In [1]:
```
import pandas as pd
```

In [2]:
```
vaccine_data=pd.read_excel("h1n1_vaccine_prediction.xlsx") #here h1n1_vaccine is dependent variable
vaccine_data
```

Out[2]:

	unique_id	h1n1_worry	h1n1_awareness	antiviral_medication	contact_avoidance	bought_face_mask	wash_hands_frequently	avoid_large_gath
0	0	1.0	0.0	0.0	0.0	0.0	0.0	
1	1	3.0	2.0	0.0	1.0	0.0	1.0	
2	2	1.0	1.0	0.0	1.0	0.0	0.0	
3	3	1.0	1.0	0.0	1.0	0.0	1.0	
4	4	2.0	1.0	0.0	1.0	0.0	1.0	
...	...	...	...	...	...	...	...	
26702	26702	2.0	0.0	0.0	1.0	0.0	0.0	
26703	26703	1.0	2.0	0.0	1.0	0.0	1.0	
26704	26704	2.0	2.0	0.0	1.0	1.0	1.0	
26705	26705	1.0	1.0	0.0	0.0	0.0	0.0	
26706	26706	0.0	0.0	0.0	1.0	0.0	0.0	

26707 rows × 34 columns

Checking null values and datatypes

In [3]:
```
vaccine_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 26707 entries, 0 to 26706
Data columns (total 34 columns):
 #   Column                       Non-Null Count  Dtype  
---  ------                       --------------  -----  
 0   unique_id                    26707 non-null  int64  
 1   h1n1_worry                   26615 non-null  float64
 2   h1n1_awareness               26591 non-null  float64
 3   antiviral_medication         26636 non-null  float64
 4   contact_avoidance            26499 non-null  float64
 5   bought_face_mask             26688 non-null  float64
 6   wash_hands_frequently        26665 non-null  float64
 7   avoid_large_gatherings       26620 non-null  float64
 8   reduced_outside_home_cont    26625 non-null  float64
 9   avoid_touch_face             26579 non-null  float64
 10  dr_recc_h1n1_vacc            24547 non-null  float64
 11  dr_recc_seasonal_vacc        24547 non-null  float64
 12  chronic_medic_condition      25736 non-null  float64
 13  cont_child_undr_6_mnths      25887 non-null  float64
 14  is_health_worker             25903 non-null  float64
 15  has_health_insur             14433 non-null  float64
 16  is_h1n1_vacc_effective       26316 non-null  float64
 17  is_h1n1_risky                26319 non-null  float64
 18  sick_from_h1n1_vacc          26312 non-null  float64
 19  is_seas_vacc_effective       26245 non-null  float64
 20  is_seas_risky                26193 non-null  float64
 21  sick_from_seas_vacc          26170 non-null  float64
 22  age_bracket                  26707 non-null  object 
 23  qualification                25300 non-null  object 
 24  race                         26707 non-null  object 
 25  sex                          26707 non-null  object 
 26  income_level                 22284 non-null  object 
 27  marital_status               25299 non-null  object 
 28  housing_status               24665 non-null  object 
 29  employment                   25244 non-null  object 
 30  census_msa                   26707 non-null  object 
 31  no_of_adults                 26458 non-null  float64
 32  no_of_children               26458 non-null  float64
 33  h1n1_vaccine                 26707 non-null  int64  
```

```
dtypes: float64(23), int64(2), object(9)
memory usage: 6.9+ MB
```

In [4]:
```
vaccine_data.isnull().sum()
```

Out[4]:
```
unique_id                     0
h1n1_worry                   92
h1n1_awareness              116
antiviral_medication         71
contact_avoidance           208
bought_face_mask             19
wash_hands_frequently        42
avoid_large_gatherings       87
reduced_outside_home_cont    82
avoid_touch_face            128
dr_recc_h1n1_vacc          2160
dr_recc_seasonal_vacc      2160
chronic_medic_condition     971
cont_child_undr_6_mnths    820
is_health_worker           804
has_health_insur           12274
is_h1n1_vacc_effective     391
is_h1n1_risky              388
sick_from_h1n1_vacc        395
is_seas_vacc_effective     462
is_seas_risky              514
sick_from_seas_vacc        537
age_bracket                  0
qualification              1407
race                         0
sex                          0
income_level               4423
marital_status             1408
housing_status             2042
employment                 1463
census_msa                   0
no_of_adults               249
no_of_children             249
h1n1_vaccine                 0
dtype: int64
```

By checking null values data has maximum of 12274 null values and some object datatype so further process deals with cleaning the data

CLEANING DATA

In [5]:
```
#removing has_health_insur because of high null values.
vac_data=vaccine_data.drop(["has_health_insur"],axis=1)
vac_data
```

Out[5]:

	unique_id	h1n1_worry	h1n1_awareness	antiviral_medication	contact_avoidance	bought_face_mask	wash_hands_frequently	avoid_large_gath
0	0	1.0	0.0	0.0	0.0	0.0	0.0	
1	1	3.0	2.0	0.0	1.0	0.0	1.0	
2	2	1.0	1.0	0.0	1.0	0.0	0.0	
3	3	1.0	1.0	0.0	1.0	0.0	1.0	
4	4	2.0	1.0	0.0	1.0	0.0	1.0	
...	...	...	...	...	...	...	...	
26702	26702	2.0	0.0	0.0	1.0	0.0	0.0	
26703	26703	1.0	2.0	0.0	1.0	0.0	1.0	
26704	26704	2.0	2.0	0.0	1.0	1.0	1.0	
26705	26705	1.0	1.0	0.0	0.0	0.0	0.0	
26706	26706	0.0	0.0	0.0	1.0	0.0	0.0	

26707 rows × 33 columns

In [6]:
```
vac_data.describe()
```

Out[6]:

	unique_id	h1n1_worry	h1n1_awareness	antiviral_medication	contact_avoidance	bought_face_mask	wash_hands_frequently	avoid_large
count	26707.000000	26615.000000	26591.000000	26636.000000	26499.000000	26688.000000	26665.000000	
mean	13353.000000	1.618486	1.262532	0.048844	0.725612	0.068982	0.825614	
std	7709.791156	0.910311	0.618149	0.215545	0.446214	0.253429	0.379448	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	6676.500000	1.000000	1.000000	0.000000	0.000000	0.000000	1.000000	
50%	13353.000000	2.000000	1.000000	0.000000	1.000000	0.000000	1.000000	
75%	20029.500000	2.000000	2.000000	0.000000	1.000000	0.000000	1.000000	
max	26706.000000	3.000000	2.000000	1.000000	1.000000	1.000000	1.000000	

8 rows × 24 columns

In [7]:
```
#finding total null values
vac_data.isnull().sum()
```

Out[7]:
```
unique_id                     0
h1n1_worry                   92
h1n1_awareness              116
antiviral_medication         71
contact_avoidance           208
bought_face_mask             19
wash_hands_frequently        42
avoid_large_gatherings       87
reduced_outside_home_cont    82
avoid_touch_face            128
dr_recc_h1n1_vacc          2160
dr_recc_seasonal_vacc      2160
chronic_medic_condition     971
cont_child_undr_6_mnths    820
is_health_worker           804
is_h1n1_vacc_effective     391
is_h1n1_risky              388
sick_from_h1n1_vacc        395
is_seas_vacc_effective     462
is_seas_risky              514
sick_from_seas_vacc        537
age_bracket                  0
qualification             1407
race                         0
sex                          0
income_level              4423
marital_status            1408
housing_status            2042
employment                1463
census_msa                   0
no_of_adults               249
no_of_children             249
h1n1_vaccine                 0
dtype: int64
```

In [8]:
```
#storing all the columns in a variable which needs to be set as 0
col_to_zero=["employment","marital_status","housing_status","income_level","qualification","sick_from_seas_vacc","is_sea
```

In [9]:
```
#filling 0 to the columns taken
vac_data[col_to_zero]=vac_data[col_to_zero].fillna(0)
```

In [10]:
```
vac_data.isnull().sum()
```

Out[10]:
```
unique_id                    0
h1n1_worry                   0
h1n1_awareness               0
antiviral_medication         0
contact_avoidance            0
bought_face_mask             0
wash_hands_frequently        0
avoid_large_gatherings       0
reduced_outside_home_cont    0
avoid_touch_face             0
dr_recc_h1n1_vacc            0
dr_recc_seasonal_vacc        0
chronic_medic_condition      0
cont_child_undr_6_mnths      0
is_health_worker             0
is_h1n1_vacc_effective       0
is_h1n1_risky                0
sick_from_h1n1_vacc          0
is_seas_vacc_effective       0
is_seas_risky                0
sick_from_seas_vacc          0
age_bracket                  0
qualification                0
race                         0
sex                          0
income_level                 0
marital_status               0
housing_status               0
employment                   0
census_msa                   0
no_of_adults               249
no_of_children             249
h1n1_vaccine                 0
dtype: int64
```

In [11]:
```
#filling mean value to one of column
vac_data["no_of_adults"]=vac_data["no_of_adults"].fillna(vac_data.no_of_adults.mean())
```

In [12]:
```
#filling median value to one of column
vac_data["no_of_children"]=vac_data["no_of_children"].fillna(vac_data.no_of_adults.median())
```

In [13]:
```
vac_data.isnull().sum()
```

Out[13]:
```
unique_id                    0
h1n1_worry                   0
h1n1_awareness               0
antiviral_medication         0
contact_avoidance            0
bought_face_mask             0
wash_hands_frequently        0
avoid_large_gatherings       0
reduced_outside_home_cont    0
avoid_touch_face             0
dr_recc_h1n1_vacc            0
dr_recc_seasonal_vacc        0
chronic_medic_condition      0
cont_child_undr_6_mnths      0
is_health_worker             0
is_h1n1_vacc_effective       0
is_h1n1_risky                0
sick_from_h1n1_vacc          0
is_seas_vacc_effective       0
is_seas_risky                0
sick_from_seas_vacc          0
age_bracket                  0
qualification                0
race                         0
sex                          0
income_level                 0
marital_status               0
housing_status               0
employment                   0
census_msa                   0
no_of_adults                 0
no_of_children               0
h1n1_vaccine                 0
dtype: int64
```

creating dummy variables

In [14]:
```
vac_data=pd.get_dummies(vac_data,drop_first=True)
```

In [15]:
```
vac_data
```

Out[15]:

	unique_id	h1n1_worry	h1n1_awareness	antiviral_medication	contact_avoidance	bought_face_mask	wash_hands_frequently	avoid_large_gath
0	0	1.0	0.0	0.0	0.0	0.0	0.0	
1	1	3.0	2.0	0.0	1.0	0.0	1.0	
2	2	1.0	1.0	0.0	1.0	0.0	0.0	
3	3	1.0	1.0	0.0	1.0	0.0	1.0	
4	4	2.0	1.0	0.0	1.0	0.0	1.0	
...	...	...	...	...	...	...	...	
26702	26702	2.0	0.0	0.0	1.0	0.0	0.0	
26703	26703	1.0	2.0	0.0	1.0	0.0	1.0	
26704	26704	2.0	2.0	0.0	1.0	1.0	1.0	
26705	26705	1.0	1.0	0.0	0.0	0.0	0.0	
26706	26706	0.0	0.0	0.0	1.0	0.0	0.0	

26707 rows × 48 columns

Now data is fully cleaned,all the null values are fixed and column with object datatypes are fixed with dummy variables so now data is set to proceed.

FINDING CORRELATION

In [16]:
```
vac_corr=vac_data.corr()
vac_corr
```

Out[16]:

	unique_id	h1n1_worry	h1n1_awareness	antiviral_medication	contact_avoidance	bought_face_mask	wash_hands_fre
unique_id	1.000000	0.017298	0.003655	-0.008458	0.011275	-0.006654	0
h1n1_worry	0.017298	1.000000	0.070812	0.089210	0.235290	0.153132	0
h1n1_awareness	0.003655	0.070812	1.000000	-0.009899	0.092918	0.029156	0
antiviral_medication	-0.008458	0.089210	-0.009899	1.000000	0.050099	0.145772	0
contact_avoidance	0.011275	0.235290	0.092918	0.050099	1.000000	0.065684	0
bought_face_mask	-0.006654	0.153132	0.029156	0.145772	0.065684	1.000000	0
wash_hands_frequently	0.010508	0.293217	0.092972	0.064248	0.334402	0.083282	1
avoid_large_gatherings	0.004455	0.253331	-0.045952	0.106314	0.226986	0.180377	0
reduced_outside_home_cont	0.009181	0.243795	-0.065273	0.127204	0.220113	0.162964	0
avoid_touch_face	0.007392	0.249087	0.090599	0.071284	0.335162	0.104725	0
dr_recc_h1n1_vacc	-0.002178	0.137611	0.094601	0.050882	0.066704	0.079887	0
dr_recc_seasonal_vacc	0.001131	0.121282	0.074284	0.031682	0.073419	0.065301	0
chronic_medic_condition	0.004753	0.090880	-0.011329	0.004637	0.040149	0.062200	0
cont_child_undr_6_mnths	-0.004993	0.049170	0.026034	0.026183	0.002066	0.036828	0
is_health_worker	-0.003372	0.033881	0.171131	0.007145	0.004244	0.065645	0
is_h1n1_vacc_effective	0.007974	0.222801	0.134903	0.019210	0.113240	0.023983	0
is_h1n1_risky	0.000122	0.362047	0.085184	0.095664	0.118234	0.120132	0
sick_from_h1n1_vacc	-0.003024	0.345678	-0.004136	0.070851	0.131410	0.097595	0
is_seas_vacc_effective	0.004801	0.212147	0.111406	0.002874	0.118067	0.024840	0
is_seas_risky	-0.006219	0.317386	0.094386	0.073957	0.132699	0.096935	0
sick_from_seas_vacc	0.008210	0.215668	-0.040411	0.074298	0.085443	0.079439	0
no_of_adults	0.000186	-0.013081	0.024104	0.044574	0.020671	0.013927	0
no_of_children	-0.003753	0.052435	0.050367	0.084977	0.043151	0.006687	0

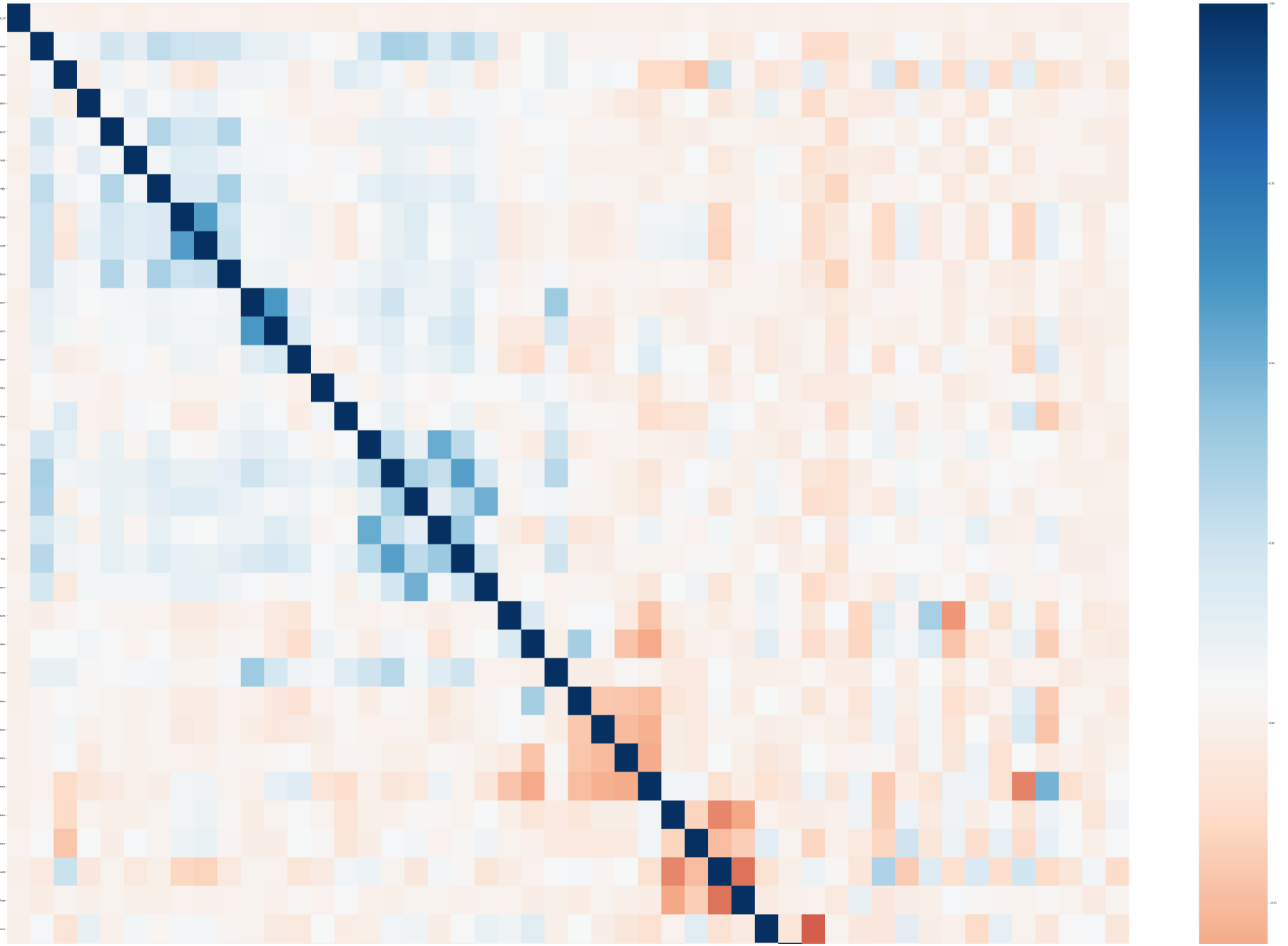
	unique_id	h1n1_worry	h1n1_awareness	antiviral_medication	contact_avoidance	bought_face_mask	wash_hands_fre
h1n1_vaccine	-0.003280	0.121664	0.117153	0.040226	0.048712	0.070413	0
age_bracket_35 - 44 Years	-0.003987	0.023401	0.050005	0.041878	0.013675	-0.000107	0
age_bracket_45 - 54 Years	0.003766	0.026791	0.076757	-0.000071	0.023699	0.004812	0
age_bracket_55 - 64 Years	0.003790	0.025974	0.061045	-0.037690	0.017630	0.004922	0
age_bracket_65+ Years	0.005368	0.010500	-0.128030	-0.060702	-0.030505	-0.001256	-0
qualification_12 Years	-0.002043	0.031211	-0.135159	0.013757	-0.007609	-0.004878	0
qualification_< 12 Years	0.012543	0.049020	-0.211514	0.054443	-0.010062	0.063551	0
qualification_College Graduate	-0.008194	-0.035290	0.261743	-0.052322	0.017636	-0.039325	-0
qualification_Some College	0.003395	-0.020343	0.021655	-0.007141	0.016145	-0.005869	-0
race_Hispanic	-0.004107	0.062742	-0.070202	0.123905	0.003860	0.074575	0
race_Other or Multiple	0.004453	0.027876	-0.032933	0.019335	-0.007555	0.042127	0
race_White	-0.002661	-0.121247	0.141172	-0.117460	0.003713	-0.095087	-0
sex_Male	0.005169	-0.126123	-0.068567	-0.006196	-0.114246	-0.049912	-0
income_level_<= $75,000, Above Poverty	0.001822	-0.014288	0.009959	-0.033920	0.015201	-0.020048	-0
income_level_> $75,000	-0.007414	-0.023029	0.193320	-0.028630	0.037895	-0.035075	0
income_level_Below Poverty	0.000503	0.069364	-0.160192	0.088680	-0.007399	0.068225	0
marital_status_Married	0.004356	0.030822	0.145337	-0.018897	0.059201	-0.015192	0
marital_status_Not Married	-0.003021	-0.029312	-0.108458	0.005924	-0.044726	0.001622	-0
housing_status_Own	0.006695	0.000314	0.152248	-0.067929	0.044978	-0.055728	0
housing_status_Rent	-0.007372	-0.000389	-0.107568	0.056575	-0.028823	0.043492	-0
employment_Employed	0.001738	-0.050091	0.160707	-0.003326	-0.000112	-0.040719	-0
employment_Not in Labor Force	0.002272	0.036781	-0.100723	-0.020534	0.010477	0.026372	0
employment_Unemployed	-0.009269	0.030796	-0.049468	0.023174	0.015771	0.007715	-0

