

# Customer Churn Linear Regression

```
In [69]: import numpy as np
import pandas as pd
import scipy.stats
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn import linear_model
import statsmodels.api as sm
import statsmodels.formula.api as smf
from scipy import stats
```

## Analysis Goals:

The goal of this analysis is to determine which, if any, of the variables collected can help predict a customer's longevity with the company. Since relationships may not be immediately clear, I have decided to include a large number of variables in my initial model.

## Reading data

```
In [8]: data = pd.read_csv('C:/Users/cynth/OneDrive/Documents/MS Data Analytics/Clean_Data_Custo
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8950 entries, 0 to 8949
Data columns (total 49 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Unnamed: 0                            8950 non-null   int64
1   Lat                                    8950 non-null   float64
2   Lng                                    8950 non-null   float64
3   Population                            8950 non-null   int64
4   Children                              8950 non-null   int64
5   Age                                    8950 non-null   int64
6   Income                                8950 non-null   float64
7   Outage_sec_perweek                    8950 non-null   float64
8   Email                                  8950 non-null   int64
9   Contacts                              8950 non-null   int64
10  Yearly equip_failure                   8950 non-null   int64
11  Techie                                 8950 non-null   int64
12  Port_modem                             8950 non-null   int64
13  Tablet                                 8950 non-null   int64
14  Phone                                  8950 non-null   int64
15  Multiple                               8950 non-null   int64
16  OnlineSecurity                         8950 non-null   int64
17  OnlineBackup                           8950 non-null   int64
18  DeviceProtection                       8950 non-null   int64
19  TechSupport                            8950 non-null   int64
20  StreamingTV                            8950 non-null   int64
21  StreamingMovies                        8950 non-null   int64
22  PaperlessBilling                       8950 non-null   int64
23  Tenure                                 8950 non-null   float64
24  MonthlyCharge                          8950 non-null   float64
25  Bandwidth_GB_Year                     8950 non-null   float64
26  Item1                                  8950 non-null   int64
```

2/12/22, 11:54 PM

CustomerChurn\_LinearRegression

27	Item2	8950	non-null	int64
28	Item3	8950	non-null	int64
29	Item4	8950	non-null	int64
30	Item5	8950	non-null	int64
31	Item6	8950	non-null	int64
32	Item7	8950	non-null	int64
33	Item8	8950	non-null	int64
34	Area_Suburban	8950	non-null	int64
35	Area_Rural	8950	non-null	int64
36	Marital_Married	8950	non-null	int64
37	Marital_Separated	8950	non-null	int64
38	Marital_NeverMarried	8950	non-null	int64
39	Marital_Divorced	8950	non-null	int64
40	Gender_Female	8950	non-null	int64
41	Gender_Nonbinary	8950	non-null	int64
42	Contract_Oneyear	8950	non-null	int64
43	Contract_TwoYear	8950	non-null	int64
44	InternetService_DSL	8950	non-null	int64
45	InternetService_None	8950	non-null	int64
46	PaymentMethod_BankTransferautomatic	8950	non-null	int64
47	PaymentMethod_MailedCheck	8950	non-null	int64
48	PaymentMethod_ElectronicCheck	8950	non-null	int64

dtypes: float64(7), int64(42)

memory usage: 3.3 MB

# Performing Stepwise Backward Multiple Linear Regression

In [9]:

```
lm1 = smf.ols('Tenure ~ Lat+Lng+Population+Children+Age+Income+Outage_sec_perweek+Email')
lm1.params
print(lm1.summary())
```

OLS Regression Results						
=====						
Dep. Variable:	Tenure	R-squared:	1.000			
Model:	OLS	Adj. R-squared:	1.000			
Method:	Least Squares	F-statistic:	1.153e+07			
Date:	Sat, 12 Feb 2022	Prob (F-statistic):	0.00			
Time:	23:36:24	Log-Likelihood:	7287.4			
No. Observations:	8950	AIC:	-1.448e+04			
Df Residuals:	8902	BIC:	-1.414e+04			
Df Model:	47					
Covariance Type:	nonrobust					
=====						
=====						
		coef	std err	t	P> t	[0.
025	0.975]					
-----						
Intercept		1.1090	0.023	48.233	0.000	1.
064	1.154					
Lat		-7.134e-05	0.000	-0.291	0.771	-0.
001	0.000					
Lng		-4.473e-05	8.52e-05	-0.525	0.599	-0.
000	0.000					
Population		-4.127e-08	9.91e-08	-0.416	0.677	-2.36e
-07	1.53e-07					
Children		-0.3758	0.001	-623.691	0.000	-0.
377	-0.375					
Age		0.0400	5.52e-05	724.132	0.000	0.
040	0.040					

Income	-3.315e-08	4.53e-08	-0.731	0.465	-1.22e
-07 5.57e-08					
Outage_sec_perweek	0.0003	0.000	0.829	0.407	-0.
000 0.001					
Email	-0.0001	0.000	-0.351	0.726	-0.
001 0.001					
Contacts	-0.0003	0.001	-0.252	0.801	-0.
003 0.002					
Yearly equip_failure	-0.0020	0.002	-1.044	0.297	-0.
006 0.002					
Techie	0.0014	0.003	0.472	0.637	-0.
005 0.007					
Port_modem	0.0034	0.002	1.505	0.132	-0.
001 0.008					
Tablet	0.0003	0.002	0.123	0.902	-0.
005 0.005					
Phone	0.0002	0.004	0.043	0.965	-0.
007 0.008					
Multiple	0.2650	0.005	55.195	0.000	0.
256 0.274					
OnlineSecurity	-0.8309	0.002	-346.068	0.000	-0.
836 -0.826					
OnlineBackup	-0.3565	0.004	-96.335	0.000	-0.
364 -0.349					
DeviceProtection	-0.5972	0.003	-212.643	0.000	-0.
603 -0.592					
TechSupport	0.3828	0.003	133.496	0.000	0.
377 0.388					
StreamingTV	-1.3013	0.006	-219.647	0.000	-1.
313 -1.290					
StreamingMovies	-0.7267	0.007	-101.666	0.000	-0.
741 -0.713					
PaperlessBilling	-0.0045	0.002	-1.957	0.050	-0.
009 7.48e-06					
MonthlyCharge	-0.0351	0.000	-271.437	0.000	-0.
035 -0.035					
Bandwidth_GB_Year	0.0122	5.26e-07	2.32e+04	0.000	0.
012 0.012					
Item1	0.0024	0.002	1.458	0.145	-0.
001 0.006					
Item2	-0.0013	0.002	-0.817	0.414	-0.
004 0.002					
Item3	0.0012	0.001	0.878	0.380	-0.
002 0.004					
Item4	0.0010	0.001	0.794	0.427	-0.
001 0.003					
Item5	-0.0004	0.001	-0.289	0.772	-0.
003 0.002					
Item6	-0.0010	0.001	-0.767	0.443	-0.
004 0.002					
Item7	-0.0010	0.001	-0.814	0.415	-0.
004 0.001					
Item8	-1.331e-05	0.001	-0.011	0.991	-0.
002 0.002					
Area_Suburban	-0.0038	0.003	-1.350	0.177	-0.
009 0.002					
Area_Rural	0.0033	0.003	1.187	0.235	-0.
002 0.009					
Marital_Married	-0.0015	0.004	-0.425	0.671	-0.
009 0.006					
Marital_Separated	0.0027	0.004	0.742	0.458	-0.
004 0.010					
Marital_NeverMarried	-0.0003	0.004	-0.075	0.940	-0.
007 0.007					
Marital_Divorced	0.0003	0.004	0.076	0.940	-0.

```

007      0.007
Gender_Female      0.7924      0.002      343.195      0.000      0.
788      0.797
Gender_Nonbinary      1.0564      0.008      137.711      0.000      1.
041      1.071
Contract_Oneyear      0.0015      0.003      0.503      0.615      -0.
004      0.007
Contract_TwoYear      0.0011      0.003      0.403      0.687      -0.
004      0.007
InternetService_DSL      -5.7518      0.004      -1570.145      0.000      -5.
759      -5.745
InternetService_None      -1.1498      0.005      -221.595      0.000      -1.
160      -1.140
PaymentMethod_BankTransferautomatic      -0.0009      0.003      -0.245      0.806      -0.
008      0.006
PaymentMethod_MailedCheck      0.0061      0.003      1.767      0.077      -0.
001      0.013
PaymentMethod_ElectronicCheck      0.0017      0.003      0.530      0.596      -0.
005      0.008
=====
Omnibus:      31231.936      Durbin-Watson:      2.006
Prob(Omnibus):      0.000      Jarque-Bera (JB):      1459.017
Skew:      -0.029      Prob(JB):      0.00
Kurtosis:      1.023      Cond. No.      9.87e+05
=====

```

## Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 9.87e+05. This might indicate that there are strong multicollinearity or other numerical problems.

```

In [10]: #Remove item 8
lm2 = smf.ols('Tenure ~ Lat+Lng+Population+Children+Age+Income+Outage_sec_perweek+Email
lm2.params
print(lm2.summary())

```

## OLS Regression Results

```

=====
Dep. Variable:      Tenure      R-squared:      1.000
Model:      OLS      Adj. R-squared:      1.000
Method:      Least Squares      F-statistic:      1.178e+07
Date:      Sat, 12 Feb 2022      Prob (F-statistic):      0.00
Time:      23:36:24      Log-Likelihood:      7287.4
No. Observations:      8950      AIC:      -1.448e+04
Df Residuals:      8903      BIC:      -1.415e+04
Df Model:      46
Covariance Type:      nonrobust
=====

```

```

=====
coef      std err      t      P>|t|      [0.
025      0.975]
-----
Intercept      1.1089      0.023      48.562      0.000      1.
064      1.154
Lat      -7.128e-05      0.000      -0.291      0.771      -0.
001      0.000
Lng      -4.473e-05      8.51e-05      -0.525      0.599      -0.
000      0.000
Population      -4.127e-08      9.91e-08      -0.416      0.677      -2.36e
-07      1.53e-07
Children      -0.3758      0.001      -623.759      0.000      -0.

```

377	-0.375					
Age		0.0400	5.52e-05	724.246	0.000	0.
040	0.040					
Income		-3.315e-08	4.53e-08	-0.731	0.465	-1.22e
-07	5.57e-08					
Outage_sec_perweek		0.0003	0.000	0.829	0.407	-0.
000	0.001					
Email		-0.0001	0.000	-0.351	0.726	-0.
001	0.001					
Contacts		-0.0003	0.001	-0.252	0.801	-0.
003	0.002					
Yearly equip_failure		-0.0020	0.002	-1.044	0.297	-0.
006	0.002					
Techie		0.0014	0.003	0.472	0.637	-0.
005	0.007					
Port_modem		0.0034	0.002	1.505	0.132	-0.
001	0.008					
Tablet		0.0003	0.002	0.123	0.902	-0.
005	0.005					
Phone		0.0002	0.004	0.043	0.965	-0.
007	0.008					
Multiple		0.2650	0.005	55.198	0.000	0.
256	0.274					
OnlineSecurity		-0.8309	0.002	-346.099	0.000	-0.
836	-0.826					
OnlineBackup		-0.3565	0.004	-96.341	0.000	-0.
364	-0.349					
DeviceProtection		-0.5972	0.003	-212.666	0.000	-0.
603	-0.592					
TechSupport		0.3828	0.003	133.504	0.000	0.
377	0.388					
StreamingTV		-1.3013	0.006	-219.668	0.000	-1.
313	-1.290					
StreamingMovies		-0.7267	0.007	-101.674	0.000	-0.
741	-0.713					
PaperlessBilling		-0.0045	0.002	-1.957	0.050	-0.
009	6.49e-06					
MonthlyCharge		-0.0351	0.000	-271.454	0.000	-0.
035	-0.035					
Bandwidth_GB_Year		0.0122	5.26e-07	2.32e+04	0.000	0.
012	0.012					
Item1		0.0024	0.002	1.464	0.143	-0.
001	0.006					
Item2		-0.0013	0.002	-0.819	0.413	-0.
004	0.002					
Item3		0.0012	0.001	0.878	0.380	-0.
002	0.004					
Item4		0.0010	0.001	0.795	0.427	-0.
001	0.003					
Item5		-0.0004	0.001	-0.289	0.772	-0.
003	0.002					
Item6		-0.0010	0.001	-0.774	0.439	-0.
004	0.002					
Item7		-0.0010	0.001	-0.818	0.413	-0.
004	0.001					
Area_Suburban		-0.0038	0.003	-1.350	0.177	-0.
009	0.002					
Area_Rural		0.0033	0.003	1.187	0.235	-0.
002	0.009					
Marital_Married		-0.0015	0.004	-0.425	0.671	-0.
009	0.006					
Marital_Separated		0.0027	0.004	0.743	0.458	-0.
004	0.010					
Marital_NeverMarried		-0.0003	0.004	-0.075	0.940	-0.
007	0.007					

2/12/22, 11:54 PM

		CustomerChurn_LinearRegression				
Marital_Divorced		0.0003	0.004	0.076	0.939	-0.
007	0.007					
Gender_Female		0.7924	0.002	343.241	0.000	0.
788	0.797					
Gender_Nonbinary		1.0564	0.008	137.726	0.000	1.
041	1.071					
Contract_Oneyear		0.0015	0.003	0.504	0.615	-0.
004	0.007					
Contract_TwoYear		0.0011	0.003	0.404	0.686	-0.
004	0.007					
InternetService_DSL		-5.7518	0.004	-1570.349	0.000	-5.
759	-5.745					
InternetService_None		-1.1498	0.005	-221.611	0.000	-1.
160	-1.140					
PaymentMethod_BankTransferautomatic		-0.0009	0.003	-0.245	0.806	-0.
008	0.006					
PaymentMethod_MailedCheck		0.0061	0.003	1.767	0.077	-0.
001	0.013					
PaymentMethod_ElectronicCheck		0.0017	0.003	0.530	0.596	-0.
005	0.008					
=====						
Omnibus:	31231.935	Durbin-Watson:	2.006			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1459.017			
Skew:	-0.029	Prob(JB):	0.00			
Kurtosis:	1.023	Cond. No.	9.81e+05			
=====						

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 9.81e+05. This might indicate that there are strong multicollinearity or other numerical problems.

```
In [11]: #Remove phone
lm3 = smf.ols('Tenure ~ Lat+Lng+Population+Children+Age+Income+Outage_sec_perweek+Email
lm3.params
print(lm3.summary())
```

OLS Regression Results						
=====						
Dep. Variable:	Tenure	R-squared:	1.000			
Model:	OLS	Adj. R-squared:	1.000			
Method:	Least Squares	F-statistic:	1.204e+07			
Date:	Sat, 12 Feb 2022	Prob (F-statistic):	0.00			
Time:	23:36:24	Log-Likelihood:	7287.4			
No. Observations:	8950	AIC:	-1.448e+04			
Df Residuals:	8904	BIC:	-1.416e+04			
Df Model:	45					
Covariance Type:	nonrobust					
=====						
=====						
		coef	std err	t	P> t	[0.
025	0.975]					
-----						
Intercept		1.1091	0.023	49.209	0.000	1.
065	1.153					
Lat		-7.129e-05	0.000	-0.291	0.771	-0.
001	0.000					
Lng		-4.472e-05	8.51e-05	-0.525	0.599	-0.
000	0.000					
Population		-4.124e-08	9.91e-08	-0.416	0.677	-2.36e
-07	1.53e-07					

Children	-0.3758	0.001	-623.795	0.000	-0.
377	-0.375				
Age	0.0400	5.52e-05	724.346	0.000	0.
040	0.040				
Income	-3.316e-08	4.53e-08	-0.732	0.464	-1.22e
-07	5.57e-08				
Outage_sec_perweek	0.0003	0.000	0.829	0.407	-0.
000	0.001				
Email	-0.0001	0.000	-0.351	0.725	-0.
001	0.001				
Contacts	-0.0003	0.001	-0.252	0.801	-0.
003	0.002				
Yearly equip_failure	-0.0020	0.002	-1.045	0.296	-0.
006	0.002				
Techie	0.0014	0.003	0.472	0.637	-0.
005	0.007				
Port_modem	0.0034	0.002	1.506	0.132	-0.
001	0.008				
Tablet	0.0003	0.002	0.124	0.902	-0.
005	0.005				
Multiple	0.2650	0.005	55.215	0.000	0.
256	0.274				
OnlineSecurity	-0.8309	0.002	-346.123	0.000	-0.
836	-0.826				
OnlineBackup	-0.3565	0.004	-96.353	0.000	-0.
364	-0.349				
DeviceProtection	-0.5972	0.003	-212.693	0.000	-0.
603	-0.592				
TechSupport	0.3828	0.003	133.516	0.000	0.
377	0.388				
StreamingTV	-1.3013	0.006	-219.700	0.000	-1.
313	-1.290				
StreamingMovies	-0.7267	0.007	-101.694	0.000	-0.
741	-0.713				
PaperlessBilling	-0.0045	0.002	-1.958	0.050	-0.
009	5.91e-06				
MonthlyCharge	-0.0351	0.000	-271.524	0.000	-0.
035	-0.035				
Bandwidth_GB_Year	0.0122	5.26e-07	2.32e+04	0.000	0.
012	0.012				
Item1	0.0024	0.002	1.464	0.143	-0.
001	0.006				
Item2	-0.0013	0.002	-0.818	0.414	-0.
004	0.002				
Item3	0.0012	0.001	0.878	0.380	-0.
002	0.004				
Item4	0.0010	0.001	0.795	0.427	-0.
001	0.003				
Item5	-0.0004	0.001	-0.289	0.773	-0.
003	0.002				
Item6	-0.0010	0.001	-0.775	0.439	-0.
004	0.002				
Item7	-0.0010	0.001	-0.818	0.413	-0.
004	0.001				
Area_Suburban	-0.0038	0.003	-1.350	0.177	-0.
009	0.002				
Area_Rural	0.0033	0.003	1.188	0.235	-0.
002	0.009				
Marital_Married	-0.0015	0.004	-0.425	0.671	-0.
009	0.006				
Marital_Separated	0.0027	0.004	0.743	0.458	-0.
004	0.010				
Marital_NeverMarried	-0.0003	0.004	-0.075	0.940	-0.
007	0.007				
Marital_Divorced	0.0003	0.004	0.076	0.939	-0.

```

007      0.007
Gender_Female      0.7924      0.002      343.264      0.000      0.
788      0.797
Gender_Nonbinary      1.0564      0.008      137.734      0.000      1.
041      1.071
Contract_Oneyear      0.0015      0.003      0.504      0.614      -0.
004      0.007
Contract_TwoYear      0.0011      0.003      0.404      0.686      -0.
004      0.007
InternetService_DSL      -5.7518      0.004      -1570.585      0.000      -5.
759      -5.745
InternetService_None      -1.1498      0.005      -221.630      0.000      -1.
160      -1.140
PaymentMethod_BankTransferautomatic      -0.0009      0.003      -0.245      0.806      -0.
008      0.006
PaymentMethod_MailedCheck      0.0061      0.003      1.767      0.077      -0.
001      0.013
PaymentMethod_ElectronicCheck      0.0017      0.003      0.531      0.596      -0.
005      0.008
=====
Omnibus:      31231.907      Durbin-Watson:      2.006
Prob(Omnibus):      0.000      Jarque-Bera (JB):      1459.020
Skew:      -0.029      Prob(JB):      0.00
Kurtosis:      1.023      Cond. No.      9.70e+05
=====

```

## Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 9.7e+05. This might indicate that there are strong multicollinearity or other numerical problems.

```

In [12]: #Remove Marital_NeverMarried
lm5 = smf.ols('Tenure ~ Lat+Lng+Population+Children+Age+Income+Outage_sec_perweek+Email
lm5.params
print(lm5.summary())

```

## OLS Regression Results

```

=====
Dep. Variable:      Tenure      R-squared:      1.000
Model:      OLS      Adj. R-squared:      1.000
Method:      Least Squares      F-statistic:      1.232e+07
Date:      Sat, 12 Feb 2022      Prob (F-statistic):      0.00
Time:      23:36:24      Log-Likelihood:      7287.4
No. Observations:      8950      AIC:      -1.448e+04
Df Residuals:      8905      BIC:      -1.417e+04
Df Model:      44
Covariance Type:      nonrobust
=====

```

```

=====
coef      std err      t      P>|t|      [0.
025      0.975]
-----
Intercept      1.1090      0.022      49.332      0.000      1.
065      1.153
Lat      -7.14e-05      0.000      -0.292      0.771      -0.
001      0.000
Lng      -4.463e-05      8.51e-05      -0.524      0.600      -0.
000      0.000
Population      -4.132e-08      9.91e-08      -0.417      0.677      -2.36e
-07      1.53e-07
Children      -0.3758      0.001      -623.841      0.000      -0.

```



377	-0.375					
Age		0.0400	5.52e-05	724.397	0.000	0.
040	0.040					
Income		-3.317e-08	4.53e-08	-0.732	0.464	-1.22e
-07	5.57e-08					
Outage_sec_perweek		0.0003	0.000	0.830	0.407	-0.
000	0.001					
Email		-0.0001	0.000	-0.351	0.725	-0.
001	0.001					
Contacts		-0.0003	0.001	-0.252	0.801	-0.
003	0.002					
Yearly equip_failure		-0.0020	0.002	-1.045	0.296	-0.
006	0.002					
Techie		0.0014	0.003	0.474	0.636	-0.
005	0.007					
Port_modem		0.0034	0.002	1.506	0.132	-0.
001	0.008					
Tablet		0.0003	0.002	0.123	0.902	-0.
005	0.005					
Multiple		0.2650	0.005	55.221	0.000	0.
256	0.274					
OnlineSecurity		-0.8309	0.002	-346.148	0.000	-0.
836	-0.826					
OnlineBackup		-0.3565	0.004	-96.359	0.000	-0.
364	-0.349					
DeviceProtection		-0.5972	0.003	-212.707	0.000	-0.
603	-0.592					
TechSupport		0.3828	0.003	133.527	0.000	0.
377	0.388					
StreamingTV		-1.3013	0.006	-219.712	0.000	-1.
313	-1.290					
StreamingMovies		-0.7267	0.007	-101.702	0.000	-0.
741	-0.713					
PaperlessBilling		-0.0045	0.002	-1.958	0.050	-0.
009	4.5e-06					
MonthlyCharge		-0.0351	0.000	-271.542	0.000	-0.
035	-0.035					
Bandwidth_GB_Year		0.0122	5.26e-07	2.32e+04	0.000	0.
012	0.012					
Item1		0.0024	0.002	1.463	0.143	-0.
001	0.006					
Item2		-0.0013	0.002	-0.817	0.414	-0.
004	0.002					
Item3		0.0012	0.001	0.878	0.380	-0.
002	0.004					
Item4		0.0010	0.001	0.793	0.428	-0.
001	0.003					
Item5		-0.0004	0.001	-0.289	0.773	-0.
003	0.002					
Item6		-0.0010	0.001	-0.776	0.438	-0.
004	0.002					
Item7		-0.0010	0.001	-0.818	0.413	-0.
004	0.001					
Area_Suburban		-0.0038	0.003	-1.350	0.177	-0.
009	0.002					
Area_Rural		0.0033	0.003	1.187	0.235	-0.
002	0.009					
Marital_Married		-0.0014	0.003	-0.444	0.657	-0.
008	0.005					
Marital_Separated		0.0028	0.003	0.896	0.370	-0.
003	0.009					
Marital_Divorced		0.0004	0.003	0.131	0.896	-0.
006	0.006					
Gender_Female		0.7924	0.002	343.283	0.000	0.
788	0.797					

Variable	Estimate	Standard Error	t-value	p-value	Pr(> t )
Gender_Nonbinary041	1.071	1.0564	0.008	137.743	0.000
Contract_OneYear004	0.007	0.0015	0.003	0.504	0.614
Contract_TwoYear004	0.007	0.0011	0.003	0.404	0.686
InternetService_DSL759	-5.745	-5.7518	0.004	-1570.689	0.000
InternetService_None160	-1.140	-1.1498	0.005	-221.677	0.000
PaymentMethod_BankTransferautomatic008	0.006	-0.0009	0.003	-0.245	0.807
PaymentMethod_MailedCheck001	0.013	0.0061	0.003	1.768	0.077
PaymentMethod_ElectronicCheck005	0.008	0.0017	0.003	0.531	0.595

Omnibus:	31231.878	Durbin-Watson:	2.006
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1459.022
Skew:	-0.029	Prob(JB):	0.00
Kurtosis:	1.023	Cond. No.	9.68e+05

#### Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 9.68e+05. This might indicate that there are strong multicollinearity or other numerical problems.

## Initial Findings

The adjusted R2 in the models above, while theoretically possible, most likely indicated a potential issue with multicollinearity. As the correlation matrix below shows Bandwidth is almost a perfect match for Tenure and it may be causing issues with my model. The simple regression below indicates that Bandwidth\_GB\_Year accounts for 98.3% of the variance in Tenure. Since the goal of this project is to find indicators of Tenure, both categorical and continuous, I have elected to remove Bandwidth\_GB\_Year from the model to see if I can find any other potentially important indicators.

