MARCH 2024

ANALYSIS OF INSURANCE COMPETITORS

A SUMMARY OF THE PERFORMANCE OF OUR COMPETITOR INSURANCE COMPANY.

Prepared by: Saeed Shirani
Business Intelligence Analyst

ATTENTION!

- The dataset exhibits several limitations. Therefore, my report solely reflects observations without proposing any strategies or delving into the causes of each condition.

- To formulate effective strategies and identify the root causes of these phenomena, additional data is necessary.

- In the conclusion of the report, I have outlined some of these limitations and suggested that if these constraints were addressed, we could consider potential decisions and strategies.

- Moreover, with access to more comprehensive data, we would be better equipped to uncover the underlying causes of each observed phenomenon.

02

IMPORTANCE OF COMPETITOR ANALYSIS

Competitor analysis is vital for strategic planning, offering insights into market trends and consumer behavior. By assessing competitors' strengths and weaknesses, businesses refine their strategies and mitigate risks. It identifies market gaps for gaining a competitive edge and enhances overall competitiveness. Vigilant and proactive analysis enables businesses to adapt and succeed in dynamic markets. In summary, competitor analysis is essential for businesses to sustain success amidst competition.

Attention

ANALYSIS OBJECTIVES

Our dataset encompasses variables including **age, sex, BMI, number of children, smoking status, region, and payment amoun**t. Initially, I will segment the dataset based on each parameter. Subsequently, I will analyze and discuss our competitors' situations within each segment. Following this, I will address the limitations of our dataset and propose potential areas for further study if additional competitor data were to be successfully collected.

03

DATA SEGMENTATION

This is the initial Dataset

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520
5	31	female	25.740	0	no	southeast	3756.62160

This is the segmented Dataset

	age	sex	bmi	children	smoker	region	charges	age_range	bmi_condition
0	19	female	27.90	0	yes	southwest	16884.9240	18-30	over_weight
1	18	male	33.77	1	no	southeast	1725.5523	18-30	over_weight
2	28	male	33.00	3	no	southeast	4449.4620	18-30	over_weight

SEGMENTS:

Our segments encompass the following values:

SEX: Male, Female
CHILDREN : 0, 1, 2, 3, 4, 5
SMOKER: Yes, No
REGION: 'southwest', 'southeast', 'northwest', 'northeast'
AGE Range: 18-30|31-40|41-50|51-60|61-70|71-80|81-90
BMI_Condition :over_weight, healthy, under_weight

04

Methodology

Our main goal is to identify the most valuable segments by evaluating their total revenue and then determining segments with the highest customer count. Additionally, we'll track the Key Performance Indicator (KPI) of Revenue per Customer for each segment.

Here's a list of comparisons that I found interesting:

- **Gender-based Purchasing Value**
- **Purchasing Value categorized by Gender and Smoking Behavior**
- **Number of Smoker Customers by Age**
- **Customer Behavior Analysis Based on Smoking Habit and Number of Children**
- **Age Segmentation of Customers Based on Smoking Habit and Age**
- **Customer Analysis by Region and BMI Condition**

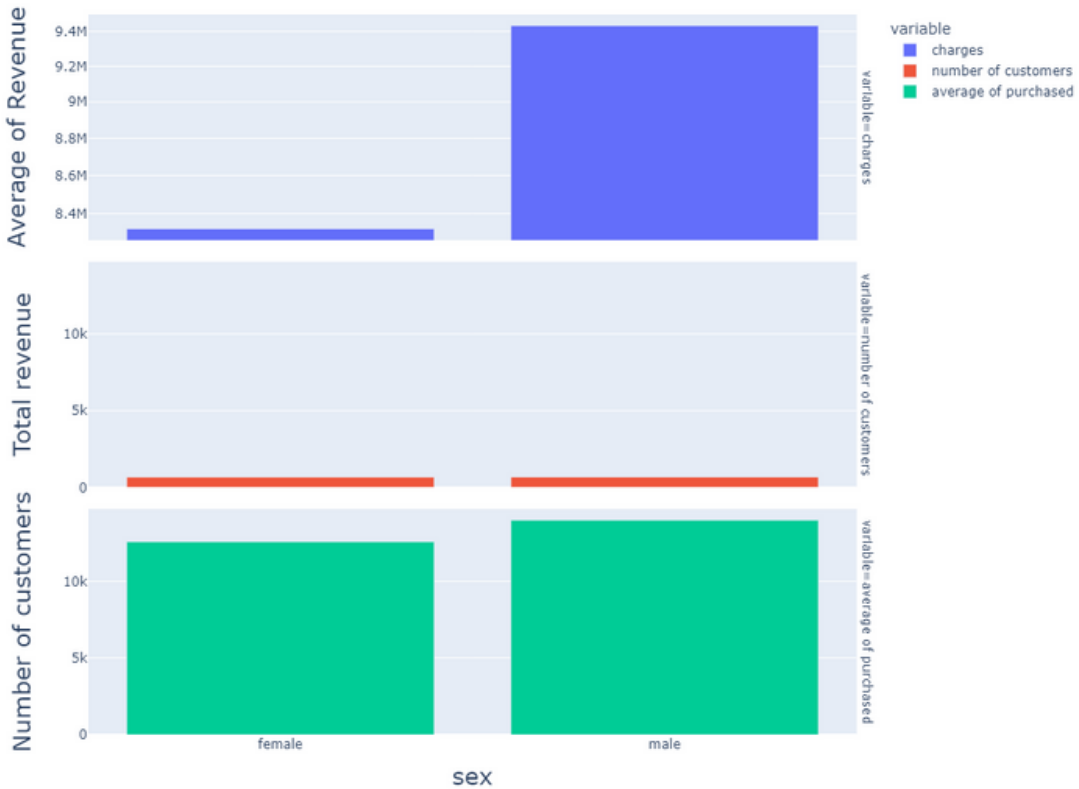
Ultimately, we'll conclude by identifying the most valuable age intervals based on the value they contribute to the company through purchases.

05

VALUE PURCHASED BY GENDER

sex	charges	number of customers	average of purchased
female	8.321061e+06	662	12569.578844
male	9.433124e+06	675	13974.998864

The Pusrchsed Value of each Gender is shown in the table above. there is no significant difference between the number of Male and Female customers, The Number of male customers is a little more than female customers, gained value and the average value of Males is more than Female customers.

