DATA ANALYST: SQL PORTFOLIO

PREPARED BY
SAMBO MOPELOLA

Professional Background

I am a graduate of The Federal Polytechnic Ado-Ekiti, Ekiti state, Nigeria. I worked as a Front Desk Officer in a Private school, for 6months and I currently work as a customer service representative at a financial technology company.

I have a certification of Excellence at thetechytrainfoundation for the techup-girls, where we learned about technology and basic digital skills. I also have a Professional Certification In Customer Service.

I love learning and I'm passionate about self growth and development. I love helping and reaching out to other with my knowledge.

I am proficient in structured query language (SQL), Microsoft Excel, data visualization, data cleaning, presentation Skills. I also have excellent communication skills, attention to detail, as well as great listening & critical thinking skills.

My Vision is to develop a career in Data Analysis, to be a disciplined and insightful Data Analyst with an eagerness to leverage big data interpreting and visualizing skills

Table of Contents

Introduction

A dataset was given, the dataset was based on students' report table, which is the Business Background. The dataset consists of information about the students' reporting table. As a data analyst, my goal is to find insights on the data given, to make the data meaningful to other people.

The business problem is to make other people understand what the data means and how it can be actioned.

The objective of this project is to:

- Find trends between the students.
- What are the specific symptoms of the problem?
- What is the hypothesis for the cause of the problem?

Root Cause Analysis

Root cause analysis (RCA) is defined as a collective term that describes a wide range of approaches, tools, and techniques used to uncover causes of problems. Some RCA approaches are geared more toward identifying true root causes than others, some are more general problem-solving techniques, and others simply offer support for the core activity of root cause analysis.

To solve the business problem, the student information, family information, their home, income and ethnicity, their region, education, wages and urban areas were analyzed accordingly.

In order to analyze the report table, these questions were asked:

- How many male and female students are there?
- What is the sum total of the student tuition of both male and female?
- How did the fcollege, mcollege, home and income affects the students education?
- What is the difference between students whose parents are college graduates and those without college graduates?
- How did wage, unemp and avg county tuition affect the students?

Root Cause Analysis

The root cause Analysis:

- **Q: Why do we have students with lower years of education?**

  A: Because their parents were not receiving a high income.

- **Q: Why were their parents not receiving a high income?**

  A: Because of the wages they received, they were affected by the county unemployment rate in 2020.

- **Q: Why did we have more students with higher student tuition?**

  A:Because we have more student with higher years of education

- **Q: Why did we have more students with higher years of education?**

A: Because their parents receive a high income based on their county average..

- **Q: Why did their parents have higher income based on their county average?**

  A: Because their parents are college graduates.

Insights

A reporting tables file consisting of four datasets was given. These datasets consist of county info, student academy, student family and student personal info.

A left join was used to join county info and student academy as one table. A right join was used to join student family and student personal data as another table, giving us two table datasets. To make the two tables into one big table, a cross join was used. This join will combine all information and attributes in each table into one whole table. The table consists of county info, student academy, student family and student personal info.

The attributes in the dataset include:

- gender: :factor indicating gender.
- ethnicity: factor indicating ethnicity (African-American, Hispanic, Asian or other).
- academic_score: student's academic score throughout high school and college
- student_tuition: cost of tuition for the student
- education: the years of education the student has received
- fcollege: factor. Is the father a college graduate?
- mcollege: factor. Is the mother a college graduate?
- home: factor. Does the family own their home?
- urban: factor. Is the school in an urban area?

Insights

- unemp: county unemployment rate in 2020
- income: high or low income household based on county average
- wage: state hourly wage in manufacturing in 1980
- distance: distance from 4-year college (in 10 miles)
- region: factor indicating region (West, East or other)
- avg_county_tuition: average state 4-year college tuition (in 1000 USD)

SQLite was used to analyze my findings from the data and Charts were created on excel.

