# JOURNAL - MARIA MIGROVA

**29/10/2021**
The first team meeting on Microsoft Teams. Idea of covering Covid-19 on a country basis as the main topic of our project.
Each team member picked a country.
Duration: 30 min

**4/11/2021**
Searching for possible datasets. Call about possible datasets found on GitHub.
Duration: 30 min

**5/11/2021**
Discussing formats of the journal, questions our dataset is going to answer. Potential use of AWS, discussing problems which can happen at the end of the project when we will connect our datasets together into one.
Duration: 30 min

**12/11/2021**
Deciding on the topic of the project. There were different suggestions.
Polina suggested using health as the main topic of the project.
Evan suggested using covid as the general topic.
We decided to use Covid 19 as the main topic of the project.
Due Dates:
- 15th November = to have datasets chosen for the work.
- 17th November = to have the Git repository established.
- 5th December = soft deadline for code section to be finished.
- 15th November = next group meeting
Duration: 30 min

**13/11/2021**
Searching for a suitable dataset which can be used for the multiple linear regression and also for the individual analysis.
It was  very difficult to find something suitable but by the end I decided to use testing and restrictions. These datasets were all downloaded from this site:
https://ourworldindata.org/coronavirus

**15/11/2021**
Git repository was created by Polina Prinii.

**16/11/2021**
Presentation by Polina Prinii about how to use GitHub.
Duration: 1 hour

**24/11/2021**
Team meeting with our lecturer. Making some changes to our project.
Aim of the project: Perform individual analysis on the chosen dataset with the final multiple linear regression performed together.
Group Action: Create final master table used for multiple linear regression.
Due date: Individual analysis - 5th December
Duration: 1 hour

**3/12/2021**
Discussing overall project approach. Deciding to use my idea about providing a finalised csv file by joining our individual datasets with Python. Polina presented her work to get some ideas.
Actions:
 6th December:
- Update main README
- Finalise csv file for merging
- Create main folder in Git

Duration: 2 hours

**4/12/2021**
Updating the main README

**5/12/2021**
Reading the separate datasets for restrictions and testing into R using read.csv() function. Cleaning the dataset and  changing the date from 1/1/2021 until 06/12/2021. Joining all the data into one final dataset which will be used for multiple linear regression. This file is called "finaldataset.csv" and contains these variables:
- Isocode
- Country
- MaxStayHome
- MeanStayHome
- MaxStringency
- MeanStringency
- MaxFacialCoverings
- MeanFacialCoverings
- MaxCancelPublicEvents
- MeanCancelPublicEvents
- MaxClosePublicTransport
- MeanClosePublicTransport
- MaxSchoolClosure
- MeanSchoolClosure

Starting to work on the individual mini analysis.

For the first analysis I decided to use the testing data for Slovakia. I decided to use this data to create a Random Forest model. This model predicts a probability of being n. Of PCR Pos bigger or smaller than PCR Neg for different age group and gender.

**6/12/2021**
Another analysis was a SVM linear regression. I decided to find a relationship between the number of cases and other restriction variables. For this I used the svm() function. I also used MAE, MSE. MAPE and RMSE to evaluate the model. The model was very strong. Few graphs were created as well.

The third analysis was about comparing a stringency index of the countries from central Europe between 1/1/2021 and 15/11/2021.
Firstly I started by filtering the data by dates and then filtering the data by the central Europe countries.
Then I used this function mutate(paired = rep(1:(n()/2),each=2),to assign each data to a group 1 or 2 (by the date). Then I created a lollipop graph with two points from this data and saved the graph using ggsave() function.

**6/12/2021**
Group members provided updates about their first phase analysis, intentions and completed analysis. Polina had issues with a date range. Evan updated the group with intention to provide a final csv file for MLR tomorrow (7th December).
Maria conducted merging of available files for MLR in Python to create 1 dataset.
Actions:
7th December:
- Review and finalise individual csv datasets with date 24th November
- Complete the README file
Duration: 3 hours

**7/12/2021**
Evan decided to leave the group, therefore  there will be only 3 members remaining for this project. Group members provided their final datasets for the MLR.
Polina decided to add the data for the latest date which contains the vaccination figures.
Maria finished the merging of the individual datasets into one final dataset.
Actions:
9th December:
- Alun to update the datasets to consists of country names equal to other datasets
- Finalise the README file
- Finalise the individual mini analysis


**8/12/2021**
Starting the multiple linear regression in R. Maria created the R Markdown file and explained the difference between R Markdown and R Script.

Maria undertook the coding in R, and achieved following results:
- Imported the csv final dataset into R
- Ran the dataset within MLR using lm() function.
- We found 3 violations of the assumptions.
- Decided to address the violation and continue the work tomorrow.

Actions:

10th December:
- Polina to address the violation of the heteroscedasticity
- Finalise the README file
- Finalise the individual mini analysis
- Continue with the regression analysis tomorrow

Duration: 2.5 hours

**9/12/2021**
Covering Polinas progress in R with heteroscedasticity.
Covered by potential issues of missing death values.
Regression model has improved by 8% from yesterday.
Maria to update the R code with correct code for removing NA values.
Duration: 4.5 hours

**10/12/2021**
Continuing the work on the individual analysis.

**13/12/2021**
Came to an agreement that the linear regression is complete. We came to the model with R squared equal to 0.94. Maria proposed to complete the regression analysis with the following functions:
MAE, MAPE, MSE.
Next Steps:
Compile functions to evaluate the fitness of the model.
Start on the report.
Compile the visualisations.
Actions:
- Maria to complete MAE, MAPE and MSE.
Report Actions:
- Abstract = Group writing
- Introduction = Alun Price
- Related Work = Polina Prinii
- Methodology = Polina Prinii
- Results and Evaluation = Maria Migrova
- Conclusion and Future work = Group writing
- Bibliography = Group writing
Alun to finish the group visualisations:
- Heatmaps

- Barcharts
- Waterfall charts

Due dates:
- Individual mini analysis - 16th December
- Project Report - 18th December
- Video - 21st December

Duration: 2 hours

**14/12/2021**

Continuing the work on the individual analysis.

**17/12/2021**

Writing the report Results and Evaluation part.

**18/12/2021**

Continuing the work on the individual analysis. Problems with creating an interactive graph in R as there were problems with saving it. Therefore I decided to use Python as well for creating 2 of my graphs.

**19/12/2021**

The report was completed by Polina Prinii by joining all our individual parts together. Alun volunteered to write the Abstract part and Maria volunteered to write the Conclusion and Future work part.

The report was finalised and submitted to Moodle by Polina Prinii.

Actions:
- Starting our work on the video on 20/12/2021
- Working on the individual journals which have to be submitted by 22nd December.