RitikaNevatia_hw3

October 23, 2018

1 CSE 519: Assignment 3

2 Ritika Nevatia

ritika.nevatia@stonybrook.edu 112007313

```
In [1]: import os
        import time
        import gc
        import warnings
        import json
        from pandas.io.json import json_normalize
        import numpy as np
        import pandas as pd
        import matplotlib.pyplot as plt
        import seaborn as sns
        color = sns.color_palette()
        from sklearn.preprocessing import LabelEncoder
        from sklearn.model_selection import KFold
        from sklearn.metrics import mean_squared_error
        import lightgbm as lgb
        from sklearn.cluster import DBSCAN
        from sklearn.preprocessing import StandardScaler
        from sklearn import metrics
        from sklearn.neighbors import NearestNeighbors
```

```
/anaconda3/lib/python3.6/site-packages/lightgbm/__init__.py:46: UserWarning: Starting from vers
This means that in case of installing LightGBM from PyPI via the ``pip install lightgbm`` comma
Instead of that, you need to install the OpenMP library, which is required for running LightGBI
You can install the OpenMP library by the following command: ``brew install libomp``.
  "You can install the OpenMP library by the following command: ``brew install libomp``.", User
```

```
In [32]: def load_df(csv_path='/Users/Shared/Work/dsf/hw3/all/train.csv', nrows=None):
             JSON_COLUMNS = ['device', 'geoNetwork', 'totals', 'trafficSource']

             df = pd.read_csv(csv_path,
```

```
                             converters={column: json.loads for column in JSON_COLUMNS},
                             dtype={'fullVisitorId': 'str'},
                             nrows=nrows)

            for column in JSON_COLUMNS:
                column_as_df = json_normalize(df[column])
                column_as_df.columns = [f"{column}.{subcolumn}" for subcolumn in column_as_df
                df = df.drop(column, axis=1).merge(column_as_df, right_index=True, left_index=
            return df
        train = load_df('/Users/Shared/Work/dsf/hw3/all/train.csv')
        test = load_df('/Users/Shared/Work/dsf/hw3/all/test.csv')
        gc.collect()

Out[32]: 28

In [54]: train.head()

Out[54]:   channelGrouping      date        fullVisitorId  \
        0  Organic Search  20160902  1131660440785968503
        1  Organic Search  20160902   377306020877927890
        2  Organic Search  20160902  3895546263509774583
        3  Organic Search  20160902  4763447161404445595
        4  Organic Search  20160902    27294437909732085


                                sessionId  socialEngagementType      visitId  \
        0  1131660440785968503_1472830385  Not Socially Engaged  1472830385
        1   377306020877927890_1472880147  Not Socially Engaged  1472880147
        2  3895546263509774583_1472865386  Not Socially Engaged  1472865386
        3  4763447161404445595_1472881213  Not Socially Engaged  1472881213
        4    27294437909732085_1472822600  Not Socially Engaged  1472822600


           visitNumber  visitStartTime device.browser         device.browserSize  \
        0            1      1472830385         Chrome  not available in demo dataset
        1            1      1472880147        Firefox  not available in demo dataset
        2            1      1472865386         Chrome  not available in demo dataset
        3            1      1472881213     UC Browser  not available in demo dataset
        4            2      1472822600         Chrome  not available in demo dataset


                 device.browserVersion device.deviceCategory  \
        0  not available in demo dataset              desktop
        1  not available in demo dataset              desktop
        2  not available in demo dataset              desktop
        3  not available in demo dataset              desktop
        4  not available in demo dataset               mobile


                   device.flashVersion  device.isMobile  \
        0  not available in demo dataset             False
        1  not available in demo dataset             False
```

```
2  not available in demo dataset                    False
3  not available in demo dataset                    False
4  not available in demo dataset                     True

                    device.language     device.mobileDeviceBranding  \
0  not available in demo dataset  not available in demo dataset
1  not available in demo dataset  not available in demo dataset
2  not available in demo dataset  not available in demo dataset
3  not available in demo dataset  not available in demo dataset
4  not available in demo dataset  not available in demo dataset

           device.mobileDeviceInfo device.mobileDeviceMarketingName  \
0  not available in demo dataset    not available in demo dataset
1  not available in demo dataset    not available in demo dataset
2  not available in demo dataset    not available in demo dataset
3  not available in demo dataset    not available in demo dataset
4  not available in demo dataset    not available in demo dataset

           device.mobileDeviceModel      device.mobileInputSelector  \
0  not available in demo dataset  not available in demo dataset
1  not available in demo dataset  not available in demo dataset
2  not available in demo dataset  not available in demo dataset
3  not available in demo dataset  not available in demo dataset
4  not available in demo dataset  not available in demo dataset

  device.operatingSystem  device.operatingSystemVersion  \
0                Windows  not available in demo dataset
1              Macintosh  not available in demo dataset
2                Windows  not available in demo dataset
3                  Linux  not available in demo dataset
4                Android  not available in demo dataset

              device.screenColors        device.screenResolution  \
0  not available in demo dataset  not available in demo dataset
1  not available in demo dataset  not available in demo dataset
2  not available in demo dataset  not available in demo dataset
3  not available in demo dataset  not available in demo dataset
4  not available in demo dataset  not available in demo dataset

                  geoNetwork.city                 geoNetwork.cityId  \
0                            Izmir  not available in demo dataset
1  not available in demo dataset  not available in demo dataset
2                           Madrid  not available in demo dataset
3  not available in demo dataset  not available in demo dataset
4  not available in demo dataset  not available in demo dataset

  geoNetwork.continent geoNetwork.country              geoNetwork.latitude  \
0                 Asia             Turkey  not available in demo dataset
```

```
1             Oceania          Australia   not available in demo dataset
2              Europe              Spain   not available in demo dataset
3                Asia          Indonesia   not available in demo dataset
4              Europe     United Kingdom   not available in demo dataset


          geoNetwork.longitude              geoNetwork.metro  \
0  not available in demo dataset                    (not set)
1  not available in demo dataset  not available in demo dataset
2  not available in demo dataset                    (not set)
3  not available in demo dataset  not available in demo dataset
4  not available in demo dataset  not available in demo dataset


  geoNetwork.networkDomain     geoNetwork.networkLocation  \
0            ttnet.com.tr  not available in demo dataset
1             dodo.net.au  not available in demo dataset
2         unknown.unknown  not available in demo dataset
3         unknown.unknown  not available in demo dataset
4         unknown.unknown  not available in demo dataset


              geoNetwork.region geoNetwork.subContinent totals.bounces  \
0                          Izmir            Western Asia              1
1  not available in demo dataset             Australasia              1
2            Community of Madrid         Southern Europe              1
3  not available in demo dataset          Southeast Asia              1
4  not available in demo dataset         Northern Europe              1


  totals.hits totals.newVisits totals.pageviews totals.transactionRevenue  \
0           1                1                1                       NaN
1           1                1                1                       NaN
2           1                1                1                       NaN
3           1                1                1                       NaN
4           1              NaN                1                       NaN


  totals.visits trafficSource.adContent  \
0             1                     NaN
1             1                     NaN
2             1                     NaN
3             1                     NaN
4             1                     NaN


  trafficSource.adwordsClickInfo.adNetworkType  \
0                                          NaN
1                                          NaN
2                                          NaN
3                                          NaN
4                                          NaN


  trafficSource.adwordsClickInfo.criteriaParameters  \
```

```
0                     not available in demo dataset
1                     not available in demo dataset
2                     not available in demo dataset
3                     not available in demo dataset
4                     not available in demo dataset


  trafficSource.adwordsClickInfo.gclId  \
0                                  NaN
1                                  NaN
2                                  NaN
3                                  NaN
4                                  NaN


  trafficSource.adwordsClickInfo.isVideoAd  \
0                                      NaN
1                                      NaN
2                                      NaN
3                                      NaN
4                                      NaN


  trafficSource.adwordsClickInfo.page trafficSource.adwordsClickInfo.slot  \
0                                 NaN                                 NaN
1                                 NaN                                 NaN
2                                 NaN                                 NaN
3                                 NaN                                 NaN
4                                 NaN                                 NaN


  trafficSource.campaign trafficSource.campaignCode  \
0             (not set)                        NaN
1             (not set)                        NaN
2             (not set)                        NaN
3             (not set)                        NaN
4             (not set)                        NaN


  trafficSource.isTrueDirect trafficSource.keyword trafficSource.medium  \
0                        NaN       (not provided)              organic
1                        NaN       (not provided)              organic
2                        NaN       (not provided)              organic
3                        NaN        google + online            organic
4                       True       (not provided)              organic


  trafficSource.referralPath trafficSource.source
0                        NaN               google
1                        NaN               google
2                        NaN               google
3                        NaN               google
4                        NaN               google
```

```
In [55]: fig, axes = plt.subplots(2,2,figsize=(15,15))
         train["device.isMobile"].value_counts().plot(kind="bar",ax=axes[0][0],rot=25,legend=":
         train["device.browser"].value_counts().head(7).plot(kind="bar",ax=axes[0][1],rot=40,le
         train["device.deviceCategory"].value_counts().plot(kind="bar",ax=axes[1][0],rot=25,leg
         train["device.operatingSystem"].value_counts().head(7).plot(kind="bar",ax=axes[1][1],i
```

Out[55]: <matplotlib.axes._subplots.AxesSubplot at 0x1edf7be160>
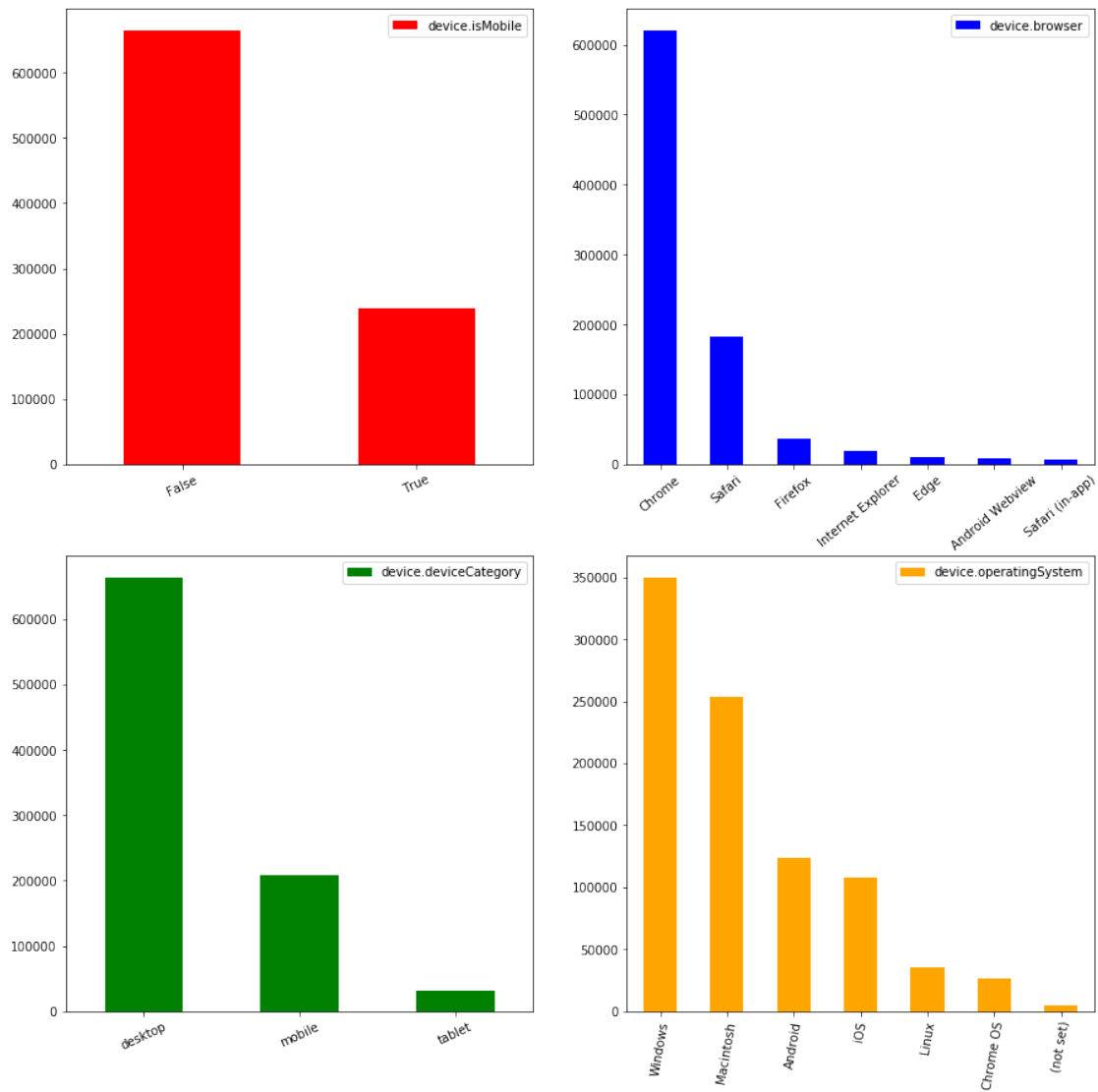

```
In [56]: fig, axes = plt.subplots(3,2,figsize=(15,15))
         fig.tight_layout()
         plt.subplots_adjust(left=None, bottom=None, right=None, top=None, wspace=0.5, hspace=(

         train["geoNetwork.continent"].value_counts().plot(kind="bar",ax=axes[0][0],rot=25,lege
```

```python
train["geoNetwork.subContinent"].value_counts().head(10).plot(kind="bar",ax=axes[0][1]
train["geoNetwork.country"].value_counts().head(10).plot(kind="bar",ax=axes[1][0],rot=
train["geoNetwork.region"].value_counts().head(10).plot(kind="bar",ax=axes[1][1],rot=8
train["geoNetwork.metro"].value_counts().head(5).plot(kind="bar",ax=axes[2][0],rot=10
train["geoNetwork.city"].value_counts().head(5).plot(kind="bar",ax=axes[2][1],rot=20,1
```

Out[56]: <matplotlib.axes._subplots.AxesSubplot at 0x1edf5879e8>
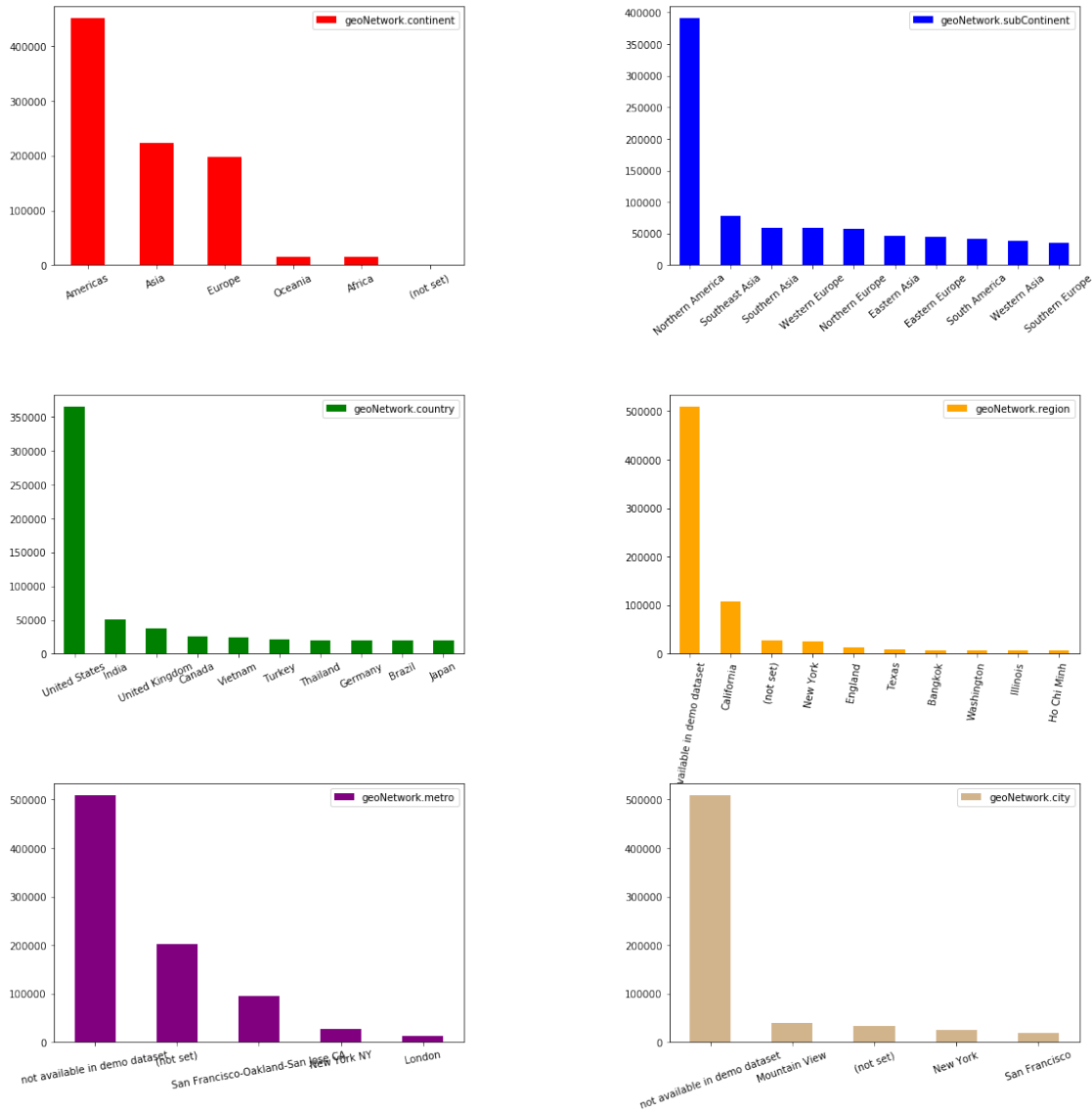

```python
In [57]: fig, axes = plt.subplots(2,2, figsize=(15,15))
train[train["geoNetwork.continent"] == "Africa"]["geoNetwork.subContinent"].value_cou
train[train["geoNetwork.continent"] == "Americas"]["geoNetwork.subContinent"].value_c
train[train["geoNetwork.continent"] == "Asia"]["geoNetwork.subContinent"].value_counts
train[train["geoNetwork.continent"] == "Europe"]["geoNetwork.subContinent"].value_cou
```

Africa Distribution

Asia Distribution

America Distribution

Europe Distribution

```
In [58]: rev = train[["totals.transactionRevenue" , "date"]].dropna()
         rev.head()
```

Out[58]:   totals.transactionRevenue   date
        752   37860000 20160902
        753   306670000 20160902
        799   68030000 20160902
        802   26250000 20160902
        859   574150000 20160902

```
In [59]: rev["totals.transactionRevenue"] = rev["totals.transactionRevenue"].astype(np.int64)
```

```
In [60]: fig,ax = plt.subplots(1,1,figsize=(10,5))
         ax.set_title("Log(visitNumbers)")
```

8

```
        ax.set_ylabel("Repetition Count")
        ax.set_xlabel("Log of number of visits")
        ax.grid(color="blue", linewidth=0.1)
        ax.hist(np.log(train.visitNumber), color="orange")
```

Out[60]: (array([7.03060e+05, 1.28391e+05, 3.84490e+04, 1.49430e+04, 9.48100e+03,
         4.35900e+03, 2.06200e+03, 1.51700e+03, 9.86000e+02, 4.05000e+02]),
         array([0.   , 0.59788858, 1.19577715, 1.79366573, 2.39155431,
         2.98944288, 3.58733146, 4.18522004, 4.78310861, 5.38099719,
         5.97888576]),
         <a list of 10 Patch objects>)
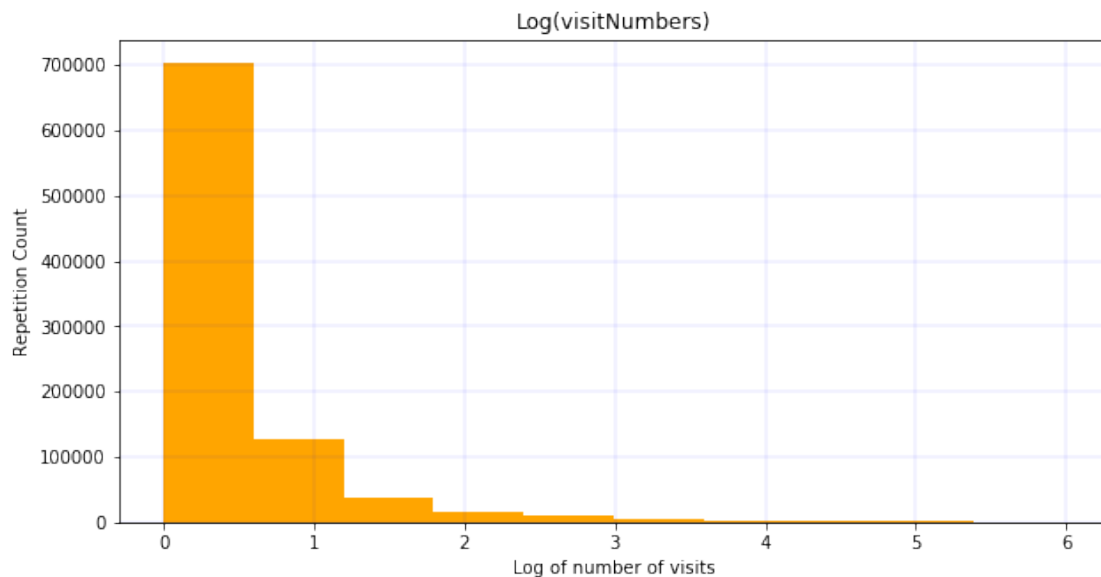

```
In [61]: train.date_ = pd.to_datetime(train.date, format="%Y%m%d")
        plt.figure(figsize=(20,6))
        train.date_.value_counts().sort_index().plot(color = "red")
        plt.legend()
```

/anaconda3/lib/python3.6/site-packages/ipykernel_launcher.py:1: UserWarning: Pandas doesn't all
  """Entry point for launching an IPython kernel.

Out[61]: <matplotlib.legend.Legend at 0x1edf20de48>
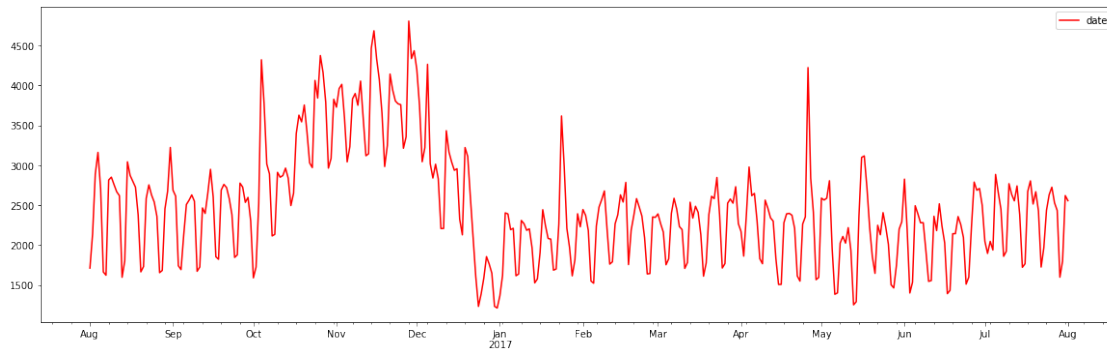
```
In [62]: test.date_ = pd.to_datetime(test.date, format="%Y%m%d")
         plt.figure(figsize=(20,6))
         test.date_.value_counts().sort_index().plot(color = "blue")
         plt.legend()
```

/anaconda3/lib/python3.6/site-packages/ipykernel_launcher.py:1: UserWarning: Pandas doesn't al
  """"Entry point for launching an IPython kernel.

Out[62]: <matplotlib.legend.Legend at 0x1edf14b2e8>
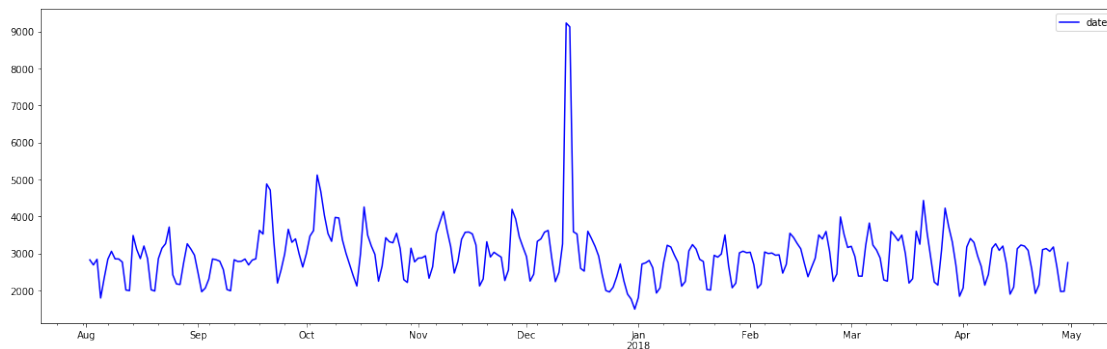

```
In [3]: gdp_df = pd.read_csv('/Users/Shared/Work/dsf/hw3/WEO_Data.xls', encoding='latin-1', sep
```

```
In [64]: gdp_df.head()
```

```
Out[64]:                    Country                                                Units  Scale  \
         0            Afghanistan  Purchasing power parity; international dollars  Units
         1                Albania  Purchasing power parity; international dollars  Units
         2                Algeria  Purchasing power parity; international dollars  Units
         3                 Angola  Purchasing power parity; international dollars  Units
         4    Antigua and Barbuda  Purchasing power parity; international dollars  Units
```

10

```
              2016         2017  Estimates Start After
0    1923276.0    1957577.0                   2016.0
1   11820515.0   12506645.0                   2011.0
2   14955261.0   15237198.0                   2016.0
3    6783266.0    6752580.0                   2015.0
4   25347436.0   26231545.0                   2011.0
```

In [65]: gdp_df.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 194 entries, 0 to 193
Data columns (total 6 columns):
Country                 194 non-null object
Units                   193 non-null object
Scale                   193 non-null object
2016                    191 non-null float64
2017                    191 non-null float64
Estimates Start After   192 non-null float64
dtypes: float64(3), object(3)
memory usage: 9.2+ KB
```

In [66]: train.head()

```
Out[66]:   channelGrouping      date       fullVisitorId  \
        0  Organic Search  20160902  1131660440785968503
        1  Organic Search  20160902   377306020877927890
        2  Organic Search  20160902  3895546263509774583
        3  Organic Search  20160902  4763447161404445595
        4  Organic Search  20160902    27294437909732085


                                sessionId  socialEngagementType     visitId  \
        0  1131660440785968503_1472830385  Not Socially Engaged  1472830385
        1   377306020877927890_1472880147  Not Socially Engaged  1472880147
        2  3895546263509774583_1472865386  Not Socially Engaged  1472865386
        3  4763447161404445595_1472881213  Not Socially Engaged  1472881213
        4    27294437909732085_1472822600  Not Socially Engaged  1472822600


           visitNumber  visitStartTime device.browser           device.browserSize  \
        0            1      1472830385         Chrome  not available in demo dataset
        1            1      1472880147        Firefox  not available in demo dataset
        2            1      1472865386         Chrome  not available in demo dataset
        3            1      1472881213     UC Browser  not available in demo dataset
        4            2      1472822600         Chrome  not available in demo dataset


              device.browserVersion device.deviceCategory  \
        0  not available in demo dataset               desktop
        1  not available in demo dataset               desktop
```

```
2   not available in demo dataset                  desktop
3   not available in demo dataset                  desktop
4   not available in demo dataset                   mobile

               device.flashVersion  device.isMobile  \
0   not available in demo dataset            False
1   not available in demo dataset            False
2   not available in demo dataset            False
3   not available in demo dataset            False
4   not available in demo dataset             True

                device.language    device.mobileDeviceBranding  \
0   not available in demo dataset  not available in demo dataset
1   not available in demo dataset  not available in demo dataset
2   not available in demo dataset  not available in demo dataset
3   not available in demo dataset  not available in demo dataset
4   not available in demo dataset  not available in demo dataset

          device.mobileDeviceInfo device.mobileDeviceMarketingName  \
0   not available in demo dataset    not available in demo dataset
1   not available in demo dataset    not available in demo dataset
2   not available in demo dataset    not available in demo dataset
3   not available in demo dataset    not available in demo dataset
4   not available in demo dataset    not available in demo dataset

          device.mobileDeviceModel    device.mobileInputSelector  \
0   not available in demo dataset  not available in demo dataset
1   not available in demo dataset  not available in demo dataset
2   not available in demo dataset  not available in demo dataset
3   not available in demo dataset  not available in demo dataset
4   not available in demo dataset  not available in demo dataset

  device.operatingSystem  device.operatingSystemVersion  \
0                Windows  not available in demo dataset
1              Macintosh  not available in demo dataset
2                Windows  not available in demo dataset
3                  Linux  not available in demo dataset
4                Android  not available in demo dataset

             device.screenColors        device.screenResolution  \
0   not available in demo dataset  not available in demo dataset
1   not available in demo dataset  not available in demo dataset
2   not available in demo dataset  not available in demo dataset
3   not available in demo dataset  not available in demo dataset
4   not available in demo dataset  not available in demo dataset

                geoNetwork.city              geoNetwork.cityId  \
0                          Izmir  not available in demo dataset
```

```
1  not available in demo dataset  not available in demo dataset
2                         Madrid  not available in demo dataset
3  not available in demo dataset  not available in demo dataset
4  not available in demo dataset  not available in demo dataset

  geoNetwork.continent geoNetwork.country           geoNetwork.latitude  \
0                 Asia             Turkey  not available in demo dataset
1              Oceania          Australia  not available in demo dataset
2               Europe              Spain  not available in demo dataset
3                 Asia          Indonesia  not available in demo dataset
4               Europe     United Kingdom  not available in demo dataset

           geoNetwork.longitude              geoNetwork.metro  \
0  not available in demo dataset                     (not set)
1  not available in demo dataset  not available in demo dataset
2  not available in demo dataset                     (not set)
3  not available in demo dataset  not available in demo dataset
4  not available in demo dataset  not available in demo dataset

  geoNetwork.networkDomain     geoNetwork.networkLocation  \
0           ttnet.com.tr  not available in demo dataset
1            dodo.net.au  not available in demo dataset
2        unknown.unknown  not available in demo dataset
3        unknown.unknown  not available in demo dataset
4        unknown.unknown  not available in demo dataset

                geoNetwork.region geoNetwork.subContinent totals.bounces  \
0                           Izmir            Western Asia              1
1  not available in demo dataset             Australasia              1
2             Community of Madrid         Southern Europe              1
3  not available in demo dataset          Southeast Asia              1
4  not available in demo dataset         Northern Europe              1

  totals.hits totals.newVisits totals.pageviews totals.transactionRevenue  \
0           1                1                1                       NaN
1           1                1                1                       NaN
2           1                1                1                       NaN
3           1                1                1                       NaN
4           1              NaN                1                       NaN

  totals.visits trafficSource.adContent  \
0             1                     NaN
1             1                     NaN
2             1                     NaN
3             1                     NaN
4             1                     NaN

  trafficSource.adwordsClickInfo.adNetworkType  \
```

```
0                                              NaN
1                                              NaN
2                                              NaN
3                                              NaN
4                                              NaN


  trafficSource.adwordsClickInfo.criteriaParameters  \
0                 not available in demo dataset
1                 not available in demo dataset
2                 not available in demo dataset
3                 not available in demo dataset
4                 not available in demo dataset


  trafficSource.adwordsClickInfo.gclId  \
0                                  NaN
1                                  NaN
2                                  NaN
3                                  NaN
4                                  NaN


  trafficSource.adwordsClickInfo.isVideoAd  \
0                                      NaN
1                                      NaN
2                                      NaN
3                                      NaN
4                                      NaN


  trafficSource.adwordsClickInfo.page trafficSource.adwordsClickInfo.slot  \
0                                  NaN                                 NaN
1                                  NaN                                 NaN
2                                  NaN                                 NaN
3                                  NaN                                 NaN
4                                  NaN                                 NaN


  trafficSource.campaign trafficSource.campaignCode  \
0            (not set)                          NaN
1            (not set)                          NaN
2            (not set)                          NaN
3            (not set)                          NaN
4            (not set)                          NaN


  trafficSource.isTrueDirect trafficSource.keyword trafficSource.medium  \
0                       NaN        (not provided)              organic
1                       NaN        (not provided)              organic
2                       NaN        (not provided)              organic
3                       NaN       google + online              organic
4                      True        (not provided)              organic
```

```
        trafficSource.referralPath trafficSource.source
      0                        NaN                google
      1                        NaN                google
      2                        NaN                google
      3                        NaN                google
      4                        NaN                google
```

In [4]: train = pd.merge(train, gdp_df, left_on=['geoNetwork.country'], right_on=['Country'],
        test = pd.merge(test, gdp_df, left_on=['geoNetwork.country'], right_on=['Country'], hou

In [5]: pd.set_option('display.max_columns', None)

        train.drop('Country', axis=1, inplace=True)
        train.drop('Units', axis=1, inplace=True)
        train.drop('Scale', axis=1, inplace=True)
        train.drop('2016', axis=1, inplace=True)
        train.drop('Estimates Start After', axis=1, inplace=True)
        train.head()

```
Out[5]:   channelGrouping        date        fullVisitorId  \
        0  Organic Search  20160902.0  1131660440785968503
        1  Organic Search  20160902.0   357659889600827884
        2  Organic Search  20160902.0  9675839814855309060
        3  Organic Search  20160902.0  9987056085213279488
        4  Organic Search  20160902.0  2936426984187397286


                           sessionId  socialEngagementType      visitId  \
        0  1131660440785968503_1472830385  Not Socially Engaged  1.472830e+09
        1   357659889600827884_1472839882  Not Socially Engaged  1.472840e+09
        2  9675839814855309060_1472811887  Not Socially Engaged  1.472812e+09
        3  9987056085213279488_1472842263  Not Socially Engaged  1.472842e+09
        4  2936426984187397286_1472833893  Not Socially Engaged  1.472834e+09


           visitNumber  visitStartTime device.browser         device.browserSize  \
        0          1.0    1.472830e+09         Chrome  not available in demo dataset
        1          1.0    1.472840e+09         Safari  not available in demo dataset
        2          1.0    1.472812e+09         Chrome  not available in demo dataset
        3          1.0    1.472842e+09         Chrome  not available in demo dataset
        4          1.0    1.472834e+09         Chrome  not available in demo dataset


                 device.browserVersion device.deviceCategory  \
        0  not available in demo dataset               desktop
        1  not available in demo dataset                mobile
        2  not available in demo dataset               desktop
        3  not available in demo dataset               desktop
        4  not available in demo dataset               desktop


                    device.flashVersion device.isMobile  \
```

15

```
0  not available in demo dataset             False
1  not available in demo dataset              True
2  not available in demo dataset             False
3  not available in demo dataset             False
4  not available in demo dataset             False


                     device.language    device.mobileDeviceBranding  \
0  not available in demo dataset  not available in demo dataset
1  not available in demo dataset  not available in demo dataset
2  not available in demo dataset  not available in demo dataset
3  not available in demo dataset  not available in demo dataset
4  not available in demo dataset  not available in demo dataset


          device.mobileDeviceInfo device.mobileDeviceMarketingName  \
0  not available in demo dataset    not available in demo dataset
1  not available in demo dataset    not available in demo dataset
2  not available in demo dataset    not available in demo dataset
3  not available in demo dataset    not available in demo dataset
4  not available in demo dataset    not available in demo dataset


          device.mobileDeviceModel    device.mobileInputSelector  \
0  not available in demo dataset  not available in demo dataset
1  not available in demo dataset  not available in demo dataset
2  not available in demo dataset  not available in demo dataset
3  not available in demo dataset  not available in demo dataset
4  not available in demo dataset  not available in demo dataset


  device.operatingSystem  device.operatingSystemVersion  \
0                Windows  not available in demo dataset
1                    iOS  not available in demo dataset
2                Windows  not available in demo dataset
3                Windows  not available in demo dataset
4                Windows  not available in demo dataset


               device.screenColors        device.screenResolution  \
0  not available in demo dataset  not available in demo dataset
1  not available in demo dataset  not available in demo dataset
2  not available in demo dataset  not available in demo dataset
3  not available in demo dataset  not available in demo dataset
4  not available in demo dataset  not available in demo dataset


                  geoNetwork.city              geoNetwork.cityId  \
0                           Izmir  not available in demo dataset
1                           Izmir  not available in demo dataset
2                        Istanbul  not available in demo dataset
3  not available in demo dataset  not available in demo dataset
4                        Istanbul  not available in demo dataset
```

```
   geoNetwork.continent geoNetwork.country          geoNetwork.latitude  \
0                 Asia            Turkey  not available in demo dataset
1                 Asia            Turkey  not available in demo dataset
2                 Asia            Turkey  not available in demo dataset
3                 Asia            Turkey  not available in demo dataset
4                 Asia            Turkey  not available in demo dataset


            geoNetwork.longitude               geoNetwork.metro  \
0  not available in demo dataset                      (not set)
1  not available in demo dataset                      (not set)
2  not available in demo dataset                      (not set)
3  not available in demo dataset  not available in demo dataset
4  not available in demo dataset                      (not set)


  geoNetwork.networkDomain     geoNetwork.networkLocation  \
0           ttnet.com.tr  not available in demo dataset
1       unknown.unknown  not available in demo dataset
2              dr.com.tr  not available in demo dataset
3           ttnet.com.tr  not available in demo dataset
4           ttnet.com.tr  not available in demo dataset


                geoNetwork.region geoNetwork.subContinent totals.bounces  \
0                           Izmir             Western Asia              1
1                           Izmir             Western Asia              1
2                        Istanbul             Western Asia              1
3  not available in demo dataset             Western Asia              1
4                        Istanbul             Western Asia            NaN


  totals.hits totals.newVisits totals.pageviews totals.transactionRevenue  \
0           1                1                1                       NaN
1           1                1                1                       NaN
2           1                1                1                       NaN
3           1                1                1                       NaN
4           2                1                2                       NaN


  totals.visits trafficSource.adContent  \
0             1                     NaN
1             1                     NaN
2             1                     NaN
3             1                     NaN
4             1                     NaN


  trafficSource.adwordsClickInfo.adNetworkType  \
0                                           NaN
1                                           NaN
2                                           NaN
3                                           NaN
4                                           NaN
```

```
   trafficSource.adwordsClickInfo.criteriaParameters  \
0                    not available in demo dataset
1                    not available in demo dataset
2                    not available in demo dataset
3                    not available in demo dataset
4                    not available in demo dataset

  trafficSource.adwordsClickInfo.gclId  \
0                                  NaN
1                                  NaN
2                                  NaN
3                                  NaN
4                                  NaN

  trafficSource.adwordsClickInfo.isVideoAd  \
0                                      NaN
1                                      NaN
2                                      NaN
3                                      NaN
4                                      NaN

  trafficSource.adwordsClickInfo.page trafficSource.adwordsClickInfo.slot  \
0                                 NaN                                 NaN
1                                 NaN                                 NaN
2                                 NaN                                 NaN
3                                 NaN                                 NaN
4                                 NaN                                 NaN

  trafficSource.campaign trafficSource.campaignCode  \
0             (not set)                         NaN
1             (not set)                         NaN
2             (not set)                         NaN
3             (not set)                         NaN
4             (not set)                         NaN

  trafficSource.isTrueDirect trafficSource.keyword trafficSource.medium  \
0                        NaN         (not provided)              organic
1                        NaN         (not provided)              organic
2                        NaN         (not provided)              organic
3                        NaN         (not provided)              organic
4                        NaN         (not provided)              organic

  trafficSource.referralPath trafficSource.source          2017
0                        NaN               google    26892871.0
1                        NaN               google    26892871.0
2                        NaN               google    26892871.0
3                        NaN               google    26892871.0
```

```
          4                       NaN              google   26892871.0

In [6]: test.drop('Country', axis=1, inplace=True)
        test.drop('Units', axis=1, inplace=True)
        test.drop('Scale', axis=1, inplace=True)
        test.drop('2016', axis=1, inplace=True)
        test.drop('Estimates Start After', axis=1, inplace=True)
        test.head()

Out[6]:    channelGrouping        date        fullVisitorId  \
        0  Organic Search   20171016.0   6167871330617112363
        1  Organic Search   20171016.0    394911561846949602
        2  Organic Search   20171016.0   8826908841852350020
        3  Organic Search   20171016.0   0455927505163133323
        4          Direct   20171016.0   2014206783685676144


                            sessionId  socialEngagementType        visitId  \
        0  6167871330617112363_1508151024  Not Socially Engaged   1.508151e+09
        1   394911561846949602_1508177948  Not Socially Engaged   1.508178e+09
        2  8826908841852350020_1508144195  Not Socially Engaged   1.508144e+09
        3  0455927505163133323_1508147676  Not Socially Engaged   1.508148e+09
        4  2014206783685676144_1508158542  Not Socially Engaged   1.508159e+09


           visitNumber   visitStartTime device.browser         device.browserSize  \
        0          2.0     1.508151e+09         Chrome  not available in demo dataset
        1          1.0     1.508178e+09         Safari  not available in demo dataset
        2          1.0     1.508144e+09         Chrome  not available in demo dataset
        3          1.0     1.508148e+09         Chrome  not available in demo dataset
        4          1.0     1.508159e+09         Chrome  not available in demo dataset


                  device.browserVersion device.deviceCategory  \
        0  not available in demo dataset               desktop
        1  not available in demo dataset                mobile
        2  not available in demo dataset               desktop
        3  not available in demo dataset               desktop
        4  not available in demo dataset               desktop


                    device.flashVersion device.isMobile  \
        0  not available in demo dataset           False
        1  not available in demo dataset            True
        2  not available in demo dataset           False
        3  not available in demo dataset           False
        4  not available in demo dataset           False


                       device.language    device.mobileDeviceBranding  \
        0  not available in demo dataset  not available in demo dataset
        1  not available in demo dataset  not available in demo dataset
        2  not available in demo dataset  not available in demo dataset
```

19

```
3  not available in demo dataset  not available in demo dataset
4  not available in demo dataset  not available in demo dataset

          device.mobileDeviceInfo device.mobileDeviceMarketingName  \
0  not available in demo dataset     not available in demo dataset
1  not available in demo dataset     not available in demo dataset
2  not available in demo dataset     not available in demo dataset
3  not available in demo dataset     not available in demo dataset
4  not available in demo dataset     not available in demo dataset

          device.mobileDeviceModel     device.mobileInputSelector  \
0  not available in demo dataset  not available in demo dataset
1  not available in demo dataset  not available in demo dataset
2  not available in demo dataset  not available in demo dataset
3  not available in demo dataset  not available in demo dataset
4  not available in demo dataset  not available in demo dataset

  device.operatingSystem  device.operatingSystemVersion  \
0              Macintosh  not available in demo dataset
1                    iOS  not available in demo dataset
2                Windows  not available in demo dataset
3              Macintosh  not available in demo dataset
4                Windows  not available in demo dataset

               device.screenColors        device.screenResolution  \
0  not available in demo dataset  not available in demo dataset
1  not available in demo dataset  not available in demo dataset
2  not available in demo dataset  not available in demo dataset
3  not available in demo dataset  not available in demo dataset
4  not available in demo dataset  not available in demo dataset

  geoNetwork.city               geoNetwork.cityId geoNetwork.continent  \
0       (not set)  not available in demo dataset                 Asia
1       (not set)  not available in demo dataset                 Asia
2       (not set)  not available in demo dataset                 Asia
3       Singapore  not available in demo dataset                 Asia
4       Singapore  not available in demo dataset                 Asia

  geoNetwork.country              geoNetwork.latitude  \
0          Singapore  not available in demo dataset
1          Singapore  not available in demo dataset
2          Singapore  not available in demo dataset
3          Singapore  not available in demo dataset
4          Singapore  not available in demo dataset

             geoNetwork.longitude geoNetwork.metro geoNetwork.networkDomain  \
0  not available in demo dataset        (not set)          myrepublic.com.sg
1  not available in demo dataset        (not set)               vqbn.com.sg
```

```
2  not available in demo dataset          (not set)         singnet.com.sg
3  not available in demo dataset          (not set)            commz.com.sg
4  not available in demo dataset          (not set)        alphawave.com.sg


       geoNetwork.networkLocation geoNetwork.region geoNetwork.subContinent  \
0  not available in demo dataset         (not set)          Southeast Asia
1  not available in demo dataset         (not set)          Southeast Asia
2  not available in demo dataset         (not set)          Southeast Asia
3  not available in demo dataset         (not set)          Southeast Asia
4  not available in demo dataset         (not set)          Southeast Asia


  totals.bounces totals.hits totals.newVisits totals.pageviews totals.visits  \
0            NaN           4              NaN                4             1
1            NaN          12                1               11             1
2            NaN           7                1                7             1
3            NaN          10                1                7             1
4            NaN           4                1                4             1


  trafficSource.adContent trafficSource.adwordsClickInfo.adNetworkType  \
0                     NaN                                          NaN
1                     NaN                                          NaN
2                     NaN                                          NaN
3                     NaN                                          NaN
4                     NaN                                          NaN


  trafficSource.adwordsClickInfo.criteriaParameters  \
0                 not available in demo dataset
1                 not available in demo dataset
2                 not available in demo dataset
3                 not available in demo dataset
4                 not available in demo dataset


  trafficSource.adwordsClickInfo.gclId  \
0                                  NaN
1                                  NaN
2                                  NaN
3                                  NaN
4                                  NaN


  trafficSource.adwordsClickInfo.isVideoAd  \
0                                      NaN
1                                      NaN
2                                      NaN
3                                      NaN
4                                      NaN


  trafficSource.adwordsClickInfo.page trafficSource.adwordsClickInfo.slot  \
0                                 NaN                                 NaN
```

```
     1                                NaN                                NaN
     2                                NaN                                NaN
     3                                NaN                                NaN
     4                                NaN                                NaN

       trafficSource.campaign trafficSource.isTrueDirect trafficSource.keyword  \
     0            (not set)                       True        (not provided)
     1            (not set)                        NaN        (not provided)
     2            (not set)                        NaN        (not provided)
     3            (not set)                        NaN        (not provided)
     4            (not set)                       True                   NaN

       trafficSource.medium trafficSource.referralPath trafficSource.source  \
     0            organic                           NaN                google
     1            organic                           NaN                google
     2            organic                           NaN                google
     3            organic                           NaN                google
     4             (none)                           NaN              (direct)

               2017
     0  93905495.0
     1  93905495.0
     2  93905495.0
     3  93905495.0
     4  93905495.0
```

```python
In [7]: def market_type(x):
            if  ('United States' in x):
                return 0
            elif  ('Canada' in x):
                return 0
            else:
                return 1

        train['marketType'] = train['geoNetwork.country'].map(lambda x:market_type(str(x)))
        test['marketType'] = test['geoNetwork.country'].map(lambda x:market_type(str(x)))

In [8]: def ad_type(x):
            if  ('google' in x):
                return 'google'
            elif  ('placement' in x) | ('placememnt' in x):
                return 'placement'
            elif '(not set)' in x or 'nan' in x:
                return x
            elif 'ad' in x:
                return 'ad'
            else:
                return 'others'
```

```
        train['adType'] = train['trafficSource.adContent'].map(lambda x:market_type(str(x)))
        test['adType'] = test['trafficSource.adContent'].map(lambda x:market_type(str(x)))


In [9]: def browser_type(x):
        browsers = ['chrome','safari','firefox','internet explorer','edge','opera','coc co
        if x in browsers:
            return x.lower()
        elif  ('android' in x) or ('samsung' in x) or ('mini' in x) or ('iphone' in x) or
            return 'mobile browser'
        elif  ('mozilla' in x) or ('chrome' in x) or ('blackberry' in x) or ('nokia' in x)
            return 'mobile browser'
        elif  ('lunascape' in x) or ('netscape' in x) or ('blackberry' in x) or ('konqueror
            return 'mobile browser'
        elif '(not set)' in x:
            return x
        else:
            return 'others'


        train['browserType'] = train['device.browser'].map(lambda x:market_type(str(x)))
        test['browserType'] = test['device.browser'].map(lambda x:market_type(str(x)))

In [73]: train.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 903686 entries, 0 to 903685
Data columns (total 59 columns):
channelGrouping                         903653 non-null object
date                                    903653 non-null float64
fullVisitorId                           903653 non-null object
sessionId                               903653 non-null object
socialEngagementType                    903653 non-null object
visitId                                 903653 non-null float64
visitNumber                             903653 non-null float64
visitStartTime                          903653 non-null float64
device.browser                          903653 non-null object
device.browserSize                      903653 non-null object
device.browserVersion                   903653 non-null object
device.deviceCategory                   903653 non-null object
device.flashVersion                     903653 non-null object
device.isMobile                         903653 non-null object
device.language                         903653 non-null object
device.mobileDeviceBranding             903653 non-null object
device.mobileDeviceInfo                 903653 non-null object
device.mobileDeviceMarketingName        903653 non-null object
device.mobileDeviceModel                903653 non-null object
device.mobileInputSelector              903653 non-null object
```

```
device.operatingSystem                                903653 non-null object
device.operatingSystemVersion                         903653 non-null object
device.screenColors                                   903653 non-null object
device.screenResolution                               903653 non-null object
geoNetwork.city                                       903653 non-null object
geoNetwork.cityId                                     903653 non-null object
geoNetwork.continent                                  903653 non-null object
geoNetwork.country                                    903653 non-null object
geoNetwork.latitude                                   903653 non-null object
geoNetwork.longitude                                  903653 non-null object
geoNetwork.metro                                      903653 non-null object
geoNetwork.networkDomain                              903653 non-null object
geoNetwork.networkLocation                            903653 non-null object
geoNetwork.region                                    903653 non-null object
geoNetwork.subContinent                              903653 non-null object
totals.bounces                                        450630 non-null object
totals.hits                                           903653 non-null object
totals.newVisits                                      703060 non-null object
totals.pageviews                                      903553 non-null object
totals.transactionRevenue                             11515 non-null object
totals.visits                                         903653 non-null object
trafficSource.adContent                               10946 non-null object
trafficSource.adwordsClickInfo.adNetworkType          21460 non-null object
trafficSource.adwordsClickInfo.criteriaParameters     903653 non-null object
trafficSource.adwordsClickInfo.gclId                  21561 non-null object
trafficSource.adwordsClickInfo.isVideoAd              21460 non-null object
trafficSource.adwordsClickInfo.page                   21460 non-null object
trafficSource.adwordsClickInfo.slot                   21460 non-null object
trafficSource.campaign                                903653 non-null object
trafficSource.campaignCode                            1 non-null object
trafficSource.isTrueDirect                            274005 non-null object
trafficSource.keyword                                 400724 non-null object
trafficSource.medium                                  903653 non-null object
trafficSource.referralPath                            330941 non-null object
trafficSource.source                                  903653 non-null object
2017                                                  869164 non-null float64
marketType                                            903686 non-null int64
adType                                                903686 non-null int64
browserType                                           903686 non-null int64
dtypes: float64(5), int64(3), object(51)
memory usage: 413.7+ MB
```

```python
In [10]: train['date_type'] = pd.to_datetime(train['date'])
         test['date_type'] = pd.to_datetime(test['date'])
         train['dayOfWeek'] = train['date_type'].dt.day_name()
         test['dayOfWeek'] = test['date_type'].dt.day_name()
         train.drop('date_type', axis=1, inplace=True)
```

```
         test.drop('date_type', axis=1, inplace=True)

In [75]: train.head()

Out[75]:   channelGrouping         date        fullVisitorId  \
         0  Organic Search  20160902.0   1131660440785968503
         1  Organic Search  20160902.0    357659889600827884
         2  Organic Search  20160902.0   9675839814855309060
         3  Organic Search  20160902.0   9987056085213279488
         4  Organic Search  20160902.0   2936426984187397286


                                sessionId  socialEngagementType      visitId  \
         0  1131660440785968503_1472830385  Not Socially Engaged  1.472830e+09
         1   357659889600827884_1472839882  Not Socially Engaged  1.472840e+09
         2  9675839814855309060_1472811887  Not Socially Engaged  1.472812e+09
         3  9987056085213279488_1472842263  Not Socially Engaged  1.472842e+09
         4  2936426984187397286_1472833893  Not Socially Engaged  1.472834e+09


            visitNumber  visitStartTime device.browser        device.browserSize  \
         0          1.0    1.472830e+09         Chrome  not available in demo dataset
         1          1.0    1.472840e+09         Safari  not available in demo dataset
         2          1.0    1.472812e+09         Chrome  not available in demo dataset
         3          1.0    1.472842e+09         Chrome  not available in demo dataset
         4          1.0    1.472834e+09         Chrome  not available in demo dataset


                   device.browserVersion device.deviceCategory  \
         0  not available in demo dataset               desktop
         1  not available in demo dataset                mobile
         2  not available in demo dataset               desktop
         3  not available in demo dataset               desktop
         4  not available in demo dataset               desktop


                     device.flashVersion device.isMobile  \
         0  not available in demo dataset           False
         1  not available in demo dataset            True
         2  not available in demo dataset           False
         3  not available in demo dataset           False
         4  not available in demo dataset           False


                       device.language    device.mobileDeviceBranding  \
         0  not available in demo dataset  not available in demo dataset
         1  not available in demo dataset  not available in demo dataset
         2  not available in demo dataset  not available in demo dataset
         3  not available in demo dataset  not available in demo dataset
         4  not available in demo dataset  not available in demo dataset


                  device.mobileDeviceInfo device.mobileDeviceMarketingName  \
         0  not available in demo dataset     not available in demo dataset
```

```
1  not available in demo dataset    not available in demo dataset
2  not available in demo dataset    not available in demo dataset
3  not available in demo dataset    not available in demo dataset
4  not available in demo dataset    not available in demo dataset


        device.mobileDeviceModel    device.mobileInputSelector  \
0  not available in demo dataset  not available in demo dataset
1  not available in demo dataset  not available in demo dataset
2  not available in demo dataset  not available in demo dataset
3  not available in demo dataset  not available in demo dataset
4  not available in demo dataset  not available in demo dataset


  device.operatingSystem  device.operatingSystemVersion  \
0                Windows  not available in demo dataset
1                    iOS  not available in demo dataset
2                Windows  not available in demo dataset
3                Windows  not available in demo dataset
4                Windows  not available in demo dataset


            device.screenColors        device.screenResolution  \
0  not available in demo dataset  not available in demo dataset
1  not available in demo dataset  not available in demo dataset
2  not available in demo dataset  not available in demo dataset
3  not available in demo dataset  not available in demo dataset
4  not available in demo dataset  not available in demo dataset


                geoNetwork.city              geoNetwork.cityId  \
0                          Izmir  not available in demo dataset
1                          Izmir  not available in demo dataset
2                       Istanbul  not available in demo dataset
3  not available in demo dataset  not available in demo dataset
4                       Istanbul  not available in demo dataset


  geoNetwork.continent geoNetwork.country              geoNetwork.latitude  \
0                 Asia             Turkey  not available in demo dataset
1                 Asia             Turkey  not available in demo dataset
2                 Asia             Turkey  not available in demo dataset
3                 Asia             Turkey  not available in demo dataset
4                 Asia             Turkey  not available in demo dataset


            geoNetwork.longitude                geoNetwork.metro  \
0  not available in demo dataset                        (not set)
1  not available in demo dataset                        (not set)
2  not available in demo dataset                        (not set)
3  not available in demo dataset  not available in demo dataset
4  not available in demo dataset                        (not set)


  geoNetwork.networkDomain      geoNetwork.networkLocation  \
```

```
0                ttnet.com.tr  not available in demo dataset
1            unknown.unknown  not available in demo dataset
2                  dr.com.tr  not available in demo dataset
3                ttnet.com.tr  not available in demo dataset
4                ttnet.com.tr  not available in demo dataset


                geoNetwork.region geoNetwork.subContinent totals.bounces  \
0                           Izmir            Western Asia              1
1                           Izmir            Western Asia              1
2                        Istanbul            Western Asia              1
3   not available in demo dataset            Western Asia              1
4                        Istanbul            Western Asia            NaN

  totals.hits totals.newVisits totals.pageviews totals.transactionRevenue  \
0           1                1                1                        NaN
1           1                1                1                        NaN
2           1                1                1                        NaN
3           1                1                1                        NaN
4           2                1                2                        NaN

  totals.visits trafficSource.adContent  \
0             1                     NaN
1             1                     NaN
2             1                     NaN
3             1                     NaN
4             1                     NaN

  trafficSource.adwordsClickInfo.adNetworkType  \
0                                           NaN
1                                           NaN
2                                           NaN
3                                           NaN
4                                           NaN

  trafficSource.adwordsClickInfo.criteriaParameters  \
0                 not available in demo dataset
1                 not available in demo dataset
2                 not available in demo dataset
3                 not available in demo dataset
4                 not available in demo dataset

  trafficSource.adwordsClickInfo.gclId  \
0                                   NaN
1                                   NaN
2                                   NaN
3                                   NaN
4                                   NaN
```

```
     trafficSource.adwordsClickInfo.isVideoAd  \
  0                                       NaN
  1                                       NaN
  2                                       NaN
  3                                       NaN
  4                                       NaN


     trafficSource.adwordsClickInfo.page trafficSource.adwordsClickInfo.slot  \
  0                                  NaN                                 NaN
  1                                  NaN                                 NaN
  2                                  NaN                                 NaN
  3                                  NaN                                 NaN
  4                                  NaN                                 NaN


     trafficSource.campaign trafficSource.campaignCode  \
  0             (not set)                          NaN
  1             (not set)                          NaN
  2             (not set)                          NaN
  3             (not set)                          NaN
  4             (not set)                          NaN


     trafficSource.isTrueDirect trafficSource.keyword trafficSource.medium  \
  0                        NaN       (not provided)              organic
  1                        NaN       (not provided)              organic
  2                        NaN       (not provided)              organic
  3                        NaN       (not provided)              organic
  4                        NaN       (not provided)              organic


     trafficSource.referralPath trafficSource.source         2017  marketType  \
  0                        NaN               google  26892871.0           1
  1                        NaN               google  26892871.0           1
  2                        NaN               google  26892871.0           1
  3                        NaN               google  26892871.0           1
  4                        NaN               google  26892871.0           1


      adType  browserType dayOfWeek
  0        1            1  Thursday
  1        1            1  Thursday
  2        1            1  Thursday
  3        1            1  Thursday
  4        1            1  Thursday

In [76]: train.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 903686 entries, 0 to 903685
Data columns (total 60 columns):
channelGrouping                                  903653 non-null object
```

```
date                                              903653 non-null float64
fullVisitorId                                     903653 non-null object
sessionId                                         903653 non-null object
socialEngagementType                              903653 non-null object
visitId                                           903653 non-null float64
visitNumber                                       903653 non-null float64
visitStartTime                                    903653 non-null float64
device.browser                                    903653 non-null object
device.browserSize                                903653 non-null object
device.browserVersion                             903653 non-null object
device.deviceCategory                             903653 non-null object
device.flashVersion                               903653 non-null object
device.isMobile                                   903653 non-null object
device.language                                   903653 non-null object
device.mobileDeviceBranding                       903653 non-null object
device.mobileDeviceInfo                           903653 non-null object
device.mobileDeviceMarketingName                  903653 non-null object
device.mobileDeviceModel                          903653 non-null object
device.mobileInputSelector                        903653 non-null object
device.operatingSystem                            903653 non-null object
device.operatingSystemVersion                     903653 non-null object
device.screenColors                               903653 non-null object
device.screenResolution                           903653 non-null object
geoNetwork.city                                   903653 non-null object
geoNetwork.cityId                                 903653 non-null object
geoNetwork.continent                              903653 non-null object
geoNetwork.country                                903653 non-null object
geoNetwork.latitude                               903653 non-null object
geoNetwork.longitude                              903653 non-null object
geoNetwork.metro                                  903653 non-null object
geoNetwork.networkDomain                          903653 non-null object
geoNetwork.networkLocation                        903653 non-null object
geoNetwork.region                                 903653 non-null object
geoNetwork.subContinent                           903653 non-null object
totals.bounces                                    450630 non-null object
totals.hits                                       903653 non-null object
totals.newVisits                                  703060 non-null object
totals.pageviews                                  903553 non-null object
totals.transactionRevenue                          11515 non-null object
totals.visits                                     903653 non-null object
trafficSource.adContent                            10946 non-null object
trafficSource.adwordsClickInfo.adNetworkType       21460 non-null object
trafficSource.adwordsClickInfo.criteriaParameters 903653 non-null object
trafficSource.adwordsClickInfo.gclId               21561 non-null object
trafficSource.adwordsClickInfo.isVideoAd           21460 non-null object
trafficSource.adwordsClickInfo.page                21460 non-null object
trafficSource.adwordsClickInfo.slot                21460 non-null object
trafficSource.campaign                            903653 non-null object
```

```
trafficSource.campaignCode                          1 non-null object
trafficSource.isTrueDirect                          274005 non-null object
trafficSource.keyword                               400724 non-null object
trafficSource.medium                                903653 non-null object
trafficSource.referralPath                          330941 non-null object
trafficSource.source                                903653 non-null object
2017                                                869164 non-null float64
marketType                                          903686 non-null int64
adType                                              903686 non-null int64
browserType                                         903686 non-null int64
dayOfWeek                                           903653 non-null object
dtypes: float64(5), int64(3), object(52)
memory usage: 420.6+ MB
```

```python
In [11]: set(train.columns).difference(set(test.columns))

Out[11]: {'totals.transactionRevenue', 'trafficSource.campaignCode'}

In [12]: cons_col = [i for i in train.columns if train[i].nunique(dropna=False)==1]
         cons_col

Out[12]: ['adType', 'browserType']

In [13]: train = train.drop(cons_col + ['trafficSource.campaignCode'], axis=1)
         test = test.drop(cons_col, axis=1)
         gc.collect()

Out[13]: 35

In [80]: train.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 903686 entries, 0 to 903685
Data columns (total 57 columns):
channelGrouping                                     903653 non-null object
date                                                903653 non-null float64
fullVisitorId                                       903653 non-null object
sessionId                                           903653 non-null object
socialEngagementType                                903653 non-null object
visitId                                             903653 non-null float64
visitNumber                                         903653 non-null float64
visitStartTime                                      903653 non-null float64
device.browser                                      903653 non-null object
device.browserSize                                  903653 non-null object
device.browserVersion                               903653 non-null object
device.deviceCategory                               903653 non-null object
device.flashVersion                                 903653 non-null object
device.isMobile                                     903653 non-null object
```

```
device.language                                      903653 non-null object
device.mobileDeviceBranding                          903653 non-null object
device.mobileDeviceInfo                              903653 non-null object
device.mobileDeviceMarketingName                     903653 non-null object
device.mobileDeviceModel                             903653 non-null object
device.mobileInputSelector                           903653 non-null object
device.operatingSystem                               903653 non-null object
device.operatingSystemVersion                        903653 non-null object
device.screenColors                                  903653 non-null object
device.screenResolution                              903653 non-null object
geoNetwork.city                                      903653 non-null object
geoNetwork.cityId                                    903653 non-null object
geoNetwork.continent                                 903653 non-null object
geoNetwork.country                                   903653 non-null object
geoNetwork.latitude                                  903653 non-null object
geoNetwork.longitude                                 903653 non-null object
geoNetwork.metro                                     903653 non-null object
geoNetwork.networkDomain                             903653 non-null object
geoNetwork.networkLocation                           903653 non-null object
geoNetwork.region                                    903653 non-null object
geoNetwork.subContinent                              903653 non-null object
totals.bounces                                       450630 non-null object
totals.hits                                          903653 non-null object
totals.newVisits                                     703060 non-null object
totals.pageviews                                     903553 non-null object
totals.transactionRevenue                            11515 non-null object
totals.visits                                        903653 non-null object
trafficSource.adContent                              10946 non-null object
trafficSource.adwordsClickInfo.adNetworkType         21460 non-null object
trafficSource.adwordsClickInfo.criteriaParameters    903653 non-null object
trafficSource.adwordsClickInfo.gclId                 21561 non-null object
trafficSource.adwordsClickInfo.isVideoAd             21460 non-null object
trafficSource.adwordsClickInfo.page                  21460 non-null object
trafficSource.adwordsClickInfo.slot                  21460 non-null object
trafficSource.campaign                               903653 non-null object
trafficSource.isTrueDirect                           274005 non-null object
trafficSource.keyword                                400724 non-null object
trafficSource.medium                                 903653 non-null object
trafficSource.referralPath                           330941 non-null object
trafficSource.source                                 903653 non-null object
2017                                                 869164 non-null float64
marketType                                           903686 non-null int64
dayOfWeek                                            903653 non-null object
dtypes: float64(5), int64(1), object(51)
memory usage: 399.9+ MB
```

```python
In [14]: def find_missing(data):
```

```
            count_missing = data.isnull().sum().values
            total = data.shape[0]
            ratio_missing = count_missing/total
            return pd.DataFrame(data={'missing_count':count_missing, 'missing_ratio':ratio_mis
        train_missing = find_missing(train)
        test_missing = find_missing(test)

In [15]: train_missing.reset_index()[['index', 'missing_ratio']]\
        .merge(test_missing.reset_index()[['index', 'missing_ratio']], on='index', how='le
        .rename(columns={'index':'columns', 'missing_ratio_x':'train_missing_ratio', 'miss
        .sort_values(['train_missing_ratio', 'test_missing_ratio'], ascending=False)\
        .query('train_missing_ratio>0')

Out[15]:                                                 columns  train_missing_ratio  \
        41                            trafficSource.adContent             0.987887
        39                          totals.transactionRevenue             0.987258
        42        trafficSource.adwordsClickInfo.adNetworkType             0.976253
        45           trafficSource.adwordsClickInfo.isVideoAd             0.976253
        46                trafficSource.adwordsClickInfo.page             0.976253
        47                trafficSource.adwordsClickInfo.slot             0.976253
        44              trafficSource.adwordsClickInfo.gclId             0.976141
        49                          trafficSource.isTrueDirect             0.696792
        52                          trafficSource.referralPath             0.633788
        50                             trafficSource.keyword             0.556567
        35                                     totals.bounces             0.501342
        37                                   totals.newVisits             0.222009
        54                                               2017             0.038201
        38                                   totals.pageviews             0.000147
        0                                     channelGrouping             0.000037
        1                                                date             0.000037
        2                                        fullVisitorId             0.000037
        3                                           sessionId             0.000037
        4                                 socialEngagementType             0.000037
        5                                             visitId             0.000037
        6                                         visitNumber             0.000037
        7                                       visitStartTime             0.000037
        8                                       device.browser             0.000037
        9                                   device.browserSize             0.000037
        10                               device.browserVersion             0.000037
        11                               device.deviceCategory             0.000037
        12                                  device.flashVersion             0.000037
        13                                     device.isMobile             0.000037
        14                                     device.language             0.000037
        15                          device.mobileDeviceBranding             0.000037
        16                              device.mobileDeviceInfo             0.000037
        17                      device.mobileDeviceMarketingName             0.000037
        18                             device.mobileDeviceModel             0.000037
        19                           device.mobileInputSelector             0.000037
```

20	device.operatingSystem	0.000037
21	device.operatingSystemVersion	0.000037
22	device.screenColors	0.000037
23	device.screenResolution	0.000037
24	geoNetwork.city	0.000037
25	geoNetwork.cityId	0.000037
26	geoNetwork.continent	0.000037
27	geoNetwork.country	0.000037
28	geoNetwork.latitude	0.000037
29	geoNetwork.longitude	0.000037
30	geoNetwork.metro	0.000037
31	geoNetwork.networkDomain	0.000037
32	geoNetwork.networkLocation	0.000037
33	geoNetwork.region	0.000037
34	geoNetwork.subContinent	0.000037
36	totals.hits	0.000037
40	totals.visits	0.000037
43	trafficSource.adwordsClickInfo.criteriaParameters	0.000037
48	trafficSource.campaign	0.000037
51	trafficSource.medium	0.000037
53	trafficSource.source	0.000037
56	dayOfWeek	0.000037

	test_missing_ratio
41	0.933155
39	NaN
42	0.933127
45	0.933127
46	0.933127
47	0.933127
44	0.933067
49	0.676267
52	0.707570
50	0.485965
35	0.476899
37	0.248965
54	0.039518
38	0.000212
0	0.000040
1	0.000040
2	0.000040
3	0.000040
4	0.000040
5	0.000040
6	0.000040
7	0.000040
8	0.000040
9	0.000040

```
10           0.000040
11           0.000040
12           0.000040
13           0.000040
14           0.000040
15           0.000040
16           0.000040
17           0.000040
18           0.000040
19           0.000040
20           0.000040
21           0.000040
22           0.000040
23           0.000040
24           0.000040
25           0.000040
26           0.000040
27           0.000040
28           0.000040
29           0.000040
30           0.000040
31           0.000040
32           0.000040
33           0.000040
34           0.000040
36           0.000040
40           0.000040
43           0.000040
48           0.000040
51           0.000040
53           0.000040
56           0.000040
```

```
In [16]: train_idx = train.fullVisitorId
         test_idx = test.fullVisitorId
         train["totals.transactionRevenue"] = train["totals.transactionRevenue"].astype('float
         train_y = train["totals.transactionRevenue"]
         train_target = np.log1p(train.groupby("fullVisitorId")["totals.transactionRevenue"].su
```

```
In [17]: train.drop(['fullVisitorId', 'sessionId', 'visitId'], axis = 1, inplace = True)
         test.drop(['fullVisitorId', 'sessionId', 'visitId'], axis = 1, inplace = True)
         num_col = ["totals.hits", "totals.pageviews", "visitNumber", "visitStartTime", 'totals
         for i in num_col:
             train[i] = train[i].astype('float').fillna(0)
             test[i] = test[i].astype('float').fillna(0)
         cat_col = [e for e in train.columns.tolist() if e not in num_col]
         cat_col.remove('date')
         cat_col.remove('totals.transactionRevenue')
```

```
for i in cat_col:
    lab_en = LabelEncoder()
    train[i] = train[i].fillna('not known')
    test[i] = test[i].fillna('not known')
    lab_en.fit(list(train[i].astype('str')) + list(test[i].astype('str')))
    train[i] = lab_en.transform(list(train[i].astype('str')))
    test[i] = lab_en.transform(test[i].astype('str'))
```

3 Probability Function

```
In [96]: prob_df = pd.DataFrame()
         prob_df["fullVisitorId"] = train_idx
         prob_df["visitNumber"] = train["visitNumber"]

         top = prob_df.groupby("fullVisitorId")["visitNumber"].sum().reset_index()
         total_count = top["visitNumber"].sum()
         top["probability_visitNumber"] = top["visitNumber"]/total_count
         sorted_top1 = top.sort_values("probability_visitNumber", ascending=False)

         sorted_top1["rank_vn"] = sorted_top1["probability_visitNumber"].rank(ascending=False)

         print(sorted_top1.head(20))
```

	fullVisitorId	visitNumber	probability_visitNumber	rank_vn
58716	0824839726118485274	63315.0	0.030935	1.0
139589	1957458976293878100	50447.0	0.024648	2.0
257343	3608475193341679870	36988.0	0.018072	3.0
132193	1856749147915772585	36688.0	0.017926	4.0
545156	7634897085866546110	14812.0	0.007237	5.0
233127	3269834865385146569	13793.0	0.006739	6.0
350356	4913801338365738862	12943.0	0.006324	7.0
148455	2082625651279391786	12828.0	0.006268	8.0
51304	0720311197761340948	11781.0	0.005756	9.0
16720	0232377434237234751	11047.0	0.005398	10.0
287807	4038076683036146727	10833.0	0.005293	11.0
67614	0949718915643445721	10432.0	0.005097	12.0
263443	3694234028523165868	9056.0	0.004425	13.0
314173	4405445121320750966	7967.0	0.003893	14.0
463962	650160476379342922	7213.0	0.003524	15.0
446231	6254908847172458133	6820.0	0.003332	16.0
429515	6018775317735347795	6768.0	0.003307	17.0
251329	3525537916960843419	6670.0	0.003259	18.0
557953	7813149961404844386	6451.0	0.003152	19.0
224558	3148617623907142276	6328.0	0.003092	20.0

```
In [99]: prob_df["totals.pageviews"] = train["totals.pageviews"]
         top = prob_df.groupby("fullVisitorId")["totals.pageviews"].sum().reset_index()
```

```python
    total_count = top["totals.pageviews"].sum()
    top["probability_pv"] = top["totals.pageviews"]/total_count
    sorted_top2 = top.sort_values("probability_pv", ascending=False)
    sorted_top2["rank_pv"] = sorted_top2["totals.pageviews"].rank(ascending=False)
    print(sorted_top2.head(20))
```

	fullVisitorId	totals.pageviews	probability_pv	rank_pv
139589	1957458976293878100	4160.0	0.001196	1.0
58716	0824839726118485274	2707.0	0.000778	2.0
132193	1856749147915772585	1715.0	0.000493	3.0
706656	9894955795481014038	1496.0	0.000430	4.0
545156	7634897085866546110	1006.0	0.000289	5.0
350356	4913801338365738862	858.0	0.000247	6.0
371377	5208937953046059083	856.0	0.000246	7.0
16720	0232377434237234751	851.0	0.000245	8.0
156451	2194592743396253647	823.0	0.000237	9.0
557953	7813149961404844386	815.0	0.000234	10.0
534288	7483600664917507409	785.0	0.000226	11.0
585497	8197879643797712877	782.0	0.000225	12.0
521958	7311242886083854158	773.0	0.000222	13.0
344724	4835082938415020542	771.0	0.000222	14.0
524342	7344519175727343086	760.0	0.000218	15.5
482512	6760732402251466726	760.0	0.000218	15.5
398871	5590743844926892757	733.0	0.000211	17.0
139507	1956307607572137989	710.0	0.000204	18.0
530500	743123551680199202	694.0	0.000200	19.0
669658	9377429831454005466	685.0	0.000197	20.0

```python
In [100]: prob_df["totals.hits"] = train["totals.hits"]
          top = prob_df.groupby("fullVisitorId")["totals.hits"].sum().reset_index()
          total_count = top["totals.hits"].sum()
          top["probability_hits"] = top["totals.hits"]/total_count
          sorted_top3 = top.sort_values("probability_hits", ascending=False)
          sorted_top3["rank_hits"] = sorted_top3["totals.hits"].rank(ascending=False)
          print(sorted_top3.head(20))
```

	fullVisitorId	totals.hits	probability_hits	rank_hits
139589	1957458976293878100	4995.0	0.001203	1.0
58716	0824839726118485274	3378.0	0.000813	2.0
132193	1856749147915772585	2174.0	0.000523	3.0
706656	9894955795481014038	1602.0	0.000386	4.0
545156	7634897085866546110	1271.0	0.000306	5.0
534288	7483600664917507409	1091.0	0.000263	6.0
16720	0232377434237234751	1067.0	0.000257	7.0
350356	4913801338365738862	1047.0	0.000252	8.0
585497	8197879643797712877	1046.0	0.000252	9.0
521958	7311242886083854158	1040.0	0.000250	10.0

```
398871   5590743844926892757          993.0            0.000239          11.0
482512   6760732402251466726          989.0            0.000238          12.0
557953   7813149961404844386          986.0            0.000237          13.0
156451   2194592743396253647          978.0            0.000235          14.0
139507   1956307607572137989          964.0            0.000232          15.0
344724   4835082938415020542          962.0            0.000232          16.0
371377   5208937953046059083          934.0            0.000225          17.0
669658   9377429831454005466          918.0            0.000221          18.0
129164   1814166460229302850          916.0            0.000221          19.0
649241   9089132392240687728          909.0            0.000219          20.0
```

In [109]: `prob_ranking_merged = pd.merge(sorted_top1.head(50), sorted_top2.head(50), left_on=[`

`prob_ranking_merged = pd.merge(prob_ranking_merged, sorted_top3.head(50), left_on=[':`

`#prob_ranking_merged["rank_vn"] = prob_ranking_merged["rank_vn"].fillna(0)`
`#prob_ranking_merged["rank_pv"] = prob_ranking_merged["rank_pv"].fillna(0)`
`#prob_ranking_merged["rank_hits"] = prob_ranking_merged["rank_hits"].fillna(0)`

`prob_ranking_merged["total_rank"] = prob_ranking_merged["rank_vn"] + prob_ranking_mei`

`prob_ranking_merged.sort_values("total_rank", ascending=True)`

Out[109]:
```
         fullVisitorId  visitNumber  probability_visitNumber  rank_vn  \
1   1957458976293878100      50447.0                 0.024648      2.0
0   0824839726118485274      63315.0                 0.030935      1.0
3   1856749147915772585      36688.0                 0.017926      4.0
4   7634897085866546110      14812.0                 0.007237      5.0
6   4913801338365738862      12943.0                 0.006324      7.0
9   0232377434237234751      11047.0                 0.005398     10.0
18  7813149961404844386       6451.0                 0.003152     19.0
21  2194592743396253647       5717.0                 0.002793     22.0
44  5208937953046059083       2526.0                 0.001234     45.0
15  6254908847172458133       6820.0                 0.003332     16.0
12  3694234028523165868       9056.0                 0.004425     13.0
31  3041133261614133977       3663.0                 0.001790     32.0
22  7498695963354635199       5650.0                 0.002761     23.0
36  4578640586284138624       3189.0                 0.001558     37.0
2   3608475193341679870      36988.0                 0.018072      3.0
5   3269834865385146569      13793.0                 0.006739      6.0
7   2082625651279391786      12828.0                 0.006268      8.0
8   0720311197761340948      11781.0                 0.005756      9.0
10  4038076683036146727      10833.0                 0.005293     11.0
11  0949718915643445721      10432.0                 0.005097     12.0
13  4405445121320750966       7967.0                 0.003893     14.0
14   650160476379342922       7213.0                 0.003524     15.0
16  6018775317735347795       6768.0                 0.003307     17.0
```

17	3525537916960843419	6670.0	0.003259	18.0
19	3148617623907142276	6328.0	0.003092	20.0
20	9801276214964695322	6240.0	0.003049	21.0
23	680206014763278315	5565.0	0.002719	24.0
24	834799491548640007	5438.0	0.002657	25.0
25	9609104828919391966	5133.0	0.002508	26.0
26	3937673380007666721	4551.0	0.002224	27.0
..	...	...	...	...
58	1956307607572137989	NaN	NaN	NaN
59	743123551680199202	NaN	NaN	NaN
60	9377429831454005466	NaN	NaN	NaN
61	9089132392240687728	NaN	NaN	NaN
62	6010250598436085923	NaN	NaN	NaN
63	1814166460229302850	NaN	NaN	NaN
64	315017261115039181	NaN	NaN	NaN
65	4309363468347582056	NaN	NaN	NaN
66	2446685875964479851	NaN	NaN	NaN
67	741993131378886687	NaN	NaN	NaN
68	9029794295932939024	NaN	NaN	NaN
69	2402527199731150932	NaN	NaN	NaN
70	7530866178634633311	NaN	NaN	NaN
71	3921649958751416379	NaN	NaN	NaN
72	7435052038545127483	NaN	NaN	NaN
73	4984366501121503466	NaN	NaN	NaN
74	7463172420271311409	NaN	NaN	NaN
75	8851822767477687842	NaN	NaN	NaN
76	7477634267430924328	NaN	NaN	NaN
77	8436426603099391262	NaN	NaN	NaN
78	1759779356678951325	NaN	NaN	NaN
79	2195063960372896546	NaN	NaN	NaN
80	596895434219823695	NaN	NaN	NaN
81	419161784882208907	NaN	NaN	NaN
82	6091498163535424271	NaN	NaN	NaN
83	7737049103084491745	NaN	NaN	NaN
84	2351329739798371879	NaN	NaN	NaN
85	7809643517607858003	NaN	NaN	NaN
86	9417857471295131045	NaN	NaN	NaN
87	0804281778755290192	NaN	NaN	NaN

	totals.pageviews	probability_pv	rank_pv	totals.hits	probability_hits	\
1	4160.0	0.001196	1.0	4995.0	0.001203	
0	2707.0	0.000778	2.0	3378.0	0.000813	
3	1715.0	0.000493	3.0	2174.0	0.000523	
4	1006.0	0.000289	5.0	1271.0	0.000306	
6	858.0	0.000247	6.0	1047.0	0.000252	
9	851.0	0.000245	8.0	1067.0	0.000257	
18	815.0	0.000234	10.0	986.0	0.000237	
21	823.0	0.000237	9.0	978.0	0.000235	

44	856.0	0.000246	7.0	934.0	0.000225
15	619.0	0.000178	25.0	752.0	0.000181
12	587.0	0.000169	29.0	709.0	0.000171
31	576.0	0.000166	31.0	707.0	0.000170
22	546.0	0.000157	36.0	680.0	0.000164
36	527.0	0.000152	40.0	659.0	0.000159
2	NaN	NaN	NaN	NaN	NaN
5	NaN	NaN	NaN	NaN	NaN
7	NaN	NaN	NaN	NaN	NaN
8	NaN	NaN	NaN	NaN	NaN
10	NaN	NaN	NaN	NaN	NaN
11	NaN	NaN	NaN	NaN	NaN
13	NaN	NaN	NaN	NaN	NaN
14	NaN	NaN	NaN	NaN	NaN
16	NaN	NaN	NaN	NaN	NaN
17	NaN	NaN	NaN	NaN	NaN
19	NaN	NaN	NaN	NaN	NaN
20	NaN	NaN	NaN	NaN	NaN
23	NaN	NaN	NaN	NaN	NaN
24	NaN	NaN	NaN	NaN	NaN
25	579.0	0.000166	30.0	NaN	NaN
26	NaN	NaN	NaN	NaN	NaN
..	...	...	...	...	...
58	710.0	0.000204	18.0	964.0	0.000232
59	694.0	0.000200	19.0	749.0	0.000180
60	685.0	0.000197	20.0	918.0	0.000221
61	670.0	0.000193	21.0	909.0	0.000219
62	662.0	0.000190	22.0	886.0	0.000213
63	640.0	0.000184	23.0	916.0	0.000221
64	621.0	0.000179	24.0	868.0	0.000209
65	609.0	0.000175	26.0	908.0	0.000219
66	598.0	0.000172	27.0	795.0	0.000191
67	591.0	0.000170	28.0	828.0	0.000199
68	574.0	0.000165	32.0	790.0	0.000190
69	559.0	0.000161	33.0	753.0	0.000181
70	557.0	0.000160	34.0	691.0	0.000166
71	553.0	0.000159	35.0	753.0	0.000181
72	544.0	0.000156	37.0	702.0	0.000169
73	540.0	0.000155	38.0	724.0	0.000174
74	539.0	0.000155	39.0	717.0	0.000173
75	526.0	0.000151	41.0	720.0	0.000173
76	513.0	0.000147	42.0	712.0	0.000171
77	501.0	0.000144	43.0	714.0	0.000172
78	499.0	0.000143	44.0	688.0	0.000166
79	498.0	0.000143	45.0	729.0	0.000176
80	497.0	0.000143	46.0	NaN	NaN
81	491.0	0.000141	47.0	694.0	0.000167
82	489.0	0.000141	48.0	691.0	0.000166

83	488.0	0.000140	49.0	NaN	NaN
84	477.0	0.000137	50.0	738.0	0.000178
85	NaN	NaN	NaN	813.0	0.000196
86	NaN	NaN	NaN	696.0	0.000168
87	NaN	NaN	NaN	664.0	0.000160

	rank_hits	total_rank
1	1.0	4.0
0	2.0	5.0
3	3.0	10.0
4	5.0	15.0
6	8.0	21.0
9	7.0	25.0
18	13.0	42.0
21	14.0	45.0
44	17.0	69.0
15	31.0	72.0
12	40.0	82.0
31	41.0	104.0
22	48.0	107.0
36	50.0	127.0
2	NaN	NaN
5	NaN	NaN
7	NaN	NaN
8	NaN	NaN
10	NaN	NaN
11	NaN	NaN
13	NaN	NaN
14	NaN	NaN
16	NaN	NaN
17	NaN	NaN
19	NaN	NaN
20	NaN	NaN
23	NaN	NaN
24	NaN	NaN
25	NaN	NaN
26	NaN	NaN
..	...	...
58	15.0	NaN
59	32.0	NaN
60	18.0	NaN
61	20.0	NaN
62	22.0	NaN
63	19.0	NaN
64	23.0	NaN
65	21.0	NaN
66	27.0	NaN
67	25.0	NaN

```
68      28.0        NaN
69      29.5        NaN
70      45.5        NaN
71      29.5        NaN
72      42.0        NaN
73      35.0        NaN
74      37.0        NaN
75      36.0        NaN
76      39.0        NaN
77      38.0        NaN
78      47.0        NaN
79      34.0        NaN
80       NaN        NaN
81      44.0        NaN
82      45.5        NaN
83       NaN        NaN
84      33.0        NaN
85      26.0        NaN
86      43.0        NaN
87      49.0        NaN

[88 rows x 11 columns]
```

```python
In [20]: y_train = np.log1p(train["totals.transactionRevenue"])
         x_train = train.drop(["totals.transactionRevenue"], axis=1)
         x_test = test.copy()
```

4 Train

```python
In [89]: folds = KFold(n_splits=5,random_state=6)
         oof_preds = np.zeros(x_train.shape[0])
         sub_preds = np.zeros(x_test.shape[0])

         start = time.time()
         valid_score = 0
         for n_fold, (trn_idx, val_idx) in enumerate(folds.split(x_train, y_train)):
             trn_x, trn_y = x_train.iloc[trn_idx], y_train[trn_idx]
             val_x, val_y = x_train.iloc[val_idx], y_train[val_idx]

             train_data = lgb.Dataset(data=trn_x, label=trn_y)
             valid_data = lgb.Dataset(data=val_x, label=val_y)

             params = {"objective" : "regression", "metric" : "rmse", 'n_estimators':10000, 'ea
                       "num_leaves" : 30, "learning_rate" : 0.01, "bagging_fraction" : 0.9,
                       "feature_fraction" : 0.3, "bagging_seed" : 0}

             lgb_model = lgb.train(params, train_data, valid_sets=[train_data, valid_data], ver
```

```python
            oof_preds[val_idx] = lgb_model.predict(val_x, num_iteration=lgb_model.best_iterat
            oof_preds[oof_preds<0] = 0
            sub_pred = lgb_model.predict(x_test, num_iteration=lgb_model.best_iteration) / fol
            sub_pred[sub_pred<0] = 0
            sub_preds += sub_pred
            print('Fold %2d RMSE : %.6f' % (n_fold + 1, np.sqrt(mean_squared_error(val_y, oof_
            valid_score += np.sqrt(mean_squared_error(val_y, oof_preds[val_idx]))
```

```
/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:113: UserWarning: Found `n_estimators
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))
/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:118: UserWarning: Found `early_stoppi
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))


Training until validation scores don't improve for 100 rounds.
Early stopping, best iteration is:
[457]         training's rmse: 1.85029        valid_1's rmse: 0.324253
Fold  1 RMSE : 0.323986


/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:113: UserWarning: Found `n_estimators
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))
/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:118: UserWarning: Found `early_stoppi
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))


Training until validation scores don't improve for 100 rounds.
[1000]        training's rmse: 1.77835        valid_1's rmse: 0.683013
Early stopping, best iteration is:
[1357]        training's rmse: 1.76207        valid_1's rmse: 0.682217
Fold  2 RMSE : 0.681136


/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:113: UserWarning: Found `n_estimators
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))
/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:118: UserWarning: Found `early_stoppi
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))


Training until validation scores don't improve for 100 rounds.
[1000]        training's rmse: 1.37219        valid_1's rmse: 2.38438
[2000]        training's rmse: 1.33423        valid_1's rmse: 2.36218
[3000]        training's rmse: 1.30862        valid_1's rmse: 2.35449
[4000]        training's rmse: 1.28711        valid_1's rmse: 2.34935
[5000]        training's rmse: 1.26936        valid_1's rmse: 2.34646
Early stopping, best iteration is:
[5089]        training's rmse: 1.26775        valid_1's rmse: 2.34623
Fold  3 RMSE : 2.344776
```

```
/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:113: UserWarning: Found `n_estimators
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))
/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:118: UserWarning: Found `early_stoppi
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))


Training until validation scores don't improve for 100 rounds.
[1000]        training's rmse: 1.31386        valid_1's rmse: 2.51139
[2000]        training's rmse: 1.2759         valid_1's rmse: 2.49224
Early stopping, best iteration is:
[2765]        training's rmse: 1.2554         valid_1's rmse: 2.4862
Fold  4 RMSE : 2.485008


/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:113: UserWarning: Found `n_estimators
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))
/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:118: UserWarning: Found `early_stoppi
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))


Training until validation scores don't improve for 100 rounds.
[1000]        training's rmse: 1.73758        valid_1's rmse: 1.03095
[2000]        training's rmse: 1.70023        valid_1's rmse: 1.02036
[3000]        training's rmse: 1.67283        valid_1's rmse: 1.01683
Early stopping, best iteration is:
[3074]        training's rmse: 1.67148        valid_1's rmse: 1.01679
Fold  5 RMSE : 1.015274
```

```python
In [90]: print('Session-level CV-score:', str(round(valid_score/folds.n_splits,4)))
         print(' ')
         train_pred = pd.DataFrame({"fullVisitorId":train_idx})
         train_pred["PredictedLogRevenue"] = np.expm1(oof_preds)
         train_pred = train_pred.groupby("fullVisitorId")["PredictedLogRevenue"].sum().reset_in
         train_pred.columns = ["fullVisitorId", "PredictedLogRevenue"]
         train_pred["PredictedLogRevenue"] = np.log1p(train_pred["PredictedLogRevenue"])
         train_rmse = np.sqrt(mean_squared_error(train_target, train_pred['PredictedLogRevenue
         print('User-level score:', str(round(train_rmse, 4)))
         print(' ')
         end = time.time()
         print('training time:', str(round((end - start)/60)), 'mins')
```

```
Session-level CV-score: 1.37


User-level score: 1.6123


training time: 10 mins
```

```
In [91]: test_pred = pd.DataFrame({"fullVisitorId":test_idx})
         test_pred["PredictedLogRevenue"] = np.expm1(sub_preds)
         test_pred = test_pred.groupby("fullVisitorId")["PredictedLogRevenue"].sum().reset_inde
         test_pred.columns = ["fullVisitorId", "PredictedLogRevenue"]
         test_pred["PredictedLogRevenue"] = np.log1p(test_pred["PredictedLogRevenue"])
         test_pred.to_csv("submit.csv", index=False)

In [92]: lgb.plot_importance(lgb_model, height=0.5, max_num_features=20, ignore_zero = False,
         plt.show()
```


Feature importance
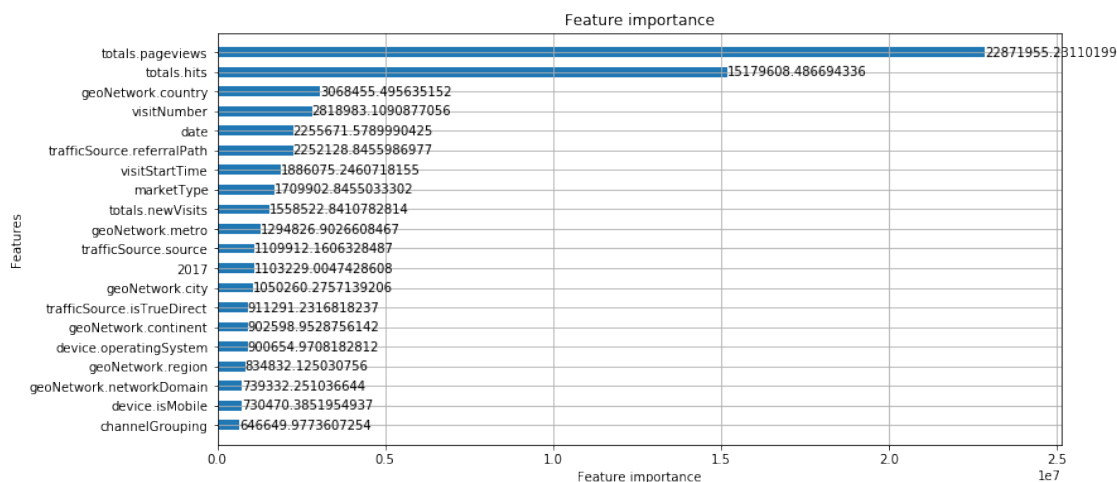
```
In [93]: train.corr(method='pearson')
```

Out[93]:

	channelGrouping	date
channelGrouping	1.000000	-0.236886
date	-0.236886	1.000000
socialEngagementType	0.011466	NaN
visitNumber	-0.036800	0.007536
visitStartTime	-0.164112	0.880936
device.browser	0.100624	-0.088072
device.browserSize	0.011466	NaN
device.browserVersion	0.011466	NaN
device.deviceCategory	-0.186820	0.118133
device.flashVersion	0.011466	NaN
device.isMobile	-0.230171	0.146148
device.language	0.011466	NaN
device.mobileDeviceBranding	0.011466	NaN
device.mobileDeviceInfo	0.011466	NaN
device.mobileDeviceMarketingName	0.011466	NaN
device.mobileDeviceModel	0.011466	NaN
device.mobileInputSelector	0.011466	NaN
device.operatingSystem	0.030185	0.001192

44

device.operatingSystemVersion	0.011466	NaN
device.screenColors	0.011466	NaN
device.screenResolution	0.011466	NaN
geoNetwork.city	-0.006282	0.051625
geoNetwork.cityId	0.011466	NaN
geoNetwork.continent	0.038730	0.007571
geoNetwork.country	-0.045936	-0.008117
geoNetwork.latitude	0.011466	NaN
geoNetwork.longitude	0.011466	NaN
geoNetwork.metro	-0.014529	0.062643
geoNetwork.networkDomain	0.110806	-0.077520
geoNetwork.networkLocation	0.011466	NaN
geoNetwork.region	0.017590	0.032589
geoNetwork.subContinent	0.062654	-0.040016
totals.bounces	0.063994	0.002479
totals.hits	-0.067364	-0.018395
totals.newVisits	0.082264	-0.043902
totals.pageviews	-0.071030	-0.012146
totals.transactionRevenue	-0.006646	0.003188
totals.visits	0.011466	NaN
trafficSource.adContent	-0.021286	-0.051095
trafficSource.adwordsClickInfo.adNetworkType	-0.032302	-0.018888
trafficSource.adwordsClickInfo.criteriaParameters	0.011466	NaN
trafficSource.adwordsClickInfo.gclId	-0.028202	-0.005973
trafficSource.adwordsClickInfo.isVideoAd	-0.032303	-0.018867
trafficSource.adwordsClickInfo.page	-0.032293	-0.018907
trafficSource.adwordsClickInfo.slot	-0.031947	-0.018385
trafficSource.campaign	-0.201301	0.013333
trafficSource.isTrueDirect	0.467864	-0.083313
trafficSource.keyword	0.282834	-0.158028
trafficSource.medium	0.802630	-0.122398
trafficSource.referralPath	-0.509481	0.024930
trafficSource.source	0.939053	-0.262463
2017	-0.224743	0.114128
marketType	0.215656	-0.085404
dayOfWeek	0.011466	NaN

	socialEngagementType \
channelGrouping	0.011466
date	NaN
socialEngagementType	1.000000
visitNumber	-0.001474
visitStartTime	-0.705206
device.browser	0.026293
device.browserSize	1.000000
device.browserVersion	1.000000
device.deviceCategory	0.015478
device.flashVersion	1.000000

```
device.isMobile                                            0.023766
device.language                                            1.000000
device.mobileDeviceBranding                                1.000000
device.mobileDeviceInfo                                    1.000000
device.mobileDeviceMarketingName                           1.000000
device.mobileDeviceModel                                   1.000000
device.mobileInputSelector                                 1.000000
device.operatingSystem                                     0.008183
device.operatingSystemVersion                              1.000000
device.screenColors                                        1.000000
device.screenResolution                                    1.000000
geoNetwork.city                                            0.004305
geoNetwork.cityId                                          1.000000
geoNetwork.continent                                       0.022398
geoNetwork.country                                         0.005586
geoNetwork.latitude                                        1.000000
geoNetwork.longitude                                       1.000000
geoNetwork.metro                                           0.004434
geoNetwork.networkDomain                                   0.002657
geoNetwork.networkLocation                                 1.000000
geoNetwork.region                                          0.004551
geoNetwork.subContinent                                    0.012515
totals.bounces                                            -0.006027
totals.hits                                               -0.002881
totals.newVisits                                          -0.011313
totals.pageviews                                          -0.003311
totals.transactionRevenue                                 -0.000195
totals.visits                                              1.000000
trafficSource.adContent                                    0.000622
trafficSource.adwordsClickInfo.adNetworkType               0.000942
trafficSource.adwordsClickInfo.criteriaParameters          1.000000
trafficSource.adwordsClickInfo.gclId                       0.000823
trafficSource.adwordsClickInfo.isVideoAd                   0.000942
trafficSource.adwordsClickInfo.page                        0.000942
trafficSource.adwordsClickInfo.slot                        0.000932
trafficSource.campaign                                     0.117426
trafficSource.isTrueDirect                                 0.003986
trafficSource.keyword                                      0.005269
trafficSource.medium                                      -0.000602
trafficSource.referralPath                                 0.003173
trafficSource.source                                       0.004904
2017                                                      -0.005053
marketType                                                 0.005273
dayOfWeek                                                  1.000000

                                           visitNumber  \
channelGrouping                               -0.036800
date                                           0.007536
```

socialEngagementType	-0.001474
visitNumber	1.000000
visitStartTime	0.002506
device.browser	-0.042104
device.browserSize	-0.001474
device.browserVersion	-0.001474
device.deviceCategory	-0.032287
device.flashVersion	-0.001474
device.isMobile	-0.037691
device.language	-0.001474
device.mobileDeviceBranding	-0.001474
device.mobileDeviceInfo	-0.001474
device.mobileDeviceMarketingName	-0.001474
device.mobileDeviceModel	-0.001474
device.mobileInputSelector	-0.001474
device.operatingSystem	0.007164
device.operatingSystemVersion	-0.001474
device.screenColors	-0.001474
device.screenResolution	-0.001474
geoNetwork.city	0.003616
geoNetwork.cityId	-0.001474
geoNetwork.continent	-0.064838
geoNetwork.country	0.052182
geoNetwork.latitude	-0.001474
geoNetwork.longitude	-0.001474
geoNetwork.metro	0.014937
geoNetwork.networkDomain	-0.046552
geoNetwork.networkLocation	-0.001474
geoNetwork.region	-0.003238
geoNetwork.subContinent	-0.025541
totals.bounces	-0.023745
totals.hits	0.041321
totals.newVisits	-0.255044
totals.pageviews	0.043337
totals.transactionRevenue	0.051367
totals.visits	-0.001474
trafficSource.adContent	-0.033052
trafficSource.adwordsClickInfo.adNetworkType	-0.001335
trafficSource.adwordsClickInfo.criteriaParameters	-0.001474
trafficSource.adwordsClickInfo.gclId	0.000322
trafficSource.adwordsClickInfo.isVideoAd	-0.001333
trafficSource.adwordsClickInfo.page	-0.001346
trafficSource.adwordsClickInfo.slot	-0.000960
trafficSource.campaign	-0.004429
trafficSource.isTrueDirect	-0.181696
trafficSource.keyword	0.003198
trafficSource.medium	-0.033188
trafficSource.referralPath	-0.029875

```
trafficSource.source                                      -0.046518
2017                                                       0.048679
marketType                                                -0.096729
dayOfWeek                                                 -0.001474


                                                     visitStartTime  \
channelGrouping                                           -0.164112
date                                                       0.880936
socialEngagementType                                      -0.705206
visitNumber                                                0.002506
visitStartTime                                             1.000000
device.browser                                            -0.066757
device.browserSize                                        -0.705206
device.browserVersion                                     -0.705206
device.deviceCategory                                      0.071518
device.flashVersion                                       -0.705206
device.isMobile                                            0.085543
device.language                                           -0.705206
device.mobileDeviceBranding                               -0.705206
device.mobileDeviceInfo                                   -0.705206
device.mobileDeviceMarketingName                          -0.705206
device.mobileDeviceModel                                  -0.705206
device.mobileInputSelector                                -0.705206
device.operatingSystem                                    -0.008923
device.operatingSystemVersion                             -0.705206
device.screenColors                                       -0.705206
device.screenResolution                                   -0.705206
geoNetwork.city                                            0.017869
geoNetwork.cityId                                         -0.705206
geoNetwork.continent                                      -0.014863
geoNetwork.country                                        -0.004350
geoNetwork.latitude                                       -0.705206
geoNetwork.longitude                                      -0.705206
geoNetwork.metro                                           0.025625
geoNetwork.networkDomain                                  -0.056860
geoNetwork.networkLocation                                -0.705206
geoNetwork.region                                          0.007656
geoNetwork.subContinent                                   -0.034235
totals.bounces                                             0.004158
totals.hits                                               -0.017608
totals.newVisits                                          -0.021115
totals.pageviews                                          -0.013448
totals.transactionRevenue                                 0.002069
totals.visits                                            -0.705206
trafficSource.adContent                                   -0.033411
trafficSource.adwordsClickInfo.adNetworkType              -0.021640
trafficSource.adwordsClickInfo.criteriaParameters         -0.705206
trafficSource.adwordsClickInfo.gclId                      -0.009520
```

```
trafficSource.adwordsClickInfo.isVideoAd            -0.021634
trafficSource.adwordsClickInfo.page                 -0.021657
trafficSource.adwordsClickInfo.slot                 -0.021203
trafficSource.campaign                              -0.073612
trafficSource.isTrueDirect                          -0.054702
trafficSource.keyword                               -0.114066
trafficSource.medium                                -0.077127
trafficSource.referralPath                           0.007839
trafficSource.source                                -0.178594
2017                                                 0.079062
marketType                                          -0.060367
dayOfWeek                                           -0.705206


                                          device.browser  \
channelGrouping                                      0.100624
date                                                -0.088072
socialEngagementType                                 0.026293
visitNumber                                         -0.042104
visitStartTime                                      -0.066757
device.browser                                       1.000000
device.browserSize                                   0.026293
device.browserVersion                                0.026293
device.deviceCategory                                0.274410
device.flashVersion                                  0.026293
device.isMobile                                      0.272339
device.language                                      0.026293
device.mobileDeviceBranding                          0.026293
device.mobileDeviceInfo                              0.026293
device.mobileDeviceMarketingName                     0.026293
device.mobileDeviceModel                             0.026293
device.mobileInputSelector                           0.026293
device.operatingSystem                               0.152662
device.operatingSystemVersion                        0.026293
device.screenColors                                  0.026293
device.screenResolution                              0.026293
geoNetwork.city                                      0.026461
geoNetwork.cityId                                    0.026293
geoNetwork.continent                                 0.042714
geoNetwork.country                                  -0.005837
geoNetwork.latitude                                  0.026293
geoNetwork.longitude                                 0.026293
geoNetwork.metro                                    -0.000761
geoNetwork.networkDomain                             0.090729
geoNetwork.networkLocation                           0.026293
geoNetwork.region                                    0.063486
geoNetwork.subContinent                              0.032524
totals.bounces                                       0.064047
totals.hits                                         -0.075645
```

```
totals.newVisits                                       0.124154
totals.pageviews                                      -0.079099
totals.transactionRevenue                             -0.015087
totals.visits                                          0.026293
trafficSource.adContent                               -0.018506
trafficSource.adwordsClickInfo.adNetworkType          -0.009192
trafficSource.adwordsClickInfo.criteriaParameters      0.026293
trafficSource.adwordsClickInfo.gclId                  -0.024861
trafficSource.adwordsClickInfo.isVideoAd              -0.009202
trafficSource.adwordsClickInfo.page                   -0.009205
trafficSource.adwordsClickInfo.slot                   -0.009737
trafficSource.campaign                                -0.027731
trafficSource.isTrueDirect                             0.106361
trafficSource.keyword                                  0.005992
trafficSource.medium                                   0.056172
trafficSource.referralPath                             0.128266
trafficSource.source                                   0.126462
2017                                                  -0.064396
marketType                                             0.084578
dayOfWeek                                              0.026293


                                             device.browserSize  \
channelGrouping                                        0.011466
date                                                        NaN
socialEngagementType                                   1.000000
visitNumber                                           -0.001474
visitStartTime                                        -0.705206
device.browser                                         0.026293
device.browserSize                                     1.000000
device.browserVersion                                  1.000000
device.deviceCategory                                  0.015478
device.flashVersion                                    1.000000
device.isMobile                                        0.023766
device.language                                        1.000000
device.mobileDeviceBranding                            1.000000
device.mobileDeviceInfo                                1.000000
device.mobileDeviceMarketingName                       1.000000
device.mobileDeviceModel                               1.000000
device.mobileInputSelector                             1.000000
device.operatingSystem                                 0.008183
device.operatingSystemVersion                          1.000000
device.screenColors                                    1.000000
device.screenResolution                                1.000000
geoNetwork.city                                        0.004305
geoNetwork.cityId                                      1.000000
geoNetwork.continent                                   0.022398
geoNetwork.country                                     0.005586
geoNetwork.latitude                                    1.000000
```

```
geoNetwork.longitude                                    1.000000
geoNetwork.metro                                        0.004434
geoNetwork.networkDomain                                0.002657
geoNetwork.networkLocation                              1.000000
geoNetwork.region                                       0.004551
geoNetwork.subContinent                                 0.012515
totals.bounces                                         -0.006027
totals.hits                                            -0.002881
totals.newVisits                                       -0.011313
totals.pageviews                                       -0.003311
totals.transactionRevenue                              -0.000195
totals.visits                                           1.000000
trafficSource.adContent                                 0.000622
trafficSource.adwordsClickInfo.adNetworkType            0.000942
trafficSource.adwordsClickInfo.criteriaParameters       1.000000
trafficSource.adwordsClickInfo.gclId                    0.000823
trafficSource.adwordsClickInfo.isVideoAd                0.000942
trafficSource.adwordsClickInfo.page                     0.000942
trafficSource.adwordsClickInfo.slot                     0.000932
trafficSource.campaign                                  0.117426
trafficSource.isTrueDirect                              0.003986
trafficSource.keyword                                   0.005269
trafficSource.medium                                   -0.000602
trafficSource.referralPath                              0.003173
trafficSource.source                                    0.004904
2017                                                   -0.005053
marketType                                              0.005273
dayOfWeek                                               1.000000


                                               device.browserVersion  \
channelGrouping                                         0.011466
date                                                         NaN
socialEngagementType                                    1.000000
visitNumber                                            -0.001474
visitStartTime                                         -0.705206
device.browser                                          0.026293
device.browserSize                                      1.000000
device.browserVersion                                   1.000000
device.deviceCategory                                   0.015478
device.flashVersion                                     1.000000
device.isMobile                                         0.023766
device.language                                         1.000000
device.mobileDeviceBranding                             1.000000
device.mobileDeviceInfo                                 1.000000
device.mobileDeviceMarketingName                        1.000000
device.mobileDeviceModel                                1.000000
device.mobileInputSelector                              1.000000
device.operatingSystem                                  0.008183
```

51

```
device.operatingSystemVersion                           1.000000
device.screenColors                                     1.000000
device.screenResolution                                 1.000000
geoNetwork.city                                         0.004305
geoNetwork.cityId                                       1.000000
geoNetwork.continent                                    0.022398
geoNetwork.country                                      0.005586
geoNetwork.latitude                                     1.000000
geoNetwork.longitude                                    1.000000
geoNetwork.metro                                        0.004434
geoNetwork.networkDomain                                0.002657
geoNetwork.networkLocation                              1.000000
geoNetwork.region                                       0.004551
geoNetwork.subContinent                                 0.012515
totals.bounces                                         -0.006027
totals.hits                                            -0.002881
totals.newVisits                                       -0.011313
totals.pageviews                                       -0.003311
totals.transactionRevenue                              -0.000195
totals.visits                                           1.000000
trafficSource.adContent                                 0.000622
trafficSource.adwordsClickInfo.adNetworkType            0.000942
trafficSource.adwordsClickInfo.criteriaParameters       1.000000
trafficSource.adwordsClickInfo.gclId                    0.000823
trafficSource.adwordsClickInfo.isVideoAd                0.000942
trafficSource.adwordsClickInfo.page                     0.000942
trafficSource.adwordsClickInfo.slot                     0.000932
trafficSource.campaign                                  0.117426
trafficSource.isTrueDirect                              0.003986
trafficSource.keyword                                   0.005269
trafficSource.medium                                   -0.000602
trafficSource.referralPath                              0.003173
trafficSource.source                                    0.004904
2017                                                   -0.005053
marketType                                              0.005273
dayOfWeek                                               1.000000


                                                device.deviceCategory  \
channelGrouping                                        -0.186820
date                                                    0.118133
socialEngagementType                                    0.015478
visitNumber                                            -0.032287
visitStartTime                                          0.071518
device.browser                                          0.274410
device.browserSize                                      0.015478
device.browserVersion                                   0.015478
device.deviceCategory                                   1.000000
device.flashVersion                                     0.015478
```

```
device.isMobile                                              0.849462
device.language                                              0.015478
device.mobileDeviceBranding                                  0.015478
device.mobileDeviceInfo                                      0.015478
device.mobileDeviceMarketingName                             0.015478
device.mobileDeviceModel                                     0.015478
device.mobileInputSelector                                   0.015478
device.operatingSystem                                      -0.046537
device.operatingSystemVersion                                0.015478
device.screenColors                                          0.015478
device.screenResolution                                      0.015478
geoNetwork.city                                              0.075380
geoNetwork.cityId                                            0.015478
geoNetwork.continent                                         0.000168
geoNetwork.country                                           0.006709
geoNetwork.latitude                                          0.015478
geoNetwork.longitude                                         0.015478
geoNetwork.metro                                             0.073592
geoNetwork.networkDomain                                    -0.008541
geoNetwork.networkLocation                                   0.015478
geoNetwork.region                                            0.082742
geoNetwork.subContinent                                     -0.014935
totals.bounces                                               0.013961
totals.hits                                                 -0.019248
totals.newVisits                                             0.018151
totals.pageviews                                            -0.018119
totals.transactionRevenue                                  -0.013956
totals.visits                                                0.015478
trafficSource.adContent                                     -0.072723
trafficSource.adwordsClickInfo.adNetworkType               -0.074480
trafficSource.adwordsClickInfo.criteriaParameters           0.015478
trafficSource.adwordsClickInfo.gclId                       -0.069977
trafficSource.adwordsClickInfo.isVideoAd                    -0.074491
trafficSource.adwordsClickInfo.page                        -0.074383
trafficSource.adwordsClickInfo.slot                        -0.076720
trafficSource.campaign                                       0.001642
trafficSource.isTrueDirect                                  -0.056104
trafficSource.keyword                                       -0.175726
trafficSource.medium                                       -0.130440
trafficSource.referralPath                                  0.185196
trafficSource.source                                       -0.184463
2017                                                         0.075270
marketType                                                 -0.058623
dayOfWeek                                                    0.015478

                                                  device.flashVersion  \
channelGrouping                                              0.011466
date                                                             NaN
```

socialEngagementType	1.000000
visitNumber	-0.001474
visitStartTime	-0.705206
device.browser	0.026293
device.browserSize	1.000000
device.browserVersion	1.000000
device.deviceCategory	0.015478
device.flashVersion	1.000000
device.isMobile	0.023766
device.language	1.000000
device.mobileDeviceBranding	1.000000
device.mobileDeviceInfo	1.000000
device.mobileDeviceMarketingName	1.000000
device.mobileDeviceModel	1.000000
device.mobileInputSelector	1.000000
device.operatingSystem	0.008183
device.operatingSystemVersion	1.000000
device.screenColors	1.000000
device.screenResolution	1.000000
geoNetwork.city	0.004305
geoNetwork.cityId	1.000000
geoNetwork.continent	0.022398
geoNetwork.country	0.005586
geoNetwork.latitude	1.000000
geoNetwork.longitude	1.000000
geoNetwork.metro	0.004434
geoNetwork.networkDomain	0.002657
geoNetwork.networkLocation	1.000000
geoNetwork.region	0.004551
geoNetwork.subContinent	0.012515
totals.bounces	-0.006027
totals.hits	-0.002881
totals.newVisits	-0.011313
totals.pageviews	-0.003311
totals.transactionRevenue	-0.000195
totals.visits	1.000000
trafficSource.adContent	0.000622
trafficSource.adwordsClickInfo.adNetworkType	0.000942
trafficSource.adwordsClickInfo.criteriaParameters	1.000000
trafficSource.adwordsClickInfo.gclId	0.000823
trafficSource.adwordsClickInfo.isVideoAd	0.000942
trafficSource.adwordsClickInfo.page	0.000942
trafficSource.adwordsClickInfo.slot	0.000932
trafficSource.campaign	0.117426
trafficSource.isTrueDirect	0.003986
trafficSource.keyword	0.005269
trafficSource.medium	-0.000602
trafficSource.referralPath	0.003173

```
trafficSource.source                                         0.004904
2017                                                        -0.005053
marketType                                                   0.005273
dayOfWeek                                                    1.000000


                                                      device.isMobile  \
channelGrouping                                             -0.230171
date                                                         0.146148
socialEngagementType                                         0.023766
visitNumber                                                 -0.037691
visitStartTime                                               0.085543
device.browser                                               0.272339
device.browserSize                                           0.023766
device.browserVersion                                        0.023766
device.deviceCategory                                        0.849462
device.flashVersion                                          0.023766
device.isMobile                                              1.000000
device.language                                              0.023766
device.mobileDeviceBranding                                  0.023766
device.mobileDeviceInfo                                      0.023766
device.mobileDeviceMarketingName                             0.023766
device.mobileDeviceModel                                     0.023766
device.mobileInputSelector                                   0.023766
device.operatingSystem                                      -0.132906
device.operatingSystemVersion                                0.023766
device.screenColors                                          0.023766
device.screenResolution                                      0.023766
geoNetwork.city                                              0.057079
geoNetwork.cityId                                            0.023766
geoNetwork.continent                                        -0.020493
geoNetwork.country                                           0.006639
geoNetwork.latitude                                          0.023766
geoNetwork.longitude                                         0.023766
geoNetwork.metro                                             0.058544
geoNetwork.networkDomain                                    -0.014757
geoNetwork.networkLocation                                   0.023766
geoNetwork.region                                            0.065362
geoNetwork.subContinent                                     -0.008374
totals.bounces                                               0.021244
totals.hits                                                 -0.030426
totals.newVisits                                             0.021380
totals.pageviews                                            -0.030250
totals.transactionRevenue                                   -0.016554
totals.visits                                                0.023766
trafficSource.adContent                                     -0.076653
trafficSource.adwordsClickInfo.adNetworkType               -0.077206
trafficSource.adwordsClickInfo.criteriaParameters           0.023766
trafficSource.adwordsClickInfo.gclId                       -0.067309
```

```
trafficSource.adwordsClickInfo.isVideoAd                        -0.077214
trafficSource.adwordsClickInfo.page                            -0.077144
trafficSource.adwordsClickInfo.slot                            -0.077914
trafficSource.campaign                                         -0.002829
trafficSource.isTrueDirect                                     -0.078572
trafficSource.keyword                                          -0.193968
trafficSource.medium                                           -0.167293
trafficSource.referralPath                                      0.215738
trafficSource.source                                           -0.228006
2017                                                            0.090522
marketType                                                     -0.069739
dayOfWeek                                                       0.023766


                                                    device.language  \
channelGrouping                                           0.011466
date                                                           NaN
socialEngagementType                                      1.000000
visitNumber                                              -0.001474
visitStartTime                                           -0.705206
device.browser                                           0.026293
device.browserSize                                       1.000000
device.browserVersion                                    1.000000
device.deviceCategory                                    0.015478
device.flashVersion                                      1.000000
device.isMobile                                          0.023766
device.language                                          1.000000
device.mobileDeviceBranding                              1.000000
device.mobileDeviceInfo                                  1.000000
device.mobileDeviceMarketingName                         1.000000
device.mobileDeviceModel                                 1.000000
device.mobileInputSelector                               1.000000
device.operatingSystem                                   0.008183
device.operatingSystemVersion                            1.000000
device.screenColors                                      1.000000
device.screenResolution                                  1.000000
geoNetwork.city                                          0.004305
geoNetwork.cityId                                        1.000000
geoNetwork.continent                                     0.022398
geoNetwork.country                                       0.005586
geoNetwork.latitude                                      1.000000
geoNetwork.longitude                                     1.000000
geoNetwork.metro                                         0.004434
geoNetwork.networkDomain                                 0.002657
geoNetwork.networkLocation                               1.000000
geoNetwork.region                                        0.004551
geoNetwork.subContinent                                  0.012515
totals.bounces                                          -0.006027
totals.hits                                             -0.002881
```

```
totals.newVisits                                        -0.011313
totals.pageviews                                        -0.003311
totals.transactionRevenue                               -0.000195
totals.visits                                            1.000000
trafficSource.adContent                                  0.000622
trafficSource.adwordsClickInfo.adNetworkType             0.000942
trafficSource.adwordsClickInfo.criteriaParameters        1.000000
trafficSource.adwordsClickInfo.gclId                     0.000823
trafficSource.adwordsClickInfo.isVideoAd                 0.000942
trafficSource.adwordsClickInfo.page                      0.000942
trafficSource.adwordsClickInfo.slot                      0.000932
trafficSource.campaign                                   0.117426
trafficSource.isTrueDirect                               0.003986
trafficSource.keyword                                    0.005269
trafficSource.medium                                    -0.000602
trafficSource.referralPath                               0.003173
trafficSource.source                                     0.004904
2017                                                    -0.005053
marketType                                               0.005273
dayOfWeek                                                1.000000


                                              device.mobileDeviceBranding  \
channelGrouping                                                 0.011466
date                                                                 NaN
socialEngagementType                                           1.000000
visitNumber                                                   -0.001474
visitStartTime                                               -0.705206
device.browser                                                 0.026293
device.browserSize                                             1.000000
device.browserVersion                                          1.000000
device.deviceCategory                                          0.015478
device.flashVersion                                            1.000000
device.isMobile                                                0.023766
device.language                                                1.000000
device.mobileDeviceBranding                                    1.000000
device.mobileDeviceInfo                                        1.000000
device.mobileDeviceMarketingName                               1.000000
device.mobileDeviceModel                                       1.000000
device.mobileInputSelector                                     1.000000
device.operatingSystem                                         0.008183
device.operatingSystemVersion                                  1.000000
device.screenColors                                            1.000000
device.screenResolution                                        1.000000
geoNetwork.city                                                0.004305
geoNetwork.cityId                                              1.000000
geoNetwork.continent                                           0.022398
geoNetwork.country                                             0.005586
geoNetwork.latitude                                            1.000000
```

```
geoNetwork.longitude                                       1.000000
geoNetwork.metro                                           0.004434
geoNetwork.networkDomain                                   0.002657
geoNetwork.networkLocation                                 1.000000
geoNetwork.region                                          0.004551
geoNetwork.subContinent                                    0.012515
totals.bounces                                            -0.006027
totals.hits                                               -0.002881
totals.newVisits                                          -0.011313
totals.pageviews                                          -0.003311
totals.transactionRevenue                                -0.000195
totals.visits                                             1.000000
trafficSource.adContent                                   0.000622
trafficSource.adwordsClickInfo.adNetworkType              0.000942
trafficSource.adwordsClickInfo.criteriaParameters         1.000000
trafficSource.adwordsClickInfo.gclId                      0.000823
trafficSource.adwordsClickInfo.isVideoAd                  0.000942
trafficSource.adwordsClickInfo.page                       0.000942
trafficSource.adwordsClickInfo.slot                       0.000932
trafficSource.campaign                                    0.117426
trafficSource.isTrueDirect                                0.003986
trafficSource.keyword                                     0.005269
trafficSource.medium                                     -0.000602
trafficSource.referralPath                                0.003173
trafficSource.source                                      0.004904
2017                                                     -0.005053
marketType                                                0.005273
dayOfWeek                                                 1.000000

                                             device.mobileDeviceInfo  \
channelGrouping                                          0.011466
date                                                          NaN
socialEngagementType                                     1.000000
visitNumber                                             -0.001474
visitStartTime                                          -0.705206
device.browser                                           0.026293
device.browserSize                                       1.000000
device.browserVersion                                    1.000000
device.deviceCategory                                    0.015478
device.flashVersion                                      1.000000
device.isMobile                                          0.023766
device.language                                          1.000000
device.mobileDeviceBranding                              1.000000
device.mobileDeviceInfo                                  1.000000
device.mobileDeviceMarketingName                         1.000000
device.mobileDeviceModel                                 1.000000
device.mobileInputSelector                               1.000000
device.operatingSystem                                   0.008183
```

device.operatingSystemVersion	1.000000
device.screenColors	1.000000
device.screenResolution	1.000000
geoNetwork.city	0.004305
geoNetwork.cityId	1.000000
geoNetwork.continent	0.022398
geoNetwork.country	0.005586
geoNetwork.latitude	1.000000
geoNetwork.longitude	1.000000
geoNetwork.metro	0.004434
geoNetwork.networkDomain	0.002657
geoNetwork.networkLocation	1.000000
geoNetwork.region	0.004551
geoNetwork.subContinent	0.012515
totals.bounces	-0.006027
totals.hits	-0.002881
totals.newVisits	-0.011313
totals.pageviews	-0.003311
totals.transactionRevenue	-0.000195
totals.visits	1.000000
trafficSource.adContent	0.000622
trafficSource.adwordsClickInfo.adNetworkType	0.000942
trafficSource.adwordsClickInfo.criteriaParameters	1.000000
trafficSource.adwordsClickInfo.gclId	0.000823
trafficSource.adwordsClickInfo.isVideoAd	0.000942
trafficSource.adwordsClickInfo.page	0.000942
trafficSource.adwordsClickInfo.slot	0.000932
trafficSource.campaign	0.117426
trafficSource.isTrueDirect	0.003986
trafficSource.keyword	0.005269
trafficSource.medium	-0.000602
trafficSource.referralPath	0.003173
trafficSource.source	0.004904
2017	-0.005053
marketType	0.005273
dayOfWeek	1.000000

	device.mobileDeviceMarketingName `
channelGrouping	0.011466
date	NaN
socialEngagementType	1.000000
visitNumber	-0.001474
visitStartTime	-0.705206
device.browser	0.026293
device.browserSize	1.000000
device.browserVersion	1.000000
device.deviceCategory	0.015478
device.flashVersion	1.000000

```
device.isMobile                                       0.023766
device.language                                       1.000000
device.mobileDeviceBranding                           1.000000
device.mobileDeviceInfo                               1.000000
device.mobileDeviceMarketingName                      1.000000
device.mobileDeviceModel                              1.000000
device.mobileInputSelector                            1.000000
device.operatingSystem                                0.008183
device.operatingSystemVersion                         1.000000
device.screenColors                                   1.000000
device.screenResolution                               1.000000
geoNetwork.city                                       0.004305
geoNetwork.cityId                                     1.000000
geoNetwork.continent                                  0.022398
geoNetwork.country                                    0.005586
geoNetwork.latitude                                   1.000000
geoNetwork.longitude                                  1.000000
geoNetwork.metro                                      0.004434
geoNetwork.networkDomain                              0.002657
geoNetwork.networkLocation                            1.000000
geoNetwork.region                                     0.004551
geoNetwork.subContinent                               0.012515
totals.bounces                                       -0.006027
totals.hits                                          -0.002881
totals.newVisits                                     -0.011313
totals.pageviews                                     -0.003311
totals.transactionRevenue                            -0.000195
totals.visits                                         1.000000
trafficSource.adContent                               0.000622
trafficSource.adwordsClickInfo.adNetworkType          0.000942
trafficSource.adwordsClickInfo.criteriaParameters     1.000000
trafficSource.adwordsClickInfo.gclId                  0.000823
trafficSource.adwordsClickInfo.isVideoAd              0.000942
trafficSource.adwordsClickInfo.page                   0.000942
trafficSource.adwordsClickInfo.slot                   0.000932
trafficSource.campaign                                0.117426
trafficSource.isTrueDirect                            0.003986
trafficSource.keyword                                 0.005269
trafficSource.medium                                 -0.000602
trafficSource.referralPath                            0.003173
trafficSource.source                                  0.004904
2017                                                 -0.005053
marketType                                            0.005273
dayOfWeek                                             1.000000


                                    device.mobileDeviceModel  \
channelGrouping                                       0.011466
date                                                       NaN
```

socialEngagementType	1.000000
visitNumber	-0.001474
visitStartTime	-0.705206
device.browser	0.026293
device.browserSize	1.000000
device.browserVersion	1.000000
device.deviceCategory	0.015478
device.flashVersion	1.000000
device.isMobile	0.023766
device.language	1.000000
device.mobileDeviceBranding	1.000000
device.mobileDeviceInfo	1.000000
device.mobileDeviceMarketingName	1.000000
device.mobileDeviceModel	1.000000
device.mobileInputSelector	1.000000
device.operatingSystem	0.008183
device.operatingSystemVersion	1.000000
device.screenColors	1.000000
device.screenResolution	1.000000
geoNetwork.city	0.004305
geoNetwork.cityId	1.000000
geoNetwork.continent	0.022398
geoNetwork.country	0.005586
geoNetwork.latitude	1.000000
geoNetwork.longitude	1.000000
geoNetwork.metro	0.004434
geoNetwork.networkDomain	0.002657
geoNetwork.networkLocation	1.000000
geoNetwork.region	0.004551
geoNetwork.subContinent	0.012515
totals.bounces	-0.006027
totals.hits	-0.002881
totals.newVisits	-0.011313
totals.pageviews	-0.003311
totals.transactionRevenue	-0.000195
totals.visits	1.000000
trafficSource.adContent	0.000622
trafficSource.adwordsClickInfo.adNetworkType	0.000942
trafficSource.adwordsClickInfo.criteriaParameters	1.000000
trafficSource.adwordsClickInfo.gclId	0.000823
trafficSource.adwordsClickInfo.isVideoAd	0.000942
trafficSource.adwordsClickInfo.page	0.000942
trafficSource.adwordsClickInfo.slot	0.000932
trafficSource.campaign	0.117426
trafficSource.isTrueDirect	0.003986
trafficSource.keyword	0.005269
trafficSource.medium	-0.000602
trafficSource.referralPath	0.003173

```
trafficSource.source                                       0.004904
2017                                                       -0.005053
marketType                                                  0.005273
dayOfWeek                                                   1.000000


                                               device.mobileInputSelector  \
channelGrouping                                             0.011466
date                                                             NaN
socialEngagementType                                        1.000000
visitNumber                                                -0.001474
visitStartTime                                             -0.705206
device.browser                                              0.026293
device.browserSize                                          1.000000
device.browserVersion                                       1.000000
device.deviceCategory                                       0.015478
device.flashVersion                                         1.000000
device.isMobile                                             0.023766
device.language                                             1.000000
device.mobileDeviceBranding                                 1.000000
device.mobileDeviceInfo                                     1.000000
device.mobileDeviceMarketingName                            1.000000
device.mobileDeviceModel                                    1.000000
device.mobileInputSelector                                  1.000000
device.operatingSystem                                      0.008183
device.operatingSystemVersion                               1.000000
device.screenColors                                         1.000000
device.screenResolution                                     1.000000
geoNetwork.city                                             0.004305
geoNetwork.cityId                                           1.000000
geoNetwork.continent                                        0.022398
geoNetwork.country                                          0.005586
geoNetwork.latitude                                         1.000000
geoNetwork.longitude                                        1.000000
geoNetwork.metro                                            0.004434
geoNetwork.networkDomain                                    0.002657
geoNetwork.networkLocation                                  1.000000
geoNetwork.region                                           0.004551
geoNetwork.subContinent                                     0.012515
totals.bounces                                             -0.006027
totals.hits                                                -0.002881
totals.newVisits                                           -0.011313
totals.pageviews                                           -0.003311
totals.transactionRevenue                                 -0.000195
totals.visits                                               1.000000
trafficSource.adContent                                     0.000622
trafficSource.adwordsClickInfo.adNetworkType                0.000942
trafficSource.adwordsClickInfo.criteriaParameters           1.000000
trafficSource.adwordsClickInfo.gclId                        0.000823
```

```
trafficSource.adwordsClickInfo.isVideoAd                    0.000942
trafficSource.adwordsClickInfo.page                         0.000942
trafficSource.adwordsClickInfo.slot                         0.000932
trafficSource.campaign                                      0.117426
trafficSource.isTrueDirect                                  0.003986
trafficSource.keyword                                       0.005269
trafficSource.medium                                       -0.000602
trafficSource.referralPath                                  0.003173
trafficSource.source                                        0.004904
2017                                                       -0.005053
marketType                                                  0.005273
dayOfWeek                                                   1.000000


                                             device.operatingSystem  \
channelGrouping                                            0.030185
date                                                       0.001192
socialEngagementType                                       0.008183
visitNumber                                                0.007164
visitStartTime                                            -0.008923
device.browser                                             0.152662
device.browserSize                                         0.008183
device.browserVersion                                      0.008183
device.deviceCategory                                     -0.046537
device.flashVersion                                        0.008183
device.isMobile                                           -0.132906
device.language                                            0.008183
device.mobileDeviceBranding                                0.008183
device.mobileDeviceInfo                                    0.008183
device.mobileDeviceMarketingName                           0.008183
device.mobileDeviceModel                                   0.008183
device.mobileInputSelector                                 0.008183
device.operatingSystem                                     1.000000
device.operatingSystemVersion                              0.008183
device.screenColors                                        0.008183
device.screenResolution                                    0.008183
geoNetwork.city                                            0.094844
geoNetwork.cityId                                          0.008183
geoNetwork.continent                                       0.109432
geoNetwork.country                                        -0.073224
geoNetwork.latitude                                        0.008183
geoNetwork.longitude                                       0.008183
geoNetwork.metro                                           0.040326
geoNetwork.networkDomain                                   0.110609
geoNetwork.networkLocation                                 0.008183
geoNetwork.region                                          0.141732
geoNetwork.subContinent                                   -0.018006
totals.bounces                                             0.071871
totals.hits                                               -0.057431
```

```
totals.newVisits                                          0.086957
totals.pageviews                                         -0.059827
totals.transactionRevenue                                -0.011926
totals.visits                                             0.008183
trafficSource.adContent                                   0.002792
trafficSource.adwordsClickInfo.adNetworkType              0.007869
trafficSource.adwordsClickInfo.criteriaParameters         0.008183
trafficSource.adwordsClickInfo.gclId                     -0.009235
trafficSource.adwordsClickInfo.isVideoAd                  0.007860
trafficSource.adwordsClickInfo.page                       0.007807
trafficSource.adwordsClickInfo.slot                       0.007880
trafficSource.campaign                                    0.028353
trafficSource.isTrueDirect                                0.114941
trafficSource.keyword                                    -0.092660
trafficSource.medium                                      0.065488
trafficSource.referralPath                                0.130801
trafficSource.source                                      0.042441
2017                                                     -0.066450
marketType                                                0.115170
dayOfWeek                                                 0.008183


                                            device.operatingSystemVersion  \
channelGrouping                                               0.011466
date                                                              NaN
socialEngagementType                                         1.000000
visitNumber                                                 -0.001474
visitStartTime                                             -0.705206
device.browser                                              0.026293
device.browserSize                                          1.000000
device.browserVersion                                       1.000000
device.deviceCategory                                       0.015478
device.flashVersion                                         1.000000
device.isMobile                                             0.023766
device.language                                             1.000000
device.mobileDeviceBranding                                 1.000000
device.mobileDeviceInfo                                     1.000000
device.mobileDeviceMarketingName                            1.000000
device.mobileDeviceModel                                    1.000000
device.mobileInputSelector                                  1.000000
device.operatingSystem                                      0.008183
device.operatingSystemVersion                               1.000000
device.screenColors                                         1.000000
device.screenResolution                                     1.000000
geoNetwork.city                                             0.004305
geoNetwork.cityId                                           1.000000
geoNetwork.continent                                        0.022398
geoNetwork.country                                          0.005586
geoNetwork.latitude                                         1.000000
```

```
geoNetwork.longitude                                    1.000000
geoNetwork.metro                                        0.004434
geoNetwork.networkDomain                                0.002657
geoNetwork.networkLocation                              1.000000
geoNetwork.region                                       0.004551
geoNetwork.subContinent                                 0.012515
totals.bounces                                         -0.006027
totals.hits                                            -0.002881
totals.newVisits                                       -0.011313
totals.pageviews                                       -0.003311
totals.transactionRevenue                              -0.000195
totals.visits                                           1.000000
trafficSource.adContent                                 0.000622
trafficSource.adwordsClickInfo.adNetworkType            0.000942
trafficSource.adwordsClickInfo.criteriaParameters       1.000000
trafficSource.adwordsClickInfo.gclId                    0.000823
trafficSource.adwordsClickInfo.isVideoAd                0.000942
trafficSource.adwordsClickInfo.page                     0.000942
trafficSource.adwordsClickInfo.slot                     0.000932
trafficSource.campaign                                  0.117426
trafficSource.isTrueDirect                              0.003986
trafficSource.keyword                                   0.005269
trafficSource.medium                                   -0.000602
trafficSource.referralPath                              0.003173
trafficSource.source                                    0.004904
2017                                                   -0.005053
marketType                                              0.005273
dayOfWeek                                               1.000000


                                      device.screenColors  \
channelGrouping                                 0.011466
date                                                 NaN
socialEngagementType                            1.000000
visitNumber                                    -0.001474
visitStartTime                                 -0.705206
device.browser                                  0.026293
device.browserSize                              1.000000
device.browserVersion                           1.000000
device.deviceCategory                           0.015478
device.flashVersion                             1.000000
device.isMobile                                 0.023766
device.language                                 1.000000
device.mobileDeviceBranding                     1.000000
device.mobileDeviceInfo                         1.000000
device.mobileDeviceMarketingName                1.000000
device.mobileDeviceModel                        1.000000
device.mobileInputSelector                      1.000000
device.operatingSystem                          0.008183
```

```
device.operatingSystemVersion                                 1.000000
device.screenColors                                           1.000000
device.screenResolution                                       1.000000
geoNetwork.city                                               0.004305
geoNetwork.cityId                                             1.000000
geoNetwork.continent                                          0.022398
geoNetwork.country                                            0.005586
geoNetwork.latitude                                           1.000000
geoNetwork.longitude                                          1.000000
geoNetwork.metro                                              0.004434
geoNetwork.networkDomain                                      0.002657
geoNetwork.networkLocation                                    1.000000
geoNetwork.region                                             0.004551
geoNetwork.subContinent                                       0.012515
totals.bounces                                               -0.006027
totals.hits                                                  -0.002881
totals.newVisits                                            -0.011313
totals.pageviews                                            -0.003311
totals.transactionRevenue                                  -0.000195
totals.visits                                                1.000000
trafficSource.adContent                                      0.000622
trafficSource.adwordsClickInfo.adNetworkType                 0.000942
trafficSource.adwordsClickInfo.criteriaParameters            1.000000
trafficSource.adwordsClickInfo.gclId                         0.000823
trafficSource.adwordsClickInfo.isVideoAd                     0.000942
trafficSource.adwordsClickInfo.page                          0.000942
trafficSource.adwordsClickInfo.slot                          0.000932
trafficSource.campaign                                       0.117426
trafficSource.isTrueDirect                                   0.003986
trafficSource.keyword                                        0.005269
trafficSource.medium                                        -0.000602
trafficSource.referralPath                                   0.003173
trafficSource.source                                         0.004904
2017                                                        -0.005053
marketType                                                   0.005273
dayOfWeek                                                    1.000000


                                                   device.screenResolution  \
channelGrouping                                                    0.011466
date                                                                    NaN
socialEngagementType                                              1.000000
visitNumber                                                      -0.001474
visitStartTime                                                  -0.705206
device.browser                                                   0.026293
device.browserSize                                               1.000000
device.browserVersion                                            1.000000
device.deviceCategory                                            0.015478
device.flashVersion                                              1.000000
```

```
device.isMobile                                         0.023766
device.language                                         1.000000
device.mobileDeviceBranding                             1.000000
device.mobileDeviceInfo                                 1.000000
device.mobileDeviceMarketingName                        1.000000
device.mobileDeviceModel                                1.000000
device.mobileInputSelector                              1.000000
device.operatingSystem                                  0.008183
device.operatingSystemVersion                           1.000000
device.screenColors                                     1.000000
device.screenResolution                                 1.000000
geoNetwork.city                                         0.004305
geoNetwork.cityId                                       1.000000
geoNetwork.continent                                    0.022398
geoNetwork.country                                      0.005586
geoNetwork.latitude                                     1.000000
geoNetwork.longitude                                    1.000000
geoNetwork.metro                                        0.004434
geoNetwork.networkDomain                                0.002657
geoNetwork.networkLocation                              1.000000
geoNetwork.region                                       0.004551
geoNetwork.subContinent                                 0.012515
totals.bounces                                         -0.006027
totals.hits                                            -0.002881
totals.newVisits                                       -0.011313
totals.pageviews                                       -0.003311
totals.transactionRevenue                              -0.000195
totals.visits                                           1.000000
trafficSource.adContent                                 0.000622
trafficSource.adwordsClickInfo.adNetworkType            0.000942
trafficSource.adwordsClickInfo.criteriaParameters       1.000000
trafficSource.adwordsClickInfo.gclId                    0.000823
trafficSource.adwordsClickInfo.isVideoAd                0.000942
trafficSource.adwordsClickInfo.page                     0.000942
trafficSource.adwordsClickInfo.slot                     0.000932
trafficSource.campaign                                  0.117426
trafficSource.isTrueDirect                              0.003986
trafficSource.keyword                                   0.005269
trafficSource.medium                                   -0.000602
trafficSource.referralPath                              0.003173
trafficSource.source                                    0.004904
2017                                                   -0.005053
marketType                                              0.005273
dayOfWeek                                               1.000000


                                            geoNetwork.city  \
channelGrouping                                    -0.006282
date                                                0.051625
```

socialEngagementType	0.004305
visitNumber	0.003616
visitStartTime	0.017869
device.browser	0.026461
device.browserSize	0.004305
device.browserVersion	0.004305
device.deviceCategory	0.075380
device.flashVersion	0.004305
device.isMobile	0.057079
device.language	0.004305
device.mobileDeviceBranding	0.004305
device.mobileDeviceInfo	0.004305
device.mobileDeviceMarketingName	0.004305
device.mobileDeviceModel	0.004305
device.mobileInputSelector	0.004305
device.operatingSystem	0.094844
device.operatingSystemVersion	0.004305
device.screenColors	0.004305
device.screenResolution	0.004305
geoNetwork.city	1.000000
geoNetwork.cityId	0.004305
geoNetwork.continent	0.029115
geoNetwork.country	-0.085380
geoNetwork.latitude	0.004305
geoNetwork.longitude	0.004305
geoNetwork.metro	0.817939
geoNetwork.networkDomain	0.065346
geoNetwork.networkLocation	0.004305
geoNetwork.region	0.710521
geoNetwork.subContinent	-0.030103
totals.bounces	0.006648
totals.hits	0.000201
totals.newVisits	0.036183
totals.pageviews	0.000804
totals.transactionRevenue	-0.003329
totals.visits	0.004305
trafficSource.adContent	-0.020175
trafficSource.adwordsClickInfo.adNetworkType	-0.034057
trafficSource.adwordsClickInfo.criteriaParameters	0.004305
trafficSource.adwordsClickInfo.gclId	-0.029989
trafficSource.adwordsClickInfo.isVideoAd	-0.034060
trafficSource.adwordsClickInfo.page	-0.034039
trafficSource.adwordsClickInfo.slot	-0.034269
trafficSource.campaign	0.007393
trafficSource.isTrueDirect	0.066719
trafficSource.keyword	-0.110182
trafficSource.medium	0.035982
trafficSource.referralPath	0.093064

```
trafficSource.source                                           0.003061
2017                                                          -0.087281
marketType                                                   -0.027961
dayOfWeek                                                     0.004305


                                                    geoNetwork.cityId  \
channelGrouping                                               0.011466
date                                                               NaN
socialEngagementType                                         1.000000
visitNumber                                                  -0.001474
visitStartTime                                               -0.705206
device.browser                                               0.026293
device.browserSize                                           1.000000
device.browserVersion                                        1.000000
device.deviceCategory                                        0.015478
device.flashVersion                                          1.000000
device.isMobile                                              0.023766
device.language                                              1.000000
device.mobileDeviceBranding                                  1.000000
device.mobileDeviceInfo                                      1.000000
device.mobileDeviceMarketingName                             1.000000
device.mobileDeviceModel                                     1.000000
device.mobileInputSelector                                   1.000000
device.operatingSystem                                       0.008183
device.operatingSystemVersion                                1.000000
device.screenColors                                          1.000000
device.screenResolution                                      1.000000
geoNetwork.city                                              0.004305
geoNetwork.cityId                                            1.000000
geoNetwork.continent                                         0.022398
geoNetwork.country                                           0.005586
geoNetwork.latitude                                          1.000000
geoNetwork.longitude                                         1.000000
geoNetwork.metro                                             0.004434
geoNetwork.networkDomain                                     0.002657
geoNetwork.networkLocation                                   1.000000
geoNetwork.region                                            0.004551
geoNetwork.subContinent                                      0.012515
totals.bounces                                               -0.006027
totals.hits                                                  -0.002881
totals.newVisits                                             -0.011313
totals.pageviews                                             -0.003311
totals.transactionRevenue                                    -0.000195
totals.visits                                                1.000000
trafficSource.adContent                                      0.000622
trafficSource.adwordsClickInfo.adNetworkType                 0.000942
trafficSource.adwordsClickInfo.criteriaParameters            1.000000
trafficSource.adwordsClickInfo.gclId                         0.000823
```

```
trafficSource.adwordsClickInfo.isVideoAd                    0.000942
trafficSource.adwordsClickInfo.page                         0.000942
trafficSource.adwordsClickInfo.slot                         0.000932
trafficSource.campaign                                      0.117426
trafficSource.isTrueDirect                                  0.003986
trafficSource.keyword                                       0.005269
trafficSource.medium                                       -0.000602
trafficSource.referralPath                                  0.003173
trafficSource.source                                        0.004904
2017                                                       -0.005053
marketType                                                  0.005273
dayOfWeek                                                   1.000000


                                            geoNetwork.continent  \
channelGrouping                                             0.038730
date                                                        0.007571
socialEngagementType                                        0.022398
visitNumber                                                -0.064838
visitStartTime                                             -0.014863
device.browser                                              0.042714
device.browserSize                                          0.022398
device.browserVersion                                       0.022398
device.deviceCategory                                       0.000168
device.flashVersion                                         0.022398
device.isMobile                                            -0.020493
device.language                                             0.022398
device.mobileDeviceBranding                                 0.022398
device.mobileDeviceInfo                                     0.022398
device.mobileDeviceMarketingName                            0.022398
device.mobileDeviceModel                                    0.022398
device.mobileInputSelector                                  0.022398
device.operatingSystem                                      0.109432
device.operatingSystemVersion                               0.022398
device.screenColors                                         0.022398
device.screenResolution                                     0.022398
geoNetwork.city                                             0.029115
geoNetwork.cityId                                           0.022398
geoNetwork.continent                                        1.000000
geoNetwork.country                                         -0.336176
geoNetwork.latitude                                         0.022398
geoNetwork.longitude                                        0.022398
geoNetwork.metro                                           -0.121983
geoNetwork.networkDomain                                    0.224018
geoNetwork.networkLocation                                  0.022398
geoNetwork.region                                           0.125180
geoNetwork.subContinent                                     0.201393
totals.bounces                                              0.176631
totals.hits                                                -0.134314
```

```
totals.newVisits                                            0.142905
totals.pageviews                                           -0.143661
totals.transactionRevenue                                  -0.025521
totals.visits                                               0.022398
trafficSource.adContent                                     0.048196
trafficSource.adwordsClickInfo.adNetworkType                0.119365
trafficSource.adwordsClickInfo.criteriaParameters           0.022398
trafficSource.adwordsClickInfo.gclId                        0.104574
trafficSource.adwordsClickInfo.isVideoAd                    0.119371
trafficSource.adwordsClickInfo.page                         0.119328
trafficSource.adwordsClickInfo.slot                         0.118171
trafficSource.campaign                                     -0.013295
trafficSource.isTrueDirect                                  0.140148
trafficSource.keyword                                      -0.065934
trafficSource.medium                                        0.089292
trafficSource.referralPath                                  0.124510
trafficSource.source                                        0.055218
2017                                                       -0.086936
marketType                                                  0.706282
dayOfWeek                                                   0.022398


                                              geoNetwork.country  \
channelGrouping                                          -0.045936
date                                                     -0.008117
socialEngagementType                                      0.005586
visitNumber                                               0.052182
visitStartTime                                           -0.004350
device.browser                                           -0.005837
device.browserSize                                        0.005586
device.browserVersion                                     0.005586
device.deviceCategory                                     0.006709
device.flashVersion                                       0.005586
device.isMobile                                           0.006639
device.language                                           0.005586
device.mobileDeviceBranding                               0.005586
device.mobileDeviceInfo                                   0.005586
device.mobileDeviceMarketingName                          0.005586
device.mobileDeviceModel                                  0.005586
device.mobileInputSelector                                0.005586
device.operatingSystem                                   -0.073224
device.operatingSystemVersion                             0.005586
device.screenColors                                       0.005586
device.screenResolution                                   0.005586
geoNetwork.city                                          -0.085380
geoNetwork.cityId                                         0.005586
geoNetwork.continent                                     -0.336176
geoNetwork.country                                        1.000000
geoNetwork.latitude                                       0.005586
```

```
geoNetwork.longitude                                    0.005586
geoNetwork.metro                                        0.116223
geoNetwork.networkDomain                               -0.207694
geoNetwork.networkLocation                              0.005586
geoNetwork.region                                      -0.187797
geoNetwork.subContinent                                -0.118254
totals.bounces                                         -0.138695
totals.hits                                             0.102283
totals.newVisits                                       -0.132127
totals.pageviews                                        0.109547
totals.transactionRevenue                               0.022577
totals.visits                                           0.005586
trafficSource.adContent                                -0.038088
trafficSource.adwordsClickInfo.adNetworkType           -0.112839
trafficSource.adwordsClickInfo.criteriaParameters       0.005586
trafficSource.adwordsClickInfo.gclId                   -0.098938
trafficSource.adwordsClickInfo.isVideoAd               -0.112845
trafficSource.adwordsClickInfo.page                    -0.112794
trafficSource.adwordsClickInfo.slot                    -0.111787
trafficSource.campaign                                  0.018427
trafficSource.isTrueDirect                             -0.115119
trafficSource.keyword                                   0.017918
trafficSource.medium                                   -0.069049
trafficSource.referralPath                             -0.114987
trafficSource.source                                   -0.064157
2017                                                    0.344886
marketType                                             -0.541062
dayOfWeek                                               0.005586

                                              geoNetwork.latitude  \
channelGrouping                                          0.011466
date                                                         NaN
socialEngagementType                                    1.000000
visitNumber                                            -0.001474
visitStartTime                                         -0.705206
device.browser                                          0.026293
device.browserSize                                      1.000000
device.browserVersion                                   1.000000
device.deviceCategory                                   0.015478
device.flashVersion                                     1.000000
device.isMobile                                         0.023766
device.language                                         1.000000
device.mobileDeviceBranding                             1.000000
device.mobileDeviceInfo                                 1.000000
device.mobileDeviceMarketingName                        1.000000
device.mobileDeviceModel                                1.000000
device.mobileInputSelector                              1.000000
device.operatingSystem                                  0.008183
```

```
device.operatingSystemVersion                              1.000000
device.screenColors                                        1.000000
device.screenResolution                                    1.000000
geoNetwork.city                                            0.004305
geoNetwork.cityId                                          1.000000
geoNetwork.continent                                       0.022398
geoNetwork.country                                         0.005586
geoNetwork.latitude                                        1.000000
geoNetwork.longitude                                       1.000000
geoNetwork.metro                                           0.004434
geoNetwork.networkDomain                                   0.002657
geoNetwork.networkLocation                                 1.000000
geoNetwork.region                                          0.004551
geoNetwork.subContinent                                    0.012515
totals.bounces                                            -0.006027
totals.hits                                              -0.002881
totals.newVisits                                         -0.011313
totals.pageviews                                         -0.003311
totals.transactionRevenue                               -0.000195
totals.visits                                             1.000000
trafficSource.adContent                                   0.000622
trafficSource.adwordsClickInfo.adNetworkType              0.000942
trafficSource.adwordsClickInfo.criteriaParameters         1.000000
trafficSource.adwordsClickInfo.gclId                      0.000823
trafficSource.adwordsClickInfo.isVideoAd                  0.000942
trafficSource.adwordsClickInfo.page                       0.000942
trafficSource.adwordsClickInfo.slot                       0.000932
trafficSource.campaign                                     0.117426
trafficSource.isTrueDirect                                0.003986
trafficSource.keyword                                     0.005269
trafficSource.medium                                     -0.000602
trafficSource.referralPath                                0.003173
trafficSource.source                                      0.004904
2017                                                     -0.005053
marketType                                                0.005273
dayOfWeek                                                 1.000000


                                                 geoNetwork.longitude  \
channelGrouping                                            0.011466
date                                                            NaN
socialEngagementType                                       1.000000
visitNumber                                              -0.001474
visitStartTime                                           -0.705206
device.browser                                            0.026293
device.browserSize                                        1.000000
device.browserVersion                                     1.000000
device.deviceCategory                                     0.015478
device.flashVersion                                       1.000000
```

```
device.isMobile                                           0.023766
device.language                                           1.000000
device.mobileDeviceBranding                               1.000000
device.mobileDeviceInfo                                   1.000000
device.mobileDeviceMarketingName                          1.000000
device.mobileDeviceModel                                  1.000000
device.mobileInputSelector                                1.000000
device.operatingSystem                                    0.008183
device.operatingSystemVersion                             1.000000
device.screenColors                                       1.000000
device.screenResolution                                   1.000000
geoNetwork.city                                           0.004305
geoNetwork.cityId                                         1.000000
geoNetwork.continent                                      0.022398
geoNetwork.country                                        0.005586
geoNetwork.latitude                                       1.000000
geoNetwork.longitude                                      1.000000
geoNetwork.metro                                          0.004434
geoNetwork.networkDomain                                  0.002657
geoNetwork.networkLocation                                1.000000
geoNetwork.region                                         0.004551
geoNetwork.subContinent                                   0.012515
totals.bounces                                           -0.006027
totals.hits                                              -0.002881
totals.newVisits                                         -0.011313
totals.pageviews                                         -0.003311
totals.transactionRevenue                               -0.000195
totals.visits                                            1.000000
trafficSource.adContent                                  0.000622
trafficSource.adwordsClickInfo.adNetworkType             0.000942
trafficSource.adwordsClickInfo.criteriaParameters        1.000000
trafficSource.adwordsClickInfo.gclId                     0.000823
trafficSource.adwordsClickInfo.isVideoAd                 0.000942
trafficSource.adwordsClickInfo.page                      0.000942
trafficSource.adwordsClickInfo.slot                      0.000932
trafficSource.campaign                                    0.117426
trafficSource.isTrueDirect                               0.003986
trafficSource.keyword                                     0.005269
trafficSource.medium                                    -0.000602
trafficSource.referralPath                               0.003173
trafficSource.source                                     0.004904
2017                                                    -0.005053
marketType                                               0.005273
dayOfWeek                                                1.000000


                                               geoNetwork.metro  \
channelGrouping                                       -0.014529
date                                                   0.062643
```

socialEngagementType	0.004434
visitNumber	0.014937
visitStartTime	0.025625
device.browser	−0.000761
device.browserSize	0.004434
device.browserVersion	0.004434
device.deviceCategory	0.073592
device.flashVersion	0.004434
device.isMobile	0.058544
device.language	0.004434
device.mobileDeviceBranding	0.004434
device.mobileDeviceInfo	0.004434
device.mobileDeviceMarketingName	0.004434
device.mobileDeviceModel	0.004434
device.mobileInputSelector	0.004434
device.operatingSystem	0.040326
device.operatingSystemVersion	0.004434
device.screenColors	0.004434
device.screenResolution	0.004434
geoNetwork.city	0.817939
geoNetwork.cityId	0.004434
geoNetwork.continent	−0.121983
geoNetwork.country	0.116223
geoNetwork.latitude	0.004434
geoNetwork.longitude	0.004434
geoNetwork.metro	1.000000
geoNetwork.networkDomain	−0.053078
geoNetwork.networkLocation	0.004434
geoNetwork.region	0.606743
geoNetwork.subContinent	−0.057936
totals.bounces	−0.041629
totals.hits	0.035971
totals.newVisits	−0.011092
totals.pageviews	0.039147
totals.transactionRevenue	0.003988
totals.visits	0.004434
trafficSource.adContent	−0.026371
trafficSource.adwordsClickInfo.adNetworkType	−0.055855
trafficSource.adwordsClickInfo.criteriaParameters	0.004434
trafficSource.adwordsClickInfo.gclId	−0.049186
trafficSource.adwordsClickInfo.isVideoAd	−0.055858
trafficSource.adwordsClickInfo.page	−0.055834
trafficSource.adwordsClickInfo.slot	−0.055629
trafficSource.campaign	−0.002450
trafficSource.isTrueDirect	0.018030
trafficSource.keyword	−0.077269
trafficSource.medium	0.015569
trafficSource.referralPath	0.015635

```
trafficSource.source                                    -0.008872
2017                                                    -0.008035
marketType                                              -0.213843
dayOfWeek                                                0.004434


                                                  geoNetwork.networkDomain  \
channelGrouping                                          0.110806
date                                                    -0.077520
socialEngagementType                                     0.002657
visitNumber                                             -0.046552
visitStartTime                                          -0.056860
device.browser                                           0.090729
device.browserSize                                       0.002657
device.browserVersion                                    0.002657
device.deviceCategory                                   -0.008541
device.flashVersion                                      0.002657
device.isMobile                                         -0.014757
device.language                                          0.002657
device.mobileDeviceBranding                              0.002657
device.mobileDeviceInfo                                  0.002657
device.mobileDeviceMarketingName                         0.002657
device.mobileDeviceModel                                 0.002657
device.mobileInputSelector                               0.002657
device.operatingSystem                                   0.110609
device.operatingSystemVersion                            0.002657
device.screenColors                                      0.002657
device.screenResolution                                  0.002657
geoNetwork.city                                          0.065346
geoNetwork.cityId                                        0.002657
geoNetwork.continent                                     0.224018
geoNetwork.country                                      -0.207694
geoNetwork.latitude                                      0.002657
geoNetwork.longitude                                     0.002657
geoNetwork.metro                                        -0.053078
geoNetwork.networkDomain                                 1.000000
geoNetwork.networkLocation                               0.002657
geoNetwork.region                                        0.155062
geoNetwork.subContinent                                  0.130431
totals.bounces                                           0.127310
totals.hits                                             -0.099445
totals.newVisits                                         0.148315
totals.pageviews                                        -0.106702
totals.transactionRevenue                               -0.020199
totals.visits                                            0.002657
trafficSource.adContent                                  0.013745
trafficSource.adwordsClickInfo.adNetworkType             0.046517
trafficSource.adwordsClickInfo.criteriaParameters        0.002657
trafficSource.adwordsClickInfo.gclId                     0.038921
```

```
trafficSource.adwordsClickInfo.isVideoAd                      0.046518
trafficSource.adwordsClickInfo.page                          0.046459
trafficSource.adwordsClickInfo.slot                          0.046118
trafficSource.campaign                                       0.004526
trafficSource.isTrueDirect                                   0.136705
trafficSource.keyword                                        0.012183
trafficSource.medium                                         0.079339
trafficSource.referralPath                                   0.156075
trafficSource.source                                         0.131221
2017                                                        -0.195755
marketType                                                   0.378584
dayOfWeek                                                    0.002657


                                          geoNetwork.networkLocation  \
channelGrouping                                               0.011466
date                                                               NaN
socialEngagementType                                         1.000000
visitNumber                                                 -0.001474
visitStartTime                                             -0.705206
device.browser                                              0.026293
device.browserSize                                          1.000000
device.browserVersion                                       1.000000
device.deviceCategory                                       0.015478
device.flashVersion                                         1.000000
device.isMobile                                             0.023766
device.language                                             1.000000
device.mobileDeviceBranding                                 1.000000
device.mobileDeviceInfo                                     1.000000
device.mobileDeviceMarketingName                            1.000000
device.mobileDeviceModel                                    1.000000
device.mobileInputSelector                                  1.000000
device.operatingSystem                                      0.008183
device.operatingSystemVersion                               1.000000
device.screenColors                                         1.000000
device.screenResolution                                     1.000000
geoNetwork.city                                             0.004305
geoNetwork.cityId                                           1.000000
geoNetwork.continent                                        0.022398
geoNetwork.country                                          0.005586
geoNetwork.latitude                                         1.000000
geoNetwork.longitude                                        1.000000
geoNetwork.metro                                            0.004434
geoNetwork.networkDomain                                    0.002657
geoNetwork.networkLocation                                  1.000000
geoNetwork.region                                           0.004551
geoNetwork.subContinent                                     0.012515
totals.bounces                                             -0.006027
totals.hits                                                -0.002881
```

```
totals.newVisits                                            -0.011313
totals.pageviews                                            -0.003311
totals.transactionRevenue                                   -0.000195
totals.visits                                                1.000000
trafficSource.adContent                                      0.000622
trafficSource.adwordsClickInfo.adNetworkType                 0.000942
trafficSource.adwordsClickInfo.criteriaParameters            1.000000
trafficSource.adwordsClickInfo.gclId                         0.000823
trafficSource.adwordsClickInfo.isVideoAd                     0.000942
trafficSource.adwordsClickInfo.page                          0.000942
trafficSource.adwordsClickInfo.slot                          0.000932
trafficSource.campaign                                       0.117426
trafficSource.isTrueDirect                                   0.003986
trafficSource.keyword                                        0.005269
trafficSource.medium                                        -0.000602
trafficSource.referralPath                                   0.003173
trafficSource.source                                         0.004904
2017                                                        -0.005053
marketType                                                   0.005273
dayOfWeek                                                    1.000000


                                              geoNetwork.region  \
channelGrouping                                        0.017590
date                                                   0.032589
socialEngagementType                                   0.004551
visitNumber                                           -0.003238
visitStartTime                                         0.007656
device.browser                                         0.063486
device.browserSize                                     0.004551
device.browserVersion                                  0.004551
device.deviceCategory                                  0.082742
device.flashVersion                                    0.004551
device.isMobile                                        0.065362
device.language                                        0.004551
device.mobileDeviceBranding                            0.004551
device.mobileDeviceInfo                                0.004551
device.mobileDeviceMarketingName                       0.004551
device.mobileDeviceModel                               0.004551
device.mobileInputSelector                             0.004551
device.operatingSystem                                 0.141732
device.operatingSystemVersion                          0.004551
device.screenColors                                    0.004551
device.screenResolution                                0.004551
geoNetwork.city                                        0.710521
geoNetwork.cityId                                      0.004551
geoNetwork.continent                                   0.125180
geoNetwork.country                                    -0.187797
geoNetwork.latitude                                    0.004551
```

