

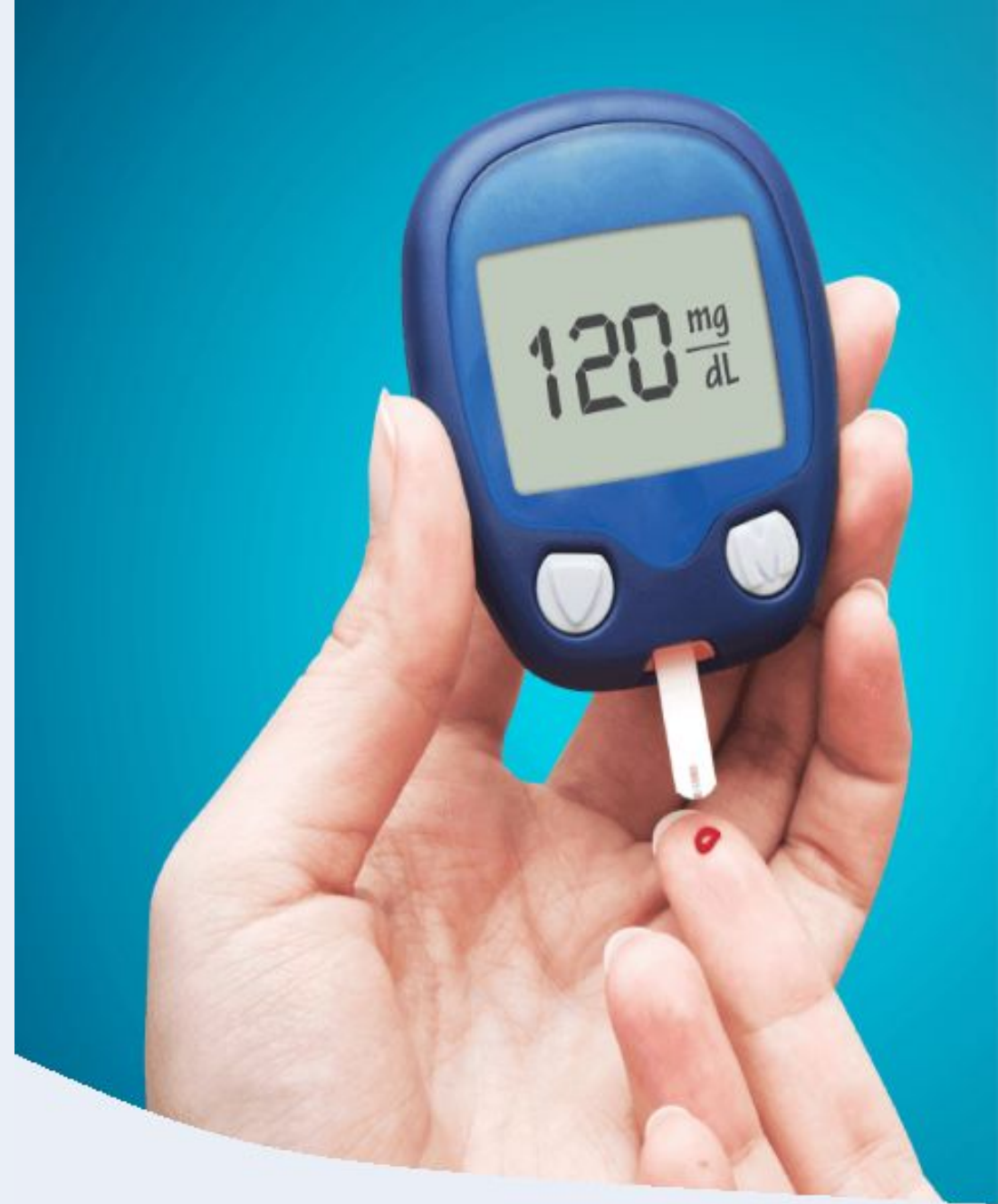
# Diabetes Mellitus en EE.UU. 2015

Factores de Riesgo y Protectores

**AUTORES:** Ángel Ciria  
Agustín Nahuel Quiroga Baigorri

# AGENDA

- **01** | Contexto
- **02** | Hipótesis/Preguntas de Interés
- **03** | Metadata
- **04** | Análisis Exploratorio
- **05** | Insights y Recomendaciones



# CONTEXTO Y AUDIENCIA

## Contexto Clínico

La diabetes mellitus es una enfermedad crónica grave por la cual las personas pierden la capacidad de regular eficazmente los niveles de glucosa en sangre y puede reducir tanto la calidad, como la esperanza de vida. El diagnóstico temprano conduce a cambios positivos en el estilo de vida y a un tratamiento más eficaz.

## Contexto Histórico

La prevalencia mundial de la diabetes mellitus ha aumentado drásticamente en los últimos 20 años. De 30 millones de casos en 1985 se ha pasado a 177 millones en el año 2000. Basándonos en las tendencias actuales, más de 360 millones de personas padecerán diabetes en el año 2030.

## Audiencia

Autoridades de salud pública como privada que a partir de lo expuesto en este estudio determinarán políticas para prevenir y combatir esta dolencia.

# PREGUNTAS DE INTERÉS

## Primarios

- ¿Se puede incidir el riesgo de padecer diabetes mellitus?
- ¿Existen factores de riesgo que desencadenan la enfermedad y hábitos saludables que ayuden a prevenirla?

## Secundarios

- ¿Cómo son los porcentajes de diabéticos en los distintos rangos etarios?
- ¿Qué información nos proporciona el estado nutricional?
- ¿Cómo es el estado de salud de las personas con diabetes?
- ¿Cómo se relaciona la diabetes con el colesterol y la hipertensión?
- ¿Cómo se distribuyen los ingresos según estatus de diabetes?
- ¿Qué hábitos se pueden considerar preventores de la enfermedad?

# RESUMEN METADATA



**Población: 320.738.994**



**Personas Encuestadas: 253.680**



**Preguntas: 22**

**Nominal: 15**

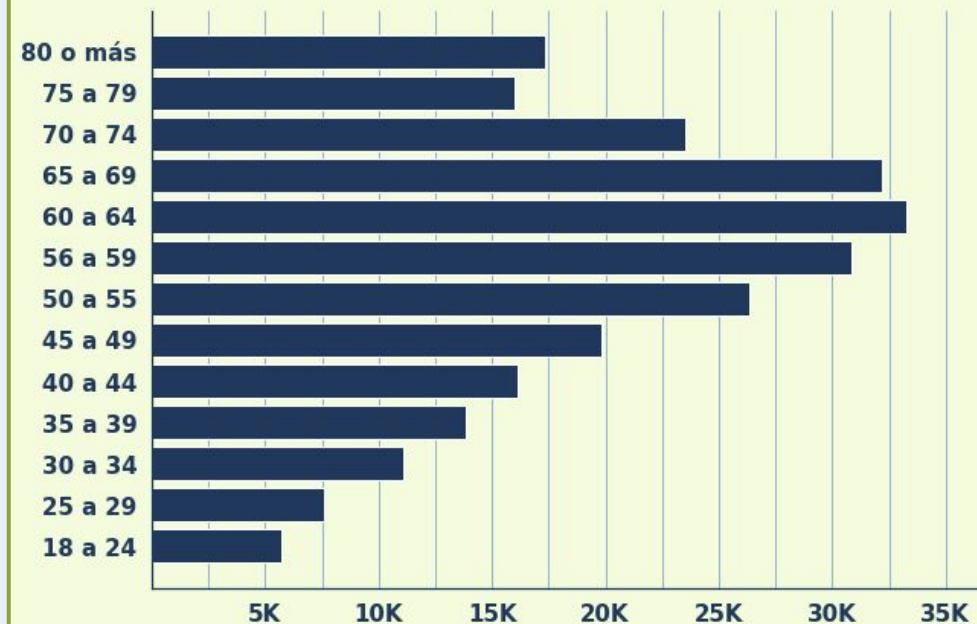
