U.S. Medical Insurance Costs

In [2]:
```python
import csv  # Import the csv module to read CSV files

def load_data():
    # Open the insurance.csv file and read its contents
    # newline='' prevents extra blank lines on some systems
    with open("insurance.csv", newline='') as insurance_csv:
        # Convert each row into a dictionary and return a list of dictionaries
        return list(csv.DictReader(insurance_csv))

def charges_by_gender(data):
    # Initialize total charges and counters for each gender
    female_total_charge = 0
    male_total_charge = 0
    number_of_women = 0
    number_of_men = 0

    # Loop through each record in the dataset
    for dictionary in data:
        # Accumulate charges and counts for females
        if dictionary['sex'] == "female":
            female_total_charge += float(dictionary['charges'])
            number_of_women += 1
        # Accumulate charges and counts for males
        elif dictionary['sex'] == 'male':
            male_total_charge += float(dictionary['charges'])
            number_of_men += 1

    # Calculate average charges for each gender
    avg_female_charge = female_total_charge / number_of_women
    avg_male_charge = male_total_charge / number_of_men

    # Print total and average charges for women and men
    print(f'Women in total are charged ${female_total_charge:,.2f} and on average are charged '
          f'${avg_female_charge:,.2f}\nMen in total are charged ${male_total_charge:,.2f} and on average are '
          f'charged ${avg_male_charge:,.2f}')
```

```python
def charges_by_region(data):
    # Initialize total charges for each region
    southwest = 0
    southeast = 0
    northwest = 0
    northeast = 0

    # Loop through each record and add charges to the appropriate region
    for dictionary in data:
        if dictionary['region'] == 'southwest':
            southwest += float(dictionary['charges'])
        elif dictionary['region'] == 'southeast':
            southeast += float(dictionary['charges'])
        elif dictionary['region'] == 'northwest':
            northwest += float(dictionary['charges'])
        elif dictionary['region'] == 'northeast':
            northeast += float(dictionary['charges'])

    # Print total charges for each region
    print(f'Total Charges Southwest: ${southwest:,.2f}\n'
          f'Total Charges Southeast: ${southeast:,.2f}\n'
          f'Total Charges Northwest: ${northwest:,.2f}\n'
          f'Total Charges Northeast: ${northeast:,.2f}')

def number_of_smokers(data):
    # Dictionaries to track smokers and non-smokers by region
    smokers_southeast = {'Smokers': 0, "Non-Smokers": 0}
    smokers_southwest = {'Smokers': 0, "Non-Smokers": 0}
    smokers_northeast = {'Smokers': 0, "Non-Smokers": 0}
    smokers_northwest = {'Smokers': 0, "Non-Smokers": 0}

    # Loop through each record to count smokers and non-smokers by region
    for dictionary in data:
        if dictionary['smoker'] == 'yes':
            if dictionary['region'] == 'southeast':
                smokers_southeast['Smokers'] += 1
            elif dictionary['region'] == 'southwest':
                smokers_southwest['Smokers'] += 1
            elif dictionary['region'] == 'northeast':
                smokers_northeast['Smokers'] += 1
```

```
            elif dictionary['region'] == 'northwest':
                smokers_northwest['Smokers'] += 1
        elif dictionary['smoker'] == 'no':
            if dictionary['region'] == 'southeast':
                smokers_southeast['Non-Smokers'] += 1
            elif dictionary['region'] == 'southwest':
                smokers_southwest['Non-Smokers'] += 1
            elif dictionary['region'] == 'northeast':
                smokers_northeast['Non-Smokers'] += 1
            elif dictionary['region'] == 'northwest':
                smokers_northwest['Non-Smokers'] += 1

    # Print smoker and non-smoker counts for each region
    print(f'There are {smokers_southeast["Smokers"]} smokers and {smokers_southeast["Non-Smokers"]} non-smokers in
          f'southeast regions.\n'
          f'There are {smokers_southwest["Smokers"]} smokers and {smokers_southwest["Non-Smokers"]} non-smokers in
          f'southwest regions.\n'
          f'There are {smokers_northeast["Smokers"]} smokers and {smokers_northeast["Non-Smokers"]} non-smokers in
          f'northeast regions.\n'
          f'There are {smokers_northwest["Smokers"]} smokers and {smokers_northwest["Non-Smokers"]} non-smokers in
          f'northwest regions.')

