DATA SCIENCE (CDA)
CLASS ASSESSMENT 1 (UNITS 1 AND 2, MODEL A)

1. From the five "the value of data" types seen in U1, which one corresponds to this example?
An insurance company predicts customer risk according to its historical incidents.

a) That data is valuable for me (out → in)
b) My data is valuable for others (in → out)
c) That data is valuable for others (out → out)
d) My data is valuable for me (in → in).

2. What is CRISP-DM?

a) A methodology for the process of data mining and knowledge discovery from data.
b) An extraction-transformation-load (ETL) tool for data silos.
c) A database specialised for Big Data .
d) A data analytics tool.

3. If we have an attribute with five possible values of increasing value: "very low", "low", "average", "high" and "very high"?

a) A numerisation "1 to n" would capture the natural gradation seen in the attribute values.
b) A numerisation "1 to 1" would capture the natural gradation seen in the attribute values.
c) A discretisation "1 to n" would capture the natural gradation seen in the attribute values.
d) A discretisation "1 to 1" would capture the natural gradation seen in the attribute values.

4. Which of the following claims is TRUE ?

a) Outlier is the name we use for referring each one of the rows in a dataset.
b) Data is unbiased if it is representative of the population of interest.
c) The selection of some features (attributes) in a dataset is called sampling.
d) The depiction of information using spatial or graphical representations is called Information Transformation.

5. Which of the following actions is NOT suitable to handle missing values of categorical variables?

a) filter/eliminate the column.
b) replace the value by the mode.
c) exchange rows and columns.
d) filter/eliminate the row.

6. What is a multidimensional datawarehouse?

a) A plot with more than three dimensions that is represented using parallel coordinates.

b) A NoSQL database, where queries are performed using a multidimensional XML query language.

c) An unstructured database where data is located in different data silos.

d) A structured database where information is represented by central facts and attributes are arranged into different dimensions.

7. What is data curation?

a) The recovery of data lost in a datawarehouse or NOSQL database.

b) A term to group those technologies in big data dealing with data integration, preparation, quality, privacy, security, metadata and general manipulation of the data.

c) The transformation of the data such that individual identities are not traceable.

d) The set of technologies that support the storage of data, the data infrastructure.

8. Which of the following visual attributes (retinal variables) is suitable to encode ordinal data (for instance, "low", "medium" and "high")?

a) Texture.

b) Size.

c) Shape.

d) Colour (Hue).

9. In a classroom of 30 students, in which 20 students support Valencia CF and 10 support Levante UD, if we apply simple random sampling to get a sample of six students, what will we get?

a) Three Valencia supporters and three Levante Supporters.

b) Four Valencia supporters and two Levante Supporters.

c) Two Valencia supporters and four Levante Supporters.

d) We can't know the proportions exactly.

10. Which of the following situations may imply a problem for data privacy or serious ethical issues?

a) A user voluntarily discl… …n Tinder.

b) A customer gets a dis… …meras are installed in their house.

c) A bank generates a list… …ikely to buy based on characteristics of the cu…

d) A mobile company is s… …people location.

QUESTION CANCELLED BECAUSE IT DOESN'T SPECIFY THE CAMERAS ARE INSIDE THE HOUSE. GRADE IS CALCULATED OVER 9 QUESTIONS.

ASSESSMENT
Answer Sheet – MODEL A

Surname:	**Name:**
Group in English: ☒	

In the following table, circle the correct answer for each question.

Question	**Answer**			
1	a	b	c	**d**
2	**a**	b	c	d
3	a	**b**	c	d
4	a	**b**	c	d
5	a	b	**c**	d
6	a	b	c	**d**
7	a	**b**	c	d
8	a	**b**	c	d
9	a	b	c	**d**
10	a	b	c	d

The result will be calculated by the statistical correction formula:

$$(\text{Right} - \text{Wrong}/3) \times 1$$

which discounts the probability of getting a right answer by chance on a question with four possibilities.

The mark is between 0 and 10.

Remember that this assessment is just 10% of the final qualification for the course.