## Simple regression

The model below will be utilized as our baseline model. Currently, Bandwidth does an excellent job of predicting customer Tenure.

```
In [13]: #For comparison reasons only, not intended as final model. Final model is Lm32.
slm = smf.ols('Tenure ~ Bandwidth_GB_Year', data = data).fit()
slm.params
print(slm.summary())
```

OLS Regression Results			
Dep. Variable:	Tenure	R-squared:	0.983
Model:	OLS	Adj. R-squared:	0.983

2/12/22, 11:54 PM

CustomerChurn\_LinearRegression

Method:Least SquaresF-statistic:5.199e+05Date:Sat, 12 Feb 2022Prob (F-statistic):0.00Time:23:36:25Log-Likelihood:-23756.No. Observations:8950AIC:4.752e+04Df Residuals:8948BIC:4.753e+04Df Model:1Covariance Type:nonrobust

	coef	std err	t	P> t	[0.025	0.975]
Intercept	-6.1272	0.067	-91.494	0.000	-6.258	-5.996
Bandwidth_GB_Year	0.0120	1.66e-05	721.067	0.000	0.012	0.012

Omnibus:269.275Durbin-Watson:1.939Prob(Omnibus):0.000Jarque-Bera (JB):223.481Skew:-0.315Prob(JB):2.96e-49Kurtosis:2.551Cond. No.:7.41e+03

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 7.41e+03. This might indicate that there are strong multicollinearity or other numerical problems.

Removing bandwidth from model

```
In [14]: lm6 = smf.ols('Tenure ~ Lat+Lng+Population+Children+Age+Income+Outage_sec_perweek+Email
lm6.params
print(lm6.summary())
```

OLS Regression Results						
=====						
Dep. Variable:	Tenure	R-squared:	0.006			
Model:	OLS	Adj. R-squared:	0.001			
Method:	Least Squares	F-statistic:	1.146			
Date:	Sat, 12 Feb 2022	Prob (F-statistic):	0.231			
Time:	23:36:26	Log-Likelihood:	-41985.			
No. Observations:	8950	AIC:	8.406e+04			
Df Residuals:	8903	BIC:	8.440e+04			
Df Model:	46					
Covariance Type:	nonrobust					
=====						
=====						
		coef	std err	t	P> t	[0.
025	0.975]					
-----						
Intercept		42.0563	5.638	7.459	0.000	31.
004	53.109					
Lat		-0.1291	0.060	-2.143	0.032	-0.
247	-0.011					
Lng		0.0144	0.021	0.688	0.492	-0.
027	0.055					
Population		-4.18e-05	2.44e-05	-1.715	0.086	-8.96e
-05	5.98e-06					
Children		-0.1165	0.148	-0.786	0.432	-0.
407	0.174					
Age		0.0169	0.014	1.244	0.214	-0.
010	0.043					
Income		-8.544e-08	1.11e-05	-0.008	0.994	-2.19e
-05	2.18e-05					
Outage_sec_perweek		0.0470	0.096	0.490	0.624	-0.

141	0.235					
Email		-0.1583	0.093	-1.703	0.089	-0.
341	0.024					
Contacts		-0.0186	0.311	-0.060	0.952	-0.
628	0.591					
Yearly equip_failure		0.3869	0.481	0.805	0.421	-0.
555	1.329					
Techie		-1.0000	0.750	-1.334	0.182	-2.
470	0.470					
Port_modem		0.3444	0.560	0.615	0.539	-0.
753	1.442					
Tablet		0.0315	0.613	0.051	0.959	-1.
170	1.233					
Phone		0.4036	0.962	0.420	0.675	-1.
482	2.289					
Multiple		0.2610	1.181	0.221	0.825	-2.
054	2.576					
OnlineSecurity		0.1712	0.590	0.290	0.772	-0.
986	1.329					
OnlineBackup		1.5183	0.910	1.668	0.095	-0.
265	3.302					
DeviceProtection		-1.3732	0.691	-1.988	0.047	-2.
727	-0.019					
TechSupport		0.0344	0.705	0.049	0.961	-1.
348	1.417					
StreamingTV		1.0017	1.457	0.687	0.492	-1.
854	3.858					
StreamingMovies		1.2789	1.758	0.727	0.467	-2.
167	4.725					
PaperlessBilling		-0.0205	0.569	-0.036	0.971	-1.
136	1.095					
MonthlyCharge		-0.0218	0.032	-0.686	0.493	-0.
084	0.041					
Item1		-0.3289	0.405	-0.811	0.417	-1.
124	0.466					
Item2		0.1692	0.378	0.447	0.655	-0.
572	0.910					
Item3		0.1221	0.345	0.354	0.723	-0.
554	0.798					
Item4		0.3011	0.309	0.974	0.330	-0.
305	0.907					
Item5		0.5708	0.321	1.777	0.076	-0.
059	1.200					
Item6		-0.1786	0.331	-0.539	0.590	-0.
828	0.470					
Item7		0.2585	0.312	0.828	0.408	-0.
354	0.871					
Item8		-0.4478	0.298	-1.502	0.133	-1.
032	0.137					
Area_Suburban		-0.0661	0.686	-0.096	0.923	-1.
410	1.278					
Area_Rural		1.0449	0.687	1.522	0.128	-0.
301	2.391					
Marital_Married		0.2657	0.892	0.298	0.766	-1.
484	2.015					
Marital_Separated		-0.1246	0.879	-0.142	0.887	-1.
849	1.599					
Marital_NeverMarried		-0.0511	0.886	-0.058	0.954	-1.
788	1.685					
Marital_Divorced		-0.7450	0.873	-0.854	0.393	-2.
456	0.966					
Gender_Female		0.5886	0.568	1.037	0.300	-0.
525	1.702					
Gender_Nonbinary		-2.3209	1.886	-1.230	0.219	-6.
019	1.377					

2/12/22, 11:54 PM

		CustomerChurn_LinearRegression				
Contract_Oneyear		0.2086	0.718	0.290	0.772	-1.
199	1.616					
Contract_TwoYear		1.3074	0.681	1.919	0.055	-0.
028	2.643					
InternetService_DSL		-0.0014	0.899	-0.002	0.999	-1.
764	1.761					
InternetService_None		-1.1177	1.276	-0.876	0.381	-3.
619	1.384					
PaymentMethod_BankTransferautomatic		-1.1791	0.854	-1.381	0.167	-2.
853	0.495					
PaymentMethod_MailedCheck		0.0784	0.849	0.092	0.926	-1.
586	1.743					
PaymentMethod_ElectronicCheck		-0.8120	0.780	-1.041	0.298	-2.
341	0.717					
=====						
Omnibus:	37293.275	Durbin-Watson:	0.167			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1139.958			
Skew:	0.076	Prob(JB):	2.89e-248			
Kurtosis:	1.258	Cond. No.	9.83e+05			
=====						

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 9.83e+05. This might indicate that there are strong multicollinearity or other numerical problems.

In [15]:

#Remove InternetService\_DSL  
lm7 = smf.ols('Tenure ~ Lat+Lng+Population+Children+Age+Income+Outage\_sec\_perweek+Email  
lm7.params  
print(lm7.summary())

OLS Regression Results						
=====						
Dep. Variable:		Tenure	R-squared:		0.006	
Model:		OLS	Adj. R-squared:		0.001	
Method:		Least Squares	F-statistic:		1.172	
Date:		Sat, 12 Feb 2022	Prob (F-statistic):		0.201	
Time:		23:36:26	Log-Likelihood:		-41985.	
No. Observations:		8950	AIC:		8.406e+04	
Df Residuals:		8904	BIC:		8.439e+04	
Df Model:		45				
Covariance Type:		nonrobust				
=====						
=====						
		coef	std err	t	P> t	[0.
025	0.975]					
-----						
Intercept		42.0526	5.063	8.306	0.000	32.
128	51.978					
Lat		-0.1291	0.060	-2.143	0.032	-0.
247	-0.011					
Lng		0.0144	0.021	0.688	0.491	-0.
027	0.055					
Population		-4.18e-05	2.44e-05	-1.715	0.086	-8.96e
-05	5.98e-06					
Children		-0.1165	0.148	-0.786	0.432	-0.
407	0.174					
Age		0.0169	0.014	1.244	0.214	-0.
010	0.043					
Income		-8.579e-08	1.11e-05	-0.008	0.994	-2.19e
-05	2.18e-05					

Outage_sec_perweek	0.0470	0.096	0.490	0.624	-0.
141 0.235					
Email	-0.1583	0.093	-1.703	0.089	-0.
341 0.024					
Contacts	-0.0186	0.311	-0.060	0.952	-0.
628 0.591					
Yearly equip_failure	0.3869	0.481	0.805	0.421	-0.
555 1.329					
Techie	-1.0000	0.750	-1.334	0.182	-2.
470 0.470					
Port_modem	0.3444	0.560	0.615	0.539	-0.
753 1.442					
Tablet	0.0315	0.613	0.051	0.959	-1.
169 1.232					
Phone	0.4036	0.962	0.420	0.675	-1.
481 2.289					
Multiple	0.2598	0.923	0.281	0.778	-1.
550 2.070					
OnlineSecurity	0.1711	0.588	0.291	0.771	-0.
981 1.323					
OnlineBackup	1.5176	0.757	2.004	0.045	0.
033 3.002					
DeviceProtection	-1.3736	0.633	-2.171	0.030	-2.
614 -0.133					
TechSupport	0.0340	0.641	0.053	0.958	-1.
223 1.291					
StreamingTV	1.0002	1.101	0.908	0.364	-1.
158 3.159					
StreamingMovies	1.2771	1.304	0.979	0.327	-1.
279 3.833					
PaperlessBilling	-0.0205	0.569	-0.036	0.971	-1.
136 1.095					
MonthlyCharge	-0.0218	0.023	-0.967	0.333	-0.
066 0.022					
Item1	-0.3288	0.405	-0.811	0.417	-1.
123 0.466					
Item2	0.1691	0.378	0.447	0.655	-0.
572 0.910					
Item3	0.1221	0.345	0.354	0.723	-0.
554 0.798					
Item4	0.3011	0.309	0.974	0.330	-0.
305 0.907					
Item5	0.5708	0.321	1.778	0.075	-0.
059 1.200					
Item6	-0.1786	0.331	-0.539	0.590	-0.
828 0.470					
Item7	0.2585	0.312	0.828	0.408	-0.
354 0.871					
Item8	-0.4478	0.298	-1.502	0.133	-1.
032 0.137					
Area_Suburban	-0.0661	0.686	-0.096	0.923	-1.
410 1.278					
Area_Rural	1.0449	0.687	1.522	0.128	-0.
301 2.391					
Marital_Married	0.2657	0.892	0.298	0.766	-1.
484 2.015					
Marital_Separated	-0.1246	0.879	-0.142	0.887	-1.
848 1.599					
Marital_NeverMarried	-0.0511	0.886	-0.058	0.954	-1.
787 1.685					
Marital_Divorced	-0.7450	0.873	-0.854	0.393	-2.
456 0.966					
Gender_Female	0.5886	0.568	1.037	0.300	-0.
524 1.702					
Gender_Nonbinary	-2.3209	1.886	-1.230	0.219	-6.

```

018      1.377
Contract_Oneyear      0.2085      0.718      0.290      0.771      -1.
199      1.616
Contract_TwoYear      1.3074      0.681      1.920      0.055      -0.
028      2.642
InternetService_None      -1.1163      0.870      -1.283      0.200      -2.
822      0.589
PaymentMethod_BankTransferautomatic      -1.1791      0.854      -1.381      0.167      -2.
853      0.494
PaymentMethod_MailedCheck      0.0783      0.849      0.092      0.926      -1.
586      1.742
PaymentMethod_ElectronicCheck      -0.8120      0.780      -1.041      0.298      -2.
341      0.717
=====
Omnibus:      37293.265      Durbin-Watson:      0.167
Prob(Omnibus):      0.000      Jarque-Bera (JB):      1139.959
Skew:      0.076      Prob(JB):      2.89e-248
Kurtosis:      1.258      Cond. No.      8.53e+05
=====

```

## Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 8.53e+05. This might indicate that there are strong multicollinearity or other numerical problems.

```

In [16]: #Remove Income
lm8 = smf.ols('Tenure ~ Lat+Lng+Population+Children+Age+Outage_sec_perweek+Email+Contac
lm8.params
print(lm8.summary())

```

## OLS Regression Results

```

=====
Dep. Variable:      Tenure      R-squared:      0.006
Model:      OLS      Adj. R-squared:      0.001
Method:      Least Squares      F-statistic:      1.198
Date:      Sat, 12 Feb 2022      Prob (F-statistic):      0.173
Time:      23:36:26      Log-Likelihood:      -41985.
No. Observations:      8950      AIC:      8.406e+04
Df Residuals:      8905      BIC:      8.438e+04
Df Model:      44
Covariance Type:      nonrobust
=====
=====

```

	coef	std err	t	P> t	[0.
025      0.975]					
-----					
Intercept	42.0492	5.043	8.338	0.000	32.
163      51.935					
Lat	-0.1291	0.060	-2.144	0.032	-0.
247      -0.011					
Lng	0.0144	0.021	0.688	0.491	-0.
027      0.055					
Population	-4.18e-05	2.44e-05	-1.715	0.086	-8.96e
-05      5.98e-06					
Children	-0.1165	0.148	-0.786	0.432	-0.
407      0.174					
Age	0.0169	0.014	1.244	0.214	-0.
010      0.043					
Outage_sec_perweek	0.0470	0.096	0.490	0.624	-0.
141      0.235					
Email	-0.1583	0.093	-1.703	0.089	-0.

341	0.024					
Contacts		-0.0186	0.311	-0.060	0.952	-0.
628	0.591					
Yearly equip failure		0.3869	0.481	0.805	0.421	-0.
555	1.329					
Techie		-1.0000	0.750	-1.334	0.182	-2.
470	0.470					
Port_modem		0.3444	0.560	0.615	0.538	-0.
753	1.442					
Tablet		0.0315	0.613	0.051	0.959	-1.
169	1.232					
Phone		0.4036	0.962	0.420	0.675	-1.
481	2.288					
Multiple		0.2598	0.923	0.281	0.778	-1.
550	2.070					
OnlineSecurity		0.1711	0.587	0.291	0.771	-0.
981	1.323					
OnlineBackup		1.5176	0.757	2.004	0.045	0.
033	3.002					
DeviceProtection		-1.3737	0.633	-2.172	0.030	-2.
614	-0.134					
TechSupport		0.0339	0.641	0.053	0.958	-1.
223	1.291					
StreamingTV		1.0002	1.101	0.908	0.364	-1.
158	3.158					
StreamingMovies		1.2770	1.304	0.979	0.327	-1.
279	3.833					
PaperlessBilling		-0.0205	0.569	-0.036	0.971	-1.
136	1.095					
MonthlyCharge		-0.0218	0.023	-0.967	0.333	-0.
066	0.022					
Item1		-0.3288	0.405	-0.811	0.417	-1.
123	0.465					
Item2		0.1691	0.378	0.447	0.655	-0.
572	0.910					
Item3		0.1221	0.345	0.354	0.723	-0.
554	0.798					
Item4		0.3012	0.309	0.975	0.330	-0.
305	0.907					
Item5		0.5707	0.321	1.778	0.075	-0.
059	1.200					
Item6		-0.1786	0.331	-0.539	0.590	-0.
827	0.470					
Item7		0.2585	0.312	0.828	0.408	-0.
354	0.871					
Item8		-0.4478	0.298	-1.502	0.133	-1.
032	0.137					
Area_Suburban		-0.0662	0.686	-0.097	0.923	-1.
410	1.278					
Area_Rural		1.0449	0.687	1.522	0.128	-0.
301	2.391					
Marital_Married		0.2656	0.892	0.298	0.766	-1.
483	2.015					
Marital_Separated		-0.1246	0.879	-0.142	0.887	-1.
848	1.599					
Marital_NeverMarried		-0.0511	0.886	-0.058	0.954	-1.
787	1.685					
Marital_Divorced		-0.7450	0.873	-0.854	0.393	-2.
455	0.965					
Gender_Female		0.5885	0.568	1.037	0.300	-0.
524	1.701					
Gender_Nonbinary		-2.3210	1.886	-1.231	0.219	-6.
018	1.376					
Contract_Oneyear		0.2085	0.718	0.290	0.772	-1.
199	1.616					



Variable	Coef	Std Err	t	P> t	[0.025	0.975]
Contract_TwoYear	1.3074	0.681	1.920	0.055	-0.025	0.975]
028	2.642					
InternetService_None	-1.1162	0.870	-1.283	0.200	-2.025	0.792
822	0.589					
PaymentMethod_BankTransferautomatic	-1.1791	0.854	-1.381	0.167	-2.025	0.678
853	0.494					
PaymentMethod_MailedCheck	0.0782	0.849	0.092	0.927	-1.025	1.173
585	1.742					
PaymentMethod_ElectronicCheck	-0.8120	0.780	-1.041	0.298	-2.025	0.401
341	0.717					

Statistic	Value
Omnibus:	37293.179
Prob(Omnibus):	0.000
Skew:	0.076
Kurtosis:	1.258
Durbin-Watson:	0.167
Jarque-Bera (JB):	1139.962
Prob(JB):	2.89e-248
Cond. No.	2.66e+05

## Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 2.66e+05. This might indicate that there are strong multicollinearity or other numerical problems.

```
In [17]: #Remove Paperless Billing
lm9 = smf.ols('Tenure ~ Lat+Lng+Population+Children+Age+Outage_sec_perweek+Email+Contacts', data=customer_churn).fit()
lm9.params
print(lm9.summary())
```

## OLS Regression Results

Variable	Coef	Std Err	t	P> t	[0.025	0.975]
Dep. Variable:	Tenure					
Model:	OLS					
Method:	Least Squares					
Date:	Sat, 12 Feb 2022					
Time:	23:36:26					
No. Observations:	8950					
Df Residuals:	8906					
Df Model:	43					
Covariance Type:	nonrobust					
R-squared:	0.006					
Adj. R-squared:	0.001					
F-statistic:	1.226					
Prob (F-statistic):	0.148					
Log-Likelihood:	-41985.					
AIC:	8.406e+04					
BIC:	8.437e+04					

Variable	Coef	Std Err	t	P> t	[0.025	0.975]
Intercept	42.0406	5.037	8.346	0.000	32.025	52.057
Lat	-0.1291	0.060	-2.145	0.032	-0.248	0.000
Lng	0.0144	0.021	0.688	0.491	-0.027	0.055
Population	-4.18e-05	2.44e-05	-1.715	0.086	-8.96e-05	6e-05
Children	-0.1166	0.148	-0.787	0.431	-0.411	0.178
Age	0.0169	0.014	1.244	0.214	-0.010	0.043
Outage_sec_perweek	0.0470	0.096	0.491	0.624	-0.145	0.239
Email	-0.1583	0.093	-1.703	0.089	-0.344	0.027
Contacts	-0.0186	0.311	-0.060	0.952	-0.635	0.591

Yearly_equip_failure	0.3868	0.481	0.805	0.421	-0.
555 1.329					
Techie	-1.0000	0.750	-1.334	0.182	-2.
469 0.469					
Port_modem	0.3442	0.560	0.615	0.539	-0.
753 1.442					
Tablet	0.0310	0.612	0.051	0.960	-1.
170 1.231					
Phone	0.4037	0.961	0.420	0.675	-1.
481 2.288					
Multiple	0.2605	0.923	0.282	0.778	-1.
549 2.070					
OnlineSecurity	0.1711	0.587	0.291	0.771	-0.
980 1.323					
OnlineBackup	1.5177	0.757	2.005	0.045	0.
034 3.002					
DeviceProtection	-1.3737	0.633	-2.172	0.030	-2.
614 -0.134					
TechSupport	0.0341	0.641	0.053	0.958	-1.
222 1.291					
StreamingTV	1.0012	1.101	0.910	0.363	-1.
156 3.159					
StreamingMovies	1.2779	1.304	0.980	0.327	-1.
277 3.833					
MonthlyCharge	-0.0218	0.023	-0.968	0.333	-0.
066 0.022					
Item1	-0.3290	0.405	-0.812	0.417	-1.
123 0.465					
Item2	0.1691	0.378	0.447	0.655	-0.
572 0.910					
Item3	0.1223	0.345	0.355	0.723	-0.
554 0.798					
Item4	0.3011	0.309	0.975	0.330	-0.
305 0.907					
Item5	0.5707	0.321	1.778	0.075	-0.
059 1.200					
Item6	-0.1784	0.331	-0.539	0.590	-0.
827 0.470					
Item7	0.2583	0.312	0.827	0.408	-0.
354 0.870					
Item8	-0.4479	0.298	-1.502	0.133	-1.
032 0.136					
Area_Suburban	-0.0660	0.685	-0.096	0.923	-1.
410 1.278					
Area_Rural	1.0450	0.686	1.522	0.128	-0.
301 2.391					
Marital_Married	0.2657	0.892	0.298	0.766	-1.
483 2.015					
Marital_Separated	-0.1243	0.879	-0.141	0.888	-1.
848 1.599					
Marital_NeverMarried	-0.0513	0.886	-0.058	0.954	-1.
787 1.685					
Marital_Divorced	-0.7444	0.872	-0.853	0.393	-2.
454 0.966					
Gender_Female	0.5882	0.568	1.036	0.300	-0.
524 1.701					
Gender_Nonbinary	-2.3203	1.886	-1.230	0.219	-6.
017 1.376					
Contract_Oneyear	0.2087	0.718	0.291	0.771	-1.
199 1.616					
Contract_TwoYear	1.3070	0.681	1.919	0.055	-0.
028 2.642					
InternetService_None	-1.1167	0.870	-1.284	0.199	-2.
822 0.589					
PaymentMethod_BankTransferautomatic	-1.1791	0.854	-1.381	0.167	-2.

```

852      0.494
PaymentMethod_MailedCheck      0.0785      0.849      0.093      0.926      -1.
585      1.742
PaymentMethod_ElectronicCheck  -0.8122      0.780     -1.042      0.298      -2.
341      0.716
=====
Omnibus:      37293.395  Durbin-Watson:      0.167
Prob(Omnibus):      0.000  Jarque-Bera (JB):      1139.957
Skew:      0.076  Prob(JB):      2.89e-248
Kurtosis:      1.258  Cond. No.      2.66e+05
=====

```

## Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 2.66e+05. This might indicate that there are strong multicollinearity or other numerical problems.

```

In [18]: #Remove Tablet
lm10 = smf.ols('Tenure ~ Lat+Lng+Population+Children+Age+Outage_sec_perweek+Email+Conta
lm10.params
print(lm10.summary())

```

```

=====
                        OLS Regression Results
=====
Dep. Variable:          Tenure      R-squared:                0.006
Model:                  OLS         Adj. R-squared:           0.001
Method:                 Least Squares   F-statistic:             1.256
Date:                  Sat, 12 Feb 2022   Prob (F-statistic):       0.125
Time:                  23:36:26         Log-Likelihood:          -41985.
No. Observations:      8950            AIC:                    8.406e+04
Df Residuals:          8907            BIC:                    8.436e+04
Df Model:              42
Covariance Type:       nonrobust
=====
=====
                                coef      std err          t      P>|t|      [0.
025      0.975]
-----
Intercept                42.0520      5.032        8.357      0.000      32.
188      51.916
Lat                    -0.1291      0.060       -2.145      0.032      -0.
247      -0.011
Lng                     0.0144      0.021        0.689      0.491      -0.
027      0.055
Population              -4.181e-05    2.44e-05     -1.716      0.086     -8.96e
-05      5.96e-06
Children                -0.1166      0.148       -0.787      0.431      -0.
407      0.174
Age                     0.0169      0.014        1.244      0.214      -0.
010      0.043
Outage_sec_perweek      0.0470      0.096        0.491      0.624      -0.
141      0.235
Email                  -0.1583      0.093       -1.703      0.089      -0.
341      0.024
Contacts               -0.0186      0.311       -0.060      0.952      -0.
628      0.591
Yearly equip_failure    0.3867      0.481        0.805      0.421      -0.
555      1.329
Techie                 -0.9996      0.750       -1.334      0.182      -2.
469      0.470
Port_modem              0.3443      0.560        0.615      0.539      -0.

