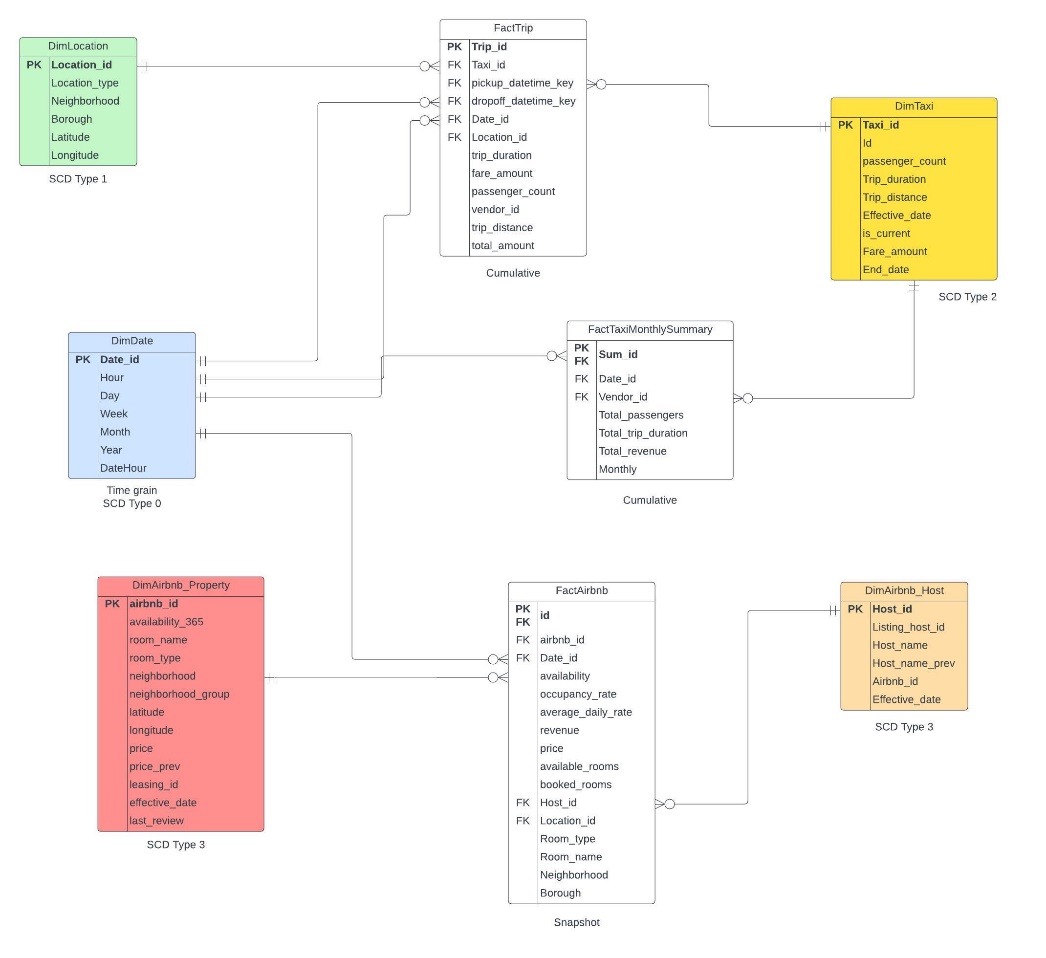
Dimension and Fact Tables Construction

**Provide code and screenshots of loading the data into the dimensions and the fact. Currently, you do not need to worry about maintenance of slowly changing dimensions, the focus is on the initial data load. If you are loading into SCD2 or SCD3, make sure to show the SCD maintenance attributes populated.**



**Staging**

ALTER TABLE Taxi\_staging ADD taxi\_datetime DATETIME;

UPDATE Taxi\_staging

SET taxi\_datetime = DATEADD(HOUR, DATEDIFF(HOUR, 0, dropoff\_datetime), 0);

(The purpose of adding taxi\_datetime is to help create the monthly function in FactTaxiMonthlySummary)

**Dimension**

1. DimAirbnb\_Property

-- Create DimAirbnb\_Property

CREATE TABLE DimAirbnb\_Property (

Airbnb\_id INT IDENTITY(1,1) PRIMARY KEY,

Leasing\_id INT,

Room\_name NVARCHAR(255),

Room\_type NVARCHAR(50),

Price INT,

Neighborhood NVARCHAR(255),

Neighborhood\_group NVARCHAR(255),

Latitude DECIMAL(9,6),

Longitude DECIMAL(9,6),

Availability\_365 INT,

Price\_prev INT

);

A screenshot of a computer

Description automatically generated

-- Insert data into DimAirbnb\_Property

INSERT INTO DimAirbnb\_Property (Airbnb\_id, Room\_name, Room\_type, Price, Neighborhood, Neighborhood\_group, Latitude, Longitude, Availability\_365)

SELECT airbnb\_id, name, room\_type, price, neighborhood, neighborhood\_group, latitude, longitude, availability\_365

FROM Airbnb\_staging;

A screenshot of a computer

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated

2. DimAirbnb\_Host

-- Create DimAirbnb\_Host

CREATE TABLE DimAirbnb\_Host (

Host\_id INT IDENTITY(1,1) PRIMARY KEY,

Listing\_host\_id INT,

Host\_name NVARCHAR(255),

Host\_name\_prev NVARCHAR(255)

);

A screenshot of a computer screen

Description automatically generated with medium confidence

-- Insert data into DimAirbnb\_Host

INSERT INTO DimAirbnb\_Host (Listing\_host\_id, Host\_name)

SELECT DISTINCT host\_id, host\_name

FROM Airbnb\_staging;

A screenshot of a computer

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated

3. DimDate

CREATE TABLE DimDate (

Date\_id INT IDENTITY(1,1) PRIMARY KEY,

DateHour DATETIME NOT NULL,

Hour INT NOT NULL,

Day INT NOT NULL,

Month INT NOT NULL,

Year INT NOT NULL,

Week INT NOT NULL

);

A screenshot of a computer

Description automatically generated with medium confidence

WITH DateRange (DateHour) AS (

SELECT CAST('2010-01-01 00:00:00' AS DATETIME)

UNION ALL

SELECT DATEADD(hour, 1, DateHour)

FROM DateRange

WHERE DateHour < '2023-12-31 23:00:00'

)

INSERT INTO DimDate (DateHour, Hour, Day, Month, Year, Week)

SELECT

DateHour,

DATEPART(hour, DateHour) AS Hour,

DAY(DateHour) AS Day,

MONTH(DateHour) AS Month,

YEAR(DateHour) AS Year,

DATEPART(wk, DateHour) AS Week

FROM DateRange

OPTION (MAXRECURSION 0);

A screenshot of a computer

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated

4. DimLocation

CREATE TABLE DimLocation (

Location\_id INT IDENTITY(1,1) PRIMARY KEY,

Location\_type VARCHAR(50),

Neighborhood VARCHAR(255),

Borough VARCHAR(255),

Longitude DECIMAL(9, 6),

Latitude DECIMAL(9, 6)

);

A screenshot of a computer

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated

INSERT INTO DimLocation (Latitude, Longitude, Location\_type)

SELECT DISTINCT

dropoff\_latitude,

dropoff\_longitude,

'Taxi Dropoff'

FROM Taxi\_staging;

(Same process for the Airbnb and Taxi Pickup, the difference is that we also need to consider about the Borough and Neighborhood from Airbnb\_staging)

5. DimTaxi

CREATE TABLE DimTaxi (

Taxi\_id INT IDENTITY(1,1) NOT NULL,

Id NVARCHAR(255),

Pickup\_datetime DATETIME,

Dropoff\_datetime DATETIME,

Passenger\_count INT,

Trip\_distance FLOAT,

pickup\_latitude FLOAT,

pickup\_longitude FLOAT,

dropoff\_latitude FLOAT,

dropoff\_longitude FLOAT,

Fare\_amount FLOAT,

Effective\_date DATE,

End\_date DATE,

Is\_current BIT

);

A screenshot of a computer

Description automatically generated with medium confidence

INSERT INTO DimTaxi (

Id,

Pickup\_datetime,

Dropoff\_datetime,

Passenger\_count,

Effective\_date,

End\_date,

Is\_current,

pickup\_latitude,

pickup\_longitude,

dropoff\_latitude,

dropoff\_longitude

)

SELECT

taxi\_id,

Pickup\_datetime,

Dropoff\_datetime,

Passenger\_count,

CAST(GETDATE() AS DATE),

NULL,

1,

pickup\_latitude,

pickup\_longitude,

dropoff\_latitude,

dropoff\_longitude

FROM

Taxi\_staging;

UPDATE DimTaxi

SET Trip\_duration =

CASE

WHEN Taxi\_staging.trip\_duration IS NOT NULL THEN Taxi\_staging.Trip\_duration

ELSE DATEDIFF(second, pickup\_datetime, dropoff\_datetime)

END

FROM DimTaxi

JOIN Taxi\_staging ON DimTaxi.Id = Taxi\_staging.taxi\_id;

A screenshot of a computer

Description automatically generated with medium confidence

UPDATE DimTaxi

SET

trip\_distance = CASE

WHEN ts.Trip\_duration IS NOT NULL THEN ts.trip\_duration \* 0.2

ELSE 0

END,

fare\_amount = CASE

WHEN ts.Trip\_duration IS NOT NULL THEN ts.trip\_duration \* 0.05 + 10

ELSE 0

END

FROM DimTaxi t

JOIN taxi\_staging ts ON t.Id = ts.taxi\_id;

**Fact**

FactAirbnb:

-- Create FactAirbnb

CREATE TABLE FactAirbnb (

Id INT IDENTITY(1,1) PRIMARY KEY NOT NULL,

Airbnb\_id INT NOT NULL FOREIGN KEY REFERENCES DimAirbnb\_Property(Airbnb\_id),

Date\_id INT NOT NULL FOREIGN KEY REFERENCES DimDate(Date\_id),

Host\_id INT NOT NULL FOREIGN KEY REFERENCES DimAirbnb\_Host(Host\_id),

Room\_name VARCHAR(300),

Room\_type VARCHAR(50),

Neighborhood VARCHAR(100),

Borough VARCHAR(50),

Availability INT,

Price FLOAT,

Revenue FLOAT,

Available\_rooms INT,

Booked\_rooms INT,

Occupancy\_rate FLOAT,

Average\_daily\_rate FLOAT

);

);

-- Insert data into FactAirbnb

INSERT INTO FactAirbnb (Airbnb\_id, Date\_id, Host\_id, Room\_name, Room\_type, Neighborhood, Borough, Availability, Price, Revenue, Occupancy\_rate, Average\_daily\_rate)

SELECT

p.Airbnb\_id,

d.Date\_id,

h.Host\_id,

p.Room\_name,

p.Room\_type,

p.Neighborhood,

p.Neighborhood\_group,

p.Availability\_365,

p.Price,

p.Price \* p.Availability\_365 AS Revenue,

NULL AS Occupancy\_rate,

NULL AS Average\_daily\_rate

FROM

DimAirbnb\_Property p

INNER JOIN DimDate d ON CAST(p.last\_review AS DATE) = CAST(d.DateHour AS DATE)

INNER JOIN DimAirbnb\_Host h ON p.Leasing\_id = h.airbnb\_id;

UPDATE FactAirbnb

SET

available\_rooms = Availability / 365,

booked\_rooms = (365 - Availability) / 365

UPDATE FactAirbnb

SET Occupancy\_rate = CASE WHEN Availability > 0 THEN (365-Availability)/365 ELSE 0 END,

Average\_daily\_rate = CASE WHEN Availability > 0 THEN Revenue/Availability ELSE 0 END;

ALTER TABLE FactAirbnb

ADD Location\_id INT;

UPDATE FactAirbnb

SET location\_id = (

SELECT TOP 1 location\_id

FROM DimLocation

WHERE Location\_type = 'Airbnb'

AND Neighborhood = FactAirbnb.Neighborhood

AND Borough = FactAirbnb.Borough

);

--- Current in use

UPDATE FactAirbnb

SET

Availability = p.availability\_365 / 365.0,

Revenue = p.availability\_365 \* p.price,

Available\_rooms = p.availability\_365,

Booked\_rooms = 365 - p.availability\_365,

Occupancy\_rate = (365 - p.availability\_365) / 365.0,

Average\_daily\_rate = (

SELECT AVG(price)

FROM DimAirbnb\_Property

WHERE Airbnb\_id = FactAirbnb.Id

)

FROM DimAirbnb\_Property p

WHERE p.Airbnb\_id = FactAirbnb.Id;

A screenshot of a computer

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated

FactTaxiMonthlySummary:

CREATE TABLE FactTaxiMonthlySummary (

Sum\_id INT IDENTITY(1,1) PRIMARY KEY NOT NULL,

Date\_id INT NOT NULL,

Taxi\_id INT NOT NULL,

Total\_passengers INT,

Total\_trip\_duration FLOAT,

Total\_revenue FLOAT,

Monthly INT,

Vendor\_id INT

);

A screenshot of a computer

Description automatically generated with medium confidence

INSERT INTO FactTaxiMonthlySummary (Date\_id, Taxi\_id, Total\_passengers, Total\_trip\_duration, Total\_revenue, Monthly)

