Medical Cost of Insurance Beneficiaries Analysis

Project Introduction:

This project focuses on analyzing the medical insurance costs of beneficiaries. I have examined various factors such as the total number of beneficiaries, categorized by the number of children, smoking status, region, and gender. The goal is to identify patterns and correlations that influence medical costs across these different demographics.

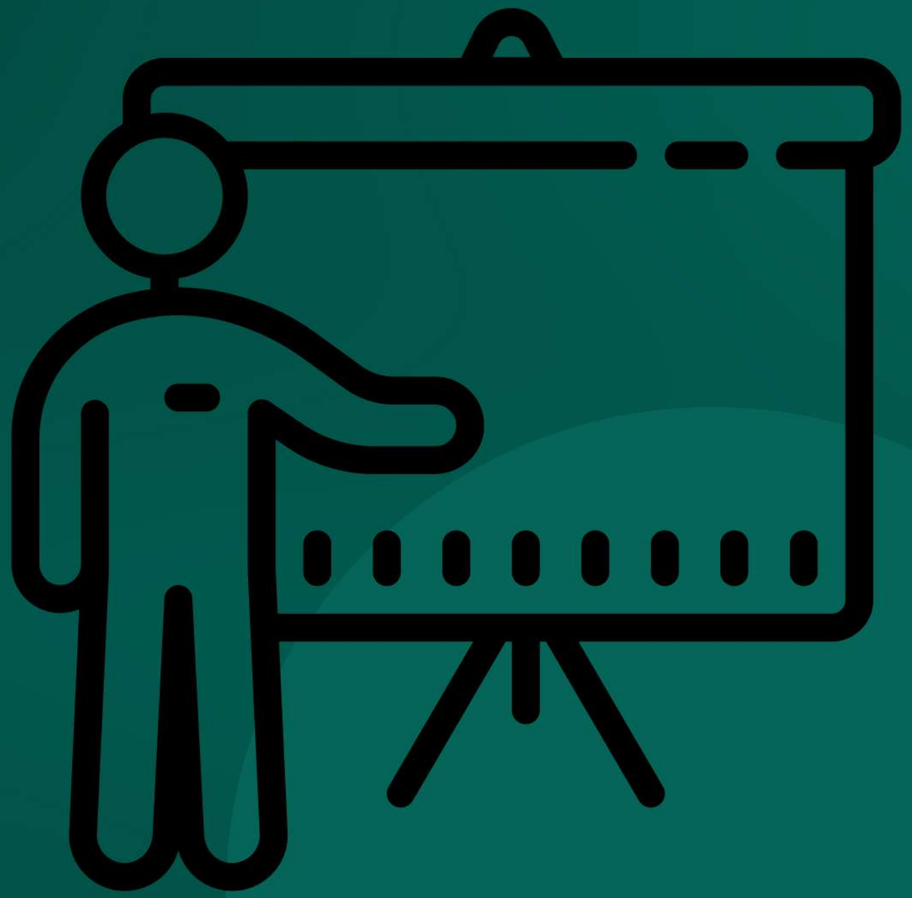
Check My Video Presentation:

YouTube **Click here to watch a detailed walkthrough of this project**
https://youtu.be/t4RYnjg7eLQ

Sumit Rathee

sumit.me1997@gmail.com

About Data



Here's a description of the data in your project:

The dataset used in this project contains the following key attributes:

- Age: The age of the primary insurance beneficiary.
- Sex: The gender of the insurance holder, categorized as either female or male.
- BMI (Body Mass Index): A measure that provides an understanding of whether an individual's body weight is relatively high or low compared to their height. It is calculated as weight in kilograms divided by the square of height in meters (kg/m^2). The ideal BMI range is between 18.5 and 24.9.
- Children: The number of children or dependents covered under the health insurance policy.
- Smoker: Indicates whether the beneficiary is a smoker.
- Region: The residential area of the beneficiary in the United States, categorized into four regions: Northeast, Southeast, Southwest, and Northwest.
- Charges: The individual medical costs billed to the beneficiary by health insurance.

