CMS_Fraud_EDA

December 17, 2025

0.1 CMS Exclusions Dataset Exploratory Data Analysis

[170]:
```
# Import necessary libraries such as numpy and pandas
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import os
```

[171]:
```
# This function will provide basic insights into any data
def quick_summary(df):
    print("=== Missing Values ===")
    print(df.isnull().sum())
    print("\n=== Basic Info ===")
    print(df.info())
    print("\n=== Sample Rows ===")
    print(df.head())
```

0.1.1 OIG Exclusions Data

[172]:
```
# Load the OIG exclusions file into a pandas dataframe
exclusions_file = 'cms_data/leie.csv'
exclusions = pd.read_csv(exclusions_file)
exclusions = exclusions[exclusions.NPI > 0].set_index('NPI')
```

```
/tmp/ipykernel_98/1029917934.py:3: DtypeWarning: Columns (3) have mixed types.
Specify dtype option on import or set low_memory=False.
  exclusions = pd.read_csv(exclusions_file)
```

[173]:
```
quick_summary(exclusions)
```

```
=== Missing Values ===
LASTNAME       524
FIRSTNAME      525
MIDNAME       2475
BUSNAME       7773
GENERAL          0
SPECIALTY      117
UPIN          6897
```

```
DOB            524
ADDRESS          0
CITY             0
STATE            0
ZIP              0
EXCLTYPE         0
EXCLDATE         0
REINDATE         0
WAIVERDATE       0
WVRSTATE      8293
dtype: int64

=== Basic Info ===
<class 'pandas.core.frame.DataFrame'>
Index: 8297 entries, 1972902351 to 1831242650
Data columns (total 17 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   LASTNAME    7773 non-null   object
 1   FIRSTNAME   7772 non-null   object
 2   MIDNAME     5822 non-null   object
 3   BUSNAME     524 non-null    object
 4   GENERAL     8297 non-null   object
 5   SPECIALTY   8180 non-null   object
 6   UPIN        1400 non-null   object
 7   DOB         7773 non-null   float64
 8   ADDRESS     8297 non-null   object
 9   CITY        8297 non-null   object
 10  STATE       8297 non-null   object
 11  ZIP         8297 non-null   int64
 12  EXCLTYPE    8297 non-null   object
 13  EXCLDATE    8297 non-null   int64
 14  REINDATE    8297 non-null   int64
 15  WAIVERDATE  8297 non-null   int64
 16  WVRSTATE    4 non-null      object
dtypes: float64(1), int64(4), object(12)
memory usage: 1.1+ MB
None

=== Sample Rows ===
           LASTNAME FIRSTNAME MIDNAME                      BUSNAME  \
NPI
1972902351      NaN       NaN     NaN  101 FIRST CARE PHARMACY INC
1922348218      NaN       NaN     NaN   184TH STREET PHARMACY CORP
1942476080      NaN       NaN     NaN    A & Y MEDICAL SUPPLY, INC
1275600959      NaN       NaN     NaN    A CARING ALTERNATIVE, INC
1891731758      NaN       NaN     NaN    A FAIR DEAL PHARMACY, INC
```

```
                    GENERAL           SPECIALTY UPIN  DOB  \
NPI
1972902351  OTHER BUSINESS            PHARMACY  NaN  NaN
1922348218  OTHER BUSINESS            PHARMACY  NaN  NaN
1942476080     DME COMPANY       DME - GENERAL  NaN  NaN
1275600959  OTHER BUSINESS  HOME HEALTH AGENCY  NaN  NaN
1891731758  OTHER BUSINESS            PHARMACY  NaN  NaN

                                  ADDRESS          CITY STATE    ZIP  \
NPI
1972902351   C/O 609 W 191ST STREET, APT D      NEW YORK    NY  10040
1922348218                   69 E 184TH ST         BRONX    NY  10468
1942476080       6310 108TH STREET, APT 6J  FOREST HILLS    NY  11375
1275600959        1229 HURON RD E, FLR 6TH     CLEVELAND    OH  44115
1891731758  C/O  P O BOX 329014, #69709-05      BROOKLYN    NY  11232

           EXCLTYPE  EXCLDATE  REINDATE  WAIVERDATE WVRSTATE
NPI
1972902351   1128b8  20220320         0           0      NaN
1922348218   1128a1  20180419         0           0      NaN
1942476080   1128b8  20170518         0           0      NaN
1275600959   1128a1  20130320         0           0      NaN
1891731758   1128b8  20170518         0           0      NaN
```

[174]:
```
# Drops unimportant columns or columns with mostly missing values
exclusions = exclusions.drop(
    columns=['UPIN', 'MIDNAME', 'WVRSTATE', 'BUSNAME', 'REINDATE',␣
 ↪'WAIVERDATE'],
    errors='ignore'
)

exclusions.head()
```

[174]:
```
           LASTNAME FIRSTNAME         GENERAL           SPECIALTY  DOB  \
NPI
1972902351      NaN       NaN  OTHER BUSINESS            PHARMACY  NaN
1922348218      NaN       NaN  OTHER BUSINESS            PHARMACY  NaN
1942476080      NaN       NaN     DME COMPANY       DME - GENERAL  NaN
1275600959      NaN       NaN  OTHER BUSINESS  HOME HEALTH AGENCY  NaN
1891731758      NaN       NaN  OTHER BUSINESS            PHARMACY  NaN

                                  ADDRESS          CITY STATE    ZIP  \
NPI
1972902351   C/O 609 W 191ST STREET, APT D      NEW YORK    NY  10040
1922348218                   69 E 184TH ST         BRONX    NY  10468
1942476080       6310 108TH STREET, APT 6J  FOREST HILLS    NY  11375
1275600959        1229 HURON RD E, FLR 6TH     CLEVELAND    OH  44115
```

```
1891731758  C/O  P O BOX 329014, #69709-05      BROOKLYN    NY  11232

           EXCLTYPE  EXCLDATE
NPI
1972902351   1128b8  20220320
1922348218   1128a1  20180419
1942476080   1128b8  20170518
1275600959   1128a1  20130320
1891731758   1128b8  20170518
```

[175]:
```
exclusions = exclusions.dropna(subset=['SPECIALTY'])
exclusions = exclusions.dropna(subset=['DOB'])
print(exclusions.shape)
```

```
(7723, 11)
```

[176]:
```
date_cols = ['DOB', 'EXCLDATE']
for col in date_cols:
    exclusions[col] = exclusions[col].astype(str).str.replace(r'\.0$', '',␣
 ↪regex=True)
    exclusions[col] = exclusions[col].replace(['nan', 'NaN', '', '00000000'],␣
 ↪np.nan)
    exclusions[col] = pd.to_datetime(exclusions[col], format='%Y%m%d',␣
 ↪errors='coerce')

exclusions.head()
```

[176]:
```
               LASTNAME FIRSTNAME             GENERAL          SPECIALTY  \
NPI
1760461826  ABAD-SANTOS  CRISELDA  PHYSICIAN (MD, DO)         PSYCHIATRY
1477537496        ABADI  JAMSHEED  PHYSICIAN (MD, DO)  INTERNAL MEDICINE
1124292966   ABARIENTOS   CRISPIN  PHYSICIAN (MD, DO)       RHEUMATOLOGY
1376108431        ABBAS     SHAFI      BUS OWNER/EXEC  DME - PROSTHETICS
1194807255      ABBASSI     JADAN  PHYSICIAN (MD, DO)   GENERAL PRACTICE

                  DOB                  ADDRESS              CITY STATE    ZIP  \
NPI
1760461826 1963-12-20           8506 N ADIR DR         WEST HILLS    CA  91304
1477537496 1939-01-10            89 WEEKS ROAD  E WILLISTON PARK    NY  11596
1124292966 1974-09-19  P O BOX 879, #26401-014              AYER    MA   1432
1376108431 1967-06-06            P O BOX 26020          BEAUMONT    TX  26020
1194807255 1944-09-19         115 NELLIS DRIVE             WAYNE    NJ   7470

           EXCLTYPE   EXCLDATE
NPI
1760461826   1128b4 2025-01-20
1477537496   1128b4 2014-05-20
```

```
1124292966   1128a1 2020-06-18
1376108431   1128a1 2025-10-20
1194807255   1128b4 2018-06-20
```

[177]:
```
exclusions['age_at_exclusion'] = (
    (exclusions['EXCLDATE'] - exclusions['DOB']).dt.days / 365.25
)
exclusions['year_excl'] = exclusions['EXCLDATE'].dt.year
```

[178]:
```
excltype_map = {
    '1128a1': 'Conviction - Medicare Fraud',
    '1128a2': 'Patient Abuse/Neglect',
    '1128a3': 'Felony - Drugs',
    '1128a4': 'Felony - Healthcare Fraud',
    '1128b1': 'Misdemeanor - Fraud',
    '1128b2': 'Default on Student Loan',
    '1128b3': 'License Revocation',
    '1128b4': 'Unlawful Claims',
    '1128b5': 'Kickbacks/Bribery',
    '1128b6': 'False Claims',
    '1128b7': 'Obstruction of Audit',
    '1128b8': 'Controlled Substances Violation',
    '1128b9': 'Insurance Fraud',
    '1128b10': 'Unlawful Billing',
    '1128b11': 'Quality of Care Violation',
    '1128b12': 'Civil Monetary Penalty',
    '1128b13': 'False Statement',
    '1128b14': 'Suspension/Exclusion',
    '1128b15': 'License Suspension',
    '1128b16': 'Federal Program Violation',
}


exclusions['EXCLTYPE'] = exclusions['EXCLTYPE'].replace(excltype_map)

exclusions['EXCLTYPE'] = exclusions['EXCLTYPE'].fillna('Other')

print(exclusions['EXCLTYPE'].value_counts())
```

```
EXCLTYPE
Conviction - Medicare Fraud    2921
Unlawful Claims                2388
Felony - Healthcare Fraud       982
Felony - Drugs                  642
Patient Abuse/Neglect           370
Suspension/Exclusion            197
Obstruction of Audit             84
Kickbacks/Bribery                51
```

```
Misdemeanor - Fraud              41
License Revocation               20
False Claims                      7
Default on Student Loan           7
1128Aa                            4
BRCH SA                           4
BRCH CIA                          3
Federal Program Violation         1
1156                              1
Name: count, dtype: int64
```

[179]:
```
quick_summary(exclusions)
```

```
=== Missing Values ===
LASTNAME            0
FIRSTNAME           1
GENERAL             0
SPECIALTY           0
DOB                 0
ADDRESS             0
CITY                0
STATE               0
ZIP                 0
EXCLTYPE            0
EXCLDATE            0
age_at_exclusion    0
year_excl           0
dtype: int64

=== Basic Info ===
<class 'pandas.core.frame.DataFrame'>
Index: 7723 entries, 1760461826 to 1831242650
Data columns (total 13 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   LASTNAME          7723 non-null   object
 1   FIRSTNAME         7722 non-null   object
 2   GENERAL           7723 non-null   object
 3   SPECIALTY         7723 non-null   object
 4   DOB               7723 non-null   datetime64[ns]
 5   ADDRESS           7723 non-null   object
 6   CITY              7723 non-null   object
 7   STATE             7723 non-null   object
 8   ZIP               7723 non-null   int64
 9   EXCLTYPE          7723 non-null   object
 10  EXCLDATE          7723 non-null   datetime64[ns]
 11  age_at_exclusion  7723 non-null   float64
 12  year_excl         7723 non-null   int32
```

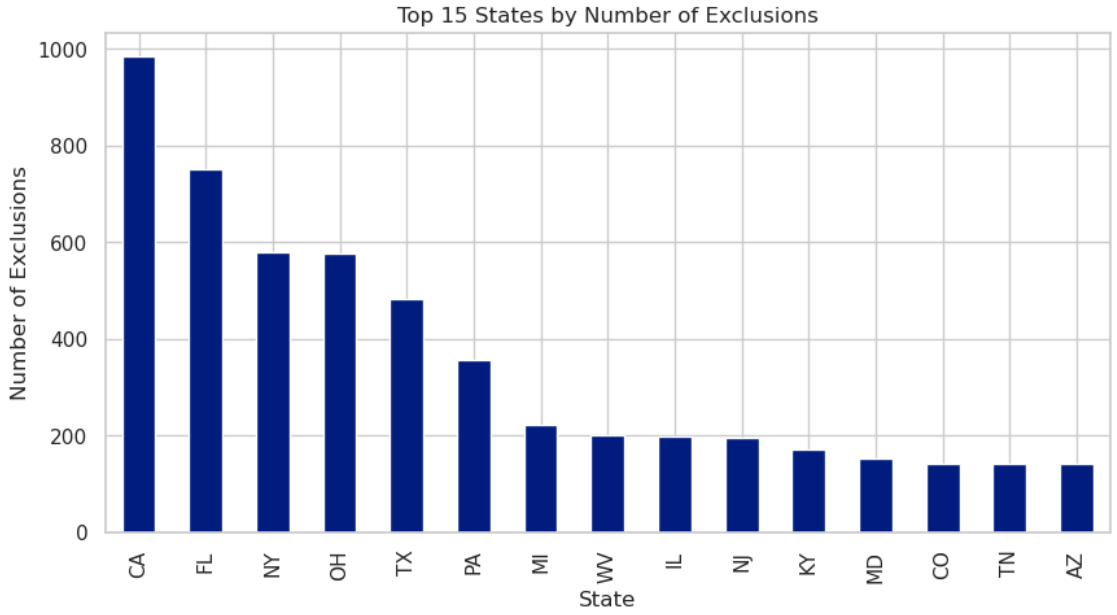
```
dtypes: datetime64[ns](2), float64(1), int32(1), int64(1), object(8)
memory usage: 814.5+ KB
None

=== Sample Rows ===
               LASTNAME FIRSTNAME             GENERAL          SPECIALTY  \
NPI
1760461826  ABAD-SANTOS  CRISELDA  PHYSICIAN (MD, DO)         PSYCHIATRY
1477537496        ABADI  JAMSHEED  PHYSICIAN (MD, DO)  INTERNAL MEDICINE
1124292966   ABARIENTOS   CRISPIN  PHYSICIAN (MD, DO)       RHEUMATOLOGY
1376108431        ABBAS     SHAFI      BUS OWNER/EXEC  DME - PROSTHETICS
1194807255      ABBASSI     JADAN  PHYSICIAN (MD, DO)   GENERAL PRACTICE

                  DOB                  ADDRESS              CITY STATE    ZIP  \
NPI
1760461826 1963-12-20           8506 N ADIR DR        WEST HILLS    CA  91304
1477537496 1939-01-10            89 WEEKS ROAD  E WILLISTON PARK    NY  11596
1124292966 1974-09-19  P O BOX 879, #26401-014              AYER    MA   1432
1376108431 1967-06-06            P O BOX 26020          BEAUMONT    TX  26020
1194807255 1944-09-19         115 NELLIS DRIVE             WAYNE    NJ   7470

                             EXCLTYPE   EXCLDATE  age_at_exclusion  \
NPI
1760461826             Unlawful Claims 2025-01-20         61.086927
1477537496             Unlawful Claims 2014-05-20         75.356605
1124292966  Conviction - Medicare Fraud 2020-06-18         45.746749
1376108431  Conviction - Medicare Fraud 2025-10-20         58.373717
1194807255             Unlawful Claims 2018-06-20         73.749487

            year_excl
NPI
1760461826       2025
1477537496       2014
1124292966       2020
1376108431       2025
1194807255       2018
```

