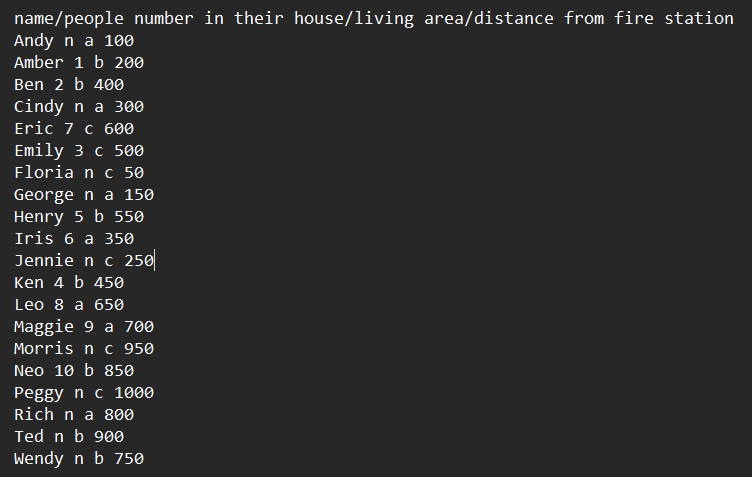
**Hw2 cc4900 Eric Chang**

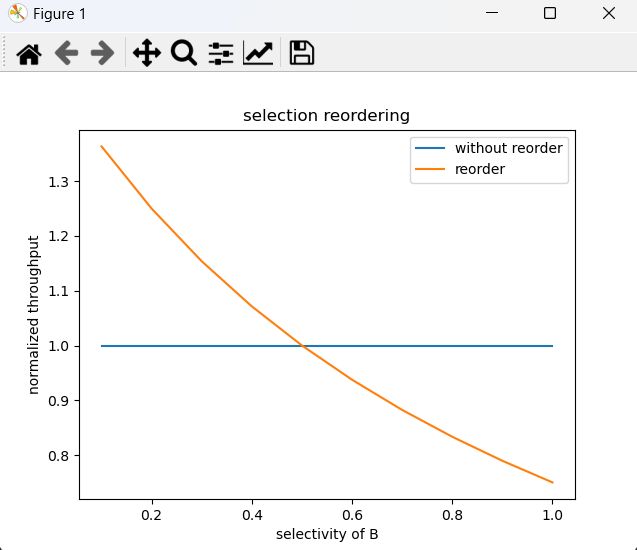
1. **Data introduction**

The picture below is a screenshot of the data. The first row to the last row represents in order: name / total number of people in the house / living area (a, b, c) / distance from the fire station. Additionally, I create 20 columns of data in total.



1. **Optimized algorithm 1: Operator Reordering**

In order to generate the comparison graph in the paper, I set some variables into the fixed number. For example, I set the cost of A and B to 1. Then, set the selectivity of A to 0.5 by using the operator “filter” to delete the data, which is not a number. Finally, use second operator “filter” to set the threshold. By using a for loop, I create ten different selectivities of B. In the end, I calculate the throughput and plot them. Below is the result.



1. **Optimized algorithm 2: Load Shedding**

In this part, I implement the shedding operator by filtering data based on the distance. In order to create different selectivity, I also deploy a for loop to generate different results. In the end, I compare the number of a,b,c by comparing to the original data in Euclidean distance. To be more specific, I normalize the distance between zero and one first and then use one minus the normalized distance as the final accuracy. Additionally, I assume C(B) is high, C(A)=shedder operator is low: C(B) = 10, C(A) = 1. Below is the result. By the way, the accuracy curve is a little different than the paper. It is because my data and the grading metric are different from the paper.

