

Kimball Lifecycle Template: **EXAMPLE**

CIS 9440 - Data Warehousing for Analytics

Week 3

Data Warehouse Project Title:

New York State Collegiate Education Data Warehouse

Description of the issues or opportunities the project will address:

New York State's school system produces many college students, but how many stay in New York for college? How many return to New York after college? How can New York State Legislators make informed decisions to keep more New York High School graduates in New York State for College?

To answer this question, the New York State Collegiate Education Data Warehouse will combine census, income, education, occupation, and recreational data to uncover insights regarding New York State students and residents.

Business Justification:

High-level Business Initiative:

Determine the key factors that keep New York State students in-state for College and the key factors that entice students to leave New York for college.

BI Sponsors and Stakeholders (who will own this project?)

New York State Higher Education Services Corporation (HESC)

What's the Business Value?

With clear factors for students' decision to stay or leave New York State for Higher Education, the Higher Education Services Corporation can recommend strategies to New York State Legislators to keep New York Staters in-state. They may also recommend strategies to entice students from other States to attend college in New York, which is important for New York's economy.

For example, if it's clear that students from Westchester are leaving New York for college at a much higher rate than other counties, maybe New York needs another university near Westchester. Or, maybe New York needs to better inform Westchester High School students about New York State colleges.

How long will this take? How much will this cost?

Currently, data to form these insights is scattered among many New York State agencies. Thus, to create an Enterprise Data Warehouse for this information will take a team of 6 data engineers 9-months. This project will cost approximately \$450,000 to stand up and \$100,000/year to maintain.

<https://www.cooladata.com/cost-of-building-a-data-warehouse/#:~:text=Assuming%20you%20want%20to%20build,can%20be%20around%20%24468K.%E2%80%9D>

Technical Justification:

Which data sources do we already have for this project?

Dataset 1: New York State College Students information

Dataset 2: New York State census data

Dataset 3: New York State income data

Dataset 4: New York State weather data

...

What new data sources do we need (if any)?

Dataset 1: Twitter sentiment datafeed

Dataset 2: College History of New York State residents (sample)

Dataset 3: Market research data by by State

Dataset 4: New York State Job Opportunities data

Is the data we have conformed, consistent, and current? (data quality)

The data is not consistent along dates. Some datasets measure by week, other's by day.

What technical skills will we need to complete this project?

1. Requirements gathering
2. Data Modeling
3. Dimensional Modeling
4. ETL Creation
5. BI Application Design and Implementation
6. Data Warehouse engineering
7. Standardized Report development

Will we need any new types of technologies?

1. ETL tool or custom development
 2. Cloud data storage
 3. Cloud data warehouse
-

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

1. College Students per 100 residents
2. New York State College Students per 100 residents
3. New York State College Students / College Students
4. Average Miles from students' High School to College
5. Average Degrees difference from High School to College climate in February
6. Average Family Income / College Student
7. Average GPA
8. YoY percent difference in New York State College Students / 100 residents