Healthcare Unsupervised Anomaly Detection

By ⇨ *Pushpendra Kumar*

Overview

Healthcare fraud diverts crucial funding from vital areas such as medication and emergency services, driven by both practitioners and patients. To combat this issue, I have initiated a project focused on healthcare data anomaly detection. Through various unsupervised techniques, we aim to uncover irregularities indicating health concerns or fraudulent behavior.

Extensive Exploratory Data Analysis (EDA) helps reveal hidden anomalies and their implications. By leveraging advanced data visualization and analytics, we can extract insights for early detection and prevention. Our ultimate goal is to develop effective anomaly detection models, which are crucial for safeguarding patient health and optimizing healthcare operations.

Importing Libraries

In [3]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline

#filtering the warnings
import warnings
warnings.filterwarnings("ignore")
```

Loading the dataset

In [25]:
```
df = pd.read_csv('Healthcare Providers.csv')
```

In [26]:
```
# Shape of the data
df.shape
```

Out[26]:
```
(100000, 27)
```

In [27]:
```
# First 5 rows
df.head()
```

Out[27]:

	index	National Provider Identifier	Last Name/Organization Name of the Provider	First Name of the Provider	Middle Initial of the Provider	Credentials of the Provider	Gender of the Provider	Entity Type of the Provider	Street Address 1 of the Provider	Street Address 2 of the Provider	...	HCPCS Code	
0	8774979	1891106191	UPADHYAYULA	SATYASREE	NaN	M.D.	F	I	1402 S GRAND BLVD	FDT 14TH FLOOR	...	99223	Ir in ty
1	3354385	1346202256	JONES	WENDY	P	M.D.	F	I	2950 VILLAGE DR	NaN	...	G0202	ma
2	3001884	1306820956	DUROCHER	RICHARD	W	DPM	M	I	20 WASHINGTON AVE	STE 212	...	99348	p v
3	7594822	1770523540	FULLARD	JASPER	NaN	MD	M	I	5746 N BROADWAY ST	NaN	...	81002	
4	746159	1073627758	PERROTTI	ANTHONY	E	DO	M	I	875 MILITARY TRL	SUITE 200	...	96372	r

5 rows × 27 columns

In [28]:
```
# Last 5 rows
df.tail()
```

Out[28]:

	index	National Provider Identifier	Last Name/Organization Name of the Provider	First Name of the Provider	Middle Initial of the Provider	Credentials of the Provider	Gender of the Provider	Entity Type of the Provider	Street Address 1 of the Provider	Street Address 2 of the Provider	...	HCPCS Code	
99995	3837311	1386938868	PAPES	JOAN	NaN	PT	F	I	324 E BALTIMORE ST	NaN	...	97162	I
99996	2079360	1215091327	HAYNER	MARGARET	S	ARNP	F	I	645 NW 4TH ST	NaN	...	99213	
99997	8927965	1902868185	VALENCIA	DANA	NaN	M.D.	M	I	3009 N BALLAS RD	SUITE 202B	...	93320	s b
99998	8854571	1891941183	GONZALEZ-LAMOS	RAFAELA	NaN	NaN	F	I	2365 BOSTON POST RD	SUITE 201	...	G0008	A
99999	3547535	1356772156	RAMEZANI	ELIIAN	NaN	NaN	F	I	444 MIDDLE NECK RD APT 1H	NaN	...	97112	

5 rows × 27 columns

In [29]:
```
# Information regarding datset
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100000 entries, 0 to 99999
Data columns (total 27 columns):
 #   Column                                                Non-Null Count   Dtype
---  ------                                                --------------   -----
 0   index                                                 100000 non-null  int64
 1   National Provider Identifier                          100000 non-null  int64
 2   Last Name/Organization Name of the Provider           100000 non-null  object
 3   First Name of the Provider                            95745 non-null   object
 4   Middle Initial of the Provider                        70669 non-null   object
 5   Credentials of the Provider                           92791 non-null   object
 6   Gender of the Provider                                95746 non-null   object
 7   Entity Type of the Provider                           100000 non-null  object
 8   Street Address 1 of the Provider                      100000 non-null  object
 9   Street Address 2 of the Provider                      40637 non-null   object
 10  City of the Provider                                  100000 non-null  object
 11  Zip Code of the Provider                              100000 non-null  float64
 12  State Code of the Provider                            100000 non-null  object
 13  Country Code of the Provider                          100000 non-null  object
 14  Provider Type                                         100000 non-null  object
 15  Medicare Participation Indicator                      100000 non-null  object
 16  Place of Service                                      100000 non-null  object
 17  HCPCS Code                                            100000 non-null  object
 18  HCPCS Description                                     100000 non-null  object
 19  HCPCS Drug Indicator                                  100000 non-null  object
 20  Number of Services                                    100000 non-null  object
 21  Number of Medicare Beneficiaries                      100000 non-null  object
 22  Number of Distinct Medicare Beneficiary/Per Day Services  100000 non-null  object
 23  Average Medicare Allowed Amount                       100000 non-null  object
 24  Average Submitted Charge Amount                       100000 non-null  object
 25  Average Medicare Payment Amount                       100000 non-null  object
 26  Average Medicare Standardized Amount                  100000 non-null  object
dtypes: float64(1), int64(2), object(24)
memory usage: 20.6+ MB
```

In [30]:
```
# Columns of the dataset
df.columns
```

Out[30]:
```
Index(['index', 'National Provider Identifier',
       'Last Name/Organization Name of the Provider',
       'First Name of the Provider', 'Middle Initial of the Provider',
       'Credentials of the Provider', 'Gender of the Provider',
       'Entity Type of the Provider', 'Street Address 1 of the Provider',
       'Street Address 2 of the Provider', 'City of the Provider',
       'Zip Code of the Provider', 'State Code of the Provider',
       'Country Code of the Provider', 'Provider Type',
       'Medicare Participation Indicator', 'Place of Service', 'HCPCS Code',
       'HCPCS Description', 'HCPCS Drug Indicator', 'Number of Services',
       'Number of Medicare Beneficiaries',
       'Number of Distinct Medicare Beneficiary/Per Day Services',
       'Average Medicare Allowed Amount', 'Average Submitted Charge Amount',
       'Average Medicare Payment Amount',
       'Average Medicare Standardized Amount'],
      dtype='object')
```

In [31]:
```
# Statistical Description of dataset
```

In [31]:
```
# Statistical description of dataset
df.describe()
```

Out[31]:

	index	National Provider Identifier	Zip Code of the Provider
count	1.000000e+05	1.000000e+05	1.000000e+05
mean	4.907646e+06	1.498227e+09	4.163820e+08
std	2.839633e+06	2.874125e+08	3.082566e+08
min	2.090000e+02	1.003001e+09	6.010000e+02
25%	2.458791e+06	1.245669e+09	1.426300e+08
50%	4.901266e+06	1.497847e+09	3.633025e+08
75%	7.349450e+06	1.740374e+09	6.819881e+08
max	9.847440e+06	1.993000e+09	9.990166e+08

Exploratory Data Analysis (EDA)

Handling Missing Values

In [32]:
```
# Checking missing values
df.isnull().sum()
```

Out[32]:
```
index                                                       0
National Provider Identifier                                0
Last Name/Organization Name of the Provider                 0
First Name of the Provider                               4255
Middle Initial of the Provider                          29331
Credentials of the Provider                              7209
Gender of the Provider                                   4254
Entity Type of the Provider                                 0
Street Address 1 of the Provider                            0
Street Address 2 of the Provider                        59363
City of the Provider                                        0
Zip Code of the Provider                                    0
State Code of the Provider                                  0
Country Code of the Provider                                0
Provider Type                                               0
Medicare Participation Indicator                            0
Place of Service                                            0
HCPCS Code                                                  0
HCPCS Description                                           0
HCPCS Drug Indicator                                        0
Number of Services                                          0
Number of Medicare Beneficiaries                            0
Number of Distinct Medicare Beneficiary/Per Day Services    0
Average Medicare Allowed Amount                             0
Average Submitted Charge Amount                             0
Average Medicare Payment Amount                             0
Average Medicare Standardized Amount                        0
dtype: int64
```

Insights: Dropping the coloumns:
'index', 'National Provider Identifier', 'Last Name/Organization Name of the Provider', 'First Name of the Provider', 'Middle Initial of the Provider', 'Street Address 1 of the Provider', 'Street Address 2 of the Provider'. As these doesn't add much towards detecting anomaly. Meanwhile filling rest missing columnns: ('Credentials of the Provider','Gender of the Provider') with their mode values.

In [33]:
```
# Dropping columns
DropColumns = ['index', 'National Provider Identifier','Last Name/Organization Name of the Provider',
       'First Name of the Provider', 'Middle Initial of the Provider','Street Address 1 of the Provider',
       'Street Address 2 of the Provider']
df = df.drop(DropColumns, axis=1)
```

In [34]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100000 entries, 0 to 99999
Data columns (total 20 columns):
 #   Column                                                Non-Null Count   Dtype
---  ------                                                --------------   -----
 0   Credentials of the Provider                           92791 non-null   object
 1   Gender of the Provider                                95746 non-null   object
 2   Entity Type of the Provider                           100000 non-null  object
 3   City of the Provider                                  100000 non-null  object
 4   Zip Code of the Provider                              100000 non-null  float64
 5   State Code of the Provider                            100000 non-null  object
 6   Country Code of the Provider                          100000 non-null  object
 7   Provider Type                                         100000 non-null  object
 8   Medicare Participation Indicator                      100000 non-null  object
 9   Place of Service                                      100000 non-null  object
 10  HCPCS Code                                            100000 non-null  object
 11  HCPCS Description                                     100000 non-null  object
 12  HCPCS Drug Indicator                                  100000 non-null  object
 13  Number of Services                                    100000 non-null  object
 14  Number of Medicare Beneficiaries                      100000 non-null  object
 15  Number of Distinct Medicare Beneficiary/Per Day Services  100000 non-null  object
 16  Average Medicare Allowed Amount                       100000 non-null  object
 17  Average Submitted Charge Amount                       100000 non-null  object
 18  Average Medicare Payment Amount                       100000 non-null  object
 19  Average Medicare Standardized Amount                  100000 non-null  object
dtypes: float64(1), object(19)
memory usage: 15.3+ MB
```

In [35]:
```
# Filling missing with mode values
df["Credentials of the Provider"] = df["Credentials of the Provider"].fillna(df["Credentials of the Provider"].
df["Gender of the Provider"] = df["Gender of the Provider"].fillna(df["Gender of the Provider"].mode()[0])
```

In [36]:
```
# Checking missing values again
df.isnull().sum()
```

Out[36]:
```
Credentials of the Provider                           0
Gender of the Provider                                0
Entity Type of the Provider                           0
City of the Provider                                  0
Zip Code of the Provider                              0
State Code of the Provider                            0
Country Code of the Provider                          0
Provider Type                                         0
Medicare Participation Indicator                      0
Place of Service                                      0
HCPCS Code                                            0
HCPCS Description                                     0
HCPCS Drug Indicator                                  0
Number of Services                                    0
Number of Medicare Beneficiaries                      0
Number of Distinct Medicare Beneficiary/Per Day Services    0
Average Medicare Allowed Amount                       0
Average Submitted Charge Amount                       0
Average Medicare Payment Amount                       0
Average Medicare Standardized Amount                  0
dtype: int64
```

Insights: Converting these object columns which are meant to be numeric:
'Number of Services', 'Number of Medicare Beneficiaries', 'Number of Distinct Medicare Beneficiary/Per Day Services', 'Average Medicare Allowed Amount', 'Average Submitted Charge Amount', 'Average Medicare Payment Amount', 'Average Medicare Standardized Amount'.
Also removing any comma separators if they are present.

In [37]:
```
# Data Cleaning

def RemoveComma(x):
    return str(x).replace(",","")  # Convert to string before replacing comma

numericCols = ['Number of Services', 'Number of Medicare Beneficiaries',
               'Number of Distinct Medicare Beneficiary/Per Day Services',
               'Average Medicare Allowed Amount', 'Average Submitted Charge Amount',
               'Average Medicare Payment Amount', 'Average Medicare Standardized Amount']

df[numericCols] = df[numericCols].applymap(RemoveComma).apply(pd.to_numeric, errors='ignore')
```

In [38]:
```
#Checking information again
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100000 entries, 0 to 99999
Data columns (total 20 columns):
 #   Column                                                Non-Null Count   Dtype
---  ------                                                --------------   -----
 0   Credentials of the Provider                           100000 non-null  object
 1   Gender of the Provider                                100000 non-null  object
 2   Entity Type of the Provider                           100000 non-null  object
 3   City of the Provider                                  100000 non-null  object
 4   Zip Code of the Provider                              100000 non-null  float64
 5   State Code of the Provider                            100000 non-null  object
 6   Country Code of the Provider                          100000 non-null  object
 7   Provider Type                                         100000 non-null  object
 8   Medicare Participation Indicator                      100000 non-null  object
 9   Place of Service                                      100000 non-null  object
 10  HCPCS Code                                            100000 non-null  object
 11  HCPCS Description                                     100000 non-null  object
 12  HCPCS Drug Indicator                                  100000 non-null  object
 13  Number of Services                                    100000 non-null  float64
 14  Number of Medicare Beneficiaries                      100000 non-null  int64
 15  Number of Distinct Medicare Beneficiary/Per Day Services  100000 non-null  int64
 16  Average Medicare Allowed Amount                       100000 non-null  float64
 17  Average Submitted Charge Amount                       100000 non-null  float64
 18  Average Medicare Payment Amount                       100000 non-null  float64
 19  Average Medicare Standardized Amount                  100000 non-null  float64
dtypes: float64(6), int64(2), object(12)
memory usage: 15.3+ MB
```

Checking Duplicate Values

In [41]:
```
# Checking for duplicates
df.duplicated().any()
```

Out[41]:
```
True
```

In [43]:
```
# Finding duplicate rows in the entire DataFrame
duplicate_rows = df[df.duplicated()]
duplicate_rows
```

Out[43]:

	Credentials of the Provider	Gender of the Provider	Entity Type of the Provider	City of the Provider	Zip Code of the Provider	State Code of the Provider	Country Code of the Provider	Provider Type	Medicare Participation Indicator	Place of Service	HCPCS Code	HCPCS Description	HC Indi
2047	MD	M	I	TAMPA	336125513.0	FL	US	Diagnostic Radiology	Y	O	84520	Urea nitrogen level to assess kidney function	

In [44]:
```
# Dropping duplicates
df = df.drop_duplicates()
```

In [45]:
```
#Checking duplicates again
df.duplicated().any()
```

Out[45]:
```
False
```

In [46]:
```
df.head()
```

Out[46]:

	Credentials of the Provider	Gender of the Provider	Entity Type of the Provider	City of the Provider	Zip Code of the Provider	State Code of the Provider	Country Code of the Provider	Provider Type	Medicare Participation Indicator	Place of Service	HCPCS Code	HCPC Descripti
0	M.D.	F	I	SAINT LOUIS	631041004.0	MO	US	Internal Medicine	Y	F	99223	Initial hospi inpatient ca typically 70
1	M.D.	F	I	FAYETTEVILLE	283043815.0	NC	US	Obstetrics & Gynecology	Y	O	G0202	Screeni mammograpl bilateral view study
2	DPM	M	I	NORTH HAVEN	64732343.0	CT	US	Podiatry	Y	O	99348	Establish patient hor visit, typica 25 n
3	MD	M	I	KANSAS CITY	641183998.0	MO	US	Internal Medicine	Y	O	81002	Urinalys manual te
4	DO	M	I	JUPITER	334585700.0	FL	US	Internal Medicine	Y	O	96372	Injecti beneath t skin or ir muscle for

In [47]:
```
# Cleaning 'Credentials of the Provider' Column
# Removing periods '.' if present
df['Credentials of the Provider'] = df['Credentials of the Provider'].str.replace('.', '')
```

In [48]:
```
df.head()
```

Out[48]:

	Credentials of the Provider	Gender of the Provider	Entity Type of the Provider	City of the Provider	Zip Code of the Provider	State Code of the Provider	Country Code of the Provider	Provider Type	Medicare Participation Indicator	Place of Service	HCPCS Code	HCPC Descripti
0	MD	F	I	SAINT LOUIS	631041004.0	MO	US	Internal Medicine	Y	F	99223	Initial hospi inpatient ca typically 70
1	MD	F	I	FAYETTEVILLE	283043815.0	NC	US	Obstetrics & Gynecology	Y	O	G0202	Screeni mammograpl bilateral view study
2	DPM	M	I	NORTH HAVEN	64732343.0	CT	US	Podiatry	Y	O	99348	Establish patient hor visit, typica 25 n
3	MD	M	I	KANSAS CITY	641183998.0	MO	US	Internal Medicine	Y	O	81002	Urinalys manual te
4	DO	M	I	JUPITER	334585700.0	FL	US	Internal Medicine	Y	O	96372	Injecti beneath t skin or ir muscle for

Data Visualization

In [67]:
```
df.describe()
```

Out[67]:

	Zip Code of the Provider	Number of Services	Number of Medicare Beneficiaries	Number of Distinct Medicare Beneficiary/Per Day Services	Average Medicare Allowed Amount	Average Submitted Charge Amount	Average Medicare Payment Amount	Average Medicare Standardized Amount
count	9.999900e+04	99999.000000	99999.000000	99999.000000	99999.000000	99999.000000	99999.000000	99999.000000
mean	4.163828e+08	239.673621	89.810008	142.116901	101.435115	354.553846	77.359516	78.031420
std	3.082580e+08	2493.199458	1109.622428	1640.235384	257.243886	1062.613042	199.719682	200.046326
min	6.010000e+02	11.000000	11.000000	11.000000	0.010000	0.010000	0.008679	0.008679
25%	1.426300e+08	21.000000	17.000000	20.000000	24.270000	57.649074	19.335749	20.122753
50%	3.633025e+08	43.000000	32.000000	40.000000	65.100000	146.000000	47.020353	47.841582
75%	6.819881e+08	118.000000	75.000000	106.000000	113.160000	298.936222	84.895746	84.880047
max	9.990166e+08	282739.000000	190306.000000	282737.000000	20494.000000	62694.000000	16067.300000	16957.148000

By observing the standard deviation, it can be concluded that the data is widley spreaded.

Correlation Matrix

In [50]:
```
df_corr = df.select_dtypes(include=['number']).corr()
df_corr
```

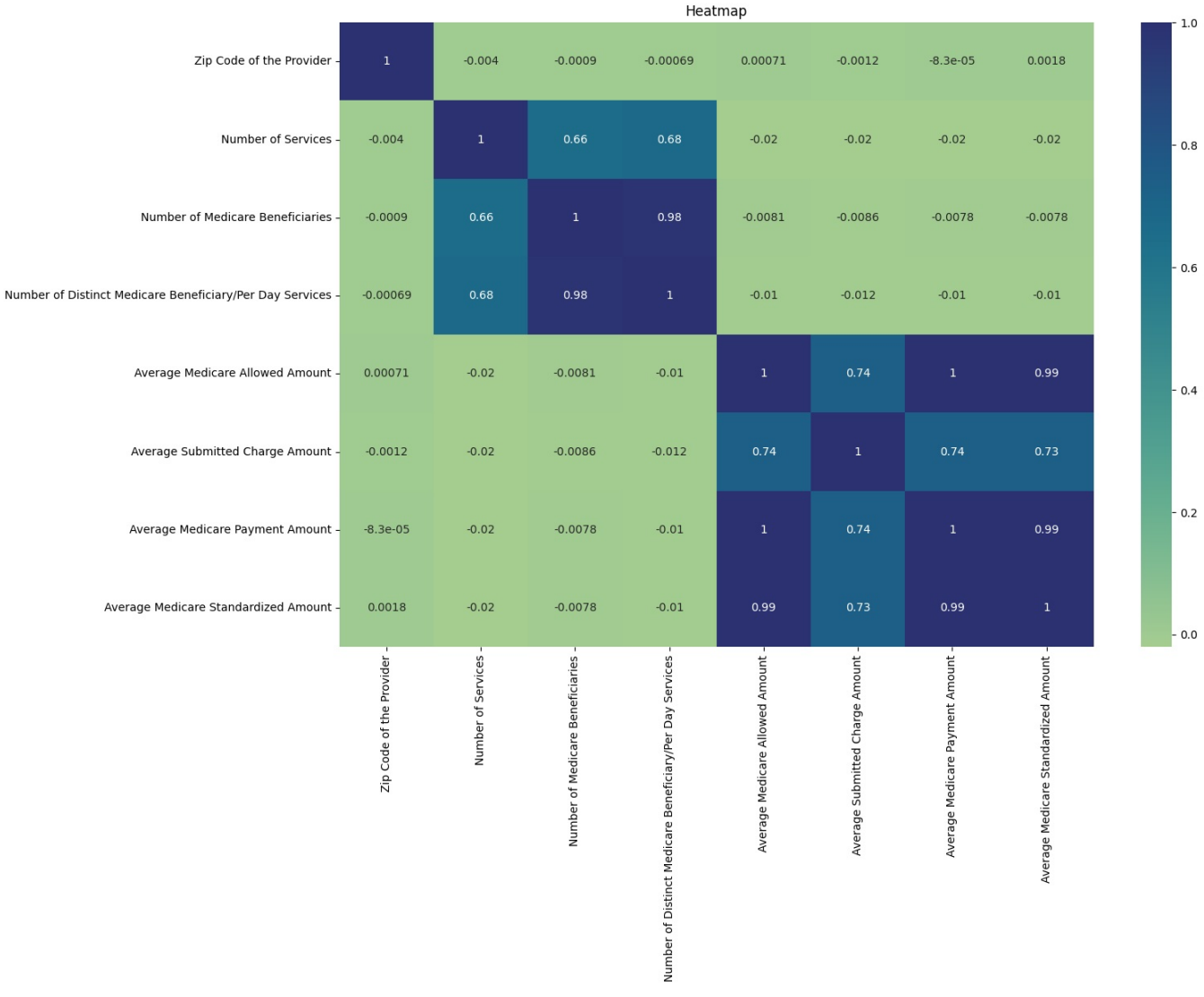
Out[50]:

