

Desenvolvimento de um sistema de apoio à decisão para gestão de uma distribuidora de recargas de dispositivos móveis utilizando análise por envoltória de dados

RESUMO

Leonardo André Colares Dantas
leonardoandreh@gmail.com
Universidade Federal do Ceará (UFC),
Fortaleza, Ceará, Brasil

Heráclito Lopes Jaguaribe Pontes
hjaguaribe@ufc.br
Universidade Federal do Ceará (UFC),
Fortaleza, Ceará, Brasil

Jônatas Breno Gomes Amarante
jonatasbrenon@gmail.com
Universidade Federal do Ceará (UFC),
Fortaleza, Ceará, Brasil

Levi Ribeiro de Abreu
leviribeiro@alu.ufc.br
Universidade Federal do Ceará (UFC),
Fortaleza, Ceará, Brasil

Ivna Praciano Teixeira
ivnapraciano@gmail.com
Universidade Federal do Ceará (UFC),
Fortaleza, Ceará, Brasil

A indústria de telecomunicações vem apresentando dificuldades nos últimos anos para conseguir elevar sua competitividade. Tendo em vista este fato, muitas empresas têm adotado medidas de corte de gastos. Assim, partindo-se da necessidade de melhorias de performance e assertividade na alocação dos recursos, foi avaliada a possibilidade de se implementar um sistema de apoio à decisão com auxílio da metodologia Análise por Envoltória de Dados (*Data Envelopment Analysis- DEA*) em uma empresa de distribuição de recargas pré-pagas situada na cidade de Fortaleza. Através do uso de ferramentas de análises de dados em *Big Data* foram mapeadas as fontes de dados necessárias para o trabalho, bem como avaliada a qualidade dos dados recebidos. A funcionalidade do sistema criado foi validada através da aplicação prática em uma base mensal de uma operadora, onde foi possível obter uma redução de 19,15% dos pontos de vendas (PDV's), sem impactar significativamente o faturamento total, além de prover recursos visuais permitindo análise dos PDV's sob óticas de diferentes objetivos.

PALAVRAS-CHAVE: Processo ETL. *Dashboard*. *Big Data*. Pesquisa Operacional. Teoria da Decisão.

INTRODUÇÃO

A indústria de telecomunicações vem apresentando dificuldades nos últimos anos para conseguir elevar seu nível de vendas. Dessa forma, grandes empresas do mercado têm adotado uma forte estratégia de redução de custos, promovendo cortes em diversos segmentos do negócio na tentativa de alavancar sua margem.

De modo semelhante, as distribuidoras responsáveis pelas recargas de aparelhos celulares têm sofrido grandes impactos, visto o baixo percentual que é repassado pelas operadoras para as distribuidoras como fonte de receita das recargas realizadas, o que implica na dificuldade de conseguir manter-se saudável no mercado.

Para fins práticos, também deve se levar em consideração a possibilidade que algumas operadoras oferecem de programas de excelência, os quais dão oportunidade ao distribuidor de aumentar a margem recebida, caso atinja certos critérios pré-estabelecidos pelas operadoras, sendo um dos critérios a capacidade, por meio da quantidade de pontos de vendas, para atingir a população como um todo.

Sendo assim, na busca por tornar suas operações mais eficientes, as distribuidoras de recarga estão cada vez mais procurando se modernizar e assim fazer uso de novas tecnologias que possam impactar suas decisões de forma positiva, trazendo embasamento científico para a gestão de suas operações.

A crescente massa de dados que são trabalhadas diariamente pelas empresas trazem consigo imensas oportunidades de alavancar as decisões relacionadas ao negócio, podendo oferecer vantagem competitiva (TAURION, 2013). A difusão de tecnologias *Big Data* ainda é assunto relativamente recente no Brasil e pouco é utilizado quando restringido ao contexto de pequenas e médias empresas. Desta forma, busca-se explorar estas aplicações dentro de área de Pesquisa Operacional (PO).

O problema proposto no trabalho foi baseado na intenção de contribuir para o aumento da lucratividade das empresas de distribuição de recarga, responsáveis pela alocação dos pontos de venda, atacando componentes de custos do negócio.

Este estudo se justifica pela necessidade de apresentar embasamento científico para analisar a performance dos pontos de vendas, atendendo a demanda existente, com o menor custo possível, alavancando oportunidades frente a concorrência. Pouco foi reportado na literatura sobre esse problema, considerando os métodos implementados e avaliando dados de várias bases diferentes como o Instituto Brasileiro de Geografia e Estatística (IBGE) e a prefeitura do município, no qual foi realizado o estudo.

O objetivo do trabalho consiste em conceber um modelo de apoio à decisão espacial, para avaliar a eficiência dos pontos de venda através da Análise por Envoltória de Dados (DEA), visando aprimorar a performance geral do sistema, através de melhor uso dos recursos disponíveis. Para o tratamento das extensas quantidades de dados, foram utilizadas ferramentas de análise de *Big Data*, como a extração, transformação e carregamentos de dados (ETL).

O sistema de apoio à decisão consiste em utilizar a ferramenta de DEA para mensurar a eficiência relativa dos pontos de venda de acordo com os recursos alocados para cada um. Para assim, diagnosticar possíveis problemas relacionados a essa estrutura e compreender onde e como esses recursos podem ser melhor utilizados.

O trabalho é composto por mais quatro seções. A segunda ilustra a fundamentação teórica, a terceira descreve a metodologia utilizada, a quarta demonstra o estudo de caso com os dados coletados e, por fim, a última seção compõe as conclusões e propostas de estudos futuros.

FUNDAMENTAÇÃO TEÓRICA

Para nortear os conceitos ilustrados no estudo de caso, faz-se necessário definir os seguintes conceitos: Teoria da decisão, Big Data, processos ETL e Análise por envoltória de dados.

TEORIA DA DECISÃO

Hansson (1994) define a teoria de decisão como sendo a teoria sobre como as decisões deveriam ser tomadas. Estas decisões irão seguir normas de racionalidade, sendo as principais regras que o decisor deve optar para a tomada de decisão.

Para Tsoukiàs (2006), cada diagnóstico, representado por um estado natural, está associado com uma probabilidade e cada tratamento, consistindo de ações potenciais, está associado a consequências. Diante disso, deve-se construir uma função de utilidade do decisor (cliente), pela qual busca-se medir a preferência dos indivíduos, sendo esta essencial para aplicação da teoria da decisão. Diante desta função, busca-se sua maximização que permitirá identificar a melhor solução dentro do cenário proposto. Esta solução irá maximizar o valor esperado pelo decisor.

Segundo Savage (1972), a existência dessa função é garantida por uma série de axiomas que enunciam aquilo que na teoria iria constituir o comportamento racional do agente decisor. Sendo assim, no contexto normativo o decisor deve adaptar seu comportamento aos axiomas estabelecidos do contrário não será considerado racional.

Como o presente artigo consiste em decidir sobre a eficiência de determinados pontos, localizados em diferentes posições geográficas, faz-se o uso de Sistemas de Apoio à Decisão Espacial (SADEs).

Para Denshaw (1991), SADEs foram concebidos para prover ao usuário do sistema, um ambiente relacionado a tomada de decisão que permita realizar análises que envolvam informações geográficas de uma maneira flexível. Com isso, são sistemas de informação destinados a auxiliar decisões baseadas em dados geográficos (posição, geometria e atributos).

Neto e Rodrigues (2001) mostraram que, nas principais estratégias praticadas de desenvolvimento de sistemas SADE, o Sistema de Informação Geográfica (SIG) é considerado o subsistema principal e a modelagem

matemática científica, o subsistema secundário. De acordo com os autores, as diferentes estratégias variam na proximidade lógica e física entre SIG e *softwares* de modelagem científica, na forma de transação de dados entre si e na proximidade lógica e física do subsistema de integração com os dois subsistemas integrados por este.

Segundo Rafaeli Neto (2004), o SADE, como um sistema de apoio a decisão, pode atuar como repositório de informação sobre o sistema em análise, além de prover ferramentas para que o decisor explore, manipule e apresente os dados de interesse.

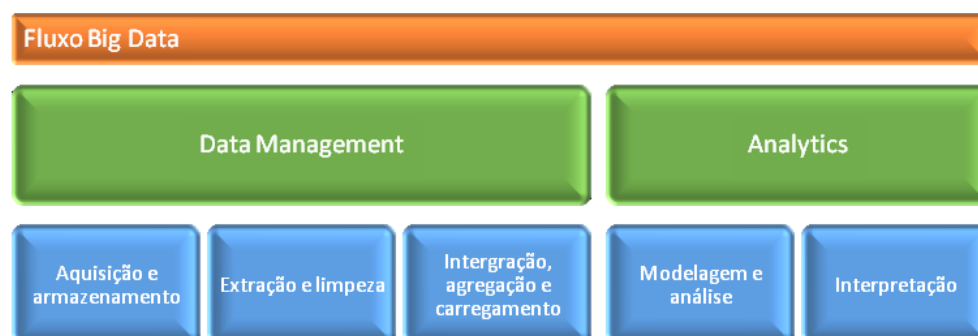
BIG DATA

Atualmente, o conceito *Big Data* vem se difundindo dentro do mercado de trabalho e comunidade acadêmica, porém sua origem ainda é bastante recente. Para Gartner (2012), *Big Data* pode ser definido como sistemas informacionais que apresentam alto grau de volume, velocidade e variedade e necessitam de ferramentas e abordagens eficientes para processar os dados e extrair insights que irão auxiliar na tomada de decisão organizacional.

De acordo com Gandomi *et al.* (2015), as estruturas dos dados podem ser classificadas em 3 categorias: estruturados, semi-estruturados e não-estruturados. Os dados estruturados referem-se às planilhas e bases de dados relacionais. Já os dados não-estruturados são representados por imagens, vídeos, áudios e textos, sendo as categorias de dados mais complexas para trabalhar apesar do recente desenvolvimento de técnicas computacionais para processar tais informações. Por último, os dados do tipo semi-estruturado como próprio nome já diz, apresentam certo grau de padronização, facilitando o acesso às informações contidas em seu meio.

