SBA Sector Data Cleaning

Project Description/Problem

I downloaded a excel file of SBA sectors from website and found the data quite messy. I used Python's pandas to clean the data.

Solution

To address this problem, the following steps were undertaken among many others:

1. Dataset: Downloaded from web.

2. Data Import and Transformation: Imported the data from an Excel file into a pandas data frame.

3. Data Format Analysis: Reviewed the provided data.

4. Meta Data and EDA: Checked data types, identified null values and duplicates, and removed records with mostly null values.

5. Cleaned the Data: Used pandas attributes and methods to clean data.

6. Save the Cleaned Data: Saved the clean data into a CSV file.

Files

1. **sba-table-of-size-standards_effective-august-19-2019_v0.xlsx**: Raw file.
2. **SBA Sector Data Cleaning.ipynb**: Python cleaning script.
3. **SBA Sectors Cleaned.csv**: Cleaned CSV file generated by the Python script as a form of output.
4. **Project Description.pdf**: Documenting the approach to come a solution.