Health Insurance Claim Analysis Project (SQL + Power BI)

Objective

To analyze health insurance claims and identify key factors influencing claim amounts such as age, gender, BMI, blood pressure, smoking habits, diabetic status, and region using MySQL for data analysis and Power BI for visualization.

Dataset Overview

Source: Kaggle - *Insurance Claim Analysis: Demographic and Health* - Dataset Link (cleaned in Excel, analyzed in MySQL, visualized in Power BI)

Total Records: 1,340 people

Key Columns: Index, PatientID, Age, Gender, BMI, BloodPressure, Diabetic, Children, Smoker, Region, Claim

Data Cleaning (Excel + SQL)

Performed initial cleaning in Excel before loading into MySQL:

- Handled missing values in region & BMI.
- Removed duplicates and ensured consistent formatting.

In MySQL, created new derived columns:

- Age_Group (18–24, 25–34, 35–44, 45–54, 55+)
- BMI_Group (Underweight, Normal, Overweight, Obese)
- BP_Category (Normal, Pre-High, High)
- High_Risk_Flag = Smoker + Diabetic + Obese

SQL Analysis Highlights

Metric	Insight
Average Claim	₹13.25K
Total Claim	₹17.76M
Total People	1,340

High-Risk Individuals	69 people with Avg. Claim ₹41.13K
Gender Trend	Males: ₹13.9K avg. vs Females: ₹12.5K avg.
Top Region	Northeast – ₹16.8K avg.
Health Risks	Smokers & Obese individuals drive claim costs
Claim Range Distribution	Most people (<₹10K), few high-claim outliers >₹30K

Power BI Dashboard Overview

Dashboard KPIs:

- Average Claim
- Total Claim
- Total People
- High-Risk Count
- High-Risk Average Claim

Key Visuals

Visual	Chart Type	Insights
Average Claim by Region	Pie Chart	Northeast region has the highest claim average (30%)
Claim by Age & Gender	Clustered Column	Males (35–44) lead in claim total; females steady across ages
Average Claim by BP Category	Column Chart	High BP → highest avg. claim (₹23K)
Claim by Diabetic Status	Donut Chart	Non-diabetics file slightly higher claims

Average Claim by BMI Group	Bar Chart	Obese → ₹15.5K avg., Underweight → lowest
Smoker vs Non-Smoker	Table Card	Smokers: ₹32K avg. vs Non-smokers: ₹8.4K

Slicers Used: BMI Group, Region, Gender, Smoker, Diabetic

Key Findings

- Smokers have 3.8× higher claims than non-smokers.
- Obese and High BP individuals show highest cost risk.
- 69 high-risk individuals contribute significantly to claim costs.
- Northeast region dominates total claims, possible lifestyle or cost factor.
- Majority of claimants (70%) file under ₹10K, showing skewed distribution.

Business Insights & Recommendations

1. Premium Segmentation: Introduce dynamic premiums based on health risk factors (BMI, smoking, diabetic).
2. Preventive Health Programs: Target obesity and BP management to reduce future claim loads.
3. Regional Strategy: Focus health campaigns in the Northeast region to lower average claims.
4. High-Risk Monitoring: track high-risk individuals monthly for early interventions.

Tools Used

Tool	Purpose
Excel	Data cleaning and preprocessing
MySQL	Data transformation & statistical queries
Power BI	Dashboard and visualization
GitHub	Project documentation & sharing

Conclusion

This project demonstrates how SQL-based data analysis combined with Power BI storytelling can uncover actionable health insights. The dashboard provides a comprehensive overview of claim behavior, enabling insurers to make data-driven policy decisions and improve cost efficiency through early risk detection.