DATA MINING LAB – 3

NAME – UDAY BOLLA

ROLL NO – 22CS3024

BRANCH – CSE

Nearest Neighbor Query

Part a)

```
CREATE TABLE store_location (
   store_id INT PRIMARY KEY,
   store_name VARCHAR(255),
   latitude DECIMAL(9,6),
   longitude DECIMAL(9,6)
);

INSERT INTO store_location (store_id, store_name, latitude, longitude)
VALUES
(1, 'Tech Store', 40.712776, -74.005974),
(2, 'Gadget Hub', 34.052235, -118.243683),
(3, 'Style Mart', 51.507351, -0.127758),
(4, 'Apparel Corner', 48.856613, 2.352222),
(5, 'Home Goods Store', 35.689487, 139.691711);


SELECT store_id, store_name, latitude, longitude,
     SQRT((latitude - 40.712776) * (latitude - 40.712776) + (longitude + 74.005974) * (longitude +
74.005974)) AS distance
FROM store_location
ORDER BY distance
LIMIT 3;
```

Output

store_id	store_name	latitude	longitude	distance
1	Tech Store	40.712776	-74.005974	0
2	Gadget Hub	34.052235	-118.243683	44.73631303517717
3	Style Mart	51.507351	-0.127758	74.66266569560237

Part b)

```
CREATE TABLE product (
   product_id INTEGER PRIMARY KEY,
   product_name TEXT NOT NULL,
   price DECIMAL(10, 2) NOT NULL,
   sales_volume INTEGER NOT NULL
```

```
);
INSERT INTO product (product_id, product_name, price, sales_volume)
VALUES
(1, 'Smartphone', 700.00, 15000),
(2, 'Laptop', 1200.00, 30000),
(3, 'Shoes', 50.00, 5000),
(4, 'T-shirt', 25.00, 2000),
(5, 'Refrigerator', 500.00, 25000);

SELECT product_id, product_name,
abs(price - 1500) as price_diff,
abs(sales_volume - 1000) as sales_diff
FROM product
ORDER BY price_diff + sales_diff asc
LIMIT 3;
```

Output

product_id	product_name	price_diff	sales_diff
4	T-shirt	1475	1000
3	Shoes	1450	4000
1	Smartphone	800	14000

Part c)

```
CREATE TABLE region_sales (
   region_name TEXT NOT NULL,
   product_name TEXT NOT NULL,
   total_sales DECIMAL(10, 2) NOT NULL
);
INSERT INTO region_sales (region_name, product_name, total_sales)
VALUES
('North', 'Smartphone', 50000.00),
('South', 'Laptop', 60000.00),
('East', 'Shoes', 45000.00),
('West', 'T-shirt', 55000.00),
('North', 'Refrigerator', 10000.00),
('Central', 'Microwave', 52000.00);

SELECT region_name, SUM(total_sales) AS total_sales,
     ABS(SUM(total_sales) - 53000.00) AS sales_difference
FROM region_sales
GROUP BY region_name
ORDER BY sales_difference
LIMIT 3;
```

Output

region_name	total_sales	sales_difference
Central	52000	1000
West	55000	2000
North	60000	7000

Top-k Query

Part b)

```
CREATE TABLE sales_data1 (
   region TEXT NOT NULL,
   profit DECIMAL(10,2) NOT NULL,
   sales_amount DECIMAL(10,2) NOT NULL,
   sale_date DATE NOT NULL
);
INSERT INTO sales_data1 (region, profit, sales_amount, sale_date)
VALUES
('North', 2000.00, 10000.00, '2024-03-15'),
('South', 3000.00, 15000.00, '2024-02-20'),
('East', 4000.00, 20000.00, '2024-01-10'),
('West', 1500.00, 8000.00, '2024-05-05'),
('Central', 5000.00, 25000.00, '2024-06-12'),
('North', 1000.00, 5000.00, '2024-07-01'),
('South', 2500.00, 12000.00, '2024-04-25');
SELECT region,
      SUM(profit) AS total_profit,
      SUM(sales_amount) AS total_sales_amount,
      (SUM(profit) / SUM(sales_amount)) * 100 AS profit_margin
FROM sales_data1
WHERE sale_date >= '2024-01-01'
GROUP BY region
ORDER BY profit_margin DESC
LIMIT 3;
```

Output

region	total_profit	total_sales_amount	profit_margin
West	1500	8000	0
South	5500	27000	0
North	3000	15000	0