Analysis Breakdown:

I divided this analysis into 10 key questions, using Excel functions and charts to visualize insights and address each aspect of the data.

1. Basic Statistics and Data Understanding
   - Objective: Summarize medical charges, BMI, and age using basic statistics.
   - Approach: Calculate averages, medians, and standard deviations.
2. Gender Analysis
   - Objective: Compare medical costs and BMI between genders.
   - Approach: Use AVERAGEIF and pivot tables to identify differences.
3. Impact of Smoking on Costs
   - Objective: Assess how smoking influences medical costs.
   - Approach: Compare average charges for smokers vs. non-smokers using pivot tables.
4. Dependents Analysis
   - Objective: Analyze the impact of dependents on medical costs.
   - Approach: Use correlation and scatter plots to explore the relationship.
5. Geographical Insights
   - Objective: Identify the region with the highest medical costs and BMI.
   - Approach: Create pivot tables and bar charts to visualize regional data.
6. Cost Distribution Analysis
   - Objective: Visualize the distribution of medical charges.
   - Approach: Create a histogram to identify common cost brackets.
7. BMI Classification
   - Objective: Categorize beneficiaries by BMI groups.
   - Approach: Use COUNTIFS and VLOOKUP for classification.
8. Advanced Correlation Analysis
   - Objective: Explore the relationship between BMI and medical charges.
   - Approach: Use scatter plots with trendlines to analyze correlation.
9. Smoking and Region Interaction
   - Objective: Examine the cost impact of smoking across regions.
   - Approach: Use pivot tables and slicers to compare costs by region.
10. Dashboard Creation for Stakeholders
    - Objective: Develop an interactive dashboard for medical cost insights.
    - Approach: Use pivot charts and slicers for a dynamic presentation.

1. Basic Statistics and Data Understanding

- Objective: Summarize medical charges, BMI, and age using basic statistics.
- Approach: Calculate averages, medians, and standard deviations.

Note: I use column names for cell references as they are easier to understand and make the formulas more dynamic and attractive.

=AVERAGE(age)
=MEDIAN(age)
=STDEV.P(age)

Functions	AGE	BMI	CHARGES
Average	39.21	30.66	13270.42
Median	39.00	30.40	9382.03
Sranderd Deviation	14.04	6.10	12105.48

2. Gender Analysis

- Objective: Compare medical costs and BMI between genders.
- Approach: Use AVERAGEIF and pivot tables to identify differences.

=AVERAGEIF(sex,E6,charges)

Row Labels	Average of bmi	Average of charges
female	30.37774924	12569.57884
male	30.9431287	13956.75118
Grand Total	**30.66339686**	**13270.42227**

Gender	Charges	BMI
female	12569.58	30.37775
male	13956.75	30.94313

Male gender has higher average costs and BMI.

3. Impact of Smoking on Costs

- Objective: Assess how smoking influences medical costs.
- Approach: Compare average charges for smokers vs. non-smokers using pivot tables.

=AVERAGEIFS(charges,smoker,D8)

Smoker	average
yes	32050.23
no	8434.268

Row Labels	Average of charges
no	8434.27
yes	32050.23
Grand Total	**13270.42227**

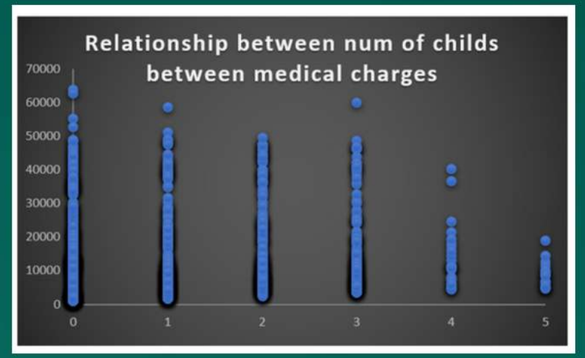
4. Dependents Analysis

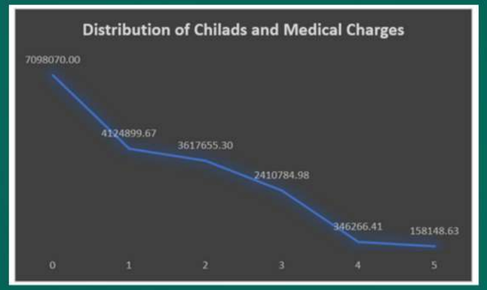
- Objective: Analyze the impact of dependents on medical costs.
- Approach: Use correlation and scatter plots to explore the relationship.

