Aarki

August 4, 2024

[127]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import statsmodels.api as sm
```

[128]:
```
from google.colab import drive
drive.mount('/content/drive/')
```

```
Drive already mounted at /content/drive/; to attempt to forcibly remount, call
drive.mount("/content/drive/", force_remount=True).
```

[129]:
```
df=pd.read_csv("/content/drive/MyDrive/Datasets/case_study.csv")
```

1 EDA

[130]:
```
len(df)
```

[130]: 8819

[131]:
```
# Removing all empty rows
df = df.dropna(how='all')
```

[132]:
```
df.tail(10)
```

[132]:
```
     LeadCreated FirstName                           Email  \
3011     7/28/09   RUSSELL          russgarland196@aol.com
3012     7/15/09   Michael        mrussell1127@hotmail.com
3013     6/16/09     jason        jasonjkauffman@gmail.com
3014      4/3/09       Kim       k-davis@cookchildrens.org
3015     8/17/09       naw             doreenpan@gmail.com
3016     6/30/09       amy                 ahokett@sjc.edu
3017     4/25/09    brandy           brandy75137@yahoo.com
3018     4/12/09  jennifer  jennifer_woods48375@yahoo.com
3019     9/23/09     debra              debraroque@att.net
3020     4/27/09    Ricard           silva3131@sbcglobal.net

                              VendorLeadID                CallStatus  \
```

```
3011  895F1933-648B-4185-9CDE-908BBF0F19BC                          NaN
3012                Z26Md0QFrkuMgZRApiJf0w  Contacted - Doesn't Qualify
3013  8967DD81-F17D-4D9F-84F2-DED051BA3BE2                          NaN
3014  B85A389B-6FB7-44FA-A758-0819BFFB4361                          NaN
3015  51DDB7F0-D901-46AC-8949-BB9F3AF83B71                          NaN
3016  F42BBDFB-D44B-4890-A1B6-D1F0A3838214                          NaN
3017  EA3703A3-61C9-40CE-92F7-61F0E02B1365                          NaN
3018  64EB2632-E29E-4EB4-B361-45F1F0C735B5                          NaN
3019  9126C7DB-F5BB-4CE1-BDCF-5E6F1841ADAD                          NaN
3020  AF369BF3-15C7-4B47-8E2D-FE218A5DBCC3                          NaN

                                     WidgetName PublisherZoneName  \
3011   w-302252-DebtReduction1-1DC-CreditSolutions    TopLeft-302252
3012  w-302252-DebtReduction1-1DC-yellowarrow-blue    TopLeft-302252
3013                   w-302252-DebtReduction1-1DC    TopLeft-302252
3014         w-300250-DebtReduction1-2DC-BlueMeter    TopLeft-302252
3015       w-302252-DebtReduction1-1DC-yellowarrow    TopLeft-302252
3016             w-302252-DebtReduction1-1DC-white    TopLeft-302252
3017             w-300250-DebtReduction1-1DC-Head2    TopLeft-302252
3018   w-300250-DebtReduction1-1DC-CreditSolutions    TopLeft-302252
3019  w-302252-DebtReduction1-1DC-yellowarrow-dark    TopLeft-302252
3020             w-300250-DebtReduction1-1DC-Head2    TopLeft-302252

     PublisherCampaignName  AddressScore  PhoneScore  … Partner  \
3011       DebtReductionInc           5.0         4.0  …  google
3012       DebtReductionInc           5.0         3.0  …  google
3013       DebtReductionInc           NaN         NaN  …  google
3014       DebtReductionInc           NaN         NaN  …   yahoo
3015       DebtReductionInc           5.0         3.0  …  Google
3016       DebtReductionInc           NaN         NaN  …   yahoo
3017       DebtReductionInc           NaN         NaN  …  Google
3018       DebtReductionInc           NaN         NaN  …   yahoo
3019       DebtReductionInc           5.0         5.0  …  Google
3020       DebtReductionInc           NaN         NaN  …  Google

                  ReferralDomain  MarketingCampaign              AdGroup  \
3011              www.google.com  Debt Holding Tank  Holding Tank - Debt
3012              www.google.com       Debt General       Lower Payments
3013              www.google.com       Debt General         Student Debt
3014            search.yahoo.com   DebtReductionInc   Debt Consolidation
3015     www.debtreductioninc.com   DebtReductionInc       Lower Payments
3016                 www.att.net   DebtReductionInc   Debt Consolidation
3017  googleads.g.doubleclick.net   DebtReductionInc         Student Debt
3018            search.yahoo.com   DebtReductionInc   Debt Consolidation
3019                         NaN              state     Debt Negotiation
3020                www.ehow.com   DebtReductionInc       Lower Payments
```

```
                    Keyword  \
3011               Debt cures
3012  Lower monthly car payments
3013        Student loan default
3014                        NaN
3015                        NaN
3016                        NaN
3017                        NaN
3018                        NaN
3019                        NaN
3020                        NaN

                                         SearchQuery  \
3011                                      debt cures
3012                        lower monthly car payment
3013                         student loans in default
3014    government debt consolidation fort worth texas
3015                                             NaN
3016                                             NaN
3017                                             NaN
3018  debt consolidation or settlement which is better
3019                                             NaN
3020                                             NaN

                                           ReferralURL  \
3011                      http://www.google.com/search
3012                         http://www.google.com/url
3013                      http://www.google.com/search
3014  http://search.yahoo.com/search;_ylt=A0geu4teHt…
3015       http://www.debtreductioninc.com/index12.html
3016                         http://www.att.net/s/s.dll
3017     http://googleads.g.doubleclick.net/pagead/ads
3018                      http://search.yahoo.com/search
3019                                               NaN
3020         http://www.ehow.com/ehow_radlinks_ads.html

                                ReferralURL Parameters  \
3011          hl=en&q=debt cures&aq=7&oq=DEBT&aqi=g10
3012  q=lower monthly car payment&url=/aclk%3Fsa%3Dl…
3013  sourceid=navclient&ie=UTF-8&rlz=1T4ADBF_enUS32…
3014  p=government debt consolidation fort worth tex…
3015  utm_source=Google&utm_medium=cpc&utm_campaign=…
3016  spage=search/error.htm&searchtype=epa&source=a…
3017  client=ca-pub-7025449865608971&dt=124069479896…
3018  ei=UTF-8&fr=yfp-t-501&SpellState=n-1665662351_…
3019                                               NaN
3020  term=Lower Monthly Car Payments&channel=fin_mo…
```

```
                                   LandingPageURL  \
3011   http://www.debtreductioninc.com/index8.html
3012   http://www.debtreductioninc.com/index8.html
3013   http://www.debtreductioninc.com/index8.html
3014   http://www.debtreductioninc.com/index8.html
3015  http://www.debtreductioninc.com/index11.html
3016   http://www.debtreductioninc.com/index8.html
3017   http://www.debtreductioninc.com/index8.html
3018   http://www.debtreductioninc.com/index8.html
3019   http://www.debtreductioninc.com/index8.html
3020   http://www.debtreductioninc.com/index8.html

                           Landing Page URL Parameters
3011  utm_source=google&utm_medium=CPC&utm_content=H…
3012  utm_source=google&utm_medium=CPC&utm_content=L…
3013  utm_source=google&utm_medium=CPC&utm_content=S…
3014  utm_source=yahoo&utm_medium=cpc&utm_campaign=D…
3015  utm_source=Google&utm_medium=cpc&utm_campaign=…
3016  utm_source=yahoo&utm_medium=cpc&utm_campaign=D…
3017  utm_source=Google&utm_medium=cpc&utm_campaign=…
3018  utm_source=yahoo&utm_medium=cpc&utm_campaign=D…
3019  utm_source=Google&utm_medium=cpc&utm_campaign=…
3020  utm_source=Google&utm_medium=cpc&utm_campaign=…

[10 rows x 24 columns]
```

[133]:

```python
# dropping IP Address column as its not needed
df = df.drop(columns=['IP Address'])
```

[134]:

```python
df
```

[134]:

```
     LeadCreated  FirstName                         Email  \
0         7/1/09    Dorinda       kanani@sandwichisles.net
1        4/13/09   Presetta          clerk2@ustconline.net
2        4/21/09       Gina         wagoner_gina@yahoo.com
3         8/3/09       Kari              usa4ley@yahoo.com
4        4/13/09  Stephanie       sr1lambert@embarqmail.com
…            …          …                              …
3016     6/30/09        amy                ahokett@sjc.edu
3017     4/25/09     brandy          brandy75137@yahoo.com
3018     4/12/09   jennifer  jennifer_woods48375@yahoo.com
3019     9/23/09      debra             debraroque@att.net
3020     4/27/09     Ricard          silva3131@sbcglobal.net

                              VendorLeadID  \
0     FDF81FDA-A649-437B-B99C-FDDE74F7971B
```

```
1     4190ACB7-5026-416C-B987-ED8AD427D5E6
2                   hFg80jf_R0CRN55hdhWILw
3                   jB01QgYZxkWArI9jWxuufw
4     D5B32074-458E-40EC-B185-1FEF20AC626D
…                                        …
3016  F42BBDFB-D44B-4890-A1B6-D1F0A3838214
3017  EA3703A3-61C9-40CE-92F7-61F0E02B1365
3018  64EB2632-E29E-4EB4-B361-45F1F0C735B5
3019  9126C7DB-F5BB-4CE1-BDCF-5E6F1841ADAD
3020  AF369BF3-15C7-4B47-8E2D-FE218A5DBCC3

                                    CallStatus  \
0                                          NaN
1                                          NaN
2     Unable to contact - Bad Contact Information
3                    Contacted - Doesn't Qualify
4                                          NaN
…                                            …
3016                                       NaN
3017                                       NaN
3018                                       NaN
3019                                       NaN
3020                                       NaN

                                     WidgetName PublisherZoneName  \
0      w-302252-DebtReduction1-1DC-CreditSolutions    TopLeft-302252
1               w-300250-DebtReduction1-1DC-Head2    TopLeft-302252
2               w-300250-DebtReduction1-1DC-Head2    TopLeft-302252
3               w-302252-DebtReduction1-1DC-white    TopLeft-302252
4           w-300250-DebtReduction1-1DC-BlueMeter    TopLeft-302252
…                                             …                …
3016            w-302252-DebtReduction1-1DC-white    TopLeft-302252
3017            w-300250-DebtReduction1-1DC-Head2    TopLeft-302252
3018  w-300250-DebtReduction1-1DC-CreditSolutions    TopLeft-302252
3019 w-302252-DebtReduction1-1DC-yellowarrow-dark    TopLeft-302252
3020            w-300250-DebtReduction1-1DC-Head2    TopLeft-302252

     PublisherCampaignName  AddressScore  PhoneScore  …      Partner  \
0          DebtReductionInc           NaN         5.0  …       google
1          DebtReductionInc           NaN         NaN  …  AdKnowledge
2          DebtReductionInc           NaN         NaN  …  AdKnowledge
3          DebtReductionInc           5.0         3.0  …       Google
4          DebtReductionInc           NaN         NaN  …       Google
…                        …             …           …  …          …
3016       DebtReductionInc           NaN         NaN  …        yahoo
3017       DebtReductionInc           NaN         NaN  …       Google
3018       DebtReductionInc           NaN         NaN  …        yahoo
```

```
3019     DebtReductionInc           5.0         5.0  …       Google
3020     DebtReductionInc           NaN         NaN  …       Google

                  ReferralDomain   MarketingCampaign             AdGroup  \
0                 www.google.com   Debt Holding Tank  Holding Tank - Debt
1                            NaN  Financial Services          Consolidate
2        us.mc582.mail.yahoo.com  Financial Services          Consolidate
3              norwich.kijiji.com    DebtReductionInc       Lower Payments
4                            NaN    DebtReductionInc       Debt Reduction
…                              …                   …                    …
3016                 www.att.net    DebtReductionInc   Debt Consolidation
3017  googleads.g.doubleclick.net    DebtReductionInc         Student Debt
3018            search.yahoo.com    DebtReductionInc   Debt Consolidation
3019                         NaN               state     Debt Negotiation
3020                www.ehow.com    DebtReductionInc       Lower Payments

                 Keyword                                        SearchQuery  \
0     Debt specialists                                    debt specialists
1                  NaN                                                 NaN
2                  NaN                                                 NaN
3                  NaN                                                 NaN
4                  NaN                                                 NaN
…                    …                                                   …
3016               NaN                                                 NaN
3017               NaN                                                 NaN
3018               NaN  debt consolidation or settlement which is better
3019               NaN                                                 NaN
3020               NaN                                                 NaN

                                          ReferralURL  \
0                       http://www.google.com/search
1                                                NaN
2        http://us.mc582.mail.yahoo.com/mc/showMessage
3     http://norwich.kijiji.com/c-Cars-vehicles-Cars…
4                                                NaN
…                                                  …
3016                      http://www.att.net/s/s.dll
3017     http://googleads.g.doubleclick.net/pagead/ads
3018                    http://search.yahoo.com/search
3019                                               NaN
3020         http://www.ehow.com/ehow_radlinks_ads.html

                              ReferralURL Parameters  \
0     sourceid=navclient&aq=1&oq=debt sp&ie=UTF-8&rl…
1                                                NaN
2     &fid=Inbox&sort=date&order=down&startMid=0&.ra…
3                                                NaN
```

```
4                                                     NaN
…                                                   …
3016   spage=search/error.htm&searchtype=epa&source=a…
3017   client=ca-pub-7025449865608971&dt=124069479896…
3018   ei=UTF-8&fr=yfp-t-501&SpellState=n-1665662351_…
3019                                                  NaN
3020   term=Lower Monthly Car Payments&channel=fin_mo…

                                    LandingPageURL  \
0      http://www.debtreductioninc.com/index8.html
1      http://www.debtreductioninc.com/index8.html
2      http://www.debtreductioninc.com/index8.html
3     http://www.debtreductioninc.com/index12.html
4      http://www.debtreductioninc.com/index8.html
…                                              …
3016   http://www.debtreductioninc.com/index8.html
3017   http://www.debtreductioninc.com/index8.html
3018   http://www.debtreductioninc.com/index8.html
3019   http://www.debtreductioninc.com/index8.html
3020   http://www.debtreductioninc.com/index8.html

                              Landing Page URL Parameters
0      utm_source=google&utm_medium=CPC&utm_content=H…
1      utm_source=AdKnowledge&utm_medium=CPC&utm_cont…
2      utm_source=AdKnowledge&utm_medium=CPC&utm_cont…
3      utm_source=Google&utm_medium=cpc&utm_campaign=…
4      utm_source=Google&utm_medium=cpc&utm_campaign=…
…                                                   …
3016   utm_source=yahoo&utm_medium=cpc&utm_campaign=D…
3017   utm_source=Google&utm_medium=cpc&utm_campaign=…
3018   utm_source=yahoo&utm_medium=cpc&utm_campaign=D…
3019   utm_source=Google&utm_medium=cpc&utm_campaign=…
3020   utm_source=Google&utm_medium=cpc&utm_campaign=…

[3021 rows x 23 columns]
```

[135]:
```
df.dtypes
```

[135]:
```
LeadCreated                    object
FirstName                      object
Email                          object
VendorLeadID                   object
CallStatus                     object
WidgetName                     object
PublisherZoneName              object
PublisherCampaignName          object
AddressScore                  float64
```

```
PhoneScore                    float64
AdvertiserCampaignName         object
State                          object
DebtLevel                      object
Partner                        object
ReferralDomain                 object
MarketingCampaign              object
AdGroup                        object
Keyword                        object
SearchQuery                    object
ReferralURL                    object
ReferralURL Parameters         object
LandingPageURL                 object
Landing Page URL Parameters    object
dtype: object
```

[136]:
```python
# converting datatype of LeadCreated from object to date for easier analysis
df['LeadCreated'] = pd.to_datetime(df['LeadCreated'])
```

```
<ipython-input-136-2cb066297cbd>:2: UserWarning: Could not infer format, so each
element will be parsed individually, falling back to `dateutil`. To ensure
parsing is consistent and as-expected, please specify a format.
  df['LeadCreated'] = pd.to_datetime(df['LeadCreated'])
```

[137]:
```python
null_counts=df.isna().sum()
null_counts
```

[137]:
```
LeadCreated                      0
FirstName                        0
Email                            0
VendorLeadID                     0
CallStatus                    2140
WidgetName                       0
PublisherZoneName                0
PublisherCampaignName            0
AddressScore                  1850
PhoneScore                    1628
AdvertiserCampaignName           0
State                            0
DebtLevel                        0
Partner                          0
ReferralDomain                 515
MarketingCampaign              272
AdGroup                        272
Keyword                       2042
SearchQuery                   1755
ReferralURL                    515
```

```
ReferralURL Parameters          738
LandingPageURL                    0
Landing Page URL Parameters       0
dtype: int64
```

[138]:
```
df.nunique()
```

[138]:
```
LeadCreated                     182
FirstName                      1695
Email                          2888
VendorLeadID                   3013
CallStatus                        7
WidgetName                       14
PublisherZoneName                 2
PublisherCampaignName             2
AddressScore                      5
PhoneScore                        5
AdvertiserCampaignName            2
State                            32
DebtLevel                        10
Partner                           6
ReferralDomain                  372
MarketingCampaign                20
AdGroup                         103
Keyword                         285
SearchQuery                     962
ReferralURL                     670
ReferralURL Parameters         2199
LandingPageURL                    5
Landing Page URL Parameters    1090
dtype: int64
```

2 1. Are we seeing any lead quality trends over time (improving, declining)? Are they statistically significant?

[139]:
```
df['CallStatus'].unique()
```

[139]:
```
array([nan, 'Unable to contact - Bad Contact Information',
       "Contacted - Doesn't Qualify", 'Closed', 'EP Received',
       'EP Confirmed', 'Contacted - Invalid Profile', 'EP Sent'],
      dtype=object)
```

[140]:
```
# Creating a new column to map each CallStatus into Bad, Good and Best Lead␣
↪Quality
status_mapping = {
```

```
    'Closed': 'Best lead quality',
    'EP Sent': 'Good lead quality',
    'EP Received': 'Good lead quality',
    'EP Confirmed': 'Good lead quality',
    'Unable to contact - Bad Contact Information': 'Bad lead quality',
    'Contacted - Invalid Profile': 'Bad lead quality',
    "Contacted - Doesn't Qualify" : 'Bad lead quality',
    "nan" :'Unkown'
}
# Map the callstatus values to the new groups
df['lead_quality'] = df['CallStatus'].map(status_mapping)

# Fill the unknown statuses
df['lead_quality'] = df['lead_quality'].fillna('Unknown')

# Display the resulting DataFrame
print(df[['CallStatus', 'lead_quality']])
```

```
                                       CallStatus      lead_quality
0                                             NaN           Unknown
1                                             NaN           Unknown
2     Unable to contact - Bad Contact Information  Bad lead quality
3                     Contacted - Doesn't Qualify  Bad lead quality
4                                             NaN           Unknown
…                                               …                 …
3016                                          NaN           Unknown
3017                                          NaN           Unknown
3018                                          NaN           Unknown
3019                                          NaN           Unknown
3020                                          NaN           Unknown

[3021 rows x 2 columns]
```

[141]:
```
df
```

[141]:
```
     LeadCreated  FirstName                        Email  \
0     2009-07-01    Dorinda     kanani@sandwichisles.net
1     2009-04-13   Presetta        clerk2@ustconline.net
2     2009-04-21       Gina       wagoner_gina@yahoo.com
3     2009-08-03       Kari           usa4ley@yahoo.com
4     2009-04-13  Stephanie     sr1lambert@embarqmail.com
…            …          …                            …
3016  2009-06-30        amy               ahokett@sjc.edu
3017  2009-04-25     brandy        brandy75137@yahoo.com
3018  2009-04-12   jennifer  jennifer_woods48375@yahoo.com
3019  2009-09-23      debra            debraroque@att.net
3020  2009-04-27     Ricard        silva3131@sbcglobal.net
```

```
                             VendorLeadID  \
0     FDF81FDA-A649-437B-B99C-FDDE74F7971B
1     4190ACB7-5026-416C-B987-ED8AD427D5E6
2                   hFg80jf_R0CRN55hdhWILw
3                   jB01QgYZxkWArI9jWxuufw
4     D5B32074-458E-40EC-B185-1FEF20AC626D
…                                        …
3016  F42BBDFB-D44B-4890-A1B6-D1F0A3838214
3017  EA3703A3-61C9-40CE-92F7-61F0E02B1365
3018  64EB2632-E29E-4EB4-B361-45F1F0C735B5
3019  9126C7DB-F5BB-4CE1-BDCF-5E6F1841ADAD
3020  AF369BF3-15C7-4B47-8E2D-FE218A5DBCC3

                                    CallStatus  \
0                                          NaN
1                                          NaN
2     Unable to contact - Bad Contact Information
3                    Contacted - Doesn't Qualify
4                                          NaN
…                                            …
3016                                       NaN
3017                                       NaN
3018                                       NaN
3019                                       NaN
3020                                       NaN

                                     WidgetName PublisherZoneName  \
0      w-302252-DebtReduction1-1DC-CreditSolutions    TopLeft-302252
1               w-300250-DebtReduction1-1DC-Head2    TopLeft-302252
2               w-300250-DebtReduction1-1DC-Head2    TopLeft-302252
3               w-302252-DebtReduction1-1DC-white    TopLeft-302252
4           w-300250-DebtReduction1-1DC-BlueMeter    TopLeft-302252
…                                             …               …
3016            w-302252-DebtReduction1-1DC-white    TopLeft-302252
3017            w-300250-DebtReduction1-1DC-Head2    TopLeft-302252
3018   w-300250-DebtReduction1-1DC-CreditSolutions    TopLeft-302252
3019 w-302252-DebtReduction1-1DC-yellowarrow-dark    TopLeft-302252
3020            w-300250-DebtReduction1-1DC-Head2    TopLeft-302252

     PublisherCampaignName  AddressScore  PhoneScore  …  \
0         DebtReductionInc           NaN         5.0  …
1         DebtReductionInc           NaN         NaN  …
2         DebtReductionInc           NaN         NaN  …
3         DebtReductionInc           5.0         3.0  …
4         DebtReductionInc           NaN         NaN  …
…                        …             …           …  …
```

```
3016     DebtReductionInc           NaN         NaN  …
3017     DebtReductionInc           NaN         NaN  …
3018     DebtReductionInc           NaN         NaN  …
3019     DebtReductionInc           5.0         5.0  …
3020     DebtReductionInc           NaN         NaN  …

                   ReferralDomain   MarketingCampaign             AdGroup  \
0                  www.google.com  Debt Holding Tank  Holding Tank - Debt
1                             NaN Financial Services          Consolidate
2         us.mc582.mail.yahoo.com Financial Services          Consolidate
3             norwich.kijiji.com    DebtReductionInc       Lower Payments
4                             NaN    DebtReductionInc       Debt Reduction
…                             …                   …                    …
3016                  www.att.net    DebtReductionInc  Debt Consolidation
3017  googleads.g.doubleclick.net    DebtReductionInc        Student Debt
3018             search.yahoo.com    DebtReductionInc  Debt Consolidation
3019                          NaN               state    Debt Negotiation
3020                 www.ehow.com    DebtReductionInc      Lower Payments

              Keyword                                       SearchQuery  \
0     Debt specialists                                 debt specialists
1                  NaN                                              NaN
2                  NaN                                              NaN
3                  NaN                                              NaN
4                  NaN                                              NaN
…                 …                                               …
3016               NaN                                              NaN
3017               NaN                                              NaN
3018               NaN  debt consolidation or settlement which is better
3019               NaN                                              NaN
3020               NaN                                              NaN

                                         ReferralURL  \
0                       http://www.google.com/search
1                                                NaN
2        http://us.mc582.mail.yahoo.com/mc/showMessage
3     http://norwich.kijiji.com/c-Cars-vehicles-Cars…
4                                                NaN
…                                                 …
3016                      http://www.att.net/s/s.dll
3017     http://googleads.g.doubleclick.net/pagead/ads
3018                    http://search.yahoo.com/search
3019                                               NaN
3020         http://www.ehow.com/ehow_radlinks_ads.html

                             ReferralURL Parameters  \
0     sourceid=navclient&aq=1&oq=debt sp&ie=UTF-8&rl…
```

```
1                                                     NaN
2     &fid=Inbox&sort=date&order=down&startMid=0&.ra…
3                                                     NaN
4                                                     NaN
…                                                   …
3016  spage=search/error.htm&searchtype=epa&source=a…
3017  client=ca-pub-7025449865608971&dt=124069479896…
3018  ei=UTF-8&fr=yfp-t-501&SpellState=n-1665662351_…
3019                                                  NaN
3020  term=Lower Monthly Car Payments&channel=fin_mo…

                                    LandingPageURL  \
0      http://www.debtreductioninc.com/index8.html
1      http://www.debtreductioninc.com/index8.html
2      http://www.debtreductioninc.com/index8.html
3     http://www.debtreductioninc.com/index12.html
4      http://www.debtreductioninc.com/index8.html
…                                              …
3016   http://www.debtreductioninc.com/index8.html
3017   http://www.debtreductioninc.com/index8.html
3018   http://www.debtreductioninc.com/index8.html
3019   http://www.debtreductioninc.com/index8.html
3020   http://www.debtreductioninc.com/index8.html

                            Landing Page URL Parameters      lead_quality
0     utm_source=google&utm_medium=CPC&utm_content=H…           Unknown
1     utm_source=AdKnowledge&utm_medium=CPC&utm_cont…           Unknown
2     utm_source=AdKnowledge&utm_medium=CPC&utm_cont…  Bad lead quality
3     utm_source=Google&utm_medium=cpc&utm_campaign=…  Bad lead quality
4     utm_source=Google&utm_medium=cpc&utm_campaign=…           Unknown
…                                                   …               …
3016  utm_source=yahoo&utm_medium=cpc&utm_campaign=D…           Unknown
3017  utm_source=Google&utm_medium=cpc&utm_campaign=…           Unknown
3018  utm_source=yahoo&utm_medium=cpc&utm_campaign=D…           Unknown
3019  utm_source=Google&utm_medium=cpc&utm_campaign=…           Unknown
3020  utm_source=Google&utm_medium=cpc&utm_campaign=…           Unknown

[3021 rows x 24 columns]
```

[142]:
```
len(df)
```

[142]: 3021

[143]:
```
# creating a new column year month to extract month from LeadCreate
df['year_month'] = df['LeadCreated'].dt.to_period('M').dt.strftime('%b')
```

[144]:
```
status_counts = df['lead_quality'].value_counts()
status_counts
```

[144]:
```
lead_quality
Unknown               2140
Bad lead quality       488
Best lead quality      245
Good lead quality      148
Name: count, dtype: int64
```

[145]:
```
#creating a new dataframe filtering unknown records for better insights w.r.t.␣
↪lead quality
df_filtered=df[df['lead_quality']!='Unknown']
df_filtered.head()
```

[145]:
```
   LeadCreated FirstName                    Email           VendorLeadID  \
2   2009-04-21      Gina   wagoner_gina@yahoo.com  hFg80jf_ROCRN55hdhWILw
3   2009-08-03      Kari        usa4ley@yahoo.com  jB01QgYZxkWArI9jWxuufw
7   2009-04-22      John      johndoe333@yahoo.com  hxFrkNSCjU6rE2u-7yH-KQ
10  2009-06-01      Juan  villalobosjgv@yahoo.com  LfatQ19SFkWfP3-hH7TVTQ
17  2009-08-01     Kandi    kandielko@verizon.net  7YvjZQLOi0aAT7DhiqDISg

                                 CallStatus  \
2   Unable to contact - Bad Contact Information
3                 Contacted - Doesn't Qualify
7   Unable to contact - Bad Contact Information
10  Unable to contact - Bad Contact Information
17  Unable to contact - Bad Contact Information

                                 WidgetName PublisherZoneName  \
2              w-300250-DebtReduction1-1DC-Head2    TopLeft-302252
3              w-302252-DebtReduction1-1DC-white    TopLeft-302252
7          w-300250-DebtReduction1-2DC-BlueMeter    TopLeft-302252
10  w-302252-DebtReduction1-1DC-CreditSolutions    TopLeft-302252
17  w-302252-DebtReduction1-1DC-CreditSolutions    TopLeft-302252

   PublisherCampaignName  AddressScore  PhoneScore  …   MarketingCampaign  \
2       DebtReductionInc           NaN         NaN  …  Financial Services
3       DebtReductionInc           5.0         3.0  …    DebtReductionInc
7       DebtReductionInc           NaN         NaN  …              Credit
10      DebtReductionInc           NaN         NaN  …    DebtReductionInc
17      DebtReductionInc           3.0         3.0  …    DebtReductionInc

                             AdGroup          Keyword      SearchQuery  \
2                        Consolidate              NaN              NaN
3                     Lower Payments              NaN              NaN
7               Debt Credit Services  Credit services  credit services
```

```
10  Credit Card Debt - high volume              NaN              NaN
17  Credit Card Debt - high volume              NaN              NaN

                                       ReferralURL  \
2       http://us.mc582.mail.yahoo.com/mc/showMessage
3   http://norwich.kijiji.com/c-Cars-vehicles-Cars…
7                        http://www.google.com/search
10      http://googleads.g.doubleclick.net/pagead/ads
17      http://googleads.g.doubleclick.net/pagead/ads

                            ReferralURL Parameters  \
2   &fid=Inbox&sort=date&order=down&startMid=0&.ra…
3                                              NaN
7   q=credit services&rls=com.microsoft:*&ie=UTF-8…
10  client=ca-pub-7277345023380563&host=pub-155622…
17  client=ca-pub-3089121361425291&dt=124917730077…

                               LandingPageURL  \
2    http://www.debtreductioninc.com/index8.html
3   http://www.debtreductioninc.com/index12.html
7    http://www.debtreductioninc.com/index8.html
10   http://www.debtreductioninc.com/index8.html
17   http://www.debtreductioninc.com/index8.html

                         Landing Page URL Parameters      lead_quality  \
2   utm_source=AdKnowledge&utm_medium=CPC&utm_cont…  Bad lead quality
3   utm_source=Google&utm_medium=cpc&utm_campaign=…  Bad lead quality
7   utm_source=google&utm_medium=CPC&utm_content=D…  Bad lead quality
10  utm_source=Google&utm_medium=cpc&utm_campaign=…  Bad lead quality
17  utm_source=Google&utm_medium=cpc&utm_campaign=…  Bad lead quality

   year_month
2         Apr
3         Aug
7         Apr
10        Jun
17        Aug

[5 rows x 25 columns]
```

[146]:
```
status_counts = df_filtered['lead_quality'].value_counts()
status_counts
```

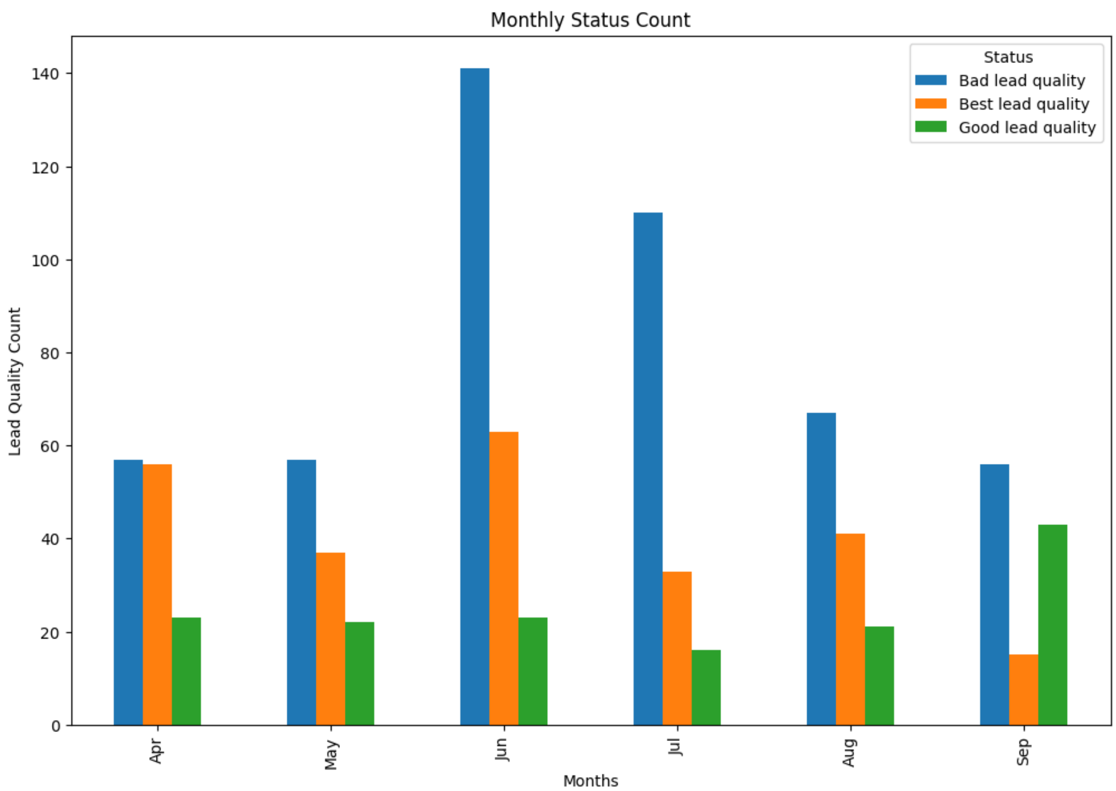
[146]:
```
lead_quality
Bad lead quality     488
Best lead quality    245
Good lead quality    148
```

```
Name: count, dtype: int64
```

[147]:
```
# Group by month and grouped_status
grouped = df_filtered.groupby(['year_month', 'lead_quality']).size().
↪unstack(fill_value=0)
grouped
```

[147]:
```
lead_quality  Bad lead quality  Best lead quality  Good lead quality
year_month
Apr                         57                 56                 23
Aug                         67                 41                 21
Jul                        110                 33                 16
Jun                        141                 63                 23
May                         57                 37                 22
Sep                         56                 15                 43
```

[148]:
```
#Sort the index by date
grouped = grouped.reindex(sorted(grouped.index, key=lambda x: pd.to_datetime(x,␣
↪format='%b')))

grouped.plot(kind='bar', stacked=False, figsize=(12, 8))

# Setting the title and labels
plt.title('Monthly Status Count')
plt.xlabel('Months')
plt.ylabel('Lead Quality Count')
plt.legend(title=' Status')

plt.show()
```



[149]:
```
def perform_regression(df, status):
    df_status = df[[status]].dropna().reset_index()
    df_status['date_ordinal'] = pd.to_datetime(df_status['year_month'],␣
 ↪format='%b').map(pd.Timestamp.toordinal)
    X = df_status['date_ordinal']
    y = df_status[status]
    X = sm.add_constant(X)  # Adds a constant term to the predictor
    model = sm.OLS(y, X).fit()
    return model

# Perform regression for each grouped status
for status in grouped.columns:
    model = perform_regression(grouped, status)
    print(f"Regression results for {status}:")
    print(model.summary())
    print("\n")
```

```
Regression results for Bad lead quality:
                            OLS Regression Results
==============================================================================
Dep. Variable:       Bad lead quality   R-squared:                       0.000
Model:                            OLS   Adj. R-squared:                 -0.250
```

```
Method:                  Least Squares   F-statistic:                 0.0004304
Date:                 Sun, 04 Aug 2024   Prob (F-statistic):              0.984
Time:                         18:03:50   Log-Likelihood:                -29.436
No. Observations:                    6   AIC:                             62.87
Df Residuals:                        4   BIC:                             62.46
Df Model:                            1
Covariance Type:             nonrobust
================================================================================
                   coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
const         4583.1210   2.17e+05      0.021      0.984   -5.98e+05    6.07e+05
date_ordinal    -0.0065      0.313     -0.021      0.984      -0.875       0.862
==============================================================================
Omnibus:                           nan   Durbin-Watson:                   1.557
Prob(Omnibus):                     nan   Jarque-Bera (JB):                0.993
Skew:                            0.891   Prob(JB):                        0.609
Kurtosis:                        2.107   Cond. No.                     9.21e+09
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 9.21e+09. This might indicate that there are
strong multicollinearity or other numerical problems.


Regression results for Best lead quality:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Best lead quality   R-squared:                       0.484
Model:                            OLS   Adj. R-squared:                  0.355
Method:                  Least Squares   F-statistic:                     3.749
Date:                 Sun, 04 Aug 2024   Prob (F-statistic):              0.125
Time:                         18:03:50   Log-Likelihood:                -23.023
No. Observations:                    6   AIC:                             50.05
Df Residuals:                        4   BIC:                             49.63
Df Model:                            1
Covariance Type:             nonrobust
================================================================================
                   coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
const         1.443e+05   7.45e+04      1.937      0.125   -6.26e+04    3.51e+05
date_ordinal    -0.2080      0.107     -1.936      0.125      -0.506       0.090
==============================================================================
Omnibus:                           nan   Durbin-Watson:                   3.136
Prob(Omnibus):                     nan   Jarque-Bera (JB):                0.553
Skew:                            0.506   Prob(JB):                        0.758
Kurtosis:                        1.910   Cond. No.                     9.21e+09
```

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

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 9.21e+09. This might indicate that there are
strong multicollinearity or other numerical problems.


Regression results for Good lead quality:
                            OLS Regression Results
==============================================================================
Dep. Variable:      Good lead quality   R-squared:                       0.269
Model:                            OLS   Adj. R-squared:                  0.086
Method:                 Least Squares   F-statistic:                     1.469
Date:                Sun, 04 Aug 2024   Prob (F-statistic):              0.292
Time:                        18:03:50   Log-Likelihood:                -20.442
No. Observations:                   6   AIC:                             44.88
Df Residuals:                       4   BIC:                             44.47
Df Model:                           1
Covariance Type:            nonrobust
================================================================================
                   coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
const        -5.871e+04   4.85e+04     -1.211      0.292   -1.93e+05    7.58e+04
date_ordinal     0.0847      0.070      1.212      0.292      -0.109       0.279
==============================================================================
Omnibus:                          nan   Durbin-Watson:                   1.523
Prob(Omnibus):                    nan   Jarque-Bera (JB):                0.306
Skew:                           0.154   Prob(JB):                        0.858
Kurtosis:                       1.937   Cond. No.                     9.21e+09
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 9.21e+09. This might indicate that there are
strong multicollinearity or other numerical problems.


/usr/local/lib/python3.10/dist-packages/statsmodels/stats/stattools.py:74:
ValueWarning: omni_normtest is not valid with less than 8 observations; 6
samples were given.
  warn("omni_normtest is not valid with less than 8 observations; %i "
/usr/local/lib/python3.10/dist-packages/statsmodels/stats/stattools.py:74:
ValueWarning: omni_normtest is not valid with less than 8 observations; 6
samples were given.
```

```
  warn("omni_normtest is not valid with less than 8 observations; %i "
/usr/local/lib/python3.10/dist-packages/statsmodels/stats/stattools.py:74:
ValueWarning: omni_normtest is not valid with less than 8 observations; 6
samples were given.
  warn("omni_normtest is not valid with less than 8 observations; %i "
```

[150]:
```python
# Perform the Chi-Square test
from scipy.stats import chi2_contingency

chi2, p, dof, expected = chi2_contingency(grouped)

# Display the results
print(f"Chi-Square statistic: {chi2}")
print(f"p-value: {p}")
print(f"Degrees of freedom: {dof}")
print("Expected frequencies:")
print(expected)

# Interpret the result
alpha = 0.05
if p < alpha:
    print("There is a significant difference in lead quality counts over time␣
 ↪(reject H0).")
else:
    print("There is no significant difference in lead quality counts over time␣
 ↪(fail to reject H0).")
```

```
Chi-Square statistic: 75.83785343992369
p-value: 3.2675999598781904e-12
Degrees of freedom: 10
Expected frequencies:
[[ 75.33257662  37.82065834  22.84676504]
 [ 64.25425653  32.25879682  19.48694665]
 [125.73893303  63.12712826  38.13393871]
 [ 88.07264472  44.21679909  26.71055619]
 [ 71.45516459  35.87400681  21.6708286 ]
 [ 63.14642452  31.70261067  19.15096481]]
There is a significant difference in lead quality counts over time (reject H0).
```

Analysis:

Bad lead quality consistently has a higher count compared to Best lead quality and Good lead quality.

If we have a month-wise view, June has the highest number of Bad Quality leads, followed by July. April and June have the highest count of Best Quality leads. There is no clear upward or downward trend in the lead quality categories over the months.

Since our p values is significant;y < 0.05 we reject the Null Hypothesis

Recommendations:

Improve Lead Quality: The consistently high number of Bad lead quality leads indicates a need for improving lead generation strategies. Focusing on targeting and qualification processes might help increase the proportion of Best and Good lead quality leads.

3 2. Whether WidgetName affects lead quality

[151]:
```
df_filtered
```

[151]:
```
     LeadCreated FirstName                         Email  \
2     2009-04-21      Gina      wagoner_gina@yahoo.com
3     2009-08-03      Kari           usa4ley@yahoo.com
7     2009-04-22      John        johndoe333@yahoo.com
10    2009-06-01      Juan      villalobosjgv@yahoo.com
17    2009-08-01     Kandi         kandielko@verizon.net
…            …         …                            …
3005  2009-09-29   Mariana         marianadit@gmail.com
3006  2009-05-09   Audelia  ada.bautista@sbcglobal.net
3009  2009-07-08     Kelly    kellybelleone@hotmail.com
3010  2009-06-19      Lucy             lucyaac@yahh.com
3012  2009-07-15   Michael    mrussell1127@hotmail.com

                 VendorLeadID                                   CallStatus  \
2     hFg80jf_ROCRN55hdhWILw  Unable to contact - Bad Contact Information
3     jB01QgYZxkWArI9jWxuufw                Contacted - Doesn't Qualify
7     hxFrkNSCjU6rE2u-7yH-KQ  Unable to contact - Bad Contact Information
10    LfatQ19SFkWfP3-hH7TVTQ  Unable to contact - Bad Contact Information
17    7YvjZQLOi0aAT7DhiqDISg  Unable to contact - Bad Contact Information
…                          …                                            …
3005  aFyR8rv4206QKiv_xj0E5g                Contacted - Doesn't Qualify
3006  cxRZiF10jUKXiaLHJPv2Ww                                       Closed
3009  vPAQIg1UhEabykCa5wpEYg                                 EP Confirmed
3010  Z-OkvmLqCE20wNaeS__oOg                Contacted - Doesn't Qualify
3012  Z26Md0QFrkuMgZRApiJf0w                Contacted - Doesn't Qualify

                                     WidgetName PublisherZoneName  \
2              w-300250-DebtReduction1-1DC-Head2    TopLeft-302252
3              w-302252-DebtReduction1-1DC-white    TopLeft-302252
7          w-300250-DebtReduction1-2DC-BlueMeter    TopLeft-302252
10    w-302252-DebtReduction1-1DC-CreditSolutions    TopLeft-302252
17    w-302252-DebtReduction1-1DC-CreditSolutions    TopLeft-302252
…                                             …                 …
3005  w-302252-DebtReduction1-1DC-CreditSolutions    TopLeft-302252
3006           w-302252-DebtReduction1-1DC-white    TopLeft-302252
3009  w-302252-DebtReduction1-1DC-CreditSolutions    TopLeft-302252
3010  w-302252-DebtReduction1-1DC-yellowarrow-blue    TopLeft-302252
```

```
3012  w-302252-DebtReduction1-1DC-yellowarrow-blue    TopLeft-302252

     PublisherCampaignName  AddressScore  PhoneScore  …   MarketingCampaign  \
2          DebtReductionInc           NaN         NaN  …  Financial Services
3          DebtReductionInc           5.0         3.0  …    DebtReductionInc
7          DebtReductionInc           NaN         NaN  …              Credit
10         DebtReductionInc           NaN         NaN  …    DebtReductionInc
17         DebtReductionInc           3.0         3.0  …    DebtReductionInc
…                         …             …           …  …                   …
3005       DebtReductionInc           5.0         1.0  …    DebtReductionInc
3006       DebtReductionInc           NaN         NaN  …    DebtReductionInc
3009       DebtReductionInc           NaN         5.0  …        Debt General
3010       DebtReductionInc           NaN         NaN  …    DebtReductionInc
3012       DebtReductionInc           5.0         3.0  …        Debt General

                           AdGroup                     Keyword  \
2                      Consolidate                         NaN
3                   Lower Payments                         NaN
7             Debt Credit Services             Credit services
10    Credit Card Debt - high volume                         NaN
17    Credit Card Debt - high volume                         NaN
…                                …                           …
3005             Debt Consolidation                         NaN
3006                   General Debt                         NaN
3009                Get Out Of Debt      Get out of credit debt
3010                   Student Debt                         NaN
3012                 Lower Payments  Lower monthly car payments

                                            SearchQuery  \
2                                                   NaN
3                                                   NaN
7                                       credit services
10                                                  NaN
17                                                  NaN
…                                                     …
3005                      goverment debts consolidation
3006                                  govt debt relief
3009  how to get out of credit card debt without paying
3010                                student loan default
3012                           lower monthly car payment

                                            ReferralURL  \
2          http://us.mc582.mail.yahoo.com/mc/showMessage
3     http://norwich.kijiji.com/c-Cars-vehicles-Cars…
7                             http://www.google.com/search
10         http://googleads.g.doubleclick.net/pagead/ads
17         http://googleads.g.doubleclick.net/pagead/ads
```

```
…                                                   …
3005                   http://search.yahoo.com/search
3006  http://search.yahoo.com/search;_ylt=Ar7g1XHyUV…
3009                     http://www.google.com/search
3010                   http://search.yahoo.com/search
3012                        http://www.google.com/url

                               ReferralURL Parameters  \
2     &fid=Inbox&sort=date&order=down&startMid=0&.ra…
3                                                 NaN
7     q=credit services&rls=com.microsoft:*&ie=UTF-8…
10    client=ca-pub-7277345023380563&host=pub-155622…
17    client=ca-pub-3089121361425291&dt=124917730077…
…                                                   …
3005  p=goverment debts consolidation&toggle=1&cop=m…
3006  p=govt debt relief&fr=att-portal-s&toggle=1&co…
3009  sourceid=navclient&aq=6&oq=how to get out of c…
3010           p=student loan default&ei=utf-8&fr=b1ie7
3012  q=lower monthly car payment&url=/aclk%3Fsa%3Dl…

                                   LandingPageURL  \
2      http://www.debtreductioninc.com/index8.html
3     http://www.debtreductioninc.com/index12.html
7      http://www.debtreductioninc.com/index8.html
10     http://www.debtreductioninc.com/index8.html
17     http://www.debtreductioninc.com/index8.html
…                                              …
3005   http://www.debtreductioninc.com/index8.html
3006   http://www.debtreductioninc.com/index8.html
3009   http://www.debtreductioninc.com/index8.html
3010   http://www.debtreductioninc.com/index8.html
3012   http://www.debtreductioninc.com/index8.html

                          Landing Page URL Parameters       lead_quality  \
2     utm_source=AdKnowledge&utm_medium=CPC&utm_cont…   Bad lead quality
3     utm_source=Google&utm_medium=cpc&utm_campaign=…   Bad lead quality
7     utm_source=google&utm_medium=CPC&utm_content=D…   Bad lead quality
10    utm_source=Google&utm_medium=cpc&utm_campaign=…   Bad lead quality
17    utm_source=Google&utm_medium=cpc&utm_campaign=…   Bad lead quality
…                                                   …                …
3005  utm_source=yahoo&utm_medium=cpc&utm_campaign=D…   Bad lead quality
3006  utm_source=yahoo&utm_medium=cpc&utm_campaign=D…  Best lead quality
3009  utm_source=google&utm_medium=CPC&utm_content=G…  Good lead quality
3010  utm_source=yahoo&utm_medium=cpc&utm_campaign=D…   Bad lead quality
3012  utm_source=google&utm_medium=CPC&utm_content=L…   Bad lead quality

     year_month
```

```
2          Apr
3          Aug
7          Apr
10         Jun
17         Aug
…        …
3005       Sep
3006       May
3009       Jul
3010       Jun
3012       Jul

[881 rows x 25 columns]
```