	Zip Code of the Provider	Number of Services	Number of Medicare Beneficiaries	Number of Distinct Medicare Beneficiary/Per Day Services	Average Medicare Allowed Amount	Average Submitted Charge Amount	Average Medicare Payment Amount	Average Medicare Standardized Amount
Zip Code of the Provider	1.000000	-0.003997	-0.000900	-0.000686	0.000706	-0.001193	-0.000083	0.001842
Number of Services	-0.003997	1.000000	0.658923	0.675076	-0.020413	-0.019698	-0.019995	-0.020093
Number of Medicare Beneficiaries	-0.000900	0.658923	1.000000	0.981072	-0.008088	-0.008557	-0.007797	-0.007815
Number of Distinct Medicare Beneficiary/Per Day Services	-0.000686	0.675076	0.981072	1.000000	-0.010471	-0.011829	-0.010246	-0.010265
Average Medicare Allowed Amount	0.000706	-0.020413	-0.008088	-0.010471	1.000000	0.739766	0.998704	0.994831
Average Submitted Charge Amount	-0.001193	-0.019698	-0.008557	-0.011829	0.739766	1.000000	0.738814	0.733502
Average Medicare Payment Amount	-0.000083	-0.019995	-0.007797	-0.010246	0.998704	0.738814	1.000000	0.994649
Average Medicare Standardized Amount	0.001842	-0.020093	-0.007815	-0.010265	0.994831	0.733502	0.994649	1.000000

Heatmap

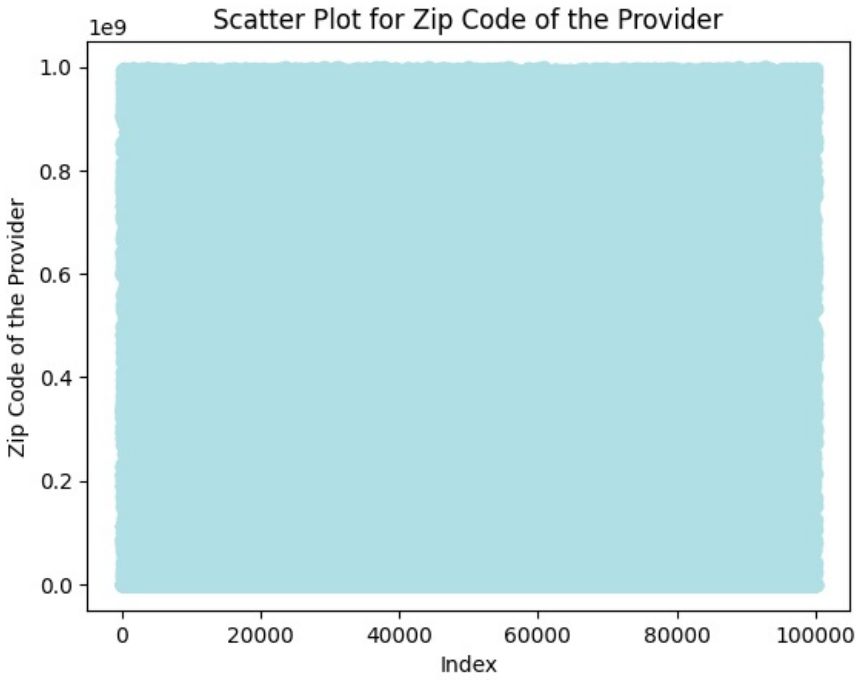
In [80]:
```
plt.figure(figsize=(15,10))
sns.heatmap(df_corr, annot=True, cmap='crest')
plt.title('Heatmap')
```

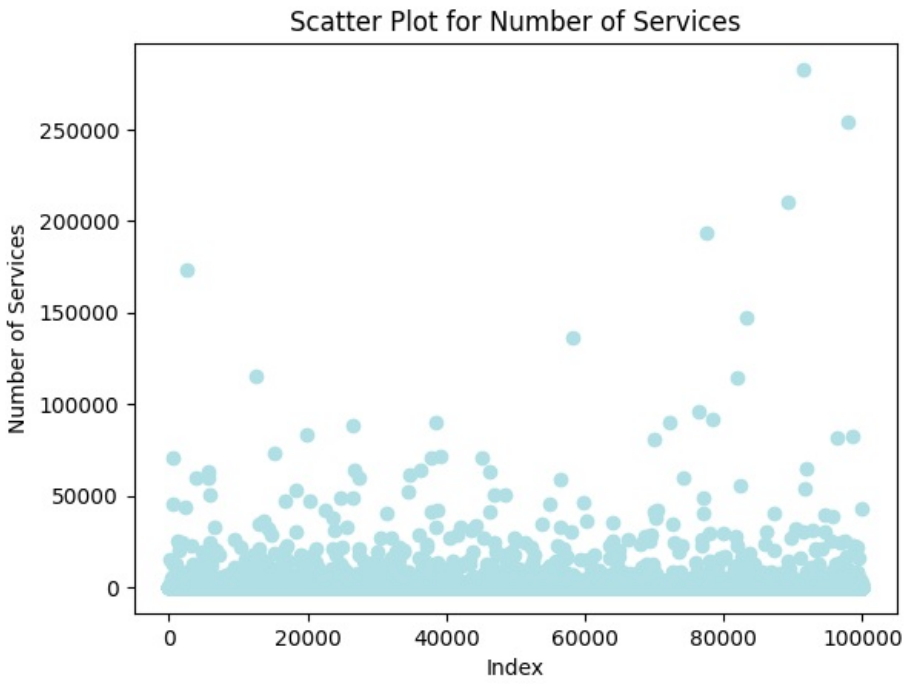
Out[80]:
```
Text(0.5, 1.0, 'Heatmap')
```

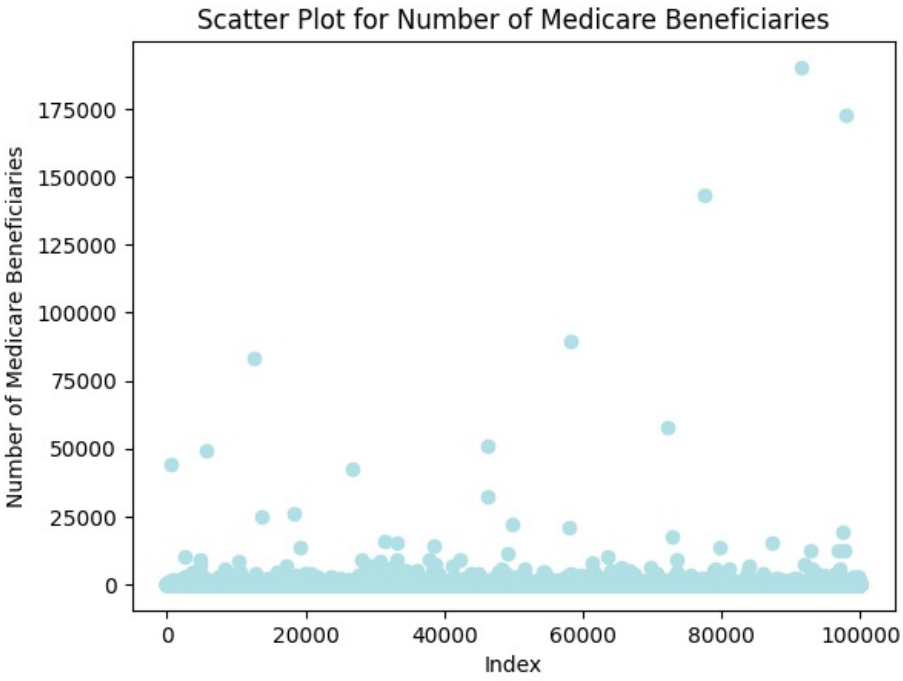


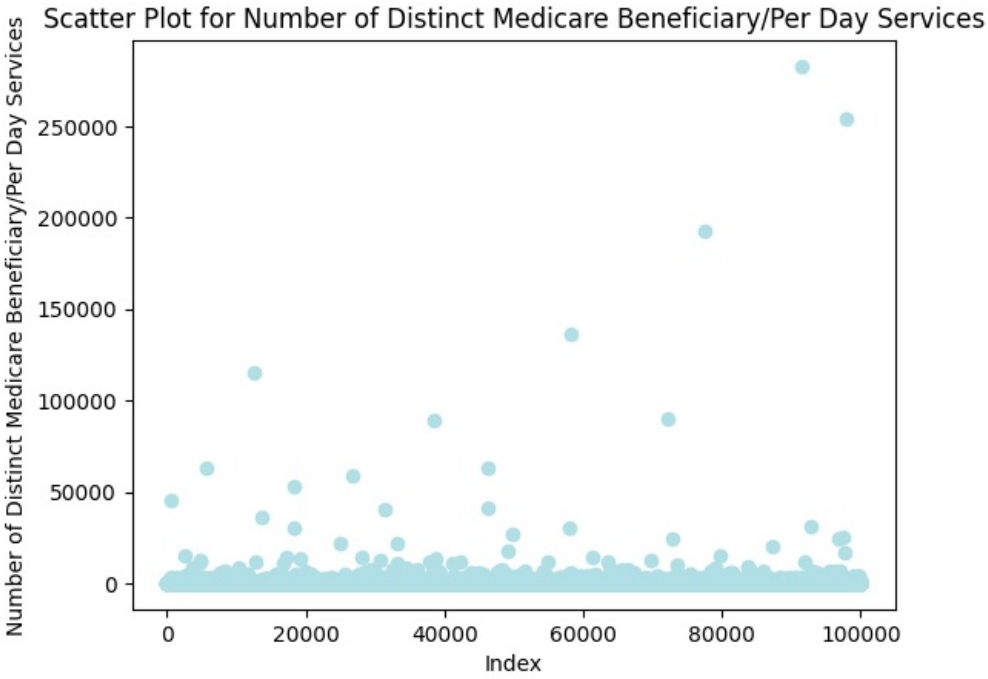
Insights: The 'Average Medicare Allowed Amount' and 'Average Medicare Payment Amount' shows a very strong correlation having value=1, suggesting that higher allowed amounts tend to result in higher payments.

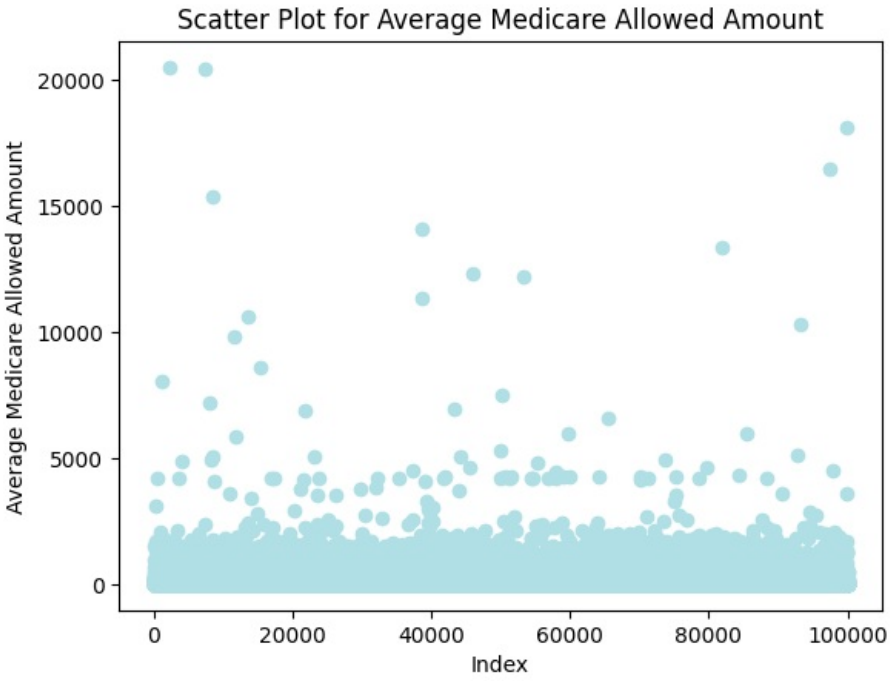
In [87]:
```
numerical_cols = df.select_dtypes(include=[np.number])

for col in numerical_cols.columns:
    plt.figure()
    plt.scatter(df.index, df[col], c='powderblue')
    plt.title(f"Scatter Plot for {col}")
    plt.xlabel("Index")
    plt.ylabel(col)

plt.show()
```

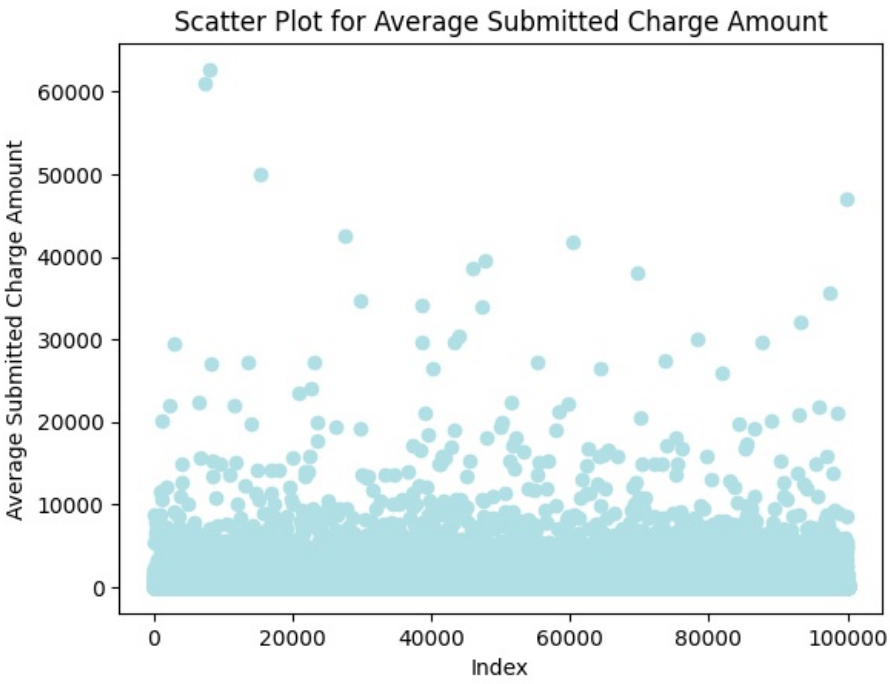


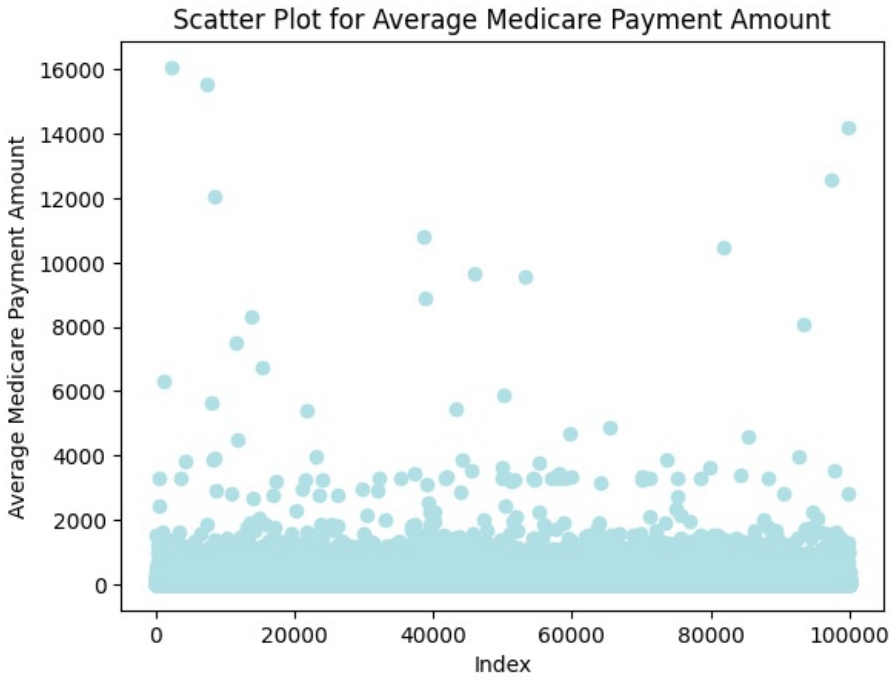


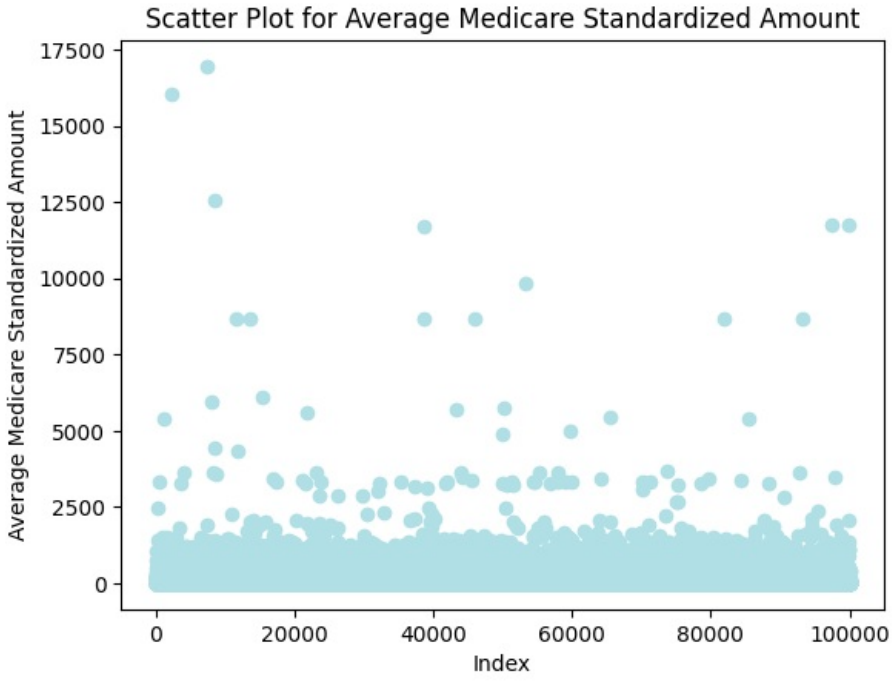

Scatter Plot for Number of Medicare Beneficiaries
Number of Medicare Beneficiaries
175000
150000
125000
100000
75000
50000
25000
0
0
20000
40000
60000
80000
100000
Index


Scatter Plot for Number of Distinct Medicare Beneficiary/Per Day Services
Number of Distinct Medicare Beneficiary/Per Day Services
250000
200000
150000
100000
50000
0
0
20000
40000
60000
80000
100000
Index


Scatter Plot for Average Medicare Allowed Amount
Average Medicare Allowed Amount
20000
15000
10000
5000
0
0
20000
40000
60000
80000
100000
Index


Scatter Plot for Average Submitted Charge Amount
Average Submitted Charge Amount
60000
50000
40000
30000
20000
10000
0
0
20000
40000
60000
80000
100000
Index





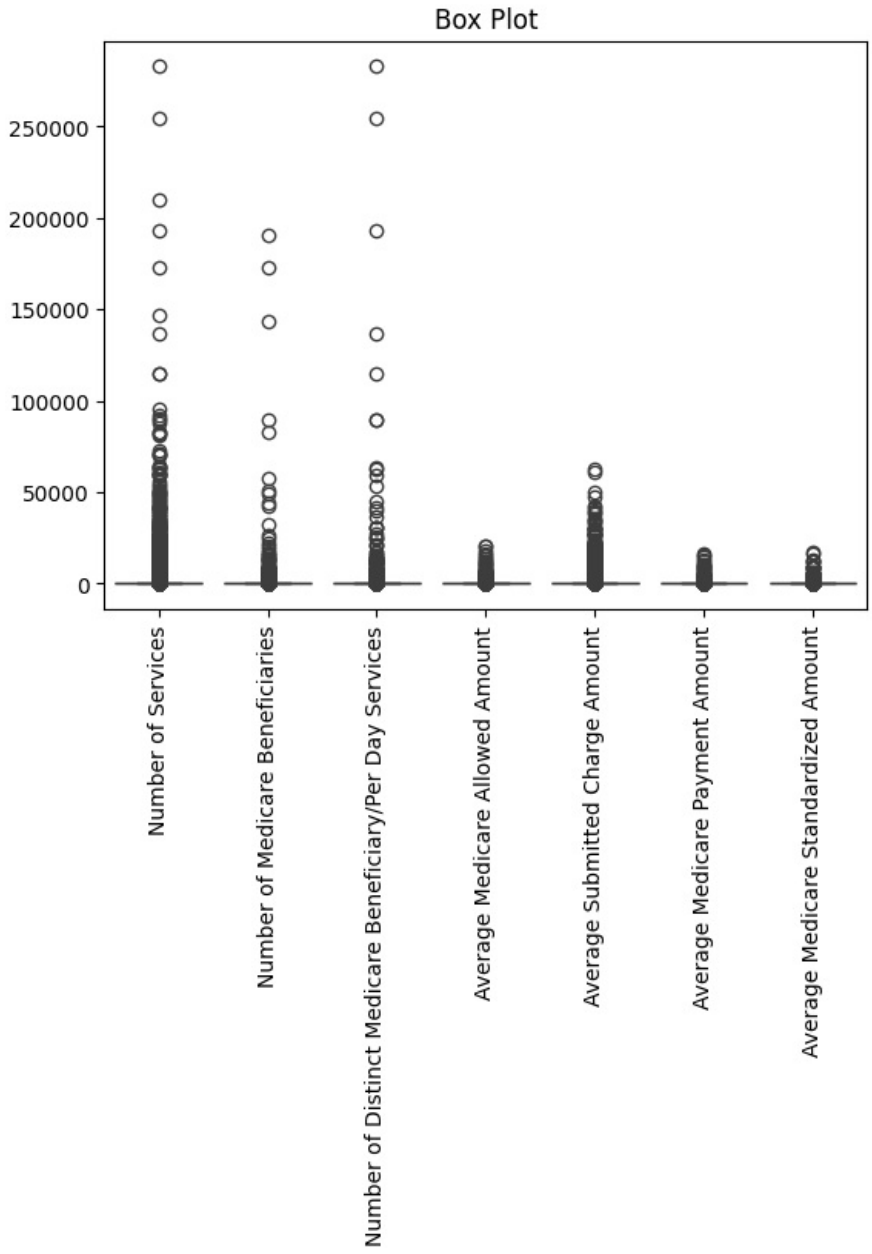
Observation: The 'Zip Code of the Provider' column has uniform distribution so it can be dropped while training the model for detecting outliers.

In [91]:
```
# Dropping Zip Code
df.drop(columns=['Zip Code of the Provider'], inplace=True)
```

Box Plot Representation

In [95]:
```
sns.boxplot(data=df)
plt.title('Box Plot')
plt.xticks(rotation=90)
```

Out[95]:
```
([0, 1, 2, 3, 4, 5, 6],
 [Text(0, 0, 'Number of Services'),
  Text(1, 0, 'Number of Medicare Beneficiaries'),
  Text(2, 0, 'Number of Distinct Medicare Beneficiary/Per Day Services'),
  Text(3, 0, 'Average Medicare Allowed Amount'),
  Text(4, 0, 'Average Submitted Charge Amount'),
  Text(5, 0, 'Average Medicare Payment Amount'),
  Text(6, 0, 'Average Medicare Standardized Amount')])
```



Insights: The columns 'Number of Services', 'Number of Medicare Beneficiaries', and 'Number of Distinct Medicare Beneficiary/Per Day Services' contains more number of outliers.