```

753	1.442					
Phone		0.4048	0.961	0.421	0.674	-1.
479	2.289					
Multiple		0.2603	0.923	0.282	0.778	-1.
549	2.070					
OnlineSecurity		0.1715	0.587	0.292	0.770	-0.
980	1.323					
OnlineBackup		1.5177	0.757	2.005	0.045	0.
034	3.002					
DeviceProtection		-1.3736	0.632	-2.172	0.030	-2.
613	-0.134					
TechSupport		0.0339	0.641	0.053	0.958	-1.
223	1.290					
StreamingTV		1.0023	1.100	0.911	0.362	-1.
155	3.159					
StreamingMovies		1.2792	1.303	0.982	0.326	-1.
275	3.834					
MonthlyCharge		-0.0218	0.023	-0.969	0.333	-0.
066	0.022					
Item1		-0.3287	0.405	-0.812	0.417	-1.
123	0.465					
Item2		0.1694	0.378	0.448	0.654	-0.
571	0.910					
Item3		0.1221	0.345	0.354	0.723	-0.
554	0.798					
Item4		0.3009	0.309	0.974	0.330	-0.
305	0.907					
Item5		0.5703	0.321	1.777	0.076	-0.
059	1.199					
Item6		-0.1786	0.331	-0.540	0.590	-0.
827	0.470					
Item7		0.2582	0.312	0.827	0.408	-0.
354	0.870					
Item8		-0.4478	0.298	-1.502	0.133	-1.
032	0.137					
Area_Suburban		-0.0662	0.685	-0.097	0.923	-1.
410	1.277					
Area_Rural		1.0444	0.686	1.522	0.128	-0.
301	2.390					
Marital_Married		0.2660	0.892	0.298	0.766	-1.
483	2.015					
Marital_Separated		-0.1234	0.879	-0.140	0.888	-1.
846	1.600					
Marital_NeverMarried		-0.0508	0.886	-0.057	0.954	-1.
787	1.685					
Marital_Divorced		-0.7440	0.872	-0.853	0.394	-2.
454	0.966					
Gender_Female		0.5885	0.567	1.037	0.300	-0.
524	1.701					
Gender_Nonbinary		-2.3204	1.886	-1.230	0.219	-6.
017	1.376					
Contract_OneYear		0.2092	0.718	0.291	0.771	-1.
198	1.616					
Contract_TwoYear		1.3071	0.681	1.920	0.055	-0.
028	2.642					
InternetService_None		-1.1167	0.870	-1.284	0.199	-2.
822	0.589					
PaymentMethod_BankTransferautomatic		-1.1793	0.854	-1.382	0.167	-2.
853	0.494					
PaymentMethod_MailedCheck		0.0784	0.849	0.092	0.926	-1.
585	1.742					
PaymentMethod_ElectronicCheck		-0.8117	0.780	-1.041	0.298	-2.
340	0.717					
=====						
Omnibus:		37293.033	Durbin-Watson:		0.167	

Prob(Omnibus):	0.000	Jarque-Bera (JB):	1139.972
Skew:	0.076	Prob(JB):	2.87e-248
Kurtosis:	1.258	Cond. No.	2.65e+05

## Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 2.65e+05. This might indicate that there are strong multicollinearity or other numerical problems.

```
In [19]: #Remove Tech support
lm11 = smf.ols('Tenure ~ Lat+Lng+Population+Children+Age+Outage_sec_perweek+Email+Contacts')
lm11.params
print(lm11.summary())
```

## OLS Regression Results

```
=====
Dep. Variable:          Tenure      R-squared:                0.006
Model:                  OLS         Adj. R-squared:           0.001
Method:                 Least Squares   F-statistic:             1.286
Date:                   Sat, 12 Feb 2022   Prob (F-statistic):      0.104
Time:                   23:36:26         Log-Likelihood:         -41985.
No. Observations:      8950            AIC:                   8.405e+04
Df Residuals:          8908            BIC:                   8.435e+04
Df Model:               41
Covariance Type:       nonrobust
=====
```

		coef	std err	t	P> t	[0.
025	0.975]					
-----						
Intercept		42.0106	4.971	8.452	0.000	32.
267	51.754					
Lat		-0.1291	0.060	-2.146	0.032	-0.
247	-0.011					
Lng		0.0144	0.021	0.688	0.491	-0.
027	0.055					
Population		-4.183e-05	2.44e-05	-1.717	0.086	-8.96e
-05	5.93e-06					
Children		-0.1166	0.148	-0.787	0.431	-0.
407	0.174					
Age		0.0169	0.014	1.245	0.213	-0.
010	0.043					
Outage_sec_perweek		0.0470	0.096	0.491	0.624	-0.
141	0.235					
Email		-0.1583	0.093	-1.703	0.089	-0.
340	0.024					
Contacts		-0.0188	0.311	-0.060	0.952	-0.
628	0.591					
Yearly equip_failure		0.3869	0.481	0.805	0.421	-0.
555	1.329					
Techie		-0.9995	0.750	-1.334	0.182	-2.
469	0.470					
Port_modem		0.3446	0.560	0.616	0.538	-0.
753	1.442					
Phone		0.4050	0.961	0.421	0.674	-1.
479	2.289					
Multiple		0.2434	0.866	0.281	0.779	-1.
454	1.941					
OnlineSecurity		0.1705	0.587	0.290	0.772	-0.
980	1.321					

OnlineBackup	1.5062	0.725	2.078	0.038	0.
086 2.927					
DeviceProtection	-1.3802	0.620	-2.226	0.026	-2.
595 -0.165					
StreamingTV	0.9807	1.022	0.960	0.337	-1.
022 2.984					
StreamingMovies	1.2521	1.198	1.045	0.296	-1.
096 3.600					
MonthlyCharge	-0.0213	0.020	-1.048	0.295	-0.
061 0.019					
Item1	-0.3283	0.405	-0.811	0.418	-1.
122 0.466					
Item2	0.1691	0.378	0.448	0.655	-0.
571 0.910					
Item3	0.1223	0.345	0.355	0.723	-0.
554 0.798					
Item4	0.3011	0.309	0.975	0.330	-0.
304 0.907					
Item5	0.5704	0.321	1.778	0.076	-0.
059 1.199					
Item6	-0.1787	0.331	-0.540	0.589	-0.
827 0.470					
Item7	0.2584	0.312	0.828	0.408	-0.
353 0.870					
Item8	-0.4477	0.298	-1.502	0.133	-1.
032 0.137					
Area_Suburban	-0.0665	0.685	-0.097	0.923	-1.
410 1.277					
Area_Rural	1.0442	0.686	1.521	0.128	-0.
301 2.389					
Marital_Married	0.2664	0.892	0.299	0.765	-1.
482 2.015					
Marital_Separated	-0.1236	0.879	-0.141	0.888	-1.
847 1.599					
Marital_NeverMarried	-0.0505	0.885	-0.057	0.954	-1.
786 1.685					
Marital_Divorced	-0.7438	0.872	-0.853	0.394	-2.
454 0.966					
Gender_Female	0.5884	0.567	1.037	0.300	-0.
524 1.701					
Gender_Nonbinary	-2.3195	1.886	-1.230	0.219	-6.
016 1.377					
Contract_Oneyear	0.2094	0.718	0.292	0.770	-1.
198 1.616					
Contract_TwoYear	1.3070	0.681	1.920	0.055	-0.
028 2.642					
InternetService_None	-1.1042	0.837	-1.319	0.187	-2.
745 0.537					
PaymentMethod_BankTransferautomatic	-1.1789	0.854	-1.381	0.167	-2.
852 0.494					
PaymentMethod_MailedCheck	0.0786	0.848	0.093	0.926	-1.
585 1.742					
PaymentMethod_ElectronicCheck	-0.8113	0.780	-1.041	0.298	-2.
340 0.717					

```

=====
Omnibus:          37293.201  Durbin-Watson:          0.167
Prob(Omnibus):    0.000  Jarque-Bera (JB):        1139.969
Skew:             0.076  Prob(JB):                2.88e-248
Kurtosis:         1.258  Cond. No.                2.61e+05
=====

```

## Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 2.61e+05. This might indicate that there are strong multicollinearity or other numerical problems.

```
In [20]: #Remove Marital Never Married
lm12 = smf.ols('Tenure ~ Lat+Lng+Population+Children+Age+Outage_sec_perweek+Email+Conta
lm12.params
print(lm12.summary())
```

```

                                OLS Regression Results
=====
Dep. Variable:                  Tenure      R-squared:                  0.006
Model:                            OLS      Adj. R-squared:              0.001
Method:                 Least Squares      F-statistic:                  1.318
Date:                 Sat, 12 Feb 2022      Prob (F-statistic):          0.0862
Time:                   23:36:26      Log-Likelihood:             -41985.
No. Observations:                8950      AIC:                        8.405e+04
Df Residuals:                    8909      BIC:                        8.434e+04
Df Model:                          40
Covariance Type:                  nonrobust
=====
=====
                                coef      std err          t      P>|t|      [0.
025      0.975]
-----
-----
Intercept                        41.9878      4.954         8.475      0.000      32.
276      51.699
Lat                          -0.1291      0.060        -2.146      0.032      -0.
247      -0.011
Lng                           0.0144      0.021         0.689      0.491      -0.
027      0.055
Population                     -4.185e-05      2.44e-05      -1.718      0.086      -8.96e
-05      5.91e-06
Children                       -0.1165      0.148        -0.787      0.431      -0.
407      0.174
Age                           0.0169      0.014         1.245      0.213      -0.
010      0.043
Outage_sec_perweek              0.0470      0.096         0.491      0.623      -0.
141      0.235
Email                         -0.1582      0.093        -1.703      0.089      -0.
340      0.024
Contacts                      -0.0187      0.311        -0.060      0.952      -0.
628      0.591
Yearly equip_failure            0.3870      0.481         0.805      0.421      -0.
555      1.329
Techie                        -0.9984      0.749        -1.333      0.183      -2.
467      0.470
Port_modem                     0.3445      0.560         0.616      0.538      -0.
753      1.442
Phone                          0.4055      0.961         0.422      0.673      -1.
478      2.289
Multiple                       0.2429      0.866         0.281      0.779      -1.
454      1.940
OnlineSecurity                 0.1703      0.587         0.290      0.772      -0.
980      1.321
OnlineBackup                   1.5063      0.725         2.079      0.038      0.
086      2.927
DeviceProtection              -1.3799      0.620        -2.226      0.026      -2.
595      -0.165
StreamingTV                    0.9811      1.022         0.960      0.337      -1.
022      2.984
StreamingMovies                1.2520      1.198         1.045      0.296      -1.
096      3.600

```

MonthlyCharge	-0.0213	0.020	-1.048	0.295	-0.
061 0.019					
Item1	-0.3285	0.405	-0.811	0.417	-1.
122 0.465					
Item2	0.1692	0.378	0.448	0.654	-0.
571 0.910					
Item3	0.1224	0.345	0.355	0.723	-0.
554 0.798					
Item4	0.3007	0.309	0.974	0.330	-0.
305 0.906					
Item5	0.5705	0.321	1.778	0.075	-0.
058 1.199					
Item6	-0.1790	0.331	-0.541	0.589	-0.
828 0.470					
Item7	0.2583	0.312	0.828	0.408	-0.
353 0.870					
Item8	-0.4475	0.298	-1.502	0.133	-1.
032 0.137					
Area_Suburban	-0.0663	0.685	-0.097	0.923	-1.
410 1.277					
Area_Rural	1.0440	0.686	1.521	0.128	-0.
301 2.389					
Marital_Married	0.2911	0.780	0.373	0.709	-1.
238 1.820					
Marital_Separated	-0.0988	0.765	-0.129	0.897	-1.
598 1.400					
Marital_Divorced	-0.7190	0.756	-0.950	0.342	-2.
202 0.764					
Gender_Female	0.5884	0.567	1.037	0.300	-0.
524 1.701					
Gender_Nonbinary	-2.3192	1.885	-1.230	0.219	-6.
015 1.377					
Contract_Oneyear	0.2092	0.718	0.292	0.771	-1.
198 1.616					
Contract_TwoYear	1.3071	0.681	1.920	0.055	-0.
027 2.642					
InternetService_None	-1.1032	0.837	-1.318	0.187	-2.
744 0.537					
PaymentMethod_BankTransferautomatic	-1.1787	0.853	-1.381	0.167	-2.
852 0.494					
PaymentMethod_MailedCheck	0.0793	0.848	0.094	0.925	-1.
584 1.742					
PaymentMethod_ElectronicCheck	-0.8109	0.780	-1.040	0.298	-2.
339 0.717					

```

=====
Omnibus:                37293.253    Durbin-Watson:           0.167
Prob(Omnibus):          0.000    Jarque-Bera (JB):       1139.968
Skew:                   0.076    Prob(JB):               2.88e-248
Kurtosis:               1.258    Cond. No.                2.60e+05
=====

```

## Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 2.6e+05. This might indicate that there are strong multicollinearity or other numerical problems.

```

In [21]: #Remove Contacts
lm13 = smf.ols('Tenure ~ Lat+Lng+Population+Children+Age+Outage_sec_perweek+Email+Year1
lm13.params
print(lm13.summary())

```

## OLS Regression Results



```

=====
Dep. Variable:          Tenure      R-squared:                0.006
Model:                  OLS         Adj. R-squared:           0.002
Method:                 Least Squares  F-statistic:             1.352
Date:                  Sat, 12 Feb 2022  Prob (F-statistic):       0.0704
Time:                  23:36:26      Log-Likelihood:          -41985.
No. Observations:      8950         AIC:                     8.405e+04
Df Residuals:          8910         BIC:                     8.433e+04
Df Model:               39
Covariance Type:       nonrobust
=====

```

```

=====
                                coef    std err          t      P>|t|      [0.
025      0.975]
-----
Intercept                41.9688      4.944      8.489      0.000      32.
278      51.660
Lat                    -0.1291      0.060     -2.146      0.032      -0.
247     -0.011
Lng                     0.0144      0.021      0.689      0.491      -0.
027      0.055
Population              -4.185e-05    2.44e-05     -1.718      0.086     -8.96e
-05      5.89e-06
Children              -0.1163      0.148     -0.785      0.432      -0.
406      0.174
Age                    0.0169      0.014      1.244      0.213      -0.
010      0.043
Outage_sec_perweek      0.0469      0.096      0.490      0.624      -0.
141      0.235
Email                 -0.1583      0.093     -1.703      0.089      -0.
340      0.024
Yearly_equip_failure      0.3869      0.481      0.805      0.421      -0.
555      1.329
Techie                 -0.9985      0.749     -1.333      0.183      -2.
467      0.470
Port_modem             0.3446      0.560      0.616      0.538      -0.
752      1.442
Phone                  0.4054      0.961      0.422      0.673      -1.
478      2.289
Multiple              0.2425      0.866      0.280      0.779      -1.
455      1.940
OnlineSecurity          0.1700      0.587      0.290      0.772      -0.
981      1.321
OnlineBackup           1.5057      0.725      2.078      0.038      0.
085      2.926
DeviceProtection       -1.3798      0.620     -2.226      0.026      -2.
595     -0.165
StreamingTV            0.9806      1.022      0.960      0.337      -1.
022      2.983
StreamingMovies         1.2510      1.198      1.045      0.296      -1.
097      3.599
MonthlyCharge          -0.0213      0.020     -1.048      0.295      -0.
061      0.019
Item1                  -0.3285      0.405     -0.811      0.417      -1.
122      0.465
Item2                  0.1689      0.378      0.447      0.655      -0.
571      0.909
Item3                  0.1228      0.345      0.356      0.722      -0.
553      0.799
Item4                  0.3006      0.309      0.973      0.330      -0.
305      0.906
Item5                  0.5702      0.321      1.778      0.076      -0.
059      1.199
Item6                 -0.1790      0.331     -0.541      0.589      -0.

```

828	0.470					
Item7		0.2580	0.312	0.827	0.408	-0.
354	0.870					
Item8		-0.4471	0.298	-1.501	0.133	-1.
031	0.137					
Area_Suburban		-0.0661	0.685	-0.096	0.923	-1.
409	1.277					
Area_Rural		1.0440	0.686	1.521	0.128	-0.
301	2.389					
Marital_Married		0.2910	0.780	0.373	0.709	-1.
238	1.820					
Marital_Separated		-0.0980	0.765	-0.128	0.898	-1.
597	1.401					
Marital_Divorced		-0.7186	0.756	-0.950	0.342	-2.
201	0.764					
Gender_Female		0.5888	0.567	1.038	0.299	-0.
523	1.701					
Gender_Nonbinary		-2.3183	1.885	-1.230	0.219	-6.
014	1.377					
Contract_OneYear		0.2088	0.718	0.291	0.771	-1.
198	1.615					
Contract_TwoYear		1.3066	0.681	1.919	0.055	-0.
028	2.641					
InternetService_None		-1.1024	0.837	-1.318	0.188	-2.
743	0.538					
PaymentMethod_BankTransferautomatic		-1.1791	0.853	-1.382	0.167	-2.
852	0.494					
PaymentMethod_MailedCheck		0.0798	0.848	0.094	0.925	-1.
583	1.743					
PaymentMethod_ElectronicCheck		-0.8107	0.780	-1.040	0.298	-2.
339	0.717					

```
=====
Omnibus:                37293.638    Durbin-Watson:           0.167
Prob(Omnibus):           0.000    Jarque-Bera (JB):       1139.955
Skew:                    0.076    Prob(JB):               2.90e-248
Kurtosis:                1.258    Cond. No.               2.60e+05
=====
```

## Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 2.6e+05. This might indicate that there are strong multicollinearity or other numerical problems.

```
In [22]: #Remove Mailed Check
lm13 = smf.ols('Tenure ~ Lat+Lng+Population+Children+Age+Outage_sec_perweek+Email+Year1
lm13.params
print(lm13.summary())
```

## OLS Regression Results

```
=====
Dep. Variable:          Tenure    R-squared:                0.006
Model:                  OLS       Adj. R-squared:           0.002
Method:                 Least Squares    F-statistic:             1.388
Date:                   Sat, 12 Feb 2022    Prob (F-statistic):      0.0570
Time:                   23:36:27          Log-Likelihood:         -41985.
No. Observations:       8950            AIC:                   8.405e+04
Df Residuals:           8911            BIC:                   8.432e+04
Df Model:               38
Covariance Type:        nonrobust
=====
```

```
=====
coef    std err          t    P>|t|    [0.
```

025	0.975]					
-----						
Intercept		42.0207	4.913	8.553	0.000	32.
390	51.651					
Lat		-0.1292	0.060	-2.147	0.032	-0.
247	-0.011					
Lng		0.0144	0.021	0.690	0.490	-0.
027	0.055					
Population		-4.189e-05	2.44e-05	-1.720	0.086	-8.96e
-05	5.86e-06					
Children		-0.1163	0.148	-0.785	0.432	-0.
406	0.174					
Age		0.0169	0.014	1.243	0.214	-0.
010	0.043					
Outage_sec_perweek		0.0468	0.096	0.489	0.625	-0.
141	0.234					
Email		-0.1584	0.093	-1.704	0.088	-0.
341	0.024					
Yearly equip failure		0.3861	0.480	0.804	0.422	-0.
556	1.328					
Techie		-0.9994	0.749	-1.334	0.182	-2.
468	0.469					
Port_modem		0.3447	0.560	0.616	0.538	-0.
752	1.442					
Phone		0.4053	0.961	0.422	0.673	-1.
478	2.289					
Multiple		0.2427	0.866	0.280	0.779	-1.
454	1.940					
OnlineSecurity		0.1699	0.587	0.289	0.772	-0.
981	1.320					
OnlineBackup		1.5059	0.724	2.079	0.038	0.
086	2.926					
DeviceProtection		-1.3796	0.620	-2.226	0.026	-2.
595	-0.165					
StreamingTV		0.9810	1.022	0.960	0.337	-1.
022	2.984					
StreamingMovies		1.2512	1.198	1.045	0.296	-1.
096	3.599					
MonthlyCharge		-0.0213	0.020	-1.048	0.294	-0.
061	0.019					
Item1		-0.3289	0.405	-0.812	0.417	-1.
123	0.465					
Item2		0.1687	0.378	0.447	0.655	-0.
572	0.909					
Item3		0.1231	0.345	0.357	0.721	-0.
553	0.799					
Item4		0.3003	0.309	0.973	0.331	-0.
305	0.905					
Item5		0.5701	0.321	1.777	0.076	-0.
059	1.199					
Item6		-0.1784	0.331	-0.539	0.590	-0.
827	0.470					
Item7		0.2579	0.312	0.826	0.409	-0.
354	0.869					
Item8		-0.4470	0.298	-1.500	0.134	-1.
031	0.137					
Area_Suburban		-0.0661	0.685	-0.096	0.923	-1.
409	1.277					
Area_Rural		1.0443	0.686	1.522	0.128	-0.
301	2.389					
Marital_Married		0.2910	0.780	0.373	0.709	-1.
238	1.820					
Marital_Separated		-0.0965	0.764	-0.126	0.899	-1.
595	1.402					

2/12/22, 11:54 PM

		CustomerChurn_LinearRegression				
Marital_Divorced		-0.7169	0.756	-0.948	0.343	-2.
199	0.765					
Gender_Female		0.5898	0.567	1.040	0.298	-0.
522	1.702					
Gender_Nonbinary		-2.3184	1.885	-1.230	0.219	-6.
014	1.377					
Contract_Oneyear		0.2097	0.718	0.292	0.770	-1.
197	1.616					
Contract_TwoYear		1.3067	0.681	1.920	0.055	-0.
028	2.641					
InternetService_None		-1.1027	0.837	-1.318	0.188	-2.
743	0.537					
PaymentMethod_BankTransferautomatic		-1.2209	0.728	-1.677	0.094	-2.
648	0.207					
PaymentMethod_ElectronicCheck		-0.8526	0.640	-1.332	0.183	-2.
107	0.402					
=====						
Omnibus:	37293.205	Durbin-Watson:	0.167			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1139.966			
Skew:	0.076	Prob(JB):	2.88e-248			
Kurtosis:	1.258	Cond. No.	2.58e+05			
=====						

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 2.58e+05. This might indicate that there are strong multicollinearity or other numerical problems.

In [23]:

#Remove Area Suburban  
lm14 = smf.ols('Tenure ~ Lat+Lng+Population+Children+Age+Outage\_sec\_perweek+Email+Year1  
lm14.params  
print(lm14.summary())

OLS Regression Results						
=====						
Dep. Variable:	Tenure	R-squared:	0.006			
Model:	OLS	Adj. R-squared:	0.002			
Method:	Least Squares	F-statistic:	1.425			
Date:	Sat, 12 Feb 2022	Prob (F-statistic):	0.0455			
Time:	23:36:27	Log-Likelihood:	-41985.			
No. Observations:	8950	AIC:	8.405e+04			
Df Residuals:	8912	BIC:	8.431e+04			
Df Model:	37					
Covariance Type:	nonrobust					
=====						
		coef	std err	t	P> t	[0.
025	0.975]					
-----						
Intercept		41.9841	4.898	8.572	0.000	32.
383	51.585					
Lat		-0.1291	0.060	-2.147	0.032	-0.
247	-0.011					
Lng		0.0144	0.021	0.690	0.490	-0.
027	0.055					
Population		-4.192e-05	2.44e-05	-1.722	0.085	-8.97e
-05	5.81e-06					
Children		-0.1163	0.148	-0.786	0.432	-0.
406	0.174					
Age		0.0169	0.014	1.243	0.214	-0.
010	0.043					

Outage_sec_perweek	0.0468	0.096	0.489	0.625	-0.
141 0.234					
Email	-0.1583	0.093	-1.704	0.088	-0.
340 0.024					
Yearly equip_failure	0.3857	0.480	0.803	0.422	-0.
556 1.327					
Techie	-0.9992	0.749	-1.334	0.182	-2.
468 0.469					
Port_modem	0.3441	0.560	0.615	0.539	-0.
753 1.441					
Phone	0.4046	0.961	0.421	0.674	-1.
479 2.288					
Multiple	0.2426	0.866	0.280	0.779	-1.
454 1.940					
OnlineSecurity	0.1690	0.587	0.288	0.773	-0.
981 1.319					
OnlineBackup	1.5065	0.724	2.080	0.038	0.
087 2.927					
DeviceProtection	-1.3794	0.620	-2.226	0.026	-2.
594 -0.164					
StreamingTV	0.9801	1.022	0.959	0.337	-1.
022 2.982					
StreamingMovies	1.2502	1.198	1.044	0.297	-1.
097 3.598					
MonthlyCharge	-0.0213	0.020	-1.048	0.295	-0.
061 0.019					
Item1	-0.3287	0.405	-0.812	0.417	-1.
122 0.465					
Item2	0.1685	0.378	0.446	0.655	-0.
572 0.909					
Item3	0.1229	0.345	0.356	0.722	-0.
553 0.799					
Item4	0.3002	0.309	0.972	0.331	-0.
305 0.905					
Item5	0.5704	0.321	1.778	0.075	-0.
058 1.199					
Item6	-0.1781	0.331	-0.539	0.590	-0.
826 0.470					
Item7	0.2578	0.312	0.826	0.409	-0.
354 0.869					
Item8	-0.4470	0.298	-1.501	0.133	-1.
031 0.137					
Area_Rural	1.0775	0.593	1.816	0.069	-0.
086 2.241					
Marital_Married	0.2919	0.780	0.374	0.708	-1.
236 1.820					
Marital_Separated	-0.0967	0.764	-0.127	0.899	-1.
595 1.401					
Marital_Divorced	-0.7178	0.756	-0.949	0.342	-2.
200 0.764					
Gender_Female	0.5890	0.567	1.039	0.299	-0.
523 1.701					
Gender_Nonbinary	-2.3206	1.885	-1.231	0.218	-6.
016 1.374					
Contract_Oneyear	0.2098	0.717	0.292	0.770	-1.
197 1.616					
Contract_TwoYear	1.3061	0.681	1.919	0.055	-0.
028 2.640					
InternetService_None	-1.1026	0.837	-1.318	0.188	-2.
743 0.537					
PaymentMethod_BankTransferautomatic	-1.2202	0.728	-1.676	0.094	-2.
648 0.207					
PaymentMethod_ElectronicCheck	-0.8519	0.640	-1.331	0.183	-2.
106 0.403					

=====

Omnibus:	37292.475	Durbin-Watson:	0.167
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1139.986
Skew:	0.076	Prob(JB):	2.85e-248
Kurtosis:	1.258	Cond. No.	2.57e+05

=====

## Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 2.57e+05. This might indicate that there are strong multicollinearity or other numerical problems.

```
In [24]: #Remove Marital Sep.
lm15 = smf.ols('Tenure ~ Lat+Lng+Population+Children+Age+Outage_sec_perweek+Email+Yearly_equip_failure+Techie+Port_modem+Phone+Multiple+OnlineSecurity+OnlineBackup', data=customer_churn).fit()
lm15.params
print(lm15.summary())
```

## OLS Regression Results

```
=====
```

Dep. Variable:	Tenure	R-squared:	0.006
Model:	OLS	Adj. R-squared:	0.002
Method:	Least Squares	F-statistic:	1.465
Date:	Sat, 12 Feb 2022	Prob (F-statistic):	0.0360
Time:	23:36:27	Log-Likelihood:	-41985.
No. Observations:	8950	AIC:	8.404e+04
Df Residuals:	8913	BIC:	8.431e+04
Df Model:	36		
Covariance Type:	nonrobust		

```
=====
```

		coef	std err	t	P> t	[0.
025	0.975]					
-----						
Intercept		41.9595	4.894	8.574	0.000	32.
366	51.553					
Lat		-0.1292	0.060	-2.147	0.032	-0.
247	-0.011					
Lng		0.0145	0.021	0.692	0.489	-0.
027	0.055					
Population		-4.187e-05	2.43e-05	-1.720	0.085	-8.96e
-05	5.85e-06					
Children		-0.1163	0.148	-0.786	0.432	-0.
406	0.174					
Age		0.0169	0.014	1.243	0.214	-0.
010	0.043					
Outage_sec_perweek		0.0465	0.096	0.486	0.627	-0.
141	0.234					
Email		-0.1584	0.093	-1.704	0.088	-0.
340	0.024					
Yearly_equip_failure		0.3875	0.480	0.807	0.420	-0.
554	1.329					
Techie		-0.9992	0.749	-1.334	0.182	-2.
467	0.469					
Port_modem		0.3436	0.560	0.614	0.539	-0.
753	1.440					
Phone		0.4036	0.961	0.420	0.674	-1.
480	2.287					
Multiple		0.2428	0.866	0.280	0.779	-1.
454	1.940					
OnlineSecurity		0.1694	0.587	0.289	0.773	-0.
981	1.320					
OnlineBackup		1.5072	0.724	2.081	0.037	0.

