Project Title: Comparative Shopping of Hardgood Tennis Equipment

Group Members: Jenette Smith, Kris Wasemiller, Rachel Podemski, Tiffany Burns

Project Description/Outline

We are going to scrape data from the vendor website, Wilson, along with two specialty stores, Dicks Sporting Goods and Academy Sports. We also found a dataset that has current pricing from Walmart that we are using as our base items to lookup. From there we will import this data into our SQL tables through Jupyter Notebook. ERD Diagram listed below.

Research Question to Answer

Is it better to buy from the vendor directly, or from a retailer?

Is buying from a specialty store a better bargain, or a big box retailer?

What type of availability is there from the vendor, big box, retailer?

Datasets to be used:

Kaggle Walmart Product Data CSV:
https://www.kaggle.com/promptcloud/walmart-product-data-2019

Web Scrape:

Dicks Sporting Goods

Academy Sports

Wilson

Rough Breakdown of Tasks:

1. Preprocessing data: Completed by all
2. Extract: Tiffany, Rachel

3. Transform/Load: Kris, Jenette

4. Final Write Up Editing: Jenette, Rachel

Everyone type up their portion of paper, and submit by Sunday.

ERD Diagram:

```
# Modify this code to update the DB schema
# To reset the sample schema, replace every
# two dots ('..' - without quotes).


product_price
-
id PK int
product_name varchar(100)
category_id int FK >- product_category.id
store_id int FK >- store.id
sale_price int
list_price int
scraped_date int
on_sale boolean
url varchar(100)
last_updated datetime

product_category
-
id PK int
category_name varchar(100)
last_updated datetime

store
----
id PK int
store_name varchar(100)
store_url varchar(100)
availability boolean
last_updated datetime
```