

Análise de Dados da Uber: Um Novo Olhar Sobre a Habitabilidade e a Mobilidade Urbana

Gisliany Alves * Ivanovitch Silva * Luciana Lima **

* Departamento de Engenharia de Computação e Automação,
Universidade Federal do Rio Grande do Norte, RN (e-mail:
gisliany.alves.094@ufrn.edu.br, ivanovitch.silva@ufrn.br).

** Departamento de Demografia e Ciências Atuariais, Universidade
Federal do Rio Grande do Norte, RN (e-mail: luciana.lima@ufrn.br)

Abstract: In 2016, the UN defined policies to foster the Sustainable Development Goals, whose implementation depends on the supervision of indicators. In this scenario, liveability is a dimension that can be monitored by indicators for that purpose, as it combines fundamental attributes of sustainable development. However, the lack of consistent and up-to-date data makes this monitoring difficult, mainly in developing and less developed countries. Hence, as a contribution to this field, this work proposes a liveability indicator that combines traditional population-based data, namely the population census, and alternative sources, such as data from Uber. For estimating the indicator, data from Uber rides, precisely the Estimated Time to Arrive, were combined with classic socioeconomic indicators. The methodology includes Factor Analysis (FA), Exploratory Data Analysis (EDA), Exploratory Spatial Data Analysis (ESDA) and Regression Analysis. Among the main results, there was a spatial variation of Uber services and the liveability indicator. One of the main conclusions is that non-traditional data, such as those from the Uber platform, can describe social and mobility inequalities between different regions of the same city.

Resumo: Em 2016, a ONU definiu políticas fomentadoras dos Objetivos de Desenvolvimento Sustentável, cuja implementação deve ser supervisionada por indicadores. Nesse cenário, a habitabilidade enquadra-se como uma dimensão a ser monitorada por meio de indicadores, posto que combina atributos pilares do desenvolvimento sustentável. Todavia, a escassez de dados consistentes e atualizados dificulta esse monitoramento, sobretudo em países em desenvolvimento e menos desenvolvidos. Como contribuição à área, este trabalho propõe um indicador de habitabilidade que combine dados tradicionais de base populacional, como os censos demográficos, e de fontes alternativas, como as disponibilizadas pela Uber. Para a construção do índice foram combinadas informações de Estimativas de Tempo até o Embarque das corridas da Uber e indicadores socioeconômicos clássicos. A metodologia inclui Análise de Fatores (AF), Análise Exploratória dos Dados (AED), Análise Exploratória dos Dados Espaciais (AEDE) e Análise de Regressão. Entre os principais resultados verificou-se variação espacial dos serviços da Uber e do indicador de habitabilidade. Uma das principais conclusões é que dados não tradicionais como os da plataforma Uber conseguem descrever as desigualdades sociais e de mobilidade entre diferentes regiões de uma mesma cidade.

Keywords: Data-Driven Approach, Uber, Liveability Indicators, Urbanization, Sustainable Urban Development.

Palavras-chaves: Abordagem Orientada a Dados, Uber, Indicadores de Habitabilidade, Urbanização, Desenvolvimento Urbano Sustentável.

1. INTRODUÇÃO

Nas últimas décadas, o mundo tem vivenciado um expressivo crescimento na parcela populacional urbana (Valcárcel-Aguiar et al., 2019). Segundo a ONU, 30% da população mundial era urbana em 1950 e estima-se que esse percentual atinja 68% em 2050 (United Nations, 2018). No Brasil, a população urbana passou de 18,8 milhões (36,8%) de habitantes em 1950 para 160,9 milhões em 2010 (84,4%), provocando problemas de infraestrutura e deficiências habitacionais (Carmo et al., 2014).

Devido à crescente complexidade econômica, desigualdade social e preocupação ambiental nas cidades, a elaboração de um planejamento metropolitano estratégico pautado no desenvolvimento sustentável tornou-se crucial (Pupphachai and Zuidema, 2017). Um termo chave que surge nesse sentido é a habitabilidade, um conjunto de atributos econômicos, sociais e físicos de uma área urbana que, quando melhorados sem causar degradação ambiental, terão um impacto positivo na qualidade de vida dos habitantes (Valcárcel-Aguiar et al., 2019). A habitabilidade foi listada pela ONU como um dos princípios e compromete-

timentos da Nova Agenda Urbana (NAU), fruto de uma campanha global que reconhece a relevância das cidades e que compromete-se com a promoção dos Objetivos do Desenvolvimento Sustentável (ODS) (United Nations, 2016).

Para tornar os ODS efetivos, é preciso monitorar o progresso de sua implementação (Barnett and Parnell, 2016). Nesse caso, indicadores de habitabilidade, por sua própria definição, podem ser vantajosos para acompanhar esse progresso (Lowe et al., 2015). O principal obstáculo reside no limite dos dados disponíveis, em termos de se conhecer o que coletar, a periodicidade da coleta para cada localidade, a qualidade dos *datasets* existentes e o seu nível de agregação (nacional, regional, etc.) (Satterthwaite, 2016).

