{

"cells": [

{

"cell\_type": "code",

"execution\_count": 1,

"id": "7ca2bebc",

"metadata": {},

"outputs": [],

"source": [

"#Import pandas. \n",

"import pandas as pd"

]

},

{

"cell\_type": "code",

"execution\_count": 2,

"id": "7bc4d4a4",

"metadata": {},

"outputs": [],

"source": [

"#Import numpy. \n",

"import numpy as np"

]

},

{

"cell\_type": "code",

"execution\_count": 3,

"id": "c7dd1e8c",

"metadata": {},

"outputs": [

{

"name": "stdout",

"output\_type": "stream",

"text": [

" sub\_icb\_location\_code sub\_icb\_location\_ons\_code \\\n",

"0 00L E38000130 \n",

"1 00L E38000130 \n",

"2 00L E38000130 \n",

"3 00L E38000130 \n",

"4 00L E38000130 \n",

"... ... ... \n",

"137788 X2C4Y E38000254 \n",

"137789 X2C4Y E38000254 \n",

"137790 X2C4Y E38000254 \n",

"137791 X2C4Y E38000254 \n",

"137792 X2C4Y E38000254 \n",

"\n",

" sub\_icb\_location\_name icb\_ons\_code \\\n",

"0 NHS North East and North Cumbria ICB - 00L E54000050 \n",

"1 NHS North East and North Cumbria ICB - 00L E54000050 \n",

"2 NHS North East and North Cumbria ICB - 00L E54000050 \n",

"3 NHS North East and North Cumbria ICB - 00L E54000050 \n",

"4 NHS North East and North Cumbria ICB - 00L E54000050 \n",

"... ... ... \n",

"137788 NHS West Yorkshire ICB - X2C4Y E54000054 \n",

"137789 NHS West Yorkshire ICB - X2C4Y E54000054 \n",

"137790 NHS West Yorkshire ICB - X2C4Y E54000054 \n",

"137791 NHS West Yorkshire ICB - X2C4Y E54000054 \n",

"137792 NHS West Yorkshire ICB - X2C4Y E54000054 \n",

"\n",

" region\_ons\_code appointment\_date actual\_duration \\\n",

"0 E40000012 01-Dec-21 31-60 Minutes \n",

"1 E40000012 01-Dec-21 21-30 Minutes \n",

"2 E40000012 01-Dec-21 6-10 Minutes \n",

"3 E40000012 01-Dec-21 Unknown / Data Quality \n",

"4 E40000012 01-Dec-21 16-20 Minutes \n",

"... ... ... ... \n",

"137788 E40000012 30-Jun-22 31-60 Minutes \n",

"137789 E40000012 30-Jun-22 21-30 Minutes \n",

"137790 E40000012 30-Jun-22 16-20 Minutes \n",

"137791 E40000012 30-Jun-22 11-15 Minutes \n",

"137792 E40000012 30-Jun-22 1-5 Minutes \n",

"\n",

" count\_of\_appointments \n",

"0 364 \n",

"1 619 \n",

"2 1698 \n",

"3 1277 \n",

"4 730 \n",

"... ... \n",

"137788 430 \n",

"137789 751 \n",

"137790 921 \n",

"137791 1439 \n",

"137792 1739 \n",

"\n",

"[137793 rows x 8 columns]\n"

]

}

],

"source": [

"# Data frame for the actual duration data.\n",

"ad = pd.read\_csv('actual\_duration.csv')\n",

" \n",

"# Print the DataFrame for actual duration.\n",

"print (ad) "

]

},

{

"cell\_type": "code",

"execution\_count": 4,

"id": "cd1ef95f",

"metadata": {},

"outputs": [

{

"name": "stdout",

"output\_type": "stream",

"text": [

" icb\_ons\_code appointment\_month appointment\_status hcp\_type \\\n",

"0 E54000034 2020-01 Attended GP \n",

"1 E54000034 2020-01 Attended GP \n",

"2 E54000034 2020-01 Attended GP \n",

"3 E54000034 2020-01 Attended GP \n",

"4 E54000034 2020-01 Attended GP \n",

"... ... ... ... ... \n",

"596816 E54000050 2022-06 Unknown Unknown \n",

"596817 E54000050 2022-06 Unknown Unknown \n",

"596818 E54000050 2022-06 Unknown Unknown \n",

"596819 E54000050 2022-06 Unknown Unknown \n",

"596820 E54000050 2022-06 Unknown Unknown \n",

"\n",

" appointment\_mode time\_between\_book\_and\_appointment \\\n",

"0 Face-to-Face 1 Day \n",

"1 Face-to-Face 15 to 21 Days \n",

"2 Face-to-Face 2 to 7 Days \n",

"3 Face-to-Face 22 to 28 Days \n",

"4 Face-to-Face 8 to 14 Days \n",

"... ... ... \n",

"596816 Unknown 2 to 7 Days \n",

"596817 Unknown 22 to 28 Days \n",

"596818 Unknown 8 to 14 Days \n",

"596819 Unknown More than 28 Days \n",

"596820 Unknown Same Day \n",

"\n",

" count\_of\_appointments \n",

"0 8107 \n",

"1 6791 \n",

"2 20686 \n",

"3 4268 \n",

"4 11971 \n",

"... ... \n",

"596816 21 \n",

"596817 8 \n",

"596818 28 \n",

"596819 17 \n",

"596820 10 \n",

"\n",

"[596821 rows x 7 columns]\n"

]

}

],

"source": [

"# Dataframe for the appointments regional data \n",

"ar = pd.read\_csv('appointments\_regional.csv')\n",

" \n",

"# Print the DataFrame for appointments regional.\n",

"print (ar) "

]

},

{

"cell\_type": "code",

"execution\_count": 5,

"id": "d9363a21",

"metadata": {},

"outputs": [

{

"name": "stdout",

"output\_type": "stream",

"text": [

" appointment\_date icb\_ons\_code \\\n",

"0 2021-08-02 E54000050 \n",

"1 2021-08-02 E54000050 \n",

"2 2021-08-02 E54000050 \n",

"3 2021-08-02 E54000050 \n",

"4 2021-08-02 E54000050 \n",

"... ... ... \n",

"817389 2022-06-30 E54000054 \n",

"817390 2022-06-30 E54000054 \n",

"817391 2022-06-30 E54000054 \n",

"817392 2022-06-30 E54000054 \n",

"817393 2022-06-30 E54000054 \n",

"\n",

" sub\_icb\_location\_name service\_setting \\\n",

"0 NHS North East and North Cumbria ICB - 00L Primary Care Network \n",

"1 NHS North East and North Cumbria ICB - 00L Other \n",

"2 NHS North East and North Cumbria ICB - 00L General Practice \n",

"3 NHS North East and North Cumbria ICB - 00L General Practice \n",

"4 NHS North East and North Cumbria ICB - 00L General Practice \n",

"... ... ... \n",

"817389 NHS West Yorkshire ICB - X2C4Y Extended Access Provision \n",

"817390 NHS West Yorkshire ICB - X2C4Y Extended Access Provision \n",

"817391 NHS West Yorkshire ICB - X2C4Y Extended Access Provision \n",

"817392 NHS West Yorkshire ICB - X2C4Y Extended Access Provision \n",

"817393 NHS West Yorkshire ICB - X2C4Y Extended Access Provision \n",

"\n",

" context\_type national\_category \\\n",

"0 Care Related Encounter Patient contact during Care Home Round \n",

"1 Care Related Encounter Planned Clinics \n",

"2 Care Related Encounter Home Visit \n",

"3 Care Related Encounter General Consultation Acute \n",

"4 Care Related Encounter Structured Medication Review \n",

"... ... ... \n",

"817389 Care Related Encounter Unplanned Clinical Activity \n",

"817390 Care Related Encounter Planned Clinics \n",

"817391 Care Related Encounter Planned Clinical Procedure \n",

"817392 Care Related Encounter General Consultation Routine \n",

"817393 Care Related Encounter General Consultation Acute \n",

"\n",

" count\_of\_appointments appointment\_month \n",

"0 3 2021-08 \n",

"1 7 2021-08 \n",

"2 79 2021-08 \n",

"3 725 2021-08 \n",

"4 2 2021-08 \n",

"... ... ... \n",

"817389 12 2022-06 \n",

"817390 4 2022-06 \n",

"817391 92 2022-06 \n",

"817392 4 2022-06 \n",

"817393 19 2022-06 \n",

"\n",

"[817394 rows x 8 columns]\n"

]

}

],

"source": [

"# Dataframe for the national categories data. \n",

"nc = pd.read\_excel('national\_categories.xlsx')\n",

" \n",

"# Print the DataFrame for national categories.\n",

"print (nc) "

]

},

{

"cell\_type": "code",

"execution\_count": 6,

"id": "28fa2960",

"metadata": {},

"outputs": [

{

"name": "stdout",

"output\_type": "stream",

"text": [

"Index(['sub\_icb\_location\_code', 'sub\_icb\_location\_ons\_code',\n",

" 'sub\_icb\_location\_name', 'icb\_ons\_code', 'region\_ons\_code',\n",

" 'appointment\_date', 'actual\_duration', 'count\_of\_appointments'],\n",

" dtype='object')\n"

]

}

],

"source": [

"#Print column names for ad\n",

"print (ad.columns)"

]

},

{

"cell\_type": "code",

"execution\_count": 7,

"id": "414bbf7e",

"metadata": {},

"outputs": [

{

"name": "stdout",

"output\_type": "stream",

"text": [

"Index(['icb\_ons\_code', 'appointment\_month', 'appointment\_status', 'hcp\_type',\n",

" 'appointment\_mode', 'time\_between\_book\_and\_appointment',\n",

" 'count\_of\_appointments'],\n",

" dtype='object')\n"

]

}

],

"source": [

"#Print column names for ar\n",

"print (ar.columns)"

]

},

{

"cell\_type": "code",

"execution\_count": 8,

"id": "e97b43a4",

"metadata": {},

"outputs": [

{

"name": "stdout",

"output\_type": "stream",

"text": [

"Index(['appointment\_date', 'icb\_ons\_code', 'sub\_icb\_location\_name',\n",

" 'service\_setting', 'context\_type', 'national\_category',\n",

" 'count\_of\_appointments', 'appointment\_month'],\n",

" dtype='object')\n"

]

}

],

"source": [

"#Print column names for nc\n",

"print (nc.columns)"

]

},

{

"cell\_type": "code",

"execution\_count": 9,

"id": "22e5a694",

"metadata": {},

"outputs": [

{

"data": {

"text/html": [

"<div>\n",

"<style scoped>\n",

" .dataframe tbody tr th:only-of-type {\n",

" vertical-align: middle;\n",

" }\n",

"\n",

" .dataframe tbody tr th {\n",

" vertical-align: top;\n",

" }\n",

"\n",

" .dataframe thead th {\n",

" text-align: right;\n",

" }\n",

"</style>\n",

"<table border=\"1\" class=\"dataframe\">\n",

" <thead>\n",

" <tr style=\"text-align: right;\">\n",

" <th></th>\n",

" <th>count\_of\_appointments</th>\n",

" </tr>\n",

" </thead>\n",

" <tbody>\n",

" <tr>\n",

" <th>count</th>\n",

" <td>137793.000000</td>\n",

" </tr>\n",

" <tr>\n",

" <th>mean</th>\n",

" <td>1219.080011</td>\n",

" </tr>\n",

" <tr>\n",

" <th>std</th>\n",

" <td>1546.902956</td>\n",

" </tr>\n",

" <tr>\n",

" <th>min</th>\n",

" <td>1.000000</td>\n",

" </tr>\n",

" <tr>\n",

" <th>25%</th>\n",

" <td>194.000000</td>\n",

" </tr>\n",

" <tr>\n",

" <th>50%</th>\n",

" <td>696.000000</td>\n",

" </tr>\n",

" <tr>\n",

" <th>75%</th>\n",

" <td>1621.000000</td>\n",

" </tr>\n",

" <tr>\n",

" <th>max</th>\n",

" <td>15400.000000</td>\n",

" </tr>\n",

" </tbody>\n",

"</table>\n",

"</div>"

