COURSE PROJECT (15%)

Objectives:

- Practice reading data from different data sources (Extract)
- Preprocess dataset(s) (Transform data into a target format)
- Loading datasets
- Apply data mining algorithms
- Visualize dataset(s) using different commands
- Analyze datasets and report results

Requirements:

- Find an online resource for COVID-19 dataset. The dataset could belong to a specific country or the whole world.
- describe the dataset and its attributes.
- Find out a business problem related to the dataset; either from the data source or from any other source in which data scientists are trying to answer.
- Describe the analytics approach that is used to answer the business question.
- Analyze the dataset.
- Describe the results and report your findings.
- **To submit:** copy all your code and description of datasets into a a MS-word file **and Save it as pdf. Submit into Blackboard together with the dataset in a zipped folder and submit into BB.**

Rubric		
Rubric	**Detail**	**Weight**
Dataset Description	• Data set described and a clear business problem discussed	**2**
Data Extraction and loading	• The correct code to extract the dataset from its source is written	**3**
Data Preprocessing	• Data preprocessed and different statistical analytics/ Visualization methods applied	**3**
Graph analysis/ Interpretation and reporting	• Different relationships between variables based on the resulting graphs are described • Is there any relationship or correlation. • If yes, what type of correlation is it? • How is the distribution of data?	**2**

Data Mining	• Describe the analytics approach that is used to answer the business question is Explained • the data mining Task/algorithm applied and data analytics model is built.	**4**
Data visualization and Reporting	• dataset(s) using different commands is Visualized • Datasets are analyzed and results reported	**1**

Variables:

- **num_positives**: the number of positives in Tokyo
  https://catalog.data.metro.tokyo.lg.jp/en/dataset/t000001d0000000011

- **num_tests**: the number of total Covid-19 tests in Tokyo
  https://catalog.data.metro.tokyo.lg.jp/en/dataset/t000010d0000000086

- **num_consult**: the number of consultations at the novel coronavirus call center in Tokyo
  https://catalog.data.metro.tokyo.lg.jp/en/dataset/t000010d0000000071

- **num_hospital_patients**: the number of hospitalized patients
  minor_moderate_sym: the number of patients whose symptom is minor or moderate
  severe_sym: the number of patients whose symptom is severe
  https://catalog.data.metro.tokyo.lg.jp/en/dataset/t000010d0000000092

- **highway_traffic**: Number of traffic on the Metropolitan Expressway.
  https://www.shutoko.co.jp/company/database/traficdata/

- **air_domestic_passenger**: the number of domestic passengers in Haneda airport.
  air_foreign_passemger: the number of foreign passengers in Haneda airport.
  air_total_passenger: the total number of passengers in Haneda airport.
  https://www.tokyo-airport-bldg.co.jp/

- **restriction_period**: whether the government or Tokyo metropolitan government announced the restriction (either State of Emergency Declaration or Semi-emergency Coronavirus Measures) in each month or not.

- **unemployment_rate**: The unemployed are people of working age who are without work, are available for work
  https://data.oecd.org/unemp/unemployment-rate.htm

- **telework_rate**: the percentage of teleworking companies in Tokyo
  between 2021-04 and 2022-08:
  https://www.metro.tokyo.lg.jp/tosei/hodohappyo/press/2022/09/12/04.html
  2020-03, 2020-04, 2020-12 & between 2021-01 and 2021-03 (mean of 2 rates; first and second half). Between 2020-05 and 2020-11 is projected from the graph, which means that it's not precise data.
  https://www.metro.tokyo.lg.jp/tosei/hodohappyo/press/2021/12/09/06.html