**Informe del Proyecto: ReDim - Análisis Multivariado y Métodos Predictivos**

**Daniel Felipe Ruiz Bermúdez y Juan Sebastián Fajardo Acevedo**

**1. Introducción**

La aplicación **ReDim** (Reducción Dimensional y Métodos Predictivos) es una herramienta interactiva diseñada para realizar análisis multivariado y aplicar técnicas de aprendizaje automático (machine learning) con un enfoque práctico, visual y accesible. Este proyecto se ha desarrollado utilizando la librería **Streamlit**, lo que permite crear interfaces gráficas dinámicas para facilitar la interacción con los modelos predictivos y las visualizaciones.

**2. Objetivo del Proyecto**

El propósito de **ReDim** es proporcionar una herramienta accesible para realizar análisis de datos complejos, aplicando métodos estadísticos y técnicas de machine learning. Los usuarios pueden realizar análisis exploratorio de datos (EDA), aplicar técnicas estadísticas como **PCA (Análisis de Componentes Principales)** y **MCA (Análisis de Correspondencias Múltiples)** para la reducción de la dimensionalidad, y entrenar modelos predictivos para tareas de **regresión** y **clasificación**.

**Objetivos específicos:**

* Realizar un análisis exploratorio de datos (EDA) que permita a los usuarios visualizar y explorar los datos de forma eficiente.
* Entrenar modelos predictivos utilizando algoritmos de **regresión** y **clasificación** como **Regresión Logística**, **QDA**, **LDA**, **Árboles de Decisión**, y **Random Forest**.
* Facilitar la **reducción de dimensionalidad** mediante **PCA** y **MCA** para mejorar la interpretación y visualización de datos complejos.
* Proporcionar una interfaz interactiva que permita realizar **predicciones manuales** y **comparación de modelos**.

**3. Público Objetivo**

**ReDim** está dirigido a:

* **Científicos de Datos** que necesitan herramientas rápidas y fáciles para análisis de datos y construcción de modelos predictivos.
* **Estudiantes de Estadística y Machine Learning** que deseen practicar con datos reales y aprender sobre análisis multivariado y modelos predictivos.
* **Investigadores** en diversas áreas (salud, marketing, ciencias sociales, etc.) que necesiten explorar grandes volúmenes de datos y aplicar técnicas estadísticas avanzadas para hacer predicciones.
* **Consultores de datos** que buscan una herramienta accesible para analizar datos y generar informes rápidos y fáciles de interpretar.

**4. Descripción General de la Aplicación**

La aplicación está organizada en varias secciones que permiten realizar análisis detallados y entrenar modelos predictivos. Estas son las principales funcionalidades y herramientas incluidas en **ReDim**:

**4.1 Carga y Preprocesamiento de Datos**

El primer paso al usar la aplicación es cargar un archivo de datos en formato **CSV**. La aplicación automáticamente identifica las variables numéricas y categóricas, y permite realizar un análisis previo con las primeras filas del archivo.

**4.2 Análisis Exploratorio de Datos (EDA)**

La aplicación permite realizar un **Análisis Exploratorio de Datos (EDA)** a través de varias herramientas estadísticas y gráficas. Entre las opciones más destacadas:

* **Estadísticas Descriptivas**: Se presentan medidas de tendencia central (media, mediana), dispersión (desviación estándar) y otras métricas clave para las variables numéricas.
* **Matriz de Valores Nulos**: Visualiza patrones de valores faltantes usando la librería **missingno**, que ayuda a identificar la calidad de los datos antes del análisis o entrenamiento de modelos.
* **Distribución de Variables**: Se generan histogramas y gráficos de caja para las variables numéricas seleccionadas, proporcionando una visualización rápida de la distribución de cada variable.
* **Conteo de Variables Categóricas**: Muestra la distribución de las categorías dentro de las variables categóricas.
* **Matriz de Correlación**: Visualiza las relaciones lineales entre las variables numéricas, utilizando **seaborn** para generar un mapa de calor con los coeficientes de correlación.

**4.3 Modelos Predictivos**

Esta sección permite a los usuarios entrenar modelos predictivos utilizando técnicas de **regresión** y **clasificación**. Entre las características clave de esta sección están:

* **Selección de Variables**: El usuario selecciona la variable dependiente (target) y las variables predictoras (X).
* **Configuración Avanzada**: Se pueden ajustar parámetros como el **tamaño del conjunto de prueba** (porcentaje de los datos que se utilizarán para probar el modelo), la **semilla aleatoria** (para asegurar la reproducibilidad), y otros parámetros específicos para cada modelo, como la **profundidad máxima** de un árbol de decisión.
* **Entrenamiento del Modelo**: La aplicación entrena el modelo seleccionado y evalúa su rendimiento utilizando métricas específicas para cada tipo de modelo:
  + **Regresión**: **MSE (Error Cuadrático Medio)**, **RMSE (Raíz del Error Cuadrático Medio)** y **R² (Coeficiente de Determinación)**.
  + **Clasificación**: **Accuracy (Exactitud)**, **Precision (Precisión)**, **Recall (Sensibilidad)**, **F1-Score**, y **AUC (Área Bajo la Curva)**.
* **Curva ROC y AUC**: Para modelos de clasificación, se generan **Curvas ROC** y se calcula el **AUC**, lo cual indica la capacidad del modelo para distinguir entre las clases positivas y negativas. Un AUC cercano a 1 indica un excelente desempeño.
* **Matriz de Confusión**: Se genera la **matriz de confusión**, que muestra los verdaderos positivos, falsos positivos, verdaderos negativos y falsos negativos. Esta matriz es útil para identificar posibles errores de clasificación.

**4.4 Comparación de Modelos**

Los usuarios pueden seleccionar varios modelos previamente entrenados para compararlos en términos de su rendimiento. La aplicación proporciona:

* **Tabla Comparativa**: Una tabla que resume las métricas clave de cada modelo, permitiendo al usuario ver cuál es el más eficiente.
* **Gráfico Comparativo**: Los resultados de los modelos se grafican en un gráfico de barras para comparar visualmente su desempeño.
* **Comparación de Curvas ROC**: Si todos los modelos seleccionados son clasificadores, la aplicación genera las **curvas ROC** para comparar la capacidad de los modelos de clasificar correctamente.

**4.5 Reducción de Dimensionalidad (PCA y MCA)**

La aplicación incluye herramientas de **reducción de dimensionalidad** que permiten a los usuarios reducir la complejidad de los datos, facilitando su visualización e interpretación.

* **PCA (Análisis de Componentes Principales)**: Permite realizar PCA para reducir la cantidad de variables y obtener componentes principales. La varianza explicada por cada componente se muestra y se genera un gráfico de dispersión con los dos primeros componentes.
* **MCA (Análisis de Correspondencias Múltiples)**: Similar a PCA, pero enfocado en datos categóricos. Permite reducir las dimensiones de los datos categóricos y visualizarlos en un espacio reducido, ayudando a identificar patrones y relaciones entre las categorías.

**4.6 Predicción Manual**

En esta sección, el usuario puede introducir manualmente valores para las variables predictoras y la aplicación calculará la predicción en tiempo real usando el modelo entrenado. Esto permite realizar predicciones sin necesidad de reentrenar el modelo, basándose en la información actual del usuario.

**5. Herramientas y Métodos Estadísticos Utilizados**

El desarrollo de esta aplicación se basa en varias técnicas estadísticas y métodos de machine learning, que se aplican a través de herramientas y librerías específicas:

**5.1 Modelos de Regresión y Clasificación**

Los **modelos de regresión** y **clasificación** utilizados en la aplicación están basados en métodos estadísticos bien establecidos.

* **Regresión Lineal**: Utiliza la relación lineal entre las variables predictoras y la variable dependiente para hacer predicciones.
* **Regresión Logística**: Es utilizada para la clasificación binaria. Aplica la función sigmoide para convertir la salida de la regresión lineal en probabilidades que pueden ser interpretadas como clases.
* **Árboles de Decisión**: Un método no paramétrico que divide los datos en subgrupos basados en reglas de decisión, ideal para clasificación y regresión.
* **Random Forest**: Un conjunto de árboles de decisión que mejora la precisión de las predicciones al promediar los resultados de varios árboles individuales.
* **LDA y QDA**: Métodos estadísticos basados en la teoría de distribución normal, utilizados para clasificación. LDA asume una covarianza común entre las clases, mientras que QDA permite que cada clase tenga su propia covarianza.

**5.2 PCA y MCA**

* **PCA (Análisis de Componentes Principales)**: Utiliza técnicas de álgebra lineal para encontrar las direcciones principales de variabilidad en los datos. Estas direcciones, conocidas como componentes principales, permiten reducir la dimensionalidad sin perder demasiada información.
* **MCA (Análisis de Correspondencias Múltiples)**: Es una extensión del análisis de correspondencias que se utiliza para trabajar con datos categóricos. MCA realiza una descomposición de los datos categóricos en componentes principales, similar a PCA pero en el contexto de variables cualitativas.

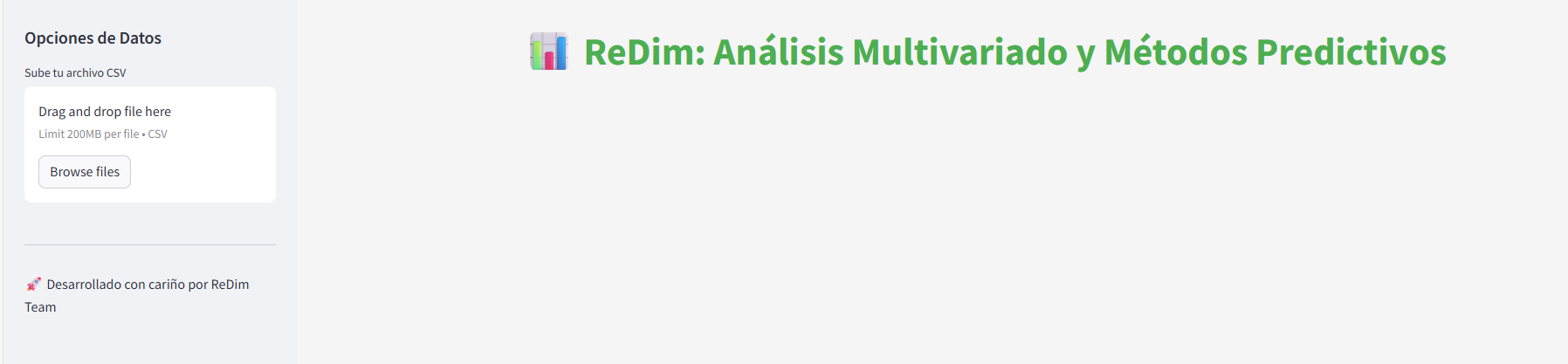
**5.3 Métodos de Evaluación**

* **Métricas de regresión**: Las métricas como **MSE**, **RMSE** y **R²** se utilizan para evaluar la precisión de los modelos de regresión. El **R²** es especialmente útil para medir la proporción de la variabilidad de la variable dependiente que es explicada por el modelo.
* **Métricas de clasificación**: Las métricas como **Accuracy**, **Precision**, **Recall**, y **F1-Score** son esenciales para evaluar el desempeño de los modelos clasificadores, especialmente en situaciones de clases desbalanceadas.
* **AUC**: El **Área Bajo la Curva (AUC)** mide la capacidad del modelo para discriminar entre las clases. Un **AUC** más cercano a 1 indica que el modelo es más efectivo en la clasificación.

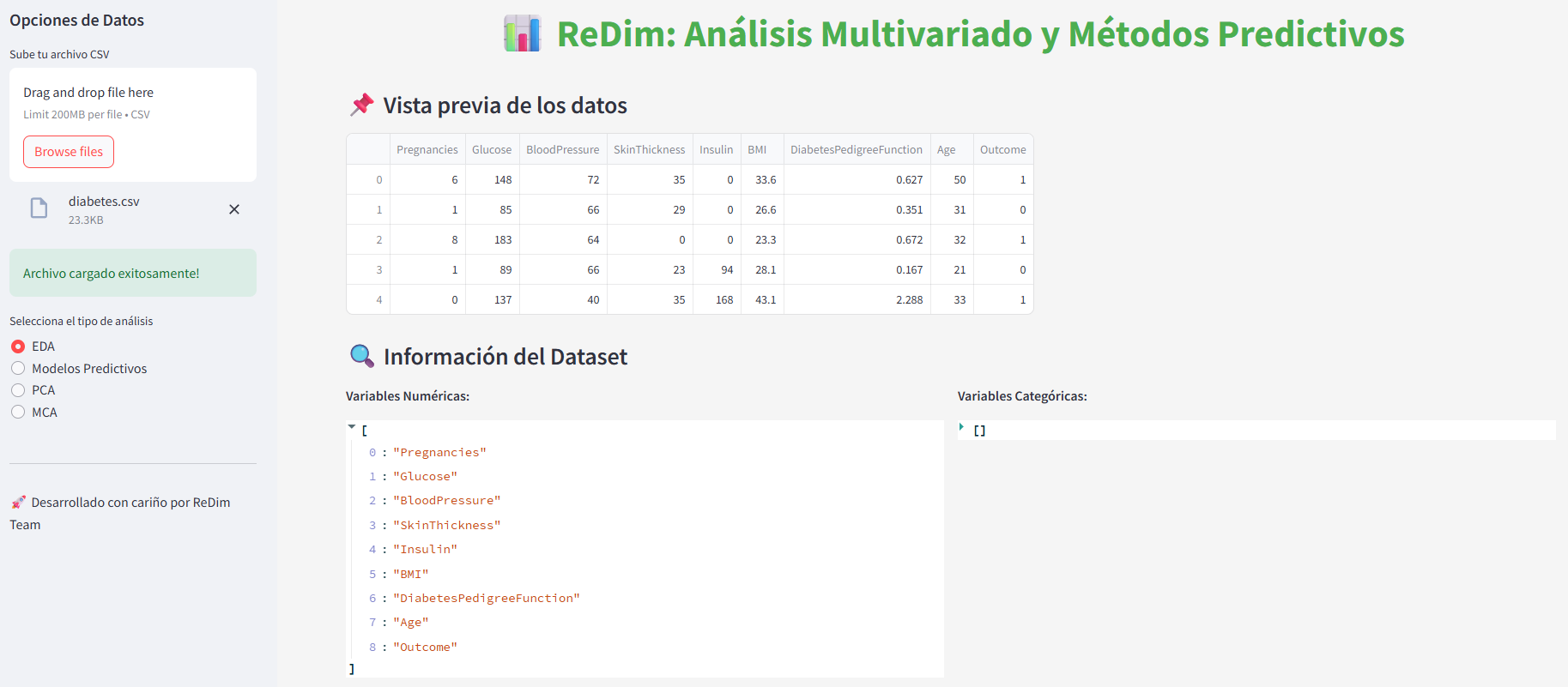
**6. Resultados**

**6.1. Inicio de la aplicación:**

En el **inicio de la aplicación se** presenta un panel interactivo desarrollado con Streamlit, una herramienta web que permite crear aplicaciones interactivas para análisis de datos de forma sencilla y eficiente. La interfaz permite cargar y explorar bases de datos, así como realizar análisis complejos como PCA y MCA. El diseño es amigable para usuarios que no necesariamente tengan conocimientos avanzados en programación, lo que hace que esta herramienta sea accesible para una amplia gama de usuarios, desde analistas de datos hasta estudiantes y profesionales en el campo de la estadística y la ciencia de datos.



**6.2. Carga de la base de datos en formato CSV.**



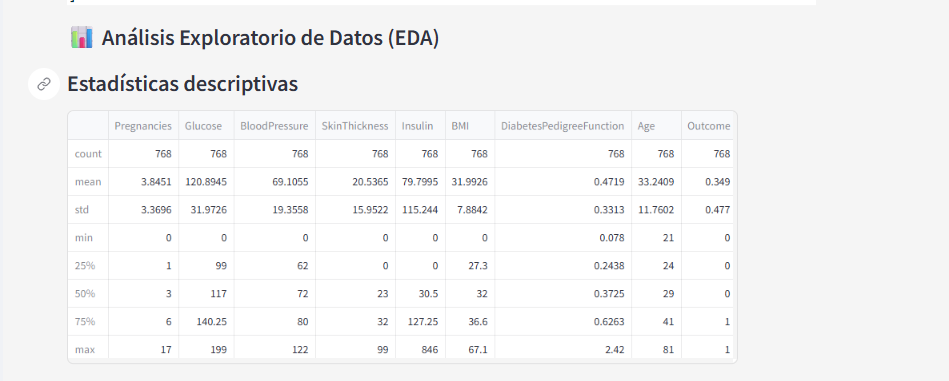
Uno de los aspectos fundamentales de la aplicación es la capacidad de cargar datos en formato **CSV**. En este caso, la base de datos utilizada es la diabetes, obtenida de **Kaggle**, la cual es una de las plataformas más populares para datasets públicos.

La base de datos contiene variables únicamente numéricas, lo que permite realizar análisis multivariados que incluyen características continuas. En particular, las variables numéricas pueden incluir características como el BMI (índice de masa corporal), edad y glucosa.

**6.3. Análisis exploratorio de datos:**

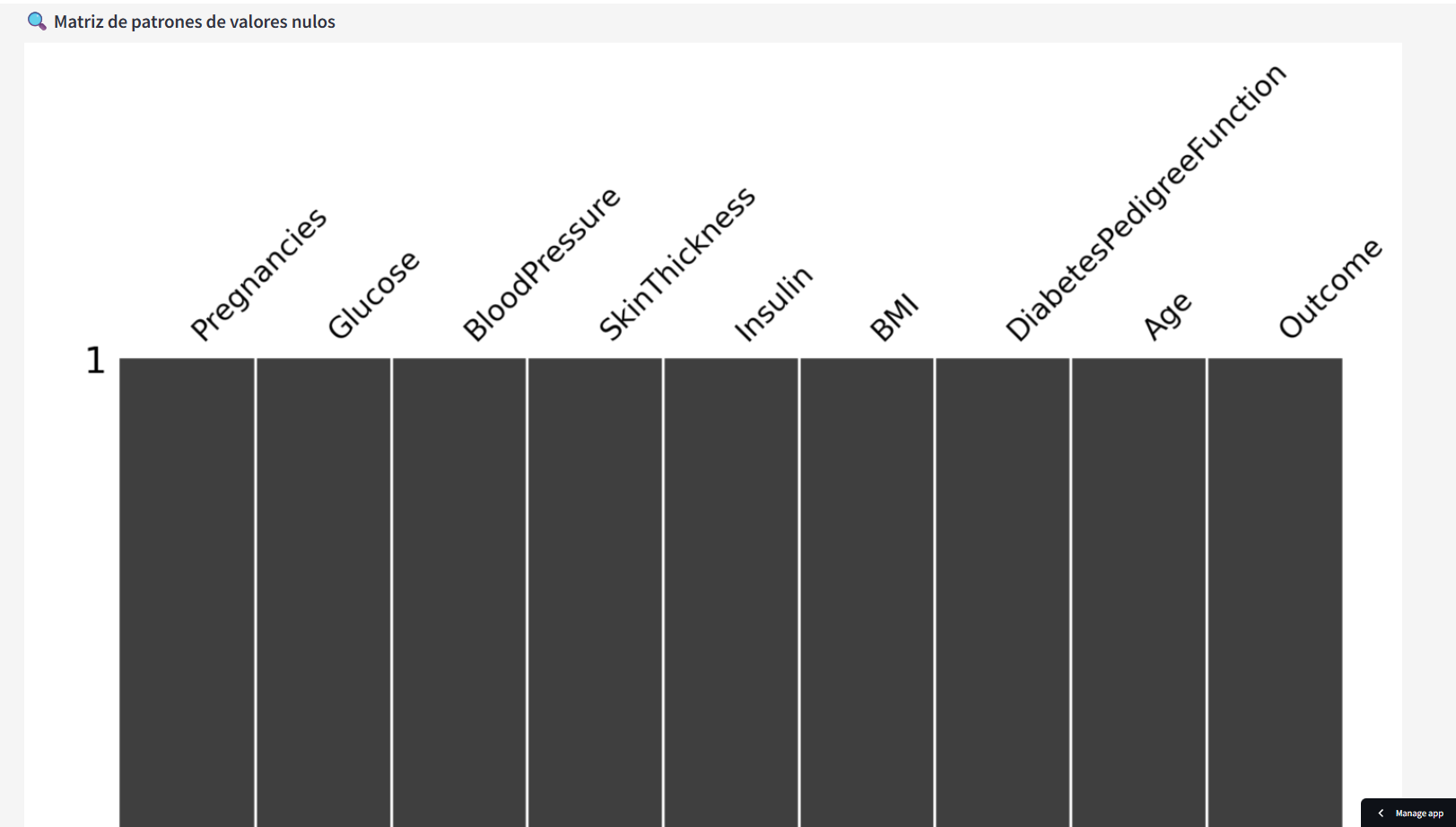
En la sección de Análisis Exploratorio de Datos (EDA), la aplicación ofrece una serie de herramientas y gráficos que permiten entender la naturaleza y estructura del conjunto de datos cargado. Aquí se incluyen:

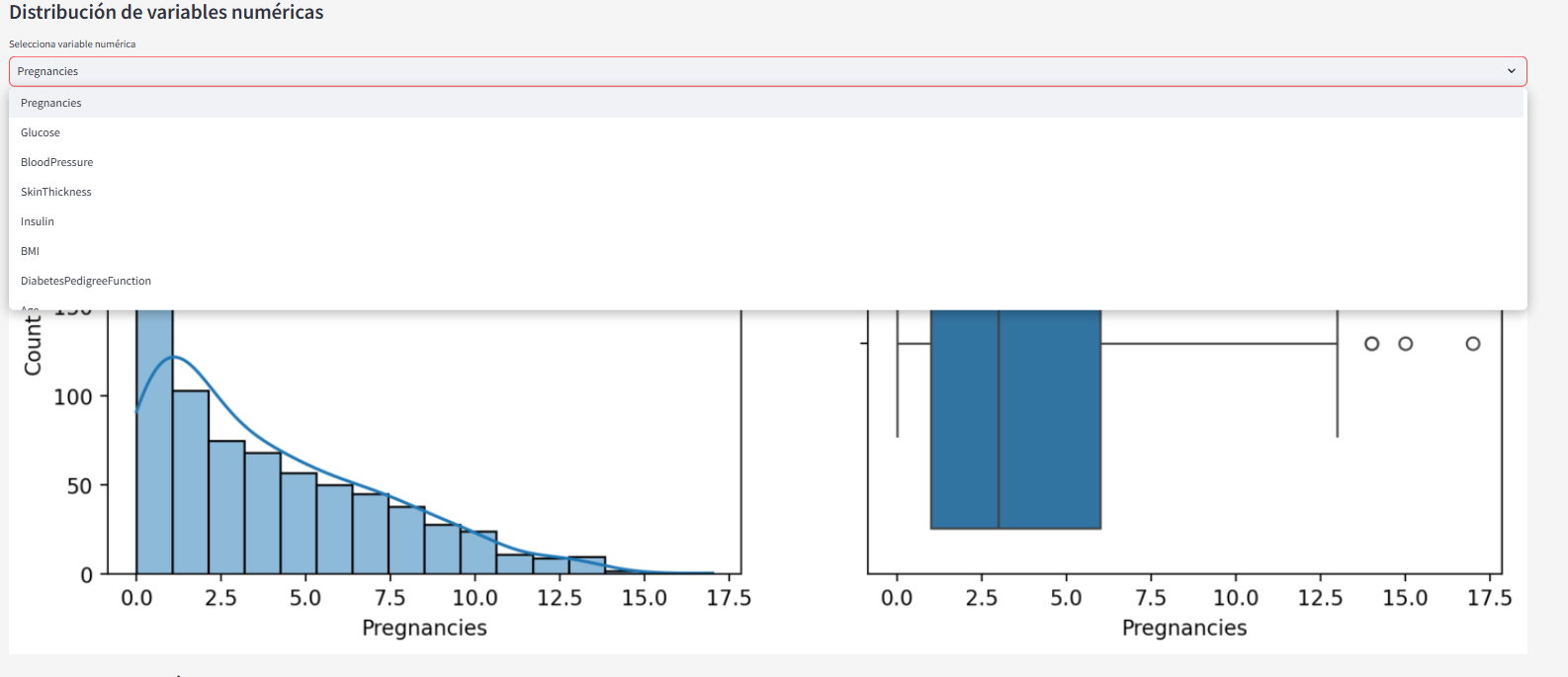
* Estadísticas descriptivas: Se genera una tabla que muestra resúmenes estadísticos de las variables numéricas del dataset, como la media, mediana, desviación estándar, máximos y mínimos, entre otros. Estos valores proporcionan una visión general de la distribución de cada variable, permitiendo identificar posibles valores atípicos o errores en los datos.



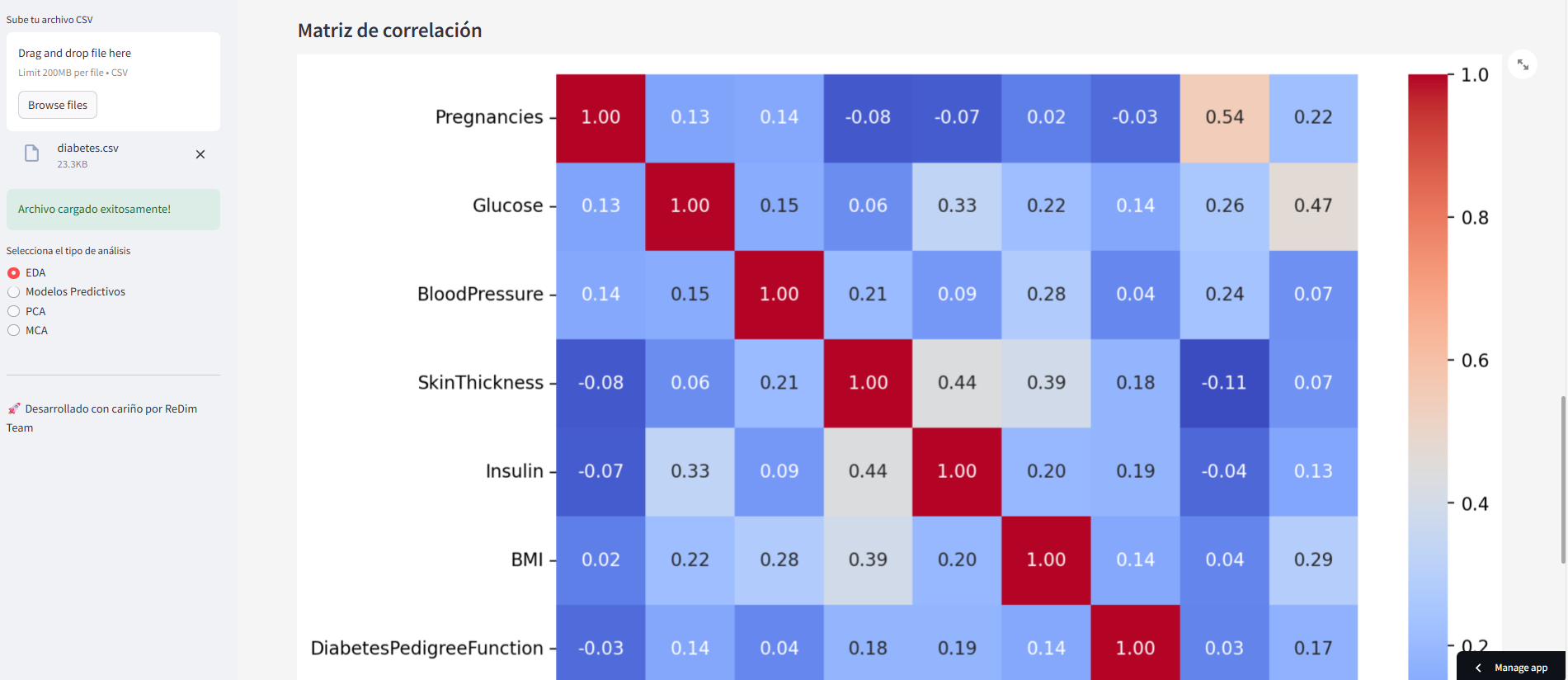
* Matriz de patrones de valores nulos: Este gráfico, proporcionado por la librería missingno, ayuda a visualizar la distribución de los valores nulos en las columnas del dataset. La ausencia de datos puede influir en la calidad de los análisis y la capacidad de los modelos predictivos, por lo que es crucial identificar dónde ocurren los valores faltantes.

La matriz visualiza la ausencia de datos de una manera gráfica, lo que facilita la toma de decisiones sobre cómo manejar estos valores, ya sea mediante imputación (reemplazo de los valores faltantes) o eliminación de las filas/columnas con demasiados valores faltantes.



* **Distribución de variables numéricas:** Aquí la aplicación permite al usuario elegir una variable numérica para visualizar su **distribución** mediante un **histograma** y un **boxplot**. El histograma muestra la frecuencia de los valores en intervalos, ayudando a identificar la forma de la distribución (por ejemplo, si es normal, sesgada, etc.). El **boxplot**, por otro lado, es útil para visualizar la **distribución** de los datos, la **mediana**, los **cuartiles** y los **valores atípicos**.  
  
* Matriz de correlación: La matriz de correlación muestra cómo las variables numéricas están relacionadas entre sí, calculando los coeficientes de correlación de Pearson. Un valor cercano a +1 o -1 indica una relación fuerte, mientras que valores cercanos a 0 sugieren una relación débil o inexistente.

La matriz de correlación es crucial para identificar posibles colinealidades entre variables. En problemas de regresión, las altas correlaciones entre variables predictoras pueden afectar la estabilidad y precisión del modelo. Esta herramienta permite que el analista identifique variables altamente correlacionadas y decida si es necesario eliminar o combinar algunas de ellas.

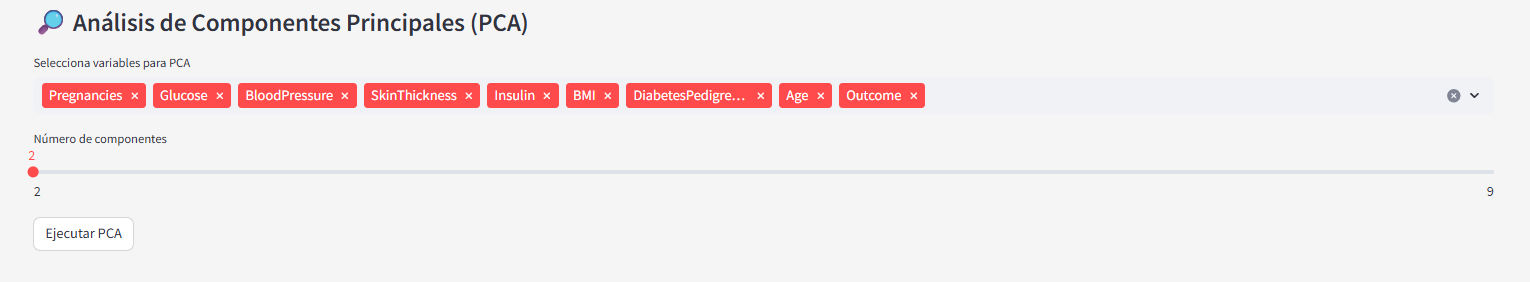


**6.3. PCA:**

El PCA es una técnica de reducción de dimensionalidad que se utiliza para reducir la cantidad de variables en un conjunto de datos mientras se conserva la mayor cantidad posible de la información original. En la app, el PCA se implementa de la siguiente manera:

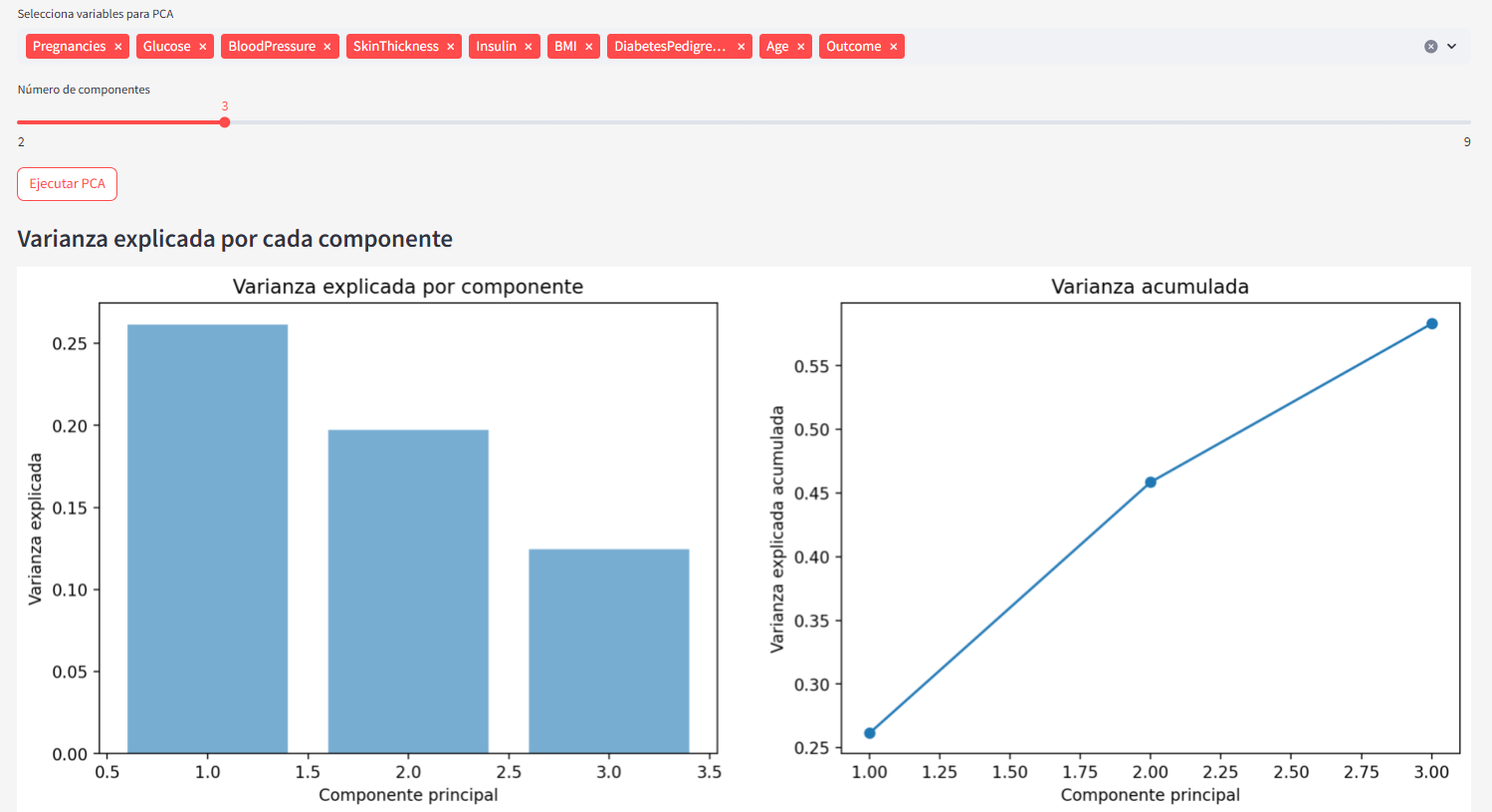
 Selección de variables: El usuario puede seleccionar las variables que desea incluir en el análisis. Solo las variables numéricas son relevantes para el PCA, ya que la técnica no se aplica directamente a variables categóricas.

 Número de componentes: El número de componentes principales a generar se puede elegir mediante un deslizador. Los componentes principales son combinaciones lineales de las variables originales que buscan maximizar la varianza explicada.

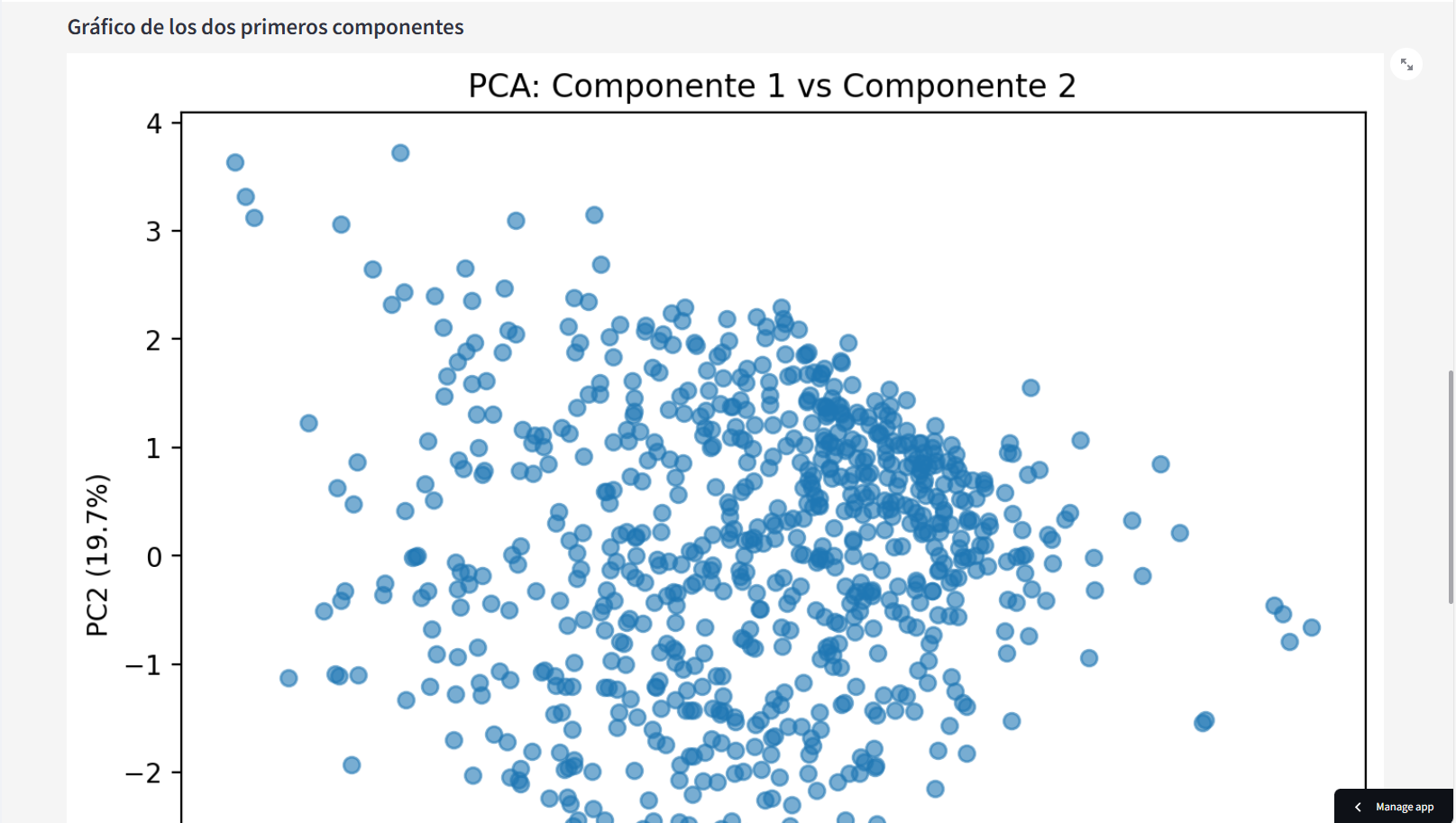


 Gráfico de varianza explicada: Este gráfico muestra la varianza explicada por cada uno de los componentes principales. El PCA ordena los componentes por la cantidad de varianza que explican, por lo que los primeros componentes retienen la mayor parte de la información. Este gráfico ayuda a determinar cuántos componentes son necesarios para capturar la mayor parte de la variabilidad del conjunto de datos.

 Gráfico de varianza acumulada: Este gráfico muestra la varianza acumulada a medida que se agregan componentes adicionales. Esto es útil para determinar el número mínimo de componentes que explican una proporción significativa de la variabilidad en los datos (por ejemplo, un 95%).



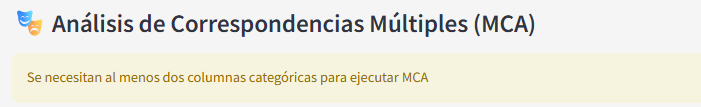
* Gráfico de componentes principales: Este gráfico proyecta los datos en los dos primeros componentes principales. Es una forma de visualizar los datos en un espacio bidimensional, lo que puede ayudar a identificar agrupamientos o patrones en los datos.

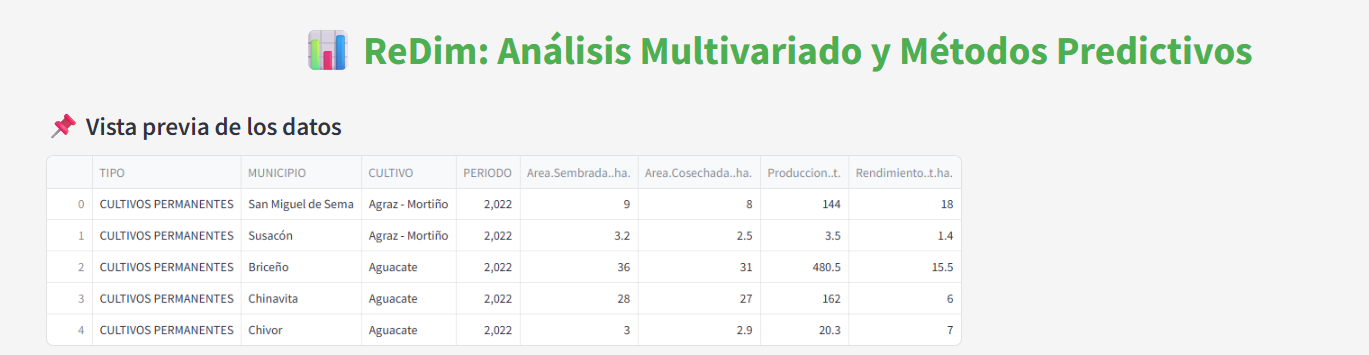


* Cargas factoriales: Las cargas factoriales muestran cómo cada variable original contribuye a cada componente principal. Las altas cargas indican que una variable tiene un impacto significativo en ese componente. Este análisis es útil para interpretar qué variables son las más importantes en los componentes principales y para comprender la estructura subyacente de los datos.



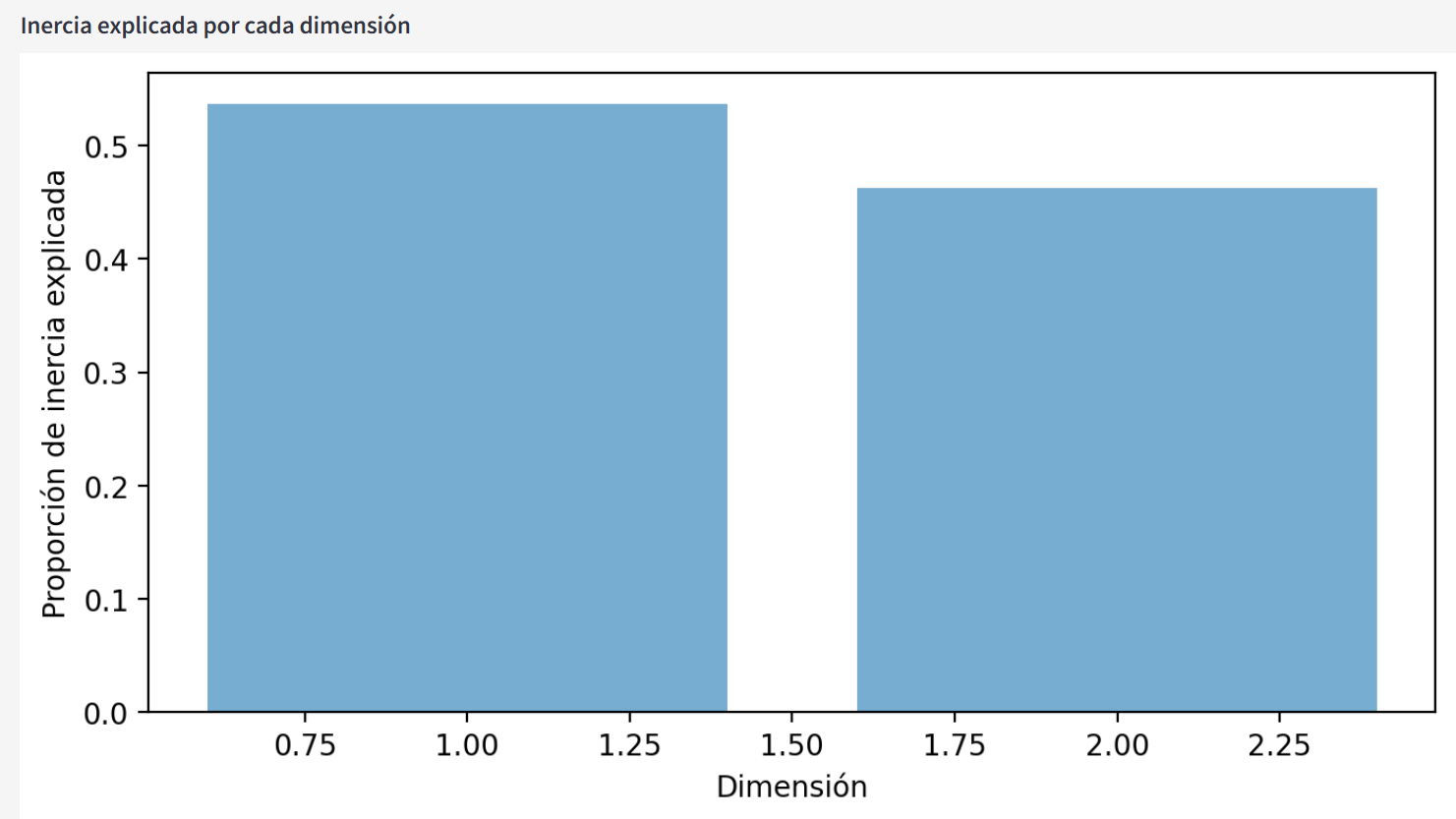
**6.3. MCA**

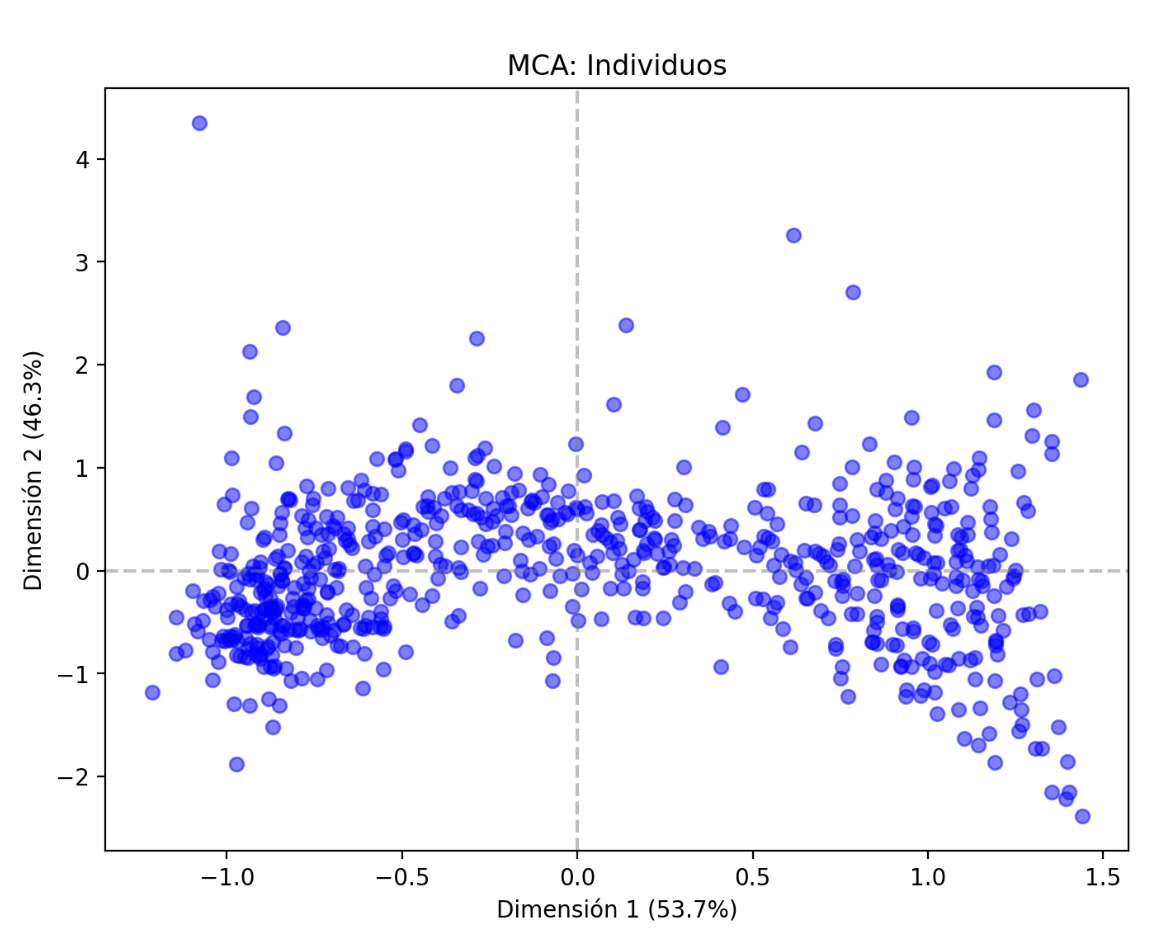
Como para el MCA se necesitan variables categóricas, usamos para probar su funcionamiento una base de datos limpia de las cosechas en los municipios de Boyacá que obtuvimos de datos abiertos Colombia. ****

****

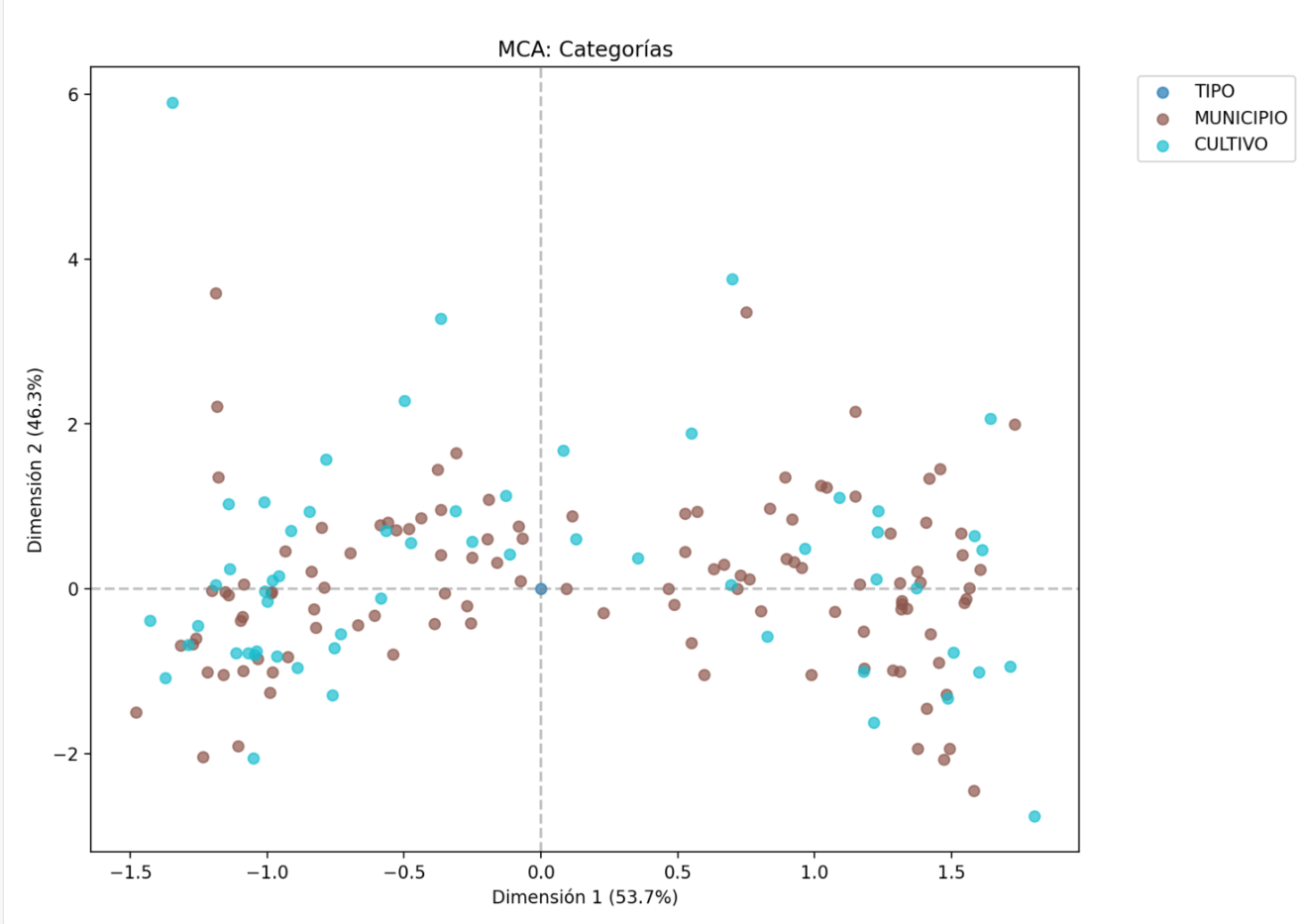
* **Selección de variables:** El usuario puede elegir las variables categóricas que desea incluir en el análisis. Al igual que con el PCA, es importante seleccionar variables relevantes para obtener un análisis útil.

****

* Gráfico de inercia explicada: Este gráfico muestra la proporción de la inercia (información) explicada por cada dimensión. Similar al PCA, el objetivo es identificar cuántas dimensiones son necesarias para retener la mayor cantidad de información posible. ****
* **Gráfico de individuos:** Este gráfico muestra la **distribución de los individuos** en las dos primeras dimensiones del MCA. Los puntos en este gráfico representan las observaciones del dataset, y la distancia entre los puntos indica la relación entre ellos en el espacio de las dimensiones reducidas.

****

* Gráfico de categorías: Este gráfico muestra la distribución de las categorías en las dos primeras dimensiones del MCA. Al igual que con los individuos, las categorías que están cerca en el gráfico están relacionadas entre sí.

****

**6.5. Métodos predictivos:**

**6.5.1 Regresión Lineal: Predicción del precio de viviendas**

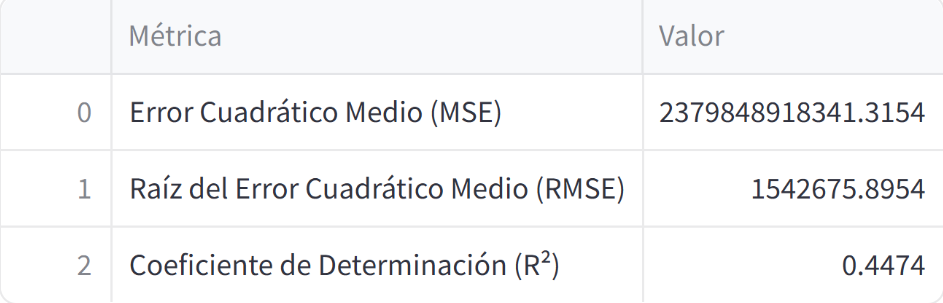
Para el modelo de regresión lineal se utilizó la base de datos Housing, con la variable price como variable dependiente y las variables área, bedrooms y bathrooms como predictoras.

Desde la interfaz de ReDim es posible seleccionar las variables mediante menús desplegables. Además, la aplicación permite ajustar el porcentaje de los datos a utilizar en el conjunto de entrenamiento, a través de un deslizador interactivo; el usuario puede modificar este parámetro según sus necesidades.

Interfaz de usuario gráfica, Aplicación

El contenido generado por IA puede ser incorrecto.

Luego del entrenamiento, se evaluó el modelo con las siguientes métricas:



Esto indica que el modelo explica el 44.7% de la variabilidad en el precio.

Finalmente, se puede observar el desempeño del modelo al comparar los valores predichos frente a los reales.

Gráfico, Gráfico de dispersión

El contenido generado por IA puede ser incorrecto.

**6.5.2 Regresión Logística**

Para este modelo se empleó la base de datos diabetes, cuyo objetivo es predecir la presencia de diabetes (1) o no (0) en pacientes, utilizando variables como BMI, Age, Insulin y Glucose. La variable respuesta fue Outcome, la cual es dicotómica.

Interfaz de usuario gráfica, Texto, Aplicación

El contenido generado por IA puede ser incorrecto.

Entrenando al modelo con el 70% de los datos, y 30% se destinó para prueba, las métricas del modelo son:

Interfaz de usuario gráfica, Aplicación

El contenido generado por IA puede ser incorrecto.

**Gráfico, Gráfico de rectángulos

El contenido generado por IA puede ser incorrecto.**

Finalmente, se generó la curva ROC, que representa la sensibilidad frente a la tasa de falsos positivos, y el área bajo la curva AUC, esta visualización es clave para evaluar la capacidad discriminativa del modelo, siendo una AUC cercana a 1 ideal y una AUC cerca de 0.5 lo mismo que un modelo aleatorio.

Gráfico, Gráfico de líneas

El contenido generado por IA puede ser incorrecto.

**6.5.3 LDA**

Para todos los modelos de clasificación vamos a utilizar diabetes y poder mostrar que hace ReDim, por eso se seleccionan las mismas variables y después comparar modelos.

Interfaz de usuario gráfica, Aplicación, Teams

El contenido generado por IA puede ser incorrecto.

Entonces, para evaluar al modelo LDA se entregan las siguientes métricas:

Interfaz de usuario gráfica, Aplicación

El contenido generado por IA puede ser incorrecto.

Matriz de confusión

Gráfico, Gráfico de rectángulos

El contenido generado por IA puede ser incorrecto.

Curva ROC

Gráfico, Gráfico de líneas

El contenido generado por IA puede ser incorrecto.

Se puede evaluar también con la curva ROC mediante la AUC, en este caso es de 0.77.

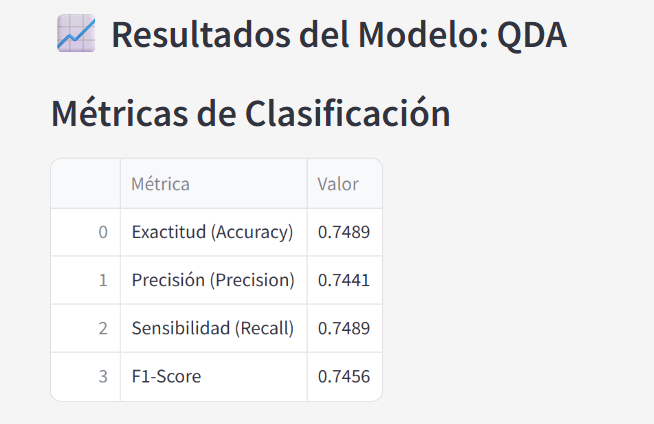
**6.5.4 QDA**

En ReDim se puede hacer modelos QDA, lo haremos también para Outcome con variables predictoras BMI, Age, Insulin, y Glucose. Se puede seleccionar el porcentaje de datos que quiero para test y una semilla.

Interfaz de usuario gráfica, Texto, Aplicación

El contenido generado por IA puede ser incorrecto.

Se entregan métricas para evaluar el modelo QDA



Matriz de confusión

Gráfico, Gráfico de rectángulos

El contenido generado por IA puede ser incorrecto.

Gráfico, Gráfico de líneas

El contenido generado por IA puede ser incorrecto.

**6.5.5 Árbol de decisión (Clasificación)**

Para mostrar el modelo de árbol de decisión en la parte de clasificación se hace lo mismo, selecciona la variable de respuesta y las predictoras, además se selecciona la profundidad máxima del árbol y el mínimo de muestras para dividir.

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

El contenido generado por IA puede ser incorrecto.

Patrón de fondo

El contenido generado por IA puede ser incorrecto.

Se hace una visualización de como se toman las decisiones en el árbol.

Diagrama

El contenido generado por IA puede ser incorrecto.

**6.5.6 Árbol de decisión (regresión)**

Para este modelo vamos a mostrarlo usando la base de datos Housing para usar como variable de respuesta a Price y predictoras a bedrooms, bathrooms y área.

Interfaz de usuario gráfica, Texto, Aplicación

El contenido generado por IA puede ser incorrecto.

Interfaz de usuario gráfica, Aplicación

El contenido generado por IA puede ser incorrecto.

Luego de seleccionar la profundidad y las muestras, se selecciona Entrenar Modelo, posteriormente se entregan las métricas para evaluar al modelo.

Interfaz de usuario gráfica, Aplicación

El contenido generado por IA puede ser incorrecto.

Y un diagrama para ver los valores reales vs los predichos

Gráfico, Gráfico de dispersión

El contenido generado por IA puede ser incorrecto.

Se entrega una visualización del árbol

Diagrama

El contenido generado por IA puede ser incorrecto.

**6.5.7 Random Forest (clasificación)**

En este modelo también se selecciona la profundidad máxima y el mínimo de muestras, también cuántos árboles.

**Interfaz de usuario gráfica, Texto, Aplicación

El contenido generado por IA puede ser incorrecto.**

**Patrón de fondo

El contenido generado por IA puede ser incorrecto.**

**6.5.8 Random Forest (regresión)**

Con ReDim podemos hacer modelos random forest también para variables numéricas

**Interfaz de usuario gráfica, Texto, Aplicación

El contenido generado por IA puede ser incorrecto.**

**Interfaz de usuario gráfica, Aplicación

El contenido generado por IA puede ser incorrecto.**

Además, se muestran métricas para evaluar el modelo

**Interfaz de usuario gráfica, Aplicación

El contenido generado por IA puede ser incorrecto.**

Y un diagrama para representar los valores reales contra los predichos.

**Gráfico, Gráfico de dispersión

El contenido generado por IA puede ser incorrecto.**

**6.6 Comparación modelos**

Para este apartado debemos seleccionar cuáles modelos queremos comparar, teniendo en cuenta los que habíamos creado podemos ver las diferencias por ejemplo entre modelos de clasificación como LDA, QDA y Regresión Logística.

**Interfaz de usuario gráfica, Aplicación

El contenido generado por IA puede ser incorrecto.**

Lo que hace ReDim es resaltar el mayor valor de cada métrica para que el usuario pueda interpretar mejor y seleccionar su modelo de acuerdo a estas métricas.

**Tabla

El contenido generado por IA puede ser incorrecto.**

Otra métrica de evaluación que se puede tener en cuenta es la AUC, por eso también se hace la comparación de las curvas ROC

**Gráfico, Gráfico de líneas

El contenido generado por IA puede ser incorrecto.**

Además, podemos comparar modelos de regresión, solo debemos seleccionarlos.

Interfaz de usuario gráfica, Aplicación

El contenido generado por IA puede ser incorrecto.

Se marca con amarillo el valor más alto, en este caso sería el error más alto.

Interfaz de usuario gráfica, Aplicación, Tabla

El contenido generado por IA puede ser incorrecto.

**6.7 Predicción manual**

En esta parte con ReDim se puede predecir manualmente con modelos como el logístico y el de regresión lineal. Lo que se hace es reemplazar el valor de cada una de las variables predictoras y luego se entrega el resultado que da ese modelo.

Predicción usando el modelo de regresión lineal con Price como variable de respuesta y área, bedrooms, bathrooms como variables predictoras:

Interfaz de usuario gráfica, Aplicación

El contenido generado por IA puede ser incorrecto.

Predicción usando el modelo de regresión logística con la variable Outcome como variable de respuesta y como variables predictoras a BMI, Age, Insulin, y Glucose:

Interfaz de usuario gráfica, Aplicación

El contenido generado por IA puede ser incorrecto.

**7. Conclusión**

La aplicación **ReDim** proporciona una solución robusta y accesible para realizar análisis multivariado, aplicar modelos predictivos y reducir la dimensionalidad de los datos. Gracias a la integración de técnicas estadísticas avanzadas como **PCA** y **MCA**, junto con modelos de **regresión** y **clasificación**, los usuarios pueden obtener resultados significativos y hacer predicciones precisas en tiempo real. **ReDim** es una herramienta ideal para científicos de datos, investigadores y estudiantes que buscan una manera práctica y visual de aplicar métodos estadísticos y machine learning.

Además, la capacidad de realizar comparaciones entre modelos, junto con la opción de realizar **predicciones manuales** y de visualizar las **curvas ROC**, ofrece una visión integral de los datos y modelos. La aplicación se puede usar para diversos sectores como la salud, finanzas, marketing y más, haciendo que el análisis de datos y la toma de decisiones basadas en datos sea más accesible para cualquier usuario.