

Columnas redundantes y/o con valores inconsistentes

<p><code>monthlyincome</code>, <code>sameasmonthlyincome</code> y <code>salary</code></p> <ul style="list-style-type: none">Todas parecen expresar ingresos mensuales o anuales en distintas unidades/formatos.<code>sameasmonthlyincome</code> explícitamente es un duplicado de <code>monthlyincome</code>.	<p>Por el momento, nos quedamos con <code>monthlyincome</code> y <code>salary</code></p> <p>✓ Eliminamos <code>sameasmonthlyincome</code></p>
<p><code>department</code>, <code>roledespartament</code> y <code>jobrole</code></p> <ul style="list-style-type: none"><code>roledespartament</code> mezcla el rol con el departamento (Manager - Research & Development), mientras que <code>department</code> y <code>jobrole</code> ya lo tiene separado.	<p>Por el momento nos quedamos con <code>department</code> y <code>jobrole</code>.</p> <p>✓ Eliminamos <code>roledespartament</code></p> <p>porque más tarde podemos fusionar las anteriores, si lo necesitamos.</p>
<p><code>employeecount</code> y <code>employeenumber</code></p> <ul style="list-style-type: none"><code>employeecount</code> es siempre 1 (columna inútil).<code>employeenumber</code> parece ser solo un identificador único → no aporta al análisis predictivo.	<p>✓ Eliminamos <code>employeecount</code></p>
<p><code>age</code> y <code>datebirth</code></p> <ul style="list-style-type: none">Son equivalentes: de <code>datebirth</code> se puede calcular la edad.	<p>✓ Eliminamos <code>age</code></p>
<p><code>numberchildren</code></p> <ul style="list-style-type: none">Está completamente vacío (0 valores válidos).	<p>✓ La eliminamos</p>
<p><code>remotework</code></p> <ul style="list-style-type: none">Tiene valores como <code>"Yes"</code>, <code>"1"</code>, <code>"0"</code>, <code>"False"</code>. Es la misma información codificada de diferentes maneras → se debe normalizar.	<p>La tenemos que normalizar a yes/no</p>