O potencial do *Big Data* é percebido quando o processo de tomada de decisão é alavancado através da sua análise. Cada vez mais as empresas estão buscando meios eficientes de transformar grandes e variados volumes de dados em poderosos *insights*. Desta forma, Labrinidis e Jajadish (2012) consideram cinco fases principais como base para uso do Big Data no processo de tomada de decisão, sendo estas fases subdivididas em 2 grupos: *Data Management* e *Analytics*. Na Figura 1 estas fases são apresentadas em seus respectivos grupos.

Figura 1 – As 5 fases principais do *Big Data*



Fonte: Adaptação Gandomi *et al.* (2015)

Para Gandomi *et al.* (2015), *Data Management* envolve as etapas de aquisição e armazenamento de informações que irão antecipar a transformação dos dados, removendo inconsistências e estruturando a base para ser utilizado na preparação de modelos e análises.

No caso de *Analytics*, os mesmos autores definem como sendo técnicas usadas para analisar os dados de forma a extrair insights que possam ser utilizados para gerir os negócios de forma mais inteligente. Este último é onde encaixa-se o termo *Big Data Analytics*.

PROCESSO ETL

De acordo com Abreu (2008), a sigla ETL tem origem na língua inglesa, sendo melhor traduzida por “*Extract, Transform and Load*”. Esta sigla representa um processo baseado em ferramentas que se destinam a extração, transformação e carregamento de dados.

Os dados trabalhados dentro deste processo podem vir das mais diversas bases, desde simples planilhas como grandes repositórios e sistemas. Quanto ao destino dado a estes dados, a carga pode ser direcionada a bancos de dados, sistemas de informação, *data warehouses* ou podem ser utilizados como carga para outros modelos e sistemas de inteligência de negócios.

Segundo Date (2000) o ETL é utilizado como processo primordial em projetos de migração de dados para sistemas de informação, *business intelligence* e aplicações de *data warehouse*. Atualmente sua utilização vem crescendo na área de análise de dados, com a etapa de transformação sendo a principal para alimentação de modelos estatísticos e de mineração de dados no intuito de extrair insights que auxiliem os gerentes e tomadores de decisão no exercício de suas funções.

Segundo Cielo (2010), o primeiro passo dentro do processo ETL é simplesmente a definição das fontes de dados e a extração delas. As origens podem ser várias e em diferentes formatos, onde poderemos encontrar desde os sistemas transacionais das empresas até simples planilhas e arquivos textos.

Nesta fase, é necessário identificar o tipo, forma de armazenamento, estrutura e modelagem dos dados a serem extraídos, além da necessidade de viabilizar através da ferramenta de extração um meio de acesso a estes dados de origem.

A segunda fase é o processo responsável pelo tratamento e transformação dos dados. Após o processo de extração, caso os dados não se encontrem em conformidade, deve-se realizar diversos procedimentos a fim de normalizar os dados e torná-los aptos a utilização na carga que será dada (CIELO, 2010).

Na obtenção dos dados em fontes muitas vezes desconhecidas ou que foram gerenciadas por sistemas de informação antigos, podendo existir falhas no projeto ou sem a utilização de um sistema gerenciador de banco de dados adequado, é comum encontrar problemas de integridade referencial ou inconsistências, como datas inválidas, atributos obrigatórios não preenchidos, somatórios numéricos inconsistentes, falta de normalização e diversos outros problemas (CIELO, 2010).

A última fase é constituída pela etapa de carga ou carregamento que deverá copiar os dados extraídos, tratados e manipulados nas etapas anteriores, em bancos de dados de destino ou em planilhas e outras extensões que possam vir a ser utilizadas pelos usuários (ABREU, 2008).

De acordo com a necessidade dos usuários finais, a fase de carga poderá ser realizada uma única vez ou de forma periódica para atualização de dados, como por exemplo em projetos de inteligência de negócios (ABREU, 2008).

ANÁLISE POR ENVOLTÓRIA DE DADOS

A Análise por Envoltória de Dados ou *Data Envelopment Analysis* (DEA) é vista como uma técnica de programação matemática, fazendo parte do conjunto de métodos encontrados na área de Pesquisa Operacional, que possibilita a análise do grau de eficiência de múltiplas unidades produtivas, chamadas de *Decision Making Units* (DMU's), nas quais avalia e compara os insumos com os produtos gerados dentro de um sistema.

Banker (1993) declara que a técnica DEA foi concebida como uma sistemática de programação matemática para análise da eficiência relativa de unidades comparativas com processos de produção similares.

O conceito básico por trás desta técnica está ligado à comparação de eficiências entre unidades semelhantes e suas efetivas operações, desta forma não faz uso de uma abordagem que busca um ideal ou produção ótima.

Cooper *et al.* (2000) cita particularidades da DEA, além de suportarem sua utilização como ferramenta de apoio à decisão e avaliação de cenários *what-if*, sendo capaz de modelar situações diversas encontradas no mercado:

- a) dados são valores reais que representam interesses dos analistas e gerentes e os valores devem ser positivos para cada DMU;
- b) as unidades de medida das diferentes entradas e saídas não precisam ser congruentes. Exemplos podem envolver número de pessoas, espaços, dinheiro gasto, etc.
- c) possui habilidade de identificar fontes e quantidades de ineficiência em cada *input* e *output* de cada entidade;
- d) possui capacidade de identificar membros de referência no conjunto eficiente para assim avaliar e identificar fontes de ineficiências.
- f) reconhece a probabilidade de que entidades *outliers* não representem somente desvio sem relação ao comportamento médio do conjunto, mas possíveis referências para benchmark a serem estudados pelas demais DMU's.

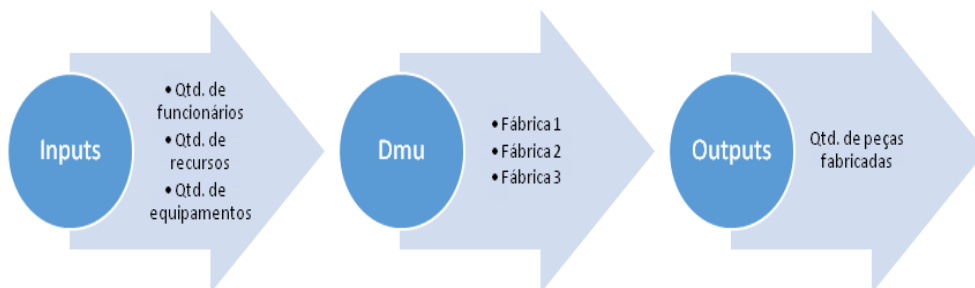
De acordo com Cooper *et al.* (2000), a programação matemática utilizada no modelo de DEA é composta basicamente de três elementos básicos:

1. **A função-objetivo:** Função linear de variáveis de decisão, que deve ser otimizada (maximizada ou minimizada).
2. **Funções Restrições:** Tratam das relações de interdependência entre as variáveis de decisão, sendo expressas por um conjunto de equações e/ou inequações lineares.

3. Variáveis do modelo: Deverão assumir valores não-negativos.

É importante notar que os itens que irão compor as aplicações DEA devem ser detalhados, de modo que cada DMU possua os mesmos insumos, referentes aos recursos empregados na produção, e os mesmos produtos, referentes à produção gerada, diferenciando-se em suas quantidades, mas sendo similares em sua natureza (FERREIRA GOMES, 2009). A Figura 2 mostra um exemplo básico de sistema com seus respectivos *inputs* e *outputs*.

Figura 2 – Exemplo geral de sistema



Fonte: Autores

A metodologia DEA permite, então, uma análise da eficiência de várias DMU's que possuem um ou mais *inputs* e/ou *outputs*, por meio da construção de uma fronteira de eficiência, linear por partes. As entidades que possuem melhor taxa "Output/Input" serão as DMU's consideradas mais eficientes dentro do conjunto analisado e irão estar situadas na fronteira, enquanto as ineficientes estarão situadas em uma região abaixo desta fronteira, denominada de envoltória convexa (SEIFORD, 1999).

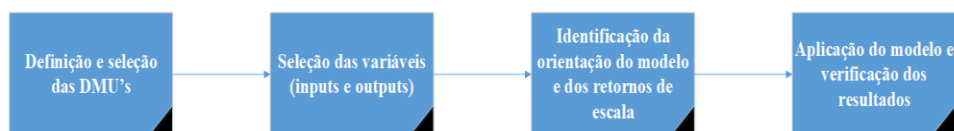
Sendo assim, cabe ressaltar a forma mais utilizada para quantificar a eficiência, mediante a razão entre a quantidade gerada de produtos e a quantidade utilizada de insumos, conforme ilustra a Equação 1 e considerando os ambientes complexos em que as organizações estão inseridas.

$$Eficiência = \frac{\sum_r u_r y_r}{\sum_i v_i x_i} \quad (1)$$

Na equação, u_r e v_i são pesos, ou seja, o grau de importância que a empresa atribui a quantidades y_r de *output* r e x_i de *input* i .

Segundo Meza (1998), na modelagem DEA devem-se seguir 4 etapas para implementar o problema. As etapas são ilustradas pela figura 3.

Figura 3 - Esquema da implementação de um modelo DEA



Fonte: Autores

O primeiro passo consiste na definição das unidades tomadoras de decisão, através do banco de dados é definido quais são as informações semelhantes que podem ser comparadas e traduzidas em alguma eficiência. Além disso, as DMU's devem ser homogêneas, possuindo mesmos insumos, referentes aos recursos

empregados na produção, e os mesmos produtos, referentes à produção gerada, que estejam trabalhando nas mesmas condições de mercado, diferenciando-se em suas quantidades, mas sendo similares em sua natureza (COOPER et al., 2004).

O segundo passo é definir quais as variáveis de *inputs* e *outputs* não sendo necessário ter atenção às unidades de medidas utilizadas, que podem ser das mais variadas possíveis (COOPER et al., 2004).

Segundo Lins e Meza (2000), quanto maior o número de variáveis em relação ao número de DMU's, mais difícil será o processo de ordenação pelas eficiências, visto a tendência de várias DMU's acabarem sendo posicionadas na fronteira de eficiência. Desta forma, deve-se definir quais serão as variáveis pertinentes para a análise e quais são dispensáveis, de modo que o modelo continue descrevendo fielmente a realidade.

