Assignment #6 – Outliers

To begin, open Jupyter Notebook through your preferred environment such as Google Colab or Anaconda. Once opened, create a new notebook. In Google Colab this is done by selecting **File** then **New Notebook**. Next, carefully write the code provided into each cell of the notebook as shown, including any comments.

Execute each cell by clicking the **Run or Play** button located on the left side of each cell. Should you have an error during execution, review the code within the cell. Pay close attention to potential typos, incorrect spacing, or misspellings, verifying each line against the provided code. Correct and rerun the cell.

Continue this process for all cells. Then save as a pdf by selecting **File** then **Print**.

Submit PDF to Canvas. Code and output must be clearly identified for full credit.

50 points

Intentional Blank Space

hw6_datacleanup_of_outliers_datawrangling.ipynb

```python
# import python packages
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

```python
# Example dataset with outliers
data_with_outliers = np.array([1, 2, 3, 4, 5, 6, 7, 100, 200, 250, 1000])

# Define a threshold for outlier removal
threshold = 100

# Remove outliers
cleaned_data = data_with_outliers[data_with_outliers < threshold]
```

```python
# Visualize data before and after outlier removal
plt.figure(figsize=(10, 5))

# Before outlier removal
plt.subplot(1, 2, 1)
sns.boxplot(y=data_with_outliers)
plt.title('Before Removing Outliers')

# After outlier removal
plt.subplot(1, 2, 2)
sns.boxplot(y=cleaned_data)
plt.title('After Removing Outliers')

plt.tight_layout()
plt.show()
```

```python
print("Original data:", data_with_outliers)
print("Cleaned data:", cleaned_data)
```