1. Get Data & Clean
   1. Time series data of infection spread - **DONE**
      1. Number of confirmed cases. (cumulative)
      2. Number of casualties (needed?)
   2. Time series data of Exogenous variables – (next phase)
      1. day wise average mobility – **DONE (Apple, Google)**
         1. <https://www.kaggle.com/roche-data-science-coalition/uncover>
      2. day wise average temperature (After Studying Researches)
      3. Pearson correlation – **NOW**
         1. **Negative correlation found for 7 days lag**
         2. **Have to test up to 14 days for other countries with smoothing**
         3. **Significance test? (13-05)**
      4. Smoothing up to +/- 7 days after extrapolation? (13-05)
   3. Merge the data sets together to train a single model (**13-05**)
   4. Normalization needed (how?) (**12-05**)
      1. First test with divide by max **NOW**
      2. For Global Forecasting (for almost unknown data)
      3. Normalization factor (N.F.)
      4. Normalize with highest value,
         1. calculate highest value using linear/logistic regressions from environment variables, if enough correlation found
         2. collect data of environment variables (Tasmia)
      5. Normalization variables – total population (state), population density (state), education level (country), healthcare index, gdp – country
      6. <https://www.kaggle.com/paultimothymooney/oxford-covid19-government-response-tracker>
   5. Collect data of BD and make short-term prediction **(14-05)**
2. Check Homo/Hetaroskedasticity of data (multi variable data?) Needed?
3. Check ACF/PACF (for determining lag/order) **- DONE**
   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>
   3. <https://medium.com/themlblog/time-series-analysis-using-recurrent-neural-networks-in-tensorflow-2a0478b00be7>
5. Train Model with FTS
   1. variations – optimize hyper-parameters by grid searching - **DONE**
   2. (hyperparam package, genetic algorithm, hyperopt library) (After project separate work)
6. Calculate RMSE - **DONE**
7. Compare – **(16 – 05)**
8. Update Presentation **done**
9. Write **(17/05 – 20/05)**

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
6. <https://www.kaggle.com/sudalairajkumar/covid19-in-italy> (ITALY)
7. <https://www.kaggle.com/sudalairajkumar/covid19-in-usa> (USA)
8. <https://www.kaggle.com/rohanrao/covid19-forecasting-metadata> (CHECK IT OUT)
9. <https://www.kaggle.com/kimjihoo/coronavirusdataset> (S. Korea)
10. <https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset>