Customer Retention: Exploratory Data Analysis

Importing Data

```
import os, pandas, numpy, matplotlib, seaborn
%matrix inline
%matrix inline
%matrix inline
```

```
Index(['Customer ID', 'Gender', 'Location', 'Tenure Months',
       'Total Long Distance Calls', 'Total Long Distance Charges',
       'Total Revenues', 'Referred a Friend', 'Number of Referrals',
       'Tenure in Months', 'Offer', 'Phone Service', 'Avg Monthly Long',
       'Multiple Lines', 'Internet Service', 'Internet Type',
       'Avg Monthly GB Download', 'Online Security', 'Online Backup',
       'Device Protection Plan', 'Premium Tech Support', 'Streaming TV',
       'Streaming Movies', 'Streaming Music', 'Unlimited Data',
       'Contract', 'Paperless Billing', 'Payment Method',
       'Monthly Charge', 'Total Charges', 'Total Refunds',
       'Total Extra Data Charges', 'Total Long Distance Charges',
       'Total Revenue', 'Satisfaction Score', 'Customer Status',
       'Churn Label', 'Churn Score', 'CLTV', 'Churn Category',
       'Churn Reason'],
      dtype=object)
```

```
# Checking for duplicates
df[df.duplicated()]
```

```
Empty DataFrame
Columns: [Customer ID, Gender, Location, Tenure Months, Total Long Distance Calls, Total Long Distance Charges, Total Revenues, Referred a Friend, Number of Referrals, Tenure in Months, Offer, Phone Service, Avg Monthly Long, Multiple Lines, Internet Service, Internet Type, Avg Monthly GB Download, Online Security, Online Backup, Device Protection Plan, Premium Tech Support, Streaming TV, Streaming Movies, Streaming Music, Unlimited Data, Contract, Paperless Billing, Payment Method, Monthly Charge, Total Charges, Total Refunds, Total Extra Data Charges, Total Long Distance Charges, Total Revenue, Satisfaction Score, Customer Status, Churn Label, Churn Score, CLTV, Churn Category, Churn Reason]
Index: []
```

```
# Checking for missing values
df.isnull().sum
```

```
Customer ID                          0
Gender                               0
Location                             0
Tenure Months                        0
Total Long Distance Calls            0
Total Long Distance Charges          0
Total Revenues                       0
Referred a Friend                    0
Number of Referrals                  0
Tenure in Months                     0
Offer                                0
Phone Service                        0
Avg Monthly Long                     0
Multiple Lines                       0
Internet Service                     0
Internet Type                        0
Avg Monthly GB Download              0
Online Security                      0
Online Backup                        0
Device Protection Plan               0
Premium Tech Support                 0
Streaming TV                         0
Streaming Movies                     0
Streaming Music                      0
Unlimited Data                       0
Contract                             0
Paperless Billing                    0
Payment Method                       0
Monthly Charge                       0
Total Charges                        0
Total Refunds                        0
Total Extra Data Charges             0
Total Long Distance Charges          0
Total Revenue                        0
Satisfaction Score                   0
Customer Status                      0
Churn Label                          0
Churn Score                          0
CLTV                                 0
Churn Category                       0
Churn Reason                         0
dtype: int64
```

```
# Checking for data types
df.dtypes
```

```
Customer ID                          object
Gender                               object
Location                             object
Tenure Months                         int64
Total Long Distance Calls             int64
Total Long Distance Charges           float64
Total Revenues                        float64
Referred a Friend                     object
Number of Referrals                   int64
Tenure in Months                      int64
Offer                                 object
Phone Service                         object
Avg Monthly Long                      int64
Multiple Lines                        object
Internet Service                      object
Internet Type                         object
Avg Monthly GB Download               int64
Online Security                       object
Online Backup                         object
Device Protection Plan                object
Premium Tech Support                  object
Streaming TV                          object
Streaming Movies                      object
Streaming Music                       object
Unlimited Data                        object
Contract                              object
Paperless Billing                     object
Payment Method                        object
Monthly Charge                        float64
Total Charges                         float64
Total Refunds                         float64
Total Extra Data Charges              int64
Total Long Distance Charges           float64
Total Revenue                         float64
Satisfaction Score                    int64
Customer Status                       object
Churn Label                           object
Churn Score                           int64
CLTV                                  int64
Churn Category                        object
Churn Reason                          object
dtype: object
```

```
# Checking for unique values
df.nunique()
```

```
Customer ID                          7043
Gender                               2
Location                             1106
Tenure Months                        72
Total Long Distance Calls            1198
Total Long Distance Charges          6381
Total Revenues                       6857
Referred a Friend                    2
Number of Referrals                  12
Tenure in Months                     72
Offer                                5
Phone Service                        2
Avg Monthly Long                     454
Multiple Lines                       2
Internet Service                     2
Internet Type                        3
Avg Monthly GB Download              38
Online Security                      2
Online Backup                        2
Device Protection Plan               2
Premium Tech Support                 2
Streaming TV                         2
Streaming Movies                     2
Streaming Music                      2
Unlimited Data                       2
Contract                             3
Paperless Billing                    2
Payment Method                       3
Monthly Charge                       1585
Total Charges                        6542
Total Refunds                        401
Total Extra Data Charges             1
Total Long Distance Charges          6381
Total Revenue                        6857
Satisfaction Score                   5
Customer Status                      3
Churn Label                          2
Churn Score                          1
CLTV                                 4
Churn Category                       5
Churn Reason                         20
dtype: int64
```

```
# Checking for shape
df.shape
```

```
(7043, 50)
```

