UNIVERSITY OF MALAYA

MIDDLE SEMESTER EXAMINATION FOR THE DEGREE OF MASTER OF DATA SCIENCE

ACADEMIC SESSION 2017/2018 : SEMESTER 2

WQD7001 : Principles of Data Science

March 2018 Time : 2 hours

INSTRUCTIONS TO CANDIDATES :

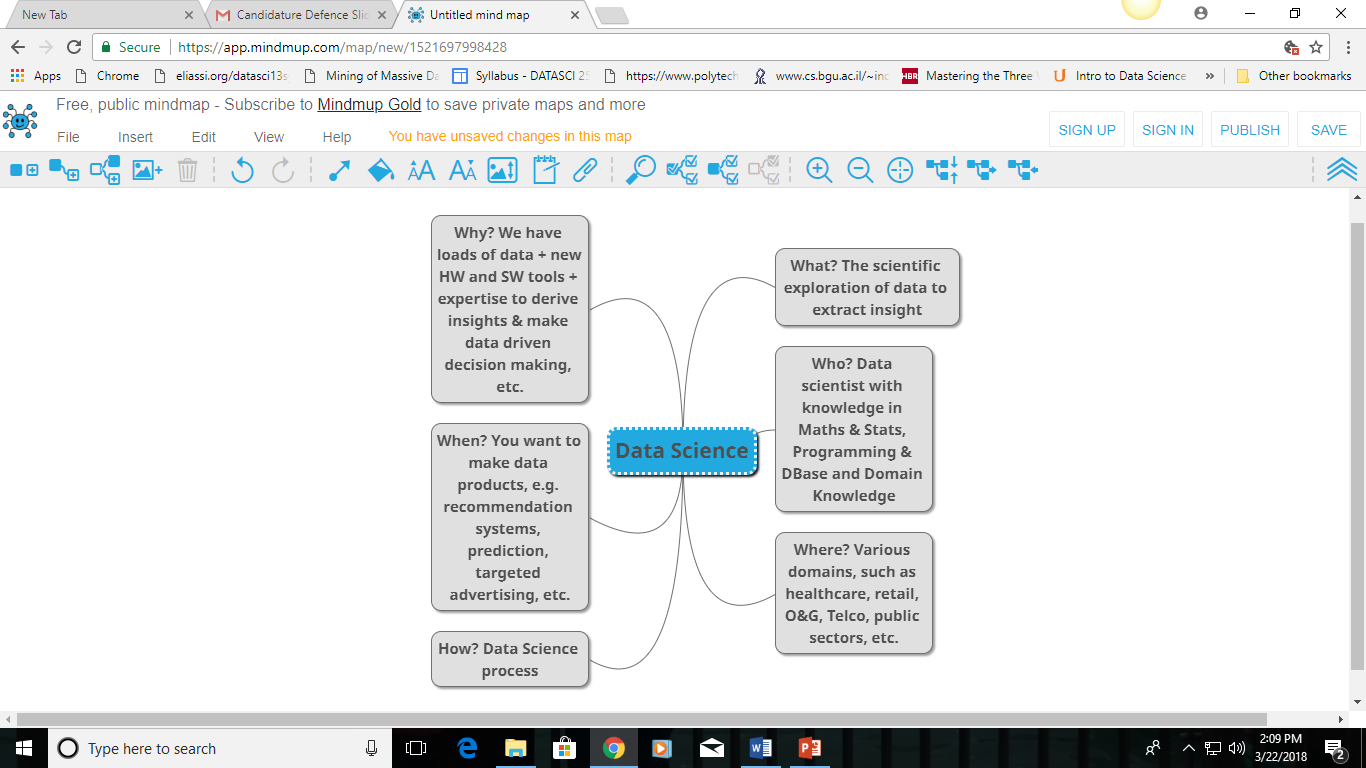
Answer **ALL** questions (50 marks).