],

"text/plain": [

" count\_of\_appointments\n",

"count 137793.000000\n",

"mean 1219.080011\n",

"std 1546.902956\n",

"min 1.000000\n",

"25% 194.000000\n",

"50% 696.000000\n",

"75% 1621.000000\n",

"max 15400.000000"

]

},

"execution\_count": 9,

"metadata": {},

"output\_type": "execute\_result"

}

],

"source": [

"#Descriptive statistics for ad\n",

"ad.describe()"

]

},

{

"cell\_type": "code",

"execution\_count": 10,

"id": "514d36ac",

"metadata": {},

"outputs": [

{

"name": "stdout",

"output\_type": "stream",

"text": [

"<class 'pandas.core.frame.DataFrame'>\n",

"RangeIndex: 137793 entries, 0 to 137792\n",

"Data columns (total 8 columns):\n",

" # Column Non-Null Count Dtype \n",

"--- ------ -------------- ----- \n",

" 0 sub\_icb\_location\_code 137793 non-null object\n",

" 1 sub\_icb\_location\_ons\_code 137793 non-null object\n",

" 2 sub\_icb\_location\_name 137793 non-null object\n",

" 3 icb\_ons\_code 137793 non-null object\n",

" 4 region\_ons\_code 137793 non-null object\n",

" 5 appointment\_date 137793 non-null object\n",

" 6 actual\_duration 137793 non-null object\n",

" 7 count\_of\_appointments 137793 non-null int64 \n",

"dtypes: int64(1), object(7)\n",

"memory usage: 8.4+ MB\n"

]

}

],

"source": [

"#Meta data for ad\n",

"ad.info()"

]

},

{

"cell\_type": "code",

"execution\_count": 11,

"id": "2229771d",

"metadata": {},

"outputs": [

{

"data": {

"text/html": [

"<div>\n",

"<style scoped>\n",

" .dataframe tbody tr th:only-of-type {\n",

" vertical-align: middle;\n",

" }\n",

"\n",

" .dataframe tbody tr th {\n",

" vertical-align: top;\n",

" }\n",

"\n",

" .dataframe thead th {\n",

" text-align: right;\n",

" }\n",

"</style>\n",

"<table border=\"1\" class=\"dataframe\">\n",

" <thead>\n",

" <tr style=\"text-align: right;\">\n",

" <th></th>\n",

" <th>count\_of\_appointments</th>\n",

" </tr>\n",

" </thead>\n",

" <tbody>\n",

" <tr>\n",

" <th>count</th>\n",

" <td>596821.000000</td>\n",

" </tr>\n",

" <tr>\n",

" <th>mean</th>\n",

" <td>1244.601857</td>\n",

" </tr>\n",

" <tr>\n",

" <th>std</th>\n",

" <td>5856.887042</td>\n",

" </tr>\n",

" <tr>\n",

" <th>min</th>\n",

" <td>1.000000</td>\n",

" </tr>\n",

" <tr>\n",

" <th>25%</th>\n",

" <td>7.000000</td>\n",

" </tr>\n",

" <tr>\n",

" <th>50%</th>\n",

" <td>47.000000</td>\n",

" </tr>\n",

" <tr>\n",

" <th>75%</th>\n",

" <td>308.000000</td>\n",

" </tr>\n",

" <tr>\n",

" <th>max</th>\n",

" <td>211265.000000</td>\n",

" </tr>\n",

" </tbody>\n",

"</table>\n",

"</div>"

],

"text/plain": [

" count\_of\_appointments\n",

"count 596821.000000\n",

"mean 1244.601857\n",

"std 5856.887042\n",

"min 1.000000\n",

"25% 7.000000\n",

"50% 47.000000\n",

"75% 308.000000\n",

"max 211265.000000"

]

},

"execution\_count": 11,

"metadata": {},

"output\_type": "execute\_result"

}

],

"source": [

"#Descriptive statistics for ar\n",

"ar.describe()"

]

},

{

"cell\_type": "code",

"execution\_count": 12,

"id": "bf7d7e04",

"metadata": {},

"outputs": [

{

"name": "stdout",

"output\_type": "stream",

"text": [

"<class 'pandas.core.frame.DataFrame'>\n",

"RangeIndex: 596821 entries, 0 to 596820\n",

"Data columns (total 7 columns):\n",

" # Column Non-Null Count Dtype \n",

"--- ------ -------------- ----- \n",

" 0 icb\_ons\_code 596821 non-null object\n",

" 1 appointment\_month 596821 non-null object\n",

" 2 appointment\_status 596821 non-null object\n",

" 3 hcp\_type 596821 non-null object\n",

" 4 appointment\_mode 596821 non-null object\n",

" 5 time\_between\_book\_and\_appointment 596821 non-null object\n",

" 6 count\_of\_appointments 596821 non-null int64 \n",

"dtypes: int64(1), object(6)\n",

"memory usage: 31.9+ MB\n"

]

}

],

"source": [

"#Meta data for ar\n",

"ar.info()"

]

},

{

"cell\_type": "code",

"execution\_count": 13,

"id": "ef1052c4",

"metadata": {},

"outputs": [

{

"data": {

"text/html": [

"<div>\n",

"<style scoped>\n",

" .dataframe tbody tr th:only-of-type {\n",

" vertical-align: middle;\n",

" }\n",

"\n",

" .dataframe tbody tr th {\n",

" vertical-align: top;\n",

" }\n",

"\n",

" .dataframe thead th {\n",

" text-align: right;\n",

" }\n",

"</style>\n",

"<table border=\"1\" class=\"dataframe\">\n",

" <thead>\n",

" <tr style=\"text-align: right;\">\n",

" <th></th>\n",

" <th>count\_of\_appointments</th>\n",

" </tr>\n",

" </thead>\n",

" <tbody>\n",

" <tr>\n",

" <th>count</th>\n",

" <td>817394.000000</td>\n",

" </tr>\n",

" <tr>\n",

" <th>mean</th>\n",

" <td>362.183684</td>\n",

" </tr>\n",

" <tr>\n",

" <th>std</th>\n",

" <td>1084.576600</td>\n",

" </tr>\n",

" <tr>\n",

" <th>min</th>\n",

" <td>1.000000</td>\n",

" </tr>\n",

" <tr>\n",

" <th>25%</th>\n",

" <td>7.000000</td>\n",

" </tr>\n",

" <tr>\n",

" <th>50%</th>\n",

" <td>25.000000</td>\n",

" </tr>\n",

" <tr>\n",

" <th>75%</th>\n",

" <td>128.000000</td>\n",

" </tr>\n",

" <tr>\n",

" <th>max</th>\n",

" <td>16590.000000</td>\n",

" </tr>\n",

" </tbody>\n",

"</table>\n",

"</div>"

],

"text/plain": [

" count\_of\_appointments\n",

"count 817394.000000\n",

"mean 362.183684\n",

"std 1084.576600\n",

"min 1.000000\n",

"25% 7.000000\n",

"50% 25.000000\n",

"75% 128.000000\n",

"max 16590.000000"

]

},

"execution\_count": 13,

"metadata": {},

"output\_type": "execute\_result"

}

],

"source": [

"#Descriptive statistics for nc\n",

"nc.describe()"

]

},

{

"cell\_type": "code",

"execution\_count": 14,

"id": "7575a169",

"metadata": {},

"outputs": [

{

"name": "stdout",

"output\_type": "stream",

"text": [

"<class 'pandas.core.frame.DataFrame'>\n",

"RangeIndex: 817394 entries, 0 to 817393\n",

"Data columns (total 8 columns):\n",

" # Column Non-Null Count Dtype \n",

"--- ------ -------------- ----- \n",

" 0 appointment\_date 817394 non-null datetime64[ns]\n",

" 1 icb\_ons\_code 817394 non-null object \n",

" 2 sub\_icb\_location\_name 817394 non-null object \n",

" 3 service\_setting 817394 non-null object \n",

" 4 context\_type 817394 non-null object \n",

" 5 national\_category 817394 non-null object \n",

" 6 count\_of\_appointments 817394 non-null int64 \n",

" 7 appointment\_month 817394 non-null object \n",

"dtypes: datetime64[ns](1), int64(1), object(6)\n",

"memory usage: 49.9+ MB\n"

]

}

],

"source": [

"#Meta data for nc\n",

"nc.info()"

]

},

{

"cell\_type": "code",

"execution\_count": 15,

"id": "4cc1d61b",

"metadata": {},

"outputs": [

{

"data": {

"text/plain": [

"NHS North West London ICB - W2U3Z 13007\n",

"NHS Kent and Medway ICB - 91Q 12637\n",

"NHS Devon ICB - 15N 12526\n",

"NHS Hampshire and Isle Of Wight ICB - D9Y0V 12171\n",

"NHS North East London ICB - A3A8R 11837\n",

" ... \n",

"NHS North East and North Cumbria ICB - 00N 4210\n",

"NHS Lancashire and South Cumbria ICB - 02G 4169\n",

"NHS Cheshire and Merseyside ICB - 01V 3496\n",

"NHS Cheshire and Merseyside ICB - 01T 3242\n",

"NHS Greater Manchester ICB - 00V 2170\n",

"Name: sub\_icb\_location\_name, Length: 106, dtype: int64"

]

},

"execution\_count": 15,

"metadata": {},

"output\_type": "execute\_result"

}

],

"source": [

"#List of the different locations in the data set\n",

"nc[\"sub\_icb\_location\_name\"].value\_counts( )"

]

},

{

"cell\_type": "code",

"execution\_count": 16,

"id": "39a8bf47",

"metadata": {},

"outputs": [

{

"data": {

"text/plain": [

"General Practice 359274\n",

"Primary Care Network 183790\n",

"Other 138789\n",

"Extended Access Provision 108122\n",

"Unmapped 27419\n",

"Name: service\_setting, dtype: int64"

]

},

"execution\_count": 16,

"metadata": {},

"output\_type": "execute\_result"

}

],

"source": [

"#List of different service settings\n",

"nc[\"service\_setting\"].value\_counts()"

]

},

{

"cell\_type": "code",

"execution\_count": 17,

"id": "330d17bc",

"metadata": {},

"outputs": [

{

"data": {

"text/plain": [

"Care Related Encounter 700481\n",

"Inconsistent Mapping 89494\n",

"Unmapped 27419\n",

"Name: context\_type, dtype: int64"

]

},

"execution\_count": 17,

"metadata": {},

"output\_type": "execute\_result"

}

],

"source": [

"#List of different context types\n",

"nc[\"context\_type\"].value\_counts()"

]

},

{

"cell\_type": "code",

"execution\_count": 18,

"id": "fe2f6501",

"metadata": {},

"outputs": [

{

"data": {

"text/plain": [

"Inconsistent Mapping 89494\n",

"General Consultation Routine 89329\n",

"General Consultation Acute 84874\n",

"Planned Clinics 76429\n",

"Clinical Triage 74539\n",

"Planned Clinical Procedure 59631\n",

"Structured Medication Review 44467\n",

"Service provided by organisation external to the practice 43095\n",

"Home Visit 41850\n",

"Unplanned Clinical Activity 40415\n",

"Patient contact during Care Home Round 28795\n",

"Unmapped 27419\n",

"Care Home Visit 26644\n",

"Social Prescribing Service 26492\n",

"Care Home Needs Assessment & Personalised Care and Support Planning 23505\n",

"Non-contractual chargeable work 20896\n",

"Walk-in 14179\n",

"Group Consultation and Group Education 5341\n",

"Name: national\_category, dtype: int64"

]

},

"execution\_count": 18,

"metadata": {},

"output\_type": "execute\_result"

}

],

"source": [

"#List of national categories\n",

"nc[\"national\_category\"].value\_counts()"

]

