhos dos destinatários, permitindo ultrapassar problemas de endereçamento e toponímia, tendo tido como *slogan* “Código Postal, Meio Caminho Andado” (Brito, 2013).

Em 1992, a até então CTT – Correios e Telecomunicações de Portugal divide a componente correio e telecomunicações e passa a sociedade anónima de capitais exclusivamente públicos com a designação atual CTT – Correios de Portugal S.A.. É nesta década que a informatização e mecanização se estendem a toda a rede e na década seguinte é notável a aposta em serviços digitais.

Em 2013 a empresa deu início à primeira fase de privatização, tendo esta sido concluído em setembro de 2014.

2.2. Missão, Visão e Valores

Os CTT são atualmente uma instituição moderna que sempre se adaptou à sociedade para responder com eficácia às exigências do século XXI. Hoje, a empresa rege-se pela seguinte missão, visão e valores.

Segundo a sua página institucional (CTT - Correios de Portugal, S.A., 2017), o Grupo CTT, com presença e impacto tanto no território nacional como em mercados internacionais, tem por missão assegurar soluções de comunicação e logística, bem como produtos financeiros, de proximidade e de excelência e estabelecer relações de confiança e inovação, traduzidas em:

- Vocação para ligar pessoas e empresas com comunicações físicas e eletrónicas;
- Soluções para dar resposta integrada às necessidades de cada cliente;
- A confiança como ativo da companhia e também como resultado;
- Estratégia empresarial e perfil de risco sustentáveis;
- A inovação como garante que se cumpre a missão hoje e no futuro.

A sua visão assenta no desejo de ser reconhecido como operador multisserviços com vocação postal e financeira, de referência mundial em qualidade, eficiência e criação de valor, caracterizado por uma:

- Operação postal de referência no mercado ibérico;
- Plataforma operacional e de distribuição líder;
- Oferta abrangente de serviços financeiros;
- Capacidade para competir internacionalmente em geografias selecionadas;
- *Performance* financeira no topo dos maiores operadores postais europeus;
- Gestão sã e prudente face ao modelo de negócio e perfil de risco;
- Responsabilidade social e ambiental.

O Grupo CTT rege-se pelos seguintes cinco valores:

- Orientação para o cliente – *o seu sucesso é o nosso sucesso*;
- Entusiasmo – *todos somos um*;
- Confiança – *cumprir sempre*;
- Excelência – *fazer sempre melhor*;
- Inovação – *criar futuro*.

2.3. Contextualização

No âmbito do estágio fui integrada na Direção de Publicidade e Soluções Empresariais, equipa de Soluções de Desmaterialização, Proximidade e *Business Intelligence* (DPI).

Apercebendo-se que a inovação é um fator de diferenciação competitiva, fonte de crescimento das receitas e contribuição para uma maior eficiência operacional e organizacional, as Soluções Empresariais apresentam-se como o elo de ligação dos CTT com as empresas, tanto no domínio físico como digital, no século XXI. Além da oferta de serviços, os CTT oferecem soluções integradas, personalizadas e de confiança, desenvolvidas para responder às necessidades concretas das empresas, com principal foco na eficiência e otimização de processos e recursos (CTT - Correios de Portugal, S.A., 2015). Através destas soluções, à rede única de proximidade às populações e capacidade de distribuição, é adicionada a oferta de soluções geográficas, comunicação digital, produção documental e digitalização de documentos, que permitem acrescentar valor aos negócios e empresas (CTT - Correios de Portugal, S.A., 2015).

As Soluções Empresariais dos CTT dividem-se em seis áreas de competência: Soluções de Comunicação, Produção Documental, Desmaterialização de Processos, Soluções Geográficas, Proximidade, Conveniência (CTT - Correios de Portugal, S.A., 2015).

A área de negócio em que estagiei - Soluções de Desmaterialização, Proximidade e *Business Intelligence* (DPI) - tem como responsabilidade criar, implementar e promover a oferta de comunicação digital, soluções geográficas e soluções de desmaterialização e proximidade que tirem partido das competências do Grupo CTT de forma global. Apesar de ter trabalhado apenas com questões relacionadas com as Soluções Geográficas, apresento resumidamente cada oferta do DPI.

As Soluções de Desmaterialização permitem a conversão de correio físico em correio digital, com soluções de tratamento de correio recebido e disponibilização

de informação sobre o negócio ao cliente, em formato digital e integrável com os seus sistemas de informação (CTT - Correios de Portugal, S.A., 2015).

As Soluções de Proximidade assentam no elevado e aprofundado conhecimento dos CTT sobre o território português e população, resultando numa relação de proximidade e confiança inigualável construída ao longo da história da empresa. Através das redes de distribuição e atendimento são disponibilizados serviços como a leitura de contadores de água e realização de questionários no terreno, bem como reconhecimento e levantamento de eventos no terreno.

As Soluções Geográficas assentam, mais uma vez, na vantagem competitiva dos CTT de conhecimento do território e permitem a criação de soluções concertadas com base neste conhecimento e a exploração geográfica de oportunidades de negócio. Através de um melhor conhecimento geográfico do mercado, apresentam-se como um suporte à decisão nos domínios empresarial, operacional, comercial e de *marketing* das empresas. Este largo espectro de ação leva a que estas soluções sejam parte integrante do domínio de *business intelligence* desta área de negócio. De entre as Soluções Geográficas há a destacar o Geoaddress, Geoindex – Georouting e Geomarketing - e Geocontacto.

Geoaddress é uma solução que visa a purificação das bases de dados de moradas de clientes, através da atualização, validação e normalização dos endereços postais, tendo como referência a Base Nacional de Endereços dos CTT (BNEG). Esta purificação é possível através do cruzamento de informação toponímica e postal das bases de dados de endereços dos clientes com informação qualificada propriedade dos CTT (CTT - Correios de Portugal, S.A., 2015). Esta solução tem dois serviços disponíveis, o Tratamento Automático de Endereços (TAE) e a Correção Assistida de Endereços (CAE). O TAE consiste num processo *batch* de tratamento da informação de endereçamento e suportado por procedimentos automáticos de análise postal (CTT - Correios de Portugal, S.A., 2015). O CAE consiste num processo manual de tratamento e validação, recorrendo a técnicos especializados (CTT - Correios de Portugal, S.A., 2015).

A solução Geindex pode ser considerada uma solução integrada, composta pelas soluções de Geomarketing e Georouting, dado serem muitas vezes desenvolvidas em conjunto para dar resposta ao cliente. O Geindex é assim um serviço de informação geográfica que permite um melhor conhecimento da área geográfica em que o cliente se move e o capacita com capacidades para identificar oportunidade e ameaças de negócio. Estas soluções assentam em recursos especializados, sistemas de informação geográfica, ferramentas de exploração e análise espacial que permitem reduzir o grau de incerteza associado à tomada de decisão e a sua posterior conversão em cartografia permite uma melhor visualização, análise, perceção e compreensão de resultados (CTT - Correios de Portugal, S.A., 2015).

A solução Geomarketing permite a determinação e caracterização de áreas de influência em torno de pontos geográficos de interesse e segmentação geográfica de acordo com critérios seletivos sociodemográficos, habitacionais e empresariais, proveniente de diversas fontes de informação. A solução Georouting consiste num serviço suportado na caracterização de redes de eixos de via, permitindo o cálculo de matrizes origem-destino, criação de áreas de influência e sua análise e otimização de percursos. Estas soluções permitem acrescentar informação às análises de negócio já realizadas pelo cliente, facilitando a identificação de mercados potenciais, padrões de consumo e análise de atratividade e acessibilidade.

A solução de Geocontacto, apesar de pertencer a outra área de negócio – Publicidade – é responsabilidade do DPI desenvolver a sua análise. Por esta razão, e por na realidade se materializar numa solução de Geomarketing, será apresentada aqui. Este é um dos vários produtos pertencentes a uma solução para campanhas publicitárias que permite às empresas contactar com o mercado de forma seletiva e eficaz. Assim, o Geocontacto consiste num produto de Correio Não Endereçado (CNE) submetido a uma distribuição segmentada, segundo uma caracterização do território com base em critérios sociodemográficos,

habitacionais e de tecido empresarial. Pode ser considerado um *upgrade* ao CNE, caracterizado pela distribuição segmentada de objetos não endereçados em frações geográficas correspondentes à organização territorial do código postal de 7 dígitos (CP7). A caracterização do território tem por base a informação censitária disponibilizada pelo Instituto Nacional de Estatística (INE) e a base de dados Ignios – diretório de empresas.

Este produto tem duas modalidades: *Geocontacto Standard* e *Geocontacto Plus*. O *Geocontacto Standard* pressupõe a utilização de apenas informação fornecida pelo cliente. A distribuição é segmentada segundo a informação fornecida e os CTT prestam apenas o serviço de distribuição de correspondência por CP7 na área geográfica indicada pelo cliente. No *Geocontacto Plus* são utilizadas bases de dados dos CTT e há necessidade de proceder a análises geográficas para definir os CP7 alvo da campanha. O cliente identifica a área geográfica a atingir (Portugal Continental e/ou Ilhas, Distritos, Concelhos, Freguesias ou código postal de 4 dígitos (CP4)) e o perfil do consumidor. Através de um modelo de geoprocessamento (detalhado no capítulo 4.2.) são identificados os CP7 mais representativos dos critérios identificados pelo cliente e é definido o número de objetos necessários a distribuir. Este serviço permite uma comunicação assente num conhecimento concertado e profundo da população, apresentando-se mais eficaz no contacto com o *target* do que um comum CNE, conduzindo assim a melhorias nas taxas de resposta e redução de custos da campanha.

3. ÂMBITO DO ESTÁGIO

3.1. Definição do Problema e Objetivos

O propósito das soluções de Geomarketing e Geocontacto é conseguir identificar a localização do público-alvo, sendo para isso necessário um profundo conhecimento do território sociodemográfico e habitacional. O atual processo que permite esta caracterização recorre a dados geográficos ao nível do código postal de 7 dígitos, constatando-se alguns constrangimentos aquando da modelação da informação que permite a caracterização sociodemográfica e habitacional do território nacional. Concretamente, estes constrangimentos são evidentes na integração dos dados censitários do INE com o código postal de 7 dígitos.

A informação censitária do INE é georreferenciada através de subsecções estatísticas. Como veremos aprofundadamente no capítulo 4.2.1., as subsecções estatísticas apresentam uma estrutura geométrica poligonal, enquanto que o código postal assenta numa filosofia geográfica linear, em regra, ao longo de artérias. Estas estruturas geométricas distintas não se adaptam diretamente, nem em toda a sua plenitude.

Para a resolução deste problema, os CTT colocaram à disposição uma base de dados com uma camada mais detalhada do código postal, ainda em fase de qualificação e controlo de qualidade, proveniente do projeto Geo10.

O desafio lançado foi assim o de desenvolver um novo modelo de geoprocessamento de informação geográfica, permitindo obter uma melhor integração entre os dados CTT e INE e alcançar resultados mais precisos de caracterização do código postal para fins de *geomarketing*.

Doravante serão apenas expostas questões relacionadas com informação fornecida pelos dados censitários do INE, dado que foi apenas esta componente que o estágio explorou.

3.2. Pertinência e Relevância

As questões levantadas pelos CTT revelaram-se, através de uma breve revisão de literatura, como um problema que poderemos considerar como praticamente global. Porém, devido às especificidades na estrutura de dados das várias entidades de cada país, bem como diversidade nos conceitos de limites administrativos, estatísticos e censitários, o problema revela-se diferente em quase todos os casos. Por consequente, a solução deste tipo de exercício difere e tem de ser adaptada a cada situação.

Primeiramente, importa definir Sistemas de Informação Geográfica (SIG), o suporte do trabalho apresentado. Goodchild *et al.* definiram-nos como *hardware* e *software* para recolha, armazenamento, processamento, análise e visualização de informação geográfica.

Tal como apontado por Serryn Eagleson *et al.*, atualmente a maioria das fronteiras espaciais são definidas de uma forma descoordenada por organizações que desenham fronteiras individuais para responder a necessidades individuais. Acrescenta ainda que com o avanço tecnológico, as fronteiras desenhadas em mapas físicos foram digitalizadas para serem incorporados com os SIG, não significando isto que a conversão seja adequada. Como resultado, os limites administrativos são produto da era em que foram desenvolvidos, sendo hoje necessário adotar novas metodologias para que os SIG consigam atingir todo o seu potencial na análise de informação geoespacial relacionada com harmonização de fronteiras (Eagleson, Escobar, & Williamson, 2003).

Robin Flowerdew *et al.* entende que um dos temas principais dos SIG é a integração de diferentes bases de dados e um dos principais problemas é a existência de diversas unidades, utilizadas para diferentes fins. É frequente ter como objetivo a comparação de dois ou mais conjuntos de dados regionais, os quais são disponibilizados para diferentes sistemas zonais. Um exemplo muito comum diz respeito a zonas administrativas, para as quais os dados são recolhidos

por censos nacionais, departamentos governamentais e zonas postais - as quais são a forma mais simples e comum de agregar dados para fins comerciais (Flowerdew & Green, 1992). De forma a comparar dados de zonas administrativas e postais, ou outro par de sistemas zonais incompatíveis, é necessário encontrar métodos que permitam estimar o valor de um conjunto de dados caso fosse representado tendo em conta outros limites (Flowerdew & Green, 1992). Goodchild *et al.* designaram esta questão como o problema de *areal interpolation*.

