**Central Information System – Zekelman School of IT**

**(Data Analytics for Business)**

**Dataset:**

The purpose of this dataset is to gather information related to the experiences, needs, and issues of students, and it was collected via an online survey. However, since the data is self-reported, there is a potential for response bias. The survey was divided into three parts: demographics, the convenience of living, and employment and scholastics. By analyzing this dataset, it is possible to gain valuable insights into the student experience, particularly in terms of their needs and issues.

However, it is important to be aware of potential biases and limitations in the survey method. Additionally, careful data cleaning and processing will be necessary to ensure the accuracy and usefulness of the data for analysis.

**Dataset assessment:**

Initial data is pushed into a Raw\_data table which acts as a landing zone for Survey data sheets, next the data is pushed into a Base\_Survey\_data table where column names are given along with correct Data types. The dataset is further broken down into seven different tables, each with a unique ID to identify and join them using Google big query.

* **"Academic"** table has columns for intake and current semester to track student progress.
* **"Demographics"** table has 10 columns for gathering demographic information to understand the student population.
* **"Accommodation"** table has three columns related to student housing situation for potential areas of improvement.
* **"Accommodation feedback"** table has five columns for scoring living and service quality.
* **"Academic feedback"** table has five columns for students to rate their course experience.
* **"Services"** table focuses on college services, gathering information on awareness, usage, and satisfaction for identifying areas of improvement.
* **“Feedback”** this table consists of data in which students can give their feedback related to their experience" means that there is a specific table within the dataset that is designed to gather feedback from students about their experiences. This feedback can then be analyzed and used to identify areas for improvement and to make data-driven decisions about how to improve the student experience.

**Data Collection and Pre-Processing – Python work**

Harshil and Jayaraj

**Ethical principles**

Rajvi

**Method and Solution**

Google Big Query will act as the Database (views) for the data to be stored, there will be connections between Google Drive (act as passage for data), Google Colab (for Python and ML work) and Tableau Visualization. Flow of data is single or both ways depending upon the use and phase in progress.

Diagram

Description automatically generated

A picture containing chart

Description automatically generated

**Future data needs and challenges**

Will discuss

**Code**

All

**References**

* **Tableau**: [Tableau Public](file:///E:\04%20-%20Semester4\01%20-%20capstone\Github\Capstone-Project\Assessment%201\Tableau%20Public), <https://www.tableau.com/learn>, [Datacamp](https://app.datacamp.com/learn/skill-tracks/tableau-fundamentals)
* **Python**: [Data science life hacks (linkedin.com)](file:///E:\04%20-%20Semester4\01%20-%20capstone\Github\Capstone-Project\Assessment%201\Data%20science%20life%20hacks%20(linkedin.com)), <https://keras.io>, <https://numpy.org>, [NLP - Datacamp](https://www.datacamp.com/tracks/natural-language-processing-in-python?)
* **Google Cloud:**
* **Data Ethics:** Content from semester 3, Ethics for analytics course – DAB 302
* **SQL:**