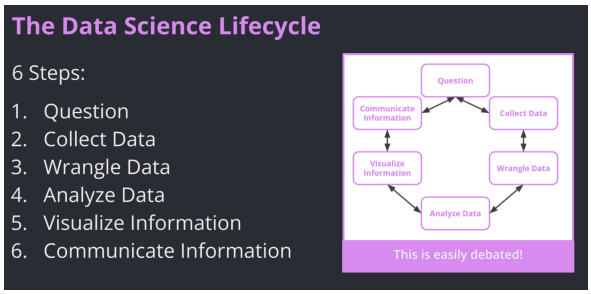
8/21/23 notes

**What Is Data Science?**

- Using advanced analytics to extract and interpret data for business
- Is used in almost all areas of business




---

9/5/23 notes

**Python Fundamentals**

- Datasets - collection of data
  - List - **ordered** and **changeable** with duplicates **allowed**
  - Dictionary - **ordered** and **changeable** with duplicates **not allowed**
  - Set - **unordered** and **unchangeable** with duplicates **not allowed**
  - Tuple - **unordered** and **unchangeable** with duplicates **allowed**
- Representing the data
  - Column-oriented - grouping my features, or column
    - Each column has values associated with the first row of that column
  - Row-oriented - grouping by observation, or row
    - Each row has the values associated with the first column of that row
- Indexing
  - List - **name[index]**
    - Index must be whole number, starts at 0 and counts up by 1
  - Dictionary **name[key]**
    - Keys can be any valad data type within used language, keys must be unique within dictionary
  - Set - **for *value* in *set***
  - Tuple - **name[index]**
- Iteration
  - While loop
    - Runs as long as condition is true
  - For loop
    - Runs through all values in a collection

- Useful functions
    - Dictionaries:
        - values()
        - items()
        - keys()
    - Lists:
        - len()
        - append()
        - sort()
    - Other:
        - range()
        - print()
        - split()
        - type()
        - int()
        - str()

---

9/5/23

**Central Tendancy**

- An attempt to use statistical measures to describe the behavbior of the collection of data
    - Mean
        - Takes the sum of all data points and divides by the number of datapoints
        - “Expected” values for data
        - Best for symmetrical data with a normal distribution
        - Can be misleading if there are outliers
    - Median
        - The middle value of the data when arranged smallest to largest
        - Works for all distributions of data, resistant to outliers
    - Mode
        - The value that shows up the most in a set of data
        - Multimodal data - Data with more than one significant modes
- Skewed data
    - Result of outliers - skews the way of the outlier(right or left)
    - Median and mode dont really change, but mean is pulled the way of the outlier