A Comprehensive Data Analysis on Medical Expenses

Medical Cost Analysis Presentation

Vinit Solanki

Click here to watch video

Basic Statistics And Data Understanding

Data Columns	AVG	Median	STDEV
Charges	13279.12	9386.16	12110.36
BMI	30.66	30.40	6.10
Age	39.22	39.00	14.04

- **Overview of Key Metrics**
  To gain an initial understanding of the dataset, we analyzed key statistical measures for **BMI**, **Age** and **Medical Charges**. These metrics offer insights into the **distribution** and **variability** of health-related factors across individuals.
- **Average Values and Trends**
  The **average BMI** is **30.66**, indicating a trend toward **overweight or obesity** among individuals in the dataset. The **average age** is **39.21**, showing a mix of **younger and older** individuals. Additionally, the **average medical charges** amount to **$13,270.42**, reflecting **significant healthcare expenses**.
- **Central Tendency and Dispersion**
  The **median values** for **BMI (30.4)**, **age (39)**, and **charges ($9,382.03)** provide insights into the **center of the data distribution**. The **standard deviation** for **BMI (6.10)**, **age (14.05)**, and **medical charges ($12,110.01)** indicate **considerable variation**, particularly in medical costs where some individuals incur **substantially higher expenses**.
- **Key Observations**
  The **high standard deviation in medical charges** suggests **significant disparities in healthcare costs**, likely influenced by factors such as **chronic conditions**, **insurance coverage**, or **hospitalization needs**. The **relatively close values of average and median age** indicate a **balanced distribution** across age groups. **BMI levels** further emphasize the importance of addressing **weight-related health risks** in the population.

Gender-Based Analysis

- **Objective**
  To compare the **average medical charges** and **BMI** between **male** and **female** beneficiaries.

- **Method**
  Used AVERAGEIF and Pivot Table functions in Excel to calculate the **average BMI** and **average medical charges** for each gender.

- **Findings**
- **Average BMI (Male)**: **30.94**
- **Average BMI (Female)**: **30.38**
- **Average Charges (Male)**: **$13,975.00**
- **Average Charges (Female)**: **$12,569.58**

- **Insights**
- **Males** have both **higher average BMI** and **higher average medical charges** compared to females.
- The difference suggests a **potential correlation** between gender, BMI, and healthcare costs.

Row Labels	Average of BMI	Average of charges
female	30.38	12569.58
male	30.94	13975.00
Grand Total	30.66	13279.12

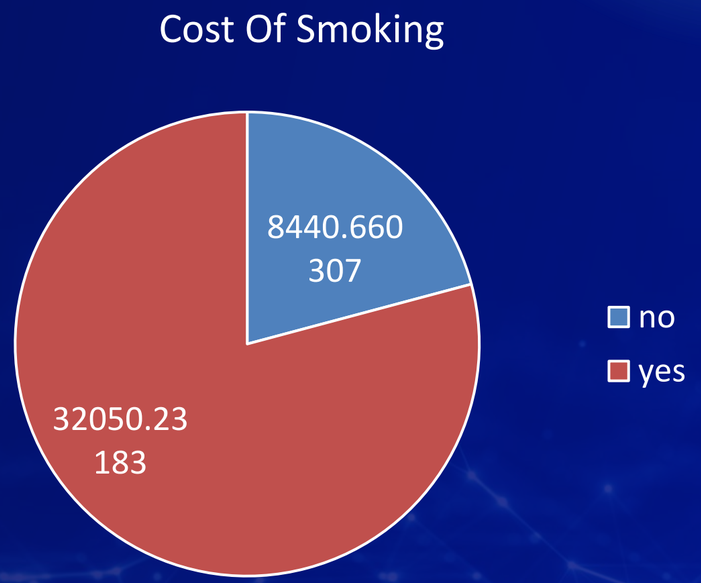
Impact of Smoking on Medical Costs

- **Objective**
  To evaluate how **smoking status** affects **average medical charges** among beneficiaries.

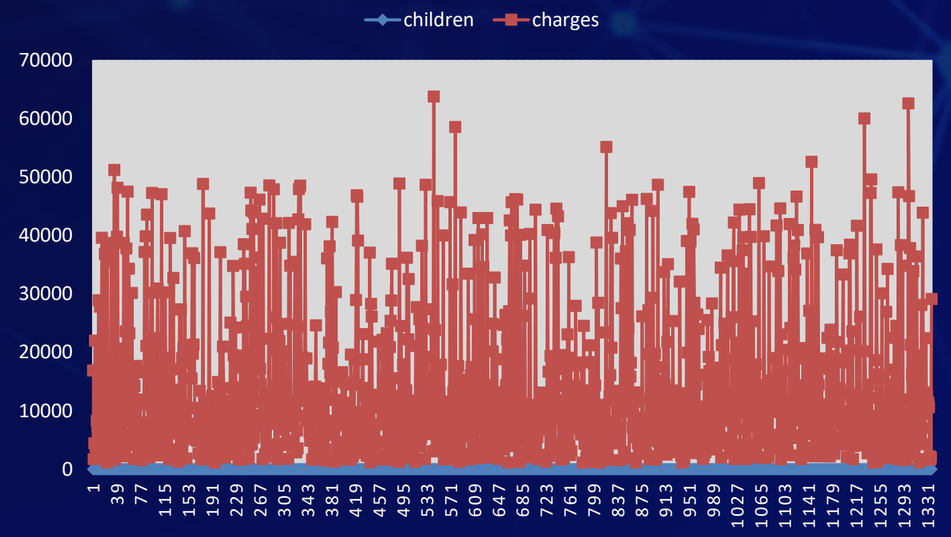
- **Method**
  Used AVERAGEIFS and Pivot Table in Excel to calculate **average charges** separately for **smokers** and **non-smokers**.

- **Findings**
- **Average Charges (Smokers)**: **$32,050.23**
- **Average Charges (Non-Smokers)**: **$8,440.66**
- **Insights**

- **Smokers incur nearly 4 times higher** average medical costs compared to non-smokers.
- This significant cost difference highlights the **financial burden of smoking-related health issues**, making smoking a **major risk factor** in healthcare expenses.



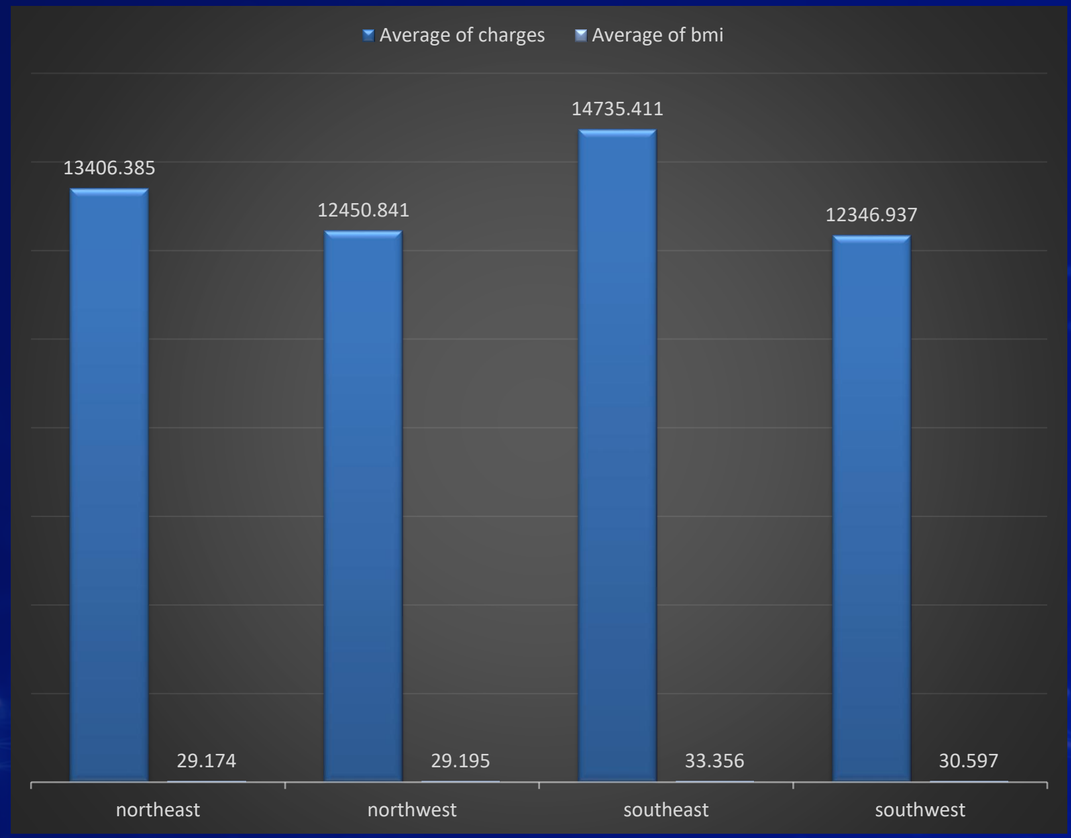
Dependents Analysis



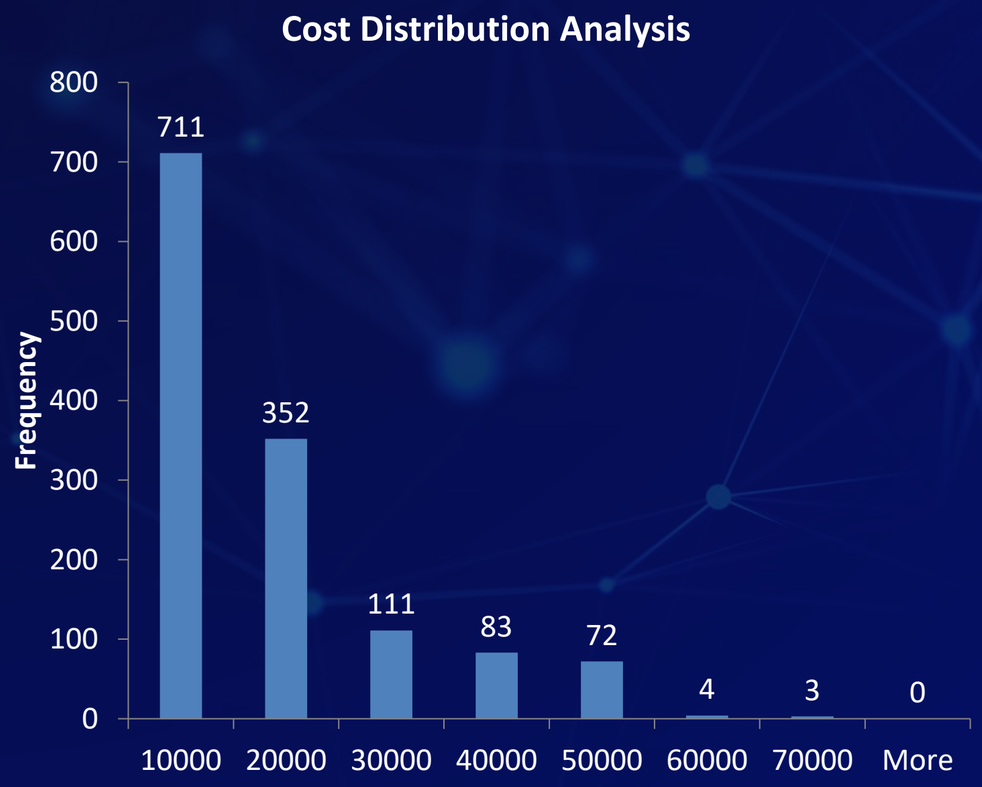
- **Objective**
  To evaluate whether the **number of children or dependents** influences **medical charges** among beneficiaries.
- **Method**
  Used Excel's CORREL function to measure the statistical relationship between **children** and **charges**. Created a **line chart** to visualize this trend.
- **Findings**
- **Correlation coefficient** between number of children and medical charges is **0.067**, indicating a **very weak positive relationship**.
- The chart shows **no clear upward or downward trend**, with medical charges fluctuating widely across all dependent levels.
- **Insights**
- There is **no strong evidence** that having more children significantly increases medical costs.
- While a slight positive trend exists, it is **statistically insignificant**, suggesting that other factors (like smoking status, age, or chronic illness) may have a greater influence on charges.

Geographical Insights

- **Objective**
  To identify which **U.S. region** has the **highest average medical charges** and **BMI** among beneficiaries.
- **Method**
  Used Excel Pivot Table and AVERAGEIFS to calculate **average charges** and **BMI** by **region**. A **bar chart** was created for visual comparison.
- **Findings**
- **Southeast** region has the **highest average medical charges** at **$14,735.41**
- It also records the **highest average BMI** of **33.36**
- **Southwest** has the **lowest average charges** at **$12,346.94**
- **Insights**
- The **Southeast** region stands out with both **elevated BMI** and **higher healthcare costs**, indicating a **potential regional health risk pattern**.
- This suggests that **geographic location** may influence both **lifestyle** and **medical expenditures**, which could be useful for **targeted public health interventions**.



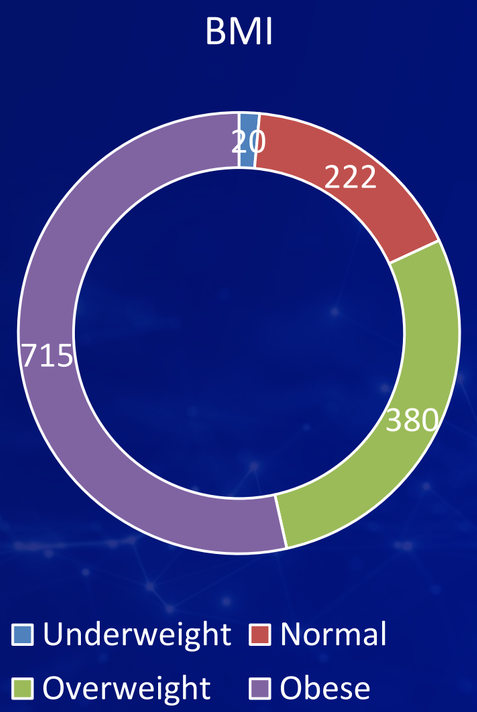
Cost Distribution Analysis



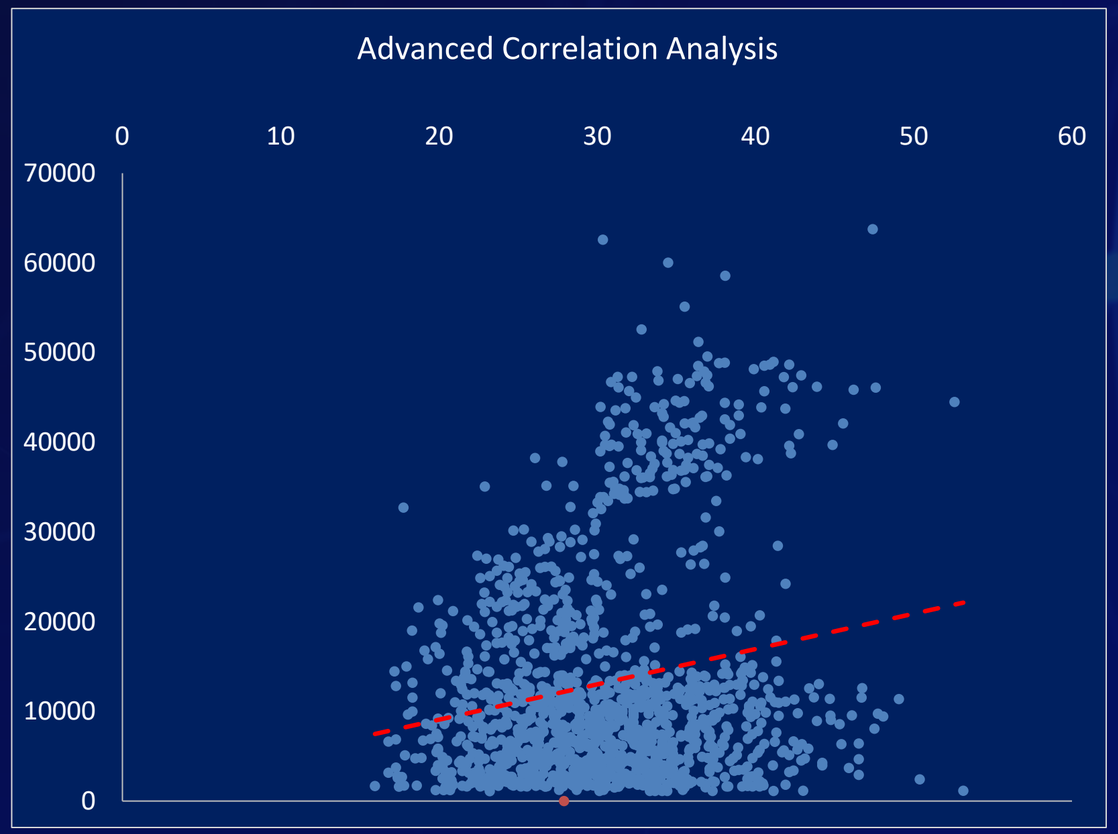
- **Objective**
  To understand how **medical charges** are distributed across beneficiaries and identify the **most common cost bracket**.
- **Method**
  Used Excel's **Histogram** tool via the **Data Analysis Toolpak** along with the FREQUENCY function to group charges into defined **bin ranges**.
- **Findings**
- The **most common cost bracket** is **$0–10,000**, with **711 beneficiaries** falling in this range.
- Frequency decreases significantly as charges increase:
  - **$10,001–20,000**: 352
  - **$20,001–30,000**: 111
  - **Above $30,000**: progressively fewer cases, only **3** above $70,000
- **Insights**
- The data is **right-skewed**, with most beneficiaries incurring **lower medical expenses**.
- A **small number** of high-cost individuals create a **long tail** in the distribution, likely due to **critical illnesses or hospitalization**.

BMI Classification

- **Objective**
  To categorize beneficiaries based on their **BMI values** and evaluate the **distribution of weight categories** within the dataset.
- **Method**
  Applied Excel's COUNTIFS and VLOOKUP with **range-based lookup** to classify BMI into four categories: **Underweight, Normal, Overweight,** and **Obese**.
- **Findings**
- **Underweight (<18.5)**: **20 beneficiaries**
- **Normal (18.5–24.9)**: **225 beneficiaries**
- **Overweight (25–29.9)**: **386 beneficiaries**
- **Obese (30+)**: **706 beneficiaries**
- **Insights**
- A **majority (706 individuals)** fall into the **Obese** category, followed by those who are **Overweight**.
- Only a **small portion** of the population is classified as **Underweight or Normal**, indicating a **potential public health concern** related to **high BMI levels**.
- These insights may support initiatives focused on **obesity prevention and management**.



Advanced Correlation Analysis



- **Objective**
  To assess the **relationship between BMI and medical charges** using a **scatter plot** and **trendline**.
- **Method**
  Created a **scatter plot** in Excel with **BMI on the x-axis** and **medical charges on the y-axis**. Added a **linear trendline** and calculated correlation using the CORREL function.
- **Findings**
- **Correlation coefficient** between **BMI and medical charges** is **0.198**, indicating a **weak positive relationship**.
- The **trendline** shows a **slight upward slope**, suggesting that higher BMI is **generally associated** with increased medical costs.
- **Insights**
- While the relationship is **not strong**, there is some evidence that **higher BMI levels** may contribute to **elevated medical expenses**.
- This supports the need for **preventive healthcare** strategies targeting **weight-related health risks**.

Smoking and Region Interaction

Sum of charges	Column Labels
Row Labels	yes
northeast	1988126.944
northwest	1751136.185
southeast	3170894.711
southwest	1871605.683
Grand Total	8781763.522

- **Objective**
  To explore how **smoking status and geographical region** together impact **medical charges**, and determine which region incurs the **highest smoker-related costs**.
- **Method**
  Created a **Pivot Table** in Excel with **region** as row labels and **smoking status** as a filter. Used AVERAGEIFS and **slicers** to isolate charges for **smokers only**.
- **Findings**
- **Southeast region** shows the **highest total medical charges** for smokers at **$3,170,894.71**
- Other regions rank as follows:
  - **Northeast**: $1,988,126.94
  - **Southwest**: $1,871,605.68
  - **Northwest**: $1,751,136.19
- **Insights**
- Smokers in the **Southeast** region account for a **significantly larger share** of healthcare expenses.
- This may reflect **regional health behaviors, access to care**, or **prevalence of smoking-related conditions**.
- Insights like these can help target **regional anti-smoking campaigns** or **healthcare resource planning**.

Dashboard for Stakeholders

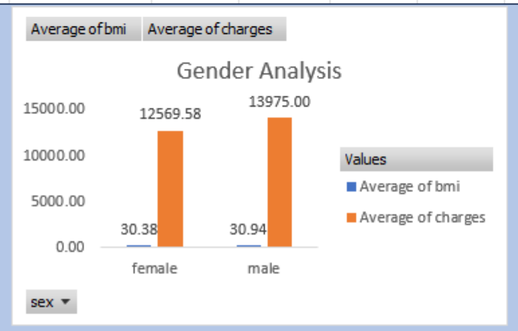
MEDICAL COST INSIGHT DASHBOARD

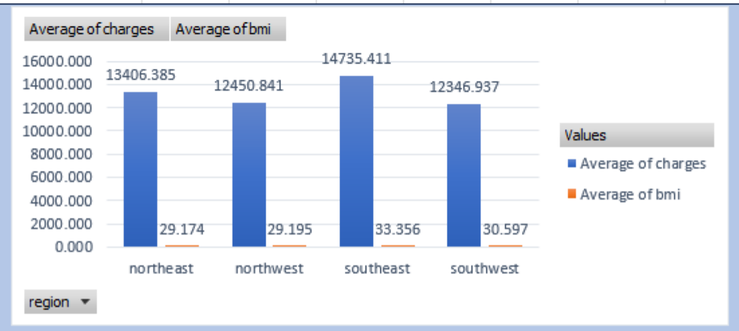
Data Columns	AVG	Median	Stdev
Charges	13279.12	9386.1613	12110.36
BMI	30.66345	30.4	6.100468
Age	39.22214	39	14.04433

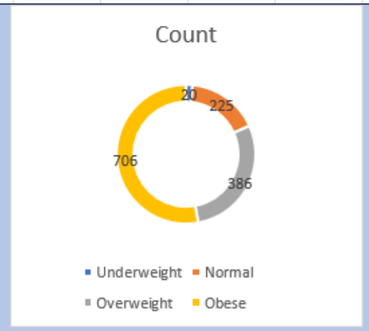
BMI categories	Range	Count
Underweight	<18.5	20
Normal	18.5-24.9	225
Overweight	25-29.9	386
Obese	30+	706

sex: female, male

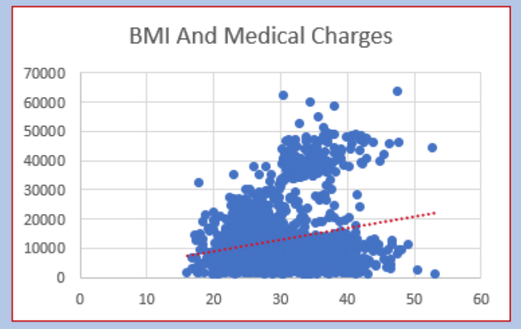
smoker: no, yes

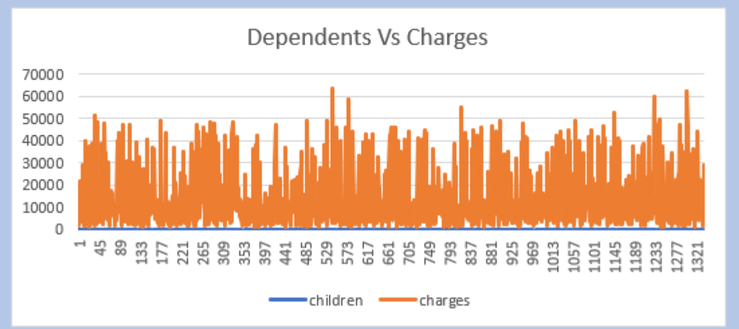


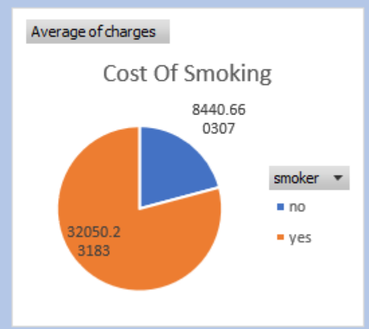




BMI categories: Normal, Obese, Overweight, Underweight







region: northeast, northwest, southeast, southwest

Thank you

Data Speaks. Insights Lead.