Data Normalization

```
import numpy
from sklearn.preprocessing import MinMaxScaler
x = df[['Tenure Months', 'Total Long Distance Calls', 'Total Long Distance Charges',
'Total Revenues', 'Number of Referrals',
'Tenure in Months', 'Avg Monthly Long', 'Avg Monthly GB Download',
'Total Refunds', 'Total Extra Data Charges', 'Total Long Distance Charges',
'Total Revenue', 'Satisfaction Score', 'Churn Score', 'CLTV']]
x
```

	Tenure Months	Total Long Distance Calls	Total Long Distance Charges	Total Revenues	Number of Referrals	Tenure in Months	Avg Monthly Long	Avg Monthly GB Download	Total Refunds	Total Extra Data Charges	Total Long Distance Charges	Total Revenue	Satisfaction Score	Churn Score	CLTV
0	12	15	73.35	2475.15	2	12	390	21	0.00	150	73.35	2475.15	3	62	3239
1	12	390	974.81	2868.15	2	12	203	10	0.00	0	974.81	2868.15	3	10	2701
2	72	203	6010.94	8684.80	8	72	494	30	45.00	1950	6055.94	8639.80	2	67	5372
3	12	494	2059.35	2059.35	0	12	234	4	0.00	0	2059.35	2059.35	2	86	2793
4	12	234	188.35	188.35	0	12	413	15	0.00	0	188.35	188.35	2	84	2235
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
7038	18	1234	2227.17	3777.60	1	18	1116	46	0.00	0	2227.17	3777.60	3	59	5306
7039	21	1307	1190.37	5370.37	1	21	1024	27	0.00	0	1190.37	5370.37	3	71	5089
7040	18	1223	307.80	1024.80	1	18	1610	43	0.00	0	307.80	1024.80	3	59	5244
7041	18	1351	577.65	1349.65	1	18	1062	11	0.00	0	577.65	1349.65	3	50	5217
7042	18	1670	5745.17	5745.17	1	18	1232	7	0.00	0	5745.17	5745.17	3	62	5369