**Intervalo: 3**

**Ordinal: 1**

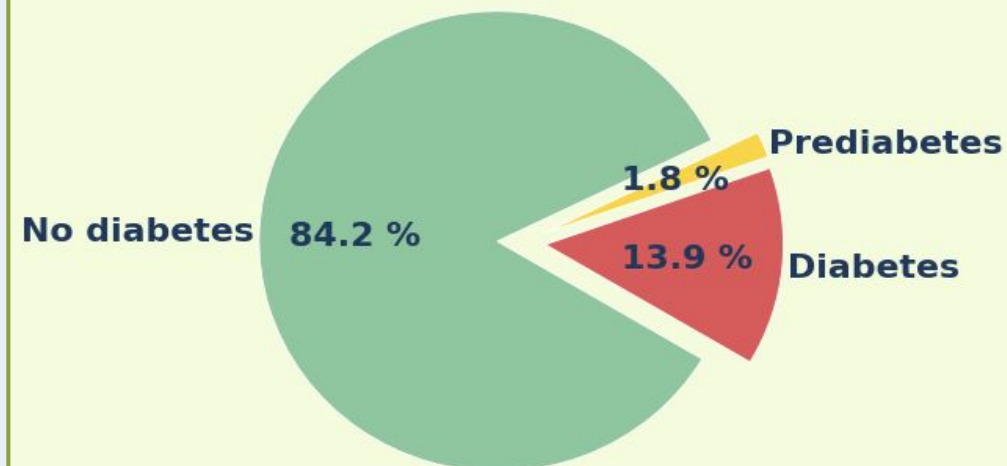
**Razón: 3**



## Distribución Etaria Encuestados



## Estatus de Diabetes



# ANÁLISIS **EXPLORATORIO**

# DIABETES & RANGOS ETARIOS

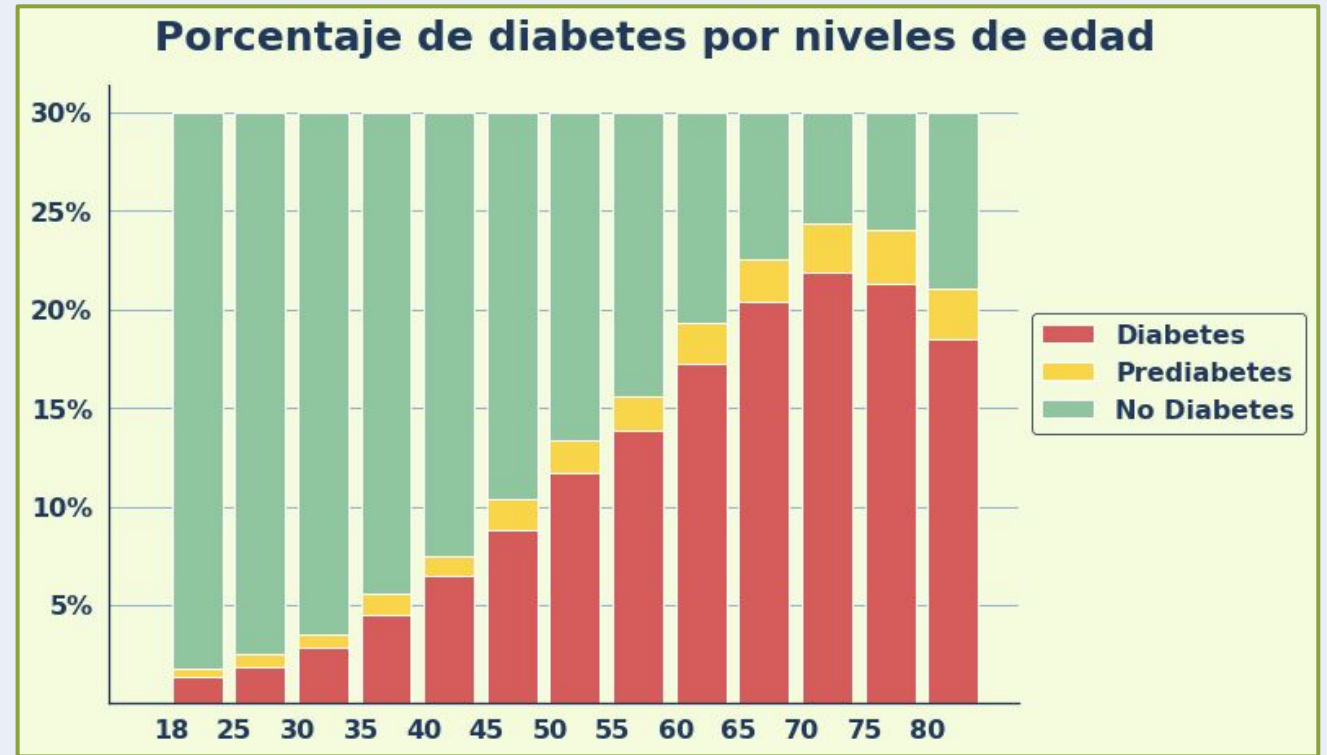
Se observa que los porcentajes de personas que padecen **prediabetes** y **diabetes** crecen al aumentar la edad.

El rango etario con mayor porcentaje es el de:

- **70 a 74 años** con un **24,41%** de prevalencia.

Luego comienza a disminuir. Esta disminución es debido a que el próximo rango etario es el de expectativa de vida: **76 años** para **hombres** y **81 años** para **mujeres**.

Las **personas con diabetes** poseen una **expectativa de vida menor** a la mayoría



55%



Expectativa de **VIDA**



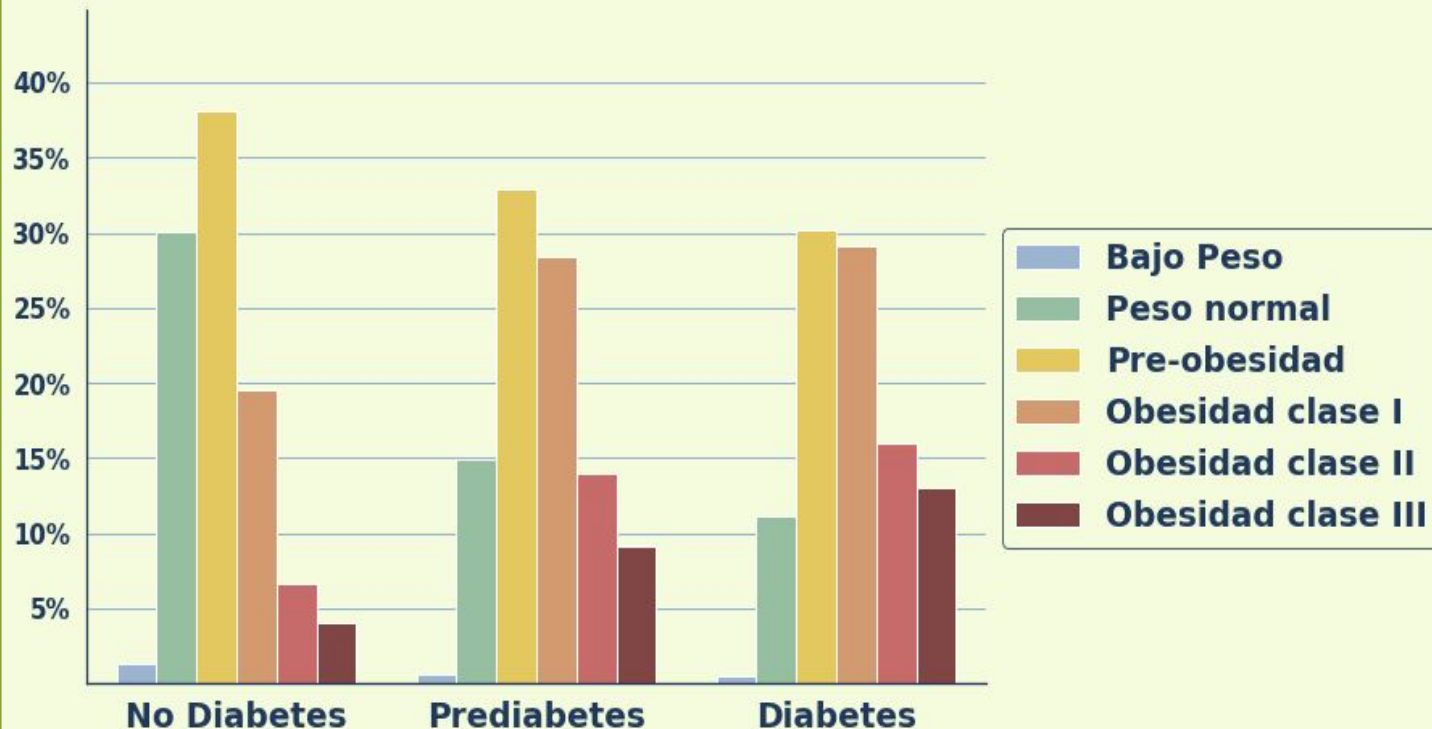
**76.3 años**



**80.8 años**

# DIABETES & ESTADO NUTRICIONAL

Estado Nutricional por Estatus de Diabetes



Del análisis se puede concluir que el porcentaje de personas con:

- **Bajo peso**
- **Peso normal**
- **Pre-obesidad**

**Disminuyen** considerablemente en el grupo con **Prediabetes** y **Diabetes**.

**Aumentando** notoriamente en estos grupos:

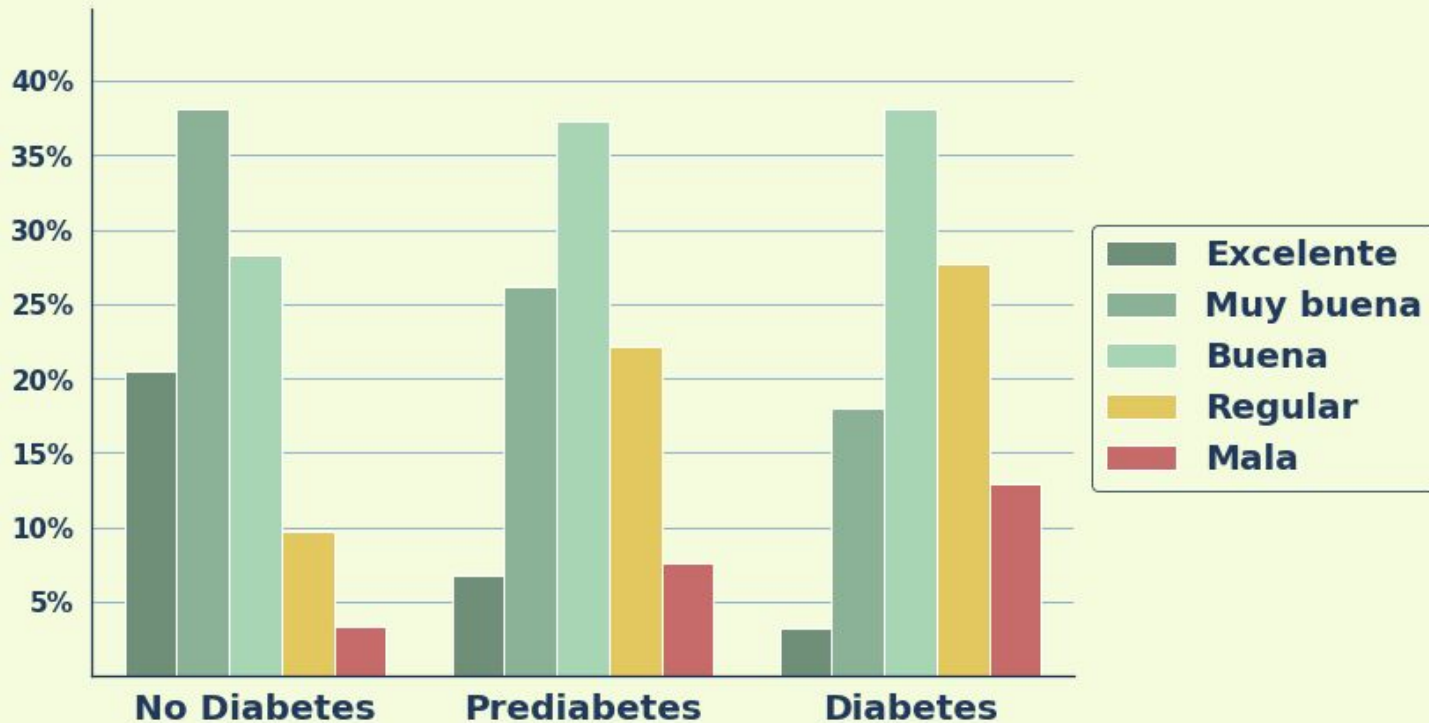
- **Obesidades clase II**
- **Obesidad clase III**

Claramente la **obesidad** puede considerarse un **factor de riesgo** para esta enfermedad.



# DIABETES & ESTADO DE SALUD

Estado de Salud por Estatus de Diabetes



Del análisis se puede concluir que el porcentaje de personas con un estado de salud :

- **Excelente**
- **Muy buena**

**Disminuyen** considerablemente en el grupo con **Prediabetes** y **Diabetes**.

**Aumentando** notoriamente en estos grupos:

- **Regular**
- **Malo**

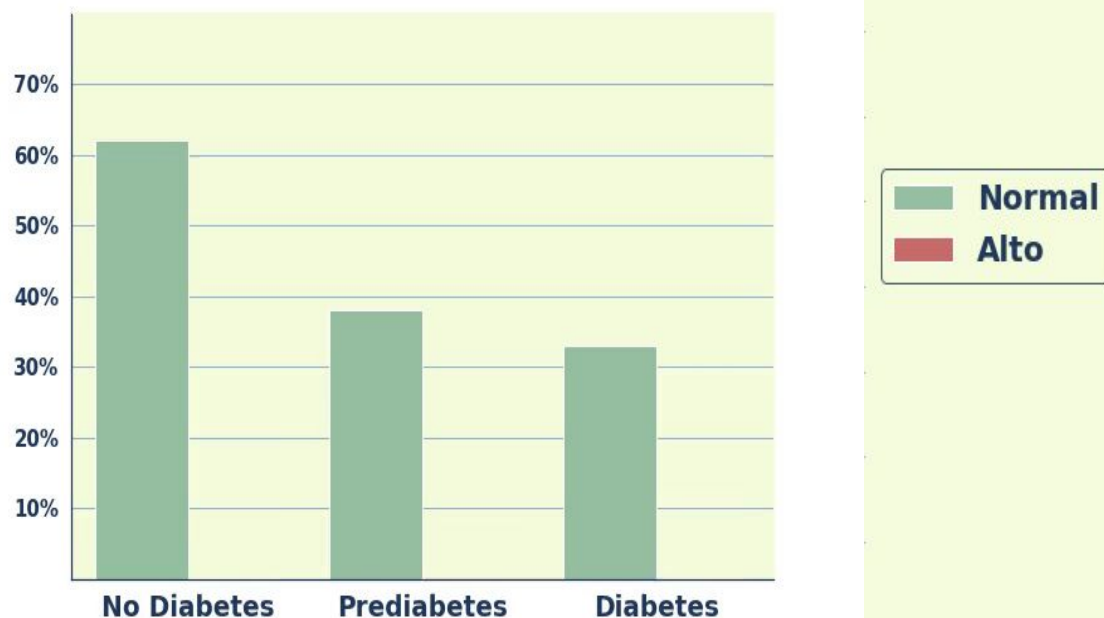
Claramente tener **diabetes** impacta **negativamente** en el **estado de salud** general de las personas.

# DIABETES & COLESTEROL HIPERTENSIÓN

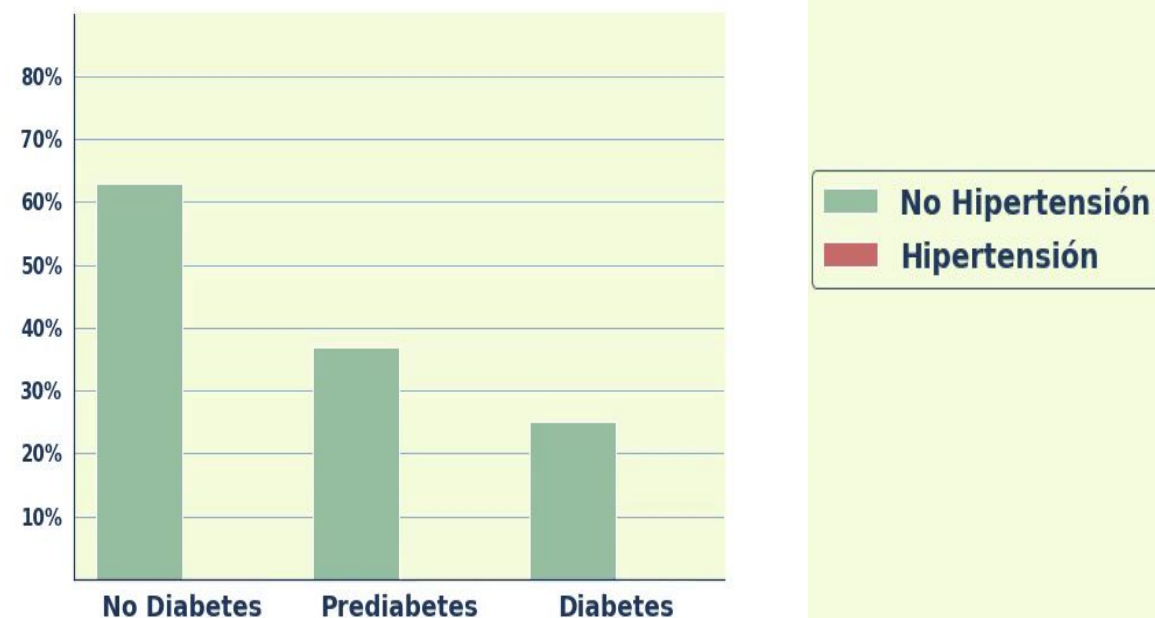
De analizar estas dos variables resulta que las personas que **no** padecen **diabetes** poseen mayoritariamente **niveles de colesterol** y **tensión normales**.

Mientras que las personas con algún tipo de **diabetes** poseen mayoritariamente **niveles altos** de **colesterol** e **hipertensión**.

## Nivel de Colesterol y Estatus de Diabetes



## Nivel de Hipertensión y Estatus de Diabetes



# DIABETES & HÁBITOS SALUDABLES

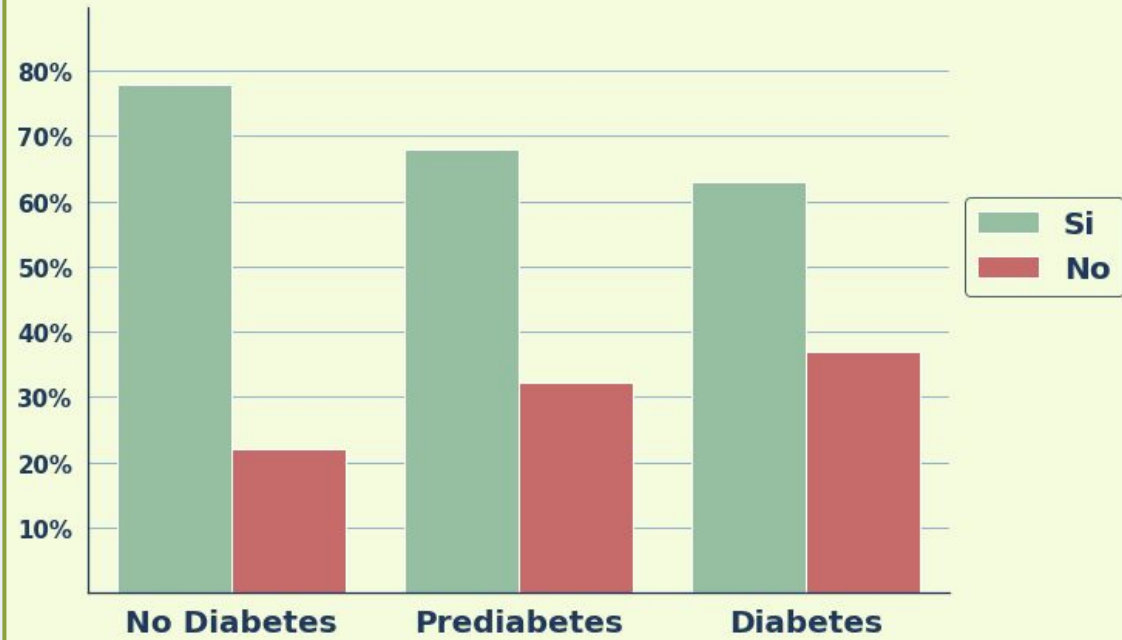
Del **total de las muestras** en **promedio** viven en hogares donde los ingresos anuales son **iguales o menores** a **\$50.000 dólares**.

De los **3 hábitos** saludables **ninguno** posee una **gran significancia** en el estatus de diabetes.

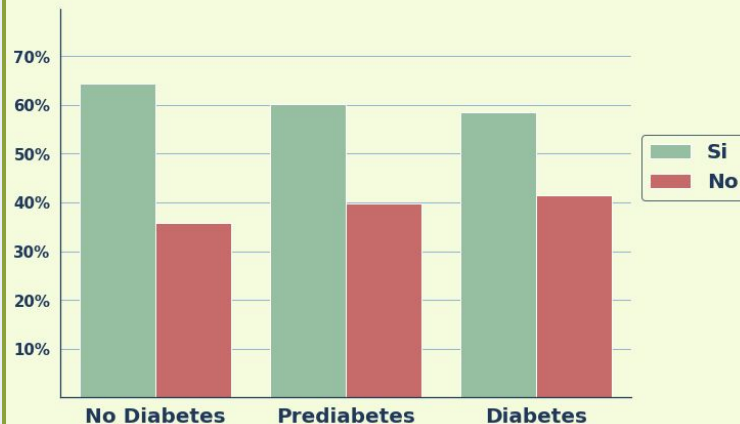
El factor en el que **mayor diferencia** se obtuvo entre las personas con diabetes y sin fue la **actividad física**.

