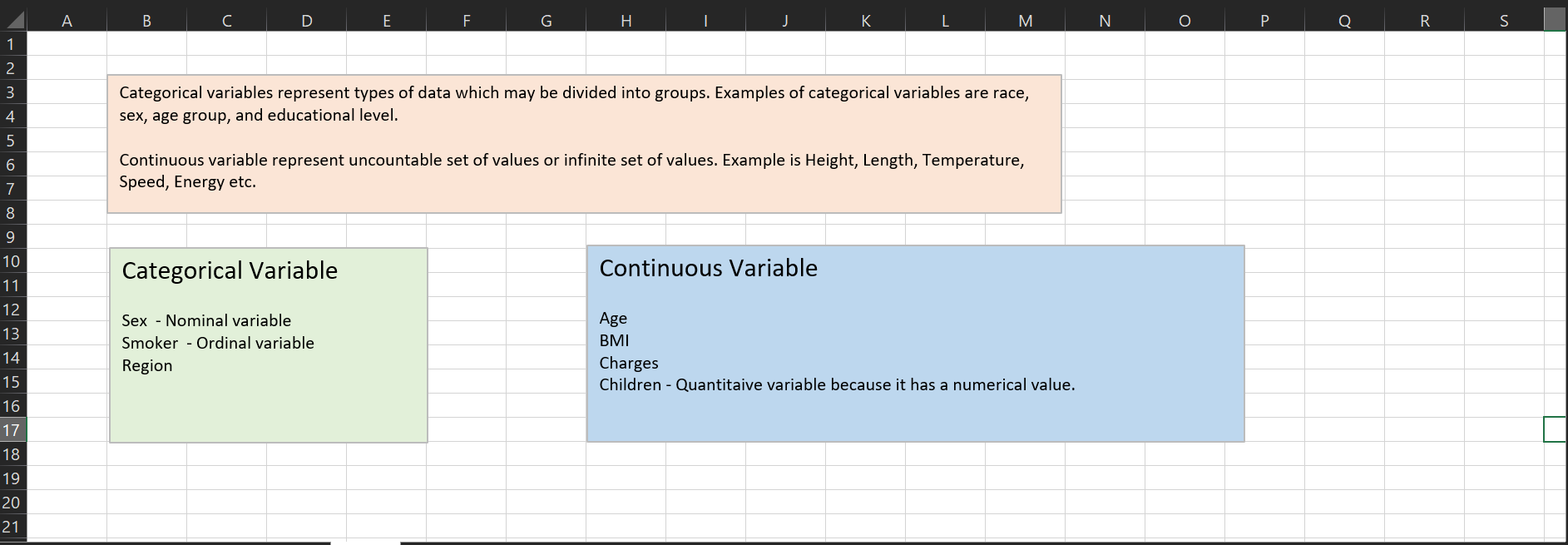
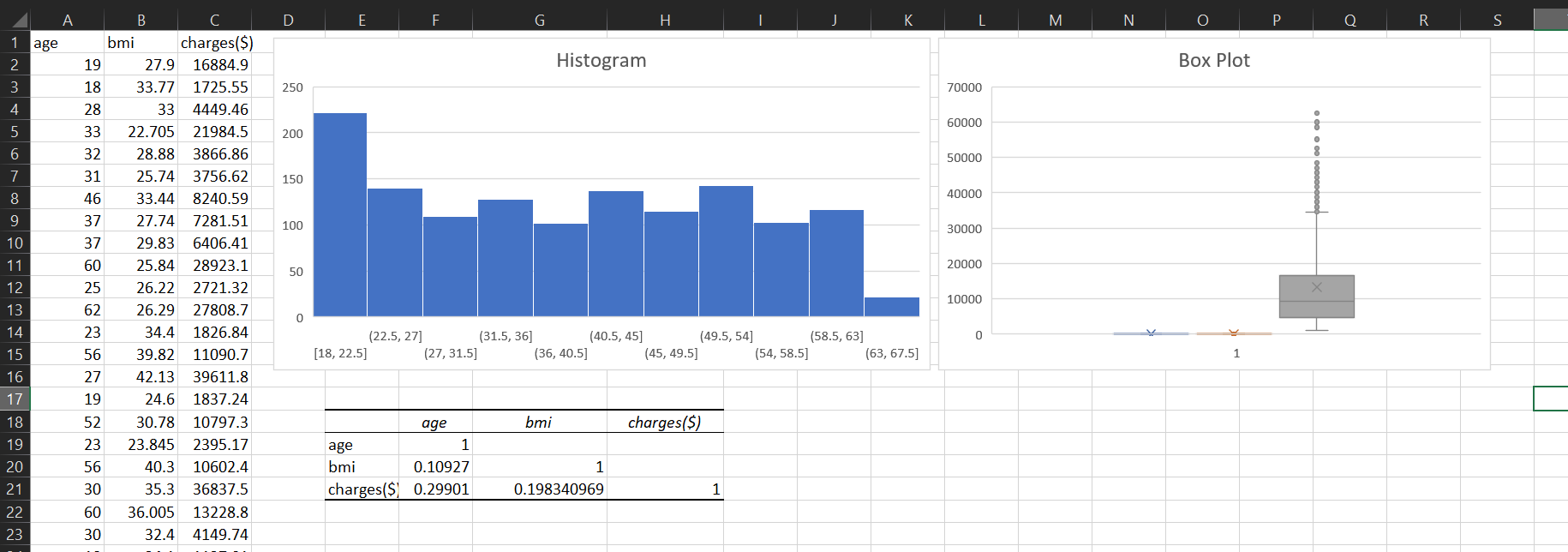