087	2.927					
DeviceProtection		-1.3805	0.620	-2.228	0.026	-2.
595	-0.166					
StreamingTV		0.9795	1.021	0.959	0.338	-1.
023	2.982					
StreamingMovies		1.2493	1.197	1.043	0.297	-1.
098	3.597					
MonthlyCharge		-0.0213	0.020	-1.048	0.295	-0.
061	0.019					
Item1		-0.3291	0.405	-0.813	0.416	-1.
123	0.464					
Item2		0.1687	0.378	0.447	0.655	-0.
571	0.909					
Item3		0.1224	0.345	0.355	0.722	-0.
553	0.798					
Item4		0.3002	0.309	0.972	0.331	-0.
305	0.905					
Item5		0.5709	0.321	1.780	0.075	-0.
058	1.200					
Item6		-0.1783	0.331	-0.539	0.590	-0.
827	0.470					
Item7		0.2579	0.312	0.827	0.408	-0.
354	0.869					
Item8		-0.4464	0.298	-1.499	0.134	-1.
030	0.137					
Area_Rural		1.0765	0.593	1.814	0.070	-0.
087	2.240					
Marital_Married		0.3244	0.736	0.441	0.660	-1.
119	1.768					
Marital_Divorced		-0.6853	0.711	-0.964	0.335	-2.
079	0.709					
Gender_Female		0.5894	0.567	1.039	0.299	-0.
522	1.701					
Gender_Nonbinary		-2.3186	1.885	-1.230	0.219	-6.
013	1.376					
Contract_Oneyear		0.2098	0.717	0.292	0.770	-1.
197	1.616					
Contract_TwoYear		1.3055	0.681	1.918	0.055	-0.
029	2.640					
InternetService_None		-1.1025	0.837	-1.318	0.188	-2.
742	0.537					
PaymentMethod_BankTransferautomatic		-1.2192	0.728	-1.675	0.094	-2.
646	0.208					
PaymentMethod_ElectronicCheck		-0.8519	0.640	-1.331	0.183	-2.
106	0.403					

```

=====
Omnibus:                37293.459    Durbin-Watson:           0.167
Prob(Omnibus):           0.000    Jarque-Bera (JB):       1139.953
Skew:                    0.076    Prob(JB):               2.90e-248
Kurtosis:                1.258    Cond. No.               2.57e+05
=====

```

## Notes:

- [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
- [2] The condition number is large, 2.57e+05. This might indicate that there are strong multicollinearity or other numerical problems.

```

In [25]: #Remove Multiple
lm16 = smf.ols('Tenure ~ Lat+Lng+Population+Children+Age+Outage_sec_perweek+Email+Year1
lm16.params
print(lm16.summary())

```

## OLS Regression Results

```

=====
Dep. Variable:          Tenure      R-squared:                0.006
Model:                  OLS         Adj. R-squared:           0.002
Method:                 Least Squares  F-statistic:             1.504
Date:                  Sat, 12 Feb 2022  Prob (F-statistic):       0.0284
Time:                  23:36:27      Log-Likelihood:          -41985.
No. Observations:      8950         AIC:                     8.404e+04
Df Residuals:          8914         BIC:                     8.430e+04
Df Model:               35
Covariance Type:       nonrobust
=====

```

```

=====
                                coef    std err          t      P>|t|      [0.
025      0.975]
-----
Intercept                41.5778      4.701        8.845      0.000      32.
364      50.792
Lat                    -0.1292      0.060       -2.149      0.032      -0.
247      -0.011
Lng                     0.0144      0.021        0.689      0.491      -0.
027      0.055
Population              -4.187e-05    2.43e-05     -1.720      0.085     -8.96e
-05      5.85e-06
Children                -0.1166      0.148       -0.788      0.431      -0.
407      0.174
Age                     0.0168      0.014        1.239      0.215      -0.
010      0.043
Outage_sec_perweek      0.0468      0.096        0.489      0.625      -0.
141      0.234
Email                  -0.1586      0.093       -1.707      0.088      -0.
341      0.024
Yearly_equip_failure    0.3879      0.480        0.808      0.419      -0.
553      1.329
Techie                 -1.0031      0.749       -1.339      0.180      -2.
471      0.465
Port_modem              0.3441      0.559        0.615      0.539      -0.
753      1.441
Phone                   0.4082      0.961        0.425      0.671      -1.
475      2.291
OnlineSecurity          0.1562      0.585        0.267      0.789      -0.
990      1.303
OnlineBackup            1.4094      0.635        2.220      0.026        0.
165      2.654
DeviceProtection       -1.4346      0.589       -2.436      0.015      -2.
589     -0.280
StreamingTV             0.7978      0.790        1.010      0.312      -0.
750      2.345
StreamingMovies         1.0241      0.888        1.153      0.249      -0.
717      2.765
MonthlyCharge          -0.0170      0.013       -1.288      0.198      -0.
043      0.009
Item1                  -0.3296      0.405       -0.814      0.416      -1.
123      0.464
Item2                   0.1663      0.377        0.440      0.660      -0.
574      0.906
Item3                   0.1232      0.345        0.358      0.721      -0.
552      0.799
Item4                   0.2991      0.309        0.969      0.333      -0.
306      0.904
Item5                   0.5709      0.321        1.780      0.075      -0.
058      1.200
Item6                  -0.1783      0.331       -0.539      0.590      -0.
827      0.470

```



2/12/22, 11:54 PM

		CustomerChurn_LinearRegression				
Item7		0.2587	0.312	0.829	0.407	-0.
353	0.870					
Item8		-0.4458	0.298	-1.497	0.134	-1.
030	0.138					
Area_Rural		1.0756	0.593	1.813	0.070	-0.
087	2.239					
Marital_Married		0.3234	0.736	0.439	0.661	-1.
120	1.767					
Marital_Divorced		-0.6845	0.711	-0.963	0.336	-2.
078	0.709					
Gender_Female		0.5866	0.567	1.035	0.301	-0.
525	1.698					
Gender_Nonbinary		-2.3176	1.885	-1.230	0.219	-6.
012	1.377					
Contract_Oneyear		0.2113	0.717	0.294	0.768	-1.
195	1.618					
Contract_TwoYear		1.3073	0.680	1.921	0.055	-0.
027	2.641					
InternetService_None		-1.0022	0.756	-1.325	0.185	-2.
484	0.480					
PaymentMethod_BankTransferautomatic		-1.2163	0.728	-1.671	0.095	-2.
643	0.211					
PaymentMethod_ElectronicCheck		-0.8511	0.640	-1.330	0.184	-2.
106	0.403					

=====

Omnibus:37295.560

Durbin-Watson:0.167

Prob(Omnibus):0.000

Jarque-Bera (JB):1139.870

Skew:0.076

Prob(JB):3.02e-248

Kurtosis:1.258

Cond. No.2.45e+05

=====

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 2.45e+05. This might indicate that there are strong multicollinearity or other numerical problems.

In [26]:

```
#Remove Online security
lm17 = smf.ols('Tenure ~ Lat+Lng+Population+Children+Age+Outage_sec_perweek+Email+Year1
lm17.params
print(lm17.summary())
```

OLS Regression Results						
=====						
Dep. Variable:		Tenure	R-squared:		0.006	
Model:		OLS	Adj. R-squared:		0.002	
Method:		Least Squares	F-statistic:		1.547	
Date:		Sat, 12 Feb 2022	Prob (F-statistic):		0.0222	
Time:		23:36:27	Log-Likelihood:		-41985.	
No. Observations:		8950	AIC:		8.404e+04	
Df Residuals:		8915	BIC:		8.429e+04	
Df Model:		34				
Covariance Type:		nonrobust				
=====						
=====						
		coef	std err	t	P> t	[0.
025	0.975]					
-----						
Intercept		41.6239	4.697	8.862	0.000	32.
416	50.831					
Lat		-0.1293	0.060	-2.150	0.032	-0.
247	-0.011					

Lng		0.0144	0.021	0.689	0.491	-0.
027	0.055					
Population		-4.183e-05	2.43e-05	-1.719	0.086	-8.96e
-05	5.88e-06					
Children		-0.1164	0.148	-0.787	0.431	-0.
407	0.174					
Age		0.0168	0.014	1.235	0.217	-0.
010	0.043					
Outage_sec_perweek		0.0466	0.096	0.487	0.626	-0.
141	0.234					
Email		-0.1591	0.093	-1.713	0.087	-0.
341	0.023					
Yearly equip_failure		0.3862	0.480	0.804	0.421	-0.
555	1.327					
Techie		-1.0047	0.749	-1.342	0.180	-2.
473	0.463					
Port_modem		0.3448	0.559	0.616	0.538	-0.
752	1.441					
Phone		0.4062	0.960	0.423	0.672	-1.
477	2.289					
OnlineBackup		1.4060	0.635	2.215	0.027	0.
162	2.650					
DeviceProtection		-1.4359	0.589	-2.438	0.015	-2.
590	-0.282					
StreamingTV		0.7889	0.789	1.000	0.317	-0.
757	2.335					
StreamingMovies		1.0137	0.887	1.142	0.253	-0.
726	2.753					
MonthlyCharge		-0.0167	0.013	-1.273	0.203	-0.
043	0.009					
Item1		-0.3297	0.405	-0.815	0.415	-1.
123	0.464					
Item2		0.1660	0.377	0.440	0.660	-0.
574	0.906					
Item3		0.1238	0.345	0.359	0.719	-0.
552	0.799					
Item4		0.2994	0.309	0.970	0.332	-0.
306	0.904					
Item5		0.5701	0.321	1.778	0.075	-0.
059	1.199					
Item6		-0.1780	0.331	-0.538	0.590	-0.
826	0.470					
Item7		0.2572	0.312	0.825	0.410	-0.
354	0.869					
Item8		-0.4452	0.298	-1.495	0.135	-1.
029	0.139					
Area_Rural		1.0759	0.593	1.814	0.070	-0.
087	2.239					
Marital_Married		0.3207	0.736	0.436	0.663	-1.
122	1.764					
Marital_Divorced		-0.6854	0.711	-0.964	0.335	-2.
079	0.708					
Gender_Female		0.5842	0.567	1.031	0.303	-0.
527	1.695					
Gender_Nonbinary		-2.3234	1.884	-1.233	0.218	-6.
017	1.371					
Contract_Oneyear		0.2113	0.717	0.295	0.768	-1.
195	1.617					
Contract_TwoYear		1.3095	0.680	1.925	0.054	-0.
024	2.643					
InternetService_None		-0.9991	0.756	-1.322	0.186	-2.
481	0.483					
PaymentMethod_BankTransferautomatic		-1.2193	0.728	-1.675	0.094	-2.
646	0.207					
PaymentMethod_ElectronicCheck		-0.8507	0.640	-1.329	0.184	-2.

105 0.404

```
=====
Omnibus:                 37295.322    Durbin-Watson:                 0.167
Prob(Omnibus):           0.000    Jarque-Bera (JB):             1139.896
Skew:                   0.076    Prob(JB):                     2.98e-248
Kurtosis:               1.258    Cond. No.                     2.45e+05
=====
```

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 2.45e+05. This might indicate that there are strong multicollinearity or other numerical problems.

```
In [27]: #Remove Contract One year
lm18 = smf.ols('Tenure ~ Lat+Lng+Population+Children+Age+Outage_sec_perweek+Email+Yearly equip_failure+Techie+Port_modem+Phone+OnlineBackup+DeviceProtection', data=customer_churn)
lm18.params
print(lm18.summary())
```

## OLS Regression Results

```
=====
Dep. Variable:            Tenure    R-squared:                0.006
Model:                    OLS       Adj. R-squared:           0.002
Method:                    Least Squares    F-statistic:             1.591
Date:                      Sat, 12 Feb 2022    Prob (F-statistic):       0.0171
Time:                      23:36:28    Log-Likelihood:          -41985.
No. Observations:          8950    AIC:                     8.404e+04
Df Residuals:              8916    BIC:                     8.428e+04
Df Model:                   33
Covariance Type:            nonrobust
=====
```

	coef	std err	t	P> t	[0.
Intercept	41.6794	4.693	8.881	0.000	32.
Lat	-0.1294	0.060	-2.152	0.031	-0.
Lng	0.0144	0.021	0.690	0.490	-0.
Population	-4.186e-05	2.43e-05	-1.720	0.086	-8.96e
Children	-0.1157	0.148	-0.782	0.434	-0.
Age	0.0167	0.014	1.233	0.218	-0.
Outage_sec_perweek	0.0465	0.096	0.487	0.627	-0.
Email	-0.1593	0.093	-1.716	0.086	-0.
Yearly equip_failure	0.3875	0.480	0.807	0.420	-0.
Techie	-1.0033	0.749	-1.340	0.180	-2.
Port_modem	0.3425	0.559	0.612	0.540	-0.
Phone	0.4107	0.960	0.428	0.669	-1.
OnlineBackup	1.4044	0.635	2.213	0.027	0.
DeviceProtection	-1.4356	0.589	-2.438	0.015	-2.

590	-0.281					
StreamingTV		0.7911	0.789	1.003	0.316	-0.
755	2.337					
StreamingMovies		1.0097	0.887	1.138	0.255	-0.
730	2.749					
MonthlyCharge		-0.0167	0.013	-1.269	0.204	-0.
042	0.009					
Item1		-0.3298	0.405	-0.815	0.415	-1.
123	0.464					
Item2		0.1645	0.377	0.436	0.663	-0.
575	0.904					
Item3		0.1256	0.345	0.364	0.715	-0.
550	0.801					
Item4		0.2990	0.309	0.969	0.333	-0.
306	0.904					
Item5		0.5705	0.321	1.779	0.075	-0.
058	1.199					
Item6		-0.1763	0.331	-0.533	0.594	-0.
824	0.472					
Item7		0.2566	0.312	0.823	0.411	-0.
355	0.868					
Item8		-0.4459	0.298	-1.497	0.134	-1.
030	0.138					
Area_Rural		1.0765	0.593	1.815	0.070	-0.
086	2.239					
Marital_Married		0.3195	0.736	0.434	0.664	-1.
123	1.762					
Marital_Divorced		-0.6883	0.711	-0.968	0.333	-2.
082	0.705					
Gender_Female		0.5852	0.567	1.032	0.302	-0.
526	1.696					
Gender_Nonbinary		-2.3222	1.884	-1.232	0.218	-6.
016	1.372					
Contract_TwoYear		1.2504	0.650	1.923	0.054	-0.
024	2.525					
InternetService_None		-0.9964	0.756	-1.318	0.188	-2.
478	0.485					
PaymentMethod_BankTransferautomatic		-1.2189	0.728	-1.675	0.094	-2.
645	0.208					
PaymentMethod_ElectronicCheck		-0.8535	0.640	-1.334	0.182	-2.
108	0.401					

```
=====
Omnibus:                37292.793    Durbin-Watson:           0.167
Prob(Omnibus):          0.000    Jarque-Bera (JB):       1139.979
Skew:                   0.076    Prob(JB):               2.86e-248
Kurtosis:               1.258    Cond. No.               2.45e+05
=====
```

## Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 2.45e+05. This might indicate that there are strong multicollinearity or other numerical problems.

```
In [28]: #Remove item 3
lm19 = smf.ols('Tenure ~ Lat+Lng+Population+Children+Age+Outage_sec_perweek+Email+Year1
lm19.params
print(lm19.summary())
```

## OLS Regression Results

```
=====
Dep. Variable:          Tenure    R-squared:                0.006
Model:                  OLS      Adj. R-squared:           0.002
=====
```

Method:	Least Squares	F-statistic:	1.637
Date:	Sat, 12 Feb 2022	Prob (F-statistic):	0.0132
Time:	23:36:28	Log-Likelihood:	-41985.
No. Observations:	8950	AIC:	8.404e+04
Df Residuals:	8917	BIC:	8.427e+04
Df Model:	32		
Covariance Type:	nonrobust		

		coef	std err	t	P> t	[0.
025		0.975]				
-----						
Intercept		41.8181	4.677	8.940	0.000	32.
649	50.987					
Lat		-0.1296	0.060	-2.155	0.031	-0.
247	-0.012					
Lng		0.0145	0.021	0.694	0.488	-0.
026	0.056					
Population		-4.193e-05	2.43e-05	-1.723	0.085	-8.96e
-05	5.78e-06					
Children		-0.1162	0.148	-0.785	0.432	-0.
406	0.174					
Age		0.0167	0.014	1.230	0.219	-0.
010	0.043					
Outage_sec_perweek		0.0461	0.096	0.482	0.630	-0.
141	0.234					
Email		-0.1591	0.093	-1.714	0.087	-0.
341	0.023					
Yearly_equip_failure		0.3851	0.480	0.802	0.422	-0.
556	1.326					
Techie		-1.0030	0.749	-1.340	0.180	-2.
471	0.465					
Port_modem		0.3410	0.559	0.610	0.542	-0.
755	1.437					
Phone		0.4103	0.960	0.427	0.669	-1.
472	2.293					
OnlineBackup		1.4012	0.635	2.208	0.027	0.
157	2.645					
DeviceProtection		-1.4394	0.589	-2.445	0.015	-2.
593	-0.285					
StreamingTV		0.7853	0.789	0.996	0.319	-0.
760	2.331					
StreamingMovies		1.0068	0.887	1.135	0.256	-0.
732	2.746					
MonthlyCharge		-0.0166	0.013	-1.266	0.206	-0.
042	0.009					
Item1		-0.2830	0.384	-0.737	0.461	-1.
035	0.469					
Item2		0.1922	0.370	0.520	0.603	-0.
533	0.917					
Item4		0.2961	0.308	0.960	0.337	-0.
309	0.901					
Item5		0.5734	0.321	1.789	0.074	-0.
055	1.202					
Item6		-0.1676	0.330	-0.508	0.611	-0.
814	0.479					
Item7		0.2626	0.311	0.843	0.399	-0.
348	0.873					
Item8		-0.4433	0.298	-1.489	0.136	-1.
027	0.140					
Area_Rural		1.0827	0.593	1.826	0.068	-0.
080	2.245					
Marital_Married		0.3163	0.736	0.430	0.667	-1.
126	1.759					

2/12/22, 11:54 PM

		CustomerChurn_LinearRegression				
Marital_Divorced		-0.6937	0.711	-0.976	0.329	-2.
087	0.699					
Gender_Female		0.5815	0.567	1.026	0.305	-0.
529	1.692					
Gender_Nonbinary		-2.3260	1.884	-1.234	0.217	-6.
020	1.368					
Contract_TwoYear		1.2524	0.650	1.927	0.054	-0.
022	2.527					
InternetService_None		-0.9872	0.756	-1.307	0.191	-2.
468	0.494					
PaymentMethod_BankTransferautomatic		-1.2246	0.728	-1.683	0.092	-2.
651	0.202					
PaymentMethod_ElectronicCheck		-0.8552	0.640	-1.337	0.181	-2.
109	0.399					
=====						
Omnibus:	37289.576	Durbin-Watson:	0.167			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1140.101			
Skew:	0.076	Prob(JB):	2.69e-248			
Kurtosis:	1.258	Cond. No.	2.44e+05			
=====						

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 2.44e+05. This might indicate that there are strong multicollinearity or other numerical problems.

In [29]:

#Remove phone  
lm21 = smf.ols('Tenure ~ Lat+Lng+Population+Children+Age+Outage\_sec\_perweek+Email+Year1  
lm21.params  
print(lm21.summary())

OLS Regression Results						
=====						
Dep. Variable:	Tenure	R-squared:	0.006			
Model:	OLS	Adj. R-squared:	0.002			
Method:	Least Squares	F-statistic:	1.684			
Date:	Sat, 12 Feb 2022	Prob (F-statistic):	0.0102			
Time:	23:36:28	Log-Likelihood:	-41985.			
No. Observations:	8950	AIC:	8.403e+04			
Df Residuals:	8918	BIC:	8.426e+04			
Df Model:	31					
Covariance Type:	nonrobust					
=====						
=====						
		coef	std err	t	P> t	[0.
025	0.975]					
-----						
Intercept		42.1793	4.600	9.169	0.000	33.
162	51.197					
Lat		-0.1296	0.060	-2.156	0.031	-0.
248	-0.012					
Lng		0.0145	0.021	0.695	0.487	-0.
026	0.056					
Population		-4.185e-05	2.43e-05	-1.720	0.086	-8.96e
-05	5.86e-06					
Children		-0.1160	0.148	-0.784	0.433	-0.
406	0.174					
Age		0.0168	0.014	1.236	0.217	-0.
010	0.043					
Outage_sec_perweek		0.0461	0.096	0.482	0.630	-0.
141	0.234					

Email		-0.1595	0.093	-1.717	0.086	-0.
341	0.023					
Yearly equip failure		0.3823	0.480	0.797	0.426	-0.
558	1.323					
Techie		-1.0026	0.749	-1.339	0.181	-2.
470	0.465					
Port_modem		0.3421	0.559	0.612	0.541	-0.
754	1.438					
OnlineBackup		1.4004	0.635	2.207	0.027	0.
157	2.644					
DeviceProtection		-1.4459	0.588	-2.457	0.014	-2.
599	-0.292					
StreamingTV		0.7835	0.789	0.994	0.320	-0.
762	2.329					
StreamingMovies		1.0068	0.887	1.135	0.256	-0.
732	2.746					
MonthlyCharge		-0.0167	0.013	-1.267	0.205	-0.
042	0.009					
Item1		-0.2868	0.384	-0.747	0.455	-1.
039	0.465					
Item2		0.1992	0.369	0.539	0.590	-0.
525	0.923					
Item4		0.2960	0.308	0.959	0.337	-0.
309	0.901					
Item5		0.5747	0.320	1.793	0.073	-0.
054	1.203					
Item6		-0.1688	0.330	-0.512	0.609	-0.
815	0.478					
Item7		0.2635	0.311	0.846	0.397	-0.
347	0.874					
Item8		-0.4425	0.298	-1.487	0.137	-1.
026	0.141					
Area_Rural		1.0854	0.593	1.831	0.067	-0.
077	2.248					
Marital_Married		0.3219	0.736	0.437	0.662	-1.
121	1.764					
Marital_Divorced		-0.6918	0.711	-0.973	0.330	-2.
085	0.701					
Gender_Female		0.5803	0.567	1.024	0.306	-0.
530	1.691					
Gender_Nonbinary		-2.3233	1.884	-1.233	0.218	-6.
017	1.370					
Contract_TwoYear		1.2529	0.650	1.927	0.054	-0.
021	2.527					
InternetService_None		-0.9831	0.755	-1.301	0.193	-2.
464	0.498					
PaymentMethod_BankTransferautomatic		-1.2240	0.728	-1.682	0.093	-2.
650	0.202					
PaymentMethod_ElectronicCheck		-0.8522	0.640	-1.332	0.183	-2.
106	0.402					

```

=====
Omnibus:                37289.755    Durbin-Watson:           0.167
Prob(Omnibus):           0.000    Jarque-Bera (JB):       1140.109
Skew:                    0.076    Prob(JB):               2.68e-248
Kurtosis:                1.258    Cond. No.               2.40e+05
=====

```

## Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 2.4e+05. This might indicate that there are strong multicollinearity or other numerical problems.

```
In [30]: #Remove Marital Married
lm22 = smf.ols('Tenure ~ Lat+Lng+Population+Children+Age+Outage_sec_perweek+Email+Yearly_equip_failure+MonthlyCharge+Item1+Item2+Item4+Item5+Item6+Item7', data=smf.data).fit()
lm22.params
print(lm22.summary())
```

### OLS Regression Results

=====						
Dep. Variable:		Tenure	R-squared:		0.006	
Model:		OLS	Adj. R-squared:		0.002	
Method:		Least Squares	F-statistic:		1.734	
Date:		Sat, 12 Feb 2022	Prob (F-statistic):		0.00773	
Time:		23:36:28	Log-Likelihood:		-41985.	
No. Observations:		8950	AIC:		8.403e+04	
Df Residuals:		8919	BIC:		8.425e+04	
Df Model:		30				
Covariance Type:		nonrobust				
=====						
=====						
			coef	std err	t	P> t
025	0.975]					[0.
-----						
Intercept			42.2144	4.599	9.178	0.000
199	51.230					33.
Lat			-0.1294	0.060	-2.153	0.031
247	-0.012					-0.
Lng			0.0143	0.021	0.686	0.493
027	0.055					-0.
Population			-4.185e-05	2.43e-05	-1.720	0.086
-05	5.86e-06					-8.96e
Children			-0.1161	0.148	-0.785	0.433
406	0.174					-0.
Age			0.0168	0.014	1.237	0.216
010	0.043					-0.
Outage_sec_perweek			0.0470	0.096	0.492	0.623
140	0.234					-0.
Email			-0.1592	0.093	-1.714	0.087
341	0.023					-0.
Yearly_equip_failure			0.3809	0.480	0.794	0.427
560	1.321					-0.
Techie			-1.0003	0.749	-1.336	0.182
468	0.467					-2.
Port_modem			0.3429	0.559	0.613	0.540
753	1.439					-0.
OnlineBackup			1.4019	0.634	2.210	0.027
158	2.646					0.
DeviceProtection			-1.4468	0.588	-2.459	0.014
600	-0.293					-2.
StreamingTV			0.7822	0.788	0.992	0.321
763	2.328					-0.
StreamingMovies			1.0077	0.887	1.136	0.256
731	2.747					-0.
MonthlyCharge			-0.0167	0.013	-1.270	0.204
042	0.009					-0.
Item1			-0.2858	0.384	-0.745	0.456
038	0.466					-1.
Item2			0.1981	0.369	0.536	0.592
526	0.922					-0.
Item4			0.2971	0.308	0.963	0.335
307	0.902					-0.
Item5			0.5772	0.320	1.801	0.072
051	1.205					-0.
Item6			-0.1703	0.330	-0.517	0.605
817	0.476					-0.
Item7			0.2632	0.311	0.845	0.398
						-0.