Visualizing using AutoViz Library

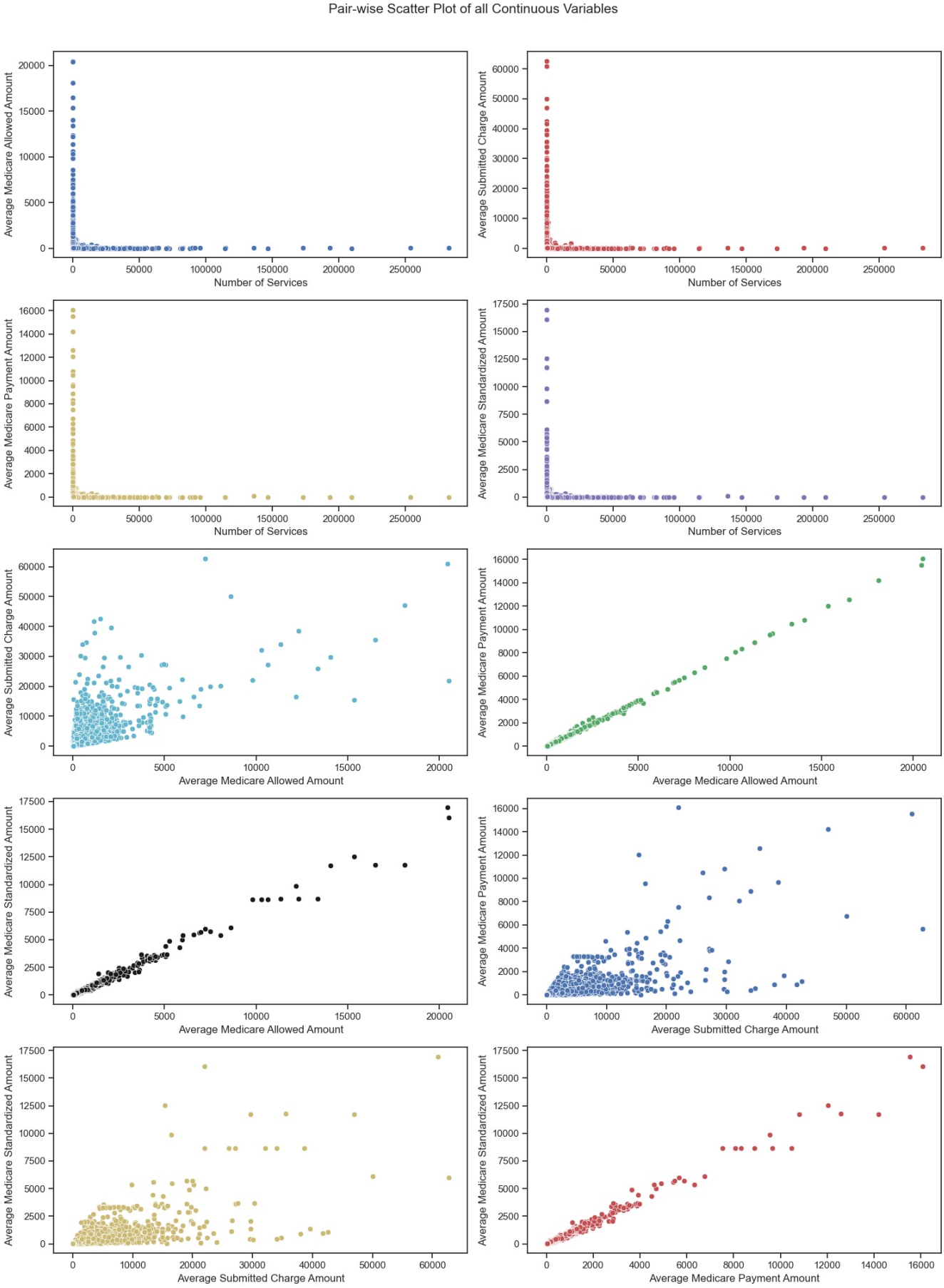
Univariate and Bivariate Analysis

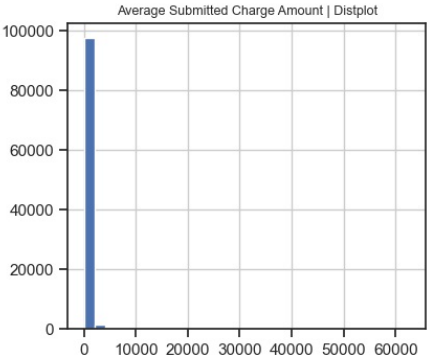
In [97]:
```
%matplotlib inline
from autoviz import AutoViz_Class
AV = AutoViz_Class()

dft = AV.AutoViz(df, chart_format="png")
```

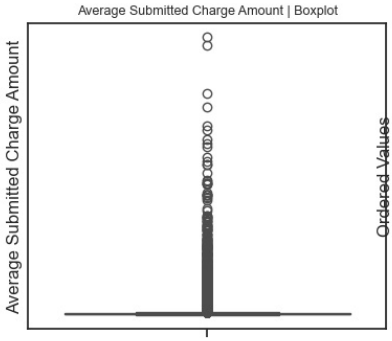
```
Shape of your Data Set loaded: (99999, 19)
###############################################################################################
###################### C L A S S I F Y I N G  V A R I A B L E S  ###################
###############################################################################################
Classifying variables in data set...
    Number of Numeric Columns =  5
    Number of Integer-Categorical Columns =  2
    Number of String-Categorical Columns =  3
    Number of Factor-Categorical Columns =  0
    Number of String-Boolean Columns =  5
    Number of Numeric-Boolean Columns =  0
    Number of Discrete String Columns =  0
    Number of NLP String Columns =  4
    Number of Date Time Columns =  0
    Number of ID Columns =  0
    Number of Columns to Delete =  0
    19 Predictors classified...
        No variables removed since no ID or low-information variables found in data set
To fix these data quality issues in the dataset, import FixDQ from autoviz...
There are 1 duplicate rows in your dataset
    Alert: Dropping duplicate rows can sometimes cause your column data types to change to object!
    All variables classified into correct types.
```

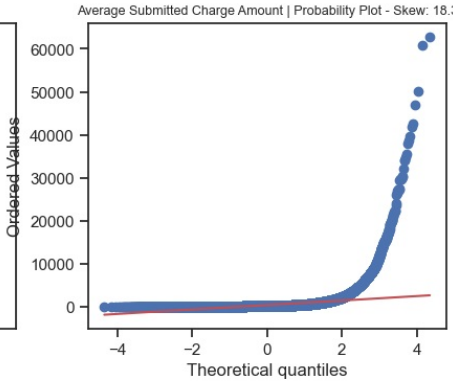
	Data Type	Missing Values%	Unique Values%	Minimum Value	Maximum Value	DQ Issue
Credentials of the Provider	object	0.000000	1			No issue
Gender of the Provider	object	0.000000	0			No issue
Entity Type of the Provider	object	0.000000	0			No issue
City of the Provider	object	0.000000	5			No issue
State Code of the Provider	object	0.000000	0			25 rare categories: Too many to list. Group them into a single category or drop the categories.
Country Code of the Provider	object	0.000000	0			3 rare categories: ['DE', 'JP', 'TR']. Group them into a single category or drop the categories.
Provider Type	object	0.000000	0			64 rare categories: Too many to list. Group them into a single category or drop the categories.
Medicare Participation Indicator	object	0.000000	0			No issue
Place of Service	object	0.000000	0			No issue
HCPCS Code	object	0.000000	2			No issue
HCPCS Description	object	0.000000	2			No issue
HCPCS Drug Indicator	object	0.000000	0			No issue
Number of Services	float64	0.000000	NA	11.000000	282739.000000	Column has 12316 outliers greater than upper bound (263.50) or lower than lower bound(-124.50). Cap them or remove them.
Number of Medicare Beneficiaries	int64	0.000000	1	11.000000	190306.000000	Column has 10473 outliers greater than upper bound (162.00) or lower than lower bound(-70.00). Cap them or remove them.
Number of Distinct Medicare Beneficiary/Per Day Services	int64	0.000000	1	11.000000	282737.000000	Column has 11895 outliers greater than upper bound (235.00) or lower than lower bound(-109.00). Cap them or remove them., Column has a high correlation with ['Number of Medicare Beneficiaries']. Consider dropping one of them.
Average Medicare Allowed Amount	float64	0.000000	NA	0.010000	20494.000000	Column has 5079 outliers greater than upper bound (246.49) or lower than lower bound(-109.06). Cap them or remove them.
Average Submitted Charge Amount	float64	0.000000	NA	0.010000	62694.000000	Column has 10673 outliers greater than upper bound (660.88) or lower than lower bound(-304.29). Cap them or remove them.
Average Medicare Payment Amount	float64	0.000000	NA	0.008679	16067.300000	Column has 6093 outliers greater than upper bound (183.24) or lower than lower bound(-79.01). Cap them or remove them., Column has a high correlation with ['Average Medicare Allowed Amount']. Consider dropping one of them.
Average Medicare Standardized Amount	float64	0.000000	NA	0.008679	16957.148000	Column has 6078 outliers greater than upper bound (182.02) or lower than lower bound(-77.01). Cap them or remove them., Column has a high correlation with ['Average Medicare Allowed Amount', 'Average Medicare Payment Amount']. Consider dropping one of them.

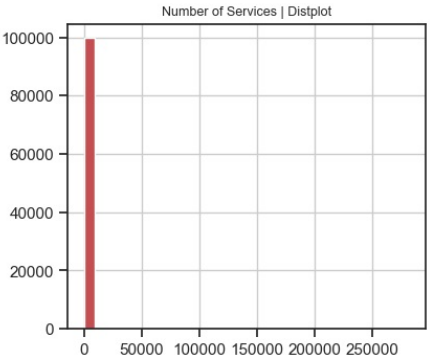
Number of All Scatter Plots = 15

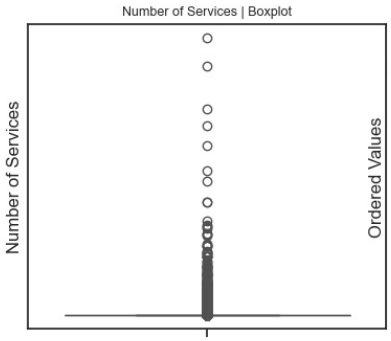


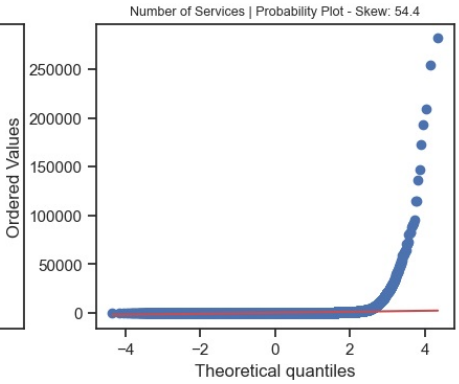


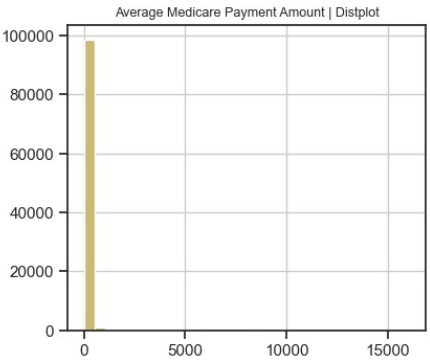


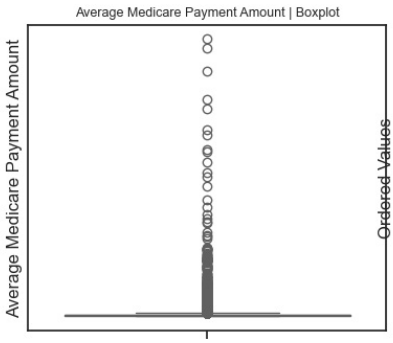


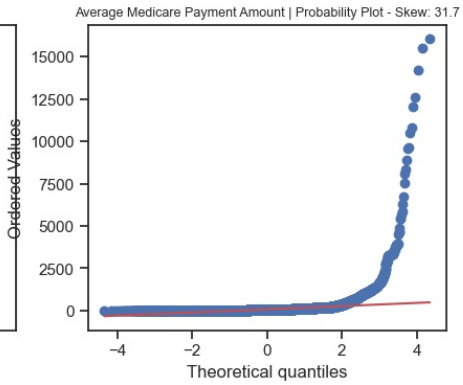


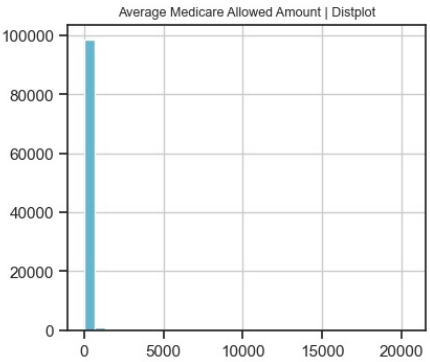


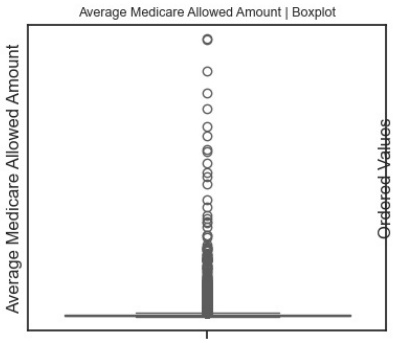


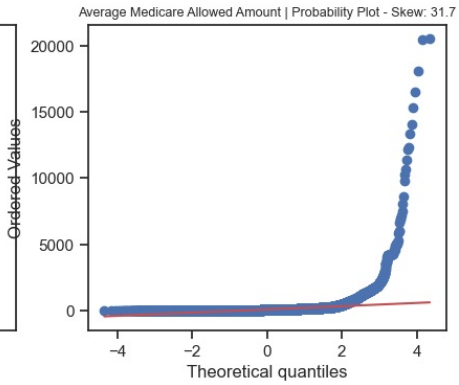


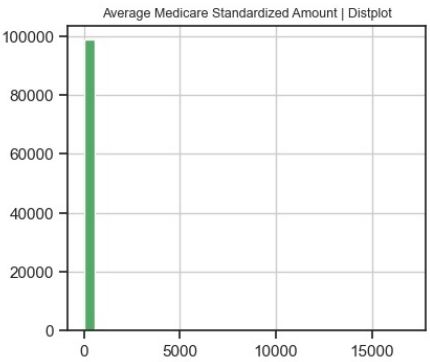


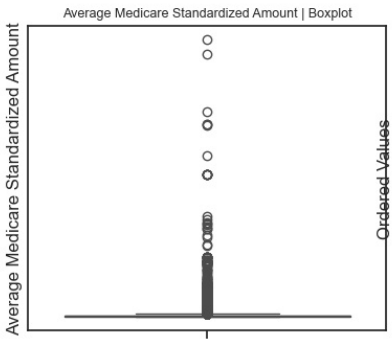


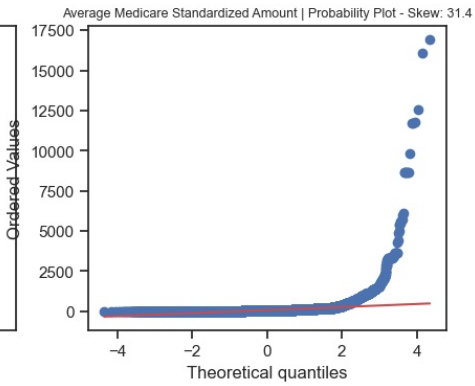




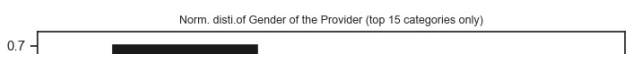


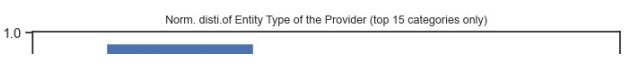


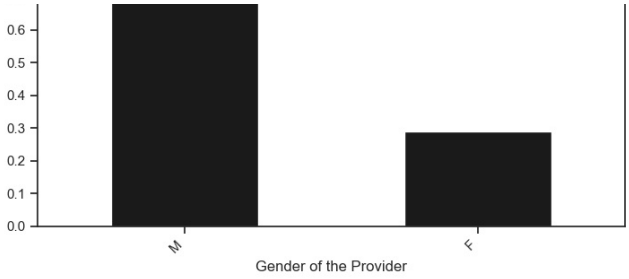


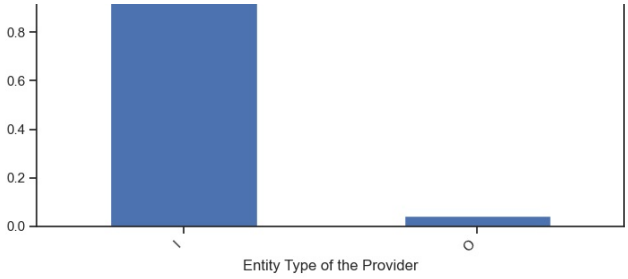


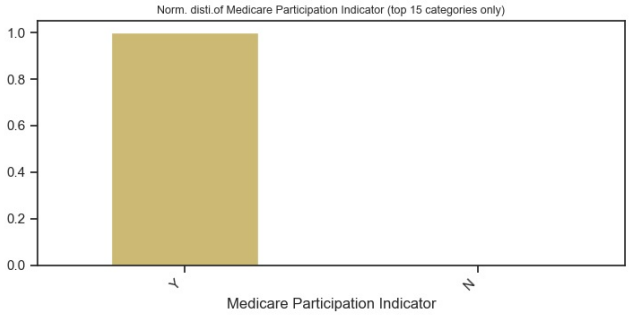
Histograms and Normalized distribitions of all variables

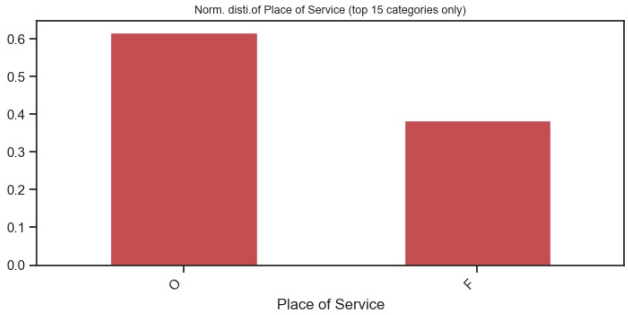


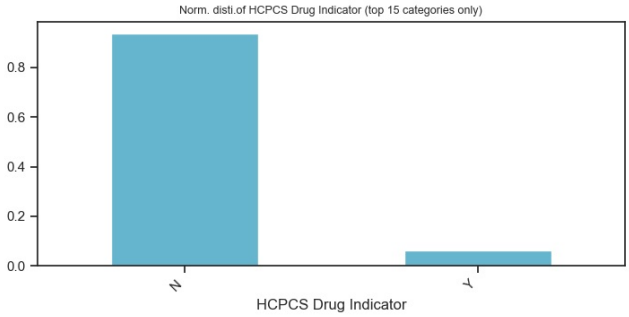


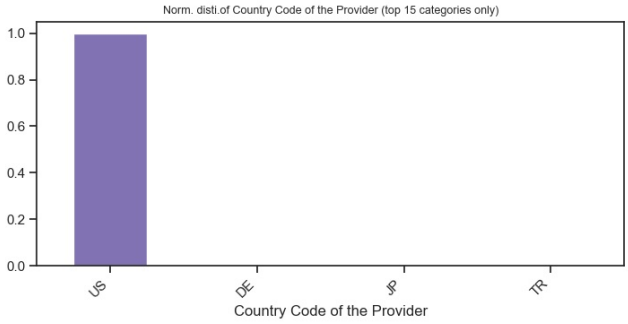

M
F
Gender of the Provider

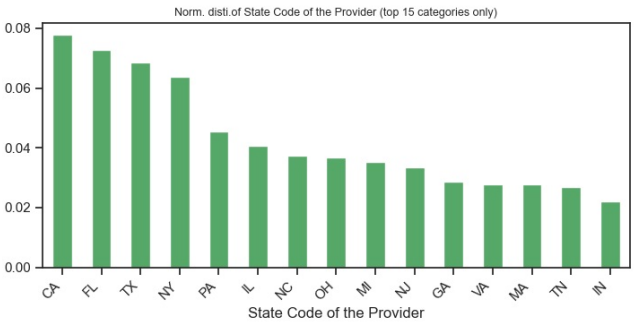

I
O
Entity Type of the Provider

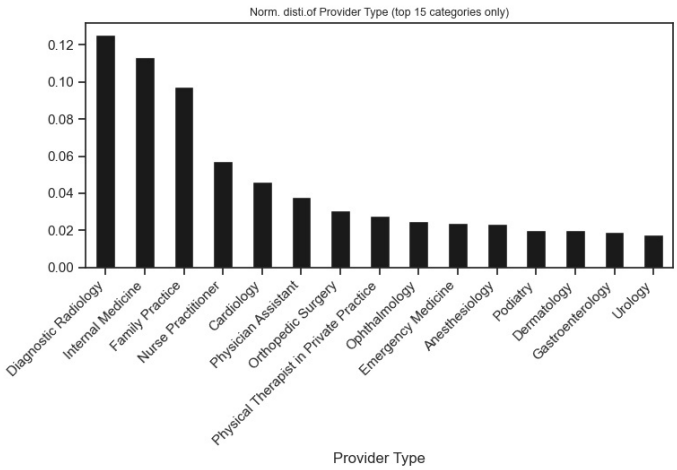

Norm. disti.of Medicare Participation Indicator (top 15 categories only)
Y
N
Medicare Participation Indicator

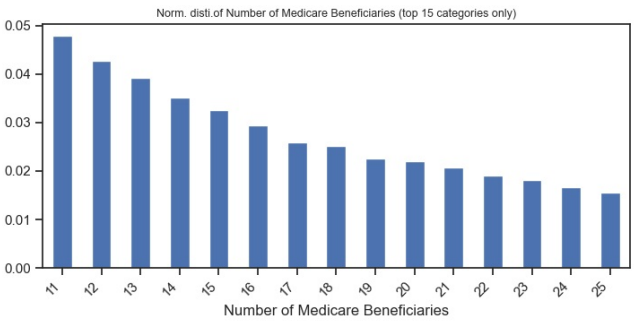

Norm. disti.of Place of Service (top 15 categories only)
O
F
Place of Service

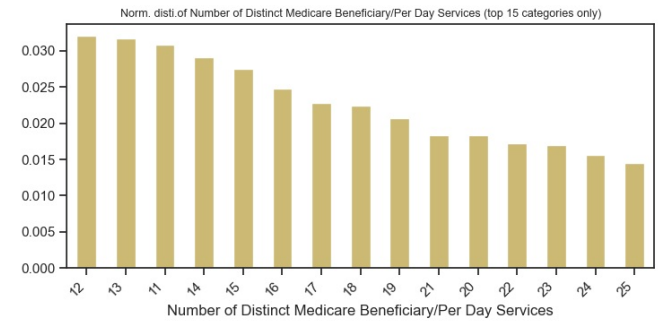

Norm. disti.of HCPCS Drug Indicator (top 15 categories only)
N
Y
HCPCS Drug Indicator


Norm. disti.of Country Code of the Provider (top 15 categories only)
US
DE
JP
TR
Country Code of the Provider

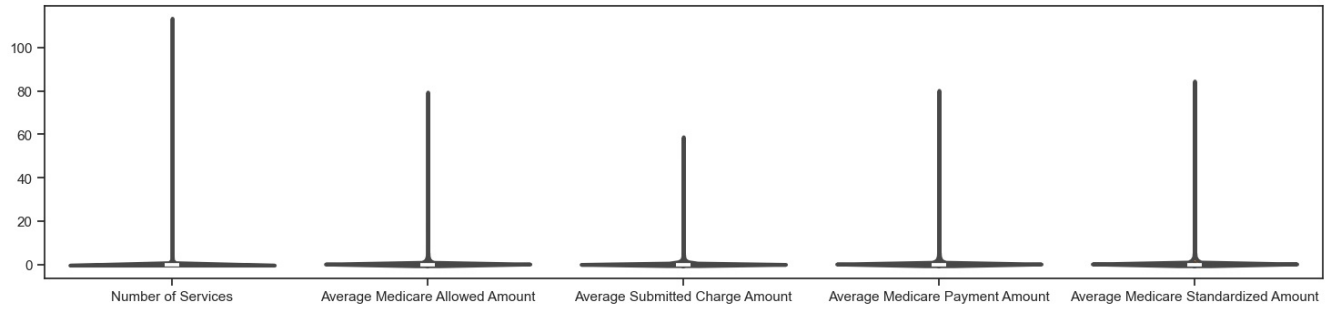

Norm. disti.of State Code of the Provider (top 15 categories only)
CA
FL
TX
NY
PA
IL
NC
OH
MI
NJ
GA
VA
MA
TN
IN
State Code of the Provider


Norm. disti.of Provider Type (top 15 categories only)
Diagnostic Radiology
Internal Medicine
Family Practice
Nurse Practitioner
Cardiology
Physician Assistant
Orthopedic Surgery
Physical Therapist in Private Practice
Ophthalmology
Emergency Medicine
Anesthesiology
Podiatry
Dermatology
Gastroenterology
Urology
Provider Type

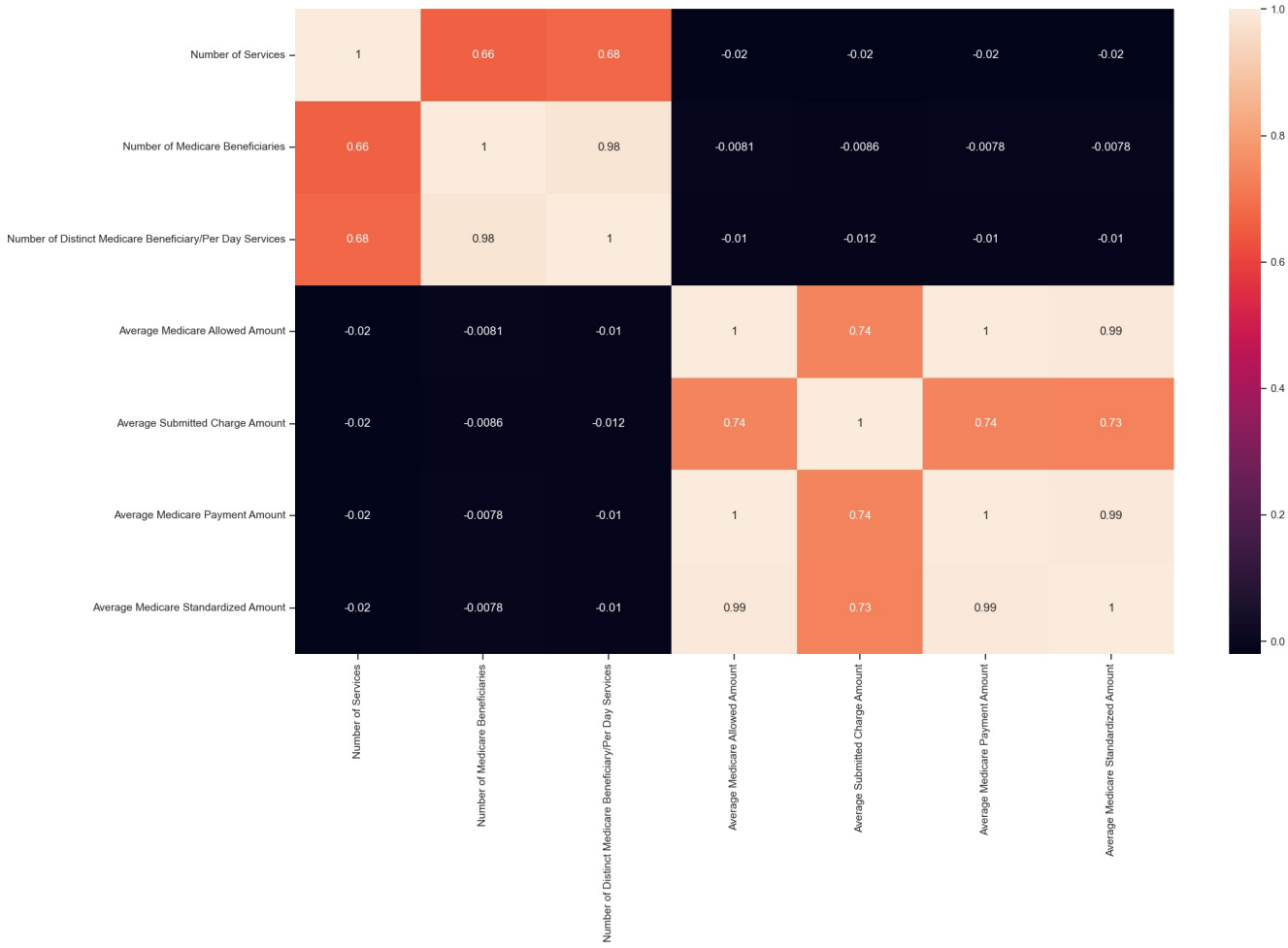

Norm. disti.of Number of Medicare Beneficiaries (top 15 categories only)
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
Number of Medicare Beneficiaries


Norm. disti.of Number of Distinct Medicare Beneficiary/Per Day Services (top 15 categories only)
12
13
11
14
15
16
17
18
19
21
20
22
23
24
25
Number of Distinct Medicare Beneficiary/Per Day Services

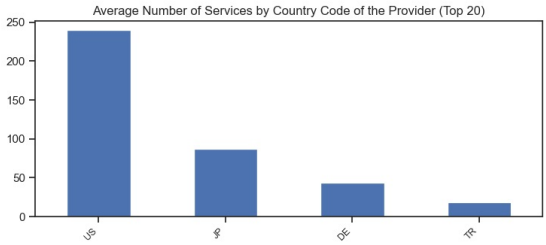
Violin Plot of all Continuous Variables

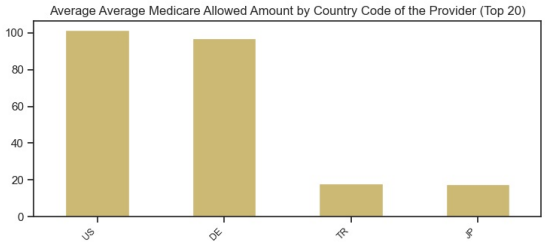


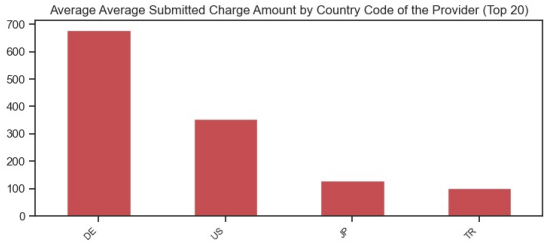
Heatmap of all Numeric Variables including target:

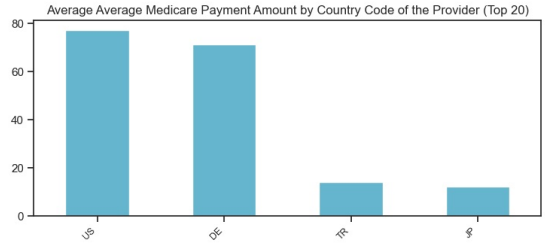


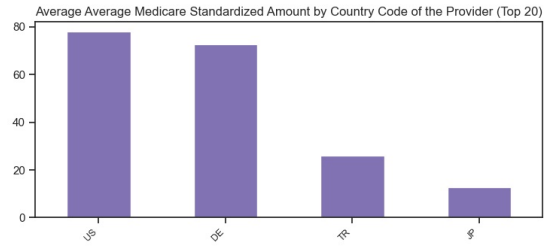
Bar plots for each Continuous by each Categorical variable

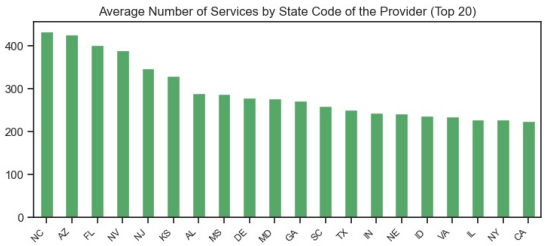

Average Number of Services by Country Code of the Provider (Top 20)
US
JP
DE
TR

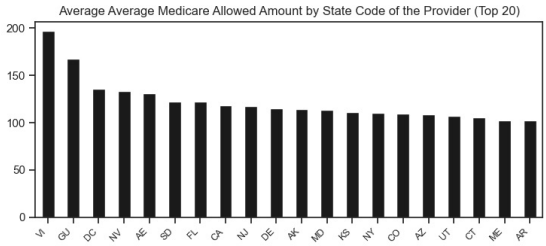

Average Average Medicare Allowed Amount by Country Code of the Provider (Top 20)
US
DE
TR
JP

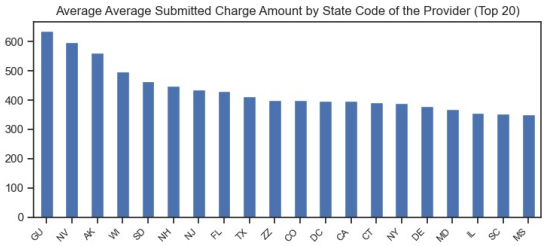

Average Average Submitted Charge Amount by Country Code of the Provider (Top 20)
DE
US
JP
TR

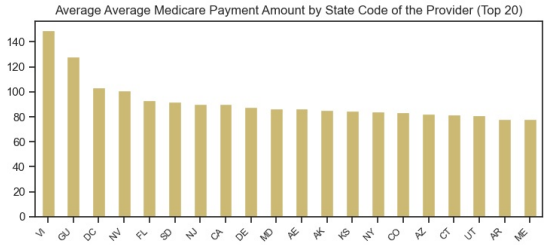

Average Average Medicare Payment Amount by Country Code of the Provider (Top 20)
US
DE
TR
JP

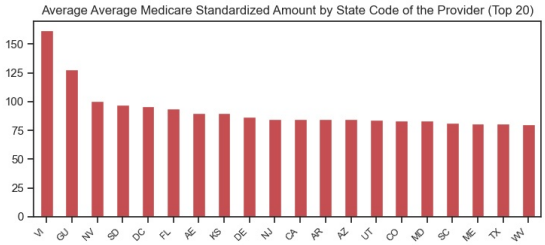

Average Average Medicare Standardized Amount by Country Code of the Provider (Top 20)
US
DE
TR
JP

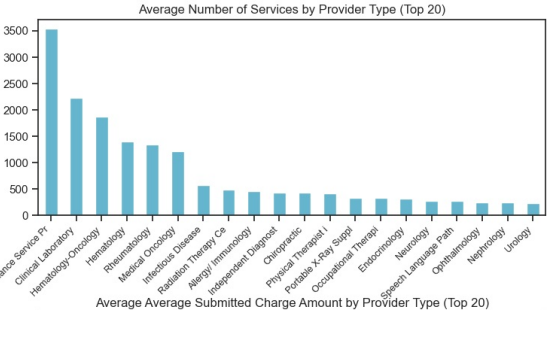

Average Number of Services by State Code of the Provider (Top 20)
NC
AZ
FL
NV
NJ
KS
AL
MS
DE
MD
GA
SC
TX
IN
NE
ID
VA
IL
NY
CA

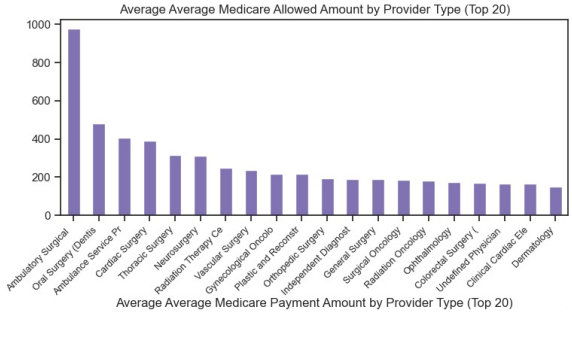

Average Average Medicare Allowed Amount by State Code of the Provider (Top 20)
VI
GU
DC
NV
AE
SD
FL
CA
NJ
DE
AK
MD
KS
NY
CO
AZ
UT
CT
ME
AR

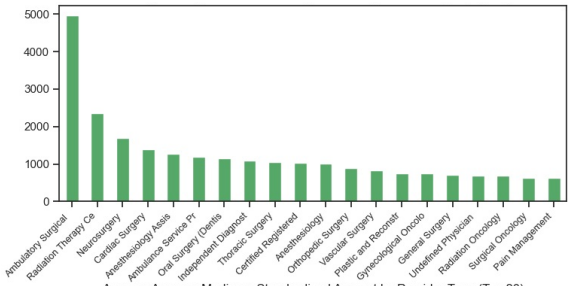

Average Average Submitted Charge Amount by State Code of the Provider (Top 20)
GU
NV
AK
WI
SD
NH
NJ
FL
TX
ZZ
CO
DC
CA
CT
NY
DE
MD
IL
SC
MS

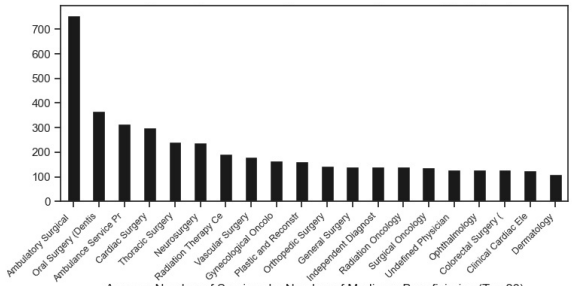

Average Average Medicare Payment Amount by State Code of the Provider (Top 20)
VI
GU
DC
NV
FL
SD
NJ
CA
DE
MD
AE
AK
KS
NY
CO
AZ
CT
UT
AR
ME


Average Average Medicare Standardized Amount by State Code of the Provider (Top 20)
VI
GU
NV
SD
DC
FL
AE
KS
DE
NJ
CA
AR
AZ
UT
CO
MD
SC
ME
TX
WV

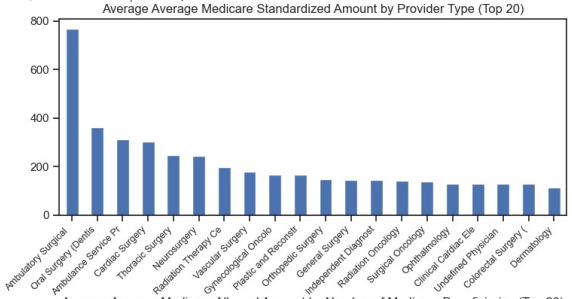

Average Number of Services by Provider Type (Top 20)
Ambulance Service Pr
Clinical Laboratory
Hematology-Oncology
Hematology
Rheumatology
Medical Oncology
Infectious Disease
Radiation Therapy Ca
Allergy/ Immunology
Independent Diagnost
Chiropractic
Physical Therapist i
Portable X-Ray Suppl
Occupational Therapi
Endocrinology
Neurology
Speech Language Path
Ophthalmology
Nephrology
Urology
Average Average Submitted Charge Amount by Provider Type (Top 20)


Average Average Medicare Allowed Amount by Provider Type (Top 20)
Ambulatory Surgical
Oral Surgery (Dentis
Ambulance Service Pr
Cardiac Surgery
Thoracic Surgery
Neurosurgery
Radiation Therapy Ce
Vascular Surgery
Gynecological Oncolo
Plastic and Reconstr
Orthopedic Surgery
Independent Diagnost
General Surgery
Surgical Oncology
Radiation Oncology
Ophthalmology
Colorectal Surgery (
Undefined Physician
Clinical Cardiac Ele
Dermatology
Average Average Medicare Payment Amount by Provider Type (Top 20)

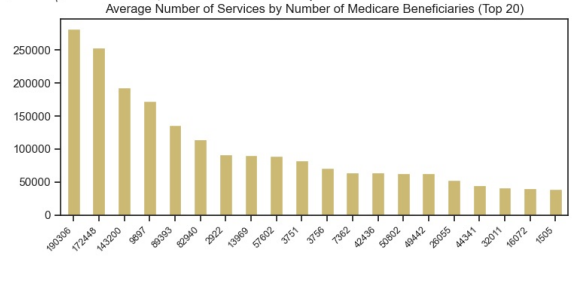




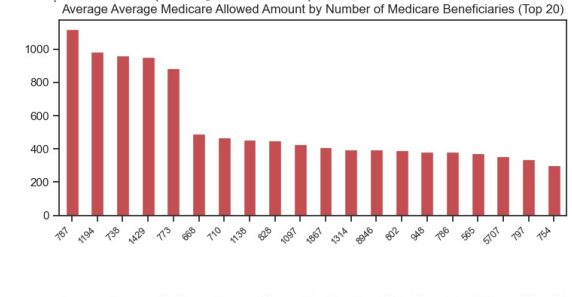
Average Average Medicare Standardized Amount by Provider Type (Top 20)



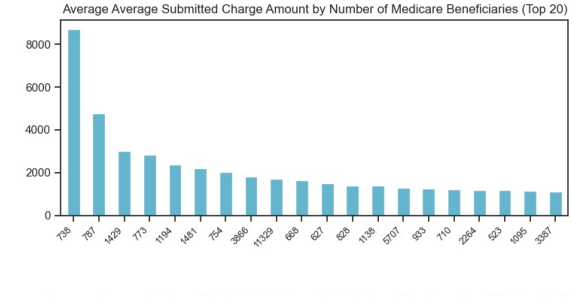
Average Number of Services by Number of Medicare Beneficiaries (Top 20)



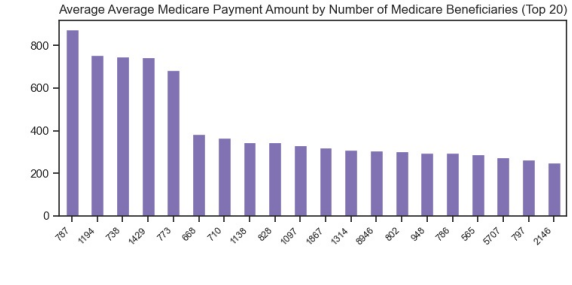
Average Average Medicare Allowed Amount by Number of Medicare Beneficiaries (Top 20)



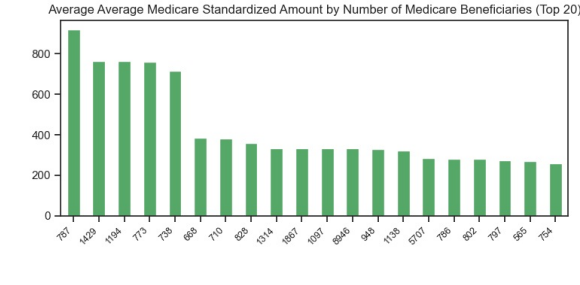
Average Average Submitted Charge Amount by Number of Medicare Beneficiaries (Top 20)



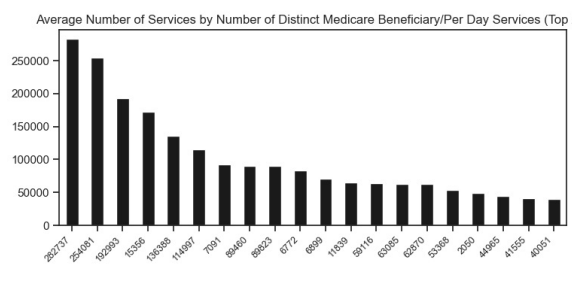
Average Average Medicare Payment Amount by Number of Medicare Beneficiaries (Top 20)



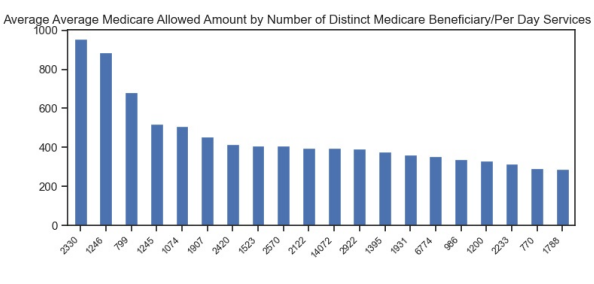
Average Average Medicare Standardized Amount by Number of Medicare Beneficiaries (Top 20)



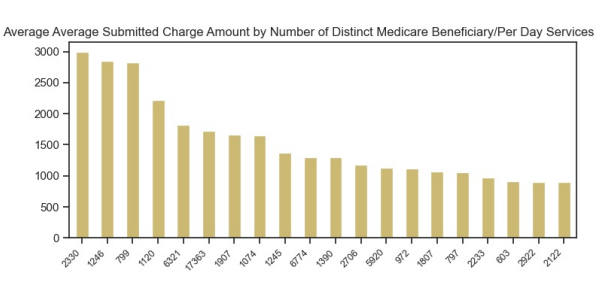
Average Number of Services by Number of Distinct Medicare Beneficiary/Per Day Services (Top 20)



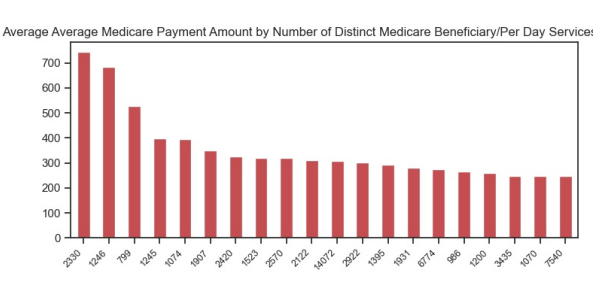
Average Average Medicare Allowed Amount by Number of Distinct Medicare Beneficiary/Per Day Services (Top 20)



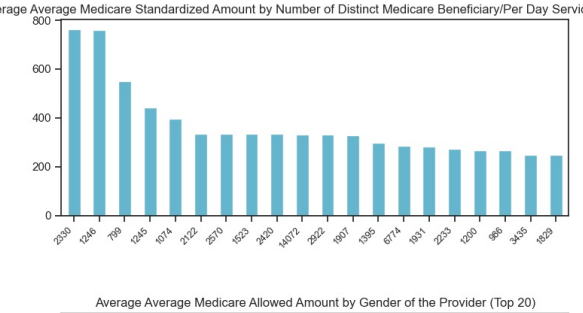
Average Average Submitted Charge Amount by Number of Distinct Medicare Beneficiary/Per Day Services (Top 20)



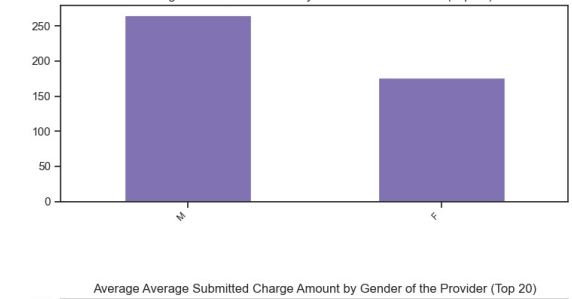
Average Average Medicare Payment Amount by Number of Distinct Medicare Beneficiary/Per Day Services (Top 20)



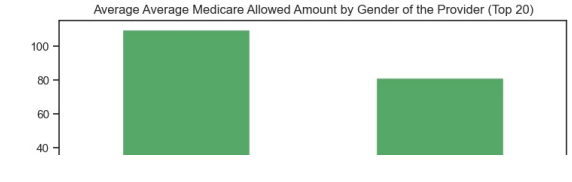
Average Average Medicare Standardized Amount by Number of Distinct Medicare Beneficiary/Per Day Services (Top 20)



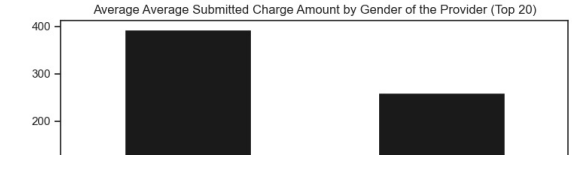
Average Number of Services by Gender of the Provider (Top 20)



Average Average Medicare Allowed Amount by Gender of the Provider (Top 20)

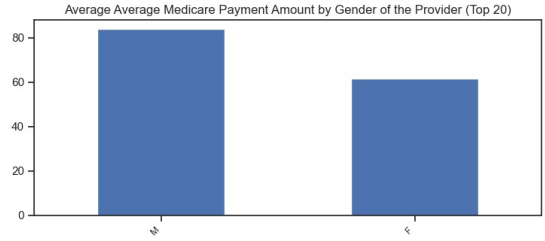


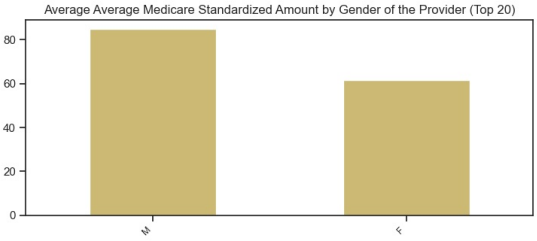
Average Average Submitted Charge Amount by Gender of the Provider (Top 20)

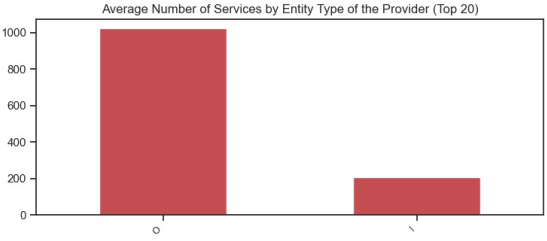


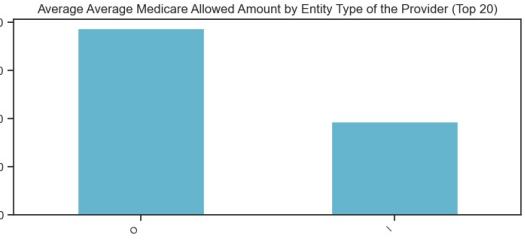

20
0
M
F

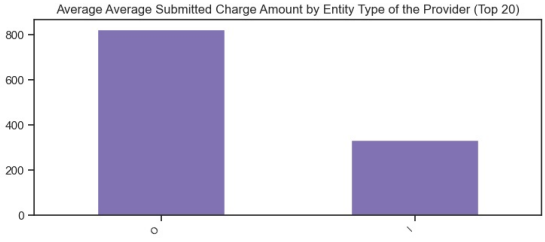

100
0
M
F

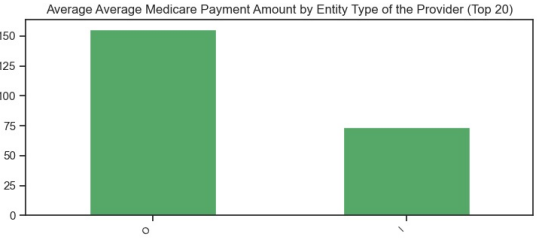

Average Average Medicare Payment Amount by Gender of the Provider (Top 20)
80
60
40
20
0
M
F

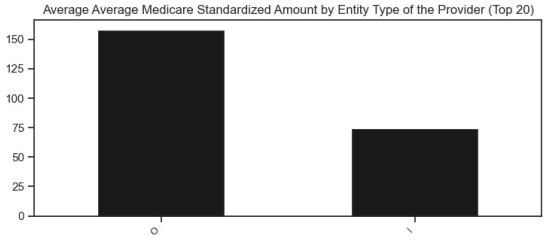

Average Average Medicare Standardized Amount by Gender of the Provider (Top 20)
80
60
40
20
0
M
F

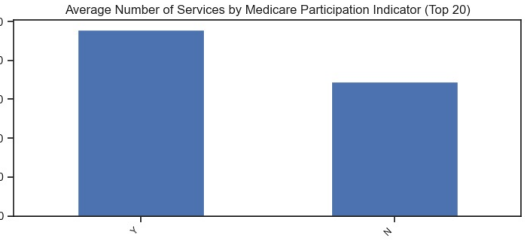

Average Number of Services by Entity Type of the Provider (Top 20)
1000
800
600
400
200
0
O
I

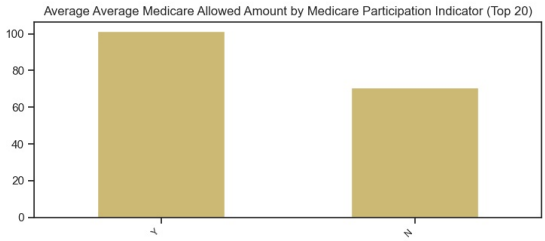

Average Average Medicare Allowed Amount by Entity Type of the Provider (Top 20)
200
150
100
50
0
O
I

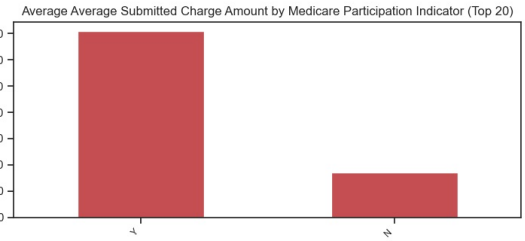

Average Average Submitted Charge Amount by Entity Type of the Provider (Top 20)
800
600
400
200
0
O
I

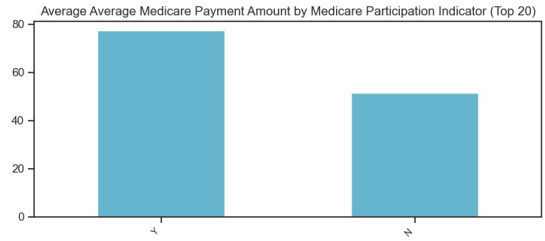

Average Average Medicare Payment Amount by Entity Type of the Provider (Top 20)
150
125
100
75
50
25
0
O
I

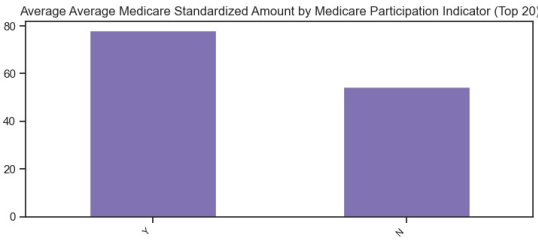

Average Average Medicare Standardized Amount by Entity Type of the Provider (Top 20)
150
125
100
75
50
25
0
O
I

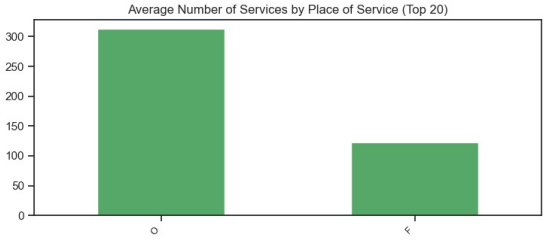

Average Number of Services by Medicare Participation Indicator (Top 20)
250
200
150
100
50
0
Y
N

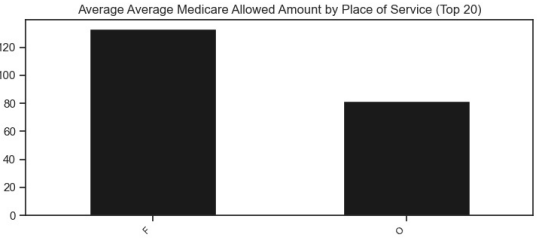

Average Average Medicare Allowed Amount by Medicare Participation Indicator (Top 20)
100
80
60
40
20
0
Y
N

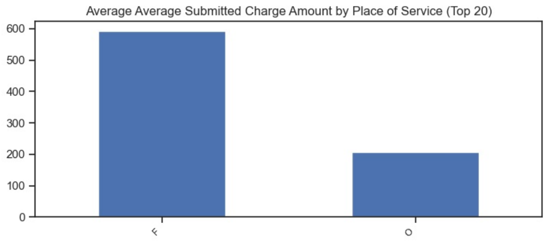

Average Average Submitted Charge Amount by Medicare Participation Indicator (Top 20)
350
300
250
200
150
100
50
0
Y
N

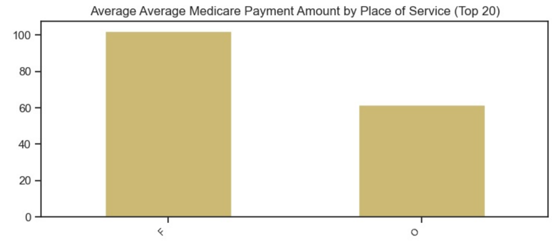

Average Average Medicare Payment Amount by Medicare Participation Indicator (Top 20)
80
60
40
20
0
Y
N

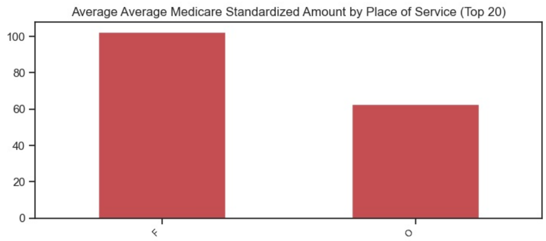

Average Average Medicare Standardized Amount by Medicare Participation Indicator (Top 20)
80
60
40
20
0
Y
N

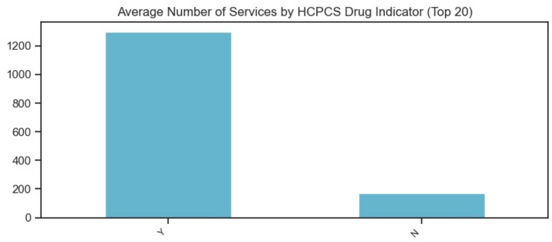

Average Number of Services by Place of Service (Top 20)
300
250
200
150
100
50
0
O
F

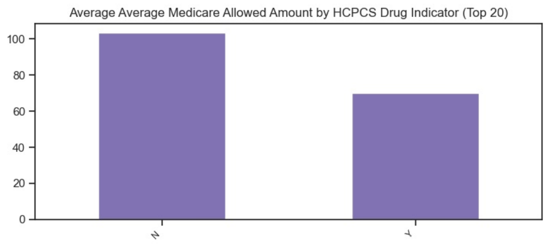

Average Average Medicare Allowed Amount by Place of Service (Top 20)
120
100
80
60
40
20
0
F
O

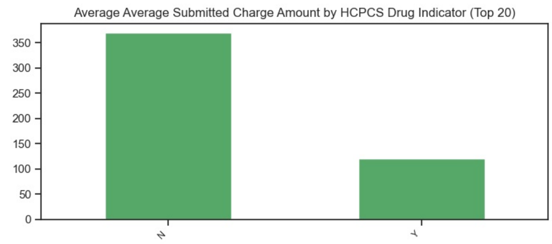

Average Average Submitted Charge Amount by Place of Service (Top 20)
0
100
200
300
400
500
600
F
O

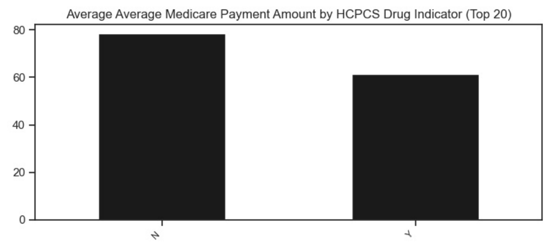

Average Average Medicare Payment Amount by Place of Service (Top 20)
0
20
40
60
80
100
F
O

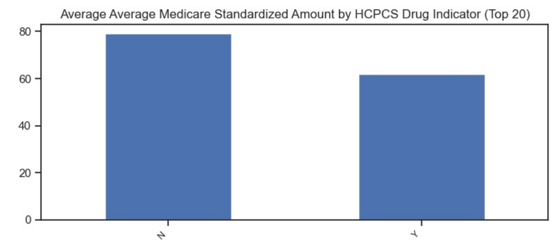

Average Average Medicare Standardized Amount by Place of Service (Top 20)
0
20
40
60
80
100
F
O

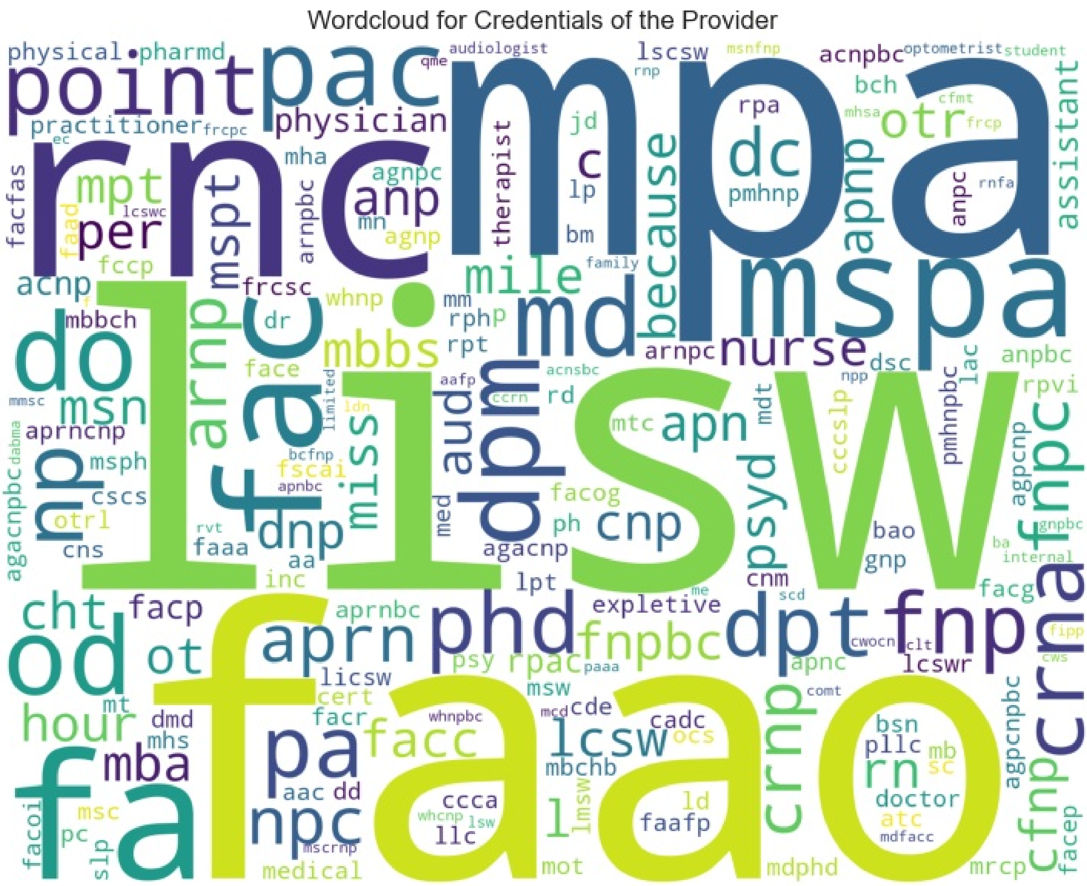

Average Number of Services by HCPCS Drug Indicator (Top 20)
0
200
400
600
800
1000
1200
Y
N

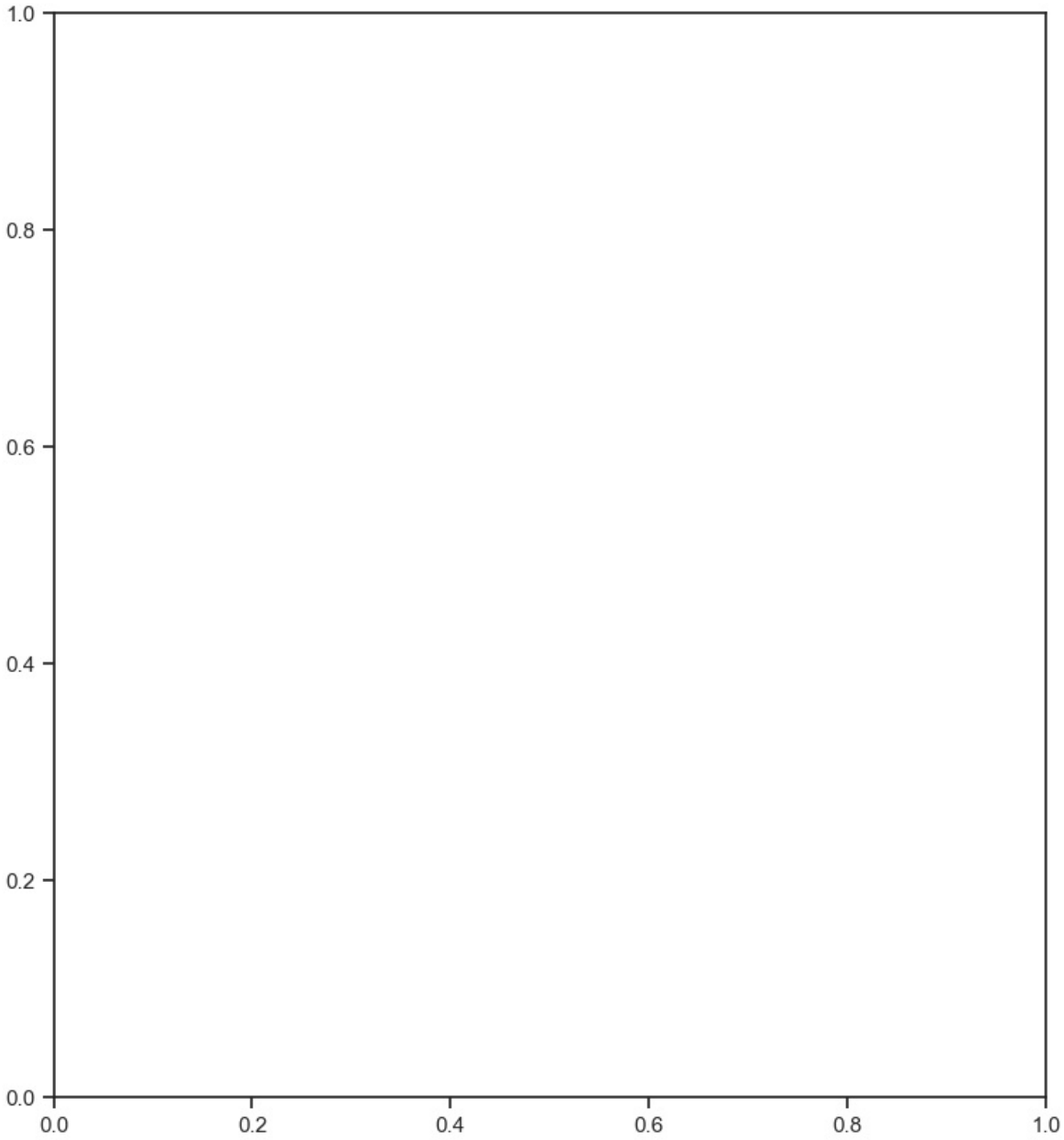

Average Average Medicare Allowed Amount by HCPCS Drug Indicator (Top 20)
0
20
40
60
80
100
N
Y


Average Average Submitted Charge Amount by HCPCS Drug Indicator (Top 20)
0
50
100
150
200
250
300
350
N
Y