É válido mencionar ainda que o desenvolvimento urbano sustentável e a habitabilidade tem relação direta com as instalações e as atividades de transporte (Wey and Huang, 2018). Os padrões de trânsito e os sistemas de transporte são capazes de refletir realidades sociais (Lucas, 2012), além de poderem melhorar a acessibilidade e beneficiar as condições de habitabilidade (Irwin, 2003). Ademais, a NAU e o ODS 11 mencionam a mobilidade urbana sustentável, segura e acessível como uma de suas metas (United Nations, 2016). Carvalho (2016) aponta que o transporte urbano sustentável baseia-se na proteção ambiental, sustentabilidade econômica e justiça social, incentivando o uso e a melhoria do transporte público coletivo.

Ainda segundo Carvalho (2016), o Brasil tem vivido poucas experiências satisfatórias de mobilidade sustentável, fato observado com os tempos maiores de deslocamento, os congestionamentos crescentes e a emissão de poluentes pelo uso de transporte individual motorizado em detrimento do transporte público coletivo. Em Natal, capital do Rio Grande do Norte (RN) e objeto do presente estudo, a queda acumulada no uso do transporte público foi de 22,28% entre os anos de 2012 e 2017, motivada por sua ineficiência, pela crise econômica, a chegada dos aplicativos de transporte e o aumento da frota de veículos particulares (Tribuna do Norte, 2018).

Face às lacunas existentes na mobilidade urbana no Brasil, o serviço da Uber entrou em cena. Por um aplicativo móvel, a empresa conecta motoristas parceiros aos usuários que solicitam uma viagem, configurando-se como transporte individual privado (Uber Technologies Inc., 2019). Quanto à sua relação com a mobilidade urbana sustentável, a Uber tem um caráter dual. Por um lado, a ideia do serviço de transporte individual privado é contrária àquela priorizada pelo desenvolvimento urbano sustentável, que incentiva o uso do transporte público coletivo, por exemplo. Todavia, políticas urbanas inteligentes impulsionam o desenvolvimento sustentável ao empregarem tecnologias de informação e comunicação, sendo o aplicativo da Uber um símbolo da realidade urbana inteligente no Brasil (Mariano et al., 2019). Telésforo (2016) ainda destaca as parcerias entre a Uber e o setor público, em que o serviço privado complementa o transporte público coletivo.

Conforme o que foi descrito sobre o serviço da Uber e a mobilidade urbana sustentável, bem como a relação entre transporte e habitabilidade, este trabalho presume que os dados provenientes da API da Uber também podem expressar a dimensão da habitabilidade. Isso posto e face

à necessidade de mais dados para a promoção dos ODS, criou-se um indicador de habitabilidade para o município brasileiro de Natal, capital do Rio Grande do Norte, que combina além de dados tradicionais do Censo Demográfico 2010 e de outras fontes de informação, os dados de viagens solicitadas pelo aplicativo da Uber. A escolha de Natal como objeto de estudo se motiva, conforme mencionado anteriormente, na existência de uma demanda reprimida por transporte na cidade, nas falhas em suas vivências de mobilidade urbana sustentável e na conseqüente necessidade de supervisionar suas condições de habitabilidade. Além disso, a capital do RN é uma das 20 maiores cidades do país em termos populacionais, com cerca de 900 mil habitantes, que vivem em sua totalidade em áreas urbanas (IBGE, 2021).

Além desta introdução, este trabalho apresenta, na Seção 2, o estado da arte referente ao uso de dados da Uber no contexto da habitabilidade. Na sequência, a Seção 3 trata da metodologia aplicada, enquanto os resultados e suas implicações são relatados na Seção 4. Por fim, as conclusões são discutidas na Seção 5.

2. TRABALHOS RELACIONADOS

A fonte de trabalhos relacionados é baseada em publicações que utilizam os dados da Uber, especialmente no contexto do desenvolvimento urbano sustentável e da habitabilidade. Jin et al. (2019) investigaram a relação espaço-temporal entre a Uber e o transporte público e avaliaram a influência do serviço da empresa sobre a equidade de distribuição do transporte urbano em Nova York (EUA) no ano de 2014. Os dados utilizados foram os locais de embarque da Uber e de táxis, além das paradas de transporte público. Os autores também analisaram a relação dos locais de embarque da Uber com dados socioeconômicos. Os resultados desse trabalho mostraram que a Uber tanto complementa quanto compete com o transporte público e que o papel desse serviço em melhorar a igualdade na distribuição do transporte é insignificante. As análises com dados socioeconômicos revelaram que o número de pontos de embarque da Uber tende a ser menor em locais de baixa renda e com maior concentração de minorias étnicas e sociais.

Hughes and MacKenzie (2016) exploraram a variabilidade espacial nos tempos de espera pelo serviço da Uber (especificamente o Uber X) coletados através da API da empresa e realizaram estudos de seu relacionamento com indicadores socioeconômicos para a cidade de Seattle (EUA) no ano de 2015. Para isso, os autores aplicaram análises de regressão ponderadas localmente e modelos de regressão espacial globais e locais. Os resultados mostraram que os tempos de espera pelo serviço eram menores em áreas com maiores densidades populacionais e de empregos, enquanto esses tempos tendiam a ser maiores em regiões onde a renda média também é maior. Quanto às áreas com maior percentual de minorias, os tempos de espera tenderam a ser curtos durante o dia e longos durante a madrugada.