[180]:
```
exclusions['SPECIALTY'].value_counts().head(15)
```

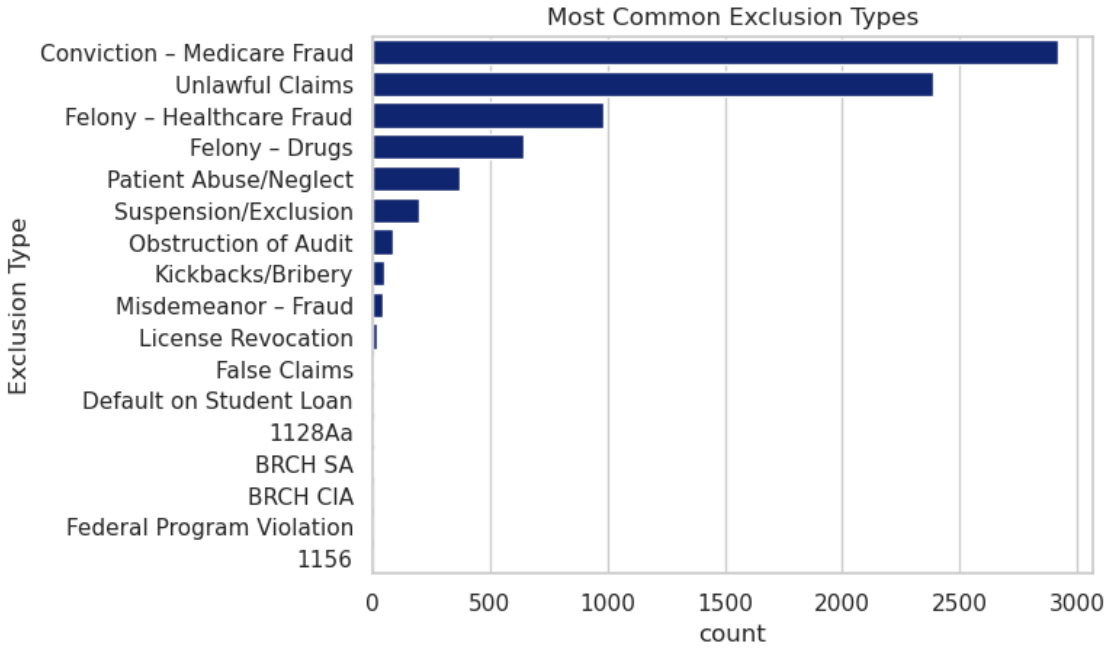
[180]:
```
SPECIALTY
NURSE/NURSES AIDE       826
GENERAL PRACTICE        753
FAMILY PRACTICE         591
INTERNAL MEDICINE       546
CHIROPRACTIC            422
COUNSELOR               397
DENTIST                 334
```

```
PHARMACIST              324
PSYCHIATRY              243
NURSE PRACTITIONER (    188
PAIN MANAGEMENT         179
SOCIAL WORKER           172
PHYSICIAN ASSISTANT     165
THERAPIST               160
PSYCHOLOGY              151
Name: count, dtype: int64
```

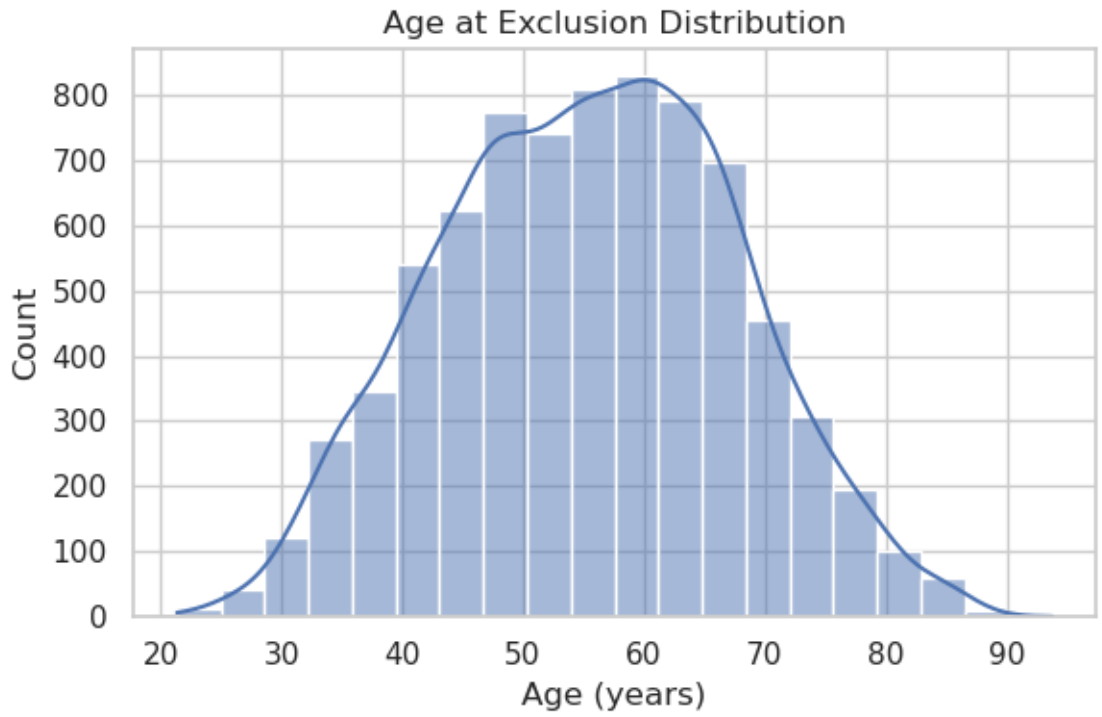
[181]:
```
plt.figure(figsize=(10,5))
exclusions['STATE'].value_counts().head(15).plot(kind='bar')
plt.title('Top 15 States by Number of Exclusions')
plt.xlabel('State')
plt.ylabel('Number of Exclusions')
plt.show()
```

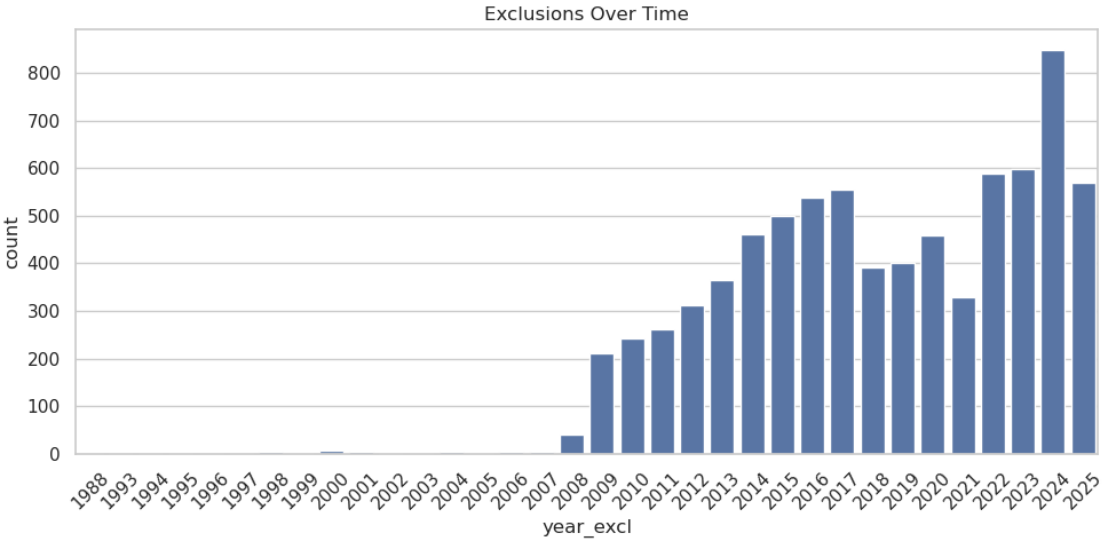


[182]:
```
sns.countplot(y='EXCLTYPE', data=exclusions, order=exclusions['EXCLTYPE'].
 ↪value_counts().index)
plt.ylabel("Exclusion Type")
plt.title('Most Common Exclusion Types')
plt.show()
```



[183]:
```
sns.set(style="whitegrid")

plt.figure(figsize=(6, 4))
sns.histplot(exclusions['age_at_exclusion'].dropna(), kde=True, bins=20)
plt.title("Age at Exclusion Distribution")
plt.xlabel("Age (years)")
plt.ylabel("Count")
plt.tight_layout()
plt.show()

plt.figure(figsize=(10, 5))
sns.countplot(
    data=exclusions,
    x='year_excl',
    order=sorted(exclusions['year_excl'].dropna().unique())
)
plt.title("Exclusions Over Time")
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

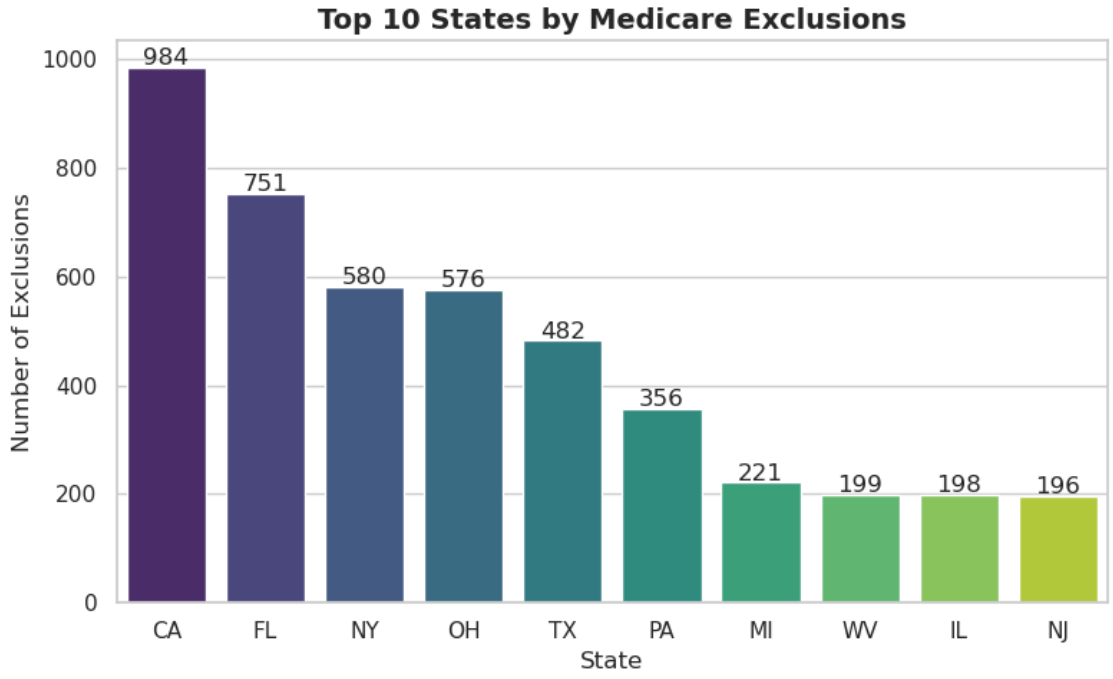




[184]:
```
top_states = (
    exclusions['STATE']
    .value_counts()
    .nlargest(10)
    .reset_index()
```

```
)

top_states.columns = ['STATE', 'COUNT']


plt.figure(figsize=(8, 5))
sns.barplot(data=top_states, x='STATE', y='COUNT', palette='viridis')

plt.title("Top 10 States by Medicare Exclusions", fontsize=14, weight='bold')
plt.xlabel("State")
plt.ylabel("Number of Exclusions")

for i, row in top_states.iterrows():
    plt.text(i, row['COUNT'] + 5, int(row['COUNT']), ha='center')

plt.tight_layout()
plt.show()
```

```
/tmp/ipykernel_98/3599476972.py:12: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in
v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same
effect.

  sns.barplot(data=top_states, x='STATE', y='COUNT', palette='viridis')
```



0.1.2 CMS Medicare Provider Data

[185]:
```
files = os.listdir('cms_data')
files
```

[185]:
```
['MUP_PHY_R25_P07_V10_D19_Prov.csv',
 'MUP_PHY_R25_P07_V10_D22_Prov.csv',
 '.ipynb_checkpoints',
 'MUP_PHY_R25_P07_V10_D21_Prov.csv',
 'leie.csv',
 'y2022_prep.csv',
 'y2019_prep.csv',
 'MUP_PHY_R25_P05_V20_D23_Prov.csv',
 'MUP_PHY_R25_P07_V10_D20_Prov.csv',
 'y2023_prep.csv']
```

[186]:
```
fil = files[6]
data = pd.read_csv('cms_data/'+fil)
```

[187]:
```
data = data.rename(columns={'Rndrng_NPI': 'NPI'})
data = data[data['NPI'] > 0].set_index('NPI')
data= data.drop('Unnamed: 0', axis=1)
```

[188]:
```
data.head()
```

[188]:

NPI	Rndrng_Prvdr_Last_Org_Name	Rndrng_Prvdr_First_Name	Rndrng_Prvdr_MI	Rndrng_Prvdr_Crdntls	Rndrng_Prvdr_Ent_Cd	Rndrng_Prvdr_St1	Rndrng_Prvdr_St2	Rndrng_Prvdr_City	Rndrng_Prvdr_State_Abrvtn
1003000134	CIBULL	THOMAS	L	M.D.	I	2650 RIDGE AVE	EVANSTON HOSPITAL	EVANSTON	IL
1003005315	SMITH	ADAM	B	MD	I	4977 SKYVIEW CT	NaN	TRAVERSE CITY	MI
1003009861	BANNA	MOUSTAFA	NaN	MD	I	5859 W. TALAVI BLVD	SUITE 100	GLENDALE	AZ
1003010570	CHOW	LING	S	M.D.	I	900 E BROADWAY AVE	NaN	BISMARCK	ND
1003017443	BLOKAR	MIRJANA	NaN	M.D.	I	65 BLEECKER ST	12TH FLOOR	NEW YORK	NY

NPI	Rndrng_Prvdr_State_FIPS	Rndrng_Prvdr_Zip5	Rndrng_Prvdr_RUCA	\
1003000134	17	60201.0	1.0	
1003005315	26	49684.0	4.0	
1003009861	4	85306.0	1.0	
1003010570	38	58501.0	1.0	
1003017443	36	10012.0	1.0	

NPI	Rndrng_Prvdr_RUCA_Desc	\
1003000134	Metropolitan area core: primary flow within an…	
1003005315	Micropolitan area core: primary flow within an…	
1003009861	Metropolitan area core: primary flow within an…	
1003010570	Metropolitan area core: primary flow within an…	
1003017443	Metropolitan area core: primary flow within an…	

NPI	Rndrng_Prvdr_Cntry	Rndrng_Prvdr_Type	\
1003000134	US	Pathology	
1003005315	US	Plastic and Reconstructive Surgery	
1003009861	US	Cardiology	
1003010570	US	Internal Medicine	
1003017443	US	Psychiatry	

NPI	Rndrng_Prvdr_Mdcr_Prtcptg_Ind	Tot_HCPCS_Cds	Tot_Benes	Tot_Srvcs	\
1003000134	Y	20	3614	7835.0	
1003005315	Y	51	71	533.0	
1003009861	Y	68	876	3581.0	
1003010570	Y	21	309	833.0	
1003017443	Y	5	64	244.0	

NPI	Tot_Sbmtd_Chrg	Tot_Mdcr_Alowd_Amt	Tot_Mdcr_Pymt_Amt	\
1003000134	1213264.0	298905.09	228078.56	
1003005315	250569.0	86114.40	67712.67	
1003009861	788610.0	361645.81	280928.83	
1003010570	173991.0	84420.40	65891.84	
1003017443	45375.0	23956.84	18245.22	

NPI	Tot_Mdcr_Stdzd_Amt	Drug_Sprsn_Ind	Drug_Tot_HCPCS_Cds	\
1003000134	218268.63	NaN	0.0	
1003005315	70489.71	*	NaN	
1003009861	286768.20	NaN	1.0	
1003010570	66566.70	NaN	0.0	

```
1003017443           16748.76           NaN                0.0

            Drug_Tot_Benes  Drug_Tot_Srvcs  Drug_Sbmtd_Chrg  \
NPI
1003000134             0.0             0.0              0.0
1003005315             NaN             NaN              NaN
1003009861            98.0           392.0          63504.0
1003010570             0.0             0.0              0.0
1003017443             0.0             0.0              0.0