},

{

"cell\_type": "code",

"execution\_count": 19,

"id": "eda5172e",

"metadata": {},

"outputs": [

{

"data": {

"text/plain": [

"Attended 232137\n",

"Unknown 201324\n",

"DNA 163360\n",

"Name: appointment\_status, dtype: int64"

]

},

"execution\_count": 19,

"metadata": {},

"output\_type": "execute\_result"

}

],

"source": [

"#List of appointment statues \n",

"ar[\"appointment\_status\"].value\_counts()"

]

},

{

"cell\_type": "code",

"execution\_count": 20,

"id": "2da9c9ee",

"metadata": {},

"outputs": [

{

"data": {

"text/html": [

"<div>\n",

"<style scoped>\n",

" .dataframe tbody tr th:only-of-type {\n",

" vertical-align: middle;\n",

" }\n",

"\n",

" .dataframe tbody tr th {\n",

" vertical-align: top;\n",

" }\n",

"\n",

" .dataframe thead th {\n",

" text-align: right;\n",

" }\n",

"</style>\n",

"<table border=\"1\" class=\"dataframe\">\n",

" <thead>\n",

" <tr style=\"text-align: right;\">\n",

" <th></th>\n",

" <th>sub\_icb\_location\_code</th>\n",

" <th>sub\_icb\_location\_ons\_code</th>\n",

" <th>sub\_icb\_location\_name</th>\n",

" <th>icb\_ons\_code</th>\n",

" <th>region\_ons\_code</th>\n",

" <th>appointment\_date</th>\n",

" <th>actual\_duration</th>\n",

" <th>count\_of\_appointments</th>\n",

" </tr>\n",

" </thead>\n",

" <tbody>\n",

" <tr>\n",

" <th>0</th>\n",

" <td>00L</td>\n",

" <td>E38000130</td>\n",

" <td>NHS North East and North Cumbria ICB - 00L</td>\n",

" <td>E54000050</td>\n",

" <td>E40000012</td>\n",

" <td>01-Dec-21</td>\n",

" <td>31-60 Minutes</td>\n",

" <td>364</td>\n",

" </tr>\n",

" <tr>\n",

" <th>1</th>\n",

" <td>00L</td>\n",

" <td>E38000130</td>\n",

" <td>NHS North East and North Cumbria ICB - 00L</td>\n",

" <td>E54000050</td>\n",

" <td>E40000012</td>\n",

" <td>01-Dec-21</td>\n",

" <td>21-30 Minutes</td>\n",

" <td>619</td>\n",

" </tr>\n",

" <tr>\n",

" <th>2</th>\n",

" <td>00L</td>\n",

" <td>E38000130</td>\n",

" <td>NHS North East and North Cumbria ICB - 00L</td>\n",

" <td>E54000050</td>\n",

" <td>E40000012</td>\n",

" <td>01-Dec-21</td>\n",

" <td>6-10 Minutes</td>\n",

" <td>1698</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3</th>\n",

" <td>00L</td>\n",

" <td>E38000130</td>\n",

" <td>NHS North East and North Cumbria ICB - 00L</td>\n",

" <td>E54000050</td>\n",

" <td>E40000012</td>\n",

" <td>01-Dec-21</td>\n",

" <td>Unknown / Data Quality</td>\n",

" <td>1277</td>\n",

" </tr>\n",

" <tr>\n",

" <th>4</th>\n",

" <td>00L</td>\n",

" <td>E38000130</td>\n",

" <td>NHS North East and North Cumbria ICB - 00L</td>\n",

" <td>E54000050</td>\n",

" <td>E40000012</td>\n",

" <td>01-Dec-21</td>\n",

" <td>16-20 Minutes</td>\n",

" <td>730</td>\n",

" </tr>\n",

" </tbody>\n",

"</table>\n",

"</div>"

],

"text/plain": [

" sub\_icb\_location\_code sub\_icb\_location\_ons\_code \\\n",

"0 00L E38000130 \n",

"1 00L E38000130 \n",

"2 00L E38000130 \n",

"3 00L E38000130 \n",

"4 00L E38000130 \n",

"\n",

" sub\_icb\_location\_name icb\_ons\_code region\_ons\_code \\\n",

"0 NHS North East and North Cumbria ICB - 00L E54000050 E40000012 \n",

"1 NHS North East and North Cumbria ICB - 00L E54000050 E40000012 \n",

"2 NHS North East and North Cumbria ICB - 00L E54000050 E40000012 \n",

"3 NHS North East and North Cumbria ICB - 00L E54000050 E40000012 \n",

"4 NHS North East and North Cumbria ICB - 00L E54000050 E40000012 \n",

"\n",

" appointment\_date actual\_duration count\_of\_appointments \n",

"0 01-Dec-21 31-60 Minutes 364 \n",

"1 01-Dec-21 21-30 Minutes 619 \n",

"2 01-Dec-21 6-10 Minutes 1698 \n",

"3 01-Dec-21 Unknown / Data Quality 1277 \n",

"4 01-Dec-21 16-20 Minutes 730 "

]

},

"execution\_count": 20,

"metadata": {},

"output\_type": "execute\_result"

}

],

"source": [

"# View the first five rows of appointment\_date for the ad DataFrame to determine the date format.\n",

"ad['appointment\_date']\n",

"ad.head(5)"

]

},

{

"cell\_type": "code",

"execution\_count": 21,

"id": "f1763690",

"metadata": {},

"outputs": [

{

"data": {

"text/html": [

"<div>\n",

"<style scoped>\n",

" .dataframe tbody tr th:only-of-type {\n",

" vertical-align: middle;\n",

" }\n",

"\n",

" .dataframe tbody tr th {\n",

" vertical-align: top;\n",

" }\n",

"\n",

" .dataframe thead th {\n",

" text-align: right;\n",

" }\n",

"</style>\n",

"<table border=\"1\" class=\"dataframe\">\n",

" <thead>\n",

" <tr style=\"text-align: right;\">\n",

" <th></th>\n",

" <th>icb\_ons\_code</th>\n",

" <th>appointment\_month</th>\n",

" <th>appointment\_status</th>\n",

" <th>hcp\_type</th>\n",

" <th>appointment\_mode</th>\n",

" <th>time\_between\_book\_and\_appointment</th>\n",

" <th>count\_of\_appointments</th>\n",

" </tr>\n",

" </thead>\n",

" <tbody>\n",

" <tr>\n",

" <th>0</th>\n",

" <td>E54000034</td>\n",

" <td>2020-01</td>\n",

" <td>Attended</td>\n",

" <td>GP</td>\n",

" <td>Face-to-Face</td>\n",

" <td>1 Day</td>\n",

" <td>8107</td>\n",

" </tr>\n",

" <tr>\n",

" <th>1</th>\n",

" <td>E54000034</td>\n",

" <td>2020-01</td>\n",

" <td>Attended</td>\n",

" <td>GP</td>\n",

" <td>Face-to-Face</td>\n",

" <td>15 to 21 Days</td>\n",

" <td>6791</td>\n",

" </tr>\n",

" <tr>\n",

" <th>2</th>\n",

" <td>E54000034</td>\n",

" <td>2020-01</td>\n",

" <td>Attended</td>\n",

" <td>GP</td>\n",

" <td>Face-to-Face</td>\n",

" <td>2 to 7 Days</td>\n",

" <td>20686</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3</th>\n",

" <td>E54000034</td>\n",

" <td>2020-01</td>\n",

" <td>Attended</td>\n",

" <td>GP</td>\n",

" <td>Face-to-Face</td>\n",

" <td>22 to 28 Days</td>\n",

" <td>4268</td>\n",

" </tr>\n",

" <tr>\n",

" <th>4</th>\n",

" <td>E54000034</td>\n",

" <td>2020-01</td>\n",

" <td>Attended</td>\n",

" <td>GP</td>\n",

" <td>Face-to-Face</td>\n",

" <td>8 to 14 Days</td>\n",

" <td>11971</td>\n",

" </tr>\n",

" </tbody>\n",

"</table>\n",

"</div>"

],

"text/plain": [

" icb\_ons\_code appointment\_month appointment\_status hcp\_type appointment\_mode \\\n",

"0 E54000034 2020-01 Attended GP Face-to-Face \n",

"1 E54000034 2020-01 Attended GP Face-to-Face \n",

"2 E54000034 2020-01 Attended GP Face-to-Face \n",

"3 E54000034 2020-01 Attended GP Face-to-Face \n",

"4 E54000034 2020-01 Attended GP Face-to-Face \n",

"\n",

" time\_between\_book\_and\_appointment count\_of\_appointments \n",

"0 1 Day 8107 \n",

"1 15 to 21 Days 6791 \n",

"2 2 to 7 Days 20686 \n",

"3 22 to 28 Days 4268 \n",

"4 8 to 14 Days 11971 "

]

},

"execution\_count": 21,

"metadata": {},

"output\_type": "execute\_result"

}

],

"source": [

"# View the first five rows of appointment\_month for the ar DataFrame to determine the date format.\n",

"ar['appointment\_month']\n",

"ar.head(5)"

]

},

{

"cell\_type": "code",

"execution\_count": 22,

"id": "92d5c6c7",

"metadata": {},

"outputs": [

{

"name": "stdout",

"output\_type": "stream",

"text": [

"01-Apr-22\n",

"31-May-22\n",

"The earliest date in the actual\_duration dataset is 01-Apr-22\n",

"The latest date in the actual\_duration dataset is 31-May-22\n"

]

}

],

"source": [

"#Determine what dates appointments were scheduled between in ad DataFrame\n",

"print(ad['appointment\_date'].min())\n",

"print(ad['appointment\_date'].max())\n",

"print ( f\"The earliest date in the actual\_duration dataset is {ad['appointment\_date'].min()}\")\n",

"print (f\"The latest date in the actual\_duration dataset is {ad['appointment\_date'].max()}\")"

]

},

{

"cell\_type": "code",

"execution\_count": 23,

"id": "6f3ae8b5",

"metadata": {},

"outputs": [

{

"name": "stdout",

"output\_type": "stream",

"text": [

"2020-01\n",

"2022-06\n",

"The earliest month in the appointments\_regional dataset is 2020-01\n",

"The latest month in the appointments\_regional dataset is 2022-06\n"

]

}

],

"source": [

"#Determine what dates appointments were scheduled between in ar DataFrame\n",

"print(ar['appointment\_month'].min())\n",

"print(ar['appointment\_month'].max())\n",

"print ( f\"The earliest month in the appointments\_regional dataset is {ar['appointment\_month'].min()}\")\n",

"print (f\"The latest month in the appointments\_regional dataset is {ar['appointment\_month'].max()}\")"

]

},

{

"cell\_type": "code",

"execution\_count": 24,

"id": "e7772d6e",

"metadata": {

"scrolled": false

},

"outputs": [

{

"name": "stdout",

"output\_type": "stream",

"text": [

"2021-08-01 00:00:00\n",

"2022-06-30 00:00:00\n",

"The earliest date in the national\_categories dataset is 2021-08-01 00:00:00\n",

"The latest date in the national\_categories dataset is 2022-06-30 00:00:00\n"

]

}

],

"source": [

"#Determine what dates appointments were scheduled between in nc DataFrame\n",

"print(nc['appointment\_date'].min())\n",

"print(nc['appointment\_date'].max())\n",

"print ( f\"The earliest date in the national\_categories dataset is {nc['appointment\_date'].min()}\")\n",

"print (f\"The latest date in the national\_categories dataset is {nc['appointment\_date'].max()}\")"

]

},

{

"cell\_type": "code",

"execution\_count": 25,

"id": "15c574e7",

"metadata": {},

"outputs": [

{

"data": {

"text/html": [

"<div>\n",

"<style scoped>\n",

" .dataframe tbody tr th:only-of-type {\n",

" vertical-align: middle;\n",

" }\n",

"\n",

" .dataframe tbody tr th {\n",

" vertical-align: top;\n",

" }\n",

"\n",

" .dataframe thead th {\n",

" text-align: right;\n",

" }\n",

"</style>\n",

"<table border=\"1\" class=\"dataframe\">\n",

" <thead>\n",

" <tr style=\"text-align: right;\">\n",

" <th></th>\n",

" <th>appointment\_date</th>\n",

" <th>icb\_ons\_code</th>\n",

" <th>sub\_icb\_location\_name</th>\n",

" <th>service\_setting</th>\n",

" <th>context\_type</th>\n",

" <th>national\_category</th>\n",

" <th>count\_of\_appointments</th>\n",

" <th>appointment\_month</th>\n",

" </tr>\n",

" </thead>\n",

" <tbody>\n",

" <tr>\n",

" <th>805547</th>\n",

" <td>2022-05-16</td>\n",

" <td>E54000027</td>\n",

" <td>NHS North West London ICB - W2U3Z</td>\n",

" <td>General Practice</td>\n",

" <td>Care Related Encounter</td>\n",

" <td>General Consultation Routine</td>\n",

" <td>15056</td>\n",

" <td>2022-05</td>\n",

" </tr>\n",

" <tr>\n",

" <th>805256</th>\n",

" <td>2022-05-09</td>\n",

" <td>E54000027</td>\n",

" <td>NHS North West London ICB - W2U3Z</td>\n",

" <td>General Practice</td>\n",

" <td>Care Related Encounter</td>\n",

" <td>General Consultation Routine</td>\n",

" <td>14964</td>\n",

" <td>2022-05</td>\n",

" </tr>\n",

" <tr>\n",

" <th>805832</th>\n",

" <td>2022-05-23</td>\n",

" <td>E54000027</td>\n",

" <td>NHS North West London ICB - W2U3Z</td>\n",

" <td>General Practice</td>\n",

" <td>Care Related Encounter</td>\n",

" <td>General Consultation Routine</td>\n",

" <td>14907</td>\n",

" <td>2022-05</td>\n",

" </tr>\n",

" <tr>\n",

" <th>805595</th>\n",

" <td>2022-05-17</td>\n",

" <td>E54000027</td>\n",

" <td>NHS North West London ICB - W2U3Z</td>\n",

" <td>General Practice</td>\n",

" <td>Care Related Encounter</td>\n",

" <td>General Consultation Routine</td>\n",

" <td>14904</td>\n",

" <td>2022-05</td>\n",

" </tr>\n",

" <tr>\n",

" <th>805311</th>\n",

" <td>2022-05-10</td>\n",

" <td>E54000027</td>\n",

" <td>NHS North West London ICB - W2U3Z</td>\n",

" <td>General Practice</td>\n",

" <td>Care Related Encounter</td>\n",

" <td>General Consultation Routine</td>\n",

" <td>14818</td>\n",

" <td>2022-05</td>\n",

" </tr>\n",

" </tbody>\n",

"</table>\n",

"</div>"

],

"text/plain": [

" appointment\_date icb\_ons\_code sub\_icb\_location\_name \\\n",

"805547 2022-05-16 E54000027 NHS North West London ICB - W2U3Z \n",

"805256 2022-05-09 E54000027 NHS North West London ICB - W2U3Z \n",

"805832 2022-05-23 E54000027 NHS North West London ICB - W2U3Z \n",

"805595 2022-05-17 E54000027 NHS North West London ICB - W2U3Z \n",

"805311 2022-05-10 E54000027 NHS North West London ICB - W2U3Z \n",

"\n",

" service\_setting context\_type \\\n",

"805547 General Practice Care Related Encounter \n",

"805256 General Practice Care Related Encounter \n",

"805832 General Practice Care Related Encounter \n",

"805595 General Practice Care Related Encounter \n",

"805311 General Practice Care Related Encounter \n",

"\n",

" national\_category count\_of\_appointments appointment\_month \n",

"805547 General Consultation Routine 15056 2022-05 \n",

"805256 General Consultation Routine 14964 2022-05 \n",

"805832 General Consultation Routine 14907 2022-05 \n",

"805595 General Consultation Routine 14904 2022-05 \n",

"805311 General Consultation Routine 14818 2022-05 "

]

},

"execution\_count": 25,

"metadata": {},

"output\_type": "execute\_result"

}

],

"source": [

"#Determine service setting that reported the most appointments in North West London from 1 January to 1 June 2022\n",

"nwlondon=nc.loc[nc['sub\_icb\_location\_name'].str.contains(\"NHS North West London ICB - W2U3Z\",case=False) &(nc['appointment\_date']>= '2022-01-01') &\\\n",

" (nc['appointment\_date']< '2022-06-02')]\n",

"\n",

"nwlondon.nlargest(n=5, columns=['count\_of\_appointments'],keep='all') "

]

},

{

"cell\_type": "code",

"execution\_count": 80,

"id": "8118444f",

"metadata": {},

"outputs": [

{

"name": "stdout",

"output\_type": "stream",

"text": [

" count\_of\_appointments\n",

"appointment\_month \n",

"2021-08 23852171\n",

"2022-04 23913060\n",

"2021-12 25140776\n",

"2022-02 25355260\n",

"2022-01 25635474\n",

"2022-06 25828078\n",

"2022-05 27495508\n",

"2021-09 28522501\n",

"2022-03 29595038\n",

"2021-10 30303834\n",

"2021-11 30405070\n"

]

}

],

"source": [

"#Count of appointments per month \n",

"monthly\_apps = nc.groupby('appointment\_month').sum()\n",

"monthly\_apps.sort\_values(by=['count\_of\_appointments'], inplace=True)\n",

"print(monthly\_apps)"

]

},

{

"cell\_type": "code",

"execution\_count": 82,

"id": "4965bd05",

"metadata": {},

"outputs": [],

"source": [

"# Import the necessary libraries.\n",

"import seaborn as sns\n",

"import matplotlib.pyplot as plt\n",

"\n",

"# Set figure size.\n",

"sns.set(rc={'figure.figsize':(15, 12)})\n",

"\n",

"# Set the plot style as white.\n",

"sns.set\_style('white')"

]

},

{

"cell\_type": "code",

"execution\_count": 1,

"id": "d5fec142",

"metadata": {},

"outputs": [

{

"ename": "NameError",

"evalue": "name 'sns' is not defined",

"output\_type": "error",

"traceback": [

"\u001b[1;31m---------------------------------------------------------------------------\u001b[0m",

"\u001b[1;31mNameError\u001b[0m Traceback (most recent call last)",

"Input \u001b[1;32mIn [1]\u001b[0m, in \u001b[0;36m<cell line: 2>\u001b[1;34m()\u001b[0m\n\u001b[0;32m 1\u001b[0m \u001b[38;5;66;03m# Create a lineplot for serivce settings.\u001b[39;00m\n\u001b[1;32m----> 2\u001b[0m time\_between\u001b[38;5;241m=\u001b[39m\u001b[43msns\u001b[49m\u001b[38;5;241m.\u001b[39mlineplot(x\u001b[38;5;241m=\u001b[39m\u001b[38;5;124m'\u001b[39m\u001b[38;5;124mappointment\_month\u001b[39m\u001b[38;5;124m'\u001b[39m, y\u001b[38;5;241m=\u001b[39m\u001b[38;5;124m'\u001b[39m\u001b[38;5;124mcount\_of\_appointments\u001b[39m\u001b[38;5;124m'\u001b[39m, hue\u001b[38;5;241m=\u001b[39m\u001b[38;5;124m'\u001b[39m\u001b[38;5;124mservice settings\u001b[39m\u001b[38;5;124m'\u001b[39m, data\u001b[38;5;241m=\u001b[39mnc, ci\u001b[38;5;241m=\u001b[39m\u001b[38;5;28;01mNone\u001b[39;00m)\n\u001b[0;32m 4\u001b[0m time\_between\u001b[38;5;241m.\u001b[39mset\_xlabel(\u001b[38;5;124m'\u001b[39m\u001b[38;5;124mMonth\u001b[39m\u001b[38;5;124m'\u001b[39m) \n\u001b[0;32m 5\u001b[0m time\_between\u001b[38;5;241m.\u001b[39mset\_ylabel(\u001b[38;5;124m'\u001b[39m\u001b[38;5;124mNo of Appointments\u001b[39m\u001b[38;5;124m'\u001b[39m)\n",

"\u001b[1;31mNameError\u001b[0m: name 'sns' is not defined"

]

}

],

"source": [

"# Create a lineplot for serivce settings.\n",

"time\_between=sns.lineplot(x='appointment\_month', y='count\_of\_appointments', hue='service settings', data=nc, ci=None)\n",

"\n",

"time\_between.set\_xlabel('Month') \n",

"time\_between.set\_ylabel('No of Appointments')\n",

"time\_between.set(title='Monthly trend in service settings') "

]

},

{

"cell\_type": "code",

"execution\_count": 88,

"id": "10b9c9fc",

"metadata": {},

"outputs": [

{

"data": {

"text/plain": [

"[Text(0.5, 1.0, 'Monthly trend in service setting')]"

]

},

"execution\_count": 88,

"metadata": {},

"output\_type": "execute\_result"

},

{

"data": {

"image/png": "\n",

"text/plain": [

"<Figure size 1080x864 with 1 Axes>"

]

},

"metadata": {},

"output\_type": "display\_data"

}

],

"source": [

"# Lineplot for service settings.\n",

"time\_between=sns.lineplot(x='appointment\_month', y='count\_of\_appointments', hue='service\_setting', data=nc, ci=None)\n",

"\n",

"time\_between.set\_xlabel('Month') \n",

"time\_between.set\_ylabel('No of appointments')\n",

"time\_between.set(title='Monthly trend in service setting') "

]

},

{

"cell\_type": "code",

"execution\_count": 89,

"id": "7c69de19",

"metadata": {},

"outputs": [

{

"data": {

"text/plain": [

"[Text(0.5, 1.0, 'Monthly trend in context type')]"

]

},

"execution\_count": 89,

"metadata": {},

"output\_type": "execute\_result"

},

{

"data": {

"image/png": "\n",

"text/plain": [

"<Figure size 1080x864 with 1 Axes>"

]

},

"metadata": {},

"output\_type": "display\_data"

}

],

"source": [

"# Lineplot context type.\n",

"time\_between=sns.lineplot(x='appointment\_month', y='count\_of\_appointments', hue='context\_type', data=nc, ci=None)\n",

"\n",

"time\_between.set\_xlabel('Month') \n",

"time\_between.set\_ylabel('No of appointments')\n",

"time\_between.set(title='Monthly trend in context type') "

]

},

{

"cell\_type": "code",

"execution\_count": 90,

"id": "9004215b",

"metadata": {},

"outputs": [

{

"data": {

"text/plain": [

"[Text(0.5, 1.0, 'Monthly trend in national categories')]"

]

},

"execution\_count": 90,

"metadata": {},

"output\_type": "execute\_result"

},

{

"data": {

"image/png": "\n",

"text/plain": [

"<Figure size 1080x864 with 1 Axes>"

]

},

"metadata": {},

"output\_type": "display\_data"

}

],

"source": [

"#Lineplot for national categories.\n",

"time\_between=sns.lineplot(x='appointment\_month', y='count\_of\_appointments', hue='national\_category', data=nc, ci=None)\n",

"\n",

"time\_between.set\_xlabel('Month') \n",

"time\_between.set\_ylabel('No of appointments')\n",

"time\_between.set(title='Monthly trend in national categories') "

]

