MGS 613 Project Report - Group 3
Real Estate Database Management System

Noodhan Gowri Shankar, Sowfia Syed Abuthaheer, Sushma Preethi Rajamani Veni, Nishanth Holla Raghavendra, Saurabh Narendra Singh

INTRODUCTION:
The housing market is currently facing a crisis, with home prices nearly doubling since 1986, making it increasingly difficult for many to afford a home. However, despite these skyrocketing costs, the demand for housing remains high, and the number of homes sold hasn't declined. Realtor agencies and agents have become central players in this booming market, facilitating the buying and selling process. They handle everything from marketing properties to negotiating deals and managing paperwork, all while earning significant commissions—often a percentage of the sale price. These commissions can result in substantial profits, especially in high-priced markets. By examining how these middlemen operate, we can better understand their impact on the housing market, their contribution to rising costs, and their role in the broader dynamics of homeownership.

DATA:
For this project, we collected and combined data from Real Trends and Kaggle to gain insights into the U.S. real estate market. The data covers various aspects of the industry, including top-performing agents, leading realtor agencies, and property listings across the country. Specifically, our sources are:

- Top 100 Real Estate Agents: A ranking of the highest-performing agents nationwide.
- Top Realtor Agencies: A list of leading realtor agencies across the country.
- Property Listings: Detailed data on properties listed for sale in the past year across various states.

We've consolidated this data into tables to streamline analysis and support future queries. Below is a description of the combined dataset.

Data Dictionary:
Table 1:

Field	**Type**	**Description**
PropertyID	Numeric	Unique ID for House
Price	Numeric	Price of House
Bedrooms	Numeric	Number of Bedrooms
Bathrooms	Numeric	Number of Bathrooms
Sqft	Numeric	Size of house in sqft
Floors	Numeric	Number of floors
YearBuilt	Numeric	Year it was built
YearRenovated	Numeric	Year it was renovated
Street	AlphaNumeric	Locality of the House
City	Character	City of the house
ZipCode	AlphaNumeric	ZipCode of the house
AgentID	Numeric	UniqueID of Agent

Table 2:

Field	**Type**	**Description**
AgentID	Numeric	Unique ID of Agent
FirstName	Character	First Name of Agent
LastName	Character	Last Name of Agent
AgentCity	Character	City of Agent
AgentState	Character	State of Agent
Volume	Numeric	Volume of Sales of Agent
AgencyID	Numeric	UniqueID of Agency

Table 3:

Field	**Type**	**Description**
AgencyID	Numeric	UniqueID of Agency
AgencyName	Character	Name of Agency
AgencyCity	Character	City of Agency
AgencyState	Character	State of Agency
Employees	Numeric	Number of Employees
Revenue	Numeric	Revenue of Agency

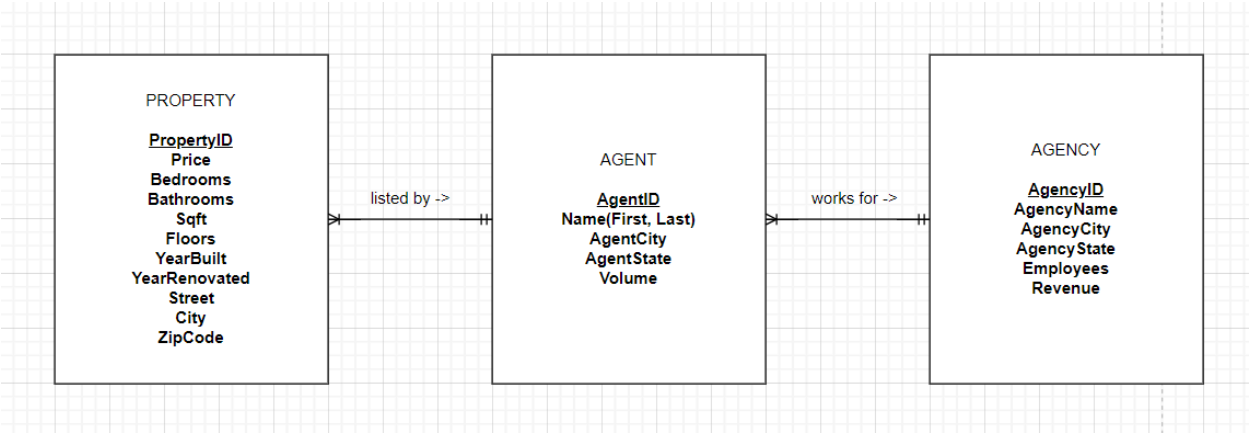

Figure 1 Entity Relationship Diagram (ERD)

Based on our ERD (Entity-Relationship Diagram), we normalized the data and developed a relational schema comprising three distinct tables: PROPERTY, AGENT, and AGENCY. This structure allows us to efficiently manage and query the data.

