

Assignment 2

Part 1:

```
import pandas as pd

# Load the dataset
df = pd.read_csv('path_to_the_downloaded_dataset.csv')

# Selecting relevant features along with CustomerID
selected_columns = ['CustomerID', 'City', 'StateProvinceName', 'AvgMonthSpend',
'BikeBuyer']

# Creating a new DataFrame with selected attributes
df_selected = df[selected_columns]

# Display the first few rows of the new DataFrame
df_selected.head()
```

Nominal: Categories without order (e.g., city names).

Ordinal: Categories with order but no consistent differences (e.g., rankings).

Interval: Ordered with consistent differences, but no true zero (e.g., temperature).

Ratio: Ordered with consistent differences and a true zero (e.g., income).

Discrete: Countable, often whole numbers (e.g., number of students).

Continuous: Measurable, with potentially infinite values (e.g., height).

Selected Attributes:

1. **CustomerID:** A unique identifier for each customer.
 - **Data Structure:** Integer
 - **Measurement Level:** Nominal (It is a label, so the values don't have a meaningful order)
2. **City:** The city where the customer resides.
 - **Data Structure:** String
 - **Measurement Level:** Nominal (Different cities are labels with no intrinsic order)
3. **StateProvinceName:** The state or province where the customer resides.
 - **Data Structure:** String
 - **Measurement Level:** Nominal (Like cities, states/provinces are also labels without a specific order)
4. **AvgMonthSpend:** The average monthly spending by the customer.
 - **Data Structure:** Float or Integer (depending on how it's stored)

- **Measurement Level:** Ratio (It has a true zero point, and the differences between values are meaningful)
- 5. **BikeBuyer:** Whether the customer has bought a bike (1 = Yes, 0 = No).
 - **Data Structure:** Integer (binary encoded)
 - **Measurement Level:** Nominal (It's a binary label indicating a category with no order)