

Um Estudo Sobre a Relação entre Smartphones e Dados Demográficos

Wandella Maia¹, Fabrício A. Silva¹, Thais R. M. Braga Silva¹

¹Instituto de Ciências Exatas e Tecnológicas – Universidade Federal de Viçosa (UFV)
Rodovia LMG 818, km 6 35690-000 – Florestal – MG - Brasil

{wandella.maia, fabricio.asilva, thais.braga}@ufv.br

Abstract. *In the recent years, we are facing an increasing in the number of mobile users and, consequently, in the amount of data collected from them and available to the industry. Therefore, companies are willing to use such data to discover insights that may help them on providing better and more personalized services. In this work, we explore mobile and census data of thousands of users in Brazil to understand how their local of residence is correlated to their smartphones. We conduct a study to evaluate the relation between the demographic variables of the residence of users and their smartphones.*

Resumo. *Nos últimos anos, tem-se percebido um aumento significativo no uso de smartphones e, conseqüentemente, de serviços móveis oferecidos a seus usuários. Os smartphones, agindo como sensores, são excelentes fontes de dados sobre os usuários e o ambiente ao qual estão inseridos. Como consequência, as empresas estão cada vez mais interessadas em extrair conhecimento útil desses dados para conhecer melhor os seus usuários e, assim, oferecer serviços melhores e mais personalizados. Neste trabalho, é feito um estudo para avaliar a relação entre dados demográficos do local de residência dos usuários e os smartphones dos mesmos.*

1. Introdução

Nos últimos anos, tem-se percebido um aumento significativo no uso de *smartphones* e, conseqüentemente, de serviços móveis oferecidos a seus usuários. Os *smartphones*, agindo como sensores, são excelentes fontes de dados sobre os usuários e o ambiente ao qual estão inseridos. Como consequência, as empresas estão cada vez mais interessadas em extrair conhecimento útil desses dados para conhecer melhor os seus usuários e, assim, oferecer serviços melhores e mais personalizados.

Dentre vários aspectos sobre os usuários móveis, podem-se destacar dois: o local de residência e o modelo de aparelho móvel utilizado. Com o local de residência, é possível inferir características demográficas dos usuários, que podem servir como base para direcionamentos de campanhas de marketing e oferta de serviços. Por outro lado, o modelo do *smartphone* e seu respectivo preço pode indicar interesses pessoais e financeiros do usuário.

O objetivo deste trabalho é entender como as características do local de residência de um usuário móvel se relaciona com o seu tipo de *smartphone*, e propor modelos para estimar o preço de aparelhos móveis de acordo com o local. Para isso, foram utilizados

dados reais de mais de 60.000 usuários do Brasil. Foram feitas caracterizações e análises de correlações com os dados. Além disso, foram testados modelos preditivos para estimar o preço dos *smartphones* de uma região. Os resultados revelam que existem algumas características demográficas dos locais de residência que podem indicar uma maior probabilidade dos usuários possuírem *smartphones Android* ou da *Apple*. Porém, não foi possível estabelecer com precisão um modelo preditivo para o preço médio dos *smartphones* com base em características da região de residência.

Este texto está organizado da seguinte forma. Na Seção 2 são discutidos os principais trabalhos que avaliam características demográficas de regiões e de usuários móveis. Em seguida, na Seção 3, os dados utilizados no trabalho são descritos. Os resultados são apresentados e discutidos na Seção 4. Finalmente, as conclusões e trabalhos futuros são listados na Seção 5.

2. Trabalhos Relacionados

O uso de diversas fontes de dados para a extração de conhecimento sobre o comportamento, características pessoais, mobilidade, interesses e aspectos demográficos tem sido bastante comum nos últimos anos. Com o grande volume de dados publicamente disponíveis, ou coletados por serviços proprietários, as empresas começaram a olhar com atenção para esses dados com o intuito de extrair informações úteis sobre seus usuários, que possam ajudar a melhorar o engajamento e a retenção dos mesmos.

Alguns trabalhos da literatura utilizam as informações de censo coletadas pelos governos para extrair informações e correlações relevantes. [Eagle et al. 2010] avaliaram a correlação entre perfis de uma rede social online e a situação econômica dos usuários com base em regiões censitárias do Reino Unido. [Wesolowski et al. 2013] utilizam dados de migração do censo para aproximar padrões de movimento humano através de escalas temporais. [Remigio et al. 2019] têm como objetivo verificar se a qualidade do bairro pode influenciar na saúde e segurança das pessoas. [De Nadai et al. 2016] visam testar a promoção de condição de vida humana em uma cidade da Itália com base no censo.

Outra fonte de informação também muito relevante é referente a dispositivos móveis e sua utilização por meio da telefonia ou acesso à Internet. Esses dados também são explorados por alguns autores em estudos sobre as características dos usuários. [Brdar et al. 2012] tiveram como desafio fazer uma exploração de dados reais de dispositivos móveis para revelar padrões de características demográficas do usuário. O artigo de [Blumenstock et al. 2015] utiliza dados de telefonia móvel de Ruanda, que possui ricas informações, e fornece correlações do uso da telefonia com estatísticas populacionais com agregação regional de censos e pesquisas domiciliares. O artigo de [Smith-Clarke et al. 2014] tem como objetivo analisar registros de assinantes de telefonia móvel e extrair correlações para ajudar o governo a monitorar regiões mais carentes. [Grauwin et al. 2015] tentam detectar padrões de comportamento humano nos principais centros financeiros com base em registros de número de chamadas, *SMS*, transferência de dados e também dados censitários. [Zhong et al. 2013] têm como objetivo criar um *framework* que prevê os dados demográficos dos usuários com base nos dados móveis. O objetivo é utilizar registros de atividades físicas, registros de usuários e contextos ambientais para prever informações como gênero, tipo de emprego, estado civil, dentre outras

características.

