**Proposal Group 6**

Team member : Hassan Mohamed, David Nguyen, Hwa Hara, Joshua Cressaty,

Hector Custodio, Maheen Abdulwaheed

**Dataset:**

**Internet Prices around 200+ countries in 2020-22**

<https://www.kaggle.com/datasets/ramjasmaurya/1-gb-internet-price>

<https://www.kaggle.com/datasets/prasertk/internet-broadband-and-mobile-speeds-by-country>

<https://www.kaggle.com/datasets/cityapiio/world-cities-average-internet-prices-2020>

<https://www.kaggle.com/datasets/sansuthi/gapminder-internet>

**Proposal:**

Through the above dataset, we would like to explore different characteristics of different countries' and cities' internet speed and price. Some countries will have faster internet speed and more users due to the availability of advanced technology, but some countries will have significantly fewer users and slower speed due to limited access to technology. Our group would like to explore these characteristics in one dataset table.

Initially, our general impression is that when internet speed is significantly higher and more users are using the internet in each country, the pricing should be less compared to countries with fewer internet users and with slower speed. With more widely available access and faster internet speed, we also assume each country will have faster mobile and broadband speed, as faster internet speed should indicate more advanced technology available in that country. As prices tend to increase in the city, the city internet prices should be higher compared to the cities’ national average internet price.

We will be using primarily country names to join the 4 data sets initially and will clean out null data as well as non-related columns through Jupyter notebook and python language. All datasets are constructed with .csv format, which we will read and transform into a data frame for further cleaning. We would like to focus on data columns that are easier to read and process. We will be left with the following columns to compare country name, city name, average internet speed in 2020 to 2022, broadband speed, number of internet plans available in that country, number of internet users, and internet usage rates. After merging and cleaning the data, it will be transformed and uploaded to PostgreSQL.