

¿Qué factores influyen en los salarios del sector IT?

Por Lara Díaz, Gino Mariani, Isaías Frellatt y Tomás Martín
GRUPO 15

INTRODUCCIÓN



Habiendo tantos sueldos como trabajadores nos preguntamos:

¿Es posible encontrar variables que definen el salario de un empleado de IT en Argentina en el 2025?

SOBRE EL DATA SET

Fuente: sysarmy.com/blog - Encuesta de sueldos 2025.1



Variables categóricas

1. **donde_estas_trabajando**
2. **dedicacion**
3. **recibis_algun_tipo_de_bono**
4. **contas_con_beneficios_adicionales**
5. **carrera**
6. **trabajo_de**
7. **cantidad_de_personas_en_tu_organizacion**
8. **modalidad_de_trabajo**
9. **maximo_nivel_de_estudios**
10. **genero**
11. **sueldo_dolarizado**
12. **seniority**

Variables numéricas [rango]

1. **anos_de_experiencia** [0,45]
2. **antiguedad_en_la_empresa_actual** [0,40]
3. **anos_en_el_puesto_actual** [0,40]
4. **cuantas_personas_tenes_a_cargo** [0,130]
5. **tengo_edad** [20,74]
6. **X_sal** [150.000, 12.960.000]

LIMPIEZA

Se filtraron las variables para quedarnos solo con las descriptas

El rango de edad se acotó a < 85 años.

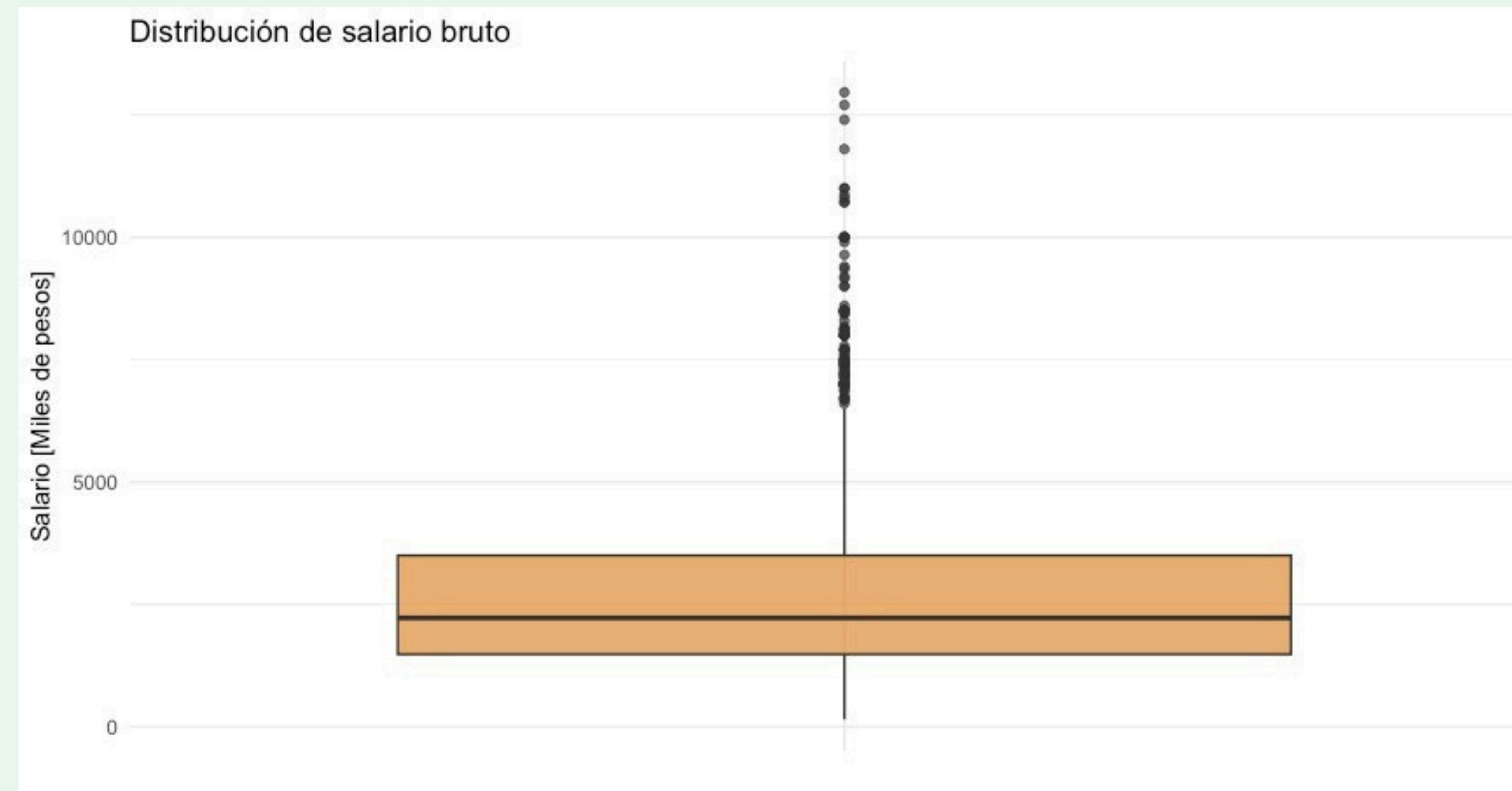
Variable salario (x_sal) se dividió por mil para mejorar la visualización de los gráficos

Se reemplazaron los valores de > 60 años de antigüedad en la empresa por NA, por considerarse valores atípicos.

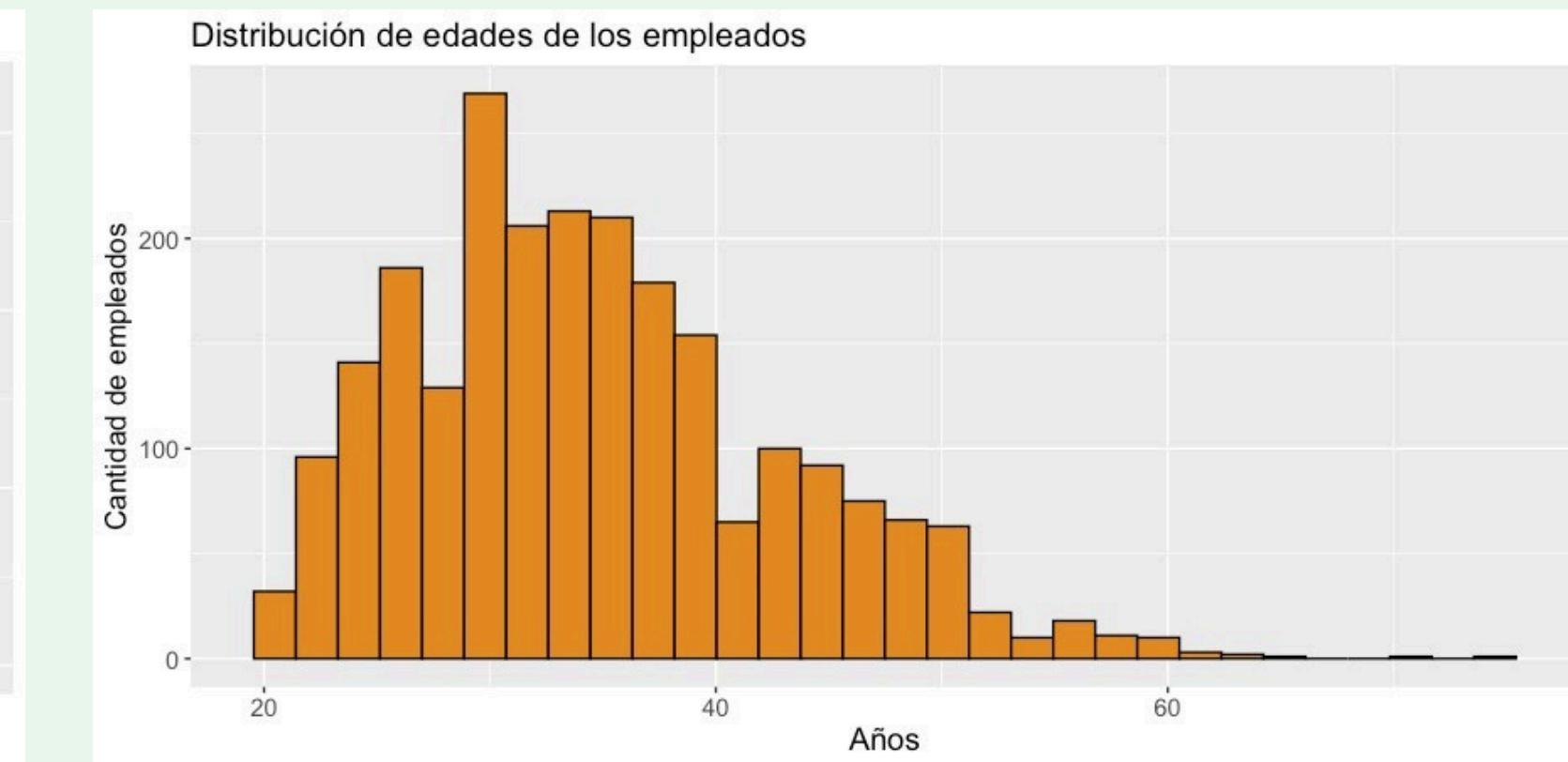
Se reemplazaron los valores "Doctorado" y "Posdoctorado" de la variable "máximo nivel de estudio", por NA, por tratarse de outliers.

La variable sobre beneficios adicionales se transformó en una variable binaria (si/no)

ANÁLISIS EXPLORATORIO

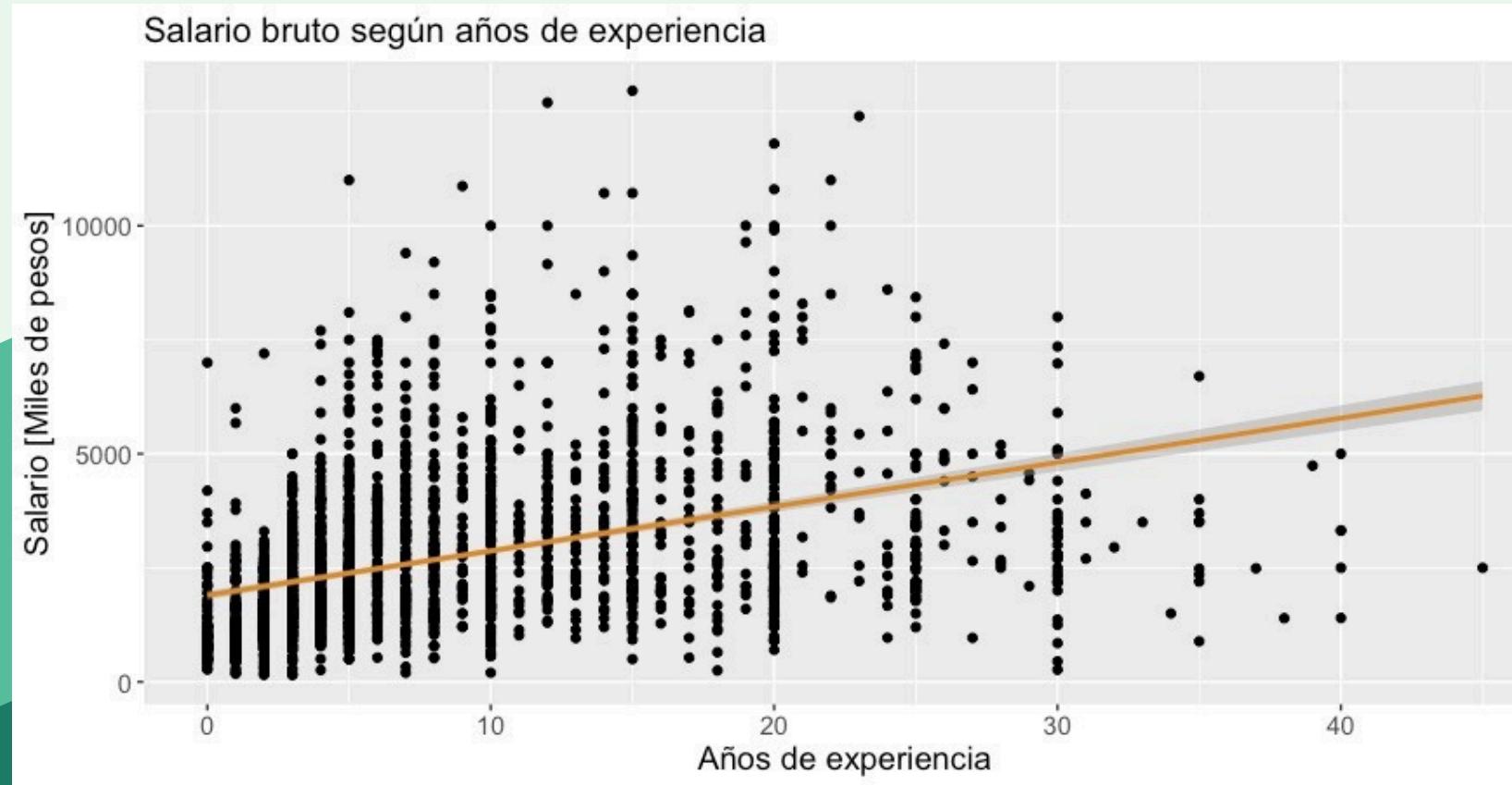
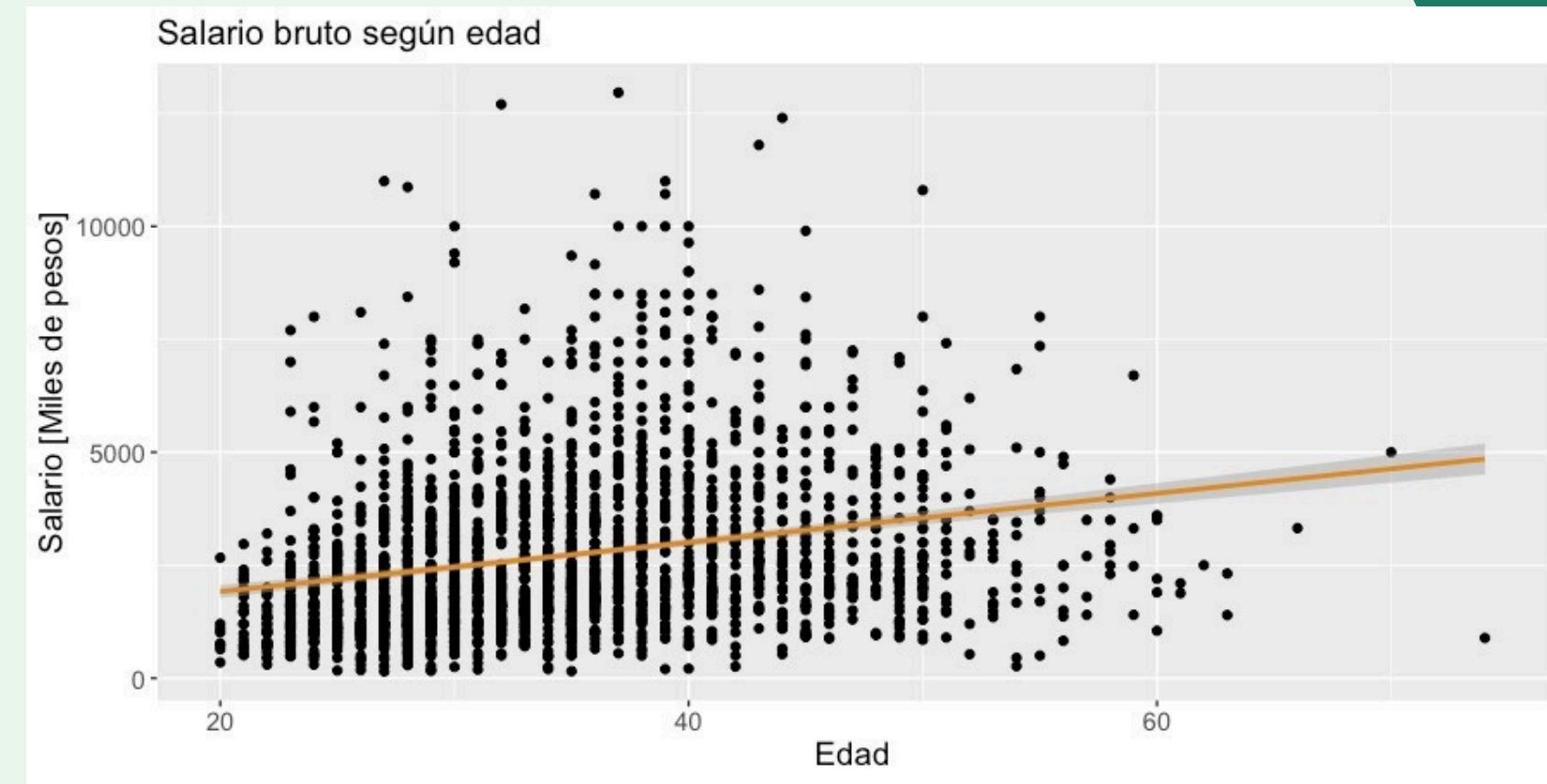
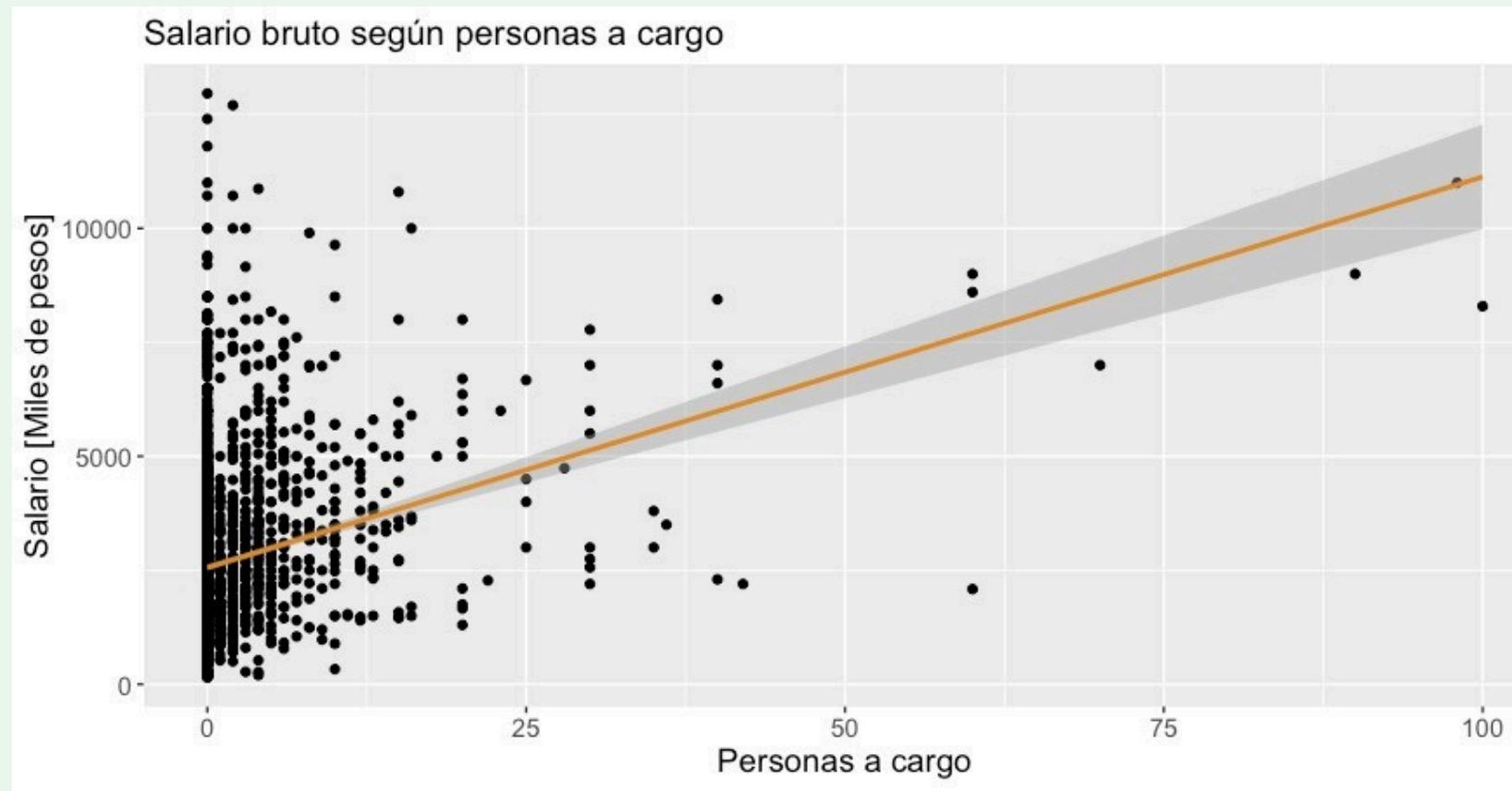


Variable target. Hay muchos outliers que distorsionan la distribución.



ANÁLISIS EXPLORATORIO

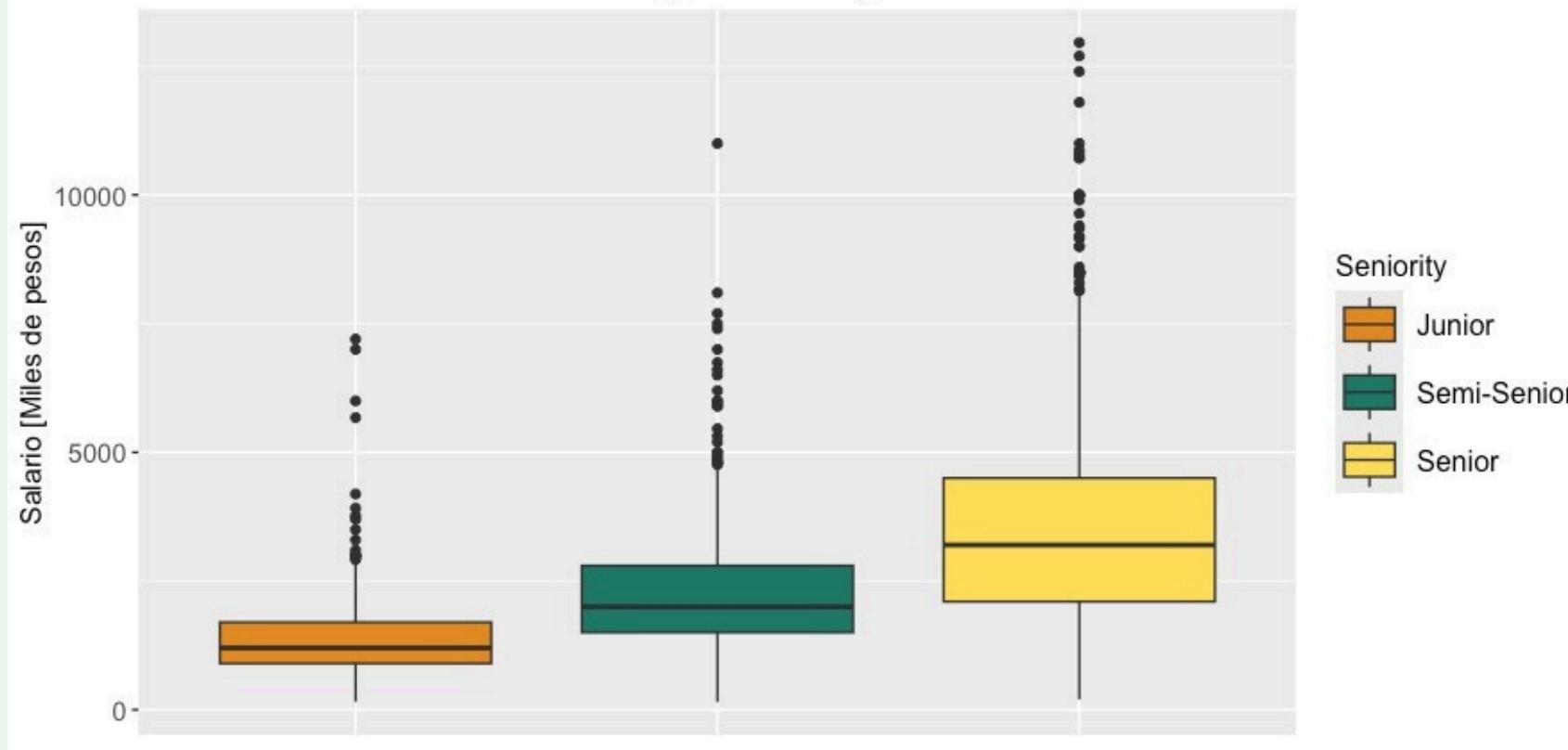
RELACIONES NUMÉRICAS



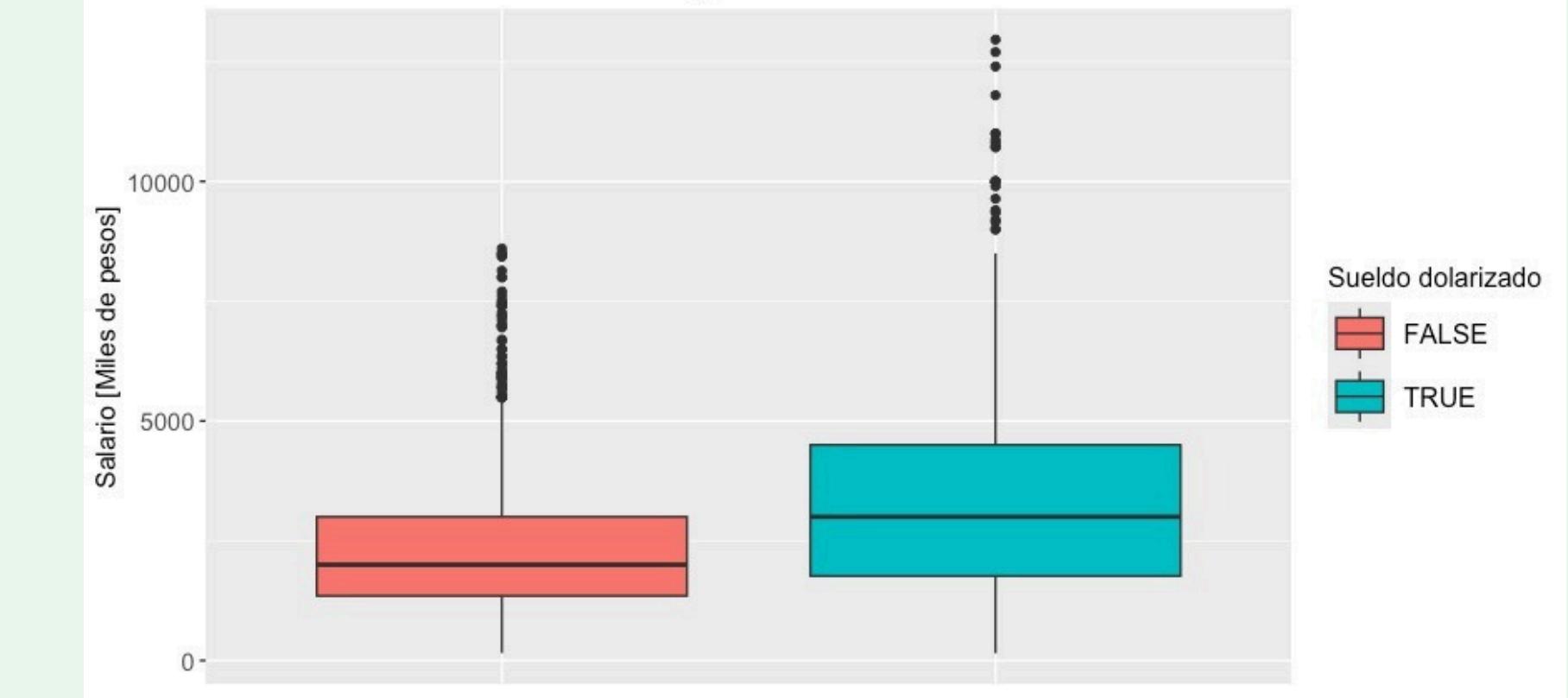
Las variables numéricas que muestran mayor relación con el salario son las personas a cargo, la edad y los años de experiencia.

ANÁLISIS DETALLADO

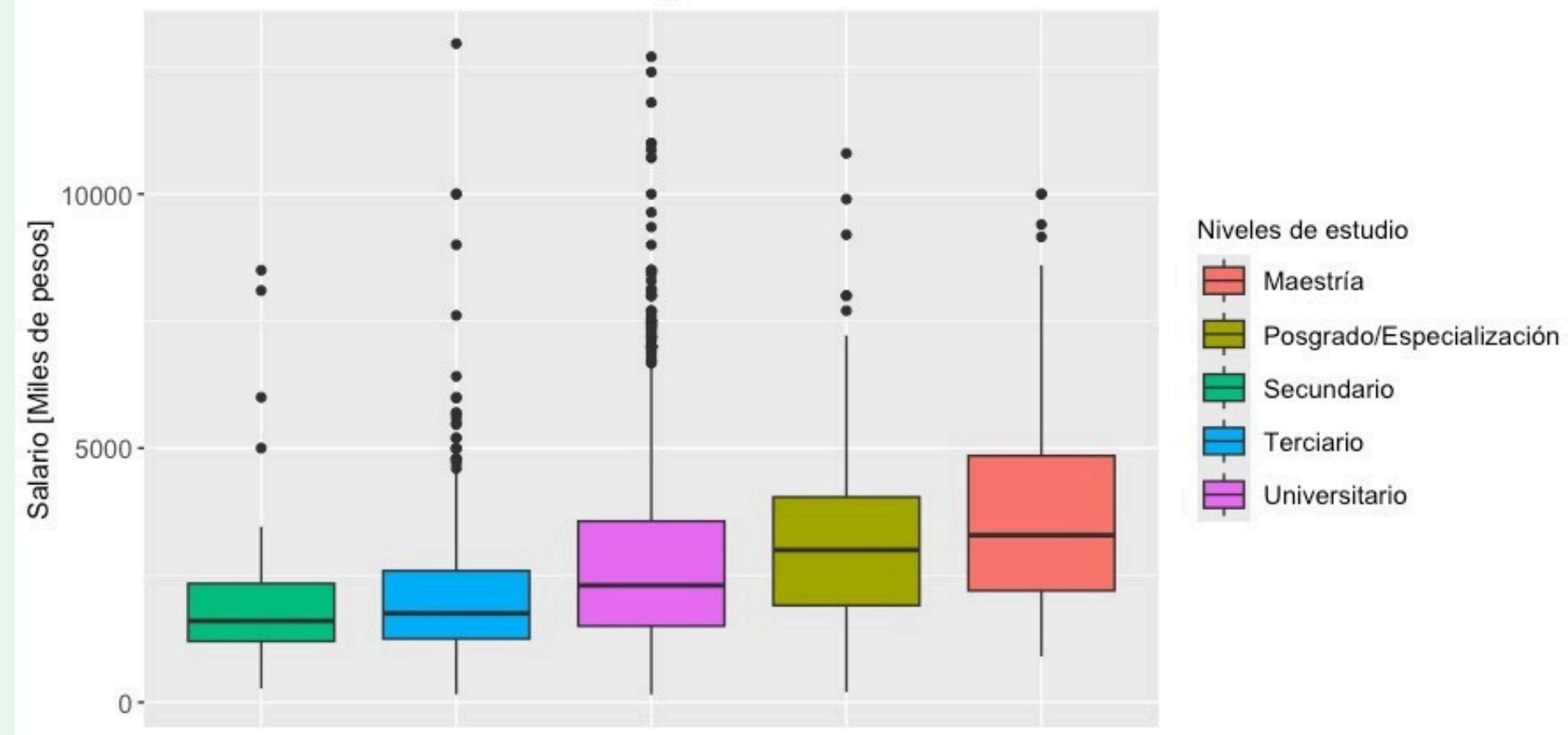
Distribución de salario bruto según seniority



Distribución de salario bruto según sueldo dolarizado

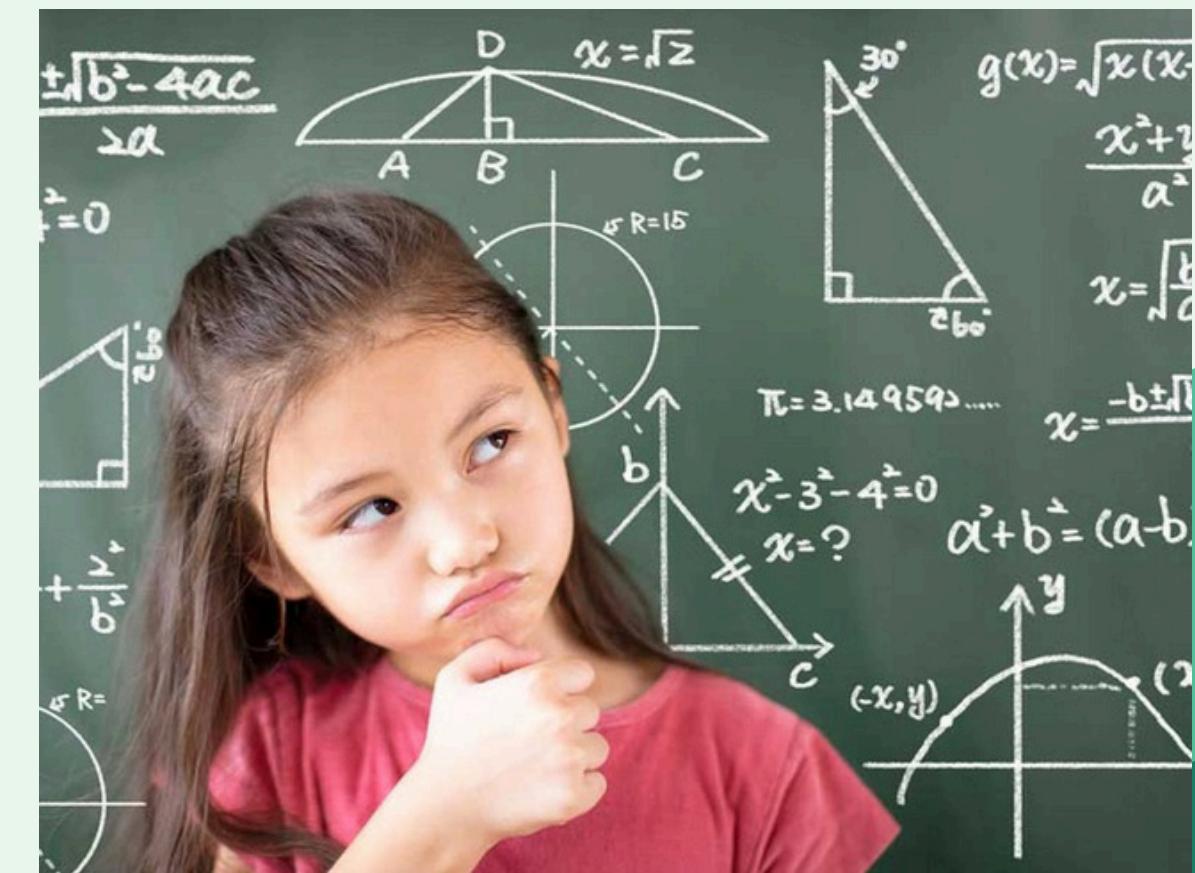


Distribución de salario bruto según máximo nivel de estudios



**Ya vimos las que son, en teoría,
las variables mas significativas...**

**¿Será posible hallar un modelo
que respalte nuestras
suposiciones?**



MODELO INICIAL

MODELO | NVARIABLES

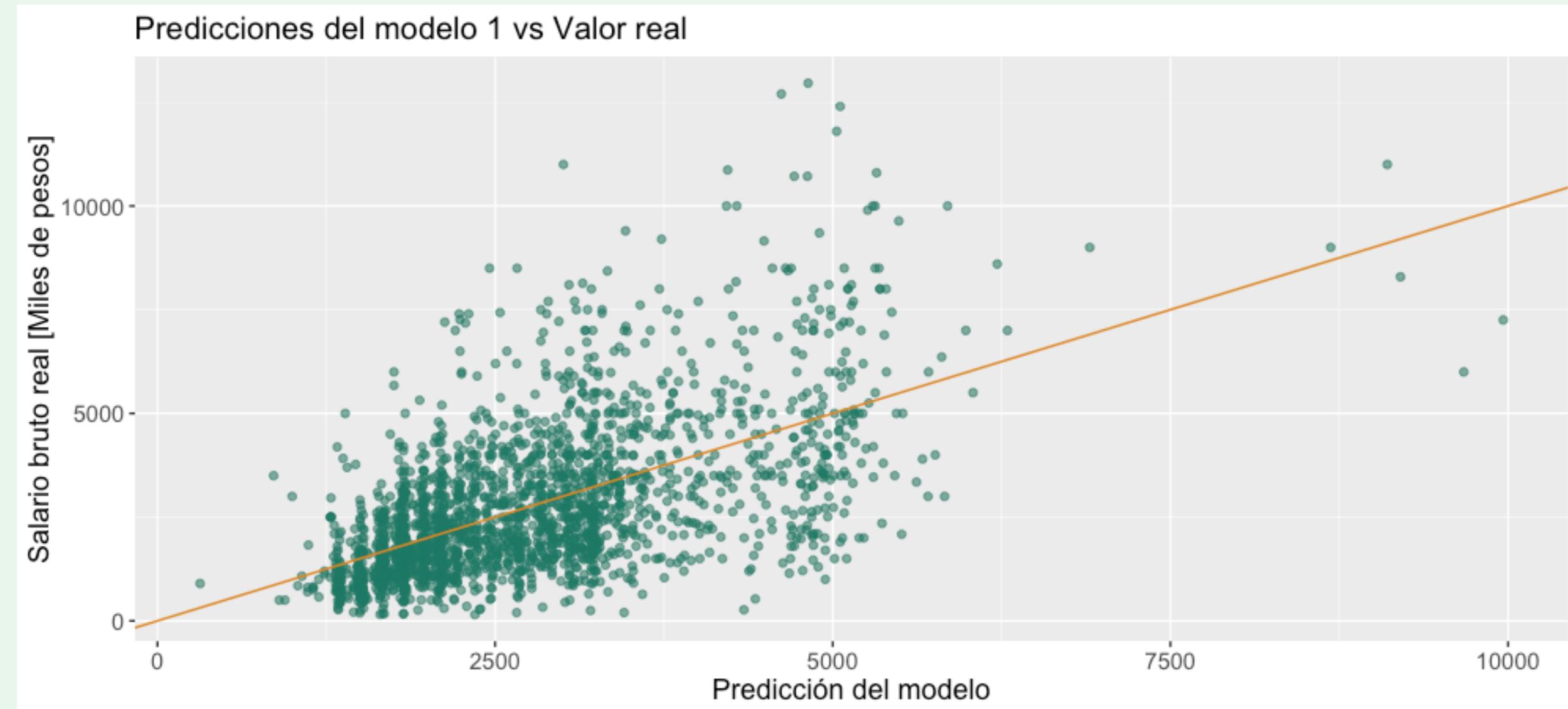
1 | 4

RMSE: 1445

ADJUSTED R-SQUARED:
0.3602

Código en R:

```
lm(data=df_nuevo, X_sal ~  
(poly(anos_de_experiencia,2,raw=TRUE)+cuantas_personas_tenes_a_cargo+tengo_edad)  
*sueldo_dolarizado)
```

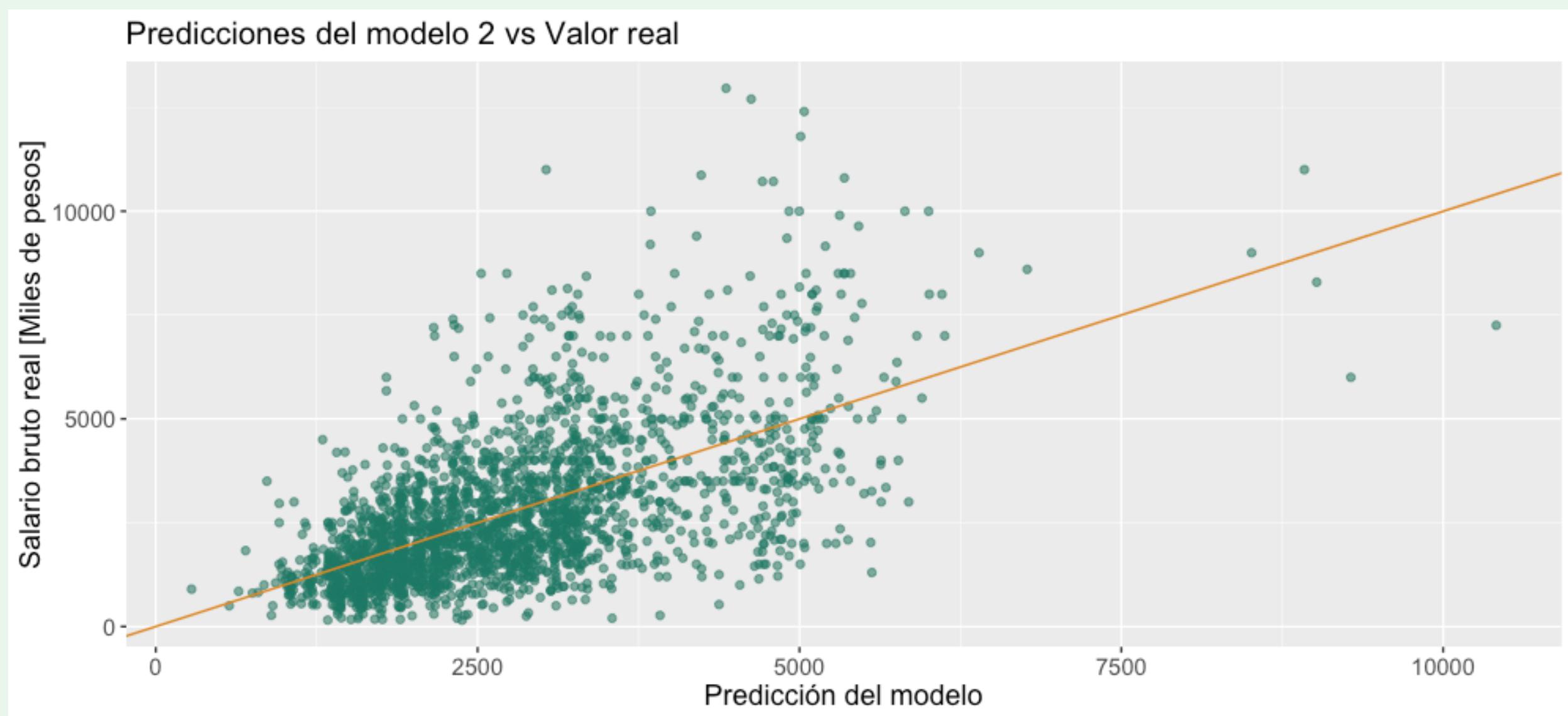


COMPARATIVA DE MODELOS

MODELO	N VARIABLES
1	4
2	5

MODELO 1 VS MODELO 2
 $p\text{-valor} = 1.249e-12$

Modelo 2 = Modelo 1 + MAXIMO NIVEL DE ESTUDIOS

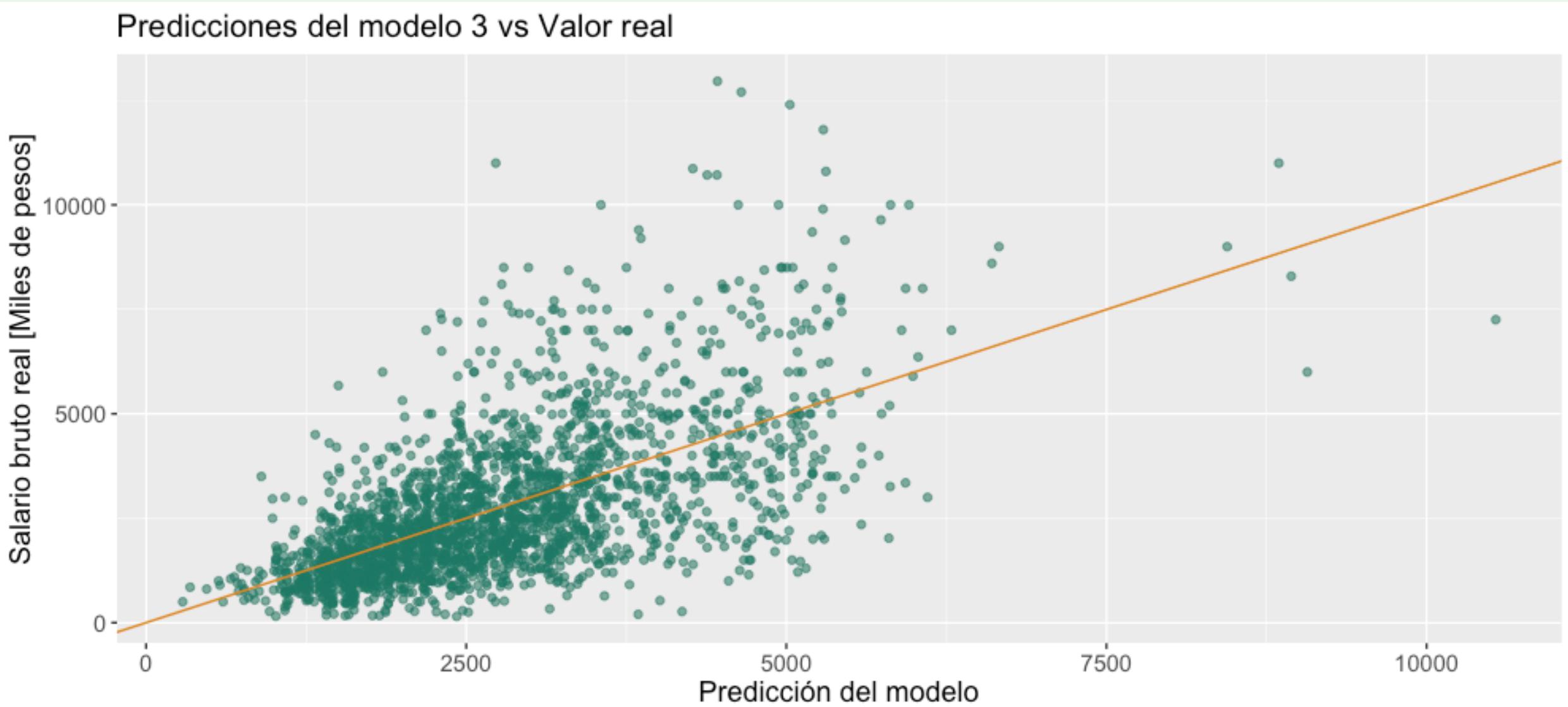


COMPARACION DE MODELOS

MODELO	N VARIABLES
1	4
2	5
3	6

MODELO 2 VS MODELO 3
p-valor = 8.977e-14

Modelo 3 = Modelo 2 + CANTIDAD DE PERSONAS EN TU
ORGANIZACION

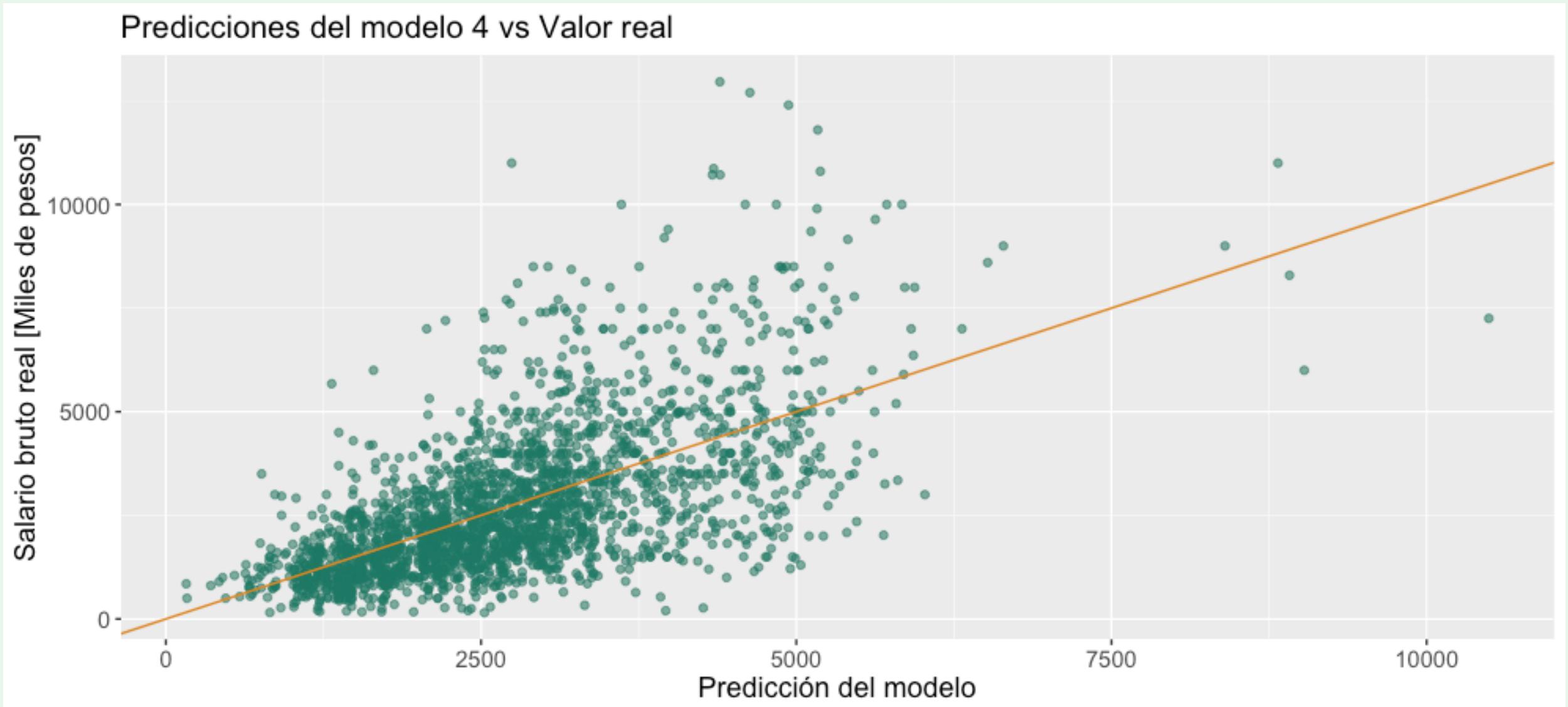


COMPARACION DE MODELOS

MODELO	N VARIABLES
1	4
2	5
3	6
4	7

MODELO 3 VS MODELO 4
p-valor = 3.484e-05

Modelo 4 = Modelo 3 + SENIORITY



MODELO FINAL

MODELO NVARIABLES

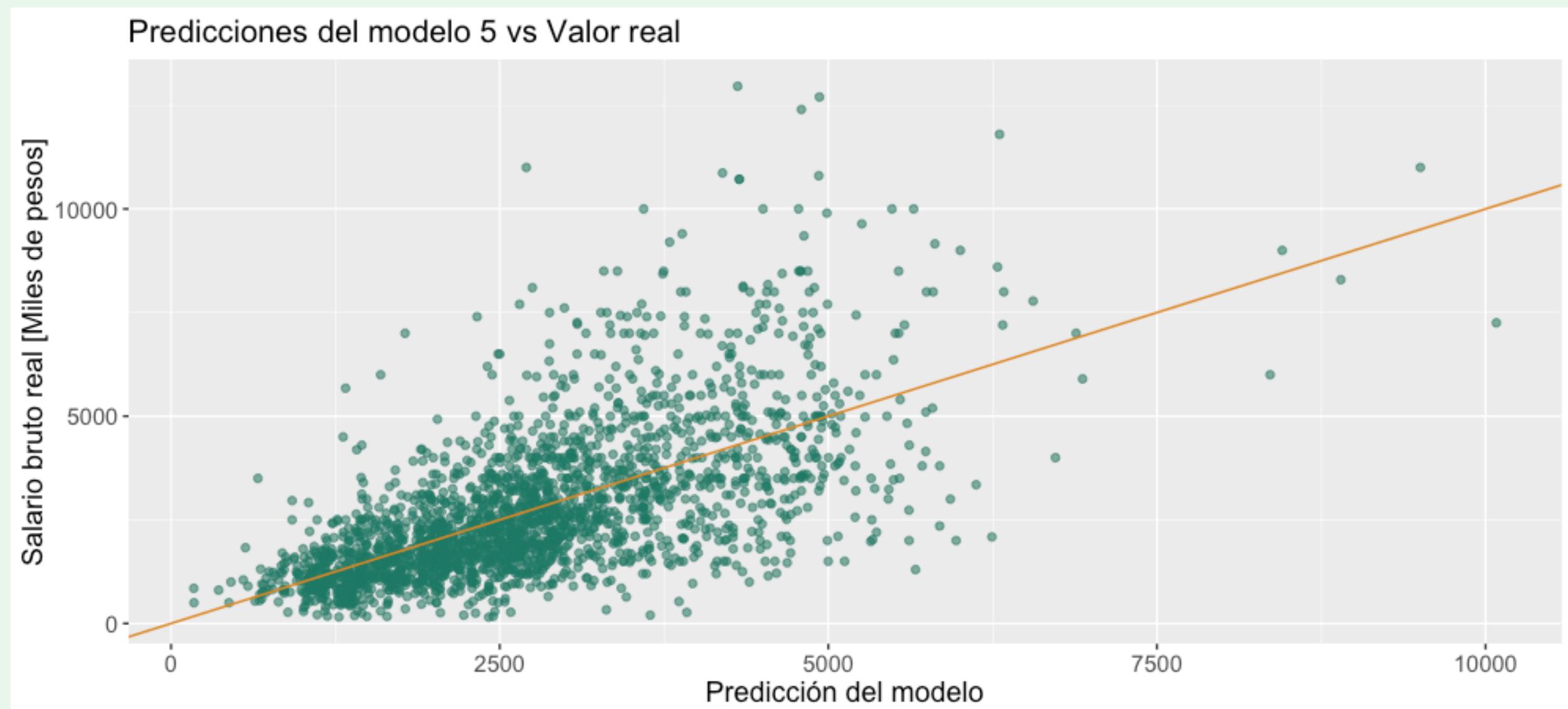
1	4
2	5
3	6
4	7
5	8

RMSE: 1368

ADJUSTED R-SQUARED:
0.4214

MODELO 4 VS MODELO FINAL
p-valor < 2.2e-16

Modelo 5 = Modelo 4 + RECIBIS ALGUN TIPO DE BONO

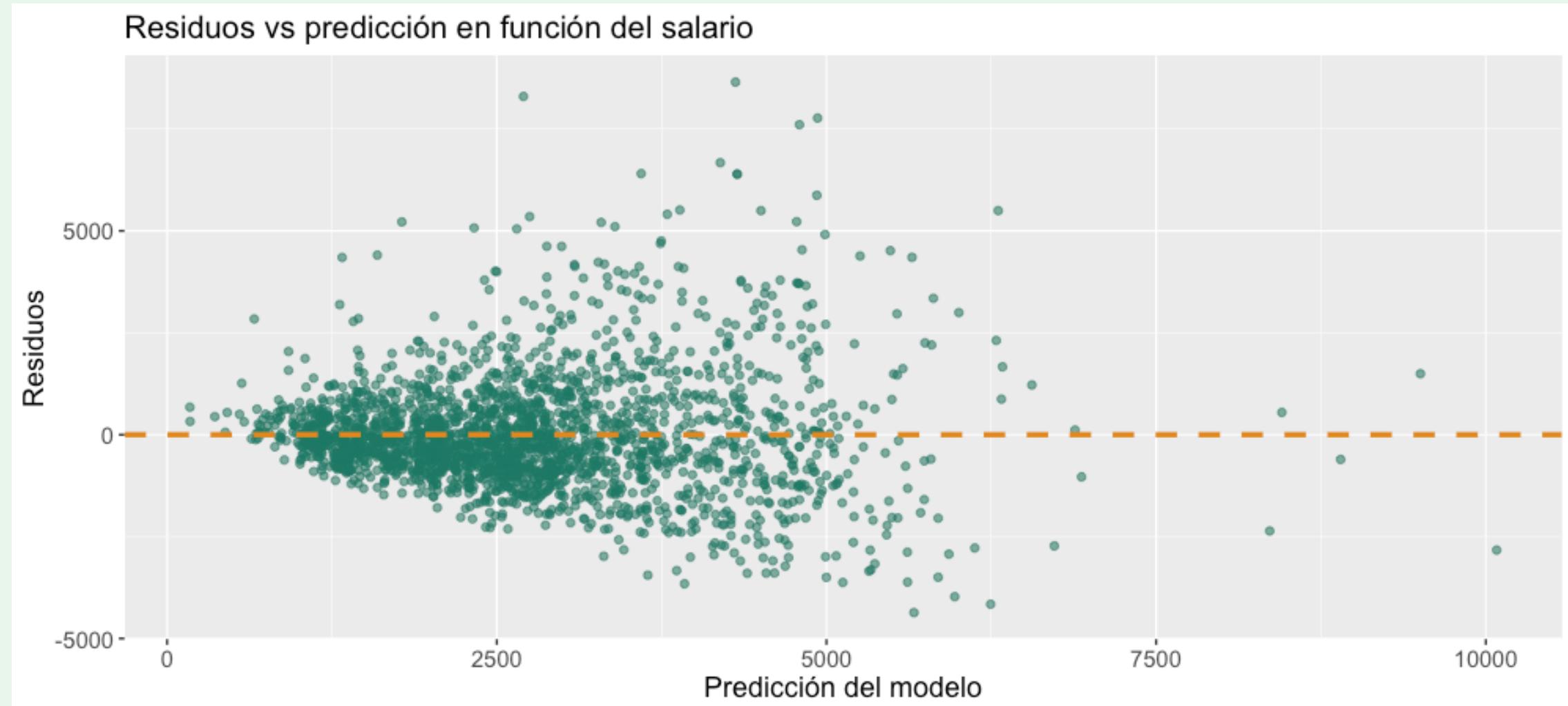


MODELO FINAL

Código en R:

```
lm(data=df_nuevo,  
X_sal ~poly(anos_de_experiencia,2, raw=TRUE)+cuantas_personas_tenes_a_cargo+tengo_edad)*sueldo_dolarizado+  
factor(maximo_nivel_de_estudios)+  
factor(cantidad_de_personas_en_tu_organizacion)+  
factor(seniority)+  
factor(recibis_algun_tipo_de_bono))
```

Parametros	Coefficientes
Intercept	687
poly(experiencia, 2)1	85
poly(experiencia, 2)2	-1,7
personas_a_cargo	39
edad	-2,4
sueldo_dolarizado (TRUE)	1005
nivel_de_estudios (Terciario)	75
nivel_de_estudios (Universitario)	388
nivel_de_estudios (Posgrado/Especializacion)	416
nivel_de_estudios (Maestria)	1005
personas_en_organizacion (51 a 1000)	267
personas_en_organizacion (mas de 1000)	364
seniority (Semi-Senior)	397
seniority (Senior)	611
recibis_algun_bono (Menos de 1 sueldo)	45
recibis_algun_bono (1 sueldo)	435
recibis_algun_bono (1 a 3 sueldos)	756
recibis_algun_bono (mas de 3 sueldos)	1443
poly(experiencia, 2)1:sueldo_dolarizado (TRUE)	248
poly(experiencia, 2)2:sueldo_dolarizado (TRUE)	-5,4
personas_a_cargo:sueldo_dolarizado (TRUE)	-7,7
edad:sueldo_dolarizado (TRUE)	-44



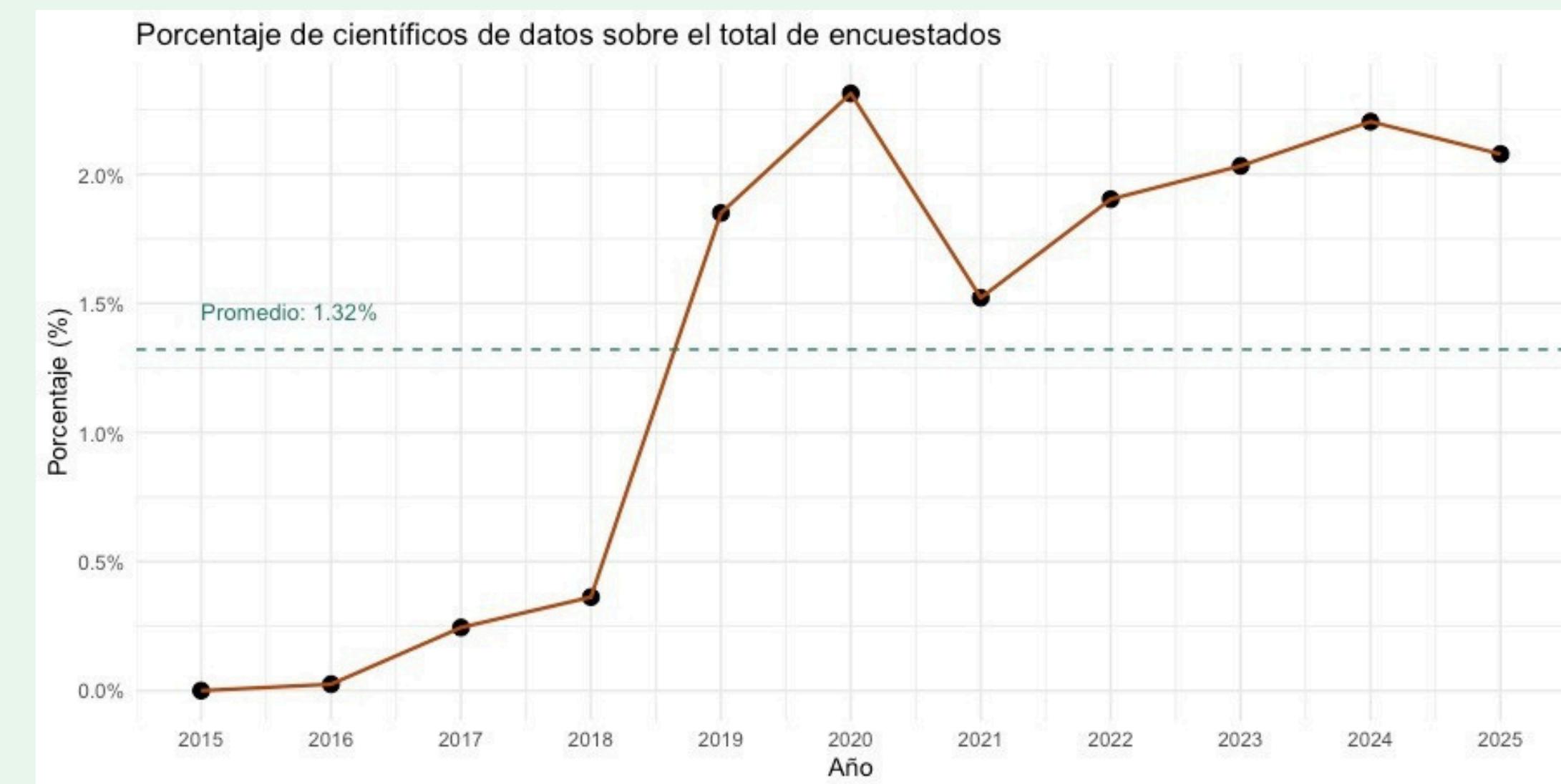
APARTADO

¿QUÉ PASA CON DATA SCIENCE?

Tomamos datasets históricos del blog de Sysarmy para hacer un análisis de la progresión de la cantidad de científicos de datos y sus salarios desde el 2015 al 2025.

AÑO	CIENTÍFICOS DE DATOS	AÑO	TOTAL DE ENCUESTADOS
2015	0	2015	1014
2016	1	2016	4001
2017	12	2017	4907
2018	7	2018	1931
2019	57	2019	3081
2020	141	2020	6095
2021	98	2021	6440
2022	102	2022	5358
2023	118	2023	5805
2024	121	2024	5490
2025	108	2025	5196

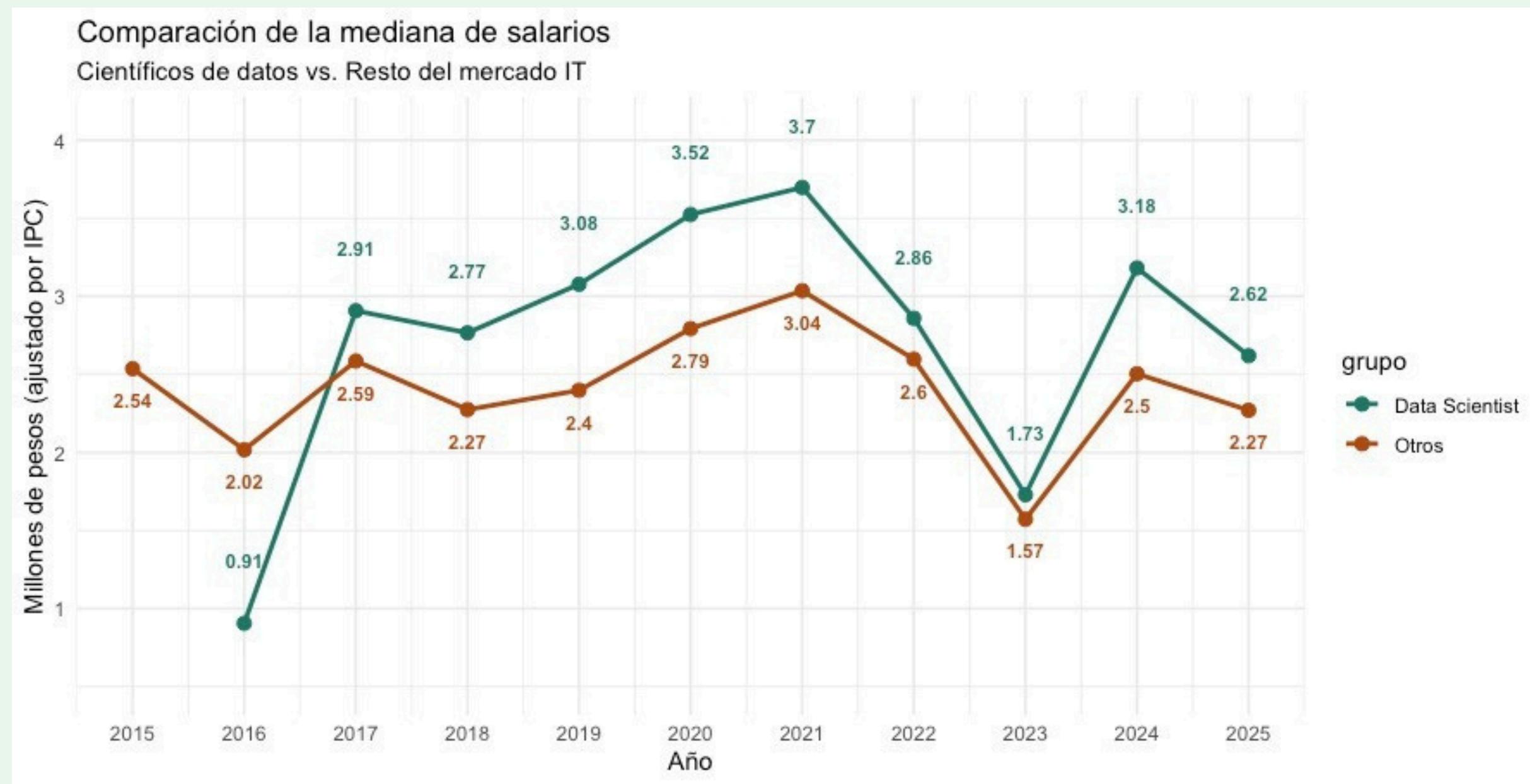
De los 49.318 encuestados entre 2015 al 2025, solamente 765 (el 1,55%) trabajan como científicos de datos. Si bien la tendencia es al alza, la cantidad podría considerarse marginal.



APARTADO

¿QUÉ PASA CON DATA SCIENCE?

Realizamos una comparativa de la mediana de los salarios de los científicos de datos frente a la mediana del resto de los encuestados, ajustando los salarios por el IPC interanual del INDEC.



CONCLUSIONES

EL MODELO FINAL EXPLICA EL 42% DE LA VARIABILIDAD DEL SALARIO, LO QUE REFLEJA LA ALTA DISPERSIÓN SALARIAL DEL SECTOR.

LAS VARIABLES MÁS INFLUYENTES FUERON:

- AÑOS DE EXPERIENCIA
- PERSONAS A CARGO
- SUELDO DOLARIZADO
- SENIORITY
- NIVEL EDUCATIVO
- TAMAÑO DE LA ORGANIZACIÓN
- TIPO DE BONO RECIBIDO

VARIABLES COMO ANTIGÜEDAD O AÑOS EN EL PUESTO ACTUAL MOSTRARON MENOR INFLUENCIA. EL MODELO TIENDE A SUBESTIMAR LOS SALARIOS MÁS ALTOS Y LOS RESIDUOS AUMENTAN CON EL NIVEL SALARIAL.

EL MODELO ES ÚTIL PARA IDENTIFICAR PATRONES ESTRUCTURALES DEL SALARIO EN IT, AUNQUE SU CAPACIDAD PREDICTIVA INDIVIDUAL ES LIMITADA DEBIDO A LA GRAN VARIABILIDAD ENTRE CASOS.

¡MUCHAS GRACIAS POR SU ATENCIÓN!



LARADIAZLM@GMAIL.COM - GINO_MARIANI@OUTLOOK.COM

ISAIAS.FRELLATT@GMAIL.COM - MARTINTOMASNAHUEL@GMAIL.COM