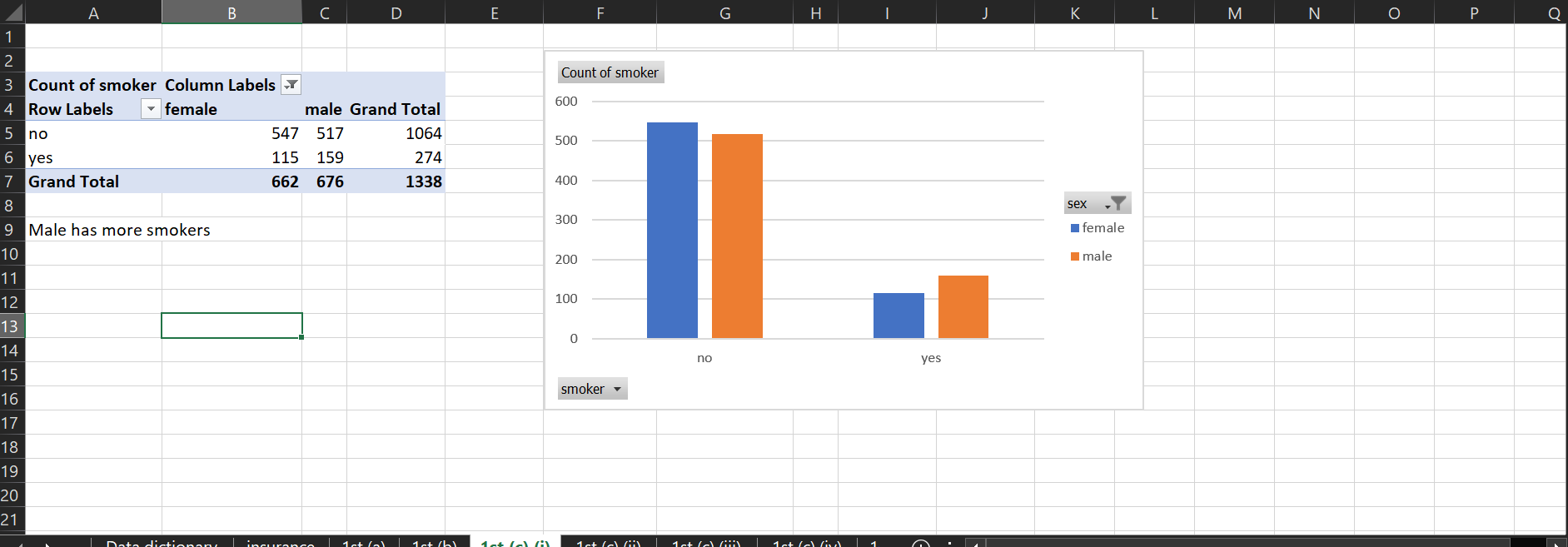
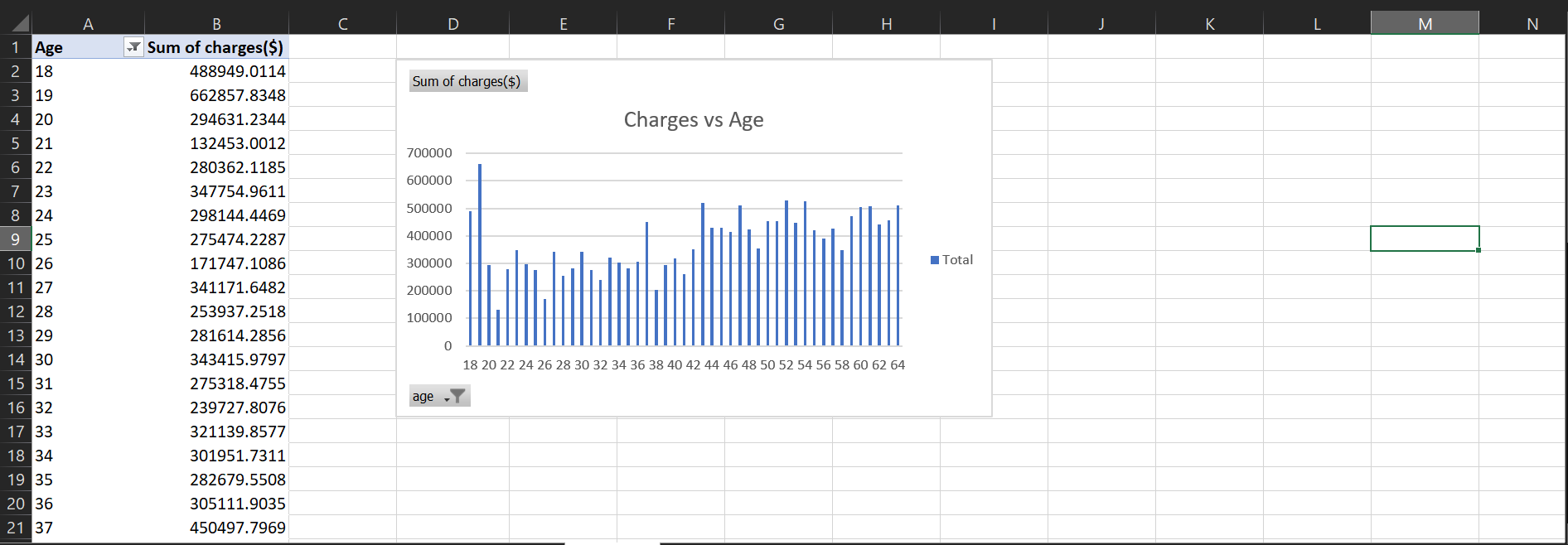
