# BUFN403 First Deliverable

Om Duggineni
University of Maryland
odo@terpmail.umd.edu

Benjamin Knight
University of Maryland
bknight5@umd.edu

Jack Doggett
University of Maryland
jdogget1@terpmail.umd.edu

March 7, 2025

## 1 Abstract

This project analyzes transportation access - including public transportation, walkability, and driving - and the relationship it has on real estate and social impacts to specific neighborhoods in the Washington D.C. metro area. Specifically, we want to determine how the economic impacts of these modes of transportation could affect bank loans, income, and other economic factors associated with an area. Public census and geographic datasets will be used to create models to infer the relationship between transportation access and key financial statistics. The ultimate goal will be to show how better metro and bus connections within an area affect indicators related to the economic activity in a neighborhood.

## 2 Data Gathering

**Goal:** Build a custom dataset which maps data about how accessible a place is to public transportation to metrics describing that region's economic activity, including data on bank deposits/loans, economic data, and census/income data.

The first step of our project consists of building a custom dataset from publicly available data sources. The sources that we plan on using are discussed below:

### 2.1 Public Transit Accessibility Data

#### 2.1.1 Background

Worldwide, public transport has been known as a positive indicator for economic development, including areas such as the hosing market [1]. Metro systems,

specifically, have a large body of research demonstrating positive economic impacts. Metro systems have been shown to increase mobility, boost industry and housing around stations, and increase property values within a large radius of each station [3]. However, metros are extremely capital intensive, limiting the scope of projects while having large costs on the communities and environments they are built in [4].

Areas better connected by metro generally have higher density, large transit oriented developments, and clusters of businesses leading to higher percentage of high income workers and larger and more connected markets. Bus lines have a wide range of ridership, with a mix of service workers and commuters. Bus lines have mixed economic benefits, often providing the only form of transportation for workers in marginalized communities, but also lacking the frequency, connectedness, and infrastructure to bring the same benefits as the metro.

Obtaining data on how well-connected a particular region is via public transportation networks is very important to our project, as understanding how well areas are connected through public transit in a standardized manner allows us to determine how this may affect the economic activity within the region. However, this is challenging, as any measurement of how accessible an area is using public transportation has to be reflective of how individual users perceive the convenience of the system. For example, if an area is well-connected through bus and train networks but navigating said network takes a long time due to slow, suboptimal routes or delays, then public transit may have a smaller economic impact on the area.

Additionally, since metro systems work in tandem with the other kinds of transportation infrastructure in an area, which means that calculating how navigable a region is with public transit also requires knowledge of the other types of transportation infrastructure in an area, including walking, bus, and biking infrastructure [7].

### 2.1.2 Approach

We plan on using a somewhat novel approach by using mapping data from routing applications in order to predict what the public transit accessibility of a particular area will look like. Two open-source projects will be of particular benefit in this case - OpenStreetMap and GTFS.

OpenStreetMap is a public dataset of mapping data built collaboratively as a community project. In addition to being quite accurate, especially for walking and biking paths, it is freely available online for anyone to download. Many projects exist that have the ability to work with this data, both for displaying maps and for GPS-based navigation apps. The dataset is also used by many big companies [8] for their own navigation apps. (As an example of a navigation app that uses exclusively OpenStreetMap (and GTFS, explained below) data, try Organic Maps)

OpenStreetMap data is available to download from a daily-updated mirror at `https://download.geofabrik.de/`, with subsections of the map focusing on different areas of the world, although we will likely be focusing on the area

surrounding Maryland for our specific analysis.

GTFS, the General Transit Feed Specification, is an open standard that public transit providers use to provide real-time and historical data on their transit networks. Specifically, it contains information such as the location of transit stops and the timings of arrivals and current locations of buses and trains.

By using OpenStreetMap data for mapping/navigation and GTFS data for the transit system, we can, essentially, use a routing engine to determine the bounded area of how far a particular user will be able to travel in a given period of time using the area's public transit system, producing a plot called an isochrone plot. The area contained within an isochrone describes how many different locations a person can reach using the public transit system, and is a good measure for how well-connected the area around that position is through public transit networks.

## 2.2 Financial and Economic Data

### 2.2.1 Census Data

The US Decennial Census provides demographic and socioeconomic statistics per census block, although some information like income is only available at the block-group level. The FCC provides an API to convert global positions to census blocks. The TIGER/Line Census datasets provide shape files for mapping geographic locations to Census areas, and is compatible with GIS software.

