SmartTransit : A Data Driven System for Navigating and Analyzing Public Transport

Explore the pulse of Boston's subway system. This presentation dives into data-driven insights and interactive visualizations of the Red, Green, Orange, and Blue lines.

Presented By: Priyanka Maan and Mitul Nakrani

Northeastern University

MOTIVATION

- **Decision Making**: City planners need quick insights for infrastructure decisions
- **Public Benefit**: Commuters want to understand their transit network better.
- **Technical Challenge**: Bridging the gap between raw data and actionable insights.
- **Environmental Impact**: Optimizing public transit reduces carbon footprint and traffic congestion.
- **Rider Fare Savings**: Better transit planning saves a lot in costs annually for the end user.
- **Accessibility for All**: Stations analysis helps people with disability to transit safely.

The MBTA network sees an average of **425,000 weekday station entries**, with approximately **200,000 on weekends**.

Northeastern University

DATA INSIGHTS & PROCESSING

MBTA GTFS Official Feed: Data fetched from Massachusetts Bay Transportation Authority Link.

Additional Dataset: Transit Times Dataset is fetched from MassDoT open dataset directory Link.

Format: General Transit Feed Specification (GTFS) - Industry standard

Processing: Converted GTFS data to CSV for analysis

- Fetch GTFS data from MBTA official feed and other sources.
- Extract relevant files (stops.txt, travel_times.csv, routes.txt, facilities.txt, connections.csv, etc.)
- Transform and Clean unstructured data into CSV format (stations, connections, locations, station_amenities)
- Calculate network metrics and other analytics as required.
- Create in-memory SQLite database for queries.

PROJECT SCOPE

AI Assistant

Ask questions in plain English and get SQL-powered answers instantly

Route Selection with Travel Time Calculation

Find routes and travel time by line, travel time, or specific station names

Analytics Dashboard

Deep dive into network efficiency, centrality, and performance stats

Accessibility Insights

Identify stations with elevators, ramps, and wheelchair access

METHODOLOGY & TECHNOLOGY USED

Data Collection: CSV files

Data Processing: Pandas, Numpy, NetworkX for graph analysis. SQLite In memory for querying.

Visualization Layer: Plotly for interactive charts

AI Integration: OpenAI GPT-4 for natural language processing

User Interface: Streamlit for web deployment

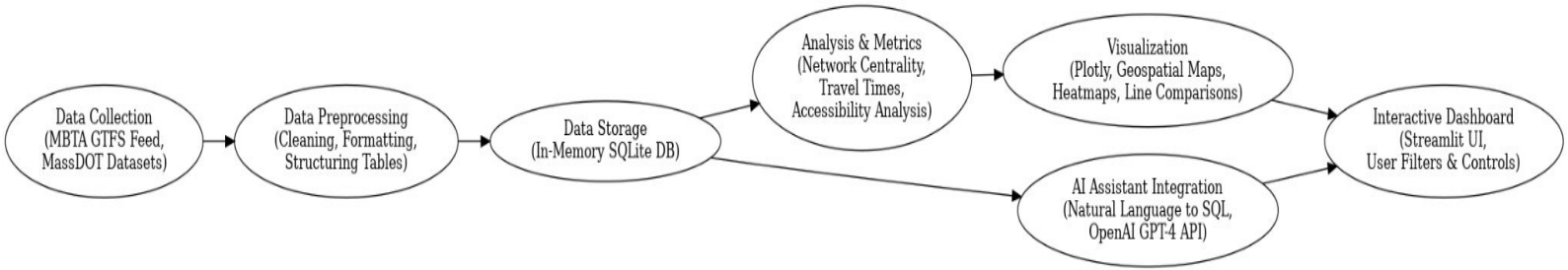
Frontend: Streamlit, Plotly, HTML/CSS

Backend: Python, Pandas, NetworkX

Database: SQLite (in-memory)

