International Student Tuition Fees Survey at Canada

Name: Pan Zhang
Student ID: 180108960
Department of Computer Science
University: Wilfrid Laurier University
Waterloo, Canada
Email: zhan8960@mylaurier.ca

Abstract—Background: As an international student, the annual tuition expenses account for most of my personal expenses. Then, I want to know more about the tuition fees charged by various provinces of Canada for international undergraduate students among different fields to compare whether my personal tuition expenses are reasonable?

Method: Firstly, data cleaning in data preprocessing is used to solve the problem of missing values. Then use histograms, time series and other techniques to visualize the data and perform data mining

Results: I found and compared tuition fees for international students across fields in Ontario and Quebec. Second, I found through the time series that the tuition fees for international students in Ontario computer science majors are increasing every year and the growth rate has increased in recent years.

Conclusion: My result provides several evidences that my personal tuition situation is lower than the average tuition fee for international undergraduate student in Ontario.

The data resource used in this report is from the government of Canada. The original open data resource can be found at https://open.canada.ca/data/en/dataset/1f02b37f-dcba-4546-816f-e592f6264e97 [1]

1. INTRODUCTION

With the trend of globalization of the world, more and more students are pursuing educational environment of different cultures. Canada, as a famous higher education country in the world, has attracted the attention of many international students. According to the data of Canada's national statistical agency, it shows "Number of international college student enrolments more than doubled over the last five years" [2]. However, tuition fees as one of the necessary expenses for international students have attracted the attention of many international students including me. Therefore, in this report, I will use various knowledge of data mining to compare the tuition fees of international undergraduates in different fields in Canadian provinces. As a comparison, I also need to compare my own data with it to know if my own tuition is reasonable?

2. DATA PREPROCESSING

First, we should perform data preprocessing before dealing with these data sets. Because Today's real-world data is highly vulnerable to incomplete (missing values for some attributes of interest), inconsistent (including code or name differences), and highly susceptible to noise (errors or outliers). In addition, datasets often come from multiple heterogeneous data sources, and low-quality data will lead to low-quality mining results as you often hear “Garbage in, Garbage out”.

```
In [1]: import pandas as pd
        import numpy as np
        df = pd.read_csv('./37100005.csv')
        df
```

Out[1]:

	REF_DATE	GEO	DGUID	Field of study	UOM	UOM_ID	SCALAR_FACTOR	SCALAR_ID	VECTOR	COORDINATE	VALUE	STATUS
0	2006/2007	Canada	2016A000011124	Total, field of study	Current dollars	75	units	0	v1001414713	1.19	13378.0	NaN
1	2006/2007	Canada	2016A000011124	Education	Current dollars	75	units	0	v62427452	1.10	11818.0	NaN
2	2006/2007	Canada	2016A000011124	Visual and performing arts, and communications...	Current dollars	75	units	0	v62427461	1.20	12784.0	NaN
3	2006/2007	Canada	2016A000011124	Humanities	Current dollars	75	units	0	v62427462	1.30	13313.0	NaN
4	2006/2007	Canada	2016A000011124	Social and behavioural sciences, and legal stu...	Current dollars	75	units	0	v62427463	1.40	12614.0	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...
2656	2020/2021	Yukon	2016A000260	Nursing	Current dollars	75	units	0	v1210490551	12.14	NaN	..
2657	2020/2021	Yukon	2016A000260	Pharmacy	Current dollars	75	units	0	v1210490552	12.15	NaN	..
2658	2020/2021	Yukon	2016A000260	Veterinary medicine	Current dollars	75	units	0	v1210490553	12.16	NaN	..
2659	2020/2021	Yukon	2016A000260	Other health, parks, recreation and fitness	Current dollars	75	units	0	v1210490554	12.17	NaN	..
2660	2020/2021	Yukon	2016A000260	Personal, protective and transportation services	Current dollars	75	units	0	v1210490555	12.18	NaN	..

2661 rows × 15 columns

Figure 1: Code for showing raw data

After explaining the main functions of data preprocessing, I think a brief introduction to data preprocessing is necessary. As figure 2 demonstrates the four main data preprocessing techniques. In terms of data cleaning, the data can have many irrelevant and missing parts. Data cleaning attempts to fill in missing values, smooth out noise and correct inconsistencies in the data. In terms of data integration, data often come from different datasets, and the integration of data at this time may cause many serious problems such as data matching problems and redundancy. Therefore, the main purpose of data integration is to reduce the inconsistency of the results so as to greatly improve the accuracy of the mining process. In terms of data reduction, while working with huge volume of data, analysis became harder in such cases. In order to get rid of this, we use data reduction technique. It aims to increase the storage efficiency and reduce data storage and analysis costs. Data reduction strategies include dimensionality reduction, numerosity reduction, data compression. One of the most well-known should be histograms that belong to numerosity reduction. In terms of data transformation, this step is taken in order to transform the data in appropriate forms suitable for mining process. The main strategies are normalization, attribute selection, discretization, concept hierarchy generation.

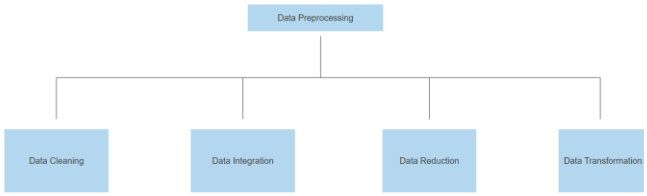

Figure 2: Four important techniques in data preprocessing

In this dataset of 2662 rows * 15 columns, it contains the fees charged by Canadian provinces to international students in different majors from 2006 to 2021. It not only includes tuition fees for international students in different provinces in Canada, but also distinguishes the majors. This allows us to find and compare tuition fees across provinces for all subjects. Meanwhile, this also causes missing values, because the problem is that some provinces do not have a certain major, such as education in Prince Edward Island, which does not exist. Dealing with this problem is also quite simple. Data cleaning in data preprocessing is used to solve this problem. As we might already know that data cleaning attempts to fill in missing values, smooth out noise and correct inconsistencies in the data. For the treatment of missing values, there are various methods of data cleaning, including ignoring tuples and regression. But for this case, filling in the missing values yourself is the most efficient way. Because there are not many missing values, and only need to fill the missing places with 0 to solve the problem.

3. DATA MINNING

It is time to explore the data. Since it is for comparison, there is no doubt that the histogram can intuitively reflect the contrast. First, I plotted the histogram which demonstrates the average cost of international students in Ontario for each major in the most recent year, 2020/2021 at figure 4. Obviously, the major with the highest tuition fee for international students is Dentistry, and the lowest tuition fee is Education. For my own major, which is computer science, I use red to distinguish it from other majors. We can find that the tuition for computer science majors ranks near the median and is almost as high as the average tuition for all majors.

```
In [2]: sub_ONTARIO = df.loc[df["GEO"] == "Ontario"]
        sub_QUEBEC = df.loc[df["GEO"] == "Quebec"]

        sub_ONTARIO_2021 = sub_ONTARIO.loc[sub_ONTARIO["REF_DATE"] == "2020/2021"]
        sub_QUEBEC_2021 = sub_QUEBEC.loc[sub_QUEBEC["REF_DATE"] == "2020/2021"]

        average_ONTARIO = sub_ONTARIO_2021["VALUE"].mean()
        average_QUEBEC = sub_QUEBEC_2021["VALUE"].mean()

        print("The average tuition fees is", average_ONTARIO, ",", average_QUEBEC)

        import matplotlib.pyplot as plt

        color = []

        sub_ONTARIO_2021 = sub_ONTARIO_2021.reset_index(drop=True)
        sub_QUEBEC_2021 = sub_QUEBEC_2021.reset_index(drop=True)

        for  i in range(sub_ONTARIO_2021.shape[0]):
            if(sub_ONTARIO_2021["Field of study"][i] == "Mathematics, computer and information sciences"):
                color.append("red")
            else:
                color.append("blue")

        plt.figure(figsize=(11,5))

        plt.subplot(121)

        plt.title("ONTARIO")

        plt.bar(sub_ONTARIO_2021["Field of study"], sub_ONTARIO_2021["VALUE"], color=color)

        plt.xticks(rotation=90)
        plt.subplot(122)
        plt.title("QUEBEC")
        plt.bar(sub_QUEBEC_2021["Field of study"], sub_QUEBEC_2021["VALUE"], color=color)
        plt.bar([1, 2], [2, 3])
        plt.xticks(rotation=90)

        plt.show()


        plt.bar(["ONTARIO", "QUEBEC"], [average_ONTARIO, average_QUEBEC], color=["red", "green"])

        plt.show()

        The average tuition fees is 45085.89473684211 , 24931.611111111111
```

Figure 3: Code for two histograms

In addition, comparisons between different provinces are also necessary. The neighboring province of Quebec is a good choice. So, I used the same method to plot the average tuition fees for international students in Quebec by major. Interestingly, like Ontario, Dentistry ranks as the highest tuition major, and Education is the lowest tuition major. It is important to note that although Figure 5 appears to be higher than Figure 4 in any area. But the Y coordinates of the two histograms are not uniform. Let's take Dentistry as an example of both. Obviously, the tuition fee for international students studying Dentistry in Ontario is close to 100k per year. However, the same discipline is placed in Quebec and only needs to make up 40k per year. By the way, the specialty of Veterinary medicine in Quebec in Figure 5 does not have data which is the missing value I mentioned in the previous

section. This means that this major is not available at any university in Quebec

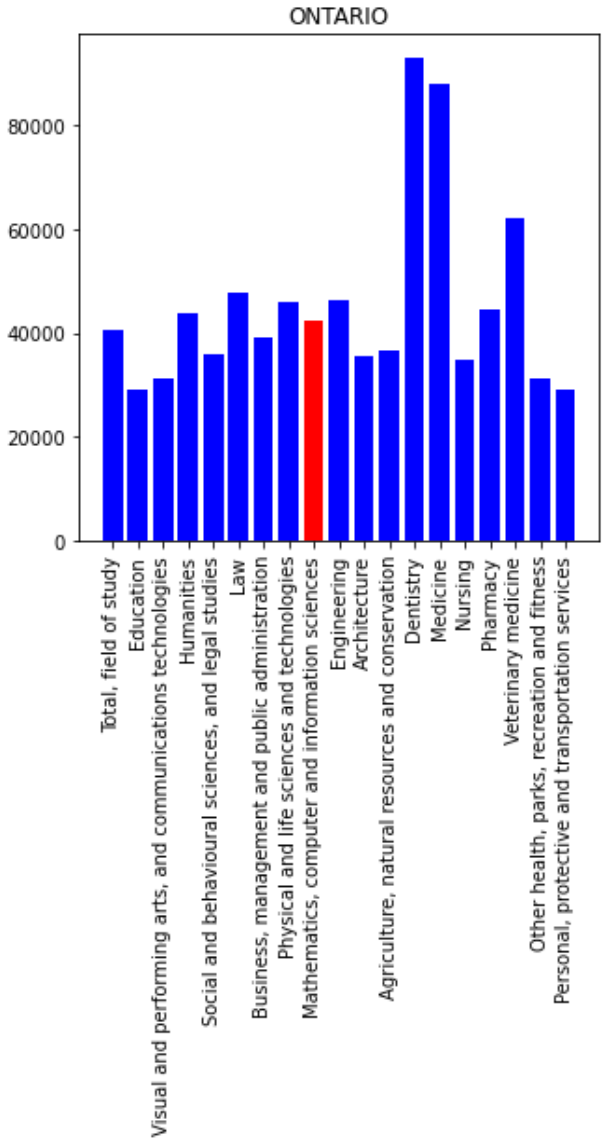

Figure 4: Average tuition fees for international students by field in Ontario

For the subject of Dentistry alone, tuition fees vary widely between Ontario and Quebec. Results like this shocked me and wondered if all professional differences were like this. Therefore, I separately calculated the average cost of all majors for international students in the two provinces and compared them at figure 6. The red bar stands for Ontario and the green bar stands for Quebec. We can find that Ontario's tuition fees are around twice as high as Quebec's per year. To be honest, the results were equally shocking to me. Because I know the universities in Quebec are excellent. And the doubling of the tuition gap in the middle makes me feel desirable.

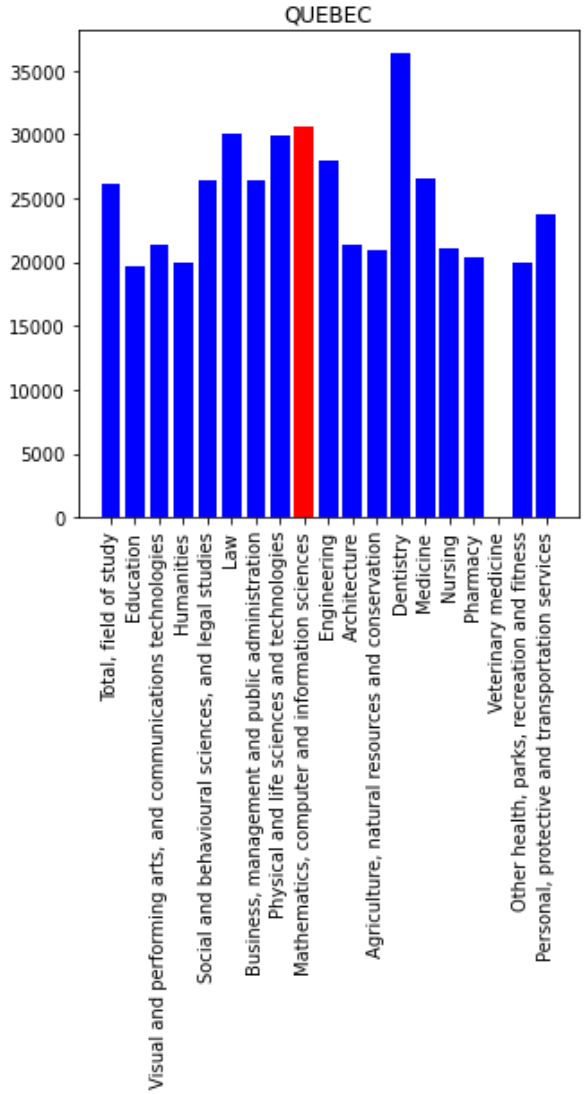

Figure 5: Average tuition fees for international students by field in Quebec

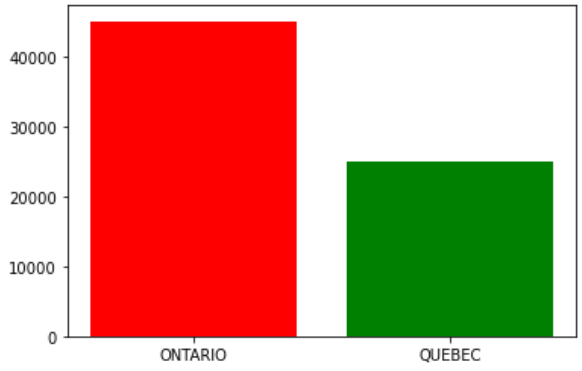

Figure 6: Average tuition fees for international undergraduate student in Ontario and Quebec

The whole dataset was recorded from 2006 to 2021. As a student major in computer science, I am interested in the changes in tuition of my major from 2006 to the present. Time

series is good to use here.

```
In [3]:
sub_ONTARIO = df.loc[df["GEO"] == "Ontario"]

sub_ONTARIO_computer = sub_ONTARIO.loc[sub_ONTARIO["Field of study"] == "Mathematics, computer and information sciences"]

import matplotlib.pyplot as plt

plt.bar(sub_ONTARIO_computer["REF_DATE"], sub_ONTARIO_computer["VALUE"])
plt.title("value")
plt.xticks(rotation=90)
plt.show()
```

Figure 6: Code for building time series

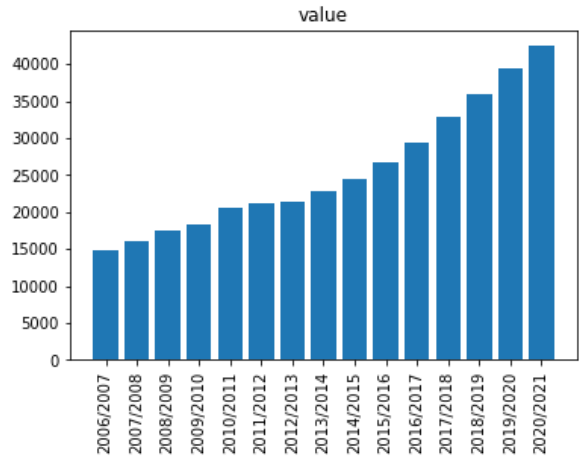

Figure 7: Time series from 2006 to 2021

Through the time series diagram in Figure 7, we can find that the tuition fees for international students in Ontario's computer science majors are increasing every year. Among them, the growth was slow from 2006 to 2015, and the growth was rapid from 2015 to 2021.

4. CONCLUSION

```
In [4]:
def search(GEO, field, time):
    dirc = {"edu":"Education", "vpact":"Visual and performing arts, and communications technologies", "hum":"Humanities",
    "sbc": "Social and behavioural sciences, and legal studies", "law":"Law", "bap":"Business, management and public administration",
    "plt":"Physical and life sciences and technologies", "mcs":"Mathematics, computer and information sciences", "eng":"Engineering",
    "arc":"Architecture", "aar":"Agriculture, natural resources and conservation", "med":"Medicine","hpr":"Other health, parks, recreation and

    a = df.loc[df["GEO"] == GEO]

    b = a.loc[a['Field of study'] == dirc[field]]

    time = time + "/" + str(int(time) + 1)

    c = b.loc[b['REF_DATE'] == time]

    c = c.reset_index(drop=True)

    return c['VALUE'][0]

geo = input("Enter your province: ")
field = input("Enter your field(remember to enter the abbreviation): ")
time = input("Enter the time you wanna search for: ")
print(search(geo,field, time))
#example testing
#Enter your province: Ontario
#Enter your field: mcs
#Enter the time you wanna search for: 2020
#42444.0
def addition(a,b,c):
    return a + b + c
#2018 35920.0
#2019 39476.0
#2020 42444.0
print(addition(35920.0, 39476.0, 42444.0))

Enter your province: Ontario
Enter your field(remember to enter the abbreviation): mcs
Enter the time you wanna search for: 2020
42444.0
117840.0
```

Figure 8: Function for searching up the tuition fees

Finally, I wrote a function as figure 8 shows to facilitate me to query the tuition fees of international students in different provinces and different majors. Back to my project question, I can get the average tuition fee for international students in Ontario since I enrolled in 2018 which is 117840.0. Then, I could go loris to check my student account and get final number 103484 which is less than the average tuition fee in Ontario.

Of course, all the data in this report is about tuition fees for international students. This does not directly represent the quality of teaching in the provinces. The professors just impart knowledge to us, and how much we can accept is our own ability. The purpose of international students coming to Canada to study is to have a better learning environment. After I learned about these high tuition data, it motivated me to focus more on my studies.

References:

[1] Government of Canada, S. C. (2021, October 28). International undergraduate tuition fees by field of study. Retrieved March 28, 2022, from https://open.canada.ca/data/en/dataset/1f02b37f-dcba-4546-816f-e592f6264e97

[2] Government of Canada, S. C. (2021, November 24). Prior to COVID-19, international students accounted for the growth in postsecondary enrolments and graduates. Retrieved March 28, 2022, from https://www150.statcan.gc.ca/n1/daily-quotidien/211124/dq211124d-eng.htm