Will Friedrichs

Policy Proposal 1

Project Motivation:

My motivation for this project is the personal experiences I have had with SEPTA’s bus system. For more than a year, I have used primarily the Google Maps app to navigate SEPTA’s system of buses, and on multiple occasions, I have waited for buses only to have them not show at all. These cancellations are displayed on the clunkier and generally less-user-friendly SEPTA App, but since Google Maps has a significant number of users in Philadelphia, it is concerning that data describing bus cancellations does not make it to Google Maps.

An additional point of analysis will be looking at buses that arrive earlier than Google Maps’ projected pickup time. Early bus arrivals have been shown to be more detrimental to the functionality of a city’s bus transit system than equivalently late pickup times.

Project Question:

Where are bus service cancellations/early pickups most likely to occur in Philadelphia?

Where are instances of Google Maps mistakes associated with cancellations most common? Where do they occur at the highest rates?

The first question is descriptive because it explores the answer to a “where” question rather than a “why” question. The second question is also descriptive, however to address it, additional context about the shortcomings of SEPTA’s real-time GTFS data stream is necessary.

*Brief Note:*

*I have done research this week into how Google Maps and SEPTA get data to make predictions associated with bus routes (e.g. delays, detours, route changes, etc.). I found that transportation agencies in the US, including SEPTA, produce feeds of real-time data in GTFS format, and Google Maps uses information from these public feeds to make real-time adjustments to their trip predictions. The GTFS data that SEPTA produces includes any delays or cancellations to bus trips, and SEPTA’s app reflects that, however Google maps does not.*

*Google’s “Transit Partners Help” page lists several possible reasons why transit agency data may have this issue going into Google Maps, and I plan to continue to try and tease out the specific problem. If I can do that, the second part of my project question currently in parentheses becomes much more feasible.*

*About two weeks ago, I was able to gain access to and scrape a webpage whose real-time updates (while not in GTFS format) did include information about bus cancellations.*

Existing relevant research:

There is not much existing academic research into Google Maps public transit accuracy problems in Philadelphia, but Google has documented information describing common issues associated with transit agency GTFS real-time feeds, and how this can translate to accuracy issues. Google Maps only fully integrated SEPTA’s transit stops in 2019 after years of partial visibility.

Datasets Identified/Methods:

I have created python scripts that can download a near-continuous stream of SEPTA’s non-GTFS real-time transit updates, including all cancellations. The next step is to run this script continuously for multiple weeks on a server capable of such a task. I expect to begin collecting data on Sunday, February 20.

Deliverables:

The final deliverable for this project will be a research paper with accompanying maps showing the frequency of bus cancellations and early pickups across Philadelphia. The research paper will identify patterns in these instances, and groups that are most affected by them (income, socio-economic status, etc.).

The research paper will also identify the mismatch between Google Maps and SEPTA real-time data, and the cause of mismatches like the one pictured below. Potential solutions could then be identified.

Additionally, time permitting, I would like to construct a dashboard that would be in working condition theoretically for as long as virtual web hosting services could be supported (unfortunately I do not intend to continue supporting these services indefinitely).

Graphical user interface, application

Description automatically generated

Policy Implications:

This will show SEPTA and any other interested parties the scale of the cancellation and early arrival issues, who is being most affected by them, and ideally, the problems with SEPTA’s GTFS real-time data feed along with a potential solution.