BookReview_EDA

December 4, 2020

[ ]:
```
! pip install --upgrade pandas-profiling
```

[12]:
```
import gdown

url = 'https://drive.google.com/uc?id=1UPZiTughL3iDtPwreoUs_SX-LfVktrI3'
output = 'BX-CSV-Dump.zip'
gdown.download(url, output, quiet=False)
```

```
Downloading...
From: https://drive.google.com/uc?id=1UPZiTughL3iDtPwreoUs_SX-LfVktrI3
To: /content/BX-CSV-Dump.zip
26.1MB [00:00, 159MB/s]
```

[12]:
```
'BX-CSV-Dump.zip'
```

[ ]:
```
! unzip BX-CSV-Dump.zip
```

```
Archive:  BX-CSV-Dump.zip
```

1 Book Crossing EDA

[388]:
```
import pandas_profiling
import scipy
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np

from sklearn.preprocessing import StandardScaler, MinMaxScaler

sns.set()
palette = sns.color_palette("icefire")

plt.style.use('ggplot')

sns.set_context("talk")
```

1.1 Loading, Cleaning and Merging the Dataset

Reading the csv files

[259]:
```
users = pd.read_csv(
    '/content/BX-Users.csv',
    names=['user_id', 'location', 'age'],
    sep=';',
    skiprows=1,
    encoding='ISO-8859-1',
    low_memory=False,
    error_bad_lines=False
)
users
```

[259]:
```
        user_id                            location   age
0             1                  nyc, new york, usa   NaN
1             2           stockton, california, usa  18.0
2             3     moscow, yukon territory, russia   NaN
3             4           porto, v.n.gaia, portugal  17.0
4             5  farnborough, hants, united kingdom   NaN
...         ...                                 ...   ...
278853   278854              portland, oregon, usa   NaN
278854   278855  tacoma, washington, united kingdom  50.0
278855   278856           brampton, ontario, canada   NaN
278856   278857           knoxville, tennessee, usa   NaN
278857   278858                dublin, n/a, ireland   NaN

[278858 rows x 3 columns]
```

parse the datatypes properly

[260]:
```
users.dtypes
```

[260]:
```
user_id       int64
location     object
age         float64
dtype: object
```

[261]:
```
users.describe().T
```

[261]:
```
            count           mean           std  ...       50%        75%
max
user_id  278858.0  139429.500000  80499.515020  ...  139429.5  209143.75
278858.0
age      168096.0      34.751434     14.428097  ...      32.0      44.00
244.0

[2 rows x 8 columns]
```

age cannot be 244 ! so let's fix that

[262]:
```
users.loc[(users.age > 100) | (users.age < 5), 'age'] = np.nan
users.age = users.age.fillna(users.age.mean())
```

[263]:
```
users['age'] = users['age'].astype(np.uint8)
```

[264]:
```
users['age'].describe()
```

[264]:
```
count    278858.000000
mean         34.446733
std          10.551712
min           5.000000
25%          29.000000
50%          34.000000
75%          35.000000
max         100.000000
Name: age, dtype: float64
```

[265]:
```
users.isna().sum()
```

[265]:
```
user_id     0
location    0
age         0
dtype: int64
```

[266]:
```
books = pd.read_csv(
    '/content/BX-Books.csv',
    names=['isbn', 'book_title', 'book_author', 'year_of_publication',␣
 ↪'publisher', 'img_s', 'img_m', 'img_l'],
    sep=';',
    skiprows=1,
    encoding='ISO-8859-1',
    low_memory=False,
    error_bad_lines=False
)
books
```

[266]:
```
              isbn  ...                                              img_l
0       0195153448  ...  http://images.amazon.com/images/P/0195153448.0...
1       0002005018  ...  http://images.amazon.com/images/P/0002005018.0...
2       0060973129  ...  http://images.amazon.com/images/P/0060973129.0...
3       0374157065  ...  http://images.amazon.com/images/P/0374157065.0...
4       0393045218  ...  http://images.amazon.com/images/P/0393045218.0...
...            ...  ...                                                ...
271374  0440400988  ...  http://images.amazon.com/images/P/0440400988.0...
271375  0525447644  ...  http://images.amazon.com/images/P/0525447644.0...
271376  006008667X  ...  http://images.amazon.com/images/P/006008667X.0...
271377  0192126040  ...  http://images.amazon.com/images/P/0192126040.0...
271378  0767409752  ...  http://images.amazon.com/images/P/0767409752.0...

[271379 rows x 8 columns]
```

parse the data types properly

[267]:
```
books.dtypes
```

[267]:
```
isbn                 object
book_title           object
book_author          object
year_of_publication  object
publisher            object
img_s                object
img_m                object
img_l                object
dtype: object
```

drop ['img_s', 'img_m', 'img_l'] since they are not useful for us

[268]:
```
books = books.drop(['img_s', 'img_m', 'img_l'], axis=1)
```

year_of_publication should be a integer

[269]:
```
books['year_of_publication'] = pd.to_numeric(books['year_of_publication'],␣
↪errors='coerce')
```

[270]:
```
books.loc[(books['year_of_publication'] == 0) | (books['year_of_publication'] >␣
↪2008), 'year_of_publication' ] = np.nan
books['year_of_publication'] = books['year_of_publication'].
↪fillna(round(books['year_of_publication'].mean()))
books['year_of_publication'] = pd.to_numeric(books['year_of_publication'],␣
↪downcast='unsigned')
```

[271]:
```
books.isna().sum()
```

[271]:
```
isbn                 0
book_title           0
book_author          1
year_of_publication  0
publisher            2
dtype: int64
```

[272]:
```
books = books.dropna()
```

[273]:
```
books.describe().T
```

[273]:
```
                      count         mean       std  ...     50%     75%
max
year_of_publication  271376.0  1993.692427  8.248715  ...  1995.0  2000.0
2008.0

[1 rows x 8 columns]
```

[274]:
```
ratings = pd.read_csv(
    '/content/BX-Book-Ratings.csv',
    names=['user_id', 'isbn', 'book_rating'],
    sep=';',
```

```
    skiprows=1,
    encoding='ISO-8859-1',
    low_memory=False,
    error_bad_lines=False
)
ratings
```

[274]:

	user_id	isbn	book_rating
0	276725	034545104X	0
1	276726	0155061224	5
2	276727	0446520802	0
3	276729	052165615X	3
4	276729	0521795028	6
...	...	...	...
1149775	276704	1563526298	9
1149776	276706	0679447156	0
1149777	276709	0515107662	10
1149778	276721	0590442449	10
1149779	276723	05162443314	8

[1149780 rows x 3 columns]

[275]:
```
ratings['book_rating'] = ratings['book_rating'].astype(np.uint8)
```

[276]:
```
ratings.dtypes
```

[276]:
```
user_id         int64
isbn           object
book_rating     uint8
dtype: object
```

[277]:
```
ratings.isna().sum()
```

[277]:
```
user_id        0
isbn           0
book_rating    0
dtype: int64
```

[278]:
```
ratings.describe().T.astype(np.int32)
```

[278]:

	count	mean	std	min	25%	50%	75%	max
user_id	1149780	140386	80562	2	70345	141010	211028	278854
book_rating	1149780	2	3	0	0	0	7	10