- The **PROPERTY** table contains detailed information about each property listed for sale and includes an AgentID as a foreign key, linking each property to the agent responsible for it.
- The **AGENT** table holds data about individual real estate agents, including their AgencyID, which serves as a foreign key to connect each agent with their respective agency.

- The **AGENCY** table acts as the parent table in this schema, storing information about the realtor agencies. Since it is the parent table, it does not have a foreign key.

This relational schema, as depicted in Figure 2, ensures that our data is organized and interconnected, facilitating complex queries and analyses while maintaining data integrity.

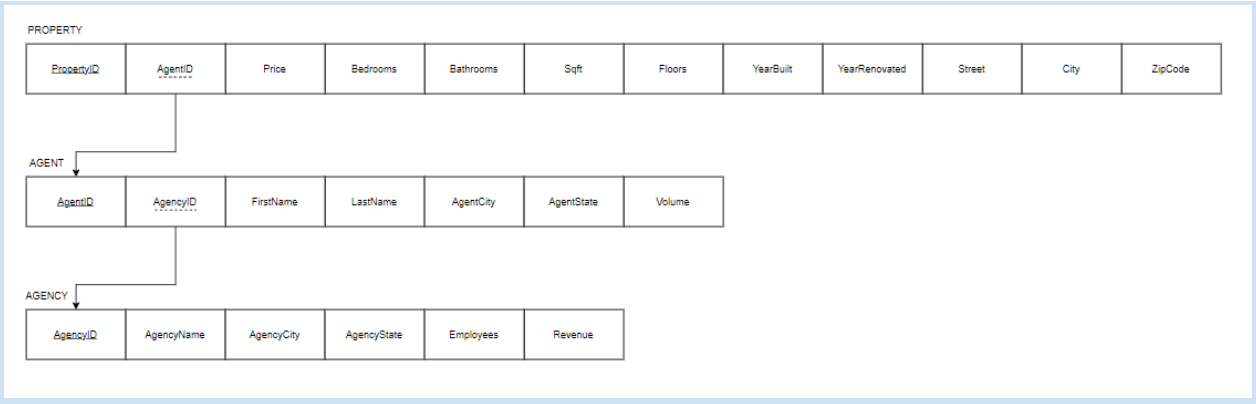

Figure 2 Relational Schema

Database Implementation

To implement the database in APEX, we wrote CREATE TABLE commands for each table in the relational schema.

Agency:

As the parent table, Agency was created and populated first:

```
CREATE TABLE Agency (
   AgencyID INT PRIMARY KEY,
   AgencyName VARCHAR(255),
   AgencyCity VARCHAR(100),
   AgencyState VARCHAR(50),
   Employees INT,
   Revenue INT);

INSERT INTO Agency (AgencyID, AgencyName, AgencyCity, AgencyState, Employees, Revenue)
VALUES
(1001, 'HomesUSA.com, Inc.', 'Dallas', 'Texas', 457, 36846957845),
(1002, 'Compass', 'New York', 'New York', 684, 184452364253),
(1003, 'ListWithfreedom.com', 'Richmond', 'Virginia', 215, 8451267594);
```

Agent:

```
CREATE TABLE Agent (
   AgentID INT PRIMARY KEY,
```

```
    FirstName VARCHAR(100),
    LastName VARCHAR(100),
    AgentCity VARCHAR(100),
    AgentState VARCHAR(50),
    Volume INT,
    AgencyID INT,
    FOREIGN KEY (AgencyID) REFERENCES Agency(AgencyID)
);

INSERT INTO Agent (AgentID, FirstName, LastName, AgentCity, AgentState, Volume, AgencyID)
VALUES
(101010, 'Ben', 'Caballero', 'Dallas', 'TX', 3638950694, 1001),
(101011, 'Alexa', 'Lambert', 'New York', 'NY', 923676714, 1002),
(101012, 'Ralph', 'M Harvey III', 'Boynton Beach', 'FL', 826461740, 1003),
(101013, 'Christian', 'Angle', 'Palm Beach', 'FL', 804599700, 1004);
```

