

26th March, 2022

EDA

.hist() → get information about features.

filtering can be used to get specific information.

`df["age"].head`

`df[df["sex"] == "female"].head()`

↑ gives only the information where the person involved is female

Indexing - `.iloc()` → integer location. First ^{argument} is of row and second argument is column.

`df.iloc[0, :]`

row 0 and all columns

Sorting - dataframe can be sorted in ascending or descending order

`df.sort_values("age", ascending=False).head()`

Grouping - `.groupby("sex").mean`.

Preprocessings

→ Once EDA is done, preprocessing is done. All rows with NAN are dropped `dropna()` and the index is reset `reset_index()`.

→ Columns not needed in the analysis is also dropped `drop()`

→ Features needed are mapped to numerical values are mapped if they are in text

`df["sex"] = df["sex"].map({"female": 0, "male": 1}).astype(int)`