Final Report

(Output)

Employment in non-profit sector

By: Sangjin (Eric) Lee

Submitted by November 27th , 2023

Page of content

Link for Github Page and details	Page 3
Output #01	Page 6
Output #02	Page 8
Output #03	Page 11
Output #04	Page 13
Output #05	Page 17
Output #06	Page 18
Output #07	Page 19
Output #08	Page 33
Output #09	Page 40
Output #10	Page 43
Output #11	Page 57
Output #12	Page 63
Output #13	Page 102
Output #14	Page 106
Output #15	Page 115
Output #16	Page 124
Output #17	Page 132
Output #18	Page 141
Output #19	Page 150

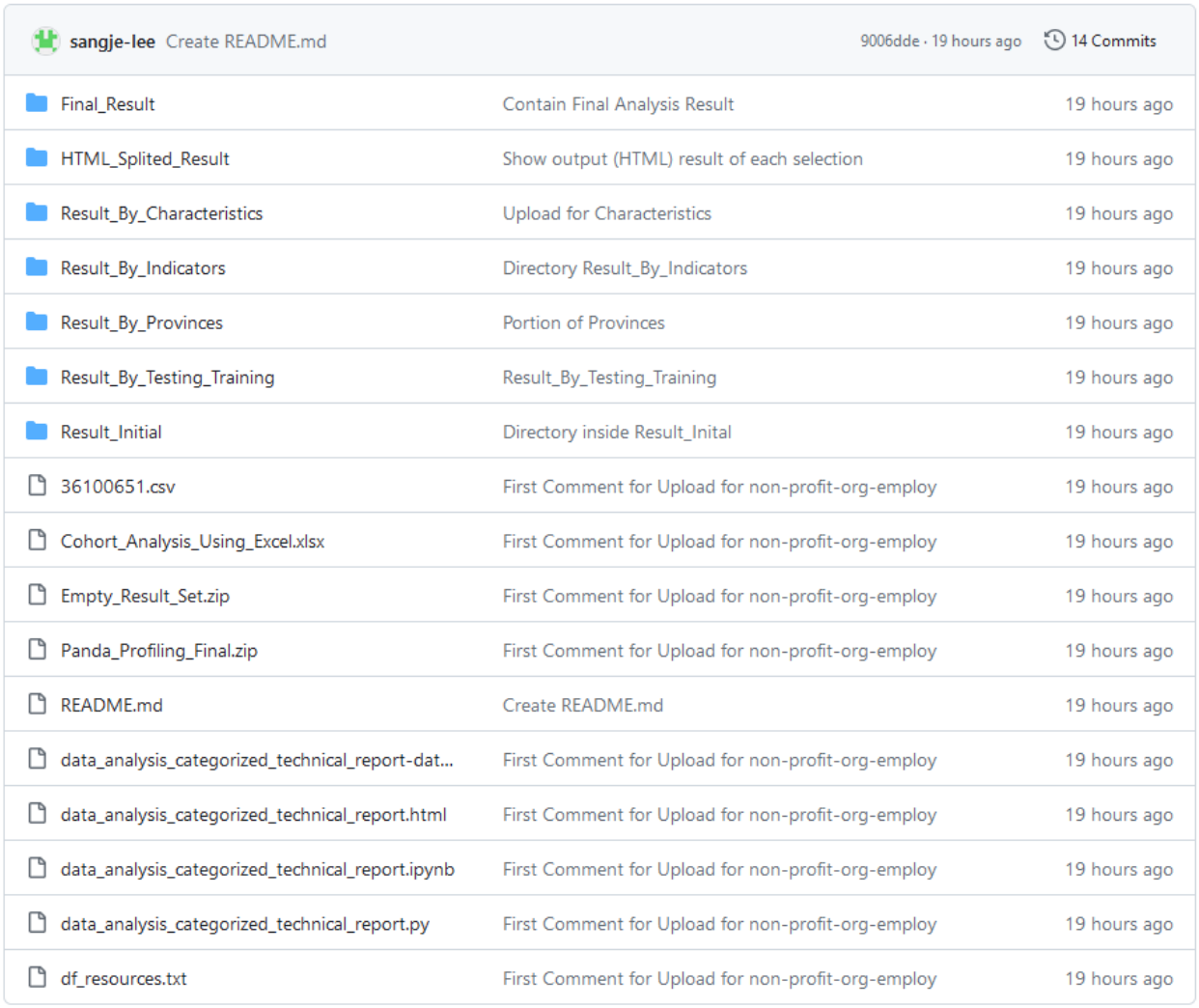
A link to a repository on GitHub.

https://github.com/sangje-lee/non-profit-org-employment

Here is the Github website where codes and results are uploaded. I have uploaded not only my progress but as well as results of each code. Since the analysis took a lot of time and effort, I could not be able to paste all the results in this report. The result itself is more than 50 pages. This includes Jupiter Notebook file, Panda Profiling, and outputting all the results in the report.

In there, I have uploaded the result that includes python code, my analysis, and the result that I haven't put in the report. In addition, I placed the project in the main directory. It will be the same file as I am submitted with this report.

Here is the detail of my Github where I put all the results below,



sangje-lee Create README.md	9006dde · 19 hours ago	14 Commits
Final_Result	Contain Final Analysis Result	19 hours ago
HTML_Splited_Result	Show output (HTML) result of each selection	19 hours ago
Result_By_Characteristics	Upload for Characteristics	19 hours ago
Result_By_Indicators	Directory Result_By_Indicators	19 hours ago
Result_By_Provinces	Portion of Provinces	19 hours ago
Result_By_Testing_Training	Result_By_Testing_Training	19 hours ago
Result_Initial	Directory inside Result_Inital	19 hours ago
36100651.csv	First Comment for Upload for non-profit-org-employ	19 hours ago
Cohort_Analysis_Using_Excel.xlsx	First Comment for Upload for non-profit-org-employ	19 hours ago
Empty_Result_Set.zip	First Comment for Upload for non-profit-org-employ	19 hours ago
Panda_Profiling_Final.zip	First Comment for Upload for non-profit-org-employ	19 hours ago
README.md	Create README.md	19 hours ago
data_analysis_categorized_technical_report-dat...	First Comment for Upload for non-profit-org-employ	19 hours ago
data_analysis_categorized_technical_report.html	First Comment for Upload for non-profit-org-employ	19 hours ago
data_analysis_categorized_technical_report.ipynb	First Comment for Upload for non-profit-org-employ	19 hours ago
data_analysis_categorized_technical_report.py	First Comment for Upload for non-profit-org-employ	19 hours ago
df_resources.txt	First Comment for Upload for non-profit-org-employ	19 hours ago

First of all, “data_analysis_categorized_technical_report.ipynb” is where I did all of my code. It is saved in Jupiter notebook format. I used Anaconda virtual environment using python 3.8.18 installing Numpy, Pandas, Fitter, and Panda-profiling.

Next “data_analysis_categorized_technical_report.html” is basically in html/PDF file from the Jupiter Notebook file. “data_analysis_categorized_technical_report.py” is the python file of Jupiter Notebook file but saved as python script. There’s “data_analysis_..._report-data-…-only.py” file, where it’s storing only the data processing of the report.

Next, “Panda Profiling Result” directory stored all the html that was generated from the Jupiter Notebook files. The html files are the main source of analysis of this report. It displays much more details than the two PDF files I mentioned.

There are two files, “df_resources.txt” and the CSV files. The txt file is just showing all the variables that I used for this analysis and from Jupiter Notebook. The CSV file is the original dataset from the website.

Here are the remaining file descriptions,

“36100651.csv”	Original dataset contain employment in non-profit organizations.
“36100651-eng.zip”	Zip file for original dataset and some description
Empty_Result_Set.zip	Contain all requirement folder with Jupiter notebook script inside. No CSV files, when extract to main directory and run the script, it will automatically add CSV files to the dedicated directory.
README.md	Read me for the Github
HTML_Splited_Result	Output that are already executed and saved in html format. The final html is also saved in there as well.
Result_By_Characteristics	Part contains about split by “Characteristics” and its portion script.
Result_By_Indicators	Part contains about split by “Indicators” and its portion script.
Result_By_Provinces	Part contains about added five provinces + modified characteristics
Result_Inital	Part contains before the split by ‘Indicators’ and removal of nonessential files. It’s the beginning of the code.
Final_Result	Script contains result of this Analysis. Important for final analysis.

For references, the Jupiter notebook file inside the directory needs main script to run before running the script inside the directory. Also, **my portion of final analysis script** is inside **Final_Result** directory. **Also, script inside the directory required to executed and have all csv file inside the directory.**

Output #01

```
print(df.info())

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 105840 entries, 0 to 105839
Data columns (total 17 columns):
 #   Column           Non-Null Count   Dtype  
---  ------           --------------   -----  
 0   REF_DATE         105840 non-null  int64  
 1   GEO              105840 non-null  object 
 2   DGUID            105840 non-null  object 
 3   Sector           105840 non-null  object 
 4   Characteristics  105840 non-null  object 
 5   Indicators       105840 non-null  object 
 6   UOM              105840 non-null  object 
 7   UOM_ID           105840 non-null  int64  
 8   SCALAR_FACTOR    105840 non-null  object 
 9   SCALAR_ID        105840 non-null  int64  
 10  VECTOR           105840 non-null  object 
 11  COORDINATE       105840 non-null  object 
 12  VALUE            102816 non-null  float64
 13  STATUS           3024 non-null    object 
 14  SYMBOL           0 non-null       float64
 15  TERMINATED       0 non-null       float64
 16  DECIMALS         105840 non-null  int64  
```

```
print(df.head(10))
   REF_DATE     GEO           DGUID                          Sector  \
0      2010  Canada  2016A000011124  Total non-profit institutions
1      2010  Canada  2016A000011124  Total non-profit institutions
2      2010  Canada  2016A000011124  Total non-profit institutions
3      2010  Canada  2016A000011124  Total non-profit institutions
4      2010  Canada  2016A000011124  Total non-profit institutions
5      2010  Canada  2016A000011124  Total non-profit institutions
6      2010  Canada  2016A000011124  Total non-profit institutions
7      2010  Canada  2016A000011124  Total non-profit institutions
8      2010  Canada  2016A000011124  Total non-profit institutions
9      2010  Canada  2016A000011124  Total non-profit institutions

    Characteristics                        Indicators      UOM  UOM_ID  \
0    Male employees                   Number of jobs     Jobs     190
1    Male employees                     Hours worked    Hours     152
2    Male employees                 Wages and salaries  Dollars      81
3    Male employees        Average annual hours worked    Hours     152
4    Male employees        Average weekly hours worked    Hours     152
5    Male employees  Average annual wages and salaries  Dollars      81
6    Male employees                Average hourly wage  Dollars      81
7  Female employees                   Number of jobs     Jobs     190
8  Female employees                     Hours worked    Hours     152
9  Female employees                 Wages and salaries  Dollars      81

  SCALAR_FACTOR  SCALAR_ID       VECTOR COORDINATE       VALUE STATUS  SYMBOL  \
0         units          0  v1273033811    1.1.1.1   642584.00    NaN     NaN
1     thousands          3  v1273033812    1.1.1.2  1048516.00    NaN     NaN
2      millions          6  v1273033813    1.1.1.3    30805.00    NaN     NaN
3         units          0  v1273033814    1.1.1.4     1632.00    NaN     NaN
4         units          0  v1273033815    1.1.1.5       31.00    NaN     NaN
5         units          0  v1273033816    1.1.1.6    47940.00    NaN     NaN
6         units          0  v1273033817    1.1.1.7       29.38    NaN     NaN
7         units          0  v1273033909    1.1.2.1  1500394.00    NaN     NaN
8     thousands          3  v1273033910    1.1.2.2  2331018.00    NaN     NaN
9      millions          6  v1273033911    1.1.2.3    60943.00    NaN     NaN

   TERMINATED  DECIMALS
0         NaN         0
1         NaN         0
2         NaN         0
3         NaN         0
4         NaN         0
5         NaN         0
6         NaN         2
7         NaN         0
8         NaN         0
9         NaN         0
```

Output #02

All of the columns are equally divided as mentioned above.

```
grouped = df.groupby(['REF_DATE'])
print(grouped['VALUE'].agg([np.size]))

grouped = df.groupby(['GEO'])
print(grouped['VALUE'].agg([np.size]))

grouped = df.groupby(['Sector'])
print(grouped['VALUE'].agg([np.size]))

grouped = df.groupby(['Characteristics'])
print(grouped['VALUE'].agg([np.size]))

grouped = df.groupby(['Indicators'])
print(grouped['VALUE'].agg([np.size]))

grouped = df.groupby(['VALUE'])
print(grouped['VALUE'].agg([np.size]))
```

```
          size
REF_DATE
2010      8820
2011      8820
2012      8820
2013      8820
2014      8820
2015      8820
2016      8820
2017      8820
2018      8820
2019      8820
2020      8820
2021      8820
```

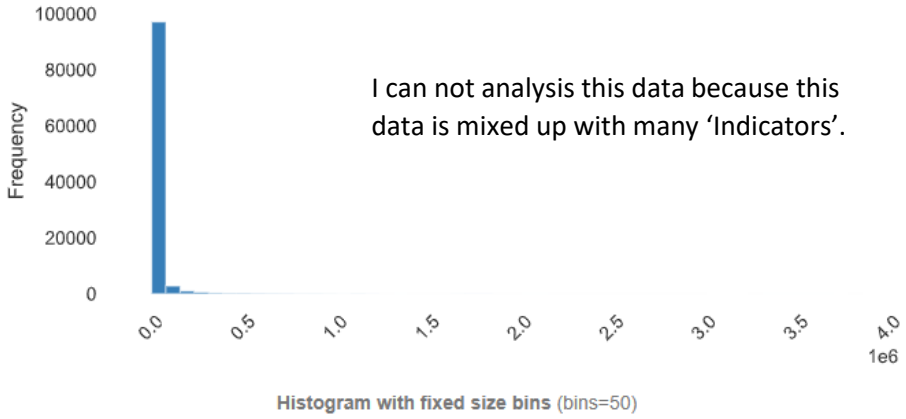
```
                          size
GEO
Alberta                   7560
British Columbia          7560
Canada                    7560
Manitoba                  7560
New Brunswick             7560
Newfoundland and Labrador 7560
Northwest Territories     7560
Nova Scotia               7560
Nunavut                   7560
Ontario                   7560
Prince Edward Island      7560
Quebec                    7560
Saskatchewan              7560
Yukon                     7560
```

Sector	size
Business non-profit institutions	21168
Government non-profit institutions	21168
Non-profit institutions serving households (com...	21168
Total non-profit institutions	21168
Total non-profit institutions excluding governm...	21168

Characteristics	size
15 to 24 years	5880
25 to 34 years	5880
35 to 44 years	5880
45 to 54 years	5880
55 to 64 years	5880
65 years old and over	5880
College diploma	5880
Female employees	5880
High school diploma and less	5880
Immigrant employees	5880
Indigenous identity employees	5880
Male employees	5880
Non-immigrant employees	5880
Non-indigenous identity employees	5880
Not a visible minority	5880
Trade certificate	5880
University degree and higher	5880
Visible minority	5880

Indicators	size
Average annual hours worked	15120
Average annual wages and salaries	15120
Average hourly wage	15120
Average weekly hours worked	15120
Hours worked	15120
Number of jobs	15120
Wages and salaries	15120

Histogram for “VALUE”



Histogram with fixed size bins (bins=50)

I can not analysis this data because this data is mixed up with many ‘Indicators’.

Output #03

```
print(df_sorted.info())
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 105840 entries, 0 to 105839
Data columns (total 8 columns):
 #   Column           Non-Null Count   Dtype
---  ------           --------------   -----
 0   REF_DATE         105840 non-null  int64
 1   GEO              105840 non-null  object
 2   Sector           105840 non-null  object
 3   Characteristics  105840 non-null  object
 4   Indicators       105840 non-null  object
 5   UOM              105840 non-null  object
 6   SCALAR_FACTOR    105840 non-null  object
 7   VALUE            102816 non-null  float64
dtypes: float64(1), int64(1), object(6)
memory usage: 6.5+ MB
None
```

```
print(df_sorted.head(20))
    REF_DATE     GEO                          Sector      Characteristics  \
0       2010  Canada  Total non-profit institutions       Male employees
1       2010  Canada  Total non-profit institutions       Male employees
2       2010  Canada  Total non-profit institutions       Male employees
3       2010  Canada  Total non-profit institutions       Male employees
4       2010  Canada  Total non-profit institutions       Male employees
5       2010  Canada  Total non-profit institutions       Male employees
6       2010  Canada  Total non-profit institutions       Male employees
7       2010  Canada  Total non-profit institutions     Female employees
8       2010  Canada  Total non-profit institutions     Female employees
9       2010  Canada  Total non-profit institutions     Female employees
10      2010  Canada  Total non-profit institutions     Female employees
11      2010  Canada  Total non-profit institutions     Female employees
12      2010  Canada  Total non-profit institutions     Female employees
13      2010  Canada  Total non-profit institutions     Female employees
14      2010  Canada  Total non-profit institutions  Immigrant employees
15      2010  Canada  Total non-profit institutions  Immigrant employees
16      2010  Canada  Total non-profit institutions  Immigrant employees
17      2010  Canada  Total non-profit institutions  Immigrant employees
18      2010  Canada  Total non-profit institutions  Immigrant employees
19      2010  Canada  Total non-profit institutions  Immigrant employees

                           Indicators      UOM SCALAR_FACTOR       VALUE
0                       Number of jobs     Jobs         units   642584.00
1                         Hours worked    Hours     thousands  1048516.00
2                    Wages and salaries  Dollars      millions    30805.00
3          Average annual hours worked    Hours         units     1632.00
4          Average weekly hours worked    Hours         units       31.00
5   Average annual wages and salaries  Dollars         units    47940.00
6                  Average hourly wage  Dollars         units       29.38
7                       Number of jobs     Jobs         units  1500394.00
8                         Hours worked    Hours     thousands  2331018.00
9                    Wages and salaries  Dollars      millions    60943.00
10         Average annual hours worked    Hours         units     1554.00
11         Average weekly hours worked    Hours         units       30.00
12  Average annual wages and salaries  Dollars         units    40618.00
13                 Average hourly wage  Dollars         units       26.14
14                      Number of jobs     Jobs         units   503236.00
```

15	Hours worked	Hours	thousands	802251.00
16	Wages and salaries	Dollars	millions	22212.00
17	Average annual hours worked	Hours	units	1594.00
18	Average weekly hours worked	Hours	units	31.00
19	Average annual wages and salaries	Dollars	units	44138.00

Output #04

These are the main reasons why I have decided to remove the following columns/attributes.

UOM_ID

UOM_ID inside this dataset is simply translate actual words into numeric codes. Since this look like having repetitive code, I have decided to remove it. Alternatively, I can use this instead of UOM.

	UOM	UOM_ID
0	Jobs	190
1	Hours	152
2	Dollars	81
3	Hours	152
4	Hours	152
5	Dollars	81
6	Dollars	81
7	Jobs	190
8	Hours	152
9	Dollars	81
10	Hours	152
11	Hours	152
12	Dollars	81
13	Dollars	81
14	Jobs	190
15	Hours	152
16	Dollars	81
17	Hours	152
18	Hours	152
19	Dollars	81

Noticed that UOM_ID for Jobs is 190, Hours = 152, and Dollars = 81.

SCALAR_ID

For similar reason why, I removed UOM_ID, I think there's a lot of repetitive inside the SCALAR_ID to SCALAR_FACTOR. If I were to remove SCALAR_FACTOR, then SCALAR_ID can be used.

	SCALAR_ID	SCALAR_FACTOR
0	0	units
1	3	thousands
2	6	millions
3	0	units
4	0	units
5	0	units
6	0	units
7	0	units
8	3	thousands
9	6	millions
10	0	units
11	0	units
12	0	units
13	0	units
14	0	units
15	3	thousands
16	6	millions
17	0	units
18	0	units
19	0	units

Noticed that, Units = 0, thousands = 3, and million = 6

‘VECTOR’

I removed VECTOR column because I don’t believe that it refers to anything that I will analysis. VECTOR columns use 11 words that usually start with “v1273”. I believed that the Vector is code that is used to refer the data.

Value	Count	Frequency (%)
v1273033811	12	< 0.1%
v1273033912	12	< 0.1%
v1273034009	12	< 0.1%
v1273034008	12	< 0.1%
v1273034007	12	< 0.1%
v1273033915	12	< 0.1%
v1273033914	12	< 0.1%
v1273033913	12	< 0.1%
v1273033911	12	< 0.1%
v1273034698	12	< 0.1%
Other values (8810)	105720	99.9%

COORDINATE

I have decided to remove ‘COORDINATE’ columns. I believed it doesn’t really need to analysis my result. From my perspective, I think Coordinate is being used as where this data is collected according to the GPS.

Value	Count	Frequency (%)
1.1.1.1	12	< 0.1%
1.1.2.4	12	< 0.1%
1.1.3.3	12	< 0.1%
1.1.3.2	12	< 0.1%
1.1.3.1	12	< 0.1%
1.1.2.7	12	< 0.1%
1.1.2.6	12	< 0.1%
1.1.2.5	12	< 0.1%
1.1.2.3	12	< 0.1%
1.1.10.6	12	< 0.1%
Other values (8810)	105720	99.9%

STATUS

I removed “STATUS” columns because the column is not being used with my analysis. In addition, it doesn’t really contain anything meaningful to analysis anything. Most of these column rows contain null (or missing value) or marked “x”.

SYMBOL

Similar reason to STATUS column. It doesn’t contain anything meaningful and do not need to analysis this column. All column rows contain null (or missing value).

TERMINATED

Similar reason to TERMINATED column. It doesn’t contain anything meaningful and do not need to analysis this column. All column rows contain missing value.

DECIMAL

I removed ‘DECIMAL’ columns because I don’t need to use this column to analysis my result. It’s there for me to check whether this row contains any decimal values or not inside “VALUES” columns.

Value	Count	Frequency (%)
0	90720	85.7%
2	15120	14.3%

Output #05

Characteristics columns:

Characteristics	Size
15 to 24 years	5880
25 to 34 years	5880
35 to 44 years	5880
45 to 54 years	5880
55 to 64 years	5880
65 years old and over	5880
College diploma	5880
Female employees	5880
High school diploma and less	5880
Immigrant employees	5880
Indigenous identity employees	5880
Male employees	5880
Non-immigrant employees	5880
Non-indigenous identity employees	5880
Not a visible minority	5880
Trade certificate	5880
University degree and higher	5880
Visible minority	5880

Used Panda-profiling and
Aggregated using Numpy.

Noticed : Data are evenly divided.

Indicator columns:

Indicators	size
Average annual hours worked	15120
Average annual wages and salaries	15120
Average hourly wage	15120
Average weekly hours worked	15120
Hours worked	15120
Number of jobs	15120
Wages and salaries	15120

Noticed that they are evenly divided.

Output #06

After filtering some of the missing value, this is the new result.

Characteristics	Before	After
15 to 24 years	5880	5376
25 to 34 years	5880	5880
35 to 44 years	5880	5880
45 to 54 years	5880	5880
55 to 64 years	5880	5880
65 years old and over	5880	5376
College diploma	5880	5880
Female employees	5880	5880
High school diploma and less	5880	5880
Immigrant employees	5880	5544
Indigenous identity employees	5880	5544
Male employees	5880	5880
Non-immigrant employees	5880	5544
Non-indigenous identity employees	5880	5544
Not a visible minority	5880	5880
Trade certificate	5880	5544
University degree and higher	5880	5544
Visible minority	5880	5880

Indicators	Before	After
Average annual hours worked	15120	14688
Average annual wages and salaries	15120	14688
Average hourly wage	15120	14688
Average weekly hours worked	15120	14688
Hours worked	15120	14688
Number of jobs	15120	14688
Wages and salaries	15120	14688

Output #07a

For Average annual hours worked,

Dataset statistics

Number of variables	**8**
Number of observations	**14688**
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	1.0 MiB
Average record size in memory	72.0 B

Based on panda-profiling.

There are 14688 observations with 8 variables to work on.

Right now, only “Indicators” columns have one value.

Variable types

Numeric	2
Categorical	6

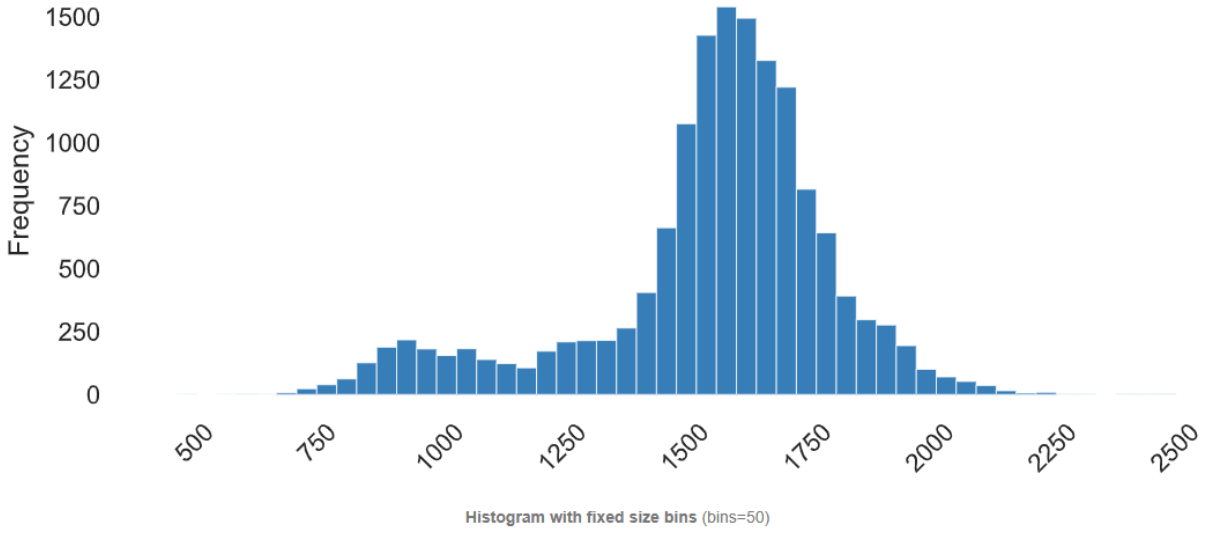
Quantile statistics (VALUE)

Minimum	**462**
5-th percentile	966.35
Q1	1480
median	**1593**
Q3	1699
95-th percentile	1892
Maximum	**2500**
Range	2038
Interquartile range (IQR)	219

Descriptive statistics (VALUE)

Standard deviation	252.78409
Coefficient of variation (CV)	0.16293554
Kurtosis	1.2934123
Mean	**1551.4361**
Median Absolute Deviation (MAD)	108
Skewness	**-1.0026593**
Sum	22787494
Variance	63899.795
Monotonicity	Not monotonic

Histogram (VALUE)



Average value is 1551.44 hours and Median value is 1593 hours. Skewed toward left.

Output #07b

For Average annual wages and salaries,

Dataset statistics

Number of variables	**8**
Number of observations	**14688**
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	1.0 MiB
Average record size in memory	72.0 B

Based on panda-profiling.

There are 14688 observations with 8 variables to work on.

Right now, only “Indicators” columns have one value.

Variable types

Numeric	2
Categorical	6

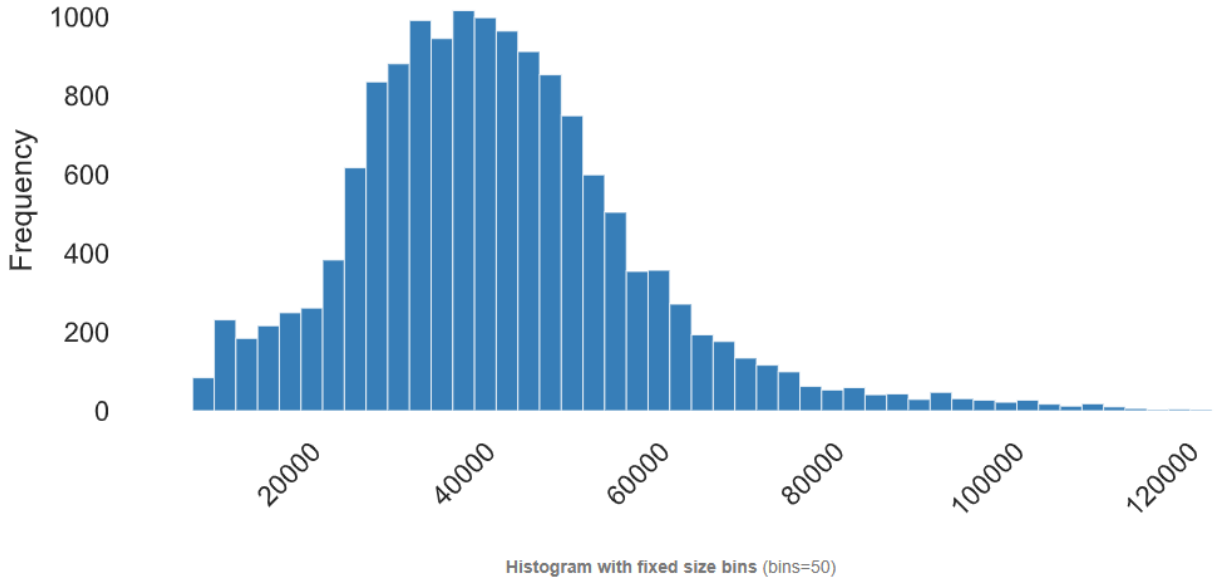
Quantile statistics (VALUE)

Minimum	**8769**
5-th percentile	18992.85
Q1	32900.5
median	**42186.5**
Q3	52319.75
95-th percentile	73362.6
Maximum	**133071**
Range	124302
Interquartile range (IQR)	19419.25

Descriptive statistics (VALUE)

Standard deviation	16620.351
Coefficient of variation (CV)	0.37941864
Kurtosis	1.9439325
Mean	**43804.783**
Median Absolute Deviation (MAD)	9684
Skewness	**0.91651225**
Sum	6.4340465 × 108
Variance	2.7623607 × 108
Monotonicity	Not monotonic

Histogram (VALUE)



The average for this section is $43804.79 and median is $42186.50. Skewed toward little right.

Output #07c

For Average hourly wages,

Dataset statistics

Number of variables	**8**
Number of observations	**14688**
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	1.0 MiB
Average record size in memory	72.0 B

Based on panda-profiling.

There are 14688 observations with 8 variables to work on.

Right now, only “Indicators” columns have one value.

Variable types

Numeric	2
Categorical	6

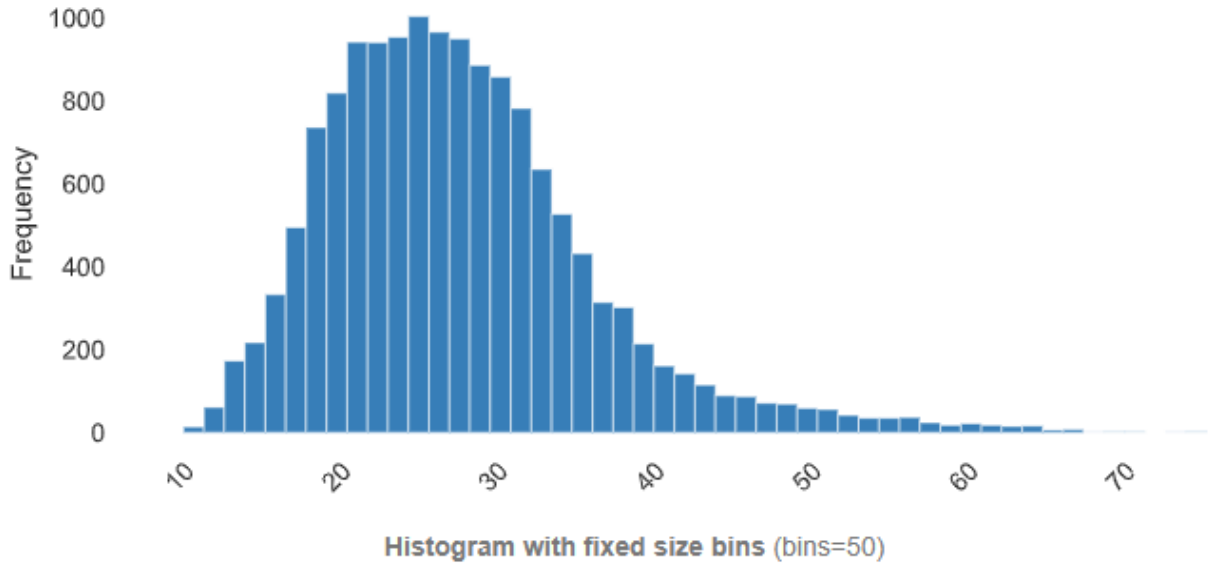
Quantile statistics (VALUE)

Minimum	**10.16**
5-th percentile	16.49
Q1	21.72
median	**26.7**
Q3	32.1
95-th percentile	43.803
Maximum	**75.37**
Range	65.21
Interquartile range (IQR)	10.38

Descriptive statistics (VALUE)

Standard deviation	8.6017205
Coefficient of variation (CV)	0.30912962
Kurtosis	2.1712535
Mean	**27.825611**
Median Absolute Deviation (MAD)	5.16
Skewness	**1.1323432**
Sum	408702.58
Variance	73.989596
Monotonicity	Not monotonic

Histogram (VALUE)



Output #07d

Average weekly hours worked,

Dataset statistics

Number of variables	**8**
Number of observations	**14688**
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	1.0 MiB
Average record size in memory	72.0 B

Based on panda-profiling.

There are 14688 observations with 8 variables to work on.

Right now, only “Indicators” columns have one value.

Variable types

Numeric	2
Categorical	6

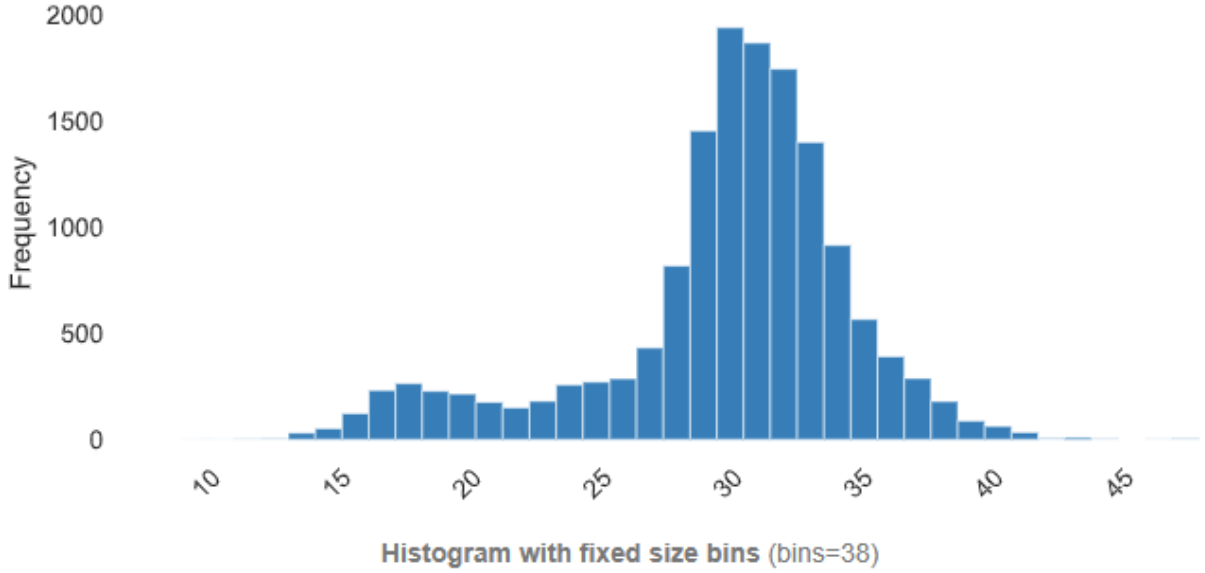
Quantile statistics (VALUE)

Minimum	**9**
5-th percentile	19
Q1	28
median	**31**
Q3	33
95-th percentile	36
Maximum	**48**
Range	39
Interquartile range (IQR)	5

Descriptive statistics (VALUE)

Standard deviation	4.8668895
Coefficient of variation (CV)	0.16314453
Kurtosis	1.2775213
Mean	**29.831767**
Median Absolute Deviation (MAD)	2
Skewness	**-0.9985335**
Sum	438169
Variance	23.686614
Monotonicity	Not monotonic

Histogram (VALUE)



Output #07e

Hours worked

Dataset statistics

Number of variables	**8**
Number of observations	**14688**
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	1.0 MiB
Average record size in memory	72.0 B

Variable types

Numeric	2
Categorical	6

Quantile statistics (VALUE)

Minimum	**6**
5-th percentile	128
Q1	1193.25
median	**9586.5**
Q3	55182.25
95-th percentile	377969.4
Maximum	**3857813**
Range	3857807
Interquartile range (IQR)	53989

Descriptive statistics (VALUE)

Standard deviation	253684.45
Coefficient of variation (CV)	3.0346138
Kurtosis	66.631407
Mean	**83596.946**
Median Absolute Deviation (MAD)	9323.5
Skewness	**7.1113237**
Sum	1.2278719 × 109
Variance	6.43558 × 1010
Monotonicity	Not monotonic

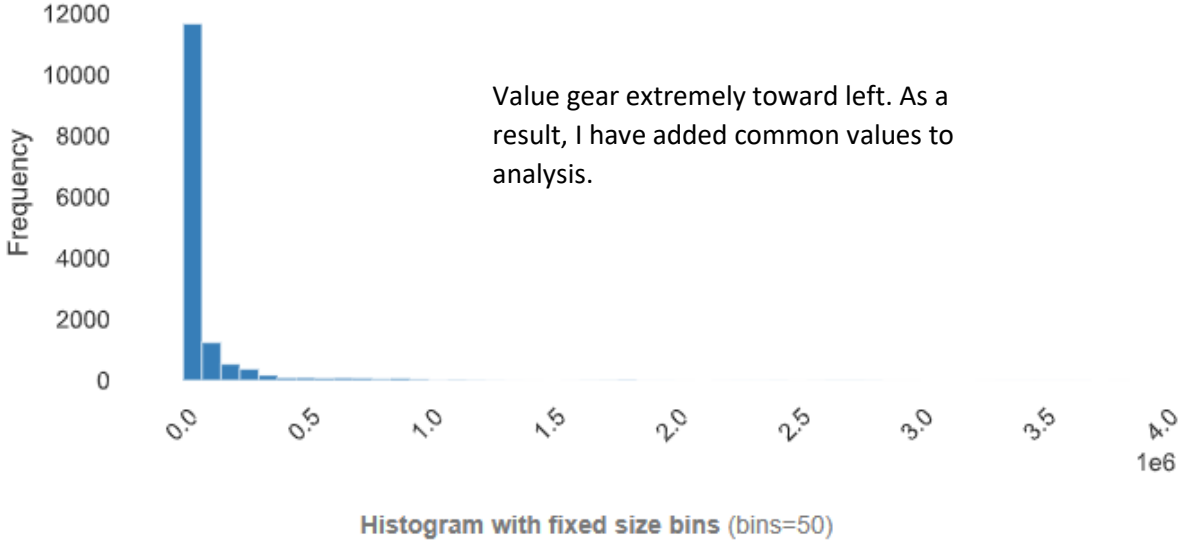
Based on panda-profiling.

There are 14688 observations with 8 variables to work on.

Right now, only “Indicators” columns have one value.

The data is skew toward left. I may not use this data afterward.

Histogram (VALUE)



Value gear extremely toward left. As a result, I have added common values to analysis.

Common Value for "hours worked".

Value	Count	Frequency (%)
37	20	0.1%
32	16	0.1%
29	14	0.1%
41	14	0.1%
54	13	0.1%
30	13	0.1%
143	12	0.1%
52	12	0.1%
262	11	0.1%
33	11	0.1%
Other values (10999)	14552	99.1%

I noticed in this section, there's a lot of extreme value in this dataset. Common values that are find in this dataset are range between 29 to 54 hours. Some worked for 6 hours while some worked at more than 370000 hours.

Output #07f

Number of Jobs

Dataset statistics

Number of variables	**8**
Number of observations	**14688**
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	1.0 MiB
Average record size in memory	72.0 B

Variable types

Numeric	2
Categorical	6

Quantile statistics (VALUE)

Minimum	**11**
5-th percentile	88
Q1	784
median	**6305.5**
Q3	35604.75
95-th percentile	243204.7
Maximum	**2428289**
Range	2428278
Interquartile range (IQR)	34820.75

Descriptive statistics (VALUE)

Standard deviation	161120.81
Coefficient of variation (CV)	3.0149252
Kurtosis	67.384586
Mean	**53441.062**
Median Absolute Deviation (MAD)	6127.5
Skewness	**7.1394116**
Sum	7.8494232 × 108
Variance	2.5959914 × 1010
Monotonicity	Not monotonic

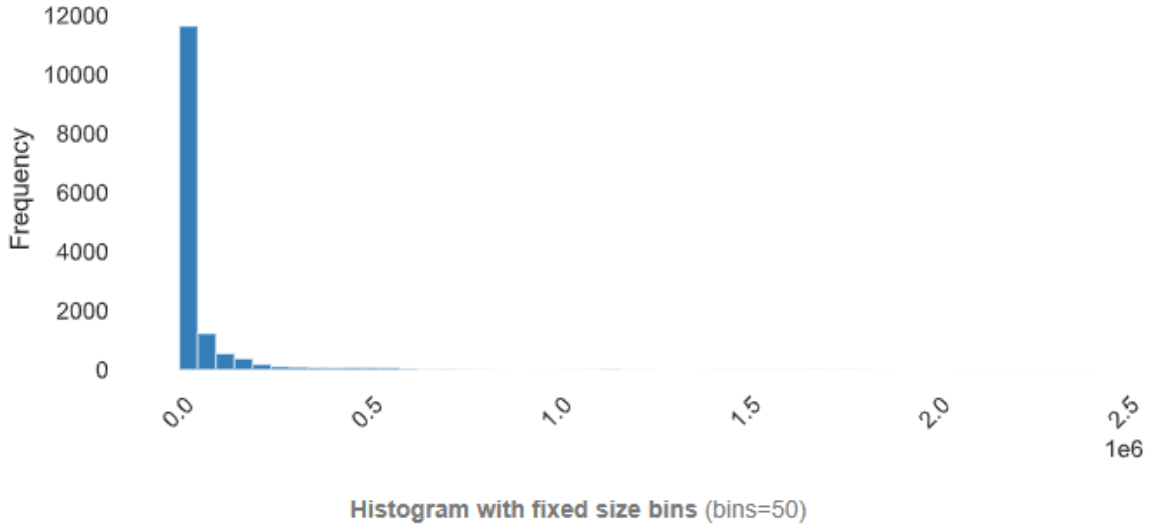
Based on panda-profiling.

There are 14688 observations with 8 variables to work on.

Right now, only “Indicators” columns have one value.

The data is skew toward left. For the purpose of this analysis, I will keep this value for time being.

Histogram (VALUE)



Common Values in VALUE

Value	Count	Frequency (%)
18	23	0.2%
36	21	0.1%
22	21	0.1%
23	20	0.1%
17	18	0.1%
129	17	0.1%
20	16	0.1%
25	15	0.1%
92	15	0.1%
29	15	0.1%
Other values (10278)	14507	98.8%

Common Values in the number of jobs in non-profit organizations range from 18 to 36 available.

Output #07g

Wages and Salaries

Dataset statistics

Number of variables	**8**
Number of observations	**14688**
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	1.0 MiB
Average record size in memory	72.0 B

Variable types

Numeric	2
Categorical	6

Quantile statistics (VALUE)

Minimum	**0**
5-th percentile	4
Q1	31
median	**224**
Q3	1455.25
95-th percentile	11603.3
Maximum	**132601**
Range	132601
Interquartile range (IQR)	1424.25

Descriptive statistics (VALUE)

Standard deviation	7977.4414
Coefficient of variation (CV)	3.2103436
Kurtosis	69.63355
Mean	**2484.9182**
Median Absolute Deviation (MAD)	217
Skewness	**7.3257394**
Sum	36498479
Variance	63639571
Monotonicity	Not monotonic

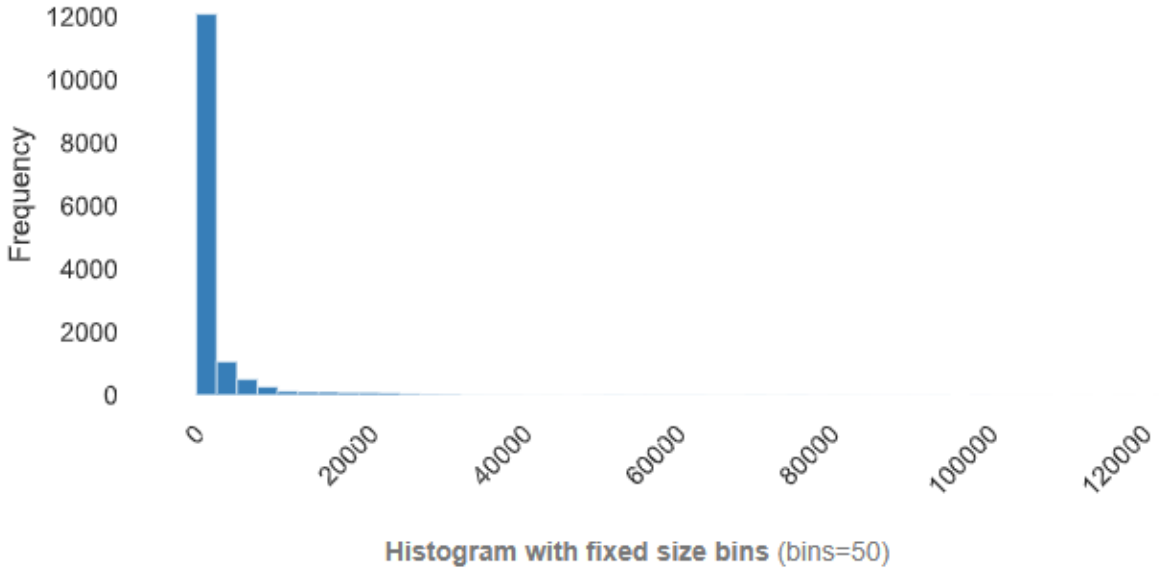
Based on panda-profiling.

There are 14688 observations with 8 variables to work on.

Right now, only “Indicators” columns have one value.

The data is skew toward left. I may not use this data afterward.

Histogram (VALUE)



Common values inside Wage and Salaries

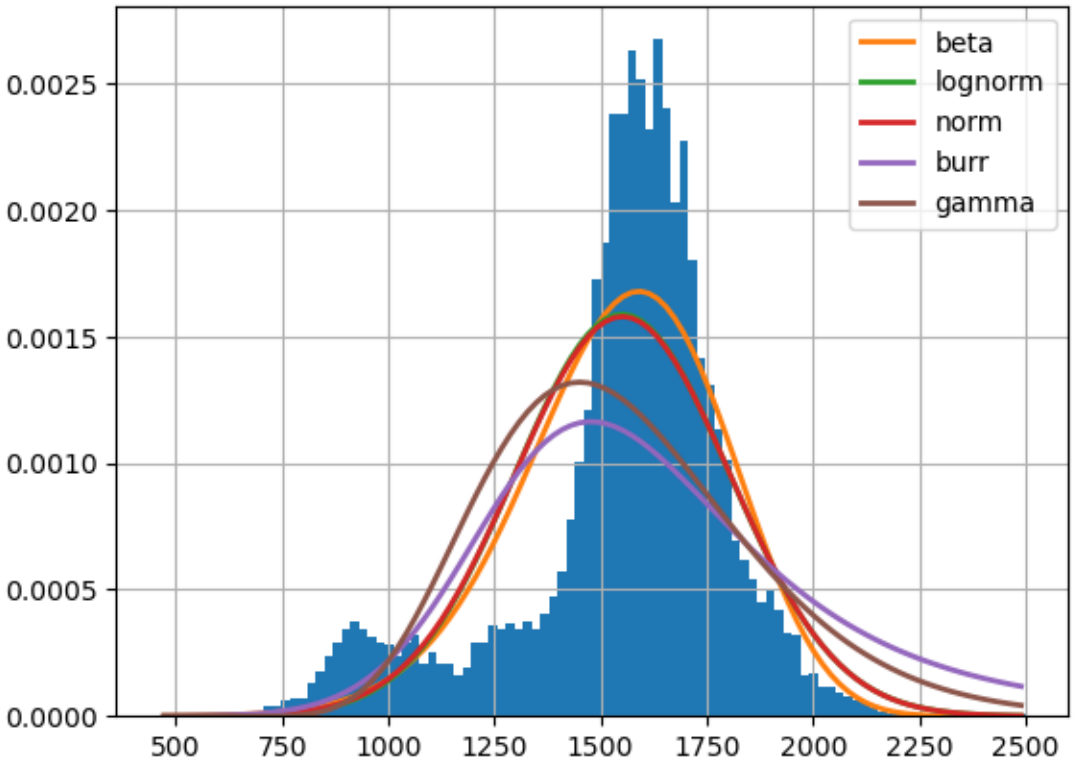
Value	Count	Frequency (%)
2	299	2.0%
1	219	1.5%
4	203	1.4%
3	201	1.4%
8	179	1.2%
5	157	1.1%
12	148	1.0%
6	146	1.0%
9	143	1.0%
11	141	1.0%
Other values (4328)	12852	87.5%

The most common wages and salaries in non-profit organizations range between $1 to $12.

Output #08a

Best distribution for "Average annual hours worked"

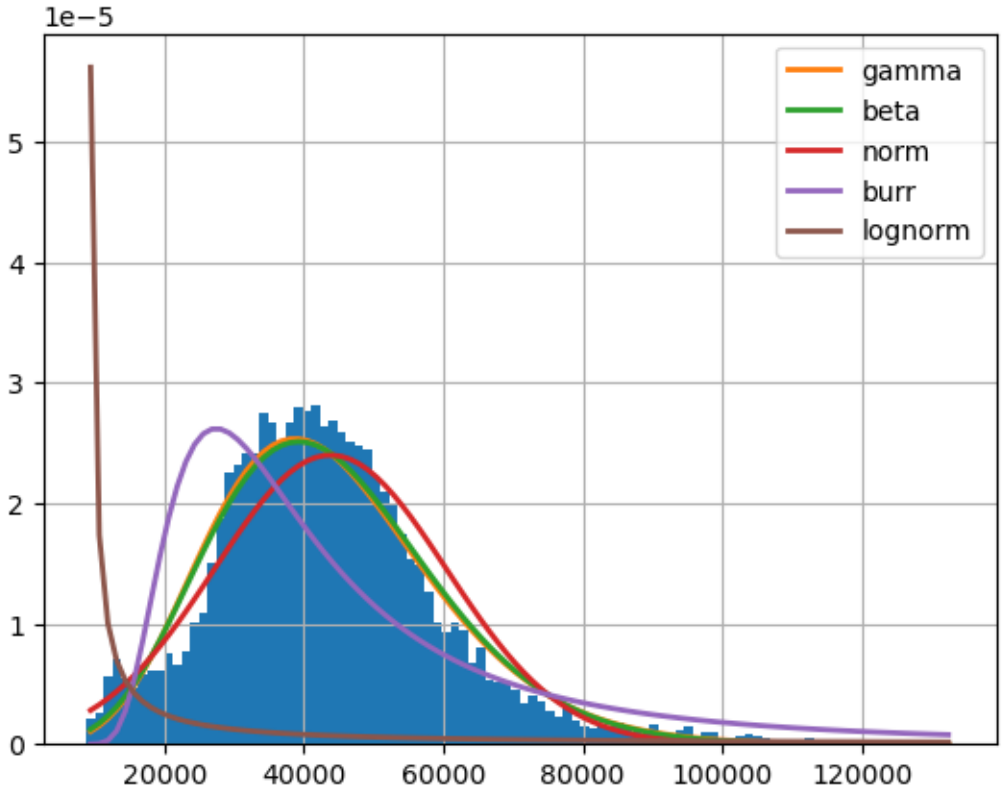
```
{'beta': {'a': 638.3317504246395,
  'b': 31.78242716784456,
  'loc': -26340.79799985259,
  'scale': 29280.128191604977}}
```



Output #08b

Best distribution for "Average annual wages and salaries"

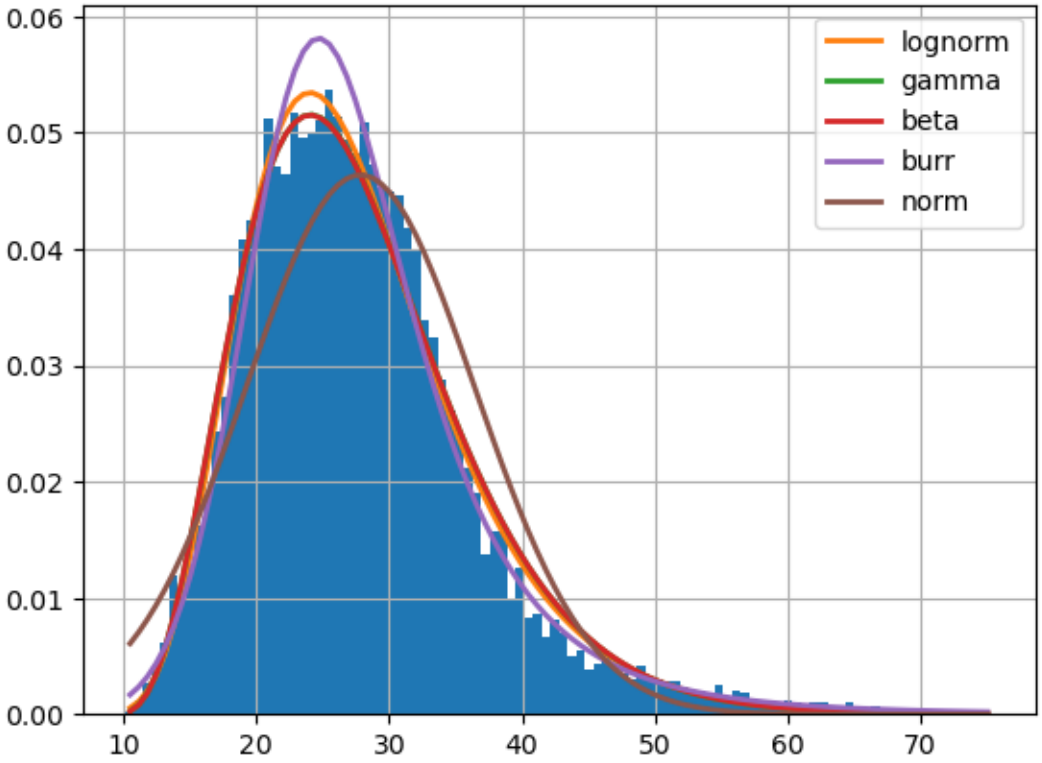
```
{'gamma': {'a': 10.317012968188816,
  'loc': -8874.841101131682,
  'scale': 5106.09266888318}}
```



Output #08c

Best distribution for "Average hourly wage"

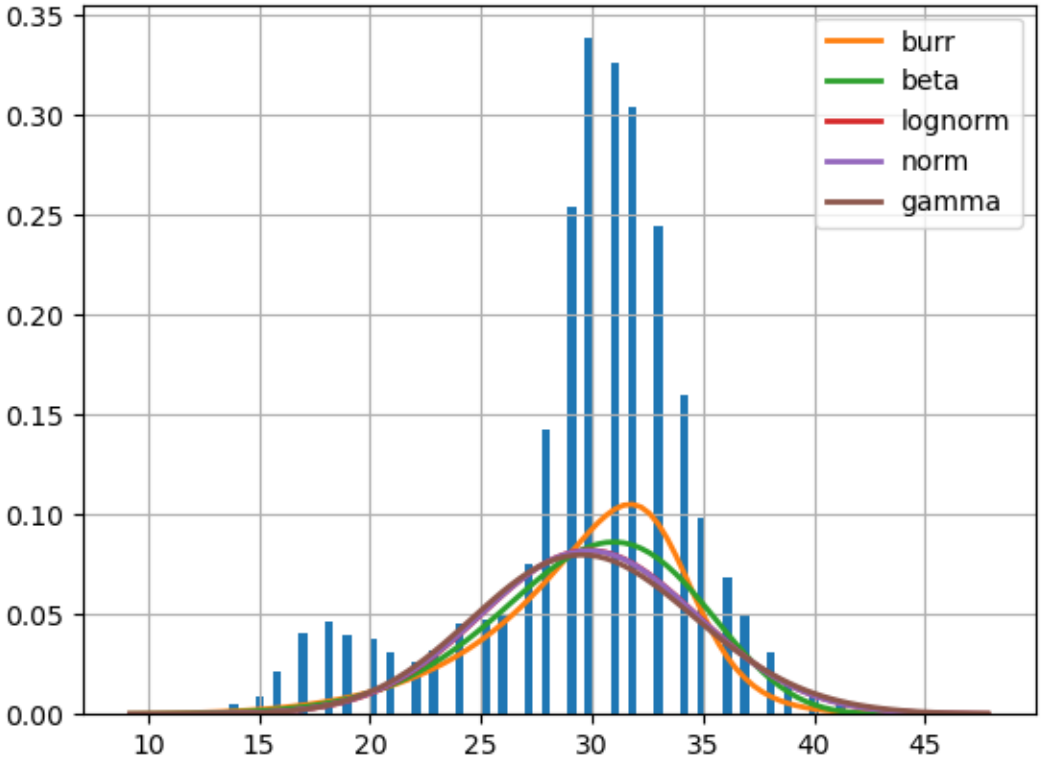
```
{'lognorm': {'s': 0.32868880665290323,
  'loc': 2.5266200749710075,
  'scale': 23.96526012956064}}
```



Output #08d

Best distribution for "Average weekly hours worked"

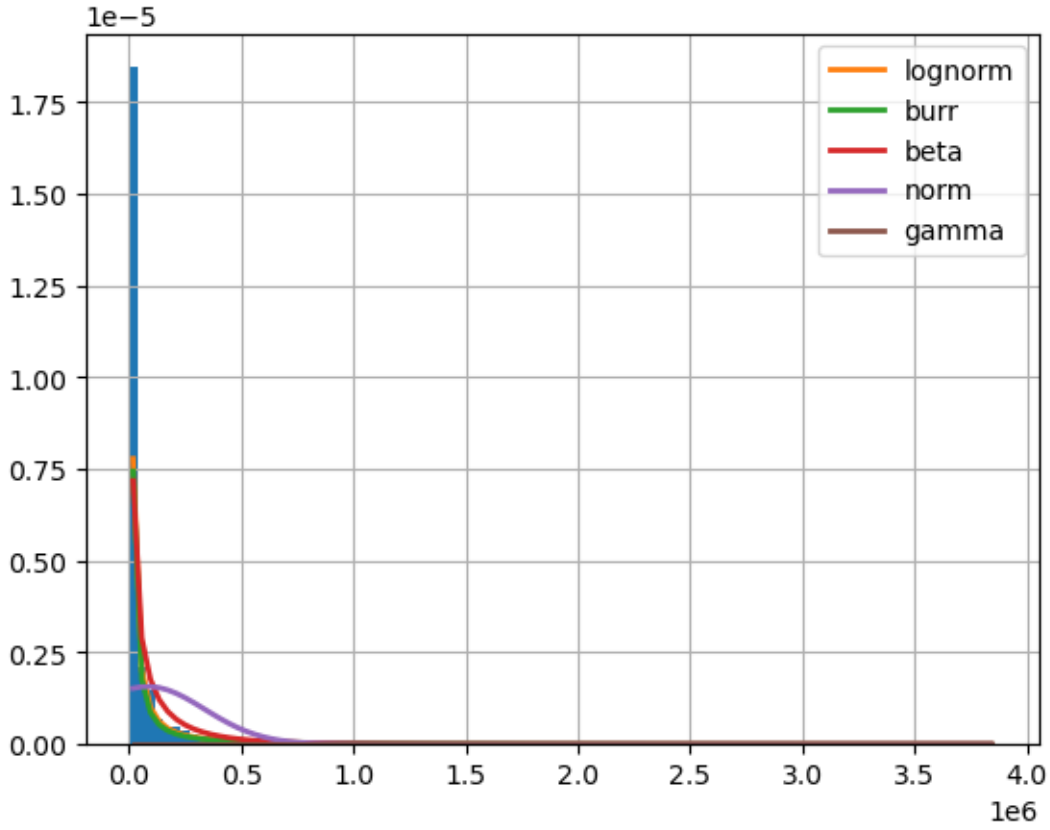
```
{'burr': {'c': 24.425610086588648,
  'd': 0.2639474187210325,
  'loc': -0.19516675888992813,
  'scale': 33.9923544234202}}
```



Output #08e

Best distribution for "Hours Worked"

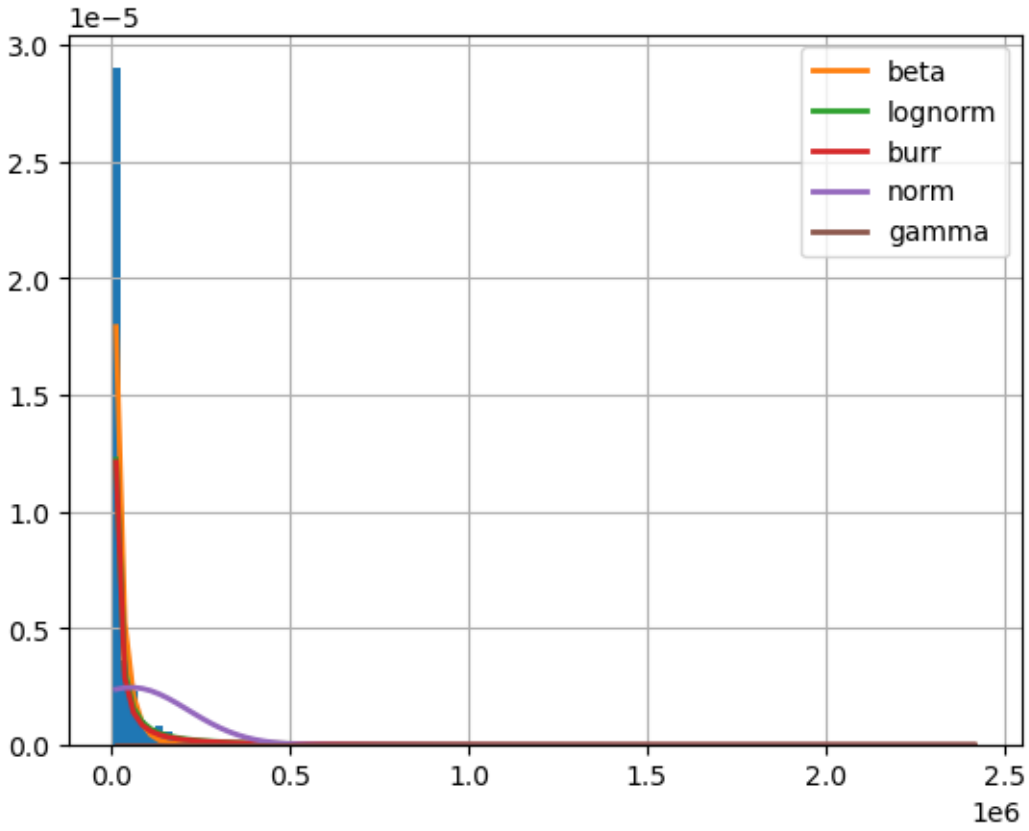
```
{'lognorm': {'s': 2.500328743903225,
  'loc': 5.890210931623885,
  'scale': 8052.849839197419}}
```



Output #08f

Best distrubution for "Number of jobs"

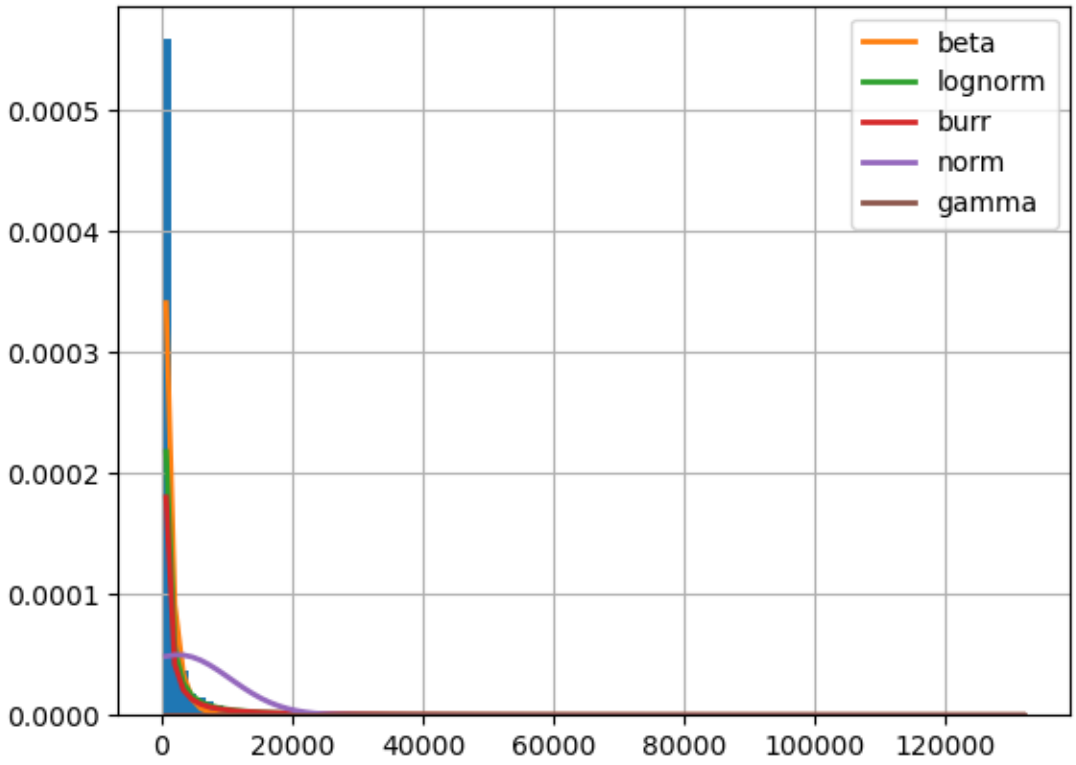
```
{'beta': {'a': 0.44093091654590877,
  'b': 478.35910204629215,
  'loc': 10.999999999999998,
  'scale': 18916817.838308237}}
```



Output #08g

Best distribution for "Wages and Salaries"

```
{'beta': {'a': 0.45806730861492995,
  'b': 504.0982556396307,
  'loc': -3.462059833380747e-23,
  'scale': 927559.1543601784}}
```



Output #09

Average annual hours worked (First one being Training Set, Second one being Testing Set)

```
                sum  size
REF_DATE
2013      1906851.0  1224
2014      1898034.0  1224
2015      1907286.0  1224
2016      1899738.0  1224
2017      1881389.0  1224
2018      1894227.0  1224

                sum  size
REF_DATE
2019      1894126.0  1224
2020      1873544.0  1224
2021      1901826.0  1224
```

Average annual wages and salaries

```
                 sum  size
REF_DATE
2013      50598135.0  1224
2014      51805889.0  1224
2015      52715143.0  1224
2016      53166285.0  1224
2017      53965359.0  1224
2018      55525920.0  1224

                 sum  size
REF_DATE
2019      56997121.0  1224
2020      60597775.0  1224
2021      60687848.0  1224
```

Average hourly wage

```
               sum  size
REF_DATE
2013      31932.19  1224
2014      32916.44  1224
2015      33362.79  1224
2016      33756.84  1224
2017      34571.41  1224
2018      35356.16  1224

               sum  size
REF_DATE
2019      36293.34  1224
2020      38980.38  1224
2021      38514.73  1224
```

Average weekly hours worked

REF_DATE	sum	size
2013	36671.0	1224
2014	36483.0	1224
2015	36678.0	1224
2016	36541.0	1224
2017	36170.0	1224
2018	36416.0	1224

REF_DATE	sum	size
2019	36400.0	1224
2020	36036.0	1224
2021	36555.0	1224

Hours Worked

REF_DATE	sum	size
2013	98935086.0	1224
2014	99777902.0	1224
2015	101927894.0	1224
2016	103980992.0	1224
2017	103906357.0	1224
2018	106885614.0	1224

REF_DATE	sum	size
2019	108384508.0	1224
2020	103897606.0	1224
2021	112280627.0	1224

Number of jobs

REF_DATE	sum	size
2013	63161481.0	1224
2014	63980381.0	1224
2015	65239031.0	1224
2016	66578386.0	1224
2017	66979678.0	1224
2018	68215341.0	1224

REF_DATE	sum	size
2019	69702012.0	1224
2020	67249987.0	1224
2021	70971515.0	1224

Wages and Salaries

REF_DATE	sum	size
2013	2746199.0	1224
2014	2844225.0	1224
2015	2965770.0	1224
2016	3043217.0	1224
2017	3118809.0	1224
2018	3263527.0	1224

REF_DATE	sum	size
2019	3437480.0	1224
2020	3532888.0	1224
2021	3764252.0	1224

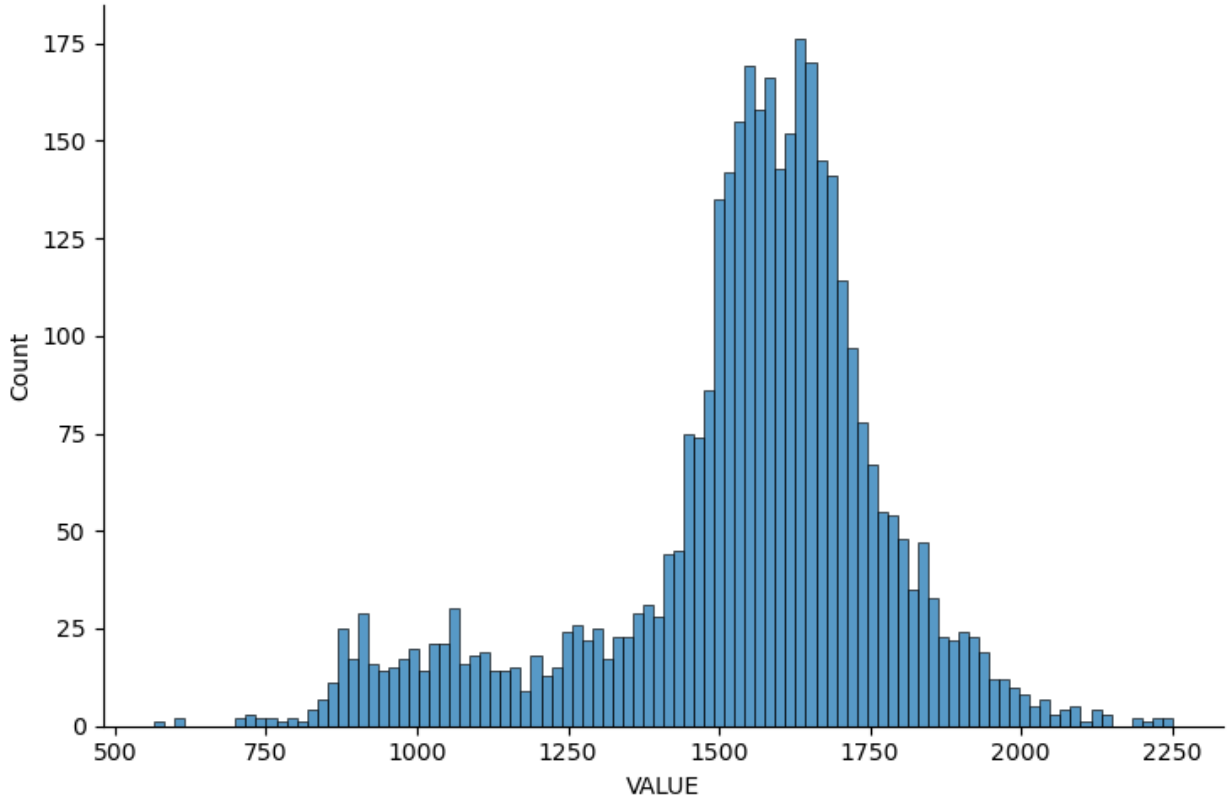
Output #10

For Average Annual hour worked,

```
testing set
                                     sum         mean    amin  median  \
Characteristics
15 to 24 years                    179872.0   936.833333   713.0   927.5
25 to 34 years                    333662.0  1588.866667  1292.0  1576.0
35 to 44 years                    371386.0  1768.504762  1424.0  1757.0
45 to 54 years                    384987.0  1833.271429  1541.0  1826.5
55 to 64 years                    351843.0  1675.442857  1377.0  1675.0
65 years old and over             206760.0  1076.875000   565.0  1076.5
College diploma                   348154.0  1657.876190  1430.0  1645.0
Female employees                  323755.0  1541.690476  1302.0  1540.0
High school diploma and less      275636.0  1312.552381  1054.0  1307.5
Immigrant employees               318818.0  1610.191919  1336.0  1589.0
Indigenous identity employees     293916.0  1484.424242  1170.0  1493.5
Male employees                    343101.0  1633.814286  1373.0  1644.0
Non-immigrant employees           310245.0  1566.893939  1315.0  1570.0
Non-indigenous identity employees 321043.0  1621.429293  1422.0  1596.5
Not a visible minority            342351.0  1630.242857  1418.0  1604.5
Trade certificate                 306174.0  1546.333333   789.0  1547.5
University degree and higher      341795.0  1726.237374  1536.0  1706.5
Visible minority                  315998.0  1504.752381  1256.0  1519.0

                                    amax  size
Characteristics
15 to 24 years                    1281.0   192
25 to 34 years                    1870.0   210
35 to 44 years                    2092.0   210
45 to 54 years                    2191.0   210
55 to 64 years                    2071.0   210
65 years old and over             1415.0   192
College diploma                   1969.0   210
Female employees                  1773.0   210
High school diploma and less      1667.0   210
Immigrant employees               2250.0   198
Indigenous identity employees     1794.0   198
Male employees                    1821.0   210
Non-immigrant employees           1767.0   198
Non-indigenous identity employees 2120.0   198
Not a visible minority            2250.0   210
Trade certificate                 1808.0   198
University degree and higher      2043.0   198
Visible minority                  1709.0   210
```

Overall,
Sum : 5669496.0
Mean : 1543.9803921568628
Min/median/max : 565.0 / 1583.0 / 2250.0
Standard Deviation : 242.19188733505
Skewnewss : -0.9940967289246635
Total size : 3672

Histogram for testing dataset

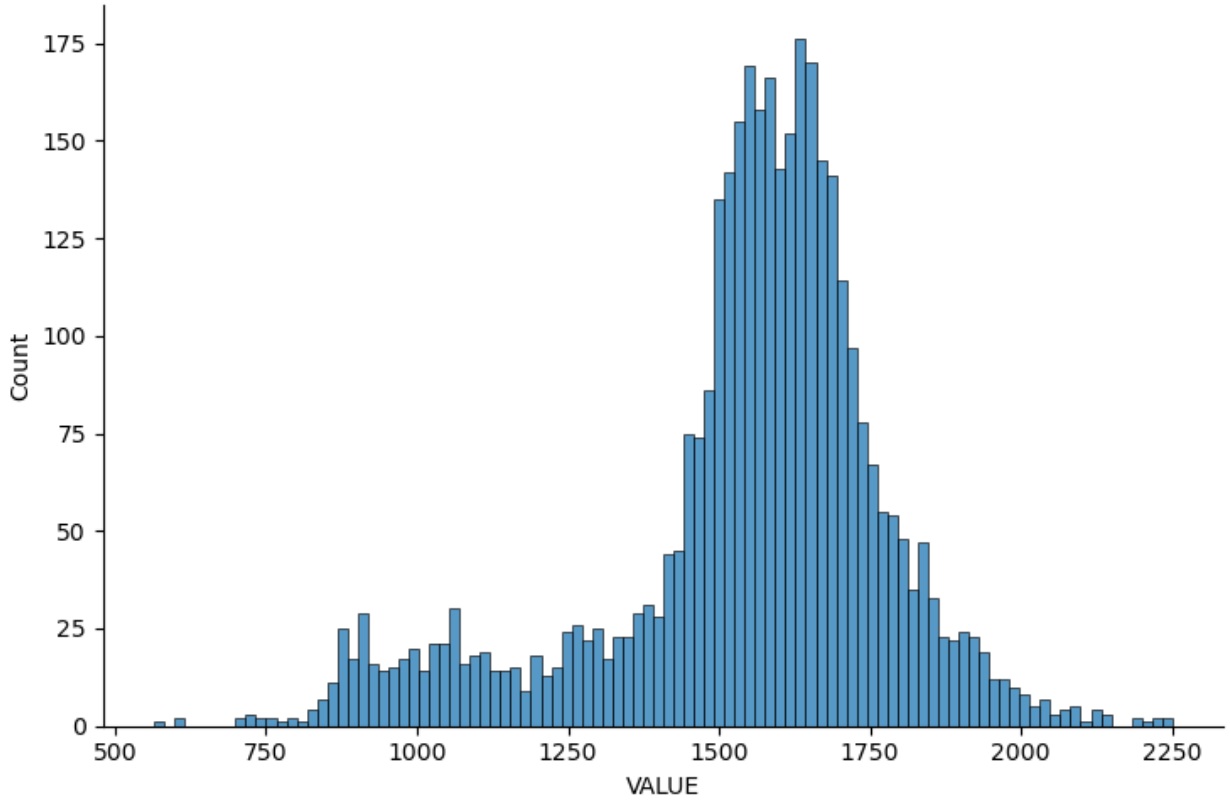


Final output for "Average annual wages and salaries"

```
testing set
                                        sum          mean     amin   median  \
Characteristics
15 to 24 years                       3482308.0  18137.020833  11093.0  16653.5
25 to 34 years                       9108041.0  43371.623810  26534.0  42901.0
35 to 44 years                      12296655.0  58555.500000  37336.0  56960.5
45 to 54 years                      13595377.0  64739.890476  39455.0  64274.5
55 to 64 years                      12592232.0  59963.009524  33296.0  56266.0
65 years old and over                6840988.0  35630.145833  18187.0  34170.0
College diploma                     10568980.0  50328.476190  30498.0  47690.5
Female employees                     9929829.0  47284.900000  28080.0  45973.5
High school diploma and less         6915927.0  32932.985714  17828.0  31071.0
Immigrant employees                  9838690.0  49690.353535  23144.0  48933.5
Indigenous identity employees        8530658.0  43084.131313  24567.0  41643.0
Male employees                      11755264.0  55977.447619  32342.0  55938.0
Non-immigrant employees              9799014.0  49489.969697  29956.0  49313.5
Non-indigenous identity employees   10861810.0  54857.626263  30440.0  51036.0
Not a visible minority              11677984.0  55609.447619  29824.0  51485.5
Trade certificate                    8609637.0  43483.015152  25221.0  42431.0
University degree and higher        12581443.0  63542.641414  38860.0  62574.0
Visible minority                     9297907.0  44275.747619  23792.0  43851.0

                                       amax  size
Characteristics
15 to 24 years                      45844.0   192
25 to 34 years                      75429.0   210
35 to 44 years                      95714.0   210
45 to 54 years                     103580.0   210
55 to 64 years                     133071.0   210
65 years old and over               76577.0   192
College diploma                    100421.0   210
Female employees                    86432.0   210
High school diploma and less        73020.0   210
Immigrant employees                 98074.0   198
Indigenous identity employees       78090.0   198
Male employees                      95273.0   210
Non-immigrant employees             85472.0   198
Non-indigenous identity employees  115680.0   198
Not a visible minority             123100.0   210
Trade certificate                   72854.0   198
University degree and higher       121264.0   198
Visible minority                    78639.0   210
```

Overall,
Sum : 178282744.0
Mean : 48551.94553376906
Min/median/max : 11093.0 / 46961.0 / 133071.0
Standard Deviation : 17487.878030435306
Skewnewss : 0.8996131427783454
Total size : 3672

Histogram for testing dataset

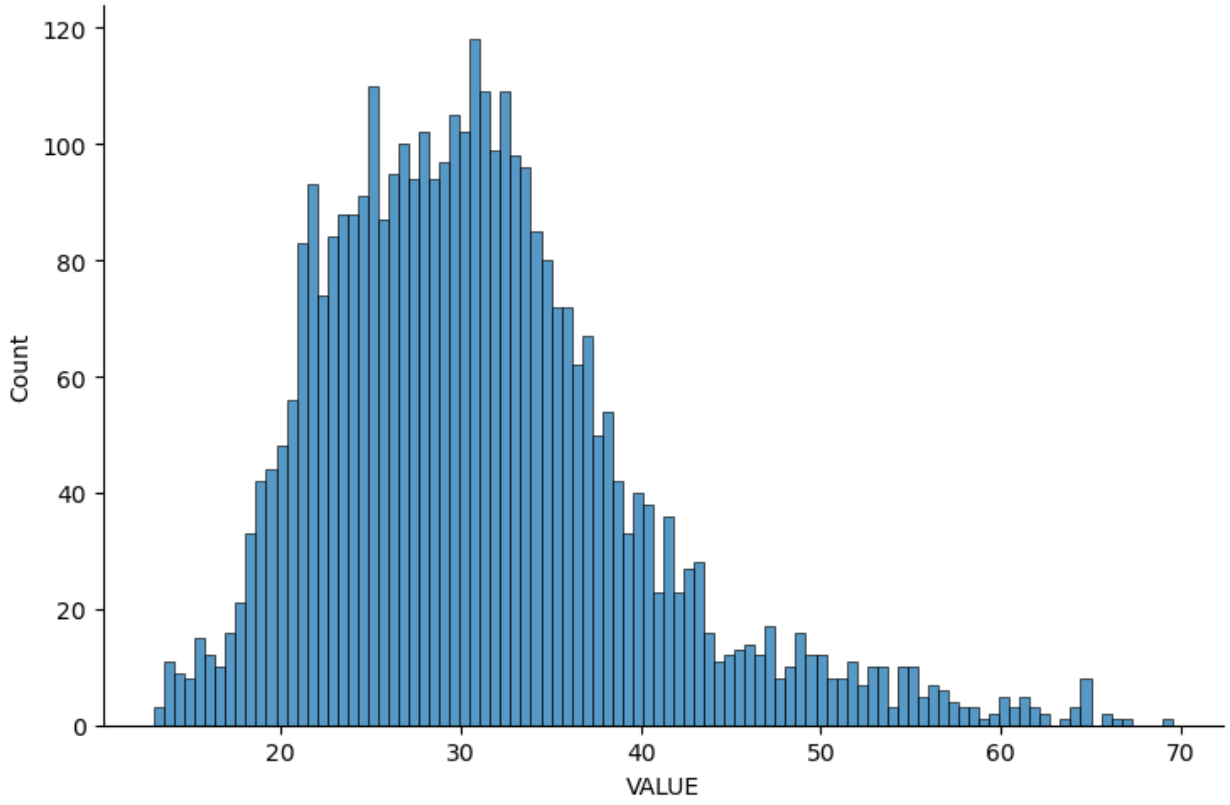


Final output for "Average hourly wage"

```
testing set
                                    sum       mean   amin  median   amax  \
Characteristics
15 to 24 years                   3689.12  19.214167  12.95  18.775  36.37
25 to 34 years                   5759.41  27.425762  17.01  27.010  45.91
35 to 44 years                   6991.24  33.291619  19.42  32.680  58.12
45 to 54 years                   7451.95  35.485476  21.54  34.965  64.54
55 to 64 years                   7475.92  35.599619  19.42  33.800  64.92
65 years old and over            6329.58  32.966563  18.32  32.140  66.63
College diploma                  6384.21  30.401000  17.61  28.755  57.48
Female employees                 6439.32  30.663429  18.34  30.200  55.06
High school diploma and less     5242.44  24.964000  15.69  23.600  44.52
Immigrant employees              6065.65  30.634596  16.93  29.925  60.18
Indigenous identity employees    5758.93  29.085505  17.34  28.020  51.61
Male employees                   7203.15  34.300714  19.31  33.585  56.59
Non-immigrant employees          6267.15  31.652273  19.24  31.000  55.32
Non-indigenous identity employees 6621.55 33.442172  19.43  32.345  63.81
Not a visible minority           7073.14  33.681619  19.19  32.220  64.48
Trade certificate                5569.30  28.127778  17.63  27.370  44.58
University degree and higher     7296.08  36.848889  21.80  36.385  69.52
Visible minority                 6170.31  29.382429  16.81  28.785  52.49

                                 size
Characteristics
15 to 24 years                    192
25 to 34 years                    210
35 to 44 years                    210
45 to 54 years                    210
55 to 64 years                    210
65 years old and over             192
College diploma                   210
Female employees                  210
High school diploma and less      210
Immigrant employees               198
Indigenous identity employees     198
Male employees                    210
Non-immigrant employees           198
Non-indigenous identity employees 198
Not a visible minority            210
Trade certificate                 198
University degree and higher      198
Visible minority                  210
```

Overall,
Sum : 113788.45
Mean : 30.988139978213507
Min/median/max : 12.95 / 30.08 / 69.52
Standard Deviation : 8.792991378623405
Skewnewss : 0.9805031115064706
Total size : 3672

Histgram for testing dataset



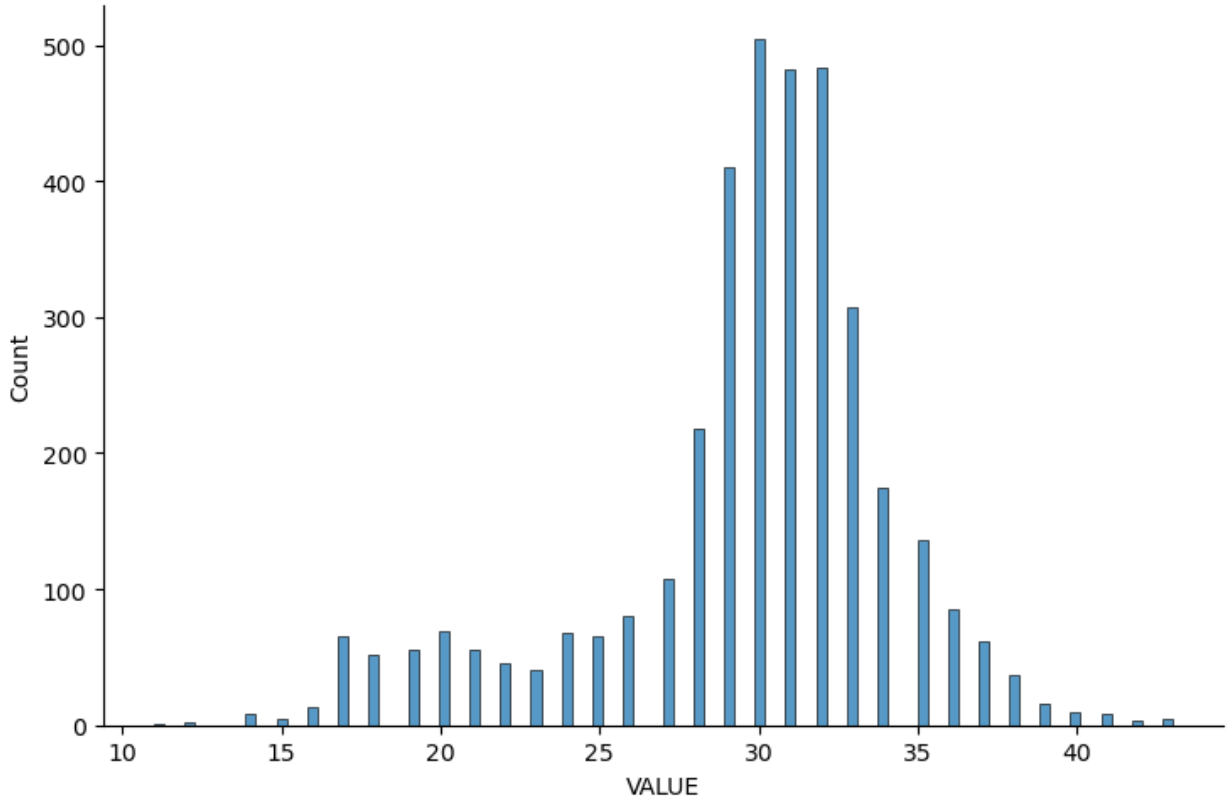
Final output for "Average weekly hours worked"

testing set

Characteristics	sum	mean	amin	median	amax	size
15 to 24 years	3465.0	18.046875	14.0	18.0	25.0	192
25 to 34 years	6413.0	30.538095	25.0	30.0	36.0	210
35 to 44 years	7139.0	33.995238	27.0	34.0	40.0	210
45 to 54 years	7406.0	35.266667	30.0	35.0	42.0	210
55 to 64 years	6765.0	32.214286	26.0	32.0	40.0	210
65 years old and over	3969.0	20.671875	11.0	21.0	27.0	192
College diploma	6691.0	31.861905	28.0	32.0	38.0	210
Female employees	6219.0	29.614286	25.0	30.0	34.0	210
High school diploma and less	5303.0	25.252381	20.0	25.0	32.0	210
Immigrant employees	6133.0	30.974747	26.0	31.0	43.0	198
Indigenous identity employees	5652.0	28.545455	22.0	29.0	34.0	198
Male employees	6590.0	31.380952	26.0	32.0	35.0	210
Non-immigrant employees	5961.0	30.106061	25.0	30.0	34.0	198
Non-indigenous identity employees	6169.0	31.156566	27.0	31.0	41.0	198
Not a visible minority	6583.0	31.347619	27.0	31.0	43.0	210
Trade certificate	5889.0	29.742424	15.0	30.0	35.0	198
University degree and higher	6568.0	33.171717	30.0	33.0	39.0	198
Visible minority	6076.0	28.933333	24.0	29.0	33.0	210

Overall,
Sum : 108991.0
Mean : 29.681644880174293
Min/median/max : 11.0 / 30.0 / 43.0
Standard Deviation : 4.658221901251303
Skewnewss : -0.9931590104885496
Total size : 3672

Histogram for testing dataset



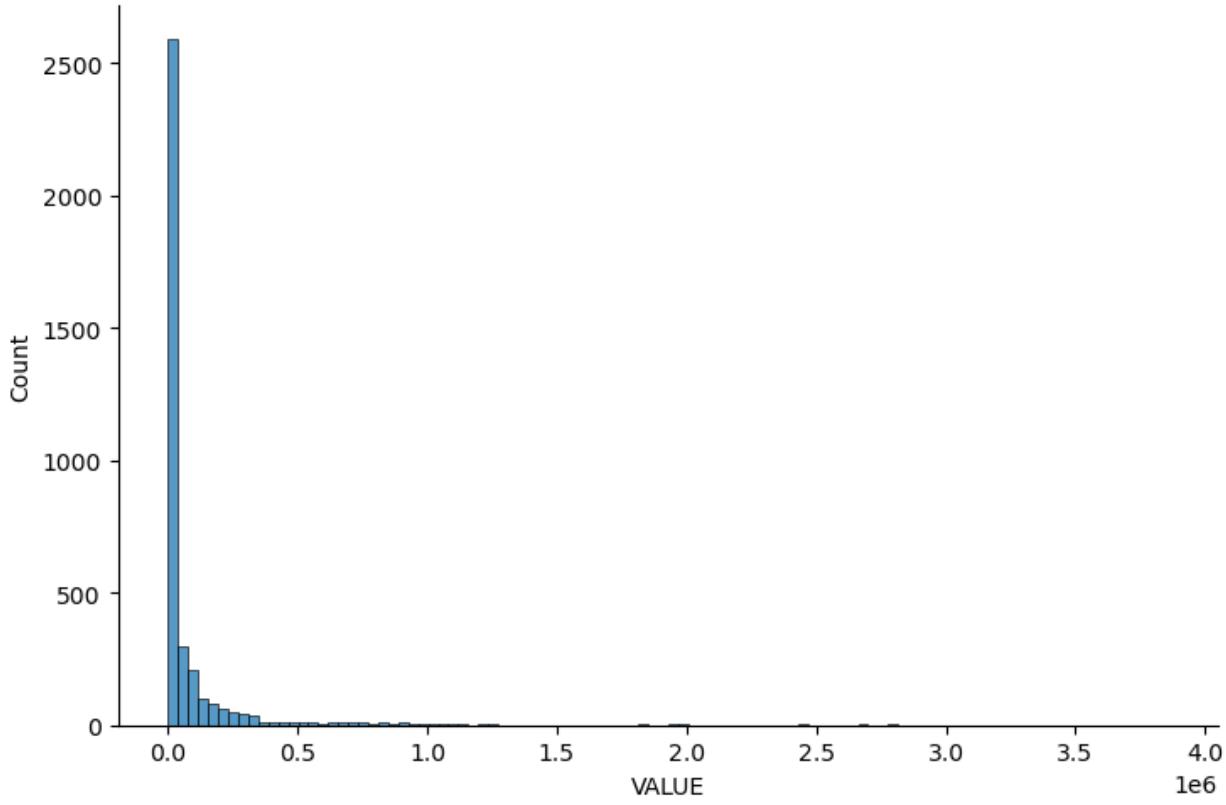
Final output for "Hours Worked"

testing set

Characteristics	sum	mean	amin	median	amax	size
15 to 24 years	3132171.0	16313.390625	29.0	3859.5	228566.0	192
25 to 34 years	11857576.0	56464.647619	32.0	10166.5	888764.0	210
35 to 44 years	12919606.0	61521.933333	33.0	10825.0	971688.0	210
45 to 54 years	12974580.0	61783.714286	31.0	10600.5	966480.0	210
55 to 64 years	10690620.0	50907.714286	25.0	8934.0	794712.0	210
65 years old and over	2528517.0	13169.359375	30.0	2913.5	183461.0	192
College diploma	13973668.0	66541.276190	31.0	10419.5	1049353.0	210
Female employees	37193214.0	177110.542857	64.0	28332.5	2809860.0	210
High school diploma and less	9285900.0	44218.571429	61.0	11984.5	658441.0	210
Immigrant employees	14355944.0	72504.767677	44.0	7122.0	1084033.0	198
Indigenous identity employees	2394002.0	12090.919192	37.0	3683.5	175862.0	198
Male employees	16912308.0	80534.800000	74.0	15656.5	1223817.0	210
Non-immigrant employees	39740524.0	200709.717172	264.0	48444.5	2949643.0	198
Non-indigenous identity employees	51656293.0	260890.368687	47.0	50464.0	3857813.0	198
Not a visible minority	39228544.0	186802.590476	39.0	30749.0	2906803.0	210
Trade certificate	3449203.0	17420.217172	20.0	3080.0	260168.0	198
University degree and higher	27393105.0	138349.015152	96.0	23963.5	2065714.0	198
Visible minority	14876966.0	70842.695238	52.0	8041.5	1126871.0	210

Overall,
Sum : 324562741.0
Mean : 88388.54602396514
Min/median/max : 20.0 / 10195.0 / 3857813.0
Standard Deviation : 267448.25074811914
Skewnewss : 7.058784198645266
Total size : 3672

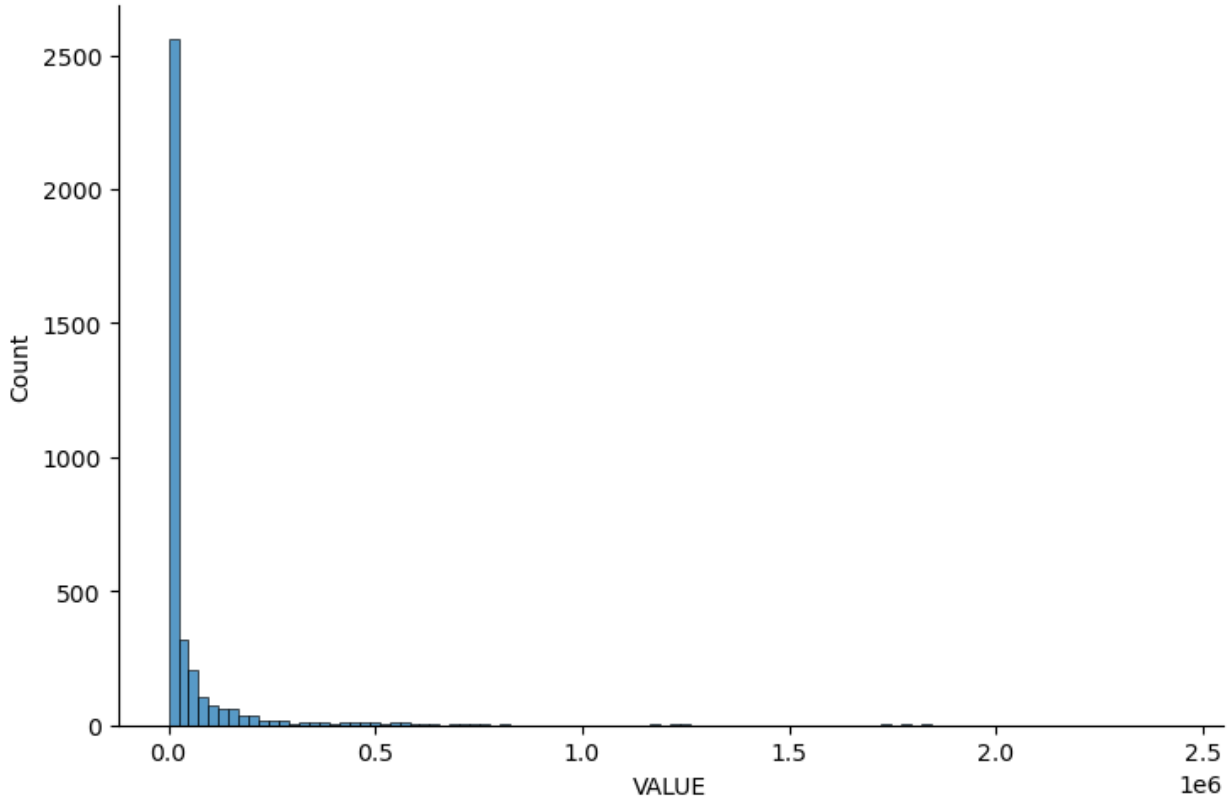
Histogram for testing dataset



Final output for "Number of jobs"

```
testing set
                                        sum           mean   amin   median  \
Characteristics
15 to 24 years                    3414206.0   17782.322917   24.0   4022.0
25 to 34 years                    7615612.0   36264.819048   19.0   6267.5
35 to 44 years                    7559436.0   35997.314286   18.0   6322.0
45 to 54 years                    7201656.0   34293.600000   16.0   5847.0
55 to 64 years                    6451160.0   30719.809524   13.0   5371.5
65 years old and over             2416851.0   12587.765625   25.0   2729.0
College diploma                   8695526.0   41407.266667   17.0   6300.0
Female employees                 24112770.0  114822.714286   37.0  19062.0
High school diploma and less      7204888.0   34308.990476   40.0   8564.5
Immigrant employees               9133147.0   46127.005051   22.0   4384.0
Indigenous identity employees     1591054.0    8035.626263   24.0   2443.5
Male employees                   10548592.0   50231.390476   43.0   9306.5
Non-immigrant employees          25522127.0  128899.631313  163.0  30374.0
Non-indigenous identity employees 33036343.0 166850.217172   23.0  31157.0
Not a visible minority           24941356.0  118768.361905   19.0  20138.5
Trade certificate                 2245738.0   11342.111111   13.0   1865.5
University degree and higher     16513042.0   83399.202020   47.0  13812.5
Visible minority                  9720010.0   46285.761905   33.0   5120.5

                                      amax  size
Characteristics
15 to 24 years                    247960.0   192
25 to 34 years                    564788.0   210
35 to 44 years                    561685.0   210
45 to 54 years                    529125.0   210
55 to 64 years                    472689.0   210
65 years old and over             173157.0   192
College diploma                   642955.0   210
Female employees                 1788569.0   210
High school diploma and less      506866.0   210
Immigrant employees               677236.0   198
Indigenous identity employees     115100.0   198
Male employees                    754820.0   210
Non-immigrant employees          1866156.0   198
Non-indigenous identity employees 2428289.0  198
Not a visible minority           1821295.0   210
Trade certificate                 167537.0   198
University degree and higher     1230063.0   198
Visible minority                  722093.0   210
```

```
Overall,
Sum :  207923514.0
Mean :  56624.05065359477
Min/median/max : 13.0 / 6759.5 / 2428289.0
Standard Deviation :  169877.4778961228
Skewnewss :  7.077329774066239
Total size :  3672

Histogram for testing dataset
```

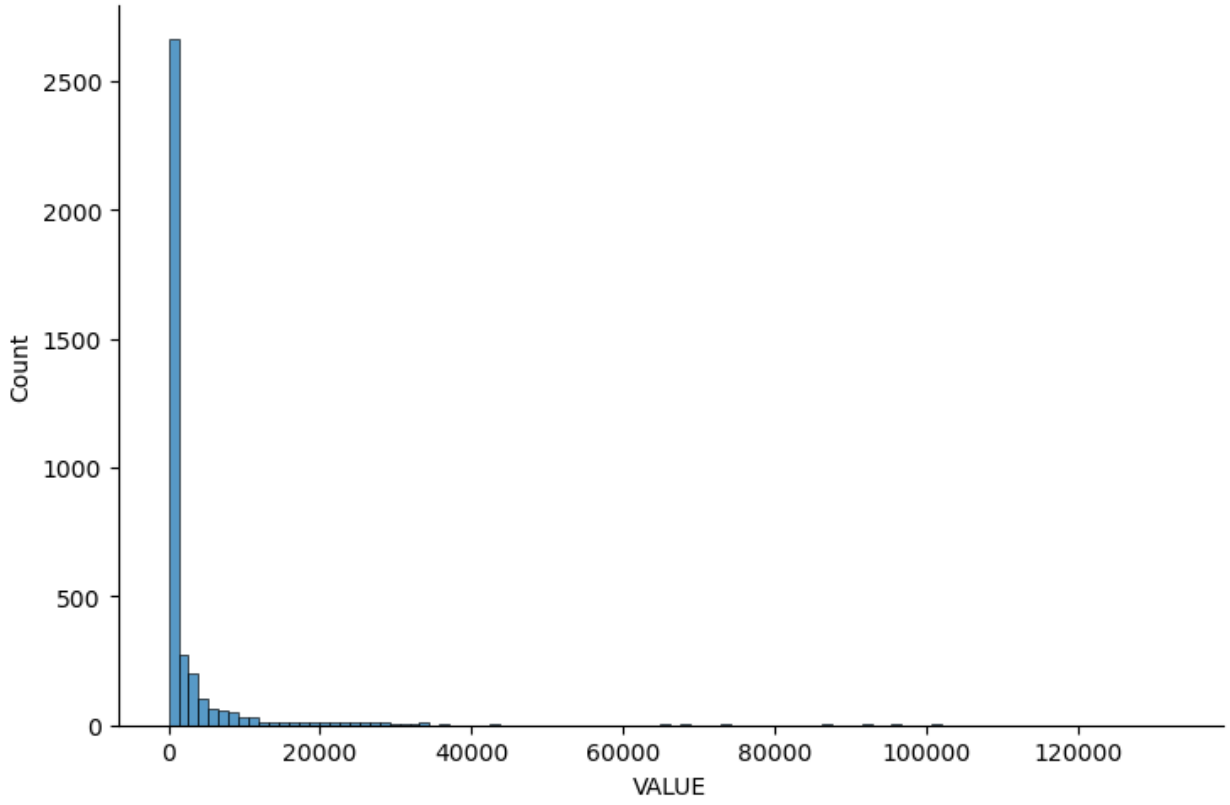


Final output for "Wages and Salaries"

```
testing set
                                     sum         mean  amin  median  \
Characteristics
15 to 24 years                     63069.0   328.484375   1.0    69.0
25 to 34 years                    334200.0  1591.428571   1.0   236.5
35 to 44 years                    441049.0  2100.233333   2.0   314.5
45 to 54 years                    474202.0  2258.104762   1.0   317.0
55 to 64 years                    386319.0  1839.614286   1.0   260.5
65 years old and over              90513.0   471.421875   1.0    88.0
College diploma                   416441.0  1983.052381   1.0   264.0
Female employees                 1179670.0  5617.476190   3.0   725.5
High school diploma and less      226638.0  1079.228571   2.0   260.5
Immigrant employees               473697.0  2392.409091   1.0   216.5
Indigenous identity employees      68927.0   348.116162   1.0    99.0
Male employees                    609748.0  2903.561905   3.0   547.0
Non-immigrant employees          1315368.0  6643.272727  10.0  1452.5
Non-indigenous identity employees 1719177.0  8682.712121   2.0  1680.5
Not a visible minority           1326971.0  6318.909524   2.0   911.0
Trade certificate                  93374.0   471.585859   1.0    85.5
University degree and higher     1052815.0  5317.247475   4.0   931.5
Visible minority                  462442.0  2202.104762   1.0   227.0

                                    amax  size
Characteristics
15 to 24 years                     4797.0   192
25 to 34 years                    25966.0   210
35 to 44 years                    34223.0   210
45 to 54 years                    36376.0   210
55 to 64 years                    29595.0   210
65 years old and over              6876.0   192
College diploma                   32249.0   210
Female employees                  92154.0   210
High school diploma and less      16432.0   210
Immigrant employees               36981.0   198
Indigenous identity employees      5232.0   198
Male employees                    45679.0   210
Non-immigrant employees          100851.0   198
Non-indigenous identity employees 132601.0   198
Not a visible minority           101650.0   210
Trade certificate                  7213.0   198
University degree and higher      81938.0   198
Visible minority                  36182.0   210
```

```
Overall,
Sum :  10734620.0
Mean :  2923.371459694989
Min/median/max : 1.0 / 261.0 / 132601.0
Standard Deviation :  9261.080456073634
Skewnewss :  7.1322192127358255
Total size :  3672
```

Histogram for testing dataset



Output #11

```
# Used by Chi-Square methods
REF_DATE                          2019  2020  2021
Characteristics
15 to 24 years                      64    64    64
25 to 34 years                      70    70    70
35 to 44 years                      70    70    70
45 to 54 years                      70    70    70
55 to 64 years                      70    70    70
65 years old and over               64    64    64
College diploma                     70    70    70
Female employees                    70    70    70
High school diploma and less        70    70    70
Immigrant employees                 66    66    66
Indigenous identity employees       66    66    66
Male employees                      70    70    70
Non-immigrant employees             66    66    66
Non-indigenous identity employees   66    66    66
Not a visible minority              70    70    70
Trade certificate                   66    66    66
University degree and higher        66    66    66
Visible minority                    70    70    70
p value is 1.0
Independent (H0 holds true)
REF_DATE                  2019  2020  2021
GEO
Alberta                     90    90    90
British Columbia            90    90    90
Canada                      90    90    90
Manitoba                    90    90    90
New Brunswick               90    90    90
Newfoundland and Labrador   86    86    86
Northwest Territories       86    86    86
Nova Scotia                 90    90    90
Nunavut                     78    78    78
Ontario                     90    90    90
Prince Edward Island        86    86    86
Quebec                      90    90    90
Saskatchewan                90    90    90
Yukon                       78    78    78
p value is 1.0
Independent (H0 holds true)
```

REF_DATE	2019	2020	2021
Characteristics			
15 to 24 years	64	64	64
25 to 34 years	70	70	70
35 to 44 years	70	70	70
45 to 54 years	70	70	70
55 to 64 years	70	70	70
65 years old and over	64	64	64
College diploma	70	70	70
Female employees	70	70	70
High school diploma and less	70	70	70
Immigrant employees	66	66	66
Indigenous identity employees	66	66	66
Male employees	70	70	70
Non-immigrant employees	66	66	66
Non-indigenous identity employees	66	66	66
Not a visible minority	70	70	70
Trade certificate	66	66	66
University degree and higher	66	66	66
Visible minority	70	70	70

p value is 1.0

Independent (H0 holds true)

REF_DATE	2019	2020	2021
GEO			
Alberta	90	90	90
British Columbia	90	90	90
Canada	90	90	90
Manitoba	90	90	90
New Brunswick	90	90	90
Newfoundland and Labrador	86	86	86
Northwest Territories	86	86	86
Nova Scotia	90	90	90
Nunavut	78	78	78
Ontario	90	90	90
Prince Edward Island	86	86	86
Quebec	90	90	90
Saskatchewan	90	90	90
Yukon	78	78	78

p value is 1.0

Independent (H0 holds true)

REF_DATE	2019	2020	2021
Characteristics			
15 to 24 years	64	64	64
25 to 34 years	70	70	70
35 to 44 years	70	70	70
45 to 54 years	70	70	70
55 to 64 years	70	70	70
65 years old and over	64	64	64
College diploma	70	70	70
Female employees	70	70	70
High school diploma and less	70	70	70
Immigrant employees	66	66	66
Indigenous identity employees	66	66	66
Male employees	70	70	70
Non-immigrant employees	66	66	66
Non-indigenous identity employees	66	66	66
Not a visible minority	70	70	70
Trade certificate	66	66	66
University degree and higher	66	66	66
Visible minority	70	70	70

p value is 1.0
Independent (H0 holds true)

REF_DATE	2019	2020	2021
GEO			
Alberta	90	90	90
British Columbia	90	90	90
Canada	90	90	90
Manitoba	90	90	90
New Brunswick	90	90	90
Newfoundland and Labrador	86	86	86
Northwest Territories	86	86	86
Nova Scotia	90	90	90
Nunavut	78	78	78
Ontario	90	90	90
Prince Edward Island	86	86	86
Quebec	90	90	90
Saskatchewan	90	90	90
Yukon	78	78	78

p value is 1.0
Independent (H0 holds true)

REF_DATE	2019	2020	2021
Characteristics			
15 to 24 years	64	64	64
25 to 34 years	70	70	70
35 to 44 years	70	70	70
45 to 54 years	70	70	70
55 to 64 years	70	70	70
65 years old and over	64	64	64
College diploma	70	70	70
Female employees	70	70	70
High school diploma and less	70	70	70
Immigrant employees	66	66	66
Indigenous identity employees	66	66	66
Male employees	70	70	70
Non-immigrant employees	66	66	66
Non-indigenous identity employees	66	66	66
Not a visible minority	70	70	70
Trade certificate	66	66	66
University degree and higher	66	66	66
Visible minority	70	70	70

p value is 1.0

Independent (H0 holds true)

REF_DATE	2019	2020	2021
GEO			
Alberta	90	90	90
British Columbia	90	90	90
Canada	90	90	90
Manitoba	90	90	90
New Brunswick	90	90	90
Newfoundland and Labrador	86	86	86
Northwest Territories	86	86	86
Nova Scotia	90	90	90
Nunavut	78	78	78
Ontario	90	90	90
Prince Edward Island	86	86	86
Quebec	90	90	90
Saskatchewan	90	90	90
Yukon	78	78	78

p value is 1.0

Independent (H0 holds true)

REF_DATE	2019	2020	2021
Characteristics			
15 to 24 years	64	64	64
25 to 34 years	70	70	70
35 to 44 years	70	70	70
45 to 54 years	70	70	70
55 to 64 years	70	70	70
65 years old and over	64	64	64
College diploma	70	70	70
Female employees	70	70	70
High school diploma and less	70	70	70
Immigrant employees	66	66	66
Indigenous identity employees	66	66	66
Male employees	70	70	70
Non-immigrant employees	66	66	66
Non-indigenous identity employees	66	66	66
Not a visible minority	70	70	70
Trade certificate	66	66	66
University degree and higher	66	66	66
Visible minority	70	70	70

p value is 1.0

Independent (H0 holds true)

REF_DATE	2019	2020	2021
GEO			
Alberta	90	90	90
British Columbia	90	90	90
Canada	90	90	90
Manitoba	90	90	90
New Brunswick	90	90	90
Newfoundland and Labrador	86	86	86
Northwest Territories	86	86	86
Nova Scotia	90	90	90
Nunavut	78	78	78
Ontario	90	90	90
Prince Edward Island	86	86	86
Quebec	90	90	90
Saskatchewan	90	90	90
Yukon	78	78	78

p value is 1.0

Independent (H0 holds true)

REF_DATE	2019	2020	2021
Characteristics			
15 to 24 years	64	64	64
25 to 34 years	70	70	70
35 to 44 years	70	70	70
45 to 54 years	70	70	70
55 to 64 years	70	70	70
65 years old and over	64	64	64
College diploma	70	70	70
Female employees	70	70	70
High school diploma and less	70	70	70
Immigrant employees	66	66	66
Indigenous identity employees	66	66	66
Male employees	70	70	70
Non-immigrant employees	66	66	66
Non-indigenous identity employees	66	66	66
Not a visible minority	70	70	70
Trade certificate	66	66	66
University degree and higher	66	66	66
Visible minority	70	70	70

p value is 1.0

Independent (H0 holds true)

REF_DATE	2019	2020	2021
GEO			
Alberta	90	90	90
British Columbia	90	90	90
Canada	90	90	90
Manitoba	90	90	90
New Brunswick	90	90	90
Newfoundland and Labrador	86	86	86
Northwest Territories	86	86	86
Nova Scotia	90	90	90
Nunavut	78	78	78
Ontario	90	90	90
Prince Edward Island	86	86	86
Quebec	90	90	90
Saskatchewan	90	90	90
Yukon	78	78	78

p value is 1.0

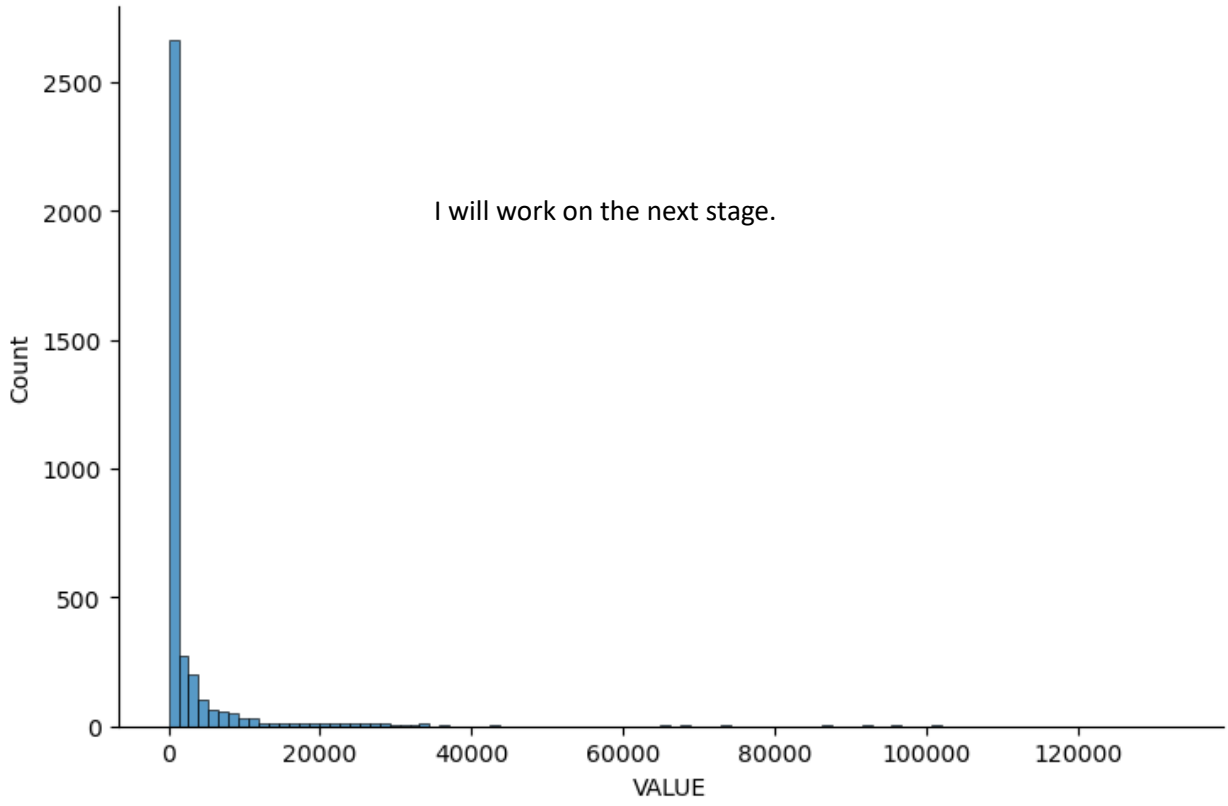
Independent (H0 holds true)

Output #12a

```
testing set By Age
                             sum         mean    amin  median    amax  size
Characteristics
15 to 24 years         179872.0   936.833333   713.0   927.5  1281.0   192
25 to 34 years         333662.0  1588.866667  1292.0  1576.0  1870.0   210
35 to 44 years         371386.0  1768.504762  1424.0  1757.0  2092.0   210
45 to 54 years         384987.0  1833.271429  1541.0  1826.5  2191.0   210
55 to 64 years         351843.0  1675.442857  1377.0  1675.0  2071.0   210
65 years old and over  206760.0  1076.875000   565.0  1076.5  1415.0   192
Overall,
Sum :  1828510.0
Mean :  1493.8807189542483
Min/median/max : 565.0 / 1633.0 / 2191.0
Standard Deviation :  357.1076307712267
Skewnewss :  -0.5829393561867995
Total size :  1224

testing set By Gender
                        sum         mean    amin  median    amax  size
Characteristics
Female employees  323755.0  1541.690476  1302.0  1540.0  1773.0   210
Male employees    343101.0  1633.814286  1373.0  1644.0  1821.0   210
Overall,
Sum :  666856.0
Mean :  1587.752380952381
Min/median/max : 1302.0 / 1596.0 / 1821.0
Standard Deviation :  93.82948373587371
Skewnewss :  -0.3050908032633471
Total size :  420

testing set By Education
                                    sum         mean    amin  median    amax  \
Characteristics
High school diploma and less  275636.0  1312.552381  1054.0  1307.5  1667.0
Trade certificate             306174.0  1546.333333   789.0  1547.5  1808.0
University degree and higher  341795.0  1726.237374  1536.0  1706.5  2043.0

                              size
Characteristics
High school diploma and less   210
Trade certificate              198
University degree and higher   198
Overall,
Sum :  923605.0
Mean :  1524.1006600660066
Min/median/max : 789.0 / 1545.0 / 2043.0
Standard Deviation :  204.02000670247472
Skewnewss :  -0.1568377438071896
Total size :  606
```

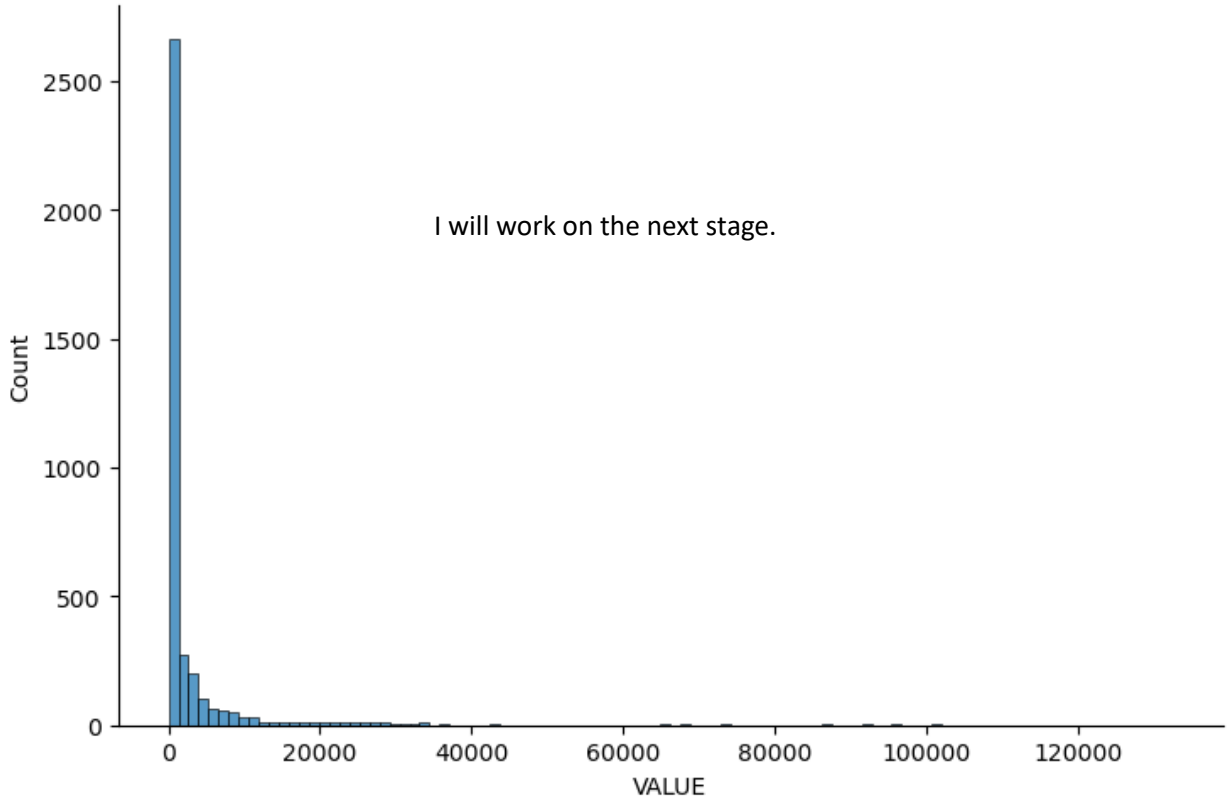
```
testing set By Immigrant
                                sum         mean    amin  median    amax  size
Characteristics
Immigrant employees      318818.0  1610.191919  1336.0  1589.0  2250.0   198
Non-immigrant employees  310245.0  1566.893939  1315.0  1570.0  1767.0   198
Overall,
Sum :  629063.0
Mean :  1588.5429292929293
Min/median/max : 1315.0 / 1580.0 / 2250.0
Standard Deviation :  112.27714481578558
Skewnewss :  1.6694520665533967
Total size :  396
```

Histogram for testing dataset by age



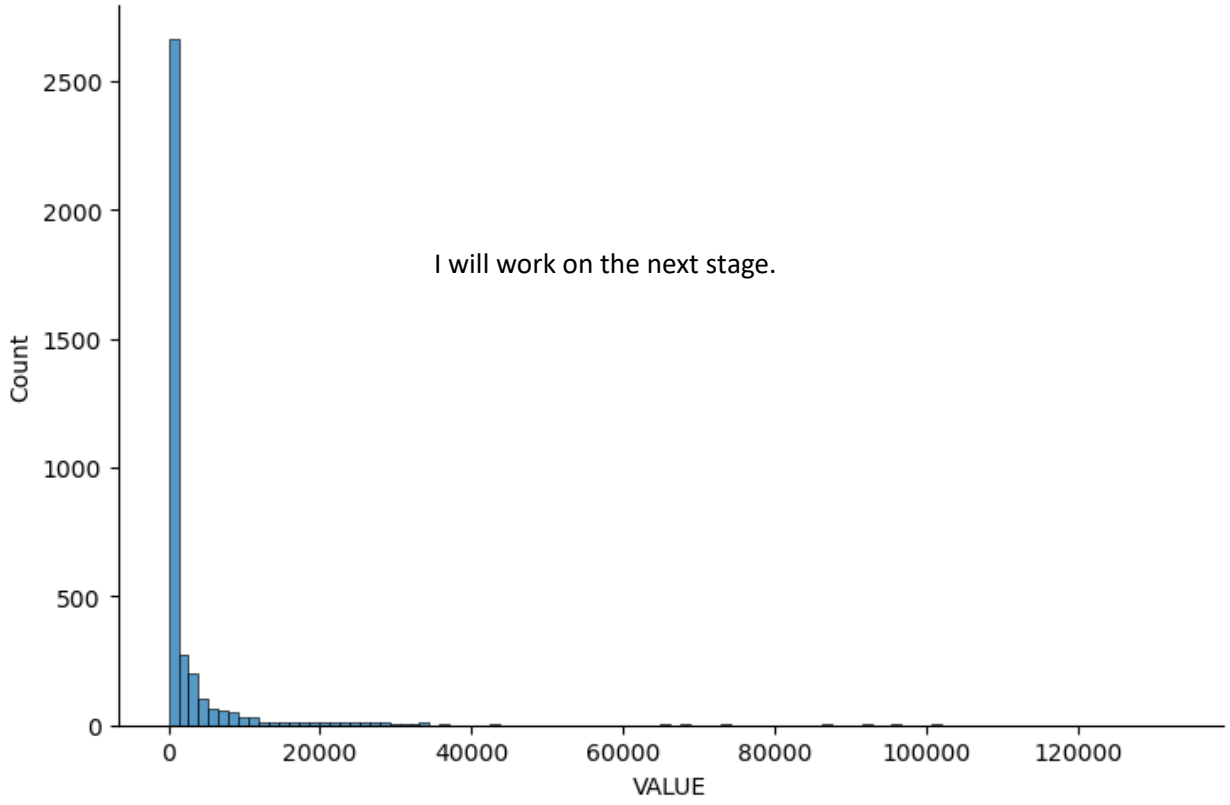
I will work on the next stage.

Histgram for testing dataset by gender



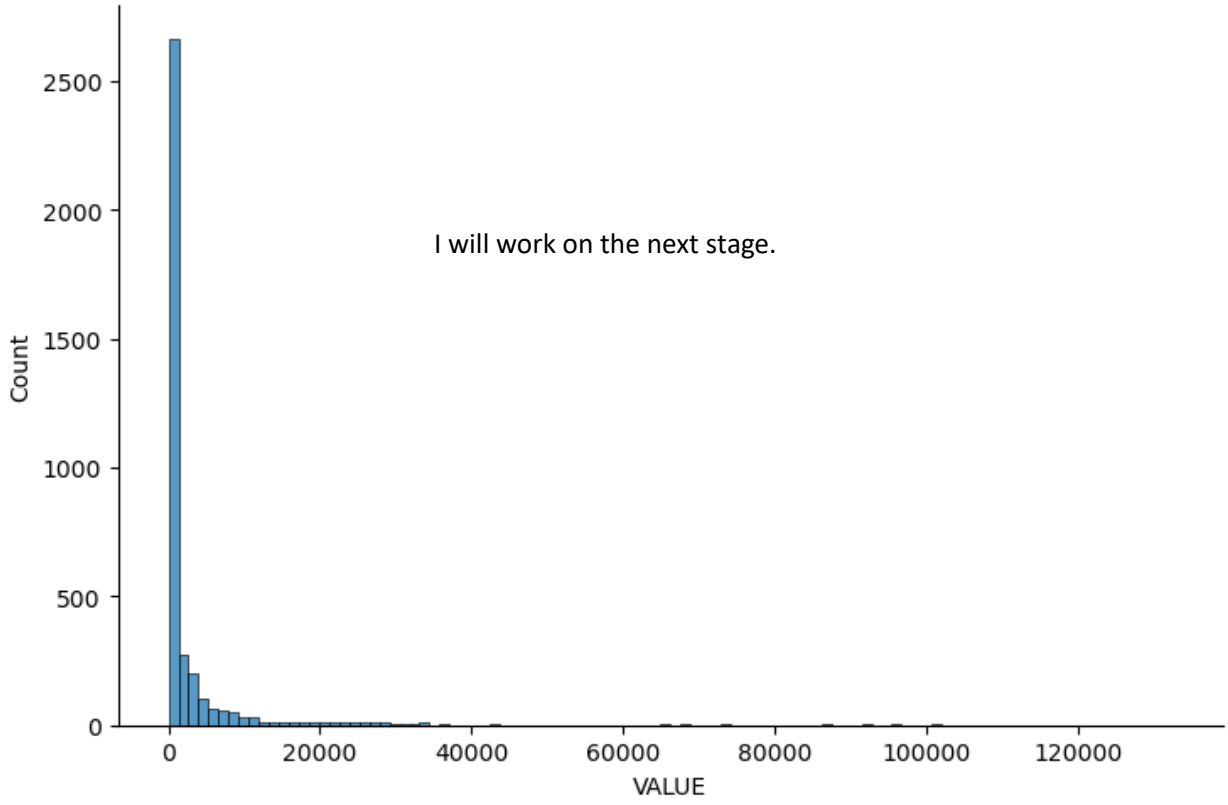
I will work on the next stage.

Histgram for testing dataset by education



I will work on the next stage.

Histgram for testing dataset by immigrant



I will work on the next stage.

Output #12b

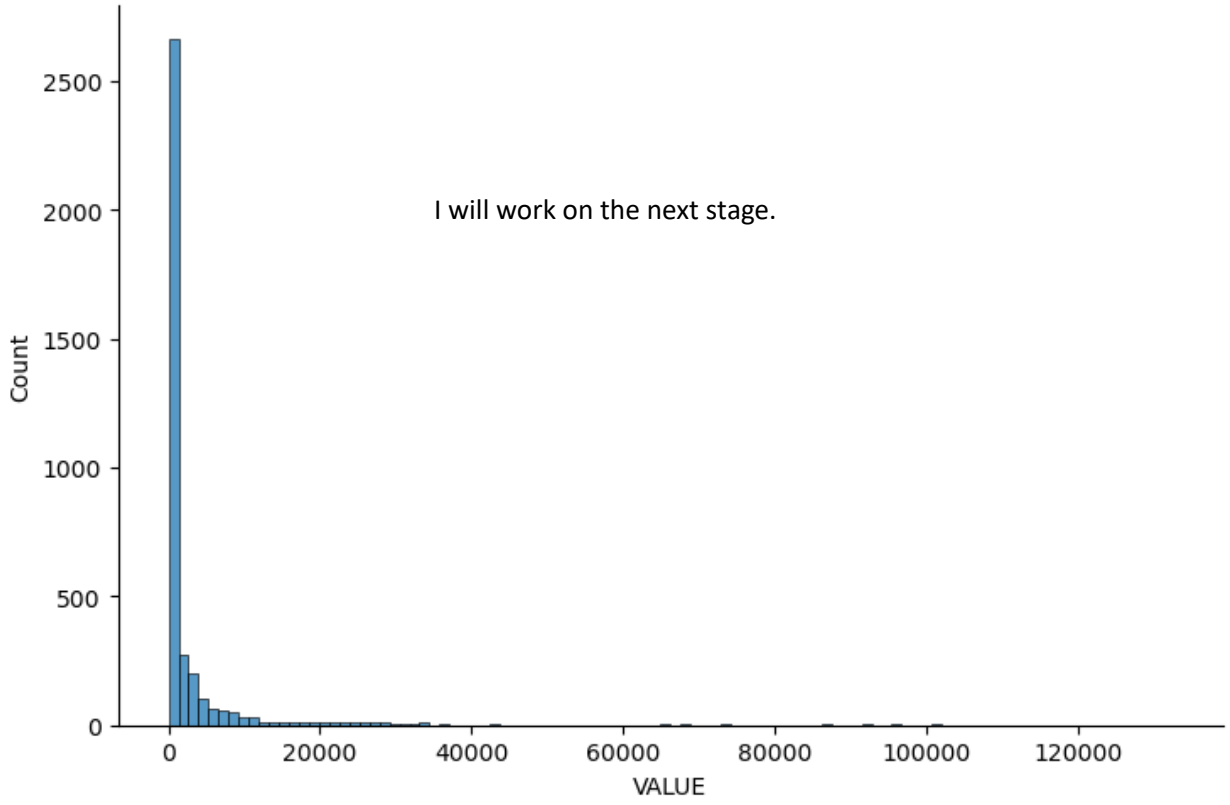
Final Output for "Average annual wages and salaries"

```
testing set By Age
                               sum          mean     amin   median      amax  \
Characteristics
15 to 24 years          3482308.0  18137.020833  11093.0  16653.5   45844.0
25 to 34 years          9108041.0  43371.623810  26534.0  42901.0   75429.0
35 to 44 years         12296655.0  58555.500000  37336.0  56960.5   95714.0
45 to 54 years         13595377.0  64739.890476  39455.0  64274.5  103580.0
55 to 64 years         12592232.0  59963.009524  33296.0  56266.0  133071.0
65 years old and over   6840988.0  35630.145833  18187.0  34170.0   76577.0

                       size
Characteristics
15 to 24 years          192
25 to 34 years          210
35 to 44 years          210
45 to 54 years          210
55 to 64 years          210
65 years old and over   192
Overall,
Sum :  57915601.0
Mean :  47316.66748366013
Min/median/max : 11093.0 / 47033.5 / 133071.0
Standard Deviation :  20534.474851841318
Skewnewss :  0.4846905588925664
Total size :  1224

testing set By Gender
                          sum          mean     amin   median     amax  size
Characteristics
Female employees   9929829.0  47284.900000  28080.0  45973.5  86432.0   210
Male employees    11755264.0  55977.447619  32342.0  55938.0  95273.0   210
Overall,
Sum :  21685093.0
Mean :  51631.17380952381
Min/median/max : 28080.0 / 49809.5 / 95273.0
Standard Deviation :  13041.836974604885
Skewnewss :  0.7219188061249169
Total size :  420
```

```
testing set By Education
                                        sum          mean     amin   median  \
Characteristics
High school diploma and less   6915927.0  32932.985714  17828.0  31071.0
Trade certificate              8609637.0  43483.015152  25221.0  42431.0
University degree and higher  12581443.0  63542.641414  38860.0  62574.0

                                   amax  size
Characteristics
High school diploma and less   73020.0   210
Trade certificate              72854.0   198
University degree and higher  121264.0   198
Overall,
Sum :  28107007.0
Mean :  46381.199669967
Min/median/max : 17828.0 / 43075.0 / 121264.0
Standard Deviation :  17647.036331311094
Skewnewss :  1.273924834730388
Total size :  606

testing set By Immigrant
                                 sum          mean     amin   median     amax  \
Characteristics
Immigrant employees      9838690.0  49690.353535  23144.0  48933.5  98074.0
Non-immigrant employees  9799014.0  49489.969697  29956.0  49313.5  85472.0

                         size
Characteristics
Immigrant employees       198
Non-immigrant employees   198
Overall,
Sum :  19637704.0
Mean :  49590.16161616162
Min/median/max : 23144.0 / 49016.0 / 98074.0
Standard Deviation :  13097.423221750063
Skewnewss :  0.9303937237175556
Total size :  396
```

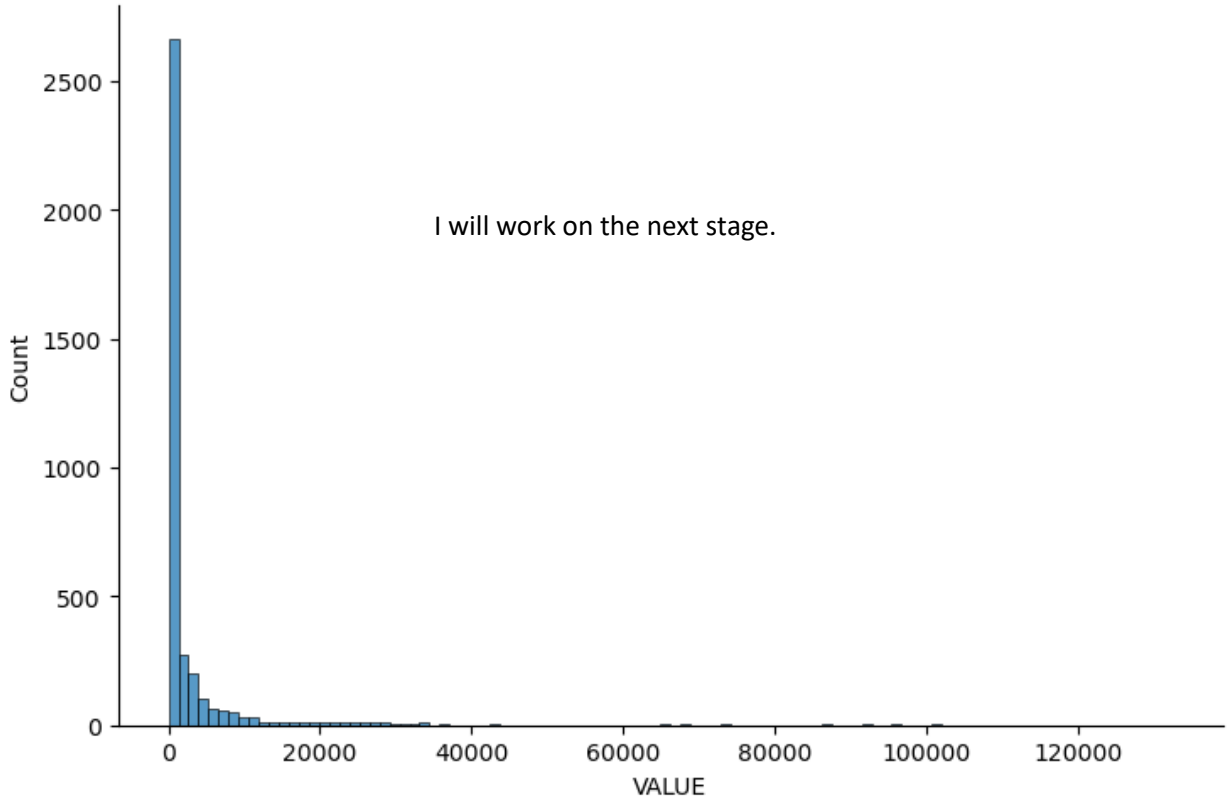
In [ ]:

Histogram for testing dataset by age



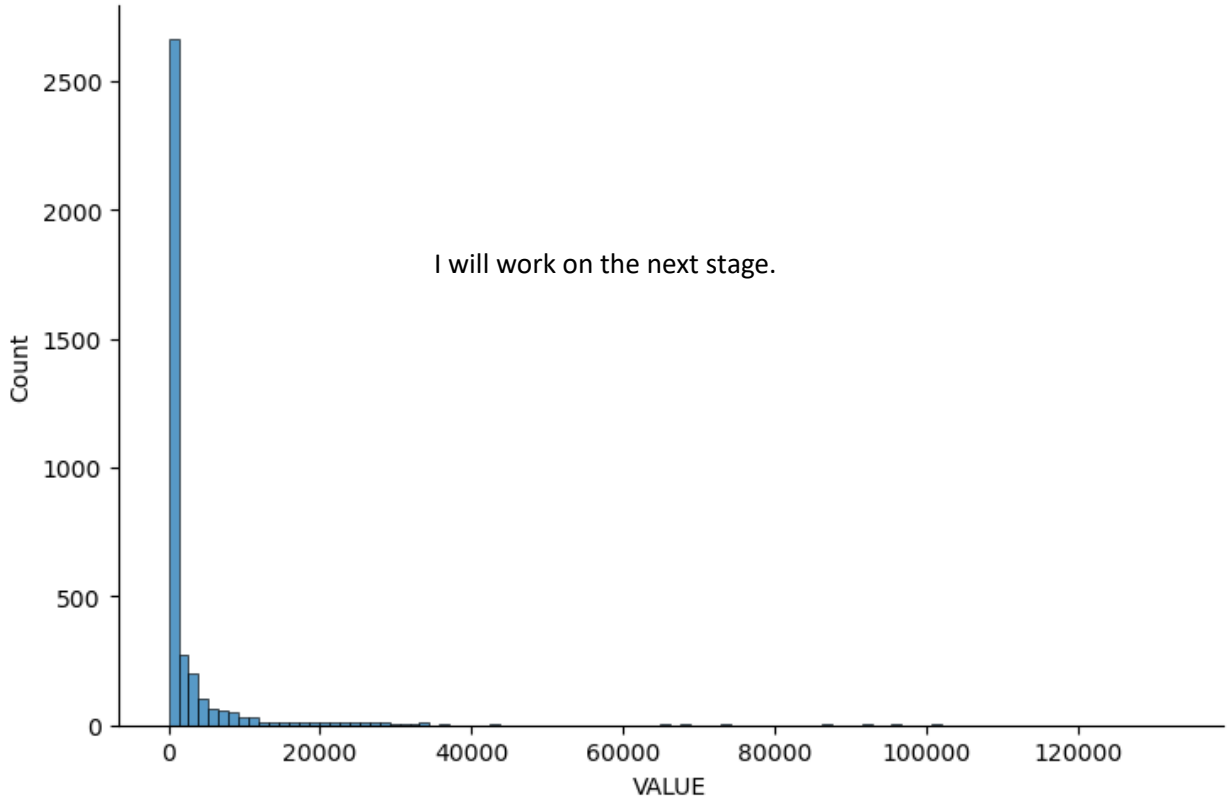
I will work on the next stage.

Histgram for testing dataset by gender



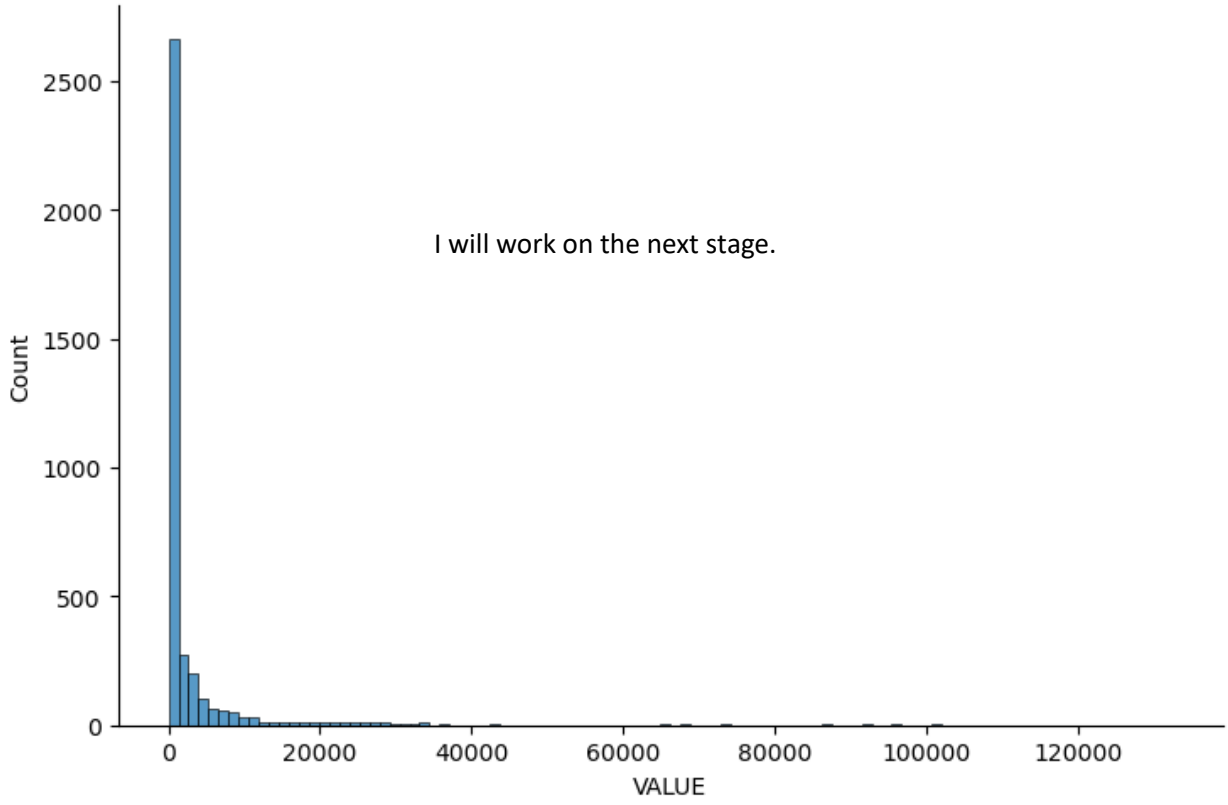
I will work on the next stage.

Histgram for testing dataset by education



I will work on the next stage.

Histgram for testing dataset by immigrant



I will work on the next stage.

Output #12c

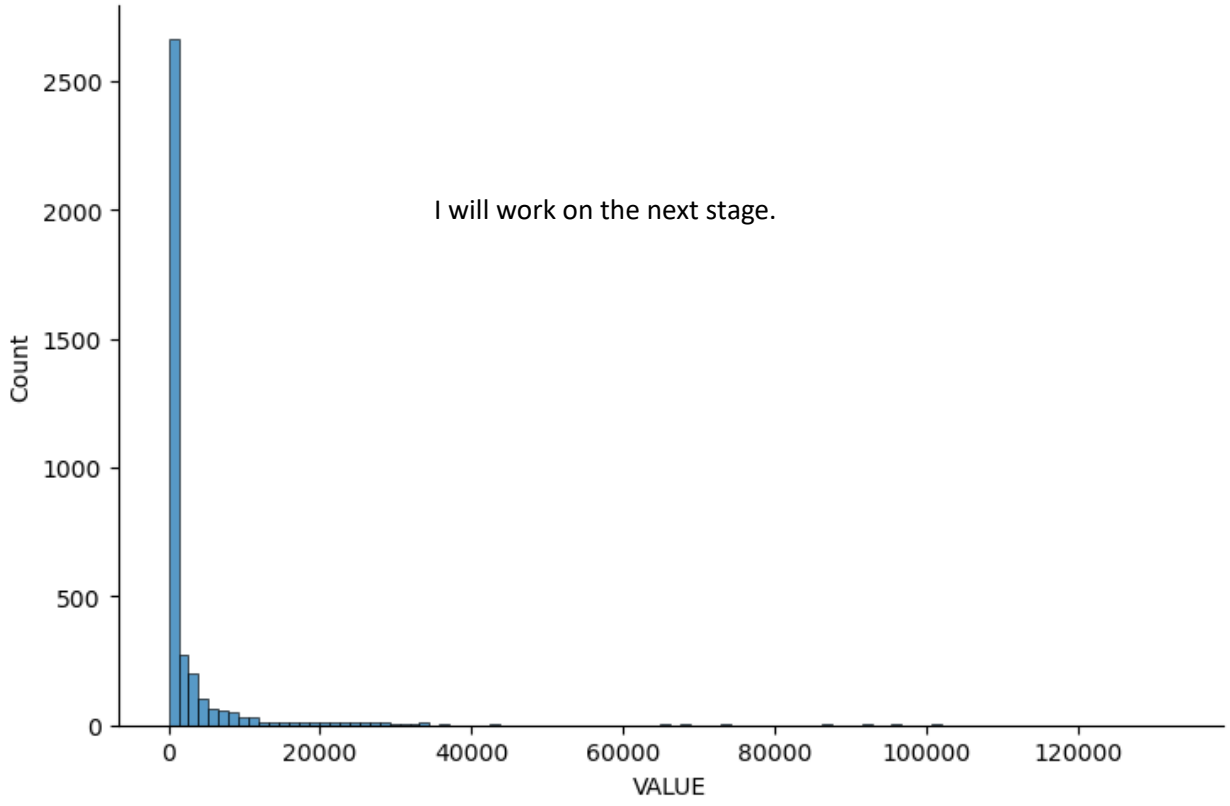
Final Output for "Average hourly wage"

```
testing set By Age
                            sum       mean   amin  median   amax  size
Characteristics
15 to 24 years         3689.12  19.214167  12.95  18.775  36.37   192
25 to 34 years         5759.41  27.425762  17.01  27.010  45.91   210
35 to 44 years         6991.24  33.291619  19.42  32.680  58.12   210
45 to 54 years         7451.95  35.485476  21.54  34.965  64.54   210
55 to 64 years         7475.92  35.599619  19.42  33.800  64.92   210
65 years old and over  6329.58  32.966563  18.32  32.140  66.63   192
Overall,
Sum :  37697.22
Mean :  30.79838235294118
Min/median/max : 12.95 / 30.425 / 66.63
Standard Deviation :  9.683878549205739
Skewnewss :  0.8175005207452904
Total size :  1224

testing set By Gender
                     sum       mean   amin  median   amax  size
Characteristics
Female employees  6439.32  30.663429  18.34  30.200  55.06   210
Male employees    7203.15  34.300714  19.31  33.585  56.59   210
Overall,
Sum :  13642.470000000001
Mean :  32.48207142857143
Min/median/max : 18.34 / 31.68 / 56.59
Standard Deviation :  7.83023476170858
Skewnewss :  0.7493470452539902
Total size :  420

testing set By Education
                                  sum       mean   amin  median   amax  size
Characteristics
High school diploma and less  5242.44  24.964000  15.69  23.600  44.52   210
Trade certificate             5569.30  28.127778  17.63  27.370  44.58   198
University degree and higher  7296.08  36.848889  21.80  36.385  69.52   198
Overall,
Sum :  18107.82
Mean :  29.88089108910891
Min/median/max : 15.69 / 27.8 / 69.52
Standard Deviation :  8.718969410933747
Skewnewss :  1.32266748675931
Total size :  606
```

```
testing set By Immigrant
                              sum       mean   amin  median   amax  size
Characteristics
Immigrant employees      6065.65  30.634596  16.93  29.925  60.18   198
Non-immigrant employees  6267.15  31.652273  19.24  31.000  55.32   198
Overall,
Sum :  12332.8
Mean :  31.143434343434343
Min/median/max : 16.93 / 30.795 / 60.18
Standard Deviation :  7.5893914744701325
Skewnewss :  0.8277016858447296
Total size :  396
```

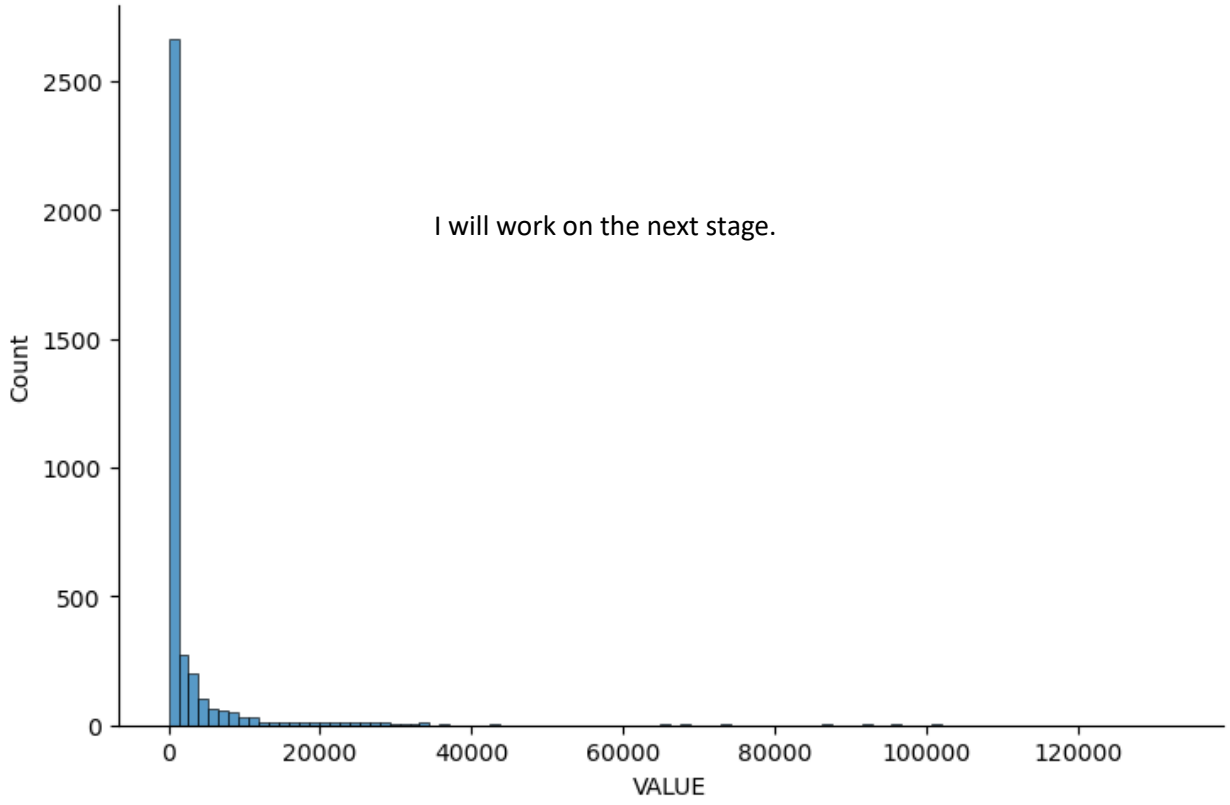
ln [ ]:

Histogram for testing dataset by age



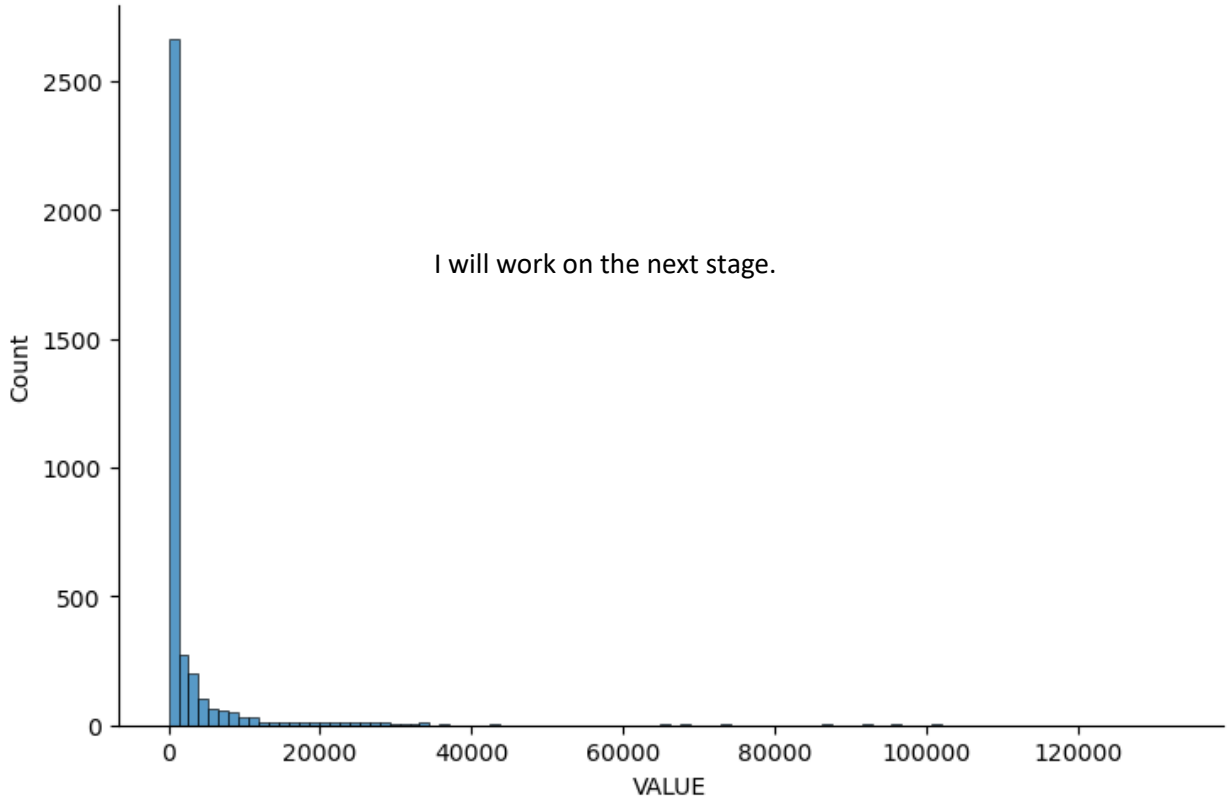
I will work on the next stage.

Histgram for testing dataset by gender



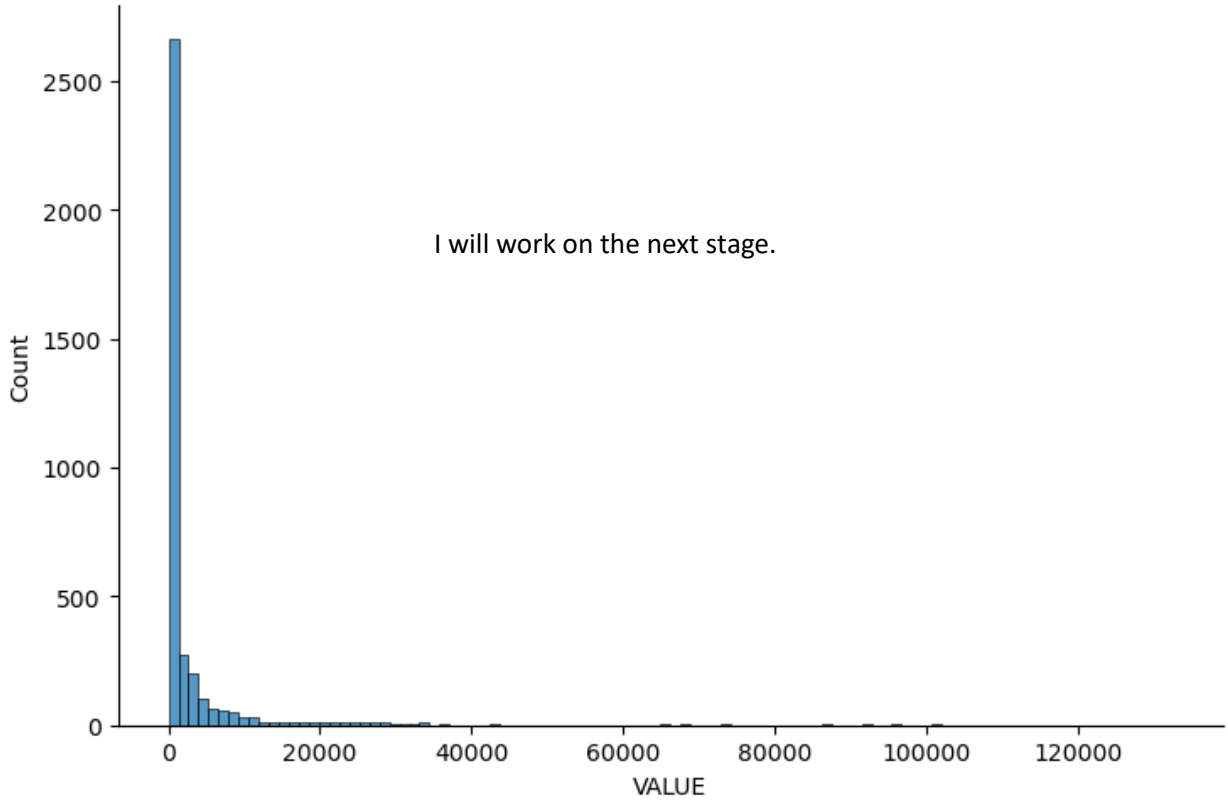
I will work on the next stage.

Histgram for testing dataset by education



I will work on the next stage.

Histgram for testing dataset by immigrant



I will work on the next stage.

Output #12d

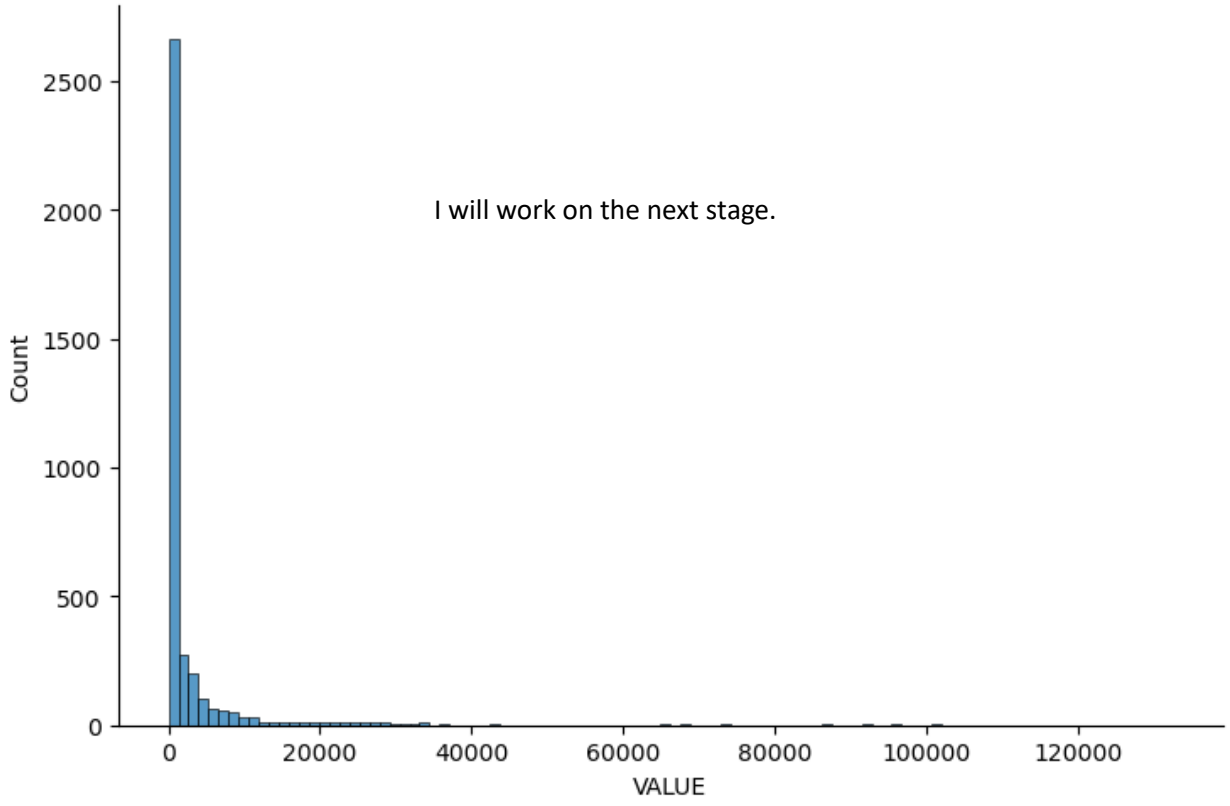
Final Output for "Average weekly hours worked"

```
testing set By Age
                           sum       mean  amin  median  amax  size
Characteristics
15 to 24 years         3465.0  18.046875  14.0    18.0  25.0   192
25 to 34 years         6413.0  30.538095  25.0    30.0  36.0   210
35 to 44 years         7139.0  33.995238  27.0    34.0  40.0   210
45 to 54 years         7406.0  35.266667  30.0    35.0  42.0   210
55 to 64 years         6765.0  32.214286  26.0    32.0  40.0   210
65 years old and over  3969.0  20.671875  11.0    21.0  27.0   192
Overall,
Sum :  35157.0
Mean :  28.723039215686274
Min/median/max : 11.0 / 31.0 / 42.0
Standard Deviation :  6.87095753369461
Skewnewss :  -0.578814861266804
Total size :  1224

testing set By Gender
                     sum       mean  amin  median  amax  size
Characteristics
Female employees  6219.0  29.614286  25.0    30.0  34.0   210
Male employees    6590.0  31.380952  26.0    32.0  35.0   210
Overall,
Sum :  12809.0
Mean :  30.497619047619047
Min/median/max : 25.0 / 31.0 / 35.0
Standard Deviation :  1.826392240670126
Skewnewss :  -0.3326933662492969
Total size :  420

testing set By Education
                                 sum       mean  amin  median  amax  size
Characteristics
High school diploma and less  5303.0  25.252381  20.0    25.0  32.0   210
Trade certificate             5889.0  29.742424  15.0    30.0  35.0   198
University degree and higher  6568.0  33.171717  30.0    33.0  39.0   198
Overall,
Sum :  17760.0
Mean :  29.306930693069308
Min/median/max : 15.0 / 30.0 / 39.0
Standard Deviation :  3.910490015779788
Skewnewss :  -0.18790375522922004
Total size :  606
```

```
testing set By Immigrant
                              sum       mean  amin  median  amax  size
Characteristics
Immigrant employees      6133.0  30.974747  26.0    31.0  43.0   198
Non-immigrant employees  5961.0  30.106061  25.0    30.0  34.0   198
Overall,
Sum :  12094.0
Mean :  30.54040404040404
Min/median/max : 25.0 / 30.0 / 43.0
Standard Deviation :  2.1837054594704566
Skewnewss :  1.6446273199481087
Total size :  396
```

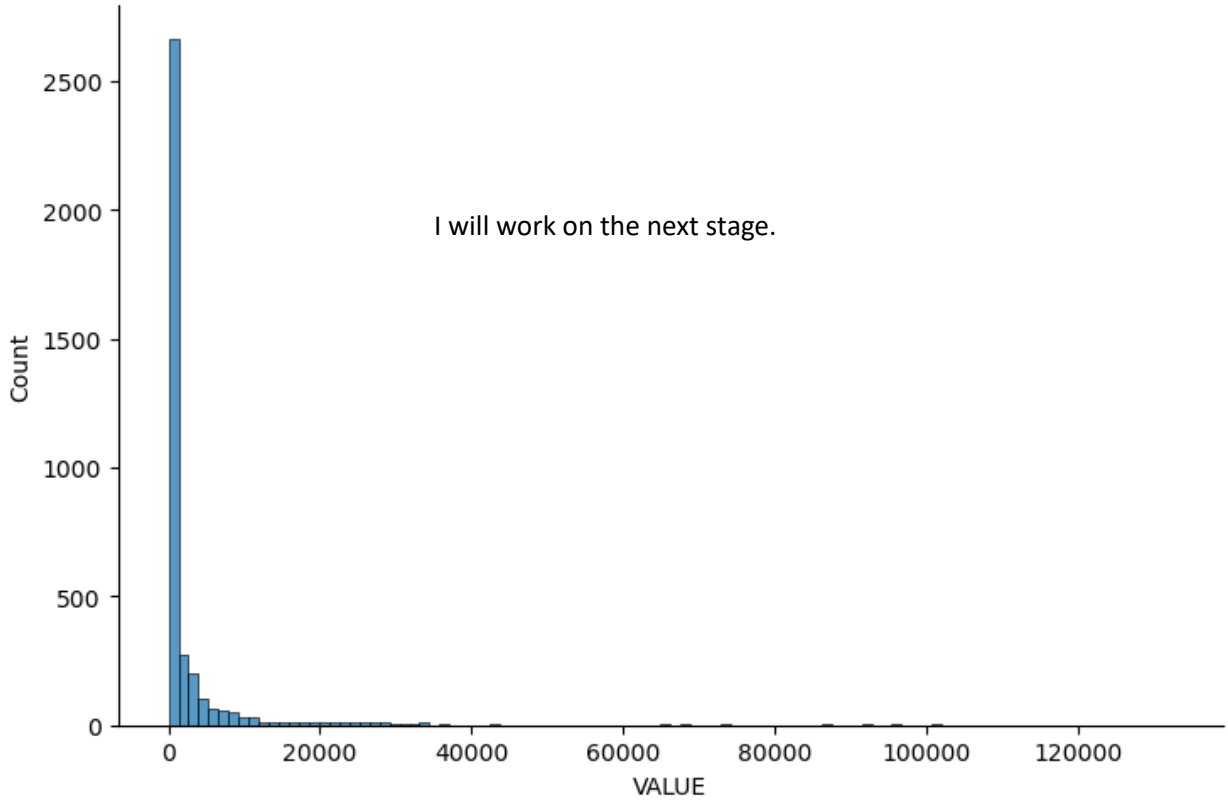
In [ ]:

```
Histogram for testing dataset by age
```



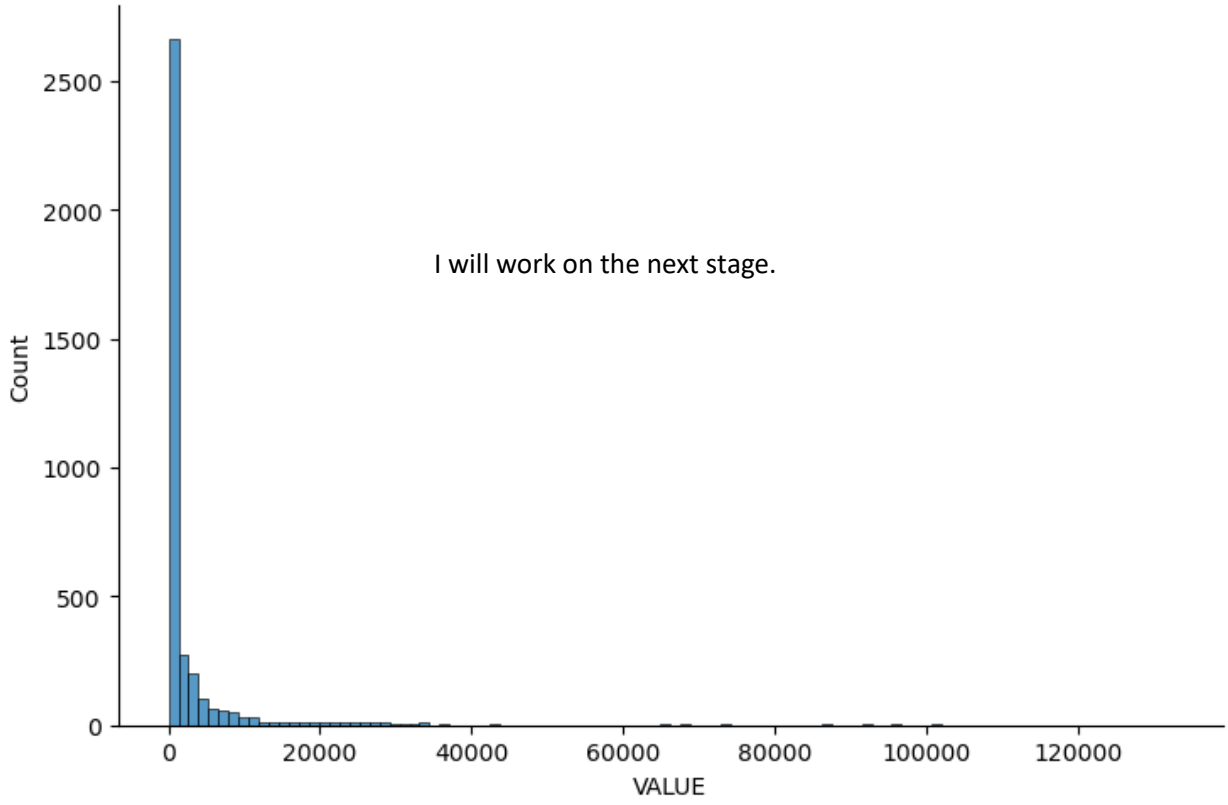
I will work on the next stage.

Histgram for testing dataset by gender



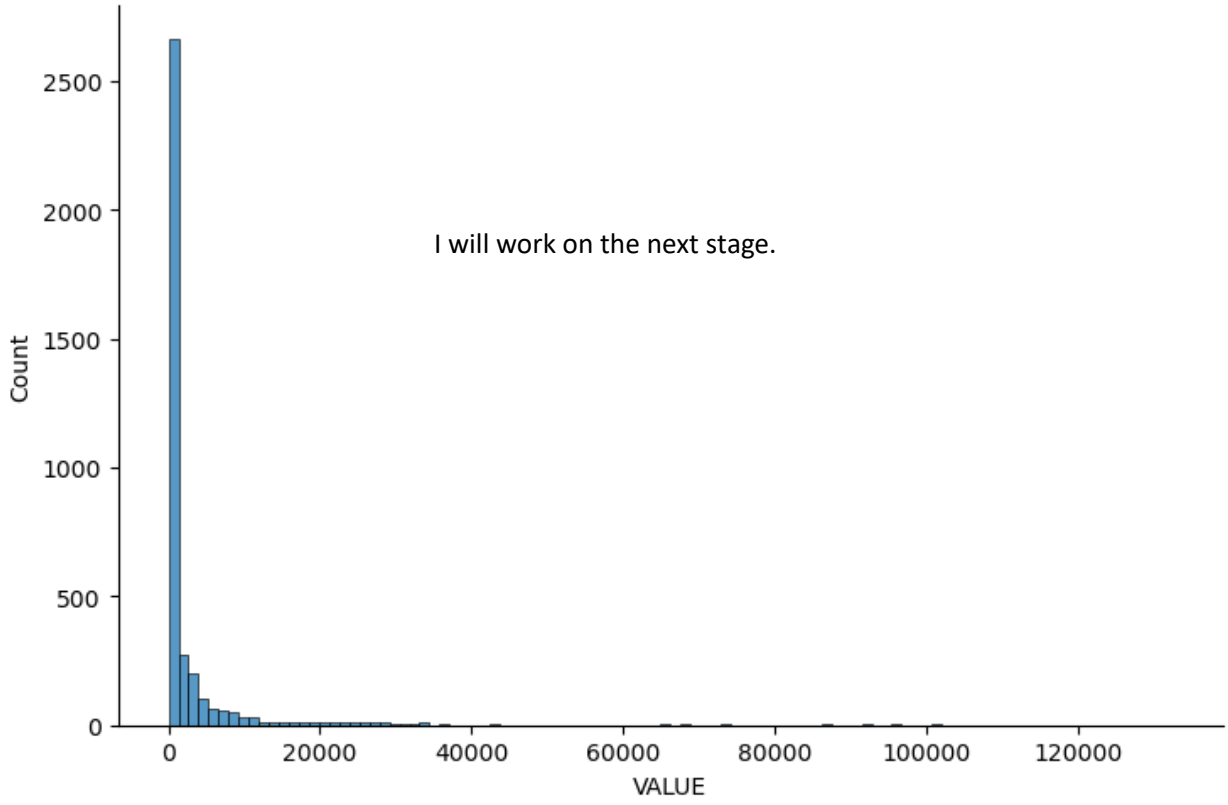
I will work on the next stage.

Histgram for testing dataset by education



I will work on the next stage.

Histgram for testing dataset by immigrant



I will work on the next stage.

Output #12e

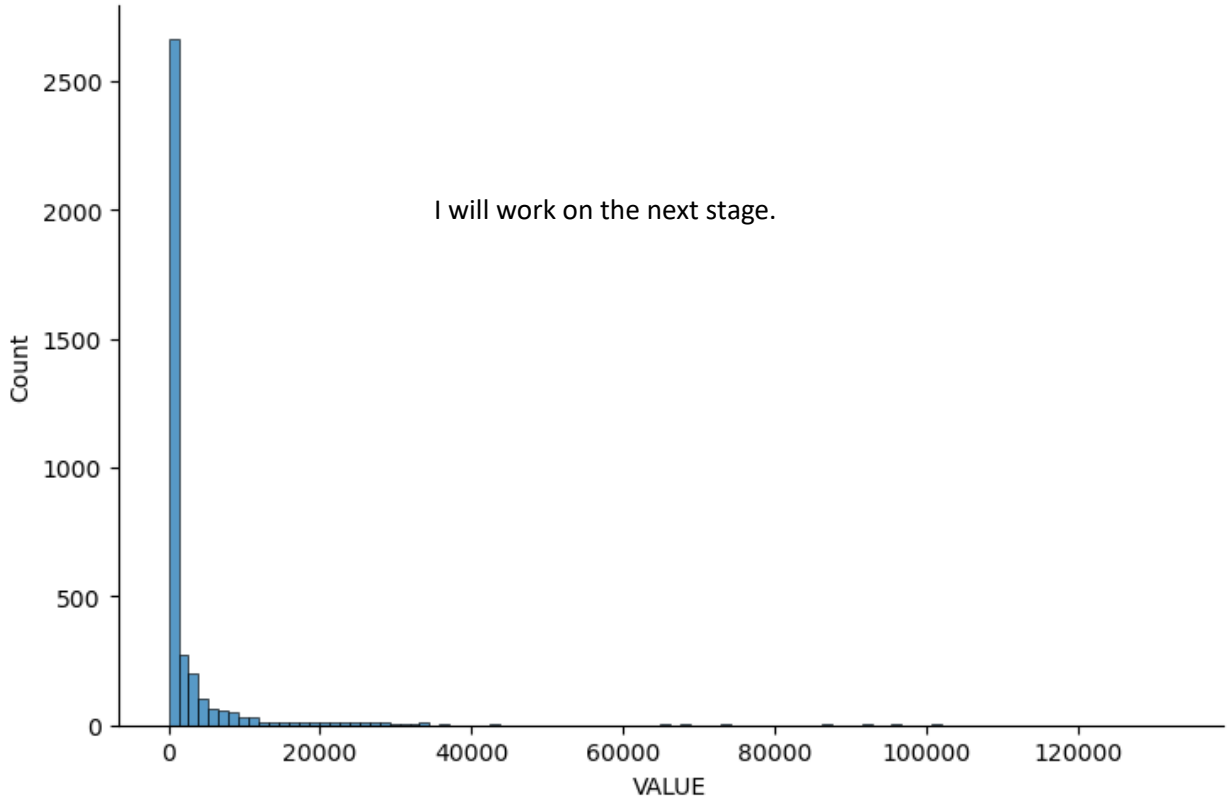
Final Output for "Hours Worked"

```
testing set By Age
                              sum          mean  amin   median      amax  size
Characteristics
15 to 24 years          3132171.0  16313.390625  29.0   3859.5  228566.0   192
25 to 34 years         11857576.0  56464.647619  32.0  10166.5  888764.0   210
35 to 44 years         12919606.0  61521.933333  33.0  10825.0  971688.0   210
45 to 54 years         12974580.0  61783.714286  31.0  10600.5  966480.0   210
55 to 64 years         10690620.0  50907.714286  25.0   8934.0  794712.0   210
65 years old and over   2528517.0  13169.359375  30.0   2913.5  183461.0   192
Overall,
Sum :  54103070.0
Mean :  44201.8545751634
Min/median/max : 25.0 / 6420.5 / 971688.0
Standard Deviation :  114827.01044742283
Skewnewss :  5.111311073923015
Total size :  1224

testing set By Gender
                          sum           mean  amin   median       amax  size
Characteristics
Female employees  37193214.0  177110.542857  64.0  28332.5  2809860.0   210
Male employees    16912308.0   80534.800000  74.0  15656.5  1223817.0   210
Overall,
Sum :  54105522.0
Mean :  128822.67142857143
Min/median/max : 64.0 / 20350.5 / 2809860.0
Standard Deviation :  325132.54715451447
Skewnewss :  5.1811837084995025
Total size :  420
```

```
testing set By Education
                                       sum           mean  amin   median  \
Characteristics
High school diploma and less   9285900.0   44218.571429  61.0  11984.5
Trade certificate              3449203.0   17420.217172  20.0   3080.0
University degree and higher  27393105.0  138349.015152  96.0  23963.5

                                    amax  size
Characteristics
High school diploma and less   658441.0   210
Trade certificate              260168.0   198
University degree and higher  2065714.0   198
Overall,
Sum :  40128208.0
Mean :  66218.16501650165
Min/median/max : 20.0 / 7837.0 / 2065714.0
Standard Deviation :  197113.4888126052
Skewnewss :  6.531158889782265
Total size :  606

testing set By Immigrant
                                  sum           mean   amin   median       amax  \
Characteristics
Immigrant employees      14355944.0   72504.767677   44.0   7122.0  1084033.0
Non-immigrant employees  39740524.0  200709.717172  264.0  48444.5  2949643.0

                         size
Characteristics
Immigrant employees       198
Non-immigrant employees   198
Overall,
Sum :  54096468.0
Mean :  136607.24242424243
Min/median/max : 44.0 / 22542.5 / 2949643.0
Standard Deviation :  342690.90436216333
Skewnewss :  5.210924241138901
Total size :  396
```

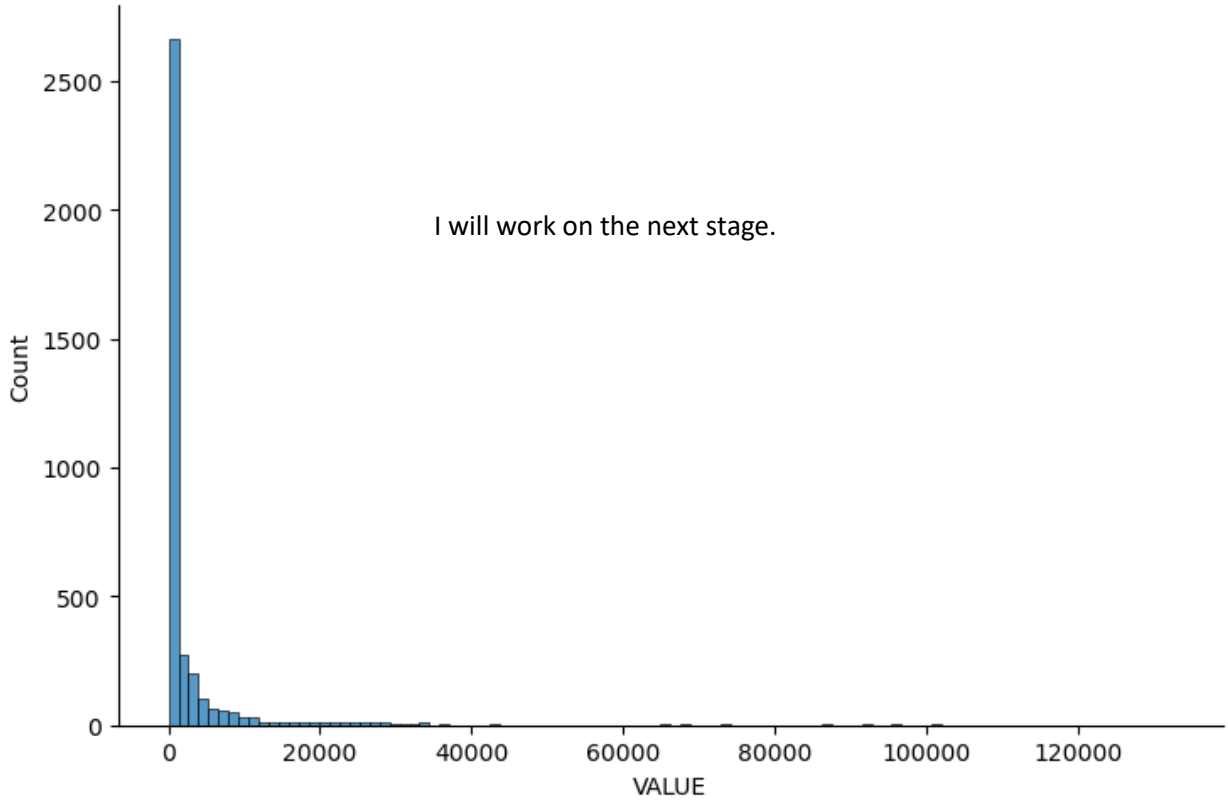
ln [ ]:

Histogram for testing dataset by age



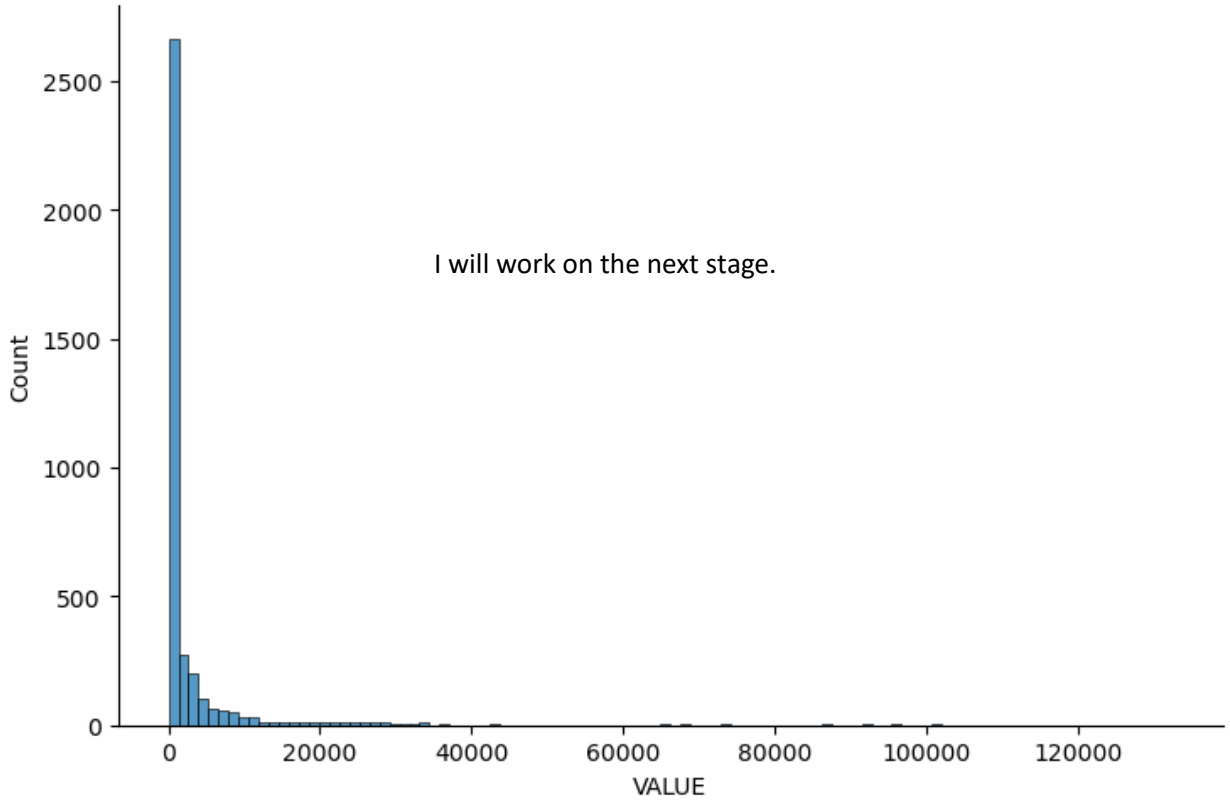
I will work on the next stage.

Histgram for testing dataset by gender



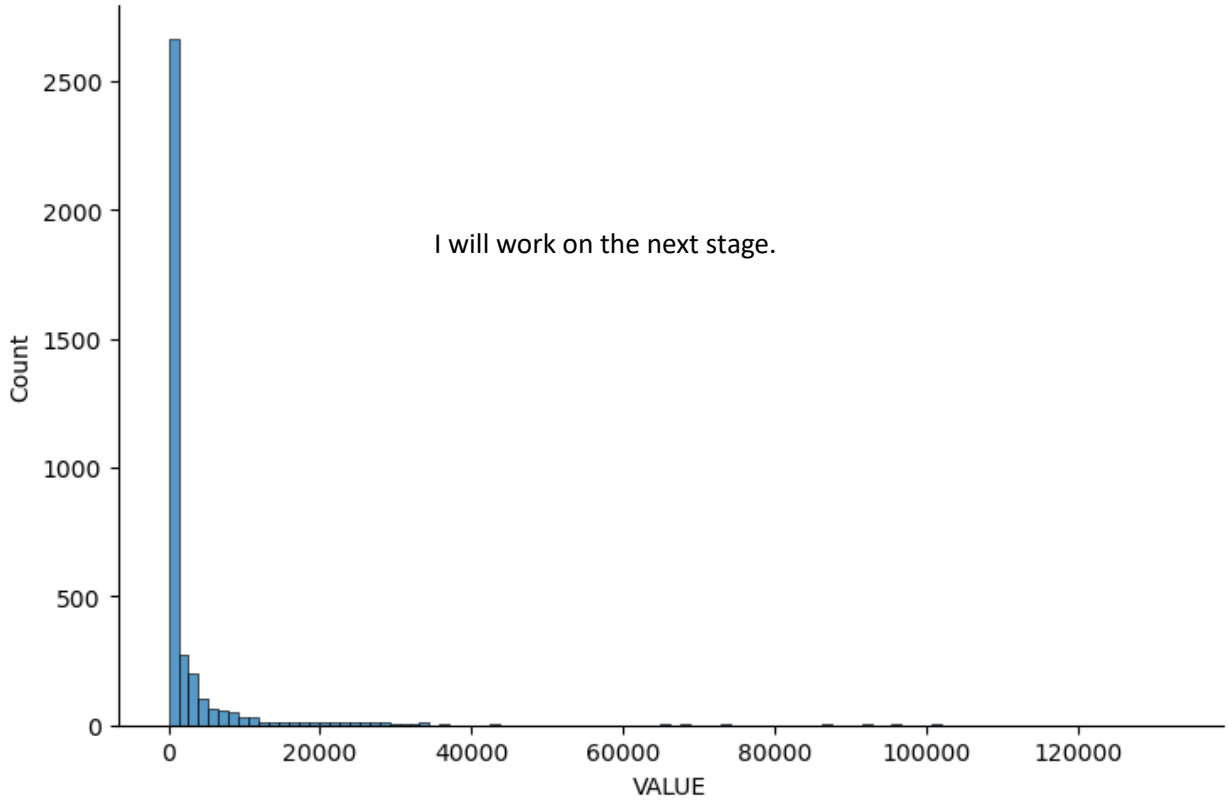
I will work on the next stage.

Histgram for testing dataset by education



I will work on the next stage.

Histgram for testing dataset by immigrant



I will work on the next stage.

Output #12f

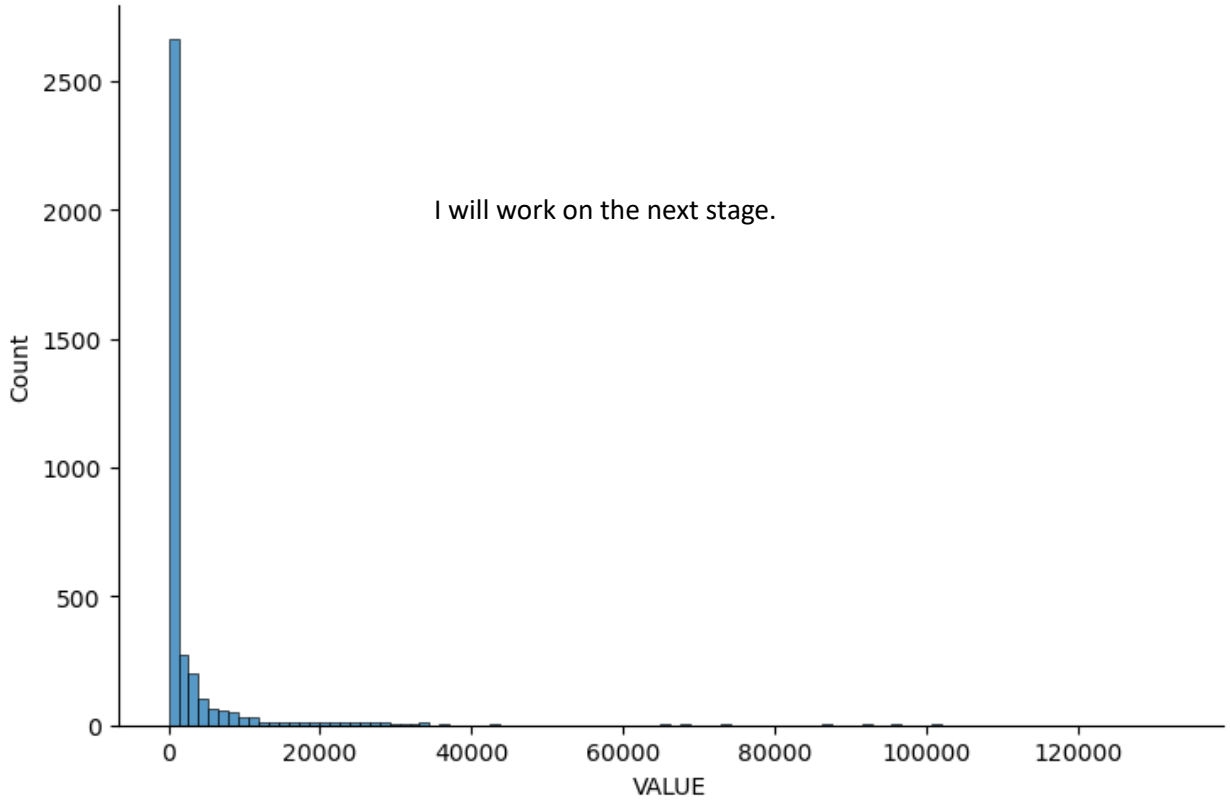
Final Output for "Number of jobs"

```
testing set By Age
                             sum          mean  amin  median      amax  size
Characteristics
15 to 24 years         3414206.0  17782.322917  24.0  4022.0  247960.0   192
25 to 34 years         7615612.0  36264.819048  19.0  6267.5  564788.0   210
35 to 44 years         7559436.0  35997.314286  18.0  6322.0  561685.0   210
45 to 54 years         7201656.0  34293.600000  16.0  5847.0  529125.0   210
55 to 64 years         6451160.0  30719.809524  13.0  5371.5  472689.0   210
65 years old and over  2416851.0  12587.765625  25.0  2729.0  173157.0   192
Overall,
Sum :  34658921.0
Mean :  28316.111928104576
Min/median/max : 13.0 / 4637.5 / 564788.0
Standard Deviation :  69539.28870799903
Skewnewss :  4.838307897340021
Total size :  1224

testing set By Gender
                         sum           mean  amin   median       amax  size
Characteristics
Female employees  24112770.0  114822.714286  37.0  19062.0  1788569.0   210
Male employees    10548592.0   50231.390476  43.0   9306.5   754820.0   210
Overall,
Sum :  34661362.0
Mean :  82527.05238095239
Min/median/max : 37.0 / 12484.5 / 1788569.0
Standard Deviation :  208846.76286467657
Skewnewss :  5.194587024584551
Total size :  420
```

```
testing set By Education
                                      sum          mean  amin   median  \
Characteristics
High school diploma and less   7204888.0  34308.990476  40.0   8564.5
Trade certificate              2245738.0  11342.111111  13.0   1865.5
University degree and higher  16513042.0  83399.202020  47.0  13812.5

                                   amax  size
Characteristics
High school diploma and less   506866.0   210
Trade certificate              167537.0   198
University degree and higher  1230063.0   198
Overall,
Sum :  25963668.0
Mean :  42844.33663366337
Min/median/max : 13.0 / 5005.0 / 1230063.0
Standard Deviation :  122102.0123600284
Skewnewss :  6.169027793480176
Total size :  606

testing set By Immigrant
                                  sum           mean   amin   median       amax  \
Characteristics
Immigrant employees       9133147.0   46127.005051   22.0   4384.0   677236.0
Non-immigrant employees  25522127.0  128899.631313  163.0  30374.0  1866156.0

                         size
Characteristics
Immigrant employees       198
Non-immigrant employees   198
Overall,
Sum :  34655274.0
Mean :  87513.31818181818
Min/median/max : 22.0 / 14330.0 / 1866156.0
Standard Deviation :  219510.15530718534
Skewnewss :  5.199524609584711
Total size :  396
```

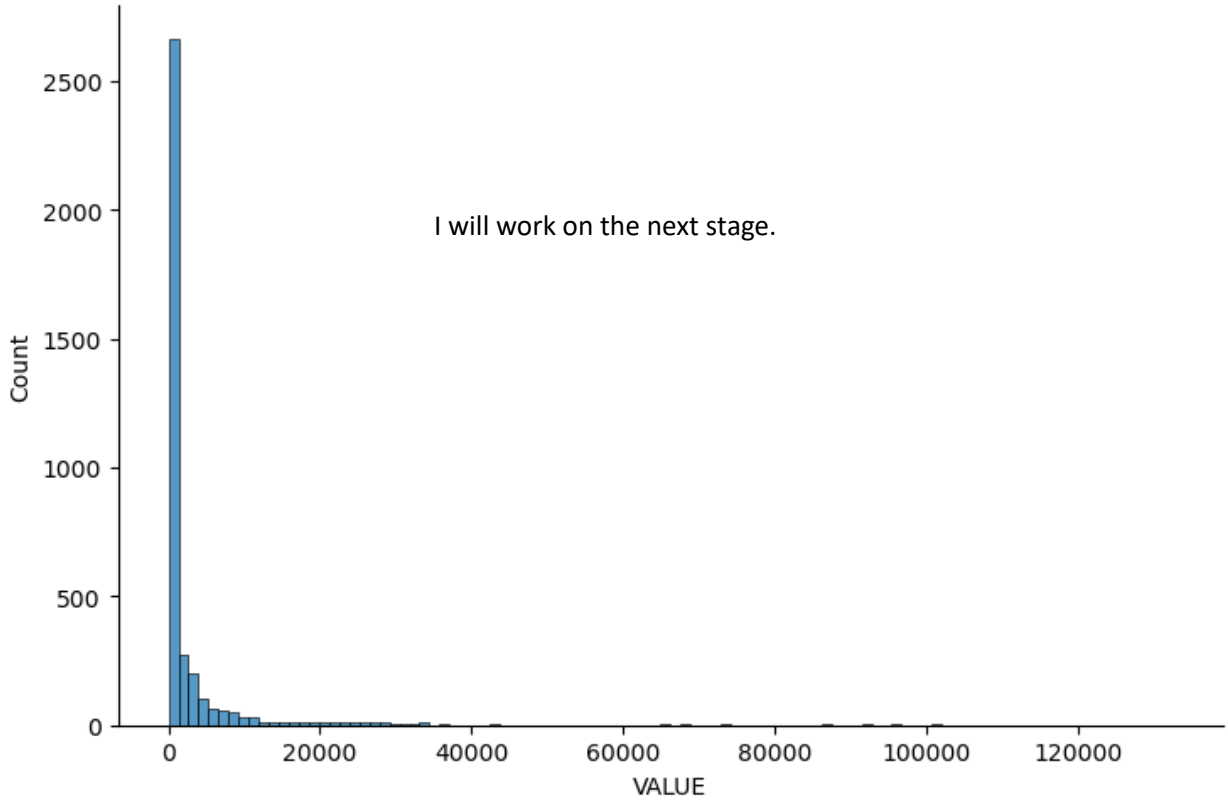
In [ ]:

Histogram for testing dataset by age



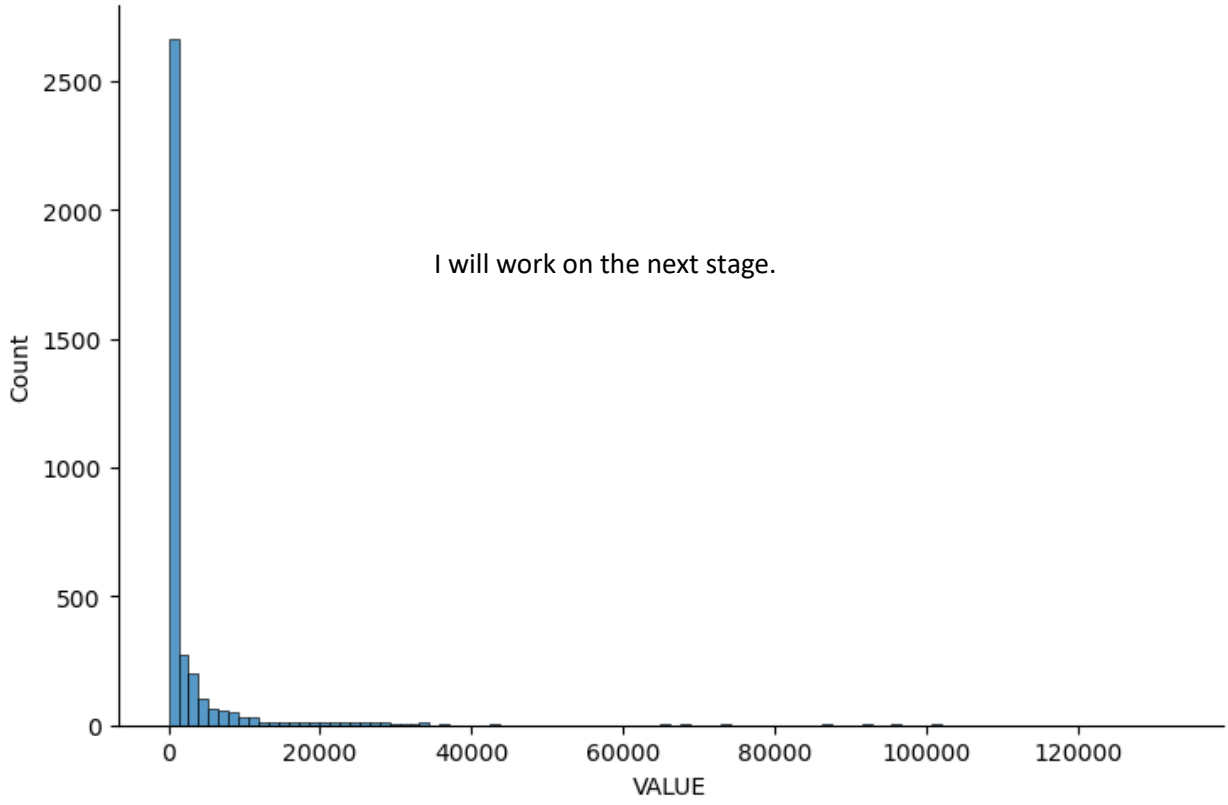
I will work on the next stage.

Histgram for testing dataset by gender



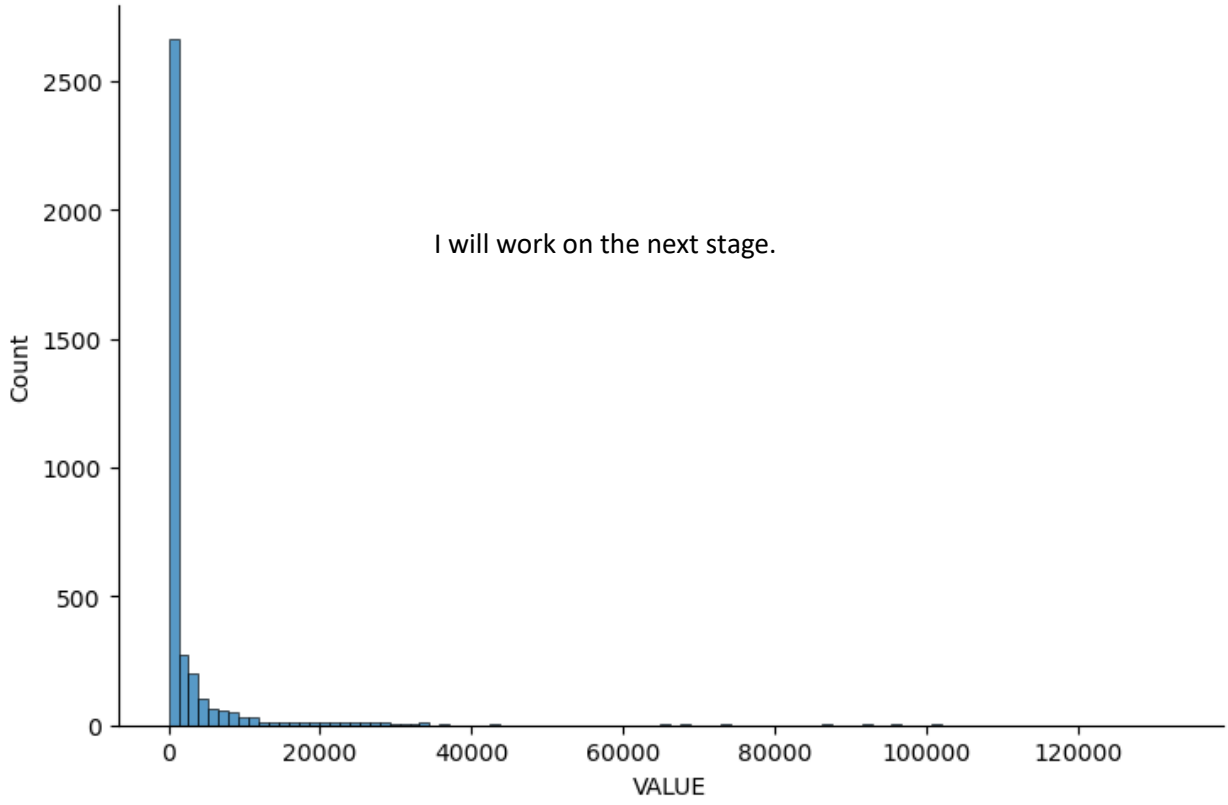
I will work on the next stage.

Histgram for testing dataset by education



I will work on the next stage.

Histgram for testing dataset by immigrant



I will work on the next stage.

Output #12g

Final Output for "Wages and Salaries"

```
testing set By Age
                             sum          mean  amin  median     amax  size
Characteristics
15 to 24 years           63069.0   328.484375   1.0    69.0   4797.0   192
25 to 34 years          334200.0  1591.428571   1.0   236.5  25966.0   210
35 to 44 years          441049.0  2100.233333   2.0   314.5  34223.0   210
45 to 54 years          474202.0  2258.104762   1.0   317.0  36376.0   210
55 to 64 years          386319.0  1839.614286   1.0   260.5  29595.0   210
65 years old and over    90513.0   471.421875   1.0    88.0   6876.0   192
Overall,
Sum :  1789352.0
Mean :  1461.888888888889
Min/median/max : 1.0 / 160.5 / 36376.0
Standard Deviation :  4021.2164121982664
Skewnewss :  5.308031050696339
Total size :  1224

testing set By Gender
                         sum          mean  amin  median     amax  size
Characteristics
Female employees  1179670.0  5617.476190   3.0   725.5  92154.0   210
Male employees     609748.0  2903.561905   3.0   547.0  45679.0   210
Overall,
Sum :  1789418.0
Mean :  4260.519047619047
Min/median/max : 3.0 / 573.0 / 92154.0
Standard Deviation :  10909.23404395687
Skewnewss :  5.063916464718029
Total size :  420
```

testing set By Education

Characteristics	sum	mean	amin	median	amax	size
High school diploma and less	226638.0	1079.228571	2.0	260.5	16432.0	210
Trade certificate	93374.0	471.585859	1.0	85.5	7213.0	198
University degree and higher	1052815.0	5317.247475	4.0	931.5	81938.0	198

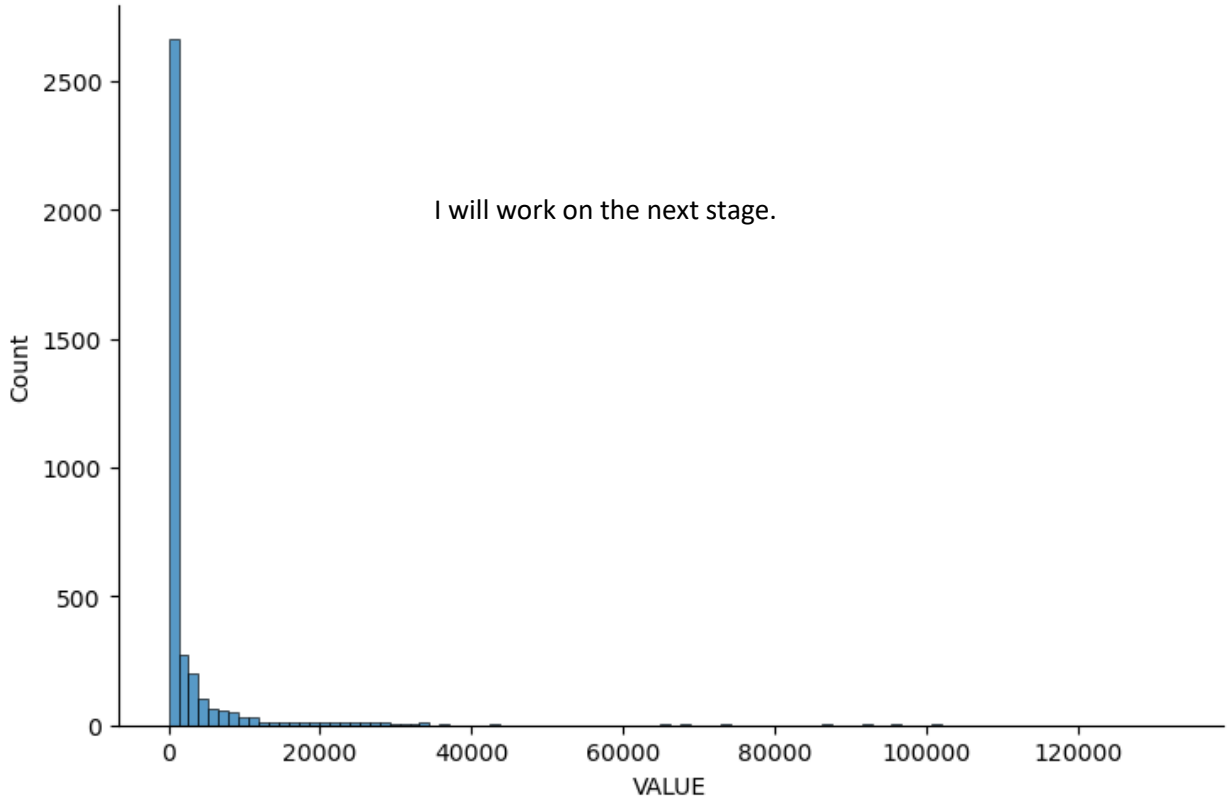
Overall,
Sum : 1372827.0
Mean : 2265.391089108911
Min/median/max : 1.0 / 215.5 / 81938.0
Standard Deviation : 7648.642116626807
Skewnewss : 7.125641720633304
Total size : 606

testing set By Immigrant

Characteristics	sum	mean	amin	median	amax	size
Immigrant employees	473697.0	2392.409091	1.0	216.5	36981.0	198
Non-immigrant employees	1315368.0	6643.272727	10.0	1452.5	100851.0	198

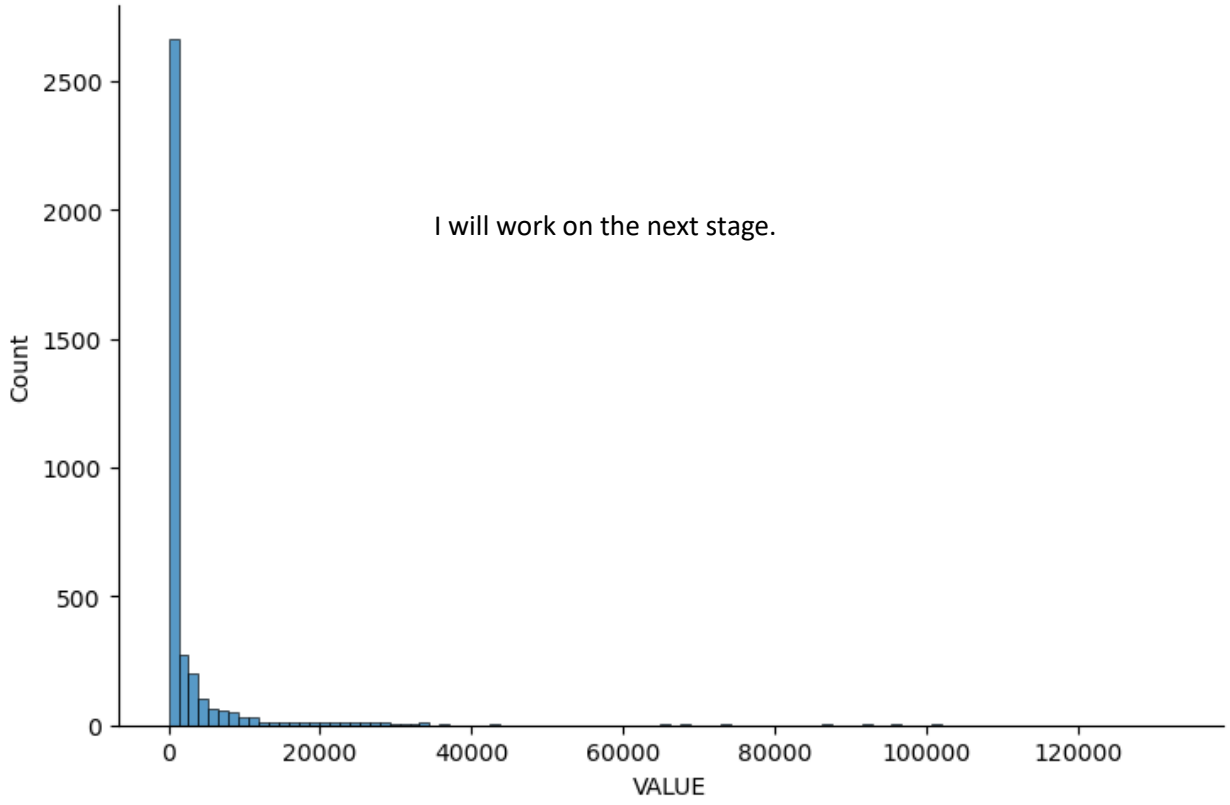
Overall,
Sum : 1789065.0
Mean : 4517.840909090909
Min/median/max : 1.0 / 578.0 / 100851.0
Standard Deviation : 11719.63396980922
Skewnewss : 5.266660488336198
Total size : 396

Histogram for testing dataset by age



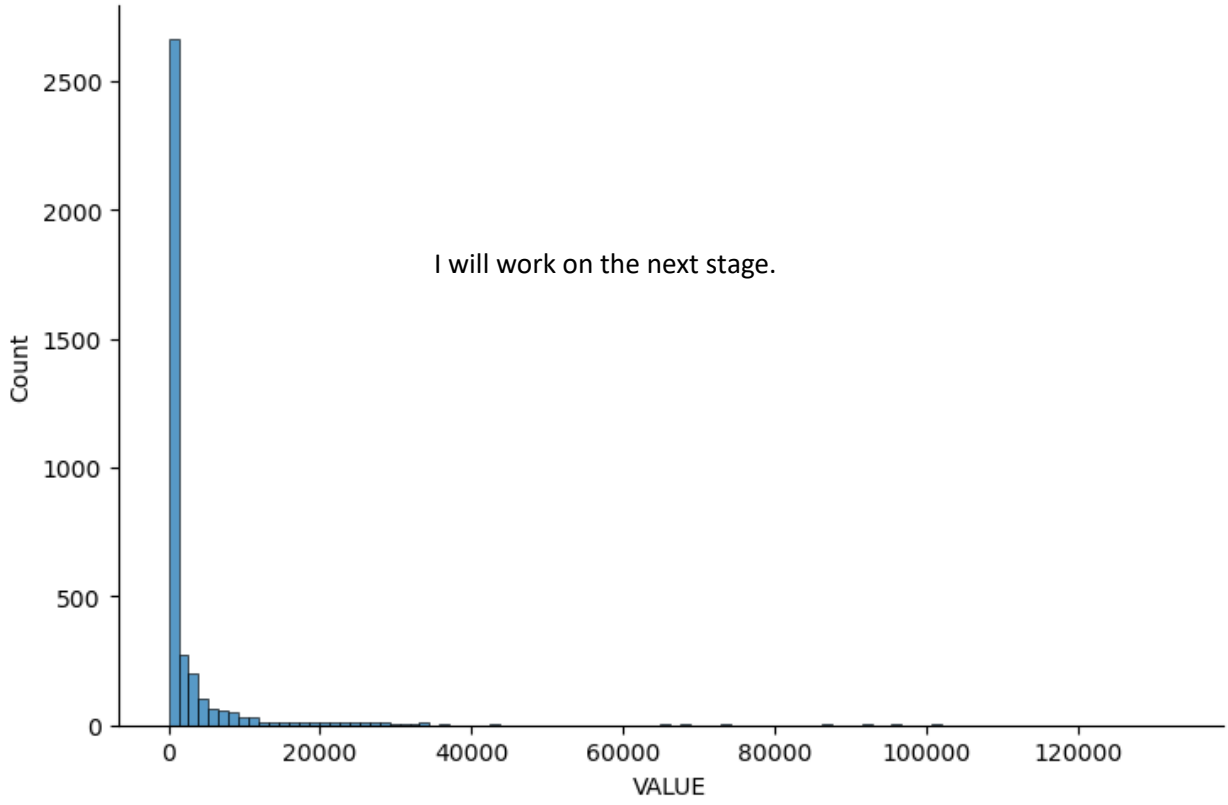
I will work on the next stage.

Histgram for testing dataset by gender



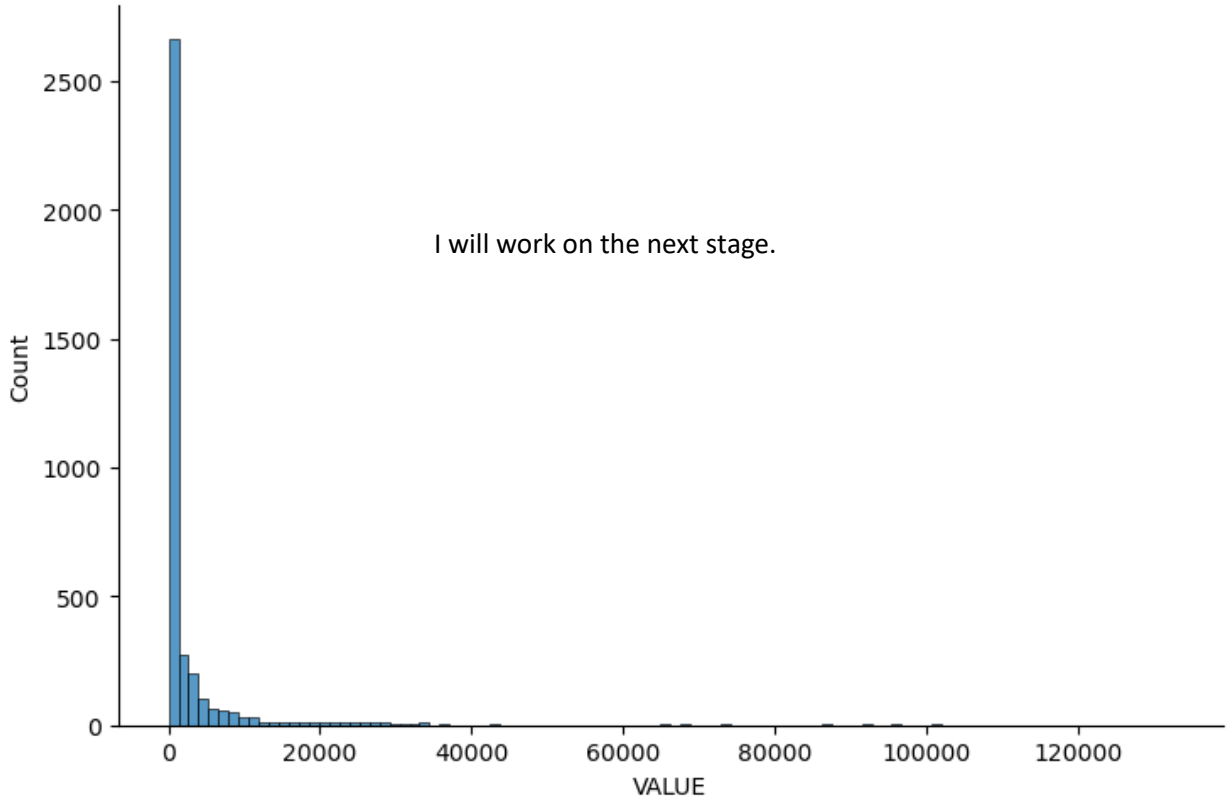
I will work on the next stage.

Histgram for testing dataset by education



I will work on the next stage.

Histgram for testing dataset by immigrant



I will work on the next stage.

Output #13

- Output for Age group after converting GEO for one hand encoding and add columns called ‘Age_group’ in numeric order.
- The Age group is grouped by youngest to oldest.
- The structure of seven indicators are look exactly like this.

```
<class 'pandas.core.frame.DataFrame'>
Index: 450 entries, 85136 to 100795
Data columns (total 14 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   REF_DATE              450 non-null    int64
 1   DGUID                 450 non-null    object
 2   Sector                450 non-null    object
 3   Characteristics       450 non-null    object
 4   Indicators            450 non-null    object
 5   UOM                   450 non-null    object
 6   SCALAR_FACTOR         450 non-null    object
 7   VALUE                 450 non-null    float64
 8   GEO_Alberta           450 non-null    bool
 9   GEO_British Columbia  450 non-null    bool
 10  GEO_Nova Scotia       450 non-null    bool
 11  GEO_Ontario           450 non-null    bool
 12  GEO_Quebec            450 non-null    bool
 13  Age_group             450 non-null    int64  ➔ [20 30 40 50 60 70]
dtypes: bool(5), float64(1), int64(2), object(6)
memory usage: 37.4+ KB
None
```

- Output for Gender group after converting GEO for one hand encoding and add columns called ‘Gender_group’ in numeric order.
- The Gender group is grouped by ‘one’ being ‘male’ and ‘zero’ being ‘female’.
- The structure of seven indicators are look exactly like this.

```
<class 'pandas.core.frame.DataFrame'>
Index: 150 entries, 85052 to 100683
Data columns (total 14 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   REF_DATE              150 non-null    int64
 1   DGUID                 150 non-null    object
 2   Sector                150 non-null    object
 3   Characteristics       150 non-null    object
 4   Indicators            150 non-null    object
 5   UOM                   150 non-null    object
 6   SCALAR_FACTOR         150 non-null    object
 7   VALUE                 150 non-null    float64
 8   GEO_Alberta           150 non-null    bool
 9   GEO_British Columbia  150 non-null    bool
 10  GEO_Nova Scotia       150 non-null    bool
 11  GEO_Ontario           150 non-null    bool
 12  GEO_Quebec            150 non-null    bool
 13  Gender_group          150 non-null    int32  ➔ [1 0]
dtypes: bool(5), float64(1), int32(1), int64(1), object(6)
memory usage: 11.9+ KB
None
```

- Output for Education level after converting GEO for one hand encoding and add columns called ‘Education_group’ in numeric order.
- The Education group is being organized from lowest (1) education to highest (3) education.
- The structure of seven indicators are look exactly like this.

```
<class 'pandas.core.frame.DataFrame'>
Index: 225 entries, 85108 to 100753
Data columns (total 14 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   REF_DATE              225 non-null    int64
 1   DGUID                 225 non-null    object
 2   Sector                225 non-null    object
 3   Characteristics       225 non-null    object
 4   Indicators            225 non-null    object
 5   UOM                   225 non-null    object
 6   SCALAR_FACTOR         225 non-null    object
 7   VALUE                 225 non-null    float64
 8   GEO_Alberta           225 non-null    bool
 9   GEO_British Columbia  225 non-null    bool
 10  GEO_Nova Scotia       225 non-null    bool
 11  GEO_Ontario           225 non-null    bool
 12  GEO_Quebec            225 non-null    bool
 13  Education_group       225 non-null    int64  ➔ [1 2 3]
dtypes: bool(5), float64(1), int64(2), object(6)
memory usage: 18.7+ KB
None
```

- Output for Immigrant status after converting GEO for one hand encoding and add columns called ‘Immigrant in numeric order.
- The Immigrant status is organized by ‘zero’ being immigrant and ‘one’ being non-immigrant.
- The structure of seven indicators are look exactly like this.

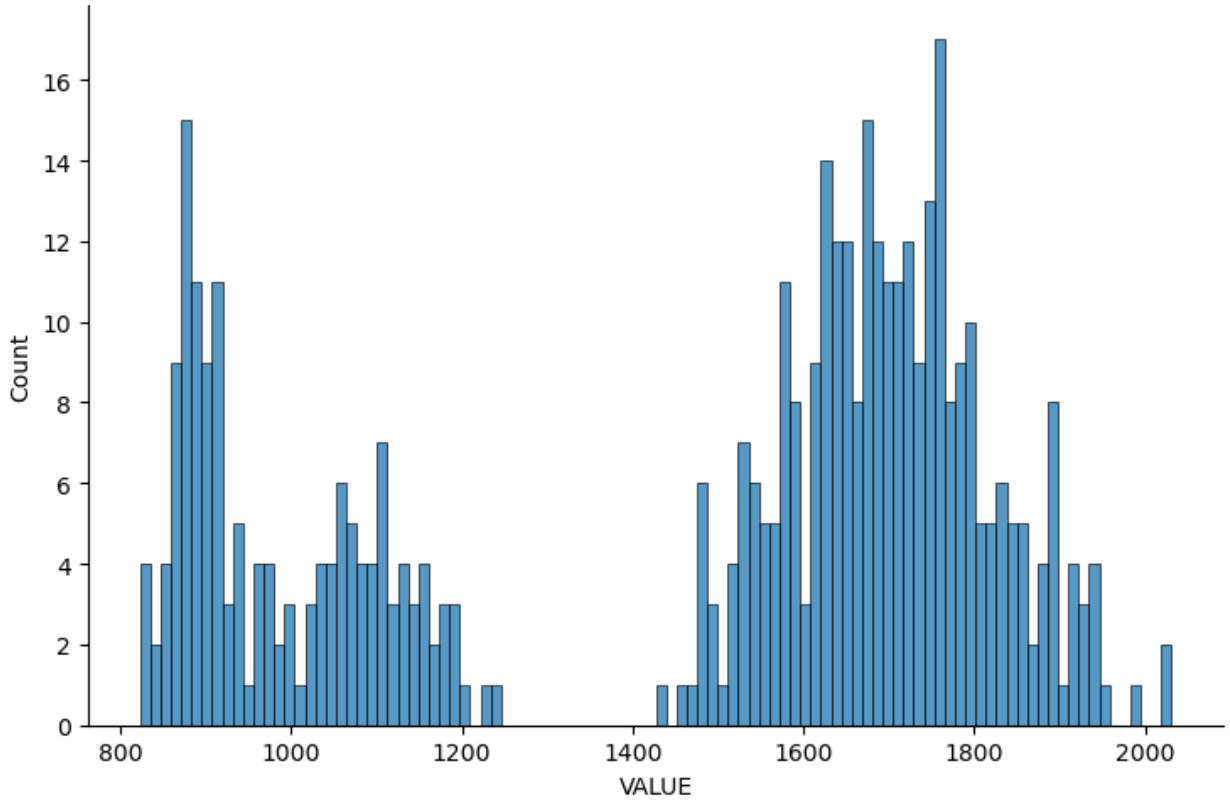
```
<class 'pandas.core.frame.DataFrame'>
Index: 150 entries, 85066 to 100697
Data columns (total 14 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   REF_DATE              150 non-null    int64
 1   DGUID                 150 non-null    object
 2   Sector                150 non-null    object
 3   Characteristics       150 non-null    object
 4   Indicators            150 non-null    object
 5   UOM                   150 non-null    object
 6   SCALAR_FACTOR         150 non-null    object
 7   VALUE                 150 non-null    float64
 8   GEO_Alberta           150 non-null    bool
 9   GEO_British Columbia  150 non-null    bool
 10  GEO_Nova Scotia       150 non-null    bool
 11  GEO_Ontario           150 non-null    bool
 12  GEO_Quebec            150 non-null    bool
 13  Immigrant_status      150 non-null    int32  ➔ [0 1]
dtypes: bool(5), float64(1), int32(1), int64(1), object(6)
memory usage: 11.9+ KB
None
```

Output #14a

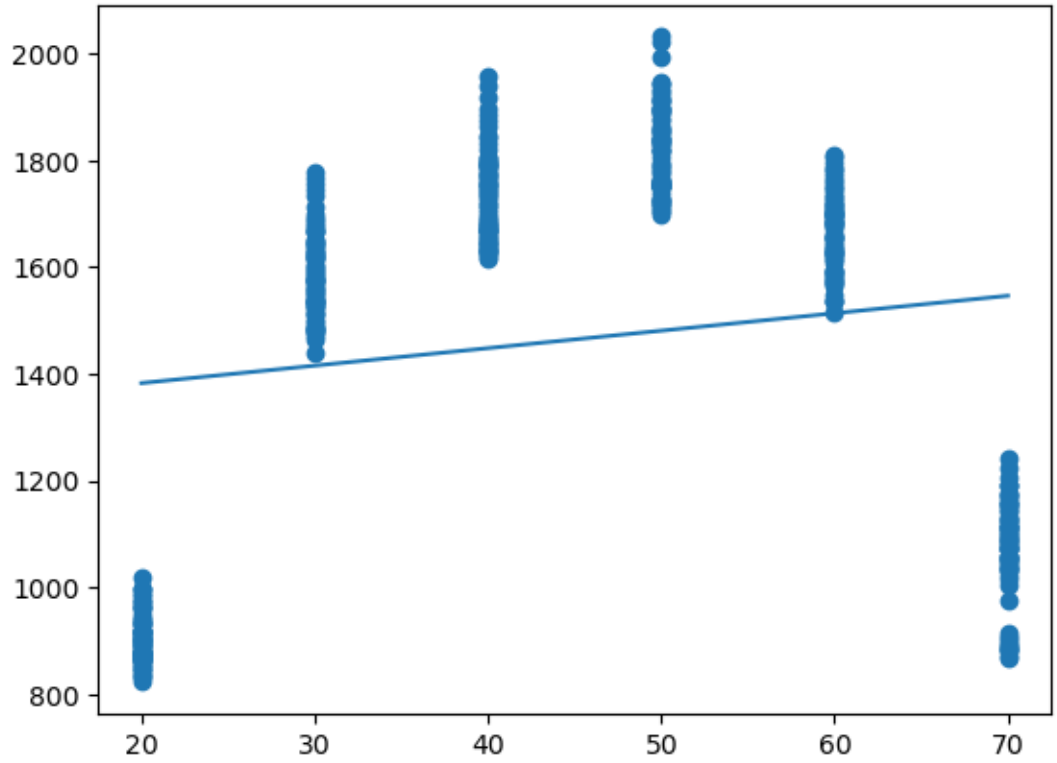
Result for testing set for 'Average annual hours worked' by Age group

```
final_testing_df_output_df_AvgAnnHrsWrk_ByAge.csv
                             sum         mean    amin  median    amax  size
Characteristics
15 to 24 years           67850.0   904.666667   823.0   897.0  1020.0    75
25 to 34 years          119194.0  1589.253333  1438.0  1579.0  1779.0    75
35 to 44 years          130076.0  1734.346667  1615.0  1732.0  1958.0    75
45 to 54 years          136664.0  1822.186667  1697.0  1821.0  2031.0    75
55 to 64 years          124987.0  1666.493333  1515.0  1676.0  1810.0    75
65 years old and over    80225.0  1069.666667   867.0  1085.0  1242.0    75
Overall,
Sum :  658996.0
Mean :  1464.4355555555555
Min/median/max : 823.0 / 1626.0 / 2031.0
Standard Deviation :  356.3536590121646
Skewnewss :  -0.5806984535776861
Total size :  450
```

final_testing_df_output_df_AvgAnnHrsWrk_ByAge.csv



final_testing_df_output_df_AvgAnnHrsWrk_ByAge.csv



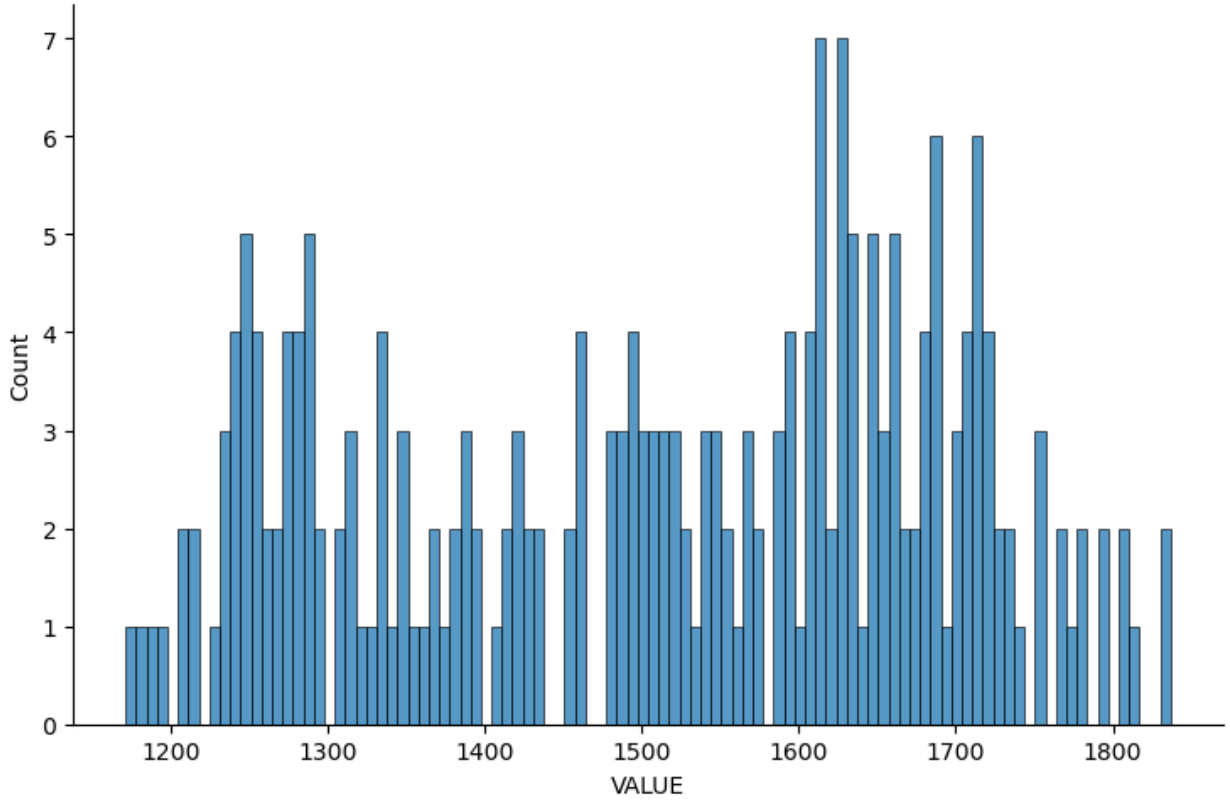
Done by Linear Regression

Output #14b

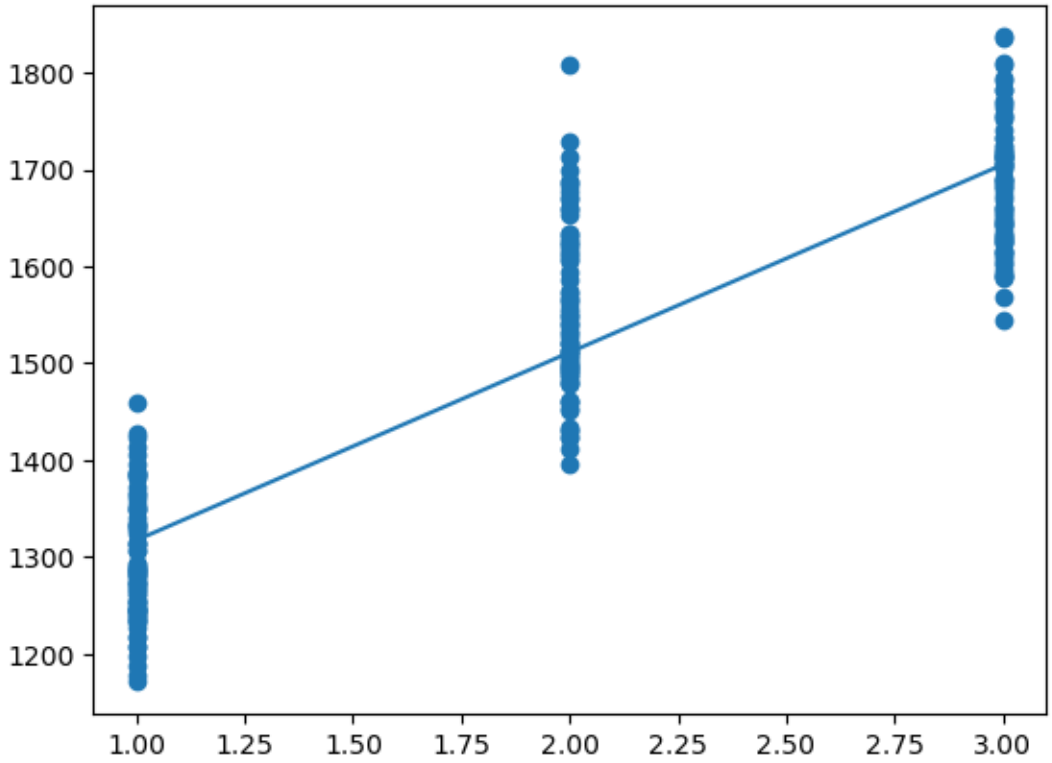
Result for testing set for 'Average annual hours worked' by Education level

```
                                sum         mean    amin  median    amax  \
Characteristics
High school diploma and less   97223.0  1296.306667  1171.0  1285.0  1460.0
Trade certificate             116506.0  1553.413333  1396.0  1538.0  1808.0
University degree and higher  126305.0  1684.066667  1544.0  1684.0  1837.0

                              size
Characteristics
High school diploma and less    75
Trade certificate               75
University degree and higher    75
Overall,
Sum :  340034.0
Mean :  1511.2622222222221
Min/median/max : 1171.0 / 1538.0 / 1837.0
Standard Deviation :  176.936517540914
Skewnewss :  -0.22126026577868005
Total size :  225
```

final_testing_df_output_df_AvgAnnHrsWrk_ByEducation.csv



final_testing_df_output_df_AvgAnnHrsWrk_ByEducation.csv



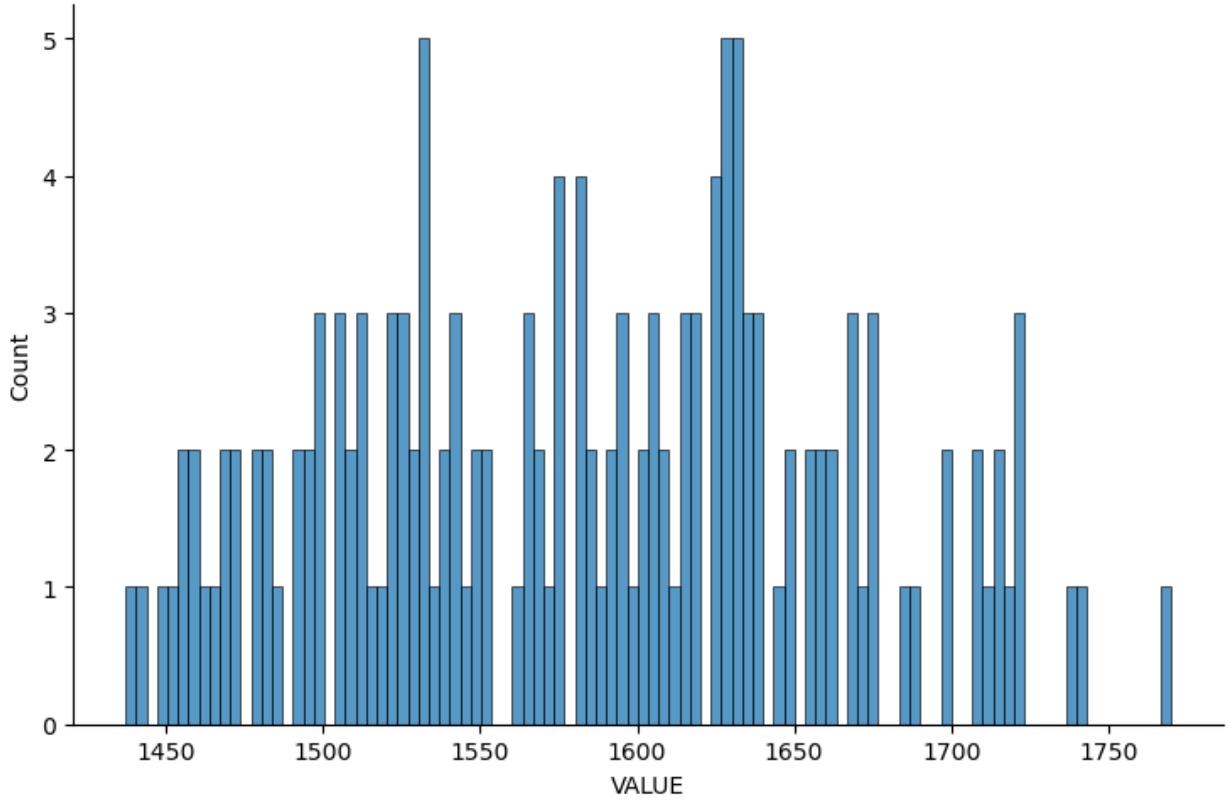
Done by Linear Regression
Higher the number, higher the education

Output #14c

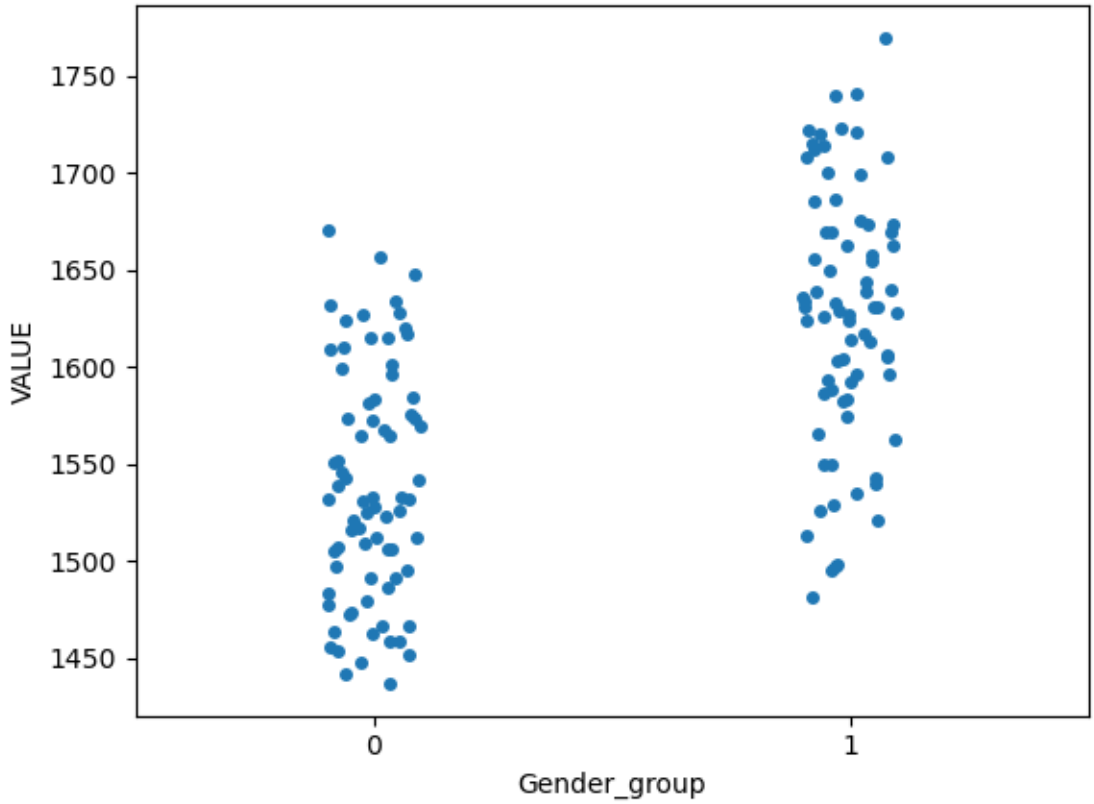
Result for testing set for 'Average annual hours worked' by Gender group

```
final_testing_df_output_df_AvgAnnHrsWrk_ByGender.csv
                       sum         mean    amin  median    amax  size
Characteristics
Female employees  115375.0  1538.333333  1437.0  1532.0  1671.0    75
Male employees    122021.0  1626.946667  1481.0  1631.0  1770.0    75
Overall,
Sum :  237396.0
Mean :  1582.64
Min/median/max : 1437.0 / 1583.5 / 1770.0
Standard Deviation :  77.0804151519697
Skewnewss :  0.11983152740265923
Total size :  150
```

final_testing_df_output_df_AvgAnnHrsWrk_ByGender.csv



final_testing_df_output_df_AvgAnnHrsWrk_ByGender.csv



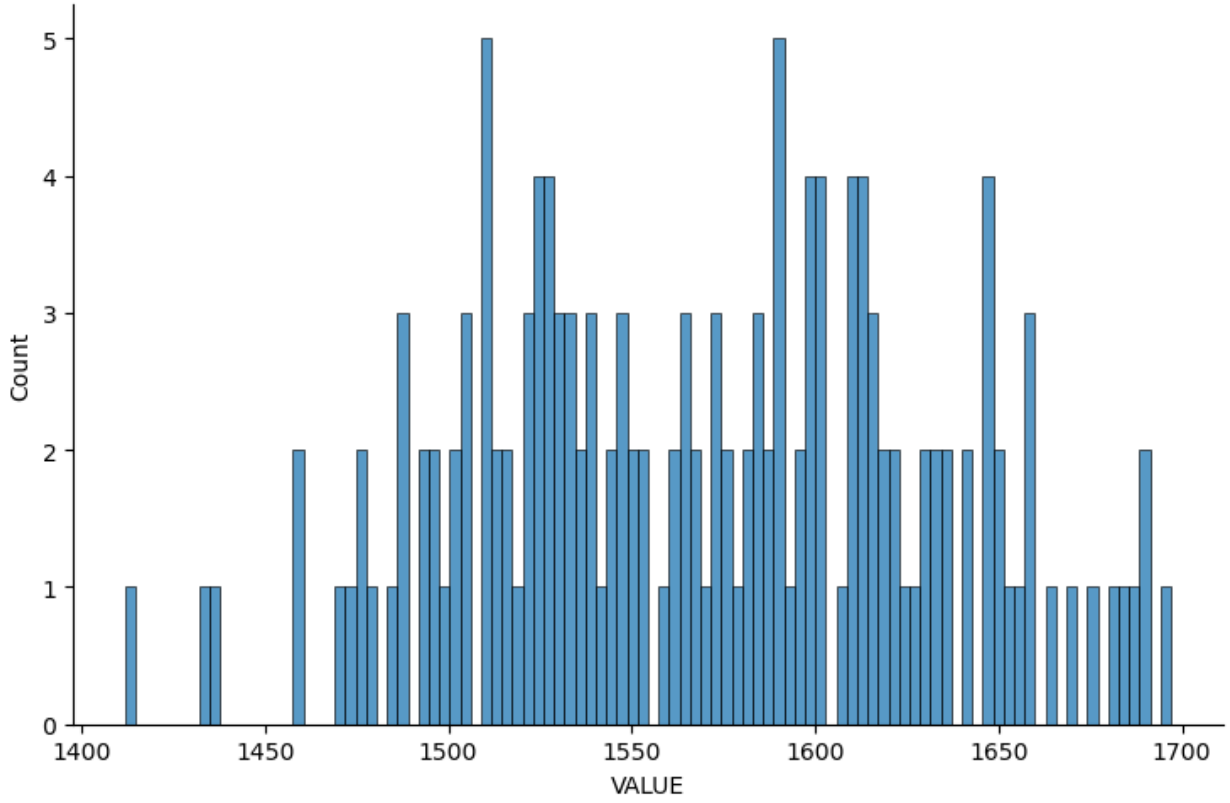
Done using Stripplot
[1, 0] = ['Male employees' 'Female employees']

Output #14d

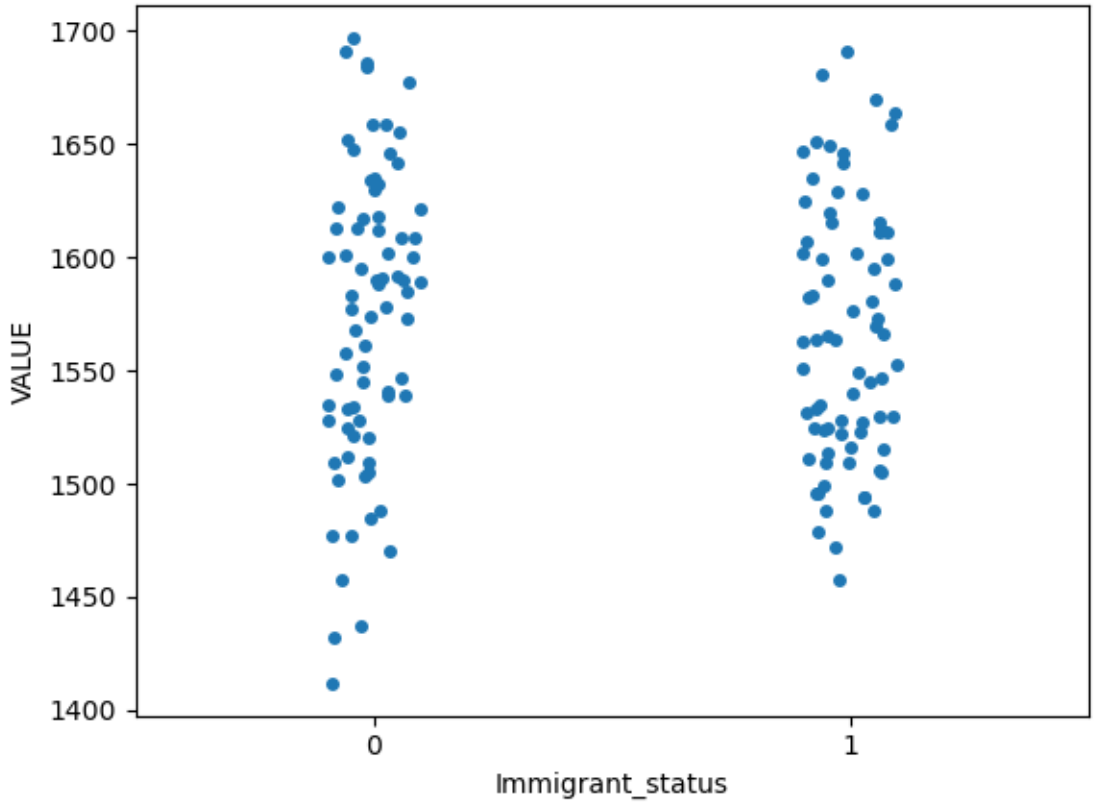
Result for testing set for 'Average annual hours worked' by Immigrant status

```
final_testing_df_output_df_AvgAnnHrsWrk_ByImmigrant.csv
                             sum         mean    amin  median    amax  size
Characteristics
Immigrant employees      117967.0  1572.893333  1412.0  1583.0  1697.0    75
Non-immigrant employees  117437.0  1565.826667  1458.0  1564.0  1691.0    75
Overall,
Sum :  235404.0
Mean :  1569.36
Min/median/max : 1412.0 / 1571.5 / 1697.0
Standard Deviation :  60.9140684352419
Skewnewss :  -0.023261406680708266
Total size :  150
```

final_testing_df_output_df_AvgAnnHrsWrk_ByImmigrant.csv



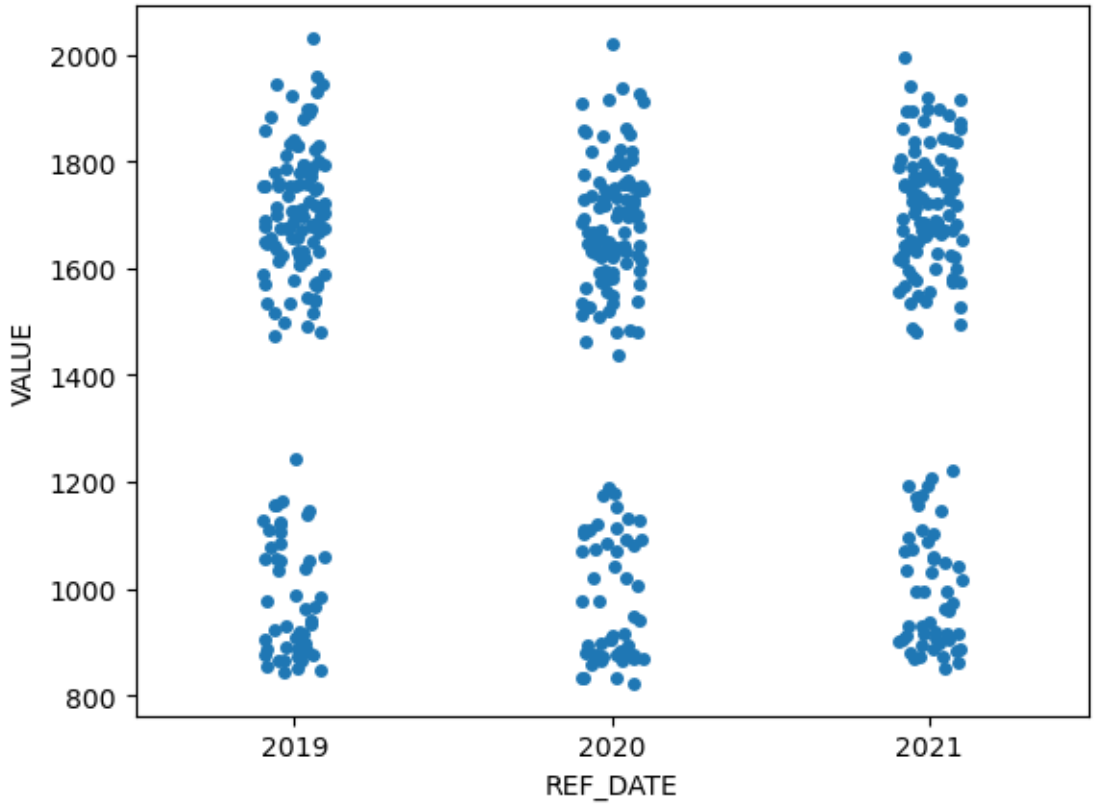
final_testing_df_output_df_AvgAnnHrsWrk_ByImmigrant.csv



Done using Stripplot
[0, 1] = ['Immigrant employees' 'Non-immigrant employees']

Result for testing set for 'Average annual hours worked' by yearly

`final_testing_df_output_df_AvgAnnHrsWrk_ByAge.csv`



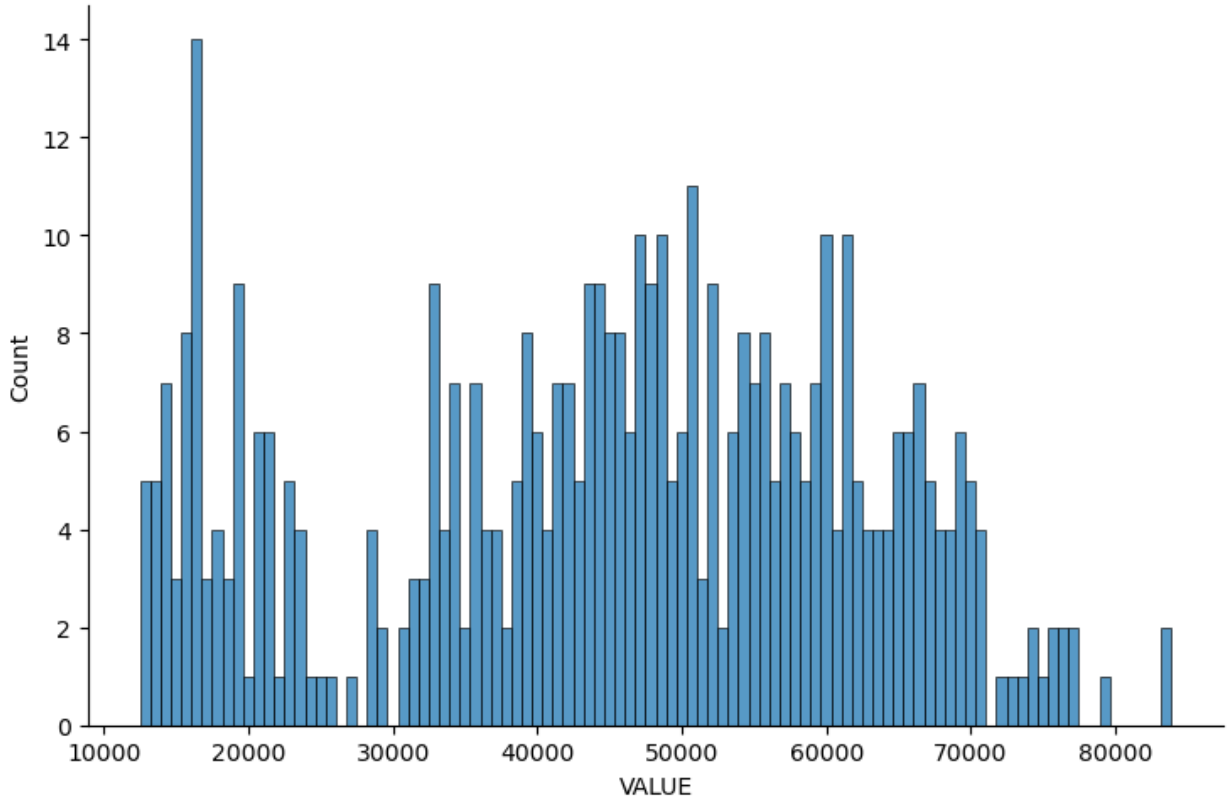
`Done using Stripplot`

Output #15a

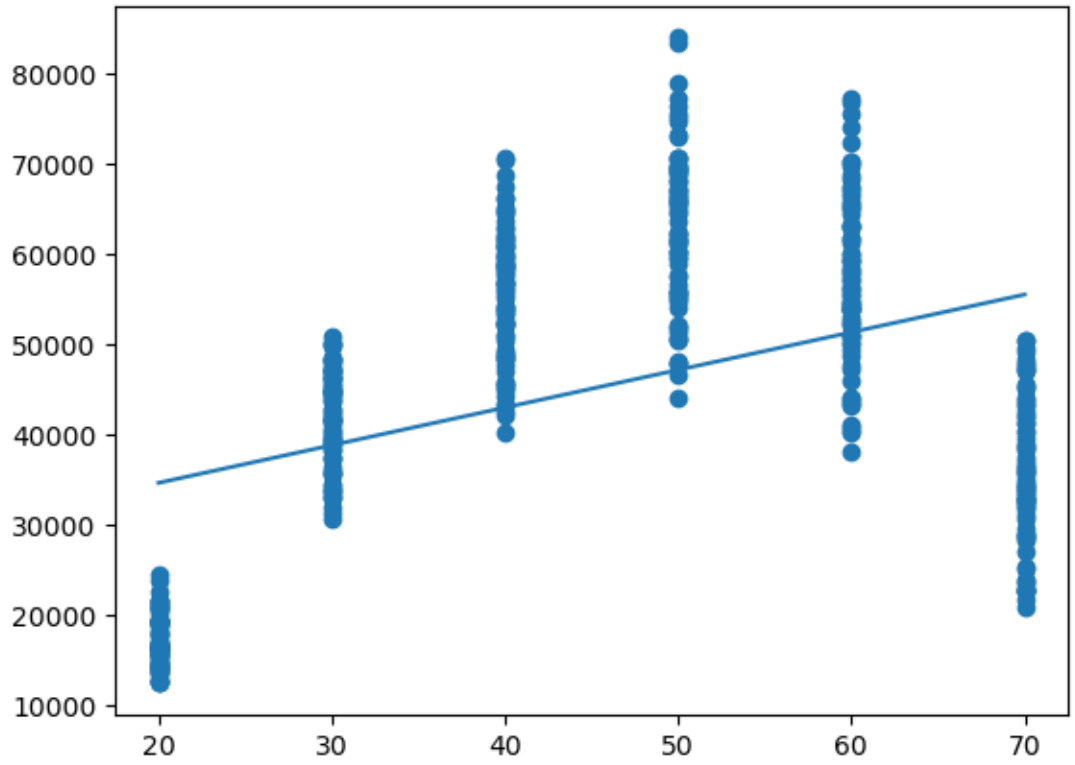
Result for testing set for 'Average annual wages and salaries' by Age group

```
final_testing_df_output_df_AvgAnnWages_ByAge.csv
                             sum          mean     amin   median     amax  \
Characteristics
15 to 24 years         1286809.0  17157.453333  12468.0  16497.0  24382.0
25 to 34 years         3134044.0  41787.253333  30721.0  42140.0  50918.0
35 to 44 years         4169234.0  55589.786667  40250.0  56501.0  70618.0
45 to 54 years         4721272.0  62950.293333  43944.0  63689.0  83894.0
55 to 64 years         4284136.0  57121.813333  38066.0  56400.0  77141.0
65 years old and over  2675845.0  35677.933333  20934.0  35117.0  50445.0

                       size
Characteristics
15 to 24 years           75
25 to 34 years           75
35 to 44 years           75
45 to 54 years           75
55 to 64 years           75
65 years old and over    75
Overall,
Sum :  20271340.0
Mean :  45047.42222222222
Min/median/max : 12468.0 / 46985.0 / 83894.0
Standard Deviation :  17188.866523407123
Skewnewss :  -0.2534546294165646
Total size :  450
```

final_testing_df_output_df_AvgAnnWages_ByAge.csv



final_testing_df_output_df_AvgAnnWages_ByAge.csv



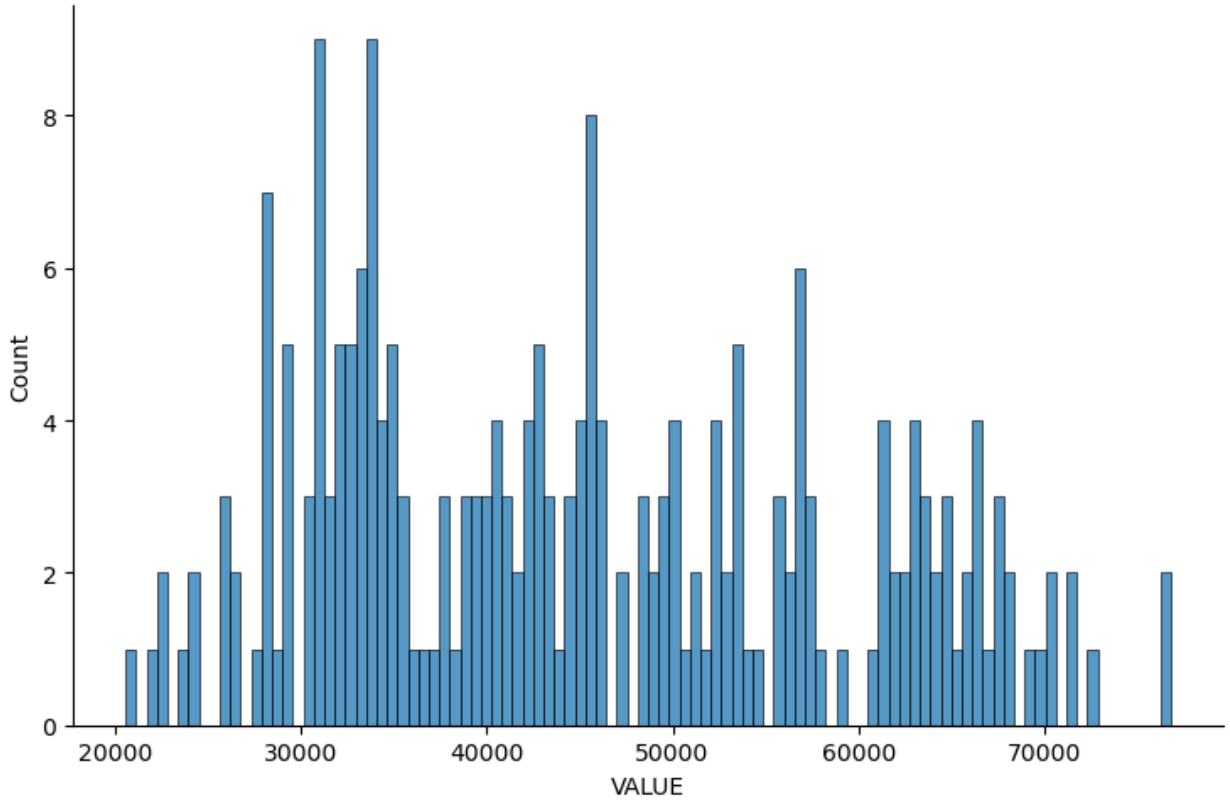
Done by Linear Regression

Output #15b

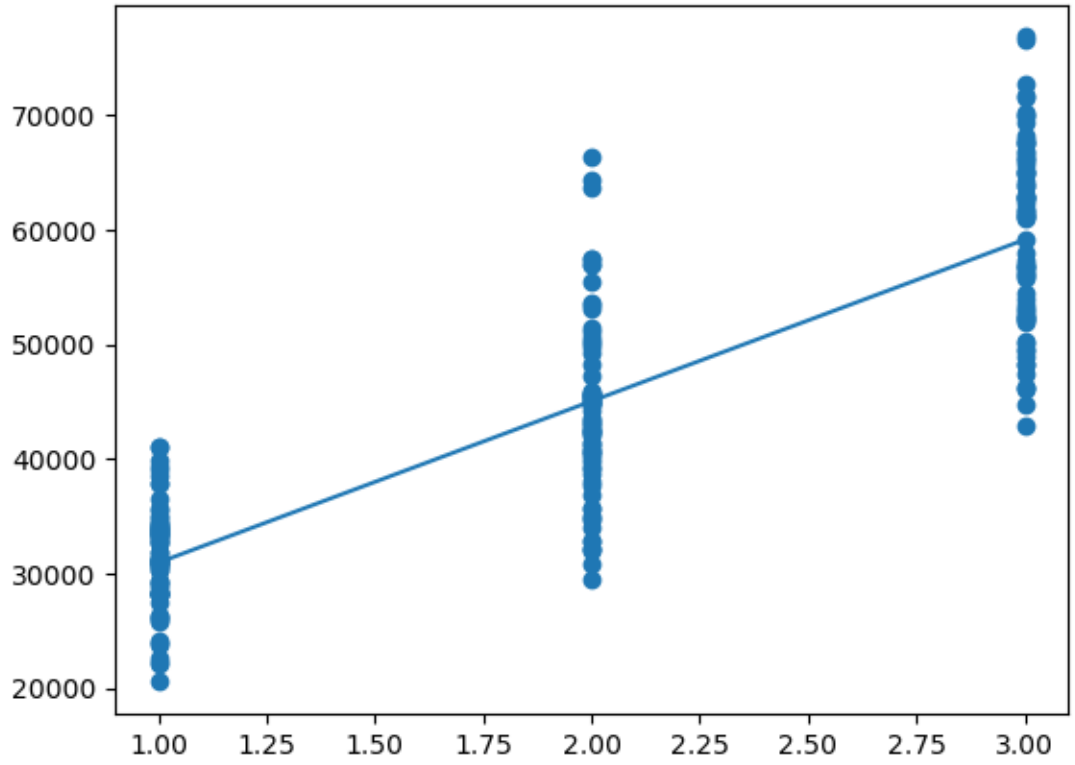
Result for testing set for 'Average annual wages and salaries' by Education level

```
final_testing_df_output_df_AvgAnnWages_ByEducation.csv
                                    sum          mean     amin   median  \
Characteristics
High school diploma and less  2355715.0  31409.533333  20536.0  31272.0
Trade certificate             3308944.0  44119.253333  29523.0  43448.0
University degree and higher  4469360.0  59591.466667  42833.0  61166.0

                                   amax  size
Characteristics
High school diploma and less  41040.0    75
Trade certificate             66374.0    75
University degree and higher  76827.0    75
Overall,
Sum :  10134019.0
Mean :  45040.084444444445
Min/median/max : 20536.0 / 43197.0 / 76827.0
Standard Deviation :  13473.728814811864
Skewnewss :  0.3611910342438643
Total size :  225
```

final_testing_df_output_df_AvgAnnWages_ByEducation.csv



final_testing_df_output_df_AvgAnnWages_ByEducation.csv



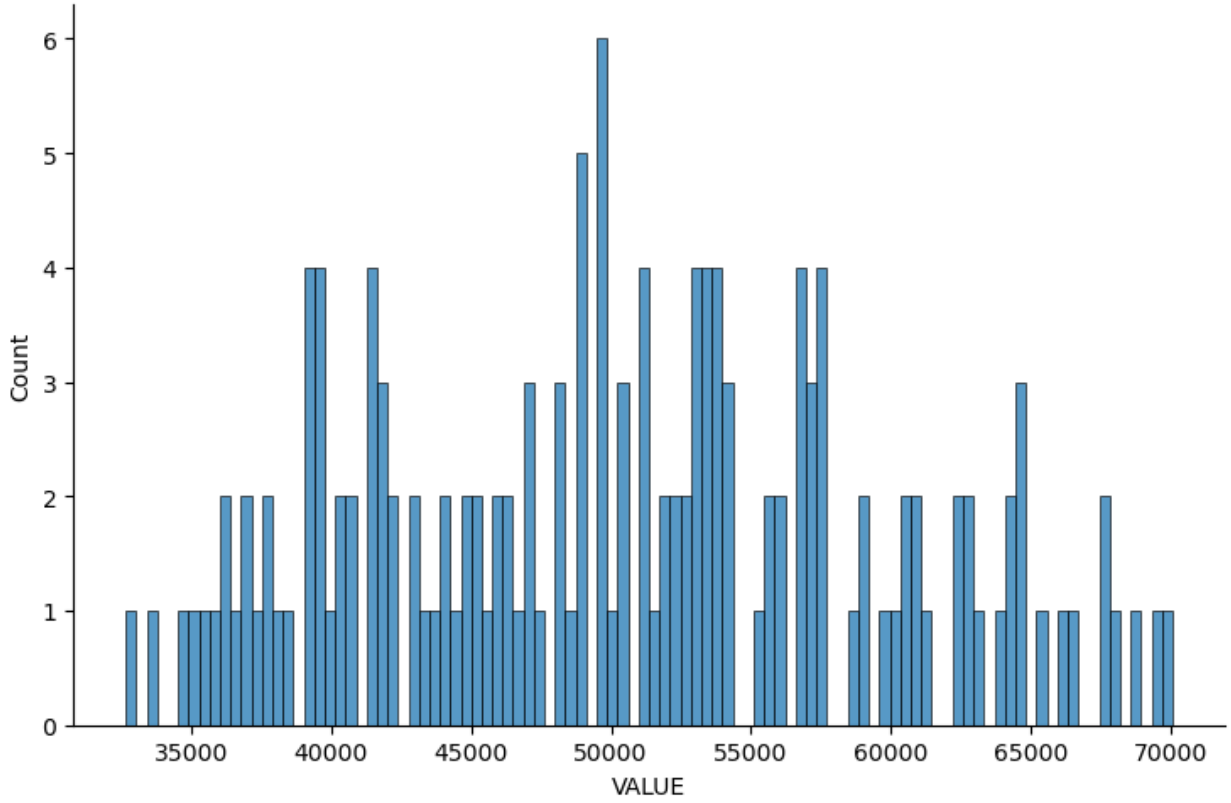
Done by Linear Regression
Higher the number, higher the education

Output #15c

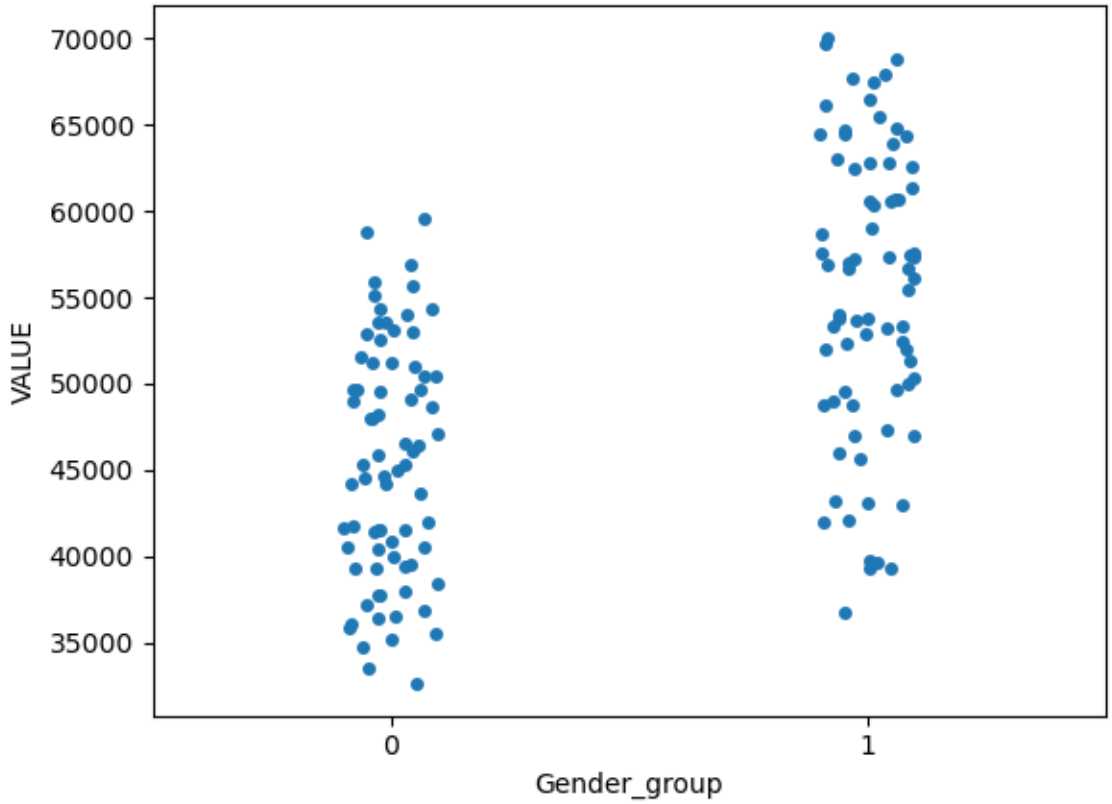
Result for testing set for 'Average annual wages and salaries' by Gender group

```
final_testing_df_output_df_AvgAnnWages_ByGender.csv
                        sum          mean     amin   median     amax  size
Characteristics
Female employees  3401364.0  45351.520000  32646.0  45290.0  59616.0    75
Male employees    4150456.0  55339.413333  36768.0  56668.0  70067.0    75
Overall,
Sum :  7551820.0
Mean :  50345.46666666667
Min/median/max : 32646.0 / 50189.0 / 70067.0
Standard Deviation :  9110.675668808664
Skewnewss :  0.1568211708397948
Total size :  150
```

final_testing_df_output_df_AvgAnnWages_ByGender.csv



final_testing_df_output_df_AvgAnnWages_ByGender.csv



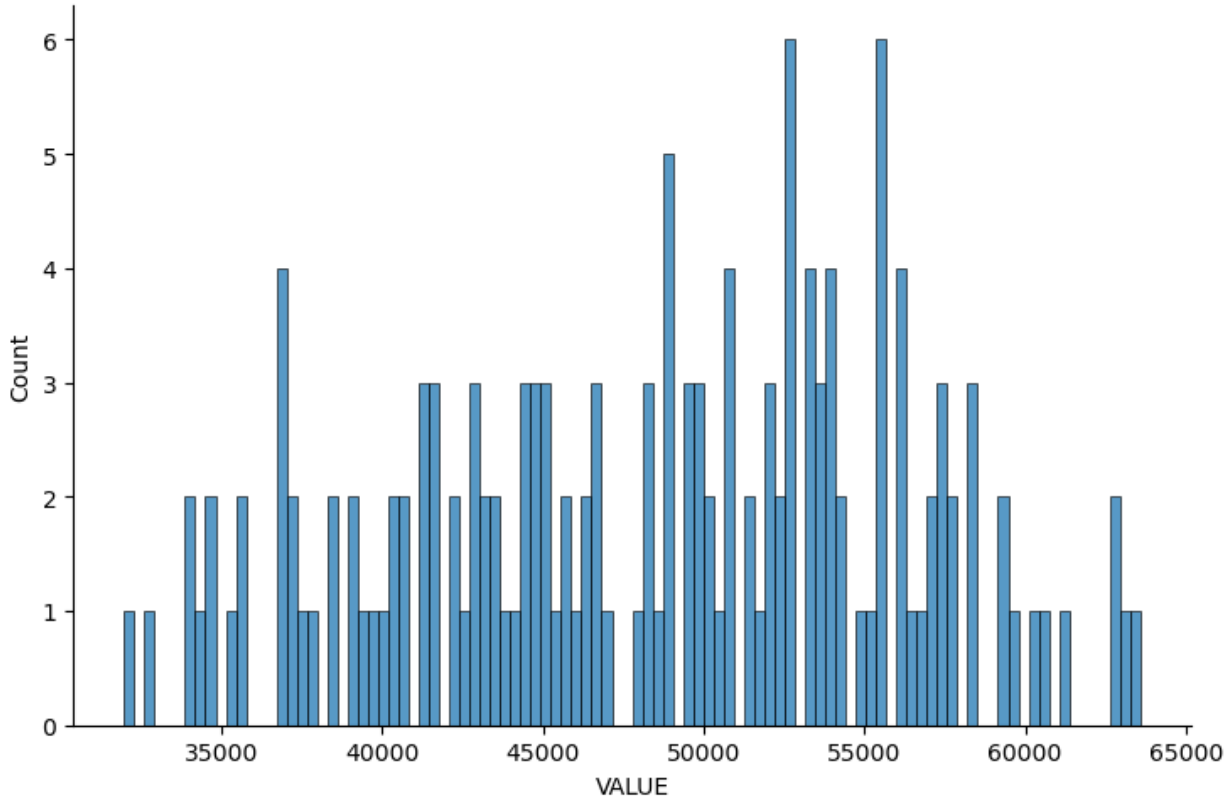
Done using Stripplot
[1, 0] = ['Male employees' 'Female employees']

Output #15c

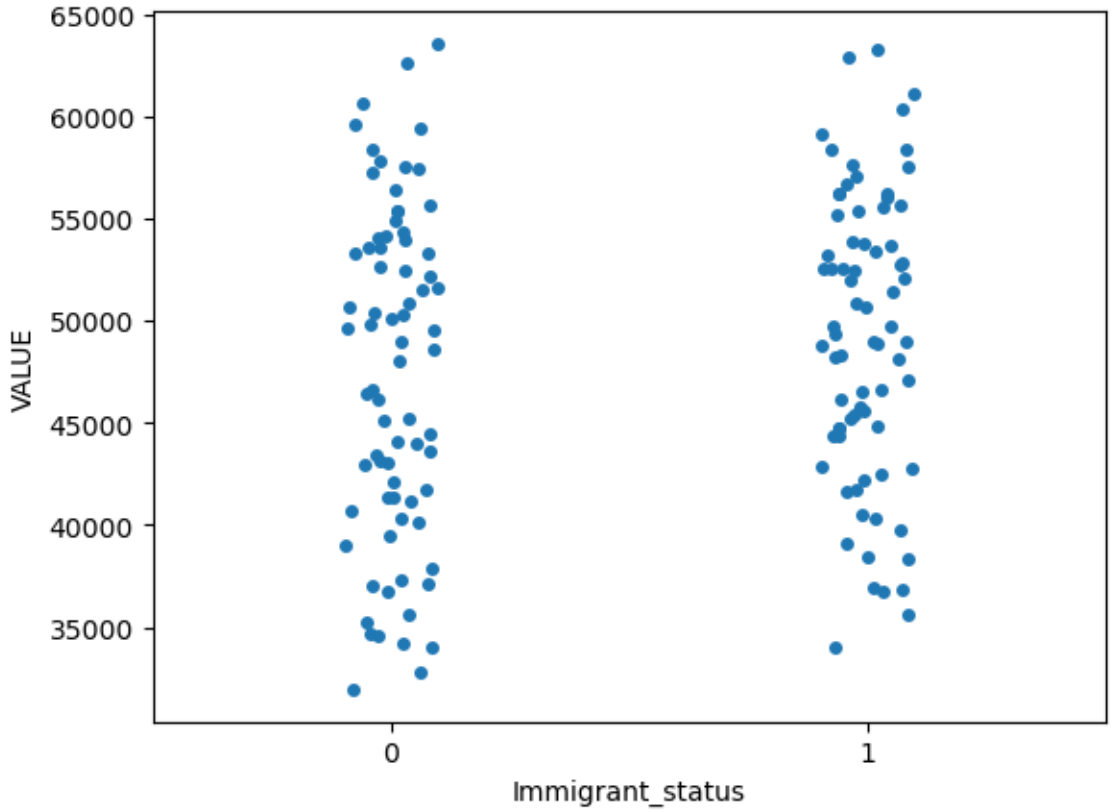
Result for testing set for 'Average annual wages and salaries' by Immigrant status

```
final_testing_df_output_df_AvgAnnWages_ByImmigrant.csv
                              sum          mean     amin   median     amax  \
Characteristics
Immigrant employees      3554950.0  47399.333333  31977.0  48615.0  63582.0
Non-immigrant employees  3690998.0  49213.306667  34013.0  49374.0  63265.0

                         size
Characteristics
Immigrant employees        75
Non-immigrant employees    75
Overall,
Sum :  7245948.0
Mean :  48306.32
Min/median/max : 31977.0 / 49016.0 / 63582.0
Standard Deviation :  7581.937589930426
Skewnewss :  -0.16449979628579625
Total size :  150
```

final_testing_df_output_df_AvgAnnWages_ByImmigrant.csv



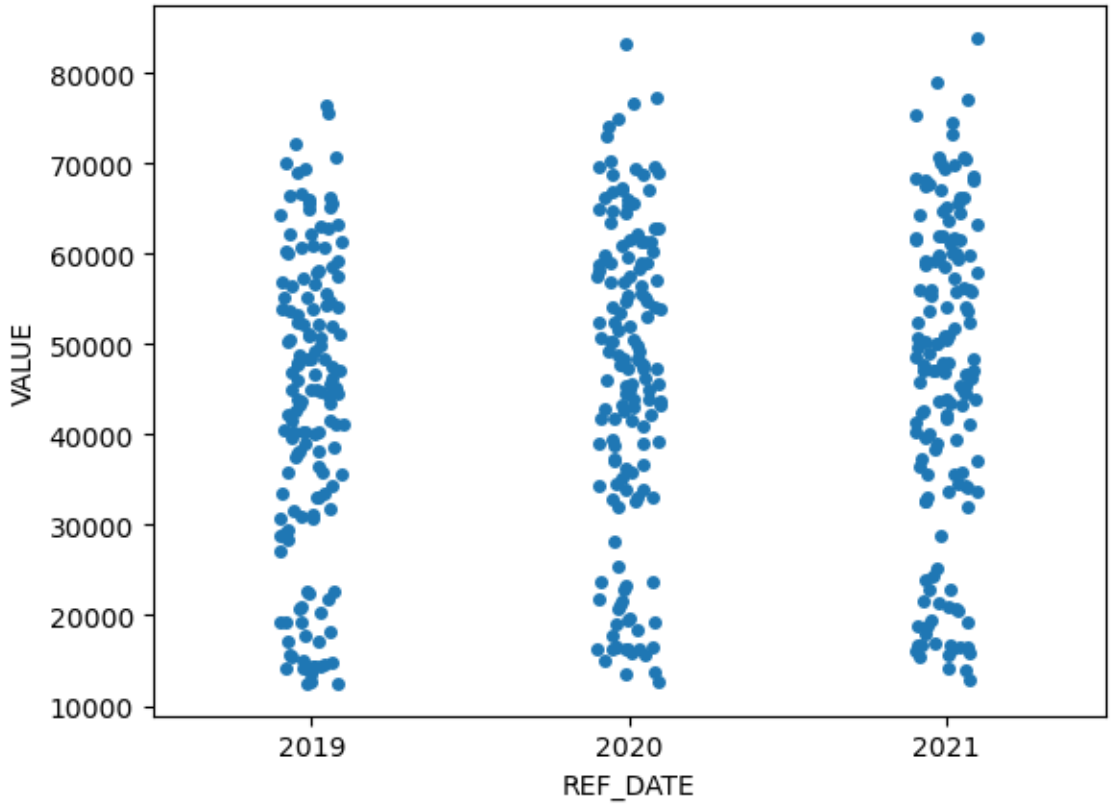
final_testing_df_output_df_AvgAnnWages_ByImmigrant.csv



Done using Stripplot
[0, 1] = ['Immigrant employees' 'Non-immigrant employees']

Result for testing set for 'Average annual wages and salaries' by yearly

`final_testing_df_output_df_AvgAnnWages_ByAge.csv`



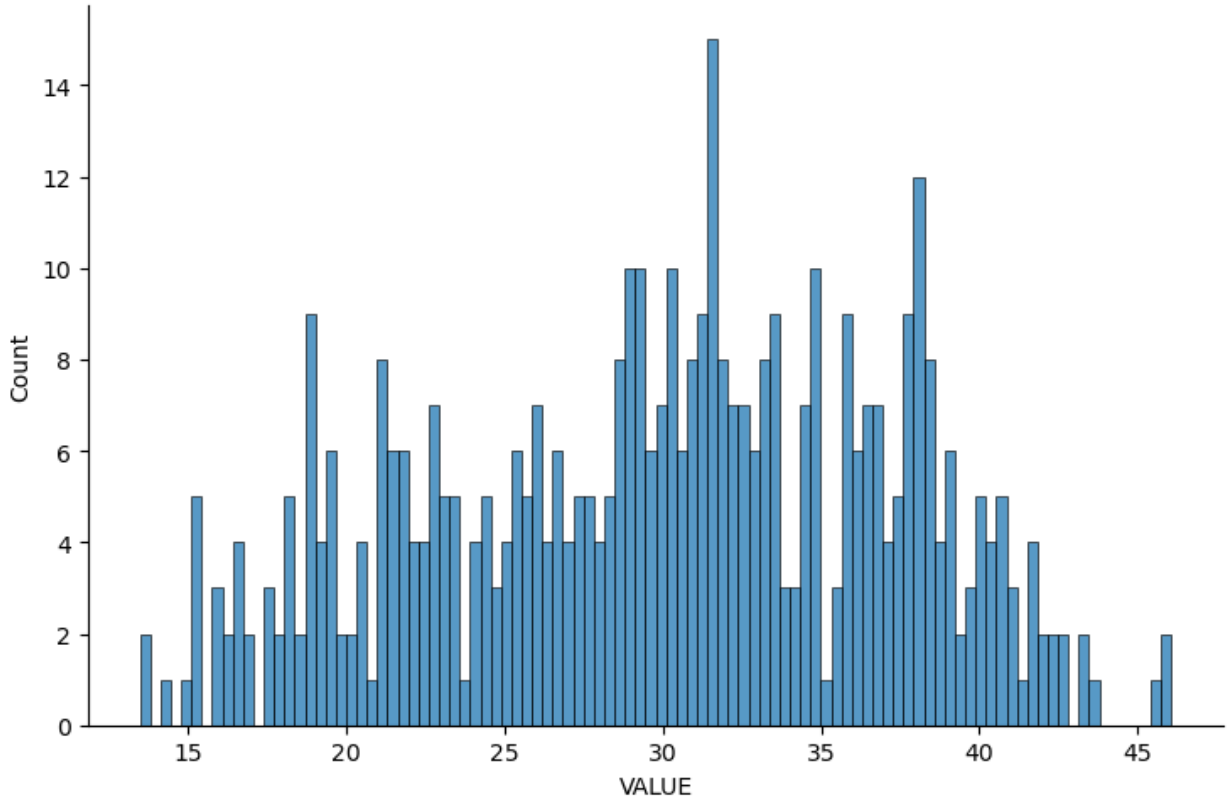
`Done using Stripplot`

Output #16a

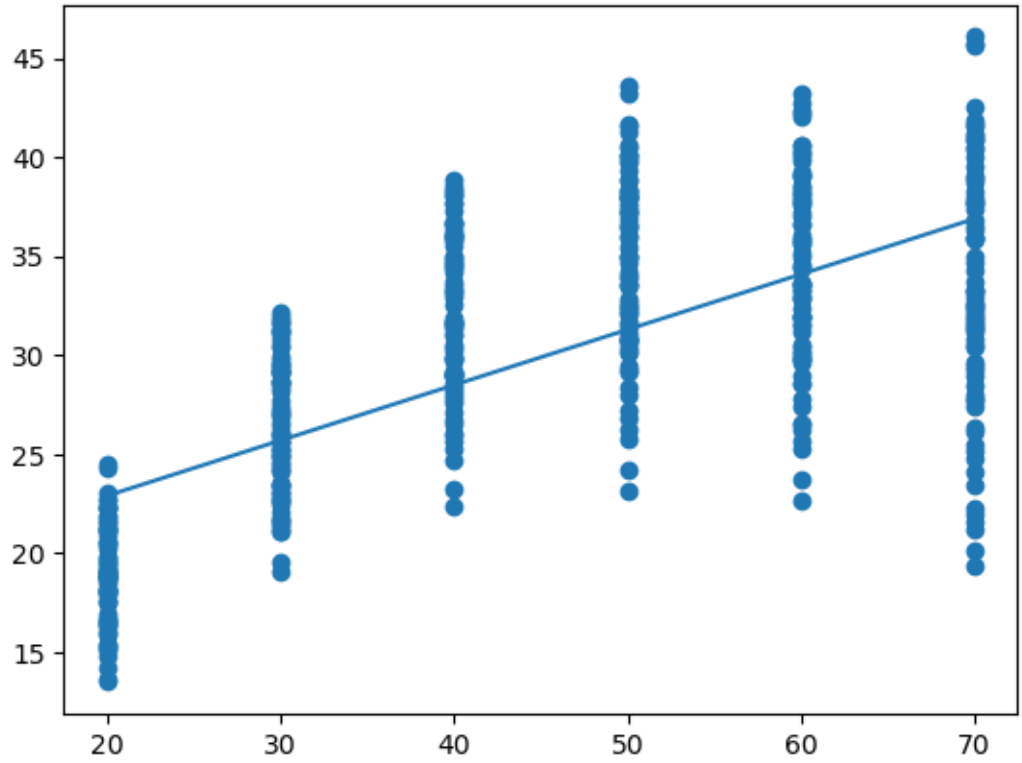
Result for testing set for 'Average hourly wage' by Age group

```
final_testing_df_output_df_AvgHrsWages_ByAge.csv
                           sum       mean   amin  median   amax  size
Characteristics
15 to 24 years         1418.04  18.907200  13.49   18.89  24.48    75
25 to 34 years         1974.53  26.327067  19.02   26.03  32.18    75
35 to 44 years         2405.70  32.076000  22.35   31.65  38.89    75
45 to 54 years         2592.01  34.560133  23.17   34.96  43.62    75
55 to 64 years         2565.26  34.203467  22.66   34.40  43.17    75
65 years old and over  2494.07  33.254267  19.36   32.92  46.07    75
Overall,
Sum :  13449.61
Mean :  29.888022222222222
Min/median/max : 13.49 / 30.490000000000002 / 46.07
Standard Deviation :  7.202130786522336
Skewnewss :  -0.2005972990040442
Total size :  450
```

final_testing_df_output_df_AvgHrsWages_ByAge.csv



final_testing_df_output_df_AvgHrsWages_ByAge.csv



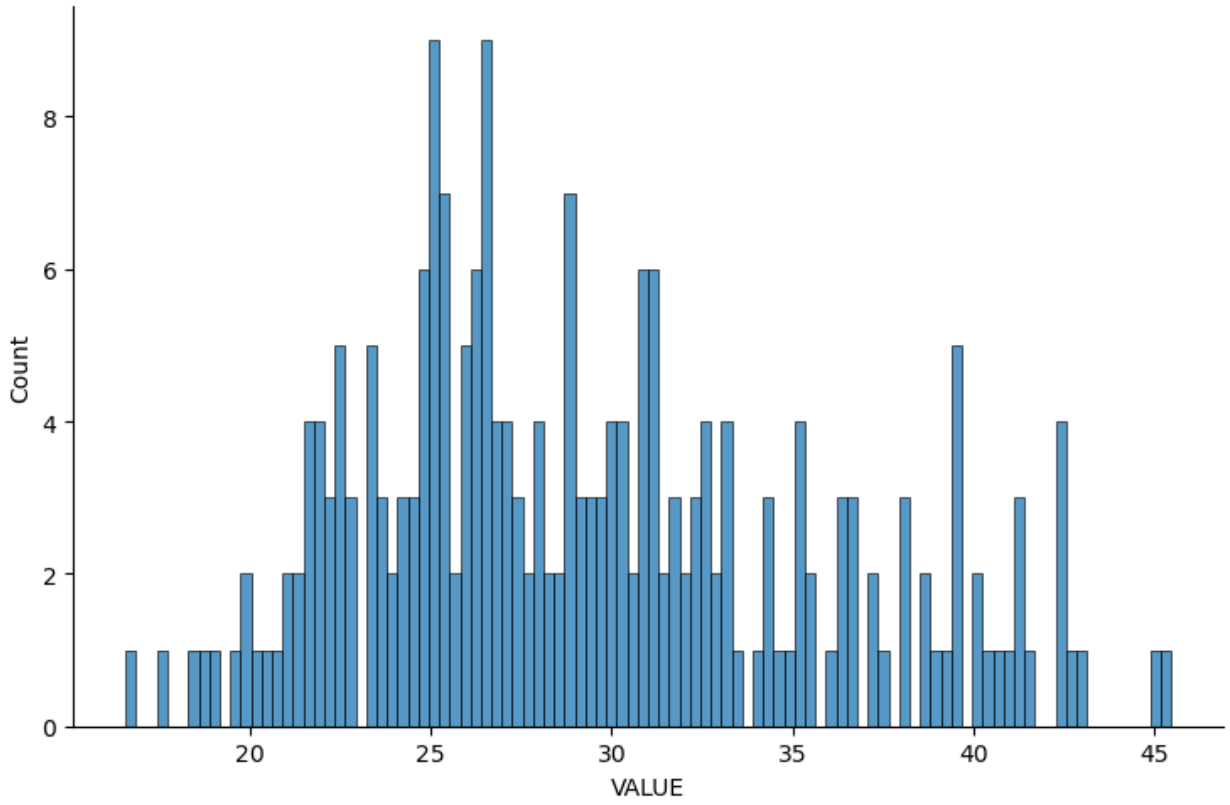
Done by Linear Regression

Output #16b

Result for testing set for 'Average hourly wage' by Education level

```
final_testing_df_output_df_AvgHrsWages_ByEducation.csv
                                sum       mean   amin  median   amax  size
Characteristics
High school diploma and less  1813.29  24.177200  16.56   24.81  29.37    75
Trade certificate             2123.45  28.312667  19.98   28.47  39.55    75
University degree and higher  2658.22  35.442933  24.19   35.58  45.47    75
Overall,
Sum :  6594.959999999999
Mean :  29.31093333333333
Min/median/max : 16.56 / 28.34 / 45.47
Standard Deviation :  6.1840297915050675
Skewnewss :  0.5213937099086743
Total size :  225
```

final_testing_df_output_df_AvgHrsWages_ByEducation.csv

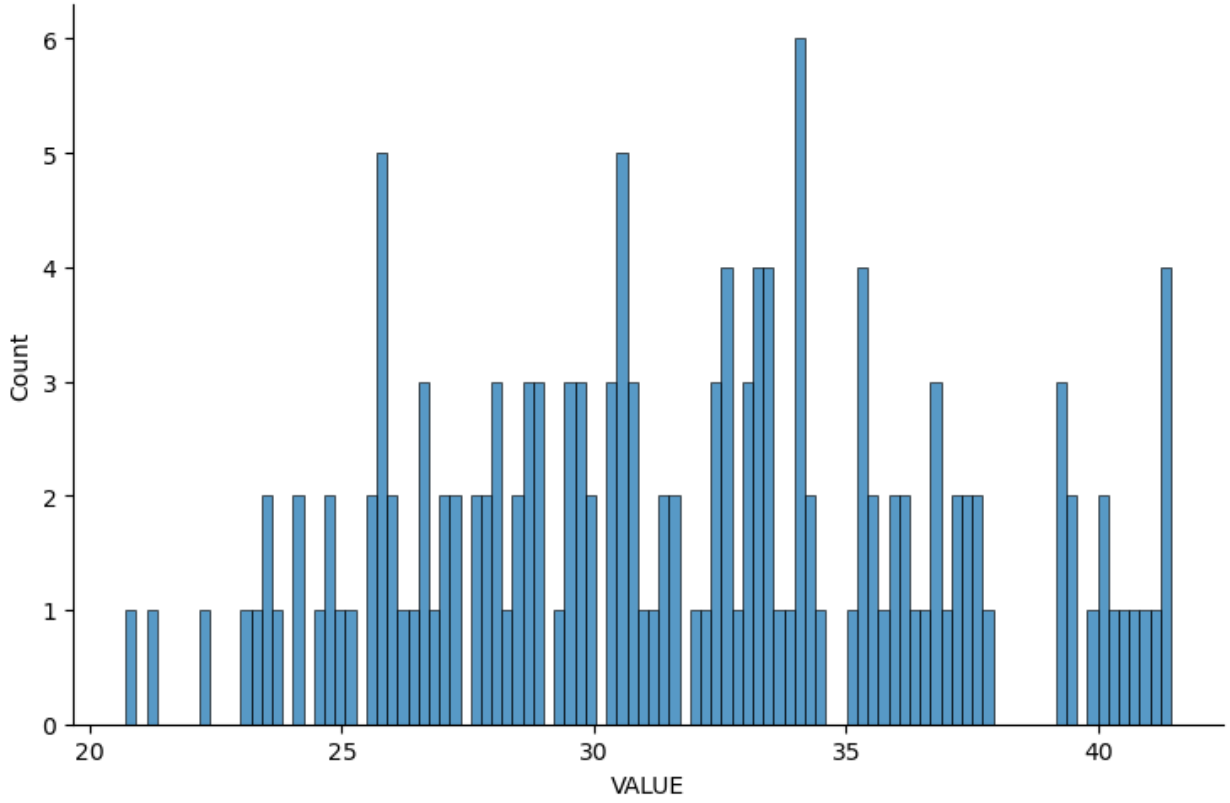


Output #16c

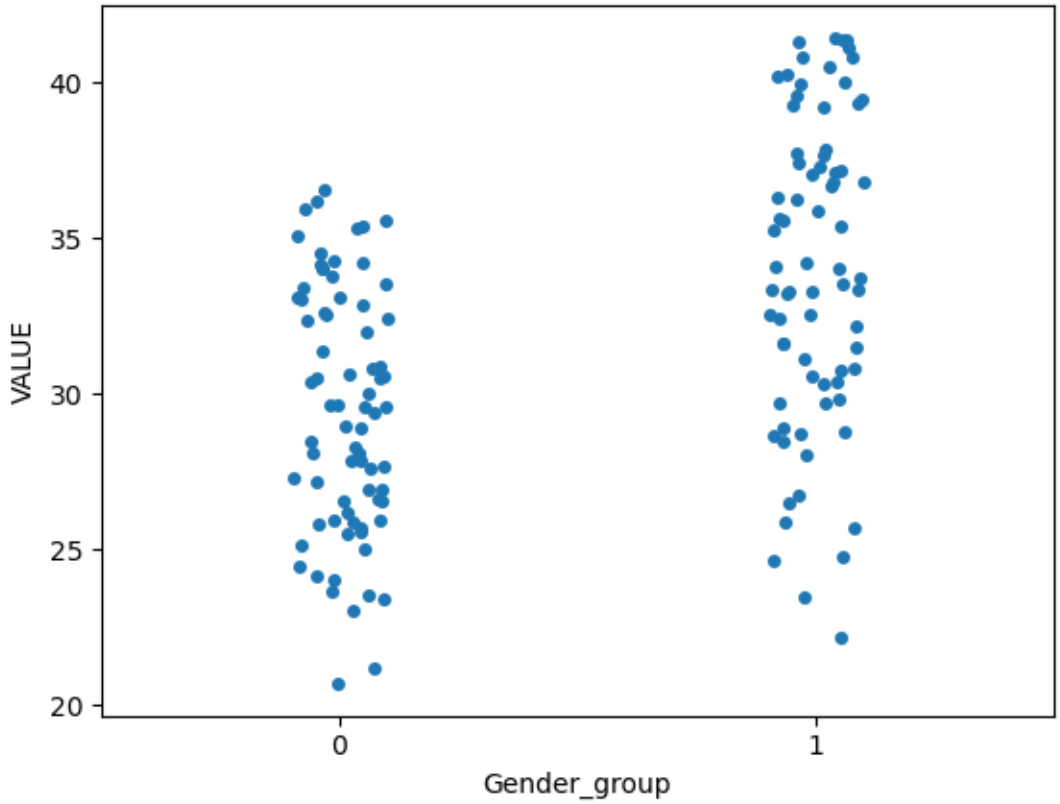
Result for testing set for 'Average hourly wage' by Gender group

```
final_testing_df_output_df_AvgHrsWages_ByGender.csv
                      sum       mean   amin  median   amax  size
Characteristics
Female employees  2207.94  29.439200  20.71   29.58  36.55    75
Male employees    2550.74  34.009867  22.20   34.03  41.44    75
Overall,
Sum :  4758.68
Mean :  31.724533333333337
Min/median/max : 20.71 / 31.595 / 41.44
Standard Deviation :  5.011556589413003
Skewnewss :  0.07868782098838015
Total size :  150
```

final_testing_df_output_df_AvgHrsWages_ByGender.csv



final_testing_df_output_df_AvgHrsWages_ByGender.csv



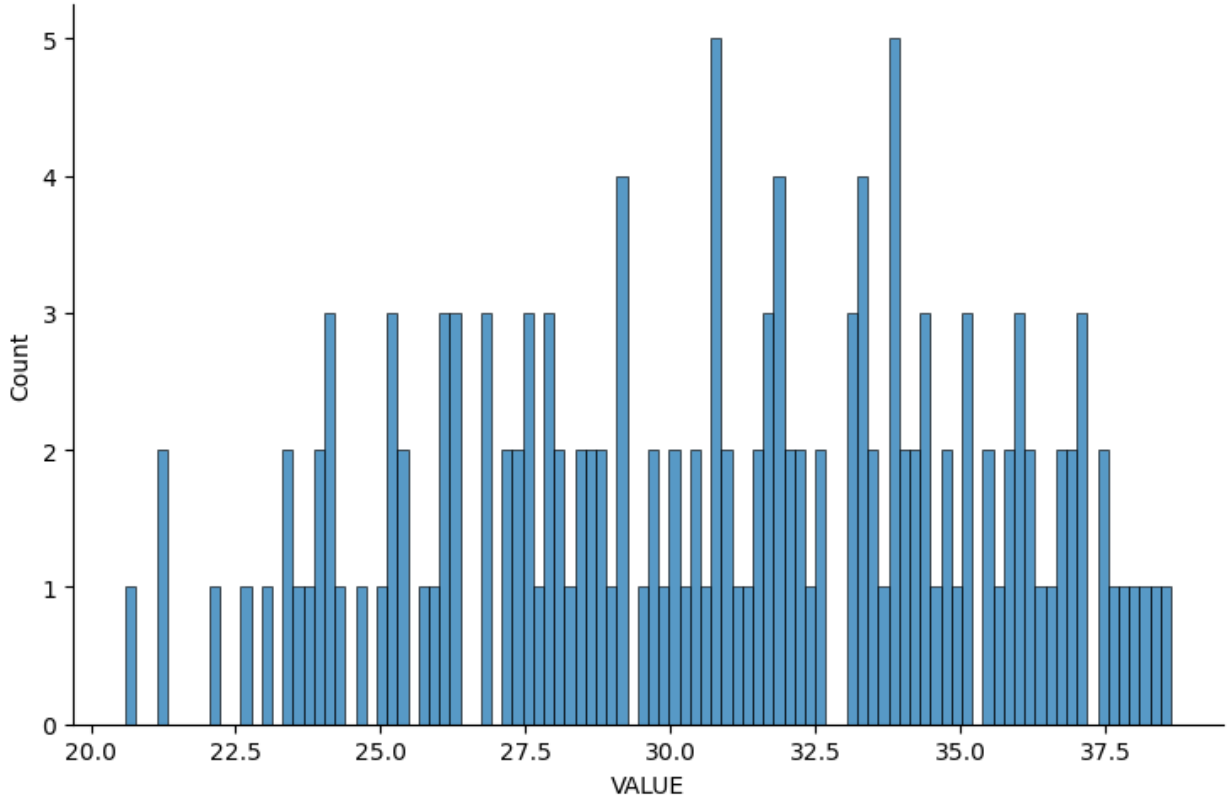
Done using Stripplot
[1, 0] = ['Male employees' 'Female employees']

Output #16c

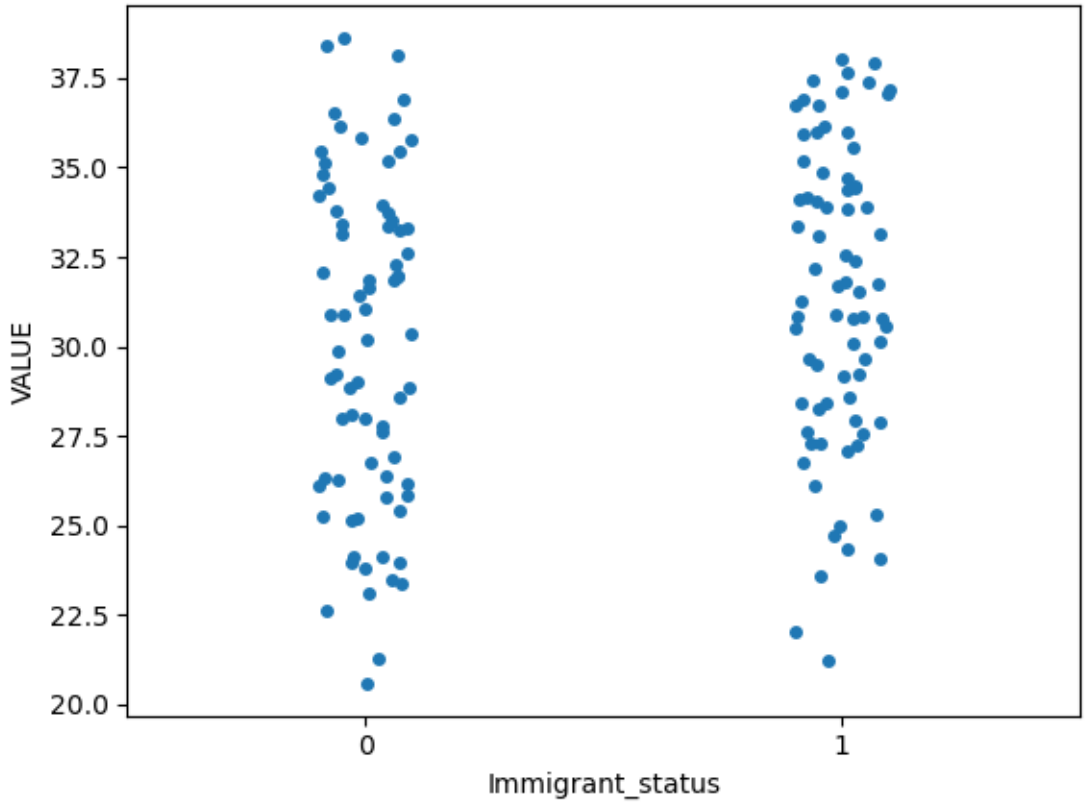
Result for testing set for 'Average hourly wage' by Immigrant status

```
final_testing_df_output_df_AvgHrsWages_ByImmigrant.csv
                            sum       mean   amin  median   amax  size
Characteristics
Immigrant employees      2254.88  30.065067  20.60   30.35  38.63    75
Non-immigrant employees  2356.06  31.414133  21.23   31.53  38.02    75
Overall,
Sum :  4610.9400000000005
Mean :  30.739600000000003
Min/median/max : 20.6 / 30.905 / 38.63
Standard Deviation :  4.419609165224154
Skewnewss :  -0.21241877283543747
Total size :  150
```

final_testing_df_output_df_AvgHrsWages_ByImmigrant.csv



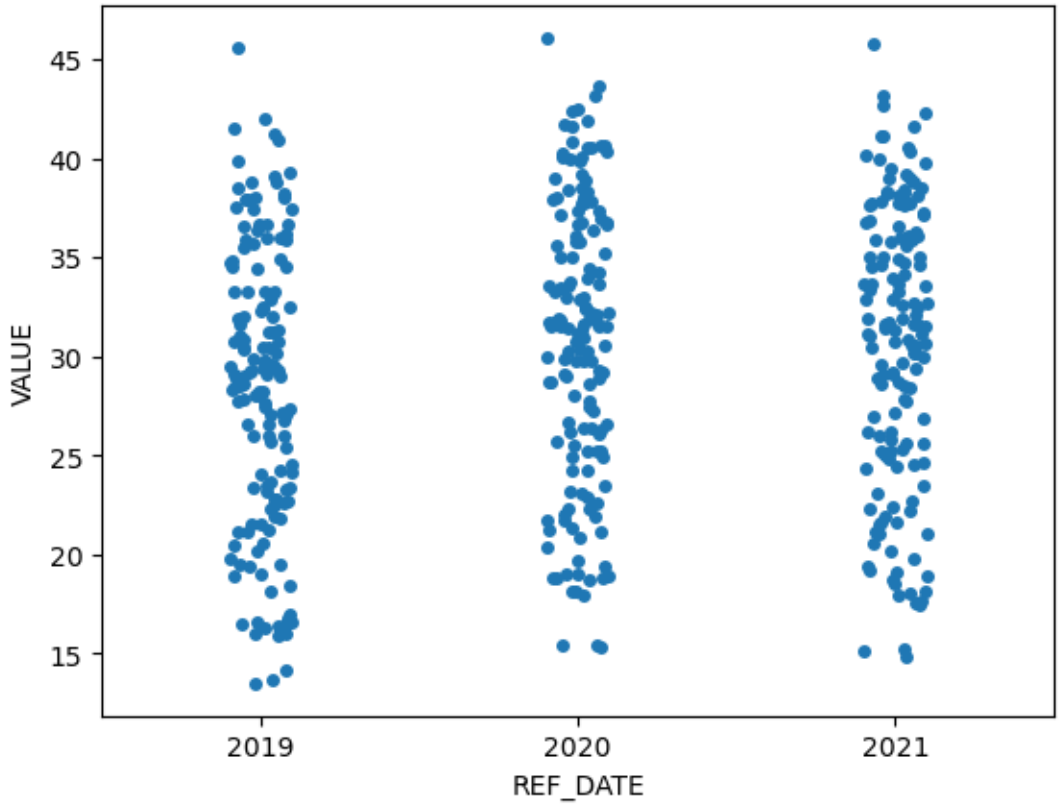
final_testing_df_output_df_AvgHrsWages_ByImmigrant.csv



Done using Stripplot
[0, 1] = ['Immigrant employees' 'Non-immigrant employees']

Result for testing set for 'Average hourly wage' by yearly

`final_testing_df_output_df_AvgHrsWages_ByAge.csv`



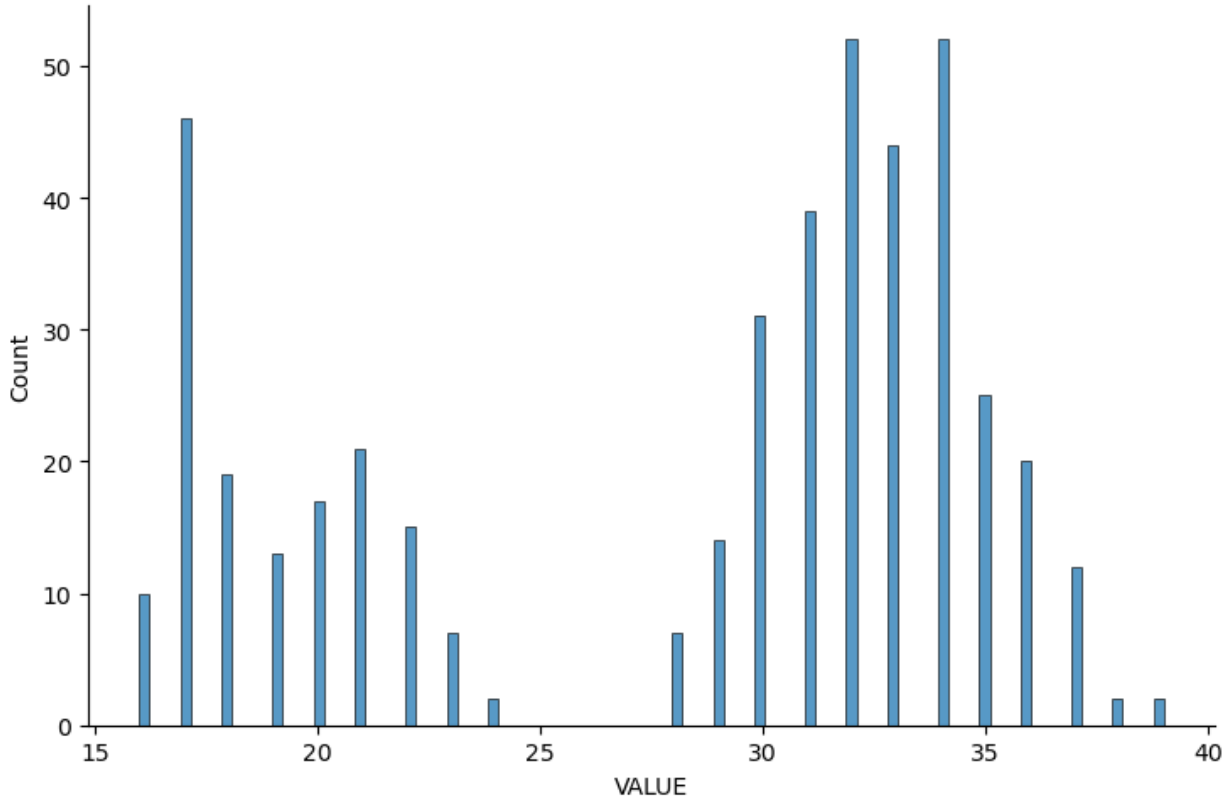
`Done using Stripplot`

Output #17a

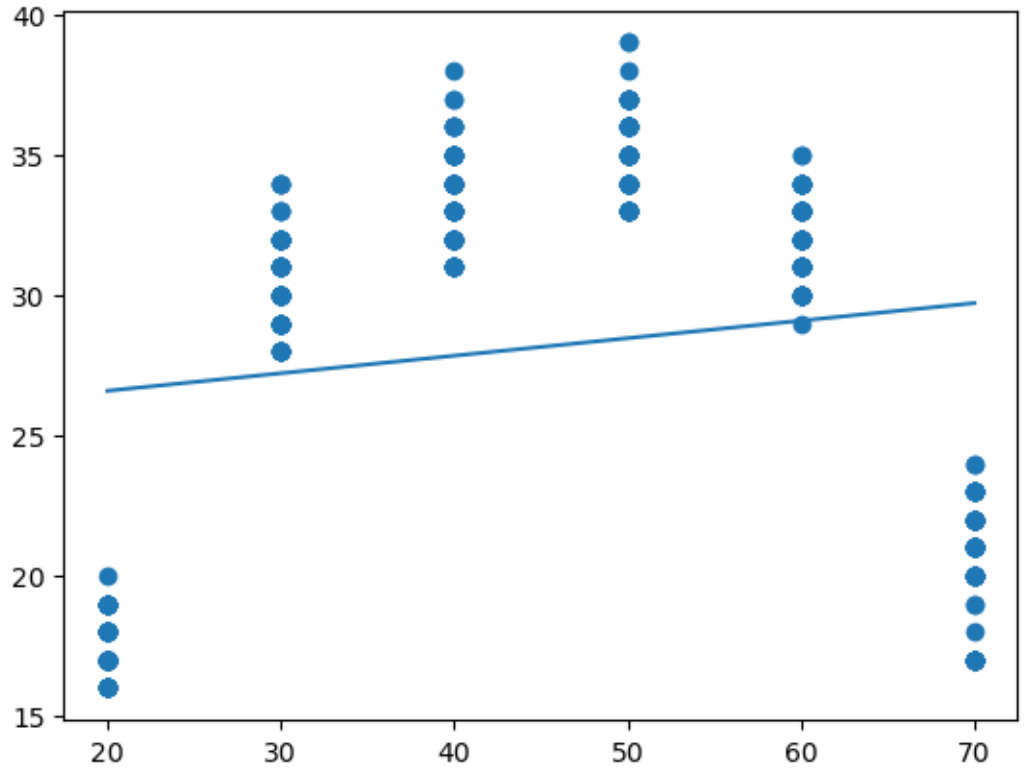
Result for testing set for 'Average weekly hours worked' by Age group

```
final_testing_df_output_df_AvgWeekHrsWrked_ByAge.csv
                          sum       mean  amin  median  amax  size
Characteristics
15 to 24 years         1308.0  17.440000  16.0    17.0  20.0    75
25 to 34 years         2290.0  30.533333  28.0    30.0  34.0    75
35 to 44 years         2499.0  33.320000  31.0    33.0  38.0    75
45 to 54 years         2626.0  35.013333  33.0    35.0  39.0    75
55 to 64 years         2403.0  32.040000  29.0    32.0  35.0    75
65 years old and over  1543.0  20.573333  17.0    21.0  24.0    75
Overall,
Sum :  12669.0
Mean :  28.153333333333332
Min/median/max : 16.0 / 31.0 / 39.0
Standard Deviation :  6.8352712699155695
Skewnewss :  -0.5809323008907273
Total size :  450
```

final_testing_df_output_df_AvgWeekHrsWrked_ByAge.csv



final_testing_df_output_df_AvgWeekHrsWrked_ByAge.csv



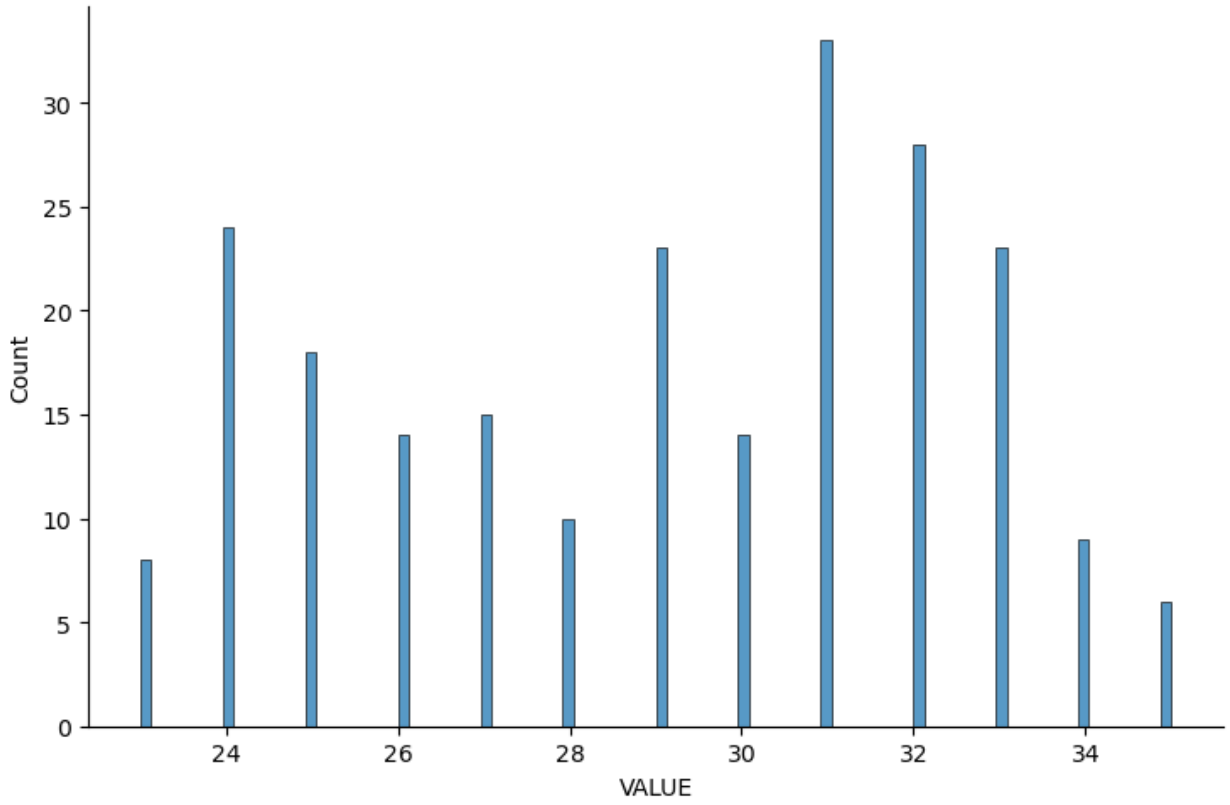
Done by Linear Regression

Output #17b

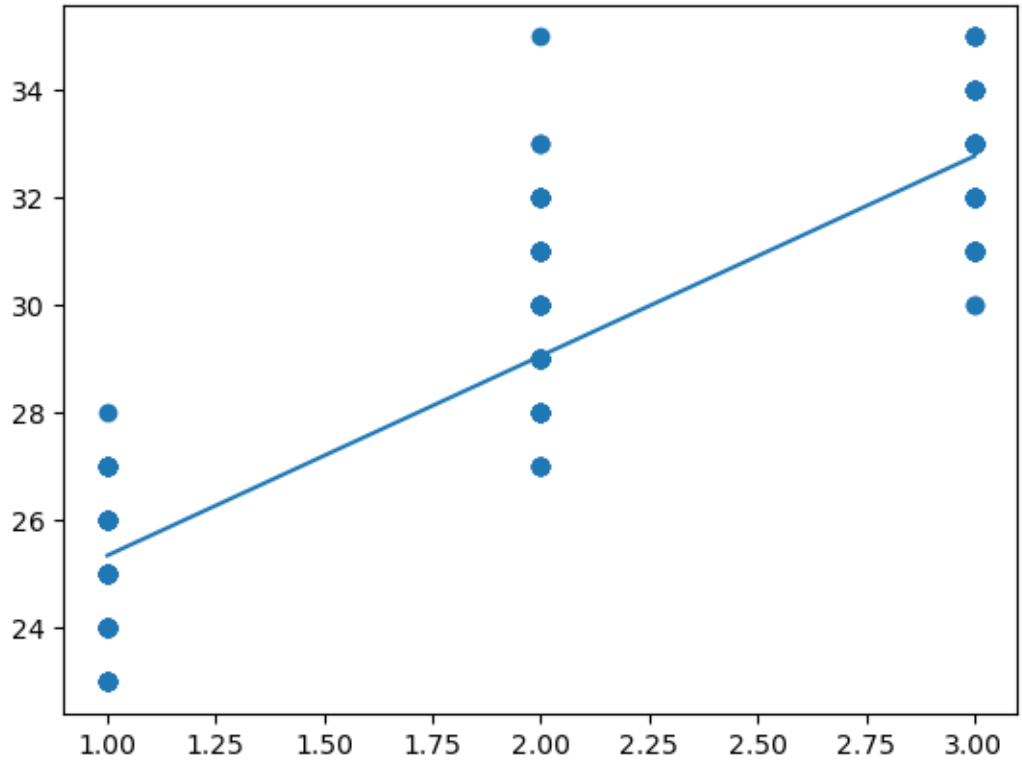
Result for testing set for 'Average weekly hours worked' by Education level

```
final_testing_df_output_df_AvgWeekHrsWrked_ByEducation.csv
                                sum       mean  amin  median  amax  size
Characteristics
High school diploma and less  1872.0  24.960000  23.0    25.0  28.0    75
Trade certificate             2239.0  29.853333  27.0    30.0  35.0    75
University degree and higher  2429.0  32.386667  30.0    32.0  35.0    75
Overall,
Sum :  6540.0
Mean :  29.066666666666666
Min/median/max : 23.0 / 30.0 / 35.0
Standard Deviation :  3.384933939942962
Skewnewss :  -0.22803984300945943
Total size :  225
```

final_testing_df_output_df_AvgWeekHrsWrked_ByEducation.csv



final_testing_df_output_df_AvgWeekHrsWrked_ByEducation.csv



Done by Linear Regression
Higher the number, higher the education

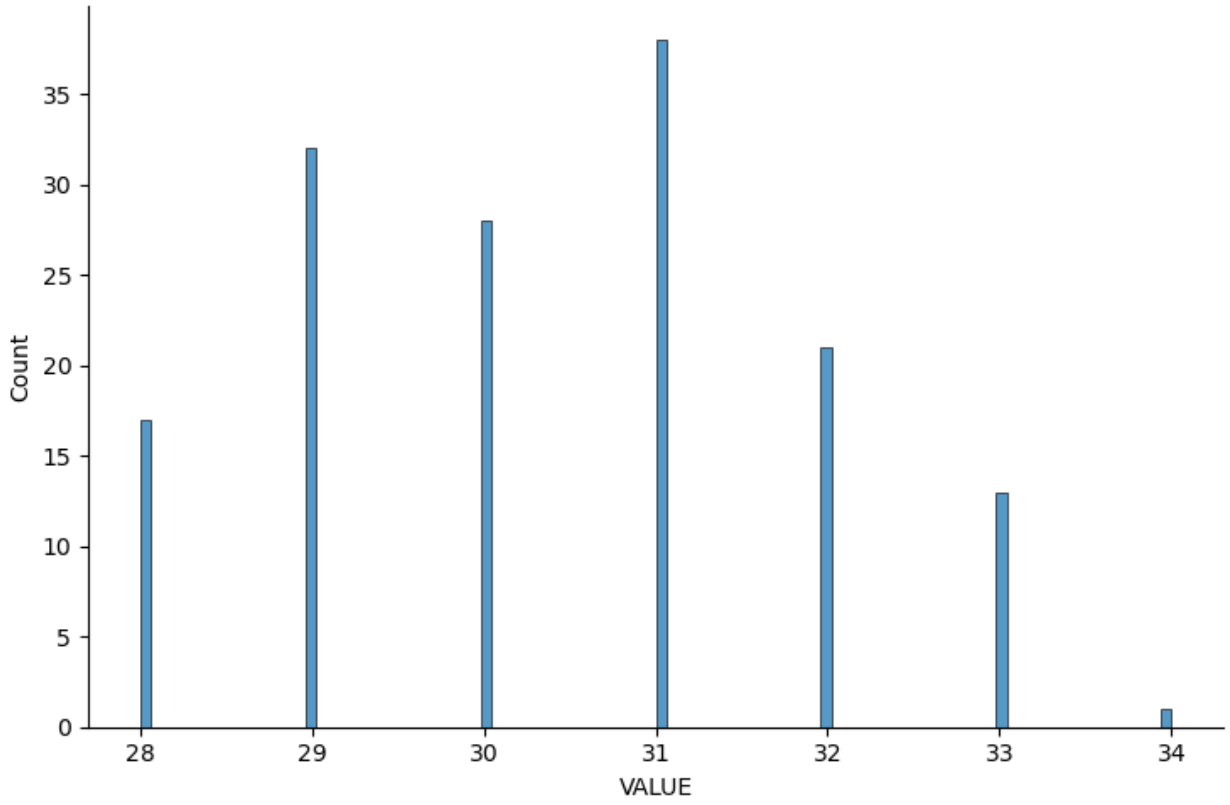
Output #17c

Result for testing set for 'Average weekly hours worked' by Gender group

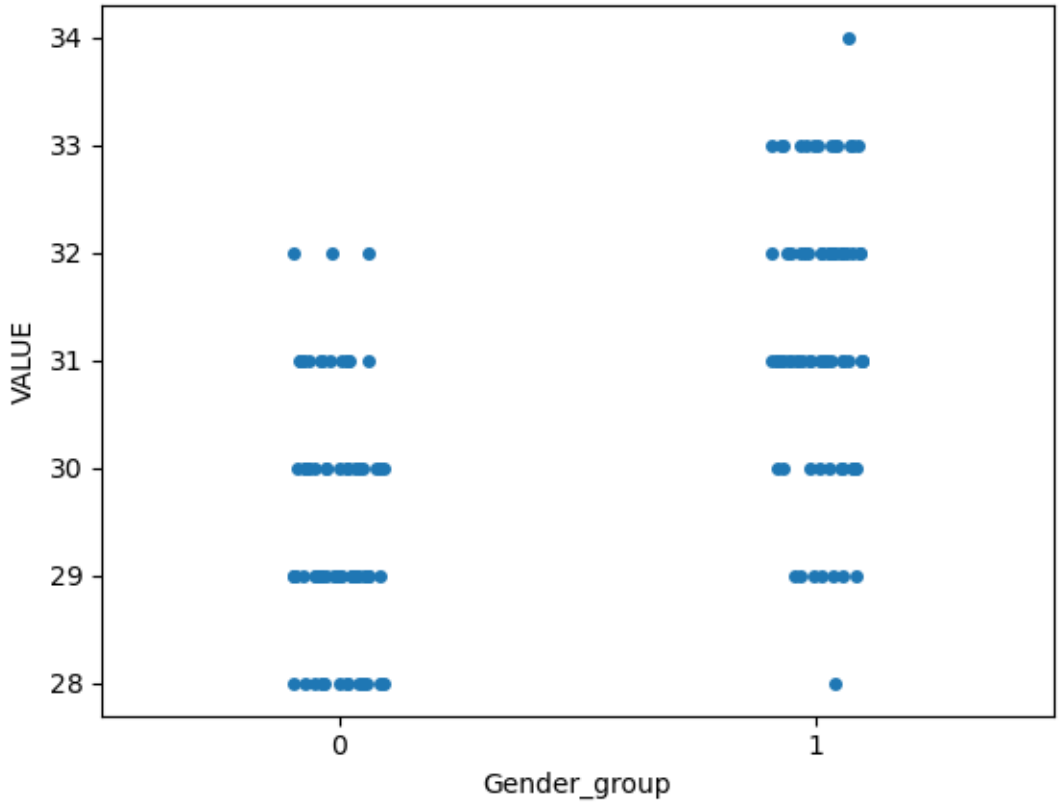
`final_testing_df_output_df_AvgWeekHrsWrked_ByGender.csv`

```
                     sum       mean  amin  median  amax  size
Characteristics
Female employees  2213.0  29.506667  28.0    29.0  32.0    75
Male employees    2344.0  31.253333  28.0    31.0  34.0    75
Overall,
Sum :  4557.0
Mean :  30.38
Min/median/max : 28.0 / 30.0 / 34.0
Standard Deviation :  1.4907268920451757
Skewnewss :  0.1316464344774555
Total size :  150
```

`final_testing_df_output_df_AvgWeekHrsWrked_ByGender.csv`



final_testing_df_output_df_AvgWeekHrsWrked_ByGender.csv



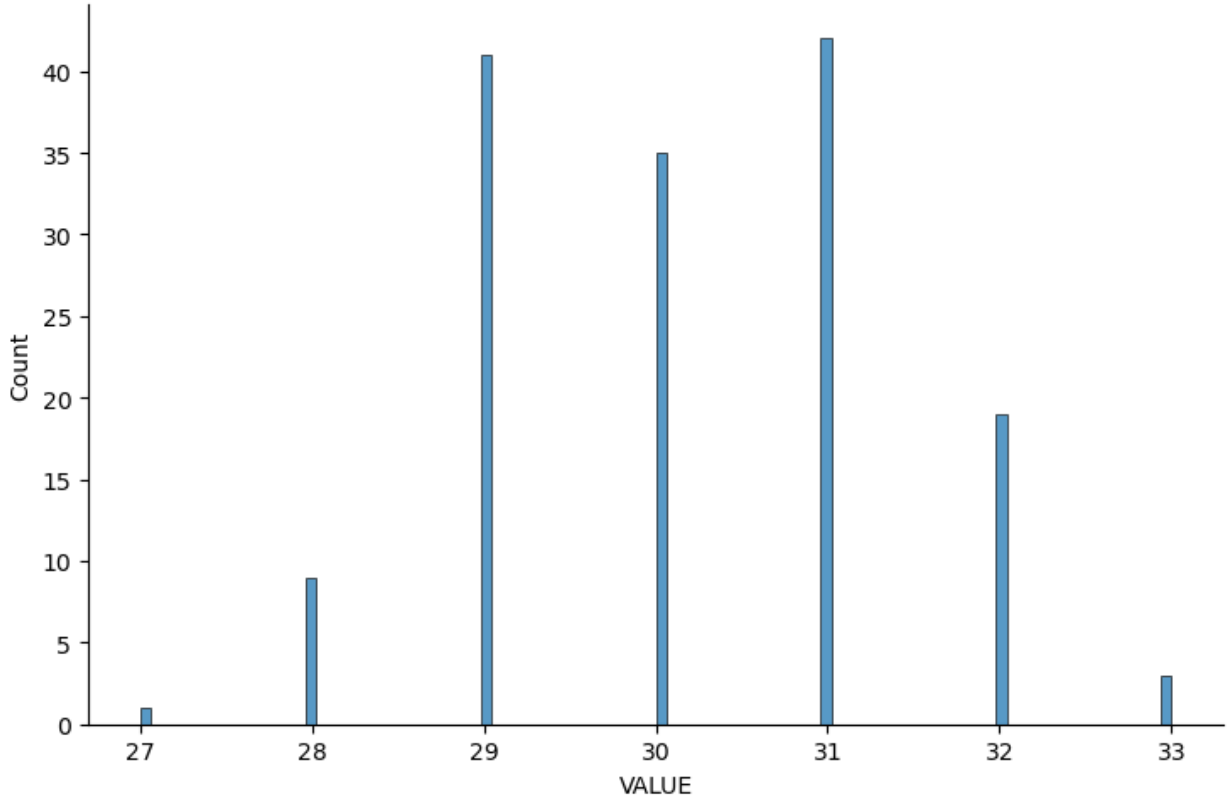
Done using Stripplot
[1, 0] = ['Male employees' 'Female employees']

Output #17c

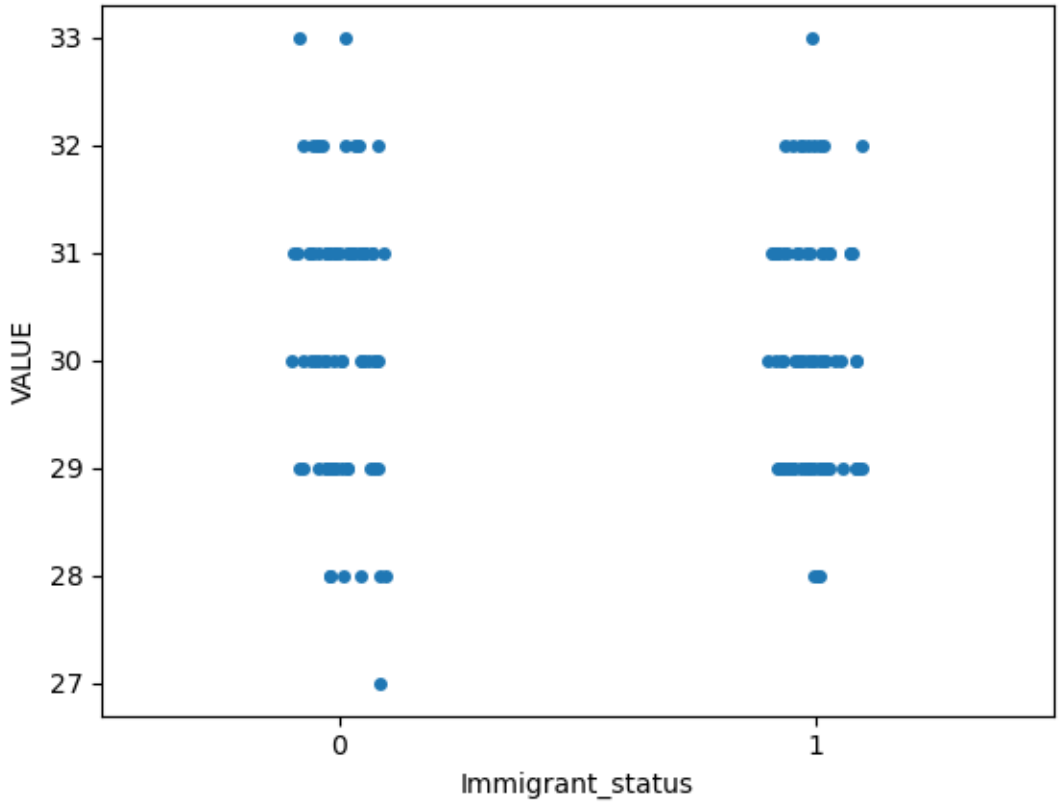
Result for testing set for 'Average weekly hours worked' by Immigrant status

```
final_testing_df_output_df_AvgWeekHrsWrked_ByImmigrant.csv
                           sum       mean  amin  median  amax  size
Characteristics
Immigrant employees      2270.0  30.266667  27.0    30.0  33.0    75
Non-immigrant employees  2257.0  30.093333  28.0    30.0  33.0    75
Overall,
Sum :  4527.0
Mean :  30.18
Min/median/max : 27.0 / 30.0 / 33.0
Standard Deviation :  1.2278436382536666
Skewnewss :  0.0436903112714573
Total size :  150
```

final_testing_df_output_df_AvgWeekHrsWrked_ByImmigrant.csv



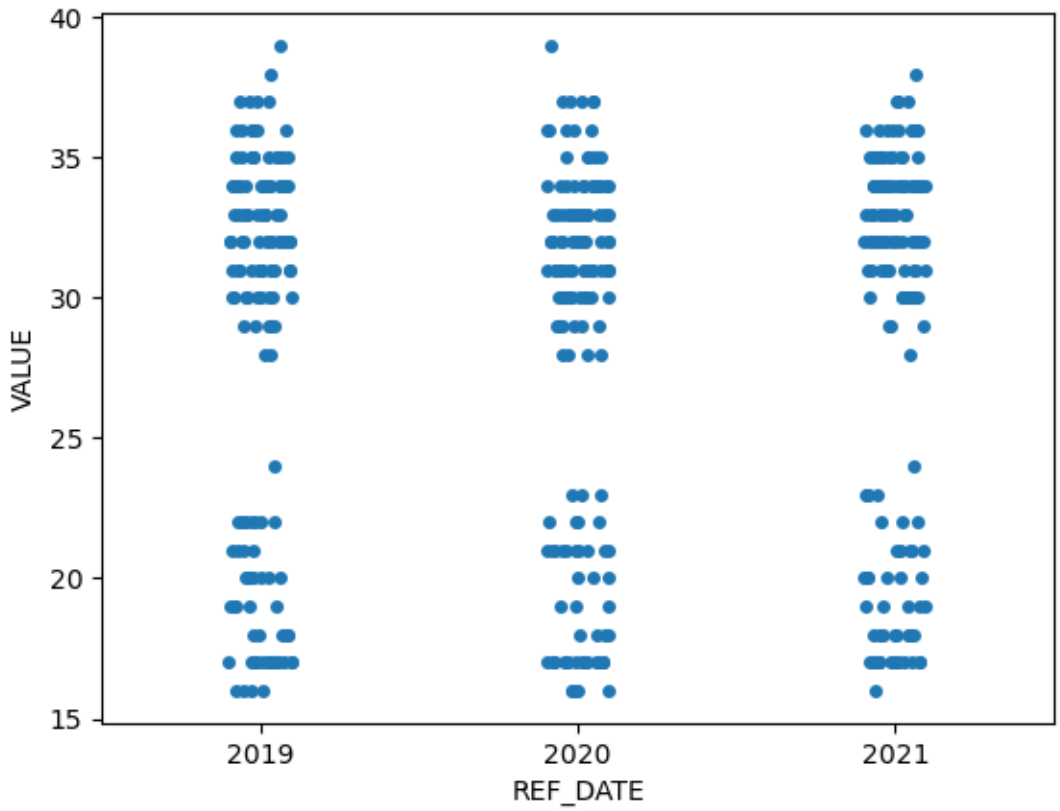
final_testing_df_output_df_AvgWeekHrsWrked_ByImmigrant.csv



Done using Stripplot
[0, 1] = ['Immigrant employees' 'Non-immigrant employees']

Result for testing set for 'Average weekly hours worked' by yearly

`final_testing_df_output_df_AvgWeekHrsWrked_ByAge.csv`



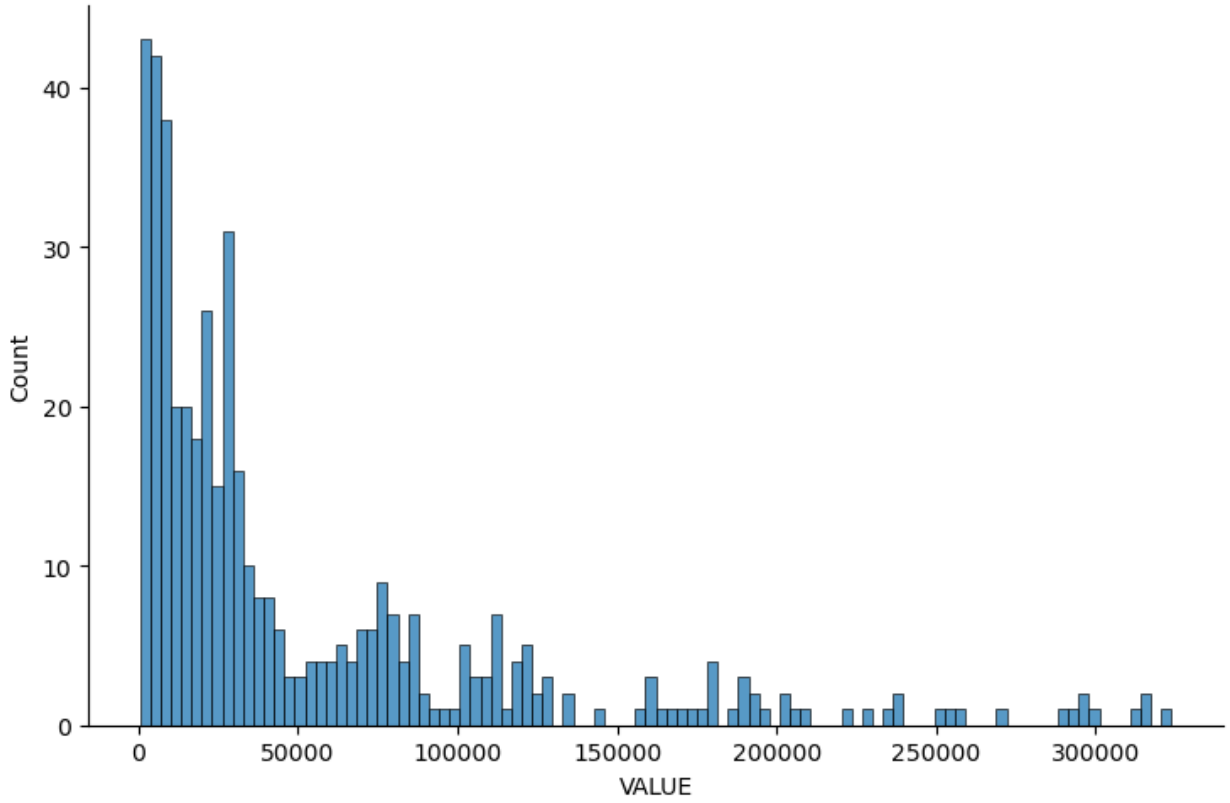
`Done using Stripplot`

Output #18a

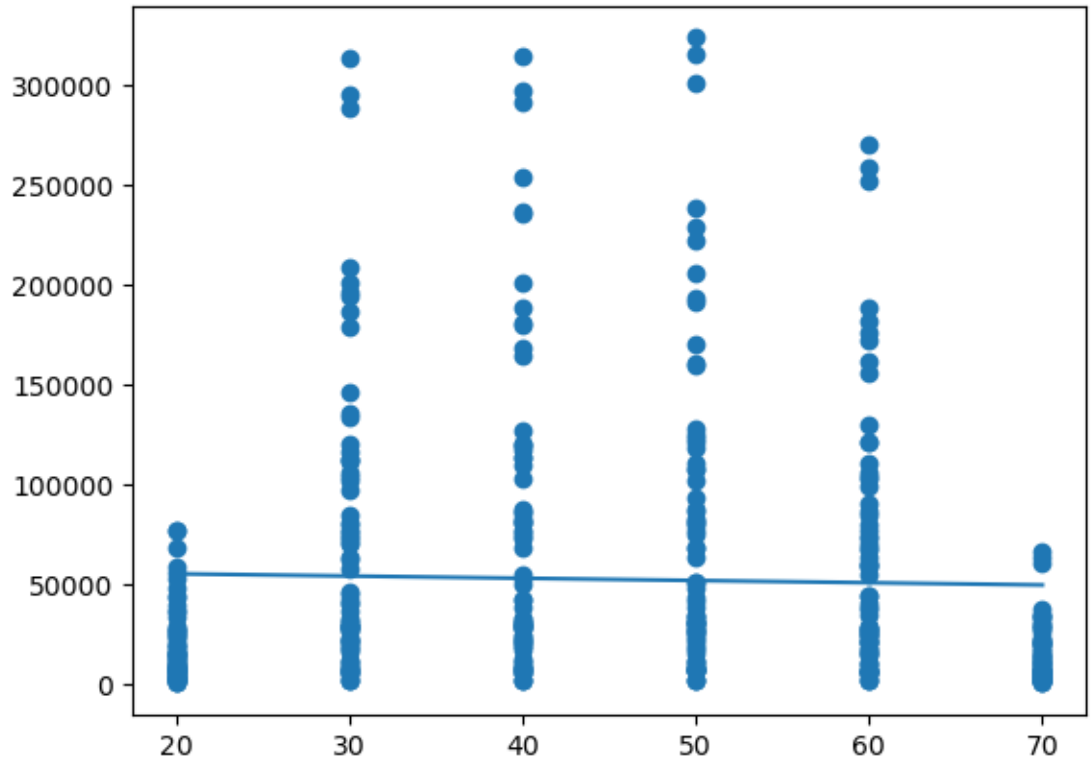
Result for testing set for 'Hours Worked' by Age group

```
final_testing_df_output_df_Hrs_Wrked_ByAge.csv
                             sum          mean    amin   median      amax  \
Characteristics
15 to 24 years         1338316.0  17844.213333   497.0  11421.0   77253.0
25 to 34 years         5209479.0  69459.720000  1434.0  40393.0  313785.0
35 to 44 years         5636717.0  75156.226667  1413.0  41945.0  314318.0
45 to 54 years         5632390.0  75098.533333  1635.0  41419.0  323859.0
55 to 64 years         4632898.0  61771.973333  1407.0  37091.0  270787.0
65 years old and over  1095950.0  14612.666667   361.0  10249.0   65892.0

                       size
Characteristics
15 to 24 years           75
25 to 34 years           75
35 to 44 years           75
45 to 54 years           75
55 to 64 years           75
65 years old and over    75
Overall,
Sum :  23545750.0
Mean :  52323.88888888889
Min/median/max : 361.0 / 26432.5 / 323859.0
Standard Deviation :  65504.80018241995
Skewnewss :  2.0444664668685038
Total size :  450
```

final_testing_df_output_df_Hrs_Wrked_ByAge.csv



final_testing_df_output_df_Hrs_Wrked_ByAge.csv



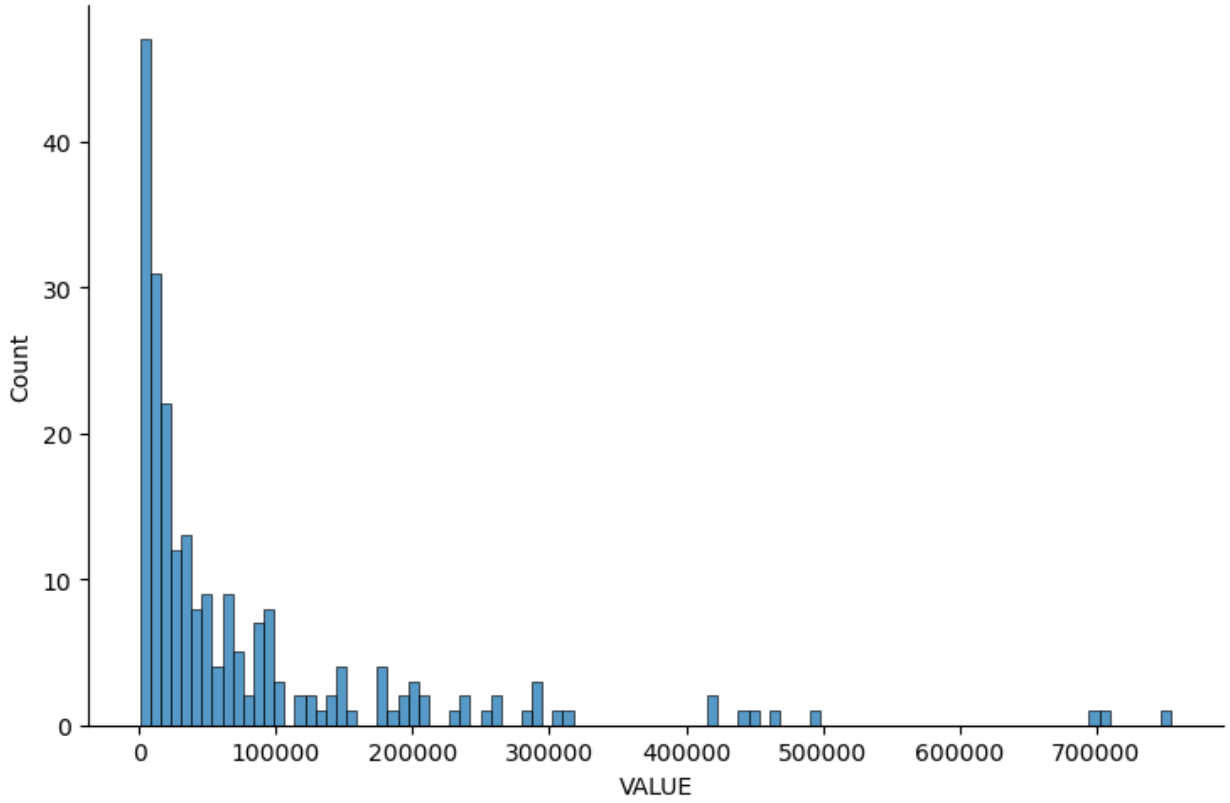
Done by Linear Regression

Output #18b

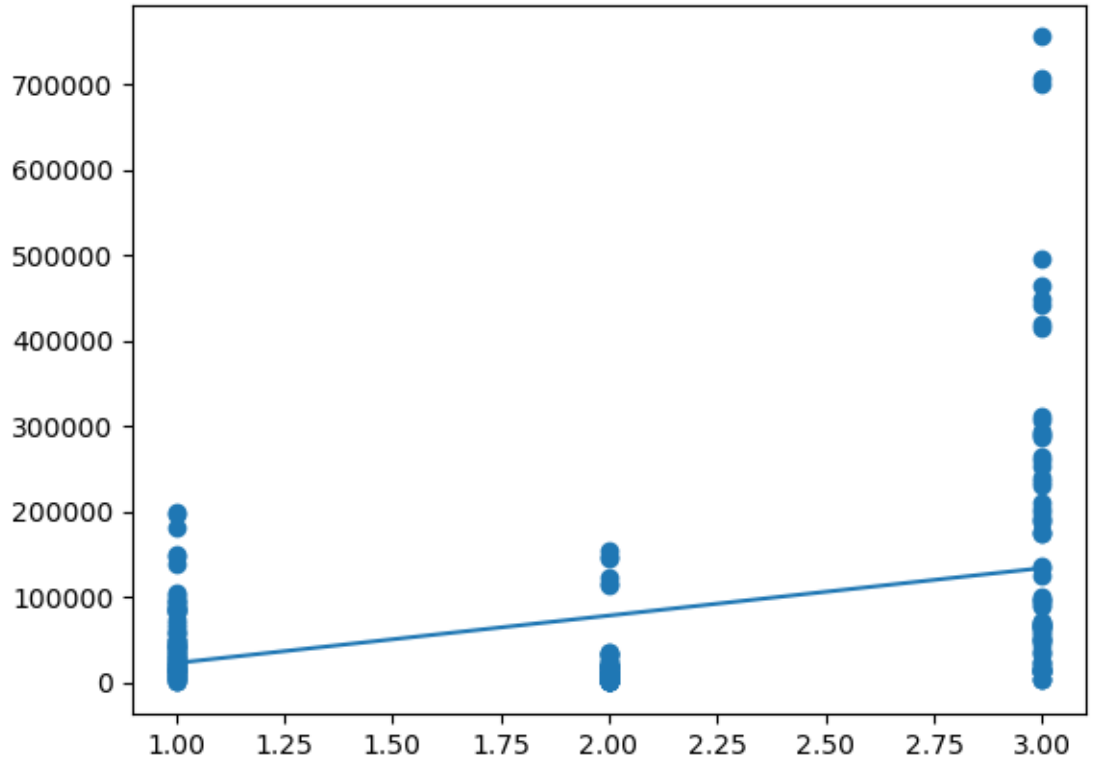
Result for testing set for 'Hours Worked' by Education level

```
final_testing_df_output_df_Hrs_Wrked_ByEducation.csv
                                      sum           mean    amin   median  \
Characteristics
High school diploma and less   3839870.0   51198.266667  1717.0  37663.0
Trade certificate              1548281.0   20643.746667   349.0   8852.0
University degree and higher  12180311.0  162404.146667  3106.0  91916.0

                                    amax  size
Characteristics
High school diploma and less  199483.0    75
Trade certificate             153953.0    75
University degree and higher  755243.0    75
Overall,
Sum :  17568462.0
Mean :  78082.05333333333
Min/median/max : 349.0 / 31865.0 / 755243.0
Standard Deviation :  120621.30059284004
Skewnewss :  3.0205628183947755
Total size :  225
```

final_testing_df_output_df_Hrs_Wrked_ByEducation.csv



final_testing_df_output_df_Hrs_Wrked_ByEducation.csv



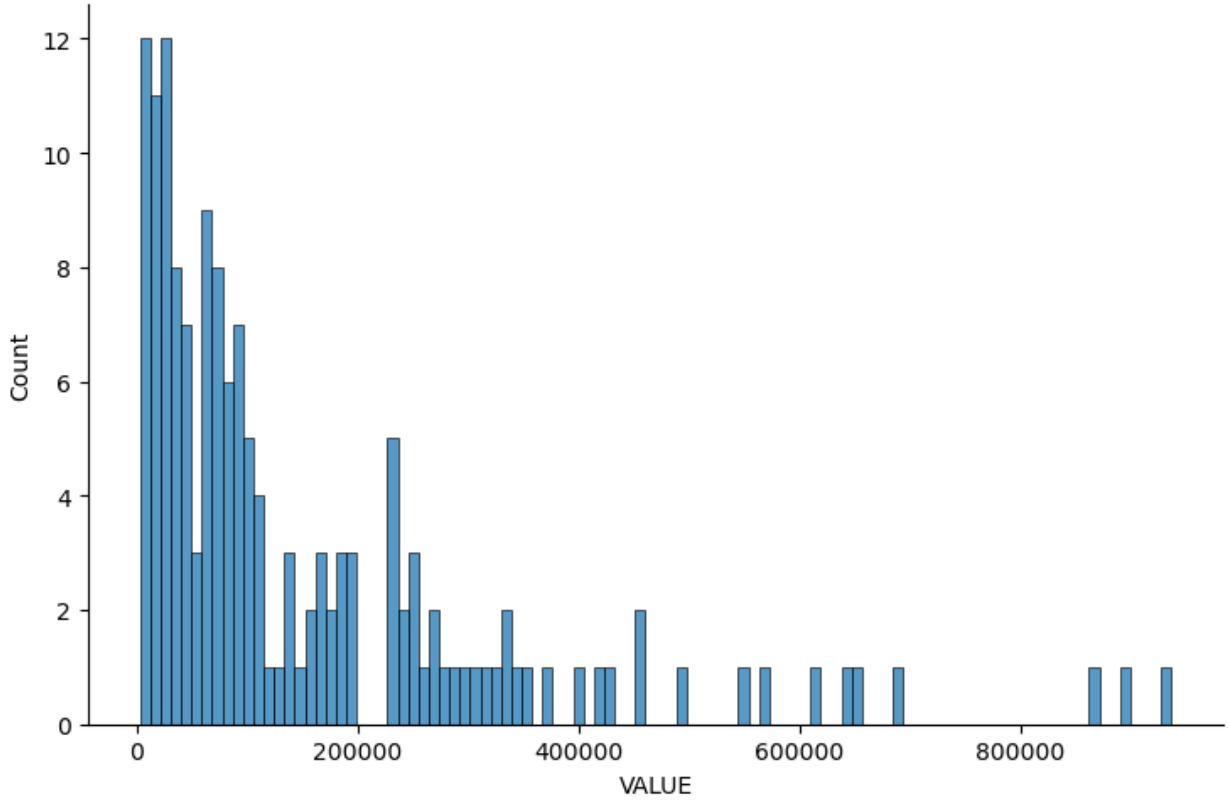
Done by Linear Regression
Higher the number, higher the education

Output #18c

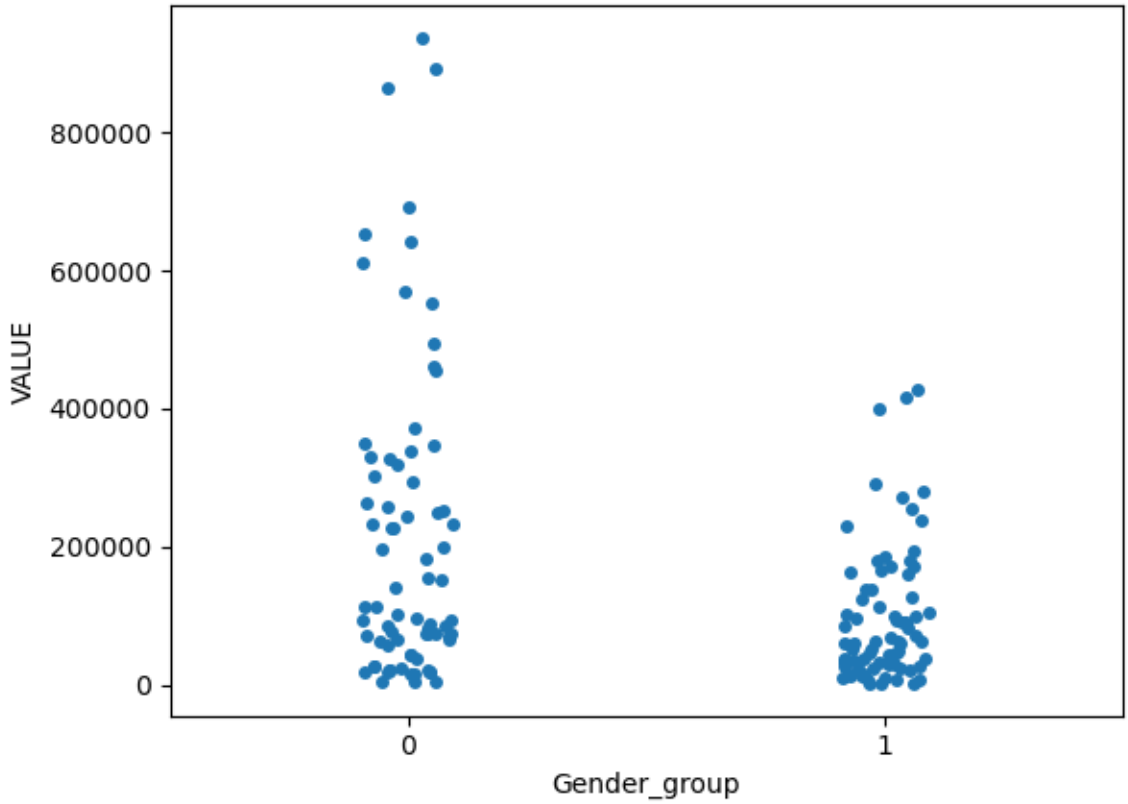
Result for testing set for 'Hours Worked' by Gender group

```
final_testing_df_output_df_Hrs_Wrked_ByGender.csv
                         sum           mean    amin    median      amax  size
Characteristics
Female employees  16144628.0  215261.706667  3830.0  114631.0  937067.0    75
Male employees     7401122.0   98681.626667  2918.0   64568.0  428111.0    75
Overall,
Sum :  23545750.0
Mean :  156971.66666666666
Min/median/max : 2918.0 / 86240.0 / 937067.0
Standard Deviation :  182769.39372530495
Skewnewss :  2.0993685578631798
Total size :  150
```

final_testing_df_output_df_Hrs_Wrked_ByGender.csv



final_testing_df_output_df_Hrs_Wrked_ByGender.csv



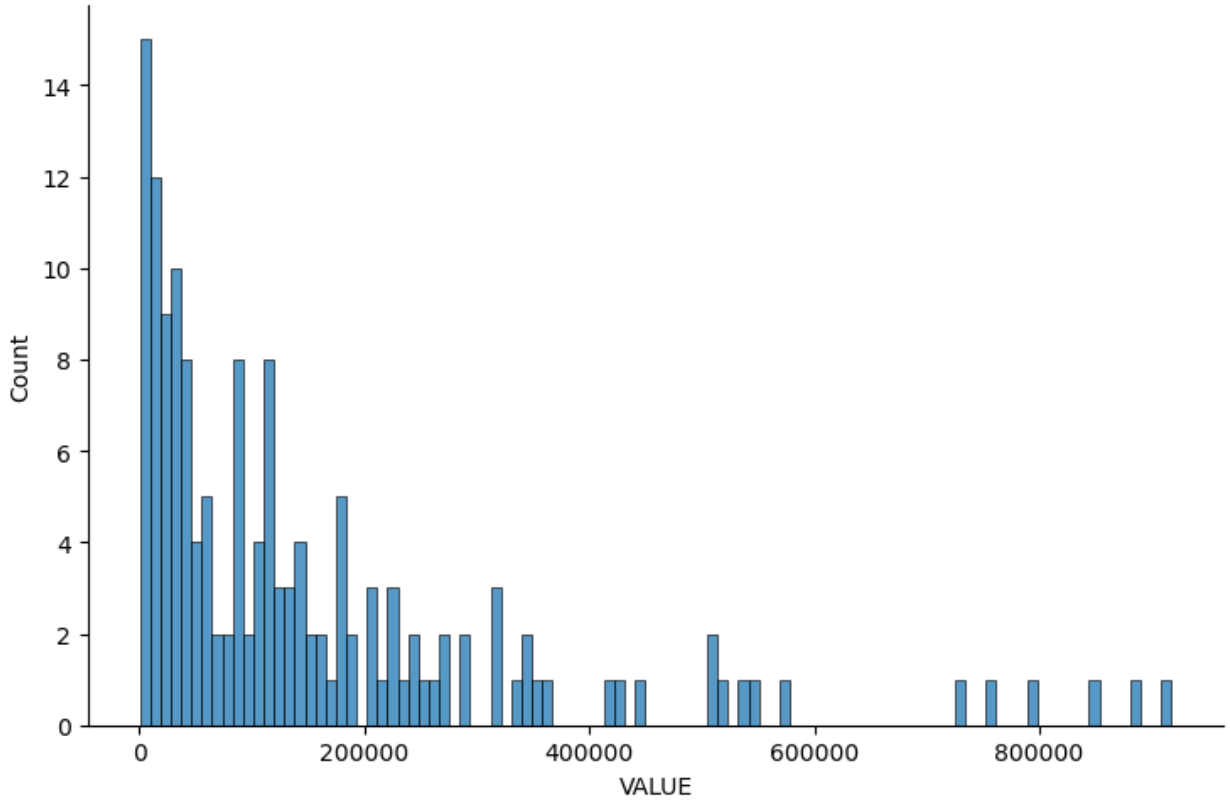
Done using Stripplot
[1, 0] = ['Male employees' 'Female employees']

Output #18c

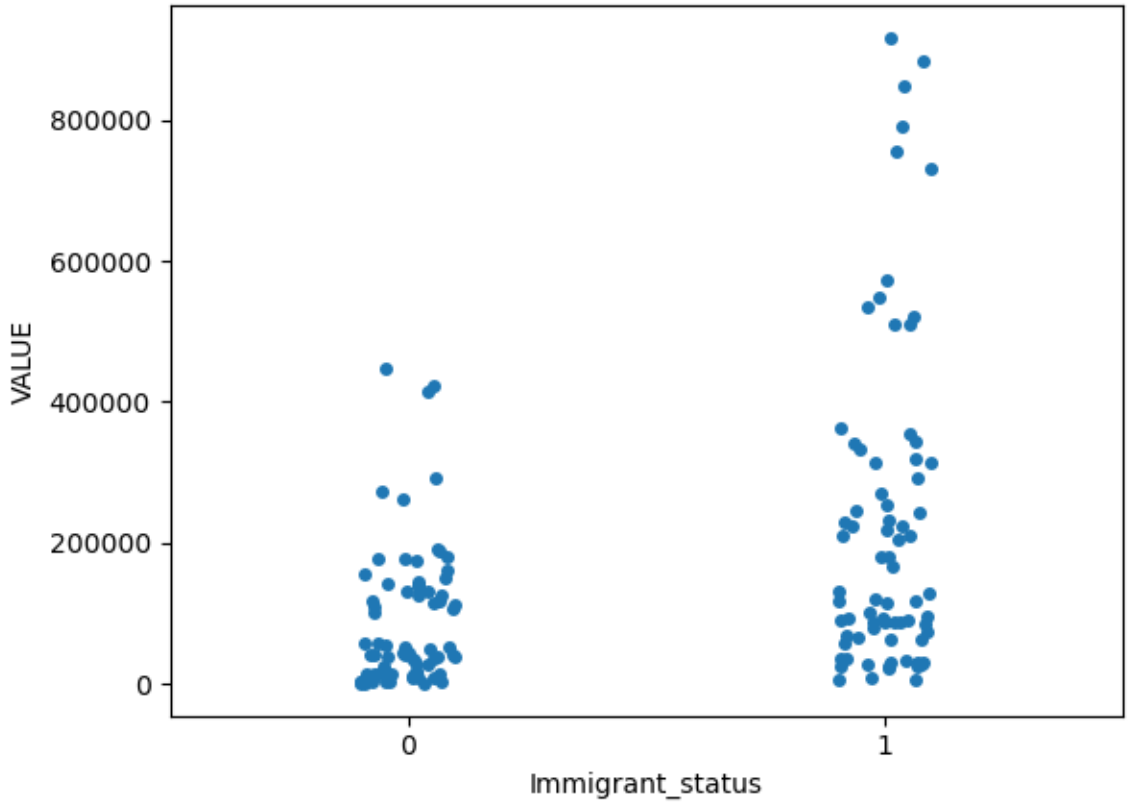
Result for testing set for 'Hours Worked' by Immigrant status

```
final_testing_df_output_df_Hrs_Wrked_ByImmigrant.csv
                                sum           mean    amin    median  \
Characteristics
Immigrant employees       6592022.0   87893.626667   550.0   43620.0
Non-immigrant employees  16953727.0  226049.693333  6197.0  127014.0

                             amax  size
Characteristics
Immigrant employees      447657.0    75
Non-immigrant employees  917521.0    75
Overall,
Sum :  23545749.0
Mean :  156971.66
Min/median/max : 550.0 / 92441.5 / 917521.0
Standard Deviation :  188873.98307114118
Skewnewss :  2.0992731555277357
Total size :  150
```

final_testing_df_output_df_Hrs_Wrked_ByImmigrant.csv



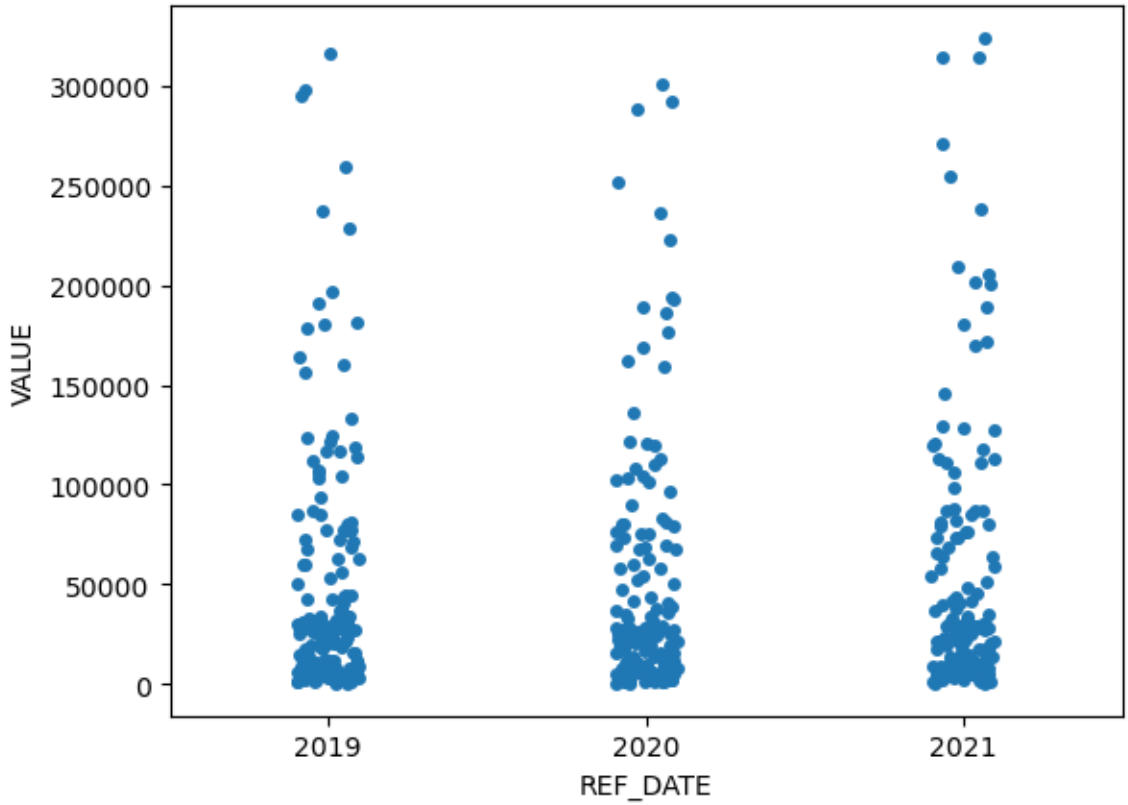
final_testing_df_output_df_Hrs_Wrked_ByImmigrant.csv



Done using Stripplot
[0, 1] = ['Immigrant employees' 'Non-immigrant employees']

Result for testing set for 'Hours Worked' by yearly

`final_testing_df_output_df_Hrs_Wrked_ByAge.csv`



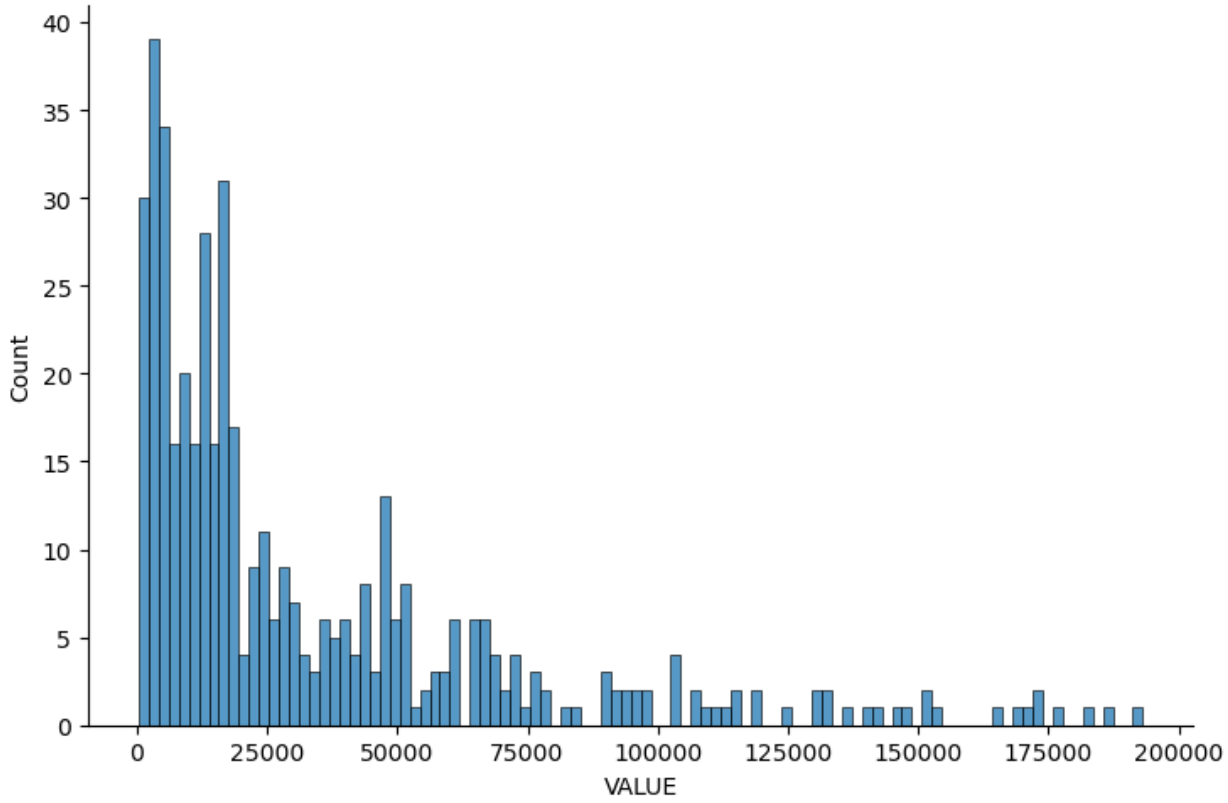
`Done using Stripplot`

Output #19a

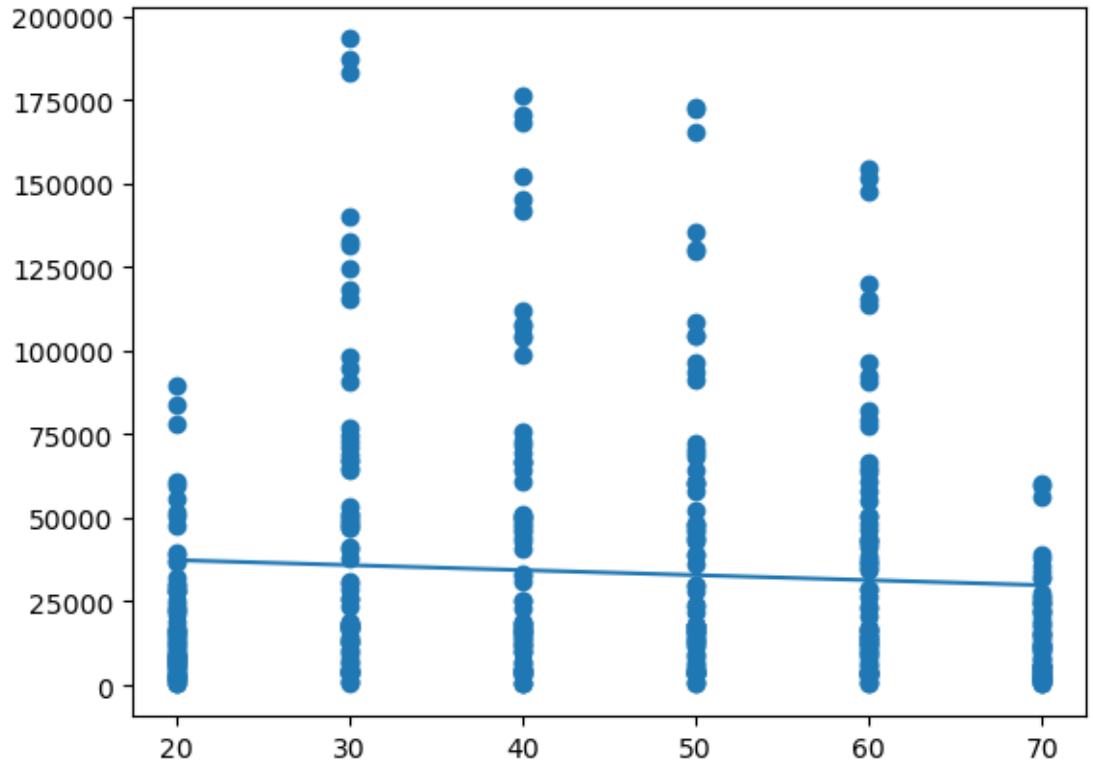
Result for testing set for 'Number of jobs' by Age group

```
final_testing_df_output_df_NumOfJob_ByAge.csv
                             sum          mean   amin   median      amax  size
Characteristics
15 to 24 years         1473846.0  19651.280000  597.0  12853.0   89352.0    75
25 to 34 years         3351530.0  44687.066667  845.0  25732.0  193127.0    75
35 to 44 years         3309827.0  44131.026667  729.0  25101.0  176247.0    75
45 to 54 years         3137650.0  41835.333333  809.0  23657.0  172644.0    75
55 to 64 years         2799581.0  37327.746667  800.0  22957.0  154501.0    75
65 years old and over  1052065.0  14027.533333  322.0  10604.0   60497.0    75
Overall,
Sum :  15124499.0
Mean :  33609.997777777775
Min/median/max : 322.0 / 17335.0 / 193127.0
Standard Deviation :  38617.96878083792
Skewnewss :  1.8801308823526393
Total size :  450
```

final_testing_df_output_df_NumOfJob_ByAge.csv



final_testing_df_output_df_NumOfJob_ByAge.csv



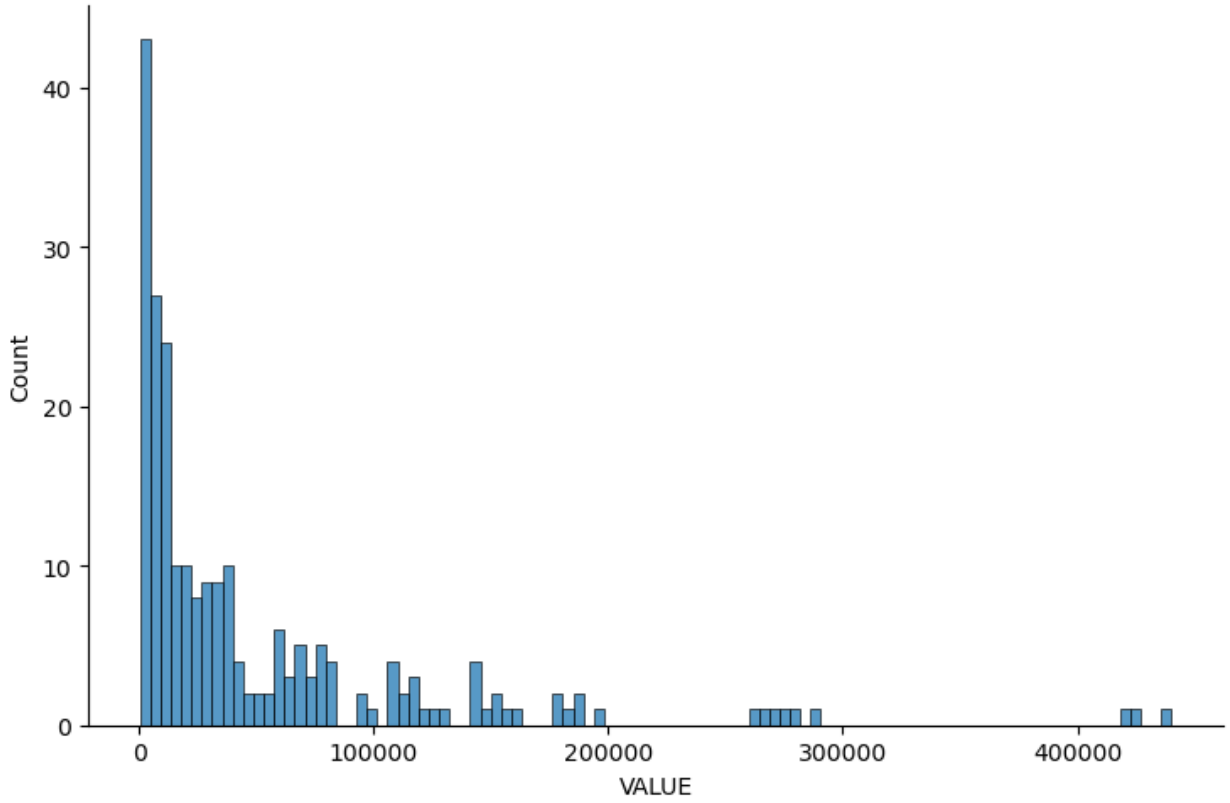
Done by Linear Regression

Output #19b

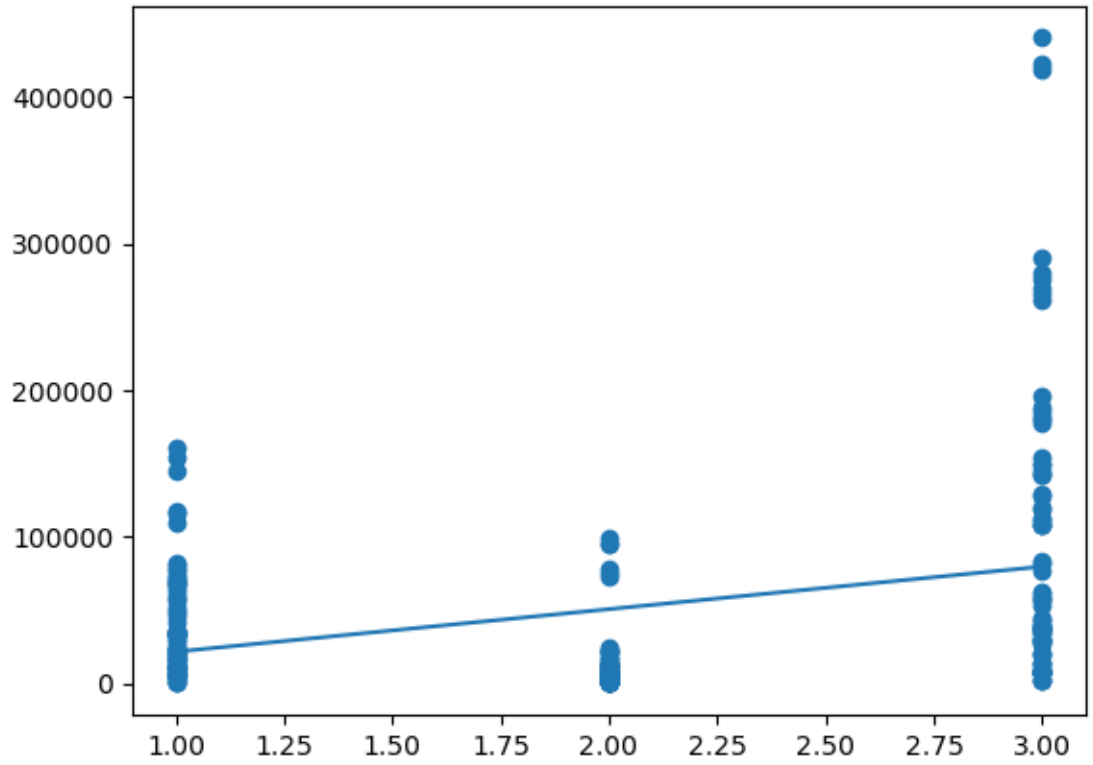
Result for testing set for 'Number of jobs' by Education level

```
final_testing_df_output_df_NumOfJob_ByEducation.csv
                                    sum          mean    amin   median  \
Characteristics
High school diploma and less  3006364.0  40084.853333  1289.0  30412.0
Trade certificate             1010387.0  13471.826667   217.0   5607.0
University degree and higher  7363505.0  98180.066667  1692.0  57742.0

                                   amax  size
Characteristics
High school diploma and less  160257.0    75
Trade certificate              99207.0    75
University degree and higher  440185.0    75
Overall,
Sum :  11380256.0
Mean :  50578.915555555555
Min/median/max : 217.0 / 21499.0 / 440185.0
Standard Deviation :  73477.22284618838
Skewnewss :  2.7652609490298943
Total size :  225
```

final_testing_df_output_df_NumOfJob_ByEducation.csv



final_testing_df_output_df_NumOfJob_ByEducation.csv



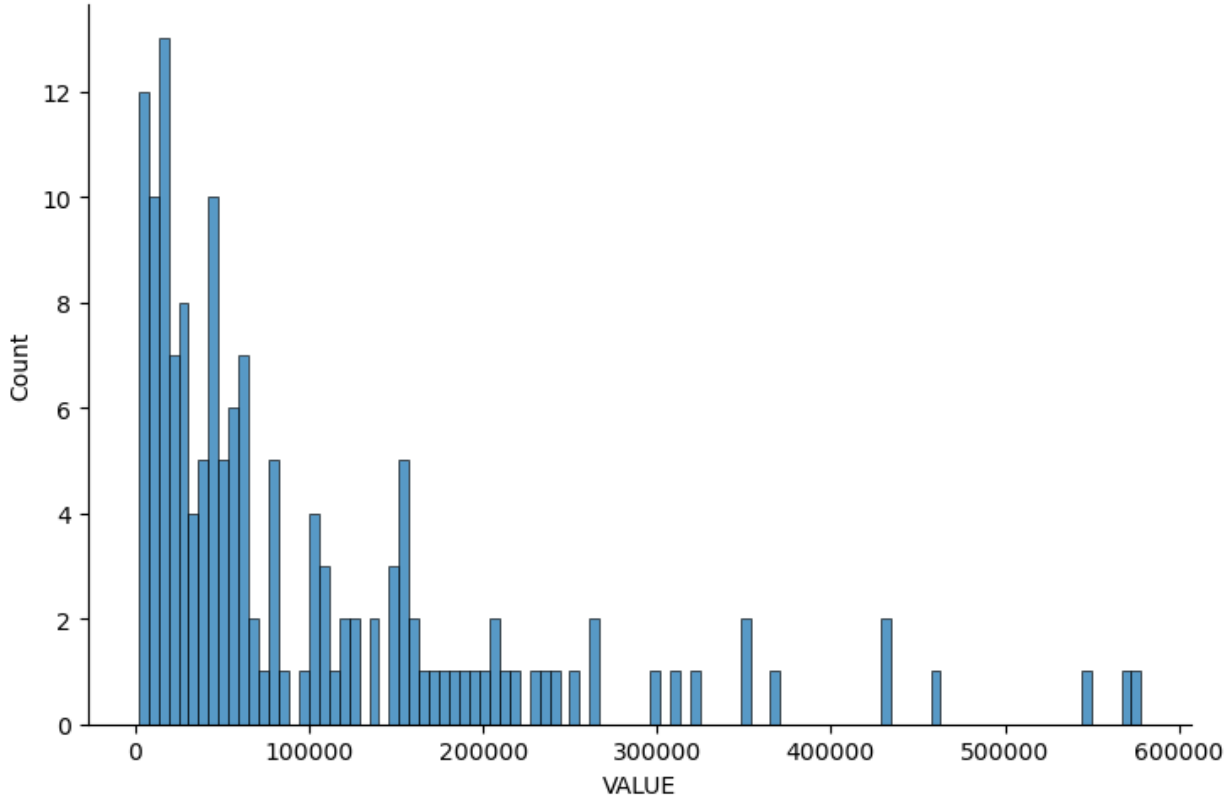
Done by Linear Regression
Higher the number, higher the education

Output #19c

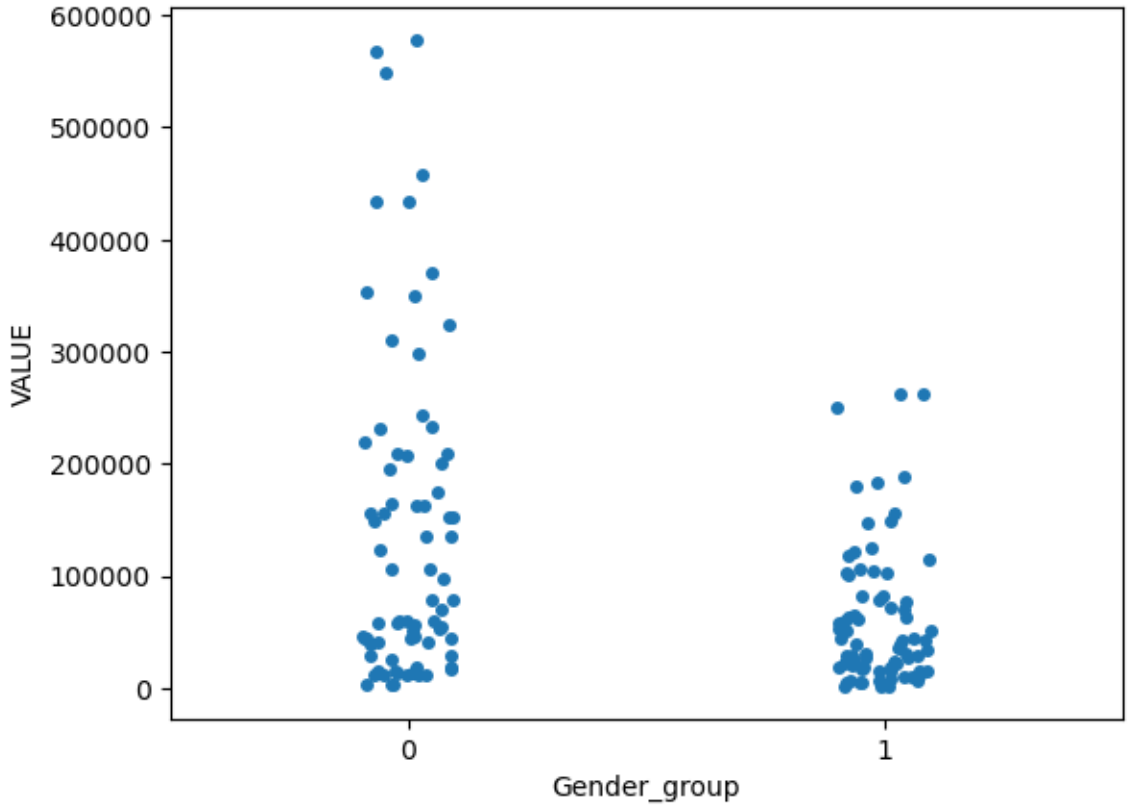
Result for testing set for 'Number of jobs' by Gender group

```
final_testing_df_output_df_NumOfJob_ByGender.csv
                         sum           mean    amin   median      amax  size
Characteristics
Female employees  10488836.0  139851.146667  2372.0  78504.0  578450.0    75
Male employees     4635657.0   61808.760000  1730.0  39774.0  262447.0    75
Overall,
Sum :  15124493.0
Mean :  100829.95333333334
Min/median/max : 1730.0 / 55510.0 / 578450.0
Standard Deviation :  117420.78828676158
Skewnewss :  2.041334832154595
Total size :  150
```

final_testing_df_output_df_NumOfJob_ByGender.csv



final_testing_df_output_df_NumOfJob_ByGender.csv



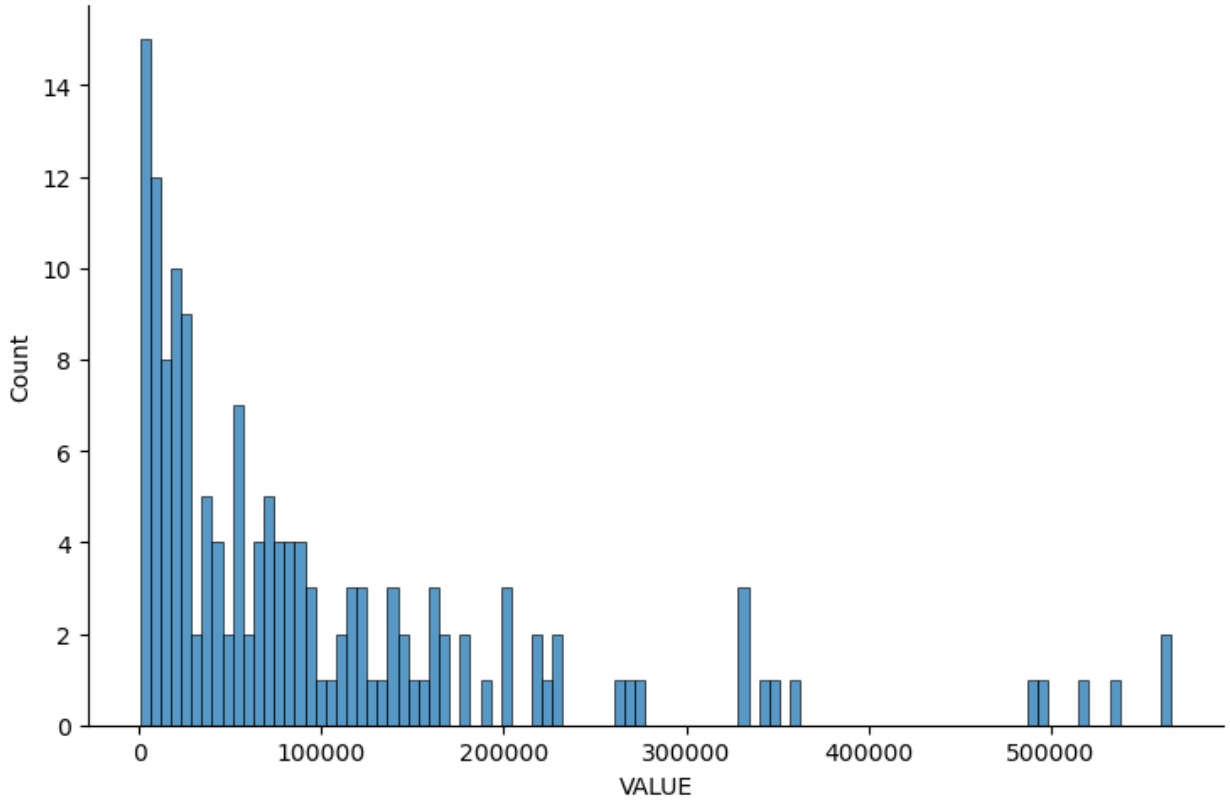
Done using Stripplot
[1, 0] = ['Male employees' 'Female employees']

Output #19d

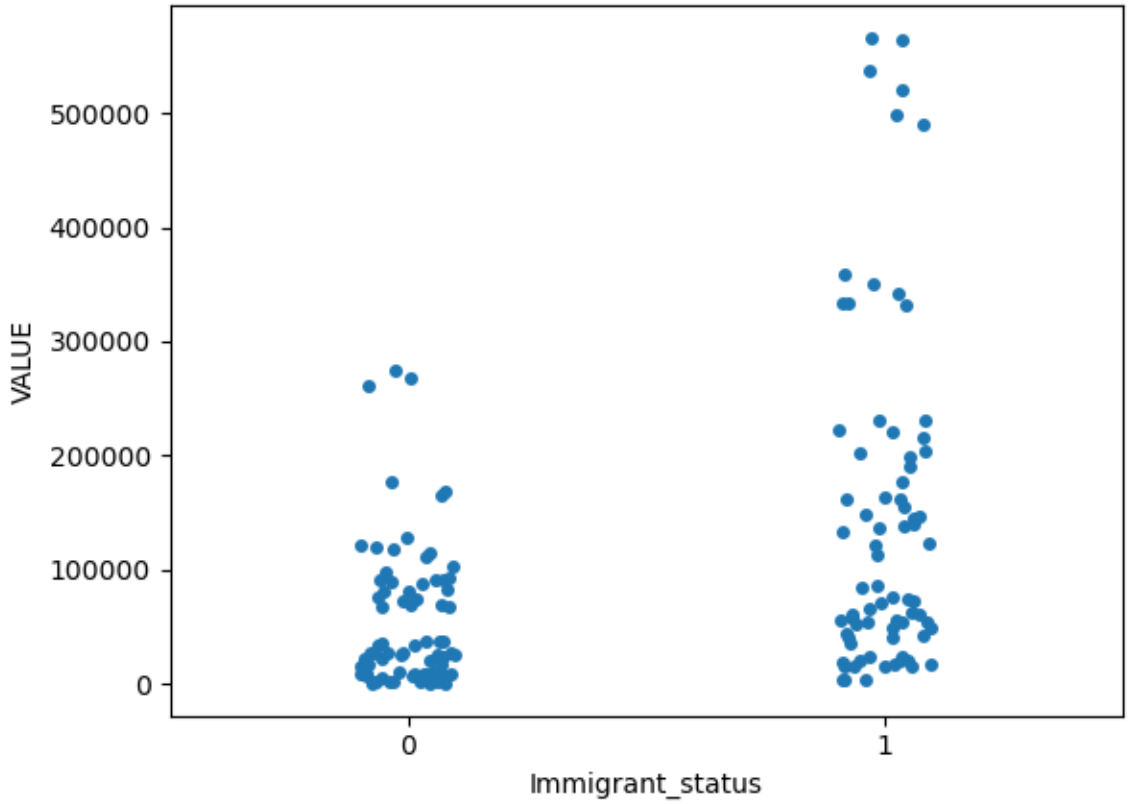
Result for testing set for 'Number of jobs' by Immigrant status

```
final_testing_df_output_df_NumOfJob_ByImmigrant.csv
                               sum           mean    amin   median      amax  \
Characteristics
Immigrant employees       4198421.0   55978.946667   337.0  27088.0  274559.0
Non-immigrant employees  10926074.0  145680.986667  3765.0  84077.0  566338.0

                         size
Characteristics
Immigrant employees        75
Non-immigrant employees    75
Overall,
Sum :  15124495.0
Mean :  100829.96666666666
Min/median/max : 337.0 / 60500.0 / 566338.0
Standard Deviation :  121238.11500027079
Skewnewss :  2.082806355312799
Total size :  150
```

`final_testing_df_output_df_NumOfJob_ByImmigrant.csv`



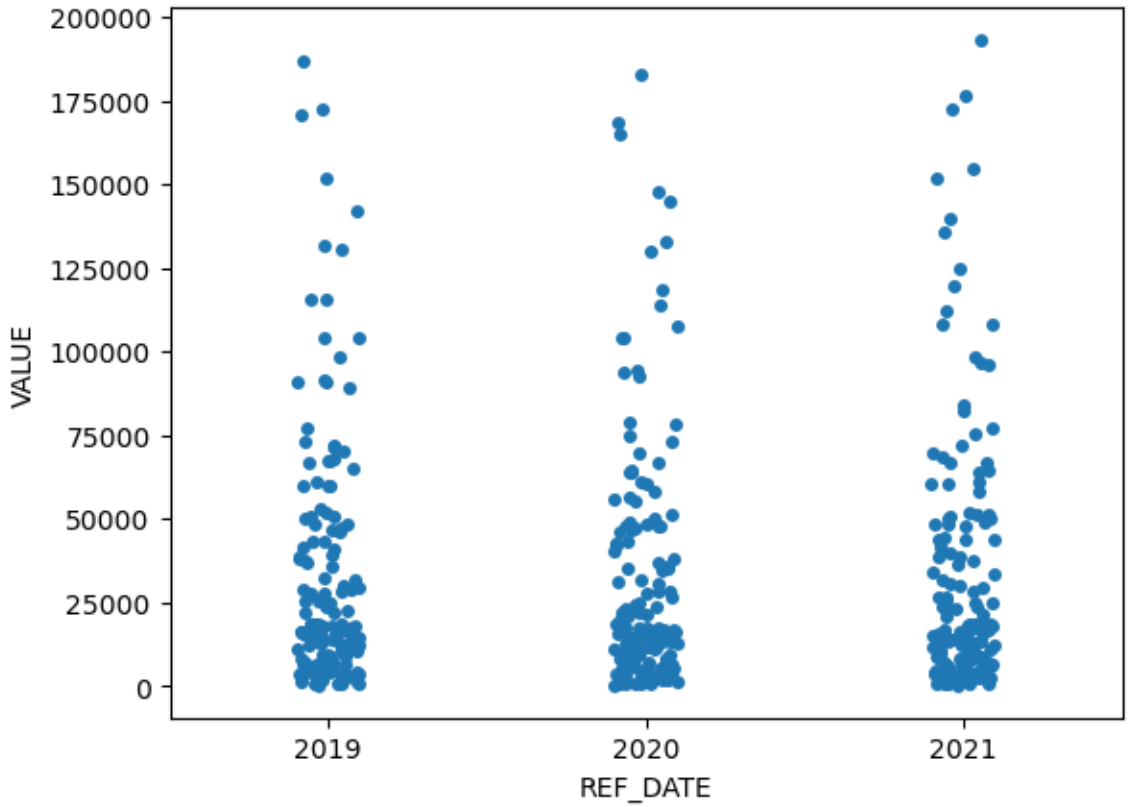
final_testing_df_output_df_NumOfJob_ByImmigrant.csv



Done using Stripplot
[0, 1] = ['Immigrant employees' 'Non-immigrant employees']

Result for testing set for 'Number of jobs' by yearly.

`final_testing_df_output_df_NumOfJob_ByAge.csv`

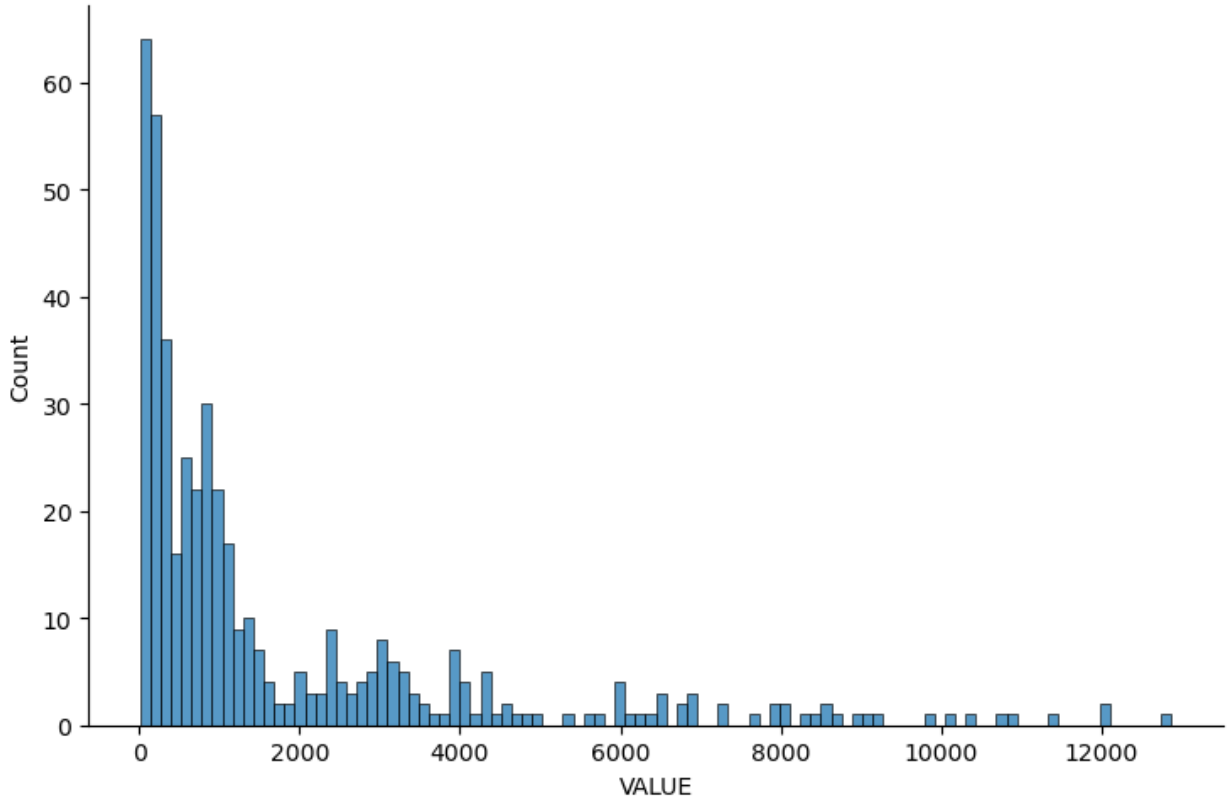


`Done using Stripplot`

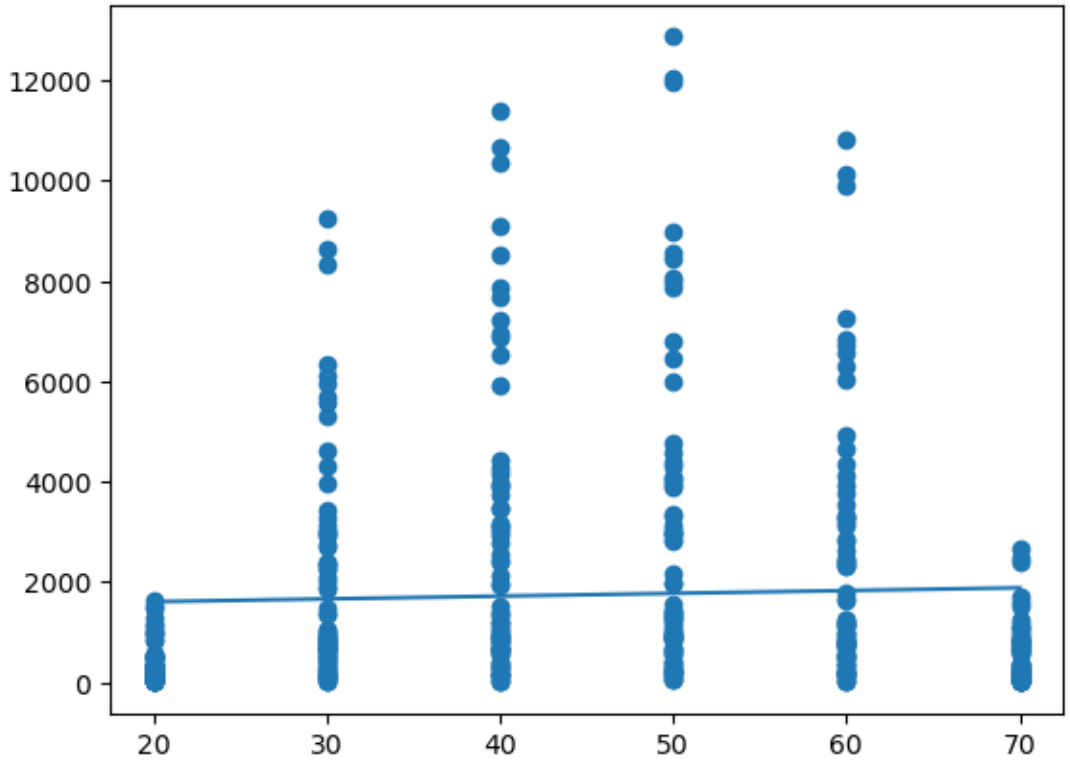
Result for testing set for ‘Wages and Salaries' by Age group

```
final_testing_df_output_df_WagesAndSalaries_ByAge.csv
                            sum         mean  amin  median     amax  size
Characteristics
15 to 24 years          27272.0   363.626667   8.0   188.0   1622.0    75
25 to 34 years         147099.0  1961.320000  32.0   921.0   9270.0    75
35 to 44 years         194164.0  2588.853333  41.0  1199.0  11414.0    75
45 to 54 years         208519.0  2780.253333  50.0  1337.0  12872.0    75
55 to 64 years         169578.0  2261.040000  42.0  1159.0  10830.0    75
65 years old and over   39976.0   533.013333   9.0   268.0   2660.0    75
Overall,
Sum :  786608.0
Mean :  1748.0177777777778
Min/median/max : 8.0 / 799.5 / 12872.0
Standard Deviation :  2388.2846414472547
Skewnewss :  2.18876652733957
Total size :  450
```

final_testing_df_output_df_WagesAndSalaries_ByAge.csv



final_testing_df_output_df_WagesAndSalaries_ByAge.csv

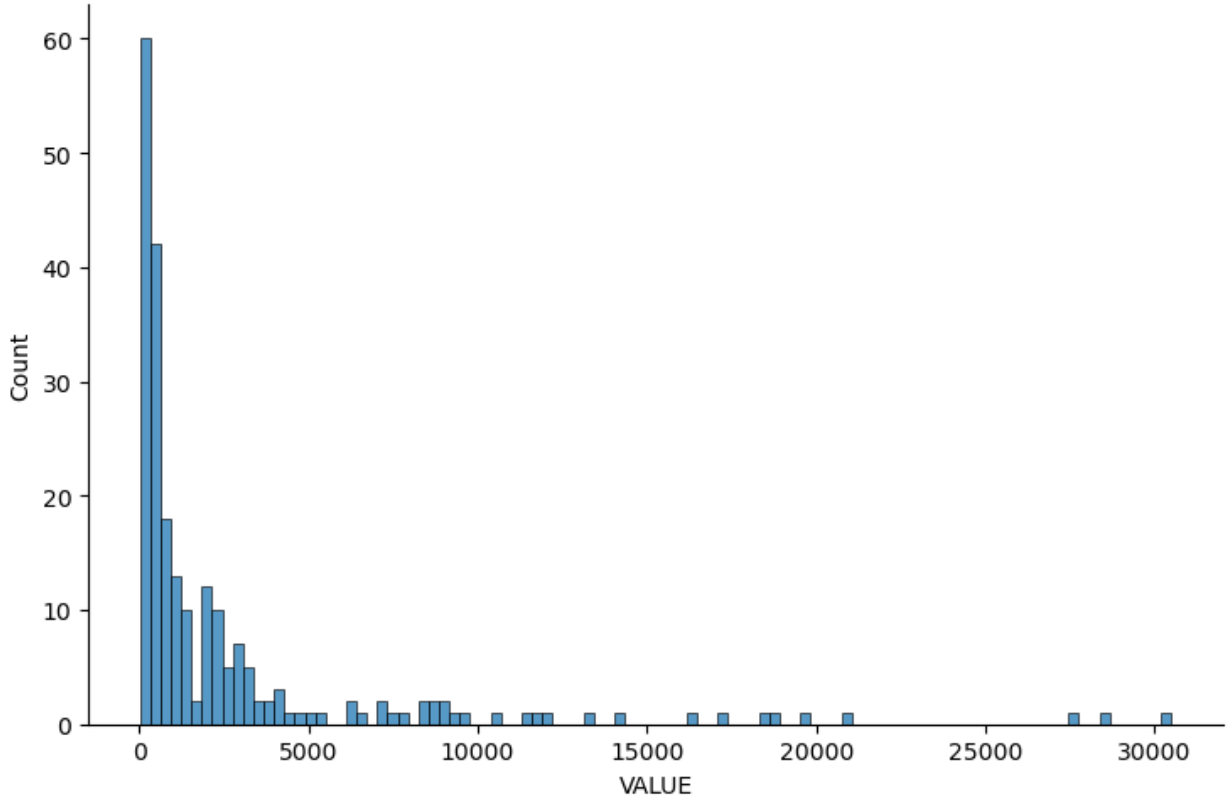


Done by Linear Regression

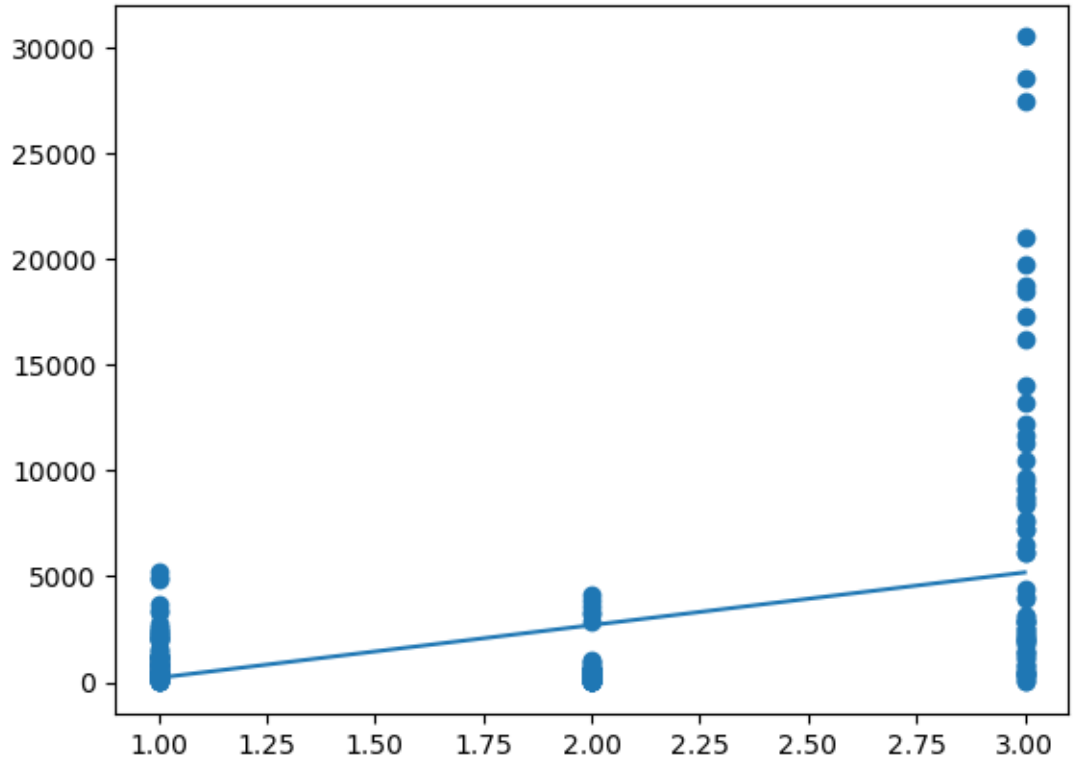
Result for testing set for 'Wages and Salaries' by Education level

```
final_testing_df_output_df_WagesAndSalaries_ByEducation.csv
                                    sum         mean  amin  median     amax  \
Characteristics
High school diploma and less   95794.0  1277.253333  37.0   861.0   5226.0
Trade certificate              42027.0   560.360000  10.0   231.0   4105.0
University degree and higher  469262.0  6256.826667  91.0  2895.0  30502.0

                              size
Characteristics
High school diploma and less    75
Trade certificate               75
University degree and higher    75
Overall,
Sum :  607083.0
Mean :  2698.1466666666665
Min/median/max : 10.0 / 801.0 / 30502.0
Standard Deviation :  4868.765589464701
Skewnewss :  3.293087045340116
Total size :  225
```

final_testing_df_output_df_WagesAndSalaries_ByEducation.csv



final_testing_df_output_df_WagesAndSalaries_ByEducation.csv

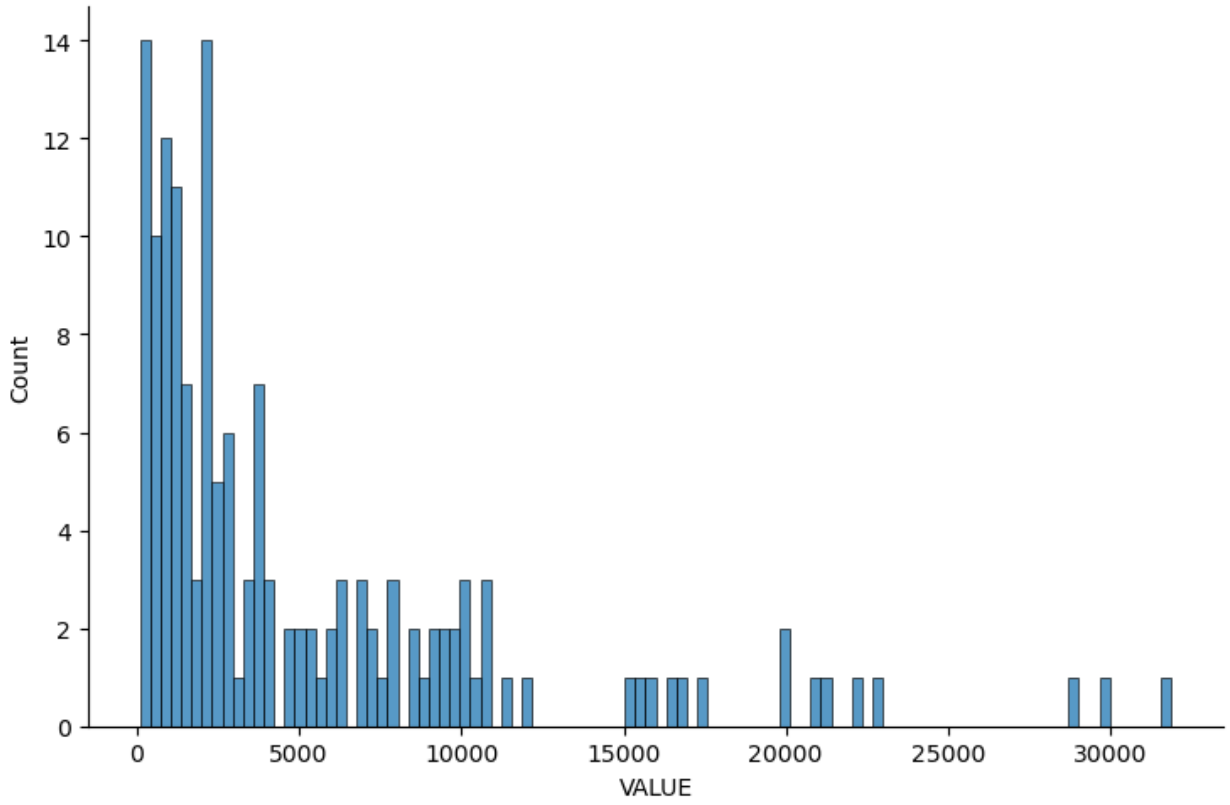


Done by Linear Regression
Higher the number, higher the education

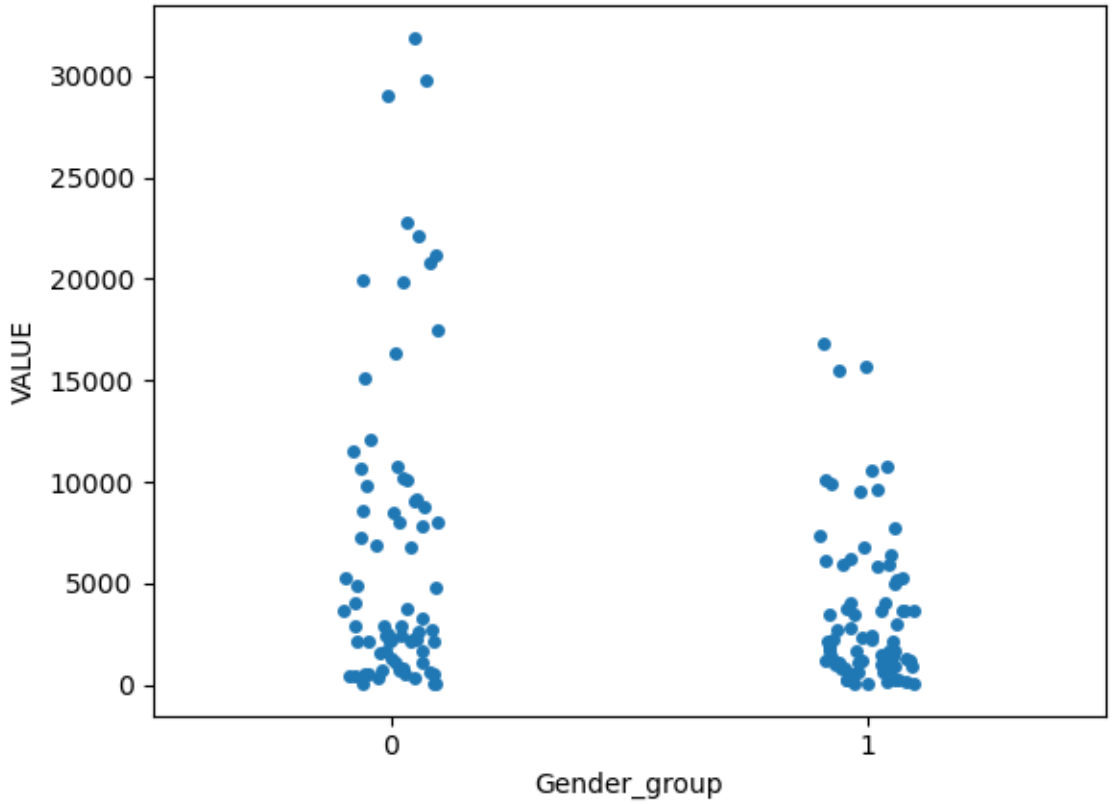
Result for testing set for 'Wages and Salaries' by Gender group

```
final_testing_df_output_df_WagesAndSalaries_ByGender.csv
                        sum         mean  amin  median     amax  size
Characteristics
Female employees  516398.0  6885.306667  94.0  2932.0  31892.0    75
Male employees    270217.0  3602.893333  84.0  2176.0  16776.0    75
Overall,
Sum :  786615.0
Mean :  5244.1
Min/median/max : 84.0 / 2563.0 / 31892.0
Standard Deviation :  6289.8685749385895
Skewnewss :  2.0515596833116616
Total size :  150
```

final_testing_df_output_df_WagesAndSalaries_ByGender.csv



final_testing_df_output_df_WagesAndSalaries_ByGender.csv

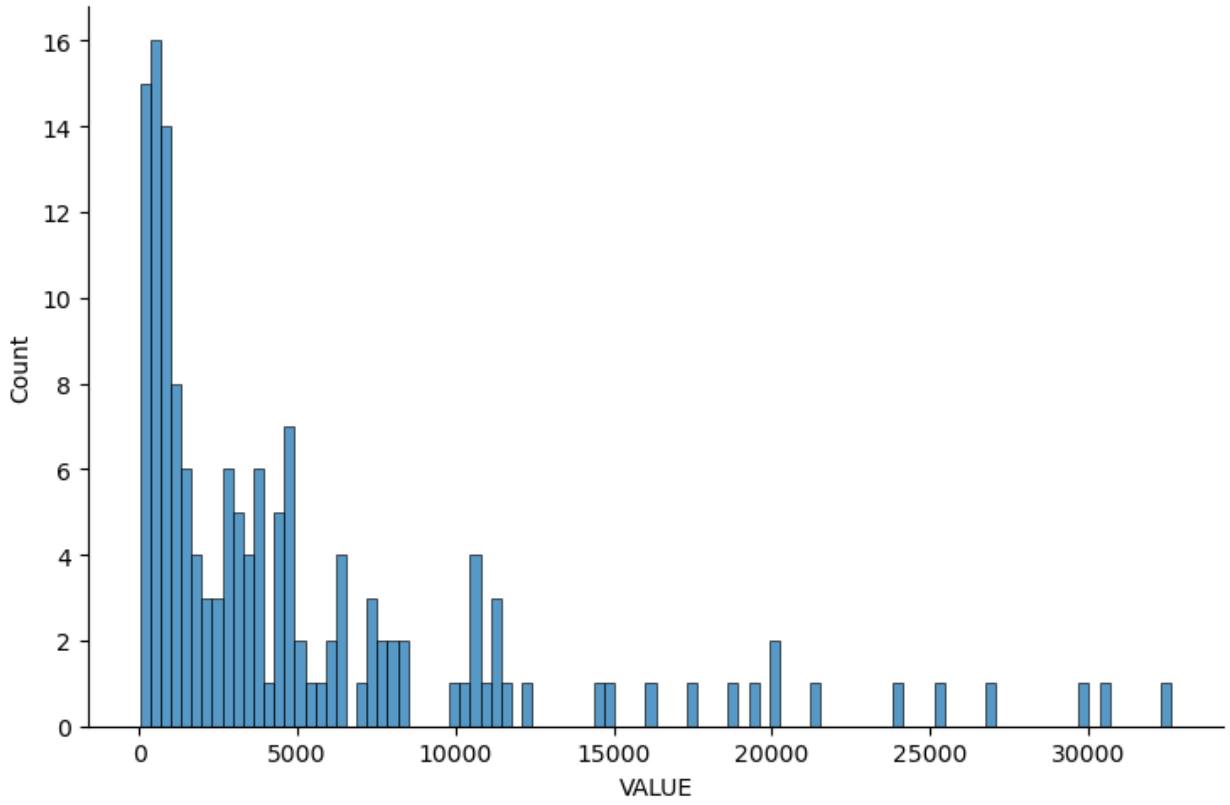


Done using Stripplot
[1, 0] = ['Male employees' 'Female employees']

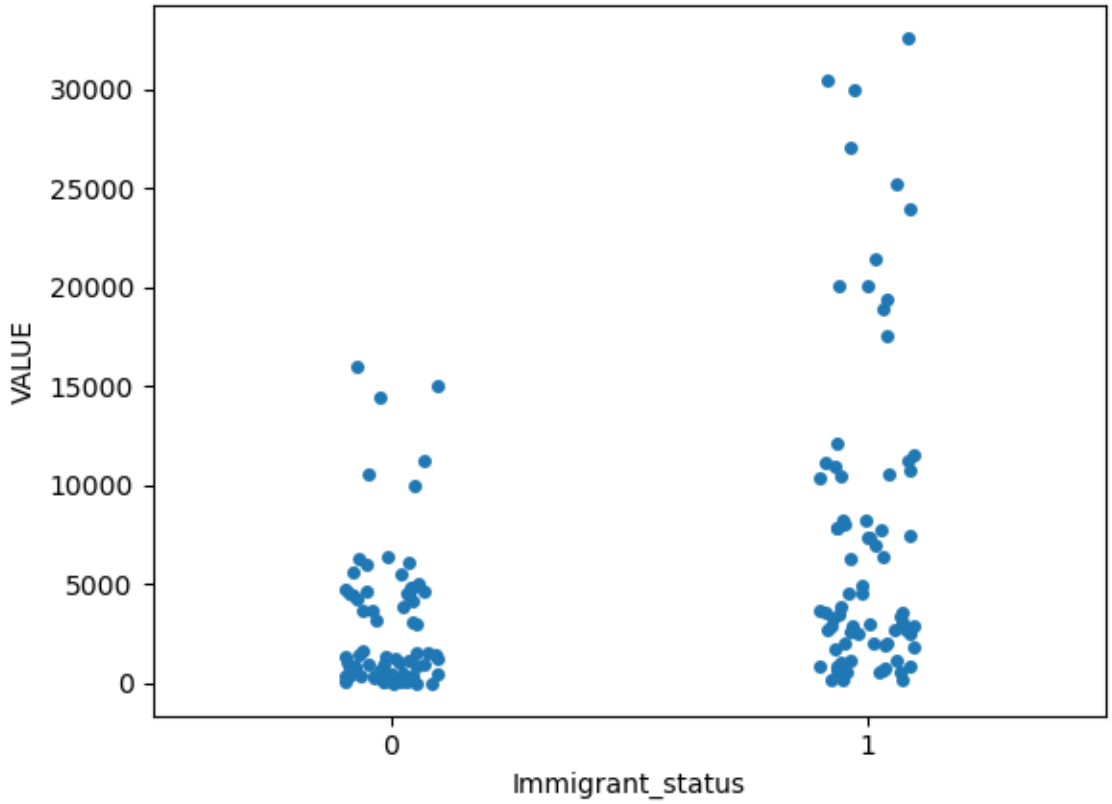
Result for testing set for 'Wages and Salaries' by Immigrant status

```
final_testing_df_output_df_WagesAndSalaries_ByImmigrant.csv
                              sum         mean   amin  median     amax  size
Characteristics
Immigrant employees      218847.0  2917.960000   14.0  1243.0  16030.0    75
Non-immigrant employees  567772.0  7570.293333  169.0  3683.0  32637.0    75
Overall,
Sum :  786619.0
Mean :  5244.126666666667
Min/median/max : 14.0 / 2977.5 / 32637.0
Standard Deviation :  6672.173255441005
Skewnewss :  2.1294707124700127
Total size :  150
```

final_testing_df_output_df_WagesAndSalaries_ByImmigrant.csv



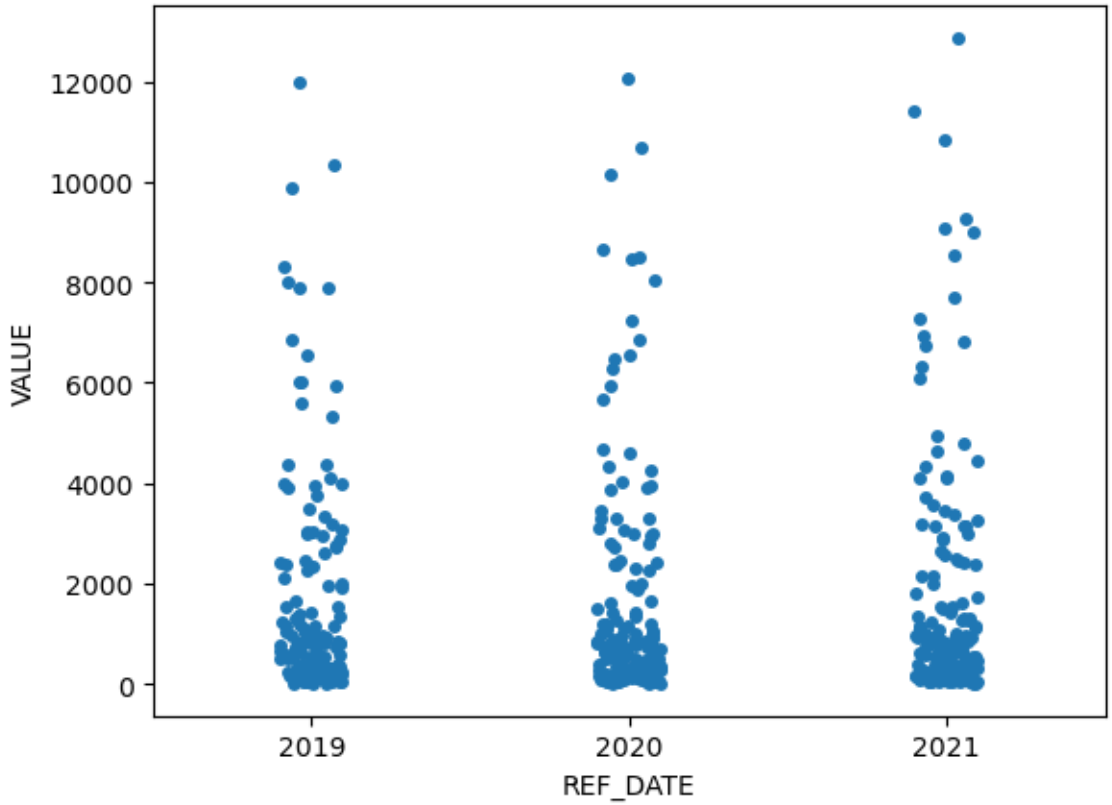
final_testing_df_output_df_WagesAndSalaries_ByImmigrant.csv



Done using Stripplot
[0, 1] = ['Immigrant employees' 'Non-immigrant employees']

Result for testing set for 'Wages and Salaries' by yearly

`final_testing_df_output_df_WagesAndSalaries_ByAge.csv`



`Done using Stripplot`