},

{

"cell\_type": "code",

"execution\_count": 33,

"id": "0bbdb1e1",

"metadata": {},

"outputs": [

{

"name": "stdout",

"output\_type": "stream",

"text": [

" tweet\_id tweet\_full\_text \\\n",

"0 1567629223795527681 As Arkansas’ first Comprehensive Stroke Certif... \n",

"1 1567582846612553728 RT @AndreaGrammer: Work-life balance is at the... \n",

"2 1567582787070304256 RT @OntarioGreens: $10 billion can go a long w... \n",

"3 1567582767625428992 RT @modrnhealthcr: 🚨#NEW:🚨 Insurance companies... \n",

"4 1567582720460570625 ICYMI: Our recent blogs on Cybersecurity in Ac... \n",

"\n",

" tweet\_entities \\\n",

"0 {'hashtags': [{'text': 'Healthcare', 'indices'... \n",

"1 {'hashtags': [{'text': 'PremiseHealth', 'indic... \n",

"2 {'hashtags': [{'text': 'Healthcare', 'indices'... \n",

"3 {'hashtags': [{'text': 'NEW', 'indices': [20, ... \n",

"4 {'hashtags': [{'text': 'blogs', 'indices': [18... \n",

"\n",

" tweet\_entities\_hashtags \\\n",

"0 #Healthcare \n",

"1 #PremiseHealth, #hiring \n",

"2 #Healthcare \n",

"3 #NEW \n",

"4 #blogs, #digitaltransformation, #cybersecurity... \n",

"\n",

" tweet\_metadata tweet\_retweet\_count \\\n",

"0 {'iso\_language\_code': 'en', 'result\_type': 're... 0 \n",

"1 {'iso\_language\_code': 'en', 'result\_type': 're... 2 \n",

"2 {'iso\_language\_code': 'en', 'result\_type': 're... 39 \n",

"3 {'iso\_language\_code': 'en', 'result\_type': 're... 5 \n",

"4 {'iso\_language\_code': 'en', 'result\_type': 're... 0 \n",

"\n",

" tweet\_favorite\_count tweet\_favorited tweet\_retweeted tweet\_lang \n",

"0 0 False False en \n",

"1 0 False False en \n",

"2 0 False False en \n",

"3 0 False False en \n",

"4 0 False False en \n"

]

}

],

"source": [

"#Data frame for the tweets data \n",

"tweets = pd.read\_csv('tweets.csv')\n",

"\n",

"#Print the dataframe for tweets \n",

"print(tweets.head(5))"

]

},

{

"cell\_type": "code",

"execution\_count": 34,

"id": "759d3def",

"metadata": {},

"outputs": [

{

"name": "stdout",

"output\_type": "stream",

"text": [

"<class 'pandas.core.frame.DataFrame'>\n",

"RangeIndex: 1174 entries, 0 to 1173\n",

"Data columns (total 10 columns):\n",

" # Column Non-Null Count Dtype \n",

"--- ------ -------------- ----- \n",

" 0 tweet\_id 1174 non-null int64 \n",

" 1 tweet\_full\_text 1174 non-null object\n",

" 2 tweet\_entities 1174 non-null object\n",

" 3 tweet\_entities\_hashtags 1007 non-null object\n",

" 4 tweet\_metadata 1174 non-null object\n",

" 5 tweet\_retweet\_count 1174 non-null int64 \n",

" 6 tweet\_favorite\_count 1174 non-null int64 \n",

" 7 tweet\_favorited 1174 non-null bool \n",

" 8 tweet\_retweeted 1174 non-null bool \n",

" 9 tweet\_lang 1174 non-null object\n",

"dtypes: bool(2), int64(3), object(5)\n",

"memory usage: 75.8+ KB\n"

]

}

],

"source": [

"#Metadata for tweets \n",

"tweets.info()"

]

},

{

"cell\_type": "code",

"execution\_count": 35,

"id": "bde69aab",

"metadata": {},

"outputs": [

{

"data": {

"text/html": [

"<div>\n",

"<style scoped>\n",

" .dataframe tbody tr th:only-of-type {\n",

" vertical-align: middle;\n",

" }\n",

"\n",

" .dataframe tbody tr th {\n",

" vertical-align: top;\n",

" }\n",

"\n",

" .dataframe thead th {\n",

" text-align: right;\n",

" }\n",

"</style>\n",

"<table border=\"1\" class=\"dataframe\">\n",

" <thead>\n",

" <tr style=\"text-align: right;\">\n",

" <th></th>\n",

" <th>tweet\_id</th>\n",

" <th>tweet\_retweet\_count</th>\n",

" <th>tweet\_favorite\_count</th>\n",

" </tr>\n",

" </thead>\n",

" <tbody>\n",

" <tr>\n",

" <th>count</th>\n",

" <td>1.174000e+03</td>\n",

" <td>1174.000000</td>\n",

" <td>1174.00000</td>\n",

" </tr>\n",

" <tr>\n",

" <th>mean</th>\n",

" <td>1.567612e+18</td>\n",

" <td>8.629472</td>\n",

" <td>0.37138</td>\n",

" </tr>\n",

" <tr>\n",

" <th>std</th>\n",

" <td>2.427553e+13</td>\n",

" <td>29.784675</td>\n",

" <td>2.04470</td>\n",

" </tr>\n",

" <tr>\n",

" <th>min</th>\n",

" <td>1.567574e+18</td>\n",

" <td>0.000000</td>\n",

" <td>0.00000</td>\n",

" </tr>\n",

" <tr>\n",

" <th>25%</th>\n",

" <td>1.567590e+18</td>\n",

" <td>0.000000</td>\n",

" <td>0.00000</td>\n",

" </tr>\n",

" <tr>\n",

" <th>50%</th>\n",

" <td>1.567611e+18</td>\n",

" <td>1.000000</td>\n",

" <td>0.00000</td>\n",

" </tr>\n",

" <tr>\n",

" <th>75%</th>\n",

" <td>1.567633e+18</td>\n",

" <td>3.000000</td>\n",

" <td>0.00000</td>\n",

" </tr>\n",

" <tr>\n",

" <th>max</th>\n",

" <td>1.567655e+18</td>\n",

" <td>303.000000</td>\n",

" <td>42.00000</td>\n",

" </tr>\n",

" </tbody>\n",

"</table>\n",

"</div>"

],

"text/plain": [

" tweet\_id tweet\_retweet\_count tweet\_favorite\_count\n",

"count 1.174000e+03 1174.000000 1174.00000\n",

"mean 1.567612e+18 8.629472 0.37138\n",

"std 2.427553e+13 29.784675 2.04470\n",

"min 1.567574e+18 0.000000 0.00000\n",

"25% 1.567590e+18 0.000000 0.00000\n",

"50% 1.567611e+18 1.000000 0.00000\n",

"75% 1.567633e+18 3.000000 0.00000\n",

"max 1.567655e+18 303.000000 42.00000"

]

},

"execution\_count": 35,

"metadata": {},

"output\_type": "execute\_result"

}

],

"source": [

"#Descriptive statistics for tweets\n",

"tweets.describe()"

]

},

{

"cell\_type": "code",

"execution\_count": 36,

"id": "17c966e8",

"metadata": {},

"outputs": [

{

"data": {

"text/plain": [

"0 526\n",

"1 215\n",

"2 114\n",

"3 70\n",

"5 35\n",

"4 27\n",

"7 18\n",

"12 16\n",

"8 15\n",

"73 14\n",

"9 13\n",

"6 12\n",

"208 12\n",

"35 10\n",

"37 6\n",

"11 6\n",

"10 5\n",

"53 5\n",

"44 4\n",

"150 4\n",

"63 4\n",

"76 3\n",

"85 3\n",

"41 3\n",

"62 3\n",

"207 3\n",

"68 3\n",

"78 2\n",

"23 2\n",

"24 2\n",

"72 2\n",

"16 2\n",

"13 1\n",

"49 1\n",

"48 1\n",

"15 1\n",

"107 1\n",

"14 1\n",

"79 1\n",

"20 1\n",

"39 1\n",

"19 1\n",

"303 1\n",

"57 1\n",

"40 1\n",

"54 1\n",

"169 1\n",

"Name: tweet\_retweet\_count, dtype: int64"

]

},

"execution\_count": 36,

"metadata": {},

"output\_type": "execute\_result"

}

],

"source": [

"#List of tweet retweet count \n",

"tweets['tweet\_retweet\_count'].value\_counts()"

]

},

{

"cell\_type": "code",

"execution\_count": 37,

"id": "4984ddfb",

"metadata": {},

"outputs": [

{

"data": {

"text/plain": [

"0 1027\n",

"1 91\n",

"2 16\n",

"3 13\n",

"4 7\n",

"5 5\n",

"6 2\n",

"17 1\n",

"12 1\n",

"10 1\n",

"8 1\n",

"13 1\n",

"11 1\n",

"7 1\n",

"20 1\n",

"28 1\n",

"14 1\n",

"18 1\n",

"9 1\n",

"42 1\n",

"Name: tweet\_favorite\_count, dtype: int64"

]

},

"execution\_count": 37,

"metadata": {},

"output\_type": "execute\_result"

}

],

"source": [

"#List of tweet favourite count \n",

"tweets['tweet\_favorite\_count'].value\_counts()"

]

},

{

"cell\_type": "code",

"execution\_count": 38,

"id": "68848074",

"metadata": {},

"outputs": [

{

"data": {

"text/plain": [

"0 As Arkansas’ first Comprehensive Stroke Certif...\n",

"1 RT @AndreaGrammer: Work-life balance is at the...\n",

"2 RT @OntarioGreens: $10 billion can go a long w...\n",

"3 RT @modrnhealthcr: 🚨#NEW:🚨 Insurance companies...\n",

"4 ICYMI: Our recent blogs on Cybersecurity in Ac...\n",

" ... \n",

"1169 RT @PotomacPhotonic: Potomac #Innovation Repor...\n",

"1170 Not a cent towards workers who would like to a...\n",

"1171 The @hfmaorg Region 9 presents \"The Value of E...\n",

"1172 Happy physiotherapy 🩺 day 🎉..\\n#bpt #physiothe...\n",

"1173 RT @KimcoStaffing: Apply now to work for #Medi...\n",

"Name: tweet\_full\_text, Length: 1174, dtype: object"

]

},

"execution\_count": 38,

"metadata": {},

"output\_type": "execute\_result"

}

],

"source": [

"#Dataframe with only text values\n",

"tweet\_text=tweets['tweet\_full\_text']\n",

"\n",

"tweet\_text"

]

},

{

"cell\_type": "code",

"execution\_count": 39,

"id": "931edce9",

"metadata": {},

"outputs": [],

"source": [

"tags=[]"

]

