

ANÁLISIS ESTIMACIÓN SALARIOS FEMENINOS

Una vez realizada la estimación de los salarios del dataset formado únicamente por las mujeres, a partir de técnicas de Machine Learning entrenando y testeando los modelos únicamente con el dataset de hombres, conseguimos una estimación del salario a partir de variables dependientes relacionadas con el trabajo realizado como nivel de estudios, actividad, tamaño del centro etc.

Si no existieran diferencias salariales por cuestiones de género a igualdad de condiciones laborales, el salario real y el estimado deberían ser muy similares.

A continuación, vamos a analizar las posibles diferencias entre lo estimado y lo real.

Comenzamos calculando el salario medio real para los hombres con tipo de jornada a tiempo completo y lo compararemos con el salario estimado que hemos calculado a través de técnicas de aprendizaje automático.

```
##           Real Estimado
## [1,] 28317.92 28160.64
```

Como podemos observar, el salario medio estimado es muy similar al salario medio real en el caso de los hombres, vamos a analizar estas dos métricas en el caso del dataset femenino.

```
##           Real Estimado diferencia DifPorcentual
## 1 25041.86 28346.75    3304.893    13.19747
```

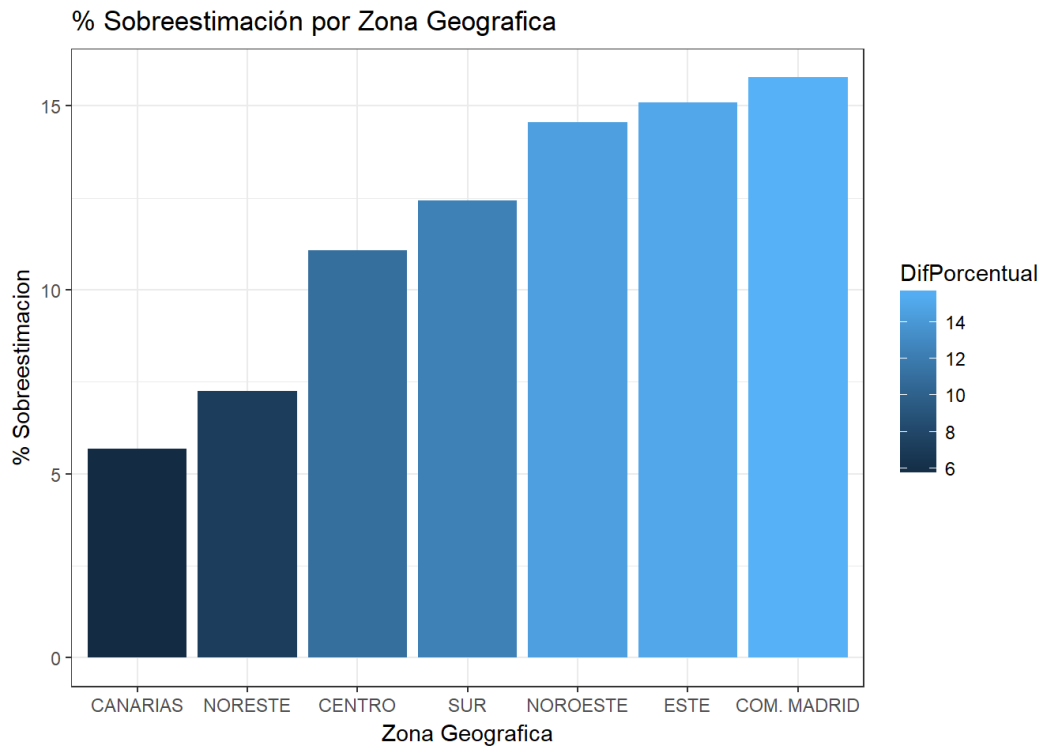
Podemos observar una diferencia considerable entre el salario Real de las mujeres y el salario Estimado (más de 3000 euros), es decir el modelo entrenado a partir del conjunto de datos masculino sobreestima el salario femenino, por lo que podemos concluir que realmente existen diferencias salariales por cuestión de género.

A continuación, vamos a analizar en qué condiciones existe mayor diferencia entre el salario Real y el Estimado.

Comenzamos analizando las zonas geográficas:

```
## # A tibble: 7 x 5
##   NUTS1           Real Estimado diferencia DifPorcentual
##   <chr>         <dbl>     <dbl>         <dbl>         <dbl>
## 1 COM. MADRID 28573.    33079.         4506.         15.8
## 2 ESTE      24492.    28190.         3698.         15.1
## 3 NOROESTE  22275.    25517.         3243.         14.6
## 4 SUR       23599.    26529.         2930.         12.4
## 5 CENTRO    23051.    25605.         2554.         11.1
## 6 NORESTE   27202.    29174.         1972.          7.25
## 7 CANARIAS  21917.    23163.         1245.          5.68
```

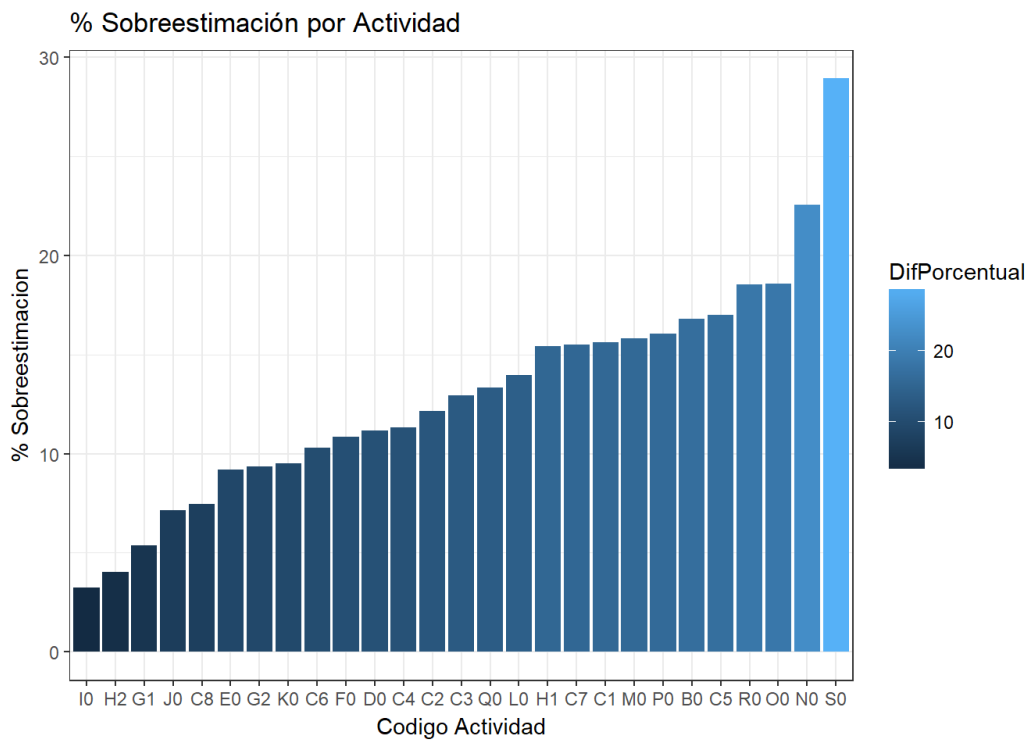
La comunidad de Madrid es dónde mayor diferencia existe entre el salario estimado y el real siendo el estimado un 15,8% superior al real, por el contrario, Canarias es la zona geográfica con menor diferencia un 5,7% superior.



Si analizamos la actividad económica nos encontramos con que otros servicios, actividades asociativas... es la actividad que mayor sobreestimación tiene, casi un 30%, seguida de actividades administrativas y servicios auxiliares. Por el contrario, Transporte y almacenamiento y hostelería son las actividades con menor diferencia entre el salario estimado y el real con menos de un 5%.

```
## # A tibble: 2 x 5
##   DES_CNACE          Real Estimado diferencia DifPorcentual
##   <chr>          <dbl>     <dbl>         <dbl>         <dbl>
## 1 Otros servicios: actividades a~ 18094.    23325.      5231.         28.9
## 2 Actividades administrativas y ~ 20037.    24555.      4518.         22.5
## # A tibble: 2 x 5
##   DES_CNACE          Real Estimado diferencia DifPorcentual
##   <chr>          <dbl>     <dbl>         <dbl>         <dbl>
## 1 Transporte y almacenamiento: a~ 24802.    25798.        996.          4.02
## 2 Hostelería: servicios de aloja~ 18181.    18770.        589.          3.24
```

El siguiente gráfico muestra la diferencia porcentual del salario real y el estimado por códigos de actividad:

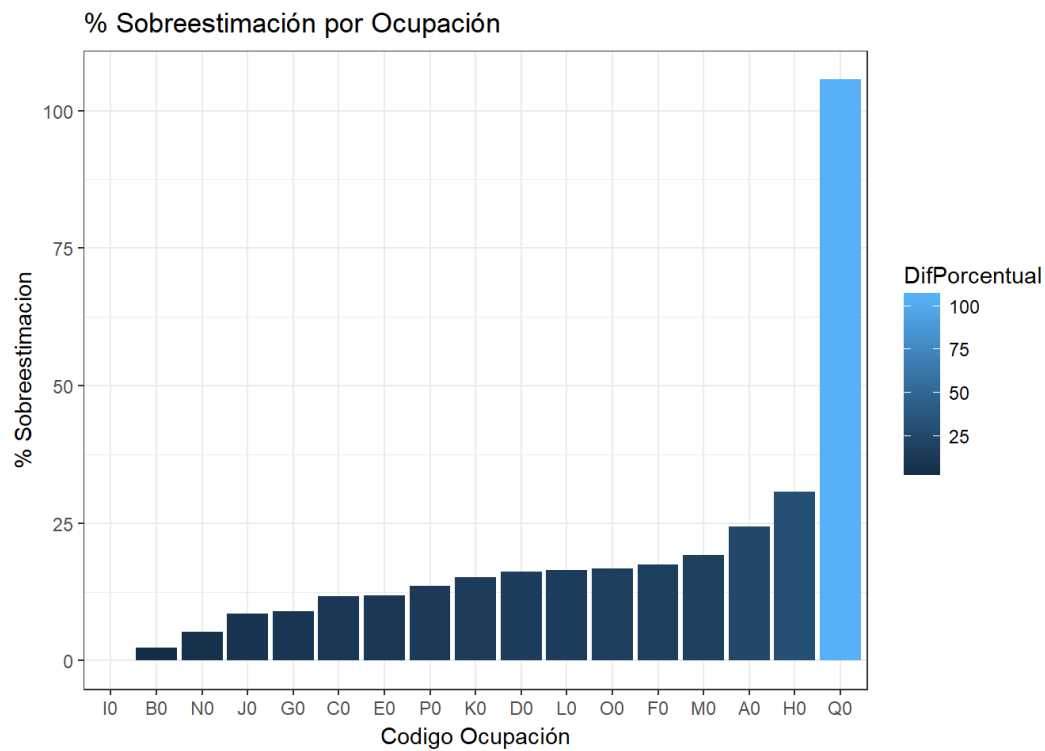


A continuación, analizamos las diferencias por ocupación.

```
## # A tibble: 2 x 5
##   DES_TCNO      Real Estimado diferencia DifPorcentual
##   <chr>      <dbl>     <dbl>      <dbl>         <dbl>
## 1 OCUPACIONES MILITARES      9681.    19910.    10228.         106.
## 2 TRABAJADORES DE LOS SERVICIOS ~ 18264.    23872.     5608.          30.7
## # A tibble: 2 x 5
##   DES_TCNO      Real Estimado diferencia DifPorcentual
##   <chr>      <dbl>     <dbl>      <dbl>         <dbl>
## 1 TÉCNICOS Y PROFESIONALES CIENT~ 33240.    34046.      806.           2.42
## 2 TRABAJADORES DE LOS SERVICIOS D~ 24633.    24650.       17.2          0.0698
```

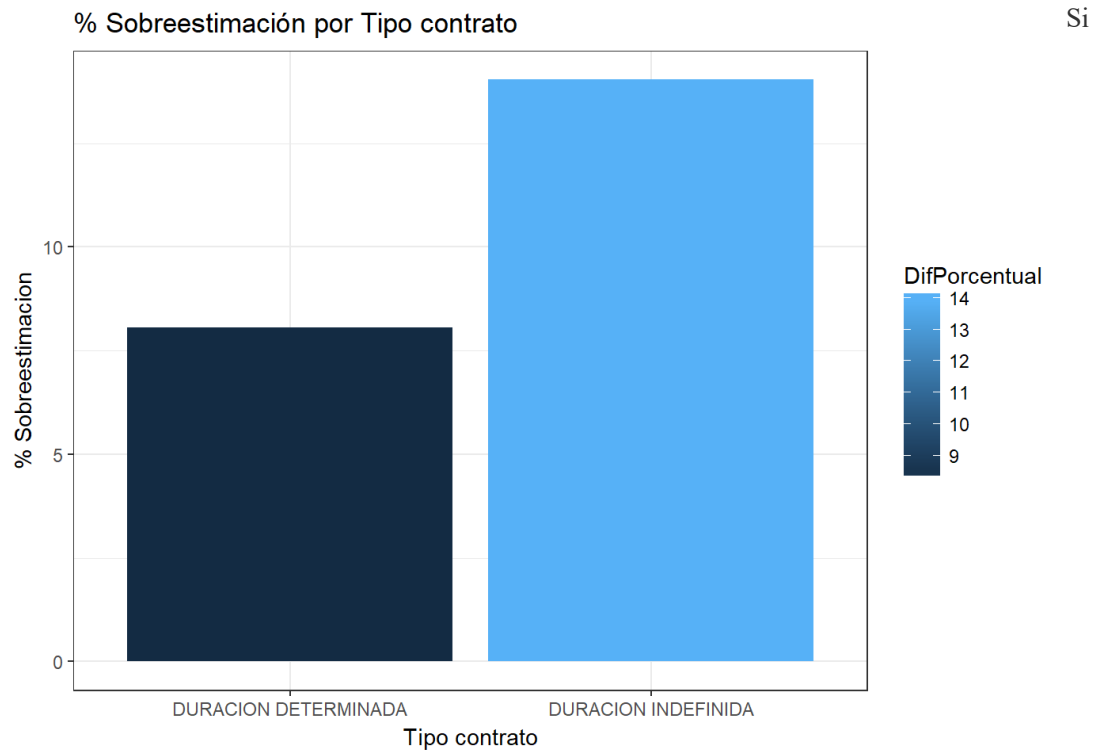
Ocupaciones militares es la que mayor diferencia presenta entre el salario estimado y el real, siendo el estimado más de 100% superior al real, le sigue los trabajadores de los servicios de salud y el ciudad de las personas con una sobreestimación de más del 30%.

Por el contrario, técnicos y profesionales científicos e intelectuales y trabajadores de los servicios de protección y seguridad son los que presentan un valor estimado más acorde con el real.



Si analizamos el tipo de contrato, la mayor diferencia entre Estimado y real se presenta en los contratos con duración indefinida.

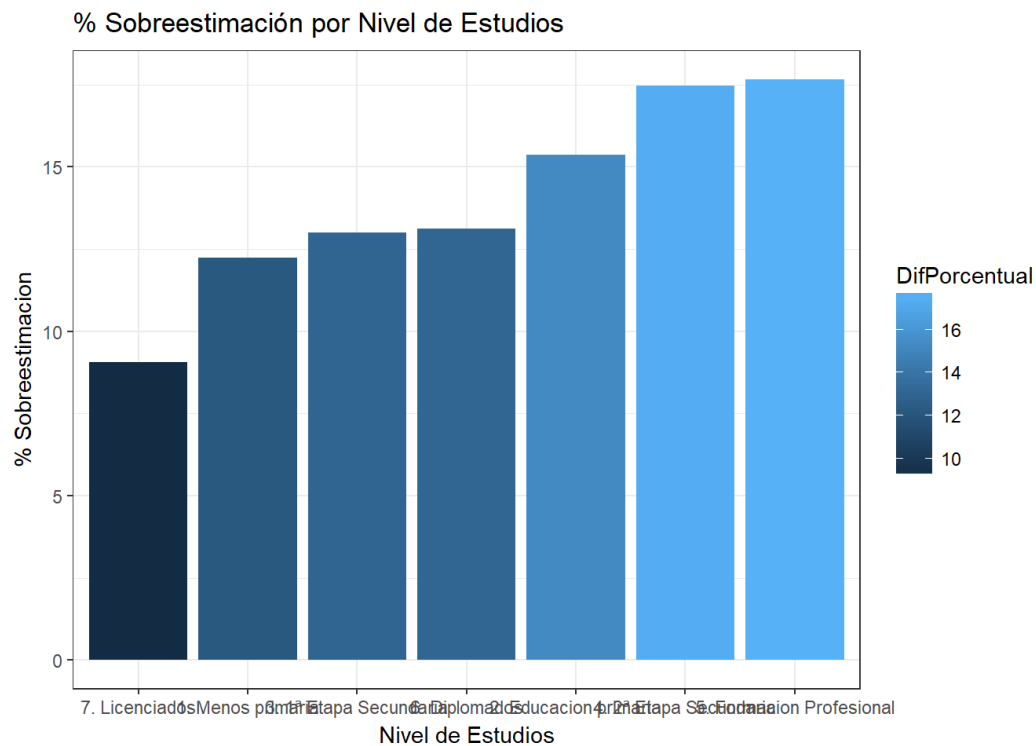
```
## # A tibble: 2 x 5
##   TIPOCON      Real Estimado diferencia DifPorcentual
##   <chr>      <dbl>    <dbl>      <dbl>      <dbl>
## 1 DURACION INDEFINIDA 25805.  29427.    3622.      14.0
## 2 DURACION DETERMINADA 21189.  22894.    1705.       8.05
```



analizamos el nivel de estudios, nos encontramos que dónde mayor diferencia salarial existe es en las enseñanzas de formación profesional de grado superior con casi un 18%, seguida de segunda etapa de educación secundaria. Los licenciados y similares son los que menor diferencia salarial entre género presentan.

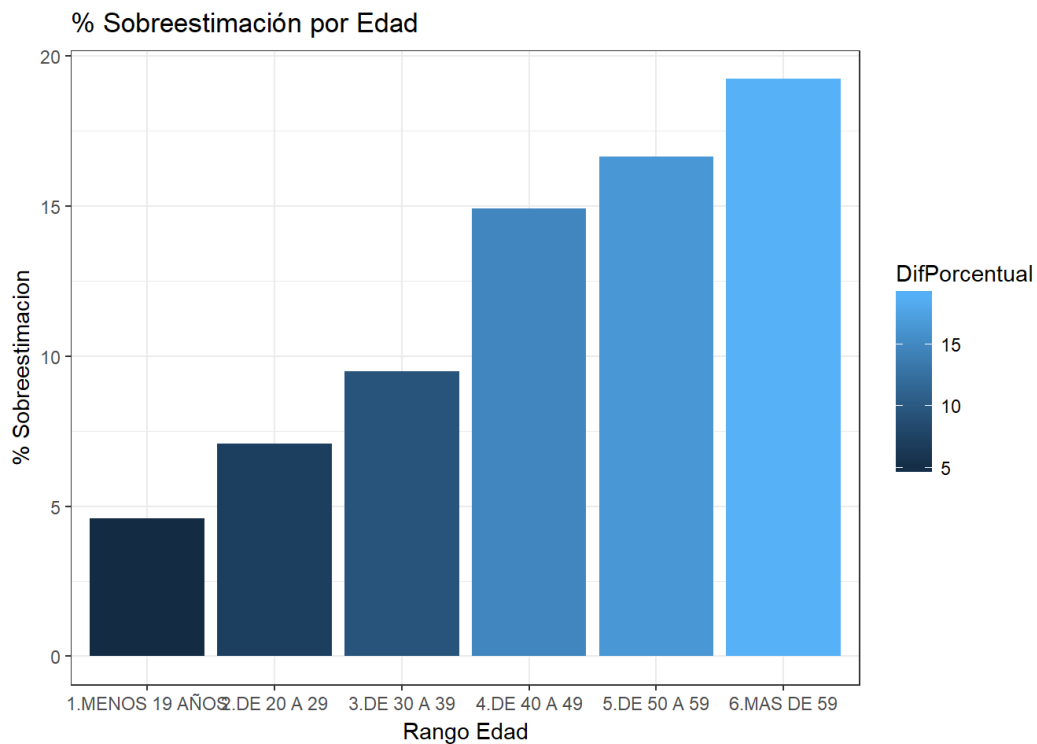
```
## # A tibble: 7 x 5
```

	ESTU	Real	Estimado	diferencia	DifPorcentual
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>
## 1	5.Enseñanzas de formación prof~	23159.	27252.	4093.	17.7
## 2	4.Segunda etapa de educacion s~	21421.	25162.	3741.	17.5
## 3	2.Educacion primaria	16873.	19465.	2593.	15.4
## 4	6.Diplomados universitarios y ~	29262.	33100.	3838.	13.1
## 5	3.Primer etapa de educacion s~	19065.	21546.	2481.	13.0
## 6	1.Menos que primaria	17287.	19400.	2114.	12.2
## 7	7.Licenciados y similares, y d~	35495.	38710.	3215.	9.06

