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

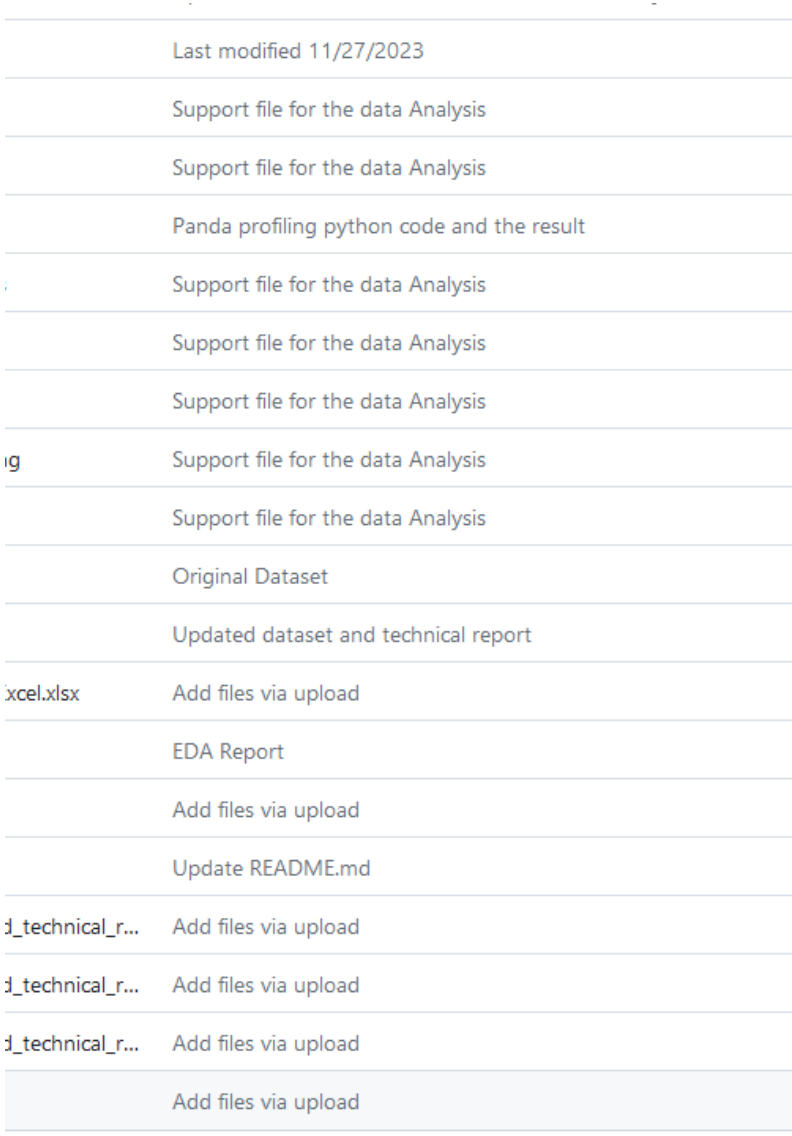
Ryerson University

Github Link

- Link

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

- Contain python script and output result in html
- Script contains in following:
  - ‘data_analysis_categorized_technical report.iypnb’.
  - ‘final_report’ directory contains final analysis.
  - Use anaconda 2022.05 with Python 3.8.18 with “Numpy”, “Pandas”, “Panda-profiling”, “Fitter”



Ryerson University

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

Initial data processing

- Removal of the unnecessary columns

```
<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
dtypes: float64(3), int64(4), object(10)
memory usage: 13.7+ MB
None
```

```
<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
```

Initial data processing

- Filter “Null” observations

	percent_in_na	num_of_na
VALUE	2.857143	3024

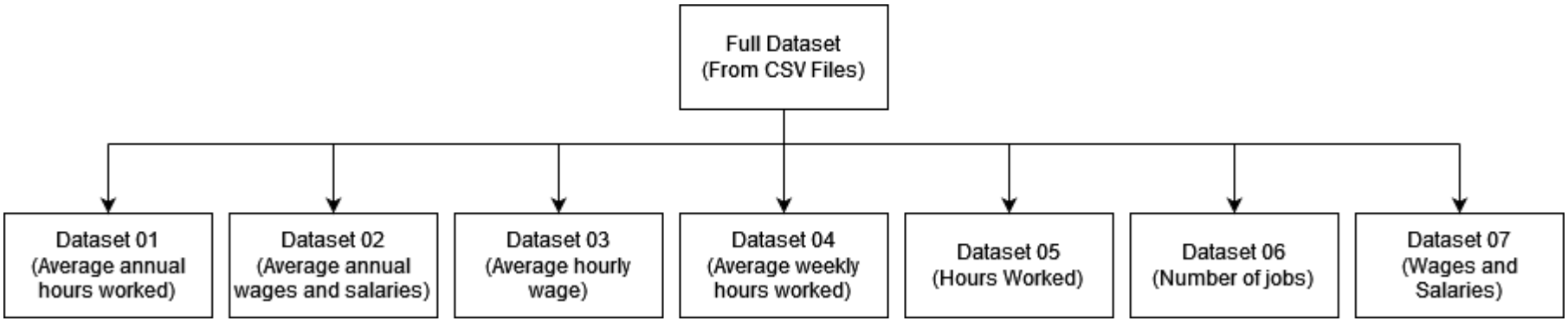
	percent_in_na
VALUE	0.0

Splitting the dataset based on ‘Indicators’

- Divided into Seven dataset, Seven indicators
- Ran into Best Fit for each.
- Measured in Annual/Weekly are distributed otherwise, skewed left.
- Did brief Cohort Analysis after splitting

Splitting the dataset based on ‘Indicators’

- Divided into Seven dataset, Seven indicators




```
Average annual hours worked
                            median         mean         std   size
Indicators
Average annual hours worked  1593.0  1551.436138  252.784087  14688
The total number of this one is  14688

Average annual wages and salaries
                                  median          mean           std   size
Indicators
Average annual wages and salaries  42186.5  43804.782748  16620.351087  14688
The total number of this one is  14688
```

… Continuing …

Spliting the dataset based on ‘Indicators’

- Ran into Best Fit for each.
  - Average annual hours worked: ‘beta’
  - Average annual wages and salaries: ‘gamma’
  - Average hourly wage: ‘lognorm’
  - Average weekly hours worked: ‘burr’
  - Hours worked: ‘lognorm’
  - Number of jobs: ‘beta’
  - Wages and Salaries: ‘beta’

Spliting 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**

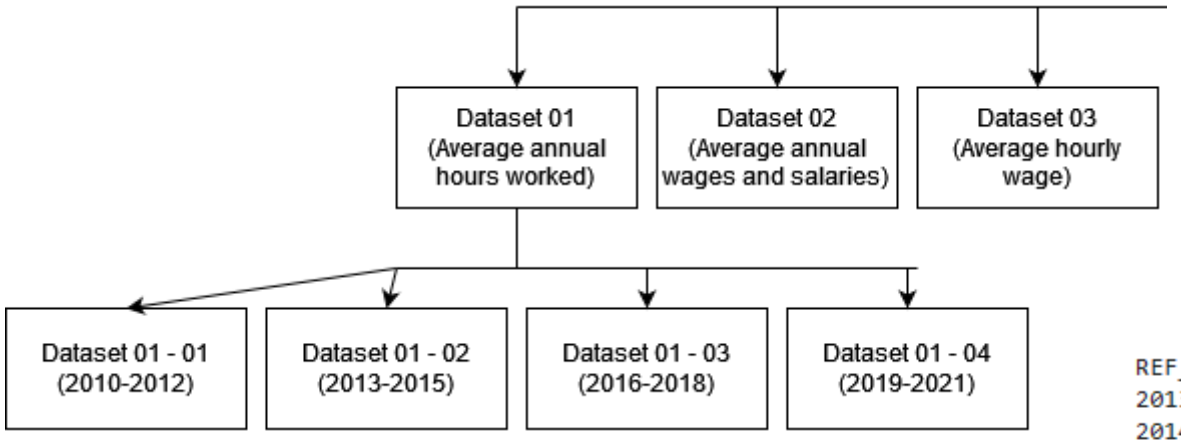
Finish it using Excel
Only ‘Average annual hours worked’

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

Spliting the Training and Testing set

- Divide into four dataset initially, 2010-2012, 2013-2015, 2016-2018, 2019-2021.

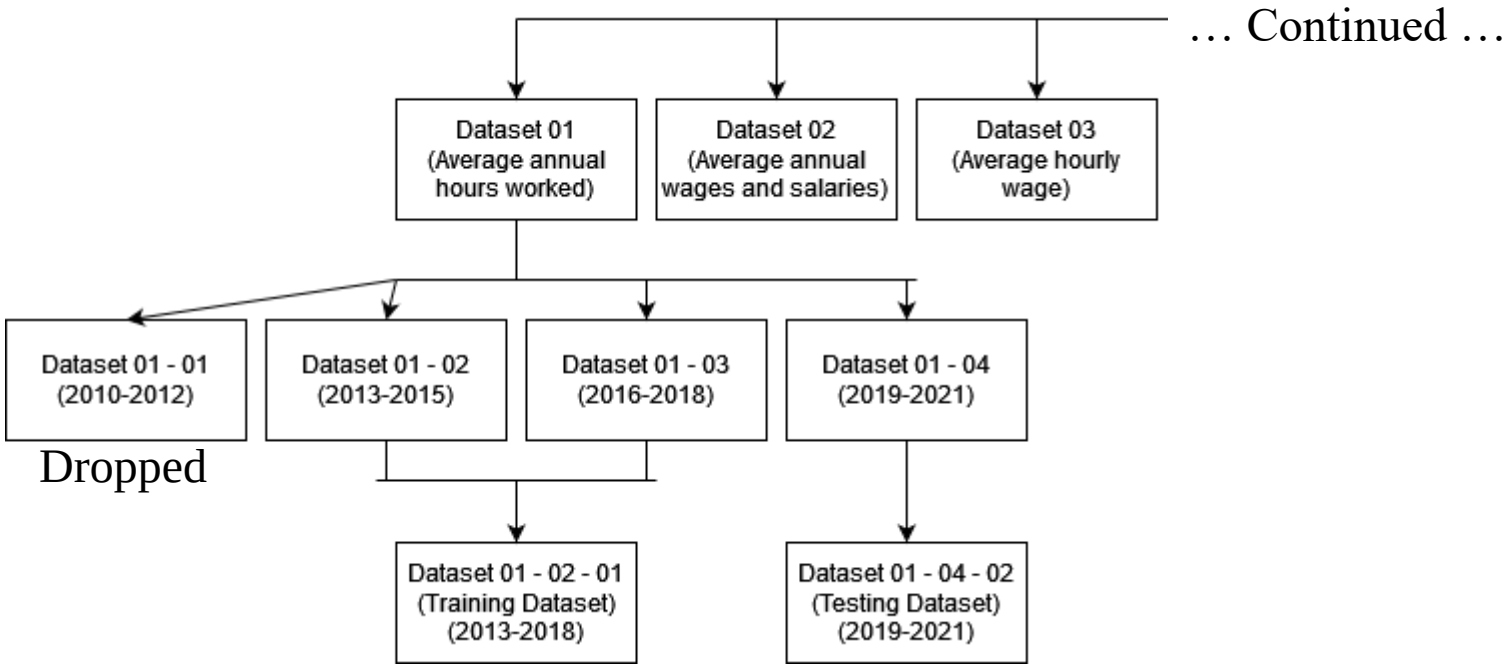


REF_DATE	sum	size
2013	36671.0	1224
2014	36483.0	1224
2015	36678.0	1224
2016	36541.0	1224
2017	36170.0	1224
2018	36416.0	1224

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

Spliting the Training and Testing set

- Drop 2010-2012 dataset
- Merge 2013-2015, 2016-2018 into training set.
- Use 2019-2021 into testing set.



Spliting the Training and Testing set

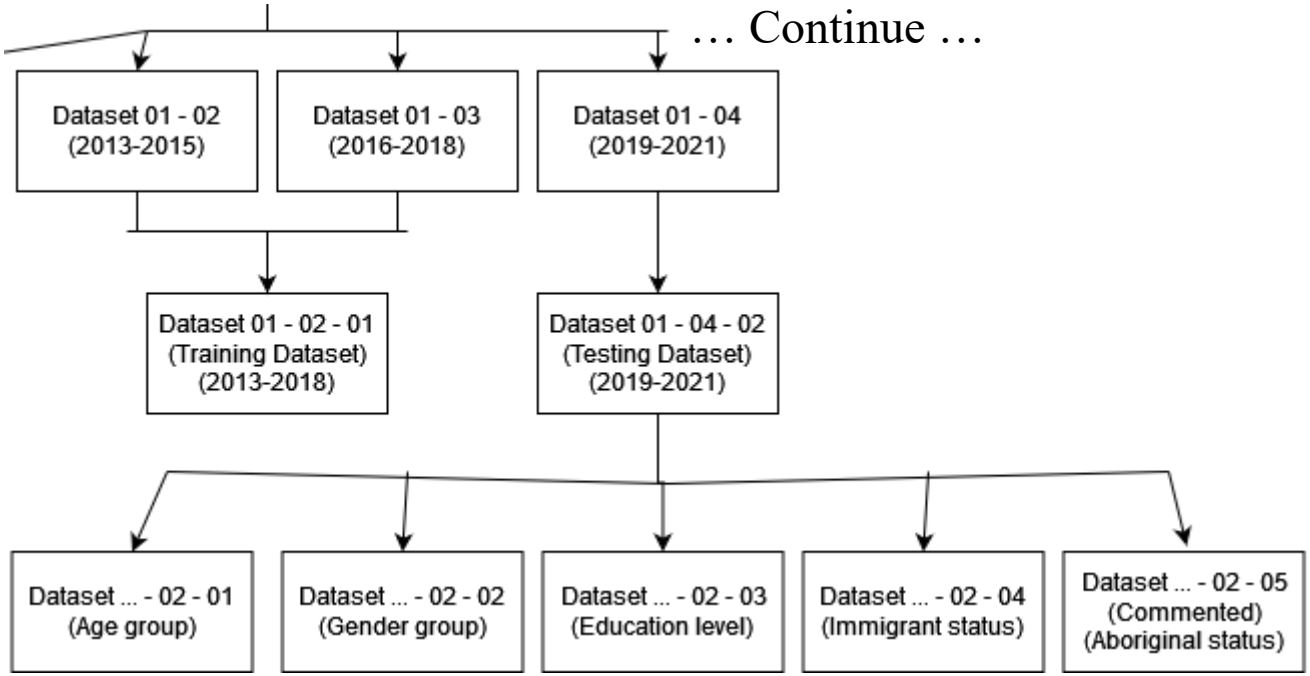
- Do chi-square before splitting
- AvgAnnHrsWrk : p value is 1.0, Independent (H0 holds true)
- AvgAnnHrsWages : p value is 1.0, Independent (H0 holds true)
- AvgWeekHrsWrked : p value is 1.0, Independent (H0 holds true)
- Hours Worked : p value is 1.0, Independent (H0 holds true)
- Number of Jobs : 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’.
  - Repeated ‘Seven times’, number of ‘indicators’ unique observations.
- Other columns not related here will be dropped.

Divide by ‘Characteristics’ columns

- Divide into four dataset based on ‘Age’, ‘Gender’, ‘Education’, ‘Immigrant’.



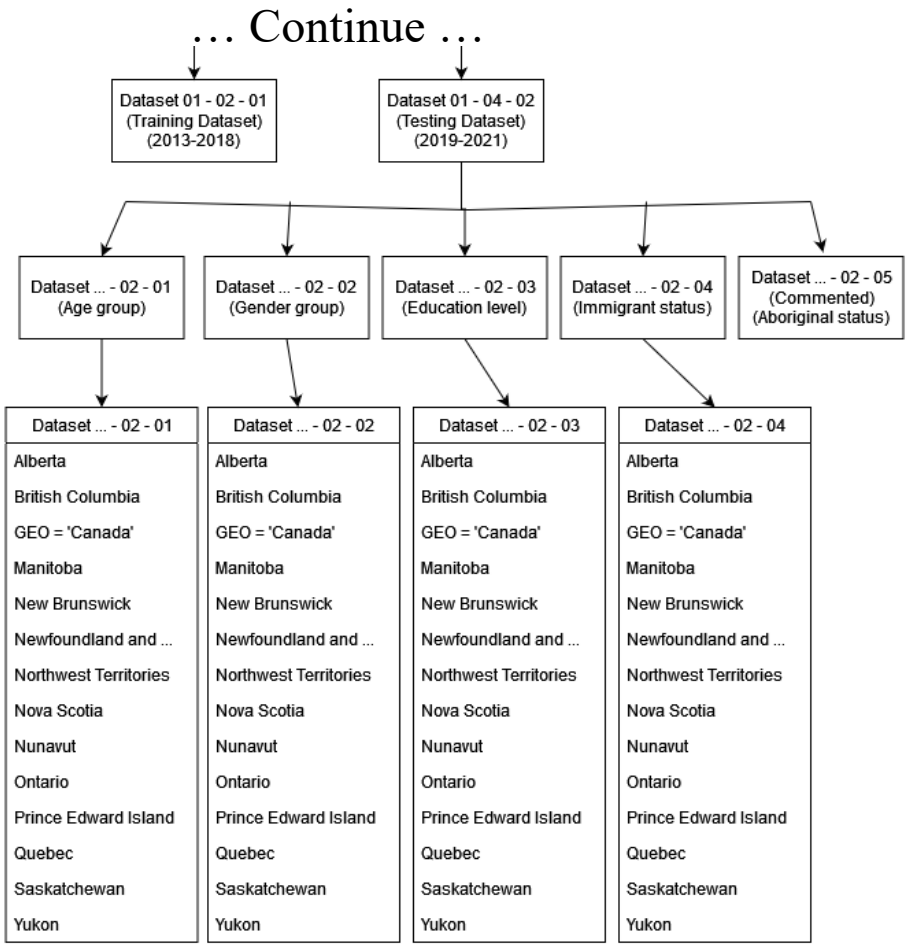

```
testing set By Age
                          sum         mean    amin  median    amax  size
Characteristics
15 to 24 years         179872.0   936.833333   713.0   927.5  1281.0   192
25 to 34 years         333662.0  1588.866667  1292.0  1576.0  1870.0   210
35 to 44 years         371386.0  1768.504762  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 :  -0.15683774387071896
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
```

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.
  - Each province is converted into binary, one hand encoding added as indicators.
- 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.

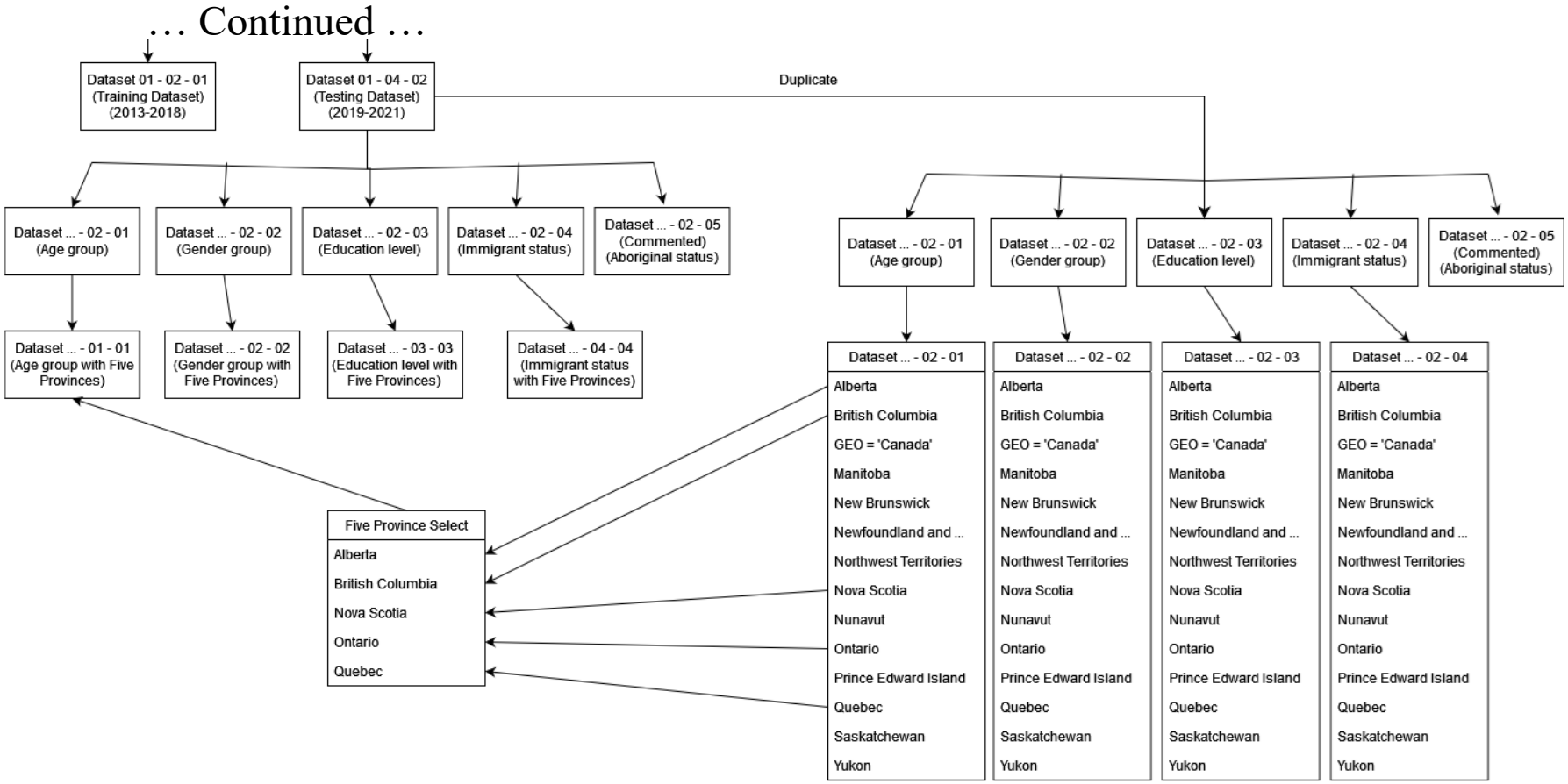
Divide by ‘Provinces’ columns

- Divide into 13 provinces provided from the previous divided dataset

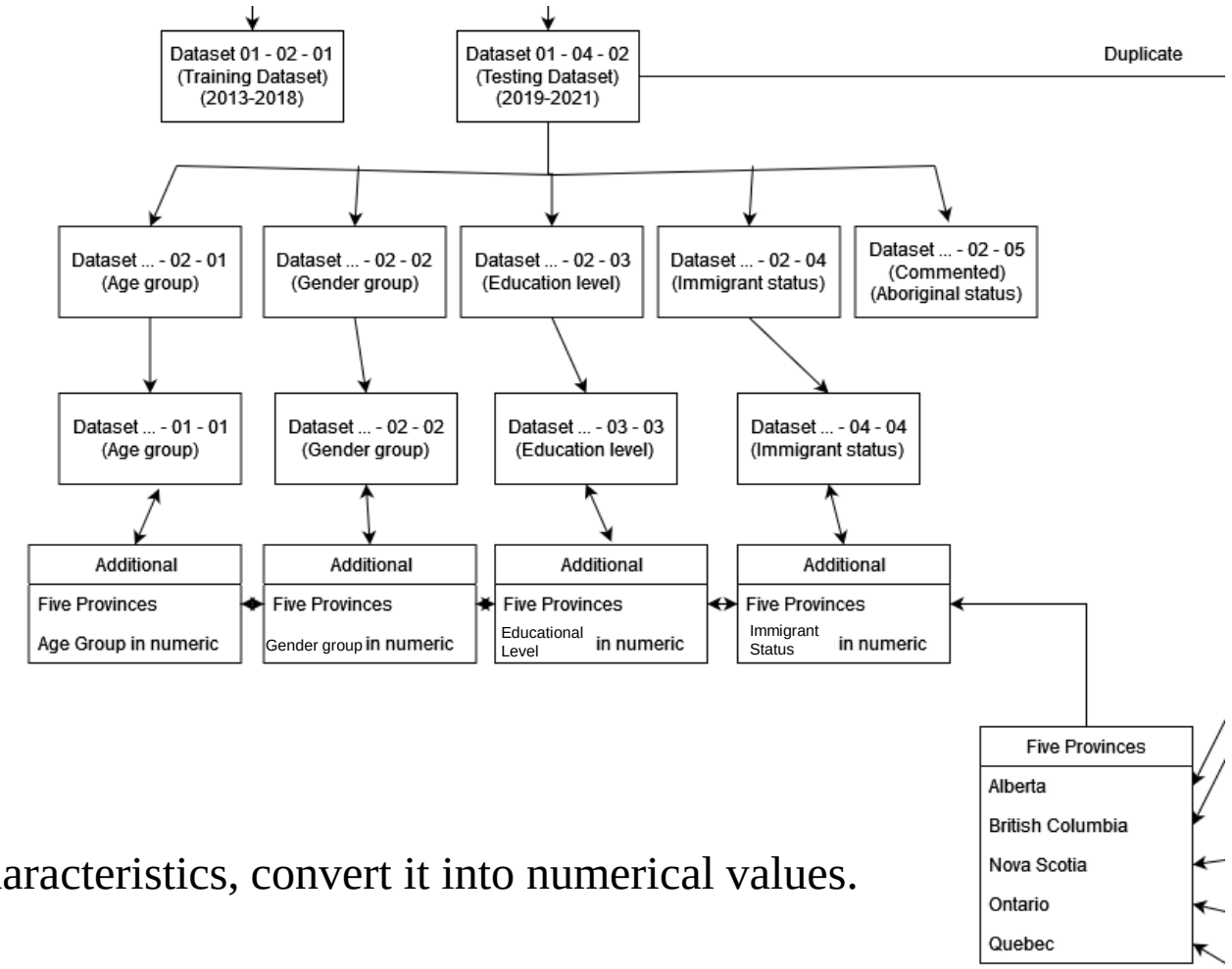


Divide by ‘Provinces’ columns

- Only select five provinces to analysis.



Divide by ‘Provinces’ columns



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

Divide by ‘Provinces’ columns

- Depend on the type of Characteristics, convert it into numerical values. (Continued)

```
<class 'pandas.core.frame.DataFrame'>
Index: 450 entries, 85137 to 100796
Data columns (total 14 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
… … … … … … … … … … … … … … … … … … … … … … … … … … … … … … … … … … …
 8   GEO_Alberta           450 non-null    bool  (“One hand encoding”)
 9   GEO_British Columbia  450 non-null    bool  (“One hand encoding”)
 10  GEO_Nova Scotia       450 non-null    bool  (“One hand encoding”)
 11  GEO_Ontario           450 non-null    bool  (“One hand encoding”)
 12  GEO_Quebec            450 non-null    bool  (“One hand encoding”)
 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
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 150 entries, 85053 to 100684
Data columns (total 14 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
… … … … … … … … … … … … … … … … … … … … … … … … … … … … … … … … … … …
 8   GEO_Alberta           150 non-null    bool    (“One hand encoding”)
 9   GEO_British Columbia  150 non-null    bool    (“One hand encoding”)
 10  GEO_Nova Scotia       150 non-null    bool    (“One hand encoding”)
 11  GEO_Ontario           150 non-null    bool    (“One hand encoding”)
 12  GEO_Quebec            150 non-null    bool    (“One hand encoding”)
 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
```

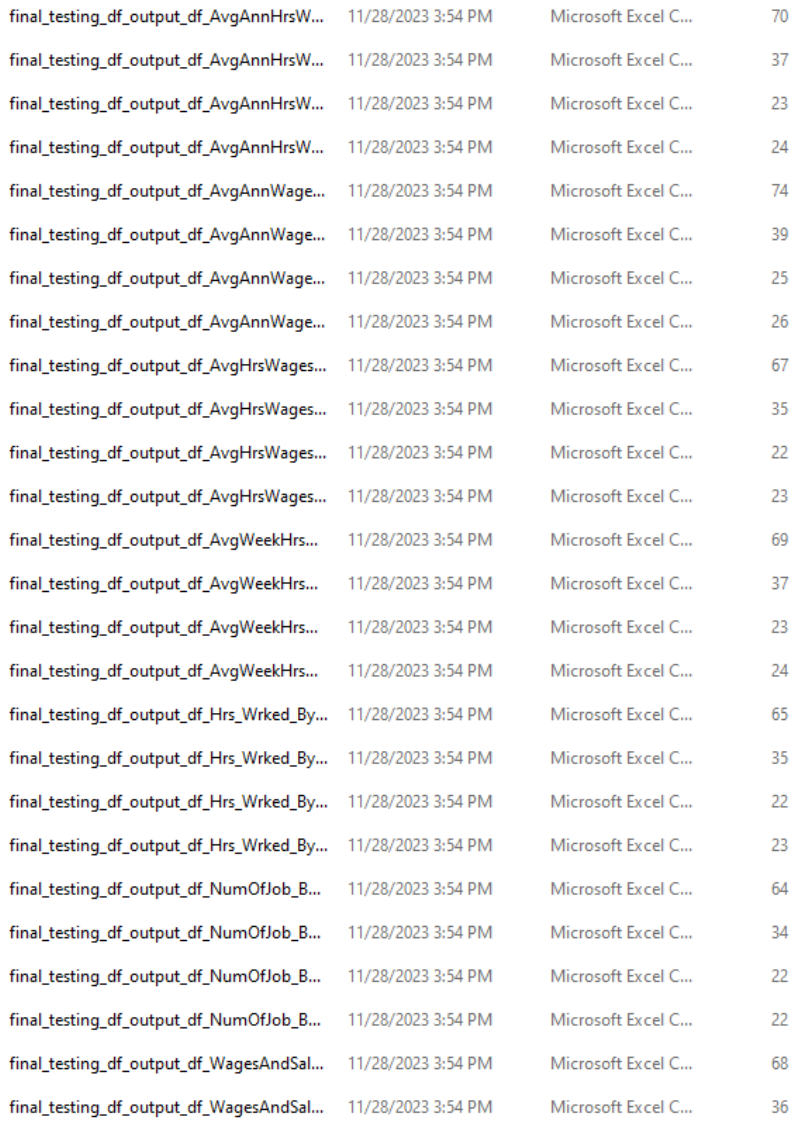
```
<class 'pandas.core.frame.DataFrame'>
Index: 225 entries, 85109 to 100754
Data columns (total 14 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
… … … … … … … … … … … … … … … … … … … … … … … … … … … … … … … … … … …
 8   GEO_Alberta           225 non-null    bool    (“One hand encoding”)
 9   GEO_British Columbia  225 non-null    bool    (“One hand encoding”)
 10  GEO_Nova Scotia       225 non-null    bool    (“One hand encoding”)
 11  GEO_Ontario           225 non-null    bool    (“One hand encoding”)
 12  GEO_Quebec            225 non-null    bool    (“One hand encoding”)
 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
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 150 entries, 85067 to 100698
Data columns (total 14 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
… … … … … … … … … … … … … … … … … … … … … … … … … … … … … … … … … … …
 8   GEO_Alberta           150 non-null    bool   (“One hand encoding”)
 9   GEO_British Columbia  150 non-null    bool   (“One hand encoding”)
 10  GEO_Nova Scotia       150 non-null    bool   (“One hand encoding”)
 11  GEO_Ontario           150 non-null    bool   (“One hand encoding”)
 12  GEO_Quebec            150 non-null    bool   (“One hand encoding”)
 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
```

Ryerson University

Divide by ‘Provinces’ columns

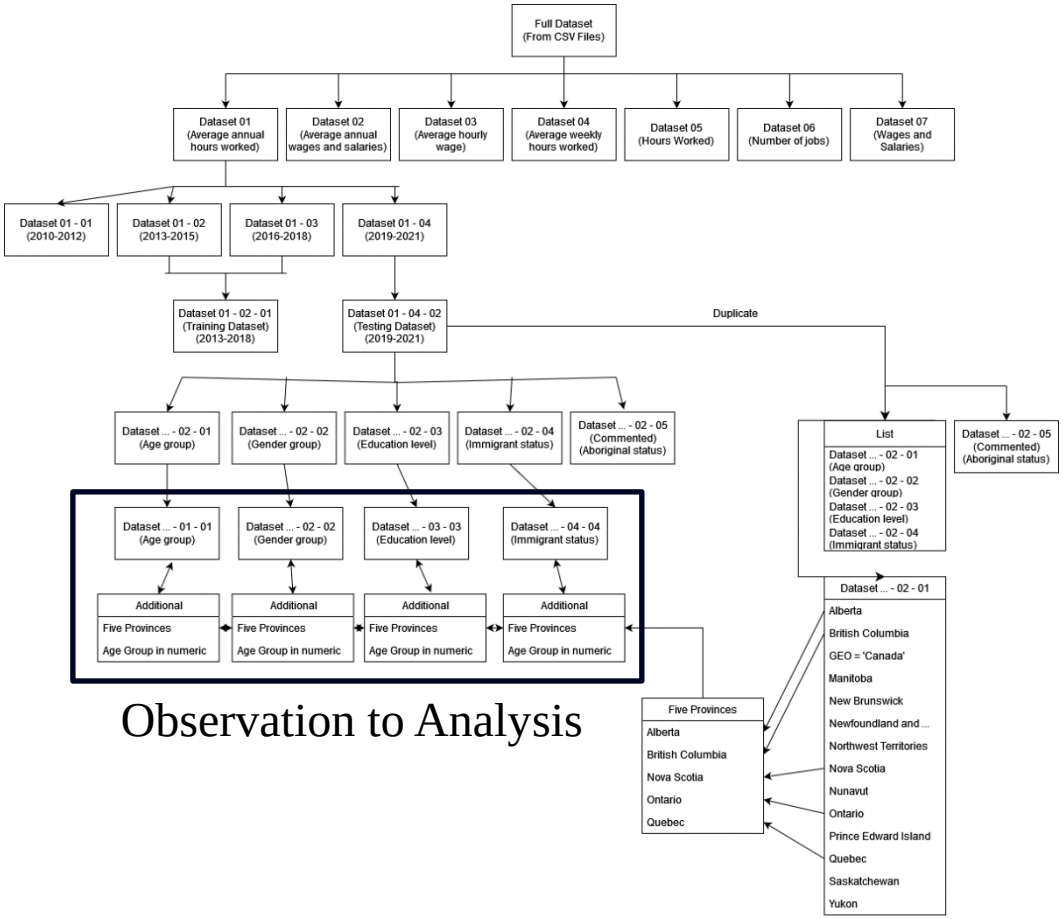
- Convert it into csv file and re-export back to final analysis in different script.



