

Exploring_Pandas_for_Data_Analysis

March 19, 2025

0.1 Exploring Pandas for Data Analysis

```
[ ]: import pandas as pd
```

```
[ ]: employee_df=pd.read_csv("/content/employee_information.csv")
employee_df
```

```
[ ]: First Name Last Name      Salary  Years with Company Postal Code \
0      Mike      Moe      5000.00              3      N94 3M0
1      Noah      Ryan     10000.00              8      N8S 14K
2      Nina      Keller    9072.02             17      S1T 4E6
3      Chanel      Steve   11072.02             12      N7T 3E6
4      Kate      Noor      5000.00             23      K8N 5H6
5      Samer      Mo      100000.00            13      J7H 3HY
6      Heba      Ismail    50000.00              7      K8Y 3M8
7      Laila      Aly      20000.00              5      J8Y 3M0
8      Joseph      Patton   2629.13              2      M6U 5U7
9      Noah      Moran      8626.96             11      K2D 4M9
```

```
                                Email
0                bird@gmail.com
1      nsmall@hotmail.com
2      azikez@gahew.mr
3      chanel@gmail.com
4      kate@hotmail.com
5      samer@gmail.com
6      heba.ismail@hotmail.com
7      Laila.a@hotmail.com
8      daafeja@boh.jm
9      guutodi@bigwoc.kw
```

```
[ ]: house_prices_df=pd.read_html('https://www.livingin-canada.com/
↳house-prices-canada.html')
house_prices_df[0]
```

```
[ ]: City \
0      Vancouver, BC
1      Toronto, Ont
```

```

2           Ottawa, Ont
3           Calgary, Alb
4           Montreal, Que
5           Halifax, NS
6           Regina, Sask
7           Fredericton, NB
8 (adsbygoogle = window.adsbygoogle || []).push(...

```

```

           Average House Price \
0           $1,036,000
1           $870,000
2           $479,000
3           $410,000
4           $435,000
5           $331,000
6           $254,000
7           $198,000
8 (adsbygoogle = window.adsbygoogle || []).push(...

```

```

           12 Month Change
0           + 2.63 %
1           +10.2 %
2           + 15.4 %
3           - 1.5 %
4           + 9.3 %
5           + 3.6 %
6           - 3.9 %
7           - 4.3 %
8 (adsbygoogle = window.adsbygoogle || []).push(...

```

```
[ ]: house_prices_df[1]
```

```

[ ]:           Province \
0           British Columbia
1           Ontario
2           Alberta
3           Quebec
4           Manitoba
5           Saskatchewan
6           Nova Scotia
7           Prince Edward Island
8           Newfoundland / Labrador
9           New Brunswick
10          Canadian Average
11 (adsbygoogle = window.adsbygoogle || []).push(...

```

```

           Average House Price \

```

```

0          $736,000
1          $594,000
2          $353,000
3          $340,000
4          $295,000
5          $271,000
6          $266,000
7          $243,000
8          $236,000
9          $183,000
10         $488,000
11  (adsbygoogle = window.adsbygoogle || []).push(...

```

```

                                12 Month Change
0                               + 7.6 %
1                               - 3.2 %
2                               - 7.5 %
3                               + 7.6 %
4                               - 1.4 %
5                               - 3.8 %
6                               + 3.5 %
7                               + 3.0 %
8                               - 1.6 %
9                               - 2.2 %
10                              - 1.3 %
11  (adsbygoogle = window.adsbygoogle || []).push(...

```

```
[ ]: employee_df.head()
```

```
[ ]:
First Name Last Name  Salary  Years with Company  Postal Code  \
0      Mike      Moe   5000.00             3      N94 3M0
1      Noah      Ryan  10000.00            8      N8S 14K
2      Nina      Keller  9072.02           17      S1T 4E6
3      Chanel      Steve  11072.02          12      N7T 3E6
4      Kate      Noor   5000.00           23      K8N 5H6

```

```

                                Email
0      bird@gmail.com
1  nsmall@hotmail.com
2      azikez@gahew.mr
3      chanel@gmail.com
4      kate@hotmail.com

```

```
[ ]: employee_df.columns
```

```
[ ]: Index(['First Name', 'Last Name', 'Salary', 'Years with Company',
           'Postal Code', 'Email'],

```

```
dtype='object')
```

```
[ ]: email_series=employee_df['Email']  
email_series
```

```
[ ]: 0      bird@gmail.com  
1      nsmall@hotmail.com  
2      azikez@gahew.mr  
3      chanel@gmail.com  
4      kate@hotmail.com  
5      samer@gmail.com  
6      heba.ismail@hotmail.com  
7      Laila.a@hotmail.com  
8      daafeja@boh.jm  
9      guutodi@bigwoc.kw  
Name: Email, dtype: object
```

```
[ ]: employee_df['Years with Company']
```

```
[ ]: 0      3  
1      8  
2     17  
3     12  
4     23  
5     13  
6      7  
7      5  
8      2  
9     11  
Name: Years with Company, dtype: int64
```

```
[ ]: Name_Salary_df=employee_df[['First Name','Salary']]  
Name_Salary_df
```

```
[ ]:   First Name    Salary  
0      Mike    5000.00  
1      Noah   10000.00  
2      Nina    9072.02  
3    Chanel   11072.02  
4      Kate    5000.00  
5      Samer  100000.00  
6      Heba   50000.00  
7      Laila   20000.00  
8    Joseph    2629.13  
9      Noah    8626.96
```

```
[ ]: employee_df['Age']=[25,30,34,35,56,44,32,35,28,29]
employee_df
```

```
[ ]: First Name Last Name      Salary  Years with Company Postal Code \
0      Mike      Moe      5000.00              3      N94 3M0
1      Noah      Ryan     10000.00              8      N8S 14K
2      Nina      Keller    9072.02             17      S1T 4E6
3      Chanel     Steve    11072.02             12      N7T 3E6
4      Kate      Noor      5000.00             23      K8N 5H6
5      Samer      Mo      100000.00            13      J7H 3HY
6      Heba      Ismail    50000.00              7      K8Y 3M8
7      Laila      Aly      20000.00              5      J8Y 3M0
8      Joseph     Patton    2629.13              2      M6U 5U7
9      Noah      Moran     8626.96             11      K2D 4M9
```

```

Email  Age
0      bird@gmail.com  25
1      nsmall@hotmail.com  30
2      azikez@gahew.mr  34
3      chanel@gmail.com  35
4      kate@hotmail.com  56
5      samer@gmail.com  44
6      heba.ismail@hotmail.com  32
7      Laila.a@hotmail.com  35
8      daafeja@boh.jm  28
9      guutodi@bigwoc.kw  29
```

```
[ ]: employee_df.loc[employee_df['First Name'] == 'Noah']
```

```
[ ]: First Name Last Name      Salary  Years with Company Postal Code \
1      Noah      Ryan     10000.00              8      N8S 14K
9      Noah      Moran     8626.96             11      K2D 4M9
```

```

Email  Age
1      nsmall@hotmail.com  30
9      guutodi@bigwoc.kw  29
```

```
[ ]: employee_df.iloc[0]
```

```
[ ]: First Name      Mike
Last Name      Moe
Salary      5000.0
Years with Company      3
Postal Code      N94 3M0
Email      bird@gmail.com
Age      25
Name: 0, dtype: object
```

```
[ ]: employee_df.iloc[2:5]
```

```
[ ]: First Name Last Name Salary Years with Company Postal Code \
2 Nina Keller 9072.02 17 S1T 4E6
3 Chanel Steve 11072.02 12 N7T 3E6
4 Kate Noor 5000.00 23 K8N 5H6

Email Age
2 azikez@gahew.mr 34
3 chanel@gmail.com 35
4 kate@hotmail.com 56
```

```
[ ]: employee_df.iloc[1,0:4]
```

```
[ ]: First Name Noah
Last Name Ryan
Salary 10000.0
Years with Company 8
Name: 1, dtype: object
```

```
[ ]: employee_df.sample(n=3,axis=0)
```

```
[ ]: First Name Last Name Salary Years with Company Postal Code \
5 Samer Mo 100000.00 13 J7H 3HY
1 Noah Ryan 10000.00 8 N8S 14K
8 Joseph Patton 2629.13 2 M6U 5U7

Email Age
5 samer@gmail.com 44
1 nsmall@hotmail.com 30
8 daafeja@boh.jm 28
```

```
[ ]: employee_df['Salary'].describe()
```

```
[ ]: count 10.000000
mean 22140.013000
std 30613.044026
min 2629.130000
25% 5906.740000
50% 9536.010000
75% 17768.005000
max 100000.000000
Name: Salary, dtype: float64
```

```
[ ]: employee_df['Salary'] = employee_df['Salary'] +1000
```

```
[ ]: employee_df['Salary'].describe()
```

```
[ ]: count      10.000000
     mean      23140.013000
     std       30613.044026
     min       3629.130000
     25%       6906.740000
     50%      10536.010000
     75%      18768.005000
     max      101000.000000
     Name: Salary, dtype: float64
```

```
[ ]: employee_df.head()
```

```
[ ]:   First Name Last Name   Salary  Years with Company Postal Code \
0      Mike      Moe    6000.00           3      N94 3M0
1      Noah      Ryan   11000.00           8      N8S 14K
2      Nina      Keller  10072.02          17      S1T 4E6
3    Chanel      Steve  12072.02          12      N7T 3E6
4      Kate      Noor    6000.00          23      K8N 5H6

      Email  Age
0    bird@gmail.com  25
1  nsmall@hotmail.com  30
2    azikez@gahew.mr  34
3   chanel@gmail.com  35
4   kate@hotmail.com  56
```

```
[ ]: employee_df.sort_values(by='Years with Company',ascending=False, inplace = True)
employee_df
```

```
[ ]:   First Name Last Name   Salary  Years with Company Postal Code \
4      Kate      Noor    6000.00          23      K8N 5H6
2      Nina      Keller  10072.02          17      S1T 4E6
5      Samer        Mo  101000.00          13      J7H 3HY
3    Chanel      Steve  12072.02          12      N7T 3E6
9      Noah      Moran   9626.96          11      K2D 4M9
1      Noah      Ryan   11000.00           8      N8S 14K
6      Heba      Ismail  51000.00           7      K8Y 3M8
7      Laila      Aly   21000.00           5      J8Y 3M0
0      Mike      Moe    6000.00           3      N94 3M0
8    Joseph      Patton   3629.13           2      M6U 5U7

      Email  Age
4   kate@hotmail.com  56
2   azikez@gahew.mr  34
5   samer@gmail.com  44
3   chanel@gmail.com  35
9  guutodi@bigwoc.kw  29
```

```

1      nsmall@hotmail.com    30
6  heba.ismail@hotmail.com   32
7      Laila.a@hotmail.com   35
0      bird@gmail.com        25
8      daafeja@boh.jm        28

```

```

[ ]: #reset index
employee_df.reset_index(drop=True, inplace=True)
employee_df

```

```

[ ]: First Name Last Name      Salary  Years with Company Postal Code \
0      Kate      Noor      6000.00           23      K8N 5H6
1      Nina      Keller    10072.02           17      S1T 4E6
2      Samer      Mo      101000.00          13      J7H 3HY
3      Chanel      Steve    12072.02           12      N7T 3E6
4      Noah      Moran      9626.96           11      K2D 4M9
5      Noah      Ryan      11000.00            8      N8S 14K
6      Heba      Ismail     51000.00            7      K8Y 3M8
7      Laila      Aly      21000.00            5      J8Y 3M0
8      Mike      Moe      6000.00            3      N94 3M0
9      Joseph     Patton    3629.13            2      M6U 5U7

```

```

      Email  Age
0      kate@hotmail.com    56
1      azikez@gahew.mr     34
2      samer@gmail.com     44
3      chanel@gmail.com    35
4      guutodi@bigwoc.kw   29
5      nsmall@hotmail.com   30
6  heba.ismail@hotmail.com   32
7      Laila.a@hotmail.com   35
8      bird@gmail.com       25
9      daafeja@boh.jm       28

```

```

[ ]: #Define a function to increase salary by 10% for all employees
def salary_update(salary):
    return salary*1.1

```

```

[ ]: employee_df['Salary']=employee_df['Salary'].apply(salary_update)
employee_df

```

```

[ ]: First Name Last Name      Salary  Years with Company Postal Code \
0      Kate      Noor      6600.000           23      K8N 5H6
1      Nina      Keller    11079.222           17      S1T 4E6
2      Samer      Mo      111100.000          13      J7H 3HY
3      Chanel      Steve    13279.222           12      N7T 3E6
4      Noah      Moran      10589.656           11      K2D 4M9

```


5	Noah	Ryan	12100.000	8	N8S 14K
6	Heba	Ismail	56100.000	7	K8Y 3M8
7	Laila	Aly	23100.000	5	J8Y 3M0
8	Mike	Moe	6600.000	3	N94 3M0
9	Joseph	Patton	3992.043	2	M6U 5U7

	Email	Age
0	kate@hotmail.com	56
1	azikez@gahew.mr	34
2	samer@gmail.com	44
3	chanel@gmail.com	35
4	guutodi@bigwoc.kw	29
5	nsmall@hotmail.com	30
6	heba.ismail@hotmail.com	32
7	Laila.a@hotmail.com	35
8	bird@gmail.com	25
9	daafeja@boh.jm	28

```
[ ]: df_loyalty=employee_df[employee_df['Years with Company']>10]
df_loyalty
```

```
[ ]: First Name Last Name      Salary  Years with Company Postal Code \
0      Kate      Noor      6600.000          23      K8N 5H6
1      Nina      Keller    11079.222          17      S1T 4E6
2      Samer      Mo      111100.000          13      J7H 3HY
3      Chanel      Steve    13279.222          12      N7T 3E6
4      Noah      Moran    10589.656          11      K2D 4M9
```

	Email	Age
0	kate@hotmail.com	56
1	azikez@gahew.mr	34
2	samer@gmail.com	44
3	chanel@gmail.com	35
4	guutodi@bigwoc.kw	29

```
[ ]: mask_1=employee_df['Years with Company']>10
mask_2=employee_df['Age']>30
df_loyalty=employee_df[mask_1 & mask_2]
df_loyalty
```

```
[ ]: First Name Last Name      Salary  Years with Company Postal Code \
0      Kate      Noor      6600.000          23      K8N 5H6
1      Nina      Keller    11079.222          17      S1T 4E6
2      Samer      Mo      111100.000          13      J7H 3HY
3      Chanel      Steve    13279.222          12      N7T 3E6
```

	Email	Age
0	kate@hotmail.com	56
1	azikez@gahew.mr	34
2	samer@gmail.com	44
3	chanel@gmail.com	35
4	guutodi@bigwoc.kw	29

```

0 kate@hotmail.com 56
1 azikez@gahew.mr 34
2 samer@gmail.com 44
3 chanel@gmail.com 35

