

## Unit 2: Assignment

This is a homework assignment for MSDS7330, File Organization and Database Management. Be sure to place your name and due date on your answer sheet and place your last name and the homework number at the beginning of the file name. For example, the filename for the homework answer sheet for homework 2 for Raghuram Srinivas should be *SrinivasMSDS7330HomeWorkAnswerSheet.docx*.

For this assignment, create a Word document. State the letter of your chosen answer for each question below and explain why the answer is correct. Note that the explanation involves also explaining why the other answer choices are not correct. Your answer sheet should be submitted for homework assignment 2 on the 2DS system.

- 1) The DBMS acts as an interface between what two components of an enterprise-class database system?

a) Database application and the database  
b) Data and the database  
c) The user and the database application  
d) Database application and SQL

In an **Enterprise-class database system**, business users interact directly with **database applications** that interact with the **DBMS**, which directly accesses the **database data**.

- 2) The following are components of a database except which of the following?

a) user data  
b) metadata  
c) reports  
d) indexes

**All the others are types of data and a report can't be put into a single table with rows of the data and columns that define the data elements.**

- 3) Table is synonymous with the term

a) record  
b) relation  
c) column  
d) row

The relational model uses a collection of **tables** to represent both data and the **relationships** among those data.

- 4) Row is synonymous with the term

a) column  
b) relation  
c) field

d) **record**

A **row** is also called a **record** or **tuple** represents a single, implicitly structured data item in a table.

- 5) Attribute is synonymous with the term

a) record  
b) relation  
c) **column**  
d) row

The **columns** of the table correspond to the **attributes** of the record type.

- 6) A database schema is

a) A plan created by the database  
b) **The logical design of the database**  
c) A particular instance of a database  
d) The snapshot of the data in the database at a given time

A **database schema** represents the **logical** configuration of all or part of a **relational database**.

- 7) The primary key is selected from the

a) composite keys  
b) determinants  
c) **candidate keys**  
d) foreign keys

A **primary key** is a **candidate key** that is chosen as a mean to identify tuples within a relation. Only one **candidate key** can be a **primary key**.

- 8) Which of the following is a group of one or more attributes that uniquely identifies a row?

a) **Key**  
b) Determinant

- c) Tuple
- d) Relation

A **key** (whether primary, candidate, or super) is a property of the entire relation, rather than of the individual tuple.

- b) determinants
- c) candidate keys
- d) foreign keys

A **foreign key** is a field that is linked to another table's primary key field in a **relationship between two tables**.

- 9) In the relational model, relationships between relations or tables are created by using:
- a) composite keys

**Question 1 : Create a dataframe with the following columns : YearBuilt , SalePrice , LotArea , BedroomAbvGr and FullBaths**

```
In [22]: new_df = df.loc[:, ['YearBuilt', 'SalePrice', 'LotArea', 'BedroomAbvGr', 'FullBaths']]
new_df.head(5)
```

```
Out[22]:
```

|   | YearBuilt | SalePrice | LotArea | BedroomAbvGr | FullBaths |
|---|-----------|-----------|---------|--------------|-----------|
| 0 | 2003      | 208500    | 8450    | 3            | NaN       |
| 1 | 1976      | 181500    | 9600    | 3            | NaN       |
| 2 | 2001      | 223500    | 11250   | 3            | NaN       |
| 3 | 1915      | 140000    | 9550    | 3            | NaN       |
| 4 | 2000      | 250000    | 14260   | 4            | NaN       |

**Question 2 : What is the avg price of single family homes**

```
In [23]: print("The avg price of single family homes is: %d"%df[df.BldgType=="1Fam"].mean()['SalePrice'])
The avg price of single family homes is: 185763
```

**Question 3 : What is the mean home price of the single family homes built after 1950**

```
In [24]: print("The Mean home price of the single family homes built after 1950 is: %d"%df[df.YearBuilt>=1950].mean()['SalePrice'])
The Mean home price of the single family homes built after 1950 is: 194468
```

**Question 4 : What is the median home price per number of bedrooms in the house?**

```
In [25]: df.groupby(df.BedroomAbvGr).median()["SalePrice"]
```

```
Out[25]: BedroomAbvGr
0      202500.0
1      145250.0
2      137250.0
3      169945.0
4      193500.0
5      161500.0
6      141000.0
8      200000.0
Name: SalePrice, dtype: float64
```

**Question 5 : What is the most expensive home in each Neighborhood**

```
In [26]: df.groupby(df.Neighborhood).max()['SalePrice']
```

```
Out[26]: Neighborhood
Blmngtn    264561
Blueste    151000
BrDale     125000
BrkSide     223500
ClearCr     328000
CollgCr     424870
Crawfor     392500
Edwards     320000
Gilbert     377500
IDOTRR      169500
MeadowV     151400
Mitchel     271000
NAmes       345000
NPKVill     155000
NWAmes      299800
NoRidge     755000
NrIdgHt     611657
OldTown     475000
SWISU       200000
Sawyer       190000
SawyerW     320000
Somerst     423000
StoneBr     556581
Timber       378500
Veenker     385000
Name: SalePrice, dtype: int64
```

**Question 6 :Sort homes by the year built**

```
In [27]: df['YearBuilt'].sort_values()
```

```
Out[27]: 1349    1872
1137    1875
747     1880
304     1880
1132    1880
630     1880
991     1882
106     1885
1416    1885
945     1890
716     1890
185     1892
1292    1892
583     1893
809     1898
711     1900
703     1900
242     1900
390     1900
1186    1900
1062    1900
520     1900
488     1900
676     1900
921     1900
841     1904
1393    1905
653     1906
184     1908
210     1908
```