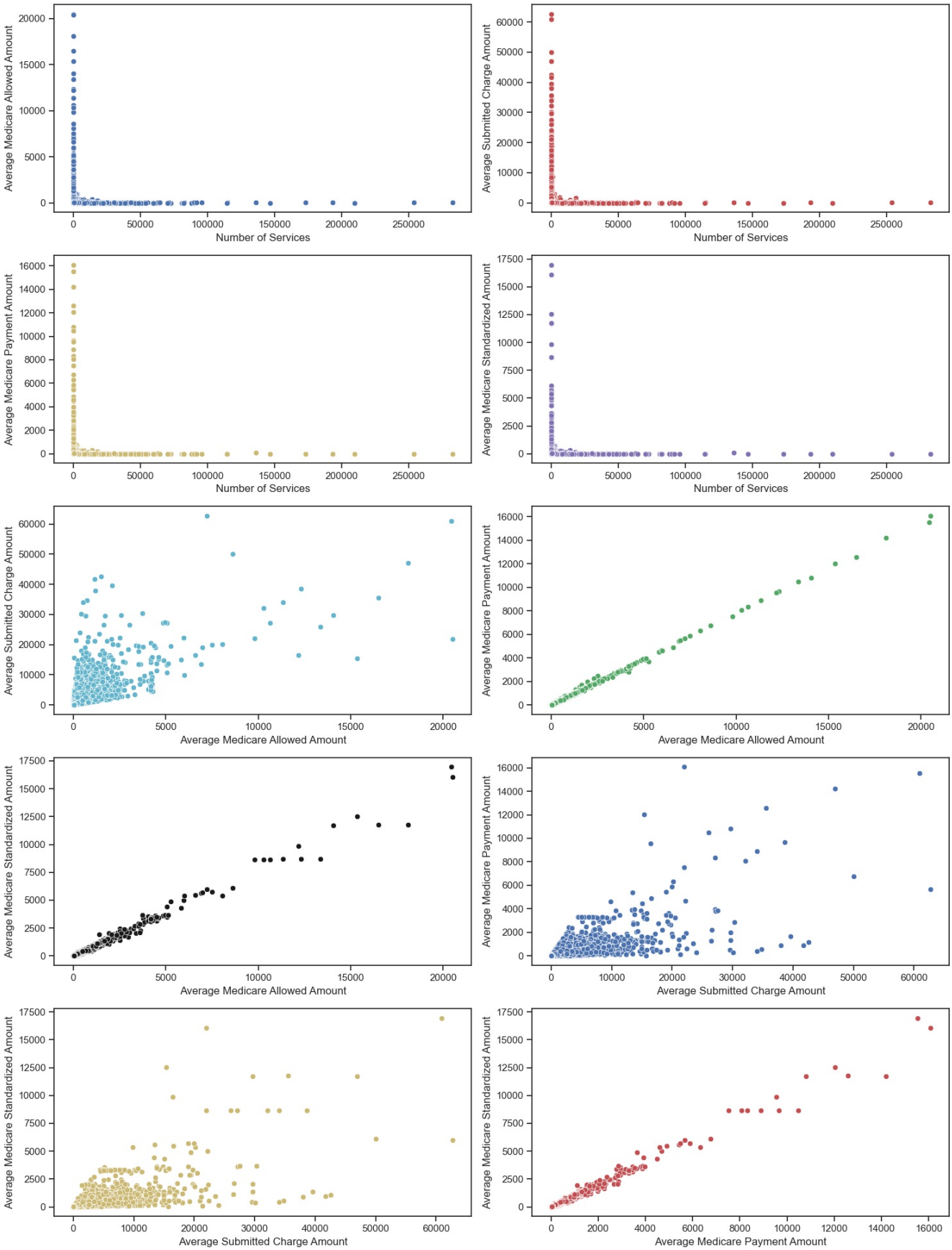

Average Average Medicare Payment Amount by HCPCS Drug Indicator (Top 20)
0
20
40
60
80
N
Y

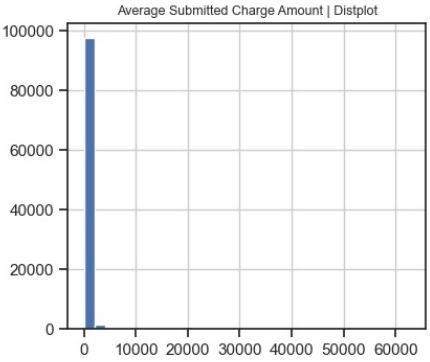

Average Average Medicare Standardized Amount by HCPCS Drug Indicator (Top 20)
0
20
40
60
80
N
Y

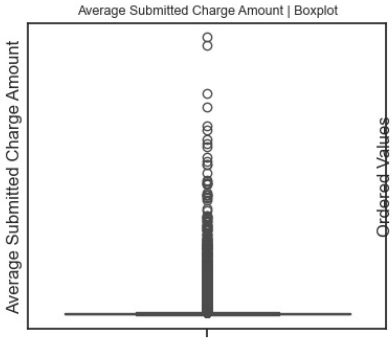

Wordcloud for Credentials of the Provider
mpa
lisw
faao
rnc
point
pac
fa
mspa
do
fac
dpm
md
od
phd
dpt
fnp
crna
aprn
pa
npc
crnp
physician
nurse
because

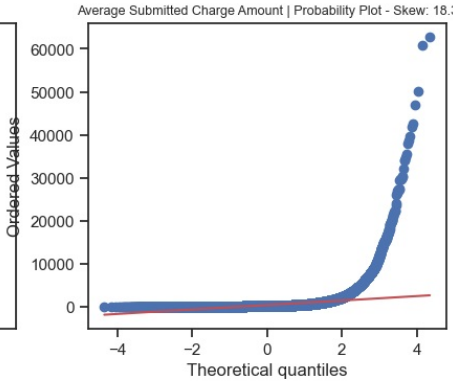

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

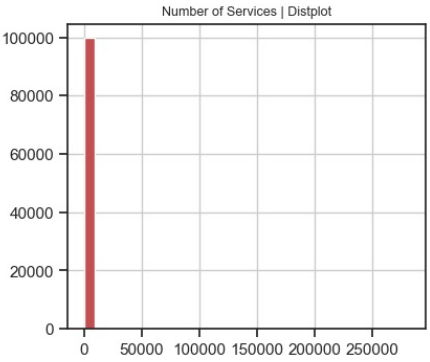
Pair-wise Scatter Plot of all Continuous Variables

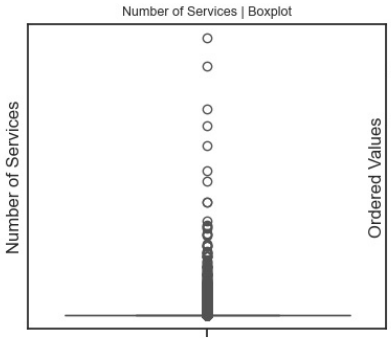


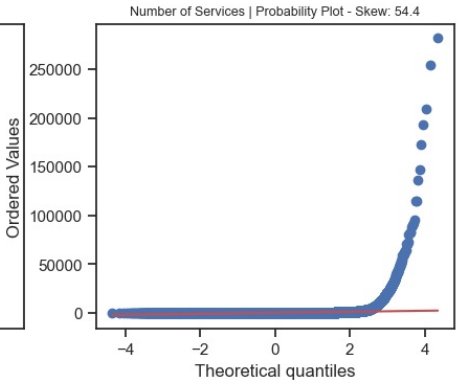


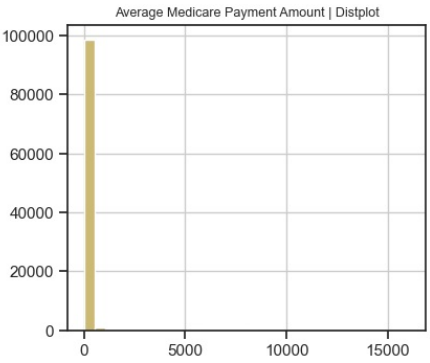


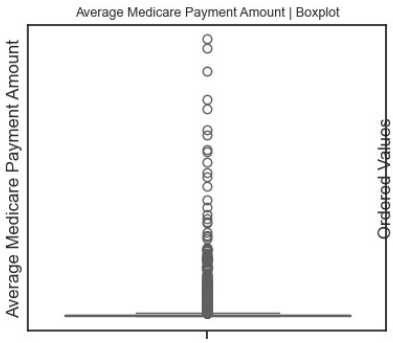


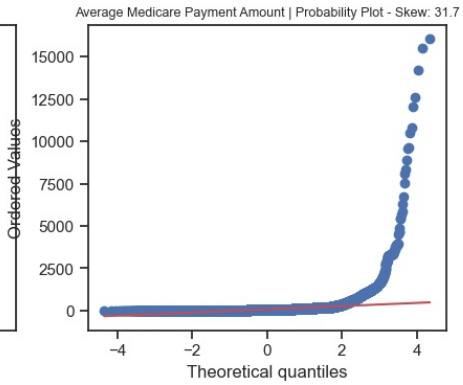


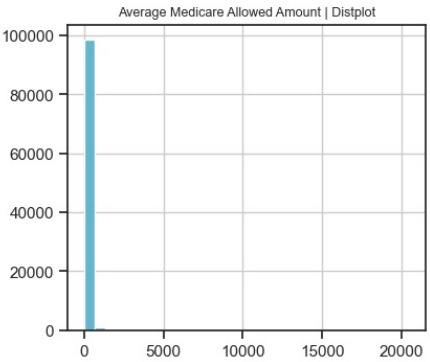


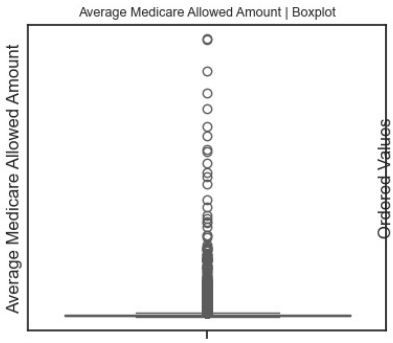


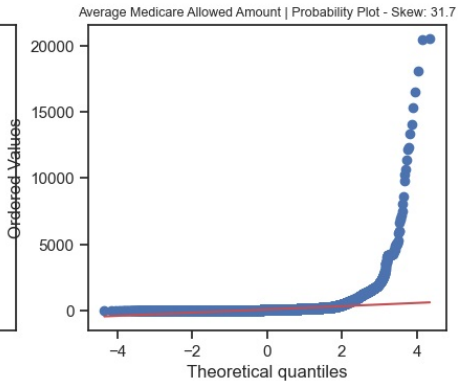


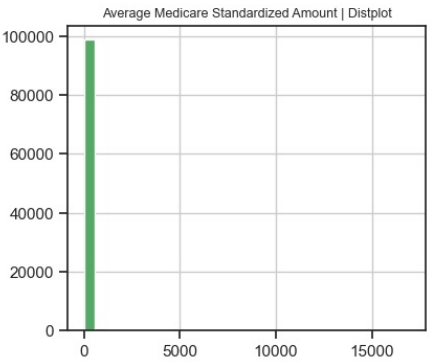


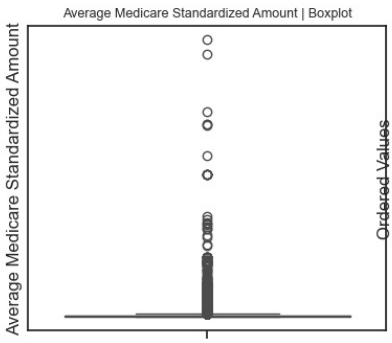


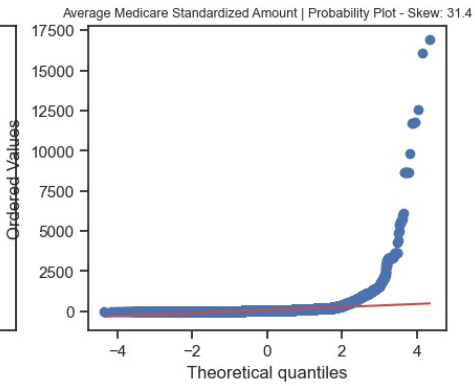




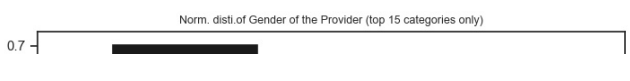


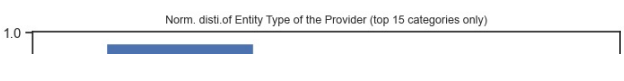


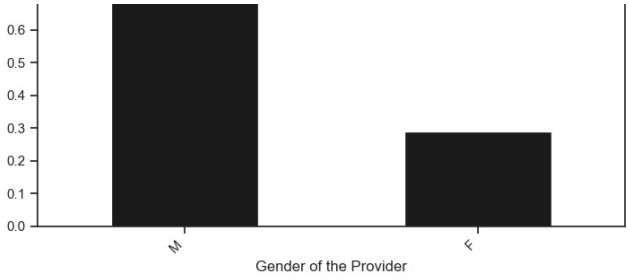


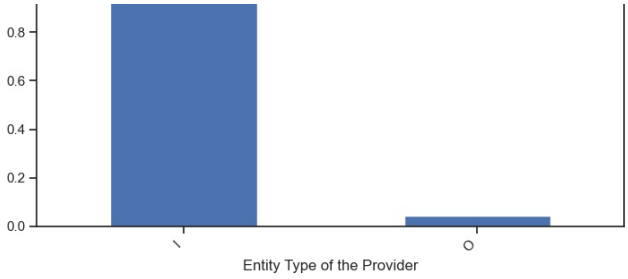


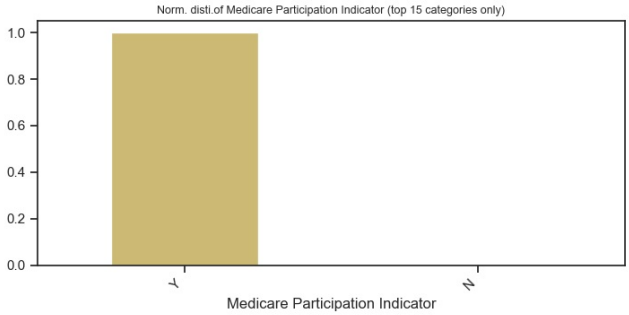
Histograms and Normalized distribitions of all variables

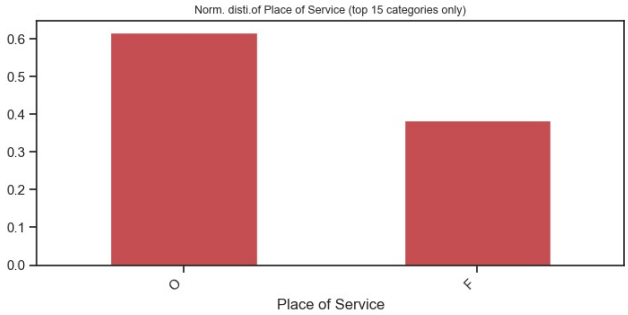


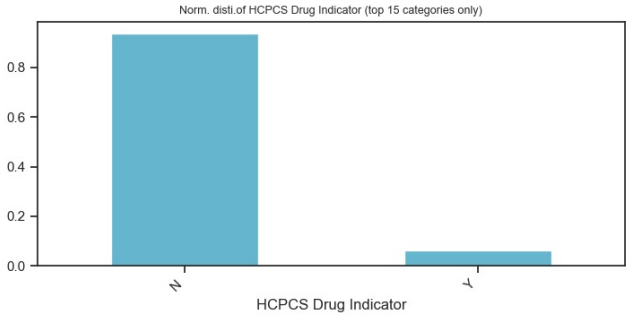


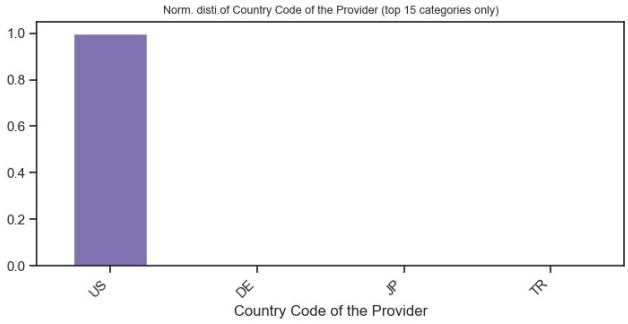

M
F
Gender of the Provider

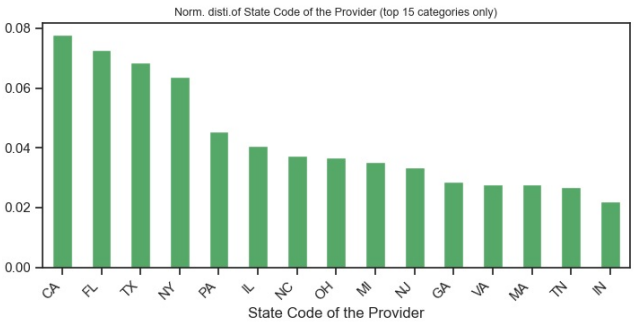

I
O
Entity Type of the Provider

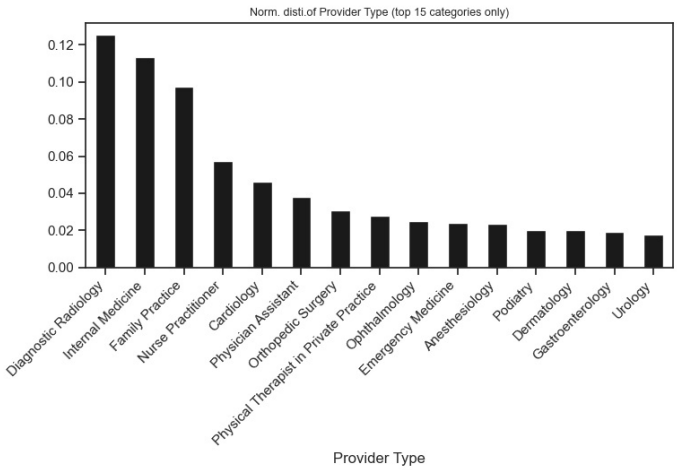

Norm. disti.of Medicare Participation Indicator (top 15 categories only)
Y
N
Medicare Participation Indicator

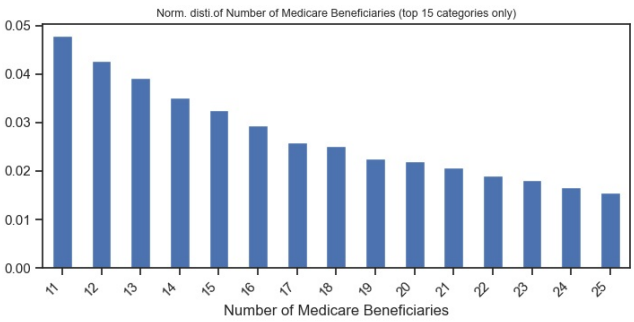

Norm. disti.of Place of Service (top 15 categories only)
O
F
Place of Service

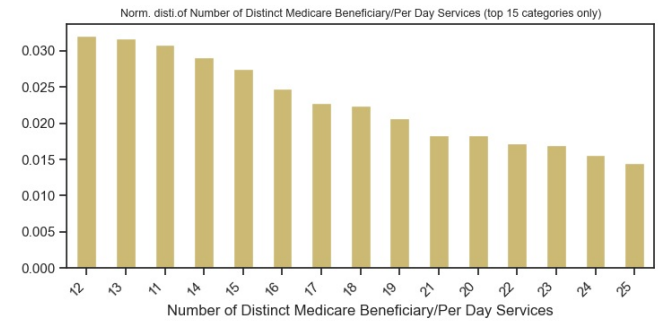

Norm. disti.of HCPCS Drug Indicator (top 15 categories only)
N
Y
HCPCS Drug Indicator


Norm. disti.of Country Code of the Provider (top 15 categories only)
US
DE
JP
TR
Country Code of the Provider

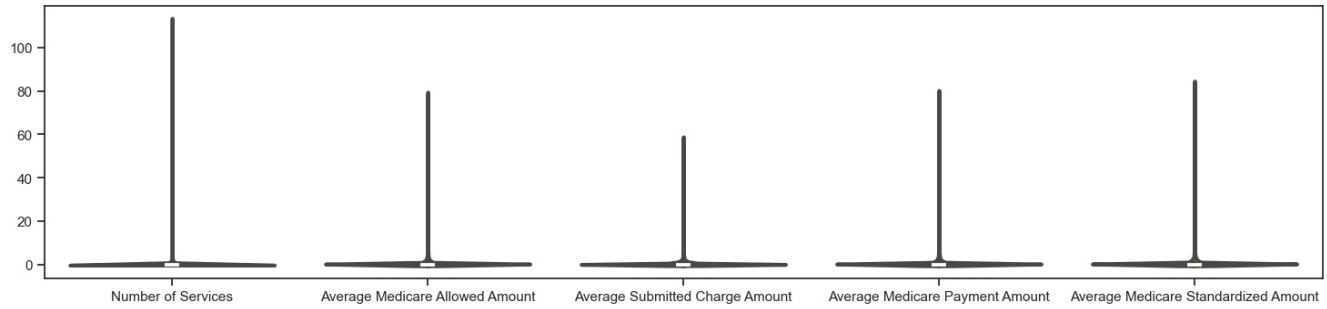

Norm. disti.of State Code of the Provider (top 15 categories only)
CA
FL
TX
NY
PA
IL
NC
OH
MI
NJ
GA
VA
MA
TN
IN
State Code of the Provider


Norm. disti.of Provider Type (top 15 categories only)
Diagnostic Radiology
Internal Medicine
Family Practice
Nurse Practitioner
Cardiology
Physician Assistant
Orthopedic Surgery
Physical Therapist in Private Practice
Ophthalmology
Emergency Medicine
Anesthesiology
Podiatry
Dermatology
Gastroenterology
Urology
Provider Type

