Coding Challenge Medical Insurance Cost Analysis

Create a table in MySQL:

```
CREATE TABLE medical_insurance (
    policy_id INT PRIMARY KEY,
    age INT,
    gender VARCHAR(10),
    bmi DECIMAL(5,2),
    children INT,
    smoker VARCHAR(3),       -- Yes / No
    region VARCHAR(20),      -- e.g., northeast, northwest, southeast, southwest
    insurance_cost DECIMAL(10,2)
);
INSERT INTO medical_insurance VALUES
(1, 19, 'Female', 27.90, 0, 'Yes', 'southwest', 16884.92),
(2, 18, 'Male', 33.77, 1, 'No', 'southeast', 1725.55),
(3, 28, 'Male', 33.00, 3, 'No', 'southeast', 4449.46),
(4, 33, 'Male', 22.70, 0, 'No', 'northwest', 21984.47),
(5, 32, 'Female', 28.88, 0, 'No', 'southeast', 3866.86),
(6, 31, 'Female', 25.74, 0, 'Yes', 'southeast', 3756.62),
(7, 46, 'Female', 33.44, 1, 'No', 'southeast', 8240.59),
(8, 37, 'Male', 27.74, 2, 'No', 'northwest', 7281.51),
(9, 37, 'Female', 29.83, 2, 'No', 'northeast', 6406.41),
(10, 60, 'Female', 25.84, 0, 'Yes', 'northwest', 28923.14);
```

Tasks

1. Find the **total number of insurance records**.
2. Retrieve the details of all **smokers above age 40**.

3. Count how many **male and female policyholders** exist in the dataset.
4. Calculate the **average insurance cost** of smokers vs. non-smokers.
5. Find the **region with the highest average insurance cost**.
6. List the **top 5 policyholders who paid the highest insurance cost**.
7. Calculate the **average insurance cost grouped by age group**:
    - Young (18–30), Middle-aged (31–50), Senior (51+).
8. Identify if **smokers with BMI > 30** are paying significantly higher insurance than non-smokers.
9. Find the **correlation-like insights** (not actual correlation in SQL, but using GROUP BY) – e.g.,
    - Avg cost by number of children.
    - Avg cost by gender.
    - Avg cost by region.
10. Write a query to find the **insurance cost difference between smokers and non-smokers in each region**.