[152]:
```
df_filtered['WidgetName'].value_counts()
```

[152]:
```
WidgetName
w-302252-DebtReduction1-1DC-CreditSolutions     304
w-300250-DebtReduction1-1DC                     136
w-302252-DebtReduction1-1DC-white               113
w-302252-DebtReduction1-1DC-yellowarrow-blue     80
w-302252-DebtReduction1-1DC                      70
w-302252-DebtReduction1-1DC-yellowarrow-dark     40
w-300250-DebtReduction1-1DC-CreditSolutions      23
w-300250-DebtReduction1-1DC-Head2                22
w-300250-DebtReduction1-2DC-BlueMeter            21
w-302252-DebtReduction1-1DC-yellowarrow          21
w-300250-DebtReduction1-2DC-CreditSolutions      20
w-300250-DebtReduction1-1DC-BlueMeter            18
w-300250-DebtReduction1-1DC-Head3                12
w-300250-DebtReduction1-1DC-white                 1
Name: count, dtype: int64
```

[153]:
```
# Group by WidgetName and lead_quality
widget_quality= df_filtered.groupby(['WidgetName', 'lead_quality']).size().
 ↪unstack(fill_value=0)
widget_quality
```

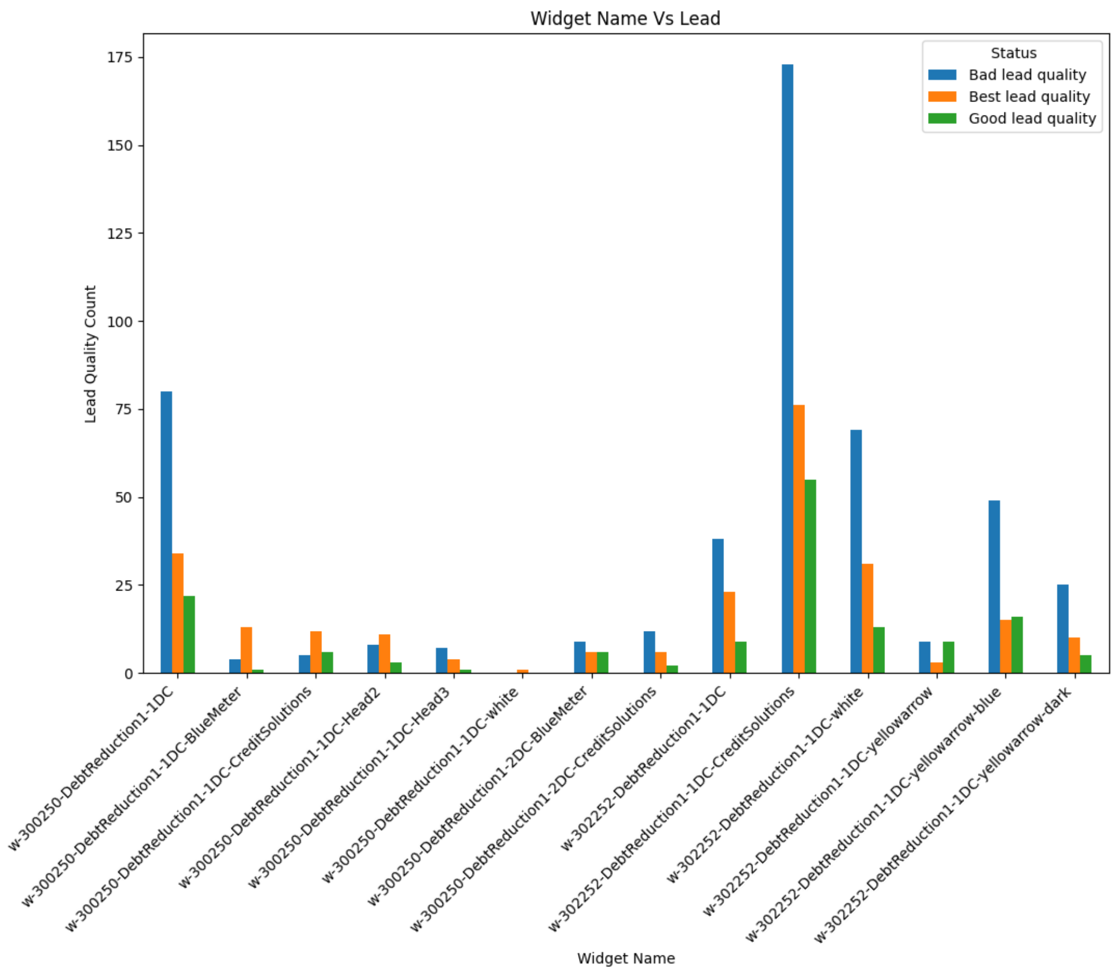
[153]:
```
lead_quality                                Bad lead quality  \
WidgetName
w-300250-DebtReduction1-1DC                               80
w-300250-DebtReduction1-1DC-BlueMeter                      4
w-300250-DebtReduction1-1DC-CreditSolutions                5
w-300250-DebtReduction1-1DC-Head2                          8
w-300250-DebtReduction1-1DC-Head3                          7
w-300250-DebtReduction1-1DC-white                          0
w-300250-DebtReduction1-2DC-BlueMeter                      9
```

```
w-300250-DebtReduction1-2DC-CreditSolutions                    12
w-302252-DebtReduction1-1DC                                    38
w-302252-DebtReduction1-1DC-CreditSolutions                   173
w-302252-DebtReduction1-1DC-white                              69
w-302252-DebtReduction1-1DC-yellowarrow                         9
w-302252-DebtReduction1-1DC-yellowarrow-blue                   49
w-302252-DebtReduction1-1DC-yellowarrow-dark                   25

lead_quality                                 Best lead quality  \
WidgetName
w-300250-DebtReduction1-1DC                                    34
w-300250-DebtReduction1-1DC-BlueMeter                          13
w-300250-DebtReduction1-1DC-CreditSolutions                    12
w-300250-DebtReduction1-1DC-Head2                              11
w-300250-DebtReduction1-1DC-Head3                               4
w-300250-DebtReduction1-1DC-white                               1
w-300250-DebtReduction1-2DC-BlueMeter                           6
w-300250-DebtReduction1-2DC-CreditSolutions                     6
w-302252-DebtReduction1-1DC                                    23
w-302252-DebtReduction1-1DC-CreditSolutions                    76
w-302252-DebtReduction1-1DC-white                              31
w-302252-DebtReduction1-1DC-yellowarrow                         3
w-302252-DebtReduction1-1DC-yellowarrow-blue                   15
w-302252-DebtReduction1-1DC-yellowarrow-dark                   10

lead_quality                                 Good lead quality
WidgetName
w-300250-DebtReduction1-1DC                                    22
w-300250-DebtReduction1-1DC-BlueMeter                           1
w-300250-DebtReduction1-1DC-CreditSolutions                     6
w-300250-DebtReduction1-1DC-Head2                               3
w-300250-DebtReduction1-1DC-Head3                               1
w-300250-DebtReduction1-1DC-white                               0
w-300250-DebtReduction1-2DC-BlueMeter                           6
w-300250-DebtReduction1-2DC-CreditSolutions                     2
w-302252-DebtReduction1-1DC                                     9
w-302252-DebtReduction1-1DC-CreditSolutions                    55
w-302252-DebtReduction1-1DC-white                              13
w-302252-DebtReduction1-1DC-yellowarrow                         9
w-302252-DebtReduction1-1DC-yellowarrow-blue                   16
w-302252-DebtReduction1-1DC-yellowarrow-dark                    5
```

[154]:

```
#Sort the index by date
widget_quality = widget_quality.reindex(sorted(widget_quality.index))

widget_quality.plot(kind='bar', stacked=False, figsize=(12, 8))
```

```
# Setting the title and labels
plt.title('Widget Name Vs Lead')
plt.xlabel('Widget Name')
plt.ylabel('Lead Quality Count')
plt.xticks(rotation=45, ha='right')

plt.legend(title=' Status')

plt.show()
```



[155]:

```
# Create a contingency table
contingency_table = pd.crosstab(df_filtered['WidgetName'],␣
↪df_filtered['lead_quality'])

# Display the contingency table to ensure it's correct
print("Contingency Table:")
print(contingency_table)
```

```python
# Perform the Chi-Square test
chi2, p, dof, expected = chi2_contingency(contingency_table)

# Display the results
print(f"Chi-Square statistic: {chi2}")
print(f"p-value: {p}")
print(f"Degrees of freedom: {dof}")
print("Expected frequencies:")
print(expected)

# Interpret the result
alpha = 0.05
if p < alpha:
    print("The variables are dependent (reject H0).")
else:
    print("The variables are independent (fail to reject H0).")
```

```
Contingency Table:
lead_quality                                  Bad lead quality  \
WidgetName
w-300250-DebtReduction1-1DC                                 80
w-300250-DebtReduction1-1DC-BlueMeter                        4
w-300250-DebtReduction1-1DC-CreditSolutions                  5
w-300250-DebtReduction1-1DC-Head2                            8
w-300250-DebtReduction1-1DC-Head3                            7
w-300250-DebtReduction1-1DC-white                            0
w-300250-DebtReduction1-2DC-BlueMeter                        9
w-300250-DebtReduction1-2DC-CreditSolutions                 12
w-302252-DebtReduction1-1DC                                 38
w-302252-DebtReduction1-1DC-CreditSolutions                173
w-302252-DebtReduction1-1DC-white                           69
w-302252-DebtReduction1-1DC-yellowarrow                      9
w-302252-DebtReduction1-1DC-yellowarrow-blue                49
w-302252-DebtReduction1-1DC-yellowarrow-dark                25

lead_quality                                  Best lead quality  \
WidgetName
w-300250-DebtReduction1-1DC                                  34
w-300250-DebtReduction1-1DC-BlueMeter                        13
w-300250-DebtReduction1-1DC-CreditSolutions                  12
w-300250-DebtReduction1-1DC-Head2                            11
w-300250-DebtReduction1-1DC-Head3                             4
w-300250-DebtReduction1-1DC-white                             1
w-300250-DebtReduction1-2DC-BlueMeter                         6
w-300250-DebtReduction1-2DC-CreditSolutions                   6
w-302252-DebtReduction1-1DC                                  23
```

```
w-302252-DebtReduction1-1DC-CreditSolutions                    76
w-302252-DebtReduction1-1DC-white                              31
w-302252-DebtReduction1-1DC-yellowarrow                         3
w-302252-DebtReduction1-1DC-yellowarrow-blue                   15
w-302252-DebtReduction1-1DC-yellowarrow-dark                   10

lead_quality                                    Good lead quality
WidgetName
w-300250-DebtReduction1-1DC                                    22
w-300250-DebtReduction1-1DC-BlueMeter                           1
w-300250-DebtReduction1-1DC-CreditSolutions                     6
w-300250-DebtReduction1-1DC-Head2                               3
w-300250-DebtReduction1-1DC-Head3                               1
w-300250-DebtReduction1-1DC-white                               0
w-300250-DebtReduction1-2DC-BlueMeter                           6
w-300250-DebtReduction1-2DC-CreditSolutions                     2
w-302252-DebtReduction1-1DC                                     9
w-302252-DebtReduction1-1DC-CreditSolutions                    55
w-302252-DebtReduction1-1DC-white                              13
w-302252-DebtReduction1-1DC-yellowarrow                         9
w-302252-DebtReduction1-1DC-yellowarrow-blue                   16
w-302252-DebtReduction1-1DC-yellowarrow-dark                    5
Chi-Square statistic: 60.784886647889806
p-value: 0.00013140709731051506
Degrees of freedom: 26
Expected frequencies:
[[7.53325766e+01 3.78206583e+01 2.28467650e+01]
 [9.97048808e+00 5.00567537e+00 3.02383655e+00]
 [1.27400681e+01 6.39614075e+00 3.86379115e+00]
 [1.21861521e+01 6.11804767e+00 3.69580023e+00]
 [6.64699205e+00 3.33711691e+00 2.01589103e+00]
 [5.53916005e-01 2.78093076e-01 1.67990919e-01]
 [1.16322361e+01 5.83995460e+00 3.52780931e+00]
 [1.10783201e+01 5.56186152e+00 3.35981839e+00]
 [3.87741203e+01 1.94665153e+01 1.17593644e+01]
 [1.68390465e+02 8.45402951e+01 5.10692395e+01]
 [6.25925085e+01 3.14245176e+01 1.89829739e+01]
 [1.16322361e+01 5.83995460e+00 3.52780931e+00]
 [4.43132804e+01 2.22474461e+01 1.34392736e+01]
 [2.21566402e+01 1.11237230e+01 6.71963678e+00]]
The variables are dependent (reject H0).
```

Analysis:

w-302252-DebtReduction1-1DC-CreditSolutions has the highest volume of leads across all categories with 173 bad leads, 76 best leads, and 55 good leads. This indicates that while this widget generates a high number of leads, a significant portion is of bad quality. Across most widgets, there is a higher number of bad quality leads compared to best and good quality leads. This trend is particularly pronounced in high volume widgets like w-302252-DebtReduction1-1DC-CreditSolutions and w-300250-DebtReduction1-1DC.

Recommendations:

Increase investment in widgets like w-300250-DebtReduction1-1DC-BlueMeter and w-302252-DebtReduction1-1DC-yellowarrow-blue, which show a better distribution of best and good quality leads. Investigate and address issues with widgets that generate a high number of bad leads, especially those with high volumes. Analyze the characteristics and strategies of widgets that generate a higher proportion of best quality leads and apply those learnings to improve underperforming widgets.

4 2.1 What can we learn about the drivers of "lead quality" from this dataset?

5 where the ad was shown?

[156]:
```
# Group by WidgetName and lead_quality
campaign_quality= df_filtered.groupby(['PublisherCampaignName','lead_quality']).
 ↪size().unstack(fill_value=0)
campaign_quality
```

[156]:
```
lead_quality              Bad lead quality  Best lead quality  \
PublisherCampaignName
DebtReductionCallCenter                 66                 26
DebtReductionInc                       422                219

lead_quality              Good lead quality
PublisherCampaignName
DebtReductionCallCenter                  18
DebtReductionInc                        130
```

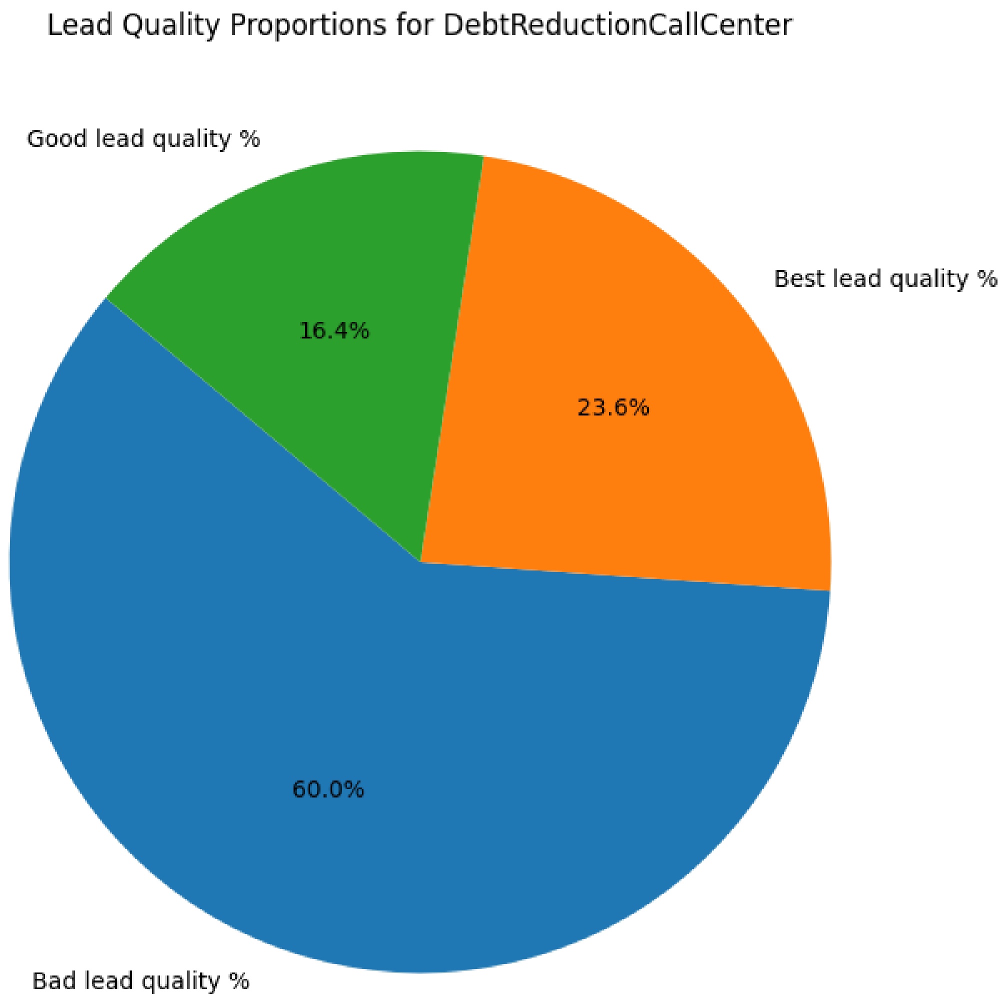
[157]:
```
# Calculate total leads for each campaign
campaign_quality['Total leads'] = campaign_quality.sum(axis=1)

# Calculate proportions
campaign_quality['Bad lead quality %'] = campaign_quality['Bad lead quality'] /␣
 ↪campaign_quality['Total leads'] * 100
campaign_quality['Best lead quality %'] = campaign_quality['Best lead quality']␣
 ↪/ campaign_quality['Total leads'] * 100
campaign_quality['Good lead quality %'] = campaign_quality['Good lead quality']␣
 ↪/ campaign_quality['Total leads'] * 100

campaign_quality
```

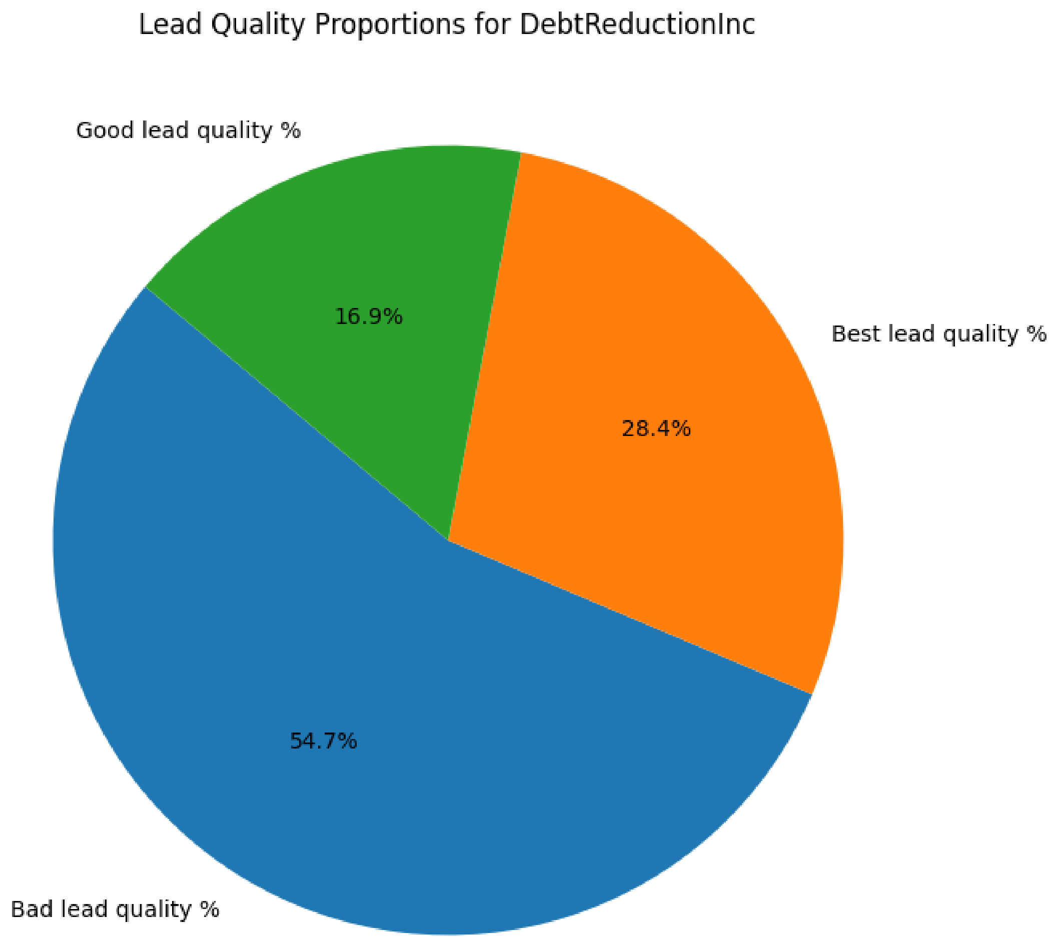
[157]:
```
lead_quality              Bad lead quality  Best lead quality  \
PublisherCampaignName
DebtReductionCallCenter                 66                 26
```

```
DebtReductionInc                          422              219

lead_quality             Good lead quality  Total leads  Bad lead quality %  \
PublisherCampaignName
DebtReductionCallCenter                 18          110           60.000000
DebtReductionInc                       130          771           54.734112

lead_quality             Best lead quality %  Good lead quality %
PublisherCampaignName
DebtReductionCallCenter            23.636364            16.363636
DebtReductionInc                   28.404669            16.861219
```

[158]:

```python
proportions = campaign_quality[['Bad lead quality %', 'Best lead quality %',␣
↪'Good lead quality %']]
def plot_pie_chart(data, campaign_name):
    # Plot the pie chart
    plt.figure(figsize=(8, 8))
    plt.pie(data, labels=data.index, autopct='%1.1f%%', startangle=140)
    plt.title(f'Lead Quality Proportions for {campaign_name}')
    plt.show()

# Plot for each campaign
for campaign in proportions.index:
    plot_pie_chart(proportions.loc[campaign], campaign)
```


Lead Quality Proportions for DebtReductionCallCenter
Good lead quality %
Best lead quality %
Bad lead quality %
16.4%
23.6%
60.0%



Analysis:

DebtReductionInc generates a significantly higher volume of leads compared to DebtReductionCallCenter. This could be due to the ease of filling out forms online versus calling an 800#.

The proportion of bad leads is relatively high for both sources, but slightly higher for the DebtReductionCallCenter. This suggests that the quality of leads from phone calls might be slightly lower than those from online forms, indicating that online forms might attract more qualified leads.

Recommendations:

Provide additional training for call center staff to help them better identify and nurture potential high-quality leads. This can include training on effective communication techniques and understanding customer needs.

Simplify and optimize the online forms to ensure they capture the necessary information without being too cumbersome for potential leads. This can help in reducing drop-offs and improving lead quality.

6 What can we learn about the drivers of "lead quality" from this dataset? What segments - where the ad was shown

[159]:
```python
df_filtered['Partner'].unique()
```

[159]:
```
array(['AdKnowledge', 'Google', 'google', 'yahoo', 'Advertise.com',
       'Call_Center'], dtype=object)
```

[160]:
```python
# combining google and Google
df_filtered['Partner']=df_filtered['Partner'].str.lower()
#df_filtered['Partner']
```

```
<ipython-input-160-716f17716d0a>:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df_filtered['Partner']=df_filtered['Partner'].str.lower()
```

[161]:
```python
# Group by partner and lead_quality
partner_quality= df_filtered.groupby(['Partner','lead_quality']).size().
 ↪unstack(fill_value=0)
partner_quality
```

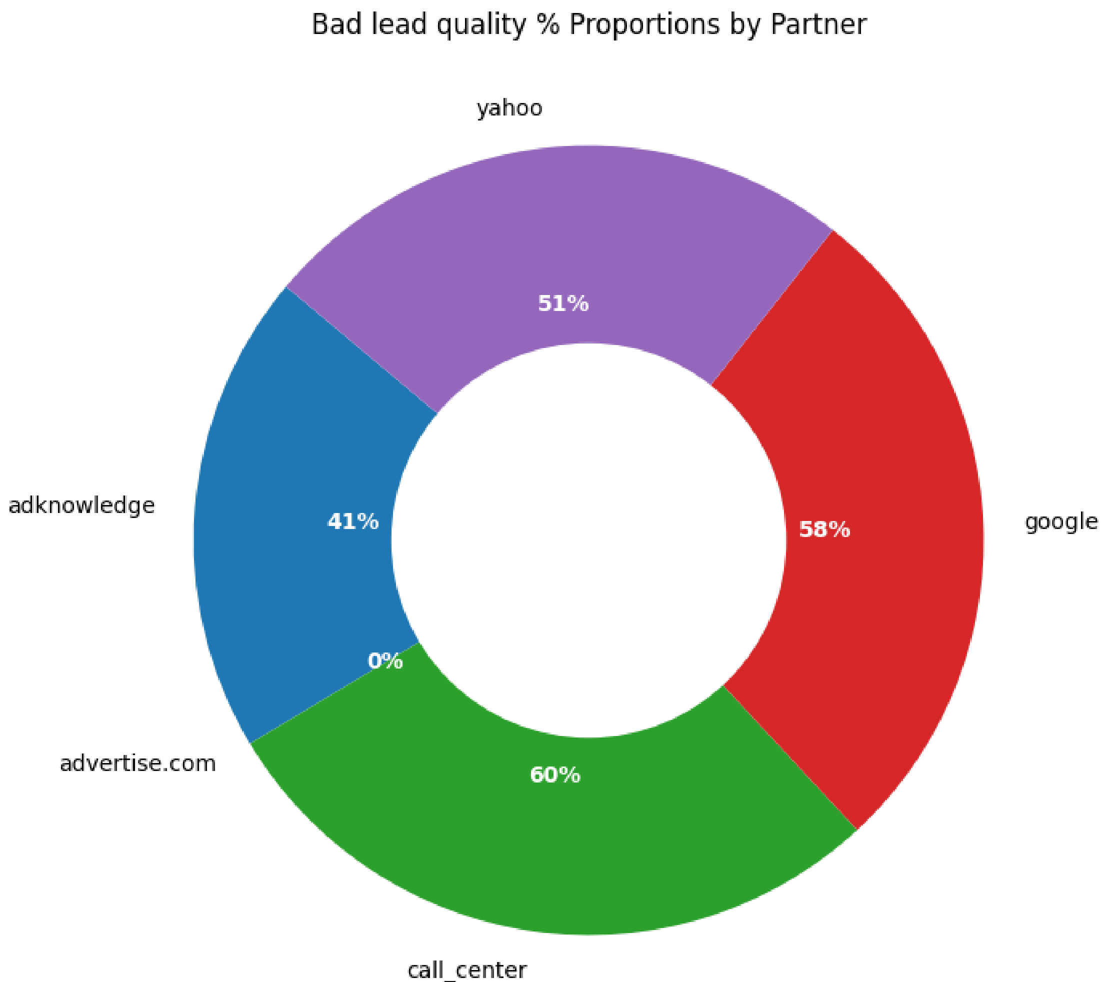
[161]:

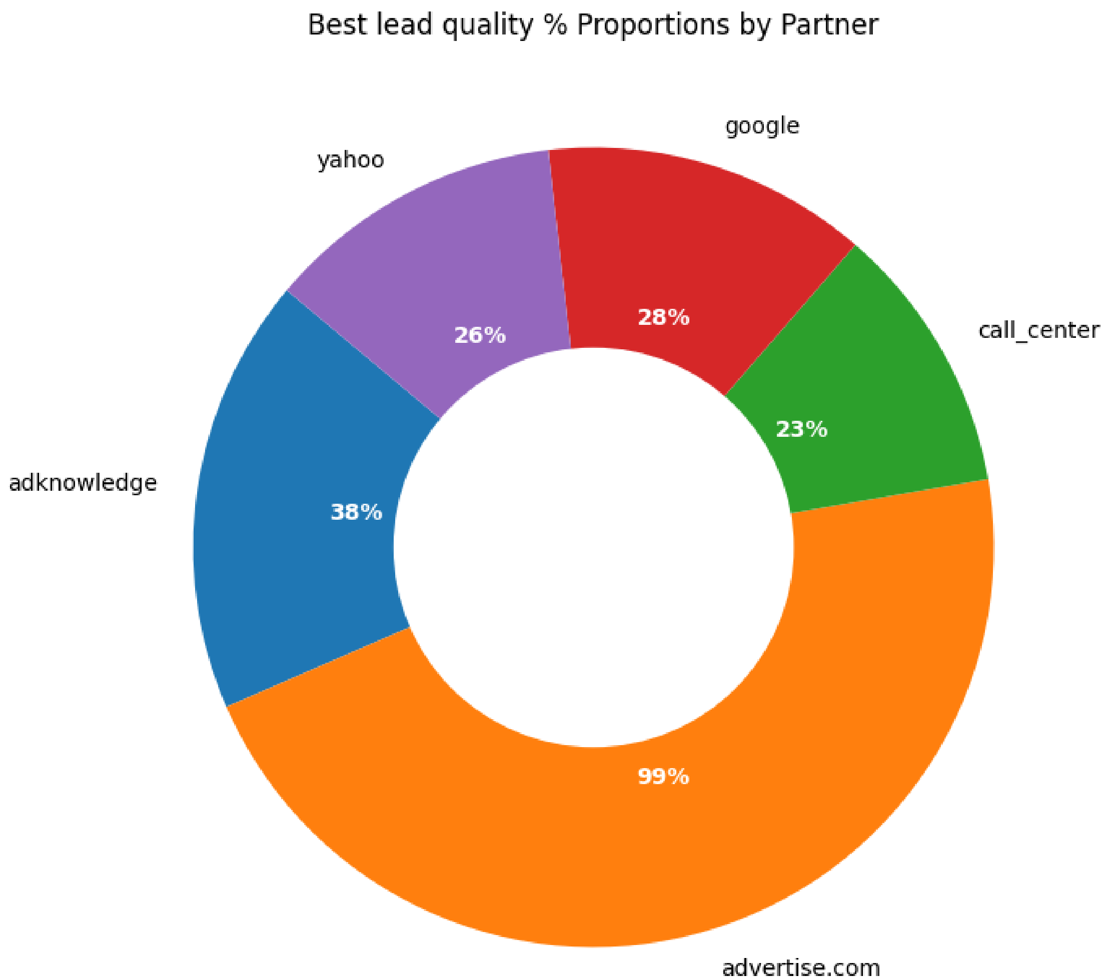
lead_quality Partner	Bad lead quality	Best lead quality	Good lead quality
adknowledge	23	21	11
advertise.com	0	1	0
call_center	66	26	18
google	252	123	57
yahoo	147	74	62

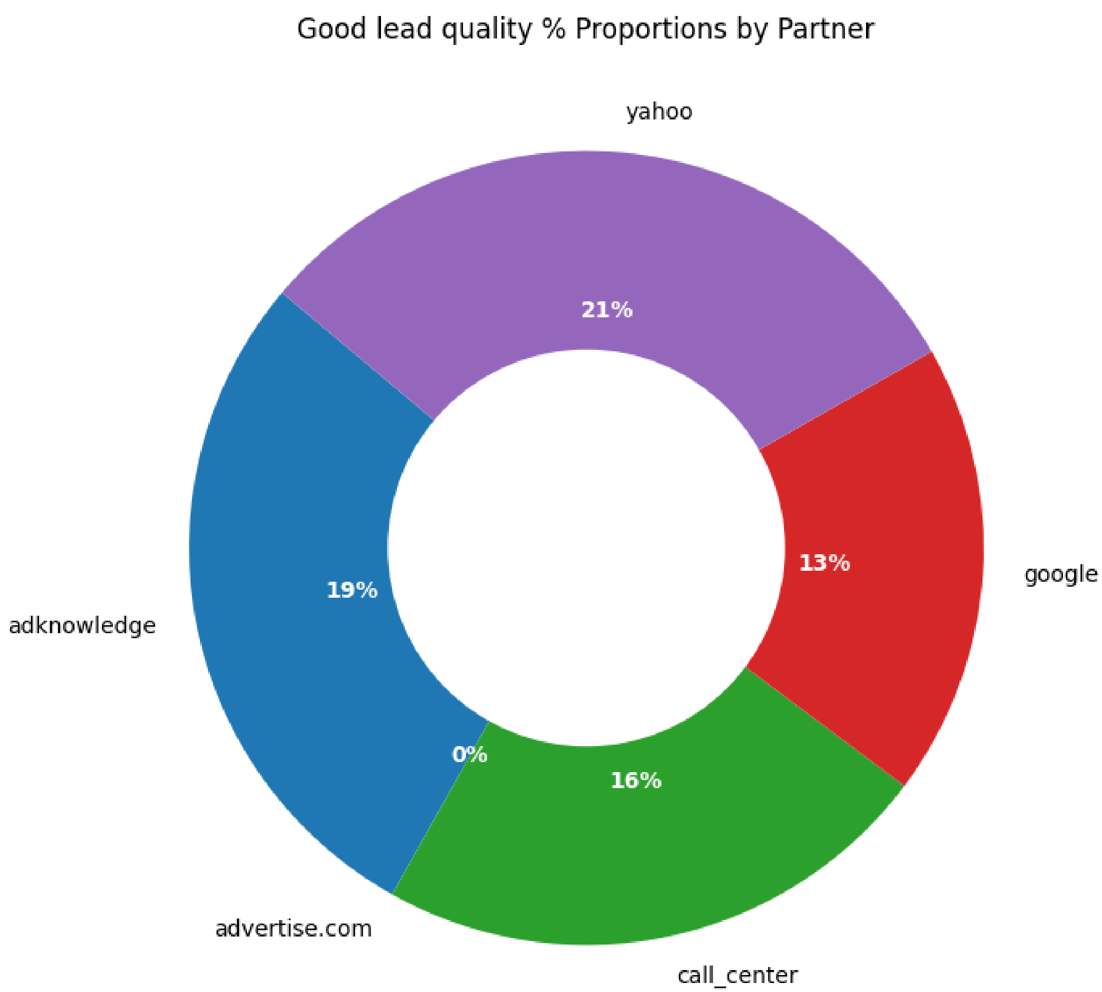
[162]:
```python
# Calculate total leads for each campaign
partner_quality['Total leads'] = partner_quality.sum(axis='columns')

# Calculate proportions
partner_quality['Bad lead quality %'] = partner_quality['Bad lead quality'] /␣
 ↪partner_quality['Total leads'] * 100
partner_quality['Best lead quality %'] = partner_quality['Best lead quality'] /␣
 ↪partner_quality['Total leads'] * 100
partner_quality['Good lead quality %'] = partner_quality['Good lead quality'] /␣
 ↪partner_quality['Total leads'] * 100

partner_quality
```

```
[162]: lead_quality   Bad lead quality  Best lead quality  Good lead quality  \
       Partner
       adknowledge                   23                 21                 11
       advertise.com                  0                  1                  0
       call_center                   66                 26                 18
       google                       252                123                 57
       yahoo                        147                 74                 62

       lead_quality   Total leads  Bad lead quality %  Best lead quality %  \
       Partner
       adknowledge             55           41.818182            38.181818
       advertise.com            1            0.000000           100.000000
       call_center            110           60.000000            23.636364
       google                 432           58.333333            28.472222
       yahoo                  283           51.943463            26.148410

       lead_quality   Good lead quality %
       Partner
       adknowledge              20.000000
       advertise.com             0.000000
       call_center              16.363636
       google                   13.194444
       yahoo                    21.908127
```

```
[190]: def plot_donut_chart(data, lead_quality_type):
           plt.figure(figsize=(8, 8))
           wedges, texts, autotexts = plt.pie(data, labels=data.index, autopct=lambda␣
        ↪p: f'{int(p*sum(data)/100)}%', startangle=140)
           plt.setp(autotexts, size=10, weight="bold", color="white")
           plt.title(f'{lead_quality_type} Proportions by Partner')
           plt.gca().add_artist(plt.Circle((0,0),0.50,fc='white')) # Creating the␣
        ↪donut hole
           plt.show()

       # Plot donut charts for each lead quality category
       for quality in ['Bad lead quality %', 'Best lead quality %', 'Good lead quality␣
        ↪%']:
           plot_donut_chart(partner_quality[quality], quality)
```


Bad lead quality % Proportions by Partner
yahoo
51%
adknowledge
41%
google
58%
0%
advertise.com
60%
call_center


Best lead quality % Proportions by Partner
google
yahoo
28%
26%
call_center
23%
adknowledge
38%
99%
advertise.com



[167]:
```
# Create a contingency table
contingency_table = pd.crosstab(df_filtered['Partner'],␣
↪df_filtered['lead_quality'])

# Perform the Chi-square test
chi2, p, dof, expected = chi2_contingency(contingency_table)

print("Chi-square Statistic:", chi2)
print("P-value:", p)
print("Degrees of Freedom:", dof)
print("Expected Frequencies:")
print(expected)
```

```
Chi-square Statistic: 17.381454040141595
P-value: 0.026373251124411908
```

```
Degrees of Freedom: 8
Expected Frequencies:
[[3.04653802e+01 1.52951192e+01 9.23950057e+00]
 [5.53916005e-01 2.78093076e-01 1.67990919e-01]
 [6.09307605e+01 3.05902384e+01 1.84790011e+01]
 [2.39291714e+02 1.20136209e+02 7.25720772e+01]
 [1.56758229e+02 7.87003405e+01 4.75414302e+01]]
```

Analysis:

Bad quality lead is highest in call center followed by adknowdledge with a Proportion of 41.8%.

Advertise.com has a proportion of 100% for best lead quality. This is due to the single lead being of the best quality, which may not be statistically significant but indicates potential. Google shows a best lead quality proportion of 28.47%, which is lower compared to Advertise.com but higher than some other partners.

Yahoo's contribution to good lead quality is 21.9%. This is the highest among the partners, indicating that Yahoo generates a balanced mix of good quality leads.

Recommendations:

Since Call Center and Google sources have the highest proportions of bad lead quality, it's crucial to review and improve the lead generation processes for these partners.

Yahoo shows a balanced lead quality distribution with the highest proportion of good lead quality leads. Focus on strategies that could convert these good quality leads into the best quality leads, optimizing the lead quality further.

7 2.2 What can we learn about the drivers of "lead quality" from this dataset? what kind of person filled out the ad

[168]:
```
df_filtered['DebtLevel'].unique()
```

[168]:
```
array(['20001-30000', '7500-15000', '7500-10000', 'More_than_100000',
       '70001-90000', '10001-15000', '30001-50000', '50001-70000',
       '15001-20000', '90000-100000'], dtype=object)
```

[169]:
```
def debt_level_to_int(debt_level):
    if debt_level == 'More_than_100000':
        return 100001
    else:
        lower, upper = map(int, debt_level.split('-'))
        return (lower + upper) // 2

# Apply the conversion to create a new column with integer debt levels
df_filtered['DebtLevel_int'] = df_filtered['DebtLevel'].apply(debt_level_to_int)

# Display the DataFrame with the new integer column
df_filtered.head()
```

```
<ipython-input-169-3eb14a9c8d8e>:9: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df_filtered['DebtLevel_int'] =
df_filtered['DebtLevel'].apply(debt_level_to_int)
```

[169]:

```
   LeadCreated FirstName                      Email            VendorLeadID  \
2   2009-04-21      Gina   wagoner_gina@yahoo.com  hFg80jf_ROCRN55hdhWILw
3   2009-08-03      Kari        usa4ley@yahoo.com  jB01QgYZxkWArI9jWxuufw
7   2009-04-22      John     johndoe333@yahoo.com  hxFrkNSCjU6rE2u-7yH-KQ
10  2009-06-01      Juan  villalobosjgv@yahoo.com  LfatQ19SFkWfP3-hH7TVTQ
17  2009-08-01     Kandi    kandielko@verizon.net  7YvjZQLOiOaAT7DhiqDISg

                                  CallStatus  \
2   Unable to contact - Bad Contact Information
3                  Contacted - Doesn't Qualify
7   Unable to contact - Bad Contact Information
10  Unable to contact - Bad Contact Information
17  Unable to contact - Bad Contact Information

                                  WidgetName PublisherZoneName  \
2           w-300250-DebtReduction1-1DC-Head2    TopLeft-302252
3           w-302252-DebtReduction1-1DC-white    TopLeft-302252
7       w-300250-DebtReduction1-2DC-BlueMeter    TopLeft-302252
10  w-302252-DebtReduction1-1DC-CreditSolutions    TopLeft-302252
17  w-302252-DebtReduction1-1DC-CreditSolutions    TopLeft-302252

   PublisherCampaignName  AddressScore  PhoneScore  …  \
2       DebtReductionInc           NaN         NaN  …
3       DebtReductionInc           5.0         3.0  …
7       DebtReductionInc           NaN         NaN  …
10      DebtReductionInc           NaN         NaN  …
17      DebtReductionInc           3.0         3.0  …

                          AdGroup          Keyword      SearchQuery  \
2                     Consolidate              NaN              NaN
3                  Lower Payments              NaN              NaN
7             Debt Credit Services  Credit services  credit services
10  Credit Card Debt - high volume              NaN              NaN
17  Credit Card Debt - high volume              NaN              NaN

                                   ReferralURL  \
2       http://us.mc582.mail.yahoo.com/mc/showMessage
3   http://norwich.kijiji.com/c-Cars-vehicles-Cars…
```

```
7                        http://www.google.com/search
10      http://googleads.g.doubleclick.net/pagead/ads
17      http://googleads.g.doubleclick.net/pagead/ads

                                  ReferralURL Parameters  \
2   &fid=Inbox&sort=date&order=down&startMid=0&.ra…
3                                                  NaN
7   q=credit services&rls=com.microsoft:*&ie=UTF-8…
10  client=ca-pub-7277345023380563&host=pub-155622…
17  client=ca-pub-3089121361425291&dt=124917730077…

                                  LandingPageURL  \
2    http://www.debtreductioninc.com/index8.html
3   http://www.debtreductioninc.com/index12.html
7    http://www.debtreductioninc.com/index8.html
10   http://www.debtreductioninc.com/index8.html
17   http://www.debtreductioninc.com/index8.html

                         Landing Page URL Parameters      lead_quality  \
2   utm_source=AdKnowledge&utm_medium=CPC&utm_cont…  Bad lead quality
3   utm_source=Google&utm_medium=cpc&utm_campaign=…  Bad lead quality
7   utm_source=google&utm_medium=CPC&utm_content=D…  Bad lead quality
10  utm_source=Google&utm_medium=cpc&utm_campaign=…  Bad lead quality
17  utm_source=Google&utm_medium=cpc&utm_campaign=…  Bad lead quality

   year_month DebtLevel_int
2         Apr         25000
3         Aug         25000
7         Apr         11250
10        Jun         25000
17        Aug          8750

[5 rows x 26 columns]
```

[170]:
```python
# Group by partner and lead_quality
debt_quality= df_filtered.groupby(['DebtLevel_int','lead_quality']).size().
 ↪unstack(fill_value=0)
debt_quality
```

[170]:

lead_quality	Bad lead quality	Best lead quality	Good lead quality
DebtLevel_int			
8750	109	22	9
11250	26	19	12
12500	49	34	20
17500	49	35	22
25000	70	40	32
40000	84	39	22

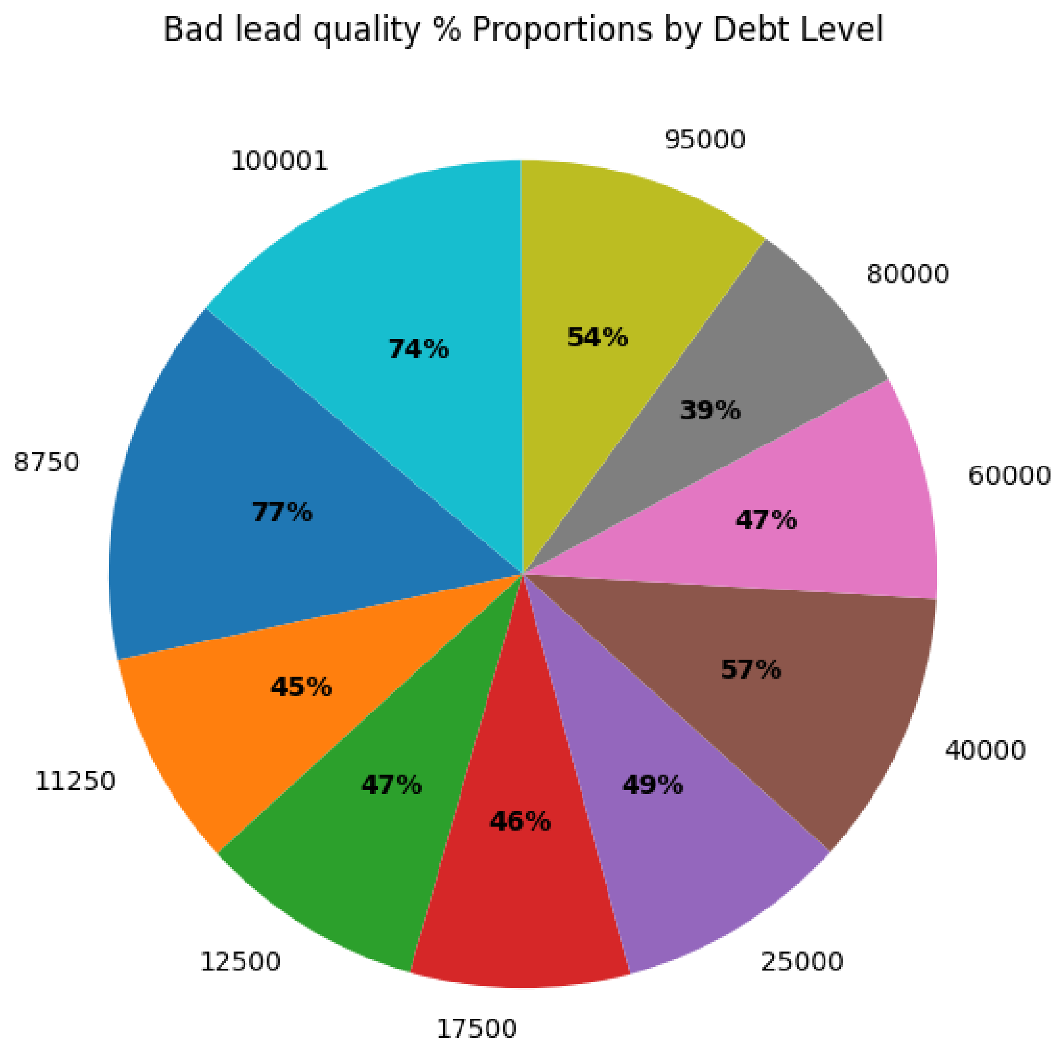
```
60000                          34                  22                  16
80000                          16                  18                   7
95000                          13                   9                   2
100001                         38                   7                   6
```

