By Mohit Kothari from MOhiteam

Lōkahi Healthcare Accelerator

INTERACTIVE HACKATHON DATA EXPLORER

Agenda

1. Introduction
2. Key Features
3. Technologies Used
4. Benefits
5. Future Scope
6. Use Cases
7. Take Action
8. Conclusion

Introduction

Empowering hackathon participants to explore and extract insights quickly.

- Purpose: To provide Hackathon participants with a tool to upload, process, and visualize large datasets efficiently.
- Key use cases: Analyzing patient data, visualizing medical research results, and exploring healthcare metrics.

Key Features

1. Data Loading
2. Data Preprocessing
3. Data Sampling
4. Dynamic Visualization Suggestions
5. Interactive Visualizations
6. Geographic Visualization

1. Supports uploading CSV and Parquet files. User-friendly file uploader interface.
2. Automatically fills missing values. Preprocesses data for analysis.
3. Samples a subset of the data for efficient visualization.
4. Suggests suitable visualization types based on data types.
5. Allows users to select and customize visualizations using Plotly.
6. Supports geographic data visualization for datasets containing a 'state' column.

Technologies USED

Streamlit

- Provides a user-friendly interface for building and sharing data applications.

PyArrow

- Efficiently handles Parquet files.

Pandas

- Handles data manipulation and analysis.

st_aggrid

- Provides an interactive grid for displaying large datasets.

Plotly

- Creates interactive and visually appealing plots.

Python

- Ease of use, rich ecosystem of libraries

BENEFITS

Data-Driven Decisions

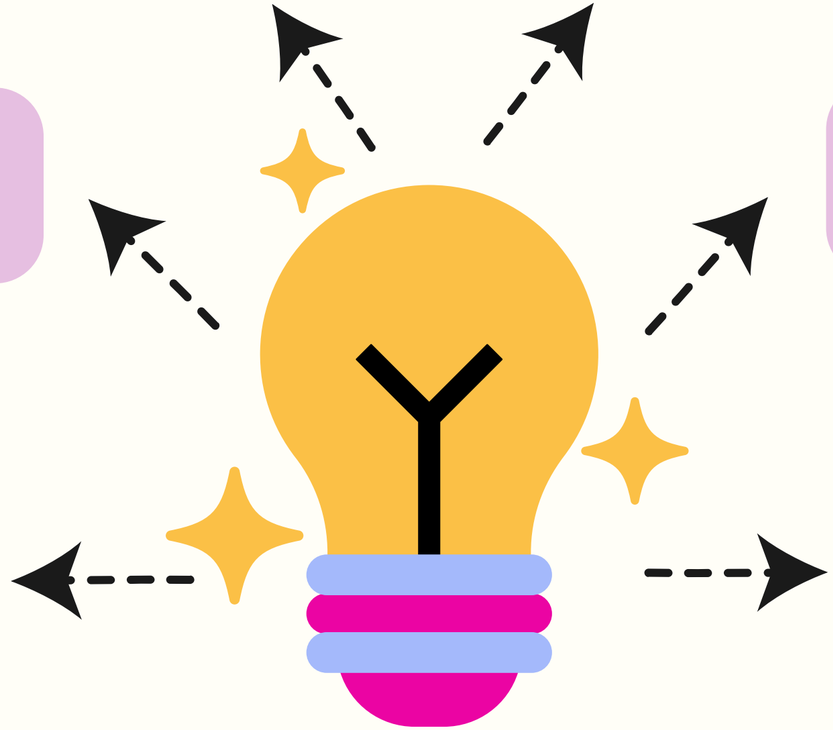
Empower users to make informed, data-driven decisions based on comprehensive data analysis.

Time Efficiency

Save time by automating data preprocessing and sampling, allowing users to focus on analysis and insights.

Save time by automating data preprocessing and sampling, allowing users to focus on analysis and insights.

User-Friendly Interface



Customizable Visualizations

Handle large datasets efficiently, making the tool suitable for various industries and applications.

Scalable Solutions

Offer a wide range of visualization options that can be customized to meet specific analytical needs.

Performance

- st_aggrid enhances performance with features like pagination and multi-row selection.

FUTURE SCOPE

- **ADVANCED VISUALIZATIONS:**

  Integrate more advanced visualization types such as 3D plots, animated charts, and dashboards.

- **COLLABORATION FEATURES:**

  Collaboration features to allow multiple users to work on the same dataset simultaneously.

- **USER AUTHENTICATION**

  Access control will be implemented to restrict access to unauthorized users.

- **CHATBOT SUPPORT**

  In the future user will be able to chat with their dataset.

USE CASES

HEALTHCARE ANALYTICS

- Analyze patient data to identify trends, patterns, and outliers.
- Visualize medical research results to gain insights and support findings.
- Explore healthcare metrics to monitor performance and make data-driven decisions.

CLINICAL RESEARCH

- Support clinical trials by providing visualizations and insights from trial data.
- Help researchers identify patterns and correlations in clinical data.

EDUCATIONAL PURPOSES

- Use the application for educational purposes to teach data analysis and visualization.
- Provide students with a hands-on tool to explore and analyze datasets.

FINANCIAL ANALYSIS

- Financial analysts can use the tool to analyze financial data, identify trends, and make data-driven investment decisions.

Take action!

GitHub: https://github.com/mohitkothari-dev/Interactive-Hackathon-Data-Explorer

Live app: https://lablabdataexplorer.streamlit.app/

If you run it locally then:

1. update the config.toml file in the ~/.streamlit directory:
[server]
maxMessageSize = 1024 # Set the browser limit to 1024 MB

2. Run the Streamlit app with the increased message size limit:
streamlit run app.py --server.maxMessageSize=1024

CONCLUSIONS

CONCLUSION 1

The Interactive Data Explorer is a powerful tool for medical professionals to analyze and visualize large datasets efficiently.

CONCLUSION 2

By leveraging technologies such as Streamlit, Pandas, Plotly, PyArrow, and st_aggrid, the application provides a user-friendly, interactive, and scalable solution for data exploration.

CONCLUSION 3

The future scope of the project includes advanced visualizations, machine learning integration, collaboration features, data security enhancements, user authentication, and mobile support, making it a comprehensive tool for various use cases in healthcare and beyond.

THANK YOU!