TRANSPORT AND TELECOMMUNICATION INSTITUTE

Modern Database Technologies

Project

Topic: US Walkability Database

Student's name: Gonzalo

Student's surname: Gamez

St. code: st83446

Lecturer: Pizika Nadežda

RIGA 2023

INTRODUCTION

The main goal of this course project is to develop an effective database solution for the business problem described in the Business case. The database will be implemented in a future project in a dynamic visual to show information to the user which will have api calls to retrieve and store data.

BUSINESS CASE

Business Case Description

A travel agency for hotels that are walkable and/or have public transit in the US for people who wish to see the United states without a car. A state has many cities, a city has many hotels, a hotel has one walk score and may have a transit score (https://www.walkscore.com/how-it-works/) . City may have an airport which may have a transit score.

➢ API requires latitude, longitude, and address for scores
➢ Scores can correspond to their description since a walk score of 25 and 30 will still mean car - dependent so it will round to nearest tens digit
➢ All hotels will not be listed, up to 100 can be listed from each city

Suggestions regarding business object descriptions:

State: state_id , name

City: city_id, state_id, name

Hotel: hotel_id, address, latitude, longitude

Airport: airport_id, address, latitude, longitude

Walk Score: walkscore_id, walk_score, score_description

Transit Score: transitcore_id, transit_score, score_description

Business Rules

Based on the analysis of the business case the following business rules have been defined.

- ➢ One state has many cities
- ➢ One city is in one state
- ➢ One city has up to 100 hotels
- ➢ One hotel is in one city
- ➢ One city may have many airports
- ➢ One airport is in one city
- ➢ One hotel has one walk score
- ➢ One hotel may have one transit score
- ➢ One airport may have one transit score
- ➢ Walk score may have hotels
- ➢ Transit score may have many hotels
- ➢ Transit score may have many airports

Possible request that could be performed against the databases:

- ➢ What is the average walk score of a city's hotels
- ➢ If there is a hotel with a very walkable score
- ➢ If a city has decent public transit from the airport and at a hotel
- ➢ Display states with low walk and public transit score from airports and hotels

Conceptual Data Model

List of Entities

State

City

Hotel

Airport
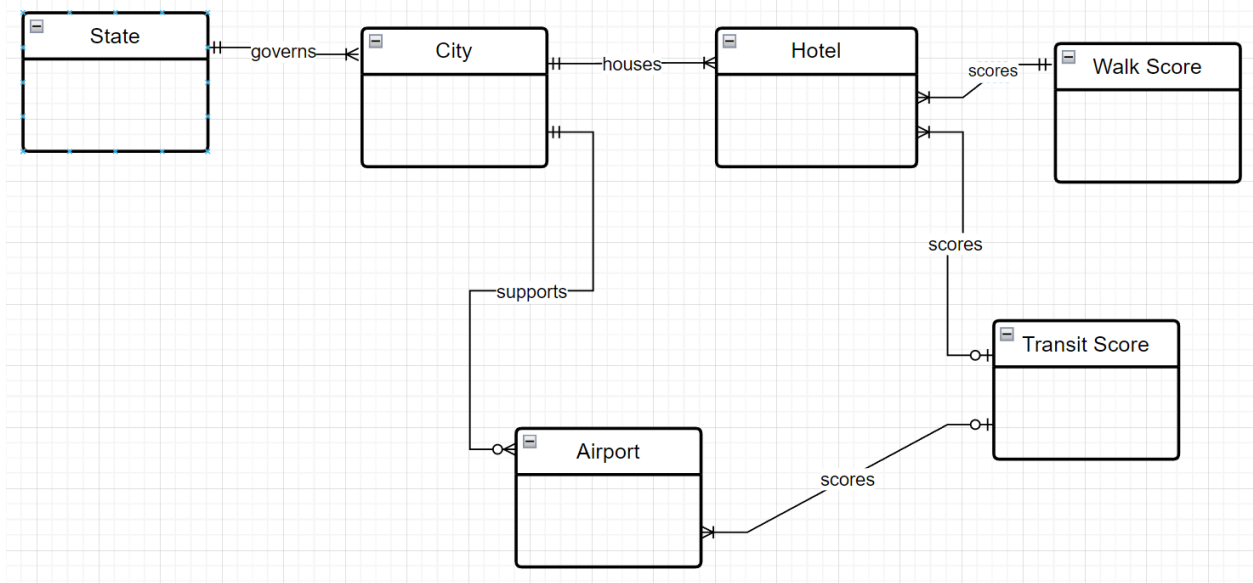
Walk Score

Transit Score

Relationships

- ○ City:State => M:1
- ○ Hotel:City => M:1
- ○ Airport:City => M:1
- ○ Walk Score: Hotel => 1:M
- ○ Transit Score: Hotel => 1:M
- ○ Transit Score: Airport => 1:M

ERD for the Conceptual Model

Attributes and Primary Keys

State: state_abbreviation_id , state_name

City: city_id, city_name, state_id, state_name

Hotel: hotel_id, hotel_name, address, city_id, walkscore_id, transitscore_id, latitude, longitude

Airport: airport_id, airport_name, address, city_id, transitscore_id, latitude, longitude

Walk Score: walkscore_id, walk_score, score_description

Transit Score: transitscore_id, transit_score, score_description

Attributes Domains. Constraints.

Entity	Attributes	Description	Data type, length	Nullable	Possible Values, Constraints, formats

State	state_id	Unique Two letter abbreviation of state	CHAR, 2	No	XX
	state_name	Name of state	VARCHAR, 20	No	

Entity	Attributes	Description	Data type, length	Nullable	Possible Values, Constraints, formats
City	city_id	Unique city id	CHAR,4	No	XXNN XX is state id
	city_name	Name of city	VARCHAR, 20	no	
	state_id	State the city resides in	CHAR,2	No	XX

Entity	Attributes	Description	Data type, length	Nullable	Possible Values, Constraints, formats
Hotel	hotel_id	Unique hotel id	CHAR,6	No	XXXXNN XXXX is city id Can't have more than 100 hotels
	hotel_name	Name of hotel	VARCHAR, 30	no	
	city_id	City the hotel resides in	CHAR,4	No	XXXX
	address	Address of hotel	VARCHAR, 50	no	
	transitscore_id	Id of transitscore	CHAR,2	yes	XX

	walkscore_id	Id of walkscore	CHAR,2	no	XX
	latitude	Latitude of hotel	DECIMAL (10,4)	no	56.98122
	longitude	Longitude of hotel	DECIMAL (10,4)	no	24.11014

Entity	Attributes	Description	Data type, length	Nullable	Possible Values, Constraints, formats
Airport	airport_id	Unique hotel id	CHAR,5	No	XXXXN XXXX is city id
	airport_name	Name of airport	VARCHAR, 30	no	
	city_id	City the hotel resides in	CHAR,4	No	XXXX
	address	Address of airport	VARCHAR, 50	no	
	transitscore_id	Id of transitscore	CHAR,2	yes	XX
	latitude	Latitude of hotel	DECIMAL (10,4)	no	56.98122
	longitude	Longitude of hotel	DECIMAL (10,4)	no	24.11014

Entity	Attributes	Description	Data type, length	Nullable	Possible Values, Constraints, formats
Walk score	walkscore_id	Walk score id	Char(2)	no	xx

Score	walkscore_value	Values by 10 for walk score	TINYINT	no	10,20,30
	walkscore_descr iption	Description of what the value means	VARCHA R, 30	no	

Entity	Attributes	Description	Data type, length	Nullable	Possible Values, Constraints, formats
Transit Score	transitscore_id	Walk score id	Char(2)	no	xx
	transitscore_val ue	Values by 10 for transit score	TINYINT	no	10,20,30
	transitscore_des cription	Description of what the value means	VARCHA R, 30	no	

Validation Against User Requests

➢ What is the average walk score of a city's hotels

Able to group by city score of hotels

➢ If there is a hotel with a very walkable score

Able to show hotels by scores that are 80 +

➢ If a city has decent public transit from the airport and at a hotel

Able to query against transit scores from hotels and airport and find one 60+

➢ Display states with low walk and public transit score from airports and hotels

Able to group by states

Logical Data Model

Resolving Relationships

Double checking that all relationships are resolved
Primary key in red
Related foreign key in purple

City:State => M:1

City:(**city_id**, city_name, **state_id**)

State:(**state_id**, state_name)

Hotel:City => M:1

Hotel:(**hotel_id**, hotel_name, **city_id**, address, transitscore_id, walkscore_id, latitude, longitude)

City:(**city_id**, state_id)

Airport:City => M:1

Airport:(**airport_id**, airport_name, **city_id**, address, transitscore_id, latitude, longitude)

City:(**city_id**, state_id)

Walk Score: Hotel => 1:M:

Walkscore:(**walkscore_id**, walkscore_value, walkscore_description)

Hotel:(**hotel_id**,hotel_name, city_id, address, transitscore_id, walkscore_id, latitude, longitude)

Transit Score: Hotel => 1:M

Transitscore:(**transitscore_id**, transitscore_value, transitscore_description

Hotel:(**hotel_id**, hotel_name, city_id, address, transitscore_id, walkscore_id, latitude, longitude)

Transit Score: Airport => 1:M

Transitcore:(**transitscore_id**, transitscore_value, transitscore_description)

Hotel:(**hotel_id**, hotel_name, city_id, address, transitscore_id, walkscore_id, latitude, longitude)

Deciding on DBMS

I want to get practice with Microsoft SQL Server to get experience since I only used PSQL before and there are no specifications in the business case on which to use.
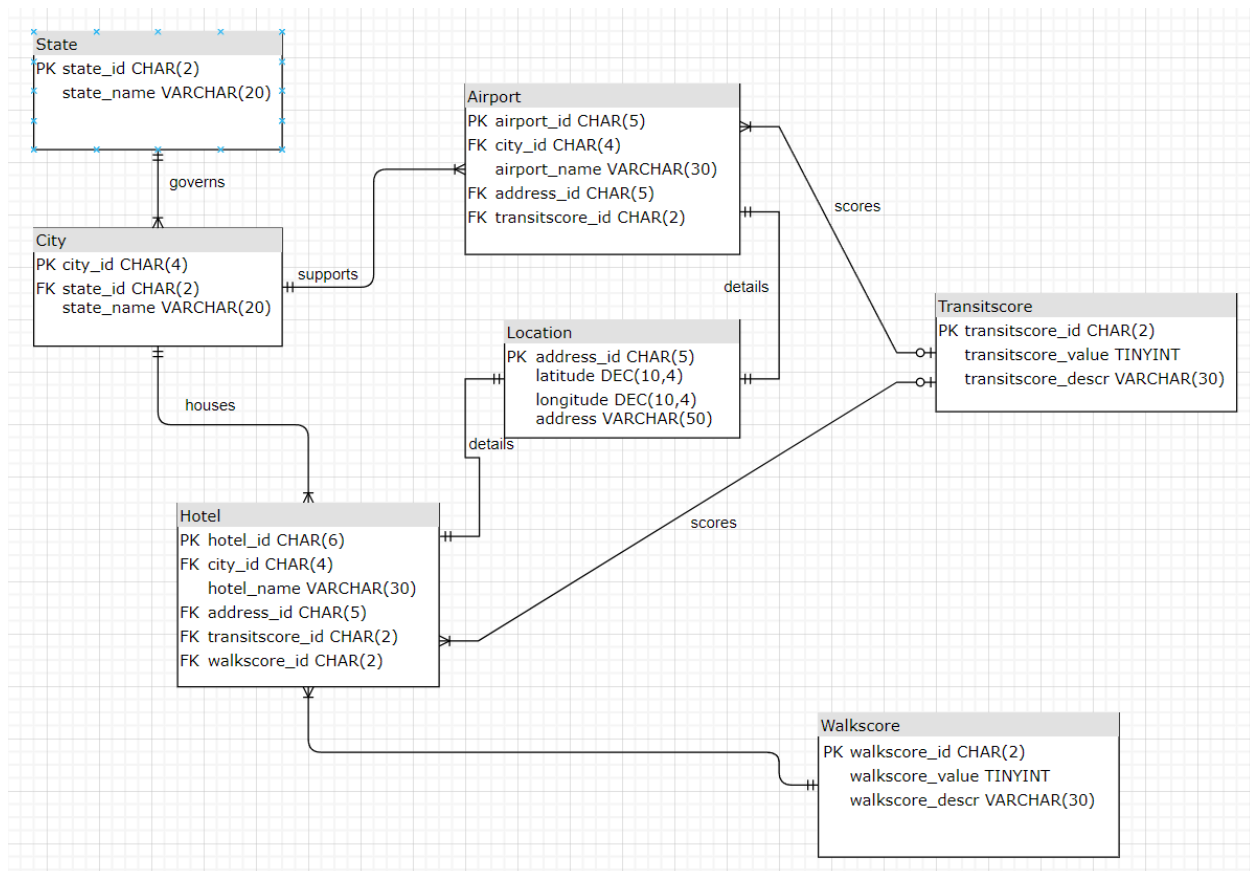
Ensure all tables are in 3NF

There are no partial or transitive dependencies except for hotel and airport.
Making the following table to resolve:
Location: (location_id, address, latitude,longitude)

Resolves both dependencies in tables.

ERD for Logical Model with data types

DB implementation

SQL DDL

Creating the database first

```
CREATE DATABASE us_walkability;
```

Now creating the 7 tables according to the logical model

```sql
CREATE TABLE State(
    state_id CHAR(2) NOT NULL PRIMARY KEY,
    state_name VARCHAR(50) NOT NULL
);

CREATE TABLE City(
    city_id CHAR(4) NOT NULL PRIMARY KEY,
    state_id CHAR(2) NOT NULL REFERENCES State(state_id),
    city_name VARCHAR(20) NOT NULL
);

CREATE TABLE Location(
    location_id CHAR(5) NOT NULL PRIMARY KEY,
    address VARCHAR(50) NOT NULL,
    latitude DEC(10,4) NOT NULL,
    longitude DEC(10,4) NOT NULL
);

CREATE TABLE Transitscore(
    transitscore_id CHAR(2) NOT NULL PRIMARY KEY,
    transitscore_value TINYINT NOT NULL,
    transitscore_descr VARCHAR(30) NOT NULL
);

CREATE TABLE Airport(
    airport_id CHAR(5) NOT NULL PRIMARY KEY,
    city_id CHAR(4) NOT NULL REFERENCES City(city_id),
    airport_name VARCHAR(30) NOT NULL,
    location_id CHAR(5) NOT NULL REFERENCES Location(location_id),
    transitscore_id CHAR(2) REFERENCES Transitscore(transitscore_id)
);

CREATE TABLE Walkscore(
    walkscore_id CHAR(2) NOT NULL PRIMARY KEY,
    walkscore_value TINYINT NOT NULL,
    walkscore_descr VARCHAR(30) NOT NULL
);

CREATE TABLE Hotel(
    hotel_id CHAR(6) NOT NULL PRIMARY KEY,
    city_id CHAR(4) NOT NULL REFERENCES City(city_id),
    hotel_name VARCHAR(30) NOT NULL,
    location_id CHAR(5) NOT NULL REFERENCES Location(location_id),
    transitscore_id CHAR(2) REFERENCES Transitscore(transitscore_id),
    walkscore_id CHAR(2) NOT NULL REFERENCES Walkscore(walkscore_id)
);
```

SQL DML

Inserting values into State:

```sql
INSERT INTO State(state_id, state_name)VALUES
    ('CA', 'California'),
    ('TX', 'Texas'),
    ('FL', 'Florida'),
    ('NY', 'New York'),
    ('PA', 'Pennsylvania'),
    ('IL', 'Illinois'),
    ('OH', 'Ohio'),
    ('NC', 'North Carolina'),
    ('GA', 'Georgia'),
    ('MD', 'Maryland');
```

Inserting values into City:

```sql
INSERT INTO City(city_id, state_id, city_name) VALUES
    ('TX00', 'TX', 'San Antonio'),
    ('TX01', 'TX', 'Austin'),
    ('CA00', 'CA', 'San Francisco'),
    ('FL00', 'FL', 'Orlando'),
    ('NY00', 'NY', 'New York City'),
    ('PA00', 'PA', 'Philadelphia'),
    ('IL00', 'IL', 'Chicago'),
    ('OH00', 'OH', 'Cleveland'),
    ('NC00', 'NC', 'Charlotte'),
    ('GA00', 'GA', 'Atlanta');
```

Inserting values into Transitscore:

```sql
INSERT INTO Transitscore VALUES
    ('T0', 00, 'No Transit'),
    ('T1', 10, 'Very Minimal Transit'),
    ('T2', 20, 'Minimal Transit'),
    ('T3', 30, 'Little Transit'),
    ('T4', 40, 'Some Transit'),
    ('T5', 50, 'Good Transit'),
    ('T6', 60, 'Great Transit'),
    ('T7', 70, 'Excellent Transit'),
    ('T8', 80, 'Almost Perfect Transit'),
    ('T9', 90, 'Riders Paradise'),
    ('TT', 100, 'Perfect Riders Paradise');
```

Inserting values into Walkscore:
These are random values, scores will be put in by API in future products.
https://www.walkscore.com/professional/research.php
https://www.walkscore.com/methodology.shtml

```sql
INSERT INTO Walkscore VALUES
    ('C0', 00, 'Extremely Car-Dependent'),
    ('C1', 10, 'Very Very Car-Dependent'),
    ('C2', 20, 'Very Car-Dependent'),
    ('C3', 30, 'Car-Dependent'),
    ('C4', 40, 'Somewhat Car-Dependent'),
    ('W0', 50, 'Somewhat Walkable'),
    ('W1', 60, 'Walkable'),
    ('W2', 70, 'Very Walkable'),
    ('W3', 80, 'Very Very Walkable'),
    ('W4', 90, 'Walkers Paradise'),
    ('W5', 100, 'Perfect Walkers Paradise');
```

Inserting values into Location:, used Google Bard to gather locations with latitude and longitude for hotels and airports since parsing by hand would take hours which will be gathered by other means in the final product. The latitude and longitude data type should be adjusted for final product for accuracy

```sql
]INSERT INTO Location VALUES
    ('SAN01','900 E Houston St',29.4279,-98.4976),
    ('SAN02','600 Alamo St',29.4229 ,-98.4945),
    ('SAN03','2100 N Interstate 35',29.4239,-98.4936),
    ('SAN04','6200 International Airport Blvd',29.4147,-98.4815),
    ('SAN05','1701 S Alamo St', 29.4246,-98.4967),
    ('SAN06','2350 N St. Marys St',29.4275  ,-98.4918),
    ('AUS01','Austin-Bergstrom International Airport',  30.2602,-97.7442),
    ('AUS02','The University of Texas at Austin - UT Tower',30.2733,-97.7841),
    ('AUS03','Zilker Park',30.2810  ,-97.7652),
    ('AUS04','Barton Creek Greenbelt',  30.2823,-97.8028);

]INSERT INTO Location VALUES
    ('SAF01', '201 5th St', 37.7749,    -122.4194),
    ('SAF02', '900 E Houston St',37.7904,   -122.4184),
    ('SAF03', '2100 N Van Ness Ave.',  37.7833,    -122.4208),
    ('SAF04', '6200 International', 37.7957,   -122.4841),
    ('SAF05', 'Pier 39, The Embarcadero',  37.7933,    -122.4192);
```

Inserting values into Airport:

```sql
]INSERT INTO Airport VALUES
    ('TX001', 'TX00', 'San Antonio International', 'SAN05', 'T2'),
    ('TX011', 'TX01', 'Austin-Bergstrom International', 'AUS01', 'T3');

]INSERT INTO Airport VALUES
    ('CA001', 'CA00', 'San Francisco International', 'SAF04', 'T7');
```

Inserting values into Hotel :

```
INSERT INTO Hotel VALUES
    ('TX0001', 'TX00', 'Omni San Antonio Hotel' , 'SAN01', 'T6', 'W3'),
    ('TX0002', 'TX00', 'Hilton San Antonio' , 'SAN02', 'T2', 'C4'),
    ('TX0003', 'TX00', 'Marriott Rivercenter' , 'SAN03', 'T5', 'W3'),
    ('TX0004', 'TX00', 'Hilton San Antonio Downtown' , 'SAN05', 'T6', 'W4'),
    ('TX0005', 'TX00', 'Marriott San Antonio Riverwalk' , 'SAN06', 'T5', 'W4'),
    ('TX0101', 'TX01', 'Kimpton Hotel Van Zandt' , 'AUS04', 'T7', 'W0'),
    ('TX0102', 'TX01', 'Hilton Austin Downtown' , 'AUS03', 'T6', 'W3'),
    ('TX0103', 'TX01', 'Courtyard Austin Downtown' , 'AUS02', 'T5', 'C3');


INSERT INTO Hotel VALUES
    ('CA0001', 'CA00', 'The Ritz-Carlton San Francisco', 'SAF01', 'T5', 'W0'),
    ('CA0002', 'CA00', 'Palace Hotel', 'SAF02', 'T7', 'W1'),
    ('CA0003', 'CA00', 'Fairmont San Francisco', 'SAF03', 'T6', 'W2'),
    ('CA0004', 'CA00', 'Zephyr at Fishermans Wharf', 'SAF05', 'T8', 'W4');
```

SQL DQL

SQL DQL to explore the data.

Following to see the full tables (results excluded):

```
SELECT * FROM State;

SELECT * FROM City;

SELECT * FROM Hotel;

SELECT * FROM Location;

SELECT * FROM Airport;

SELECT * FROM Walkscore;

SELECT * FROM Transitscore;
```

Following query to see number of table entries:

```
SELECT T.name as Table_Name, max(s.rowcnt) as Total_Records
FROM SYS.tables T
INNER JOIN SYS.sysindexes S
ON T.object_id=S.id
group by t.schema_id, t.name;
```

Table_Name	Total_Records
Airport	3
City	10
Hotel	12
Location	15
State	10
Transitscore	11
Walkscore	11

Following query to see relevant information to see info of all hotels:

```sql
SELECT h.hotel_name "Hotel Name", l.address "Address",
c.city_name "City", s.state_name "State",
w.walkscore_value "Walkscore", w.walkscore_descr "Walk Description",
t.transitscore_value "Transitscore", t.transitscore_descr "Transit Description"
FROM Hotel h
LEFT JOIN Location l
ON h.location_id=l.location_id
LEFT JOIN City c
ON h.city_id=c.city_id
LEFT JOIN State s
ON s.state_id=c.state_id
LEFT JOIN Walkscore w
ON w.walkscore_id=h.walkscore_id
LEFT JOIN Transitscore t
ON t.transitscore_id=h.transitscore_id;
```

Hotel Name	Address	City	State	Walkscore	Walk Description	Transitscore	Transit Description
The Ritz-Carlton San Francisco	201 5th St	San Francisco	California	50	Somewhat Walkable	50	Good Transit
Palace Hotel	900 E Houston St	San Francisco	California	60	Walkable	70	Excellent Transit
Fairmont San Francisco	2100 N Van Ness Ave.	San Francisco	California	70	Very Walkable	60	Great Transit
Zephyr at Fishermans Wharf	Pier 39, The Embarcadero	San Francisco	California	90	Walkers Paradise	80	Almost Perfect Transit
Omni San Antonio Hotel	900 E Houston St	San Antonio	Texas	80	Very Very Walkable	60	Great Transit
Hilton San Antonio	600 Alamo St	San Antonio	Texas	40	Somewhat Car-Dependent	20	Minimal Transit
Marriott Rivercenter	2100 N Interstate 35	San Antonio	Texas	80	Very Very Walkable	50	Good Transit
Hilton San Antonio Downtown	1701 S Alamo St	San Antonio	Texas	90	Walkers Paradise	60	Great Transit
Marriott San Antonio Riverwalk	2350 N St. Marys St	San Antonio	Texas	90	Walkers Paradise	50	Good Transit
Kimpton Hotel Van Zandt	Barton Creek Greenbelt	Austin	Texas	50	Somewhat Walkable	70	Excellent Transit
Hilton Austin Downtown	Zilker Park	Austin	Texas	80	Very Very Walkable	60	Great Transit
Courtyard Austin Downtown	The University of Texas at Austin - UT Tower	Austin	Texas	30	Car-Dependent	50	Good Transit

Following query to see relevant information to see info of all airports:

```sql
SELECT a.airport_name "Airport Name", l.address "Address",
c.city_name "City", s.state_name "State",
t.transitscore_value "Transitscore", t.transitscore_descr "Transit Description"
FROM Airport A
LEFT JOIN Location l
ON a.location_id=l.location_id
LEFT JOIN City c
ON a.city_id=c.city_id
LEFT JOIN State s
ON s.state_id=c.state_id
LEFT JOIN Transitscore t
ON t.transitscore_id=a.transitscore_id;
```

Airport Name	Address	City	State	Transitscore	Transit Description
San Francisco International	6200 International	San Francisco	California	70	Excellent Transit
San Antonio International	1701 S Alamo St	San Antonio	Texas	20	Minimal Transit
Austin-Bergstrom International	Austin-Bergstrom International Airport	Austin	Texas	30	Little Transit

Now to use SQL DQL to implement the following user queries.

➢ **What is the average walk score of a city's hotels**

```sql
SELECT c.city_name AS 'City', AVG(w.walkscore_value) AS 'Average walk score'
FROM City c
LEFT JOIN Hotel h
ON c.city_id= h.city_id
LEFT JOIN Walkscore w
ON h.walkscore_id=w.walkscore_id
WHERE h.hotel_id IS NOT NULL
GROUP BY c.city_name;
```

	City	Average walk score
1	Austin	53
2	San Antonio	76
3	San Francisco	67

➢ **If there is a hotel with a very walkable score**

Will use a score with 70 and above

```sql
SELECT h.hotel_name AS 'Hotel', c.city_name AS 'City', c.state_id AS 'State', w.walkscore_descr AS 'Description'
FROM City c
LEFT JOIN Hotel h
ON c.city_id= h.city_id
LEFT JOIN Walkscore w
ON h.walkscore_id=w.walkscore_id
WHERE w.walkscore_value >= 70;
```

Hotel	City	State	Description
Fairmont San Francisco	San Francisco	CA	Very Walkable
Zephyr at Fishermans Wharf	San Francisco	CA	Walkers Paradise
Omni San Antonio Hotel	San Antonio	TX	Very Very Walkable
Marriott Rivercenter	San Antonio	TX	Very Very Walkable
Hilton San Antonio Downtown	San Antonio	TX	Walkers Paradise
Marriott San Antonio Riverwalk	San Antonio	TX	Walkers Paradise
Hilton Austin Downtown	Austin	TX	Very Very Walkable