def average_age_of_patients(data):
    # Initialize total age and patient counter
    total_age = 0
    number_of_patients = 0

    # Loop through each patient and sum their ages
    for patient in data:
        total_age += int(patient['age'])
        number_of_patients += 1

    # Calculate the average age
    avg_age = total_age / number_of_patients

    # Return a formatted string with the average age
    return f'The average age across all patients is roughly {round(avg_age)} years old'

def most_populated_region(data):
    # Dictionary to count population per region
    population_by_region = {'southeast': 0, 'southwest': 0, 'northeast': 0, 'northwest': 0}
```

```
    # Count how many people belong to each region
    for dictionary in data:
        if dictionary['region'] == 'southeast':
            population_by_region['southeast'] += 1
        elif dictionary['region'] == 'southwest':
            population_by_region['southwest'] += 1
        elif dictionary['region'] == 'northeast':
            population_by_region['northeast'] += 1
        elif dictionary['region'] == 'northwest':
            population_by_region['northwest'] += 1

    # Determine the region with the largest population
    largest_population = max(population_by_region, key=population_by_region.get)

    # Print the most populated region and its population count
    print(f'The most populated region is the {largest_population} region which has '
          f'{population_by_region[largest_population]} people')

def different_costs(data):
    # Initialize counters and total costs for smokers and non-smokers
    number_of_smokers = 0
    smokers_total_costs = 0
    number_of_non_smokers = 0
    non_smokers_total_costs = 0

    # Loop through each record and separate costs by smoking status
    for dictionary in data:
        if dictionary['smoker'] == 'yes':
            number_of_smokers += 1
            smokers_total_costs += float(dictionary['charges'])
        elif dictionary['smoker'] == 'no':
            number_of_non_smokers += 1
            non_smokers_total_costs += float(dictionary['charges'])
        else:
            continue

    # Calculate average costs
    avg_smoker_cost = smokers_total_costs / number_of_smokers
    avg_non_smoker_cost = non_smokers_total_costs / number_of_non_smokers
    avg_difference = avg_smoker_cost - avg_non_smoker_cost
```

```
    # Print cost comparison results
    print(f'\nThere are {number_of_smokers} smokers and {number_of_non_smokers:,.0f} non-smokers.\nSmokers altogeth
          f'are charged ${smokers_total_costs:,.2f} and non-smokers are charged ${non_smokers_total_costs:,.2f}.'
          f'\nThus, on average smokers are charged ${avg_smoker_cost:,.2f} per smoker whereas non-smokers are charg
          f'on average ${avg_non_smoker_cost:,.2f}.\nThat is a ${avg_difference:,.2f} average cost difference!')

def average_age_with_at_least_one_child(data):
    # Initialize total ages and counter for people with at least one child
    total_ages = 0
    number_of_people = 0

    # Loop through the data and include only those with children
    for dictionary in data:
        if dictionary['children'] != '0':
            total_ages += int(dictionary['age'])
            number_of_people += 1

    # Calculate the average age
    average_age = total_ages / number_of_people

    # Return a formatted string with the average age
    return f'Average Age: {average_age:.0f}'


# Load the dataset from the CSV file
data = load_data()

# Run each analysis function using the loaded data
charges_by_gender(data)
charges_by_region(data)
number_of_smokers(data)
print(average_age_of_patients(data))
most_populated_region(data)
different_costs(data)
print(average_age_with_at_least_one_child(data))
```

```
Women in total are charged $8,321,061.19 and on average are charged $12,569.58
Men in total are charged $9,434,763.80 and on average are charged $13,956.75
Total Charges Southwest: $4,012,754.65
Total Charges Southeast: $5,363,689.76
Total Charges Northwest: $4,035,712.00
Total Charges Northeast: $4,343,668.58
There are 91 smokers and 273 non-smokers in the southeast regions.
There are 58 smokers and 267 non-smokers in the southwest regions.
There are 67 smokers and 257 non-smokers in the northeast regions.
There are 58 smokers and 267 non-smokers in the northwest regions.
The average age across all patients is roughly 39 years old
The most populated region is the southeast region which has 364 people

There are 274 smokers and 1,064 non-smokers.
Smokers altogether are charged $8,781,763.52 and non-smokers are charged $8,974,061.47.
Thus, on average smokers are charged $32,050.23 per smoker whereas non-smokers are charged on average $8,434.27.
That is a $23,615.96 average cost difference!
Average Age: 40
```

In [ ]: