

```
In [25]: import pandas as pd

In [26]: testdataframe=pd.read_csv('../2021_Competition_Holdout.csv',low_memory=False)

In [27]: testdataframe.head()
```

	Unnamed: 0	ID	auth_3mth_post_acute_dia	rx_gpi2_72_pmpm_cost_6to9m_b4	atlas_pct_laccess_child15	atlas_recfacp
0	0	5M9OS1T3dLc09Ybc7013AI9f		0	9.735346	0.13
1	1	0M8f45b4a4Of6aS3TL6Ya9AI		0	6.264524	0.07
2	2	12MOS912Tb204LYacA2flc22		0	6.136065	0.10
3	3	M6O69S24T36LYec4AI5d0b56		0	1.244630	0.00
4	4	45947aM37Oc8S3dTLYA00lca		0	8.009543	0.06

5 rows × 367 columns

```
In [28]: #Saving missing values in a variable
a = testdataframe.isnull().sum()/len(testdataframe)*100
variables = testdataframe.columns
# saving column names in a variable
variables_no_missing=[]
variables_missing=[]
for i in range(len(variables)):
    if a[i]<=70:
        variables_no_missing.append(variables[i])
    else:
        variables_missing.append(variables[i])
```

```
In [29]: variables_no_missing
```

```
Out[29]: ['Unnamed: 0',
'ID',
'auth_3mth_post_acute_dia',
'rx_gpi2_72_pmpm_cost_6to9m_b4',
'atlas_pct_laccess_child15',
'atlas_recfacpth14',
'atlas_pct_fmrkt_frveg16',
'atlas_pct_free_lunch14',
'bh_ip_snf_net_paid_pmpm_cost_9to12m_b4',
'auth_3mth_acute_ckd',
'bh_ncal_pmpm_ct',
'src_div_id',
'total_bh_copay_pmpm_cost_t_9-6-3m_b4',
'bh_ip_snf_net_paid_pmpm_cost_3to6m_b4',
'cons_chmi',
'mcc_ano_pmpm_ct_t_9-6-3m_b4',
'auth_3mth_post_acute_trm',
'rx_maint_pmpm_cost_t_12-9-6m_b4',
'auth_3mth_post_acute_rsk',
'cons_ltmedicr',
'rx_gpi4_6110_pmpm_ct',
'atlas_pc_snapben15',
'credit_bal_nonmtgcredit_60dpd',
'rx_bh_mbr_resp_pmpm_cost_9to12m_b4',
'rx_nonbh_pmpm_cost_t_9-6-3m_b4',
'atlas_pct_laccess_nhna15',
'auth_3mth_acute_vco',
'credit_hh_nonmtgcredit_60dpd',
'rx_bh_pmpm_ct_0to3m_b4',
'auth_3mth_dc_ltac',
'cons_lwcm10',
'auth_3mth_post_acute_inj',
'atlas_fsrpth14',
'auth_3mth_dc_home',
'atlas_wicspth12',
'rx_gpi2_17_pmpm_cost_t_12-9-6m_b4',
'cons_hxmioc',
'rx_generic_pmpm_cost_t_6-3-0m_b4',
'cmsd2_sns_digest_abdomen_pmpm_ct',
'atlas_ghveg_farms12',
'credit_hh_bankcardcredit_60dpd',
'total_outpatient_allowed_pmpm_cost_6to9m_b4',
'cons_cwht',
'atlas_netmigrationrate1016',
'atlas_pct_laccess_snap15',
```

'bh_ncdm_ind',
'rx_nonmaint_mbr_resp_pmpm_cost_9to12m_b4',
'atlas_retirement_destination_2015_upda',
'rx_overall_mbr_resp_pmpm_cost_t_6-3-0m_b4',
'atlas_naturalchangerate1016',
'ccsp_236_pct',
'bh_ip_snf_mbr_resp_pmpm_cost_6to9m_b4',
'rx_overall_dist_gpi6_pmpm_ct_t_6-3-0m_b4',
'auth_3mth_post_acute_ben',
'atlas_pct_laccess_hisp15',
'auth_3mth_dc_no_ref',
'rx_overall_mbr_resp_pmpm_cost',
'rx_overall_gpi_pmpm_ct_0to3m_b4',
'auth_3mth_dc_snf',
'rx_phar_cat_humana_pmpm_ct_t_9-6-3m_b4',
'atlas_pct_laccess_hhnv15',
'auth_3mth_acute_ccs_048',
'bh_ip_snf_net_paid_pmpm_cost_0to3m_b4',
'auth_3mth_acute_end',
'auth_3mth_psychic',
'atlas_hiamenity',
'auth_3mth_bh_acute',
'credit_bal_consumerfinance',
'auth_3mth_acute_chf',
'rx_overall_gpi_pmpm_ct_t_6-3-0m_b4',
'rwjf_uninsured_pct',
'mcc_chf_pmpm_ct_t_9-6-3m_b4',
'rx_mail_mbr_resp_pmpm_cost_0to3m_b4',
'bh_urgent_care_copay_pmpm_cost_t_12-9-6m_b4',
'auth_3mth_hospice',
'auth_3mth_acute_bld',
'atlas_pct_wic15',
'ccsp_193_pct',
'auth_3mth_dc_hospice',
'auth_3mth_acute_ccs_030',
'atlas_pct_fmrkt_baked16',
'rx_nonmaint_mbr_resp_pmpm_cost',
'auth_3mth_acute_skn',
'atlas_veg_farms12',
'atlas_vlfoodsec_13_15',
'rx_gpi2_34_dist_gpi6_pmpm_ct',
'bh_ip_snf_net_paid_pmpm_cost',
'credit_hh_bankcard_severederog',
'rx_hum_16_pmpm_ct',
'est_age',
'rx_maint_pmpm_cost_t_6-3-0m_b4',
'cnt_cp_webstatement_pmpm_ct',
'atlas_pct_laccess_seniors15',
'phy_em_px_pct',
'atlas_percapitainc',
'rwjf_uninsured_adults_pct',
'rx_generic_mbr_resp_pmpm_cost_0to3m_b4',
'auth_3mth_acute_neo',
'rwjf_air_pollute_density',
'rx_gpi2_02_pmpm_cost',
'atlas_recfac14',
'cons_mobplus',
'lab_albumin_loinc_pmpm_ct',
'atlas_pct_obese_adults13',
'rx_maint_net_paid_pmpm_cost_t_12-9-6m_b4',
'rev_pm_obsrm_pmpm_ct',
'atlas_pct_sfsp15',
'total_physician_office_net_paid_pmpm_cost_9to12m_b4',
'atlas_pc_dirsales12',
'med_ip_snf_admit_days_pmpm',
'rej_med_outpatient_visit_ct_pmpm_t_6-3-0m_b4',
'auth_3mth_post_acute_vco',
'cms_tot_partd_payment_amt',
'rx_nonotc_dist_gpi6_pmpm_ct',
'rx_nonmaint_pmpm_ct',
'rx_nonbh_mbr_resp_pmpm_cost_6to9m_b4',
'cons_stlnindx',
'atlas_hipov_1115',
'auth_3mth_post_acute_dig',
'rx_nonbh_mbr_resp_pmpm_cost',
'atlas_redemp_snaps16',
'atlas_berry_farms12',
'rej_med_ip_snf_coins_pmpm_cost_t_9-6-3m_b4',
'rwjf_inactivity_pct',
'rx_gpi2_72_pmpm_ct_6to9m_b4',
'cons_n2pmr',
'med_physician_office_allowed_pmpm_cost_t_9-6-3m_b4',
'auth_3mth_acute_res',
'rev_cms_ct_pmpm_ct',
'atlas_foodhub16',
'total_physician_office_copay_pmpm_cost',
'auth_3mth_acute_dig',
'auth_3mth_dc_acute_rehab',
'atlas_pct_fmrkt_anmlprod16',

