**PROJECT SUMMARY**

Project Proposal (Group 6 )  
The objective of this ETL project  is  to extract two unrelated data set from [www.kaggle.com](http://www.kaggle.com/)  for selected colleges in the US states; while focusing on the “tuition”, “projected salaries after college”, and  any other “values records”, we plan to transform data by removing irrelevant, inaccurate, and duplicate  records, validate data for accuracy,  present  the data in preferred grouping and merging two data sets into a single data. The merged data will be loaded into a SQL database.

**BRANDON**

Extraction - Source Data (Python/Panda)  
<https://www.kaggle.com/wsj/college-salaries#salaries-by-college-type.csv>  
<https://www.kaggle.com/theriley106/university-statistics#schoolInfo.json>Transform

**Timothy Ayoade**

(Python/Panda or SQL)  
•    Remove incomplete and inaccurate records  
•    Remove duplicates  
•    Data standardization (Uniform spacing, All upper case and proper indentation)  
•    Aggregate data as required (present data in the preferred grouping)  
•    Merge the data together.

**Mohamed Abdi**Load (SQL)  
•    Create SQL Database  
•    Load the Merged data into SQL

**Timothy Ayoade**

**Post - Load (SQL)**

* Read the Data loaded in SQL from Python
* Created and engine to connect to PGAdmin database
* And read the table from the database.