final_testing_df_output_df_AvgAnnHrsW...	11/28/2023 3:54 PM	Microsoft Excel C...	70
final_testing_df_output_df_AvgAnnHrsW...	11/28/2023 3:54 PM	Microsoft Excel C...	37
final_testing_df_output_df_AvgAnnHrsW...	11/28/2023 3:54 PM	Microsoft Excel C...	23
final_testing_df_output_df_AvgAnnHrsW...	11/28/2023 3:54 PM	Microsoft Excel C...	24
final_testing_df_output_df_AvgAnnWage...	11/28/2023 3:54 PM	Microsoft Excel C...	74
final_testing_df_output_df_AvgAnnWage...	11/28/2023 3:54 PM	Microsoft Excel C...	39
final_testing_df_output_df_AvgAnnWage...	11/28/2023 3:54 PM	Microsoft Excel C...	25
final_testing_df_output_df_AvgAnnWage...	11/28/2023 3:54 PM	Microsoft Excel C...	26
final_testing_df_output_df_AvgHrsWages...	11/28/2023 3:54 PM	Microsoft Excel C...	67
final_testing_df_output_df_AvgHrsWages...	11/28/2023 3:54 PM	Microsoft Excel C...	35
final_testing_df_output_df_AvgHrsWages...	11/28/2023 3:54 PM	Microsoft Excel C...	22
final_testing_df_output_df_AvgHrsWages...	11/28/2023 3:54 PM	Microsoft Excel C...	23
final_testing_df_output_df_AvgWeekHrs...	11/28/2023 3:54 PM	Microsoft Excel C...	69
final_testing_df_output_df_AvgWeekHrs...	11/28/2023 3:54 PM	Microsoft Excel C...	37
final_testing_df_output_df_AvgWeekHrs...	11/28/2023 3:54 PM	Microsoft Excel C...	23
final_testing_df_output_df_AvgWeekHrs...	11/28/2023 3:54 PM	Microsoft Excel C...	24
final_testing_df_output_df_Hrs_Wrked_By...	11/28/2023 3:54 PM	Microsoft Excel C...	65
final_testing_df_output_df_Hrs_Wrked_By...	11/28/2023 3:54 PM	Microsoft Excel C...	35
final_testing_df_output_df_Hrs_Wrked_By...	11/28/2023 3:54 PM	Microsoft Excel C...	22
final_testing_df_output_df_Hrs_Wrked_By...	11/28/2023 3:54 PM	Microsoft Excel C...	23
final_testing_df_output_df_NumOfJob_B...	11/28/2023 3:54 PM	Microsoft Excel C...	64
final_testing_df_output_df_NumOfJob_B...	11/28/2023 3:54 PM	Microsoft Excel C...	34
final_testing_df_output_df_NumOfJob_B...	11/28/2023 3:54 PM	Microsoft Excel C...	22
final_testing_df_output_df_NumOfJob_B...	11/28/2023 3:54 PM	Microsoft Excel C...	22
final_testing_df_output_df_WagesAndSal...	11/28/2023 3:54 PM	Microsoft Excel C...	68
final_testing_df_output_df_WagesAndSal...	11/28/2023 3:54 PM	Microsoft Excel C...	36

Ryerson University

Overall dataset division



Data Analysis

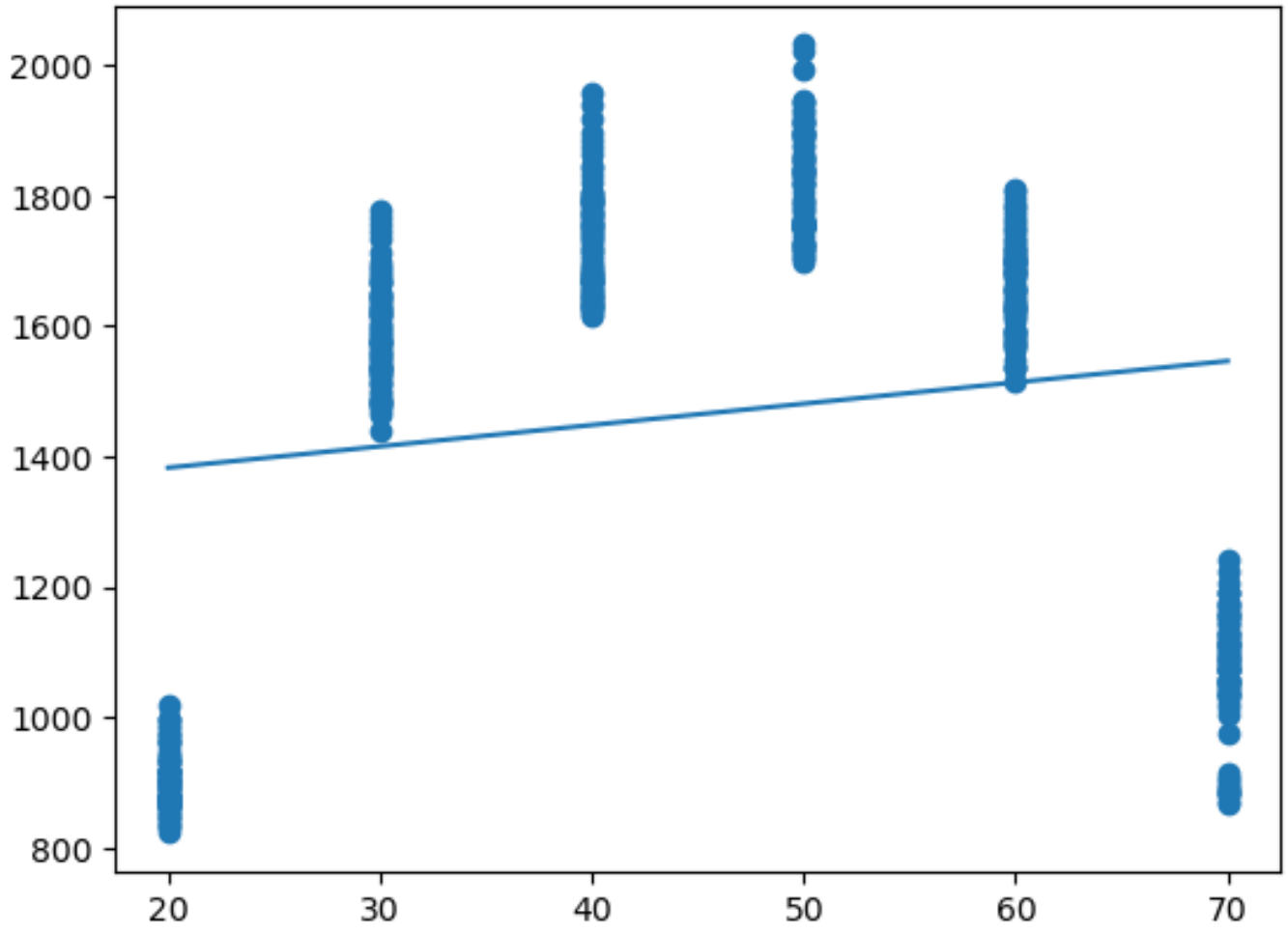
Overall Conclusion

- 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’

- Employees who are older generally work more.
- Employees in 20s and 70s work less.
- Employees in 30s to 60s works the most.
- Employees in 50s work the most.
- Employees in 20s worked 800-1100 hours/annually while employees 30s worked 1400-1800 hours/annually.
- Employees in 60s worked 1400-1800 hours/annually while employees in 70s worked 850-1300 hours/annually.

Annual Hour Worked based on ‘Age group’



Annual Hour Worked based on ‘Age group’

final_testing_df_output_df_AvgAnnHrsWrk_ByAge.csv

Characteristics	sum	mean	amin	median	amax	size
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

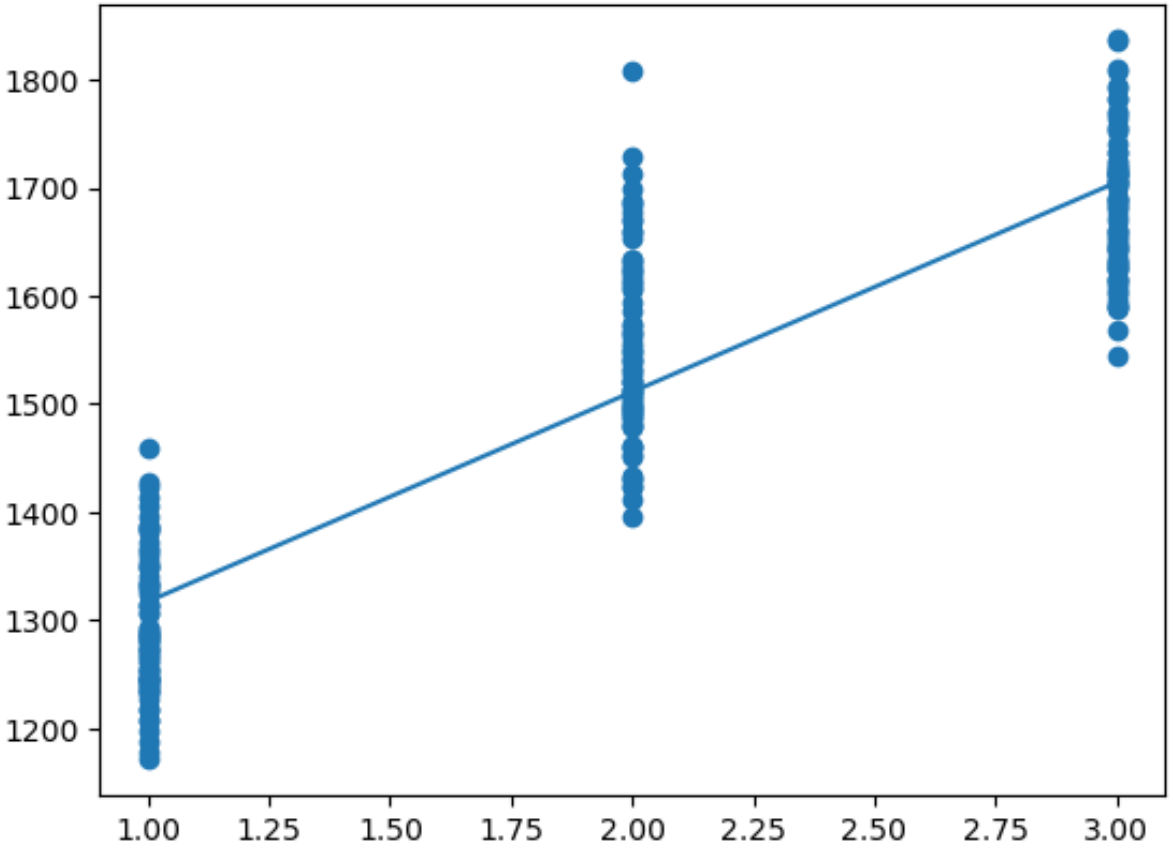
Total size : 450

Annual Hour Worked based on ‘Education level’

- Employees who get higher education work more than lower education.
- Employees with higher (university) degrees work around 1500 hours/annually to 1800 hours/annually.
- Employees with lower (high school) education work around 1000 hours/annually to 1500 hours/annually.
- Those with trade certification have a bigger gap of working hours, bigger differences between highest to their average.

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

Characteristics	sum	mean	amin	median	amax	\
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	

Characteristics	size
High school diploma and less	75
Trade certificate	75
University degree and higher	75

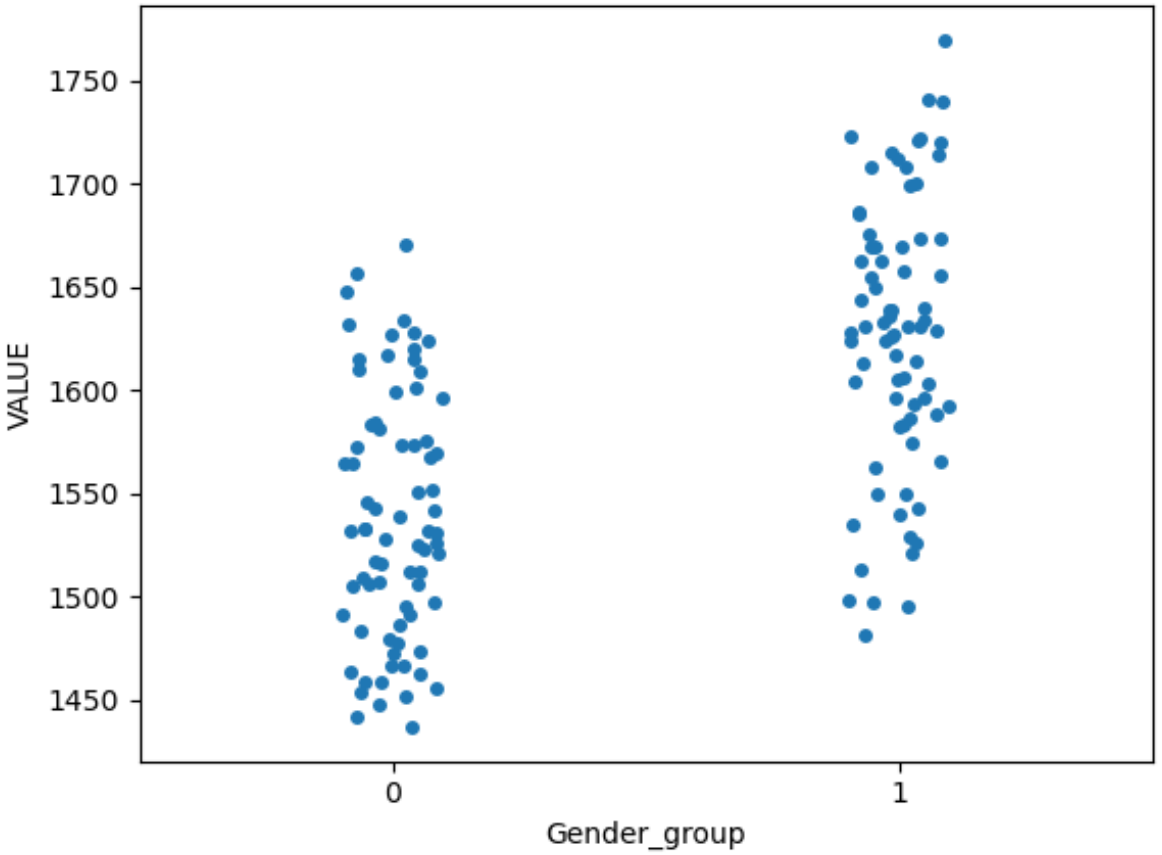
Overall,
Sum : 340034.0
Mean : 1511.2622222222221
Min/median/max : 1171.0 / 1538.0 / 1837.0
Standard Deviation : 176.936517540914
Skewnewss : -0.22126026577868005
Total size : 225

Annual Hour Worked based on ‘Gender group’

- Male employees are work more than their fellow female employees annually.
- Approximately 16 female employees work less than the lowest male employees work (1500 hours/annually).
- Approximately 19 male employees who work more than highest female work (1450 hours/annually).