Part c)

```
CREATE TABLE sales_data2 (
   customer_id INT NOT NULL,
   customer_name TEXT NOT NULL,
   product_name TEXT NOT NULL,
   sales_amount DECIMAL(10,2) NOT NULL
);
INSERT INTO sales_data2 (customer_id, customer_name, product_name, sales_amount)
VALUES
(1, 'A', 'Laptop', 1200.00),
(2, 'B', 'Smartphone', 800.00),
(3, 'C', 'Headphones', 150.00),
```

```
(4, 'D', 'Smartwatch', 250.00),
(5, 'E', 'Tablet', 500.00),
(6, 'F', 'Smartphone', 750.00),
(7, 'G', 'Tablet', 400.00),
(8, 'H', 'Smartwatch', 300.00),
(9, 'I', 'Laptop', 1200.00),
(10, 'J', 'Smartphone', 650.00),
(11, 'K', 'Laptop', 1100.00),
(12, 'L', 'Headphones', 120.00),
(13, 'M', 'Smartwatch', 280.00),
(14, 'N', 'Smartphone', 900.00),
(15, 'O', 'Tablet', 450.00);

SELECT customer_id,
    customer_name,
    SUM(sales_amount) AS total_purchase_amount
FROM sales_data2
GROUP BY customer_id, customer_name
ORDER BY total_purchase_amount DESC
LIMIT 10;
```

Output

customer_id	customer_name	total_purchase_amount
1	A	1200
9	I	1200
11	K	1100
14	N	900
2	B	800

2	B	800
6	F	750
10	J	650
5	E	500
15	O	450
7	G	400

Star Schema

```
DROP TABLE IF EXISTS sales_fact;
DROP TABLE IF EXISTS product;
DROP TABLE IF EXISTS store;
DROP TABLE IF EXISTS customer;
DROP TABLE IF EXISTS time;


CREATE TABLE product (
   product_id INTEGER PRIMARY KEY,
   product_name TEXT NOT NULL,
   category TEXT NOT NULL,
   brand TEXT NOT NULL
```

```
);


CREATE TABLE store (
   store_id INTEGER PRIMARY KEY,
   store_name TEXT NOT NULL,
   region TEXT NOT NULL,
   city TEXT NOT NULL
);


CREATE TABLE customer (
   customer_id INTEGER PRIMARY KEY,
   customer_name TEXT NOT NULL,
   age INTEGER NOT NULL,
   gender TEXT NOT NULL
);


CREATE TABLE time (
   time_id INTEGER PRIMARY KEY,
   sales_date DATE NOT NULL,
   year INTEGER NOT NULL,
   quarter INTEGER NOT NULL,
   month INTEGER NOT NULL
);


CREATE TABLE sales_fact (
   sales_id INTEGER PRIMARY KEY,
   product_id INTEGER,
   store_id INTEGER,
   customer_id INTEGER,  -- Corrected reference
   time_id INTEGER,
   quantity_sold INTEGER NOT NULL,
   sales_amount DECIMAL(10, 2) NOT NULL,
   FOREIGN KEY (product_id) REFERENCES product (product_id),
   FOREIGN KEY (store_id) REFERENCES store (store_id),
   FOREIGN KEY (customer_id) REFERENCES customer (customer_id), -- Fixed spelling
   FOREIGN KEY (time_id) REFERENCES time (time_id)
);


INSERT INTO product (product_id, product_name, category, brand)
VALUES
(1, 'Smartphone', 'Electronics', 'Brand A'),
(2, 'Laptop', 'Electronics', 'Brand B'),
(3, 'Shoes', 'Fashion', 'Brand C'),
(4, 'T-shirt', 'Fashion', 'Brand D'),
(5, 'Refrigerator', 'Home Appliances', 'Brand E');
```

```
INSERT INTO store (store_id, store_name, region, city)
VALUES
(1, 'Tech Store', 'North', 'New York'),
(2, 'Gadget Hub', 'South', 'Los Angeles'),
(3, 'Style Mart', 'East', 'Chicago'),
(4, 'Apparel Corner', 'West', 'San Francisco'),
(5, 'Home Goods Store', 'Central', 'Dallas');

INSERT INTO customer (customer_id, customer_name, age, gender)
VALUES
(1, 'John Doe', 30, 'Male'),
(2, 'Jane Smith', 25, 'Female'),
(3, 'Michael Johnson', 45, 'Male'),
(4, 'Emily Davis', 35, 'Female'),
(5, 'Sarah Brown', 50, 'Female');


INSERT INTO time (time_id, sales_date, year, quarter, month)
VALUES
(1, '2024-01-15', 2024, 1, 1),
(2, '2024-02-20', 2024, 1, 2),
(3, '2024-03-01', 2024, 1, 3),
(4, '2024-04-05', 2024, 2, 4),
(5, '2024-05-10', 2024, 2, 5);

INSERT INTO sales_fact (sales_id, product_id, store_id, customer_id, time_id, quantity_sold,
sales_amount)
VALUES
(1, 1, 1, 1, 1, 3, 1500.00),
(2, 2, 2, 2, 2, 2, 2400.00),
(3, 3, 3, 3, 3, 4, 200.00),
(4, 4, 4, 4, 4, 5, 125.00),
(5, 5, 5, 5, 5, 1, 500.00);


SELECT
   p.category,
   s.region,
   SUM(sf.sales_amount) AS total_sales
FROM
   sales_fact sf
JOIN
   product p ON sf.product_id = p.product_id
JOIN
   store s ON sf.store_id = s.store_id
GROUP BY
   p.category, s.region
ORDER BY
   total_sales DESC;
```

Output

category	region	total_sales
Electronics	South	2400
Electronics	North	1500
Home Appliances	Central	500
Fashion	East	200
Fashion	West	125