Abdul P Najah
Harish K Chandran
Abhishek Mishra

TA : Bhavesh Neekhra

How the Story unfolded

- Scraped all the <a> tags from CCD's website using Selenium, Python. <a> contained links to individual pages of CCD outlets
- Cleaned the data obtained in the last step to obtain all the links
- Used a script to visit all the pages and scrape addresses from them
- Cleaned the data obtained in the last step to obtain only the address and remove all the HTML texts
- Used Excel formula to obtain Pin Code from addresses
- At this stage, one of the problems that we faced: Data of three states - Haryana, Chhattisgarh, and Odisha were missing. Haryana's data was found under "Hariyana", Chhattisgarh's was under "Chattisgarh" and Odisha's data was under "Orissa".
- After we extracted the pincodes from the CCD, we downloaded district wise pin code data from data.gov.in
- We merged both the datasets and obtained the district and state information for all the pin codes of CCD using the **Vlookup** function.
- At this stage, we found 81 error messages because the pin code data of CCD had many anomalies.
- As our datasets had only 1500 data points, an anomaly of 81 data points would greatly influence the conclusions that we derive from the data.
- Therefore, each of the 81 data points was manually corrected.
- After that, we used a pivot table to analyze the data and obtain state wise and district wise rankings.
- After that, we downloaded the district wise wealth index data from NHFS and used a pivot table to obtain state wise and district wise data.
- We imported the both CCD and NFHS data set to tableau and blended the data on the basis of district.
- We created a wealth index map of India in tableau using the latitude and longitude generated by tableau and the NFHS data on the affluent population of each districts.
- Also another Indian map where we plotted all the CCD stores on the basis of its pincode.
- Then we merged both the maps using the *dual axis map* option in tableau.
- We saved the map in *Tableau Public platform* to embed it in our website as an interactive map.