

Trabajo de simulación 3

Rinaldi, Flavio Sabato, Ángel Shifman, Iván Zárate, Marcelo

Índice de Contenidos

| | |
|---|-----------|
| 1 Consigna | 1 |
| 2 Nuestra resolución: <i>Optimización salarial</i> | 2 |
| 2.1 Inspección y limpieza de datos | 2 |
| 2.2 El perfil del top 10% | 6 |
| 2.3 El “salto salarial” en la trayectoria profesional | 7 |
| 2.4 Evolución del salario a través de los años | 7 |
| 2.5 Visualizaciones comparativas | 8 |
| 2.6 La combinación ganadora | 10 |
| 3 Machine Learning: <i>Random Forest Regressor</i> | 11 |
| 3.1 Importancia de las variables | 11 |
| 3.2 Visualización: predicciones vs. realidad | 13 |
| 3.3 Cálculo de la importancia de las variables | 14 |
| 3.4 Análisis de residuos | 17 |
| 3.5 Conclusiones del modelo de Machine Learning | 17 |
| 4 Conclusiones | 18 |

List of Figures

1 Consigna

Para el último trabajo de simulación, vamos a hacer un análisis de datos a partir de la encuesta de salarios de Sysarmy del primer trimestre de 2025 (si están publicados los del segundo trimestre o posterior, podemos usar esos, pero a agosto de 2025 todavía no estaban).

Los datos oficiales se encuentran [acá](#). Por si en algún momento el enlace se cae o cambia de ubicación, una versión alojada en mi drive puede encontrarse [acá](#).

Las consignas de este trabajo no son tan dirigidas como las de los trabajos anteriores, pues en el análisis de datos, siempre hay libertad y margen para la creatividad y la producción personal. Sin embargo, les compartimos algunas pautas de lo que debe tener, como mínimo, este trabajo.

Pautas generales y **OBLIGATORIAS** para la aprobación de la entrega:

- Debe replicarse, como mínimo, un análisis similar al aquí presentado, para estos datos (SysArmy 2025 o cualquier otro dataset que sea de su interés). Por “replicar” nos referimos a que el análisis debe incluir: inspección y limpieza de los datos, descripción y visualización, o estudio de alguna variable de interés a partir de alguna hipótesis o conjectura.

- Debe escribirse en formato “informe”, es decir, no se trata de exhibir código y gráficos, sino de explicar qué se observa y por qué es relevante observar eso. El informe es requisito **excluyente**. No se aprueba el trabajo de simulación sin él. Este informe breve debe entregarse en pdf, en esta entrega **NO** se evalúa el *colab*, sino el reporte. No tiene que ser largo, al contrario, tiene que ser de calidad. Como dice el dicho: lo bueno -si breve- dos veces bueno...
- El trabajo debe contener, como mínimo, **una conjetura que sea sometida a prueba y de la que se exhiba alguna conclusión fundamentada**, como se hizo en el caso de los datos de 2020 para el salario medio bruto por género y para hombres y mujeres con nivel universitario completo. Por ejemplo, frente a la pregunta de si el salario medio de mujeres y hombres es igual, podríamos poner en práctica lo que estudiamos sobre convergencia para, de alguna forma, darnos una idea de cuán probable es observar lo que efectivamente estamos observando. Este es un “coqueteo” con la estadística inferencial, que no estudiamos formalmente en la materia, pero que es válido comenzar a encarar con todo lo que hemos estudiado. Esta conjetura puede hacerse con datos propios, si es que eligen trabajar con otro dataset.

El resto de la producción queda a criterio de los grupos. Esperamos que haya un interés genuino en tratar de extraer información a partir de estos datos. ¡Muchos éxitos!

PD: El formato “informe” puede ser cambiado por el formato “póster/infografía” si es que así prefieren.

Source: [Article Notebook](#)

2 Nuestra resolución: *Optimización salarial*

Para realizar nuestro análisis tomamos la perspectiva de un profesional del sector IT que necesita transformar su carrera por alguna razón. Sabemos que esta industria se caracteriza por un nivel de rotación alto y los recorridos profesionales suelen ser muy variados. La pregunta central de nuestro trabajo es *¿cómo puedo maximizar mi salario?*, es decir, *¿cuál es la combinación ganadora que se podría obtener como insight de este dataset?* Los resultados obtenidos podrían ser de utilidad tanto para el *junior*, que necesita alguna “brújula” para ver por dónde orientarse dentro del mercado laboral, como para el *senior* que quizás se siente estancado y necesite reorientar su desarrollo profesional dentro del sector.

Source: [Article Notebook](#)

Source: [Article Notebook](#)

2.1 Inspección y limpieza de datos

Comenzamos importando las librerías necesarias y cargando el dataset. Luego inspeccionamos los datos para ver qué variables tenemos disponibles y qué tipo de datos contienen.

Source: [Article Notebook](#)

```
*** Archivo cargado exitosamente
Dimensiones: 5196 filas x 50 columnas
```

Source: [Article Notebook](#)

Luego procesamos los datos para dejarlos en un formato adecuado para el análisis. Esto incluye manejar valores faltantes, convertir variables categóricas en numéricas (si es necesario) y asegurarnos de que todas las variables estén en el tipo de dato correcto.

Source: [Article Notebook](#)

[PASO 1] Calculando salarios en USD con TC individual...

- * 927 registros con TC reportado individual
- * TC mediano de la encuesta: \$1,060.00
- * 4269 registros usando TC mediano

Dataset preparado: (5196, 53)

Registros con salario USD: 5196

Source: [Article Notebook](#)

Luego de todo el procesamiento “de rigor” que exige un análisis de datos, nos queda un dataset limpio y listo para analizar. En este punto comenzamos con la búsqueda de “la combinación ganadora”, es decir, la combinación de variables que maximiza el salario.

Source: [Article Notebook](#)

[INFO] Variables analizadas: 16

[INFO] Registros validos: 5196

Source: [Article Notebook](#)

En esta parte realizamos un análisis por variable individual. Tratamos de ir de a poco analizando la información que nos permita responder a la pregunta central del trabajo.

Source: [Article Notebook](#)

[1] UBICACION GEOGRAFICA:

| | Media | Mediana | Q75 | Max | Cantidad |
|----------------|---------|---------|---------|------------|----------|
| es_caba | | | | | |
| Resto del pais | 5324.79 | 2124.29 | 3275.33 | 3885135.14 | 5196 |

GANADOR: Resto del pais

Mediana: \$2,124 USD

[2] TIPO DE CONTRATO:

| | Media | Mediana | \ |
|--|----------|------------|---|
| tipo_contrato | | | |
| Contractor | 11316.76 | 2641.51 | |
| Staff (planta permanente) | 4477.06 | 2169.81 | |
| Participación societaria en una cooperativa | 2206.28 | 2075.47 | |
| Tercerizado (trabajo a través de consultora o a... | 1869.14 | 1600.73 | |
| Freelance | 1895.63 | 1391.51 | |
| | Q75 | Max | \ |
| tipo_contrato | | | |
| Contractor | 4379.25 | 3500000.00 | |
| Staff (planta permanente) | 3154.64 | 3885135.14 | |
| Participación societaria en una cooperativa | 2650.60 | 4245.28 | |
| Tercerizado (trabajo a través de consultora o a... | 2264.15 | 9433.96 | |
| Freelance | 2405.66 | 9433.96 | |
| | Cantidad | | |
| tipo_contrato | | | |

| | |
|---|------|
| Contractor | 859 |
| Staff (planta permanente) | 3768 |
| Participación societaria en una cooperativa | 29 |
| Tercerizado (trabajo a través de consultora o a...) | 403 |
| Freelance | 137 |

GANADOR: Contractor

Mediana: \$2,642 USD

[3] DEDICACION:

| | Media | Mediana | Q75 | Max | Cantidad |
|-------------------|---------|---------|---------|------------|----------|
| dedicacion | | | | | |
| Full-Time | 5507.10 | 2169.81 | 3301.89 | 3885135.14 | 4974 |
| Part-Time | 1240.13 | 849.06 | 1509.43 | 6603.77 | 222 |

GANADOR: Full-Time

Mediana: \$2,170 USD

[4] MODALIDAD DE TRABAJO:

| | Media | Mediana | Q75 | Max | Cantidad |
|-------------------------------|---------|---------|---------|------------|----------|
| modalidad | | | | | |
| Híbrido (presencial y remoto) | 4282.31 | 2207.55 | 3286.79 | 3500000.00 | 2056 |
| 100% remoto | 6650.02 | 2169.81 | 3396.23 | 3885135.14 | 2733 |
| 100% presencial | 1692.13 | 1415.09 | 2125.50 | 6477.50 | 407 |

GANADOR: Híbrido (presencial y remoto)

Mediana: \$2,208 USD

[5] SENIORITY:

| | Media | Mediana | Q75 | Max | Cantidad |
|------------------|---------|---------|---------|------------|----------|
| seniority | | | | | |
| Senior | 7332.15 | 2830.19 | 4050.30 | 3885135.14 | 2648 |
| Semi-Senior | 3716.79 | 1886.79 | 2641.51 | 2600000.00 | 1634 |
| Junior | 2383.88 | 1224.81 | 1698.11 | 909118.18 | 914 |

GANADOR: Senior

Mediana: \$2,830 USD

[6] PUESTO (Top 10):

| | Media | Mediana | Q75 | Max | Cantidad |
|--------------------------|---------|---------|---------|----------|----------|
| puesto | | | | | |
| Smart contracts engineer | 7264.15 | 7264.15 | 7264.15 | 7264.15 | 1 |
| embedded engineer | 6750.00 | 6750.00 | 6750.00 | 6750.00 | 1 |
| Engineer | 5566.04 | 5566.04 | 5566.04 | 5566.04 | 1 |
| Staff Engineer | 4961.24 | 4961.24 | 5813.95 | 6666.67 | 2 |
| VP / C-Level | 4715.65 | 4205.97 | 6376.50 | 10849.06 | 40 |
| AI Engineer | 4079.94 | 4096.19 | 4980.28 | 8867.92 | 7 |
| Architect | 4117.63 | 3773.58 | 5188.68 | 12549.02 | 125 |
| CIO | 3773.58 | 3773.58 | 3773.58 | 3773.58 | 1 |
| GeneXus Analyst | 3415.09 | 3415.09 | 3415.09 | 3415.09 | 1 |
| Technical Leader | 3580.64 | 3301.89 | 4525.02 | 11981.13 | 367 |

GANADOR: Smart contracts engineer
 Mediana: \$7,264 USD

[7] FORMA DE PAGO:

| forma_pago | Media | Mediana | \ |
|---|----------|------------|---|
| Cobro todo el salario en dólares | 11217.87 | 3301.89 | |
| Cobro parte del salario en dólares | 13133.81 | 2340.42 | |
| Mi sueldo está dolarizado (pero cobro en moneda...) | 5355.33 | 2075.47 | |
| | Q75 | Max | \ |
| forma_pago | | | |
| Cobro todo el salario en dólares | 4866.50 | 3500000.00 | |
| Cobro parte del salario en dólares | 3487.89 | 3885135.14 | |
| Mi sueldo está dolarizado (pero cobro en moneda...) | 3292.92 | 909118.18 | |
| | Cantidad | | |
| forma_pago | | | |
| Cobro todo el salario en dólares | 811 | | |
| Cobro parte del salario en dólares | 714 | | |
| Mi sueldo está dolarizado (pero cobro en moneda...) | 323 | | |

GANADOR: Cobro todo el salario en dólares

Mediana: \$3,302 USD

[8] TAMANO DE EMPRESA:

| tamano_empresa | Media | Mediana | Q75 | Max | Cantidad |
|--------------------------|----------|---------|---------|------------|----------|
| De 2001a 5000 personas | 2912.02 | 2641.51 | 3720.12 | 11132.08 | 364 |
| De 5001 a 10000 personas | 2927.20 | 2547.17 | 3438.03 | 47142.86 | 247 |
| Más de 10000 personas | 2817.12 | 2547.17 | 3492.45 | 12000.00 | 589 |
| De 1001 a 2000 personas | 2874.95 | 2358.49 | 3584.91 | 10362.69 | 361 |
| De 501 a 1000 personas | 10830.66 | 2358.49 | 3590.61 | 3500000.00 | 432 |
| De 201 a 500 personas | 2633.27 | 2169.81 | 3301.89 | 12226.42 | 675 |
| De 101 a 200 personas | 12609.64 | 2169.81 | 3301.89 | 3500000.00 | 611 |
| De 51 a 100 personas | 8008.99 | 1886.79 | 2995.27 | 3885135.14 | 687 |
| 1 (solamente yo) | 2685.63 | 1839.62 | 3820.75 | 10000.00 | 44 |
| De 11 a 50 personas | 3222.85 | 1720.10 | 2667.92 | 909118.18 | 880 |
| De 2 a 10 personas | 1863.54 | 1379.25 | 2172.89 | 10109.43 | 306 |

GANADOR: De 2001a 5000 personas

Mediana: \$2,642 USD

[9] PERSONAS A CARGO:

| tiene_equipo | Media | Mediana | Q75 | Max | Cantidad |
|--------------|---------|---------|---------|------------|----------|
| Con equipo | 8443.76 | 2830.19 | 4084.91 | 3500000.00 | 1498 |
| Sin equipo | 4061.35 | 1901.51 | 2869.27 | 3885135.14 | 3698 |

GANADOR: Con equipo

Mediana: \$2,830 USD

Source: [Article Notebook](#)

2.2 El perfil del top 10%

Si bien nuestro objetivo es encontrar la combinación ganadora, creemos que un buen punto de partida es analizar el perfil del top 10% con el mejor salario. La idea es entender qué características tienen en común los profesionales que están en este grupo y ver si podemos extraer alguna conclusión útil para nuestro análisis.

Source: [Article Notebook](#)

*** Salario minimo Top 10%: \$4,717 USD
*** Trabajadores en Top 10%: 530

[Distribucion Top 10%]

Ubicacion:

```
es_caba
Resto del pais    100.0
Name: proportion, dtype: float64
```

Tipo de Contrato:

```
tipo_contrato
Staff (planta permanente)      59.6
Contractor                      35.1
Tercerizado (trabajo a través de consultora o agencia)  2.8
Freelance                         2.5
Name: proportion, dtype: float64
```

Seniority:

```
seniority
Senior        88.1
Semi-Senior   10.4
Junior         1.5
Name: proportion, dtype: float64
```

Top 5 Puestos:

```
puesto
Developer          129
Manager / Director 116
Technical Leader    81
SysAdmin / DevOps / SRE 49
Architect           39
Name: count, dtype: int64
```

Experiencia promedio: 14.3 años
Edad promedio: 38.8 años
Antiguedad promedio: 4.6 años

Source: [Article Notebook](#)

2.3 El “salto salarial” en la trayectoria profesional

De a poco vemos emerger la información relevante. Una vez que analizamos las variables anteriores, podemos abordar la cuestión de la “trayectoria profesional” para encontrar el momento del **salto salarial**, *¿suele tardar en llegar?, ¿llega en algún momento o tiende más a bien a ser estable?*. Esto es lo que intentamos responder a continuación.

Source: [Article Notebook](#)

EVOLUCIÓN SALARIAL POR EXPERIENCIA:

| rango_exp | Cantidad | Mínimo | Q25 | Mediana | Media | Q75 | \ |
|------------------------|----------|--------|---------|---------|---------|---------|---|
| 0-2 años (Junior) | 754 | 145.94 | 935.49 | 1273.16 | 2637.31 | 1756.84 | |
| 3-5 años (Semi-Senior) | 1634 | 1.00 | 1389.47 | 1886.79 | 3716.79 | 2641.51 | |
| 6-8 años (Senior) | 685 | 1.25 | 1840.20 | 2688.68 | 8085.04 | 3653.85 | |
| 9-12 años (Senior+) | 608 | 1.37 | 1886.79 | 2754.73 | 3124.81 | 3920.56 | |
| 12+ años (Expert) | 1355 | 145.99 | 1981.13 | 2978.77 | 8839.40 | 4339.62 | |

| rango_exp | Máximo | Desv_Est | Crecimiento_% | Acumulado_% |
|------------------------|------------|-----------|---------------|-------------|
| 0-2 años (Junior) | 909118.18 | 33065.23 | NaN | 0.000000 |
| 3-5 años (Semi-Senior) | 2600000.00 | 64278.19 | 48.197399 | 48.197399 |
| 6-8 años (Senior) | 3500000.00 | 133633.65 | 42.500225 | 111.181627 |
| 9-12 años (Senior+) | 12679.25 | 1822.43 | 2.456596 | 116.369506 |
| 12+ años (Expert) | 3885135.14 | 141888.30 | 8.132920 | 133.966666 |

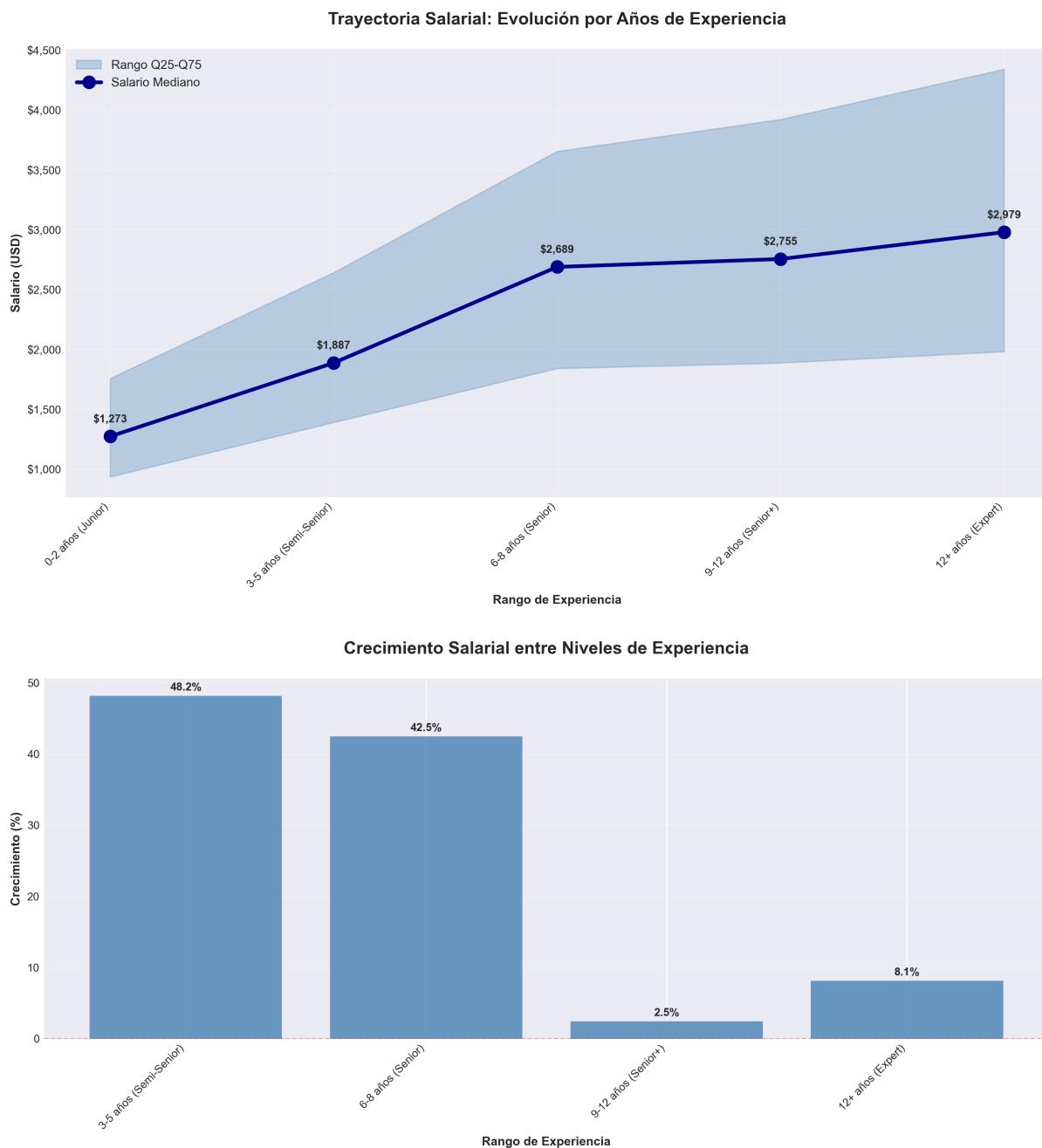
MAYOR SALTO SALARIAL: 3-5 años (Semi-Senior) (+48.2%)

Source: [Article Notebook](#)

2.4 Evolución del salario a través de los años

Podemos visualizar la evolución del salario mediano a lo largo de los años de experiencia para observar tendencias y patrones en el crecimiento salarial. También podemos analizar la distribución de salarios en diferentes rangos de experiencia para ver cómo varía el salario entre profesionales con diferentes niveles.

Source: [Article Notebook](#)

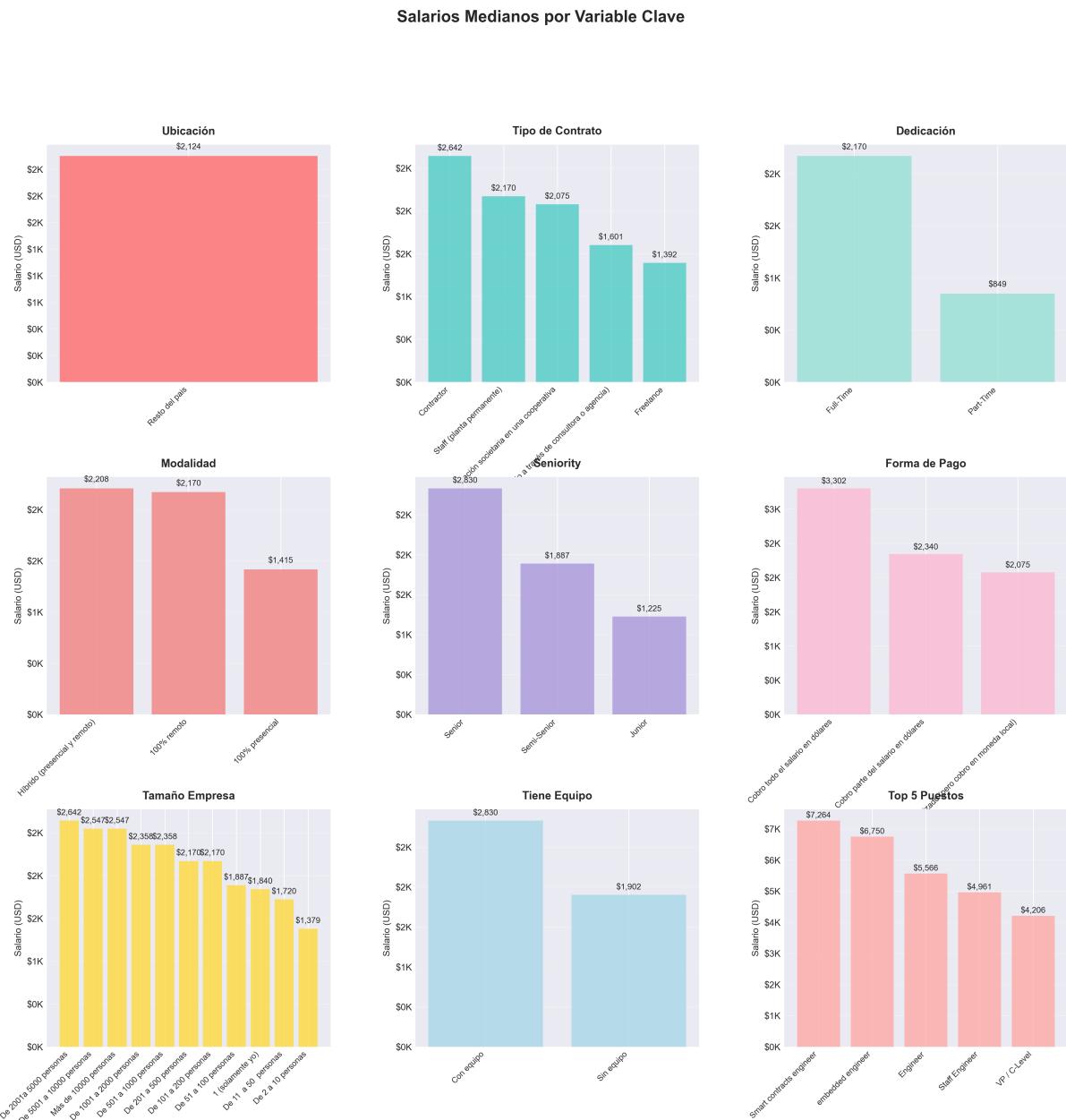


Source: [Article Notebook](#)

2.5 Visualizaciones comparativas

Ahora bien, una vez que tenemos toda esta información, podemos intentar combinar las variables para encontrar la combinación ganadora que maximiza el salario.

Source: [Article Notebook](#)



[TOP 10 COMBINACIONES]

Top 10 Combinaciones (Ubicación + Contrato + Dedicación):

| es_caba | tipo_contrato | dedicacion | Media \ |
|----------------|--|------------|----------|
| Resto del pais | Contractor | Full-Time | 12040.33 |
| | Staff (planta permanente) | Full-Time | 4563.36 |
| | Participación societaria en una cooperativa | Full-Time | 2268.23 |
| | Freelance | Full-Time | 2309.57 |
| | Tercerizado (trabajo a través de consultora o a... | Full-Time | 1912.36 |
| | Staff (planta permanente) | Part-Time | 1343.89 |
| | Contractor | Part-Time | 1324.06 |
| | Tercerizado (trabajo a través de consultora o a... | Part-Time | 887.62 |
| | Freelance | Part-Time | 1049.34 |

Mediana \

| es_caba | tipo_contrato | dedicacion | |
|----------------|---|------------|---------|
| Resto del pais | Contractor | Full-Time | 2830.19 |
| | Staff (planta permanente) | Full-Time | 2169.81 |
| | Participación societaria en una cooperativa | Full-Time | 2124.69 |
| | Freelance | Full-Time | 1836.32 |
| | Tercerizado (trabajo a través de consultora o a...) | Full-Time | 1608.49 |
| | Staff (planta permanente) | Part-Time | 1008.90 |
| | Contractor | Part-Time | 825.13 |
| | Tercerizado (trabajo a través de consultora o a...) | Part-Time | 801.89 |
| | Freelance | Part-Time | 745.28 |
| | | Cantidad | |
| es_caba | tipo_contrato | dedicacion | |
| Resto del pais | Contractor | Full-Time | 801 |
| | Staff (planta permanente) | Full-Time | 3667 |
| | Participación societaria en una cooperativa | Full-Time | 28 |
| | Freelance | Full-Time | 92 |
| | Tercerizado (trabajo a través de consultora o a...) | Full-Time | 386 |
| | Staff (planta permanente) | Part-Time | 101 |
| | Contractor | Part-Time | 58 |
| | Tercerizado (trabajo a través de consultora o a...) | Part-Time | 17 |
| | Freelance | Part-Time | 45 |

Source: [Article Notebook](#)

2.6 La combinación ganadora

Después de todo este análisis, podemos finalmente presentar la combinación ganadora que maximiza el salario. Esta combinación se basa en las variables que hemos analizado y las conclusiones que hemos extraído a lo largo de este “análisis exploratorio de datos” (EDA).

Source: [Article Notebook](#)

PERFIL DE LOS TRABAJADORES MEJOR PAGOS

- [1] UBICACION: Resto del pais
Mediana: \$2,124 USD
- [2] TIPO CONTRATO: Contractor
Mediana: \$2,642 USD
- [3] DEDICACION: Full-Time
Mediana: \$2,170 USD
- [4] MODALIDAD: Híbrido (presencial y remoto)
Mediana: \$2,208 USD
- [5] SENIORITY: Senior
Mediana: \$2,830 USD
- [6] PUESTO: Smart contracts engineer
Mediana: \$7,264 USD

[7] FORMA PAGO: Cobro todo el salario en dólares
Mediana: \$3,302 USD

[8] TAMANO EMPRESA: De 2001a 5000 personas
Mediana: \$2,642 USD

[9] LIDERAZGO: Con equipo
Mediana: \$2,830 USD

EVOLUCION ESPERADA DEL SALARIO

0-2 años: \$1,273 USD
3-5 años: \$1,887 USD (+48.2%)
6-8 años: \$2,689 USD (+42.5%)
9-12 años: \$2,755 USD (+2.5%)
12+ años: \$2,979 USD (+8.1%)

RECOMENDACIONES PARA MAXIMIZAR SALARIO

1. Ubicarse en Resto del pais
2. Buscar contratos tipo Contractor
3. Trabajar Full-Time
4. Priorizar modalidad Híbrido (presencial y remoto)
5. Desarrollarse hasta Senior
6. Especializarse en roles como Smart contracts engineer
7. Negociar modalidad "Cobro todo el salario en dólares"
8. Apuntar a empresas De 2001a 5000 personas
9. Desarrollar capacidad de liderazgo (Con equipo)
10. Acumular ~14 años de experiencia

Objetivo: Alcanzar Top 10% (\$4,717+ USD/mes)

Source: [Article Notebook](#)

3 Machine Learning: *Random Forest Regressor*

En esta parte del trabajo, buscamos ir más allá del análisis descriptivo y explorar la importancia relativa de las variables en la predicción del salario utilizando técnicas de Machine Learning. Nuestro objetivo es identificar cuáles son las características más influyentes que determinan el salario de los profesionales en el sector IT y ver si existe una coincidencia entre estas variables y las que hemos identificado en nuestro análisis previo como parte de la “combinación ganadora”. Creemos que al sumar este enfoque, podemos obtener una visión más completa y robusta de los factores que impactan en el salario, validando o complementando nuestras conclusiones anteriores con un análisis basado en datos y modelos predictivos.

Source: [Article Notebook](#)

3.1 Importancia de las variables

En primer lugar, debemos preparar los datos para el modelo de Machine Learning. Esto incluye seleccionar las variables relevantes, manejar valores faltantes y dividir los datos en conjuntos de entrenamiento y prueba. El modelo que vamos a aplicar es un Random Forest Regressor,

muy utilizado para problemas de regresión que puede manejar tanto variables numéricas como categóricas.

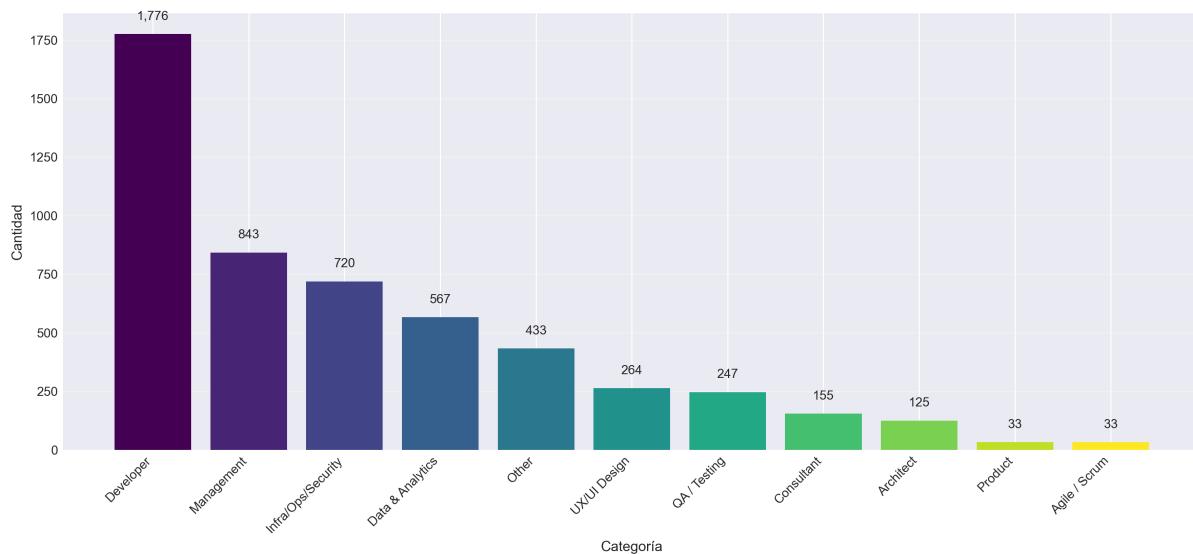
Source: [Article Notebook](#)

[INFO] Categorías de puestos creadas

puesto_categoria

| | |
|---------------------------|------|
| Developer | 1776 |
| Management | 843 |
| Infra/Ops/Security | 720 |
| Data & Analytics | 567 |
| Other | 433 |
| UX/UI Design | 264 |
| QA / Testing | 247 |
| Consultant | 155 |
| Architect | 125 |
| Product | 33 |
| Agile / Scrum | 33 |
| Name: count, dtype: int64 | |

Distribución de Puestos por Categoría



[INFO] Tamaño empresa simplificado:

tamano_empresa

| | |
|---------------------------|------|
| Pequeña (1-100) | 1917 |
| Mediana (101-1000) | 1718 |
| Grande (1000+) | 1561 |
| Name: count, dtype: int64 | |

[INFO] Registros iniciales: 1848

[LIMPIEZA DE DATOS]

Después de eliminar infinitos: 1848 registros

Después de eliminar salarios <= 0: 1848 registros

Rango válido: \$259 - \$12,198 USD

Después de filtrar outliers extremos: 1828 registros

Después de validar rangos de variables: 1827 registros

```
[INFO] Registros finales para modelo: 1827  
[INFO] Variables predictoras: 11  
[INFO] Salario USD - Min: $261 | Max: $12,106  
[INFO] Salario USD - Media: $3,126 | Mediana: $2,642
```

```
[INFO] Features después de encoding: 26  
[VALIDACIÓN] NaN en X: 0  
[VALIDACIÓN] Infinitos en X: 0  
[VALIDACIÓN] NaN en y: 0  
[VALIDACIÓN] Infinitos en y: 0
```

```
[MODELO] Entrenamiento: 1461 | Test: 366  
[MODELO] Entrenado exitosamente
```

MÉTRICAS DEL MODELO

```
MAE Train: $1,129 USD  
MAE Test: $1,097 USD  
RMSE Test: $1,559 USD  
R2 Train: 0.4104 (41.04%)  
R2 Test: 0.3647 (36.47%)
```

El modelo generaliza bien (diferencia R²: 0.046)

VALIDACIÓN CRUZADA (5-FOLD):

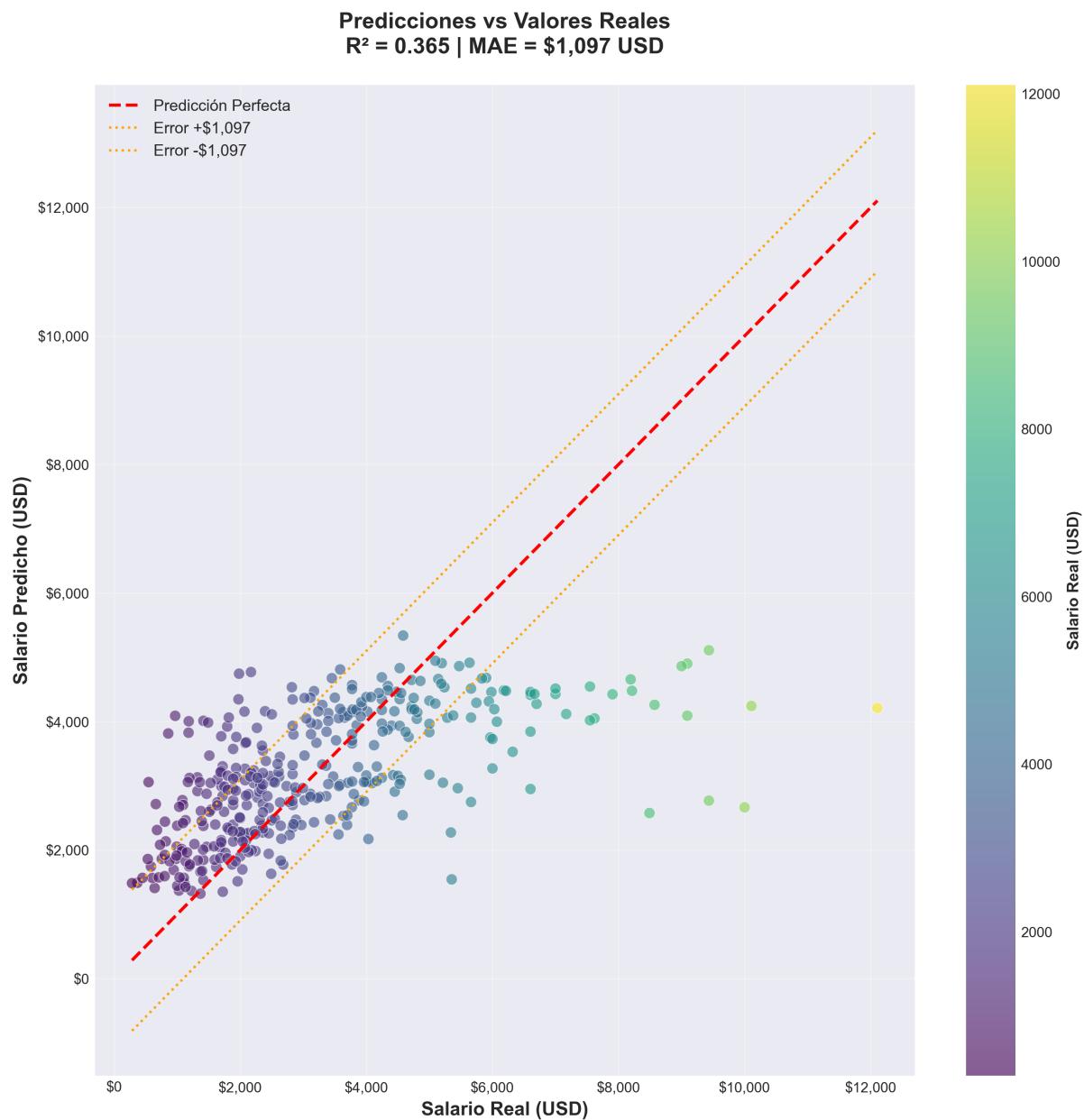
```
R2 promedio: 0.3445 ( $\pm 0.0251$ )  
R2 por fold: ['0.371', '0.369', '0.340', '0.341', '0.301']  
MAE promedio: $1,177 ( $\pm \$56$ )
```

Source: [Article Notebook](#)

3.2 Visualización: predicciones vs. realidad

El modelo entrenado nos permite hacer predicciones sobre los salarios basándonos en las características de los profesionales. Para evaluar el rendimiento del modelo, podemos comparar las predicciones con los valores reales de salario en el conjunto de prueba. Una forma efectiva de visualizar esta comparación es mediante un gráfico de dispersión (scatter plot), donde cada punto representa un profesional, con su salario real en el eje Y y su salario predicho por el modelo en el eje X.

Source: [Article Notebook](#)



Source: [Article Notebook](#)

3.3 Cálculo de la importancia de las variables

En este caso, el modelo nos proporciona una medida de la importancia de cada variable en la predicción del salario. Podemos ordenar estas importancias y visualizarlas para identificar cuáles son las características más influyentes.

Source: [Article Notebook](#)

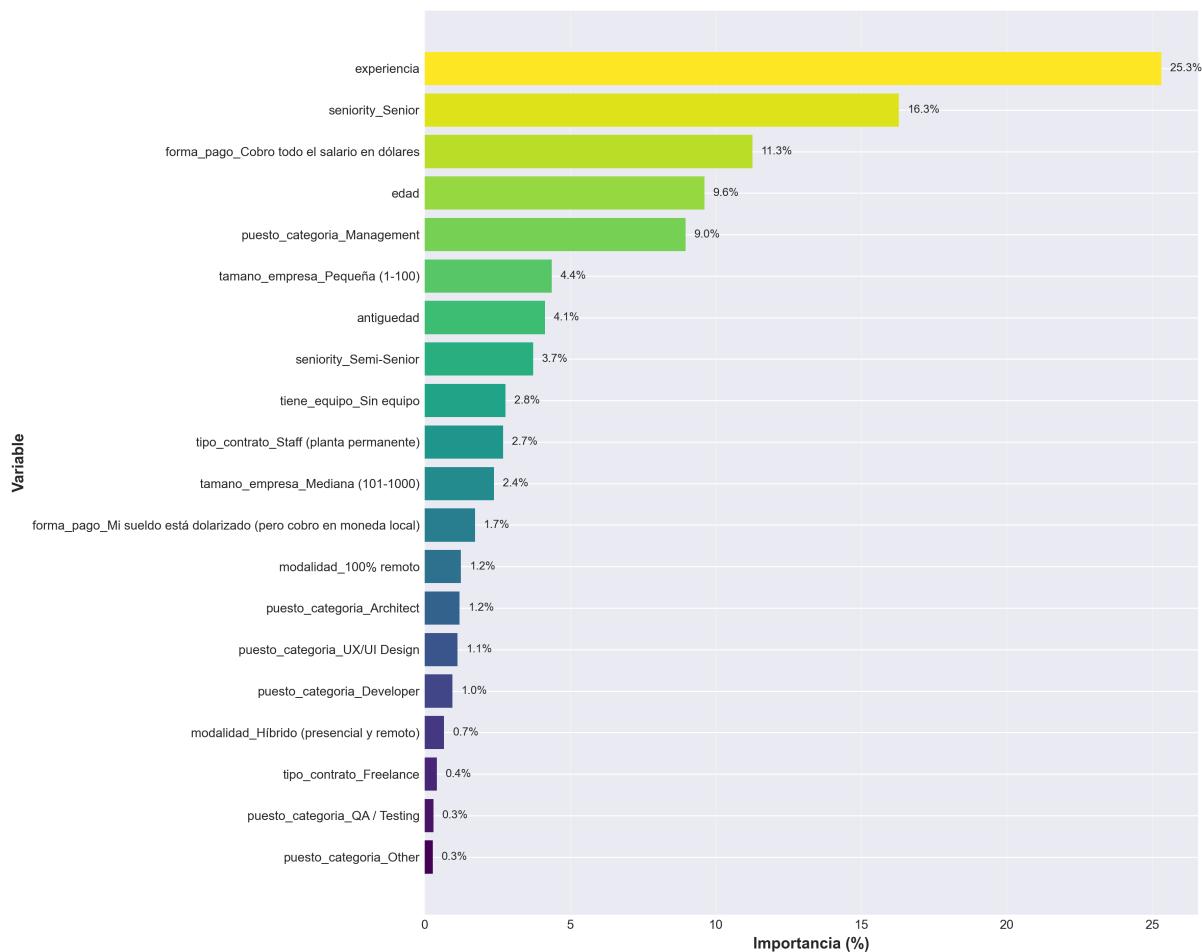
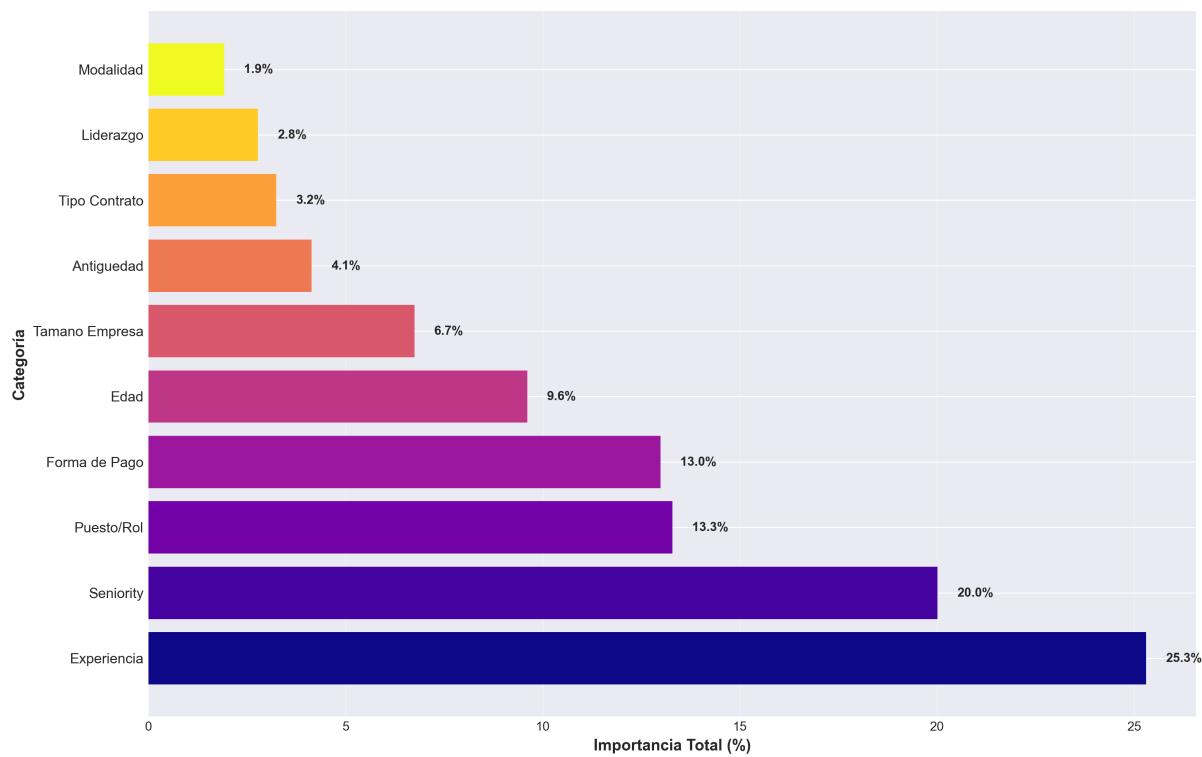
TOP 15 VARIABLES MÁS IMPORTANTES:

| | | Variable | Importancia_% |
|----|---|------------------|---------------|
| 0 | | experiencia | 25.301486 |
| 4 | | seniority_Senior | 16.287362 |
| 23 | forma_pago_Cobro todo el salario en dólares | | 11.255804 |
| 2 | | edad | 9.608281 |

| | | |
|----|--|----------|
| 18 | puesto_categoria_Management | 8.963168 |
| 12 | tamano_empresa_Pequeña (1-100) | 4.362377 |
| 1 | antiguedad | 4.133615 |
| 3 | seniority_Semi-Senior | 3.722887 |
| 25 | tiene_equipo_Sin equipo | 2.774539 |
| 9 | tipo_contrato_Staff (planta permanente) | 2.697445 |
| 11 | tamano_empresa_Mediana (101-1000) | 2.380986 |
| 24 | forma_pago_Mi sueldo está dolarizado (pero cob...) | 1.727312 |
| 5 | modalidad_100% remoto | 1.248300 |
| 13 | puesto_categoria_Architect | 1.203275 |
| 22 | puesto_categoria_UX/UI Design | 1.131085 |

IMPORTANCIA POR CATEGORÍA:

| Categoría | Importancia % |
|----------------|---------------|
| Experiencia | 25.301486 |
| Seniority | 20.010249 |
| Puesto/Rol | 13.288836 |
| Forma de Pago | 12.983116 |
| Edad | 9.608281 |
| Tamano Empresa | 6.743363 |
| Antiguedad | 4.133615 |
| Tipo Contrato | 3.240205 |
| Liderazgo | 2.774539 |
| Modalidad | 1.916310 |

Top 20 Variables MÁS Importantes para Predecir el Salario**Importancia Agrupada por Categoría de Variable**

Source: [Article Notebook](#)

3.4 Análisis de residuos

Finalmente, podemos analizar los residuos del modelo, es decir, la diferencia entre los salarios reales y los predichos. Esto nos permite identificar patrones en los errores del modelo y entender mejor su rendimiento.

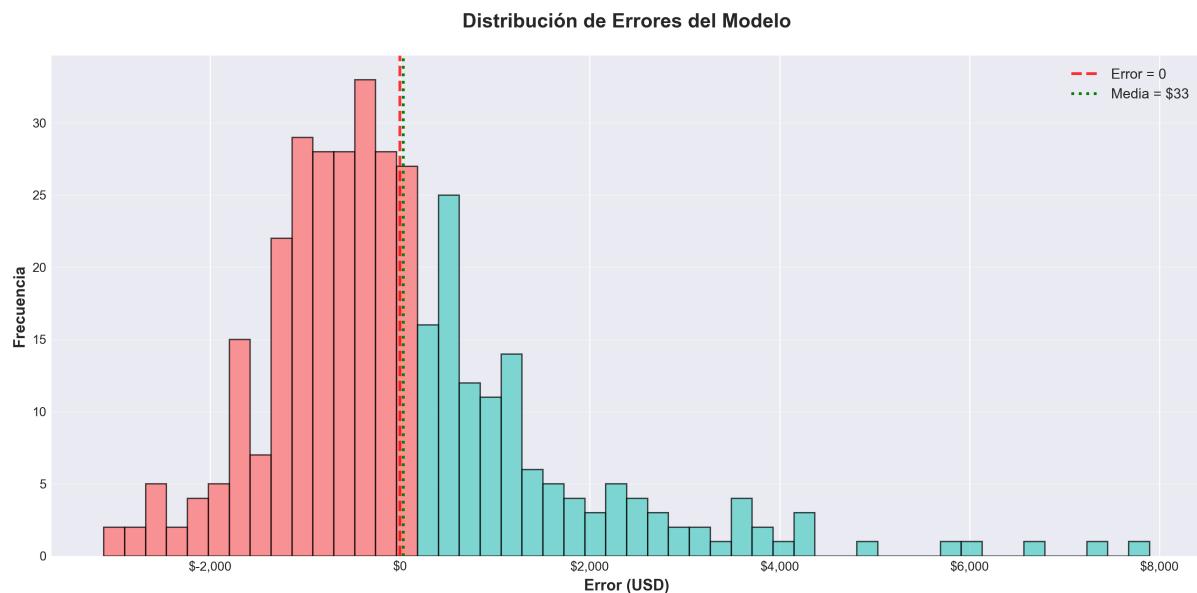
Source: [Article Notebook](#)

ESTADISTICAS DE LOS RESIDUOS:

Media: \$33 USD

Mediana: \$-249 USD

Desviación Estándar: \$1,561 USD



Source: [Article Notebook](#)

3.5 Conclusiones del modelo de Machine Learning

El modelo identifica correctamente que la experiencia, edad y forma de pago en dólares son los predictores más importantes del salario. Esto coincide con nuestro análisis previo, donde también destacamos estas variables como parte de la combinación ganadora. De esta forma validamos nuestras conclusiones y reforzamos la idea de que estas características son clave para maximizar el salario en el sector IT.

Source: [Article Notebook](#)

El modelo Random Forest revela que:

- * Precision (R² test): 36.5% - Explica 36.5% de la variabilidad salarial
- * Precision (R² CV): 34.5% - Validación cruzada confirma estabilidad
- * Error promedio: \$1,097 USD/mes ($\pm \56)
- * Overfitting: Controlado (diferencia R²: 0.046)
- * Factor mas determinante: Experiencia (25.3% de importancia)

Top 3 variables individuales:

1. experiencia: 25.30%

2. seniority_Senior: 16.29%
3. forma_pago_Cobro todo el salario en dólares: 11.26%

Source: [Article Notebook](#)

4 Conclusiones

A modo de cierre, luego del EDA (“Análisis exploratorio de Datos”) y del análisis utilizando Machine Learning, podemos llegar a la conclusión de que la mejor combinación de variables para maximizar el salario en el sector IT es la siguiente: en primer lugar, la experiencia laboral. Podríamos decir que “cuanto antes, mejor”, es decir, comenzar a trabajar lo antes posible para acumular experiencia, incluso mientras se está estudiando, ya que no parece ser un mercado laboral que exija títulos universitarios para acceder a mejores salarios. La mejora constante sí es algo necesario pero no necesariamente a través de títulos formales. La dolarización del salario también es un factor clave, ya que los salarios en dólares tienden a ser más altos en comparación con los salarios en moneda local. Por último, algo que también está vinculado con la seniority/experiencia es el liderazgo, aquellas personas con “gente a cargo” que dirigen equipos suelen tener salarios más altos. Esta última variable se vincula directamente con las llamadas “soft skills”, lo que coincide con la tendencia actual del mercado laboral a valorar cada vez más estas habilidades interpersonales y de gestión.

Source: [Article Notebook](#)