7043 rows × 15 columns

```
import numpy as np
from sklearn.preprocessing import MinMaxScaler
x = np.array(x)
scaler = MinMaxScaler()
scaled_data = scaler.fit_transform(x)
print("Original data:\n", x[:5])
print("Scaled data:\n", scaled_data[:5])
```

```
Original data:
 [[  12   15   73.35 2475.15    2   12  390   21    0.   150   73.35 2475.15
    3   62 3239]
 [  12  390  974.81 2868.15    2   12  203   10    0.     0  974.81 2868.15
    3   10 2701]
 [  72  203 6010.94 8684.80    8   72  494   30   45.  1950 6055.94 8639.80
    2   67 5372]
 [  12  494 2059.35 2059.35    0   12  234    4    0.     0 2059.35 2059.35
    2   86 2793]
 [  12  234  188.35  188.35    0   12  413   15    0.     0  188.35  188.35
    2   84 2235]]
Scaled data:
 [[0.16666667 0.01252086 0.01221001 0.34511153 0.28571429 0.16666667
 0.32441471 0.21       0.         0.15       0.01221001 0.34511153
 0.66666667 0.62       0.44766267]
 [0.16666667 0.32471261 0.16233567 0.39985376 0.28571429 0.16666667
 0.16888451 0.1        0.         0.         0.16233567 0.39985376
 0.66666667 0.1        0.37662488]
 [1.         0.16888451 0.83406592 1.         1.         1.
 0.41097421 0.3        0.45       1.         0.83406592 1.
 0.33333333 0.67       0.74916736]
 [0.16666667 0.41097421 0.28571429 0.28571429 0.         0.16666667
 0.19467221 0.04       0.         0         0.28571429 0.28571429
 0.33333333 0.86       0.38929288]
 [0.16666667 0.19467221 0.02611408 0.02611408 0.         0.16666667
 0.34359101 0.15       0.         0         0.02611408 0.02611408
 0.33333333 0.84       0.31153631]]
```

Mathematical Mistation:

Xnew = (Xold-Mean)/standard deviation

Exploratory Data Analysis

Log Transformation View Of Customer Loyalty

```
import matplotlib as mp
import numpy as np
import matplotlib.pyplot as plt
plt.plot(df['Customer Status'].value_counts(), df['Customer Status'].value_counts(),'ro')
plt.title('Log Transformed View Of Dataset')
```


Export to plotly »

The graph starts from 1.61, indicating first respondent's response. After that, it saw a steady rise till ninth respondent, and fall slightly at the 10th respondent's response. Indicating lesser customers. The graph then rises steadily till the 20th respondent and the line is crooked for all the respondents, indicating many less loyal customers. The last 9 customers seem to be quite loyal pulling the curve steadily upwards.

Outliers Detection

```
import numpy as np
import matplotlib.pyplot as plt
plt.boxplot(df['Tenure Months'])
plt.title('Tenure Months')
plt.show()
plt.boxplot(df['Total Long Distance Calls'])
plt.title('Total Long Distance Calls')
plt.show()
```

Intended of respondent

2 Kids and pets?

3 What is the Pole Cause of where you shop online from?

3 Since time using Pay and Shopping Online ?

4 How many times you have made an online purchase in the past 1 year?

7 How do you access the internet while shopping on-line?

8 Where device do you use to access the online shopping?

9 What is the screen size of your mobile device?

10 What is the operating system (OS) of your device?

11 What browser do you run on your device to access the website?

12 Which channel did you follow to arrive at your favorite online store for the first time?

13 After this visit, how do you count the online retail store?

14 How much time do you explore the e- retail store before making a purchase decision?

15 What is your preferred payment Option?

16 How frequently do you abandon (selecting an items and leaving without making payment) your shopping cart?

17 Why did you abandon the "Bag", "Shopping Cart"?

18 The content of the website must be easy to read and understand.

19 Information on similar product to the one highlighted is important for product comparison

20 Complete information on listed seller and product being offered is important for purchase decision.

21 All relevant information on listed products must be stated clearly.

22 Ease of navigation in website

23 Loading and processing speed

24 User Friendly Interface of the website

25 Customized Payment methods

26 Trust that the online retail store will fulfill its part of the transaction at the stipulated time

27 Empathy (readiness to assist with queries) towards the customers

28 Being able to guarantee the privacy of the customer

29 Responsiveness, availability of several communication channels (email, sender, sms, hotline, phone etc.)

30 Online shopping gives monetary benefit and discounts

31 Impersonal is derived from shopping online

32 Shopping online is convenient and flexible

33 Return and replacement policy of the e-store is inspiration for purchase decision

34 Gaining access to loyalty programs is a benefit of shopping online

35 Displaying quality information on the website improves satisfaction of customers

36 User dernier satisfaction while shopping on a good quality website or application

37 Net Benefit derived from shopping online can lead to users satisfaction

38 User satisfaction cannot exist without trust

39 Offering a wide variety of latest products in several category

40 Provision of complete and relevant product information

41 Monetary savings

42 The convenience of patronizing the online retailer

43 Shopping on the website gives you the sense of adventure

44 Shopping on your preferred e-tailer enhances your social status

45 You feel gratification shopping on use flexible trailer

46 Shopping on the website helps you fulfill certain roles

47 Gaining value for money spent

1.What do do you shop online from? ,encoded

2.In the following, that any (or all) of the online retailers you have shopped from? ,encoded

