



AniPython

pandas查看数据的基本信息

pandas查看数据的基本信息

DataFrame:	<code>df.info()</code> :	打印摘要
	<code>df.describe()</code> :	描述性统计信息

样本标准差

样本标准差(std)公式:

$$\sqrt{\frac{1}{N-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

样本标准差

样本标准差(std)公式: $\sqrt{\frac{1}{N-1} \sum_{i=1}^n (x_i - \bar{x})^2}$

[1, 1, 2, 5, 6] 均值: $\bar{x} = \frac{1+1+2+5+6}{5} = 3$

$$(1-3)^2 = 4$$

$$(1-3)^2 = 4$$

$$(2-3)^2 = 1$$

$$(5-3)^2 = 4$$

$$(6-3)^2 = 9$$

求和: 22

$$\sqrt{\frac{22}{5-1}} = \sqrt{5.5} = 2.345$$

百分位数

百分位数 25% \rightarrow 10 的意思: 有25%的数小于10

百分位数 50% \rightarrow 30 的意思: 中位数是30

50% \rightarrow 5.5

[0, 1, 2, 5, 6, 100, 200, 500]

25% \rightarrow 1.75

75% \rightarrow 125

pandas查看数据的基本信息

DataFrame:	<code>df.info()</code> :	打印摘要
	<code>df.describe()</code> :	描述性统计信息
	<code>df.values</code> :	数据 <ndarray>
	<code>df.to_numpy()</code>	数据 <ndarray> (推荐)
	<code>df.shape</code> :	形状 (行数, 列数)
	<code>df.columns</code> :	列标签 <Index>
	<code>df.columns.values</code> :	列标签 <ndarray>
	<code>df.index</code> :	行标签 <Index>
	<code>df.index.values</code> :	行标签 <ndarray>
	<code>df.head(n)</code> :	前n行
	<code>df.tail(n)</code> :	尾n行
	<code>pd.options.display.max_columns=n</code> :	最多显示n列
	<code>pd.options.display.max_rows=n</code> :	最多显示n行
	<code>df.memory_usage()</code> :	占用内存(字节B)