O terceiro passo é a escolha da orientação do modelo e dos retornos de escala. A orientação consiste em definir como os *benchmarks* das unidades ineficientes são determinados na projeção da fronteira de eficiência. Existem basicamente dois tipos, que são selecionados dependendo das características do problema, que são: orientação a *inputs*, quando a eficiência é atingida por uma redução proporcional de entradas, mantidas as saídas constantes; e orientação a *outputs*, quando se deseja maximizar os resultados sem diminuir os recursos (COOPER et al., 2004).

A relação entre *inputs* e *outputs* é denominada retorno de escala. Segundo Mello et al. (2004) existem dois tipos básicos de modelos, conhecidos como retorno de escala constante e retorno de escala variável, eles são conhecidos também pelo nome de seus autores Charnes, Cooper e Rhodes (CCR) e Banker, Charnes e Cooper (BCC); respectivamente.

O modelo CCR tem como propriedade principal a proporcionalidade entre *inputs* e *outputs* na fronteira, ou seja, o aumento (decremento) na quantidade dos *inputs* provocará acréscimo (redução) proporcional no valor dos *outputs*. Já o modelo BCC é invariante à translação a *outputs* quando é orientado a *inputs* e vice-versa, além disso, a DMU que tiver o menor valor de um determinado *input* ou o maior valor de certo *output* será eficiente (MELLO et al., 2004).

O último passo é a execução do modelo, utilizando os dados já pré-processados e com as informações de *input* e *output*. Por fim, com as eficiências relativas de cada DMU's é possível analisar a qualidade do modelo gerado verificando se ele represente a realidade ou não, bem como comparar as DMU's eficientes e verificar o as ineficientes precisam melhorar para se tornarem eficientes (COOPER et al., 2004).

METODOLOGIA

A metodologia utilizada para a coleta, tratamento, utilização dos dados e a modelagem por DEA para a geração de informações, foi baseada na metodologia *Agile* como estratégia de desenvolvimento do estudo.

A metodologia de gerenciamento de projetos *Agile*, proposta por Larson e Chang (2016), possui diversas aplicações para o gerenciamento de projetos

relacionados a tecnologia e inovação, representando uma boa alternativa para ser aplicada em projetos relacionados a *Big Data*.

Segundo Beck *et al.* (2001), a metodologia possui os seguintes princípios: interações sobre processos e ferramentas de forma individualizada, construção de *softwares* com documentação compreensível e a colaboração do cliente para eventuais mudanças no contrato. As cinco etapas que constituem o método *Agile* são descritas a seguir:

1. **Descoberta:** realização das atividades de levantamento e brainstorming de ideias, mapeamento das fontes de dados necessárias e o processo de *data profiling*, no qual busca-se uma compreensão da qualidade dos dados e sua estrutura. Nessa etapa, foram coletadas informações com os principais gerentes sobre o problema em questão, além de consultadas importantes base de dados do IBGE.
2. **Arquitetura:** execução de três atividades principais, começando pelo mapeamento das ferramentas utilizadas, descrição de escopo base do projeto e a prova de conceito.
3. **Concepção/Desenvolvimento:** descrição do modelo criado, dividido pelo processo ETL, aplicação de análise por envoltória de dados e metodologia para remoção de pontos de vendas. Apresentado toda a sistemática utilizada para construção bem como as considerações a serem realizadas.
4. **Teste/Produção:** criação de métricas para avaliação de desempenho da solução e disposição dos outputs gerados pelo modelo.
5. **Aplicação Prática:** aplicação visando a redução do número de pontos de vendas. Análise do cenário atual de uma distribuidora de recarga e sugestão de solução a partir do nível de faturamento realizado. Por fim, são comparados os resultados.

ESTUDO DE CASO

A empresa em estudo atua no segmento de distribuição de recargas pré-pagas por meio da prestação de serviços em tecnologia da informação. A mesma participa em mais de dois bilhões de transações de recarga ao ano, também desenvolve *softwares* que permitem o levantamento e o controle de dados em toda a cadeia de distribuição de recarga de celulares pré-pagos. Desta forma, ela busca oferecer às operadoras e distribuidoras, seus clientes diretos, serviços que as auxiliem a melhorar a gestão de seus negócios.

Sendo assim, esta empresa opera com um grande volume de dados gerados por cada transação de recarga feita, buscando formas de analisá-los e processá-los e criando modelos de soluções que possam ser implementadas pelos seus clientes.

A problemática abordada neste projeto derivou-se de uma situação de engenharia reversa, na qual após visto que tipos de dados e objetivos a empresa têm, analisa-se o que poderia ser desenvolvido com estes mesmos dados.

ETAPAS DO ESTUDO

O estudo foi iniciado com a escolha do tema a ser abordado, o qual consiste em conceber um sistema de apoio à decisão espacial que possa ser utilizado pela companhia para propor serviços de análise de dados georreferenciados relacionado a gestão dos pontos de vendas de distribuidores de recargas e operadoras e, posteriormente, realizar a aplicação da metodologia DEA nos Pontos de Vendas (PDV's).

É importante ressaltar que visto a necessidade de envolver paralelamente áreas de negócio e desenvolvimento, buscando uma metodologia eficiente de projeto, foi adotado uma estratégia de desenvolvimento *Agile*.

Feito isso, foi dado início ao desenvolvimento do sistema de apoio a decisão para problemas georreferenciados, visto que no modelo atual esse tipo de problemática era tratado a partir de esforços manuais e sem muitos recursos para análise espacial.

Desta forma, um novo modelo abrangendo todas as etapas de extração, processamento, tratamento e carregamento dos dados com a utilização da análise por envoltórias de dados para priorização de ações em bairros, foi desenvolvido dentro de um mesmo ambiente. O estudo foi dividido, com base na metodologia *Agile*, em 5 etapas já destacadas na metodologia do estudo.

DESCOBERTA

Esta etapa consiste na realização de três etapas: o levantamento e *brainstorming* de ideias, já válido no início do projeto, o mapeamento das fontes de dados necessárias e o processo de *data profiling*.

O mapeamento foi realizado com o auxílio do gerente da área, que visto o levantamento das demandas, foi traçado as informações necessárias para a execução do projeto e então as bases nas quais encontravam-se estas informações.

Sendo assim, as bases utilizadas para o processo resumem-se basicamente à 3 fontes distintas: Interno, com dados relacionados aos pontos de vendas, suas coordenadas geográficas e faturamento diário; Instituto Brasileiro de Geografia e Estatística (IBGE), com dados relacionados à censo demográfico; Portal Fortaleza Dados Abertos, com dados relacionados aos bairros de Fortaleza e seus polígonos para visualização espacial.

Após mapear as bases que seriam utilizadas, foi realizada a atividade de *data profiling* necessária para entender melhor a qualidade das métricas que compunham as bases. Nesta etapa, as bases foram todas analisadas a fim de que se pudesse obter dados mais bem estruturados para o estudo em questão.

A etapa de descoberta, portanto, foi bastante útil como ponto de partida para entender como as fontes necessárias para o trabalho podem ser trabalhadas para adquirir um conjunto de dados mais limpo e com uma estrutura rigorosamente adequada para aplicação de modelos analíticos.

ARQUITETURA

Caracterização da Arquitetura

Após adquiridas todas as bases, iniciou-se o tratamento e manipulação das bases de forma a obter um *dataset* limpo para utilização em análises espaciais.

Neste *dataset* irá ser aplicado um modelo de DEA, para avaliar a eficiência de cada bairro segundo variáveis demográficas e internas a empresa, no intuito de obter uma lista de priorizações de ações que seriam estabelecidas sobre os bairros considerados ineficientes, além de projeções de eficiência para melhorar a performance destes bairros.

Os resultados do modelo anterior foram novamente juntados a base principal, agora possuindo campo relacionado a eficiência dos bairros. A partir desta aplicação será avaliada a distribuição dos pontos por faixas de referência para a eficiência, visualizando as regiões de priorização de ações vide resultado do modelo DEA.

Desta forma, tendo rastreado as necessidades tecnológicas para desenvolvimento do projeto os *softwares*/ferramentas escolhidas para utilização foram: *Alteryx Designer*, *software* responsável por todo o processo de extração, manipulação e carregamento dos dados que serão utilizados, Linguagem R, escolhida para implementação do modelo de análise por envoltória de dados, *CartoDB* e *Tableau Software*, *softwares* escolhidos para visualização de dados georreferenciados e não georreferenciados, respectivamente.

Para a prova de conceito, foi avaliada a possibilidade de uso das ferramentas dentro de um processo ETL de menor escala com aplicação de DEA simplificada. A prova de conceito iniciou-se pela extração de uma amostra da base relacionada aos pontos de vendas de distribuidoras de recarga, ainda não havendo nesta informação geográfica.

Nessa primeira etapa foi utilizado um total de 20 competências para a primeira tabela consolidada de dados. Ainda na extração, vale ressaltar que visto a grande quantidade de registros (+10 Milhões) no banco utilizado, foi necessário aplicar certos filtros para agilizar a conclusão da extração. Desta forma, aplicou-se um filtro para o estado apenas do Ceará, mais especificamente na cidade de Fortaleza.

Ao final da extração, gerou-se uma base com 315.000 mil registros e 72 campos distintos. Dentro desta base consolidada, possui dados a respeito de cada Ponto de Venda (PDV), mostrando seu faturamento ao longo de cada competência existente, além de informações relacionadas a sua localização e ao seu proprietário.

Esta base serviu de ponto de partida para análise, visto todas as explicações e esclarecimentos a respeito da forma de negócio da empresa, suas fontes de receitas e seus custos. Como passo inicial, buscou-se formas de agregar estes dados, através de campos que tivessem grande importância dentro do contexto do negócio.