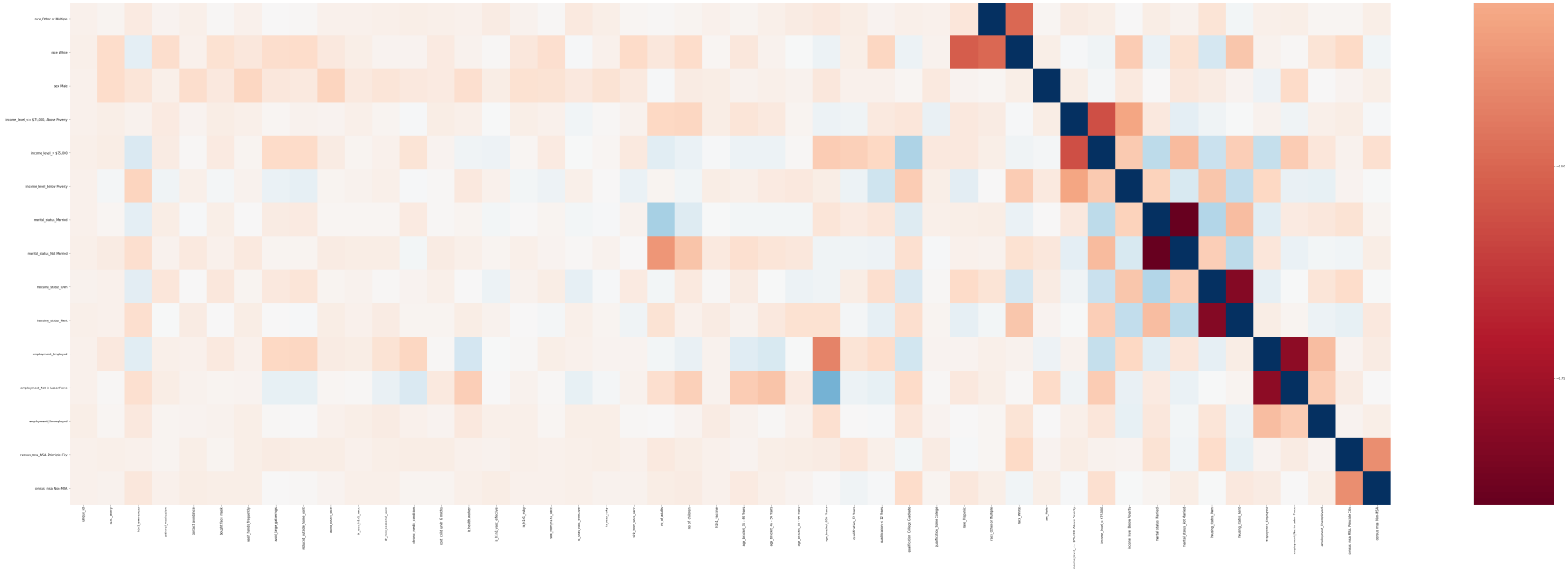
	unique_id	h1n1_worry	h1n1_awareness	antiviral_medication	contact_avoidance	bought_face_mask	wash_hands_fre
census_msa_MSA, Principle City	0.002355	-0.005119	0.002187	0.020956	-0.016241	0.021048	-0
census_msa_Non-MSA	0.002916	0.005844	-0.056981	0.000532	-0.020726	-0.014384	-0

48 rows × 48 columns

In [17]:
```
import seaborn as sns
import matplotlib.pyplot as plt
```

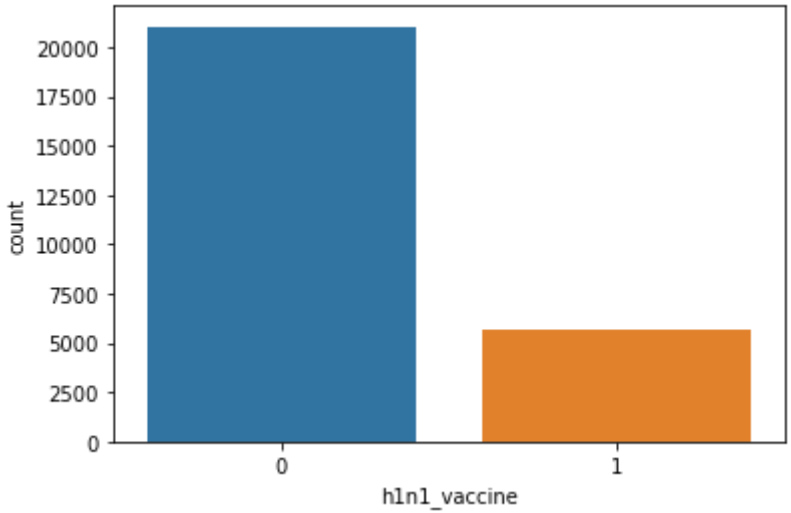
In [18]:
```
plt.subplots(figsize=(100,100))
sns.heatmap(vac_corr,cmap="RdBu");
```





most of them are negatively correlated only dr_recc_h1n1_vacc(doctor recommended) is highly correlated

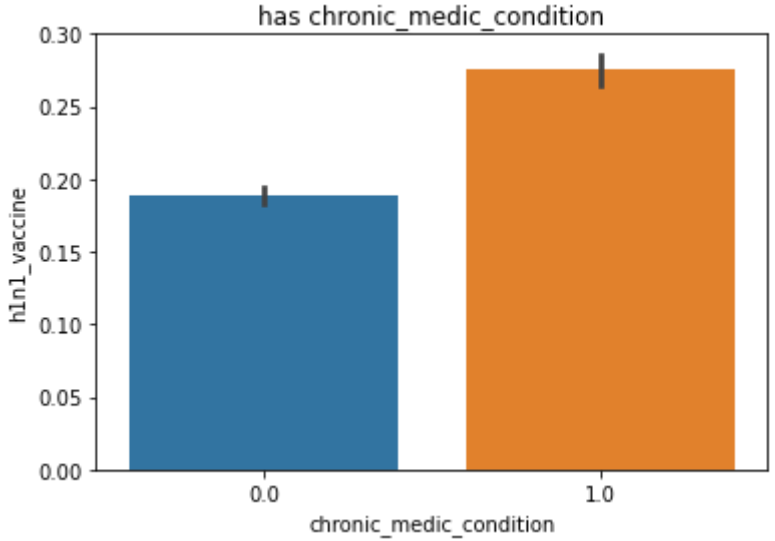
In [19]:
```
sns.countplot(x="h1n1_vaccine",data=vac_data);
```



data is highly unequal

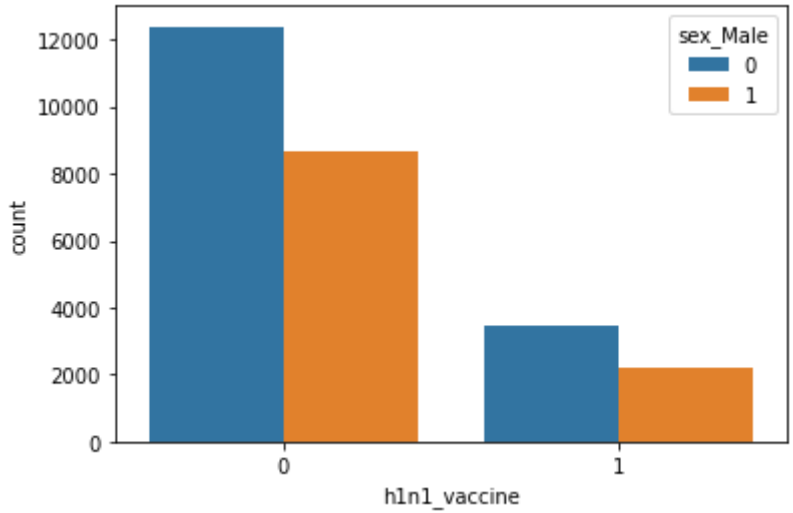
person who recived vaccine is more less than didn't recived

In [20]:
```
sns.barplot(x="chronic_medic_condition",y="h1n1_vaccine",data=vac_data);
plt.title("has chronic_medic_condition");
```



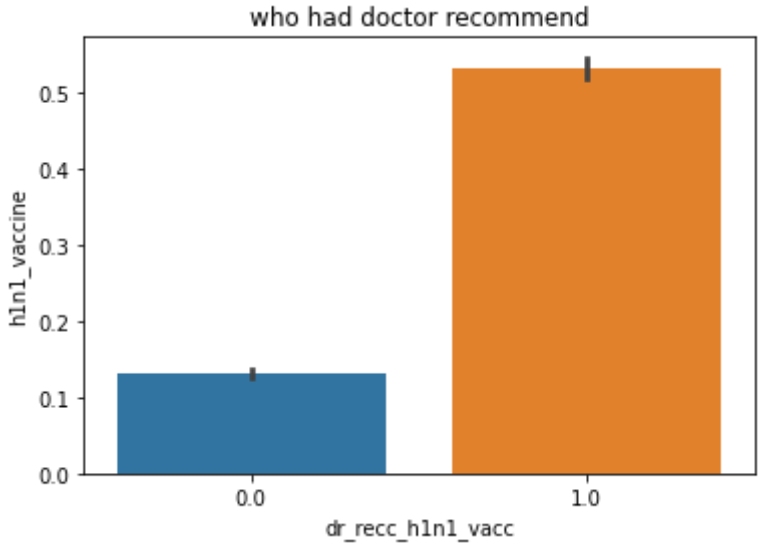
person has chronic_medic_condition recived vaccine higher than who didn't

In [21]:
```
sns.countplot(x="h1n1_vaccine",data=vac_data,hue="sex_Male");
```



male is highly vaccined and also highly didn't got vaccined

In [22]:
```
sns.barplot(x="dr_recc_h1n1_vacc",y="h1n1_vaccine",data=vac_data);
plt.title("who had doctor recommend");
```



after doctor recommendation, many took the vaccine

Seperating independent and dependent variables

In [23]:
```
y_dep=vac_data.h1n1_vaccine
x_indep=vac_data.drop("h1n1_vaccine",axis=1)
```

In [24]:
```
x_indep
```

Out[24]:

	unique_id	h1n1_worry	h1n1_awareness	antiviral_medication	contact_avoidance	bought_face_mask	wash_hands_frequently	avoid_large_gath
0	0	1.0	0.0	0.0	0.0	0.0	0.0	
1	1	3.0	2.0	0.0	1.0	0.0	1.0	
2	2	1.0	1.0	0.0	1.0	0.0	0.0	
3	3	1.0	1.0	0.0	1.0	0.0	1.0	
4	4	2.0	1.0	0.0	1.0	0.0	1.0	
...	...	...	...	...	...	...	...	
26702	26702	2.0	0.0	0.0	1.0	0.0	0.0	
26703	26703	1.0	2.0	0.0	1.0	0.0	1.0	
26704	26704	2.0	2.0	0.0	1.0	1.0	1.0	
26705	26705	1.0	1.0	0.0	0.0	0.0	0.0	
26706	26706	0.0	0.0	0.0	1.0	0.0	0.0	

26707 rows × 47 columns

MACHINE LEARNING

In [25]:
```
import sklearn
from sklearn import model_selection
from sklearn.model_selection import train_test_split
```

In [71]:
```
x_train, x_test, y_train, y_test = train_test_split(x_indep, y_dep, test_size=0.3, random_state=1)
```

In [27]:
```
x_train.loc[:,"qualification_12 Years":]
```

Out[27]:

	qualification_12 Years	qualification_< 12 Years	qualification_College Graduate	qualification_Some College	race_Hispanic	race_Other or Multiple	race_White	sex_Male	income_level_<= $75,000, Above Poverty
26181	0	1	0	0	0	0	0	1	0
24965	0	0	0	1	0	0	1	0	0
17856	0	0	1	0	0	0	1	1	0
10353	0	0	1	0	0	0	1	0	1
21029	1	0	0	0	0	0	1	0	0
...	...	...	...	...	...	...	...	...	...
10955	0	0	0	1	0	0	1	0	1
17289	0	0	1	0	0	0	1	0	1
5192	0	0	1	0	0	0	1	0	1
12172	0	0	0	1	0	0	1	0	1
235	0	0	1	0	0	0	1	1	0

18694 rows × 20 columns

checking p and AIC value

In [28]:
```
import statsmodels.api as sm
model_st=sm.Logit(y_train,x_train).fit()
```

```
Optimization terminated successfully.
         Current function value: 0.416462
         Iterations 6
```

In [29]:
```
model_st.summary2()
```

Out[29]:

Model:	Logit	Pseudo R-squared:	0.197
Dependent Variable:	h1n1_vaccine	AIC:	15664.6928
Date:	2021-09-01 15:56	BIC:	16032.9828
No. Observations:	18694	Log-Likelihood:	-7785.3
Df Model:	46	LL-Null:	-9693.7
Df Residuals:	18647	LLR p-value:	0.0000
Converged:	1.0000	Scale:	1.0000
No. Iterations:	6.0000		

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
unique_id	-0.0000	0.0000	-6.5073	0.0000	-0.0000	-0.0000
h1n1_worry	-0.0952	0.0268	-3.5498	0.0004	-0.1478	-0.0427
h1n1_awareness	0.0031	0.0359	0.0862	0.9313	-0.0673	0.0735
antiviral_medication	0.1567	0.0935	1.6766	0.0936	-0.0265	0.3400
contact_avoidance	-0.1308	0.0504	-2.5967	0.0094	-0.2295	-0.0321
bought_face_mask	0.2142	0.0782	2.7410	0.0061	0.0610	0.3674
wash_hands_frequently	-0.2590	0.0608	-4.2576	0.0000	-0.3782	-0.1398
avoid_large_gatherings	-0.1472	0.0536	-2.7483	0.0060	-0.2523	-0.0422
reduced_outside_home_cont	-0.0480	0.0545	-0.8808	0.3784	-0.1547	0.0588
avoid_touch_face	-0.0048	0.0497	-0.0972	0.9226	-0.1023	0.0926
dr_recc_h1n1_vacc	2.0263	0.0611	33.1763	0.0000	1.9066	2.1460
dr_recc_seasonal_vacc	-0.4534	0.0599	-7.5675	0.0000	-0.5708	-0.3359
chronic_medic_condition	0.1455	0.0470	3.0935	0.0020	0.0533	0.2377
cont_child_undr_6_mnths	0.1505	0.0723	2.0806	0.0375	0.0087	0.2922
is_health_worker	0.7704	0.0618	12.4768	0.0000	0.6494	0.8915
is_h1n1_vacc_effective	0.2666	0.0234	11.4104	0.0000	0.2208	0.3124

is_h1n1_risky	0.3262	0.0196	16.6670	0.0000	0.2878	0.3645
sick_from_h1n1_vacc	-0.0530	0.0182	-2.9133	0.0036	-0.0887	-0.0174
is_seas_vacc_effective	-0.0347	0.0219	-1.5825	0.1135	-0.0777	0.0083
is_seas_risky	0.1816	0.0188	9.6710	0.0000	0.1448	0.2184
sick_from_seas_vacc	-0.1415	0.0180	-7.8513	0.0000	-0.1768	-0.1061
no_of_adults	-0.2436	0.0318	-7.6607	0.0000	-0.3060	-0.1813
no_of_children	-0.2130	0.0276	-7.7272	0.0000	-0.2670	-0.1589
age_bracket_35 - 44 Years	-0.3957	0.0731	-5.4151	0.0000	-0.5389	-0.2525
age_bracket_45 - 54 Years	-0.4977	0.0679	-7.3312	0.0000	-0.6307	-0.3646
age_bracket_55 - 64 Years	-0.2611	0.0686	-3.8053	0.0001	-0.3956	-0.1266
age_bracket_65+ Years	-0.1925	0.0726	-2.6502	0.0080	-0.3349	-0.0501
qualification_12 Years	-0.6462	0.2006	-3.2206	0.0013	-1.0394	-0.2529
qualification_< 12 Years	-0.8581	0.2093	-4.1002	0.0000	-1.2682	-0.4479
qualification_College Graduate	-0.4774	0.1996	-2.3910	0.0168	-0.8687	-0.0861
qualification_Some College	-0.6310	0.2001	-3.1527	0.0016	-1.0232	-0.2387
race_Hispanic	-0.5250	0.1034	-5.0756	0.0000	-0.7278	-0.3223
race_Other or Multiple	-0.5509	0.1028	-5.3579	0.0000	-0.7524	-0.3493
race_White	-0.7530	0.0679	-11.0852	0.0000	-0.8861	-0.6198
sex_Male	-0.1472	0.0427	-3.4496	0.0006	-0.2308	-0.0636
income_level_<= $75,000, Above Poverty	-0.0052	0.0738	-0.0707	0.9436	-0.1499	0.1395
income_level_> $75,000	0.2179	0.0831	2.6215	0.0088	0.0550	0.3809
income_level_Below Poverty	0.1411	0.0978	1.4421	0.1493	-0.0507	0.3328
marital_status_Married	-0.1133	0.2179	-0.5199	0.6031	-0.5404	0.3138
marital_status_Not Married	-0.5237	0.2173	-2.4098	0.0160	-0.9497	-0.0978
housing_status_Own	-0.2271	0.1311	-1.7324	0.0832	-0.4841	0.0298
housing_status_Rent	-0.4312	0.1373	-3.1402	0.0017	-0.7003	-0.1621
employment_Employed	-0.6238	0.2143	-2.9109	0.0036	-1.0437	-0.2038

employment_Not in Labor Force	-0.5309	0.2153	-2.4663	0.0137	-0.9529	-0.1090
employment_Unemployed	-0.6490	0.2337	-2.7774	0.0055	-1.1070	-0.1910
census_msa_MSA, Principle City	-0.1339	0.0482	-2.7811	0.0054	-0.2283	-0.0395
census_msa_Non-MSA	-0.0438	0.0503	-0.8712	0.3836	-0.1423	0.0547

In [30]:
```
#variables have high p values=h1n1_awareness,antiviral_medication,avoid_touch_face,reduced_outside_home_cont,is_seas_vac
#income_level_<=$75,000,Above Poverty, income_level_Below Poverty,census_msa_Non-MSA
```

In [31]:
```
#removing h1n1_awareness because of high p-value,expecting AIC value will reduce
#LAST AIC VALUE =15664
```

In [72]:
```
x_train=x_train.drop("h1n1_awareness",axis=1)
```

In [55]:
```
model_st=sm.Logit(y_train,x_train).fit()
```

```
Optimization terminated successfully.
         Current function value: 0.416463
         Iterations 6
```

In [34]:
```
model_st.summary2()
```

Out[34]:

Model:	Logit	Pseudo R-squared:	0.197
Dependent Variable:	h1n1_vaccine	AIC:	15662.7003
Date:	2021-09-01 15:56	BIC:	16023.1543
No. Observations:	18694	Log-Likelihood:	-7785.4
Df Model:	45	LL-Null:	-9693.7
Df Residuals:	18648	LLR p-value:	0.0000
Converged:	1.0000	Scale:	1.0000
No. Iterations:	6.0000		

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
unique_id	-0.0000	0.0000	-6.5088	0.0000	-0.0000	-0.0000
h1n1_worry	-0.0951	0.0268	-3.5512	0.0004	-0.1476	-0.0426
antiviral_medication	0.1567	0.0935	1.6767	0.0936	-0.0265	0.3400
contact_avoidance	-0.1307	0.0503	-2.5953	0.0095	-0.2293	-0.0320
bought_face_mask	0.2145	0.0781	2.7457	0.0060	0.0614	0.3676
wash_hands_frequently	-0.2587	0.0607	-4.2595	0.0000	-0.3777	-0.1397
avoid_large_gatherings	-0.1473	0.0536	-2.7488	0.0060	-0.2523	-0.0423
reduced_outside_home_cont	-0.0481	0.0544	-0.8828	0.3773	-0.1548	0.0586
avoid_touch_face	-0.0046	0.0496	-0.0917	0.9269	-0.1018	0.0927
dr_recc_h1n1_vacc	2.0264	0.0611	33.1832	0.0000	1.9067	2.1461
dr_recc_seasonal_vacc	-0.4532	0.0599	-7.5679	0.0000	-0.5706	-0.3358
chronic_medic_condition	0.1456	0.0470	3.0945	0.0020	0.0534	0.2378
cont_child_undr_6_mnths	0.1505	0.0723	2.0822	0.0373	0.0088	0.2923
is_health_worker	0.7711	0.0613	12.5707	0.0000	0.6508	0.8913
is_h1n1_vacc_effective	0.2668	0.0233	11.4460	0.0000	0.2211	0.3124
is_h1n1_risky	0.3262	0.0196	16.6671	0.0000	0.2878	0.3645

