



PONTIFICIA
UNIVERSIDAD
CATÓLICA
DE CHILE

Curso: Análisis de Datos en Python

Quiz sumativo 4

Profesor

Alejandro Cataldo

Utilice el archivo DatosQS4.xlsx. Realice sus códigos, cálculos, desarrollos y análisis en Jupyter Notebook de Python para responder las siguientes preguntas:

Pregunta 1

¿Cuál de las siguientes alternativas permita hacer un gráfico de torta con la distribución de la cantidad de estudiantes por tipo de establecimiento educacional?

- (1) `escuelas[['Tipo','Cantidad de estudiantes']].sum().plot.pie(y='Cantidad de estudiantes', figsize = (8,8))`
`plt.ylabel("")`
`plt.show()`
- (2) `escuelas[['Tipo','Cantidad de estudiantes']].count('Tipo').sum().plot.pie(y='Cantidad de estudiantes', figsize = (8,8))`
`plt.ylabel("")`
`plt.show()`
- (3) `escuelas[['Tipo','Cantidad de estudiantes']].groupby('Tipo').sum().plot.pie(y='Cantidad de estudiantes', figsize = (8,8))`
`plt.ylabel("")`
`plt.show()`
- (4) `escuelas[['Tipo','Cantidad de estudiantes']].groupby('Tipo').count().plot.pie(y='Cantidad de estudiantes', figsize = (8,8))`
`plt.ylabel("")`
`plt.show()`

Pregunta 2

Considere el siguiente código:

```
p2 = furgones[['Furgón', 'Escuela a la que pertenece']].groupby('Escuela a la que pertenece').count()
p2['Conteo'] = p2.apply(lambda row: row.Furgón if row.Furgón < 5 else 5, axis = 1)
p2.groupby([pd.cut(p2["Conteo"], np.arange(0, 6, 1), include_lowest = True)]).count().plot(y = 'Conteo', kind='bar', figsize = (10,7))
plt.ylabel('Cantidad de escuelas')
plt.grid(alpha = 0.5)
plt.show()
```

¿Cuál de las siguientes alternativas indica correctamente lo que hace este código?

Gráfico de barra con la cantidad de escuelas que tienen 0, 1, 2, 3, 4 o 5+ furgones (que pertenecen)

- (1) Un gráfico de torta con la proporción de escuelas a las que pertenecen 0, 1, 2, 3... y así sucesivamente, furgones.
- (2) Un gráfico de torta con la cantidad de furgones que hay en cada escuela.
- (3) Un gráfico de barra con la proporción de escuelas a las que pertenecen 0, 1, 2, 3, 4 y 5 furgones.
- (4) Un gráfico de barra con la proporción de escuelas a las que pertenecen 0, 1, 2, 3, 4, y 5 o más furgones.

Pregunta 3

¿Cuál de las siguientes alternativas corresponde al código para construir un gráfico de dispersión de las ubicaciones de los colegios, y en el que el tamaño del círculo que indica la posición dimensiona la cantidad de estudiantes en ese colegio?

- (1) `escuelas.plot.scatter(s='Cantidad de estudiantes', c = 'Cantidad de estudiantes', colormap = 'plasma',
figsize = (15,8))
plt.show()`
- (2) `escuelas.plot.scatter(x='Ubicación en x', y='Ubicación en y', s='Cantidad de estudiantes', c = 'Cantidad
de estudiantes', colormap = 'plasma', figsize = (15,8))
plt.show()`
- (3) `escuelas.plot.scatter(x, y, s='Cantidad de estudiantes', c = 'Cantidad de estudiantes', colormap =
'plasma', figsize = (15,8))
plt.show()`
- (4) `escuelas.plot.scatter(x='Ubicación en x', y='Ubicación en y', c = 'Cantidad de estudiantes', colormap
= 'plasma', figsize = (15,8))
plt.show()`

Pregunta 4

Considere el siguiente código:

```
furgones.columns = ['Furgón', 'Escuela_pertenece', 'Capacidad', 'Costo_alumno_año']  
  
furgones['Costo_total'] = furgones.apply(lambda row: row.Capacidad*row.Costo_alumno_año, axis  
= 1)
```

```
furgones[['Escuela_pertenece', 'Costo_total']].groupby('Escuela_pertenece').sum().sort_values(by =
'Costo_total', ascending = True).plot(kind = 'bar', figsize = (10,6))

plt.ylabel('Costo total [MM$]')

plt.show()
```

¿cuál de las siguientes alternativas indica claramente lo que hace este código?

- (1) Muestra, a través de un gráfico de barras, el costo total anual que pagan todas las escuelas por transportar estudiantes.
- (2) Muestra, a través de un gráfico de líneas, el costo total que pagaría cada escuela si cada furgón que le pertenece ocupa toda su capacidad transportando niños.
- (3) Muestra, a través de un gráfico de barras, el costo total anual que pagaría cada escuela si cada furgón que le pertenece ocupa toda su capacidad transportando niños.
- (4) Muestra, a través de un gráfico de barras, el costo total anual que gasta cada furgón transportando niños.

Pregunta 5

Usted quiere hacer un gráfico del tipo MapTree, en el que el primer nivel contiene los tres tipos de capacidad de furgones (20, 30 y 35) y el segundo nivel contiene la cantidad de escuelas diferentes que tienen furgones de esa capacidad (ojo que hay escuelas que tiene más de un furgón del mismo tipo (caso Escuela 2)). ¿Cuál de las siguientes alternativas construye este gráfico?

- (1)

```
p5 = furgones[['Capacidad', 'Escuela_pertenece']].groupby(['Escuela_pertenece', 'Capacidad']).sum().reset_index()
p5 = p5.groupby('Capacidad').count().reset_index()
fig = px.treemap(p5, path=['Capacidad', 'Escuela_pertenece'], values='Escuela_pertenece')
fig.update_layout(margin = dict(t=50, l=25, r=25, b=25))
fig.show()
```
- (2)

```
p5 = furgones[['Capacidad', 'Escuela_pertenece']].groupby(['Escuela_pertenece', 'Capacidad']).sum().reset_index()
p5 = p5.groupby('Capacidad').count().reset_index()
fig = px.treemap(p5, path=['Capacidad', 'Escuela_pertenece'], values='Escuela_pertenece')
fig.update_layout(margin = dict(t=50, l=25, r=25, b=25))
fig.show()
```
- (3)

```
p5 = furgones[['Capacidad', 'Escuela_pertenece']].groupby(['Escuela_pertenece', 'Capacidad']).sum().reset_index()
p5 = p5.groupby('Capacidad')
fig = px.treemap(p5, path=['Capacidad', 'Escuela_pertenece'], values='Escuela_pertenece')
fig.update_layout(margin = dict(t=50, l=25, r=25, b=25))
fig.show()
```

```
(4) p5 = furgones[['Capacidad', 'Escuela_pertenece']].groupby(['Escuela_pertenece',  
    'Capacidad']).sum().reset_index()  
p5 = p5.groupby('Capacidad').count().reset_index()  
fig = px.treemap(p5, path=['Escuela_pertenece'], values='Escuela_pertenece')  
fig.update_layout(margin = dict(t=50, l=25, r=25, b=25))  
fig.show()
```