Este é o processo de estimar valores sobre uma ou mais variáveis numa *target area*, com base em valores conhecidos existentes na *source area* (Hawley & Moellering, 2005). O método de sobreposição, tal como descrito por Lam (1983), permite a interpolação de variáveis com base na área de interseção entre as duas áreas, *target* e *source*. Os valores da *target area* são estimados com base nos valores da *source area* e a proporção da interseção com a *source area* (Lam, 1983 *as cited in* (Hawley & Moellering, 2005). Este processo de interpolação pode ser agrupado em duas classes: métodos de interpolação com recursos a dados auxiliares e métodos de interpolação sem recurso a outros dados (Okabe & Sadahiro, 1995).

De entre os métodos sem recursos a dados auxiliares destacam-se os métodos desenvolvidos por Flowerdew and Openshaw (1987), Flowerdew and Green (1991), Martin (1989), Burrough (1986) e Tobler (1979). Apesar de distintos, apresentam as suas bases teóricas na representação da *target area* como um ponto, assumindo por vezes características de centroide, e a *source area* como um polígono.

Os métodos de interpolação que recorrem a dados auxiliares apresentam-se muito mais eficazes e fiáveis. Apesar dos avanços tecnológicos no domínio da deteção remota e disponibilização de dados, bem como a evolução na qualidade das imagens de satélite, estes métodos são ainda morosos e dispendiosos. Os dados auxiliares consistem em dados, geralmente referentes ao território, tal como cartas de uso do solo e rede viária, que permitem, por exemplo, estudos

demográficos. As cartas de uso do solo permitem a exclusão, à partida, de determinadas porções do território como sendo habitadas. Os dados relativos a redes viárias permitem aprofundar esta análise, dado que a existência de estradas é indicativo de desenvolvimento territorial e presença de aglomerados urbanos.

Exemplos de utilização deste tipo de métodos são os de Wright (1936), o qual mapeou a densidade populacional utilizando cartas topográficas e classificando áreas como inabitadas ou com diferentes densidades, tendo por base uma estimativa controlada (Hawley & Moellering, 2005). Fisher e Langford (1995) adaptaram o método de Wright (1936) e criaram o método dasimétrico, o qual divide binariamente os tipos de uso do solo, recorrendo a dados auxiliares.

Os constrangimentos na utilização de informação censitária são já um tema recorrente na literatura. Tal como Martin (2003) explicita, um dos desafios diz respeito à compatibilização da geografia da informação censitária com outras geografias. Esta necessidade pode ocorrer quando se pretende comparar dados publicados em diferentes momentos temporais ou quando há necessidade de agregar ou desagregar a informação para se adequar a diferentes meios de investigação e análise (Martin, 2003). Esta situação resulta em incongruidades espaciais, um obstáculo sentido principalmente na integração de informação social e do território (Voss, Long, & Hammer, 1999).

O desafio é assim encontrar a melhor correspondência entre duas áreas geográficas, sendo comum recorrer a técnicas de interpolação, tal como as já descritas (Martin, 2003).

No caso em estudo, dado que códigos postais, áreas censitárias e fronteiras naturais são incompatíveis entre si, é necessário recorrer a métodos de *areal interpolation* de forma a ser possível integrar todos estes dados de diferentes fontes. Na área do *geomarketing*, um dos métodos mais recorridos é o *point-in-polygon*, no qual é analisada a inclusão ou exclusão de um ponto relativamente a um polígono com informação conhecida. O valor de um atributo na *source area* é transferido para a *target area* se um ponto da *target area* estiver inserido na

source area (Okabe & Sadahiro, 1995). Dado o elevado número de pontos de distribuição postal, é comum alocar a localização dos pontos a um centroide, representando este uma zona com um determinado código postal (Shi, 2007).

Foi esta a metodologia adotada até hoje pelos CTT para representar os códigos postais de Portugal. Como veremos, e tal como citado por Shi (2007), esta alocação pode incorrer em erros significativos. Além da questão da utilização de centroides, quando a informação censitária é agregada em polígonos, todos os pontos dentro desse polígono terão o mesmo valor, resultando na perda de especificidades locais (Robinson & Zubrow, 1997). Com a possibilidade de utilização da base de dados Geo10, a questão da utilização do centroide como meio de georreferenciação pode ser melhorada. Não obstante, e apesar deste método preservar volume, assume que a variável apresenta uma distribuição homogénea. Deste modo, é inevitavelmente incerto, dado que envolve a estimação de valores baseada em suposições sobre a distribuição de objetos espaciais (Sadahiro, 1999). É assim necessária bastante precaução aquando do uso deste tipo de métodos.

Como veremos, foi assim com base no método de *point-in-polygon* que se desenvolveu a solução apresentada aos CTT.

Apresentada a diversidade de constrangimentos e soluções encontradas aquando da integração de dados de fontes distintas, importa entender porque razão o desenvolvimento de um modelo habilitante de integração do *know-how* CTT e outras entidades nacionais se revela tão importante no sucesso das soluções relacionadas com *geomarketing*.

Geomarketing foi definido por Yrigoyen (2003) como sendo um conjunto de técnicas que permitem analisar a realidade socioeconómica de um ponto de vista geográfico, através de instrumentos cartográficos e ferramentas da estatística espacial (Ramos, Oliveira, Epiphany, & Ferreira, 2011). As investigações neste campo iniciaram-se na década de 90, podendo considerar-se como precursor os modelos desenvolvidos por Von Thünen (1826) e Weber (1909) – *teoria da localização*, Reilly (1929) – *modelo gravitacional*, ou Huff (1964) e Applebaum

(1966) – *teoria das áreas de influência*, no domínio da análise económica espacial (Cardoso, 2011). As bases de dados fundamentais no *geomarketing* são os dados censitários, comportamento dos consumidores, bem como localização de consumidores e concorrência.

Cliquet (2006) afirma que a aplicabilidade deste domínio se estende desde adaptar a oferta comercial de um espaço de comércio otimizando desempenhos, trabalhar em micro-mercados, reestruturar redes, determinar novas localizações, fixar objetivos relativamente a potencial, apontar os melhores clientes usando campanhas específicas, otimizar investimentos em *marketing* direto e analisar a competitividade (Cliquet, 2006 *as cited in* (Cardoso, 2011)).

Tal como Jones (1990) afirmou, *existem três pontos essenciais para o sucesso de um negócio: localização, localização, localização*. Através das soluções de *geomarketing*, os CTT providenciam ao mercado a possibilidade de integrarem este domínio de estudo nas suas análises de negócio.

3.3. Metodologia Adotada

O estágio e relatório realizados desenvolveram-se em 6 fases.

Numa primeira fase a ênfase foi dada a uma integração harmoniosa no contexto empresarial e, em específico, no departamento de Soluções de Desmaterialização, Proximidade e *Business Intelligence*. Foi aqui essencial o tempo despendido a entender as componentes de trabalho do departamento, a oferta de soluções e serviços, bem como a sua dinâmica com a restante Direção de Publicidade e Soluções Empresariais. Esta aprendizagem deve em muito à amabilidade dos colegas para transmissão de conhecimentos e leitura de documentação informativa.

Numa fase seguinte procedeu-se a uma análise do *estado da arte* das Soluções Geográficas dos CTT, identificando-se não só os seus aspetos positivos,

constrangimentos e questões passíveis de melhoria, bem como analisando todo o seu potencial no futuro. A partir deste momento ficou claro qual o objetivo do trabalho a desenvolver e qual o produto final ambicionado.

Na terceira fase procedeu-se ao enquadramento do tema em análise, explorando bibliograficamente a base de conhecimento sobre o problema. Esta fase apresentou-se mais conturbada do que o esperado, dado que, no contexto do problema identificado, cada país apresenta soluções diferentes com o mesmo fim, não tendo encontrado na verdade nenhuma problemática com os pressupostos e constrangimentos precisamente iguais aos do caso dos CTT. Deste modo, a pesquisa bibliográfica e aprofundamento sobre o *estado da arte* revelou-se de extrema importância, não para encontrar diretamente a resposta ao problema, mas sim para elucidar sobre os vários constrangimentos relacionados com a geografia, o território e as fronteiras artificiais concetualizadas pelo Homem, as suas incompatibilidades, e potenciar o pensamento crítico no assunto e dar início a um processo de tentativa e erro.

O início da fase seguinte pode realmente ser caracterizado como um processo de tentativa e erro de várias metodologias, tendo por base a pesquisa bibliográfica. Foi a partir desta etapa que, tendo já entendido as especificidades do caso em estudo, procedi à cimentação dos conhecimentos, e tendo sempre estes como guia no contexto científico, foi formulada uma proposta de novo modelo de geoprocessamento.

A quinta fase teve uma componente de implementação, análise e validação de resultados, tendo o modelo sido alterado, reavaliado e reformulado várias vezes até ser considerado o modelo final. O modelo foi desenvolvido em *software* ArcGIS e Microsoft Access. No final desta etapa realizou-se uma apresentação a vários colaboradores dos CTT, nomeadamente das Direções Publicidade e Soluções Empresariais e Estratégia e Desenvolvimento Corporativo, dando a conhecer o trabalho desenvolvido e as propostas de melhorias a implementar no futuro.

A última fase correspondeu à redação deste relatório, com o intuito de demonstrar todo o trabalho desenvolvido durante o período de estágio, bem como o contributo dado às Soluções DPI dos CTT.

3.3.1. Área de Estudo

A área de estudo selecionada para a incidência dos trabalhos desenvolvidos ao longo do estágio foi a unidade territorial estatística de nível III (NUT III), correspondente à Comunidade Intermunicipal do Oeste – OesteCIM, constituída por 12 concelhos (fig. 1). O distrito de Santarém e Oceano Atlântico fazem fronteira a este e oeste, respetivamente, e os concelhos de Lourinhã, Cadaval, Torres Vedras, Alenquer, Sobral de Monte Agraço e Arruda dos Vinhos pertencem ao distrito de Lisboa, enquanto que Bombarral, Peniche, Óbidos, Caldas da Rainha, Alcobaca e Nazaré pertencem ao distrito de Leiria.

Foi sobre este território que os CTT disponibilizaram os seus dados, sendo que toda informação apresentada doravante diz respeito apenas a estes concelhos.



Figura 1 – Área de Estudo: OesteCIM

4. GEOMARKETING DOS CTT

4.1. Dados de Suporte

Procede-se agora à apresentação de todos os dados de suporte utilizados ao longo do estágio, dividindo-se estes em internos, propriedade dos CTT e cedidos para este efeito, e os de fontes externas.

4.1.1. Dados de Suporte Internos

4.1.1.1. DOL – Distribuição OnLine

A DOL – Distribuição OnLine é a plataforma de apoio à distribuição dos CTT, construída com dados próprios e atualizada mensalmente.

Contem informação sobre os Pontos de Entrega (PE), Pontos de Distribuição (PD), alojamentos residenciais e funcionais por código postal de 7 dígitos, identificação do Centro de Distribuição Postal (CDP), Giro do Carteiro e identificação de Pontos de Entrega com dístico *Publicidade Aqui Não* associado.

Pontos de Entrega (PE) podem ser definidos como recetáculos postais, isto é, caixas de correio, de cada alojamento. Assim, a relação PE-alojamento deverá ser de 1 para 1, sendo que um CP7 poderá ter apenas um ou vários alojamentos. Os PE estão devidamente caracterizados segundo condições de habitação do alojamento associado, adquirindo as seguintes características:

- PE Residencial
- PE Coletivo
- PE Comercial
- PE Degradado
- PE Desabitado
- PE Férias

Os PE são ainda identificados como tendo, ou não, associado um dístico proibindo a colocação de publicidade não endereçada.

Os Pontos de Distribuição (PD) referem-se à entrada do edifício onde será entregue correspondência. Este é identificado como ativo ou inativo e poderá ter associado vários Pontos de Entrega, apenas um, ou até nenhum.

4.1.1.2. Geindex

Geindex² é a base de dados utilizada para georreferenciar o código postal. Define assim as coordenadas do código postal de 7 dígitos, o qual, de um modo geral, representa uma rua, um conjunto de ruas ou um lugar (CTT - Correios de Portugal, S.A., s.d.). Apesar de ser utilizada como geometria do código postal, esta não localiza todas as portas pertencentes a um código postal, mas sim o seu centro geométrico, o centroide (fig. 2). O centroide do código postal foi definido segundo as fachadas dos quarteirões e artérias, sendo um ponto georreferenciado sobre uma artéria, representativo dessa rua, segmento de rua, conjunto de ruas ou lugar, também comumente designado por coordenadas médias do código postal de 7 dígitos (CTT - Correios de Portugal, S.A., s.d.).