Join the three datasets based on `user_id` and `isbn`

[279]:
```
temp = pd.merge(users, ratings, on='user_id')
temp = pd.merge(temp, books, on='isbn')
dataset = temp.copy()
```

[280]:
```
dataset
```

[280]:
```
         user_id  ...                 publisher
0              2  ...  Oxford University Press
1              8  ...    HarperFlamingo Canada
2          11400  ...    HarperFlamingo Canada
3          11676  ...    HarperFlamingo Canada
4          41385  ...    HarperFlamingo Canada
...          ...  ...                       ...
1031167   278851  ...     Simon &amp;amp; Schuster
1031168   278851  ...           Broadway Books
1031169   278851  ...          Lone Star Books
1031170   278851  ...               Kqed Books
1031171   278851  ...  American Map Corporation

[1031172 rows x 9 columns]
```

Split the location into `city`, `state` and `country` and replacing missing location details with just `n/a`

[281]:
```
location = dataset['location'].str.split(', ', n=2, expand=True)
location.columns = ['city', 'state', 'country']
location = location.fillna('n/a')
```

[282]:
```
dataset['city'] = location['city'] ; dataset['state'] = location['state'] ;␣
↪dataset['country'] = location['country']
```

[283]:
```
dataset = dataset.drop(['location'], axis=1)
```

[284]:
```
dataset.describe().T.astype(np.int32)
```

[284]:
```
                      count    mean    std  ...     50%     75%     max
user_id             1031172  140594  80524  ...  141210  211426  278854
age                 1031172      36     10  ...      34      41     100
book_rating         1031172       2      3  ...       0       7      10
year_of_publication 1031172    1995      7  ...    1997    2001    2008

[4 rows x 8 columns]
```

[285]:
```
dataset.isna().sum()
```

[285]:
```
user_id               0
age                   0
isbn                  0
book_rating           0
book_title            0
book_author           0
year_of_publication   0
publisher             0
city                  0
state                 0
country               0
dtype: int64
```

[286]:
```
# profile = pandas_profiling.ProfileReport(dataset)
# profile
```

[287]:
```
dataset.shape
```

[287]:
```
(1031172, 11)
```

[288]:
```
dataset.dtypes
```

[288]:
```
user_id                int64
age                    uint8
isbn                  object
book_rating            uint8
book_title            object
book_author           object
year_of_publication   uint16
publisher             object
city                  object
state                 object
country               object
dtype: object
```

This will be the final dataset we will be working with !

[289]:
```
dataset.head(5)
```

[289]:

	user_id	age	isbn	...	city	state	country
0	2	18	0195153448	...	stockton	california	usa
1	8	34	0002005018	...	timmins	ontario	canada
2	11400	49	0002005018	...	ottawa	ontario	canada
3	11676	34	0002005018	...	n/a	n/a	n/a
4	41385	34	0002005018	...	sudbury	ontario	canada

[5 rows x 11 columns]

[290]:
```
dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1031172 entries, 0 to 1031171
Data columns (total 11 columns):
 #   Column               Non-Null Count    Dtype
---  ------               --------------    -----
 0   user_id              1031172 non-null  int64
 1   age                  1031172 non-null  uint8
 2   isbn                 1031172 non-null  object
 3   book_rating          1031172 non-null  uint8
 4   book_title           1031172 non-null  object
 5   book_author          1031172 non-null  object
 6   year_of_publication  1031172 non-null  uint16
 7   publisher            1031172 non-null  object
 8   city                 1031172 non-null  object
 9   state                1031172 non-null  object
```

```
 10  country              1031172 non-null  object
dtypes: int64(1), object(7), uint16(1), uint8(2)
memory usage: 74.7+ MB
```

[291]:
```
# cleaned_data = dataset.copy()
```

[365]:
```
# dataset = cleaned_data.copy()
```

1.2 Analyzing the Feature Space

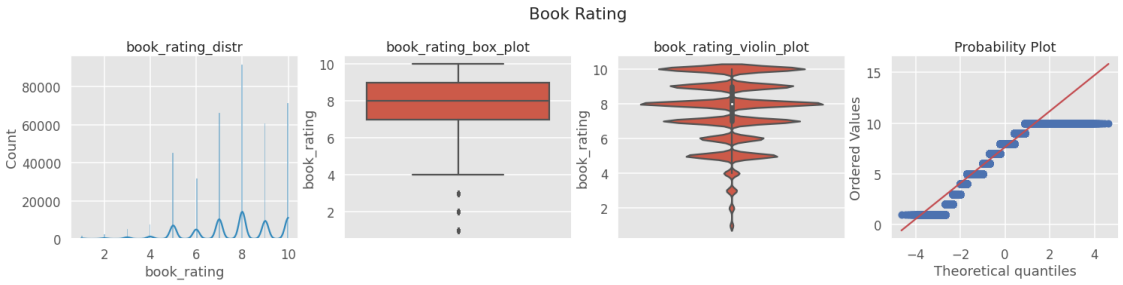
[373]:
```
f, axes = plt.subplots(ncols = 4, figsize=(25, 5))
sns.histplot(x="book_rating", data=dataset, kde=True, ax=axes[0]).
 ↪set_title('rating_distr')
sns.histplot(x="book_rating", data=dataset[dataset['book_rating'] != 0],␣
 ↪kde=True, ax=axes[1]).set_title('explicit rating_distr')
sns.boxplot(y="book_rating", data=dataset, orient='v', ax=axes[2]).
 ↪set_title('rating_distr')
sns.boxplot(y="book_rating", data=dataset[dataset['book_rating'] != 0],␣
 ↪orient='v', ax=axes[3]).set_title('explicit rating_distr')
plt.show()
```

We can remove 0 ratings, since these are unrated, and why would someone rate a book as 0 ?

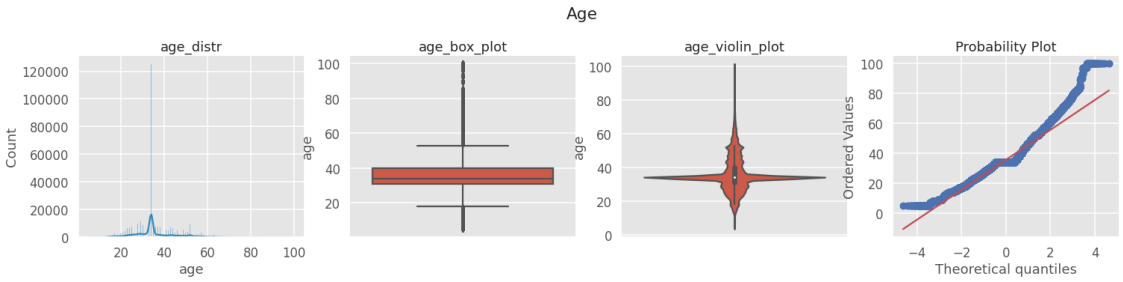
[374]:
```
dataset = dataset[dataset['book_rating'] != 0]
```

[450]:
```
def plot_univariate(dataset, column_name, suptitle = None):
    f, axes = plt.subplots(ncols = 4, figsize=(25, 5))
    sns.histplot(x=column_name, data=dataset, kde=True, ax=axes[0]).
 ↪set_title(f'{column_name}_distr')
    sns.boxplot(y=column_name, data=dataset, orient='v', ax=axes[1]).
 ↪set_title(f'{column_name}_box_plot')
    sns.violinplot(y=column_name, data=dataset, orient='v', ax=axes[2]).
 ↪set_title(f'{column_name}_violin_plot')
    scipy.stats.probplot(dataset[column_name], dist="norm", plot=axes[3])
    if suptitle:
        plt.suptitle(suptitle)
        plt.subplots_adjust(top=0.80)
    plt.show()
```

[451]:
```
plot_univariate(dataset=dataset, column_name='book_rating', suptitle='Book␣
↪Rating')
```



[452]:
```
plot_univariate(dataset=dataset, column_name='age', suptitle='Age')
```



1.3 Data Transformation

1.3.1 Min-Max Normalization

$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}$

[453]:
```
scaler = MinMaxScaler()
scaled = scaler.fit_transform(dataset['book_rating'].values.reshape(-1, 1)).
↪reshape(-1)
```

[454]:
```
scaled_rating = pd.DataFrame(data=scaled, columns=['book_rating'])
scaled_rating
```

[454]:

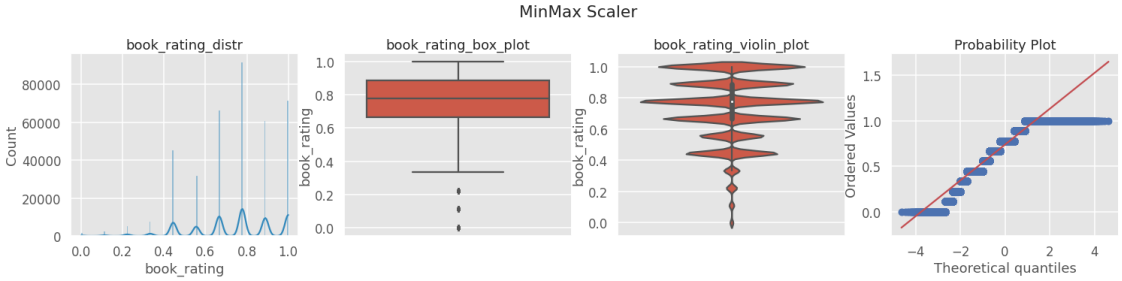
	book_rating
0	0.444444
1	0.777778
2	0.777778
3	0.888889
4	0.888889
...	...
383844	0.666667
383845	0.444444

```
383846     0.666667
383847     0.666667
383848     1.000000

[383849 rows x 1 columns]
```

[455]:
```
scaled_rating.describe().T
```

[455]:
```
                count     mean       std  min       25%       50%       75%  max
book_rating  383849.0   0.7363  0.204593  0.0  0.666667  0.777778  0.888889  1.0
```

[456]:
```
plot_univariate(dataset=scaled_rating, column_name='book_rating',␣
↪suptitle='MinMax Scaler')
```



1.3.2 Z-Score Standardization

$z = \frac{x-\mu}{\sigma}$

[457]:
```
scaler = StandardScaler()
scaled = scaler.fit_transform(dataset['book_rating'].values.reshape(-1, 1)).
↪reshape(-1)
```

[458]:
```
scaled_rating = pd.DataFrame(data=scaled, columns=['book_rating'])
scaled_rating
```

[458]:
```
        book_rating
0         -1.426522
1          0.202732
2          0.202732
3          0.745817
4          0.745817
...             ...
383844    -0.340353
383845    -1.426522
383846    -0.340353
383847    -0.340353
383848     1.288902

[383849 rows x 1 columns]
```

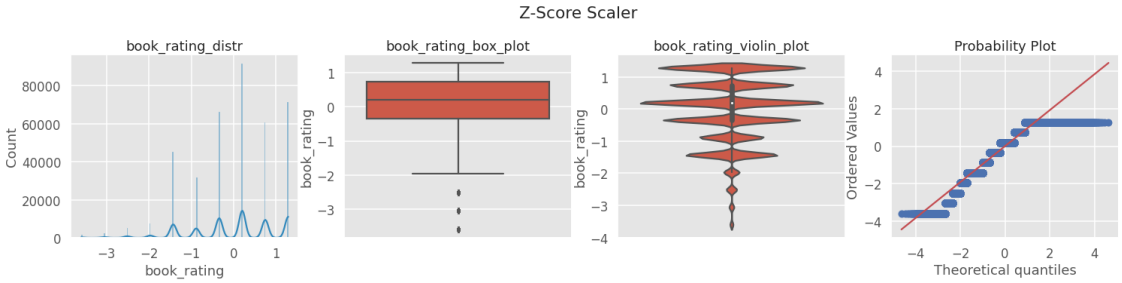
[459]:
```
scaled_rating.describe().T
```

[459]:

	count	mean	std	...	50%	75%	max
book_rating	383849.0	2.953229e-14	1.000001	...	0.202732	0.745817	1.288902

[1 rows x 8 columns]

[460]:
```
plot_univariate(dataset=scaled_rating, column_name='book_rating',␣
↪suptitle='Z-Score Scaler')
```



1.3.3 Decimal Scaling

$v_i' = \frac{v_i}{10^j}$

[461]:
```
p = dataset['book_rating'].max()
q = len(str(abs(p)))
scaled = dataset['book_rating'].values / 10 ** q
```

[462]:
```
scaled_rating = pd.DataFrame(data=scaled, columns=['book_rating'])
scaled_rating
```

[462]:

	book_rating
0	0.05
1	0.08
2	0.08
3	0.09
4	0.09
...	...
383844	0.07
383845	0.05
383846	0.07
383847	0.07
383848	0.10

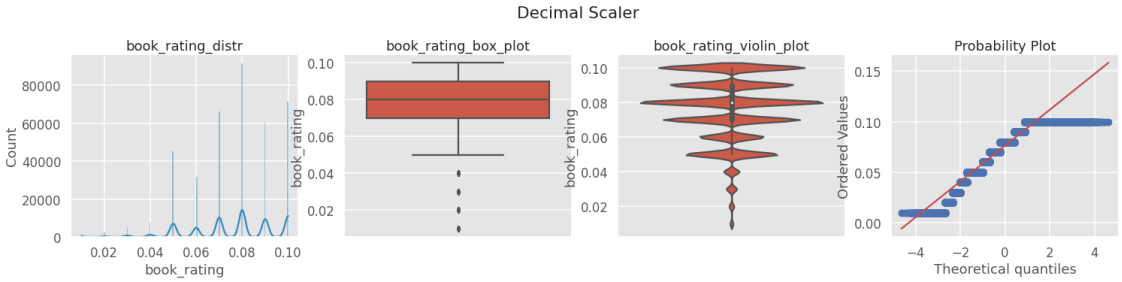
[383849 rows x 1 columns]

[463]:
```
scaled_rating.describe().T
```

[463]:

	count	mean	std	min	25%	50%	75%	max
book_rating	383849.0	0.076267	0.018413	0.01	0.07	0.08	0.09	0.1

[464]:
```
plot_univariate(dataset=scaled_rating, column_name='book_rating',␣
↪suptitle='Decimal Scaler')
```



1.4 Data Normality

1.4.1 Natural Log Transform

[465]:
```
transformed = np.log(dataset['book_rating'])
```

[466]:
```
trans_rating = pd.DataFrame(data=transformed, columns=['book_rating'])
trans_rating
```

[466]:

	book_rating
1	1.609375
3	2.080078
5	2.080078
8	2.197266
9	2.197266
...	...
1031166	1.946289
1031168	1.609375
1031169	1.946289
1031170	1.946289
1031171	2.302734

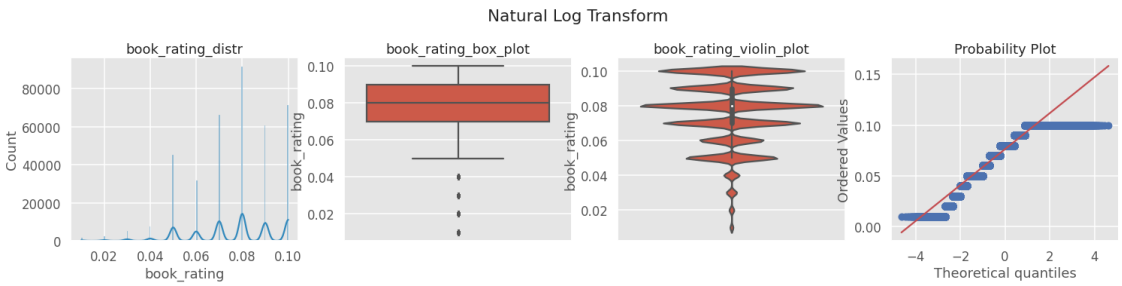
[383849 rows x 1 columns]

[467]:
```
trans_rating.describe().T
```

[467]:

	count	mean	std	min	25%	50%	75%	max
book_rating	383849.0	NaN	0.0	0.0	1.946289	2.080078	2.197266	2.302734

[468]:
```
plot_univariate(dataset=scaled_rating, column_name='book_rating',␣
↪suptitle='Natural Log Transform')
```



1.4.2 Square Root Transform

[469]:
```
transformed = np.sqrt(dataset['book_rating'])
```

[470]:
```
trans_rating = pd.DataFrame(data=transformed, columns=['book_rating'])
trans_rating
```

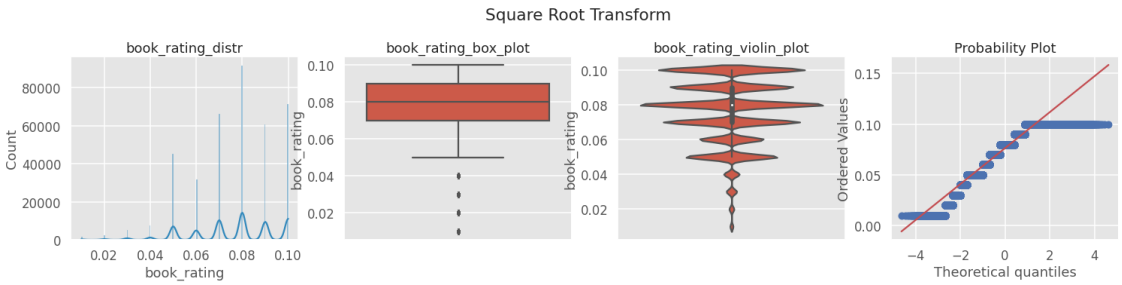
[470]:
```
         book_rating
1           2.236328
3           2.828125
5           2.828125
8           3.000000
9           3.000000
...              ...
1031166     2.646484
1031168     2.236328
1031169     2.646484
1031170     2.646484
1031171     3.162109

[383849 rows x 1 columns]
```

[471]:
```
trans_rating.describe().T
```

[471]:

	count	mean	std	min	25%	50%	75%	max
book_rating	383849.0	NaN	0.0	1.0	2.646484	2.828125	3.0	3.162109

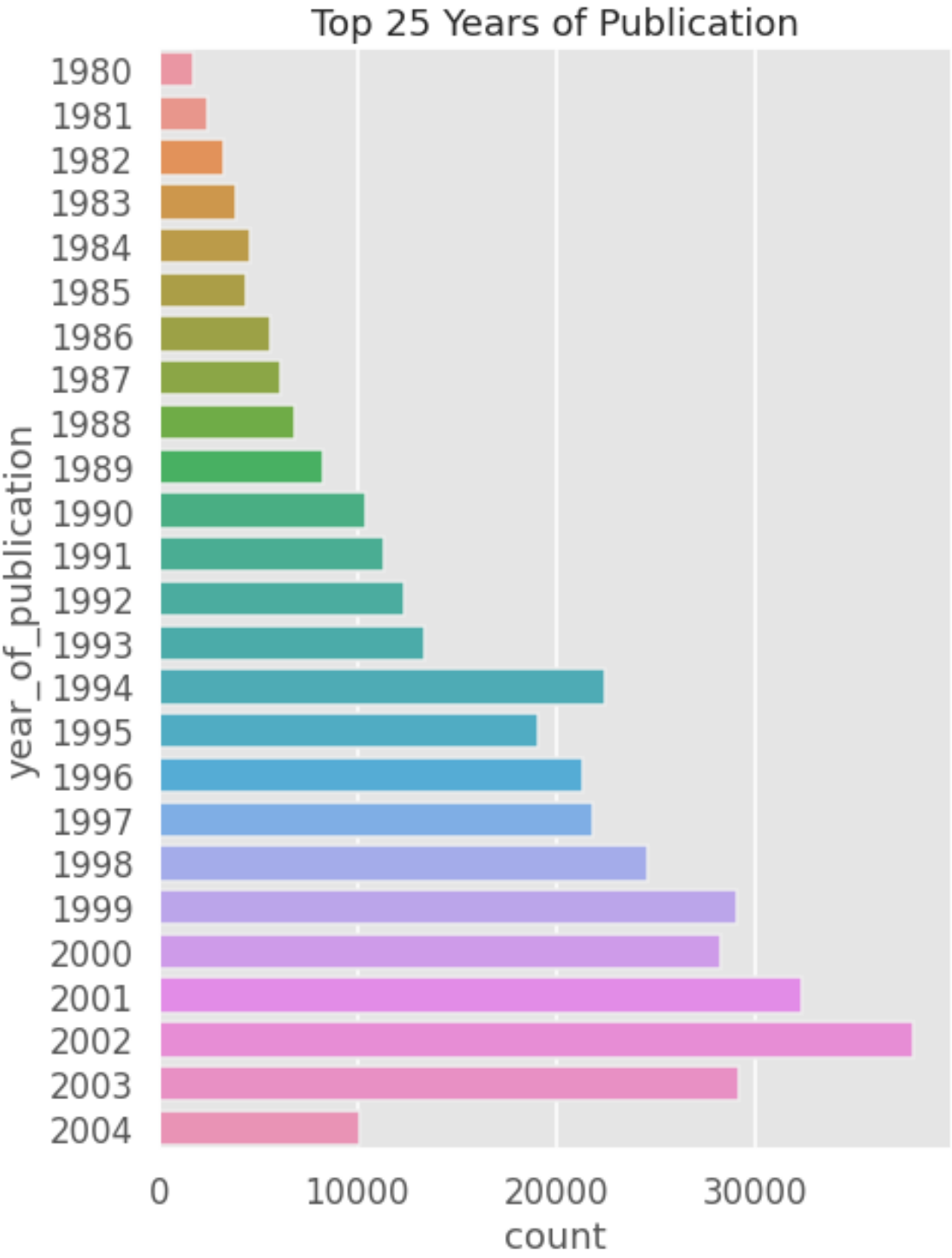
[472]:
```
plot_univariate(dataset=scaled_rating, column_name='book_rating',␣
↪suptitle='Square Root Transform')
```



1.5 Exploratory Data Analysis

[485]:
```
eda = dataset['year_of_publication'].value_counts().head(25).reset_index()
eda.columns = ['year_of_publication', 'count']
```

[493]:
```
plt.figure(figsize=(7, 10))
sns.barplot(x='count', y='year_of_publication', data=eda, orient='h')
plt.title('Top 25 Years of Publication')
plt.show()
```


Top 25 Years of Publication
year_of_publication
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
0
10000
20000
30000
count

1.6 Convert this Notebook to PDF

[494]:
```
! apt update
! apt install texlive-xetex texlive-fonts-recommended␣
 ↪texlive-generic-recommended
```

```
Ign:1 https://developer.download.nvidia.com/compute/cuda/repos/ubuntu1804/x86_64
InRelease
Get:2 https://cloud.r-project.org/bin/linux/ubuntu bionic-cran40/ InRelease
[3,626 B]
Get:3 http://security.ubuntu.com/ubuntu bionic-security InRelease [88.7 kB]
Get:4 http://ppa.launchpad.net/c2d4u.team/c2d4u4.0+/ubuntu bionic InRelease
[15.9 kB]
Ign:5 https://developer.download.nvidia.com/compute/machine-
learning/repos/ubuntu1804/x86_64  InRelease
Get:6 https://developer.download.nvidia.com/compute/cuda/repos/ubuntu1804/x86_64
Release [697 B]
Hit:7 http://archive.ubuntu.com/ubuntu bionic InRelease
Hit:8 https://developer.download.nvidia.com/compute/machine-
learning/repos/ubuntu1804/x86_64  Release
Get:9 https://developer.download.nvidia.com/compute/cuda/repos/ubuntu1804/x86_64
Release.gpg [836 B]
Get:10 http://archive.ubuntu.com/ubuntu bionic-updates InRelease [88.7 kB]
Get:11 http://ppa.launchpad.net/graphics-drivers/ppa/ubuntu bionic InRelease
[21.3 kB]
Get:12 http://archive.ubuntu.com/ubuntu bionic-backports InRelease [74.6 kB]
Ign:14
https://developer.download.nvidia.com/compute/cuda/repos/ubuntu1804/x86_64
Packages
Get:14
https://developer.download.nvidia.com/compute/cuda/repos/ubuntu1804/x86_64
Packages [444 kB]
Get:15 http://ppa.launchpad.net/c2d4u.team/c2d4u4.0+/ubuntu bionic/main Sources
[1,692 kB]
Get:16 http://security.ubuntu.com/ubuntu bionic-security/main amd64 Packages
[1,815 kB]
Get:17 http://archive.ubuntu.com/ubuntu bionic-updates/restricted amd64 Packages
[265 kB]
Get:18 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 Packages
[2,243 kB]
Get:19 http://security.ubuntu.com/ubuntu bionic-security/restricted amd64
Packages [236 kB]
Get:20 http://security.ubuntu.com/ubuntu bionic-security/universe amd64 Packages
[1,370 kB]
Get:21 http://ppa.launchpad.net/c2d4u.team/c2d4u4.0+/ubuntu bionic/main amd64
Packages [866 kB]
Get:22 http://archive.ubuntu.com/ubuntu bionic-updates/universe amd64 Packages
[2,134 kB]
```

```
Get:23 http://archive.ubuntu.com/ubuntu bionic-updates/multiverse amd64 Packages
[54.3 kB]
Get:24 http://ppa.launchpad.net/graphics-drivers/ppa/ubuntu bionic/main amd64
Packages [46.5 kB]
Fetched 11.5 MB in 3s (3,670 kB/s)
Reading package lists... Done
Building dependency tree
Reading state information... Done
68 packages can be upgraded. Run 'apt list --upgradable' to see them.
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following additional packages will be installed:
  fonts-droid-fallback fonts-lato fonts-lmodern fonts-noto-mono fonts-texgyre
  javascript-common libcupsfilters1 libcupsimage2 libgs9 libgs9-common
  libijs-0.35 libjbig2dec0 libjs-jquery libkpathsea6 libpotrace0 libptexenc1
  libruby2.5 libsynctex1 libtexlua52 libtexluajit2 libzzip-0-13 lmodern
  poppler-data preview-latex-style rake ruby ruby-did-you-mean ruby-minitest
  ruby-net-telnet ruby-power-assert ruby-test-unit ruby2.5
  rubygems-integration t1utils tex-common tex-gyre texlive-base
  texlive-binaries texlive-latex-base texlive-latex-extra
  texlive-latex-recommended texlive-pictures texlive-plain-generic tipa
Suggested packages:
  fonts-noto apache2 | lighttpd | httpd 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-reader
  | pdf-viewer texlive-fonts-recommended-doc texlive-latex-base-doc
  python-pygments icc-profiles libfile-which-perl
  libspreadsheet-parseexcel-perl texlive-latex-extra-doc
  texlive-latex-recommended-doc texlive-pstricks dot2tex prerex ruby-tcltk
  | libtcltk-ruby texlive-pictures-doc vprerex
The following NEW packages will be installed:
  fonts-droid-fallback fonts-lato fonts-lmodern fonts-noto-mono fonts-texgyre
  javascript-common libcupsfilters1 libcupsimage2 libgs9 libgs9-common
  libijs-0.35 libjbig2dec0 libjs-jquery libkpathsea6 libpotrace0 libptexenc1
  libruby2.5 libsynctex1 libtexlua52 libtexluajit2 libzzip-0-13 lmodern
  poppler-data preview-latex-style rake ruby ruby-did-you-mean ruby-minitest
  ruby-net-telnet ruby-power-assert ruby-test-unit ruby2.5
  rubygems-integration t1utils tex-common tex-gyre texlive-base
  texlive-binaries texlive-fonts-recommended texlive-generic-recommended
  texlive-latex-base texlive-latex-extra texlive-latex-recommended
  texlive-pictures texlive-plain-generic texlive-xetex tipa
0 upgraded, 47 newly installed, 0 to remove and 68 not upgraded.
Need to get 146 MB of archives.
After this operation, 460 MB of additional disk space will be used.
Get:1 http://archive.ubuntu.com/ubuntu bionic/main amd64 fonts-droid-fallback
all 1:6.0.1r16-1.1 [1,805 kB]
```

```
Get:2 http://archive.ubuntu.com/ubuntu bionic/main amd64 fonts-lato all 2.0-2
[2,698 kB]
Get:3 http://archive.ubuntu.com/ubuntu bionic/main amd64 poppler-data all
0.4.8-2 [1,479 kB]
Get:4 http://archive.ubuntu.com/ubuntu bionic/main amd64 tex-common all 6.09
[33.0 kB]
Get:5 http://archive.ubuntu.com/ubuntu bionic/main amd64 fonts-lmodern all
2.004.5-3 [4,551 kB]
Get:6 http://archive.ubuntu.com/ubuntu bionic/main amd64 fonts-noto-mono all
20171026-2 [75.5 kB]
Get:7 http://archive.ubuntu.com/ubuntu bionic/universe amd64 fonts-texgyre all
20160520-1 [8,761 kB]
Get:8 http://archive.ubuntu.com/ubuntu bionic/main amd64 javascript-common all
11 [6,066 B]
Get:9 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 libcupsfilters1
amd64 1.20.2-0ubuntu3.1 [108 kB]
Get:10 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 libcupsimage2
amd64 2.2.7-1ubuntu2.8 [18.6 kB]
Get:11 http://archive.ubuntu.com/ubuntu bionic/main amd64 libijs-0.35 amd64
0.35-13 [15.5 kB]
Get:12 http://archive.ubuntu.com/ubuntu bionic/main amd64 libjbig2dec0 amd64
0.13-6 [55.9 kB]
Get:13 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 libgs9-common
all 9.26~dfsg+0-0ubuntu0.18.04.13 [5,092 kB]
Get:14 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 libgs9 amd64
9.26~dfsg+0-0ubuntu0.18.04.13 [2,263 kB]
Get:15 http://archive.ubuntu.com/ubuntu bionic/main amd64 libjs-jquery all
3.2.1-1 [152 kB]
Get:16 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 libkpathsea6
amd64 2017.20170613.44572-8ubuntu0.1 [54.9 kB]
Get:17 http://archive.ubuntu.com/ubuntu bionic/main amd64 libpotrace0 amd64
1.14-2 [17.4 kB]
Get:18 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 libptexenc1
amd64 2017.20170613.44572-8ubuntu0.1 [34.5 kB]
Get:19 http://archive.ubuntu.com/ubuntu bionic/main amd64 rubygems-integration
all 1.11 [4,994 B]
Get:20 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 ruby2.5 amd64
2.5.1-1ubuntu1.7 [48.6 kB]
Get:21 http://archive.ubuntu.com/ubuntu bionic/main amd64 ruby amd64 1:2.5.1
[5,712 B]
Get:22 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 rake all
12.3.1-1ubuntu0.1 [44.9 kB]
Get:23 http://archive.ubuntu.com/ubuntu bionic/main amd64 ruby-did-you-mean all
1.2.0-2 [9,700 B]
Get:24 http://archive.ubuntu.com/ubuntu bionic/main amd64 ruby-minitest all
5.10.3-1 [38.6 kB]
Get:25 http://archive.ubuntu.com/ubuntu bionic/main amd64 ruby-net-telnet all
0.1.1-2 [12.6 kB]
```

```
Get:26 http://archive.ubuntu.com/ubuntu bionic/main amd64 ruby-power-assert all
0.3.0-1 [7,952 B]
Get:27 http://archive.ubuntu.com/ubuntu bionic/main amd64 ruby-test-unit all
3.2.5-1 [61.1 kB]
Get:28 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 libruby2.5
amd64 2.5.1-1ubuntu1.7 [3,068 kB]
Get:29 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 libsynctex1
amd64 2017.20170613.44572-8ubuntu0.1 [41.4 kB]
Get:30 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 libtexlua52
amd64 2017.20170613.44572-8ubuntu0.1 [91.2 kB]
Get:31 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 libtexluajit2
amd64 2017.20170613.44572-8ubuntu0.1 [230 kB]
Get:32 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 libzzip-0-13
amd64 0.13.62-3.1ubuntu0.18.04.1 [26.0 kB]
Get:33 http://archive.ubuntu.com/ubuntu bionic/main amd64 lmodern all 2.004.5-3
[9,631 kB]
Get:34 http://archive.ubuntu.com/ubuntu bionic/main amd64 preview-latex-style
all 11.91-1ubuntu1 [185 kB]
Get:35 http://archive.ubuntu.com/ubuntu bionic/main amd64 t1utils amd64 1.41-2
[56.0 kB]
Get:36 http://archive.ubuntu.com/ubuntu bionic/universe amd64 tex-gyre all
20160520-1 [4,998 kB]
Get:37 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 texlive-
binaries amd64 2017.20170613.44572-8ubuntu0.1 [8,179 kB]
Get:38 http://archive.ubuntu.com/ubuntu bionic/main amd64 texlive-base all
2017.20180305-1 [18.7 MB]
Get:39 http://archive.ubuntu.com/ubuntu bionic/universe amd64 texlive-fonts-
recommended all 2017.20180305-1 [5,262 kB]
Get:40 http://archive.ubuntu.com/ubuntu bionic/universe amd64 texlive-plain-
generic all 2017.20180305-2 [23.6 MB]
Get:41 http://archive.ubuntu.com/ubuntu bionic/universe amd64 texlive-generic-
recommended all 2017.20180305-1 [15.9 kB]
Get:42 http://archive.ubuntu.com/ubuntu bionic/main amd64 texlive-latex-base all
2017.20180305-1 [951 kB]
Get:43 http://archive.ubuntu.com/ubuntu bionic/main amd64 texlive-latex-
recommended all 2017.20180305-1 [14.9 MB]
Get:44 http://archive.ubuntu.com/ubuntu bionic/universe amd64 texlive-pictures
all 2017.20180305-1 [4,026 kB]
Get:45 http://archive.ubuntu.com/ubuntu bionic/universe amd64 texlive-latex-
extra all 2017.20180305-2 [10.6 MB]
Get:46 http://archive.ubuntu.com/ubuntu bionic/universe amd64 tipa all 2:1.3-20
[2,978 kB]
Get:47 http://archive.ubuntu.com/ubuntu bionic/universe amd64 texlive-xetex all
2017.20180305-1 [10.7 MB]
Fetched 146 MB in 5s (27.4 MB/s)
Extracting templates from packages: 100%
Preconfiguring packages ...
Selecting previously unselected package fonts-droid-fallback.
```

```
(Reading database ... 144793 files and directories currently installed.)
Preparing to unpack .../00-fonts-droid-fallback_1%3a6.0.1r16-1.1_all.deb ...
Unpacking fonts-droid-fallback (1:6.0.1r16-1.1) ...
Selecting previously unselected package fonts-lato.
Preparing to unpack .../01-fonts-lato_2.0-2_all.deb ...
Unpacking fonts-lato (2.0-2) ...
Selecting previously unselected package poppler-data.
Preparing to unpack .../02-poppler-data_0.4.8-2_all.deb ...
Unpacking poppler-data (0.4.8-2) ...
Selecting previously unselected package tex-common.
Preparing to unpack .../03-tex-common_6.09_all.deb ...
Unpacking tex-common (6.09) ...
Selecting previously unselected package fonts-lmodern.
Preparing to unpack .../04-fonts-lmodern_2.004.5-3_all.deb ...
Unpacking fonts-lmodern (2.004.5-3) ...
Selecting previously unselected package fonts-noto-mono.
Preparing to unpack .../05-fonts-noto-mono_20171026-2_all.deb ...
Unpacking fonts-noto-mono (20171026-2) ...
Selecting previously unselected package fonts-texgyre.
Preparing to unpack .../06-fonts-texgyre_20160520-1_all.deb ...
Unpacking fonts-texgyre (20160520-1) ...
Selecting previously unselected package javascript-common.
Preparing to unpack .../07-javascript-common_11_all.deb ...
Unpacking javascript-common (11) ...
Selecting previously unselected package libcupsfilters1:amd64.
Preparing to unpack .../08-libcupsfilters1_1.20.2-0ubuntu3.1_amd64.deb ...
Unpacking libcupsfilters1:amd64 (1.20.2-0ubuntu3.1) ...
Selecting previously unselected package libcupsimage2:amd64.
Preparing to unpack .../09-libcupsimage2_2.2.7-1ubuntu2.8_amd64.deb ...
Unpacking libcupsimage2:amd64 (2.2.7-1ubuntu2.8) ...
Selecting previously unselected package libijs-0.35:amd64.
Preparing to unpack .../10-libijs-0.35_0.35-13_amd64.deb ...
Unpacking libijs-0.35:amd64 (0.35-13) ...
Selecting previously unselected package libjbig2dec0:amd64.
Preparing to unpack .../11-libjbig2dec0_0.13-6_amd64.deb ...
Unpacking libjbig2dec0:amd64 (0.13-6) ...
Selecting previously unselected package libgs9-common.
Preparing to unpack .../12-libgs9-common_9.26~dfsg+0-0ubuntu0.18.04.13_all.deb
...
Unpacking libgs9-common (9.26~dfsg+0-0ubuntu0.18.04.13) ...
Selecting previously unselected package libgs9:amd64.
Preparing to unpack .../13-libgs9_9.26~dfsg+0-0ubuntu0.18.04.13_amd64.deb ...
Unpacking libgs9:amd64 (9.26~dfsg+0-0ubuntu0.18.04.13) ...
Selecting previously unselected package libjs-jquery.
Preparing to unpack .../14-libjs-jquery_3.2.1-1_all.deb ...
Unpacking libjs-jquery (3.2.1-1) ...
Selecting previously unselected package libkpathsea6:amd64.
Preparing to unpack .../15-libkpathsea6_2017.20170613.44572-8ubuntu0.1_amd64.deb
```

```
...
Unpacking libkpathsea6:amd64 (2017.20170613.44572-8ubuntu0.1) ...
Selecting previously unselected package libpotrace0.
Preparing to unpack .../16-libpotrace0_1.14-2_amd64.deb ...
Unpacking libpotrace0 (1.14-2) ...
Selecting previously unselected package libptexenc1:amd64.
Preparing to unpack .../17-libptexenc1_2017.20170613.44572-8ubuntu0.1_amd64.deb
...
Unpacking libptexenc1:amd64 (2017.20170613.44572-8ubuntu0.1) ...
Selecting previously unselected package rubygems-integration.
Preparing to unpack .../18-rubygems-integration_1.11_all.deb ...
Unpacking rubygems-integration (1.11) ...
Selecting previously unselected package ruby2.5.
Preparing to unpack .../19-ruby2.5_2.5.1-1ubuntu1.7_amd64.deb ...
Unpacking ruby2.5 (2.5.1-1ubuntu1.7) ...
Selecting previously unselected package ruby.
Preparing to unpack .../20-ruby_1%3a2.5.1_amd64.deb ...
Unpacking ruby (1:2.5.1) ...
Selecting previously unselected package rake.
Preparing to unpack .../21-rake_12.3.1-1ubuntu0.1_all.deb ...
Unpacking rake (12.3.1-1ubuntu0.1) ...
Selecting previously unselected package ruby-did-you-mean.
Preparing to unpack .../22-ruby-did-you-mean_1.2.0-2_all.deb ...
Unpacking ruby-did-you-mean (1.2.0-2) ...
Selecting previously unselected package ruby-minitest.
Preparing to unpack .../23-ruby-minitest_5.10.3-1_all.deb ...
Unpacking ruby-minitest (5.10.3-1) ...
Selecting previously unselected package ruby-net-telnet.
Preparing to unpack .../24-ruby-net-telnet_0.1.1-2_all.deb ...
Unpacking ruby-net-telnet (0.1.1-2) ...
Selecting previously unselected package ruby-power-assert.
Preparing to unpack .../25-ruby-power-assert_0.3.0-1_all.deb ...
Unpacking ruby-power-assert (0.3.0-1) ...
Selecting previously unselected package ruby-test-unit.
Preparing to unpack .../26-ruby-test-unit_3.2.5-1_all.deb ...
Unpacking ruby-test-unit (3.2.5-1) ...
Selecting previously unselected package libruby2.5:amd64.
Preparing to unpack .../27-libruby2.5_2.5.1-1ubuntu1.7_amd64.deb ...
Unpacking libruby2.5:amd64 (2.5.1-1ubuntu1.7) ...
Selecting previously unselected package libsynctex1:amd64.
Preparing to unpack .../28-libsynctex1_2017.20170613.44572-8ubuntu0.1_amd64.deb
...
Unpacking libsynctex1:amd64 (2017.20170613.44572-8ubuntu0.1) ...
Selecting previously unselected package libtexlua52:amd64.
Preparing to unpack .../29-libtexlua52_2017.20170613.44572-8ubuntu0.1_amd64.deb
...
Unpacking libtexlua52:amd64 (2017.20170613.44572-8ubuntu0.1) ...
Selecting previously unselected package libtexluajit2:amd64.
```

```
Preparing to unpack
.../30-libtexluajit2_2017.20170613.44572-8ubuntu0.1_amd64.deb ...
Unpacking libtexluajit2:amd64 (2017.20170613.44572-8ubuntu0.1) ...
Selecting previously unselected package libzzip-0-13:amd64.
Preparing to unpack .../31-libzzip-0-13_0.13.62-3.1ubuntu0.18.04.1_amd64.deb ...
Unpacking libzzip-0-13:amd64 (0.13.62-3.1ubuntu0.18.04.1) ...
Selecting previously unselected package lmodern.
Preparing to unpack .../32-lmodern_2.004.5-3_all.deb ...
Unpacking lmodern (2.004.5-3) ...
Selecting previously unselected package preview-latex-style.
Preparing to unpack .../33-preview-latex-style_11.91-1ubuntu1_all.deb ...
Unpacking preview-latex-style (11.91-1ubuntu1) ...
Selecting previously unselected package t1utils.
Preparing to unpack .../34-t1utils_1.41-2_amd64.deb ...
Unpacking t1utils (1.41-2) ...
Selecting previously unselected package tex-gyre.
Preparing to unpack .../35-tex-gyre_20160520-1_all.deb ...
Unpacking tex-gyre (20160520-1) ...
Selecting previously unselected package texlive-binaries.
Preparing to unpack .../36-texlive-
binaries_2017.20170613.44572-8ubuntu0.1_amd64.deb ...
Unpacking texlive-binaries (2017.20170613.44572-8ubuntu0.1) ...
Selecting previously unselected package texlive-base.
Preparing to unpack .../37-texlive-base_2017.20180305-1_all.deb ...
Unpacking texlive-base (2017.20180305-1) ...
Selecting previously unselected package texlive-fonts-recommended.
Preparing to unpack .../38-texlive-fonts-recommended_2017.20180305-1_all.deb ...
Unpacking texlive-fonts-recommended (2017.20180305-1) ...
Selecting previously unselected package texlive-plain-generic.
Preparing to unpack .../39-texlive-plain-generic_2017.20180305-2_all.deb ...
Unpacking texlive-plain-generic (2017.20180305-2) ...
Selecting previously unselected package texlive-generic-recommended.
Preparing to unpack .../40-texlive-generic-recommended_2017.20180305-1_all.deb
...
Unpacking texlive-generic-recommended (2017.20180305-1) ...
Selecting previously unselected package texlive-latex-base.
Preparing to unpack .../41-texlive-latex-base_2017.20180305-1_all.deb ...
Unpacking texlive-latex-base (2017.20180305-1) ...
Selecting previously unselected package texlive-latex-recommended.
Preparing to unpack .../42-texlive-latex-recommended_2017.20180305-1_all.deb ...
Unpacking texlive-latex-recommended (2017.20180305-1) ...
Selecting previously unselected package texlive-pictures.
Preparing to unpack .../43-texlive-pictures_2017.20180305-1_all.deb ...
Unpacking texlive-pictures (2017.20180305-1) ...
Selecting previously unselected package texlive-latex-extra.
Preparing to unpack .../44-texlive-latex-extra_2017.20180305-2_all.deb ...
Unpacking texlive-latex-extra (2017.20180305-2) ...
Selecting previously unselected package tipa.
```

```
Preparing to unpack .../45-tipa_2%3a1.3-20_all.deb ...
Unpacking tipa (2:1.3-20) ...
Selecting previously unselected package texlive-xetex.
Preparing to unpack .../46-texlive-xetex_2017.20180305-1_all.deb ...
Unpacking texlive-xetex (2017.20180305-1) ...
Setting up libgs9-common (9.26~dfsg+0-0ubuntu0.18.04.13) ...
Setting up libkpathsea6:amd64 (2017.20170613.44572-8ubuntu0.1) ...
Setting up libjs-jquery (3.2.1-1) ...
Setting up libtexlua52:amd64 (2017.20170613.44572-8ubuntu0.1) ...
Setting up fonts-droid-fallback (1:6.0.1r16-1.1) ...
Setting up libsynctex1:amd64 (2017.20170613.44572-8ubuntu0.1) ...
Setting up libptexenc1:amd64 (2017.20170613.44572-8ubuntu0.1) ...
Setting up tex-common (6.09) ...
update-language: texlive-base not installed and configured, doing nothing!
Setting up poppler-data (0.4.8-2) ...
Setting up tex-gyre (20160520-1) ...
Setting up preview-latex-style (11.91-1ubuntu1) ...
Setting up fonts-texgyre (20160520-1) ...
Setting up fonts-noto-mono (20171026-2) ...
Setting up fonts-lato (2.0-2) ...
Setting up libcupsfilters1:amd64 (1.20.2-0ubuntu3.1) ...
Setting up libcupsimage2:amd64 (2.2.7-1ubuntu2.8) ...
Setting up libjbig2dec0:amd64 (0.13-6) ...
Setting up ruby-did-you-mean (1.2.0-2) ...
Setting up t1utils (1.41-2) ...
Setting up ruby-net-telnet (0.1.1-2) ...
Setting up libijs-0.35:amd64 (0.35-13) ...
Setting up rubygems-integration (1.11) ...
Setting up libpotrace0 (1.14-2) ...
Setting up javascript-common (11) ...
Setting up ruby-minitest (5.10.3-1) ...
Setting up libzzip-0-13:amd64 (0.13.62-3.1ubuntu0.18.04.1) ...
Setting up libgs9:amd64 (9.26~dfsg+0-0ubuntu0.18.04.13) ...
Setting up libtexluajit2:amd64 (2017.20170613.44572-8ubuntu0.1) ...
Setting up fonts-lmodern (2.004.5-3) ...
Setting up ruby-power-assert (0.3.0-1) ...
Setting up texlive-binaries (2017.20170613.44572-8ubuntu0.1) ...
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 texlive-base (2017.20180305-1) ...
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/config/pdftexconfig.tex
Setting up texlive-fonts-recommended (2017.20180305-1) ...
Setting up texlive-plain-generic (2017.20180305-2) ...
Setting up texlive-generic-recommended (2017.20180305-1) ...
Setting up texlive-latex-base (2017.20180305-1) ...
Setting up lmodern (2.004.5-3) ...
Setting up texlive-latex-recommended (2017.20180305-1) ...
Setting up texlive-pictures (2017.20180305-1) ...
Setting up tipa (2:1.3-20) ...
Regenerating '/var/lib/texmf/fmtutil.cnf-DEBIAN'... done.
Regenerating '/var/lib/texmf/fmtutil.cnf-TEXLIVEDIST'... done.
update-fmtutil has updated the following file(s):
        /var/lib/texmf/fmtutil.cnf-DEBIAN
        /var/lib/texmf/fmtutil.cnf-TEXLIVEDIST
If you want to activate the changes in the above file(s),
you should run fmtutil-sys or fmtutil.
Setting up texlive-latex-extra (2017.20180305-2) ...
Setting up texlive-xetex (2017.20180305-1) ...
Setting up ruby2.5 (2.5.1-1ubuntu1.7) ...
Setting up ruby (1:2.5.1) ...
Setting up ruby-test-unit (3.2.5-1) ...
Setting up rake (12.3.1-1ubuntu0.1) ...
Setting up libruby2.5:amd64 (2.5.1-1ubuntu1.7) ...
Processing triggers for mime-support (3.60ubuntu1) ...
Processing triggers for libc-bin (2.27-3ubuntu1.2) ...
/sbin/ldconfig.real: /usr/local/lib/python3.6/dist-
packages/ideep4py/lib/libmkldnn.so.0 is not a symbolic link

Processing triggers for man-db (2.8.3-2ubuntu0.1) ...
Processing triggers for fontconfig (2.12.6-0ubuntu2) ...
Processing triggers for tex-common (6.09) ...
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.
```

[ ]:
```
import re, pathlib, shutil

# Get a list of all your Notebooks
notebooks = [x for x in pathlib.Path("/content/drive/My Drive/Colab Notebooks").
 ↪iterdir() if
             re.search(r"\.ipynb", x.name, flags = re.I)]
```

```
for i, n in enumerate(notebooks):
    print(f"\nProcessing  [{i+1:{len(str(len(notebooks)))}d}/{len(notebooks)}] ␣
 ↪{n.name}\n")

    # Optionally copy your notebooks from gdrive to your vm
    shutil.copy(n, n.name)
    n = pathlib.Path(n.name)

    !jupyter nbconvert "{n.as_posix()}" --to pdf --output "{n.stem.replace(" ",␣
 ↪"_")}"
```

[495]:
```
import subprocess
```

[ ]:
```
s = subprocess.Popen(shlex.split(
    f'jupyter nbconvert "{n.as_posix()}" --to pdf --output "{n.stem.replace("␣
 ↪", "_")}"'
    ), shell = False, stdout = subprocess.PIPE, stderr = subprocess.PIPE)
s.wait()
s.stdout.read()
```