sick_from_h1n1_vacc	-0.0530	0.0182	-2.9135	0.0036	-0.0887	-0.0174
is_seas_vacc_effective	-0.0346	0.0219	-1.5802	0.1141	-0.0775	0.0083
is_seas_risky	0.1816	0.0188	9.6717	0.0000	0.1448	0.2184
sick_from_seas_vacc	-0.1415	0.0180	-7.8600	0.0000	-0.1768	-0.1062
no_of_adults	-0.2436	0.0318	-7.6620	0.0000	-0.3059	-0.1813
no_of_children	-0.2128	0.0275	-7.7342	0.0000	-0.2668	-0.1589
age_bracket_35 - 44 Years	-0.3955	0.0730	-5.4149	0.0000	-0.5387	-0.2523
age_bracket_45 - 54 Years	-0.4973	0.0677	-7.3436	0.0000	-0.6300	-0.3645
age_bracket_55 - 64 Years	-0.2608	0.0685	-3.8055	0.0001	-0.3952	-0.1265
age_bracket_65+ Years	-0.1926	0.0726	-2.6515	0.0080	-0.3350	-0.0502
qualification_12 Years	-0.6460	0.2006	-3.2197	0.0013	-1.0392	-0.2527
qualification_< 12 Years	-0.8585	0.2092	-4.1033	0.0000	-1.2686	-0.4484
qualification_College Graduate	-0.4763	0.1993	-2.3900	0.0168	-0.8670	-0.0857
qualification_Some College	-0.6304	0.2000	-3.1514	0.0016	-1.0224	-0.2383
race_Hispanic	-0.5243	0.1031	-5.0862	0.0000	-0.7263	-0.3223
race_Other or Multiple	-0.5501	0.1025	-5.3686	0.0000	-0.7510	-0.3493
race_White	-0.7519	0.0668	-11.2634	0.0000	-0.8827	-0.6210
sex_Male	-0.1473	0.0427	-3.4523	0.0006	-0.2309	-0.0637
income_level_<= $75,000, Above Poverty	-0.0052	0.0738	-0.0700	0.9442	-0.1498	0.1395
income_level_> $75,000	0.2181	0.0831	2.6235	0.0087	0.0552	0.3810
income_level_Below Poverty	0.1407	0.0977	1.4398	0.1499	-0.0508	0.3323
marital_status_Married	-0.1132	0.2179	-0.5195	0.6034	-0.5404	0.3139
marital_status_Not Married	-0.5236	0.2173	-2.4092	0.0160	-0.9496	-0.0976
housing_status_Own	-0.2271	0.1311	-1.7325	0.0832	-0.4841	0.0298
housing_status_Rent	-0.4313	0.1373	-3.1408	0.0017	-0.7004	-0.1621
employment_Employed	-0.6234	0.2143	-2.9097	0.0036	-1.0434	-0.2035
employment_Not in Labor Force	-0.5307	0.2153	-2.4654	0.0137	-0.9527	-0.1088

employment_Unemployed	-0.6489	0.2337	-2.7769	0.0055	-1.1069	-0.1909
census_msa_MSA, Principle City	-0.1339	0.0482	-2.7799	0.0054	-0.2282	-0.0395
census_msa_Non-MSA	-0.0438	0.0502	-0.8724	0.3830	-0.1423	0.0547

AIC value is decreasing as 15662 so this variable will affect the accuracy so this variable should be removed

In [40]:
```
#last AIC VALUE=15662
x_train=x_train.drop("antiviral_medication",axis=1)
```

In [41]:
```
model_st=sm.Logit(y_train,x_train).fit()
```

```
Optimization terminated successfully.
         Current function value: 0.416537
         Iterations 6
```

In [37]:
```
model_st.summary2()
```

Out[37]:

Model:	Logit	Pseudo R-squared:	0.197
Dependent Variable:	h1n1_vaccine	AIC:	15663.4777
Date:	2021-09-01 15:57	BIC:	16016.0958
No. Observations:	18694	Log-Likelihood:	-7786.7
Df Model:	44	LL-Null:	-9693.7
Df Residuals:	18649	LLR p-value:	0.0000
Converged:	1.0000	Scale:	1.0000
No. Iterations:	6.0000		

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
unique_id	-0.0000	0.0000	-6.5064	0.0000	-0.0000	-0.0000
h1n1_worry	-0.0946	0.0268	-3.5318	0.0004	-0.1471	-0.0421
contact_avoidance	-0.1302	0.0503	-2.5862	0.0097	-0.2289	-0.0315
bought_face_mask	0.2277	0.0777	2.9311	0.0034	0.0755	0.3800
wash_hands_frequently	-0.2569	0.0607	-4.2307	0.0000	-0.3759	-0.1379
avoid_large_gatherings	-0.1451	0.0535	-2.7096	0.0067	-0.2500	-0.0401
reduced_outside_home_cont	-0.0432	0.0543	-0.7946	0.4268	-0.1497	0.0633
avoid_touch_face	-0.0030	0.0496	-0.0613	0.9511	-0.1002	0.0942
dr_recc_h1n1_vacc	2.0269	0.0611	33.1905	0.0000	1.9072	2.1466
dr_recc_seasonal_vacc	-0.4527	0.0599	-7.5597	0.0000	-0.5701	-0.3354
chronic_medic_condition	0.1456	0.0470	3.0953	0.0020	0.0534	0.2377
cont_child_undr_6_mnths	0.1513	0.0723	2.0925	0.0364	0.0096	0.2930
is_health_worker	0.7703	0.0613	12.5599	0.0000	0.6501	0.8905
is_h1n1_vacc_effective	0.2668	0.0233	11.4465	0.0000	0.2211	0.3125
is_h1n1_risky	0.3269	0.0196	16.7041	0.0000	0.2885	0.3652
sick_from_h1n1_vacc	-0.0532	0.0182	-2.9231	0.0035	-0.0889	-0.0175

is_seas_vacc_effective	-0.0348	0.0219	-1.5908	0.1117	-0.0778	0.0081
is_seas_risky	0.1818	0.0188	9.6839	0.0000	0.1450	0.2186
sick_from_seas_vacc	-0.1410	0.0180	-7.8320	0.0000	-0.1762	-0.1057
no_of_adults	-0.2420	0.0318	-7.6176	0.0000	-0.3043	-0.1797
no_of_children	-0.2120	0.0275	-7.7078	0.0000	-0.2659	-0.1581
age_bracket_35 - 44 Years	-0.3956	0.0730	-5.4178	0.0000	-0.5388	-0.2525
age_bracket_45 - 54 Years	-0.5005	0.0677	-7.3940	0.0000	-0.6331	-0.3678
age_bracket_55 - 64 Years	-0.2653	0.0685	-3.8739	0.0001	-0.3995	-0.1311
age_bracket_65+ Years	-0.1988	0.0725	-2.7409	0.0061	-0.3410	-0.0566
qualification_12 Years	-0.6502	0.2006	-3.2409	0.0012	-1.0435	-0.2570
qualification_< 12 Years	-0.8607	0.2092	-4.1136	0.0000	-1.2707	-0.4506
qualification_College Graduate	-0.4805	0.1993	-2.4110	0.0159	-0.8711	-0.0899
qualification_Some College	-0.6344	0.2000	-3.1717	0.0015	-1.0265	-0.2424
race_Hispanic	-0.5168	0.1030	-5.0192	0.0000	-0.7186	-0.3150
race_Other or Multiple	-0.5513	0.1024	-5.3810	0.0000	-0.7521	-0.3505
race_White	-0.7532	0.0668	-11.2834	0.0000	-0.8841	-0.6224
sex_Male	-0.1464	0.0427	-3.4325	0.0006	-0.2300	-0.0628
income_level_<= $75,000, Above Poverty	-0.0054	0.0738	-0.0727	0.9420	-0.1500	0.1393
income_level_> $75,000	0.2175	0.0831	2.6173	0.0089	0.0546	0.3804
income_level_Below Poverty	0.1456	0.0977	1.4909	0.1360	-0.0458	0.3371
marital_status_Married	-0.1125	0.2181	-0.5159	0.6060	-0.5400	0.3150
marital_status_Not Married	-0.5206	0.2175	-2.3935	0.0167	-0.9470	-0.0943
housing_status_Own	-0.2282	0.1310	-1.7413	0.0816	-0.4850	0.0286
housing_status_Rent	-0.4325	0.1372	-3.1513	0.0016	-0.7015	-0.1635
employment_Employed	-0.6212	0.2143	-2.8982	0.0038	-1.0413	-0.2011
employment_Not in Labor Force	-0.5290	0.2154	-2.4564	0.0140	-0.9511	-0.1069
employment_Unemployed	-0.6469	0.2337	-2.7678	0.0056	-1.1051	-0.1888

census_msa_MSA, Principle City	-0.1320	0.0481	-2.7422	0.0061	-0.2263	-0.0376
census_msa_Non-MSA	-0.0432	0.0502	-0.8594	0.3901	-0.1417	0.0553

AIC value is increasing so this variable should not be dropped

In [73]:
```
x_train=x_train.drop("avoid_touch_face",axis=1)
```

In [45]:
```
model_st=sm.Logit(y_train,x_train).fit()
```

```
Optimization terminated successfully.
         Current function value: 0.416463
         Iterations 6
```

In [46]:
```
model_st.summary2()
```

Out[46]:

Model:	Logit	Pseudo R-squared:	0.197
Dependent Variable:	h1n1_vaccine	AIC:	15660.7087
Date:	2021-09-01 15:59	BIC:	16013.3268
No. Observations:	18694	Log-Likelihood:	-7785.4
Df Model:	44	LL-Null:	-9693.7
Df Residuals:	18649	LLR p-value:	0.0000
Converged:	1.0000	Scale:	1.0000
No. Iterations:	6.0000		