The SQL queries/Commands, window functions used to analyze the Datasets are:

- LEFT JOIN
- RIGHT JOIN
- CROSS JOIN
- ORDER BY
- WHERE
- AND
- SUM()
- COUNT()
- RANK
- GROUP BY
- LIMIT()

Insights

A left join was used to join county info and student academy as one table

```
CREATE TABLE county_studentacdc_info AS
SELECT *
FROM county_info
LEFT JOIN student_academic_info
ON county_info.id = student_academic_info.id
```

A right join was used to join student family and student personal data as another table,

```
CREATE TABLE student_fam_persnl_details AS
SELECT *
FROM student_family_details
RIGHT JOIN student_personal_details
ON student_family_details.id = student_personal_details.id
```

To make the two tables into one big table, a cross join was used. This join will combine all information and attributes in each table into one whole table. The table consists of county info, student academy, student family and student personal info.

```
CREATE TABLE report_student_colleges AS
SELECT *
FROM county_studentacdc_info
CROSS JOIN student_fam_persnl_details
ON county_studentacdc_info.id = student_fam_persnl_details.id
```

Insights

Female students whose parents are college graduates, own their own home, have high income with years of education above 14, their urban ethnicity.

```sql
SELECT academic_score, student_tuition, gender, income, home, ethnicity, education, mcollege, fco
FROM reports_student_colleges
WHERE gender = 'female'
AND income = 'high'
AND mcollege = 'yes'
AND fcollege = 'yes'
AND home = 'yes'
AND education > 14
AND urban = 'yes'
ORDER BY education DESC
```

Male students whose parents are college graduates, own their own home, have high income with years of education above 14, their urban ethnicity.

```sql
SELECT academic_score, student_tuition, gender, income, home, ethnicity, education, mcollege, fcol
FROM reports_student_colleges
WHERE gender = 'male'
AND income = 'high'
AND mcollege = 'yes'
AND fcollege = 'yes'
AND home = 'yes'
AND education > 14
AND urban = 'yes'
ORDER BY education DESC
```

Female students whose parents are not college graduates, do not own their own home, have high income with years of education above 14, their urban ethnicity.

```sql
SELECT academic_score, student_tuition, gender, income, home, ethnicity, education, mcollege, fcol
FROM reports_student_colleges
WHERE gender = 'female'
AND income = 'high'
AND mcollege = 'no'
AND fcollege = 'no'
AND home = 'no'
AND education > 14
AND urban = 'no'
ORDER BY education DESC
```

Insights

Male students whose parents are not college graduates, do not own their own home, have high income with years of education above 14, their urban ethnicity.

```sql
SELECT academic_score, student_tuition, gender, income, home, ethnicity, education, mcollege, fc
FROM reports_student_colleges
WHERE gender = 'male'
AND income = 'high'
AND mcollege = 'no'
AND fcollege = 'no'
AND home = 'no'
AND education > 14
AND urban = 'no'
ORDER BY education DESC
```

To know the avg_county_tuition by student_tuition rounded off into 3 decimal places. The statement below was used.

```sql
-- avg_county tition by student_tuition.
SELECT gender, avg_county_tuition, student_tuition,
Round(avg(avg_county_tuition) OVER (PARTITION BY student_tuition), 3) AS avg_county_tuition_studen
FROM reports_student_colleges
```

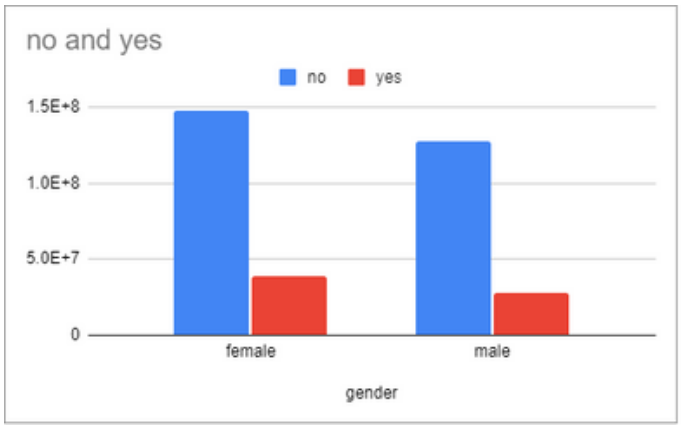
To rank the student_score, student_tuition, education by date of birth, this statement was written.

```sql
SELECT *,
rank() OVER (PARTITION BY academic_score,student_tuition, education ORDER BY dob DESC) AS rank_by_
FROM reports_student_colleges
```

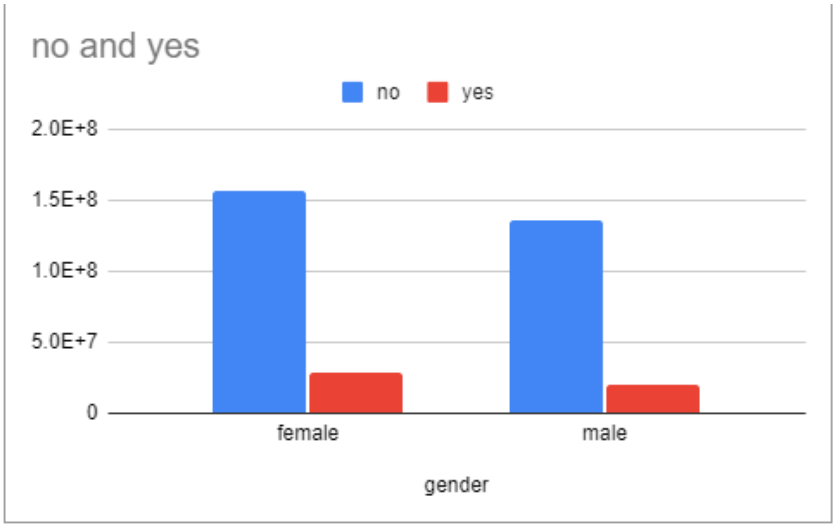
Insights

The chart was created on Google sheets.

SUM of student_ fcollege			
gender	no	yes	Grand Total
female	147827370	38375226	186202596
male	128256614	27458826	155715440
Grand Total	**276083984**	**65834052**	**341918036**



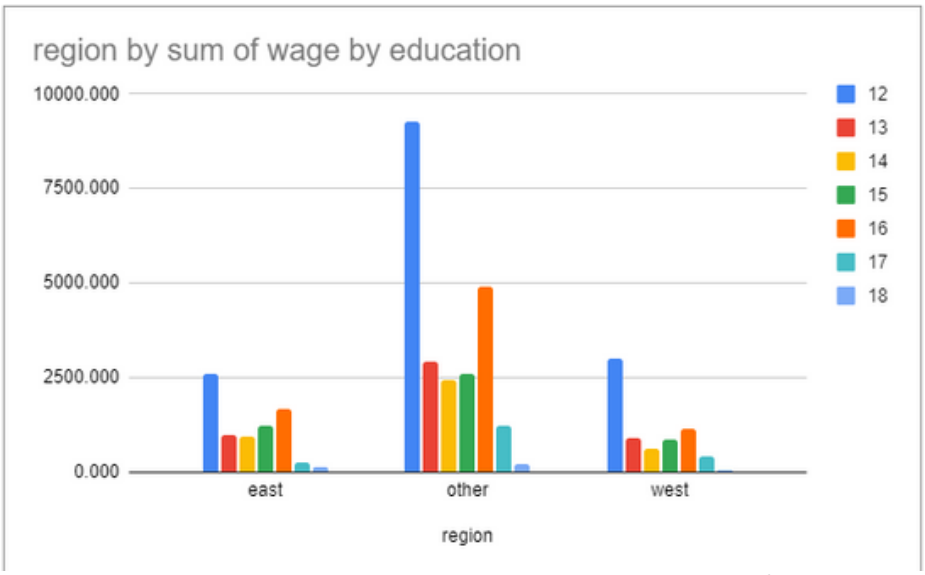
SUM of student_ mcollege			
gender	no	yes	Grand Total
female	157408862	28793734	186202596
male	135650979	20064461	155715440
Grand Total	**293059841**	**48858195**	**341918036**



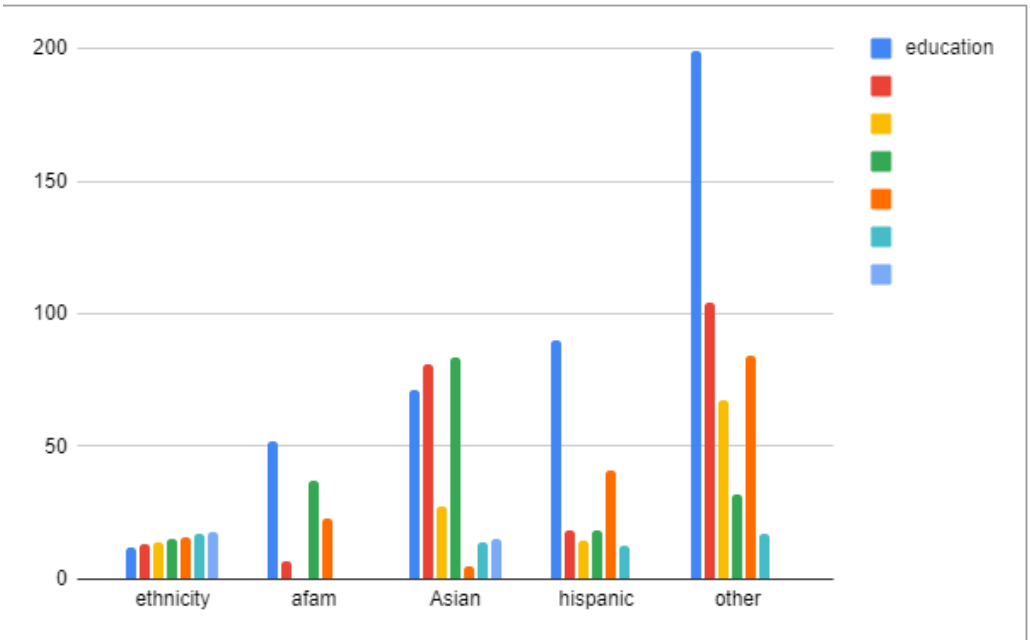
The Table and charts above shows that the sum total of student_tuition for students(male and female) whose parents attended college is higher than students whose parents did not attend college.

Insights

The chart created shows that students from other regions have the highest sum of wages and also receive the lowest years of education.

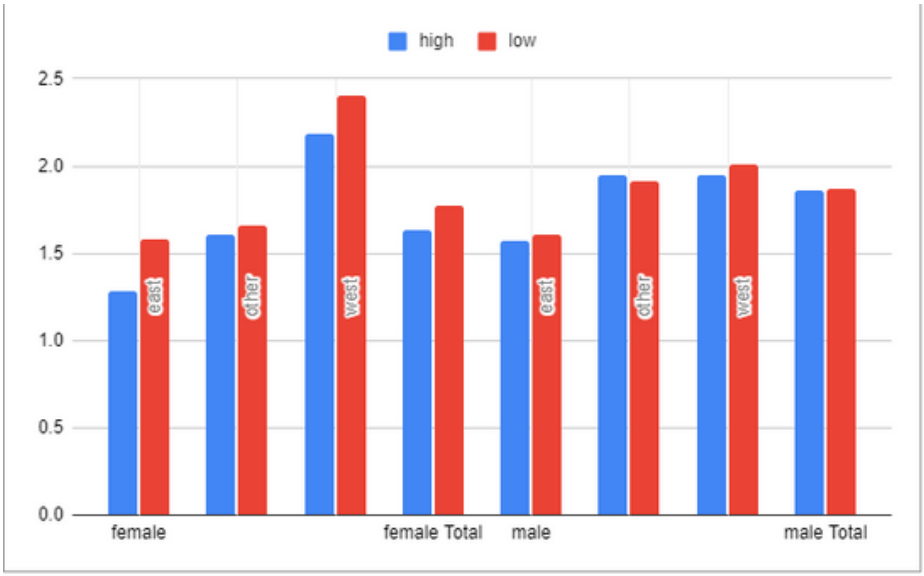


The chart demonstrates that students from other ethnicities with 12 years of education being their highest years of education they have received. This also shows that the parents of students from other ethnicities have the highest sum total of unemployment rate in 2020, which results in the years of education the students have received.



Insights

AVERAGE of distance		income		
gender	region	high	low	Grand Total
female	east	1.28	1.58	1.49
	other	1.61	1.66	1.64
	west	2.18	2.41	2.34
female Total		1.63	1.77	1.73
male	east	1.57	1.61	1.60
	other	1.95	1.91	1.92
	west	1.94	2.01	1.99
male Total		1.86	1.87	1.86
Grand Total		**1.73**	**1.81**	**1.79**



This chart shows the relationship between students, their region, their parents income and their distance from 4-year college.

This chart tells us that female students from the east, west,and other regions, have low income. With female students from other regions ranking the top lowest income.

It also informs us that male students from the east and west regions have low income but have high income from other regions. Male students' total income from all regions is not entirely low to the total of the high income. This resulted in the factor affecting their distance from 4-year college.

Findings and Recommendations

Findings from the Analysis are shown below:

The total number of the students is 5375.

The total number of male students is 2456.

The total number of female students is 2919.

The highest student_tuition is 119987.

The lowest student_tuition is 50030.

Total number of students whose parents attended college and own their homes, receives high income. And students whose parents did not attend college do not own their home and do not receive high income.

Fig. 1.0

gender	f/mcollege(yes)	f/mcollege(no)
Male	168	344
Female	265	434

A table consisting of student_tuition, academic_score, ethnicity, income, home, mcollege, fcollege, education, urban and gender.

Findings and Recommendations

Findings from the Analysis are shown below:

Fig. 1.1

academic_score	student_tuition	gender	income	home	ethnicity	education	mcollege	fcollege	urban
40.5999984 7	75288	female	high	yes	afam	18	yes	yes	yes
40.5999984 7	75288	female	high	yes	afam	18	yes	yes	yes
40.5999984 7	75288	female	high	yes	afam	18	yes	yes	yes
59.3899993 9	51029	female	high	yes	afam	16	yes	yes	yes
59.3899993 9	51029	female	high	yes	afam	16	yes	yes	yes
58	107531	female	high	yes	Asian	16	yes	yes	yes
62.1399993 9	104646	female	high	yes	hispanic	16	yes	yes	yes
58	107531	female	high	yes	Asian	16	yes	yes	yes
58.9099998 5	87350	female	high	yes	other	15	yes	yes	yes
47.1500015 3	64522	female	high	yes	Asian	15	yes	yes	yes
58.9099998 5	87350	female	high	yes	other	15	yes	yes	yes
47.1500015 3	64522	female	high	yes	Asian	15	yes	yes	yes
58.9099998 5	87350	female	high	yes	other	15	yes	yes	yes

Fig. 1.2 and Fig.1.3 shows students with higher years of education, whose parents attended colleague, owned their home, as high icome, etc **Fig. 1.2**

academic_score	student_tuition	gender	income	home	ethnicity	education	mcollege	fcollege	urban
45.11999893	67070	male	high	yes	Asian	18	yes	yes	yes
45.11999893	67070	male	high	yes	Asian	18	yes	yes	yes
52.93999863	80100	male	high	yes	other	17	yes	yes	yes
62.24000168	58972	male	high	yes	Asian	17	yes	yes	yes
53.93999863	97664	male	high	yes	afam	15	yes	yes	yes

Findings and Recommendations

Findings from the Analysis are shown below:

Fig. 1.3

academic_score	student_tuition	gender	income	home	ethnicity	education	mcollege	fcollege	urban
40.5999984 7	75288	female	high	yes	afam	18	yes	yes	yes
40.5999984 7	75288	female	high	yes	afam	18	yes	yes	yes
40.5999984 7	75288	female	high	yes	afam	18	yes	yes	yes
59.3899993 9	51029	female	high	yes	afam	16	yes	yes	yes
59.3899993 9	51029	female	high	yes	afam	16	yes	yes	yes
58	107531	female	high	yes	Asian	16	yes	yes	yes
62.1399993 9	104646	female	high	yes	hispanic	16	yes	yes	yes
58	107531	female	high	yes	Asian	16	yes	yes	yes
58.9099998 5	87350	female	high	yes	other	15	yes	yes	yes
47.1500015 3	64522	female	high	yes	Asian	15	yes	yes	yes
58.9099998 5	87350	female	high	yes	other	15	yes	yes	yes
47.1500015 3	64522	female	high	yes	Asian	15	yes	yes	yes
58.9099998 5	87350	female	high	yes	other	15	yes	yes	yes

Fig. 1.4 and Fig.1.5 shows students with higher years of education, whose parents did not attend college, did not own their home, but received high income, etc.

Fig. 1.4

academic_score	student_tuition	gender	income	home	ethnicity	education	mcollege	fcollege	urban
61.7700 0046	88936	male	high	no	Asian	16	no	no	no
62.9399 9863	101813	male	high	no	Asian	16	no	no	no
51.2799 9878	67818	male	high	no	other	16	no	no	no
51.2799 9878	67818	male	high	no	other	16	no	no	no
61.7700 0046	88936	male	high	no	Asian	16	no	no	no
62.9399 9863	101813	male	high	no	Asian	16	no	no	no
60.259 99832	61274	male	high	no	other	16	no	no	no

Findings and Recommendations

Findings from the Analysis are shown below:

Fig. 1.5

academic_score	student_tuition	gender	income	home	ethnicity	education	mcollege	fcollege	urban
53.20999908	86214	female	high	no	other	18	no	no	no
65.11000061	76001	female	high	no	Asian	17	no	no	no
65.11000061	76001	female	high	no	Asian	17	no	no	no
49.00999832	105953	female	high	no	other	17	no	no	no
55.27999878	103170	female	high	no	hispanic	16	no	no	no
60.04999924	99045	female	high	no	afam	16	no	no	no
55.04000092	109976	female	high	no	hispanic	16	no	no	no
56.86000061	114594	female	high	no	other	16	no	no	no
53.77999878	118360	female	high	no	Asian	16	no	no	no
56.13999939	116732	female	high	no	Asian	16	no	no	no
55.27999878	103170	female	high	no	hispanic	16	no	no	no
60.04999924	99045	female	high	no	afam	16	no	no	no
64.86000061	114961	female	high	no	other	16	no	no	no
55.27999878	103170	female	high	no	hispanic	16	no	no	no
53.77999878	118360	female	high	no	Asian	16	no	no	no
56.13999939	116732	female	high	no	Asian	16	no	no	no
55.27999878	103170	female	high	no	hispanic	16	no	no	no
62.75999832	111858	female	high	no	other	16	no	no	no
55.27999878	103170	female	high	no	hispanic	16	no	no	no
65.62000275	107816	female	high	no	afam	15	no	no	no
53.74000168	55440	female	high	no	Asian	15	no	no	no
65.62000275	107816	female	high	no	afam	15	no	no	no
65.62000275	107816	female	high	no	afam	15	no	no	no
63.75999832	119253	female	high	no	Asian	15	no	no	no
51.56000137	63591	female	high	no	Asian	15	no	no	no
53.74000168	55440	female	high	no	Asian	15	no	no	no
65.62000275	107816	female	high	no	afam	15	no	no	no

Findings and Recommendations

This report shows that, although female students are more than the male students given in the reporting table, we found out that students whose parents attended college have more student tuition than students whose parents did not attend college. Also, students from other regions have the highest sum of wages and also receive the lowest years of education.

Parents of students from other ethnicities have the highest sum total of unemployment rate in 2020, which results in the years of education the students have received. The student majorly receives 12 years of education.

This report also tells us that parents of female students from the east, west,and other regions, have low income. With female students from other regions ranking the top lowest income.
It also informs us that male students from the east and west regions have low income but have high income from other regions. Male students' total income from all regions is not entirely low to the total of the high income. This resulted in the factor affecting their distance from 4-year college.

Findings and Recommendations

The report also shows that regardless of the fact that some parents did not attend college, they have more students attending schools. This report shows students whose parents attended college and those whose parents did not attend college received a high level of education and this could be said according to this report, that their level of education is dependent on the fact their parents receive high income.

This report recommends that parents who receive higher wages should focus more on their children's education, since income is entirely not a factor for reducing or delaying their years of education. Also, region, urban, home, ethnicity, mcollgue, fcollogue, according to this report, has shown that all of this should not be considered a factor that affects the years of education a student should receive.

Conclusion

Conclusively, a reporting table consisting of 4 datasets were given. county info, student academy, student family and student personal info. The purpose of this analysis was to find insights, make other people understand what the data means and how it can be actioned.

The analysis carried out on this report, showed the relationship between the students, their mcollege , fcollege, region, urban, ethnicity, wages, student_tuition, student_score, distance affects the students years of Education.

This report also made us understand that the parent literacy(whether the parents attended college or not) is not and should not be a factor that should affect the years of education the student receives.