```
geoNetwork.longitude                                        0.004551
geoNetwork.metro                                            0.606743
geoNetwork.networkDomain                                    0.155062
geoNetwork.networkLocation                                  0.004551
geoNetwork.region                                           1.000000
geoNetwork.subContinent                                     0.004792
totals.bounces                                              0.067495
totals.hits                                                -0.039878
totals.newVisits                                            0.106419
totals.pageviews                                           -0.043175
totals.transactionRevenue                                 -0.006862
totals.visits                                               0.004551
trafficSource.adContent                                    -0.014586
trafficSource.adwordsClickInfo.adNetworkType              -0.011856
trafficSource.adwordsClickInfo.criteriaParameters          0.004551
trafficSource.adwordsClickInfo.gclId                      -0.012804
trafficSource.adwordsClickInfo.isVideoAd                  -0.011867
trafficSource.adwordsClickInfo.page                       -0.011853
trafficSource.adwordsClickInfo.slot                       -0.012724
trafficSource.campaign                                     0.017587
trafficSource.isTrueDirect                                 0.115301
trafficSource.keyword                                     -0.108637
trafficSource.medium                                      0.044304
trafficSource.referralPath                               0.186288
trafficSource.source                                     0.034500
2017                                                     -0.123690
marketType                                               0.149249
dayOfWeek                                                 0.004551


                                                 geoNetwork.subContinent  \
channelGrouping                                          0.062654
date                                                    -0.040016
socialEngagementType                                    0.012515
visitNumber                                            -0.025541
visitStartTime                                         -0.034235
device.browser                                          0.032524
device.browserSize                                      0.012515
device.browserVersion                                   0.012515
device.deviceCategory                                  -0.014935
device.flashVersion                                     0.012515
device.isMobile                                        -0.008374
device.language                                         0.012515
device.mobileDeviceBranding                             0.012515
device.mobileDeviceInfo                                 0.012515
device.mobileDeviceMarketingName                        0.012515
device.mobileDeviceModel                                0.012515
device.mobileInputSelector                              0.012515
device.operatingSystem                                 -0.018006
```

```
device.operatingSystemVersion                              0.012515
device.screenColors                                        0.012515
device.screenResolution                                    0.012515
geoNetwork.city                                           -0.030103
geoNetwork.cityId                                          0.012515
geoNetwork.continent                                       0.201393
geoNetwork.country                                        -0.118254
geoNetwork.latitude                                        0.012515
geoNetwork.longitude                                       0.012515
geoNetwork.metro                                          -0.057936
geoNetwork.networkDomain                                   0.130431
geoNetwork.networkLocation                                 0.012515
geoNetwork.region                                          0.004792
geoNetwork.subContinent                                    1.000000
totals.bounces                                             0.072579
totals.hits                                               -0.062258
totals.newVisits                                           0.066573
totals.pageviews                                          -0.066021
totals.transactionRevenue                                -0.009146
totals.visits                                             0.012515
trafficSource.adContent                                   0.008051
trafficSource.adwordsClickInfo.adNetworkType              0.041720
trafficSource.adwordsClickInfo.criteriaParameters         0.012515
trafficSource.adwordsClickInfo.gclId                      0.036448
trafficSource.adwordsClickInfo.isVideoAd                  0.041722
trafficSource.adwordsClickInfo.page                       0.041711
trafficSource.adwordsClickInfo.slot                       0.041264
trafficSource.campaign                                    -0.018876
trafficSource.isTrueDirect                                 0.070414
trafficSource.keyword                                     -0.010269
trafficSource.medium                                      0.057555
trafficSource.referralPath                                0.053611
trafficSource.source                                      0.077840
2017                                                     -0.091829
marketType                                                0.244379
dayOfWeek                                                 0.012515

                                          totals.bounces  \
channelGrouping                                 0.063994
date                                            0.002479
socialEngagementType                           -0.006027
visitNumber                                    -0.023745
visitStartTime                                  0.004158
device.browser                                  0.064047
device.browserSize                             -0.006027
device.browserVersion                          -0.006027
device.deviceCategory                           0.013961
device.flashVersion                            -0.006027
```

```
device.isMobile                                       0.021244
device.language                                      -0.006027
device.mobileDeviceBranding                          -0.006027
device.mobileDeviceInfo                              -0.006027
device.mobileDeviceMarketingName                     -0.006027
device.mobileDeviceModel                             -0.006027
device.mobileInputSelector                           -0.006027
device.operatingSystem                                0.071871
device.operatingSystemVersion                        -0.006027
device.screenColors                                  -0.006027
device.screenResolution                              -0.006027
geoNetwork.city                                       0.006648
geoNetwork.cityId                                    -0.006027
geoNetwork.continent                                  0.176631
geoNetwork.country                                   -0.138695
geoNetwork.latitude                                  -0.006027
geoNetwork.longitude                                 -0.006027
geoNetwork.metro                                     -0.041629
geoNetwork.networkDomain                              0.127310
geoNetwork.networkLocation                           -0.006027
geoNetwork.region                                     0.067495
geoNetwork.subContinent                               0.072579
totals.bounces                                        1.000000
totals.hits                                          -0.371110
totals.newVisits                                      0.078903
totals.pageviews                                     -0.404498
totals.transactionRevenue                            -0.032204
totals.visits                                        -0.006027
trafficSource.adContent                              -0.000703
trafficSource.adwordsClickInfo.adNetworkType          0.052667
trafficSource.adwordsClickInfo.criteriaParameters    -0.006027
trafficSource.adwordsClickInfo.gclId                  0.045179
trafficSource.adwordsClickInfo.isVideoAd              0.052635
trafficSource.adwordsClickInfo.page                   0.052723
trafficSource.adwordsClickInfo.slot                   0.050303
trafficSource.campaign                               -0.017127
trafficSource.isTrueDirect                            0.044522
trafficSource.keyword                                 0.051043
trafficSource.medium                                  0.018919
trafficSource.referralPath                            0.127284
trafficSource.source                                  0.082105
2017                                                 -0.133959
marketType                                            0.257215
dayOfWeek                                            -0.006027

                                                     totals.hits  \
channelGrouping                                      -0.067364
date                                                 -0.018395
```

```
socialEngagementType                                      -0.002881
visitNumber                                                0.041321
visitStartTime                                            -0.017608
device.browser                                           -0.075645
device.browserSize                                       -0.002881
device.browserVersion                                    -0.002881
device.deviceCategory                                    -0.019248
device.flashVersion                                      -0.002881
device.isMobile                                          -0.030426
device.language                                          -0.002881
device.mobileDeviceBranding                              -0.002881
device.mobileDeviceInfo                                  -0.002881
device.mobileDeviceMarketingName                         -0.002881
device.mobileDeviceModel                                 -0.002881
device.mobileInputSelector                               -0.002881
device.operatingSystem                                   -0.057431
device.operatingSystemVersion                            -0.002881
device.screenColors                                      -0.002881
device.screenResolution                                  -0.002881
geoNetwork.city                                           0.000201
geoNetwork.cityId                                        -0.002881
geoNetwork.continent                                     -0.134314
geoNetwork.country                                        0.102283
geoNetwork.latitude                                      -0.002881
geoNetwork.longitude                                     -0.002881
geoNetwork.metro                                          0.035971
geoNetwork.networkDomain                                 -0.099445
geoNetwork.networkLocation                               -0.002881
geoNetwork.region                                        -0.039878
geoNetwork.subContinent                                  -0.062258
totals.bounces                                           -0.371110
totals.hits                                               1.000000
totals.newVisits                                         -0.115162
totals.pageviews                                          0.983201
totals.transactionRevenue                                0.154332
totals.visits                                            -0.002881
trafficSource.adContent                                  -0.000842
trafficSource.adwordsClickInfo.adNetworkType             -0.040489
trafficSource.adwordsClickInfo.criteriaParameters        -0.002881
trafficSource.adwordsClickInfo.gclId                     -0.035042
trafficSource.adwordsClickInfo.isVideoAd                 -0.040475
trafficSource.adwordsClickInfo.page                      -0.040541
trafficSource.adwordsClickInfo.slot                      -0.038758
trafficSource.campaign                                    0.004590
trafficSource.isTrueDirect                               -0.083970
trafficSource.keyword                                    -0.046718
trafficSource.medium                                     -0.027634
trafficSource.referralPath                               -0.102781
```

```
trafficSource.source                                     -0.080420
2017                                                      0.097031
marketType                                              -0.196305
dayOfWeek                                               -0.002881


                                                       totals.newVisits  \
channelGrouping                                            0.082264
date                                                     -0.043902
socialEngagementType                                    -0.011313
visitNumber                                             -0.255044
visitStartTime                                          -0.021115
device.browser                                           0.124154
device.browserSize                                      -0.011313
device.browserVersion                                   -0.011313
device.deviceCategory                                    0.018151
device.flashVersion                                     -0.011313
device.isMobile                                          0.021380
device.language                                         -0.011313
device.mobileDeviceBranding                             -0.011313
device.mobileDeviceInfo                                 -0.011313
device.mobileDeviceMarketingName                        -0.011313
device.mobileDeviceModel                                -0.011313
device.mobileInputSelector                              -0.011313
device.operatingSystem                                   0.086957
device.operatingSystemVersion                           -0.011313
device.screenColors                                     -0.011313
device.screenResolution                                 -0.011313
geoNetwork.city                                          0.036183
geoNetwork.cityId                                       -0.011313
geoNetwork.continent                                     0.142905
geoNetwork.country                                      -0.132127
geoNetwork.latitude                                     -0.011313
geoNetwork.longitude                                    -0.011313
geoNetwork.metro                                        -0.011092
geoNetwork.networkDomain                                 0.148315
geoNetwork.networkLocation                              -0.011313
geoNetwork.region                                        0.106419
geoNetwork.subContinent                                  0.066573
totals.bounces                                           0.078903
totals.hits                                             -0.115162
totals.newVisits                                         1.000000
totals.pageviews                                        -0.117104
totals.transactionRevenue                               -0.041159
totals.visits                                           -0.011313
trafficSource.adContent                                  0.068379
trafficSource.adwordsClickInfo.adNetworkType             0.039858
trafficSource.adwordsClickInfo.criteriaParameters       -0.011313
trafficSource.adwordsClickInfo.gclId                     0.030189
```

```
trafficSource.adwordsClickInfo.isVideoAd                 0.039866
trafficSource.adwordsClickInfo.page                      0.039889
trafficSource.adwordsClickInfo.slot                      0.038405
trafficSource.campaign                                  -0.036202
trafficSource.isTrueDirect                               0.615951
trafficSource.keyword                                   -0.008296
trafficSource.medium                                     0.055883
trafficSource.referralPath                               0.181201
trafficSource.source                                     0.123211
2017                                                    -0.126007
marketType                                               0.230105
dayOfWeek                                               -0.011313


                                              totals.pageviews  \
channelGrouping                                      -0.071030
date                                                 -0.012146
socialEngagementType                                 -0.003311
visitNumber                                           0.043337
visitStartTime                                       -0.013448
device.browser                                       -0.079099
device.browserSize                                   -0.003311
device.browserVersion                                -0.003311
device.deviceCategory                                -0.018119
device.flashVersion                                  -0.003311
device.isMobile                                      -0.030250
device.language                                      -0.003311
device.mobileDeviceBranding                          -0.003311
device.mobileDeviceInfo                              -0.003311
device.mobileDeviceMarketingName                     -0.003311
device.mobileDeviceModel                             -0.003311
device.mobileInputSelector                           -0.003311
device.operatingSystem                               -0.059827
device.operatingSystemVersion                        -0.003311
device.screenColors                                  -0.003311
device.screenResolution                              -0.003311
geoNetwork.city                                       0.000804
geoNetwork.cityId                                    -0.003311
geoNetwork.continent                                 -0.143661
geoNetwork.country                                    0.109547
geoNetwork.latitude                                  -0.003311
geoNetwork.longitude                                 -0.003311
geoNetwork.metro                                      0.039147
geoNetwork.networkDomain                             -0.106702
geoNetwork.networkLocation                           -0.003311
geoNetwork.region                                    -0.043175
geoNetwork.subContinent                              -0.066021
totals.bounces                                       -0.404498
totals.hits                                           0.983201
```

```
totals.newVisits                                      -0.117104
totals.pageviews                                       1.000000
totals.transactionRevenue                              0.155589
totals.visits                                         -0.003311
trafficSource.adContent                               -0.001107
trafficSource.adwordsClickInfo.adNetworkType          -0.044694
trafficSource.adwordsClickInfo.criteriaParameters     -0.003311
trafficSource.adwordsClickInfo.gclId                  -0.038596
trafficSource.adwordsClickInfo.isVideoAd              -0.044679
trafficSource.adwordsClickInfo.page                   -0.044750
trafficSource.adwordsClickInfo.slot                   -0.042823
trafficSource.campaign                                 0.004792
trafficSource.isTrueDirect                            -0.084334
trafficSource.keyword                                 -0.050762
trafficSource.medium                                  -0.028537
trafficSource.referralPath                            -0.109132
trafficSource.source                                  -0.085482
2017                                                   0.102916
marketType                                            -0.209779
dayOfWeek                                             -0.003311


                                   totals.transactionRevenue  \
channelGrouping                                    -0.006646
date                                                0.003188
socialEngagementType                               -0.000195
visitNumber                                         0.051367
visitStartTime                                      0.002069
device.browser                                     -0.015087
device.browserSize                                 -0.000195
device.browserVersion                              -0.000195
device.deviceCategory                              -0.013956
device.flashVersion                                -0.000195
device.isMobile                                    -0.016554
device.language                                    -0.000195
device.mobileDeviceBranding                        -0.000195
device.mobileDeviceInfo                            -0.000195
device.mobileDeviceMarketingName                   -0.000195
device.mobileDeviceModel                           -0.000195
device.mobileInputSelector                         -0.000195
device.operatingSystem                             -0.011926
device.operatingSystemVersion                      -0.000195
device.screenColors                                -0.000195
device.screenResolution                            -0.000195
geoNetwork.city                                    -0.003329
geoNetwork.cityId                                  -0.000195
geoNetwork.continent                               -0.025521
geoNetwork.country                                  0.022577
geoNetwork.latitude                                -0.000195
```

```
geoNetwork.longitude                                     -0.000195
geoNetwork.metro                                          0.003988
geoNetwork.networkDomain                                 -0.020199
geoNetwork.networkLocation                               -0.000195
geoNetwork.region                                        -0.006862
geoNetwork.subContinent                                  -0.009146
totals.bounces                                           -0.032204
totals.hits                                               0.154332
totals.newVisits                                         -0.041159
totals.pageviews                                          0.155589
totals.transactionRevenue                                 1.000000
totals.visits                                            -0.000195
trafficSource.adContent                                   0.000635
trafficSource.adwordsClickInfo.adNetworkType             -0.000835
trafficSource.adwordsClickInfo.criteriaParameters        -0.000195
trafficSource.adwordsClickInfo.gclId                     -0.000797
trafficSource.adwordsClickInfo.isVideoAd                 -0.000834
trafficSource.adwordsClickInfo.page                      -0.000843
trafficSource.adwordsClickInfo.slot                      -0.000750
trafficSource.campaign                                   -0.003364
trafficSource.isTrueDirect                               -0.030819
trafficSource.keyword                                     0.012867
trafficSource.medium                                     -0.009195
trafficSource.referralPath                               -0.026764
trafficSource.source                                     -0.008539
2017                                                      0.014508
marketType                                               -0.034688
dayOfWeek                                                -0.000195

                                        totals.visits  \
channelGrouping                              0.011466
date                                              NaN
socialEngagementType                         1.000000
visitNumber                                 -0.001474
visitStartTime                              -0.705206
device.browser                               0.026293
device.browserSize                           1.000000
device.browserVersion                        1.000000
device.deviceCategory                        0.015478
device.flashVersion                          1.000000
device.isMobile                              0.023766
device.language                              1.000000
device.mobileDeviceBranding                  1.000000
device.mobileDeviceInfo                      1.000000
device.mobileDeviceMarketingName             1.000000
device.mobileDeviceModel                     1.000000
device.mobileInputSelector                   1.000000
device.operatingSystem                       0.008183
```

```
device.operatingSystemVersion                          1.000000
device.screenColors                                    1.000000
device.screenResolution                                1.000000
geoNetwork.city                                        0.004305
geoNetwork.cityId                                      1.000000
geoNetwork.continent                                   0.022398
geoNetwork.country                                     0.005586
geoNetwork.latitude                                    1.000000
geoNetwork.longitude                                   1.000000
geoNetwork.metro                                       0.004434
geoNetwork.networkDomain                               0.002657
geoNetwork.networkLocation                             1.000000
geoNetwork.region                                      0.004551
geoNetwork.subContinent                                0.012515
totals.bounces                                        -0.006027
totals.hits                                           -0.002881
totals.newVisits                                      -0.011313
totals.pageviews                                      -0.003311
totals.transactionRevenue                             -0.000195
totals.visits                                          1.000000
trafficSource.adContent                                0.000622
trafficSource.adwordsClickInfo.adNetworkType           0.000942
trafficSource.adwordsClickInfo.criteriaParameters      1.000000
trafficSource.adwordsClickInfo.gclId                   0.000823
trafficSource.adwordsClickInfo.isVideoAd               0.000942
trafficSource.adwordsClickInfo.page                    0.000942
trafficSource.adwordsClickInfo.slot                    0.000932
trafficSource.campaign                                 0.117426
trafficSource.isTrueDirect                             0.003986
trafficSource.keyword                                  0.005269
trafficSource.medium                                  -0.000602
trafficSource.referralPath                             0.003173
trafficSource.source                                   0.004904
2017                                                  -0.005053
marketType                                             0.005273
dayOfWeek                                              1.000000


                                           trafficSource.adContent  \
channelGrouping                                          -0.021286
date                                                     -0.051095
socialEngagementType                                      0.000622
visitNumber                                              -0.033052
visitStartTime                                           -0.033411
device.browser                                           -0.018506
device.browserSize                                        0.000622
device.browserVersion                                     0.000622
device.deviceCategory                                    -0.072723
device.flashVersion                                       0.000622
```

device.isMobile	-0.076653
device.language	0.000622
device.mobileDeviceBranding	0.000622
device.mobileDeviceInfo	0.000622
device.mobileDeviceMarketingName	0.000622
device.mobileDeviceModel	0.000622
device.mobileInputSelector	0.000622
device.operatingSystem	0.002792
device.operatingSystemVersion	0.000622
device.screenColors	0.000622
device.screenResolution	0.000622
geoNetwork.city	-0.020175
geoNetwork.cityId	0.000622
geoNetwork.continent	0.048196
geoNetwork.country	-0.038088
geoNetwork.latitude	0.000622
geoNetwork.longitude	0.000622
geoNetwork.metro	-0.026371
geoNetwork.networkDomain	0.013745
geoNetwork.networkLocation	0.000622
geoNetwork.region	-0.014586
geoNetwork.subContinent	0.008051
totals.bounces	-0.000703
totals.hits	-0.000842
totals.newVisits	0.068379
totals.pageviews	-0.001107
totals.transactionRevenue	0.000635
totals.visits	0.000622
trafficSource.adContent	1.000000
trafficSource.adwordsClickInfo.adNetworkType	0.347390
trafficSource.adwordsClickInfo.criteriaParameters	0.000622
trafficSource.adwordsClickInfo.gclId	0.278785
trafficSource.adwordsClickInfo.isVideoAd	0.347298
trafficSource.adwordsClickInfo.page	0.347470
trafficSource.adwordsClickInfo.slot	0.341126
trafficSource.campaign	-0.196791
trafficSource.isTrueDirect	-0.015821
trafficSource.keyword	0.073362
trafficSource.medium	0.093777
trafficSource.referralPath	-0.054012
trafficSource.source	0.028720
2017	-0.030086
marketType	0.061509
dayOfWeek	0.000622

trafficSource.adwordsClickInfo.adN

channelGrouping
date

```
socialEngagementType
visitNumber
visitStartTime
device.browser
device.browserSize
device.browserVersion
device.deviceCategory
device.flashVersion
device.isMobile
device.language
device.mobileDeviceBranding
device.mobileDeviceInfo
device.mobileDeviceMarketingName
device.mobileDeviceModel
device.mobileInputSelector
device.operatingSystem
device.operatingSystemVersion
device.screenColors
device.screenResolution
geoNetwork.city
geoNetwork.cityId
geoNetwork.continent
geoNetwork.country
geoNetwork.latitude
geoNetwork.longitude
geoNetwork.metro
geoNetwork.networkDomain
geoNetwork.networkLocation
geoNetwork.region
geoNetwork.subContinent
totals.bounces
totals.hits
totals.newVisits
totals.pageviews
totals.transactionRevenue
totals.visits
trafficSource.adContent
trafficSource.adwordsClickInfo.adNetworkType
trafficSource.adwordsClickInfo.criteriaParameters
trafficSource.adwordsClickInfo.gclId
trafficSource.adwordsClickInfo.isVideoAd
trafficSource.adwordsClickInfo.page
trafficSource.adwordsClickInfo.slot
trafficSource.campaign
trafficSource.isTrueDirect
trafficSource.keyword
trafficSource.medium
trafficSource.referralPath
```

```
trafficSource.source
2017
marketType
dayOfWeek


                                                          trafficSource.adwordsClickInfo.crit

channelGrouping
date
socialEngagementType
visitNumber
visitStartTime
device.browser
device.browserSize
device.browserVersion
device.deviceCategory
device.flashVersion
device.isMobile
device.language
device.mobileDeviceBranding
device.mobileDeviceInfo
device.mobileDeviceMarketingName
device.mobileDeviceModel
device.mobileInputSelector
device.operatingSystem
device.operatingSystemVersion
device.screenColors
device.screenResolution
geoNetwork.city
geoNetwork.cityId
geoNetwork.continent
geoNetwork.country
geoNetwork.latitude
geoNetwork.longitude
geoNetwork.metro
geoNetwork.networkDomain
geoNetwork.networkLocation
geoNetwork.region
geoNetwork.subContinent
totals.bounces
totals.hits
totals.newVisits
totals.pageviews
totals.transactionRevenue
totals.visits
trafficSource.adContent
trafficSource.adwordsClickInfo.adNetworkType
trafficSource.adwordsClickInfo.criteriaParameters
trafficSource.adwordsClickInfo.gclId
```

```
trafficSource.adwordsClickInfo.isVideoAd
trafficSource.adwordsClickInfo.page
trafficSource.adwordsClickInfo.slot
trafficSource.campaign
trafficSource.isTrueDirect
trafficSource.keyword
trafficSource.medium
trafficSource.referralPath
trafficSource.source
2017
marketType
dayOfWeek
```

	trafficSource.adwordsClickInfo.gcl...
channelGrouping	-0.02820
date	-0.00597
socialEngagementType	0.00082
visitNumber	0.00032
visitStartTime	-0.00952
device.browser	-0.02486
device.browserSize	0.00082
device.browserVersion	0.00082
device.deviceCategory	-0.06997
device.flashVersion	0.00082
device.isMobile	-0.06730
device.language	0.00082
device.mobileDeviceBranding	0.00082
device.mobileDeviceInfo	0.00082
device.mobileDeviceMarketingName	0.00082
device.mobileDeviceModel	0.00082
device.mobileInputSelector	0.00082
device.operatingSystem	-0.00923
device.operatingSystemVersion	0.00082
device.screenColors	0.00082
device.screenResolution	0.00082
geoNetwork.city	-0.02998
geoNetwork.cityId	0.00082
geoNetwork.continent	0.10457
geoNetwork.country	-0.09893
geoNetwork.latitude	0.00082
geoNetwork.longitude	0.00082
geoNetwork.metro	-0.04918
geoNetwork.networkDomain	0.03892
geoNetwork.networkLocation	0.00082
geoNetwork.region	-0.01280
geoNetwork.subContinent	0.03644
totals.bounces	0.04517
totals.hits	-0.03504

totals.newVisits	0.03018
totals.pageviews	-0.03859
totals.transactionRevenue	-0.00079
totals.visits	0.00082
trafficSource.adContent	0.27878
trafficSource.adwordsClickInfo.adNetworkType	0.86931
trafficSource.adwordsClickInfo.criteriaParameters	0.00082
trafficSource.adwordsClickInfo.gclId	1.00000
trafficSource.adwordsClickInfo.isVideoAd	0.86946
trafficSource.adwordsClickInfo.page	0.86937
trafficSource.adwordsClickInfo.slot	0.85995
trafficSource.campaign	-0.36545
trafficSource.isTrueDirect	-0.04004
trafficSource.keyword	0.11904
trafficSource.medium	0.12411
trafficSource.referralPath	-0.07148
trafficSource.source	0.03800
2017	-0.06641
marketType	0.14710
dayOfWeek	0.00082

	trafficSource.adwordsClickInfo.isV...
channelGrouping	-0.0
date	-0.0
socialEngagementType	0.0
visitNumber	-0.0
visitStartTime	-0.0
device.browser	-0.0
device.browserSize	0.0
device.browserVersion	0.0
device.deviceCategory	-0.0
device.flashVersion	0.0
device.isMobile	-0.0
device.language	0.0
device.mobileDeviceBranding	0.0
device.mobileDeviceInfo	0.0
device.mobileDeviceMarketingName	0.0
device.mobileDeviceModel	0.0
device.mobileInputSelector	0.0
device.operatingSystem	0.0
device.operatingSystemVersion	0.0
device.screenColors	0.0
device.screenResolution	0.0
geoNetwork.city	-0.0
geoNetwork.cityId	0.0
geoNetwork.continent	0.1
geoNetwork.country	-0.1
geoNetwork.latitude	0.0

geoNetwork.longitude	0.(
geoNetwork.metro	-0.(
geoNetwork.networkDomain	0.(
geoNetwork.networkLocation	0.(
geoNetwork.region	-0.(
geoNetwork.subContinent	0.(
totals.bounces	0.(
totals.hits	-0.(
totals.newVisits	0.(
totals.pageviews	-0.(
totals.transactionRevenue	-0.(
totals.visits	0.(
trafficSource.adContent	0.3
trafficSource.adwordsClickInfo.adNetworkType	0.9
trafficSource.adwordsClickInfo.criteriaParameters	0.(
trafficSource.adwordsClickInfo.gclId	0.8
trafficSource.adwordsClickInfo.isVideoAd	1.(
trafficSource.adwordsClickInfo.page	0.9
trafficSource.adwordsClickInfo.slot	0.9
trafficSource.campaign	-0.4
trafficSource.isTrueDirect	-0.(
trafficSource.keyword	0.1
trafficSource.medium	0.1
trafficSource.referralPath	-0.(
trafficSource.source	0.(
2017	-0.(
marketType	0.1
dayOfWeek	0.(

	trafficSource.adwordsClickInfo.page
channelGrouping	-0.03229
date	-0.01890
socialEngagementType	0.00094
visitNumber	-0.00134
visitStartTime	-0.02165
device.browser	-0.00920
device.browserSize	0.00094
device.browserVersion	0.00094
device.deviceCategory	-0.07438
device.flashVersion	0.00094
device.isMobile	-0.07714
device.language	0.00094
device.mobileDeviceBranding	0.00094
device.mobileDeviceInfo	0.00094
device.mobileDeviceMarketingName	0.00094
device.mobileDeviceModel	0.00094
device.mobileInputSelector	0.00094
device.operatingSystem	0.00780

device.operatingSystemVersion	0.00094
device.screenColors	0.00094
device.screenResolution	0.00094
geoNetwork.city	-0.034039
geoNetwork.cityId	0.00094
geoNetwork.continent	0.119328
geoNetwork.country	-0.112794
geoNetwork.latitude	0.00094
geoNetwork.longitude	0.00094
geoNetwork.metro	-0.055834
geoNetwork.networkDomain	0.046459
geoNetwork.networkLocation	0.00094
geoNetwork.region	-0.011853
geoNetwork.subContinent	0.04171
totals.bounces	0.05272
totals.hits	-0.04054
totals.newVisits	0.039889
totals.pageviews	-0.04475
totals.transactionRevenue	-0.00084
totals.visits	0.00094
trafficSource.adContent	0.34747
trafficSource.adwordsClickInfo.adNetworkType	0.99966
trafficSource.adwordsClickInfo.criteriaParameters	0.00094
trafficSource.adwordsClickInfo.gclId	0.86937
trafficSource.adwordsClickInfo.isVideoAd	0.99970
trafficSource.adwordsClickInfo.page	1.00000
trafficSource.adwordsClickInfo.slot	0.98800
trafficSource.campaign	-0.41405
trafficSource.isTrueDirect	-0.043419
trafficSource.keyword	0.13441
trafficSource.medium	0.14212
trafficSource.referralPath	-0.081859
trafficSource.source	0.04351
2017	-0.075415
marketType	0.167630
dayOfWeek	0.00094

	trafficSource.adwordsClickInfo.slot
channelGrouping	-0.031947
date	-0.018385
socialEngagementType	0.00093
visitNumber	-0.000960
visitStartTime	-0.021203
device.browser	-0.00973
device.browserSize	0.00093
device.browserVersion	0.00093
device.deviceCategory	-0.076720
device.flashVersion	0.00093

device.isMobile	-0.077914
device.language	0.000932
device.mobileDeviceBranding	0.000932
device.mobileDeviceInfo	0.000932
device.mobileDeviceMarketingName	0.000932
device.mobileDeviceModel	0.000932
device.mobileInputSelector	0.000932
device.operatingSystem	0.007880
device.operatingSystemVersion	0.000932
device.screenColors	0.000932
device.screenResolution	0.000932
geoNetwork.city	-0.034269
geoNetwork.cityId	0.000932
geoNetwork.continent	0.118171
geoNetwork.country	-0.111787
geoNetwork.latitude	0.000932
geoNetwork.longitude	0.000932
geoNetwork.metro	-0.055629
geoNetwork.networkDomain	0.046118
geoNetwork.networkLocation	0.000932
geoNetwork.region	-0.012724
geoNetwork.subContinent	0.041264
totals.bounces	0.050303
totals.hits	-0.038758
totals.newVisits	0.038405
totals.pageviews	-0.042823
totals.transactionRevenue	-0.000750
totals.visits	0.000932
trafficSource.adContent	0.341126
trafficSource.adwordsClickInfo.adNetworkType	0.988773
trafficSource.adwordsClickInfo.criteriaParameters	0.000932
trafficSource.adwordsClickInfo.gclId	0.859951
trafficSource.adwordsClickInfo.isVideoAd	0.988972
trafficSource.adwordsClickInfo.page	0.988003
trafficSource.adwordsClickInfo.slot	1.000000
trafficSource.campaign	-0.409520
trafficSource.isTrueDirect	-0.043783
trafficSource.keyword	0.133024
trafficSource.medium	0.140594
trafficSource.referralPath	-0.080980
trafficSource.source	0.043050
2017	-0.074703
marketType	0.166017
dayOfWeek	0.000932

	trafficSource.campaign \
channelGrouping	-0.201301
date	0.013333

socialEngagementType	0.117426
visitNumber	-0.004429
visitStartTime	-0.073612
device.browser	-0.027731
device.browserSize	0.117426
device.browserVersion	0.117426
device.deviceCategory	0.001642
device.flashVersion	0.117426
device.isMobile	-0.002829
device.language	0.117426
device.mobileDeviceBranding	0.117426
device.mobileDeviceInfo	0.117426
device.mobileDeviceMarketingName	0.117426
device.mobileDeviceModel	0.117426
device.mobileInputSelector	0.117426
device.operatingSystem	0.028353
device.operatingSystemVersion	0.117426
device.screenColors	0.117426
device.screenResolution	0.117426
geoNetwork.city	0.007393
geoNetwork.cityId	0.117426
geoNetwork.continent	-0.013295
geoNetwork.country	0.018427
geoNetwork.latitude	0.117426
geoNetwork.longitude	0.117426
geoNetwork.metro	-0.002450
geoNetwork.networkDomain	0.004526
geoNetwork.networkLocation	0.117426
geoNetwork.region	0.017587
geoNetwork.subContinent	-0.018876
totals.bounces	-0.017127
totals.hits	0.004590
totals.newVisits	-0.036202
totals.pageviews	0.004792
totals.transactionRevenue	-0.003364
totals.visits	0.117426
trafficSource.adContent	-0.196791
trafficSource.adwordsClickInfo.adNetworkType	-0.414174
trafficSource.adwordsClickInfo.criteriaParameters	0.117426
trafficSource.adwordsClickInfo.gclId	-0.365457
trafficSource.adwordsClickInfo.isVideoAd	-0.414190
trafficSource.adwordsClickInfo.page	-0.414052
trafficSource.adwordsClickInfo.slot	-0.409520
trafficSource.campaign	1.000000
trafficSource.isTrueDirect	0.050308
trafficSource.keyword	0.020013
trafficSource.medium	-0.211061
trafficSource.referralPath	0.096924

```
trafficSource.source                                         -0.132040
2017                                                          0.033423
marketType                                                  -0.031533
dayOfWeek                                                    0.117426


                                               trafficSource.isTrueDirect  \
channelGrouping                                             0.467864
date                                                       -0.083313
socialEngagementType                                       0.003986
visitNumber                                               -0.181696
visitStartTime                                            -0.054702
device.browser                                             0.106361
device.browserSize                                         0.003986
device.browserVersion                                      0.003986
device.deviceCategory                                     -0.056104
device.flashVersion                                        0.003986
device.isMobile                                           -0.078572
device.language                                            0.003986
device.mobileDeviceBranding                                0.003986
device.mobileDeviceInfo                                    0.003986
device.mobileDeviceMarketingName                           0.003986
device.mobileDeviceModel                                   0.003986
device.mobileInputSelector                                 0.003986
device.operatingSystem                                     0.114941
device.operatingSystemVersion                              0.003986
device.screenColors                                        0.003986
device.screenResolution                                    0.003986
geoNetwork.city                                            0.066719
geoNetwork.cityId                                          0.003986
geoNetwork.continent                                       0.140148
geoNetwork.country                                        -0.115119
geoNetwork.latitude                                        0.003986
geoNetwork.longitude                                       0.003986
geoNetwork.metro                                           0.018030
geoNetwork.networkDomain                                   0.136705
geoNetwork.networkLocation                                 0.003986
geoNetwork.region                                          0.115301
geoNetwork.subContinent                                    0.070414
totals.bounces                                             0.044522
totals.hits                                               -0.083970
totals.newVisits                                           0.615951
totals.pageviews                                          -0.084334
totals.transactionRevenue                                 -0.030819
totals.visits                                              0.003986
trafficSource.adContent                                   -0.015821
trafficSource.adwordsClickInfo.adNetworkType              -0.043484
trafficSource.adwordsClickInfo.criteriaParameters          0.003986
trafficSource.adwordsClickInfo.gclId                      -0.040047
```

```
trafficSource.adwordsClickInfo.isVideoAd              -0.043495
trafficSource.adwordsClickInfo.page                   -0.043419
trafficSource.adwordsClickInfo.slot                   -0.043783
trafficSource.campaign                                 0.050308
trafficSource.isTrueDirect                             1.000000
trafficSource.keyword                                 -0.228375
trafficSource.medium                                   0.613083
trafficSource.referralPath                            -0.018384
trafficSource.source                                   0.496873
2017                                                  -0.139147
marketType                                             0.224356
dayOfWeek                                              0.003986


                                        trafficSource.keyword  \
channelGrouping                                        0.282834
date                                                  -0.158028
socialEngagementType                                   0.005269
visitNumber                                            0.003198
visitStartTime                                        -0.114066
device.browser                                         0.005992
device.browserSize                                     0.005269
device.browserVersion                                  0.005269
device.deviceCategory                                 -0.175726
device.flashVersion                                    0.005269
device.isMobile                                       -0.193968
device.language                                        0.005269
device.mobileDeviceBranding                            0.005269
device.mobileDeviceInfo                                0.005269
device.mobileDeviceMarketingName                       0.005269
device.mobileDeviceModel                               0.005269
device.mobileInputSelector                             0.005269
device.operatingSystem                                -0.092660
device.operatingSystemVersion                          0.005269
device.screenColors                                    0.005269
device.screenResolution                                0.005269
geoNetwork.city                                       -0.110182
geoNetwork.cityId                                      0.005269
geoNetwork.continent                                  -0.065934
geoNetwork.country                                     0.017918
geoNetwork.latitude                                    0.005269
geoNetwork.longitude                                   0.005269
geoNetwork.metro                                      -0.077269
geoNetwork.networkDomain                               0.012183
geoNetwork.networkLocation                             0.005269
geoNetwork.region                                     -0.108637
geoNetwork.subContinent                               -0.010269
totals.bounces                                         0.051043
totals.hits                                           -0.046718
```

```
totals.newVisits                                       -0.008296
totals.pageviews                                       -0.050762
totals.transactionRevenue                               0.012867
totals.visits                                           0.005269
trafficSource.adContent                                 0.073362
trafficSource.adwordsClickInfo.adNetworkType            0.134476
trafficSource.adwordsClickInfo.criteriaParameters       0.005269
trafficSource.adwordsClickInfo.gclId                    0.119049
trafficSource.adwordsClickInfo.isVideoAd                0.134487
trafficSource.adwordsClickInfo.page                     0.134412
trafficSource.adwordsClickInfo.slot                     0.133024
trafficSource.campaign                                  0.020013
trafficSource.isTrueDirect                             -0.228375
trafficSource.keyword                                   1.000000
trafficSource.medium                                   -0.206920
trafficSource.referralPath                             -0.457794
trafficSource.source                                    0.239420
2017                                                   -0.131276
marketType                                              0.079308
dayOfWeek                                               0.005269

                                            trafficSource.medium  \
channelGrouping                                         0.802630
date                                                   -0.122398
socialEngagementType                                   -0.000602
visitNumber                                            -0.033188
visitStartTime                                         -0.077127
device.browser                                          0.056172
device.browserSize                                     -0.000602
device.browserVersion                                  -0.000602
device.deviceCategory                                  -0.130440
device.flashVersion                                    -0.000602
device.isMobile                                        -0.167293
device.language                                        -0.000602
device.mobileDeviceBranding                            -0.000602
device.mobileDeviceInfo                                -0.000602
device.mobileDeviceMarketingName                       -0.000602
device.mobileDeviceModel                               -0.000602
device.mobileInputSelector                             -0.000602
device.operatingSystem                                  0.065488
device.operatingSystemVersion                          -0.000602
device.screenColors                                    -0.000602
device.screenResolution                                -0.000602
geoNetwork.city                                         0.035982
geoNetwork.cityId                                      -0.000602
geoNetwork.continent                                    0.089292
geoNetwork.country                                     -0.069049
geoNetwork.latitude                                    -0.000602
```

```
geoNetwork.longitude                                     -0.000602
geoNetwork.metro                                          0.015569
geoNetwork.networkDomain                                  0.079339
geoNetwork.networkLocation                               -0.000602
geoNetwork.region                                         0.044304
geoNetwork.subContinent                                   0.057555
totals.bounces                                            0.018919
totals.hits                                              -0.027634
totals.newVisits                                          0.055883
totals.pageviews                                         -0.028537
totals.transactionRevenue                                -0.009195
totals.visits                                            -0.000602
trafficSource.adContent                                   0.093777
trafficSource.adwordsClickInfo.adNetworkType              0.142156
trafficSource.adwordsClickInfo.criteriaParameters        -0.000602
trafficSource.adwordsClickInfo.gclId                      0.124112
trafficSource.adwordsClickInfo.isVideoAd                  0.142162
trafficSource.adwordsClickInfo.page                       0.142120
trafficSource.adwordsClickInfo.slot                       0.140594
trafficSource.campaign                                   -0.211061
trafficSource.isTrueDirect                                0.613083
trafficSource.keyword                                    -0.206920
trafficSource.medium                                      1.000000
trafficSource.referralPath                               -0.373959
trafficSource.source                                      0.789478
2017                                                     -0.139850
marketType                                                0.173327
dayOfWeek                                                -0.000602

                                            trafficSource.referralPath  \
channelGrouping                                          -0.509481
date                                                      0.024930
socialEngagementType                                      0.003173
visitNumber                                              -0.029875
visitStartTime                                            0.007839
device.browser                                            0.128266
device.browserSize                                        0.003173
device.browserVersion                                     0.003173
device.deviceCategory                                     0.185196
device.flashVersion                                       0.003173
device.isMobile                                           0.215738
device.language                                           0.003173
device.mobileDeviceBranding                               0.003173
device.mobileDeviceInfo                                   0.003173
device.mobileDeviceMarketingName                          0.003173
device.mobileDeviceModel                                  0.003173
device.mobileInputSelector                                0.003173
device.operatingSystem                                    0.130801
```

```
device.operatingSystemVersion                               0.003173
device.screenColors                                         0.003173
device.screenResolution                                     0.003173
geoNetwork.city                                             0.093064
geoNetwork.cityId                                           0.003173
geoNetwork.continent                                        0.124510
geoNetwork.country                                         -0.114987
geoNetwork.latitude                                         0.003173
geoNetwork.longitude                                        0.003173
geoNetwork.metro                                            0.015635
geoNetwork.networkDomain                                    0.156075
geoNetwork.networkLocation                                  0.003173
geoNetwork.region                                           0.186288
geoNetwork.subContinent                                     0.053611
totals.bounces                                              0.127284
totals.hits                                                -0.102781
totals.newVisits                                            0.181201
totals.pageviews                                           -0.109132
totals.transactionRevenue                                 -0.026764
totals.visits                                               0.003173
trafficSource.adContent                                    -0.054012
trafficSource.adwordsClickInfo.adNetworkType              -0.081880
trafficSource.adwordsClickInfo.criteriaParameters          0.003173
trafficSource.adwordsClickInfo.gclId                      -0.071486
trafficSource.adwordsClickInfo.isVideoAd                  -0.081883
trafficSource.adwordsClickInfo.page                       -0.081859
trafficSource.adwordsClickInfo.slot                       -0.080980
trafficSource.campaign                                      0.096924
trafficSource.isTrueDirect                                -0.018384
trafficSource.keyword                                     -0.457794
trafficSource.medium                                      -0.373959
trafficSource.referralPath                                 1.000000
trafficSource.source                                      -0.357615
2017                                                      -0.050264
marketType                                                 0.141114
dayOfWeek                                                   0.003173

                                              trafficSource.source  \
channelGrouping                                          0.939053
date                                                    -0.262463
socialEngagementType                                     0.004904
visitNumber                                             -0.046518
visitStartTime                                          -0.178594
device.browser                                           0.126462
device.browserSize                                       0.004904
device.browserVersion                                    0.004904
device.deviceCategory                                   -0.184463
device.flashVersion                                      0.004904
```

```
device.isMobile                                            -0.228006
device.language                                             0.004904
device.mobileDeviceBranding                                 0.004904
device.mobileDeviceInfo                                     0.004904
device.mobileDeviceMarketingName                            0.004904
device.mobileDeviceModel                                    0.004904
device.mobileInputSelector                                  0.004904
device.operatingSystem                                      0.042441
device.operatingSystemVersion                               0.004904
device.screenColors                                         0.004904
device.screenResolution                                     0.004904
geoNetwork.city                                             0.003061
geoNetwork.cityId                                           0.004904
geoNetwork.continent                                        0.055218
geoNetwork.country                                         -0.064157
geoNetwork.latitude                                         0.004904
geoNetwork.longitude                                        0.004904
geoNetwork.metro                                           -0.008872
geoNetwork.networkDomain                                    0.131221
geoNetwork.networkLocation                                  0.004904
geoNetwork.region                                           0.034500
geoNetwork.subContinent                                     0.077840
totals.bounces                                              0.082105
totals.hits                                                -0.080420
totals.newVisits                                            0.123211
totals.pageviews                                           -0.085482
totals.transactionRevenue                                 -0.008539
totals.visits                                              0.004904
trafficSource.adContent                                     0.028720
trafficSource.adwordsClickInfo.adNetworkType               0.043528
trafficSource.adwordsClickInfo.criteriaParameters          0.004904
trafficSource.adwordsClickInfo.gclId                       0.038003
trafficSource.adwordsClickInfo.isVideoAd                   0.043530
trafficSource.adwordsClickInfo.page                        0.043517
trafficSource.adwordsClickInfo.slot                        0.043050
trafficSource.campaign                                     -0.132040
trafficSource.isTrueDirect                                 0.496873
trafficSource.keyword                                      0.239420
trafficSource.medium                                       0.789478
trafficSource.referralPath                                -0.357615
trafficSource.source                                       1.000000
2017                                                       -0.250089
marketType                                                 0.251033
dayOfWeek                                                  0.004904

                                             2017  marketType  \
channelGrouping                         -0.224743    0.215656
date                                     0.114128   -0.085404
```

socialEngagementType	-0.005053	0.005273
visitNumber	0.048679	-0.096729
visitStartTime	0.079062	-0.060367
device.browser	-0.064396	0.084578
device.browserSize	-0.005053	0.005273
device.browserVersion	-0.005053	0.005273
device.deviceCategory	0.075270	-0.058623
device.flashVersion	-0.005053	0.005273
device.isMobile	0.090522	-0.069739
device.language	-0.005053	0.005273
device.mobileDeviceBranding	-0.005053	0.005273
device.mobileDeviceInfo	-0.005053	0.005273
device.mobileDeviceMarketingName	-0.005053	0.005273
device.mobileDeviceModel	-0.005053	0.005273
device.mobileInputSelector	-0.005053	0.005273
device.operatingSystem	-0.066450	0.115170
device.operatingSystemVersion	-0.005053	0.005273
device.screenColors	-0.005053	0.005273
device.screenResolution	-0.005053	0.005273
geoNetwork.city	-0.087281	-0.027961
geoNetwork.cityId	-0.005053	0.005273
geoNetwork.continent	-0.086936	0.706282
geoNetwork.country	0.344886	-0.541062
geoNetwork.latitude	-0.005053	0.005273
geoNetwork.longitude	-0.005053	0.005273
geoNetwork.metro	-0.008035	-0.213843
geoNetwork.networkDomain	-0.195755	0.378584
geoNetwork.networkLocation	-0.005053	0.005273
geoNetwork.region	-0.123690	0.149249
geoNetwork.subContinent	-0.091829	0.244379
totals.bounces	-0.133959	0.257215
totals.hits	0.097031	-0.196305
totals.newVisits	-0.126007	0.230105
totals.pageviews	0.102916	-0.209779
totals.transactionRevenue	0.014508	-0.034688
totals.visits	-0.005053	0.005273
trafficSource.adContent	-0.030086	0.061509
trafficSource.adwordsClickInfo.adNetworkType	-0.075434	0.167677
trafficSource.adwordsClickInfo.criteriaParameters	-0.005053	0.005273
trafficSource.adwordsClickInfo.gclId	-0.066412	0.147106
trafficSource.adwordsClickInfo.isVideoAd	-0.075438	0.167686
trafficSource.adwordsClickInfo.page	-0.075415	0.167630
trafficSource.adwordsClickInfo.slot	-0.074703	0.166017
trafficSource.campaign	0.033423	-0.031533
trafficSource.isTrueDirect	-0.139147	0.224356
trafficSource.keyword	-0.131276	0.079308
trafficSource.medium	-0.139850	0.173327
trafficSource.referralPath	-0.050264	0.141114

```
trafficSource.source                                -0.250089    0.251033
2017                                                 1.000000   -0.417797
marketType                                          -0.417797    1.000000
dayOfWeek                                           -0.005053    0.005273


                                                    dayOfWeek
channelGrouping                                      0.011466
date                                                      NaN
socialEngagementType                                 1.000000
visitNumber                                         -0.001474
visitStartTime                                      -0.705206
device.browser                                       0.026293
device.browserSize                                   1.000000
device.browserVersion                                1.000000
device.deviceCategory                                0.015478
device.flashVersion                                  1.000000
device.isMobile                                      0.023766
device.language                                      1.000000
device.mobileDeviceBranding                          1.000000
device.mobileDeviceInfo                              1.000000
device.mobileDeviceMarketingName                     1.000000
device.mobileDeviceModel                             1.000000
device.mobileInputSelector                           1.000000
device.operatingSystem                               0.008183
device.operatingSystemVersion                        1.000000
device.screenColors                                  1.000000
device.screenResolution                              1.000000
geoNetwork.city                                      0.004305
geoNetwork.cityId                                    1.000000
geoNetwork.continent                                 0.022398
geoNetwork.country                                   0.005586
geoNetwork.latitude                                  1.000000
geoNetwork.longitude                                 1.000000
geoNetwork.metro                                     0.004434
geoNetwork.networkDomain                             0.002657
geoNetwork.networkLocation                           1.000000
geoNetwork.region                                    0.004551
geoNetwork.subContinent                              0.012515
totals.bounces                                      -0.006027
totals.hits                                         -0.002881
totals.newVisits                                    -0.011313
totals.pageviews                                    -0.003311
totals.transactionRevenue                           -0.000195
totals.visits                                        1.000000
trafficSource.adContent                              0.000622
trafficSource.adwordsClickInfo.adNetworkType         0.000942
trafficSource.adwordsClickInfo.criteriaParameters    1.000000
trafficSource.adwordsClickInfo.gclId                 0.000823
```

```
trafficSource.adwordsClickInfo.isVideoAd          0.000942
trafficSource.adwordsClickInfo.page               0.000942
trafficSource.adwordsClickInfo.slot               0.000932
trafficSource.campaign                            0.117426
trafficSource.isTrueDirect                        0.003986
trafficSource.keyword                             0.005269
trafficSource.medium                             -0.000602
trafficSource.referralPath                        0.003173
trafficSource.source                              0.004904
2017                                             -0.005053
marketType                                        0.005273
dayOfWeek                                         1.000000
```

In [18]: plt.subplots(figsize=(10,10))
         corr = train.corr(method='pearson')
         sns.heatmap(corr, xticklabels=corr.columns,yticklabels=corr.columns, cmap="YlGnBu")

Out[18]: <matplotlib.axes._subplots.AxesSubplot at 0x1be0360d30>
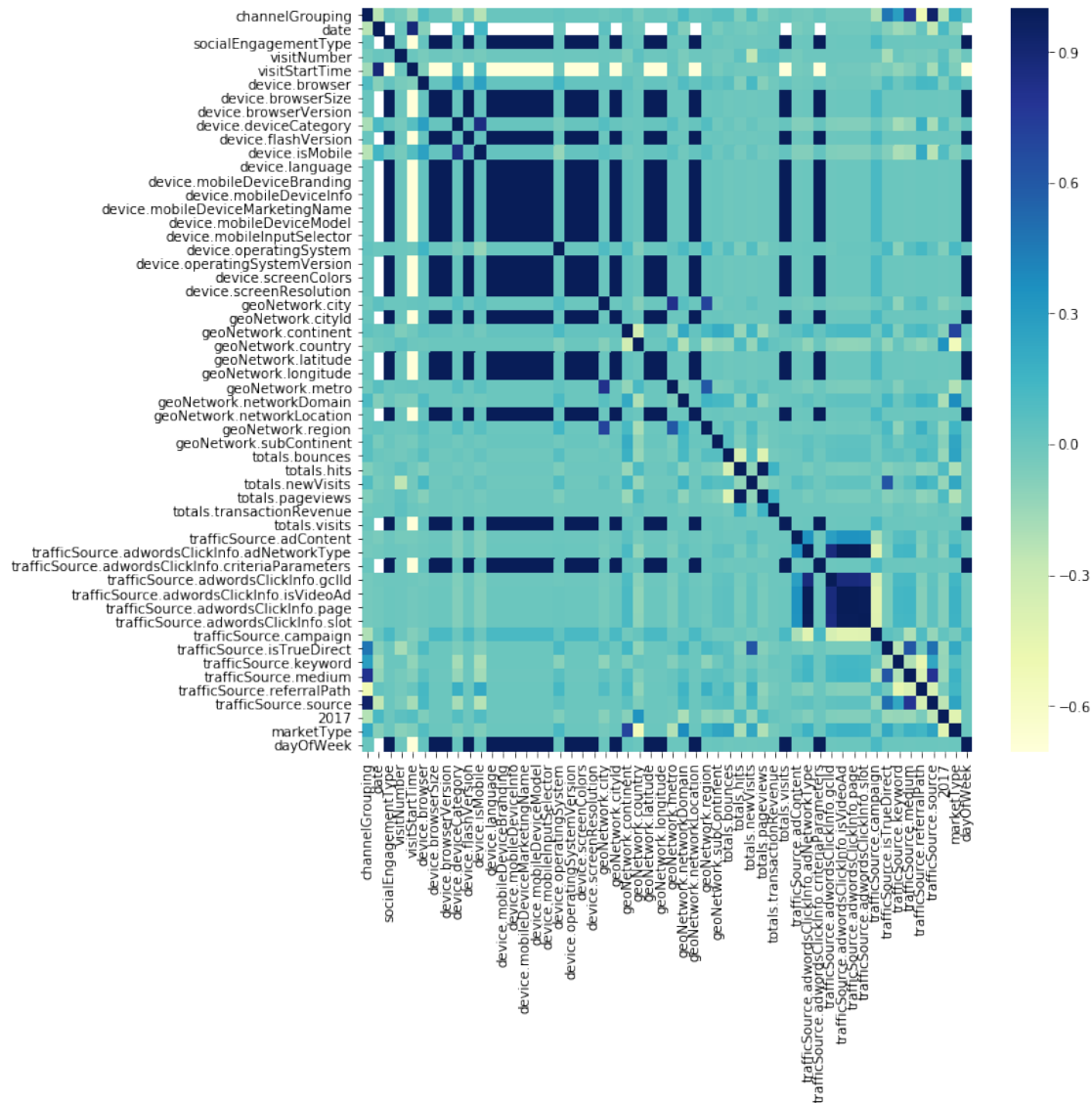
In [28]: sub_train = train[["totals.pageviews", "totals.hits", "geoNetwork.country", "visitNum

         plt.subplots(figsize=(10,10))
         corr = sub_train.corr(method='pearson')
         sns.heatmap(corr, xticklabels=corr.columns,yticklabels=corr.columns, cmap="YlGnBu")

Out[28]: <matplotlib.axes._subplots.AxesSubplot at 0x1a3c551240>
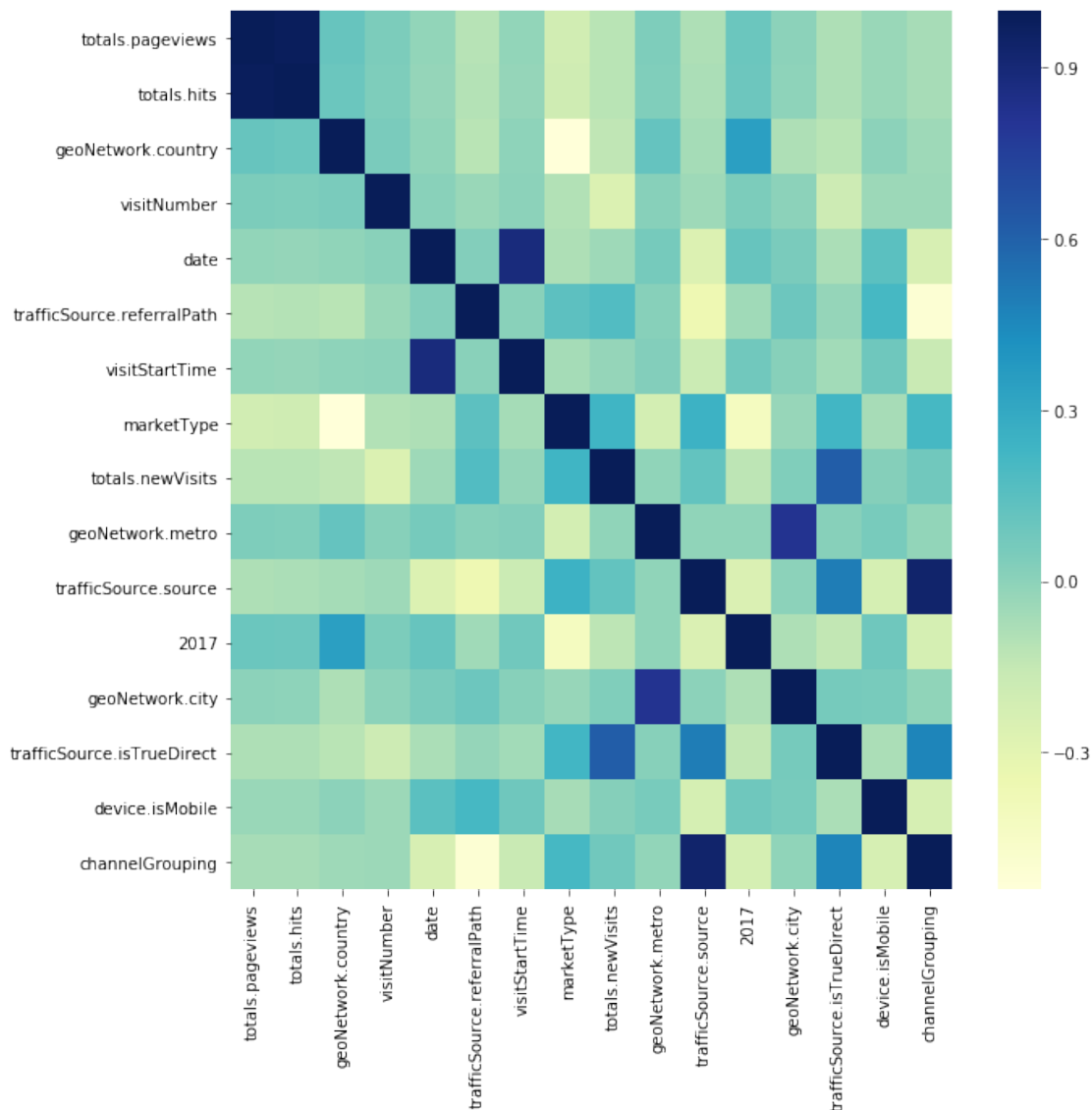
```
In [110]: print(type(lgb_model))

<class 'lightgbm.basic.Booster'>


In [ ]: import numpy as np
        import matplotlib.pyplot as plt
        from sklearn.svm import SVC
        from sklearn.model_selection import StratifiedKFold
        from sklearn.model_selection import permutation_test_score
        from sklearn import datasets

        n_classes = np.unique(train_y).size
```

```python
# Some noisy data not correlated
random = np.random.RandomState(seed=0)
E = random.normal(size=(len(X), 2200))

# Add noisy data to the informative features for make the task harder
#train = np.c_[train, E]

svm = SVC(kernel='linear')
cv = StratifiedKFold(2)

train = np.nan_to_num(train)

score, permutation_scores, pvalue = permutation_test_score(svm, train, train_y, scoring

print("Classification score %s (pvalue : %s)" % (score, pvalue))

# #############################################################################
# View histogram of permutation scores
plt.hist(permutation_scores, 20, label='Permutation scores',
         edgecolor='black')
ylim = plt.ylim()
# BUG: vlines(..., linestyle='--') fails on older versions of matplotlib
# plt.vlines(score, ylim[0], ylim[1], linestyle='--',
#            color='g', linewidth=3, label='Classification Score'
#            ' (pvalue %s)' % pvalue)
# plt.vlines(1.0 / n_classes, ylim[0], ylim[1], linestyle='--',
#            color='k', linewidth=3, label='Luck')
plt.plot(2 * [score], ylim, '--g', linewidth=3,
         label='Classification Score'
         ' (pvalue %s)' % pvalue)
plt.plot(2 * [1. / n_classes], ylim, '--k', linewidth=3, label='Luck')

plt.ylim(ylim)
plt.legend()
plt.xlabel('Score')
plt.show()
```

```
/anaconda3/lib/python3.6/site-packages/sklearn/model_selection/_split.py:605: Warning: The leas
  % (min_groups, self.n_splits)), Warning)
```

```python
In [21]: x_train_no_pv = x_train.copy(deep=True)
         x_train_no_pv.drop('totals.pageviews', axis = 1, inplace = True)
         x_test = test.copy()
         x_test.drop('totals.pageviews', axis = 1, inplace = True)

         folds = KFold(n_splits=5,random_state=6)
```

```python
        oof_preds = np.zeros(x_train_no_pv.shape[0])
        sub_preds = np.zeros(x_test.shape[0])

        start = time.time()
        valid_score = 0
        for n_fold, (trn_idx, val_idx) in enumerate(folds.split(x_train_no_pv, y_train)):
            trn_x, trn_y = x_train_no_pv.iloc[trn_idx], y_train[trn_idx]
            val_x, val_y = x_train_no_pv.iloc[val_idx], y_train[val_idx]

            train_data = lgb.Dataset(data=trn_x, label=trn_y)
            valid_data = lgb.Dataset(data=val_x, label=val_y)

            params = {"objective" : "regression", "metric" : "rmse", 'n_estimators':10000, 'ea
                    "num_leaves" : 30, "learning_rate" : 0.01, "bagging_fraction" : 0.9,
                    "feature_fraction" : 0.3, "bagging_seed" : 0}

            lgb_model = lgb.train(params, train_data, valid_sets=[train_data, valid_data], ve

            oof_preds[val_idx] = lgb_model.predict(val_x, num_iteration=lgb_model.best_iterat:
            oof_preds[oof_preds<0] = 0
            sub_pred = lgb_model.predict(x_test, num_iteration=lgb_model.best_iteration) / fol
            sub_pred[sub_pred<0] = 0
            sub_preds += sub_pred
            print('Fold %2d RMSE : %.6f' % (n_fold + 1, np.sqrt(mean_squared_error(val_y, oof_
            valid_score += np.sqrt(mean_squared_error(val_y, oof_preds[val_idx]))


        print('Session-level CV-score:', str(round(valid_score/folds.n_splits,4)))
        print(' ')
        train_pred = pd.DataFrame({"fullVisitorId":train_idx})
        train_pred["PredictedLogRevenue"] = np.expm1(oof_preds)
        train_pred = train_pred.groupby("fullVisitorId")["PredictedLogRevenue"].sum().reset_ir
        train_pred.columns = ["fullVisitorId", "PredictedLogRevenue"]
        train_pred["PredictedLogRevenue"] = np.log1p(train_pred["PredictedLogRevenue"])
        train_rmse = np.sqrt(mean_squared_error(train_target, train_pred['PredictedLogRevenue
        print('User-level score:', str(round(train_rmse, 4)))
        print(' ')
        end = time.time()
        print('training time:', str(round((end - start)/60)), 'mins')

        test_pred = pd.DataFrame({"fullVisitorId":test_idx})
        test_pred["PredictedLogRevenue"] = np.expm1(sub_preds)
        test_pred = test_pred.groupby("fullVisitorId")["PredictedLogRevenue"].sum().reset_inde
        test_pred.columns = ["fullVisitorId", "PredictedLogRevenue"]
        test_pred["PredictedLogRevenue"] = np.log1p(test_pred["PredictedLogRevenue"])
        test_pred.to_csv("perm_1.csv", index=False)


/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:113: UserWarning: Found `n_estimators
```

```
   warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))
/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:118: UserWarning: Found `early_stopp:
   warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))


Training until validation scores don't improve for 100 rounds.
Early stopping, best iteration is:
[643]        training's rmse: 1.89172        valid_1's rmse: 0.325545
Fold  1 RMSE : 0.325233


/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:113: UserWarning: Found `n_estimators
   warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))
/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:118: UserWarning: Found `early_stopp:
   warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))


