

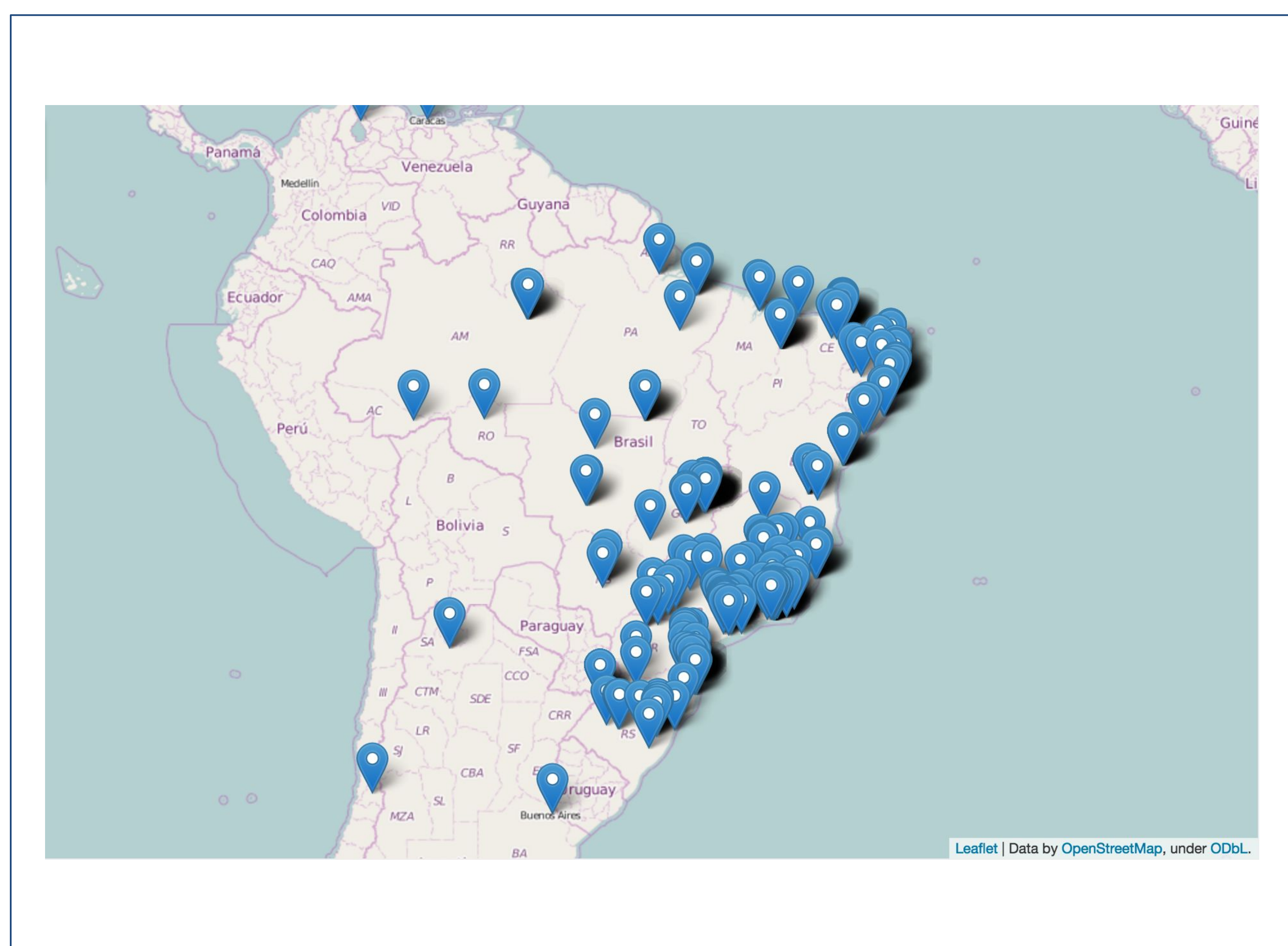
## MINERAÇÃO DE DADOS APLICADO À REDE SOCIAL *TWITTER* UTILIZANDO A LINGUAGEM DE PROGRAMAÇÃO PYTHON