=CORREL(child,charges)
=SUMIF(child,B10,charges)

With correl formula
0.067998

Childs	chargis
0	7098070.00
1	4124899.67
2	3617655.30
3	2410784.98
4	346266.41
5	158148.63





:- The near-zero correlation (0.068) indicates that the number of children has little to no impact on medical charges.

:- A line chart verification shows that as the number of children increases, medical costs slightly decrease, indicating a weak inverse relationship.

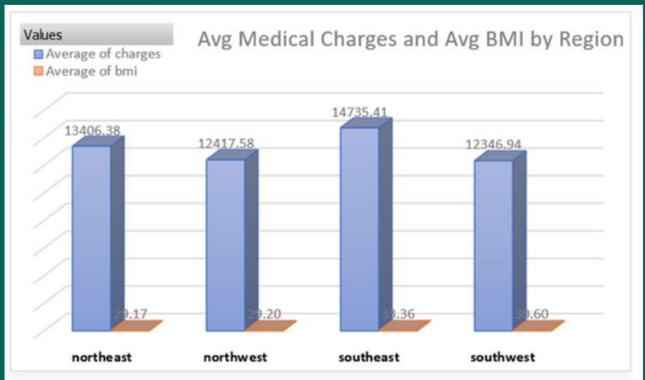
5. Geographical Insights

- Objective: Identify the region with the highest medical costs and BMI.
- Approach: Create pivot tables and bar charts to visualize regional data.

=AVERAGEIFS(charges,region,B6)

Region	avg of Charges	Avg of BMI
southwest	12346.94	30.60
southeast	14735.41	33.36
northwest	12417.58	29.20
northeast	13406.38	29.17

Row Labels	Average of charges	Average of bmi
northeast	13406.38	29.17
northwest	12417.58	29.20
southeast	14735.41	33.36
southwest	12346.94	30.60
Grand Total	**13270.42**	**30.66**



:- The Southeast region has the highest medical costs and BMI, while the Southwest has the lowest.

:- Regions with higher BMIs tend to have higher medical charges.

6. Cost Distribution Analysis

- Objective: Visualize the distribution of medical charges.
- Approach: Create a histogram to identify common cost brackets.

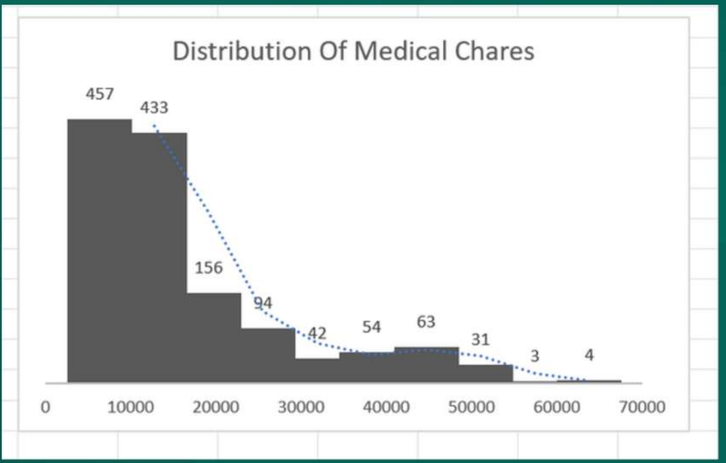
=FREQUENCY(charges,B7:B21)

I am using the Data Analysis Tool in Excel to create a frequency distribution and histogram.