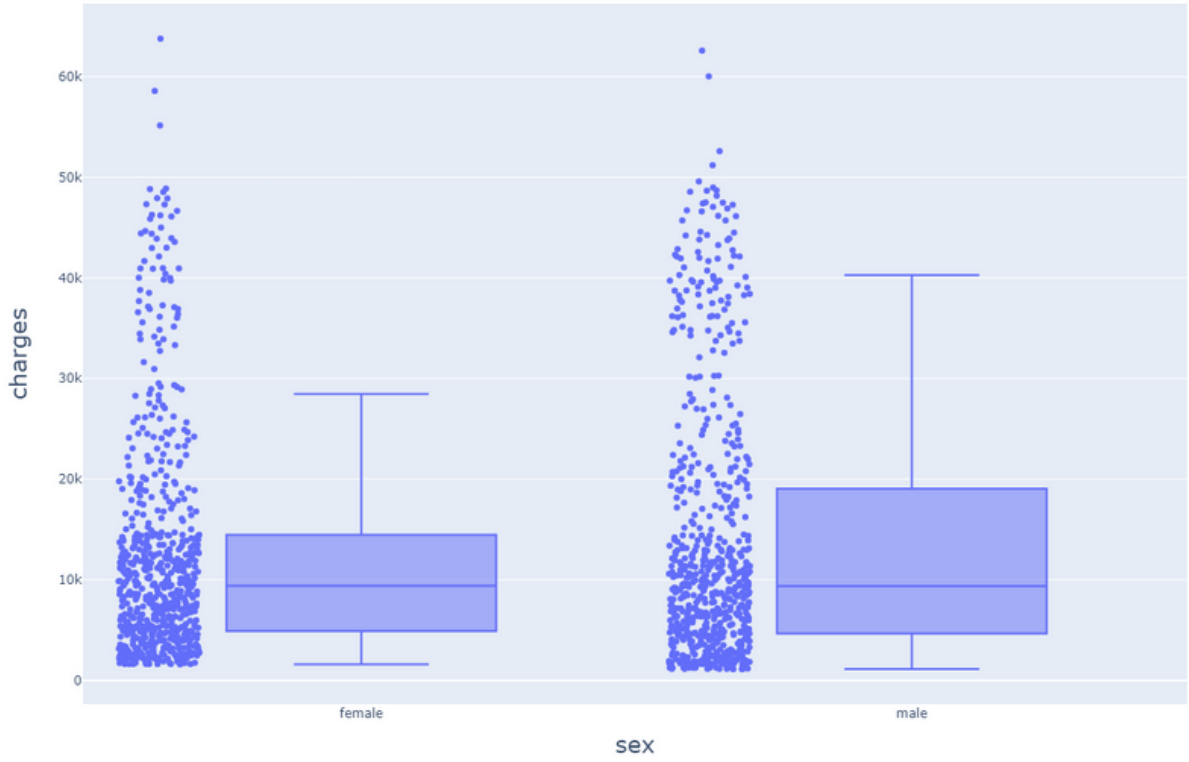


0 6

As observed, there is a significant difference in the average revenue between the Male and Female segments, despite the total revenue of both segments and the total number of customers in both segments being approximately equal. It is imperative to investigate the underlying cause of this variation.

It's highly likely that certain users and specific categories of male customers are paying substantial amounts for their insurance.

Let's examine the diversity of our data within each gender category:



The payment values of all males significantly surpass those of females. The third quartile of males is approximately $20,000, compared to around $15,000 for females. While the median payment for males slightly exceeds that of females, the noteworthy observation lies in the outliers among males. There are substantially more outliers and data points near the upper bound in the male segment compared to the female segment. This suggests that men are paying more for insurance on average.

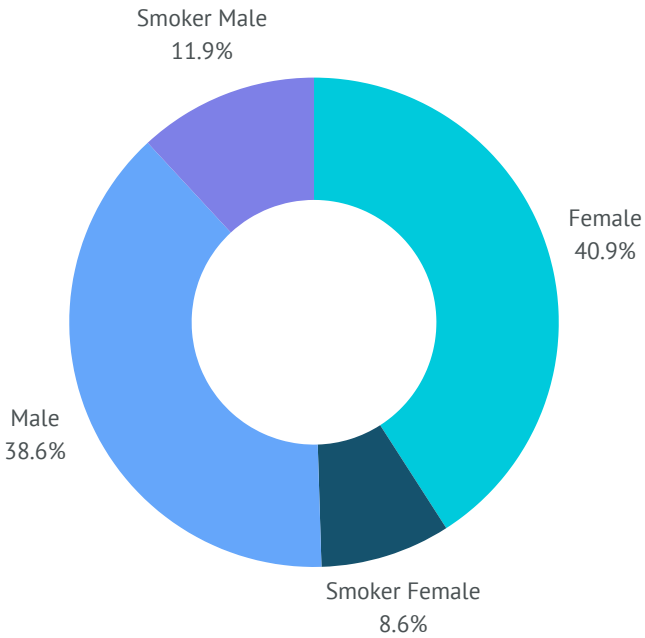
07

GENDER ANALYSIS BASED ON SMOKING BEHAVIOR

Let's separate individuals into smoker and non-smoker categories based on their gender.

After segregating users based on their smoking habits and gender, we obtained the following table.

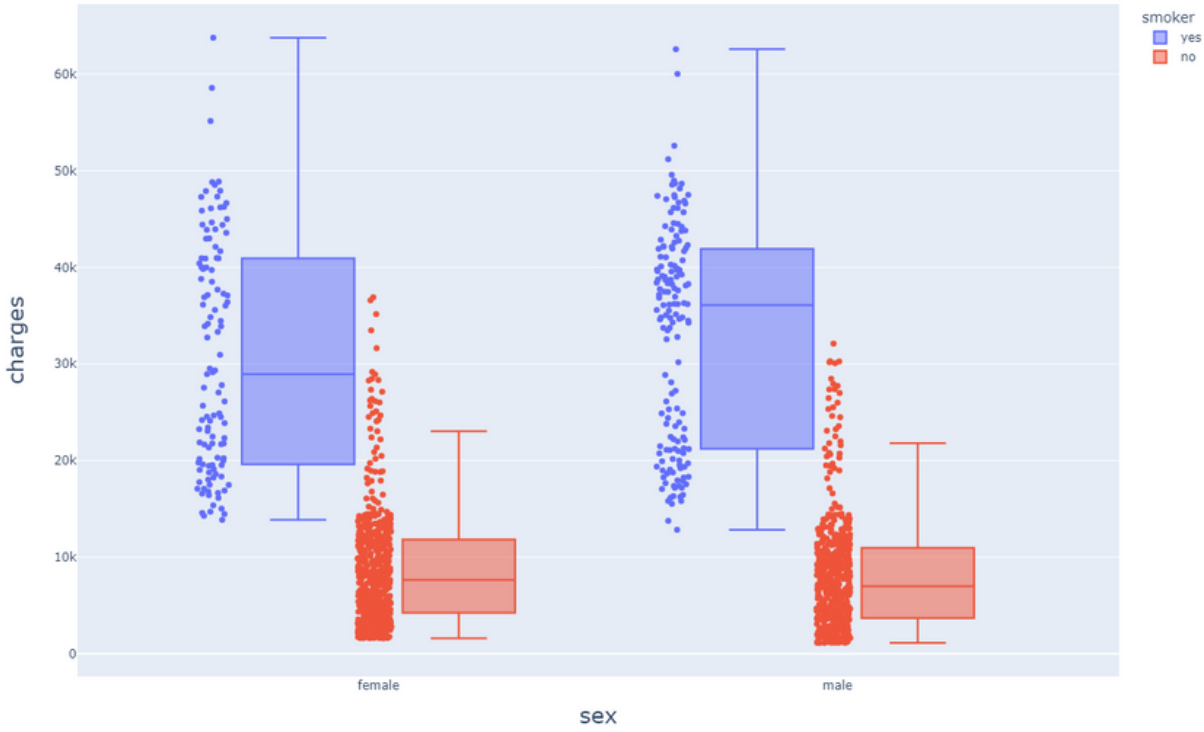
sex	smoker	count	total_revenue
Female	No	547	4792977
Female	Yes	115	352808
Male	No	516	417944
Male	Yes	159	5253679