},

{

"cell\_type": "code",

"execution\_count": 40,

"id": "88caa5d1",

"metadata": {},

"outputs": [

{

"data": {

"text/plain": [

"['#healthcare',\n",

" '#premisehealth',\n",

" '#hiring',\n",

" '#healthcare',\n",

" '🚨#new:🚨',\n",

" 'look!\\n\\n#blogs',\n",

" '#digitaltransformation',\n",

" '#cybersecurity',\n",

" '#accounting',\n",

" '#finance',\n",

" '#healthcare',\n",

" 'https://t.co/jrgqeqdme4\\n.\\n#firstcoastcna',\n",

" '#cnaexam',\n",

" '#cnaexampreparation',\n",

" '#jacksonville',\n",

" '#cnatraining',\n",

" '#nurse',\n",

" '#nursing',\n",

" '#nurselife',\n",

" '#nursepractitioner',\n",

" '#nurseproblems',\n",

" '#nursingschool',\n",

" '#healthcare',\n",

" '🚨#new:🚨',\n",

" '#disparities.',\n",

" '@karahartnett\\n#healthcare',\n",

" '#alert',\n",

" '#insurance',\n",

" '#data\\nhttps://t.co/h9hlamr7p9',\n",

" '#healthcare',\n",

" '#healthcare',\n",

" '❤️\\n\\n#healthcare',\n",

" '#healthcare',\n",

" '#hcldr',\n",

" '#premisehealth',\n",

" '#hiring',\n",

" '#premisehealth',\n",

" '#hiring',\n",

" '#healthcare',\n",

" '#qualitypatientcare',\n",

" '#jobs',\n",

" '#job',\n",

" '#ascp2022',\n",

" '#ascp100.',\n",

" '#healthcare',\n",

" '#healthsecretary',\n",

" '#healthcare',\n",

" '#ai,',\n",

" '#sdoh,',\n",

" '#healthcare',\n",

" '\\nhttps://t.co/ewe7jntj1e\\n#tropicana',\n",

" '#real',\n",

" '#juice',\n",

" '#healthcare',\n",

" '#watch',\n",

" '#worms',\n",

" '#fruits',\n",

" '#healthtips',\n",

" '#tips',\n",

" '#healthcare',\n",

" '#thewoodlands,',\n",

" '#healthcare',\n",

" '#chicago,',\n",

" '#healthcare',\n",

" '#telehealth',\n",

" '#healthcare',\n",

" '#virtualcare',\n",

" '#mediqueststaffing',\n",

" '#hospital',\n",

" '#shift',\n",

" '(#newportbeach)',\n",

" '#job',\n",

" '#mediqueststaffing',\n",

" '#hospital',\n",

" '#shift',\n",

" '(#newportbeach)',\n",

" '#job',\n",

" '#jobs',\n",

" '#employment',\n",

" '#healthcare',\n",

" '#job',\n",

" '#hiring',\n",

" '#antiracist',\n",

" '#healthcare:',\n",

" '#newhealthcare',\n",

" '#healthcare',\n",

" '#hospitals',\n",

" '#physicians',\n",

" 'of\\xa0#healthcare\\xa0delivery:',\n",

" '#newhealthcare\\xa0',\n",

" '#healthcare',\n",

" '#cloud',\n",

" '#ehr',\n",

" '#datasecurity',\n",

" '#options',\n",

" '#optionstrading',\n",

" '#pfizer',\n",

" '#optionsflow',\n",

" '#stocks',\n",

" '#stock',\n",

" '#stockmarket',\n",

" '#investing',\n",

" '#investment',\n",

" '#healthcare',\n",

" '#healthcare',\n",

" '#womenleaders',\n",

" '#healthcare',\n",

" '#albuquerque,',\n",

" '#healthcare',\n",

" '#healthcare',\n",

" 'future\\n\\nhttps://t.co/hctoc1vady\\n\\n#telehealth',\n",

" '#telemedicine',\n",

" '#healthcare',\n",

" '#therapy',\n",

" '#primarycare',\n",

" '#home',\n",

" '#hit',\n",

" '#covid19',\n",

" '#usa',\n",

" '#selfcare',\n",

" '#coronavirus',\n",

" '#photooftheday',\n",

" '#healthcare.',\n",

" '#health.',\n",

" '#internalmedicine',\n",

" '#tipsfornewdocs',\n",

" '#meded',\n",

" '#medtwitter',\n",

" '#medicine',\n",

" '#medical',\n",

" '#medica…',\n",

" '#healthcare',\n",

" '#covid!',\n",

" '#stra…',\n",

" '#strategy',\n",

" '#competitiveintellige…',\n",

" '#strate…',\n",

" '#strategy',\n",

" '#competi…',\n",

" '#strategy',\n",

" '#competitiveintelligence',\n",

" '#marketing',\n",

" '#healthcare',\n",

" '#biotech',\n",

" '#pharma',\n",

" '#competitivemarketing',\n",

" '#pharmaceutical',\n",

" '#strategy',\n",

" '#competitiveintelligence',\n",

" '#marketing',\n",

" '#pharmaceutical',\n",

" '#competitivemarketing',\n",

" '#biotech',\n",

" '#healthcare',\n",

" '#pharma',\n",

" '#internalmedicine',\n",

" '#tipsfornewdocs',\n",

" '#meded',\n",

" '#medtwitter',\n",

" '#medicine',\n",

" '#medical',\n",

" '#medica…',\n",

" '#strategy',\n",

" '#competitiveintelligence',\n",

" '#marketing',\n",

" '#pharmaceutical',\n",

" '#biotech',\n",

" '#competitivemarketing',\n",

" '#pharma',\n",

" '#healthcare',\n",

" '#strategy',\n",

" '#competitiveintelligence',\n",

" '#marketing',\n",

" '#pharmaceutical',\n",

" '#pharma',\n",

" '#biotech',\n",

" '#competitivemarketing',\n",

" '#healthcare',\n",

" '#strategy',\n",

" '#competitiveintelligence',\n",

" '#marketing',\n",

" '#healthcare',\n",

" '#pharmaceutical',\n",

" '#competitivemarketing',\n",

" '#pharma',\n",

" '#biotech',\n",

" '#strategy',\n",

" '#competitiveintelligence',\n",

" '#marketing',\n",

" '#healthcare',\n",

" '#biotech',\n",

" '#competitivemarketing',\n",

" '#pharma',\n",

" '#pharmaceutical',\n",

" '#strategy',\n",

" '#competitiveintelligence',\n",

" '#marketing',\n",

" '#pharma',\n",

" '#healthcare',\n",

" '#biotech',\n",

" '#competitivemarketing',\n",

" '#pharmaceutical',\n",

" '#strategy',\n",

" '#competitiveintelligence',\n",

" '#marketing',\n",

" '#pharma',\n",

" '#competitivemarketing',\n",

" '#pharmaceutical',\n",

" '#healthcare',\n",

" '#biotech',\n",

" '#strategy',\n",

" '#competitiveintelligence',\n",

" '#marketing',\n",

" '#healthcare',\n",

" '#biotech',\n",

" '#pharma',\n",

" '#productmarketing',\n",

" '#pharmaceutical',\n",

" '#strategy',\n",

" '#competitiveintelligence',\n",

" '#marketing',\n",

" '#productmarketing',\n",

" '#healthcare',\n",

" '#pharma',\n",

" '#biotech',\n",

" '#pharmaceutical',\n",

" '#strategy',\n",

" '#competitiveintelligence',\n",

" '#marketing',\n",

" '#biotech',\n",

" '#pharmaceutical',\n",

" '#healthcare',\n",

" '#pharma',\n",

" '#productmarketing',\n",

" '#lenexa,',\n",

" '#healthcare',\n",

" '#healthcare',\n",

" 'delivery.\\nhttps://t.co/etq4tbb6vu\\n#hippotechnologies',\n",

" '#virtualcare',\n",

" '#headset',\n",

" '#wearabletech',\n",

" '#wearables',\n",

" '#healthcare',\n",

" '#firstresponder',\n",

" '#teletriage',\n",

" '#worldhealthorganization',\n",

" '#ascp2022',\n",

" '#ascp100.',\n",

" '#healthcare.',\n",

" '\\n\\n#healthcare',\n",

" '#revenuecycle',\n",

" '#healthit',\n",

" '#saskatchewan',\n",

" 'https://t.co/xxahgpsvsz\\n#healthcare',\n",

" '#business',\n",

" '#datasecurity',\n",

" '#compliance\\ncloud',\n",

" '#vaccine',\n",

" '#covid',\n",

" '#coronavirus',\n",

" '#pandemic',\n",

" '#health',\n",

" '#corona',\n",

" '#vaccines',\n",

" '#vaccination',\n",

" '#virus',\n",

" '#medicine',\n",

" '#lockdown',\n",

" '#science',\n",

" '#qua…',\n",

" '#vaccine',\n",

" '#covid',\n",

" '#coronavirus',\n",

" '#pandemic',\n",

" '#health',\n",

" '#corona',\n",

" '#vaccines',\n",

" '#vaccination',\n",

" '#virus',\n",

" '#medicine',\n",

" '#lockdown',\n",

" '#science',\n",

" '#qua…',\n",

" '#vaccine',\n",

" '#covid',\n",

" '#coronavirus',\n",

" '#pandemic',\n",

" '#health',\n",

" '#corona',\n",

" '#vaccines',\n",

" '#vaccination',\n",

" '#virus',\n",

" '#medicine',\n",

" '#lockdown',\n",

" '#science',\n",

" '#qua…',\n",

" '#epic',\n",

" '#identitybridge…',\n",

" '#epic',\n",

" '#identitybridge…',\n",

" '#medicaid',\n",

" '#heartfailure',\n",

" '#barostim',\n",

" '@lluhealth!\\n\\n#outsmarttheheart',\n",

" '#cardiology',\n",

" '#healthcare',\n",

" '#medtech',\n",

" '#innovation',\n",

" '#familydoctor.',\n",

" '#globalaussieawards',\n",

" '#vr',\n",

" '#ar',\n",

" '#healthcare',\n",

" '#healthcare',\n",

" 'itself\\n\\n#cdnpoli',\n",

" '#holyoke,',\n",

" '#healthcare',\n",

" '#healthcare',\n",

" '#mammogr…',\n",

" '#healthcare',\n",

" '#speechpathology',\n",

" '#healthcare',\n",

" 'https://t.co/bmvwpsfker\\n\\n#fitness',\n",

" '#beuty',\n",

" '#healthcare',\n",

" '#fitnesslife',\n",

" '#beauty',\n",

" '#beautytips',\n",

" '#beaut',\n",

" '#healthcare',\n",

" '#birmingham,',\n",

" '#privacyofficer',\n",

" '#healthcare',\n",

" '#webinar',\n",

" 'come!!\\n\\n#race4…',\n",

" '#healthcareisahumanright',\n",

" '#prep',\n",

" '#hiv',\n",

" '#healthcare',\n",

" '#bluecrew',\n",

" '#voteblue',\n",

" '#nursetwitter',\n",

" '#healthcare',\n",

" '#epic',\n",

" '#identitybridge',\n",

" '\\n\\n#itsecurity',\n",

" '#cybersecurity',\n",

" '#iam',\n",

" '#datasecurity',\n",

" '#healthcare',\n",

" '#hipaa',\n",

" '#healthcare',\n",

" '#prep',\n",

" '#healthcare',\n",

" 'https://t.co/luvbnj8opf\\n\\n#healthcare',\n",

" '#innovation',\n",

" '#customerexperience',\n",

" '#healthcare',\n",

" '#healthcare',\n",

" '#security',\n",

" '#healthcare',\n",

" '#centralized',\n",

" '\\n#humanrights',\n",

" '#healthcare',\n",

" '\\n#truth',\n",

" '#truth',\n",

" '\\n#nazi',\n",

" '#vaccine',\n",

" '#covid',\n",

" '#coronavirus',\n",

" '#pandemic',\n",

" '#health',\n",

" '#corona',\n",

" '#vaccines',\n",

" '#vaccination',\n",

" '#virus',\n",

" '#medicine',\n",

" '#lockdown',\n",

" '#science',\n",

" '#qua…',\n",

" '#vaccine',\n",

" '#covid',\n",

" '#coronavirus',\n",

" '#pandemic',\n",

" '#health',\n",

" '#corona',\n",

" '#vaccines',\n",

" '#vaccination',\n",

" '#virus',\n",

" '#medicine',\n",

" '#lockdown',\n",

" '#science',\n",

" '#qua…',\n",

" '#vaccine',\n",

" '#covid',\n",

" '#coronavirus',\n",

" '#pandemic',\n",

" '#health',\n",

" '#corona',\n",

" '#vaccines',\n",

" '#vaccination',\n",

" '#virus',\n",

" '#medicine',\n",

" '#lockdown',\n",

" '#science',\n",

" '#quarantine',\n",

" '#influenza',\n",

" '#news',\n",

" '#doctor',\n",

" '#stayhome',\n",

" '#healthcare',\n",

" '#who',\n",

" '#flu',\n",

" '#immunization',\n",

" '#truth',\n",

" '#social-distancing',\n",

" '\\n#detroit',\n",

" '#vaccine',\n",

" '#covid',\n",

" '#coronavirus',\n",

" '#pandemic',\n",

" '#health',\n",

" '#corona',\n",

" '#vaccines',\n",

" '#vaccination',\n",

" '#virus',\n",

" '#medicine',\n",

" '#lockdown',\n",

" '#science',\n",

" '#qua…',\n",

" '#vaccine',\n",

" '#covid',\n",

" '#coronavirus',\n",

" '#pandemic',\n",

" '#health',\n",

" '#corona',\n",

" '#vaccines',\n",

" '#vaccination',\n",

" '#virus',\n",

" '#medicine',\n",

" '#lockdown',\n",

" '#science',\n",

" '#quarantine',\n",

" '#influenza',\n",

" '#news',\n",

" '#doctor',\n",

" '#stayhome',\n",

" '#healthcare',\n",

" '#who',\n",

" '#flu',\n",

" '#immunization',\n",

" '#truth',\n",

" '#social-distancing',\n",

" '\\n#detroit',\n",

" '#vaccine',\n",

" '#covid',\n",

" '#coronavirus',\n",

" '#pandemic',\n",

" '#health',\n",

" '#corona',\n",

" '#vaccines',\n",

" '#vaccination',\n",

" '#virus',\n",

" '#medicine',\n",

" '#lockdown',\n",

" '#science',\n",

" '#quarantine',\n",

" '#influenza',\n",

" '#news',\n",

" '#doctor',\n",

" '#stayhome',\n",

" '#healthcare',\n",

" '#who',\n",

" '#flu',\n",

" '#immunization',\n",

" '#truth',\n",

" '#social-distancing',\n",

" '\\n#detroit',\n",

" '#data',\n",

" '#healthcare?',\n",

" '#healthcare',\n",

" '#healthcare',\n",

" '#ontariohealthcare',\n",

" '#healthcare',\n",

" '#regulation',\n",

" '#itsm',\n",

" '#health',\n",

" '#healthcare',\n",

" '#tech…',\n",

" '#healthcare',\n",

" '#palliative',\n",

" '#care',\n",

" '#healthcare',\n",

" '#congenital',\n",

" '#cardiotwitter',\n",

" '#doctors',\n",

" '#hospital',\n",

" '#clinics',\n",

" '#health',\n",

" '#healthcareit',\n",

" '#ehealth',\n",

" '#healthcare',\n",

" '#healthcare',\n",

" '#ai',\n",

" '#healthcare',\n",

" '#technology',\n",

" '#soaak',\n",

" '#healthcare',\n",

" '#mentalhealth',\n",

" '#therapy',\n",

" '#anxiety',\n",

" '#stress',\n",

" '#emergingtech',\n",

" '#coruzant',\n",

" '#thedigitalexecutive',\n",

" '#dental',\n",

" '#practice',\n",

" '#patients',\n",

" 'team.\\nhttps://t.co/jt60d4ab4v\\n#dentalcare',\n",

" '#dentistry',\n",

" '#healthtech',\n",

" '#healthcareit',\n",

" '#healthcare',\n",

" '#digitalhealth',\n",

" '#healthcare',\n",

" '#technology',\n",

" '#healthcare',\n",

" '#ontario',\n",

" '#generalstrike',\n",

" '#workersunite',\n",

" '#onpoli',\n",

" '#cdnpoli',\n",

" '#education',\n",

" '#healthcare',\n",

" '#healthcare',\n",

" '#chicago,',\n",

" '#specialneedsparenting',\n",

" '#healthcare',\n",

" '\\n\\n#skincare',\n",

" '#technology',\n",

" '#cosmetics',\n",

" '#beauty',\n",

" '#healt…',\n",

" '#healthcare',\n",

" '#healthcare',\n",

" '#ai',\n",

" '#cheverly,',\n",

" '#healthcare',\n",

" '\\n#lifescience',\n",

" '#healthcare',\n",

" '#quantiphi',\n",

" '#ai-led',\n",

" '#medicalimaging:',\n",

" '\\n#solvingwhatmatters',\n",

" '#healthcare',\n",

" '#aiinhealthcare',\n",

" '#ai',\n",

" '#healthcare',\n",

" '#drugdiscovery\\n@closedloopai',\n",

" '#patientdata…',\n",

" '#monkeypox',\n",

" '\\n#monkeypoxcases',\n",

" '#monkeypoxvaccine',\n",

" '#deathrate',\n",

" '#vaccine',\n",

" '#healthcare',\n",

" '#monkeypox',\n",

" '\\n#monkeypoxcases',\n",

" '#monkeypoxvaccine',\n",

" '#deathrate',\n",

" '#vaccine',\n",

" '#healthcare',\n",

" '#antiracist',\n",

" '#healthcare:',\n",

" '#örebro',\n",

" '#healthcare',\n",

" '#career',\n",

" '–#transportation/#construction/#telecommunications/#lawenforcement/#healthcare',\n",

" '#veterans',\n",

" '#healthcare',\n",

" '#mammogram',\n",

" '#mobilehealth',\n",

" '#healthsecutity',\n",

" '#breastcancerawareness',\n",

" '#ruralhealth',\n",

" '#healthcare',\n",

" '\\n#medicine',\n",

" '#doctor',\n",

" '#healthcare…',\n",

" '#primarycare',\n",

" '#healthcare',\n",

" '#maryland',\n",

" '#healthcare,',\n",

" '#healthcare',\n",

" '#covid19',\n",

" 'further👉https://t.co/zkx4vl4sxf\\n\\n#clubbing',\n",

" '#nail',\n",

" '#diseases',\n",

" '#health',\n",

" '#healthcare',\n",

" '#healthylifestyle',\n",

" '\\n\\nhttps://t.co/n8mboeqnbk\\n\\n#healthcare',\n",

" '#personalfinance',\n",

" '#medical',\n",

" '#healthcare…',\n",

" '#ai',\n",

" '#healthcare',\n",

" '#drugdiscovery\\n@closedloopai',\n",

" '#patientdata\\n@aitherightway',\n",

" '#eyedisease\\n@cleerly\_inc',\n",

" '#imaging\\n@owkinscience',\n",

" '#personalizedmedicine\\n@deepcell\_inc',\n",

" '#cellbiology',\n",

" '#research',\n",

" '#memphis:',\n",

" '#media',\n",

" '#socialmedia',\n",

" '#healthcare',\n",

" '#business.',\n",

" 'firm.\\n\\n#healthcare',\n",

" '#medical',\n",

" '#healthcare',\n",

" '#n95',\n",

" '#masks',\n",

" '#cdnpoli',\n",

" '#healthcare',\n",

" '#goals',\n",

" '#potential.',\n",

" '#nursingcare',\n",

" '#linkinbio',\n",

" '#nursing',\n",

" '#ohio',\n",

" '#health',\n",

" '#healthcare.',\n",

" 'planting...\\n\\n#appleevent',\n",

" '#iphone',\n",

" '#ems',\n",

" '#healthcare',\n",

" '#genzeon',\n",

" '#healthcare',\n",

" '#automation',\n",

" 't-square\\n#financialastrology',\n",

" '#gold',\n",

" '#nasdaq',\n",

" '#hea…',\n",

" '#healthcare',\n",

" '#healthcare',\n",

" '#chicago,',\n",

" '#substanceabuse',\n",

" '#healthcare',\n",

" '#healthcare.',\n",

" 'needs.\\n\\n#appleevent',\n",

" '#match',\n",

" '#healthcare',\n",

" '#education',\n",

" '#usa',\n",

" '#ruralhealth',\n",

" 'system!#match2023',\n",

" '#match2023madness',\n",

" '#usmle',\n",

" '#img',\n",

" '#cloud',\n",

" '#healthcare',\n",

" '#healthcare',\n",

" '#chicago,',\n",

" '#healthcare',\n",

" '#saskatchewan',\n",

" '#ai',\n",

" '#healthcare!\\n\\ninfographic',\n",

" '@ingliguori\\n\\n#python',\n",

" '#nodejs',\n",

" '#machinelearning',\n",

" '#100daysofmlcode',\n",

" '#html5',\n",

" '#rstats',\n",

" '#s…',\n",

" '#healthcare',\n",

" '#losangeles,',\n",

" '\\n\\n#comedy',\n",

" '#adulting',\n",

" '#healthcare',\n",

" '#esmo22:',\n",

" '#esmo22:',\n",

" '#ai',\n",

" '#healthcare!\\n\\ninfographic',\n",

" '@ingliguori\\n\\n#python',\n",

" '#nodejs',\n",

" '#machinelearning',\n",

" '#100daysofmlcode',\n",

" '#html5',\n",

" '#rstats',\n",

" '#s…',\n",

" '#esmo22:',\n",

" '#ai',\n",

" '#healthcare!\\n\\ninfographic',\n",

" '@ingliguori\\n\\n#python',\n",

" '#nodejs',\n",

" '#machinelearning',\n",

" '#100daysofmlcode',\n",

" '#html5',\n",

" '#rstats',\n",

" '#s…',\n",

" '#esmo22:',\n",

" '#ai',\n",

" '#healthcare!\\n\\ninfographic',\n",

" '@ingliguori\\n\\n#python',\n",

" '#nodejs',\n",

" '#machinelearning',\n",

" '#100daysofmlcode',\n",

" '#html5',\n",

" '#rstats',\n",

" '#s…',\n",

" '#esmo22:',\n",

" '#strategy',\n",

" '#competitiveintelligence',\n",

" '#marketing',\n",

" '#healthcare',\n",

" '#pharmaceutical',\n",

" '#productmarketing',\n",

" '#biotech',\n",

" '#pharma',\n",

" '#esmo22:',\n",

" '#strategy',\n",

" '#competitiveintelligence',\n",

" '#marketing',\n",

" '#pharma',\n",

" '#biotech',\n",

" '#productmarketing',\n",

" '#healthcare',\n",

" '#pharmaceutical',\n",

" '#esmo22:',\n",

" '#strategy',\n",

" '#competitiveintelligence',\n",

" '#marketing',\n",

" '#healthcare',\n",

" '#pharmaceutical',\n",

" '#productmarketing',\n",

" '#biotech',\n",

" '#pharma',\n",

" '#esmo22:',\n",

" '#strategy',\n",

" '#competitiveintelligence',\n",

" '#marketing',\n",

" '#productmarketing',\n",

" '#biotech',\n",

" '#healthcare',\n",

" '#pharma',\n",

" '#pharmaceutical',\n",

" '#premisehealth',\n",

" '#hiring',\n",

" '#healthcare',\n",

" '#qualitypatientcare',\n",

" '#jobs',\n",

" '#job',\n",

" '#healthcare',\n",

" '#doctors',\n",

" '#mediqueststaffing',\n",

" '#medical',\n",

" '#assistant!',\n",

" '(#suncity)',\n",

" '#job',\n",

" '#jobs',\n",

" '#employment',\n",

" '#healthcare',\n",

" '#job',\n",

" '#hiring',\n",

" '#mediqueststaffing',\n",

" '#jobsearch',\n",

" '#portolavalley,',\n",

" '#healthcare',\n",

" 'open\\n\\n#gvwire',\n",

" '#news',\n",

" '#politics',\n",

" '#healthcare',\n",

" '#economy',\n",

" '#tech',\n",

" '#califor…',\n",

" 'initiatives\\n\\n#healthcare',\n",

" '#mentalhealth',\n",

" '#pandemic',\n",

" '#capitalism',\n",

" 'us.\\n\\n#infrastructure',\n",

" '\\n#climateemergency',\n",

" '\\n#utilities',\n",

" '\\n#foodinsecurity\\n#water\\n#healthcare',\n",

" '#buffalo,',\n",

" '#healthcare',\n",

" '#womeninmedicine',\n",

" '#healthcare',\n",

" '#womeninmedicine',\n",

" '#healthcare',\n",

" '#patientexperience',\n",

" '#tech',\n",

" '#healthcareprofessionals',\n",

" '#5g',\n",

" '#healthcare.',\n",

" '#healthtech',\n",

" '#healthcare',\n",

" '#medtech',\n",

" '#internalmedicine',\n",

" '#tipsfornewdocs',\n",

" '#meded',\n",

" '#medtwitter',\n",

" '#medicine',\n",

" '#medical',\n",

" '#medica…',\n",

" '#hospitalstalktolovedones',\n",

" '#education',\n",

" '#ptsafet…',\n",

" '#brexit',\n",

" '#eu',\n",

" '🌟\\n\\n👉👉https://t.co/de14oseogo\\n\\n#indianbusinesslisting',\n",

" '#localsearchlisting',\n",

" '#localseo',\n",

" '#healthcare',\n",

" '#lifestyle',\n",

" '#theclairvoyantregister',\n",

" '#dating',\n",

" '#theadultentertainment',\n",

" '#healthcare',\n",

" '#chicago,',\n",

" '#artificialintelligence’s',\n",

" '#healthcare\\nby',\n",

" 'https://t.co/rro0ioptmm\\n\\n#ai',\n",

" '#iot',\n",

" '#bigdata',\n",

" '#machinelearning\\n\\ncc:',\n",

" '\\nhttps://t.co/4h1wip7jbs\\n\\n#centurycollege',\n",

" '#nursingassistant',\n",

" '#nursingassistanttraining',\n",

" '#na',\n",

" '#nahha',\n",

" '#hha',\n",

" '#mnnursingassistant',\n",

" '#healthcare',\n",

" '#healthcarecareers',\n",

" '#obra',\n",

" '#online',\n",

" '#hybridlearning',\n",

" '#startups',\n",

" '#leaders',\n",

" '#investors.\\nhttps://t.co/ck3hgpimt4\\n#innovations',\n",

" '#femtech',\n",

" '#womenentrepreneurs',\n",

" '#healthcare',\n",

" '#robotics',\n",

" '#5g',\n",

" '#iot',\n",

" '#healthcare.',\n",

" '#mobileff',\n",

" '#palmcoast,',\n",

" '#healthcare',\n",

" 'information\\n#healthcare',\n",

" '#health',\n",

" '\\n#santshriasharamjibapu',\n",

" '#hiv/#aids',\n",

" '“#christian”',\n",

" '#universal',\n",

" '#healthcare',\n",

" '#healthcare?',\n",

" '#hospitalstalktolovedones',\n",

" '#education',\n",

" '#ptsafet…',\n",

" '#digitalhealth!\\n-',\n",

" '#medtech',\n",

" '#hcldr',\n",

" '#healthcare',\n",

" '(#matthews,',\n",

" '#job',\n",

" '#doctorofveterinarymedicine',\n",

" '#doctorofveterinarymedicine',\n",

" '#preventivecare',\n",

" '#formularymanagement',\n",

" '#medicalrecords',\n",

" '#hospitaloperations',\n",

" '#wellness',\n",

" '#healthcare',\n",

" '#patientcare',\n",

" 'death.\\n\\n#healthfacts',\n",

" '#healthylifestyletips',\n",

" '#healthcare',\n",

" '#toomuchsleep',\n",

" '#sittingandhealth',\n",

" '#sittingtoomuch',\n",