347	0.874					
Item8		-0.4412	0.298	-1.482	0.138	-1.0
025	0.142					
Area_Rural		1.0857	0.593	1.831	0.067	-0.0
076	2.248					
Marital_Divorced		-0.7690	0.688	-1.117	0.264	-2.0
118	0.580					
Gender_Female		0.5783	0.567	1.021	0.307	-0.0
532	1.689					
Gender_Nonbinary		-2.3252	1.884	-1.234	0.217	-6.0
018	1.368					
Contract_TwoYear		1.2499	0.650	1.923	0.055	-0.0
024	2.524					
InternetService_None		-0.9837	0.755	-1.302	0.193	-2.0
464	0.497					
PaymentMethod_BankTransferautomatic		-1.2221	0.727	-1.680	0.093	-2.0
648	0.204					
PaymentMethod_ElectronicCheck		-0.8494	0.640	-1.328	0.184	-2.0
103	0.404					
=====						
Omnibus:	37284.620	Durbin-Watson:	0.167			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1140.303			
Skew:	0.076	Prob(JB):	2.43e-248			
Kurtosis:	1.258	Cond. No.	2.40e+05			
=====						

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 2.4e+05. This might indicate that there are strong multicollinearity or other numerical problems.

In [31]:

#Remove Outages Sec perweek  
lm23 = smf.ols('Tenure ~ Lat+Lng+Population+Children+Age+Email+Yearly equip\_failure+Tec  
lm23.params  
print(lm23.summary())

OLS Regression Results						
=====						
Dep. Variable:	Tenure	R-squared:	0.006			
Model:	OLS	Adj. R-squared:	0.003			
Method:	Least Squares	F-statistic:	1.785			
Date:	Sat, 12 Feb 2022	Prob (F-statistic):	0.00588			
Time:	23:36:28	Log-Likelihood:	-41985.			
No. Observations:	8950	AIC:	8.403e+04			
Df Residuals:	8920	BIC:	8.424e+04			
Df Model:	29					
Covariance Type:	nonrobust					
=====						
		coef	std err	t	P> t	[0.
025	0.975]					
-----						
Intercept		42.7129	4.486	9.522	0.000	33.
920	51.506					
Lat		-0.1294	0.060	-2.152	0.031	-0.
247	-0.012					
Lng		0.0145	0.021	0.696	0.487	-0.
026	0.056					
Population		-4.168e-05	2.43e-05	-1.713	0.087	-8.94e
-05	6.02e-06					
Children		-0.1150	0.148	-0.778	0.437	-0.

405	0.175					
Age		0.0167	0.014	1.230	0.219	-0.
010	0.043					
Email		-0.1592	0.093	-1.715	0.086	-0.
341	0.023					
Yearly equip failure		0.3827	0.480	0.798	0.425	-0.
558	1.323					
Techie		-0.9999	0.749	-1.336	0.182	-2.
467	0.467					
Port_modem		0.3451	0.559	0.617	0.537	-0.
751	1.441					
OnlineBackup		1.4014	0.634	2.209	0.027	0.
158	2.645					
DeviceProtection		-1.4416	0.588	-2.450	0.014	-2.
595	-0.288					
StreamingTV		0.7859	0.788	0.997	0.319	-0.
759	2.331					
StreamingMovies		1.0108	0.887	1.140	0.254	-0.
728	2.750					
MonthlyCharge		-0.0167	0.013	-1.269	0.204	-0.
042	0.009					
Item1		-0.2901	0.384	-0.756	0.449	-1.
042	0.462					
Item2		0.1994	0.369	0.540	0.589	-0.
524	0.923					
Item4		0.2964	0.308	0.961	0.337	-0.
308	0.901					
Item5		0.5754	0.320	1.796	0.073	-0.
053	1.203					
Item6		-0.1705	0.330	-0.517	0.605	-0.
817	0.476					
Item7		0.2609	0.311	0.838	0.402	-0.
349	0.871					
Item8		-0.4383	0.298	-1.473	0.141	-1.
022	0.145					
Area_Rural		1.0819	0.593	1.825	0.068	-0.
080	2.244					
Marital_Divorced		-0.7655	0.688	-1.112	0.266	-2.
115	0.584					
Gender_Female		0.5743	0.567	1.014	0.311	-0.
536	1.685					
Gender_Nonbinary		-2.3223	1.884	-1.233	0.218	-6.
015	1.371					
Contract_TwoYear		1.2539	0.650	1.929	0.054	-0.
020	2.528					
InternetService_None		-0.9810	0.755	-1.299	0.194	-2.
462	0.500					
PaymentMethod_BankTransferautomatic		-1.2254	0.727	-1.684	0.092	-2.
651	0.201					
PaymentMethod_ElectronicCheck		-0.8510	0.640	-1.330	0.183	-2.
105	0.403					

```

=====
Omnibus:                37284.855   Durbin-Watson:           0.167
Prob(Omnibus):           0.000   Jarque-Bera (JB):       1140.310
Skew:                    0.076   Prob(JB):               2.43e-248
Kurtosis:                1.258   Cond. No.               2.34e+05
=====

```

## Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 2.34e+05. This might indicate that there are strong multicollinearity or other numerical problems.

```
In [32]: # Remove item 6
lm24 = smf.ols('Tenure ~ Lat+Lng+Population+Children+Age+Email+Yearly equip_failure+Tec
lm24.params
print(lm24.summary())
```

### OLS Regression Results

=====						
Dep. Variable:		Tenure	R-squared:		0.006	
Model:		OLS	Adj. R-squared:		0.003	
Method:		Least Squares	F-statistic:		1.840	
Date:		Sat, 12 Feb 2022	Prob (F-statistic):		0.00445	
Time:		23:36:28	Log-Likelihood:		-41985.	
No. Observations:		8950	AIC:		8.403e+04	
Df Residuals:		8921	BIC:		8.423e+04	
Df Model:		28				
Covariance Type:		nonrobust				
=====						
			coef	std err	t	P> t
025	0.975]					[0.
-----						
Intercept			42.4272	4.452	9.531	0.000
701	51.153					33.
Lat			-0.1289	0.060	-2.145	0.032
247	-0.011					-0.
Lng			0.0147	0.021	0.704	0.482
026	0.056					-0.
Population			-4.172e-05	2.43e-05	-1.715	0.086
-05	5.98e-06					-8.94e
Children			-0.1147	0.148	-0.776	0.438
405	0.175					-0.
Age			0.0166	0.014	1.227	0.220
010	0.043					-0.
Email			-0.1596	0.093	-1.719	0.086
342	0.022					-0.
Yearly equip_failure			0.3840	0.480	0.800	0.424
556	1.324					-0.
Techie			-1.0008	0.749	-1.337	0.181
468	0.467					-2.
Port_modem			0.3462	0.559	0.619	0.536
750	1.442					-0.
OnlineBackup			1.3989	0.634	2.205	0.027
155	2.642					0.
DeviceProtection			-1.4451	0.588	-2.456	0.014
598	-0.292					-2.
StreamingTV			0.7825	0.788	0.993	0.321
763	2.328					-0.
StreamingMovies			1.0043	0.887	1.132	0.258
734	2.743					-0.
MonthlyCharge			-0.0166	0.013	-1.261	0.207
042	0.009					-0.
Item1			-0.3281	0.376	-0.872	0.383
066	0.410					-1.
Item2			0.1770	0.367	0.483	0.629
542	0.896					-0.
Item4			0.2782	0.306	0.908	0.364
322	0.879					-0.
Item5			0.6090	0.314	1.941	0.052
006	1.224					-0.
Item7			0.2330	0.307	0.760	0.447
368	0.834					-0.
Item8			-0.4571	0.295	-1.548	0.122
036	0.122					-1.
Area_Rural			1.0831	0.593	1.827	0.068
						-0.

079	2.245					
Marital_Divorced		-0.7665	0.688	-1.114	0.265	-2.1
116	0.583					
Gender_Female		0.5768	0.566	1.018	0.309	-0.1
534	1.687					
Gender_Nonbinary		-2.3394	1.884	-1.242	0.214	-6.1
032	1.353					
Contract_TwoYear		1.2541	0.650	1.930	0.054	-0.1
020	2.528					
InternetService_None		-0.9747	0.755	-1.291	0.197	-2.1
455	0.506					
PaymentMethod_BankTransferautomatic		-1.2244	0.727	-1.683	0.092	-2.1
650	0.202					
PaymentMethod_ElectronicCheck		-0.8491	0.640	-1.328	0.184	-2.1
103	0.405					
=====						
Omnibus:	37279.250	Durbin-Watson:	0.167			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1140.519			
Skew:	0.077	Prob(JB):	2.19e-248			
Kurtosis:	1.258	Cond. No.	2.32e+05			
=====						

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 2.32e+05. This might indicate that there are strong multicollinearity or other numerical problems.

In [33]:

# Remove item 2  
lm25 = smf.ols('Tenure ~ Lat+Lng+Population+Children+Age+Email+Yearly equip\_failure+Tec  
lm25.params  
print(lm25.summary())

OLS Regression Results						
=====						
Dep. Variable:	Tenure	R-squared:	0.006			
Model:	OLS	Adj. R-squared:	0.003			
Method:	Least Squares	F-statistic:	1.899			
Date:	Sat, 12 Feb 2022	Prob (F-statistic):	0.00329			
Time:	23:36:29	Log-Likelihood:	-41985.			
No. Observations:	8950	AIC:	8.403e+04			
Df Residuals:	8922	BIC:	8.423e+04			
Df Model:	27					
Covariance Type:	nonrobust					
=====						
=====						
		coef	std err	t	P> t	[0.
025	0.975]					
-----						
Intercept		42.5413	4.445	9.570	0.000	33.
828	51.255					
Lat		-0.1287	0.060	-2.141	0.032	-0.
247	-0.011					
Lng		0.0145	0.021	0.695	0.487	-0.
026	0.055					
Population		-4.17e-05	2.43e-05	-1.714	0.087	-8.94e
-05	6e-06					
Children		-0.1138	0.148	-0.770	0.441	-0.
404	0.176					
Age		0.0167	0.014	1.233	0.218	-0.
010	0.043					
Email		-0.1597	0.093	-1.720	0.085	-0.

342	0.022					
Yearly_equip_failure		0.3803	0.480	0.793	0.428	-0.
560	1.321					
Techie		-1.0000	0.749	-1.336	0.182	-2.
467	0.467					
Port_modem		0.3482	0.559	0.623	0.533	-0.
748	1.444					
OnlineBackup		1.4058	0.634	2.217	0.027	0.
163	2.649					
DeviceProtection		-1.4406	0.588	-2.449	0.014	-2.
594	-0.288					
StreamingTV		0.7863	0.788	0.997	0.319	-0.
759	2.331					
StreamingMovies		1.0109	0.887	1.140	0.254	-0.
727	2.749					
MonthlyCharge		-0.0167	0.013	-1.271	0.204	-0.
042	0.009					
Item1		-0.2195	0.302	-0.727	0.467	-0.
811	0.372					
Item4		0.2763	0.306	0.902	0.367	-0.
324	0.877					
Item5		0.6150	0.313	1.962	0.050	0.
001	1.229					
Item7		0.2491	0.305	0.817	0.414	-0.
348	0.847					
Item8		-0.4466	0.295	-1.517	0.129	-1.
024	0.131					
Area_Rural		1.0813	0.593	1.824	0.068	-0.
081	2.243					
Marital_Divorced		-0.7647	0.688	-1.111	0.267	-2.
114	0.584					
Gender_Female		0.5779	0.566	1.020	0.308	-0.
532	1.688					
Gender_Nonbinary		-2.3442	1.884	-1.245	0.213	-6.
036	1.348					
Contract_TwoYear		1.2558	0.650	1.933	0.053	-0.
018	2.530					
InternetService_None		-0.9801	0.755	-1.298	0.194	-2.
460	0.500					
PaymentMethod_BankTransferautomatic		-1.2237	0.727	-1.682	0.093	-2.
650	0.202					
PaymentMethod_ElectronicCheck		-0.8433	0.639	-1.319	0.187	-2.
097	0.410					

```
=====
Omnibus:          37275.854    Durbin-Watson:          0.167
Prob(Omnibus):    0.000    Jarque-Bera (JB):      1140.655
Skew:             0.077    Prob(JB):              2.04e-248
Kurtosis:         1.258    Cond. No.               2.32e+05
=====
```

## Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 2.32e+05. This might indicate that there are strong multicollinearity or other numerical problems.

```
In [34]: #Remove port modem
lm26 = smf.ols('Tenure ~ Lat+Lng+Population+Children+Age+Email+Yearly_equip_failure+Tec
lm26.params
print(lm26.summary())
```

## OLS Regression Results

```
=====
```

Dep. Variable:	Tenure	R-squared:	0.006
Model:	OLS	Adj. R-squared:	0.003
Method:	Least Squares	F-statistic:	1.958
Date:	Sat, 12 Feb 2022	Prob (F-statistic):	0.00251
Time:	23:36:29	Log-Likelihood:	-41985.
No. Observations:	8950	AIC:	8.402e+04
Df Residuals:	8923	BIC:	8.422e+04
Df Model:	26		
Covariance Type:	nonrobust		

=====						
=====		coef	std err	t	P> t	[0.
025	0.975]					
-----						
Intercept		42.6975	4.438	9.621	0.000	33.
998	51.397					
Lat		-0.1292	0.060	-2.150	0.032	-0.
247	-0.011					
Lng		0.0142	0.021	0.682	0.496	-0.
027	0.055					
Population		-4.166e-05	2.43e-05	-1.712	0.087	-8.94e
-05	6.03e-06					
Children		-0.1131	0.148	-0.765	0.444	-0.
403	0.177					
Age		0.0167	0.014	1.233	0.218	-0.
010	0.043					
Email		-0.1591	0.093	-1.714	0.087	-0.
341	0.023					
Yearly_equip_failure		0.3833	0.480	0.799	0.424	-0.
557	1.323					
Techie		-1.0048	0.748	-1.342	0.179	-2.
472	0.462					
OnlineBackup		1.4065	0.634	2.218	0.027	0.
163	2.650					
DeviceProtection		-1.4413	0.588	-2.451	0.014	-2.
594	-0.288					
StreamingTV		0.7868	0.788	0.998	0.318	-0.
758	2.332					
StreamingMovies		1.0157	0.887	1.145	0.252	-0.
722	2.754					
MonthlyCharge		-0.0167	0.013	-1.273	0.203	-0.
042	0.009					
Item1		-0.2197	0.302	-0.728	0.467	-0.
811	0.372					
Item4		0.2776	0.306	0.906	0.365	-0.
323	0.878					
Item5		0.6146	0.313	1.961	0.050	0.
000	1.229					
Item7		0.2492	0.305	0.818	0.414	-0.
348	0.847					
Item8		-0.4476	0.295	-1.520	0.129	-1.
025	0.130					
Area_Rural		1.0859	0.593	1.832	0.067	-0.
076	2.248					
Marital_Divorced		-0.7608	0.688	-1.106	0.269	-2.
110	0.588					
Gender_Female		0.5712	0.566	1.009	0.313	-0.
539	1.681					
Gender_Nonbinary		-2.3488	1.883	-1.247	0.212	-6.
041	1.343					
Contract_TwoYear		1.2579	0.650	1.936	0.053	-0.
016	2.532					
InternetService_None		-0.9803	0.755	-1.298	0.194	-2.
460	0.500					

2/12/22, 11:54 PM

CustomerChurn\_LinearRegression

PaymentMethod_BankTransferautomatic	-1.2247	0.727	-1.684	0.092	-2.
650 0.201					
PaymentMethod_ElectronicCheck	-0.8458	0.639	-1.323	0.186	-2.
099 0.408					
=====					
Omnibus:	37260.927	Durbin-Watson:		0.167	
Prob(Omnibus):	0.000	Jarque-Bera (JB):		1141.173	
Skew:	0.077	Prob(JB):		1.58e-248	
Kurtosis:	1.257	Cond. No.		2.32e+05	
=====					

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 2.32e+05. This might indicate that there are strong multicollinearity or other numerical problems.

In [35]:

#Remove lng  
lm27 = smf.ols('Tenure ~ Lat+Population+Children+Age+Email+Yearly equip\_failure+Techie+  
lm27.params  
print(lm27.summary())

OLS Regression Results						
=====						
Dep. Variable:	Tenure	R-squared:		0.006		
Model:	OLS	Adj. R-squared:		0.003		
Method:	Least Squares	F-statistic:		2.017		
Date:	Sat, 12 Feb 2022	Prob (F-statistic):		0.00192		
Time:	23:36:29	Log-Likelihood:		-41986.		
No. Observations:	8950	AIC:		8.402e+04		
Df Residuals:	8924	BIC:		8.421e+04		
Df Model:	25					
Covariance Type:	nonrobust					
=====						
		coef	std err	t	P> t	[0.
025	0.975]					
-----						
Intercept		41.3902	4.002	10.343	0.000	33.
546	49.235					
Lat		-0.1293	0.060	-2.151	0.031	-0.
247	-0.011					
Population		-4.238e-05	2.43e-05	-1.743	0.081	-9e
-05	5.27e-06					
Children		-0.1116	0.148	-0.755	0.450	-0.
401	0.178					
Age		0.0168	0.014	1.239	0.215	-0.
010	0.043					
Email		-0.1587	0.093	-1.710	0.087	-0.
341	0.023					
Yearly equip_failure		0.3765	0.480	0.785	0.432	-0.
563	1.316					
Techie		-1.0026	0.748	-1.340	0.180	-2.
470	0.464					
OnlineBackup		1.4127	0.634	2.228	0.026	0.
170	2.656					
DeviceProtection		-1.4335	0.588	-2.438	0.015	-2.
586	-0.281					
StreamingTV		0.7814	0.788	0.991	0.322	-0.
764	2.326					
StreamingMovies		1.0197	0.887	1.150	0.250	-0.
718	2.758					

2/12/22, 11:54 PM

		CustomerChurn_LinearRegression				
MonthlyCharge		-0.0167	0.013	-1.272	0.204	-0.
042	0.009					
Item1		-0.2141	0.302	-0.710	0.478	-0.
805	0.377					
Item4		0.2780	0.306	0.907	0.364	-0.
323	0.879					
Item5		0.6142	0.313	1.960	0.050	-0.
000	1.229					
Item7		0.2483	0.305	0.815	0.415	-0.
349	0.846					
Item8		-0.4487	0.294	-1.524	0.128	-1.
026	0.129					
Area_Rural		1.0870	0.593	1.834	0.067	-0.
075	2.249					
Marital_Divorced		-0.7606	0.688	-1.105	0.269	-2.
109	0.588					
Gender_Female		0.5656	0.566	0.999	0.318	-0.
544	1.676					
Gender_Nonbinary		-2.3564	1.883	-1.251	0.211	-6.
048	1.335					
Contract_TwoYear		1.2622	0.650	1.943	0.052	-0.
011	2.536					
InternetService_None		-0.9852	0.755	-1.305	0.192	-2.
465	0.495					
PaymentMethod_BankTransferautomatic		-1.2109	0.727	-1.665	0.096	-2.
636	0.214					
PaymentMethod_ElectronicCheck		-0.8339	0.639	-1.305	0.192	-2.
087	0.419					
=====						
Omnibus:	37257.476	Durbin-Watson:	0.167			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1141.325			
Skew:	0.077	Prob(JB):	1.46e-248			
Kurtosis:	1.257	Cond. No.	2.09e+05			
=====						

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 2.09e+05. This might indicate that there are strong multicollinearity or other numerical problems.