The percentage of customers who smoke is significantly lower compared to non-smokers. It's essential to examine the value purchased by each segment.

08

ANALYZING PAYMENTS MADE BY SMOKERS:



A significant portion of our customer base comprises non-smokers. However, the value of purchases made by smoking customers, irrespective of gender, greatly exceeds that of non-smoking customers. iin table in blow you can see all important data from each box plot

	Lower	Q1	Mean	Median	Q3	Upper
Female smoker	13844.51	19696.43	30679.00	28950.47	40918.31	63770.43
Female Non-smoker	1607.51	4265.60	8762.30	7639.42	11786.17	36910.61
Male Smoker	12829.46	21241.53	33042.01	36085.22	41797.59	62592.88
Male Non-Smoker	1121.87	3684.91	8099.70	6986.10	10960.86	32108.66

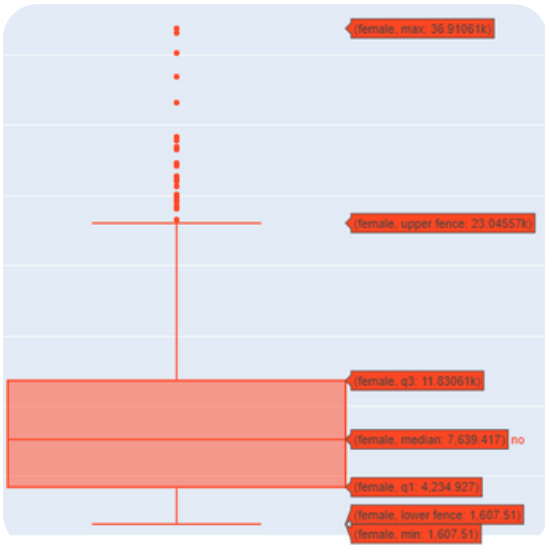
9

LETS LOOK AT EACH BOXPLOT SEPRATELY:

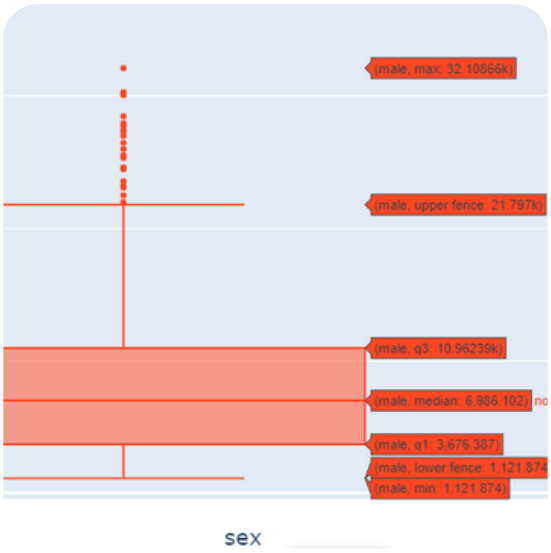
Female - Smoker

Male - Smoker

Female - Non Smoker



Male - Non Smoker



10

AGE GROUP:

WHICH AGE GROUP HAS THE HIGHEST AVERAGE PAYMENT?

