**Case Context**

A fictional telecom company provided home phone and Internet services to 7043 customers in California in Q3. The company has provided information for those 7043 customers.

**Variable Description**

7043 observations with 33 variables

**CustomerID**: A unique ID that identifies each customer.

**Count:** A value used in reporting/dashboarding to sum up the number of customers in a filtered set.

**Country:** The country of the customer’s primary residence.

**State:** The state of the customer’s primary residence.

**City:** The city of the customer’s primary residence.

**Zip Code:** The zip code of the customer’s primary residence.

**Lat Long:** The combined latitude and longitude of the customer’s primary residence.

**Latitude:** The latitude of the customer’s primary residence.

**Longitude:** The longitude of the customer’s primary residence.

**Gender:** The customer’s gender: Male, Female

**Senior Citizen:** Indicates if the customer is 65 or older: Yes, No

**Partner:** Indicate if the customer has a partner: Yes, No

**Dependents:** Indicates if the customer lives with any dependents: Yes, No. Dependents could be children, parents, grandparents, etc.

**Tenure Months:** Indicates the total amount of months that the customer has been with the company by the end of the quarter specified above.

**Phone Service:** Indicates if the customer subscribes to home phone service with the company: Yes, No

**Multiple Lines:** Indicates if the customer subscribes to multiple telephone lines with the company: Yes, No

**Internet Service:** Indicates if the customer subscribes to Internet service with the company: No, DSL, Fiber Optic, Cable.

**Online Security:** Indicates if the customer subscribes to an additional online security service provided by the company: Yes, No

**Online Backup:** Indicates if the customer subscribes to an additional online backup service provided by the company: Yes, No

**Device Protection:** Indicates if the customer subscribes to an additional device protection plan for their Internet equipment provided by the company: Yes, No

**Tech Support:** Indicates if the customer subscribes to an additional technical support plan from the company with reduced wait times: Yes, No

**Streaming TV:** Indicates if the customer uses their Internet service to stream television programing from a third party provider: Yes, No. The company does not charge an additional fee for this service.

**Streaming Movies:** Indicates if the customer uses their Internet service to stream movies from a third party provider: Yes, No. The company does not charge an additional fee for this service.

**Contract:** Indicates the customer’s current contract type: Month-to-Month, One Year, Two Year.

**Paperless Billing:** Indicates if the customer has chosen paperless billing: Yes, No

**Payment Method:** Indicates how the customer pays their bill: Bank Withdrawal, Credit Card, Mailed Check

**Monthly Charge:** Indicates the customer’s current total monthly charge for all their services from the company.

**Total Charges:** Indicates the customer’s total charges, calculated to the end of the quarter specified above.

**Churn Label:** Yes = the customer left the company this quarter. No = the customer remained with the company. Directly related to Churn Value.

**Churn Value:** 1 = the customer left the company this quarter. 0 = the customer remained with the company. Directly related to Churn Label.

**Churn Score:** A value from 0-100 that is calculated using the predictive tool IBM SPSS Modeler. The model incorporates multiple factors known to cause churn. The higher the score, the more likely the customer will churn.

**CLTV:** Customer Lifetime Value. A predicted CLTV is calculated using corporate formulas and existing data. The higher the value, the more valuable the customer. High-value customers should be monitored for churn.

**Churn Reason:** A customer’s specific reason for leaving the company. Directly related to Churn Category.

**Exercises on Probability and conditional probability:**

**Probability and Distributions**

1. In the Telco churn dataset, what are the variables discrete random variables and continuous random variables?

**Binomial Distributions**

1. What is the probability that in the coming quarter, among 500 customers exactly 350 customers will stay with the operator?
2. What is the probability that in the coming quarter, among 500 customers 160 customers will leave the operator?
3. What is the probability that in the coming quarter among 1000 customers less than 300 customers will leave the operator?
4. What is the probability that in the coming quarter among 1000 customers more than 300 customers will leave the operator?
5. What is the probability that out of 500 senior citizens 200 will stay with the operator in the coming quarter?
6. What is the probability that out of 500 senior citizens exactly 285 customers will leave the operator in the coming quarter?
7. What is the probability that out of 500 senior citizens more than 285 customers will leave the operator in the coming quarter?
8. Out of total 500 senior citizen customers, calculate the probabilities of stay for all the senior citizen customers from 160 to 260 and plot those values in a graph to show the distribution. **(Home Exercise)**

**Normal Distribution**

1. Calculate the probability of customers having less than 30 months of tenure with the operator.
2. Draw the normal distribution curve of the ‘tenure months’ variable by using GGPLOT package.
3. Calculate the probability of customers having more than 47 months of tenure with the operator.
4. What is the probability that the customers are staying with the operator in between 30 to 40 months?
5. Calculate the probability of female customers having CLTV of more than 4000? **(Home exercise)**
6. What is standard normal curve? Calculate the probability of z values -1.96, -1.64, 0, 1.64, and 1.96 from a standard normal curve?
7. Now, calculate the z values on the standard normal curve when probability values are 0.025, 0.5, 0.95, and 0.975.

**Estimation of Population Parameters**

1. Calculate the population mean interval estimate values of the variable ‘Monthly Charges’ considering 95% confidence interval.
2. Calculate the population mean interval estimate values of the variable ‘Monthly Charges’ considering 99% confidence interval.
3. Why is t-distribution more important than z-distribution during real-world case analysis?
4. Calculate the population mean value of the variable ‘Monthly Charges’ considering 95% confidence interval as interval estimate and follows t-distribution.
5. Calculate the interval estimate of the population mean of the variable ‘Total Charges’ following t-distribution and 90% confidence interval.
6. Separately, calculate the interval estimate of the population mean of the variable ‘Tenure Months’ for customers who churned out of the operator and for those who stayed with the operator. Consider that the variable followed the t-distribution graph and the confidence interval is 95%. Explain your findings. **(Home Exercise)**