



Máster Data Science, Big Data & Business Analytics

Curso académico 2023-2024

Trabajo fin de Master

**Hacia un turismo inteligente en el Valle del Cauca: Sistema de
recomendación personalizada basado en aprendizaje no
supervisado y redes neuronales.**

Autor:

Alejandra Otero Leyton

Resumen

El desarrollo de un turismo inteligente en el Valle del Cauca es esencial para mejorar la distribución equitativa de los flujos turísticos y el aprovechamiento integral de su riqueza natural y cultural. Este Trabajo de Fin de Máster propone un sistema de recomendación personalizada basado en técnicas de aprendizaje no supervisado y redes neuronales, el cual busca enriquecer la experiencia de los turistas al ofrecerles recomendaciones más precisas y diversificadas que les permitan descubrir una amplia gama de actividades y lugares de interés en la región.

Tabla de contenido

1. Introducción	3
2. Objetivo General.....	3
2.1. Objetivos Específicos.....	3
3. Antecedentes y contexto de la investigación	4
4. Marco Teórico	5
5. Metodología.....	6
6. Resultados.....	11
6.1. Preparación y procesamiento de los datos	11
6.2. Análisis descriptivo.....	11
6.3. Análisis de correspondencia simple	12
6.3.1. Análisis de correspondencia ente Rango de edad y Actividades turísticas ..	14
6.3.2. Análisis de correspondencia entre Genero y Actividades turísticas.....	15
6.3.3. Análisis de correspondencia entre Procedencia y Actividades turísticas.....	16
6.3.4. Análisis de correspondencia con variables suplementarias	17
6.4. Red Neuronal Convolutiva	18
6.5. Aplicativo para recomendación de actividades basadas en la edad.....	19
6.5.1. Implementación del sistema en tiempo real	20
7. Conclusiones.....	20
8. Recomendaciones.....	21
9. Bibliografía.....	22
10. Anexos	22

1. Introducción

El Valle del Cauca es un departamento colombiano conocido por su riqueza natural y atractivo turístico, este ofrece una amplia variedad de paisajes que incluyen playas, montañas, valles y ríos. Sin embargo, a pesar de su gran potencial, la región enfrenta desafíos debido a la falta de recursos y tecnología, lo que ha limitado la implementación de estrategias avanzadas de gestión y promoción turística. Esta situación ha llevado a que la experiencia turística se concentre en unas pocas áreas conocidas, sin aprovechar al máximo la diversidad de atractivos disponibles. Como resultado, muchas zonas con gran potencial permanecen desconocidas para los visitantes.

Ante esta realidad, se pretende dar el primer paso en la creación de un sistema de recomendación personalizada para los visitantes que lleguen al Valle del Cauca. De esta manera, se podrán considerar más lugares para realizar actividades, beneficiando a todos los integrantes de la cadena turística del sector.

Para lograr este objetivo, se llevó a cabo un análisis de correspondencia simple (AC) que permitió identificar relaciones significativas entre variables sociodemográficas y actividades turísticas. A partir de esta información, se seleccionó la variable más relevante para entrenar una red neuronal convolucional, y se recopiló y etiquetó un conjunto de imágenes de diversos tipos de personas para mejorar la precisión del sistema.

La implementación de este sistema no solo beneficia a los turistas al proporcionarles una experiencia más completa y enriquecedora, sino que también impulsa el desarrollo económico y social del Valle del Cauca.

2. Objetivo General

El propósito de este Trabajo de Fin de Máster es desarrollar un sistema de recomendación personalizada basado en técnicas de aprendizaje no supervisado y redes neuronales, con el objetivo de promover un turismo inteligente en el Valle del Cauca.

2.1. Objetivos Específicos

- a) Obtención y organización de datos de turismo receptivo del Valle del Cauca
 - Extraer y consolidar datos relevantes del Sistema de Información Turística.
 - Limpiar y reprocesar los datos para garantizar su calidad y consistencia.

- b) Análisis de Correspondencia Simple (ACS):
 - Aplicar un análisis de correspondencia simple entre cada variable sociodemográfica (Rango Edad, Género, Procedencia) y las actividades turísticas para identificar relaciones significativas.
 - Seleccionar la variable sociodemográfica más relevante y significativa identificada en el AC para el análisis posterior.

- c) Desarrollo de un Sistema de Recomendación Personalizado:
 - Recopilar y etiquetar imágenes de diversos tipos de personas
 - Entrenar una red neuronal convolucional utilizando la variable sociodemográfica seleccionada como etiqueta.
 - Integrar los resultados del análisis de correspondencia y la red neuronal en un sistema de recomendación.

3. Antecedentes y contexto de la investigación

El sector turístico del Valle del Cauca se ha desarrollado con la participación de diversas entidades, como las secretarías de turismo de los municipios y gremios como Cotelco Valle, que opera el Sistema de Información Turística del Valle del Cauca (SITUR Valle). Este sistema, creado en 2014, complementa al Centro de Información Turística Nacional (CITUR) y tiene como objetivo principal recopilar datos sobre el turismo en la región. SITUR Valle recolecta información de los prestadores de servicios turísticos y realiza mediciones del turismo receptivo, que son la base para este estudio. Aunque las entidades del sector público y privado utilizan los datos de SITUR para desarrollar estrategias de promoción turística, muchas veces no se aprovecha toda la información disponible. Este análisis busca identificar las preferencias y actividades de los turistas según características sociodemográficas, lo que permitirá a los actores del sector alinear sus esfuerzos y entender mejor lo que identifica al Valle del Cauca como destino turístico.

4. Marco Teórico

Turismo Inteligente

Gretzel y Jamal (2020) definen el turismo inteligente como un enfoque que integra tecnologías emergentes para ofrecer experiencias más adaptadas a las necesidades del turista, mejorando tanto la sostenibilidad como la eficiencia de los destinos turísticos. La creación de un destino inteligente debe basarse en las necesidades reales del destino, considerando diversos factores como su ecosistema económico e infraestructura. De esta manera, es posible elaborar proyectos que enriquezcan las experiencias de los visitantes y de los residentes.

Aprendizaje Automatizado / Machine Learning

El Aprendizaje Automatizado (*Machine Learning*; ML) es una rama de la inteligencia artificial, en gran parte inspirada en el razonamiento humano, que comprende el aprendizaje a partir de experiencia (Sammut y Webb 2011)(Sammut y Webb 2011). El aprendizaje automático aborda dos tipos de aprendizaje: el aprendizaje supervisado y el aprendizaje no supervisado.

Aprendizaje No supervisado

Para este trabajo de investigación nos centraremos en los algoritmos que hacen parte del aprendizaje no supervisado. Para tener un mejor contexto, aquí solo le damos las características al algoritmo, nunca las etiquetas. Queremos que nos agrupe los datos que le dimos según sus características. El algoritmo solo sabe que como los datos comparten ciertas características, de esa forma asume que pueda que pertenezcan al mismo grupo. Dentro de este tipo de aprendizaje nos encontramos con diferentes algoritmos, entre los principales se encuentran: el análisis de componentes principales (PCA), métodos de clustering, análisis de correspondencia simple (AC) y el análisis de correspondencia múltiple (ACM).

Análisis de Correspondencia Simple (AC)

El análisis de correspondencia (AC) es una técnica exploratoria multivariada, que permite analizar datos de un conjunto de variables cualitativas, dándonos la posibilidad de estudiar la asociación entre variables categóricas, permitiéndonos identificar con esto, agrupaciones de individuos. El análisis de correspondencia se divide de acuerdo a su grado de

complejidad, el análisis de correspondencia simple (AC) y análisis de correspondencia múltiple (ACM). El conjunto de datos de entrada en un Análisis de Correspondencias (AC) es una tabla de contingencia, que es una tabla de frecuencias para dos variables categóricas. Esta tabla se elabora cuando se observa una muestra en la que se registran dos o más características simultáneamente, cada una con varias categorías o clases. El objetivo del análisis de correspondencias es representar las filas y columnas de esta tabla, mostrando las relaciones entre ellas. Usualmente, los resultados del AC se presentan en un gráfico que muestra los perfiles de las filas y columnas como puntos en un plano cartesiano. Para construir estos gráficos, el análisis utiliza una descomposición singular de la matriz, que implica crear matrices diagonales para las filas y las columnas. Para la interpretación de los resultados del análisis es importante tener en cuenta los ejes significativos. Los ejes principales permiten analizar las relaciones entre los perfiles de fila y columna. En la mayoría de los casos prácticos, es suficiente con estudiar los dos primeros ejes principales, ya que estos representan un alto porcentaje de la variabilidad de los perfiles, conocida como inercia en el AC. Esto significa que los ejes principales capturan la mayor parte de la información de la tabla de contingencia, y la representación de los perfiles en dimensiones reducidas será más precisa cuanto mayor sea la inercia explicada por los ejes. (Greenacre, 2007).

Redes neuronales convolucionales (CNN)

Las redes neuronales convolucionales son un tipo de red neuronal diseñada para procesar datos con una estructura de grilla, como imágenes. Su arquitectura incluye capas de convolución que aplican filtros a los datos de entrada para extraer características esenciales, como bordes y texturas en el caso de imágenes. Estas redes son especialmente eficaces en tareas de clasificación y detección de patrones visuales gracias a su capacidad para capturar relaciones espaciales y reducir la dimensionalidad de los datos. Las CNN permiten construir modelos más precisos y eficientes, facilitando mejores predicciones.

5. Metodología

Para realizar este trabajo, utilizamos como fuente de datos la medición de turismo receptivo realizada por el sistema de información turística del Valle del Cauca. Esta medición se enfoca en caracterizar el comportamiento, la satisfacción y otros componentes de los

turistas que visitan la región. La recolección de esta información se lleva a cabo mediante encuestas presenciales en los diferentes puntos turísticos de los municipios del Valle del Cauca.

En la medición de turismo receptivo se cuenta con un total de 69 variables, sin embargo, para este análisis, solo se trabajará con las variables asociadas a las actividades que realizan los visitantes y las variables sociodemográficas. En total, se utilizarán 34 variables con información recolectada desde el año 2017 hasta abril de 2024. La mayoría de esta información se recoge mensualmente, aunque en algunos periodos se recolectó de manera trimestral.

Las principales variables sociodemográficas que se considerarán en el análisis son el rango de edad, la procedencia y el género. En cuanto a las actividades realizadas por los visitantes en el Valle del Cauca, se incluyen las siguientes: visita a centros comerciales; recorrer calles y parques del casco urbano; visita a museos, casas de la cultura, iglesias, santuarios y monumentos; haciendas y/o casas históricas; actividades religiosas; asistencia a espectáculos artísticos y festivales; avistamiento de aves; avistamiento de ballenas; visita a discotecas, bares, karaoke; visita a parques temáticos o parques de atracciones; compra de bienes de consumo duradero; realizar inversiones o reuniones de negocios; práctica de deportes; asistencia a competencias o muestras deportivas; asistencia a conferencias y congresos; asistencia a fiestas y ferias de pueblos y municipios; visita a fábricas o haciendas; visita a casinos y otros sitios de juegos de azar; visita a ríos, lagos y cascadas; miradores paisajísticos; zoológicos; proceso del café; proceso del azúcar dulce; actividades de aventura como rappel, parapente, bungee, escalada, espelelismo y canotaje; senderismo; reservas o parques; playas; asistir a bodas, primera comunión, 15 años, bautizos, entre otros; visita a familiares o amigos; y actividades gastronómicas.

Una vez obtenida la información, se procede con el procesamiento de los datos para asegurar su calidad y fiabilidad. Este procesamiento incluye la limpieza de los datos, el tratamiento de los valores faltantes y un análisis descriptivo inicial que permite observar el comportamiento preliminar de las variables. Posteriormente, se lleva a cabo el análisis de correspondencia (AC) para cada una de las variables sociodemográficas (como rango de edad, género y procedencia) en relación con las actividades turísticas, con el fin de identificar patrones y asociaciones en un plano bidimensional.

El análisis de correspondencia se inicia con la creación de las tablas de contingencia, que resumen las frecuencias de cada combinación de variables sociodemográficas y actividades turísticas. Estas tablas se descomponen mediante la función CA de la librería FactoMineR en R (se debe descargar la extensión de R en google colab), que permite realizar el AC de manera eficiente.

Técnicamente, el proceso incluye los siguientes pasos:

1. **Descomposición de la Matriz:** La función CA descompone la tabla de contingencia de forma similar a una descomposición en valores singulares (SVD). Esto produce una matriz de componentes principales donde los valores propios indican la cantidad de varianza explicada por cada dimensión.
2. **Generación de Matrices de Factores:** El proceso genera matrices de coordenadas para las filas (perfiles de las variables sociodemográficas) y las columnas (perfiles de las actividades turísticas). Estas coordenadas representan las posiciones de cada perfil en un espacio de menor dimensión (generalmente en dos dimensiones para facilitar la visualización).
3. **Valores Propios:** Los valores propios resultantes de la descomposición proporcionan una medida de la inercia o varianza explicada por cada dimensión. Estos valores permiten identificar los ejes principales que capturan la mayor parte de la variabilidad de los datos.
4. **Asignación de Coordenadas:** Se asignan coordenadas a cada perfil de fila y columna en el nuevo espacio dimensional, facilitando la visualización de las relaciones entre variables.
5. **Visualización en Dos Dimensiones:** Finalmente, se representa gráficamente las filas y columnas en un espacio bidimensional, lo que permite interpretar las asociaciones y diferencias entre variables de forma intuitiva. Los perfiles más cercanos en el gráfico indican asociaciones más fuertes, mientras que los perfiles más distantes sugieren relaciones menos significativas.

En el caso particular de las variables analizadas (rango de edad, género y procedencia), se observó que en algunas situaciones, una sola componente explicaba prácticamente toda la varianza, lo que imposibilitaba la representación de múltiples dimensiones. En estos casos, se optó por graficar las coordenadas de los perfiles en lugar de las componentes principales, con el fin de capturar mejor el comportamiento de los datos y sus asociaciones.

Una vez obtenido el análisis de correspondencia (AC) para cada variable, se seleccionará aquella que muestre la mejor relación con las actividades turísticas, tomando en cuenta la proximidad de las variables en el espacio bidimensional. Posteriormente, se calcularán las distancias euclidianas entre cada actividad y los perfiles de la variable seleccionada, para asignar cada actividad al grupo sociodemográfico más cercano. Con esta información, se construirá una tabla de asociación que servirá como base para entrenar un modelo de red neuronal, con el objetivo de predecir y clasificar actividades turísticas según las características sociodemográficas de los visitantes.

Una vez completado el análisis de correspondencia (AC) y tras la selección de la variable sociodemográfica más adecuada, se procedió a entrenar un modelo de predicción utilizando redes neuronales. Para este fin, se descargó el dataset **UTKFace** desde Kaggle, configurando las credenciales de la API de Kaggle mediante el archivo kaggle.json para facilitar la autenticación y descarga de los datos. El dataset se compone de imágenes faciales acompañadas de información sociodemográfica relevante, lo que lo hace adecuado para entrenar el modelo. Posteriormente, se organizó y exploró el contenido del dataset, asegurando que los datos estuvieran preparados para el preprocesamiento.

El siguiente paso consistió en cargar las imágenes y realizar un preprocesamiento de las mismas. Las imágenes se redimensionaron a un tamaño estándar de 128x128 píxeles y se normalizaron, dividiendo los valores de los píxeles por 255 para garantizar la uniformidad de las entradas al modelo. La información sociodemográfica correspondiente a cada imagen se extrajo a partir del nombre de los archivos y se almacenó junto con la imagen preprocesada, facilitando su manipulación y uso durante el entrenamiento.

Para garantizar la robustez del modelo, los datos se dividieron en conjuntos de entrenamiento y prueba, utilizando una proporción del 80% para el entrenamiento y el 20% para la prueba. El conjunto de entrenamiento se subdividió a su vez en un subconjunto de validación, asignando el 20% de los datos originales para monitorizar el rendimiento del modelo y prevenir el sobreajuste. Además, se aplicó una estrategia de aumentación de datos en el conjunto de entrenamiento, introduciendo transformaciones aleatorias como rotaciones, flips horizontales y zooms. Estas transformaciones aumentaron la diversidad de las imágenes, lo que permitió mejorar la capacidad del modelo para generalizar.

El modelo de red neuronal convolucional (CNN) diseñado para este proyecto consta de dos capas de convolución. La primera capa convolucional tiene 64 filtros con un kernel de 5x5, seguida de una capa de **MaxPooling** para reducir las dimensiones del mapa de características. La segunda capa convolucional tiene 128 filtros con un kernel de 3x3, también seguida de una capa de **MaxPooling**. Tras las capas convolucionales, el modelo pasa a una capa **Flatten**, que convierte los mapas de características en un vector de una dimensión, el cual es procesado por dos capas densas (de 256 y 128 neuronas respectivamente). Para prevenir el sobreajuste, se implementa una capa de **Dropout** con una tasa del 50%. Finalmente, una capa de salida densa con una sola neurona es usada para predecir el valor continuo de la edad.

El modelo fue compilado utilizando el optimizador **Adam** y la métrica de error absoluto medio (**MAE**), lo que permitió evaluar el rendimiento del modelo de manera efectiva. Además, se implementó **EarlyStopping** para detener el entrenamiento cuando el rendimiento en el conjunto de validación dejara de mejorar después de varias épocas consecutivas, asegurando así un entrenamiento eficiente.

Una vez entrenado el modelo, este se evaluó utilizando el conjunto de prueba para medir su capacidad predictiva. El modelo final fue guardado en un archivo **.h5** para su implementación futura.

Para la fase de implementación, se desarrolló una funcionalidad que permite capturar imágenes faciales mediante la cámara web del usuario. Estas imágenes se preprocesaron de manera similar a las imágenes utilizadas en el entrenamiento y se utilizaron para hacer predicciones en tiempo real. El valor predicho de la variable sociodemográfica se ajustó a una categoría predefinida, y, con base en esta clasificación, se recomendaron actividades específicas para cada grupo.

Finalmente, para la implementación del modelo en un entorno interactivo, se desarrolló una aplicación web utilizando **Streamlit** y **ngrok** para exponerla de manera accesible desde cualquier navegador. El modelo previamente entrenado fue cargado en la aplicación, y se diseñó una interfaz que permite al usuario subir una imagen. Dicha imagen es procesada utilizando OpenCV para ajustarse al tamaño adecuado y normalizar los valores de los píxeles. Una vez preprocesada, la imagen se pasa al modelo, que genera una predicción sobre la variable sociodemográfica seleccionada. A partir de esta predicción, se clasifica al

usuario en un rango predefinido y se le muestran recomendaciones personalizadas de actividades, basadas en la variable seleccionada. Para hacer accesible la aplicación en línea, se utilizó **ngrok** para crear un túnel hacia el puerto donde se ejecuta Streamlit, permitiendo acceder a la aplicación desde una URL pública. El sistema fue configurado para mantener el túnel activo hasta que se detenga manualmente el proceso.

6. Resultados

6.1. Preparación y procesamiento de los datos

Se cargó la base de datos "Datos Receptivo" en Google Colab, que contiene 48.810 registros recopilados entre el año 2027 y abril de 2024. De las 34 variables disponibles, se trabajará específicamente con Procedencia, Rango de edad, Género y 30 actividades turísticas. La variable Género incluye 2 categorías: Masculino y Femenino; el Rango de edad abarca 5 grupos etarios: 15-16, 27-38, 39-50, 51-62, y más de 63 años; y la variable Procedencia también incluye 2 categorías, esta se se clasifica en Nacional (Colombia) y Extranjero (otros países). Las actividades turísticas se registran como 0 (no realizada) o 1 (realizada), permitiendo que un turista participe en múltiples actividades. En cuanto a los tipos de datos, las variables sociodemográficas son de tipo categórico mientras que las actividades turísticas están mayormente registradas como numéricas. Posteriormente, se decidió convertir en categóricas todas las variables. En la limpieza de datos, se verificó que las variables sociodemográficas cumplen con los valores esperados, mientras que en las actividades turísticas se corrigieron categorías incorrectas asignando NaN a valores diferentes de 0 y 1. En el tratamiento de valores nulos, se asignó un valor de 0 a los registros vacíos en las actividades, interpretando la ausencia de participación en la actividad. Para las variables sociodemográficas, se eliminaron filas con valores nulos (365 en Género y 2.592 en Rango de edad) para evitar sesgos. Finalmente, se consolidó una base de datos con 46.109 registros para realizar los análisis posteriores.

6.2. Análisis descriptivo

Para realizar este análisis se utilizaron varios indicadores que nos ayudan a comprender la distribución y las características principales de los datos categóricos, entre estos se encuentra: *count* que representa el número total de registros no nulos en cada variable,

unique que es el número de valores únicos por cada variable, *top* que representa el valor más frecuente, *freq* que es la frecuencia del valor más común, y *% freq* que es la proporción que representa la frecuencia del valor más común sobre el total de registros. Ahora bien, en cuanto al género, se observa que la mayoría de los turistas son de género femenino, con 23,658 registros (51.3%). Respecto a las actividades turísticas, la mayoría de los turistas no participan en todas las actividades disponibles, siendo la opción "0" la más común en la mayoría de las categorías. Por ejemplo, actividades como la Visita a museos/casas de la cultura/iglesias, santuarios y monumentos tienen un alto porcentaje de no participación, con 34,671 registros (75.2%) indicando que los turistas no realizaron esta actividad. En términos de procedencia, la mayoría de los turistas provienen de Colombia, representando el 95.6% del total de registros. Finalmente, el grupo etario más representado es el de 27-38 años, que conforma el 39.9% del total.

Estos resultados nos muestran que, aunque existe una amplia oferta de actividades en la región, la participación es selectiva y predominan los turistas nacionales y jóvenes adultos. En la tabla1 se presentan los descriptivos de algunas variables seleccionadas; los demás descriptivos están disponibles en los anexos.

Tabla 1 Estadísticas descriptivas de algunas variables

Descriptivos	Genero	Procedencia	Rango edad	Visita a centros comerciales	Recorrer calles y parques del casco urbano
count	46109	46109	46109	46109	46109
unique	2	2	5	2	2
top	Femenino	Nacional	27-38	0	0
freq	23658	42905	18401	30248	24490
% freq	51%	93%	40%	66%	53%

6.3. Análisis de correspondencia simple

Para realizar el Análisis de Correspondencias Simples (AC), se procedió a trabajar con cada una de las variables sociodemográficas (Rango de edad, Género y Procedencia) de manera independiente. Este análisis tiene como objetivo visualizar las relaciones entre las actividades turísticas y cada una de estas variables. Antes de realizar el AC, todas las actividades turísticas fueron renombradas para facilitar la representación gráfica. Se enumeraron de y1 a y31, según la siguiente tabla:

Tabla 2. Renombramiento Actividades turísticas

Actividad	Renombre
Visita a centros comerciales	y1
Recorrer calles y parques del casco urbano	y2
Visita a museos / casas de la cultura / iglesias, santuarios y monumentos	y3
Haciendas y/o casas históricas	y4
Actividades religiosas	y5
Asistencia a espectáculos artísticos y festivales	y6
Avistamiento de aves	y7
Avistamiento de ballenas	y8
Visita a discotecas; bares; karaoke	y9
Visita a parques temáticos/parques de atracciones	y10
Compra de bienes de consumo duradero	y11
Realizar inversión/reunión de negocio	y12
Práctica de deportes	y13
Asistencia a competencias o muestras deportivas	y14
Asistencia a conferencias/congresos	y15
Asistencia a fiestas y ferias de pueblos municipios	y16
Visita a fábricas/haciendas	y17
Visita a casinos y otros sitios de juegos de azar	y18
Visita a ríos, lagos y cascadas	y19
Miradores paisajísticos	y20
Zoológicos	y21
Proceso del Café	y22
Proceso del Azúcar Dulce	y23
Actividades de aventura (Rappel, parapente, bungee, escalada, espelelismo, canotaje)	y24
Senderismo (caminatas)	y25
Reservas/parques	y26
Playas	y27
Otro	y28
Asistir a Boda/Primera comunión/15 años/ bautizo entre otros	y29
Visita a familiares/Amigos	y30
Actividades gastronómicas	y31

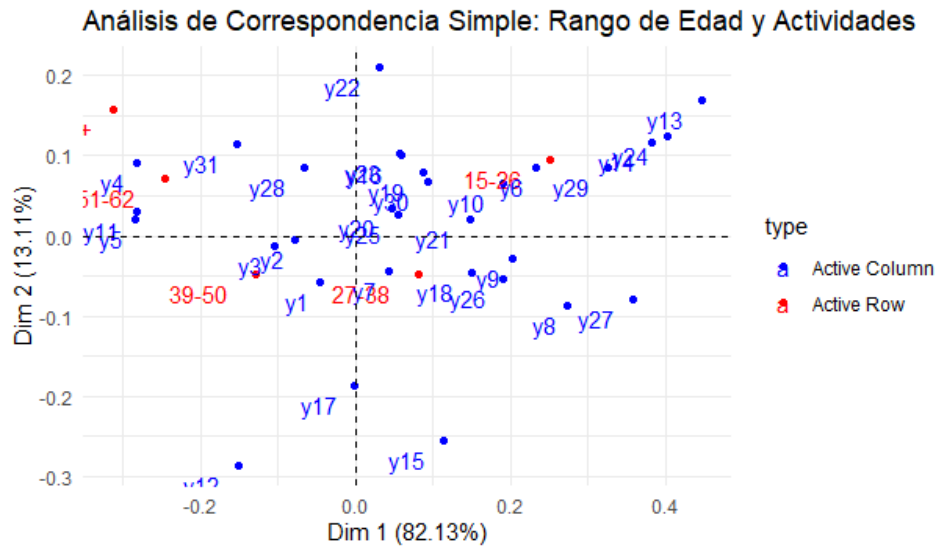
El AC se realizó primero para la variable Rango de edad. El proceso incluyó la creación de una tabla de contingencia que cruza esta variable con cada una de las actividades turísticas. Luego, se aplicó el análisis de correspondencias utilizando la librería FactoMineR en R. Los gráficos resultantes se centraron en visualizar la relación entre las dimensiones principales del análisis y cómo las diferentes categorías de Rango de edad se asocian con las actividades enumeradas. El mismo proceso fue repetido para las otras variables sociodemográficas, Género y Procedencia. En cada caso, se generaron gráficos

que representan la relación entre las categorías de la variable sociodemográfica en cuestión y las actividades turísticas.

6.3.1. Análisis de correspondencia ente Rango de edad y Actividades turísticas

En la figura 1 se puede observar cómo las preferencias de actividades turísticas en el Valle del Cauca están influenciadas por el rango de edad de los visitantes, de acuerdo a el grafico se puede ver que los turistas más jóvenes, entre 15 y 26 años, tienden a buscar experiencias dinámicas y entretenidas, con una fuerte inclinación hacia la asistencia a espectáculos artísticos y festivales, visitas a parques temáticos y participación en actividades de aventura como el parapente, bungee y canotaje, lo que sugiere una búsqueda constante de emociones y diversión durante su estancia en la región. Por otro lado, los turistas en el rango de 27 a 38 años parecen estar más interesados en actividades relacionadas con la naturaleza y el ecoturismo, mostrando afinidad por el senderismo, la visita a miradores paisajísticos y zoológicos, lo que indica un deseo de explorar el entorno natural y conectar con el paisaje local. En cuanto a los turistas de mediana edad, específicamente aquellos entre 39 y 50 años, se observa una preferencia por actividades más culturales y de exploración urbana, como visitar museos, casas de la cultura, iglesias, monumentos, así como recorrer calles y parques del casco urbano, lo que sugiere que valoran profundamente la riqueza cultural e histórica del destino. Además, los turistas en el rango de 51 a 62 años se inclinan por actividades más tradicionales y relacionadas con la historia y la religión, tales como la visita a haciendas históricas y la participación en actividades religiosas, reflejando un interés por conectar con las tradiciones y la historia de la región. Finalmente, los turistas de 63 años o más muestran una marcada preferencia por actividades de ocio, especialmente la visita a centros comerciales, lo que indica que buscan experiencias más relajadas y accesibles, centradas en el confort y la conveniencia durante su estancia.

Figura 1. ACS Rango de edad y actividades turísticas



6.3.2. Análisis de correspondencia entre Género y Actividades turísticas

El Análisis de Correspondencias Simples (AC) reveló que toda la variabilidad en la relación entre género y actividades se explica en una única dimensión como se puede ver en la Tabla 3, esto puede estar sugiriendo que existe una estructura simple en la asociación entre estas variables. Aunque esto limita la visualización en un gráfico bidimensional, las coordenadas en la dimensión única aún permiten interpretar cómo se relacionan las actividades con los géneros como se puede ver en la Figura 2.

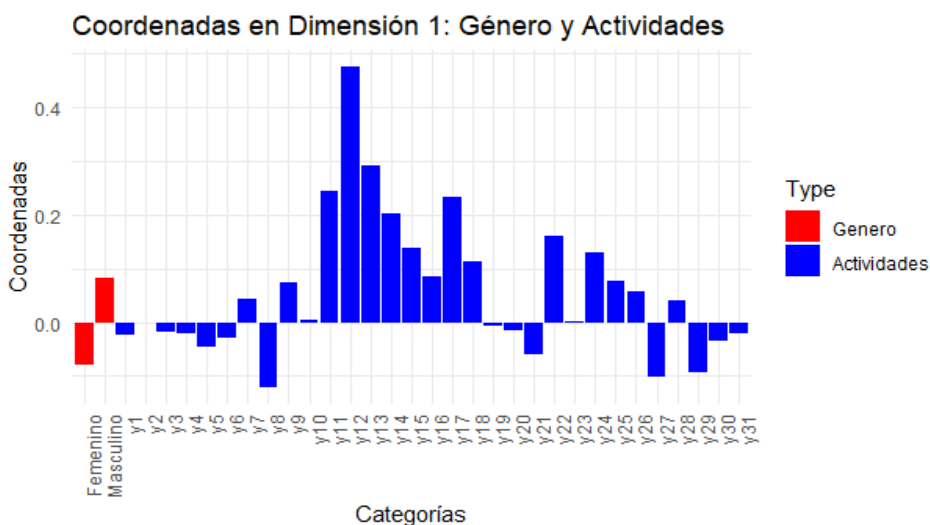
Tabla 3. Varianza Explicada Genero y Actividades

Dimensiones	Valor propio	Varianza explicada	Varianza explicada Acumulada
Dim 1	0.006440774	100%	100%

En la figura 2 se muestra que tanto el género femenino como el masculino tienen coordenadas muy cercanas a cero, con el género femenino ligeramente negativo y el masculino ligeramente positivo. Esta cercanía al origen sugiere que el género en sí mismo no está fuertemente asociado con las actividades turísticas representadas en esta dimensión, lo que significa que las diferencias en participación entre géneros son mínimas en este contexto. Aunque algunas actividades tienen coordenadas más extremas en la Dimensión 1, como Realizar inversión/reunión de negocio (y12), la práctica de deportes (y13), y la compra de bienes de consumo duradero (y11), que están más asociadas con el lado positivo, y otras como el avistamiento de ballenas (y8) o la visita a playas (y27) que se

encuentran en el extremo negativo, estas asociaciones no parecen estar fuertemente influenciadas por el género.

Figura 2. Coordenadas de Género y Actividades



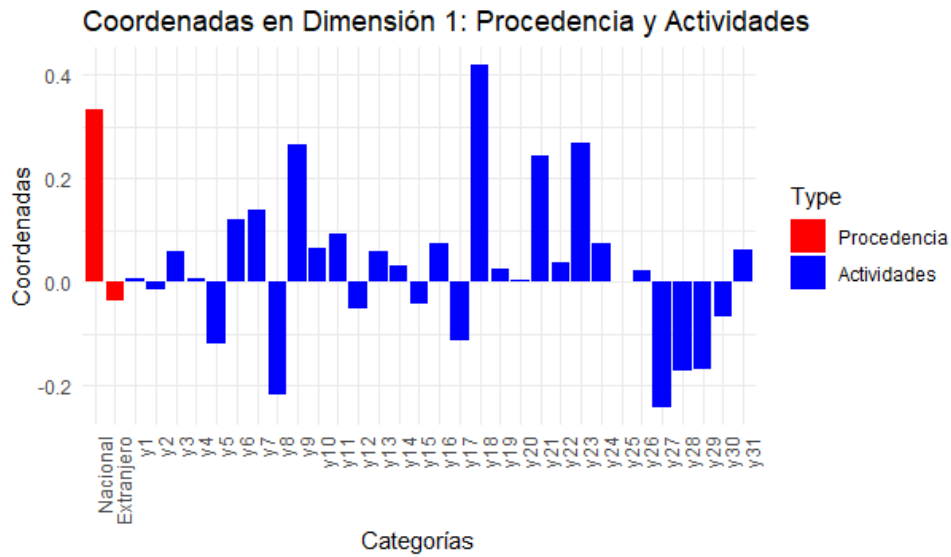
6.3.3. Análisis de correspondencia entre Procedencia y Actividades turísticas

El Análisis de Correspondencias Simples (AC) entre procedencia y actividades se explica en una única dimensión, como se observa en la Tabla 4, lo que sugiere una asociación relativamente simple entre estas variables. En la Figura 3, se observa que los turistas extranjeros tienen una coordenada muy cercana a cero, lo que sugiere que no están fuertemente asociados con ninguna actividad en particular en la Dimensión 1. Esto implica que, en general, las actividades no varían significativamente en función de si el turista es extranjero o no. Por otro lado, los turistas nacionales tienen una coordenada más positiva, aunque no tan extrema, pero puede indicar una mayor asociación con ciertas actividades, específicamente, las actividades y18 (Visita a casinos y otros sitios de juegos de azar) y y24 (Actividades de aventura como rappel, parapente, bungee, etc.)

Tabla 4. Varianza explicada Procedencia y actividades

Dimensiones	Valor propio	Varianza explicada	Varianza explicada Acumulada
Dim 1	0.011	100%	100%

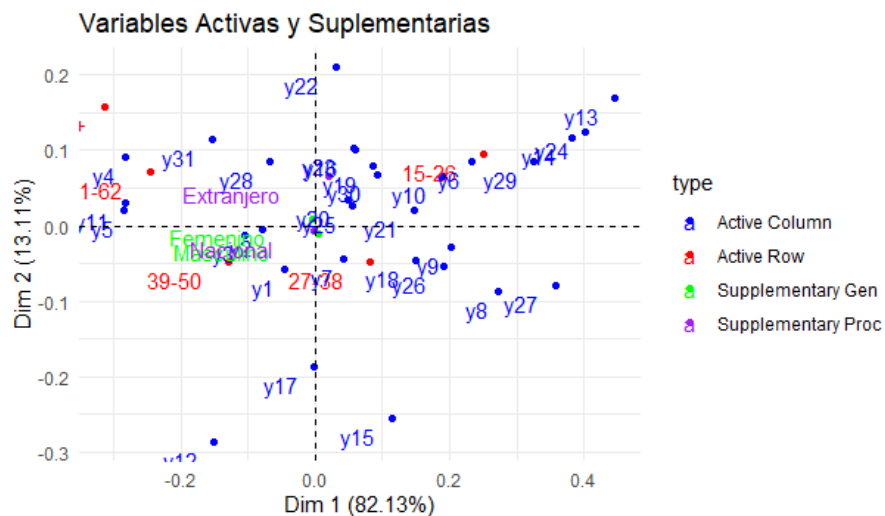
Figura 3. Coordenadas de Procedencia y Actividades



6.3.4. Análisis de correspondencia con variables suplementarias

A partir de análisis de correspondencias individuales, se identificó el rango de edad como la variable más informativa y significativa para comprender las actividades turísticas realizadas por los visitantes. Para complementar el análisis, se incorporaron las variables suplementarias de género, procedencia en función del rango de edad como variable activa, mostrando cómo estas dimensiones adicionales se alinean en el espacio de actividades y subrayando la importancia del rango de edad en la predicción de comportamientos turísticos como se ve en la figura 4.

Figura 4. Análisis de correspondencia variables activas y suplementarias



Con lo anterior, la variable rango de edad proporciona una base sólida para entrenar un modelo de red neuronal, ya que captura mejor la variabilidad en el comportamiento de los turistas con respecto a las actividades que realizan en la región. Para seleccionar las actividades más relevantes para cada rango de edad, se calculó la distancia entre las actividades y los diferentes rangos de edad en el espacio generado por el Análisis de Correspondencias Simple. Cada actividad fue luego asignada al rango de edad más cercano, permitiendo identificar las actividades que son más representativas para cada grupo etario. En la tabla 5 se puede observar cómo se distribuyen estas actividades con respecto a la edad, ya que no se le asignó ninguna actividad a los mayores de 63 años, se les recomendará las mismas actividades que a los turistas que tienen entre 50-62 años.

Tabla 5. Asignación de actividades por rango de edad

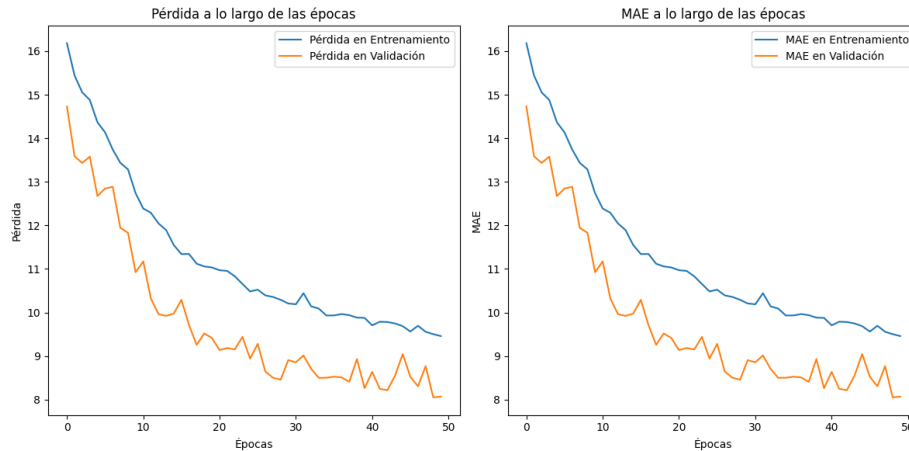
Rango de Edad	Actividades
15-26	y6, y8, y10, y13, y14, y22, y24, y27, y29
27-38	y7, y9, y15, y16, y17, y18, y19, y20, y21, y23, y25, y26, y30
39-50	y1, y2, y3, y12, y28
51-62 y más de 63	y4, y5, y11, y31

6.4. Red Neuronal Convolutiva

Se subió el archivo kaggle.json para autenticar el acceso al dataset UTKFace en Kaggle. Tras descargar el archivo comprimido y descomprimirlo, se trabajó con la carpeta *UTKFace*, que contiene un total de 23,708 imágenes faciales. Estas imágenes fueron utilizadas para entrenar y evaluar el modelo de predicción de edad utilizando redes neuronales convolucionales.

Durante las primeras épocas, se observó una mejora constante en la métrica de error absoluto medio (MAE) tanto para el conjunto de entrenamiento como para el de validación. En las primeras diez épocas, el MAE en el conjunto de validación se redujo rápidamente de un valor inicial de 14.73 hasta alcanzar aproximadamente 10.92. Esta mejora progresiva continuó de forma más gradual hasta la época 50, donde el modelo logró un MAE de 8.06 para el conjunto de validación. La pérdida (loss) también mostró un descenso significativo a lo largo de las épocas, disminuyendo de 14.7 a aproximadamente 8.06 para la validación. (Figura 5)

Figura 5. Indicadores de rendimiento a lo largo de las épocas



El uso de la técnica de EarlyStopping fue crucial para prevenir el sobreajuste del modelo, asegurando que el entrenamiento se detuviera cuando no se observaron mejoras relevantes en el conjunto de validación. A lo largo de las 50 épocas, se aplicaron técnicas de aumentación de datos, como rotaciones, desplazamientos y zooms, lo que permitió mejorar la capacidad del modelo para generalizar mejor frente a nuevos datos.

Al evaluar el rendimiento final en el conjunto de prueba, el modelo alcanzó un MAE de 7.84, por lo que se puede predecir la edad con una precisión razonable, con un error promedio de menos de 8 años.

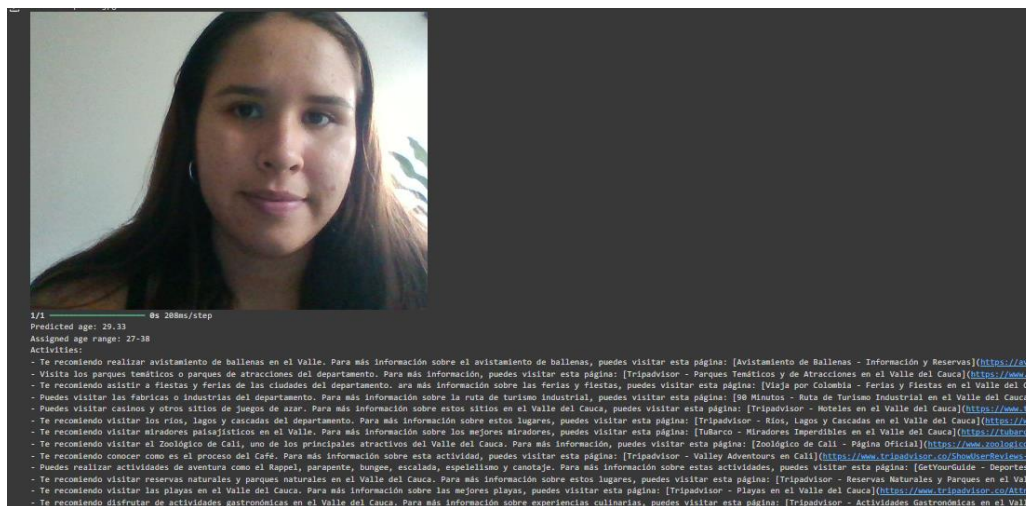
6.5. Aplicativo para recomendación de actividades basadas en la edad

Al implementar el aplicativo, se logró que el sistema capture una imagen del usuario a través de su cámara web y, utilizando el modelo de red neuronal se predijo la edad de la persona en tiempo real. A partir de la edad predicha, el sistema asigna al usuario a un rango de edad específico y ofrece una lista personalizada de actividades turísticas para realizar en el Valle del Cauca.

Al realizar la prueba con mi foto, el sistema predijo una edad de **29.33 años**, lo que clasificó al usuario en el rango de **27-38 años**, por lo que me clasifiqué en un rango de edad adecuado ya que tengo 28 años. Este grupo etario tiene preferencia por actividades relacionadas con la naturaleza, el turismo de aventura y eventos culturales. Las actividades recomendadas incluyeron visitas a parques temáticos, avistamiento de ballenas, participación en ferias y fiestas locales, entre otras. Además, cada actividad recomendada vino acompañada de un

enlace que lleva a sitios web relevantes donde el usuario puede obtener más información detallada. (Figura 6)

Figura 6. Prueba del Aplicativo para recomendaciones personalizadas



6.5.1. Implementación del sistema en tiempo real

Para facilitar el uso del modelo de predicción de edad en un entorno práctico, se desarrolló una aplicación web utilizando Streamlit. Los usuarios pueden generar su propio enlace para acceder a la aplicación mediante la ejecución del código provisto. Al correr el código en su entorno, se les proporcionará un enlace único generado con la herramienta Ngrok. Para generar este enlace y realizar la prueba, es necesario correr el código que integra Streamlit y Ngrok, lo que permitirá subir una foto y obtener las recomendaciones personalizadas.

Una vez que el usuario accede a este enlace, puede subir una fotografía que será procesada por el modelo para predecir su edad y así mismo se le recomendará las actividades que puede realizar en la región.

7. Conclusiones

Predicción de la edad y las recomendaciones personalizadas: El modelo de red neuronal convolucional desarrollado para predecir la edad de los turistas a partir de imágenes faciales mostró resultados razonablemente precisos, con un margen de error promedio inferior a 8 años. A través de la integración de este modelo con un sistema de

recomendaciones personalizadas, se logró ofrecer actividades turísticas específicas basadas en el rango de edad del usuario, lo que proporciona una experiencia más personalizada para los visitantes del Valle del Cauca.

Relación entre Rango de Edad y Preferencias Turísticas: El análisis de correspondencia reveló que las actividades turísticas están claramente segmentadas por rango de edad. Los turistas jóvenes (15-26 años) prefieren actividades dinámicas y de aventura, mientras que los turistas de más de 63 años se inclinan por actividades más relajadas como visitas a centros comerciales. Esta segmentación ofrece una oportunidad para ajustar las estrategias de promoción turística de acuerdo a los intereses particulares de cada grupo etario.

Optimización del Modelo: Aunque el modelo de predicción de edad funcionó de manera aceptable, se observó que requiere múltiples intentos para predecir una edad cercana a la real. Esto sugiere que el modelo aún puede beneficiarse de más ajustes en la arquitectura o en el entrenamiento para mejorar su precisión.

8. Recomendaciones

Mejorar la Precisión del Modelo: Dado que el modelo de predicción de edad requiere varios intentos para generar una predicción precisa, es recomendable realizar más iteraciones de ajuste de hiperparámetros, aumentar el conjunto de datos de entrenamiento, o explorar arquitecturas más avanzadas para mejorar la precisión del modelo.

Expansión de la Base de Datos de Actividades: Sería beneficioso integrar más información detallada sobre las actividades turísticas y los lugares específicos donde se pueden realizar, como restaurantes, parques o puntos turísticos clave del Valle del Cauca. Esto no solo mejoraría la calidad de las recomendaciones, sino que también fomentaría una mayor diversidad de opciones para los turistas.

Desarrollo de una Aplicación Móvil: Se recomienda desarrollar una versión móvil del aplicativo que no solo utilice la predicción de edad, sino que también permita al usuario ingresar sus preferencias, intereses y motivaciones, lo que permitirá generar recomendaciones más personalizadas. Esta aplicación podría incluir funciones de retroalimentación para que los usuarios puedan calificar actividades, lo que contribuiría a mejorar continuamente las recomendaciones basadas en datos reales.

Promoción y Uso de Datos en Tiempo Real: Es importante que la información obtenida de los turistas y sus preferencias se utilice para generar estrategias de marketing en tiempo real, ofreciendo ofertas y promociones que se alineen con los intereses de cada grupo demográfico, fomentando así un turismo más sostenible y segmentado en la región.

9. Bibliografía

Gretzel, U., & Jamal, T. (2020). *Turismo inteligente y destinos inteligentes: Una integración de tecnologías emergente*.

Sammut, C., & Webb, G. I. (Eds.). (2011). *Encyclopedia of Machine Learning*. Springer.

Greenacre, M. (2007). *Correspondence Analysis in Practice*. CRC Press.

10. Anexos

Tabla descriptiva con todas las variables

Descriptivos	Genero	Procedencia	Rango edad	Visita a centros comerciales	Recorrer calles y parques del casco urbano	Visita a museos / casas de la cultura / iglesias, santuarios y monumentos.	Haciendas y/o casas históricas
count	46109	46109	46109	46109	46109	46109	46109
unique	2	2	5	2	2	2	2
top	Femenino	Nacional	27-38	0	0	0	0
freq	23658	42905	18401	30248	24490	34671	43151
% freq	51%	93%	40%	66%	53%	75%	94%

Descriptivos	Asistencia a espectáculos artísticos y festivales	Avistamiento de Aves	Avistamiento de Ballenas	Visita a discotecas; bares; karaoke	Visita a parque temáticos/parque de atracciones	Compra de bienes de consumo duradero	Realizar inversión/reunión de negocio
count	46109	46109	46109	46109	46109	46109	46109
unique	2	2	2	2	2	2	2
top	0	0	0	0	0	0	0
freq	43777	45226	43775	41818	44531	45774	44695
% freq	95%	98%	95%	91%	97%	99%	97%

LinkCodigo: https://colab.research.google.com/drive/134VRTbHe85GXN94-kGm5rFpLcxEA8exY?usp=drive_link

```
# Tratamiento de datos
# =====

import numpy as np
import pandas as pd
from sklearn.datasets import make_blobs

# Gráficos
# =====

import matplotlib.pyplot as plt
from matplotlib import style
import seaborn as sns # Adicional para gráficos más complejos
style.use('ggplot') or plt.style.use('ggplot')
plt.rcParams["figure.figsize"] = (12,10)

# Preprocesado y modelado
# =====

from sklearn.cluster import KMeans
from sklearn.preprocessing import scale
from sklearn.metrics import silhouette_score, silhouette_samples
!pip install prince
import prince # Análisis de correspondencias (AC y ACM)

# Red neuronal y procesamiento de imágenes
# =====

from google.colab import files
import os
import zipfile
import cv2

from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Conv2D, MaxPooling2D, Flatten, Dense, Dropout
from tensorflow.keras.preprocessing.image import ImageDataGenerator
from sklearn.model_selection import train_test_split
from tensorflow.keras.callbacks import EarlyStopping

# Kaggle para descargar datasets
!pip install kaggle

# Aplicativo (Streamlit y Ngrok)
# =====

!pip install streamlit
```

```

!pip install pyngrok
from pyngrok import ngrok

# Funciones de visualización en Colab
# =====

from IPython.display import display, Image, Javascript
from google.colab.output import eval_js
from base64 import b64decode

# Configuración warnings
# =====

import warnings
warnings.filterwarnings('ignore')

# Lectura de datos
from google.colab import files
uploaded = files.upload()

#Cargue de datos
df = pd.read_excel('Datos Receptivo.xlsx')

#Visualizacion de la base
pd.option_context('display.max_columns', None)
df.head()

# Informacion de las variables
df.info()

#Valores unicos de las variables
df.nunique()

# Lista de columnas con menos de 13 valores distintos. Potenciales factores
to_factor = list(df.loc[:,df.nunique() <= 13]);

# Podemos cambiar el tipo de todas ellas a factor de una vez
df[to_factor] = df[to_factor].astype('category')

# Informacion de las variables despues de transofrmadas
df.info()

#Limpieza de datos
actividades_columns = df.columns[1:31]
for col in actividades_columns:
    print(f"Activity: {col}")
    print(f"Unique values: {df[col].unique()}")
for col in df.columns[1:31]:
    unique_values = df[col].unique()

    # Verificar si la columna tiene más de 2 categorías
    if len(unique_values) > 3:
        # Reemplazar los valores que no son 0 o 1 con NaN
        df[col] = df[col].apply(lambda x: x if x in [0, 1, np.nan] else np.nan)

#Tratamiento de valores nulos

```

```

df.isnull().sum()
#Imputacion de datos

#Actividades que realizan los turistas en el valle, todos los vacios imputar con 0
actividades = df.columns[1:32]
df.loc[:,actividades]=df.loc[:,actividades].fillna(0)

# Variables objetivos
columnas a imputar = ['Rango edad', 'Genero']

# Eliminar las filas que tengan valores NaN en las columnas especificadas
df = df.dropna(subset=columnas_a_imputar)
#Evidencia de que no queden valores nulos
df.isnull().sum()

# Configurar pandas para mostrar todas las columnas
pd.set_option('display.max_columns', None)

# Obtenemos el análisis descriptivo general de todas las columnas categóricas
descriptive table = df.describe(include='category')

# Mostrar la tabla descriptiva
descriptive_table
from google.colab import drive
drive.mount('/content/drive')
#Exportar la base para cargarla en un espacio de R
df.to_excel('base_cluster.xlsx')
# Instalar R en Google Colab
!apt-get install r-base -y

# Instalar rpy2 para permitir la ejecución de código R en Google Colab
!pip install rpy2
# Cargar la extensión R en Google Colab
%load_ext rpy2.ipython

%%R

# Instalar y cargar paquetes necesarios
install.packages("reshape2")
install.packages("FactoMineR")
install.packages("ggplot2")
install.packages("readxl")
library(readxl)
library(reshape2)
library(FactoMineR)
library(ggplot2)

```

```

base_cluster <- read_excel('/content/base_cluster.xlsx')
# Convertir la base de datos a formato largo
actividades <- setdiff(names(base_cluster), c("Rango edad", "Genero", "Procedencia"))
base_long <- melt(base_cluster, id.vars = "Rango edad", measure.vars = actividades)
base_long <- base_long[base_long$value == 1, ]

# Crear la tabla de contingencia
combined_table <- table(base_long`Rango edad`, base_long$variable)
colnames(combined_table) <- paste0("y", 1:ncol(combined_table))

# Mostrar la tabla de contingencia
print(combined_table)

# Realizar el análisis de correspondencias
ca_result <- CA(combined_table, graph = FALSE)

# Ver los resultados
summary(ca_result)

# Gráfico de variables activas
plot(ca_result, invisible = "row")

# Extraer coordenadas de las variables activas
active_rows <- ca_result$row$coord
active_cols <- ca_result$col$coord

# Agregar etiquetas
active_rows <- data.frame(active_rows, label = rownames(active_rows), type = "Active Row")
active_cols <- data.frame(active_cols, label = rownames(active_cols), type = "Active
Column")

# Unir los data frames
df_combined <- rbind(active_rows, active_cols)

# Gráfico combinado
ggplot(df_combined, aes(x = Dim.1, y = Dim.2, color = type, label = label)) +
  geom_point() +
  geom_text(vjust = 1.5, hjust = 1.5) +
  theme_minimal() +
  ggtitle("Análisis de Correspondencia Simple: Rango de Edad y Actividades") +
  xlab(paste0("Dim 1 (", round(ca_result$eig[1, 2], 2), "%)")) +
  ylab(paste0("Dim 2 (", round(ca_result$eig[2, 2], 2), "%)")) +
  geom_hline(yintercept = 0, linetype = "dashed") +
  geom_vline(xintercept = 0, linetype = "dashed") +

```

```

    scale_color_manual(values = c("blue", "red"))
%%R
# Calcular distancias entre cada actividad y cada rango de edad
distances <- as.matrix(dist(df_combined[, c("Dim.1", "Dim.2")]))

# Filtrar las filas correspondientes a los rangos de edad
age_rows <- rownames(df_combined[df_combined$type == "Active Row", ])

# Filtrar las columnas correspondientes a las actividades
activity_cols <- rownames(df_combined[df_combined$type == "Active Column", ])

# Crear un dataframe para almacenar la asociación de actividades con rangos de edad
activity_age_association <- data.frame(Activity = character(), Rango Edad = character())

# Asignar cada actividad al rango de edad más cercano
for (activity in activity_cols) {
  closest_age <- age_rows[which.min(distances[activity, age_rows])]
  activity_age_association <- rbind(activity_age_association, data.frame(Activity =
activity, Rango Edad = closest_age))
}

# Mostrar las asociaciones
print(activity_age_association)

# Exportar la tabla `activity_age_association` a CSV
write.csv(activity_age_association, "/content/activity_age_association.csv", row.names =
TRUE)
%%R

actividades <- setdiff(names(base_cluster), c("Rango edad", "Genero", "Procedencia"))
base_long <- melt(base_cluster, id.vars = "Genero", measure.vars = actividades)
base_long <- base_long[base_long$value == 1, ]

# Crear la tabla de contingencia
combined_table <- table(base_long$`Genero`, base_long$variable)
colnames(combined_table) <- paste0("y", 1:ncol(combined_table))

# Mostrar la tabla de contingencia
print(combined_table)

# Realizar el análisis de correspondencias
library(FactoMineR)
ca_result <- CA(combined_table, graph = FALSE)

##Inercia
ca_result$eig

# Extraer coordenadas como vectores

```

```

row_coords <- as.vector(ca_result$row$coord) # Coordinadas para Procedencia
col_coords <- as.vector(ca_result$col$coord) # Coordinadas para Actividades

# Crear DataFrame para Procedencia
df_procedencia <- data.frame(
  Coords = row_coords,
  Type = "Genero",
  Label = c("Femenino", "Masculino") # Etiquetas manuales para procedencia
)

# Crear DataFrame para Actividades
df_actividades <- data.frame(
  Coords = col_coords,
  Type = "Actividades",
  Label = paste0("y", 1:length(col_coords)) # Etiquetas automáticas para actividades
)

# Combinar ambos DataFrames
df_combined <- rbind(df_procedencia, df_actividades)

# Convertir las etiquetas a un factor con el orden numérico deseado
df_combined$Label <- factor(df_combined$Label, levels = c("Femenino", "Masculino",
paste0("y", 1:31)))

ggplot(df_combined, aes(x = Label, y = Coords, fill = Type)) +
  geom_bar(stat = "identity", position = "dodge") +
  theme_minimal() +
  labs(title = "Coordinadas en Dimensión 1: Género y Actividades", x = "Categorías", y =
"Coordinadas") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  scale_fill_manual(values = c("Genero" = "red", "Actividades" = "blue"))

%%R

actividades <- setdiff(names(base_cluster), c("Rango edad", "Genero", "Procedencia"))
base_long <- melt(base_cluster, id.vars = "Procedencia", measure.vars = actividades)
base_long <- base_long[base_long$value == 1, ]

# Crear la tabla de contingencia
combined_table <- table(base_long$Procedencia, base_long$variable)
colnames(combined_table) <- paste0("y", 1:ncol(combined_table))

# Mostrar la tabla de contingencia

```

```

print(combined table)

# Realizar el análisis de correspondencias
library(FactoMineR)
ca result <- CA(combined table, graph = FALSE)

##Inercia
ca_result$eig

# Extraer coordenadas como vectores
row_coords <- as.vector(ca_result$row$coord) # Coordenadas para Procedencia
col_coords <- as.vector(ca_result$col$coord) # Coordenadas para Actividades

# Crear DataFrame para Procedencia
df_procedencia <- data.frame(
  Coords = row_coords,
  Type = "Procedencia",
  Label = c("Nacional", "Extranjero") # Etiquetas manuales para procedencia
)

# Crear DataFrame para Actividades
df_actividades <- data.frame(
  Coords = col_coords,
  Type = "Actividades",
  Label = paste0("y", 1:length(col_coords)) # Etiquetas automáticas para actividades
)

# Combinar ambos DataFrames
df_combined <- rbind(df_procedencia, df_actividades)

# Verificar el DataFrame combinado
print(df_combined)

library(ggplot2)

# Convertir las etiquetas a un factor con el orden numérico deseado
df_combined$Label <- factor(df_combined$Label, levels = c("Nacional", "Extranjero",
paste0("y", 1:31)))

# Crear el gráfico con ggplot2
ggplot(df_combined, aes(x = Label, y = Coords, fill = Type)) +
  geom_bar(stat = "identity", position = "dodge") +
  theme_minimal() +

```

```

    labs(title = "Coordenadas en Dimensión 1: Procedencia y Actividades", x = "Categorías", y
= "Coordenadas") +
    theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
    scale_fill_manual(values = c("Procedencia" = "red", "Actividades" = "blue"))
from google.colab import files
import os
from google.colab import files
files.upload()

os.makedirs('/root/.kaggle', exist_ok=True)
os.rename('kaggle.json', '/root/.kaggle/kaggle.json')
!kaggle datasets download -d jangedoo/utkface-new
import zipfile

zip_path = '/content/utkface-new.zip'
extract_path = '/content/UTKFace/'

os.makedirs(extract_path, exist_ok=True)

with zipfile.ZipFile(zip_path, 'r') as zip_ref:
    zip_ref.extractall(extract_path)
import os

# Definir las rutas de las carpetas
base_path = '/content/UTKFace/'
folders = ['crop_part1', 'utkface_aligned_cropped', 'UTKFace']

# Verificar contenido de cada carpeta
for folder in folders:
    folder_path = os.path.join(base_path, folder)
    files = os.listdir(folder_path)
    print(f"Carpeta: {folder}")
    print(f"Total de archivos en la carpeta: {len(files)}")
    print(files[:10]) # Mostrar los primeros 10 archivos para verificar
    print()

# Ruta a la carpeta 'UTKFace'
utkface_path = '/content/UTKFace/UTKFace'
files = os.listdir(utkface_path)

print(f"Total de archivos en la carpeta 'UTKFace': {len(files)}")
print(files[:10]) # Mostrar los primeros 10 archivos para verificar
import numpy as np
import cv2
import os

```

```

from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Conv2D, MaxPooling2D, Flatten, Dense, Dropout
from tensorflow.keras.preprocessing.image import ImageDataGenerator
from sklearn.model_selection import train_test_split
from tensorflow.keras.callbacks import EarlyStopping

# Cargar las imágenes y las edades
images = []
ages = []

for file in files:
    img_path = os.path.join(utkface_path, file)
    image = cv2.imread(img_path)
    if image is not None:
        # Redimensionar y normalizar la imagen
        image = cv2.resize(image, (128, 128))
        image = image / 255.0 # Normalizar los valores de píxel al rango [0, 1]
        images.append(image)

        # Extraer edad del nombre del archivo
        split_var = file.split(' ')
        ages.append(int(split_var[0]))

# Convertir listas a arrays de NumPy
images = np.array(images)
ages = np.array(ages)

# Dividir los datos en conjuntos de entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(images, ages, test_size=0.2,
random_state=42)

# Dividir el conjunto de entrenamiento en un conjunto de entrenamiento y un conjunto de
validación
X_train, X_val, y_train, y_val = train_test_split(X_train, y_train, test_size=0.2,
random_state=42)

# Aumentación de datos y preprocesamiento
datagen_train = ImageDataGenerator(
    rotation_range=20,
    width_shift_range=0.2,
    height_shift_range=0.2,
    horizontal_flip=True,
    zoom_range=0.2
)

```

```

datagen_val = ImageDataGenerator() # No se aplica aumento a los datos de validación

datagen_train.fit(X_train)
datagen_val.fit(X_val)

# Crear el modelo con más complejidad
model = Sequential([
    Conv2D(64, (5, 5), activation='relu', input_shape=(128, 128, 3)),
    MaxPooling2D(2, 2),
    Conv2D(128, (3, 3), activation='relu'),
    MaxPooling2D(2, 2),
    Flatten(),
    Dense(256, activation='relu'),
    Dense(128, activation='relu'),
    Dropout(0.5), # Regularización para prevenir sobreajuste
    Dense(1) # Una sola neurona para la salida (edad)
])

model.compile(optimizer='adam', loss='mean absolute error', metrics=['mae'])

# Implementar Early Stopping
early_stopping = EarlyStopping(monitor='val_loss', patience=10, restore_best_weights=True)

# Entrenar el modelo usando el generador de datos aumentado
history = model.fit(
    datagen_train.flow(X_train, y_train, batch_size=32),
    validation_data=datagen_val.flow(X_val, y_val),
    epochs=50,
    callbacks=[early_stopping]
)

# Evaluar el modelo
test_loss, test_mae = model.evaluate(X_test, y_test, verbose=2)
print(f'Test MAE: {test_mae:.2f}')

# Guardar el modelo entrenado
model.save('age_prediction_model.h5')
import matplotlib.pyplot as plt

# Graficar la pérdida (loss) y el MAE en función de las épocas
def plot_training_history(history):
    # Gráfica de loss
    plt.figure(figsize=(12, 6))

    # Gráfico de la pérdida (loss)

```

```

plt.subplot(1, 2, 1)
plt.plot(history.history['loss'], label='Pérdida en Entrenamiento')
plt.plot(history.history['val_loss'], label='Pérdida en Validación')
plt.title('Pérdida a lo largo de las épocas')
plt.xlabel('Épocas')
plt.ylabel('Pérdida')
plt.legend()

# Gráfico del MAE
plt.subplot(1, 2, 2)
plt.plot(history.history['mae'], label='MAE en Entrenamiento')
plt.plot(history.history['val_mae'], label='MAE en Validación')
plt.title('MAE a lo largo de las épocas')
plt.xlabel('Épocas')
plt.ylabel('MAE')
plt.legend()

# Mostrar las gráficas
plt.tight layout()
plt.show()

# Llamar a la función para graficar la historia de entrenamiento
plot_training_history(history)

from google.colab import drive
drive.mount('/content/drive')

from google.colab import files

# Guardar el modelo
model.save('age_prediction_model.h5')
# Descargar el modelo
files.download('age_prediction_model.h5')
from google.colab import files
uploaded = files.upload() # Esto te permitirá subir el archivo .h5 desde tu computadora
import pandas as pd
from tensorflow.keras.models import load_model
import cv2
import numpy as np

# Diccionario de mapeo de actividades
activity_names = {
    'y1': 'Te recomiendo visitar los centros comerciales de las diferentes ciudades del
departamento. Para más información sobre centros comerciales, puedes visitar esta página:
[Tripadvisor - Centros Comerciales en el Valle del

```

Cauca](https://www.tripadvisor.co/Attractions-g2021548-Activities-c26-t143-Valle_del_Cauca_Department.html).',

'y2': 'Te recomiendo recorrer las calles y parques de las diferentes ciudades del departamento',

'y3': 'Puedes visitar museos, casas de la cultura, iglesias, santuarios y monumentos. Para más información sobre museos, iglesias y demas en el deparamento, puedes visitar esta página :[Tripadvisor - Museos, Iglesias y Monumentos en el Valle del Cauca]([https://www.tripadvisor.co/Attractions-g2021548-Activities-c47-Valle del Cauca Department.html](https://www.tripadvisor.co/Attractions-g2021548-Activities-c47-Valle_del_Cauca_Department.html)).',

'y4': 'Puedes visitar Haciendas y/o casas históricas. Para más información sobre estas haciendas, puedes visitar esta página: [TurismoI - Haciendas en el Valle](<https://turismoI.co/tours/haciendas-coloniales-un-viaje-a-las-tradiciones-del-valle>).',

'y5': 'Te recomiendo realizar actividades religiosas. Para más información sobre que hacer, puedes visitar esta página: [VisitValle - Tradición y Ancestralidad en el Valle del Cauca](<https://visitvalle.travel/blogs-tradiccion-ancestralidad.html>).',

'y6': 'Puedes asistir a espectáculos artísticos y festivales de las diferentes ciudades en el Valle. Para más información sobre estas festividades, puedes visitar esta página: [LiveValle - Ferias y Fiestas en el Valle del Cauca](<https://www.livevalledelcauca.com/ferias-y-fiestas.html>).',

'y7': 'Te recomiendo realizar avistamiento de aves en el departamento. ara más información sobre las mejores rutas, puedes visitar esta página: [Wikiloc - Rutas de Observación de Aves en el Valle del Cauca](<https://es.wikiloc.com/rutas/observacion-de-aves/colombia/valle-del-cauca>).',

'y8': 'Te recomiendo realizar avistamiento de ballenas en el Valle. Para más información sobre el avistamiento de ballenas, puedes visitar esta página: [Avistamiento de Ballenas - Información y Reservas](<https://avistamientodeballenas.co/>).',

'y9': 'Puedes ir a discotecas, bares o karaokes en el Valle del Cauca. Para más información sobre estos lugares, puedes visitar esta página: [Tripadvisor - Bares y Discotecas en el Valle del Cauca](https://www.tripadvisor.co/Attractions-g2021548-Activities-c20-t99-Valle_del_Cauca_Department.html).',

'y10': 'Visita los parques temáticos o parques de atracciones del departamento. Para más información, puedes visitar esta página: [Tripadvisor - Parques Temáticos y de Atracciones en el Valle del Cauca](https://www.tripadvisor.es/Attractions-g2021548-Activities-c52-t98-Valle_del_Cauca_Department.html).',

'y11': 'Te recomiendo ir de compras. Para más información sobre las mejores opciones de compras, puedes visitar esta página: [Minube - Compras en el Valle del Cauca](<https://www.minube.com.co/top/compras-valle-del-cauca-z4216>).',

'y12': 'El destino es un lugar en donde puedes realizar reuniones de negocio. Para más información sobre la industria de reuniones, puedes visitar esta página: [Colombia Travel - Industria de Reuniones en Cali](<https://colombia.travel/es/reuniones/cali-industria-de-reuniones>).',

'y13': 'El destino es un lugar en donde puedes practicar deportes.',

'y14': 'Te recomiendo asistir a competencias o muestras deportivas. Para más información sobre eventos deportivos y recreativos, puedes visitar esta página: [Cali - Secretaría del

Deporte y la Recreación](https://www.cali.gov.co/calendario/categoria/790/secretaria-del-deporte-y-la-recreacion/).',

'y15': 'Te recomiendo asistir a conferencias o congresos. Para más información sobre eventos y conferencias, puedes visitar esta página: [Centro de Eventos Valle del Pacífico - Calendario](https://valledelpacifico.co/calendario/).',

'y16': 'Te recomiendo asistir a fiestas y ferias de las ciudades del departamento. Para más información sobre las ferias y fiestas, puedes visitar esta página: [Viaja por Colombia - Ferias y Fiestas en el Valle del Cauca](https://viajaporcolombia.com/ferias-y-fiestas/valle-del-cauca/).',

'y17': 'Puedes visitar las fabricas o industrias del departamento. Para más información sobre la ruta de turismo industrial, puedes visitar esta página: [90 Minutos - Ruta de Turismo Industrial en el Valle del Cauca](https://90minutos.co/valle-del-cauca/ruta-de-turismo-industrial-31-05-2024/).',

'y18': 'Puedes visitar casinos y otros sitios de juegos de azar. Para más información sobre estos sitios en el Valle del Cauca, puedes visitar esta página: [Tripadvisor - Hoteles en el Valle del Cauca](https://www.tripadvisor.co/Hotels-g2021548-zffl4-Valle-del-Cauca-Department-Hotels.html).',

'y19': 'Te recomiendo visitar los ríos, lagos y cascadas del departamento. Para más información sobre estos lugares, puedes visitar esta página: [Tripadvisor - Ríos, Lagos y Cascadas en el Valle del Cauca](https://www.tripadvisor.co/Attractions-g2021548-Activities-c57-t162-Valle-del-Cauca-Department.html).',

'y20': 'Te recomiendo visitar miradores paisajísticos en el Valle. Para más información sobre los mejores miradores, puedes visitar esta página: [TuBarco - Miradores Imperdibles en el Valle del Cauca](https://tubarco.news/los-miradores-imperdibles-para-conocer-en-el-valle-del-cauca/).',

'y21': 'Te recomiendo visitar el Zoológico de Cali, uno de los principales atractivos del Valle del Cauca. Para más información, puedes visitar esta página: [Zoológico de Cali - Página Oficial](https://www.zoologicodecali.com.co/).',

'y22': 'Te recomiendo conocer como es el proceso del Café. Para más información sobre esta actividad, puedes visitar esta página: [Tripadvisor - Valley Adventours en Cali](https://www.tripadvisor.co/ShowUserReviews-g297475-d9465730-r568683635-Valley_Adventours-Cali_Valle_del_Cauca_Department.html).',

'y23': 'Te recomiendo conocer el proceso del azúcar dulce visitando las haciendas de caña de azúcar en el Valle del Cauca. Para más información sobre estos tours, puedes visitar esta página: [Civitatis - Tour por Haciendas de Caña de Azúcar en Cali](https://www.civitatis.com/es/cali/tour-haciendas-cana-azucar-cali/?srsltid=AfmBOora0SyG7LcUs8mbV4T79Af_TGFWBJHw0PcjBp5fYpn6Dsnnikq4).',

'y24': 'Puedes realizar actividades de aventura como el Rappel, parapente, bungee, escalada, espelelismo y canotaje. Para más información sobre estas actividades, puedes visitar esta página: [GetYourGuide - Deportes Extremos y Adrenalina en el Valle del Cauca](https://www.getyourguide.es/valle-del-cauca-11212/deportes-extremos-y-adrenalina-tc85/).',

'y25': 'Te recomiendo realizar senderismo en el Valle del Cauca. Para más información sobre las mejores rutas de senderismo, puedes visitar esta página: [Wikiloc - Rutas de

```

Senderismo en el Valle del Cauca](https://es.wikiloc.com/rutas/senderismo/colombia/valle-del-cauca).',

'y26': 'Te recomiendo visitar reservas naturales y parques naturales en el Valle del Cauca. Para más información sobre estos lugares, puedes visitar esta página: [Tripadvisor - Reservas Naturales y Parques en el Valle del Cauca](https://www.tripadvisor.co/Attractions-g2021548-Activities-c57-t67-Valle_del_Cauca_Department.html).',

'y27': 'Te recomiendo visitar las playas en el Valle del Cauca. Para más información sobre las mejores playas, puedes visitar esta página: [Tripadvisor - Playas en el Valle del Cauca](https://www.tripadvisor.co/Attractions-g2021548-Activities-c61-t52-Valle_del_Cauca_Department.html).',

'y29': 'El destino es un lugar en donde puedes realizar bodas o eventos familiares. Para más información sobre salones de fiestas y eventos, puedes visitar esta página: [Matrimonio.com.co - Salones de Fiestas en el Valle del Cauca](https://www.matrimonio.com.co/salones-de-fiestas/valle-del-cauca).',

'y30': 'Puedes visitar el departamento con familiares y amigos',

'y31': 'Te recomiendo disfrutar de actividades gastronómicas en el Valle del Cauca. Para más información sobre experiencias culinarias, puedes visitar esta página: [Tripadvisor - Actividades Gastronómicas en el Valle del Cauca](https://www.tripadvisor.co/Attractions-g2021548-Activities-c36-Valle del Cauca Department.html).'
}

# Cargar la tabla de asociación
activity_age_association = pd.read_csv("/content/activity_age_association.csv")

# Reemplazar los identificadores yX con los nombres reales de las actividades
activity_age_association['Activity'] =
activity_age_association['Activity'].map(activity_names)

# Función para capturar la imagen de la cámara web
from IPython.display import display, Image, Javascript
from google.colab.output import eval_js
from base64 import b64decode

def take_photo(filename='photo.jpg', quality=0.8):
    js = Javascript('''
        async function takePhoto(quality) {
            const div = document.createElement('div');
            const capture = document.createElement('button');
            capture.textContent = 'Capture';
            div.appendChild(capture);

            const video = document.createElement('video');
            video.style.display = 'block';
            const stream = await navigator.mediaDevices.getUserMedia({video: true});

```

```

        document.body.appendChild(div);
        div.appendChild(video);
        video.srcObject = stream;
        await video.play();

        google.colab.output.setIframeHeight(document.documentElement.scrollHeight, true);

        await new Promise((resolve) => capture.onclick = resolve);

        const canvas = document.createElement('canvas');
        canvas.width = video.videoWidth;
        canvas.height = video.videoHeight;
        canvas.getContext('2d').drawImage(video, 0, 0);
        stream.getVideoTracks()[0].stop();
        div.remove();
        return canvas.toDataURL('image/jpeg', quality);
    }
    '')
    display(js)
    data = eval_js('takePhoto({})'.format(quality))
    binary = b64decode(data.split(',')[1])
    with open(filename, 'wb') as f:
        f.write(binary)
    return filename

# Cargar el modelo y hacer predicción
model = load_model('/content/age_prediction_model.h5')

try:
    filename = take_photo()
    print('Saved to {}'.format(filename))

    # Mostrar la imagen capturada
    display(Image(filename))

    # Cargar y procesar la imagen para la predicción
    img = cv2.imread(filename)
    img = cv2.resize(img, (128, 128)) # Redimensionar para que coincida con el tamaño de
    entrada del modelo
    img = img / 255.0 # Normalizar la imagen
    img = np.expand_dims(img, axis=0) # Agregar dimensión de batch

    # Realizar la predicción
    prediction = model.predict(img)
    predicted age = prediction[0][0]

```

```

print(f'Predicted age: {predicted_age:.2f}')

# Asignar el rango de edad según la predicción
def assign_age_range(predicted_age):
    if predicted_age <= 26:
        return "15-26"
    elif predicted_age <= 38:
        return "27-38"
    elif predicted_age <= 50:
        return "39-50"
    else:
        return "51-62" # Incluye a los de 63+

assigned_range = assign_age_range(predicted_age)

# Filtrar las actividades asociadas al rango de edad asignado
activities_for_age = activity_age_association[activity_age_association['Rango Edad'] ==
assigned_range]['Activity'].tolist()

print(f'Assigned age range: {assigned_range}')
print('Activities:')

# Imprimir cada actividad en una nueva línea
for activity in activities_for_age:
    print(f'- {activity}')

except Exception as err:
    print(str(err))
!pip install pyngrok

!pip install streamlit
code = '''
import streamlit as st
from tensorflow.keras.models import load_model
import cv2
import numpy as np
import pandas as pd

model = load_model('/content/age_prediction_model.h5')
activity_age_association = pd.read_csv("/content/activity_age_association.csv")
activity_names = {
    'y1': 'Te recomiendo visitar los centros comerciales de las diferentes ciudades del
departamento. Para más información sobre centros comerciales, puedes visitar esta página:
[Tripadvisor - Centros Comerciales en el Valle del

```

Cauca](https://www.tripadvisor.co/Attractions-g2021548-Activities-c26-t143-Valle_del_Cauca_Department.html).',

'y2': 'Te recomiendo recorrer las calles y parques de las diferentes ciudades del departamento',

'y3': 'Puedes visitar museos, casas de la cultura, iglesias, santuarios y monumentos. Para más información sobre museos, iglesias y demas en el deparamento, puedes visitar esta página :[Tripadvisor - Museos, Iglesias y Monumentos en el Valle del Cauca]([https://www.tripadvisor.co/Attractions-g2021548-Activities-c47-Valle del Cauca Department.html](https://www.tripadvisor.co/Attractions-g2021548-Activities-c47-Valle_del_Cauca_Department.html)).',

'y4': 'Puedes visitar Haciendas y/o casas históricas. Para más información sobre estas haciendas, puedes visitar esta página: [TurismoI - Haciendas en el Valle](<https://turismoI.co/tours/haciendas-coloniales-un-viaje-a-las-tradiciones-del-valle>).',

'y5': 'Te recomiendo realizar actividades religiosas. Para más información sobre que hacer, puedes visitar esta página: [VisitValle - Tradición y Ancestralidad en el Valle del Cauca](<https://visitvalle.travel/blogs-tradiccion-ancestralidad.html>).',

'y6': 'Puedes asistir a espectáculos artísticos y festivales de las diferentes ciudades en el Valle. Para más información sobre estas festividades, puedes visitar esta página: [LiveValle - Ferias y Fiestas en el Valle del Cauca](<https://www.livevalledelcauca.com/ferias-y-fiestas.html>).',

'y7': 'Te recomiendo realizar avistamiento de aves en el departamento. ara más información sobre las mejores rutas, puedes visitar esta página: [Wikiloc - Rutas de Observación de Aves en el Valle del Cauca](<https://es.wikiloc.com/rutas/observacion-de-aves/colombia/valle-del-cauca>).',

'y8': 'Te recomiendo realizar avistamiento de ballenas en el Valle. Para más información sobre el avistamiento de ballenas, puedes visitar esta página: [Avistamiento de Ballenas - Información y Reservas](<https://avistamientodeballenas.co/>).',

'y9': 'Puedes ir a discotecas, bares o karaokes en el Valle del Cauca. Para más información sobre estos lugares, puedes visitar esta página: [Tripadvisor - Bares y Discotecas en el Valle del Cauca](https://www.tripadvisor.co/Attractions-g2021548-Activities-c20-t99-Valle_del_Cauca_Department.html).',

'y10': 'Visita los parques temáticos o parques de atracciones del departamento. Para más información, puedes visitar esta página: [Tripadvisor - Parques Temáticos y de Atracciones en el Valle del Cauca](https://www.tripadvisor.es/Attractions-g2021548-Activities-c52-t98-Valle_del_Cauca_Department.html).',

'y11': 'Te recomiendo ir de compras. Para más información sobre las mejores opciones de compras, puedes visitar esta página: [Minube - Compras en el Valle del Cauca](<https://www.minube.com.co/top/compras-valle-del-cauca-z4216>).',

'y12': 'El destino es un lugar en donde puedes realizar reuniones de negocio. Para más información sobre la industria de reuniones, puedes visitar esta página: [Colombia Travel - Industria de Reuniones en Cali](<https://colombia.travel/es/reuniones/cali-industria-de-reuniones>).',

'y13': 'El destino es un lugar en donde puedes practicar deportes.',

'y14': 'Te recomiendo asistir a competencias o muestras deportivas. Para más información sobre eventos deportivos y recreativos, puedes visitar esta página: [Cali - Secretaría del

Deporte y la Recreación](<https://www.cali.gov.co/calendario/categoria/790/secretaria-del-deporte-y-la-recreacion/>).',

'y15': 'Te recomiendo asistir a conferencias o congresos. Para más información sobre eventos y conferencias, puedes visitar esta página: [Centro de Eventos Valle del Pacífico - Calendario](<https://valledelpacifico.co/calendario/>).',

'y16': 'Te recomiendo asistir a fiestas y ferias de las ciudades del departamento. Para más información sobre las ferias y fiestas, puedes visitar esta página: [Viaja por Colombia - Ferias y Fiestas en el Valle del Cauca](<https://viajaporcolombia.com/ferias-y-fiestas/valle-del-cauca/>).',

'y17': 'Puedes visitar las fábricas o industrias del departamento. Para más información sobre la ruta de turismo industrial, puedes visitar esta página: [90 Minutos - Ruta de Turismo Industrial en el Valle del Cauca](<https://90minutos.co/valle-del-cauca/ruta-de-turismo-industrial-31-05-2024/>).',

'y18': 'Puedes visitar casinos y otros sitios de juegos de azar. Para más información sobre estos sitios en el Valle del Cauca, puedes visitar esta página: [Tripadvisor - Hoteles en el Valle del Cauca](<https://www.tripadvisor.co/Hotels-g2021548-zffl4-Valle-del-Cauca-Department-Hotels.html>).',

'y19': 'Te recomiendo visitar los ríos, lagos y cascadas del departamento. Para más información sobre estos lugares, puedes visitar esta página: [Tripadvisor - Ríos, Lagos y Cascadas en el Valle del Cauca](<https://www.tripadvisor.co/Attractions-g2021548-Activities-c57-t162-Valle-del-Cauca-Department.html>).',

'y20': 'Te recomiendo visitar miradores paisajísticos en el Valle. Para más información sobre los mejores miradores, puedes visitar esta página: [TuBarco - Miradores Imperdibles en el Valle del Cauca](<https://tubarco.news/los-miradores-imperdibles-para-conocer-en-el-valle-del-cauca/>).',

'y21': 'Te recomiendo visitar el Zoológico de Cali, uno de los principales atractivos del Valle del Cauca. Para más información, puedes visitar esta página: [Zoológico de Cali - Página Oficial](<https://www.zoologicodecali.com.co/>).',

'y22': 'Te recomiendo conocer como es el proceso del Café. Para más información sobre esta actividad, puedes visitar esta página: [Tripadvisor - Valley Adventours en Cali](https://www.tripadvisor.co/ShowUserReviews-g297475-d9465730-r568683635-Valley_Adventours-Cali_Valle_del_Cauca_Department.html).',

'y23': 'Te recomiendo conocer el proceso del azúcar dulce visitando las haciendas de caña de azúcar en el Valle del Cauca. Para más información sobre estos tours, puedes visitar esta página: [Civitatis - Tour por Haciendas de Caña de Azúcar en Cali](https://www.civitatis.com/es/cali/tour-haciendas-cana-azucar-cali/?srsltid=AfmBOora0SyG7LcUs8mbV4T79Af_TGFWBJHw0PcjBp5fYpn6Dsnnikq4).',

'y24': 'Puedes realizar actividades de aventura como el Rappel, parapente, bungee, escalada, espelelismo y canotaje. Para más información sobre estas actividades, puedes visitar esta página: [GetYourGuide - Deportes Extremos y Adrenalina en el Valle del Cauca](<https://www.getyourguide.es/valle-del-cauca-11212/deportes-extremos-y-adrenalina-tc85/>).',

'y25': 'Te recomiendo realizar senderismo en el Valle del Cauca. Para más información sobre las mejores rutas de senderismo, puedes visitar esta página: [Wikiloc - Rutas de

```

Senderismo en el Valle del Cauca](https://es.wikiloc.com/rutas/senderismo/colombia/valle-del-cauca).',

'y26': 'Te recomiendo visitar reservas naturales y parques naturales en el Valle del Cauca. Para más información sobre estos lugares, puedes visitar esta página: [Tripadvisor - Reservas Naturales y Parques en el Valle del Cauca](https://www.tripadvisor.co/Attractions-g2021548-Activities-c57-t67-Valle_del_Cauca_Department.html).',

'y27': 'Te recomiendo visitar las playas en el Valle del Cauca. Para más información sobre las mejores playas, puedes visitar esta página: [Tripadvisor - Playas en el Valle del Cauca](https://www.tripadvisor.co/Attractions-g2021548-Activities-c61-t52-Valle_del_Cauca_Department.html).',

'y29': 'El destino es un lugar en donde puedes realizar bodas o eventos familiares. Para más información sobre salones de fiestas y eventos, puedes visitar esta página: [Matrimonio.com.co - Salones de Fiestas en el Valle del Cauca](https://www.matrimonio.com.co/salones-de-fiestas/valle-del-cauca).',

'y30': 'Puedes visitar el departamento con familiares y amigos',

'y31': 'Te recomiendo disfrutar de actividades gastronómicas en el Valle del Cauca. Para más información sobre experiencias culinarias, puedes visitar esta página: [Tripadvisor - Actividades Gastronómicas en el Valle del Cauca](https://www.tripadvisor.co/Attractions-g2021548-Activities-c36-Valle del Cauca Department.html).'
}

activity_age_association['Activity'] =
activity_age_association['Activity'].map(activity_names)

def assign_age_range(predicted_age):
    if predicted_age <= 26:
        return "15-26"
    elif predicted_age <= 38:
        return "27-38"
    elif predicted_age <= 50:
        return "39-50"
    else:
        return "51-62"

st.title("Age Prediction and Activity Recommendation")

uploaded_file = st.file_uploader("Upload an image...", type=["jpg", "jpeg", "png"])
if uploaded_file is not None:
    image = cv2.imdecode(np.frombuffer(uploaded_file.read(), np.uint8), 1)
    st.image(image, channels="BGR")

    img_resized = cv2.resize(image, (128, 128))
    img_normalized = img_resized / 255.0
    img_expanded = np.expand_dims(img_normalized, axis=0)

    prediction = model.predict(img_expanded)

```

```

    predicted_age = prediction[0][0]
    st.write(f"Predicted age: {predicted_age:.2f}")

    assigned_range = assign_age_range(predicted_age)
    activities_for_age = activity_age_association[activity_age_association['Rango Edad'] ==
assigned_range]['Activity'].tolist()

    st.write(f"Assigned age range: {assigned_range}")
    st.write("Recommended activities:")
    for activity in activities_for_age:
        st.write(f"- {activity}")
'''

with open('/content/app.py', 'w') as f:
    f.write(code)
from pyngrok import ngrok
ngrok.set_auth_token("21aUUOUJkc7VGD8FqIujhlmUvmT 4WLth2wY7bEwbFuEpFBfY")

import subprocess
from pyngrok import ngrok

# Inicia Streamlit en segundo plano
process = subprocess.Popen(['streamlit', 'run', '/content/app.py', '--server.port', '8501'])

# Crear un túnel con ngrok al puerto 8501
public_url = ngrok.connect(8501)
print(f"Streamlit app is live at: {public_url}")

# Mantén el túnel activo hasta que detengas el proceso
try:
    process.wait() # Espera a que el proceso de Streamlit termine
except KeyboardInterrupt:
    print("Streamlit app has been stopped.")
    process.terminate()

```