Wang and Mu (2018) examinaram as disparidades espaciais da acessibilidade para a cidade de Atlanta (EUA) em 2016, utilizando os tempos de espera de dois serviços da Uber: o Uber X (mais popular) e o Uber Black (de mais alto padrão). Os dados foram coletados também a partir da

API da Uber e a acessibilidade tratada é medida em termos da média e do desvio padrão dos tempos de cada serviço. Foram utilizados modelos autorregressivos espaciais para analisar a influência de variáveis socioeconômicas sobre a acessibilidade da Uber. Os resultados mostraram que a riqueza e a taxa de minorias não se associaram de maneira significativa à acessibilidade dos serviços, enquanto que maiores densidades populacionais e densidades da rede viária e menores tempos de deslocamento até o trabalho se relacionaram com maiores valores de acessibilidade.

É importante observar que a maioria dos trabalhos mencionados abrange regiões desenvolvidas, havendo uma escassez na literatura de estudos voltados para regiões em desenvolvimento, onde se supõe que o alcance dos ODS seja uma demanda de maior urgência. Pensando nisso, este trabalho aplica sua metodologia em um município situado na região nordeste do Brasil, que é um país em desenvolvimento. A principal contribuição, portanto, reside em propor um indicador de habitabilidade composto para o município brasileiro de Natal (RN), que combina dados tradicionais, como os do Censo Demográfico 2010, com os dados da Uber. Apesar da especificidade para Natal (RN), a abordagem pode ser expandida para outros municípios.

3. METODOLOGIA

A metodologia adotada neste trabalho possui duas fases: a preparatória e a de implementação (Figura 1). A fase preparatória envolve as etapas de coleta, limpeza e análise dos dados, enquanto a fase de implementação compreende a criação do indicador composto.

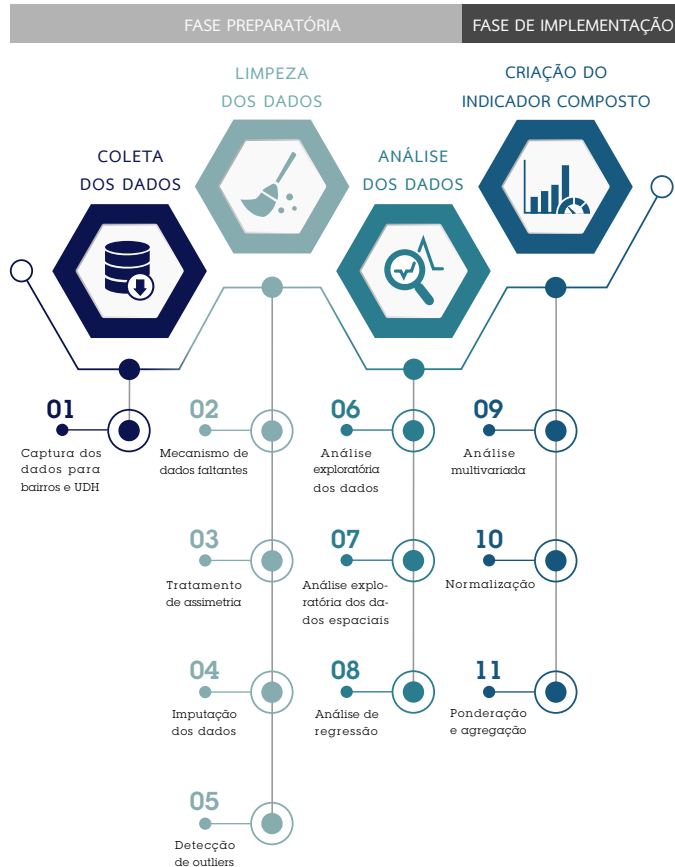


Figura 1. Visão geral da metodologia aplicada.

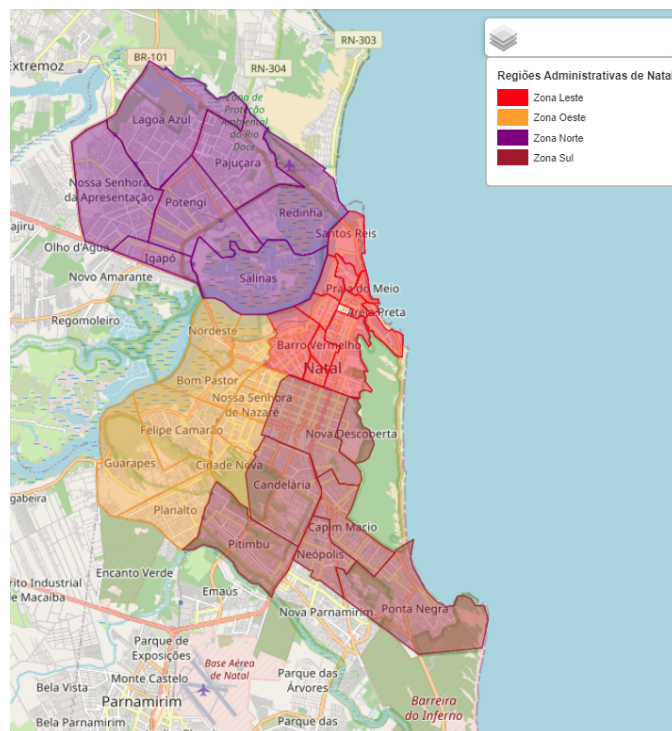


Figura 2. Mapa administrativo de Natal (RN). Fonte: Pereira et al. (2019); Google (2021).