Easy to use website or application, encoded

Fresh appealing web page layout, encoded

Wide variety of product or offer, encoded

Complete, relevant description information of products, encoded

Fast loading website speed of website and application, encoded

Reliability of the website or application, encoded

Sportsman is a variety purchase_encoded

↓

Availability of several payment options_encoded

↓

Speedy order delivery _encoded

↓

Privacy of customer information_encoded

↓

Security of customer financial information_encoded

↓

Personal Transactions_encoded

↓

Presence of online assistance through crude channel_encoded

↓

Longer time to get logged in promotion_sales period_encoded

↓

Longer time in displaying graphics and photos promotion_sales period_encoded

↓

Late declaration of price promotion_sales period_encoded

↓

Longer page loading time promotion_sales period_encoded

↓

Limited mode of payment or even products promotion_sales period_encoded

↓

Longer delivery period_encoded

↓

Change in website/application design_encoded

↓

Frequent disruption when moving from one page to another_encoded

↓

Website is as efficient as before_encoded

↓

Most of the time online dealer would you recommend to a friend?_encoded

↓

There seem very few outliers in the data. Major disintegration in website/app design, Frequent disruption, Most of the time about websites, example, question 46. Response for value for money is equally positive by all the respondents. As a conclusion, customers are satisfied with the existing more or less. Because there is some outliers in the responses, Some additional things that can be done for customer loyalty is providing more product information and including more variety (because of some outliers are detected in these responses).

Group By View Of Data

```
[nan]
[nan]
[nan]
```

Question_1

Question_2

Question_3

1.1E-01 BERGOGLIO
1.1E-01 BERTONI
1.1E-01 CAMPOS
1.1E-01 CAMPOS CTA
1.1E-01 DEMPSAL
1.1E-01 FERNANDEZ
1.1E-01 FERNANDEZ CTA
1.1E-01 FUSCO
1.1E-01 FUSCO CTA
1.1E-01 GARCIA
1.1E-01 GARCIA CTA
1.1E-01 GONZALEZ
1.1E-01 GONZALEZ CTA
1.1E-01 MARTINEZ
1.1E-01 MARTINEZ CTA
1.1E-01 PEREZ
1.1E-01 PEREZ CTA
1.1E-01 RODRIGUEZ
1.1E-01 RODRIGUEZ CTA
1.1E-01 TOMAS
1.1E-01 TOMAS CTA
1.1E-01 VIDAL
1.1E-01 VIDAL CTA
1.1E-01 YAPOLSKY
1.1E-01 YAPOLSKY CTA
1.1E-01 ZAMORA
1.1E-01 ZAMORA CTA
1.1E-01 ZUCCHI
1.1E-01 ZUCCHI CTA
1.1E-01 ZUCCHI CTA
1.1E-01 ZUCCHI CTA

Question_28

0.0%
0.0%
0.0%

Question_56

Question_57

Question_58

Question_59

Question_60

Question_61

Question_62

Question_63

Question_64

Question_65

Question_66

Question_67

Question_68

Question_69

Question_70

Question_71

The above pie charts summarize responses of each question's single-select multiple-choice answer responses for that question. Question and responses alongside percentage coverage are pinned with each pie chart.

Scatter Plots

```
Figure(1,1,1,
```

There is seen rising correlation between loyalty and paired response for most of the questions. For example, gender, why did you abandon bag, shopping cart, responsiveness availability, product information, product quality, shipping, price. There is seen correlation between loyalty and paired response for most of the questions. For example, loading and processing speed, user friendly, ease of use, ease of navigation, product information, etcetera. There seems no correlation between loyalty and paired response for most of the questions. For example, how old are you, the content on the website must be easy to read and understand.

