# Mobilidade Urbana (São Paulo) com PySpark

O projeto **“Mobilidade Urbana”** propõe o desenvolvimento de um pipeline de análise de dados voltado à compreensão dos padrões de deslocamento na cidade de São Paulo contemplando a construção de processos de **ETL**, manipulação e exploração de dados por meio de **consultas SQL**, a aplicação de **métodos de visualização de dados** e a implementação de **Structured Streaming** para lidar com fluxos de informação em tempo quase real. O produto final será composto por um **Notebook técnico** com a implementação do processamento, um **Relatório analítico** com os resultados alcançados e uma **Apresentação executiva** destinada à comunicação dos achados para diferentes públicos.

**Público (stakeholders)**

**Justificativa**

A mobilidade urbana é um dos maiores desafios enfrentados por metrópoles contemporâneas, em especial em São Paulo, onde a densidade populacional, a extensão territorial e a alta demanda por transporte público e privado geram sobrecarga nas infraestruturas existentes. Nesse contexto, a análise de grandes volumes de dados oriundos de registros de transporte público, aplicativos de mobilidade e sensores urbanos representa uma oportunidade estratégica para compreender padrões de deslocamento, identificar gargalos e apoiar a formulação de políticas públicas mais eficientes.

Do ponto de vista tecnológico, a escolha pelo PySpark se justifica pela sua robustez no processamento distribuído de dados em larga escala, pela integração nativa com linguagens de consulta SQL e pela capacidade de lidar tanto com cargas batch quanto streaming, oferecendo um ambiente adequado para análises complexas e cenários de dados dinâmicos. O projeto, portanto, combina relevância social com inovação tecnológica, criando subsídios práticos e teóricos para aplicações futuras em Big Data Analytics voltadas à mobilidade urbana.

**Objetivo**

Desenvolver um ecossistema de análise de Big Data, baseado em PySpark, para identificar e interpretar padrões de mobilidade urbana na cidade de São Paulo, utilizando abordagens batch e streaming, com foco em subsidiar estudos e decisões estratégicas relacionados à gestão de transporte e deslocamento urbano.

**Objetivos Específicos**

1. Construir um pipeline de ETL batch para ingestão, tratamento e integração de dados de mobilidade.
2. Aplicar consultas SQL distribuídas para a exploração e análise dos dados transformados.
3. Desenvolver visualizações analíticas que permitam interpretar os padrões de deslocamento de forma clara e objetiva.
4. Implementar Structured Streaming para monitorar fluxos de dados em tempo quase real, possibilitando análises contínuas.
5. Produzir documentação técnica (Notebook) e relatórios que consolidem as evidências obtidas e suportem a tomada de decisão em diferentes níveis.

**Requisitos funcionais**

**Requisitos Não Funcionais**