```
# pip install numpy pandas matplotlob
```

```
import pandas as pd
```

importing csv file customer_support_tickets.csv

```
# read the CSV data as pandas dataframe
transaction_df = pd.read_csv('customer_support_tickets.csv')

#looking ak the first 5 rows of the dataframe
transaction_df.head()
```

	Ticket ID	Customer Name	Customer Email	Customer Age	Customer Gender	Product Purchased	Date of Purchase	Ticket Type	Ticket Subject	Ticket Description	Ticke Statu
0	1	Marisa Obrien	carrollallison@example.com	32	Other	GoPro Hero	2021-03-22	Technical issue	Product setup	I'm having an issue with the {product_purchase...	Pendir Custom Respons
1	2	Jessica Rios	clarkeashley@example.com	42	Female	LG Smart TV	2021-05-22	Technical issue	Peripheral compatibility	I'm having an issue with the {product_purchase...	Pendir Custom Respons
2	3	Christopher Robbins	gonzalestracy@example.com	48	Other	Dell XPS	2020-07-14	Technical issue	Network problem	I'm facing a problem with my {product_purchase...	Close
3	4	Christina Dillon	bradleyolson@example.org	27	Female	Microsoft Office	2020-11-13	Billing inquiry	Account access	I'm having an issue with the {product_purchase...	Close
4	5	Alexander Carroll	bradleymark@example.com	67	Female	Autodesk AutoCAD	2020-02-04	Billing inquiry	Data loss	I'm having an issue with the {product_purchase...	Close

Next steps: Generate code with transaction_df View recommended plots New interactive sheet

Double-click (or enter) to edit

```
# printing the shape of the dataframe
transaction_df.shape
```

```
(8469, 17)
```

```
#checking for the null values
transaction_df.isnull().sum()
```

```
                                0
Ticket ID                       0
Customer Name                   0
Customer Email                  0
Customer Age                    0
Customer Gender                 0
Product Purchased               0
Date of Purchase                0
Ticket Type                     0
Ticket Subject                  0
Ticket Description              0
Ticket Status                   0
Resolution                   5700
Ticket Priority                 0
Ticket Channel                  0
First Response Time          2819
Time to Resolution           5700
Customer Satisfaction Rating 5700

dtype: int64
```

*Filling the null vaues *

filling the null values with the help of pandas function #dropna

```
#Dropping the null values
transaction_df.dropna(inplace=True)
```

```
# check for null. values
transaction_df.isnull().sum()
```

	0
Ticket ID	0
Customer Name	0
Customer Email	0
Customer Age	0
Customer Gender	0
Product Purchased	0
Date of Purchase	0
Ticket Type	0
Ticket Subject	0
Ticket Description	0
Ticket Status	0
Resolution	0
Ticket Priority	0
Ticket Channel	0
First Response Time	0
Time to Resolution	0
Customer Satisfaction Rating	0

dtype: int64

```
# printing a concise summary of the data frame
transaction_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 2769 entries, 2 to 8467
Data columns (total 17 columns):
 #   Column                        Non-Null Count  Dtype  
---  ------                        --------------  -----  
 0   Ticket ID                     2769 non-null   int64  
 1   Customer Name                 2769 non-null   object 
 2   Customer Email                2769 non-null   object 
 3   Customer Age                  2769 non-null   int64  
 4   Customer Gender               2769 non-null   object 
 5   Product Purchased             2769 non-null   object 
 6   Date of Purchase              2769 non-null   object 
 7   Ticket Type                   2769 non-null   object 
 8   Ticket Subject                2769 non-null   object 
 9   Ticket Description            2769 non-null   object 
 10  Ticket Status                 2769 non-null   object 
 11  Resolution                    2769 non-null   object 
 12  Ticket Priority               2769 non-null   object 
 13  Ticket Channel                2769 non-null   object 
 14  First Response Time           2769 non-null   object 
 15  Time to Resolution            2769 non-null   object 
 16  Customer Satisfaction Rating  2769 non-null   float64
dtypes: float64(1), int64(2), object(14)
memory usage: 389.4+ KB
```

```
# Convert date/time columns
# Convert date/time columns
transaction_df['Date of Purchase'] = pd.to_datetime(transaction_df['Date of Purchase'])

transaction_df['First Response Time'] = pd.to_timedelta(transaction_df['First Response Time'],errors='coerce')
transaction_df['Time to Resolution'] = pd.to_timedelta(transaction_df['Time to Resolution'],errors='coerce')

# Optional: Convert text columns to category
cat_cols = ['Ticket Type', 'Ticket Priority', 'Ticket Status', 'Ticket Channel', 'Customer Gender']
transaction_df[cat_cols] = transaction_df[cat_cols].astype('category')
# Optional: Convert text columns to category
cat_cols = ['Ticket Type', 'Ticket Priority', 'Ticket Status', 'Ticket Channel', 'Customer Gender']
transaction_df[cat_cols] = transaction_df[cat_cols].astype('category')
```

```
#Exploratory Data Analysis (EDA)  Summary Statistics
print(transaction_df.describe(include='all'))
```

```
min               NaN            NaN          2020-01-01 00:00:00
25%               NaN            NaN          2020-06-29 00:00:00
50%               NaN            NaN          2020-12-26 00:00:00
75%               NaN            NaN          2021-07-01 00:00:00
max               NaN            NaN          2021-12-30 00:00:00
std               NaN            NaN                          NaN

           Ticket Type   Ticket Subject  \
count             2769             2769
unique               5               16
top     Refund request  Network problem
freq               596              201
mean               NaN              NaN
min                NaN              NaN
25%                NaN              NaN
50%                NaN              NaN
75%                NaN              NaN
max                NaN              NaN
std                NaN              NaN

                                      Ticket Description Ticket Status  \
count                                               2769          2769
unique                                              2680             1
top     I'm having an issue with the {product_purchase...        Closed
freq                                                  12          2769
mean                                                 NaN           NaN
min                                                  NaN           NaN
25%                                                  NaN           NaN
50%                                                  NaN           NaN
75%                                                  NaN           NaN
max                                                  NaN           NaN
std                                                  NaN           NaN

                Resolution Ticket Priority Ticket Channel  \
count                 2769            2769           2769
unique                2769               4              4
top     We seat culture plan.        Critical          Email
freq                     1             726            720
mean                   NaN             NaN            NaN
min                    NaN             NaN            NaN
25%                    NaN             NaN            NaN
50%                    NaN             NaN            NaN
75%                    NaN             NaN            NaN
max                    NaN             NaN            NaN
std                    NaN             NaN            NaN

       First Response Time Time to Resolution  Customer Satisfaction Rating
count                    0                  0                   2769.000000
unique                 NaN                NaN                           NaN
top                    NaN                NaN                           NaN
freq                   NaN                NaN                           NaN
mean                   NaT                NaT                      2.991333
min                    NaT                NaT                      1.000000
25%                    NaT                NaT                      2.000000
50%                    NaT                NaT                      3.000000
75%                    NaT                NaT                      4.000000
max                    NaT                NaT                      5.000000
std                    NaT                NaT                      1.407016
```

```
#Ticket Distribution
transaction_df['Ticket Priority'].value_counts()
transaction_df['Ticket Status'].value_counts()
transaction_df['Ticket Channel'].value_counts()
```

	count
Ticket Channel	
Email	720
Phone	691
Social media	684
Chat	674

dtype: int64

```
#Time-Based Analysis
transaction_df['Resolution Days'] = transaction_df['Time to Resolution'].dt.total_seconds() / 86400
transaction_df['Response Hours'] = transaction_df['First Response Time'].dt.total_seconds() / 3600

# Average by priority
transaction_df.groupby('Ticket Priority')[['Response Hours', 'Resolution Days']].mean()
```

```
<ipython-input-12-8c71eb846632>:6: FutureWarning: The default of observed=False is deprecated and will be changed to True in a future ve
  transaction_df.groupby('Ticket Priority')[['Response Hours', 'Resolution Days']].mean()
```

	Response Hours	Resolution Days
Ticket Priority		
Critical	NaN	NaN
High	NaN	NaN
Low	NaN	NaN
Medium	NaN	NaN

**Text Analysis with NLTK*

Goal: Analyze Ticket Description or Ticket Subject

```
import nltk
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from collections import Counter
import string

nltk.download('punkt')
nltk.download('stopwords')
# Download the punkt_tab resource
nltk.download('punkt_tab')
# Basic Preprocessing
```