3.0   T-SQL STATEMENTS

3.1   CREATING THE DATABASE

The school survey database was created to house all the data and schema used in this analysis using the following lines of code:

```
--CREATING DATABASE
USE master;
GO
CREATE DATABASE VietnamSchoolSurvey;

--CHOOSE THE DATABASE AS THE DATABASE IN USE
USE VietnamSchoolSurvey;
GO
```

3.2   Inserting the data into Microsoft Sequel Server

The data was saved as an excel file and was imported into SQL Server by following the steps below:

- Right click on VietnamSchoolSurvey database| Task | import Data then follow the prompts
- Set Microsoft Excel as the data source, browse to the file path and click on the file and click on next
- Choose Sequel Server Native Client 11.0 as the destination and click on next four times and then click on finish
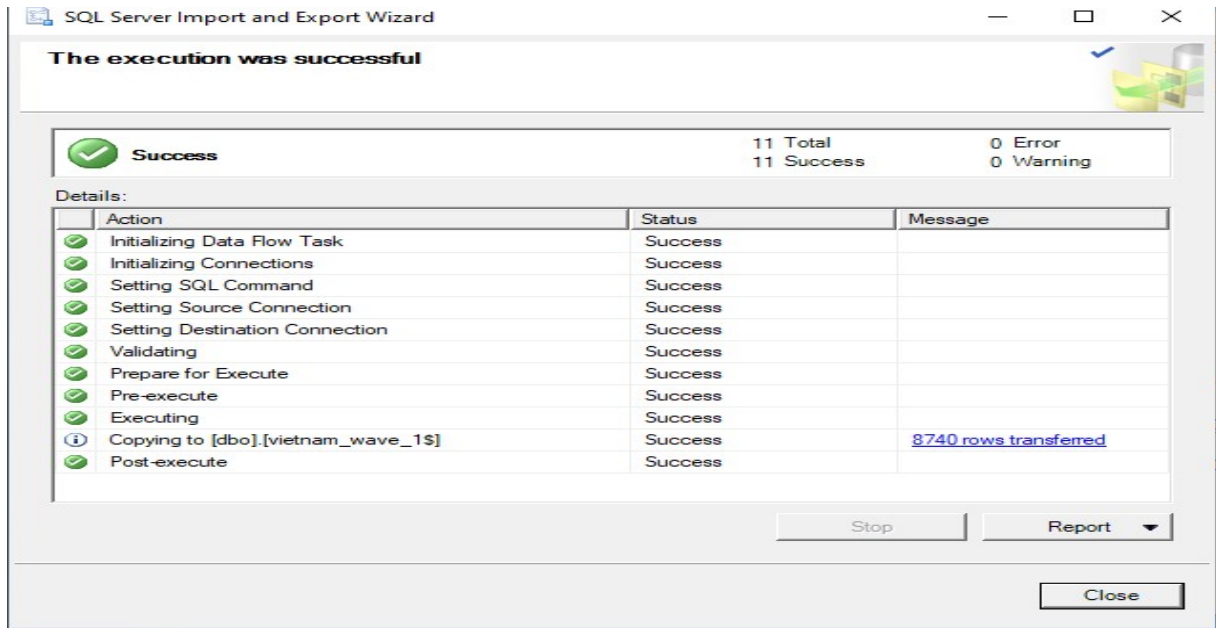
Figure 2: Inserting data into SQL Server

3.3 CREATING TABLES AND VIEWS

The table previously imported into SQL Server was without a primary key and columns that will ideally be of an Integer data type were of float data type so a new table comprising of all columns relevant for this analysis was created with the **'UNIQUEID'** column as the primary key and all required data from the previous table was imported into this new table. See script below:

```
CREATE TABLE [dbo].[VietnamDataset](
    [UNIQUEID] [nvarchar](150) NOT NULL,
    [SCHOOLID] [int] NULL,
    [CLASSID] [int] NULL,
    [STUDENTID] [int] NULL,
    [YLCHILDID] [nvarchar](150) NULL,
    [PROVINCE] [int] NULL,
    [DISTRICTCODE] [int] NULL,
    [LOCALITY] [int] NULL,
    [GENDER] [int] NULL,
    [AGE] [int] NULL,
```

```sql
[ETHNICITY] [int] NULL,
[ABSENT_DAYS] [int] NULL,
[MOM_READ] [int] NULL,
[MOM_EDUC] [int] NULL,
[DAD_READ] [int] NULL,
[DAD_EDUC] [int] NULL,
[STDYLCHD] [int] NULL,
[STDCMPLT] [nvarchar](150) NULL,
[STDDINT] [nvarchar](150) NULL,
[STDLIV] [int] NULL,
[STDLNGHM] [int] NULL,
[STDMEAL] [int] NULL,
[STDHLTH1] [int] NULL,
[STDHLTH2] [int] NULL,
[STDHLTH3] [int] NULL,
[STDHLTH4] [int] NULL,
[STDHLTH5] [int] NULL,
[STDHLTH6] [int] NULL,
[STDHLTH0] [int] NULL,
[STPPLHM] [int] NULL,
[STSIBOLD] [int] NULL,
[STSIBYNG] [int] NULL,
[STNMBOOK] [int] NULL,
[STPLSTDY] [int] NULL,
[STHVMTEL] [int] NULL,
[STHVRADO] [int] NULL,
[STHVTELE] [int] NULL,
[STHVBIKE] [int] NULL,
[STHVMTBK] [int] NULL,
[STHVDESK] [int] NULL,
[STHVCHR] [int] NULL,
[STHVLAMP] [int] NULL,
[STHVEFAN] [int] NULL,
[STHVAIRC] [int] NULL,
[STHVCAR] [int] NULL,
[STHVCOMP] [int] NULL,
[STHVINTR] [int] NULL,
[STHVFRDG] [int] NULL,
[STHVMCRO] [int] NULL,
```

```
[STHVDVD] [int] NULL,
[STHVCBLE] [int] NULL,
[STTMSCH] [int] NULL,
[STEATLNC] [int] NULL,
[STPAYLNC] [int] NULL,
[STRPTCL1] [int] NULL,
[STRPTCL6] [int] NULL,
[STRPTCL10] [int] NULL,
[STTLTSCH] [int] NULL,
[STAGEENG] [int] NULL,
[STITMOW1] [int] NULL,
[STITMOW2] [int] NULL,
[STITMOW3] [int] NULL,
[STITMOW4] [int] NULL,
[STITMOW5] [int] NULL,
[STITMOW6] [int] NULL,
[STITMOW7] [int] NULL,
[STITMOW8] [int] NULL,
[STBRWBK] [int] NULL,
[STREADFN] [int] NULL,
[STREADCH] [int] NULL,
[STREADLR] [int] NULL,
[STPLHLRD] [int] NULL,
[STPLHL01] [int] NULL,
[STPLHL02] [int] NULL,
[STPLHL03] [int] NULL,
[STPLHL04] [int] NULL,
[STPLHL05] [int] NULL,
[STPLHL06] [int] NULL,
[STPLHL07] [int] NULL,
[STSPEN01] [int] NULL,
[STSPEN02] [int] NULL,
[STSPEN03] [int] NULL,
[STSPEN04] [int] NULL,
[STSPEN05] [int] NULL,
[STATEN01] [int] NULL,
[STATEN02] [int] NULL,
[STATEN03] [int] NULL,
[STATEN04] [int] NULL,
```

```
[STNONSCL] [int] NULL,
[STGR1001] [int] NULL,
[STGR1002] [int] NULL,
[STGR1003] [int] NULL,
[STGR1004] [int] NULL,
[STGR1005] [int] NULL,
[STGR1006] [int] NULL,
[STGR1007] [int] NULL,
[STGR1008] [int] NULL,
[STGR1009] [int] NULL,
[STGR1011] [int] NULL,
[STHGHGRD] [int] NULL,
[ENG_TEST] [nvarchar](50) NULL,
[ENG_RAWSCORE] [int] NULL,
[MATH_TEST] [nvarchar](50) NULL,
[MATH_RAWSCORE] [int] NULL,
[GRLENRL] [int] NULL,
[BOYENRL] [int] NULL,
[TTLENRL] [int] NULL,
[TGRLENRL] [int] NULL,
[TBOYENRL] [int] NULL,
[TTTLENRL] [int] NULL,
[ATDTMSY] [int] NULL,
[ATDDFSY] [int] NULL,
[TGMTHENG] [int] NULL,
[SCALLCT] [int] NULL,
[GRPABLTY] [nvarchar](50) NULL,
[CLSORD1] [nvarchar](50) NULL,
[CLSORD2] [nvarchar](50) NULL,
[SCAVLB1] [int] NULL,
[SCAVLB2] [int] NULL,
[SCAVLB3] [int] NULL,
[SCAVLB4] [int] NULL,
[SCAVLB5] [int] NULL,
[SCAVLB6] [int] NULL,
[SCAVLB7] [int] NULL,
[SCAVLB8] [int] NULL,
[SCAVLB9] [int] NULL,
[SCAVLB10] [int] NULL,
```

```
    [SCPRDDAY] [int] NULL,
    [SCLNONPR] [int] NULL,
    [SCMNMTIN] [int] NULL,
    [SCMNENIN] [int] NULL,
    [SCTXTMTH] [int] NULL,
    [SCTXTENG] [int] NULL,
    [NUMG10CLS] [int] NULL,
 CONSTRAINT [PK_VietnamDataset] PRIMARY KEY CLUSTERED
(
    [UNIQUEID] ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF,
IGNORE_DUP_KEY = OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON)
ON [PRIMARY]
) ON [PRIMARY]
GO
```

3.4   INSERTING DATA INTO TABLE

The data needed for this analysis was inserted into the newly created table using the following script:

```
INSERT INTO [dbo].[VietnamDataset](
    [UNIQUEID],
    [SCHOOLID],
    [CLASSID],
    [STUDENTID],
    [YLCHILDID],
    [PROVINCE],
    [DISTRICTCODE],
    [LOCALITY],
    [GENDER],
    [AGE],
    [ETHNICITY],
    [ABSENT_DAYS],
    [MOM_READ],
    [MOM_EDUC],
```

[DAD_READ],
[DAD_EDUC],
[STDYLCHD],
[STDCMPLT],
[STDDINT],
[STDLIV],
[STDLNGHM],
[STDMEAL],
[STDHLTH1],
[STDHLTH2],
[STDHLTH3],
[STDHLTH4],
[STDHLTH5],
[STDHLTH6],
[STDHLTH0],
[STPPLHM],
[STSIBOLD],
[STSIBYNG],
[STNMBOOK],
[STPLSTDY],
[STHVMTEL],
[STHVRADO],
[STHVTELE],
[STHVBIKE],
[STHVMTBK],
[STHVDESK],
[STHVCHR],
[STHVLAMP],
[STHVEFAN],
[STHVAIRC],
[STHVCAR],
[STHVCOMP],
[STHVINTR],
[STHVFRDG],
[STHVMCRO],
[STHVDVD],
[STHVCBLE],
[STTMSCH],
[STEATLNC],

[STPAYLNC],
[STRPTCL1],
[STRPTCL6],
[STRPTCL10],
[STTLTSCH],
[STAGEENG],
[STITMOW1],
[STITMOW2],
[STITMOW3],
[STITMOW4],
[STITMOW5],
[STITMOW6],
[STITMOW7],
[STITMOW8],
[STBRWBK],
[STREADFN],
[STREADCH],
[STREADLR],
[STPLHLRD],
[STPLHL01],
[STPLHL02],
[STPLHL03],
[STPLHL04],
[STPLHL05],
[STPLHL06],
[STPLHL07],
[STSPEN01],
[STSPEN02],
[STSPEN03],
[STSPEN04],
[STSPEN05],
[STATEN01],
[STATEN02],
[STATEN03],
[STATEN04],
[STNONSCL],
[STGR1001],
[STGR1002],
[STGR1003],

[STGR1004],
[STGR1005],
[STGR1006],
[STGR1007],
[STGR1008],
[STGR1009],
[STGR1011],
[STHGHGRD],
[ENG_TEST],
[ENG_RAWSCORE],
[MATH_TEST],
[MATH_RAWSCORE],
[GRLENRL],
[BOYENRL],
[TTLENRL],
[TGRLENRL],
[TBOYENRL],
[TTTLENRL],
[ATDTMSY],
[ATDDFSY],
[TGMTHENG],
[SCALLCT],
[GRPABLTY],
[CLSORD1],
[CLSORD2],
[SCAVLB1],
[SCAVLB2],
[SCAVLB3],
[SCAVLB4],
[SCAVLB5],
[SCAVLB6],
[SCAVLB7],
[SCAVLB8],
[SCAVLB9],
[SCAVLB10],
[SCPRDDAY],
[SCLNONPR],
[SCMNMTIN],
[SCMNENIN],

```sql
        [SCTXTMTH],
        [SCTXTENG],
        [NUMG10CLS]
SELECT
--COLUMNS FROM PARENT TABLE
        [UNIQUEID],
        [SCHOOLID],
        [CLASSID],
        [STUDENTID],
        [YLCHILDID],
        [PROVINCE],
        [DISTRICTCODE],
        [LOCALITY],
        [GENDER],
        [AGE],
        [ETHNICITY],
        [ABSENT_DAYS],
        [MOM_READ],
        [MOM_EDUC],
        [DAD_READ],
        [DAD_EDUC],
        [STDYLCHD],
        [STDCMPLT],
        [STDDINT],
        [STDLIV],
        [STDLNGHM],
        [STDMEAL],
        [STDHLTH1],
        [STDHLTH2],
        [STDHLTH3],
        [STDHLTH4],
        [STDHLTH5],
        [STDHLTH6],
        [STDHLTH0],
        [STPPLHM],
        [STSIBOLD],
        [STSIBYNG],
        [STNMBOOK],
        [STPLSTDY],
```

[STHVMTEL],
[STHVRADO],
[STHVTELE],
[STHVBIKE],
[STHVMTBK],
[STHVDESK],
[STHVCHR],
[STHVLAMP],
[STHVEFAN],
[STHVAIRC],
[STHVCAR],
[STHVCOMP],
[STHVINTR],
[STHVFRDG],
[STHVMCRO],
[STHVDVD],
[STHVCBLE],
[STTMSCH],
[STEATLNC],
[STPAYLNC],
[STRPTCL1],
[STRPTCL6],
[STRPTCL10],
[STTLTSCH],
[STAGEENG],
[STITMOW1],
[STITMOW2],
[STITMOW3],
[STITMOW4],
[STITMOW5],
[STITMOW6],
[STITMOW7],
[STITMOW8],
[STBRWBK],
[STREADFN],
[STREADCH],
[STREADLR],
[STPLHLRD],
[STPLHL01],

[STPLHL02],
[STPLHL03],
[STPLHL04],
[STPLHL05],
[STPLHL06],
[STPLHL07],
[STSPEN01],
[STSPEN02],
[STSPEN03],
[STSPEN04],
[STSPEN05],
[STATEN01],
[STATEN02],
[STATEN03],
[STATEN04],
[STNONSCL],
[STGR1001],
[STGR1002],
[STGR1003],
[STGR1004],
[STGR1005],
[STGR1006],
[STGR1007],
[STGR1008],
[STGR1009],
[STGR1011],
[STHGHGRD],
[ENG_TEST],
[ENG_RAWSCORE],
[MATH_TEST],
[MATH_RAWSCORE],
[GRLENRL],
[BOYENRL],
[TTLENRL],
[TGRLENRL],
[TBOYENRL],
[TTTLENRL],
[ATDTMSY],
[ATDDFSY],

```
       [TGMTHENG],
       [SCALLCT],
       [GRPABLTY],
       [CLSORD1],
       [CLSORD2],
       [SCAVLB1],
       [SCAVLB2],
       [SCAVLB3],
       [SCAVLB4],
       [SCAVLB5],
       [SCAVLB6],
       [SCAVLB7],
       [SCAVLB8],
       [SCAVLB9],
       [SCAVLB10],
       [SCPRDDAY],
       [SCLNONPR],
       [SCMNMTIN],
       [SCMNENIN],
       [SCTXTMTH],
       [SCTXTENG],
       [NUMG10CLS]
FROM VietnamDF
```

3.5   CREATING VIEWS

Each numerical value in columns of the table created has real meaning which is contained in the data dictionary. So to make the data more sensible views were created with the appropriate meaning of these numerical values. In light of this, two views were created using the following script:

```
--CREATE VIEW CONTAINING STUDENT INFO AND SCORES
CREATE VIEW [dbo].[students_scores]
AS
Select Uniqueid,
case when GENDER=1 then 'Male' else 'Female'
end as Gender,
CASE WHEN ETHNICITY = 1 THEN 'Kinh'
```

```sql
WHEN ETHNICITY = 2 THEN 'H'Mong'
WHEN ETHNICITY = 3 THEN 'Cham-HRoi'
WHEN ETHNICITY = 4 THEN 'Ede'
WHEN ETHNICITY = 5 THEN 'Ba Na'
WHEN ETHNICITY = 6 THEN 'Nung'
WHEN ETHNICITY = 7 THEN 'Tay'
WHEN ETHNICITY = 8 THEN 'Dio'
WHEN ETHNICITY = 9 THEN 'Giay'
WHEN ETHNICITY = 10 THEN 'Others'
END AS ETHNICITY,
case when LOCALITY = 1 then 'Rural'
else 'Urban'
end as Locality,
CASE WHEN PROVINCE = 1 THEN 'Ben Tre'
WHEN PROVINCE = 2 THEN 'Da Nang'
WHEN PROVINCE = 3 THEN 'Hung Yen'
WHEN PROVINCE = 4 THEN 'Lao Cai'
ELSE 'Phu Yen'
END AS PROVINCE,
SCHOOLID,
Math_Rawscore, Eng_Rawscore
From VietnamDataset
```

Figure 3: Child's basic information

```sql
--VIEW CONTAIN PARENTS EDUCATIONAL INFO
CREATE VIEW [dbo].[parent_education]
AS
SELECT Uniqueid,
case when GENDER=1 then 'Male' else 'Female'
end as Gender,
case when MOM_READ=0 then 'Yes'
when MOM_READ=1 then 'Yes'
else 'Null'
end as MOM_READ,
CASE WHEN MOM_EDUC = 0 THEN 'No_edu'
WHEN MOM_EDUC = 1 THEN 'Grade 1-5'
WHEN MOM_EDUC = 2 THEN 'lower sec sch'
WHEN MOM_EDUC = 3 THEN 'voc training'
WHEN MOM_EDUC = 4 THEN 'upper sec sch'
WHEN MOM_EDUC = 5 THEN 'higher edu'
WHEN MOM_EDUC = 6 THEN 'Dont know'
ELSE 'Null'
```

```
END AS MOM_EDUC,
case when DAD_READ=0 then 'Yes'
when DAD_READ=1 then 'Yes'
else 'Null'
end as DAD_READ,
CASE WHEN DAD_EDUC = 0 THEN 'No_edu'
WHEN DAD_EDUC = 1 THEN 'Grade 1-5'
WHEN DAD_EDUC = 2 THEN 'lower sec sch'
WHEN DAD_EDUC = 3 THEN 'voc training'
WHEN DAD_EDUC = 4 THEN 'upper sec sch'
WHEN DAD_EDUC = 5 THEN 'higher edu'
WHEN DAD_EDUC = 6 THEN 'Dont know'
ELSE 'Null'
end as DAD_EDUC
FROM VietnamDataset
```

	Uniqueid	Gender	MOM_READ	MOM_EDUC	DAD_READ	DAD_EDUC
1	VN1203101	Female	Yes	Dont know	Yes	Dont know
2	VN1203102	Male	Yes	lower sec sch	Yes	lower sec sch
3	VN1203103	Female	Yes	Grade 1-5	Yes	lower sec sch
4	VN1203104	Female	Yes	lower sec sch	Yes	Grade 1-5
5	VN1203105	Male	Yes	upper sec sch	Yes	lower sec sch
6	VN1203106	Female	Yes	Grade 1-5	Yes	Grade 1-5
7	VN1203107	Female	Yes	Dont know	Yes	Dont know
8	VN1203108	Female	Yes	lower sec sch	Yes	lower sec sch
9	VN1203109	Female	Yes	Grade 1-5	Yes	Dont know
10	VN1203110	Female	Yes	Dont know	Yes	Dont know
11	VN1203111	Female	Yes	Dont know	Yes	Dont know
12	VN1203112	Male	Yes	Dont know	Yes	Dont know
13	VN1203113	Female	Yes	lower sec sch	Yes	Grade 1-5
14	VN1203114	Female	Yes	No_edu	Yes	lower sec sch
15	VN1203115	Female	Yes	Grade 1-5	Yes	Grade 1-5

Query executed successfully.   DESKTOP-DLCD2E0\SQLEXPRESS

Figure 4: Showing parents educational information

4.0   REPORT DESIGN

To create a simple reporting system to enable the client measure and understand the inequality in the education sector in Vietnam, the following summarized reports were created:

- Performance by province
- Performance by gender
- Inequalities by ethnic group
- Performance by locality

INEQUALITIES BY PROVINCE

To understand education inequality in Vietnam, the average performance of students in English and Mathematics in the various province in which this survey was conduction was analyzed to ascertain the gap in learning outcomes between the various Provinces. This was obtained using the following query:

```
CREATE VIEW [dbo].[performance_by_province]
AS
SELECT PROVINCE, COUNT(DISTINCT SCHOOLID) AS no_of_schs,
avg(Math_Rawscore) as avg_mathscore, AVG(Eng_Rawscore) AS
avg_engscore
FROM students_scores
GROUP BY PROVINCE
```

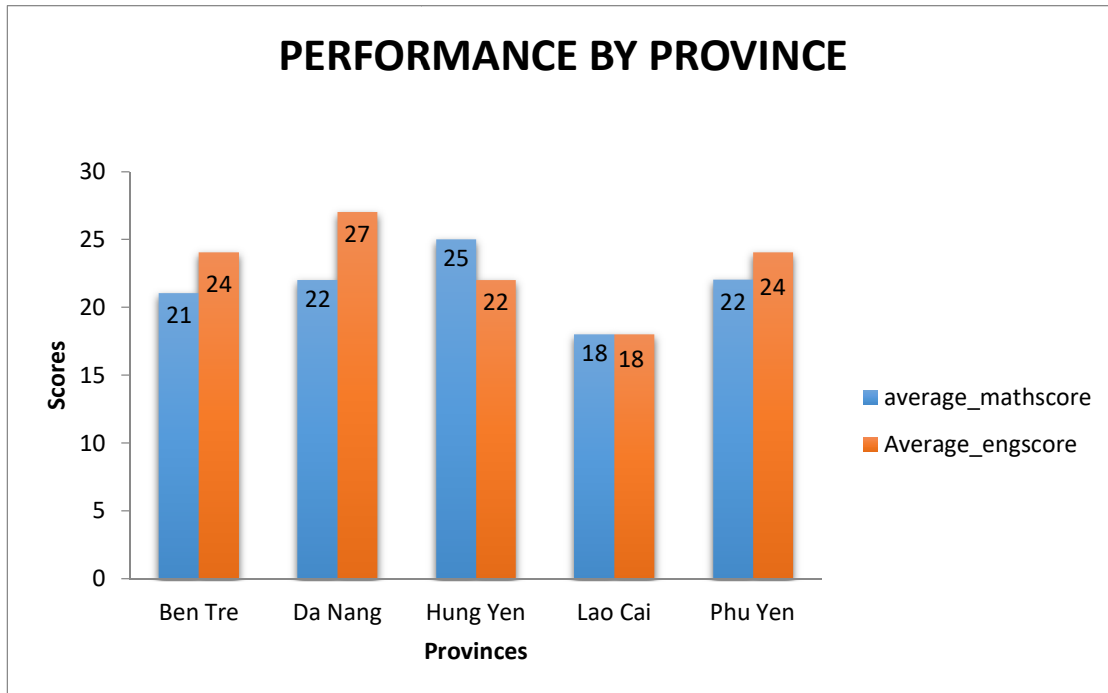| Provinces	average_mathscore	Average_engscore
Ben Tre	21	24
Da Nang	22	27
Hung Yen	25	22
Lao Cai	18	18
Phu Yen	22	24

Figure 5: Students learning outcome performance by province

A subtle gap in learning outcome was observed between the various provinces. My analysis shows that in the entire area surveyed, the general performance of students in English language is better than their performance in Mathematics with an average score of 22.2 and 20.2 respectively. On a provincial level, on an average students in Da Nang performed best in English Language followed by Ben Tre and Phu Yen which had an equal average score. They were followed by Hung Yen while Lao Cai had the least performance in English. In mathematics, Hung Yen performed best while Phu Yen and Da Nang were on a par. Ben Tre was next in the herachty while Lao Cai was the least performer.

PERFORMANCE BY LOCALITY

To further understand how much effect locality has on students learning outcome a simple report was generated using the following query:

```sql
CREATE VIEW [dbo].[performance_by_locality]
AS
Select LOCALITY,
PROVINCE, count(Uniqueid) as counts,
avg(Math_Rawscore) as math_score, avg(Eng_Rawscore) as eng_score
from students_scores
group by Locality, PROVINCE
```

	LOCALITY	PROVINCE	counts	math_score	eng_score
1	Rural	Lao Cai	768	17	18
2	Urban	Lao Cai	968	18	18
3	Rural	Phu Yen	1174	21	23
4	Urban	Hung Yen	393	17	17
5	Rural	Ben Tre	1172	23	25
6	Urban	Phu Yen	1188	22	25
7	Urban	Ben Tre	539	17	20
8	Urban	Da Nang	1639	22	27
9	Rural	Hung Yen	899	28	25

Query executed successfully. | DESKTO
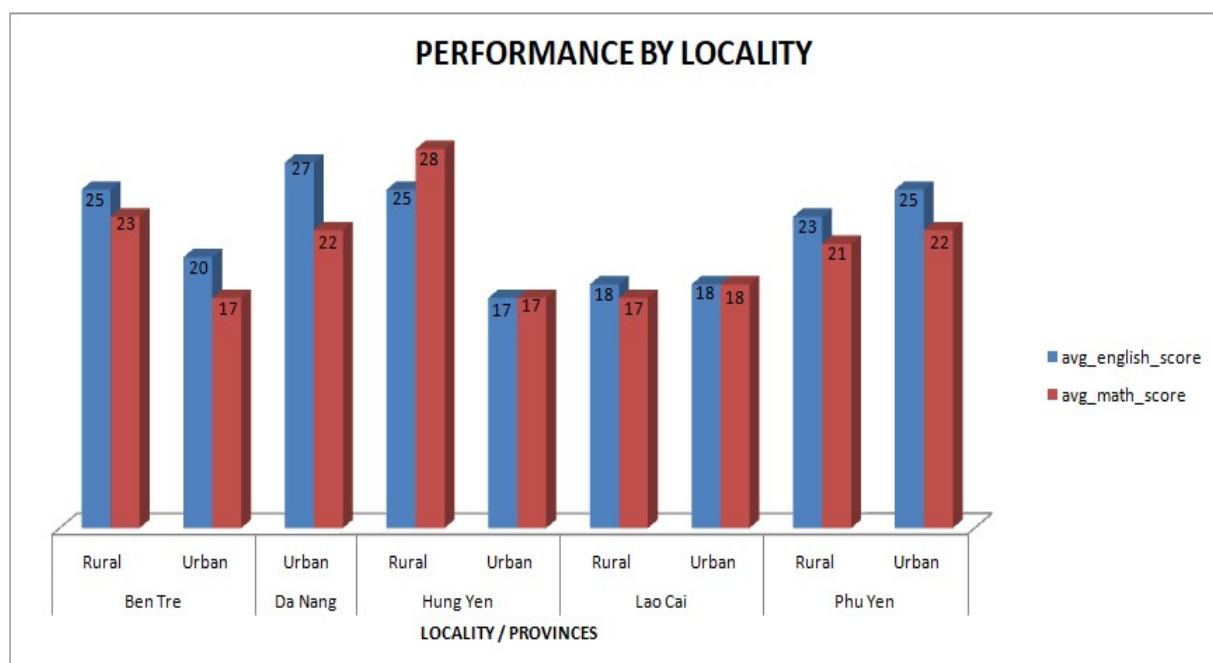

Figure 6: Learning outcome by Locality

My analysis shows a gap between the performance of students in the Rural and Urban area of Ben Tre and Hung Yen while no gap exists between students learning outcome in the rural and urban area of Lao Cai and Phu Yen. Da Nang is typically an urban area and it shows a very high students performance.

PERFORMANCE BY GENDER

To analyze students performance by gender across the entire surveyed area, students who obtained the score of 20 (which is the average score) were considered to have passed the exam. Based on this a view comprising of students who passed was created and students' performance by gender was analyzed using the codes below:

```
---STUDENTS WHO PASS THE EXAMINATION
CREATE VIEW [dbo].[std_passed]
AS
SELECT a.GENDER, a.std_pass_eng, b.std_pass_math
FROM
(SELECT GENDER, COUNT(ENG_RAWSCORE) AS std_pass_eng
FROM students_scores
WHERE ENG_RAWSCORE >= 20 AND GENDER IS NOT NULL
GROUP BY GENDER) as a,
(SELECT GENDER, COUNT(MATH_RAWSCORE) as std_pass_math
FROM students_scores
WHERE MATH_RAWSCORE >= 20 AND GENDER IS NOT NULL
GROUP BY GENDER) AS b
WHERE a.GENDER = b.GENDER
```

Another view was created to show the gap between students learning outcome by gender using the following lines of code:

```
---STUDENTS PERFORMANCE BY GENDER
CREATE VIEW [dbo].[perf_by_gender]
AS
SELECT a.Gender,
std_pass_eng,
std_pass_math,
```

```
total_std
FROM
(SELECT * FROM std_passed) AS a,
(select GENDER, count(UNIQUEID) AS total_std
FROM students_scores
where GENDER IS NOT NULL
group by GENDER) AS b
WHERE a.GENDER= b.GENDER
```

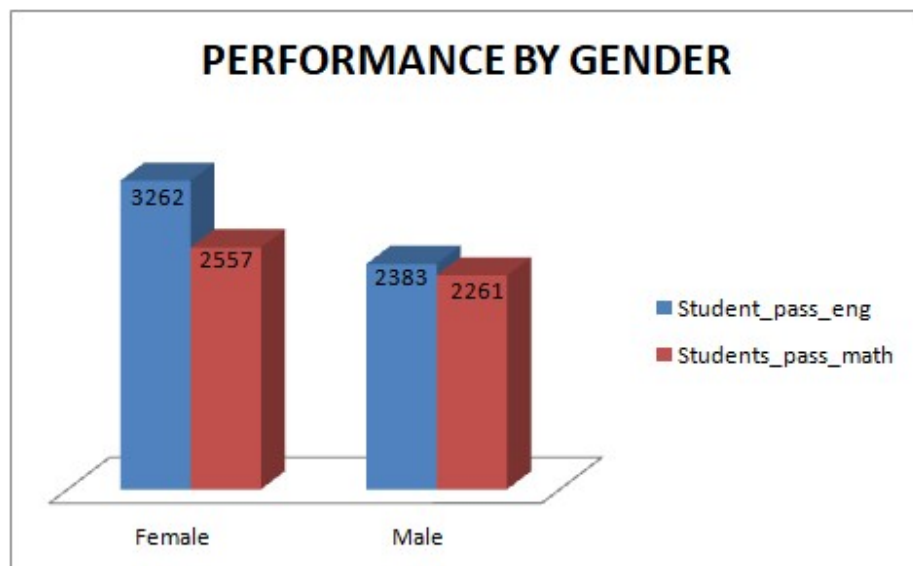|	Gender	std_pass_eng	std_pass_math	total_std
1	Male	2383	2261	4185
2	Female	3262	2557	4555

Figure 7: Learning outcome by Gender

My analysis shows that females perform better than males in both mathematics and English. A total of 4555 female students participated in the test and 3262 students scored at least 20 marks in the English test while 2557 students scored at least 20marks in mathematics. A total of 4185 male students participated in the test and 2383 students passed English while 2261 students passed mathematics.

INEQUALITY BY ETHNICITY

To understand the inequality in education among the various ethnic groups in Vietnam, a view was created to count the number of students for each ethnic group with the aim of ascertaining the ethnic group with the most dominant student population. This was done using the following line of code:

```sql
--STUDENT POPULATION BY ETHNIC GROUP
CREATE VIEW [dbo].[std_by_ethnicity]
as
select temp.ETHNICITY, temp.no_of_stds_per_ethnic,
temp.Total_std, round((temp.no_of_stds_per_ethnic *
100.0/temp.Total_std),2) AS std_percent
from
(select ETHNICITY,
count(uniqueid) as no_of_stds_per_ethnic,
 (
    SELECT count(*)
    FROM VietnamDataset
    WHERE ETHNICITY IS NOT NULL
  ) AS Total_std
from students_scores
WHERE ETHNICITY IS NOT NULL
GROUP BY ETHNICITY) as temp
```

	ETHNICITY	no_of_stds_per_ethnic	Total_std	std_percent
1	Ede	22	8413	0.260000000000
2	Others	52	8413	0.620000000000
3	Dio	235	8413	2.790000000000
4	Giay	141	8413	1.680000000000
5	Kinh	7368	8413	87.580000000000
6	Cham-HRoi	37	8413	0.440000000000
7	Tay	98	8413	1.160000000000
8	Ba Na	4	8413	0.050000000000
9	H'Mong	376	8413	4.470000000000
10	Nung	80	8413	0.950000000000

Query executed successfully. | DESKTOP-DLCL