**Homework – Mini Project on COVID data “Look what I have discovered”**

**FINAL PRODUCT**: *Power point presentation with plots of timeseries of selected COVID statistics – the title of presentation says it all “****Look what I have discovered!”.***  *Tell us what you did, what worked and what did not. This is fine. It makes presentation more interesting and we all can learn from your experiences.*

*Make sure you at least attempt to explain your results in your ppts. Include sql queries. If you want include any other files, if you feel they are important. But ppt is a must.*

*Since students love rubrics. To get an A – make sure you have at least 6 plots and good explanation why you selected them and how did you preprocess the data. Attempt to interpret what you found.*

*Best work will be presented by their authors. It is individual project. We will of course select the top presentations for our 336* ***hall of fame.***

**What to do?**

There is a multitude of online data about the COVID-19 disease. In this project you will use the data provided by <https://ourworldindata.org/policy-responses-covid>,

More specifically, first you will have to download the data, create your own database schema and import the data to your database on your local mysql server.

You are required to to use more information about countries from any other sources you may find on internet (and there is plenty of them). Any socio-political, economical, or health and even covid related categorizations can be used. You should then bring this info to your database and join in with the covid data you already have.

Bring at least one such external “attribute”, called it E (for external) which applies to all countries.

For example, you might find what covid measures different countries applied (lockdown, partial lockdown, no lockdown). Or, government type attribute – what type of political system they have? What economic category countries belong to?

Your objective it to investigate if your external attribute, E, has had any impact on the progression of pandemic. Better select E which does! Otherwise it will be….boring.

You never know, you might discover something which is not known yet! It is time to be creative.

Next step is to select two statistics which you will plot (say cases and deaths) for different values of your attribute E. For different values of E, you should built aggregated time series of stats which you selected. Aggregated time series will add daily stats of all countries which share the same value of attribute E. Say, if E was a the covid mitigation method, you will aggregate all time series for E= ‘lockdown”, or E =’partial lockdown’.

Finally, you will have to plot these aggregated time series for stats which you selected and pick the most interesting plots for your presentation to contrast (look how differently Covid progressed in G20 countries than in the remaining countries)

**TOOLS:** Mysql, Plotting of your choice. Highcharts are nice. But it is up to you. Also we do not care if you plot off excel or csv etc. This project does not require our live access to your database. Use the database, show how you used it, but show in your power points show only final results. And make it best you can, so you will be one of the presenters.