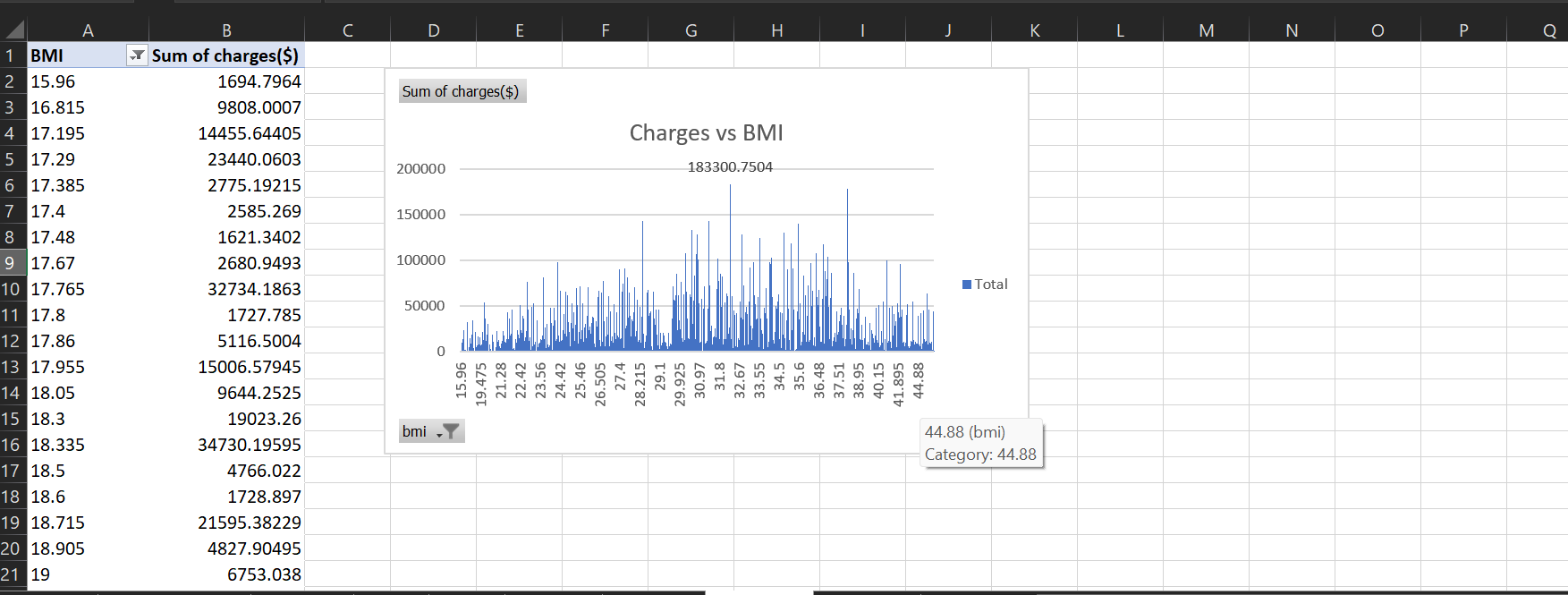
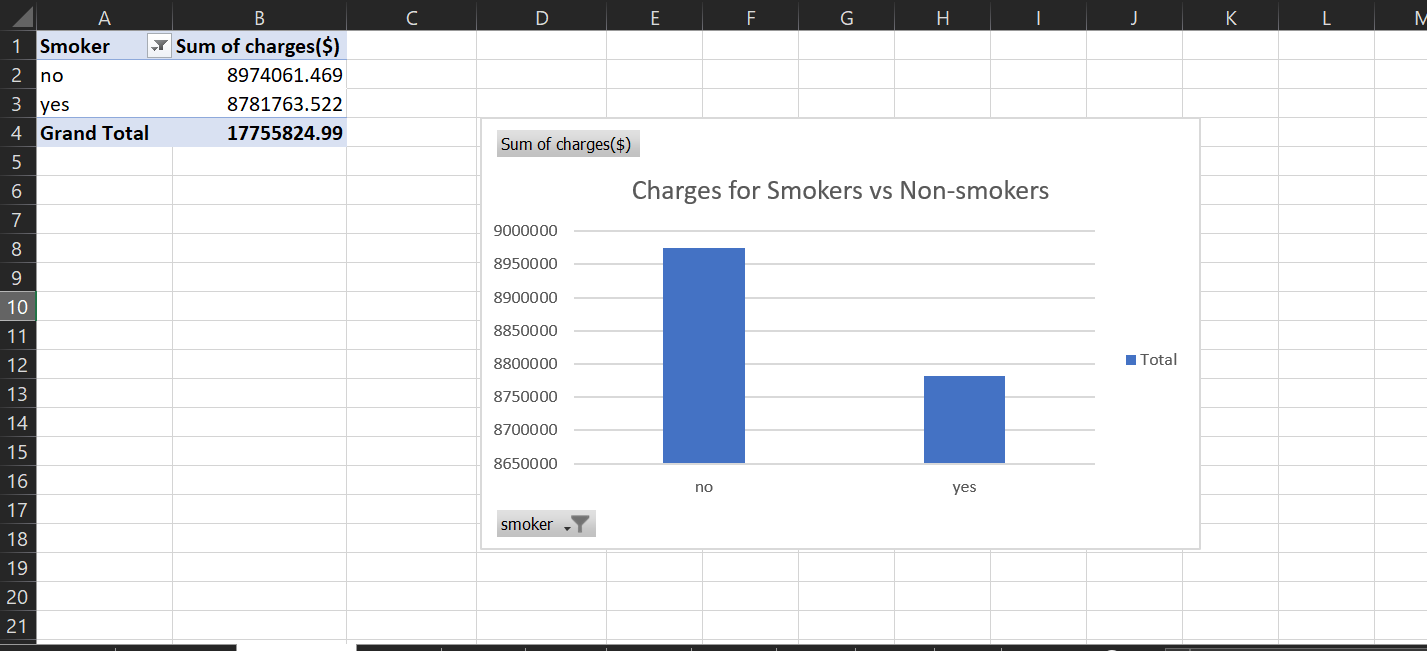
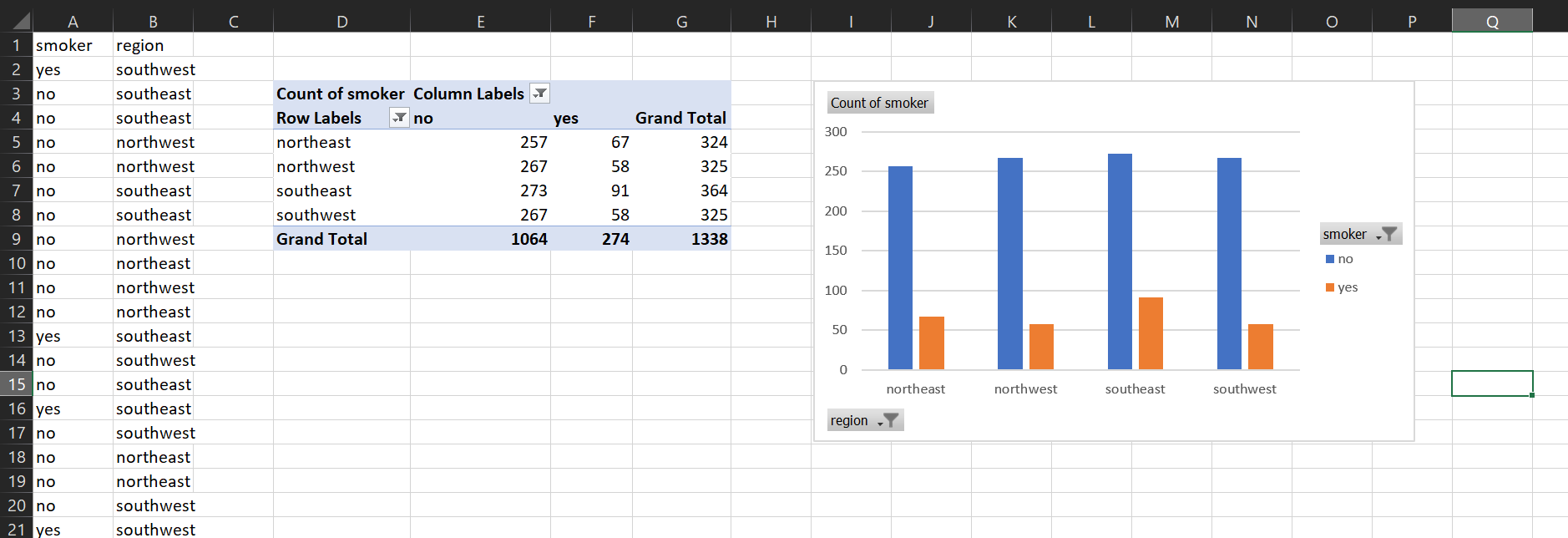
