

Employment in non-profit sector

By: Sangjin
(Eric) Lee

Ryerson
University

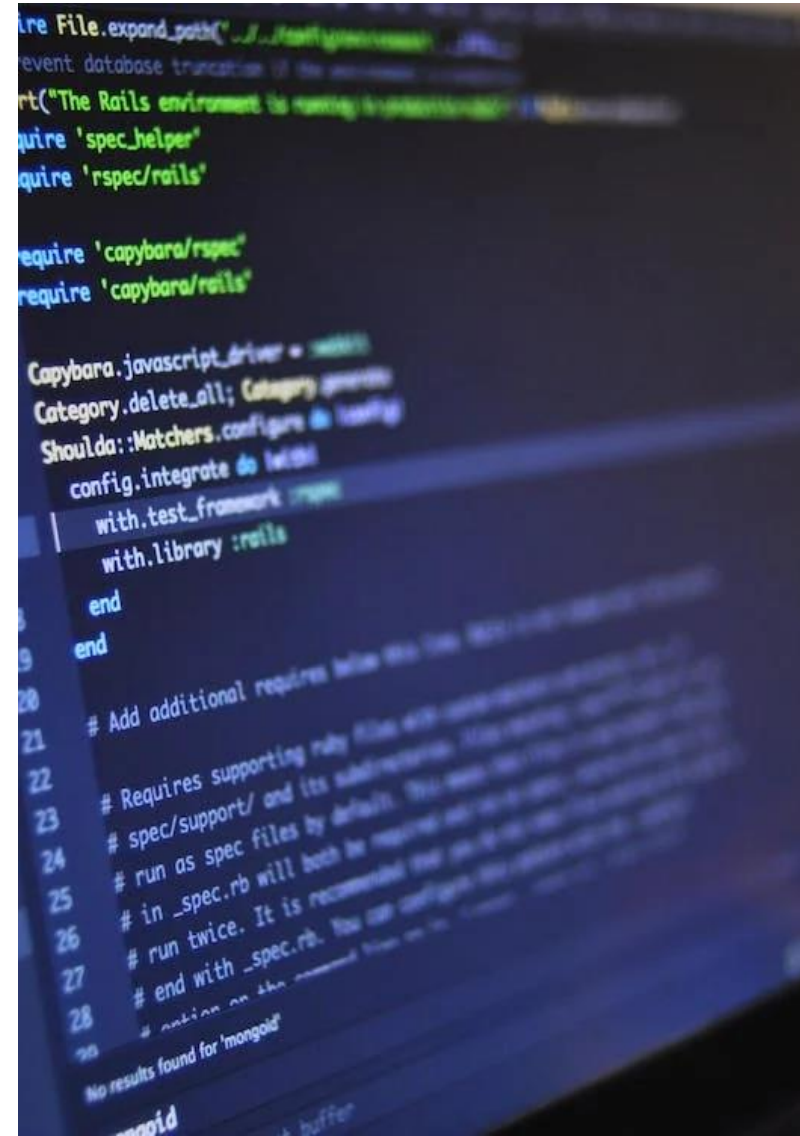


Objective and Goals

- How did the working condition in non-profit organizations in Canada has improved?
- What's the current trends for employment in term of salary, wages available?
- Which age group, gender group, education levels, and immigrant status has benefit and drawback in non-profit organizations?

Source of the dataset

- Open Government website provided by Statistics of Canada
- Link:
<https://open.canada.ca/data/en/dataset/edc0fe3c-23a3-4ccf-929b-8a385b62f6c3>



Models and Techniques used

- Data classification
 - Decision Trees
- Data clustering
- Data Mining

Data processing and preparation

Initial data processing

- Upload the ‘csv files’ into the dataset
- Remove unnecessary columns from the dataset
- Filter “Null” observations from the dataset

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 105840 entries, 0 to 105839
```

```
Data columns (total 9 columns):
```

#	Column	Non-Null Count	Dtype
0	REF_DATE	105840 non-null	int64
1	DGUID	105840 non-null	object
2	GEO	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	SCALAR_FACTOR	105840 non-null	object
8	VALUE	102816 non-null	float64

```
dtypes: float64(1), int64(1), object(7)
```

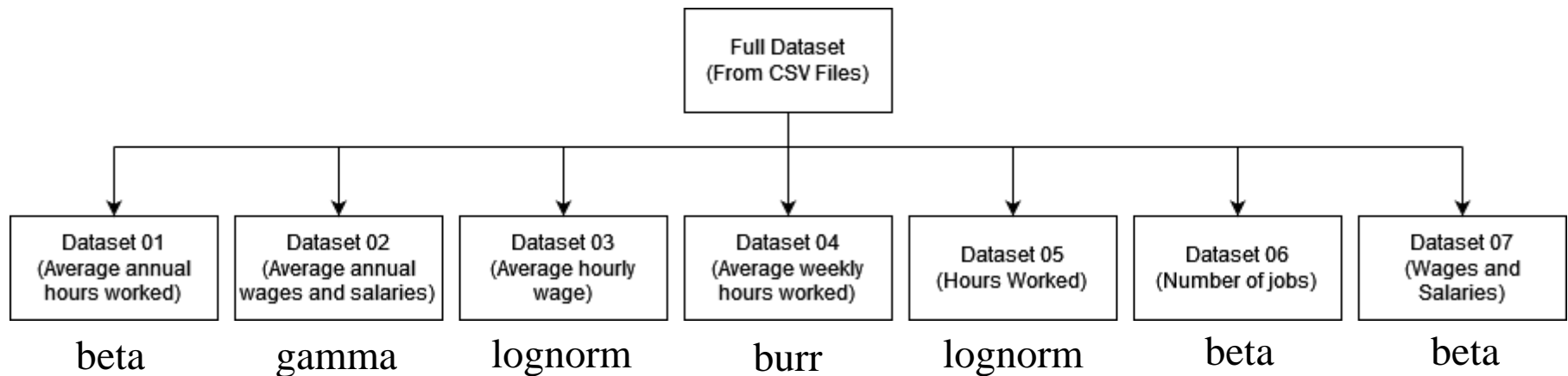
```
memory usage: 7.3+ MB
```

```
None
```

	percent_in_na	num_of_na
VALUE	2.857143	3024
	↓	
VALUE	percent_in_na	0.0

Splitting the dataset based on 'Indicators'

- Divided into Seven dataset/indicators and run best fit.
 - Measured in Annual/Weekly are distributed
 - Otherwise, skewed left.
- Did brief Cohort Analysis after splitting



Run best fit for each (Not all them are normally distributed)

Splitting the dataset based on 'Indicators'

- Did brief Cohort Analysis after splitting

	Cohort	Month 0							
2010	1/31/2010	1224							
2011	2/28/2010	1224							
2012	3/31/2010	1224							
2013	4/30/2010	1224							
2014	5/31/2010	1224							
2015	6/30/2010	1224							
2016	7/31/2010	1224							
2017	8/31/2010	1224							
2018	9/30/2010	1224							
2019	10/31/2010	1224							
2020	11/30/2010	1224							
2021	12/31/2010	1224							
	All	9792	0	0	0	0	0	0	0

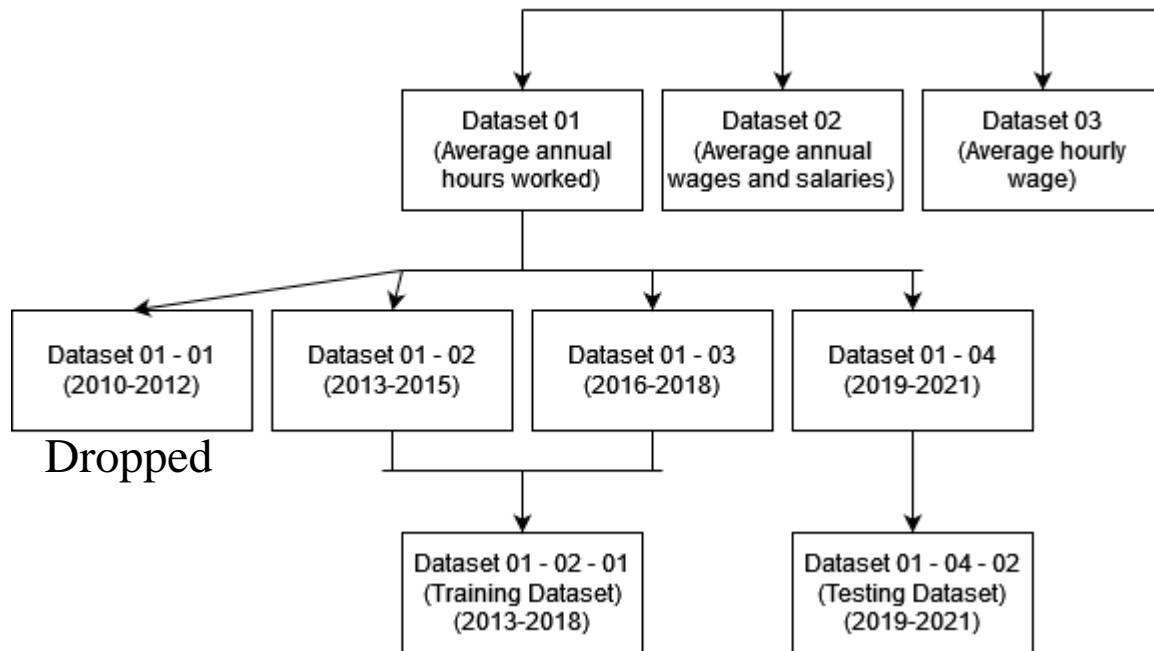
Finish it using Excel
Only 'Average annual hours worked'

Splitting the Training and Testing set

- Divide into four dataset initially, 2010-2012, 2013-2015, 2016-2018, 2019-2021.
- Drop 2010-2012 dataset
- Merge 2013-2015, 2016-2018 into training set.
- Use 2019-2021 into testing set.
- Do Chi-Square before do another splitting

Splitting the Training and Testing set

... Continued ...



	sum	size
REF_DATE		
2013	36671.0	1224
2014	36483.0	1224
2015	36678.0	1224
2016	36541.0	1224
2017	36170.0	1224
2018	36416.0	1224
	sum	size
REF_DATE		
2019	36400.0	1224
2020	36036.0	1224
2021	36555.0	1224

Chi-square:

AvgAnnHrsWrk : p value is 1.0, Independent (H0 holds true)

AvgAnnHrsWages : p value is 1.0, Independent (H0 holds true)

AvgWeekHrsWrked : p value is 1.0, Independent (H0 holds true)

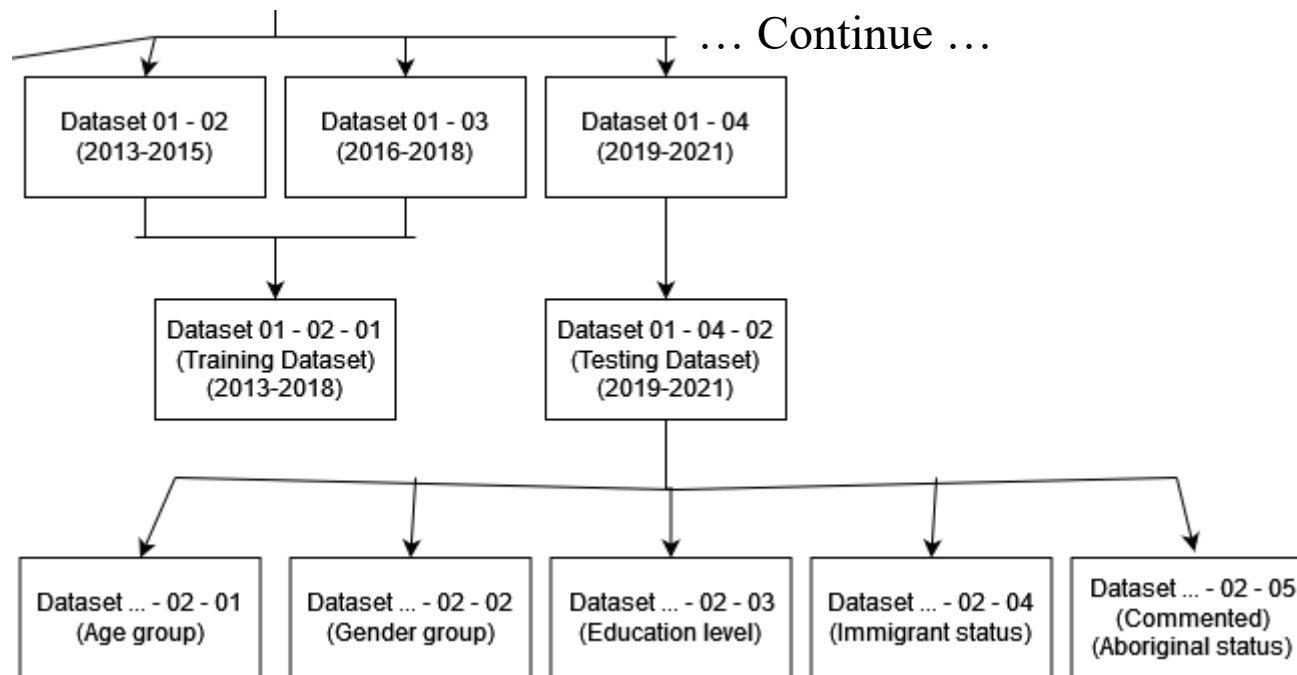
Hours Worked : p value is 1.0, Independent (H0 holds true)

Number of Jobs : p value is 1.0, Independent (H0 holds true)

Wages and Salaries : p value is 1.0, Independent (H0 holds true)

Divide by 'Characteristics' columns

- Divide into four dataset based on 'Age', 'Gender', 'Education', 'Immigrant', repeat seven times.