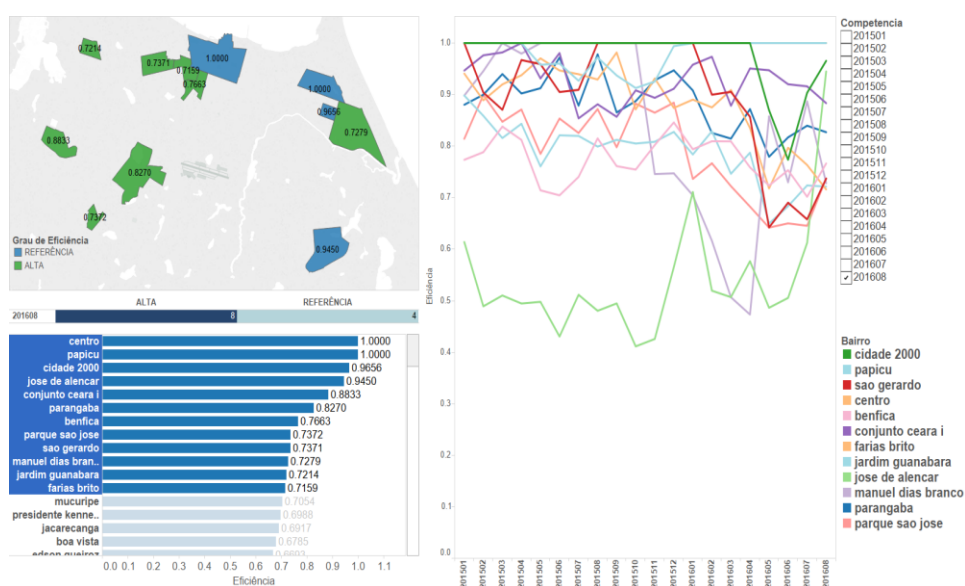
Para melhor compreender a evolução dos dados, buscou-se primeiramente a aplicação da metodologia DEA nos grupos agregados de bairros da cidade de

Fortaleza. Desta forma, todos os PDV's foram agregados por bairros, obtendo a quantidade de PDV's e o faturamento total de cada bairro.

Neste início buscou-se aplicar o modelo DEA para cada competência, ou seja, mensalmente, como pode ser visto na figura 4. No final de todo o processo, foi feita a união de todas as bases processadas mês a mês pelo modelo DEA e gerado um novo arquivo base para análise.

A partir deste arquivo foi criado o *Dashboard* mostrado na Figura 4 com o intuito de avaliar a evolução dos níveis de eficiência de cada bairro ao longo dos dois últimos anos, que vai de 0, pior grau de eficiência relativa, até 1, eficiência relativa máxima.

Figura 4 - *Dashboard* histórico de eficiência I



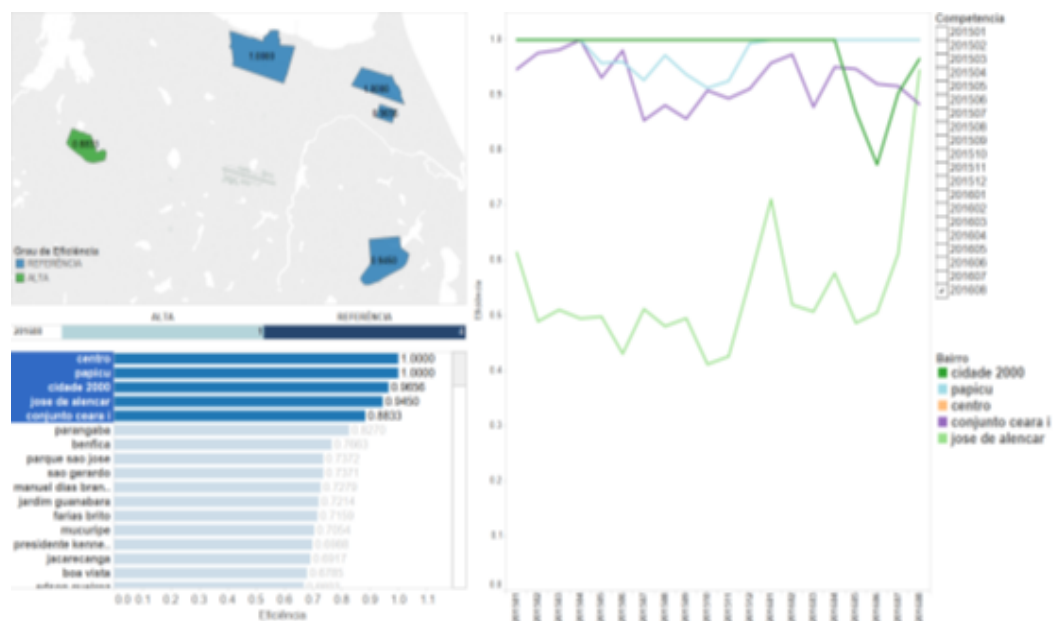
Fonte: Autores

Pela Figura 4, selecionando apenas os 12 bairros mais eficientes no mês de agosto/2016, observa-se que alguns bairros permaneceram várias competências com eficiência em torno do grau de referência.

Porém, nota-se também que vários bairros vêm decaindo em grau de eficiência ao longo dos últimos meses o que deve ser levado em consideração para analisar as causas e diagnosticar os problemas relacionados a esse decaimento. Por último, têm-se também os bairros que se mantiveram em oscilação como o Conjunto Ceará I, a Parangaba e o Benfica.

A Figura 5 busca analisar se de fato existem aquelas entidades que podem ser consideradas como referência para um *benchmarking* comparativo, estas entidades seriam aquelas que continuam se mantendo com alto grau de eficiência relativa diante dos outros bairros.

Figura 5 - Dashboard histórico de eficiência II



Fonte: Autores

No gráfico em linhas, na Figura 5, tem-se uma análise histórica dos 5 bairros mais eficientes no mês de agosto/2016. Assim, é possível notar que existem dois bairros que conseguem se manter quase que constantemente com grau de eficiência relativa máxima, visto que na maioria das competências sua linha está indicando eficiência igual a 1 pela escala de eficiência no eixo Y do gráfico. Estes bairros são Papicó e Cidade 2000.

Fora estes, existe ainda o centro sendo o único bairro que se mantém em todas as competências com eficiência relativa máxima igual a 1.

O bairro José de Alencar é um caso interessante, pois apesar de possuir sua eficiência variando entre 0.5-0.6 em grande parte das competências analisadas, nos últimos meses teve forte evolução. Já o Conjunto Ceará, apesar de obter alto grau de eficiência, possui comportamento extremamente irregular ao longo de seu histórico.

A partir dos resultados obtidos, foi validado pela empresa e pelos autores do trabalho, a real possibilidade de desenvolver o projeto proposto com o auxílio das ferramentas mapeadas, buscando conceber um sistema de apoio a decisão com uso de DEA para priorização de ações.

CONCEPÇÃO/DESENVOLVIMENTO

Para o projeto em questão as atividades que compuseram a etapa de concepção ou desenvolvimento foram: Desenvolvimento de processo ETL, aplicação da metodologia DEA e criação de macro para análise dos pontos de vendas passíveis de remoção.

Processo ETL

O processo ETL iniciou pela extração e aquisição de todas as bases necessárias para desenvolvimento do projeto. Desta forma, foi criado o fluxo para extração das bases internas da empresa. Em seguida, foi feita a preparação dos dados para posterior carregamento.

Assim como nas bases internas da empresa, foi realizada a junção das bases externas com informações relacionadas aos bairros que serão adicionadas a base gerada no fluxo anterior, para enriquecimento. Estas informações são necessárias para os modelos DEA que serão aplicados posteriormente.

As bases relacionadas a bairros foram adquiridas a partir dos próprios sites do IBGE e do portal de dados da cidade de Fortaleza - CE.

O último tratamento a ser realizado na base foi referente às coordenadas geográficas de cada PDV que, de acordo com a atividade de *data profiling* realizada, possuíam muitas linhas nulas além de coordenadas fora das fronteiras de referência para cidade de Fortaleza.

Para os registros nulos, optou-se por remover da base as linhas que não possuíam informações confiáveis, sendo necessário posterior análise juntos as distribuidoras

A última etapa do processo ETL está relacionado a agregação dos dados para adquirir a quantidade de PDV's no par competência-bairro e seus respectivos faturamentos. Após isso os dados foram carregados para o modelo de acordo com a competência escolhida pelo usuário.

Aplicação da Metodologia DEA

Na aplicação da metodologia DEA, as DMU's da simulação são os próprios bairros de Fortaleza - CE. O modelo em questão ordena os bairros que possuem maior faturamento diante das condições demográficas da localidade e da quantidade de PDV's.

Análise dos Pontos de Vendas Passíveis de Remoção

Nesta etapa busca-se a criação de um fluxo para processar os registros existentes na base e avaliar quais pontos estariam passíveis de serem removidos vide o critério de percentual de significância no faturamento total do CEP analisado e a eficiência relativa referente ao modelo DEA.

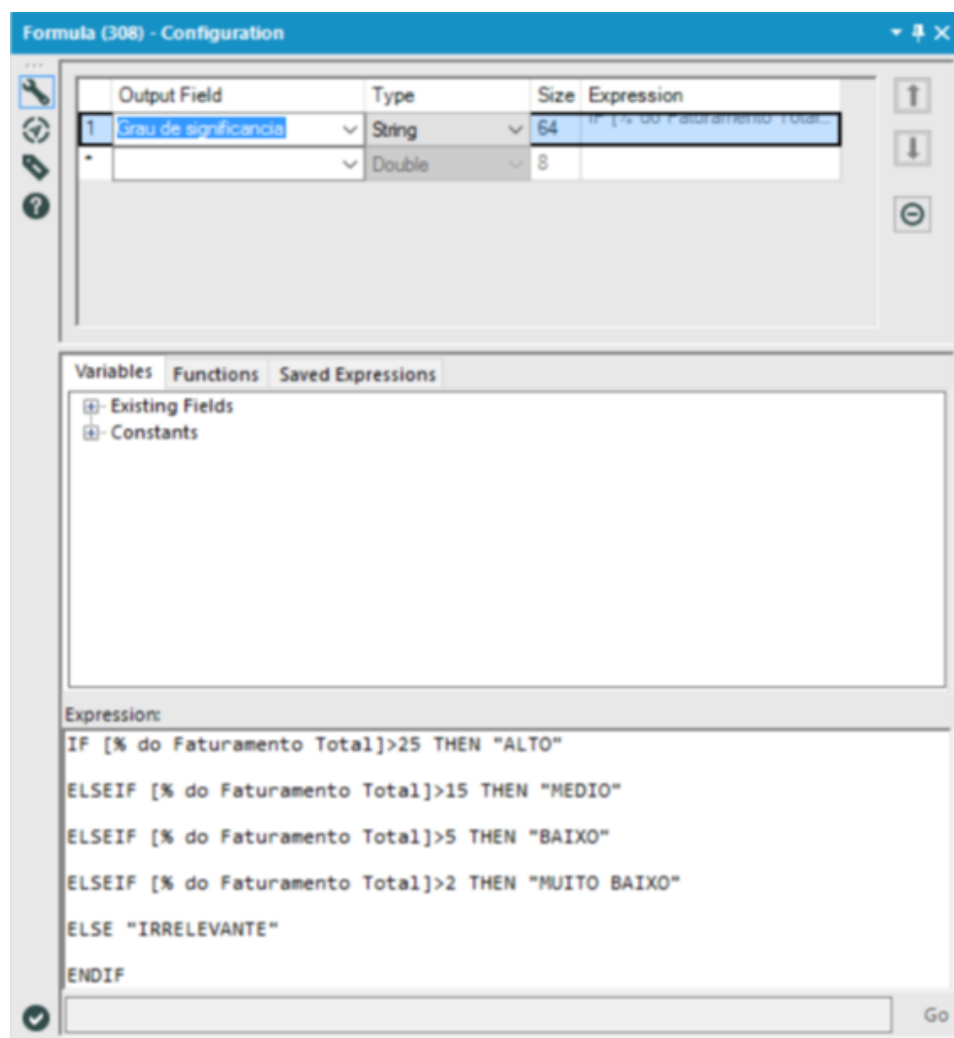
A partir da base consolidada e dos valores resultados da aplicação DEA, todos os PDV's representando bairros considerados eficientes foram filtrados.