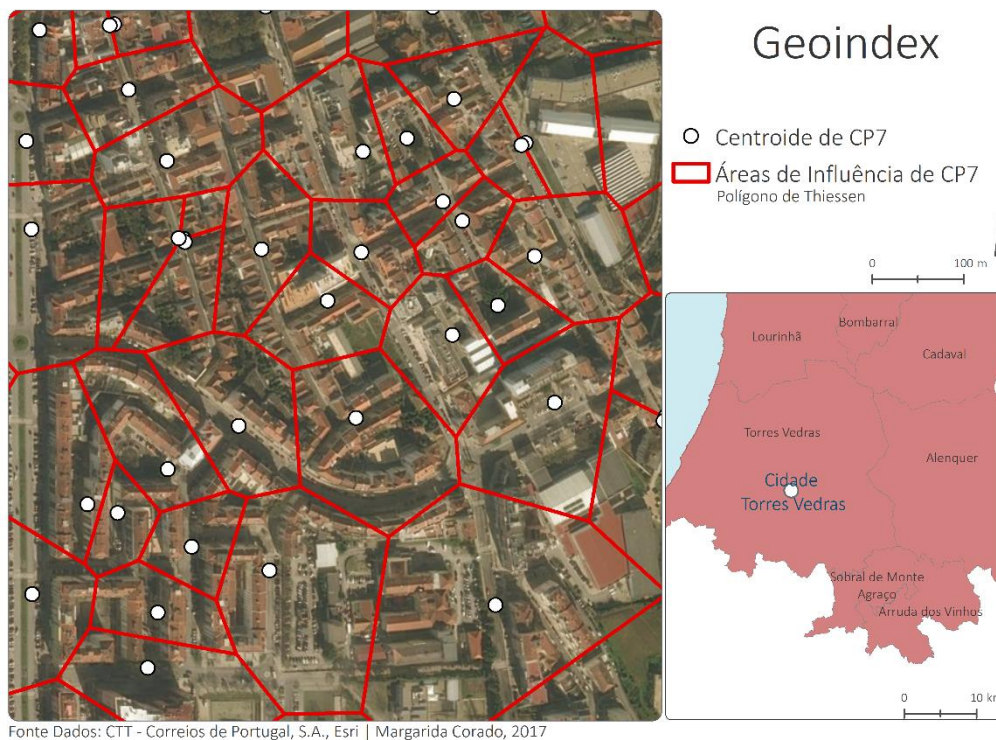


Figura 2 - Geindex: Centroide de CP7 e Áreas de Influência

² Apesar da mesma denominação, não confundir com a Solução Geográfica Geindex.

Como veremos no capítulo 4.2. esta é a base de dados utilizada atualmente para georreferenciar e caracterizar o código postal de 7 dígitos nos estudos geográficos.

4.1.1.3. BNE – Base Nacional de Endereços

A Base Nacional de Endereços dos CTT é o suporte de todas as bases de dados relacionadas com o código postal. Nesta consta a identificação de alojamentos funcionais e residenciais por CP7, a designação da porta, ou seja, o número de polícia ou nome do edifício, identificação do Ponto de Entrega, tal como 1ºEsq., 18ºDrt., enquadramento e toponímia das artérias e identificação do Ponto de Distribuição como ativo ou inativo.

4.1.1.4. Geo10

O Geo10 foi um projeto inicialmente idealizado em 2009 e iniciado em 2011 com o objetivo de georreferenciar e caracterizar todas as ruas, portas e domicílios a nível nacional. O trabalho no terreno já terminou no Continente, estando agora em fase de qualificação e controlo de qualidade. Nas Ilhas será dado agora início aos levantamentos no terreno.

Além da informação já disponibilizada pela BNE e DOL, o Geo10 permite a caracterização da porta e edifício.

As portas podem adquirir as seguintes características:

- Porta
- Garagem
- Janela/Montra/Outra
- " (em branco)

Os edifícios podem ser caracterizados como:

- Normal
- Em Construção
- Fechado/Desabitado
- Não Avaliado

- Em Ruína
- " (em branco)

O Geo10 será o sucessor do Geindex ao permitir a georreferenciação de todas os PD (fig. 3) invés da representação aproximada do centro médio de um CP7, e a visualização da sua distribuição ao longo das artérias.

Foi idealizado com o objetivo de dotar os CTT de maior competitividade na área de negócios dos endereços, permitindo maior eficiência na gestão do território postal (Carvalho & Marcelino, 2015).

Como veremos com este relatório, dotará os CTT de uma vantagem competitiva única e inigualável ao melhorar substancialmente os pressupostos e resultados das Soluções Geográficas.

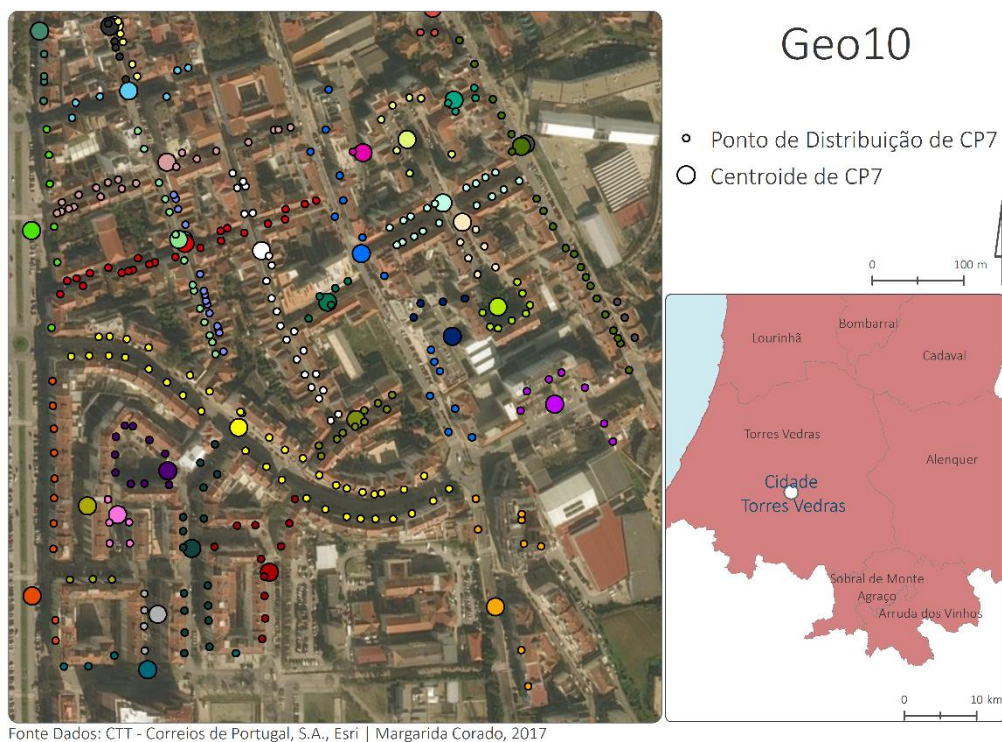


Figura 3 - Geo10: Centroide de CP7 e Ponto de Distribuição

4.1.1.5. O Código Postal

Importa referir que o *core* de informação dos CTT advém do código postal e que é este o suporte das bases de dados apresentadas anteriormente.

O código postal foi criado pelos CTT não só com o intuito, já referido, de ultrapassar problemas de ambiguidade de endereçamento e toponímia, bem como automatizar os processos, até então manuais, de classificação, encaminhamento e distribuição de correspondência (Barradas & Ferrão, 1996).

Até 1998 o código postal era composto por apenas 4 dígitos, indicativos da localidade de destino.

Procedendo à divisão do território nacional em 9 (fig. 4), com início em Lisboa e prosseguindo para norte, em espiral, no sentido dos ponteiros do relógio, obteve-se o primeiro algarismo de cada código postal de 4 dígitos (CP4) (Barradas & Ferrão, 1996).

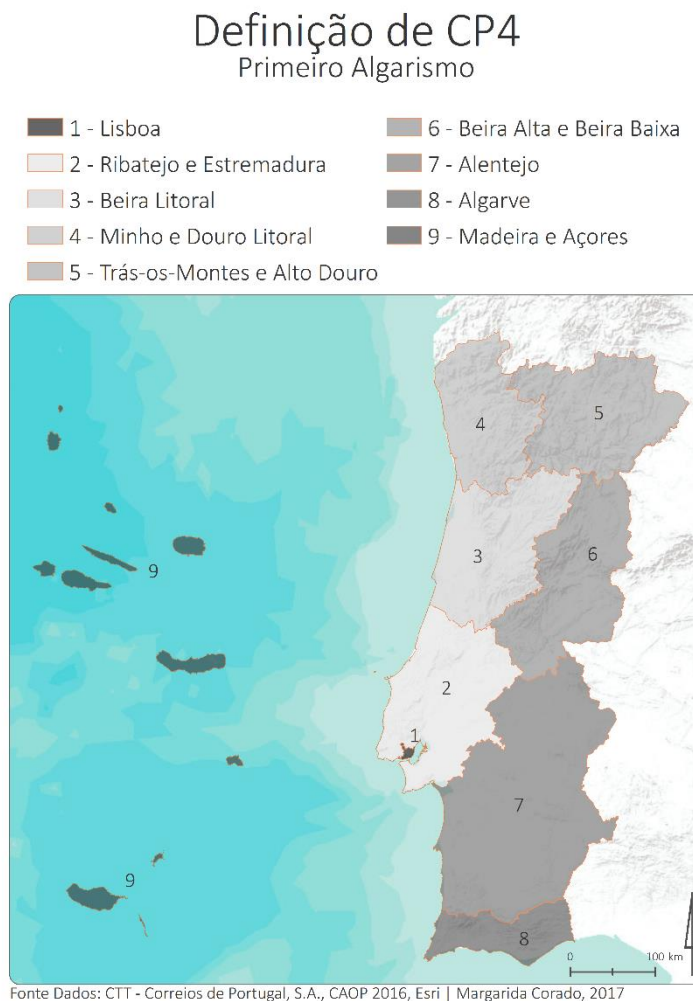


Figura 4 - Divisão do território nacional para definição de primeiro algarismo de CP4

Em 1998 é introduzido o código postal de 7 dígitos (CP7), hoje utilizado. Os 3 novos dígitos dizem respeito a ruas ou segmentos de rua, identificando as frentes de quarteirão. Quando foi idealizado, as ruas de Lisboa, cidade-piloto do projeto,

foram “cortadas” a cada 400 metros, obtendo-se uma planta dividida em arcos e nós, com uma numeração atribuída a cada divisão (Barradas & Ferrão, 1996).

Hoje a filosofia mantém-se, podendo a divisão ser mais flexível consoante as especificidades da localidade. As zonas rurais, por apresentarem menor volume de tráfego postal, poderão ter associadas um código postal de 7 dígitos único para toda a localidade.

O código postal foi criado com o propósito único de auxiliar a atividade de então dos CTT, distribuição de correio. Dada a sua capacidade de georreferenciação precisa, pode hoje auxiliar muitos mais segmentos de negócio da empresa, como veremos.

4.1.2. Dados de Suporte Externos

4.1.2.1. Censos

Os Recenseamentos da População e Habitação, comumente designados Censos, realizados de 10 em 10 anos pelo Instituto Nacional de Estatística (INE), têm como objetivo recensear todos os cidadãos e famílias residentes no território português no momento censitário, bem como todos os alojamentos e edifícios destinados a habitação (INE, 2011).

Os Censos são uma importante fonte de informação sobre as características da população e do parque habitacional de um país, entrando no detalhe da freguesia ou até mesmo quarteirão, permitindo comparar a evolução e dinâmica de fenómenos ao longo do tempo (Oliveira, 2008). A nível internacional, são de extrema importância para comparação da estrutura social e económica, bem como análise de evolução e tendências, com outros países.

É possível segmentar a informação proveniente dos Censos em três chavões – edificado, parque habitacional e população. A nível informativo referem-se as

variáveis, das 136 disponibilizadas nos Censos 2011, mais requisitadas para os estudos levados a cabo pelos CTT:

- Em relação ao edificado, variáveis como número de edifícios clássicos com 1 ou 2 pisos, número de edifícios clássicos com 3 ou 4 pisos, número de edifícios construídos após o ano 2000, número de edifícios exclusivamente residenciais, número de edifícios principalmente não residenciais;
- Em relação ao parque habitacional, variáveis como número de alojamentos familiares, número de alojamentos vagos, número de alojamentos de residência habitual, número de alojamentos de residência secundária, número de residência habitual com 1 ou 2 divisões, número de residência habitual com 3 ou 4 divisões, número de residência habitual cujo ocupante é o proprietário, número de residência habitual arrendada;
- Em relação à população, no que se refere a dados sobre família, variáveis como número de famílias clássicas com 1 ou 2 pessoas, número de famílias clássicas com um residente com idade superior a 65 anos, número de famílias clássicas sem desempregados, número de núcleos familiares com pelo menos um filho com idade inferior a 6 anos;
- Em relação à população, no que se refere a dados sobre os residentes, variáveis como número de indivíduos masculinos ou femininos com idade superior a 64 anos, número de indivíduos masculinos ou femininos em idade ativa, número de indivíduos que não sabem ler nem escrever, número de indivíduos com ensino superior completo, número de indivíduos desempregados à procura do primeiro emprego, número de indivíduos empregados, número de indivíduos desempregados, número de indivíduos pensionistas ou reformados, número de indivíduos empregados no setor primário/secundário/terciário.

Aliada aos Censos existe a Base Geográfica de Referenciação de Informação (BGRI), a qual permite que a informação censitária alfanumérica seja georreferenciada. A BGRI desenvolve-se segundo uma estrutura poligonal

hierárquica cuja unidade elementar de representação é a subsecção estatística (Geirinhas, 2001). A partir desta unidade é possível, por agregação, construir qualquer representação de nível hierárquico superior (Geirinhas, 2001). Além das divisões administrativas – freguesia, concelho e distrito – e dos três níveis da Nomenclatura de Unidades Territoriais para fins Estatísticos – NUTS I, NUTS II, NUTS III – a BGRI é constituída por outros quatro tipos de unidades espaciais – Subsecção Estatística, Secção Estatística, Lugar e Base Mínima Comum (Geirinhas, 2001).

