1. Study 1 publication, write down the related works (0.5 page)
   1. need 1 publication to justify that Corona takes 7-14 days to show symptoms
   2. got 2, read, and written, done
   3. got 2 more read and written, done
   4. will check 3 more
2. Catch up with previous codes, and set code ready for 3 countries with recent data – Brazil/US/India + Bangladesh
   1. doing
   2. normalize data starting from 7 days before first confirmed case and with max value (x and y axes)
   3. correlate with Bd
   4. merge data (how?) and develop model
   5. need to develop model with separate countries and compare result with BD (?)
3. Write down the experiment procedure and result for ACF, FTs, mobility correlation, etc. (2 page)
   1. checked mobility correlation for Brazil/Italy/India, 0.5, too low to be significant
   2. check again for values of 2 - 8 days lag with more recent data and find out the one with highest correlation value
4. Add casualty prediction

Fugure Work:

1. Prepare Machine Learning code for PNN+cf or, RNN for comparison
2. Write down experimental procedure (2+1 page)

June 4:

1. Related Works (1.5 page)
2. Introduce the data sets (1 page)
3. Introduce our work (Abstract, Introduction) – 1 page
4. Experimental Procedure (3 page)
5. Present the results (2 page)
6. Conclusion + References (1 page)

Updated Work Plan:

1. Time Series data of infection spread – Collected
2. <https://www.kaggle.com/rohanrao/covid19-forecasting-metadata> **NOW**
3. Time Series data of Normalization factors –
   1. Healthcare index (Ordinal) – Collected
   2. Population – total, density – (<https://www.kaggle.com/paultimothymooney/oxford-covid19-government-response-tracker> )
   3. Govt. measures effectiveness - ??? **NOW**
   4. Mobility effect - <https://www.kaggle.com/rohanrao/covid-19-forecasting-modelling-with-mobility> (not accepting)
4. Normalization factor calculation – 13/05
   1. Linear/Logistic regression
   2. Check with real data
   3. RMSE
5. Merge normalized data for training FTS models – 14/05
   1. Endogenous variable only
   2. De-normalize data, compare result and RMSE with actual result
6. Update Kaggle Notebook – 15/05, 16/05
7. Writing results – 17/05, 18/05

Past Work Plan:

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 – **Studying**
         1. **Temporal relations?**
         2. **Negative correlation found for 7 days lag**
         3. **Have to test up to 14 days for other countries with smoothing**
         4. **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. **Logarithmic scale**
      3. For Global Forecasting (for almost unknown data)
      4. Normalization factor (N.F.)
      5. 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)
      6. Normalization variables – total population (state), population density (state), education level (country), healthcare index, gdp – country
      7. <https://www.kaggle.com/paultimothymooney/oxford-covid19-government-response-tracker>
   5. Collect data of BD and make short-term prediction **(14-05)**
2. **Check the dampening effect of mobility reduction on exponential growth**
   1. [**https://www.kaggle.com/agneshui/worldwide-measures-implementations**](https://www.kaggle.com/agneshui/worldwide-measures-implementations)
3. Check Homo/Hetaroskedasticity of data (multi variable data?) Needed?
4. Check ACF/PACF (for determining lag/order) **- DONE**
   1. increasing order improves accuracy, but slows down calculation
5. 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>
6. Train Model with FTS
   1. variations – optimize hyper-parameters by grid searching - **DONE**
   2. (hyperparam package, genetic algorithm, hyperopt library) (After project separate work)
7. Calculate RMSE - **DONE**
8. Compare – **(16 – 05)**
9. Update Presentation **done**
10. 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>