

2. ANÁLISIS DEL COMPORTAMIENTO DE POSTULACIÓN A LA CARRERA DE INGENIERÍA DE SISTEMAS DE LA UNIVERSIDAD NACIONAL DEL ALTIPLANO⁷

Behavior Analysis of Applying to the Systems Engineering Career at The National University of the Altiplano

Adolfo Carlos Jimenez Chura⁸

Charles Ignacio Mendoza Mollocondo⁹

Julio Cesar Villata Pacori¹⁰

Ralph Miranda Castillo¹¹

Martín Condori Concha¹²

Pares evaluadores: Red de Investigación en Educación, Empresa y Sociedad – REDIEES.¹³

⁷ Derivado del proyecto de investigación: *Análisis del comportamiento de postulación al programa de estudios de Medicina Humana a la Universidad Nacional del Altiplano, período 2018 - 2020*

⁸ Universidad Nacional del Altiplano, Puno-Perú, <https://orcid.org/0000-0003-3125-8197>, ajimenez@unap.edu.pe.

⁹ Universidad Nacional del Altiplano, Puno-Perú, <https://orcid.org/0000-0002-4766-2701>, cmendoza@unap.edu.pe.

¹⁰ Universidad Nacional del Altiplano, Puno-Perú, <https://orcid.org/0000-0002-9412-2222>, juvipal1@unap.edu.pe.

¹¹ Universidad Nacional Amazónica de Madre de Dios, Puerto Maldonado-Perú, <https://orcid.org/0000-0002-3510-1421>, rmiranda@unamad.edu.pe.

¹² Universidad Nacional del Altiplano, Puno-Perú, <https://orcid.org/0000-0003-4601-219X>, mcondoric@unap.edu.pe.

¹³ Red de Investigación en Educación, Empresa y Sociedad – REDIEES. www.rediees.org

ANÁLISIS DEL COMPORTAMIENTO DE POSTULACIÓN A LA CARRERA DE INGENIERÍA DE SISTEMAS DE LA UNIVERSIDAD NACIONAL DEL ALTIPLANO

Adolfo Carlos Jimenez Chura, Charles Ignacio Mendoza Mollocondo, Julio Cesar Villalta Pacori, Ralph Miranda Castillo, Martín Condori Concha.

RESUMEN

Toda universidad cuenta con una base de datos transaccional donde se almacena información de los postulantes que desean ingresar a una carrera específica. El análisis investiga factores como la cantidad de postulaciones, rango de edades, entre otros.

El objetivo principal es comprender el comportamiento de postulación y utilizar esta información para mejorar los procesos de admisión y la planificación académica en la universidad. Se usó el lenguaje de programación Python con las librerías Numpy, Pandas y Matplotlib que permitieron realizar el análisis de datos y las operaciones estadísticas de manera eficiente. El entorno de desarrollo integrado (IDE) fue Visual Studio Code. Se obtuvo la base de datos en un archivo csv de los procesos de admisión cepreuna y general comprendidos entre los años 2018 y 2022. Para el análisis estadístico se consideró de tres (3) a diez (10) postulaciones por postulante. Se tuvo 1838 postulantes (169 mujeres) y 124 ingresantes (19 mujeres), que contenían las variables de estudio como: nombres, tipo de proceso, género, fecha de nacimiento, tipo de colegio de la carrera de ingeniería de sistemas. Como resultado se obtuvo 256 postulantes que se presentaron 3 veces de los cuales obtuvieron una vacante 71 de ellos; un postulante se presentó hasta 10 oportunidades para finalmente obtener el ingreso a la universidad. Un 60.5% de ingresantes proceden de colegios estatales y oscilan en 14 y 18 años, frente a un 5.6% de colegios privados. Se concluye que el análisis estadístico y las capacidades de Python permitieron realizar una exploración exhaustiva de los datos y así obtener información relevante encontrando patrones y tendencias de los postulantes.

Palabras Clave: postulante; Python; análisis del comportamiento; librerías.

ABSTRACT

Every university has a transactional database where information on applicants who wish to enter a specific career is stored. The analysis investigates factors such as the number of applications, age range, among others.

