Structured vs Unstructured Data

Can be organized into a table format?

**Structured Data**

- Yes

**Unstructured Data**

- No

### Different Types of Variables

**Variable Types**

- **Categorical**
  - non-ordered (nominal)
  - ordered (ordinal)

- **Quantitative (numeric)**
  - continuous
  - discrete

**Source:** Bank Transaction Data on Kaggle (https://www.kaggle.com/apoorvwatsky/bank-transaction-data)

**Source:** Chibird.com
Practice: Identify the type of data

<table>
<thead>
<tr>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TRANSACTION DETAILS</td>
<td>VALUE DATE</td>
<td>WITHDRAWAL AMT</td>
<td>DEPOSIT AMT</td>
</tr>
<tr>
<td>2</td>
<td>TRF FROM Indiaforensic SERVICES</td>
<td>29-Jun-17</td>
<td>1,000,000.00</td>
<td>1,000,000.00</td>
</tr>
<tr>
<td>3</td>
<td>TRF FROM Indiaforensic SERVICES</td>
<td>5-Jul-17</td>
<td>1,000,000.00</td>
<td>2,000,000.00</td>
</tr>
<tr>
<td>4</td>
<td>FDR/INTERNAL FUND TRANSFE</td>
<td>18-Jul-17</td>
<td>500,000.00</td>
<td>2,500,000.00</td>
</tr>
<tr>
<td>5</td>
<td>TRF FRM Indiaforensic SERVICES</td>
<td>1-Aug-17</td>
<td>3,000,000.00</td>
<td>5,500,000.00</td>
</tr>
<tr>
<td>6</td>
<td>FDR/INTERNAL FUND TRANSFE</td>
<td>16-Aug-17</td>
<td>500,000.00</td>
<td>6,000,000.00</td>
</tr>
<tr>
<td>7</td>
<td>FDR/INTERNAL FUND TRANSFE</td>
<td>16-Aug-17</td>
<td>500,000.00</td>
<td>6,500,000.00</td>
</tr>
<tr>
<td>8</td>
<td>FDR/INTERNAL FUND TRANSFE</td>
<td>16-Aug-17</td>
<td>500,000.00</td>
<td>7,000,000.00</td>
</tr>
<tr>
<td>9</td>
<td>FDR/INTERNAL FUND TRANSFE</td>
<td>16-Aug-17</td>
<td>500,000.00</td>
<td>7,500,000.00</td>
</tr>
<tr>
<td>10</td>
<td>FDR/INTERNAL FUND TRANSFE</td>
<td>16-Aug-17</td>
<td>500,000.00</td>
<td>8,000,000.00</td>
</tr>
<tr>
<td>11</td>
<td>FDR/INTERNAL FUND TRANSFE</td>
<td>16-Aug-17</td>
<td>500,000.00</td>
<td>8,500,000.00</td>
</tr>
</tbody>
</table>

{  
  "student": [  
    {  
      "id": 12345,  
      "name": "Harry Potter",  
      "age": 11  
    },  
    {  
      "id": 12346,  
      "name": "Ron Weasley",  
      "age": 11  
    }  
  ]}  

Practice: Identify the type of the variable of interest.

20 runners run a mile as fast as they can. Their times are recorded.  
Non-ordered Categorical, Ordered Categorical, Discrete, or Continuous (circle one)

50 Students are asked what their major is.  
Non-ordered Categorical, Ordered Categorical, Discrete, or Continuous (circle one)

100 Married Couples are asked how many children they have.  
Non-ordered Categorical, Ordered Categorical, Discrete, or Continuous (circle one)

20 runners are asked to run a mile as fast as they can. Next to each runner’s name, the coach records “yes” or “no” to indicate whether or not they broke the 5-minute mark.  
Non-ordered Categorical, Ordered Categorical, Discrete, or Continuous (circle one)

Judges score musicians across a number of different criteria using four choices: “superior,” “excellent,” “good,” or “needs work.”  
Non-ordered Categorical, Ordered Categorical, Discrete, or Continuous (circle one)

To-do:  
- Finish Lab 1, commit and push the lab using git commands!  
- Join Piazza if you haven’t done so!  
- Log in to PrairieLearn using your NetID login and add “STAT/CS/IS 107” to your course list.