Calculators may be used in this examination provided they are not capable of being used to store alphabetical information other than hexadecimal numbers

UNIVERSITY OF BIRMINGHAM

School of Computer Science

LM Storing and Managing Data

Main January Examinations 2023

Time allowed: 2 hours

[Answer all questions]

Turn Over

Note

Answer ALL questions. Each question will be marked out of 20. The paper will be marked out of 80, which will be rescaled to a mark out of 100.

Question 1

(a) A small company keeps a database about the restaurants in town to provide advice about food. The database keeps track of all restaurants via a unique code, and for each restaurant it records information about the name, address and telephone number; in addition to this, the database keeps track of all restaurant owners via their national insurance number (NINo), their name and surname. More importantly, the database records all information about the menu that each restaurant offers, including the name of the dish, its price, the cuisine (e.g., Spanish, Chinese, Italian, etc.), the category (e.g., gluten free, vegan, meat, halal, etc.), and the ingredients required. Given the following schema corresponding to the database described above, where primary and foreign keys are denoted via PK and FK, respectively, normalise it to the third normal form (3NF):

Restaurant (RCode PK, RName, Address, Telephone, OwnerNINo, OwnerName, OwnerSurname)
Menu (RCode PK FK, DishName PK, Price, Cuisine, DishCategory, Ingredients)

[7 marks]

Given the following schema corresponding to a database about museums and paintings, where primary and foreign keys are denoted via PK and FK, respectively:

Museum (MCode PK, MName, Street, City, Director FK)
Room (MCode PK FK, RoomNo PK, FloorArea, NumberOfDoors, Guardian FK)
Painting (PCode PK, PName, Painter, DateOfPurchase, Museum FK, RoomNo FK)
Employee (NINo PK, EName, ESurname, StartDate)

Produce the SQL queries to answer the following:

(b) Retrieve the total floor area for each museum and the corresponding museum code.

[5 marks]

(c) Retrieve the painter with the largest number of paintings held across all the museums in the city of Miami.

[8 marks]

Question 2

Dr. Anthony is a Professor in a UK University. During the last one year, he has been gathering the health data of his colleagues suffering from hypertension to conduct research. The data he has collected contains their name, nationality, age, height, email address, heart rate (ECG), body weight and blood pressure. He has only collected the data and has not used the data for research. At the time of this data collection, the GDPR still applies in the UK. Does the GPDR apply to Dr. Anthony when he collects this data? Explain your answer. Merely a yes or no answer without an explanation will not result in any points.

Answer the question using the IRAC format.

[20 marks]

Hint: IRAC stands for Issue, Rule, Application, Conclusion.

Question 3

An analyst is preparing a Document Store to keep information about restaurant food quality as assessed by customers. The analyst has found online reviews from 3 customers for 2 different restaurants.

- The address of the first restaurant, named "Banti Birmingham" is: Street: 26 Moat Ln, Postcode: B5 5BD. It is in the City Centre, and it serves African cuisine. It has been graded by 2 different customers; the first submitted a score of 2 stars on the following date: 22 December 2022 at 23:15; the second submitted a score of 3 stars on the following date: 10/01/2023 at 20:00.
- The second restaurant, named "Toyosu Heaven" is in Harborne at 103 High Str, B17 9NR, and it serves Japanese cuisine. It has been graded by a customer who submitted a score of 4 stars on the following date: 30 December 2022 at 19:30.

(a) Prepare 2 JSON documents accumulating the restaurant information provided above, i.e., their address, location (i.e., borough), cuisine, name, and grades from customers. **[10 marks]**

(b) Using MongoDB syntax, make a query to your restaurant collection to retrieve the restaurant name(s), as well as type of the cuisine they serve, of those restaurants that received at least one score of 3 stars. **[10 marks]**

You can store dates using ISODate(), e.g.

```
1 ISODate("2021-01-03T23:30:15.123Z"), or {"$date": "2021-01-03T23:30:15.123Z"}.
```

Question 4

A study in obesity in a small town measured the values in Table 1 from the study participants. The scatter plot of height vs weight data is shown in Figure 1. The regression model as fitted with least squares is:

$$Weight[kg] = \beta_1 \cdot Height[cm] + \beta_0 + \varepsilon \tag{1}$$

with $\beta_1 = 0.33163$, $\beta_0 = 19.814$, the parameters of the model and ε the innovation or dispersion term.

