

آنالیز داده

فایل `product.csv` شامل اطلاعات بخشی از کالاهای فروشگاه اینترنتی دیجی کالا است. هر سطر در این فایل بیانگر اطلاعات از یک کالای خاص در این فروشگاه است. این اطلاعات شامل نام محصول، برند، طبقه بندی کالا و .. است. ستون اول این فایل شامل id محصولات است. در این سوال شما باید برنامه ای بنویسید که انتظارات زیر را برآورده کند:

- در ابتدا باید برند ۱۰۰ id که در متغیر `prod_id` به شما داده می شود از روی دیتافریم `df` خوانده و نام برند هریک را از ستون `"brand_name_en"` به ترتیب در لیستی با عنوان `brand` ذخیره کند.
- تعداد برند های متمایز در ستون `"brand_name_en"` (روی همه داده ها) را در متغیر `brandUniq` ذخیره کند.
- تعداد دسته بندی های متمایز در ستون `"category_title_fa"` (روی همه داده ها) را در متغیر `categUniq` ذخیره کند.
- تعداد برندهای متمایز در ستون `"brand_name_en"` که نام دسته بندی آن های در ستون `"category_title_fa"` برابر **کفش مردانه** است را در متغیر `BrCat` ذخیره کند.

در نهایت باید متغیرهای بالا توسط این تابع `return` (به ترتیب ذکر شده) شود. برای سهولت در کدنویسی پیکره کد در پایین صفحه برای شما قرار داده شده است. توجه کنید **اسم تابع، ترتیب خروجی ها و تابع `read_input`** را تغییر ندهید و فقط در محلی که برای شما مشخص شده است کد خود را وارد کنید.

```
import pandas as pd

def predict(filename):
    #your dataset
    df = pd.read_csv("product.csv")
    #IDs you should find their corresponding brands
    prod_id = read_input(filename)
    #write your code here
```

```
#####
```

```
return brand, brandUniq, categUniq, BrCat
```

```
def read_input(filename):  
    contents = []  
    f = open(filename, "r")  
    for line in f:  
        contents.append(int(line))  
    return contents
```

توجه: نام فایل پایتون شما می بایست **solution.py** باشد؛ سپس آن را با فرمت **zip** ارسال کنید.

در ادامه یک نمونه ورودی خروجی برای کمک به فهم بیشتر آورده شده است.

ورودی

شما می‌توانید یک نمونه از فایل ورودی را در [اینجا](#) ببینید.

خروجی

شما می‌توانید یک نمونه از فایل خروجی را در [اینجا](#) ببینید.

نکته: در این سوال می‌توانید از کتابخانه *pandas* استفاده کنید.