"We want to determine which Age interval has the highest number of customers and generates the most revenue."

11

INTERPRETATION OF THE DIAGRAM :

Let's take a closer look at the data:

	age_range	total_revenue	total_costumer	average_revenue
0	18-30	3.827458e+06	416	9200.619154
1	31-40	3.016868e+06	257	11738.784117
2	41-50	4.017378e+06	279	14399.203564
3	51-60	4.470208e+06	271	16495.232665
4	61-70	2.422274e+06	114	21248.021885
5	71-80	0.000000e+00	0	0.000000
6	81-90	0.000000e+00	0	0.000000

Most of the revenue is generated from the age intervals 51-60, 41-50, and 18-30. However, it's worth noting that the number of customers in the 18-30 interval is significantly higher compared to the other two intervals, which likely contributes to its higher revenue. The average charges for individuals in the 41-60 age interval are notably higher than other age groups.

Another observed trend is the lowest revenue coming from the 61-70 age interval, despite having the highest payment average. This is likely due to the relatively low number of customers in this segment.

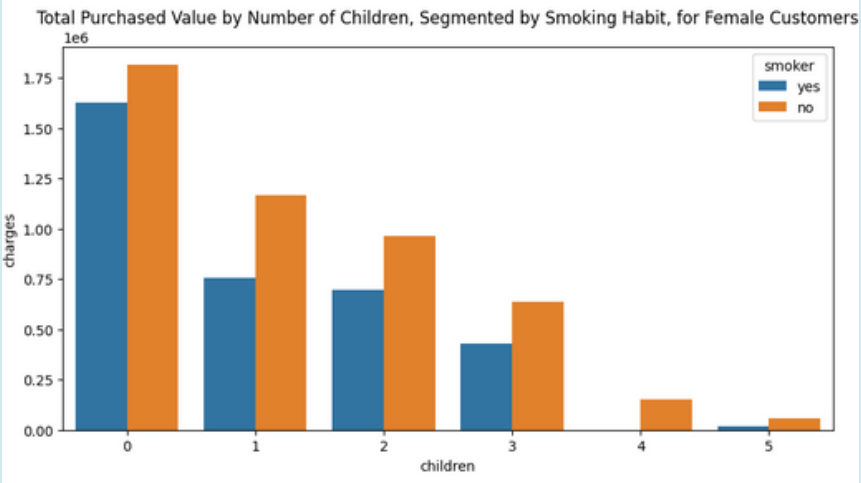
Additionally, there's a notable observation of low revenue and a small number of customers in the 31-40 age interval. With more data, further investigation into this trend could be conducted.

12

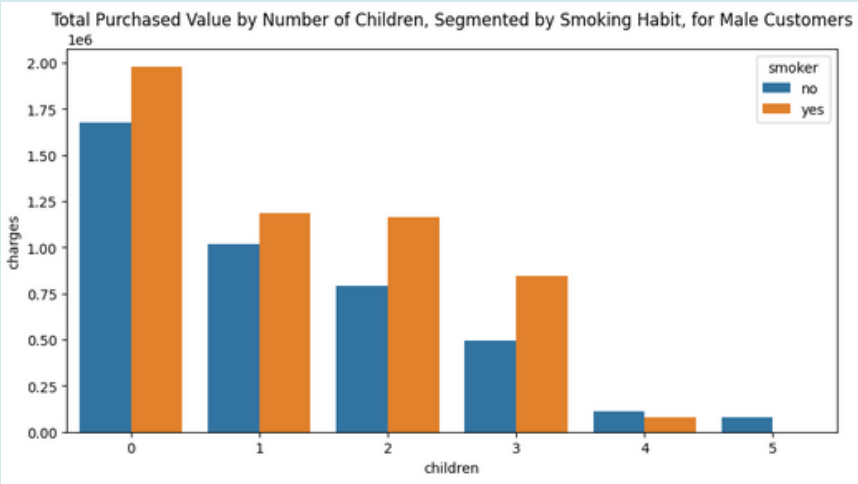
SMOKERS WITH CHILDREN:

WE WANT TO SEE WHAT IS POPULATION AND GAINED VALUE OF COSTUMERS WITH DIFFERENT NUMBER OF CHILDREN.

For better analysis, it's preferable to segregate customers into **male** and **female** categories.



