

SENAC – FACULDADE DE TECNOLOGIA E INOVAÇÃO
CURSO – CIÊNCIA DE DADOS
DISCIPLINA – FUNDAMENTOS DA ESTATÍSTICA

PROJETO (BI)

Análise Estatística de Acidentes Aéreos no Brasil

ALUNOS: Paulo Cesar de Souza

Simone Cândida Rodrigues

Victor Miranda Veloso

DEFINIÇÃO DO PROJETO:

Nome do projeto: Análise Estatística de Acidentes Aéreos no Brasil

Arquivo para análise: V_OCORRENCIA_AMPLA.JSON ou CSV (Fonte: ANAC)

Colunas a serem analisadas: Numero_da_Ocorrencia; Descricao_do_Tipo; Data_da_Ocorrencia; Municipio; UF; Nome_do_Fabricante, Modelo.

OBJETIVOS GERAIS DO PROJETO:

Mostrar através de gráficos:

- Os modelos/fabricantes de aeronaves com o maior índice de ocorrências (quantidade);
- Os tipos de ocorrência com maior frequência;
- Os locais (cidades e unidades federativas) com maior ocorrência de acidentes (quantidade);
- Os modelos/fabricantes de aeronaves com maior quantidade por tipo de ocorrência;
- Os cálculos estatísticos: média, mediana, moda, desvio padrão (por dia, ano a ano).
- Outros questionamentos que se fizerem necessários durante a apresentação do projeto.

INTRODUÇÃO

O transporte aéreo, vital para a mobilidade e a economia global, inerentemente carrega consigo riscos que demandam análise e compreensão aprofundadas. No contexto brasileiro, a segurança da aviação civil é um tema de relevância tanto para órgãos reguladores quanto para a sociedade em geral. A ocorrência de acidentes aéreos, embora relativamente rara, possui um impacto significativo, suscitando a necessidade de investigações estatísticas rigorosas para identificar padrões, fatores de risco e áreas que demandam maior atenção em termos de prevenção.

Nesse sentido, o presente projeto de Estatística (BI) propõe uma análise detalhada das ocorrências aeronáuticas no Brasil, utilizando dados estruturados em um banco de dados MySQL Workbench. Essa abordagem permitirá uma gestão eficiente e organizada das informações provenientes da Força Aérea Brasileira (FAB) por meio do Centro de Investigação e Prevenção de Acidentes Aeronáuticos (CENIPA) e enriquecidas pela Agência Nacional de Aviação Civil (ANAC). Através da aplicação de consultas SQL, da integração com a linguagem Python e do Power BI (MS) usados para análise estatística e visualização, busca-se identificar tendências relevantes relacionadas aos tipos de ocorrências, locais com maior incidência, fabricantes e modelos de aeronaves envolvidos, bem como a evolução temporal desses eventos. A análise explorará dados distribuídos em quatro tabelas distintas dentro do ambiente MySQL Workbench, otimizando a organização e o relacionamento das informações relevantes.

DESENVOLVIMENTO

A etapa inicial deste projeto envolverá a configuração e a conexão ao banco de dados MySQL Workbench, onde os dados das ocorrências aeronáuticas estarão armazenados em quatro tabelas distintas. A estrutura dessas tabelas, incluindo seus campos e relacionamentos, será fundamental para a elaboração de consultas SQL eficientes. A linguagem de consulta estruturada (SQL) será utilizada para extrair, filtrar e agregar os dados necessários para a análise, substituindo a leitura direta de um arquivo .JSON ou .CSV.

A análise exploratória dos dados (AED) será conduzida através da execução de *queries* SQL específicas para obter *insights* preliminares sobre o conjunto de dados organizado nas tabelas. Serão calculadas estatísticas descritivas, como média, mediana, moda e desvio padrão, para as variáveis quantitativas relevantes ao longo dos dias, ano a ano, utilizando funções agregadas do SQL. A identificação dos tipos de ocorrência mais frequentes, dos municípios e Unidades Federativas com maior e menor número de registros de acidentes será realizada através de consultas de agrupamento e ordenação.

A integração com a linguagem Python se dará através de bibliotecas como `mysql.connector` ou `SQLAlchemy`, permitindo a execução de *queries* SQL a partir de *scripts* Python e a transferência dos resultados para estruturas de dados como `DataFrames` do `Pandas`. A visualização de dados continuará a desempenhar um papel fundamental, com a construção de gráficos estatísticos utilizando as bibliotecas `Matplotlib` e `Seaborn` a partir dos dados extraídos do banco de dados. Gráficos de barras, pizza e linhas serão empregados para comparar quantidades, proporções e evoluções temporais, respectivamente, tanto utilizando a linguagem Python, quanto o aplicativo `Power BI` (MS). A capacidade do MySQL Workbench de gerenciar relacionamentos entre tabelas permitirá uma análise mais sofisticada, como a relação entre modelo/fabricante da aeronave por tipo de ocorrência, explorando *joins* e outras operações relacionais em SQL.

A análise estatística avançada poderá envolver a aplicação de técnicas diretamente no MySQL (quando apropriado) ou através da manipulação dos dados em Python após a extração. A interpretação dos resultados obtidos em cada etapa, desde a consulta SQL inicial até a visualização final, será cuidadosamente documentada, buscando fornecer *insights* relevantes para a compreensão do cenário das ocorrências aeronáuticas no Brasil, agora com o benefício da estrutura e da eficiência proporcionadas por um banco de dados relacional.

REFERÊNCIAS

DADOS ABERTOS DO GOVERNO. **Ocorrências Aeronáuticas**. Disponível em: <https://dados.gov.br/dados/conjuntos-dados/ocorrencias-aeronauticas>. Última alteração nos metadados: 13 jun. 2024. Acesso em: [abril 2025].

FERRAMENTAS UTILIZADAS:

- MySQL Workbench – Banco de Dados Relacional
- Power BI (MS) para a geração de *dashboard* interativos
- Linguagem de programação Python com o uso das Bibliotecas:
 - Pandas para análise de dados;
 - Mysql Connector, para a conexão com o MySQL Workbench;
 - Matplotlib e Seaborn para a geração dos gráficos;