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
unique_id	-0.0000	0.0000	-6.5100	0.0000	-0.0000	-0.0000
h1n1_worry	-0.0952	0.0267	-3.5620	0.0004	-0.1476	-0.0428
antiviral_medication	0.1566	0.0935	1.6753	0.0939	-0.0266	0.3398
contact_avoidance	-0.1315	0.0494	-2.6607	0.0078	-0.2284	-0.0346
bought_face_mask	0.2143	0.0781	2.7444	0.0061	0.0613	0.3674
wash_hands_frequently	-0.2600	0.0590	-4.4051	0.0000	-0.3757	-0.1443
avoid_large_gatherings	-0.1476	0.0535	-2.7590	0.0058	-0.2524	-0.0427
reduced_outside_home_cont	-0.0486	0.0542	-0.8969	0.3698	-0.1547	0.0576
dr_recc_h1n1_vacc	2.0264	0.0611	33.1833	0.0000	1.9067	2.1461
dr_recc_seasonal_vacc	-0.4533	0.0599	-7.5712	0.0000	-0.5707	-0.3360
chronic_medic_condition	0.1457	0.0470	3.0979	0.0019	0.0535	0.2378
cont_child_undr_6_mnths	0.1505	0.0723	2.0814	0.0374	0.0088	0.2922
is_health_worker	0.7707	0.0612	12.5857	0.0000	0.6507	0.8908
is_h1n1_vacc_effective	0.2667	0.0233	11.4500	0.0000	0.2210	0.3123
is_h1n1_risky	0.3262	0.0196	16.6675	0.0000	0.2878	0.3645
sick_from_h1n1_vacc	-0.0530	0.0182	-2.9132	0.0036	-0.0887	-0.0174

is_seas_vacc_effective	-0.0346	0.0219	-1.5810	0.1139	-0.0775	0.0083
is_seas_risky	0.1816	0.0188	9.6717	0.0000	0.1448	0.2184
sick_from_seas_vacc	-0.1415	0.0180	-7.8617	0.0000	-0.1768	-0.1062
no_of_adults	-0.2436	0.0318	-7.6630	0.0000	-0.3059	-0.1813
no_of_children	-0.2129	0.0275	-7.7381	0.0000	-0.2668	-0.1590
age_bracket_35 - 44 Years	-0.3957	0.0730	-5.4194	0.0000	-0.5388	-0.2526
age_bracket_45 - 54 Years	-0.4975	0.0676	-7.3552	0.0000	-0.6301	-0.3650
age_bracket_55 - 64 Years	-0.2611	0.0685	-3.8146	0.0001	-0.3953	-0.1270
age_bracket_65+ Years	-0.1929	0.0726	-2.6577	0.0079	-0.3351	-0.0506
qualification_12 Years	-0.6463	0.2006	-3.2215	0.0013	-1.0394	-0.2531
qualification_< 12 Years	-0.8588	0.2092	-4.1050	0.0000	-1.2688	-0.4487
qualification_College Graduate	-0.4765	0.1993	-2.3909	0.0168	-0.8671	-0.0859
qualification_Some College	-0.6306	0.2000	-3.1529	0.0016	-1.0226	-0.2386
race_Hispanic	-0.5242	0.1031	-5.0855	0.0000	-0.7262	-0.3222
race_Other or Multiple	-0.5501	0.1025	-5.3684	0.0000	-0.7509	-0.3493
race_White	-0.7519	0.0668	-11.2634	0.0000	-0.8827	-0.6210
sex_Male	-0.1470	0.0425	-3.4556	0.0005	-0.2303	-0.0636
income_level_<= $75,000, Above Poverty	-0.0051	0.0738	-0.0693	0.9448	-0.1498	0.1396
income_level_> $75,000	0.2182	0.0831	2.6256	0.0086	0.0553	0.3811
income_level_Below Poverty	0.1408	0.0977	1.4404	0.1498	-0.0508	0.3324
marital_status_Married	-0.1133	0.2179	-0.5201	0.6030	-0.5405	0.3138
marital_status_Not Married	-0.5237	0.2173	-2.4098	0.0160	-0.9497	-0.0978
housing_status_Own	-0.2271	0.1311	-1.7320	0.0833	-0.4840	0.0299
housing_status_Rent	-0.4312	0.1373	-3.1405	0.0017	-0.7004	-0.1621
employment_Employed	-0.6235	0.2142	-2.9101	0.0036	-1.0434	-0.2036
employment_Not in Labor Force	-0.5307	0.2153	-2.4654	0.0137	-0.9526	-0.1088
employment_Unemployed	-0.6489	0.2337	-2.7769	0.0055	-1.1069	-0.1909

census_msa_MSA, Principle City	-0.1339	0.0482	-2.7805	0.0054	-0.2283	-0.0395
census_msa_Non-MSA	-0.0439	0.0502	-0.8740	0.3821	-0.1424	0.0546

AIC value is decreasing as 15660 so this variable will affect the accuracy so this variable should be removed

In [74]:
```
#last AIC value=15660
x_train=x_train.drop("reduced_outside_home_cont",axis=1)
```

In [51]:
```
model_st=sm.Logit(y_train,x_train).fit()
```

```
Optimization terminated successfully.
         Current function value: 0.416484
         Iterations 6
```

In [52]:
```
model_st.summary2()
```

Out[52]:

Model:	Logit	Pseudo R-squared:	0.197
Dependent Variable:	h1n1_vaccine	AIC:	15659.5139
Date:	2021-09-01 15:59	BIC:	16004.2960
No. Observations:	18694	Log-Likelihood:	-7785.8
Df Model:	43	LL-Null:	-9693.7
Df Residuals:	18650	LLR p-value:	0.0000
Converged:	1.0000	Scale:	1.0000
No. Iterations:	6.0000		

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
unique_id	-0.0000	0.0000	-6.5188	0.0000	-0.0000	-0.0000
h1n1_worry	-0.0966	0.0267	-3.6205	0.0003	-0.1489	-0.0443
antiviral_medication	0.1518	0.0933	1.6272	0.1037	-0.0310	0.3347
contact_avoidance	-0.1353	0.0493	-2.7471	0.0060	-0.2319	-0.0388
bought_face_mask	0.2109	0.0780	2.7038	0.0069	0.0580	0.3638
wash_hands_frequently	-0.2627	0.0589	-4.4565	0.0000	-0.3782	-0.1472
avoid_large_gatherings	-0.1708	0.0468	-3.6458	0.0003	-0.2626	-0.0790
dr_recc_h1n1_vacc	2.0264	0.0611	33.1859	0.0000	1.9068	2.1461
dr_recc_seasonal_vacc	-0.4540	0.0599	-7.5829	0.0000	-0.5713	-0.3366
chronic_medic_condition	0.1452	0.0470	3.0885	0.0020	0.0531	0.2374
cont_child_undr_6_mnths	0.1504	0.0723	2.0796	0.0376	0.0087	0.2921
is_health_worker	0.7710	0.0612	12.5902	0.0000	0.6510	0.8910
is_h1n1_vacc_effective	0.2665	0.0233	11.4429	0.0000	0.2208	0.3121
is_h1n1_risky	0.3259	0.0196	16.6580	0.0000	0.2876	0.3643
sick_from_h1n1_vacc	-0.0534	0.0182	-2.9343	0.0033	-0.0891	-0.0177
is_seas_vacc_effective	-0.0343	0.0219	-1.5656	0.1174	-0.0772	0.0086

is_seas_risky	0.1815	0.0188	9.6705	0.0000	0.1448	0.2183
sick_from_seas_vacc	-0.1419	0.0180	-7.8847	0.0000	-0.1772	-0.1066
no_of_adults	-0.2433	0.0318	-7.6563	0.0000	-0.3056	-0.1811
no_of_children	-0.2126	0.0275	-7.7283	0.0000	-0.2665	-0.1587
age_bracket_35 - 44 Years	-0.3967	0.0730	-5.4333	0.0000	-0.5397	-0.2536
age_bracket_45 - 54 Years	-0.4980	0.0676	-7.3627	0.0000	-0.6306	-0.3654
age_bracket_55 - 64 Years	-0.2621	0.0684	-3.8298	0.0001	-0.3963	-0.1280
age_bracket_65+ Years	-0.1951	0.0725	-2.6898	0.0071	-0.3373	-0.0529
qualification_12 Years	-0.6454	0.2006	-3.2165	0.0013	-1.0386	-0.2521
qualification_< 12 Years	-0.8586	0.2092	-4.1037	0.0000	-1.2687	-0.4485
qualification_College Graduate	-0.4725	0.1993	-2.3711	0.0177	-0.8631	-0.0819
qualification_Some College	-0.6282	0.2000	-3.1405	0.0017	-1.0202	-0.2361
race_Hispanic	-0.5235	0.1031	-5.0785	0.0000	-0.7255	-0.3215
race_Other or Multiple	-0.5494	0.1025	-5.3619	0.0000	-0.7502	-0.3486
race_White	-0.7498	0.0667	-11.2423	0.0000	-0.8806	-0.6191
sex_Male	-0.1481	0.0425	-3.4835	0.0005	-0.2314	-0.0648
income_level_<= $75,000, Above Poverty	-0.0042	0.0738	-0.0563	0.9551	-0.1488	0.1405
income_level_> $75,000	0.2200	0.0831	2.6482	0.0081	0.0572	0.3829
income_level_Below Poverty	0.1392	0.0977	1.4247	0.1543	-0.0523	0.3308
marital_status_Married	-0.1165	0.2180	-0.5346	0.5929	-0.5438	0.3107
marital_status_Not Married	-0.5261	0.2174	-2.4200	0.0155	-0.9522	-0.1000
housing_status_Own	-0.2266	0.1311	-1.7284	0.0839	-0.4836	0.0304
housing_status_Rent	-0.4322	0.1373	-3.1473	0.0016	-0.7013	-0.1630
employment_Employed	-0.6215	0.2143	-2.9006	0.0037	-1.0414	-0.2015
employment_Not in Labor Force	-0.5304	0.2153	-2.4636	0.0138	-0.9523	-0.1084
employment_Unemployed	-0.6490	0.2337	-2.7770	0.0055	-1.1070	-0.1910
census_msa_MSA, Principle City	-0.1336	0.0481	-2.7751	0.0055	-0.2280	-0.0392

census_msa_Non-MSA	-0.0442	0.0502	-0.8799	0.3789	-0.1427	0.0543

AIC value is decreasing as 15659 so this variable will affect the accuracy so this variable should be removed

In [61]:
```
#last AIC value=15659
x_train=x_train.drop("is_seas_vacc_effective",axis=1)
```

In [62]:
```
model_st=sm.Logit(y_train,x_train).fit()
```

```
Optimization terminated successfully.
         Current function value: 0.416550
         Iterations 6
```

In [63]:
```
model_st.summary2()
```

Out[63]:

Model:	Logit	Pseudo R-squared:	0.197
Dependent Variable:	h1n1_vaccine	AIC:	15659.9572
Date:	2021-09-01 16:01	BIC:	15996.9034
No. Observations:	18694	Log-Likelihood:	-7787.0
Df Model:	42	LL-Null:	-9693.7
Df Residuals:	18651	LLR p-value:	0.0000
Converged:	1.0000	Scale:	1.0000
No. Iterations:	6.0000		

