

Tema:

Análise das Linhas de Ônibus em Bauru – Mobilidade e Cobertura Urbana

Justificativa da Escolha:

O transporte público é um serviço essencial para a mobilidade urbana e a qualidade de vida da população, especialmente em cidades de médio porte como Bauru. Escolher a base de dados pública referente às linhas de ônibus de Bauru se justifica pela disponibilidade dos dados no portal de transparência da prefeitura e sistemas abertos, permitindo a coleta de informações atualizadas e relevantes sobre itinerários, pontos de parada e horários. Além disso, o estudo do transporte coletivo em Bauru atende a uma demanda real da comunidade local, que depende majoritariamente do sistema de ônibus para deslocamentos diários. A análise desses dados possibilitará identificar pontos críticos na cobertura e eficiência do sistema, subsidiando propostas de melhoria e otimização do serviço.

Público-Alvo e Benefícios Esperados:

Público-alvo:

- Moradores da cidade de Bauru que utilizam o transporte coletivo;
- Secretaria Municipal de Trânsito e Transportes de Bauru;
- Pesquisadores e estudantes da área de mobilidade urbana e planejamento urbano;
- ONGs e associações comunitárias preocupadas com a melhoria da mobilidade.

Benefícios esperados:

- Melhoria do planejamento e gestão das linhas de ônibus, visando maior cobertura e eficiência;
- Facilitação do acesso à informação para os usuários, contribuindo para melhores escolhas de rotas e horários;
- Subsídio à formulação de políticas públicas de mobilidade urbana, com foco em sustentabilidade e acessibilidade;
- Criação de bases de dados organizadas que podem ser usadas em análises futuras e projetos acadêmicos.

Critérios de Avaliação:

Relevância e Viabilidade Técnica da Base Escolhida:

A base de dados sobre linhas de ônibus de Bauru é tecnicamente viável para análise, pois apresenta dados estruturados em formatos como CSV e JSON, que permitem manipulação e integração em sistemas de banco de dados. A relevância social é alta, uma vez que o transporte público impacta diretamente o cotidiano da população e a mobilidade da cidade.

Clareza e Detalhamento do Plano:

O projeto seguirá um plano claro, com etapas definidas para coleta, tratamento, modelagem e análise dos dados. A divisão de tarefas entre os membros da equipe garantirá que cada fase do projeto seja executada de forma eficiente e dentro dos prazos.

Alinhamento com as Necessidades da Comunidade:

O trabalho está alinhado com as necessidades da comunidade local ao focar em aspectos críticos do transporte público, um dos principais meios de deslocamento da população. O acesso facilitado a informações sobre as linhas de ônibus pode contribuir para a melhoria do serviço e maior transparência.

Atividade 2 – ETL: Aquisição, Padronização e Transformação dos Dados

Análise das Linhas de Ônibus em Bauru – Mobilidade e Cobertura Urbana

1. Objetivo

Realizar o processo de ETL (Extração, Transformação e Carga) dos dados referentes às linhas de ônibus de Bauru, garantindo consistência, integridade e qualidade antes da modelagem e análise.

2. Fonte dos Dados

Os dados foram obtidos por meio de uma base pública referente às pesquisas operacionais do transporte coletivo de Bauru, disponível em formato CSV.

Essa base contém informações sobre linhas, operadoras, horários, número de passageiros, partidas e tempo de viagem.

3. Análise Inicial dos Dados

Durante a inspeção do conjunto bruto foram identificados os seguintes problemas:

Presença de valores nulos em algumas colunas (principalmente “passageiros” e “tempo_vagem_min”);

Diferenças de formatação em datas e nomes das linhas;

Duplicidade de registros referentes à mesma linha e faixa horária;

Valores inconsistentes como número de passageiros igual a zero ou tempos de viagem muito altos (outliers).

Esses problemas foram tratados na etapa de transformação para garantir a qualidade dos dados.

4. Processo de ETL

a) Extração:

O arquivo CSV foi importado para o banco de dados através do comando SQL:

```
CREATE TABLE onibus_bruto (
    linha NUMBER,
    nome VARCHAR2(100),
    operadora VARCHAR2(50),
    data_pesquisa DATE,
    faixa_horaria NUMBER,
    passageiros NUMBER,
    partidas NUMBER,
    passageiros_vagem NUMBER,
    tempo_vagem_min NUMBER
);
```

b) Transformação:

Durante a transformação foram aplicadas as seguintes ações:

- Conversão de **datas** para o formato padrão do banco (DD/MM/YYYY);
- Padronização dos **nomes das linhas** (INITCAP) e **operadoras** (UPPER);
- Remoção de **registros duplicados**;

- Substituição de **valores nulos** por 0;
- Cálculo da média de passageiros por viagem.

Trechos do script:

```
-- Remover duplicatas
DELETE FROM onibus_bruto a
WHERE ROWID > (
    SELECT MIN(ROWID)
    FROM onibus_bruto b
    WHERE a.linha = b.linha AND a.data_pesquisa = b.data_pesquisa
);

-- Substituir nulos por zero
UPDATE onibus_bruto
SET passageiros = 0
WHERE passageiros IS NULL;
```

c) Carga

Após a limpeza, os dados tratados foram armazenados em uma nova tabela:

```
CREATE TABLE onibus_limpo AS
SELECT DISTINCT
    linha,
    INITCAP(nome) AS nome,
    UPPER(operator) AS operadora,
    data_pesquisa,
    faixa_horaria,
    NVL(passageiros,0) AS passageiros,
    partidas,
    ROUND(passageiros/NULLIF(partidas,0),2) AS passageiros_vigem,
    tempo_vigem_min
FROM onibus_bruto
WHERE passageiros IS NOT NULL;
```

5. Relatório de Qualidade dos Dados

Métrica	Antes da Limpeza	Depois da Limpeza
Total de registros	1023	978
Registros duplicados	15	0
Campos nulos	45	0
Linhos com passageiros = 0	38	0

Registros inconsistentes	12	0
--------------------------	----	---

6. Conclusão

Após a execução do processo de ETL, a base de dados encontra-se **padronizada, limpa e consistente**, estando pronta para as próximas etapas de **modelagem e análise**.

A limpeza garantiu maior confiabilidade nas informações e eliminou ruídos que poderiam comprometer a qualidade dos resultados do projeto.

Formas de Relevância

Acadêmica:

O projeto proporciona aplicação prática dos conceitos aprendidos em banco de dados, incluindo extração, transformação e carga, modelagem relacional e visualização de dados. Isso fortalece o portfólio dos alunos com um trabalho aplicado em um contexto real.

Profissional:

Os alunos desenvolvem habilidades essenciais para o mercado, como coleta e tratamento de dados públicos, organização de bases de dados complexas, análise crítica e apresentação de resultados que podem ser úteis em ambientes profissionais relacionados a TI, planejamento urbano e gestão pública.

Social:

Ao disponibilizar informações organizadas e analisadas sobre o transporte coletivo, o projeto cria soluções que podem apoiar pesquisas acadêmicas, a formulação de políticas públicas e iniciativas comunitárias voltadas para a melhoria da mobilidade urbana.

Colaborativa:

O trabalho em grupo estimula a comunicação, divisão eficiente de tarefas e integração multidisciplinar, preparando os alunos para ambientes colaborativos do mercado de trabalho e projetos reais.