            Drug_Mdcr_Alowd_Amt  Drug_Mdcr_Pymt_Amt  Drug_Mdcr_Stdzd_Amt  \
NPI
1003000134                0.00                0.00                 0.00
1003005315                 NaN                 NaN                  NaN
1003009861            21905.24            17472.35             18074.19
1003010570                0.00                0.00                 0.00
1003017443                0.00                0.00                 0.00

           Med_Sprsn_Ind  Med_Tot_HCPCS_Cds  Med_Tot_Benes  Med_Tot_Srvcs  \
NPI
1003000134           NaN               20.0         3614.0         7835.0
1003005315             #                NaN            NaN            NaN
1003009861           NaN               67.0          876.0         3189.0
1003010570           NaN               21.0          309.0          833.0
1003017443           NaN                5.0           64.0          244.0

            Med_Sbmtd_Chrg  Med_Mdcr_Alowd_Amt  Med_Mdcr_Pymt_Amt  \
NPI
1003000134       1213264.0           298905.09          228078.56
1003005315             NaN                 NaN                NaN
1003009861        725106.0           339740.57          263456.48
1003010570        173991.0            84420.40           65891.84
1003017443         45375.0            23956.84           18245.22

            Med_Mdcr_Stdzd_Amt  Bene_Avg_Age  Bene_Age_LT_65_Cnt  \
NPI
1003000134           218268.63            76                97.0
1003005315                 NaN            73                 NaN
1003009861           268694.01            73                91.0
1003010570            66566.70            76                39.0
1003017443            16748.76            83                 NaN

            Bene_Age_65_74_Cnt  Bene_Age_75_84_Cnt  Bene_Age_GT_84_Cnt  \
NPI
1003000134              1623.0              1333.0               561.0
1003005315                29.0                20.0                 NaN
1003009861               400.0               294.0                91.0
```

```
1003010570               95.0              100.0               75.0
1003017443                NaN               18.0               31.0

            Bene_Feml_Cnt  Bene_Male_Cnt  Bene_Race_Wht_Cnt  \
NPI
1003000134         1767.0         1847.0             3312.0
1003005315           44.0           27.0                NaN
1003009861          410.0          466.0              709.0
1003010570          167.0          142.0              294.0
1003017443           42.0           22.0               48.0

            Bene_Race_Black_Cnt  Bene_Race_API_Cnt  Bene_Race_Hspnc_Cnt  \
NPI
1003000134                41.0               62.0                 43.0
1003005315                 NaN                NaN                  NaN
1003009861                35.0               20.0                 80.0
1003010570                 NaN                NaN                  NaN
1003017443                 NaN                0.0                  NaN

            Bene_Race_NatInd_Cnt  Bene_Race_Othr_Cnt  Bene_Dual_Cnt  \
NPI
1003000134                   0.0               156.0          195.0
1003005315                   NaN                 NaN           14.0
1003009861                  11.0                21.0          164.0
1003010570                   NaN                 NaN           77.0
1003017443                   0.0                 NaN            NaN

            Bene_Ndual_Cnt  Bene_CC_BH_ADHD_OthCD_V1_Pct  \
NPI
1003000134          3419.0                           0.0
1003005315            57.0                           NaN
1003009861           712.0                           NaN
1003010570           232.0                           NaN
1003017443             NaN                           NaN

            Bene_CC_BH_Alcohol_Drug_V1_Pct  Bene_CC_BH_Tobacco_V1_Pct  \
NPI
1003000134                             2.0                        3.0
1003005315                             NaN                        NaN
1003009861                            13.0                       21.0
1003010570                             8.0                       17.0
1003017443                             NaN                        NaN

            Bene_CC_BH_Alz_NonAlzdem_V2_Pct  Bene_CC_BH_Anxiety_V1_Pct  \
NPI
1003000134                              4.0                       13.0
1003005315                              NaN                       35.0
```

```
1003009861                          12.0                      29.0
1003010570                          26.0                      42.0
1003017443                          63.0                      45.0

            Bene_CC_BH_Bipolar_V1_Pct  Bene_CC_BH_Mood_V2_Pct  \
NPI
1003000134                        2.0                    16.0
1003005315                        NaN                    32.0
1003009861                        5.0                    31.0
1003010570                        9.0                    46.0
1003017443                        NaN                    75.0

            Bene_CC_BH_Depress_V1_Pct  Bene_CC_BH_PD_V1_Pct  \
NPI
1003000134                       14.0                   1.0
1003005315                       30.0                   NaN
1003009861                       28.0                   NaN
1003010570                       45.0                   NaN
1003017443                       75.0                   NaN

            Bene_CC_BH_PTSD_V1_Pct  Bene_CC_BH_Schizo_OthPsy_V1_Pct  \
NPI
1003000134                    0.0                             1.0
1003005315                    NaN                             NaN
1003009861                    3.0                             3.0
1003010570                    NaN                             6.0
1003017443                    NaN                            23.0

            Bene_CC_PH_Asthma_V2_Pct  Bene_CC_PH_Afib_V2_Pct  \
NPI
1003000134                       6.0                   16.0
1003005315                       NaN                   24.0
1003009861                      15.0                   38.0
1003010570                      17.0                   35.0
1003017443                       NaN                   30.0

            Bene_CC_PH_Cancer6_V2_Pct  Bene_CC_PH_CKD_V2_Pct  \
NPI
1003000134                       18.0                   15.0
1003005315                       17.0                   25.0
1003009861                       19.0                   42.0
1003010570                       22.0                   45.0
1003017443                       23.0                   30.0

            Bene_CC_PH_COPD_V2_Pct  Bene_CC_PH_Diabetes_V2_Pct  \
NPI
1003000134                    10.0                        21.0
```

```
1003005315                   17.0                         45.0
1003009861                   34.0                         40.0
1003010570                   37.0                         42.0
1003017443                   36.0                         41.0

           Bene_CC_PH_HF_NonIHD_V2_Pct  Bene_CC_PH_Hyperlipidemia_V2_Pct  \
NPI
1003000134                        11.0                              72.0
1003005315                         NaN                              63.0
1003009861                        37.0                              75.0
1003010570                        42.0                              75.0
1003017443                        44.0                              75.0

           Bene_CC_PH_Hypertension_V2_Pct  Bene_CC_PH_IschemicHeart_V2_Pct  \
NPI
1003000134                            64.0                             20.0
1003005315                            75.0                             23.0
1003009861                            75.0                             61.0
1003010570                            75.0                             49.0
1003017443                            75.0                             61.0

           Bene_CC_PH_Osteoporosis_V2_Pct  Bene_CC_PH_Parkinson_V2_Pct  \
NPI
1003000134                            15.0                          3.0
1003005315                             NaN                          NaN
1003009861                            13.0                          2.0
1003010570                            23.0                          NaN
1003017443                            20.0                          NaN

           Bene_CC_PH_Arthritis_V2_Pct  Bene_CC_PH_Stroke_TIA_V2_Pct  \
NPI
1003000134                        42.0                           6.0
1003005315                        48.0                           NaN
1003009861                        50.0                          22.0
1003010570                        59.0                          18.0
1003017443                        70.0                          31.0

           Bene_Avg_Risk_Scre LASTNAME FIRSTNAME MIDNAME BUSNAME  \
NPI
1003000134             1.1246      NaN       NaN     NaN     NaN
1003005315             1.6177    SMITH      ADAM  BRYANT     NaN
1003009861             2.1074      NaN       NaN     NaN     NaN
1003010570             2.1762      NaN       NaN     NaN     NaN
1003017443             2.2548      NaN       NaN     NaN     NaN

                     GENERAL        SPECIALTY UPIN         DOB  \
NPI
```

```
1003000134                 NaN              NaN  NaN         NaN
1003005315  PHYSICIAN (MD, DO)  PLASTIC SURGERY  NaN  19800519.0
1003009861                 NaN              NaN  NaN         NaN
1003010570                 NaN              NaN  NaN         NaN
1003017443                 NaN              NaN  NaN         NaN

                      ADDRESS        CITY STATE      ZIP EXCLTYPE  \
NPI
1003000134                NaN         NaN   NaN      NaN      UNK
1003005315  2136 FORD PKWY 5103  SAINT PAUL    MN  55116.0   1128a1
1003009861                NaN         NaN   NaN      NaN      UNK
1003010570                NaN         NaN   NaN      NaN      UNK
1003017443                NaN         NaN   NaN      NaN      UNK

              EXCLDATE  REINDATE  WAIVERDATE WVRSTATE  \
NPI
1003000134         NaN       NaN         NaN      NaN
1003005315  20221220.0       0.0         0.0      NaN
1003009861         NaN       NaN         NaN      NaN
1003010570         NaN       NaN         NaN      NaN
1003017443         NaN       NaN         NaN      NaN

            rat_Drug_Mdcr_Alowd_Amt_Drug_Mdcr_Pymt_Amt  \
NPI
1003000134                                        NaN
1003005315                                        NaN
1003009861                                   1.253709
1003010570                                        NaN
1003017443                                        NaN

            rat_Tot_Mdcr_Alowd_Amt_Med_Sbmtd_Chrg  \
NPI
1003000134                               0.246364
1003005315                                    NaN
1003009861                               0.498749
1003010570                               0.485200
1003017443                               0.527974

            rat_Drug_Tot_Benes_Tot_Benes  excluded      rand
NPI
1003000134                      0.000000         0  0.994100
1003005315                           NaN         1  0.253627
1003009861                      0.111872         0  0.994092
1003010570                      0.000000         0  0.994872
1003017443                      0.000000         0  0.991242
```

[189]:
```
quick_summary(data)
```

```
=== Missing Values ===
Rndrng_Prvdr_Last_Org_Name                        0
Rndrng_Prvdr_First_Name                         649
Rndrng_Prvdr_MI                                4210
Rndrng_Prvdr_Crdntls                           1302
Rndrng_Prvdr_Ent_Cd                               0
                                                 …
rat_Drug_Mdcr_Alowd_Amt_Drug_Mdcr_Pymt_Amt     9475
rat_Tot_Mdcr_Alowd_Amt_Med_Sbmtd_Chrg          1410
rat_Drug_Tot_Benes_Tot_Benes                   1410
excluded                                          0
rand                                              0
Length: 102, dtype: int64

=== Basic Info ===
<class 'pandas.core.frame.DataFrame'>
Index: 12470 entries, 1003000134 to 1992999759
Columns: 102 entries, Rndrng_Prvdr_Last_Org_Name to rand
dtypes: float64(70), int64(5), object(27)
memory usage: 9.8+ MB
None

=== Sample Rows ===
           Rndrng_Prvdr_Last_Org_Name Rndrng_Prvdr_First_Name Rndrng_Prvdr_MI  \
NPI
1003000134                   CIBULL                  THOMAS               L
1003005315                    SMITH                    ADAM               B
1003009861                    BANNA                MOUSTAFA             NaN
1003010570                     CHOW                    LING               S
1003017443                   BLOKAR                 MIRJANA             NaN

           Rndrng_Prvdr_Crdntls Rndrng_Prvdr_Ent_Cd     Rndrng_Prvdr_St1  \
NPI
1003000134                M.D.                    I       2650 RIDGE AVE
1003005315                  MD                    I      4977 SKYVIEW CT
1003009861                  MD                    I  5859 W. TALAVI BLVD
1003010570                M.D.                    I   900 E BROADWAY AVE
1003017443                M.D.                    I       65 BLEECKER ST

             Rndrng_Prvdr_St2 Rndrng_Prvdr_City Rndrng_Prvdr_State_Abrvtn  \
NPI
1003000134  EVANSTON HOSPITAL          EVANSTON                         IL
1003005315                NaN     TRAVERSE CITY                         MI
1003009861          SUITE 100          GLENDALE                         AZ
1003010570                NaN          BISMARCK                         ND
1003017443         12TH FLOOR          NEW YORK                         NY

            Rndrng_Prvdr_State_FIPS  Rndrng_Prvdr_Zip5  Rndrng_Prvdr_RUCA  \
```

```
NPI
1003000134                       17            60201.0                1.0
1003005315                       26            49684.0                4.0
1003009861                        4            85306.0                1.0
1003010570                       38            58501.0                1.0
1003017443                       36            10012.0                1.0

                                      Rndrng_Prvdr_RUCA_Desc  \
NPI
1003000134  Metropolitan area core: primary flow within an…
1003005315  Micropolitan area core: primary flow within an…
1003009861  Metropolitan area core: primary flow within an…
1003010570  Metropolitan area core: primary flow within an…
1003017443  Metropolitan area core: primary flow within an…

           Rndrng_Prvdr_Cntry                   Rndrng_Prvdr_Type  \
NPI
1003000134                 US                           Pathology
1003005315                 US  Plastic and Reconstructive Surgery
1003009861                 US                         Cardiology
1003010570                 US                  Internal Medicine
1003017443                 US                         Psychiatry

           Rndrng_Prvdr_Mdcr_Prtcptg_Ind  Tot_HCPCS_Cds  Tot_Benes  Tot_Srvcs  \
NPI
1003000134                             Y             20       3614     7835.0
1003005315                             Y             51         71      533.0
1003009861                             Y             68        876     3581.0
1003010570                             Y             21        309      833.0
1003017443                             Y              5         64      244.0

            Tot_Sbmtd_Chrg  Tot_Mdcr_Alowd_Amt  Tot_Mdcr_Pymt_Amt  \
NPI
1003000134       1213264.0           298905.09          228078.56
1003005315        250569.0            86114.40           67712.67
1003009861        788610.0           361645.81          280928.83
1003010570        173991.0            84420.40           65891.84
1003017443         45375.0            23956.84           18245.22

            Tot_Mdcr_Stdzd_Amt Drug_Sprsn_Ind  Drug_Tot_HCPCS_Cds  \
NPI
1003000134          218268.63            NaN                 0.0
1003005315           70489.71              *                 NaN
1003009861          286768.20            NaN                 1.0
1003010570           66566.70            NaN                 0.0
1003017443           16748.76            NaN                 0.0

            Drug_Tot_Benes  Drug_Tot_Srvcs  Drug_Sbmtd_Chrg  \
```

```
NPI
1003000134             0.0             0.0             0.0
1003005315             NaN             NaN             NaN
1003009861            98.0           392.0         63504.0
1003010570             0.0             0.0             0.0
1003017443             0.0             0.0             0.0

            Drug_Mdcr_Alowd_Amt  Drug_Mdcr_Pymt_Amt  Drug_Mdcr_Stdzd_Amt  \
NPI
1003000134                 0.00                0.00                 0.00
1003005315                  NaN                 NaN                  NaN
1003009861             21905.24            17472.35             18074.19
1003010570                 0.00                0.00                 0.00
1003017443                 0.00                0.00                 0.00

           Med_Sprsn_Ind  Med_Tot_HCPCS_Cds  Med_Tot_Benes  Med_Tot_Srvcs  \
NPI
1003000134           NaN               20.0         3614.0         7835.0
1003005315             #                NaN            NaN            NaN
1003009861           NaN               67.0          876.0         3189.0
1003010570           NaN               21.0          309.0          833.0
1003017443           NaN                5.0           64.0          244.0

            Med_Sbmtd_Chrg  Med_Mdcr_Alowd_Amt  Med_Mdcr_Pymt_Amt  \
NPI
1003000134       1213264.0           298905.09          228078.56
1003005315             NaN                 NaN                NaN
1003009861        725106.0           339740.57          263456.48
1003010570        173991.0            84420.40           65891.84
1003017443         45375.0            23956.84           18245.22

            Med_Mdcr_Stdzd_Amt  Bene_Avg_Age  Bene_Age_LT_65_Cnt  \
NPI
1003000134           218268.63            76                97.0
1003005315                 NaN            73                 NaN
1003009861           268694.01            73                91.0
1003010570            66566.70            76                39.0
1003017443            16748.76            83                 NaN

            Bene_Age_65_74_Cnt  Bene_Age_75_84_Cnt  Bene_Age_GT_84_Cnt  \
NPI
1003000134              1623.0              1333.0               561.0
1003005315                29.0                20.0                 NaN
1003009861               400.0               294.0                91.0
1003010570                95.0               100.0                75.0
1003017443                 NaN                18.0                31.0

            Bene_Feml_Cnt  Bene_Male_Cnt  Bene_Race_Wht_Cnt  \
```

```
NPI
1003000134         1767.0         1847.0             3312.0
1003005315           44.0           27.0                NaN
1003009861          410.0          466.0              709.0
1003010570          167.0          142.0              294.0
1003017443           42.0           22.0               48.0

            Bene_Race_Black_Cnt  Bene_Race_API_Cnt  Bene_Race_Hspnc_Cnt  \
NPI
1003000134                41.0               62.0                 43.0
1003005315                 NaN                NaN                  NaN
1003009861                35.0               20.0                 80.0
1003010570                 NaN                NaN                  NaN
1003017443                 NaN                0.0                  NaN

            Bene_Race_NatInd_Cnt  Bene_Race_Othr_Cnt  Bene_Dual_Cnt  \
NPI
1003000134                   0.0               156.0          195.0
1003005315                   NaN                 NaN           14.0
1003009861                  11.0                21.0          164.0
1003010570                   NaN                 NaN           77.0
1003017443                   0.0                 NaN            NaN

            Bene_Ndual_Cnt  Bene_CC_BH_ADHD_OthCD_V1_Pct  \
NPI
1003000134          3419.0                           0.0
1003005315            57.0                           NaN
1003009861           712.0                           NaN
1003010570           232.0                           NaN
1003017443             NaN                           NaN

            Bene_CC_BH_Alcohol_Drug_V1_Pct  Bene_CC_BH_Tobacco_V1_Pct  \
NPI
1003000134                             2.0                        3.0
1003005315                             NaN                        NaN
1003009861                            13.0                       21.0
1003010570                             8.0                       17.0
1003017443                             NaN                        NaN

            Bene_CC_BH_Alz_NonAlzdem_V2_Pct  Bene_CC_BH_Anxiety_V1_Pct  \
NPI
1003000134                              4.0                       13.0
1003005315                              NaN                       35.0
1003009861                             12.0                       29.0
1003010570                             26.0                       42.0
1003017443                             63.0                       45.0

            Bene_CC_BH_Bipolar_V1_Pct  Bene_CC_BH_Mood_V2_Pct  \
```

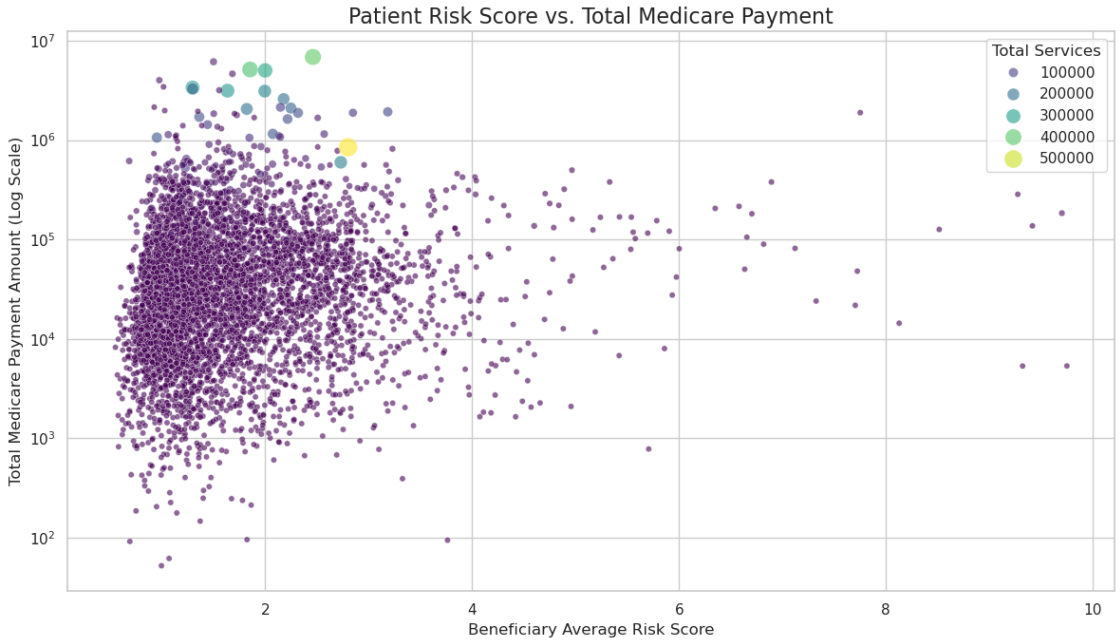
```
NPI
1003000134                       2.0                    16.0
1003005315                       NaN                    32.0
1003009861                       5.0                    31.0
1003010570                       9.0                    46.0
1003017443                       NaN                    75.0

            Bene_CC_BH_Depress_V1_Pct  Bene_CC_BH_PD_V1_Pct  \
NPI
1003000134                       14.0                   1.0
1003005315                       30.0                   NaN
1003009861                       28.0                   NaN
1003010570                       45.0                   NaN
1003017443                       75.0                   NaN

            Bene_CC_BH_PTSD_V1_Pct  Bene_CC_BH_Schizo_OthPsy_V1_Pct  \
NPI
1003000134                    0.0                             1.0
1003005315                    NaN                             NaN
1003009861                    3.0                             3.0
1003010570                    NaN                             6.0
1003017443                    NaN                            23.0

            Bene_CC_PH_Asthma_V2_Pct  Bene_CC_PH_Afib_V2_Pct  \
NPI
1003000134                       6.0                    16.0
1003005315                       NaN                    24.0
1003009861                      15.0                    38.0
1003010570                      17.0                    35.0
1003017443                       NaN                    30.0

            Bene_CC_PH_Cancer6_V2_Pct  Bene_CC_PH_CKD_V2_Pct  \
NPI
1003000134                       18.0                   15.0
1003005315                       17.0                   25.0
1003009861                       19.0                   42.0
1003010570                       22.0                   45.0
1003017443                       23.0                   30.0

            Bene_CC_PH_COPD_V2_Pct  Bene_CC_PH_Diabetes_V2_Pct  \
NPI
1003000134                    10.0                        21.0
1003005315                    17.0                        45.0
1003009861                    34.0                        40.0
1003010570                    37.0                        42.0
1003017443                    36.0                        41.0

            Bene_CC_PH_HF_NonIHD_V2_Pct  Bene_CC_PH_Hyperlipidemia_V2_Pct  \
```

```
NPI
1003000134                        11.0                                72.0
1003005315                         NaN                                63.0
1003009861                        37.0                                75.0
1003010570                        42.0                                75.0
1003017443                        44.0                                75.0

           Bene_CC_PH_Hypertension_V2_Pct  Bene_CC_PH_IschemicHeart_V2_Pct  \
NPI
1003000134                           64.0                             20.0
1003005315                           75.0                             23.0
1003009861                           75.0                             61.0
1003010570                           75.0                             49.0
1003017443                           75.0                             61.0

           Bene_CC_PH_Osteoporosis_V2_Pct  Bene_CC_PH_Parkinson_V2_Pct  \
NPI
1003000134                           15.0                          3.0
1003005315                            NaN                          NaN
1003009861                           13.0                          2.0
1003010570                           23.0                          NaN
1003017443                           20.0                          NaN

           Bene_CC_PH_Arthritis_V2_Pct  Bene_CC_PH_Stroke_TIA_V2_Pct  \
NPI
1003000134                        42.0                           6.0
1003005315                        48.0                           NaN
1003009861                        50.0                          22.0
1003010570                        59.0                          18.0
1003017443                        70.0                          31.0

           Bene_Avg_Risk_Scre LASTNAME FIRSTNAME MIDNAME BUSNAME  \
NPI
1003000134             1.1246      NaN       NaN     NaN     NaN
1003005315             1.6177    SMITH      ADAM  BRYANT     NaN
1003009861             2.1074      NaN       NaN     NaN     NaN
1003010570             2.1762      NaN       NaN     NaN     NaN
1003017443             2.2548      NaN       NaN     NaN     NaN

                     GENERAL        SPECIALTY UPIN         DOB  \
NPI
1003000134               NaN              NaN  NaN         NaN
1003005315  PHYSICIAN (MD, DO)  PLASTIC SURGERY  NaN  19800519.0
1003009861               NaN              NaN  NaN         NaN
1003010570               NaN              NaN  NaN         NaN
1003017443               NaN              NaN  NaN         NaN

                      ADDRESS        CITY STATE      ZIP EXCLTYPE  \
```

```
NPI
1003000134                  NaN         NaN   NaN      NaN      UNK
1003005315  2136 FORD PKWY 5103  SAINT PAUL    MN  55116.0   1128a1
1003009861                  NaN         NaN   NaN      NaN      UNK
1003010570                  NaN         NaN   NaN      NaN      UNK
1003017443                  NaN         NaN   NaN      NaN      UNK

              EXCLDATE  REINDATE  WAIVERDATE WVRSTATE  \
NPI
1003000134         NaN       NaN         NaN      NaN
1003005315  20221220.0       0.0         0.0      NaN
1003009861         NaN       NaN         NaN      NaN
1003010570         NaN       NaN         NaN      NaN
1003017443         NaN       NaN         NaN      NaN

            rat_Drug_Mdcr_Alowd_Amt_Drug_Mdcr_Pymt_Amt  \
NPI
1003000134                                         NaN
1003005315                                         NaN
1003009861                                    1.253709
1003010570                                         NaN
1003017443                                         NaN

            rat_Tot_Mdcr_Alowd_Amt_Med_Sbmtd_Chrg  \
NPI
1003000134                               0.246364
1003005315                                    NaN
1003009861                               0.498749
1003010570                               0.485200
1003017443                               0.527974

            rat_Drug_Tot_Benes_Tot_Benes  excluded      rand
NPI
1003000134                      0.000000         0  0.994100
1003005315                           NaN         1  0.253627
1003009861                      0.111872         0  0.994092
1003010570                      0.000000         0  0.994872
1003017443                      0.000000         0  0.991242
```

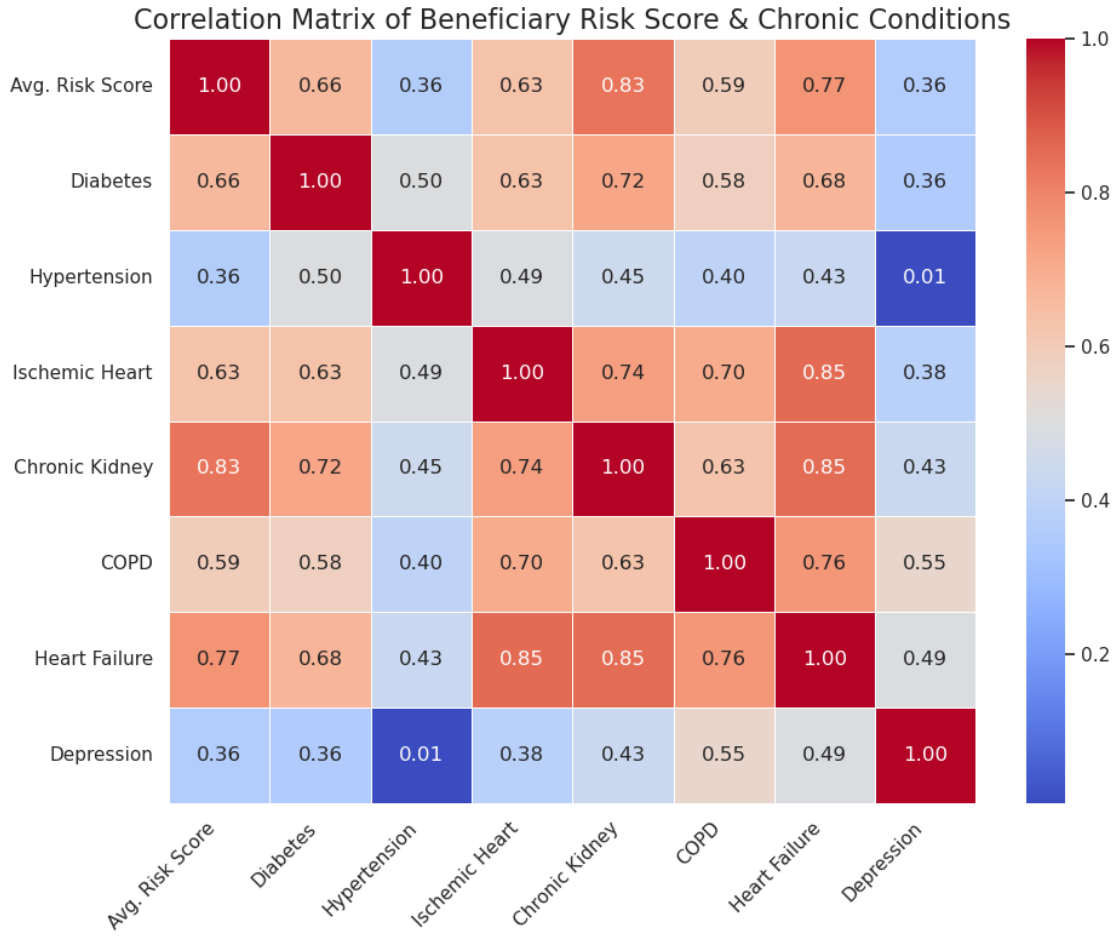
[190]:
```
plt.figure(figsize=(12, 7))

sample_df = data.sample(n=5000, random_state=42)

ax = sns.scatterplot(
    data=sample_df,
    x='Bene_Avg_Risk_Scre',
    y='Tot_Mdcr_Pymt_Amt',
```

```
    hue='Tot_Srvcs',
    size='Tot_Srvcs',
    sizes=(20, 200),
    palette='viridis',
    alpha=0.6
)

ax.set_yscale('log')
plt.title('Patient Risk Score vs. Total Medicare Payment', fontsize=16)
plt.xlabel('Beneficiary Average Risk Score', fontsize=12)
plt.ylabel('Total Medicare Payment Amount (Log Scale)', fontsize=12)
plt.legend(title='Total Services')
plt.tight_layout()
plt.style.use("seaborn-v0_8-dark-palette")
plt.show()
```



[191]:
```
cc_cols_of_interest = [
    'Bene_Avg_Risk_Scre',
    'Bene_CC_PH_Diabetes_V2_Pct',
    'Bene_CC_PH_Hypertension_V2_Pct',
    'Bene_CC_PH_IschemicHeart_V2_Pct',
    'Bene_CC_PH_CKD_V2_Pct',
    'Bene_CC_PH_COPD_V2_Pct',
    'Bene_CC_PH_HF_NonIHD_V2_Pct',
    'Bene_CC_BH_Depress_V1_Pct'
]
```

```
corr_matrix = data[cc_cols_of_interest].corr()


label_map = {
    'Bene_Avg_Risk_Scre': 'Avg. Risk Score',
    'Bene_CC_PH_Diabetes_V2_Pct': 'Diabetes',
    'Bene_CC_PH_Hypertension_V2_Pct': 'Hypertension',
    'Bene_CC_PH_IschemicHeart_V2_Pct': 'Ischemic Heart',
    'Bene_CC_PH_CKD_V2_Pct': 'Chronic Kidney',
    'Bene_CC_PH_COPD_V2_Pct': 'COPD',
    'Bene_CC_PH_HF_NonIHD_V2_Pct': 'Heart Failure',
    'Bene_CC_BH_Depress_V1_Pct': 'Depression'
}


corr_matrix_renamed = corr_matrix.rename(columns=label_map, index=label_map)

plt.figure(figsize=(10, 8))
sns.heatmap(
    corr_matrix_renamed,
    annot=True,
    cmap='coolwarm',
    fmt=".2f",
    linewidths=.5
)

plt.title('Correlation Matrix of Beneficiary Risk Score & Chronic Conditions',␣
 ↪fontsize=16)
plt.xticks(rotation=45, ha='right')
plt.yticks(rotation=0)
plt.tight_layout()
plt.style.use("seaborn-v0_8-dark-palette")
plt.show()
```



1 Linear Regression/Modeling

[192]:
```
df = pd.merge(data, exclusions, on='NPI',how='left')
```

[193]:
```
df['excluded'].unique()
```

[193]:
```
array([0, 1])
```

[194]:
```
len(df)
```

[194]:
```
12496
```

[195]:
```
df.columns
```

[195]:
```
Index(['Rndrng_Prvdr_Last_Org_Name', 'Rndrng_Prvdr_First_Name',
       'Rndrng_Prvdr_MI', 'Rndrng_Prvdr_Crdntls', 'Rndrng_Prvdr_Ent_Cd',
       'Rndrng_Prvdr_St1', 'Rndrng_Prvdr_St2', 'Rndrng_Prvdr_City',
```

```
       'Rndrng_Prvdr_State_Abrvtn', 'Rndrng_Prvdr_State_FIPS',
       …
       'SPECIALTY_y', 'DOB_y', 'ADDRESS_y', 'CITY_y', 'STATE_y', 'ZIP_y',
       'EXCLTYPE_y', 'EXCLDATE_y', 'age_at_exclusion', 'year_excl'],
      dtype='object', length=115)
```

[196]:
```
df = df.copy()

# 1 if excluded (has an exclusion date), 0 otherwise
df['is_excluded'] = df['EXCLDATE_y'].notna().astype(int)

# keep rows where provider type is known
df = df[df['Rndrng_Prvdr_Type'].notna()]
df.head()
```

[196]:

NPI	Rndrng_Prvdr_Last_Org_Name	Rndrng_Prvdr_First_Name	Rndrng_Prvdr_MI	Rndrng_Prvdr_Crdntls	Rndrng_Prvdr_Ent_Cd	Rndrng_Prvdr_St1	Rndrng_Prvdr_St2	Rndrng_Prvdr_City	Rndrng_Prvdr_State_Abrvtn	Rndrng_Prvdr_State_FIPS	Rndrng_Prvdr_Zip5	Rndrng_Prvdr_RUCA
1003000134	CIBULL	THOMAS	L	M.D.	I	2650 RIDGE AVE	EVANSTON HOSPITAL	EVANSTON	IL	17	60201.0	1.0
1003005315	SMITH	ADAM	B	MD	I	4977 SKYVIEW CT	NaN	TRAVERSE CITY	MI	26	49684.0	4.0
1003009861	BANNA	MOUSTAFA	NaN	MD	I	5859 W. TALAVI BLVD	SUITE 100	GLENDALE	AZ	4	85306.0	1.0
1003010570	CHOW	LING	S	M.D.	I	900 E BROADWAY AVE	NaN	BISMARCK	ND	38	58501.0	1.0
1003017443	BLOKAR	MIRJANA	NaN	M.D.	I	65 BLEECKER ST	12TH FLOOR	NEW YORK	NY	36	10012.0	1.0

```
                                    Rndrng_Prvdr_RUCA_Desc  \
NPI
1003000134  Metropolitan area core: primary flow within an…
1003005315  Micropolitan area core: primary flow within an…
1003009861  Metropolitan area core: primary flow within an…
1003010570  Metropolitan area core: primary flow within an…
1003017443  Metropolitan area core: primary flow within an…

           Rndrng_Prvdr_Cntry                   Rndrng_Prvdr_Type  \
NPI
1003000134                 US                          Pathology
1003005315                 US  Plastic and Reconstructive Surgery
1003009861                 US                        Cardiology
1003010570                 US                  Internal Medicine
1003017443                 US                         Psychiatry

           Rndrng_Prvdr_Mdcr_Prtcptg_Ind  Tot_HCPCS_Cds  Tot_Benes  Tot_Srvcs  \
NPI
1003000134                             Y             20       3614     7835.0
1003005315                             Y             51         71      533.0
1003009861                             Y             68        876     3581.0
1003010570                             Y             21        309      833.0
1003017443                             Y              5         64      244.0

            Tot_Sbmtd_Chrg  Tot_Mdcr_Alowd_Amt  Tot_Mdcr_Pymt_Amt  \
NPI
1003000134       1213264.0           298905.09          228078.56
1003005315        250569.0            86114.40           67712.67
1003009861        788610.0           361645.81          280928.83
1003010570        173991.0            84420.40           65891.84
1003017443         45375.0            23956.84           18245.22

            Tot_Mdcr_Stdzd_Amt Drug_Sprsn_Ind  Drug_Tot_HCPCS_Cds  \
NPI
1003000134          218268.63            NaN                 0.0
1003005315           70489.71              *                 NaN
1003009861          286768.20            NaN                 1.0
1003010570           66566.70            NaN                 0.0
1003017443           16748.76            NaN                 0.0

            Drug_Tot_Benes  Drug_Tot_Srvcs  Drug_Sbmtd_Chrg  \
NPI
1003000134             0.0             0.0              0.0
1003005315             NaN             NaN              NaN
1003009861            98.0           392.0          63504.0
1003010570             0.0             0.0              0.0
1003017443             0.0             0.0              0.0
```

NPI	Drug_Mdcr_Alowd_Amt	Drug_Mdcr_Pymt_Amt	Drug_Mdcr_Stdzd_Amt
1003000134	0.00	0.00	0.00
1003005315	NaN	NaN	NaN
1003009861	21905.24	17472.35	18074.19
1003010570	0.00	0.00	0.00
1003017443	0.00	0.00	0.00

NPI	Med_Sprsn_Ind	Med_Tot_HCPCS_Cds	Med_Tot_Benes	Med_Tot_Srvcs
1003000134	NaN	20.0	3614.0	7835.0
1003005315	#	NaN	NaN	NaN
1003009861	NaN	67.0	876.0	3189.0
1003010570	NaN	21.0	309.0	833.0
1003017443	NaN	5.0	64.0	244.0

NPI	Med_Sbmtd_Chrg	Med_Mdcr_Alowd_Amt	Med_Mdcr_Pymt_Amt
1003000134	1213264.0	298905.09	228078.56
1003005315	NaN	NaN	NaN
1003009861	725106.0	339740.57	263456.48
1003010570	173991.0	84420.40	65891.84
1003017443	45375.0	23956.84	18245.22

NPI	Med_Mdcr_Stdzd_Amt	Bene_Avg_Age	Bene_Age_LT_65_Cnt
1003000134	218268.63	76	97.0
1003005315	NaN	73	NaN
1003009861	268694.01	73	91.0
1003010570	66566.70	76	39.0
1003017443	16748.76	83	NaN

NPI	Bene_Age_65_74_Cnt	Bene_Age_75_84_Cnt	Bene_Age_GT_84_Cnt
1003000134	1623.0	1333.0	561.0
1003005315	29.0	20.0	NaN
1003009861	400.0	294.0	91.0
1003010570	95.0	100.0	75.0
1003017443	NaN	18.0	31.0

NPI	Bene_Feml_Cnt	Bene_Male_Cnt	Bene_Race_Wht_Cnt
1003000134	1767.0	1847.0	3312.0
1003005315	44.0	27.0	NaN
1003009861	410.0	466.0	709.0
1003010570	167.0	142.0	294.0

```
1003017443          42.0          22.0               48.0

            Bene_Race_Black_Cnt  Bene_Race_API_Cnt  Bene_Race_Hspnc_Cnt  \
NPI
1003000134                41.0               62.0                 43.0
1003005315                 NaN                NaN                  NaN
1003009861                35.0               20.0                 80.0
1003010570                 NaN                NaN                  NaN
1003017443                 NaN                0.0                  NaN

            Bene_Race_NatInd_Cnt  Bene_Race_Othr_Cnt  Bene_Dual_Cnt  \
NPI
1003000134                  0.0               156.0          195.0
1003005315                  NaN                 NaN           14.0
1003009861                 11.0                21.0          164.0
1003010570                  NaN                 NaN           77.0
1003017443                  0.0                 NaN            NaN

            Bene_Ndual_Cnt  Bene_CC_BH_ADHD_OthCD_V1_Pct  \
NPI
1003000134          3419.0                           0.0
1003005315            57.0                           NaN
1003009861           712.0                           NaN
1003010570           232.0                           NaN
1003017443             NaN                           NaN

            Bene_CC_BH_Alcohol_Drug_V1_Pct  Bene_CC_BH_Tobacco_V1_Pct  \
NPI
1003000134                             2.0                        3.0
1003005315                             NaN                        NaN
1003009861                            13.0                       21.0
1003010570                             8.0                       17.0
1003017443                             NaN                        NaN

            Bene_CC_BH_Alz_NonAlzdem_V2_Pct  Bene_CC_BH_Anxiety_V1_Pct  \
NPI
1003000134                              4.0                       13.0
1003005315                              NaN                       35.0
1003009861                             12.0                       29.0
1003010570                             26.0                       42.0
1003017443                             63.0                       45.0

            Bene_CC_BH_Bipolar_V1_Pct  Bene_CC_BH_Mood_V2_Pct  \
NPI
1003000134                        2.0                    16.0
1003005315                        NaN                    32.0
1003009861                        5.0                    31.0
```

```
1003010570                      9.0                    46.0
1003017443                      NaN                    75.0

           Bene_CC_BH_Depress_V1_Pct  Bene_CC_BH_PD_V1_Pct  \
NPI
1003000134                      14.0                   1.0
1003005315                      30.0                   NaN
1003009861                      28.0                   NaN
1003010570                      45.0                   NaN
1003017443                      75.0                   NaN

           Bene_CC_BH_PTSD_V1_Pct  Bene_CC_BH_Schizo_OthPsy_V1_Pct  \
NPI
1003000134                    0.0                              1.0
1003005315                    NaN                              NaN
1003009861                    3.0                              3.0
1003010570                    NaN                              6.0
1003017443                    NaN                             23.0

           Bene_CC_PH_Asthma_V2_Pct  Bene_CC_PH_Afib_V2_Pct  \
NPI
1003000134                      6.0                   16.0
1003005315                      NaN                   24.0
1003009861                     15.0                   38.0
1003010570                     17.0                   35.0
1003017443                      NaN                   30.0

           Bene_CC_PH_Cancer6_V2_Pct  Bene_CC_PH_CKD_V2_Pct  \
NPI
1003000134                      18.0                   15.0
1003005315                      17.0                   25.0
1003009861                      19.0                   42.0
1003010570                      22.0                   45.0
1003017443                      23.0                   30.0

           Bene_CC_PH_COPD_V2_Pct  Bene_CC_PH_Diabetes_V2_Pct  \
NPI
1003000134                   10.0                        21.0
1003005315                   17.0                        45.0
1003009861                   34.0                        40.0
1003010570                   37.0                        42.0
1003017443                   36.0                        41.0

           Bene_CC_PH_HF_NonIHD_V2_Pct  Bene_CC_PH_Hyperlipidemia_V2_Pct  \
NPI
1003000134                        11.0                              72.0
1003005315                         NaN                              63.0
```

```
1003009861                        37.0                                75.0
1003010570                        42.0                                75.0
1003017443                        44.0                                75.0

           Bene_CC_PH_Hypertension_V2_Pct  Bene_CC_PH_IschemicHeart_V2_Pct  \
NPI
1003000134                           64.0                              20.0
1003005315                           75.0                              23.0
1003009861                           75.0                              61.0
1003010570                           75.0                              49.0
1003017443                           75.0                              61.0

           Bene_CC_PH_Osteoporosis_V2_Pct  Bene_CC_PH_Parkinson_V2_Pct  \
NPI
1003000134                           15.0                           3.0
1003005315                            NaN                           NaN
1003009861                           13.0                           2.0
1003010570                           23.0                           NaN
1003017443                           20.0                           NaN

           Bene_CC_PH_Arthritis_V2_Pct  Bene_CC_PH_Stroke_TIA_V2_Pct  \
NPI
1003000134                        42.0                           6.0
1003005315                        48.0                           NaN
1003009861                        50.0                          22.0
1003010570                        59.0                          18.0
1003017443                        70.0                          31.0

           Bene_Avg_Risk_Scre LASTNAME_x FIRSTNAME_x MIDNAME BUSNAME  \
NPI
1003000134             1.1246        NaN         NaN     NaN     NaN
1003005315             1.6177      SMITH        ADAM  BRYANT     NaN
1003009861             2.1074        NaN         NaN     NaN     NaN
1003010570             2.1762        NaN         NaN     NaN     NaN
1003017443             2.2548        NaN         NaN     NaN     NaN

                  GENERAL_x      SPECIALTY_x UPIN       DOB_x  \
NPI
1003000134              NaN              NaN  NaN         NaN
1003005315  PHYSICIAN (MD, DO)  PLASTIC SURGERY  NaN  19800519.0
1003009861              NaN              NaN  NaN         NaN
1003010570              NaN              NaN  NaN         NaN
1003017443              NaN              NaN  NaN         NaN

                   ADDRESS_x      CITY_x STATE_x    ZIP_x EXCLTYPE_x  \
NPI
1003000134               NaN         NaN     NaN      NaN        UNK
```

```
1003005315  2136 FORD PKWY 5103  SAINT PAUL      MN  55116.0     1128a1
1003009861                 NaN         NaN     NaN      NaN        UNK
1003010570                 NaN         NaN     NaN      NaN        UNK
1003017443                 NaN         NaN     NaN      NaN        UNK

           EXCLDATE_x  REINDATE  WAIVERDATE WVRSTATE  \
NPI
1003000134        NaN       NaN         NaN      NaN
1003005315  20221220.0       0.0         0.0      NaN
1003009861        NaN       NaN         NaN      NaN
1003010570        NaN       NaN         NaN      NaN
1003017443        NaN       NaN         NaN      NaN

           rat_Drug_Mdcr_Alowd_Amt_Drug_Mdcr_Pymt_Amt  \
NPI
1003000134                                        NaN
1003005315                                        NaN
1003009861                                   1.253709
1003010570                                        NaN
1003017443                                        NaN

           rat_Tot_Mdcr_Alowd_Amt_Med_Sbmtd_Chrg  \
NPI
1003000134                              0.246364
1003005315                                   NaN
1003009861                              0.498749
1003010570                              0.485200
1003017443                              0.527974

           rat_Drug_Tot_Benes_Tot_Benes  excluded      rand LASTNAME_y  \
NPI
1003000134                     0.000000         0  0.994100        NaN
1003005315                          NaN         1  0.253627      SMITH
1003009861                     0.111872         0  0.994092        NaN
1003010570                     0.000000         0  0.994872        NaN
1003017443                     0.000000         0  0.991242        NaN

          FIRSTNAME_y           GENERAL_y      SPECIALTY_y      DOB_y  \
NPI
1003000134         NaN                 NaN              NaN        NaT
1003005315        ADAM  PHYSICIAN (MD, DO)  PLASTIC SURGERY 1980-05-19
1003009861         NaN                 NaN              NaN        NaT
1003010570         NaN                 NaN              NaN        NaT
1003017443         NaN                 NaN              NaN        NaT

                   ADDRESS_y      CITY_y STATE_y    ZIP_y  \
NPI
```

```
1003000134                  NaN         NaN     NaN      NaN
1003005315  2136 FORD PKWY 5103  SAINT PAUL      MN  55116.0
1003009861                  NaN         NaN     NaN      NaN
1003010570                  NaN         NaN     NaN      NaN
1003017443                  NaN         NaN     NaN      NaN

                            EXCLTYPE_y EXCLDATE_y  age_at_exclusion  \
NPI
1003000134                         NaN        NaT               NaN
1003005315  Conviction - Medicare Fraud 2022-12-20         42.587269
1003009861                         NaN        NaT               NaN
1003010570                         NaN        NaT               NaN
1003017443                         NaN        NaT               NaN

            year_excl  is_excluded
NPI
1003000134        NaN            0
1003005315     2022.0            1
1003009861        NaN            0
1003010570        NaN            0
1003017443        NaN            0
```

[197]:
```
by_type = (
    df.groupby('Rndrng_Prvdr_Type')['is_excluded']
      .agg(['sum', 'count'])
      .rename(columns={'sum': 'n_excluded', 'count': 'n_total'})
      .sort_values('n_total', ascending=False)
)

print(by_type.head(20))
print(by_type.tail(20))
```

```
                                                n_excluded  n_total
Rndrng_Prvdr_Type
Nurse Practitioner                                      87     1482
Internal Medicine                                      116     1026
Family Practice                                        150     1016
Physician Assistant                                     23      934
Physical Therapist in Private Practice                  20      605
Certified Registered Nurse Anesthetist (CRNA)           16      499
Emergency Medicine                                      35      478
Anesthesiology                                          35      427
Chiropractic                                            33      397
Diagnostic Radiology                                     8      310
Optometry                                                3      310
Mass Immunizer Roster Biller                             0      307
Obstetrics & Gynecology                                 28      263
General Surgery                                         15      251
```

