Exploratory Data Analysis (EDA)

There are a number of powerful tools like Pandas Profiling and SweetViz that can make EDA fast and repeatable.

Pandas Profiling

Pandas Profiling is an automated EDA tool that generates rich HTML reports from pandas dataframes. It can be a very nice way to show early progress to a customer when doing data engineering.

SweetViz

SweetViz is a visualisation tool for Python that generates comparisons of data frames. The primary use case is comparison of test and train sets to ensure that they are similar but it could be used for other purposes such as comparing annotated data from different sources.

Revision #3
Created 2022-03-31 20:09:18 UTC by James
Updated 2024-01-21 14:49:27 UTC by James