**Student Number: 10241485**

**Name: Jeslyn Ho Ka Yan   
Date: 24 Oct 2024**

**Assignment 2, Task 1**

**A computer screen shot of a computer program

Description automatically generated with medium confidence**

1. **1) Sepcification of Data Cube:**

**Fact**

1. **endpoint1 and endpoint2:** These represent the starting and ending points of each segment.
2. **Trip-name:** Identifier for each trip.
3. **Licence-number:** Identifier for each driver.
4. **registration:** Identifier for each bus.
5. **Lot-number**: Lot number associated with the bus’s location.
6. **street:** Identifier for the street where the bus is located.
7. **city:** Identifier for the city where the bus is located.
8. **Date:** Date for time-based analysis (day, month, year)

**Measure**

1. **Kilometers Traveled(length) :** Derived from the length of each segment.
2. **Fuel Consumption:** Derived from the MODEL table based on the bus.
3. **Number of Passengers:** Taken from numOfPassenger in the Segment Trip table.
4. **Total Trips Performed:** Count of trips for each bus and driver over time.
5. **Average Duration:** If available, calculate the average time per segment or trip.
6. **Number of Drivers per Trip:** Count of unique drivers assigned to each trip.

**Names of Dimensions and the hierarchies**

**1. Date Dimension (**date, day, month, quarter, year)

* **Hierarchy**: Day → Month → Quarter → Year

**2. Bus Dimension (**registration, model-name, manufacturer, depot)

* **Hierarchy**: Bus → Model → Manufacturer

**3. Trip Dimension (**trip-name, total-segments, total-length)

* **Hierarchy**: Trip (no further levels)

**4. Segment Dimension (**segment-id, endpoint1, endpoint2, length)

* **Hierarchy**: Segment (no further levels)

**5. Driver Dimension (**licence-number, full-name, years-of-experience)

* **Hierarchy**: Driver (no further levels)

**6. Depot Dimension (**depot-id, lot-number, street, city)

* **Hierarchy**: Street → City

**7. Passenger Dimension (**num-of-passengers, average-passengers)

* **Hierarchy**: Passenger Count (no further levels)

**Explanation on how to obtain the measurements**

**(i)Total Kilometers Traveled by Each Bus per Year, Month, Day**

SELECT

registration AS bus\_id,

YEAR(date) AS year,

MONTH(date) AS month,

DAY(date) AS day,

SUM(length) AS total\_kilometers

FROM

Segment\_Trip

GROUP BY

registration, YEAR(date), MONTH(date), DAY(date);

**Explanation:**

By adding up all of the trip segments that are listed in the Segment\_Trip table and grouping them by bus ID and date (year, month, and day), this query determines the total distance that each bus has travelled.

**(ii)Total Number of Trips Performed per Bus, Driver, Year, Month, Day**

SELECT

registration AS bus\_id,

licence\_number AS driver\_id,

YEAR(date) AS year,

MONTH(date) AS month,

DAY(date) AS day,

COUNT(DISTINCT trip\_name) AS total\_trips

FROM

Fact\_Table

GROUP BY

registration, licence\_number, YEAR(date), MONTH(date), DAY(date);

**Explanation:**

The total number of journeys made by each bus and driver over particular time periods (year, month, and day) is determined by this query, which counts individual trips in the Fact\_Table.

**(iii)Total Number of Drivers per Trip**

SELECT

trip\_name,

COUNT(DISTINCT licence\_number) AS total\_drivers

FROM

Fact\_Table

GROUP BY

trip\_name;

**Explanation:**

This query provides information about driver assignment and utilisation by counting the number of distinct drivers allocated to each trip.

**(iv)Total Number of Buses Traveling per Trip Segment, Trip, Year, Month, Day**

SELECT

trip\_name,

endpoint1,

endpoint2,

YEAR(date) AS year,

MONTH(date) AS month,

DAY(date) AS day,

COUNT(DISTINCT registration) AS number\_of\_buses

FROM

Segment\_Trip

GROUP BY

trip\_name, endpoint1, endpoint2, YEAR(date), MONTH(date), DAY(date);

**Explanation:**

By calculating the number of buses used for each trip segment, this query enables examination of fleet usage across route segments.