Além de dados de censos e de dispositivos móveis, outros tipos de dados também têm sido utilizados na análise de perfil de usuários. [Steiger et al. 2015] fazem uma análise espaço-temporal das redes sociais sobre as pessoas e suas atividades sociais em áreas urbanas. Os dados foram obtidos através de serviço de *stream* oferecido pela rede social *Twitter* e para validação, utilizaram dados do censo do Reino Unido. [Lenormand and Ramasco 2016] utilizam aplicações empíricas de *Big Data* para estudar a sistemática de cidades e seus problemas de mobilidade. Com isso, utilizou-se 20 milhões de *tweets* de vários lugares do mundo para quantificar e comparar as 58 áreas mais influentes no mundo, levando em conta a área coberta pelos usuários. [Bogomolov et al. 2014] utilizam dados sobre o comportamento humano agregado e anônimo, derivados da atividade da rede móvel, para lidar com o problema da previsão de crime. Sendo assim, foram utilizados para análise três conjuntos de dados: um com a localização de 124.119 pessoas de Londres, um com todos os crimes reportados do Reino Unido e o último com as métricas da população de Londres.

Diferente dos trabalhos existentes, o presente estudo analisa dados de residência dos usuários, com suas respectivas características demográficas, e o aparelho móvel que o usuário possui. O objetivo é investigar se existe alguma correlação entre esses dois aspectos, e se essa correlação pode ser usada para se conhecer melhor o perfil do usuário. Esta é a primeira vez que esse tipo de correlação é feita com uma amostra real com dezenas de milhares de usuários, e com informações detalhadas da região de residência.

3. Os Dados

3.1. Usuários

Para a realização deste trabalho, foi disponibilizada por uma empresa parceira que atua na área de serviços móveis, uma amostra de dados com 64.108 usuários. Esses usuários representam cerca de 1.000 usuários *Apple* e 63.000 *Android*, contendo a localização aproximada da residência do usuário e o modelo do *smartphone* utilizado. Os dados foram coletados com o consentimento dos usuários, e disponibilizados para a pesquisa sob as penas de um termo de confidencialidade assinado entre as partes. Esses usuários estão presentes em 3.665 municípios do Brasil, e distribuídos no mapa conforme mostra a Figura 1.

3.2. Características dos *Smartphones*

A base de usuários possui apenas um nome técnico indicando o modelo do dispositivo móvel. Como são muitos dispositivos distintos, a coleta das características de cada um manualmente tornaria o trabalho inviável. Portanto, foi criada uma ferramenta chamada *CrawMobi*¹ para complementar a base do trabalho.

O *CrawMobi* é um *Web Service* cujo objetivo principal é auxiliar o usuário a pesquisar na Web informações técnicas sobre *smartphones*. Para isso, a ferramenta lança mão de técnicas como recuperação da informação, mecanismos de busca e *web scraping*. A recuperação da informação é realizada por meio de navegação automática na Web para percorrer as páginas de interesse e extrair os conteúdos relacionados.

¹Agradecemos Kaio Igor Vieira Alves pela contribuição no *CrawMobi*.

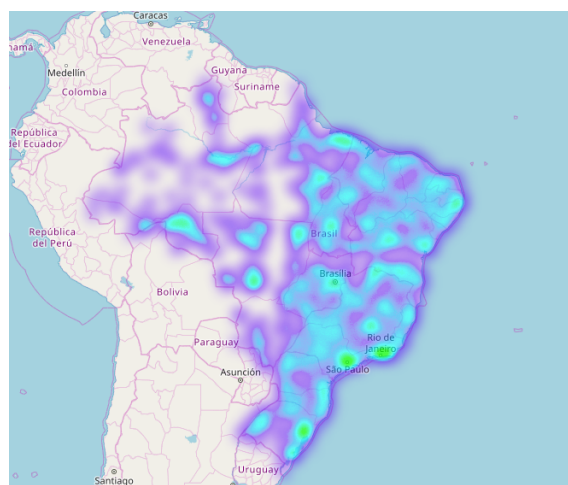


Figura 1. Mapa de distribuição dos 64.108 usuários ao longo do Brasil.

A figura 2 apresenta a arquitetura do *Crawmobi*, contendo todos os seus componentes e o fluxo da informação desde a requisição do cliente até o retorno com a caracterização dos dados. A interação com a aplicação é através de uma interface. Primeiramente, é necessário criar um arquivo em formato de planilha com os nomes comerciais dos aparelhos desejados para a coleta, descritos linha após linha na primeira coluna do arquivo. Os nomes comerciais devem ser compostos pela marca do aparelho seguido de seu modelo. Após isso, deve-se fazer o *upload* do arquivo na página do *Web Service*. Logo, o serviço retorna um arquivo contendo 22 características dos aparelhos listados, que são recuperados de dois sites diferentes para manter a consistência das informações. Além disso, um terceiro site é utilizado para a recuperação do preço do aparelho em valores atualizados. Após a coleta, os dados são agregados e formatados antes de serem enviados ao cliente interessado.

Para este trabalho, foram recuperadas as informações de mais de 400 aparelhos de diferentes marcas. Os principais atributos recuperados são marca, tamanho da tela, capacidade de memória, se possui Bluetooth e NFC, capacidade da bateria, sistema operacional e preço. No escopo deste artigo, serão utilizados apenas o sistema operacional (se Android ou iOS) e o preço. A Figura 3a ilustra a quantidade de usuários por marca de aparelho. Percebe-se uma predominância da marca *Samsung* com mais da metade dos usuários, sendo que 695 usuários utilizam um aparelho da marca *Apple*. A Figura 3b apresenta um histograma mostrando a distribuição dos preços, separados por aparelhos das marcas que usam *Android* e da *Apple*. Pode-se perceber que a maioria dos aparelhos se encontram na faixa de preço de R\$1.000, sendo que, proporcionalmente, aparelhos da *Apple* são mais caros.

3.3. Setor Censitário

Os dados com os setores censitários foram obtidos através do *site* governamental do IBGE (Instituto Brasileiro de Geografia e Estatística) que disponibiliza os resultados do censo do ano de 2010. O IBGE é um instituto que realiza pesquisas sobre a população brasileira e esses dados buscam detalhar as condições em que a população vive. O setor censitário é uma área menor (existem vários setores por cidade) de controle onde o objetivo é saber informações cadastrais daquela área. O arquivo no formato *shapefile* espacial foi obtido

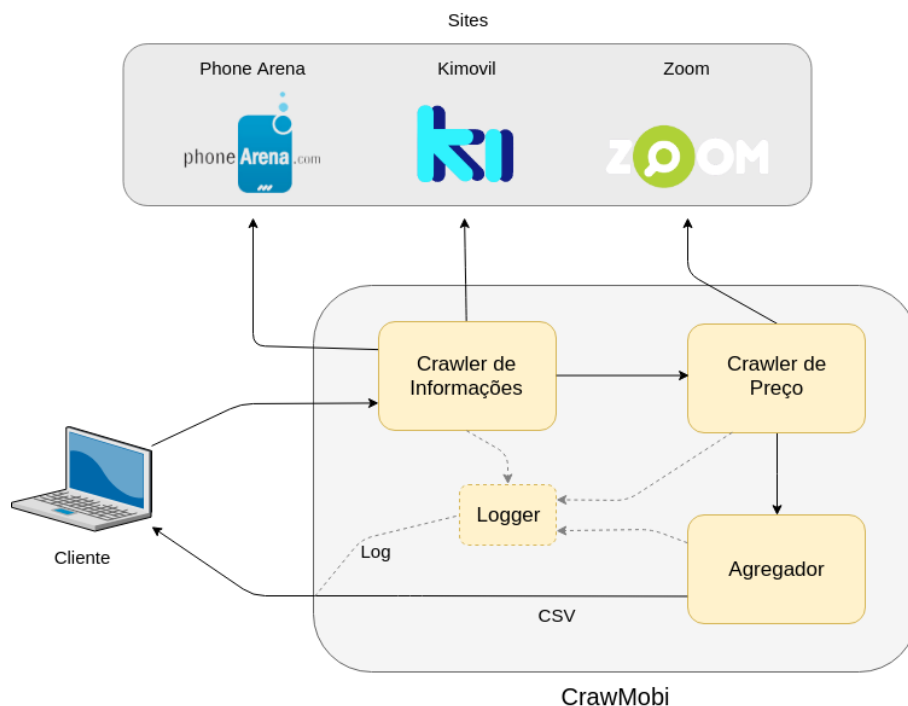


Figura 2. Arquitetura do CrawMobi

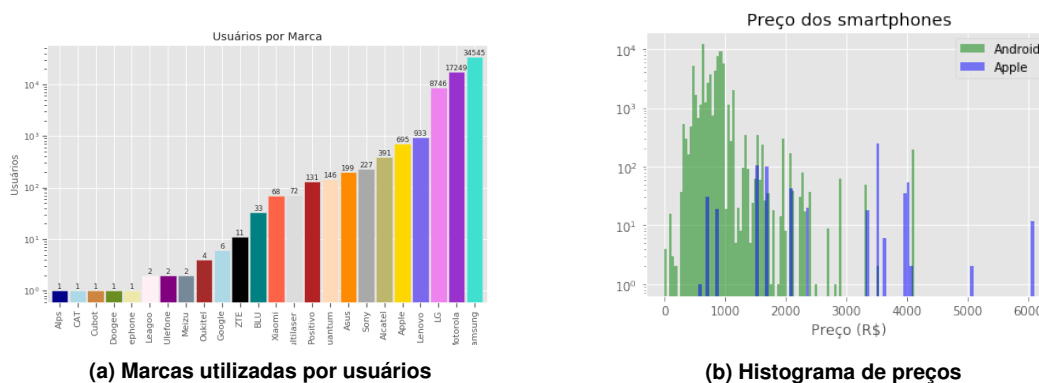


Figura 3. Características dos smartphones dos usuários, sendo o histograma dos preços e a quantidade de usuários por marca.

e contém polígonos com toda a divisão dos municípios cobertos pelos dados, totalizando 310.120 setores. O uso de setores censitários e não de cidades é um diferencial deste trabalho, pois um setor representa mais precisamente as características demográficas, com poucas variações entre seus residentes.

4. Análise e Resultados

Inicialmente, foi feita uma caracterização dos dados para que seja possível extrair conhecimentos sobre a relação entre as características do local de residência com as características do smartphone, em especial aparelhos Android (de várias marcas) ou iOS (Apple). Em seguida, foi feita uma análise com regressão linear para se tentar estimar o preço do smartphone com base nas características dos setores censitários. Por fim, foram criados modelos de classificação para avaliar a capacidade de se classificar a faixa de

preço de um *smartphone* com base nas características de uma região.

4.1. Caracterização

A Figura 4a apresenta como estão distribuídos os *smartphones* pelo número de moradores por domicílio, diferenciando-os pelo sistema operacional (o valor zero no gráfico indica ausência da informação para o usuário). Como o número de aparelhos com iOS é inferior ao de *Android* (ver Seção 3.2), é esperado que a curva descrita no gráfico para o sistema operacional da *Apple* esteja abaixo da do concorrente, o que de fato ocorre, e que o comportamento das duas sejam próximos, o que não acontece. De fato, setores censitários com média de cinco ou mais moradores por domicílio não apresentam o número de *smartphones* com *iOS* que deveriam, o que mostra uma correlação negativa entre essas variáveis. Ou seja, é possível supor que quanto mais moradores há em uma residência, menores são as chances de que algum deles tenha um celular com *iOS*, e que usuários da *Apple* normalmente dividem sua residência com outras poucas pessoas.

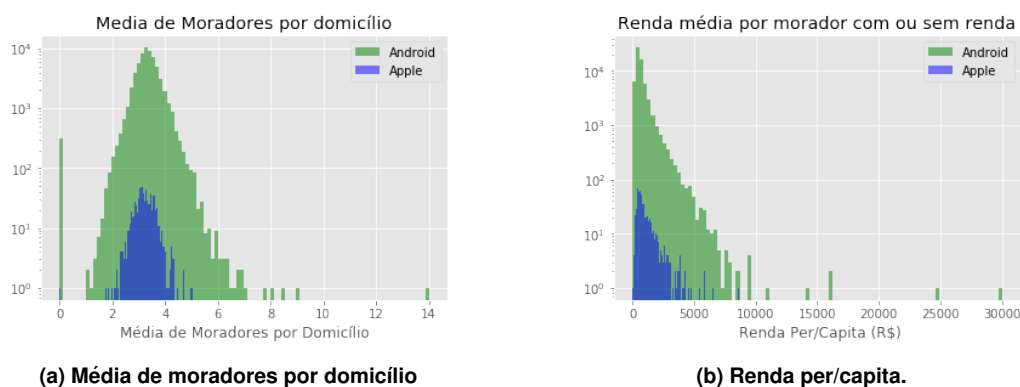


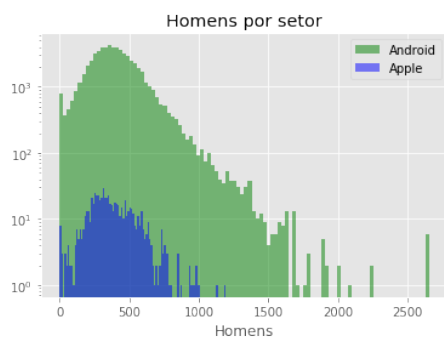
Figura 4. Histogramas do número de moradores por domicílio e da renda per/capita.

A Figura 4b apresenta as distribuições de renda média per capita dos setores em termos do sistema operacional dos usuários. Apesar de os aparelhos da *Apple* serem em geral mais caros (como mostra a Figura 3b), não foi possível observar que os usuários *Apple* estejam em regiões com renda/capita maior. Apesar de termos uma amostra menor desses usuários, ainda assim era esperado um comportamento diferente.

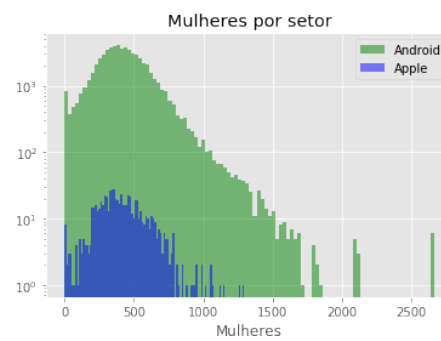
Em relação ao gênero, a Figura 5 mostra que em setores censitários em que o número de pessoas do gênero masculino é maior, a proporção de usuários *Apple* é menor. Ou seja, em locais com uma predominância um pouco maior de mulheres, a proporção de usuários *Apple* é maior.

Em relação às características de cor e raça, pode-se perceber na Figura 6 que há muita diversidade nas curvas apresentadas. Dentre esses resultados, pode-se destacar que em setores em que a taxa de negros é menor, maior é a proporção de usuários *Apple*. Por exemplo, não foi registrado nenhum usuário *Apple* em setores com mais de 50% de negros. O mesmo é válido para amarelos, sendo que a maior proporção de usuários *Apple* está nos setores com predominância parda ou branca.

Por fim, em termos dos atributos de idade e alfabetização pode-se observar outras características interessantes. A Figura 7 mostra que quanto menos alfabetizado for o

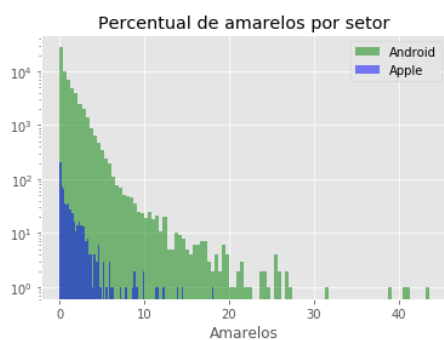


(a) Histograma do gênero masculino

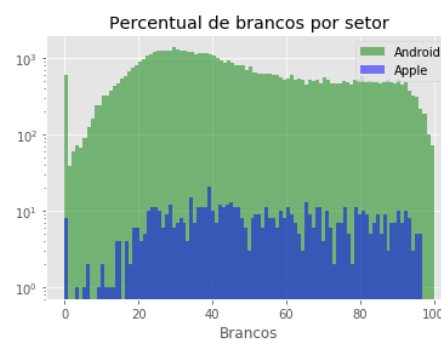


(b) Histograma do gênero feminino

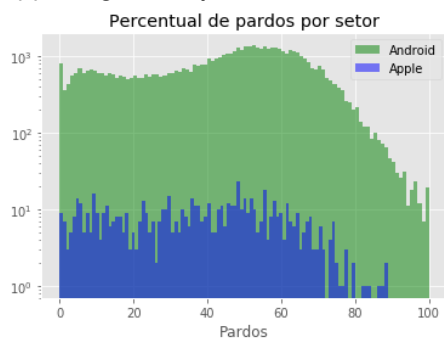
Figura 5. Histograma da taxa média de (a) homens e (b) mulheres



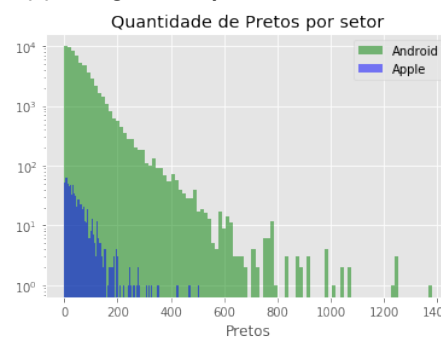
(a) Histograma de pessoas de cor amarela



(b) Histograma de pessoas de cor branca



(c) Histograma de pessoas de cor parda

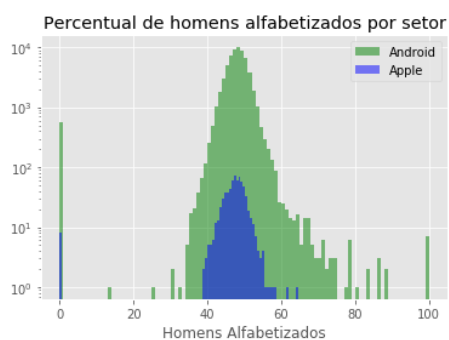


(d) Histograma de pessoas de cor preta

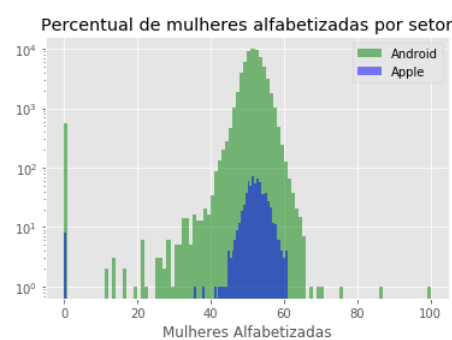
Figura 6. Histograma das 4 principais cores da população

setor, maior é a proporção dos usuários *Android*. Além disso, na Figura 8 percebe-se uma distribuição similar em termos das idades dos moradores para usuários *Android* e *Apple*.

Com base nos resultados obtidos, é possível perceber que existe alguma correlação entre as características de uma região e a proporção de usuários *Apple* e *Android* dessa região. Em resumo, foram observadas diferenças nas proporções para características de alfabetização, gênero, número de moradores e cor. Diferentemente do esperado, a renda per capita do setor não foi um fator relevante, mesmo considerando que aparelhos da marca *Apple* são mais caros que marcas que adotam *Android* em geral.

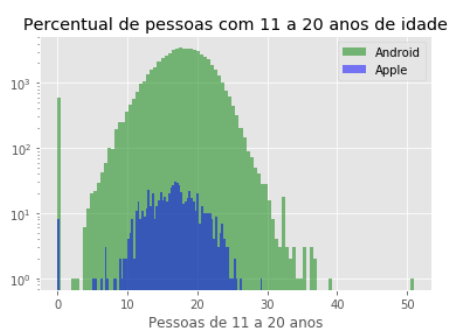


(a) Histograma de homens alfabetizados

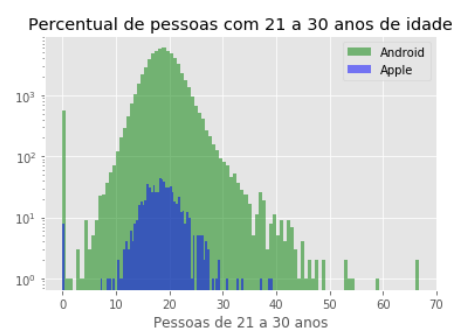


(b) Histograma de mulheres alfabetizadas

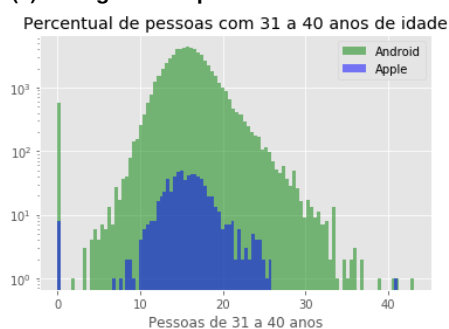
Figura 7. Histograma da taxa média de (a) homens e (b) mulheres alfabetizados



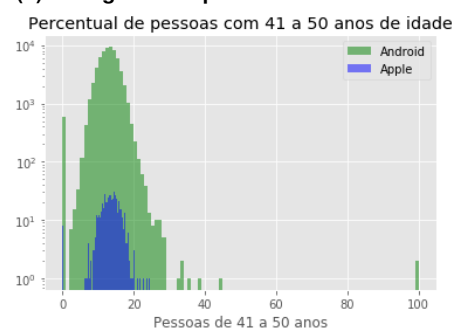
(a) Histograma de pessoas de 11 a 20 anos



(b) Histograma de pessoas de 21 a 30 anos



(c) Histograma de pessoas de 31 a 40 anos



(d) Histograma de pessoas de 41 a 50 anos

Figura 8. Histograma de faixas de idade

4.2. Modelo de Regressão

Nesta etapa, foram utilizados modelos de regressão linear com o objetivo de estimar o preço médio dos *smartphones* de um setor censitário com base em suas características. Neste trabalho, aplicou-se diferentes combinações de atributos da base de dados de setores censitários como variáveis explicativas. Essas combinações foram realizadas inicialmente em pares, e em seguida foram feitas várias alternativas com potencial de bons resultados.

Na Figura 9, são apresentados os principais resultados de correlação par-a-par. Os atributos do setor censitário referentes à taxa de alfabetização, taxa de mulheres, taxa de pessoas negras e rendimento per capita foram os que tiveram melhores correlações com o preço médio dos *smartphones* do setor. Porém, é possível perceber que, pelo menos visualmente, a correlação não é nítida para aparelhos *Android* e *iOS*.

