**打卡二-数据可视化（EDA）**

**数据导入：**

import pandas as pd

df\_train = pd.read\_csv('./train.csv')

df\_train = pd.read\_csv('./test.csv')

**读取前多少行：**

df\_train\_sample = pd.read\_csv("./train.csv",nrows=5)

**分块读取：**

#设置chunksize参数，来控制每次迭代数据的大小

chunker = pd.read\_csv("./train.csv",chunksize=5)

for item in chunker:

print(type(item))

#<class 'pandas.core.frame.DataFrame'>

print(len(item))

**总体了解：**

df\_train.shape#了解形状

df\_test.shape#了解形状

df\_train.columns#列出所有属性名

df\_test.columns#列出所有属性名

df\_train.head()#显示前n行

df\_train.tail()#显示最后n行

df\_test.tail()

df\_test.head()

df\_train.info()#显示数据集的统计情况，通过info（）我们可以了解到有什么属性存在着缺失值。