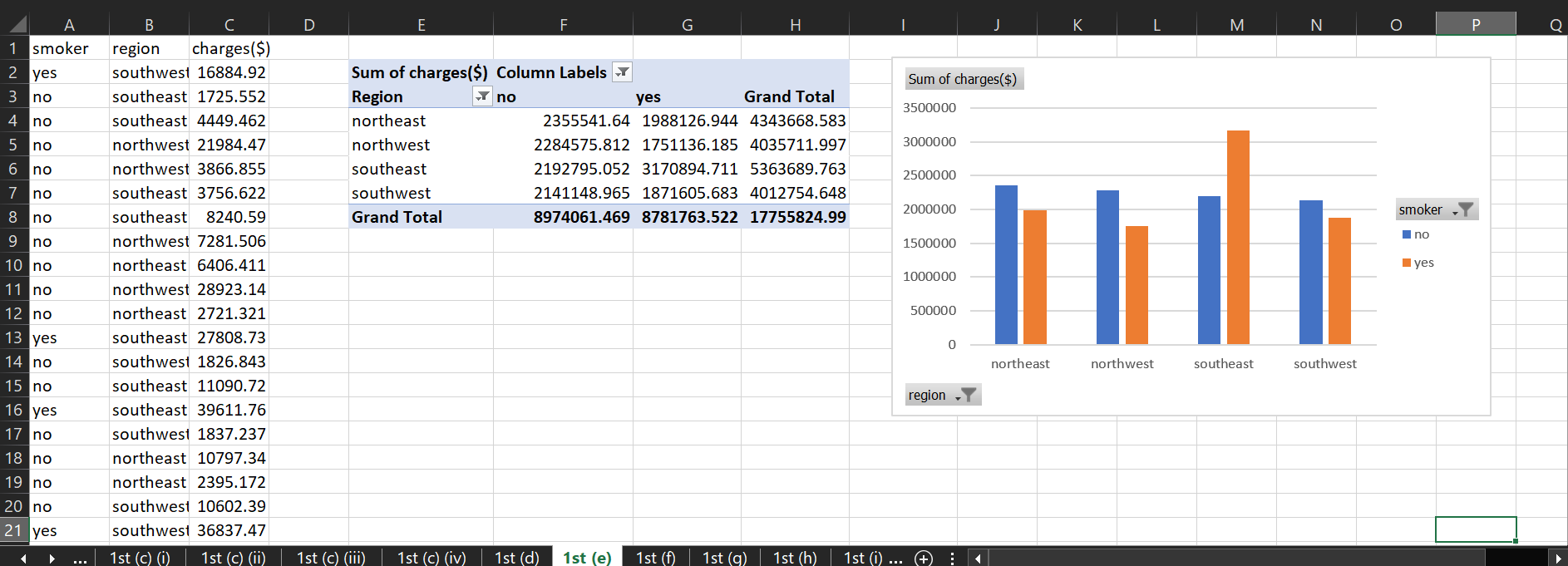
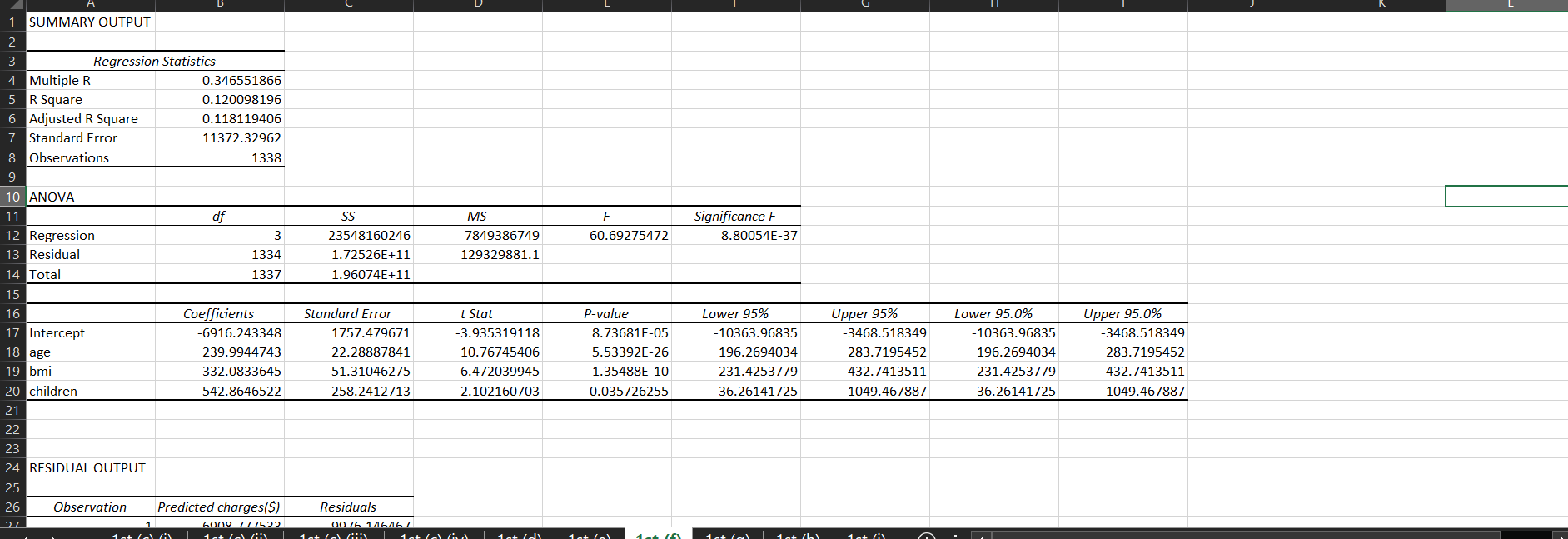
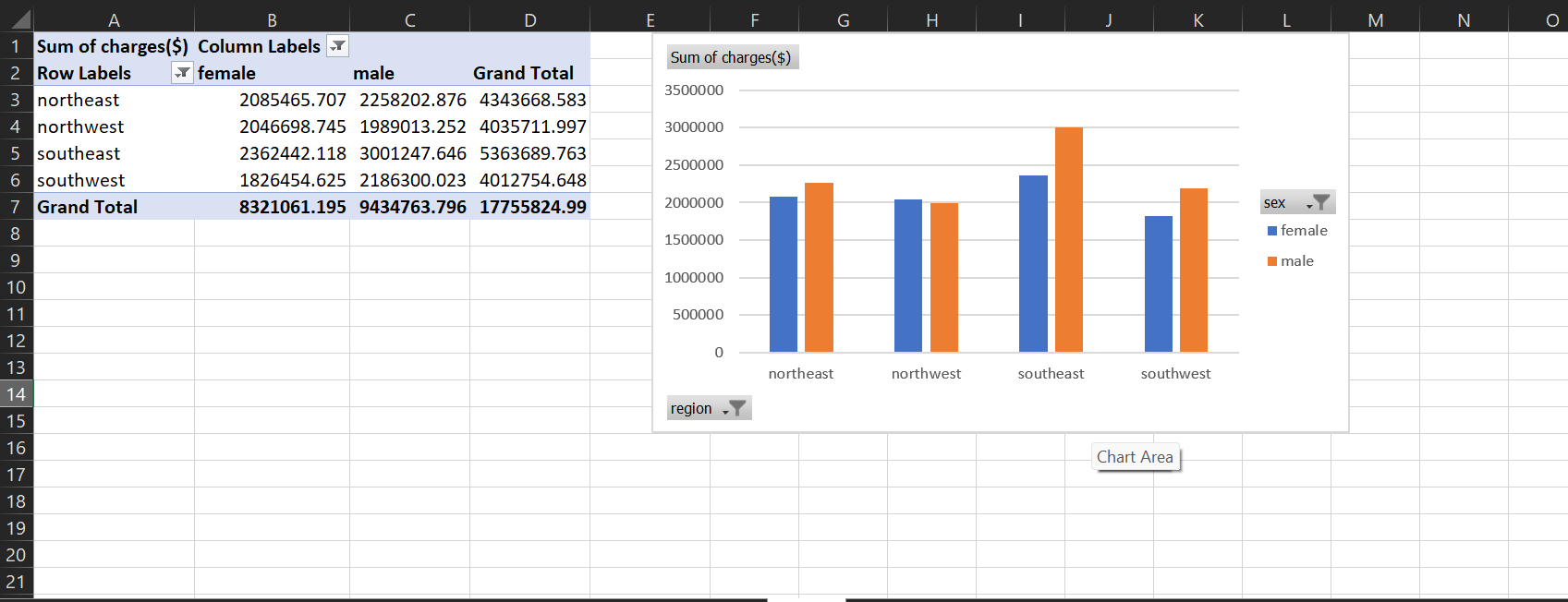
Business Report – Insurance Claim (Amit Varma)