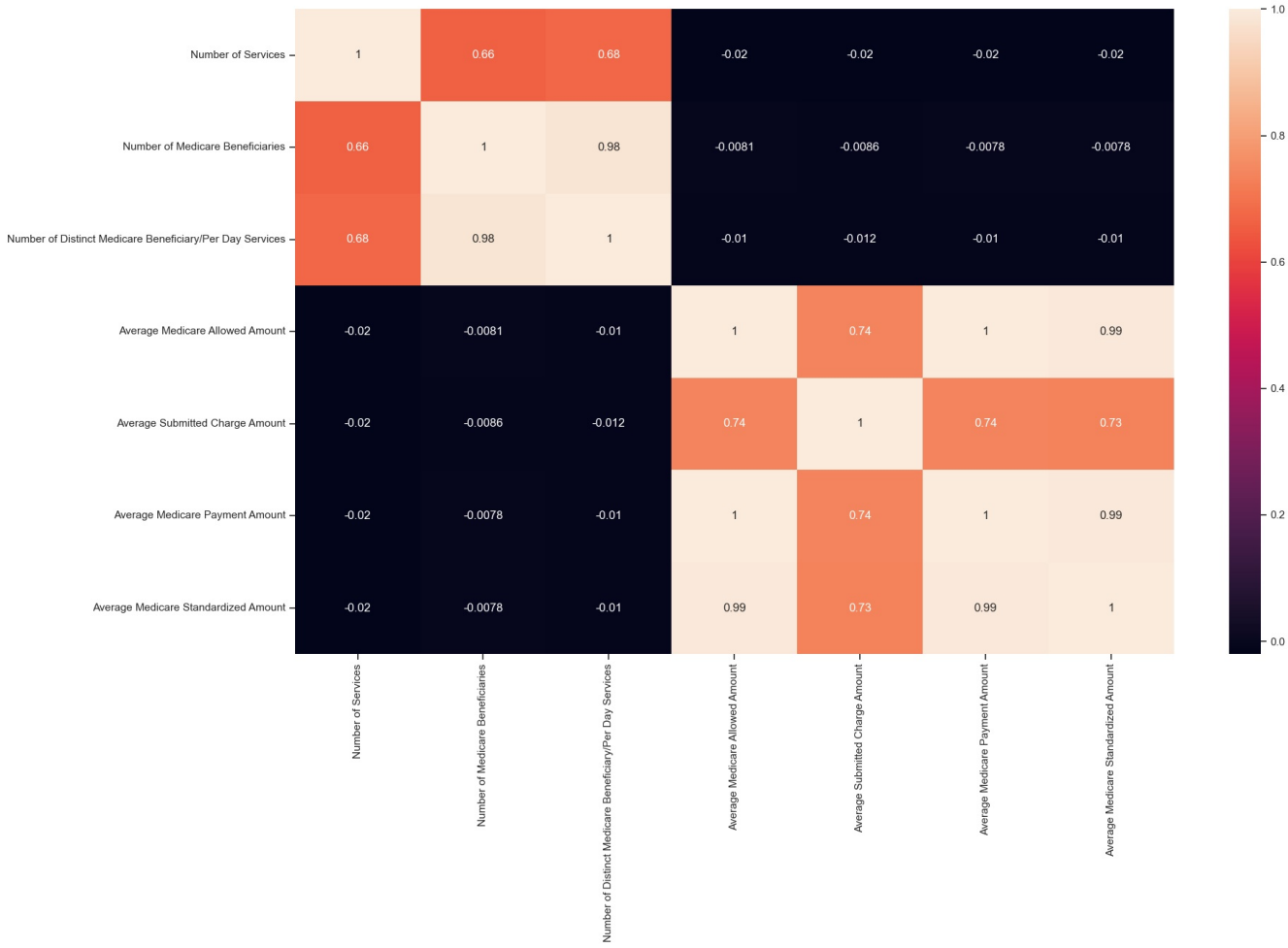

Norm. disti.of Number of Medicare Beneficiaries (top 15 categories only)
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
Number of Medicare Beneficiaries


Norm. disti.of Number of Distinct Medicare Beneficiary/Per Day Services (top 15 categories only)
12
13
11
14
15
16
17
18
19
21
20
22
23
24
25
Number of Distinct Medicare Beneficiary/Per Day Services

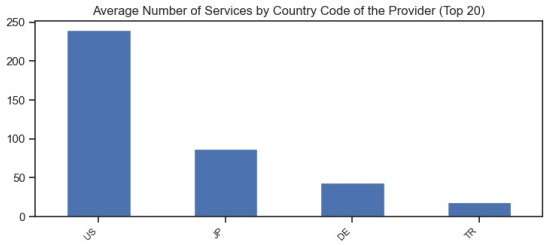
Violin Plot of all Continuous Variables

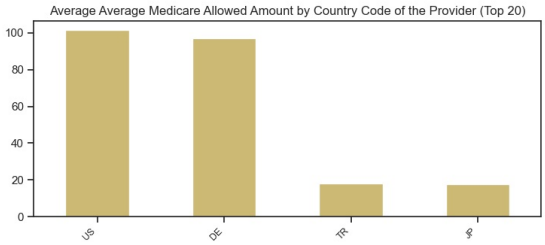


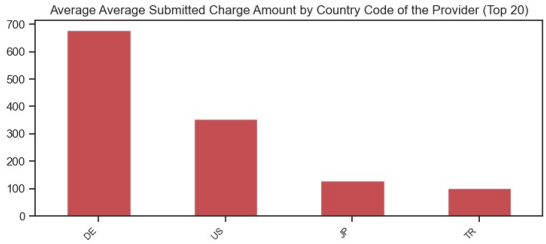
Heatmap of all Numeric Variables including target:

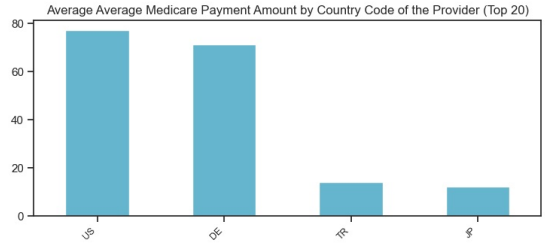


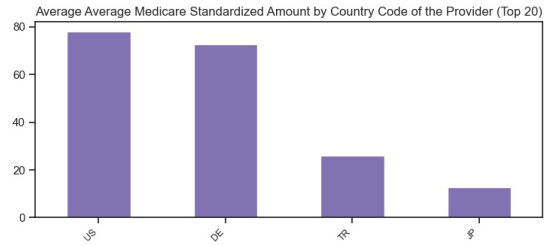
Bar plots for each Continuous by each Categorical variable

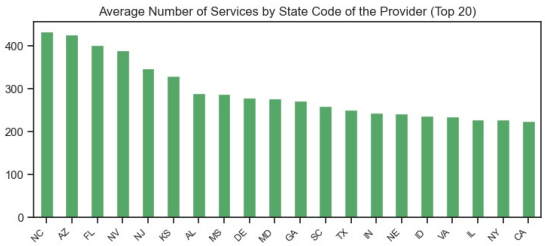

Average Number of Services by Country Code of the Provider (Top 20)
US
JP
DE
TR

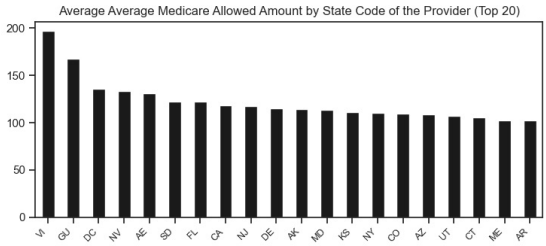

Average Average Medicare Allowed Amount by Country Code of the Provider (Top 20)
US
DE
TR
JP

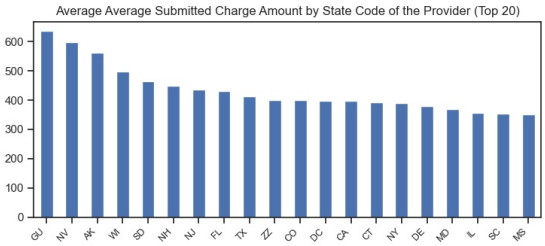

Average Average Submitted Charge Amount by Country Code of the Provider (Top 20)
DE
US
JP
TR

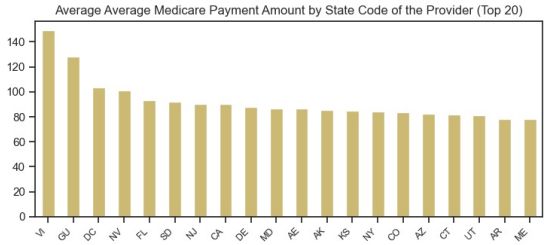

Average Average Medicare Payment Amount by Country Code of the Provider (Top 20)
US
DE
TR
JP

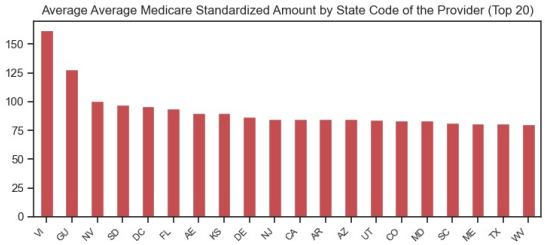

Average Average Medicare Standardized Amount by Country Code of the Provider (Top 20)
US
DE
TR
JP

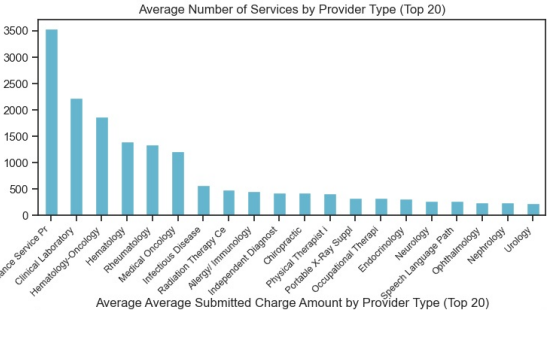

Average Number of Services by State Code of the Provider (Top 20)
NC
AZ
FL
NV
NJ
KS
AL
MS
DE
MD
GA
SC
TX
IN
NE
ID
VA
IL
NY
CA

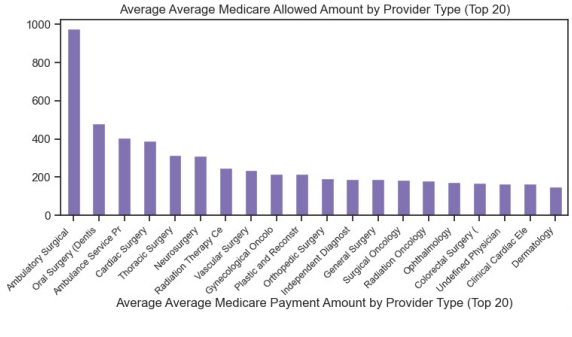

Average Average Medicare Allowed Amount by State Code of the Provider (Top 20)
VI
GU
DC
NV
AE
SD
FL
CA
NJ
DE
AK
MD
KS
NY
CO
AZ
UT
CT
ME
AR

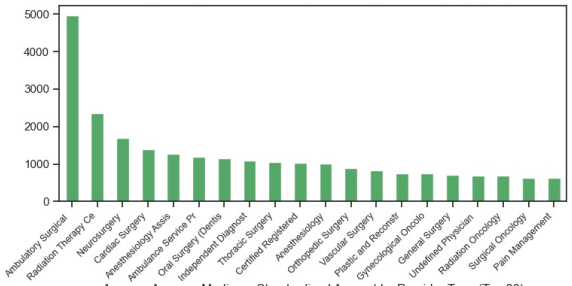

Average Average Submitted Charge Amount by State Code of the Provider (Top 20)
GU
NV
AK
WI
SD
NH
NJ
FL
TX
ZZ
CO
DC
CA
CT
NY
DE
MD
IL
SC
MS

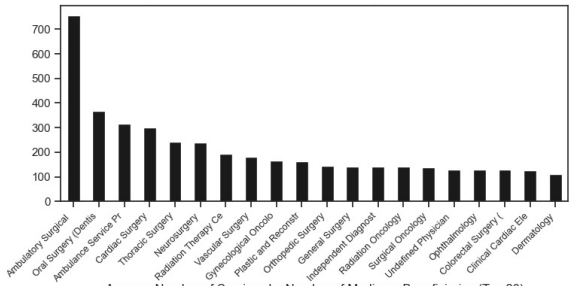

Average Average Medicare Payment Amount by State Code of the Provider (Top 20)
VI
GU
DC
NV
FL
SD
NJ
CA
DE
MD
AE
AK
KS
NY
CO
AZ
CT
UT
AR
ME


Average Average Medicare Standardized Amount by State Code of the Provider (Top 20)
VI
GU
NV
SD
DC
FL
AE
KS
DE
NJ
CA
AR
AZ
UT
CO
MD
SC
ME
TX
WV

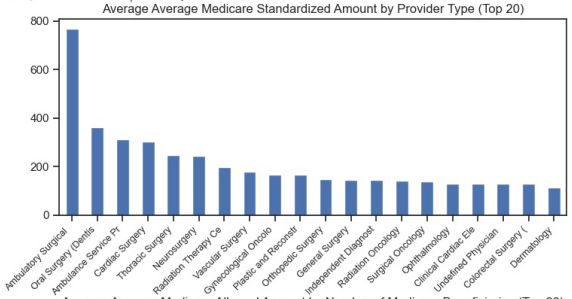

Average Number of Services by Provider Type (Top 20)
Ambulance Service Pr
Clinical Laboratory
Hematology-Oncology
Hematology
Rheumatology
Medical Oncology
Infectious Disease
Radiation Therapy Ca
Allergy/ Immunology
Independent Diagnost
Chiropractic
Physical Therapist i
Portable X-Ray Suppl
Occupational Therapi
Endocrinology
Neurology
Speech Language Path
Ophthalmology
Nephrology
Urology
Average Average Submitted Charge Amount by Provider Type (Top 20)


Average Average Medicare Allowed Amount by Provider Type (Top 20)
Ambulatory Surgical
Oral Surgery (Dentis
Ambulance Service Pr
Cardiac Surgery
Thoracic Surgery
Neurosurgery
Radiation Therapy Ce
Vascular Surgery
Gynecological Oncolo
Plastic and Reconstr
Orthopedic Surgery
Independent Diagnost
General Surgery
Surgical Oncology
Radiation Oncology
Ophthalmology
Colorectal Surgery (
Undefined Physician
Clinical Cardiac Ele
Dermatology
Average Average Medicare Payment Amount by Provider Type (Top 20)

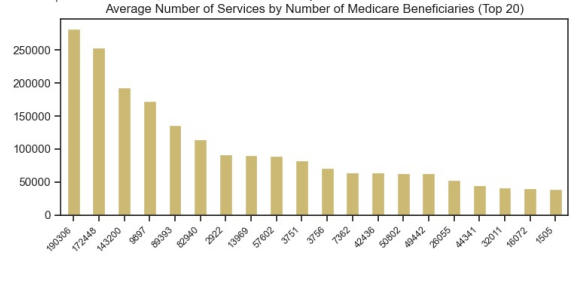




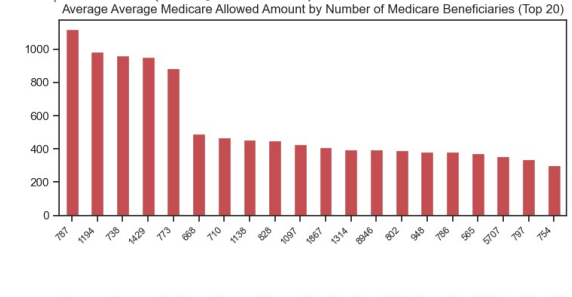
Average Average Medicare Standardized Amount by Provider Type (Top 20)



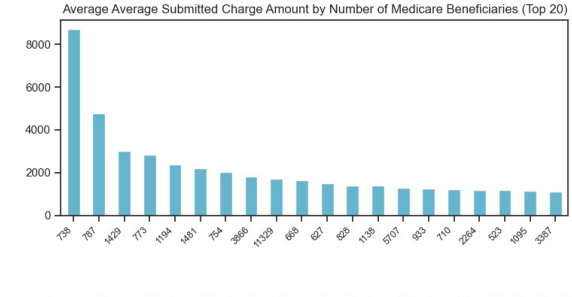
Average Number of Services by Number of Medicare Beneficiaries (Top 20)



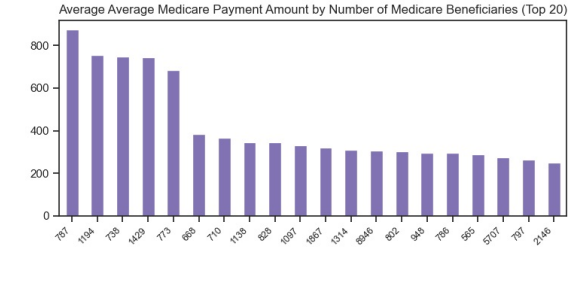
Average Average Medicare Allowed Amount by Number of Medicare Beneficiaries (Top 20)



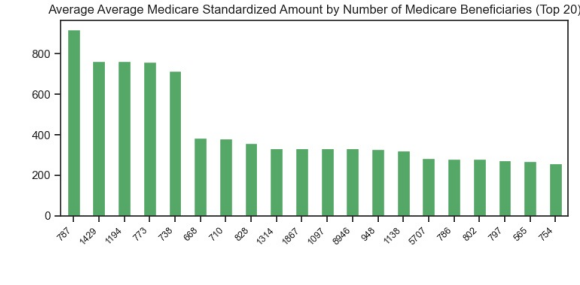
Average Average Submitted Charge Amount by Number of Medicare Beneficiaries (Top 20)



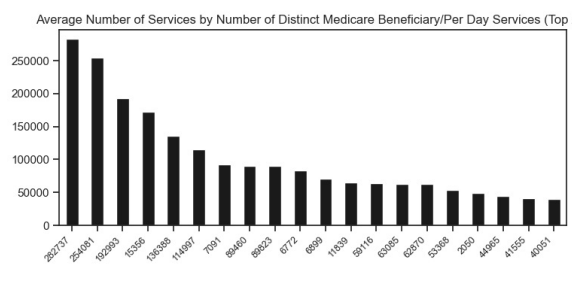
Average Average Medicare Payment Amount by Number of Medicare Beneficiaries (Top 20)



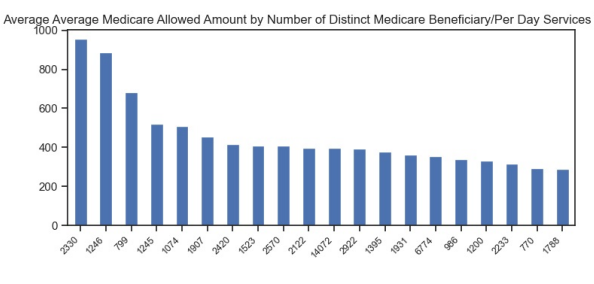
Average Average Medicare Standardized Amount by Number of Medicare Beneficiaries (Top 20)



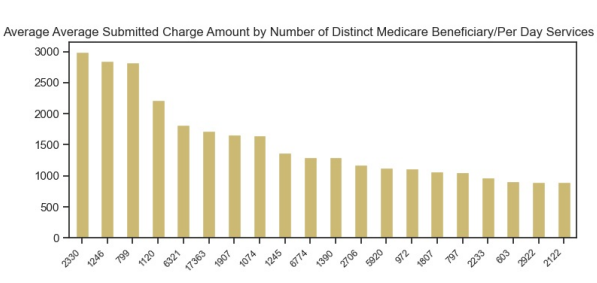
Average Number of Services by Number of Distinct Medicare Beneficiary/Per Day Services (Top 20)



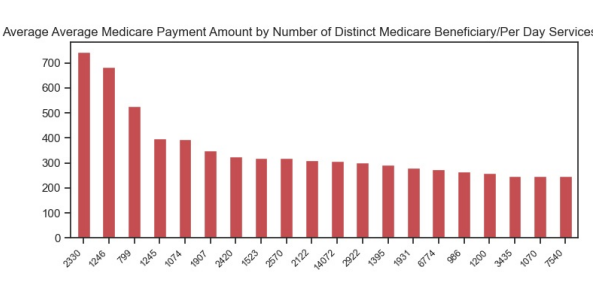
Average Average Medicare Allowed Amount by Number of Distinct Medicare Beneficiary/Per Day Services (Top 20)



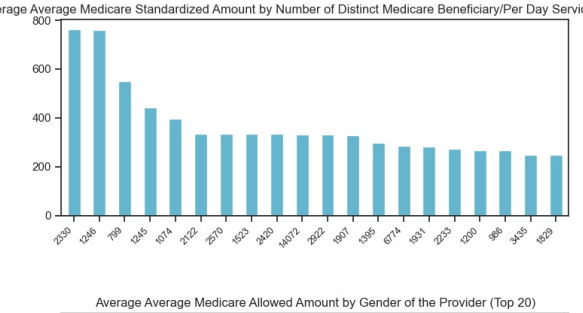
Average Average Submitted Charge Amount by Number of Distinct Medicare Beneficiary/Per Day Services (Top 20)



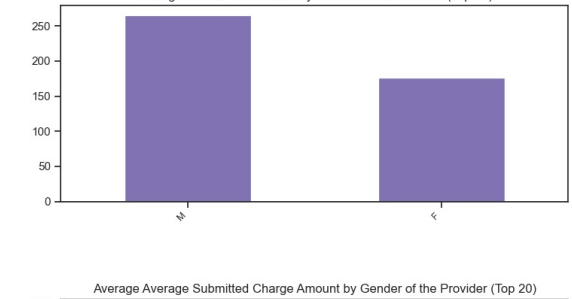
Average Average Medicare Payment Amount by Number of Distinct Medicare Beneficiary/Per Day Services (Top 20)



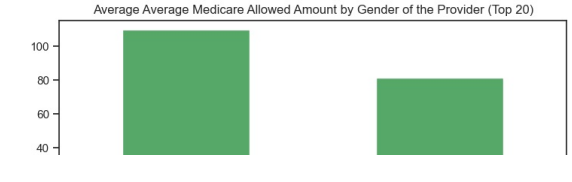
Average Average Medicare Standardized Amount by Number of Distinct Medicare Beneficiary/Per Day Services (Top 20)



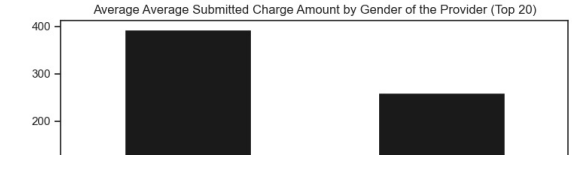
Average Number of Services by Gender of the Provider (Top 20)



Average Average Medicare Allowed Amount by Gender of the Provider (Top 20)

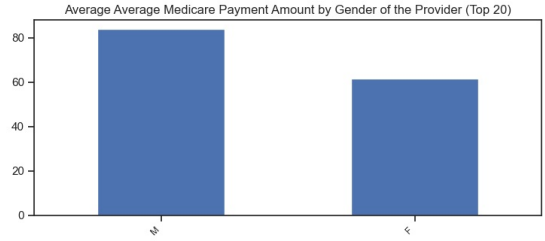


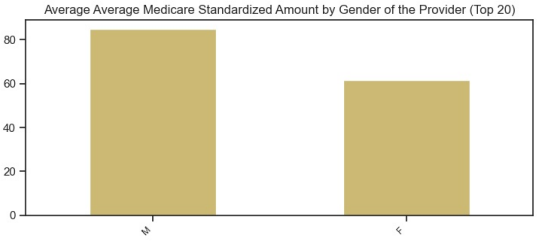
Average Average Submitted Charge Amount by Gender of the Provider (Top 20)

