# Final Project

## Data Programming – BDAT 1004​

## Group 2:

**Summary:**

This process involves reading a Data Set which was found on Kaggle and was stored on a local computer and then imported into a Pandas DataFrame and connecting to an Azure SQL Database using a connection string. The dataset is related to book reviews and includes information such as book names, authors, average star ratings, total ratings, reviews, and distribution of ratings.

The CSV file was read from the local computer. The SQL Database contains a "Books" table, and the process demonstrates loading the first 1000 rows of data into this table.

The subsequent steps include connecting to the Azure SQL Database through a connection string and utilizing a Pandas Data Frame for various computations and aggregations. We choose these Five specific queries that are performed and visualized in Power BI based on this processed data.

The five Queries that were chosen are as follows:

a) Find the books with the highest average star rating.

b) Calculate the average star rating for all books.

c) Calculate the percentage of 5-star ratings for each book:

d) Filter books with an average rating greater than 4.0

e) Sort the Dataframe by the number of reviews in descending order

In conclusion, this comprehensive process seamlessly merges local CSV file data into a Pandas DataFrame and establishes a connection to an Azure SQL Database, housing a "Books" table. The dataset's informative attributes, spanning book details, ratings, and reviews, offer rich insights. By employing strategic Power BI visualizations driven by five specific queries, this process aptly transforms and illuminates the book review data, harmonizing usage of Pandas, SQL Database, and Power BI tools for meaningful analysis.