Python Data Dive Activities Unit 2.3

1. Visualizing Iris Dataset: Analyze the famous Iris dataset to understand flower species classification.
   a. Create boxplots and violin plots to explore data distribution for each petal and sepal feature.
   b. Use pairplots to analyze relationships between features and identify potential clustering.
   c. Build a heatmap to visualize correlation between all features.

   Dataset: archive.ics.uci.edu/dataset/53/iris

2. Exploring Tips Dataset: Analyze restaurant tipping data to discover patterns and trends.

   a. Create a distribution plot of tip amounts and compare across bill amounts.

   b. Construct a bar chart to see how day of the week affects tip percentage.

   c. Plot a scatter plot with hue based on smoker/non-smoker to investigate any tipping differences.

   Dataset: kaggle.com/code/sanjanabasu/tips-dataset

3. Movie Ratings Analysis: Examine movie ratings data to understand user preferences and trends.

   a. Build a joint distribution plot for user ratings and movie genres.

   b. Use a boxplot to compare average ratings for different directors.

   c. Create a violin plot to analyze rating distribution across decades.

   Dataset: grouplens.org/datasets/movielens/100k/

4. Analyzing Flight Delays: Explore flight delay data to identify factors contributing to late arrivals.
   a. Create a heatmap to visualize delay distribution across airlines and origin/destination.
   b. Use a bar chart to analyze the impact of day of the week and weather conditions on delays.
   c. Build a boxplot to compare average delay durations for different types of flights (e.g., domestic, international).

   Dataset: kaggle.com/datasets/giovamata/airlinedelaycauses

5. Clustering Customer Segments: Apply KMeans clustering to identify distinct customer groups based on purchase behavior.
   - Activities:
     - Use pairplots to explore relationships between purchase metrics (e.g., amount, frequency).
     - Perform KMeans clustering based on selected features and visualize clusters using a scatter plot.
     - Analyze cluster characteristics by comparing average values of purchase metrics across groups.

   Dataset: kaggle.com/datasets/prajitdatta/movielens-100k-dataset