Project Proposal

## Group Members:

Tyler Sutherland (tsuther4@kennesaw.view.usg.edu),

Muhammad Sohel (msohel@kennesaw.view.usg.edu),

Rushitha Buddharaju (rbuddhar@kennesaw.view.usg.edu)

## What is the problem?

Investigating Workforce Metrics through HR Analytics

How can we use company data to analyze employee performance?

Can we identify patterns that contribute to high performance and employee retention?

Can we identify patterns and trends in employee turnover?

## Related Works

A related study by Hamoud and others focused on creating an HR data mart using a dataset of 484 retired employees of an Oil Company, with data on demographics, job details, and compensation. The data was imported into an SQL Server for cleaning and integration. A data mart is chosen over a data warehouse due to the small, departmental scope, with potential for future expansion and use. The implementation involves a bottom-up approach while focusing on efficient ETL processes and the method of Slowly Changing Dimensions (SCD). The final product is an OLAP cube that enables fast, multidimensional analysis to support HR decision-making for the company.

In the study of OLAP Analytical Solution for Human Resource Management by Debeljacki and others, they created a system for measuring and assessing HR management (HRM) performance to help in improving HR processes. They did this by enabling efficient tracking and analysis of HR data itself. They analyzed data such as employee motivation, productivity, compensation, training, and performance assessment. The model's architecture includes data sources, ETL processes, a data warehouse, and OLAP software for creating multidimensional reports and analyses. Data included employee attrition, compensation, recruitment, workforce development, and productivity, with KPIs like attrition rate, average salary, hiring rate, training costs, and overall productivity. The OLAP-based system provided ad-hoc reports, KPI monitoring, and a dashboard.

### Current and existing methods-

As previously talked about in our related works, we are seeing a fair bit of methods that use OLAP and ETL configurations to generate reports for us to look through and to develop answers to our questions.

## Our Project

As we proceed through this semester we will be creating a way to analyze and organize company data in order to generate performance metrics of how employees perform within a company structure. Our method of generating performance reports will be to implement the following:

**Methodology**:

Implement an OLAP (Online Analytical Processing) model for multi-dimensional analysis of HR data

Data Warehouse Construction:

* ETL Process: Extract, Transform, Load (ETL) the IBM dataset into a data warehouse.
* Dimension of the data warehouse:
  + Employee Demographics: Education, age, gender.
  + Job Roles: Titles, levels, status.
  + Departments: Names, sizes.
  + Performance Metrics: Environment Satisfaction, Job Involvement, Job Satisfaction, Performance Rating, Relationship Satisfaction, Work-Life Balance.
* OLAP model:
* Creating cubes to analyze turnover, performance trends, and compensation patterns.
* Data Analysis:
  + OLAP operations to perform multi-dimensional analysis and visualization of data
  + When it comes to our data, we will be using a couple different datasets that were used in our related works that we found during our research phase. The datasets are structured about the same with some distinction in some of the columns.The datasets include:

**Dataset:**

IBM HR Dataset from Kaggle, Includes 35 columns of employee data. HR Analytics Dataset from Kaggle, Includes 38 columns of employee data. Included in these columns are a couple of columns that we will be using for this project:

* Education: Below College, College, Bachelor, Master, Doctor
* Environment Satisfaction: Low, Medium, High, Very High
* Job Involvement: Low, Medium, High, Very High
* Job Satisfaction: Low, Medium, High, Very High
* Performance Rating: Low, Good, Excellent, Outstanding
* Relationship Satisfaction: Low, Medium, High, Very High
* Work-Life Balance: Bad, Good, Better, Best

Attached to this document are the two databases that we will be using for our project. We will be using the datasets to evaluate the results of our algorithm’s performance. As well as using works that use these datasets to analyze our evaluation results and comparing their outcomes to ours in order to get a concrete answer on how our system performs. These comparisons can be made using a few different metrics specified below.

**Evaluation:**

OLAP models can be evaluated through various metrics like accuracy, consistency, efficiency, and usability.

**Analysis:**

We will be checking out some of the other works that use these databases to analyze how our version does versus other versions out there. Here are the links to all the entries that use these databases.

IBM HR Analytics Employee Attrition & Performance: <https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset/code>

HR Analytics Dataset:

<https://www.kaggle.com/datasets/saadharoon27/hr-analytics-dataset/code>

## Citations

1. Hamoud, A. K., Abd Ulkareem, M., Hussain, H. N., Mohammed, Z. A., & Salih, G. M. (2020, May). Improve HR decision-making based on data mart and OLAP. In Journal of Physics: Conference Series (Vol. 1530, No. 1, p. 012058). IOP Publishing.
2. Debeljački, R., & Grljević, O. (2014). Olap Analytical Solution for Human Resource Management Performance Measurement And Evaluation: From Theoretical Concepts To Application.