3.1 Coleta dos Dados

A coleta de dados é realizada por uma rotina em Python, que faz requisições para a API de solicitações de viagem da Uber (<https://developer.uber.com/>), considerando um conjunto de localizações de embarque no município brasileiro de Natal (RN). As divisões administrativas do município são exibidas na Figura 2. O período de tempo analisado é de fevereiro de 2018 à maio de 2019, abrangendo os serviços da Uber disponíveis na cidade: *Uber X* (mais popular) e *Uber Select* (padrão mais elevado). Os dados coletados foram as Estimativas de Tempo até o Embarque (ETE), que representam a estimativa do tempo em segundos que o motorista parceiro leva para chegar ao local de embarque selecionado pelo usuário. Os dados coletados são então agregados à nível de Unidades de Desenvolvimento Humano (UDH) para posterior análise (ver Tabela 1).

Tabela 1. Amostra dos dados das ETE por UDH para Natal de 02/2018 à 05/2019.

Data/Hora	UDH	Região	ETEX ^a	ETES ^b
2018-05-15 09:40:00	1240810200035	east	140,0	153,0
2019-05-08 00:20:00	1240810200056	north	547,0	507,0
2018-10-19 12:30:00	1240810200016	south	190,0	320,0
2018-12-04 20:50:00	1240810200007	north	120,0	360,0
2018-10-17 09:20:00	1240810200021	south	120,0	240,0

^a ETE para o Uber X; ^b ETE para o Uber Select.

3.2 Limpeza dos Dados

A etapa de limpeza dos dados é crucial durante a fase preparatória. Nela, busca-se avaliar a qualidade dos dados coletados, determinar a quantidade e o padrão dos dados faltantes, além de se verificar a existência de *outliers*. A

identificação e a caracterização do mecanismo de ausência de dados auxilia na escolha da estratégia de imputação, que pode se dar por extirpação de dados, imputação simples ou imputação múltipla. A análise da distribuição dos dados para tratar possíveis assimetrias, por sua vez, evita imprecisões nas estimativas geradas durante esse tratamento de dados omissos. Ao final da etapa de limpeza, o conjunto de dados torna-se propício para os passos subsequentes de processamento.

3.3 Análise dos Dados

A última etapa da fase preparatória é a análise dos dados, que inclui a Análise Exploratória dos Dados (AED), a Análise Exploratória dos Dados Espaciais (AEDE) e a Análise de Regressão. Na AED, busca-se obter uma visão geral dos dados, avaliando o comportamento do serviço da Uber de acordo com as influências do ambiente e em diferentes épocas, além de sua correlação com variáveis socioeconômicas. Na AEDE são aplicados métodos da estatística espacial, a fim de verificar se há relação entre as ETE próximas espacialmente e se existe a formação de agrupamentos. A Análise de Regressão objetiva criar modelos que evidenciem como os dados socioeconômicos influenciam as ETE. As variáveis socioeconômicas que passam as análises de dados são descritas na Tabela 2. Em geral, a ideia desta etapa é verificar o grau de sensibilidade dos dados da Uber ao contexto e sua relação com atributos ligados à dimensão da habitabilidade.

Tabela 2. Variáveis socioeconômicas.

Variável	Descrição
PESOTOT ^a	População residente total
GINI ^a	Índice de Gini
ESPVIDA ^a	Esperança de vida ao nascer
RAZDEP ^a	Razão entre a população dependente e a potencialmente ativa
RDPC ^a	Renda per capita média
T_ENV ^a	Taxa de envelhecimento
IDHM ^a	Índice de Desenvolvimento Humano Municipal
IDHM_E ^a	IDHM - Dimensão Educação
IDHM_L ^a	IDHM - Dimensão Longevidade
IDHM_R ^a	IDHM - Dimensão Renda
ROUBOS_CRR ^b	Ocorrências de roubo de carros em 2018

^a Fonte dos dados: Atlas do Desenvolvimento Humano no Brasil (Baseado em dados do Censo Demográfico 2010).

^b Fonte dos dados: Observatório da Violência do Rio Grande do Norte - OBVIO (2018).

As principais bibliotecas utilizadas para manipulação e visualização de dados foram, respectivamente, *Pandas* (McKinney, 2010) e *matplotlib* (Hunter, 2007), ambas do Python. Durante a AEDE foram utilizadas as bibliotecas do Python *geopandas* (Jordahl et al., 2019), para a manipulação e visualização dos dados espaciais, e *pysal* (Rey and Anselin, 2007) para realizar a análise espacial.

3.4 Criação do Indicador Composto

Na fase de implementação se dá a etapa de criação do indicador composto que contempla alguns dos passos descritos por OECD et al. (2008). Tais passos são a análise multivariada, a normalização dos dados e a aplicação de métodos de ponderação e agregação.

A **análise multivariada** estuda a estrutura do conjunto de dados para identificar agrupamentos de indicadores similares. A técnica escolhida para essa etapa foi a Análise de Fatores (AF), que de acordo com OECD et al. (2008), objetiva descrever Q variáveis observadas x_1, x_2, \dots, x_Q em termos de um número menor de m fatores F_1, F_2, \dots, F_m não observados, evidenciando a relação entre as variáveis e esses fatores. Formalmente, isso significa que:

$$\begin{aligned} x_1 &= \alpha_{11}F_1 + \alpha_{12}F_2 + \dots + \alpha_{1m}F_m + e_1 \\ x_2 &= \alpha_{21}F_1 + \alpha_{22}F_2 + \dots + \alpha_{2m}F_m + e_2 \\ &\dots \\ x_Q &= \alpha_{Q1}F_1 + \alpha_{Q2}F_2 + \dots + \alpha_{Qm}F_m + e_Q \end{aligned} \quad (1)$$

Onde x_i ($i = 1, 2, \dots, Q$) são as variáveis observadas normalizadas, $\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{im}$ são as cargas fatoriais da variável x_i , F_1, F_2, \dots, F_m são os m fatores comuns não-correlacionados e e_1, e_2, \dots, e_Q são os Q fatores independentes e identicamente distribuídos com média zero.

Figueiredo Filho and Silva Júnior (2010) descrevem a AF em três passos: i) verificação da adequabilidade do conjunto de dados; ii) determinação e aplicação da técnica de extração e do número de fatores a ser extraído; iii) escolha do tipo de rotação dos fatores. Segundo os autores, no primeiro passo deve-se averiguar se o tamanho da amostra é razoável, se o modelo de AF é adequado ao tratamento dos dados e se as variáveis são bem correlacionadas.

Quanto à escolha do método de extração dos fatores, neste trabalho optou-se pela Análise de Componentes Principais (ACP). Para a definição do número de fatores escolhidos, as três práticas de OECD et al. (2008) foram adotadas: (i) os autovalores resultantes da ACP serem maiores que uma unidade; (ii) os autovalores também contribuam individualmente em mais de 10% com a variância geral; (iii) e os autovalores também terem, cumulativamente, uma contribuição superior a 60% da variância geral.

O último passo da AF é a rotação dos fatores, cujo objetivo é simplificar a interpretação do resultado obtido (Costello and Osborne, 2005). De acordo com OECD et al. (2008), o método mais usado para a rotação é o chamado *varimax*, sendo ele o adotado neste trabalho. Ao final da AF, obtém-se um conjunto reduzido de fatores que resumem as variáveis observadas.

A **normalização** é aplicada tanto na AF, como anteriormente às etapas de ponderação e agregação do indicador. Para este trabalho, o método de *z-score* foi adotado por ser pré-requisito para a AF e o método *Min-Max* foi o escolhido para ser aplicado antes das etapas de ponderação e agregação, por ser um dos mais utilizados na literatura, empregado inclusive pelo IDH (OECD et al., 2008).

A **ponderação** diz respeito à atribuição de pesos para cada indicador individualmente, enquanto a **agregação** finaliza a criação do indicador composto (OECD et al., 2008). Como durante a etapa de análise multivariada optou-se por uma Análise de Fatores, o método de ponderação também se baseará nessa técnica e será referenciado como ACP/AF. Basicamente, essa técnica utiliza a matriz de cargas fatoriais resultante da AF para calcular os pesos, conforme descrito por OECD et al. (2008).

A técnica de agregação aplicada é híbrida (Talukder et al., 2017) e utiliza, inicialmente, a agregação geométrica, for-

malmente descrita pela Equação 2. Para os casos em que a agregação geométrica retornar zero, aplica-se a agregação linear, descrita pela Equação 3. Em ambas as equações, IC é o Indicador Composto, n é o número de indicadores individuais, x_i representa cada indicador individual e w_i é o peso associado a cada um deles, de modo que $\sum_i w_i = 1$.

$$IC = \prod_{i=1}^n x_i^{w_i} \quad (2) \quad IC = \sum_{i=1}^n w_i x_i \quad (3)$$

4. RESULTADOS

Após a etapa de coleta, foi obtido um conjunto de dados com 4.049.799 observações com as médias das ETE para as UDH da cidade natalense e para ambos os serviços da Uber disponíveis. Quanto à etapa de limpeza, primeiramente observou-se cerca de 36,6104% de dados faltantes para a ETEX e 39,1628% para a ETES. Como a ausência de dados nas variáveis das ETE da Uber pode estar relacionada aos dados observados em diferentes turnos do dia e para UDH distintas, pressupôs-se o MAR (*Missing At Random*) como mecanismo de ausência de dados. Feito isso, uma correção de assimetria na distribuição das ETE foi realizada, aplicando-se a função logarítmica para prepará-las para a imputação de dados. Dado que o mecanismo de ausência foi o MAR e que há mais de 10% de dados faltantes, a técnica de imputação múltipla através de uma regressão linear Bayesiana foi escolhida. Para implementar essa estratégia de imputação, utilizou-se a biblioteca *MICE* (Van Buuren and Groothuis-Oudshoorn, 2011) do R. Concluída a imputação, a transformação logarítmica foi revertida e os *outliers* foram removidos.