PROPERTY:

```
CREATE TABLE Property (
    PropertyID INT PRIMARY KEY,
    Price INT,
    Bedrooms INT,
    Bathrooms INT,
    Sqft INT,
    Floors INT,
    YearBuilt INT,
    YearRenovated INT,
    Street VARCHAR(255),
    City VARCHAR(100),
    ZipCode VARCHAR(20),
    AgentID INT,
    FOREIGN KEY (AgentID) REFERENCES Agent(AgentID)
);

INSERT INTO Property (PropertyID, Price, Bedrooms, Bathrooms, Sqft, Floors, YearBuilt,
YearRenovated, Street, City, ZipCode, AgentID)
VALUES
(1, 376000, 3, 2, 2724, 3, 2008, 0, '9245-9249 Fremont Ave N', 'Seattle', 'WA 98103', 101039);
```

ANALYSIS:

This analysis aims to provide insights into the performance of realtors and agencies, market leadership, and geographical trends in the real estate industry.

Question 1: Top 3 Realtors by Total Sales Volume and Market Share

Which three realtors have sold the most volume(total sales in dollars) of houses in the database? What percentage do they make up in the entire volume of all realtors?

To achieve this we composed a simple query that calculates the total sales volume for each realtor. The query then identifies the top 3 realtors with the highest sales volume and calculates the percentage of the entire volume these three realtors represent out of the total sales volume generated by all realtors in the database.

```
WITH AgentSales AS (
   SELECT
      A.AgentID,
      A.FirstName,
      A.LastName,
      SUM(P.Price) AS TotalSales
   FROM
      Agent A
   JOIN
      Property P ON A.AgentID = P.AgentID
   GROUP BY
      A.AgentID, A.FirstName, A.LastName
),
TotalVolume AS (
   SELECT
      SUM(TotalSales) AS OverallVolume
   FROM
      AgentSales
)
SELECT
   ASales.AgentID,
   ASales.FirstName,
   ASales.LastName,
   ASales.TotalSales,
   ROUND((ASales.TotalSales / TV.OverallVolume) * 100,4) AS PercentageOfTotal
FROM
   AgentSales ASales,
   TotalVolume TV
ORDER BY
   ASales.TotalSales DESC
FETCH FIRST 3 ROWS ONLY;
```

The results of this query are shown below. The top three realtors are Susan Breitenbach, Leslie McElwreath and Pollena Forsman. The table also displays the percentage they make up in the entire volume of all realtors.

AGENTID	FIRSTNAME	LASTNAME	TOTALSALES	PERCENTAGEOFTOTAL
101069	Susan	Breitenbach	50524150	2.2066
101031	Leslie	McElwreath	33856934	1.4787
101094	Pollena	Forsman	32766281	1.431

Figure 3 Total Sales and Percentage by Realtors

Question 2: Top Agencies' Market Share in Real Estate Revenue

Which two or three agencies(duopoly) have been the major market leader in the real estate business? What percentage of total share do they own in the entire revenue?

To determine which agencies have been the major market leaders in the real estate business, we composed a query that calculates the total revenue generated by each agency. The query identifies the top three agencies with the highest revenue and calculates their combined percentage of the total market share.

```
WITH AgencyRevenues AS (
   SELECT
      A.AgencyID,
      A.AgencyName,
      SUM(A.Revenue) AS TotalRevenue
   FROM
      AgencY A
   GROUP BY
      A.AgencyID, A.AgencyName
),
TotalMarketRevenue AS (
   SELECT
      SUM(TotalRevenue) AS OverallRevenue
   FROM
      AgencyRevenues
)
SELECT
   AR.AgencyID,
   AR.AgencyName,
   AR.TotalRevenue,
   ROUND((AR.TotalRevenue / TMR.OverallRevenue) * 100,4) AS MarketSharePercentage
FROM
   AgencyRevenues AR,
```

```
    TotalMarketRevenue TMR
ORDER BY
    AR.TotalRevenue DESC
FETCH FIRST 3 ROWS ONLY;
```

AGENCYID	AGENCYNAME	TOTALREVENUE	MARKETSHAREPERCENTAGE
1002	Compass	184452364253	22.9524
1048	eXp Realty LLC	143578000000	17.8662
1036	Green Banker	75864759845	9.4403

Figure 4 Total Revenue and Market Share Percentage of Agencies

The top three agencies that have been the major market leader in the real estate business are Compass, eXp Realty LLC and Green Banker.

Question 3: Agency Producing Most Successful Realtors by Average Sales

Which agency has the habit of creating the most successful realtors?
This is calculated by dividing the total volume of an agency by the total number of realtors belonging to it.

```
WITH AgencyPerformance AS (
    SELECT
        A.AgencyID,
        A.AgencyName,
        COUNT(AG.AgentID) AS NumRealtors,
        SUM(AG.Volume) AS TotalAgencyVolume,
        SUM(AG.Volume) / COUNT(AG.AgentID) AS AvgVolumePerRealtor
    FROM
        Agency A
    JOIN
        Agent AG ON A.AgencyID = AG.AgencyID
    GROUP BY
        A.AgencyID, A.AgencyName
)
SELECT
    AgencyID,
    AgencyName,
    AvgVolumePerRealtor
FROM
    AgencyPerformance
ORDER BY
    AvgVolumePerRealtor DESC
FETCH FIRST 1 ROW ONLY;
```

AGENCYID	AGENCYNAME	AVGVOLUMEPERREALTOR
1001	HomesUSA.com, Inc.	3638950694

Figure 5 Agency with most successful realtors

Question 4: Top States for Agent Returns vs. Average Property Prices

Which states provide the most returns for the agents? How high are they when compared to the average price of properties across the board?

To determine which states provide the most returns for agents, we composed a query that calculates the average property price in each state and compares it to the overall average property price. The query identifies the states with the highest returns by measuring how much their average property prices exceed the national average, providing insight into the regions where agents can achieve the best returns.

```
WITH StateReturns AS (
    SELECT
        AG.AgentState,
        SUM(P.Price) AS TotalSalesVolume,
        COUNT(P.PropertyID) AS NumProperties,
        Round(AVG(P.Price),4) AS AvgPriceInState
    FROM
        Agent AG
    JOIN
        Property P ON AG.AgentID = P.AgentID
    GROUP BY
        AG.AgentState
),
OverallAverage AS (
    SELECT
        Round(AVG(Price),4) AS AvgPriceOverall
    FROM
        Property
)
SELECT
    SR.AgentState,
    SR.TotalSalesVolume,
    SR.AvgPriceInState,
    OA.AvgPriceOverall,
    Round((SR.AvgPriceInState - OA.AvgPriceOverall) / OA.AvgPriceOverall * 100,4) AS
PriceComparisonPercentage
FROM
    StateReturns SR,
    OverallAverage OA
ORDER BY
    SR.TotalSalesVolume DESC;
```

AGENTSTATE	TOTALSALESVOLUME	AVGPRICEINSTATE	AVGPRICEOVERALL	PRICECOMPARISONPERCENTAGE
CA	703965323	546132.9116	553062.8792	-1.253
FL	377393064	559930.362	553062.8792	1.2417
NY	373190132	626157.9396	553062.8792	13.2164
CO	138247999	497294.9604	553062.8792	-10.0835
TX	130881339	529883.9636	553062.8792	-4.191
MA	125226471	491084.2	553062.8792	-11.2064
AZ	85272748	519955.7805	553062.8792	-5.9861
IL	77342389	608995.189	553062.8792	10.1132
CT	52163034	511402.2941	553062.8792	-7.5327
HI	42560064	575136	553062.8792	3.9911

Figure 5 Agent State Sales and Pricing Analysis

California leads the nation in total sales volume, providing real estate agents with significant opportunities due to the high number of transactions. This bustling market allows agents to earn substantial returns from frequent deals. In contrast, New York boasts the highest average property prices, exceeding the national average. This results in higher commissions for agents on each sale, making it a lucrative market despite fewer overall transactions. In essence, California offers volume-based earnings, while New York offers high-value commissions per transaction.