Histograms

```
numerical_paired_frame.shape[1]/3.0(int(numerical_paired_frame.shape[1]/3.0),
figsize=(20,
sharex=False,
sharey=True,
dpi=100,
facecolor='w',
edgecolor='k')
for i in range(0,numerical_paired_frame.shape[1]):
    ax = plt.subplot(numerical_paired_frame.shape[1]/3.0,3,i+1)
    ax.set_title(numerical_paired_frame.columns[i],
    numerical_paired_frame[numerical_paired_frame.columns[i]].hist(
    bins=10,
    alpha=0.5,
    color='red')
```

Whole data seems standard normally spread in the range -2 to +2

```
numerical_paired_frame.describe()
```

	Gender of respondent	How old are you?	1 Must be easy to read and understand	2 The content on the website must be easy to read and understand	3 Ease of navigation	4 Ease of use	5 User friendly	6 Loading and processing speed	7 Product information	8 Product quality	9 Shipping	10 Price	11 Responsiveness	12 Availability	13 Why did you abandon bag, shopping cart, or wishlist?	14 Why do you shop at this website?
count	1000.000	1000.000	1000.000	1000.000	1000.000	1000.000	1000.000	1000.000	1000.000	1000.000	1000.000	1000.000	1000.000	1000.000	1000.000	1000.000
mean	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000
std	0.500250	0.500250	0.500250	0.500250	0.500250	0.500250	0.500250	0.500250	0.500250	0.500250	0.500250	0.500250	0.500250	0.500250	0.500250	0.500250
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
50%	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000
75%	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
max	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000

2 rows × 15 columns

```
numerical_paired_frame[numerical_paired_frame[[0]].sort_values]
numerical_paired_frame.shape
```

```
(1000, 15)
Gender of respondent                                                    float64
How old are you?                                                        float64
1 Must be easy to read and understand                                   float64
2 The content on the website must be easy to read and understand        float64
3 Ease of navigation                                                    float64
4 Ease of use                                                           float64
5 User friendly                                                         float64
6 Loading and processing speed                                          float64
7 Product information                                                   float64
8 Product quality                                                       float64
9 Shipping                                                              float64
10 Price                                                                float64
11 Responsiveness                                                       float64
12 Availability                                                         float64
13 Why did you abandon bag, shopping cart, or wishlist?                 float64
14 Why do you shop at this website?                                     float64
dtype: object
```

The above table represents correlation among dataset. There seems some multicollinearity due to labels present in multicollinear label. Dealing with these by including y vector and ANOVA test can help us find most relevant features for prediction.

Correlation Heatmaps Of Entire Dataset

```
plt.figure(figsize=(20,20))
sns.heatmap(numerical_paired_frame.corr(), cmap='Reds', linewidth=0.1, linecolor='black')
plt.show()
```

Some of the multicollinearity pairs that can be spotted from this heatmap are q20 and q18, q22 and q23, q26 and q32, etcetera.

Descriptive Statistics

```
numerical_paired_frame.describe()
```

	Gender of respondent	How old are you?	1 Must be easy to read and understand	2 The content on the website must be easy to read and understand	3 Ease of navigation	4 Ease of use	5 User friendly	6 Loading and processing speed	7 Product information	8 Product quality	9 Shipping	10 Price	11 Responsiveness	12 Availability	13 Why did you abandon bag, shopping cart, or wishlist?	14 Why do you shop at this website?
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000
std	0.500250	0.500250	0.500250	0.500250	0.500250	0.500250	0.500250	0.500250	0.500250	0.500250	0.500250	0.500250	0.500250	0.500250	0.500250	0.500250
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
50%	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000	0.500000
75%	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
max	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000