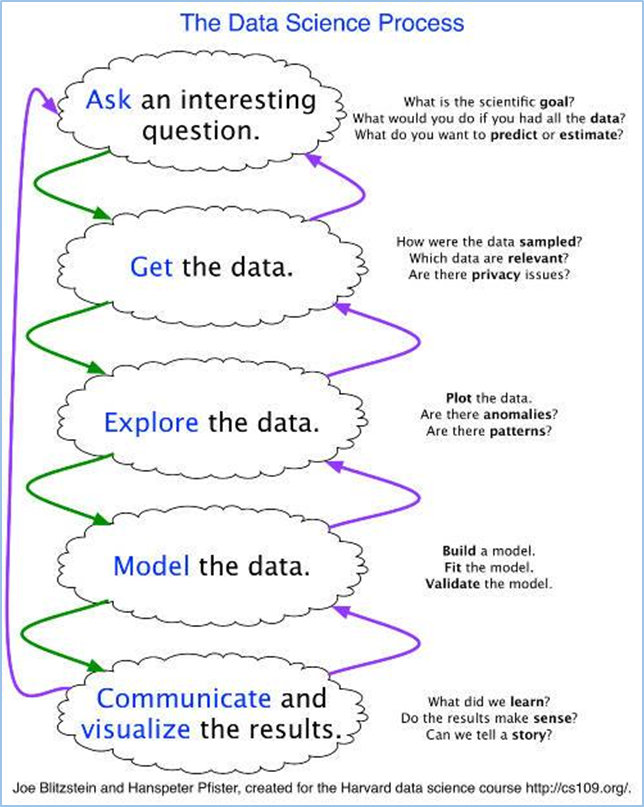
**ANSWER SCHEME**

(This question paper consists of 5 questions on 2 printed pages)

1. Based on your own understanding and using your own words, explain what data science is. You must use mind map or 5W 1H or drawing or any best and interesting way to assist your explanation, instead of just words.

[10 marks]





1. The FCSIT has loads of students data. The faculty needs to unlock some valuable insights from the data. Given a set of students’ data in the appendix, suggest what insights that could be derived from this dataset. Do state any assumptions, such as creating variable(s), that you make in this learning analytics task.

[10 marks]

Answers varies.

1. (a) How datafication and analytics are related?

[4 marks]

Datafication is putting data in a quantified format so it can be tabulated, analyzed, and enables prediction.

Analytics is the science of understanding the cognitive processes an analyst uses to understand problems and explore data in meaningful ways.

Data explosion is the fuel for analytics. Data on every aspect of our life is now tracked and stored in a quantified format, in databases and analytics allows us to turn this data into insights. For example, wearable devices, such as smart watches, Google glass and pedometers collect data.

(b) Missing data in general is one of the trickier issues that is dealt with when cleaning data. The simplest solution is to delete rows with missing values. What is your advice to a data scientist who chose this option?

[3 marks]

The first is that this approach only makes sense if the number of rows with missing data is relatively small compared to the dataset. If you are finding that you will be deleting more than around 10% of your dataset due to rows having missing values, you may need to reconsider.

The second issue is that in order to delete the rows containing missing data, you have to be confident that the rows you are deleting do not contain information that is not contained in other rows.

(c) You are finding some json files data for your project. What would you add to your Google query to look for files with a json extension?

[3 marks]

Add "+filetype:json"

1. You are a CIO who is hiring a few data scientists to help shape the future of your company. You intend to look at their portfolios when making hiring decisions. Anticipate at least **FOUR** things that you expect to see in an impressive data science portfolio.

[10 marks]

When employers hire, they're looking for someone who can add value to their business. They will look at their Qualificiation, Knowledge, Skills, Attitude components. Most importantly their real world experience in DS.

4 components:

i. Candidate’s particulars, like in CV.