SELECT

d.Date\_id,

tx.Taxi\_id,

SUM(tx.Passenger\_count) AS Total\_passengers,

SUM(tx.Trip\_duration) AS Total\_trip\_duration,

SUM(tx.Fare\_amount) AS Total\_revenue,

FORMAT(d.DateHour, 'yyyy-MM') AS Monthly

FROM

DimTaxi tx

JOIN Taxi\_staging t ON t.taxi\_id = tx.Id

JOIN DimDate d ON CAST(t.taxi\_datetime AS DATE) = CAST(d.DateHour AS DATE)

GROUP BY

d.Date\_id,

tx.Taxi\_id,

FORMAT(d.DateHour, 'yyyy-MM');

ALTER TABLE FactTrip

ALTER COLUMN Pickup\_datetime\_key DATETIME;

ALTER TABLE FactTrip

ALTER COLUMN Dropoff\_datetime\_key DATETIME;

A screenshot of a computer

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

FactTrip:

CREATE TABLE FactTrip (

Trip\_id INT IDENTITY(1,1) PRIMARY KEY NOT NULL,

Taxi\_id INT NOT NULL,

Location\_id INT NOT NULL,

Date\_id INT NOT NULL,

Pickup\_datetime\_key INT,

Dropoff\_datetime\_key INT,

Pickup\_location INT,

Dropoff\_location INT,

Passenger\_count INT,

Trip\_distance FLOAT,

Trip\_duration INT,

Fare\_amount FLOAT,

Total\_amount FLOAT

);

INSERT INTO FactTrip (Taxi\_id, Location\_id, Date\_id, Pickup\_datetime\_key, Dropoff\_datetime\_key, Passenger\_count, Trip\_distance, Trip\_duration, Fare\_amount, Vendor\_id)

SELECT DISTINCT

tx.Taxi\_id,

l.Location\_id,

d.Date\_id,

ts.pickup\_datetime,

ts.dropoff\_datetime,

tx.Passenger\_count,

tx.Trip\_distance,

tx.Trip\_duration,

tx.Fare\_amount,

ts.vendor\_id

FROM

DimTaxi tx

JOIN Taxi\_staging ts ON ts.taxi\_id = tx.Id

JOIN (

SELECT l1.Location\_id, l1.latitude

FROM DimLocation l1

JOIN Taxi\_staging ts1 ON ts1.dropoff\_latitude = l1.latitude AND l1.Location\_type = 'Taxi Dropoff'

UNION

SELECT l2.Location\_id, l2.latitude

FROM DimLocation l2

JOIN Taxi\_staging ts2 ON ts2.pickup\_latitude = l2.latitude AND l2.Location\_type = 'Taxi Pickup'

) AS l ON l.latitude = ts.dropoff\_latitude OR l.latitude = ts.pickup\_latitude

JOIN DimDate d ON CAST(ts.taxi\_datetime AS DATE) = CAST(d.DateHour AS DATE)

UNION

SELECT DISTINCT

tx.Taxi\_id,

l.Location\_id,

d.Date\_id,

ts.pickup\_datetime,

ts.dropoff\_datetime,

tx.Passenger\_count,

tx.Trip\_distance,

tx.Trip\_duration,

tx.Fare\_amount,

ts.vendor\_id

FROM

DimTaxi tx

JOIN Taxi\_staging ts ON ts.taxi\_id = tx.Id

JOIN (

SELECT l1.Location\_id, l1.longitude

FROM DimLocation l1

JOIN Taxi\_staging ts1 ON ts1.dropoff\_longitude = l1.longitude AND l1.Location\_type = 'Taxi Dropoff'

UNION

SELECT l2.Location\_id, l2.longitude

FROM DimLocation l2

JOIN Taxi\_staging ts2 ON ts2.pickup\_longitude = l2.longitude AND l2.Location\_type = 'Taxi Pickup'

) AS l ON l.longitude = ts.dropoff\_longitude OR l.longitude = ts.pickup\_longitude

JOIN DimDate d ON CAST(ts.taxi\_datetime AS DATE) = CAST(d.DateHour AS DATE)

A screenshot of a computer

Description automatically generated

UPDATE FactTrip

SET Total\_amount = subquery.Total\_amount

FROM (

SELECT Taxi\_id, SUM(Fare\_amount) as Total\_amount

FROM FactTrip

GROUP BY Taxi\_id

) AS subquery

WHERE FactTrip.Taxi\_id = subquery.Taxi\_id;

**A screenshot of a computer

Description automatically generated with medium confidence**

A screenshot of a computer

Description automatically generated