In [36]:

```
#Remove Item 1
lm28 = smf.ols('Tenure ~ Lat+Population+Children+Age+Email+Yearly_equip_failure+Techie+
lm28.params
print(lm28.summary())
```

OLS Regression Results					
=====					
Dep. Variable:	Tenure	R-squared:	0.006		
Model:	OLS	Adj. R-squared:	0.003		
Method:	Least Squares	F-statistic:	2.081		
Date:	Sat, 12 Feb 2022	Prob (F-statistic):	0.00148		
Time:	23:36:29	Log-Likelihood:	-41986.		
No. Observations:	8950	AIC:	8.402e+04		
Df Residuals:	8925	BIC:	8.420e+04		
Df Model:	24				
Covariance Type:	nonrobust				
=====					
=====					
		coef	std err	t	P> t
025	0.975]				[0.
-----					
-----					



Intercept	41.1125	3.982	10.323	0.000	33.
306 48.919					
Lat	-0.1297	0.060	-2.158	0.031	-0.
247 -0.012					
Population	-4.253e-05	2.43e-05	-1.750	0.080	-9.02e
-05 5.11e-06					
Children	-0.1123	0.148	-0.760	0.447	-0.
402 0.177					
Age	0.0169	0.014	1.246	0.213	-0.
010 0.043					
Email	-0.1584	0.093	-1.706	0.088	-0.
340 0.024					
Yearly equip_failure	0.3714	0.479	0.775	0.439	-0.
568 1.311					
Techie	-1.0096	0.748	-1.349	0.177	-2.
477 0.457					
OnlineBackup	1.4224	0.634	2.244	0.025	0.
180 2.665					
DeviceProtection	-1.4344	0.588	-2.439	0.015	-2.
587 -0.282					
StreamingTV	0.7904	0.788	1.003	0.316	-0.
754 2.335					
StreamingMovies	1.0325	0.886	1.165	0.244	-0.
705 2.770					
MonthlyCharge	-0.0169	0.013	-1.290	0.197	-0.
043 0.009					
Item4	0.2912	0.306	0.952	0.341	-0.
308 0.891					
Item5	0.5947	0.312	1.905	0.057	-0.
017 1.207					
Item7	0.1824	0.290	0.628	0.530	-0.
387 0.751					
Item8	-0.4974	0.286	-1.737	0.082	-1.
059 0.064					
Area_Rural	1.0899	0.593	1.839	0.066	-0.
072 2.252					
Marital_Divorced	-0.7595	0.688	-1.104	0.270	-2.
108 0.589					
Gender_Female	0.5575	0.566	0.985	0.325	-0.
552 1.667					
Gender_Nonbinary	-2.3772	1.883	-1.262	0.207	-6.
068 1.314					
Contract_TwoYear	1.2643	0.650	1.946	0.052	-0.
009 2.538					
InternetService_None	-0.9822	0.755	-1.301	0.193	-2.
462 0.498					
PaymentMethod_BankTransferautomatic	-1.2165	0.727	-1.673	0.094	-2.
642 0.209					
PaymentMethod_ElectronicCheck	-0.8432	0.639	-1.320	0.187	-2.
096 0.409					

```

=====
Omnibus:                37248.366   Durbin-Watson:           0.167
Prob(Omnibus):           0.000   Jarque-Bera (JB):       1141.668
Skew:                    0.077   Prob(JB):               1.23e-248
Kurtosis:                1.257   Cond. No.               2.08e+05
=====

```

## Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 2.08e+05. This might indicate that there are strong multicollinearity or other numerical problems.

```
In [37]: #Remove Item 7
lm29 = smf.ols('Tenure ~ Lat+Population+Children+Age+Email+Yearly equip_failure+Techie+
lm29.params
print(lm29.summary())
```

### OLS Regression Results

=====						
Dep. Variable:		Tenure	R-squared:		0.006	
Model:		OLS	Adj. R-squared:		0.003	
Method:		Least Squares	F-statistic:		2.154	
Date:		Sat, 12 Feb 2022	Prob (F-statistic):		0.00108	
Time:		23:36:29	Log-Likelihood:		-41986.	
No. Observations:		8950	AIC:		8.402e+04	
Df Residuals:		8926	BIC:		8.419e+04	
Df Model:		23				
Covariance Type:		nonrobust				
=====						
=====						
		coef	std err	t	P> t	[0.
025	0.975]					
-----						
Intercept		41.6538	3.888	10.713	0.000	34.
032	49.275					
Lat		-0.1290	0.060	-2.147	0.032	-0.
247	-0.011					
Population		-4.267e-05	2.43e-05	-1.756	0.079	-9.03e
-05	4.97e-06					
Children		-0.1126	0.148	-0.761	0.446	-0.
402	0.177					
Age		0.0170	0.014	1.252	0.211	-0.
010	0.044					
Email		-0.1577	0.093	-1.699	0.089	-0.
340	0.024					
Yearly equip_failure		0.3644	0.479	0.760	0.447	-0.
575	1.304					
Techie		-1.0051	0.748	-1.343	0.179	-2.
472	0.462					
OnlineBackup		1.4206	0.634	2.241	0.025	0.
178	2.663					
DeviceProtection		-1.4434	0.588	-2.456	0.014	-2.
596	-0.291					
StreamingTV		0.7801	0.788	0.990	0.322	-0.
764	2.324					
StreamingMovies		1.0238	0.886	1.155	0.248	-0.
714	2.761					
MonthlyCharge		-0.0168	0.013	-1.282	0.200	-0.
043	0.009					
Item4		0.3054	0.305	1.002	0.317	-0.
292	0.903					
Item5		0.5602	0.307	1.823	0.068	-0.
042	1.163					
Item8		-0.4616	0.281	-1.645	0.100	-1.
012	0.088					
Area_Rural		1.0917	0.593	1.842	0.065	-0.
070	2.253					
Marital_Divorced		-0.7554	0.688	-1.098	0.272	-2.
104	0.593					
Gender_Female		0.5587	0.566	0.987	0.324	-0.
551	1.668					
Gender_Nonbinary		-2.3713	1.883	-1.259	0.208	-6.
062	1.320					
Contract_TwoYear		1.2726	0.650	1.959	0.050	-0.
001	2.546					
InternetService None		-0.9811	0.755	-1.300	0.194	-2.

```

461      0.499
PaymentMethod_BankTransferautomatic -1.2191      0.727      -1.677      0.094      -2.
644      0.206
PaymentMethod_ElectronicCheck -0.8406      0.639      -1.316      0.188      -2.
093      0.412
=====
Omnibus:      37245.487      Durbin-Watson:      0.166
Prob(Omnibus):      0.000      Jarque-Bera (JB):      1141.763
Skew:      0.077      Prob(JB):      1.17e-248
Kurtosis:      1.257      Cond. No.      2.03e+05
=====

```

## Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 2.03e+05. This might indicate that there are strong multicollinearity or other numerical problems.

```

In [38]: #Remove Yearly equip failure
lm30 = smf.ols('Tenure ~ Lat+Population+Children+Age+Email+Techie+OnlineBackup+DevicePr
lm30.params
print(lm30.summary())

```

## OLS Regression Results

```

=====
Dep. Variable:      Tenure      R-squared:      0.005
Model:      OLS      Adj. R-squared:      0.003
Method:      Least Squares      F-statistic:      2.226
Date:      Sat, 12 Feb 2022      Prob (F-statistic):      0.000827
Time:      23:36:30      Log-Likelihood:      -41986.
No. Observations:      8950      AIC:      8.402e+04
Df Residuals:      8927      BIC:      8.418e+04
Df Model:      22
Covariance Type:      nonrobust
=====
=====

```

	coef	std err	t	P> t	[0.
Intercept	41.8041	3.883	10.766	0.000	34.
Lat	-0.1289	0.060	-2.145	0.032	-0.
Population	-4.286e-05	2.43e-05	-1.764	0.078	-9.05e
Children	-0.1108	0.148	-0.749	0.454	-0.
Age	0.0170	0.014	1.251	0.211	-0.
Email	-0.1584	0.093	-1.707	0.088	-0.
Techie	-1.0086	0.748	-1.348	0.178	-2.
OnlineBackup	1.4176	0.634	2.236	0.025	0.
DeviceProtection	-1.4465	0.588	-2.461	0.014	-2.
StreamingTV	0.7827	0.788	0.994	0.320	-0.
StreamingMovies	1.0260	0.886	1.158	0.247	-0.
MonthlyCharge	-0.0169	0.013	-1.284	0.199	-0.

043	0.009					
Item4		0.3044	0.305	0.998	0.318	-0.
293	0.902					
Item5		0.5602	0.307	1.823	0.068	-0.
042	1.163					
Item8		-0.4637	0.281	-1.652	0.098	-1.
014	0.086					
Area_Rural		1.0936	0.593	1.846	0.065	-0.
068	2.255					
Marital_Divorced		-0.7558	0.688	-1.098	0.272	-2.
104	0.593					
Gender_Female		0.5614	0.566	0.992	0.321	-0.
548	1.671					
Gender_Nonbinary		-2.3319	1.882	-1.239	0.215	-6.
022	1.358					
Contract_TwoYear		1.2712	0.650	1.957	0.050	-0.
002	2.544					
InternetService_None		-0.9847	0.755	-1.304	0.192	-2.
464	0.495					
PaymentMethod_BankTransferautomatic		-1.2156	0.727	-1.672	0.095	-2.
640	0.209					
PaymentMethod_ElectronicCheck		-0.8322	0.639	-1.303	0.193	-2.
084	0.420					
=====						
Omnibus:	37243.108	Durbin-Watson:	0.166			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1141.860			
Skew:	0.077	Prob(JB):	1.12e-248			
Kurtosis:	1.257	Cond. No.	2.03e+05			
=====						

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 2.03e+05. This might indicate that there are strong multicollinearity or other numerical problems.

In [39]:

#Remove Children  
lm31 = smf.ols('Tenure ~ Lat+Population+Age+Email+Techie+OnlineBackup+DeviceProtection+  
lm31.params  
print(lm31.summary())

OLS Regression Results						
=====						
Dep. Variable:		Tenure	R-squared:		0.005	
Model:		OLS	Adj. R-squared:		0.003	
Method:		Least Squares	F-statistic:		2.305	
Date:		Sat, 12 Feb 2022	Prob (F-statistic):		0.000620	
Time:		23:36:30	Log-Likelihood:		-41987.	
No. Observations:		8950	AIC:		8.402e+04	
Df Residuals:		8928	BIC:		8.417e+04	
Df Model:		21				
Covariance Type:		nonrobust				
=====						
		coef	std err	t	P> t	[0.
025	0.975]					
-----						
Intercept		41.5434	3.867	10.743	0.000	33.
963	49.124					
Lat		-0.1289	0.060	-2.145	0.032	-0.
247	-0.011					
Population		-4.289e-05	2.43e-05	-1.765	0.078	-9.05e

-05	4.75e-06				
Age		0.0171	0.014	1.266	0.206
009	0.044				-0.
Email		-0.1583	0.093	-1.706	0.088
340	0.024				-0.
Techie		-1.0031	0.748	-1.341	0.180
470	0.464				-2.
OnlineBackup		1.4115	0.634	2.227	0.026
169	2.654				0.
DeviceProtection		-1.4544	0.588	-2.475	0.013
606	-0.302				-2.
StreamingTV		0.7783	0.788	0.988	0.323
766	2.323				-0.
StreamingMovies		1.0132	0.886	1.144	0.253
724	2.750				-0.
MonthlyCharge		-0.0167	0.013	-1.274	0.203
042	0.009				-0.
Item4		0.3079	0.305	1.010	0.313
290	0.906				-0.
Item5		0.5628	0.307	1.831	0.067
040	1.165				-0.
Item8		-0.4618	0.281	-1.646	0.100
012	0.088				-1.
Area_Rural		1.0859	0.592	1.833	0.067
075	2.247				-0.
Marital_Divorced		-0.7566	0.688	-1.100	0.272
105	0.592				-2.
Gender_Female		0.5645	0.566	0.997	0.319
545	1.674				-0.
Gender_Nonbinary		-2.3503	1.882	-1.249	0.212
039	1.339				-6.
Contract_TwoYear		1.2622	0.649	1.944	0.052
011	2.535				-0.
InternetService_None		-0.9858	0.755	-1.306	0.192
466	0.494				-2.
PaymentMethod_BankTransferautomatic		-1.2154	0.727	-1.672	0.095
640	0.210				-2.
PaymentMethod_ElectronicCheck		-0.8312	0.639	-1.301	0.193
083	0.421				-2.
=====					
Omnibus:	37228.682	Durbin-Watson:		0.166	
Prob(Omnibus):	0.000	Jarque-Bera (JB):		1142.385	
Skew:	0.077	Prob(JB):		8.60e-249	
Kurtosis:	1.257	Cond. No.		2.02e+05	
=====					

## Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 2.02e+05. This might indicate that there are strong multicollinearity or other numerical problems.

```
In [40]: #Remove streaming tv
lm32 = smf.ols('Tenure ~ Lat+Population+Age+Email+Techie+OnlineBackup+DeviceProtection+
lm32.params
print(lm32.summary())
```

## OLS Regression Results

=====			
Dep. Variable:	Tenure	R-squared:	0.005
Model:	OLS	Adj. R-squared:	0.003
Method:	Least Squares	F-statistic:	2.372
Date:	Sat, 12 Feb 2022	Prob (F-statistic):	0.000524

Time: 23:36:30 Log-Likelihood: -41987.  
 No. Observations: 8950 AIC: 8.402e+04  
 Df Residuals: 8929 BIC: 8.417e+04  
 Df Model: 20  
 Covariance Type: nonrobust

=====						
		coef	std err	t	P> t	[0.
025	0.975]					
-----						
Intercept		40.6257	3.754	10.822	0.000	33.
267	47.984					
Lat		-0.1281	0.060	-2.133	0.033	-0.
246	-0.010					
Population		-4.295e-05	2.43e-05	-1.767	0.077	-9.06e
-05	4.68e-06					
Age		0.0171	0.014	1.265	0.206	-0.
009	0.044					
Email		-0.1579	0.093	-1.702	0.089	-0.
340	0.024					
Techie		-1.0013	0.748	-1.338	0.181	-2.
468	0.465					
OnlineBackup		1.2035	0.598	2.013	0.044	0.
032	2.375					
DeviceProtection		-1.5710	0.576	-2.729	0.006	-2.
700	-0.443					
StreamingMovies		0.5290	0.738	0.717	0.474	-0.
918	1.976					
MonthlyCharge		-0.0076	0.009	-0.815	0.415	-0.
026	0.011					
Item4		0.3100	0.305	1.017	0.309	-0.
288	0.908					
Item5		0.5644	0.307	1.837	0.066	-0.
038	1.167					
Item8		-0.4578	0.281	-1.632	0.103	-1.
008	0.092					
Area_Rural		1.0918	0.592	1.843	0.065	-0.
069	2.253					
Marital_Divorced		-0.7543	0.688	-1.096	0.273	-2.
103	0.594					
Gender_Female		0.5551	0.566	0.981	0.327	-0.
554	1.664					
Gender_Nonbinary		-2.3424	1.882	-1.245	0.213	-6.
032	1.347					
Contract_TwoYear		1.2580	0.649	1.937	0.053	-0.
015	2.531					
InternetService_None		-0.7574	0.719	-1.054	0.292	-2.
166	0.651					
PaymentMethod_BankTransferautomatic		-1.2204	0.727	-1.679	0.093	-2.
645	0.204					
PaymentMethod_ElectronicCheck		-0.8331	0.639	-1.304	0.192	-2.
085	0.419					
=====						
Omnibus:	37210.733	Durbin-Watson:	0.166			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1143.021			
Skew:	0.077	Prob(JB):	6.25e-249			
Kurtosis:	1.256	Cond. No.	1.96e+05			
=====						

## Notes:

- [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
- [2] The condition number is large, 1.96e+05. This might indicate that there are strong multicollinearity or other numerical problems.

```
In [41]: #Remove streaming movies
lm33 = smf.ols('Tenure ~ Lat+Population+Age+Email+Techie+OnlineBackup+DeviceProtection+
lm33.params
print(lm33.summary())
```

```

=====
                        OLS Regression Results
=====
Dep. Variable:          Tenure      R-squared:                0.005
Model:                  OLS        Adj. R-squared:            0.003
Method:                 Least Squares    F-statistic:              2.470
Date:                  Sat, 12 Feb 2022    Prob (F-statistic):       0.000376
Time:                  23:36:30      Log-Likelihood:          -41987.
No. Observations:      8950          AIC:                     8.401e+04
Df Residuals:          8930          BIC:                     8.416e+04
Df Model:              19
Covariance Type:       nonrobust
=====
=====
                                coef      std err          t      P>|t|      [0.
025      0.975]
-----
-----
Intercept                40.1637      3.698      10.861      0.000      32.
915      47.413
Lat                    -0.1282      0.060      -2.135      0.033      -0.
246      -0.010
Population              -4.273e-05    2.43e-05    -1.759      0.079     -9.04e
-05      4.9e-06
Age                     0.0172      0.014       1.268      0.205      -0.
009      0.044
Email                  -0.1579      0.093      -1.701      0.089      -0.
340      0.024
Techie                 -1.0027      0.748      -1.340      0.180      -2.
469      0.464
OnlineBackup            1.1091      0.583       1.902      0.057      -0.
034      2.252
DeviceProtection       -1.6197      0.572      -2.833      0.005      -2.
740      -0.499
MonthlyCharge          -0.0032      0.007      -0.458      0.647      -0.
017      0.011
Item4                   0.3107      0.305       1.019      0.308      -0.
287      0.908
Item5                   0.5665      0.307       1.844      0.065      -0.
036      1.169
Item8                  -0.4563      0.281      -1.627      0.104      -1.
006      0.094
Area_Rural              1.0879      0.592       1.837      0.066      -0.
073      2.249
Marital_Divorced       -0.7477      0.688      -1.087      0.277      -2.
096      0.601
Gender_Female           0.5514      0.566       0.974      0.330      -0.
558      1.661
Gender_Nonbinary       -2.3492      1.882      -1.248      0.212      -6.
038      1.340
Contract_TwoYear        1.2548      0.649       1.932      0.053      -0.
018      2.528
InternetService_None   -0.6506      0.703      -0.925      0.355      -2.
028      0.727
PaymentMethod_BankTransfereautomatic -1.2319      0.727      -1.695      0.090      -2.
656      0.193
PaymentMethod_ElectronicCheck -0.8331      0.639      -1.304      0.192      -2.
085      0.419
=====
=====
```

Omnibus:37209.687

Durbin-Watson:0.166

Prob(Omnibus):0.000

Jarque-Bera (JB):1143.073

Skew:0.077

Prob(JB):6.09e-249

Kurtosis:1.256

Cond. No.1.93e+05

=====

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 1.93e+05. This might indicate that there are strong multicollinearity or other numerical problems.

In [42]:

#Remove monthly charge  
lm34 = smf.ols('Tenure ~ Lat+Population+Age+Email+Techie+OnlineBackup+DeviceProtection+  
lm34.params  
print(lm34.summary())

OLS Regression Results						
=====						
Dep. Variable:	Tenure	R-squared:	0.005			
Model:	OLS	Adj. R-squared:	0.003			
Method:	Least Squares	F-statistic:	2.595			
Date:	Sat, 12 Feb 2022	Prob (F-statistic):	0.000239			
Time:	23:36:30	Log-Likelihood:	-41988.			
No. Observations:	8950	AIC:	8.401e+04			
Df Residuals:	8931	BIC:	8.415e+04			
Df Model:	18					
Covariance Type:	nonrobust					
=====						
=====						
		coef	std err	t	P> t	[0.
025	0.975]					
-----						
Intercept		39.6415	3.517	11.270	0.000	32.
746	46.537					
Lat		-0.1282	0.060	-2.135	0.033	-0.
246	-0.010					
Population		-4.272e-05	2.43e-05	-1.758	0.079	-9.03e
-05	4.91e-06					
Age		0.0171	0.014	1.261	0.207	-0.
009	0.044					
Email		-0.1582	0.093	-1.705	0.088	-0.
340	0.024					
Techie		-1.0047	0.748	-1.343	0.179	-2.
471	0.462					
OnlineBackup		1.0369	0.561	1.847	0.065	-0.
063	2.137					
DeviceProtection		-1.6637	0.563	-2.953	0.003	-2.
768	-0.559					
Item4		0.3110	0.305	1.020	0.308	-0.
287	0.909					
Item5		0.5687	0.307	1.851	0.064	-0.
034	1.171					
Item8		-0.4568	0.281	-1.628	0.103	-1.
007	0.093					
Area_Rural		1.0886	0.592	1.838	0.066	-0.
073	2.250					
Marital_Divorced		-0.7515	0.688	-1.093	0.275	-2.
100	0.597					
Gender_Female		0.5543	0.566	0.979	0.327	-0.
555	1.664					
Gender_Nonbinary		-2.3516	1.882	-1.250	0.211	-6.



```

040      1.337
Contract_TwoYear      1.2537      0.649      1.931      0.054      -0.
019      2.526
InternetService_None      -0.5700      0.681      -0.838      0.402      -1.
904      0.764
PaymentMethod_BankTransferautomatic      -1.2299      0.727      -1.693      0.091      -2.
654      0.194
PaymentMethod_ElectronicCheck      -0.8292      0.639      -1.298      0.194      -2.
081      0.423
=====
Omnibus:      37210.191      Durbin-Watson:      0.166
Prob(Omnibus):      0.000      Jarque-Bera (JB):      1143.059
Skew:      0.077      Prob(JB):      6.14e-249
Kurtosis:      1.256      Cond. No.      1.83e+05
=====

```

## Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 1.83e+05. This might indicate that there are strong multicollinearity or other numerical problems.

```

In [43]: #Remove Internet service none
lm35 = smf.ols('Tenure ~ Lat+Population+Age+Email+Techie+OnlineBackup+DeviceProtection+
lm35.params
print(lm35.summary())

```

## OLS Regression Results

```

=====
Dep. Variable:      Tenure      R-squared:      0.005
Model:      OLS      Adj. R-squared:      0.003
Method:      Least Squares      F-statistic:      2.707
Date:      Sat, 12 Feb 2022      Prob (F-statistic):      0.000176
Time:      23:36:31      Log-Likelihood:      -41988.
No. Observations:      8950      AIC:      8.401e+04
Df Residuals:      8932      BIC:      8.414e+04
Df Model:      17
Covariance Type:      nonrobust
=====
=====
                                coef      std err      t      P>|t|      [0.
025      0.975]
-----
-----
Intercept      39.5417      3.515      11.248      0.000      32.
651      46.433
Lat      -0.1292      0.060      -2.151      0.032      -0.
247      -0.011
Population      -4.287e-05      2.43e-05      -1.764      0.078      -9.05e
-05      4.76e-06
Age      0.0170      0.014      1.253      0.210      -0.
010      0.044
Email      -0.1579      0.093      -1.702      0.089      -0.
340      0.024
Techie      -0.9978      0.748      -1.334      0.182      -2.
464      0.469
OnlineBackup      1.0406      0.561      1.854      0.064      -0.
060      2.141
DeviceProtection      -1.6621      0.563      -2.950      0.003      -2.
767      -0.558
Item4      0.3099      0.305      1.017      0.309      -0.
288      0.907
Item5      0.5690      0.307      1.852      0.064      -0.

```

2/12/22, 11:54 PMCustomerChurn\_LinearRegression

033	1.171					
Item8		-0.4527	0.280	-1.614	0.107	-1.
003	0.097					
Area_Rural		1.0868	0.592	1.835	0.067	-0.
074	2.248					
Marital_Divorced		-0.7483	0.688	-1.088	0.277	-2.
097	0.600					
Gender_Female		0.5544	0.566	0.980	0.327	-0.
555	1.664					
Gender_Nonbinary		-2.3673	1.882	-1.258	0.208	-6.
056	1.321					
Contract_TwoYear		1.2627	0.649	1.945	0.052	-0.
010	2.535					
PaymentMethod_BankTransferautomatic		-1.2262	0.727	-1.688	0.092	-2.
650	0.198					
PaymentMethod_ElectronicCheck		-0.8325	0.639	-1.303	0.192	-2.
085	0.419					
=====						
Omnibus:	37208.884	Durbin-Watson:		0.166		
Prob(Omnibus):	0.000	Jarque-Bera (JB):		1143.097		
Skew:	0.077	Prob(JB):		6.02e-249		
Kurtosis:	1.256	Cond. No.		1.83e+05		
=====						

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 1.83e+05. This might indicate that there are strong multicollinearity or other numerical problems.

In [44]:

#Remove Gender female  
lm36 = smf.ols('Tenure ~ Lat+Population+Age+Email+Techie+OnlineBackup+DeviceProtection'+  
lm36.params  
print(lm36.summary())

OLS Regression Results						
=====						
Dep. Variable:	Tenure	R-squared:		0.005		
Model:	OLS	Adj. R-squared:		0.003		
Method:	Least Squares	F-statistic:		2.816		
Date:	Sat, 12 Feb 2022	Prob (F-statistic):		0.000140		
Time:	23:36:31	Log-Likelihood:		-41988.		
No. Observations:	8950	AIC:		8.401e+04		
Df Residuals:	8933	BIC:		8.413e+04		
Df Model:	16					
Covariance Type:	nonrobust					
=====						
=====						
		coef	std err	t	P> t	[0.
025	0.975]					
-----						
Intercept		39.7715	3.508	11.339	0.000	32.
896	46.647					
Lat		-0.1284	0.060	-2.138	0.033	-0.
246	-0.011					
Population		-4.252e-05	2.43e-05	-1.750	0.080	-9.01e
-05	5.1e-06					
Age		0.0169	0.014	1.247	0.212	-0.
010	0.043					
Email		-0.1593	0.093	-1.716	0.086	-0.
341	0.023					
Techie		-0.9930	0.748	-1.327	0.184	-2.

459	0.473					
OnlineBackup		1.0449	0.561	1.861	0.063	-0.
055	2.145					
DeviceProtection		-1.6743	0.563	-2.972	0.003	-2.
779	-0.570					
Item4		0.3205	0.305	1.052	0.293	-0.
277	0.918					
Item5		0.5758	0.307	1.875	0.061	-0.
026	1.178					
Item8		-0.4552	0.280	-1.623	0.105	-1.
005	0.095					
Area_Rural		1.0904	0.592	1.841	0.066	-0.
071	2.252					
Marital_Divorced		-0.7596	0.688	-1.105	0.269	-2.
108	0.588					
Gender_Nonbinary		-2.6547	1.859	-1.428	0.153	-6.
298	0.989					
Contract_TwoYear		1.2551	0.649	1.933	0.053	-0.
017	2.528					
PaymentMethod_BankTransferautomatic		-1.2267	0.727	-1.688	0.091	-2.
651	0.198					
PaymentMethod_ElectronicCheck		-0.8420	0.639	-1.319	0.187	-2.
094	0.410					
=====						
Omnibus:	37193.879	Durbin-Watson:		0.166		
Prob(Omnibus):	0.000	Jarque-Bera (JB):		1143.645		
Skew:	0.077	Prob(JB):		4.58e-249		
Kurtosis:	1.256	Cond. No.		1.83e+05		
=====						

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 1.83e+05. This might indicate that there are strong multicollinearity or other numerical problems.