```
Psychiatry                                                39      242
Licensed Clinical Social Worker                           17      230
Orthopedic Surgery                                        12      229
Cardiology                                                 8      193
Podiatry                                                  26      190
Psychologist, Clinical                                     8      179
                                                      n_excluded  n_total
Rndrng_Prvdr_Type
Gynecological Oncology                                         0       12
Maxillofacial Surgery                                          0       12
Pharmacy                                                       0       11
Speech Language Pathologist                                    0       11
Hematology                                                     0        9
Sports Medicine                                                1        9
Preventive Medicine                                            2        7
Undefined Physician type                                       1        6
Nuclear Medicine                                               1        6
Sleep Medicine                                                 0        4
Advanced Heart Failure and Transplant Cardiology               0        4
Osteopathic Manipulative Medicine                              0        4
Public Health or Welfare Agency                                0        3
Addiction Medicine                                             2        3
Portable X-Ray Supplier                                        0        3
Hematopoietic Cell Transplantation and Cellular…               0        2
Neuropsychiatry                                                1        2
Dentist                                                        0        1
Mammography Center                                             0        1
Radiation Therapy Center                                       0        1
```

[198]:
```python
min_n = 30

# mark small or perfectly separated types
small_or_perfect = by_type[
    (by_type['n_total'] < min_n) |
    (by_type['n_excluded'] == 0) |
    (by_type['n_excluded'] == by_type['n_total'])
].index

# create a cleaned copy
df_clean = df.copy()
df_clean = df_clean[df_clean['Rndrng_Prvdr_Type'].notna()].copy()

# collapse problematic types
df_clean['prov_type_collapsed'] = df_clean['Rndrng_Prvdr_Type'].where(
    ~df_clean['Rndrng_Prvdr_Type'].isin(small_or_perfect),
    other='Other'
)
```

```
print(df_clean['prov_type_collapsed'].value_counts())
```

```
prov_type_collapsed
Nurse Practitioner                                1482
Other                                             1136
Internal Medicine                                 1026
Family Practice                                   1016
Physician Assistant                                934
Physical Therapist in Private Practice             605
Certified Registered Nurse Anesthetist (CRNA)      499
Emergency Medicine                                 478
Anesthesiology                                     427
Chiropractic                                       397
Diagnostic Radiology                               310
Optometry                                          310
Obstetrics & Gynecology                            263
General Surgery                                    251
Psychiatry                                         242
Licensed Clinical Social Worker                    230
Orthopedic Surgery                                 229
Cardiology                                         193
Podiatry                                           190
Psychologist, Clinical                             179
Ophthalmology                                      166
Gastroenterology                                   153
Neurology                                          147
Hospitalist                                        134
Dermatology                                        128
Pathology                                          117
Otolaryngology                                     115
Physical Medicine and Rehabilitation               113
Pulmonary Disease                                  102
Hematology-Oncology                                100
Nephrology                                          98
Urology                                             88
Occupational Therapist in Private Practice          77
Infectious Disease                                  72
General Practice                                    68
Neurosurgery                                        63
Endocrinology                                       62
Rheumatology                                        46
Critical Care (Intensivists)                        43
Pain Management                                     38
Plastic and Reconstructive Surgery                  37
Radiation Oncology                                  35
Medical Oncology                                    34
Allergy/ Immunology                                 32
```

```
Clinical Cardiac Electrophysiology                31
Name: count, dtype: int64
```

[199]:
```
import statsmodels.api as sm
import patsy

y, X = patsy.dmatrices('is_excluded ~ C(prov_type_collapsed)', df_clean,␣
 ↪return_type='dataframe')
logit_reg = sm.Logit(y, X).fit_regularized()
print(logit_reg.summary())
```

```
Optimization terminated successfully    (Exit mode 0)
            Current function value: 0.22972037051460392
            Iterations: 237
            Function evaluations: 238
            Gradient evaluations: 237
                           Logit Regression Results
==============================================================================
Dep. Variable:            is_excluded   No. Observations:                12496
Model:                          Logit   Df Residuals:                    12451
Method:                           MLE   Df Model:                           44
Date:                Wed, 17 Dec 2025   Pseudo R-squ.:                 0.07320
Time:                        22:15:43   Log-Likelihood:                -2870.6
converged:                       True   LL-Null:                       -3097.3
Covariance Type:            nonrobust   LLR p-value:                 2.193e-69
==============================================================================
==========================================================
coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
------------------------------------------------------------
Intercept
-2.7291      0.737     -3.703      0.000      -4.174      -1.284
C(prov_type_collapsed)[T.Anesthesiology]
0.3128      0.758      0.413      0.680      -1.173       1.798
C(prov_type_collapsed)[T.Cardiology]
-0.4148      0.821     -0.505      0.613      -2.024       1.194
C(prov_type_collapsed)[T.Certified Registered Nurse Anesthetist (CRNA)]
-0.6776      0.780     -0.869      0.385      -2.206       0.850
C(prov_type_collapsed)[T.Chiropractic]
0.3292      0.759      0.434      0.665      -1.159       1.817
C(prov_type_collapsed)[T.Clinical Cardiac Electrophysiology]
-0.6557      1.249     -0.525      0.600      -3.105       1.793
C(prov_type_collapsed)[T.Critical Care (Intensivists)]
-0.9906      1.245     -0.796      0.426      -3.431       1.449
C(prov_type_collapsed)[T.Dermatology]
-2.1358      1.254     -1.704      0.088      -4.593       0.322
C(prov_type_collapsed)[T.Diagnostic Radiology]
-0.8997      0.819     -1.098      0.272      -2.506       0.706
```

```
C(prov_type_collapsed)[T.Emergency Medicine]
0.1911      0.758      0.252      0.801      -1.294       1.676
C(prov_type_collapsed)[T.Endocrinology]
-0.2511      0.945     -0.266      0.791      -2.104       1.602
C(prov_type_collapsed)[T.Family Practice]
0.9759      0.742      1.315      0.189      -0.479       2.431
C(prov_type_collapsed)[T.Gastroenterology]
-0.4683      0.846     -0.553      0.580      -2.127       1.191
C(prov_type_collapsed)[T.General Practice]
1.7822      0.785      2.270      0.023       0.243       3.321
C(prov_type_collapsed)[T.General Surgery]
-0.0254      0.784     -0.032      0.974      -1.561       1.511
C(prov_type_collapsed)[T.Hematology-Oncology]
-0.2113      0.868     -0.244      0.808      -1.912       1.490
C(prov_type_collapsed)[T.Hospitalist]
-2.1830      1.254     -1.741      0.082      -4.641       0.275
C(prov_type_collapsed)[T.Infectious Disease]
-0.8215      1.027     -0.800      0.424      -2.835       1.192
C(prov_type_collapsed)[T.Internal Medicine]
0.6693      0.744      0.900      0.368      -0.788       2.127
C(prov_type_collapsed)[T.Licensed Clinical Social Worker]
0.2008      0.779      0.258      0.797      -1.326       1.728
C(prov_type_collapsed)[T.Medical Oncology]
-0.7494      1.248     -0.601      0.548      -3.195       1.696
C(prov_type_collapsed)[T.Nephrology]
-1.8577      1.251     -1.485      0.138      -4.310       0.595
C(prov_type_collapsed)[T.Neurology]
0.2152      0.801      0.269      0.788      -1.355       1.785
C(prov_type_collapsed)[T.Neurosurgery]
0.4803      0.853      0.563      0.573      -1.191       2.152
C(prov_type_collapsed)[T.Nurse Practitioner]
-0.0456      0.745     -0.061      0.951      -1.506       1.415
C(prov_type_collapsed)[T.Obstetrics & Gynecology]
0.6021      0.764      0.788      0.430      -0.895       2.099
C(prov_type_collapsed)[T.Occupational Therapist in Private Practice]
-1.6029      1.248     -1.284      0.199      -4.049       0.843
C(prov_type_collapsed)[T.Ophthalmology]
-0.5532      0.846     -0.654      0.513      -2.212       1.105
C(prov_type_collapsed)[T.Optometry]
-1.9044      0.939     -2.028      0.043      -3.745      -0.064
C(prov_type_collapsed)[T.Orthopedic Surgery]
-0.1695      0.795     -0.213      0.831      -1.727       1.388
C(prov_type_collapsed)[T.Other]
-0.7796      0.758     -1.029      0.304      -2.265       0.706
C(prov_type_collapsed)[T.Otolaryngology]
0.1344      0.823      0.163      0.870      -1.479       1.748
C(prov_type_collapsed)[T.Pain Management]
2.5156      0.806      3.121      0.002       0.936       4.095
```

```
C(prov_type_collapsed)[T.Pathology]
-1.3144      1.024     -1.284      0.199      -3.321       0.692
C(prov_type_collapsed)[T.Physical Medicine and Rehabilitation]
1.4184       0.772      1.837      0.066      -0.095       2.932
C(prov_type_collapsed)[T.Physical Therapist in Private Practice]
-0.6464      0.771     -0.838      0.402      -2.158       0.865
C(prov_type_collapsed)[T.Physician Assistant]
-0.9506      0.767     -1.240      0.215      -2.453       0.552
C(prov_type_collapsed)[T.Plastic and Reconstructive Surgery]
0.6192       0.907      0.682      0.495      -1.159       2.398
C(prov_type_collapsed)[T.Podiatry]
0.8868       0.767      1.157      0.247      -0.616       2.390
C(prov_type_collapsed)[T.Psychiatry]
1.0797       0.758      1.425      0.154      -0.405       2.564
C(prov_type_collapsed)[T.Psychologist, Clinical]
-0.3367      0.821     -0.410      0.682      -1.946       1.273
C(prov_type_collapsed)[T.Pulmonary Disease]
-1.1757      1.025     -1.147      0.251      -3.184       0.832
C(prov_type_collapsed)[T.Radiation Oncology]
-0.7790      1.247     -0.625      0.532      -3.223       1.665
C(prov_type_collapsed)[T.Rheumatology]
0.8321       0.857      0.971      0.332      -0.848       2.512
C(prov_type_collapsed)[T.Urology]
-0.3154      0.897     -0.352      0.725      -2.074       1.443
==============================================================================
==============================================================
```

[200]:
```
print(
    df_clean.groupby('prov_type_collapsed')['is_excluded']
            .agg(['sum', 'count'])
)
```

```
                                                sum  count
prov_type_collapsed
Allergy/ Immunology                               2     32
Anesthesiology                                   35    427
Cardiology                                        8    193
Certified Registered Nurse Anesthetist (CRNA)    16    499
Chiropractic                                     33    397
Clinical Cardiac Electrophysiology                1     31
Critical Care (Intensivists)                      1     43
Dermatology                                       1    128
Diagnostic Radiology                              8    310
Emergency Medicine                               35    478
Endocrinology                                     3     62
Family Practice                                 150   1016
Gastroenterology                                  6    153
General Practice                                 19     68
```

```
General Surgery                                    15    251
Hematology-Oncology                                 5    100
Hospitalist                                         1    134
Infectious Disease                                  2     72
Internal Medicine                                 116   1026
Licensed Clinical Social Worker                    17    230
Medical Oncology                                    1     34
Nephrology                                          1     98
Neurology                                          11    147
Neurosurgery                                        6     63
Nurse Practitioner                                 87   1482
Obstetrics & Gynecology                            28    263
Occupational Therapist in Private Practice          1     77
Ophthalmology                                       6    166
Optometry                                           3    310
Orthopedic Surgery                                 12    229
Other                                              33   1136
Otolaryngology                                      8    115
Pain Management                                    17     38
Pathology                                           2    117
Physical Medicine and Rehabilitation               24    113
Physical Therapist in Private Practice             20    605
Physician Assistant                                23    934
Plastic and Reconstructive Surgery                  4     37
Podiatry                                           26    190
Psychiatry                                         39    242
Psychologist, Clinical                              8    179
Pulmonary Disease                                   2    102
Radiation Oncology                                  1     35
Rheumatology                                        6     46
Urology                                             4     88
```

[201]:
```
logit_model = smf.logit(
    'is_excluded ~ C(prov_type_collapsed)',
    data=df_clean
).fit()

print(logit_model.summary())

# odds ratios
odds_ratios = np.exp(logit_model.params)
print(odds_ratios.sort_values())
```

```
Optimization terminated successfully.
         Current function value: 0.229720
         Iterations 9
                           Logit Regression Results
==============================================================================
```

```
Dep. Variable:            is_excluded   No. Observations:                12496
Model:                          Logit   Df Residuals:                    12451
Method:                           MLE   Df Model:                           44
Date:                Wed, 17 Dec 2025   Pseudo R-squ.:                 0.07320
Time:                        22:15:45   Log-Likelihood:                -2870.6
converged:                       True   LL-Null:                       -3097.3
Covariance Type:            nonrobust   LLR p-value:                 2.189e-69
==============================================================================
                                                                              
==============================================================
coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
                                                                              
