**Analisando a História da Fórmula 1 com Dados e Visualizações Interativas**

A Fórmula 1 é um dos esportes mais emocionantes e ricos em dados do mundo. Cada corrida gera uma quantidade enorme de informações sobre pilotos, equipes, tempos de volta, estratégias e muito mais. Como entusiasta de F1 e dados, decidi explorar essa história de maneira interativa, criando um dashboard que permite visualizar estatísticas da categoria de forma dinâmica.

Para isso, utilizei Python e algumas bibliotecas poderosas para extrair, tratar e visualizar os dados. Neste artigo, compartilho o processo de extração de informações direto do site oficial da Fórmula 1, armazenando os dados em arquivos estruturados para facilitar a análise.

**Extração de dados**

Para realizar essa extração, utilizei as seguintes bibliotecas:

1. **playwright** – Biblioteca de automação de navegadores utilizada para acessar e extrair informações de sites dinâmicos, como a página de resultados da F1.
2. **pandas** – Ferramenta essencial para manipulação de dados. Depois da extração, os dados são organizados em DataFrames e salvos em arquivos CSV.
3. **datetime** – Usada para obter dinamicamente o ano atual e garantir que a extração sempre alcance os dados mais recentes.
4. **StringIO** – Facilita a conversão do HTML extraído das tabelas para um formato que o Pandas possa interpretar como DataFrame.
5. **re** - Expressões Regulares

**Processo de Extração de Dados**

O código automatiza a coleta de estatísticas da Fórmula 1 a partir do site oficial. O fluxo segue estes passos:

1. **Configuração do Navegador**
   * O Playwright inicia um navegador **headless** (sem interface gráfica) para otimizar a extração de dados.
   * Um **contexto** é criado para gerenciar as sessões de navegação.
2. **Iteração sobre os anos**
   * A extração ocorre desde **1950 até o ano atual** para corridas, pilotos e voltas mais rápidas.
   * Para equipes, a busca começa em **1958**, pois o Campeonato de Construtores foi introduzido nesse ano.
   * Para cada ano, o código gera uma URL correspondente e acessa a página de resultados.
3. **Coleta e Conversão dos Dados**
   * A tabela da página é localizada usando **seletor XPath** (//table[@class='f1-table f1-table-with-data w-full']).
   * O conteúdo da tabela é extraído como **HTML bruto**.
   * O Pandas converte essa estrutura para um DataFrame.
   * Adiciona-se uma coluna "Year" para manter a referência do ano correspondente.
4. **Armazenamento dos Dados**
   * Cada DataFrame é salvo em um arquivo CSV para análise posterior.

A computer screen shot of a program code

AI-generated content may be incorrect.

**Arquivos Gerados**

Após a extração, os seguintes arquivos são gerados:

1. **formula1\_race\_results.csv** → Dados das corridas (GPs e vencedores).
2. **formula1\_drivers\_results.csv** → Estatísticas dos pilotos.
3. **formula1\_teams\_results.csv** → Dados das equipes (construtores).
4. **formula1\_fastest\_laps\_results.csv** → Voltas mais rápidas de cada ano.

**2. Tratamento e Organização dos Dados**

Após a extração, os dados passaram por um processo de limpeza e estruturação. As principais etapas foram:

* **Correção de valores ausentes:** Algumas temporadas antigas tinham dados inconsistentes que precisaram ser ajustados.
* **Conversão de formatos:**
  + Tempos de volta foram convertidos para segundos para facilitar cálculos.
  + Anos foram padronizados como valores inteiros.
* **Criação de colunas derivadas:**
  + Número total de vitórias por piloto e equipe.
  + Tempo total de corrida por pista ao longo dos anos.
  + Média de pontos por corrida para cada piloto.

Essa etapa garantiu que os dados estivessem prontos para visualização no dashboard.

**3. Construção do Dashboard (Arquivo: dashboard.py)**

Para tornar a análise mais acessível, utilizei o **Streamlit** para criar um dashboard interativo com três seções principais:

**Seasons 🏁**

* Permite selecionar um ano específico e visualizar:
  + Ranking de pilotos baseado em pontos.
  + Ranking de equipes com mais pontos na temporada.
  + Lista de voltas mais rápidas registradas no ano.
  + Destaques do ano, como o piloto com mais vitórias e a corrida mais longa.

**Rankings 📊**

* Exibe estatísticas históricas gerais, como:
  + Piloto que mais venceu corridas.
  + Corridas mais disputadas ao longo da história.
  + Pilotos com mais voltas rápidas registradas.
  + Tempo total acumulado de corrida por circuito.

**Interactions 📈**

* Permite explorar dados de forma interativa:
  + Selecionar um Grande Prêmio e visualizar quais pilotos mais venceram.
  + Para cada piloto, listar os anos em que venceu e quem fez a volta mais rápida nesses anos.
  + Top 5 voltas mais rápidas já registradas no circuito selecionado.

As bibliotecas **Plotly** e **Pandas** foram utilizadas para criar gráficos dinâmicos e tabelas organizadas.