UDC 004

Mykola Turchyn, Master's Student,
Olena Chyzhmotria, Senior Lecturer,
Iryna Dmytrenko, Senior Lecturer
Zhytomyr Polytechnic State University

SEABORN AS A TOOL FOR EFFECTIVE WORK WITH NUMERICAL DATA

Effective data visualization is indispensable in modern analysis, helping researchers confirm theories and communicate ideas. When working with numerical data, such as age or income, graphical representation makes it easy to detect trends and relationships. Seaborn, a powerful library based on Matplotlib, offers excellent tools for this purpose. Due to its system-oriented implementation, it appeals to users as it works seamlessly with Pandas DataFrames and offers high-level functions for better visualization readability [1].

Numerical Data

Entities such as age, income, or sales can be considered as numerical data. Graphical representation of numbers makes it easy to detect trends and relations within data sets. Understanding how frequently, uniformly, or how many modes a data set contains in its frequency distribution can be done through the aid of histograms. In this way, they enable analysts to see at a glance how counts or measurements smear overall and so are significant for defining of the attributes of the data [2]. An example is shown in Fig. 1.

Scatter plots are another key visualization tool for numerical data, allowing analysts to explore relationships between two variables. By plotting numerical values on both axes, scatter plots highlight correlations, clusters, and trends by plotting numerical values on both axes enabling insights into linear or non-linear relationships. This visualization is particularly valuable for identifying whether two numerical variables influence one another.

However, *line plots* are commonly deployed to visualize changes over time. They are especially helpful in the performance of time series studies, for instance, in analyzing sales, variations in temperature or the stock market. Many data points are connected with lines and thus have definite trends and fluctuations, making them suitable for temporal analysis and forecasting.

Pair plots, are a useful tool given by Seaborn, as they allow the user to visualize multiple scatter plots of numerical relationships within groups at once. Pair plots help the user visualize scatter plots and histograms of all variable combinations, allowing the user to more thoroughly investigate pairwise relationships in an effort to locate correlations, clusters, or outliers in massive data sets.

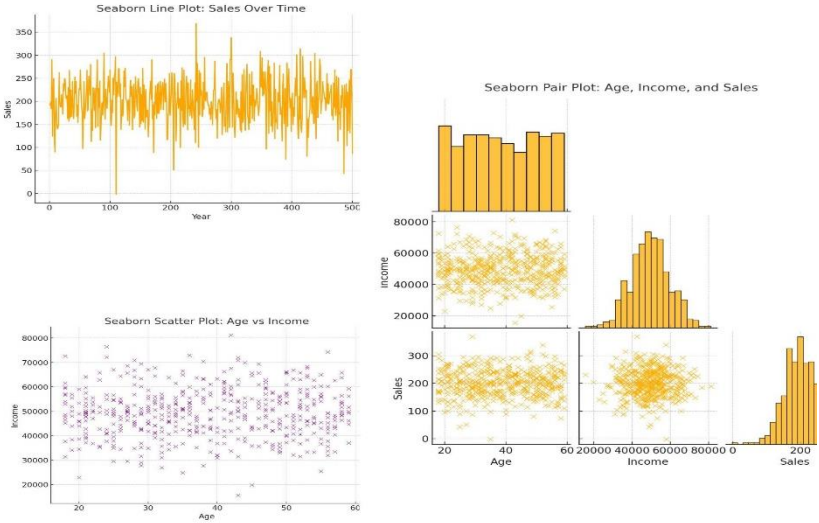

Fig. 1 – Example of visualization of numerical data

Summary

Seaborn can be useful for visualizing and analyzing numerical data. It features friendly syntax, integrates well with Pandas, and has good high-level plotting functions, making things easier when creating informative visualizations. If you are interested in studying data relations or trends, Seaborn provides the tools, such as histograms, scatter plots, and line plots, to expose and present interesting findings. By learning the essential functions for numerical data, you will be able to convert unorganized data into clear and compelling messages that aid in making decisions and guiding narratives.

References:

1. Wes M. Python for Data Analysis Data Wrangling with pandas, NumPy, and Jupyter / McKinney Wes., 2022. – 561 c.

2. Rishabh S. Seaborn: A comprehensive guide to statistical data visualization in Python [Information resource] / Singh Rishabh. - 2024. - Resource access mode: https://medium.com/@RobuRishabh/seaborn-a-comprehensive-guide-to-statistical-data-visualization-in-python-60f0d7c1de33.