Primeiramente foi feita uma agregação por bairro e depois por CEP para obter o faturamento total por cada par Bairro-CEP. Logo após organizou-se a base em ordem decrescente de faturamento.

Em seguida é feito a junção novamente das bases pelos campos Bairro e CEP. O intuito é após esse passo obter na mesma base os campos de faturamento do PDV e faturamento total do CEP em que o PDV se encontra.

Por fim, foi calculado o campo “% do faturamento total” que servirá como base para classificação dos PDV’s. Desta forma, os PDV’s serão categorizados de acordo com as faixas apresentadas na Figura 6.

Figura 6 - Categorias de faturamento



Fonte: Autores

O filtro aplicado elimina todos os pontos de vendas encaixados na faixa abaixo de 2% do faturamento total, considerados de baixa significância vide faturamento total.

É importante ressaltar que todos estes valores são parâmetros que podem ser modificados de acordo com o cenário específico que uma empresa pode-se encontrar.

TESTE/PRODUÇÃO

Para esta etapa foram criadas algumas métricas para avaliar a eficiência da solução proposta no caso analisado, visando entender as reduções resultantes tanto na quantidade de PDV’s por bairro como nos seus respectivos faturamentos.

Novamente foram realizadas agregações no nível de bairro para obter a soma do faturamento e a quantidade de PDV's após remoção. Depois disso, a base é anexada a base gerada, com o intuito de buscar as informações de faturamento e quantidade de PDV's por bairro antes de remoção dos pontos, afins de comparação. Esta junção é feita pelo campo de "Bairro".

No caso do problema supracitado, algumas métricas foram calculadas para avaliar o desempenho da solução gerada. Estas métricas estão apresentadas no Quadro 1.

Quadro 1 - Definição de métricas do teste

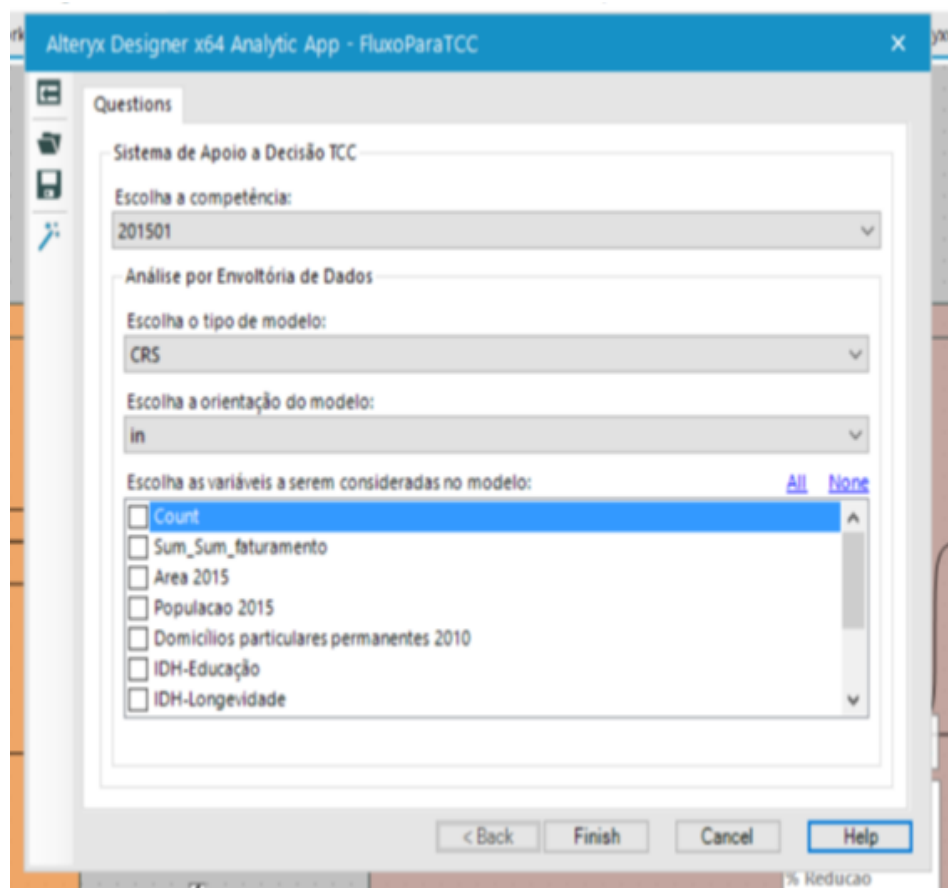
Métricas	Definição de Cálculo
Redução PDV	Representa a redução da quantidade de PDV's em valor absoluto. Calculada reduzindo-se o valor inicial do valor obtido após aplicação do fluxo completo.
% Redução PDV	Representa o percentual de redução da quantidade de PDV's. Calculada pela relação entre a métrica "Redução PDV" e o valor inicial de quantidade de PDV's.
Redução Faturamento	Representa a redução do faturamento em valor absoluto. Calculado reduzindo-se o valor inicial de faturamento do valor obtido após aplicação do fluxo completo.
% Redução Faturamento	Representa o percentual de redução do faturamento. Calculado pela relação entre a métrica "Redução faturamento" e o valor inicial de faturamento.
Gap PDV	Representa o valor faltante de PDV's a serem reduzidos para chegar ao valor de referência obtido através do resultado da aplicação DEA. Calculado pela subtração da projeção de eficiência PDV e a quantidade de PDV's removidos pelo fluxo.

Fonte: Autores

Como fase final foram adicionados importantes *outputs* para análise desta situação e tomada de decisão, de forma a visualizar os pontos de venda a partir de um mapa. Estes são descritos a seguir: *Output* Geral, mapa dos pontos de vendas antes de qualquer manipulação (Situação inicial), *Output* Eficientes, apenas os pontos de vendas pertencentes aos bairros considerados eficientes pelo modelo, *Output* Ineficientes, apenas os pontos de vendas pertencentes aos bairros considerados ineficientes pelo modelo, *Output* Removidos, todos os pontos que foram considerados como irrelevante pelo modelo e *Output* Final, mapa dos pontos de vendas após aplicado o fluxo por completo (Situação final).

Na Figura 7 é mostrada a interface da aplicação para selecionar parâmetros relativos ao modelo.

Figura 7 - Interface usuário



Fonte: Autores

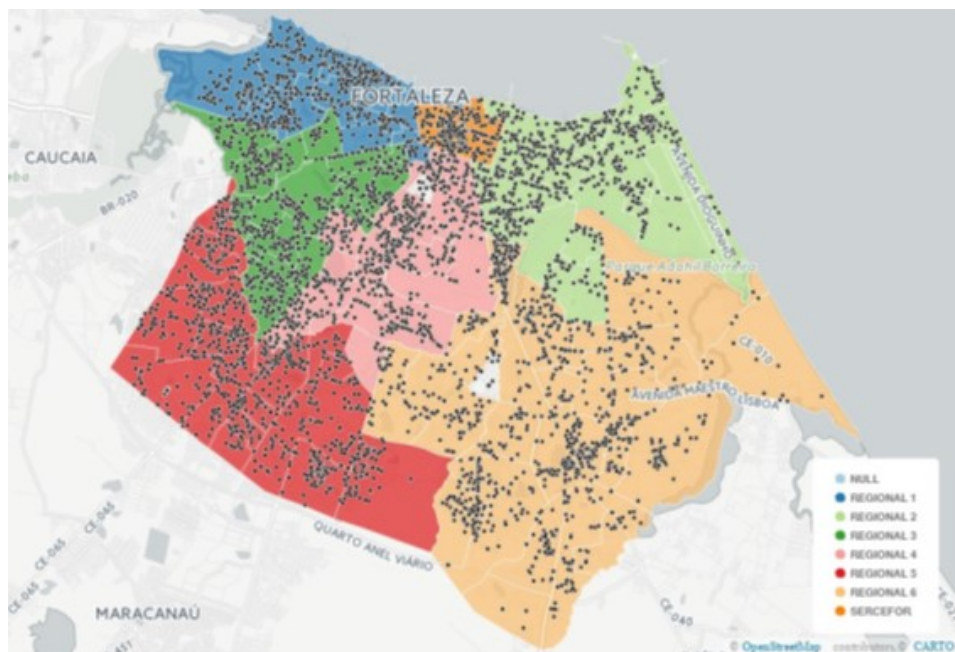
APLICAÇÃO PRÁTICA

Para aplicação prática, escolheu a base mensal de janeiro de 2015 pois apresentava maior quantidade de registros com dados georreferenciados. Também foi adotado apenas análise de uma das distribuidoras, sendo escolhida aquela com maior representatividade em ponto de vendas.

