Overview of Factors Affecting Public Transit Ridership

Course: Data Mining CU Boulder
Course Code: (CSCI 5502-872)

Team:
Yashwant Gandham
Diego Olaya
Sasidhar Reddy

18 October, 2024

Abstract:

Transportation access in modern society is critical for people to gain access to job markets, social settings, and other essential services. Public transit systems are known to reduce traffic congestion, decrease individual transportation costs, and decrease emission compared to personal vehicles. Although the United States invests in public transit infrastructure, ridership is lower compared to other developed countries where transit is more common, like in Western Europe and Japan. The purpose of this study is to identify the factors most strongly related to public transit ridership through an analysis of data from the National Transit Database. In this study, we investigate the influence of capital expenditures, service frequency, reliability, and other variables on ridership levels with a view to inform effective policy decisions for developing or improving transit systems.

Data Collection and Preparation:

Source:

The principal data source for this analysis is the National Transit Database (NTD) which is maintained by the Federal Transit Administration. The NTD is the central repository of data on transit systems in the United States, including information on annual expenditures, service area populations, and operational statistics. Reporting to the NTD has been required by law since 1974.

Selection of Data:

We focus on capital expenditures and associated metrics from 2016 to 2022, since this period gives the most relevant and recent data to understand prevailing trends. The data includes information on:

Annual Capital Expenditures: The sum of money that is spent annually by different transit agencies on various capital projects.
Service Area Population: The population which is served by every transit agency.
Input Categories: Spending in various breakdowns given by specific categories defined by the FTA, which may be guideways, passenger vehicles, facilities, and other communication systems.

Extraction:

The data were extracted for each year. There was variation in the formats and definitions of the data provided:

2016-2021: Data were presented in Excel files with column definitions detailed in accompanying Excel sheets.
2022: Data format and definitions were directly available on the NTD website.

Consolidation:

The collected data for the years shown were combined into one dataset:

- Column Alignment: Standardization of column names and formats across years.
- Exclusion of Questionable Data: Removed columns marked as questionable or with a large portion of missing data.
- Consistent Identifiers: Ensured agency identifiers and names were consistent throughout to enable proper aggregation and analysis.

Data Cleaning

Cleaning the data was one of the most important steps toward maintaining integrity and, as such, reliability within the analysis:

- Handling Missing Values: Records and columns with a large amount of missing or questionable data, as noted by the NTD.
- Removing Duplicates: Duplicated records were checked for and removed to prevent biased results.
- Standardizing Data Types: Numerical data was correctly typed (e.g., expenditures as floating-point), and categorical (e.g., agency names, states) was standardized.
- Removing Inconsistencies: Various year agency names and codes were normalized, and reporting practices differing across years and have been corrected for consistency.

Visualization:

Trends, pattern, and insight exploration of the data by Visualization:

1. Total Capital Expenditures Over Time

Description: Line chart of total capital expenditures for U.S. transit agencies from 2016 to 2022.
- Insight: From this chart, it is obvious that the total expenditures have been on a continuous rise, increasing by almost 7 billion dollars in six years. This could mean increased investment in the United States on public transit infrastructures.

2. Per Capita Capital Expenditures Over Time

- Description: Line chart of the capital expenditures per service area population for the same period.
Insight - The per capita spending increases even after adjusting for population growth in service areas. This indicates that the increase in spending is higher than inflation and cannot be due to serving more citizens.

3. State-Level Per Capita Expenditures

Box plots of the per capita spending on transit for every state for relevant yearly periods.

State-by-State Variation: Spending is very different from state to state. While most states spend around $100 per capita, there is a considerable number of outliers.
High-Spending Outliers: Some states, such as California, consistently report way higher per-capita spending due to big transit projects like metro expansions.
This means that, even though investment in public transportation is growing at the national level, investment distributes very differently and concentrates on different levels for each state.

4. Capital Expenditure Categories

Bar charts showing the distribution of spending of the capital expenditures in nine predefined categories among the transit agencies, including but not limited to guideways, passenger vehicles, and stations.

Frequency of Expenditures:

Most of the commonly reported expenses are passenger vehicles, communication systems, and maintenance buildings. It has to be taken into notice, however that investment in guideways is a much more

seldom exercised activity, with the high expense entailed by this type of structure. Let's analyze what is being said, translated: "Essential components are those most frequently updated, which also require regular maintenance.".
Infrastructure Expansion: Large guideway expenditures are indicative of infrastructure expansion, and though rare, they have a large effect on overall spending.

Conclusion

From this exploratory data analysis, it becomes clear that, while the United States is indeed investing more in its public transit infrastructure, such growth is very uneven among states and expenditure categories. It also becomes apparent that higher spending does not always equate to more ridership, thus calling for a more informed understanding of public transit usage factors. The spending indicates a priority on the basic elements of a transit system, but the more intensive infrastructure projects like guideways are more expensive. This kind of analysis needs to be taken to the next level by studying the association between these outputs with ridership data and other variables of service frequency, reliability, and accessibility, in an effort to understand what comprises the drivers of successful transit systems.

References

- National Transit Database - NTD: https://www.transit.dot.gov/ntd
- Strong Towns: https://www.strongtowns.org
- Not Just Bikes - YouTube: https://www.youtube.com/c/NotJustBikes
- RM Transit (YouTube): https://www.youtube.com/c/RMTransit