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
unique_id	-0.0000	0.0000	-6.5537	0.0000	-0.0000	-0.0000
h1n1_worry	-0.0982	0.0267	-3.6823	0.0002	-0.1505	-0.0459
antiviral_medication	0.1529	0.0933	1.6387	0.1013	-0.0300	0.3358
contact_avoidance	-0.1379	0.0492	-2.8025	0.0051	-0.2344	-0.0415
bought_face_mask	0.2121	0.0780	2.7188	0.0066	0.0592	0.3649
wash_hands_frequently	-0.2662	0.0589	-4.5212	0.0000	-0.3816	-0.1508
avoid_large_gatherings	-0.1710	0.0468	-3.6510	0.0003	-0.2628	-0.0792
dr_recc_h1n1_vacc	2.0305	0.0610	33.2922	0.0000	1.9110	2.1501
dr_recc_seasonal_vacc	-0.4579	0.0598	-7.6593	0.0000	-0.5751	-0.3407
chronic_medic_condition	0.1448	0.0470	3.0811	0.0021	0.0527	0.2370
cont_child_undr_6_mnths	0.1492	0.0723	2.0633	0.0391	0.0075	0.2909
is_health_worker	0.7701	0.0612	12.5758	0.0000	0.6501	0.8901
is_h1n1_vacc_effective	0.2505	0.0209	11.9861	0.0000	0.2095	0.2915
is_h1n1_risky	0.3263	0.0196	16.6795	0.0000	0.2880	0.3646
sick_from_h1n1_vacc	-0.0550	0.0182	-3.0267	0.0025	-0.0906	-0.0194
is_seas_risky	0.1749	0.0183	9.5697	0.0000	0.1391	0.2107

sick_from_seas_vacc	-0.1411	0.0180	-7.8403	0.0000	-0.1764	-0.1058
no_of_adults	-0.2452	0.0318	-7.7162	0.0000	-0.3075	-0.1829
no_of_children	-0.2135	0.0275	-7.7628	0.0000	-0.2674	-0.1596
age_bracket_35 - 44 Years	-0.3991	0.0730	-5.4674	0.0000	-0.5421	-0.2560
age_bracket_45 - 54 Years	-0.5022	0.0676	-7.4305	0.0000	-0.6347	-0.3698
age_bracket_55 - 64 Years	-0.2694	0.0683	-3.9447	0.0001	-0.4032	-0.1355
age_bracket_65+ Years	-0.2061	0.0722	-2.8541	0.0043	-0.3476	-0.0646
qualification_12 Years	-0.6558	0.2003	-3.2739	0.0011	-1.0484	-0.2632
qualification_< 12 Years	-0.8685	0.2089	-4.1572	0.0000	-1.2780	-0.4591
qualification_College Graduate	-0.4821	0.1990	-2.4234	0.0154	-0.8721	-0.0922
qualification_Some College	-0.6379	0.1997	-3.1941	0.0014	-1.0293	-0.2465
race_Hispanic	-0.5316	0.1030	-5.1632	0.0000	-0.7334	-0.3298
race_Other or Multiple	-0.5565	0.1024	-5.4359	0.0000	-0.7571	-0.3558
race_White	-0.7604	0.0664	-11.4584	0.0000	-0.8905	-0.6303
sex_Male	-0.1501	0.0425	-3.5321	0.0004	-0.2334	-0.0668
income_level_<= $75,000, Above Poverty	-0.0039	0.0738	-0.0532	0.9576	-0.1486	0.1407
income_level_> $75,000	0.2204	0.0831	2.6536	0.0080	0.0576	0.3832
income_level_Below Poverty	0.1431	0.0977	1.4647	0.1430	-0.0484	0.3346
marital_status_Married	-0.1267	0.2175	-0.5823	0.5604	-0.5530	0.2997
marital_status_Not Married	-0.5388	0.2169	-2.4844	0.0130	-0.9638	-0.1137
housing_status_Own	-0.2272	0.1311	-1.7331	0.0831	-0.4841	0.0297
housing_status_Rent	-0.4333	0.1373	-3.1564	0.0016	-0.7024	-0.1642
employment_Employed	-0.6289	0.2138	-2.9419	0.0033	-1.0480	-0.2099
employment_Not in Labor Force	-0.5399	0.2148	-2.5134	0.0120	-0.9608	-0.1189
employment_Unemployed	-0.6547	0.2333	-2.8066	0.0050	-1.1120	-0.1975
census_msa_MSA, Principle City	-0.1357	0.0481	-2.8187	0.0048	-0.2300	-0.0413
census_msa_Non-MSA	-0.0448	0.0502	-0.8915	0.3727	-0.1433	0.0537

AIC value didn't got changed so this variable should not be dropped

In [68]:
```
#last AIC value=15659
x_train=x_train.drop("income_level_Below Poverty",axis=1)
```

In [69]:
```
model_st=sm.Logit(y_train,x_train).fit()
```

```
Optimization terminated successfully.
         Current function value: 0.416539
         Iterations 6
```

In [70]:
```
model_st.summary2()
```

Out[70]:

Model:	Logit	Pseudo R-squared:	0.197
Dependent Variable:	h1n1_vaccine	AIC:	15659.5436
Date:	2021-09-01 16:01	BIC:	15996.4897
No. Observations:	18694	Log-Likelihood:	-7786.8
Df Model:	42	LL-Null:	-9693.7
Df Residuals:	18651	LLR p-value:	0.0000
Converged:	1.0000	Scale:	1.0000
No. Iterations:	6.0000		

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
unique_id	-0.0000	0.0000	-6.5274	0.0000	-0.0000	-0.0000
h1n1_worry	-0.0963	0.0267	-3.6096	0.0003	-0.1486	-0.0440
antiviral_medication	0.1561	0.0933	1.6731	0.0943	-0.0268	0.3389
contact_avoidance	-0.1350	0.0493	-2.7397	0.0061	-0.2315	-0.0384
bought_face_mask	0.2121	0.0780	2.7181	0.0066	0.0592	0.3650
wash_hands_frequently	-0.2641	0.0589	-4.4824	0.0000	-0.3796	-0.1486
avoid_large_gatherings	-0.1693	0.0468	-3.6153	0.0003	-0.2610	-0.0775
dr_recc_h1n1_vacc	2.0265	0.0611	33.1880	0.0000	1.9068	2.1461
dr_recc_seasonal_vacc	-0.4541	0.0599	-7.5846	0.0000	-0.5714	-0.3367
chronic_medic_condition	0.1475	0.0470	3.1402	0.0017	0.0555	0.2396
cont_child_undr_6_mnths	0.1495	0.0723	2.0677	0.0387	0.0078	0.2913
is_health_worker	0.7695	0.0612	12.5669	0.0000	0.6495	0.8895
is_h1n1_vacc_effective	0.2678	0.0233	11.5102	0.0000	0.2222	0.3134
is_h1n1_risky	0.3261	0.0196	16.6674	0.0000	0.2877	0.3644
sick_from_h1n1_vacc	-0.0534	0.0182	-2.9321	0.0034	-0.0890	-0.0177
is_seas_vacc_effective	-0.0351	0.0219	-1.6023	0.1091	-0.0780	0.0078

is_seas_risky	0.1820	0.0188	9.6941	0.0000	0.1452	0.2188
sick_from_seas_vacc	-0.1418	0.0180	-7.8786	0.0000	-0.1771	-0.1065
no_of_adults	-0.2427	0.0318	-7.6365	0.0000	-0.3050	-0.1804
no_of_children	-0.2102	0.0275	-7.6545	0.0000	-0.2640	-0.1563
age_bracket_35 - 44 Years	-0.3984	0.0730	-5.4575	0.0000	-0.5415	-0.2553
age_bracket_45 - 54 Years	-0.4994	0.0676	-7.3859	0.0000	-0.6320	-0.3669
age_bracket_55 - 64 Years	-0.2649	0.0684	-3.8722	0.0001	-0.3990	-0.1308
age_bracket_65+ Years	-0.2025	0.0723	-2.7985	0.0051	-0.3443	-0.0607
qualification_12 Years	-0.6384	0.2002	-3.1894	0.0014	-1.0307	-0.2461
qualification_< 12 Years	-0.8428	0.2085	-4.0421	0.0001	-1.2515	-0.4341
qualification_College Graduate	-0.4703	0.1989	-2.3651	0.0180	-0.8601	-0.0806
qualification_Some College	-0.6237	0.1996	-3.1249	0.0018	-1.0149	-0.2325
race_Hispanic	-0.5223	0.1031	-5.0676	0.0000	-0.7243	-0.3203
race_Other or Multiple	-0.5530	0.1024	-5.3990	0.0000	-0.7538	-0.3523
race_White	-0.7545	0.0666	-11.3274	0.0000	-0.8851	-0.6240
sex_Male	-0.1486	0.0425	-3.4956	0.0005	-0.2319	-0.0653
income_level_<= $75,000, Above Poverty	-0.0667	0.0588	-1.1335	0.2570	-0.1820	0.0486
income_level_> $75,000	0.1619	0.0719	2.2503	0.0244	0.0209	0.3029
marital_status_Married	-0.1171	0.2177	-0.5379	0.5907	-0.5438	0.3096
marital_status_Not Married	-0.5203	0.2171	-2.3972	0.0165	-0.9457	-0.0949
housing_status_Own	-0.1884	0.1283	-1.4681	0.1421	-0.4398	0.0631
housing_status_Rent	-0.3841	0.1331	-2.8861	0.0039	-0.6449	-0.1232
employment_Employed	-0.6082	0.2139	-2.8438	0.0045	-1.0274	-0.1890
employment_Not in Labor Force	-0.5137	0.2148	-2.3918	0.0168	-0.9347	-0.0928
employment_Unemployed	-0.6263	0.2330	-2.6882	0.0072	-1.0830	-0.1697
census_msa_MSA, Principle City	-0.1337	0.0481	-2.7770	0.0055	-0.2281	-0.0393
census_msa_Non-MSA	-0.0406	0.0502	-0.8096	0.4181	-0.1390	0.0577

AIC value didn't got changed so this variable should not be dropped

In [75]:
```
#last AIC value=15659
x_train=x_train.drop("census_msa_Non-MSA",axis=1)
```

In [76]:
```
model_st=sm.Logit(y_train,x_train).fit()
```

```
Optimization terminated successfully.
         Current function value: 0.416505
         Iterations 6
```

In [77]:
```
model_st.summary2()
```

Out[77]:

Model:	Logit	Pseudo R-squared:	0.197
Dependent Variable:	h1n1_vaccine	AIC:	15658.2891
Date:	2021-09-01 16:02	BIC:	15995.2353
No. Observations:	18694	Log-Likelihood:	-7786.1
Df Model:	42	LL-Null:	-9693.7
Df Residuals:	18651	LLR p-value:	0.0000
Converged:	1.0000	Scale:	1.0000
No. Iterations:	6.0000		