ii. Candidate’s DS knowledge & skills & attitude

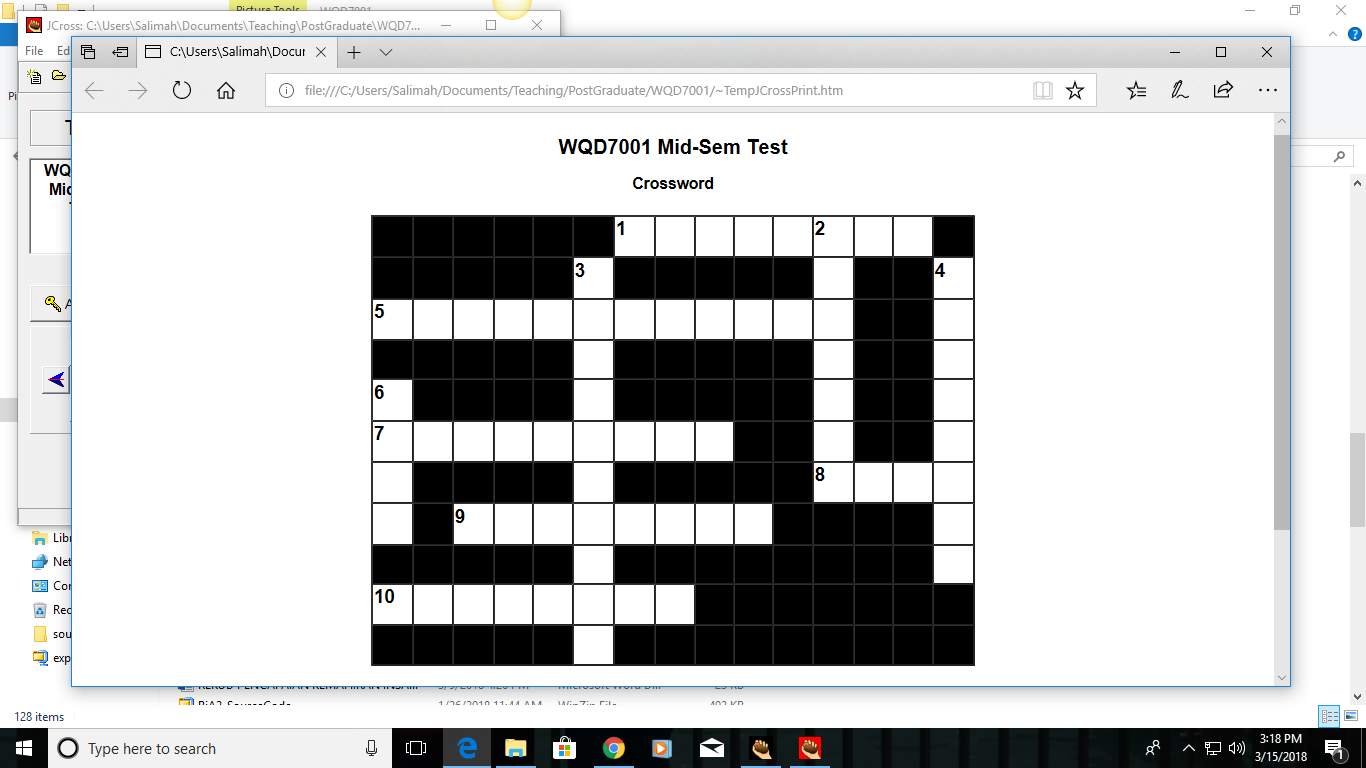
iii. Candidate’s real world experience doing DS work

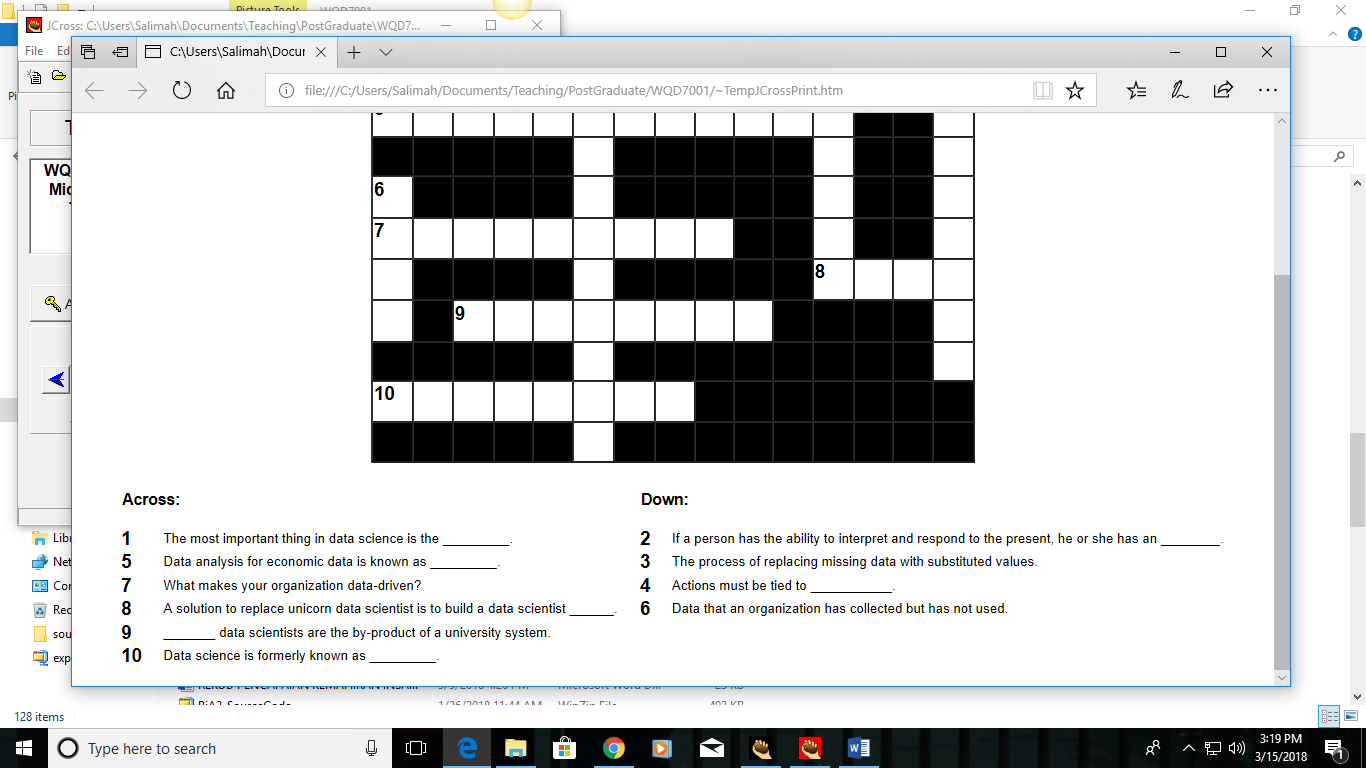
iv. Candidate’s contribution on DS, like participating in competition, bootcamp, talk, etc.

