## horizontal line



ICS4U COMPUTER STUDIES SAS PROJECT: REPORT

30.05.2017

**─**

BY ARYAN KUKREJA AND CHRIS CHANCE

Earl Haig Secondary School

100 Princess Avenue

North York, ON, M2N3R7

Table of Contents

[Title Page 0](#_Toc483867139)

[Introduction 1](#_Toc483867140)

[Major Goals of Project 3](#_Toc483867141)

[Pre-research Hypothesis 3](#_Toc483867142)

[Discussion - Analysis of Data Sets 4](#_Toc483867143)

[Beginner Questions 4](#_Toc483867144)

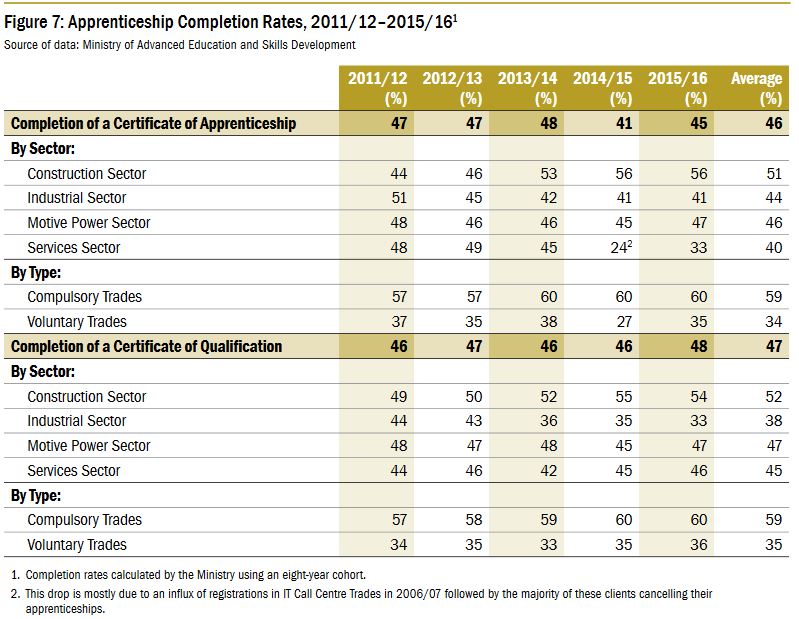
[Intermediate Questions 7](#_Toc483867146)

[Conclusion 11](#_Toc483867148)

Sources...................................................................................................................12

# Introduction

The Auditor General’s Report for 2016 revealed that over 5 fiscal years starting from 2011-2012 to 2015-2016, only 47% of apprenticeship trainees got a Certificate of Apprenticeship, and only 46% of them got a Certificate of Qualification. Power line technicians had the highest completion rate of 75%, while the lowest rate of completion was the IT customer service agents had the lowest rate of completion at 4%. A full breakdown of the completion rates is provided in the chart below:



Even though improving this rate has been coined as a top priority in the 2016 report, the task was proved to be much more tiresome since there has been no data collection through exit surveys on *why* the apprentices leave their program midway. Nevertheless, the task of improving the completion rate has been handed over to MAESD (Ministry of Advanced Education and Skills Development).

As a summative project for high school students, 4 datasets that provide information on apprenticeships have been provided to students so that they can analyze them and find some reason for the low rates of apprenticeship completion.

This project was done in a joint collaboration between Aryan Kukreja and Chris Chance, and the aim of this project was to find a plausible explanation for the high dropout rates from apprenticeship programs across the province.

# Major Goals of Project

1. Import all 4 datasets (.csv format) into the SAS Environment
2. Compress the data so as each apprenticeship trade only has 1 observation in a dataset (this pertains to the JVWS dataset)
3. Sort the data, and output it neatly into a .csv file for export into Excel
4. Filter the data based on further analysis.
5. Answer questions provided in problem statement.
6. Answer main project question.
7. Create final report that covers all findings and research
8. Create presentation that goes over entire report, and present to class.

# Pre-research Hypothesis

Prior to conducting any research, we came up with the following hypothesis:

* *Many apprentices may be dropping out of their particular trade training because their hourly wage relative to the number of hours they put in every week may not be rewarding enough to them, or they fear that the job market for their field is shrinking, which could have them jobless.*

Based on information obtained directly from the Auditor General’s Report for 2016, the construction sector had a much higher graduate turnout than other sectors such as the service sector. Given that the pay scales in the services sector can vary (not every employee in the service industry would be getting a good tip to offset the lower minimum wage), and the work hours can be quite frustrating, it can be assumed that many apprentices get second thoughts about continuing the program. On the other hand, employees in the construction sector have fixed timings. They are unable to work in harsh winter conditions most of the time, and given the construction boom taking place across Ontario’s metropolitan cities, there is no shortage in jobs for skilled workers in the construction sector; hence, leading to the high completion rate.

# Discussion - Analysis of Data Sets

The following are a list of questions provided in the Problem Statement of the project. It is subdivided into 3 categories that gauge the difficulty level of each question: Beginner, Intermediate, and Advanced. The purpose of these questions is to provide a pathway that will assist in answering the main purpose of the project, and they may open up a possible reasoning as to why the completion rates of apprenticeship programs is so low. In order to answer these question, the SAS programming language was used to analyze the data to get the relevant results.

## Beginner

1. How many Apprentices are there in Ontario?

*In order to answer this question, the SUM command in the PROC PRINT procedure was implemented in the unfiltered output of the RAIS.csv dataset in the SAS environment.The SUM command provided the totals for the variables: total\_cont, total\_comp, total\_NA, total\_reg and total. The results of compiling this code gave the following results:*

Answer:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **2011** | **2012** | **2013** | **2014** | **2015** |
| **Not Applicable for Direct Qualification** | 0 | 0 | 0 | 0 | 0 |
| **Continuing Program** | 134,676 | 143,498 | 121,788 | 125,231 | 94,767 |
| **Completed Program** | 10,158 | 11,064 | 17,706 | 11,619 | 9,915 |
| **Discontinued Program** | 8,470 | 9,424 | 32,666 | 7555 | 28,406 |
| **Total Individuals** | 153,767 | 164,426 | 172,559 | 144,827 | 133,449 |
| **Total Excluding Discontinued** | 144,834 | 154,562 | 139,494 | 136,850 | 104,682 |

# 2. What is the mean age of participants in the program?

*In order to answer this question, the PROC MEANS procedure was implemented on the mean\_age variable from the RAIS.csv dataset. The mean value of the PORC MEANS output is given below.*

Answer:

|  |  |
| --- | --- |
| **Mean Age of Individual in Sampled Record** | |
| **2011** | 30.93 years old |
| **2012** | 31.14 years old |
| **2013** | 31.56 years old |
| **2014** | 31.61 years old |
| **2015** | 31.52 years old |

3. What trade has the most women registered?

Answer:

*To get both the most trade where females choose to enroll and the remaining individuals in that apprenticeship we had to split the data into two DATA SETS, one for male and one for female. Once we do this, we sort them out in the PROC SORT step to both variables: regyr - Registration Year and appr\_trade\_name - Apprenticeship Trade Name. Then merge the two for an easy read on the comparison of which has the highest count of registered females in a specific trade and for males as well.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Most Registered Trade of Female Apprenticeship** | | | | |
| **Year** | **Name of Trade** | **Total Females** | **Remaining Total** | **Total** |
| **2011** | Hairstylist | 9602 | 1257 | 10859 |
| **2012** | Hairstylist | 9592 | 1262 | 10854 |
| **2013** | Hairstylist | 8421 | 1121 | 9542 |
| **2014** | Hairstylist | 8113 | 1077 | 9190 |
| **2015** | Hairstylist | 7417 | 936 | 8353 |

4. What is the most popular trade?

Answer:

*In order to calculate this, a new variable was created that only tool in the number of individuals that completed a program or are still in the program. The results of compiling the code provided the following results:*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Most Popular Trade by Persons Continuing/Completed Program** | | | | |
|  | **Name of Trade** | **Total Completed** | **Total Continuing** | **Total** |
| **2011** | Automotive Service Technician | 1,212 individuals | 10,644 individuals | 11,624 individuals |
| **2012** | Automotive Service Technician | 1,338 individuals | 11,215 individuals | 12,201 individuals |
| **2013** | Electrician - Construction and Maintenance | 2,581 individuals | 9,265 individuals | 10,230 individuals |
| **2014** | Automotive Service Technician | 1,221 individuals | 9,472 individuals | 10,326 individuals |
| **2015** | Electrician - Construction and Maintenance | 1,314 individuals | 7,258 individuals | 9,258 individuals |

5. Which trade has the highest rate of completion?

Answer: *In order to calculate this, we derived a formula that used the variables total\_reg and total\_comp, to find the percentage of students that complete their program against those that registered in it. The formula was added in the DATA step, and a PROC MEANS procedure was implemented to find the highest rate of completion. A scale of 0% to 100% was used. It is important to note that this value does not take into account the number of individuals currently registered in an apprenticeship. The results per year are as follows:*

|  |  |  |
| --- | --- | --- |
| **Highest Apprenticeship Completion Rates by Year** | | |
|  | **Name of Apprenticeship** | **% Completion Rate** |
| **2011** | Construction Craft Worker | 77.8% |
| **2012** | Electrician (Signal Maintenance) (TTC) and Light Rail Overhead Contact Systems Line-Person | 50.0% |
| **2013** | Ironworker - Structural and Ornamental | 66.7% |
| **2014** | Information Technology - Contact Centre Technical Support Agent | 50.0% |
| **2015** | Tractor-Trailer Commercial Driver | 40.0% |

## Intermediate

1. Are age and gender groups equally represented across trades and/or sectors?

Answer:

*What we had to do in order to answer this question is to look at both the male and female mean age for each apperienship trade though years 2011-2015 on the RAIS\_supressed.csv file. There are two ways to see if age and gender groups are equally represented across trades or sectors. Looking at the overall mean in PROC MEANS and splitting the two genders into their own variable (ex. instead of just mean\_ age you split it into mean\_age\_F and mean\_age\_M ). This way we can see the mean age for both genders and see if their similar toward each other. For the register total for each gender we also just MERGE the two separated DATA SETS (A similar method with the female trade question) to use in our PROC PRINT STEP to format a report showing both registered gender as well as the total with the use of SUM.*

*Age and gender does have a similar mean towards each other, both in PROC MEANS and observation from the printed table. Even the min and max values seem the same despite the higher count of males. There is also one problem with the results as most of the mean age for some registered apprentices haven’t provide about half the information for only on females. Which greatly affects the question ‘if these groups are equally represented across these trades’.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Age & Gender Groups in Registered Trade** | | | | | |
|  | Female Mean Age | Male Mean Age | Total Registered Female | Total Registered Male | Total Registered |
| 2011 | 32 Years old | 30 Years old | 36,990 | 117,264 | 154,254 |
| 2012 | 32 Years old | 31 Years old | 40,166 | 124,524 | 164,690 |
| 2013 | 32 Years old | 31 Years old | 40,981 | 131,885 | 172,866 |
| 2014 | 33 Years old | 31 Years old | 34,484 | 110,470 | 144,954 |
| 2015 | 32 Years old | 31 Years old | 29,248 | 105,156 | 134,404 |

2. How have registrations changed over time?

Answer:

*Before this question is answered, it is important to note that 2 outliers in the RAIS.csv dataset were observed. The first was in the reporting year 2013, and the second one in 2015. The reason for the outlier total registration values in the fiscal year 2013-2014 is due to the transfer in responsibility of maintaining and developing the Apprenticeship Training and Curriculum Standards from the MTCU (Ministry of Training, College and University) to the OCT (Ontario College of Trades). Meanwhile, the outlier total registration values in 2015-2016 is because the provincial budget for this fiscal year introduced drastic cuts to the Ontario Apprenticeship Tax Credit, provided to small and medium-sized businesses for taking in and training new apprentices. This cut reduced the incentive for many businesses to hire new apprentices, which could also have led to the low completion rates for 2015-2016.*

*Consider the table below from Beginner Question 1. The total number of individuals who registered into an apprentice in 2011 was 153,767. In 2012, the value went up to 164,426. In 2014, the registered number of apprentices goes down further from the sharp drop in 2013 from 139,494 to 136,850. In 2015, it witnesses another sharp drop to 104,682.*

*Therefore, even if we do not consider the years 2015-2016 due to their outlier data, we can safely say that the number of apprentices who did not discontinue their apprenticeship program has been dropping slowly.*

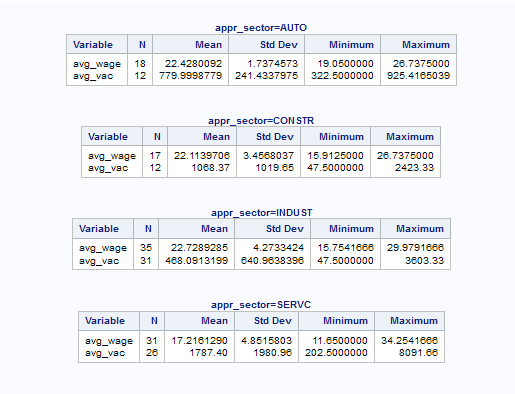
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **2011** | **2012** | **2013** | **2014** | **2015** |
| **Not Applicable for Direct Qualification** | 0 | 0 | 0 | 0 | 0 |
| **Continuing Program** | 134,676 | 143,498 | 121,788 | 125,231 | 94,767 |
| **Completed Program** | 10,158 | 11,064 | 17,706 | 11,619 | 9,915 |
| **Discontinued Program** | 8,470 | 9,424 | 32,666 | 7555 | 28,406 |
| **Total Individuals** | 153,767 | 164,426 | 172,559 | 144,827 | 133,449 |
| **Total Excluding Discontinued** | 144,834 | 154,562 | 139,494 | 136,850 | 104,682 |

# 3. How do wages vary by sector?

Answer:

*In order to answer this question, 2 datafiles were merged; the OCTAA.csv file and the JVWS.csv datafile. This way, we could compare the wages data found in the JVWS datafile to the Various sectors each individual trade fell under. It is important to note that only 156 of the 500 records of the JVWS file could be analyzed, as the OCTAA.csv file only has 156 apprenticeship trade codes to match with those in the JVWS file. The next step we took was to sort the output by every sector; this way, we could calculate the mean pay per sector and compare it with the other sectors.*

*Here is the output from our code for the average wages (and job vacancies) per sector. The maximum and minimum values per sector are also listed in it:*



# Compiling Data Gathered from Discussion to Verify Hypothesis

*Based on this table, even though the industrial sector (INDUST) has the highest average wage for the years 2015 and 2016 combined, it has the least number of job opportunities when compared with the other mean job vacancies of other sectors. The Service sector has the lowest mean pay, even though its maximum pay in one record is the highest over all 4 sectors, and it has the highest mean number of jobs in comparison to the other trades. The Automotive sector (AUTO) and the Construction sector (CONSTR) have a healthy number of jobs and average mean hourly wages.*

*These observations tie in with our hypothesis. The Service sector, with its ill-suited working hours and its extreme-ended pay scale can make it quite unattractive for apprentices, which is why it has a high dropout rate. The Industrial sector, which is well known today for moving into automation, now just in Ontario, but across Canada and even the world, has witnessed a decline in demand for human labour. This is undermined by the fact that it has the lowest number of jobs in comparison to other sectors, again making it an unattractive future career for many apprentices, even for those that are enrolled in a training program.*

# Conclusion

## Hypothesis Verification: True or False?

Based on the reasoning in the previous section, we believe that out hypothesis is indeed verified. Sectors that were found to have a low completion rate (Industrial and Services sectors respectively) did indeed either have a low number of mean jobs or an over-stressful work culture, without a rewarding incentive to keep trainees attached to training. On the other hand, sectors such as the Automotive sector (the Automotive Service Technician was the most popular trade amongst over the years 2011, 2012, and 2014, only excluding 2013 and 2015), and the Construction sector (Construction Craft Worker and Ironworker had the highest completion rates for the years 2011 and 2013 respectively, based on other questions) had a high completion rate, as mentioned in the Auditor General’s Report for 2016. This is undermined by the relatively high and non-extreme-ended pay scales (The maximum and minimum pay scales were relatively high for both of these sectors when analyzed), and their better work conditions as discussed above.

While we have found out hypothesis to be true, we would also like to agree that this would not be the only reason for the low completion rates. Other factors, involving sponsors, factors concerning genders of trainees, etc... Could definitely have an impact, if not a more profound one, than what we have found with our research.

## Sources of Error in Our Research

We would also like to point out some sources of errors in our research. Many of the data sets had incomplete values, and some observations had no numeric values for their variables at all. This could have quite possibly affected our results, though we do not expect it to have a profound effect on our findings. Minor changes to the values in the tables above is all we are expecting from these missing values.

# Bibliography

* "Manufacturing." Statcan.gc.ca. Goverment of Canada, 07 Oct. 2016. Web. 30 May 2017.
* "2016 Auditor General Annual Report." 2016 Annual Report 1 (2016): 240-70. Auditor.on.ca. The Provincial Government of Ontario. Web. 30 May 2017.