

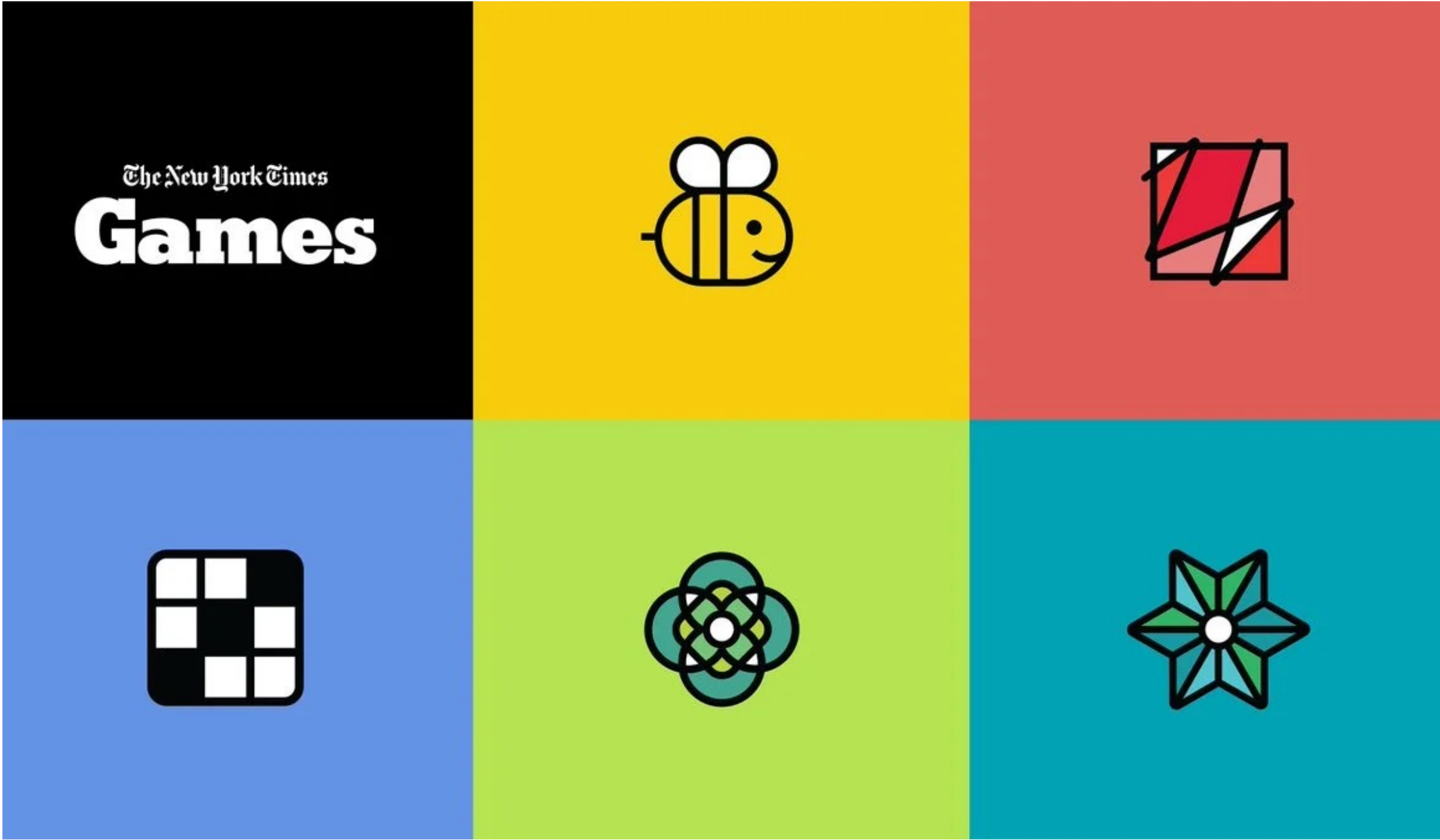
EPPS 6354 - INFORMATION MANAGEMENT

HOMEWORK 1

Olivia Wilson-Pietrzak

QUESTION 1

NAME AND DESCRIBE THREE APPLICATIONS YOU HAVE USED THAT EMPLOYED A DATABASE SYSTEM TO STORE AND ACCESS PERSISTENT DATA.