The main objective is to understand the application behavior and use this information to improve the admission processes and academic planning in the university. The Python programming language was used with the Numpy, Pandas and Matplotlib libraries that allowed data analysis and statistical operations to be carried out efficiently. The integrated development environment (IDE) was Visual Studio Code. The database was obtained in a csv file of the Cepreuna and general admission processes between the years 2018 and 2022. For the statistical analysis, three (3) to ten (10) applications per applicant were considered. There were 1838 applicants (169 women) and 124 entrants (19 women), which contained the study variables such as: names, type of process, gender, date of birth, type of college of the systems engineering career. As a result, 256 applicants were obtained who applied 3 times, of which 71 of them obtained a vacancy; an applicant presented up to 10 opportunities to finally gain admission to the university. 60.5% of entrants come from state schools and range from 14 to 18 years old, compared to 5.6% from private schools. It is concluded that the statistical analysis and the capabilities of Python allowed an exhaustive exploration of the data and thus obtain relevant information, finding patterns and trends of the applicants.

Keywords: applicant; python; behavior analysis; library.

INTRODUCCIÓN

El proceso de análisis de datos estadísticos consiste en interpretar la información recopilada con el fin de tomar decisiones eficientes en cualquier organización (Ortiz Aguilar et al., 2021). Existen diversas técnicas de análisis que abarcan desde el cálculo de medidas de tendencia central hasta la utilización de modelos estadísticos para inferir relaciones dentro de los datos. En los últimos años, el aprendizaje automático (machine learning) ha surgido como una técnica de inteligencia artificial ampliamente utilizada en diversos contextos, incluyendo los procesos de admisión de una universidad (Unsihuay y Flores, 2022). Esta técnica involucra y aprovecha los datos, los cuales se describen y se utilizan dentro de un marco estadístico. Aunque muchos modelos estadísticos pueden realizar predicciones, no todos logran hacerlo con precisión. Para alcanzar la precisión deseada, es necesario crear o utilizar algoritmos específicos que cumplan con esta tarea. Por ejemplo, en el caso de predecir el precio de una vivienda basado en múltiples variables, se requiere un enfoque orientado al machine learning.

En el ámbito universitario, es posible emplear estas metodologías para examinar grandes volúmenes de datos relacionados con los estudiantes y predecir la probabilidad de que un estudiante tenga éxito académico en la universidad. Por ejemplo, un modelo basado en aprendizaje automático puede analizar información sobre las calificaciones de los estudiantes, sus actividades extracurriculares y otros factores relevantes para anticipar la probabilidad de que un estudiante logre obtener su grado en un período de tiempo específico. El estudio del comportamiento de postulación a la carrera de Ingeniería de Sistemas implica la aplicación de una serie de técnicas estadísticas para analizar las variables de estudio, tales como género, fecha de nacimiento, tipo de colegio y cantidad de postulaciones. El propósito de este estudio es contribuir con conocimientos valiosos para la universidad, permitiendo comprender mejor los patrones y tendencias de los postulantes.

Python, como lenguaje de programación, ha adquirido un amplio reconocimiento en diversos ámbitos, como el académico y la investigación tanto en el sector privado como público. Además, ha ganado popularidad en comunidades de software libre debido a su facilidad de uso y su capacidad para abordar una amplia gama de problemas. Este lenguaje

ofrece ventajas como la portabilidad y el acceso a una variedad de bibliotecas necesarias para realizar tareas de preparación, modelado, análisis de datos y visualización (Manrique, 2020).

MATERIAL Y MÉTODOS

En esta investigación se recopilaron datos de postulantes a la carrera de Ingeniería de Sistemas de la universidad Nacional del Altiplano. Estos datos incluyeron información de nombres, fecha nacimiento, lugar de procedencia, género y cualquier otra variable relevante para el estudio. Se utilizó Python debido a su popularidad y versatilidad en el análisis de datos. Python proporciona una amplia variedad de bibliotecas especializadas, como NumPy, pandas y matplotlib, que facilitan la manipulación y visualización de datos (Rivas y Castillo, 2022). Se seleccionó Visual Studio Code ya que ofrece una interfaz intuitiva y personalizable, lo que facilita la escritura, ejecución y depuración de código Python.

Se llevó a cabo un proceso de preparación y limpieza de los datos recopilados para asegurar su calidad y consistencia. Esto incluyó la eliminación de valores nulos o inconsistentes, la estandarización de formatos y cualquier otra transformación necesaria para su posterior análisis. Se aplicaron técnicas de análisis estadístico a los datos utilizando bibliotecas. Esto incluyó el cálculo de medidas descriptivas, la exploración de las relaciones entre variables pertinentes para el estudio. Utilizando la biblioteca matplotlib, se generaron gráficos y visualizaciones que ayudaron a comprender y presentar los resultados del análisis. Estas visualizaciones incluyeron histogramas y cualquier otro tipo de gráfico relevante para mostrar los patrones y tendencias identificados (Rincón Valencia, 2022).

Finalmente, se interpretaron los resultados obtenidos a través del análisis de datos. Se examinaron las relaciones identificadas, las tendencias observadas y cualquier otro hallazgo significativo con respecto al comportamiento de postulación a la carrera de Ingeniería de Sistemas.

RESULTADOS

La población de estudio se consideró a los postulantes con mayor de dos (2) postulaciones a la carrera profesional de Ingeniería de Sistemas, obteniendo un total de 1838 postulantes y 124 ingresantes de los procesos cepreuna y general entre los años 2018 y 2022.

Tabla 1.

Cantidad de postulantes e ingresantes según tipo de colegio y género

Tipo de colegio	Postulantes		Ingresantes	
	Público	Privado	Público	Privado
	1657	181	110	14
Género	Masculino	Femenino	Masculino	Femenino
	1669	169	105	19

Nota. Esta tabla muestra las cantidades por tipo de colegio, género de los postulantes de los procesos cepreuna y general entre los años 2018-2022. Elaboración propia.

Para el entorno de trabajo se usó Python en Visual Studio Code. Ambas herramientas facilitaron la exploración, limpieza, análisis de datos y presentación de resultados.

La investigación es de tipo descriptiva (Sampieri y Mendoza, 2018). Se aplicó un análisis estadístico descriptivo con las variables género, año de egreso, cantidad de postulaciones, tipo de colegio de procedencia y edad que tuvo al momento de postular. Los datos recolectados fueron preprocesados, modelizados y desplegados (Alvaro, 2020) utilizando las cuatro librerías de Python para ciencia de datos más usadas: Numpy, Pandas, Matplotlib, Seaborn, entre otras.

Para la carga de datos se importó a partir de un formato en csv donde se consignaba las variables a ser estudiadas (categóricas y numéricas). Se realizó un análisis de los posibles datos atípicos (outliers) que distorsionaban los datos de las variables, esto fue posible realizando un análisis del rango intercuartílico (IQR) (Anchino et al., 2019), aquellos datos fueron reemplazados con la media. La variable edad y año de egreso fue discretizada usando la función cut de la librería pandas de forma personalizada:

```
edad["edad3"] = pd.cut(edad["edad2"],
                       bins = [14, 18, 22, 27, 32, np.inf],
                       labels = ["14-18", "18-22", "22-27", "27+"])
```

Respecto a la cantidad de postulaciones se ha considerado sólo a los que se presentaron más de dos veces en los diferentes procesos de admisión cepreuna y general.

Se implementó el siguiente script para la presentación de los datos usando la función countplot de la librería seaborn con la variable edad ya discretizada.

```
plt.figure(figsize=(8,5))
ax = sns.countplot(var, data=edad)
plt.title(titulo)
plt.xlabel(xtitle)
ax2=ax.twinx()
ax2.yaxis.tick_left()
ax.yaxis.tick_right()
ax.yaxis.set_label_position('right')
ax2.yaxis.set_label_position('left')
ax2.set_ylabel(labLeft)
for p in ax.patches:
    x=p.get_bbox().get_points()[:,0]
    y=p.get_bbox().get_points()[1,1]
    ax.annotate('{:.1f}%'.format(100.*y/len(data)), (x.mean(), y),
               ha='center', va='bottom')
ax.yaxis.set_major_locator(ticker.LinearLocator(11))
ax2.set_ylim(0,100)
ax.set_ylim(0,len(edad))
ax2.yaxis.set_major_locator(ticker.MultipleLocator(10))
```

Para establecer la cantidad de postulaciones de los postulantes e ingresantes se creó el siguiente script con la librería matplotlib.

```

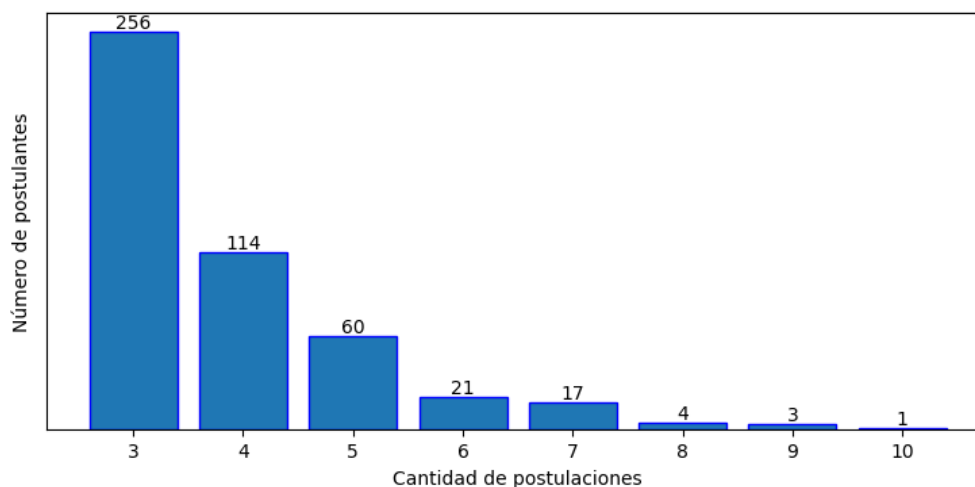
plt.clf()
xs = np.arange(len(data.cantPost))
ys = data.cant
plt.figure(figsize = (9,4))
plt.bar(xs,ys, edgecolor = "b")
for x,y in zip(xs,ys):
    plt.annotate(y, (x,y),
                 textcoords="offset points",
                 xytext=(0,2),
                 fontsize=10,
                 ha='center')
plt.xticks(xs, data, color='black', fontsize=10)
plt.yticks([])
plt.xlabel(countPost, fontsize=10)
plt.ylabel(numberPost, fontsize=10)
plt.title(titlePost, fontsize=10)
plt.show()

```

En la siguiente figura se puede apreciar la cantidad de postulaciones que realizaron los postulantes entre los años 2018 – 2022.

Figura 1.

Cantidad de postulantes por cantidad de postulaciones al cepreuna y general 2018 – 2022.



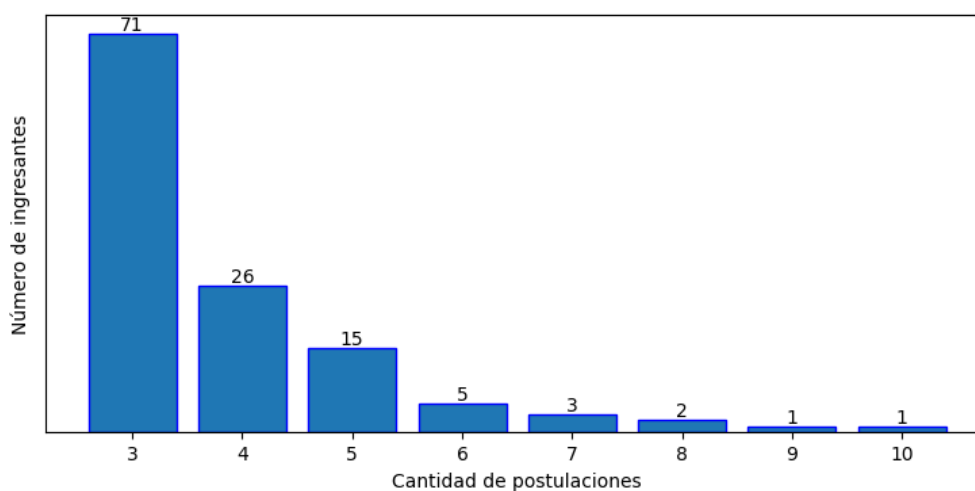
Nota. Elaboración propia.

En general, 476 postulantes se presentaron entre tres (3) y diez (10) veces a rendir el examen de admisión a la carrera de Ingeniería de Sistemas. Se puede apreciar que 256 postulantes se presentaron 3 veces y un (1) postulante hasta 10 veces.

La siguiente figura muestra la cantidad de ingresantes que realizaron varios intentos para obtener una vacante a lo largo de los procesos realizados durante el 2018 – 2022.

Figura 2.