```

```
[ ]: df_high_salary=employee_df[employee_df['Salary']>60000]
df_high_salary
```

```
[ ]: First Name Last Name Salary Years with Company Postal Code \
2 Samer Mo 111100.0 13 J7H 3HY

Email Age
2 samer@gmail.com 44
```

```
[ ]: employee_df
```

```
[ ]: First Name Last Name Salary Years with Company Postal Code \
0 Kate Noor 6600.000 23 K8N 5H6
1 Nina Keller 11079.222 17 S1T 4E6
2 Samer Mo 111100.000 13 J7H 3HY
3 Chanel Steve 13279.222 12 N7T 3E6
4 Noah Moran 10589.656 11 K2D 4M9
5 Noah Ryan 12100.000 8 N8S 14K
6 Heba Ismail 56100.000 7 K8Y 3M8
7 Laila Aly 23100.000 5 J8Y 3M0
8 Mike Moe 6600.000 3 N94 3M0
9 Joseph Patton 3992.043 2 M6U 5U7

Email Age
0 kate@hotmail.com 56
1 azikez@gahew.mr 34
2 samer@gmail.com 44
3 chanel@gmail.com 35
4 guutodi@bigwoc.kw 29
5 nsmall@hotmail.com 30
6 heba.ismail@hotmail.com 32
7 Laila.a@hotmail.com 35
8 bird@gmail.com 25
9 daafeja@boh.jm 28
```

```
[ ]: df_filtered = employee_df[employee_df['First Name']=='Mike']
df_filtered
```

```
[ ]: First Name Last Name Salary Years with Company Postal Code \
8 Mike Moe 6600.0 3 N94 3M0

Email Age
```

```
8 bird@gmail.com 25
```

```
[ ]: mask = employee_df['Last Name'].isin(['Moe', 'Ryan'])
df_filtered = employee_df[mask]
df_filtered
```

```
[ ]: First Name Last Name Salary Years with Company Postal Code \
5 Noah Ryan 12100.0 8 N8S 14K
8 Mike Moe 6600.0 3 N94 3M0

Email Age
5 nsmall@hotmail.com 30
8 bird@gmail.com 25
```

```
[ ]: employee_df[employee_df['Salary'].between(5000,9000)]
```

```
[ ]: First Name Last Name Salary Years with Company Postal Code \
0 Kate Noor 6600.0 23 K8N 5H6
8 Mike Moe 6600.0 3 N94 3M0

Email Age
0 kate@hotmail.com 56
8 bird@gmail.com 25
```

```
[ ]: employee_df['First Name'].duplicated(keep=False)
```

```
[ ]: 0 False
1 False
2 False
3 False
4 True
5 True
6 False
7 False
8 False
9 False
Name: First Name, dtype: bool
```

```
[ ]: #reversing
mask = ~employee_df['First Name'].duplicated(keep=False)
mask
```

```
[ ]: 0 True
1 True
2 True
3 True
4 False
```

```

5    False
6     True
7     True
8     True
9     True
Name: First Name, dtype: bool

```

```
[ ]: employee_df[mask]
```

```
[ ]:
First Name Last Name      Salary  Years with Company Postal Code \
0      Kate      Noor    6600.000           23      K8N 5H6
1      Nina      Keller  11079.222           17      S1T 4E6
2      Samer       Mo   111100.000           13      J7H 3HY
3      Chanel     Steve  13279.222           12      N7T 3E6
6      Heba      Ismail  56100.000            7      K8Y 3M8
7      Laila      Aly   23100.000            5      J8Y 3M0
8      Mike      Moe    6600.000            3      N94 3M0
9      Joseph    Patton  3992.043            2      M6U 5U7

      Email  Age
0    kate@hotmail.com  56
1    azikez@gahew.mr  34
2    samer@gmail.com  44
3    chanel@gmail.com  35
6  heba.ismail@hotmail.com  32
7    Laila.a@hotmail.com  35
8    bird@gmail.com  25
9    daafeja@boh.jm  28

```

```
[ ]: sales_df=pd.read_csv('/content/ecommerce_sales.csv', encoding='unicode_escape')
sales_df
```

```
[ ]:
InvoiceNo StockCode      Description  Quantity \
0      536365      85123A  WHITE HANGING HEART T-LIGHT HOLDER      6
1      536365      71053           WHITE METAL LANTERN      6
2      536365      84406B    CREAM CUPID HEARTS COAT HANGER      8
3      536365      84029G  KNITTED UNION FLAG HOT WATER BOTTLE      6
4      536365      84029E    RED WOOLLY HOTTIE WHITE HEART.      6
...      ...      ...      ...      ...
541904    581587      22613    PACK OF 20 SPACEBOY NAPKINS      12
541905    581587      22899    CHILDREN'S APRON DOLLY GIRL      6
541906    581587      23254    CHILDRENS CUTLERY DOLLY GIRL      4
541907    581587      23255    CHILDRENS CUTLERY CIRCUS PARADE      4
541908    581587      22138    BAKING SET 9 PIECE RETROSPOT      3

      InvoiceDate  UnitPrice  CustomerID      Country
0      12/1/2010 8:26      2.55      17850.0  United Kingdom

```

1	12/1/2010 8:26	3.39	17850.0	United Kingdom
2	12/1/2010 8:26	2.75	17850.0	United Kingdom
3	12/1/2010 8:26	3.39	17850.0	United Kingdom
4	12/1/2010 8:26	3.39	17850.0	United Kingdom
...
541904	12/9/2011 12:50	0.85	12680.0	France
541905	12/9/2011 12:50	2.10	12680.0	France
541906	12/9/2011 12:50	4.15	12680.0	France
541907	12/9/2011 12:50	4.15	12680.0	France
541908	12/9/2011 12:50	4.95	12680.0	France

[541909 rows x 8 columns]

```
[ ]: sales_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 541909 entries, 0 to 541908
Data columns (total 8 columns):
#   Column          Non-Null Count  Dtype
---  -
0   InvoiceNo        541909 non-null object
1   StockCode        541909 non-null object
2   Description      540455 non-null object
3   Quantity         541909 non-null int64
4   InvoiceDate      541909 non-null object
5   UnitPrice        541909 non-null float64
6   CustomerID       406829 non-null float64
7   Country          541909 non-null object
dtypes: float64(2), int64(1), object(5)
memory usage: 33.1+ MB
```

```
[ ]: sales_df['InvoiceDate']=pd.to_datetime(sales_df['InvoiceDate'])
original_sales=sales_df.copy()
sales_df.head()
```

```
[ ]: InvoiceNo StockCode Description Quantity \
0   536365   85123A  WHITE HANGING HEART T-LIGHT HOLDER      6
1   536365   71053    WHITE METAL LANTERN                    6
2   536365   84406B    CREAM CUPID HEARTS COAT HANGER        8
3   536365   84029G  KNITTED UNION FLAG HOT WATER BOTTLE      6
4   536365   84029E    RED WOOLLY HOTTIE WHITE HEART.        6

InvoiceDate UnitPrice CustomerID Country
0 2010-12-01 08:26:00      2.55   17850.0 United Kingdom
1 2010-12-01 08:26:00      3.39   17850.0 United Kingdom
2 2010-12-01 08:26:00      2.75   17850.0 United Kingdom
3 2010-12-01 08:26:00      3.39   17850.0 United Kingdom
```

4 2010-12-01 08:26:00 3.39 17850.0 United Kingdom

```
[ ]: sales_df.isnull().sum()
```

```
[ ]: InvoiceNo      0
      StockCode    0
      Description  1454
      Quantity     0
      InvoiceDate   0
      UnitPrice    0
      CustomerID   135080
      Country      0
      dtype: int64
```

```
[ ]: sales_df.nunique()
```

```
[ ]: InvoiceNo      25900
      StockCode     4070
      Description   4223
      Quantity      722
      InvoiceDate   23260
      UnitPrice     1630
      CustomerID    4372
      Country       38
      dtype: int64
```

```
[ ]: sales_df['Country'].unique()
```

```
[ ]: array(['United Kingdom', 'France', 'Australia', 'Netherlands', 'Germany',
          'Norway', 'EIRE', 'Switzerland', 'Spain', 'Poland', 'Portugal',
          'Italy', 'Belgium', 'Lithuania', 'Japan', 'Iceland',
          'Channel Islands', 'Denmark', 'Cyprus', 'Sweden', 'Austria',
          'Israel', 'Finland', 'Bahrain', 'Greece', 'Hong Kong', 'Singapore',
          'Lebanon', 'United Arab Emirates', 'Saudi Arabia',
          'Czech Republic', 'Canada', 'Unspecified', 'Brazil', 'USA',
          'European Community', 'Malta', 'RSA'], dtype=object)
```

```
[ ]: sales_df.groupby('Country')['Quantity'].sum().sort_values(ascending=False)
```

```
[ ]: Country
      United Kingdom      4263829
      Netherlands         200128
      EIRE                 142637
      Germany              117448
      France               110480
      Australia            83653
      Sweden               35637
```

Switzerland	30325
Spain	26824
Japan	25218
Belgium	23152
Norway	19247
Portugal	16180
Finland	10666
Channel Islands	9479
Denmark	8188
Italy	7999
Cyprus	6317
Singapore	5234
Austria	4827
Hong Kong	4769
Israel	4353
Poland	3653
Unspecified	3300
Canada	2763
Iceland	2458
Greece	1556
USA	1034
United Arab Emirates	982
Malta	944
Lithuania	652
Czech Republic	592
European Community	497
Lebanon	386
Brazil	356
RSA	352
Bahrain	260
Saudi Arabia	75

Name: Quantity, dtype: int64

```
[ ]: mean_sales = sales_df.groupby('Country')['UnitPrice'].mean()
mean_sales
```

```
[ ]: Country
Australia      3.220612
Austria        4.243192
Bahrain        4.556316
Belgium        3.644335
Brazil         4.456250
Canada         6.030331
Channel Islands 4.932124
Cyprus         6.302363
Czech Republic 2.938333
Denmark        3.256941
```

EIRE	5.911077
European Community	4.820492
Finland	5.448705
France	5.028864
Germany	3.966930
Greece	4.885548
Hong Kong	42.505208
Iceland	2.644011
Israel	3.633131
Italy	4.831121
Japan	2.276145
Lebanon	5.387556
Lithuania	2.841143
Malta	5.244173
Netherlands	2.738317
Norway	6.012026
Poland	4.170880
Portugal	8.582976
RSA	4.277586
Saudi Arabia	2.411000
Singapore	109.645808
Spain	4.987544
Sweden	3.910887
Switzerland	3.403442
USA	2.216426
United Arab Emirates	3.380735
United Kingdom	4.532422
Unspecified	2.699574

Name: UnitPrice, dtype: float64

```
[ ]: min_sales = sales_df.groupby('Country')['UnitPrice'].min().
      ↪sort_values(ascending=False)
min_sales
```

```
[ ]: Country
Bahrain      1.25
Lithuania    1.25
Brazil       0.85
Lebanon      0.55
European Community 0.55
USA          0.42
Saudi Arabia 0.42
United Arab Emirates 0.29
Czech Republic 0.29
Iceland      0.25
Japan        0.21
Hong Kong    0.21
```


Denmark	0.21
Malta	0.19
Unspecified	0.19
Poland	0.19
Sweden	0.19
Channel Islands	0.19
Singapore	0.19
Greece	0.14
Portugal	0.12
Italy	0.12
Austria	0.12
Finland	0.12
Cyprus	0.12
Belgium	0.12
Canada	0.10
Israel	0.06
Netherlands	0.00
Norway	0.00
Germany	0.00
France	0.00
RSA	0.00
EIRE	0.00
Spain	0.00
Switzerland	0.00
Australia	0.00
United Kingdom	-11062.06

Name: UnitPrice, dtype: float64

```
[ ]: sales_df.groupby(['Country', 'InvoiceDate'])['UnitPrice'].mean().
      ↪sort_values(ascending=False)
```

```
[ ]: Country      InvoiceDate
United Kingdom 2011-06-10 15:31:00    38970.00
               2011-12-05 11:36:00    17836.46
               2011-01-05 09:55:00    16888.02
               2011-01-05 09:57:00    16453.71
               2010-12-07 15:49:00    13541.33
               ...
               2011-07-28 14:58:00         0.00
               2011-03-30 17:24:00         0.00
               2011-03-29 09:36:00         0.00
               2011-08-12 14:52:00   -11062.06
               2011-08-12 14:51:00   -11062.06
Name: UnitPrice, Length: 23616, dtype: float64
```

```
[ ]: #reset index
sales_df.reset_index(drop=True, inplace=True)
```

```
sales_df
```

```
[ ]: InvoiceNo StockCode Description Quantity \
0      536365      85123A WHITE HANGING HEART T-LIGHT HOLDER      6
1      536365      71053          WHITE METAL LANTERN      6
2      536365      84406B          CREAM CUPID HEARTS COAT HANGER      8
3      536365      84029G KNITTED UNION FLAG HOT WATER BOTTLE      6
4      536365      84029E          RED WOOLLY HOTTIE WHITE HEART.      6
...      ...      ...      ...      ...
541904      581587      22613          PACK OF 20 SPACEBOY NAPKINS      12
541905      581587      22899          CHILDREN'S APRON DOLLY GIRL      6
541906      581587      23254          CHILDRENS CUTLERY DOLLY GIRL      4
541907      581587      23255          CHILDRENS CUTLERY CIRCUS PARADE      4
541908      581587      22138          BAKING SET 9 PIECE RETROSPOT      3
```

```
InvoiceDate UnitPrice CustomerID Country
0      2010-12-01 08:26:00      2.55      17850.0 United Kingdom
1      2010-12-01 08:26:00      3.39      17850.0 United Kingdom
2      2010-12-01 08:26:00      2.75      17850.0 United Kingdom
3      2010-12-01 08:26:00      3.39      17850.0 United Kingdom
4      2010-12-01 08:26:00      3.39      17850.0 United Kingdom
...      ...      ...      ...
541904 2011-12-09 12:50:00      0.85      12680.0 France
541905 2011-12-09 12:50:00      2.10      12680.0 France
541906 2011-12-09 12:50:00      4.15      12680.0 France
541907 2011-12-09 12:50:00      4.15      12680.0 France
541908 2011-12-09 12:50:00      4.95      12680.0 France
```

```
[541909 rows x 8 columns]
```

```
[ ]: sales_df.set_index(keys=['Country', 'InvoiceDate'], inplace=True)
sales_df
```

```
[ ]: InvoiceNo StockCode \
Country InvoiceDate
United Kingdom 2010-12-01 08:26:00      536365      85123A
                2010-12-01 08:26:00      536365      71053
                2010-12-01 08:26:00      536365      84406B
                2010-12-01 08:26:00      536365      84029G
                2010-12-01 08:26:00      536365      84029E
...      ...
France      2011-12-09 12:50:00      581587      22613
                2011-12-09 12:50:00      581587      22899
                2011-12-09 12:50:00      581587      23254
                2011-12-09 12:50:00      581587      23255
                2011-12-09 12:50:00      581587      22138
```

Country	InvoiceDate	Description \
United Kingdom	2010-12-01 08:26:00	WHITE HANGING HEART T-LIGHT HOLDER
	2010-12-01 08:26:00	WHITE METAL LANTERN
	2010-12-01 08:26:00	CREAM CUPID HEARTS COAT HANGER
	2010-12-01 08:26:00	KNITTED UNION FLAG HOT WATER BOTTLE
	2010-12-01 08:26:00	RED WOOLLY HOTTIE WHITE HEART.
...		...
France	2011-12-09 12:50:00	PACK OF 20 SPACEBOY NAPKINS
	2011-12-09 12:50:00	CHILDREN'S APRON DOLLY GIRL
	2011-12-09 12:50:00	CHILDRENS CUTLERY DOLLY GIRL
	2011-12-09 12:50:00	CHILDRENS CUTLERY CIRCUS PARADE
	2011-12-09 12:50:00	BAKING SET 9 PIECE RETROSPOT

