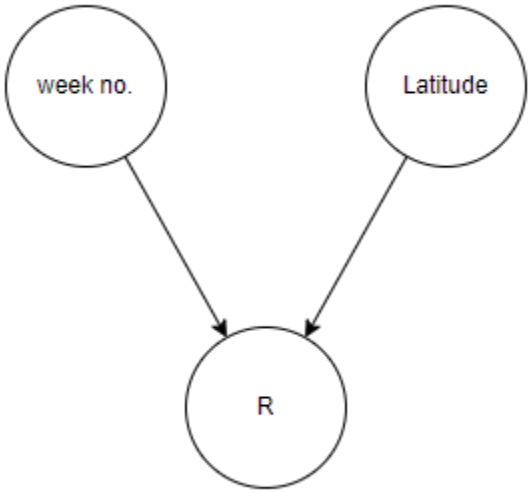
Hypothesis Proposal

- **Arun Gaonkar**

1. **Covid19 dataset – https://covid19.hpccsystems.com/**

   Hypothesis: "Geo-location and seasonality effect on COVID-19 infection rate"
   Variables:
   1. Week number: Obtained from the start date (Ranging from 01-52)
   2. Latitude : Have to obtain based on the country location
   3. R value : Infection rate, available from level1.csv

   Proposed Causal models:



Assumption: Week number is a constant factor of the year and therefore can be used as an approximation for the season of the year. The country location has an impact on the spreading of the COVID-19.

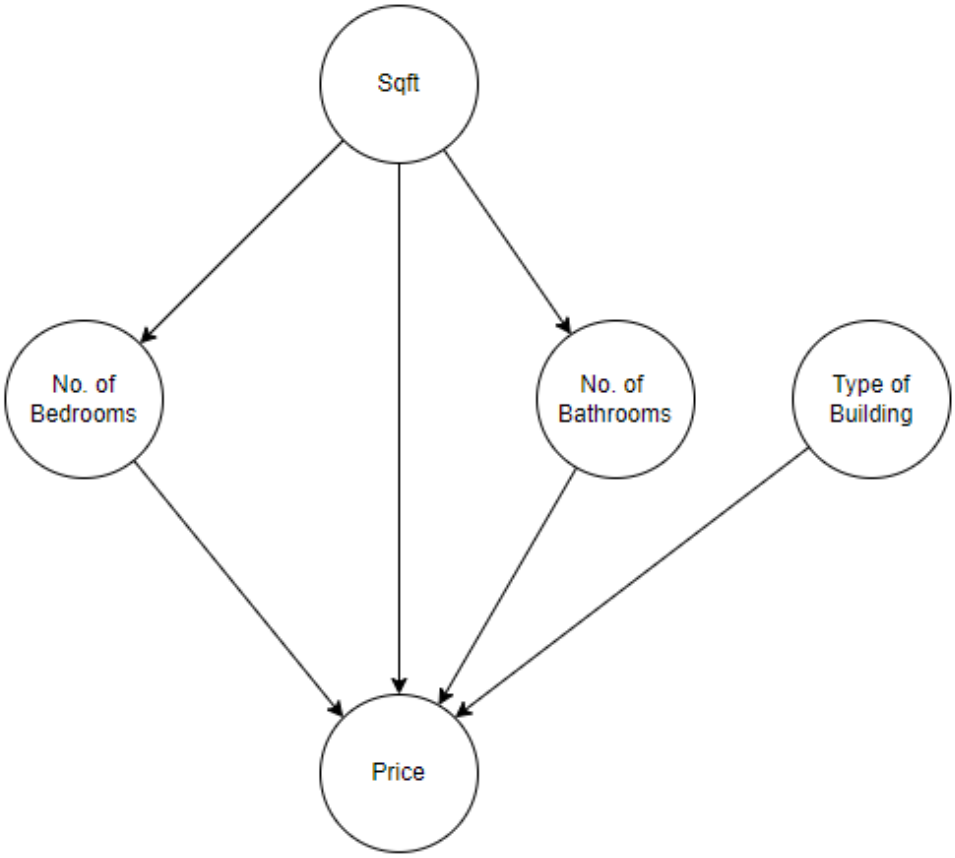
2. **Housing Dataset**: https://www.kaggle.com/datasets/austinreese/usa-housing-listings

Hypothesis : "Selling Price is affected by number of bedrooms, number of bathrooms, sqft of the house and also type of the house"

Variables:

1. Number of bedrooms : No. of bedrooms available in the building
2. Number of bathrooms: no. of bathrooms available in the building
3. Sqft of the house : Size of the building in sqft
4. Type of the house : Type of the building (Apartment, house, duplex, condo, loft, land etc.)
5. Price : Selling price of the house

Proposed Causal Model:



Assumption:

1. Sqft of the house restricts the number of bedrooms and bathrooms.
2. Type of the building can significantly increase or decrease the price. An e.g., would be apartment might be costlier than the duplex for the same sqft.
3. When buying a house, number of bathrooms, bedrooms and available space will be considered. And even type of the house might have impact.