PEACOCK (NBC STREAMING)

- Media catalog, individual watching habits, geolocation information, purchase information, engagement information, ad information

NYT GAMES

- Game catalog, player statistics, purchase information

LIBBY (AUDIOBOOKS)

- Media catalog, member institution list, member institution media catalog, individual borrowing records, bookmarks

QUESTION 2

PROPOSE THREE APPLICATIONS IN DOMAIN PROJECTS . BE SURE YOU INCLUDE:

I. PURPOSE

II. FUNCTIONS

III. SIMPLE INTERFACE DESIGN

R CODE FOR SDAR CLASSES

- **Purpose:** As SDAR does not have a requirement for coding skills, it could be useful to have a designated resource for students to learn about specific R packages and commands used in a specific class. It could be built out to be as complex as the instructor would want (e.g., listing only the packages needed for a class vs. providing the mathematical logic behind a command or package)
- **Functions:** provide a reliable resource for students to source packages/ commands from; allows instructors to provide rationale for preference for specific packages; allows for “on-demand” assistance in understanding logic behind certain packages.

SDAR CLASS CODE → SECTION/PROFESSOR → PACKAGE LIST → COMMANDS →

- LOGIC OF COMMANDS
- WHY THE PROF PREFERS
- CORRELATED READINGS

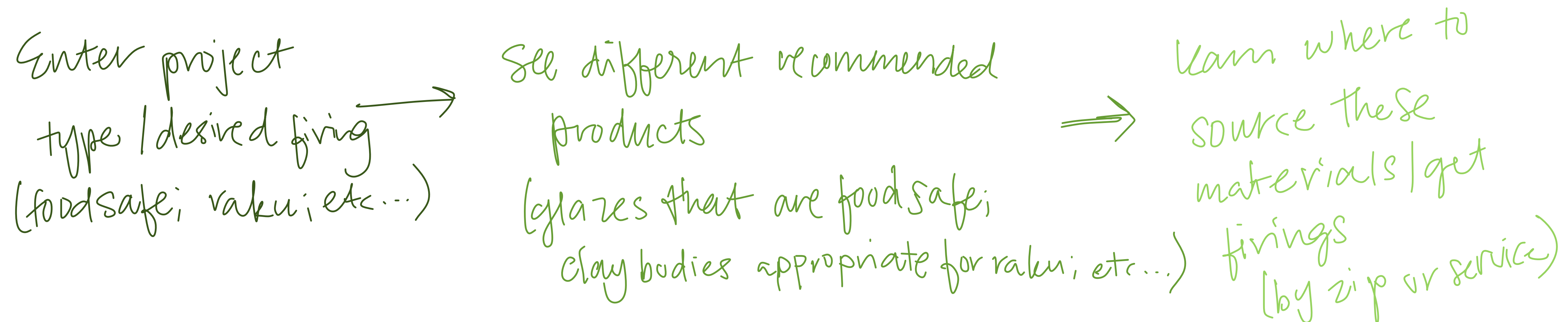
EDUCATION DATASETS

- **Purpose:** I am interested in studying education, particularly from a labor economics viewpoint. It would be helpful to have a resource with links to good data sources, along with accompanying metadata that would allow you to check if the data source has variables you are interested in working with.
- **Functions:** provide a reliable resource for quality datasets for students; department could review and approve; incorporate resources from TSP.

Topic / Research Question → age/type of dataset → select dataset that works best for project
(use metadata or group by the kind of question that the data could answer)
(survey, admin, etc...)

FUN ANSWER: CLAY DATABASE

- **Purpose:** I like to make pottery, and there are a lot of different choices to make regarding clay bodies and finishes. Not all clay bodies or glazes are appropriate for all purposes, and not all clay shops carry the same products, so it could be cool to create a database that tracked all of the offerings for clay retailers in a certain area, and what products are best/safest for certain applications.
- **Function:** to provide a complete resource for potters in an area so they can understand what clay bodies, finishes, and firing resources are available. Also allows potters to understand what finishes are safe for different applications.



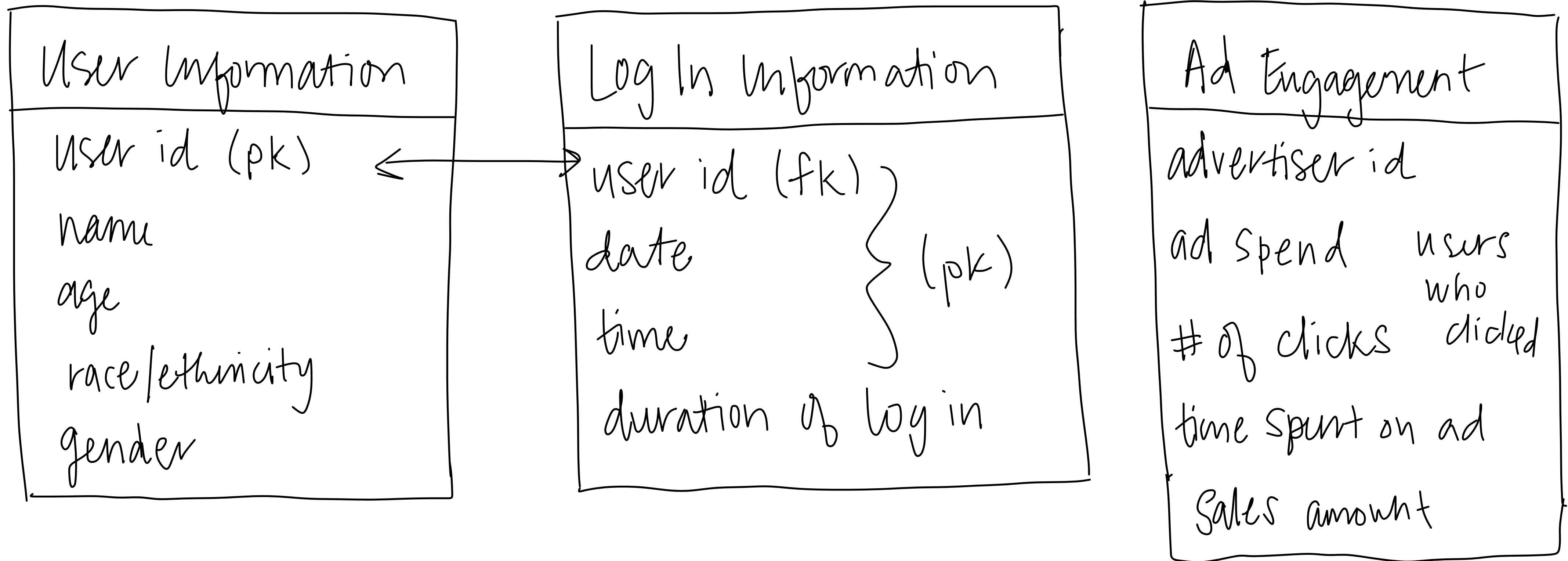
QUESTION 3

IF DATA CAN BE RETRIEVED EFFICIENTLY AND EFFECTIVELY, WHY DATA MINING IS NEEDED?

- Data mining is still needed because it allows us to understand what the data means.
 - If we just extract data without looking for patterns or doing any analysis, the data is not valuable.
 - We need to be able to critically think about the data, and use different statistical methods to analyze what is happening.
 - The analysis allows us to draw an actionable insight faster than the list of individuals. This is the benefit of knowledge mining.
- Ex: let's pretend we are candy makers, and we want to understand what product we should develop a new flavor for. We have a sample of 1000 people's preference for candy: do they like Swedish Fish or Snickers bars more? It is of limited helpfulness to list out each individual and whether they like Swedish Fish or Snickers more. It's far more helpful to look at our data, analyze it , and say "60% of people like Swedish Fish" or "Swedish Fish are more popular than Snickers in the sample".

QUESTION 6

DESCRIBE AT LEAST THREE TABLES THAT MIGHT BE USED TO STORE INFORMATION IN A SOCIAL-NETWORK/SOCIAL MEDIA SYSTEM SUCH AS TWITTER OR REDDIT.



★ An example of 3 tables that could be used in a social media company → user info & login info are closely related, but I thought it would be interesting to examine a more revenue oriented table (Ad Engagement) as well! ★