Thiago M. de SOUZA; Valmei ABREU Júnior valmeijr@gmail.com

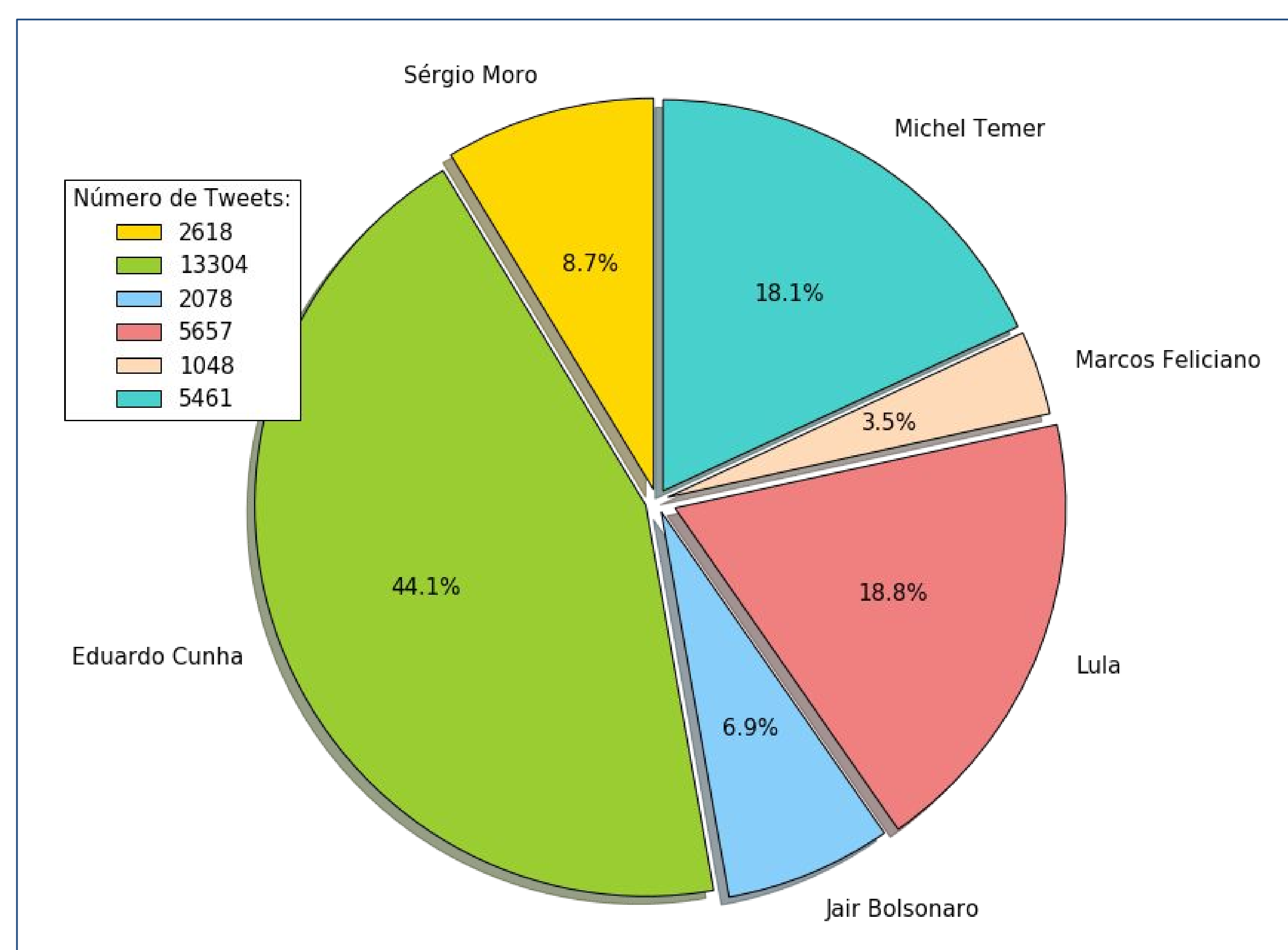
### Resumo

**Introdução:** A rede social *Twitter* é caracterizada como um serviço de microblog que permite uma breve comunicação entre pessoas. Devido a este aspecto *Twitter* é utilizado para troca de mensagens imediatas, onde uma enorme quantidade de dados é gerado por interações de usuários. A grande proporcionalidade destes dados é disponibilizado pela rede social permitindo, então, a sua análise com o intuito de gerar conhecimento útil. **Objetivo:** Utilizar técnicas e algoritmos de *data mining*, para análise e mineração de dados provenientes da rede social *Twitter*, utilizando os recursos e bibliotecas que a linguagem de programação Python possui. Apresentar gráficos e planilhas para a visualização e interpretação dos dados minerados. **Métodos:** Foi desenvolvido uma solução utilizando Python, que se beneficia da API do *Twitter*, para coletar todas as publicações referentes a *hashtag* '#ImpeachmentDay' durante o dia de votação no Congresso Brasileiro, para a continuidade do processo de Impeachment da Presidente Dilma Rousseff. Após a coleta de dados, fez-se necessário a limpeza dos dados, onde é eliminado informações desnecessárias como *dirty data*. **Resultados:** Através da implementação de filtros e técnicas de mineração de dados é possível encontrar alguns padrões e outras informações proveniente das publicações. Os idiomas e os países que trazem mais *tweets*, as *hashtags* mais utilizadas e os nomes mais mencionados, são ilustrados através de gráficos onde se quantifica em números e porcentagens a disponibilidade dos *tweets*. Os picos de horários em que houveram a maior ocorrência de *tweets* e a localização geográfica de algumas dessas publicações também são apresentadas neste trabalho. **Conclusão:** A possibilidade de utilizar a linguagem Python desde a coleta de *tweets*, autenticação da rede social através da sua API, a normalização dos dados e posteriormente filtrá-los e apresentá-los de diversas maneiras para visualização, torna a linguagem uma excelente ferramenta para as tarefas de mineração de dados.

**Palavras chaves:** data mining, twitter, python.



Localização geográfica de *tweets*  
Fonte: Elaborado pelo autor



Nomes relevantes com mais *tweets*  
Fonte: Elaborado pelo autor

### REFERÊNCIAS BIBLIOGRÁFICAS

- HAN, J. et al. **Data Mining: Concepts and Techniques**. Elsevier, 2012.
- MCKINNEY, W. **Python for Data Analysis**. O'Reilly, 2013.
- RUSSELL, M. A. **Mining the Social Web**. O'Reilly Media, 2013. ISBN 978-1- 449-36761-9.
- SFERRA, H. H.; CORREA, A. M. C. J. **Conceitos e aplicações de data mining**. 2003.