--------------------------------------------------------------
Intercept
-2.7081      0.730     -3.708      0.000      -4.139      -1.277
C(prov_type_collapsed)[T.Anesthesiology]
0.2921      0.751      0.389      0.697      -1.180       1.765
C(prov_type_collapsed)[T.Cardiology]
-0.4329      0.815     -0.531      0.595      -2.030       1.164
C(prov_type_collapsed)[T.Certified Registered Nurse Anesthetist (CRNA)]
-0.6994      0.773     -0.904      0.366      -2.215       0.816
C(prov_type_collapsed)[T.Chiropractic]
0.3074      0.753      0.408      0.683      -1.168       1.782
C(prov_type_collapsed)[T.Clinical Cardiac Electrophysiology]
-0.6931      1.252     -0.554      0.580      -3.146       1.760
C(prov_type_collapsed)[T.Critical Care (Intensivists)]
-1.0296      1.248     -0.825      0.409      -3.475       1.416
C(prov_type_collapsed)[T.Dermatology]
-2.1361      1.241     -1.721      0.085      -4.569       0.297
C(prov_type_collapsed)[T.Diagnostic Radiology]
-0.9229      0.813     -1.135      0.257      -2.517       0.671
C(prov_type_collapsed)[T.Emergency Medicine]
0.1698      0.751      0.226      0.821      -1.302       1.642
C(prov_type_collapsed)[T.Endocrinology]
-0.2709      0.940     -0.288      0.773      -2.113       1.572
C(prov_type_collapsed)[T.Family Practice]
0.9548      0.736      1.298      0.194      -0.487       2.397
C(prov_type_collapsed)[T.Gastroenterology]
-0.4906      0.841     -0.584      0.560      -2.138       1.157
C(prov_type_collapsed)[T.General Practice]
1.7607      0.779      2.261      0.024       0.234       3.287
C(prov_type_collapsed)[T.General Surgery]
-0.0477      0.777     -0.061      0.951      -1.571       1.476
C(prov_type_collapsed)[T.Hematology-Oncology]
-0.2364      0.862     -0.274      0.784      -1.927       1.454
C(prov_type_collapsed)[T.Hospitalist]
-2.1823      1.241     -1.758      0.079      -4.615       0.251
C(prov_type_collapsed)[T.Infectious Disease]
-0.8473      1.024     -0.828      0.408      -2.853       1.159
```

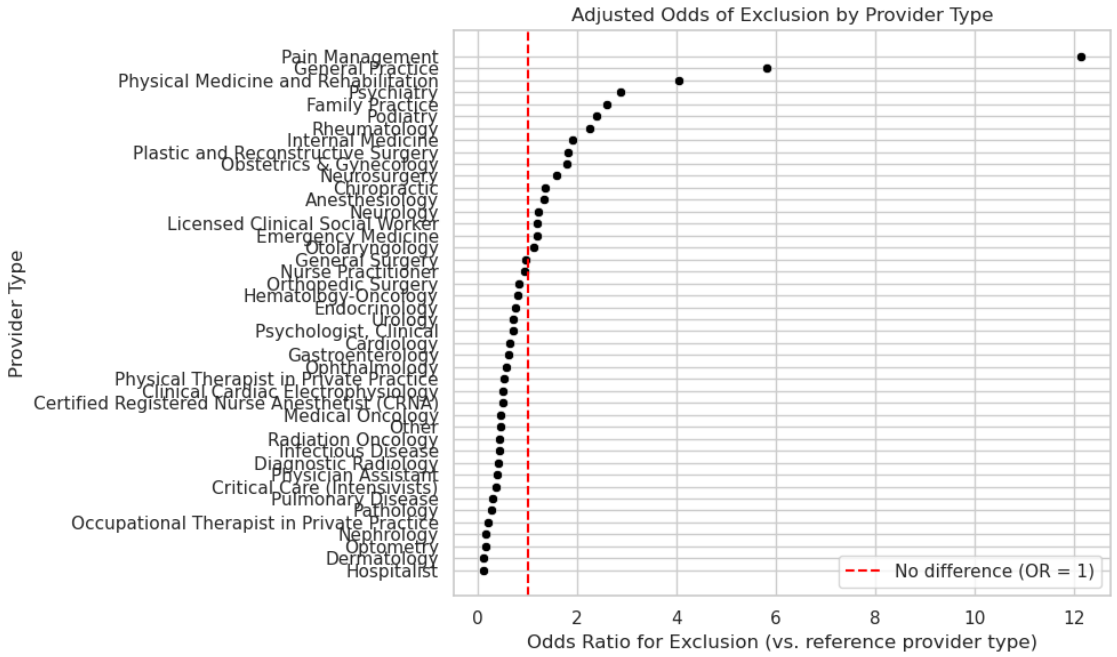
```
C(prov_type_collapsed)[T.Internal Medicine]
0.6482      0.737      0.880      0.379      -0.796       2.093
C(prov_type_collapsed)[T.Licensed Clinical Social Worker]
0.1800      0.773      0.233      0.816      -1.334       1.694
C(prov_type_collapsed)[T.Medical Oncology]
-0.7885      1.250     -0.631      0.528      -3.239       1.662
C(prov_type_collapsed)[T.Nephrology]
-1.8667      1.242     -1.502      0.133      -4.302       0.568
C(prov_type_collapsed)[T.Neurology]
0.1933      0.795      0.243      0.808      -1.364       1.751
C(prov_type_collapsed)[T.Neurosurgery]
0.4568      0.847      0.539      0.590      -1.203       2.117
C(prov_type_collapsed)[T.Nurse Practitioner]
-0.0667      0.739     -0.090      0.928      -1.514       1.381
C(prov_type_collapsed)[T.Obstetrics & Gynecology]
0.5807      0.757      0.767      0.443      -0.903       2.065
C(prov_type_collapsed)[T.Occupational Therapist in Private Practice]
-1.6227      1.244     -1.305      0.192      -4.060       0.815
C(prov_type_collapsed)[T.Ophthalmology]
-0.5754      0.840     -0.685      0.494      -2.222       1.072
C(prov_type_collapsed)[T.Optometry]
-1.9202      0.933     -2.059      0.040      -3.748      -0.092
C(prov_type_collapsed)[T.Orthopedic Surgery]
-0.1869      0.788     -0.237      0.813      -1.732       1.358
C(prov_type_collapsed)[T.Other]
-0.8012      0.751     -1.066      0.286      -2.274       0.671
C(prov_type_collapsed)[T.Otolaryngology]
0.1147      0.817      0.140      0.888      -1.487       1.716
C(prov_type_collapsed)[T.Pain Management]
2.4967      0.800      3.121      0.002       0.929       4.064
C(prov_type_collapsed)[T.Pathology]
-1.3437      1.021     -1.316      0.188      -3.344       0.657
C(prov_type_collapsed)[T.Physical Medicine and Rehabilitation]
1.3975      0.766      1.825      0.068      -0.103       2.898
C(prov_type_collapsed)[T.Physical Therapist in Private Practice]
-0.6678      0.765     -0.873      0.383      -2.167       0.831
C(prov_type_collapsed)[T.Physician Assistant]
-0.9710      0.760     -1.277      0.202      -2.461       0.519
C(prov_type_collapsed)[T.Plastic and Reconstructive Surgery]
0.5978      0.902      0.663      0.507      -1.170       2.366
C(prov_type_collapsed)[T.Podiatry]
0.8663      0.760      1.140      0.254      -0.624       2.356
C(prov_type_collapsed)[T.Psychiatry]
1.0584      0.751      1.409      0.159      -0.413       2.530
C(prov_type_collapsed)[T.Psychologist, Clinical]
-0.3542      0.815     -0.435      0.664      -1.951       1.243
C(prov_type_collapsed)[T.Pulmonary Disease]
-1.2040      1.021     -1.179      0.239      -3.206       0.798
```

```
C(prov_type_collapsed)[T.Radiation Oncology]
-0.8183      1.250     -0.655      0.513      -3.268       1.632
C(prov_type_collapsed)[T.Rheumatology]
0.8109      0.851      0.952      0.341      -0.858       2.480
C(prov_type_collapsed)[T.Urology]
-0.3365      0.892     -0.377      0.706      -2.084       1.411
==============================================================================
==============================================================
Intercept
0.066667
C(prov_type_collapsed)[T.Hospitalist]
0.112782
C(prov_type_collapsed)[T.Dermatology]
0.118110
C(prov_type_collapsed)[T.Optometry]
0.146580
C(prov_type_collapsed)[T.Nephrology]
0.154639
C(prov_type_collapsed)[T.Occupational Therapist in Private Practice]
0.197368
C(prov_type_collapsed)[T.Pathology]
0.260870
C(prov_type_collapsed)[T.Pulmonary Disease]
0.300000
C(prov_type_collapsed)[T.Critical Care (Intensivists)]
0.357143
C(prov_type_collapsed)[T.Physician Assistant]
0.378705
C(prov_type_collapsed)[T.Diagnostic Radiology]
0.397351
C(prov_type_collapsed)[T.Infectious Disease]
0.428571
C(prov_type_collapsed)[T.Radiation Oncology]
0.441176
C(prov_type_collapsed)[T.Other]
0.448776
C(prov_type_collapsed)[T.Medical Oncology]
0.454545
C(prov_type_collapsed)[T.Certified Registered Nurse Anesthetist (CRNA)]
0.496894
C(prov_type_collapsed)[T.Clinical Cardiac Electrophysiology]
0.500000
C(prov_type_collapsed)[T.Physical Therapist in Private Practice]
0.512821
C(prov_type_collapsed)[T.Ophthalmology]
0.562500
C(prov_type_collapsed)[T.Gastroenterology]
0.612245
```

```
C(prov_type_collapsed)[T.Cardiology]
0.648649
C(prov_type_collapsed)[T.Psychologist, Clinical]
0.701754
C(prov_type_collapsed)[T.Urology]
0.714286
C(prov_type_collapsed)[T.Endocrinology]
0.762712
C(prov_type_collapsed)[T.Hematology-Oncology]
0.789474
C(prov_type_collapsed)[T.Orthopedic Surgery]
0.829493
C(prov_type_collapsed)[T.Nurse Practitioner]
0.935484
C(prov_type_collapsed)[T.General Surgery]
0.953390
C(prov_type_collapsed)[T.Otolaryngology]
1.121495
C(prov_type_collapsed)[T.Emergency Medicine]
1.185102
C(prov_type_collapsed)[T.Licensed Clinical Social Worker]
1.197183
C(prov_type_collapsed)[T.Neurology]
1.213235
C(prov_type_collapsed)[T.Anesthesiology]
1.339286
C(prov_type_collapsed)[T.Chiropractic]
1.359890
C(prov_type_collapsed)[T.Neurosurgery]
1.578947
C(prov_type_collapsed)[T.Obstetrics & Gynecology]
1.787234
C(prov_type_collapsed)[T.Plastic and Reconstructive Surgery]
1.818182
C(prov_type_collapsed)[T.Internal Medicine]
1.912088
C(prov_type_collapsed)[T.Rheumatology]
2.250000
C(prov_type_collapsed)[T.Podiatry]
2.378049
C(prov_type_collapsed)[T.Family Practice]
2.598152
C(prov_type_collapsed)[T.Psychiatry]
2.881773
C(prov_type_collapsed)[T.Physical Medicine and Rehabilitation]
4.044944
C(prov_type_collapsed)[T.General Practice]
5.816327
```

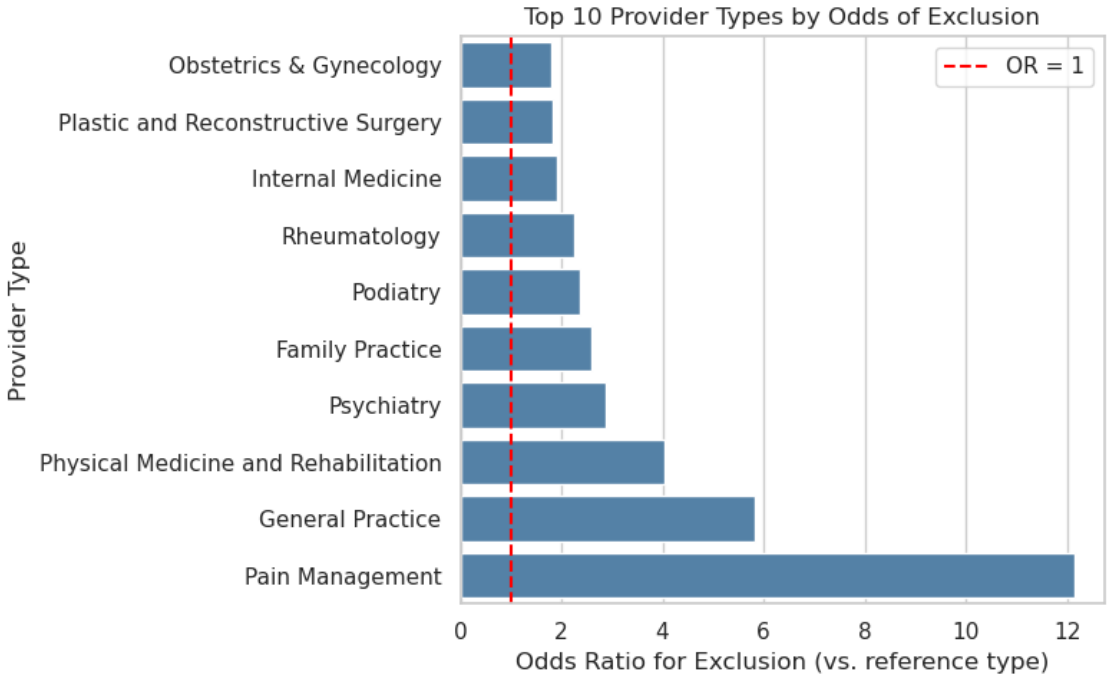
```
C(prov_type_collapsed)[T.Pain Management]
12.142857
dtype: float64
```

[202]:
```python
# drop the intercept
or_series = odds_ratios.drop('Intercept')

# build a DataFrame
or_df = (
    or_series
    .reset_index()
    .rename(columns={'index': 'term', 0: 'or'})
)

# extract provider type name
or_df['provider_type'] = (
    or_df['term']
    .str.replace('C(prov_type_collapsed)[T.', '', regex=False)
    .str.rstrip(']')
)

or_df = or_df.sort_values('or', ascending=False)

plt.figure(figsize=(10, 6))
sns.scatterplot(
    data=or_df,
    x='or',
    y='provider_type',
    color='black'
)

plt.axvline(1.0, color='red', linestyle='--', label='No difference (OR = 1)')
plt.xlabel('Odds Ratio for Exclusion (vs. reference provider type)')
plt.ylabel('Provider Type')
plt.title('Adjusted Odds of Exclusion by Provider Type')
plt.legend()
plt.tight_layout()
plt.show()
```



[203]:
```
# Drop intercept and build DataFrame
or_series = odds_ratios.drop('Intercept')

or_df = (
    or_series.reset_index()
             .rename(columns={'index': 'term', 0: 'or'})
)

# Extract provider type names
or_df['provider_type'] = (
    or_df['term']
    .str.replace('C(prov_type_collapsed)[T.', '', regex=False)
    .str.rstrip(']')
)

# Keep top 10 by odds ratio
top10 = or_df.sort_values('or', ascending=False).head(10)


max_len = 40
top10['label'] = top10['provider_type'].str.slice(0, max_len)

# horizontal bar chart, largest at top
plt.figure(figsize=(8, 5))
sns.barplot(
```

```
    data=top10.sort_values('or', ascending=True),  # ascending so highest at top
    x='or',
    y='label',
    color='steelblue'
)

plt.axvline(1.0, color='red', linestyle='--', label='OR = 1')
plt.xlabel('Odds Ratio for Exclusion (vs. reference type)')
plt.ylabel('Provider Type')
plt.title('Top 10 Provider Types by Odds of Exclusion')
plt.legend()
plt.tight_layout()
plt.show()
```



[204]:
```
model_df = df[[
    'excluded',                  # 0/1 target
    'Rndrng_Prvdr_Type',         # provider type
    'Rndrng_Prvdr_State_Abrvtn', # state
    'Tot_Mdcr_Pymt_Amt'          # funding
]].dropna()

# Make sure excluded is numeric 0/1
model_df['excluded'] = model_df['excluded'].astype(int)
```

```
# Log-transform funding
model_df['log_funding'] = np.log1p(model_df['Tot_Mdcr_Pymt_Amt'])

# Collapse rare provider types
type_counts = model_df['Rndrng_Prvdr_Type'].value_counts()
rare_types = type_counts[type_counts < 50].index
model_df['Rndrng_Prvdr_Type_simplified'] = model_df['Rndrng_Prvdr_Type'].where(
    ~model_df['Rndrng_Prvdr_Type'].isin(rare_types),
    other='Other'
)

# Collapse rare states
state_counts = model_df['Rndrng_Prvdr_State_Abrvtn'].value_counts()
rare_states = state_counts[state_counts < 50].index
model_df['Rndrng_Prvdr_State_simplified'] =␣
 ↪model_df['Rndrng_Prvdr_State_Abrvtn'].where(
    ~model_df['Rndrng_Prvdr_State_Abrvtn'].isin(rare_states),
    other='Other'
)
```

[205]:
```
y = model_df['excluded'].astype(float)

X = model_df[[
    'Rndrng_Prvdr_Type_simplified',
    'Rndrng_Prvdr_State_simplified',
    'log_funding'
]]

# One-hot encode categoricals and ensure float dummies
X = pd.get_dummies(
    X,
    columns=['Rndrng_Prvdr_Type_simplified', 'Rndrng_Prvdr_State_simplified'],
    drop_first=True,
    dtype=float          # this avoids bool columns
)

