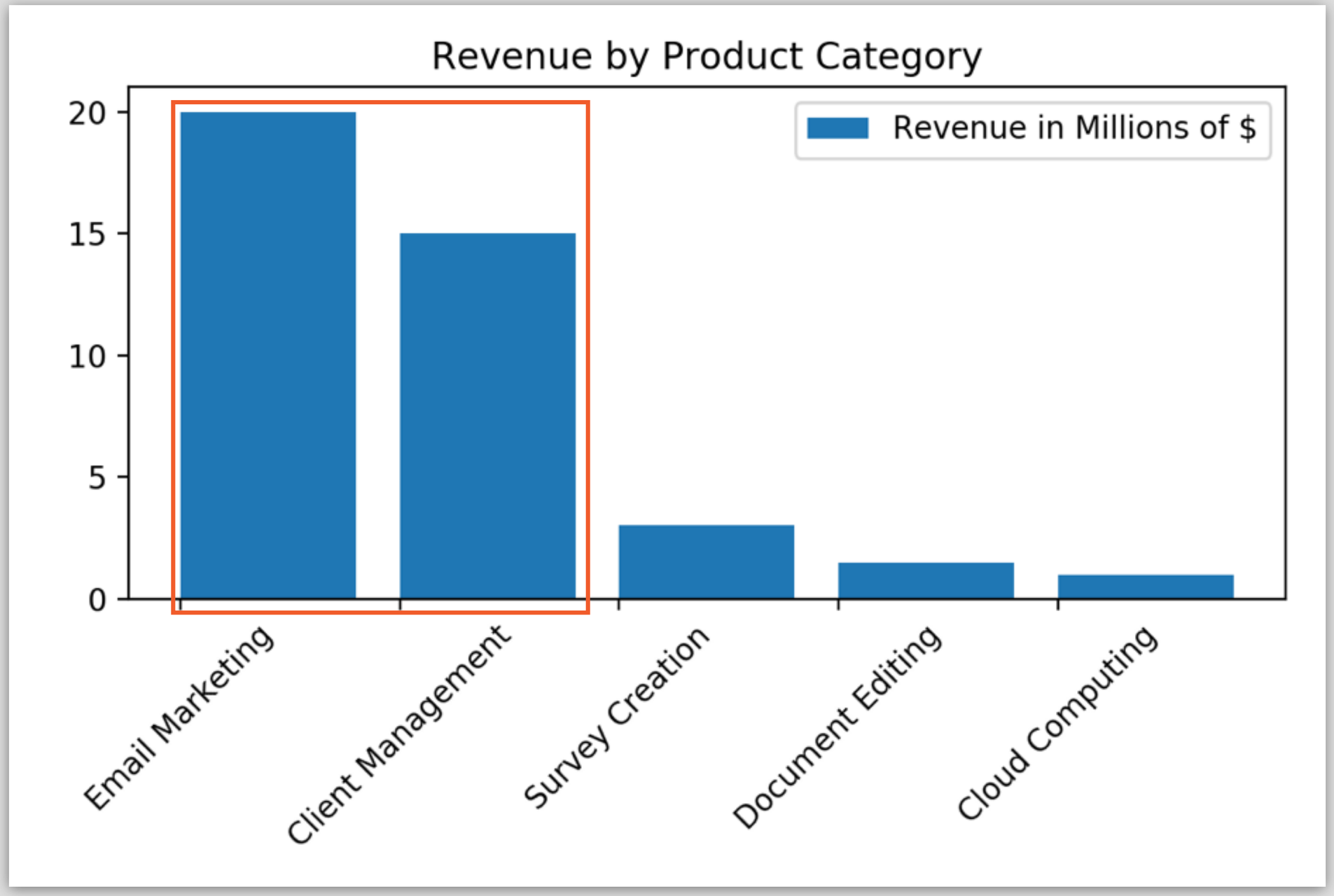
# Comparing Data with Bar Graphs

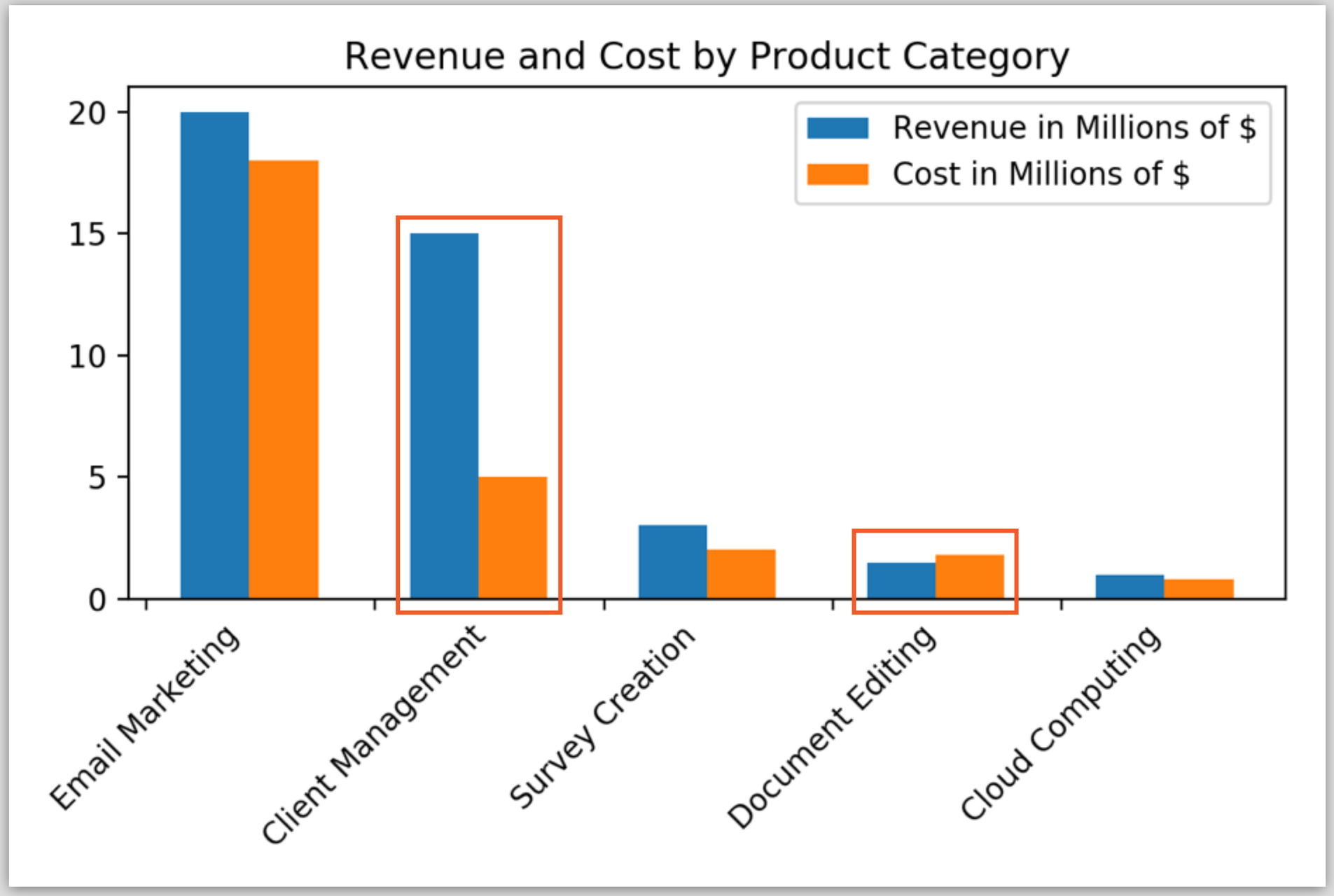
**YK Sugishita**
SOFTWARE DEVELOPER / DATA SCIENTIST
www.csdojo.io

# Bar Graphs

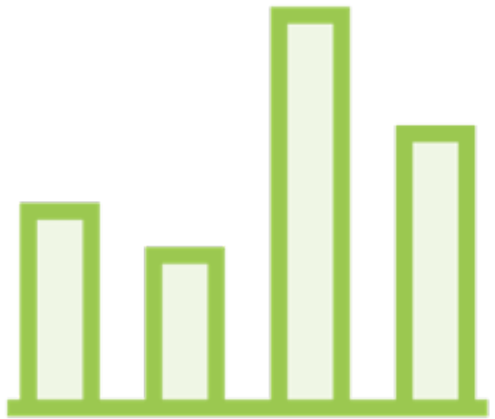
A convenient way to compare numeric values of several groups

# When to Use Bar Graphs


Revenue by Product Category
Revenue in Millions of $
20
15
10
5
0
Email Marketing
Client Management
Survey Creation
Document Editing
Cloud Computing


Revenue and Cost by Product Category
Revenue in Millions of $
Cost in Millions of $
20
15
10
5
0
Email Marketing
Client Management
Survey Creation
Document Editing
Cloud Computing

# Why Bar Graphs



**Helps you compare numeric values**

**Good for comparing multiple values**

**Sometimes gives you a new insight**

# Creating Bar Graphs with Matplotlib

# Problem: Compare the Populations of the 10 Most Populous Countries in 2007

# countries.csv

| country | continent | year | lifeExpectancy | population | gdpPerCapita |
|---|---|---|---|---|---|
| Afghanistan | Asia | 1952 | 28.801 | 8425333 | 779.4453145 |
| Afghanistan | Asia | 1957 | 30.332 | 9240934 | 820.8530296 |
| Afghanistan | Asia | 1962 | 31.997 | 10267083 | 853.10071 |
| | | | ⋮ | | |
| Zimbabwe | Africa | 1997 | 46.809 | 11404948 | 792.4499603 |
| Zimbabwe | Africa | 2002 | 39.989 | 11926563 | 672.0386227 |
| Zimbabwe | Africa | 2007 | 43.487 | 12311143 | 469.7092981 |

# Demo

**Creating Bar Graphs with Matplotlib**

- Sort a DataFrame with sort_values()
- Create bar graphs with plt.bar()

# Practice Problem 4 - Bar Graphs

# Problem: Compare the GDP of the 10 Most Populous Countries in 2007

# countries.csv

| country | continent | year | lifeExpectancy | population | gdpPerCapita |
|---|---|---|---|---|---|
| Afghanistan | Asia | 1952 | 28.801 | 8425333 | 779.4453145 |
| Afghanistan | Asia | 1957 | 30.332 | 9240934 | 820.8530296 |
| Afghanistan | Asia | 1962 | 31.997 | 10267083 | 853.10071 |
| ⋮ | | | | | |
| Zimbabwe | Africa | 1997 | 46.809 | 11404948 | 792.4499603 |
| Zimbabwe | Africa | 2002 | 39.989 | 11926563 | 672.0386227 |
| Zimbabwe | Africa | 2007 | 43.487 | 12311143 | 469.7092981 |

# Compare the GDP of the 10 Most Populous Countries in 2007

- Find GDP with data.gdpPerCapita * data.population
- Use subplot() to show multiple plots on the same graph (review the histogram module)
- Pause video

# Demo

## Example solution