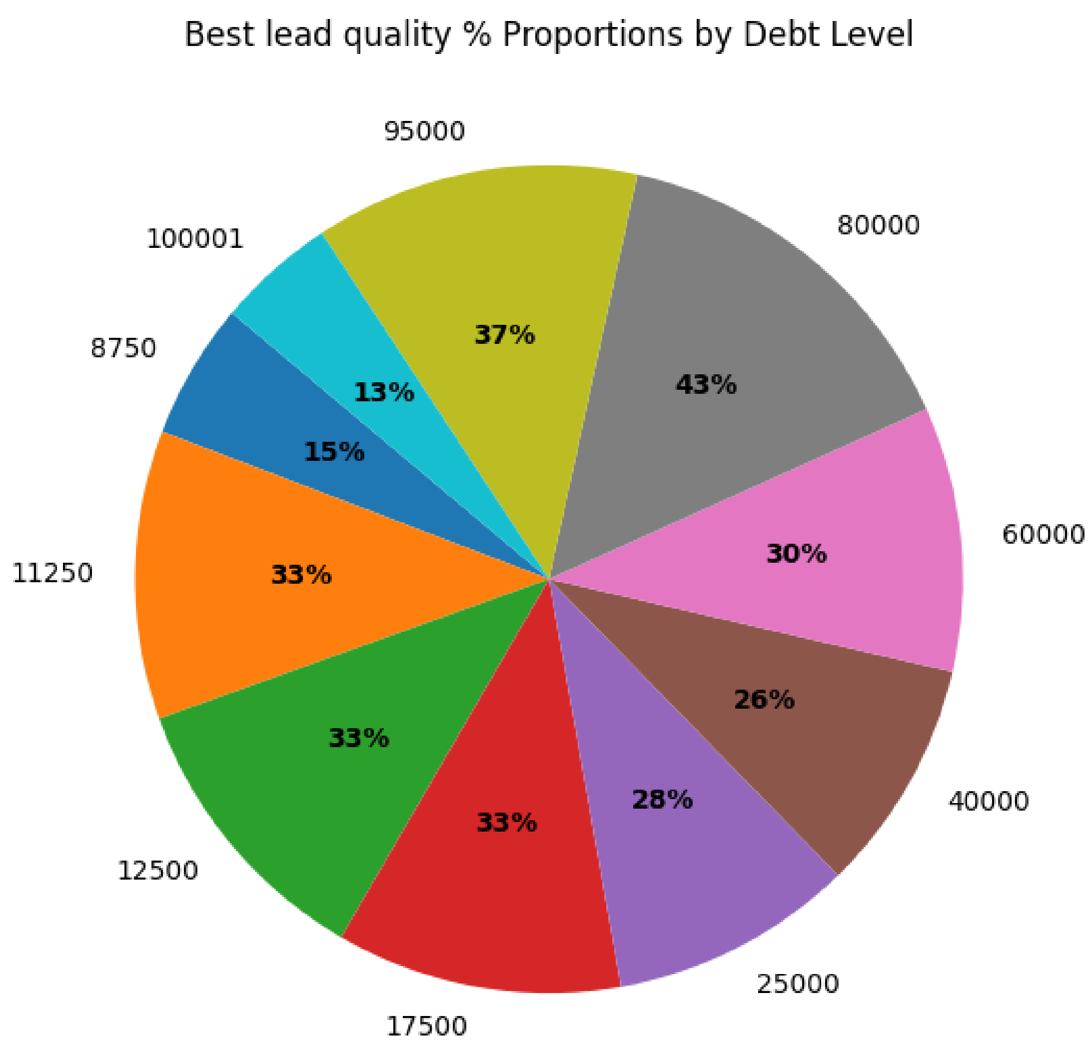
[171]:
```python
# Calculate total leads for each debt level
debt_quality['Total leads'] = debt_quality.sum(axis=1)

# Calculate proportions
debt_quality['Bad lead quality %'] = debt_quality['Bad lead quality'] /␣
 ↪debt_quality['Total leads'] * 100
debt_quality['Best lead quality %'] = debt_quality['Best lead quality'] /␣
 ↪debt_quality['Total leads'] * 100
debt_quality['Good lead quality %'] = debt_quality['Good lead quality'] /␣
 ↪debt_quality['Total leads'] * 100
debt_quality
```

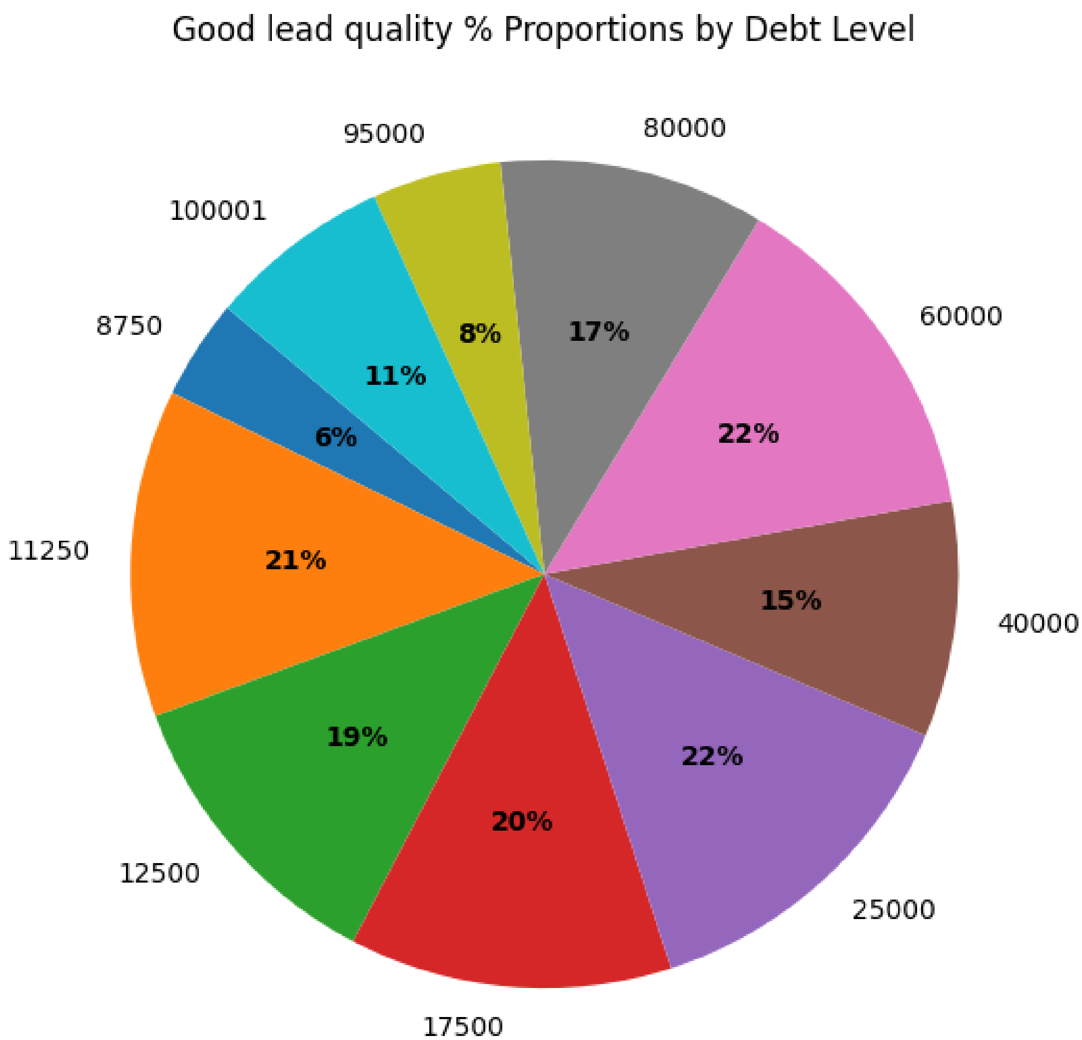
[171]:
```
lead_quality   Bad lead quality  Best lead quality  Good lead quality  \
DebtLevel_int
8750                        109                 22                  9
11250                        26                 19                 12
12500                        49                 34                 20
17500                        49                 35                 22
25000                        70                 40                 32
40000                        84                 39                 22
60000                        34                 22                 16
80000                        16                 18                  7
95000                        13                  9                  2
100001                       38                  7                  6

lead_quality   Total leads  Bad lead quality %  Best lead quality %  \
DebtLevel_int
8750                   140           77.857143            15.714286
11250                   57           45.614035            33.333333
12500                  103           47.572816            33.009709
17500                  106           46.226415            33.018868
25000                  142           49.295775            28.169014
40000                  145           57.931034            26.896552
60000                   72           47.222222            30.555556
80000                   41           39.024390            43.902439
95000                   24           54.166667            37.500000
100001                  51           74.509804            13.725490

lead_quality   Good lead quality %
DebtLevel_int
8750                      6.428571
11250                    21.052632
```

```
12500                         19.417476
17500                         20.754717
25000                         22.535211
40000                         15.172414
60000                         22.222222
80000                         17.073171
95000                          8.333333
100001                        11.764706
```

[191]:
```
# Function to plot pie chart for each lead quality
def plot_pie_chart(data, lead_quality_type):
    plt.figure(figsize=(10, 7))
    wedges, texts, autotexts = plt.pie(data, labels=data.index, autopct=lambda␣
 ↪p: f'{int(p*sum(data)/100)}%', startangle=140)
    plt.setp(autotexts, size=10, weight="bold", color="black")
    plt.title(f'{lead_quality_type} Proportions by Debt Level')
    plt.show()

# Plot pie charts for each lead quality category
for quality in ['Bad lead quality %', 'Best lead quality %', 'Good lead quality␣
 ↪%']:
    plot_pie_chart(debt_quality[quality], quality)
```


Bad lead quality % Proportions by Debt Level
100001
95000
80000
60000
40000
25000
17500
12500
11250
8750
74%
54%
39%
47%
57%
49%
46%
47%
45%
77%


Best lead quality % Proportions by Debt Level
95000
37%
80000
43%
100001
13%
8750
15%
60000
30%
11250
33%
12500
33%
17500
33%
25000
28%
40000
26%



Analysis:

Lower debt levels (e.g, 8,750) tend to generate a higher proportion of bad quality leads and lower proportions of best quality leads.

Higher debt levels (e.g, 40,000, 60,000) generally produce a more balanced mix of good and best quality leads.

Recommendations:

Focus on improving lead quality strategies or re-evaluating lead generation methods for lower debt levels to reduce the proportion of bad quality leads.

Allocate more resources and focus on debt levels that produce high proportions of best quality leads (e.g., 25,000, 40,000) to maximize overall lead quality and conversion rates.

[173]:
```
df_filtered['State'].unique()
```

[173]:
```
array(['NY', 'WA', 'IA', 'TX', 'IL', 'MD', 'AZ', 'VA', 'CA', 'MI', 'OR',
       'DC', 'FL', 'MA', 'IN', 'AL', 'WV', 'CO', 'PA', 'NM', 'OK', 'SD',
       'NE', 'AR', 'NV', 'CT', 'ND', 'HI', 'MT', 'MO', 'LA', 'AK'],
      dtype=object)
```

[174]:
```
# Define East Coast and West Coast states
east_coast_states = ['NY', 'MD', 'VA', 'DC', 'FL', 'MA', 'PA', 'CT']
west_coast_states = ['WA', 'CA', 'OR']

# Function to categorize states
def categorize_state(state):
    if state in east_coast_states:
        return 'East Coast'
    elif state in west_coast_states:
        return 'West Coast'
    else:
        return 'Other'

# Apply the categorization function to create a new column
df_filtered['coast'] = df_filtered['State'].apply(categorize_state)

# Display the DataFrame with the new column
df_filtered['coast'].head()
```

```
<ipython-input-174-4369236a9803>:15: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df_filtered['coast'] = df_filtered['State'].apply(categorize_state)
```

[174]:
```
2     East Coast
3     West Coast
7          Other
10         Other
17         Other
Name: coast, dtype: object
```

[175]:
```
# Group by partner and lead_quality
coast_quality= df_filtered.groupby(['coast','lead_quality']).size().
 ↪unstack(fill_value=0)
coast_quality
```

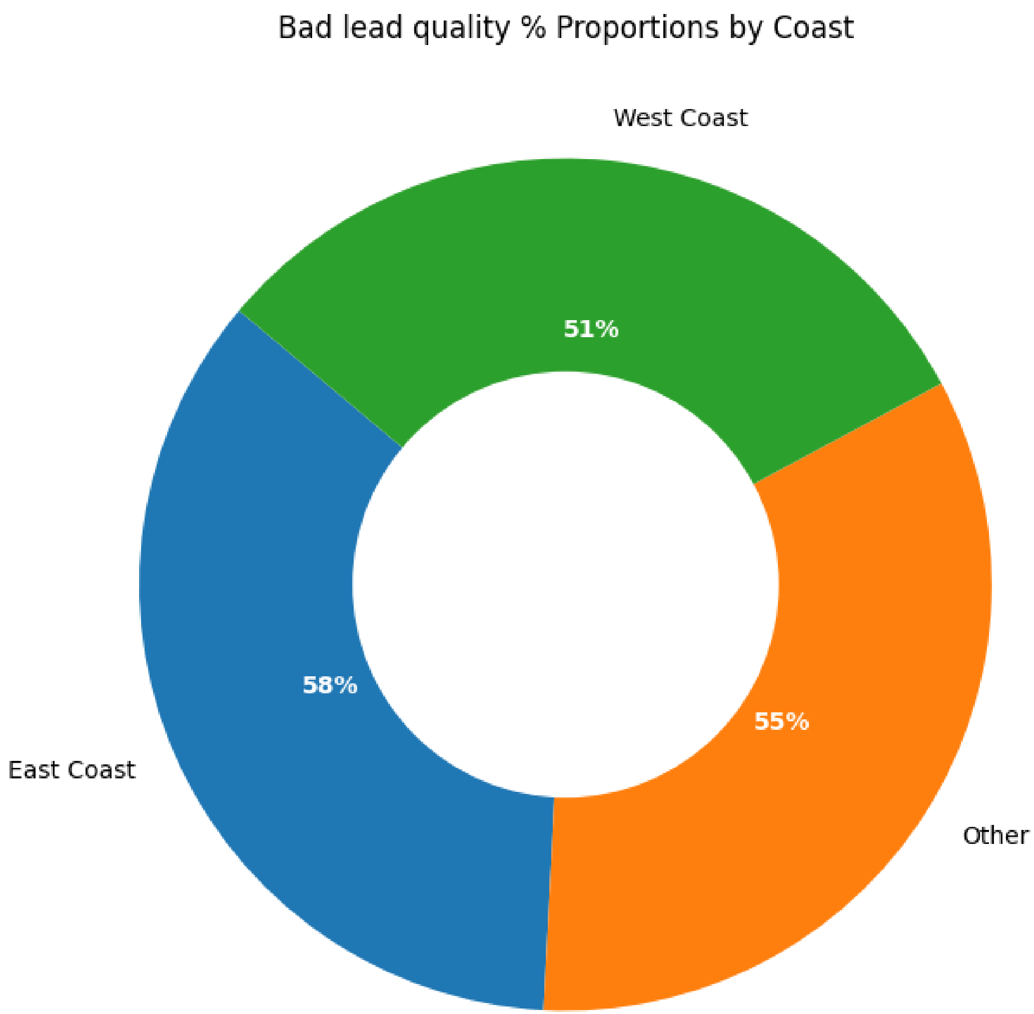
[175]:
```
lead_quality  Bad lead quality  Best lead quality  Good lead quality
coast
East Coast                 158                 66                 48
Other                      241                130                 64
West Coast                  89                 49                 36
```