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
unique_id	-0.0000	0.0000	-6.5458	0.0000	-0.0000	-0.0000
h1n1_worry	-0.0971	0.0267	-3.6391	0.0003	-0.1494	-0.0448
antiviral_medication	0.1511	0.0933	1.6197	0.1053	-0.0317	0.3339
contact_avoidance	-0.1344	0.0492	-2.7298	0.0063	-0.2310	-0.0379
bought_face_mask	0.2120	0.0780	2.7175	0.0066	0.0591	0.3649
wash_hands_frequently	-0.2631	0.0589	-4.4638	0.0000	-0.3786	-0.1476
avoid_large_gatherings	-0.1722	0.0468	-3.6782	0.0002	-0.2639	-0.0804
dr_recc_h1n1_vacc	2.0269	0.0611	33.1917	0.0000	1.9072	2.1466
dr_recc_seasonal_vacc	-0.4539	0.0599	-7.5815	0.0000	-0.5713	-0.3366
chronic_medic_condition	0.1459	0.0470	3.1036	0.0019	0.0538	0.2380
cont_child_undr_6_mnths	0.1491	0.0723	2.0619	0.0392	0.0074	0.2907
is_health_worker	0.7705	0.0612	12.5816	0.0000	0.6504	0.8905
is_h1n1_vacc_effective	0.2661	0.0233	11.4320	0.0000	0.2205	0.3118
is_h1n1_risky	0.3258	0.0196	16.6543	0.0000	0.2875	0.3642
sick_from_h1n1_vacc	-0.0533	0.0182	-2.9303	0.0034	-0.0890	-0.0177
is_seas_vacc_effective	-0.0344	0.0219	-1.5723	0.1159	-0.0773	0.0085

is_seas_risky	0.1816	0.0188	9.6748	0.0000	0.1448	0.2184
sick_from_seas_vacc	-0.1421	0.0180	-7.8972	0.0000	-0.1774	-0.1069
no_of_adults	-0.2432	0.0318	-7.6509	0.0000	-0.3055	-0.1809
no_of_children	-0.2133	0.0275	-7.7587	0.0000	-0.2672	-0.1594
age_bracket_35 - 44 Years	-0.3972	0.0730	-5.4410	0.0000	-0.5403	-0.2541
age_bracket_45 - 54 Years	-0.4995	0.0676	-7.3879	0.0000	-0.6320	-0.3670
age_bracket_55 - 64 Years	-0.2648	0.0684	-3.8732	0.0001	-0.3988	-0.1308
age_bracket_65+ Years	-0.1976	0.0725	-2.7271	0.0064	-0.3396	-0.0556
qualification_12 Years	-0.6469	0.2005	-3.2269	0.0013	-1.0398	-0.2540
qualification_< 12 Years	-0.8615	0.2090	-4.1214	0.0000	-1.2712	-0.4518
qualification_College Graduate	-0.4704	0.1991	-2.3626	0.0181	-0.8606	-0.0802
qualification_Some College	-0.6285	0.1999	-3.1448	0.0017	-1.0202	-0.2368
race_Hispanic	-0.5253	0.1030	-5.0982	0.0000	-0.7273	-0.3234
race_Other or Multiple	-0.5546	0.1023	-5.4225	0.0000	-0.7550	-0.3541
race_White	-0.7561	0.0663	-11.4073	0.0000	-0.8860	-0.6262
sex_Male	-0.1486	0.0425	-3.4961	0.0005	-0.2319	-0.0653
income_level_<= $75,000, Above Poverty	-0.0058	0.0738	-0.0793	0.9368	-0.1505	0.1388
income_level_> $75,000	0.2221	0.0831	2.6746	0.0075	0.0593	0.3849
income_level_Below Poverty	0.1349	0.0976	1.3823	0.1669	-0.0564	0.3262
marital_status_Married	-0.1181	0.2178	-0.5422	0.5877	-0.5451	0.3088
marital_status_Not Married	-0.5275	0.2173	-2.4282	0.0152	-0.9534	-0.1017
housing_status_Own	-0.2260	0.1311	-1.7244	0.0846	-0.4830	0.0309
housing_status_Rent	-0.4309	0.1373	-3.1390	0.0017	-0.7000	-0.1619
employment_Employed	-0.6262	0.2140	-2.9256	0.0034	-1.0457	-0.2067
employment_Not in Labor Force	-0.5340	0.2151	-2.4826	0.0130	-0.9555	-0.1124
employment_Unemployed	-0.6509	0.2335	-2.7872	0.0053	-1.1086	-0.1932
census_msa_MSA, Principle City	-0.1185	0.0450	-2.6325	0.0085	-0.2067	-0.0303

AIC value is decreasing as 15658 so this variable will affect the accuracy so this variable should be removed

variables which affect AIC=h1n1_awareness,avoid_touch_face,reduced_outside_home_cont,cens MSA

AIC value of 15658 is the least by finally removing all the variables which have high p value

.....Removing the columns in x_test which is also removed in x_train.....

In [78]:
```
col_to_drop=["h1n1_awareness","avoid_touch_face","reduced_outside_home_cont","census_msa_Non-MSA"]
```

In [79]:
```
x_test=x_test.drop(col_to_drop,axis=1)
```

In [80]:
```
x_test
```

Out[80]:

	unique_id	h1n1_worry	antiviral_medication	contact_avoidance	bought_face_mask	wash_hands_frequently	avoid_large_gatherings	dr_recc_h
25567	25567	2.0	0.0	1.0	0.0	1.0	0.0	
6023	6023	2.0	0.0	1.0	0.0	1.0	0.0	
22055	22055	3.0	0.0	0.0	0.0	0.0	0.0	
7914	7914	3.0	0.0	1.0	0.0	0.0	1.0	
12380	12380	2.0	0.0	0.0	0.0	1.0	1.0	
...	...	...	...	...	...	...	...	
17077	17077	1.0	0.0	1.0	0.0	1.0	0.0	
6729	6729	0.0	1.0	1.0	0.0	1.0	0.0	
22700	22700	0.0	0.0	1.0	0.0	1.0	0.0	
17911	17911	1.0	0.0	1.0	0.0	1.0	0.0	
13912	13912	1.0	0.0	0.0	0.0	1.0	0.0	

8013 rows × 43 columns

LOGISTIC REGRESSION

In [83]:
```
#importing log.reg and fitting the model to predict
from sklearn.linear_model import LogisticRegression
model1=LogisticRegression()
model1.fit(x_train,y_train)
```

Out[83]:
```
LogisticRegression()
```

In [82]:
```
import warnings
warnings.filterwarnings("ignore")
```

In [84]:
```
y_pred=model1.predict(x_test)
y_pred
```

Out[84]:
```
array([0, 0, 1, ..., 0, 0, 0], dtype=int64)
```

In [85]:
```
#checking actual values with machine predicted values
from sklearn.metrics import confusion_matrix,accuracy_score
```

In [86]:
```
confusion_matrix(y_test,y_pred)
```

Out[86]:
```
array([[5985,  345],
       [1041,  642]], dtype=int64)
```

True negative=5985,False positive=345,False nagative=1041,True positive=642

here it seems machine predicted is good

In [87]:
```
accuracy_score(y_test,y_pred)
```

Out[87]:
```
0.8270310745039311
```

"Accuracy seems to be high by using logistic regression,so this model will predict with very less chance of error occurance".

even though accuracy is 0.82 there are chances it might increase by using optimizer

ROC curve-(Reciver Operating Characterstic curve)

roc is mainly used because we can set a threshold point from where we can diffrentiate 0 and 1

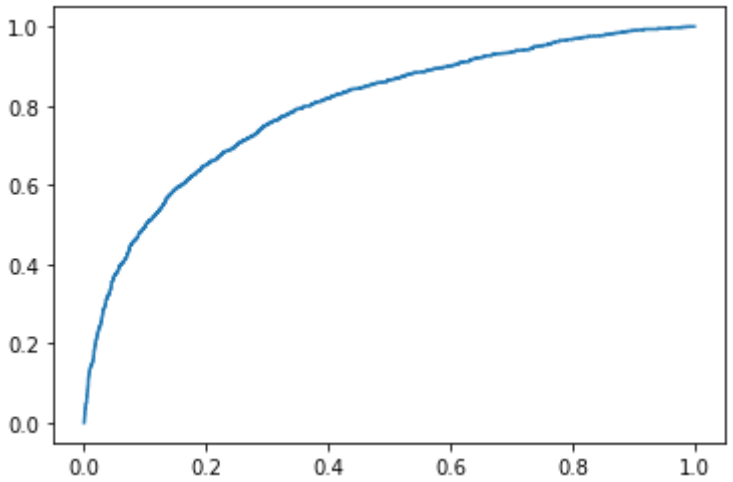
In [88]:
```
from sklearn.metrics import roc_auc_score
from sklearn.metrics import roc_curve
```

In [90]:
```
logis_roc_auc=roc_auc_score(y_test,y_pred)
```

In [89]:
```
fpr,tpr,thresholds=roc_curve(y_test,model1.predict_proba(x_test)[:,1])
```

In [91]:
```
plt.plot(fpr,tpr,logis_roc_auc)
```

Out[91]:
```
[<matplotlib.lines.Line2D at 0x25f497e1e50>,
 <matplotlib.lines.Line2D at 0x25f497e1f10>]
```



In [ ]:
```
#setting threshold as o.8 curve
```

In [98]:
```
roc_t=LogisticRegression(class_weight="balanced")
roc_t.fit(x_train,y_train)
THRESHOLD=0.8
```

In [100]:
```
import numpy as np
```

In [101]:
```
y_pred_roc=np.where(roc_t.predict_proba(x_test)[:,1]>THRESHOLD,1,0)
```

In [102]:
```
accuracy_score(y_test,y_pred_roc)
```

Out[102]:
```
0.8299014102084113
```

"After using ROC i got a slight increase in accuracy value than the accuracy value calculated using logistic regression"

In [ ]:
```
#to check models performance here classification_report is used.
```

In [103]:
```
from sklearn.metrics import classification_report
```

In [104]:
```
c_Report=classification_report(y_test,y_pred_roc)
```

In [105]:
```
print(c_Report)
```

```
              precision    recall  f1-score   support

           0       0.85      0.96      0.90      6330
           1       0.69      0.35      0.46      1683

    accuracy                           0.83      8013
   macro avg       0.77      0.65      0.68      8013
weighted avg       0.81      0.83      0.81      8013
```

In [ ]:
```

```