pandas

May 24, 2022

[ ]:
```
# imporing libraries
import pandas as pd
import seaborn as sns
import numpy as np
```

[ ]:
```
mall_data = pd.read_csv("./data/Mall_Customers.csv")
```

[ ]:
```
mall_data.head()
```

[ ]:

	CustomerID	Genre	Age	Annual Income (k$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40

[ ]:
```
mall_data['annual_income'] = mall_data['Annual Income (k$)']*1000
# mall_data['Annual Income (k$)']*1000
```

[ ]:
```
average_income = mall_data['annual_income'].mean()
average_income
```

[ ]: 60560.0

[ ]:
```
mall_data['above_average_income'] = (mall_data['annual_income'] -␣
 ↪average_income) > 0
mall_data.sample(5)
```

[ ]:

	CustomerID	Genre	Age	Annual Income (k$)	Spending Score (1-100)	\
149	150	Male	34	78	90	
174	175	Female	52	88	13	
102	103	Male	67	62	59	
68	69	Male	19	48	59	
127	128	Male	40	71	95	

	annual_income	above_average_income
149	78000	True
174	88000	True

```
102          62000                 True
68           48000                False
127          71000                 True
```

[ ]:
```
mall_data.to_csv('./data/mall_data_processed.csv', index=False)
```

[ ]:
```
!ls data/
```

```
500_Person_Gender_Height_Weight_Index.csv
Mall_Customers.csv
mall_data_processed.csv
```