2 rows × 15 columns

```
MULTICOLLINEAR_LABELS=[0,2,3,4,5,6,7,8,9,10,11,12,13,14]
numerical_paired_frame.corr()
```

```
Gender of respondent                                                    True
How old are you?                                                        True
1 Must be easy to read and understand                                   True
2 The content on the website must be easy to read and understand        True
3 Ease of navigation                                                    True
4 Ease of use                                                           True
5 User friendly                                                         True
6 Loading and processing speed                                          True
7 Product information                                                   True
8 Product quality                                                       True
9 Shipping                                                              True
10 Price                                                                True
11 Responsiveness                                                       True
12 Availability                                                         True
13 Why did you abandon bag, shopping cart, or wishlist?                 True
14 Why do you shop at this website?                                     True
dtype: object
```

```
CORRELATION_VALUES=[0,1]
numerical_paired_frame.corr()[numerical_paired_frame.columns[0]]
```

```
/usr/local/lib/python3.7/dist-packages/numpy/lib/function_base.py:2534: RuntimeWarning: invalid value encountered in true_divide
  c /= stddev[None, :]
/usr/local/lib/python3.7/dist-packages/numpy/lib/function_base.py:2535: RuntimeWarning: invalid value encountered in true_divide
  c /= stddev[:, None]
```

```
Gender of respondent                                                    1.000000
How old are you?                                                        0.000000
1 Must be easy to read and understand                                   0.000000
2 The content on the website must be easy to read and understand        0.000000
3 Ease of navigation                                                    0.000000
4 Ease of use                                                           0.000000
5 User friendly                                                         0.000000
6 Loading and processing speed                                          0.000000
7 Product information                                                   0.000000
8 Product quality                                                       0.000000
9 Shipping                                                              0.000000
10 Price                                                                0.000000
11 Responsiveness                                                       0.000000
12 Availability                                                         0.000000
13 Why did you abandon bag, shopping cart, or wishlist?                 0.000000
14 Why do you shop at this website?                                     0.000000
Name: Gender of respondent, dtype: float64
```

```
numerical_paired_frame[numerical_paired_frame.columns[CORRELATION_VALUES]]
```

	How old are you?	1 Must be easy to read and understand	2 The content on the website must be easy to read and understand	3 Ease of navigation	4 Ease of use	5 User friendly	6 Loading and processing speed	7 Product information	8 Product quality	9 Shipping	10 Price	11 Responsiveness	12 Availability	13 Why did you abandon bag, shopping cart, or wishlist?	14 Why do you shop at this website?
0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
3	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
4	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
5	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
6	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
7	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
8	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
9	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
10	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
11	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
12	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
13	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
14	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

15 rows × 15 columns

```
numerical_paired_frame[numerical_paired_frame.columns[CORRELATION_VALUES]].describe()
```

	How old are you?	1 Must be easy to read and understand	2 The content on the website must be easy to read and understand	3 Ease of navigation	4 Ease of use	5 User friendly	6 Loading and processing speed	7 Product information	8 Product quality	9 Shipping	10 Price	11 Responsiveness	12 Availability	13 Why did you abandon bag, shopping cart, or wishlist?	14 Why do you shop at this website?
count	15.000000	15.000000	15.000000	15.000000	15.000000	15.000000	15.000000	15.000000	15.000000	15.000000	15.000000	15.000000	15.000000	15.000000	15.000000
mean	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
std	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
50%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
75%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
max	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

```
numerical_paired_frame[numerical_paired_frame.columns[CORRELATION_VALUES]].shape
```

```
(15, 15)
```

```
numerical_paired_frame[numerical_paired_frame.columns[CORRELATION_VALUES]].dtypes
```

```
How old are you?                                                        float64
1 Must be easy to read and understand                                   float64
2 The content on the website must be easy to read and understand        float64
3 Ease of navigation                                                    float64
4 Ease of use                                                           float64
5 User friendly                                                         float64
6 Loading and processing speed                                          float64
7 Product information                                                   float64
8 Product quality                                                       float64
9 Shipping                                                              float64
10 Price                                                                float64
11 Responsiveness                                                       float64
12 Availability                                                         float64
13 Why did you abandon bag, shopping cart, or wishlist?                 float64
14 Why do you shop at this website?                                     float64
dtype: object
```

```
numerical_paired_frame[numerical_paired_frame.columns[CORRELATION_VALUES]].corr()
```

