

The Data collected is from an .The owner lives in the same apartment as me. Getting the data initially was a challenge, she was reluctant in sharing things like her fee structure which was differential in nature. I kept looking elsewhere but thankfully to the bona-fide certificate issued, she agreed instantly.

The Data I finally got was a collection of her “bank account” and “student fee and place of residence” information. The latter of which I had to enter from a handwritten copy she gave me. However during cleaning the real challenge was the bank data where I had to decipher which transaction was from which person. There was enough information in the transaction description that we could get the names of sender however due to the variety of payment methods people used, the format was problematic. I used text to columns to get the names separated and then sorted some columns with the most names. This gave me clusters of same names from where I used find, cut and paste to organize the rest manually.

Coming to the inference, from the household wise bank data (because some households sent more than 1 children ) I found that 20 percent of households gave way to 40 percent of the income. This is understandable because the prices are not that different (Rs2400 vs Rs2000) in most cases and even then the top ones are either people who have been customers for a long time or are foreigners. From the student residence/fee information I was able to get things like income streams from different cities and city wise average fee.