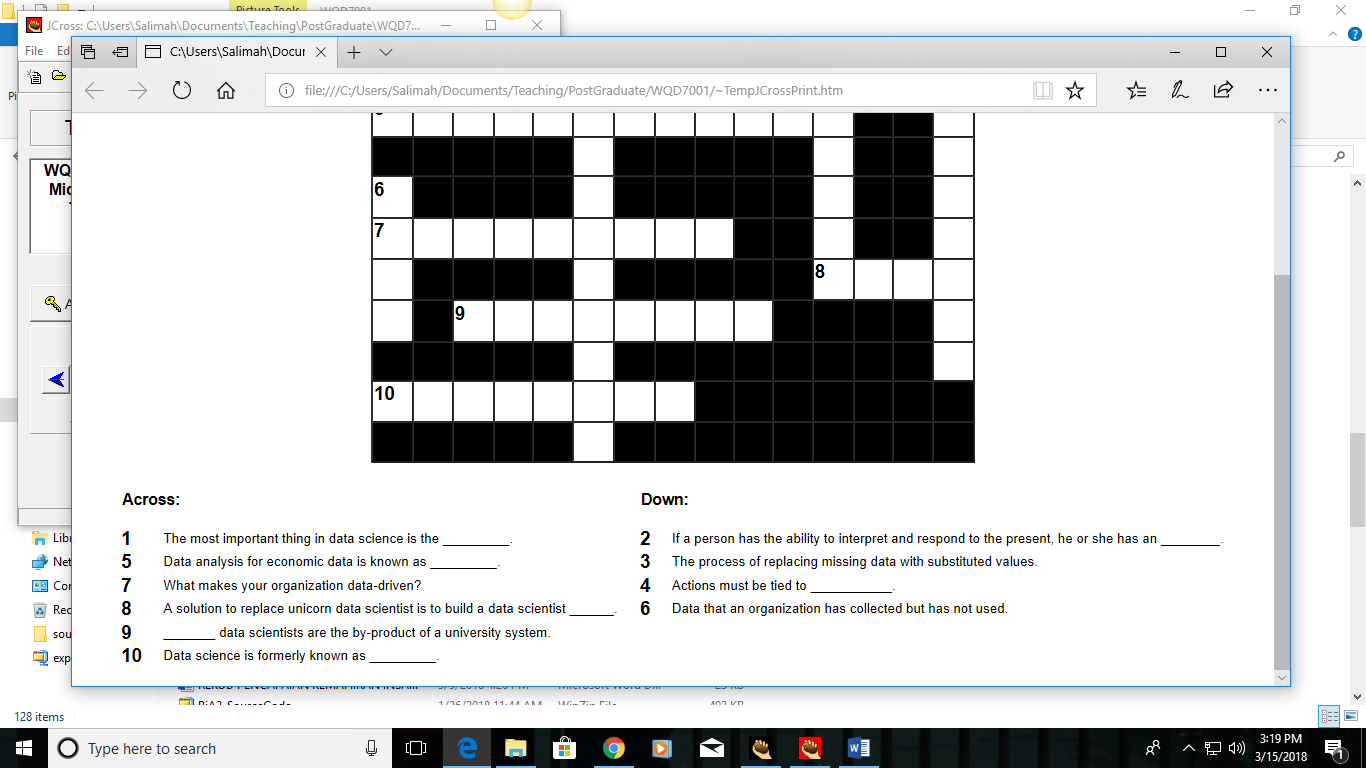
The number one way to build trust with a hiring manager is to prove you can do the work that they need you to do. With data science, this comes down to building a portfolio of projects. The more "real-world" the projects are, the more the hiring manager will trust that you'll be an asset to the business.