Situação Atual

A Figura 8, mostra como estão distribuídos os diversos pontos de vendas da base em questão analisada. O gráfico em questão foi gerado pelo *output* Geral.

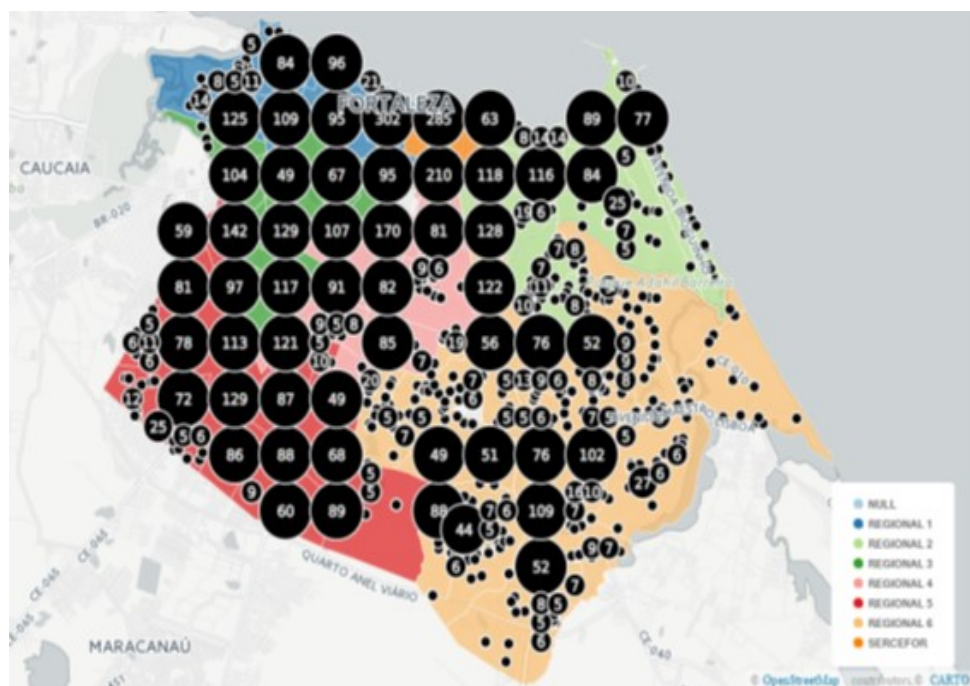
Figura 8 - Situação Janeiro/2015



Fonte: Autores

A criação de “clusters” facilitou o entendimento das concentrações de pontos de vendas. Esta análise é apresentada na figura 9.

Figura 9 - Clusters Janeiro/2015



Fonte: Autores

Situação Proposta

Para seleção das variáveis que iriam compor o modelo, procurou-se aquelas que mais tinham influência no faturamento agregado mensal.

Entre as variáveis existentes no parâmetro anteriormente criado optou-se por 4 variáveis referentes aos bairros analisados, sendo elas: quantidade de pontos de vendas, população, área e quantidade de domicílios existentes.

Para aplicação, o modelo BCC foi considerado o modelo mais propício visto que acréscimos nos *inputs*, podem promover ou não acréscimos no *output*, e esta relação não é proporcional. Da mesma forma, ele se adequa bem à heterogeneidade da amostra e ao porte relativamente não uniforme das unidades analisadas.

Foi adotada a orientação a *inputs*, uma vez que se busca a minimização das variáveis que impactam o custo da empresa, ressaltando que a única que pode ser de fato reduzida é a variável de quantidade de PDV's, pois faz parte das operações internas a empresa.

Após decisão sobre as variáveis e configurações a serem adotadas no modelo DEA, basta rodar o fluxo e analisar os resultados obtidos.

Na Figura 10, tem-se o resultado do modelo DEA aplicado a competência de janeiro/2015, para mostrar os campos gerados pelo modelo.

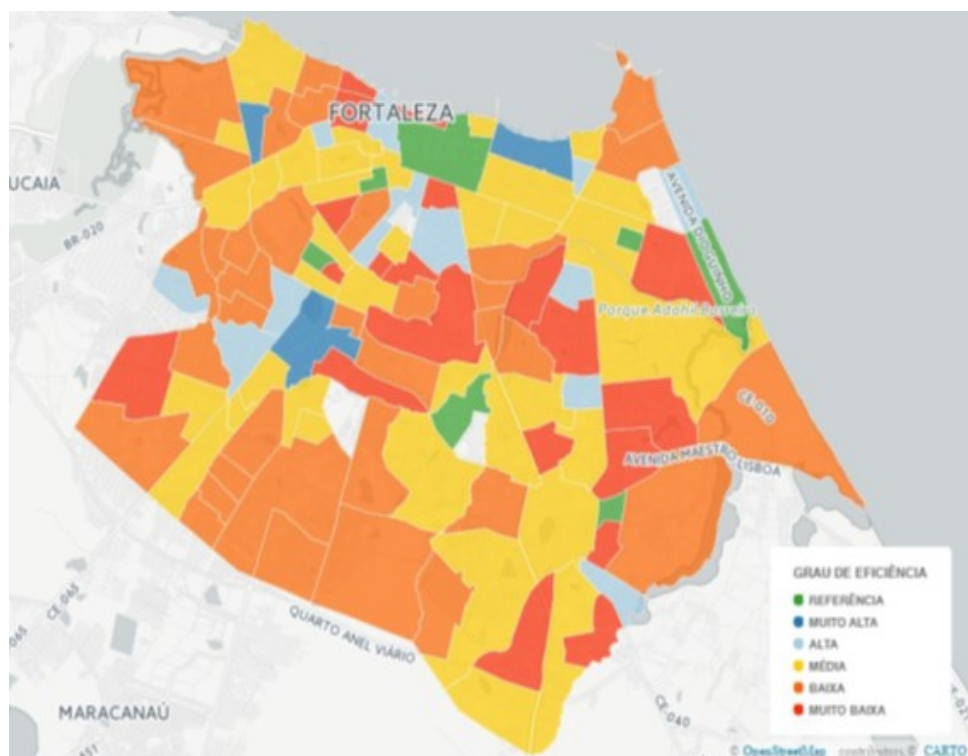
Figura 10 - Resultado aplicação DEA

Bairro	Eficiência	Projeção Quantidade PDV	% Projeção PDV	Projeção Área	% Projeção Área	Projeção População	% Projeção População	Projeção Domicílios	% Projeção Domicílios
boa vista	1	0	0,00%	0	0,00%	0	0,00%	0	0,00%
centro	1	0	0,00%	0	0,00%	0	0,00%	0	0,00%
cidade 2000	1	0	0,00%	0	0,00%	0	0,00%	0	0,00%
curti	1	0	0,00%	0	0,00%	0	0,00%	0	0,00%
gen americano	1	0	0,00%	0	0,00%	0	0,00%	0	0,00%
parque arca	1	0	0,00%	0	0,00%	0	0,00%	0	0,00%
praia do futuro II	1	0	0,00%	0	0,00%	0	0,00%	0	0,00%
meatim	0,94494	2.697737	1,52%	90905,73883	3,50%	1294,318040	3,50%	44.460210	3,50%
pedrin rocheta	0,877178	5.770223	12,28%	130231,355411	12,28%	2846,130871	12,28%	81.849205	12,28%
pedras	0,82384	1.917965	17,44%	747205,678837	17,44%	233,891701	17,44%	443.1236	17,44%
parangaba	0,822213	18.55834	17,78%	714994,783376	17,78%	9472,862572	17,78%	1846,084040	17,78%
berficia	0,814811	11.461726	18,52%	174632,722781	18,52%	1639,679732	18,52%	388,837963	18,52%
vareja	0,779402	7.843428	22,41%	118767,129721	22,41%	1885,326964	22,41%	625,681467	22,41%
Somacuco	0,754851	25.789303	24,31%	412603,728176	24,31%	9996,301703	24,31%	285,407189	24,31%
praia do futuro I	0,744054	7.413726	25,54%	419446,272882	25,54%	1682,660178	25,54%	492,118025	25,54%
paracangá	0,672839	16.7531	32,72%	411467,848452	32,72%	4063,618329	32,72%	1363,278388	32,72%
dama	0,672588	6.840118	32,74%	285383,848358	32,74%	3301,346708	32,74%	1155,182627	32,74%
conjunto casa I	0,624208	63.548143	37,38%	588678,284254	37,38%	7176,863216	37,38%	2445,779811	37,38%
guararapes	0,623304	1.882471	37,63%	509623,781867	37,63%	1862,618463	37,63%	382,060044	37,63%
vila alary	0,618543	6.484779	38,13%	17614,862436	38,13%	2897,210584	38,13%	874,500287	38,13%
fátima	0,587223	33.019541	46,27%	1330623,348448	46,27%	6113,121788	46,27%	2418,813354	46,27%
mesquita	0,580243	88.884146	41,63%	254983,28929	41,63%	17442,861278	41,63%	9189,872605	41,63%
maraponga	0,576381	27.848368	42,34%	724634,552440	42,34%	4294,612258	42,34%	1287,342605	42,34%
monte castelo	0,575209	21.234620	42,47%	326667,176489	42,47%	5610,184155	42,47%	1629,962909	42,47%
capiminas	0,574534	6.592033	42,55%	1406336,941871	42,55%	6101,048187	42,55%	1877,126811	42,55%
passaia	0,573051	22.086411	42,69%	306045,284843	42,69%	21995,30075	42,69%	3385,812080	42,69%
parque marburg	0,568184	6.477214	43,18%	544611,718861	43,18%	3251,129529	43,18%	88,890108	43,18%
praia de naama	0,558124	16.163145	44,19%	229432,848308	44,19%	1379,329328	44,19%	481,202838	44,19%
mutaípe	0,558061	22.53707	44,19%	389466,438221	44,19%	6068,658452	44,19%	1963,144109	44,19%
pedra amarela	0,557881	6.724627	44,20%	539423,551683	44,20%	5868,898005	44,20%	1881,188885	44,20%
joquei clube	0,555294	12.897462	44,47%	754823,85843	44,47%	8381,717813	44,47%	252,168005	44,47%
pedrin casanova	0,542864	5.027373	46,70%	362396,148874	46,70%	4895,833136	46,70%	1328,887816	46,70%
parque sao joão	0,52888	26.88940	47,19%	262936,514571	47,19%	4942,888231	47,19%	1423,782167	47,19%
sao geraldo	0,523863	11.426328	47,61%	686280,421781	47,61%	6881,02988	47,61%	2103,482167	47,61%
antonio bezerra	0,516791	63.30033	48,31%	1057616,115216	48,31%	12449,387028	48,31%	3613,454102	48,31%
passaia fazenda	0,514801	41.230977	48,51%	853346,812281	48,51%	11232,212834	48,51%	3386,738161	48,51%
centro II	0,510379	9.8623	48,46%	1344717,109419	48,46%	3764,548504	48,46%	1041,422287	48,46%
parque verde	0,504727	25.306201	48,53%	620851,248227	48,53%	7058,623468	48,53%	2181,088174	48,53%
camoquinha	0,494056	34.875806	51,94%	195031,703035	51,94%	20818,209645	51,94%	1834,848329	51,94%
bon jardim	0,481264	99.203254	51,87%	1232748,837114	51,87%	18178,839176	51,87%	5322,394229	51,87%

Fonte: Autores

Desta forma, pelos resultados acima mostrados, observa-se que sete bairros foram considerados como eficientes pelo modelo e irão servir como referência para os demais bairros considerados ineficientes. A partir deste resultado, o *output* geral foi usado para gerar a Figura 11.

