Capstone Project 1 - Data Wrangling

Steps undertaken to clean capstone project data set:

Register for api.football-data.org API:

I created an account to receive an API key needed to collect data for my data set.

Import Python Modules:

Pandas, requests, http.client, and json were imported into ipython notebook.

Collect market values:

Market values were copied from TransferMarkt.com and pasted into an excel file with five sheet, each for one the leagues. Used pandas to read each table and append it to a data frame. Manually made some changes to club names so to work with the names from api.football-data.org. Lastly reordered, and rename columns in the data frame to look better.

From api.football-data.org using HTTPConnection:

League Date Frame

Removed characters from the name string to remove year information and leave only club names. Included a column to show the country of the league. Reordered, and rename columns in the data frame to look better.

Team Data Frame

Used a-for loop to extract data for each team and append it to a teams\_df. Added a column to include the league\_id for each team. Reordered, and rename columns in the data frame to look better.

Team Standing Data Frame

Retrieved performance statistics for each team including venue data (home vs away games)

Player Data Frame

Used a for-loop to add every player data from every team in all five league into a data frame. Had to divide the players into two data frame which I merged after requesting the data. Reordered, and rename columns in the data frame to look better.

With all of the data imported, I created five table each with a key to connect them. Some of the data columns types where changed to integer types to perform calculations later.

Data Tables:









