

Reflection Report – Collecting Mobility Data Using GPS

Ashok Erukonda

1 What GPS Mobility Data Means and Why It's Important

GPS (Global Positioning System) helps us find exact locations anywhere on Earth. It works by using signals from satellites to record where something is such as a person, car, or bus at a specific time.

The goal of this lab was to learn how GPS data can be collected, cleaned, and turned into useful information. We worked with raw GPS files (called **.gsd** files) that contained numbers showing position, time, speed, and altitude. These numbers didn't make much sense at first, so we used Python to process them into clear data that could be analyzed.

This kind of data is used in many ways in the real world:

- To study how people or vehicles move in a city.
- To help in transport planning or find better travel routes.
- To track deliveries or taxi routes.
- To design smart cities and improve traffic flow.

In short, GPS data helps us understand movement, where, when, and how fast things travel.

2 How the Data Was Processed in the Lab

The raw GPS data we got was not ready to use. It needed cleaning and proper formatting. Here's what we did step by step:

Reading and Converting the Data

The coordinates in the files were not in normal decimal degrees. For example, “60267285” means 60 degrees and 26.7285 minutes. We converted these into proper latitude and longitude values in decimal form, so that mapping tools could understand them.

Adding IDs and Organizing Trips

Each GPS file represented one trip. We added a “Trip ID” and “Point ID” so that every recorded point could be traced back to the trip it belonged to.

Calculating New Values

Using Python, we calculated:

- **Speed (km/h):** How fast the object was moving.
- **Distance (km):** The gap between two GPS points using the Haversine formula.
- **Time Difference (seconds):** The time between two recordings.
- **Acceleration (m/s²):** How quickly speed was changing.

Merging All Trips Together

After processing each trip, we combined all the individual CSV files into one large file called **All_Trips_Combined.csv**. This single file contained every trip’s cleaned data, including latitude, longitude, time, speed, distance, and acceleration. We used this combined dataset to create our visual graphs for Task 2. It allowed us to see all the trips together and understand overall movement patterns in one view.

Summarizing Trips

For each trip, we found the total distance, trip duration, average speed, and average acceleration. This helped us understand the behavior of each journey.

Visualizing the Data

Finally, we made four main plots:

- A map showing all trip routes (trajectories).
- A version of the same map where colors showed speed.
- A heatmap showing where the GPS points were most concentrated.
- A time graph showing how speed changed during trips.

These steps changed messy, unreadable data into something structured and meaningful.

3 How Reliable GPS Data Is (and Its Problems)

GPS is a powerful and widely used technology, but it isn't perfect. While it gives us valuable location and movement information, many factors can affect how accurate or reliable the readings are. Understanding these problems helps explain why cleaning and checking GPS data is so important.

- **Signal Problems and Obstruction:** GPS works by receiving signals from satellites orbiting the Earth. When a person or vehicle moves through tunnels, near tall buildings, or under thick trees, these signals can be blocked. This causes missing data points or sudden jumps in location. In city areas, this is often called the “urban canyon effect.”
- **Reflected Signals (Multipath Errors):** Sometimes, GPS signals bounce off surfaces like buildings, water, or metal before reaching the receiver. This can make the position appear a few meters away from the true location.
- **Weather and Atmospheric Changes:** The GPS signals pass through the atmosphere, and changes in humidity, temperature, or pressure can slightly delay them.
- **Device Accuracy and Sampling Rate:** Professional devices record more precise locations, while cheap or old devices might have an error of several meters. If the GPS only records a point every few seconds or minutes, it can miss small movements.
- **Power and Data Gaps:** Battery loss or app crashes can stop recording and create large gaps in time and space.

Because of all these issues, GPS data should never be used directly without checking. The reliability of GPS can change depending on location, environment, and device quality. That is why the preprocessing and cleaning we did in this lab, such as checking speeds, times, and distances are so necessary before doing any analysis or visualization.

4 Checking and Improving the Quality of GPS Data

In this lab, we focused on making sure our GPS data was correct and realistic. Here are the main ways to check quality:

- Check for missing or duplicate points.
- Look for impossible speeds (e.g., a person moving at 200 km/h).
- Compare distances and times to catch inconsistencies.
- Convert coordinates correctly (small mistakes shift points by kilometers).
- Use graphs (maps, heatmaps, time plots) to spot issues quickly.

By applying these checks, the data became clean and trustworthy. We could then confidently use it to analyze movement.

5 What I Learned and Reflections

This assignment taught me much more than just using Python. I learned how important data cleaning and quality checking are even for something as common as GPS data.

At first, the raw files looked confusing, but after processing, they revealed clear movement patterns. Making visual plots helped me “see” the data and understand its story.

I also learned that every dataset, especially from sensors like GPS, can have errors and that the quality of our conclusions depends entirely on how well we clean and prepare the data.

This lab made me realize how location data supports so many technologies around us from ride-sharing and delivery apps to research in smart cities and public transport.

6 Conclusion

Collecting and analyzing GPS mobility data is a powerful way to study how people and vehicles move. But raw GPS data is never perfect it must be cleaned, checked, and processed before it becomes useful.

In this lab, I practiced converting coordinates, calculating speeds and distances, merging all trip files into one dataset, and creating visual maps that showed overall movement patterns. This experience gave me a strong foundation in data quality management and showed how small errors can affect the big picture.

Good data is the key to good insights and this lab helped me understand exactly how to achieve that.