Country	InvoiceDate	Quantity	UnitPrice	CustomerID
United Kingdom	2010-12-01 08:26:00	6	2.55	17850.0
	2010-12-01 08:26:00	6	3.39	17850.0
	2010-12-01 08:26:00	8	2.75	17850.0
	2010-12-01 08:26:00	6	3.39	17850.0
	2010-12-01 08:26:00	6	3.39	17850.0
...	
France	2011-12-09 12:50:00	12	0.85	12680.0
	2011-12-09 12:50:00	6	2.10	12680.0
	2011-12-09 12:50:00	4	4.15	12680.0
	2011-12-09 12:50:00	4	4.15	12680.0
	2011-12-09 12:50:00	3	4.95	12680.0

[541909 rows x 6 columns]

```
[ ]: sales_df.sort_index(ascending=True,inplace=True)
sales_df.head()
```

```
[ ]:
Country InvoiceDate InvoiceNo StockCode \
Australia 2010-12-01 10:03:00 536389 22941
          2010-12-01 10:03:00 536389 21622
          2010-12-01 10:03:00 536389 21791
          2010-12-01 10:03:00 536389 35004C
          2010-12-01 10:03:00 536389 35004G

Country InvoiceDate Description Quantity \
Australia 2010-12-01 10:03:00 CHRISTMAS LIGHTS 10 REINDEER 6
          2010-12-01 10:03:00 VINTAGE UNION JACK CUSHION COVER 8
          2010-12-01 10:03:00 VINTAGE HEADS AND TAILS CARD GAME 12
          2010-12-01 10:03:00 SET OF 3 COLOURED FLYING DUCKS 6
```

		UnitPrice	CustomerID
Country	InvoiceDate		
Australia	2010-12-01 10:03:00	8.50	12431.0
	2010-12-01 10:03:00	4.95	12431.0
	2010-12-01 10:03:00	1.25	12431.0
	2010-12-01 10:03:00	5.45	12431.0
	2010-12-01 10:03:00	6.35	12431.0

```
[ ]: sales_df.index.names
```

```
[ ]: FrozenList(['Country', 'InvoiceDate'])
```

```
[ ]: sales_df.index[0]
```

```
[ ]: ('Australia', Timestamp('2010-12-01 10:03:00'))
```

```
[ ]: sales_df.index.get_level_values(0)
```

```
[ ]: Index(['Australia', 'Australia', 'Australia', 'Australia', 'Australia',
          'Australia', 'Australia', 'Australia', 'Australia', 'Australia',
          ...
          'Unspecified', 'Unspecified', 'Unspecified', 'Unspecified',
          'Unspecified', 'Unspecified', 'Unspecified', 'Unspecified',
          'Unspecified', 'Unspecified'],
          dtype='object', name='Country', length=541909)
```

```
[ ]: sales_df.index.get_level_values('Country')
```

```
[ ]: Index(['Australia', 'Australia', 'Australia', 'Australia', 'Australia',
          'Australia', 'Australia', 'Australia', 'Australia', 'Australia',
          ...
          'Unspecified', 'Unspecified', 'Unspecified', 'Unspecified',
          'Unspecified', 'Unspecified', 'Unspecified', 'Unspecified',
          'Unspecified', 'Unspecified'],
          dtype='object', name='Country', length=541909)
```

```
[ ]: sales_df.index.get_level_values(1)
```

```
[ ]: DatetimeIndex(['2010-12-01 10:03:00', '2010-12-01 10:03:00',
                   '2010-12-01 10:03:00', '2010-12-01 10:03:00',
                   '2010-12-01 10:03:00', '2010-12-01 10:03:00',
                   '2010-12-01 10:03:00', '2010-12-01 10:03:00',
                   ...
                   '2011-11-24 14:55:00', '2011-11-24 14:55:00',
```

```

        '2011-11-24 14:55:00', '2011-11-24 14:55:00',
        '2011-11-24 14:55:00', '2011-11-24 14:55:00',
        '2011-11-24 14:55:00', '2011-11-24 14:55:00',
        '2011-11-24 14:55:00', '2011-11-24 14:55:00'],
        dtype='datetime64[ns]', name='InvoiceDate', length=541909,
        freq=None)

```

```
[ ]: sales_df.index.get_level_values('InvoiceDate')
```

```
[ ]: DatetimeIndex(['2010-12-01 10:03:00', '2010-12-01 10:03:00',
        '2010-12-01 10:03:00', '2010-12-01 10:03:00',
        '2010-12-01 10:03:00', '2010-12-01 10:03:00',
        '2010-12-01 10:03:00', '2010-12-01 10:03:00',
        ...,
        '2011-11-24 14:55:00', '2011-11-24 14:55:00',
        '2011-11-24 14:55:00', '2011-11-24 14:55:00',
        '2011-11-24 14:55:00', '2011-11-24 14:55:00',
        '2011-11-24 14:55:00', '2011-11-24 14:55:00'],
        dtype='datetime64[ns]', name='InvoiceDate', length=541909,
        freq=None)

```

```
[ ]: sales_df.index.set_names(names=['Transaction Location', 'Transaction Date'],
        inplace = True)
sales_df
```

```
[ ]:
Transaction Location Transaction Date InvoiceNo StockCode \
Australia          2010-12-01 10:03:00    536389    22941
                  2010-12-01 10:03:00    536389    21622
                  2010-12-01 10:03:00    536389    21791
                  2010-12-01 10:03:00    536389    35004C
                  2010-12-01 10:03:00    536389    35004G
...
Unspecified        2011-11-24 14:55:00    578539    22560
                  2011-11-24 14:55:00    578539    23570
                  2011-11-24 14:55:00    578539    23571
                  2011-11-24 14:55:00    578539    84992
                  2011-11-24 14:55:00    578539    23309

```

```

Transaction Location Transaction Date Description \
Australia          2010-12-01 10:03:00    CHRISTMAS LIGHTS 10 REINDEER
                  2010-12-01 10:03:00    VINTAGE UNION JACK CUSHION COVER
                  2010-12-01 10:03:00    VINTAGE HEADS AND TAILS CARD GAME
                  2010-12-01 10:03:00    SET OF 3 COLOURED FLYING DUCKS

```

```

                2010-12-01 10:03:00                SET OF 3 GOLD FLYING DUCKS
...
Unspecified    2011-11-24 14:55:00                TRADITIONAL MODELLING CLAY
                2011-11-24 14:55:00                TRADITIONAL PICK UP STICKS GAME
                2011-11-24 14:55:00                TRADITIONAL NAUGHTS & CROSSES
                2011-11-24 14:55:00                72 SWEETHEART FAIRY CAKE CASES
                2011-11-24 14:55:00                SET OF 60 I LOVE LONDON CAKE CASES

                Quantity  UnitPrice  CustomerID
Transaction Location Transaction Date
Australia        2010-12-01 10:03:00           6         8.50       12431.0
                2010-12-01 10:03:00           8         4.95       12431.0
                2010-12-01 10:03:00          12         1.25       12431.0
                2010-12-01 10:03:00           6         5.45       12431.0
                2010-12-01 10:03:00           4         6.35       12431.0
...
Unspecified    2011-11-24 14:55:00           24         1.25         NaN
                2011-11-24 14:55:00           12         1.25         NaN
                2011-11-24 14:55:00           12         1.65         NaN
                2011-11-24 14:55:00           24         0.55         NaN
                2011-11-24 14:55:00           24         0.55         NaN

```

[541909 rows x 6 columns]

```
[ ]: sales_df.loc['Australia','1/11/2011 9:47']
```

```

[ ]:
InvoiceNo StockCode Description \
Transaction Date
2011-01-11 09:47:00 540700 21581 SKULLS DESIGN COTTON TOTE BAG
2011-01-11 09:47:00 540700 22619 SET OF 6 SOLDIER SKITTLES
2011-01-11 09:47:00 540700 84997B RED 3 PIECE RETROSPOT CUTLERY SET
2011-01-11 09:47:00 540700 20727 LUNCH BAG BLACK SKULL.
2011-01-11 09:47:00 540700 20726 LUNCH BAG WOODLAND
2011-01-11 09:47:00 540700 22383 LUNCH BAG SUKI DESIGN
2011-01-11 09:47:00 540700 21249 WOODLAND HEIGHT CHART STICKERS
2011-01-11 09:47:00 540700 22378 WALL TIDY RETROSPOT
2011-01-11 09:47:00 540700 22175 PINK OWL SOFT TOY
2011-01-11 09:47:00 540700 22176 BLUE OWL SOFT TOY
2011-01-11 09:47:00 540700 84997C BLUE 3 PIECE POLKADOT CUTLERY SET
2011-01-11 09:47:00 540700 20728 LUNCH BAG CARS BLUE
2011-01-11 09:47:00 540700 22382 LUNCH BAG SPACEBOY DESIGN
2011-01-11 09:47:00 540700 21915 RED HARMONICA IN BOX
2011-01-11 09:47:00 540700 22549 PICTURE DOMINOES
2011-01-11 09:47:00 540700 21578 WOODLAND DESIGN COTTON TOTE BAG
2011-01-11 09:47:00 540700 21577 SAVE THE PLANET COTTON TOTE BAG
2011-01-11 09:47:00 540700 22245 HOOK, 1 HANGER ,MAGIC GARDEN
2011-01-11 09:47:00 540700 22244 3 HOOK HANGER MAGIC GARDEN

```

Transaction Date	Quantity	UnitPrice	CustomerID
2011-01-11 09:47:00	6	2.25	12393.0
2011-01-11 09:47:00	8	3.75	12393.0
2011-01-11 09:47:00	6	3.75	12393.0
2011-01-11 09:47:00	20	1.65	12393.0
2011-01-11 09:47:00	20	1.65	12393.0
2011-01-11 09:47:00	10	1.65	12393.0
2011-01-11 09:47:00	6	2.95	12393.0
2011-01-11 09:47:00	20	0.85	12393.0
2011-01-11 09:47:00	12	2.95	12393.0
2011-01-11 09:47:00	12	2.95	12393.0
2011-01-11 09:47:00	6	3.75	12393.0
2011-01-11 09:47:00	20	1.65	12393.0
2011-01-11 09:47:00	20	1.65	12393.0
2011-01-11 09:47:00	12	1.25	12393.0
2011-01-11 09:47:00	12	1.45	12393.0
2011-01-11 09:47:00	12	2.25	12393.0
2011-01-11 09:47:00	12	2.25	12393.0
2011-01-11 09:47:00	12	0.85	12393.0
2011-01-11 09:47:00	12	1.95	12393.0

```
[ ]: sales_df.iloc[0]
```

```
[ ]: InvoiceNo          536389
      StockCode         22941
      Description      CHRISTMAS LIGHTS 10 REINDEER
      Quantity          6
      UnitPrice         8.5
      CustomerID       12431.0
      Name: (Australia, 2010-12-01 10:03:00), dtype: object
```

```
[ ]: sales_df.iloc[0,2]
```

```
[ ]: 'CHRISTMAS LIGHTS 10 REINDEER'
```

```
[ ]: Transposed_df = sales_df.transpose()
      Transposed_df
```

```
[ ]: Transaction Location          Australia \
      Transaction Date          2010-12-01 10:03:00
      InvoiceNo                  536389
      StockCode                  22941
      Description      CHRISTMAS LIGHTS 10 REINDEER
      Quantity                   6
      UnitPrice                  8.5
```

CustomerID	12431.0
Transaction Location	\
Transaction Date	2010-12-01 10:03:00
InvoiceNo	536389
StockCode	21622
Description	VINTAGE UNION JACK CUSHION COVER
Quantity	8
UnitPrice	4.95
CustomerID	12431.0

Transaction Location	\
Transaction Date	2010-12-01 10:03:00
InvoiceNo	536389
StockCode	21791
Description	VINTAGE HEADS AND TAILS CARD GAME
Quantity	12
UnitPrice	1.25
CustomerID	12431.0

Transaction Location	\
Transaction Date	2010-12-01 10:03:00
InvoiceNo	536389
StockCode	35004C
Description	SET OF 3 COLOURED FLYING DUCKS
Quantity	6
UnitPrice	5.45
CustomerID	12431.0

Transaction Location	\	
Transaction Date	2010-12-01 10:03:00	2010-12-01 10:03:00
InvoiceNo	536389	536389
StockCode	35004G	85014B
Description	SET OF 3 GOLD FLYING DUCKS	RED RETROSPOT UMBRELLA
Quantity	4	6
UnitPrice	6.35	5.95
CustomerID	12431.0	12431.0

Transaction Location	\	
Transaction Date	2010-12-01 10:03:00	2010-12-01 10:03:00
InvoiceNo	536389	536389
StockCode	85014A	22193
Description	BLACK/BLUE POLKADOT UMBRELLA	RED DINER WALL CLOCK
Quantity	3	2
UnitPrice	5.95	8.5
CustomerID	12431.0	12431.0

Transaction Location			\
Transaction Date	2010-12-01 10:03:00	2010-12-01 10:03:00	
InvoiceNo	536389	536389	
StockCode	22726	22727	
Description	ALARM CLOCK BAKELIKE GREEN	ALARM CLOCK BAKELIKE RED	
Quantity	4	4	
UnitPrice	3.75	3.75	
CustomerID	12431.0	12431.0	

Transaction Location	...	Unspecified	\
Transaction Date	...	2011-11-24 14:55:00	
InvoiceNo	...	578539	
StockCode	...	23569	
Description	...	TRADITIONAL ALPHABET STAMP SET	
Quantity	...	4	
UnitPrice	...	4.95	
CustomerID	...	NaN	

Transaction Location			\
Transaction Date		2011-11-24 14:55:00	
InvoiceNo		578539	
StockCode		23507	
Description		MINI PLAYING CARDS BUFFALO BILL	
Quantity		20	
UnitPrice		0.42	
CustomerID		NaN	

Transaction Location			\
Transaction Date	2011-11-24 14:55:00	2011-11-24 14:55:00	
InvoiceNo	578539	578539	
StockCode	21889	21914	
Description	WOODEN BOX OF DOMINOES	BLUE HARMONICA IN BOX	
Quantity	12	12	
UnitPrice	1.25	1.25	
CustomerID	NaN	NaN	

Transaction Location			\
Transaction Date	2011-11-24 14:55:00	2011-11-24 14:55:00	
InvoiceNo	578539	578539	
StockCode	22619	22560	
Description	SET OF 6 SOLDIER SKITTLES	TRADITIONAL MODELLING CLAY	
Quantity	4	24	
UnitPrice	3.75	1.25	
CustomerID	NaN	NaN	

Transaction Location			\
Transaction Date		2011-11-24 14:55:00	

InvoiceNo	578539
StockCode	23570
Description	TRADITIONAL PICK UP STICKS GAME
Quantity	12
UnitPrice	1.25
CustomerID	NaN

Transaction Location	\
Transaction Date	2011-11-24 14:55:00
InvoiceNo	578539
StockCode	23571
Description	TRADITIONAL NAUGHTS & CROSSES
Quantity	12
UnitPrice	1.65
CustomerID	NaN

Transaction Location	\
Transaction Date	2011-11-24 14:55:00
InvoiceNo	578539
StockCode	84992
Description	72 SWEETHEART FAIRY CAKE CASES
Quantity	24
UnitPrice	0.55
CustomerID	NaN

Transaction Location	
Transaction Date	2011-11-24 14:55:00
InvoiceNo	578539
StockCode	23309
Description	SET OF 60 I LOVE LONDON CAKE CASES
Quantity	24
UnitPrice	0.55
CustomerID	NaN

[6 rows x 541909 columns]

```
[ ]: sales_df.loc(['Australia','2010-12-01 10:03:00'],'UnitPrice']
```

```
[ ]: Transaction Date
2010-12-01 10:03:00    8.50
2010-12-01 10:03:00    4.95
2010-12-01 10:03:00    1.25
2010-12-01 10:03:00    5.45
2010-12-01 10:03:00    6.35
2010-12-01 10:03:00    5.95
2010-12-01 10:03:00    5.95
2010-12-01 10:03:00    8.50
```

```

2010-12-01 10:03:00    3.75
2010-12-01 10:03:00    3.75
2010-12-01 10:03:00    8.50
2010-12-01 10:03:00    8.50
2010-12-01 10:03:00    1.65
2010-12-01 10:03:00    0.85
Name: UnitPrice, dtype: float64

```

```
[ ]: sales_df.loc[('Australia','2010-12-01 10:03:00'):('Belgium','2010-12-01 10:03:00'), 'UnitPrice']
```

```
[ ]: Transaction Location Transaction Date
      Australia      2010-12-01 10:03:00    8.50
                        2010-12-01 10:03:00    4.95
                        2010-12-01 10:03:00    1.25
                        2010-12-01 10:03:00    5.45
                        2010-12-01 10:03:00    6.35
                        ...
      Bahrain      2011-05-09 13:49:00    4.25
                        2011-05-19 17:47:00    9.95
                        2011-05-19 17:47:00    1.45
                        2011-05-19 17:47:00    9.95
                        2011-05-19 17:47:00    2.95
Name: UnitPrice, Length: 1679, dtype: float64
```

```
[ ]: import datetime as dt

my_date = dt.date(2020,3,22)
my_date
```

```
[ ]: datetime.date(2020, 3, 22)
```

```
[ ]: str(my_date)
```

```
[ ]: '2020-03-22'
```

```
[ ]: my_date.day
```

```
[ ]: 22
```

```
[ ]: my_date.month
```

```
[ ]: 3
```

```
[ ]: my_date.year
```

```
[ ]: 2020
```

```
[ ]: my_datetime = dt.datetime(2020,3,22,10,30,0)
my_datetime
```

```
[ ]: datetime.datetime(2020, 3, 22, 10, 30)
```

```
[ ]: str(my_datetime)
```

```
[ ]: '2020-03-22 10:30:00'
```

```
[ ]: my_datetime.hour
```

```
[ ]: 10
```

```
[ ]: import calendar
print(calendar.month(2021,3))
```

```
    March 2021
Mo Tu We Th Fr Sa Su
 1  2  3  4  5  6  7
 8  9 10 11 12 13 14
15 16 17 18 19 20 21
22 23 24 25 26 27 28
29 30 31
```

```
[ ]: dates = pd.Series(['2020/03/22', '2020-10-25', 'February 20th,2020'])
dates
```

```
[ ]: 0      2020/03/22
1      2020-10-25
2      February 20th,2020
dtype: object
```

```
[ ]: my_dates = pd.to_datetime(dates,format='mixed')
my_dates
```

```
[ ]: 0      2020-03-22
1      2020-10-25
2      2020-02-20
dtype: datetime64[ns]
```

```
[ ]: pd.Timestamp('2020-3-22') # Use hyphens or slashes instead of commas
```

```
[ ]: Timestamp('2020-03-22 00:00:00')
```

```
[ ]: day_1 = pd.Timestamp('2020-3-22')
day_2 = pd.Timestamp('2020-3-25')
```

```
day_2-day_1
```

```
[ ]: Timedelta('3 days 00:00:00')
```

```
[ ]: date_1=dt.date(2020,3,22)
      date_2=dt.date(2020,4,22)
      date_3=dt.date(2020,4,22)
      dates_list = [date_1,date_2,date_3]
      dates_list
```

```
[ ]: [datetime.date(2020, 3, 22),
      datetime.date(2020, 4, 22),
      datetime.date(2020, 4, 22)]
```

```
[ ]: dates_index=pd.DatetimeIndex(dates_list)
      dates_index
```

```
[ ]: DatetimeIndex(['2020-03-22', '2020-04-22', '2020-04-22'],
                    dtype='datetime64[ns]', freq=None)
```

```
[ ]: sales = pd.Series(data=[100,200,300],index=dates_index)
      sales
```

```
[ ]: 2020-03-22    100
      2020-04-22    200
      2020-04-22    300
      dtype: int64
```

```
[ ]: my_dates = pd.date_range(start='2020-3-22',end='2020-4-22',freq='D')
      my_dates
```

```
[ ]: DatetimeIndex(['2020-03-22', '2020-03-23', '2020-03-24', '2020-03-25',
                    '2020-03-26', '2020-03-27', '2020-03-28', '2020-03-29',
                    '2020-03-30', '2020-03-31', '2020-04-01', '2020-04-02',
                    '2020-04-03', '2020-04-04', '2020-04-05', '2020-04-06',
                    '2020-04-07', '2020-04-08', '2020-04-09', '2020-04-10',
                    '2020-04-11', '2020-04-12', '2020-04-13', '2020-04-14',
                    '2020-04-15', '2020-04-16', '2020-04-17', '2020-04-18',
                    '2020-04-19', '2020-04-20', '2020-04-21', '2020-04-22'],
                    dtype='datetime64[ns]', freq='D')
```

```
[ ]: my_dates = pd.date_range(start='2020-3-22',end='2020-6-22',freq='ME')
      my_dates
```

```
[ ]: DatetimeIndex(['2020-03-31', '2020-04-30', '2020-05-31'],
                    dtype='datetime64[ns]', freq='ME')
```

```
[ ]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
[ ]: df=pd.read_csv('/content/text_data.csv')
df
```

```
[ ]:
rating      date      variation \
0          5  31-Jul-18  Charcoal Fabric
1          5  31-Jul-18  Charcoal Fabric
2          4  31-Jul-18   Walnut Finish
3          5  31-Jul-18  Charcoal Fabric
4          5  31-Jul-18  Charcoal Fabric
...
3145      5  30-Jul-18    Black Dot
3146      5  30-Jul-18    Black Dot
3147      5  30-Jul-18    Black Dot
3148      5  30-Jul-18    White Dot
3149      4  29-Jul-18    Black Dot

verified_reviews  feedback
0              Love my Echo!          1
1              Loved it!             1
2  Sometimes while playing a game, you can answer...      1
3  I have had a lot of fun with this thing. My 4 ...      1
4              Music                 1
...
3145  Perfect for kids, adults and everyone in betwe...      1
3146  Listening to music, searching locations, check...      1
3147  I do love these things, i have them running my...      1
3148  Only complaint I have is that the sound qualit...      1
3149              Good                1
```

[3150 rows x 5 columns]

```
[ ]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3150 entries, 0 to 3149
Data columns (total 5 columns):
#   Column              Non-Null Count  Dtype
---  -
0   rating              3150 non-null  int64
1   date                3150 non-null  object
2   variation            3150 non-null  object
3   verified_reviews    3149 non-null  object
4   feedback            3150 non-null  int64
```

```
dtypes: int64(2), object(3)
memory usage: 123.2+ KB
```

```
[ ]: df.describe()
```

```
[ ]:
count    3150.000000    3150.000000
mean      4.463175      0.918413
std       1.068506      0.273778
min       1.000000      0.000000
25%       4.000000      1.000000
50%       5.000000      1.000000
75%       5.000000      1.000000
max       5.000000      1.000000
```

```
[ ]: df['variation'].nunique()
```

```
[ ]: 16
```

```
[ ]: df['verified_reviews'].str.lower()
```

```
[ ]: 0          love my echo!
    1          loved it!
    2    sometimes while playing a game, you can answer...
    3    i have had a lot of fun with this thing. my 4 ...
    4          music
    ...
    3145    perfect for kids, adults and everyone in betwe...
    3146    listening to music, searching locations, check...
    3147    i do love these things, i have them running my...
    3148    only complaint i have is that the sound qualit...
    3149          good
    Name: verified_reviews, Length: 3150, dtype: object
```

```
[ ]: df.columns = df.columns.str.upper()
    df
```

```
[ ]:
RATING  DATE  VARIATION \
0      5  31-Jul-18  Charcoal Fabric
1      5  31-Jul-18  Charcoal Fabric
2      4  31-Jul-18   Walnut Finish
3      5  31-Jul-18  Charcoal Fabric
4      5  31-Jul-18  Charcoal Fabric
...    ...    ...
3145    5  30-Jul-18   Black Dot
3146    5  30-Jul-18   Black Dot
3147    5  30-Jul-18   Black Dot
```

```

3148      5  30-Jul-18      White  Dot
3149      4  29-Jul-18      Black  Dot

```

```

                                VERIFIED_REVIEWS  FEEDBACK
0                                Love my Echo!          1
1                                Loved it!             1
2      Sometimes while playing a game, you can answer...  1
3      I have had a lot of fun with this thing. My 4 ...  1
4                                Music                 1
...                                ...                 ...
3145   Perfect for kids, adults and everyone in betwe...  1
3146   Listening to music, searching locations, check...  1
3147   I do love these things, i have them running my...  1
3148   Only complaint I have is that the sound qualit...  1
3149                                Good                 1

```

[3150 rows x 5 columns]

```
[ ]: df.columns = df.columns.str.lower()
df
```

```
[ ]:
   rating  date      variation \
0      5  31-Jul-18  Charcoal Fabric
1      5  31-Jul-18  Charcoal Fabric
2      4  31-Jul-18   Walnut Finish
3      5  31-Jul-18  Charcoal Fabric
4      5  31-Jul-18  Charcoal Fabric
...
3145      5  30-Jul-18      Black  Dot
3146      5  30-Jul-18      Black  Dot
3147      5  30-Jul-18      Black  Dot
3148      5  30-Jul-18      White  Dot
3149      4  29-Jul-18      Black  Dot

```

```

                                verified_reviews  feedback
0                                Love my Echo!          1
1                                Loved it!             1
2      Sometimes while playing a game, you can answer...  1
3      I have had a lot of fun with this thing. My 4 ...  1
4                                Music                 1
...                                ...                 ...
3145   Perfect for kids, adults and everyone in betwe...  1
3146   Listening to music, searching locations, check...  1
3147   I do love these things, i have them running my...  1
3148   Only complaint I have is that the sound qualit...  1
3149                                Good                 1

```


[3150 rows x 5 columns]

```
[ ]: negative_df = df[df['feedback'] == 0]
negative_df
```

```
[ ]:      rating      date      variation \
46         2  30-Jul-18   Charcoal Fabric
111        2  30-Jul-18   Charcoal Fabric
141        1  30-Jul-18   Charcoal Fabric
162        1  30-Jul-18   Sandstone Fabric
176        2  30-Jul-18  Heather Gray Fabric
...      ...      ...      ...
3047       1  30-Jul-18   Black Dot
3048       1  30-Jul-18   White Dot
3067       2  30-Jul-18   Black Dot
3091       1  30-Jul-18   Black Dot
3096       1  30-Jul-18   White Dot

      verified_reviews  feedback
46  It's like Siri, in fact, Siri answers more acc...      0
111  Sound is terrible if u want good music too get...      0
141              Not much features.      0
162  Stopped working after 2 weeks ,didn't follow c...      0
176              Sad joke. Worthless.      0
...      ...      ...
3047  Echo Dot responds to us when we aren't even ta...      0
3048              NOT CONNECTED TO MY PHONE PLAYLIST :(      0
3067  The only negative we have on this product is t...      0
3091              I didn't order it      0
3096  The product sounded the same as the emoji spea...      0
```

[257 rows x 5 columns]

```
[ ]: positive_df = df[df['feedback'] == 1]
positive_df
```

```
[ ]:      rating      date      variation \
0         5  31-Jul-18   Charcoal Fabric
1         5  31-Jul-18   Charcoal Fabric
2         4  31-Jul-18   Walnut Finish
3         5  31-Jul-18   Charcoal Fabric
4         5  31-Jul-18   Charcoal Fabric
...      ...      ...      ...
3145       5  30-Jul-18   Black Dot
3146       5  30-Jul-18   Black Dot
3147       5  30-Jul-18   Black Dot
3148       5  30-Jul-18   White Dot
```

3149	4	29-Jul-18	Black Dot	
------	---	-----------	-----------	--

		verified_reviews	feedback
0		Love my Echo!	1
1		Loved it!	1
2	Sometimes while playing a game, you can answer...		1
3	I have had a lot of fun with this thing. My 4 ...		1
4	Music		1
...	
3145	Perfect for kids, adults and everyone in betwe...		1
3146	Listening to music, searching locations, check...		1
3147	I do love these things, i have them running my...		1
3148	Only complaint I have is that the sound qualit...		1
3149	Good		1

[2893 rows x 5 columns]

```
[ ]: poor_rating_df=df[df['rating']==1]
poor_rating_df
```

```
[ ]:
rating    date    variation \
141      1  30-Jul-18  Charcoal Fabric
162      1  30-Jul-18  Sandstone Fabric
341      1  28-Jul-18  Charcoal Fabric
350      1  31-Jul-18           Black
361      1  29-Jul-18           Black
...      ...      ...      ...
3024     1  30-Jul-18  Black Dot
3047     1  30-Jul-18  Black Dot
3048     1  30-Jul-18  White Dot
3091     1  30-Jul-18  Black Dot
3096     1  30-Jul-18  White Dot
```

		verified_reviews	feedback
141		Not much features.	0
162	Stopped working after 2 weeks ,didn't follow c...		0
341	Alexa hardly came on..		0
350	Item no longer works after just 5 months of us...		0
361	This thing barely works. You have to select 3r...		0
...	
3024	I was really happy with my original echo so i ...		0
3047	Echo Dot responds to us when we aren't even ta...		0
3048	NOT CONNECTED TO MY PHONE PLAYLIST :(0
3091	I didn't order it		0
3096	The product sounded the same as the emoji spea...		0

[161 rows x 5 columns]

```
[ ]: good_rating_df=df[df['rating']==5]
good_rating_df
```

```
[ ]:      rating      date      variation \
0         5  31-Jul-18   Charcoal Fabric
1         5  31-Jul-18   Charcoal Fabric
3         5  31-Jul-18   Charcoal Fabric
4         5  31-Jul-18   Charcoal Fabric
5         5  31-Jul-18  Heather Gray Fabric
...
3144      5  30-Jul-18      Black Dot
3145      5  30-Jul-18      Black Dot
3146      5  30-Jul-18      Black Dot
3147      5  30-Jul-18      Black Dot
3148      5  30-Jul-18      White Dot

      verified_reviews  feedback
0              Love my Echo!      1
1              Loved it!      1
3  I have had a lot of fun with this thing. My 4 ...      1
4              Music      1
5  I received the echo as a gift. I needed anothe...      1
...
3144              love it      1
3145  Perfect for kids, adults and everyone in betwe...      1
3146  Listening to music, searching locations, check...      1
3147  I do love these things, i have them running my...      1
3148  Only complaint I have is that the sound qualit...      1
```