In [45]:

#Remove item 4  
lm37 = smf.ols('Tenure ~ Lat+Population+Age+Email+Techie+OnlineBackup+DeviceProtection+  
lm37.params  
print(lm37.summary())

OLS Regression Results						
=====						
Dep. Variable:	Tenure	R-squared:		0.005		
Model:	OLS	Adj. R-squared:		0.003		
Method:	Least Squares	F-statistic:		2.930		
Date:	Sat, 12 Feb 2022	Prob (F-statistic):		0.000115		
Time:	23:36:31	Log-Likelihood:		-41989.		
No. Observations:	8950	AIC:		8.401e+04		
Df Residuals:	8934	BIC:		8.412e+04		
Df Model:	15					
Covariance Type:	nonrobust					
=====						
=====						
		coef	std err	t	P> t	[0.
025	0.975]					
-----						
Intercept		41.2501	3.214	12.836	0.000	34.
951	47.550					
Lat		-0.1277	0.060	-2.126	0.034	-0.
245	-0.010					
Population		-4.291e-05	2.43e-05	-1.767	0.077	-9.05e

-05	4.7e-06					
Age		0.0170	0.014	1.255	0.209	-0.
010	0.044					
Email		-0.1600	0.093	-1.725	0.085	-0.
342	0.022					
Techie		-0.9975	0.748	-1.333	0.182	-2.
464	0.469					
OnlineBackup		1.0421	0.561	1.856	0.063	-0.
058	2.142					
DeviceProtection		-1.6804	0.563	-2.983	0.003	-2.
785	-0.576					
Item5		0.4420	0.280	1.581	0.114	-0.
106	0.990					
Item8		-0.4301	0.279	-1.539	0.124	-0.
978	0.118					
Area_Rural		1.0933	0.592	1.846	0.065	-0.
068	2.254					
Marital_Divorced		-0.7461	0.688	-1.085	0.278	-2.
094	0.602					
Gender_Nonbinary		-2.6311	1.859	-1.416	0.157	-6.
274	1.012					
Contract_TwoYear		1.2614	0.649	1.943	0.052	-0.
011	2.534					
PaymentMethod_BankTransferautomatic		-1.2326	0.727	-1.696	0.090	-2.
657	0.192					
PaymentMethod_ElectronicCheck		-0.8452	0.639	-1.324	0.186	-2.
097	0.407					
=====						
Omnibus:	37174.253	Durbin-Watson:	0.165			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1144.330			
Skew:	0.077	Prob(JB):	3.25e-249			
Kurtosis:	1.255	Cond. No.	1.67e+05			
=====						

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 1.67e+05. This might indicate that there are strong multicollinearity or other numerical problems.

```
In [46]: #Remove marital divorced
lm38 = smf.ols('Tenure ~ Lat+Population+Age+Email+Techie+OnlineBackup+DeviceProtection+
lm38.params
print(lm38.summary())
```

OLS Regression Results					
=====					
Dep. Variable:	Tenure	R-squared:	0.005		
Model:	OLS	Adj. R-squared:	0.003		
Method:	Least Squares	F-statistic:	3.055		
Date:	Sat, 12 Feb 2022	Prob (F-statistic):	9.57e-05		
Time:	23:36:31	Log-Likelihood:	-41990.		
No. Observations:	8950	AIC:	8.401e+04		
Df Residuals:	8935	BIC:	8.412e+04		
Df Model:	14				
Covariance Type:	nonrobust				
=====					
		coef	std err	t	P> t
025	0.975]				[0.
-----					
Intercept		41.0404	3.208	12.794	0.000
					34.

752	47.329					
Lat		-0.1275	0.060	-2.123	0.034	-0.
245	-0.010					
Population		-4.266e-05	2.43e-05	-1.756	0.079	-9.03e
-05	4.96e-06					
Age		0.0168	0.014	1.244	0.213	-0.
010	0.043					
Email		-0.1594	0.093	-1.718	0.086	-0.
341	0.023					
Techie		-0.9938	0.748	-1.329	0.184	-2.
460	0.473					
OnlineBackup		1.0504	0.561	1.871	0.061	-0.
050	2.151					
DeviceProtection		-1.6682	0.563	-2.962	0.003	-2.
772	-0.564					
Item5		0.4469	0.280	1.599	0.110	-0.
101	0.995					
Item8		-0.4235	0.279	-1.516	0.130	-0.
971	0.124					
Area_Rural		1.0839	0.592	1.830	0.067	-0.
077	2.245					
Gender_Nonbinary		-2.6264	1.859	-1.413	0.158	-6.
270	1.017					
Contract_TwoYear		1.2665	0.649	1.951	0.051	-0.
006	2.539					
PaymentMethod_BankTransferautomatic		-1.2362	0.727	-1.701	0.089	-2.
661	0.188					
PaymentMethod_ElectronicCheck		-0.8473	0.639	-1.327	0.185	-2.
099	0.405					
=====						
Omnibus:	37151.896	Durbin-Watson:	0.165			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1145.157			
Skew:	0.077	Prob(JB):	2.15e-249			
Kurtosis:	1.254	Cond. No.	1.67e+05			
=====						

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 1.67e+05. This might indicate that there are strong multicollinearity or other numerical problems.

In [47]:

#Remove age  
lm39 = smf.ols('Tenure ~ Lat+Population+Email+Techie+OnlineBackup+DeviceProtection+Item  
lm39.params  
print(lm39.summary())

OLS Regression Results						
=====						
Dep. Variable:	Tenure	R-squared:	0.005			
Model:	OLS	Adj. R-squared:	0.003			
Method:	Least Squares	F-statistic:	3.171			
Date:	Sat, 12 Feb 2022	Prob (F-statistic):	9.00e-05			
Time:	23:36:31	Log-Likelihood:	-41990.			
No. Observations:	8950	AIC:	8.401e+04			
Df Residuals:	8936	BIC:	8.411e+04			
Df Model:	13					
Covariance Type:	nonrobust					
=====						
		coef	std err	t	P> t	[0.
025	0.975]					
-----						

-----						
Intercept		41.9276	3.128	13.405	0.000	35.
797	48.059					
Lat		-0.1278	0.060	-2.128	0.033	-0.
246	-0.010					
Population		-4.216e-05	2.43e-05	-1.736	0.083	-8.98e
-05	5.45e-06					
Email		-0.1594	0.093	-1.718	0.086	-0.
341	0.022					
Techie		-0.9968	0.748	-1.333	0.183	-2.
463	0.469					
OnlineBackup		1.0507	0.561	1.872	0.061	-0.
050	2.151					
DeviceProtection		-1.6562	0.563	-2.941	0.003	-2.
760	-0.552					
Item5		0.4443	0.280	1.590	0.112	-0.
104	0.992					
Item8		-0.4181	0.279	-1.496	0.135	-0.
966	0.130					
Area_Rural		1.0787	0.592	1.821	0.069	-0.
082	2.240					
Gender_Nonbinary		-2.6839	1.858	-1.444	0.149	-6.
326	0.958					
Contract_TwoYear		1.2659	0.649	1.950	0.051	-0.
007	2.538					
PaymentMethod_BankTransferautomatic		-1.2187	0.726	-1.678	0.093	-2.
643	0.205					
PaymentMethod_ElectronicCheck		-0.8437	0.639	-1.321	0.187	-2.
096	0.408					
=====						
Omnibus:	37132.605	Durbin-Watson:	0.165			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1145.890			
Skew:	0.077	Prob(JB):	1.49e-249			
Kurtosis:	1.254	Cond. No.	1.63e+05			
=====						

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 1.63e+05. This might indicate that there are strong multicollinearity or other numerical problems.

In [48]:

#Remove payment method electronic check  
lm40 = smf.ols('Tenure ~ Lat+Population+Email+Techie+OnlineBackup+DeviceProtection+Item  
lm40.params  
print(lm40.summary())

OLS Regression Results					
=====					
Dep. Variable:	Tenure	R-squared:	0.004		
Model:	OLS	Adj. R-squared:	0.003		
Method:	Least Squares	F-statistic:	3.289		
Date:	Sat, 12 Feb 2022	Prob (F-statistic):	9.01e-05		
Time:	23:36:31	Log-Likelihood:	-41991.		
No. Observations:	8950	AIC:	8.401e+04		
Df Residuals:	8937	BIC:	8.410e+04		
Df Model:	12				
Covariance Type:	nonrobust				
=====					
=====					
		coef	std err	t	P> t  [0.
025	0.975]				
-----					

```
-----
Intercept                41.5285      3.113      13.340      0.000      35.
426      47.631
Lat                      -0.1281      0.060      -2.133      0.033      -0.
246      -0.010
Population               -4.181e-05    2.43e-05    -1.721      0.085     -8.94e
-05      5.8e-06
Email                   -0.1572      0.093      -1.694      0.090      -0.
339      0.025
Techie                  -0.9912      0.748      -1.325      0.185      -2.
458      0.475
OnlineBackup            1.0445      0.561      1.861      0.063      -0.
056      2.145
DeviceProtection       -1.6605      0.563      -2.949      0.003      -2.
764      -0.557
Item5                   0.4472      0.280      1.600      0.110      -0.
101      0.995
Item8                  -0.4153      0.279      -1.487      0.137      -0.
963      0.132
Area_Rural             1.0686      0.592      1.804      0.071      -0.
092      2.230
Gender_Nonbinary       -2.6967      1.858      -1.451      0.147      -6.
339      0.946
Contract_TwoYear       1.2578      0.649      1.938      0.053      -0.
015      2.530
PaymentMethod_BankTransferautomatic -0.8490      0.670      -1.266      0.205      -2.
163      0.465
=====
```

```
Omnibus:                37116.595    Durbin-Watson:           0.164
Prob(Omnibus):          0.000    Jarque-Bera (JB):       1146.432
Skew:                   0.077    Prob(JB):               1.14e-249
Kurtosis:               1.253    Cond. No.               1.62e+05
=====
```

## Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 1.62e+05. This might indicate that there are strong multicollinearity or other numerical problems.

```
In [49]: #Remove PaymentMethod_BankTransferautomatic
lm41 = smf.ols('Tenure ~ Lat+Population+Email+Techie+OnlineBackup+DeviceProtection+Item
lm41.params
print(lm41.summary())
```

## OLS Regression Results

```
=====
Dep. Variable:          Tenure      R-squared:                0.004
Model:                  OLS         Adj. R-squared:           0.003
Method:                 Least Squares   F-statistic:              3.442
Date:                  Sat, 12 Feb 2022   Prob (F-statistic):       8.44e-05
Time:                  23:36:32         Log-Likelihood:          -41992.
No. Observations:      8950            AIC:                    8.401e+04
Df Residuals:          8938            BIC:                    8.409e+04
Df Model:              11
Covariance Type:       nonrobust
=====
```

	coef	std err	t	P> t	[0.025	0.975]
Intercept	41.3442	3.110	13.294	0.000	35.248	47.440
Lat	-0.1287	0.060	-2.142	0.032	-0.246	-0.011
Population	-4.184e-05	2.43e-05	-1.723	0.085	-8.95e-05	5.77e-06
Email	-0.1556	0.093	-1.678	0.093	-0.337	0.026

Techie	-0.9937	0.748	-1.328	0.184	-2.460	0.473
OnlineBackup	1.0460	0.561	1.864	0.062	-0.054	2.146
DeviceProtection	-1.6717	0.563	-2.969	0.003	-2.776	-0.568
Item5	0.4456	0.280	1.594	0.111	-0.102	0.993
Item8	-0.4140	0.279	-1.482	0.138	-0.962	0.134
Area_Rural	1.0797	0.592	1.823	0.068	-0.081	2.241
Gender_Nonbinary	-2.7136	1.858	-1.460	0.144	-6.356	0.929
Contract_TwoYear	1.2557	0.649	1.934	0.053	-0.017	2.528

```
=====
Omnibus:                37090.878    Durbin-Watson:                0.164
Prob(Omnibus):          0.000    Jarque-Bera (JB):            1147.408
Skew:                   0.077    Prob(JB):                    6.98e-250
Kurtosis:               1.253    Cond. No.                    1.62e+05
=====
```

## Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 1.62e+05. This might indicate that there are strong multicollinearity or other numerical problems.

```
In [50]: #Remove Techie
lm42 = smf.ols('Tenure ~ Lat+Population+Email+OnlineBackup+DeviceProtection+Item5+Item8
lm42.params
print(lm42.summary())
```

## OLS Regression Results

```
=====
Dep. Variable:          Tenure    R-squared:                0.004
Model:                  OLS      Adj. R-squared:           0.003
Method:                 Least Squares    F-statistic:             3.610
Date:                  Sat, 12 Feb 2022    Prob (F-statistic):       8.28e-05
Time:                  23:36:32    Log-Likelihood:          -41993.
No. Observations:      8950    AIC:                     8.401e+04
Df Residuals:          8939    BIC:                     8.409e+04
Df Model:              10
Covariance Type:       nonrobust
=====
```

	coef	std err	t	P> t	[0.025	0.975]
Intercept	41.1271	3.106	13.242	0.000	35.039	47.215
Lat	-0.1282	0.060	-2.134	0.033	-0.246	-0.010
Population	-4.158e-05	2.43e-05	-1.712	0.087	-8.92e-05	6.03e-06
Email	-0.1535	0.093	-1.655	0.098	-0.335	0.028
OnlineBackup	1.0412	0.561	1.855	0.064	-0.059	2.142
DeviceProtection	-1.6863	0.563	-2.995	0.003	-2.790	-0.583
Item5	0.4448	0.280	1.591	0.112	-0.103	0.993
Item8	-0.4101	0.279	-1.468	0.142	-0.958	0.138
Area_Rural	1.0805	0.592	1.825	0.068	-0.080	2.241
Gender_Nonbinary	-2.6912	1.858	-1.448	0.148	-6.334	0.951
Contract_TwoYear	1.2630	0.649	1.945	0.052	-0.010	2.536

```
=====
Omnibus:                37056.809    Durbin-Watson:                0.164
Prob(Omnibus):          0.000    Jarque-Bera (JB):            1148.471
Skew:                   0.076    Prob(JB):                    4.10e-250
Kurtosis:               1.252    Cond. No.                    1.62e+05
=====
```

## Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 1.62e+05. This might indicate that there are strong multicollinearity or other numerical problems.



```
In [51]: #Remove Gener nonbinary
lm43 = smf.ols('Tenure ~ Lat+Population+Email+OnlineBackup+DeviceProtection+Item5+Item8')
lm43.params
print(lm43.summary())
```

```

=====
                        OLS Regression Results
=====
Dep. Variable:          Tenure      R-squared:                0.004
Model:                  OLS         Adj. R-squared:           0.003
Method:                 Least Squares   F-statistic:              3.777
Date:                   Sat, 12 Feb 2022   Prob (F-statistic):       9.13e-05
Time:                   23:36:32         Log-Likelihood:          -41994.
No. Observations:      8950            AIC:                    8.401e+04
Df Residuals:          8940            BIC:                    8.408e+04
Df Model:               9
Covariance Type:       nonrobust
=====
                        coef      std err          t      P>|t|      [0.025      0.975]
-----
Intercept              41.0554        3.106     13.220     0.000     34.968     47.143
Lat                   -0.1284        0.060     -2.138     0.033     -0.246     -0.011
Population            -4.162e-05    2.43e-05    -1.714     0.087    -8.92e-05    5.99e-06
Email                 -0.1528        0.093     -1.647     0.100     -0.335     0.029
OnlineBackup           1.0220        0.561      1.821     0.069     -0.078     2.122
DeviceProtection       -1.6670        0.563     -2.962     0.003     -2.770     -0.564
Item5                  0.4467        0.280      1.598     0.110     -0.101     0.995
Item8                 -0.4093        0.279     -1.465     0.143     -0.957     0.138
Area_Rural             1.0766        0.592      1.818     0.069     -0.084     2.238
Contract_TwoYear       1.2691        0.649      1.955     0.051     -0.003     2.542
=====
Omnibus:                37001.335    Durbin-Watson:           0.163
Prob(Omnibus):           0.000    Jarque-Bera (JB):        1150.413
Skew:                    0.076    Prob(JB):                1.55e-250
Kurtosis:                1.250    Cond. No.                1.62e+05
=====
```

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 1.62e+05. This might indicate that there are strong multicollinearity or other numerical problems.

```
In [52]: #Remove item 8
lm44 = smf.ols('Tenure ~ Lat+Population+Email+OnlineBackup+DeviceProtection+Item5+Area_')
lm44.params
print(lm44.summary())
```

```

=====
                        OLS Regression Results
=====
Dep. Variable:          Tenure      R-squared:                0.004
Model:                  OLS         Adj. R-squared:           0.003
Method:                 Least Squares   F-statistic:              3.981
Date:                   Sat, 12 Feb 2022   Prob (F-statistic):       0.000101
Time:                   23:36:32         Log-Likelihood:          -41995.
No. Observations:      8950            AIC:                    8.401e+04
Df Residuals:          8941            BIC:                    8.407e+04
Df Model:               8
Covariance Type:       nonrobust
=====
                        coef      std err          t      P>|t|      [0.025      0.975]
-----
```

Intercept	39.2803	2.859	13.737	0.000	33.675	44.886
Lat	-0.1268	0.060	-2.111	0.035	-0.245	-0.009
Population	-4.184e-05	2.43e-05	-1.722	0.085	-8.95e-05	5.78e-06
Email	-0.1509	0.093	-1.627	0.104	-0.333	0.031
OnlineBackup	1.0256	0.561	1.827	0.068	-0.075	2.126
DeviceProtection	-1.6751	0.563	-2.976	0.003	-2.778	-0.572
Item5	0.5239	0.275	1.908	0.056	-0.014	1.062
Area_Rural	1.0630	0.592	1.795	0.073	-0.098	2.224
Contract_TwoYear	1.2787	0.649	1.970	0.049	0.006	2.551

Omnibus:	36971.504	Durbin-Watson:	0.162
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1151.349
Skew:	0.075	Prob(JB):	9.72e-251
Kurtosis:	1.249	Cond. No.	1.49e+05

## Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 1.49e+05. This might indicate that there are strong multicollinearity or other numerical problems.

```
In [53]: #Remove email
lm45 = smf.ols('Tenure ~ Lat+Population+OnlineBackup+DeviceProtection+Item5+Area_Rural+
lm45.params
print(lm45.summary())
```

## OLS Regression Results

Dep. Variable:	Tenure	R-squared:	0.003
Model:	OLS	Adj. R-squared:	0.002
Method:	Least Squares	F-statistic:	4.171
Date:	Sat, 12 Feb 2022	Prob (F-statistic):	0.000135
Time:	23:36:32	Log-Likelihood:	-41996.
No. Observations:	8950	AIC:	8.401e+04
Df Residuals:	8942	BIC:	8.407e+04
Df Model:	7		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
Intercept	37.3712	2.608	14.331	0.000	32.259	42.483
Lat	-0.1245	0.060	-2.073	0.038	-0.242	-0.007
Population	-4.243e-05	2.43e-05	-1.747	0.081	-9e-05	5.18e-06
OnlineBackup	1.0360	0.561	1.846	0.065	-0.064	2.136
DeviceProtection	-1.6785	0.563	-2.982	0.003	-2.782	-0.575
Item5	0.5265	0.275	1.918	0.055	-0.012	1.065
Area_Rural	1.0582	0.592	1.787	0.074	-0.103	2.219
Contract_TwoYear	1.2840	0.649	1.978	0.048	0.011	2.557

Omnibus:	36932.089	Durbin-Watson:	0.162
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1152.717
Skew:	0.075	Prob(JB):	4.91e-251
Kurtosis:	1.248	Cond. No.	1.36e+05

## Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 1.36e+05. This might indicate that there are strong multicollinearity or other numerical problems.

```
In [54]: #Remove population
lm46 = smf.ols('Tenure ~ Lat+OnlineBackup+DeviceProtection+Item5+Area_Rural+Contract_Two
lm46.params
print(lm46.summary())
```

#### OLS Regression Results

```
=====
Dep. Variable:          Tenure      R-squared:                0.003
Model:                  OLS        Adj. R-squared:            0.002
Method:                 Least Squares    F-statistic:            4.356
Date:                  Sat, 12 Feb 2022    Prob (F-statistic):      0.000213
Time:                  23:36:33      Log-Likelihood:         -41998.
No. Observations:      8950        AIC:                   8.401e+04
Df Residuals:          8943        BIC:                   8.406e+04
Df Model:               6
Covariance Type:       nonrobust
=====
```

	coef	std err	t	P> t	[0.025	0.975]
Intercept	36.1641	2.515	14.380	0.000	31.234	41.094
Lat	-0.1024	0.059	-1.745	0.081	-0.218	0.013
OnlineBackup	1.0267	0.561	1.829	0.067	-0.073	2.127
DeviceProtection	-1.6740	0.563	-2.973	0.003	-2.778	-0.570
Item5	0.5260	0.275	1.916	0.055	-0.012	1.064
Area_Rural	1.0544	0.592	1.780	0.075	-0.107	2.215
Contract_TwoYear	1.2663	0.649	1.950	0.051	-0.006	2.539

```
=====
Omnibus:                 36904.363    Durbin-Watson:           0.161
Prob(Omnibus):           0.000        Jarque-Bera (JB):        1153.792
Skew:                    0.075        Prob(JB):                2.87e-251
Kurtosis:                1.247        Cond. No.                354.
=====
```

#### Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

```
In [55]: #Remove Lat
lm47 = smf.ols('Tenure ~ OnlineBackup+DeviceProtection+Item5+Area_Rural+Contract_TwoYea
lm47.params
print(lm47.summary())
```

#### OLS Regression Results

```
=====
Dep. Variable:          Tenure      R-squared:                0.003
Model:                  OLS        Adj. R-squared:            0.002
Method:                 Least Squares    F-statistic:            4.618
Date:                  Sat, 12 Feb 2022    Prob (F-statistic):      0.000328
Time:                  23:36:33      Log-Likelihood:         -41999.
No. Observations:      8950        AIC:                   8.401e+04
Df Residuals:          8944        BIC:                   8.405e+04
Df Model:               5
Covariance Type:       nonrobust
=====
```

	coef	std err	t	P> t	[0.025	0.975]
Intercept	32.2039	1.082	29.757	0.000	30.082	34.325
OnlineBackup	1.0316	0.561	1.838	0.066	-0.069	2.132
DeviceProtection	-1.6879	0.563	-2.998	0.003	-2.792	-0.584
Item5	0.5252	0.275	1.912	0.056	-0.013	1.064
Area_Rural	1.0518	0.592	1.776	0.076	-0.109	2.213
Contract_TwoYear	1.2600	0.649	1.940	0.052	-0.013	2.533

```
=====
```

Omnibus:	36853.829	Durbin-Watson:	0.161
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1155.678
Skew:	0.075	Prob(JB):	1.12e-251
Kurtosis:	1.246	Cond. No.	15.6

=====

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

```
In [56]: #Remove Area Rural
lm48 = smf.ols('Tenure ~ OnlineBackup+DeviceProtection+Item5+Contract_TwoYear', data =
lm48.params
print(lm48.summary())
```

```

=====
                        OLS Regression Results
=====
Dep. Variable:          Tenure      R-squared:          0.002
Model:                  OLS         Adj. R-squared:       0.002
Method:                 Least Squares   F-statistic:         4.983
Date:                   Sat, 12 Feb 2022   Prob (F-statistic):   0.000520
Time:                   23:36:33          Log-Likelihood:      -42001.
No. Observations:      8950             AIC:                8.401e+04
Df Residuals:          8945             BIC:                8.405e+04
Df Model:               4
Covariance Type:       nonrobust
=====

```

	coef	std err	t	P> t	[0.025	0.975]
Intercept	32.5609	1.064	30.616	0.000	30.476	34.646
OnlineBackup	1.0280	0.561	1.831	0.067	-0.073	2.128
DeviceProtection	-1.6986	0.563	-3.017	0.003	-2.802	-0.595
Item5	0.5253	0.275	1.913	0.056	-0.013	1.064
Contract_TwoYear	1.2585	0.649	1.938	0.053	-0.015	2.531

```

=====
Omnibus:          36807.122      Durbin-Watson:          0.160
Prob(Omnibus):    0.000          Jarque-Bera (JB):       1157.410
Skew:             0.075          Prob(JB):              4.70e-252
Kurtosis:         1.245          Cond. No.              15.2
=====

```

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

```
In [57]: #Remove Area Rural
lm49= smf.ols('Tenure ~ DeviceProtection+Item5+Contract_TwoYear', data = data).fit()
lm49.params
print(lm49.summary())
```

```

=====
                        OLS Regression Results
=====
Dep. Variable:          Tenure      R-squared:          0.002
Model:                  OLS         Adj. R-squared:       0.002
Method:                 Least Squares   F-statistic:         5.525
Date:                   Sat, 12 Feb 2022   Prob (F-statistic):   0.000870
Time:                   23:36:33          Log-Likelihood:      -42003.
No. Observations:      8950             AIC:                8.401e+04
Df Residuals:          8946             BIC:                8.404e+04
Df Model:               3
Covariance Type:       nonrobust
=====

```

	coef	std err	t	P> t	[0.025	0.975]
Intercept	32.9882	1.038	31.789	0.000	30.954	35.022
DeviceProtection	-1.6859	0.563	-2.994	0.003	-2.790	-0.582
Item5	0.5343	0.275	1.945	0.052	-0.004	1.073
Contract_TwoYear	1.2557	0.649	1.933	0.053	-0.017	2.529
=====						
Omnibus:	36754.208		Durbin-Watson:		0.159	
Prob(Omnibus):	0.000		Jarque-Bera (JB):		1159.329	
Skew:	0.075		Prob(JB):		1.80e-252	
Kurtosis:	1.243		Cond. No.		14.7	

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

```
In [58]: #Remove Device Contract 2 year
lm50= smf.ols('Tenure ~ DeviceProtection+Item5', data = data).fit()
lm50.params
print(lm50.summary())
```

OLS Regression Results						
=====						
Dep. Variable:	Tenure	R-squared:	0.001			
Model:	OLS	Adj. R-squared:	0.001			
Method:	Least Squares	F-statistic:	6.416			
Date:	Sat, 12 Feb 2022	Prob (F-statistic):	0.00164			
Time:	23:36:33	Log-Likelihood:	-42005.			
No. Observations:	8950	AIC:	8.402e+04			
Df Residuals:	8947	BIC:	8.404e+04			
Df Model:	2					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]
-----						
Intercept	33.2769	1.027	32.399	0.000	31.264	35.290
DeviceProtection	-1.6931	0.563	-3.006	0.003	-2.797	-0.589
Item5	0.5407	0.275	1.968	0.049	0.002	1.079
=====						
Omnibus:	36703.300	Durbin-Watson:	0.158			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1161.262			
Skew:	0.075	Prob(JB):	6.84e-253			
Kurtosis:	1.242	Cond. No.	14.5			
=====						

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

```
In [59]: #Remove Adding back in Bandwidth
lm51= smf.ols('Tenure ~ DeviceProtection+Item5+Bandwidth_GB_Year', data = data).fit()
lm51.params
print(lm51.summary())
```

OLS Regression Results						
=====						
Dep. Variable:	Tenure	R-squared:	0.984			
Model:	OLS	Adj. R-squared:	0.984			
Method:	Least Squares	F-statistic:	1.778e+05			
Date:	Sat, 12 Feb 2022	Prob (F-statistic):	0.00			
Time:	23:36:34	Log-Likelihood:	-23641.			