```

testing set By Age
sum      mean  amin  median  amax  size
Characteristics
15 to 24 years  179872.0  936.833333  715.0  927.5  1281.0  192
25 to 34 years  333662.0  1508.866667  1292.0  1576.0  1878.0  210
35 to 44 years  371386.0  1768.584762  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  286760.0  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
Skewness : -0.5829393561867995
Total size : 1224

testing set By Gender
sum      mean  amin  median  amax  size
Characteristics
Female employees  123755.0  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
Skewness : -0.305090883263471
Total size : 420

testing set By Education
sum      mean  amin  median  amax  size
Characteristics
High school diploma and less  275636.0  1312.552381  1054.0  1307.5  1667.0  210
Trade certificate  306174.0  1546.133333  789.0  1547.5  1808.0  198
University degree and higher  341795.0  1726.237374  1536.0  1706.5  2043.0  198
Overall,
Sum : 923605.0
Mean : 1524.1006608660866
Min/median/max : 789.0 / 1545.0 / 2043.0
Standard Deviation : 294.02000670247472
Skewness : -0.1568377438071896
Total size : 606

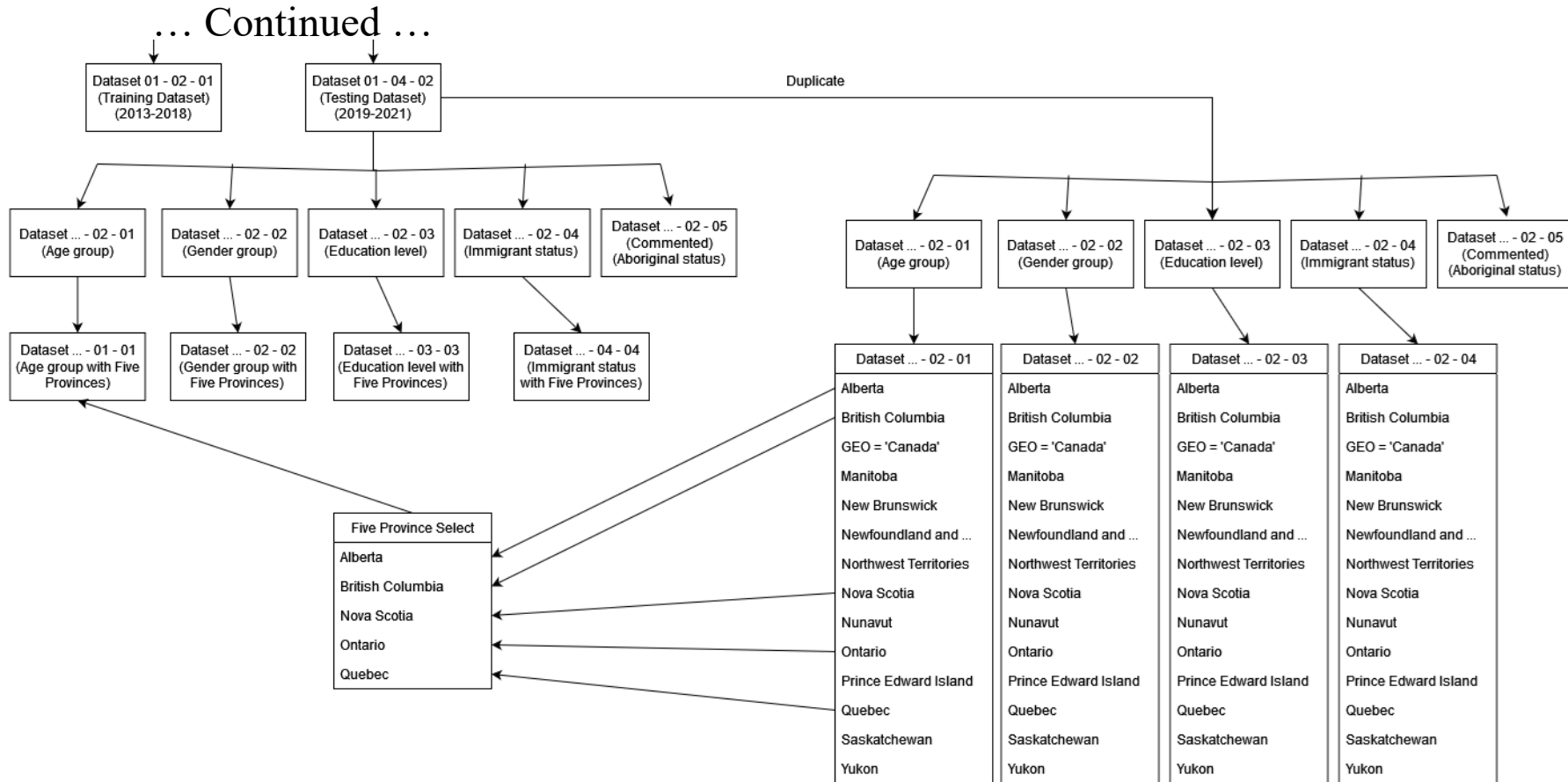
testing set By Immigrant
sum      mean  amin  median  amax  size
Characteristics
Immigrant employees  138818.0  1610.191919  1336.0  1589.0  2250.0  198
Non-immigrant employees  310245.0  1566.893939  1315.0  1578.0  1767.0  198
Overall,
Sum : 629063.0
Mean : 1588.5429202020202
Min/median/max : 1315.0 / 1580.0 / 2250.0
Standard Deviation : 112.27714481578558
Skewness : 1.6694528665533967
Total size : 396
  
```

Divide by 'Provinces' columns

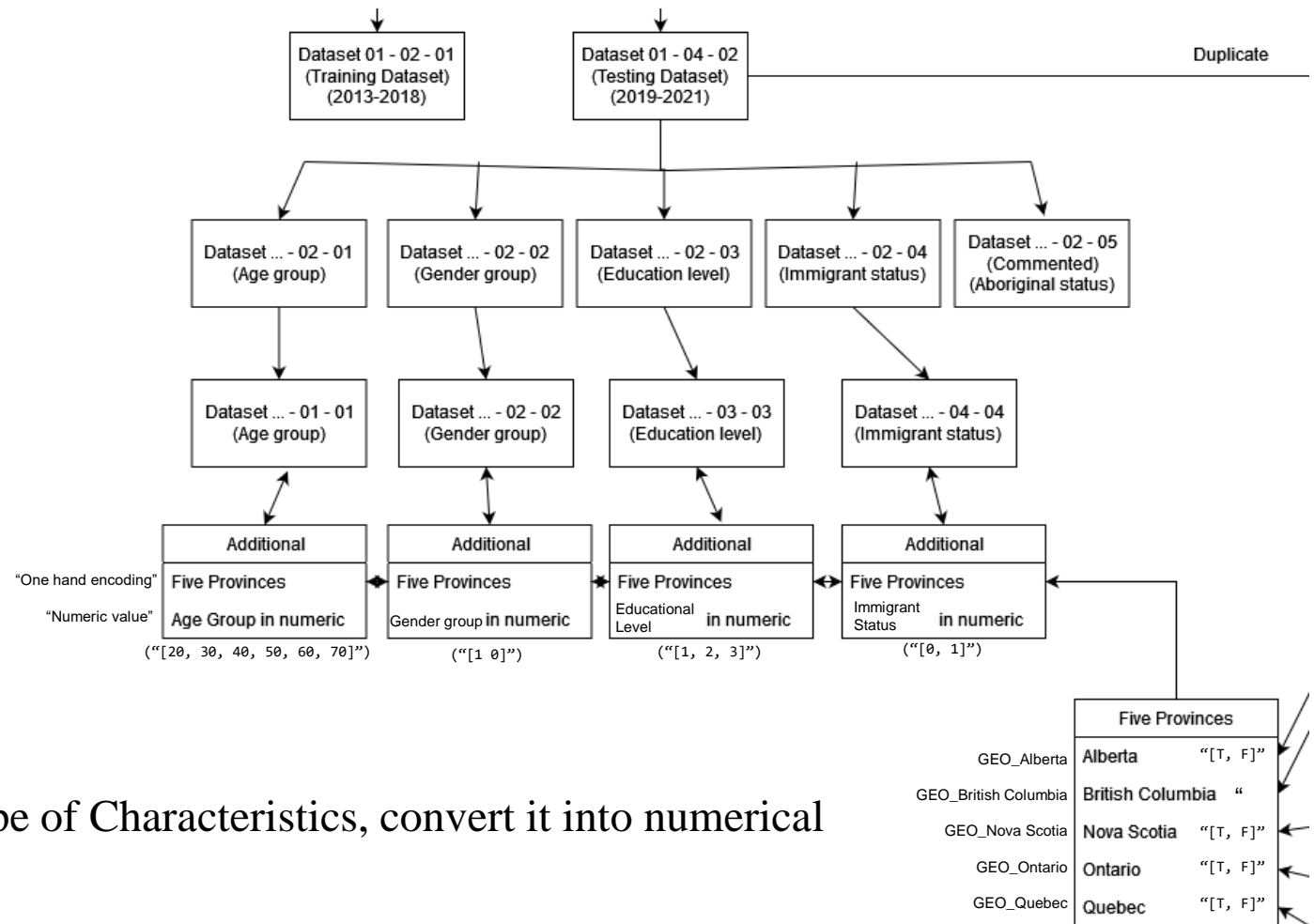
- Divide into 13 provinces provided from the previous divided dataset
- Select only five provinces to analysis and add these columns to previous divided dataset using one hand encoding, binary.
- Depend on the type of Characteristics, convert it into numerical values.
- Convert it into csv file and re-export back to final analysis in different script.

Divide by 'Provinces' columns

- Divided by provinces and select five of them to analysis.

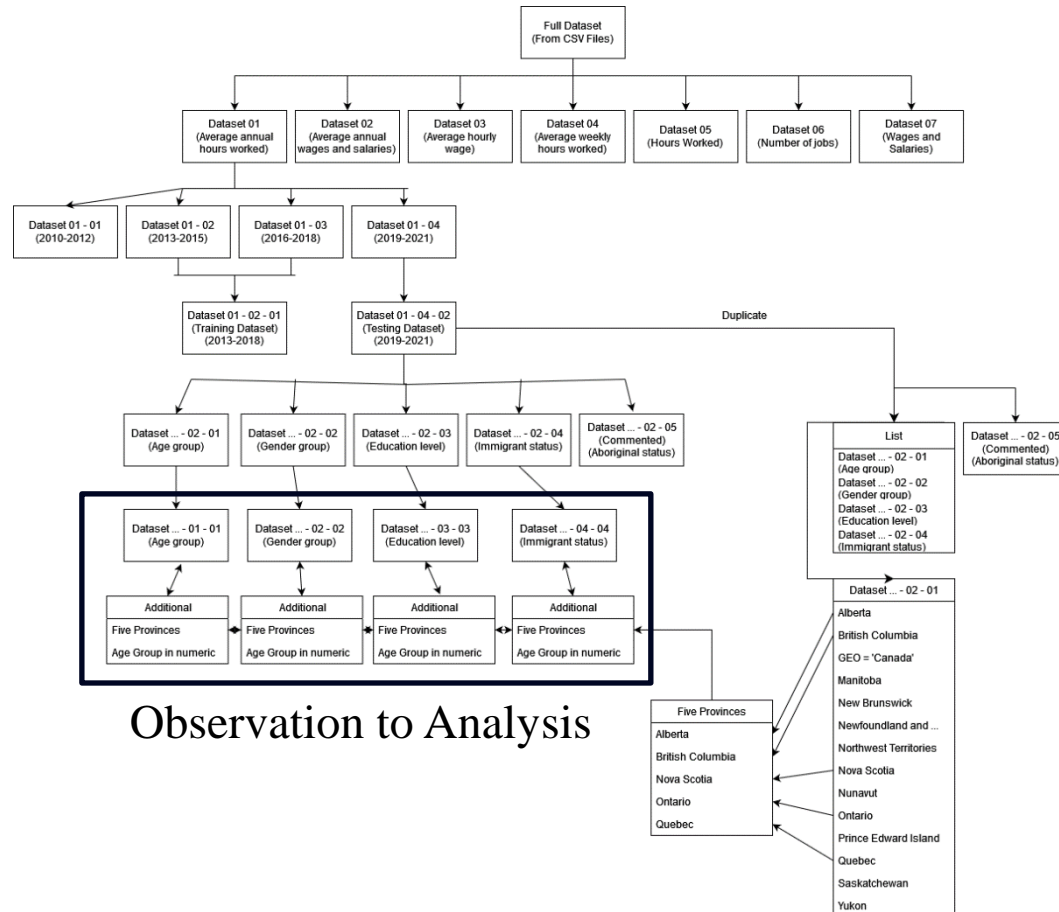


Divide by 'Provinces' columns



Depend on the type of Characteristics, convert it into numerical values.

Overall dataset division



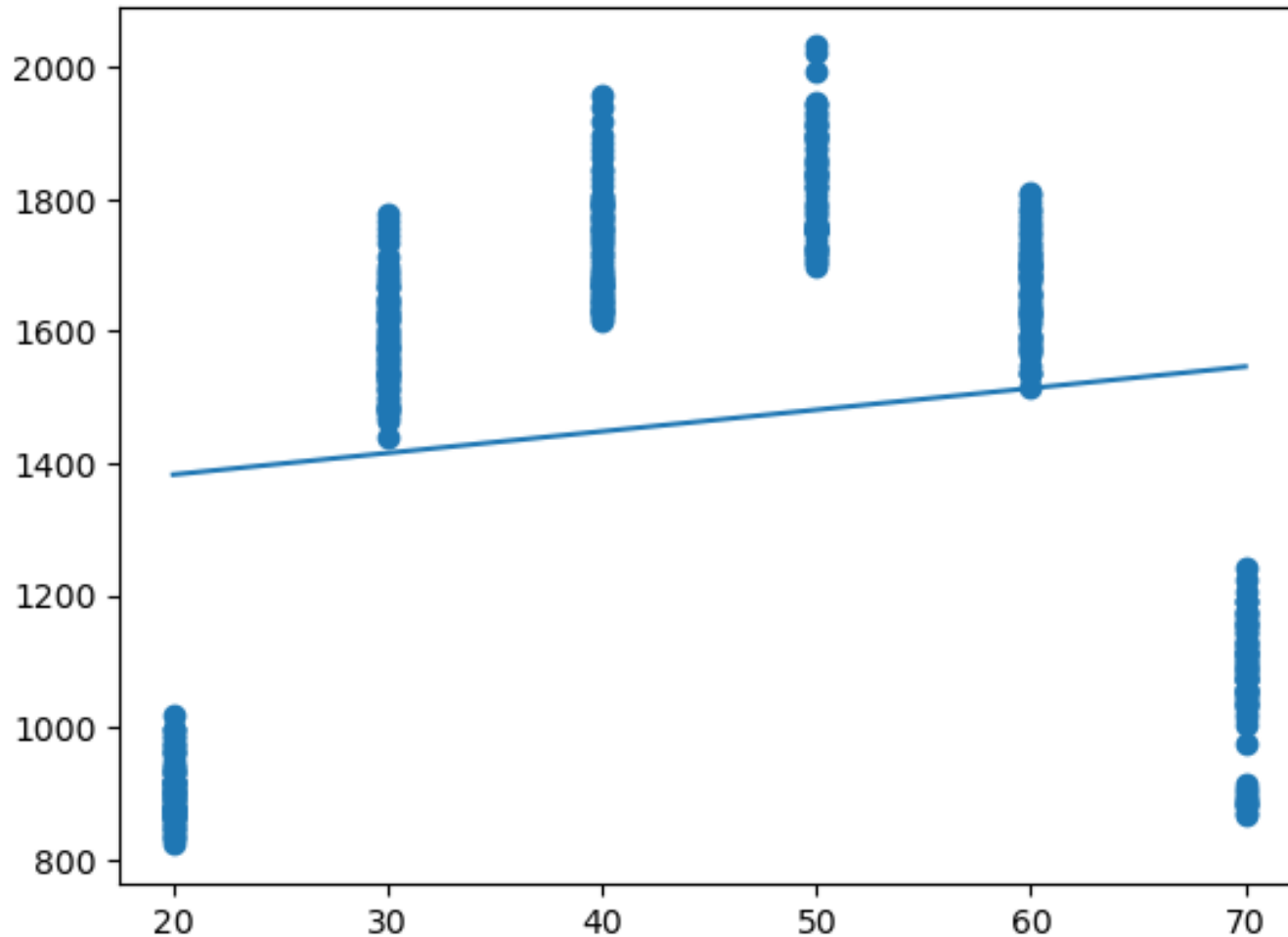
Observation to Analysis

Data Analysis

Overall Conclusion

- Employees who are older get to work higher and pay higher with exception of their 20s and 60s and older.
- Employees who have higher education work longer and pay higher except for trade who work longer term.
- Female employees work more and pay higher but as a long term, male employees work more and pay higher.
- Non-immigrant employees work more and pay higher but as a long term, Immigrant employees work more and pay higher.

Annual Hour Worked based on 'Age group'



Annual Hour Worked based on '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

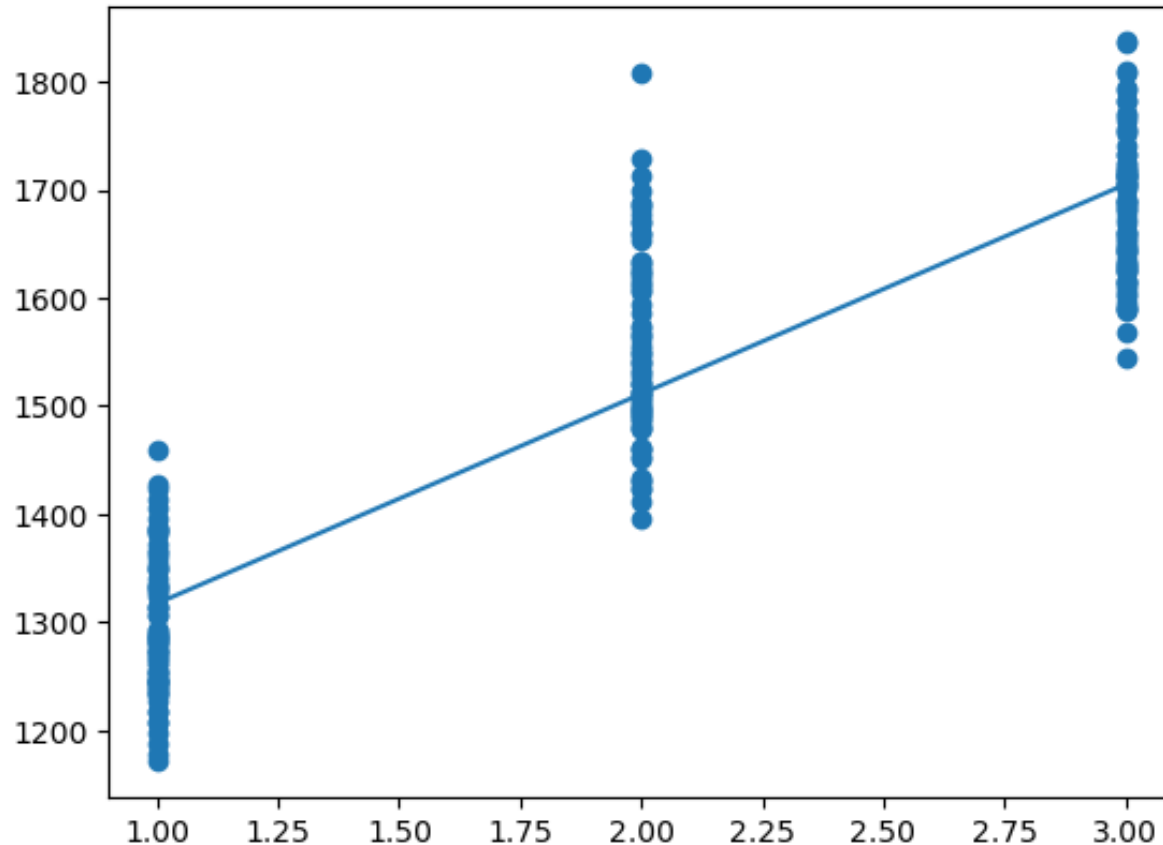
Standard Deviation : 356.3536590121646

Skewnewss : -0.5806984535776861

Total size : 450

Annual Hour Worked based on 'Education level'

Higher number = Higher education



Annual Hour Worked based on 'Education level'