**(v)Average Duration of Bus Travel per Trip Segment, Trip, Year, Month, Day**

SELECT

trip\_name,

endpoint1,

endpoint2,

YEAR(start\_date) AS year,

MONTH(start\_date) AS month,

DAY(start\_date) AS day,

AVG(TIMESTAMPDIFF(MINUTE, start\_date\_endpoint1, end\_date\_endpoint2)) AS average\_duration\_minutes

FROM

Segment\_Trip

GROUP BY

trip\_name, endpoint1, endpoint2, YEAR(start\_date), MONTH(start\_date), DAY(start\_date);

**Explanation:**

This query determines the average bus travel time for each journey segment, offering information on efficiency and travel times.

**(vi)Total Fuel Consumption per Trip Segment, Trip, Bus Model, Manufacturer, Year, Month, Day**

SELECT

s.trip\_name,

s.endpoint1,

s.endpoint2,

m.model\_name,

manuf.name AS manufacturer,

YEAR(date) AS year,

MONTH(date) AS month,

DAY(date) AS day,

SUM(m.fuel\_consumption \* s.length) AS total\_fuel\_consumption

FROM

Segment\_Trip

JOIN

Bus b ON s.registration = b.registration

JOIN

Model m ON b.model\_name = m.model\_name

JOIN

Manufacturer manuf ON m.manufacturer\_id = manuf.id

GROUP BY

s.trip\_name, s.endpoint1, s.endpoint2, m.model\_name, manuf.name, YEAR(date), MONTH(date), DAY(date);

**Explanation:**

This query calculates the total fuel consumption for each trip segment, trip, bus model, and manufacturer, giving insights into fuel efficiency and usage patterns.

**(vii)Total Number of Trips per Bus, Depot, City**

SELECT

b.registration AS bus\_id,

d.lot\_number AS depot\_id,

d.city,

COUNT(DISTINCT trip\_name) AS total\_trips

FROM

Fact\_Table f

JOIN

Bus b ON f.registration = b.registration

JOIN

Depot d ON b.depot\_id = d.id

GROUP BY

b.registration, d.lot\_number, d.city;

**Explanation:**

This query counts the total number of trips performed by each bus at different depots and cities, helping analyze bus utilization across locations.

**(viii)Total Number of Passengers per Segment, Trip, Year, Month, Day**

SELECT

trip\_name,

endpoint1,

endpoint2,

YEAR(date) AS year,

MONTH(date) AS month,

DAY(date) AS day,

SUM(numOfPassenger) AS total\_passengers

FROM

Segment\_Trip

GROUP BY

trip\_name, endpoint1, endpoint2, YEAR(date), MONTH(date), DAY(date);

**Explanation:**

This query provides information on passenger volume and demand on various routes by adding up the number of passengers for each section of a journey.

**(ix)Largest Number of Passengers per Bus, Trip**

SELECT

registration AS bus\_id,

trip\_name,

MAX(numOfPassenger) AS max\_passengers

FROM

Segment\_Trip

GROUP BY

registration, trip\_name;

**Explanation:**

This query identifies the maximum number of passengers that traveled on each trip per bus, highlighting peak usage.

**(x)Average Number of Passengers per Trip**

SELECT

trip\_name,

AVG(numOfPassenger) AS average\_passengers

FROM

Segment\_Trip

GROUP BY

trip\_name;

**Explanation:**

In order to understand the typical load on various routes, this query determines the average number of passengers for each journey.

**2) Three-dimensional Data Cube:**

Date, Bus, and Trip can be used as axes to see this sample cube in three dimensions. The total passengers for a given combination of date, bus, and trip are represented by each cell in the cube.

**Measure**

* A cube with numbers and a bus id

  Description automatically generated**Total Passengers**: Number of passengers recorded on each trip segmen