No. Observations: 8950 AIC: 4.729e+04  
 Df Residuals: 8946 BIC: 4.732e+04  
 Df Model: 3  
 Covariance Type: nonrobust

```
=====
              coef      std err          t      P>|t|      [0.025      0.975]
-----
Intercept      -5.6204      0.142     -39.483      0.000     -5.899     -5.341
DeviceProtection -1.1020      0.072     -15.226      0.000     -1.244     -0.960
Item5           -0.0044      0.035      -0.125      0.900     -0.074      0.065
Bandwidth_GB_Year 0.0120     1.64e-05     729.897      0.000      0.012      0.012
=====
Omnibus:            301.677   Durbin-Watson:           1.943
Prob(Omnibus):      0.000   Jarque-Bera (JB):        239.061
Skew:               -0.318   Prob(JB):                1.23e-52
Kurtosis:           2.515   Cond. No.                1.64e+04
=====
```

#### Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 1.64e+04. This might indicate that there are strong multicollinearity or other numerical problems.

```
In [60]: #Remove item 5
lm52= smf.ols('Tenure ~ DeviceProtection+Bandwidth_GB_Year', data = data).fit()
lm52.params
print(lm52.summary())
```

#### OLS Regression Results

```
=====
Dep. Variable:      Tenure      R-squared:            0.984
Model:              OLS        Adj. R-squared:         0.984
Method:             Least Squares   F-statistic:         2.668e+05
Date:               Sat, 12 Feb 2022   Prob (F-statistic):    0.00
Time:               23:36:34      Log-Likelihood:       -23641.
No. Observations:   8950          AIC:                4.729e+04
Df Residuals:       8947          BIC:                4.731e+04
Df Model:           2
Covariance Type:    nonrobust
=====
              coef      std err          t      P>|t|      [0.025      0.975]
-----
Intercept      -5.6356      0.074     -76.593      0.000     -5.780     -5.491
DeviceProtection -1.1021      0.072     -15.228      0.000     -1.244     -0.960
Bandwidth_GB_Year 0.0120     1.64e-05     730.098      0.000      0.012      0.012
=====
Omnibus:            301.736   Durbin-Watson:           1.943
Prob(Omnibus):      0.000   Jarque-Bera (JB):        239.079
Skew:               -0.318   Prob(JB):                1.22e-52
Kurtosis:           2.515   Cond. No.                9.81e+03
=====
```

#### Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 9.81e+03. This might indicate that there are strong multicollinearity or other numerical problems.

## Comparison with Simple regression

```
In [61]: #Remove item 5
lm53= smf.ols('Tenure ~ Bandwidth_GB_Year', data = data).fit()
lm53.params
print(lm53.summary())
```

```

                                OLS Regression Results
=====
Dep. Variable:                  Tenure      R-squared:                0.983
Model:                            OLS      Adj. R-squared:           0.983
Method:                 Least Squares      F-statistic:              5.199e+05
Date:                  Sat, 12 Feb 2022     Prob (F-statistic):       0.00
Time:                  23:36:34             Log-Likelihood:          -23756.
No. Observations:                8950      AIC:                    4.752e+04
Df Residuals:                    8948      BIC:                    4.753e+04
Df Model:                        1
Covariance Type:                nonrobust
=====
                                coef      std err          t      P>|t|      [0.025      0.975]
-----
Intercept                   -6.1272      0.067     -91.494      0.000     -6.258     -5.996
Bandwidth_GB_Year           0.0120     1.66e-05     721.067      0.000      0.012      0.012
=====
Omnibus:                    269.275    Durbin-Watson:           1.939
Prob(Omnibus):              0.000    Jarque-Bera (JB):        223.481
Skew:                       -0.315    Prob(JB):                2.96e-49
Kurtosis:                   2.551    Cond. No.                7.41e+03
=====
```

#### Notes:

- [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
- [2] The condition number is large, 7.41e+03. This might indicate that there are strong multicollinearity or other numerical problems.

## 2. Justifying a statistically based variable selection procedure and a model evaluation metric to reduce the initial model in a way that aligns with the research question.

Initially, I began by performing a backwards stepwise regression, however, after running a few models, the adjusted R2 was consistently 1. This indicated potential problems with my model. I ran a correlation matrix on my data and found that Bandwidth\_GB\_Year has a 0.99 correlation with Tenure indicating an almost perfect linear relationship. For the purpose of this project, I decided to remove Bandwidth\_GB\_Year from the regression and try to find any variables that might be significant indicators of Bandwidth. These variables were also reduced by removing the least significant variable at each stage of the model reduction. Ultimately I was left with two variables with p-values less than 0.05. Since these variables accounted for 0.001 of the variability in the data, I added Bandwidth\_GB\_Year into the regression and removed item 5 due to its p-value. This left me with a model that explained 0.984 of the variability in the data.

## 3. Providing a reduced multiple regression model that includes both categorical and continuous variables.

Tenure ~ -1.1021(DeviceProtection)+ 0.0120(Bandwidth\_GB\_Year)-5.6356

```
In [62]: linearmodeldata= data[['DeviceProtection', 'Bandwidth_GB_Year', 'Tenure']]
linearmodeldata.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8950 entries, 0 to 8949
Data columns (total 3 columns):
#   Column                Non-Null Count  Dtype
---  -
0   DeviceProtection      8950 non-null   int64
1   Bandwidth_GB_Year     8950 non-null   float64
2   Tenure                8950 non-null   float64
dtypes: float64(2), int64(1)
memory usage: 209.9 KB
```

## E. Analyzing the data set using the reduced multiple regression model

### 1. Explaining data analysis process by comparing the initial and reduced multiple regression models, including the following elements:

- the logic of the variable selection technique
- the model evaluation metric

The logic behind selecting so many variables initially was that there may be relationships that were not intuitively related to Tenure. As a result, I included anything related to customer demographics, customer habits, and customer surveys. Initially, I decided to drop 'CaseOrder', 'Customer\_id', 'Interaction', 'UID', 'City', 'TimeZone', 'State', 'County', 'Zip', 'Churn', and 'Job' from the original data set since I felt some of these variables were not necessarily useful such as CaseOrder', 'Customer\_id', 'Interaction', 'UID'. I also removed some columns which I felt duplicated location information such as 'TimeZone', 'City', 'State', 'County', and 'Zip' and I instead kept the location variable with the highest level of granularity, lat and lng. Finally, I removed job and Churn from the data set because I felt like the job column was not useful without further natural language processing and I removed Churn since it was a response variable not a predictor.

Once the initial model was constructed I began to refine the model by performing a stepwise regression. At every stage the least significant variable (based on P value) was eliminated and a new model was generated. As explained above, the model appeared to have some issues caused by the high correlation with Bandwidth\_GB\_Year and Tenure. This variable was removed from the regression and the remaining variables were reduced one at a time based on the P-value. Once I was left with only significant variables I re-added Bandwidth\_GB\_Year and I reduced the last remaining variable with a P value above 0.05.

- a residual plot

Please see the figures below.

### 2. Providing the output and any calculations of the analysis performed, including the model's residual error.



Please see the calculations below for the model's residual error and the code above for the regression outputs.

### 3. Providing the code used to support the implementation of the multiple regression models.

Please see the code above.

## Part V: Data Summary and Implications

### F. Summarizing the findings and assumptions

#### 1. Discussing the results of the data analysis, including the following elements:

- a regression equation for the reduced model

$$y = b_1x_1 + b_2x_2$$

$$\text{Tenure} \sim -1.1021(\text{DeviceProtection}) + 0.0120(\text{Bandwidth\_GB\_Year}) - 5.6356$$

- an interpretation of coefficients of the statistically significant variables of the model

For every one unit of Tenure, you can expect to decrease by an average of 1.1021 and you can expect Bandwidth\_GB\_Year to increase by 0.0120.

- the statistical and practical significance of the model

The P-values provided in the final model suggest that all variables in the final model are statistically significant. While comparing the final model with the simple regression, the AIC appears to be better in the simple regression indicating that it may be better at predicting Tenure than the model which includes Device Protection.

- the limitations of the data analysis

As previously mentioned, one limitation of this model is that the AIC for the simple regression is better than the AIC that includes the additional categorical variable. While more variance is explained in the final model which includes device protection, the simple regression will probably perform better with other data. The residuals plotted below indicate that the distribution of errors is fairly normal, with some issues around the tails. The variance is also not constant, it appears to increase and decrease throughout the partial residuals.

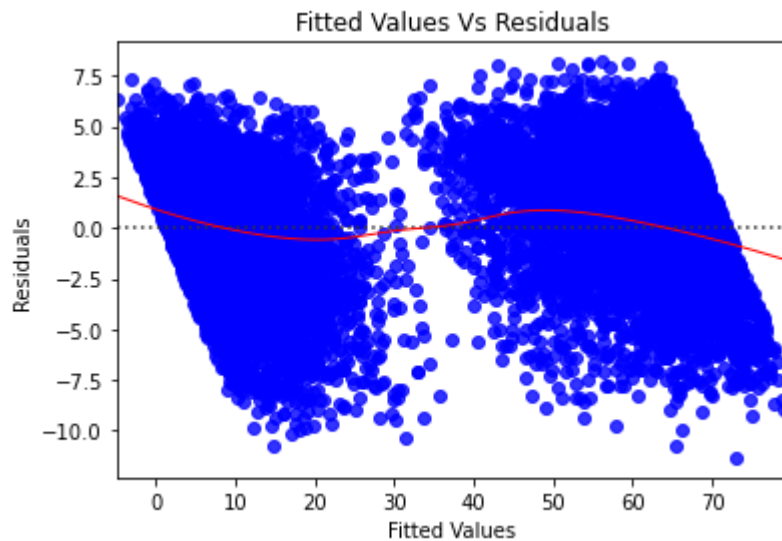
#### 2. Recommend a course of action based on your results.

Based on the residuals, I might try to apply a different model to see if the behavior can be better predicted, however, for the sake of this company I would suggest that they use the simple

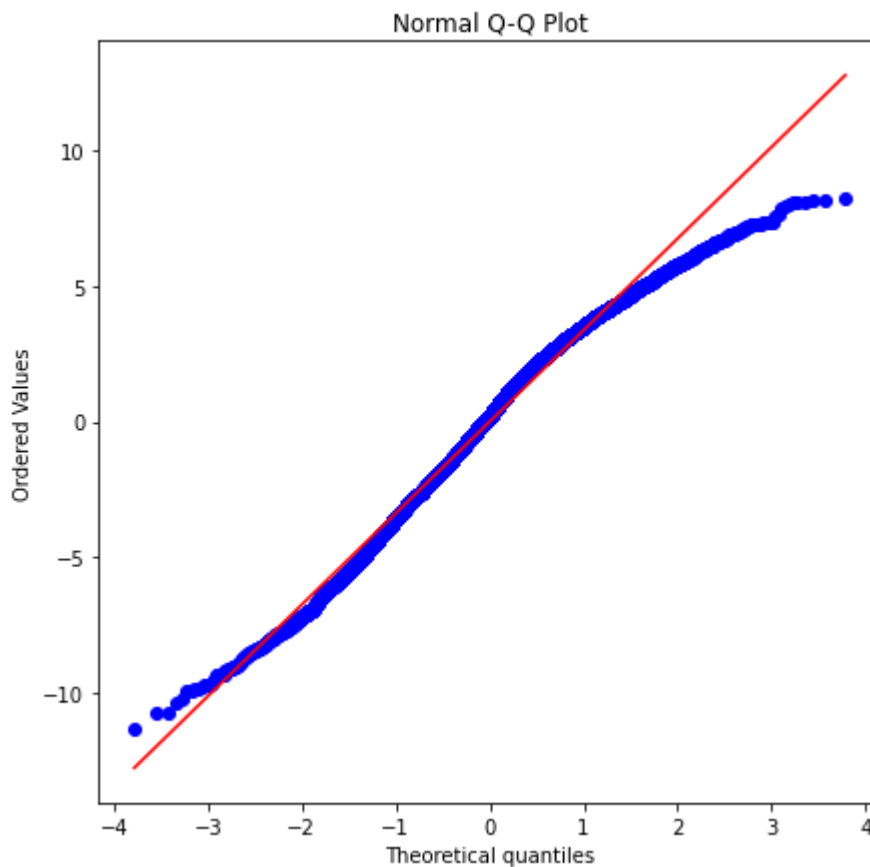
regression model to predict Tenure.

```
In [63]: ##3rd party code. See Residual and Prediction Plots In Python.  
         uhat=lm52.resid  
         yhat = lm52.fittedvalues
```

```
In [64]: #3rd party code, see Luhaniwa.  
         sns.residplot(x=yhat, y=uhat, lowess=True, color='blue', line_kws={'color': 'red', 'lw'  
         plt.xlabel('Fitted Values')  
         plt.ylabel('Residuals')  
         plt.title("Fitted Values Vs Residuals")  
         plt.show()
```



```
In [65]: #3rd party code, see Luhaniwa.  
  
         residuals = data.Tenure - yhat  
         plt.figure(figsize=(7,7))  
         stats.probplot(residuals, dist="norm", plot=plt)  
         plt.title("Normal Q-Q Plot")  
         plt.show()
```



The qq plot above seems to indicate that the assumption of normally distributed residuals is not adequately met. According to the article below, this graph suggest that the residuals have thinner tails than expected.

[https://www.ucd.ie/ecomodel/Resources/QQplots\\_WebVersion.html](https://www.ucd.ie/ecomodel/Resources/QQplots_WebVersion.html)

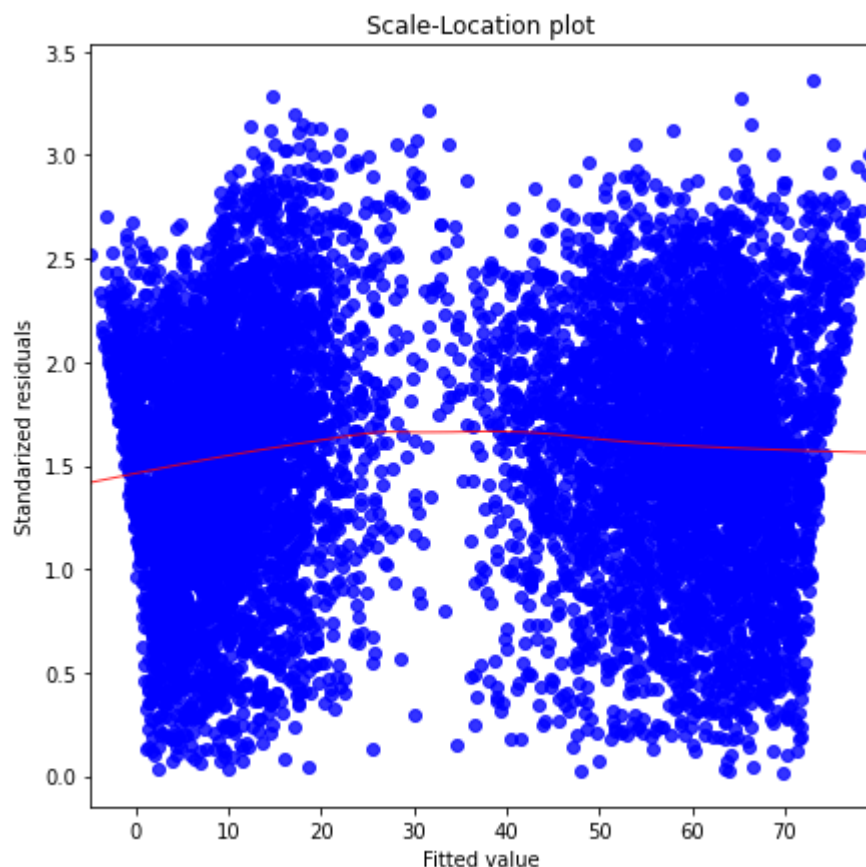
```
In [66]: #3rd party code, see Luhaniwa.

model_norm_residuals_abs_sqrt=np.sqrt(np.abs(residuals))

plt.figure(figsize=(7,7))
sns.regplot(x=yhat, y=model_norm_residuals_abs_sqrt,
            scatter=True,
            lowess=True,color="blue",
            line_kws={'color': 'red', 'lw': 1, 'alpha': 0.8})
plt.ylabel("Standarized residuals")
plt.xlabel("Fitted value")
plt.title("Scale-Location plot")

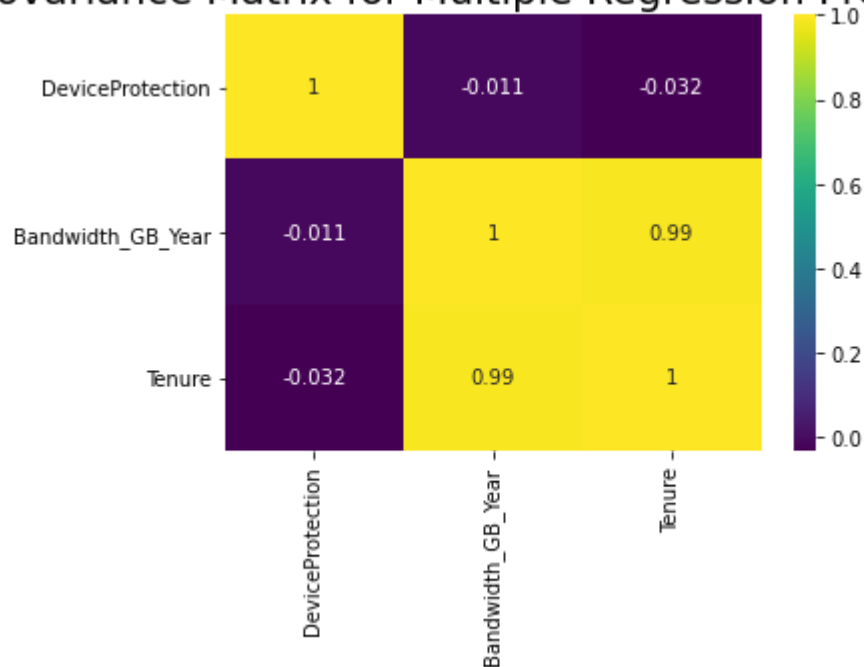
# https://towardsdatascience.com/simulating-replicating-r-regression-plot-in-python-usi
```

```
Out[66]: Text(0.5, 1.0, 'Scale-Location plot')
```



```
In [67]: # Multicollinearity
# 3rd party code, see seaborn heatmap
corr = linearmodeldata.corr()
sns.heatmap(corr, annot=True, cmap='viridis')
plt.rcParams['figure.figsize'] = [15, 10]
plt.title('Covariance Matrix for Multiple Regression Predictors', fontsize=20)
plt.show()
# variables not standardized, therefore covariance shown below.
```

Covariance Matrix for Multiple Regression Predictors



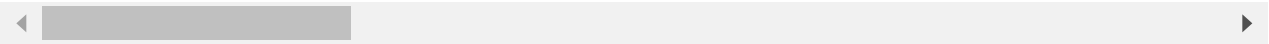
In [68]: data.corr()

Out[68]:

	Unnamed: 0	Lat	Lng	Population	Children	Age
<b>Unnamed: 0</b>	1.000000	-0.018236	0.003201	-0.016163	-0.009421	0.01250
<b>Lat</b>	-0.018236	1.000000	0.006847	-0.210420	-0.000661	-0.00763
<b>Lng</b>	0.003201	0.006847	1.000000	-0.043112	0.015032	0.00809
<b>Population</b>	-0.016163	-0.210420	-0.043112	1.000000	0.002039	0.01755
<b>Children</b>	-0.009421	-0.000661	0.015032	0.002039	1.000000	-0.01927
<b>Age</b>	0.012500	-0.007636	0.008095	0.017553	-0.019278	1.00000
<b>Income</b>	-0.000956	0.012180	0.003657	-0.012791	0.008053	-0.00432
<b>Outage_sec_perweek</b>	-0.004429	-0.001116	0.017198	0.012753	0.015712	-0.01400
<b>Email</b>	-0.011416	-0.027069	0.004544	0.020289	-0.000699	-0.00056
<b>Contacts</b>	-0.002592	-0.008075	0.006280	0.008257	-0.028512	0.01008
<b>Yearly_equip_failure</b>	0.003645	0.005080	-0.019834	-0.011282	0.016259	-0.00200
<b>Techie</b>	-0.017860	-0.003455	0.005277	-0.007577	-0.009704	-0.00299
<b>Port_modem</b>	0.005150	-0.014387	-0.022159	0.006359	0.008484	0.00025
<b>Tablet</b>	0.000633	0.008252	0.014097	-0.004764	-0.001250	-0.00900
<b>Phone</b>	-0.001871	-0.002295	0.000160	0.008429	0.002409	0.01219
<b>Multiple</b>	0.007000	-0.008121	-0.003038	-0.001752	-0.014049	-0.00479
<b>OnlineSecurity</b>	-0.000618	-0.007688	-0.000286	0.006535	0.003126	-0.01329
<b>OnlineBackup</b>	0.024523	-0.004785	0.017435	0.010121	0.007762	0.00001
<b>DeviceProtection</b>	-0.028176	0.014046	0.023552	-0.007475	0.014090	0.01806
<b>TechSupport</b>	0.002749	-0.007794	-0.003566	-0.015930	-0.001104	0.02351
<b>StreamingTV</b>	0.005045	0.011651	-0.011326	-0.010643	-0.002786	0.00583
<b>StreamingMovies</b>	-0.003230	-0.004550	0.014075	0.009357	0.014210	0.01429
<b>PaperlessBilling</b>	-0.003053	0.018495	0.003579	0.003007	0.010913	-0.00205
<b>Tenure</b>	0.832870	-0.018762	0.006475	-0.013353	-0.007834	0.01191
<b>MonthlyCharge</b>	0.003002	-0.003626	0.015715	0.000007	0.002886	0.01570
<b>Bandwidth_GB_Year</b>	0.825691	-0.019639	0.006555	-0.014148	0.020176	-0.01896
<b>Item1</b>	-0.005678	0.006214	0.026688	0.005561	0.002571	-0.00278
<b>Item2</b>	0.003010	0.009850	0.004409	0.004098	0.009366	0.00822
<b>Item3</b>	0.000503	-0.001275	0.019970	-0.002042	-0.003870	-0.00422
<b>Item4</b>	0.000972	0.011344	0.000181	-0.015931	-0.013314	0.01181
<b>Item5</b>	0.013881	0.001842	-0.000287	0.000937	-0.002719	-0.01024
<b>Item6</b>	-0.009588	-0.008492	-0.005162	0.004876	-0.004582	0.01107

	Unnamed: 0	Lat	Lng	Population	Children	Age
Item7	0.004040	0.014398	0.004845	-0.011001	-0.005531	0.01361
Item8	-0.016094	-0.019631	0.001140	0.009268	-0.009745	0.01768
Area_Suburban	-0.012631	-0.015272	0.003023	0.014162	-0.006477	0.00188
Area_Rural	0.021712	0.002301	0.001952	0.003129	0.016887	-0.00725
Marital_Married	0.010307	0.007252	-0.018527	0.001967	-0.001587	-0.00105
Marital_Separated	0.008118	0.005497	-0.007044	-0.012272	0.001123	-0.00342
Marital_NeverMarried	-0.009006	-0.003240	0.000554	0.018075	-0.006011	0.00211
Marital_Divorced	-0.006747	-0.000093	0.000811	-0.009529	0.001491	0.00978
Gender_Female	0.008082	0.010377	-0.014877	0.011124	-0.010395	-0.00239
Gender_Nonbinary	-0.015772	0.001940	-0.004082	0.000879	0.014595	-0.02508
Contract_Oneyear	-0.007709	-0.006041	-0.000912	-0.006646	0.010080	-0.00757
Contract_TwoYear	0.012564	0.005551	0.009793	0.014128	0.018055	-0.00064
InternetService_DSL	0.004336	-0.014990	-0.012420	-0.003764	-0.001239	-0.00746
InternetService_None	0.000294	0.017852	-0.012695	0.002966	0.008403	0.00885
PaymentMethod_BankTransferautomatic	-0.013118	0.007662	0.019057	-0.000902	-0.000279	0.01909
PaymentMethod_MailedCheck	0.008527	-0.015018	-0.009638	-0.002208	0.000136	-0.01754
PaymentMethod_ElectronicCheck	-0.008028	0.003423	0.018777	-0.010974	-0.001088	-0.00347

49 rows × 49 columns



In [ ]:

In [ ]: