

Ashvin Vaseeharan

Mr. Sufi

MDM4U1

June 7, 2019

Exposing Ontario's Public Salary Distribution Between Males and Females

Table of Contents

Introduction -----	(3)
Process -----	(4 - 6)
One Variable Statistics -----	(6)
Measure of Central Tendency -----	(7-8)
Two Variable Statistics -----	(8-9)
Males vs Females -----	(9-11)
Evaluation of Techniques -----	(12)
Conclusion -----	(12)
Work Cited -----	(13)
Appendix -----	(14-17)

Introduction

Taxes; something that is taken off of each and every paycheck. Even if it is disliked by many, taxes are essential for the government to function. Looking specifically at the provincial government, taxes help fund the salary of all the provincial public sector workers. Since it is the tax payers money that goes to fund these salaries, the public should be aware of how their money is being used. This is why the government of Ontario releases a Public Sector Salary Disclosure every year, also known as Ontario's "Sunshine List". This list contains information such as the job sector, full names, salary paid, and the employers of the government workers making \$100,000 or more. The purpose of this project is to analyse the data found on the Sunshine List and come to a conclusion if major part of the tax money is being equally distributed between male and females. I hypothesize that there are significantly more males than females on the Sunshine list. Another prediction is that men make more money than women based on this population.

Ontario's Financial Accountability Officer, Peter Weltman estimated that there were about 650,000 public servants working in Ontario in 2018. Weltman stated The Government of Ontario estimates to pays about \$41.4 Billion in salary and wages combined to all of the public servants. Using this information also found on appendix page 16, I would also like to examine the pay gap between the public servants on the list and off the list.

Process

One can easily find the Public Sector Salary Disclosure online on the Government of Ontario's website (see appendix page 15). There is a web interface where one could view the spreadsheet and look up individuals in the table. For the years of 2016, 2017 and 2018, there is an option to download the data as a spreadsheet (.xlsx file). For the 2014 and 2015 lists, I had to download the data as CSV files and use Microsoft Excel to convert them into Excel Spreadsheets. Since the 2018 Sunshine List had information about 151,198 individuals and all the data I collected from 2014-2018 totaled 635,121 individuals, I will include the links to the complete data in my works cited and placing a larger sample of the list in the Appendix.

	A	B	C	D	E	F	G	H
1	Sector	Last Name	First Name	Salary Paid	Taxable Benefits	Employer	Job Title	Calendar Year
2	Colleges	Jensen	Cheryl	\$ 337,533.24	\$ 3,408.12	Algonquin College	President	2018
3	Colleges	Brule	Claude	\$ 230,976.76	\$ 374.96	Algonquin College	Senior Vice President Academic	2018
4	Colleges	Wotherspoon	Doug	\$ 230,788.38	\$ 186.58	Algonquin College	Vice President Innovation and Strategy	2018
5	Colleges	Mcnaire	Duane	\$ 221,744.64	\$ 366.12	Algonquin College	Vice President Finance and Administration	2018
6	Colleges	Stanbra	Laura	\$ 221,744.64	\$ 366.12	Algonquin College	Vice President Student Services	2018
7	Colleges	Anderson	Scott	\$ 194,628.02	\$ 331.04	Algonquin College	Executive Director Communications Marketing External F	2018
8	Colleges	Mclester	Ronald	\$ 191,833.46	\$ 161.82	Algonquin College	Executive Director Truth Reconciliation Indigenization	2018
9	Colleges	Pollock	Lois	\$ 177,265.71	\$ 273.41	Algonquin College	Chief Digital Officer	2018
10	Colleges	Leduc	Mark	\$ 169,073.33	\$ 310.33	Algonquin College	Executive Director Academic Operations and Planning	2018
11	Colleges	Frederick	Catherine	\$ 168,347.03	\$ 131.30	Algonquin College	Vice President Human Resources	2018
12	Colleges	Mccutcheon	Diane	\$ 166,846.12	\$ 188.64	Algonquin College	Vice President Human Resources	2018
13	Colleges	Schonewille	Todd	\$ 164,726.07	\$ 304.41	Algonquin College	Director Physical Resources	2018
14	Colleges	Janzen	Christopher	\$ 163,593.81	\$ 293.89	Algonquin College	Dean Faculty Technology and Trades	2018
15	Colleges	Foulds	Barbara	\$ 160,742.83	\$ 137.91	Algonquin College	Dean Faculty Health Public Safety and Community Stud	2018
16	Colleges	Tosh	Marlene	\$ 160,044.73	\$ 0.00	Algonquin College	Chair Speciality Nursing Programs and Contract Training	2018
17	Colleges	Devey	Patrick	\$ 159,588.61	\$ 137.19	Algonquin College	Dean Centre for Continuing and Online Learning	2018
18	Colleges	Kyle	James Gregory	\$ 158,873.70	\$ 290.78	Algonquin College	Dean School of Hospitality and Tourism	2018
19	Colleges	Davies	Karen	\$ 158,873.70	\$ 290.78	Algonquin College	Dean Algonquin College In The Ottawa Valley	2018
20	Colleges	Donaldson	Dave	\$ 158,719.79	\$ 136.87	Algonquin College	Dean School of Business	2018

Figure 1

Figure 1 shows a small sample of the data I had extracted from the Sunshine list. While exploring the original data, I realised I would not be able to verify my hypothesis of the number of males on the list versus the number of females as there were no indicator on the sex of the employees. As I have background knowledge in the data analysis software Python, I decided to look online for a tool that would identify the gender based on the first name provided. I

understood that this is not 100% accurate. However after I randomly selected a small sample and reviewed the gender

identified along with the first name, I'm more confident on the methodology. I came across multiple different gender identification tools, but I decided to use the one that was created by professor Igor Korostil that was licensed by the Massachusetts Institute of Technology (MIT).

```
# Add a SEX column in each of the excel sheets (2014-2018)
for i in range(2014,2019):
    file = pd.ExcelFile(f"SunshineList{i}.xlsx")
    data = pd.read_excel(file)
    data['SEX'] = data['First Name'].apply(lambda x: gd.gender(x))
    data.to_excel(f"ModifiedSunshineList{i}.xlsx")
    print(data.SEX.value_counts())
    print(f"Year: {i}'s excel file is completed! ")
print(" ")
print("DONE EVERYTHING! ")
print(" ")
```

Figure 2

Figure 2 shows part of the script that I wrote in order to add the necessary gender column to the existing Sunshine Lists. Rest of the code can be found on appendix page 14.

