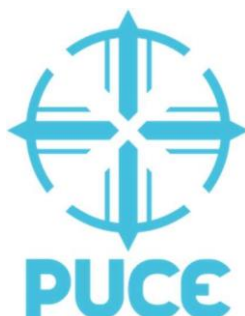


Pontificia Universidad Católica Del Ecuador
Facultad de Ingeniería



Trabajo previo a la obtención del título de Magister en Sistemas de Información mención Data Science.

Tema

Factores que influyen en la formación de jóvenes sisi en Ecuador durante el año 2023 mediante el análisis clúster.

Autor

Barahona Ocaña Henry Fabricio

Tutor

Msc. Damián Aníbal Nicolalde Rodríguez

Quito, Marzo – 2025

Dedicatoria

A Dios, por ser mi guía en los momentos difíciles.

A mis padres y hermanos, cuyo apoyo incondicional me permitió culminar esta etapa.

Tampoco puedo dejar de lado a mis compañeros de clase y maestros, quienes supieron transmitir sólidos conocimientos, contribuyendo a la formación de profesionales con criterio propio y capaces de tomar decisiones fundamentadas.

Con su experiencia, dedicación, tiempo y exigencia, me ayudaron a formarme como profesional.

Agradecimiento

A Dios, por ser mi fortaleza en todo momento.

A mis padres mi gratitud, Octavio B. y Celia O., por su apoyo incondicional a lo largo de este camino.

A mis maestros, por su paciencia y dedicación en la enseñanza, reconociendo en cada uno de ellos el valioso y arduo trabajo del docente.

Este logro no habría sido posible sin el respaldo y la motivación de mi tutor durante el desarrollo de la tesis.

Finalmente, expreso mi más sincero agradecimiento a la Pontificia Universidad Católica del Ecuador, por brindarme la oportunidad de formarme como profesional.

Resumen

Las herramientas de análisis de datos son actualmente concebidas como primordiales en diferentes campos o ciencias, una de ellas las sociales. Históricamente los datos han permitido conocer el comportamiento de cierto grupo o grupos cohesionados y a partir de ello, los datos elevados a la categoría información han permitido la creación y aplicación de políticas públicas en pro de los diversos grupos de interés y de los 5 ejes principales que por lo general tiene un gobierno. Es frecuente que los jóvenes estudien y trabajen a la vez debido a limitaciones económicas o ideológicas, lo que los obliga a ser autosuficientes. Esta investigación busca aplicar la ciencia de datos en la búsqueda y definición de los factores que influyen en la formación de jóvenes sisi.

Abstract

Data analysis tools are currently conceived as essential in various fields or sciences, one of them being the social sciences. Historically, data has allowed us to understand the behavior of certain group or cohesive groups, and from this, data elevated to the category of information has enabled the creation and implementation of public policies in favor of the various stakeholders and the 5 main axes that a government usually has. It is frequent that young people study and work at the same time due to economic or ideological limitations, which forces them to be self-sufficient. This research seeks to apply data science in the search and definition of the factors that influence the formation of 'NiNi' youth

Tabla de Contenidos

Capítulo I: Introducción	12
1.1. Introducción	12
1.2. Planteamiento del problema.....	13
1.3. Objetivos	15
1.3.1. Objetivo general.....	15
1.3.2. Objetivos específicos	15
1.4. Justificación.....	15
1.5. Alcance.....	17
1.6. Unidad de análisis	18
Capítulo II: Marco Teórico y Conceptual	19
2.1. Antecedentes	19
2.2. Marco teórico	20
2.2.1. Capital humano	20
2.2.2. La juventud ecuatoriana.....	23
2.2.3. La educación y los jóvenes	23
2.2.4. Empleo juvenil	25
2.2.5. Mercado laboral en el Ecuador	29
2.2.6. Transición de los jóvenes de la escuela al mercado laboral.....	30
2.2.7. Jóvenes que ni estudian ni trabajan “NINI”	32
2.2.8. Jóvenes que estudian y trabajan “SISI”	34
2.2.9. El clúster análisis	36
2.3. Marco conceptual	38

2.3.1. Análisis clúster	38
2.3.2. Jóvenes nini	38
2.3.3. Jóvenes Sisi	38
2.3.4. Empleo	39
2.3.5. Educación	39
2.3.6. Índice de silhoutte	39
3.1. Instrumentos de investigación	41
3.1.1. Análisis descriptivo	41
3.1.2. K means o cluster jerarquico	43
3.1.3. Análisis de componentes principales	45
3.1.4. Índice de silhouette	50
3.2. Fuente de datos	53
3.3. Modalidades de investigación	55
3.3.1. Bibliográfico - documental	55
3.3.2. De escritorio	55
3.4. Limitaciones metodológicas	55
Capitulo IV: Resultados	58
4.1. Comprensión de los datos	58
4.2. Recopilación de los datos iniciales	59
4.3. Exploración de los datos	59
4.3.1. Análisis factorial de correspondencias	69
4.3.2. Índice de silhouette	76
Capítulo V: Conclusiones y Recomendaciones	80

Conclusiones	80
Recomendaciones.....	81
Referencias.....	84
ANEXO.....	96

Índice de Tablas

Tabla 1 Matriz de componentes principales	49
Tabla 2 Valores calculados	52
Tabla 3 Librerías utilizadas.....	54
Tabla 4 Estado civil de los jóvenes según sexo declarado.....	60
Tabla 5 Porcentaje de estado civil de los jóvenes según sexo declarado.....	60
Tabla 6 Clasificación de los jóvenes según estado civil y edad	61
Tabla 7 Porcentaje de Clasificación de los jóvenes según estado civil y edad.....	62
Tabla 8 Etnia de los jóvenes según sexo declarado	63
Tabla 9 Porcentaje de etnia de los jóvenes según sexo declarado	63
Tabla 10 Estadísticos de edad	64
Tabla 11 Grupos de ocupación	64
Tabla 12 Determinación de k-medias	66
Tabla 13 K-medias	66
Tabla 14 Medias por grupos	67
Tabla 15 Caracterización de clusters	67
Tabla 16 Análisis factorial de correspondencias	69
Tabla 17 Eigen valores.....	71
Tabla 18 Ponderación de filas.....	72
Tabla 19 Ponderación columnas	73
Tabla 20 Eigenvalores.....	75
Tabla 21 Individuos	75
Tabla 22 Variables	76

Tabla 23 Clústers	77
-------------------------	----

Índice de Figuras

Figura 1 <i>Conceptos principales</i>	40
Figura 2 <i>Elección de clusters</i>	45
Figura 3 <i>Pasos para la selección de componentes principales</i>	48
Figura 4 <i>Plano factorial</i>	49
Figura 5 <i>Algoritmo para la agrupación de clúster</i>	52
Figura 6 <i>Aproximación gráfica a el agrupamiento jerárquico con k.medias. Heatmap</i> .	65
Figura 7 <i>Gráfico CPA</i>	71
Figura 8 <i>Gráfico variables-CPA</i>	73
Figura 9 <i>Indice de silhouette</i>	78
Figura 10 <i>Indice de Dunn</i>	78

Capítulo I: Introducción

1.1. Introducción

El clúster análisis es una herramienta útil desde incluso antes de formalizarse como un instrumento científico, el agrupamiento de poblaciones, siempre se da desde el contexto de características homogéneas entre individuos y han servido para clasificar a la sociedad por grupos en función de sus ingresos, su etnia, su ubicación geográfica y demás. Con el tiempo y con la formalización de esta técnica, la misma se ha tornado de suma relevancia en todas las áreas de investigación, uno de ellas el análisis de poblaciones pares, pero con realidades distintas como jóvenes que estudian y trabajan donde incluso, dentro de esta misma se pueden formar subgrupos.

Este estudio de investigación tiene como objetivo principal determinar los elementos que influyen en los grupos de jóvenes sisi en el Ecuador que permitan comprender a este grupo muy poco estudiado dentro del contexto del país, estos a través de metodologías estadísticas que permiten ratificar la validez de los clúster formados.

En el primer capítulo resume los objetivos y propósitos del trabajo de investigación dando visibilidad al problema y la importancia del mismo. Mientras tanto en el segundo capítulo se buscan bases sólidas teóricas que refuerzan el espectro de la investigación, bases y principios fundamentales de las ciencias de datos, así también teorías sociales relacionadas a la importancia de los jóvenes en la estructura social antigua y actual. El tercer capítulo se centra en la construcción metodológica a llevar dentro de la investigación, aquí se indican las técnicas e instrumentos que se usan el capítulo precedente de resultados; por último, se llega a las conclusiones y recomendaciones.

1.2. Planteamiento del problema

La terminología joven sisi hace alusión a aquellos jóvenes que estudian y trabajan al mismo tiempo. Este término se utiliza para describir un grupo específico de la población juvenil que se enfrenta a las responsabilidades, por un lado, la necesidad de continuar su educación formal y, por otro, la obligación de contribuir económicamente a sus familias o adquirir experiencia laboral.

En años precedentes, varias exploraciones han centrado su atención en el puente de los jóvenes a el periodo de adultez, prestando gran atención a su inclusión en el mercado profesional. La relevancia de estos estudios radica en hallazgos tales como el número considerable de jóvenes interrumpen su formación educativa al concluir la educación básica, optando por ingresar al ámbito laboral. Así mismo, se identifica a un sector juvenil que se encuentra fuera del sistema educativo y del mercado laboral, o que combina ambas actividades, lo que define a los grupos conocidos como jóvenes 'sisi' y 'nini'.

Los jóvenes constituyen un sector de la población heterogéneo; su acceso al mercado laboral o desarrollo académico están influenciados por características personales, grupales y las oportunidades que les ofrece la colectividad. La población de jóvenes sisi son un fenómeno existente en todos los países, mucho más en aquellos que se conocen como subdesarrollados sumado a ello que no existe un verdadero interés por una población con una formación alta, pero existe la necesidad de una fuerza laboral joven que aún emplee la fuerza como medio de producción y de ingreso personal, acción de la que Ecuador no está exenta.

En Ecuador, los jóvenes sisi presentan características y comportamientos que requieren una comprensión detallada para abordar eficazmente los desafíos asociados. Sin embargo, la falta de un análisis detallado de los factores que influyen en su formación limita la efectividad de las

políticas y programas existentes, por consecuencia la falta de precisión al momento de identificar dichos factores puede llegar a demeritar las intervenciones de origen gubernamental o no gubernamental con el fin de componer una estructura ideal donde los jóvenes sean capaces de valerse por sus propios medios sin la necesidad de desertar de los estudios sino más bien buscar una escenario ideal donde exista flexibilidad del empleador y de la academia para que los jóvenes se sustenten y estudien a la par.

Según menciona Ramírez (2022), para el caso ecuatoriano:

La probabilidad que los jóvenes decidan solo estudiar, solo trabajar, estudiar y trabajar o no trabajar ni estudiar se atribuye a las características propias de los jóvenes y de su hogar. Así, variables como: área de residencia, edad, género, estado civil, tienen mayor influencia en la participación o ausencia en actividades educativas o laborales de los jóvenes. (p. 39)

El desarrollo de los jóvenes sisi en Ecuador durante el año 2023 es un fenómeno que ha generado considerable preocupación y debate debido a sus implicaciones profundas y variadas. Los jóvenes sisi, definidos por sus comportamientos y características distintivas, representan un segmento de la población con patrones y necesidades específicas que pueden influir en su integración y desarrollo social. Comprender estos patrones y las causas subyacentes es esencial para diseñar políticas públicas y estrategias de intervención efectivas. Por ejemplo una intervención en la reformulación de las prácticas preprofesionales que permitan al joven desarrollarse realmente dentro de su ámbito de estudio, esto basado en un salario digno casi equiparable al de un empleado común o que además se creen beneficios para las empresas que empleen gente joven como promoción de la empleabilidad juvenil y del uso del talento humano como un recurso en constante capacitación y entrenamiento.

1.3. Objetivos

1.3.1. Objetivo general

Analizar los factores que influyen en la formación de jóvenes sisi en Ecuador durante el año 2023 mediante el análisis clúster para comprender mejor las características y patrones asociados.

1.3.2. Objetivos específicos

- Compilar datos sobre características sociodemográficas, educativas y psicológicas tales como edad, nivel de instrucción, etnia, ubicación geográfica, tipo de empleo, sexo de jóvenes sisi con el fin de identificar los perfiles económicos y sociales de la población objetivo a través de tablas de frecuencias, porcentajes, medidas de tendencia central, gráficos.
- Identificar el grupo de jóvenes denominados como sisi de acuerdo a su característica sociodemográfica a través de las técnicas de clustering como K-means o clustering jerárquico obtenido mediante índice de Silhouette para determinar el número óptimo de clusters.
- Determinar los factores predominantes en cada cluster que permitan comprender su influencia en la formación de jóvenes sisi a través de métodos multivariantes como el análisis de componentes principales.
- Evaluar la validez y la precisión de los clusters y los factores identificados mediante técnicas de validación como el índice de Silhouette.

1.4. Justificación

La formación de jóvenes sisi en Ecuador es un fenómeno complejo que requiere una comprensión profunda de los factores subyacentes. El análisis clúster ofrece una metodología

robusta para segmentar y analizar estos jóvenes en función de características comunes, facilitando la identificación de patrones y factores influyentes. Esta perspectiva posibilitará que las políticas públicas y los programas de intervención desarrollen estrategias más a medida, atendiendo a las necesidades únicas de cada grupo de jóvenes 'sisi'.

En las ciencias sociales, el análisis clúster ha sido una herramienta ampliamente utilizada para agrupar poblaciones y descubrir patrones en datos complejos. Por ejemplo, en estudios de pobreza, el clustering ha permitido identificar subgrupos de hogares que comparten características socioeconómicas similares, facilitando así la formulación de políticas.

Otro aspecto relevante se encuentra en la investigación sobre el comportamiento juvenil, donde el análisis clúster ha sido utilizado para categorizar a los jóvenes según sus estilos de vida, actitudes y comportamientos de riesgo. Esto ha permitido a los responsables de políticas diseñar programas de intervención que abordan necesidades específicas, como prevención de violencia o promoción de la salud mental.

Por lo tanto, el análisis clúster ha demostrado ser efectivo en estudios sobre migración, donde se ha empleado para segmentar migrantes en función de factores como motivaciones, perfiles demográficos y destinos preferidos, mejorando la comprensión de las dinámicas migratorias y la formulación de estrategias adecuadas.

En el contexto de los jóvenes sisi en Ecuador, aplicar esta metodología no solo ofrece una justificación robusta para segmentar a esta población, sino que también brinda la oportunidad de construir en base a evidencias y datos concretos sobre las características y necesidades de cada grupo. Esto puede resultar en un uso más eficiente de los recursos y un mayor impacto en la optimización del bienestar de los jóvenes. Dentro de los objetivos de desarrollo del gobierno nacional, este estudio puede contribuir a la modificación de las políticas de empleo juvenil, con

el fin de incrementar los salarios de quienes se incorporan al mercado laboral por primera vez y que a la par vayan acompañados de una capacitación continua que impulse y desarrolle sus capacidades haciendo que la población joven se convierta realmente en una fuerza laboral preparada y no solo en aquella que se limita a un salario mínimo y a un desarrollo deficiente de sus capacidades.

1.5. Alcance

Los estudios sociológicos han recurrido extensamente al análisis de clústeres con el fin de dividir a las poblaciones en grupos diferenciados y comprender sus patrones de conducta particulares. En el contexto de jóvenes, estudios previos han aplicado técnicas de clustering para identificar patrones de comportamiento y factores influyentes. Sin embargo, la aplicación de estas técnicas en el contexto de jóvenes sí es limitada, lo que destaca la relevancia de este estudio.

En sí, la aplicabilidad del análisis clúster se ve limitada por la carencia del manejo o intervención focalizada para los diferentes grupos pues las políticas públicas no se centran en cada uno sino suele tornarse generalizada. Aspecto similar lo describe Franco (2022), pues enuncia que la política pública es compleja de manejar debido a el gran tamaño de las situaciones que intenta abordar y resolver. Por este motivo las políticas públicas pueden apoyarse en mecanismos que orienten las acciones y los recursos asegurando el cumplimiento de los objetivos. Sin embargo, mucha de estas iniciativas se ven cuartadas por la falta de políticas focalizadas en un sector diferente al productivo de manera diferenciada, sino que también involucre a actores locales y existan condiciones necesarias para la incorporación de nuevos participantes a través del desarrollo de políticas públicas innovadoras que se centren en los territorios y en la transferencia de conocimientos.

El contexto de la investigación se da desde el entorno nacional donde a través de información secundaria de fuentes oficiales se analizará el fenómeno de estudio con la metodología de clúster de ciencias de datos.

1.6. Unidad de análisis

Este estudio se enfoca en la población de jóvenes ecuatorianos que tienen entre 18 y 29 años.

Capítulo II: Marco Teórico y Conceptual

2.1. Antecedentes

La investigación sociológica y el estudio del comportamiento juvenil han encontrado en el análisis de clústeres una herramienta de gran valor. La aplicación específica de estas técnicas al fenómeno de los jóvenes sisi representan una oportunidad para avanzar en la comprensión de este grupo particular y desarrollar estrategias de intervención más efectivas. Este estudio busca llenar una brecha en la investigación al aplicar técnicas de clustering para analizar y entender los elementos que determinan el desarrollo de jóvenes sisi en Ecuador.

Ward (2014), en el estudio denominado *"Agrupamiento jerárquico para optimizar una función objetivo."* *Revista de la Asociación Estadounidense de Estadística.*", indica el uso efectivo de cluster para crear comunidades de grupos bacterianos en el lodo esto a través de análisis estadísticos y de técnicas tales como el índice de diversidad de Shanon y análisis de escalamiento óptimo, al igual se aplicaron técnicas multidimensionales, incluyendo el análisis de factores principales, y los resultados demostraron que la integración de la estadística en la biología, facilita la identificación de comunidades bacterianas.

Otro estudio donde se utilizan cluster es el realizado por Hassan (2017), en el estudio denominado *"Clustering of social media users: Analyzing behavior and interactions"* *Social Networks*. En este estudio se examina el clúster en redes sociales donde se decidió analizar las redes estimando a la cantidad de participación de los usuarios dentro de estas plataformas, aquí se agrupó a los usuarios en función de sus interacciones.

En el ámbito social Kapila y Saraj (2023), realizan un análisis de agrupamiento de los estudiantes en relación de sus factores socioeconómicos tomando además variables como género, educación, ocupación de los padres y demás. La metodología que se utiliza en este estudio es el

de k-medias, homogeneidad intracluster y heterogeneidad intercluster además se utilizan pruebas estadísticas para determinar la significancia de los cluster. Esta investigación contribuye a la formulación de políticas para reducir las brechas en el logro educativo de los sectores desfavorecidos de la sociedad.

En un estudio realizado por Salgado et al. (2009), se indica que a partir de las técnicas de clúster es posible la comprensión de los individuos, su conducta, sus emociones y control de las mismas; además en el ámbito psicológico. Por otra parte, los clústers sirven también para determinar el movimiento de las organizaciones, pues para esta se desarrolle es necesario un amplio interés de la comunidad académica con una amplia gestión de difusión y apoyo para el desarrollo de la diversificación del conocimiento de la utilidad de la elaboración de clúster (Perez et al, 2009). Realmente las investigaciones de clúster inician dentro del áreas de botánica y biología, debido a la necesidad de agrupar plantas y animales, pero puede ser utilizada en cualquier otro tipo de ámbitos, uno de ellos el marketing en el cual su principal función es la segmentación de mercados (Gomez, 2002).

2.2. Marco teórico

2.2.1. *Capital humano*

El recurso humano está determinado como el saber de cada persona que adquiere, cultiva y acumula en el transcurso de su formación académica y experiencia en el mercado laboral junto con las cualidades personales que tiene, por ejemplo la lealtad, versatilidad, y adaptabilidad, entre otras, puede influir en el valor que el individuo aporta a la empresa (Vasquez et al., 1998).

El capital humano se compone de elementos intangibles como conocimientos, valores, habilidades y actitudes, lo que dificulta su identificación y medición a través de métodos cuantitativos, aunque una gestión efectiva implica conocer de manera oportuna y precisa los

activos que la componen, su cantidad y su comportamiento a lo largo del tiempo, el conocimiento, comprendido como la formación, se considera el elemento primordial, seguido por las habilidades, que engloban las destrezas, y posteriormente las actitudes y el aprendizaje. Sin embargo, se observa una falta de atención hacia las características demográficas del personal y una notable subestimación de los valores en la mayoría de los estudios (Gallego & Naranjo, 2020).

La formación en Ecuador desde un análisis longitudinal muestra que de 2000 a 2018 ha existido un incremento respecto a los años promedio de escolaridad de jóvenes de hasta 24 años donde se aprecia un promedio de 10,15 años, esto se configura como un buen antecedente para obtener un mayor stock de capital humano que es la base del desarrollo (Guaita, 2022).

En las últimas décadas, las organizaciones han atravesado diversas transformaciones para enfrentar las repercusiones del mundo globalizado, que demanda una mejora en los niveles de competitividad. En este contexto, se destaca la importancia del talento o capital humano como un factor esencial para aumentar la productividad de las organizaciones. El capital humano se basa en características innatas y adquiridas; en relación a estas últimas, es prioridad resaltar la relevancia de la capacitación y formación como una inversión que la organización realiza en sus empleados, con el fin de proporcionarles los conocimientos necesarios para participar en la toma de decisiones (Oliveira et al., 2019).

Complementando lo anteriormente expuesto, Salas (2021) manifiestan que la motivación es fundamental para aumentar la productividad y el rendimiento de los empleados, considerándose un elemento esencial para el éxito de la empresa. Un informe realizado por el Foro Económico Mundial para el año 2017 puntúa a el Ecuador con un índice de capital humano

de 59,87 lo cual significa que la inversión en capital humano se encuentra en la media con respecto a los demás países (Datosmacro, 2024).

Para Bohórquez et al. (2020), el talento humano se considera como el activo de una empresa ya que al tener al personal satisfecho se puede mejorar notoriamente la productividad, esto debido a que el empleado genera una cultura de compromiso y su comportamiento se mantiene alineado a los objetivos que las jefaturas asignen. En este contexto, es fundamental que el jefe adopte el papel del líder y fomente el respeto mutuo, una comunicación efectiva y el reconocimiento del esfuerzo, entre otros aspectos. Por el contrario, si hay personal insatisfecho, es probable que su rendimiento disminuya y la calidad del trabajo se vea comprometida.

La capacitación y el desarrollo de los empleados son, sin lugar a duda la inversión y elección más acertadas para las organizaciones que buscan destacarse en el mercado. Esto se debe a que, a medida que incrementan el conocimiento y las habilidades de cada colaborador, la empresa tiene más oportunidades de crear nuevas ideas, tomar decisiones correctas, ofrecer productos y servicios de calidad, establecer vínculos sólidos con clientes y partes interesadas, optimizar sus recursos tanto tangibles como intangibles, y, sobre todo, generar ventajas competitivas que les permitan enfrentar retos del negocio. Se establece, por tanto, una relación directa (Bernal et al., 2020).

Finalmente, es preciso importante considerar que la aportación de la gestión del capital humano se ve afectada por el ambiente de trabajo que se experimenta dentro de la organización; por ende, es importante que las jefaturas promuevan de forma sustancial un clima laboral que fortalezca e incentive emociones positivas en su personal, con el objetivo de contar con un capital humano motivado, comprometido, identificado y productivo (Pedraza, 2020).

2.2.2. *La juventud ecuatoriana*

La Ley de la Juventud define como “joven” a quienes tienen entre 18 y 29 años. Sin embargo, las diferentes categorías de edad utilizadas por el Instituto Nacional de Estadística y Censo (INEC), no presentan una variable única que establezca claramente esta delimitación. Esto distorsiona la información sobre las tasas de empleo juvenil y complica la comparación con otros instrumentos públicos de medición (De Domingo Soler et al., 2020). Por otra parte, para Ayala (2011) según la perspectiva sociológica más común en Ecuador, la población “joven” abarca a quienes tienen entre 15 y 29 años. Teóricamente, los 15 años marcan el comienzo de un proceso psicológico y social de transición hacia la adultez, que debería concluir idealmente a los 30 años, momento en el cual la persona habría logrado su autonomía, definido su identidad y se habría integrado de manera exitosa en la vida social como un adulto responsable.

En Ecuador, se acepta de manera general que hay un sector significativo de la población que está formado por jóvenes. La realidad de que aproximadamente el 30% de la población total tiene menos de 30 años lleva a declaraciones como, por ejemplo, que “el Ecuador es un país de jóvenes” (Ayala, 2011).

2.2.3. *La educación y los jóvenes*

La educación se ve como un pilar esencial para el progreso social de una comunidad; su acceso garantiza el avance social y económico de una nación. Esta disponibilidad de educación es crucial para el desarrollo sostenible, la disminución de la pobreza que se transmite de generación en generación, la mejora en la salud, la igualdad de género y la disminución de la desigualdad social en general. En el Ecuador, a partir de la Constituyente llevada a cabo en el año 2008 y la Ley de Educación Superior decretada en el año 2010 el camino a la formación en todos los niveles tuvo un beneficio superlativo.

El nuevo enfoque de las políticas ecuatorianas intentaba garantizar el acceso a la educación de los distintos grados formativos, y esto ha implementado sólidas políticas de inclusión desde la promulgación de la Constitución, dando lugar a una transformación educativa poco común en los países de América Latina (Burneo & Yunga, 2020).

De acuerdo con la Organización de las Naciones Unidas (ONU), la educación constituye la base para el desarrollo personal, favoreciendo un progreso sostenible y una mejora en la calidad de vida. Por esta razón, el tema de la calidad educativa es un punto central en los debates internacionales y está presente en distintos compromisos adquiridos por los gobiernos. Una educación de calidad debe ser inclusiva y equitativa, ya que esto ayuda a eliminar obstáculos y proporciona a las personas las herramientas necesarias para crear soluciones innovadoras a los desafíos actuales del mundo (Contreras, 2022).

Los jóvenes consideran la educación como una inversión valiosa y efectiva para su futuro, así como una oportunidad para el desarrollo personal, lo que les permite diferenciarse de los grupos sociales y familiares con los que se identifican. Por ello, se concentran en alcanzar un nivel educativo que les asegure el éxito en sus carreras, buenos ingresos y avance profesional (Firsova et al., 2020).

Para Burneo (2020), en el Ecuador en los últimos años existieron muchos avances con respecto a la apertura de los jóvenes a los distintos grados de educación sin embargo también hubo retrocesos. Por un lado, se han intensificado los esfuerzos para promover la inclusión de jóvenes, y se puede afirmar que se han logrado parcialmente los objetivos de garantizar un acceso equitativo a las universidades en Ecuador, ya que se observa un aumento en la cantidad de jóvenes inscritos en diversas instituciones educativas.

Hay un incremento en la solicitud de cupos sin una verificación eficaz de la verdadera capacidad de las universidades para educar a más personas. Se debe considerar además, por ejemplo, que también es importante considerar el avance de la educación en sectores considerados desfavorecidos. En estas áreas, es necesario implementar programas de apoyo educativo dirigidos a jóvenes con antecedentes criminales, así como programas de nivelación para aquellos con bajo rendimiento académico, con el objetivo de prevenir la deserción escolar a gran escala.

En la actualidad, el mundo avanza hacia una economía fundamentada en el conocimiento, y la educación debe ser un servicio accesible para todos, así como un derecho humano universal. La formación de los jóvenes ya sea en el nivel universitario o en cualquiera de sus etapas, garantiza una participación equitativa en la política, la sociedad y la economía. Por esta razón, uno de los objetivos prioritarios del Estado debe ser generar auténticas oportunidades de accesibilidad e inclusión que se materialicen en la realidad del país, más allá de lo que estipulan las leyes.

2.2.4. Empleo juvenil

La juventud es una etapa de cambio en la vida de cada persona que requiere ser valorada, atendida y fomentada (Olaguibe, 2021). No obstante, en la actualidad, la perspectiva educativa, sociológica y política sostiene que los jóvenes son un recurso humano valioso para la sociedad actual, lo que indica un alejamiento gradual de la visión idealista que los veía como un potencial para el futuro, y un enfoque en sus realidades y dinámicas actuales (Avendaño et al., 2020). Un informe de la Organización Internacional del Trabajo (OIT) indica que, en la actualidad, los jóvenes tienen una probabilidad de estar desempleados tres veces mayor que la de los adultos (Organización Internacional del Trabajo, 2024).

Dentro de este informe de la OIT en contraste con un estudio anterior del año 2004 los resultados indican que las tendencias mundiales del empleo juvenil, los indicadores y proyecciones se siguen centrando en la inestabilidad del mercado laboral para hombres y mujeres jóvenes, por ejemplo, algunas cifras determinaron que la participación juvenil en la fuerza de trabajo disminuyó considerablemente en los últimos 20 años y muchos jóvenes desempleados llevaban un año o más sin trabajar, y por ende existe una clara tendencia de conseguir trabajos menos estables (Bermúdez, 2023).

El Ministerio de Trabajo (2021), en el estudio de empleabilidad del año 2021 señala que el empleo juvenil se ha convertido en una de las prioridades para las agendas de desarrollo a nivel nacional e internacional priorizando no solo las intervenciones parciales sino la creación de una cultura de impulso de políticas a largo plazo que permita tener a los jóvenes un empleo digno y adecuado con la participación del sector público y privado. Así también Ramírez (2022), menciona que el empleo es una forma de incorporación social de los jóvenes dentro del mercado laboral que afecta de manera positiva en la capacidad de autonomía de los jóvenes y de poder de adquisición de estos más sin embargo los jóvenes se hallan en una época donde el contexto es cambiante.

El empleo juvenil constituye uno de los desafíos globales más significativos, ya que sus repercusiones trascienden el ámbito laboral, lo que genera una gran preocupación entre gobiernos y organizaciones internacionales. Las limitadas expectativas laborales y, en particular, la escasez de oportunidades que enfrentan tanto mujeres como hombres al finalizar su educación tienen efectos profundos en la sociedad. Esto se agrava si, por diversas razones, deciden abandonar la educación de manera anticipada.

Es crucial establecer un vínculo sólido entre lo personal y lo profesional en esta etapa, ya que influye tanto en el desarrollo individual como social. Según la OIT, hay 70 millones de jóvenes en el mundo en busca de empleo; más de 160 millones de jóvenes están empleados, pero viven en condiciones de pobreza; 2 de cada 5 jóvenes son desempleados o trabajan en situación de pobreza; y en términos comparativos, la tasa de desempleo juvenil es tres veces superior a la de los adultos (Olaguibe, 2021).

Para López (2021), la situación del empleo juvenil se desarrolla en un contexto internacional sumamente complejo, en el que las cifras mundiales de desempleo van a continuar creciendo en los próximos años. Se debe tomar en cuenta que existe un importante problema relativo a la desvinculación de los jóvenes frente a su formación académica y de su orientación hacia al mercado de trabajo. Por ejemplo, se calcula que, 1300 millones de jóvenes a nivel mundial, no están ni en el mercado de trabajo, ni están estudiando. Igualmente, conviene destacar que casi 190 millones de mujeres en edad joven se encuentran afectadas por las dificultades de acceso al mercado laboral; en especial, el problema que tienen las mujeres jóvenes de acceder al mercado de trabajo radica en una posición doblemente desfavorable ya que se valora su edad y su género.

Existe una creciente inquietud por las dificultades que enfrentan los jóvenes graduados al ingresar al mundo laboral. Esta preocupación se fundamenta en varios factores: en primer lugar, la formación superior a menudo no se ajusta a las vacantes disponibles. En segundo lugar, se suma la precariedad salarial. Y, finalmente, la falta de creación de empleos suficientes para estos trabajadores jóvenes formados. Como consecuencia, se desaprovecha el potencial del colectivo juvenil en la sociedad, lo que a su vez merma la propia capacidad de este grupo.

En América Latina y el Caribe, se observa una paradoja: las empresas exigen experiencia laboral junto con la formación académica, pero los jóvenes carecen de oportunidades para adquirirla. En países como Chile, donde las prácticas profesionales son la principal vía de entrada al mercado laboral, la escasez de estas oportunidades aumenta el riesgo de desempleo juvenil. Por otro lado, en el contexto ecuatoriano, las empresas requieren que los jóvenes cuenten con una maestría y una experiencia laboral de dos a tres años para poder ingresar al mercado laboral (Martínez, 2023).

El gobierno ecuatoriano ha mostrado preocupación por el desajuste laboral que afecta a los jóvenes, incorporando esta problemática en sus políticas públicas. Sin embargo, la precarización laboral persiste, a pesar de los esfuerzos institucionales. Aunque el Estado y las empresas privadas han realizado diversas iniciativas, estas no han sido suficientes para generar condiciones óptimas de empleo para los jóvenes, quienes continúan enfrentando obstáculos significativos para su inserción laboral (Organización Internacional del Trabajo, 2024).

Puntualmente en Ecuador, para Pillajo y Sierra (2023), a pesar de la considerable inversión estatal, los programas 'Empleo Joven' y 'Mi Primer Empleo' no han logrado reducir el desempleo y el subempleo juvenil en el país. Se ha evidenciado que estos programas, dirigidos a jóvenes de 15 a 24 años, carecen de la estructura necesaria para fomentar la creación de empleo. Para mejorar la situación, se sugiere simplificar las regulaciones para la creación de empresas y reducir los impuestos a las nuevas compañías, lo que incentivaría la generación de empleo juvenil. Además, se destaca la importancia de que las universidades adapten sus planes de estudio a las demandas del mercado laboral global.

En cuanto a la calidad del empleo, se muestra que, en diciembre de 2020, el 66,7% de jóvenes tienen un empleo no pleno y el 9,4% se encuentra desempleado. Esta misma

situación se observa en marzo 2021, el 64,9% de los jóvenes tienen un empleo no pleno y el 9,7% se encuentra desempleado. Siguiendo la tendencia, para mayo de 2021, el 66,2% de los jóvenes tiene empleo no pleno y el 11,7% de los jóvenes se encuentra desempleado. (Ministerio de Trabajo, 2021, p. 6)

2.2.5. Mercado laboral en el Ecuador

El mercado laboral es un sistema complejo, influenciado por diversos factores. Estos incluyen la participación de la población en la fuerza laboral, tanto hombres como mujeres, la edad promedio de los trabajadores, sus niveles educativos, las demandas y cambios en los sectores productivos y sociales, los requisitos de formación, la calidad del empleo (considerando el subempleo y la informalidad), y la distribución de los ingresos laborales (Bernal et al., 2022). Según Rodríguez (2019), A lo largo de los años, el mercado laboral en Ecuador ha presentado una tendencia hacia la generación de empleos de baja calidad. Los informes muestran altos niveles de subempleo, discriminación laboral que afecta a grupos vulnerables como jóvenes, mujeres y minorías étnicas, y una limitada efectividad de las políticas implementadas para mejorar la situación laboral.

El mercado laboral, como se ha mencionado, opera bajo los principios de oferta y demanda de empleo. Sin embargo, este mercado está sujeto a regulaciones legales y a la supervisión de instituciones y organismos que protegen tanto a los trabajadores como a las empresas. Estas regulaciones establecen requisitos salariales y otras condiciones laborales, limitando la capacidad de las empresas para operar libremente en la interacción entre la oferta y la demanda de trabajadores.

En la economía ecuatoriana, un conjunto de sectores clave concentra la mayor parte de la generación de empleo, representando en total un 83.2%. Destacan particularmente la agricultura,

el comercio, la industria manufacturera, los servicios sociales, de salud y educación, la hostelería, la construcción y el transporte (Bernal et al., 2022).

Los jóvenes perciben el mercado laboral como centrado en los adultos, inflexible a sus necesidades y aspiraciones, y deficiente en ofrecer oportunidades que valoren su preparación, a pesar de su falta de experiencia. Esta percepción se justifica por la alta tasa de empleo precario (72%) entre los jóvenes y su significativa proporción (40.4%) en el total de desempleados (Olmedo, 2018).

2.2.6. Transición de los jóvenes de la escuela al mercado laboral

La transición de la educación al empleo es un momento crítico en la vida de los jóvenes. Aquellos que abandonan sus estudios a menudo carecen de las habilidades necesarias para el empleo formal, lo que los lleva a aceptar trabajos inestables y mal remunerados en el sector informal, o a enfrentarse al desempleo, afectando negativamente sus futuros ingresos. Sin embargo, existe una teoría económica que sugiere que trabajar mientras se estudia permite adquirir habilidades prácticas que la educación formal no ofrece. Además, la experiencia laboral temprana demuestra motivación y productividad, lo que facilita el acceso al empleo y permite a los jóvenes financiar su educación (Ubfal, 2022).

Comprender la transición entre la formación académica y el mundo laboral es crucial para analizar el desarrollo de la vida de los individuos. Este análisis posibilita el desarrollo de hipótesis sobre la creación de oportunidades en la sociedad (Rodríguez E. , 2019). Este proceso de transición se puede entender como el inicio de la vida adulta, ya que implica que los jóvenes alcancen la independencia económica, generalmente a través del empleo, la formación y la creación de un hogar propio. Este desarrollo ocurre mediante la participación en diversas estructuras y entornos sociales, lo que también fomenta el crecimiento personal y comunitario.

Este período de transición puede interpretarse como el inicio de la vida adulta, ya que implica que los jóvenes logren la independencia financiera, generalmente a través del trabajo, la formación y el establecimiento de un hogar independiente. Este proceso se desarrolla mediante la participación en diversas estructuras y entornos sociales, lo que también impulsa el desarrollo individual y colectivo (Albert et al. 1998).

Es esencial realizar un análisis detallado de los antecedentes relacionados con la integración socio laboral de los jóvenes. Si bien los programas, las políticas gubernamentales y los elementos básicos para la vida adulta son útiles, la inserción laboral efectiva, la formación integral y el acceso a procesos formativos vinculados al empleo, la vivienda digna y el apoyo comunitario son fundamentales y aún están pendientes. Aún queda mucho por investigar y analizar, y es crucial recordar que estos jóvenes son el futuro de la sociedad global. Invertir en su inclusión es apostar por una sociedad tolerante, diversa e inclusiva, donde todos nos beneficiamos (Jiménez, 2022).

Para Vargas (2019), en América Latina la educación enfrenta dificultades para asegurar la inserción laboral de los jóvenes, con millones sin empleo. Ecuador comparte esta problemática, donde un título de bachiller ya no garantiza un trabajo digno. Al salir del sistema educativo, muchos jóvenes se encuentran en una situación de 'no formación', similar al analfabetismo en niveles educativos inferiores. Incluso los jóvenes con educación profesional enfrentan dificultades para encontrar empleos de calidad, lo que los lleva a aceptar trabajos precarios por temor al desempleo. Esto refleja que el sistema educativo no cumple con las expectativas de los jóvenes: aunque la educación no asegura un futuro mejor, su ausencia empeora las condiciones laborales.

La transición de los jóvenes de la academia al trabajo se relaciona directamente con el hecho de el avance del nivel de instrucción. La población económicamente activa joven en Ecuador es mayormente formada por aquellos jóvenes con título, pues los jóvenes que tienen título y no trabajan tienen una menor participación dentro de este segmento; el 2020, los jóvenes con título y sin trabajo eran el 9,77% (Mideros et al., 2022).

2.2.7. Jóvenes que ni estudian ni trabajan “NINI”

El término "nini" se originó en el Reino Unido a fines de los años 90 como NEET (ni en empleo, ni en educación, ni en formación), sustituyendo a la denominación previa de "status zero". Este grupo se compone de tres categorías principales: desempleados sin estudios, inactivos económicos (como amas de casa o personas con discapacidad) y aquellos clasificados como "otros". Este último grupo representa el núcleo duro de los nini, caracterizado por exclusión social y dificultades de integración, y es el foco principal de análisis que a menudo engloban a todo el colectivo (Assusa, 2019).

Una realidad preocupante en la sociedad actual es la exclusión social que enfrentan los jóvenes. En América Latina, alrededor de 30 millones de jóvenes se encuentran marginados de los sistemas económicos y sociales. Esta situación se debe a que no participan ni en la educación ni en el empleo, lo que los sume en un estado de desesperanza y los hace sentir culpables de su propia situación. Las estadísticas revelan que uno de cada cinco jóvenes latinoamericanos no tiene acceso a empleo, educación o capacitación. Esta problemática se agrava en el caso de las mujeres jóvenes, donde el 30% se encuentra en esta situación, en comparación con el 11% de los hombres. Muchas de estas mujeres jóvenes realizan labores de cuidado no remuneradas, las cuales no son reconocidas como trabajo (Vargas & Bracchi, 2019).

La magnitud del fenómeno "nini" es el resultado de una combinación de factores tanto estructurales como microeconómicos, los cuales requieren un análisis profundo para fundamentar las políticas públicas dirigidas a diagnosticar y abordar este problema. No existe una edad específica definida para este grupo de jóvenes; algunos estudios los sitúan entre los 15 y 24 años, mientras que otros amplían el rango hasta los 29. Esta variación depende del contexto social, político e histórico, así como de los objetivos del análisis. Los niveles de ingreso y la situación económica del hogar son factores determinantes en la probabilidad de que los jóvenes se conviertan en "nini". En el caso de América Latina, lo común en el contexto del estudio es que los jóvenes de hogares pobres o de menor ingreso enfrentan mayores barreras en su inserción laboral y menores oportunidades para alargar su permanencia en el sistema educativo (Sánchez R. , 2021).

Para Sánchez (2021), el fenómeno de los "nini" trasciende la esfera individual, afectando no solo las perspectivas de vida de los jóvenes, sino también su desarrollo personal, la realización de sus capacidades y el ejercicio de sus libertades. Además, representa un obstáculo para la productividad, el desarrollo y el crecimiento económico a largo plazo. Los "nini" evidencian las limitaciones de una sociedad para proporcionar oportunidades equitativas y la ineficacia en la utilización de los recursos disponibles.

Los llamados "nini" están en toda América Latina, y en la realidad de la sociedad ecuatoriana se puede identificar como grupos que son vulnerables y se están enfrentando día a día no solo a la exclusión laboral sino también social, pues al tener mayor tiempo de ocio hace que estos jóvenes caigan en situaciones de vandalismo, drogas, armas, etc; sumando un problema más a la sociedad; para ello existen distintos tipos de medidas que se deben tomar a favor de la juventud y que pueden ayudar al éxito de combatir el desempleo y a la población actualmente

inactiva. Para garantizar dichas medidas y normas a nuestra juventud se necesitan programas y leyes constitucionales coherentes y realizables, es esencial la inversión social en las cualificaciones y capacidades de las personas, aumentando así sus posibilidades a la inserción de un mercado laboral digno (López & Mondragón, 2019). Para el contexto local “del total de la PEI, obteniendo el porcentaje de jóvenes que ni trabajan ni estudian, se concluye que de forma general hay un pequeño porcentaje de nini, siendo el mayor en 2021, con el 7,42%” (Mideros et al., 2022, p. 7).

Las medidas destinadas a los “*nini*” deben ser flexibles, intentando alcanzar resultados sostenibles y positivos a largo plazo en el mercado laboral, tanto para los jóvenes como para las empresas, pues es más rentable tener un empleado ya capacitado para desempeñar una actividad dentro de la empresa que tener que capacitar a uno. Este tipo de prácticas incentiva a la enseñanza de calidad en el trabajo, los periodos de prácticas y becas para mejorar la transición sostenible de la enseñanza y la formación al trabajo, fomentando las cualificaciones pertinentes para el mercado laboral.

2.2.8. Jóvenes que estudian y trabajan “SISI”

Existe un grupo de jóvenes, conocidos como la generación 'sisi', que combina sus estudios con el trabajo. Estos jóvenes, en edad escolar o académica, se incorporan al mercado laboral debido a diversas circunstancias personales o familiares (Ramirez, 2022). Para Cruz (2017), desde la perspectiva de las estrategias de supervivencia familiar, la combinación de estudios y trabajo por parte de los jóvenes se presenta como una táctica para asegurar la estabilidad económica de los hogares en situaciones de limitados recursos.

Cruz et al. (2017), determinan que para los jóvenes en esta situación se genera claramente una duplicidad de roles, donde puede suceder diversas situaciones, una de las más observadas es

donde los jóvenes sin terminar sus estudios se insertan en el mercado laboral teniendo esta la particularidad de ser reversible o no, es decir que se pueda revertir su situación. Al igual que los jóvenes *nini*, este grupo de personas puede tener un efecto negativo a mediano plazo, este principalmente observado en la posición económica de los individuos a futuro.

La práctica de combinar estudios y trabajo conlleva riesgos, especialmente en dos escenarios. Primero, puede aumentar el abandono escolar, ya que los estudiantes que trabajan en zonas urbanas tienen casi el doble de probabilidades de dejar la escuela secundaria en comparación con aquellos que solo estudian. También se debe mencionar que ciertos trabajos, como los de alta intensidad y en las industrias manufactureras y de la construcción afectan en mayor medida el abandono escolar y, por lo tanto, impiden la combinación de los estudios con el trabajo.

En segundo lugar, iniciar la vida laboral antes de completar la educación básica puede tener consecuencias negativas en la estabilidad económica de los jóvenes en el futuro. Entonces, ¿cuál es la razón por la que los jóvenes deciden ingresar al mundo laboral? La respuesta radica en que el empleo juvenil se ha convertido en una estrategia de supervivencia para las familias, impulsada por la disminución de los ingresos reales y los cambios sociales de los últimos años (Cruz et al., 2017).

El panorama laboral juvenil en América Latina presenta características particulares, especialmente en lo que respecta al aumento de la participación femenina. Este fenómeno se relaciona con cambios en los patrones de comportamiento entre generaciones durante las décadas de 1980 y 1990. En ese período, los jóvenes de 18 a 24 años experimentaban tasas de desempleo significativamente más altas, de tres a cuatro veces superiores a las de los adultos de mediana edad. Además, enfrentaban mayores niveles de precariedad laboral y menor acceso a la

seguridad social. En ese sentido, Quispe et al. (2019) mencionan los jóvenes trabajadores con bajos ingresos se encuentran en dos situaciones laborales precarias: algunos laboran en establecimientos fijos, expuestos al riesgo de despido, y otros trabajan por cuenta propia en el sector informal. La escasez de empresas que ofrezcan prestaciones por desempleo dificulta la búsqueda de empleo, lo que impacta negativamente en el desarrollo de los países, incrementa la población de trabajadores pobres y ralentiza el progreso en la reducción de la pobreza.

Se puede afirmar que las características de los jóvenes 'sisi' en Ecuador están estrechamente ligadas a factores socioeconómicos del entorno, como la edad, el género, el nivel educativo y la disposición para trabajar. Sin embargo, el género sigue siendo un factor discriminatorio en la contratación por parte de las empresas industriales, ya que las mujeres enfrentan mayores obstáculos para acceder al empleo formal, lo que las empuja hacia la informalidad. Por otro lado, los incentivos económicos que ofrece el Estado a las empresas por la contratación de jóvenes 'sisi' de entre 18 y 26 años representan un beneficio para el sector empresarial (Quispe et al., 2019).

2.2.9. El clúster análisis

El clúster análisis comúnmente hace referencia a el agrupamiento de datos en base a una característica o características en común; de ahí que detrás de ello existe una metodología a seguir mediante la cual se puede obtener un resultado óptimo.

Clasificación o identificación es el proceso o acto de asignar un nuevo objeto u observación en su lugar correspondiente dentro de un conjunto de categorías establecido. Los atributos esenciales de cada categoría son conocidos, aunque haya algunas incertidumbres a la hora de asignar alguna observación dada. (Universidad de Granada, 2022, p. 58)

El análisis clúster es un procedimiento estadístico multivariante en el cual se utiliza un conjunto de datos que se organizan en grupos homogéneos y dan lugar a lo que se denomina clúster. La relevancia del análisis de clúster radica en su validez como método científico, el cual persigue cuatro objetivos principales: establecer clasificaciones, desarrollar esquemas conceptuales para la agrupación de entidades, formular hipótesis a través de la exploración de datos y verificar dichas hipótesis (Universidad de Granada, 2022).

Para Cruz y Texeira (2014), mencionan tres elementos relevantes en las investigaciones sobre clusters:

1. La cercanía geográfica entre componentes
2. Las redes sociales, que circunscriben red de conexiones dentro de los cluster que dan lugar a las proximidades
3. La cultura.

Noissan (2021), menciona que es fundamental comprender que un clúster no es una entidad estática, sino que evoluciona a lo largo del tiempo. Diversos factores pueden provocar cambios en su desarrollo, desde el avance y retroceso hasta el estancamiento. Estas transformaciones pueden ocurrir de forma natural, como resultado de las características inherentes del clúster, o ser inducidas por acciones deliberadas y planificadas.

El análisis clúster es ampliamente utilizado en todos los contextos, uno de los más utilizados es en área social, así como en el ámbito comercial, un estudio realizado por Gómez (2002) , muestra que el área del marketing es uno de los más aprovechados de esta herramienta pues se conoce que se utiliza desde los años 90. Otro estudio donde se utiliza clúster análisis es el realizado por Rodríguez (2022) quien analiza los clúster de investigadores en función de algunas variables.

Cada persona investigadora tiene sus propios intereses y necesidades, en función de su desarrollo cognitivo, del entorno en el que se desenvuelve y de su experiencia vital, que la hacen única, de los perfiles de investigadores se pueden determinar el nivel de interacción entre estos, en función de los campos de experiencia recogidos en su perfil, el nivel de compatibilidad o la distancia entre ellos, los clústeres de investigadores que responden a los parámetros definidos en su perfil. (Rodríguez, 2022, p. 6)

2.3. Marco conceptual

2.3.1. *Análisis clúster*

“El análisis clúster es un conjunto de técnicas multivariantes utilizadas para clasificar a un conjunto de individuos en grupos homogéneos” (Lejarza, 2018, p. 1). En el contexto de los estudios sociales los clústeres análisis son importantes para organizaciones públicas y privadas, conglomerados sociales, estudiantiles y demás, por ejemplo, en un estudio realizado por Aguado et al. (2019), se indica la importancia del uso de los instrumentos de agrupamiento estadístico para entidades municipales.

2.3.2. *Jóvenes nini*

“La generación nini es un grupo poblacional conformado por jóvenes que ni estudian ni trabajan, debido a las limitaciones económicas y sociales que tienen para culminar sus estudios o conseguir un empleo” (Balarezo, 2019, p. 2266). Los nini son la antítesis de quienes estudian y trabajan, este clúster posee características diferentes a el objeto de estudio sin embargo sirve al momento de analizar la unidad de estudio.

2.3.3. *Jóvenes Sisi*

Esta palabra se ha formado por analogía con nini y, de manera opuesta a esta última, designa a quien sí estudia y sí trabaja, en particular si se trata de jóvenes. Se

documenta su uso como sustantivo («un sisi») y como adjetivo («Repuntan los jóvenes sisis»). (RAE, 2023)

Esta población constituye la unidad de análisis de la investigación, estos se analizan en función de sus variables sociodemográficas, socioeconómicas que constituyen factores dirimentes a la hora de crear clúster análisis.

2.3.4. Empleo

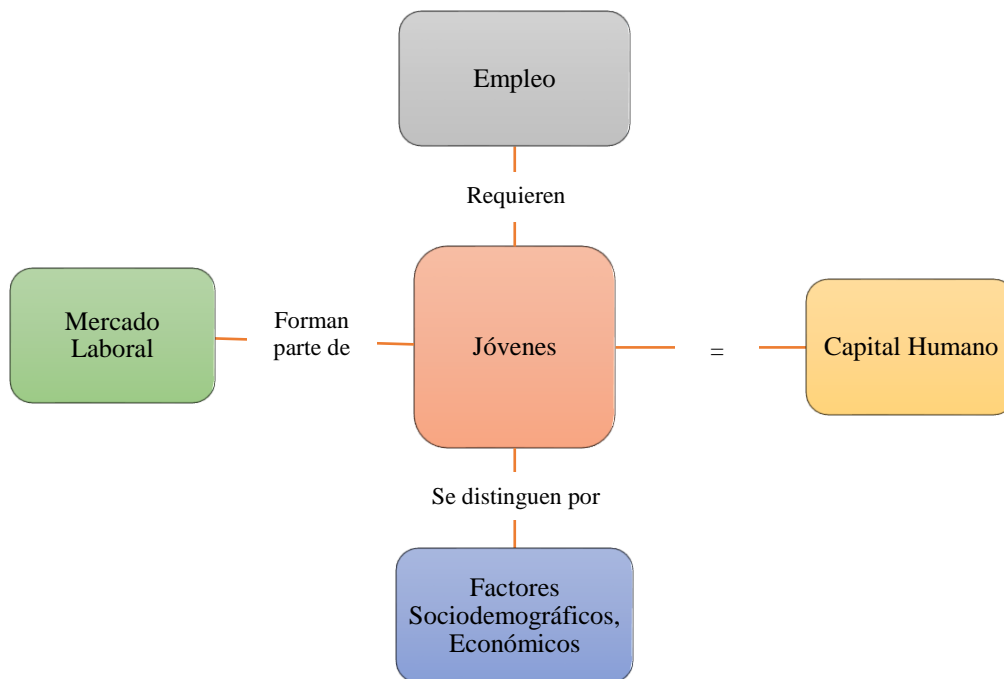
La Organización Internacional del Trabajo (OIT) entiende el “pleno empleo” como el escenario donde a) hay trabajo para todas las personas que quieren trabajar y están en busca de él; b) tal empleo es tan productivo como es posible; y c) los individuos tienen la libertad de elegir el empleo. (Enriquez & Galindo, 2015, p. 2)

2.3.5. Educación

La educación es un intento humano racional, intencional de concebirse y perfeccionarse en el ser natural total. Este intento implica apoyarse en el poder de la razón, empleando recursos humanos para continuar el camino del hombre natural hacia el ser cultural. Cada ser humano/ hombre/mujer termina siendo a través de la educación una cultura individual en sí mismo. (León, 2007, p. 598)

2.3.6. Índice de silhouette

“El coeficiente de Silueta es una herramienta que sirve para valorar la eficacia de la calidad de agrupamiento formados por algoritmos de clustering. Su principal función es identificar cuál es el número óptimo de agrupamientos” (Bancero, 2015, p. 1)

Figura 1*Conceptos principales*

Capítulo III: Marco Metodológico

3.1. Instrumentos de investigación

Los instrumentos de investigación descritos a continuación aportan para la respuesta de los objetivos del estudio, se realiza una secuencia de elección de métodos de manera escalada que permiten analizar, describir, determinar, seleccionar y evaluar las características de los jóvenes que si estudian y trabajan, el análisis descriptivo permite revisar los primeras asociaciones de las categorías de variables lo que muestra a breves rasgos como pueden estar caracterizados los clusters, sin embargo esto se asegura a través de las técnicas de agrupamiento como el Análisis de componentes principales y el método de k medias, mismo que por último y es validado por el índice de silhouette dando así la eficacia y respaldo suficiente de utilización óptima de los métodos estadísticos.

3.1.1. Análisis descriptivo

La estadística descriptiva se enfoca en organizar, resumir y analizar datos mediante diversas técnicas, limitándose a describir el conjunto de datos específico. Al aplicarla, se obtienen medidas básicas, y se evitan datos no esenciales para enfocarse en elementos clave que permitan una caracterización general de los datos procesados (Epidat, 2014).

Además de resumir datos, la estadística descriptiva ofrece guías para presentar la información de manera visual mediante cuadros, tablas, gráficos o figuras. Es crucial recordar que antes de realizar cualquier análisis descriptivo, es necesario revisar los objetivos de la investigación y determinar las escalas de medición de las variables registradas en el estudio. El objetivo de las tablas o cuadros es proporcionar información puntual de los resultados. Las gráficas muestran las tendencias y pueden ser histogramas, representaciones en “pastel”, “cajas con bigotes”, gráficos de líneas o de puntos de dispersión. Una vez que se han recordado los objetivos de la investigación, es necesario identificar las escalas de medición de las variables

registradas. Para fines prácticos, estas escalas se pueden clasificar en dos categorías principales: cuantitativas (métricas) y cualitativas (categóricas).

Las Cuantitativas se definen por la existencia de una unidad de medición, que puede ser contable (unidades enteras), medible o ponderada por algún atributo físico con algún instrumento y las cualitativas se caracterizan por clasificar a los individuos o fenómenos solo con relación a sus atributos. (Rendón et al., 2016, p. 398)

El Análisis Descriptivo actúa como una herramienta reveladora, exponiendo el razonamiento de los investigadores. Más allá de una simple reflexión, este proceso de pensamiento impulsa la revisión de las bases teóricas en la práctica científica (Sánchez et al., 2010).

De las tablas y gráficos Cañadas (2010), menciona que:

La ciencia utiliza representaciones semióticas externas como tablas, gráficas, diagramas, ecuaciones, ilustraciones o enunciados para construir y comunicar los conceptos abstractos. Por tanto, el aprendizaje de los conceptos científicos está ligado al de estas representaciones y al de sus procesos de construcción y transformación. Estas representaciones se usan también en las ciencias como puente entre los datos experimentales y las formalizaciones científicas y ayudan a determinar las relaciones entre las variables que intervienen en los fenómenos, para poder modelizarlos. (p. 5)

Las tablas y gráficos ayudan a una mejor comprensión de los datos transformados en información, a partir de ellos se puede comprender las características del grupo poblacional, en este caso la información recabada de los jóvenes “sisi”, los cruces de variables en las tablas de contingencia, especialmente las bivariada o multivariada indican las primeras características de segmentación de los grupos los cuales a partir de las categorías de variables indican como, de

manera resumida existe una determinada frecuencia que permite evaluar o dar rasgos generalizados de como la población se va agrupando de manera “orgánica”. Es decir, por ejemplo, en una tabla se puede visualizar la cantidad de jóvenes por sexo y edad o sexo, nivel de instrucción, o por disposición de ubicación geográfica o constructo familiar.

Respecto a las tablas de contingencia Bouza (2021), menciona que:

Analizar la distribución de una variable con relación a otras es corriente en la búsqueda de patrones de asociación. Este es el objetivo de muchas investigaciones sobre posibles causales y casales. En el estudio de como influencia una variable) sobre otra conforma los problemas del análisis bivariado. (p. 5)

Las tablas y gráficos que se usaran serán en su mayoría escalas nominales y de intervalo por las características de las variables a utilizar que son de carácter cualitativo, se procurará transformar las variables cuantitativas en variables de intervalo, en función de esa transformación y de las características de las variables, las tablas y gráficos van a aportar para dimensionar las características más importantes de la comunidad de los jóvenes que estudian y trabajan dando el primer indicio de como se pueden formar los clusters en función de la frecuencia de las características más importantes detalladas a través de la clasificación de las variables. La lectura del gráfico empieza con una caracterización externa del tema a través del título del gráfico o tabla, luego se observa las dimensiones relevantes para luego percibir la relación entre los niveles individuales de cada extensión.

3.1.2. K means o cluster jerarquico

“El análisis de conglomerados es una técnica multivariante que permite agrupar los casos o variables de un archivo de datos en función del parecido o similitud que existente entre ellos” (Universidad Carlos III de Madrid, 2021, p. 461).

El método K – means o K- medias es un procedimiento iterativo mediante el cual cada observación se asigna al clúster más cercano. Inicialmente se establecen de forma aleatoria las ubicaciones centrales de los grupos o centroides, luego se calculan las distancias de los distintos individuos a estos centroides y se agrupan cada uno en el clúster más cercano. Una vez las observaciones han sido clasificadas en el grupo cuyo centroide es más cercano, se recalculan los centroides y se repite el proceso de reagrupación hasta que los centroides se estabilizan. (Universitat de València, 2024, p. 3)

El método jerárquico es idóneo para determinar el número óptimo de conglomerados existentes en los datos y el contenido de los mismos. El método K-means ofrece la ventaja de procesar grandes volúmenes de datos, pero se limita a un único método de agrupación y requiere la definición previa del número de grupos deseados. Ambos métodos de análisis comparten la característica de ser aglomerativos, es decir, parten del análisis individual de cada caso para luego agruparlos en conglomerados homogéneos. (Universidad Carlos III de Madrid, 2021). El análisis de conglomerados K-medias resulta particularmente útil cuando se manejan conjuntos de datos extensos. Esta técnica ofrece flexibilidad, permitiendo su aplicación tanto para explorar datos y determinar la ubicación de los centroides de manera iterativa, como para clasificar datos utilizando centroides previamente definidos por el usuario.

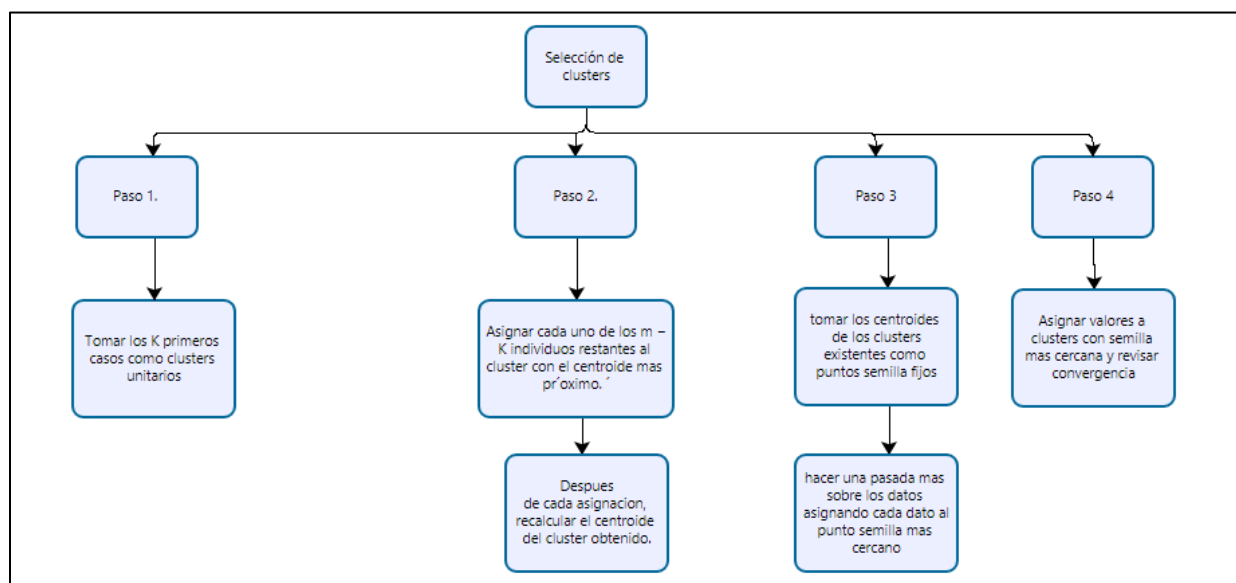
Aunque el método K-medias es eficaz para agrupar grandes conjuntos de datos, presenta limitaciones con ciertos tipos de datos. La media, utilizada como representante natural de un conjunto de datos, se justifica por minimizar la dispersión cuadrática media (Perucha, 2022). El análisis de conglomerados K-medias emplea consistentemente la distancia euclidiana para medir la separación entre los puntos de datos. Esta métrica representa la longitud de la línea recta que conecta dos puntos. La distancia euclidiana se calcula de una manera específica:

$$d_{ii} = \sqrt{\sum_j (X_{ij} - X_{i'j})^2}$$

De acuerdo a la metodología de las K medias se menciona la manera de elegir los clusters de la siguiente manera:

Figura 2

Elección de clusters



3.1.3. *Análisis de componentes principales*

El análisis de componentes principales es una técnica estadística multivariante de simplificación, que permite transformar un conjunto de variables originales correlacionadas entre sí, en un conjunto sintético de variables no correlacionados denominados factores o componentes principales. En esta transformación no se establecen jerarquías entre variables y se elimina la información repetida. “Las nuevas variables son combinaciones linealmente independientes de las variables originales, ordenadas de acuerdo a la representación de dispersión respecto a la nube total de información recogida en las muestras” (Zapotitla, 2024, p. 45).

“Esta técnica fue inicialmente desarrollada por Pearson a finales del siglo XIX y posteriormente fueron estudiadas por Hotelling en los años 30 del siglo XX. Sin embargo, hasta la aparición de los ordenadores no se empezaron a popularizar” (Universidad Carlos III de Madrid, 2023, p. 5).

El análisis de componentes principales es una técnica estadística descriptiva que analiza datos de múltiples variables medidas en un conjunto de individuos. Por esta razón, se clasifica como una herramienta multivariante.

La técnica del análisis de componentes principales construye una transformación lineal que escoge un nuevo sistema de coordenadas para el conjunto original de datos, en el cual la varianza de mayor tamaño del conjunto de datos es capturada en el primer eje (llamado el Primer Componente Principal), la segunda varianza más grande es el segundo eje, y así sucesivamente. (Ruiz, 2022, p. 47)

Una de las fortalezas de esta herramienta estadística radica en su capacidad para simplificar conjuntos de datos complejos, al concentrarse en las características que más contribuyen a su variabilidad (Ruiz, 2022).

Es importante reducir el número de variables para desechar información redundante y optimizar el proceso. El Análisis de Componentes Principales (ACP) propone la transformación a un nuevo conjunto sintético de variables (los componentes principales), que no están correlacionados y se encuentran ordenados de tal forma que los primeros conservan la mayor parte de la variación presente en todas las variables originales (Zapotitla, 2024, p. 44)

Un problema importante del análisis de componentes principales es que sus resultados dependen de la escala de medida de las variables originales. Así, si se cambia de escala

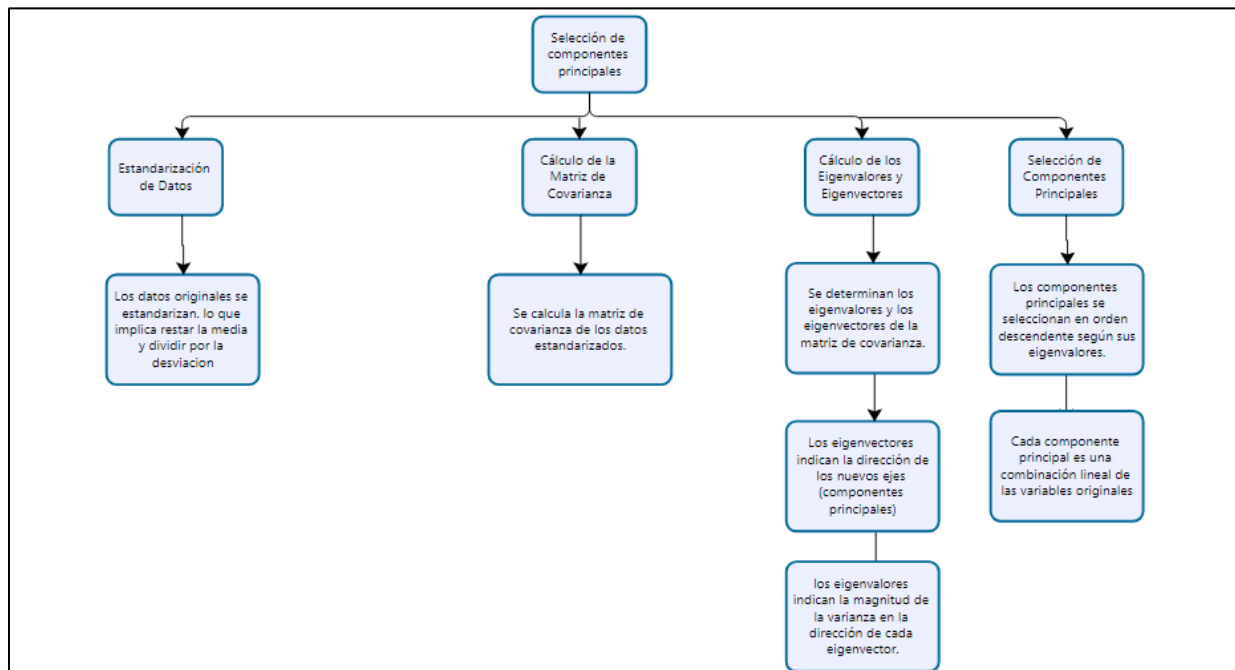
una variable (dejando las demás fijas) las componentes principales se modifican, cambiando incluso la posible interpretación de las mismas. (Sánchez C. , 2008, p. 58)

Existen dos formas básicas de aplicar el ACP:

- 1. Método basado en la matriz de correlación:** cuando los datos presentan heterogeneidad, ya sea por diferencias en las unidades de medida de las variables o por variaciones en la magnitud de las mismas, es necesario aplicar una estandarización previa.
- 2. Método basado en la matriz de covarianzas:** se emplea cuando los datos son homogéneos en sus dimensiones y presentan valores promedio similares. Se basa en la determinación de la matriz de varianza-covarianza sin realizar una estandarización previa de la información (Ruiz, 2022).

Cuando se busca entender las relaciones entre múltiples variables correlacionadas, se puede recurrir a la transformación de estas variables originales en un nuevo conjunto de variables independientes, denominadas componentes principales. Estas nuevas variables, que son combinaciones lineales de las anteriores, se ordenan según su capacidad para explicar la variabilidad total de los datos. Uno de los objetivos clave de este análisis es identificar la información esencial que cada componente principal resume de la muestra (Fuente, 2021).

Partiendo de este hecho el análisis de componentes principales nos servirá para observar las características de diferentes clusters formados a partir de la metodología del acp donde por ejemplo habrá presencia de grupos de jóvenes sisi que tenga una determinada etnia y rango de edad y que posiblemente tenga un nivel de estudio en común que le permita acceder a un determinado tipo de trabajo y existirá otro donde posiblemente la determinación del cluster se dará a partir de su ubicación geográfica o de su constructo familiar.

Figura 3*Pasos para la selección de componentes principales*

Un ejemplo de aplicación del método de análisis de componentes principales es la investigación realizada por Restrepo et al. (2012), para la evaluación de tres variedades de pasto muestra que la elección del método R , utilizando el criterio de Kaiser a través del programa SAS se muestran 2 componentes los cuales son los más importantes en función de cada tipo de variedad de pasto, en este caso las variables que definieron al factor uno fueron la altura desde el piso, ancho de la parte media de la última hoja, y la biomasa y del componente dos el porcentaje de establecimiento.

Tabla 1

Matriz de componentes principales

Variables ¹	Variedad 1		Variedad 2		Variedad 3	
	CP1 ² (Crec. ³)	CP2 ² Calidad ³	CP1 Crec.	CP2 Calidad	CP1 Crec.	CP2 Calidad.
A	0.9273	-0.1995	0.8168	-0.2941	0.7870	-0.2906
B	0.8712	-0.2782	0.6749	-0.5271	0.3326	-0.6271
C	0.9137	-0.1795	0.8619	-0.0220	0.8634	-0.0593
D	0.3380	0.1401	0.5438	-0.4818	0.4481	-0.4401
E	-0.5561	0.5277	-0.6207	0.0107	-0.5185	0.3738
F	0.7064	0.6254	0.7846	0.5721	0.7769	0.5312
G	0.4944	0.8003	0.5792	0.7507	0.6902	0.6762

¹Altura desde el piso, cm (A), Longitud de la última hoja completamente elongada, cm (B), Ancho de la parte media de la última hoja completamente elongada, cm (C), Número de hojas por plantas (D), Relación Hoja:Tallos, g/g (E), Biomasa, Kg MS/m² (F), Establecimiento, % (G).²CP = Componente principal. ³Crec. = Crecimiento vegetal, Calidad de semilla corresponde al nombre de los factores.

Figura 4

Plano factorial

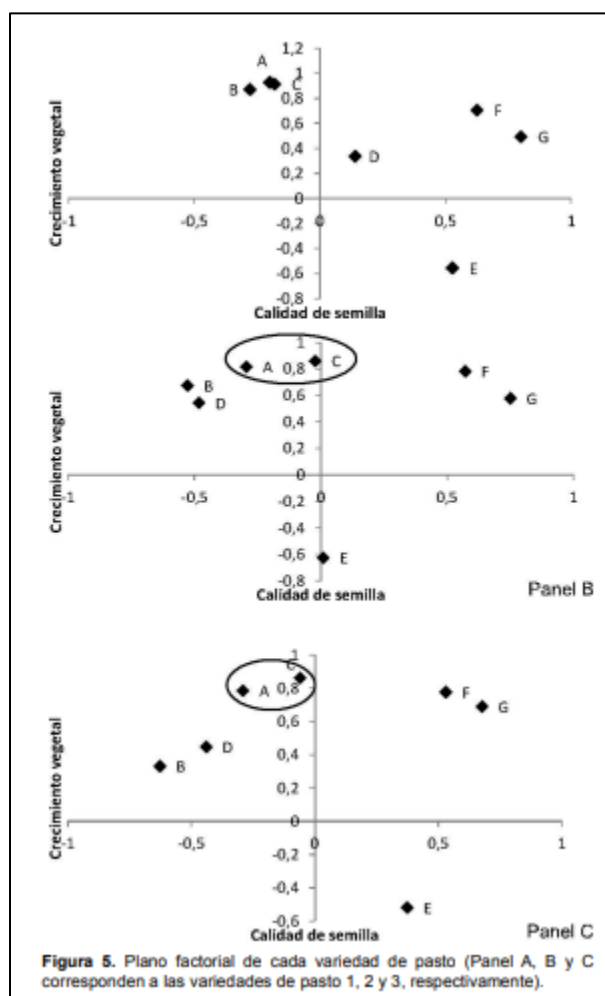


Figura 5. Plano factorial de cada variedad de pasto (Panel A, B y C corresponden a las variedades de pasto 1, 2 y 3, respectivamente).

En la figura 4, durante el trabajo de campo se obtuvieron ocho variables, identificadas desde la A hasta la G. A partir del análisis de componentes principales fue posible sintetizar la información o reducir el número de variables. La denominación de los componentes principales fue deducida tras observar la relación de ellos con las variables iniciales, de tal forma que a partir de variables observables se obtuvieron variables no observables o componentes. Las variables no observables generadas se denominaron crecimiento vegetal y calidad de la semilla, las cuales representaron el primer y segundo componente principal, respectivamente. (Restrepo et al., 2012, p. 264)

3.1.4. Índice de silhouette

El coeficiente de Silueta es una métrica para evaluar la calidad del agrupamiento obtenido con algoritmos de clustering. El objetivo de Silueta es identificar cuál es el número óptimo de agrupamientos. En los algoritmos de aprendizaje no supervisado, la cantidad de grupos puede ser un parámetro de entrada del algoritmo o puede ser determinado automáticamente por el algoritmo. (Banchero, 2015, p. 1)

El índice de Silhouette permite evaluar la bondad de los resultados del agrupamiento, este resulta de relevancia pues ayuda para evitar encontrar patrones en datos aleatorios. el índice de Silhouette es útil para la validación interna de clusters, validación externa de clusters, validación relativa de clusters, este último que se refiere a la evaluación de la estructura de agrupamiento dependiendo del número de grupos (Nabi, 2021).

$$s(i) = \frac{b - a}{\max(a, b)}$$

Donde:

a es el promedio de las disimilitudes (o distancias) de la observación i con las demás observaciones del cluster al que pertenece i.

b es la distancia mínima a otro cluster que no es el mismo en el que está la observación i . Ese cluster es la segunda mejor opción para i y se lo denomina vecindad de i .

El valor de $s(i)$ puede ser obtenido combinando los valores de a y b como se muestra a continuación:

$$s(i) = \begin{cases} 1 - \frac{a}{b}, & \text{si } a < b \\ 0, & \text{si } a = b \\ \frac{b}{a} - 1, & \text{si } a > b \end{cases}$$

El coeficiente de Silhouette está comprendido ente -1 y 1

$s(i) \approx 1$ la observación i está bien asignada al cluster

$s(i) \approx 0$ la observación i está cercana a dos clusters

$s(i) \approx -1$ la observación i está mal asignada al cluster

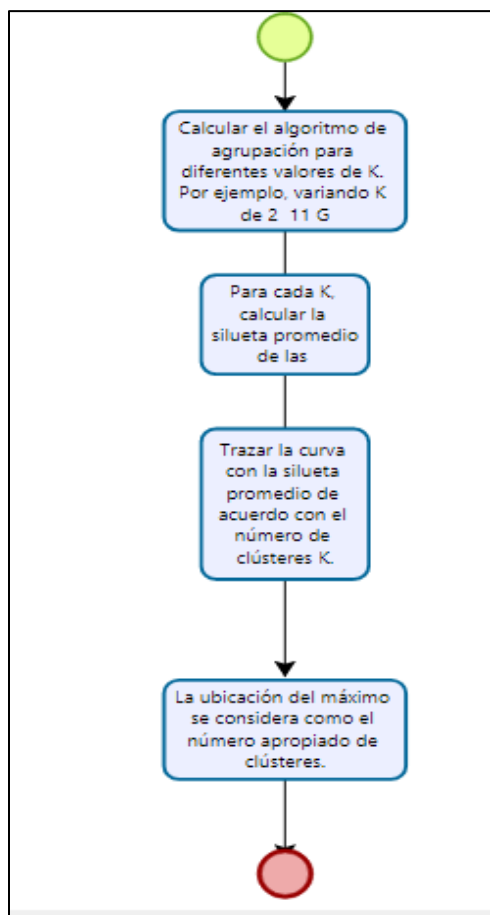
Según Schiatti (2017), en el contexto de evaluar la eficacia de distintos algoritmos de agrupación para identificar grupos de individuos con debilidades conceptuales similares, el índice de Silhouette se utiliza como una medida de confianza. Este índice representa el promedio de los valores de Silhouette para cada observación, indicando la fiabilidad de la asignación de cada punto de datos a su respectivo grupo. Valores cercanos a 1 señalan una alta confianza en la agrupación realizada, mientras que valores próximos a -1 sugieren que la agrupación no es fiable.

Para evaluar la calidad de los resultados obtenidos con diferentes algoritmos de agrupación, se utilizan índices de validación interna. Estos índices miden la eficacia de la estructura interna de los grupos formados por cada algoritmo, y en este caso se consideran los índices de Dunn y Silhouette. Dado que estos índices evalúan la separación entre grupos y la cohesión dentro de los grupos respectivamente, se busca maximizar sus valores. Por lo tanto, se considera que el algoritmo más preciso es aquel que presenta los índices más altos.

Tabla 2*Valores calculados*

	Jerárquico Distancia Euclidiana Método Ward	Jerárquico Distancia Euclidiana Método Ward	K-means (MacQueen)
Indice de Dunn	1	3	0.2926338
Indice de Silhouette	0.9673971	0.974685	0.5512583

En la tabla 2 consideramos los valores mostrados en la tabla, se tiene que el algoritmo que obtiene los mejores índices de validación, es decir, los más altos.

Figura 5*Algoritmo para la agrupación de clúster*

3.2. Fuente de datos

El presente trabajo de investigación utilizara como fuente de datos la encuesta Nacional de Empleo, desempleo y subempleo ENEMDU año 2023.

Antes de la aplicación de los instrumentos utilizados en la investigación se realizará un análisis exploratorio de los datos, muchos más en aquellos cuantitativos donde se observará los valores atípicos y la injerencia de estos en los cambios del comportamiento de la población para luego tomar la decisión de incluir o no estos datos en función de su impacto; además el análisis exploratorio permite producir datos de mejor calidad.

Para la construcción de factores es necesario normalizar los datos extraídos de la fuente porque la normalización está en las técnicas de reducción de la dimensionalidad, como el ACP, este identifica la dirección en la que se maximiza la varianza de los datos. La normalización de la media por sí sola no es suficiente, porque el ACP considera tanto la media como la varianza, y diferentes escalas de rasgos distorsionarían el análisis

Se trabajará mediante el paquete estadístico R-studio, utilizando librerías tales como las descritas a continuación:

Tabla 3*Librerías utilizadas*

Librerías	Características
haven	Haven permite que R lea y escriba varios formatos de datos utilizados por otros paquetes estadísticos al incluir la biblioteca ReadStat C escrita por Evan Miller. Haven es parte de tidyverse. Actualmente admite: SAS: read_sas() lee . sas7bdat +.
readxl	El paquete readxl de R facilita la importación de datos desde archivos Excel, ya sean .xls o .xlsx. Para lograr esto, el paquete readxl proporciona la función read_excel(), diseñada específicamente para leer datos de archivos Excel
dplyr	Dplyr es un componente esencial del ecosistema tidyverse en R, y se destaca como uno de los paquetes más utilizados. Su potencia radica en la facilidad con la que permite manipular datos, abarcando desde datos nativos de R hasta bases de datos (sin requerir conocimientos de SQL) e incluso entornos como Spark
gapminder	Esta herramienta facilita la correlación entre distintos indicadores, permitiendo, a través de una representación animada, el análisis de su evolución a lo largo del tiempo.
summarytools	Esta herramienta, ideal para quienes se inician en R, ofrece una manera sencilla y potente de resumir datos. Entre sus cuatro funciones útiles, destaca 'dfSummary', que genera un resumen gráfico y detallado de las variables presentes en un dataframe
FactoMineR	FactoMineR es una herramienta versátil para análisis multivariados, que abarca desde la agrupación jerárquica hasta el análisis factorial. Para comenzar, es necesario instalar y cargar los paquetes requeridos en R, y luego asignar los datos a un objeto de R para su manipulación.
ggplot2	ggplot2 es un paquete de R diseñado para crear gráficos visualmente atractivos. A diferencia de otros paquetes, ggplot2 utiliza la gramática de los gráficos, lo que permite construir gráficos combinando elementos independientes. Este paquete está diseñado para trabajar de forma iterativa, permitiendo comenzar con una capa que muestra los datos brutos y luego agregar capas adicionales, como anotaciones y estadísticos
factoextra	Este paquete ofrece las herramientas necesarias para la representación visual de los resultados del análisis de componentes principales, incluyendo técnicas como el gráfico de sedimentación y el biplot, entre otras opciones de visualización.

3.3. Modalidades de investigación

3.3.1. Bibliográfico - documental

Se caracteriza por la utilización de los datos secundarios como fuente de información. Su objetivo principal es dirigir la investigación desde dos aspectos, primeramente, relacionando datos ya existentes que proceden de distintas fuentes y posteriormente proporcionando una visión panorámica y sistemática de una determinada cuestión elaborada en múltiples fuentes. (Reyes & Alvarado, 2020, p. 1)

3.3.2. De escritorio

La investigación documental es un enfoque que se basa en la recopilación de datos, información y conocimientos a partir de fuentes ya existentes. Esto implica utilizar materiales publicados, como informes, documentos, datos de encuestas previas, y recursos disponibles en bibliotecas y sitios web.

La investigación secundaria, también conocida como investigación de escritorio, se basa en el análisis de datos ya existentes y validados, lo que aumenta la confiabilidad de la información. Aunque estos datos pueden no coincidir completamente con los objetivos del investigador, su propósito es proporcionar una comprensión más profunda del tema. Entre las fuentes comunes de investigación secundaria se incluyen datos de agencias gubernamentales y no gubernamentales, bibliotecas, informes de instituciones educativas, publicaciones periódicas y recursos en línea (Velázquez, 2024).

3.4. Limitaciones metodológicas

Dentro de las posibles limitaciones metodológicas se puede encontrar los sesgos en los datos y los errores de validación en los clústeres.

La detección de sesgos en los datos es esencial para la validez y confiabilidad de cualquier estudio. Para minimizar estos sesgos, se emplea la aleatorización, un método que asigna participantes o datos a grupos de forma completamente aleatoria, garantizando así la igualdad de oportunidades de asignación y reduciendo la posibilidad de sesgos en los grupos. Asimismo, el tamaño de la muestra es un factor crítico; una muestra inadecuada puede resultar en estudios con baja potencia y falta de detección de efectos significativos, mientras que muestras excesivamente grandes pueden consumir recursos sin aportar un valor sustancial (Appinio, 2024).

En el caso puntual de los sesgos de los datos una de las opciones más adecuadas es la transformación de la escala de la variable, además de técnicas de validación cruzada como el bootstrapping que permite reducir el sesgo de análisis.

El bootstrap es un tipo de técnica de remuestreo de datos que permite resolver problemas relacionados con la estimación de intervalos de confianza o la prueba de significación estadística (...) el método permite aproximar la distribución de muestreo de un estadístico y de sus propiedades mediante un procedimiento muy simple: Crear un gran número de muestras con reposición de los datos observados. (Ledesma, 2008, pp. 52-53)

La determinación de los hallazgos en relación a los factores que trabajan sobre los jóvenes “sisi” deben enfocarse en la promoción y ejecución de políticas de estado que primero garanticen que este grupo de jóvenes deben dedicarse de manera exclusiva a estudiar durante su etapa formativa. Sin embargo, de este grupo de jóvenes también refleja una nueva realidad completamente diferente para Ecuador y Latinoamérica. Ahora, es fundamental que estas políticas o leyes procuren determinar que el número de jóvenes sisi, se equipare con el de los nini. Y este cambio, podría impulsar diversos factores, como la necesidad de obtener ingresos

adicionales, la búsqueda de desarrollo profesional o la adaptación a un mercado laboral cada vez más demandante de perfiles calificados. Pero, en cualquier caso, los resultados deben promover una tendencia realmente positiva para el desarrollo del país, pues no sólo se encuentra formando profesionales, sino impulsando su economía, al fomentar el consumo y la creación de nuevas empresas.

Al analizar los datos de encuestas, es crucial reconocer los posibles sesgos y limitaciones derivadas de la presencia de datos faltantes. Estos datos ausentes pueden introducir distorsiones significativas, ya que su distribución no suele ser aleatoria y puede estar correlacionada con variables de interés. Por ejemplo, individuos con menor acceso a servicios o mayor vulnerabilidad podrían ser menos propensos a participar en encuestas, generando una subrepresentación de sus experiencias.

Además, las respuestas a las encuestas están sujetas a sesgos de autoselección y deseabilidad social, donde los participantes pueden ofrecer respuestas que consideran más aceptables en lugar de reflejar su realidad. La falta de información en ciertas variables clave, como ingresos o condiciones laborales, limita la capacidad de realizar análisis exhaustivos y puede llevar a conclusiones sesgadas sobre la situación de la población joven. Por lo tanto, es fundamental emplear técnicas de imputación y análisis de sensibilidad para mitigar estos sesgos, y reconocer las limitaciones inherentes a los datos de encuestas al interpretar los resultados

Capítulo IV: Resultados

4.1. Comprensión de los datos

A partir de la base de datos obtenida para la investigación sobre la situación socio – demográfica de los jóvenes sisi, se puede identificar variables tales como: estado civil, sexo, edad, etnia, nivel de instrucción, y demás datos que a continuación se detalla:

- Estado Civil
- Sexo
- Edad
- Etnia
- Nivel de Instrucción
- Ocupación
- Años de trabajo
- Número de trabajos en la semana
- Provincia
- Recibe el bono de desarrollo humano
- Total de horas trabajadas en la semana

Respecto a la importancia de las variables escogidas, la variable de relevancia para la segmentación corresponde a la edad en años de las personas que si estudian y que si trabajan, a partir de ellas se pretende investigar el o los perfiles sociodemográficos de la población en investigación, un perfil sociodemográfico es un conjunto de características demográficas y socioeconómicas de el o los individuos que describan a el grupo poblacional y es de utilidad pues contribuye a entender las necesidades, comportamientos y tendencias de una población, además colabora de manera importante en la creación de clusters a través de la identificación de

variables relevantes, además direcciona la agrupación de individuos que por las características asociadas a las variables se pueden observar similares pues a su vez se divisa patrones y tendencias primarias , principio fundamental en la creación de clusters.

4.2. Recopilación de los datos iniciales

La base de datos para este estudio se obtiene de la información suministrada por el INEC en su MANUAL DEL ENCUESTADOR/A ENCUESTA NACIONAL DE EMPLEO, DESEMPLEO Y SUBEMPLEO ENEMDU con un corte de información a diciembre del 2023. La Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU) es una herramienta estadística clave dentro del Sistema Integrado de Encuestas de Hogares (SIEH). Su diseño metodológico la convierte en un instrumento fundamental para el estudio del empleo en el país, la caracterización del mercado laboral, la actividad económica de los ecuatorianos y sus fuentes de ingreso. Además, permite dimensionar fenómenos sociodemográficos, proporcionando datos e indicadores sobre las condiciones de vida de hombres y mujeres en sus hogares, especialmente relevantes para el grupo de estudio de esta investigación.

4.3. Exploración de los datos

Para llevar a cabo con la exploración de los datos se ha empleado la herramienta de visualización R- Studio. Esta facilita la visualización gráfica y detallada de estadísticas correspondientes a cada variable que integrará el análisis.

A partir de lo indicado inicialmente se obtuvo los siguientes resultados:

- Identificación sociodemográfica
- #cruce de variables entre sexo y estado civil
- `table(p02,p06)`
- `tablasexo_edad <-table(p02,p06)`
- `prop.table(tablasexo_edad)`

Tabla 4*Estado civil de los jóvenes según sexo declarado*

Estado Civil	Casado(a)	Separado (a)	Divorciado (a)	Viudo (a)	Unión Libre	Soltero (a)	TOTAL
Sexo							
Hombres	104	11	8	1	258	2448	2830
Mujeres	107	100	18	4	232	1853	2314
Total	211	111	26	5	490	4301	5144

Tabla 5*Porcentaje de estado civil de los jóvenes según sexo declarado*

Estado Civil	Casado(a)	Separado (a)	Divorciado (a)	Viudo (a)	Unión Libre	Soltero (a)	TOTAL
Sexo							
Hombres	0.02	0.00	0.00	0.00	0.05	0.48	0.55
Mujeres	0.02	0.02	0.00	0.00	0.05	0.36	0.45
Total	0.04	0.02	0.00	0.00	0.1	0.84	1

En la tabla 5, se muestra la población joven comprendida entre 18 a 29 años es un 10% más representada por hombres, en general esta población se caracteriza por ser soltera y además independientemente del sexo que declaran, en los dos hay más solteros que cualquier otro. De acuerdo a este primer resultado se observa que los jóvenes que estudian y trabajan posiblemente viven en casa de sus padres, al ser solteros, otros cuantos mientras tanto, probablemente están independizados. El Centro Nacional de Planeamiento Estratégico del Perú en 2016 mencionaba cifras similares a las anteriormente expuestas, revelando que su población joven que estudia y trabaja son 54% hombres y 46% mujeres, además que la mayoría son solteros sin responsabilidades conyugales o cargas familiares (Centro Nacional de Planeamiento Estratégico, 2016).

```
#cruce de variables entre edad sexo y estado civil
```

```
fable(xtabs(~ p03+p02+p06, data=SISI))
```

```
tablas_e_e<-fable(xtabs(~ p03+p02+p06, data=SISI))
```

prop.table(tablas_e_e)

Tabla 6*Clasificación de los jóvenes según estado civil y edad*

Estado Civil	Sexo	Casado (a)	Separado (a)	Divorciado (a)	Viudo (a)	Unión Libre	Soltero (a)	Total
Edad								
18	Hombre	0	0	0	0	16	283	299
	Mujer	1	3	0	0	5	162	171
19	Hombre	0	0	0	0	10	21	31
	Mujer	0	3	0	0	7	203	213
20	Hombre	2	0	0	0	14	286	302
	Mujer	2	5	0	0	12	206	225
21	Hombre	3	0	0	0	18	251	272
	Mujer	6	5	0	0	15	215	241
22	Hombre	5	2	0	0	21	283	311
	Mujer	4	6	0	1	24	254	289
23	Hombre	5	1	0	0	23	290	319
	Mujer	8	6	0	0	27	182	223
24	Hombre	5	0	0	0	22	226	253
	Mujer	14	9	1	0	23	175	222
25	Hombre	11	1	2	0	20	155	189
	Mujer	4	8	2	0	15	135	164
26	Hombre	12	2	0	0	26	154	194
	Mujer	12	19	2	1	26	111	171
27	Hombre	17	1	3	0	29	115	165
	Mujer	13	16	3	0	25	91	148
28	Hombre	18	0	1	0	16	114	149
	Mujer	21	16	3	2	31	69	142
29	Hombre	26	4	2	1	43	70	146
	Mujer	22	4	7	0	22	50	105
Total		211	111	26	5	490	4101	4944

Tabla 7*Porcentaje de Clasificación de los jóvenes según estado civil y edad*

Estado Civil	Sexo	Casado o (a)	Separado (a)	Divorciado (a)	Viudo (a)	Unión Libre	Soltero (a)	Total I
18	Hombre	0,00	0,00	0,00	0,00	0,00	0,06	0,06
	Mujer	0,00	0,00	0,00	0,00	0,00	0,03	0,03
19	Hombre	0,00	0,00	0,00	0,00	0,00	0,04	0,04
	Mujer	0,00	0,00	0,00	0,00	0,00	0,04	0,04
20	Hombre	0,00	0,00	0,00	0,00	0,00	0,06	0,06
	Mujer	0,00	0,00	0,00	0,00	0,00	0,04	0,04
21	Hombre	0,00	0,00	0,00	0,00	0,00	0,05	0,05
	Mujer	0,00	0,00	0,00	0,00	0,00	0,04	0,04
22	Hombre	0,00	0,00	0,00	0,00	0,00	0,06	0,06
	Mujer	0,00	0,00	0,00	0,00	0,00	0,05	0,05
23	Hombre	0,00	0,00	0,00	0,00	0,00	0,06	0,06
	Mujer	0,00	0,00	0,00	0,00	0,01	0,04	0,05
24	Hombre	0,00	0,00	0,00	0,00	0,00	0,04	0,04
	Mujer	0,00	0,00	0,00	0,00	0,00	0,03	0,03
25	Hombre	0,00	0,00	0,00	0,00	0,00	0,03	0,03
	Mujer	0,00	0,00	0,00	0,00	0,00	0,03	0,03
26	Hombre	0,00	0,00	0,00	0,00	0,01	0,03	0,04
	Mujer	0,00	0,00	0,00	0,00	0,01	0,02	0,03
27	Hombre	0,00	0,00	0,00	0,00	0,01	0,02	0,03
	Mujer	0,00	0,00	0,00	0,00	0,00	0,02	0,02
28	Hombre	0,00	0,00	0,00	0,00	0,00	0,02	0,02
	Mujer	0,00	0,00	0,00	0,00	0,01	0,01	0,02
29	Hombre	0,01	0,00	0,00	0,00	0,09	0,01	0,11

Mujer	0,01	0,00	0,00	0,00	0,00	0,00	0,01	0,02
TOTAL	0,02	0,00	0,00	0,00	0,00	0,14	0,84	1,00

#cruce de variables entre sexo y etnia

fable(xtabs(~ p02+p15, data=SISI))

tabla_sex_ni<-fable(xtabs(~ p02+p15, data=SISI))

prop.table(tabla_sex_ni)

Tabla 8

Etnia de los jóvenes según sexo declarado

Etnia	Indígena	Afroecuatoriano (a)	Negro (a)	Mulato (a)	Montubio (a)	Mestizo (a)	Blanco (a)	Otro	Total
Sexo									
Hombre	218	39	23	15	20	2492	23	0	2830
Mujer	174	17	9	15	18	2065	15	1	2314
Total	392	56	32	30	38	4557	38	1	5144

Tabla 9

Porcentaje de etnia de los jóvenes según sexo declarado

Etnia	Indígena	Afroecuatoriano (a)	Negro (a)	Mulato (a)	Montubio (a)	Mestizo (a)	Blanco (a)	Otro	Total
Sexo									
Hombre	0,04	0,01	0,01	0,00	0,00	0,50	0,01	0,00	0,57
Mujer	0,03	0,00	0,00	0,00	0,00	0,40	0,00	0,00	0,43
Total	0,07	0,01	0,01	0	0	0,9	0,01	0	1

En la tabla 9, de acuerdo a lo revisado, se puede observar que la mayoría de los jóvenes insertados en el mercado laboral y en el estudio son mestizos, además probablemente este porcentaje se debe a que la mayoría de la población en el Ecuador son mestizas, seguido de esto, los jóvenes que estudian y trabajan son los indígenas, y las demás nacionalidades representan el 1%.

```
# estadísticos edad
```

```
sum(p03)
```

```
median(p03)
```

```
min(p03)
```

```
max(p03)
```

Tabla 10

Estadísticos de edad

Suma	Media	Mínimo	Máximo
117.468	23	18	29

```
#GRUPO DE OCUPACION
```

```
install.packages("summarytools")
```

```
library(summarytools)
```

```
tabocup<-freq(SISIS$p42)
```

```
tabocup
```

Tabla 11

Grupos de ocupación

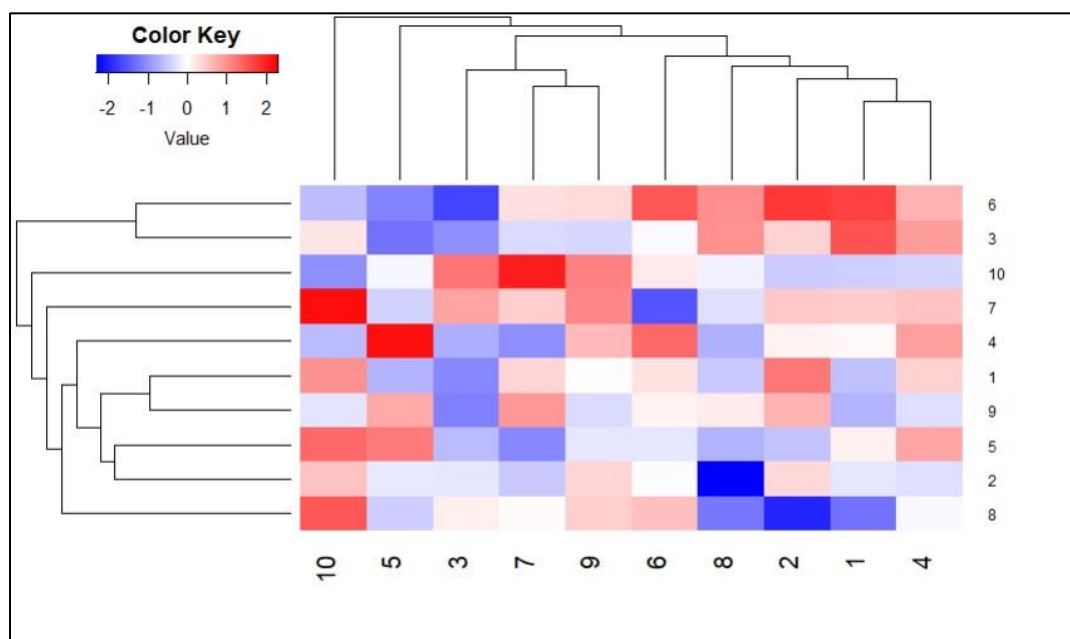
	Frecuencia	Porcentaje	Porcentaje acumulado	Total	Total Acumulado
Empleado público	332	6,45	6,45	6,45	6,45
Empleado privado	2449	47,60	54,06	47,61	54,06
Empleado tercerizado	1	0,02	54,08	0,02	54,08
Jornalero	261	5,07	59,16	5,07	59,16
Patrón	32	0,62	59,78	0,62	59,78
Cuenta Propia	831	16,16	75,99	16,16	75,93
Trabajador del hogar	1109	21,56	97,49	21,56	97,49

Trabajador del hogar en otro hogar	22	0,43	97,92	0,43	97,92
Ayudante	15	0,29	98,21	0,29	98,21
Otros	92	1,79	100	1,78	100
Total	5144	100	100	100	100

En la tabla 11, con respecto a la ocupación del grupo estudiado, y a las plazas de empleo, estos se desempeñan mayormente en el sector privado como obrero u empleado, esto quiere decir que comúnmente están en puestos operativos, o auxiliares de alguna área situación que coincide con el estudio de CEPLAN, que manifiesta que los jóvenes se desempeñan en el sector privado.

Figura 6

Aproximación gráfica a el agrupamiento jerárquico con k.medias. Heatmap



En la figura 6 es un heatmap que visualiza una matriz de datos a través de colores, donde el azul representa valores negativos, el rojo valores positivos y el blanco valores cercanos a cero. Los dendrogramas en los ejes de filas y columnas muestran agrupaciones jerárquicas, revelando similitudes entre los elementos. La intensidad del color en cada celda refleja la magnitud de los datos en esa intersección, permitiendo identificar patrones y relaciones entre filas y columnas.

En resumen, el heatmap, junto con los dendrogramas, ofrece una visión clara de la distribución y las relaciones en los datos, destacando agrupaciones y valores extremos. Esta representación visual facilita la identificación de patrones y la comprensión de la estructura subyacente de los datos.

Tabla 12

Determinación de k-medias

Cuadro Metodológico k medias	
Instrumento	k-medias
Selección del número de clusters	se inicia con k=3 por defecto
Criterios de inclusión	los valores medios se trabajan a traves de medias ponderadas
Herramienta de análisis	R-studio

```
#kmedias
```

```
kmeans(BASEACP,centers = 3,iter.max = 10,nstart =1 )
```

```
fviz_nbclust(BASEACP, kmeans, method = "wss") + geom_vline(xintercept = 4,
linetype = 2)
```

Tabla 13

K-medias

Cluster means	Edad	Sexo	Seguro social	Estado Civil	Disp electrónicos	Etnia	Lugar nacimiento
1	21,89	1,47	8,82	5,73	3,27	5,37	1,28
2	23,68	1,44	5,31	5,48	2,68	5,72	1,29
3	23,75	1,35	6,51	5,39	2,42	5,58	1,34

Tabla 14*Medias por grupos*

Total de horas trabajadas en la semana	Ocupación	Años de Trabajo	Cuantos trabajos tuvo durante la semana	Recibe el bono	Provincia
15,77	4,97	3,29	1,03	1,98	11,32
38,42	2,99	2,54	1,04	1,99	11,56
54,48	3,27	2,981	1,14	2,00	12,14

Tabla 15*Caracterización de clusters*

Clúster	Caracterización del clúster
Clúster 1	La edad de los individuos dentro del clúster fluctúa de entre los 21 a 22 años, siendo indiferente el sexo de estos, cuentan con seguros privados sin hospitalización, además el clúster se caracteriza por una identificación de estado civil soltera. En cuanto a su disponibilidad de instrumentos electrónicos para el desarrollo de sus labores educativas, este grupo no cuenta con los equipos electrónicos para realizar esta actividad y generalmente viven en parroquias rurales, además tienen una jornada laboral mínima, por lo cual se puede concluir que son trabajadores de medio tiempo con una experiencia laboral de alrededor de 3 años, en una tercerizadora.
Clúster 2	El cluster 2 mientras tanto se caracteriza por personas entre 23 y 24 años hombres y mujeres que cuentan con seguros privados con hospitalización, solteros y con una disponibilidad limitada de recursos electrónicos para desempeñar sus labores académicas, además residen normalmente en su lugar de nacimiento y se podría deducir además que trabajan a tiempo completo, en el sector privado con alrededor de dos años y medio de trabajo
Clúster 3	El tercer clúster se caracteriza por tener a individuos de entre 23 y 24 años solteros, con seguro privado sin hospitalización, solteros, que cuentan con un equipo electrónico de uso compartido para la realización de sus tareas mayormente de etnias mestizas y que residen en el mismo lugar de nacimiento, además son individuos que tienen una jornada entera de trabajo.

Contrastando la realidad de Ecuador con el estudio de CEPLAN, Perú, los peruanos que estudian y trabajan en edad joven tienen una participación escasa en el sector públicos, al igual que la realidad ecuatoriana, su participación es mayor en el mercado privado, trabajando en unidades productivas pequeñas como jornaleros o empleados, además algo que refleja el estudio del vecino país es que los jóvenes mencionan que en general la calidad de sus empleos les resulta malo. Mientras tanto un estudio realizado en 2023 en Europa revela los siguientes resultados para la población joven universitaria:

(...) 61,7% de los estudiantes trabajaban con un contrato a tiempo completo o parcial, pero un 38,3% estaba sujeto a condiciones precarias o ilegales, propias de un mercado de trabajo precarizado y especialmente ensañado con los más jóvenes, que llevan muchos años sufriendo tasas de desempleo e infraempleo de las más altas de la Unión Europea. (Sofia et al., 2024, p. 6)

De acuerdo a la tabla 15, los resultados se pueden determinar que la mayoría de las políticas públicas deben ir encaminadas a la implementación de programas de acceso a medios electrónicos, conectividad y desarrollo de habilidades digitales. Así mismos incentivos a las empresas que otorguen a los jóvenes trabajos de calidad, además de programas de capacitación laboral. En otro aspecto relacionado a la salud, cuidado de la salud mental y física de los jóvenes ampliando las coberturas de salud, hecho que ayudara también a la recesión económica del Instituto Ecuatoriano de Seguridad Social. Descentralizar los posibles programas de desarrollo y procurar que lleguen a los sectores rurales generando oportunidades de empleo, entre otros.

4.3.1. Análisis factorial de correspondencias

El análisis de correspondencias permite seleccionar las categorías de las variables que contribuyen significativamente a un factor y que sirve para caracterizar el mismos dando paso a una segmentación de las categorías de las variables e individuos que por ciertas características forman parte del factor, contribuye también a un análisis correlacionar de las distintas variables y categorías. El análisis de correspondencias y el k.medias se vinculan de tal modo que permite la identificación de patrones, selecciona las variables relevantes y permite de mejor manera la interpretación de los clusters pues el análisis mejora la calidad de los clusters al presentar una representación más precisa y compacta de los datos.

```

tablaacp<-table(BASEACP$prov,p06)

tablaacp

ACPSISI<- CA(tablaacp, graph =TRUE)

print(ACPSISI)

summary(ACPSISI)

```

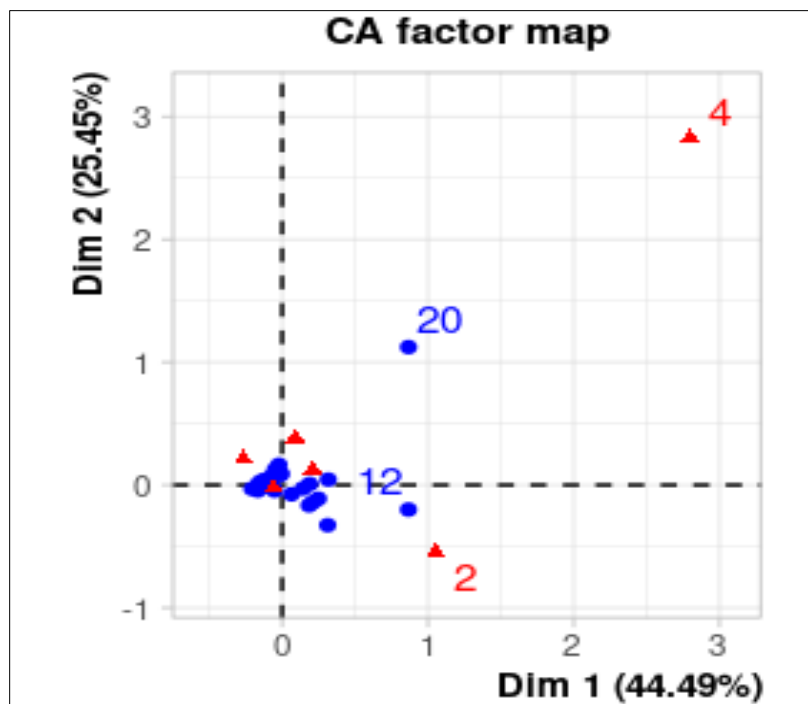
Tabla 16

Análisis factorial de correspondencias

Estado Civil	Casado (a)	Separado (a)	Divorciado (a)	Viudo (a)	Unión Libre	Soltero (a)
Provincia						
AZUAY	21	3	6	1	28	509
BOLIVAR	1	5	1	1	15	80
CAÑAR	8	2	3	0	5	68
CARCHI	0	0	0	0	10	60
COTOPAXI	5	1	0	0	7	140
CHIMBORAZO	4	0	0	0	6	124
EL ORO	34	8	3	0	66	449
ESMERALDAS	6	1	1	0	28	131
GUAYAS	24	37	2	0	56	512

IMBABURA	6	4	2	0	17	172
LOJA	6	1	1	0	15	160
LOS RIOS	1	9	0	1	7	47
MANABI	3	5	0	0	11	71
MORONA SANTIAGO	1	2	0	0	16	51
NAPO	0	2	0	0	14	69
PASTAZA	8	1	1	0	20	145
PICHINCHA	32	3	2	0	53	580
TUNGURAHUA	20	2	2	0	29	433
ZAMORA CHINCHIPE	4	0	1	0	16	67
GALAPAGOS	10	2	0	2	11	29
SUCUMBIOS	4	4	0	0	18	89
ORELLANA	2	1	1	0	18	116
SANTO DOMINGO	1	8	0	0	9	78
SANTA ELENA	10	10	0	0	15	121

Figura 7
Gráfico CPA



```
#correspondencia multiple
```

```
mult<- PCA(BASEACP, graph = TRUE)
```

```
print(mult)
```

```
summary(mult)
```

Tabla 17

Eigen valores

Eigenvalues	Dim .1	Dim.2	Dim.3	Dim.4	Dim.5
Variance	0,039	0,022	0,015	0,007	0,004
% of var	44,493	25,454	17,341	7,703	5,009
Cumulative % of var	44,493	69,948	87,289	94,991	100

En la tabla 17, el análisis factorial permite revisar que a través de la varianza acumulada los tres primeros factores son los que recogen la mayor cantidad de varianza, es decir son los más representativos, en especial la primera que tiene más del 44% de la varianza explicada dentro del plano factorial, es decir la mayoría de factores se encuentran en ese plano o factor.

Tabla 18*Ponderación de filas*

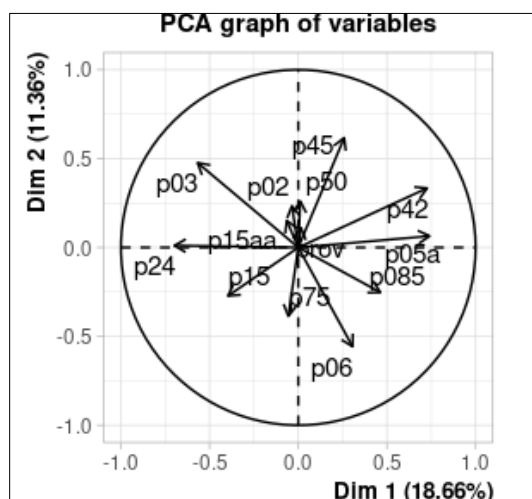
ROWS (the 10 first)	Iner*1000	Dim 1	Ctr	cos2	Dim2	ctr	cos2	Dim3	ctr	cos2
1	5,061	-0,151	6,417	0,496	0,024	0,292	0,013	0,125	11,266	0,34
2	3,429	0,317	5,134	0,586	0,044	0,173	0,011	-0,064	0,53	0,024
3	4,347	-0,047	0,093	0,008	0,127	1,200	0,062	0,222	5,39	0,189
4	1,265	-0,097	0,325	0,1	-0,015	0,014	0,002	-0,241	5,169	0,623
5	1,528	-0,166	2,097	0,537	-0,045	0,272	0,040	0,096	1,811	0,181
6	1,743	-0,206	2,831	0,635	-0,031	0,112	0,014	0,087	1,293	0,113
7	2,143	-0,010	0,028	0,005	0,083	3,372	0,352	-0,057	2,348	0,167
8	2,314	-0,009	0,006	0,001	0,095	1,317	0,127	-0,247	12,934	0,853
9	8,183	0,185	10,715	0,512	-0,166	15,064	0,412	0,05	1,998	0,037
10	0,412	-0,051	0,264	0,25	-0,046	0,366	0,199	0,012	0,040	0,015

Tabla 19*Ponderación columnas*

Columns	Iner*1000	Dim 1	Ctr	cos2	Dim2	ctr	cos2	Dim3	ctr	cos2
1	11,501	0,092	0,886	0,030	0,383	26,808	0,522	0,191	9,801	0,13
2	31,306	1,053	61,129	0,764	-0,548	28,897	0,207	0,180	4,575	0,022
3	5,331	-0,265	0,905	0,066	0,211	1,006	0,042	0,226	1,691	0,048
4	19,214	2,797	19,437	0,396	2,825	34,656	0,404	1,216	9,417	0,075
5	16,691	0,209	10,659	0,25	0,124	6,591	0,088	-0,339	71,894	0,657
6	3,910	-0,057	6,984	0,699	-0,023	2,041	0,117	0,022	2,622	0,102

Figura 8

Gráfico variables-CPA



De acuerdo a la figura 8, las variables más representativas son las que se encuentran más lejos del origen y más cerca de los vértices, por tanto para el primer factor la variable más representativa es la p05a que corresponde a la variable Seguro social, otra de las que está cerca es la categoría de ocupación y el estado civil de las personas, además para la segunda dimensión una variable que resulta representativa es la p45 correspondiente a los años de trabajo, así también la variable p50 que es el número de trabajos diferentes en la semana. Por tanto, al

realizar una comparación con los cluster realizados con k medias se determina que existe una congruencia con los dos métodos de segmentación. A partir de ello se analiza que el acceso limitado a los artículos electrónicos y los empleos sin los beneficios necesarios no ayudan a los jóvenes a desarrollarse en ninguno de los dos ámbitos, el sector público es un limitante de generación de empleo joven por lo cual el sector privado representa un mayor porcentaje de inmersión de la demanda laboral joven que no necesariamente indica que tiene las condiciones necesarias e ideales para que los jóvenes puedan tener experiencia y profesionalización.

El análisis factorial y la comparación con el clúster de k-medias revelan patrones significativos de vulnerabilidad laboral y limitaciones en el desarrollo de la población joven: la precariedad laboral, evidenciada por la falta de protección social y la alta movilidad laboral, se correlaciona fuertemente con la ausencia de seguro social, la inestabilidad ocupacional y la dependencia del empleo privado en condiciones deficientes; además, el acceso limitado a artículos electrónicos y empleos sin beneficios obstaculiza el desarrollo profesional y personal, situación agravada por la restricción del sector público en la generación de empleo juvenil.

Para abordar estos desafíos, se requieren políticas públicas integrales que fortalezcan la inspección laboral, incentiven empleos de calidad en el sector privado, promuevan la formación profesional y la inclusión digital, fomenten el desarrollo profesional a través de mentorías y pasantías, y faciliten el emprendimiento mediante microcréditos; la implementación efectiva de estas políticas exige un seguimiento continuo, la colaboración entre sectores y la participación activa de los jóvenes en su diseño, con el fin de mejorar significativamente sus oportunidades de empleo, desarrollo y bienestar.

Tabla 20*Eigenvalores*

Eigenvalues	Dim .1	Dim.2	Dim.3	Dim.4	Dim.5	Dim .6
Variance	2,426	1,477	1,172	1,083	0,997	0,969
% of var	18,659	11,365	9,012	8,329	7,669	7,454
Cumulative % of var	18,659	30,024	39,036	47,366	55,035	62,489

	Dim.7	Dim.8	Dim.9	Dim.10	Dim.11	Dim.12	Dim.13
Variance	0,908	0,868	0,819	0,735	0,57	0,544	0,432
% of var	6,987	6,68	6,298	5,652	4,385	4,186	3,324
Cumulative % of var	69,476	76,156	82,454	88,105	92,49	96,676	100

Tabla 21*Individuos*

Individuals (the 10 first)	Dist	Dim 1	ctr	cos2	Dim2	Ctr	cos2	Dim3	ctr	cos2
1	2,851	1,505	0,018	0,279	-0,456	0,003	0,026	-0,163	0,000	0,003
2	4,128	-3,043	0,074	0,544	0,764	0,008	0,034	-1,253	0,026	0,092
3	3,036	-1,03	0,008	0,115	-0,548	0,004	0,033	-0,984	0,016	0,105
4	3,298	-1,062	0,009	0,104	-0,129	0,000	0,002	-1,029	0,018	0,097
5	10,044	0,982	0,008	0,010	2,071	0,056	0,043	-4,403	0,322	0,192
6	2,895	0,744	0,004	0,066	-2,002	0,053	0,478	-0,241	0,001	0,007
7	2,890	-1,634	0,021	0,320	-0,53	0,004	0,034	-0,042	0,000	0,000
8	3,517	-2,105	0,036	0,358	-1,007	0,013	0,082	-0,460	0,004	0,017
9	3,300	-2,492	0,05	0,570	-0,371	0,002	0,013	0,334	0,002	0,010
10	4,053	-2,417	0,047	0,356	-0,978	0,013	0,058	-0,355	0,002	0,008

Tabla 22*Variables*

Variables (the 10 first)	Dim 1	Ctr	cos2	Dim2	Ctr	cos2	Dim3	Ctr	cos2
Edad	-0,568	13,291	0,322	0,476	15,358	0,227	0,035	0,104	0,001
Sexo	-0,036	0,053	0,001	0,234	3,698	0,055	-0,631	34,107	0,399
Seguro Social	0,74	22,579	0,548	0,064	0,280	0,004	-0,076	0,489	0,006
Estado Civil	0,306	3,862	0,094	-0,557	21,004	0,310	0,202	3,485	0,041
Dispositivo elec	0,463	8,820	0,214	-0,253	4,316	0,064	0,024	0,050	0,001
Etnia	-0,396	6,467	0,157	-0,272	5,005	0,074	-0,315	8,449	0,099
Lugar nacimiento	-0,065	0,173	0,004	0,151	1,551	0,023	-0,078	0,519	0,006
Total de horas trabajadas en la semana	-0,698	20,102	0,488	0,011	0,008	0,000	0,293	7,349	0,086
Ocupación	0,726	21,747	0,528	0,334	7,545	0,111	-0,026	0,057	0,001
Años de Trabajo	0,259	2,762	0,067	0,617	25,743	0,38	0,35	10,432	0,122

4.3.2. Índice de silhouette

```
# Ejecutar k-means y almacenar el resultado en km
set.seed(123) # Fijar semilla para reproducibilidad

km <- kmeans(BASEACP, centers = 3, iter.max = 10, nstart = 1)

# Calcular distancias entre observaciones

dist_matrix <- dist(BASEACP)

# Calcular el índice de silhouette

sil <- silhouette(km$cluster, dist_matrix)

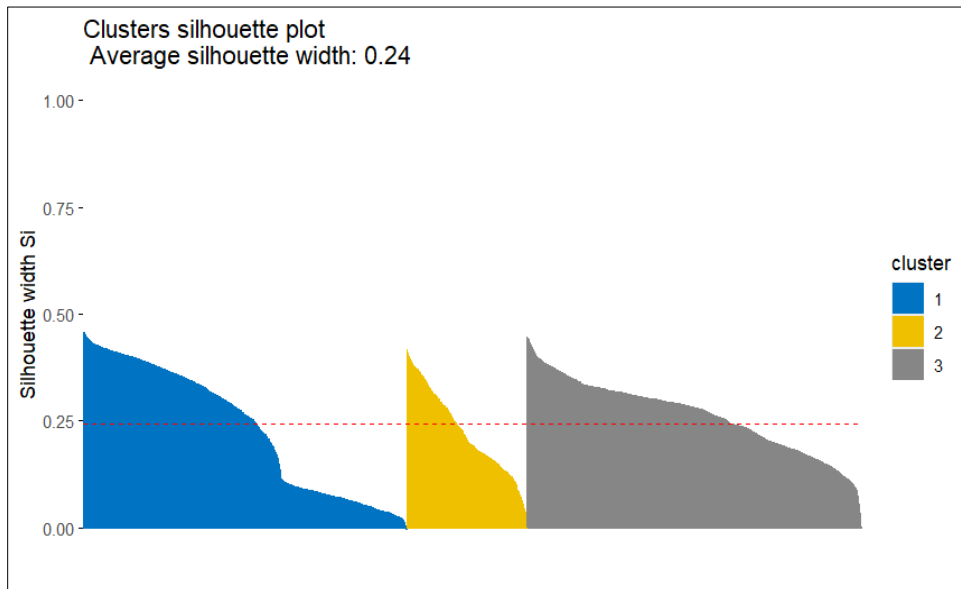
# Visualizar el índice de silhouette

fviz_silhouette(sil, palette = "jco")
```

Tabla 23*Clústers*

Cluster	Size	ave.sil.width
1	2135	0.23
2	795	0.22
3	2214	0.26

En la tabla 23, respecto a las distancias euclidianas para la división de los individuos para la formación de clusters el índice de silhouette indica que si bien los agrupamientos no se encuentran alejados de los promedios respecto al mismo grupo puede realizarse un mayor agrupamiento formando más de 3 grupos diferentes que puedan distinguir de mejor manera las características de los clusters en mención., el parámetro para determinar si los clúster se encuentran bien definidos es que primero sea positive, y mientras más cercano a uno este es más óptimo.

Figura 9*Indice de silhouette***Figura 10***Indice de Dunn*

Name	Type	Value
km_indices	list [34]	List of length 34
n	integer [1]	150
cluster.number	integer [1]	3
cluster.size	double [3]	50 36 64
min.cluster.size	double [1]	36
noisen	double [1]	0
diameter	double [3]	5.03 3.34 3.57
average.distance	double [3]	1.18 1.25 1.28
median.distance	double [3]	0.988 1.157 1.213
separation	double [3]	1.553 0.267 0.267
average.toothor	double [3]	3.65 3.11 2.78
separation.matrix	double [3 x 3]	0.000 2.992 1.553 2.992 0.000 0.267 1.553 0.267 0.000 ...
ave.between.matrix	double [3 x 3]	0.00 4.36 3.25 4.36 0.00 2.13 3.25 2.13 0.00 ...
average.between	double [1]	3.168981
average.within	double [1]	1.238354
n.between	integer [1]	7304
n.within	integer [1]	3871
max.diameter	double [1]	5.034198
min.separation	double [1]	0.2667788
within.cluster.ss	double [1]	142.627
clus.avg.silwidths	double [3]	0.640 0.397 0.354
avg.silwidth	double [1]	0.459889
g2	NULL	Pairlist of length 0
g3	NULL	Pairlist of length 0
pearsongamma	double [1]	0.6945761

En la figura 10 el índice de Dunn es una métrica utilizada en el análisis de clústeres para evaluar la calidad de la agrupación de datos. Su definición se basa en la relación entre la distancia mínima entre los clústeres y la distancia máxima dentro de un clúster. La importancia de este índice radica en su capacidad para determinar qué tan bien separados están los clústeres entre sí y cuán compactos son los puntos dentro de cada clúster. Un valor alto del índice de Dunn indica una mejor calidad de la agrupación, ya que sugiere que los clústeres están bien separados y los puntos dentro de ellos son muy similares. Su utilidad principal es ayudar a seleccionar el número óptimo de clústeres en un conjunto de datos, permitiendo a los analistas elegir la partición que mejor representa la estructura inherente de los datos.

De acuerdo a esta definición y utilizando este método como ratificación y símil del índice de Silhouette, respecto a la salida de R, indica que el número de clusters determinaos, en este caso 3 es una segmentación óptima para determinar los factores que influyen en la formación de los jóvenes sisi, el valor del índice es positivo y más cercano a 0,5 por lo cual la afirmación anterior respecto a la parametrización se muestra correctas.

Capítulo V: Conclusiones y Recomendaciones

Conclusiones

- Las características sociodemográficas de los jóvenes que trabajan y estudian en el Ecuador para el año 2023 muestra una población joven distribuida uniformemente respecto a su edad sin embargo existe un porcentaje representativo de los jóvenes que estudian y trabajan que cursan por la edad de 29 años, además se determina que los jóvenes identificados como sisi en su mayoría son solteros y de etnia mestiza.
- Mediante la técnica de k medias a partir de la determinación de 3 clasificaciones jerárquicas se pudo evidenciar que respecto a las características sociodemográficos los jóvenes están agrupados de acuerdo a su edad y su ingreso, además respecto a su ocupación muestra que se desempeñan en áreas generalmente operativas en trabajos generalmente no bien remunerados , además respecto a su nivel de instrucción se encuentra grupos que poseen bachillerato o en su defecto cursan niveles de educación para una formación técnica profesional.
- Los factores predominantes obtenidos a través del ACP y a ACFM indican que existen tres factores que tienen el mayor grado de varianza dentro de las categorías de las variables, además que los jóvenes muestran tres cluster diferenciados caracterizados de igual manera, por su edad, nivel de instrucción, estado civil, la aplicación del ACP junto con la congruencia del método de k medias indica que no existe aún un verdadero interés por el desarrollo paralelo de la población joven como fuerza de trabajo y como promesa de capacidades y ámbitos ligados a mejorar el país en miras de poder aportar y resurgir sectores productivos que hasta el momento no se han trabajado.

- El índice de silhouette indica que la creación de 3 cluster jerárquicos es bueno por lo cual se ratifica la utilización de herramienta estadística, para este caso la aplicación de un índice que refleje la confiabilidad del método permite identificar la buena elección de la metodología y la descripción del problema de investigación con métodos altamente comprobados.

Recomendaciones

- Es necesario la inclusión de políticas públicas diferenciadas que permitan manejar los grupos de acuerdo a las prioridades establecidas a través de investigaciones como la llevada a cabo con el fin de obtener Jóvenes productivos y bien capacitados, no solo teniendo mano de obra relacionada con la capacidad física sino más bien con la intelectual, conocer las características de los grupos de jóvenes y en general de los problemas sociales permite establecer políticas focalizadas que permitan distribuir los recursos de manera tal que se atiende a los sectores respecto a sus carencias y conflictos propios
- Esta investigación revela patrones significativos por lo cual debería tomarse en cuenta como fuente importante de futuras investigaciones a nivel local y nacional además priorizar a los jóvenes como el presente del país y encaminar todos los recursos necesarios para su formación e inmersión en el mercado laboral con políticas de pasantías y de primer empleo con salarios dignos y horarios flexibles que permitan una profesionalización y un recurso humano capacitado y con la experiencia necesaria para afrontar la vida laboral. Se podría realizar programas de empleo juvenil como lo está realizando al momento Perú con una política Nacional de empleo decente formulada desde el 2021 y contemplada en la agenda del 2030 en el marco de la sustentabilidad en

todos los ámbitos posibles enmarcada en los objetivos de desarrollo sostenible. Además, se podría utilizar otros métodos de machine learning como por ejemplo el método de k medoides que es un método que se puede decir es más robusto e igual de confiable.

- La creación e implementación de las políticas públicas siempre debe estar basado en un reconocimiento científico del estado situacional de lo que atañe a diferentes grupos de interés, las herramientas estadísticas que sirven para la toma de decisiones debe estar precedido de un plan de acción donde se plane se haga se verifique y se actúe, es decir se cree , se lleve a cabo, se evalúa y se dé oportunidad a mejorar el bienestar de los jóvenes que pretenden un futuro mejor para sí mismos y su familia. Se deben crear planes que impacten en la manera de encontrar trabajo de los jóvenes habiendo una concordancia entre la oferta y demanda laboral, además incentivar a el sector privado para que las oportunidades que se dan al momento, para los jóvenes, puedan mejorar significativamente ampliando a du vez los beneficios sociales de este grupo y creando también espacios de igualdad.
- Se pueden utilizar otros índices para poder evaluar la validez del instrumento de investigación como puede ser el caso del índice de Davies - Bouldin que mide la similitud promedio de cada cluster y que además cumple con el mismo propósito del índice utilizado en la investigación. Además podría resultar de utilidad realizar un análisis longitudinal de estos grupos para revisar el impacto de las políticas en este segmento de la sociedad y revisar la eficiencia de estas , se podría realizar un análisis de series de tiempo o enfocar un análisis comparativo de un grupo reducido de jóvenes con un método no paramétrico para revisar muestras pareadas y demostrar que la situación en

un momento dado respecto a su empleo y desarrollo académico, permitió su avance en su calidad de vida.

Referencias

- Aguado, I., Echeberria, C., & Barrutia, J. (2019). Aplicación de un análisis clúster para el estudio de la segregación social en el municipio de Bilbao. *Boletín de la Asociación*, 81(2763), 1–35. doi:<http://dx.doi.org/10.21138/bage.2763>
- Albert, C., Suarez, P., Sanchez, R., & Toharia, L. (1998). *Las transformaciones de los jóvenes de la escuela al mercado de trabajo: un análisis de flujos*. Instituto Valenciano de Investigaciones Económicas, S.A.
- Appinio. (25 de junio de 2024). *Ocho tipos de sesgo en la investigación y cómo evitarlos*. Obtenido de <https://www.appinio.com/es/blog/investigacion-de-mercados/sesgo-investigacion#%C2%BFC%C3%B3mo%20evitar%20sesgos%20en%20la%20investigaci%C3%B3n?>
- Assusa, G. (2019). Ni jóvenes, ni desempleados, ni peligrosos, ni novedosos. Una crítica sociológica del concepto de “jóvenes nini” en torno los casos de España, México y Argentina. *Cuadernos de Relaciones Laborales*, 37(1), 91-111. doi:<http://dx.doi.org/10.5209/CRLA.63821>
- Avendaño, W., Luna, H., & Quintero, L. (2020). La política colombiana de emprendimiento en educación y su impacto en el acceso al empleo de jóvenes. *El Ágora USB*, 20(2), 158-171. doi:[10.21500/16578031.5137](https://doi.org/10.21500/16578031.5137)
- Ayala, P. (2011). Políticas Públicas sobre juventud en Ecuador desde 1980. *PROCESOS*, 2(34), 126-144. Obtenido de <https://revistas.uasb.edu.ec/index.php/procesos/article/view/1879/1669>

- Balarezo, G. (2019). Generación Nini , jóvenes que ni estudian ni trabajan. *Paidela xxi*, 9(1), 2271-7770. Obtenido de <https://repositorioacademico.upc.edu.pe/bitstream/handle/10757/626408/Generación%20Nini%20Jóvenes%20que%20ni%20Estudian%20ni%20Trabajan.pdf>
- Bancero, S. (12 de octubre de 2015). *Coeficiente de Silueta*. Obtenido de <https://www.labredes.unlu.edu.ar/sites/www.labredes.unlu.edu.ar/files/site/data/bdm/coeficiente-silueta.pdf>
- Banchero, S. (2015). *Repositorio Universidad Nacional de Lujan*. Obtenido de Repositorio Universidad Nacional de Lujan: <https://www.labredes.unlu.edu.ar/sites/www.labredes.unlu.edu.ar/files/site/data/bdm/coeficiente-silueta.pdf>
- Bermúdez, K. (2023). *Cuestiones actuales de la seguridad social*. In Universidad Externado de Colombia. Obtenido de <https://bdigital.uexternado.edu.co/server/api/core/bitstreams/2fafa6ef-134e-4796-be1b-a270e136d99a/content>
- Bernal, I., Pedraza, N., & Castillo, L. (2020). El capital humano y su relación con el desempeño organizacional. *Revista Espacios*, 41(22), 213-227. Obtenido de <https://revistaespacios.com/a20v41n22/a20v41n22p14.pdf>
- Bernal, J., Fuentes, C., Sion, N., & Tapia, E. (2022). Análisis de los principales indicadores del mercado laboral y las ramas de ocupación en el Ecuador. *Ciencia Latina*, 6(5), 903-924. doi:https://doi.org/10.37811/cl_rcm.v6i5.3151

- Bohorquez, E., Perez, M., & Caiche , W. (2020). La motivación y el desempeño laboral: el capital humano como factor clave en una organización. *Universidad y Sociedad*, 12(3), 385-390. Obtenido de <http://scielo.sld.cu/pdf/rus/v12n3/2218-3620-rus-12-03-385.pdf>
- Bouza, C. (2021). Medidas para análisis basados en las tablas de contingencia. *Research Gate*, 1-35. Obtenido de https://www.researchgate.net/publication/354343092_MEDIDAS_PARA_ANALISIS_BASADOS_EN_TABLAS_DE_CONTINGENCIA
- Burneo, A., & Yunga, D. (2020). Acceso de los jóvenes a la educación universitaria en el Ecuador: reformas, políticas y progreso. *Sisyphus*, 8(2), 70-85. doi:<https://doi.org/10.25749/sis.20259>
- Cañadas, R. (2010). *Las tablas de contingencia en la formación de profesionales de psicología*. [Tesis de Maestría, Universidad de Granada], Repositorio Universidad de Granada. Obtenido de <https://www.ugr.es/~batanero/pages/ARTICULOS/Gustavo2.pdf>
- Centro Nacional de Planeamiento Estratégico. (19 de agosto de 2016). *Los “NiNis” y los “SiSis”: una generación en riesgo*. Obtenido de <https://cdn.www.gob.pe/uploads/document/file/1057126/NINIs-y-SISIs-FINAL25-08-201620200728-16199-6uhogq.pdf>
- Cruz, R., Vargas , E., Hernandez, A., & Rodriguez, O. (2017). Adolescentes que estudian y trabajan: factores sociodemográficos y contextuales. *Revista Mexicana de Sociología*, 73(3), 571-604. Obtenido de <https://www.redalyc.org/pdf/321/32151631005.pdf>
- Cruz, R., Vargas, E., Hernandez, A., & Rodriguez, O. (2017). Adolescentes que estudian y trabajan: factores sociodemográficos y contextuales. *Revista Mexicana de Sociología* 79,

- 1(3), 572-604. Obtenido de <https://www.scielo.org.mx/pdf/rms/v79n3/0188-2503-rms-79-03-00571.pdf>
- Cruz, S., & Texeira, A. (2014). The Evolution of the Cluster Literature: Shedding Light on the Regional Studies–Regional Science Debate. *Regional Studies*, 44(9), 1263-1288. doi:<http://dx.doi.org/10.1080/00343400903234670>
- Datosmacro. (11 de junio de 2024). *Ecuador - Índice de Capital Humano*. Obtenido de <https://datosmacro.expansion.com/mercado-laboral/capital-humano/ecuador>
- De Domingo Soler, C., Naveda, M., Rodríguez, M., & Muñoz, M. (2020). Juventud, academia y empleo. Análisis de una desconexión. *PODIUM*, 1(37), 129-146. Obtenido de <http://scielo.senescyt.gob.ec/pdf/podium/n37/2588-0969-podium-37-129.pdf>
- Enriquez, A., & Galindo, M. (2015). Empleo. *Mexico como vamos*, 1(1), 1-9. Obtenido de https://scholar.harvard.edu/files/vrios/files/201508_mexicoemployment.pdf
- Epidat. (10 de 2014). *Análisis descriptivo*. Servizo Galego de Saúde. Obtenido de https://www.sergas.es/Saude-publica/Documents/1891/Ayuda_Epidat_4_Analisis_descriptivo_Octubre2014.pdf
- Firsova, I., Lukashenko, I., & Svetlana, A. (2020). *La importancia socioeconómica de la educación en una economía basada en el conocimiento*. Springer.
- Franco, A. (2022). EL CLÚSTER COMO MECANISMO DE POLÍTICA PÚBLICA PARA INCENTIVARLA COMPETITIVIDAD A PARTIR DEL PAPEL DE LOS SISTEMAS DE INNOVACIÓN. *Revista Inclusiones*, 9(es), 221-243. Obtenido de <https://revistainclusiones.com/carga/wp-content/uploads/2022/03/10-Alejandro-et-al-VOL-9-NUM-POL-PUB-MX-ABRILJUNIO2022INCL-1.pdf>

Fuente, S. (11 de julio de 2021). *Multivariante*. Obtenido de <https://www.fuenterrebollo.com/Economicas/ECONOMETRIA/MULTIVARIANTE/ACP/ACP.pdf>

Gallego, C., & Naranjo, C. (2020). El capital humano de la empresa: una propuesta de medición. *Ciencia y Tecnología*, 16(2), 70 - 89. doi:<https://doi.org/10.18041/1900-3803/entramado.2.6544>

Gomez, M. (2002). *El análisis cluster en investigación de marketing: metodología y crítica*. UAM. Obtenido de https://www.researchgate.net/publication/283070714_EL_ANALISIS_CLUSTER_EN_INVESTIGACION_DE_MARKETING_METODOLOGIA_Y_CRITICA

Gomez, M. (2002). *Research gate*. Obtenido de Research gate: Research gate

Guaita, A. (2022). *La incidencia del capital humano en el crecimiento económico del Ecuador en el período 2000-2018*. [Tesis de Grado, Universidad Técnica de Ambato], Repositorio Institucional UTA. Obtenido de <https://repositorio.uta.edu.ec/server/api/core/bitstreams/13fd34dd-2c7b-47d9-9eb9-53c4c7337f0d/content>

Hassan, D. (2017). A Clustering Approach for Evaluation of UserInteraction on Facebook Social Network. *International Journal of Computer & Organization Trends (IJCOT)*, 39(1), 4-8. doi:<http://dx.doi.org/10.14445/22492593/IJCOT-V39P302>

Jiménez, M. (2022). Transición a la vida adulta e inserción sociolaboral de jóvenes que han migrado de forma autónoma. *Mediterráneo Económico*, 1(36), 369-379. Obtenido de <https://dialnet.unirioja.es/servlet/articulo?codigo=8814872>

- Kapila, D., & Saraj, R. (2023). Cluster analysis of socio-economic factors and academic performance of school students. *Indonesian Journal of Electrical Engineering and Computer Science*, 31(3), 1568-1577. doi:<http://dx.doi.org/10.11591/ijeecs.v31.i3.pp1568-1577>
- Ledesma, R. (2008). *Researchgate*. Obtenido de Researchgate: https://www.researchgate.net/publication/49619436_Introduccion_al_Bootstrap_Desarrollo_de_un_ejemplo_acompanado_de_software_de_aplicacion#:~:text=El%20bootstrap%20es%20un%20tipo,la%20prueba%20de%20significaci%C3%B3n%20estadística.
- Lejarza, I. (2018). *Analisis clúster*. Universitat de València. Obtenido de <https://www.uv.es/mlejarza/actuariales/tam/cluster.pdf>
- León, A. (2007). Qué es la educación. *Educere*, 11(39), 595-604. Obtenido de <https://www.redalyc.org/pdf/356/35603903.pdf>
- López, J. (2021). Reflexiones sobre la situación del empleo juvenil y el fomento de su empleabilidad: mercado de trabajo y políticas de empleo en la Unión europea y España. *Estudios Latinoamericanos de Relaciones Laborales y Protección Social*, 1(12), 21-37.
- López, J., & Mondragón, J. (2019). *Jóvenes construyendo futuro: un medio para la inclusión laboral de los NINIS*. [Tesis de Pregrado, Universidad Católica de Colombia], Repositorio Institucional Universidad Católica de Colombia. Obtenido de <https://repository.ucatolica.edu.co/server/api/core/bitstreams/eb9120df-37ed-472b-bcf6-96e559c7147f/content>
- Martínez, A. (2023). *Mapa de Brecha de Evidencia a partir de la revisión sistemática de las estrategias para mejorar el empleo juvenil en América Latina y el Caribe*. [Tesis de

- Maestría, Universidad de los Andes], Repositorio Digital de la Universidad de los Andes. Obtenido de <https://repositorio.uniandes.edu.co/server/api/core/bitstreams/a47e697b-83f6-444b-8450-6388219f9922/content>
- Mideros, A., Sánchez, C., & Ramos, C. (2022). *Jóvenes con y sin título de educación superior en el mercado laboral: Un análisis descriptivo*. Pontificia Universidad Católica del Ecuador. Obtenido de <https://iie-puce.com/wp-content/uploads/2022/10/Jovenes-con-y-sin-titulo-de-educacion-superior-en-el-mercado-laboral.-Un-analisis-descriptivo-1.pdf>
- Ministerio de Trabajo. (2021). *Estudio de empleabilidad juvenil*. Dirección de Investigación y Estudios del Trabajo. Obtenido de https://www.trabajo.gob.ec/wp-content/uploads/2024/01/8.4.-mdt-diet-084-2021-estudio_empleabilidad_juvenil_vf-signed-signed.pdf
- Nabi, I. (11 de abril de 2021). *CONSTRUCCIÓN E INTERPRETACIÓN DEL COEFICIENTE SILUETA EN AGRUPAMIENTOS POR K-MEDIAS*. Obtenido de <https://marxistphilosophyofscience.com/wp-content/uploads/2022/04/construccion-e-interpretacion-del-coeficiente-silueta-en-agrupamientos-por-k-medias-isadore-nabi-1.pdf>
- Noussan, R. (2021). *Las etapas de evolución de un cluster y el ciclo de vida organizacional. Análisis de caso*. Secretaría de Internacionalización UNVM. Obtenido de <https://www.aacademica.org/xcongresodeadministraciondelcentrodela-republica/136.pdf>
- Olaguibe, J. (2021). La transición de los jóvenes hacia el trabajo decente: política social y empleo juvenil en España. *Estudios Latinoamericanos de Relaciones Laborales y Protección Social*, 1(12), 39-50. Obtenido de <https://dialnet.unirioja.es/servlet/articulo?codigo=8276486>

- Oliveira, A., Passos, C., & Ribeiro, C. (2019). El marketing interno como estrategia para motivar a los empleados: un estudio con PYMES de la Región de Lafões. *Gestión y desarrollo*, 1(27), 81-97. doi:<https://doi.org/10.7559/gestaoedesenvolvimento.2019.375>
- Olmedo, P. (16 de mayo de 2018). *El empleo en el Ecuador - Una mirada a la situación y perspectivas para el mercado laboral actual*. Obtenido de <https://library.fes.de/pdf-files/bueros/quito/14525.pdf>
- Organización Internacional del Trabajo. (12 de agosto de 2024). *El número de jóvenes que están sin empleo y que no están estudiando o en formación es motivo de preocupación, a pesar del descenso de la tasa de desocupación*. Obtenido de <https://www.ilo.org/es/node/666121>
- Pedraza, N. (2020). El clima y la satisfacción laboral del capital humano: factores diferenciados en organizaciones privadas y públicas. *INNOVAR*, 30(76), 9-24. doi:<https://orcid.org/0000-0001-9566-2880>
- Perez, A., Castillo, A., Barcelo, M., & León, J. (2009). Importancia de los clusters del conocimiento como estructura que favorece la gestión del conocimiento. *Intangible Capital*, 5(1), 33-64. Obtenido de <https://www.redalyc.org/pdf/549/54912872002.pdf>
- Perucha, C. (07 de 2022). *EL METODO DE K MEDIAS*. Obtenido de <https://uvadoc.uva.es/bitstream/handle/10324/58229/TFG-G5999.pdf?sequence=1>
- Pillajo, H., & Sierra, P. (2023). La eficacia de las políticas públicas del empleo juvenil en Ecuador. *Vínculos*, 8(1), 13-26. doi:<https://doi.org/10.24133/vinculospe.v8i1.3111>

- Quispe, G., Sanchez, L., & Arellano, O. (2019). Los costos del contrato juvenil en las empresas industriales del Ecuador. *Cumbres*, 5(1), 55-68. Obtenido de <https://dialnet.unirioja.es/servlet/articulo?codigo=7210416>
- RAE. (11 de agosto de 2023). *Sisi*. Obtenido de <https://www.fundeu.es/recomendacion/sisi-neologismo-valido/>
- Ramirez, K. (2022). *La generación SISI del Ecuador*. [Tesis de Grado, Escuela Politécnica Nacional], Repositorio Digital de la EPN. Obtenido de <https://bibdigital.epn.edu.ec/bitstream/15000/22852/1/CD%2012314.pdf>
- Rendón, M., Villasis, M., & Miranda, M. (2016). Estadística descriptiva. *ALERGIA MEXICO*, 397-407. Obtenido de <https://www.redalyc.org/pdf/4867/486755026009.pdf>
- Restrepo, L., Posada, S., & Noguera, R. (2012). Aplicación del análisis por componentes principales en la evaluación de tres variedades de pasto. *Revista colombiana de ciencias pecuarias*, 25(2), 258-266. Obtenido de <https://www.redalyc.org/pdf/2950/295023555011.pdf>
- Reyes, L., & Alvarado, C. (2020). *Investigación Documental*. Universidad Simón Bolívar. Obtenido de <https://bonga.unisimon.edu.co/server/api/core/bitstreams/2af35a4b-2abf-4f78-a550-0a4e4764e674/content>
- Rodríguez, E. (2019). La transición al primer empleo de los egresados de institutos técnicos superiores en la provincia de Córdoba, Argentina. *Estudios Demográficos y Urbanos*, 34(2), 427-442. doi:<http://dx.doi.org/10.24201/edu.v34i2.1823>

- Rodríguez, G. (2022). Método de algoritmo de clúster para el análisis del perfil de investigadores científicos. *e-Ciencias de la Información*, 12(2), 1-24. doi:<http://dx.doi.org/10.15517/eci.v12i2.50456>
- Rodriguez, S. (2019). *Transiciones en el mercado laboral en Ecuador período 2007-2016*. [Tesis de Grado, Pontificia Universidad Católica del Ecuador], Repositorio Nacional PUCE. Obtenido de <https://repositorio.puce.edu.ec/handle/123456789/34164>
- Ruiz, G. (2022). Estadística multivariable análisis de componenetes principales. *Varianza*, 1(5), 44-61. Obtenido de http://www.revistasbolivianas.ciencia.bo/pdf/rv/n15/n15_a07.pdf
- Salas, H. (20 de marzo de 2021). *Endomarketing: Una herramienta para la gestión efectiva del capital humano*. Obtenido de <http://scielo.iics.una.py/pdf/riics/v17n1/2226-4000-riics-17-01-126.pdf>
- Salgado, A., Velasco, C., Franco, D., Gonzalez, A., & Olmus, A. (2009). *Experiencias utilizando el behavioral cluster analysis en diferentes contextos*. Universidad de Buenos Aires. Obtenido de <https://www.academica.org/000-020/434.pdf>
- Sánchez, C. (2008). *Tema 5. Análisis de componentes principales*. Obtenido de http://eio.usc.es/eipc1/base/basemaster/formularios-php/materialesmaster/Mat_14_master0809multi-tema5.pdf
- Sánchez, M., Blas, H., & Tujague, P. (2010). El Análisis Descriptivo como recurso necesario en Ciencias Sociales y Humanas. *Fundamentos en Humanidades*, 101-114. Obtenido de <https://www.redalyc.org/pdf/184/18419812007.pdf>
- Sánchez, R. (2021). *Los jóvenes que no estudian ni trabajan (NINIS) en Colombia*. Centro de Investigaciones para el Desarrollo - CID. Obtenido de

<https://fce.unal.edu.co/media/files/CentroEditorial/documentos/documentosEE/documentos-economia-118.pdf>

Schiatti, L. (2017). *Estudio comparativo de diferentes algoritmos de clustering para la estimación de grupos de evaluados que comparten debilidades conceptuales similares*. [Tesis de Grado, Universidad Tecnológica de Bolívar], Repositorio Digital de la Universidad Tecnológica de Bolívar. Obtenido de <https://biblioteca.utb.edu.co/notas/tesis/0069812.pdf>

Shah, M. (2014). Evaluación y análisis de comunidades bacterianas de diferentes plantas de tratamiento de aguas residuales mediante electroforesis en gel de gradiente desnaturizante con ARNr 16s específico de grupo. *Revista internacional de biorremediación y biodegradación ambiental*, 2(3), 100-111. doi:10.12691/ijebb-2-3-2

Sofia, F., Parra, P., & Carbonel, J. (2024). ¿Estudias y trabajas?, perfiles del estudiantado. *Revista Española de Sociología*, 1578-2824. Obtenido de <file:///C:/Users/favioalejandro/Downloads/Dialnet-EstudiasYTrabajasPerfilesDelEstudiantadoUniversita-9429157.pdf>

Ubfal, D. (2022). *Facilitar la transición de la escuela al trabajo de las mujeres jóvenes*. Banco Mundial. Obtenido de <https://thedocs.worldbank.org/en/doc/5d1fe018df2aa5eb859c5491c1564eec-0350072023/original/Facilitar-la-transicion-de-la-escuela-al-trabajo-de-las-mujeres-jovenes.pdf>

- Universidad Carlos III de Madrid. (21 de mayo de 2021). *Análisis de Conglomerados (I): El procedimiento conglomerados de K medias*. Obtenido de <https://halweb.uc3m.es/esp/personal/personas/jmmarin/esp/guiaspss/21conglk.pdf>
- Universidad Carlos III de Madrid. (11 de abril de 2023). *Temas 3: Análisis de componentes principales*. Obtenido de <https://halweb.uc3m.es/esp/personal/personas/jmmarin/esp/amult/tema3am.pdf>
- Universidad de Granada. (11 de abril de 2022). *Capítulo 6: Introducción al Análisis Cluster. Consideraciones generales*. Obtenido de <https://www.ugr.es/~gallardo/pdf/cluster-g.pdf>
- Universitat de València. (11 de junio de 2024). *Agrupación o clustering*. Obtenido de <https://www.uv.es/mlejarza/datamine/T4.pdf>
- Vargas, Y., & Bracchi, C. (2019). Jóvenes en Colombia y Brasil en una sociedad individualista en busca de trabajo digno y seguridad social. *IUSTA*, 1(51), 115-129. doi:<https://doi.org/10.15332/25005286.5035>
- Vasquez, C., Fernández, E., & Montes, J. (1998). Los recursos intangibles como factores de la competitividad de la empresa. *Dirección y Organización*, 20(1), 20-98. doi:<https://doi.org/10.37610/dyo.v0i20.318>
- Velázquez, A. (2024). *Qué es Desk Research o Investigación de Escritorio*. Obtenido de <https://www.questionpro.com/blog/es/desk-research/>
- Zapotitla, J. (2024). *ANÁLISIS DE COMPONENTES PRINCIPALES*. Obtenido de <http://www.ptolomeo.unam.mx:8080/xmlui/bitstream/handle/132.248.52.100/139/A7.pdf?sequence=7&isAllowed=y>

ANEXO

```
library(haven)

BDDenemdu_personas_2023_anual <- read_sav("BDDenemdu_personas_2023_anual.sav")

View(BDDenemdu_personas_2023_anual)

install.packages("readxl")

library(readxl)

attach(BDDenemdu_personas_2023_anual)

install.packages("dplyr")

library(dplyr)

install.packages("gapminder")

library(gapminder)

glimpse(BDDenemdu_personas_2023_anual)

SISI<- BDDenemdu_personas_2023_anual %>% filter(p03>=18 & p03<=29 & p07==1 &
p20==1)

View(SISI)

attach(SISI)

#cruce de variables entre sexo y estado civil

table(p02,p06)

tablasexo_edad <-table(p02,p06)

prop.table(tablasexo_edad)
```

```
#cruce de variables entre edad sexo y estado civil

fable(xtabs(~ p03+p02+p06, data=SISI))

tablas_e_e<-fable(xtabs(~ p03+p02+p06, data=SISI))

prop.table(tablas_e_e)
```

```
#cruce de variables entre sexo y etnia

fable(xtabs(~ p02+p15, data=SISI))

tabla_sex_ni<-fable(xtabs(~ p02+p15, data=SISI))

prop.table(tabla_sex_ni)
```

```
# estadisticos edad

sum(p03)

median(p03)

min(p03)

max(p03)
```

```
#GRUPO DE OCUPACION

install.packages("summarytools")

library(summarytools)

tabocup<-freq(SISI$p42)
```

```
tabocup
```

```
#acp
```

```
BASEACP<-select(SISI,p03,p02,p05a,p06,p085,p15,p15aa,p24,p42,p45,p50,p75,prov)
```

```
BASEACP
```

```
View(BASEACP)
```

```
attach(BASEACP)
```

```
install.packages("FactoMineR")
```

```
library(FactoMineR)
```

```
install.packages("ggplot2")
```

```
library(ggplot2)
```

```
install.packages("factoextra")
```

```
library(factoextra)
```

```
install.packages("tidyverse")
```

```
library(tidyverse)
```

```
install.packages("cluster")
```

```
library(cluster)
```

```
install.packages("NbClust")
```

```
library(NbClust)
```

```
install.packages("naniar")
```

```
library(naniar)
```

```
install.packages("corrplot")
```

```
library(corrplot)
```

```
#correspondencia simple

tablaacp<-table(BASEACP$prov,p06)

tablaacp

ACPSISI<- CA(tablaacp, graph =TRUE)

print(ACPSISI)

summary(ACPSISI)

#correspondencia multiple

mult<- PCA(BASEACP, graph = TRUE)

print(mult)

summary(mult)

#kmedias

kmeans(BASEACP,centers = 3,iter.max = 10,nstart =1 )

fviz_nbclust(BASEACP, kmeans, method = "wss") + geom_vline(xintercept = 4, linetype =
2)

# Ejecutar k-means y almacenar el resultado en km

set.seed(123) # Fijar semilla para reproducibilidad

km <- kmeans(BASEACP, centers = 3, iter.max = 10, nstart = 1)
```

```
# Calcular distancias entre observaciones

dist_matrix <- dist(BASEACP)

# Calcular el índice de silhouette

sil <- silhouette(km$cluster, dist_matrix)

# Visualizar el índice de silhouette

fviz_silhouette(sil, palette = "jco")

library(gplots)

set.seed(123)

datos <- matrix(rnorm(100), nrow = 10) # Crea una matriz aleatoria

heatmap.2(x = datos, scale = "none", col = bluered(256),
          distfun = function(x) dist(x, method = "euclidean"),
          hclustfun = function(x) hclust(x, method = "average"),
          density.info = "none",
          trace = "none", cexRow = 0.7)

install.packages("fpc")

library(fpc)

# Se emplean los datos iris excluyendo la variable Species

datos <- scale(iris[, -5])
```

```
# K-means clustering con k = 3

set.seed(321)

km_clusters <- kmeans(x = dist(datos, method = "euclidean"), centers = 3,
                      nstart = 50)

# Cálculo de índices (se calculan un total de 34 índices y parámetros)

km_indices <- cluster.stats(d = dist(datos, method = "euclidean"),
                           clustering = km_clusters$cluster)

# Medidas de homogeneidad y separación

km_indices$average.within

## [1] 1.238354
```