```
final_testing_df_output_df_AvgAnnHrsWrk_ByEducation.csv
```

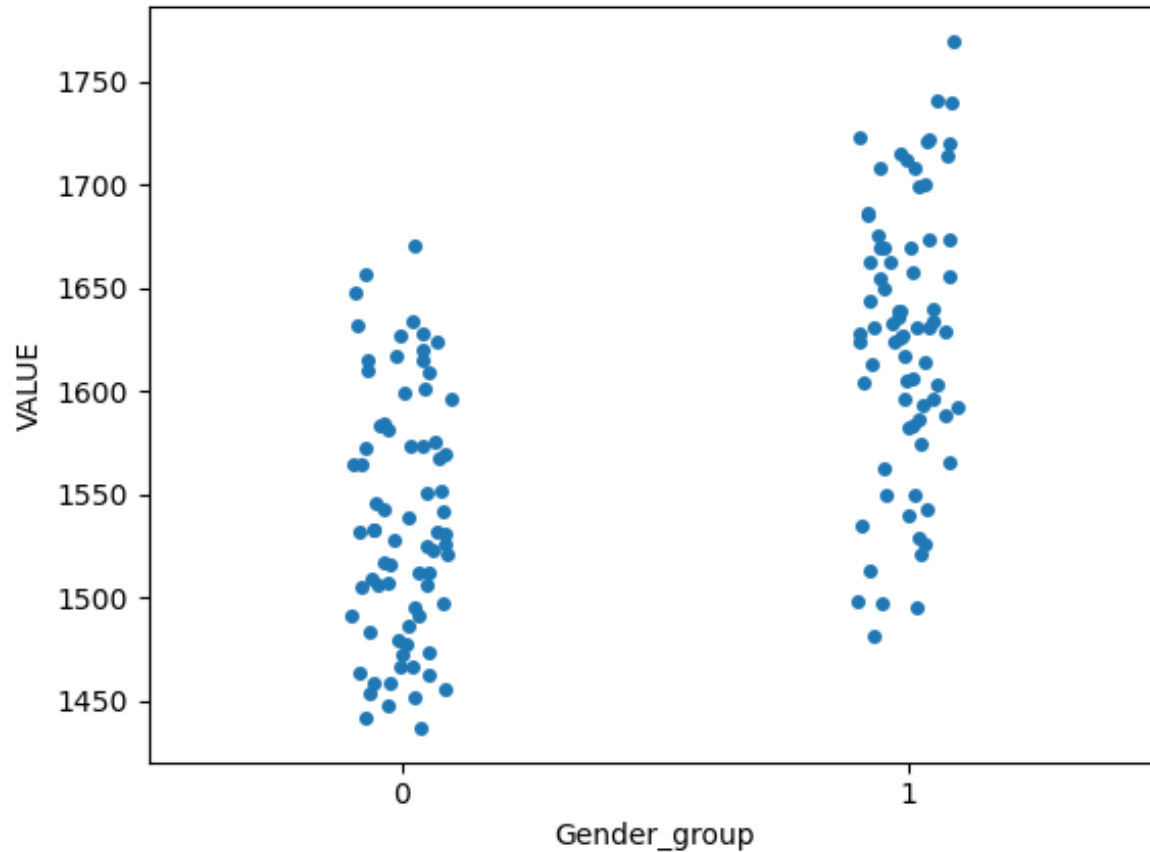
	sum	mean	amin	median	amax	\
Characteristics						
High school diploma and less	97223.0	1296.306667	1171.0	1285.0	1460.0	
Trade certificate	116506.0	1553.413333	1396.0	1538.0	1808.0	
University degree and higher	126305.0	1684.066667	1544.0	1684.0	1837.0	

```
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
```

Annual Hour Worked based on 'Gender group'

0 = Female

1 = Male

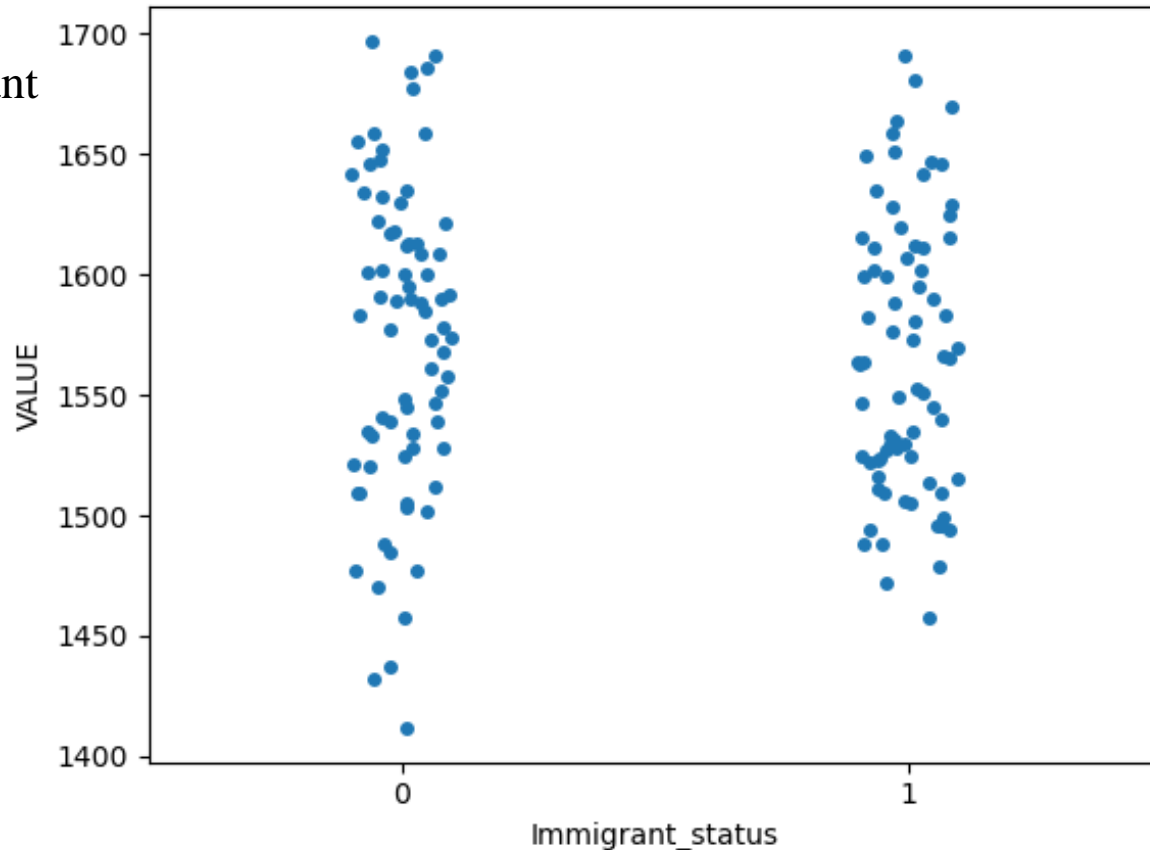


Annual Hour Worked based on 'Gender group'

```
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
```

Annual Hour Worked based on 'Immigrant group'

0 = Immigrant
1 = Non-immigrant



Annual Hour Worked based on 'Immigrant group'

```
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

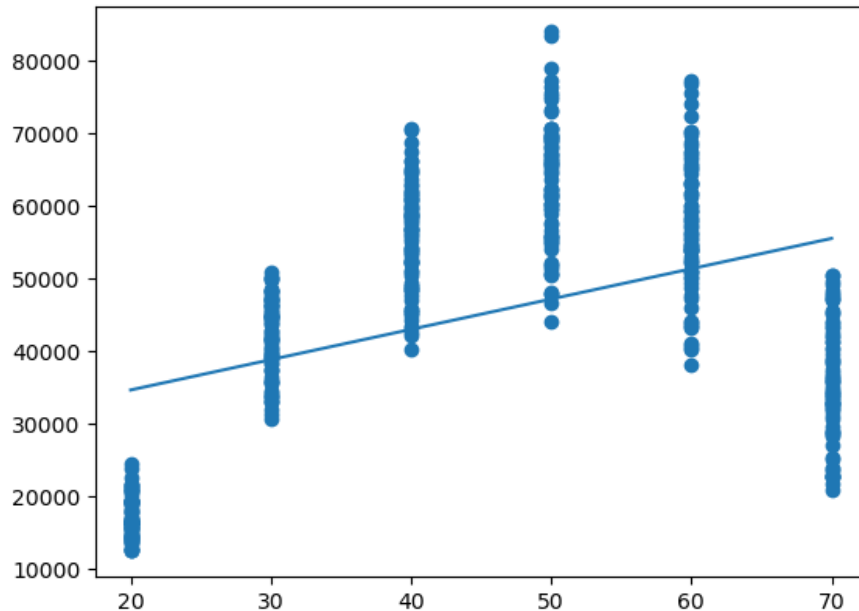
Min/median/max : 1412.0 / 1571.5 / 1697.0

Standard Deviation : 60.9140684352419

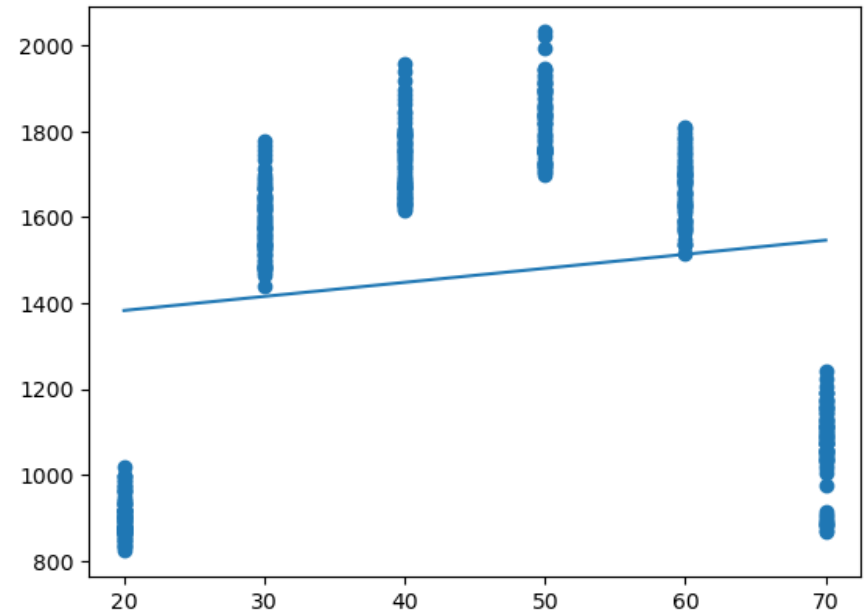
Skewnewss : -0.023261406680708266

Total size : 150

Average Annual Wage based on 'age group'