➢ **If a city has decent public transit from the airport and at a hotel**

Will use a score with 50 and above

17

```
SELECT h.hotel_name AS 'Hotel Name', c.city_name AS 'city', c.state_id AS 'State', 'has decent transport here and from airport'
FROM City c
LEFT JOIN Hotel h
ON c.city_id= h.city_id
LEFT JOIN Transitscore t
ON h.transitscore_id=t.transitscore_id
WHERE t.transitscore_value >= 50 AND c.city_name IN
(SELECT city_name
FROM City c
LEFT JOIN Airport a
ON c.city_id= a.city_id
LEFT JOIN Transitscore t
ON a.transitscore_id=t.transitscore_id
WHERE t.transitscore_value >= 50);
```

	Hotel Name	city	State	(No column name)
1	The Ritz-Carlton San Francisco	San Francisco	CA	has decent transport here and from airport
2	Palace Hotel	San Francisco	CA	has decent transport here and from airport
3	Fairmont San Francisco	San Francisco	CA	has decent transport here and from airport
4	Zephyr at Fishermans Wharf	San Francisco	CA	has decent transport here and from airport

➢ **Display states with low walk and public transit score from airports and hotels**

Will display those with amount of airports less of transit score of 50 and number of hotels with less of walkscore of 50

```
SELECT s.state_id AS "State", count(*) AS "Number of low walkscore hotels"
FROM State s
LEFT JOIN City c
ON s.state_id=c.state_id
RIGHT JOIN Hotel h
ON h.city_id=c.city_id
LEFT JOIN Walkscore w
ON h.walkscore_id=w.walkscore_id
WHERE w.walkscore_value < 50
GROUP BY s.state_id;
```

	State	Number of low walkscore hotels
1	TX	2

```
SELECT s.state_id AS "State", count(*) AS "Number of low transit aiports"
FROM State s
LEFT JOIN City c
ON s.state_id=c.state_id
RIGHT JOIN airport a
ON a.city_id=c.city_id
LEFT JOIN Transitscore t
ON a.transitscore_id=t.transitscore_id
WHERE t.transitscore_value < 50
GROUP BY s.state_id;
```

	State	Number of low transit aiports
1	TX	2

18

Conclusion

The main achievements of the project work is that a future project will have a database scheme ready to go to implement. Joining now makes more logical sense when and how to use along with foreign keys in a project I made.

Problems that I came across with are from the queries, since I did not have a client to clarify the queries I made them easy to be executed to be possible in the database. Otherwise I would have to adjust table names.

As stated earlier, I used Google Bard to gather location, latitude, address, hotel name, and I picked random walk scores and transit scores for the locations since it would have taken hours to complete the project. Also the need for latitude and location to be increased to DEC(10,5) for better accuracy or Google Bard might have given incorrect information but will be using APIs or data scraping for hotel information.

References

Erkec, Esat. "SQL Multiple Joins for Beginners with Examples." *SQL Shack - Articles about Database Auditing, Server Performance, Data Recovery, and More*, 16 Oct. 2019, www.sqlshack.com/sql-multiple-joins-for-beginners-with-examples/.

Google. "Bard - Chat Based AI Tool from Google, Powered by PaLM 2." *Bard.google.com*, bard.google.com/chat.

"How to Get Record Count for All the Tables in SQL Sever | Get List of All Tables & Their Row Count." *Www.youtube.com*, www.youtube.com/watch?v=_jYmcaWOBZs. Accessed 18 Dec. 2023.

Stack Overflow. "Stack Overflow - Where Developers Learn, Share, & Build Careers." *Stack Overflow*, 2022, stackoverflow.com/.

W3Schools. "W3Schools Online Web Tutorials." *W3schools.com*, W3Schools, 2023,

www.w3schools.com/.

Walk Score. "Find Apartments for Rent and Rentals - Get Your Walk Score." *Walk Score*, 2018,

www.walkscore.com/.