Figure 3

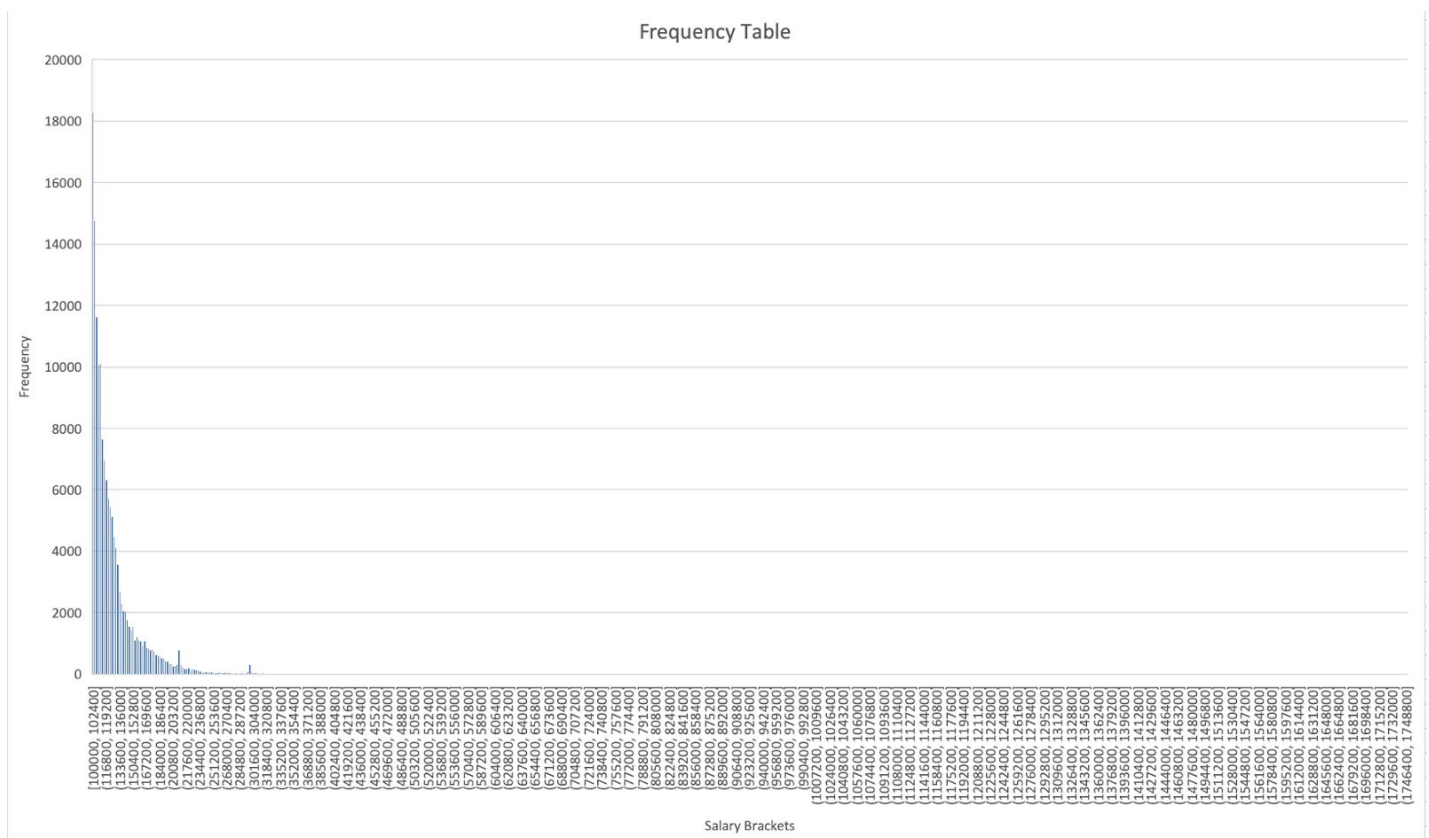
	A	B	C	D	E	F	G	H	I	J
1		Sector	Last Name	First Name	Salary Paid	Available Bene	Employer	Job Title	Calendar Year	SEX
2	0	Colleges	Jensen	Cheryl	337533.2	3408.12	Algonquin	President	2018	f
3	1	Colleges	Brule	Claude	230976.8	374.96	Algonquin	Senior Vice	2018	m
4	2	Colleges	Wotherspo	Doug	230788.4	186.58	Algonquin	Vice Presid	2018	m
5	3	Colleges	Mcnaire	Duane	221744.6	366.12	Algonquin	Vice Presid	2018	m
6	4	Colleges	Stanbra	Laura	221744.6	366.12	Algonquin	Vice Presid	2018	f
7	5	Colleges	Anderson	Scott	194628	331.04	Algonquin	Executive D	2018	m
8	6	Colleges	Mclester	Ronald	191833.5	161.82	Algonquin	Executive D	2018	m
9	7	Colleges	Pollock	Lois	177265.7	273.41	Algonquin	Chief Digita	2018	f
10	8	Colleges	Leduc	Mark	169073.3	310.33	Algonquin	Executive D	2018	m
11	9	Colleges	Frederick	Catherine	168347	131.3	Algonquin	Vice Presid	2018	f
12	10	Colleges	Mccutcheo	Diane	166846.1	188.64	Algonquin	Vice Presid	2018	f
13	11	Colleges	Schonewill	Todd	164726.1	304.41	Algonquin	Director PH	2018	m
14	12	Colleges	Janzen	Christophe	163593.8	293.89	Algonquin	Dean Facul	2018	m
15	13	Colleges	Foulds	Barbara	160742.8	137.91	Algonquin	Dean Facul	2018	f
16	14	Colleges	Tosh	Marlene	160044.7	0	Algonquin	Chair Speci	2018	f
17	15	Colleges	Devey	Patrick	159588.6	137.19	Algonquin	Dean Centr	2018	m
18	16	Colleges	Kyte	James Greg	158873.7	290.78	Algonquin	Dean Schoc	2018	m
19	17	Colleges	Davies	Karen	158873.7	290.78	Algonquin	Dean Algon	2018	f
20	18	Colleges	Donaldson	Dave	158719.8	136.87	Algonquin	Dean Schoc	2018	m

Figure 3 shows a sample of the updated Sunshine lists that includes the gender detail. I had collected enough secondary data to make conclusions at this point. Another procedure that I had considered but did not end up using was a survey. I had planned on using a Voluntary-Response Sample as my sampling technique for my survey to collect primary data. However, there was not enough data to come up with a valid conclusion. I discovered that salary is personal information that many people do not feel comfortable sharing with others.

One Variable Statistics

I first decided to analyse the Salary Paid data that I had from the Sunshine List.

Figure 4



As shown in Figure 4, creating a frequency table reveals a lot of information about the salaries paid by the Ontario Government. We can see that the salaries range from \$100,000 to over \$1.7 million. The data however is highly positively skewed. Most people on the list get paid closer to \$100,000. The frequency of the group \$100,000 to \$102,400 was 18,269. After \$300,000 the frequency of the groups ranged between 1 and 20. It is also interesting to note that only 4 employees made more than \$800,000. The salaries were \$803,552, \$962,749, \$989,303, and \$1,746,825.

Measure of Central Tendency

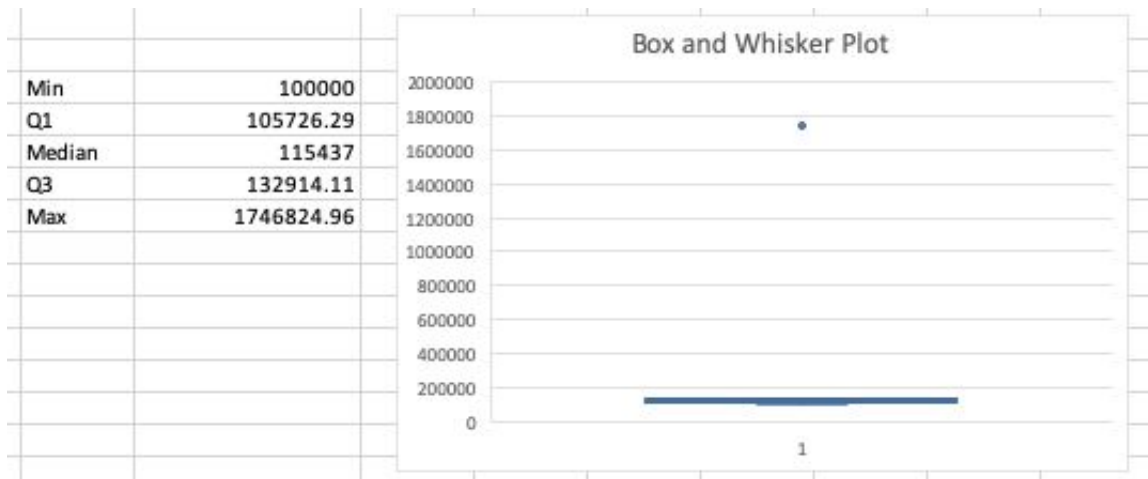
Figure 5

Mean	127168.1089
Median	115439.34
Mode	100000
Sum	19227436560
Count	151197
Max	1746824.96
Min	100000
Std Dev	36563.56017
Q1	105727.13
Q2	115439.34
Q3	132919.99
Q4	1746824.96
IQR	27192.86

As shown in figure 5, the average salary paid in the Sunshine list was \$127,168. The median of the salaries paid was \$115,439. It makes sense that the mean is higher than the median because there are a lot of high values skewing the data. The mode of the salaries was \$100,000. This shows that this is the most common pay for government workers on the list. An interesting point is that the sum of the 151,197 workers was \$19.2 Billion. Since the government stated they are spending \$41.4 Billion on 650,000 public servants, that means there is a remaining \$22.2 Billion to be spent on 498,803 workers. This equates to an average salary of \$44,507 to the workers not on the list. This is very surprising considering that the max salary paid was over \$1.7

Million to the President of Ontario Power Generation. Finally, the high standard deviation of 36,563 means that there are many data points far from the mean.

Figure 6



The Box and Whisker plot shown in Figure 6 really helps to visualize the range between the max and min salaries the government pays. You can see the IQR is \$27192, which means that most people on the Sunshine list are paid within the IQR of each other. The Box and Whisker plot once again shows much variance there is in pay in the upper end. The very high outlier makes the IQR area very small.

Two Variable Statistics

Job Sector	Count of Sector	Average of Salary Paid
Colleges	6584	116539.6173
Crown Agencies	6528	132395.6275
Government of Ontario - Judiciary	633	210918.5804
Government of Ontario - Legislative Assembly and Offices	347	133250.1159
Government of Ontario - Ministries	16544	126367.8918
Hospitals and Boards of Public Health	18317	125047.8211
Municipalities and Services	44467	120769.5948
Ontario Power Generation	7919	144819.598
Other Public Sector Employers	6906	131618.113
School Boards	23281	111672.4062
Seconded (Advanced Education and Skills Development)*	2	206710.26
Seconded (Children and Youth Services)*	6	106569.7167
Seconded (Community and Social Services)*	1	138517.41
Seconded (Community Safety and Correctional Services)*	31	120600.2739
Seconded (Education)*	116	131040.479
Seconded (Labour)*	1	114876.85
Universities	19514	153164.2403
Grand Total	151197	127168.1089



Figure 7

Figure 8

Using the pivot table in figure 7 and bar chart in figure 8, we can conclude that the different job sectors in the Ontario public sector have a significant pay variance. Some sectors such as the Seconded sectors do not have enough data points. Therefore, they are not included in the analysis. If we look at the other sectors, we see that the Judiciary Sector is paid the highest on average at \$210,918. The sector with the lowest average salary is the Public School Boards with average of \$111,672. Causes for this may have to do with qualification needed for the job, number of experienced workers and supply and demand of the workers for the particular positions.

```
m 66308
f 40774
Name: SEX, dtype: int64
Year: 2014's excel file is completed!
m 66044
f 45089
Name: SEX, dtype: int64
Year: 2015's excel file is completed!
m 69446
f 49591
Name: SEX, dtype: int64
Year: 2016's excel file is completed!
m 73195
f 52784
Name: SEX, dtype: int64
Year: 2017's excel file is completed!
m 81011
f 63029
Name: SEX, dtype: int64
Year: 2018's excel file is completed!
```

Figure 9

By comparing the results of my python scripts in figure 9, every year there are more males on the Sunshine list than females. This also supports that males are paid more than females

since there is a significantly less amount of females appearing on the list. Appendix on page 14 contains the full code and implementation.

Males vs Females

	A	B	C	D
1	Sector	Male Salary	Female Salary	Difference
2	Colleges	116562.9617	116763.6582	-200.696546
3	Crown Agencies	135612.3627	131084.4899	4527.872758
4	Government of Ontario - Judiciary	218388.2246	204897.3663	13490.85828
5	Government of Ontario - Legislative Assembly and Offices	134239.9066	132183.0794	2056.827171
6	Government of Ontario - Ministries	126704.6393	126589.9712	114.6681394
7	Hospitals and Boards of Public Health	139456.4127	120113.089	19343.32369
8	Municipalities and Services	121260.7953	119668.6285	1592.166786
9	Ontario Power Generation	146857.6402	136095.598	10762.04218
10	Other Public Sector Employers	139619.4921	125700.0033	13919.48886
11	School Boards	112685.0387	111087.7088	1597.329894
12	Universities	158296.0976	145705.8217	12590.27589

Figure 10

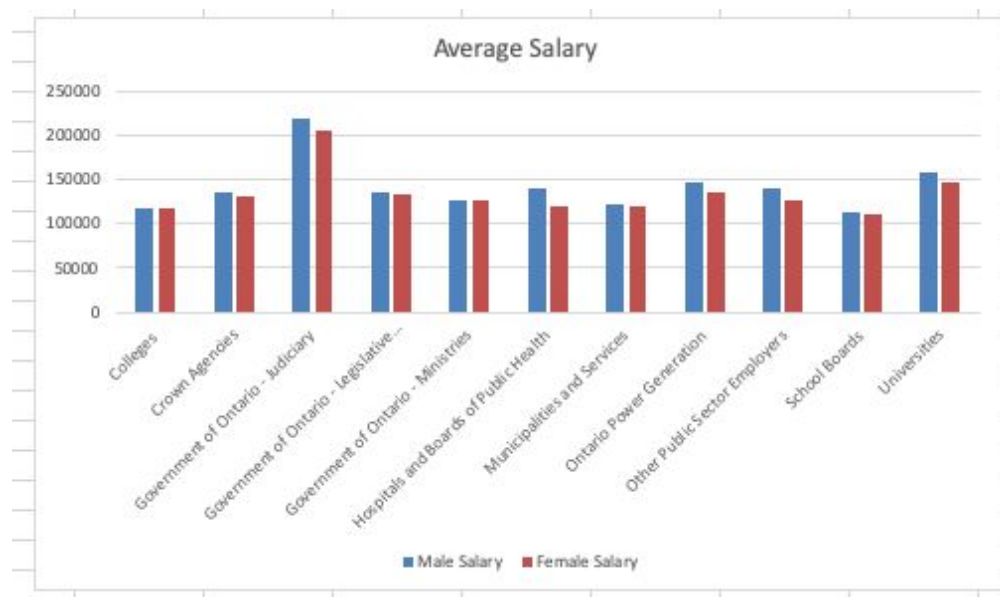


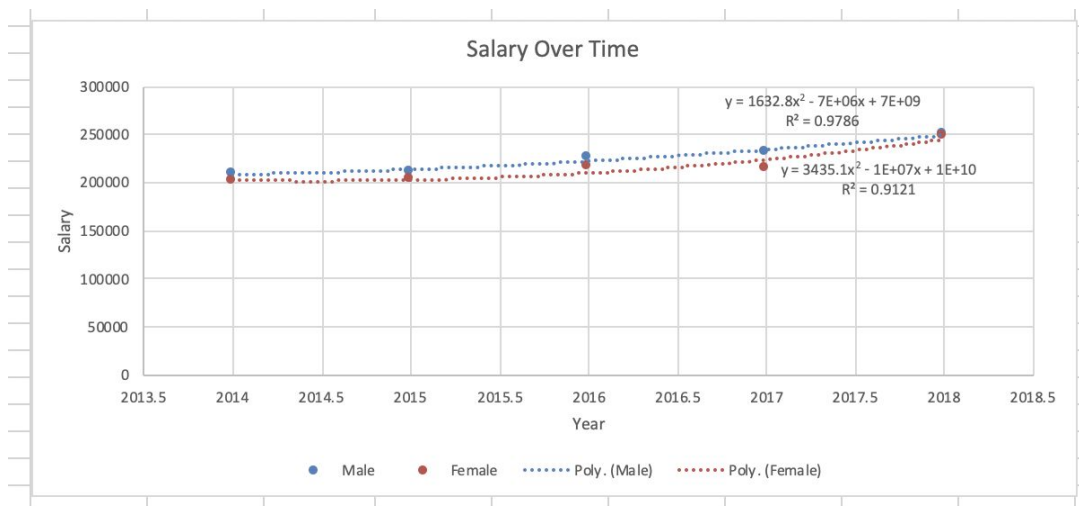
Figure 11

Now when comparing the average salaries by each sector between males and females using figures 10 and 11, we start to notice some differences. In 10 out of the 11 public sectors,

females are paid less on average than males. Only in Colleges females are paid about \$200 more on average than their male co-workers. The gap is the greatest in the Judiciary Sector, Other Public Sector and the University sector where women earn about \$13,000 less than their male counterparts. The pay gap may be due to different roles and job positions men and women have in these sectors, but even in sectors where there are traditionally more women than men such as the Hospital Sector, women still make less than males. This evidence can be used to conclude that women are unfairly paid less than males in today's public sector.

Year	Male	Female
2014	209055	201177
2015	211192.5	203314.5
2016	225979.06	216042.81
2017	232320.51	215371.1
2018	250110.03	248253.96
R =	0.97235423	0.89195916

Figure 13



The final analysis is to compare 2 workers; one male, one female and their salary changes over time. To keep everything consistent, I made sure I chose one female and one male from the

University Sector randomly. Furthermore both the subjects are Professors of Law at the University of Toronto. As shown in figure 12, I also tried to make sure they had similar starting salaries. When performing a linear regression on the data, we can see the Correlation Coefficient is 0.9724 for the male and 0.8920 for the female. Both of these values are classified as having strong positive linear correlation but we can say that the male data has a stronger correlation therefore fitting the linear equation better. The linear equation for the male was $y = 10324x - 2E+07$ while the linear equation for the female was $y = 10621x - 2E+07$. Surprisingly the equation for female has a higher slope meaning if the trend continues, the female will eventually surpass the male's salary. While looking at the data closely, it seems that a non-linear regression will better model the growth as shown in figure 13. For both the male and female subjects, performing a polynomial regression best fits the data. The Coefficient of Determination for the male was 0.9786 while the female subject had 0.9121. The equations of the line of best fit for the male and female respectively were: $y = 1632.8x^2 - 7E+06x + 7E+09$ and $y = 3435.1x^2 - 1E+07x + 1E+10$. To see if the trend will continue we can perform extrapolation on the data for this year by substituting 2019 in for x. I think that for both the male and female data there is a Cause and Effect relationship because a change in time produces a change in salary (because more experience has been gained teaching).

Evaluation of Technique

I think that my report is free from bias because I used the information provided by the Ontario Government. For most of my analysis I used all of the data available to be more accurate so sampling techniques were not needed. To perform my over time analysis I used a Multi Stage Sampling Technique so I would choose Law professors from the University of Toronto. Possible limitations of this project is that I am not considering the population who works in the public sector with salary below \$100,000. A question of validity would be the accuracy of the Python gender identification tool. Since there was no other reasonable means of finding out the gender of each person, I can justify that using the Python tool was the best alternative.

Conclusion

In summary, the government publicly disclosing the Sunshine list allows for the population to perform their own critiques about the government. I found that there is an unnecessary gap between the salaries of male and female employees that should be looked into. There were more males on the list compared to females for each of the years that I had analysed. Also in most sectors, males made more income than their female coworkers. Hopefully, this trend goes away and we see more equality in the public sector in the future. By looking at the one variable statistic results, I would suggest that the government should reduce the salary of the outliers and allocate some of the total funding toward the workforce making under \$100,000. There is no need for a public serving CEO to be making over \$1.7 million unless he can justify his worth. The lower paid workers are hardworking as well. Therefore, so there needs to be a more even distribution of the funding.

Works Cited

- Gender. (n.d.). Retrieved from <https://pypi.org/project/gender/>
- Public sector salary disclosure 2018. (n.d.). Retrieved from <https://www.ontario.ca/page/public-sector-salary-disclosure-2018-all-sectors-and-second-ed-employees>
- Public sector salary disclosure 2017. (n.d.). Retrieved from <https://www.ontario.ca/page/public-sector-salary-disclosure-2017-all-sectors-and-second-ed-employees>
- Public sector salary disclosure 2016. (n.d.). Retrieved from <https://www.ontario.ca/page/public-sector-salary-disclosure-2016-all-sectors-and-second-ed-employees>
- Public sector salary disclosure 2015. (n.d.). Retrieved from <https://www.ontario.ca/page/public-sector-salary-disclosure-2015-all-sectors-and-second-ed-employees>
- Public Sector Salary Disclosure Act: Disclosures for 2014. (n.d.). Retrieved from <https://www.ontario.ca/page/public-sector-salary-disclosure-act-disclosures-2014>
- ArtusoMore, A., & Artuso, A. (2018, July 25). Ontario government employs 11% of province's salaried workers: FAO. Retrieved from <https://torontosun.com/news/provincial/ontario-government-employs-11-of-provinces-salaried-workers-fao>
-

Appendix

```
In [1]: # DATA CULMINATING
```

```
In [2]: # Install Needed Packages
!pip3 install gender
!pip3 install gender-guesser
```

```
Requirement already satisfied: gender in /Users/vaseeharan/anaconda3/lib/python3.7/site-packages (0.0.12)
Requirement already satisfied: unicode in /Users/vaseeharan/anaconda3/lib/python3.7/site-packages (from gender) (1.0.23)
Requirement already satisfied: gender-guesser in /Users/vaseeharan/anaconda3/lib/python3.7/site-packages (0.4.0)
```

```
In [3]: # Python Libraries Used
import numpy as np
import pandas as pd
from gender import GenderDetector
import gender_guesser.detector as gender2
```

```
In [4]: # Create Gender Detectors
gd = GenderDetector()
d = gender2.Detector()
```

```
In [5]: # Test Detectors with Generic Names
names = ["Bob", "Anna", "John", "Smith", "Lucy", "David", "Mike", "Sally"]
genders1 = []
genders2 = []

for name in names:
    genders1.append(gd.gender(name))
    genders2.append(d.get_gender(name))

print(genders1)
print(genders2)

# Can conclude that both gender detectors are accurate
# Can conclude that gender detector from package "gender" is more consistent
# gd.gender() will be used

['m', 'f', 'm', 'm', 'f', 'm', 'm', 'f']
['male', 'female', 'male', 'unknown', 'female', 'male', 'male', 'female']
```

```
In [6]: %timeit -n 1 -r 1

# Add a SEX column in each of the excel sheets (2014-2018)
for i in range(2014,2019):
    file = pd.ExcelFile(f"SunshineList{i}.xlsx")
    data = pd.read_excel(file)
    data['SEX'] = data['First Name'].apply(lambda x: gd.gender(x))
    data.to_excel(f"ModifiedSunshineList{i}.xlsx")
    print(data.SEX.value_counts())
    print(f"Year: {i}'s excel file is completed! ")
print("")
print("DONE EVERYTHING! ")
print("")

m    66308
f    40774
Name: SEX, dtype: int64
Year: 2014's excel file is completed!
m    66044
f    45089
Name: SEX, dtype: int64
Year: 2015's excel file is completed!
m    69446
f    49591
Name: SEX, dtype: int64
Year: 2016's excel file is completed!
m    73195
f    52784
Name: SEX, dtype: int64
Year: 2017's excel file is completed!
m     81011
f     63029
```

```
Year: 2017's excel file is completed!
m      81011
f      63029
Name: SEX, dtype: int64
Year: 2018's excel file is completed!

DONE EVERYTHING!

1h 56min 21s ± 0 ns per loop (mean ± std. dev. of 1 run, 1 loop each)
```

*Excel Files containing data are over 500pgs long. Links added below to entire excel files

- <https://www.ontario.ca/page/public-sector-salary-disclosure-2018-all-sectors-and-second-ed-employees>
- <https://www.ontario.ca/page/public-sector-salary-disclosure-2017-all-sectors-and-second-ed-employees>
- <https://www.ontario.ca/page/public-sector-salary-disclosure-2016-all-sectors-and-second-ed-employees>
- <https://www.ontario.ca/page/public-sector-salary-disclosure-2015-all-sectors-and-second-ed-employees>
- <https://www.ontario.ca/page/public-sector-salary-disclosure-act-disclosures-2014>

Ontario government employs 11% of province's salaried workers: FAO

Antonella Artuso
More from Antonella Artuso

Published:
July 25, 2018

Updated:
July 25, 2018 1:34 PM EDT

Filed Under:
Toronto SUN > News >
Ontario



RELATED



The Ontario government employs more than one in 10 salaried workers in the province, paying them \$41.4 billion in salary and wages.

Peter Weltman, Ontario's new Financial Accountability Officer (FAO), released those facts in a commentary Wednesday, raising serious questions about Premier Doug Ford's stated plan to balance the province's books without laying anybody off.

Finance Minister Vic Fedeli said the government is currently contracting a line-by-line audit of government spending that will help find efficiencies without costing public sector workers their jobs.

"As an example, the auditor general reviewed only 14 agencies in December of last year, and found \$1 billion in savings ... and not one single job would be cut to save the billion dollars that she outlined," Fedeli said. "If you took that and just stretched that right across the broader public sector service, you can see that there are going to be efficiencies found that don't affect the employees at all."

The minister gave the example of changes his government brought in for OHIP+, the public prescription drug program for young people, which he said saved tens of millions of dollars by involving private sector plans but still leaving all those under age 25 covered.

The FAO commentary estimates there are about 650,000 public servants in Ontario this year, of which 88,000 are employed in public administration in government ministries.

Paid for largely through provincial transfer payments, another 341,000 workers are in the education sector, while 221,000 are in the hospital sector, the FAO reports. That represents about 11% of salaried employees in Ontario, the FAO says.

“The average earnings for Ontario public sector workers is estimated to be \$1,227 per week in 2018 or roughly \$63,800 annually,” the FAO says. “Among public sector workers, Ontario public administration employees have the highest average weekly pay at \$1,598 per week, followed by hospital and education workers at \$1,158 and \$1,176 per week, respectively.”

Public wages exploded between 2001-09, rising at the rate of 6.7% on average every year.

After the recession of 2009, and until about two years ago, government restraint kept a lid on increases, but that trend has changed, the FAO found.

“More recently, growth in Ontario’s employee compensation expense has accelerated sharply, averaging 4.4% in 2017 and 2018,” the report says.

Collective agreements inked recently point to an acceleration in these wages over the coming years, Weltman reported.

As well, a competitive employment market puts upward pressure on public sector wages, his commentary says.

aartuso@postmedia.com

 **20** Comments
Share your thoughts