A subsecção estatística é possivelmente a unidade mais abstrata e de difícil compreensão. Pode ser entendida como a menor área homogénea, construída ou não, dentro de uma secção estatística. A secção estatística corresponde à agregação de subsecções que configurem uma área contínua com cerca de 300 alojamentos. A subsecção estatística caracteriza-se por estar associada ao código e topónimo, nas zonas rurais, do lugar ou parte de lugar, e nas zonas urbanas, do quarteirão em que se insere.

Os Censos e BGRI têm um papel preponderante nos estudos desenvolvidos no âmbito das Soluções Geográficas, ao permitirem a caracterização e segmentação territorial em função da população, habitação e edificado.

4.1.2.2. Ignios

Apesar de não ter sido estudo deste estágio, apresenta-se a base de dados Ignios, por esta permitir complementar a informação fornecida pelos Censos com informação de carácter empresarial.

Consiste num diretório de todas as empresas em território português, fornecendo informação sobre o número de empregados, volume de faturação e identificação do CAE – Classificação de Atividades Económicas.

Permite assim a caracterização e segmentação territorial em função do tecido empresarial.

4.2. Modelo de Geoprocessamento Atual

4.2.1. Apresentação do Modelo Atual

Os estudos subjacentes às Soluções Geográficas dos CTT assentam numa caracterização do território, passível de ser analisada, e na transformação de informação em conhecimento útil aos seus clientes. A caracterização do território tem por base a informação proveniente das fontes já mencionadas, nomeadamente a informação censitária do INE e os códigos postais dos CTT. Sendo o código postal o meio preferencial de trabalho dos CTT, dada a sua exatidão, precisão, minuciosidade e refinamento, é a caracterização deste o fator diferenciador que a empresa detém. Através de uma caracterização minuciosa do código postal é possível obter uma visão sobre o território bastante concertada.

O modelo atual procede à análise espacial através da associação da informação censitária, georreferenciada pelas subsecções estatísticas, e o código postal, georreferenciado pelo Geindex.

Como já mencionado, o Geindex georreferencia o código postal de 7 dígitos através do seu centroide. Ao contrário de vários modelos adotados pelas entidades postais de outros países, os CTT não adotaram o conceito de polígonos de código postal, uma forma direta de saber a real distribuição e abrangência de um código postal. Desta forma, para que seja considerada a representatividade de um código postal ao longo do território, utilizam-se áreas de influência dos centroides do CP7, definidas através da construção de Polígonos de Thiessen (fig. 5). Os Polígonos de Thiessen são criados com base numa fórmula matemática que garante que cada ponto dentro desse polígono está mais perto do seu centroide do que de qualquer outro centroide (Smith, Goodchild, & Longley, 2015).

Para associar a informação censitária aos centroides recorre-se à ferramenta *spatial join*, em *software* ArcGIS, utilizando as subsecções estatísticas e os Polígonos de Thiessen. Esta ferramenta, tendo dois conjuntos de dados de objetos multidimensionais com representação num espaço euclidiano, encontra todos os

pares de objetos que satisfazem uma determinada relação (Jacox & Samet, 2006). Através da análise de sobreposição das duas bases de dados, com uma relação n-n, a cada centroide de CP7 é alocada informação proveniente de uma ou várias subsecções estatísticas. O *spatial join* distingue-se assim de uma simples relação entre dois conjuntos de dados ao permitir que a localização assuma um fator determinante (Jacox & Samet, 2006).

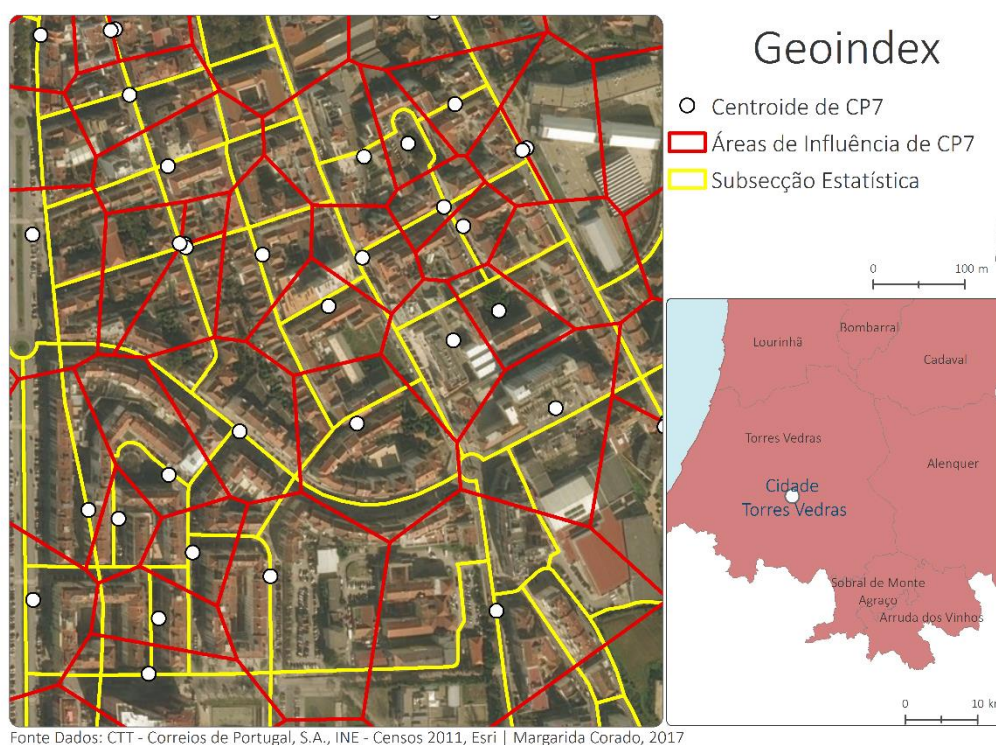


Figura 5 – Dados Geoindex: Relação CP7 e Subsecção Estatística, através de Centroides e Áreas de Influência

Dado que muitas vezes o objetivo do estudo é apenas selecionar os códigos postais de 7 dígitos mais representativos de determinados critérios, apresenta-se aqui o pressuposto definido para considerar a representatividade de cada código postal.

Para tal recorre-se a uma medida de localização, as quais têm como objetivo comparar a expressão de uma determinada atividade, por norma económica, em territórios distintos, adotando uma perspetiva setorial (INE, 2002). Estas medidas de análise regional, tradicionalmente utilizadas na análise dos tecidos produtivos regionais e em análises económicas setoriais à escala regional e sub-regional têm

visto o seu âmbito ser alargado para análises do tecido social e para níveis de desagregação mais finas, nomeadamente, em análises de *geomarketing* em espaço urbano (INE, 2002), adaptando-se ao espectro de variáveis disponibilizadas nos Censos.

A medida utilizada é o Quociente de Localização (QL), que permite avaliar o grau relativo de concentração de uma determinada atividade. Neste cálculo é comparada a importância da atividade j na região r , com a importância que essa mesma atividade tem numa região padrão p (INE, 2002).

$$QL_{rj} = \frac{X_{rj}}{X_r} / \frac{X_{pj}}{X_p} \quad (0 \leq QL_{rj} \leq \infty) \quad (1)$$

sendo:

X_{rj} - O valor da atividade j na região r

X_r - O valor do total das atividades consideradas, na região r

X_{pj} - O valor da atividade j na região padrão p

X_p - O valor do total das atividades consideradas, na região padrão p

A região padrão é a região, ou agregado de regiões, de referência, podendo ser o espaço composto pelo conjunto de regiões em análise ou outra região distinta (INE, 2002). O QL pode assim ser sintetizado como um índice comparativo da importância de uma parte de um fenómeno em relação ao todo desse mesmo fenómeno, numa determinada área geográfica. Relativamente aos resultados, valores inferiores a 1 indicam uma expressão do setor j na região r inferior à expressão desse setor na região padrão; por seu lado, valores superiores a 1 indicam uma expressão do setor j superior à verificada na região padrão (INE, 2002).

No caso em estudo, será comparada a representatividade de uma variável dos Censos v na subsecção estatística s em relação à sua variável de relativização³ r , com a importância que essa mesma variável tem no território nacional em relação à sua variável de relativização.

Por razão desconhecida, apesar da fórmula utilizada ser a do QL, esta não está devidamente implementada, tendo sido dividida em dois passos distintos. Será agora demonstrado o processo adotado no contexto empresarial, o qual foi revisto e otimizado no modelo proposto, como será apresentado no capítulo 4.3..

Primeiramente é hierarquizada a representatividade da variável em estudo pelas diferentes subsecções estatísticas, através da fórmula 2. Esta fórmula foi anteriormente aplicada a todas as variáveis e subsecções estatísticas, existindo hoje uma base de dados com um registo por subsecção estatística e as variáveis dos Censos recalculadas.

$$\frac{V_s}{V_{rs}} \times 100 \quad (2)$$

sendo:

V_s - valor da variável por subsecção estatística

V_{rs} – valor da variável de relativização por subsecção estatística

³ Variável de relativização refere-se à variável base que é segmentada segundo diferentes características e classes. A título de exemplo, a variável de relativização da variável *número de edifícios com 1 ou 2 pisos* é a variável *número de edifícios clássicos; número de indivíduos residentes com mais de 64 anos* tem como variável de relativização *número de indivíduos residentes*.

Os valores resultantes são posteriormente comparados com um valor de corte, obtido da seguinte forma:

$$\frac{V_t}{V_{rt}} \times 100 \quad (3)$$

sendo:

V_t - valor total nacional da variável

V_{rt} – valor total nacional da variável de relativização

Assim, é comparado o valor resultante da fórmula 2 com o resultado da fórmula 3. Caso seja inferior ao valor de corte, é considerado que a representatividade da variável em estudo na subsecção estatística é inferior à representatividade dessa variável no território nacional.

Através da relação subsecção estatística – Polígono de Thiessen de CP7, os CP7 associados a subsecções estatísticas com valores inferiores ao valor de corte não são selecionadas para as campanhas de *geomarketing*.

4.2.2. Aplicação do Modelo – Estudo Geocontacto

Será agora simulado um estudo de Geocontacto *Plus*.

A *Associação Crescer*⁴ abriu mais um espaço social para promoção de relações intergeracionais, onde o objetivo é juntar diariamente seniores e crianças em idade pré-escolar para atividades conjuntas, criando um espaço de centro de dia e infantário.

Pretende assim criar uma campanha de distribuição de folhetos publicitários, tendo optado por não seguir com uma campanha de Correio Não Endereçado,

⁴ Caso fictício.

dado ter um público alvo bastante específico e preferir uma campanha segmentada, aplicando-se um estudo de Geocontacto *Plus*.

A Associação definiu o seu público alvo como residentes seniores, segundo o INE população com idade superior a 64 anos, e residentes com idades compreendidas entre os 5 e os 9 anos. Definiu ainda a sua área geográfica de ação como os concelhos de Torres Vedras, Alenquer, Sobral de Monte Agraço, Arruda dos Vinhos e Caldas da Rainha.

Tendo já a base de dados que associa a subsecção estatística ao código postal de 7 dígitos, é necessário encontrar os CP7 com maior predominância do público alvo descrito. Para este processo recorre-se a *software Access*, onde se ligam três tabelas e se corre uma *query*. As tabelas utilizadas dizem respeito a:

- Tabela com ligação subsecção estatística – Polígono de Thiessen de CP7;
- Tabela DOL – Distribuição OnLine;
- Tabela com Censos recalculados por subsecção estatística, segundo fórmula 2.

A tabela DOL – Distribuição OnLine é utilizada para obter a informação sobre CP4, concelho, freguesia e número de Pontos de Entrega Residencial (PER) por CP7 (dado que o número de objetos por campanha é definido pelo número de PE, neste caso residencial e não funcional, existentes nos CP7 alvo). A ligação entre esta tabela e a informação censitária é feita através da tabela com a correspondência entre a subsecção estatística e o CP7.

É agora necessário calcular o valor de corte (fórmula 3) para restringir a seleção dos CP7 com variáveis iguais ou superiores ao valor de referência.

Valor de corte para a variável *número de indivíduos residentes com idade compreendida entre os 5 e 9 anos*:

$$\frac{525\ 087}{10\ 562\ 178} \times 100 = 4,97 \quad (4)$$

Valor de corte para a variável *número de indivíduos residentes com idade superior a 64 anos*:

$$\frac{2\,010\,064}{10\,562\,178} \times 100 = 19,03 \quad (5)$$

Com esta informação procede-se à construção da *query*, exemplificada na fig. 6.

