



PAPER

Proyecto Final: Brecha salarial de género en Glassdoor

Luis Rafael Castillo Rico¹

¹rafacastillo@ciencias.unam.mx

Resumen

En este trabajo se realizó un modelo predictivo capaz de calcular el sueldo de un perfil de empleado ficticio. Tras generar tres pares de perfiles, dónde la diferencia entre cada par era únicamente el género del perfil, se encontró que la brecha salarial predicha por el modelo fue en promedio del 2.79 %. Esto se logró mediante un ajuste basado en un conjunto de datos del sitio web de empleos Glassdoor. En la exploración se encontraron relaciones entre el sueldo total de los empleados y diversos parámetros como su nivel educativo, el departamento en el cual trabajan o su experiencia laboral, que revelan una brecha salarial que depende fuertemente del género. También se discutieron las consecuencias y el contexto de dicha brecha.

Key words: brecha salarial, regresión lineal

Introducción

Objetivos

Objetivo general

El objetivo de este trabajo es predecir el sueldo de un perfil ficticio y verificar si existe una brecha salarial al cambiar el género del empleado.

Objetivos particulares

- Definir el concepto de brecha salarial
- Discutir el contexto social
- Explorar y discutir el conjunto de datos de Glassdoor
- Realizar el un modelo lineal eficiente que prediga el sueldo de un perfil ficticio
- Comparar el sueldo predicho del perfil con su equivalente de distinto género

Brecha salarial

La brecha salarial de género es la diferencia promedio entre la remuneración de hombres y mujeres que están trabajando. En general la brecha salarial está a favor del género masculino, es decir, los hombres perciben un mayor salario que las mujeres por realizar el mismo trabajo. En los Estados Unidos, por ejemplo, el salario medio anual de la mujer es el 95 % del salario medio anual del hombre [1]. Por lo que la brecha salarial en Estados Unidos es del 5 %. Diversos

estudios afirman que en México la brecha salarial es del 13 % [2].

La brecha salarial no del todo una medida directa de la discriminación, sino una consecuencia de la misma. Ya que es causada en parte por la segregación ocupacional (con más hombres en industrias mejor pagadas y mujeres en industrias peor pagadas), segregación vertical (menos mujeres en puestos superiores y, por lo tanto, mejor pagados), legislación ineficaz sobre igualdad salarial, horas de trabajo remuneradas en general de las mujeres, y barreras de entrada al mercado laboral (como el nivel de educación y que, por ejemplo, la cantidad de madres solteras es mayor que la de padres solteros) [3]. En este trabajo se analizarán algunos de estos puntos en la exploración del conjunto de datos y se corroborará la brecha salarial por medio de un modelo de regresión lineal.

Consecuencias de la brecha salarial

El análisis realizado por el Banco Mundial sobre la naturaleza cambiante del trabajo [4] relaciona los ingresos con la acumulación de habilidades, lo que sugiere que la brecha salarial ocasione que las mujeres también acumulen menos capital humano (habilidades y conocimientos) en el trabajo y a lo largo de sus carreras. Por ejemplo, en Venezuela, por cada año adicional de trabajo, los salarios del género masculino aumentan en promedio un 2.2 %, en comparación con solo el 1.5 % del género femenino. Esto genera una especie de círculo vicioso en la brecha salarial ya que es una consecuencia y causa de la misma.

La brecha salarial de género puede ser también un problema

Cuadro 1. Columnas del conjunto de datos

Columna (original)	Columna (traducción)	Descripción	Tipo	Valores
JobTitle	Puesto	Puesto del empleado/a	Categorico	10 valores distintos*
Gender	Género	Género del empleado/a	Categorico	Femenino o masculino
Age	Edad	Edad del empleado/a	Numérico	Edad en años
PerfEval	Evaluación del Perfil	Evaluación del empleado/a dada por la empresa	Numérico	1, 2, 3, 4 ó 5
Education	Educación	Último grado de estudios del empleado/a	Categorico	Bachillerato, Licenciatura, Maestría y Doctorado
Dept	Departamento	Departamento de la empresa donde trabaja	Categorico	Departamento en la empresa **
Seniority	Años en la empresa	Años trabajados en la empresa	Numérico	Años trabajados
BasePay	Sueldo base	Salario anual del empleado/a	Numérico	Salario anual (en dólares estadounidenses)
Bonus	Bonos	Bono por rendimiento	Numérico	Bonos anuales (en dólares estadounidenses)

* Data Scientist, Driver, Financial Analyst, Graphic Designer, IT, Manager, Marketing Associate, Sales Associate, Software Engineer, Warehouse Associate

** Administration, Engineering, Management, Operations, Sales

desde la perspectiva de las políticas públicas, ya que reduce la producción económica y significa que es más probable que las mujeres dependan de los pagos de asistencia social, especialmente en la vejez [5], incrementando la probabilidad de que las mujeres empobrezcan con el paso de los años.

¿Se han tomado acciones para combatir la brecha salarial?

Según la edición de 2008 del informe de Perspectivas de empleo de la Organización para la cooperación económica y el desarrollo (OECD por sus siglas en inglés), casi todos los países de la OECD (México no incluido) han establecido leyes para combatir la discriminación por motivos de género. Ejemplos de esto son la Ley de Igualdad Salarial de 1963 y el Título VII de la Ley de Derechos Civiles de 1964 [6]. Sin embargo, la prohibición legal del comportamiento discriminatorio solo puede ser efectiva si se hace cumplir y, desgraciadamente, aunque muchos países de la OECD han establecido agencias especializadas contra la discriminación, solo en algunos de ellos estas agencias están efectivamente facultadas para investigar empresas, tomar acciones contra empleadores sospechosos de operar prácticas discriminatorias, y sancionarlos cuando encuentren evidencia de discriminación [7].

Los grupos de la sociedad civil organizan campañas de concienciación que incluyen actividades como el Día de la Igualdad de Pago o el movimiento de igualdad de salario por trabajo igual para aumentar la atención pública que recibe la brecha salarial de género. Por la misma razón, varios grupos publican informes periódicos sobre el estado actual de las diferencias salariales de género. Estos actos divulgativos son importantes para disminuir la discriminación hacia el género femenino ya que, por palabras de Catalina Ruiz Navarro en su plática "¿Por qué es importante hablar de feminismo?" [8]: La lucha feminista es y ha sido un movimiento ideológico, por lo tanto los ideales deben ser discutidos e insertados en la sociedad para generar un cambio en la misma, la tarea divulgativa es crucial en este proceso.

Glassdoor

Glassdoor es un sitio web estadounidense en el que los empleados actuales y anteriores califican a las empresas de forma anónima, también permite a los usuarios compartir sus salarios y compararlos con otros, así como buscar y solicitar puestos de trabajo ofrecidos por distintas empresas en su plataforma. Por lo que Glassdoor tiene grandes cantidades de información sobre las tendencias del mer-

cado laboral. En este trabajo se utilizará un conjunto de datos de Glassdoor para analizar las diferencias laborales, poniendo énfasis en el salario y analizando las diferencias de éste entre el género masculino y el femenino.

Descripción del conjunto de datos

El conjunto de datos [Glassdoor Gender Pay Gap](#) está disponible en la página de Kaggle y se centra en los ingresos de distintos puestos de trabajo según el género del empleado/a [9]. Este conjunto de datos resulta útil para verificar, y en su caso identificar, la profundidad de la brecha salarial basada en el género.

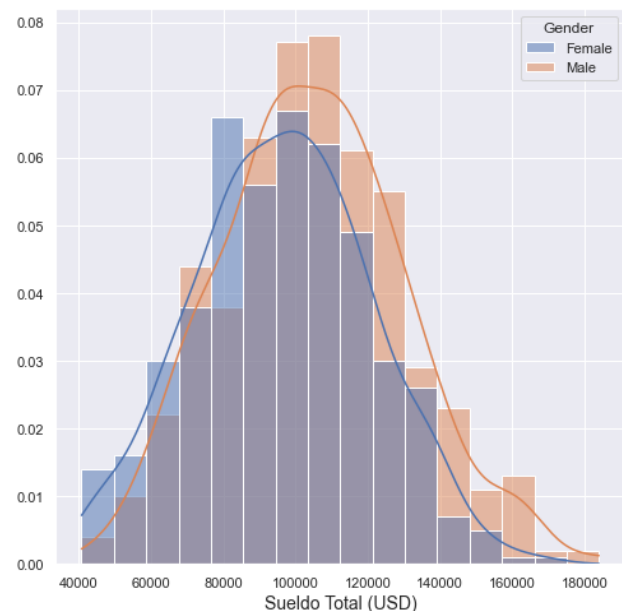


Figura 1. Porcentaje de empleados con un determinado sueldo

Limpieza de los datos

En el cuadro 1 se muestra la información principal de los atributos del conjunto de datos, que son las características de un conjunto

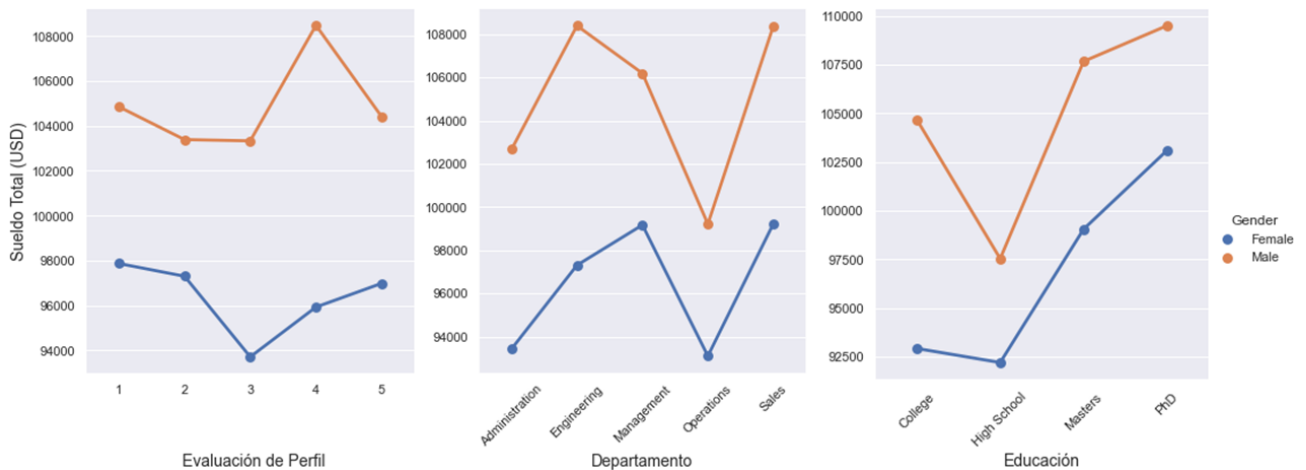


Figura 2. Distintos parámetros donde se presenta la brecha salarial

de 1,000 empleados elegidos al azar, donde 53.2 % son de género masculino y 46.8 % de género femenino. Para reducir posibles errores generados por la desigualdad en la cantidad de datos, se eliminaron aleatoriamente 64 datos de hombres, terminando con un conjunto de 946 datos, mitad hombres y mitad mujeres.

También se generó un nuevo atributo denominado "Sueldo Total", donde se sumó el sueldo base y los bonos para obtener todo el ingreso anual del empleado/a.

Análisis y exploración

Al hacer un histograma del sueldo total diferenciando el género, como el que se muestra en la figura 1, distintos factores en la gráfica resaltan lo discutido anteriormente. Por un lado el ajuste muestra claramente que los sueldos menores a 80,000 dólares anuales corresponden mayormente a mujeres, mientras que los sueldos más altos, por encima de los 80,000 dólares anuales, corresponden a hombres. Habiendo incluso un pequeño salto en la gráfica en el millón de dólares anuales. Para ambos géneros la gráfica tiene una forma gaussiana, con un corrimiento hacia la derecha (sueldos más altos) para la gráfica de los hombres, mostrando de manera gráfica la brecha salarial.

En el cuadro 2 se muestran los promedios de los valores numéricos al agrupar el conjunto de datos por género y se aprecia una brecha salarial del 8 % a favor del género masculino. Cabe también destacar que la evaluación del perfil y los años trabajados son también mas bajos para las mujeres.

Cuadro 2. Promedios dependiendo del género

Género	Edad	Evaluación de Perfil	Años Trabajados	Sueldo Total*
Masculino	41.83 años	2.94	3.01	96416.83
Femenino	41.01 años	3.13	2.93	104918.68

* En dólares estadounidenses

Al graficar otras variables como las mostradas en la figura 2, es apreciable que la brecha salarial se observa independientemente de la evaluación del perfil, el departamento de la empresa o la educación de los empleados. Se observa también que, aunque la brecha es independiente de estos factores, la magnitud de esta si puede variar para distintas variables de los mismos. En la gráfica central, que muestra los sueldos promedio por departamento, es claro que en el área de ingeniería, la brecha salarial es aún mayor que para los demás departamentos. Se muestra también que la brecha salarial

se amplifica para los empleados con un título de licenciatura y para los empleados con evaluaciones de 3 y 4.

Otra relación encontrada, que puede tener una influencia en la brecha salarial, es la que se da al entre la edad y la evaluación que la empresa da del empleado/a. Es apreciable en la figura 3 que la evaluación otorgada no varía dependiendo de la edad para los hombres, pero muestra una conducta decreciente con respecto a la edad para las mujeres.

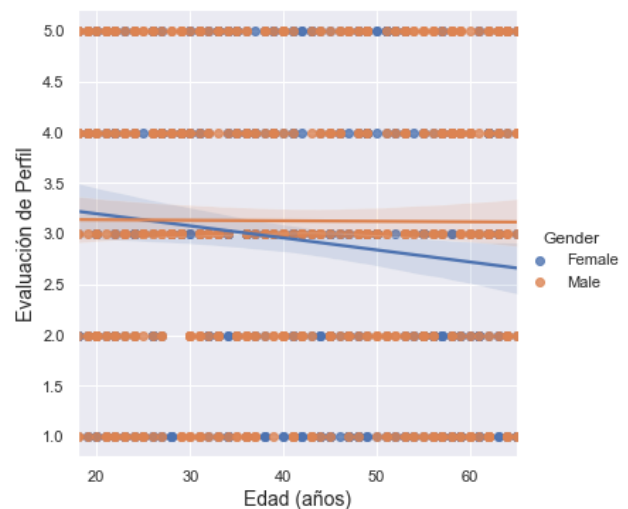


Figura 3. Dependencia de la evaluación del perfil con respecto a la edad

Las mujeres jóvenes, de entre 18 y 25 años, tienen en promedio calificaciones iguales, incluso superiores (para las mujeres más jóvenes) a las de los hombres de su misma edad, mientras que las mujeres mayores son evaluadas con calificaciones inferiores. Este hecho puede deberse a diversos factores relacionados con la discriminación que no serán discutidos en este trabajo o a factores no discriminatorios que pueden afectar el desempeño de una mujer en su trabajo como el permiso por maternidad.

La última relación encontrada es entre la edad y el sueldo base y se muestra en la figura 5. En ella se aprecia que para ambos géneros el sueldo es mayor conforme la edad aumenta, sin embargo se vuelve a percibir la brecha salarial a favor de los hombres, se muestran dos ajustes lineales con rectas casi paralelas. Cabe destacar que los datos son para empleados diferentes, no se tomó en consideración un empleado a una cierta edad y el mismo empleado

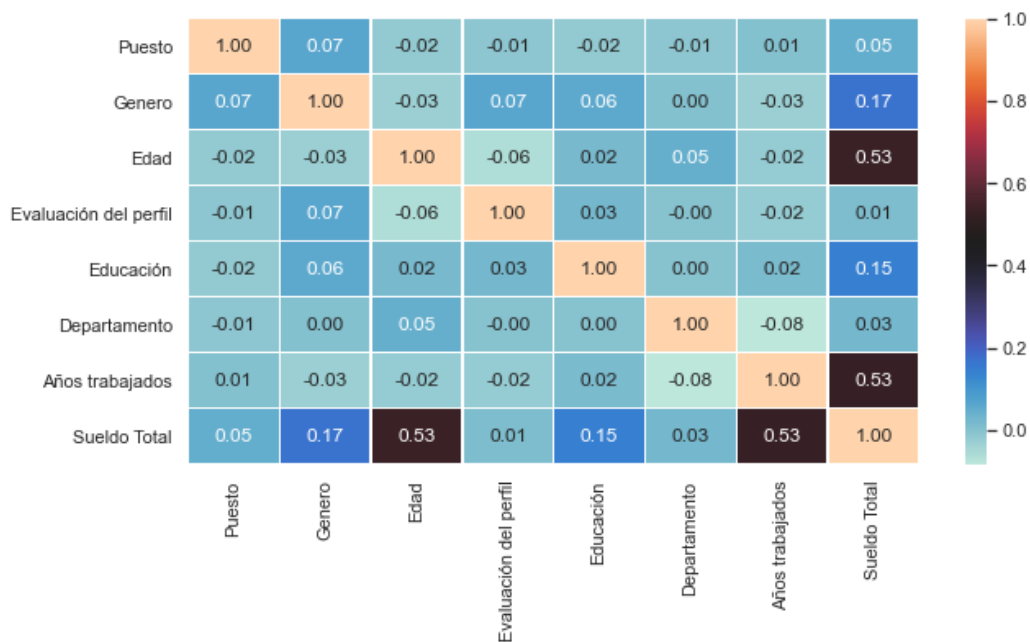


Figura 4. Matriz de correlaciones

algunos años después, sería interesante analizar una base de datos con ese tipo de recopilación para verificar si los empleados hombres van aumentando su salario más rápido que las mujeres, en dicho caso se esperaría que la recta ajustada para el salario de los hombres tuviera una pendiente mayor.

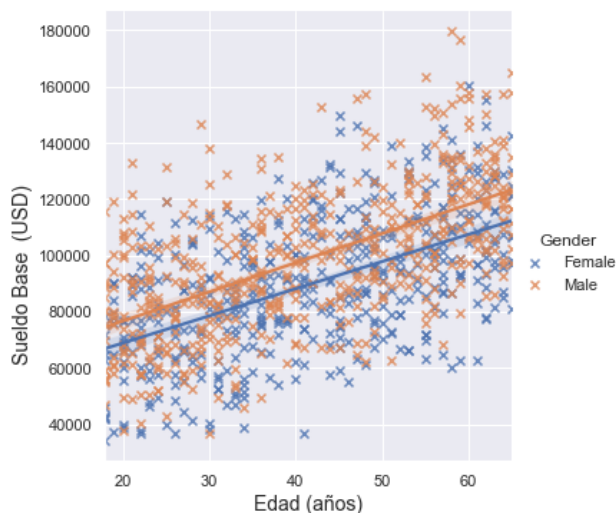


Figura 5. Dependencia del sueldo base con respecto a la edad

Para obtener las relaciones entre el género de un empleado/a y los atributos descritos en el conjunto de datos analizado, resulta útil obtener la matriz de correlaciones. Esta matriz se muestra en la figura 4. En dicha matriz destaca que el género es un parámetro que se relaciona más que la educación del empleado/a o el puesto de éste/a con el salario total que el/la empleado/a percibe. Lo cual es un hecho alarmante.

Se encuentran también algunas relaciones con el sueldo total que se esperarían, como los años trabajados o la edad del empleado/a.

Modelado

Regresión lineal

Ya se ha mostrado la relación directa que hay entre el sueldo total y distintos atributos en el conjunto de datos, incluso se han mostrado ajustes lineales que puedan predecir el sueldo dependiendo de algún atributo en específico. Sin embargo para retratar la brecha salarial resulta aún más evidente predecir el sueldo de una persona con un conjunto de características fijas y comparando su sueldo predicho al cambiar únicamente el género. En este trabajo se modeló la relación entre las variables descritas en el conjunto de datos mediante el ajuste de una ecuación lineal para hacer la predicción del sueldo. El modelo fue entrenado con el conjunto de datos de Glassdoor y se utilizaron dos estrategias de regularización: Ridge y Lasso.

Preparación de los datos

Como primer paso, se separaron los atributos en numéricos y categóricos como se muestra en la tabla 1, con excepción del atributo educación, al cual se le aplicó la estrategia LabelEncoding, asignándole una escala ordinaria a las variables:

- I. High-School (bachillerato)
- II. College (licenciatura)
- III. Master (maestría)
- IV. PhD (doctorado)

De esta forma se organizaron las variables para que los números más altos correspondieran con grados de educación superiores.

Posteriormente, las variables numéricas se estandarizaron (StandardScaler) y se convirtieron a un arreglo polinomial de grado 3 (PolynomialFeatures(degree=3)). A las variables categóricas se les aplicó la técnica One-Hot encoding. Finalmente, el conjunto de datos se separó para el conjunto de entrenamiento y para el de prueba en una relación 70 % - 30 % de manera aleatoria y se aplicó un cross-validation dentro de un grid search para obtener la regularización con los hiper-parámetros más confiable para predecir el sueldo. Las puntuaciones más altas, junto con los hiper-parámetros

para cada regularización, se muestran en la tabla 3.

Cuadro 3. Puntuaciones R^2 promedio para los modelos

Parámetros	Lasso	Ridge
α	37	8
solver	auto	-
selection	cyclic	-
Puntuación	0.8328	0.8202

Para el cross validation se tomó $cv=10$

La regularización con mejor puntuación resultó ser Lasso, por lo que la predicción del salario dada por este método se considerará más confiable, aunque dadas las puntuaciones no se espera que las diferencias entre las predicciones de ambos modelos difieran de manera considerable.

Prediciendo un sueldo

Para este ejercicio se crearon entradas nuevas, dándole valores a los atributos analizados y se solicitó al modelo que predijera el sueldo total. Los datos ingresados al modelo Lasso ajustado y los sueldos que predijo el mismo se muestran en el cuadro 4.

Para los tres datos se predice una brecha salarial a favor del género masculino, en promedio la brecha para los tres datos fue del 2.79 %, mostrando que los datos de Glassdoor reflejan una brecha salarial entre géneros, favoreciendo siempre al género masculino.

Cuadro 4. Sueldos predichos por los modelos

Puesto	Dato 1 Data scientist	Dato 2 Manager	Dato 3 Software Engineer
Edad	27	31	33
Evaluación	4	5	3
Educación	College	Master	PhD
Departamento	Administration	Operations	Sales
Años trabajados	1	5	3
Sueldo (Mujer)	65,119.22	145,196.90	100,496.73
Sueldo (Hombre)	67,264.40	148,373.76	103,673.59
Brecha Salarial	3.18 %	2.14 %	3.06 %

Discusión

Para futuros trabajos se propone realizar una predicción para todas las posibles combinaciones de los atributos en el conjunto de datos y obtener el valor promedio de la brecha salarial. Factores como el departamento o el puesto pueden jugar un factor relevante que puede ser evidente en las predicciones. El programa desarrollado también podría ser útil para ver que tan competitivo es un sueldo dado o como una guía para saber qué sueldo solicitar en una entrevista de trabajo.

Por otro lado, aunque la discriminación hacia el género femenino es un hecho, cabe destacar que la brecha salarial también se debe a que ciertos empleos son mejor pagados para hombres por factores que nada tienen que ver con la discriminación. Por ejemplo, en un estudio realizado por la universidad de Stanford [10], con una base de datos de más de un millón de conductores de Uber en Estados Unidos, se mostró que la brecha salarial de género es de

un 7 % a favor de los hombres. El algoritmo de Uber no distingue entre hombres y mujeres, por lo que en el estudio esta brecha se le adjudica a factores como la experiencia o incluso la malicia de los conductores masculinos al aceptar viajes a ciertas horas o en lugares donde la tarifa dinámica de Uber los favorezca (se le denomina tarifa dinámica al aumento en el costo del viaje cuando se solicitan muchos choferes al mismo tiempo en un determinado lugar). Por ejemplo a la salida de un concierto. Sin embargo en este trabajo no se considerarán esas atribuciones como válidas al ser meramente subjetivas y no contar con un sustento estadístico. Un hecho sustentado estadísticamente, que puede explicar dicha brecha y también es mostrado en el estudio, es que los conductores masculinos manejan 2.2 % más rápido y son 5.6 % menos propensos a cancelar viajes. En este trabajo no se busca profundizar en el motivo de este comportamiento, sino que se toma como una justificación objetiva para explicar la brecha salarial entre los choferes de Uber por factores que no tienen que ver con su empleador.

Conclusiones

En este trabajo se observó que el conjunto de datos de Glassdoor [9] refleja una brecha salarial que favorece al género masculino independientemente de parámetros como la educación, el departamento de la empresa, o la evaluación que la empresa le da a sus empleados/as, además, se observó que la evaluación dada por la empresa tiene una dependencia con respecto a la edad para el género femenino que no se presenta para el género masculino. También se obtuvo la mejor regularización (Lasso) con su configuración de parámetros más eficiente para predecir el sueldo de una persona, observando que la brecha salarial se mantiene en las predicciones con un promedio de 2.79 %. Se propuso también una metodología para futuros trabajos, tomando en consideración todas las posibles combinaciones entre los atributos del conjunto de datos para obtener una brecha salarial promedio con el menor sesgo posible.

Referencias

1. Andrew Chamberlain, Daniel Zhao, Amanda Stansell, (27 de marzo del 2019), Progress on the Gender Pay Gap: 2019, <https://www.glassdoor.com/research/gender-pay-gap-2019/>
2. Índice de competitividad internacional 2021, Instituto Mexicano para la Competitividad (IMCO), sitio web: <https://imco.org.mx/indice-de-competitividad-internacional-2021/:~:text=M%C3%A9xico%20obtuvo%20la%20posici%C3%B3n%2037,M%C3%A9xico%20muestra%20brechas%20de%20competitividad>
3. Denis Leythienne, Piotr Ronkowski, (2018), A decomposition of the unadjusted gender pay gap using Structure of Earnings Survey data, Statistical Working Papers.
4. "World Bank World Development Report 2019: The Changing Nature of Work", (2019), <https://documents1.worldbank.org/curated/en/816281518818814423/pdf/2019-WDR-Report.pdf>
5. Klasen, Stephan (5 de Octubre del 2018). "The Impact of Gender Inequality on Economic Performance in Developing Countries". Annual Review of Resource Economics. 10 (1): 279–298. doi:10.1146/annurev-resource-100517-023429. ISSN 1941-1340
6. Tharenou, Phyllis (7 de Octubre del 2012). "The Work of Feminists is Not Yet Done: The Gender Pay Gap—a Stubborn Anachronism". Sex Roles. 68 (3–4), 198–206. doi:10.1007/s11199-012-0221-8. S2CID 144449088
7. U.S. Government Accountability Office. Women's Earnings: Federal Agencies Should Better Monitor Their Performance in Enforcing Anti-Discrimination Laws. (1 de Abril del 2011)
8. Catalina Ruiz Navarro, (2017), ¿Por qué es importante hablar de feminismo?, TEDx Talks,

<https://www.youtube.com/watch?v=CnXizUTJfYQ>

9. Glassdoor Gender Pay Gap, Kaggle Dataset, sitio web:
<https://www.kaggle.com/nilimajauhari/glassdoor-analyze-gender-pay-gap?select=Glassdoor+Gender+Pay+Gap.csv>,
revisado el 25 de octubre del 2021
10. "The Gender Earnings Gap in the Gig Economy: Evidence from over a Million Rideshare Drivers", (8 de Marzo del 2018), Stanford University.