1. Solve the cross-word puzzle.

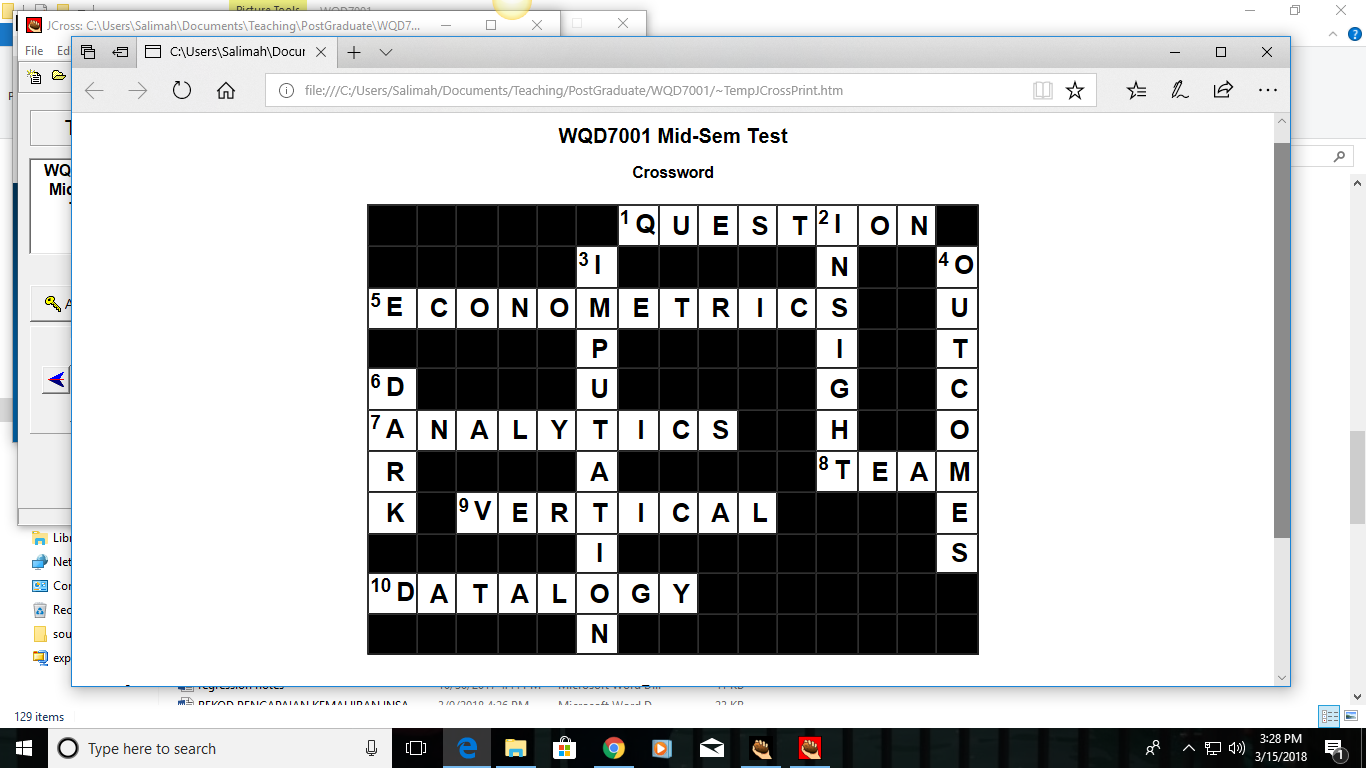
[10 marks]







**SOLUTION:**



**END**