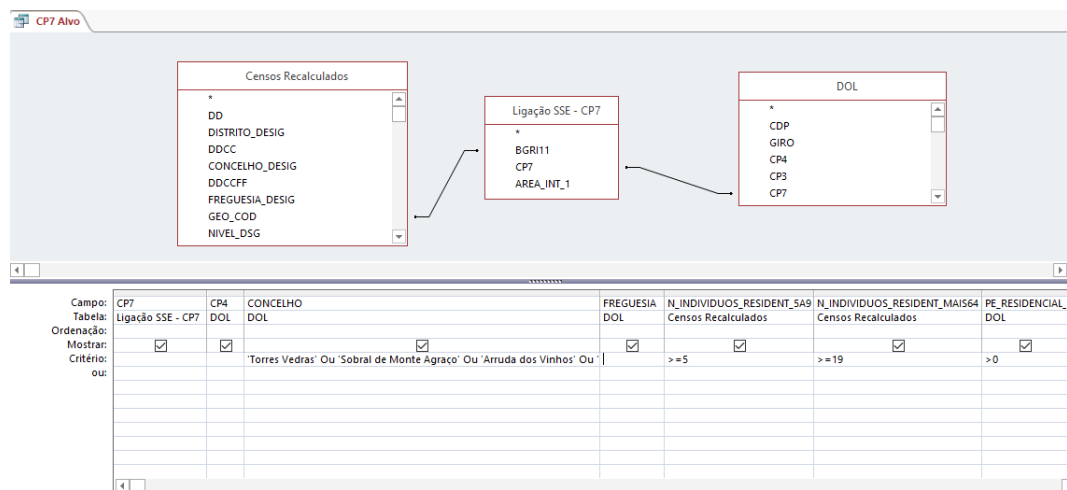


Figura 6 – Modelo Atual: *query*, em *software* Access, para estudo Geocontacto

Dado que o mesmo Polígono de Thiessen de CP7 pode intersestar várias subsecções estatísticas, é necessário recorrer a um *group by* ao campo CP7 para eliminar possíveis registos duplicados. Daqui obtêm-se os CP7 alvo da campanha com informação sobre número de PER associados⁵.

O resultado desta análise resulta na seleção de 1 411 CP7 alvo, equivalendo a 49 388 PER. Dado que o número de objetos a distribuir é arredondado ao milheiro, o cliente adjudicará uma campanha de 50 000 objetos.

⁵ Por forma a manter a confidencialidade dos dados detidos pelos CTT, apresentar-se-á apenas os resultados finais.

4.2.3. Análise de Resultados

Procede-se agora à análise dos resultados obtidos da segmentação territorial com base em variáveis sociais. A forma mais fiável de analisar a exatidão na escolha dos CP7 alvo é através da representação cartográfica dos dados disponíveis – subsecções estatísticas, informação censitária e áreas de influência do código postal (fig. 7).

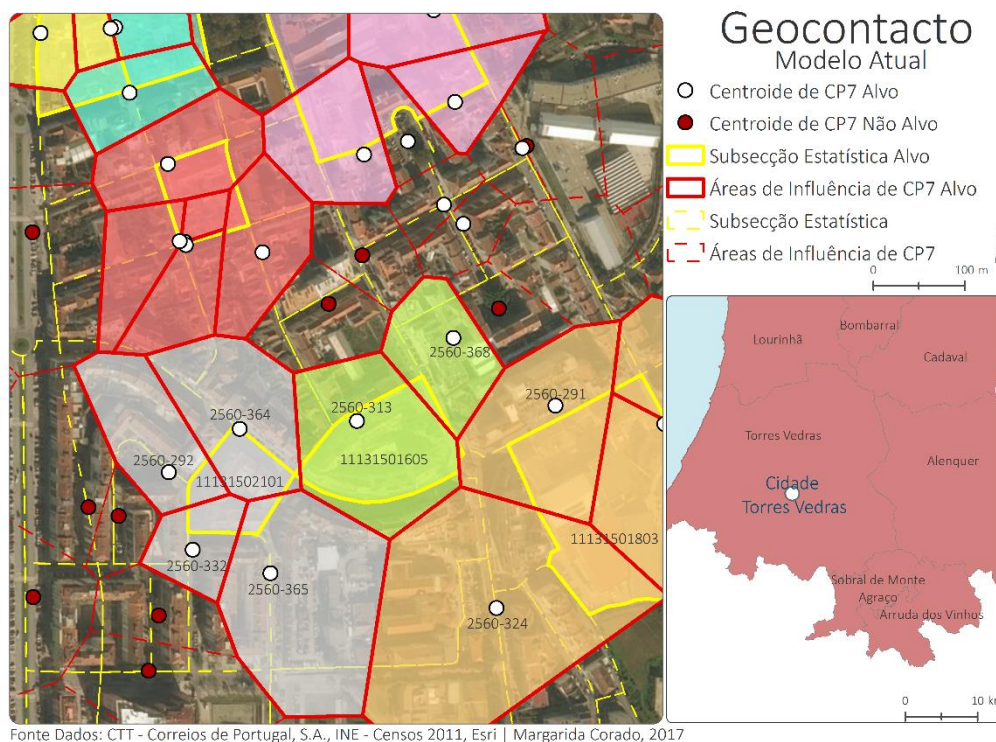


Figura 7 – Modelo Atual: Exemplo de resultado de estudo Geocontacto

Através da análise cartográfica é facilmente perceptível como os códigos postais de 7 dígitos se relacionam com as subsecções estatísticas. Primeiramente são selecionadas as subsecções estatísticas que contenham informação censitária que se enquadre no estudo. As áreas de influência de CP7 que intersetem estas subsecções estatísticas definem que o seu código postal correspondente será incluído na campanha. Qualquer código postal cuja área de influência não intersete a subsecção estatística é expurgado da campanha.

A partir da cartografia é claro que as dimensões e extensão das áreas de influência não se assemelham nem enquadram à geometria linear que o código postal apresenta. Esta questão pode resultar na seleção de CP7 indevidos, dado que a distribuição das suas portas pode não coincidir de forma alguma com a área de influência obtida através da definição de Polígonos de Thiessen.

No capítulo 4.4. veremos em pormenor as dificuldades do modelo em selecionar os CP7 alvo.

4.3. Modelo de Geoprocessamento Proposto

Apresentar-se-á agora o modelo de geoprocessamento proposto.

Primeiramente recorre-se ao capítulo 4.3.1. Tratamento de Dados de forma a explicitar a necessidade de tratamento dos dados provenientes do projeto Geo10 e de seguida, no capítulo 4.3.2. apresenta-se o desenvolvimento do modelo.

4.3.1. Tratamento de Dados

O projeto Geo10 encontra-se ainda em fase de qualificação e controlo de qualidade, pelo que se procedeu ao tratamento dos dados, de forma a serem integrados no modelo apenas dados de qualidade e de interesse para concretização do seu propósito - Pontos de Distribuição passíveis de receber correspondência e de terem associado a si variáveis censitárias, nomeadamente capacidade de serem habitados. Foram analisados 214 397 Pontos de Distribuição na região OesteCIM.

Foram expurgados os registos com PD não ativos, identificados na base de dados com um 0, equivalendo à exclusão de 40 345 registos. Estes PD não seriam considerados pelo carteiro aquando da distribuição de correspondência, pelo que apenas interessa integrar no modelo os PD ativos.

Analisando a informação associada a cada PD, foram seleccionados apenas os PD com alojamentos residenciais associados, podendo ainda ter associado alojamentos funcionais ou não, tendo sido expurgados 182 688 registos, os quais representavam apenas alojamentos funcionais.

Quanto à caracterização do edifício, foram apenas tidos em conta os caracterizados como *normal* ou " (em branco), e excluídos os *em construção, fechado/desabitado, não avaliado, em ruína*, tendo sido excluídos 11 372 registos. Considerou-se que edifícios *em construção, fechado/desabitado* ou *em ruína* não são passíveis de ser habitados. Os PD em edificado caracterizado como " (em branco) foram incluídos no modelo dado que, devido ao elevado número de

registos assim caracterizados, considerou-se que a sua caracterização deveria ser *by default* e equivaleriam a um *normal*.

Relativamente à caracterização da porta associada, foram excluídos os caracterizados como *garagem, janela/montra/outra*, ou " (em branco), sendo apenas mantido *portas*. Estes 7 389 registos foram expurgados por este tipo de porta não corresponder a uma entrada principal de alojamento residencial.

Além disto, 18 405 PD não tinham informação sobre o CP7 corretamente associada, apresentando o campo CP3 em branco. Procedeu-se à sua eliminação, dado que a informação correta sobre o CP7 é vital para o modelo.

Aquando do *spatial join* descrito no capítulo seguinte, alguns registos foram identificados como não estando completamente dentro de uma subsecção estatística, tendo como informação associada o registo *NULL*. Procedendo à sua análise individual concluiu-se que estes se sobrepõem ao limite da subsecção estatística, não estando em conformidade com o critério *completely within* definido. Dado o seu valor residual, estes foram expurgados.

A *query* que resume o processo de tratamento de dados descrito é a exposta na figura 8. Daqui resultam 139 788 PD a considerar no modelo de geoprocessamento.

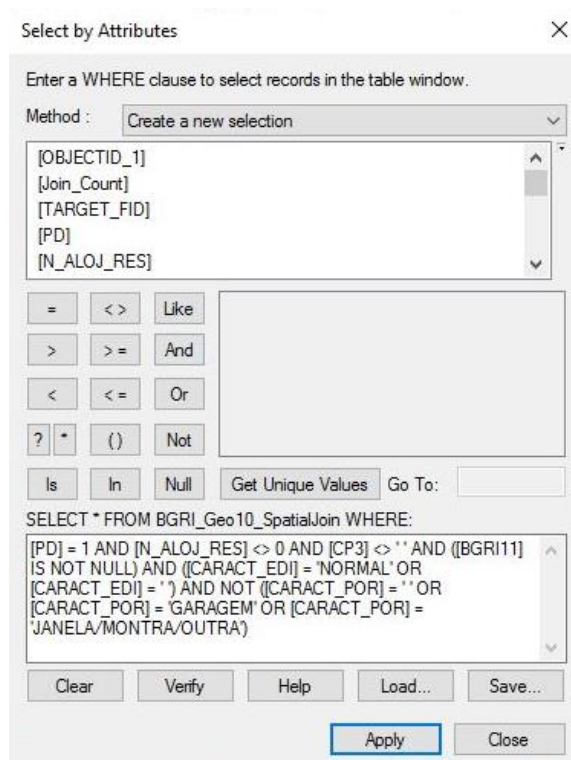


Figura 8 – Query inserida na ferramenta *select by attributes*, em software ArcGIS, para tratamento de dados

Estes seriam os registos a processar. Porém, aquando do primeiro estudo realizado, identificaram-se incoerências entre as bases de dados DOL – Distribuição OnLine e Geo10. Registou-se que alguns PD georreferenciados dentro dos limites administrativos do OesteCIM, aquando da associação da informação alfanumérica presente na DOL, eram identificados como pertencentes a concelhos limítrofes da área de estudo, nomeadamente Azambuja, Mafra e Marinha Grande. Estes registos, equivalendo a 19 CP7, 68 PD e 77 alojamentos residenciais, foram considerados como residuais, expurgados da análise e reportados para análise e posterior resolução.

Assim sendo, dos 214 397 Pontos de Distribuição do OesteCIM foram processados no modelo 139 741.

Paralelamente foi criado um modelo (fig. 9), utilizando a ferramenta *Model Builder*, em *software* ArcGIS, que permite replicar todo o processo para qualquer área de estudo, bastando para isso carregar os dados necessários (fig. 10).

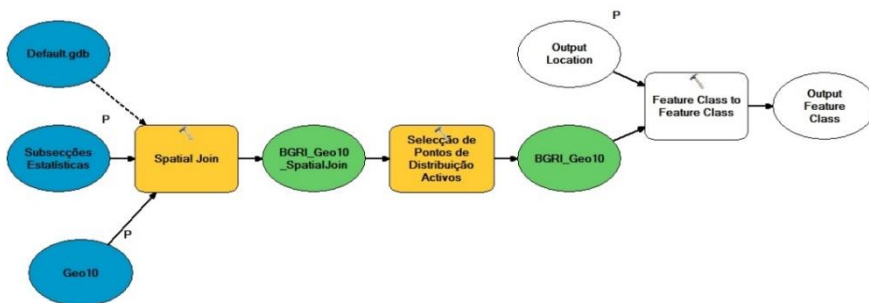


Figura 9 - Modelo, construído em *Model Builder*, *software* ArcGIS, para tratamento de dados

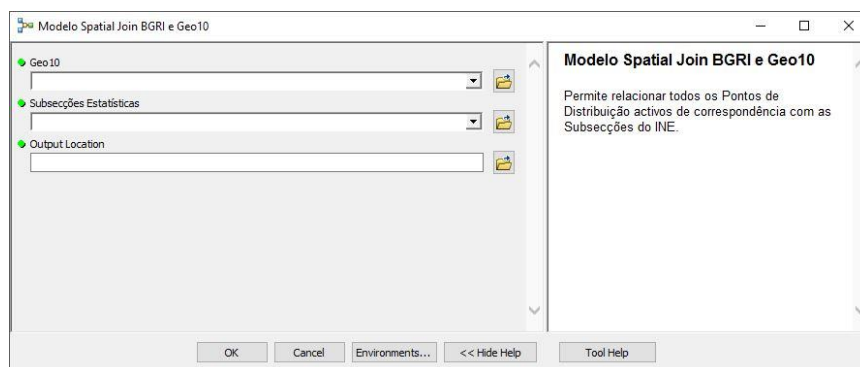


Figura 10 – Formulário para utilização do modelo anterior

4.3.2. Apresentação do Modelo Proposto

O modelo de geoprocessamento proposto está totalmente dependente da utilização da base de dados Geo10, dado que a alteração chave ocorre na forma como a informação censitária e o código postal se relacionam. A adoção da base de dados Geo10 para resolução de questões de *geomarketing* permite descartar a relação entre as subsecções estatísticas e o código postal de 7 dígitos, representado pelos Polígonos de Thiessen.