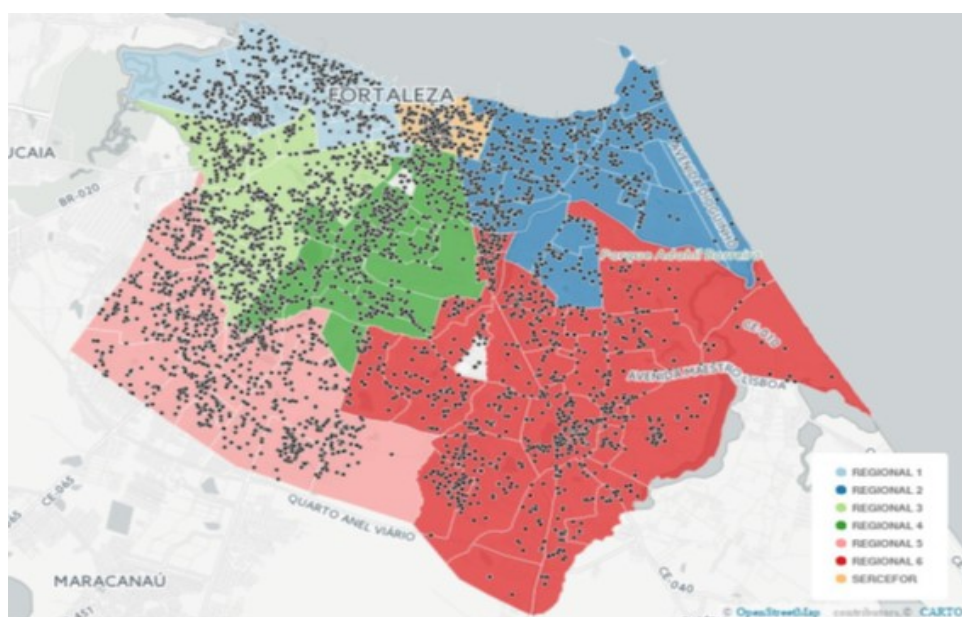
Figura 11 - Visualização grau de eficiência



Fonte: Autores

É possível notar que há maior predominância de bairros com maior eficiência nas regiões 2,3 e 4. Porém, a faixa de eficiência dominante encontra-se entre média e baixa (0.2-0.6). Vale ressaltar que a regional 5 é a única que não possui nenhum bairro considerado acima da faixa média (0.4-0.6) de eficiência.

Figura 12 - Situação proposta pelo modelo



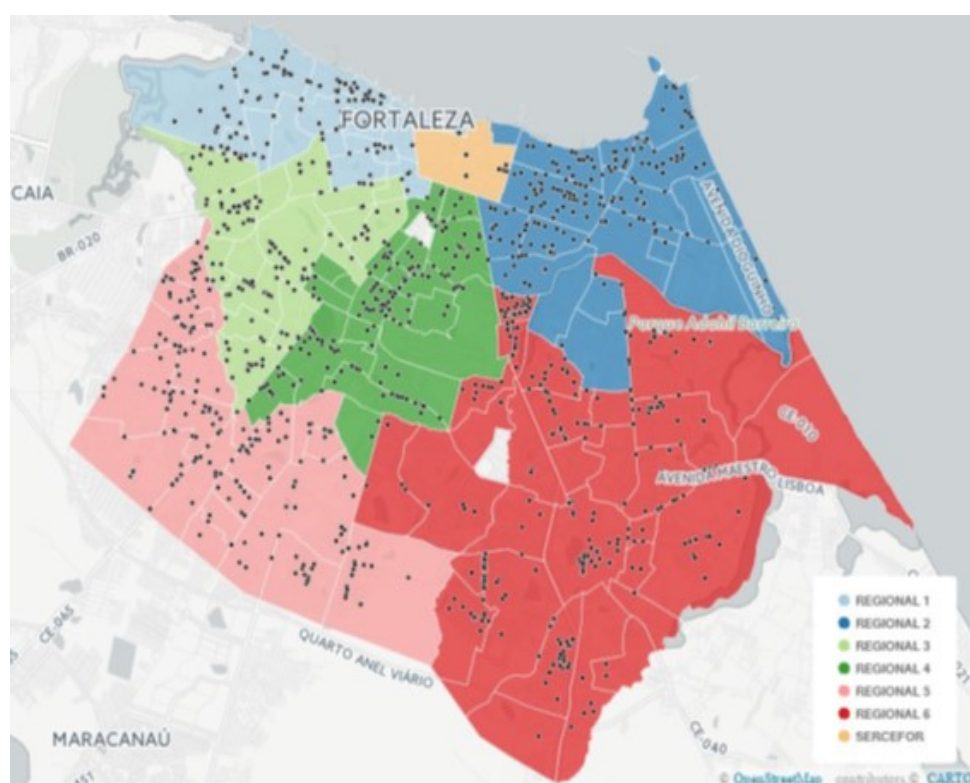
Fonte: Autores

A partir desses resultados a Figura 12 mostra o cenário obtido através do modelo, quando removido os PDV's que não possuem valor significativo de faturamento diante do agregado para seu CEP. Esta situação é representada pelo *output* final.

Pela figura 12 tem-se que apesar de retirar 1.247 pontos de vendas com o modelo, a distribuição continua bastante satisfatória, abrangendo a totalidade do território estudado.

A figura 13 mostra apenas aqueles pontos considerados como passíveis de remoção. Este gráfico foi gerado pelo *output* removidos.

Figura 13- PDV's removidos



Fonte: Autores

Para finalizar as figuras 14 e 15 mostram uma comparação do bairro aldeota nas duas situações. Este bairro foi escolhido como exemplo por ter sido um dos bairros que mais reduziu em quantidade de PDV's. Vale ressaltar que com o modelo implementado, a aplicação de filtros para bairros se torna processo prático dentro do ambiente de visualização.

Figura 14 - Bairro Aldeota situação inicial



Fonte: Autores

Figura 15 - Bairro Aldeota situação final



Fonte: Autores

Comparação dos Resultados

A figura 16 mostra a relação das métricas criadas para avaliar a performance da solução aplicada após fluxo completo.

Figura 16 - Avaliação métricas performance

Bairro	Reducao PDV	% Reducao PDV	% Projecao PDV	Reducao Faturamento	% Reducao Faturamento	Gap PDV
aerolandia	49	46.67%	67.39%	5866	13.08%	21.758839
aeroporto	0	0.00%	83.28%	0	0.00%	3.331169
aldeota	67	38.73%	58.04%	6929	4.41%	33.439494
alto da balanca	7	16.28%	66.90%	575	1.75%	21.766661
alvaro weyne	8	13.11%	66.44%	858	1.62%	32.530711
amadeo furtado	0	0.00%	86.21%	0	0.00%	6.896539
ancuri	35	59.32%	81.83%	2295	8.63%	13.280143
antonio bezerra	34	25.95%	48.03%	4427	3.41%	28.918384
arraial moura brasil	0	0.00%	89.16%	0	0.00%	4.458076
autran nunes	3	13.04%	74.47%	302	1.76%	14.128268
barra do ceara	26	21.31%	57.31%	2818	1.85%	43.922421
barroso	0	0.00%	72.89%	0	0.00%	33.529742
bela vista	4	10.53%	56.05%	385	0.86%	17.299333
benfica	10	16.13%	18.91%	837	0.95%	1.721473
bom futuro	0	0.00%	80.85%	0	0.00%	16.169243
bom jardim	37	18.97%	49.51%	4480	2.84%	59.550064
bonsucesso	15	14.15%	24.46%	1920	2.00%	10.922422
cais do porto	8	26.67%	60.15%	791	2.25%	10.04608
cajezeiras	4	20.00%	41.83%	490	1.43%	4.366149
cambeba	3	15.00%	48.54%	243	0.90%	6.708976
canindezinho	11	15.94%	49.44%	1207	1.18%	23.116833
carlito pamplona	2	7.69%	87.80%	123	1.32%	20.827181
cidade dos funcionarios	27	29.35%	57.94%	2443	3.26%	26.319345
coacu	0	0.00%	53.98%	0	0.00%	1.61952
coco	0	0.00%	58.91%	0	0.00%	11.193828
conjunto ceara i	42	24.71%	37.27%	3738	2.99%	21.350903
conjunto esperanca	22	37.29%	62.43%	2127	4.34%	14.833504
conjunto palmeiras	10	22.73%	76.21%	1004	3.25%	23.533539
couto fernandes	0	0.00%	94.06%	0	0.00%	2.821743
criso redentor	12	24.00%	59.94%	1368	2.62%	17.9723
damas	4	14.81%	32.74%	320	0.64%	4.839568
democrito rocha	8	18.18%	52.88%	734	1.66%	15.268145
dias macedo	12	33.33%	55.20%	1173	2.96%	7.87146
dionisio tomes	9	19.57%	50.87%	737	1.29%	14.398207
domlustosa	0	0.00%	64.19%	0	0.00%	3.209588
edson queiroz	43	40.57%	58.27%	4510	5.26%	18.764172
faras britto	7	28.00%	54.71%	741	2.24%	6.676781
fatima	31	37.80%	38.92%	4135	3.58%	0.915317
floresta	0	0.00%	71.43%	0	0.00%	5.000325
genibau	16	18.60%	67.54%	1420	1.90%	42.084261

Fonte: Autores

Para a competência analisada, houve uma redução total de 19,15% dos PDV's contemplados na base, sem redução significativa no faturamento.