Training until validation scores don't improve for 100 rounds.
Early stopping, best iteration is:
[515]        training's rmse: 1.88378        valid_1's rmse: 0.691198
Fold  2 RMSE : 0.691039


/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:113: UserWarning: Found `n_estimators
   warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))
/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:118: UserWarning: Found `early_stopp:
   warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))


Training until validation scores don't improve for 100 rounds.
[1000]        training's rmse: 1.42499        valid_1's rmse: 2.46021
[2000]        training's rmse: 1.39813        valid_1's rmse: 2.44509
[3000]        training's rmse: 1.38192        valid_1's rmse: 2.44247
Early stopping, best iteration is:
[3678]        training's rmse: 1.37281        valid_1's rmse: 2.44156
Fold  3 RMSE : 2.440392


/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:113: UserWarning: Found `n_estimators
   warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))
/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:118: UserWarning: Found `early_stopp:
   warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))


Training until validation scores don't improve for 100 rounds.
[1000]        training's rmse: 1.36519        valid_1's rmse: 2.5884
[2000]        training's rmse: 1.3378       valid_1's rmse: 2.57724
Early stopping, best iteration is:
[2409]        training's rmse: 1.32953        valid_1's rmse: 2.57632
Fold  4 RMSE : 2.574860
```

```
/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:113: UserWarning: Found `n_estimators
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))
/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:118: UserWarning: Found `early_stopp:
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))


Training until validation scores don't improve for 100 rounds.
[1000]          training's rmse: 1.81527          valid_1's rmse: 1.0672
[2000]          training's rmse: 1.79471          valid_1's rmse: 1.06425
Early stopping, best iteration is:
[2081]          training's rmse: 1.79351          valid_1's rmse: 1.06421
Fold  5 RMSE : 1.062379
Session-level CV-score: 1.4188


User-level score: 1.6763


training time: 8 mins
```

```python
In [22]: x_train_no_pv = x_train.copy(deep=True)
         x_train_no_pv.drop('visitStartTime', axis = 1, inplace = True)
         x_test = test.copy()
         x_test.drop('visitStartTime', axis = 1, inplace = True)

         folds = KFold(n_splits=5,random_state=6)
         oof_preds = np.zeros(x_train_no_pv.shape[0])
         sub_preds = np.zeros(x_test.shape[0])

         start = time.time()
         valid_score = 0
         for n_fold, (trn_idx, val_idx) in enumerate(folds.split(x_train_no_pv, y_train)):
             trn_x, trn_y = x_train_no_pv.iloc[trn_idx], y_train[trn_idx]
             val_x, val_y = x_train_no_pv.iloc[val_idx], y_train[val_idx]

             train_data = lgb.Dataset(data=trn_x, label=trn_y)
             valid_data = lgb.Dataset(data=val_x, label=val_y)

             params = {"objective" : "regression", "metric" : "rmse", 'n_estimators':10000, 'ea
                       "num_leaves" : 30, "learning_rate" : 0.01, "bagging_fraction" : 0.9,
                       "feature_fraction" : 0.3, "bagging_seed" : 0}

             lgb_model = lgb.train(params, train_data, valid_sets=[train_data, valid_data], ve:

             oof_preds[val_idx] = lgb_model.predict(val_x, num_iteration=lgb_model.best_iterat:
             oof_preds[oof_preds<0] = 0
             sub_pred = lgb_model.predict(x_test, num_iteration=lgb_model.best_iteration) / fo:
             sub_pred[sub_pred<0] = 0
             sub_preds += sub_pred
```

```
            print('Fold %2d RMSE : %.6f' % (n_fold + 1, np.sqrt(mean_squared_error(val_y, oof_
            valid_score += np.sqrt(mean_squared_error(val_y, oof_preds[val_idx]))


        print('Session-level CV-score:', str(round(valid_score/folds.n_splits,4)))
        print(' ')
        train_pred = pd.DataFrame({"fullVisitorId":train_idx})
        train_pred["PredictedLogRevenue"] = np.expm1(oof_preds)
        train_pred = train_pred.groupby("fullVisitorId")["PredictedLogRevenue"].sum().reset_i
        train_pred.columns = ["fullVisitorId", "PredictedLogRevenue"]
        train_pred["PredictedLogRevenue"] = np.log1p(train_pred["PredictedLogRevenue"])
        train_rmse = np.sqrt(mean_squared_error(train_target, train_pred['PredictedLogRevenue
        print('User-level score:', str(round(train_rmse, 4)))
        print(' ')
        end = time.time()
        print('training time:', str(round((end - start)/60)), 'mins')

        test_pred = pd.DataFrame({"fullVisitorId":test_idx})
        test_pred["PredictedLogRevenue"] = np.expm1(sub_preds)
        test_pred = test_pred.groupby("fullVisitorId")["PredictedLogRevenue"].sum().reset_ind
        test_pred.columns = ["fullVisitorId", "PredictedLogRevenue"]
        test_pred["PredictedLogRevenue"] = np.log1p(test_pred["PredictedLogRevenue"])
        test_pred.to_csv("perm_2.csv", index=False)

/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:113: UserWarning: Found `n_estimators
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))
/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:118: UserWarning: Found `early_stoppi
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))


Training until validation scores don't improve for 100 rounds.
Early stopping, best iteration is:
[450]        training's rmse: 1.84495        valid_1's rmse: 0.323605
Fold  1 RMSE : 0.323248


/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:113: UserWarning: Found `n_estimators
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))
/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:118: UserWarning: Found `early_stoppi
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))


Training until validation scores don't improve for 100 rounds.
[1000]        training's rmse: 1.77899        valid_1's rmse: 0.684919
Early stopping, best iteration is:
[1117]        training's rmse: 1.77291        valid_1's rmse: 0.684661
Fold  2 RMSE : 0.683880
```

```
/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:113: UserWarning: Found `n_estimators
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))
/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:118: UserWarning: Found `early_stopp
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))


Training until validation scores don't improve for 100 rounds.
[1000]        training's rmse: 1.37336        valid_1's rmse: 2.38341
[2000]        training's rmse: 1.3364         valid_1's rmse: 2.36091
[3000]        training's rmse: 1.31206        valid_1's rmse: 2.3529
[4000]        training's rmse: 1.29143        valid_1's rmse: 2.34804
[5000]        training's rmse: 1.2743         valid_1's rmse: 2.34579
Early stopping, best iteration is:
[5530]        training's rmse: 1.26605        valid_1's rmse: 2.34518
Fold  3 RMSE : 2.343569


/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:113: UserWarning: Found `n_estimators
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))
/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:118: UserWarning: Found `early_stopp
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))


Training until validation scores don't improve for 100 rounds.
[1000]        training's rmse: 1.31599        valid_1's rmse: 2.51
[2000]        training's rmse: 1.27897        valid_1's rmse: 2.48946
Early stopping, best iteration is:
[2702]        training's rmse: 1.26024        valid_1's rmse: 2.48316
Fold  4 RMSE : 2.481859


/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:113: UserWarning: Found `n_estimators
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))
/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:118: UserWarning: Found `early_stopp
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))


Training until validation scores don't improve for 100 rounds.
[1000]        training's rmse: 1.7617         valid_1's rmse: 1.05176
[2000]        training's rmse: 1.73368        valid_1's rmse: 1.04756
Early stopping, best iteration is:
[2697]        training's rmse: 1.71803        valid_1's rmse: 1.04651
Fold  5 RMSE : 1.045300
Session-level CV-score: 1.3756


User-level score: 1.6141


training time: 12 mins
```

```
In [23]: x_train_no_pv = x_train.copy(deep=True)
         x_train_no_pv.drop('channelGrouping', axis = 1, inplace = True)
         x_test = test.copy()
         x_test.drop('channelGrouping', axis = 1, inplace = True)

         folds = KFold(n_splits=5,random_state=6)
         oof_preds = np.zeros(x_train_no_pv.shape[0])
         sub_preds = np.zeros(x_test.shape[0])

         start = time.time()
         valid_score = 0
         for n_fold, (trn_idx, val_idx) in enumerate(folds.split(x_train_no_pv, y_train)):
             trn_x, trn_y = x_train_no_pv.iloc[trn_idx], y_train[trn_idx]
             val_x, val_y = x_train_no_pv.iloc[val_idx], y_train[val_idx]

             train_data = lgb.Dataset(data=trn_x, label=trn_y)
             valid_data = lgb.Dataset(data=val_x, label=val_y)

             params = {"objective" : "regression", "metric" : "rmse", 'n_estimators':10000, 'ea
                       "num_leaves" : 30, "learning_rate" : 0.01, "bagging_fraction" : 0.9,
                       "feature_fraction" : 0.3, "bagging_seed" : 0}

             lgb_model = lgb.train(params, train_data, valid_sets=[train_data, valid_data], ver

             oof_preds[val_idx] = lgb_model.predict(val_x, num_iteration=lgb_model.best_iterati
             oof_preds[oof_preds<0] = 0
             sub_pred = lgb_model.predict(x_test, num_iteration=lgb_model.best_iteration) / fol
             sub_pred[sub_pred<0] = 0
             sub_preds += sub_pred
             print('Fold %2d RMSE : %.6f' % (n_fold + 1, np.sqrt(mean_squared_error(val_y, oof_
             valid_score += np.sqrt(mean_squared_error(val_y, oof_preds[val_idx]))


         print('Session-level CV-score:', str(round(valid_score/folds.n_splits,4)))
         print(' ')
         train_pred = pd.DataFrame({"fullVisitorId":train_idx})
         train_pred["PredictedLogRevenue"] = np.expm1(oof_preds)
         train_pred = train_pred.groupby("fullVisitorId")["PredictedLogRevenue"].sum().reset_i
         train_pred.columns = ["fullVisitorId", "PredictedLogRevenue"]
         train_pred["PredictedLogRevenue"] = np.log1p(train_pred["PredictedLogRevenue"])
         train_rmse = np.sqrt(mean_squared_error(train_target, train_pred['PredictedLogRevenue
         print('User-level score:', str(round(train_rmse, 4)))
         print(' ')
         end = time.time()
         print('training time:', str(round((end - start)/60)), 'mins')

         test_pred = pd.DataFrame({"fullVisitorId":test_idx})
         test_pred["PredictedLogRevenue"] = np.expm1(sub_preds)
```

114

```
                test_pred = test_pred.groupby("fullVisitorId")["PredictedLogRevenue"].sum().reset_ind
                test_pred.columns = ["fullVisitorId", "PredictedLogRevenue"]
                test_pred["PredictedLogRevenue"] = np.log1p(test_pred["PredictedLogRevenue"])
                test_pred.to_csv("perm_3.csv", index=False)
```

/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:113: UserWarning: Found `n_estimators
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))
/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:118: UserWarning: Found `early_stopp:
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))

Training until validation scores don't improve for 100 rounds.
Early stopping, best iteration is:
[411]   training's rmse: 1.84768   valid_1's rmse: 0.324209
Fold 1 RMSE : 0.323935

/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:113: UserWarning: Found `n_estimators
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))
/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:118: UserWarning: Found `early_stopp:
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))

Training until validation scores don't improve for 100 rounds.
[1000]   training's rmse: 1.77554   valid_1's rmse: 0.684748
Early stopping, best iteration is:
[1144]   training's rmse: 1.76717   valid_1's rmse: 0.68446
Fold 2 RMSE : 0.683559

/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:113: UserWarning: Found `n_estimators
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))
/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:118: UserWarning: Found `early_stopp:
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))

Training until validation scores don't improve for 100 rounds.
[1000]   training's rmse: 1.37033   valid_1's rmse: 2.38123
[2000]   training's rmse: 1.33249   valid_1's rmse: 2.35923
[3000]   training's rmse: 1.30659   valid_1's rmse: 2.35104
[4000]   training's rmse: 1.28412   valid_1's rmse: 2.34571
[5000]   training's rmse: 1.26697   valid_1's rmse: 2.34363
Early stopping, best iteration is:
[5218]   training's rmse: 1.26362   valid_1's rmse: 2.34323
Fold 3 RMSE : 2.341761

/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:113: UserWarning: Found `n_estimators
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))
```

```
/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:118: UserWarning: Found `early_stopp:
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))


Training until validation scores don't improve for 100 rounds.
[1000]          training's rmse: 1.31231          valid_1's rmse: 2.50825
[2000]          training's rmse: 1.27421          valid_1's rmse: 2.4894
[3000]          training's rmse: 1.24857          valid_1's rmse: 2.48299
Early stopping, best iteration is:
[3068]          training's rmse: 1.24721          valid_1's rmse: 2.48268
Fold  4 RMSE : 2.481508


/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:113: UserWarning: Found `n_estimators
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))
/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:118: UserWarning: Found `early_stopp:
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))


Training until validation scores don't improve for 100 rounds.
[1000]          training's rmse: 1.7575          valid_1's rmse: 1.05179
[2000]          training's rmse: 1.72727          valid_1's rmse: 1.04759
[3000]          training's rmse: 1.7037          valid_1's rmse: 1.04515
[4000]          training's rmse: 1.68309          valid_1's rmse: 1.04419
Early stopping, best iteration is:
[3938]          training's rmse: 1.68429          valid_1's rmse: 1.04409
Fold  5 RMSE : 1.042785
Session-level CV-score: 1.3747

User-level score: 1.6133

training time: 13 mins
```

```python
In [24]: x_train_no_pv = x_train.copy(deep=True)
         x_train_no_pv.drop('geoNetwork.country', axis = 1, inplace = True)
         x_test = test.copy()
         x_test.drop('geoNetwork.country', axis = 1, inplace = True)

         folds = KFold(n_splits=5,random_state=6)
         oof_preds = np.zeros(x_train_no_pv.shape[0])
         sub_preds = np.zeros(x_test.shape[0])

         start = time.time()
         valid_score = 0
         for n_fold, (trn_idx, val_idx) in enumerate(folds.split(x_train_no_pv, y_train)):
             trn_x, trn_y = x_train_no_pv.iloc[trn_idx], y_train[trn_idx]
             val_x, val_y = x_train_no_pv.iloc[val_idx], y_train[val_idx]
```

```python
            train_data = lgb.Dataset(data=trn_x, label=trn_y)
            valid_data = lgb.Dataset(data=val_x, label=val_y)

            params = {"objective" : "regression", "metric" : "rmse", 'n_estimators':10000, 'ea
                        "num_leaves" : 30, "learning_rate" : 0.01, "bagging_fraction" : 0.9,
                        "feature_fraction" : 0.3, "bagging_seed" : 0}

            lgb_model = lgb.train(params, train_data, valid_sets=[train_data, valid_data], ve

            oof_preds[val_idx] = lgb_model.predict(val_x, num_iteration=lgb_model.best_iterati
            oof_preds[oof_preds<0] = 0
            sub_pred = lgb_model.predict(x_test, num_iteration=lgb_model.best_iteration) / fo
            sub_pred[sub_pred<0] = 0
            sub_preds += sub_pred
            print('Fold %2d RMSE : %.6f' % (n_fold + 1, np.sqrt(mean_squared_error(val_y, oof_
            valid_score += np.sqrt(mean_squared_error(val_y, oof_preds[val_idx]))


        print('Session-level CV-score:', str(round(valid_score/folds.n_splits,4)))
        print(' ')
        train_pred = pd.DataFrame({"fullVisitorId":train_idx})
        train_pred["PredictedLogRevenue"] = np.expm1(oof_preds)
        train_pred = train_pred.groupby("fullVisitorId")["PredictedLogRevenue"].sum().reset_i
        train_pred.columns = ["fullVisitorId", "PredictedLogRevenue"]
        train_pred["PredictedLogRevenue"] = np.log1p(train_pred["PredictedLogRevenue"])
        train_rmse = np.sqrt(mean_squared_error(train_target, train_pred['PredictedLogRevenue
        print('User-level score:', str(round(train_rmse, 4)))
        print(' ')
        end = time.time()
        print('training time:', str(round((end - start)/60)), 'mins')

        test_pred = pd.DataFrame({"fullVisitorId":test_idx})
        test_pred["PredictedLogRevenue"] = np.expm1(sub_preds)
        test_pred = test_pred.groupby("fullVisitorId")["PredictedLogRevenue"].sum().reset_ind
        test_pred.columns = ["fullVisitorId", "PredictedLogRevenue"]
        test_pred["PredictedLogRevenue"] = np.log1p(test_pred["PredictedLogRevenue"])
        test_pred.to_csv("perm_4.csv", index=False)
/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:113: UserWarning: Found `n_estimators
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))
/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:118: UserWarning: Found `early_stoppi
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))


Training until validation scores don't improve for 100 rounds.
Early stopping, best iteration is:
[703]        training's rmse: 1.81849        valid_1's rmse: 0.324095
Fold  1 RMSE : 0.322757
```

```
/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:113: UserWarning: Found `n_estimators
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))
/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:118: UserWarning: Found `early_stopp
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))


Training until validation scores don't improve for 100 rounds.
Early stopping, best iteration is:
[181]        training's rmse: 1.90486        valid_1's rmse: 0.691789
Fold  2 RMSE : 0.691787


/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:113: UserWarning: Found `n_estimators
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))
/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:118: UserWarning: Found `early_stopp
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))


Training until validation scores don't improve for 100 rounds.
[1000]       training's rmse: 1.37176        valid_1's rmse: 2.38296
[2000]       training's rmse: 1.33361        valid_1's rmse: 2.36042
[3000]       training's rmse: 1.30806        valid_1's rmse: 2.35213
[4000]       training's rmse: 1.28596        valid_1's rmse: 2.34688
Early stopping, best iteration is:
[4419]       training's rmse: 1.27839        valid_1's rmse: 2.34595
Fold  3 RMSE : 2.344816


/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:113: UserWarning: Found `n_estimators
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))
/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:118: UserWarning: Found `early_stopp
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))


Training until validation scores don't improve for 100 rounds.
[1000]       training's rmse: 1.31383        valid_1's rmse: 2.51115
[2000]       training's rmse: 1.27557        valid_1's rmse: 2.49121
Early stopping, best iteration is:
[2724]       training's rmse: 1.2555         valid_1's rmse: 2.48507
Fold  4 RMSE : 2.484056


/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:113: UserWarning: Found `n_estimators
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))
/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:118: UserWarning: Found `early_stopp
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))


Training until validation scores don't improve for 100 rounds.
[1000]       training's rmse: 1.76426        valid_1's rmse: 1.0541
```

```
[2000]          training's rmse: 1.73626          valid_1's rmse: 1.04982
[3000]          training's rmse: 1.71463          valid_1's rmse: 1.04671
Early stopping, best iteration is:
[3323]          training's rmse: 1.70837          valid_1's rmse: 1.04638
Fold  5 RMSE : 1.044879
Session-level CV-score: 1.3777


User-level score: 1.6171


training time: 10 mins
```

```python
In [25]: x_train_no_pv = x_train.copy(deep=True)
         x_train_no_pv.drop('date', axis = 1, inplace = True)
         x_test = test.copy()
         x_test.drop('date', axis = 1, inplace = True)

         folds = KFold(n_splits=5,random_state=6)
         oof_preds = np.zeros(x_train_no_pv.shape[0])
         sub_preds = np.zeros(x_test.shape[0])

         start = time.time()
         valid_score = 0
         for n_fold, (trn_idx, val_idx) in enumerate(folds.split(x_train_no_pv, y_train)):
             trn_x, trn_y = x_train_no_pv.iloc[trn_idx], y_train[trn_idx]
             val_x, val_y = x_train_no_pv.iloc[val_idx], y_train[val_idx]

             train_data = lgb.Dataset(data=trn_x, label=trn_y)
             valid_data = lgb.Dataset(data=val_x, label=val_y)

             params = {"objective" : "regression", "metric" : "rmse", 'n_estimators':10000, 'ea
                     "num_leaves" : 30, "learning_rate" : 0.01, "bagging_fraction" : 0.9,
                     "feature_fraction" : 0.3, "bagging_seed" : 0}

             lgb_model = lgb.train(params, train_data, valid_sets=[train_data, valid_data], ver

             oof_preds[val_idx] = lgb_model.predict(val_x, num_iteration=lgb_model.best_iterat
             oof_preds[oof_preds<0] = 0
             sub_pred = lgb_model.predict(x_test, num_iteration=lgb_model.best_iteration) / fol
             sub_pred[sub_pred<0] = 0
             sub_preds += sub_pred
             print('Fold %2d RMSE : %.6f' % (n_fold + 1, np.sqrt(mean_squared_error(val_y, oof_
             valid_score += np.sqrt(mean_squared_error(val_y, oof_preds[val_idx]))

         print('Session-level CV-score:', str(round(valid_score/folds.n_splits,4)))
         print(' ')
         train_pred = pd.DataFrame({"fullVisitorId":train_idx})
```

```
train_pred["PredictedLogRevenue"] = np.expm1(oof_preds)
train_pred = train_pred.groupby("fullVisitorId")["PredictedLogRevenue"].sum().reset_i
train_pred.columns = ["fullVisitorId", "PredictedLogRevenue"]
train_pred["PredictedLogRevenue"] = np.log1p(train_pred["PredictedLogRevenue"])
train_rmse = np.sqrt(mean_squared_error(train_target, train_pred['PredictedLogRevenue
print('User-level score:', str(round(train_rmse, 4)))
print(' ')
end = time.time()
print('training time:', str(round((end - start)/60)), 'mins')

test_pred = pd.DataFrame({"fullVisitorId":test_idx})
test_pred["PredictedLogRevenue"] = np.expm1(sub_preds)
test_pred = test_pred.groupby("fullVisitorId")["PredictedLogRevenue"].sum().reset_ind
test_pred.columns = ["fullVisitorId", "PredictedLogRevenue"]
test_pred["PredictedLogRevenue"] = np.log1p(test_pred["PredictedLogRevenue"])
test_pred.to_csv("perm_5.csv", index=False)
```

/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:113: UserWarning: Found `n_estimators
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))
/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:118: UserWarning: Found `early_stoppi
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))

Training until validation scores don't improve for 100 rounds.
Early stopping, best iteration is:
[405]   training's rmse: 1.85209   valid_1's rmse: 0.324492
Fold 1 RMSE : 0.324260

/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:113: UserWarning: Found `n_estimators
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))
/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:118: UserWarning: Found `early_stoppi
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))

Training until validation scores don't improve for 100 rounds.
[1000]   training's rmse: 1.77805   valid_1's rmse: 0.684338
Early stopping, best iteration is:
[1168]   training's rmse: 1.76861   valid_1's rmse: 0.683877
Fold 2 RMSE : 0.682915

/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:113: UserWarning: Found `n_estimators
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))
/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:118: UserWarning: Found `early_stoppi
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))

Training until validation scores don't improve for 100 rounds.
[1000]   training's rmse: 1.37252   valid_1's rmse: 2.38275
```

```
[2000]          training's rmse: 1.33555         valid_1's rmse: 2.36072
[3000]          training's rmse: 1.31076         valid_1's rmse: 2.35206
[4000]          training's rmse: 1.28894         valid_1's rmse: 2.34573
Early stopping, best iteration is:
[4554]          training's rmse: 1.28019         valid_1's rmse: 2.34503
Fold  3 RMSE : 2.343587


/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:113: UserWarning: Found `n_estimators
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))
/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:118: UserWarning: Found `early_stopp
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))


Training until validation scores don't improve for 100 rounds.
[1000]          training's rmse: 1.31496         valid_1's rmse: 2.50795
[2000]          training's rmse: 1.27765         valid_1's rmse: 2.48746
Early stopping, best iteration is:
[2724]          training's rmse: 1.25822         valid_1's rmse: 2.48178
Fold  4 RMSE : 2.480462


/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:113: UserWarning: Found `n_estimators
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))
/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:118: UserWarning: Found `early_stopp
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))


Training until validation scores don't improve for 100 rounds.
[1000]          training's rmse: 1.76338         valid_1's rmse: 1.05377
[2000]          training's rmse: 1.73602         valid_1's rmse: 1.04969
[3000]          training's rmse: 1.71523         valid_1's rmse: 1.04789
[4000]          training's rmse: 1.69721         valid_1's rmse: 1.04644
[5000]          training's rmse: 1.68164         valid_1's rmse: 1.04538
Early stopping, best iteration is:
[5406]          training's rmse: 1.67583         valid_1's rmse: 1.04495
Fold  5 RMSE : 1.042971
Session-level CV-score: 1.3748

User-level score: 1.6143

training time: 13 mins


In [26]: x_train_no_pv = x_train.copy(deep=True)

         x_train_no_pv.drop('device.isMobile', axis = 1, inplace = True)
         x_train_no_pv.drop('channelGrouping', axis = 1, inplace = True)
         x_train_no_pv.drop('geoNetwork.networkDomain', axis = 1, inplace = True)
```

```python
        x_train_no_pv.drop('geoNetwork.region', axis = 1, inplace = True)


x_test = test.copy()
x_test.drop('device.isMobile', axis = 1, inplace = True)
x_test.drop('channelGrouping', axis = 1, inplace = True)
x_test.drop('geoNetwork.networkDomain', axis = 1, inplace = True)
x_test.drop('geoNetwork.region', axis = 1, inplace = True)

folds = KFold(n_splits=5,random_state=6)
oof_preds = np.zeros(x_train_no_pv.shape[0])
sub_preds = np.zeros(x_test.shape[0])

start = time.time()
valid_score = 0
for n_fold, (trn_idx, val_idx) in enumerate(folds.split(x_train_no_pv, y_train)):
    trn_x, trn_y = x_train_no_pv.iloc[trn_idx], y_train[trn_idx]
    val_x, val_y = x_train_no_pv.iloc[val_idx], y_train[val_idx]

    train_data = lgb.Dataset(data=trn_x, label=trn_y)
    valid_data = lgb.Dataset(data=val_x, label=val_y)

    params = {"objective" : "regression", "metric" : "rmse", 'n_estimators':10000, 'ea
              "num_leaves" : 30, "learning_rate" : 0.01, "bagging_fraction" : 0.9,
              "feature_fraction" : 0.3, "bagging_seed" : 0}

    lgb_model = lgb.train(params, train_data, valid_sets=[train_data, valid_data], ver

    oof_preds[val_idx] = lgb_model.predict(val_x, num_iteration=lgb_model.best_iterat
    oof_preds[oof_preds<0] = 0
    sub_pred = lgb_model.predict(x_test, num_iteration=lgb_model.best_iteration) / fo
    sub_pred[sub_pred<0] = 0
    sub_preds += sub_pred
    print('Fold %2d RMSE : %.6f' % (n_fold + 1, np.sqrt(mean_squared_error(val_y, oof_
    valid_score += np.sqrt(mean_squared_error(val_y, oof_preds[val_idx]))


print('Session-level CV-score:', str(round(valid_score/folds.n_splits,4)))
print(' ')
train_pred = pd.DataFrame({"fullVisitorId":train_idx})
train_pred["PredictedLogRevenue"] = np.expm1(oof_preds)
train_pred = train_pred.groupby("fullVisitorId")["PredictedLogRevenue"].sum().reset_i
train_pred.columns = ["fullVisitorId", "PredictedLogRevenue"]
train_pred["PredictedLogRevenue"] = np.log1p(train_pred["PredictedLogRevenue"])
train_rmse = np.sqrt(mean_squared_error(train_target, train_pred['PredictedLogRevenue
print('User-level score:', str(round(train_rmse, 4)))
print(' ')
end = time.time()
```

```python
        print('training time:', str(round((end - start)/60)), 'mins')

        test_pred = pd.DataFrame({"fullVisitorId":test_idx})
        test_pred["PredictedLogRevenue"] = np.expm1(sub_preds)
        test_pred = test_pred.groupby("fullVisitorId")["PredictedLogRevenue"].sum().reset_ind
        test_pred.columns = ["fullVisitorId", "PredictedLogRevenue"]
        test_pred["PredictedLogRevenue"] = np.log1p(test_pred["PredictedLogRevenue"])
        test_pred.to_csv("perm_6.csv", index=False)
```

```
/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:113: UserWarning: Found `n_estimators
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))
/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:118: UserWarning: Found `early_stoppi
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))


Training until validation scores don't improve for 100 rounds.
Early stopping, best iteration is:
[470]        training's rmse: 1.83814        valid_1's rmse: 0.323722
Fold  1 RMSE : 0.323353


/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:113: UserWarning: Found `n_estimators
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))
/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:118: UserWarning: Found `early_stoppi
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))


Training until validation scores don't improve for 100 rounds.
Early stopping, best iteration is:
[366]        training's rmse: 1.82904        valid_1's rmse: 0.684849
Fold  2 RMSE : 0.684744


/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:113: UserWarning: Found `n_estimators
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))
/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:118: UserWarning: Found `early_stoppi
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))


Training until validation scores don't improve for 100 rounds.
[1000]        training's rmse: 1.36969        valid_1's rmse: 2.3781
[2000]        training's rmse: 1.33606        valid_1's rmse: 2.3598
[3000]        training's rmse: 1.31257        valid_1's rmse: 2.35395
[4000]        training's rmse: 1.29235        valid_1's rmse: 2.34884
[5000]        training's rmse: 1.27648        valid_1's rmse: 2.34685
Early stopping, best iteration is:
[5347]        training's rmse: 1.27136        valid_1's rmse: 2.34586
Fold  3 RMSE : 2.344373
```

```
/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:113: UserWarning: Found `n_estimators
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))
/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:118: UserWarning: Found `early_stopp:
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))


Training until validation scores don't improve for 100 rounds.
[1000]        training's rmse: 1.31116        valid_1's rmse: 2.50784
[2000]        training's rmse: 1.27759        valid_1's rmse: 2.49153
[3000]        training's rmse: 1.25433        valid_1's rmse: 2.48597
[4000]        training's rmse: 1.23519        valid_1's rmse: 2.48321
Early stopping, best iteration is:
[4026]        training's rmse: 1.23486        valid_1's rmse: 2.4832
Fold  4 RMSE : 2.481716


/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:113: UserWarning: Found `n_estimators
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))
/anaconda3/lib/python3.6/site-packages/lightgbm/engine.py:118: UserWarning: Found `early_stopp:
  warnings.warn("Found `{}` in params. Will use it instead of argument".format(alias))


Training until validation scores don't improve for 100 rounds.
[1000]        training's rmse: 1.73958        valid_1's rmse: 1.0336
[2000]        training's rmse: 1.70149        valid_1's rmse: 1.0233
[3000]        training's rmse: 1.67823        valid_1's rmse: 1.02023
[4000]        training's rmse: 1.65982        valid_1's rmse: 1.01797
[5000]        training's rmse: 1.64269        valid_1's rmse: 1.01613
Early stopping, best iteration is:
[5591]        training's rmse: 1.63336        valid_1's rmse: 1.0154
Fold  5 RMSE : 1.013236
Session-level CV-score: 1.3695

User-level score: 1.6108

training time: 16 mins
```

```python
In [34]: import random
         import squarify
         number_of_colors = 20
         color = ["#"+''.join([random.choice('A089B14C32D65E7F') for j in range(6)]) for i in
         country_tree = train["geoNetwork.country"].value_counts() #counting
         country_tree = round((train["geoNetwork.country"].value_counts()[:25] \
                               / len(train['geoNetwork.country']) * 100),2)
         plt.figure(figsize=(14,5))
         g = squarify.plot(sizes=country_tree.values, label=country_tree.index,
                           value=country_tree.values,
                           alpha=.4, color=color)
```

```
g.set_title("Top 25 Countries - Percent size of total",fontsize=20)
g.set_axis_off()
plt.show()
```

Top 25 Countries - Percent size of total

	Mexico 1.46	Italy 1.25	Singapore 0.79	Romania 0.71 / Malaysia 0.71
	France 1.75	Netherlands 1.27	Philippines 1.02	Ireland 0.72
United States 40.36		Spain 1.29	Poland 1.07	Indonesia 1.03
	Japan 2.18	Taiwan 1.44	Australia 1.41	Russia 1.29
	Thailand 2.23	Germany 2.21		Brazil 2.19
	Canada 2.86	Vietnam 2.72		Turkey 2.27
	India 5.66		United Kingdom 4.14	