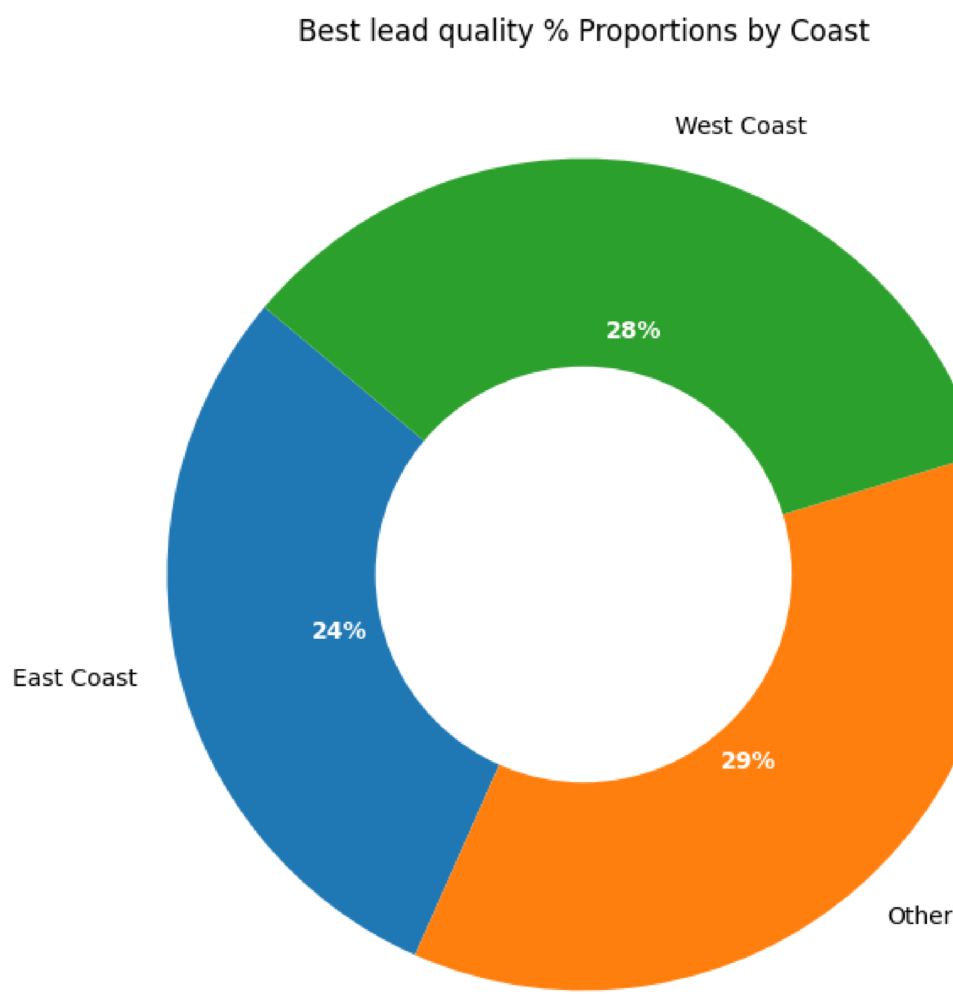
[176]:
```
# Calculate total leads for each debt level
coast_quality['Total leads'] = coast_quality.sum(axis=1)

# Calculate proportions
coast_quality['Bad lead quality %'] = coast_quality['Bad lead quality'] /␣
↪coast_quality['Total leads'] * 100
coast_quality['Best lead quality %'] = coast_quality['Best lead quality'] /␣
↪coast_quality['Total leads'] * 100
coast_quality['Good lead quality %'] = coast_quality['Good lead quality'] /␣
↪coast_quality['Total leads'] * 100
coast_quality
```

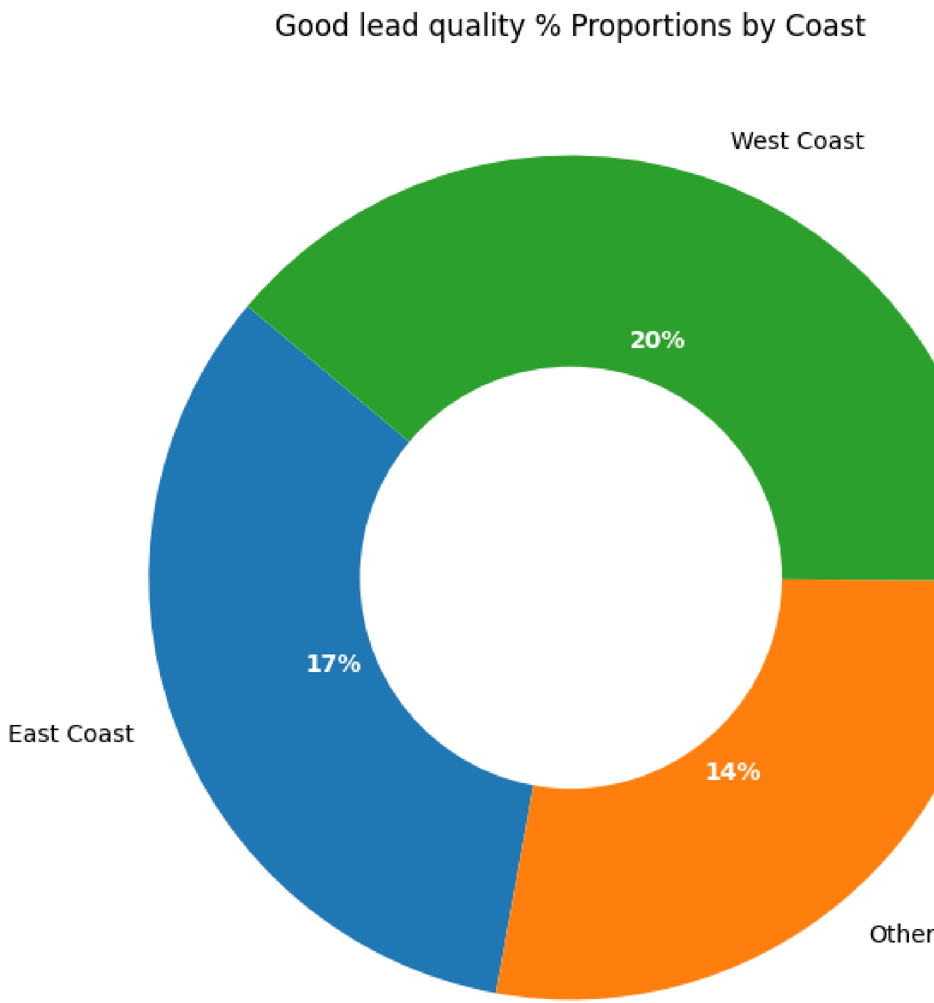
[176]:
```
lead_quality  Bad lead quality  Best lead quality  Good lead quality  \
coast
East Coast                 158                 66                 48
Other                      241                130                 64
West Coast                  89                 49                 36

lead_quality  Total leads  Bad lead quality %  Best lead quality %  \
coast
East Coast            272           58.088235            24.264706
Other                 435           55.402299            29.885057
West Coast            174           51.149425            28.160920

lead_quality  Good lead quality %
coast
East Coast              17.647059
Other                   14.712644
West Coast              20.689655
```

[192]:
```
def plot_donut_chart(data, lead_quality_type):
    plt.figure(figsize=(8, 8))
    wedges, texts, autotexts = plt.pie(data, labels=data.index, autopct=lambda␣
↪p: f'{int(p*sum(data)/100)}%', startangle=140)
    plt.setp(autotexts, size=10, weight="bold", color="white")
    plt.title(f'{lead_quality_type} Proportions by Coast')
    plt.gca().add_artist(plt.Circle((0,0),0.50,fc='white')) # Creating the␣
↪donut hole
    plt.show()

# Plot donut charts for each lead quality category
```

```
for quality in ['Bad lead quality %', 'Best lead quality %', 'Good lead quality␣
 ↪%']:
    plot_donut_chart(coast_quality[quality], quality)
```




Best lead quality % Proportions by Coast
West Coast
28%
East Coast
24%
29%
Other



Analysis:

East Coast has the higher % of bad quality lead and it indicates that leads from the East Coast are more likely to be of lower quality compared to other regions.

Other states comparatively have good amount of Best quality lead than east coast and west Coast.

West Coast exhibits the highest proportion of good lead quality at 20.7%. This implies that leads from the West Coast are more balanced, with a substantial portion being of good quality.

Recommendations:

Since the East Coast has the highest proportion of bad quality leads, it's crucial to review and enhance lead generation strategies specific to this region. Implement stricter filtering criteria and targeted marketing efforts to improve lead quality from this region.

With a higher proportion of best quality leads, the strategies used in "Other" regions can provide valuable insights. Analyze and replicate successful lead generation tactics from these regions to other regions, particularly the East Coast.

8 2.3 What can we learn about the drivers of "lead quality" from this dataset?what kind of ad did they see

[178]:
```
df_filtered['AdvertiserCampaignName'].unique()
```

[178]:
```
array(['Debt Settlement1 Master', 'creditsolutions-branded-shortform'],
      dtype=object)
```

[179]:
```
# Group by partner and lead_quality
ad_quality= df_filtered.groupby(['AdvertiserCampaignName','lead_quality']).
 ↪size().unstack(fill_value=0)
ad_quality
```

[179]:
```
lead_quality                       Bad lead quality  Best lead quality  \
AdvertiserCampaignName
Debt Settlement1 Master                         298                151
creditsolutions-branded-shortform               190                 94

lead_quality                       Good lead quality
AdvertiserCampaignName
Debt Settlement1 Master                           85
creditsolutions-branded-shortform                 63
```

Analysis:

Debt Settlement1 Master has a significant number of bad quality leads at 298. Indicates that this campaign may have issues in targeting the right audience or filtering leads effectively.

Debt Settlement1 Master also has 151 best quality leads. This shows that despite the high number of bad leads, there is also a considerable number of high-quality leads, indicating potential in the campaign if optimizations are made.

Recommendations:

Given the high number of bad quality leads, it is essential to review and improve the targeting strategies and lead qualification criteria.

Implement strategies to convert good quality leads into the best quality category, such as personalized follow-ups, targeted offers, and improved customer engagement tactics.

9 3. If the advertiser says they will increase our CPL by 20% (i.e., $30 to $33) if we increase our lead quality by 20% (i.e., from 8.0% to 9.6%), do we see any opportunities to do that here? What kinds of things could we do?

[181]:
```
good_leads = ["Best lead quality", "Good lead quality"]  # Closed and leads␣
 ↪considered good
bad_leads = ["Bad lead quality"]
```

[182]:
```
total_leads = len(df_filtered)
good_leads_count = df_filtered[df_filtered['lead_quality'].isin(good_leads)].
 ↪shape[0]
```

Lead Quality Rate=(Total Number of Leads/Number of High-Quality Leads) * 100

[183]:
```
current_quality_rate = (good_leads_count / total_leads) * 100
print(f"Current Lead Quality Rate: {current_quality_rate:.2f}%")
```

```
Current Lead Quality Rate: 44.61%
```

[184]:
```
#Calculate the target lead quality rate after a 20% increase.
target_quality_rate = current_quality_rate * 1.20
print(f"Target Lead Quality Rate: {target_quality_rate:.2f}%")
```

```
Target Lead Quality Rate: 53.53%
```

[185]:
```
# Distribution of lead dispositions
lead_distribution = df_filtered['lead_quality'].value_counts()
print(lead_distribution)

# Calculate the number of leads needed to meet the target quality rate
needed_good_leads_count = (target_quality_rate / 100) * total_leads
current_good_leads_needed = max(0, needed_good_leads_count - good_leads_count)
print(f"Additional Good Leads Needed: {current_good_leads_needed:.0f}")
```

```
lead_quality
Bad lead quality     488
Best lead quality    245
Good lead quality    148
Name: count, dtype: int64
Additional Good Leads Needed: 79
```

[186]:
```
# Print results
print(f"Current Lead Quality Rate: {current_quality_rate:.2f}%")
print(f"Target Lead Quality Rate: {target_quality_rate:.2f}%")
print(f"Additional Good Leads Needed: {current_good_leads_needed:.0f}")
```

```
# Lead distribution
lead_distribution = df_filtered['lead_quality'].value_counts()
print("Lead Distribution:")
print(lead_distribution)
```

```
Current Lead Quality Rate: 44.61%
Target Lead Quality Rate: 53.53%
Additional Good Leads Needed: 79
Lead Distribution:
lead_quality
Bad lead quality     488
Best lead quality    245
Good lead quality    148
Name: count, dtype: int64
```

Analysis:

We can see that if we increase our current lead quality rate by 20%, we need additional 79 more good leads in order to fulfill the target.

Recommendations:

The suggestions for improving lead quality include:

Improving the contact rate for 'Unable to Contact' leads. Validating lead profiles more thoroughly to reduce 'Invalid Profile' leads. Adjusting targeting criteria to reduce 'Doesn't Qualify' leads. Reviewing and following up on 'Unknown' leads to determine their quality.

[ ]:
```
!sudo apt-get install texlive-xetex texlive-fonts-recommended
↪texlive-plain-generic pandoc
```

```
Reading package lists… Done
Building dependency tree… Done
Reading state information… Done
The following additional packages will be installed:
  dvisvgm fonts-droid-fallback fonts-lato fonts-lmodern fonts-noto-mono
  fonts-texgyre fonts-urw-base35 libapache-pom-java
  libcmark-gfm-extensions0.29.0.gfm.3 libcmark-gfm0.29.0.gfm.3
  libcommons-logging-java libcommons-parent-java libfontbox-java libfontenc1
  libgs9 libgs9-common libidn12 libijs-0.35 libjbig2dec0 libkpathsea6
  libpdfbox-java libptexenc1 libruby3.0 libsynctex2 libteckit0 libtexlua53
  libtexluajit2 libwoff1 libzzip-0-13 lmodern pandoc-data poppler-data
  preview-latex-style rake ruby ruby-net-telnet ruby-rubygems ruby-webrick
  ruby-xmlrpc ruby3.0 rubygems-integration t1utils teckit tex-common tex-gyre
  texlive-base texlive-binaries texlive-latex-base texlive-latex-extra
  texlive-latex-recommended texlive-pictures tipa xfonts-encodings
  xfonts-utils
Suggested packages:
  fonts-noto fonts-freefont-otf | fonts-freefont-ttf libavalon-framework-java
  libcommons-logging-java-doc libexcalibur-logkit-java liblog4j1.2-java
```