Na etapa de análise de dados, a hipótese de que os dados da Uber podem retratar a habitabilidade ganhou força. Observou-se que esses dados acompanham as características da dinâmica urbana durante a AED, tendem a formar agrupamentos espaciais na AEDE e apresentam associações com fatores sociais, econômicos e infraestruturais na análise de regressão. Esses resultados podem ser visualizados em mais detalhes em Oliveira (2020). Estudos preliminares deste trabalho, considerando os bairros de Natal também são descritos em Bezerra et al. (2019).

Chega-se então à fase de implementação, que utiliza os dados da Uber e os indicadores listados na Tabela 2. Inicialmente, a AF foi empreendida e dois fatores foram extraídos da matriz de cargas fatoriais obtida. O Fator 1 é composto pelo IDH e suas dimensões, além da ESPVIDA, GINI e RDPC. Portanto, será nominado como Desenvolvimento Humano. O Fator 2 é composto pelas médias das ETEX e ETES por UDH e, portanto, será entendido como Qualidade do Serviço da Uber.

A Figura 3 apresenta a distribuição das UDH de Natal (identificadas pelos dois últimos dígitos de seus códigos) a partir da interação entre esses dois fatores. É interessante observar que, no quadrante superior direito, onde os dois fatores de Desenvolvimento Humano e Qualidade do Serviço da Uber são acima da média, há uma concentração de UDH que fazem parte das Regiões Administrativas Sul e Leste. Por outro lado, no quadrante inferior esquerdo, onde ambos os fatores são abaixo da média, a maior parte das UDH são das Regiões Administrativas

Norte e Oeste. No quadrante superior esquerdo, onde o Desenvolvimento Humano é abaixo da média, mas a Qualidade do Serviço da Uber é acima da média, concentram-se predominantemente UDH da Região Leste em que há intensa movimentação comercial. Por fim, no quadrante inferior direito, onde o Desenvolvimento Humano é acima da média e a Qualidade do Serviço da Uber é abaixo da média, observam-se UDH em geral circundadas ou na vizinhança de áreas mais vulneráveis em termos dos outros dois fatores.

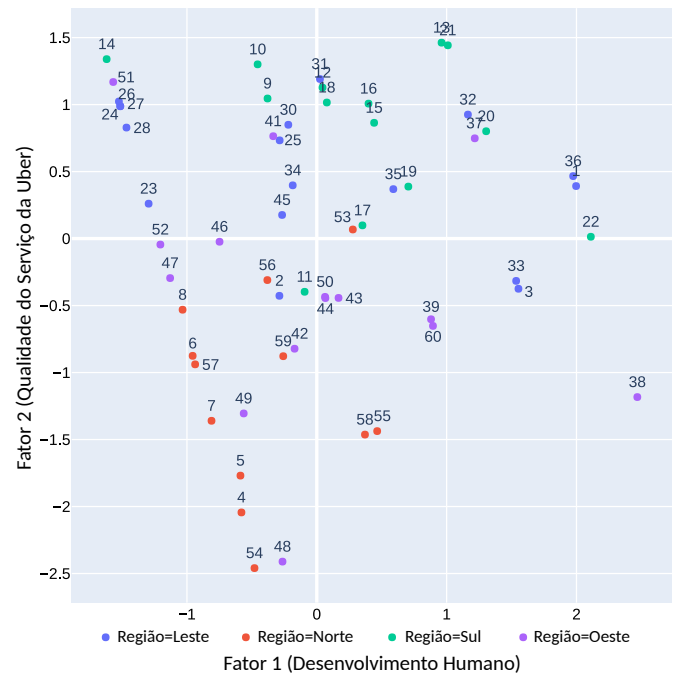


Figura 3. Distribuição das UDH considerando o Desenvolvimento Humano e a Qualidade do Serviço da Uber.

Antes de partir para a etapa final de ponderação e agregação, os dados foram normalizados com o método *Min-Max*. Em seguida, aplicou-se a técnica ACP/AF para a obtenção dos pesos a partir da matriz de cargas fatoriais e, por fim, as variáveis foram agregadas pelo método híbrido para compor o indicador de habitabilidade. A Figura 4 exibe os valores dos indicadores de habitabilidade compostos para as UDH. Quanto mais vermelha for a área, menor é o valor do indicador e, quanto mais azul, maior é o indicador.

Observa-se que a UDH 14 foi a que apresentou a pior medida para o indicador de habitabilidade proposto (0,0726), enquanto a UDH 38 foi a que exibiu o melhor valor (0,8408). O indicador criado avalia a habitabilidade em termos da dimensão do desenvolvimento humano, considerando as facetas de educação, renda e longevidade, e em termos da qualidade do serviço da Uber.

O mapa coroplético da Figura 4 expõe, de certa forma, que o município natalense é segregado espacialmente, quanto às condições de habitabilidade de seus moradores. Enquanto a maior parte das UDH das regiões Sul e Leste, que são as mais abastadas de Natal, se sobressaem com melhores indicadores, as UDH situadas predominantemente nas periferias das regiões Norte e Oeste são mais vulneráveis nesse ponto.