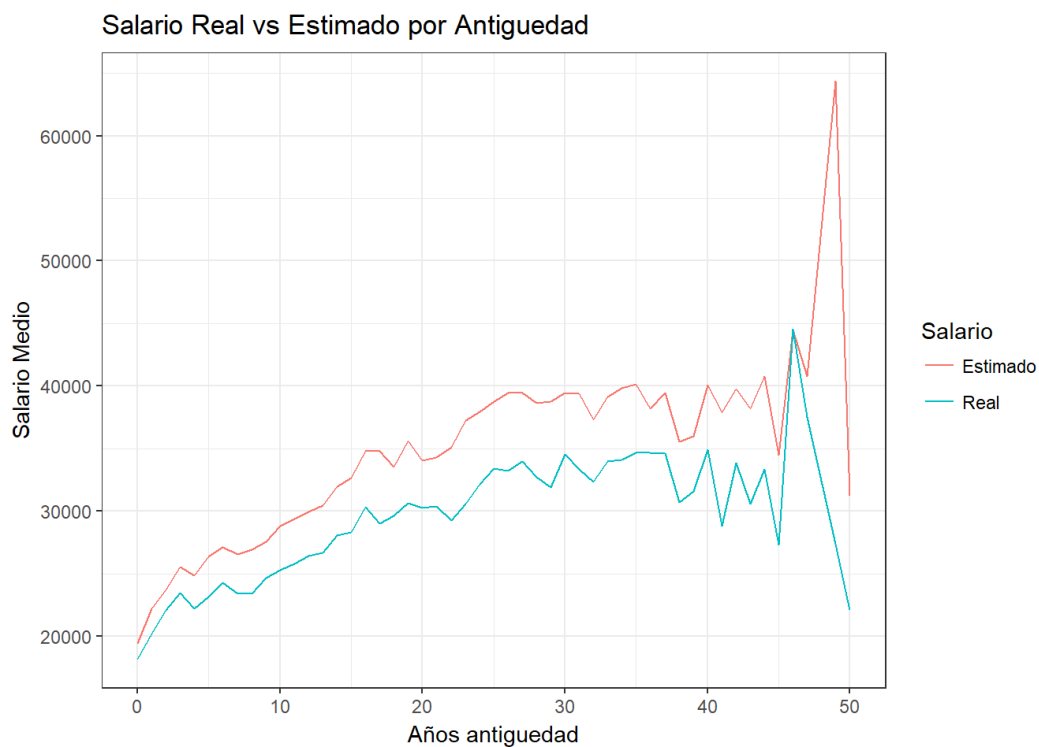


Al analizar la edad, nos encontramos que las diferencias salariales van aumentando a medida que aumenta la edad, llegando a superar el 19% de sobreestimación en las mujeres con más de 59 años.

```
## # A tibble: 6 x 5
##   ANOS2      Real Estimado diferencia DifPorcentual
##   <chr>      <dbl>      <dbl>      <dbl>      <dbl>
## 1 6.MAS DE 59 26809.    31963.    5154.     19.2
## 2 5.DE 50 A 59 27383.    31938.    4555.     16.6
## 3 4.DE 40 A 49 26610.    30577.    3967.     14.9
## 4 3.DE 30 A 39 23908.    26179.    2272.      9.50
## 5 2.DE 20 A 29 18570.    19884.    1314.      7.07
## 6 1.MENOS 19 AÑOS 14310.    14966.     656.      4.58
```



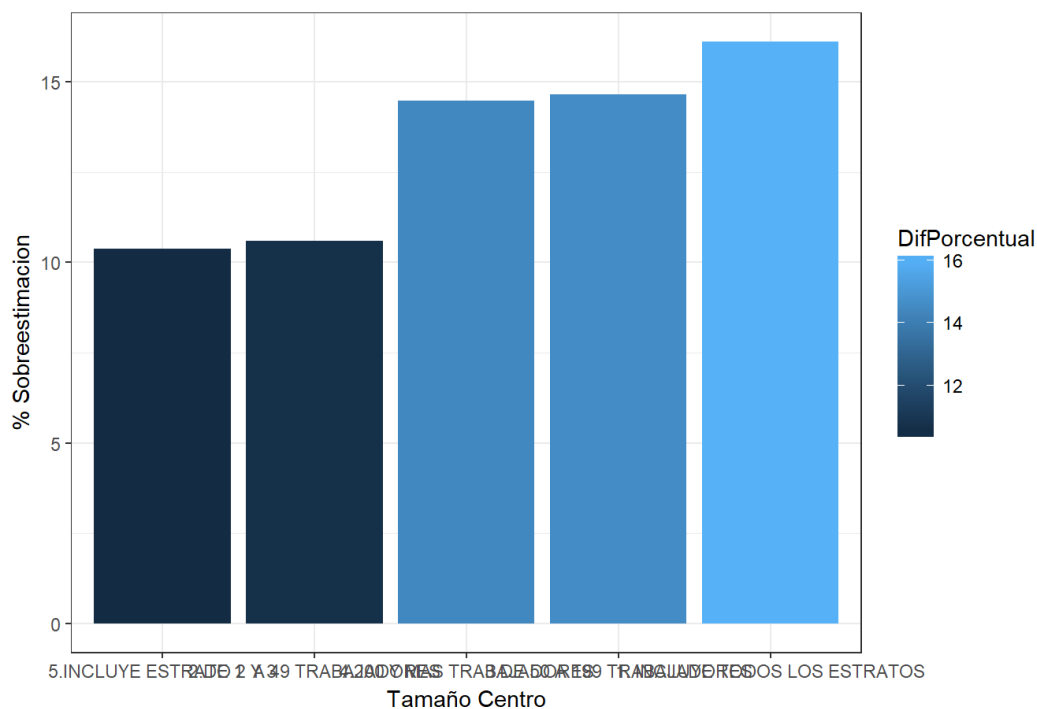
En el siguiente gráfico observamos la diferencia salarial entre Estimado y Real por los años de antigüedad en la empresa, dónde parece observarse que en antigüedad de pocos años es dónde menor diferencia existe.



Si estudiamos el tamaño del centro, podemos observar que con más de 59 trabajadores existe una sobre estimación de más de 14,5%, mientras que en las empresas con menos de 50 trabajadores se sitúa en 10,6%.

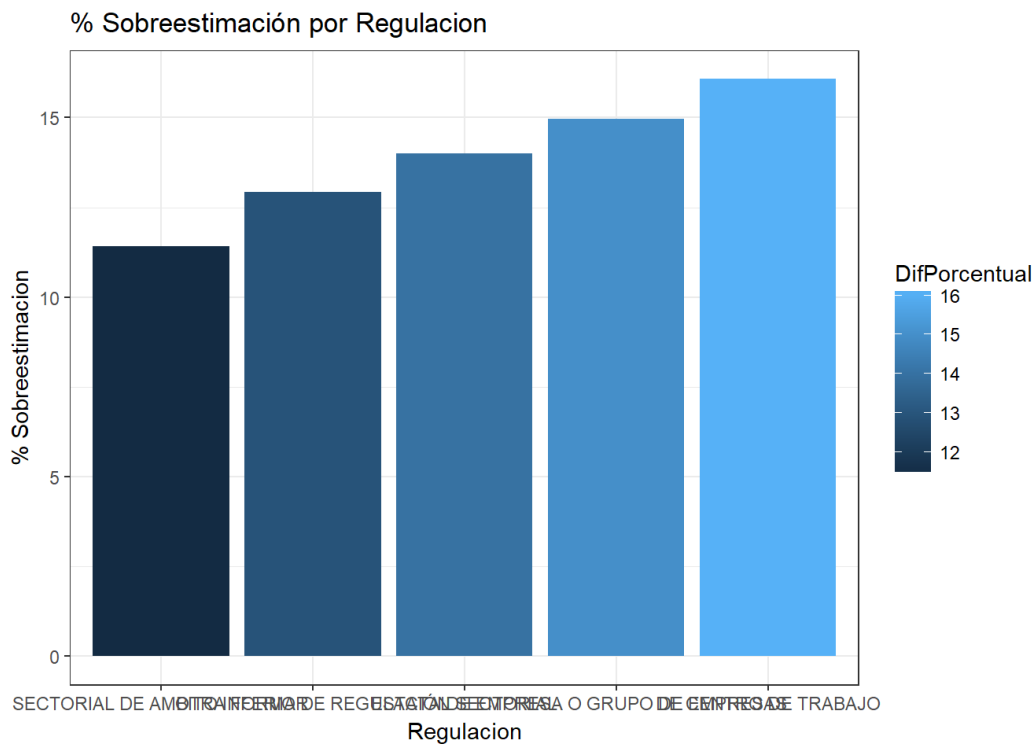
```
## # A tibble: 5 x 5
##   ESTRATO2      Real Estimado diferencia DifPorcentual
##   <chr>      <dbl>    <dbl>      <dbl>      <dbl>
## 1 1. INCLUYE TODOS LOS ESTRATOS 21367.    24811.     3445.      16.1
## 2 3.DE 50 A 199 TRABAJADORES  25362.    29079.     3717.      14.7
## 3 4.200 Y MAS TRABAJADORES    28914.    33101.     4187.      14.5
## 4 2.DE 1  A 49 TRABAJADORES   20955.    23178.     2223.      10.6
## 5 5.INCLUYE ESTRATO 2 Y 3      25422.    28062.     2640.      10.4
```

% Sobreestimación por Tamaño del Centro



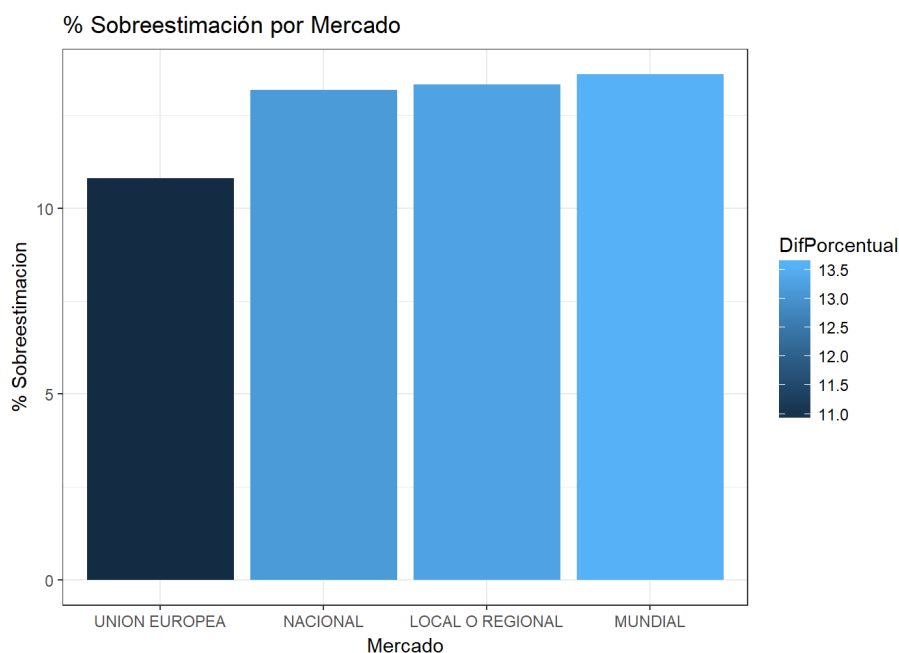
Si estudiamos el salario por el tipo de convenio colectivo, tenemos que de centro de trabajo presenta una sobreestimación de más de un 16% siendo el tipo de convenio con mayor diferencia entre el salario estimado y el real, por el contrario, sectorial de ámbito inferior con un 11,4% de sobreestimación es el tipo de convenio con menor diferencia.

```
## # A tibble: 5 x 5
##   REGULACION      Real Estimado diferencia DifPorcentual
##   <chr>      <dbl>    <dbl>      <dbl>      <dbl>
## 1 DE CENTRO DE TRABAJO      26668.    30954.     4286.      16.1
## 2 DE EMPRESA O GRUPO DE EMPRESAS 28523.    32793.     4270.      15.0
## 3 ESTATAL SECTORIAL        24520.    27955.     3435.      14.0
## 4 OTRA FORMA DE REGULACIÓN    28871.    32604.     3733.      12.9
## 5 SECTORIAL DE AMBITO INFERIOR  21932.    24438.     2506.      11.4
```

Por último, analizamos el tipo de mercado de destino de la empresa, dónde nos encontramos que si el destino es Union Europea presenta menor sobreestimación que el resto, situándose en un 10,8% mientras que el resto de mercado supera el 13%.

```
## # A tibble: 4 x 5
##   MERCADO      Real Estimado diferencia DifPorcentual
##   <chr>      <dbl>    <dbl>      <dbl>      <dbl>
## 1 MUNDIAL    30948.   35159.     4212.       13.6
## 2 LOCAL O REGIONAL 23577.   26719.     3142.       13.3
## 3 NACIONAL   25821.   29225.     3403.       13.2
## 4 UNION EUROPEA  25417.   28163.     2746.       10.8
```



CONCLUSIONES

Según el análisis realizado, se puede determinar que existe una diferencia salarial de género que afecta negativamente a las mujeres.

El modelo realizado sobreestima el salario femenino con más de 3.000 euros de diferencia en el salario medio, es decir, se sobreestima en más de un 13%.

Esta diferencia salarial se ve incrementada en la Comunidad de Madrid y el Este de España, en actividades relacionadas con otros servicios, actividades asociativas y administrativas. En ocupaciones Militares y trabajadores de los servicios de salud.

Además, en tipo de contratos con duración indefinida y en aquellos asalariados con nivel de estudios formación profesional de grado superior.

La edad afecta en la diferencia salarial, dándose mayor diferencia a medida que aumenta la edad de las asalariadas.

En empresas con más de 50 empleados existe también mayor diferencia salarial y en aquellas empresas con tipo de convenio colectivo de centro de trabajo.

Por el contrario, se puede decir que no existe esa diferencia salarial o es bastante inferior en Canarias y el Noreste de España, en actividades de hostelería y transporte y almacenamiento, en técnicos y profesionales científicos y trabajadores de los servicios de salud y cuidado de las personas.

Tampoco existe tal diferencia en Licenciados y similares o en los asalariados con menos de 19 años.