

# Final Project (Advanced Database Systems)

Jong-Kyou Kim, PhD

2022-11-01

## 1 Introduction

- Analyzing practical size data using SQL and a general purpose programming
- Size matters

## 2 Schedule

- Nov 16 (Wed): Proposal submission due
- Nov 30 (Wed), Dec 7 (Wed): Presentation
- Dec 21 (Wed): Final report

## 3 Requirements

- Minimum requirement (60%)
  - Convert a given small-size dataset into SQL
  - Descriptive analytics
    - \* Basic statistics (min, max, average, median)
    - \* Visualize the description (such as correlation heatmap)
  - Submit the database as sqlite3 with SQL statements to generate the results
- Intermediate requirements (20%)
  - Design a schema that reduces the redundancy of analysis
  - Apply to the intermediate-size set
  - Analyze the operations and add indices
  - Submit the schema, the ER diagram and SQL statements to generate the results
  - Submit the database as sqlite3
- Advanced requirement (A. use an intermediate-size dataset) (20%)
  - Choose a model such as the linear regression or the artificial neural network

- Convert the data to fit the model
  - Save 20% of data for testing (the test dataset)
  - Evaluate the predictive model using the test dataset
  - Submit the code and the converted data
- Advanced requirement (B. use a large-size dataset) (20%)
  - Choose an advanced database
  - Migrate the given data to the advanced database
  - Provide descriptive analytics for the large-size dataset
  - Visualize the description
  - Submit the codes that are used to generate the analytics and the visuals
- Bonus (10%)
  - Effectiveness of schema design
  - Interesting analytics and/or visuals
  - The performance of a predictive model