✦

- **Most medical charges cluster below 20,000, with a sharp decline as charges increase.**
- **High-cost cases above 50,000 are rare, reflecting a skewed distribution.**

Bins	Frequency	Bins	Frequency
6378	457	6378	457
12756	433	12756	433
19134	157	19134	156
25512	94	25512	94
31890	42	31890	42
38268	54	38268	54
44646	63	44646	63
51024	31	51024	31
57402	3	57402	3
63780	4	63780	4
	0	More	0



7. BMI Classification

- Objective: Categorize beneficiaries by BMI groups.
- Approach: Use COUNTIFS and IF for classification.

=IF(C2<18.5,"Underweight",IF(AND(C2>=18.5,C2<=24.9),"Normal"
IF(AND(C2>=25,C2<=29.5),"Overweight","Obese")))

=COUNTIFS(category,A12)

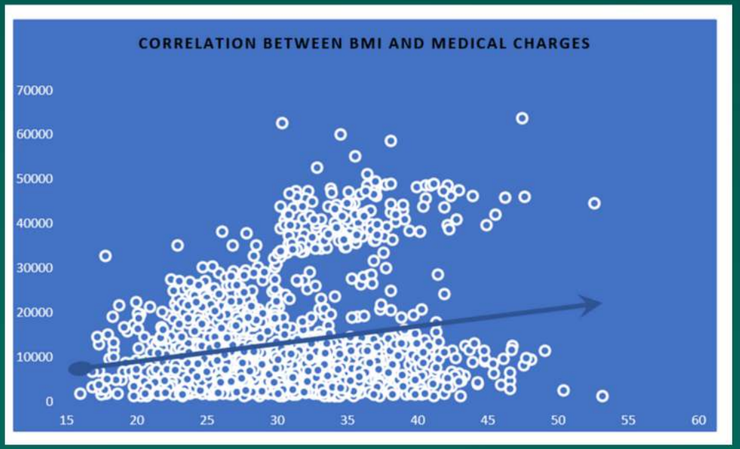
With Countifs	
Category	Counts
Overweight	339
Obese	757
Normal	222
Underweight	20

8. Advanced Correlation Analysis

- Objective: Explore the relationship between BMI and medical charges.
- Approach: Use scatter plots with trendlines to analyze correlation.

=CORREL(charges,bmi)

Correl
0.198341



- **Higher BMI leads to higher medical costs, showing a clear upward trend.**
- **Medical charges vary more widely for individuals with a higher BMI.**

9. Smoking and Region Interaction

- Objective: Examine the cost impact of smoking across regions.
- Approach: Use pivot tables and slicers to compare costs by region.

=UNIQUE(region)

=AVERAGEIFS(charges,region,$Y11,smoker,LOWER(Z$10))

With Averageifs

Region/Smoker	Yes	No
southwest	32269.06	8019.28
southeast	34845.00	8032.22
northwest	30192.00	8556.46
northeast	29673.54	9165.53

With Pivot Table

yes | no

northeast | northwest | southeast | southwest

Average of charges	Column Labels	
Row Labels	no	yes
northeast	9165.53	29673.54
northwest	8556.46	30192.00
southeast	8032.22	34845.00
southwest	8019.28	32269.06
Grand Total	**8434.268**	**32050.232**

