Business Analysis Report

Shalini Krishnan | Exploratory Data Analysis | 06-03-2022

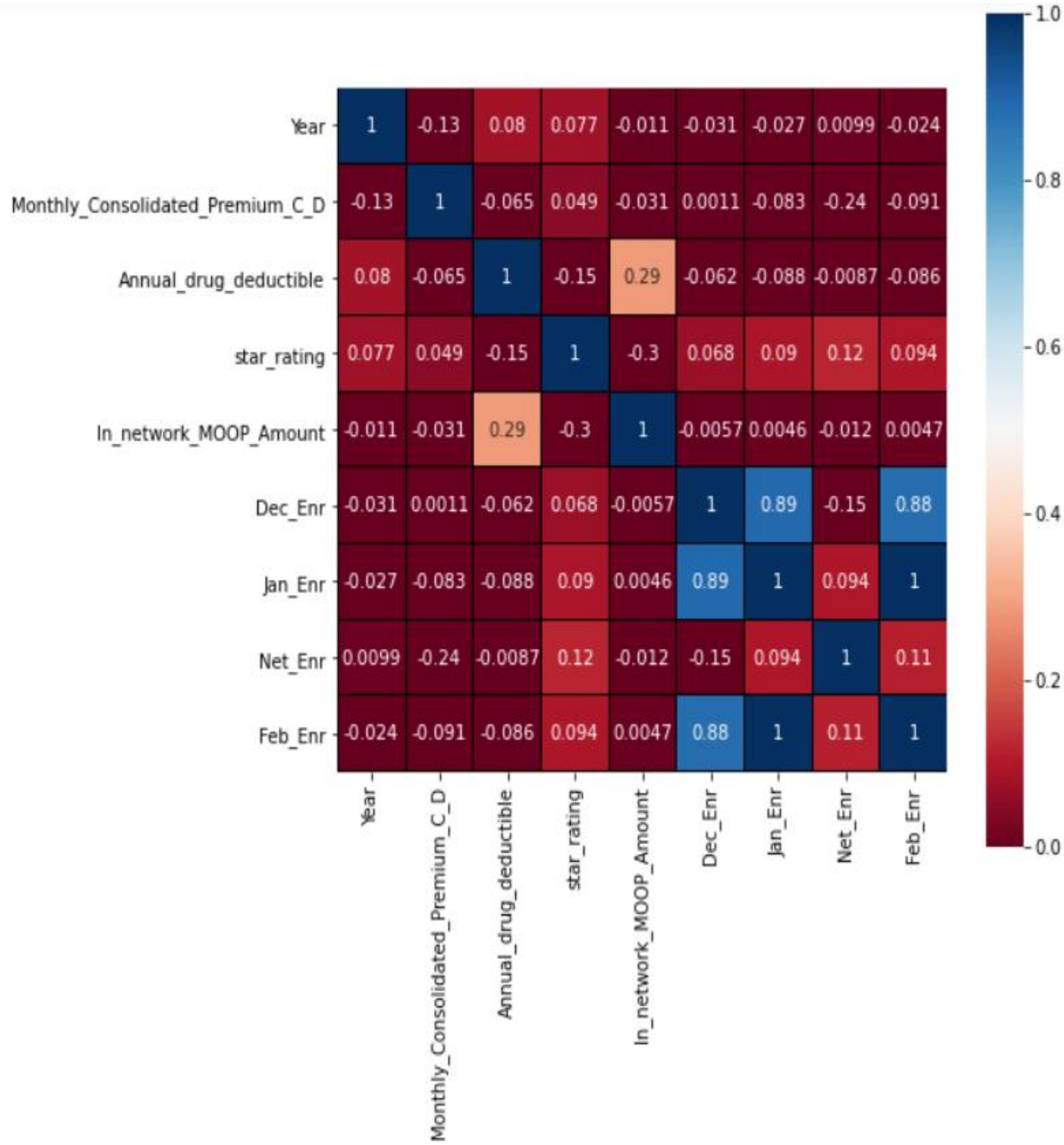
Descriptive Statistics

- One of the major steps to fine-tune the given data set in a different form of analysis to understand the insights of the key characteristics of various entities of the data set like column, rows by applying Statistical Methods and Data visualization packages.
- The given dataset is of **218960** rows and **70** columns. It contains discrete, categorical and continuous variables.
- On further analysis, it is clear that the data set doesn't have any anomalies like missing values, outliers and duplicates. (Refer source code)
- The following table shows the descriptive statistics of the numerical values.

