**MGS 4020 Business Intelligence**

Midterm Examination, Spring 2019

Dr. Satish Nargundkar

Assume that the following is a complete list of fields you want to store in your sales database.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Customer\_ID | Cust\_Fname | Cust\_Lname | Cust\_address | Cust\_City |
| Cust\_State | Cust\_ZIP | Invoice\_Number | Item\_Number | Quantity |
| Price/unit | Purchase\_Date | Store\_ID | Store\_Location | Supplier\_ID |
| Supplier\_Name | Supplier\_Address | Supplier\_City | Supplier\_State | Supplier\_Zip |

**Assumptions:**

1. The same invoice numbers may be repeated in a different store. Within a store, invoice numbers are unique for each transaction.
2. A customer transaction may involve multiple items listed on the same invoice.
3. A given invoice number can only have one customer associated with it, for a given store.
4. A given item number can have only one supplier associated with it, though one supplier can supply multiple items.
5. Prices are the same across all stores for a given item. An item is only listed once on a given invoice.
6. If all of the above fields were put in a flat file (1st Normal form), what would be the primary key? Explain briefly why you think so. **[10 points]**
7. Convert into the 2nd Normal Form. Show how the fields would be put into different tables based on the primary key components. **[10 points]**
8. Convert into 3rd Normal Form. Separate out any transitive dependencies. Show only the tables that would be split from the 2nd Normal Form. **[10 points]**
9. Consider the following two tables of data.

**Table 1**

|  |  |
| --- | --- |
| **ID\_Number** | **First\_Name** |
| 111 | John |
| 222 | Sally |
| 333 | James |
| 444 | Janet |

**Table 2**

|  |  |
| --- | --- |
| **ID\_Number** | **Alma Mater** |
| 111 | Clemson |
| 222 | UGA |
| 333 | Georgia State |
| 555 | UNC |

What would the results of an **inner join** and a **right outer join** of these two tables look like? Use as many rows of the tables below as needed to write out the answers. Assume Table 1 is on the left. **[10 points]**

The **inner join** of Tables 1 and 2 would create the following output:

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

The **right outer join** of Tables 1 and 2 would create the following output:

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

1. Assume there is an Employee Table with the following fields in it:

Emp\_ID, Emp\_Firstname, Emp\_Lastname, Department, and Salary.

Assume further that there is data on 10 people for these fields, and no two salary values are the same. What fields and records (columns and rows) will each of the following SQL statements return as the output? **[10 points]**

1. **Select** \* **from** employee

**Where** salary = (**select** Max(salary) **from** employee)

1. **Select Max**(salary)

**From** employee

**Where** salary **Not In** (**select** Max(salary) **from** employee)

1. Why might one create **Forms** in a database? Give at least two reasons forms are used. **[4 points]**
2. Compare and contrast Operational Data and a Data Warehouse, on the following bases for comparison. Write a phrase or a sentence under the two columns, for each of the bases.

**[6 points]**

|  |  |  |
| --- | --- | --- |
| **Basis for comparison** | **Operational data** | **Warehouse data** |
| Key Purpose |  |  |
| Size |  |  |
| Response Times |  |  |

1. How do Dashboards help an organization? **[4 points]**
2. Compute the Mean and Standard Deviation for the following sample data that show the scores out of 10 points for 5 students on a math question. Show the computations in the table.

**[5 points]**

|  |  |  |
| --- | --- | --- |
| **Student Scores** |  |  |
| 8 |  |  |
| 6 |  |  |
| 7 |  |  |
| 6 |  |  |
| 8 |  |  |
|  |  |  |
|  |  |  |

1. Draw a Box and Whisker plot for the following raw data **[10 points]**

[5, 6, 6, 7, 7, 8, 9, 12, 13, 13, 15, 16, 17, 18, 18, 20, 20, 21, 25, 25, 26, 27, 28, 29, 30]

1. In the Analytics Ecosystem, who are the Technology Accelerators? Briefly explain what that means and give a couple of examples. **[5 points]**
2. What is the difference between Descriptive, Predictive, and Prescriptive Analytics? Give examples of techniques used in each. **[6 points]**

|  |  |  |
| --- | --- | --- |
| Type | What it means | Example Techniques |
| Descriptive |  |  |
| Predictive |  |  |
| Prescriptive |  |  |

1. Consider the spreadsheet below showing the GPAs of some students.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | A | B | C | D |
| 1 | **First** | **Last** | **GPA** | **Scholarship** |
| 2 | Bruce | Willis | 2.80 |  |
| 3 | Clint | Eastwood | 3.20 |  |
| 4 | Taraji | Henson | 3.50 |  |
| 5 | Jackie | Chan | 3.80 |  |
| 6 | Jennifer | Lawrence | 2.90 |  |
| 7 | Jessica | Chastain | 4.00 |  |

The college offers scholarships of $5,000 to all students with a GPA of 3.5 or above. What formula would you type in cell D2 so that it would show $5,000 for the appropriate students, and show $0 for the rest, if the formula were copied down the column? **[5 points]**

Formula in Cell D2: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

If there was now a scholarship of $2,000 for students with a GPA of 3.0 or above, as well as the $5,000 scholarship for those with a GPA of 3.5 or above, how would you now handle the new situation? What function would you type cell D2? What other changes would you make in the spreadsheet in keeping with good spreadsheet design? **[5 points]**

Function in Cell D2: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Other changes needed to spreadsheet: