1. Get Data
   1. Time series data of infection spread
      1. Number of confirmed cases. (cumulative)
      2. Number of casualties (?)
      3. Hubei(China)  **– get data from combined .csv - done**
      4. More – Italy, India, US (next phase 2)
   2. Time series data of Exogenous variables – (next phase)
      1. day wise average mobility (next phase 1)
      2. day wise average temperature (next phase 1)
      3. precipitation (next phase 2)
      4. cross-entropy correlation or other correlation measurements (next phase 1)
   3. Merge the data sets together to train a single model (next phase 2)
   4. Normalization needed (how?) (next phase 2)
      1. For Global Forecast (for almost unknown data)
      2. Normalization factor (N.F.) ~ Fuzzification + Defuzzification
      3. Normalize with highest value, calculate highest value using linear/logistic regressions from environment variables, if enough correlation found
      4. Normalization variables – total population (state), population density (state), education level (country), healthcare index, gdp – country
   5. Collect data of BD and make short-term prediction **(after normalization)**
2. Check Homo/Hetaroskedasticity of data (multi variable data?) (next phase 1)
3. Check ACF/PACF (for determining lag/order) (Next phase 1)
   1. increasing order improves accuracy, but slows down calculation
4. Train Model with ML/DM algoms (Next phase 1 & 2 – CNN, Ensemble, RNN) **(Sayem)**
   1. <https://towardsdatascience.com/3-steps-to-forecast-time-series-lstm-with-tensorflow-keras-ba88c6f05237>
   2. <https://towardsdatascience.com/the-complete-guide-to-time-series-analysis-and-forecasting-70d476bfe775>
5. Train Model with FTS
   1. variations – optimize hyper-parameters by grid searching **(NOW) done**
   2. (hyperparam package, genetic algorithm, hyperopt library) (Next phase 2)
6. Calculate RMSE (next phase 1)
7. Compare (next phase 2)
8. Presentation **(NOW) done**
9. Write (next phase 3)

Some Datasets:

1. <https://data.world/databrett/tracking-the-covid-19-death-rate-by-age>
2. <https://www.kaggle.com/tarunkr/covid-19-case-study-analysis-viz-comparisons>
   1. dataset + visualization + prediction
3. <https://www.kaggle.com/roche-data-science-coalition/uncover>
4. <https://www.kaggle.com/data/139140>
5. <https://www.kaggle.com/unanimad/corona-virus-brazil>
   1. time series data of infection in Brazil