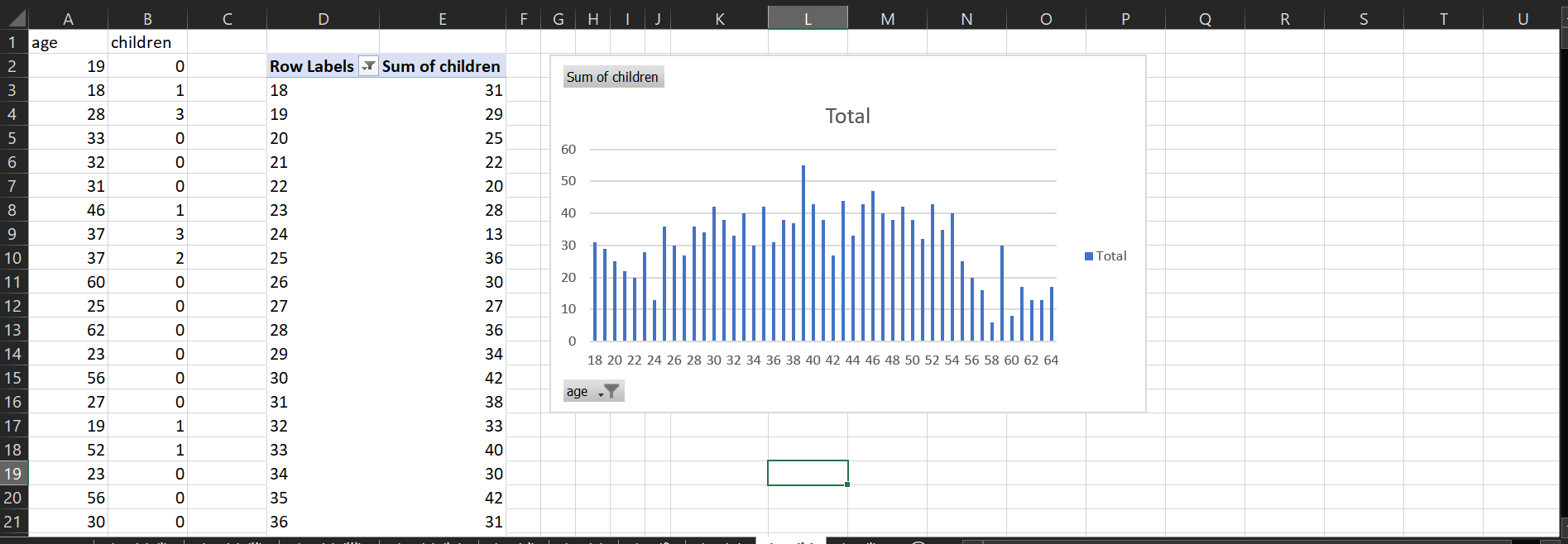
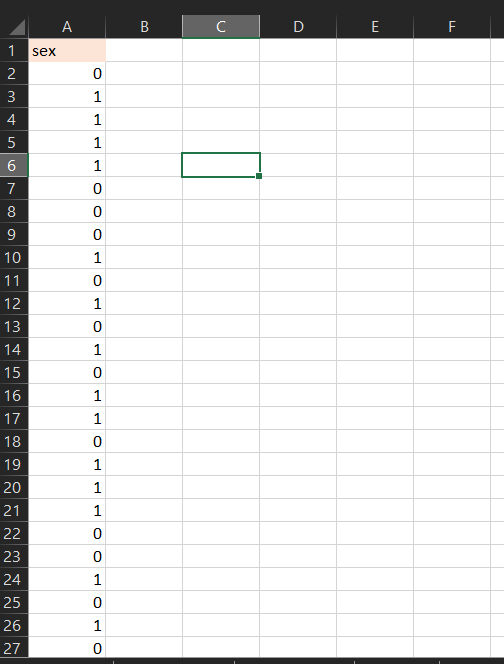
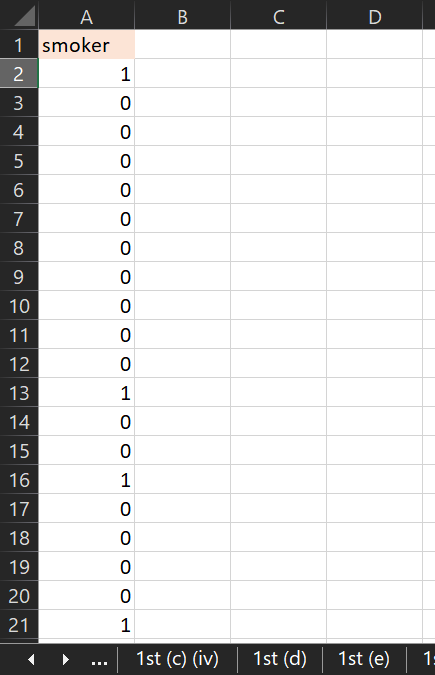
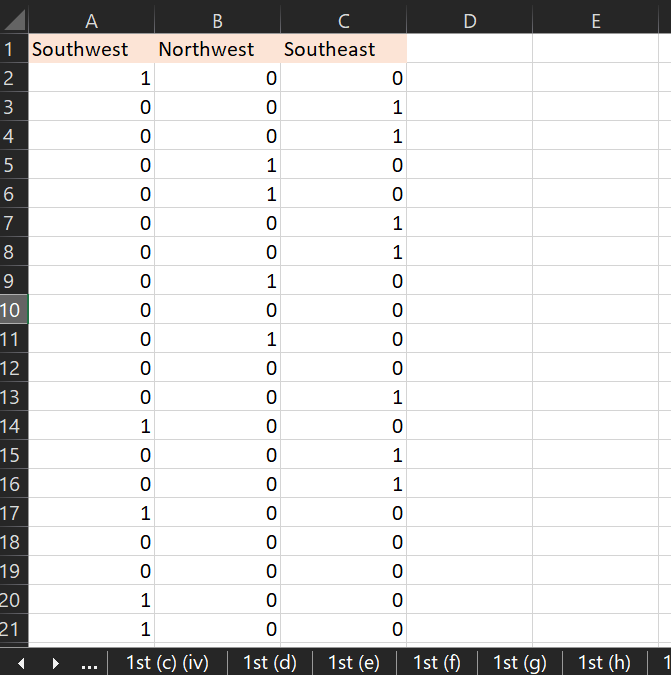
Question 1(a): Categorical and Continuous Variable

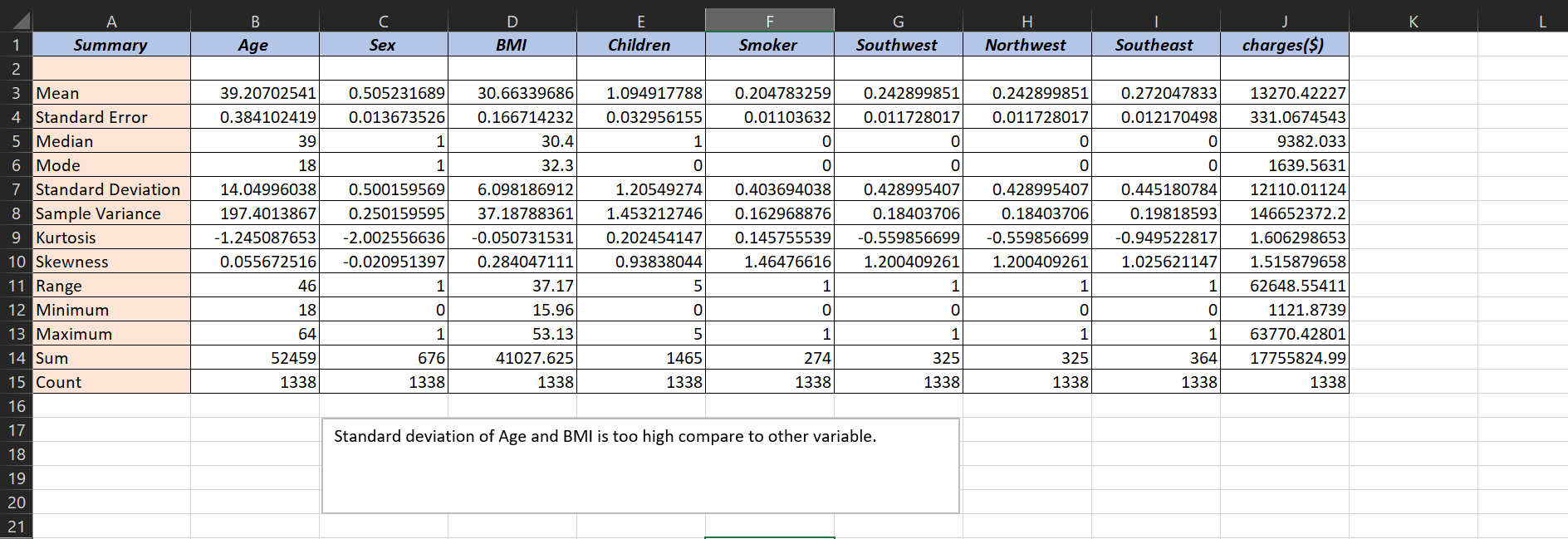
  
  
Here Sex, Smoker, Region is categorical variable because sex and region is nominal and ordinal variable this type of variable divided in to group.

Age, BMI, Charges and Children are continuous variable because he has numerical value.  
  
  
Question 1(b): Histogram, Box plot and Correlation.

  
  
Univariate analysis of continuous variable Age, BMI, Charges and multivariate analysis of correlation.  
  
My understanding on point b is Age and Body mass index ( 18, 22.5 )  
  
charges are high and correlation between Age, BMI, Charges.  
  
BMI has ( 0.10927 ) and charges ( 0.19834 ).  
  
  
Question 1(c): Pivot table and Pivot chart  
  
Male/Female ratio of smoker  
  
  
  
Male/Female ratio of smoker – here ratio of male smoker has more compare to female smoker.

But the good thing is that the number of non-smokers male is also high.  
  
Charges vs Age  
  
  
Charges of age between 18 to 20 is too high to claim insurance.  
  
Charges vs BMI  
  
  
  
Body mass index 31.8 to 32.67 charges is too high to claim insurance.  
  
  
Charges for Smokers vs Non-smokers  
  
  
  
Charges of non-smoker is high compare to smoker.  
  
  
  
  
  
Question 1(d):  
Region-wise smokers vs non-smokers  
  
  
  
Here southeast and northeast has more smokers and non-smoker has approx. equal to all region.  
  
  
Question 1(e):  
Region-wise charges for smokers vs non-smokers  
  
  
  
Here southeast has more smokers and his charges is 5363689.763 and second is northeast has more smokers and his charges is 4343668.583.  
Northeast has also more non-smoker  
Total charges of all region is 17755824.99.  
  
  
  
Question 1(f):  
  
  
  
Taking charges as y range and Age, BMI and children as x range and do regression analysis with residual plot.  
and R Square and Adjusted R Square is approx. similar.  
  
Question 1(g):  
  
  
  
Taking all region, sex and charges and created pivot chart and table, region as column, sex as rows and charges as values and result is that – male is more compare to female and his charges are high.

Question 1(h):  
  
  
  
Question 2(a)  
  
Replace all the male with 1 and female with 0.  
  
  
  
  
Question 2(b):  
Replace all the smokers with 1 and non-smokers with 0.  
  
  
  
  
Question 2(c):  
  
Replace whether northwest, southwest, southeast with 1 otherwise 0.  
  
using conditional statement formula.  
  
=IF(A2 = “northeast”, 1,0)   
  
  
  
  
Question 3:

Descriptive summary analysis  
  
  
  
  
  
  
  
  
  
Multiple Linear Regression analysis to identify which variables decide the insurance charges/billed insurance claim.  
  