This diagram pertains to female customers. It's evident that the total revenue from female customers decreases as the number of children increases. Additionally, within each group of customers with a specific number of children, smokers tend to pay less than non-smokers.

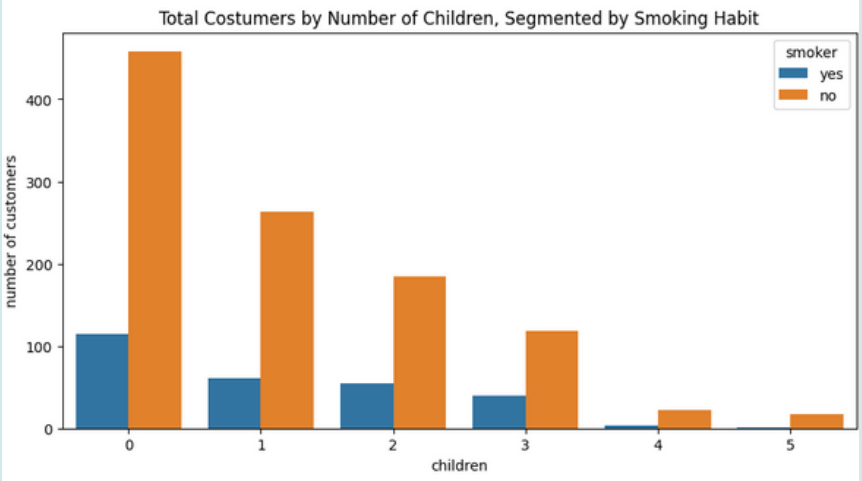


This diagram pertains to male customers. Notably, the total revenue of female customers decreases as the number of children increases. Furthermore, within each group of customers with a specific number of children, smokers tend to pay less than non-smokers. However, there are some differences: the total purchased value of males in all segments (except for those with 4 and 5 children) is greater than that of females. Unlike females, the gained value from men with 1 and 2 children is approximately the same.

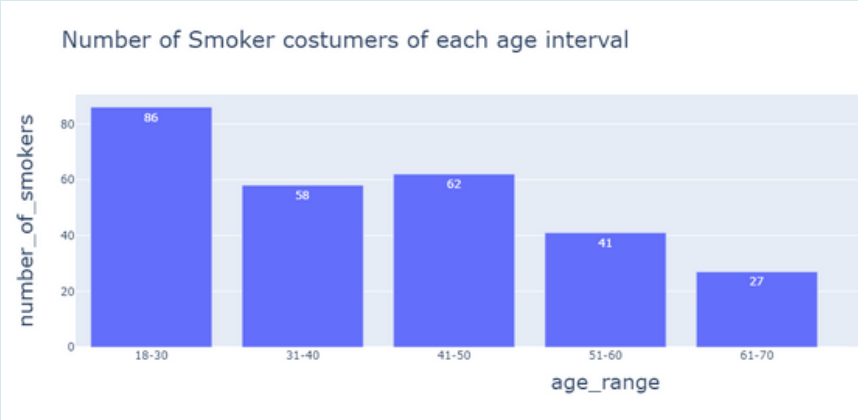
13

DISTRIBUTION OF CUSTOMERS BY CHILDREN AND SMOKING HABITS

LETS SEE WHAT ARE NUMBER OF COSTUMERS IN EACH SEGMENT BASED ON NUMBER OF CHILDREN AND SMOKING HABIT



As observed, the number of non-smoker customers significantly outweighs that of smoker customers. Moreover, with an increase in the number of children, the count of both smoker and non-smoker customers declines.