	How old are you?	1 Must be easy to read and understand	2 The content on the website must be easy to read and understand	3 Ease of navigation	4 Ease of use	5 User friendly	6 Loading and processing speed	7 Product information	8 Product quality	9 Shipping	10 Price	11 Responsiveness	12 Availability	13 Why did you abandon bag, shopping cart, or wishlist?	14 Why do you shop at this website?
How old are you?	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1 Must be easy to read and understand	NaN	1.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000
2 The content on the website must be easy to read and understand	NaN	-0.000000	1.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000
3 Ease of navigation	NaN	-0.000000	-0.000000	1.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000
4 Ease of use	NaN	-0.000000	-0.000000	-0.000000	1.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000
5 User friendly	NaN	-0.000000	-0.000000	-0.000000	-0.000000	1.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000
6 Loading and processing speed	NaN	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	1.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000
7 Product information	NaN	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	1.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000
8 Product quality	NaN	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	1.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000
9 Shipping	NaN	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	1.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000
10 Price	NaN	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	1.000000	-0.000000	-0.000000	-0.000000	-0.000000
11 Responsiveness	NaN	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	1.000000	-0.000000	-0.000000	-0.000000
12 Availability	NaN	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	1.000000	-0.000000	-0.000000
13 Why did you abandon bag, shopping cart, or wishlist?	NaN	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	1.000000	-0.000000
14 Why do you shop at this website?	NaN	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	-0.000000	1.000000

15 rows × 15 columns

```
Range   2.2E06  2.3E06  2.4E06  2.5E06  2.6E06  2.7E06  2.8E06  2.9E06  3.0E06  3.1E06
```

Code 75 Variables

```
Variable(s) ignored and imported: Customer_ID, Churn, Contract, tenure, PaperlessBilling, MonthlyCharges, TotalCharges
Factor variable(s) ignored and imported: PhoneService, InternetService, OnlineSecurity, OnlineBackup, DeviceProtection, TechSupport, StreamingTV, StreamingMovies, gender, SeniorCitizen, Partner, Dependents, MultipleLines, PaymentMethod
Character variable(s) ignored and imported: TotalCharges
Column(s) with all missing values: TotalCharges
Column(s) with zero variance: SeniorCitizen
Column(s) with near zero variance: PhoneService, MultipleLines, OnlineBackup, DeviceProtection, StreamingTV, StreamingMovies
Column(s) with low variance: gender, Partner, Dependents, PaperlessBilling
Column(s) with high skewness: tenure, MonthlyCharges
Column(s) with high kurtosis: tenure, MonthlyCharges
Column(s) with high missing values: OnlineSecurity, TechSupport
Column(s) with high correlation: OnlineSecurity, TechSupport
Column(s) with high collinearity: OnlineSecurity, TechSupport
Column(s) with high cardinality: PaymentMethod
Column(s) with high zero inflation: tenure, MonthlyCharges, TotalCharges
Column(s) with high outliers: tenure, MonthlyCharges, TotalCharges
Column(s) with high dispersion: tenure, MonthlyCharges, TotalCharges
Column(s) with low dispersion: SeniorCitizen
Column(s) with multicollinearity: MonthlyCharges, TotalCharges
Column(s) with linear dependency: MonthlyCharges, TotalCharges
Column(s) with non-linear dependency: tenure, MonthlyCharges, TotalCharges
Column(s) with low information: gender, SeniorCitizen, Partner, Dependents, PhoneService, MultipleLines, OnlineSecurity, OnlineBackup, DeviceProtection, TechSupport, StreamingTV, StreamingMovies, PaperlessBilling, PaymentMethod, MonthlyCharges, TotalCharges
Column(s) with low dispersion: SeniorCitizen
```

1. Skewness is 0 and within +/-0.65 for:

```
0 Enjoyable is service that shopping online  0.0004972
0 Displaying quality information on the website improves satisfaction of customers  0.0100887
1 Since they ship the shopping online I  0.0147006
0 Offering a wide variety of listed product in several category  0.0100400
0 Shipping on the website helps you fulfill certain needs  0.0040322
```

```
"At the frequently at you wonder interesting on items and leaving without being complete? you shopping online  0.0100887
```

```
1 Lack delivered of price sometimes, solve period_product_2  0.0100801
0 You feel gratification shopping at your favorite a tailor  0.0051162
0 Shopping on the website gives you the sense of adventure  0.0151581
0 Shopping at your preference I social otherwise your social circle  0.1002429
0 Shopping online is convenient and friendly  0.1000000
"Considering all other factors, I am likely to buy online soon"  0.1004000
1 Longer delivery period_product_2  0.0070229
"Frequent disruptive ads making them pop up to center product_2  0.0004731
1 Control lack of support or best products (promotion, solve period_product_2  0.0017123
1 Now not one get  0.0010392
```

