COMP 333 — Week 9 Recap

Exploratory Data Analysis

In Week 8 and Week 9 we cover Exploratory Data Analysis(EDA).

EDA involves Data Wrangling, including Data Cleaning

EDA involves repeated Descriptive Data Analysis (DDA)

EDA is about *exploring* the data,
the relationships between variables,
to gain *insight*
so you can answer your questions
and achieve your goals.

EDA builds *models* to capture the insight
to predict outcomes in new situations
as aids to decision-making.

You should revisit Example 3 of Week 2 on the Titanic
as it is a very good introduction to EDA.
A Gentle Introduction to Exploratory Data Analysis by Daniel Bourke
`https://towardsdatascience.com/a-gentle-introduction-to-exploratory-data-analysis-f11d84`

We will return to the topic of modeling
in the lectures on Machine Learning.

READ the files marked READ.

For these topics, it is very worthwhile to also read/watch the supplementary material.

For Future Reference

There is an extensive, long tutorial on EDA

Exploratory Data Analysis in Python
By Chloe Mawer and Jonathan Whitmore,
Silicon Valley Data Science, 2017 ay PyCon.
`https://www.kdnuggets.com/2017/07/exploratory-data-analysis-python.html`

with an almost 3-hour video taking you through the notebooks at
`https://github.com/cmawer/pycon-2017-eda-tutorial`

You do not need to read/watch/work through the tutorial for this course.

However, it shows a ton of exploration,
with lots of "questions" being explored
with real world datasets.

So, for revision, or for more learning after the course,

when you have lots of time,

it is here for your future reference.