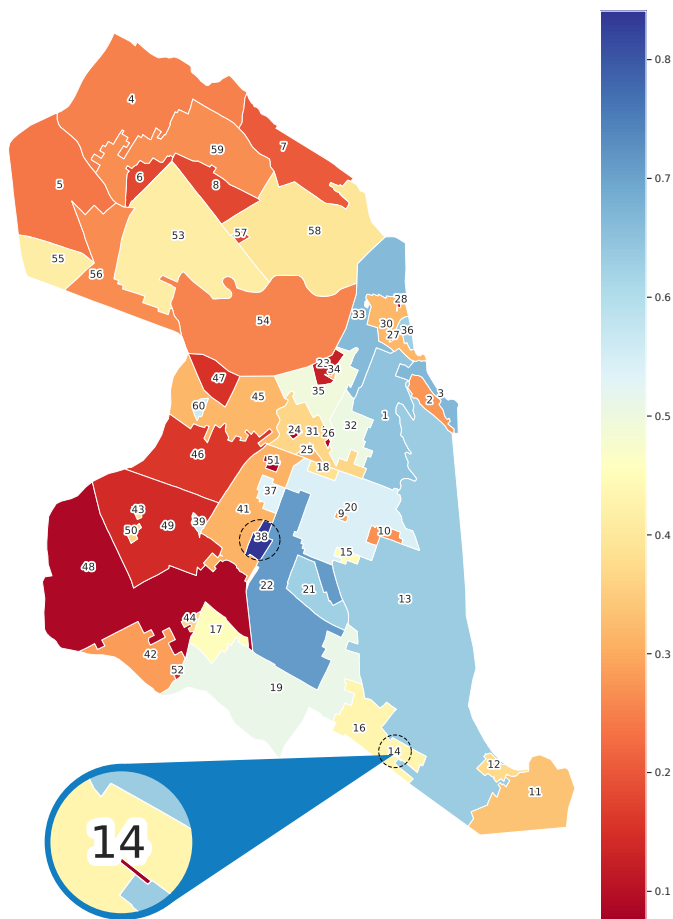


Figura 4. Mapa coroplético de Natal (RN), com as UDH coloridas conforme o valor do indicador composto.

5. CONCLUSÃO

Neste trabalho foi proposta uma metodologia orientada a dados para a criação de um indicador de habitabilidade composto para o município brasileiro de Natal, capital do Rio Grande do Norte e uma das grandes cidades do país. Esse indicador combina tanto dados obtidos tradicionalmente, como informações censitárias, quanto os que advêm de fontes alternativas, como os dados da API da Uber, que neste trabalho correspondem às Estimativas de Tempo até o Embarque. A abordagem proposta tratou das etapas de coleta, limpeza (manipulação) e análise de dados, além dos passos para a criação do indicador composto. O indicador de habitabilidade criado revelou uma predominância de melhores condições de habitabilidade nas UDH das Regiões Administrativas Sul e Leste, enquanto as Regiões Administrativas Norte e Oeste tendem a vivenciar as piores experiências nesse sentido. Além disso, a criação do indicador evidenciou o potencial das fontes de dados alternativas em ser tão consistentes quanto os dados tradicionais, abrindo um campo de possibilidades de descrição da realidade por meio de dados mais interativos, atualizados, dinâmicos e para pequenas áreas como as UDH.

Este trabalho trouxe como principais contribuições a criação de um indicador de habitabilidade para o município de Natal por meio do uso de dados de uma tecnologia recente, o emprego de uma estrutura própria de coleta de

dados de baixo custo e a aplicação de uma metodologia que pode ser replicada para outros centros urbanos. Quanto às suas limitações, pode-se ressaltar a impossibilidade de estimação dos pesos dos indicadores quando os dados não são bem correlacionados, desvantagem advinda do método de ponderação baseado em Análise de Fatores. Como trabalhos futuros, pretende-se melhorar alguns pontos metodológicos pelo emprego de outras técnicas estatísticas, bem como atualizar o indicador com os dados provenientes do próximo Censo Demográfico ou de outras fontes de informação. Vale salientar que, embora vários dos indicadores sociodemográficos utilizados sejam provenientes do Censo Demográfico 2010, em um período anterior ao dos dados da Uber, alguns dados são menos susceptíveis a alterações bruscas ao longo do tempo. Assim, mesmo havendo um distanciamento temporal, os dados sociodemográficos utilizados são suficientes para fornecer uma dimensão aproximada da realidade, e assim como os dados da Uber, mostraram coerência com a descrição da realidade local.

