Kimball Lifecycle Proposal Template

CIS 9440 - Data Warehousing for Analytics
Final Project Milestone 1
Group Number - <#6 – Change mind would like to do it alone please>
Student(s) - <Leandro Coimbra>

This Proposal is the beginning of your semester-long Final Project. The goal of the project is to develop a working Data Warehouse using a commercial database management system. Your project will use data from at least 2 sources, dimensionally model the data inside your Data Warehouse, and connect to a Business Intelligence application to produce valuable, actionable insights.

For motivation on project ideas, **think about interesting problems, opportunities, or insights that could be shown, solved, or highlighted with data**. Search for public datasets that interest you and your group. You may need to combine datasets to address your desired problem or opportunity. Below are just a few (of many) public data sources:

1. NYC Open Data
2. Google's Public Datasets

To complete this Milestone, please fill in all bolded sections below:

Data Warehouse Project Title:

New York Education and Budget Correlation

Motivation for Project idea:

The increase of debt after COVID and the low numbers of school evasion in the year of 2020. I work at a New York Private University and have daily contact with the decrease of search for higher education and wonder if it is across all the schools in the state.

Description of the issues or opportunities the project will address:

With the impact of the NY government increase debt and how with that, did it drive students to evade school? With less resources for students and the education budget reallocation due to the pandemic caused students to flee? The need to answer these questions is prominent and current and must be answer prompted in order to help students to get back to class and furthermore better the base of our community.

Business Justification:

<u>High-level Business Initiative:</u>

Determine the key factors causing school evasion and the correlation to debt. I look to Improve education retention in the state of New York, increase educational budget and spend it wiser.

<u>BI Sponsors and Stakeholders (who will own this project?)</u>

New York State Department of Education

<u>What's the Business Value?</u>

Find the major factor for student evasion and lower ir, decrease public debt, increase school resources therefore bringing students back to the classroom. Sound the alert to Governments so they may adjust their education budget with strategies to target the spending and at the same time attract students with a better learning opportunity in a batter facility with well paid staff.

<u>How long will this take? How much will this cost?</u>

The project will be finished as soon as possible with a 6-month target goal to gather information that is found in different New York state departments. The cost will be $500.00 USD and it will cost $50000 a year to maintain (– Donated student's time)

Technical Justification:

<u>Which data sources do we already have for this project?</u>

Dataset 1: Y2020 Local Law 16 Final Report DD.xlsx

Dataset 2: Data_Dictionary_COVIDSpendingbyDate.xlsx
Dataset 2: Data_Dictionary_COVIDSpendingbyDate.xlsx

Dataset 3: Demographic_Data_in_NYC_Public_Schools_-_Suppressed_(1).xlsx
Dataset 3: Demographic Data In NYC Schools Supressed DD.xlsx

Dataset 4: international_education
Dataset 5: schooldistrictunified_2015_5yr

Dataset 6: Country graduation percentage
Dataset 7: Education Index - Wikipedia

<u>What new data sources do we need (if any)?</u>

Dataset 1: Country graduation percentage (Retention)

Dataset 2: Education Index - Wikipedia

<u>Is the data we have conformed, consistent, and current? (data quality)</u>

On first look over the data seems cleaned and confined to the project. However, the data may present unexpected issues that will be dealt by the team if any arouse without increase of budget or deadline for the project.

<u>What technical skills will we need to complete this project?</u>

Knowledge and experience in the following code languages: SQL, python and Tableau

<u>Will we need any new types of technologies?</u>

The technology for the project already exists and will be implement.

Key Performance Indicators (KPI's) your Data Warehouse will display:

1. NY (New York) Revenue growth
2. NY Profit margin
3. Student retention rate
4. Student satisfaction
5. Education Instructional Costs
6. Administrative Costs Per Student
7. NY's Education Budgets
8. Student Financial Aid
9. Ny's Graduation Rates
10. Retention Rates: Understand and monitor the factors that affect students
11. Student Outcomes: Keep track of students after graduation
12. Passing Rates