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

Characteristics	sum	mean	amin	median	amax	size
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

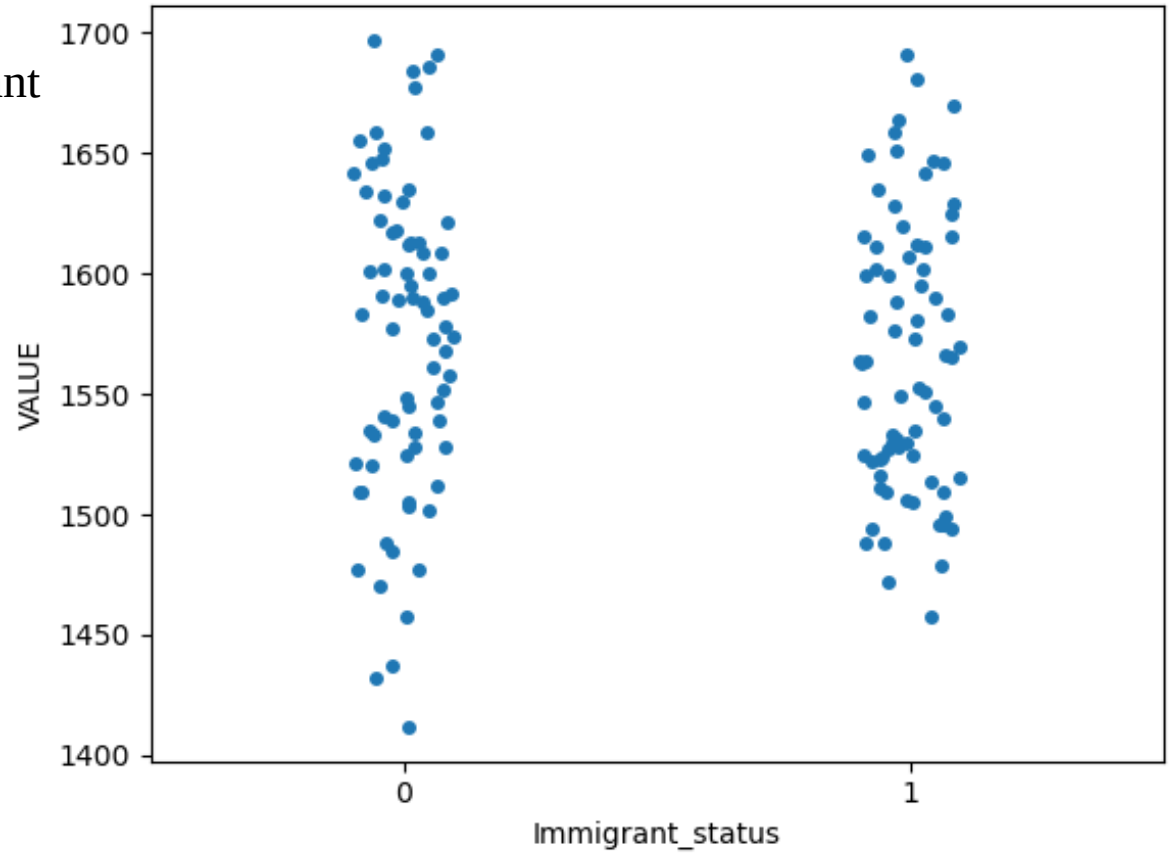
Overall,
Sum : 237396.0
Mean : 1582.64
Min/median/max : 1437.0 / 1583.5 / 1770.0
Standard Deviation : 77.0804151519697
Skewnewss : 0.1198315274026592 3
Total size : 150

Annual Hour Worked based on ‘Immigrant group’

- Both highest and lowest work employees are immigrant.
- There’s some immigrant employees who work more than highest non-immigrant worked (1691 hours/annually).
- There’s approximately three immigrant employees who work less than lowest non-immigrant worked (1458 hours/annually).

Annual Hour Worked based on ‘Immigrant group’

0 = Immigrant
1 = Non-immigrant



Ryerson University

Annual Hour Worked based on ‘Immigrant group’

final_testing_df_output_df_AvgAnnHrsWrk_ByImmigrant.csv

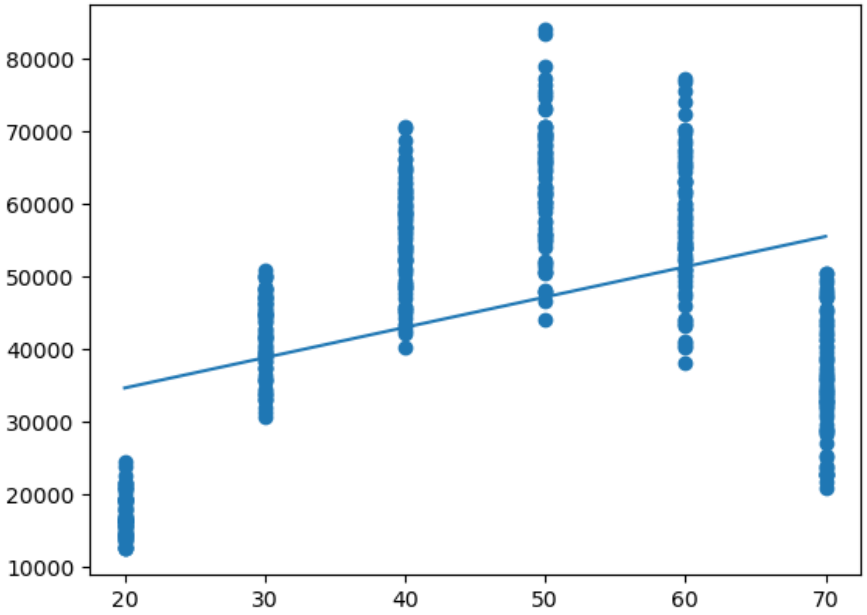
Characteristics	sum	mean	amin	median	amax	size
Immigrant employees	117967.0	1572.893333	1412.0	1583.0	1697.0	75
Non-immigrant employees	117437.0	1565.826667	1458.0	1564.0	1691.0	75

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

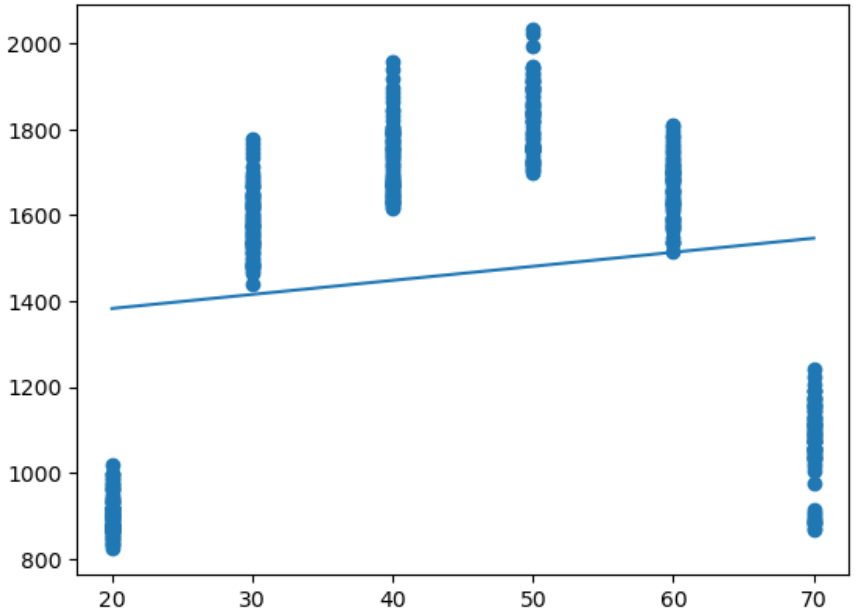
Average Annual Wage based on ‘age group’

- Employees who get older get more wages annually
- Employees between 40s to 60s get highest pay.
- There’s bigger wage gap between 20s and 30s annually.
- There’s also gap between 60s and 70s as well.
- Approximately 58 employees in 60s get more than highest wages in their 70s.

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

Characteristics	sum	mean	amin	median	amax
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

Characteristics	size
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

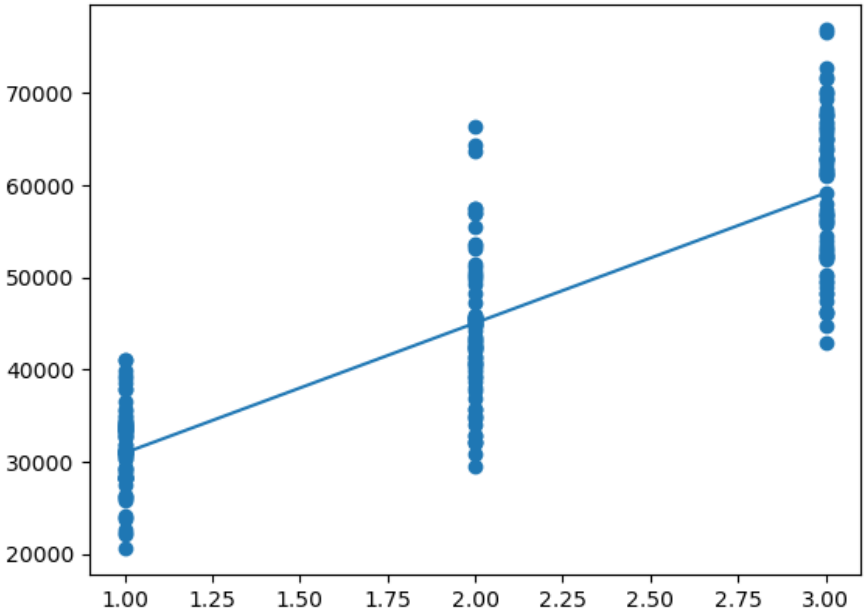
Overall,
Sum : 20271340.0
Mean : 45047.42222222222
Min/median/max : 12468.0 / 46985.0 / 83894.0
Standard Deviation : 17188.866523407123
Skewnewss : -0.2534546294165646
Total size : 450

Average Annual Wages for ‘education’

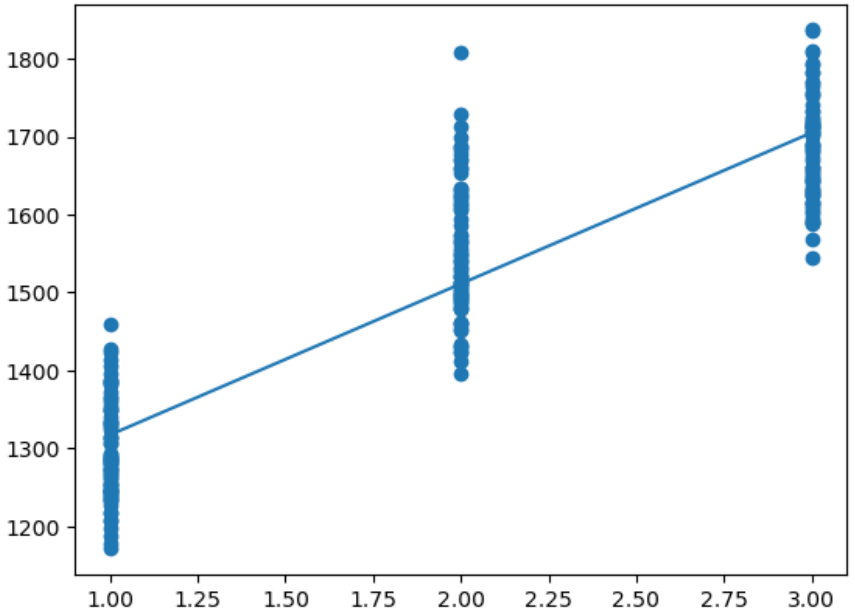
- Employees with higher education get more wages than those who have less in annually.
- There’s several employees who work more than the highest earning trade certificate does.
- Employees with trade certificate have some gap between average wages for trade certificate.

Average Annual Wages for ‘education’

Higher number = Higher education



By Average Annual Wages



By Annual work hours

Average Annual Wages for ‘education’

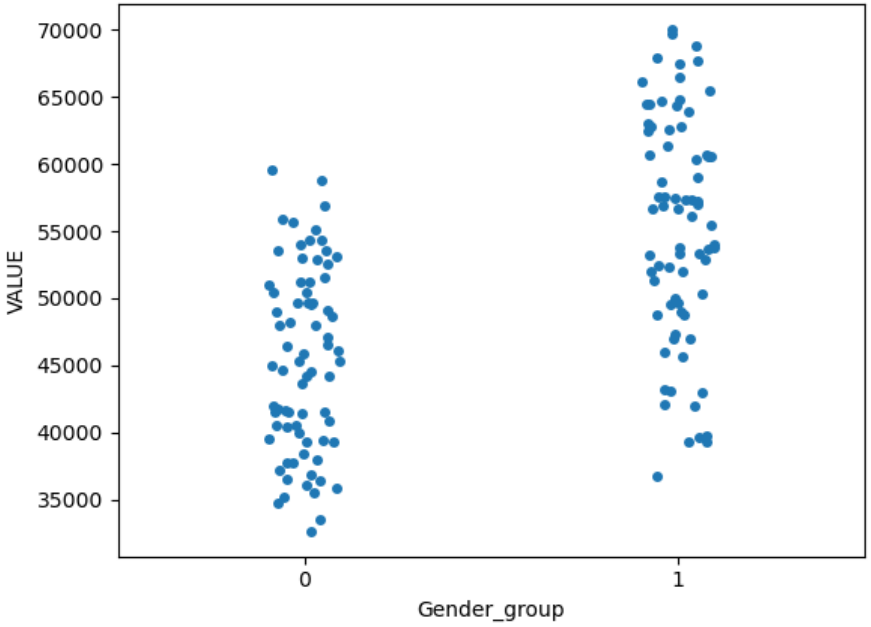
```
final_testing_df_output_df_AvgAnnWages_ByEducation.csv
                                        sum           mean     amin   median  \
Characteristics
High school diploma and less  2355715.0  31409.533333  20536.0  31272.0
Trade certificate             3308944.0  44119.253333  29523.0  43448.0
University 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’

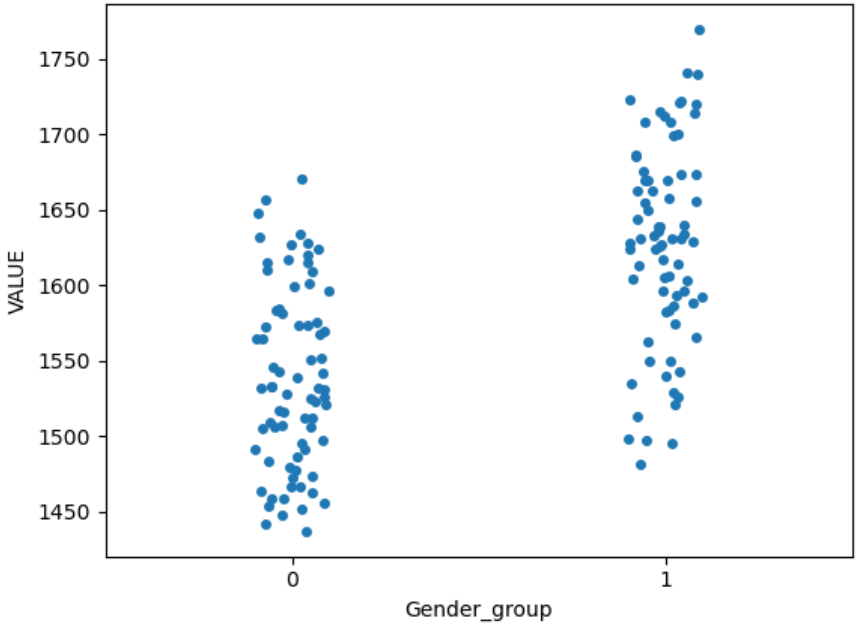
- Male employees get more wages annually than the female employees do.
- Some female employees get less than lowest male employees do.
- Approximately 26 male employees get more than the highest female employees do.

Average Annual Wages for ‘gender’

0 = Female
1 = Male



By Average Annual Wages



By Annual Hour worked

Average Annual Wages for ‘gender’

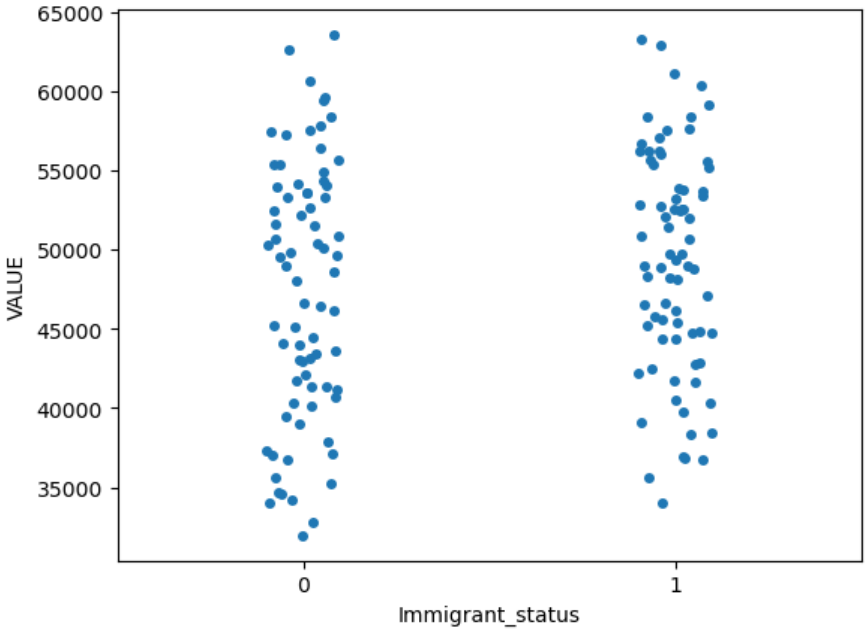
```
final_testing_df_output_df_AvgAnnWages_ByImmigrant.csv
                                  sum          mean     amin   median      amax  \
Characteristics
Immigrant employees           3554950.0  47399.333333  31977.0  48615.0  63582.0
Non-immigrant employees       3690998.0  49213.306667  34013.0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