REFERÊNCIAS

- Barnett, C. and Parnell, S. (2016). Ideas, implementation and indicators: epistemologies of the post-2015 urban agenda. *Environment and Urbanization*, 28(1), 87–98.
- Bezerra, A., Alves, G., Silva, I., Rosati, P., Endo, P.T., and Lynn, T. (2019). A Preliminary Exploration of Uber Data as an Indicator of Urban Liveability. In *2019 International Conference on Cyber Situational Awareness, Data Analytics And Assessment (Cyber SA)*, 1–8. IEEE.
- Carmo, R.L.d., Dagnino, R.d.S., and Johansen, I.C. (2014). Transição demográfica e transição do consumo urbano de água no Brasil. *Revista Brasileira de Estudos de População*, 31(1), 169–190.
- Carvalho, C.H.R.d. (2016). *Mobilidade Urbana Sustentável: conceitos, tendências e reflexões*. Ipea.
- Costello, A.B. and Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical assessment, research, and evaluation*, 10(1), 7.
- Figueiredo Filho, D.B. and Silva Júnior, J.A.d. (2010). Visão além do alcance: uma introdução à análise fatorial. *Opinião pública*, 16(1), 160–185.
- Google (2021). Natal. URL <https://www.google.pt/maps/search/natal/@-5.8002645,-35.2221761,12z/data=!3m1!4b1>.
- Hughes, R. and MacKenzie, D. (2016). Transportation network company wait times in Greater Seattle, and relationship to socioeconomic indicators. *Journal of Transport Geography*, 56, 36–44.
- Hunter, J.D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in science & engineering*, 9(3), 90.
- IBGE (2021). Instituto Brasileiro de Geografia e Estatística. IBGE Cidades. <https://cidades.ibge.gov.br/>. Accessed: 2021-05-20.
- Irwin, N. (2003). Urban transportation: An essential factor. In *Annual Conference of the Transportation Association of Canada*.
- Jin, S.T., Kong, H., and Sui, D.Z. (2019). Uber, Public Transit, and Urban Transportation Equity: A Case Study in New York City. *The Professional Geographer*, 71(2), 315–330.

- Jordahl, K., den Bossche, J.V., Wasserman, J., McBride, J., Gerard, J., Tratner, J., and et al. (2019). *geopandas/geopandas: v0.4.1*. doi:10.5281/zenodo.2585849.
- Lowe, M., Whitzman, C., Badland, H., Davern, M., Aye, L., Hes, D., Butterworth, I., and Giles-Corti, B. (2015). Planning healthy, liveable and sustainable cities: How can indicators inform policy? *Urban Policy and Research*, 33(2), 131–144.
- Lucas, K. (2012). Transport and social exclusion: Where are we now? *Transport Policy*, 20.
- Mariano, A.M., Ramírez-Correa, P., Alfaro-Pérez, J., Painén-Aravena, G., and Machorro-Ramos, F. (2019). O Papel da Aceitação da Tecnologia nas Cidades Inteligentes: Um estudo das percepções dos usuários do Uber Brasil. *Revista Ibérica de Sistemas e Tecnologias de Informação*, 17, 571–583.
- McKinney, W. (2010). Data Structures for Statistical Computing in Python. In S. van der Walt and J. Millman (eds.), *Proceedings of the 9th Python in Science Conference*, 51 – 56.
- OECD, European Union, and Joint Research Centre - European Commission (2008). *Handbook on Constructing Composite Indicators: Methodology and User Guide*. OECD Paris.
- Oliveira, G.L.A.d. (2020). *Uma abordagem orientada a dados para a criação de um indicador de habitabilidade baseado na API da UBER*. Master's thesis, UFRN.
- Pereira, R., Gonçalves, C., et al. (2019). *geobr: Loads Shapefiles of Official Spatial Data Sets of Brazil*. URL <https://github.com/ipeaGIT/geobr>.
- Pupphachai, U. and Zuidema, C. (2017). Sustainability indicators: A tool to generate learning and adaptation in sustainable urban development. *Ecological Indicators*, 72, 784–793.
- Rey, S.J. and Anselin, L. (2007). PySAL: A Python Library of Spatial Analytical Methods. *The Review of Regional Studies*, 37(1), 5–27.
- Satterthwaite, D. (2016). A new urban agenda? *Environment and Urbanization*, 28(1), 3–12.
- Talukder, B., W Hipel, K., W vanLoon, G., et al. (2017). Developing composite indicators for agricultural sustainability assessment: Effect of normalization and aggregation techniques. *Resources*, 6(4), 66.
- Telésforo, R.L. (2016). *Uber: Inovação disruptiva e ciclos de intervenção regulatória*. Master's thesis, Escola de Direito do Rio de Janeiro da Fundação Getúlio Vargas.
- Tribuna do Norte (2018). Número de passageiros de ônibus em Natal cai 22% em cinco anos. URL <http://www.tribunadonorte.com.br/noticia/naomero-de-passageiros-de-a-nibus-em-natal-cai-22-em-cinco-anos/423417>.
- Uber Technologies Inc. (2019). Fatos e Dados sobre a Uber. URL <https://www.uber.com/pt-BR/newsroom/fatos-e-dados-sobre-uber/>.
- United Nations (2016). New urban agenda.
- United Nations (2018). World urbanization prospects: The 2018 revision. Research report, United Nations, Department of Economic and Social Affairs,.
- Valcárcel-Aguiar, B., Murias, P., and Rodríguez-González, D. (2019). Sustainable urban liveability: A practical proposal based on a composite indicator. *Sustainability*, 11(1), 86.
- Van Buuren, S. and Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, 45(3), 1–67. URL <https://www.jstatsoft.org/v45/i03/>.
- Wang, M. and Mu, L. (2018). Spatial disparities of Uber accessibility: An exploratory analysis in Atlanta, USA. *Computers, Environment and Urban Systems*, 67, 169–175.
- Wey, W.M. and Huang, J.Y. (2018). Urban sustainable transportation planning strategies for livable city's quality of life. *Habitat International*, 82, 9–27.