DS065-Advanced EDA

By: Eng. Esraa Madhi

Exploring data is a key step in any data analysis or data science project, as it helps analysts understand their data better, spot trends and connections, and catch any unusual data points or outliers.

While the Python library Pandas is a popular choice for this task—offering a strong array of features for data manipulation, including cleaning, transforming, and charting—the process can still be quite labor-intensive and slow.

Fortunately, using supplementary tools that build upon Pandas' capabilities can make data exploration much more efficient, saving valuable time by simplifying and accelerating various tasks in the data exploration stage.

Pandas Profiling (ydata profiling):

https://github.com/ydataai/ydata-profiling

GitHub - ydataai/ydata-profiling: 1 Line of code data quality profiling & exploratory data analysis for Pandas ...

How to use it:

- https://www.datacamp.com/tutorial/pandas-profiling-ydata-profiling-in-python-guide
- https://docs.profiling.ydata.ai/latest/getting-started/quickstart/

Dtale:

https://github.com/man-group/dtale

GitHub - man-group/dtale: Visualizer for pandas data structures • github.com

How to use it:

https://analyticsindiamag.com/dtale-tutorial-guide-to-visualize-pandas-data-structure/

Dtale Tutorial - Guide To Visualize Pandas Data Structure • analyticsindiamag.com

https://www.analyticsvidhya.com/blog/2020/11/data-exploration-dtale/

Data Exploration with the dtale Library in Python • www.analyticsvidhya.com

---

Sweetviz

https://github.com/fbdesignpro/sweetviz

GitHub - fbdesignpro/sweetviz: Visualize and compare datasets, target values and associations, with one lin...

How to use it:

https://towardsdatascience.com/sweetviz-automated-eda-in-python-a97e4cabacde

Sweetviz: Automated EDA in Python • towardsdatascience.com

---

ipyvizzu

https://github.com/vizzuhq/ipyvizzu

GitHub - vizzuhq/ipyvizzu: Build animated charts in Jupyter Notebook and similar environments with a simpl...

How to use it:

https://towardsdatascience.com/ipyvizzu-a-fast-library-to-build-charts-animations-in-python-126c100522c5

ipyvizzu: A Fast Library to Build Charts' Animations in Python • towardsdatascience.com

---

Image exploration:

https://github.com/Renumics/spotlight

GitHub - Renumics/spotlight: Interactively explore unstructured datasets from your dataframe. • github.com

How to use it:

https://itnext.io/how-to-explore-and-visualize-ml-data-for-object-detection-in-images-88e074f46361

How to Explore and Visualize ML-Data for Object Detection in Images • itnext.io

---

Resources:

- https://medium.com/gustavorsantos/4-good-ways-to-explore-your-data-6a0f4360a254
- https://www.malicksarr.com/top-10-exploratory-data-analysis-eda-libraries/
- https://towardsdatascience.com/comparing-five-most-popular-eda-tools-dccdef05aa4c
- https://medium.com/geekculture/10-automated-eda-libraries-at-one-place-ea5d4c162bbb
- https://builtin.com/data-science/EDA-python
- https://towardsdatascience.com/data-frame-eda-packages-comparison-pandas-profiling-sweetviz-and-pandasgui-bbab4841943b
-