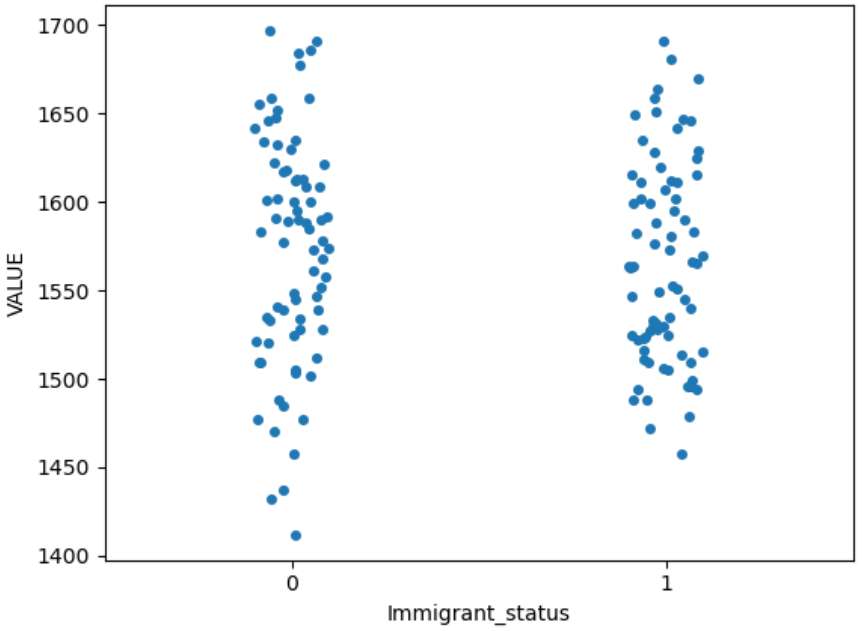
- Both immigrant and non-immigrant employees make about similar annual wages.
- Highest annual wages between immigrant and non-immigrant employees about $300 dollars differences.
- Lowest annual wages between immigrant and non-immigrant employees about $300 dollars differences too.

Average Annual Wages for ‘immigrant’

0 = Immigrant
1 = Non-immigrant



By Average annual wages



By Annual hours worked

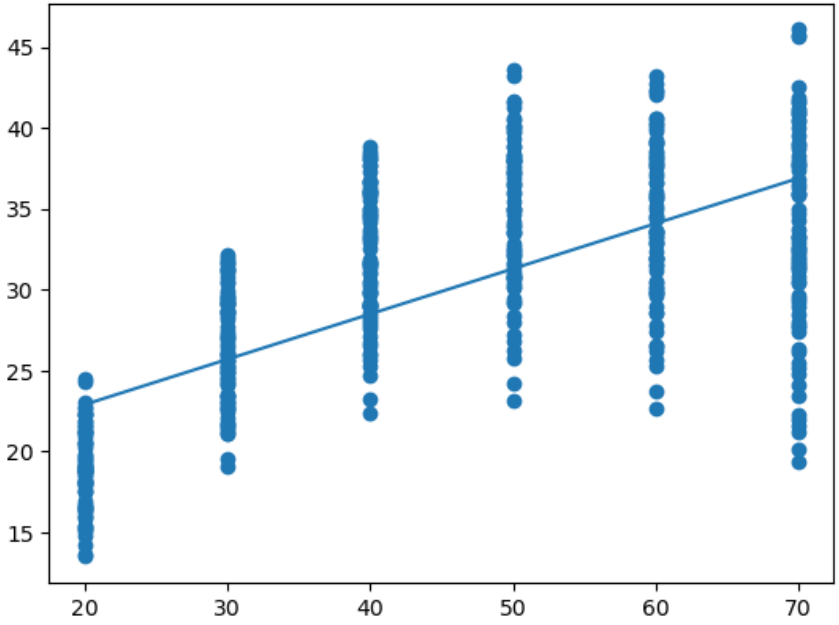
Average Annual Wages for ‘immigrant’

```
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
```

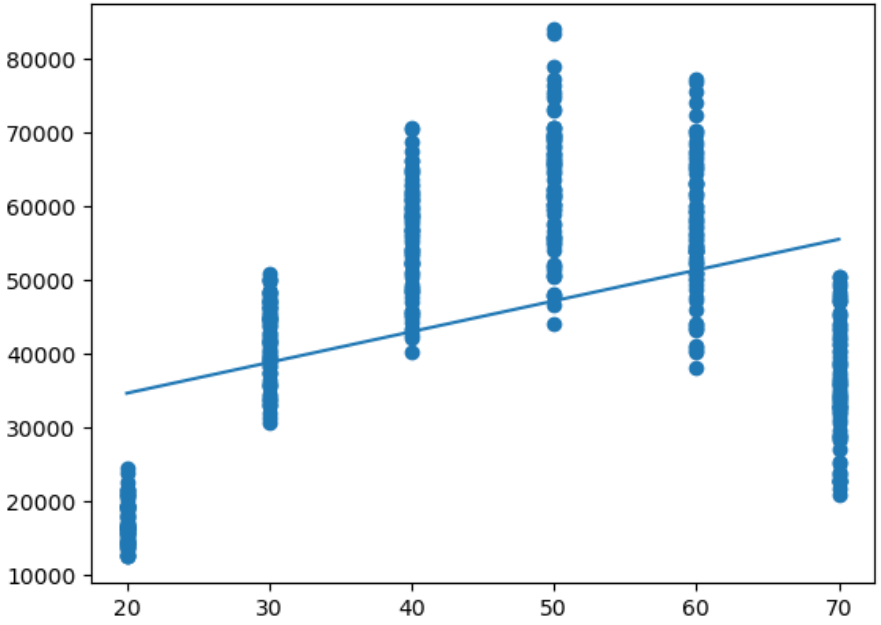
Average hourly wages for ‘age group’

- Employees in 20s still get the lowest wages per hours.
- Employees in 50s and 60s still make similar higher salaries.
- Some employees in 70s make more in hourly wages than employees in others 70s and highest 50s.
- There’s still employees in 70s who make less as well.

Average hourly wages for ‘age group’



By Hourly



By Annually

Average hourly wages for ‘age group’

final_testing_df_output_df_AvgHrsWages_ByAge.csv

Characteristics	sum	mean	amin	median	amax	size
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

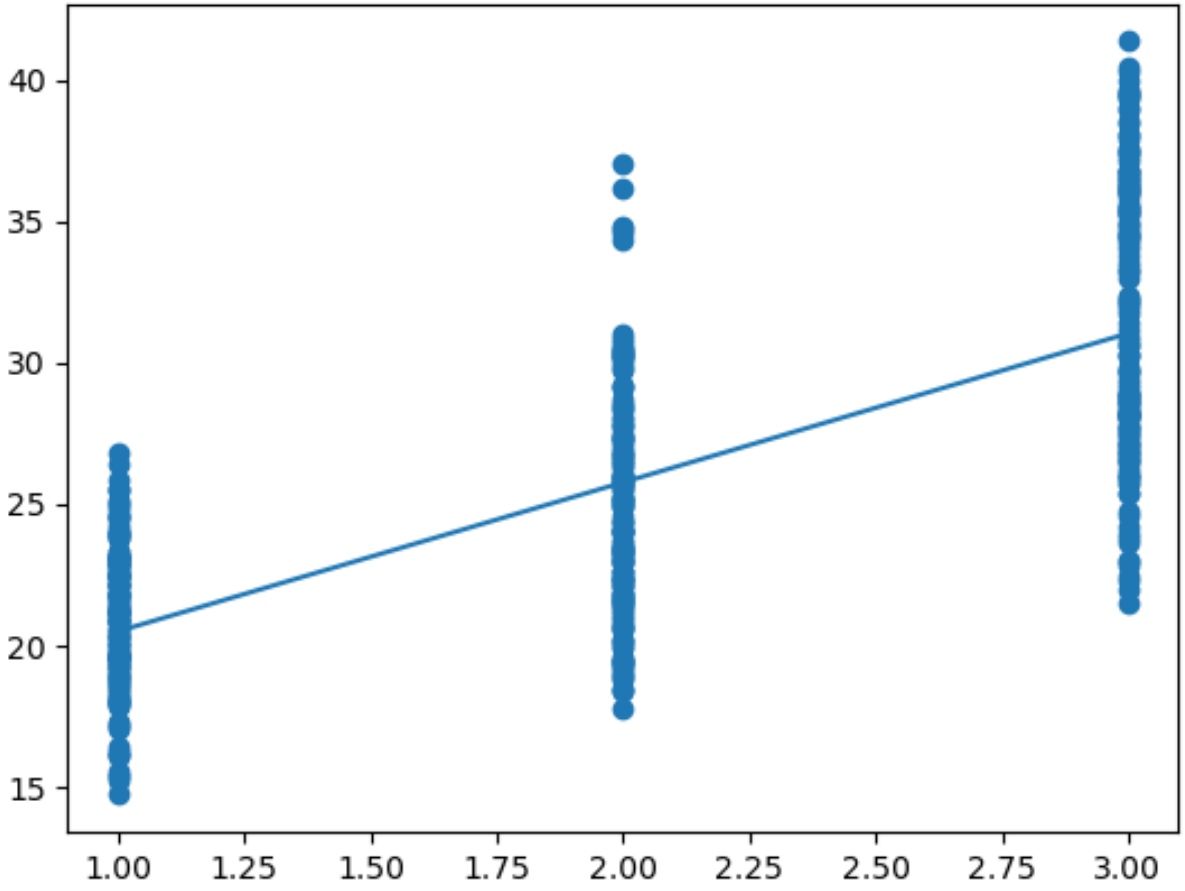
Total size : 450

Average hourly wages for ‘Education’

- Employees who are higher in education get more wages hourly than the employees in lower in education.
- Some who have trade certification make more than their fellow trade certificate workers but not as higher as university degrees.
- This is similar concept as annual wages for education.

Average hourly wages for ‘Education’

Higher number = Higher education



Used training set instead

Average hourly wages for ‘Education’

final_testing_df_output_df_AvgHrsWages_ByEducation.csv

Characteristics	sum	mean	amin	median	amax	size
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

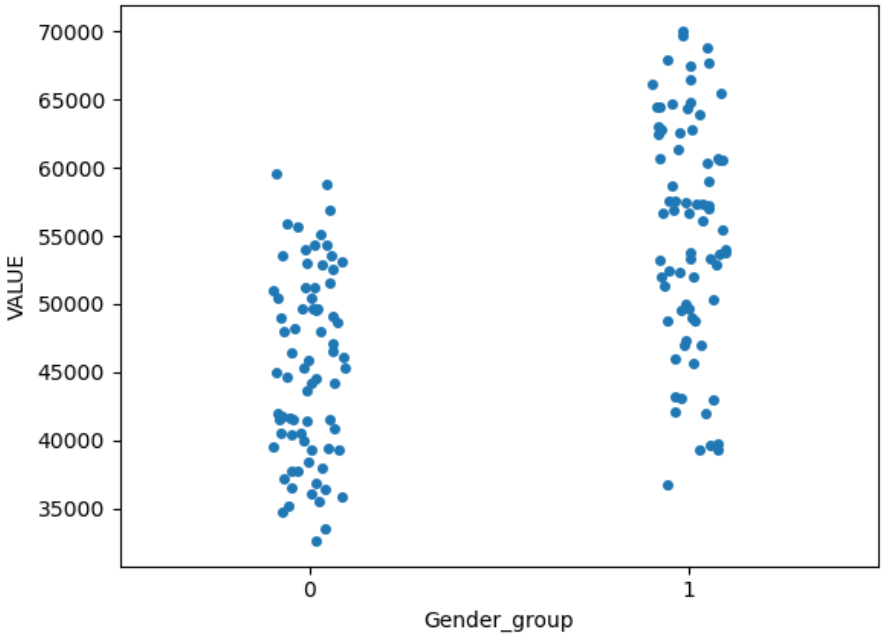
Total size : 225

Average hourly wages for ‘Gender group’

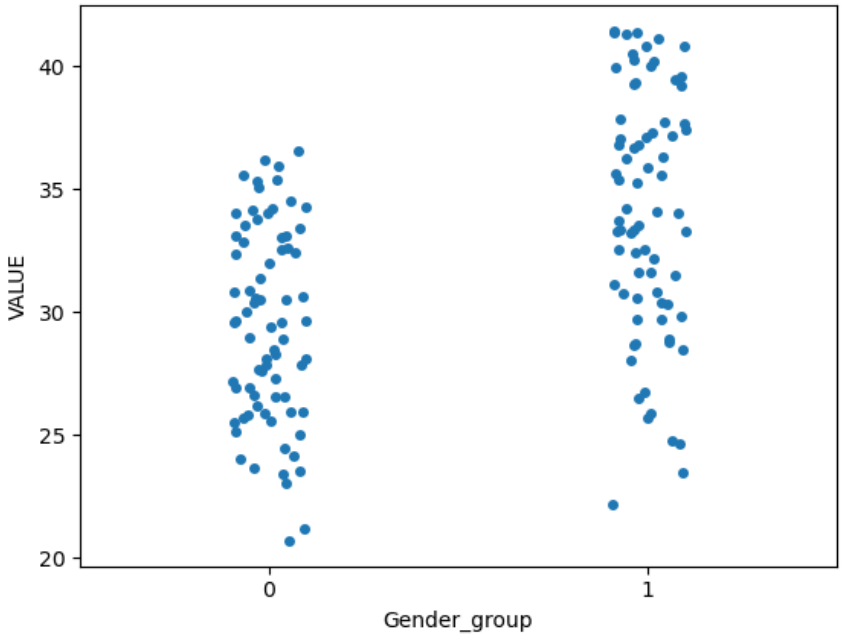
- Male employees still make more money per hours than female employees do.
- There isn’t any differences between annual to hourly in term of trend.
- Only differences are there’s fewer female employees (approx. two female employees) who make less than male lowest hourly age.

Average hourly wages for ‘Gender group’

0 = Female
1 = Male



By Annually



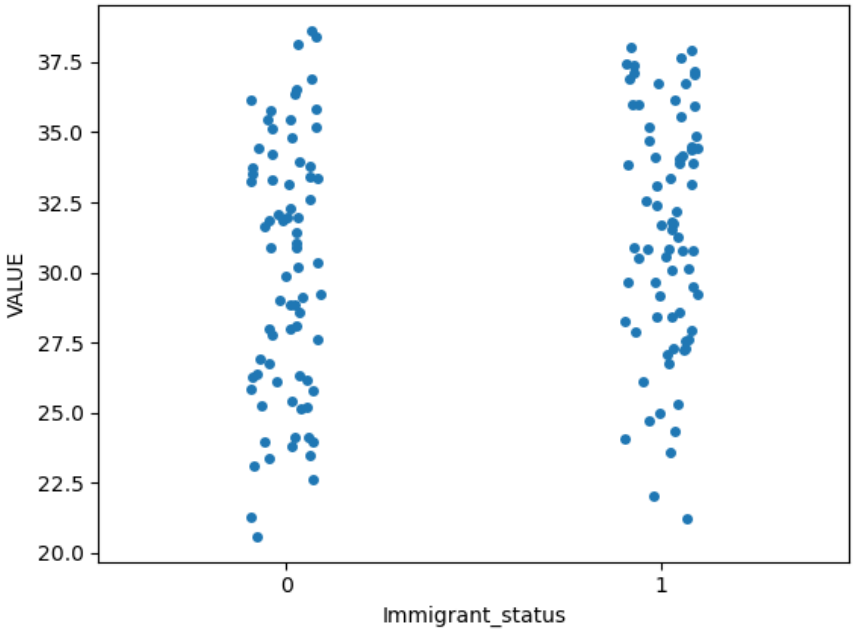
By hourly wages

Average hourly wages for immigrant

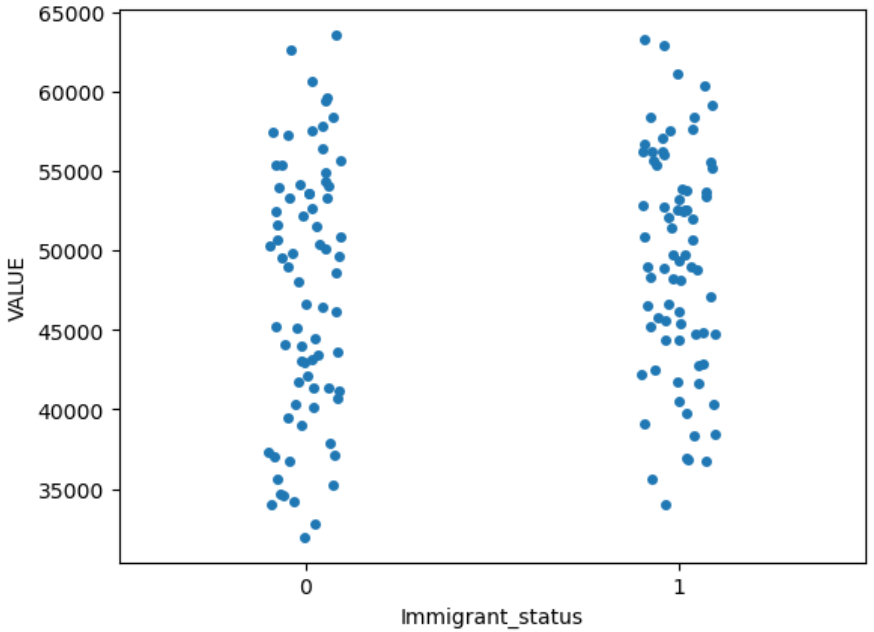
- Both non-immigrant and immigrant both make minimum and maximum hourly wages.
- Only difference is .50 to $1.

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

Characteristics	sum	mean	amin	median	amax	size
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

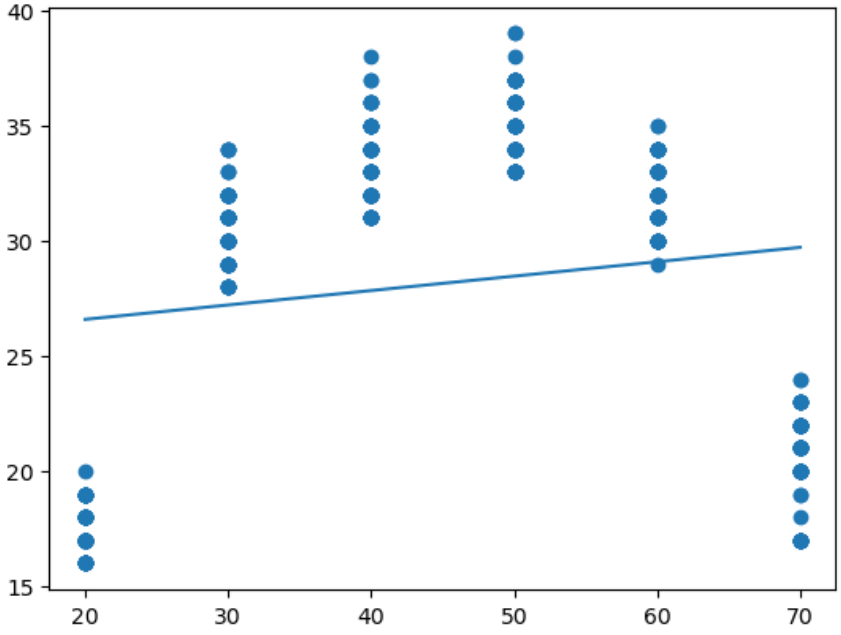
Skewnewss : -0.21241877283543747

Total size : 150

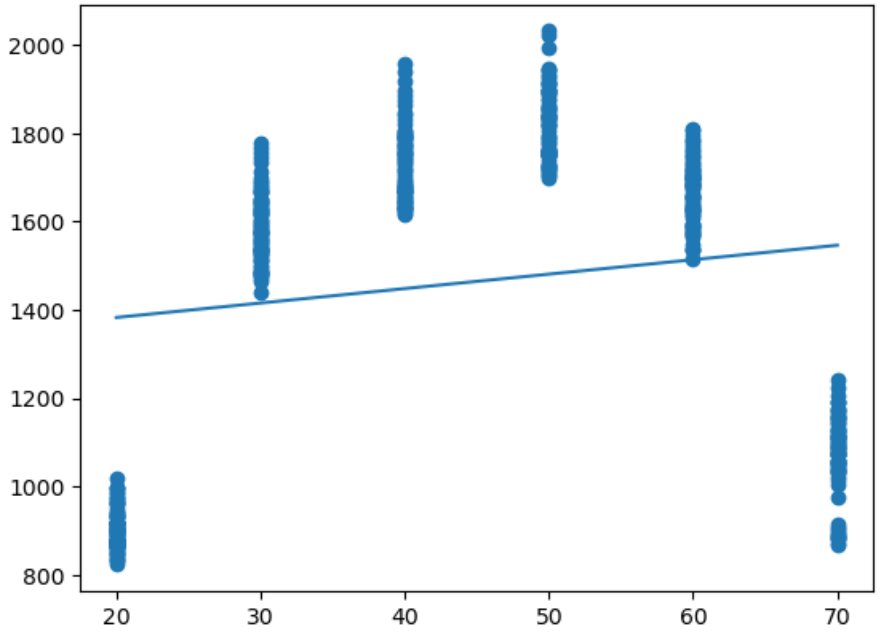
Average weekly hours worked by ‘Age group’

- Like annually, employees in 20s and 70s still work less.
- Like annually, employees in 30s to 60s work more.
- Noticed that the data is organized categorically because of that the plot is showing small gap between each others.

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

Characteristics	sum	mean	amin	median	amax	size
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 : 6.835271269915569 5

Skewnewss : -0.5809323008907273

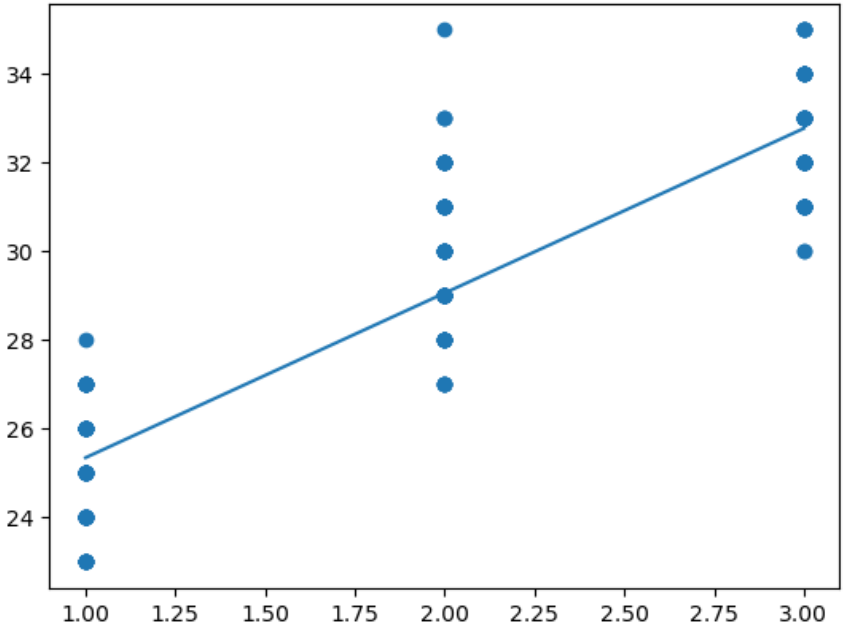
Total size : 450

Average weekly hours worked based on ‘Education group’

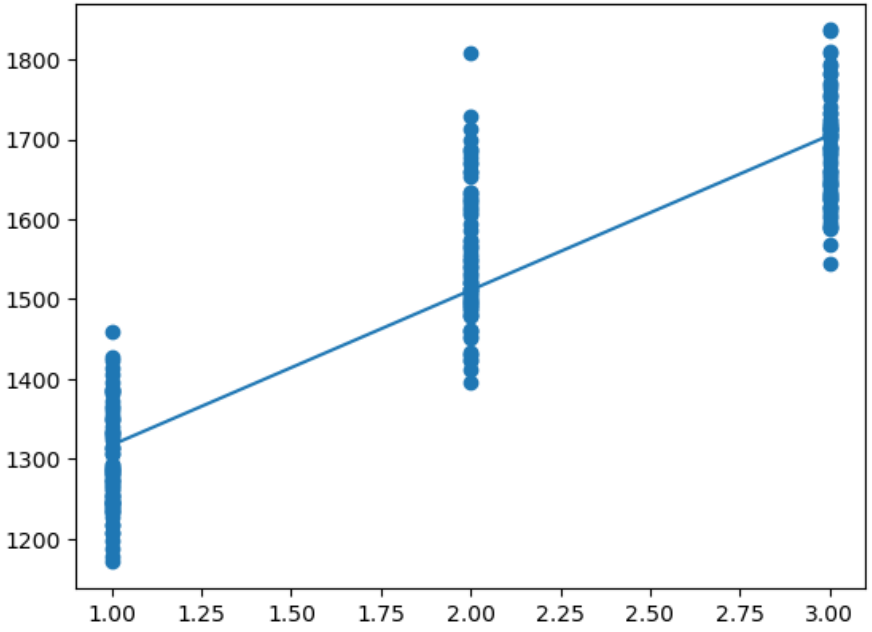
- Those who have highest educations work more than the lowest educations weekly.
- All the sample given are approximately a hour and two hours differences.
- Trade certification like annual work hours still have gap and only exception to 1-2 hours differences rules.
- For references, highest work hours for lowest educations is around 28 hours weekly. Lowest work hours for highest educations is around 30 hours weekly.

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

Characteristics	sum	mean	amin	median	amax	size
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

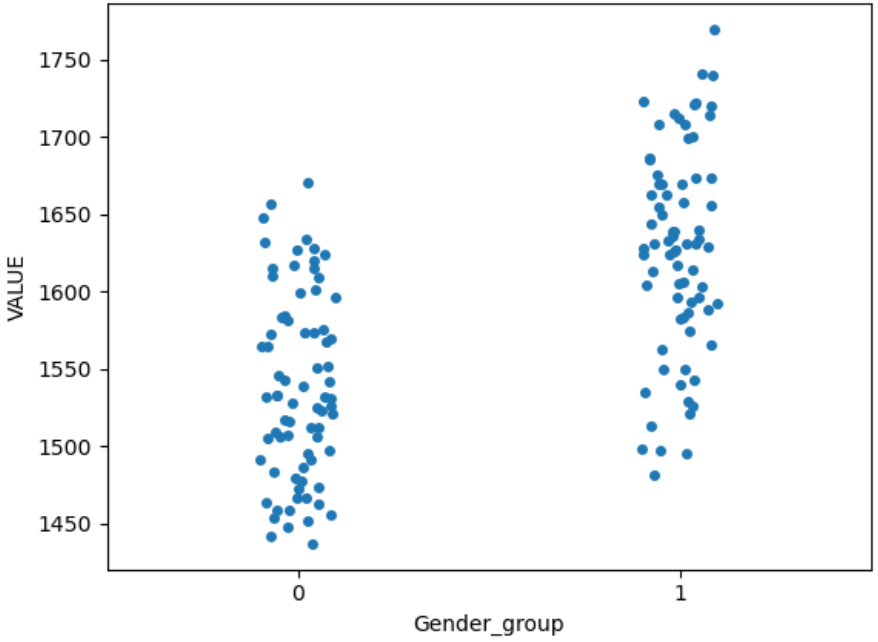
Average weekly work hours for on ‘Gender group’

- More female employees work higher than male employees.
- Male employees (34 hours weekly) are less likely to work lowest than female employees (32 hours weekly) do.
- Also, no female employees work less than male employees who work lowest. (approx. 28 hours weekly)

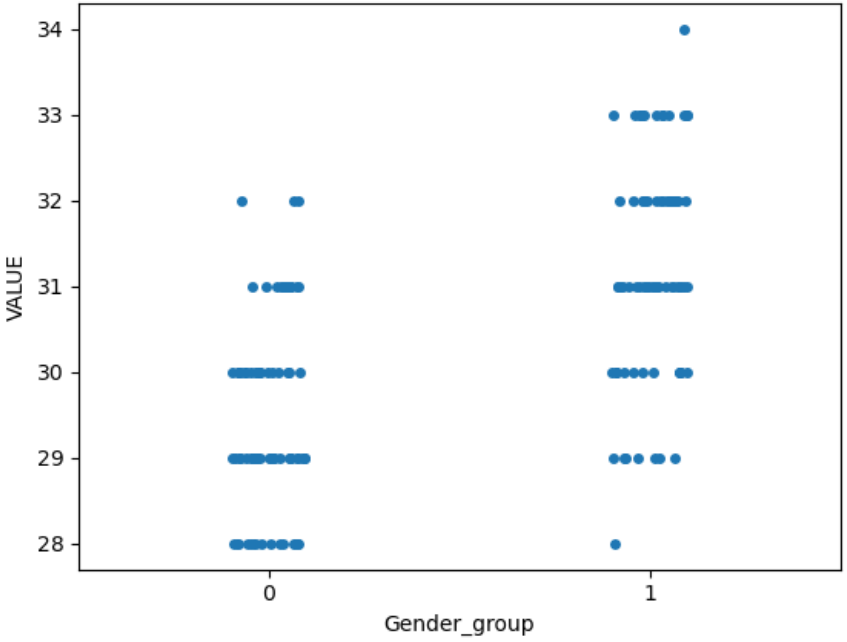
Average weekly work hours for on ‘Gender group’

0 = Female

1 = Male



By annually



By Weekly

Average weekly work hours for on ‘Gender group’

final_testing_df_output_df_AvgWeekHrsWrked_ByGender.csv

Characteristics	sum	mean	amin	median	amax	size
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 : 1.490726892045175 7

Skewnewss : 0.1316464344774555

Total size : 150

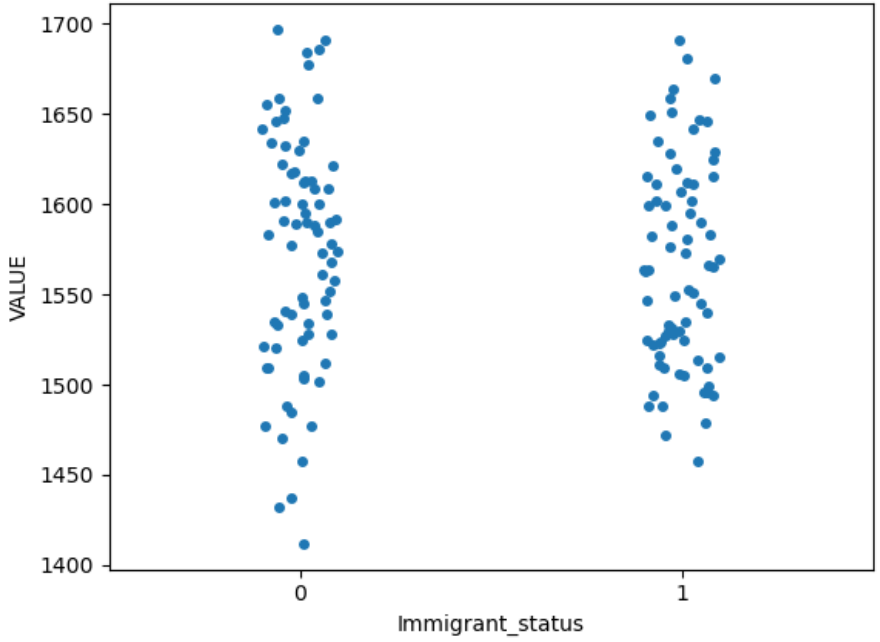
Average weekly Hour Worked based on ‘Immigrant group’

- Both immigrant and non-immigrant work highest.
- Immigrant employees work less than the non-immigrant employees.
- Lowest immigrant worked 27 hours weekly. Lowest non-immigrant worked 28 hours weekly.

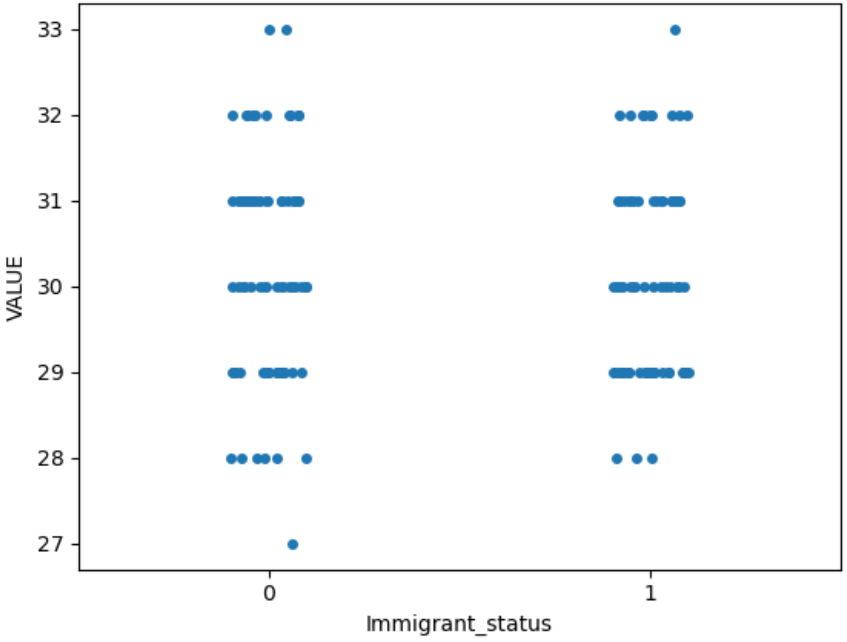
Average weekly Hour Worked based on ‘Immigrant group’

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

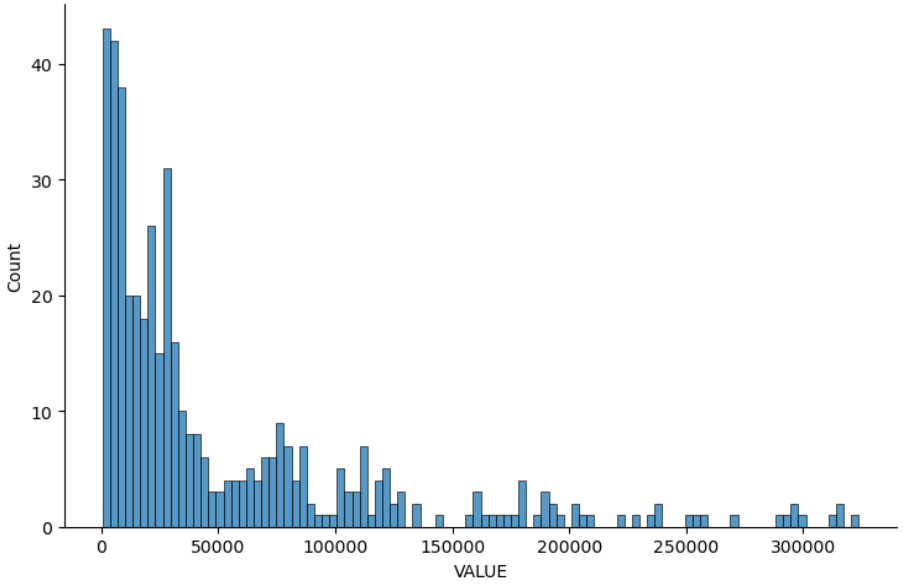
Characteristics	sum	mean	amin	median	amax	size
Immigrant employees	2270.0	30.266667	27.0	30.0	33.0	75
Non-immigrant employees	2257.0	30.093333	28.0	30.0	33.0	75

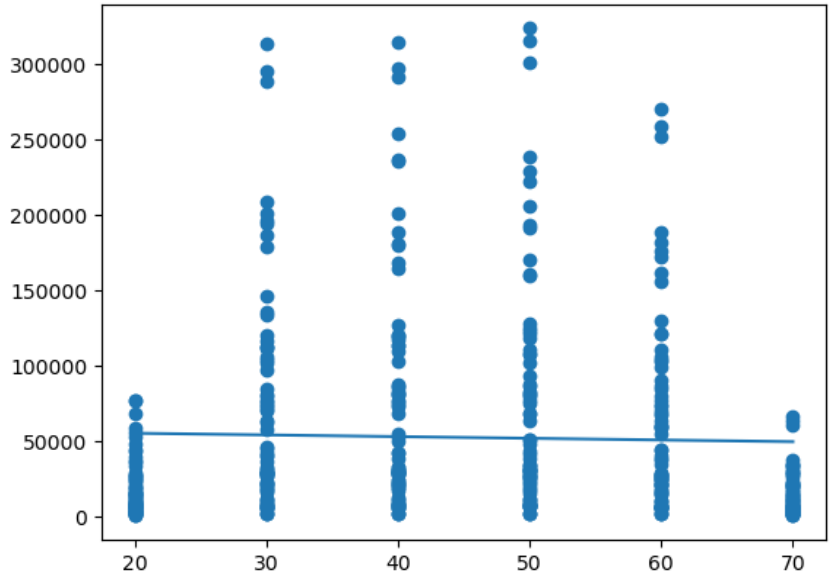
Overall,
Sum : 4527.0
Mean : 30.18
Min/median/max : 27.0 / 30.0 / 33.0
Standard Deviation : 1.2278436382536666
Skewnewss : 0.04369031127145 73
Total size : 150

Hours worked in ‘Age group’

- Employees in 20s and 70s have least number of working hours.
- Employees in 20s and 30s, 60s and 70s have most gap and highest working hours.
- Unlike annual or weekly hour worked which show distributed histogram, they are skew left.
- Most employees worked less (150000 hours) with exception of few employees.

Hours worked in ‘Age group’





Hours worked in ‘Age group’

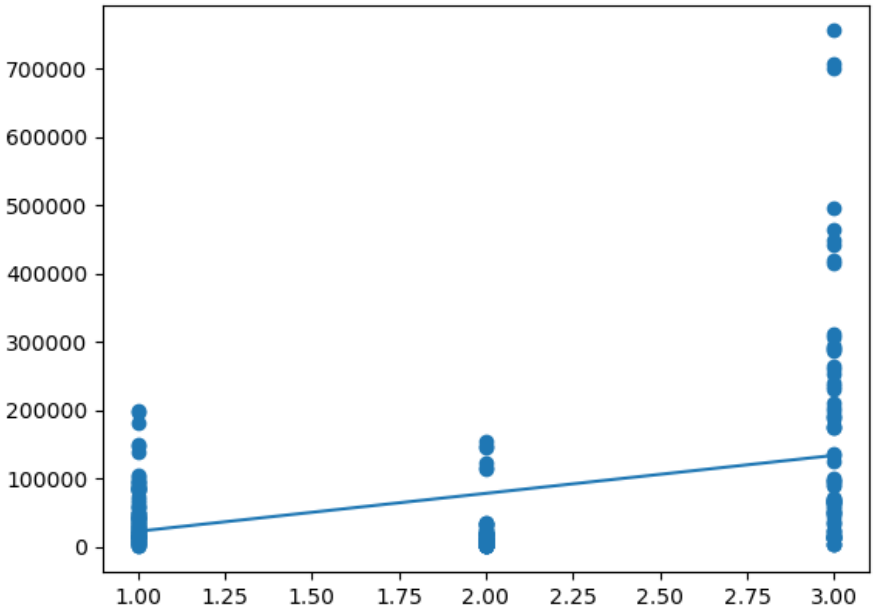
```
final_testing_df_output_df_Hrs_Wrked_ByAge.csv
                              sum          mean    amin   median      amax  \
Characteristics
15 to 24 years          1338316.0  17844.213333   497.0  11421.0   77253.0
25 to 34 years          5209479.0  69459.720000  1434.0  40393.0  313785.0
35 to 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’

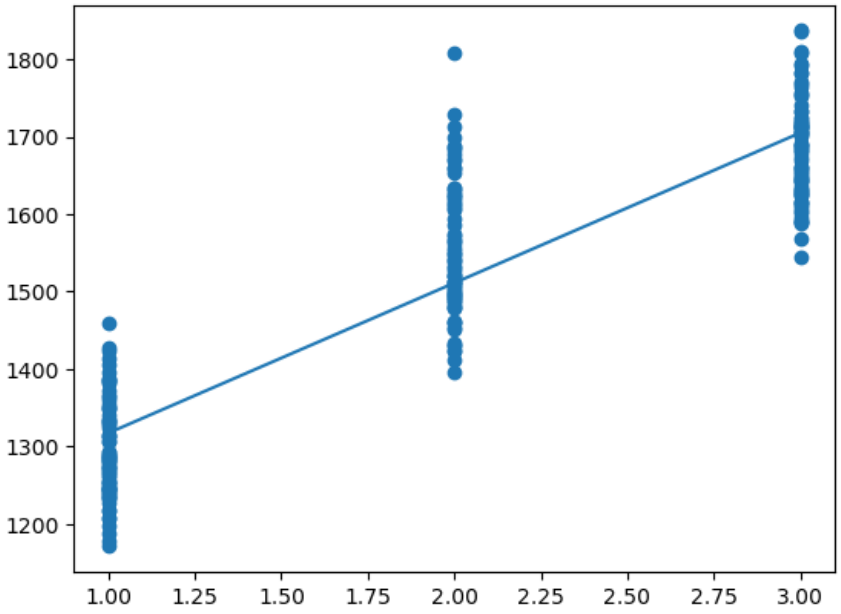
- Higher education worked more than people who are in lower education.
- For higher education, it is more clears that some of highest educations works significantly more than everyone else.
- For trade certification, they work less than those who are in highest education or lowest education.

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

Characteristics	sum	mean	amin	median	amax	size
High school diploma and less	3839870.0	51198.266667	1717.0	37663.0	199483.0	75
Trade certificate	1548281.0	20643.746667	349.0	8852.0	153953.0	75
University degree and higher	12180311.0	162404.146667	3106.0	91916.0	755243.0	75

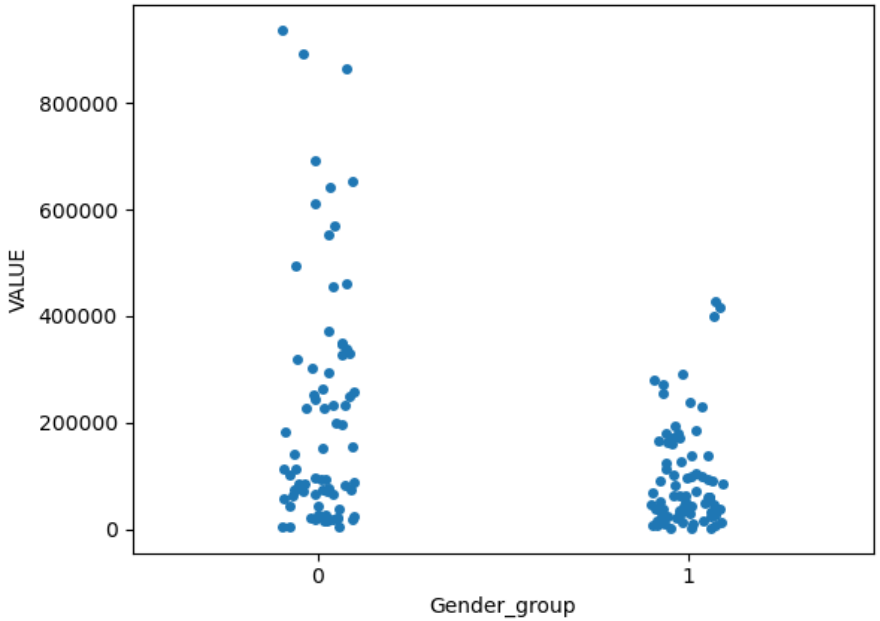
Overall,
Sum : 17568462.0
Mean : 78082.05333333333
Min/median/max : 349.0 / 31865.0 / 755243.0
Standard Deviation : 120621.30059284004
Skewnewss : 3.02056281839477 55
Total size : 225

Hour worked for ‘By Gender’

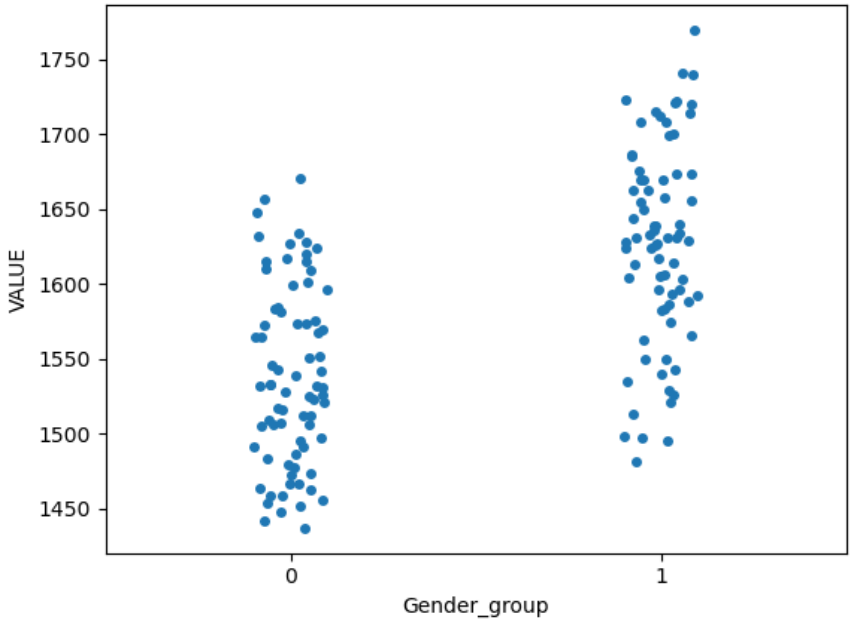
- Female employees work more than male employees.
- As references, male employees work more than female employees in term of annually or weekly.
- Both female and male employees work less as well.

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

Characteristics	sum	mean	amin	median	amax	size
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

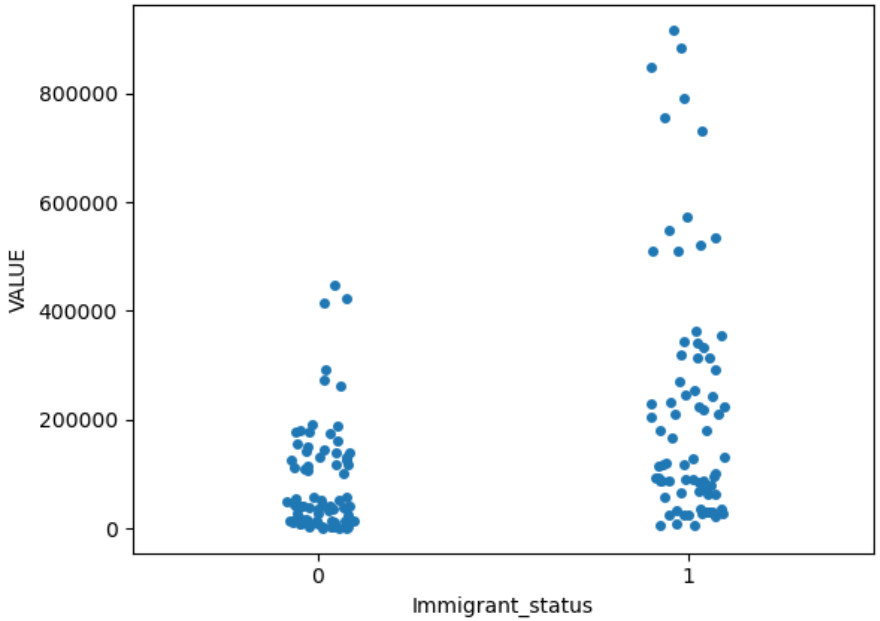
Overall,
Sum : 23545750.0
Mean : 156971.66666666666
Min/median/max : 2918.0 / 86240.0 / 937067.0
Standard Deviation : 182769.39372530495
Skewnewss : 2.0993685578631798
Total size : 150

Hour worked by ‘Immigrant status’

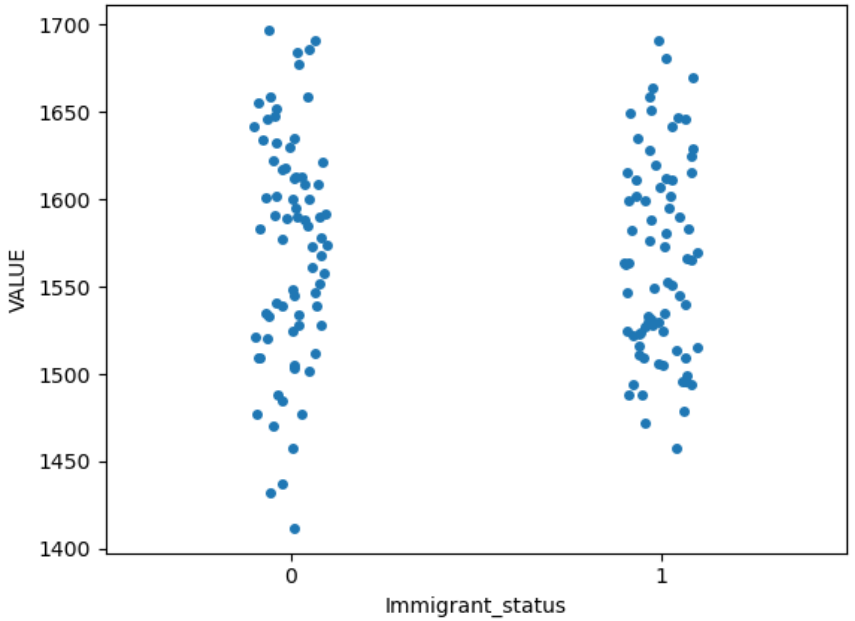
- Non-immigrant employees (about 12 employees) work more than immigrant employees, esp. highest working employees.
- For annual/weekly, immigrant employees work more.

Hour worked by ‘Immigrant status’

0 = Immigrant
1 = Non-immigrant



By hour worked



By annual hour worked

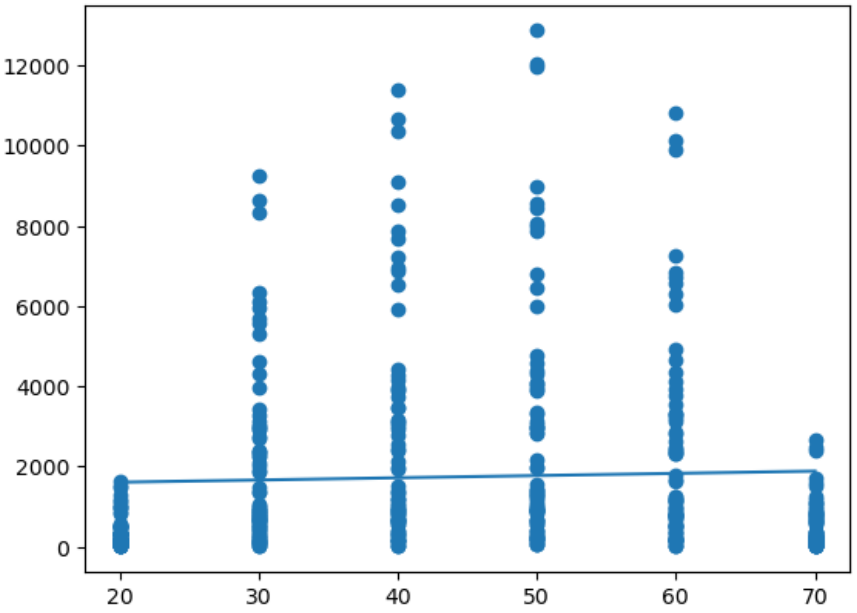
Hour worked by ‘Immigrant status’