Dado que o Geo10 representa cada Ponto de Distribuição, ou seja, cada entrada de edifício, a associação direta entre o CP7 e a informação censitária é possível. Para tal, mais uma vez recorreu-se à ferramenta *spatial join*, tendo sido definida uma relação 1-1, dado que para cada PD existe apenas uma subsecção estatística. A opção de correspondência (*match option*) é *completely within*, na qual a correspondência é validada quando o *target feature* - neste caso o Geo10 - está completamente dentro dos limites do *join features* - os polígonos das subsecções estatísticas.

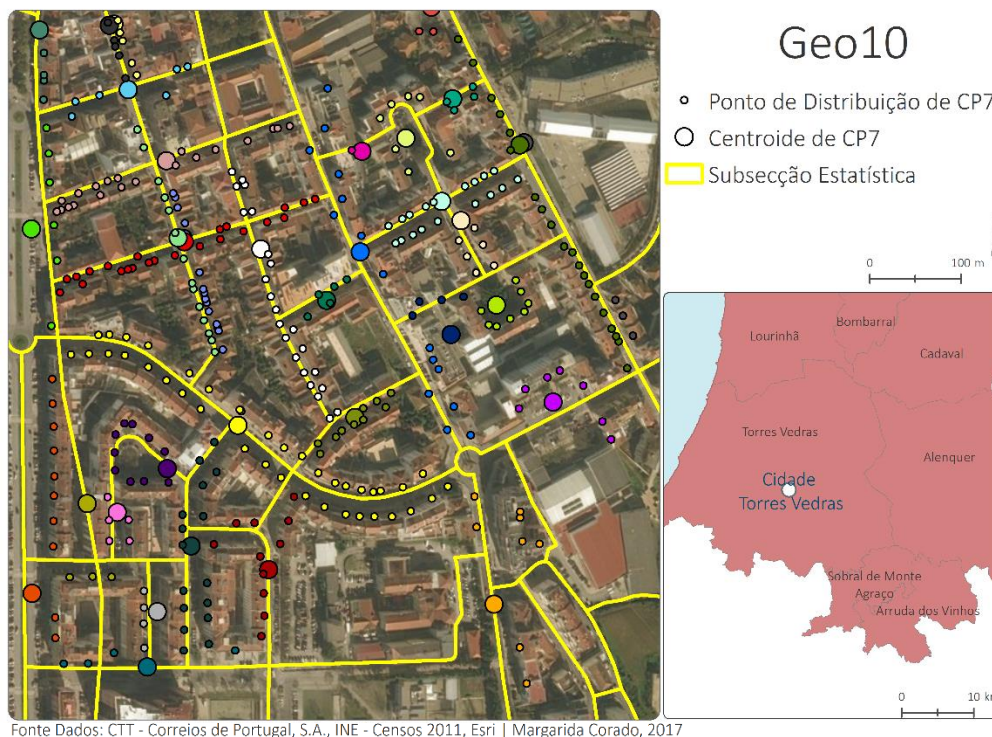


Figura 11 – Dados Geo10: Relação CP7 e Subsecção Estatística, através de Ponto de Distribuição

Apresenta-se agora o modelo desenvolvido para capacitar o CP7 de caracterização proveniente dos Censos. O desafio, mais uma vez, foi conciliar as diferentes geometrias dos dados. Tal como perceptível pela figura 11, o que acontece na realidade é que uma subsecção estatística pode abranger apenas um CP7 ou vários CP7, tal como um CP7 pode distribuir-se por apenas uma subsecção estatística ou várias. Deste modo, é necessário distribuir a informação censitária presente nas subsecções estatísticas pelos diferentes Pontos de Distribuição, de variados CP7, que a intersetam.

O modelo de geoprocessamento proposto tem como objetivo ratear a informação pelos CP7, segundo o número de alojamentos residenciais de cada. A fórmula criada é a seguinte:

$$\sum_{CP7} \left(\frac{V_s}{\sum A_s} \times \sum A_{cs} \right) \quad (6)$$

sendo:

V_s - valor da variável por subsecção estatística

A_s – alojamentos por subsecção estatística

A_{cs} – alojamentos por CP7 na subsecção estatística

O primeiro passo passou por proceder a uma distribuição linear por subsecção estatística, onde o valor de cada variável é dividido pelo somatório de alojamentos residenciais presentes na subsecção estatística, independentemente do seu CP7. Deste modo obtém-se a informação censitária por alojamento.

A distribuição não poderia ficar por este primeiro passo, dado que cada CP7 é diferente, tendo “pesos” diferentes consoante o número de Pontos de Distribuição e Pontos de Entrega. Para que esta diversidade seja tida em conta é necessário proceder a uma distribuição ponderada. Atendendo às especificidades do negócio, determinou-se que o fator de ponderação deveria ser o número de alojamentos residenciais por CP7. Um CP7 com mais alojamentos residenciais,

materializem-se estes em construção vertical ou edifícios unifamiliares, terá um peso superior aquando da distribuição da população. Para tal, multiplica-se o valor obtido da distribuição linear pelo somatório de alojamentos por CP7 na subsecção estatística. Procedendo ao somatório por CP7 é finalizado o processo, o qual pode ser sintetizado como o mapeamento dos dados censitários ao código postal, isto é, a caracterização sociodemográfica e habitacional do código postal.

Comprova-se que a informação foi devidamente distribuída por todos os alojamentos residenciais pois o seu somatório ao código postal de 7 dígitos é igual ao somatório por subsecção estatística.

Tendo já os CP7 com informação censitária associada, e dado que, tal como explicado anteriormente, o objetivo muitas vezes é apenas selecionar os códigos postais de 7 dígitos mais representativos de determinados critérios, é agora necessário implementar o Quociente de Localização.

Apercebendo-me dos constrangimentos no cálculo de valor de corte implementado atualmente, e nas dificuldades sentidas na interpretação destes valores, tomou-se a decisão de otimizar o processo.

Desta vez será comparada a representatividade de uma variável dos Censos v no código postal de 7 dígitos c , em relação à sua variável de relativização r , com a importância que essa mesma variável tem na área de estudo em relação à sua variável de relativização. A fórmula é aplicada à base de dados que contem a caracterização sociodemográfica e habitacional por CP7.

$$QL_{rc} = \frac{V_c}{V_{rc}} / \frac{V_t}{V_{rt}} \quad (0 \leq QL_{rc} \leq \infty) \quad (7)$$

sendo:

V_c - valor da variável por CP7

V_{rc} – valor da variável de relativização por CP7

V_t - valor total da área de estudo da variável

V_{rt} – valor total da área de estudo da variável de relativização

Aplicando esta fórmula, os resultados são analisados consoante a sua variação em relação ao valor 1, o valor de corte. Valores inferiores a 1 indicam uma expressão da variável em estudo, no CP7, inferior à representatividade dessa variável na área de estudo, podendo esta ser o território nacional ou área geográfica do estudo. Valores superiores a 1 indicam uma expressão da variável em estudo, no CP7, superior à representatividade dessa variável na área de estudo, pelo que serão os CP7 selecionados para a campanha como CP7 alvo.

Tendo como perspetiva as utilizações futuras, acrescentou-se ainda à fórmula do Quociente de Localização alguns extras, que permitem o refinamento da fórmula.

A primeira variante é a possibilidade de ponderação dos critérios de seleção, permitindo ao cliente segmentar o território atribuindo maior ou menor importância a determinados critérios. A fórmula que o permite:

$$QL_{rc} = p \left(\frac{V_c}{V_{rc}} / \frac{V_t}{V_{rt}} \right) \cap p \left(\frac{V_c}{V_{rc}} / \frac{V_t}{V_{rt}} \right) \cap p \left(\frac{V_c}{V_{rc}} / \frac{V_t}{V_{rt}} \right) \quad (0 \leq QL_{rc} \leq \infty) \quad (8)$$

sendo:

p - peso atribuído a cada critério

V_c - valor da variável por CP7

V_{rc} - valor da variável de relativização por CP7

V_t - valor total da área de estudo da variável

V_{rt} - valor total da área de estudo da variável de relativização

Outra variante é a ordenação dos CP7, como um *scoring*, ao número de recetáculos postais. Dado que muitas vezes o resultado da campanha na seleção de CP7 é superior ao comportável pelo cliente, os CP7 são eliminados da campanha até atingir o número de objetos pretendidos. Deste modo é possível dar prioridade aos CP7 com mais recetáculos postais, os mais densos. Em termos

operacionais apresenta-se também como uma vantagem ao permitir a otimização da distribuição por parte dos carteiros.

Além disto pode ainda ser considerada a concentração ou dispersão espacial dos CP7, permitindo customizar a campanha, alcançando uma população alvo concentrada ou vários núcleos dessa mesma população.

4.3.3. Aplicação do Modelo – Estudo Geocontacto

Mais uma vez será tomado como exemplo a *Associação Crescer* e replicado o estudo de Geocontacto *Plus* apresentado anteriormente para demonstração do modelo de geoprocessamento proposto e posterior comparação de resultados.

Para o estudo define-se novamente o público alvo como residentes seniores, segundo o INE população com idade superior a 64 anos, e residentes com idades compreendidas entre os 5 e os 9 anos, e a área geográfica é definida pelos concelhos de Torres Vedras, Alenquer, Sobral de Monte Agraço, Arruda dos Vinhos e Caldas da Rainha.

É necessário encontrar os CP7 com maior predominância do público alvo descrito. Para este processo recorre-se a *software Access*, onde se ligam duas tabelas e se corre uma *query*. As tabelas utilizadas dizem respeito a:

- Tabela DOL – Distribuição OnLine;
- Tabela com Quociente de Localização calculado para cada variável dos Censos, por CP7.

A tabela DOL – Distribuição OnLine é utilizada para obter a informação sobre CP4, concelho, freguesia e número de Pontos de Entrega Residencial (PER) por CP7 (dado que o número de objetos por campanha é definido pelo número de PE, neste caso residencial e não funcional, existentes nos CP7 alvo). Ao contrário do método atual, a ligação entre a DOL e a informação censitária é direta, dado que

o nível de análise é apenas a camada do CP7 e as subsecções estatísticas são descartadas.

Na metodologia atual o passo seguinte seria proceder ao cálculo do valor de corte. Na metodologia proposta isto não será necessário, dado que a base de dados Quociente de Localização, só por si, permite a seleção dos valores acima ou abaixo do valor de corte – 1, sendo este critério explicitado na *query*.

Com esta informação procede-se à construção da *query*, exemplificada na figura 12.

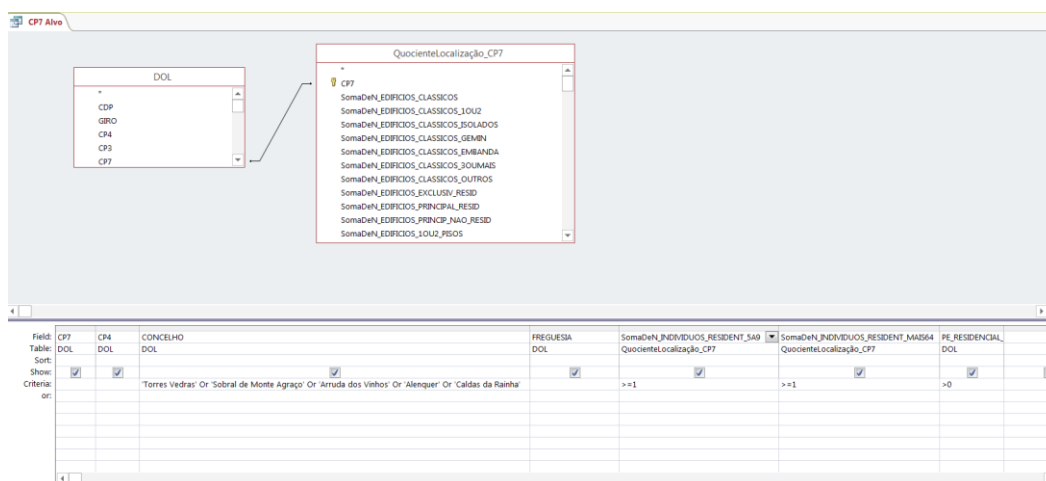


Figura 12 – Modelo Proposto: *query*, em software Access, para estudo Geocontacto

Daqui obtêm-se os CP7 alvo da campanha com informação sobre número de PER associados⁶. O resultado desta análise resulta na seleção de 505 CP7 alvo, equivalendo a 15 843 PER. Mais uma vez arredondando ao milheiro, o cliente adjudicará uma campanha de 16 000 objetos.

4.3.4. Análise de Resultados

Neste caso, a questão colocada anteriormente, na apresentação do modelo atual, não se coloca, dado que cada porta de cada CP7 é caracterizada por si só,

⁶ Por forma a manter a confidencialidade dos dados detidos pelos CTT, apresentar-se-á apenas os resultados finais.

independentemente das restantes, do centroide do CP7 e das áreas de influência. Após a caracterização de cada porta, procede-se à agregação por CP7. A caracterização de um CP7 tem assim em conta todas as suas portas, agregando informação de várias subsecções estatísticas.

Na área de estudo selecionada para representação cartográfica dos modelos são claras as alterações, tendo a maioria dos códigos postais de 7 dígitos sido expurgados da campanha (fig. 13).

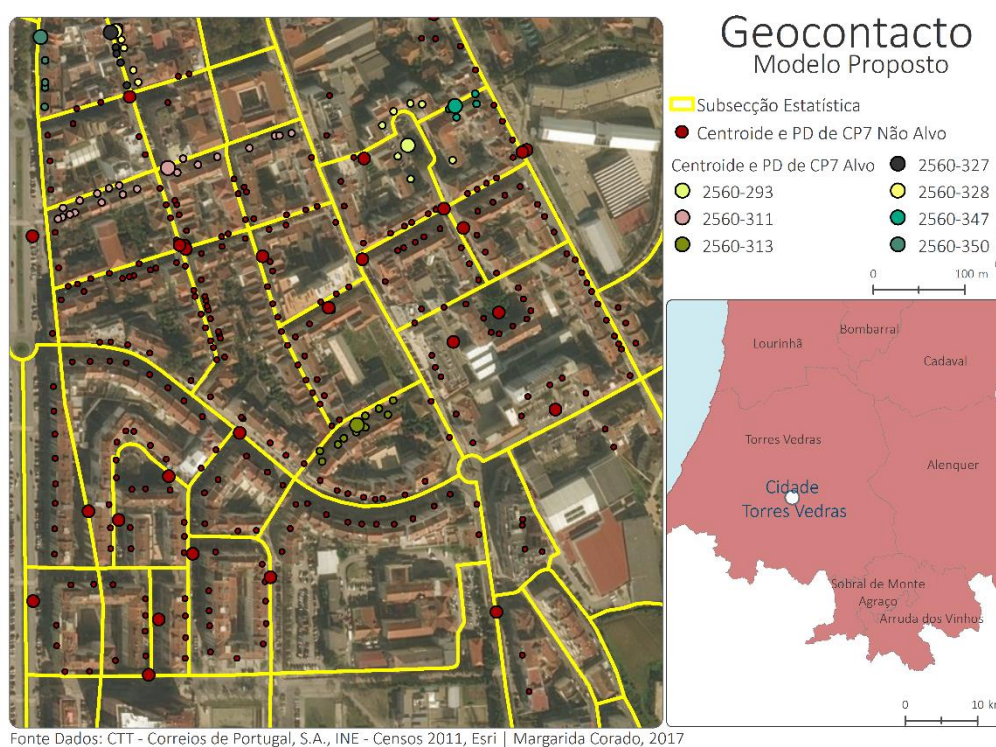


Figura 13 – Modelo Proposto: Exemplo de resultado de estudo Geocontacto

No capítulo seguinte procede-se à comparação dos resultados obtidos através dos dois métodos apresentados.

4.4. Comparação dos Modelos

A diferença entre os dois modelos é sonante. O modelo proposto tem a capacidade de refinar os resultados, por norma expurgando CP7 que não se enquadram no estudo e não deveriam ter sido selecionados. Não obstante, por vezes acrescenta códigos postais que o modelo atual havia deixado escapar.

No estudo desenvolvido, o modelo proposto expurga 986 códigos postais e acrescenta 80, reduzindo em cerca de 34 000 os Pontos de Entrega Residenciais, ou seja, os objetos necessários a produzir e distribuir.

A relação restrita um-para-um, entre centroide de CP7 e subsecção estatística é perdida e substituída pela relação um-para-muitos, entre uma subsecção estatística e várias portas. Não é assim viável proceder a uma comparação direta e literal entre os dois modelos através da representação cartográfica, dado que no modelo proposto o conceito de subsecção estatística se dissolve em informação censitária por porta. Porém, podemos comprovar, justificar e validar cartograficamente, o refinamento verificado entre os dois modelos. Para tal, à cartografia referente ao modelo proposto acrescenta-se a *layer* do Geo10, com a representação de cada porta por CP7. Deste modo, é possível a abstração do centroide (representado pelos círculos maiores) e área de influência e a análise verídica da existência de interseção, ou não, das portas/PD (representados pelos círculos menores) com as subsecções estatísticas alvo (fig. 14).

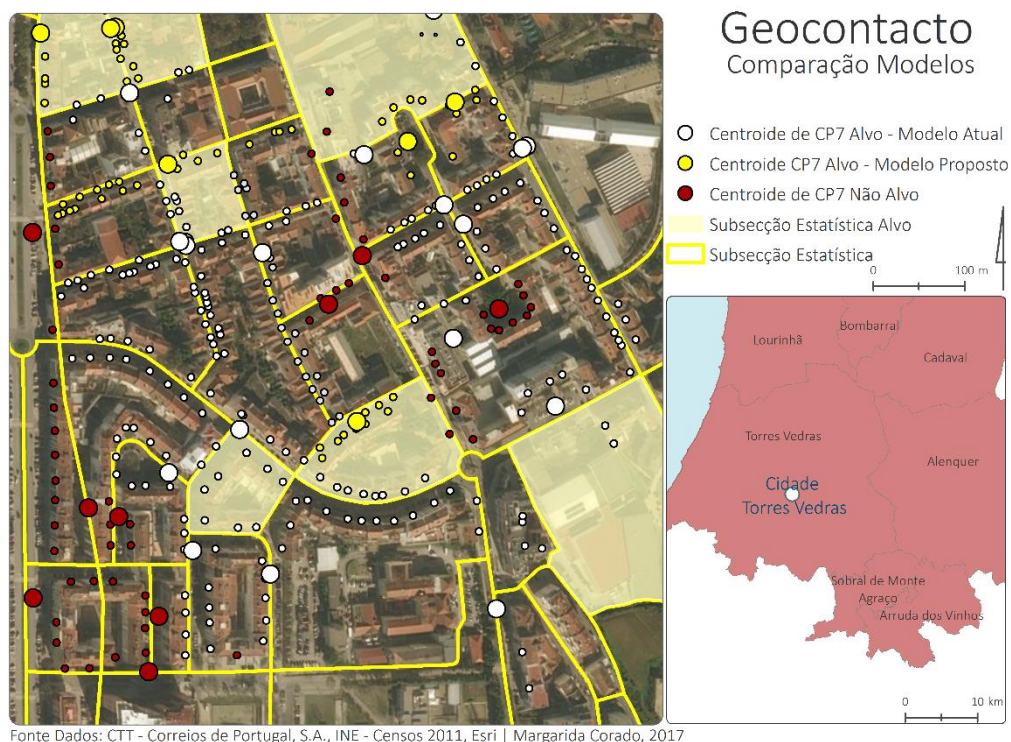


Figura 14 – Comparação de modelos através de representação de relação de Geo10 e Geoindex com Subsecções Estatísticas

Os CP7 a vermelho não foram selecionados em nenhum dos modelos e os brancos foram selecionados no modelo atual. A amarelo estão os CP7 selecionados apenas através do modelo proposto. Ou seja, todos os CP7 não amarelos, foram expurgados pelo modelo. Visualmente, podemos comprovar que realmente a maioria das portas que compõem estes CP7 não intersectam nenhuma subsecção alvo, pelo que são corretamente expurgados.

A precisão do modelo proposto permite que, mesmo que algumas portas de um código postal de 7 dígitos intersectem subsecções estatísticas alvo, este possa ser expurgado caso a expressão global do CP7 não esteja acima do Quociente de Localização. Ou seja, a relação espacial por si só não define os resultados obtidos, como se sucedia no modelo atual.

Da mesma forma, apesar de com menor expressão, alguns CP7 são acrescentados. Estes são CP7 cuja expressão de algumas portas numa subsecção estatística, em

relação aos restantes, permite que todo o código postal seja seleccionado para a campanha.

Procedeu-se agora à selecção de duas áreas aleatórias na área de estudo para análise do comportamento do modelo proposto em relação ao modelo atual, no estudo de Geocontacto em curso.

Na figura 15, o modelo proposto volta a seleccionar a maior parte dos CP7 já seleccionados anteriormente. Porém, elimina ainda um valor considerável de códigos postais. De ressaltar o facto de o modelo proposto acrescentar como alvo da campanha dois CP7 que anteriormente não eram considerados.

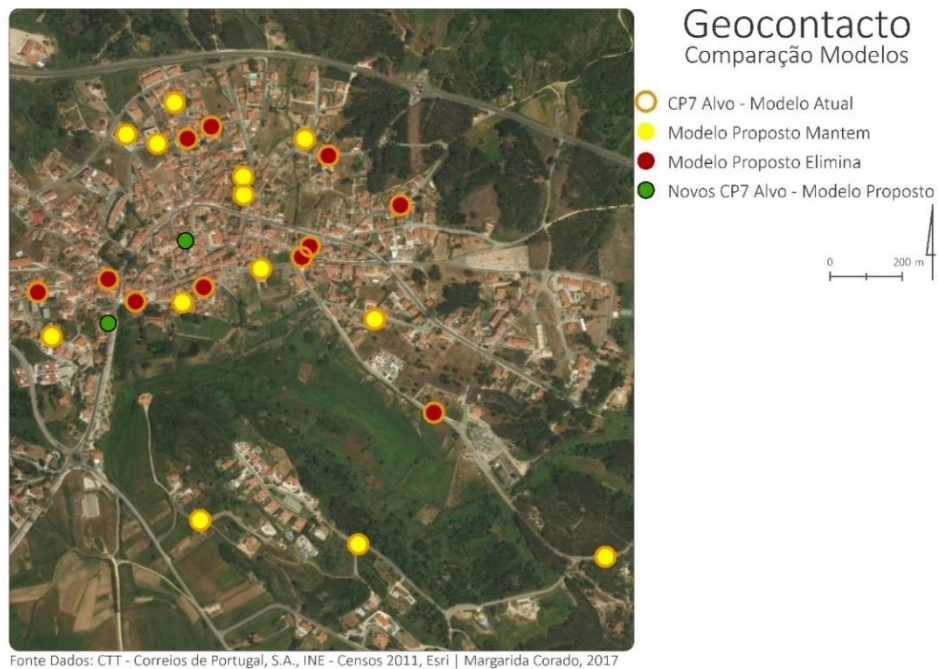


Figura 15 – Comparação de modelos – CP7 mantidos, eliminados e acrescentados pelo Modelo Proposto

O facto de na proximidade de um CP7 expurgado ser acrescentado um novo permite suportar a tese de que o modelo proposto refina a análise e, por conseguinte, os resultados obtidos.

Na segunda área aleatoriamente escolhida (fig. 16), o comportamento difere. Neste caso o modelo proposto valida todos os CP7 anteriormente seleccionados,

não expurgando nenhum. Não obstante, além dos CP7 já considerados, acrescenta 7 códigos postais como de interesse para a campanha.

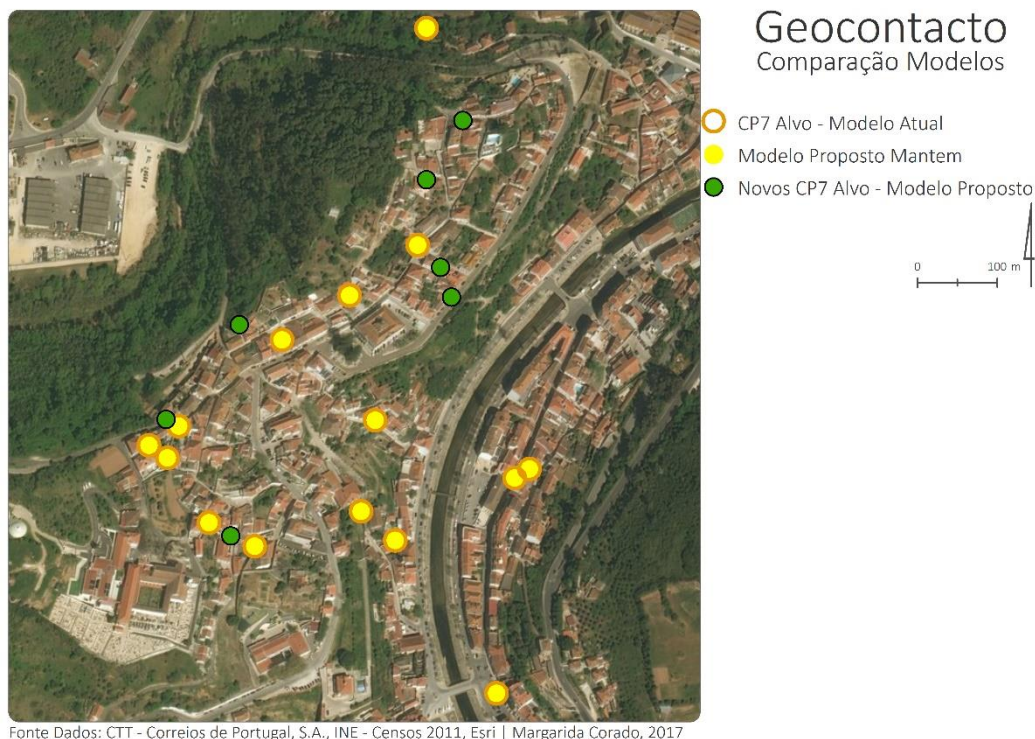


Figura 16 – Comparação de modelos – CP7 mantidos e acrescentados pelo Modelo Proposto

Esta breve análise cartográfica, e meramente visual, permite entender melhor o comportamento diferenciador do modelo proposto perante o modelo atual. Por norma a sua ação é de reafirmação dos CP7 selecionados, expurgando um ou outro esporadicamente, e refinando o detalhe da análise, permitindo a seleção de novos CP7 que até então escapavam ao filtro menos fino do modelo atualmente utilizado.

