

# Exploratory Research – Problem Formulation & Data Collection Planning

Ashwin Ajoy Dharmavaram - 1RV22AI011

## 1 Problem Statement

Improvement of the Bangalore Metro Scheduling.

The bangalore metro, is invaluable for me since I have to travel a long distance to the college. However, it takes a long time, has many stops, and sometimes, no one gets in. The goal is to see that for a scheduling algorithm that optimises time spent in the metro for most people, how would it behave in the real world.

I won't be taking any particular scheduling algorithm, just analysis on if anything can be improved.

---

## 2 Research Design

Of the 2 main design strategies both might be needed:

### Primary Research

Due to the lack of publically available datasets on metro scheduling especially with regards to the number of people coming on and off at stations, number of people in a day, data on metro card swipes; it will be necessary to gather data directly from sources such as **Stored Databases**.

### Secondary Research

The metro schedules are actually public (on Android), when the trains come at each station can be recorded at each time of the day through scraping, and are valuable information.

Along with that, the metro routes are public information, and for the sake of numerical simulation that might be needed through the project, the data regarding distances between stations and other physical information can be gotten.

## 3 Sampling and Data Collection

Since the topic is primarily an analysis/algorithm topic, the data that needs to be collected from people as an **interview** needs to be able to answer questions that cannot be inferred from cameras/databases.

1. How many people miss trains every 5 minutes (the train frequency)?
2. How many people are on the platform for more than 10 minutes?
3. What's the average time the train stops at the station at each hour?
4. On sunday mornings, the train frequency is 15 minutes since no one has work or college. How many passengers get on each train?
  - (a) 1-5
  - (b) 6-10

- (c) 10-20
  - (d) 21+
5. Would a scheduling algorithm where the train skips some stations be preferred by them?
  6. What's the maximum number of people in the station that can be handled at a particular time?
  7. If the train doors are open a little less (3 seconds), would the rush be balanced by the speed of the train?
  8. (For the points where train lines intersect) how is the shift from 1 line to the next, over the hours of the day?

## 4 Analysis

With data regarding the number of people who get on at particular stations, along with their time of day, it will be possible to discover patterns of rush hour at different stations. This makes algorithm design much easier, as parameters such as the amount of time a train spends at a particular station is directly dependant on the number of people at the station.

## 5 Report

The presentation (apart from the algorithm itself) will be the justification on why different parts of the algorithm are the way they are, along with simulation on what might happen on different edge cases.