# Introduction to Data Mining
# Individual Project II: Association Rule

*Due: 10/26/2015 - 10:30am*

## What to Deliver

You should submit:

- A *Readme.txt* file, explaining how to run your script. (If there is any ambiguity in working with your functions)
- A report clearly showing the requested items in the **Project Descirption** and **Dataset** parts.
- A .R script, containing every function definintion. By running this script, we should be able to get your reported results.

Submit your files as a .zip file: **FirstnameLastname_UFID.zip**.

## Project Description

For the given datasets, you should apply two different association rule mining techniques you learned in the class: **Apriori** and **FP-Growth**. For each dataset and method the following steps are mandatory:

- Finding all the rules. ( *It is not needed to be in the final report, but it is mandatory that your program produces it.*)
- Finding rules which meet the requrested criteria:
  - Support = 0.01
  - Confidence = 0.90
- Removing redundant rules and listing unique ones.
- Sorting the remained rules based on their *lifts*.

As an example:

```
#Load your packages
...
#Read Datasets
...
#Printing all of the rules
rules <- myAssociationRuleMining(myData)
inspect(rules)
#Rules which meet the criteria (such as support and confidence)
betterRules <- myAssociationRuleMining(myData, parameter = list(supp = 0.01, conf = 0.9))
inspect(betterRules)
#Removing redundant rules
redundant <- findRedundantRules(betterRules)
uniqueBetterRules <- betterRules[!redundant]
```

```r
# Sorting based on lift
sortedUniqueBetterRules <- sort(uniqueBetterRules, by = "lift")
inspect(sortedUniqueBetterRules)
#Detaching packages
...
```

## Datasets

There are three different datasets, for each case (as well as reporting the above-mentioned results) you should provide some extra information.

### Titanic

This is an easy example just to get you ready. In this dataset there are four columns: *Class*, *Sex*, *Age*, and *Survived*.
Report those rules which have *Survived* (either YES or NO) in their right hand side. This should be an output of your script too.

### Retail

In this dataset, there is a list of 10000 transactions in a grocery store.
Report those rules which have *Beverage* or *Meat* or *PersonalCare* on their right hand side. Which items (or lack thereof) would result in purchasing (or forgoing) such items.
Do not forget to exclude *TransactionNo* for your analysis.

### Game of Thrones

The list of all characters in the first five books of the Game of Thrones is provided. The features are

- Name (Exclude this for your analysis)
- House: the allegiance
- Gender: male or female
- Nobility: True or False
- Survives: True or False
- $Book_i$: True or False (mentioned in the book $i$)

Report the rules which have *Survives* in the right hand side.
Does nobility play any role in survival? What about gender?
In the first five books, *Jon Snow* lives. Do you find any rule which says otherwise? (like his fate in the Game of Thrones TV-Series) Select your own favorite character (someone other than *Jon Snow*) and see if you have any rule for his/her survival?

Good luck and Valar Morghulis