**Medical Insurance Cost Analysis**

**1. Project Overview**

This project analyzes medical insurance data from a Kaggle dataset to uncover insights into insurance charges, risk factors, and demographic patterns. The goal is to guide insurance policy design, cost reduction strategies, and targeted customer interventions.

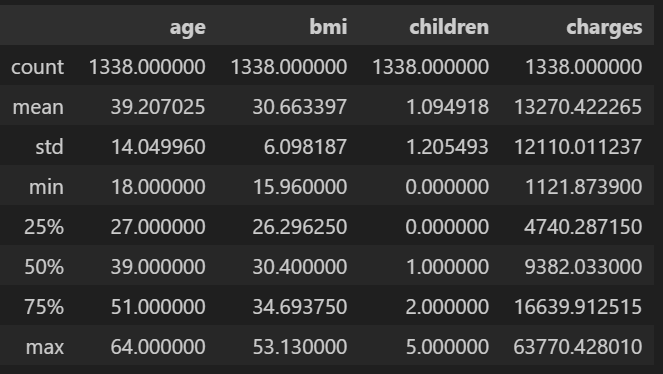
**2. Dataset Summary**

* **Rows:** 1,338
* **Columns:** 10
* **Key Features:**
  + Customer demographics: **age, sex, region, children, smoker**
  + Health metrics**: bmi, age\_group, bmi\_group**
  + Insurance financials: charges
* **Missing Data:** None after cleaning
* **Feature Engineering:**
  + Created **age\_group** (Young Adult, Adult, Middle-aged, Senior)
  + Created **bmi\_group** (Underweight, Normal, Overweight, Obese)

**3. Exploratory Data Analysis using Python**

We began with data preparation and cleaning in Python:

* **Data Loading:** Imported the dataset using pandas.
* **Initial Exploration:** Checked structure with df.info() and summary statistics with df.describe().



* **Missing Data Handling:** No missing values after preprocessing.
* **Feature Engineering:**
  + age\_group and bmi\_group for categorization.
  + Children categories (No/One Child, Few Children, Many Children).
* **Data Consistency Check:** Verified relationships between smoker status, BMI, and charges.
* **Database Integration:** Connected to MySQL using SQLAlchemy and loaded the cleaned DataFrame for SQL analysis.

A graph of blue and orange dots

AI-generated content may be incorrect.

A graph of different colored rectangular shapes

AI-generated content may be incorrect.

A graph of smoking status

AI-generated content may be incorrect.

A graph of different colored squares

AI-generated content may be incorrect.

A diagram of different colored squares

AI-generated content may be incorrect.

A graph of different colored squares

AI-generated content may be incorrect.

A graph of different colored squares

AI-generated content may be incorrect.

A graph of different colored squares

AI-generated content may be incorrect.

A diagram of a number of numbers

AI-generated content may be incorrect.

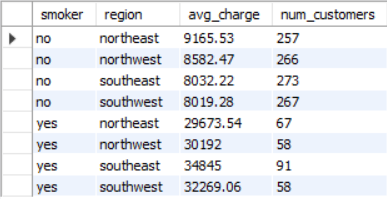
A graph of a distribution of medical charges

AI-generated content may be incorrect.

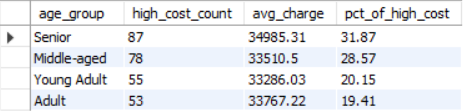
**4. Data Analysis using SQL (Insurance Metrics)**

**Key Questions and Insights**

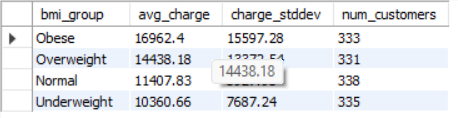
1. **Average Charges by Smoker Status and Region**



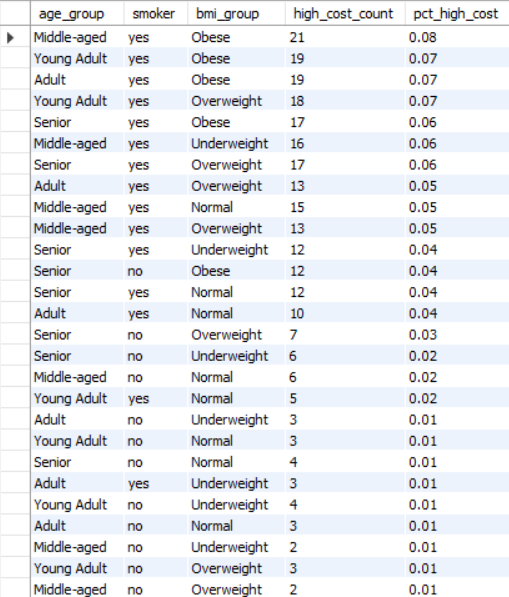
1. **Age Correlation with Charges & High-Cost Contributors**

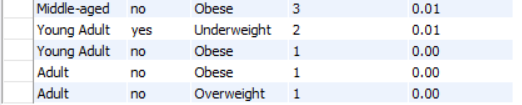


1. **BMI Categories Effect on Charges**



1. **High Charges (>20,000) by Age, Smoker, BMI**





1. **Projected Cost Reduction if Smoker Prevalence Drops 10%**

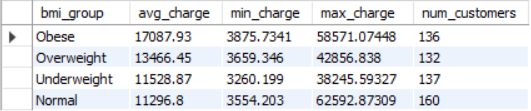


1. **Children Categories Effect on Charges by Region**

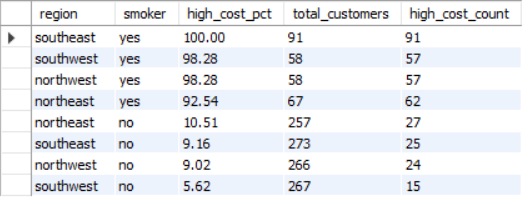
A screenshot of a table

AI-generated content may be incorrect.

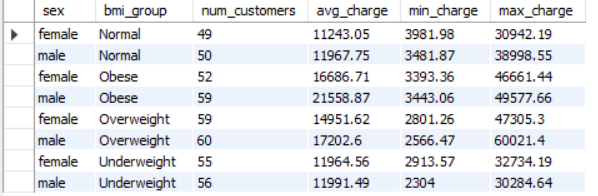
1. **Charges Trend Across BMI for Ages 30–50**



1. **High-Cost Claims by Region and Smoker Status**



1. **Charges by Gender within BMI for Households with ≥2 Children**



1. **Projected Reduction if Obese Customers Reduce BMI by 5%**



**5. Dashboard in Power BI**

* Visualizations created for:
  + Charges distribution by smoker, BMI, and age group
  + High-cost claim concentration per region
  + Children categories vs. average charges
* Filters for region, smoker, BMI group, children group and age group for dynamic exploration.

A screenshot of a insurance policy

AI-generated content may be incorrect.

A screenshot of a computer screen

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a report

AI-generated content may be incorrect.

**6. Business Recommendations**

* **Smoking Cessation Programs:** Promote health campaigns targeting smokers.
* **Targeted Wellness Programs:** Reduce BMI among high-risk groups.
* **Family-Based Interventions:** Consider children categories for risk-adjusted pricing.
* **Age & Region Segmentation:** Design policies focusing on Middle-aged/Senior, high-cost regions.
* **Data-Driven Pricing:** Adjust premiums based on demographic, BMI, and smoking status insights.