**ABSTRACT**

There are a lot of inequality in the education sector of the world in general. The purpose of this analysis is to analyze the inequality in the education sector in Vietnam. The resulting findings from this analysis is to initiate a project helping children from low-income families. The reports are analysis and generated using Young lives: School Survey, Vietnam, 2016-2017 dataset.

**INTRODUCTION**

Young Lives is a longitudinal cohort study investigating the changing nature of childhood poverty in four low- and middle-income countries. The project gathers both household and child-level data to improve understanding of the causes and consequences of childhood poverty and examine how policies affect children’s well-being, in order to inform the development of future policy and to target child welfare interventions more effectively. I makes use of T-SQL and Microsoft excel to carries out the analysis, designs and reports. It also covers data protection issues.

The datasets used are from the Uk Data service websites. They are in an archive forms. Young Lives administers makes use of two sets of questionnaires in every survey round: a household questionnaire typically answered by the child’s main caregiver and a child questionnaire answered by the Young Lives child.

The constructed files are combined sub-sets of variables from Rounds 1 to 5 of the Young Lives household and child surveys conducted from 2002 to 2016. This analysis will create tables and vies using t-SQL. It will also create various summarized reports in excel. Also filtering, sorting and grouping functionality in the reports.

**DESIGN RATIONALE**

| **Table1.1 process for design of Young lives Project** | |
| --- | --- |
| Process | Description |
| 1. Research | It comprises of a school survey in Vietnam in the year 2016-2017 dataset. It main elements are child questionnaire household questionnaire and community questionnaire. The total number of variables in the dataset is 406 and the total number of cases is 8740 |
| 1. refining | Relevant data is check from the files with the use of T-SQL |
| 1. Explore | Exploration is also done with the use of T-SQL. |
| 1. Goals and Objectives | To initiate a project helping children from low-income families. |
| 1. Infrastructure for the project | There are many files in the dataset |

**5.0 Design considerations**

5.1 **Database Normalizations**

There are a lot of row that needs to be normalize because when importing the files into SSMS it bringing out error. The image below shows one of the error it bringing out.

I went back to the database and deleted from the system and started the process all over.

Also there was also an error when generating the second file which is truncated error, it was generated because the files has a lot of empty rows, it was ignored because there is nothing that can be done.

The data was also change to char in other to get ACCURATE RESULTS

**Constraints**

There are a lot of constraint in the file used. The following are the list of constraint in the files.

1. Null Constraints: there are no null files in both data sets.
2. Unique constraint: I make sure that the columns are all different when working all the data. All duplicates were deleted.

**Data Validation**

Data validation was done on using **T-SQL.**

1. Firstly the data was view and it consist of many rows and columns.
2. The data was counted and a total of 8740 rows were found.
3. We also check if there is no null values. There are no null value in the data.
4. Also there is no duplicate values in the system.
5. After joining the two data together we get 3101 rows.

**Error Handling**

1. I am make sure I stated the SQL statements properly and then put into Try-Catch block to get error.
2. Also I create a log table for storing errors generated by the stored procedures.
3. Also I provide operator with a recall function to open the error log file and display errors.
4. Also I handle logs all the error, not just the ones that caused the error resolving to occur.

**Security**

A database server was created using the command prompt. There series of in the localdb created .the features includes encrypted communication over SSL/TLS, the Windows Data Protection API (DPAPI) used to encrypt data at rest, authentication and authorization.

**Comments**

A comment is written before each query. It is begin with a slash and an asterisk(/\*). The comments explain the query so as other can understand what is being done.

**T-SQL Statements**

**Tables**

Tables was created using some set of query. The tables created were gender, student details, and Head teachers details.

The tables query started with the join of the two dataset provided which is the inner join. Then create tables was then use to create the tables.

**Views**

The tables created were then view using the create tables query. The view is a SQL statement that is stored in the database with a associated name. it is a composition of a table in the form of a predefined SQL query which is the table created.it contains all rows in the tables created.

**Database Security**

Database Security is defined as the ranges of tools, measures and control designed to establish and preserve database confidentiality integrity, and availability.

These are some of the techniques used in the database security techniques:

1. Access Control Mechanisms: for example when someone tries to access data Object, it check the rights of the user AGAINST set of authorizations. The was used when accessing the Young Lives website, the sites ask for authority in order to acess the files on the dataset.
2. Content based access control: Access control should be based on contents of the data. This is also applicable to Young Lives site

Because with filling a form someone will not have access to the data.

1. Mandatory Access Control Model: Mandatory access is based on the classification of data objects and users. The sites is also based that way. You can only access a particular dataset when someone fill the forms.

Data.

**Database Backup and Restore Strategy**

A backup and recovery strategy addresses both data growth and security issues.

It provides a similar level of service for all platforms, enabling a single database administrator to control both backup and recovery operations and the backup files themselves. A database backup was done on this project, which includes the following:

1. A new folder was created and it was called project.
2. SSMS was created and connect to an instance of SQL Server.
3. In Object Explorer, the server tree was expanded and the databases’
4. It was then right-click and tasks, Back Up was selected respectively.
5. The appropriate selection was done to get the database backup.
6. It was then backup to a disk.

**Report Designs**

Report Designs is the act of visualizing data and also reporting it on a dashboard.

I make use of Microsoft excel as a presentation and also provide front-end design to link the file together

Pivot table is used, it involve the use of filtering, sorting and grouping to visualize the data.

The first report the school id and gender. it shows the relationship between school id and gender.

It reveals that female student are more in some school than another.

The second report shows the age of the students. It shows that most of the student are of the age 15 years, but some are other than 15years, it shows that there is some inequality in the education sectors in Vietnam.

The third report shows how often the teacher gives mathematics assignment in each school.

It shows that some teacher do not give assignment regularly. So there are inequality in the Vietnam city.

**Data Science/Business Intelligence Techniques**

A Interactive Report was created in other to be able to visualize the data. A lot of patterns where created with this. A report was then reported about the data.

**Data Privacy, Ethical and legal issues**

Ethical issues are not governed by a set of rules and thereby are not punishable by law. All ethical are observed when doing the project.

Data Privacy are also observed when downloading the data. A form was filled before getting access to the data.

**Conclusion**

In conclusion, there are lot of inequality in the education sector in Vietnam. These are some of the ways to stop the inequality in the education sector:

1. The Teachers should have a schedule for the number of assignment in each weeks.
2. There should be a regular age for the students in each school.
3. They should be a balance in the gender of the in each school.