[2286 rows x 5 columns]

```
[ ]: df['reviews_length']=df['verified_reviews'].str.len()
df
```

```
[ ]:      rating      date      variation \
0         5  31-Jul-18   Charcoal Fabric
1         5  31-Jul-18   Charcoal Fabric
2         4  31-Jul-18   Walnut Finish
3         5  31-Jul-18   Charcoal Fabric
4         5  31-Jul-18   Charcoal Fabric
...
3145      5  30-Jul-18      Black Dot
3146      5  30-Jul-18      Black Dot
3147      5  30-Jul-18      Black Dot
3148      5  30-Jul-18      White Dot
3149      4  29-Jul-18      Black Dot
```

	verified_reviews	feedback \
0	Love my Echo!	1
1	Loved it!	1
2	Sometimes while playing a game, you can answer...	1
3	I have had a lot of fun with this thing. My 4 ...	1
4	Music	1
...
3145	Perfect for kids, adults and everyone in betwe...	1
3146	Listening to music, searching locations, check...	1
3147	I do love these things, i have them running my...	1
3148	Only complaint I have is that the sound qualit...	1
3149	Good	1

	reviews_length
0	13.0
1	9.0
2	195.0
3	172.0
4	5.0
...	...
3145	50.0
3146	135.0
3147	441.0
3148	380.0
3149	4.0

[3150 rows x 6 columns]

```
[ ]: min_char = df['verified_reviews'].str.len().min()
min_char
```

```
[ ]: 1.0
```

```
[ ]: max_char = df['verified_reviews'].str.len().max()
max_char
```

```
[ ]: 2851.0
```

```
[ ]: df[df['reviews_length'] == min_char]
```

	rating	date	variation	verified_reviews	feedback \
60	5	30-Jul-18	Heather Gray Fabric		1
85	5	30-Jul-18	Heather Gray Fabric		1
183	3	29-Jul-18	Heather Gray Fabric		1
219	5	29-Jul-18	Sandstone Fabric		1
374	1	26-Jul-18	Black		0
...

3114	3	30-Jul-18	Black	Dot	1
3120	5	30-Jul-18	Black	Dot	1
3123	4	30-Jul-18	Black	Dot	1
3126	5	30-Jul-18	Black	Dot	1
3141	3	30-Jul-18	Black	Dot	1

reviews_length	
60	1.0
85	1.0
183	1.0
219	1.0
374	1.0
...	...
3114	1.0
3120	1.0
3123	1.0
3126	1.0
3141	1.0

[81 rows x 6 columns]

```
[ ]: df[df['reviews_length'] == max_char]
```

```
[ ]:      rating      date  variation \
2016      5  20-Jul-18  Black  Plus

                                verified_reviews  feedback \
2016  Incredible piece of technology.I have this rig...      1

      reviews_length
2016              2851.0
```

```
[ ]: df[df['reviews_length'] == max_char]['verified_reviews'].iloc[0]
```

```
[ ]: "Incredible piece of technology.I have this right center of my living room on an
island kitchen counter. The mic and speaker goes in every direction and the
quality of the sound is quite good. I connected the Echo via Bluetooth to my
Sony soundbar on my TV but find the Echo placement and 360 sound more appealing.
It's no audiophile equipment but there is good range and decent bass. The sound
is more than adequate for any indoor entertaining and loud enough to bother
neighbors in my building. The knob on the top works great for adjusting volume.
This is my first Echo device and I would imagine having to press volume buttons
(on the Echo 2) a large inconvenience and not as precise. For that alone I would
recommend this over the regular Echo (2nd generation).The piece looks quality
and is quite sturdy with some weight on it. The rubber material on the bottom
has a good grip on the granite counter-- my cat can even rub her scent on it
without tipping it over.This order came with a free Philips Hue Bulb which I
```

installed along with an extra one I bought. I put the 2 bulbs into my living room floor lamp, turned on the light, and all I had to do was say "Alexa, connect my devices". The default names for each bulb was assigned as "First light" and "Second light", so I can have a dimmer floor lamp if I just turned on/off one of the lights by saying "Alexa, turn off the second light". In the Alexa app, I created a 'Group' with "First light" and "Second light", and named the group "The light", so to turn on the lamp with both bulbs shining I just say "Alexa, turn on The light". I was surprised how easily the bulbs connected to the Echo Plus with its built in hub. I thought I would have to buy a hub bridge to connect to my floor lamp power plug. Apparently there is some technology built directly inside the bulb! I was surprised by that. Awesome. You will feel like Tony Stark on this device. I added quite a few "Skills" like 'Thunderstorm sounds' and 'Quote of the day'. Alexa always loads them up quickly. Adding songs that you hear to specific playlists on Amazon Music is also a great feature. I can go on and on and this is only my second day of ownership. I was lucky to buy this for \$100 on Prime Day, but I think for \$150 is it pretty expensive considering the Echo 2 is only \$100. In my opinion, you will be paying a premium for the Echo Plus and you have to decide if the value is there for you: 1) Taller and 360 sound unit. 2) Volume knob on top that you spin (I think this is a huge benefit over buttons) 3) Built in hub for Hue bulbs. After researching more, there are some cons to this setup if you plan on having more advanced light setups. For me and my floor lamp, it's just perfect. I highly recommend it and will buy an Echo dot for my bedroom now."

```
[ ]: # Replace NaN values in 'verified_reviews' with an empty string before applying
      ↳ the mask
mask = df['verified_reviews'].fillna('').str.lower().str.endswith('love')
df_filtered = df[mask]
df_filtered
```

```
[ ]:      rating      date      variation  verified_reviews  feedback  \
438         5    7-Jul-18         Black             Love          1
2018        5   19-Jul-18    Black Plus    Love, Love, Love          1

      reviews_length
438                 4.0
2018                16.0
```

```
[ ]: # Replace NaN values in 'verified_reviews' with an empty string before applying
      ↳ the mask
# and fill NaN values in the mask with False
mask = df['verified_reviews'].fillna('').str.lower().str.contains('bad').
      ↳ fillna(False)

# Now the mask should only contain True/False values, and you can use it to
      ↳ filter the DataFrame
```

```
df_filtered = df[mask]
df_filtered
```

```
[ ]:      rating      date      variation \
113         5  30-Jul-18      Charcoal Fabric
288         5  29-Jul-18      Charcoal Fabric
567         1  11-Jun-18              White
601         3   5-Jun-18              Black
808         5  30-Jul-18      Charcoal Fabric
983         5  29-Jul-18      Charcoal Fabric
1233        5  26-Jul-18      Black Spot
1338        5  19-Jul-18      White Spot
1440         4  30-Jul-18      White Show
1448         1  30-Jul-18      White Show
1786         4  30-Jul-18      Black Plus
1809         5  29-Jul-18      Black Plus
1967         5  25-Jul-18      White Plus
1983         2  23-Jul-18      Black Plus
1994         5  21-Jul-18      Black Plus
2021         1  19-Jul-18      Black Plus
2226         3  30-Jul-18  Configuration: Fire TV Stick
2651         4  30-Jul-18      Black Dot
2688         2  30-Jul-18      Black Dot
3002         4  30-Jul-18      Black Dot
3039         2  30-Jul-18      Black Dot

      verified_reviews  feedback \
113  Easy to set up and use. Too bad it has to be p...      1
288  Mainly use it for the music but I'm learning m...      1
567  Very bad this device, I don't know if it's bec...      0
601              Not bad bad speaker for sound      1
808  Easy to set up and use. Too bad it has to be p...      1
983  Mainly use it for the music but I'm learning m...      1
1233  I purchased this on prime day mostly as a pres...      1
1338  Omg where do I start, I LOVE THIS THING! None ...      1
1440  Love the product idea. Super easy to set up & ...      1
1448  Don't waste your money. I own an echo gen 1, ...      0
1786  Not that much different than the one we had. ...      1
1809  I love my echo plus. I haven't one bad thing t...      1
1967  I have been hesitant to purchase anything smar...      1
1983  I'm sure the Echo is as cool as everyone says...      0
1994  I have an original Echo and really like it. I ...      1
2021  I would like to tell you that I have a reason ...      0
2226  Works very well for Amazon and Netflix but wil...      1
2651  It isn't bad for what it is. Have issues with ...      1
2688  Weak sound. Compared to the Google Home Mini t...      0
3002  It isn't bad for what it is. Have issues with ...      1
```

3039 Weak sound. Compared to the Google Home Mini t...

0

	reviews_length
113	127.0
288	99.0
567	205.0
601	29.0
808	127.0
983	99.0
1233	727.0
1338	57.0
1440	978.0
1448	770.0
1786	282.0
1809	684.0
1967	882.0
1983	460.0
1994	259.0
2021	1124.0
2226	626.0
2651	203.0
2688	300.0
3002	203.0
3039	300.0

```
[ ]: df['verified_reviews'].str.split(' ')
```

```
[ ]: 0          [Love, my, Echo!]
      1          [Loved, it!]
      2  [Sometimes, while, playing, a, game,, you, can...
      3  [I, have, had, a, lot, of, fun, with, this, th...
      4          [Music]

      ...
3145  [Perfect, for, kids,, adults, and, everyone, i...
3146  [Listening, to, music,, searching, locations,,...
3147  [I, do, love, these, things,, i, have, them, r...
3148  [Only, complaint, I, have, is, that, the, soun...
3149          [Good]
Name: verified_reviews, Length: 3150, dtype: object
```

```
[ ]: df['verified_reviews'].str.split(' ').str.get(0)
```

```
[ ]: 0      Love
      1     Loved
      2  Sometimes
      3         I
      4     Music
```



```

...
3145    Perfect
3146    Listening
3147         I
3148        Only
3149        Good
Name: verified_reviews, Length: 3150, dtype: object

```

```
[ ]: import string
      string.punctuation
```

```
[ ]: '!"#$%&\'()*+,-./:;<=>?@[\\]^_`{|}~'
```

```
[ ]: Test = '$I l@ove P$andas &Data Analytics!!!'
```

```
[ ]: Test_punc_removed = [char for char in Test if char not in string.punctuation]
      Test_punc_removed
```

```
[ ]: ['I',
      ' ',
      'l',
      'o',
      'v',
      'e',
      ' ',
      'P',
      'a',
      'n',
      'd',
      'a',
      's',
      ' ',
      'D',
      'a',
      't',
      'a',
      ' ',
      'A',
      'n',
      'a',
      'l',
      'y',
      't',
      'i',
      'c',
      's']
```

```
[ ]: Test_punc_removed_join=''.join(Test_punc_removed)
Test_punc_removed_join
```

```
[ ]: 'I love Pandas Data Analytics'
```

```
[ ]: def remove_punc(message):
    # Convert message to string first, handling NaN
    message = str(message)
    Test_punc_removed = [char for char in message if char not in string.
↪punctuation]
    Test_punc_removed_join=''.join(Test_punc_removed)
    return Test_punc_removed_join
```

```
[ ]: df['verified_reviews'].apply(remove_punc)
```

```
[ ]: 0                Love my Echo
1                Loved it
2    Sometimes while playing a game you can answer ...
3    I have had a lot of fun with this thing My 4 y...
4                Music

...

3145    Perfect for kids adults and everyone in between
3146    Listening to music searching locations checkin...
3147    I do love these things i have them running my ...
3148    Only complaint I have is that the sound qualit...
3149                Good
Name: verified_reviews, Length: 3150, dtype: object
```

```
[ ]: !pip install gensim
!pip install nltk
```

```
Requirement already satisfied: gensim in /usr/local/lib/python3.10/dist-packages
(4.3.3)
Requirement already satisfied: numpy<2.0,>=1.18.5 in
/usr/local/lib/python3.10/dist-packages (from gensim) (1.26.4)
Requirement already satisfied: scipy<1.14.0,>=1.7.0 in
/usr/local/lib/python3.10/dist-packages (from gensim) (1.13.1)
Requirement already satisfied: smart-open>=1.8.1 in
/usr/local/lib/python3.10/dist-packages (from gensim) (7.0.5)
Requirement already satisfied: wrapt in /usr/local/lib/python3.10/dist-packages
(from smart-open>=1.8.1->gensim) (1.17.0)
Requirement already satisfied: nltk in /usr/local/lib/python3.10/dist-packages
(3.9.1)
Requirement already satisfied: click in /usr/local/lib/python3.10/dist-packages
(from nltk) (8.1.7)
Requirement already satisfied: joblib in /usr/local/lib/python3.10/dist-packages
(from nltk) (1.4.2)
```

Requirement already satisfied: regex>=2021.8.3 in
 /usr/local/lib/python3.10/dist-packages (from nltk) (2024.9.11)
 Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-packages
 (from nltk) (4.66.6)

```
[ ]: import nltk
from nltk.corpus import stopwords
nltk.download('stopwords')
import gensim
from gensim.utils import simple_preprocess
```

[nltk_data] Downloading package stopwords to /root/nltk_data...
 [nltk_data] Unzipping corpora/stopwords.zip.

```
[ ]: stop_words = stopwords.words('english')
stop_words.
↳extend(['Amazon', 'amazon', 'alexa', 'echo', 'Alexa', 'Device', 'Dot', 'dot'])
```

```
[ ]: gensim.utils.simple_preprocess(df['verified_reviews'][0])
```

```
[ ]: ['love', 'my', 'echo']
```

```
[ ]: def preprocess(text):
    text = str(text)
    result = []
    for token in gensim.utils.simple_preprocess(text):
        if token not in gensim.parsing.preprocessing.STOPWORDS and len(token)>3 and
↳token not in stop_words:
            result.append(token)
    return result
```

```
[ ]: df['verified_reviews_nopunc_nostopwords']=df['verified_reviews'].
↳apply(preprocess)
df
```

```
[ ]:
rating      date      variation \
0          5  31-Jul-18  Charcoal Fabric
1          5  31-Jul-18  Charcoal Fabric
2          4  31-Jul-18   Walnut Finish
3          5  31-Jul-18  Charcoal Fabric
4          5  31-Jul-18  Charcoal Fabric
...
3145      5  30-Jul-18    Black Dot
3146      5  30-Jul-18    Black Dot
3147      5  30-Jul-18    Black Dot
3148      5  30-Jul-18    White Dot
3149      4  29-Jul-18    Black Dot
```

		verified_reviews	feedback \
0		Love my Echo!	1
1		Loved it!	1
2	Sometimes while playing a game, you can answer...		1
3	I have had a lot of fun with this thing. My 4 ...		1
4		Music	1
...	
3145	Perfect for kids, adults and everyone in betwe...		1
3146	Listening to music, searching locations, check...		1
3147	I do love these things, i have them running my...		1
3148	Only complaint I have is that the sound qualit...		1
3149		Good	1

	reviews_length	verified_reviews_nopunc_nostopwords
0	13.0	[love]
1	9.0	[loved]
2	195.0	[playing, game, answer, question, correctly, s...
3	172.0	[thing, learns, dinosaurs, control, lights, pl...
4	5.0	[music]
...
3145	50.0	[perfect, kids, adults]
3146	135.0	[listening, music, searching, locations, check...
3147	441.0	[love, things, running, entire, home, lights, ...
3148	380.0	[complaint, sound, quality, great, commands, h...
3149	4.0	[good]

[3150 rows x 7 columns]

```
[ ]: df['verified_reviews'][38]
```

```
[ ]: 'This thing is way cool! You should get one. If you want to be cool, that is.'
```

```
[ ]: df['verified_reviews_nopunc_nostopwords'][38]
```

```
[ ]: ['thing', 'cool', 'want', 'cool']
```

```
[ ]: df['verified_reviews_nopunc_nostopwords_join'] =_
      ↪df['verified_reviews_nopunc_nostopwords'].apply(lambda x: ' '.join(x))
df
```

	rating	date	variation \
0	5	31-Jul-18	Charcoal Fabric
1	5	31-Jul-18	Charcoal Fabric
2	4	31-Jul-18	Walnut Finish
3	5	31-Jul-18	Charcoal Fabric
4	5	31-Jul-18	Charcoal Fabric

...
3145	5	30-Jul-18	Black Dot
3146	5	30-Jul-18	Black Dot
3147	5	30-Jul-18	Black Dot
3148	5	30-Jul-18	White Dot
3149	4	29-Jul-18	Black Dot

	verified_reviews	feedback	\
0	Love my Echo!	1	
1	Loved it!	1	
2	Sometimes while playing a game, you can answer...	1	
3	I have had a lot of fun with this thing. My 4 ...	1	
4	Music	1	
...	
3145	Perfect for kids, adults and everyone in betwe...	1	
3146	Listening to music, searching locations, check...	1	
3147	I do love these things, i have them running my...	1	
3148	Only complaint I have is that the sound qualit...	1	
3149	Good	1	

	reviews_length	verified_reviews_nopunc_nostopwords	\
0	13.0	[love]	
1	9.0	[loved]	
2	195.0	[playing, game, answer, question, correctly, s...	
3	172.0	[thing, learns, dinosaurs, control, lights, pl...	
4	5.0	[music]	
...	
3145	50.0	[perfect, kids, adults]	
3146	135.0	[listening, music, searching, locations, check...	
3147	441.0	[love, things, running, entire, home, lights, ...	
3148	380.0	[complaint, sound, quality, great, commands, h...	
3149	4.0	[good]	

	verified_reviews_nopunc_nostopwords_join
0	love
1	loved
2	playing game answer question correctly says wr...
3	thing learns dinosaurs control lights play gam...
4	music
...	...
3145	perfect kids adults
3146	listening music searching locations checking t...
3147	love things running entire home lights thermos...
3148	complaint sound quality great commands hooked ...
3149	good

[3150 rows x 8 columns]

```
[ ]: df['length']=df['verified_reviews_nopunc_nostopwords_join'].str.len()
df
```

```
[ ]:      rating      date      variation \
0         5  31-Jul-18  Charcoal Fabric
1         5  31-Jul-18  Charcoal Fabric
2         4  31-Jul-18   Walnut Finish
3         5  31-Jul-18  Charcoal Fabric
4         5  31-Jul-18  Charcoal Fabric
...
3145      5  30-Jul-18      Black Dot
3146      5  30-Jul-18      Black Dot
3147      5  30-Jul-18      Black Dot
3148      5  30-Jul-18      White Dot
3149      4  29-Jul-18      Black Dot
```

```
                                verified_reviews  feedback \
0                                Love my Echo!           1
1                                Loved it!             1
2      Sometimes while playing a game, you can answer...  1
3      I have had a lot of fun with this thing. My 4 ...  1
4                                Music                 1
...
3145  Perfect for kids, adults and everyone in betwe...  1
3146  Listening to music, searching locations, check...  1
3147  I do love these things, i have them running my...  1
3148  Only complaint I have is that the sound qualit...  1
3149                                Good                1
```

```
      reviews_length      verified_reviews_nopunc_nostopwords \
0              13.0                                [love]
1              9.0                                [loved]
2             195.0  [playing, game, answer, question, correctly, s...
3             172.0  [thing, learns, dinosaurs, control, lights, pl...
4              5.0                                [music]
...
3145             50.0                                [perfect, kids, adults]
3146            135.0  [listening, music, searching, locations, check...
3147            441.0  [love, things, running, entire, home, lights, ...
3148            380.0  [complaint, sound, quality, great, commands, h...
3149             4.0                                [good]
```

```
      verified_reviews_nopunc_nostopwords_join  length
0                                love            4
1                                loved           5
2      playing game answer question correctly says wr...  89
3      thing learns dinosaurs control lights play gam...  89
```

```

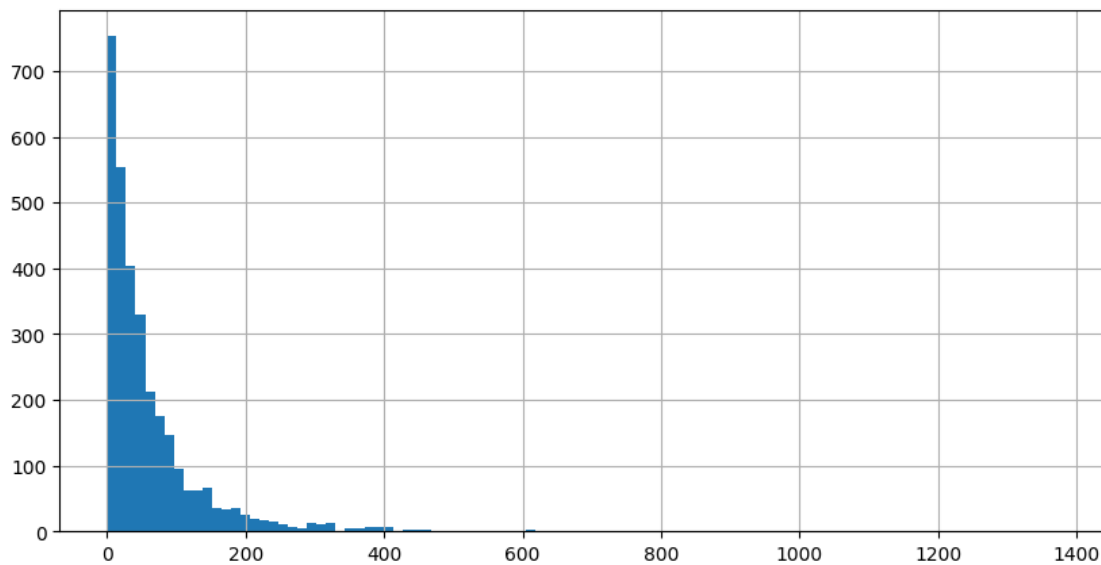
4                                     music      5
...
3145                                perfect kids adults    19
3146  listening music searching locations checking t...    83
3147  love things running entire home lights thermos...    220
3148  complaint sound quality great commands hooked ...    148
3149                                good      4

```

[3150 rows x 9 columns]

```
[ ]: df['length'].hist(bins=100,figsize=(10,5))
```

```
[ ]: <Axes: >
```



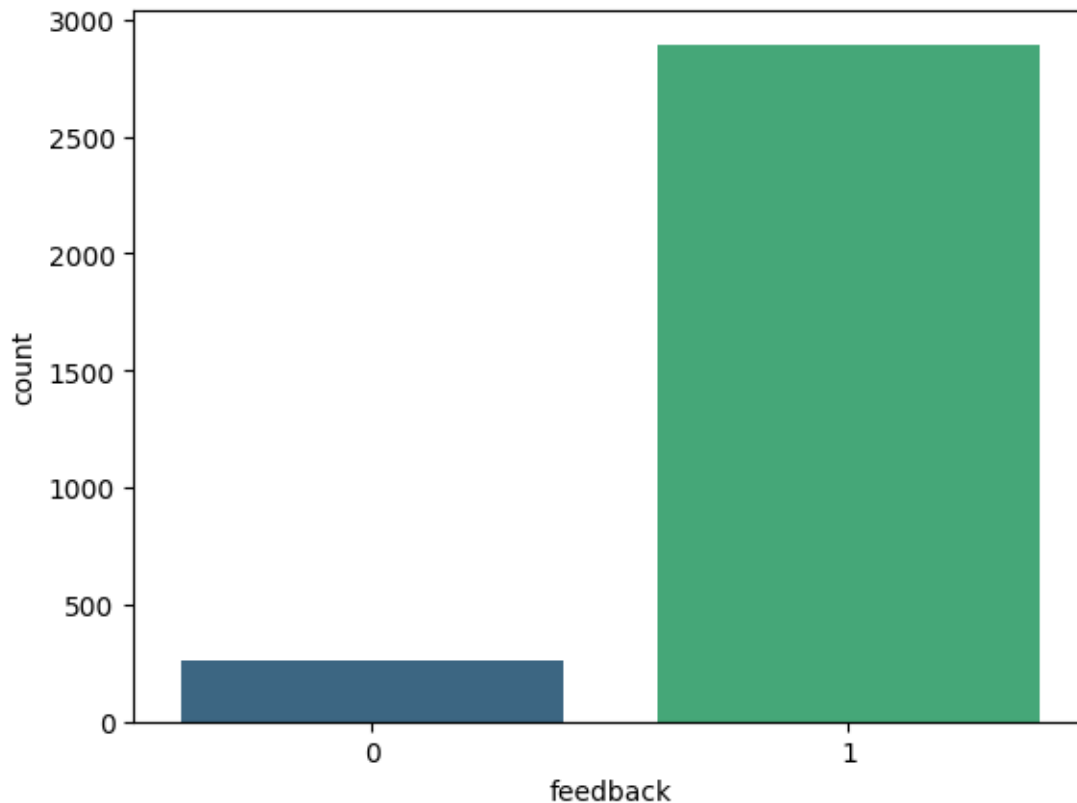
```
[ ]: sns.countplot(x='feedback', data=df, palette='viridis')
```

<ipython-input-184-aa0e6830abb0>:1: FutureWarning:

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

```
sns.countplot(x='feedback', data=df, palette='viridis')
```

```
[ ]: <Axes: xlabel='feedback', ylabel='count'>
```

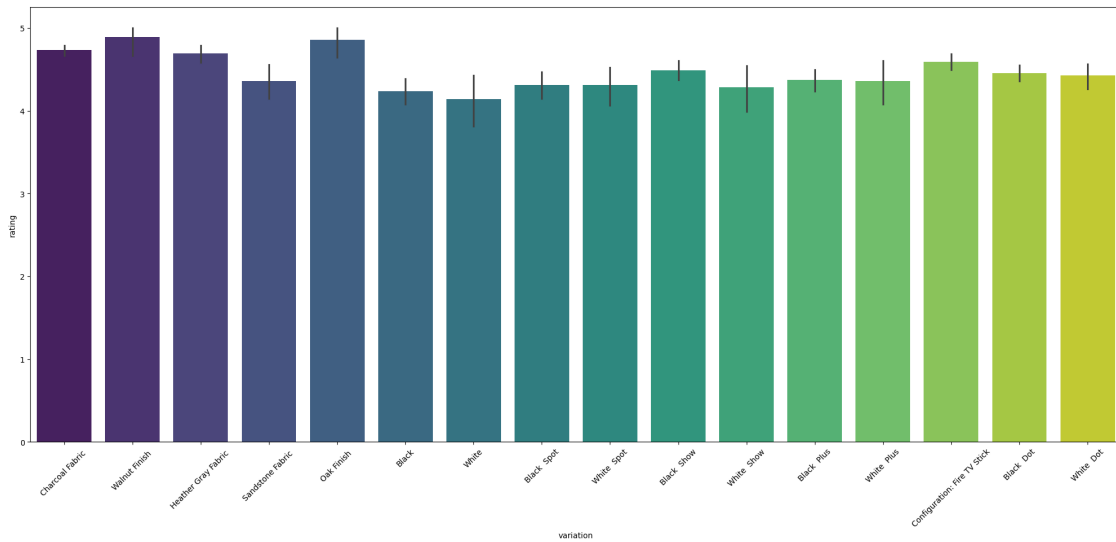


```
[ ]: plt.figure(figsize=(25,10))
sns.barplot(x='variation', y='rating', data=df, hue='variation',
           palette='viridis', legend=False) # Updated line
plt.xticks(rotation=45)
```

```
[ ]: ([0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15],
      [Text(0, 0, 'Charcoal Fabric '),
       Text(1, 0, 'Walnut Finish '),
       Text(2, 0, 'Heather Gray Fabric '),
       Text(3, 0, 'Sandstone Fabric '),
       Text(4, 0, 'Oak Finish '),
       Text(5, 0, 'Black'),
       Text(6, 0, 'White'),
       Text(7, 0, 'Black Spot'),
       Text(8, 0, 'White Spot'),
       Text(9, 0, 'Black Show'),
       Text(10, 0, 'White Show'),
       Text(11, 0, 'Black Plus'),
       Text(12, 0, 'White Plus'),
       Text(13, 0, 'Configuration: Fire TV Stick'),
       Text(14, 0, 'Black Dot')],
```



```
Text(15, 0, 'White Dot']])
```



```
[1]: !sudo apt-get update
      !sudo apt-get install texlive-xetex pandoc
```

```
0% [Working]                               Hit:1 http://archive.ubuntu.com/ubuntu jammy
InRelease
0% [Connecting to security.ubuntu.com (185.125.190.81)] [Connected to cloud.r-p
Get:2 http://archive.ubuntu.com/ubuntu jammy-updates InRelease [128 kB]
0% [2 InRelease 15.6 kB/128 kB 12%] [Waiting for headers] [Connected to cloud.r
Get:3 https://cloud.r-project.org/bin/linux/ubuntu jammy-cran40/ InRelease
[3,632 B]
0% [2 InRelease 107 kB/128 kB 83%] [Waiting for headers] [3 InRelease 0 B/3,632
0% [2 InRelease 110 kB/128 kB 86%] [Waiting for headers] [Waiting for headers]
0% [Waiting for headers] [Waiting for headers] [Connected to developer.download
Get:4 http://archive.ubuntu.com/ubuntu jammy-backports InRelease [127 kB]
Get:5 http://security.ubuntu.com/ubuntu jammy-security InRelease [129 kB]
Get:6 https://r2u.stat.illinois.edu/ubuntu jammy InRelease [6,555 B]
Get:7 https://developer.download.nvidia.com/compute/cuda/repos/ubuntu2204/x86_64
InRelease [1,581 B]
Hit:8 https://ppa.launchpadcontent.net/deadsnakes/ppa/ubuntu jammy InRelease
Hit:9 https://ppa.launchpadcontent.net/graphics-drivers/ppa/ubuntu jammy
InRelease
Get:10 https://cloud.r-project.org/bin/linux/ubuntu jammy-cran40/ Packages [69.9
kB]
Hit:11 https://ppa.launchpadcontent.net/ubuntugis/ppa/ubuntu jammy InRelease
Get:12 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 Packages [3,000
kB]
Get:13 http://archive.ubuntu.com/ubuntu jammy-updates/universe amd64 Packages
```

```

[1,535 kB]
Get:14 https://r2u.stat.illinois.edu/ubuntu jammy/main all Packages [8,754 kB]
Get:15
https://developer.download.nvidia.com/compute/cuda/repos/ubuntu2204/x86_64
Packages [1,378 kB]
Get:16 https://r2u.stat.illinois.edu/ubuntu jammy/main amd64 Packages [2,675 kB]
Get:17 http://security.ubuntu.com/ubuntu jammy-security/main amd64 Packages
[2,692 kB]
Get:18 http://security.ubuntu.com/ubuntu jammy-security/universe amd64 Packages
[1,236 kB]
Fetched 21.7 MB in 8s (2,673 kB/s)
Reading package lists... Done
W: Skipping acquire of configured file 'main/source/Sources' as repository
'https://r2u.stat.illinois.edu/ubuntu jammy InRelease' does not seem to provide
it (sources.list entry misspelt?)
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
The following additional packages will be installed:
  dvisvgm fonts-droid-fallback fonts-lato fonts-lmodern fonts-noto-mono
  fonts-texgyre fonts-urw-base35 libapache-pom-java
  libcmark-gfm-extensions0.29.0.gfm.3 libcmark-gfm0.29.0.gfm.3
  libcommons-logging-java libcommons-parent-java libfontbox-java libfontenc1
  libgs9 libgs9-common libidn12 libijs-0.35 libjbig2dec0 libkpathsea6
  libpdfbox-java libptexenc1 libruby3.0 libsynchronet2 libteckit0 libtexlua53
  libtexluaajit2 libwoff1 libzip-0-13 lmodern pandoc-data poppler-data
  preview-latex-style rake ruby ruby-net-telnet ruby-rubygems ruby-webrick
  ruby-xmlrpc ruby3.0 rubygems-integration tclutils teckit tex-common tex-gyre
  texlive-base texlive-binaries texlive-fonts-recommended texlive-latex-base
  texlive-latex-extra texlive-latex-recommended texlive-pictures
  texlive-plain-generic tipa xfonts-encodings xfonts-utils
Suggested packages:
  fonts-noto fonts-freefont-otf | fonts-freefont-ttf libavalon-framework-java
  libcommons-logging-java-doc libexcalibur-logkit-java liblog4j1.2-java
  texlive-luatex pandoc-citeproc context wkhtmltopdf librsvg2-bin groff ghc
  nodejs php python libjs-mathjax libjs-katex citation-style-language-styles
  poppler-utils ghostscript fonts-japanese-mincho | fonts-ipafont-mincho
  fonts-japanese-gothic | fonts-ipafont-gothic fonts-arphic-ukai
  fonts-arphic-uming fonts-nanum ri ruby-dev bundler debhelper gv
  | postscript-viewer perl-tk xpdf | pdf-viewer xzdec
  texlive-fonts-recommended-doc texlive-latex-base-doc python3-pygments
  icc-profiles libfile-which-perl libspreadsheet-parseexcel-perl
  texlive-latex-extra-doc texlive-latex-recommended-doc texlive-pstricks
  dot2tex prerex texlive-pictures-doc vprerex default-jre-headless tipa-doc
The following NEW packages will be installed:
  dvisvgm fonts-droid-fallback fonts-lato fonts-lmodern fonts-noto-mono
  fonts-texgyre fonts-urw-base35 libapache-pom-java
  libcmark-gfm-extensions0.29.0.gfm.3 libcmark-gfm0.29.0.gfm.3

```

```

libcommons-logging-java libcommons-parent-java libfontbox-java libfontenc1
libgs9 libgs9-common libidn12 libijs-0.35 libjbig2dec0 libkpathsea6
libpdfbox-java libptexenc1 libruby3.0 libsynchronet2 libteckit0 libtexlua53
libtexluajit2 libwoff1 libzip-0-13 lmodern pandoc pandoc-data poppler-data
preview-latex-style rake ruby ruby-net-telnet ruby-rubygems ruby-webrick
ruby-xmlrpc ruby3.0 rubygems-integration t1utils teckit tex-common tex-gyre
texlive-base texlive-binaries texlive-fonts-recommended texlive-latex-base
texlive-latex-extra texlive-latex-recommended texlive-pictures
texlive-plain-generic texlive-xetex tipa xfonts-encodings xfonts-utils
0 upgraded, 58 newly installed, 0 to remove and 35 not upgraded.
Need to get 202 MB of archives.
After this operation, 728 MB of additional disk space will be used.
Get:1 http://archive.ubuntu.com/ubuntu jammy/main amd64 fonts-droid-fallback all
1:6.0.1r16-1.1build1 [1,805 kB]
Get:2 http://archive.ubuntu.com/ubuntu jammy/main amd64 fonts-lato all 2.0-2.1
[2,696 kB]
Get:3 http://archive.ubuntu.com/ubuntu jammy/main amd64 poppler-data all
0.4.11-1 [2,171 kB]
Get:4 http://archive.ubuntu.com/ubuntu jammy/universe amd64 tex-common all 6.17
[33.7 kB]
Get:5 http://archive.ubuntu.com/ubuntu jammy/main amd64 fonts-urw-base35 all
20200910-1 [6,367 kB]
Get:6 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libgs9-common
all 9.55.0~dfsg1-0ubuntu5.10 [752 kB]
Get:7 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libidn12 amd64
1.38-4ubuntu1 [60.0 kB]
Get:8 http://archive.ubuntu.com/ubuntu jammy/main amd64 libijs-0.35 amd64
0.35-15build2 [16.5 kB]
Get:9 http://archive.ubuntu.com/ubuntu jammy/main amd64 libjbig2dec0 amd64
0.19-3build2 [64.7 kB]
Get:10 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libgs9 amd64
9.55.0~dfsg1-0ubuntu5.10 [5,031 kB]
Get:11 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libkpathsea6
amd64 2021.20210626.59705-1ubuntu0.2 [60.4 kB]
Get:12 http://archive.ubuntu.com/ubuntu jammy/main amd64 libwoff1 amd64
1.0.2-1build4 [45.2 kB]
Get:13 http://archive.ubuntu.com/ubuntu jammy/universe amd64 dvisvgm amd64
2.13.1-1 [1,221 kB]
Get:14 http://archive.ubuntu.com/ubuntu jammy/universe amd64 fonts-lmodern all
2.004.5-6.1 [4,532 kB]
Get:15 http://archive.ubuntu.com/ubuntu jammy/main amd64 fonts-noto-mono all
20201225-1build1 [397 kB]
Get:16 http://archive.ubuntu.com/ubuntu jammy/universe amd64 fonts-texgyre all
20180621-3.1 [10.2 MB]
Get:17 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libapache-pom-java
all 18-1 [4,720 B]
Get:18 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libcmark-
gfm0.29.0.gfm.3 amd64 0.29.0.gfm.3-3 [115 kB]

```

Get:19 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 libcmark-gfm-extensions0.29.0.gfm.3 amd64 0.29.0.gfm.3-3 [25.1 kB]
Get:20 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 libcommons-parent-java all 43-1 [10.8 kB]
Get:21 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 libcommons-logging-java all 1.2-2 [60.3 kB]
Get:22 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 libfontenc1 amd64 1:1.1.4-1build3 [14.7 kB]
Get:23 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 libptexenc1 amd64 2021.20210626.59705-1ubuntu0.2 [39.1 kB]
Get:24 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 rubygems-integration all 1.18 [5,336 B]
Get:25 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 ruby3.0 amd64 3.0.2-7ubuntu2.8 [50.1 kB]
Get:26 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 ruby-rubygems all 3.3.5-2 [228 kB]
Get:27 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 ruby amd64 1:3.0~exp1 [5,100 B]
Get:28 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 rake all 13.0.6-2 [61.7 kB]
Get:29 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 ruby-net-telnet all 0.1.1-2 [12.6 kB]
Get:30 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 ruby-webrick all 1.7.0-3ubuntu0.1 [52.1 kB]
Get:31 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 ruby-xmlrpc all 0.3.2-1ubuntu0.1 [24.9 kB]
Get:32 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 libruby3.0 amd64 3.0.2-7ubuntu2.8 [5,113 kB]
Get:33 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 libsynchronet2 amd64 2021.20210626.59705-1ubuntu0.2 [55.6 kB]
Get:34 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 libteckit0 amd64 2.5.11+ds1-1 [421 kB]
Get:35 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 libtexlua53 amd64 2021.20210626.59705-1ubuntu0.2 [120 kB]
Get:36 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 libtexluajit2 amd64 2021.20210626.59705-1ubuntu0.2 [267 kB]
Get:37 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 libzip-0-13 amd64 0.13.72+dfsg.1-1.1 [27.0 kB]
Get:38 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 xfonts-encodings all 1:1.0.5-0ubuntu2 [578 kB]
Get:39 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 xfonts-utils amd64 1:7.7+6build2 [94.6 kB]
Get:40 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 lmodern all 2.004.5-6.1 [9,471 kB]
Get:41 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 pandoc-data all 2.9.2.1-3ubuntu2 [81.8 kB]
Get:42 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 pandoc amd64 2.9.2.1-3ubuntu2 [20.3 MB]

```

Get:43 http://archive.ubuntu.com/ubuntu jammy/universe amd64 preview-latex-style
all 12.2-1ubuntu1 [185 kB]
Get:44 http://archive.ubuntu.com/ubuntu jammy/main amd64 t1utils amd64
1.41-4build2 [61.3 kB]
Get:45 http://archive.ubuntu.com/ubuntu jammy/universe amd64 teckit amd64
2.5.11+ds1-1 [699 kB]
Get:46 http://archive.ubuntu.com/ubuntu jammy/universe amd64 tex-gyre all
20180621-3.1 [6,209 kB]
Get:47 http://archive.ubuntu.com/ubuntu jammy-updates/universe amd64 texlive-
binaries amd64 2021.20210626.59705-1ubuntu0.2 [9,860 kB]
Get:48 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-base all
2021.20220204-1 [21.0 MB]
Get:49 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-fonts-
recommended all 2021.20220204-1 [4,972 kB]
Get:50 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-latex-base
all 2021.20220204-1 [1,128 kB]
Get:51 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libfontbox-java all
1:1.8.16-2 [207 kB]
Get:52 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libpdfbox-java all
1:1.8.16-2 [5,199 kB]
Get:53 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-latex-
recommended all 2021.20220204-1 [14.4 MB]
Get:54 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-pictures
all 2021.20220204-1 [8,720 kB]
Get:55 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-latex-extra
all 2021.20220204-1 [13.9 MB]
Get:56 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-plain-
generic all 2021.20220204-1 [27.5 MB]
Get:57 http://archive.ubuntu.com/ubuntu jammy/universe amd64 tipa all 2:1.3-21
[2,967 kB]
Get:58 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-xetex all
2021.20220204-1 [12.4 MB]
Fetched 202 MB in 3s (62.8 MB/s)
debconf: unable to initialize frontend: Dialog
debconf: (No usable dialog-like program is installed, so the dialog based
frontend cannot be used. at /usr/share/perl5/Debconf/FrontEnd/Dialog.pm line 78,
<> line 58.)
debconf: falling back to frontend: Readline
debconf: unable to initialize frontend: Readline
debconf: (This frontend requires a controlling tty.)
debconf: falling back to frontend: Teletype
dpkg-preconfigure: unable to re-open stdin:
Selecting previously unselected package fonts-droid-fallback.
(Reading database ... 125044 files and directories currently installed.)
Preparing to unpack .../00-fonts-droid-fallback_1%3a6.0.1r16-1.1build1_all.deb
...
Unpacking fonts-droid-fallback (1:6.0.1r16-1.1build1) ...
Selecting previously unselected package fonts-lato.

```

```

Preparing to unpack .../01-fonts-lato_2.0-2.1_all.deb ...
Unpacking fonts-lato (2.0-2.1) ...
Selecting previously unselected package poppler-data.
Preparing to unpack .../02-poppler-data_0.4.11-1_all.deb ...
Unpacking poppler-data (0.4.11-1) ...
Selecting previously unselected package tex-common.
Preparing to unpack .../03-tex-common_6.17_all.deb ...
Unpacking tex-common (6.17) ...
Selecting previously unselected package fonts-urw-base35.
Preparing to unpack .../04-fonts-urw-base35_20200910-1_all.deb ...
Unpacking fonts-urw-base35 (20200910-1) ...
Selecting previously unselected package libgs9-common.
Preparing to unpack .../05-libgs9-common_9.55.0~dfsg1-0ubuntu5.10_all.deb ...
Unpacking libgs9-common (9.55.0~dfsg1-0ubuntu5.10) ...
Selecting previously unselected package libidn12:amd64.
Preparing to unpack .../06-libidn12_1.38-4ubuntu1_amd64.deb ...
Unpacking libidn12:amd64 (1.38-4ubuntu1) ...
Selecting previously unselected package libijs-0.35:amd64.
Preparing to unpack .../07-libijs-0.35_0.35-15build2_amd64.deb ...
Unpacking libijs-0.35:amd64 (0.35-15build2) ...
Selecting previously unselected package libjbig2dec0:amd64.
Preparing to unpack .../08-libjbig2dec0_0.19-3build2_amd64.deb ...
Unpacking libjbig2dec0:amd64 (0.19-3build2) ...
Selecting previously unselected package libgs9:amd64.
Preparing to unpack .../09-libgs9_9.55.0~dfsg1-0ubuntu5.10_amd64.deb ...
Unpacking libgs9:amd64 (9.55.0~dfsg1-0ubuntu5.10) ...
Selecting previously unselected package libkpathsea6:amd64.
Preparing to unpack .../10-libkpathsea6_2021.20210626.59705-1ubuntu0.2_amd64.deb
...
Unpacking libkpathsea6:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Selecting previously unselected package libwoff1:amd64.
Preparing to unpack .../11-libwoff1_1.0.2-1build4_amd64.deb ...
Unpacking libwoff1:amd64 (1.0.2-1build4) ...
Selecting previously unselected package dvisvgm.
Preparing to unpack .../12-dvisvgm_2.13.1-1_amd64.deb ...
Unpacking dvisvgm (2.13.1-1) ...
Selecting previously unselected package fonts-lmodern.
Preparing to unpack .../13-fonts-lmodern_2.004.5-6.1_all.deb ...
Unpacking fonts-lmodern (2.004.5-6.1) ...
Selecting previously unselected package fonts-noto-mono.
Preparing to unpack .../14-fonts-noto-mono_20201225-1build1_all.deb ...
Unpacking fonts-noto-mono (20201225-1build1) ...
Selecting previously unselected package fonts-texgyre.
Preparing to unpack .../15-fonts-texgyre_20180621-3.1_all.deb ...
Unpacking fonts-texgyre (20180621-3.1) ...
Selecting previously unselected package libapache-pom-java.
Preparing to unpack .../16-libapache-pom-java_18-1_all.deb ...
Unpacking libapache-pom-java (18-1) ...

```

```

Selecting previously unselected package libcmark-gfm0.29.0.gfm.3:amd64.
Preparing to unpack .../17-libcmark-gfm0.29.0.gfm.3_0.29.0.gfm.3-3_amd64.deb ...
Unpacking libcmark-gfm0.29.0.gfm.3:amd64 (0.29.0.gfm.3-3) ...
Selecting previously unselected package libcmark-gfm-
extensions0.29.0.gfm.3:amd64.
Preparing to unpack .../18-libcmark-gfm-
extensions0.29.0.gfm.3_0.29.0.gfm.3-3_amd64.deb ...
Unpacking libcmark-gfm-extensions0.29.0.gfm.3:amd64 (0.29.0.gfm.3-3) ...
Selecting previously unselected package libcommons-parent-java.
Preparing to unpack .../19-libcommons-parent-java_43-1_all.deb ...
Unpacking libcommons-parent-java (43-1) ...
Selecting previously unselected package libcommons-logging-java.
Preparing to unpack .../20-libcommons-logging-java_1.2-2_all.deb ...
Unpacking libcommons-logging-java (1.2-2) ...
Selecting previously unselected package libfontenc1:amd64.
Preparing to unpack .../21-libfontenc1_1%3a1.1.4-1build3_amd64.deb ...
Unpacking libfontenc1:amd64 (1:1.1.4-1build3) ...
Selecting previously unselected package libptexenc1:amd64.
Preparing to unpack .../22-libptexenc1_2021.20210626.59705-1ubuntu0.2_amd64.deb
...
Unpacking libptexenc1:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Selecting previously unselected package rubygems-integration.
Preparing to unpack .../23-rubygems-integration_1.18_all.deb ...
Unpacking rubygems-integration (1.18) ...
Selecting previously unselected package ruby3.0.
Preparing to unpack .../24-ruby3.0_3.0.2-7ubuntu2.8_amd64.deb ...
Unpacking ruby3.0 (3.0.2-7ubuntu2.8) ...
Selecting previously unselected package ruby-rubygems.
Preparing to unpack .../25-ruby-rubygems_3.3.5-2_all.deb ...
Unpacking ruby-rubygems (3.3.5-2) ...
Selecting previously unselected package ruby.
Preparing to unpack .../26-ruby_1%3a3.0~exp1_amd64.deb ...
Unpacking ruby (1:3.0~exp1) ...
Selecting previously unselected package rake.
Preparing to unpack .../27-rake_13.0.6-2_all.deb ...
Unpacking rake (13.0.6-2) ...
Selecting previously unselected package ruby-net-telnet.
Preparing to unpack .../28-ruby-net-telnet_0.1.1-2_all.deb ...
Unpacking ruby-net-telnet (0.1.1-2) ...
Selecting previously unselected package ruby-webrick.
Preparing to unpack .../29-ruby-webrick_1.7.0-3ubuntu0.1_all.deb ...
Unpacking ruby-webrick (1.7.0-3ubuntu0.1) ...
Selecting previously unselected package ruby-xmlrpc.
Preparing to unpack .../30-ruby-xmlrpc_0.3.2-1ubuntu0.1_all.deb ...
Unpacking ruby-xmlrpc (0.3.2-1ubuntu0.1) ...
Selecting previously unselected package libruby3.0:amd64.
Preparing to unpack .../31-libruby3.0_3.0.2-7ubuntu2.8_amd64.deb ...
Unpacking libruby3.0:amd64 (3.0.2-7ubuntu2.8) ...

```

```

Selecting previously unselected package libsyntax2:amd64.
Preparing to unpack .../32-libsyntax2_2021.20210626.59705-1ubuntu0.2_amd64.deb
...
Unpacking libsyntax2:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Selecting previously unselected package libteckit0:amd64.
Preparing to unpack .../33-libteckit0_2.5.11+ds1-1_amd64.deb ...
Unpacking libteckit0:amd64 (2.5.11+ds1-1) ...
Selecting previously unselected package libtexlua53:amd64.
Preparing to unpack .../34-libtexlua53_2021.20210626.59705-1ubuntu0.2_amd64.deb
...
Unpacking libtexlua53:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Selecting previously unselected package libtexluaajit2:amd64.
Preparing to unpack
.../35-libtexluaajit2_2021.20210626.59705-1ubuntu0.2_amd64.deb ...
Unpacking libtexluaajit2:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Selecting previously unselected package libzip-0-13:amd64.
Preparing to unpack .../36-libzip-0-13_0.13.72+dfsg.1-1.1_amd64.deb ...
Unpacking libzip-0-13:amd64 (0.13.72+dfsg.1-1.1) ...
Selecting previously unselected package xfonts-encodings.
Preparing to unpack .../37-xfonts-encodings_1%3a1.0.5-0ubuntu2_all.deb ...
Unpacking xfonts-encodings (1:1.0.5-0ubuntu2) ...
Selecting previously unselected package xfonts-utils.
Preparing to unpack .../38-xfonts-utils_1%3a7.7+6build2_amd64.deb ...
Unpacking xfonts-utils (1:7.7+6build2) ...
Selecting previously unselected package lmodern.
Preparing to unpack .../39-lmodern_2.004.5-6.1_all.deb ...
Unpacking lmodern (2.004.5-6.1) ...
Selecting previously unselected package pandoc-data.
Preparing to unpack .../40-pandoc-data_2.9.2.1-3ubuntu2_all.deb ...
Unpacking pandoc-data (2.9.2.1-3ubuntu2) ...
Selecting previously unselected package pandoc.
Preparing to unpack .../41-pandoc_2.9.2.1-3ubuntu2_amd64.deb ...
Unpacking pandoc (2.9.2.1-3ubuntu2) ...
Selecting previously unselected package preview-latex-style.
Preparing to unpack .../42-preview-latex-style_12.2-1ubuntu1_all.deb ...
Unpacking preview-latex-style (12.2-1ubuntu1) ...
Selecting previously unselected package t1utils.
Preparing to unpack .../43-t1utils_1.41-4build2_amd64.deb ...
Unpacking t1utils (1.41-4build2) ...
Selecting previously unselected package teckit.
Preparing to unpack .../44-teckit_2.5.11+ds1-1_amd64.deb ...
Unpacking teckit (2.5.11+ds1-1) ...
Selecting previously unselected package tex-gyre.
Preparing to unpack .../45-tex-gyre_20180621-3.1_all.deb ...
Unpacking tex-gyre (20180621-3.1) ...
Selecting previously unselected package texlive-binaries.
Preparing to unpack .../46-texlive-
binaries_2021.20210626.59705-1ubuntu0.2_amd64.deb ...

```



```

Unpacking texlive-binaries (2021.20210626.59705-1ubuntu0.2) ...
Selecting previously unselected package texlive-base.
Preparing to unpack .../47-texlive-base_2021.20220204-1_all.deb ...
Unpacking texlive-base (2021.20220204-1) ...
Selecting previously unselected package texlive-fonts-recommended.
Preparing to unpack .../48-texlive-fonts-recommended_2021.20220204-1_all.deb ...
Unpacking texlive-fonts-recommended (2021.20220204-1) ...
Selecting previously unselected package texlive-latex-base.
Preparing to unpack .../49-texlive-latex-base_2021.20220204-1_all.deb ...
Unpacking texlive-latex-base (2021.20220204-1) ...
Selecting previously unselected package libfontbox-java.
Preparing to unpack .../50-libfontbox-java_1%3a1.8.16-2_all.deb ...
Unpacking libfontbox-java (1:1.8.16-2) ...
Selecting previously unselected package libpdfbox-java.
Preparing to unpack .../51-libpdfbox-java_1%3a1.8.16-2_all.deb ...
Unpacking libpdfbox-java (1:1.8.16-2) ...
Selecting previously unselected package texlive-latex-recommended.
Preparing to unpack .../52-texlive-latex-recommended_2021.20220204-1_all.deb ...
Unpacking texlive-latex-recommended (2021.20220204-1) ...
Selecting previously unselected package texlive-pictures.
Preparing to unpack .../53-texlive-pictures_2021.20220204-1_all.deb ...
Unpacking texlive-pictures (2021.20220204-1) ...
Selecting previously unselected package texlive-latex-extra.
Preparing to unpack .../54-texlive-latex-extra_2021.20220204-1_all.deb ...
Unpacking texlive-latex-extra (2021.20220204-1) ...
Selecting previously unselected package texlive-plain-generic.
Preparing to unpack .../55-texlive-plain-generic_2021.20220204-1_all.deb ...
Unpacking texlive-plain-generic (2021.20220204-1) ...
Selecting previously unselected package tipa.
Preparing to unpack .../56-tipa_2%3a1.3-21_all.deb ...
Unpacking tipa (2:1.3-21) ...
Selecting previously unselected package texlive-xetex.
Preparing to unpack .../57-texlive-xetex_2021.20220204-1_all.deb ...
Unpacking texlive-xetex (2021.20220204-1) ...
Setting up fonts-lato (2.0-2.1) ...
Setting up fonts-noto-mono (20201225-1build1) ...
Setting up libwoff1:amd64 (1.0.2-1build4) ...
Setting up libtexlua53:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Setting up libijs-0.35:amd64 (0.35-15build2) ...
Setting up libtexluaajit2:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Setting up libfontbox-java (1:1.8.16-2) ...
Setting up rubygems-integration (1.18) ...
Setting up libzip-0-13:amd64 (0.13.72+dfsg.1-1.1) ...
Setting up fonts-urw-base35 (20200910-1) ...
Setting up poppler-data (0.4.11-1) ...
Setting up tex-common (6.17) ...
debconf: unable to initialize frontend: Dialog
debconf: (No usable dialog-like program is installed, so the dialog based

```

```

frontend cannot be used. at /usr/share/perl5/Debconf/FrontEnd/Dialog.pm line
78.)
debconf: falling back to frontend: Readline
update-language: texlive-base not installed and configured, doing nothing!
Setting up libfontenc1:amd64 (1:1.1.4-1build3) ...
Setting up libjbig2dec0:amd64 (0.19-3build2) ...
Setting up libteckit0:amd64 (2.5.11+ds1-1) ...
Setting up libapache-pom-java (18-1) ...
Setting up ruby-net-telnet (0.1.1-2) ...
Setting up xfonts-encodings (1:1.0.5-0ubuntu2) ...
Setting up t1utils (1.41-4build2) ...
Setting up libidn12:amd64 (1.38-4ubuntu1) ...
Setting up fonts-texgyre (20180621-3.1) ...
Setting up libkpathsea6:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Setting up ruby-webrick (1.7.0-3ubuntu0.1) ...
Setting up libcmark-gfm0.29.0.gfm.3:amd64 (0.29.0.gfm.3-3) ...
Setting up fonts-lmodern (2.004.5-6.1) ...
Setting up libcmark-gfm-extensions0.29.0.gfm.3:amd64 (0.29.0.gfm.3-3) ...
Setting up fonts-droid-fallback (1:6.0.1r16-1.1build1) ...
Setting up pandoc-data (2.9.2.1-3ubuntu2) ...
Setting up ruby-xmlrpc (0.3.2-1ubuntu0.1) ...
Setting up libsynchronet2:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Setting up libgs9-common (9.55.0~dfsg1-0ubuntu5.10) ...
Setting up teckit (2.5.11+ds1-1) ...
Setting up libpdfbox-java (1:1.8.16-2) ...
Setting up libgs9:amd64 (9.55.0~dfsg1-0ubuntu5.10) ...
Setting up preview-latex-style (12.2-1ubuntu1) ...
Setting up libcommons-parent-java (43-1) ...
Setting up dvisvgm (2.13.1-1) ...
Setting up libcommons-logging-java (1.2-2) ...
Setting up xfonts-utils (1:7.7+6build2) ...
Setting up libptexenc1:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Setting up pandoc (2.9.2.1-3ubuntu2) ...
Setting up texlive-binaries (2021.20210626.59705-1ubuntu0.2) ...
update-alternatives: using /usr/bin/xdvi-xaw to provide /usr/bin/xdvi.bin
(xdvi.bin) in auto mode
update-alternatives: using /usr/bin/bibtex.original to provide /usr/bin/bibtex
(bibtex) in auto mode
Setting up lmodern (2.004.5-6.1) ...
Setting up texlive-base (2021.20220204-1) ...
/usr/bin/ucfr
/usr/bin/ucfr
/usr/bin/ucfr
/usr/bin/ucfr
mktexlsr: Updating /var/lib/texmf/ls-R-TEXLIVEDIST...
mktexlsr: Updating /var/lib/texmf/ls-R-TEXMFMAIN...
mktexlsr: Updating /var/lib/texmf/ls-R...
mktexlsr: Done.

```

```

tl-paper: setting paper size for dvips to a4:
/var/lib/texmf/dvips/config/config-paper.ps
tl-paper: setting paper size for dvipdfmx to a4:
/var/lib/texmf/dvipdfmx/dvipdfmx-paper.cfg
tl-paper: setting paper size for xdvi to a4: /var/lib/texmf/xdvi/XDvi-paper
tl-paper: setting paper size for pdftex to a4: /var/lib/texmf/tex/generic/tex-
ini-files/pdftexconfig.tex
debconf: unable to initialize frontend: Dialog
debconf: (No usable dialog-like program is installed, so the dialog based
frontend cannot be used. at /usr/share/perl5/Debconf/FrontEnd/Dialog.pm line
78.)
debconf: falling back to frontend: Readline
Setting up tex-gyre (20180621-3.1) ...
Setting up texlive-plain-generic (2021.20220204-1) ...
Setting up texlive-latex-base (2021.20220204-1) ...
Setting up texlive-latex-recommended (2021.20220204-1) ...
Setting up texlive-pictures (2021.20220204-1) ...
Setting up texlive-fonts-recommended (2021.20220204-1) ...
Setting up tipa (2:1.3-21) ...
Setting up texlive-latex-extra (2021.20220204-1) ...
Setting up texlive-xetex (2021.20220204-1) ...
Setting up rake (13.0.6-2) ...
Setting up libruby3.0:amd64 (3.0.2-7ubuntu2.8) ...
Setting up ruby3.0 (3.0.2-7ubuntu2.8) ...
Setting up ruby (1:3.0~exp1) ...
Setting up ruby-rubygems (3.3.5-2) ...
Processing triggers for man-db (2.10.2-1) ...
Processing triggers for mailcap (3.70+nmulubuntu1) ...
Processing triggers for fontconfig (2.13.1-4.2ubuntu5) ...
Processing triggers for libc-bin (2.35-0ubuntu3.8) ...
/sbin/ldconfig.real: /usr/local/lib/libtbb.so.12 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbbind_2_5.so.3 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbbind.so.3 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libur_adapter_opencl.so.0 is not a symbolic
link

/sbin/ldconfig.real: /usr/local/lib/libtbbbind_2_0.so.3 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbmalloc.so.2 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libumf.so.0 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libur_loader.so.0 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtcm.so.1 is not a symbolic link

```

```
/sbin/ldconfig.real: /usr/local/lib/libhwloc.so.15 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtcm_debug.so.1 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbmalloc_proxy.so.2 is not a symbolic
link

/sbin/ldconfig.real: /usr/local/lib/libur_adapter_level_zero.so.0 is not a
symbolic link

Processing triggers for tex-common (6.17) ...
debconf: unable to initialize frontend: Dialog
debconf: (No usable dialog-like program is installed, so the dialog based
frontend cannot be used. at /usr/share/perl5/Debconf/FrontEnd/Dialog.pm line
78.)
debconf: falling back to frontend: Readline
Running updmap-sys. This may take some time... done.
Running mktexlsr /var/lib/texmf ... done.
Building format(s) --all.
    This may take some time... done.
```