### 2.2.2 Foursquare OS Places + Safegraph/Advan Points of Interest

These two datasets contain many geospatial points of interest along with demographic / economic information. This dataset was originally found based on Elliot Lee's project. We plan to use this to see if we can use the demographic/economic insight from this data in order to find any trends associated with public transit development.

### 2.2.3 Bank Summary of Deposits Data

Banks which are insured by the FDIC are required to provide certain basic data on all the deposits and loans made by the bank. Can be downloaded from the FDIC website ([6]). Provides a lot of useful insight into whether banks mainly focus on accepting deposits or lending to small businesses.

This dataset was also originally found based on Elliot Lee's project. We plan to obtain an updated copy of it from the FDIC website for our analysis.

# 3 Potential Algorithmic/Data Challenges

The main challenges associated with this project will be (in order):

1. Finding and gaining access to the datasets which we want to integrate into our project.
2. Determining our locations of interest based on the data that we have access to, as well as what granularity we should use to measure them.
3. Using a routing engine to determine public transit accessibility in these locations of interest using OpenStreetMap data
4. Using data analysis and potentially AI to determine how public transit accessibility represents the economic factors collected about our locations of interest.
5. Creating a visualization of the results and making it available online.

# 4 Data Analysis

After our group finishes this collection, machine learning algorithms will be used to determine feature importance and generate insights that will be displayed on our final product. The exact techniques may vary depending on the type of data we collect, but will be determined further as we go further into our analysis.

## 4.1 User Interface + Data Visualization

Our final product will be an interactive visualization that users can interact with to determine how various factors across these datasets contribute to bank loans, income, and other economic factors associated with an area. Users will be able to view our data, as well as general trends and conclusions that we observed in the process of analyzing it. This user interface will be hosted online and will be publicly accessible.

We plan on using React (may change) to design the visualization and use mapping tools to show map data in a user-interactive way. We will also have static graphs and text describing the statistical analyses we have done. We aim to host the final product online using GitHub Pages, where we will also include access to the datasets we used in our project.

# 5 Discussion

## 5.1 Intended Audience

Our intended audience are stakeholders in real estate and the communities in which we are modeling. Our dashboard will be public and designed to communicate our insights clearly and powerfully with all users. We are especially interested in needs within the real estate industry and welcome collaboration in our dashboard design process.

### 5.2 Potential Advisors

Our group is looking for advisors in real estate who can help us understand needs in the industry and guide our research. This may give us better insight on how to analyze the economic data we obtain.

Additionally, we may consider briefly meeting with someone involved in transportation research at UMD to understand how public transit can affect economic development. Groups we are considering for this include the Maryland Transportation Institute and the CATT (Center for Advanced Transportation Technology) Lab.

## 6 Team Roles

It's possible that these roles may change in the future (this will be communicated clearly).

| Member | Role |
|---|---|
| Om Duggineni | Data Cleaning - Obtaining Data, Initial Processing of data |
| Benjamin Knight | Combining Data Sources + Data Analysis |
| Jack Doggett | Distributed Compute + User Interface |

## References


[1] Dorantes, L. M., Paez, A., & Vassallo, J. M. (2011). Analysis of House Prices to Assess Economic Impacts of New Public Transport Infrastructure: Madrid Metro Line 12. Transportation Research Record, 2245(1), 962-978. https://doi.org/10.1177/13634607221107827

[2] General Transit Feed Specification. (n.d.). Retrieved March 7, 2025, from https://gtfs.org/

[3] Lin, D., Broere, W., & Cui, J. (2022). Metro systems and urban development: Impacts and implications. Tunnelling and underground space technology, 125, 104509.

[4] TIWARI, G. (2013). Metro Rail and the City: Derailing Public Transport. Economic and Political Weekly, 48(48), 65–76. http://www.jstor.org/stable/23528925

[5] OpenStreetMap. (n.d.). OpenStreetMap. Retrieved March 7, 2025, from https://www.openstreetmap.org/

[6] Summary of deposits. (2023, June 1). FDIC. https://www.fdic.gov/bank-financial-reports/summary-deposits

[7] Truong, R., Gkountouna, O., Pfoser, D., & Züfle, A. (2018). Towards a better understanding of public transportation traffic: A case study of the Washington, DC metro. Urban Science, 2(3), 65.

[8] Who uses OpenStreetMap? (n.d.). Retrieved March 7, 2025, from https://welcome.openstreetmap.org/about-osm-community/consumers/