As a result of the ongoing pandemic, our dependence on quality access to the Internet as well as overall online footprint has significantly increased. While people in high income and urban areas are able to take advantage of high speed and high quality of internet to perform activities ranging from learning and working from home, telemedicine and keeping in touch with distant relatives, such benefits remain inaccessible to a large number of people from disadvantaged regions. This division in access to the internet between different populations is often referred to as digital inequity or divide.

Our project aims to better understand the various dimensions of this digital inequity that exist in the USA. We have thus far gathered internet performance data from a large speedtest entity, Ookla and analysed how internet quality changes across different demographics and locations around the country. The second part of our project aims to tackle the issue of cost of access to the internet itself across different locations in the USA. There are few industries that pose the challenge of scarce publicly available data to the same extent as broadband Internet especially in the USA. To the best of our knowledge, there exists no public data set that provides internet price related information at the granularity of house address in the USA. It is our goal to create such a data set to better understand how the price/cost of access varies across different parts of the country. This we believe can better help us and policymakers guide in our attempts to bridge the ever increasing digital divide.

In order to obtain this price data from major US Internet Service Providers, our research group at the University of California Santa Barbara, in collaboration with a research team from Princeton university, have built a tool that is capable of obtaining internet bundle price information that is offered by the ISPs at a particular location. This tool can crawl the ISP APIs and scrape the necessary information (speed and price offered at a particular location) for us to conduct further analysis and reveal any interesting results/outcomes. A limitation of this approach is that by making these queries from a single IP address, we run the risk of getting either rate limited or permanently banned by the APIs. In order to prevent that, we are seeking collaboration with Bright Data to leverage the proxies and achieve our set goals.

The findings of this project will be a submission in form of a manuscript in the upcoming Internet Measurement Conference (due May 26, 2021). Additionally, our findings will have a far reaching effect in properly understanding various factors that might potentially contribute to digital divide and aid policy makers in making effective decisions.