**Titulo**

**RESUMO**

**Palavras-chave**: *Mobilidade urbana, Simulação de dados, Big Data Analytics, Estatística Espacial.*

# 

# **1. INTRODUÇÃO**

**1.1. Problema da Pesquisa**

**1.2. Justificativas**

**2. REVISÃO DE LITERATURA**

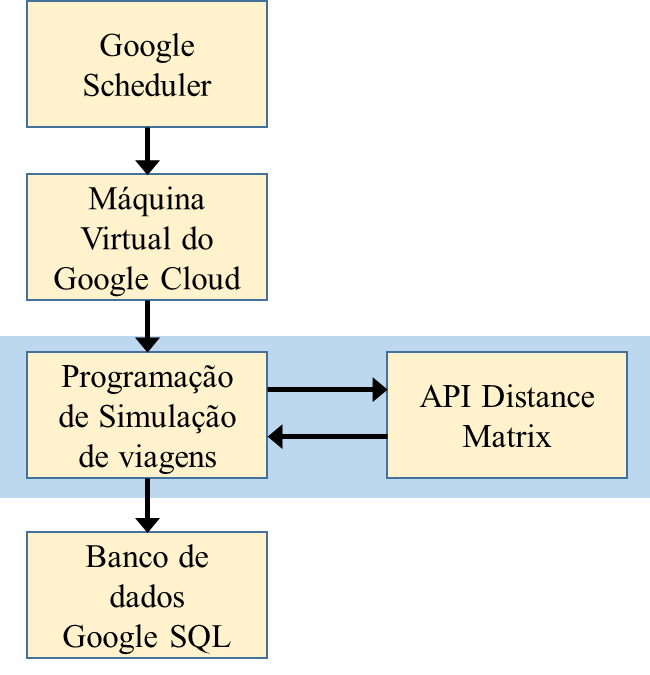
**2.1. *Big Data***

**2.2. Mobilidade Urbana**

# **3. MÉTODO**

**3.1. Simulação de viagens**

A simulação foi feita em duas etapas: primeiro, a geração de um banco de endereços, e depois, a simulação das viagens propriamente ditas. Algumas considerações precisam ser feitas. A relativa alta complexidade de simulações locais que consigam captar o comportamento em tempo real – com informações de trânsito – da mobilidade implicou o uso de alguma ferramenta de previsão de tempos de viagem já estabelecida e acessível por meio remoto. A escolha feita (pela API Distance Matrix da Google) implicou em um número limitado de requisições de viagens, por questões de custo; essa limitação levou a escolhas para reduzir o número de viagens “perdidas” na simulação, incorridas quando as coordenadas usadas na API não correspondiam ou não podiam ser aproximadas a endereços válidos, como no caso de coordenadas nas represas de São Paulo. Ao mesmo tempo, foi feita a opção por usar a computação em nuvem para a simulação, o que levou ao esforço de reduzir a computação necessária para evitar problemas relacionados ao desempenho. Essas limitações definiram o processo de definição de endereços. Procurou-se sortear endereços em regiões mais densamente povoadas para evitar possíveis perdas, seguindo um modelo de densidade de probabilidade da população. Ao mesmo tempo, para reduzir o esforço computacional, foi montada uma base de coordenadas *offline*, que foi usada para sortear os endereços das viagens. O processo de geração da base foi feito nas seguintes etapas: 1) O mapa do município de São Paulo (em formato *shapefile*) foi dividido por uma grade com quadrículas de 500 metros de lado; 2) Foram calculadas as populações de cada quadrícula com dados do Censo Demográfico de 2010 e retiradas as quadrículas com população igual a zero; 3) As quadrículas foram divididas em quintis de densidade populacional, e sorteados aleatoriamente pontos geográficos dentro de cada quadrícula, de acordo com o quintil: 5 pontos para o quintil mais populoso, 4 para o 2º quintil, 3 para o 3º, 2 para o 2º e 1 ponto para o quintil menos populoso. 4) O conjunto de pontos resultante foi usado como base para o sorteio dos endereços de origem. Essa primeira etapa foi realizada utilizando bases cartográficas abertas do município de São Paulo e o software aberto QGIS. A simulação das viagens foi feita a partir de um programa desenvolvido em Python, executado no serviço de computação em nuvem da Google. A estrutura do programa seguiu a arquitetura apresentada na Figura 1. Foi usada uma ferramenta de agendamento (Google Scheduler) de ativação ligada a uma máquina virtual no ambiente em nuvem da Google, para que a chamada ao serviço de viagem do Google Maps fosse realizada nos dias úteis da semana, a cada hora cheia, das cinco da manhã até às nove da noite. A intenção do espaçamento era obter amostragens de viagens em diferentes horários para comparar periodicidades diárias e horárias nas viagens. O programa seguiu as etapas descritas na Figura 1.

* Abrir uma conexão com o banco de dados SQL da nuvem da Google
* Carregar o banco de coordenadas
* Sortear dez coordenadas de origem e dez coordenadas de destino
* Chamar a API Distance Matrix com as dez origens e os dez destinos, para viagens de transporte público
* Processar os resultados devolvidos pela API e armazenar em um vetor auxiliar
* Chamar novamente a API Distance Matrix com as mesmas dez origens e os dez destinos, para viagens de transporte privado
* Processar os resultados devolvidos pela API e anexar ao vetor auxiliar
* Submeter o vetor auxiliar à função que insere os dados no Banco de Dados hospedado na nuvem.

**Figura 1.**Etapas e Estrutura funcional do programa elaborado para simular os dados  
Fonte: Os autores

Cada chamada da API Distance Matrix retornou uma lista com duzentas viagens com dados de horário e dia da semana, coordenadas da origem e do destino, endereços da origem e do destino da viagem, duração, distância e tarifa da viagem. Para cada par origem-destino houve registro de viagem de carro e de transporte público. O período de simulação foi entre os dias 11 de fevereiro de 2019 a 5 de junho de 2019. O total de viagens armazenadas no banco de dados nesse período foi de 257.400 viagens, sendo 253.450 viagens válidas – 128.700 (100% de aproveitamento) das viagens de carro e 126.725 das viagens transporte público (98,47% de aproveitamento).

**3.2. Estrutura Analítica do Banco de Dados**

# **4. RESULTADOS E ANÁLISE**

# **5. DISCUSSÃO FINAL E PRINCIPAIS CONCLUSÕES**

# **REFERÊNCIAS BIBLIOGRÁFICAS**