Good evening everyone. we’are excited to share with you today a project we’ve been working on that combines data analytics, AI, and interactive visualization to transform how we understand and analyze public transit systems. This project focuses on the Boston subway network, but the techniques and tools we've developed can be applied to any transit system.

Slide 2

So why did we build this? We've always been fascinated by public transit systems, but we noticed a significant gap. Most transit data is either too technical for the average person or too simplified to be useful for analysis. Meanwhile, commuters want to understand things like 'which station is the most connected?' or 'what's the most efficient route?

But beyond individual convenience, there are larger implications. Optimizing public transit directly impacts our environment - every improvement that makes transit more efficient encourages more people to leave their cars at home, reducing emissions and traffic.

Slide 3

Let's dive into the data itself. Our data comes directly from the MBTA's official GTFS feed - that's the Massachusetts Bay Transportation Authority's General Transit Feed Specification data. GTFS is the industry standard format that transit agencies worldwide use to share their data. We fetched this authoritative data and processed it into a format optimized for our analysis.

We also create an in-memory SQLite database that our AI assistant can query. This preprocessing of official MBTA data enables real-time analysis without performance penalties while ensuring our insights are based on the most accurate, up-to-date information available.

Slide 4

Let us walk you through how we approached building this system. We started with core CSV files containing station data, connection information, geographic coordinates, transit data, facilities data, routes data and many other data files. Using Pandas for data manipulation and NetworkX for graph analysis, we processed this raw data into meaningful metrics. The visualization layer uses Plotly to create interactive, responsive charts. But here's where it gets interesting - we integrated OpenAI's GPT-4 to translate natural language questions into SQL queries, making the data accessible to non-technical users. Everything is wrapped in a Streamlit interface, which allows for rapid deployment and easy updates.

Slide 5

We've built five main features into the application. The AI Assistant is the star of the show - users can ask questions like 'Which line has the most stations?' and get instant answers with the SQL query shown for transparency. The Network Visualization provides an map where users can explore the subway system visually. The Analytics Dashboard offers deep insights with multiple charts and metrics. The Connection Explorer allows users to filter connections by line, travel time, or search for specific stations. And importantly, we've included Accessibility Insights that analyze which stations have elevators, ramps, and other accessibility features - this helps mobility-impaired users plan their routes and helps transit authorities identify gaps in accessibility infrastructure. Each feature serves a different user need while maintaining a cohesive experience.