- **78% sin diabetes.**
- **63% con diabetes.**

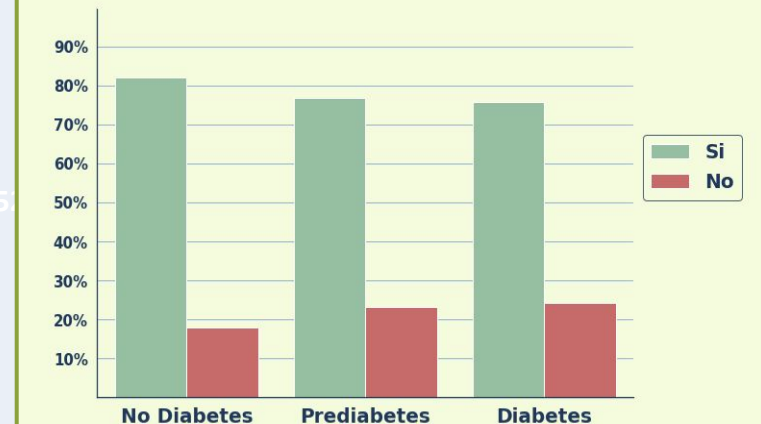
## Actividad Física y Estatus de Diabetes



## Consumo de Frutas y Estatus de Diabetes



## Consumo de Vegetales y Estatus de Diabetes



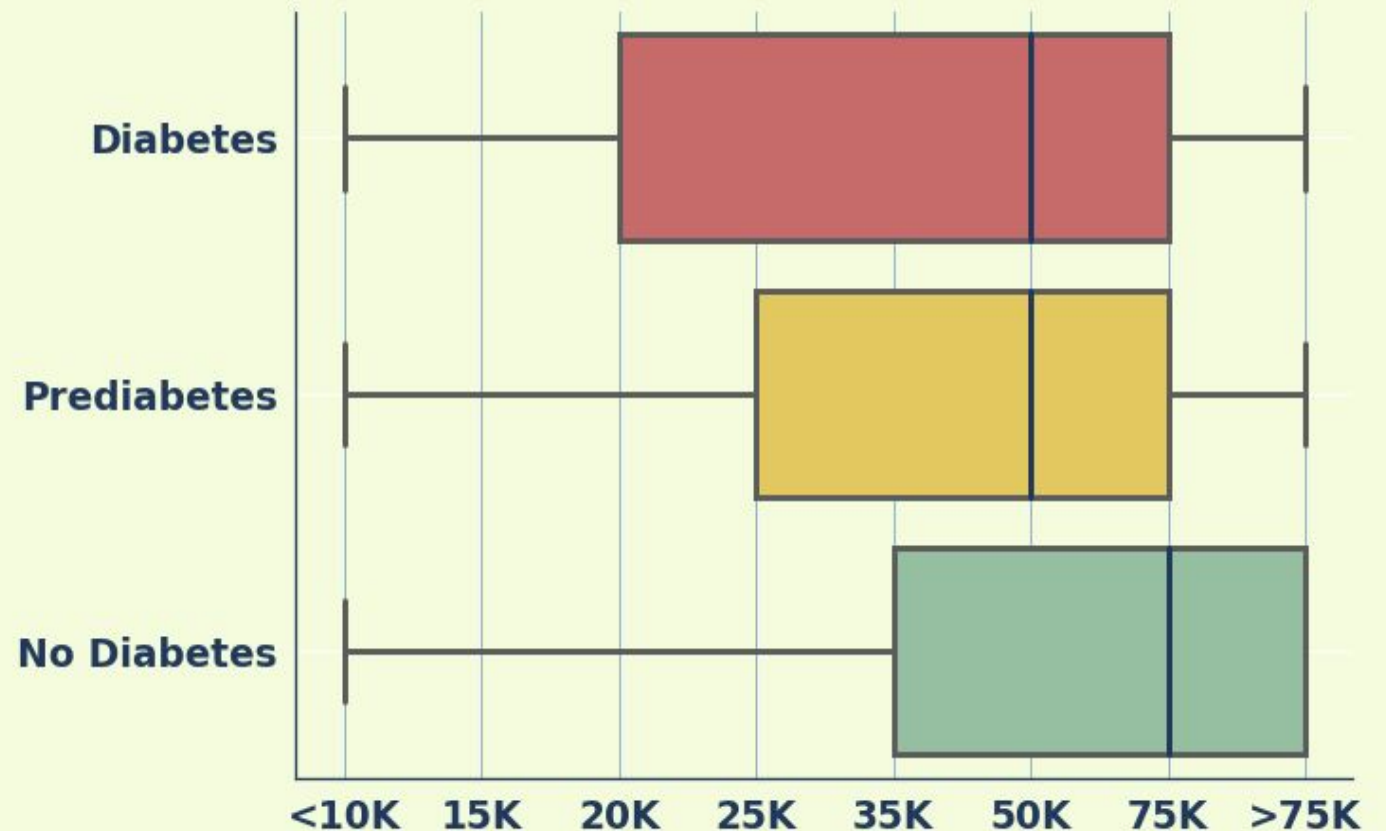
# DIABETES & NIVEL DE INGRESOS

Del **total** de las **muestras** en **promedio** viven en hogares donde los ingresos anuales son **iguales** o **menores** a **\$50.000 dólares**.

De los tres estatus de diabetes, las personas con **diabetes** y **prediabetes** viven en hogares con **menores ingresos** que las personas **sin diabetes**. Lo que indica que es una patología que tiene **mayor incidencia** en **poblaciones vulnerables**.

Se observa un gran sesgo a la izquierda en la población que sufre **diabetes**, el **25%** viven en hogares con **ingresos menores** a **\$20.000 dólares**.

## Estatus de Diabetes y Nivel de Ingresos [USD]



# APLICACIÓN DE **ALGORITMOS DE ML**

# EVALUACIÓN DE MODELOS

Se probaron los modelos:

- **DecisionTree**
- **Random Forest**
- **LightGBM**
- **XGBoost**
- **KNN**
- **Bagging Classifier**

En el dataset **general** y dividiendo las muestras en 3 rangos etarios: **menores**, **intermedios** y **mayores**, en todos los casos las muestras se encontraban balanceadas.

## GENERAL

Modelo/Métrica	Accuracy	Precisión	Recall	F1 Score	Tiempo de inferencia
Random Forest	0.929	0.931	<b>0.929</b>	0.929	9.821
Bagging Classifier	0.90	0.906	0.905	0.904	1.22
KNN	0.873	0.888	0.873	0.867	1357.61
Decision Tree	0.864	0.863	0.863	0.863	0.113
LightGBM	0.777	0.772	0.777	0.771	4.612
XGBoost	0.714	0.709	0.714	0.711	2.766

## MENORES

Modelo/Métrica	Accuracy	Precisión	Recall	F1 Score	Tiempo de inferencia
Random Forest	0.987	0.987	<b>0.987</b>	0.987	0.877
Bagging Classifier	0.98	0.98	0.98	0.98	0.115
KNN	0.929	0.935	0.929	0.926	39.086
Decision Tree	0.965	0.965	0.965	0.965	0.018
LightGBM	0.962	0.963	0.962	0.962	0.705
XGBoost	0.813	0.809	0.812	0.81	0.381

## INTERMEDIOS

Modelo/Métrica	Accuracy	Precisión	Recall	F1 Score	Tiempo de inferencia
Random Forest	0.953	0.954	<b>0.952</b>	0.953	1.222
Bagging Classifier	0.932	0.932	0.931	0.931	0.145
KNN	0.886	0.9	0.886	0.88	48.894
Decision Tree	0.896	0.896	0.896	0.896	0.015
LightGBM	0.875	0.874	0.874	0.872	0.761
XGBoost	0.771	0.765	0.77	0.765	0.427

En estos dos subgrupos se obtienen mejores métricas en todos los algoritmos comparado al dataset sin agrupar por edades. Si analizamos detalladamente el **Recall**, **Random Forest**, **Bagging Classifier** y **Decision Tree** tuvieron los mejores resultados siendo el de **Random Forest 0.987** y **0.952** para **menores** e **intermedios** respectivamente.

# EVALUACIÓN DE MODELOS

Para este subgrupo se obtienen peores métricas en todos los algoritmos comparado al dataset sin agrupar por edades. Si analizamos detalladamente el **Recall**, **Random Forest**, **Bagging Classifier** y **Decision Tree** tuvieron los mejores resultados siendo el de **Random Forest 0.892**.

En las edades intermedias y menores las métricas resultaron ser mejores. Es recomendable dividir el dataset por edades.

## MAYORES

Modelo/Métrica	Accuracy	Precisión	Recall	F1 Score	Tiempo de inferencia
Random Forest	0.892	0.895	<b>0.892</b>	0.891	4.175
Bagging Classifier	0.857	0.858	0.857	0.855	0.428
KNN	0.835	0.85	0.835	0.824	249.71
Decision Tree	0.809	0.807	0.809	0.808	0.061
LightGBM	0.771	0.766	0.771	0.762	1.929
XGBoost	0.705	0.698	0.705	0.7	0.953



# INSIGHTS & **RECOMENDACIONES**

## INSIGHTS & RECOMENDACIONES

- La **cantidad** de encuestados por **rango etario** son muy dispares.
- El rango etario con mayor porcentaje de personas que sufren diabetes es el de **70 a 74 años** con un **24,41%**.
- La **obesidad** puede considerarse un factor de riesgo para esta enfermedad.
- Las personas con algún tipo de **diabetes** poseen, mayoritariamente, **niveles altos** de **colesterol** e **hipertensión**.
- De los **3 hábitos** saludables analizados ninguno posee una **gran significancia** en el estatus de diabetes.
- Es una patología que tiene mayor incidencia en poblaciones vulnerables.
- Segmentar la población por **rangos etarios** y a la hora de implementar los modelos predictivos mejora las métricas.
- En base a los resultados obtenidos, los factores **estado nutricional**, **colesterol alto** e **hipertensión** son factores de riesgo de la enfermedad.
- Ningún hábito saludable de la encuesta trajo aparejado una gran significancia en el estatus de diabetes. por lo tanto no es un factor preventivo de la enfermedad.
- De los modelos empleados, Radom Forest fue el que mejores resultados otorgó.
- Si se pondera el **tiempo de entrenamiento**, el mejor modelo es el **Bagging Classifier** siendo un 88% más rápido que **Random Forest**.
- El modelo predice mejor a **pre diabéticos**, seguido de los no diabéticos, y por último a los diabéticos.