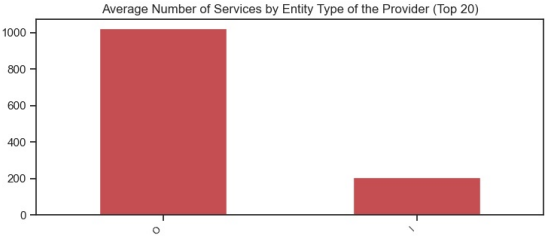


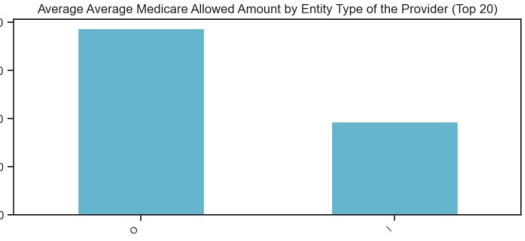

20
0
M
F

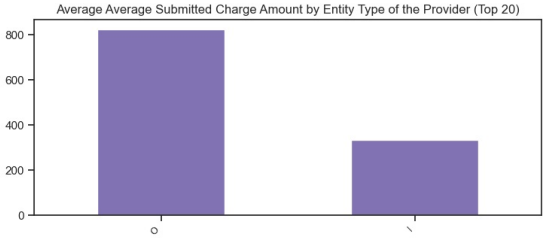

100
0
M
F

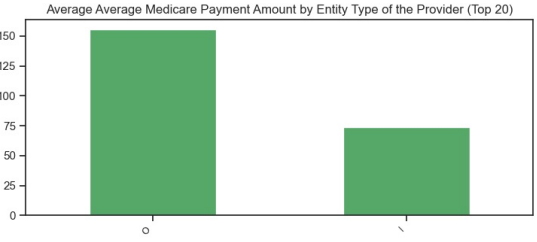

Average Average Medicare Payment Amount by Gender of the Provider (Top 20)
80
60
40
20
0
M
F

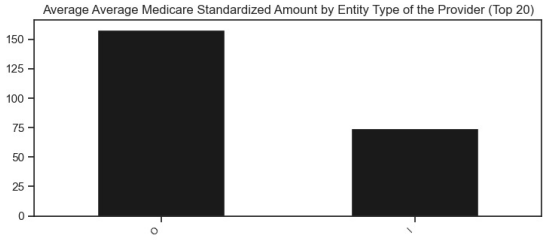

Average Average Medicare Standardized Amount by Gender of the Provider (Top 20)
80
60
40
20
0
M
F

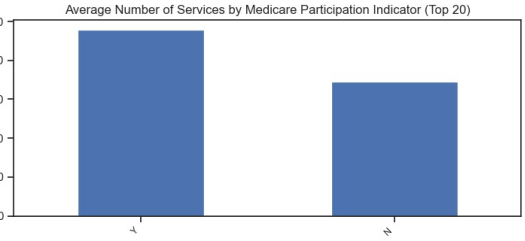

Average Number of Services by Entity Type of the Provider (Top 20)
1000
800
600
400
200
0
O
I

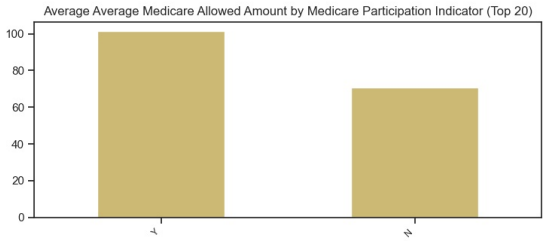

Average Average Medicare Allowed Amount by Entity Type of the Provider (Top 20)
200
150
100
50
0
O
I

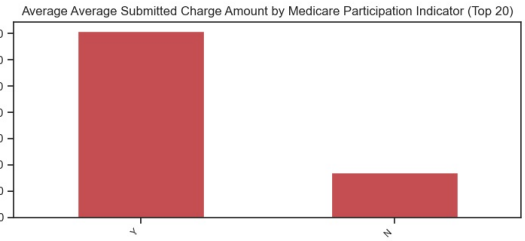

Average Average Submitted Charge Amount by Entity Type of the Provider (Top 20)
800
600
400
200
0
O
I

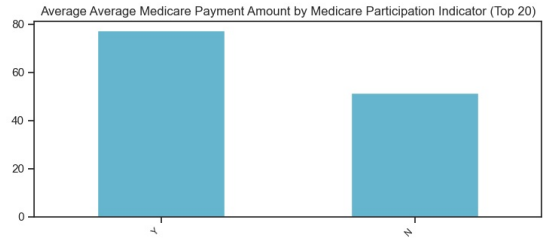

Average Average Medicare Payment Amount by Entity Type of the Provider (Top 20)
150
125
100
75
50
25
0
O
I

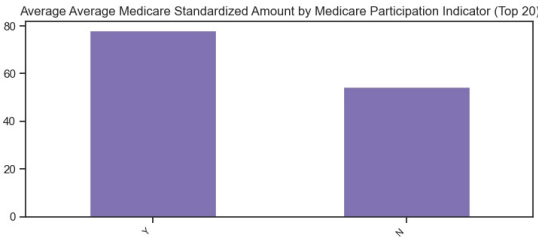

Average Average Medicare Standardized Amount by Entity Type of the Provider (Top 20)
150
125
100
75
50
25
0
O
I

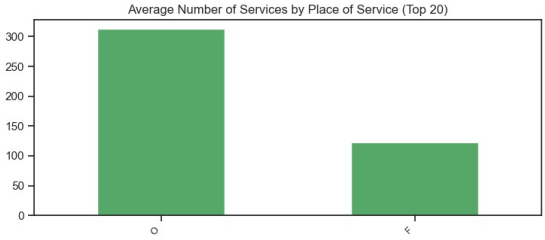

Average Number of Services by Medicare Participation Indicator (Top 20)
250
200
150
100
50
0
Y
N

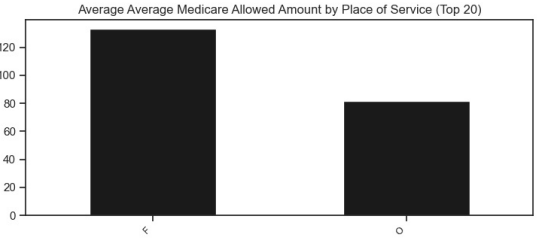

Average Average Medicare Allowed Amount by Medicare Participation Indicator (Top 20)
100
80
60
40
20
0
Y
N

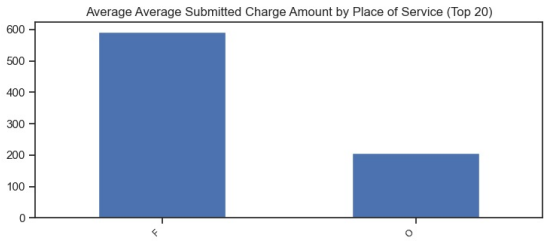

Average Average Submitted Charge Amount by Medicare Participation Indicator (Top 20)
350
300
250
200
150
100
50
0
Y
N

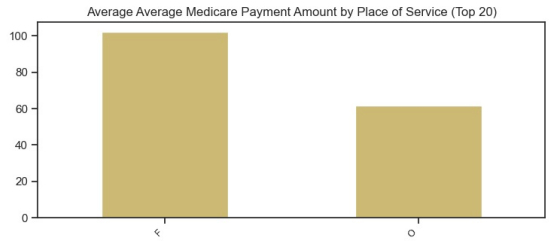

Average Average Medicare Payment Amount by Medicare Participation Indicator (Top 20)
80
60
40
20
0
Y
N

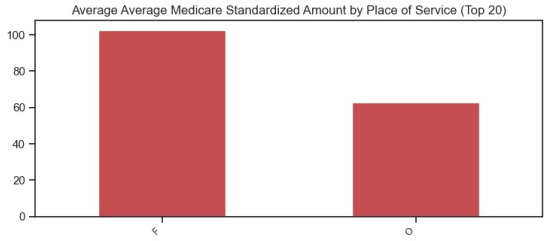

Average Average Medicare Standardized Amount by Medicare Participation Indicator (Top 20)
80
60
40
20
0
Y
N

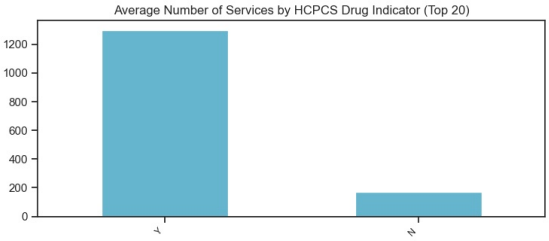

Average Number of Services by Place of Service (Top 20)
300
250
200
150
100
50
0
O
F

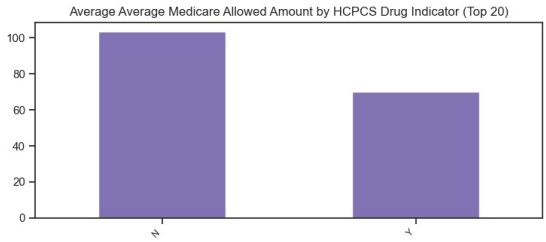

Average Average Medicare Allowed Amount by Place of Service (Top 20)
120
100
80
60
40
20
0
F
O

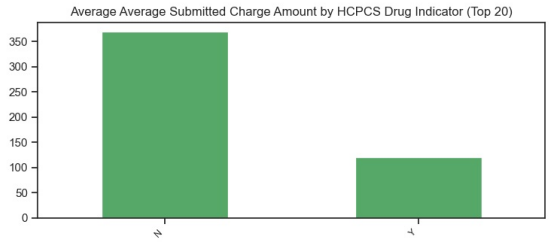


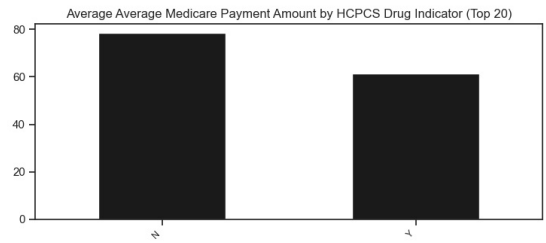


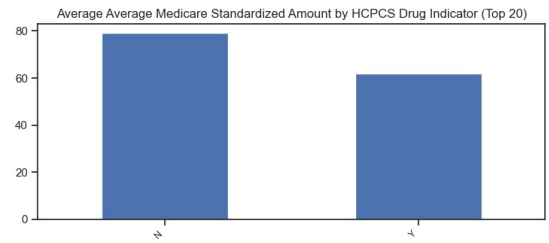






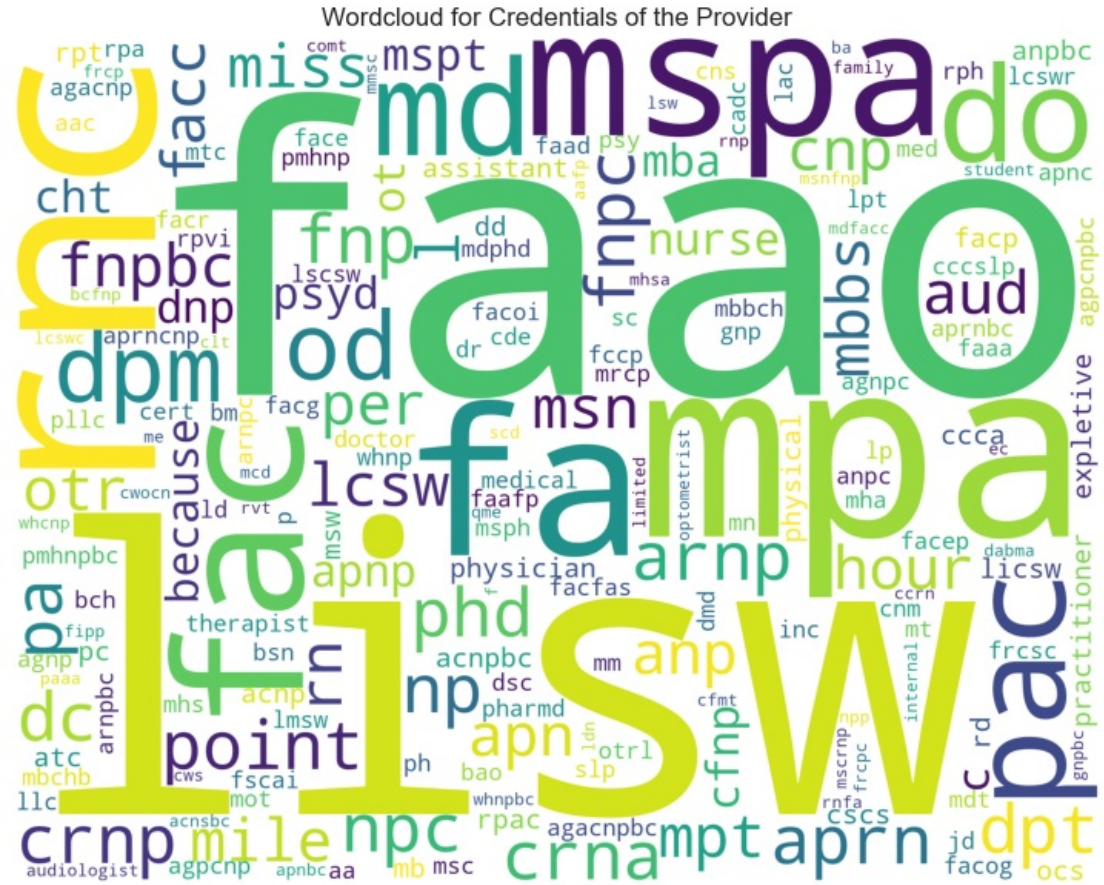






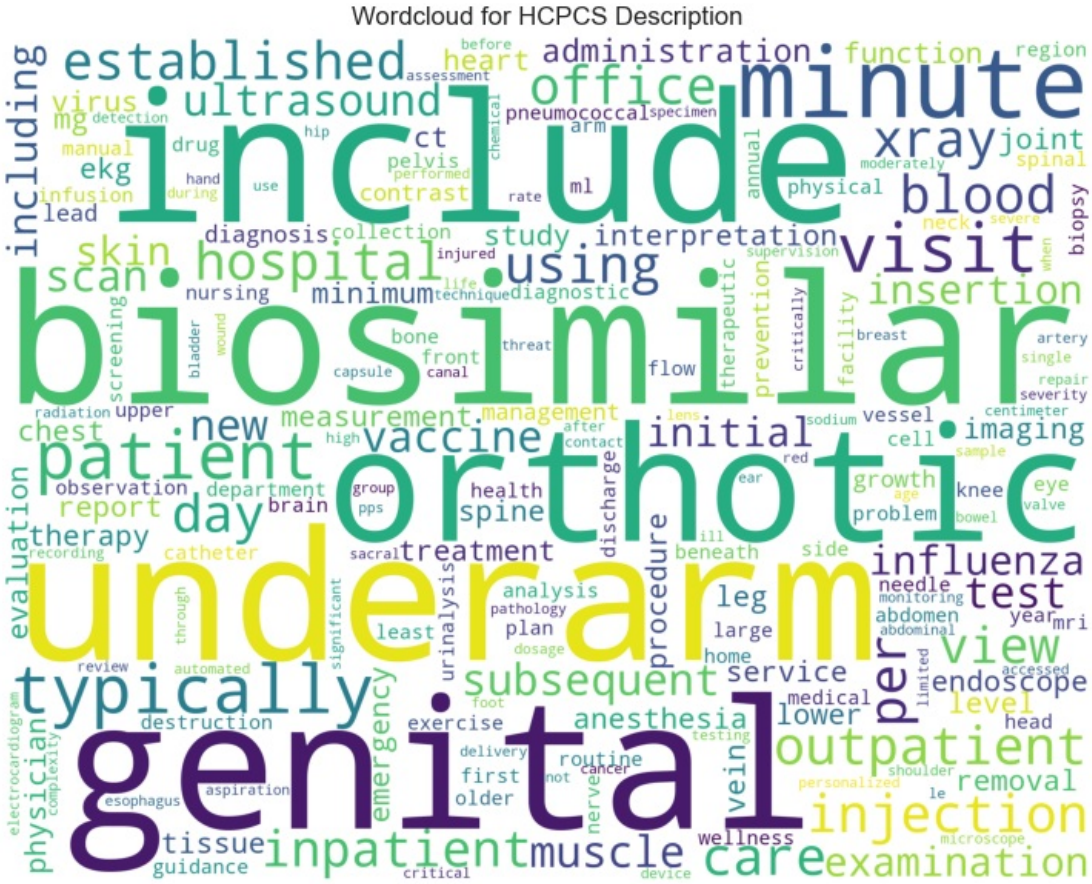

```
[nltk_data] Downloading collection 'popular'
[nltk_data]    |
[nltk_data]    | Downloading package cmudict to
[nltk_data]    |     C:\Users\Pushpendra.PUSHPENDRA-
[nltk_data]    |     PC.001\AppData\Roaming\nltk_data...
[nltk_data]    |   Package cmudict is already up-to-date!
[nltk_data]    | Downloading package gazetteers to
[nltk_data]    |     C:\Users\Pushpendra.PUSHPENDRA-
[nltk_data]    |     PC.001\AppData\Roaming\nltk_data...
[nltk_data]    |   Package gazetteers is already up-to-date!
[nltk_data]    | Downloading package genesis to
[nltk_data]    |     C:\Users\Pushpendra.PUSHPENDRA-
[nltk_data]    |     PC.001\AppData\Roaming\nltk_data...
[nltk_data]    |   Package genesis is already up-to-date!
[nltk_data]    | Downloading package gutenberg to
[nltk_data]    |     C:\Users\Pushpendra.PUSHPENDRA-
[nltk_data]    |     PC.001\AppData\Roaming\nltk_data...
[nltk_data]    |   Package gutenberg is already up-to-date!
[nltk_data]    | Downloading package inaugural to
[nltk_data]    |     C:\Users\Pushpendra.PUSHPENDRA-
[nltk_data]    |     PC.001\AppData\Roaming\nltk_data...
[nltk_data]    |   Package inaugural is already up-to-date!
[nltk_data]    | Downloading package movie_reviews to
[nltk_data]    |     C:\Users\Pushpendra.PUSHPENDRA-
[nltk_data]    |     PC.001\AppData\Roaming\nltk_data...
[nltk_data]    |   Package movie_reviews is already up-to-date!
[nltk_data]    | Downloading package names to
[nltk_data]    |     C:\Users\Pushpendra.PUSHPENDRA-
[nltk_data]    |     PC.001\AppData\Roaming\nltk_data...
[nltk_data]    |   Package names is already up-to-date!
[nltk_data]    | Downloading package shakespeare to
[nltk_data]    |     C:\Users\Pushpendra.PUSHPENDRA-
[nltk_data]    |     PC.001\AppData\Roaming\nltk_data...
[nltk_data]    |   Package shakespeare is already up-to-date!
[nltk_data]    | Downloading package stopwords to
[nltk_data]    |     C:\Users\Pushpendra.PUSHPENDRA-
[nltk_data]    |     PC.001\AppData\Roaming\nltk_data...
[nltk_data]    |   Package stopwords is already up-to-date!
[nltk_data]    | Downloading package treebank to
[nltk_data]    |     C:\Users\Pushpendra.PUSHPENDRA-
[nltk_data]    |     PC.001\AppData\Roaming\nltk_data...
[nltk_data]    |   Package treebank is already up-to-date!
```

```
[nltk_data]     | Downloading package twitter_samples to
[nltk_data]     |     C:\Users\Pushpendra.PUSHPENDRA-
[nltk_data]     |     PC.001\AppData\Roaming\nltk_data...
[nltk_data]     |   Package twitter_samples is already up-to-date!
[nltk_data]     | Downloading package omw to
[nltk_data]     |     C:\Users\Pushpendra.PUSHPENDRA-
[nltk_data]     |     PC.001\AppData\Roaming\nltk_data...
[nltk_data]     |   Package omw is already up-to-date!
[nltk_data]     | Downloading package omw-1.4 to
[nltk_data]     |     C:\Users\Pushpendra.PUSHPENDRA-
[nltk_data]     |     PC.001\AppData\Roaming\nltk_data...
[nltk_data]     |   Package omw-1.4 is already up-to-date!
[nltk_data]     | Downloading package wordnet to
[nltk_data]     |     C:\Users\Pushpendra.PUSHPENDRA-
[nltk_data]     |     PC.001\AppData\Roaming\nltk_data...
[nltk_data]     |   Package wordnet is already up-to-date!
[nltk_data]     | Downloading package wordnet2021 to
[nltk_data]     |     C:\Users\Pushpendra.PUSHPENDRA-
[nltk_data]     |     PC.001\AppData\Roaming\nltk_data...
[nltk_data]     |   Package wordnet2021 is already up-to-date!
[nltk_data]     | Downloading package wordnet31 to
[nltk_data]     |     C:\Users\Pushpendra.PUSHPENDRA-
[nltk_data]     |     PC.001\AppData\Roaming\nltk_data...
[nltk_data]     |   Package wordnet31 is already up-to-date!
[nltk_data]     | Downloading package wordnet_ic to
[nltk_data]     |     C:\Users\Pushpendra.PUSHPENDRA-
[nltk_data]     |     PC.001\AppData\Roaming\nltk_data...
[nltk_data]     |   Package wordnet_ic is already up-to-date!
[nltk_data]     | Downloading package words to
[nltk_data]     |     C:\Users\Pushpendra.PUSHPENDRA-
[nltk_data]     |     PC.001\AppData\Roaming\nltk_data...
[nltk_data]     |   Package words is already up-to-date!
[nltk_data]     | Downloading package maxent_ne_chunker to
[nltk_data]     |     C:\Users\Pushpendra.PUSHPENDRA-
[nltk_data]     |     PC.001\AppData\Roaming\nltk_data...
[nltk_data]     |   Package maxent_ne_chunker is already up-to-date!
[nltk_data]     | Downloading package punkt to
[nltk_data]     |     C:\Users\Pushpendra.PUSHPENDRA-
[nltk_data]     |     PC.001\AppData\Roaming\nltk_data...
[nltk_data]     |   Package punkt is already up-to-date!
[nltk_data]     | Downloading package snowball_data to
[nltk_data]     |     C:\Users\Pushpendra.PUSHPENDRA-
[nltk_data]     |     PC.001\AppData\Roaming\nltk_data...
[nltk_data]     |   Package snowball_data is already up-to-date!
[nltk_data]     | Downloading package averaged_perceptron_tagger to
[nltk_data]     |     C:\Users\Pushpendra.PUSHPENDRA-
[nltk_data]     |     PC.001\AppData\Roaming\nltk_data...
[nltk_data]     |   Package averaged_perceptron_tagger is already up-
[nltk_data]     |       to-date!
[nltk_data]     |
[nltk_data]   Done downloading collection popular
```



Wordcloud for City of the Provider

Wordcloud for HCPCS Code




```
All Plots done
Time to run AutoViz = 181 seconds

 ###################### AUTO VISUALIZATION Completed ########################
```

In [ ]:

In [ ]:

In [ ]: