

**ANLY482 Analytics Practicum**

**Interim Report**

Team 15

Nguyen Le Hong Ngoc

Poh Jin Hui

Zhao Yazhi

Table of Contents

Changes in project sponsor 3

Problem introduction 3

Business Objectives 3

Analytical Problems 3

Scope of work & Deliverables 4

Methodology 4

Tools 4

Data Preparation 4

Analytics cube 5

Exploratory Data Analysis 5

Revised work plan 8

References: 9

Appendix A: Change in project sponsor 10

# Changes in project sponsor

After much consideration, our team has decided to withdraw from the KTPH project due to insufficient richness of data provided and other resource constraints. (Please refer to Appendix A for full account). Moving forth, our team will be focusing on the analysis of PISA survey data related to education in Singapore, sponsored by Professor Kam Tin Seong.

# Problem introduction

The Ministry of Education (MOE) of Singapore collects and analyses data from schools island wide to continually improve policies and practices in Education. However, most of this data are not publicly available for research and analysis by those outside the Ministry. Hence, the sponsor seeks to gain insights about education in Singapore from the publicly available data collected by the OECD through their “Programme for International Student Assessment” (PISA) survey. The PISA is a triennial international survey that aims to evaluate education systems worldwide by testing the skills and knowledge of 15-year-old students. The most recently published results are from the assessment in 2012, which will be used in our analysis.

# Business Objectives

The project aims to investigate the effects of various personal and environmental factors on the cognitive abilities of secondary school students in Singapore, especially their family background and overall academic environment. Key findings in this project can serve as insights for policy makers and educators to further improve current practices for the creation of better learning environments and processes.

# Analytical Problems

To achieve the business objectives, the following questions need to be addressed:

*In terms of school resource allocation:*

* How are resources in schools allocated to different aspects, such as teachers, computers, network, learning facilities and activities?
* What is the outlook of student-teacher relationship and how might it affect student’s emotional well-being at school?
* What are the prominent problems schools in Singapore are facing and their relationship with student academic performance?
* Does student-teacher ratio matter to student well-being and teacher morale? Overall does it affect student performance in school?

*In terms of student profile*

* Is student performance correlated with their socio-economic status and parents’ education level?
* Is student performance affected by emotional well-being or vice-versa?
* What lifestyle factors might contribute to students’ emotional well-being in school?

# Scope of work & Deliverables

The scope of this project is based on data available from the PISA2012 survey, specifically for Singapore. The PISA survey samples groups of 35-40 students from each of the 172 selected Singapore secondary schools. Data collected includes information for family background, parents’ education level, student possession at home, school funding, staff headcounts and profile, facilities and prominent issues (truancy and shortage of resources). A 2-hour test conducted by PISA measures students’ competency in Mathematics, Science and Reading. Overall, the PISA2012 survey results will serve as a rich source of data to conduct our analysis for the education landscape in Singapore.

At the end of our project, we aim to deliver a storyboard of all key findings from the PISA2012 data and possible recommendations for improved school resource management and practices, based on insights from our analysis.

# Methodology

*Frequency Analysis*

* To examine resources given to different secondary schools in Singapore
* To understand the purposes for which school resources are used

*Correlation Analysis*

* + To examine the extent that resource availability in schools affect student performance/ To evaluate if problems of shortages in school (teaching staff, facilities) are associated with low performance
  + To examine if family background and socio-economic status are linked to student performance
  + Is emotional well-being necessarily associated with good performance?

## Tools

The main tool for analysis is SAS Enterprise Guide. In our project, we will make use of features such as Data exploration, Correlation Analysis, Frequency Analysis, Sorting and Filtering, and Query Builder.

# Data Preparation

Text data is retrieved from the PISA2012 database found on the PISA official website. The PISA survey results comprises of items from 5 questionnaires: student, school, parent, cognitive item and score cognitive item. In our project, we only require data relevant to Singapore.

The format of text data is not readily readable by SAS or any analytics tool. Each record is stored as a long sequence of characters, where a fixed length of character would represent a value. In order to convert the data into format that SAS Enterprise Guide can read, a program is created in SAS to read the character sequences, separate them into smaller parts and assign to appropriate attributes. We then extracted and exported the data for Singapore into a separate set of tables for our analysis.

## Analytics cube

*Tables of data to be used in the analysis*

* Student questionnaire data (stu): this table contains student demographic information, parents’ education, interest in school subjects, engagement with activities outside school, access to information communication technology (ICT), familiarity with academics concept and sense of belonging in school.
* School questionnaire data (sch): this table contains information about school sources of funding, size of student population, staff headcounts, availability of ICT for student and teacher use, shortage of resources (if any), co-curricular activities, parent participation, teacher morale and learning hindrance (if any).
* Score cognitive item data (cogs): this table records the score of students subjected to PISA tests in Reading, Mathematics and Science. Since student grades are not available to us as far as the scope is concerned, we will use this table as a measure for student academic performance

*Integrating tables to support more queries*

* **SchoolID** and **StudentID** can be used to link the 3 tables above. This integration is crucial in discovering possible correlation between various aspects of school environment and student performances, which offers valuable insights to education policy makers such as MOE and school management.

# Exploratory Data Analysis

*Preliminary findings*

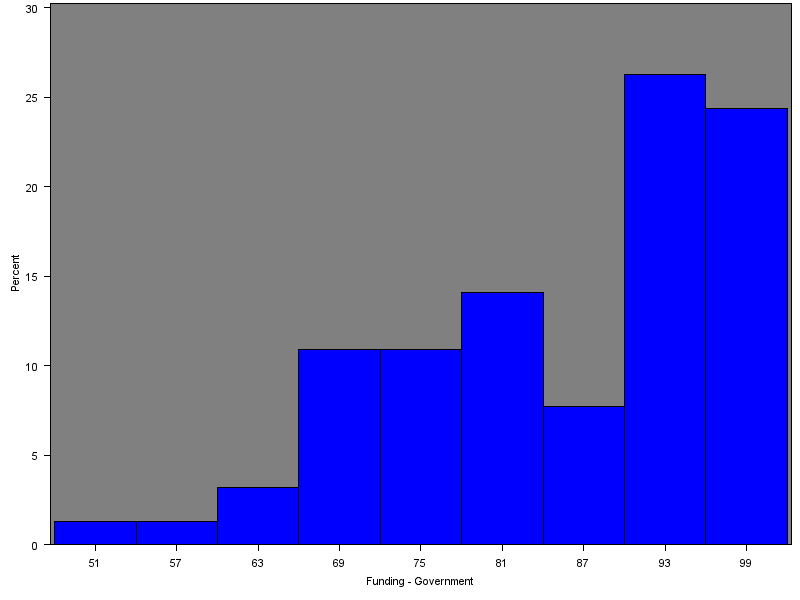


Fig 1: Percentage of government funding received by public schools

From Fig 1 above, we observe that there seems to be a significant discrepancy in government funding provided to the public schools. This is a potentially useful observation, as differences in funding will affect the amount of resources available to the schools and hence contributes to different learning environments and opportunities available to students. Student performance may be affected by such difference in allocation of education resources.

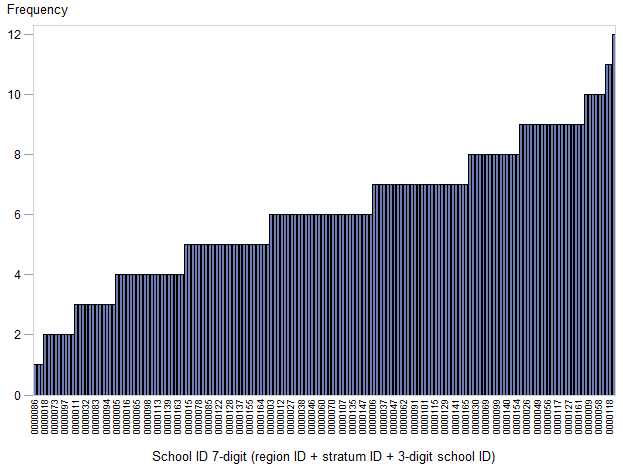


Fig 2: Distribution of scores across students from different schools

From Fig 2, we observe that scores of students differ significantly across secondary schools. In addition, secondary schools in Singapore do not allocate resources efficiently by PISA standards, with a mean index of -0.36797 (range is -0.80 to 9999.0). There is also a slight negative correlation between percentage funding given by government and index of school responsibility for resource allocation.

# Revised work plan

|  |  |  |  |
| --- | --- | --- | --- |
| **Week** | **Dates** | **Actions** | **Deliverables** |
| Week 8 | 22 Feb – 28 Feb | 1.     Group meeting to discuss findings on small area estimation for KTPH project  2.     Consultation with Professor to discuss approach to small area estimation  3.     Group meeting and consultation with professor to discuss change of project and new project requirements  4.     Group research on PISA 2012 reports  5.     Clean and extract PISA data for Singapore  6.     Conduct exploratory data analysis (EDA)  7.     Group meeting to brainstorm on possible project scope and preliminary findings of EDA | 28 Feb 11.59PM  Submission of interim report and update wikipage |
| Week 9 | 29 Feb – 6 Mar | 1.     Conduct EDA  2.     Group meeting to discuss EDA findings and further direction of analysis  3.     Sponsor meeting to confirm scope and discuss EDA findings |  |
| Week 10 | 7 Mar – 13 Mar | 1.     Iterative refinement of analytical questions to address business objectives  2. Group meeting to review findings & brainstorming sessions  3. Sponsor meeting to share findings |  |
| Week 11 | 14 Mar – 20 Mar | 1.     Integrate data tables to facilitate more complex queries. Additional analytical questions can be raised.  2. Group meeting to review findings & brainstorming sessions  3. Sponsor meeting to share findings |  |
| Week 12 | 21 Mar – 27 Mar | 1. Explore additional analytical techniques to gather insights  2. Business objectives and analytical problems finalized  3. Sponsor meeting to update progress |  |
| Week 13 | 28 Mar – 3 Apr | 1.     Compilation of findings for analytical problems  2. Selection of key findings to present in storyboard  3. Sponsor meeting to discuss content of storyboard and final report |  |
| Week 14 | 4 Apr – 10 Apr | 1.     Final presentation slides  2.     Final report  3.     Final wiki page update |  |
| Week 15 | 11 Apr – 17 Apr | 1.     Final presentation slides  2.     Final report  3.     Final wiki page update  4.     Poster | Final report submission  Final presentation  Poster submission |

# References:

http://www.oecd.org/pisa/aboutpisa/

# Appendix A: Change in project sponsor

After much consideration, our team has decided to withdraw from the KTPH project due to insufficient richness of data provided and resource constraints. For ANLY 482, teams are expected to discover understanding or insights from the data given using appropriate data analytics techniques. Unfortunately, after reviewing the data given, we discovered that the data set provided by KTPH was not rich enough and hence not appropriate for meeting the expectations of ANLY482.  Survey data related to intervention was also not ready for analysis, making it difficult to analyze the data to improve intervention efforts. We also considered exploring Small Area Estimation, a relatively advanced statistical method for geospatial analysis, to further improve the geospatial segment of the current dashboard. However, the team has no experience in this area since we have not taken IS415 Geospatial Analytics for Business Intelligence. This makes it difficult to produce a satisfactory solution given the shortage of time. Hence, we have decided to take on a different project to apply our analysis in the remaining time given. As a team, we understand that team management needs to be improved and frequent communication highly essential to the project success, Hence, moving forth, our team has included regular and more frequent internal meetings to brainstorm and critique each other’s analysis and viewpoints. We will also communicate more frequently with the sponsor to update about our project progress and get their views.