'auth_3mth_post_acute_hdz',
'bh_ip_snf_mbr_resp_pmpm_cost_3to6m_b4',
'auth_3mth_acute_ccs_172',
'credit_num_agencyfirstmtg',
'total_physician_office_net_paid_pmpm_cost_t_9-6-3m_b4',
'auth_3mth_acute_ccs_154',
'atlas_type_2015_mining_no',
'atlas_agritrsm_rct12',
'rx_days_since_last_script',
'atlas_pct_laccess_pop15',
'auth_3mth_post_acute_res',
'auth_3mth_acute_inf',
'rx_gpi2_01_pmpm_cost_0to3m_b4',
'atlas_povertyallagespct',
'rwjf_uninsured_child_pct',
'rx_branded_pmpm_ct_t_6-3-0m_b4',
'med_outpatient_deduct_pmpm_cost_t_9-6-3m_b4',
'credit_bal_mtgcredit_new',
'atlas_low_employment_2015_update',
'atlas_pct_diabetes_adults13',
'atlas_pct_laccess_nhasian15',
'atlas_deep_pov_all',
'atlas_net_international_migration_rate',
'atlas_deep_pov_children',
'bh_ncdm_pct',
'auth_3mth_non_er',
'atlas_foodinsec_child_03_11',
'rx_branded_mbr_resp_pmpm_cost',
'atlas_pc_wic_redemp12',
'rwjf_mv_deaths_rate',
'auth_3mth_acute_cad',
'atlas_pct_reduced_lunch14',
'cons_nwperadult',
'total_allowed_pmpm_cost_t_9-6-3m_b4',
'rx_hum_28_pmpm_cost',
'mabh_seg',
'cms_orig_reas_entitle_cd',
'atlas_totalocchu',
'med_physician_office_ds_clm_6to9m_b4',
'bh_ip_snf_mbr_resp_pmpm_cost_9to12m_b4',
'atlas_pct_loc1farm12',
'rx_generic_mbr_resp_pmpm_cost',
'total_outpatient_mbr_resp_pmpm_cost_6to9m_b4',
'auth_3mth_post_acute_cir',
'rx_gpi4_3400_pmpm_ct',
'auth_3mth_post_acute_cer',
'lab_dist_loinc_pmpm_ct',
'atlas_pct_nslp15',
'rx_generic_pmpm_ct_0to3m_b4',
'oontwk_mbr_resp_pmpm_cost_t_6-3-0m_b4',
'atlas_pct_laccess_lowi15',
'bh_ncal_ind',
'auth_3mth_post_acute_mus',
'atlas_pct_fmrkt_sfmnp16',
'hum_region',
'rx_nonmail_dist_gpi6_pmpm_ct_t_9-6-3m_b4',
'atlas_pct_loc1sale12',
'bh_ip_snf_net_paid_pmpm_cost_6to9m_b4',
'rej_med_er_net_paid_pmpm_cost_t_9-6-3m_b4',
'credit_bal_autobank',
'med_outpatient_mbr_resp_pmpm_cost_t_9-6-3m_b4',
'rx_overall_mbr_resp_pmpm_cost_0to3m_b4',
'rx_tier_2_pmpm_ct_3to6m_b4',
'rx_nonbh_net_paid_pmpm_cost',
'rx_maint_pmpm_ct_9to12m_b4',
'rx_nonbh_net_paid_pmpm_cost_t_6-3-0m_b4',
'atlas_type_2015_recreation_no',
'auth_3mth_post_acute_sns',
'rx_gpi2_39_pmpm_cost_t_6-3-0m_b4',
'atlas_type_2015_update',
'cms_risk_adjustment_factor_a_amt',
'total_ip_maternity_net_paid_pmpm_cost_t_12-9-6m_b4',
'rx_generic_pmpm_cost',
'cmsd2_eye_retina_pmpm_ct',
'auth_3mth_acute_can',
'auth_3mth_post_acute',
'auth_3mth_facility',
'rx_days_since_last_script_0to3m_b4',
'atlas_population_loss_2015_update',
'rx_maint_pmpm_ct_t_6-3-0m_b4',
'auth_3mth_post_acute_men',
'auth_3mth_acute_mean_los',
'credit_num_autofinance',
'cons_rxmaint',
'rx_mail_net_paid_pmpm_cost_t_6-3-0m_b4',
'auth_3mth_home',
'rx_maint_mbr_resp_pmpm_cost_6to9m_b4',
'cons_hxwearbl',
'total_physician_office_mbr_resp_pmpm_cost_t_9-6-3m_b4',

'atlas_pct_laccess_black15',
'atlas_hh65plusalonepct',
'atlas_farm_to_school13',
'auth_3mth_acute_inj',
'auth_3mth_acute_ccs_153',
'rej_days_since_last_clm',
'auth_3mth_transplant',
'bh_outpatient_net_paid_pmpm_cost',
'atlas_dirsales_farms12',
'rx_generic_pmpm_cost_6to9m_b4',
'rev_cms_ansth_pmpm_ct',
'atlas_convspth14',
'total_med_allowed_pmpm_cost_9to12m_b4',
'rx_mail_mbr_resp_pmpm_cost_t_9-6-3m_b4',
'med_outpatient_visit_ct_pmpm_t_12-9-6m_b4',
'rx_nonbh_pmpm_ct_t_9-6-3m_b4',
'auth_3mth_acute',
'rx_nonbh_pmpm_ct_0to3m_b4',
'atlas_pc_ffrsales12',
'auth_3mth_dc_left_ama',
'credit_bal_bankcard_severederog',
'atlas_povertyunder18pct',
'rx_tier_1_pmpm_ct_0to3m_b4',
'auth_3mth_acute_ccs_227',
'cons_estinv30_rc',
'auth_3mth_bh_acute_men',
'rx_gpi2_34_pmpm_ct',
'auth_3mth_dc_custodial',
'atlas_veg_acrespth12',
'atlas_grocpth14',
'total_med_net_paid_pmpm_cost_t_6-3-0m_b4',
'rx_gpi2_90_dist_gpi6_pmpm_ct_9to12m_b4',
'atlas_csa12',
'sex_cd',
'rx_gpi2_62_pmpm_cost_t_9-6-3m_b4',
'rx_overall_gpi_pmpm_ct_t_12-9-6m_b4',
'auth_3mth_ltac',
'cons_hhcomp',
'auth_3mth_acute_hdz',
'cons_rxadhs',
'auth_3mth_acute_men',
'atlas_pct_fmrkt_snap16',
'met_obese_diag_pct',
'cms_partd_ra_factor_amt',
'atlas_pct_sbp15',
'rwjf_resident_seg_black_inx',
'atlas_pct_cacfp15',
'auth_3mth_rehab',
'pdc_lip',
'atlas_ffrpth14',
'credit_num_autobank_new',
'auth_3mth_acute_ccs_086',
'rx_tier_2_pmpm_ct',
'cons_n2pwh',
'rx_nonmaint_dist_gpi6_pmpm_ct_t_12-9-6m_b4',
'atlas_berry_acrespth12',
'atlas_pct_fmrkt_credit16',
'atlas_slhouse12',
'atlas_pc_frsales12',
'credit_hh_1stmtgcredit',
'auth_3mth_snf_post_hsp',
'atlas_pct_fmrkt_wiccash16',
'atlas_foodinsec_13_15',
'auth_3mth_acute_cer',
'cons_rxadhm',
'atlas_fmrktptth16',
'rx_nonotc_pmpm_cost_t_6-3-0m_b4',
'cci_dia_m_pmpm_ct',
'auth_3mth_acute_trm',
'cons_n2phi',
'bh_physician_office_copay_pmpm_cost_6to9m_b4',
'rwjf_income_inequ_ratio',
'rej_total_physician_office_visit_ct_pmpm_0to3m_b4',
'auth_3mth_acute_dia',
'credit_num_nonmtgcredit_60dpd',
'auth_3mth_snf_direct',
'credit_bal_autofinance_new',
'auth_3mth_acute_ccs_067',
'auth_3mth_acute_ccs_043',
'rwjf_men_hlth_prov_ratio',
'auth_3mth_dc_home_health',
'rx_gpi2_56_dist_gpi6_pmpm_ct_3to6m_b4',
'cmsd2_sns_genitourinary_pmpm_ct',
'auth_3mth_acute_cir',
'auth_3mth_acute_ner',
'auth_3mth_acute_ccs_094',
'med_ambulance_coins_pmpm_cost_t_9-6-3m_b4',
'hedis_dia_hbalc_ge9',
'bh_ncal_pct',

```
'atlas_pct_snap16',
'ccsp_227_pct',
'atlas_ghveg_sqftpth12',
'rx_days_since_last_script_6to9m_b4',
'atlas_orchard_acrespth12',
'atlas_persistentchildpoverty_1980_2011',
'auth_3mth_post_acute_cad',
'atlas_pct_laccess_multir15',
'cons_cgqs',
'ccsp_065_pmpm_ct',
'auth_3mth_acute_ccs_044',
'atlas_medhhinc',
'rx_maint_net_paid_pmpm_cost_t_9-6-3m_b4',
'rwjf_mental_distress_pct',
'bh_ip_snf_admit_days_pmpm_t_9-6-3m_b4',
'rx_phar_cat_cvs_pmpm_ct_t_9-6-3m_b4',
'zip_cd',
'auth_3mth_post_acute_ckd',
'atlas_pct_laccess_nhpi15',
'auth_3mth_post_acute_ner',
'auth_3mth_post_er',
'credit_num_consumerfinance_new',
'rx_gpi2_49_pmpm_cost_0to3m_b4',
'cons_chva',
'atlas_avghhsize',
'rx_overall_net_paid_pmpm_cost_6to9m_b4',
'atlas_ownhomepct',
'atlas_orchard_farms12',
'total_physician_office_visit_ct_pmpm_t_6-3-0m_b4',
'atlas_pct_fmrkt_wic16',
'rx_gpi2_33_pmpm_ct_0to3m_b4',
'auth_3mth_post_acute_chf',
'rwjf_social_associate_rate',
'atlas_freshveg_farms12',
'auth_3mth_acute_ccs_042',
'auth_3mth_post_acute_inf',
'auth_3mth_acute_sns',
'days_since_last_clm_0to3m_b4',
'auth_3mth_dc_other',
'auth_3mth_bh_acute_mean_los',
'mcc_end_pct',
'auth_3mth_post_acute_gus',
'cons_lwcm07',
'atlas_pct_fmrkt_otherfood16',
'auth_3mth_post_acute_end',
'auth_3mth_acute_mus',
'atlas_perpov_1980_0711',
'atlas_pct_laccess_white15',
'auth_3mth_post_acute_mean_los',
'rx_gpi2_66_pmpm_ct',
'auth_3mth_acute_gus',
'rx_generic_dist_gpi6_pmpm_ct_t_9-6-3m_b4',
'atlas_low_education_2015_update',
'race_cd']
```

In [30]: variables_missing

Out[30]: ['lang_spoken_cd']

In [31]: len(variables_no_missing)

Out[31]: 366

In [32]: testdataframe.drop(['lang_spoken_cd'], axis=1, inplace=True)
testdataframe.head()

Out[32]:

	Unnamed: 0	ID	auth_3mth_post_acute_dia	rx_gpi2_72_pmpm_cost_6to9m_b4	atlas_pct_laccess_child15	atlas_recfacp
0	0	5M9OS1T3dLc09Ybc7013AI9f	0	0.0	9.735346	0.13
1	1	0M8f45b4a4Of6aS3TL6Ya9AI	0	0.0	6.264524	0.07
2	2	12MOS912Tb204LYacA2flc22	0	0.0	6.136065	0.10
3	3	M6O69S24T36LYec4AI5d0b56	0	0.0	1.244630	0.00
4	4	45947aM37Oc8S3dTLYA00lca	0	0.0	8.009543	0.06

5 rows × 366 columns

```
In [33]: #Imputation
for col in variables_no_missing:
    testdataframe[col].fillna(testdataframe[col].mode()[0], inplace=True)
```

```
In [34]: #categorical Columns
categ_cols = testdataframe.select_dtypes(include='object')
categ_cols
```

```
Out[34]:
```

	ID	auth_3mth_post_acute_dia	bh_ip_snf_net_paid_pmpm_cost_9to12m_b4	auth_3mth_acute_ckd	src_div_id	to
0	5M9OS1T3dLc09Ybc7013AI9f		0	0.0	0	000
1	0M8f45b4a4Of6aS3TL6Ya9AI		0	0.0	0	000
2	12MOS912Tb204LYacA2flc22		0	0.0	0	000
3	M6O69S24T36LYec4AI5d0b56		0	0.0	0	000
4	45947aM37Oc8S3dTLYA00lca		0	0.0	0	000
...
525153	9M7eOcedbS2f0cT72Le4YA7I		0	0.0	0	000
525154	6MOS9e391T29eLY3A1I93bdd		0	0.0	0	000
525155	3996Mad5bcOS8TL567Y8fAI8		0	0.0	0	000
525156	M2b6OS0TLY9cAeedI3e11cb0		0	0.0	0	000
525157	M8282dbf2Od8SbTL06Y2A31I		0	0.0	0	001

525158 rows x 107 columns

```
In [35]: #Printing cardinality of each categorical column
categ_cols.nunique()
```

```
Out[35]: ID 525158
auth_3mth_post_acute_dia 2
bh_ip_snf_net_paid_pmpm_cost_9to12m_b4 2
auth_3mth_acute_ckd 2
src_div_id 13
...
auth_3mth_acute_sns 3
auth_3mth_dc_other 3
auth_3mth_bh_acute_mean_los 3
auth_3mth_post_acute_gus 2
rx_generic_dist_gpi6_pmpm_ct_t_9-6-3m_b4 12
Length: 107, dtype: int64
```

```
In [36]: categ_cols.columns
```

```
Out[36]: Index(['ID', 'auth_3mth_post_acute_dia',
                'bh_ip_snf_net_paid_pmpm_cost_9to12m_b4', 'auth_3mth_acute_ckd',
                'src_div_id', 'total_bh_copay_pmpm_cost_t_9-6-3m_b4',
                'bh_ip_snf_net_paid_pmpm_cost_3to6m_b4', 'mcc ano_pmpm_ct_t_9-6-3m_b4',
                'auth_3mth_post_acute_trm', 'rx_maint_pmpm_cost_t_12-9-6m_b4',
                ...
                'rx_phar_cat_cvs_pmpm_ct_t_9-6-3m_b4', 'auth_3mth_post_er',
                'total_physician_office_visit_ct_pmpm_t_6-3-0m_b4',
                'rx_gpi2_33_pmpm_ct_0to3m_b4', 'auth_3mth_post_acute_chf',
                'auth_3mth_acute_sns', 'auth_3mth_dc_other',
                'auth_3mth_bh_acute_mean_los', 'auth_3mth_post_acute_gus',
                'rx_generic_dist_gpi6_pmpm_ct_t_9-6-3m_b4'],
                dtype='object', length=107)
```

```
In [37]: testdataframe['auth_3mth_post_acute_dia'] = testdataframe['auth_3mth_post_acute_dia'].astype(str)
```

```
In [38]: from sklearn.preprocessing import LabelEncoder

testdataframe['auth_3mth_post_acute_dia'] = LabelEncoder().fit_transform(testdataframe['auth_3mth_post_acute_dia'])
```

```
In [39]: testdataframe['auth_3mth_post_acute_dia'].head(10220)
```

```
Out[39]: 0      1
          1      1
          2      1
          3      1
          4      1
          ..
10215    1
10216    1
10217    1
10218    1
10219    1
Name: auth_3mth_post_acute_dia, Length: 10220, dtype: int32
```

```
In [40]: list1 = ['ID', 'auth_3mth_post_acute_dia',
                  'bh_ip_snf_net_paid_pmpm_cost_9to12m_b4', 'auth_3mth_acute_ckd',
                  'src_div_id', 'total_bh_copay_pmpm_cost_t_9-6-3m_b4',
                  'bh_ip_snf_net_paid_pmpm_cost_3to6m_b4', 'mcc_an0_pmpm_ct_t_9-6-3m_b4',
                  'auth_3mth_post_acute_trm', 'rx_maint_pmpm_cost_t_12-9-6m_b4']

for col in list1:
    testdataframe[col] = testdataframe[col].astype(str)

for col in list1:
    testdataframe[col] = LabelEncoder().fit_transform(testdataframe[col])
```

```
In [41]: list2 = ['rx_phar_cat_cvs_pmpm_ct_t_9-6-3m_b4', 'auth_3mth_post_er',
                  'total_physician_office_visit_ct_pmpm_t_6-3-0m_b4',
                  'rx_gpi2_33_pmpm_ct_0to3m_b4', 'auth_3mth_post_acute_chf',
                  'auth_3mth_dc_other', 'auth_3mth_bh_acute_mean_los',
                  'auth_3mth_post_acute_gus', 'auth_3mth_acute_mus',
                  'rx_generic_dist_gpi6_pmpm_ct_t_9-6-3m_b4']

for col in list2:
    testdataframe[col] = testdataframe[col].astype(str)

for col in list2:
    testdataframe[col] = LabelEncoder().fit_transform(testdataframe[col])
```

```
In [42]: #categorical Columns
categ_cols = testdataframe.select_dtypes(include='object')
categ_cols.columns
```

```
Out[42]: Index(['rx_gpi4_6110_pmpm_ct', 'rx_nonbh_pmpm_cost_t_9-6-3m_b4',
                'auth_3mth_acute_vco', 'rx_bh_pmpm_ct_0to3m_b4', 'auth_3mth_dc_ltac',
                'auth_3mth_post_acute_inj', 'auth_3mth_dc_home',
                'rx_gpi2_17_pmpm_cost_t_12-9-6m_b4', 'rx_generic_pmpm_cost_t_6-3-0m_b4',
                'rx_overall_mbr_resp_pmpm_cost_t_6-3-0m_b4',
                'bh_ip_snf_mbr_resp_pmpm_cost_6to9m_b4',
                'rx_overall_dist_gpi6_pmpm_ct_t_6-3-0m_b4', 'auth_3mth_dc_no_ref',
                'auth_3mth_dc_snf', 'rx_phar_cat_humana_pmpm_ct_t_9-6-3m_b4',
                'bh_ip_snf_net_paid_pmpm_cost_0to3m_b4', 'auth_3mth_psychic',
                'auth_3mth_bh_acute', 'rx_overall_gpi_pmpm_ct_t_6-3-0m_b4',
                'mcc_chf_pmpm_ct_t_9-6-3m_b4',
                'bh_urgent_care_copay_pmpm_cost_t_12-9-6m_b4', 'auth_3mth_acute_bld',
                'auth_3mth_acute_skn', 'rx_gpi2_34_dist_gpi6_pmpm_ct',
                'rx_maint_pmpm_cost_t_6-3-0m_b4', 'cons_mobplus',
                'lab_albumin_loinc_pmpm_ct', 'rx_maint_net_paid_pmpm_cost_t_12-9-6m_b4',
                'rej_med_outpatient_visit_ct_pmpm_t_6-3-0m_b4',
                'rej_med_ip_snf_coins_pmpm_cost_t_9-6-3m_b4',
                'rx_gpi2_72_pmpm_ct_6to9m_b4',
                'med_physician_office_allowed_pmpm_cost_t_9-6-3m_b4',
                'auth_3mth_acute_dig', 'auth_3mth_dc_acute_rehab',
                'bh_ip_snf_mbr_resp_pmpm_cost_3to6m_b4',
                'total_physician_office_net_paid_pmpm_cost_t_9-6-3m_b4',
                'rx_branded_pmpm_ct_t_6-3-0m_b4',
                'med_outpatient_deduct_pmpm_cost_t_9-6-3m_b4', 'auth_3mth_non_er',
                'total_allowed_pmpm_cost_t_9-6-3m_b4', 'mabh_seg',
                'bh_ip_snf_mbr_resp_pmpm_cost_9to12m_b4', 'auth_3mth_post_acute_cer',
                'oontwk_mbr_resp_pmpm_cost_t_6-3-0m_b4', 'hum_region',
                'rx_nonmail_dist_gpi6_pmpm_ct_t_9-6-3m_b4',
                'bh_ip_snf_net_paid_pmpm_cost_6to9m_b4',
                'rej_med_er_net_paid_pmpm_cost_t_9-6-3m_b4',
                'med_outpatient_mbr_resp_pmpm_cost_t_9-6-3m_b4',
                'rx_nonbh_net_paid_pmpm_cost_t_6-3-0m_b4',
                'rx_gpi2_39_pmpm_cost_t_6-3-0m_b4',
                'total_ip_maternity_net_paid_pmpm_cost_t_12-9-6m_b4',
                'auth_3mth_acute_can', 'auth_3mth_post_acute', 'auth_3mth_facility',
                'rx_maint_pmpm_ct_t_6-3-0m_b4',
                'rx_mail_net_paid_pmpm_cost_t_6-3-0m_b4', 'auth_3mth_home',
                'total_physician_office_mbr_resp_pmpm_cost_t_9-6-3m_b4',
                'rev_cms_ansth_pmpm_ct', 'rx_mail_mbr_resp_pmpm_cost_t_9-6-3m_b4',
                'med_outpatient_visit_ct_pmpm_t_12-9-6m_b4',
```

```
'rx_nonbh_pmpm_ct_t_9-6-3m_b4', 'auth_3mth_acute',
'auth_3mth_acute_ccs_227', 'total_med_net_paid_pmpm_cost_t_6-3-0m_b4',
'rx_gpi2_90_dist_gpi6_pmpm_ct_9to12m_b4', 'sex_cd',
'rx_gpi2_62_pmpm_cost_t_9-6-3m_b4',
'rx_overall_gpi_pmpm_ct_t_12-9-6m_b4', 'auth_3mth_ltac', 'cons_hhcomp',
'rx_nonmaint_dist_gpi6_pmpm_ct_t_12-9-6m_b4', 'auth_3mth_snf_post_hsp',
'auth_3mth_acute_cer', 'rx_nonotc_pmpm_cost_t_6-3-0m_b4',
'auth_3mth_acute_trm',
'rej_total_physician_office_visit_ct_pmpm_0to3m_b4',
'auth_3mth_dc_home_health', 'rx_gpi2_56_dist_gpi6_pmpm_ct_3to6m_b4',
'auth_3mth_acute_cir', 'auth_3mth_acute_ner',
'med_ambulance_coins_pmpm_cost_t_9-6-3m_b4', 'hedis_dia_hba1c_ge9',
'ccsp_065_pmpm_ct', 'rx_maint_net_paid_pmpm_cost_t_9-6-3m_b4',
'bh_ip_snf_admit_days_pmpm_t_9-6-3m_b4', 'auth_3mth_acute_sns'],
dtype='object')
```

In [43]:

```
list3 = ['rx_gpi4_6110_pmpm_ct', 'rx_nonbh_pmpm_cost_t_9-6-3m_b4',
'auth_3mth_acute_vco', 'rx_bh_pmpm_ct_0to3m_b4', 'auth_3mth_dc_ltac',
'auth_3mth_post_acute_inj', 'auth_3mth_dc_home',
'rx_gpi2_17_pmpm_cost_t_12-9-6m_b4', 'rx_generic_pmpm_cost_t_6-3-0m_b4',
'rx_overall_mbr_resp_pmpm_cost_t_6-3-0m_b4',
'bh_ip_snf_mbr_resp_pmpm_cost_6to9m_b4',
'rx_overall_dist_gpi6_pmpm_ct_t_6-3-0m_b4', 'auth_3mth_dc_no_ref',
'auth_3mth_dc_snf', 'rx_phar_cat_humana_pmpm_ct_t_9-6-3m_b4',
'bh_ip_snf_net_paid_pmpm_cost_0to3m_b4', 'auth_3mth_psychic',
'auth_3mth_bh_acute', 'auth_3mth_acute_chf',
'rx_overall_gpi_pmpm_ct_t_6-3-0m_b4', 'mcc_chf_pmpm_ct_t_9-6-3m_b4',
'bh_urgent_care_copay_pmpm_cost_t_12-9-6m_b4', 'auth_3mth_acute_bld',
'rx_gpi2_34_dist_gpi6_pmpm_ct', 'rx_maint_pmpm_cost_t_6-3-0m_b4',
'cons_mobplus', 'lab_albumin_loinc_pmpm_ct',
'rx_maint_net_paid_pmpm_cost_t_12-9-6m_b4',
'rej_med_outpatient_visit_ct_pmpm_t_6-3-0m_b4',
'rej_med_ip_snf_coins_pmpm_cost_t_9-6-3m_b4',
'rx_gpi2_72_pmpm_ct_6to9m_b4',
'med_physician_office_allowed_pmpm_cost_t_9-6-3m_b4',
'auth_3mth_acute_res', 'auth_3mth_acute_dig',
'auth_3mth_dc_acute_rehab', 'bh_ip_snf_mbr_resp_pmpm_cost_3to6m_b4',
'total_physician_office_net_paid_pmpm_cost_t_9-6-3m_b4',
'rx_branded_pmpm_ct_t_6-3-0m_b4',
'med_outpatient_deduct_pmpm_cost_t_9-6-3m_b4', 'auth_3mth_non_er',
'total_allowed_pmpm_cost_t_9-6-3m_b4', 'mabh_seg',]

for col in list3:
    testdataframe[col] = testdataframe[col].astype(str)

for col in list3:
    testdataframe[col] = LabelEncoder().fit_transform(testdataframe[col])
```

In [44]:

```
list4 = ['bh_ip_snf_mbr_resp_pmpm_cost_9to12m_b4', 'auth_3mth_post_acute_cer',
'oontwk_mbr_resp_pmpm_cost_t_6-3-0m_b4', 'auth_3mth_post_acute_mus',
'hum_region', 'rx_nonmail_dist_gpi6_pmpm_ct_t_9-6-3m_b4',
'bh_ip_snf_net_paid_pmpm_cost_6to9m_b4',
'rej_med_er_net_paid_pmpm_cost_t_9-6-3m_b4',
'med_outpatient_mbr_resp_pmpm_cost_t_9-6-3m_b4',
'rx_nonbh_net_paid_pmpm_cost_t_6-3-0m_b4', 'auth_3mth_post_acute_sns',
'rx_gpi2_39_pmpm_cost_t_6-3-0m_b4',
'total_ip_maternity_net_paid_pmpm_cost_t_12-9-6m_b4',
'auth_3mth_acute_can', 'auth_3mth_post_acute', 'auth_3mth_facility',
'rx_maint_pmpm_ct_t_6-3-0m_b4', 'auth_3mth_post_acute_men',
'rx_mail_net_paid_pmpm_cost_t_6-3-0m_b4', 'auth_3mth_home',
'total_physician_office_mbr_resp_pmpm_cost_t_9-6-3m_b4',
'auth_3mth_transplant', 'rev_cms_ansth_pmpm_ct',
'rx_mail_mbr_resp_pmpm_cost_t_9-6-3m_b4',
'med_outpatient_visit_ct_pmpm_t_12-9-6m_b4',
'rx_nonbh_pmpm_ct_t_9-6-3m_b4', 'auth_3mth_acute',
'auth_3mth_dc_left_ama', 'auth_3mth_acute_ccs_227',
'auth_3mth_dc_custodial', 'total_med_net_paid_pmpm_cost_t_6-3-0m_b4',
'rx_gpi2_90_dist_gpi6_pmpm_ct_9to12m_b4', 'sex_cd',
'rx_gpi2_62_pmpm_cost_t_9-6-3m_b4',
'rx_overall_gpi_pmpm_ct_t_12-9-6m_b4', 'auth_3mth_ltac', 'cons_hhcomp',
'rx_nonmaint_dist_gpi6_pmpm_ct_t_12-9-6m_b4', 'auth_3mth_snf_post_hsp',
'rx_nonotc_pmpm_cost_t_6-3-0m_b4', 'auth_3mth_acute_trm',
'rej_total_physician_office_visit_ct_pmpm_0to3m_b4',
'auth_3mth_snf_direct', 'auth_3mth_dc_home_health',
'rx_gpi2_56_dist_gpi6_pmpm_ct_3to6m_b4', 'auth_3mth_acute_ner',
'med_ambulance_coins_pmpm_cost_t_9-6-3m_b4', 'hedis_dia_hba1c_ge9',
'ccsp_065_pmpm_ct', 'rx_maint_net_paid_pmpm_cost_t_9-6-3m_b4',
'bh_ip_snf_admit_days_pmpm_t_9-6-3m_b4']

for col in list4:
    testdataframe[col] = testdataframe[col].astype(str)

for col in list4:
    testdataframe[col] = LabelEncoder().fit_transform(testdataframe[col])
```



```
In [ ]:
```

```
In [50]: categ_cols = testdataframe.select_dtypes(include='object')
         categ_cols.columns
```

```
Out[50]: Index([], dtype='object')
```

```
In [49]: list5 = ['auth_3mth_acute_skn', 'auth_3mth_acute_cer', 'auth_3mth_acute_cir',
                 'auth_3mth_acute_sns']

         for col in list5:
             testdataframe[col] = testdataframe[col].astype(str)

         for col in list5:
             testdataframe[col] = LabelEncoder().fit_transform(testdataframe[col])
```

```
In [51]: testdataframe.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 525158 entries, 0 to 525157
Columns: 366 entries, Unnamed: 0 to race_cd
dtypes: float64(202), int32(117), int64(47)
memory usage: 1.2 GB
```

```
In [52]: testdataframe.head()
         #giving training and test data
```

```
Out[52]:
```

	Unnamed: 0	ID	auth_3mth_post_acute_dia	rx_gpi2_72_pmpm_cost_6to9m_b4	atlas_pct_laccess_child15	atlas_recfacpth14	atlas_pct_fmrrt
0	0	121252	1	0.0	9.735346	0.131562	
1	1	11804	1	0.0	6.264524	0.075427	
2	2	24182	1	0.0	6.136065	0.106704	
3	3	268083	1	0.0	1.244630	0.000000	
4	4	92461	1	0.0	8.009543	0.061276	

5 rows × 366 columns

```
In [53]: from sklearn.preprocessing import MinMaxScaler
         df_std = MinMaxScaler().fit_transform(testdataframe)
         df_new = pd.DataFrame(df_std, columns=testdataframe.columns)
         df_new.head()
```

```
Out[53]:
```

	Unnamed: 0	ID	auth_3mth_post_acute_dia	rx_gpi2_72_pmpm_cost_6to9m_b4	atlas_pct_laccess_child15	atlas_recfacpth14	atlas_pct_fmrrt
0	0.000000	0.230887	1.0	0.0	0.312471	0.215479	
1	0.000002	0.022477	1.0	0.0	0.201069	0.123538	
2	0.000004	0.046047	1.0	0.0	0.196946	0.174766	
3	0.000006	0.510482	1.0	0.0	0.039948	0.000000	
4	0.000008	0.176064	1.0	0.0	0.257079	0.100361	

5 rows × 366 columns

```
In [54]: df_new.to_csv("preprocessed_holdout.csv")
```

```
In [ ]:
```