Cantidad de ingresantes por cantidad de postulaciones al cepreuna y general 2018 - 2022

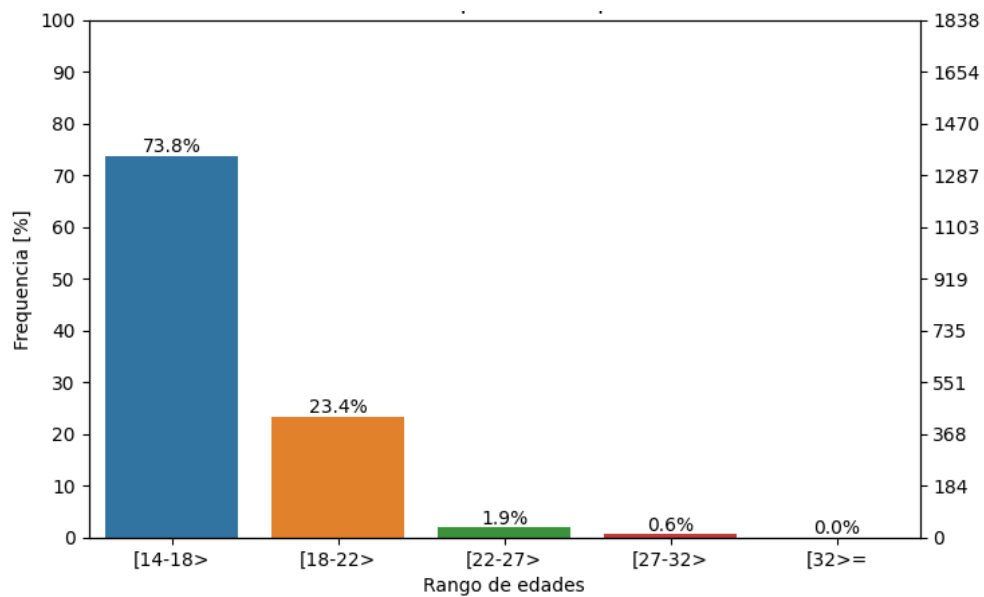


Nota. Elaboración propia.

En la figura 2 se contabiliza que 124 ingresantes se presentaron entre tres (3) y diez (10) veces a rendir el examen obteniendo una vacante de ingreso. También se puede apreciar que 71 postulantes en su tercer intento lograron una vacante y, que como máximo 1 postulantes en su intento número 10 obtuvo una vacante.

Figura 3.

Distribución de postulantes por edades.

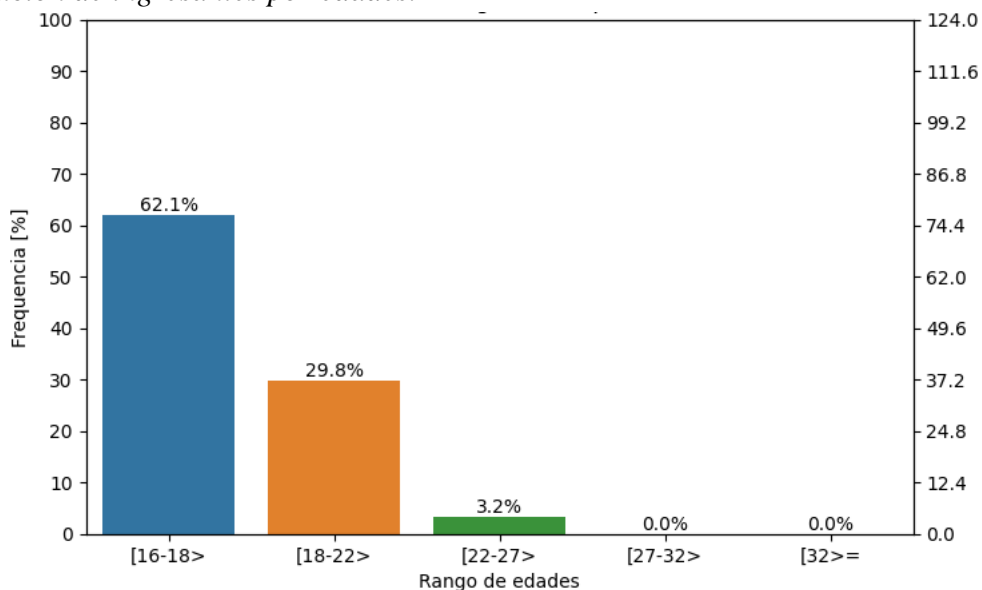


Nota. Elaboración propia.

En la figura 3, se aprecia un pico que corresponde a los intervalos de edades entre 14 y 18 años. Esto indica que el 73.8% de postulantes son menores de edad y 0.6% son mayores a 27 años.

Figura 4.

Distribución de ingresantes por edades.

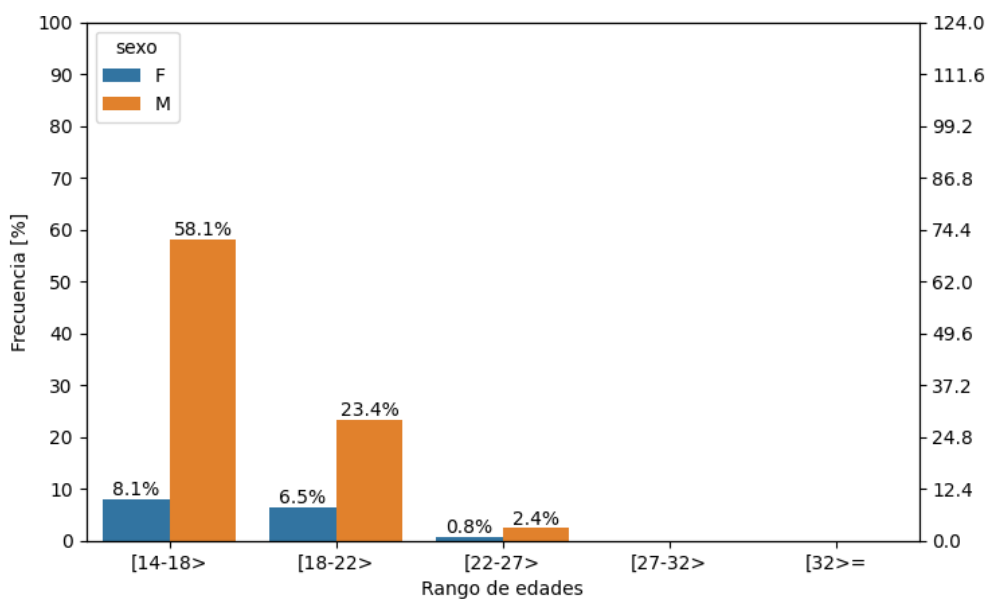


Nota. Elaboración propia.

En la figura 4, se observa que el 62.1% de ingresantes tiene una edad que está comprendido entre 16 y 18 años, mientras que existe un mínimo de 3.2% de ingresantes tiene una edad comprendida entre 22 y 27 años.

Figura 5.

Distribución de ingresantes por edad y género.

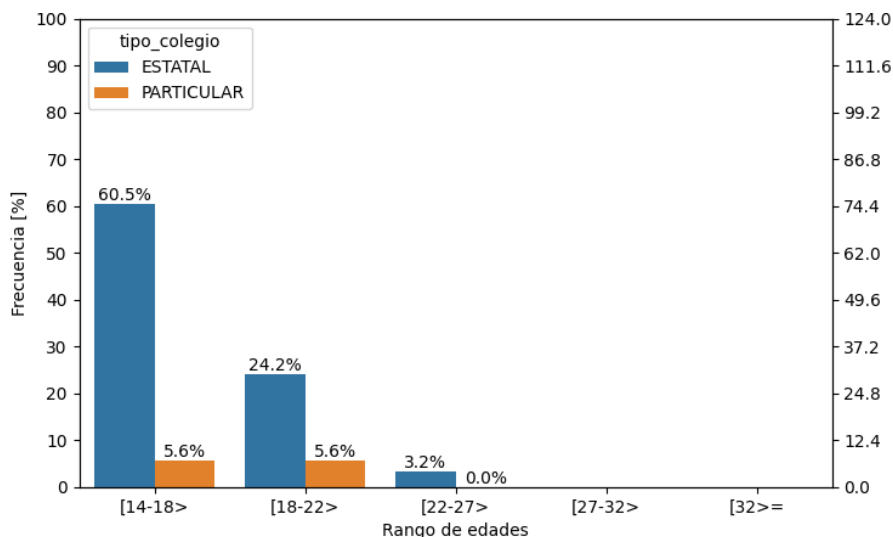


Nota. Elaboración propia.

En la figura 5, se puede apreciar que la mayor cantidad de ingresantes son del género masculino. El 58% ingresan con una edad comprendidos entre 14 y 18 años. En general la mayor cantidad de ingresantes son masculinos.

Figura 6.

Distribución de ingresantes por edad y tipo de colegio.



Nota. Elaboración propia.

En la figura 6, la mayor cantidad de ingresantes provienen de un colegio público. El 60.5% de ingresantes provienen de un colegio estatal, 5.6% provienen de un colegio privado y sus edades están entre 14 y 18 años. El 3.2% provienen de una institución pública y sus edades están comprendidos entre 22 y 27 años.

DISCUSIÓN Y CONCLUSIONES

Al emplear Python en Visual Studio Code, se puede realizar un procesamiento eficiente de los datos y aplicar diversas técnicas de análisis de datos para identificar patrones, tendencias y correlaciones relevantes con su amplia gama de bibliotecas y paquetes específicamente diseñados para el análisis de datos, manipulación, visualización y modelado de los datos recopilados.

Visual Studio Code, por su parte, proporciona un entorno de desarrollo integrado (IDE) poderoso y versátil que permite escribir, depurar y ejecutar código Python de manera

eficiente facilitando el proceso de análisis de datos en investigación aplicando algoritmos a problemas de nuestro quehacer diario tal como indica Bustamante y Useche (2022).

A partir del análisis de los datos los resultados obtenidos son válidos y únicos para la carrera de Ingeniería de Sistemas de la Universidad Nacional del Altiplano de Puno. De los ingresantes se concluye que el 88% provienen de instituciones públicas y 12% de instituciones privadas; el 85% son del género masculino frente a un 15% del género femenino; para obtener una vacante como máximo se presentaron hasta diez (10) oportunidades.

REFERENCIAS BIBLIOGRÁFICAS

- Alvaro. (27 de junio de 2020). *El ciclo de vida de la ciencia de datos*. [Artículo en página web]. [El ciclo de vida de la ciencia de datos \(machinelearningparatodos.com\)](https://machinelearningparatodos.com)
- Bustamante Narváez, L. F. y Rodríguez Useche, V. M. (2022). *Resolución de problemas a través de programación en Python centrado en la Teoría de Situaciones Didácticas dirigido a estudiantes de grado noveno*. [Tesis de maestría]. Repositorio Universidad de Antioquia. <https://hdl.handle.net/10495/30755>
- Caymes Scutari, P., Tardivo, M. L., Blanchini, G. y Méndez Garabetti, M. (2019). IQR: una medida estadística como modelo para la sintonización computacional. *XXI Workshop de Investigadores En Ciencias de La Computación*, 971–974. [IQR: una medida estadística como modelo para la sintonización computacional \(unlp.edu.ar\)](https://unlp.edu.ar)
- Hernández-Sampieri, R. y Mendoza Torres, C. P. (2018). *Metodología de la investigación: las rutas cuantitativa, cualitativa y mixta*. McGraw-Hill.
- Manrique Rojas, E. (2020). Machine Learning: análisis de lenguajes de programación y herramientas para desarrollo. *RISTI*, (28), 586-599. [Machine Learning: análisis de lenguajes de programación y herramientas para desarrollo - ProQuest](https://proquest.com/docview/258444444)
- Ortiz Aguilar, W., Ortega Chávez, W., Valencia Cruzaty, L. E., González Vásquez, Á. E., Gamarra Mendoza, S., Ortiz Aguilar, W., Ortega Chávez, W., Valencia Cruzaty, L. E., González Vásquez, Á. E. y Gamarra Mendoza, S. (2021). La educación estadística del ingeniero: reto de la educación superior. *Revista Universidad y Sociedad*, 13(5), 307–318. <https://bit.ly/3JdnIe>
- Rincón Valencia, K. J. (2022). Desarrollo de un Prototipo de Software en Python con Técnicas de Machine Learning para el Análisis de Datos Astronómicos de Exoplanetas Recopilados por la NASA. *Perfiles*, 1, 4-10. <https://doi.org/10.47187/perf.v1i27.142>

Rivas, J. G. R. y Castillo, S. R. (2022). Uso de Python para el análisis de datos aplicado en la investigación. *Investigación y Ciencia Aplicada a La Ingeniería*, 5(34), 33–40.
<https://bit.ly/3Xd4qpi>

Unsihuay, J. E. G. y Flores, J. W. S. (2022). Prediction of academic situation in undergraduate students using machine learning algorithms. *Perfiles*, 1(27), 4–10.
<https://doi.org/10.47187/PERF.V1I27.142>