Based on Annual wages



Based on Annual hours worked

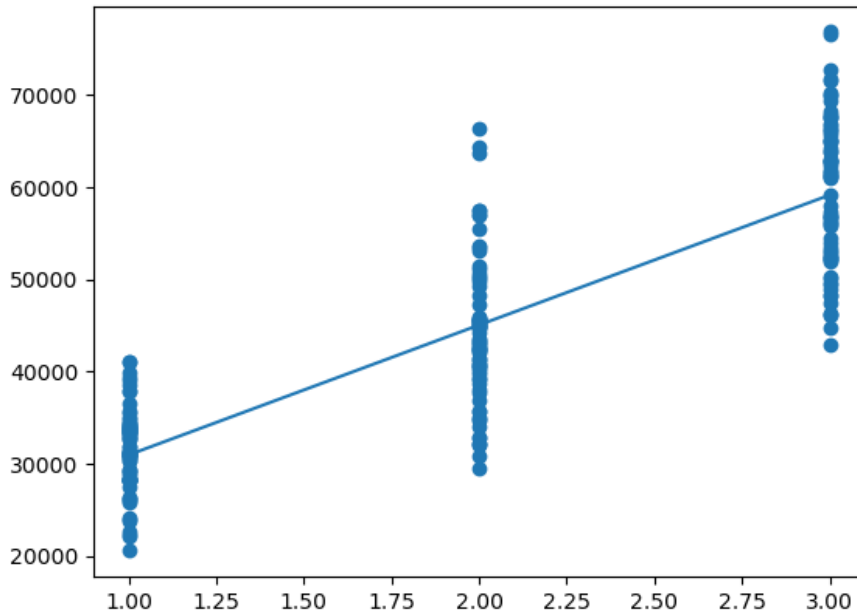
Average Annual Wage based on 'age group'

```
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

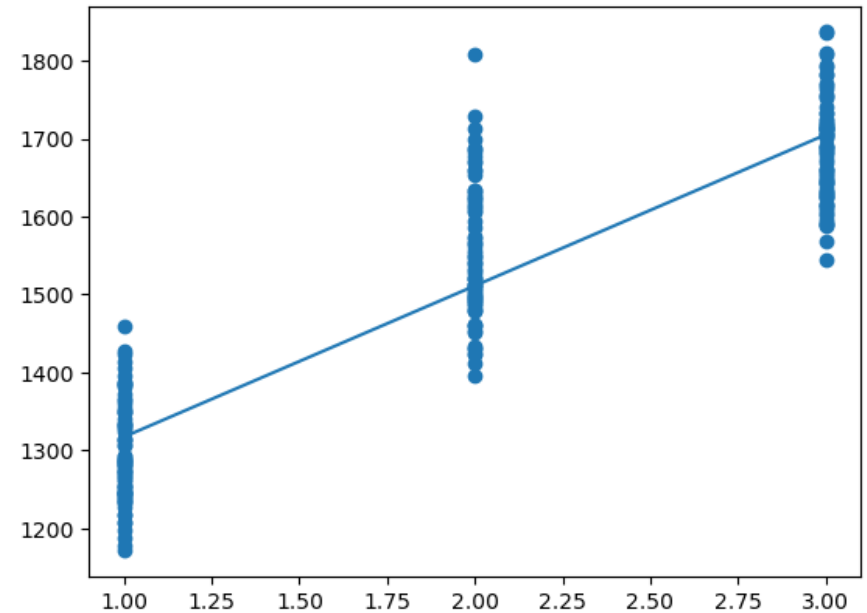
      size
Characteristics
15 to 24 years      75
25 to 34 years      75
35 to 44 years      75
45 to 54 years      75
55 to 64 years      75
65 years old and over  75
Overall,
Sum :  20271340.0
Mean :  45047.4222222222
Min/median/max : 12468.0 / 46985.0 / 83894.0
Standard Deviation : 17188.866523407123
Skewnewss :  -0.2534546294165646
Total size :  450
```

Average Annual Wages for 'education'

Higher number = Higher education



By Average Annual Wages



By Annual work hours

Average Annual Wages for 'education'

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	

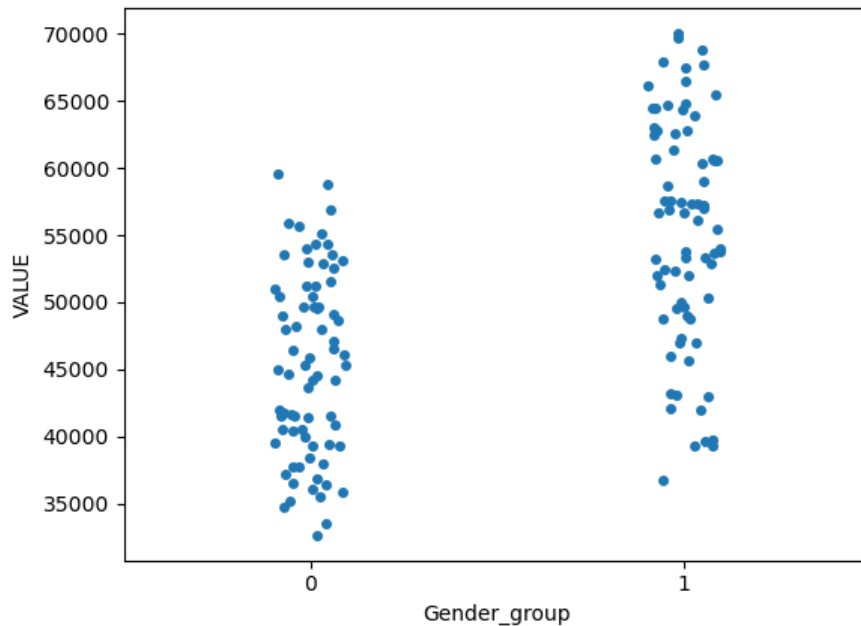
	amax	size
Characteristics		
High school diploma and less	41040.0	75
Trade certificate	66374.0	75
University degree and higher	76827.0	75

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

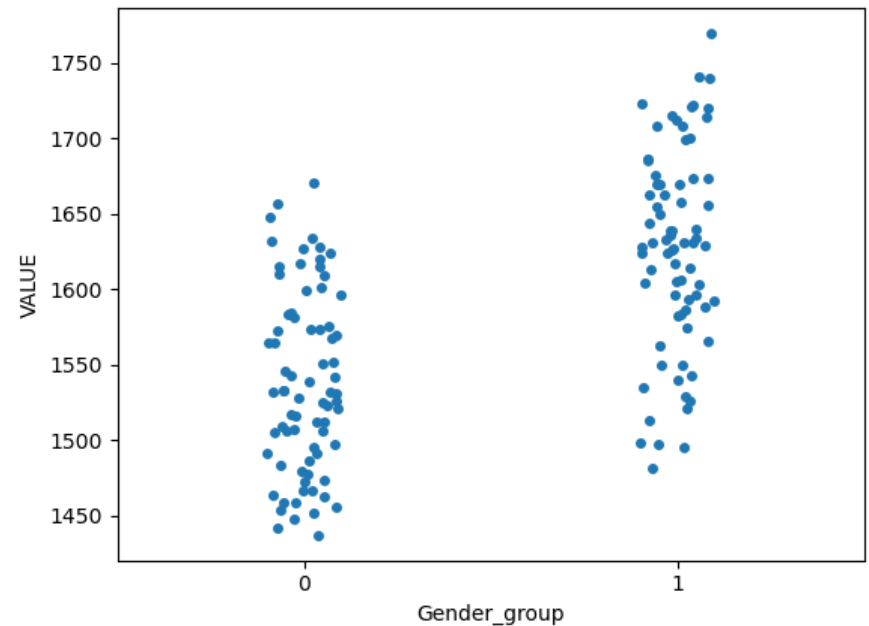
Average Annual Wages for 'gender'

0 = Female

1 = Male



By Average Annual Wages



By Annual Hour worked

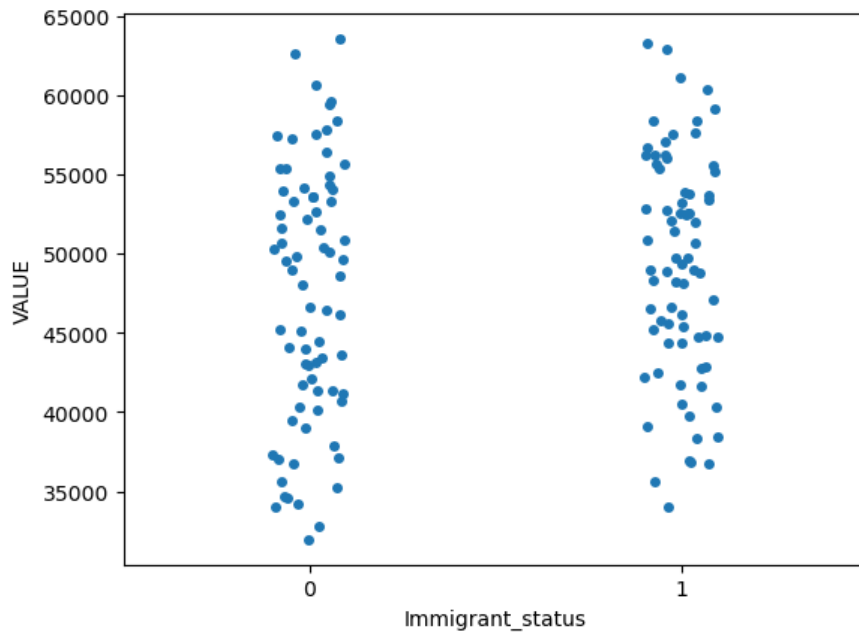
Average Annual Wages for 'gender'

```
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

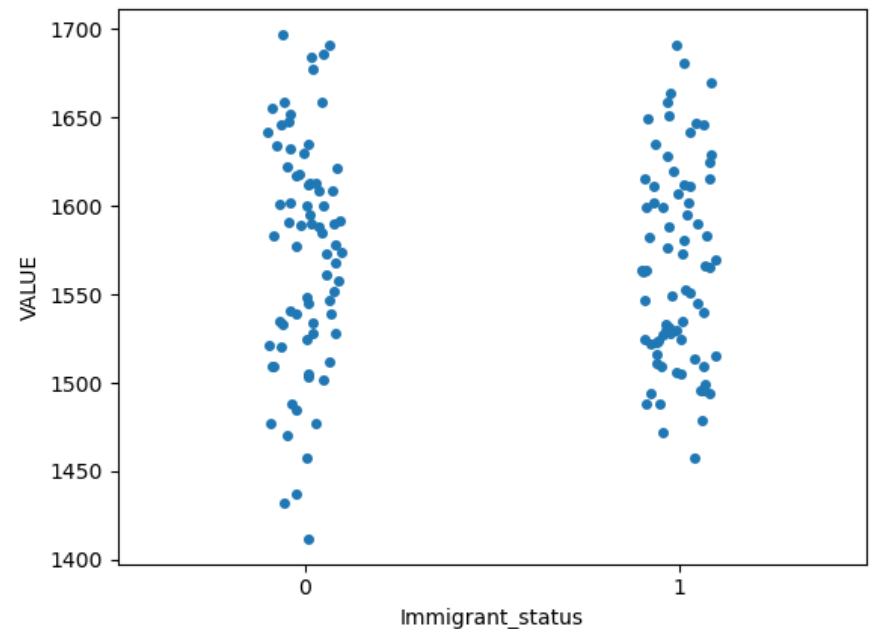
              size
Characteristics
Immigrant employees      75
Non-immigrant employees   75
Overall,
Sum :    7245948.0
Mean :    48306.32
Min/median/max : 31977.0 / 49016.0 / 63582.0
Standard Deviation :    7581.937589930426
Skewnewss :    -0.16449979628579625
Total size :    150
```

Average Annual Wages for 'immigrant'

0 = Immigrant
1 = Non-immigrant



By Average annual wages



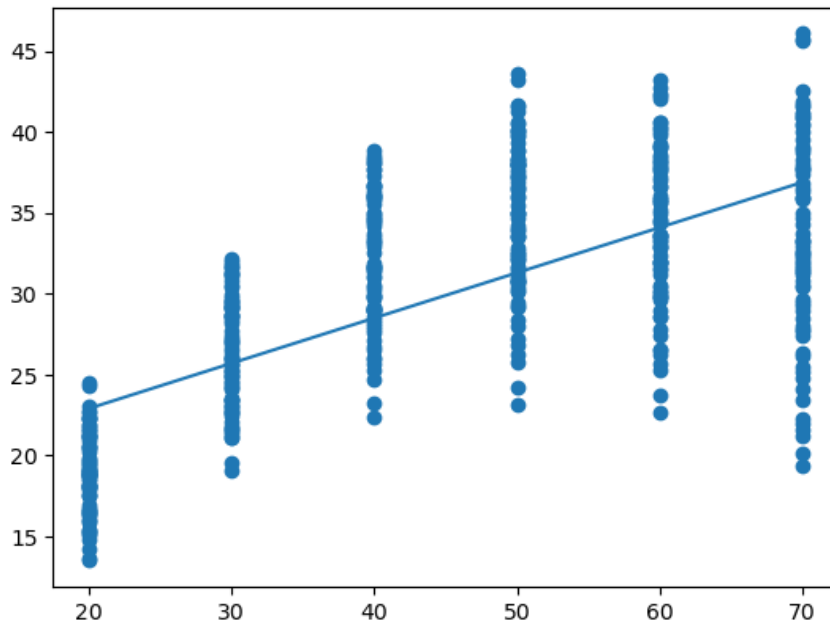
By Annual hours worked

Average Annual Wages for 'immigrant'

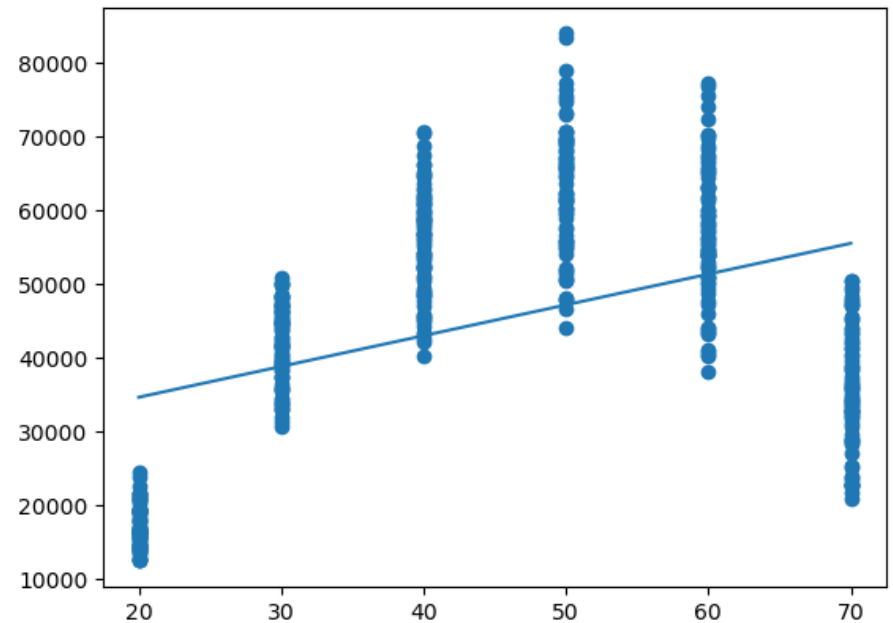
```
final_testing_df_output_df_AvgAnnWages_ByImmigrant.csv
              sum              mean              amin              median              amax  \
Characteristics
Immigrant employees      3554950.0  47399.333333  31977.0  48615.0  63582.0
Non-immigrant employees  3690998.0  49213.306667  34013.0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

Average hourly wages for 'age group'



By Hourly



By Annually

Average hourly wages for '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

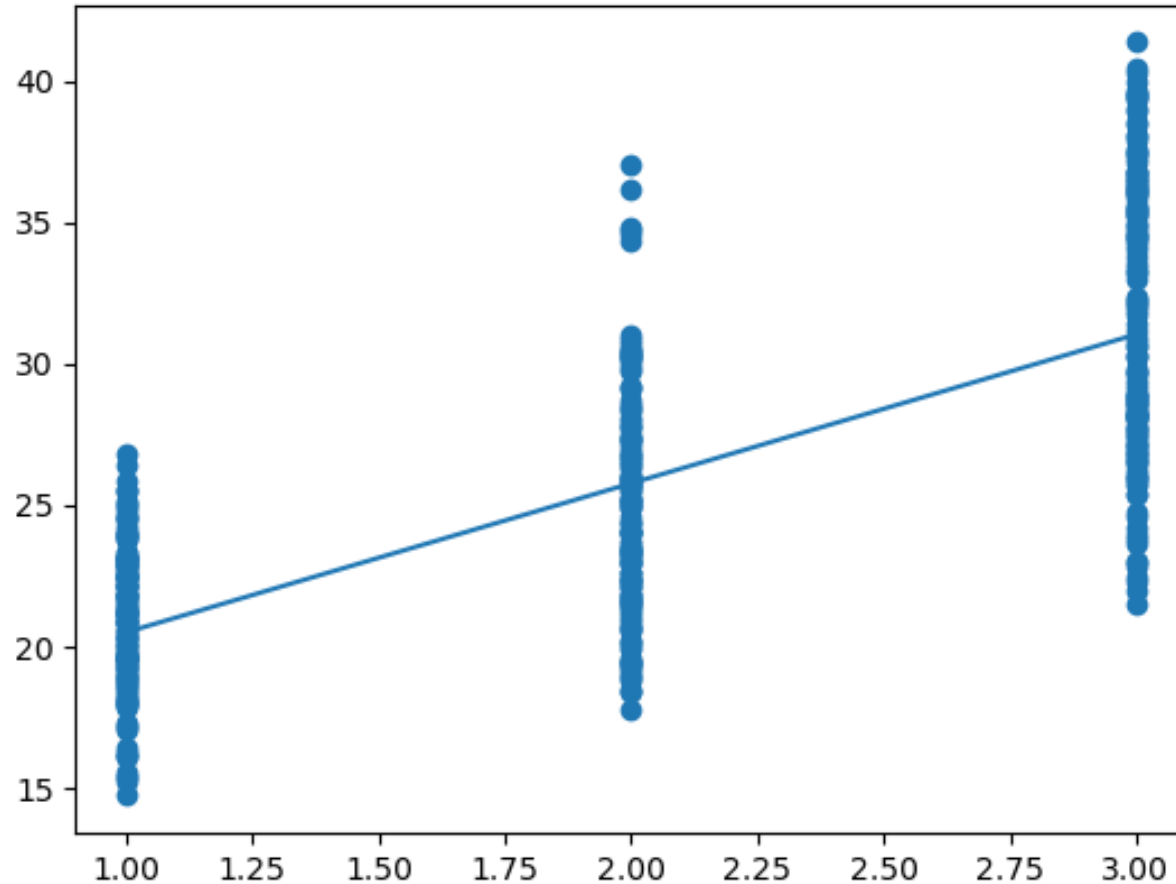
Standard Deviation : 7.202130786522336

Skewnewss : -0.2005972990040442

Total size : 450

Average hourly wages for 'Education'

Higher number = Higher education



Used training set instead due to debugging

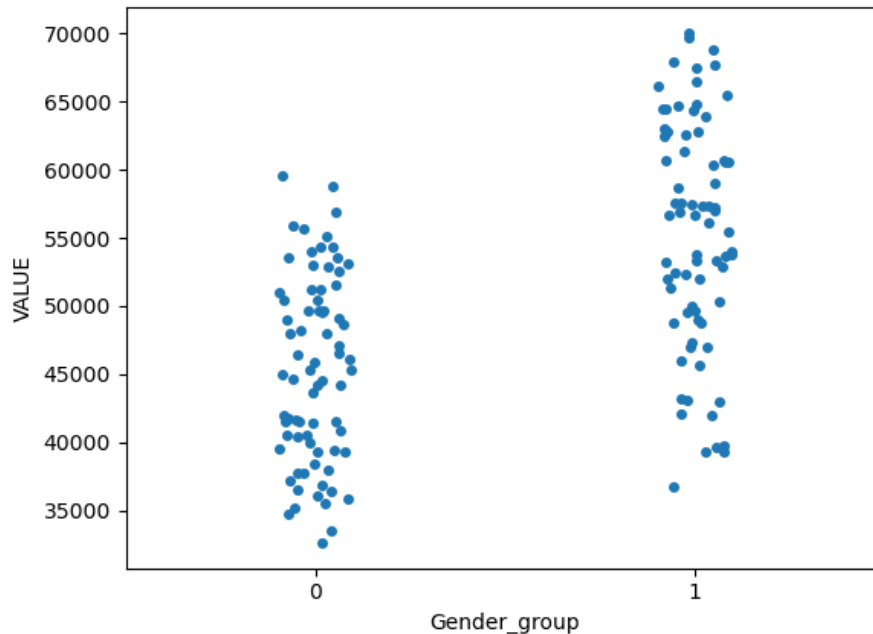
Average hourly wages for 'Education'

```
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 :  29.310933333333333
Min/median/max : 16.56 / 28.34 / 45.47
Standard Deviation :  6.1840297915050675
Skewnewss :  0.5213937099086743
Total size :  225
```

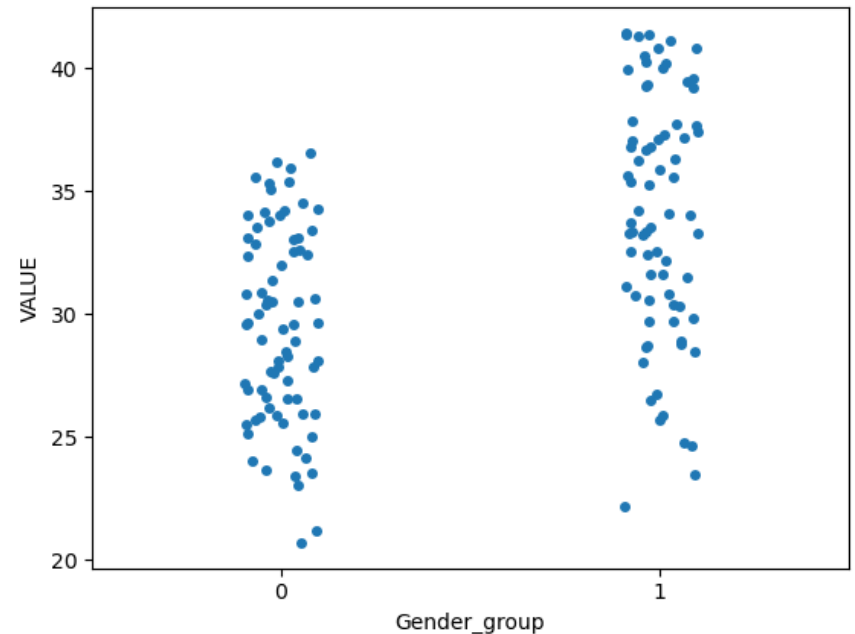
Average hourly wages for 'Gender group'

0 = Female

1 = Male



By Annually

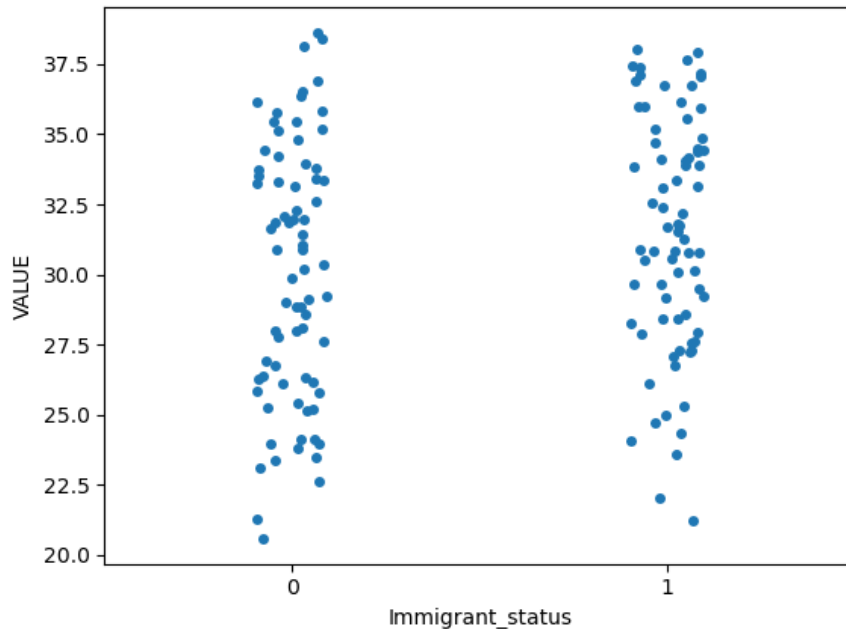


By hourly wages

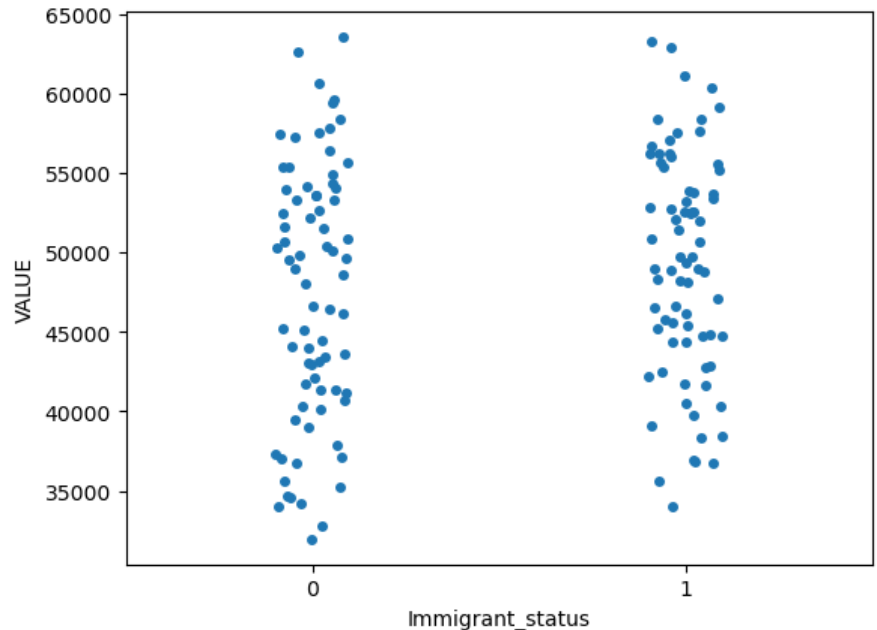
Average hourly wages for immigrant

0 = Immigrant

1 = Non-immigrant



By hourly

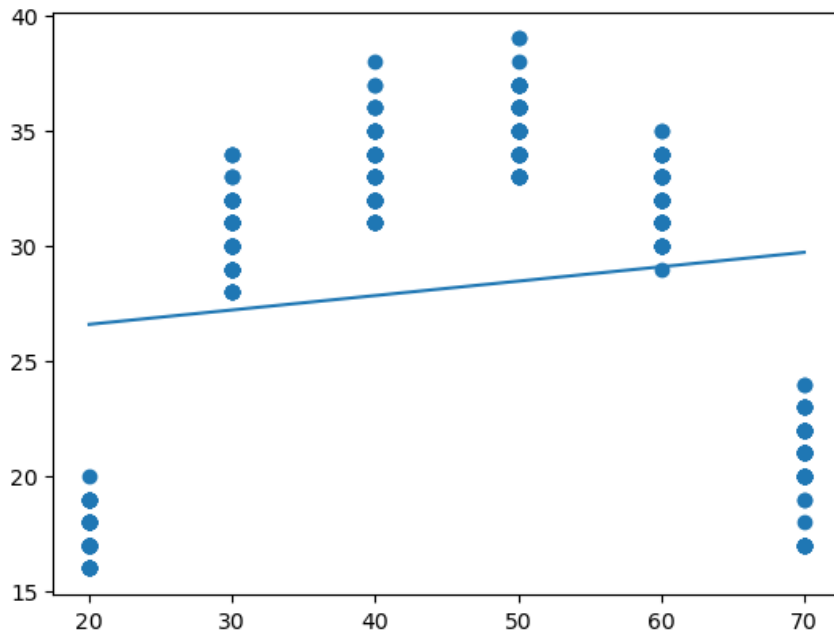


By Annually

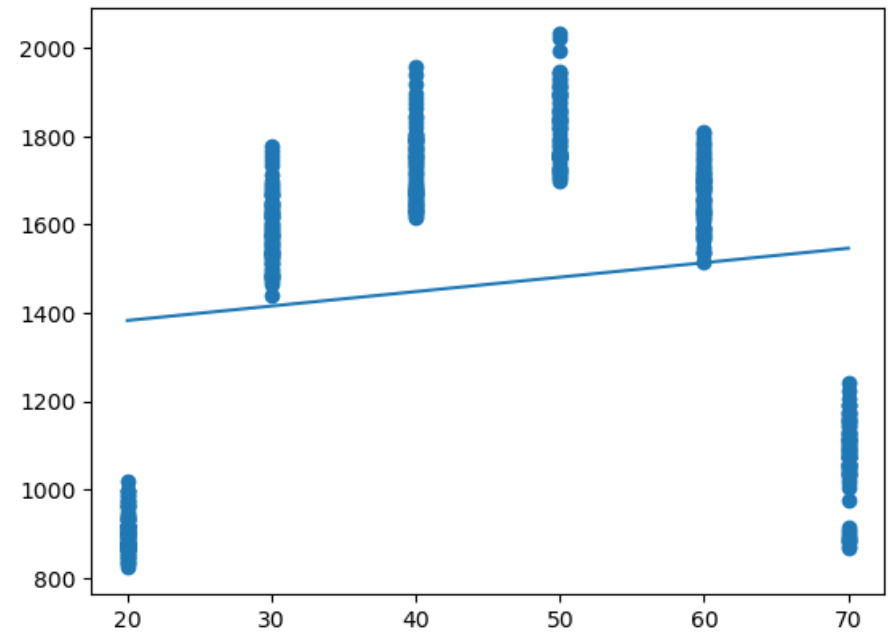
Average hourly wages for immigrant

```
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 :  4610.94000000000005
Mean :  30.739600000000003
Min/median/max : 20.6 / 30.905 / 38.63
Standard Deviation :  4.419609165224154
Skewnewss :  -0.21241877283543747
Total size :  150
```


Average weekly hours worked by 'Age group'



By Weekly



By Annually

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

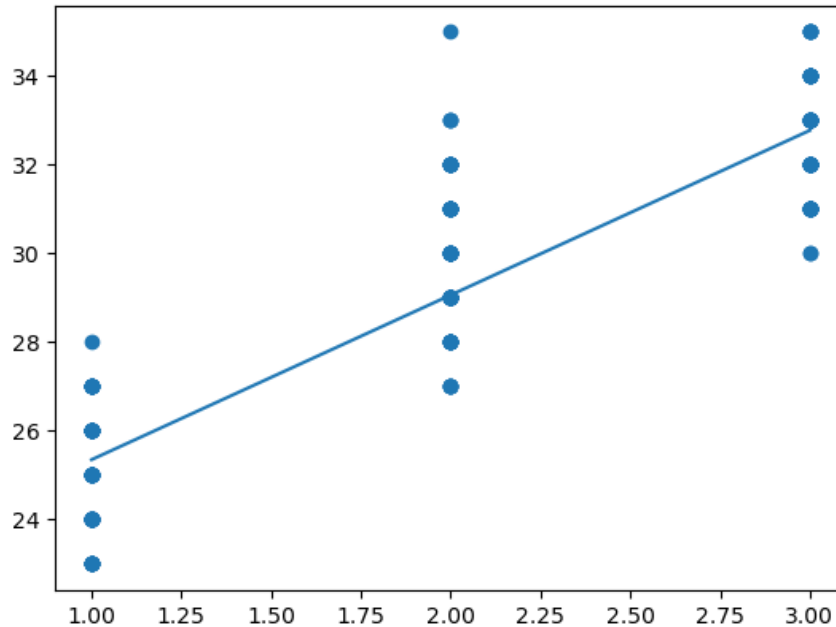
Standard Deviation : 6.8352712699155695

Skewnewss : -0.5809323008907273

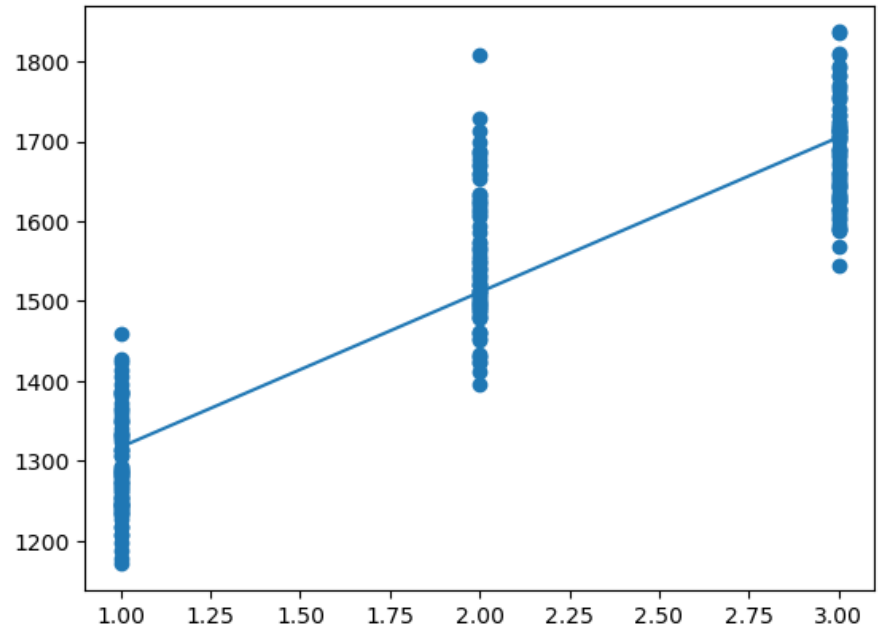
Total size : 450

Average weekly hours worked based on 'Education group'

Higher number = Higher education



By weekly



By Annually

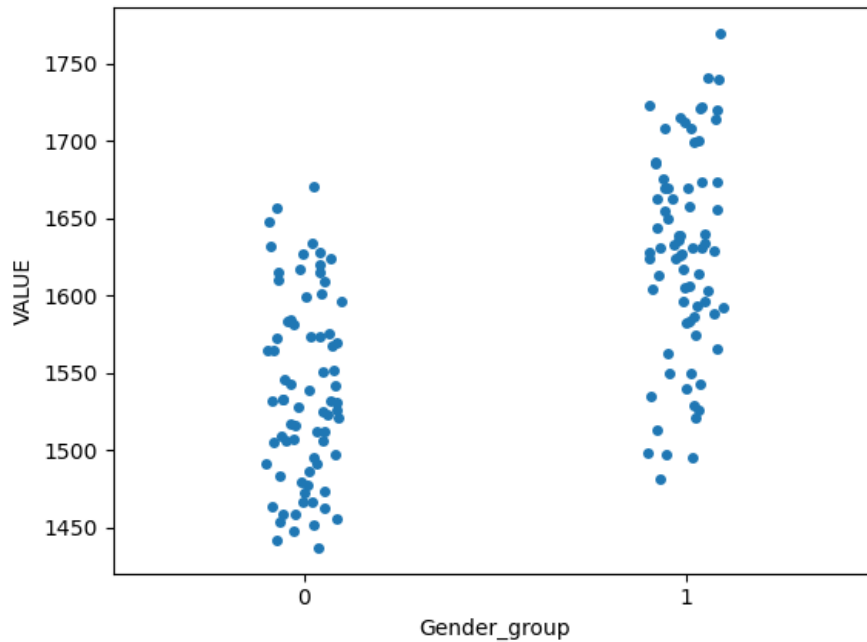
Average weekly hours worked based on 'Education group'

```
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
```

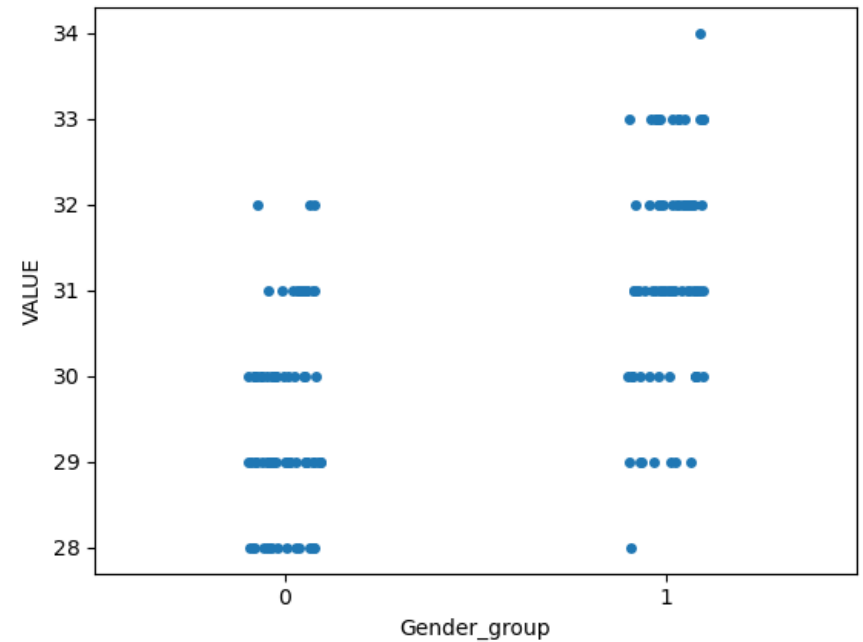
Average weekly work hours for on 'Gender group'

0 = Female

1 = Male



By annually



By Weekly

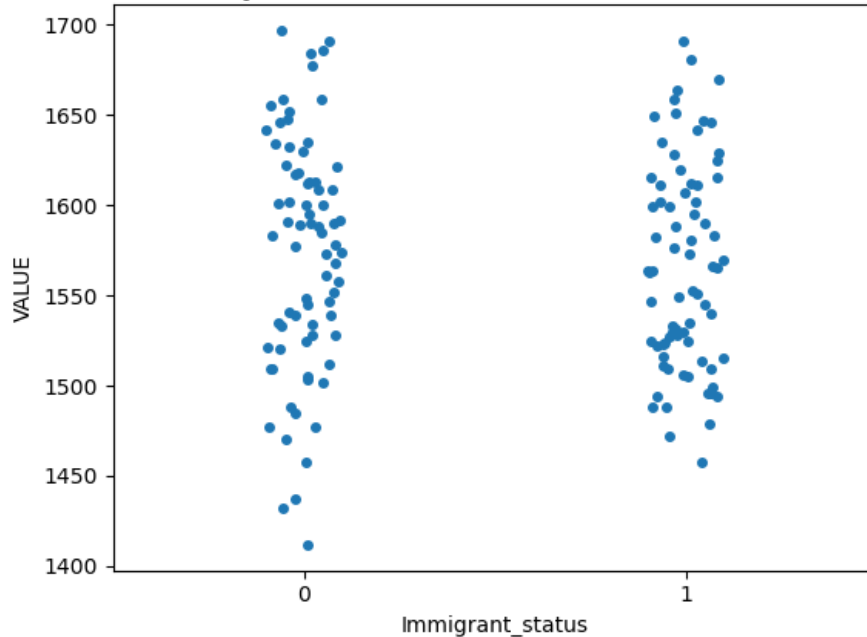
Average weekly work hours for on 'Gender group'

```
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
```

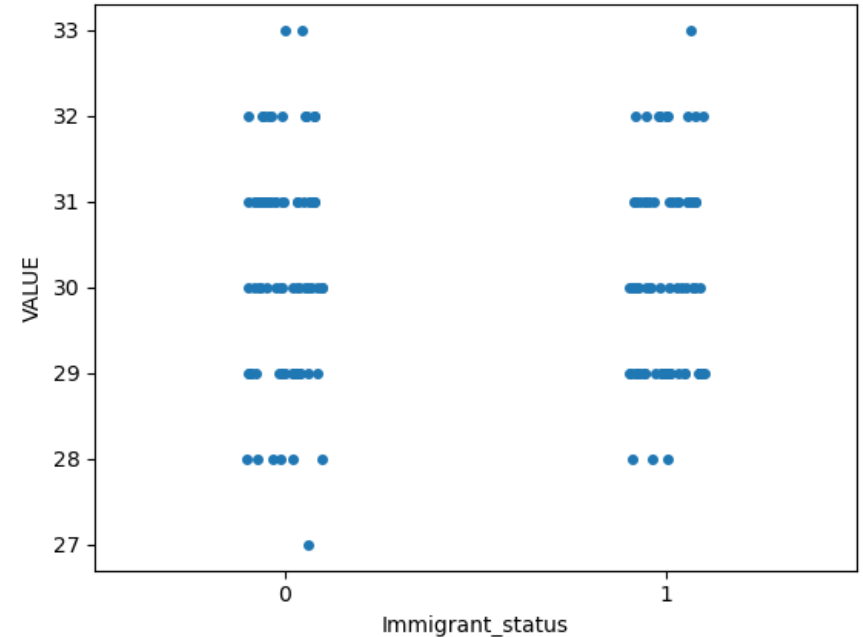
Average weekly Hour Worked based on 'Immigrant group'

0 = Immigrant

1 = Non-immigrant



By annually

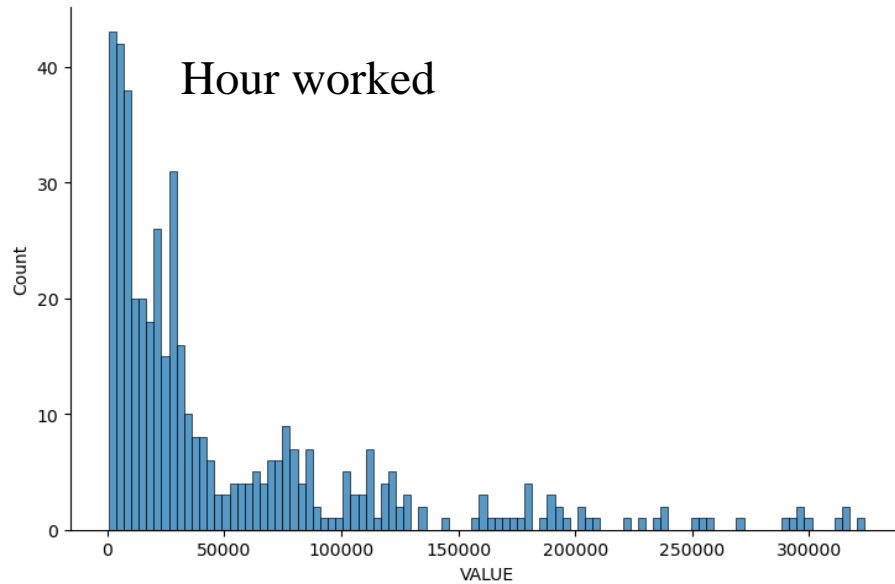


By weekly

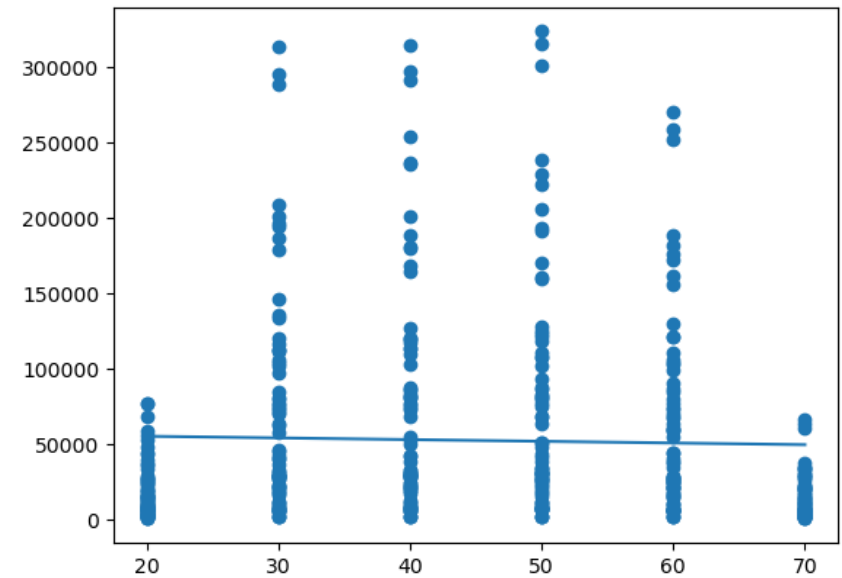
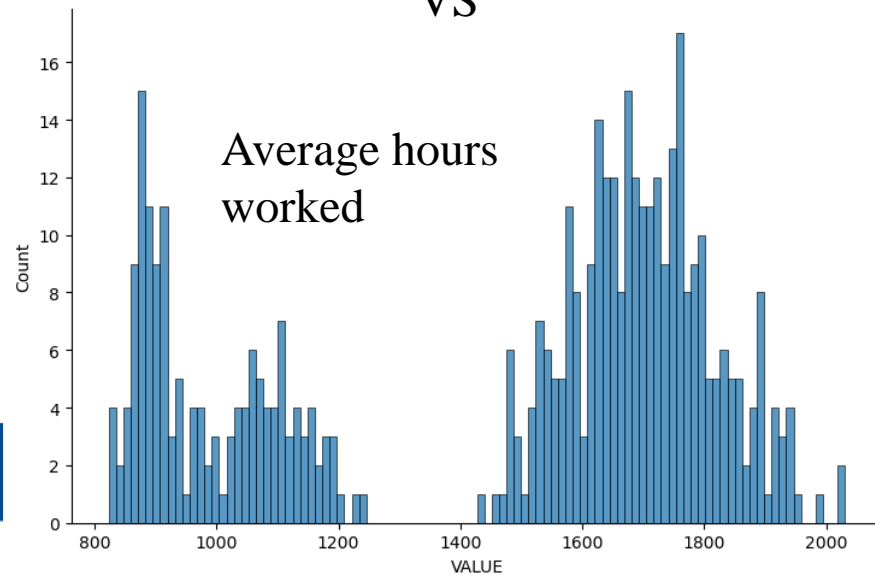
Average weekly Hour Worked based on 'Immigrant group'

```
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
```


Hours worked in 'Age group'



VS



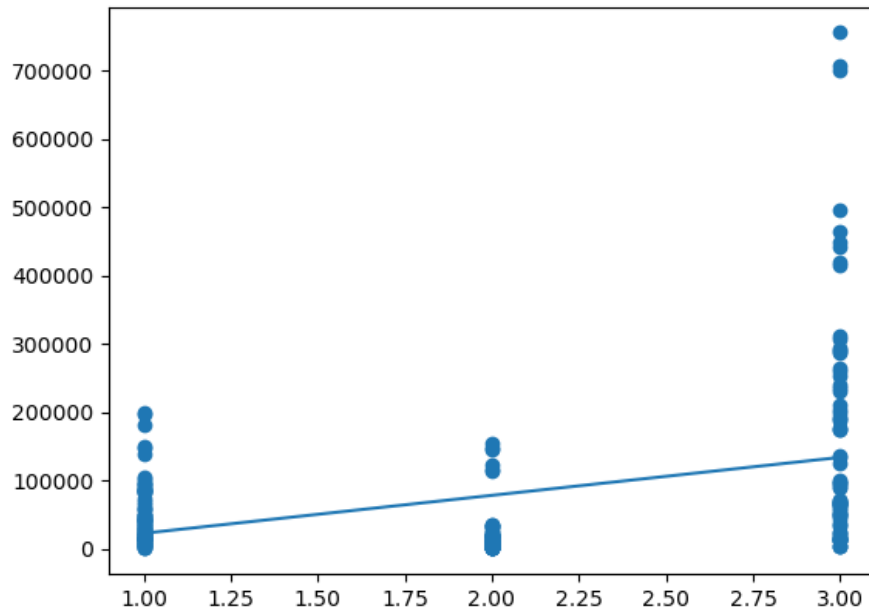
Hours worked in 'Age group'

```
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

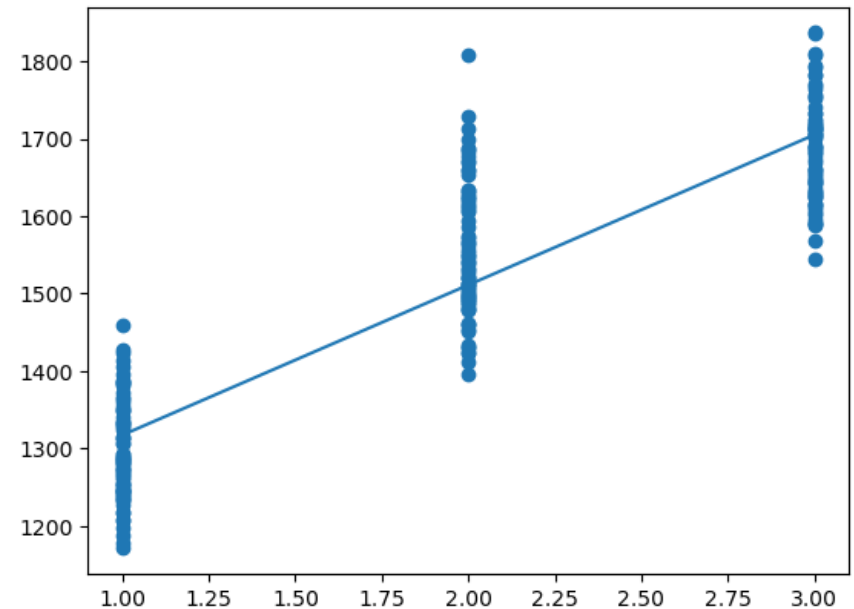
      size
Characteristics
15 to 24 years      75
25 to 34 years      75
35 to 44 years      75
45 to 54 years      75
55 to 64 years      75
65 years old and over 75
Overall,
Sum : 23545750.0
Mean : 52323.88888888889
Min/median/max : 361.0 / 26432.5 / 323859.0
Standard Deviation : 65504.80018241995
Skewnewss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

Hours Worked by 'Education level'

Higher number = Higher education



For hours worked



For annually hours worked

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

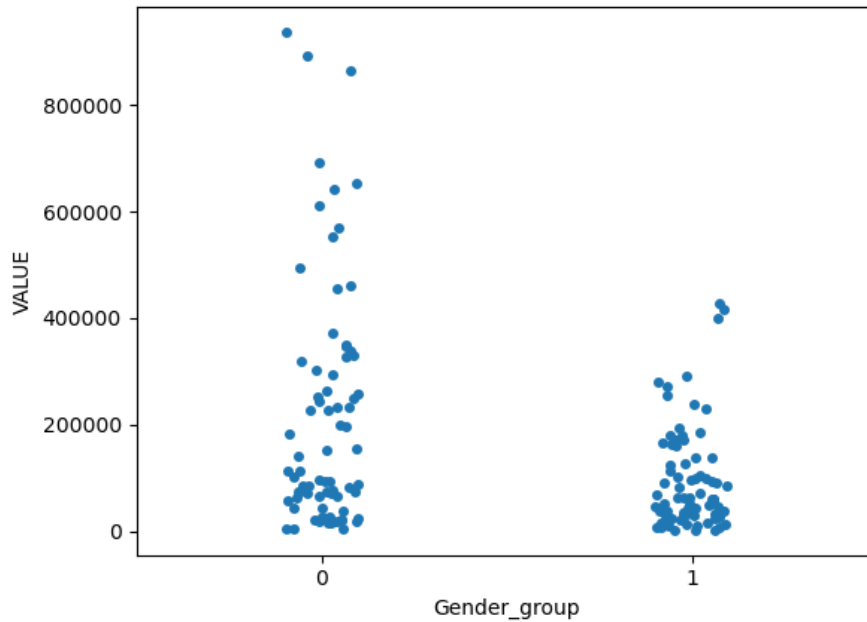
Skewnewss : 3.0205628183947755

Total size : 225

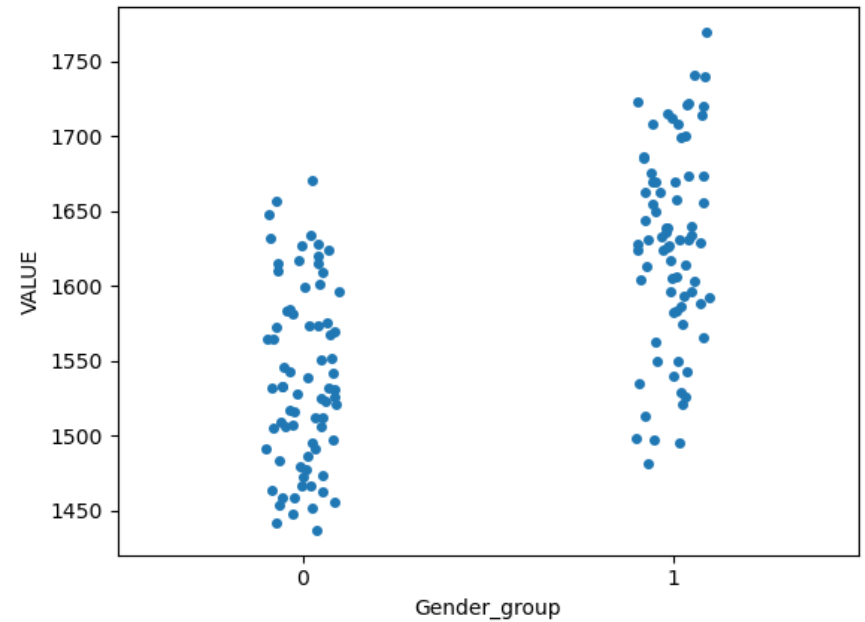
Hour worked for 'By Gender'

0 = Female

1 = Male



By hour worked



By annual hour worked

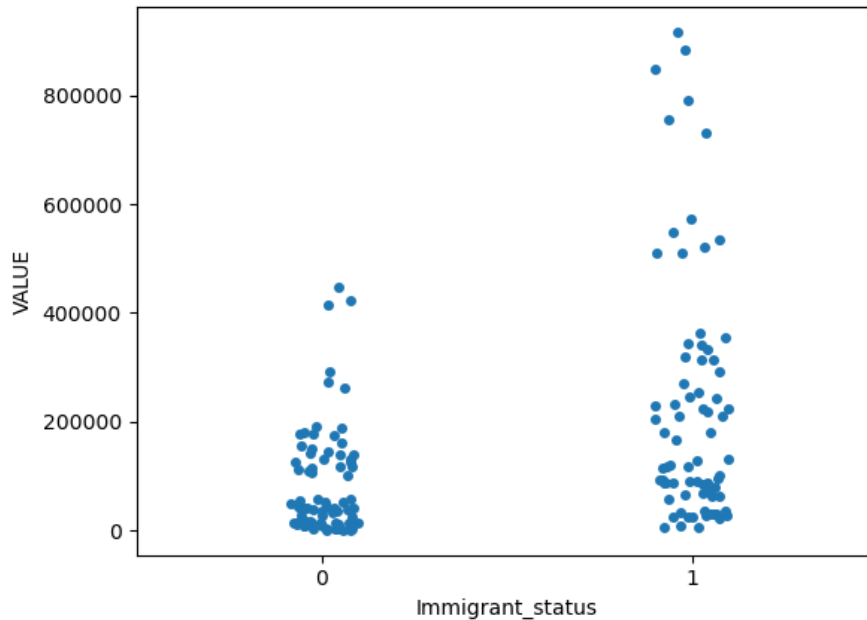
Hour worked for 'By Gender'

```
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
```

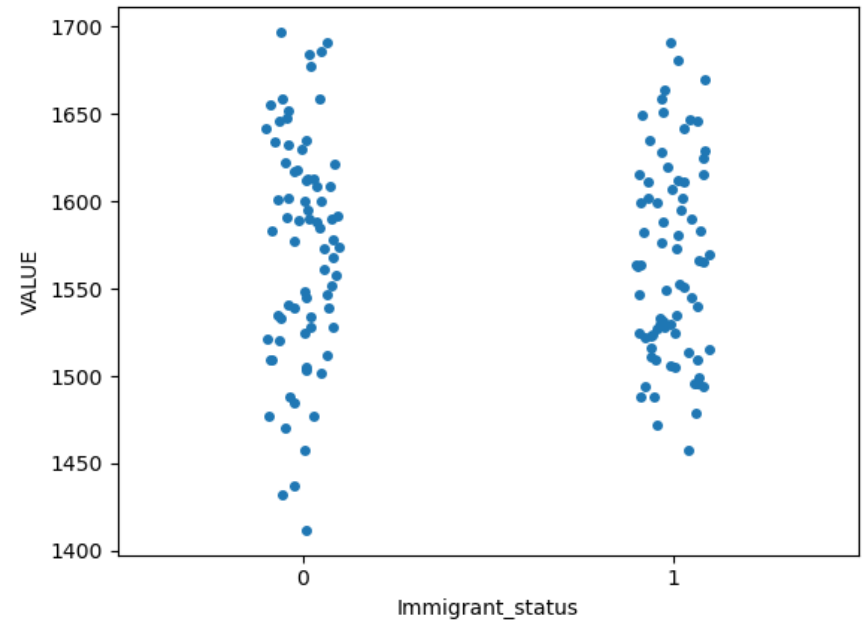
Hour worked by 'Immigrant status'

0 = Immigrant

1 = Non-immigrant



By hour worked



By annual hour worked

Hour 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

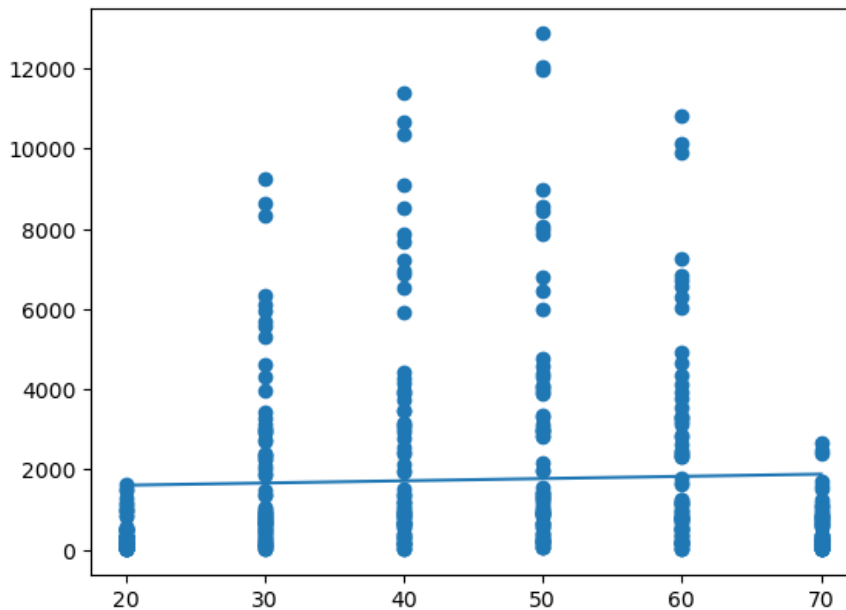
Min/median/max : 550.0 / 92441.5 / 917521.0

Standard Deviation : 188873.98307114118

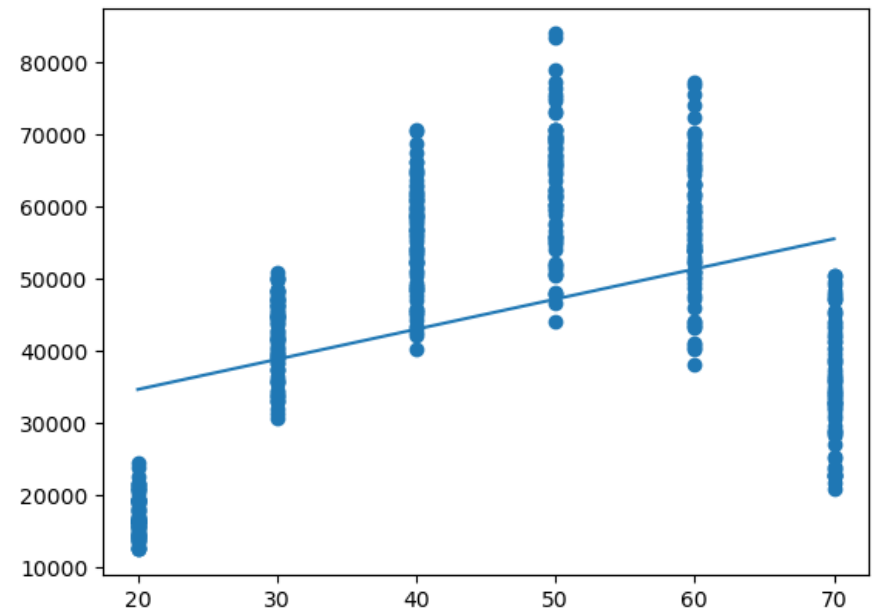
Skewnewss : 2.0992731555277357

Total size : 150

Wages and Salaries for Age group



By wages and salaries



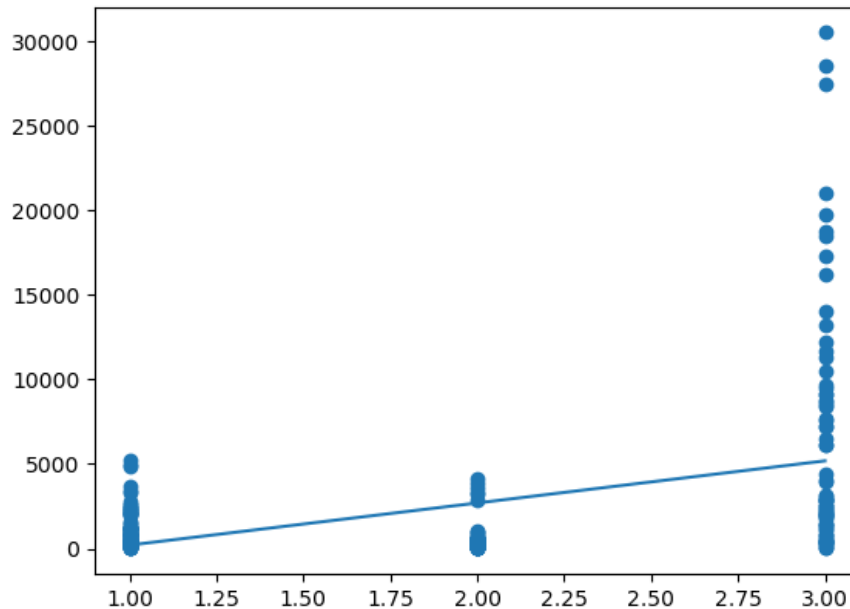
By average annual wages

Wages and Salaries for Age group

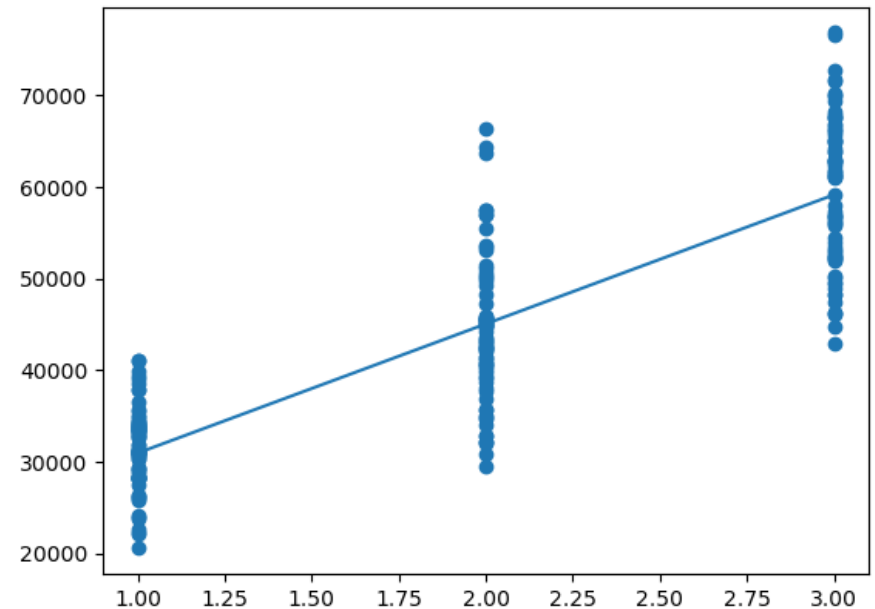
```
final_testing_df_output_df_WagesAndSalaries_ByAge.csv
              sum          mean  amin  median          amax  size
Characteristics
15 to 24 years      27272.0    363.626667    8.0    188.0    1622.0    75
25 to 34 years     147099.0   1961.320000   32.0    921.0    9270.0    75
35 to 44 years     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
```

Wages and Salaries by Education

Higher number = Higher education



By wages and salaries



By average annual wages

Wages and Salaries by Education

```
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	

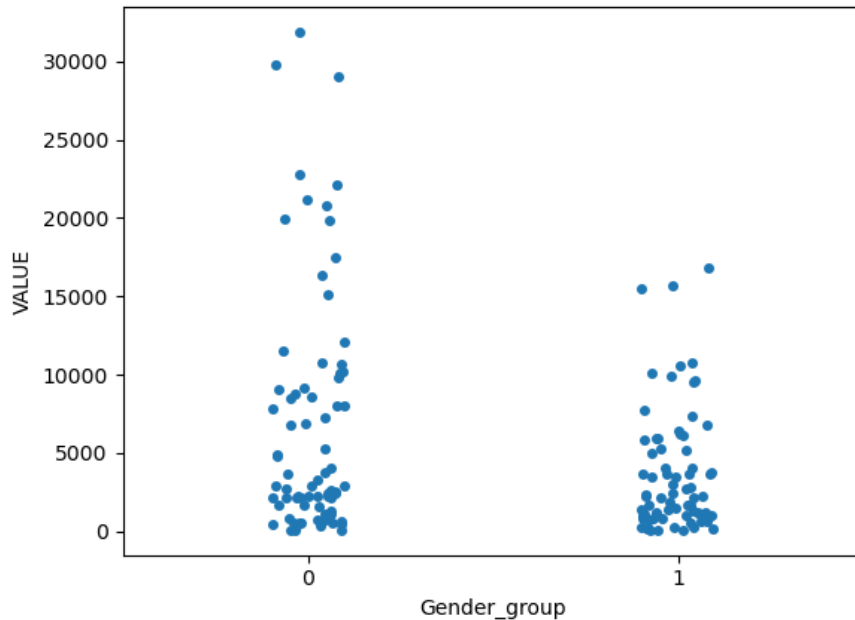
size

```
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
```

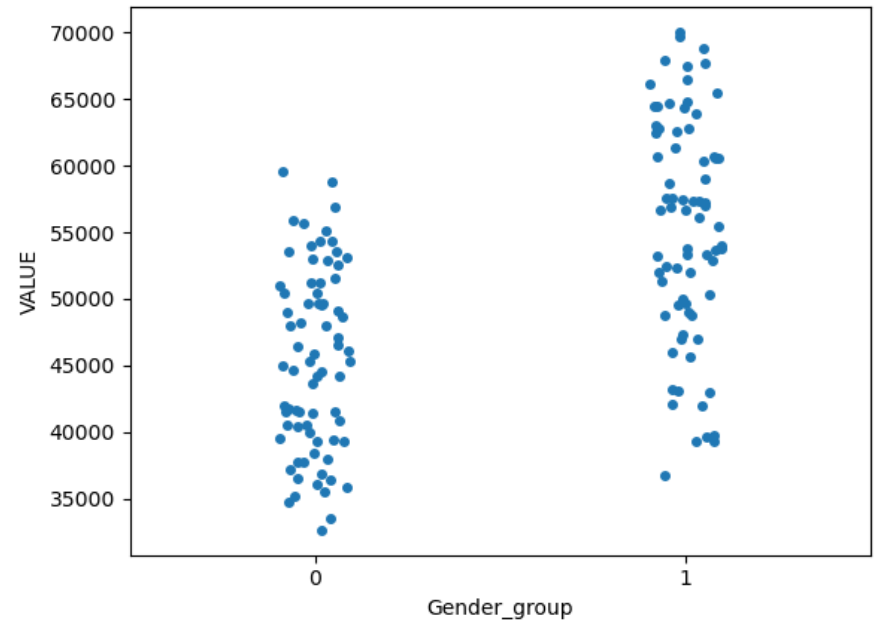
Wages and Salaries by Gender

0 = Female

1 = Male



By wages and salaries



By average annual wages

Wages and Salaries by Gender

```
final_testing_df_output_df_WagesAndSalaries_ByGender.csv
```

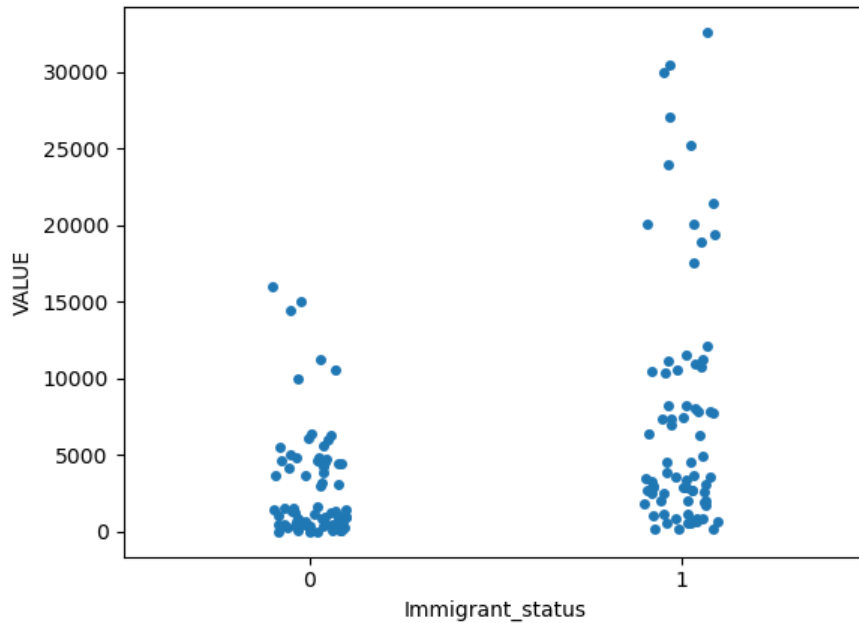
	sum	mean	amin	median	amax	size
Characteristics						
Female employees	516398.0	6885.306667	94.0	2932.0	31892.0	75
Male employees	270217.0	3602.893333	84.0	2176.0	16776.0	75

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

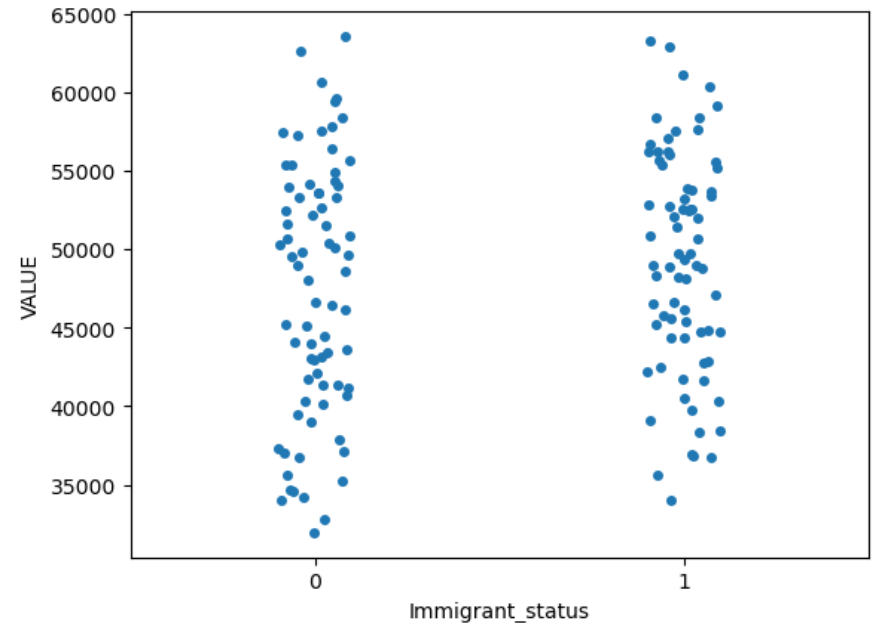
Wages and Salaries by Immigrant

0 = Immigrant

1 = Non-immigrant



By wages and salaries



By average annual wages

Wages and Salaries by Immigrant

```
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

Conclusion

Conclusion

- Employees who are younger and have at least a trade certification had potential to working for non-profit organizations.
- Although 20s make a lot less works at same time make less as get older and have more educations, there's significant wages increased and probably work more.
- Employees who are older than 60s are encouraged to avoid working for non-profit organizations as they get less salaries but with less working hours.
- Male and immigrant employees will earn more work with more wages more than a week at non-profit organizations.

Limitation

- The dataset given from the site is confused and complicated.
- Some indicators given are almost impossible to compare.
- Some indicators inside dataset are useless or repetitive.
- The dataset given is not clear for me to understand.



Challenge

- There wasn't enough time to fully analysis as most of my time spend here is mostly debugging or data preparation.
- The dataset that was provided are all categorical exception to the VALUE indicator.
 - Converting some indicators to numeric was required.
- Some indicators and columns in the dataset are impossible to understand and compare for time given.
- As more data preparation occurred, lot of datasets with lot of scripts occur.



Thank you