

ESTADÍSTICA DE
NEGOCIOS



TAREA SEMANA 3

MODELADO DE DATOS CONTINUOS

Modelado de datos continuos

- Se ha hablado mucho durante los últimos años sobre la necesidad de que el departamento de RRHH sea más estratégico.
- Aparte de cualquier debate filosófico sobre la verdadera naturaleza de la "estrategia" y el pensamiento estratégico, lo que esto normalmente significa es que los líderes empresariales quieren que los departamentos de RRHH sean más como sus departamentos de marketing: analíticos, orientados a los resultados y capaces de vincular directamente sus actividades con los resultados deseados por la empresa.
- Las empresas de marketing de hoy dependen bastante del análisis predictivo y la minería de datos/ciencia de datos.
 - Es probable que el futuro de los recursos humanos se encuentre en este mismo camino.
- Este ejercicio explora un problema potencial con el que un profesional de RRHH estratégico-analítico podría razonablemente esperar enfrentarse.

Problema: Creative HR

- La empresa Creative HR está analizando los factores de pérdida de personal (**Attrition**).
- El departamento de recursos humanos a recopilado información demográfica y de educación de sus actuales y antiguos empleados.
- El departamento de recursos humanos pretende modelar la edad de sus empleados siguiendo una distribución normal.

Archivo de datos:

- [DL_data.csv](#)
- Lea el archivo csv.

```
import pandas as pd
import seaborn as sns
```

```
data = pd.read_csv("../data/CHR_data.csv") # si la estructura es la del repositorio de GitHub
```

```
# si está trabajando en local y el archivo de datos y el notebook están en el mismo directorio
# data = pd.read_csv("CHR_data.csv")
```

También puede leer el archivo desde el repositorio de GitHub.

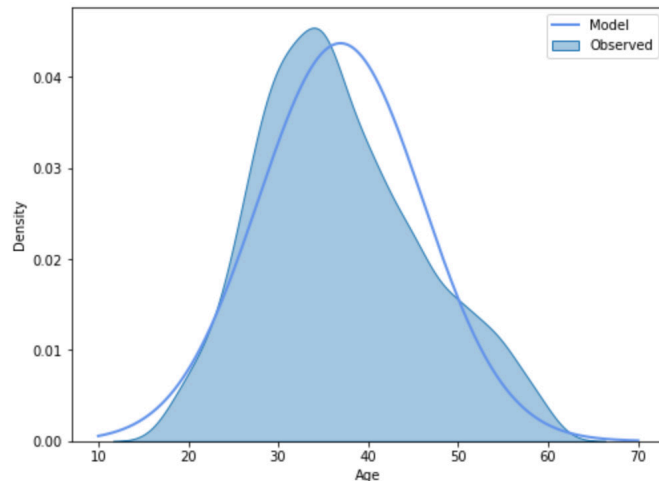
```
#data = pd.read_csv("https://raw.githubusercontent.com/marsgr6/EN-online/8a1cee296279c274d8f05f3905fcea97c1f7636a2/data/CHR_data.csv")
```

1. (2 pt) Cuál es la distribución de la edad (**Age**).

■ Escriba el código Python para replicar el gráfico a continuación.

■ El diagrama de densidad para la variable **Age**.

Una distribución normal $X \sim N(\mu, \sigma)$, donde μ es la media y σ la desviación estándar de los datos (**Age**).



2. (2 pt) Cuál es la probabilidad de que un trabajador de Creative HR tenga 30 años o menos.

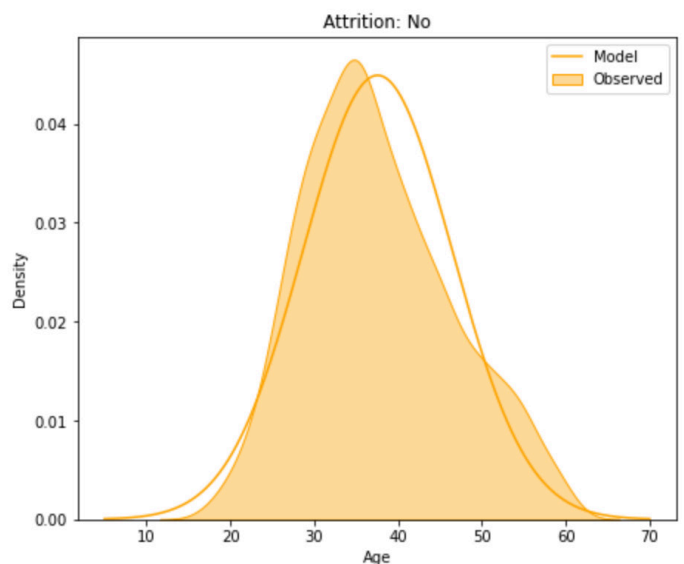
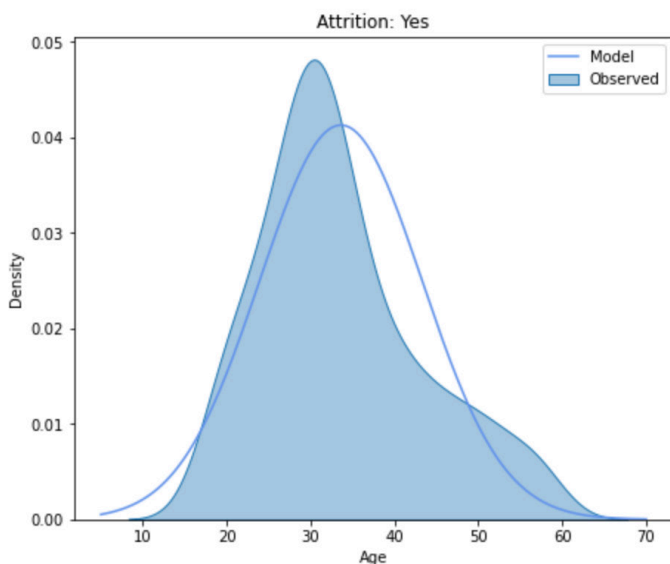
■ Responda usando los datos observados y el modelo asumido.

3. (2 pt) Analice la edad de los trabajadores (**Age**) con respecto a si el trabajador se encuentra o no en la empresa (**Attrition**).

■ Escriba el código Python para replicar el gráfico a continuación.

El diagrama de densidad para la variable **Age**.

Una distribución normal $X \sim N(\mu, \sigma)$, donde μ es la media y σ la desviación estándar de la edad (**Age**) de acuerdo a si se encuentra aún en la empresa (**Attrition**).



4. **(2 pt)** Cuál es la probabilidad de que un trabajador que ya no se encuentra en Creative HR tenga 30 años o menos.

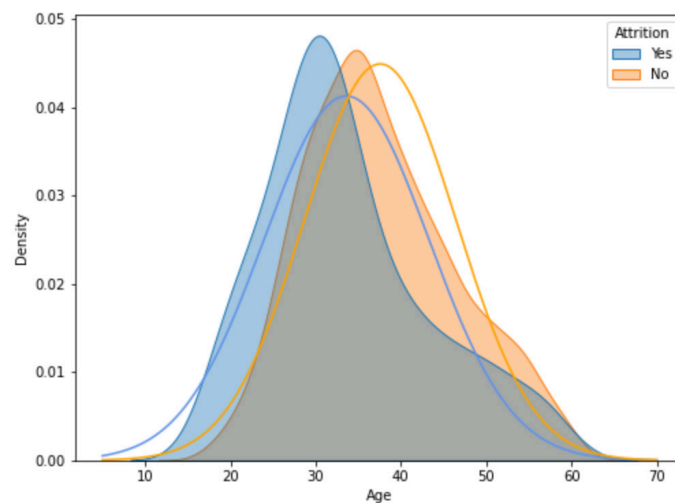
- Responda usando los datos observados y el modelo asumido.

5. **(2 pt)** Discuta qué tan adecuada considera la distribución usada para modelar nuestros datos observados.

- Responda usando los datos observados y el modelo asumido.

1. Opcional

- Grafique los datos en un mismo panel.



Fuentes y recursos:

- <https://www.kaggle.com/jamestollefson/modeling-the-business-cost-of-retention/notebook>