```
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
```

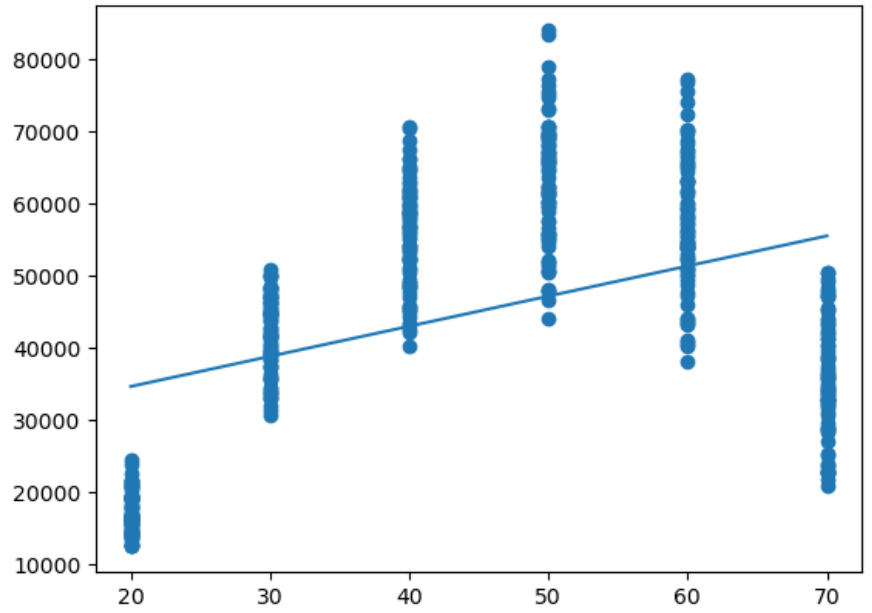
Wages and Salaries for Age group

- Employees in 20s and 70s still get the lowest pay than employees in 30s to 60s.
- Employees in 50s still get the highest salary.
- Employees between 20s and 30s, 60s and 70s still have gap between.

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

Characteristics	sum	mean	amin	median	amax	size
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

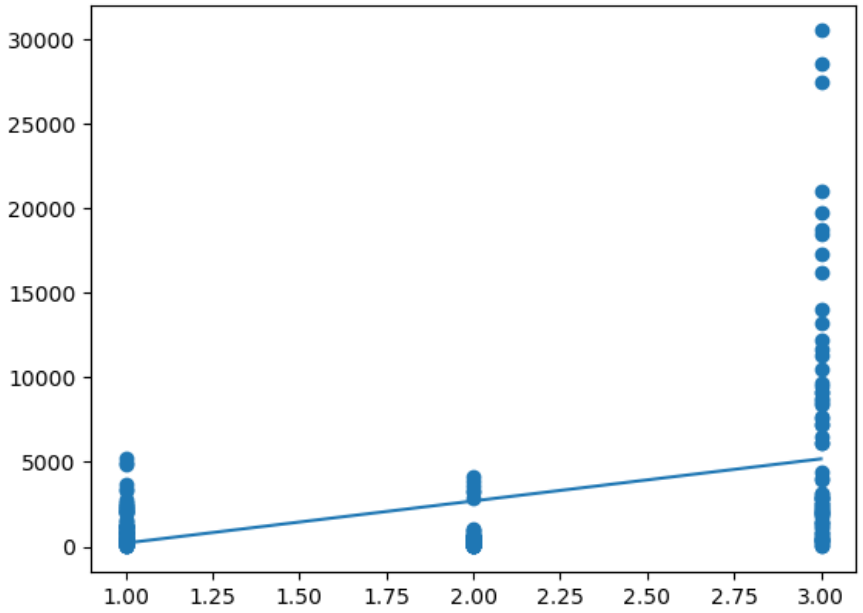
Total size : 450

Wages and Salaries by Education

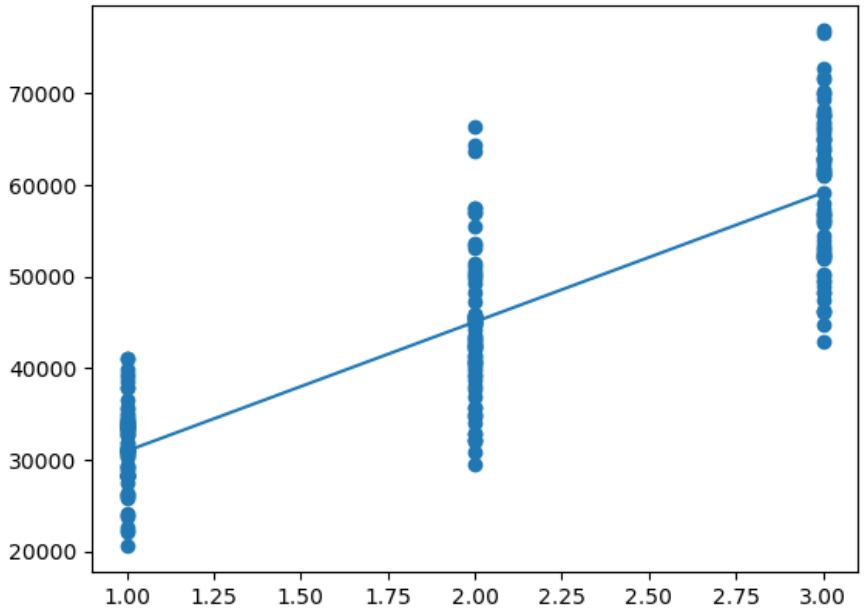
- Highest education still get the more wages than lowest education.
- Trade certificate employees get the lowest highest wages compare to those are higher or lower education.
  - Also, average trade certificate employees are also lower too.

Wages and Salaries by Education

Higher number = Higher education



By wages and salaries



By average annual wages

Wages and Salaries by Education

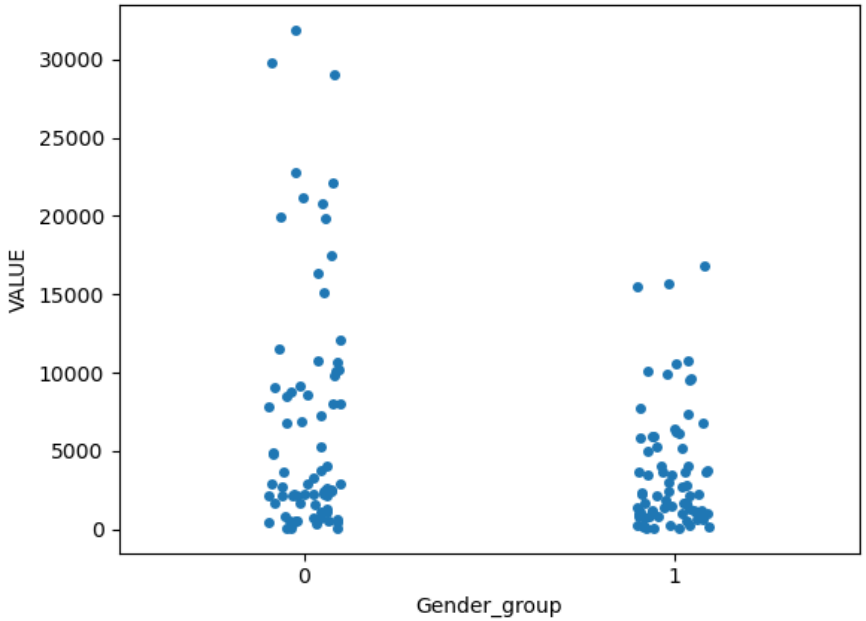
```
final_testing_df_output_df_WagesAndSalaries_ByEducation.csv
                                    sum         mean  amin  median     amax  \
Characteristics
High school diploma and less    95794.0  1277.253333  37.0   861.0   5226.0
Trade certificate               42027.0   560.360000  10.0   231.0   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

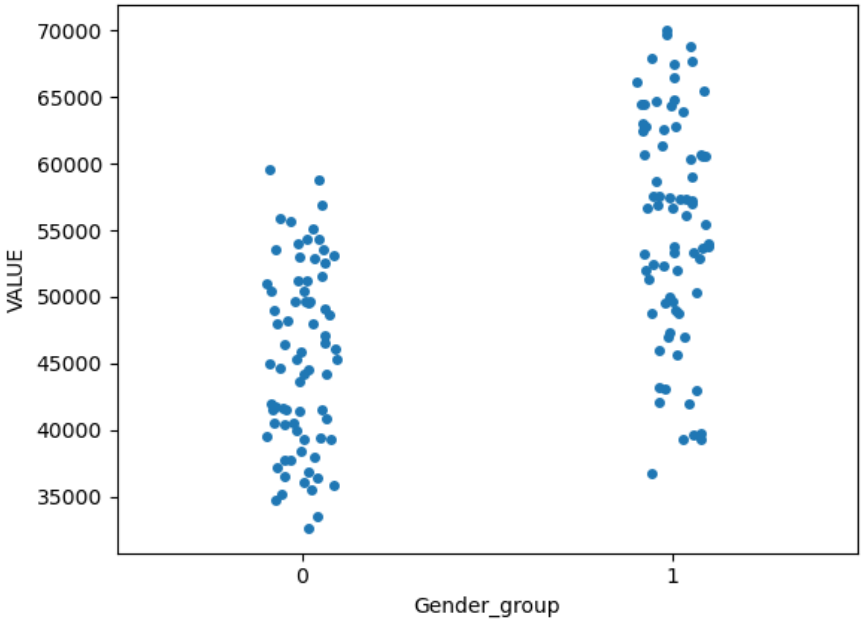
- Like counterpart (hour worked), female employees get more wages than male employees.

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

Characteristics	sum	mean	amin	median	amax	size
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

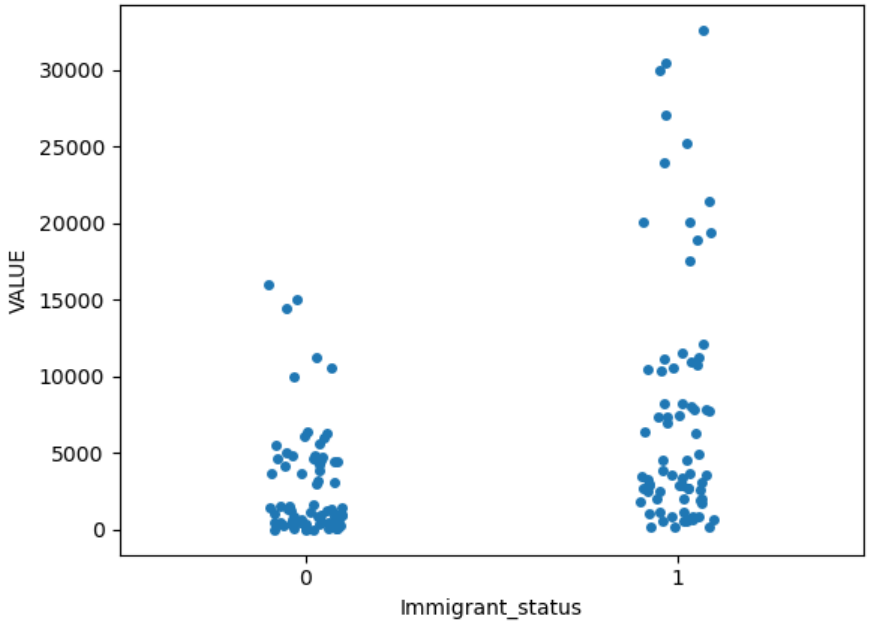
Overall,
Sum : 786615.0
Mean : 5244.1
Min/median/max : 84.0 / 2563.0 / 31892.0
Standard Deviation : 6289.8685749385895
Skewnewss : 2.0515596833116616
Total size : 150

Wages and Salaries by Immigrant

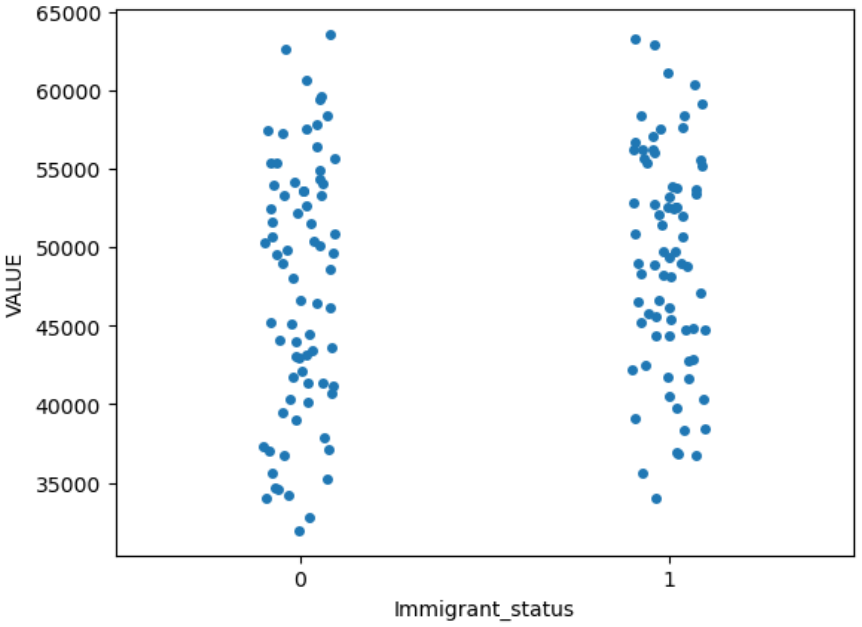
- Non-immigrant employees make more wages than those who aren’t immigrant.
- There’s small gap between non-immigrant and immigrant employees.

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

Characteristics	sum	mean	amin	median	amax	size
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 have 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 or hard to understand.
- Some indicators inside dataset are useless or repetitive or even very confused to analysis.
- 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.
- As mentions before, Some value are impossible to compare in short amount of time
  - For example, annual wages vs wages in million.
- Some indicators and columns in the dataset are impossible to understand for time given.
- As more data preparation, I am starting to seeing a lot of dataset or ten pages of scripts, I need to deal with. This increase time of my analysis also possibility of mislead inaccurate result.