Sendo assim, pelos resultados mostrados na figura 16, vemos que apesar de apenas 4 bairros ter atingido o valor de referência para as projeções de eficiência (Jacarecanga, Meireles, Parangaba e Pedras), muitos obtiveram valor satisfatório de redução (considerando a projeção como referência), sem impactar o nível de faturamento.

Visto modelo desenvolvido e aplicação prática estabelecida, o Quadro 2, irá resumir os principais benefícios ganhos com a tecnologia desenvolvida no trabalho.

Quadro 2 - Benefícios gerados no trabalho

Benefícios Gerados
Integração dos ambientes relacionados ao processamento ETL, aplicação de modelo analítico e carregamento para visualizações.
A partir da base gerada pelo modelo DEA outras problemáticas podem ser abordadas utilizando como referência o ranking de eficiência para priorização de ações.
Visualizações espaciais, permitindo avaliar diversos ângulos do negócio, como bairros eficientes, ineficientes, diferentes cenários, faixas de referência para faturamento e eficiência.
Interface permitindo controle sobre a configuração do modelo DEA e competência a ser analisada.
Tratamento nos campos de latitude/longitude dando a possibilidade de realizar diagnóstico sobre os erros de localização e analisar possíveis ações a serem tomadas.
Tratamento realizado nas categorias de bairros, tornando possível o uso da base com maior grau de confiança e validando maior número de registros para análises posteriores.

Fonte: Autores

A maior base utilizada no estudo apresentava mais de 3 milhões e o tempo total de processamento do fluxo criado, levando em considerações todas as etapas avaliadas no projeto levou em média 10.5 segundos de execução, o que gera flexibilidade na parametrização para testes de possíveis outros cenários que possa vir a ser interessante.

Cabe enfatizar aqui, que apesar do uso de algumas ferramentas sofisticadas para visualização, a empresa em questão possuía ainda poucas competências no que diz respeito a processamento de dados e visualizações espaciais.

Desta forma a ferramenta, mostra-se como uma proposta de serviço da empresa a seus clientes (distribuidoras e operadoras), podendo fornecer *insights* valiosos para um bom gerenciamento dos recursos envolvidos.

CONCLUSÃO

O presente trabalho teve o objetivo de elaborar uma ferramenta de apoio à decisão, utilizando o método de análise por envoltória de dados para o controle e avaliação dos postos de uma distribuidora de recarga de dispositivos móveis.

Diante do exposto, o presente estudo possibilitou uma análise de como o *software* Alteryx e Tableau, podem ser tecnologias relevantes para impactar positivamente as decisões. Com efeito, a partir dos dados obtidos pela DEA, pode-se extrair *insights* que irão auxiliar na tomada de decisão organizacional, contribuindo para tornar as operações mais eficientes e para a redução de custos. Além disso, por meio de uma comparação de processos similares, esses dados permitiram a classificação de pontos de vendas para remoção, gerando resultados satisfatórios no que diz respeito a redução geral (19,15% dos PDV's) sem impactos significativos no faturamento.

Foi identificado os principais fatores e parâmetros que devem ser considerados na formulação do problema, este objetivo foi obtido por meio da colaboração com pessoas da área de negócio e com as ferramentas estatísticas, por exemplo, o *software R*, utilizado no estudo de caso.

O estudo apresentou algumas limitações vide a qualidade dos dados georreferenciados presentes na base. Estes dados são obtidos pelos próprios supervisores responsáveis pelos pontos de venda, porém existem diferenças significativas entre os dados obtidos e a realidade dos fatos. Limitação semelhante é a escassez de informações georreferenciadas diante do total de registros da base, haja vista que apenas algumas distribuidoras mantêm esse tipo de dados em suas bases. De forma análoga, necessita-se melhor preenchimento dos campos relacionados às categorias de bairros.

No que se refere a estudos futuros, recomenda-se: a utilização de tratamento nos casos de pontos de vendas que apresentem latitude e longitude fora das fronteiras do seu bairro; Análise de critérios que possam ser vinculados ao modelo para avaliar quão significativo é um ponto de venda; Criação de uma interface acessível aos usuários deste modelo para facilitar a parametrização dos critérios utilizados como Nível de significância e outros que possam vir a ser considerados.

Development of a decision support system for the management of a mobile device recharge distribution using data envelopment analysis

ABSTRACT

The telecommunications industry has been struggling in recent years to raise its competitiveness. In view of this fact, many companies have adopted cost-cutting measures. Thus, based on the need for performance improvements and assertiveness in the allocation of resources, the possibility of implementing a decision support system was evaluated using the Data Envelopment Analysis (DEA) methodology in a company of mobile device recharge distribution in the city of Fortaleza. Through the use of data analysis tools in Big Data were mapped the data sources required for the work, as well as evaluated the quality of the data received. The functionality of the system was validated through the practical application on a monthly base of a telephone provider, where it was possible to obtain a reduction of 19.15% of the sales points (POS), without significantly affecting the total billing, besides providing visual aids allowing analysis of the POS's under different objectives.

KEYWORDS: ETL Process. Dashboard. Big Data. Operational Research. Decision Theory.

REFERÊNCIAS

ABREU, F. S. G. da G. **Desmistificando o conceito de ETL**. Disponível em: http://www.fsma.edu.br/si/Artigos/V2_Artigo1.pdf.

BANKER, R. D. Maximum likelihood, consistency and Data Envelopment Analysis: A statistical foundation. **Management Science**, Vol. 39, nº 10, pp. 1265-1273, 1993. **crossref**

BECK, K., BEEDLE M., VAN, B. A., COCKBURN, A. **Manifesto for agile software development**. 2001.

CIELO, I. **ETL – Extração, Transformação e Carga de Dados**. Disponível em: <http://www.datawarehouse.inf.br/etl.htm>. 2010

COOPER, William W.; SEIFORD, Lawrence M.; TONE, Kaoru. **Introduction to data envelopment analysis and its uses: with DEA-solver software and references**. Springer Science & Business Media, 2006.

COOPER, William W.; SEIFORD, Lawrence M.; ZHU, Joe. Data envelopment analysis: History, models, and interpretations. In: **Handbook on data envelopment analysis**. Springer, Boston, MA, 2011. p. 1-39. **crossref**

DATE, C. J. **Introdução a Sistemas de Banco de Dados**. Rio de Janeiro: Campus, 2000. P. 803.

DENSHAW, P. **Spatial decision support systems**. In: Maguire, D. J.; Goodchild, M. F.; Rhind, D. W., *Geographical Information Systems: principles and applications*, New York, Longman, vol. 1, 1991, 403-412.

FERREIRA, C. M. de C.; GOMES, A. P. **Introdução à análise envoltória de dados: teoria, modelos e aplicações**. Viçosa – MG: Editora UFV, 2009.

GANDOMI, A.; HAIDER, M. Beyond the hype: big data concepts, methods, and analytics, 2015. **Int J Inf Manag** 35 (2):137–144 **crossref**

Gartner IT Glossary. **Definition of Business Intelligence**, 2014. Retrieved from: <http://www.gartner.com/it-glossary/business-intelligence-bi/>

HANSSON, S. **Decision Theory: A Brief Introduction**, 1994.

LABRINIDIS, A.; JAGADHIS, H. V. **Challenges and opportunities with big data**, 2012. *Proc. VLDB Endow.* 5, 12 (August 2012), 2032-2033. **crossref**

LINS, M.P.E.; MEZA, L.A. **Análise envoltória de dados e perspectivas de integração no meio ambiente de apoio à decisão**. Rio de Janeiro: COPPE, 2000.

SOARES DE MELLO, João Carlos CB et al. Suavização da fronteira DEA: o caso BCC tridimensional. **Investigação Operacional**, v. 24, n. 1, p. 89-107, 2004.

NETO, S. L. R.; RODRIGUES, M. **Um modelo conceitual para integração de modelos científicos e informação geográfica**. In: III Workshop Brasileiro de Geoinformática – GEOINFO, 3., Rio de Janeiro, 2001. Anais. Rio de Janeiro, Sociedade Brasileira de Computação. 2001. 71-78.

RAFAELI NETO, S. L. **Sistemas de Apoio à Decisão Espacial: uma contribuição à teoria em geoprocessamento**, 2004.

SAVAGE, L. J. **The Foundations of Statistics**. J. Wiley, New York, 1954. second revised edition, 1972.

SEIFORD, L.M.; ZHU, J. An investigation of returns to scale under data envelopment analysis. **International Journal of Management Science**, v. 27, p.1–11. 1999. **crossref**

TAURION, C. Big Data. Rio de Janeiro: **Brasport**, 2013.
<https://pt.scribd.com/doc/259741402/Big-Data-Cezar-Taurion>.

TSOUKIÀS, A. “**De la théorie de la décision à l'aide à la décision**”, in D. Bouyssou, D. Dubois, M. Pirlot, H. Prade (eds.), **Concepts et Méthodes pour l'Aide à la Décision**, Hermès, Paris, 25 - 69, 2006.

Recebido: 07 Mar. 2019

Aprovado: 11 Out. 2020

DOI: 10.3895/gi.v16n3.9763

Como citar:

DANTAS, L.A.C. et al. Desenvolvimento de um sistema de apoio à decisão para gestão de uma distribuidora de recargas de dispositivos móveis utilizando análise por envoltória de dados. **R. Gest. Industr.**, Ponta Grossa, v. 16, n. 3, p. 147-174, Jul./Set. 2020. Disponível em: <https://periodicos.utfpr.edu.br/revistagi>.

Correspondência:

Leonardo André Colares Dantas

Universidade Federal do Ceará (UFC), Fortaleza, Ceará, Brasil

Direito autorial: Este artigo está licenciado sob os termos da Licença Creative Commons-Atribuição 4.0 Internacional.