2. Acceptable skewness is +/- 0.65 and skewness for bell shaped curve should be 0.

3. Right skewness is observed in:

```
1 Website is as efficient as before_product_2,0.0100400700010
1 Pricing of customerself information_product_2,0.7100002570400
"If day to you decide the Amazon's Wednesday Cyrcle !"
0.0100400700010
1 Availability of several payment options_product_2,0.7001010010010
1 Trust displaying web page layout_product_2,0.0010107070700
1 Complete, relevant description information of products_product_2,0.0100001010000
1 Speedy order delivery _product_2,1.0010010010007
0 Which device do you use to access the online shopping?,0.5310701010001
1 Wide variety of products on offer_product_2,1.0000710010700
"Considering all other factors, I am likely to buy online soon"
0.0100400700010
0 What is the free time of years you shop online from?,0.7001010010007
1 It would interest you and follow to online at your favorite online store for the first time?
0.0100400700010
"If day to you decide the Amazon's Wednesday Cyrcle !"
0.7100002570400
```

4. Left skewness is observed in:

```
0 Impacts creativeness to social with quantity towards the customers,0.2100010010007
"Is Secure and transparent ability of the retailer is important for purchase
intentions,0.0010010010007
0 The control on the website must be easy to read and understand,0.2301010010070
0 Professionals, specialists of several communication channels (email, online has, twitter, phone
etc.),0.0001001010700
1 Lack delivered of price sometimes,0.0010010010007
0 Lack of easy navigation of the website,0.0010010010001
0 Lack delivered of price sometimes shopping at a good quality website of
satisfaction,0.0000101010701
0 Wish satisfaction cannot solve without trust,0.0010700101071
0 Do not increase information or filled products must be stated clearly,0.7100101000101
0 Convenient Payment methods,0.0010700001000
0 Lack of and processing speed,0.0010100000100
0 Lack of and processing speed (out of your control device)
,0.0100700100000
0 Being able to guarantee the privacy of the customer,0.0107010700001
0 Trust that the online retail store will fulfill its part of the transaction or the stipulated
conditions,0.0000000000000
0 Remaining savings,0.0010700000101
0 Not Wasted personal time shopping online can lead to cyber satisfaction,0.1001000700107
0 Shopping online is convenient and friendly,0.1310001700001
0 Online shopping gives monetary benefits and discounts,0.0107010700101
0 Provision of complete and relevant product information,0.0010101000001
0 Complete information on filled order and product being offered is important for purchase
intentions,0.0000000000000
0 Saving money is equally important as a benefit of shopping online,0.0101000701070
0 The more time do you explore the e- retail store before making a purchase decision?
,0.0000000000000
0 Intelligence or creative product is the highlighter is important for product
intentions,0.0010701000101
Gender of respondents,0.7100000101001
"Longer page loading time (promotion, solve period_product_2,0.7001010010101
```

Conclusion

How can Data Science help in customer retention?

```
- Data science can help this dataset in classification problem to predict if the customer will
stay with the current seller.
- Various Machine learning and ensemble unsupervised machine learning techniques can be applied and
a best fitting model with maximum accuracy is available. Also, it can be processed further for CV and
hyper tuning.
```

What kind of predictions can data scientists make with this data and can this explain if the customers will stay with the company for a long time? In other terms, what characteristic have a more influence on product?

```
- Data science can help this dataset in classification problem to predict if the customer will
stay with the current seller.
- Various Machine learning and ensemble unsupervised machine learning techniques can be applied and
a best fitting model with maximum accuracy is available. Also, it can be processed further for CV and
hyper tuning.
- Feature Engineering to highlighter or Correlation Analysis, regression analysis and Descriptive
Analysis can be performed to see which feature is more influence on product. Also, Feature
Engineering can be performed to see if the current transformed features are better than the
features. Note that the new features are transformed features and not new features. Feature Engineering
can be performed to see if the current features are better than the current features.
- Of course and MORE analysis can help to further specifications improve and feature selection or
optimization methods. Finally, the model can be deployed on Microsoft Azure or the cloud.
```

Credits: Flipredo, DataTrained and Python Documentation.