Subject ID	Height [cm]	Weight [kg]	Subject ID	Height [cm]	Weight [kg]
1	164	71	11	171	86
2	173	72	12	158	68
3	168	69	13	210	93
4	179	84	14	164	78
5	189	63	15	162	78
6	161	72	16	189	85
7	170	76	17	165	76
8	174	86	18	176	74
9	179	83	19	168	71
10	182	84	20	177	81

Table 1: Obesity study

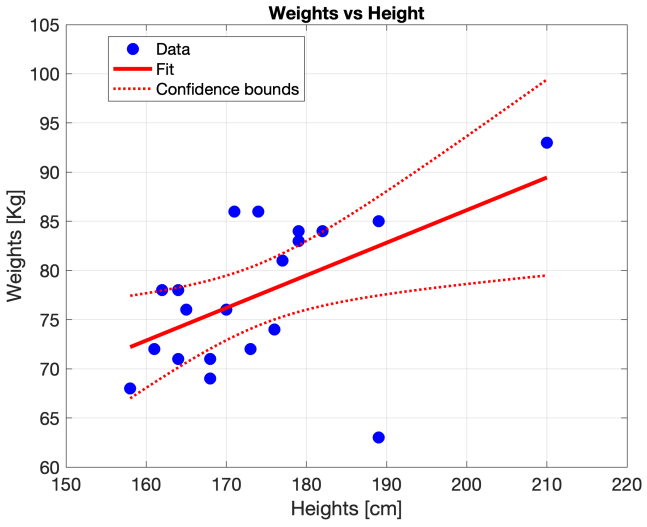

Figure 1: Obesity study scatterplot.

(a) Given the regression model, calculate the residuals for subjects 1, 2 and 3. Show the intermediate calculations. **[4 marks]**

(b) Indicate whether there are any outliers. If there are, then also indicate the Subject ID of the outlier or outliers and the reason why you think they are outliers. **[4 marks]**

(c) Calculate the body mass index statistic for subjects 1, 2 and 3. The body mass index (BMI) can be calculated according to Equation 2

$$BMI = \frac{weight\ [kg]}{height^2\ [m]} \tag{2}$$

Round to the first decimal for convenience. **[6 marks]**

(d) The boxplot of the BMI is shown in Figure 2. Indicate whether the outliers in the original features space (height and weight), match those in the derived features (BMI). **[6 marks]**

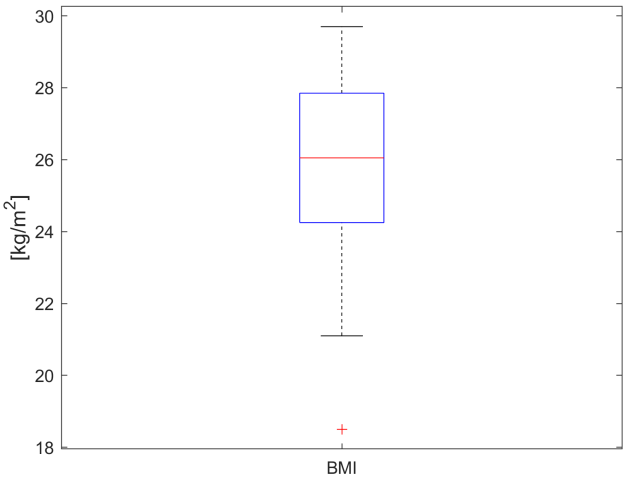

Figure 2: Obesity study boxplot.

End of Paper

Do not complete the attendance slip, fill in the front of the answer book or turn over the question paper until you are told to do so

Important Reminders

- Coats/outwear should be placed in the designated area.
- Unauthorised materials (e.g. notes or Tippex) must be placed in the designated area.
- Check that you do not have any unauthorised materials with you (e.g. in your pockets, pencil case).
- Mobile phones and smart watches **must** be switched off and placed in the designated area or under your desk. They must not be left on your person or in your pockets.
- You are not permitted to use a mobile phone as a clock. If you have difficulty seeing a clock, please alert an Invigilator.
- You are not permitted to have writing on your hand, arm or other body part.
- Check that you do not have writing on your hand, arm or other body part – if you do, you must inform an Invigilator immediately
- Alert an Invigilator immediately if you find any unauthorised item upon you during the examination.

Any students found with non-permitted items upon their person during the examination, or who fail to comply with Examination rules may be subject to Student Conduct procedures.