

Análisis para el establecimiento de ideas para instituciones educativas que deseen adaptar sus programas de formación

Barboza, Marcos; Reniero Isaías

Universidad Nacional de Entre Ríos

Facultad de Ciencias de la Administración

Licenciatura en Sistemas

Trabajo práctico integrador de Probabilidad y Estadística

1. Introducción

En el presente trabajo integrador se analizará una base de datos de empleados del rubro IT con el objetivo de buscar formas de alinear los programas educativos de la disciplina Licenciatura en Sistemas de la Facultad de Ciencias de la Administración (FCAD), de la Universidad Nacional de Entre Ríos (UNER), con las habilidades y áreas de especialización demandadas en el mercado laboral. Esta tarea resulta esencial para preparar a los estudiantes y profesionales en formación, brindándoles las herramientas necesarias para enfrentarse al entorno laboral actual. El objetivo final es asegurarse de que los programas educativos estén alineados con las necesidades y tendencias actuales y futuras del mercado laboral tecnológico, proporcionando oportunidades de aprendizaje para cerrar cualquier brecha identificada entre las habilidades requeridas y las habilidades que poseen los estudiantes o profesionales en formación.

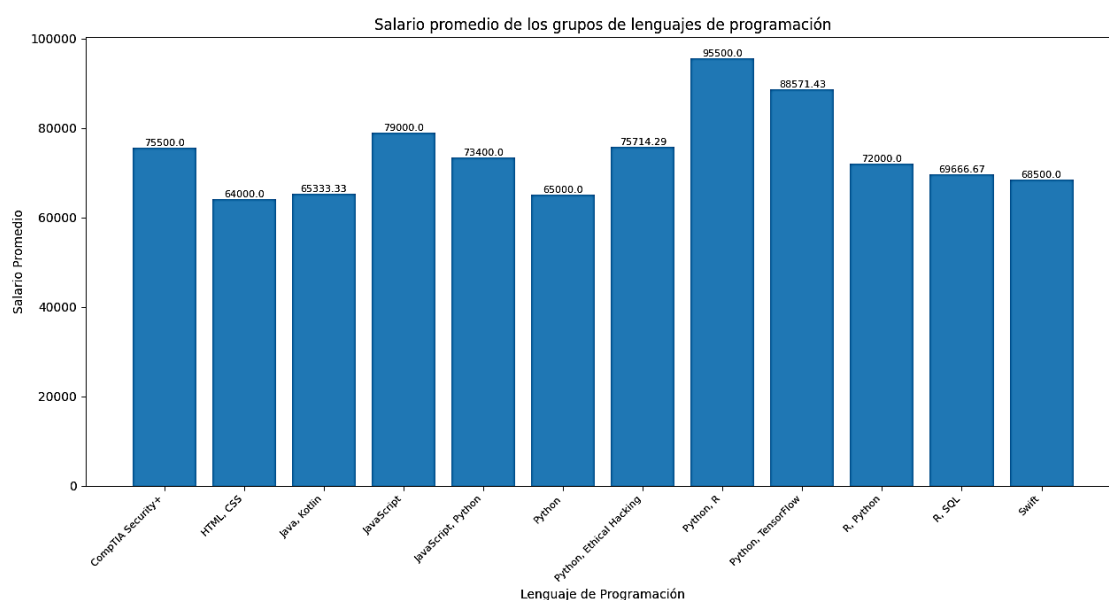
1.1 Análisis descriptivo

En este análisis se expondrán tablas y gráficos estadísticos esenciales para el análisis de datos, ya que permiten resumir y visualizar los datos de manera que puedan ser fácilmente interpretados.

1.1.1 Salario promedio de los grupos de lenguajes de programación

El diagrama de barras refleja los promedios de los salarios. En el eje X se representan los lenguajes de programación (variable cuantitativa), y en el eje Y se representan los valores de salarios (variable cualitativa) de cada grupo de lenguaje de programación. Cada barra representa un grupo de lenguaje de programación y su altura indica el promedio de ganancia que obtienen los profesionales al implementar dicho lenguaje. Gracias a este gráfico se pueden identificar los lenguajes de programación más demandados en relación con los salarios.

En un primer análisis, en la base de datos se identificaron repeticiones de lenguajes de programación, donde la única variación radicaba en el orden de los nombres de estos lenguajes. Además, se observó que, en ciertos conjuntos de lenguajes de programación con más de dos elementos, el tercer lenguaje carecía de relevancia para determinar el área de aplicación del grupo en cuestión. Basándonos en esta justificación, se decidió realizar una “limpieza” en la base de datos con el objetivo de evitar la inclusión de datos innecesarios en el gráfico. Esto permitió obtener una estimación más precisa del promedio salarial, eliminando así la aparición de barras redundantes en la representación gráfica.



En la siguiente tabla se presentan los grupos de lenguajes de programación y la especialización en la cual son implementados.

Lenguajes de Programación	Especialización
R, Python	Análisis de Datos
R, SQL	Análisis de Datos
Java, Kotlin	Desarrollo Móvil
Swift	Desarrollo Móvil
Python, R	Inteligencia Artificial
Python, TensorFlow	Inteligencia Artificial
HTML, CSS	Desarrollo Web
JavaScript	Desarrollo Web
JavaScript, Python	Desarrollo Web
Python	Seguridad
CompTIA Security+	Ciberseguridad
Python, Ethical Hacking	Ciberseguridad

La combinación de los lenguajes R y Python muestra el salario promedio más alto, alcanzando los \$95.500, seguida por Python y TensorFlow con un promedio de \$88.571,43, y JavaScript con \$79.000. En contraste, la combinación de HTML y CSS, utilizada en el desarrollo web, refleja la remuneración promedio más baja, llegando a \$64.000.

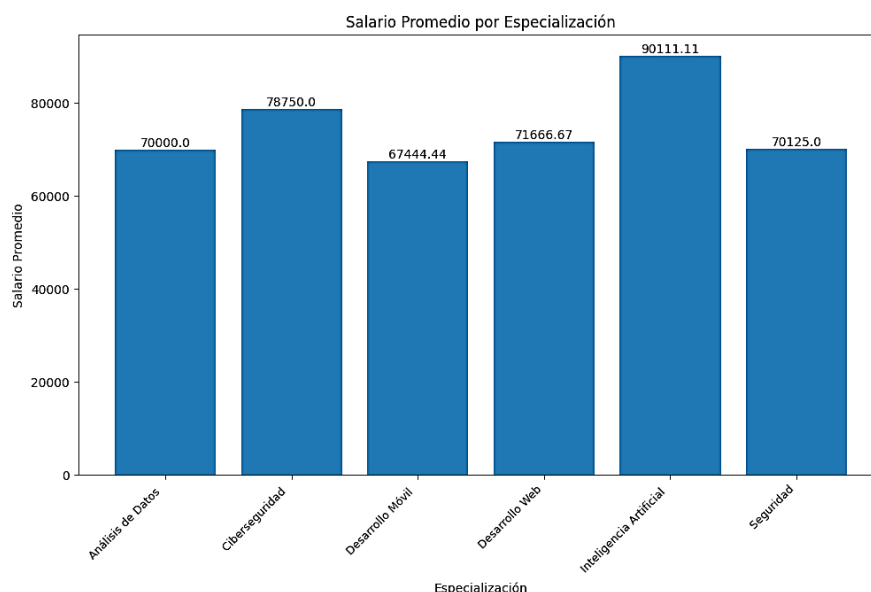
Estos hallazgos proporcionan una guía para el diseño de talleres especializados en los lenguajes más demandados en distintas áreas. Por ejemplo, Python destaca por su versatilidad, ya que,

como muestra la tabla, se aplica en el análisis de datos, la inteligencia artificial, el desarrollo web y la ciberseguridad.

1.2.2 Salario promedio por especialización

En este gráfico, es posible identificar las habilidades más demandadas en relación con los salarios promedio de cada una. Además, se puede enfocar la atención en áreas de especialización que ofrecen mayores oportunidades de crecimiento y remuneración. Según la información observada en la tabla, se destaca que la especialización en "Inteligencia Artificial" presenta el salario promedio más elevado, seguido por la ciberseguridad y, en último lugar, el desarrollo web. Estas tres especializaciones son las más destacadas.

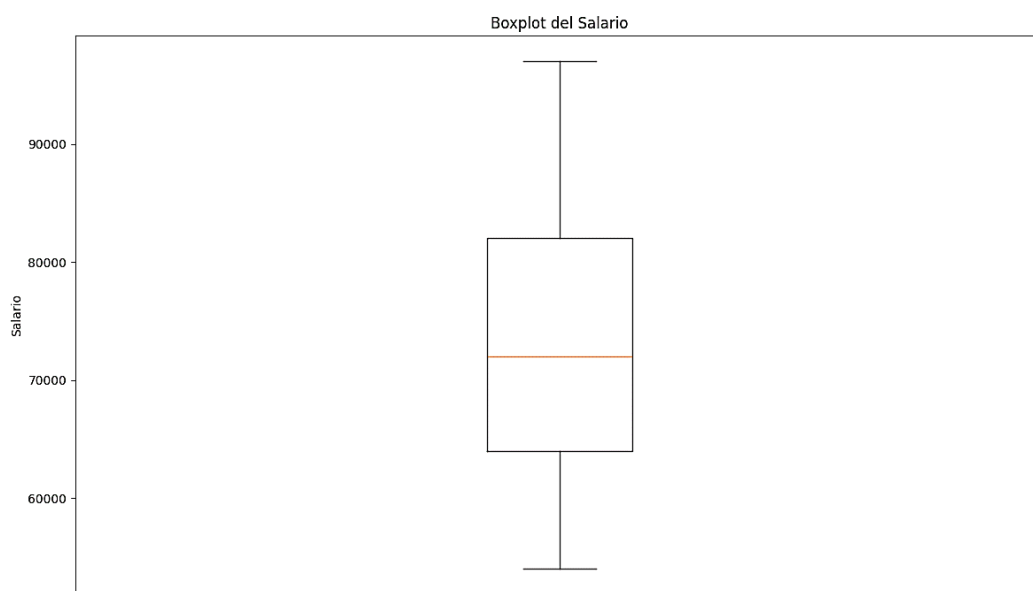
Una propuesta consistiría en ofrecer cursos de crédito que introduzcan a los estudiantes en estas especializaciones, permitiéndoles descubrir lo que realmente les apasiona y maximizar así el provecho de su carrera profesional.



1.2.3 Boxplot del salario

En el diagrama de caja analizado, la mediana representa el salario central de la muestra. Esto indica que la mitad de los profesionales en el rubro de IT tienen un salario igual o menor a \$72.000. Además, se observa que los salarios superiores a esta mediana tienden a estar más dispersos, lo que sugiere una mayor variabilidad en las remuneraciones más altas.

La distribución salarial muestra una tendencia hacia la derecha, indicando que los salarios más altos se encuentran menos concentrados y más dispersos en comparación con los salarios más bajos. Esta asimetría revela que hay una proporción menor de profesionales con salarios significativamente más altos que la mediana.



Tamaño de la muestra	49
Media	\$ 74.367,00
Mediana	\$ 74.367,00
Desviación estándar	12.468,00
Q₁	\$ 64.000,00
Q₃	\$ 82.000,00
Rango intercuartil	\$ 72.000,00
Máximo	\$ 97.000,00
Mínimo	\$ 54.000,00

1.2.4 Proyectos en curso de los profesionales

En la siguiente tabla se presentan las frecuencias absolutas simples que indican la cantidad de veces que ciertos proyectos tecnológicos están siendo desarrollados por profesionales actualmente. Este análisis permite identificar los proyectos más comunes y populares en el campo de la tecnología en la actualidad.

Según la tabla proporcionada, se destaca que los proyectos relacionados con aplicaciones móviles tienen la mayor frecuencia, seguidos por proyectos de robótica y en tercer lugar, proyectos vinculados al análisis de datos en el ámbito de la salud. Esta información revela que la creación de aplicaciones móviles es un área de interés significativo y podría ser propuesta como un taller desafiante para estudiantes interesados en programación, aunque no sea un requisito dentro de la carrera. En la facultad, actualmente se ofrecen cursos de robótica; sin embargo, considerando la alta demanda de estos proyectos, se podrían mejorar estos cursos para abordar las necesidades actuales del mercado. Por último, el análisis de datos en el ámbito de la salud es un proyecto más específico. Se propone la idea de desarrollar un curso de crédito enfocado en el análisis de datos en general, centrado en aspectos prácticos. Esto complementaría los temas relacionados con datos que se cubren en la carrera y brindaría una perspectiva más amplia sobre el análisis de datos en diversas áreas.

Especialización del proyecto	Cantidad de proyectos
Aplicación Móvil	9
Robótica	5
Análisis de Datos de Salud	5
Sistemas Embebidos	4
Procesamiento de Lenguaje Natural	4
Plataformas de Comercio Electrónico	4
Seguridad en la Nube	4
Ciberseguridad	3
E-commerce	3
Análisis de Seguridad	3
Juegos en Línea	2
Otras: Análisis de Datos, IoT, Realidad Aumentada.	1

2. Desarrollo

En esta sección, se iniciará con la estimación del salario promedio de los profesionales del rubro de IT. Posteriormente, se buscará determinar el promedio de años de experiencia de estos profesionales. Seguido de esto, se llevará a cabo una prueba de hipótesis enfocada en los salarios promedio considerando dos especializaciones. El análisis continuará con un estudio de regresión del salario y la experiencia laboral, con el propósito de evaluar si la experiencia laboral guarda una relación significativa con los salarios de los profesionales de IT. Para concluir este estudio, se realizará un análisis de correlación entre la edad y la experiencia laboral en el ámbito tecnológico, con el fin de determinar la existencia de una correlación notable entre estos dos factores.

2.1 Estimación por intervalo

Dado que la base de datos que estamos analizando cuenta con más de 30 datos de distintos profesionales, utilizaremos la siguiente metodología para la construcción de intervalos de confianza para la media poblacional:

El Teorema del Límite Central establece que, incluso si las poblaciones muestrales no siguen una distribución normal, las distribuciones muestrales de los estadísticos (las medias aritméticas de Salario y Años de experiencia) serán aproximadamente normales cuando el tamaño muestral n sea grande ($n \geq 30$). Por lo tanto, las distribuciones muestrales de la media muestral de los salarios y la media muestral de los años de experiencia tienen distribuciones que pueden aproximarse a una distribución normal.

En ambos análisis que se realizarán, se aplicará un nivel de confianza de 0,95, es decir, se utilizará un intervalo de confianza del 95%.

Se utilizará la siguiente fórmula para la construcción de los intervalos de confianza:

$$\bar{x} \pm z \cdot \frac{\sigma}{\sqrt{n}}$$

Dado que desconocemos la desviación estándar de la media poblacional σ , utilizaremos la desviación estándar de la media muestral s . Esto introduce cierta incertidumbre ya que s puede variar de una muestra a otra y por lo tanto modifica la distribución muestral de la media muestral.

2.1.1 El salario promedio de profesionales del área de IT de la población mundial

Se presentan los datos necesarios para reemplazarlos en la fórmula mostrada anteriormente:

Media del salario muestral = 74367,34694

Margen de error = 5%

Desviación estándar muestral (s) = 12467,78161

Tamaño de la muestra = 49

- Límite inferior: $74367,34694 - 1,96 * (12467,78161/\sqrt{49}) = 70.876,36809$
- Límite superior: $74367,34694 + 1,96 * (12467,78161/\sqrt{49}) = 77.858,32579$

La estimación por intervalo nos brinda una visión del rango probable para el salario promedio poblacional, con un nivel de confianza del 95%. En este caso, según los cálculos, el salario medio para profesionales del área de IT a nivel mundial se encuentra estimado entre \$70.876,37 y \$77.858,33 dólares. Esto sugiere que, en repeticiones de muestreo, existe una probabilidad del 95% de que el verdadero salario promedio, representado por μ , se encuentre dentro de este intervalo.

2.1.2 Los años de experiencia promedio de profesionales del área de IT de la población mundial

Se presentan los datos necesarios para reemplazarlos en la fórmula mostrada anteriormente:

Media del salario muestral = 9,326530612

Margen de error = 5%

Desviación estándar muestral (s) = 3,2299881

Tamaño de la muestra = 49

- Límite inferior: $9,326530612 - 1,96 * (3,2299881/\sqrt{49}) = 8,422133944$
- Límite superior: $9,326530612 + 1,96 * (3,2299881/\sqrt{49}) = 10,23092728$

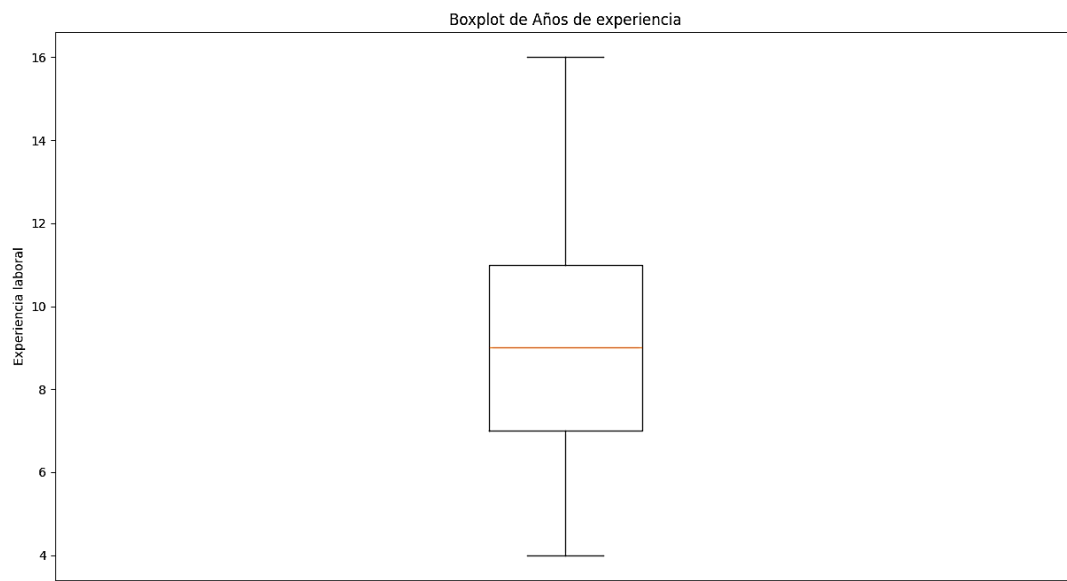
El intervalo de confianza del 95% para el promedio de años de experiencia de profesionales del área de IT se sitúa entre 8,4 y 10,2 años. Esto indica que existe una alta probabilidad, del 95%, de que la media real de años de experiencia de esta población se encuentre dentro de este rango establecido.

2.3 Análisis de regresión

2.3.1 Boxplot de años de experiencia

El análisis muestra que la mediana de 9 años indica un punto medio en la distribución de la experiencia laboral. La presencia de individuos con una vasta experiencia contribuye a una

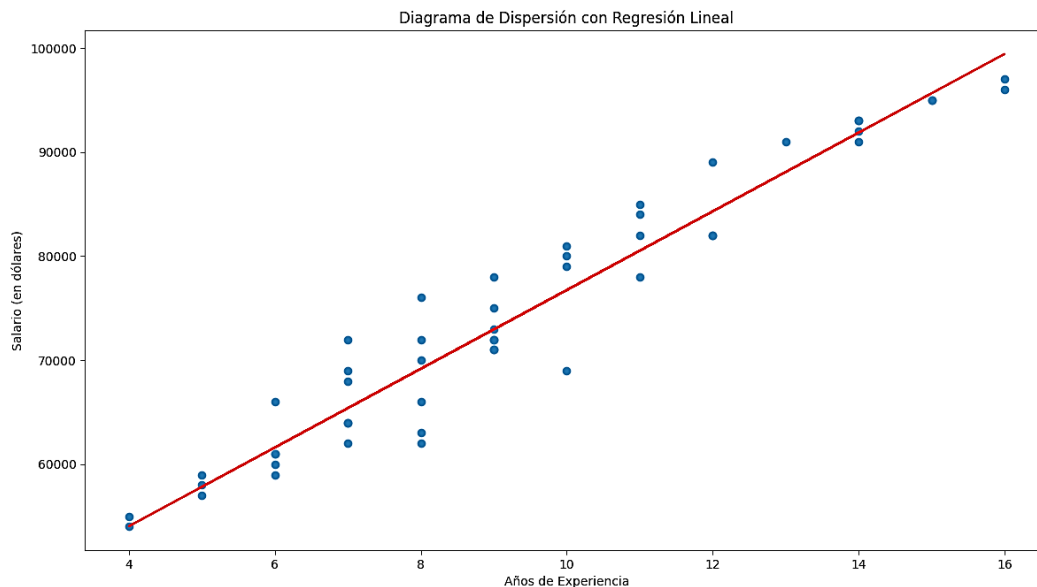
dispersión significativa en el extremo superior, mientras que la mayoría se concentra alrededor de los 7 a 11 años. Esto sugiere que, aunque la mayoría tiene una experiencia laboral moderada, hay un grupo notable con una experiencia significativamente alta que afecta a la distribución, generando una asimetría positiva y elevando la mediana por encima de la moda, pero por debajo de la media en términos de experiencia laboral.



Tamaño de la muestra	49
Media	9.326531
Mediana	9
Desviación estándar	3.229988
Q₁	7
Q₃	11
Rango intercuartil	4
Máximo	16
Mínimo	4

2.3.2 Diagrama de dispersión con regresión lineal

El diagrama de dispersión que se muestra a continuación tiene como objetivo representar de manera visual la relación entre dos variables cuantitativas: salario y años de experiencia de los profesionales en el campo de la informática. Mediante el uso de datos extraídos de la base de datos (muestra), se pretende identificar una línea de tendencia, también conocida como “línea de mejor ajuste”, que permita realizar proyecciones o predicciones sobre estas variables para toda la población.



Coefficiente de determinación (R-squared): 0,92

Coefficiente de pendiente: 3782,43

Ordenada al origen: 38907,9

La fórmula de la regresión lineal es: $\hat{y} = 3782.43x + 38907.9$

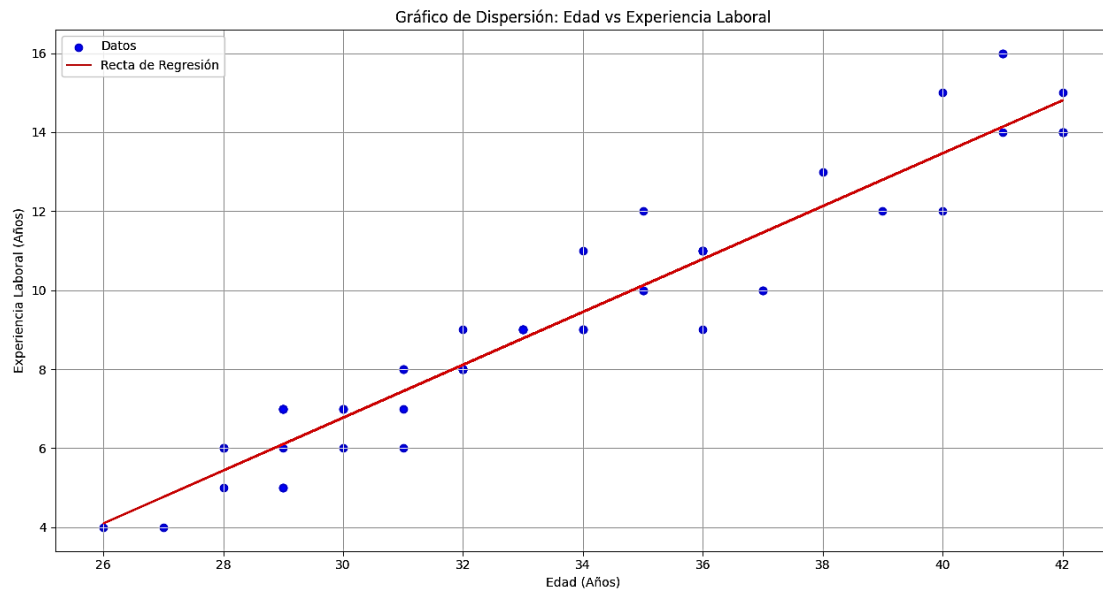
El coeficiente de determinación, con un valor cercano a 1, indica que una gran parte de la variabilidad de la variable “Salario” está siendo explicada por la variable “Años de experiencia”. Además, el coeficiente de pendiente de la recta es positivo, lo que indica una relación creciente entre ambas variables. En otras palabras, a medida que aumenta la variable independiente (los años de experiencia), también lo hace la variable dependiente (el salario). Esta relación sugiere que a medida que una persona gana más experiencia laboral, es probable que su salario aumente de manera correspondiente.

Conclusiones:

El análisis de regresión revela una relación significativa entre la experiencia laboral y el salario. Esto sugiere que, en términos generales, una mayor experiencia en el campo de la tecnología se asocia con salarios más altos. Sin embargo, dado que nuestro análisis se basa únicamente en dos variables, no podemos afirmar que esta correlación implique causalidad. Es posible que existan una o más variables adicionales que influyan en esta relación. En este contexto, los profesionales con más años de experiencia tienden a percibir salarios más elevados en comparación con aquellos con menos experiencia. Estos hallazgos pueden ser útiles para la toma de decisiones en términos de desarrollo profesional y estrategias de formación para mejorar las oportunidades salariales.

2.4 Análisis de correlación

El propósito de este análisis de correlación es medir la fuerza y la dirección de la relación lineal entre dos variables: la edad y la experiencia laboral. Se busca evaluar de qué manera una variable varía en relación con los cambios de la otra variable.



Covarianza entre Experiencia Laboral y Edad: 14.059975010412327

Coefficiente de Correlación Lineal: 0.9595733968778531

Coefficiente de Determinación (R^2): 0.9207811039957019

La ecuación de la recta de regresión es: $\hat{y} = 0.6693x + (-13.3066)$

El coeficiente de correlación lineal, con un valor cercano a 1, indica una correlación lineal positiva y fuerte entre las variables “Edad” y “Experiencia laboral”. En otras palabras, existe una alta tendencia de las variables a moverse juntas en la misma dirección: cuando la edad aumenta, la experiencia laboral también tiende a aumentar de manera proporcional, y viceversa. Además, el coeficiente de determinación al ser cercano a 1, señala que el modelo puede predecir mejor los cambios en la experiencia laboral según la edad. Por lo tanto, se vuelve más confiable para decir cómo la experiencia laboral puede cambiar a medida que las personas envejecen.

Conclusiones:

Este análisis presenta una correlación positiva fuerte entre la edad y la experiencia laboral, esto significa que a medida que las personas en el campo tecnológico envejecen, tienden a acumular más experiencia laboral en términos de años trabajados en el sector. En este caso, se espera que los profesionales más viejos tengan una experiencia laboral mayor en tecnología.

Esta correlación puede ser útil para planificar programas de capacitación o desarrollo profesional. Si existe una fuerte correlación entre edad y experiencia laboral, las estrategias educativas podrían diseñarse considerando que a edades más avanzadas, se posee una mayor experiencia acumulada. También podría influir en las políticas de contratación o en la toma de decisiones para asignar roles o responsabilidades en función de la experiencia acumulada a lo largo del tiempo.

Es importante considerar que, aunque se encuentre una correlación significativa, existen otros factores que pueden influir en la experiencia laboral, como la formación educativa, el tipo de proyectos en los que se ha trabajado, las oportunidades laborales, entre otros. La correlación no

implica necesariamente una relación causal directa entre la edad y la experiencia laboral en tecnología.

3. Conclusiones

A partir de los análisis hechos en este trabajo se podrían ofrecer oportunidades de aprendizaje como:

- Ofrecer cursos cortos, talleres o certificaciones en áreas identificadas como más implementadas en los proyectos actuales para los estudiantes o profesionales en formación.
- Facilitar programas de mentoría donde estudiantes o profesionales en formación puedan ser orientados por expertos en las áreas específicas donde necesitan mejorar.
- Fomentar la participación en proyectos prácticos que permitan a los estudiantes aplicar los conocimientos adquiridos a situaciones del mundo real.
- Ofrecer oportunidades de aprendizaje continuo para graduados que deseen mantenerse actualizados en un entorno tecnológico en constante cambio.
- Establecer colaboraciones con empresas del sector para ofrecer programas educativos conjuntos, pasantías o programas de formación adaptados a las necesidades específicas de la industria.

4. Referencias

Mendenhall, W., Beaver, R., & Beaver, B. (2010). *Introducción a la probabilidad y estadística* (13^a ed.). McGraw-Hill.

IsaiasOR. (2023). IntegradorPE [Repositorio de GitHub]. Recuperado de <https://github.com/IsaiasOR/IntegradorPE>

Universidad Nacional de Entre Ríos (s. f.). Curso de créditos en la Facultad de Ciencias de la Alimentación (UNER) [Página web]. Recuperado de <https://www.fcad.uner.edu.ar/estudiantes/creditos-academicos/>