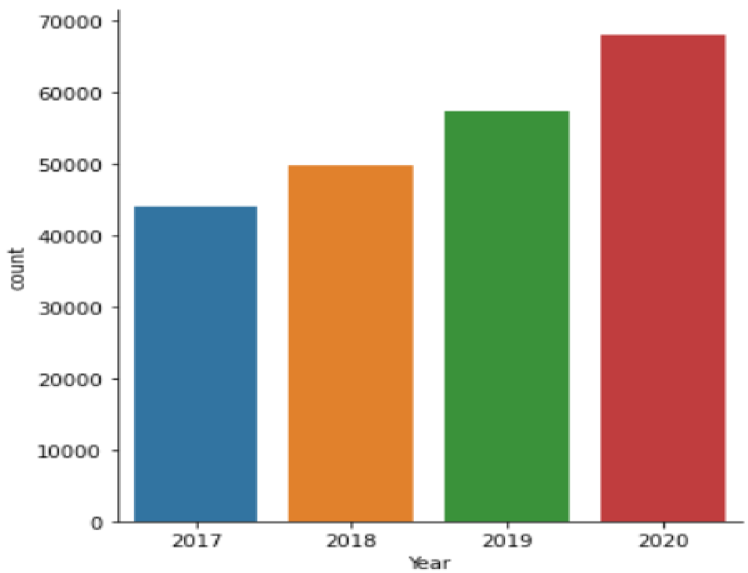
	count	mean	std	min	25 %	50%	75%	max
Year	218960.0	2018.681111	1.113882	2017.0	2018.0	2019.0	2020.0	2020.0
Monthly_Consolidated_Premium_C_D	218960.0	43.703227	54.597898	0.0	0.0	27.9	65.0	375.9
Annual_drug_deductible	178830.0	211.847654	170.237253	0.0	0.0	200.0	400.0	435.0
star_rating	195538.0	3.832176	0.509887	2.0	3.5	4.0	4.0	5.0
In_network_MOOP_Amount	203748.0	5418.554926	1495.350041	0.0	3950.0	6050.0	6700.0	6700.0
Dec Enr	174912.0	289.192497	1521.109356	0.0	0.0	27.0	141.0	168835.0
Jan Enr	218960.0	312.110833	1544.444934	0.0	0.0	39.0	166.0	169142.0
Net Enr	174912.0	16.153289	186.877726	-10040.0	-2.0	0.0	11.0	13394.0

	count	mean	std	min	25 %	50%	75%	max
feb enr	218805.0	315.593844	1551.552656	0.0	0.0	41.0	169.0	169903.0

- There is a high positive correlation between January Enrollment, December Enroll ment and February Enrollment.



Univariate Analysis



- The frequency count of the year 2020 is 31%, 2019 is 26%, 2018 is 22% and 2017 is 20%. The plans keep on incrementing year over year. Over all the years 2018, 2019 and 2020 shows a steady increase in plans where 2020 is in the highest position, which shows that more people are inclined towards enrolling into the plan. It proves the market sentiment is heading towards positive direction.
- Among the States, the following is the statistic of top 5 states which shows higher values of enrollment.

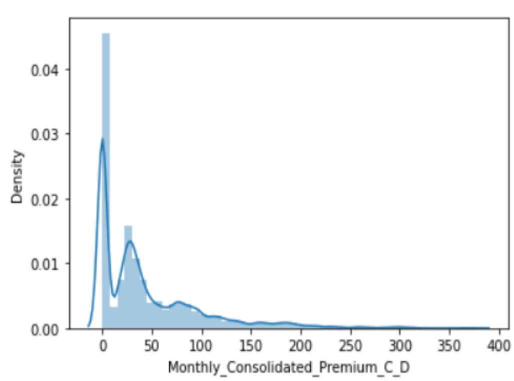
States	Percentage of Plans Enrolled
Texas	7%
Puerto Rico	6%
Ohio	5.7%
Georgia	5.4%
Pennsylvania	5%

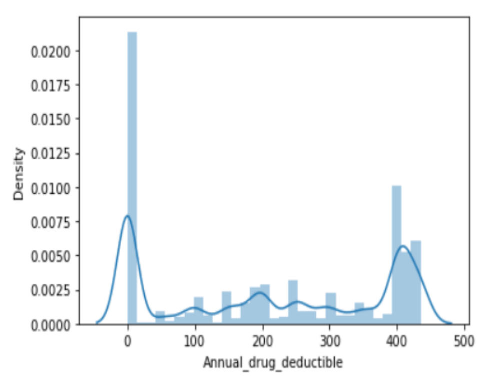
- Among the Counties, the first 5 which has the highest frequency of plans is shown below.

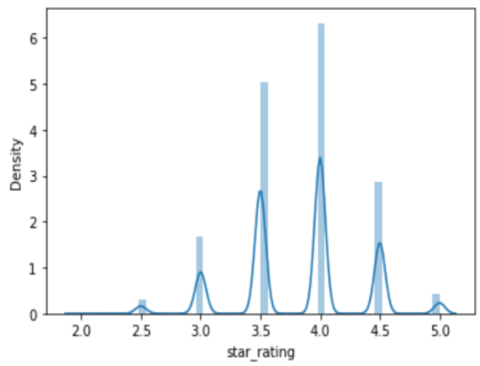
Counties	Percentage of Plans Enrolled
Washington	0.9%
Jefferson	0.9%
Franklin	0.7%
Montgomery	0.7%
Jackson	0.6%

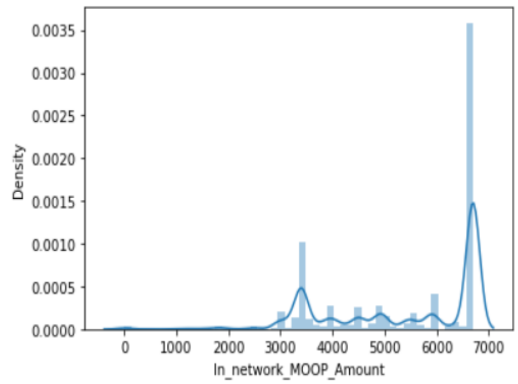
- On Organization level, the below table furnishes the top 5 positions whose counts of plans are actively more.

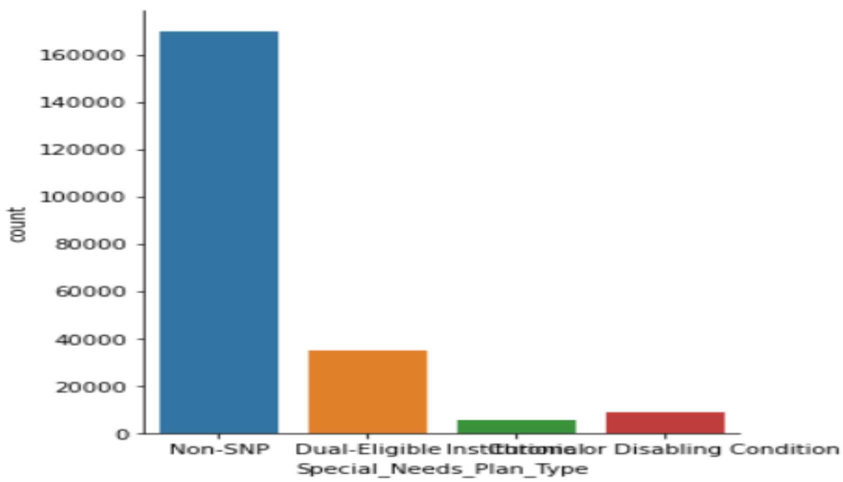
Organization	Percentage of Plans Enrolled
Humana Inc.	25 %
UnitedHealth Group, Inc.	15%
CVS Health Corporation	6%
WellCare Health Plans, Inc.	4%
Anthem Inc.	4%











- The Monthly Consolidated Premium ranges between an average of 43 dollars to maximum 375 dollars per month. Also, Annual Drug deductible ranges between an average of 211 dollars to maximum 435 dollars per annum.

- As the above graph depicts ,37% people had given 4-star ratings to the plans and only 2% people had given full 5-star ratings, which means somewhere in the borderline, deep analysis is needed on what goes missing with the plans to impress the customers.

Star Ratings	People %
4.0	37%
3.5	30%
4.5	17%
3.0	10%
5.0	2%
2.5	1.8%
2.0	0.2%

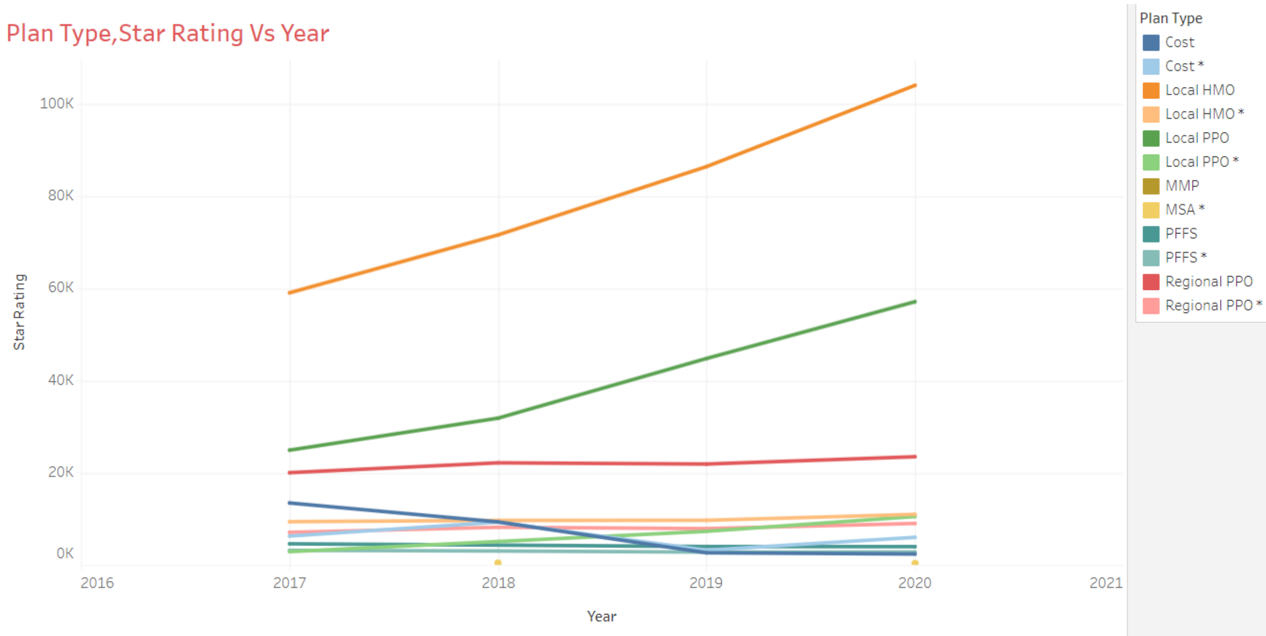
- The Annual MOOP Value ranges between an average of 5418 dollars to maximum 6700 dollars.

  - Overall, 77% of people had availed Non-SNP plan type, whereas 16% for Dual Eligible, 3% for Chronic or Disabling Condition and only 2% for Institutional type. It is clearly shown in the above bar plot.

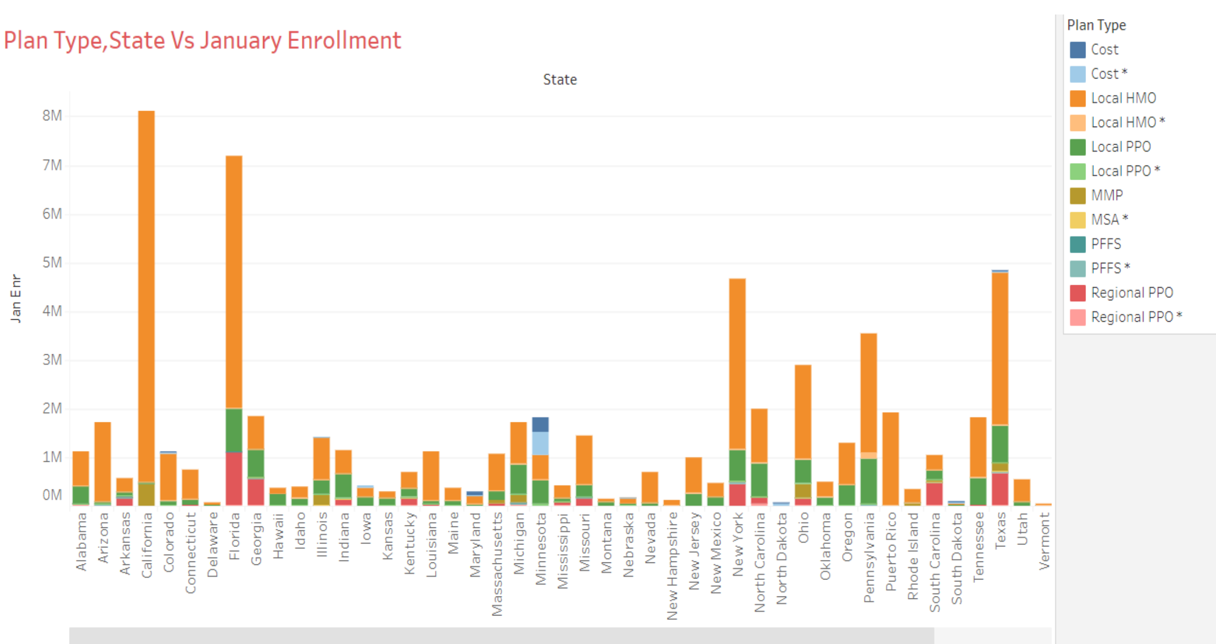
  - The following table shows the type of plan and percentage enrolled for the same. According to the data, Local HMO seems to be in the highest position.

Plan Type	Percentage Enrolled
Local HMO	42%
Local PPO	21%
Regional PPO	11%
Local HMO *	4.6%
Regional PPO *	4.3%
Local PPO *	3%
Cost	2.9%
Cost *	2.8%
PFFS	2.2%
MSA *	1.6%
PFFS *	1.6%
MMP	0.7%

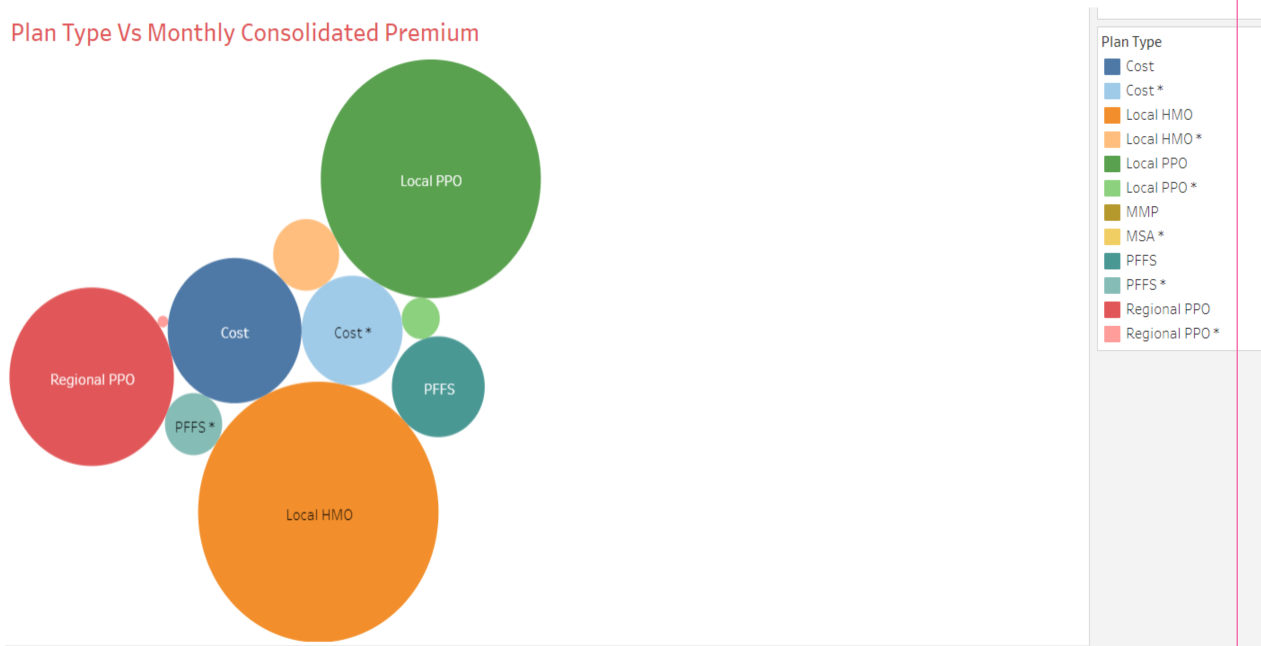
Bi-Variate & Multivariate Analysis



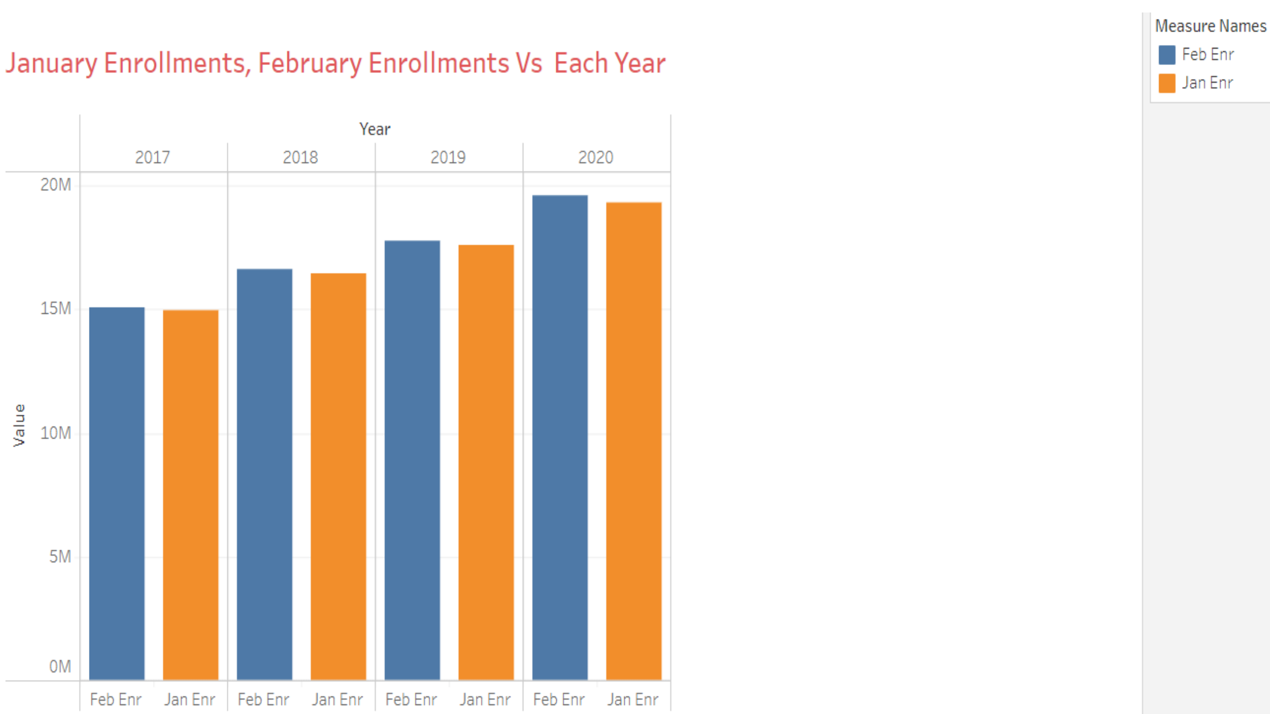
- Local HMO Plans shows a steady increase in star ratings from the year 2017 to 2020 and is the topmost rated plan
- More or less the other plans too converge closely up until 2020.But Cost Plan type shows a heavy decrease after the year 2018 and constantly declines thereafter. Further inspection is needed to figure out the fall of ratings in Cost Plan type.



- Though Texas shows the high percentage of plans enrolled (7%), with respect to January Enrollment, California tops the other States with the high percentage for Local HMO plan type.

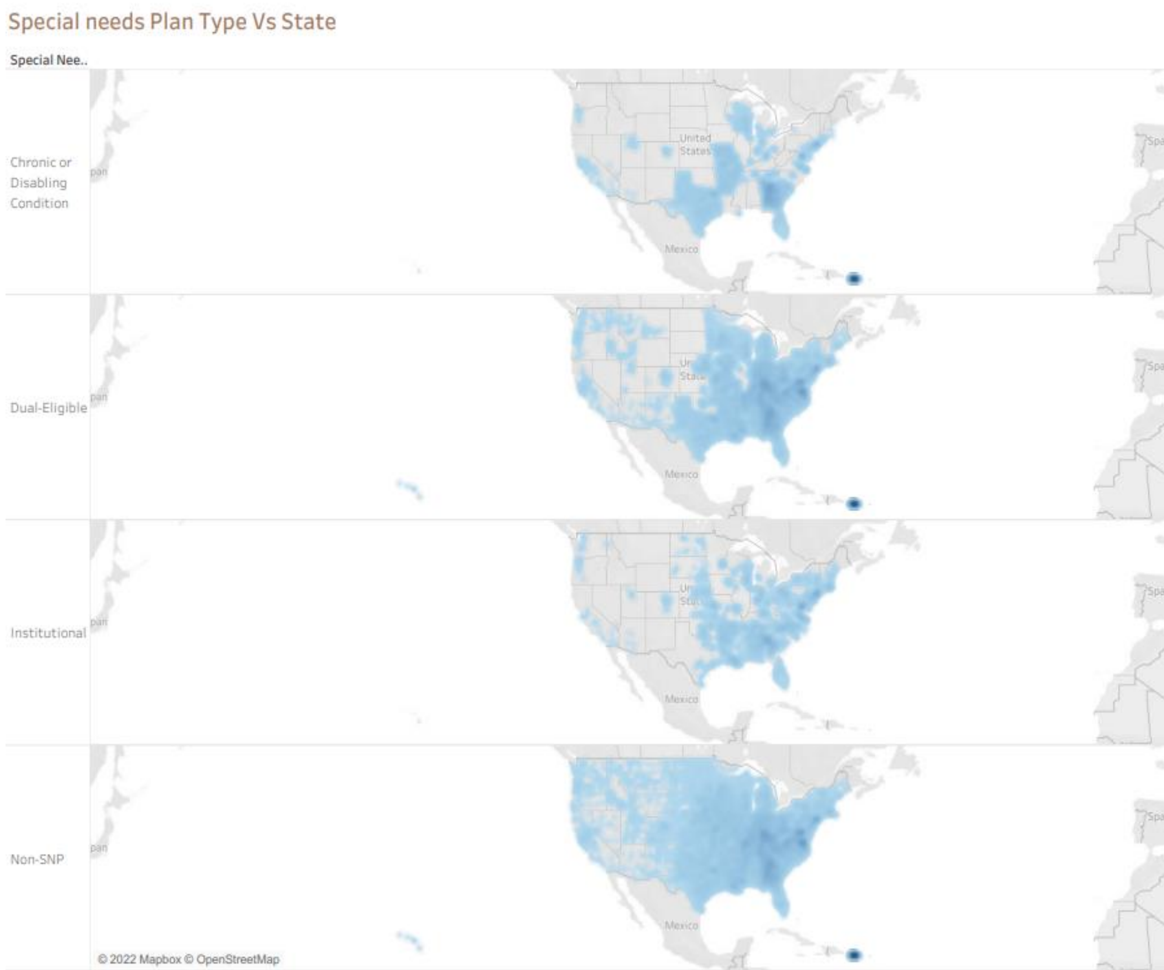


- Monthly Consolidated Premium is high for Local HMO plan type, followed by Local PPO plan type.



YEAR	JANUARY ENROLLMENT	FEBRUARY ENROLLMENT
2017	14,931,146	15,061,512
2018	16,470,904	16,602,525
2019	17,606,375	17,792,984
2020	19,331,363	19,596,390

- The enrollment trend graph seems to be steadily increasing from month over month (Jan- Feb) and also from year over year (2017-2020). More number of beneficiaries are adding up year over year.



- From the above geoplot it is evident that the beneficiaries count percentage is more for Non-SNP Plan Types when compared to the other specific plans. The State of North Dakota and Williams County shows the maximum count of Non-SNP plan.

Business Recommendations

- Proper groundwork should be carried out to analyze why there is a sudden decrease in Cost Plan type Enrollment over the four-year interval. Necessary actions and ideal plan for transitioning it to be a WIN-WIN scenario should be in the track.

- Though there are few special needs plan types, beneficiaries are more pitched towards the non-SNP plan type. On par with non- SNP type, the other special types of plans can be reviewed and refactored in such a way that it too reaches more audience.

- The overall market sentiment seems to be moving towards the positive direction year over year, but close monitoring and analysis is needed to figure out for any anomalies in the future trend.

- On the other hand, marketing campaigns, referral program, video testimonial share, contests shall be organized to promote the Plan types which are declining from its original position, to capture larger market area.

- When needs arise, new Partner branding can help in a long way to create hype over the plans, but in longer run it is only the elements of any plan which keeps itself intact with the beneficiary needs. Hence it is always better to design a plan in such a way it reaches more people.

- Hence, we sometimes need to analyze the other type of plans, the neighboring countries and other brands follow, and try to leverage ours from it. Constant upgradation of any plans with respect to the people sentiment analysis is always directly proportional to the highest success rate.

Source Code

```
# Importing the required Libraries #

import numpy as np

import pandas as pd

import seaborn as sn

import matplotlib as mpl

import matplotlib.pyplot as plt

import warnings

warnings.filterwarnings('ignore')

df=pd.read_excel("interview case study data.xlsx")

df.shape

df=df.rename(columns={"Dec Enr":"Dec_Enr"})

df=df.rename(columns={"Jan Enr":"Jan_Enr"})

df=df.rename(columns={"feb enr":"Feb_Enr"})

df=df.rename(columns={"Net Enr":"Net_Enr"})

df.info()

duplicate_rows=df[df.duplicated()]

print("Number of duplicate rows :  ",duplicate_rows.shape)

df.describe().T

unique_counts = pd.DataFrame.from_records([(col, df[col].nunique()) for col in
df.columns],

                        columns=['Column_Name',
'Num_Unique']).sort_values(by=['Num_Unique'])

unique_counts

def findoutliers(column):
```

```
    outliers=[]
    Q1=column.quantile(.25)
    Q3=column.quantile(.75)
    IQR=Q3-Q1
    lower_limit=Q1-(1.5*IQR)
    upper_limit=Q3+(1.5*IQR)
    for outlier in column:
        if outlier>upper_limit or outlier<lower_limit:
            outliers.append(outlier)
        return outliers
# Finding Missing Values #

df.isnull().sum()
corr= df.corr(method='spearman')
plt.figure(figsize=(8,8))
sn.heatmap(corr,vmin=0.0,vmax=1.0,linewidths=0.01,square=True,annot=True,cmap='RdBu',linecolor='black')
# Univariate Analysis #
 # Year #
df.boxplot(column=['Year'])
df.Year.value_counts()
df.Year.value_counts(normalize=True)
sn.factorplot('Year',data= df,kind='count')
df.Year.unique()
# State #
df.State.value_counts()
```

```
df.State.value_counts(normalize=True)

sn.factorplot('State',data= df,kind='count')

my_tab = pd.crosstab(index = df["State"],  # Make a crosstab

                     columns="count")      # Name the count column


my_tab.plot.bar()

df.State.unique()

# County #

df.County.value_counts()

df.County.value_counts(normalize=True)

sn.factorplot('County',data= df,kind='count')

df.County.unique()

# Bid_ID #

df.Bid_ID.value_counts()

df.Bid_ID.value_counts(normalize=True)

sn.factorplot('Bid_ID',data= df,kind='count')

my_tab = pd.crosstab(index = df["Bid_ID"],  # Make a crosstab

                     columns="count")      # Name the count column


my_tab.plot.bar()

df.Bid_ID.unique()

# Plan name #

df.plan_name.value_counts()

df.plan_name.value_counts(normalize=True)

sn.factorplot('plan_name',data= df,kind='count')

plt.scatter(df[:,0], df[:,1], c=labels, cmap = 'viridis')
```

```
plt.show()

df.plan_name.unique()

# Organization name #

df.Organization_Name.value_counts()

df.Organization_Name.value_counts(normalize=True)

sn.factorplot('Organization_Name',data= df,kind='count')

df.Organization_Name.unique()

# Monthly consolidated Premium_C_D #

df.Monthly_Consolidated_Premium_C_D.value_counts()

df.Monthly_Consolidated_Premium_C_D.value_counts(normalize=True)

df.Monthly_Consolidated_Premium_C_D.unique()

import seaborn as sns

sns.distplot(df["Monthly_Consolidated_Premium_C_D"],hist=True)

# Annual drug Deductible #

df.Annual_drug_deductible.value_counts()

df.Annual_drug_deductible.value_counts(normalize=True)

sns.distplot(df["Annual_drug_deductible"],hist=True)

# Star Rating #

df.star_rating.value_counts()

df.star_rating.value_counts(normalize=True)

sns.distplot(df["star_rating"],hist=True)

# In Network MOOP Amount #

df.In_network_MOOP_Amount.value_counts()

df.In_network_MOOP_Amount.value_counts(normalize=True)

sns.distplot(df["In_network_MOOP_Amount"],hist=True)
```

```
# Special Needs Plan Type #

df.Special_Needs_Plan_Type.value_counts()
```

```
Non-SNP                              169839
Dual-Eligible                         35158
Chronic or Disabling Condition         8691
Institutional                          5272
Name: Special_Needs_Plan_Type, dtype: int64
```

```
df.Special_Needs_Plan_Type.value_counts(normalize=True)

sn.factorplot('Special_Needs_Plan_Type',data= df,kind='count')

# Plan Type #

df.plan_type.value_counts()

df.plan_type.value_counts(normalize=True)

sn.factorplot('plan_type',data= df,kind='count')

# December Enrollment #

df.Dec_Enr.value_counts()

df.Dec_Enr.value_counts(normalize=True)

df.Dec_Enr.unique()

sns.distplot(df["Dec_Enr"],hist=True)

# January Enrollments #

df.Jan_Enr.value_counts()

df.Jan_Enr.value_counts(normalize=True)

df.Jan_Enr.unique()

sns.distplot(df["Jan_Enr"],hist=True)

# February Enrollments #

df.Feb_Enr.value_counts()

df.Feb_Enr.value_counts(normalize=True)

df.Feb_Enr.unique()

sns.distplot(df["Feb_Enr"],hist=True)
```

```
# Net Enrollments #
df.Net_Enr.value_counts()
df.Net_Enr.value_counts(normalize=True)
df.Net_Enr.unique()
sns.distplot(df["Net_Enr"],hist=True)
# Bi Variate Analysis #
# Pattern between Year and Plan name #
plt.scatter(df.Year, df.plan_name)
plt.title('Year Vs. Plan Name')
plt.xlabel('Year')
plt.ylabel('plan_name')
df.boxplot(column="Year",        # Column to plot
           by= "plan_name",         # Column to split upon
           figsize= (8,8))        # Figure size
grouped = df.groupby(['Year','plan_name'])
grouped.size()
table = pd.crosstab(index=df["Year"],
                    columns=df["plan_name"])

table
table.plot(kind="bar",
           figsize=(8,8),
           stacked=True)
```