MIS2502: Exam 2 Study Guide (Spring 2022)

Instructor: Youngjin Kwon

The exam will be a combination of multiple-choice and short-answer questions. It is a closed-book, closed-notes exam.

The following is a list of items that you should review in preparation for the exam. Note that *not every item on this list may be on the exam, and there may be items on the exam not on this list.*

**SQL Out: Advanced Queries (Joins, Subselects)**

* Given the schema of a database, be able to create the SQL statements that
  + Require a join of multiple tables
  + Contain a Subselect statement  
    (i.e., determine the customers with the highest sales)
* Understand how to use different statements (e.g., WHERE, GROUP BY, ORDER BY) on top of join/subselect statements

**Semi-Structured Data**

* What is semi-structured data? Examples? What does it mean to have no formal data model?
* What is unstructured data? Examples?
* Compare csv, XML, and JSON data formats and explain advantages/disadvantages of each
* Construct a csv, XML, and JSON data file from raw data

**NoSQL**

* Understand how RDBMS and NoSQL database are different
* Understand how different operators work and what is the correct order to use them
  + $project, $match, $group, $sort, $limit
  + $and, $or, $eq, $gt, $lt
* Understand how to query array and embedded documents and join collections
  + Dot notation
  + $all, $size, $lookup

**ETL**

* What is it? Why is it important?
* Explain the purpose of each component (Extract, Transform, Load)
* How do inconsistencies in data get resolved?

**Data Visualization**

* Be able to assess a visualization by applying data visualization principles.
  + Tell a story
  + Graphical integrity
  + Minimize graphical complexity
* Explain how a visualization can be improved based on those principles.
* Understand basic chart types. Be able to choose an appropriate chart type given a scenario.

**\*\*Advanced Analytics and Python will be covered in the third exam.**