# PROJECT PHASE I-SPECIFICATION

# Nov 24th (Load your report to [DYS](mailto:tsuzek@gmail.com) until midnight AND bring 1 print-out to class)

# **Do not put your print-out in plastic bag, 50 points (out of 100) will be deducted.**

1. **For each day of delayed report, 10 points (out of 100) will be deducted.**
2. **Not inserting 1 row of data to every table in your mwb file, 100 points will be deducted.**

The final project in this class will give you the opportunity to design and implement a small database system in a **domain of interest** to you.

Each group will select a separate task particularly relevant to their outside interests, such as an investment portfolio database, a medical database, an astronomy database, a pharmaceutical database or a small business database.

The purpose of the first phase of the project is for groups to specify which domains and tasks they plan to cover (to confirm that they are suitable and sufficiently challenging) and to begin the database design and implementation process.

**You should have presented your data and taken prior permission from Dr.Tugba during the recitation class.**

In general, your final project will constitute an interface for data input (either from the user entering the data manually using stored procedures, the MySQL WWW interface or Java or Php-based data-extraction from text or on-line data sources), and an interface for data output (through stored procedures and views to provide customized perspectives on the database for different users).

# PHASE I requirements:

In your **written** report, please respond to **all 8** questions:

1) Who are your team members? Each team may consist of up to 3 members. 1 or 2 person teams are not allowed and you will get 0 points unless you have a 3 member team.

2) Explain your data:

* What’s the name/purpose of your project? Why did you pick this topic?
* Provide the links of the **raw data** of your project (e.g. kaggle address, ftp site address etc).
* How many files exist?
* How many rows and columns do they contain?
* How many string(non-numeric) columns do you have?
* Provide the links of the **raw data** of your project (e.g. kaggle address, ftp site address etc).
* **Provide the Bitbucket URLs of your ER diagram (.mwb files) in your report AND UPLOAD THE REPORT TO DYS.**

3) Give a reasonably comprehensive and representative list of the English questions you would like your system to be able to answer (find at least 10 questions containing min/max/avg/sum/count ).

For example,

“Compute the mean expense spent to hobbies by family members, grouped by adult or children"

4) Design and show a relational data model (ER diagram) in MySQL Workbench.

In your ER diagram **insert one row of sample values for each relation**. Define all your primary keys and indicate all referential constraints with arrows between foreign keys and the relation/attribute they reference. **Again do not forget to provide one row of sample values from your data!!**

5) Write at least 5 SQL statements that will implement the English questions from of your target queries (from Question 3). **You will need to provide at least 1 aggregate function with group by.**

6) Provide a proposal of how you will load the database with values.

* If you plan to extract/import data from on-line sources, briefly describe what are the sources (e.g. personal data, or WWW URL's) and what any format conversion issues you expect to encounter are.
* If you plan to input your data primarily through a WWW or form-based inter-face, briefly describe this interface and the issues involved.

7) Are you going to implement any front-end user interface? If not, at the minimal you need to indicate here that you will at least implement 1 view and 1 stored procedure with IN, OUT and INOUT.

8) Describe the database platform you plan to use (i.e. MySQL 5.0 on your home computer Pentium II with 26 MB of memory, etc.), including any relevant implementation details or challenging issues.