" '#digitaltransformation',\n",

" '#healthcare.\\n\\n#healthtech',\n",

" '#iot',\n",

" '#dataanalytics',\n",

" '#data',\n",

" '#medtech',\n",

" '#tech',\n",

" '#digitaltransformation',\n",

" '#healthcare?',\n",

" '#administration',\n",

" '#heal…',\n",

" '#healthcare',\n",

" '#education.',\n",

" '(#rockhill,',\n",

" '#job',\n",

" '#doctorofveterinarymedicine',\n",

" '#doctorofveterinarymedicine',\n",

" '#preventivecare',\n",

" '#formularymanagement',\n",

" '#medicalrecords',\n",

" '#hospitaloperations',\n",

" '#wellness',\n",

" '#healthcare',\n",

" '#patientcare',\n",

" '#cloud',\n",

" '#healthcare',\n",

" '#cloud',\n",

" '#healthcare',\n",

" '#healthcare',\n",

" '#patientcare',\n",

" '#d2o',\n",

" '#healthcare',\n",

" '#artificialintelligence',\n",

" '#supplychain',\n",

" '#monkeypox',\n",

" '\\n#monkeypoxcases',\n",

" '#monkeypoxvaccine',\n",

" '#deathrate',\n",

" '#vaccine',\n",

" '#healthcare',\n",

" '#monkeypox',\n",

" '\\n#monkeypoxcases',\n",

" '#monkeypoxvaccine',\n",

" '#deathrate',\n",

" '#vaccine',\n",

" '#healthcare',\n",

" '(#charlotte,',\n",

" '#job',\n",

" '#doctorofveterinarymedicine',\n",

" '#doctorofveterinarymedicine',\n",

" '#preventivecare',\n",

" '#formularymanagement',\n",

" '#medicalrecords',\n",

" '#hospitaloperations',\n",

" '#wellness',\n",

" '#healthcare',\n",

" '#patientcare',\n",

" '#healthcare',\n",

" '#hiv/#aids',\n",

" '“#christian”',\n",

" '#universal',\n",

" '#healthcare',\n",

" '#faithregional',\n",

" '#healthcare',\n",

" '#healthcare',\n",

" '#worcesterma',\n",

" '(#winstonsalem,',\n",

" '#job',\n",

" '#doctorofveterinarymedicine',\n",

" '#doctorofveterinarymedicine',\n",

" '#preventivecare',\n",

" '#formularymanagement',\n",

" '#medicalrecords',\n",

" '#hospitaloperations',\n",

" '#wellness',\n",

" '#healthcare',\n",

" '#patientcare',\n",

" '#hospitalstalktolovedones',\n",

" '#education',\n",

" '#ptsafet…',\n",

" '#new:',\n",

" '#coronavirusvaccines',\n",

" 'to:\\nhttps://t.co/wukk9hk1vz\\n\\n#hospital',\n",

" '#healthcare',\n",

" '#catholic',\n",

" '#catholichospitals',\n",

" '#hospitalstalktolovedones',\n",

" '#education',\n",

" '#ptsafet…',\n",

" 'implementation.\\n#healthcare',\n",

" '#trangendercare',\n",

" '#genderaffirmingcare',\n",

" '#…',\n",

" '(#wesleychapel,',\n",

" '#job',\n",

" '#doctorofveterinarymedicine',\n",

" '#doctorofveterinarymedicine',\n",

" '#preventivecare',\n",

" '#formularymanagement',\n",

" '#medicalrecords',\n",

" '#hospitaloperations',\n",

" '#wellness',\n",

" '#healthcare',\n",

" '#patientcare',\n",

" '#healthcare?',\n",

" '#hiv/#aids',\n",

" '“#christian”',\n",

" '#universal',\n",

" '#healthcare',\n",

" '#medicareforall',\n",

" '#voteblue',\n",

" '#votebluetosavedemocracy',\n",

" '#healthcare',\n",

" '#mutualaidrequest?',\n",

" '#healthcare',\n",

" '#healthcare',\n",

" ...]"

]

},

"execution\_count": 40,

"metadata": {},

"output\_type": "execute\_result"

}

],

"source": [

"#Loop through the messages to createa list of values containing hashtags\n",

"for y in [x.split(' ') for x in tweets['tweet\_full\_text'].values]:\n",

" for z in y:\n",

" if '#' in z:\n",

" # Change to lowercase.\n",

" tags.append(z.lower())\n",

" \n",

"#Display all tags \n",

"tags"

]

},

{

"cell\_type": "code",

"execution\_count": 41,

"id": "b77dac02",

"metadata": {},

"outputs": [],

"source": [

"hashtags=pd.Series(tags).value\_counts()"

]

},

{

"cell\_type": "code",

"execution\_count": 94,

"id": "29387bae",

"metadata": {},

"outputs": [

{

"data": {

"text/plain": [

"#healthcare 716\n",

"#health 80\n",

"#medicine 41\n",

"#ai 40\n",

"#job 38\n",

"#medical 35\n",

"#strategy 30\n",

"#pharmaceutical 28\n",

"#digitalhealth 25\n",

"#pharma 25\n",

"#marketing 25\n",

"#medtwitter 24\n",

"#biotech 24\n",

"#competitiveintelligence 24\n",

"#meded 23\n",

"#vaccine 18\n",

"#hiring 18\n",

"#news 17\n",

"#machinelearning 17\n",

"#technology 17\n",

"#coronavirus 16\n",

"#womeninmedicine 16\n",

"#covid 16\n",

"#competitivemarketing 16\n",

"#wellness 15\n",

"#healthtech 15\n",

"#doctorofveterinarymedicine 14\n",

"#science 14\n",

"#medicare 14\n",

"#covid19 14\n",

"dtype: int64"

]

},

"execution\_count": 94,

"metadata": {},

"output\_type": "execute\_result"

}

],

"source": [

"hashtags.iloc[:30]"

]

},

{

"cell\_type": "code",

"execution\_count": 127,

"id": "ca97b515",

"metadata": {},

"outputs": [

{

"data": {

"text/plain": [

"#healthcare 716\n",

"#health 80\n",

"#medicine 41\n",

"#ai 40\n",

"#job 38\n",

"#medical 35\n",

"#strategy 30\n",

"#pharmaceutical 28\n",

"#digitalhealth 25\n",

"#pharma 25\n",

"dtype: int64"

]

},

"execution\_count": 127,

"metadata": {},

"output\_type": "execute\_result"

}

],

"source": [

"hashtags.head(10)"

]

},

{

"cell\_type": "code",

"execution\_count": 128,

"id": "2c6706f8",

"metadata": {},

"outputs": [

{

"data": {

"text/html": [

"<div>\n",

"<style scoped>\n",

" .dataframe tbody tr th:only-of-type {\n",

" vertical-align: middle;\n",

" }\n",

"\n",

" .dataframe tbody tr th {\n",

" vertical-align: top;\n",

" }\n",

"\n",

" .dataframe thead th {\n",

" text-align: right;\n",

" }\n",

"</style>\n",

"<table border=\"1\" class=\"dataframe\">\n",

" <thead>\n",

" <tr style=\"text-align: right;\">\n",

" <th></th>\n",

" <th>0</th>\n",

" </tr>\n",

" </thead>\n",

" <tbody>\n",

" <tr>\n",

" <th>0</th>\n",

" <td>#healthcare</td>\n",

" </tr>\n",

" <tr>\n",

" <th>1</th>\n",

" <td>#premisehealth</td>\n",

" </tr>\n",

" <tr>\n",

" <th>2</th>\n",

" <td>#hiring</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3</th>\n",

" <td>#healthcare</td>\n",

" </tr>\n",

" <tr>\n",

" <th>4</th>\n",

" <td>🚨#new:🚨</td>\n",

" </tr>\n",

" <tr>\n",

" <th>...</th>\n",

" <td>...</td>\n",

" </tr>\n",

" <tr>\n",

" <th>4328</th>\n",

" <td>#mediqueststaffing</td>\n",

" </tr>\n",

" <tr>\n",

" <th>4329</th>\n",

" <td>#hospital</td>\n",

" </tr>\n",

" <tr>\n",

" <th>4330</th>\n",

" <td>#shift</td>\n",

" </tr>\n",

" <tr>\n",

" <th>4331</th>\n",

" <td>(#newportbeach)</td>\n",

" </tr>\n",

" <tr>\n",

" <th>4332</th>\n",

" <td>#job</td>\n",

" </tr>\n",

" </tbody>\n",

"</table>\n",

"<p>4333 rows × 1 columns</p>\n",

"</div>"

],

"text/plain": [

" 0\n",

"0 #healthcare\n",

"1 #premisehealth\n",

"2 #hiring\n",

"3 #healthcare\n",

"4 🚨#new:🚨\n",

"... ...\n",

"4328 #mediqueststaffing\n",

"4329 #hospital\n",

"4330 #shift\n",

"4331 (#newportbeach)\n",

"4332 #job\n",

"\n",

"[4333 rows x 1 columns]"

]

},

"metadata": {},

"output\_type": "display\_data"

}

],

"source": [

"hashtags=pd.DataFrame (tags)\n",

"display (hashtags)"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"id": "c6b3dc0a",

"metadata": {},

"outputs": [],

"source": []

}

],

"metadata": {

"kernelspec": {

"display\_name": "Python 3 (ipykernel)",

"language": "python",

"name": "python3"

},

"language\_info": {

"codemirror\_mode": {

"name": "ipython",

"version": 3

},

"file\_extension": ".py",

"mimetype": "text/x-python",

"name": "python",

"nbconvert\_exporter": "python",

"pygments\_lexer": "ipython3",

"version": "3.9.12"

}

},

"nbformat": 4,

"nbformat\_minor": 5

}