**Project Report (50 points)**

**The report describes the original data sources and how the data were formatted. (10 points)**

* Position Info
  1. Data was collected from an excel file created by operations management where all the positions and hourly wages are tracked
* Associates
  1. Data was extracted from ADP in csv files
* SKU Master
  1. Source: Data was extracted from an actual pricing model of a manufacturing company in an excel file. Some data was changed to protect the company.
* Volume
  1. Source: Data was extracted from SAP in csv files collecting the production volume of a manufacturing plant for the last 2 years. Some data was changed to protect the company.
  2. Formatting: Columns names were changed using pandas in python to the standard names used in other tables within the final data based
* Payroll
  1. Source: The data was extracted from ADP as csv files. One csv file was received from ADP per day. The project covers a 3-month time frame of labor data
  2. Formatting: Using VBA, the data collected from ADP was cleaned and formatted in Using Pandas, the data was cleaned and the unnecessary columns were eliminated

**The report explains what data cleaning or transformation was performed and why it was needed. (20 points)**

* SKU Master
  1. Using pandas, the data was cleaned eliminating unnecessary columns and standardizing column names in order to match with column names used in other tables in athe data base
* Volume
  1. Data was changed using a Unique Identifying Key to replace Item ID. Columns names were changed using pandas in python to the standard names used in other tables within the final data based.
* Payroll
  1. Using the VBA code to properly structure the payroll data, the Payroll data is structured into a CSV with the following columns: *Date, Name, ID, Account, Pay Code, Money, Hours, Days, Wages*.
  2. These files are saved in a folder, where PowerQuery was used to query all the CSV files into an Excel file. All associates are assigned a Unique Identifying Key in a new column, and wage data was adjusted. The Query left the columns: *Date, Pay Code, Hours, Wages, Associate UIK.* This was converted to a CSV file.

**The report describes the structure of the final database along with any improvements that could be made in the future. (20 points)**

Structure of the relational SQL database

* Tables
  1. Position Info
     1. Associate UIK
     2. Shift
     3. Position
     4. Base wage
     5. Incentive
     6. Base OT
     7. Incentive OT
  2. Associates *(Delete Account)*
     1. Associate UIK
     2. Position
  3. SKU Master
     1. SKU ID
     2. Run Rate
     3. Hourly Run Rate
     4. Case Cost
  4. Volume
     1. SKU ID
     2. Production Volume
     3. Date
  5. Payroll
     1. Date
     2. Pay Code
     3. Hours
     4. Wages
     5. Associate UIK