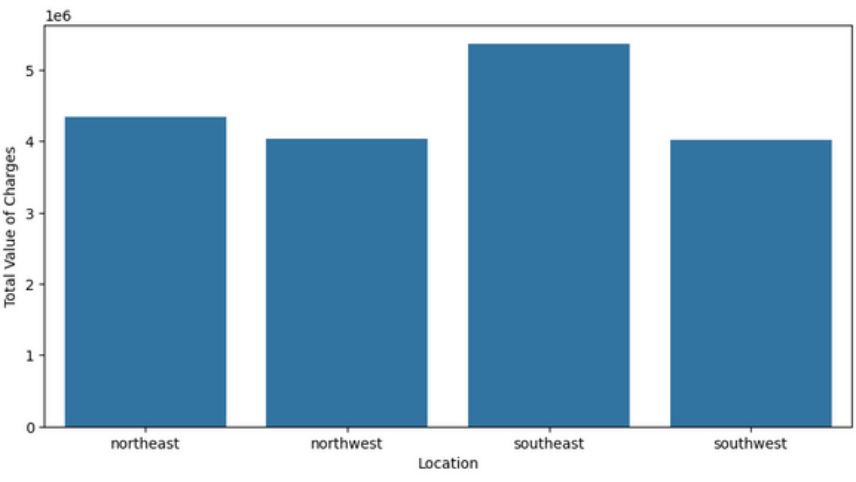
Total Smokers based by age

In the entire dataset, we have a total of 1337 customers, among whom 274 are smokers. The largest smoker group comprises 83 individuals aged between 18 and 30, followed by 62 smokers in the 41-50 age range, and 58 persons in the age range of 31-40.

14

COSTUMERS BASED ON REGION :

WE WANT TO SEE WHAT IS POPULATION AND GAINED VALUE OF COSTUMERS WITH DIFFERENT NUMBER OF CHILDREN.



As you can observe, there isn't a significant difference in the gained value based on location. However, due to limitations in the dataset, I'd prefer not to discuss location.

KPI TRACKING:

sex	smoker	bmi_condition	children	charges_count	charges_avg	revenue	rev_per_costumer
female	no	over_weight	0	195	7728.144666	1.506988e+06	2.938627e+08
male	no	over_weight	0	180	7892.383349	1.420629e+06	2.557132e+08
female	no	over_weight	1	109	8761.713400	9.550268e+05	1.040979e+08
male	no	over_weight	1	110	7924.412547	8.716854e+05	9.588539e+07
male	yes	over_weight	0	51	34356.377681	1.752175e+06	8.936094e+07
female	no	over_weight	2	79	10341.478413	8.169768e+05	6.454117e+07
female	yes	over_weight	0	41	34139.822789	1.399733e+06	5.738904e+07
male	no	over_weight	2	72	9283.379826	6.684033e+05	4.812504e+07
male	yes	over_weight	1	34	33681.703812	1.145178e+06	3.893605e+07
female	no	over_weight	3	52	10547.546353	5.484724e+05	2.852057e+07
male	yes	over_weight	2	27	38872.013417	1.049544e+06	2.833770e+07
male	no	over_weight	3	46	9518.802867	4.378649e+05	2.014179e+07

This is a segment of the table where I've defined the revenue per customer KPI for each segment. By tracking this KPI, we can monitor the situation of each segment effectively.

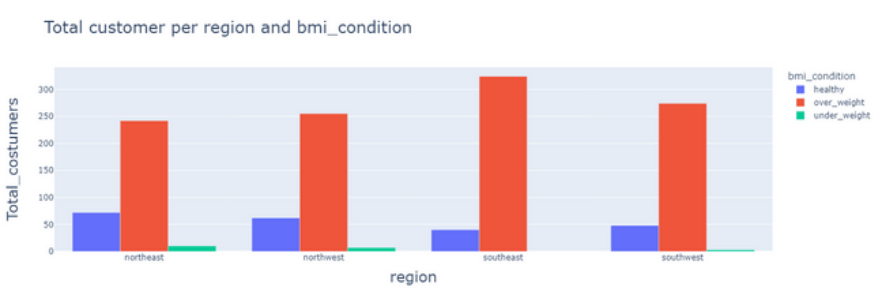
Within each segment, we have documented the number of customers and the revenue generated. By dividing the total revenue of the segment by the number of customers within that segment, we can establish a key performance indicator (KPI) known as "revenue per customer."

15

BMI CONDITION:

WE'RE INTERESTED IN EXAMINING THE POPULATION AND REVENUE OF CUSTOMERS WITH DIFFERENT BMI CONDITIONS ACROSS VARIOUS REGIONS.

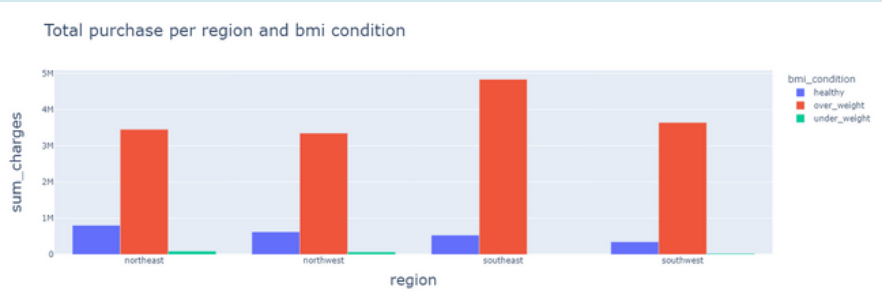
Based on the fact that the healthy BMI range falls between 18.5 to 24.9, the BMI condition in our dataset is categorized into **underweight, healthy,** and **overweight**.



In all regions, the majority of our customers are overweight. This could be due to two possible reasons:

Firstly, the majority of the population in Target society may be overweight, And this is normalizing the condition.

Secondly, another possibility is that overweight individuals are more concerned with being covered by insurance.



Here is the revenue gained based on BMI condition and region.

16

TOP 5 REVENUE:

INSURANCE COMPANY GAINED MOST FROM THESE SEGMENTS:

sex	smoker	bmi_condition	children	charges_count	charges_avg	revenue	rev_per_costumer
male	yes	over_weight	0	51	34356.377681	1.752175e+06	8.936094e+07
female	no	over_weight	0	195	7728.144666	1.506988e+06	2.938627e+08
male	no	over_weight	0	180	7892.383349	1.420629e+06	2.557132e+08
female	yes	over_weight	0	41	34139.822789	1.399733e+06	5.738904e+07
male	yes	over_weight	1	34	33681.703812	1.145178e+06	3.893605e+07

The majority of revenue is generated by male smokers who are overweight, which aligns with our observations from previous charts. However, what's intriguing is that 51 customers in this segment have contributed more value compared to 195 or 180 customers in other segments. This raises several questions: Are they solely paying for their insurance, or are some of them also covering their family members? Alternatively, perhaps they've opted for a specific payment plan, or maybe the company has offered them only one payment plan option?

Determine the most valuable age range within the top-ranked segment of our customer base.

In the table provided above, the initial row denotes non-smoking males with zero children. I will proceed to compute the average revenue for each age range within this specific segment. By doing so, we can identify the most valuable age range within each segment.

MOST VALUABLE AGE RANGE:

Age interval	Number of Clients	Average Payment
18-30	26	29834.68
31-40	6	31086.94
41-50	5	40794.06
51-60	6	37031.42
61-70	8	45474.15
71-80	0	0
81-90	0	0

For overweight male smokers with no children, I calculated the average payment within each age interval.

The segment with the highest average payment comprises individuals aged 61-70, followed by those aged 41-50. An interesting observation is that although the majority of customers in this segment fall within the age interval of 18-30, they have the lowest average payment.

16 LIMITATION OF PROJECT:

WE REQUIRE MORE DATA TO ENHANCE OUR COMPETITOR ANALYSIS CAPABILITIES.

HERE IS A LIST OF THE DATA WE REQUIRE:

1. What geographic area does the data belong to?
   City - Province - Country -Region - Continent - World

2. What are the characteristics of the target area? What are the characteristics of the population living in the target area? How many households are there? What is the crime history in the target area?

3. Is any information available on the public health of the target area?

4. What is the per capita income of people in this area?

5. What are the company's advertising methods?

6. What is the variety of plans that the insurance company offers?

7. How did people become familiar with our company?

8. When did each person join the insurance? What plan do they use?

9. What is the occupation and income of the insured person?

10. Has the insured person been insured before? If yes, which company? Why did they leave the previous company?

11. Is the customer still covered by the company? If not, why and when did they leave?

12. What underlying medical conditions does the insured person have?

13. Has the insured person been hospitalized before? Why?

14. Does the insured person live with their family or alone?

15. What is the marital status of the person? Single, married, divorced?

16. Does the company customer pay for insurance for themselves or for multiple people? If for multiple people, are these people the customer's family or employees of a company?

17. How much does each person from each category of insured people cost the insurance?

17

POSSIBLE NEXT STEPS WITH SUITABLE DATA:

1. By knowing the general status of the target area, it is possible to evaluate the overall performance of the company in each category of customers.

2. Find the appropriate advertising method for the company.

3. Calculate the cost of adding each new customer.

4. Calculate the risks and costs of adding customers with special conditions.

5. It is possible to create a variety of plans for customers.

6. Measure the churn rate of customers and investigate the factors affecting churn.

7. Measure the probability of churn of current customers and prevent it.

8. Review the profitability and loss of the company in different periods.

9. It is possible to create a series of fair and affordable plans for the poor who cannot afford them so that the company also makes a profit.

10. Measure the public health status of the target market.