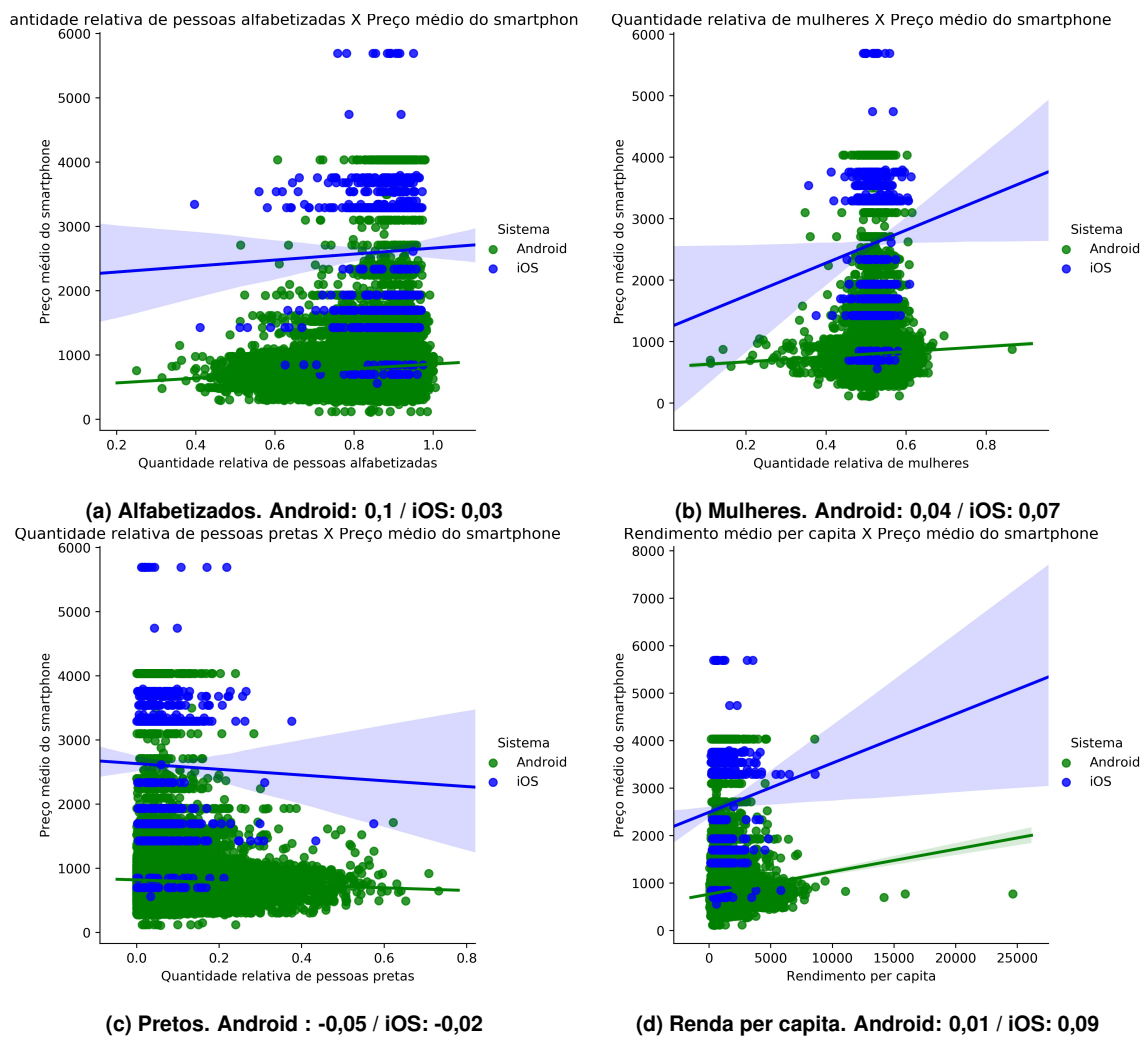


Figura 9. Correlações analisadas em relação ao preço do smartphone com coeficiente de Pearson para smartphones com Android e iOS.

Para verificar se uma combinação com mais variáveis explicativas pudesse estar correlacionada ao preço do *smartphone*, foram criados diferentes modelos de regressão linear multi-variada, alternando os atributos dos setores utilizados como entrada. A tabela 1 mostra as nove combinações de variáveis explicativas que tiveram os menores erros. Pode-se perceber que mesmo essas combinações tiveram um valor de R^2 muito baixo, o que mostra que o preço não é explicado pelas variáveis na grande maioria das vezes. Além disso, os erros em geral foram significativamente altos, como é possível perceber na coluna de desvio médio quadrático, e o algoritmo de validação cruzada também apresentou resultados ruins. Tudo isso aconteceu independente da combinação de atributos usada, o que reforça a tese de que essas variáveis, em geral, não apresentam um grau de significância suficiente com esse conjunto de dados para a previsão do preço do *smartphone*. Em resumo, não foi possível encontrar um modelo para estimar o preço médio dos *smartphones* dos usuários com base nas características demográficas do local de residência dos mesmos.

Atributos usados	R^2	Desvio médio quadrático
idade/mulheres/pretos/amarelos	0,02	381,10
moradores domicílio/mulheres/pretos/amarelos	0,02	381,13
mulheres/pretos/amarelos	0,02	381,13
idade/moradores domicílio/mulheres/amarelos	0,02	381,14
idade/mulheres/amarelos	0,02	381,14
moradores domicílio/mulheres/amarelos	0,02	381,16
mulheres/amarelos	0,02	381,16
idade/moradores domicílio/pretos/amarelos	0,02	381,21
idade/pretos/amarelos	0,02	381,22

Tabela 1. Resultados para diferentes combinações de atributos do setor censitário como entrada.

4.3. Modelo de Classificação

4.3.1. Classes de preço

De acordo com o site de tecnologia Tech Tudo [Amazon 2019], os *smartphones* podem ser classificados em modelos básicos, intermediários e intermediários *premium*. Os básicos permitem utilizar aplicativos de mensagens, de redes sociais, de navegação e de músicas. Logo, pode-se encontrar modelos simples na faixa de R\$500. Os *smartphones* intermediários permitem geralmente utilizar redes sociais, tirar fotos, gravar vídeos, apresentam mais memória, câmeras melhores e são um pouco mais rápidos. Logo, eles abrangem a faixa de R\$500 a R\$1.000. Os modelos *premium* possuem telas melhores, câmeras de ótima resolução para fotos e vídeos, mais espaço de armazenamento e ótimo desempenho. Logo, eles ocupam a faixa de preço de R\$1.000 a R\$2.5000.

Neste trabalho, além das três classes descritas acima, em que as faixas de preço foram classificadas em *barato*, *moderado*, *caro*, também foi acrescentada uma quarta categoria chamada *muito caro* para os aparelhos com preço acima de R\$2.500.

4.3.2. Modelos de predição e escolha

O objetivo do modelo é prever a faixa do preço (barato, moderado, caro ou muito caro) do preço médio de *smartphones* com base em características demográficas de um setor censitário. Foi apresentado anteriormente que as correlações entre as variáveis não foram muito atrativas, e um modelo de regressão não foi capaz de estimar o preço com precisão. Agora, será verificado se algum algoritmo de classificação supervisionada poderá ser aplicado para alcançar o objetivo.

Para isso, cinco algoritmos foram testados: Regressão Logística, Árvore de Decisão, *Bagging* (variação da árvore de decisão), *Random Forest* e *KNN*. Tais algoritmos são muito conhecidos na área de ciência dos dados e são do tipo supervisionados. Logo, todos eles aprendem com os dados através de uma variável objetivo, que no caso, é a faixa do preço.

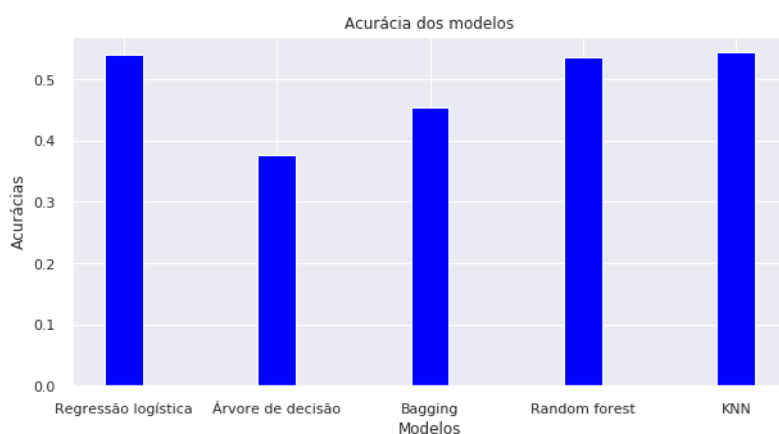


Figura 10. Resultados das acurácias de cada modelo

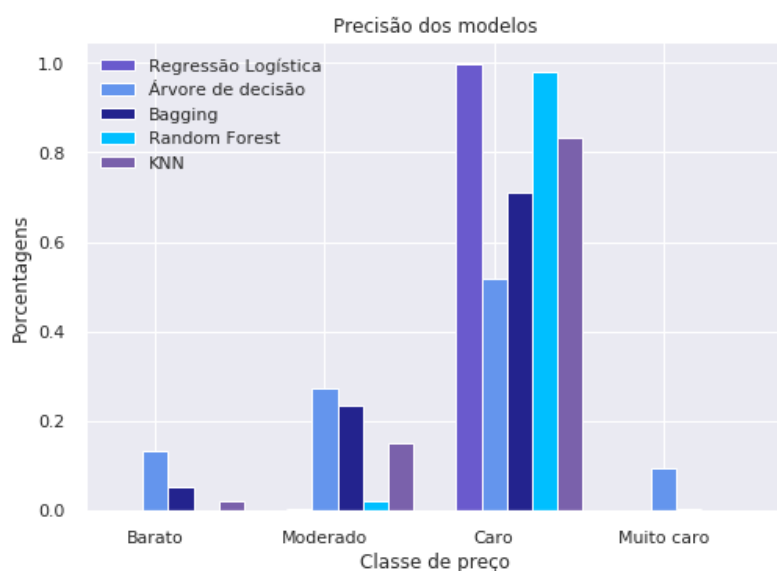


Figura 11. Resultados da precisão de cada modelo

A acurácia é uma das métricas mais simples, pois se trata do número de acertos diante de todos os casos. A precisão é a porcentagem de resultados classificados corretamente para cada categoria individualmente, e a revocação é a porcentagem dos resultados relevantes que foram classificados corretamente.

A Figura 10 apresenta a acurácia para cada um dos modelos. Como pode ser visto, os modelos de Regressão Logística, *Random Forest* e *KNN* apresentaram os melhores resultados de acurácia. No entanto, a acurácia geral pode ser distorcida quando um modelo é muito eficiente para alguma classe com maior número de entradas. Para verificar esse detalhe, as Figuras 11 e 12 apresentam a precisão e revocação para cada classe individual.

Como pode ser percebido, a classe *Caro* apresenta os melhores resultados, que estão fazendo com que a acurácia fica distorcida. Fica claro pelas figuras que todos os modelos foram muito ineficientes para classificar *smartphones* das classes *barato*, *moderado* e *muito caro*. Um dos motivos para esses resultados é o fato de se ter uma distribuição de *smartphones* de diferentes preços em diferentes locais com características similares, fa-

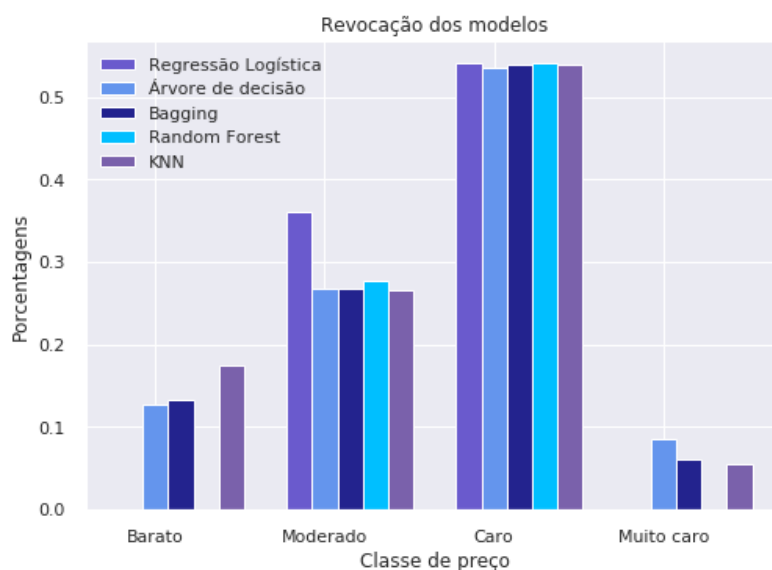


Figura 12. Resultados do *recall*(revocação) de cada modelo

zendo com que o modelo não consiga separar bem essas características que representam bem o preço.

Dentre todos os modelos, o de Árvore de Decisão foi o que obteve melhores resultados em geral, por serem mais balanceados. Por exemplo, a Regressão Logística foi excelente para a categoria *caro*, obteve precisão próxima de zero para as outras. Então, a seguir são apresentados alguns resultados extras específicos para esse modelo

Na Figura 13, é possível analisar melhor os resultados apresentados. Com isso, nota-se que há um bom desempenho para classe de preço se tratando da curva AUC. A medida de *micro-average* (micro-média) permite a possibilidade de desenho da curva ROC considerando cada elemento na matriz do indicador de etiqueta com uma previsão binária. Já no *macro-average*(macro-média) é atribuído um peso igual a classificação de cada rótulo.

A Figura 14 apresenta a matriz de confusão que é conhecida por permitir a visualização do desempenho de um modelo supervisionado. Ela permite ver como o modelo classificou cada classe. Sendo assim, pode-se observar que o modelo conseguiu classificar todas classes, alcançando um balanceamento melhor que os outros modelos. Apesar de ter alguns modelos terem alcançado uma precisão muito alta para a categoria *caro*, em geral os modelos não foram capazes de prever com boa precisão a faixa de preços dos aparelhos com base em dados demográficos dos moradores.

5. Conclusão

Este trabalho teve como objetivo avaliar possíveis correlações entre características demográficas do setor censitário de residência e as características do *smartphone* de um grande volume de usuários móveis. Para isso, foram feitas caracterizações em termos de renda, cor/raça, gênero, alfabetização, idade e número de moradores. Foi possível verificar que algumas características do setor fazem com que a proporção de usuários *Apple* seja diferente dos usuários *Android*.

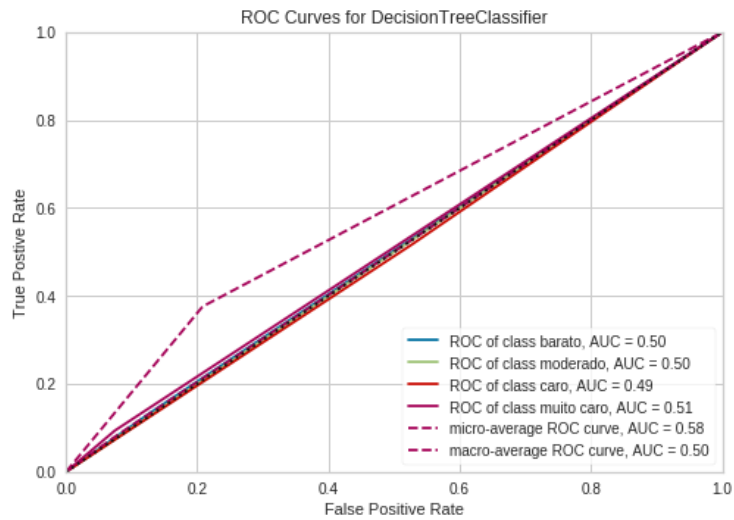


Figura 13. ROC Árvore de decisão

DecisionTreeClassifier Confusion Matrix

True Class \ Predicted Class	barato	moderado	caro	muito caro
barato	275	558	1080	124
moderado	568	1248	2330	329
caro	1187	2487	4700	667
muito caro	140	309	611	91

Figura 14. Resultado da matriz de confusão para classe de preço no modelo de Árvore de decisão

Após a caracterização, foram feitas análises com modelos de regressão com o objetivo de prever o preço médio de *smartphones* de um setor censitário com base em suas características. Ao se tentar várias combinações de variáveis explicativas, verificou-se que não existe uma correlação clara e, conseqüentemente, não foi possível criar um modelo de regressão preciso para prever o preço dos *smartphones*. Além disso, também não foi possível treinar um modelo supervisionado eficiente para classificar a faixa de preço de um *smartphone* com base nos dados demográficos. Este trabalho é importante pois as empresas têm buscado oferecer melhores serviços a seus usuários. Portanto, esta análise pode auxiliá-las a direcionar as ofertas e promoções de serviços e produtos ao público-alvo que potencialmente terá interesse. Como trabalhos futuros, pretende-se avaliar outros modelos de aprendizagem para estimar o preço e o sistema operacional dos usuários. Além disso, pretende-se também fazer análises em relação às características das cidades e estados.

Agradecimentos

Este trabalho contou com o apoio do CNPq e da Fapemig.

Referências

- Amazon (2019). *Melhores smartphones em todas as faixas de preço em 2019*. Amazon.
- Blumenstock, J., Cadamuro, G., and On, R. (2015). Predicting poverty and wealth from mobile phone metadata. *Science*, 350(6264):1073–1076.
- Bogomolov, A., Lepri, B., Staiano, J., Oliver, N., Pianesi, F., and Pentland, A. (2014). Once upon a crime: Towards crime prediction from demographics and mobile data. In *Proceedings of the 16th International Conference on Multimodal Interaction, ICMI '14*, pages 427–434, New York, NY, USA. ACM.
- Brdar, S., Culibrk, D., and Crnojevic, V. (2012). Demographic attributes prediction on the real-world mobile data. In *Proc. Mobile Data Challenge by Nokia Workshop, in Conjunction with Int. Conf. on Pervasive Computing, Newcastle, UK*.
- De Nadai, M., Staiano, J., Larcher, R., Sebe, N., Quercia, D., and Lepri, B. (2016). The death and life of great italian cities: a mobile phone data perspective. In *Proceedings of the 25th international conference on world wide web*, pages 413–423. International World Wide Web Conferences Steering Committee.
- Eagle, N., Macy, M., and Claxton, R. (2010). Network diversity and economic development. *Science*, 328(5981):1029–1031.
- Grauwin, S., Sobolevsky, S., Moritz, S., Gódor, I., and Ratti, C. (2015). Towards a comparative science of cities: Using mobile traffic records in new york, london, and hong kong. In *Computational approaches for urban environments*, pages 363–387. Springer.
- Lenormand, M. and Ramasco, J. J. (2016). Towards a better understanding of cities using mobility data. *Built Environment*, 42(3):356–364.
- Remigio, R. V., Zulaika, G., Rabello, R. S., Bryan, J., Sheehan, D. M., Galea, S., Carvalho, M. S., Rundle, A., and Lovasi, G. S. (2019). A local view of informal urban environments: a mobile phone-based neighborhood audit of street-level factors in a brazilian informal community. *Journal of Urban Health*, pages 1–12.
- Smith-Clarke, C., Mashhadi, A., and Capra, L. (2014). Poverty on the cheap: Estimating poverty maps using aggregated mobile communication networks. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 511–520. ACM.
- Steiger, E., Westerholt, R., Resch, B., and Zipf, A. (2015). Twitter as an indicator for whereabouts of people? correlating twitter with uk census data. *Computers, Environment and Urban Systems*, 54:255–265.
- Wesolowski, A., Buckee, C. O., Pindolia, D. K., Eagle, N., Smith, D. L., Garcia, A. J., and Tatem, A. J. (2013). The use of census migration data to approximate human movement patterns across temporal scales. *PLOS ONE*, 8(1):1–8.
- Zhong, E., Tan, B., Mo, K., and Yang, Q. (2013). User demographics prediction based on mobile data. *Pervasive and Mobile Computing*, 9(6):823 – 837. Mobile Data Challenge.