A exatidão/precisão dos modelos pode também ser avaliada através do cálculo da probabilidade em atingir o público alvo (tabela 1). A fórmula passa pela divisão dos casos favoráveis pelos casos possíveis, sendo adaptada para que se compare o segmento sociodemográfico em estudo pelo total de residentes.

No caso da modalidade Correio Não Endereçado (CNE), no qual seriam distribuídos objetos por todos os concelhos de interesse, foi dividido o total de população alvo pelo total de população residente nesses concelhos. A distribuição massificada,

por não recorrer a qualquer tipo de segmentação, apresenta a menor taxa de exatidão/precisão no público alvo – 25%.

Entre o modelo atual e o modelo proposto regista-se uma ligeira melhoria na taxa de exatidão. Apesar de ligeira, permite, mais uma vez, validar o refinamento e melhorias que o método proposto acrescenta a esta solução CTT.

	CNE	Modelo Atual	Modelo Proposto
Indivíduos – 5 a 9 anos	5,34	5,55	6,08
Indivíduos – superior a 64 anos	19,22	21,71	23,74
Total	24,56	27,26	29,82

Tabela 1 – Probabilidade em atingir público-alvo em cada modelo

Para o cliente, esta melhoria resulta numa maior facilidade em comunicar com o seu público alvo, bem com numa diminuição do desperdício de objetos entregues a público não-alvo. Poderá até, com o mesmo orçamento, estender a sua campanha a outros territórios e expandir o impacto da mesma.

4.5. Breves Conclusões

O modelo proposto apresenta-se como uma forma viável e fiável de melhorar a oferta de ações de *geomarketing* dos CTT.

A sua precisão e rigor superiores na seleção dos códigos postais de 7 dígitos são claros e, apesar da adoção do modelo proposto representar um decréscimo no valor de cada campanha, este foi validado e aceite positivamente no seio da empresa.

A modelação implica a representação abstrata de um sistema do mundo real, isto é, é um modo de simplificação de fenómenos. É assim expetável que o modelo desenvolvido não espelhe inteiramente a realidade, mas sim uma aproximação que se crê fidedigna.

O constrangimento que se revelou mais determinante relaciona-se com os dados. Primeiramente, as bases de dados provêm de diferentes fontes, com diferentes periodicidades, potenciando incompatibilidades. Porém, a incompatibilidade mais marcante reside na morfologia dos dados geográficos. Tendo, aliás, sido este o fator que originou este trabalho. Não obstante os esforços empreendidos, o resultado da relação entre o código postal de 7 dígitos e os Censos dificilmente corresponderá completamente à realidade.

O facto de a base de dados Geo10 se encontrar ainda em fase de qualificação e controlo de qualidade pode também representar um constrangimento ao modelo ideal, dado que vários registos tiveram de ser expurgados, como já explicitado. Deste modo, a informação censitária não foi rateada por todos os Pontos de Distribuição existentes, mas sim por aqueles validados, pelo que aquando da implementação formal do modelo, os resultados poderão diferir dos apresentados neste relatório.

Por último, o facto da informação censitária, informação alfanumérica que permite toda a caracterização sociodemográfica, de edificado e parque habitacional, ser atualizada apenas de 10 em 10 anos, apresenta-se como um défice profundo na fiabilidade da informação obtida pelo modelo. No que se refere a variáveis demográficas, esta questão pode ser fraturante, dado que quando se pretende realizar uma campanha cujo público alvo deverá ter determinada idade, são na realidade devolvidos CP7 onde há 10 anos esta situação se verificava, sendo que hoje essa população residente terá mais 10 anos que o pretendido.

O desafio de entender e resolver um problema de integração de dados geográficos, em concreto a integração da informação geográfica e alfanumérica do INE e os códigos postais dos CTT, foi devidamente encarado e o objetivo alcançado, tendo o modelo superado as expetativas existentes relativamente às melhorias possíveis no modelo atualmente em uso.

Através do aperfeiçoamento do atual modelo de geoprocessamento, os CTT terão a oportunidade de potenciar a inteligência geográfica que os SIG proporcionam, aumentando significativamente a sua capacidade de segmentação do território, a sua exatidão geográfica e concretizando uma nova abordagem à caracterização sociodemográfica e habitacional dos códigos postais. Os seus serviços de *geomarketing* poderão ser ainda mais qualificados e diferenciados de tudo o que existe no mercado, sendo o novo modelo de geoprocessamento encarado como uma vantagem competitiva.

5. CONCLUSÕES

Os CTT detêm um ativo relevante e crucial para ações de *geomarketing*, com potencial para crescerem como um *player* ainda mais interessante no mercado. Este ativo é trabalhado todos os dias e advém da atividade mais intrínseca da empresa, a distribuição postal. Através desta, todos os dias, milhares de carteiros apuram a mais ínfima informação sobre o território, elevando estes dados a informação geográfica. Tendo o projeto Geo10, o qual tem como objetivo georreferenciar e caracterizar todas as ruas, portas e domicílios a nível nacional, terminado o trabalho de campo no Continente, revelou-se necessário ultrapassar os constrangimentos existentes na integração dos dados geográficos provenientes de diferentes fontes utilizados nos estudos de *geomarketing*, desenvolvidos pela equipa de Soluções de Desmaterialização, Proximidade e *Business Intelligence*.

Foi assim desenvolvido um modelo de geoprocessamento e caracterização geográfica do território português, com recurso a dados provenientes do projeto Geo10 e informação censitária. Considera-se que o trabalho desenvolvido conseguiu ultrapassar os constrangimentos habituais aquando do trabalho com dados geográficos de diferentes fontes e capacitou os CTT de uma vantagem competitiva inigualável. O resultado do modelo pode ser sintetizado como o mapeamento dos dados censitários ao código postal de 7 dígitos, ou seja, a caracterização sociodemográfica e habitacional do código postal.

O trabalho desenvolvido não é um produto em si, mas sim a base de tratamento e processamento de dados e informação, o *background* que permitirá obter conhecimento de elevada qualidade e *profiling* de informação, passível de ser utilizado nos mais variados estudos. Através deste modelo, a capacidade de segmentação territorial, base de várias ações de *geomarketing*, foi substancialmente refinada, aprimorada e melhorada. Esta questão apresenta-se como o estímulo necessário para que novos produtos e serviços sejam desenvolvidos de forma a responder à procura crescente no mercado.

Também os estudos de Geocontacto sofreram uma análise minuciosa, tendo alguns processos sido reformulados. Um dos processos com maior impacto foi a reformulação do cálculo do Quociente de Localização para apuramento dos CP7 alvo, permitindo a otimização e diminuição de propensão para erros na elaboração de estudos de Geocontacto.

Além disto, foi proposta a possibilidade de customização do público-alvo a atingir prioritariamente, com a possibilidade de ordenação dos CP7 selecionados segundo a densidade de recetáculos postais ou considerando a concentração ou dispersão espacial da população. Foi também sugerida a possibilidade de, aquando da segmentação segundo mais do que um variável, atribuir pesos diferentes a cada critério. É assim possível oferecer ao cliente uma campanha com uma customização superior.

Na ótica do cliente, este aperceber-se-á que terá campanhas mais eficazes, com melhores taxas de resposta dos seus consumidores e poderá criar campanhas com maior alcance, com os mesmos recursos. Para os CTT haverá uma melhoria clara na qualidade dos seus produtos e serviços, alcançará uma vantagem competitiva potenciada pelas suas próprias bases de dados e *know-how*, obterá resultados mais refinados e que permitirão novos produtos.

Mais uma vez se salienta o facto de se tratar de um modelo, não sendo uma representação 100% fiel da realidade, apresenta algumas limitações já referidas.

Como perspetivas futuras, espera-se a disponibilização dos dados Geo10 a nível nacional, de forma a implementar o modelo a todo o território português, e não apenas à região Oeste. Infelizmente, não foi possível em tempo útil de estágio. Não obstante, precavendo as limitações de tempo, todo o processo foi automatizado para que o trabalho possa ter continuidade.

Numa apresentação formal a colaboradores dos CTT, nomeadamente das Direções de Publicidade e Soluções Empresariais e Estratégia e Desenvolvimento Corporativo, foi demonstrado o interesse em implementar o novo modelo de

geoprocessamento, tendo sido reconhecido o seu valor para a empresa, bem como as limitações do modelo atual e potencial para atingir um nível superior de excelência.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- BARRADAS, S., & FERRÃO, F. (1996). *Código Postal*. Lisboa: ISEGI.
- BRITO, A. (2013). CTT: uma empresa onde se lê a história do país. *Público*. Obtido em 24 de Março de 2017, de <https://www.publico.pt/2013/12/04/economia/noticia/ctt-uma-empresa-onde-se-le-a-historia-do-pais-1614866>
- CARDOSO, C. E. (2011). *Geomarketing como Suporte de Decisão em Gestão do Território*. Lisboa: Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa.
- CARVALHO, R., & MARCELINO, I. (2015). Todo o potencial de negócio à distância de um mapa. *Diário Económico*, 2-4.
- CTT - Correios de Portugal, S.A. (2012). Desenvolver competências para melhores desempenhos. *17º Encontro Nacional de Formadores & Coaches*. Lisboa: Associação Portuguesa dos Gestores e Técnicos dos Recursos Humanos (APG). Obtido em 31 de Março de 2017, de http://www.apg.pt/downloads/file414_pt.pdf
- CTT - Correios de Portugal, S.A. (2015). Soluções Empresariais.
- CTT - Correios de Portugal, S.A. (2017). *Missão, Visão e Valores*. Obtido de CTT: <https://www.ctt.pt/ctt-e-investidores/a-empresa/estrategia-empresarial/missao-visao-valores.html>
- CTT - Correios de Portugal, S.A. (s.d.). *Perguntas Frequentes*. Obtido em 31 de Março de 2017, de Geindex dos CTT: <http://geindex.ctt.pt/>
- EAGLESON, S., ESCOBAR, F., & WILLIAMSON, I. (2003). Automating the administration boundary design process using hierarchical spatial reasoning theory and geographical information systems. *International*

Journal of Geographical Information Science, 99 - 118. Obtido de <http://dx.doi.org/10.1080/713811748>

FLOWERDEW, R., & GREEN, M. (1992). Development in areal interpolation methods and GIS. *Annals of Regional Science*. 26, pp. 67 - 78. Springer Berlin Heidelberg.

Fundação Portuguesa das Comunicações. (2012). Vencer a Distância - Cinco Séculos dos Correios em Portugal. *Vencer a Distância - Cinco Séculos dos Correios em Portugal*. Portugal. Obtido em 24 de Março de 2017, de <http://www.fpc.pt/Portals/0/PDF%20Exposicoes/Percurso%20de%20Correios.pdf>

GEIRINHAS, J. (2001). Conceitos e Metodologias: BGRI - Base Geográfica de Referenciação de Informação. *Revista de Estudos Regionais - Região de Lisboa e Vale do Tejo*, pp. 67 - 73. Obtido em 31 de Março de 2017

HAWLEY, K., & MOELLERING, H. (2005). A Comparative Analysis of Areal Interpolation Methods. *Cartography and Geographic Information Science*, 32, 411 - 423. Obtido de <http://dx.doi.org/10.1559/152304005775194818>

Instituto Nacional de Estatística. (2002). Conceitos e Metodologias: Medidas de Especialização Regional. *Revista de Estudos Regionais - Região de Lisboa e Vale do Tejo*, pp. 65 - 72. Obtido em 6 de Abril de 2017, de http://idinis.weebly.com/uploads/5/6/3/9/5639534/medidas_de_especializacao_ine.pdf

Instituto Nacional de Estatística. (2011). *Censos 2011*. Obtido em 31 de Março de 2017, de https://www.ine.pt/publicacoes/Folheto_censos_portal/index.html

JACOX, E. H., & SAMET, H. (Novembro de 2006). Spatial Join Techniques. *ACM Transactions on Database Systems*, V, pp. 1 - 45. Obtido de <http://www.cs.umd.edu/~hjs/pubs/jacoxtds07.pdf>

Obtido

de

<http://www.tandfonline.com/doi/full/10.1080/13658810600924211#aHR0cDovL3d3dy50YW5kZm9ubGluZS5jb20vZG9pL3BkZi8xMC4xMDgwLzEzNjU4ODEwNjAwOTI0MjExP25lZWRBY2Nlc3M9dHJ1ZUBAQDA=>

SMITH, M. J., GOODCHILD, M. F., & LONGLEY, P. A. (2015). *Geospatial Analysis - A Comprehensive Guide to Principles, Techniques and Software Tools* (5ª ed.). Winchelsea, UK: The Winchelsea Press.

VOSS, P. R., LONG, D. D., & HAMMER, R. B. (1999). *When census geography doesn't work: using ancillary information to improve the spatial interpolation of demographic data*. Center for Demographic and Ecology, University of Wisconsin - Madison.