**Data Pipeline Documentation:**

**Overview**

The data pipeline outlined below is designed to collect, store, and analyze data from student surveys. The pipeline utilizes Microsoft Forms for survey creation, Power Automate for automation, Azure SQL Database for data storage, and Power BI for visualization.

**Components**

1. **Survey Creation with Microsoft Forms:**

* Surveys are created using Microsoft Forms to collect data from students.

1. **Automation with Power Automate:**

* Initially, the pipeline was set up to connect Microsoft Forms with Google Sheets using Power Automate to automatically update the Google Sheets with new responses.
* However, after consultation with Professor Manjari and further group discussion, the decision was made to utilize Microsoft's ecosystem entirely.

1. **Data Storage with Azure SQL Database:**

* A decision was made to use Azure SQL Database for data storage. This involved:
* Creating a new test survey in Microsoft Forms.
* Setting up an Azure SQL Database.
* Establishing a connection between the survey and the Azure SQL Database.
* Creating three separate database tables to store data related to demographics, accommodation, and academics.
* Configuring Power Automate to map survey variables to the appropriate database tables.

1. **Data Retrieval with PyODBC:**

* Data retrieval from the Azure SQL Database is accomplished using the PyODBC library in Python. This involves:
* Defining connection parameters.
* Establishing a connection to the Azure SQL Database.
* Executing SQL queries to load data into Python files for further processing.

1. **Visualization with Power BI:**

* The final step involves connecting the Azure SQL Database with Power BI. This will be achieved through automation to ensure that new responses or data updates in the database trigger automatic updates to graphs, charts, dashboards, and reports in Power BI.

**Workflow**

1. Students complete surveys created in Microsoft Forms.
2. Power Automate is utilized to automatically update an Azure SQL Database with new survey responses.
3. PyODBC is used to retrieve data from the Azure SQL Database into Python files for analysis.
4. Power BI is connected to the Azure SQL Database to visualize the data and generate reports.
5. Automation ensures that any new data is seamlessly integrated into the visualization process.

**Future Considerations**

* Continuous monitoring and optimization of the pipeline for performance and scalability.
* Regular backups and data integrity checks to ensure the reliability of the pipeline.
* Exploration of additional features and functionalities offered by Microsoft's ecosystem to enhance the pipeline's capabilities.

**Conclusion**

The data pipeline outlined above provides a comprehensive solution for collecting, storing, and analyzing student survey data. By leveraging Microsoft's suite of tools and services, the pipeline offers automation, scalability, and integration capabilities essential for efficient data management and decision-making processes.