Task 3

Sub task – 1:

Think through what key drivers of churn could be for our client

```
proc logistic data=work.inner_join01;
model churn(event='1') =  cons_12m cons_gas_12m cons_last_month forecast_base_bill_ele
                          forecast_base_bill_year forecast_bill_12m forecast_cons forecast_cons_12m
                          forecast_cons_year forecast_discount_energy forecast_price_energy_p1
                          forecast_meter_rent_12m forecast_price_energy_p2 forecast_price_pow_p1
                          margin_gross_pow_ele margin_net_pow_ele nb_prod_act net_margin num_years_antig
                          pow_max / link=probit rsquare;
run;
```

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-1.2531	0.3203	15.3056	<.0001
cons_12m	1	2.508E-8	2.816E-7	0.0079	0.9290
cons_gas_12m	1	-7.48E-7	3.099E-7	5.8216	0.0158
cons_last_month	1	-2.5E-6	2.679E-6	0.8704	0.3508
forecast_base_bill_e	1	-0.00002	0.000293	0.0062	0.9375
forecast_base_bill_y	0	0	.	.	.
forecast_bill_12m	1	0.000025	0.000053	0.2193	0.6396
forecast_cons	1	3.295E-6	0.000456	0.0001	0.9942
forecast_cons_12m	1	-0.00004	0.000062	0.3590	0.5490
forecast_cons_year	1	7.326E-7	0.000034	0.0005	0.9829
forecast_discount_en	1	-0.00548	0.00544	1.0121	0.3144
forecast_price_energ	1	1.1050	1.7367	0.4048	0.5246
forecast_meter_rent_	1	0.00150	0.000626	5.7805	0.0162
forecast_price_energ	1	0.4619	0.8493	0.2958	0.5866
forecast_price_pow_p	1	0.00857	0.00693	1.5289	0.2163
margin_gross_pow_ele	1	0.00457	0.00156	8.5666	0.0034
margin_net_pow_ele	1	0.000365	0.00120	0.0926	0.7609
nb_prod_act	1	0.00624	0.0458	0.0185	0.8917
net_margin	1	0.000118	0.000078	2.2734	0.1316
num_years_antig	1	-0.1190	0.0233	26.0703	<.0001
pow_max	1	-0.00257	0.00316	0.6630	0.4155

Base on Maximum likelihood estimation, we find that num_years_antig has the significant relation with churn variable.

```
proc freq data=work.only_churn nlevels;
tables churn*num_years_antig;
run;
```

Table of churn by num_years_antig										
	num_years_antig									
churn	2	3	4	5	6	7	8	10	11	Total
1	1 0.20 0.20 100.00	161 31.82 31.82 100.00	196 38.74 38.74 100.00	72 14.23 14.23 100.00	71 14.03 14.03 100.00	2 0.40 0.40 100.00	1 0.20 0.20 100.00	1 0.20 0.20 100.00	1 0.20 0.20 100.00	506 100.00
Total	1 0.20	161 31.82	196 38.74	72 14.23	71 14.03	2 0.40	1 0.20	1 0.20	1 0.20	506 100.00

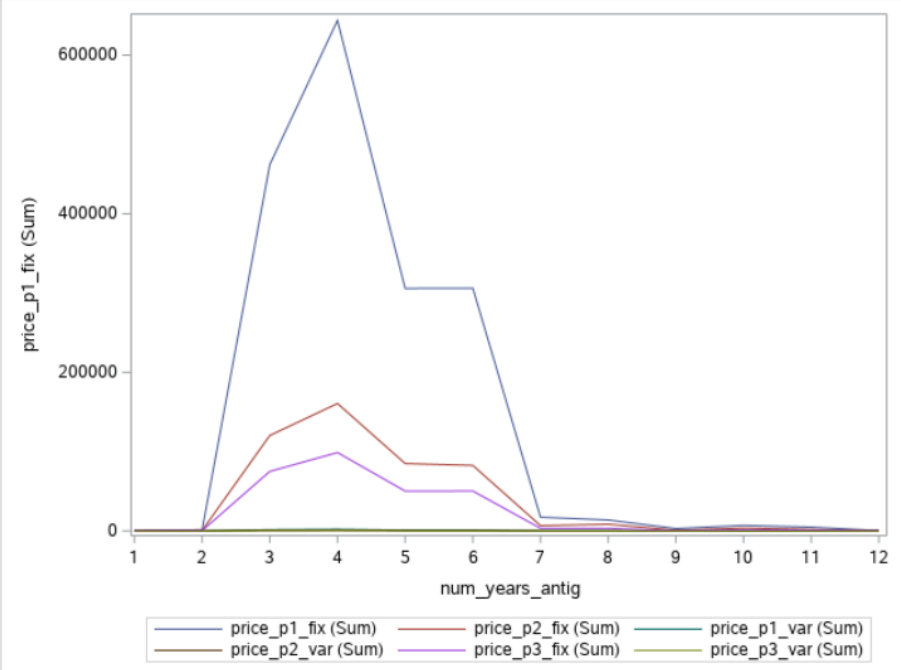
Observing the table, it shows little positive skewed distributions where we have maximum churn rate when antiquity of the client is 3 to 4 years old.

```
proc logistic data=work.price_inner_join;
model num_years_antig = price_p1_fix price_p1_var price_p2_fix
                        price_p2_var price_p3_fix price_p3_var;
run;
```

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	1	-13.6442	0.3622	1418.9355	<.0001
Intercept	2	1	-11.9381	0.1703	4915.9391	<.0001
Intercept	3	1	-6.0884	0.0786	6005.2600	<.0001
Intercept	4	1	-4.5044	0.0779	3342.3974	<.0001
Intercept	5	1	-3.6260	0.0775	2190.4638	<.0001
Intercept	6	1	-1.3594	0.0723	353.9363	<.0001
Intercept	7	1	-0.8264	0.0700	139.3874	<.0001
Intercept	8	1	-0.1907	0.0682	7.8237	0.0052
Intercept	9	1	1.3482	0.0809	277.7379	<.0001
Intercept	10	1	2.1305	0.1043	417.5453	<.0001
Intercept	11	1	4.4339	0.2927	229.5158	<.0001
price_p1_fix		1	0.0771	0.00341	511.0723	<.0001
price_p1_var		1	10.9316	0.8404	169.2051	<.0001
price_p2_fix		1	-0.1144	0.00660	300.5977	<.0001
price_p2_var		1	-1.9031	0.4409	18.6333	<.0001
price_p3_fix		1	0.0650	0.00862	56.7506	<.0001
price_p3_var		1	33.3734	3.2206	107.3838	<.0001

Now this table shows the significant relations between the prices of power and energy of different period with regards to antiquity of the client (in number of years).

```
proc sgplot data = work.price_inner_join;
     vline num_years_antig/response = price_p1_fix;
     vline num_years_antig/response = price_p2_fix;
     vline num_years_antig/response = price_p1_var;
     vline num_years_antig/response = price_p2_var;
     vline num_years_antig/response = price_p3_fix;
     vline num_years_antig/response = price_p3_var;
run;
```



Above graph shows prices of the power and energy is highest in the 3 to 5 num_years_antig where maximum number of churn rate occur.

Sub-task 2

The following input variables or features must include when building a model are:

1. Price_p1_fix
2. Price_p2_fix
3. Price_p3_fix
4. Num_year_antig
5. margin_gross_pow_ele
6. forecast_meter_rent_12m
7. forecast_meter_rent_12m

these variable show significant probability when comparing with churn rate. First table.