**Sample & Count Assignment**

**Problem Statement:**

Sam’s next exam would be on sample\_n(), sample\_frac() & count() functions from the dplyr package. You would be asked questions on what you’ve learnt from the respective module.

Questions:

1. Extract 333 random records from the customer\_churn dataframe & store the result in

‘customer\_333’

> sample\_n(customer\_churn, 100) -> random\_100

> head(random\_100)

customerID gender SeniorCitizen Partner Dependents tenure PhoneService MultipleLines InternetService

1 3026-ATZYV Female 0 Yes Yes 37 Yes No DSL

2 8496-DMZUK Male 0 No No 30 Yes Yes Fiber optic

3 5480-TBGPH Female 0 Yes No 24 Yes No Fiber optic

4 6317-YPKDH Female 0 No No 1 No No phone service DSL

5 6591-QGOYB Male 0 No No 17 Yes Yes Fiber optic

6 3259-QMXUN Male 0 Yes No 58 Yes Yes DSL

OnlineSecurity OnlineBackup DeviceProtection TechSupport StreamingTV StreamingMovies Contract

1 Yes Yes No No Yes Yes One year

2 Yes Yes Yes No No No One year

3 No Yes Yes No Yes Yes Month-to-month

4 No Yes No No No No Month-to-month

5 Yes Yes Yes Yes Yes No Month-to-month

6 Yes Yes Yes No Yes Yes Two year

PaperlessBilling PaymentMethod MonthlyCharges TotalCharges Churn

1 No Bank transfer (automatic) 75.10 2658.80 No

2 No Bank transfer (automatic) 90.40 2820.65 No

3 Yes Credit card (automatic) 99.30 2431.35 Yes

4 No Bank transfer (automatic) 29.95 29.95 Yes

5 Yes Bank transfer (automatic) 106.65 1672.10 No

6 Yes Electronic check 86.10 4890.50 No

2. Extract 1000 random records from the customer\_churn dataframe & store the result in

‘customer\_1000’

> sample\_n(customer\_churn,1000) -> customer\_1000

> head(customer\_1000)

customerID gender SeniorCitizen Partner Dependents tenure PhoneService MultipleLines InternetService

1 0233-FTHAV Female 0 No No 60 Yes Yes Fiber optic

2 8217-QYOHV Male 0 No No 5 Yes No DSL

3 8950-MTZNV Male 0 No No 1 Yes No DSL

4 4729-XKASR Male 0 No Yes 1 No No phone service DSL

5 8400-WZICQ Female 0 Yes Yes 55 Yes No DSL

6 8035-PWSEV Female 0 No No 6 Yes No Fiber optic

OnlineSecurity OnlineBackup DeviceProtection TechSupport StreamingTV StreamingMovies Contract

1 No No Yes No No No One year

2 Yes No No Yes No No Month-to-month

3 No No No No No No Month-to-month

4 No No No No No No Month-to-month

5 Yes Yes No No No No Month-to-month

6 No No Yes Yes No Yes Month-to-month

PaperlessBilling PaymentMethod MonthlyCharges TotalCharges Churn

1 Yes Bank transfer (automatic) 79.20 4765.00 No

2 Yes Mailed check 55.75 266.95 No

3 Yes Mailed check 44.95 44.95 No

4 No Electronic check 24.75 24.75 Yes

5 No Electronic check 55.70 3131.80 No

6 Yes Electronic check 89.25 487.05 No

3. Randomly extract 23% of the records from the customer\_churn dataframe & store the result in ‘customer\_23\_percent’

> sample\_n(customer\_churn, 0.23) -> customer\_23percentage

> head(customer\_23percentage)

[1] customerID gender SeniorCitizen Partner Dependents tenure

[7] PhoneService MultipleLines InternetService OnlineSecurity OnlineBackup DeviceProtection

[13] TechSupport StreamingTV StreamingMovies Contract PaperlessBilling PaymentMethod

[19] MonthlyCharges TotalCharges Churn

<0 rows> (or 0-length row.names)

4. Get the count of different levels from the ‘PaymentMethod’ column

> count(customer\_churn, PaymentMethod)

# A tibble: 4 x 2

PaymentMethod n

*<fct>* *<int>*

1 Bank transfer (automatic) 1544

2 Credit card (automatic) 1522

3 Electronic check 2365

4 Mailed check 1612

5. Get the count of different levels from the ‘Churn’ column

> count(customer\_churn, Churn)

# A tibble: 2 x 2

Churn n

*<fct>* *<int>*

1 No 5174

2 Yes 1869

s