# Drop zero-variance columns
X = X.loc[:, X.nunique() > 1]

# Add intercept and ensure float
X = sm.add_constant(X).astype(float)

print(X.dtypes.head())   # should now show all float64

logit_model = sm.Logit(y, X).fit()
print(logit_model.summary())
```

```
const                                          float64
```

```
log_funding                                                    float64
Rndrng_Prvdr_Type_simplified_Ambulatory Surgical Center        float64
Rndrng_Prvdr_Type_simplified_Anesthesiology                    float64
Rndrng_Prvdr_Type_simplified_Audiologist                       float64
dtype: object
Warning: Maximum number of iterations has been exceeded.
         Current function value: 0.225812
         Iterations: 35
                           Logit Regression Results
==============================================================================
Dep. Variable:                excluded   No. Observations:                12496
Model:                           Logit   Df Residuals:                    12411
Method:                            MLE   Df Model:                           84
Date:                 Wed, 17 Dec 2025   Pseudo R-squ.:                  0.1041
Time:                         22:15:51   Log-Likelihood:                -2821.7
converged:                       False   LL-Null:                       -3149.5
Covariance Type:             nonrobust   LLR p-value:                 2.199e-89
==============================================================================
==============================================================
 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
--------------------------------------------------------------
const
-3.9108      0.586     -6.676      0.000      -5.059      -2.763
log_funding
0.1677      0.028      5.904      0.000       0.112       0.223
Rndrng_Prvdr_Type_simplified_Ambulatory Surgical Center
-21.8424   2.29e+04     -0.001      0.999   -4.49e+04    4.49e+04
Rndrng_Prvdr_Type_simplified_Anesthesiology
0.3795      0.438      0.866      0.387      -0.480       1.239
Rndrng_Prvdr_Type_simplified_Audiologist
-16.1469   1836.395     -0.009      0.993   -3615.416    3583.122
Rndrng_Prvdr_Type_simplified_Cardiology
-0.5194      0.538     -0.965      0.334      -1.574       0.535
Rndrng_Prvdr_Type_simplified_Centralized Flu
-16.9419   1959.253     -0.009      0.993   -3857.007    3823.123
Rndrng_Prvdr_Type_simplified_Certified Registered Nurse Anesthetist (CRNA)
-0.5385      0.478     -1.128      0.259      -1.474       0.397
Rndrng_Prvdr_Type_simplified_Chiropractic
0.7153      0.446      1.605      0.109      -0.158       1.589
Rndrng_Prvdr_Type_simplified_Dermatology
-2.3001      1.081     -2.129      0.033      -4.418      -0.182
Rndrng_Prvdr_Type_simplified_Diagnostic Radiology
-0.9920      0.536     -1.851      0.064      -2.043       0.059
Rndrng_Prvdr_Type_simplified_Emergency Medicine
0.2336      0.438      0.534      0.594      -0.624       1.092
Rndrng_Prvdr_Type_simplified_Endocrinology
-0.1920      0.716     -0.268      0.789      -1.596       1.212
```

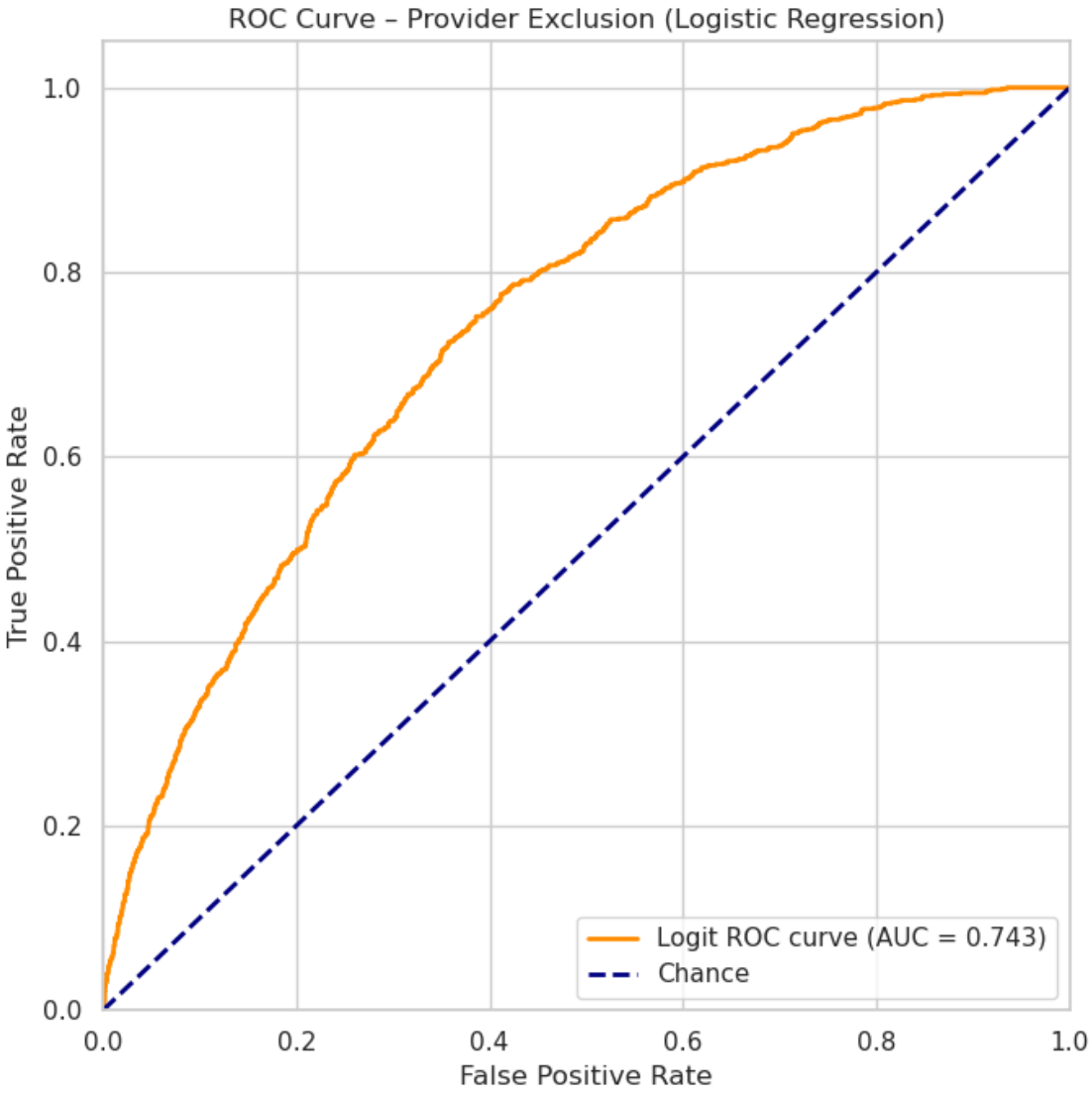
```
Rndrng_Prvdr_Type_simplified_Family Practice
1.0588      0.410      2.581      0.010       0.255       1.863
Rndrng_Prvdr_Type_simplified_Gastroenterology
-0.4835      0.577     -0.837      0.402      -1.615       0.648
Rndrng_Prvdr_Type_simplified_General Practice
1.8904      0.492      3.844      0.000       0.927       2.854
Rndrng_Prvdr_Type_simplified_General Surgery
-0.0128      0.481     -0.027      0.979      -0.956       0.930
Rndrng_Prvdr_Type_simplified_Hematology-Oncology
-0.3763      0.610     -0.616      0.538      -1.573       0.820
Rndrng_Prvdr_Type_simplified_Hospitalist
-2.0079      1.081     -1.857      0.063      -4.127       0.111
Rndrng_Prvdr_Type_simplified_Infectious Disease
-0.8228      0.823     -1.000      0.317      -2.436       0.790
Rndrng_Prvdr_Type_simplified_Internal Medicine
0.6818      0.411      1.657      0.097      -0.124       1.488
Rndrng_Prvdr_Type_simplified_Licensed Clinical Social Worker
0.5876      0.479      1.226      0.220      -0.352       1.527
Rndrng_Prvdr_Type_simplified_Mass Immunizer Roster Biller
-21.2897   1.05e+04     -0.002      0.998   -2.07e+04    2.06e+04
Rndrng_Prvdr_Type_simplified_Nephrology
-2.0378      1.082     -1.884      0.060      -4.158       0.083
Rndrng_Prvdr_Type_simplified_Neurology
0.2117      0.508      0.416      0.677      -0.785       1.208
Rndrng_Prvdr_Type_simplified_Neurosurgery
0.4226      0.588      0.719      0.472      -0.729       1.574
Rndrng_Prvdr_Type_simplified_Nurse Practitioner
0.2124      0.420      0.506      0.613      -0.610       1.035
Rndrng_Prvdr_Type_simplified_Obstetrics & Gynecology
0.8845      0.454      1.949      0.051      -0.005       1.774
Rndrng_Prvdr_Type_simplified_Occupational Therapist in Private Practice
-1.3569      1.085     -1.250      0.211      -3.484       0.771
Rndrng_Prvdr_Type_simplified_Ophthalmology
-0.7786      0.577     -1.350      0.177      -1.909       0.352
Rndrng_Prvdr_Type_simplified_Optometry
-1.7136      0.707     -2.425      0.015      -3.099      -0.329
Rndrng_Prvdr_Type_simplified_Orthopedic Surgery
-0.2645      0.497     -0.532      0.595      -1.239       0.710
Rndrng_Prvdr_Type_simplified_Other
0.5608      0.416      1.349      0.177      -0.254       1.376
Rndrng_Prvdr_Type_simplified_Otolaryngology
0.1444      0.543      0.266      0.790      -0.920       1.209
Rndrng_Prvdr_Type_simplified_Pathology
-1.2742      0.819     -1.557      0.120      -2.878       0.330
Rndrng_Prvdr_Type_simplified_Physical Medicine and Rehabilitation
1.3806      0.462      2.986      0.003       0.474       2.287
Rndrng_Prvdr_Type_simplified_Physical Therapist in Private Practice
-0.4926      0.461     -1.069      0.285      -1.395       0.410
```

```
Rndrng_Prvdr_Type_simplified_Physician Assistant
-0.6620      0.456     -1.450      0.147      -1.557       0.233
Rndrng_Prvdr_Type_simplified_Podiatry
0.8173      0.452      1.807      0.071      -0.069       1.704
Rndrng_Prvdr_Type_simplified_Psychiatry
1.3622      0.441      3.087      0.002       0.497       2.227
Rndrng_Prvdr_Type_simplified_Psychologist, Clinical
-0.1819      0.541     -0.336      0.737      -1.243       0.879
Rndrng_Prvdr_Type_simplified_Pulmonary Disease
-1.2712      0.819     -1.553      0.120      -2.876       0.333
Rndrng_Prvdr_Type_simplified_Urology
-0.4419      0.650     -0.680      0.497      -1.716       0.832
Rndrng_Prvdr_State_simplified_AR
-0.9486      0.497     -1.910      0.056      -1.922       0.025
Rndrng_Prvdr_State_simplified_AZ
-0.8966      0.373     -2.401      0.016      -1.628      -0.165
Rndrng_Prvdr_State_simplified_CA
-0.4254      0.271     -1.570      0.116      -0.956       0.106
Rndrng_Prvdr_State_simplified_CO
-1.9966      0.565     -3.534      0.000      -3.104      -0.889
Rndrng_Prvdr_State_simplified_CT
-1.2161      0.442     -2.750      0.006      -2.083      -0.349
Rndrng_Prvdr_State_simplified_FL
-0.3421      0.277     -1.236      0.217      -0.885       0.200
Rndrng_Prvdr_State_simplified_GA
-0.9316      0.344     -2.707      0.007      -1.606      -0.257
Rndrng_Prvdr_State_simplified_IA
-1.1106      0.492     -2.258      0.024      -2.075      -0.147
Rndrng_Prvdr_State_simplified_ID
-1.2467      0.769     -1.621      0.105      -2.754       0.260
Rndrng_Prvdr_State_simplified_IL
-0.7329      0.310     -2.364      0.018      -1.341      -0.125
Rndrng_Prvdr_State_simplified_IN
-2.4720      0.633     -3.903      0.000      -3.713      -1.231
Rndrng_Prvdr_State_simplified_KS
-1.7287      0.640     -2.703      0.007      -2.982      -0.475
Rndrng_Prvdr_State_simplified_KY
-0.3963      0.359     -1.104      0.270      -1.100       0.307
Rndrng_Prvdr_State_simplified_LA
-0.3042      0.381     -0.799      0.424      -1.050       0.442
Rndrng_Prvdr_State_simplified_MA
-1.4332      0.388     -3.698      0.000      -2.193      -0.674
Rndrng_Prvdr_State_simplified_MD
-1.1798      0.411     -2.870      0.004      -1.986      -0.374
Rndrng_Prvdr_State_simplified_ME
-0.1915      0.477     -0.402      0.688      -1.126       0.743
Rndrng_Prvdr_State_simplified_MI
-0.4832      0.304     -1.591      0.112      -1.078       0.112
```

```
Rndrng_Prvdr_State_simplified_MN
-2.0310      0.564     -3.598      0.000      -3.137      -0.925
Rndrng_Prvdr_State_simplified_MO
-0.4958      0.347     -1.429      0.153      -1.176       0.184
Rndrng_Prvdr_State_simplified_MS
-0.1067      0.393     -0.272      0.786      -0.876       0.663
Rndrng_Prvdr_State_simplified_NC
-1.3120      0.371     -3.541      0.000      -2.038      -0.586
Rndrng_Prvdr_State_simplified_NE
-1.1256      0.575     -1.958      0.050      -2.252       0.001
Rndrng_Prvdr_State_simplified_NH
-15.6234    710.685     -0.022      0.982   -1408.541    1377.294
Rndrng_Prvdr_State_simplified_NJ
-1.2470      0.353     -3.534      0.000      -1.939      -0.555
Rndrng_Prvdr_State_simplified_NM
0.1483      0.431      0.344      0.731      -0.697       0.994
Rndrng_Prvdr_State_simplified_NV
-0.3481      0.455     -0.765      0.444      -1.240       0.544
Rndrng_Prvdr_State_simplified_NY
-0.7345      0.285     -2.580      0.010      -1.293      -0.176
Rndrng_Prvdr_State_simplified_OH
-0.3735      0.292     -1.279      0.201      -0.946       0.199
Rndrng_Prvdr_State_simplified_OK
-0.6450      0.417     -1.548      0.122      -1.462       0.172
Rndrng_Prvdr_State_simplified_OR
-0.2742      0.367     -0.748      0.455      -0.993       0.445
Rndrng_Prvdr_State_simplified_Other
-0.5573      0.319     -1.748      0.081      -1.182       0.068
Rndrng_Prvdr_State_simplified_PA
-0.6464      0.294     -2.199      0.028      -1.223      -0.070
Rndrng_Prvdr_State_simplified_PR
-0.3715      0.500     -0.744      0.457      -1.351       0.608
Rndrng_Prvdr_State_simplified_SC
-1.5645      0.487     -3.212      0.001      -2.519      -0.610
Rndrng_Prvdr_State_simplified_TN
-0.0384      0.308     -0.125      0.901      -0.641       0.565
Rndrng_Prvdr_State_simplified_TX
-0.3451      0.277     -1.248      0.212      -0.887       0.197
Rndrng_Prvdr_State_simplified_UT
-1.9977      0.761     -2.624      0.009      -3.490      -0.506
Rndrng_Prvdr_State_simplified_VA
-0.7633      0.338     -2.257      0.024      -1.426      -0.100
Rndrng_Prvdr_State_simplified_WA
-0.6014      0.339     -1.776      0.076      -1.265       0.062
Rndrng_Prvdr_State_simplified_WI
-1.3261      0.424     -3.128      0.002      -2.157      -0.495
Rndrng_Prvdr_State_simplified_WV
-0.2315      0.509     -0.455      0.649      -1.229       0.766
```

```
==============================================================================
==============================================================

/opt/conda/lib/python3.12/site-packages/statsmodels/base/model.py:607:
ConvergenceWarning: Maximum Likelihood optimization failed to converge. Check
mle_retvals
  warnings.warn("Maximum Likelihood optimization failed to "
```

[206]:
```python
from sklearn.metrics import roc_curve, auc

y_score = logit_model.predict(X)

fpr, tpr, _ = roc_curve(y, y_score)
roc_auc = auc(fpr, tpr)

plt.figure(figsize=(7, 7))
plt.plot(fpr, tpr, color='darkorange', lw=2,
         label=f'Logit ROC curve (AUC = {roc_auc:0.3f})')
plt.plot([0, 1], [0, 1], color='navy', lw=2, linestyle='--', label='Chance')

plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.05])
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('ROC Curve - Provider Exclusion (Logistic Regression)')
plt.legend(loc="lower right")
plt.tight_layout()
plt.show()
```


ROC Curve – Provider Exclusion (Logistic Regression)
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Logit ROC curve (AUC = 0.743)
Chance