

**Topic: US.Regional Analysis of Medical Insurance cost**

**Files: CIS41\_Project**

**Name : Tayyaba Fatima**

**\*\*Tasks:** EDA - pivot table, groupby table,countplot,histogram, FacetGrid,Regplot, pairplot, Implot,iplot or plotly, boxplot, violinplot,barplot, countplot,swarmplot,heatmap,ECDF Hypothesis Testing -ANOVA,Chi-squared,Normal-Test, pearson-correlation-coefficient,Z-Test Conclusion (findings) Summary:

1. This is a data of 1338 people for medical insurance in which most of them have age below 20.
2. Female smoke less than the male
3. There are more young males than young females whereas there are more old female than young male
4. If a person is smoker then their charges increased as smoking is positively correlated with the charges.
5. South east region have maximum number of people who smoke where as Northeast have less smokers out of 4 regions.
6. South east region have maximum people with 0 children.
7. South east region have 0 people with 5 children.That mean this region have more young people.
8. Children and charges have weak relationship.
9. People who have no children smokes more may be that's why there charges are more
10. BMI and Age have little effect on charges as it has weak strong relationship with charges.
11. Charges and sex have negative relationship.
12. The data shows that having more children does'nt increase charges.
13. People who have no children smokes more may be that's why there charges are more
14. 70% percent charges are between 10,000 to 20,000 in all the regions
15. Out of All the four regions South east people have more charges.

In [ ]: ▶

1

