Parcial_Pandas_2C_1

October 15, 2020

[31]:
```
import pandas as pd
df_original = pd.read_csv('C:\\Sebas\\Facu\\75.06 Organizacion de␣
↪Datos\\datos_parcial_2C_1.csv', usecols=['uid','sessionid','event','value'])
df_original.head(10)
```

[31]:

	uid	sessionid	event	value
0	1	1	checkout	1
1	1	1	view-product	5
2	1	1	conversion	3
3	1	2	checkout	3
4	1	2	checkout	1
5	2	3	view-product	5
6	2	3	view-product	2
7	2	3	checkout	3
8	2	4	view-product	5
9	3	5	conversion	6

[32]:
```
df_group_x_id_session = df_original.groupby(['uid','sessionid']).agg({'event':
↪'count'})
df_group_x_id_session = df_group_x_id_session.reset_index()
sesiones_x_usuario = df_group_x_id_session.groupby('uid').agg({'sessionid':
↪'count'})
sesiones_x_usuario['promedio_sesiones'] = sesiones_x_usuario['sessionid'].mean()
sesiones_x_usuario
```

[32]:

uid	sessionid	promedio_sesiones
1	2	2.111111
2	2	2.111111
3	3	2.111111
4	2	2.111111
5	2	2.111111
6	1	2.111111
7	3	2.111111
8	1	2.111111
9	3	2.111111

[33]:
```
sesiones_x_usuario.columns = ['cantidad_sesiones','promedio_sesiones_x_usuario']
sesiones_x_usuario
```

[33]:

uid	cantidad_sesiones	promedio_sesiones_x_usuario
1	2	2.111111
2	2	2.111111
3	3	2.111111
4	2	2.111111
5	2	2.111111
6	1	2.111111
7	3	2.111111
8	1	2.111111
9	3	2.111111

[34]:
```
solucion_a = sesiones_x_usuario[sesiones_x_usuario['cantidad_sesiones'] >␣
↪sesiones_x_usuario['promedio_sesiones_x_usuario']]
solucion_a
```

[34]:

uid	cantidad_sesiones	promedio_sesiones_x_usuario
3	3	2.111111
7	3	2.111111
9	3	2.111111

[43]:
```
sesiones_x_usuario = sesiones_x_usuario[sesiones_x_usuario['cantidad_sesiones']␣
↪> 25] #reemplazar 2 por 25
eventos_x_usuario = df_original.merge(sesiones_x_usuario,␣
↪on='uid')[['uid','event','value']]
eventos_x_usuario
```

[43]:

	uid	event	value
0	3	conversion	6
1	3	view-product	2
2	3	conversion	4
3	3	view-product	7
4	3	checkout	2
5	3	view-product	3
6	7	view-product	5
7	7	view-product	4
8	7	conversion	5
9	7	view-product	7
10	7	view-product	5
11	7	checkout	6
12	9	conversion	8
13	9	checkout	2
14	9	checkout	3

```
15    9  view-product      1
16    9  view-product      4
17    9      checkout      3
```

[53]:
```
promedios_x_usuario = eventos_x_usuario.groupby(['uid', 'event']).agg({'value':
↪'mean'})
promedios_x_usuario
```

[53]:
```
                      value
uid event
3   checkout       2.000000
    conversion     5.000000
    view-product   4.000000
7   checkout       6.000000
    conversion     5.000000
    view-product   5.250000
9   checkout       2.666667
    conversion     8.000000
    view-product   2.500000
```

[54]:
```
promedios_x_usuario = promedios_x_usuario.unstack()
promedios_x_usuario
```

[54]:
```
          value
event  checkout conversion view-product
uid
3      2.000000        5.0         4.00
7      6.000000        5.0         5.25
9      2.666667        8.0         2.50
```

[55]:
```
promedios_x_usuario.columns =␣
↪['ecommerce_checkout_mean','ecommerce_conversion_mean','ecommerce_view_product_mean']
promedios_x_usuario
```

[55]:
```
     ecommerce_checkout_mean  ecommerce_conversion_mean  \
uid
3                   2.000000                        5.0
7                   6.000000                        5.0
9                   2.666667                        8.0

     ecommerce_view_product_mean
uid
3                           4.00
7                           5.25
9                           2.50
```