AI/ML: OpenAI GPT-4 API

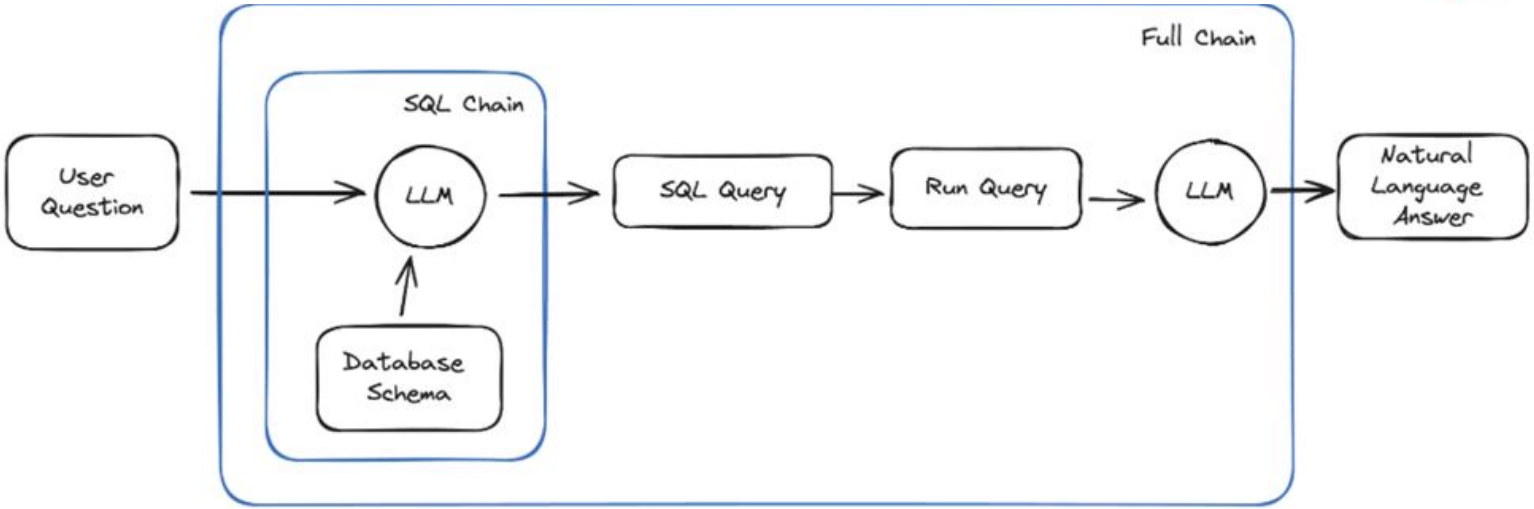
Analytics: NumPy, SciPy for calculations



Northeastern University

AI ASSISTANT

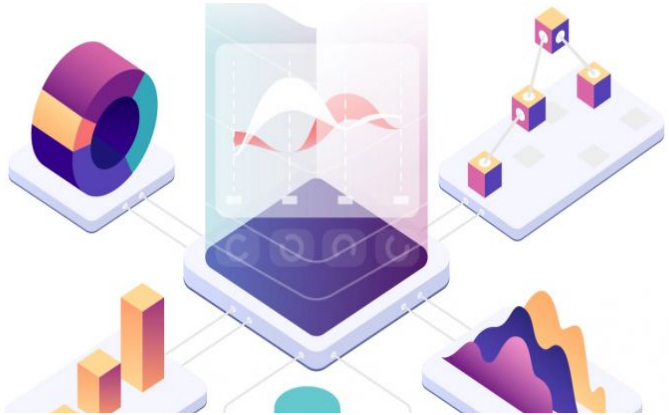
- ➔ GPT - 4 powered AI Assistant
- ➔ Natural Language Queries to Insightful results & visualizations
- ➔ SQLite database



Northeastern University

VISUALIZATIONS

- **Network Graphs & Centrality Heatmaps**: Identify hubs, connectivity patterns, and bottlenecks
- **Multi-panel Line Comparisons**: Compare efficiency, stations, connections across lines
- **Travel Time Analysis**: Optimize schedules and identify slow connections
- **Geographic Scatter Plots:** Analyze spatial distribution and accessibility
- **Network Visualizations**: Immersive exploration of network density and patterns



DEMO

LIMITATIONS

- **Static and Mode-Limited Data:** The platform relies on scheduled GTFS feeds and static datasets, focusing only on MBTA's subway network without multi-modal integration.
- **Data Quality and Completeness:** Accessibility and facility information may be outdated or incomplete due to reliance on public datasets.
- **Technical Constraints:** In-memory storage limits scalability for large datasets, and the AI query interface may misinterpret complex or ambiguous queries.

CONCLUSION AND FUTURE WORK

Conclusion

- ➔ Built comprehensive analytics platform for Boston subway network
- ➔ Integrated AI to make data analysis accessible to everyone
- ➔ Created interactive visualizations for deeper insights
- ➔ Demonstrated potential for smarter transit planning

Future Efforts

- ➔ **Real-time GTFS Integration:** Live transit data feeds
- ➔ **Predictive Analytics:** ML models for delay prediction
- ➔ **Multi-modal Analysis:** Include bus and commuter rail
- ➔ **Historical Analysis:** Track network changes over time
- ➔ **API Development:** RESTful API for third-party developers

Northeastern University

THANK YOU!!

GITHUB REPOSITORY LINK -

https://github.com/MitulNakrani003/DS5110IDMP-Project

Northeastern University