10. Dashboard Creation for Stakeholders

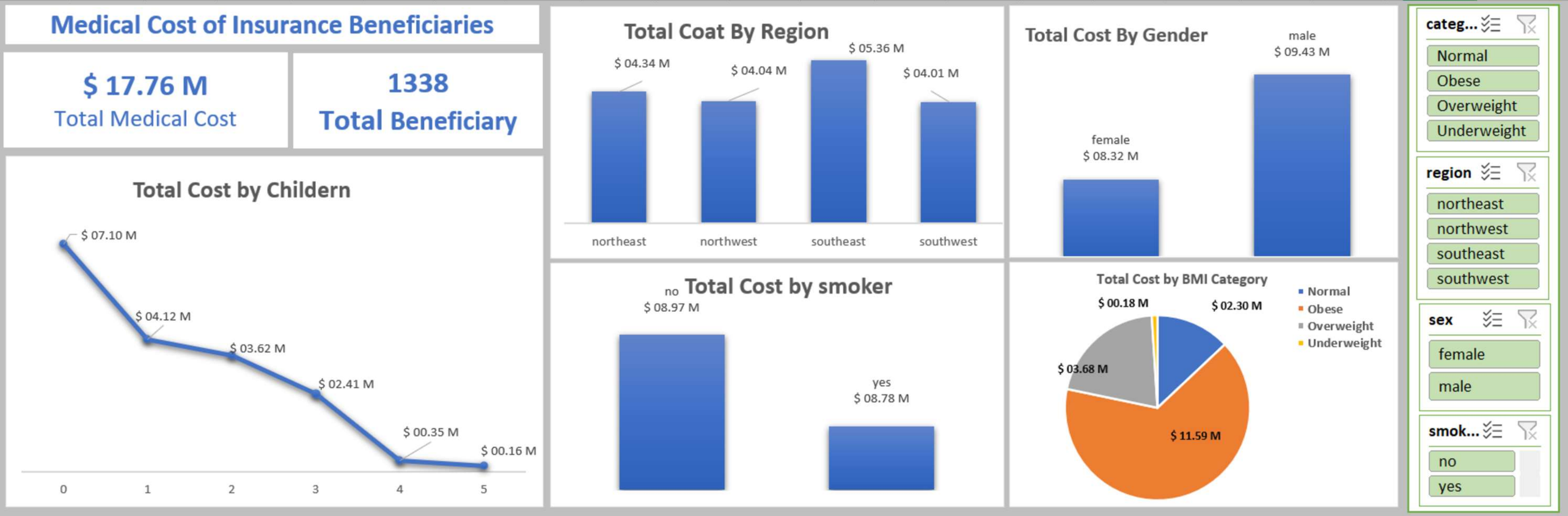
- Objective: Develop an interactive dashboard for medical cost insights.
- Approach: Use pivot charts and slicers for a dynamic presentation.

Row Labels	Sum of charges
no	$ 08.97 M
yes	$ 08.78 M
Grand Total	**$ 17.76 M**

Row Labels	Sum of charges
0	$ 07.10 M
1	$ 04.12 M
2	$ 03.62 M
3	$ 02.41 M
4	$ 00.35 M
5	$ 00.16 M
Grand Total	**$ 17.76 M**

Row Labels	Sum of charges
Normal	$ 02.30 M
Obese	$ 11.59 M
Overweight	$ 03.68 M
Underweight	$ 00.18 M
Grand Total	**$ 17.76 M**

Row Labels	Sum of charges
Normal	$ 02.30 M
Obese	$ 11.59 M
Overweight	$ 03.68 M
Underweight	$ 00.18 M
Grand Total	**$ 17.76 M**



Insights from the Dashboard:

- Cost by Region: The Southeast region incurs the highest medical costs at $5.36 million, indicating a need for targeted healthcare strategies in this area.
- Gender Disparity in Costs: Male beneficiaries have higher medical costs ($9.43 million) compared to females ($8.32 million).
- Impact of Smoking: Smokers contribute to nearly half of the total medical costs, underscoring the financial burden of smoking-related health issues.
- BMI and Medical Costs: The 'Obese' BMI category accounts for the highest medical costs at $11.59 million, suggesting a strong link between obesity and healthcare expenses.
- Children and Cost: Families with no children or one child have the highest medical costs, which drastically decrease with more children.

Suggestions:

- Focus on High-Risk Groups: Males and smokers should be the focus of intensive health education and preventive care to mitigate future medical expenses.
- Family Planning and Healthcare: Explore why costs decrease with more children and consider incentives or programs for smaller families to manage healthcare effectively
- Preventive Measures for Obesity: Given the high costs associated with obesity, consider rolling out preventive healthcare programs focusing on weight management and healthy living.

Thank You!

https://github.com/sumit-me-97/Hospital-Cost-Of-Insurance

https://youtu.be/t4RYnjg7eLQ

https://www.linkedin.com/in/sumit-rathee-73049b196/