

NHANES2009-2012

Preparación y visualización de datos

William Andres Gomez Roa
Santiago Camilo Rey Benavides
Daniel Alfredo Vidal de León

nhanes

The Nation's Mobile Health Survey

Pregunta de Investigación

¿Cómo influyen los factores socioeconómicos (ingreso, educación, propiedad de vivienda) en la prevalencia de consumo de marihuana y tabaco?

Cargar y Explorar el Dataset

- La revisión de la información del dataset se realizó utilizando el método `data.info()`.
- Rango de Índices: Desde 0 hasta 9999, indicando que el dataset tiene 10,000 entradas.
- Columnas Totales: 35 columnas.
- Resumen de Columnas:
 - Nombre de la Columna: El nombre de cada columna en el dataset.
 - Cantidad de Valores No Nulos: La cantidad de entradas no nulas en cada columna.
 - Tipo de Dato: El tipo de dato de cada columna (e.g., object, int64, float64).

Data columns (total 35 columns):			
#	Column	Non-Null Count	Dtype
0	Gender	10000 non-null	object
1	Age	10000 non-null	int64
2	Race1	10000 non-null	object
3	Education	7221 non-null	object
4	MaritalStatus	7231 non-null	object
5	HHIncome	9189 non-null	object
6	HHIncomeMid	9189 non-null	float64
7	Poverty	9274 non-null	float64
8	HomeRooms	9931 non-null	float64
9	HomeOwn	9937 non-null	object
10	Weight	9922 non-null	float64
11	Length	543 non-null	float64
12	Height	9647 non-null	float64
13	BMI	9634 non-null	float64
14	Diabetes	9858 non-null	object
15	DiabetesAge	629 non-null	float64
16	Depressed	1427 non-null	object
17	nPregnancies	2604 non-null	float64
18	nBabies	2416 non-null	float64
19	Age1stBaby	1884 non-null	float64
20	TVHrsDay	4859 non-null	object
21	CompHrsDay	4863 non-null	object
22	AlcoholDay	4914 non-null	float64
23	AlcoholYear	5922 non-null	float64
24	SmokeNow	3211 non-null	object
25	SmokeAge	3080 non-null	float64
26	Marijuana	4941 non-null	object
27	AgeFirstMarij	2891 non-null	float64
28	HardDrugs	5765 non-null	object
29	SexEver	5767 non-null	object
30	SexAge	5540 non-null	float64
31	SexNumPartnLife	5725 non-null	float64
32	SexNumPartYear	4928 non-null	float64
33	SexOrientation	4842 non-null	object
34	ID	10000 non-null	int64

Detalle de las columnas



Columna	Descripción
Gender	Género del participante (masculino o femenino).
Age	Edad del participante en años.
Race1	Raza/etnia del participante.
Education	Nivel educativo alcanzado por el participante.
MaritalStatus	Estado civil del participante (e.g., casado, soltero, etc.).
HHIncome	Rango de ingresos familiares anuales.
HHIncomeMid	Valor medio del rango de ingresos familiares anuales.
Poverty	Índice de pobreza del participante.
HomeRooms	Número de habitaciones en la vivienda del participante.
HomeOwn	Estado de propiedad de la vivienda (e.g., propia, alquilada).
Weight	Peso del participante en kilogramos.
Length	Longitud del participante en centímetros.
Height	Altura del participante en centímetros.
BMI	Índice de Masa Corporal del participante (peso en kg dividido por el cuadrado de la altura en metros).
Diabetes	Indica si el participante tiene diabetes (sí o no).
DiabetesAge	Edad a la que se diagnosticó la diabetes.
Depressed	Indica si el participante se siente deprimido (sí o no).

Detalle de las columnas



Columna	Descripción
nPregnancies	Número de embarazos que ha tenido el participante.
nBabies	Número de bebés nacidos del participante.
Age1stBaby	Edad a la que el participante tuvo su primer bebé.
TVHrsDay	Número de horas al día que el participante ve televisión.
CompHrsDay	Número de horas al día que el participante usa la computadora.
AlcoholDay	Consumo promedio de alcohol por día (en gramos).
AlcoholYear	Consumo promedio de alcohol por año (en gramos).
SmokeNow	Indica si el participante fuma actualmente (sí o no).
SmokeAge	Edad a la que el participante comenzó a fumar.
Marijuana	Indica si el participante ha consumido marihuana (sí o no).
AgeFirstMarij	Edad a la que el participante consumió marihuana por primera vez.
HardDrugs	Indica si el participante ha consumido drogas duras (sí o no).
SexEver	Indica si el participante ha tenido relaciones sexuales (sí o no).
SexAge	Edad a la que el participante tuvo su primera relación sexual.
SexNumPartnLife	Número de parejas sexuales que el participante ha tenido en su vida.
SexNumPartYear	Número de parejas sexuales que el participante ha tenido en el último año.
SexOrientation	Orientación sexual del participante (e.g., heterosexual, homosexual, bisexual).
ID	Identificador único del participante.

Preparación y limpieza de datos

- Registros duplicados
 - Filas con exactamente los mismos datos en cada una de las celdas
- Datos faltantes en algunas celdas
 - Las filas con datos faltantes de las columnas importantes ('HomeOwn', 'HHIncomeMid', 'Education', 'Marijuana', 'SmokeNow') fueron eliminadas.
- Revisar valores únicos en las variables categóricas, que no hubiera errores tipográficos.

```
Número de registros duplicados: 3221  
Número de registros duplicados después de la limpieza: 0
```

```
HomeOwn      44  
HHIncomeMid  585  
Education    2134  
Marijuana    3696  
SmokeNow     4723  
dtype: int64  
HomeOwn      0  
HHIncomeMid  0  
Education    0  
Marijuana    0  
SmokeNow     0  
dtype: int64
```

```
Valores únicos en HomeOwn: ['Rent' 'Own' 'Other']  
Valores únicos en Education: ['HighSchool' '8thGrade' 'SomeCollege' 'CollegeGrad' '9_11thGrade']  
Valores únicos en Marijuana: ['Yes' 'No']  
Valores únicos en SmokeNow: ['No' 'Yes']
```

Análisis exploratorio



La mayoría de los encuestados (54.48%) son propietarios de sus viviendas.



El ingreso medio es aproximadamente \$51,270, con una alta variabilidad (desviación estándar de \$32,817). Los ingresos van desde \$2,500 hasta \$100,000.

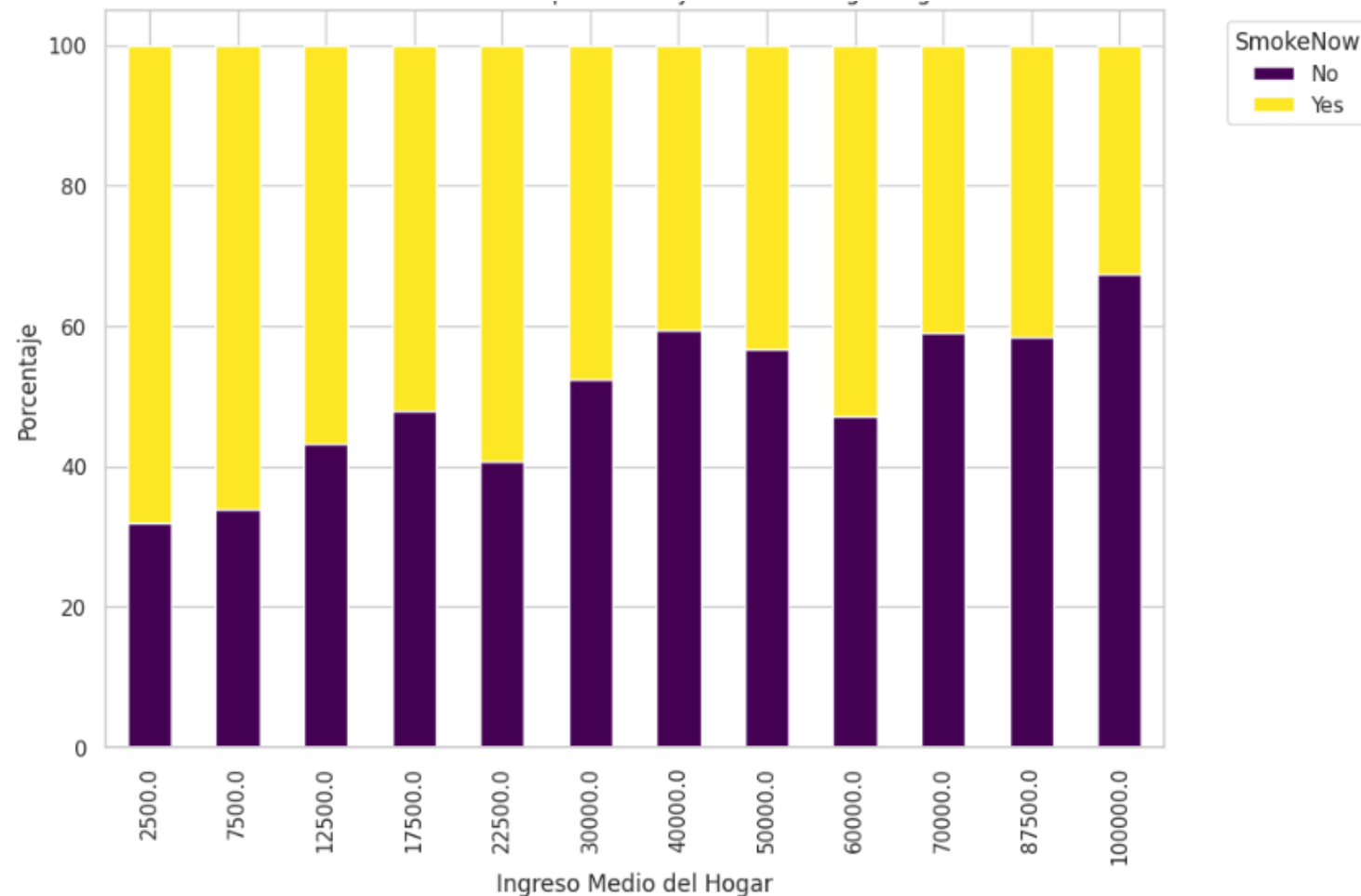
- El nivel educativo más común es "SomeCollege" (Alguna Educación Universitaria), representando el 31.27% de la muestra.
- Un 78.34% de los encuestados reportaron haber consumido marihuana.
- Un 57.21% de los encuestados indicaron que fuman actualmente.

	HomeOwn	HHIncomeMid	Education	Marijuana	SmokeNow
count	1145	1145.000000	1145	1145	1145
unique	3	NaN	5	2	2
top	Own	NaN	SomeCollege	Yes	Yes
freq	624	NaN	358	897	655
mean	NaN	51270.742358	NaN	NaN	NaN
std	NaN	32817.019850	NaN	NaN	NaN
min	NaN	2500.000000	NaN	NaN	NaN
25%	NaN	22500.000000	NaN	NaN	NaN
50%	NaN	40000.000000	NaN	NaN	NaN
75%	NaN	87500.000000	NaN	NaN	NaN
max	NaN	100000.000000	NaN	NaN	NaN

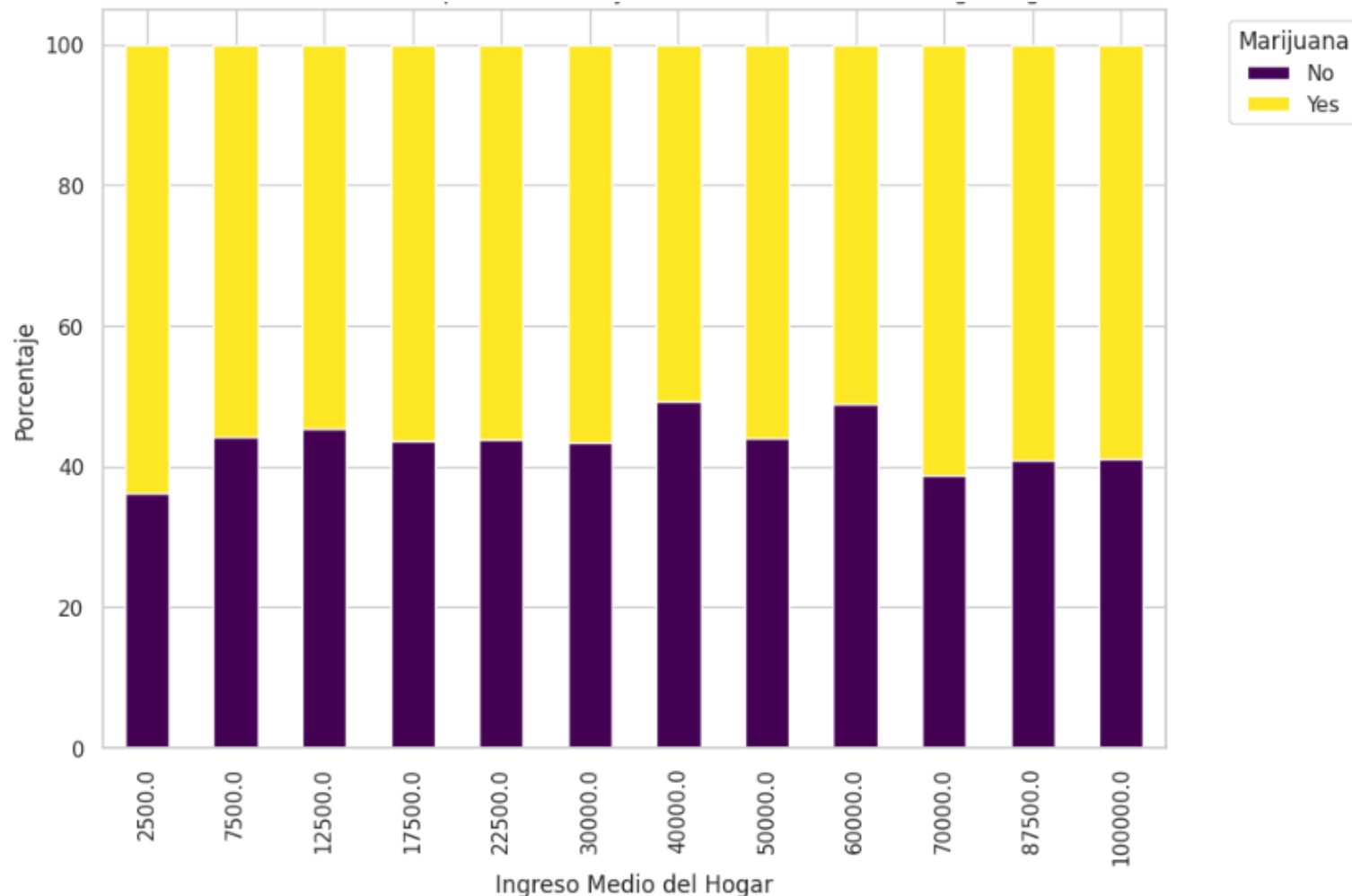
Visualizaciones



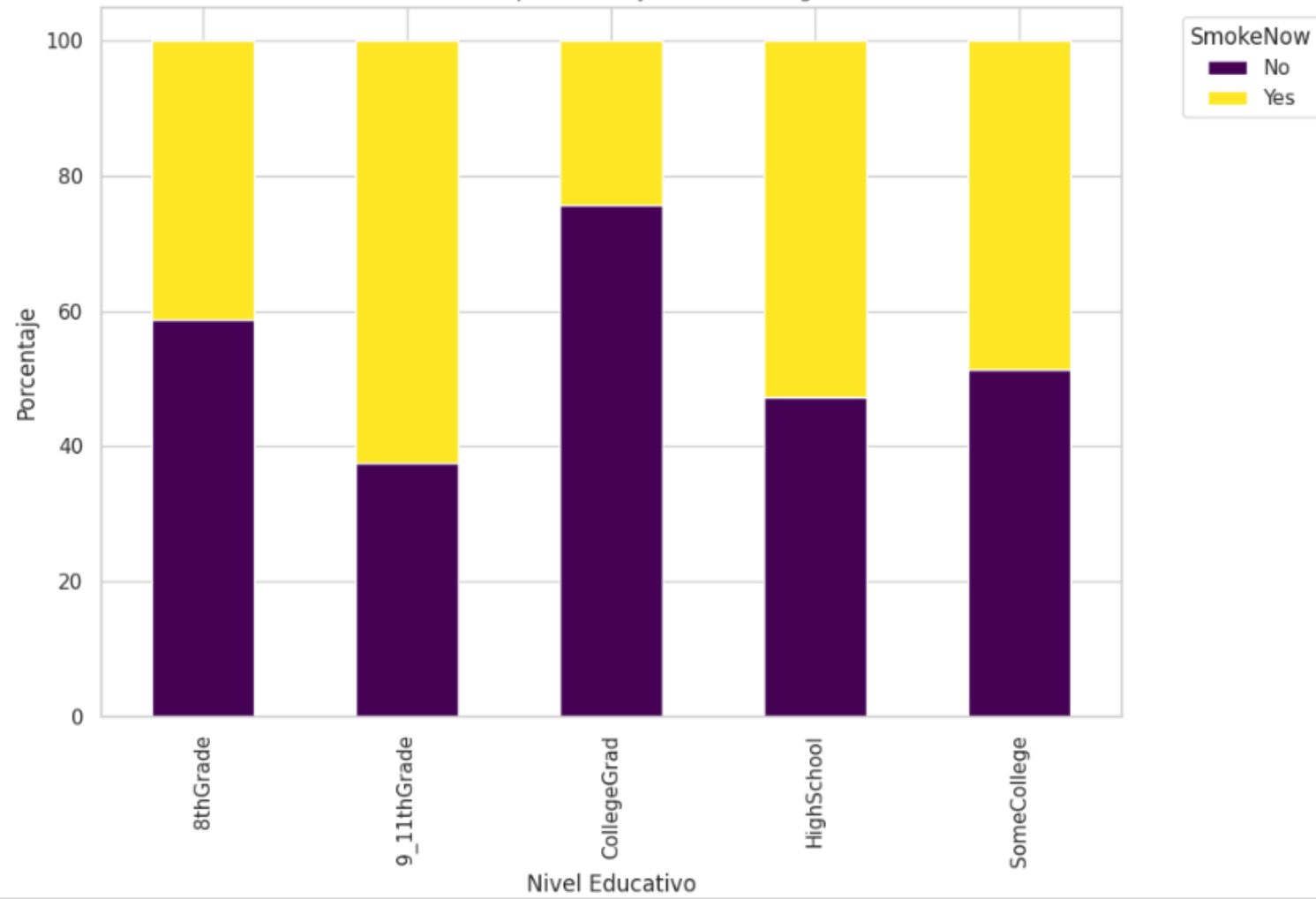
Cantidad de personas que fuman y no fuman según ingreso



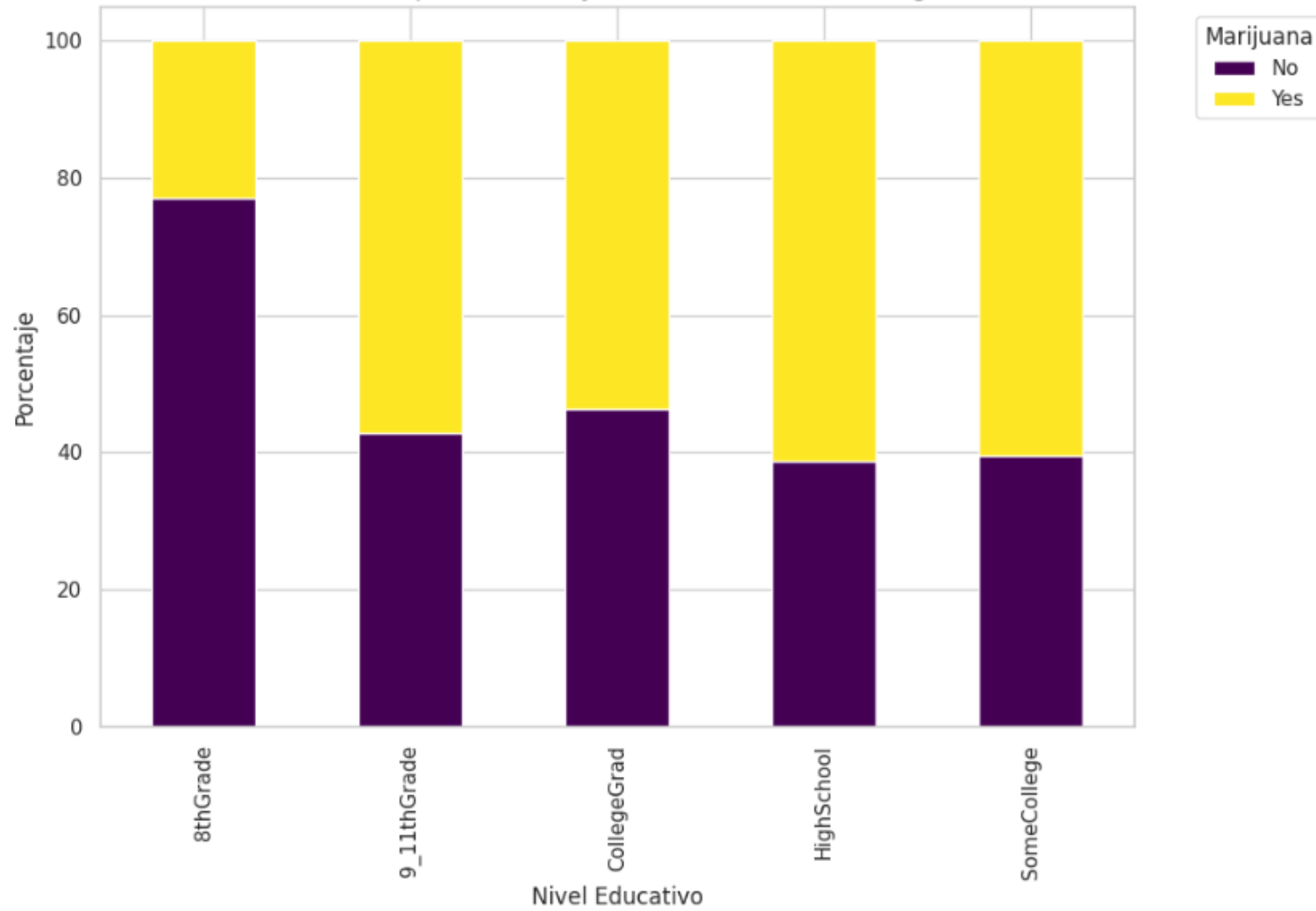
Cantidad de personas que consumen y no consumen marihuana según ingreso



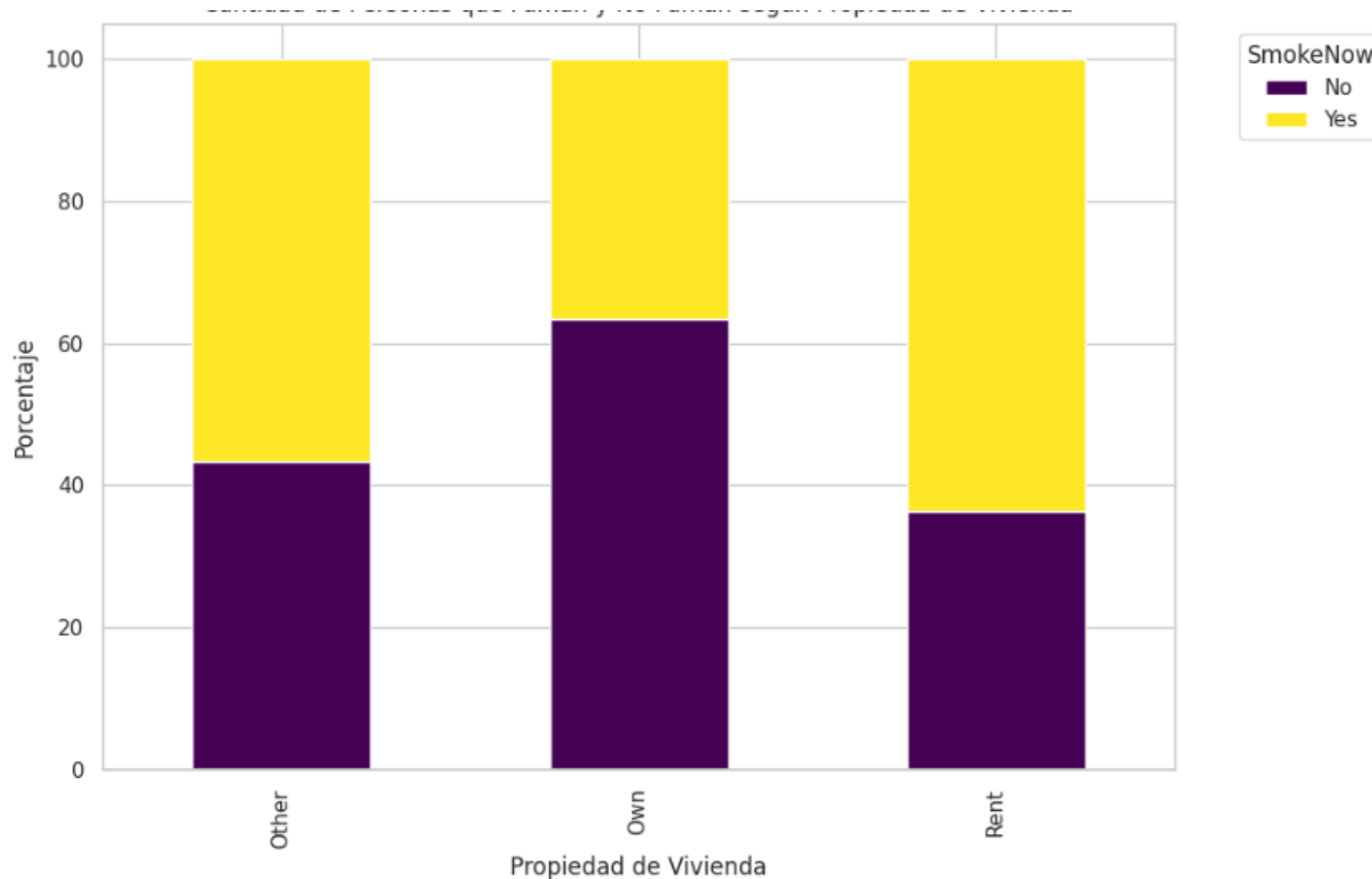
Cantidad de personas que fuman y no fuman según Educación



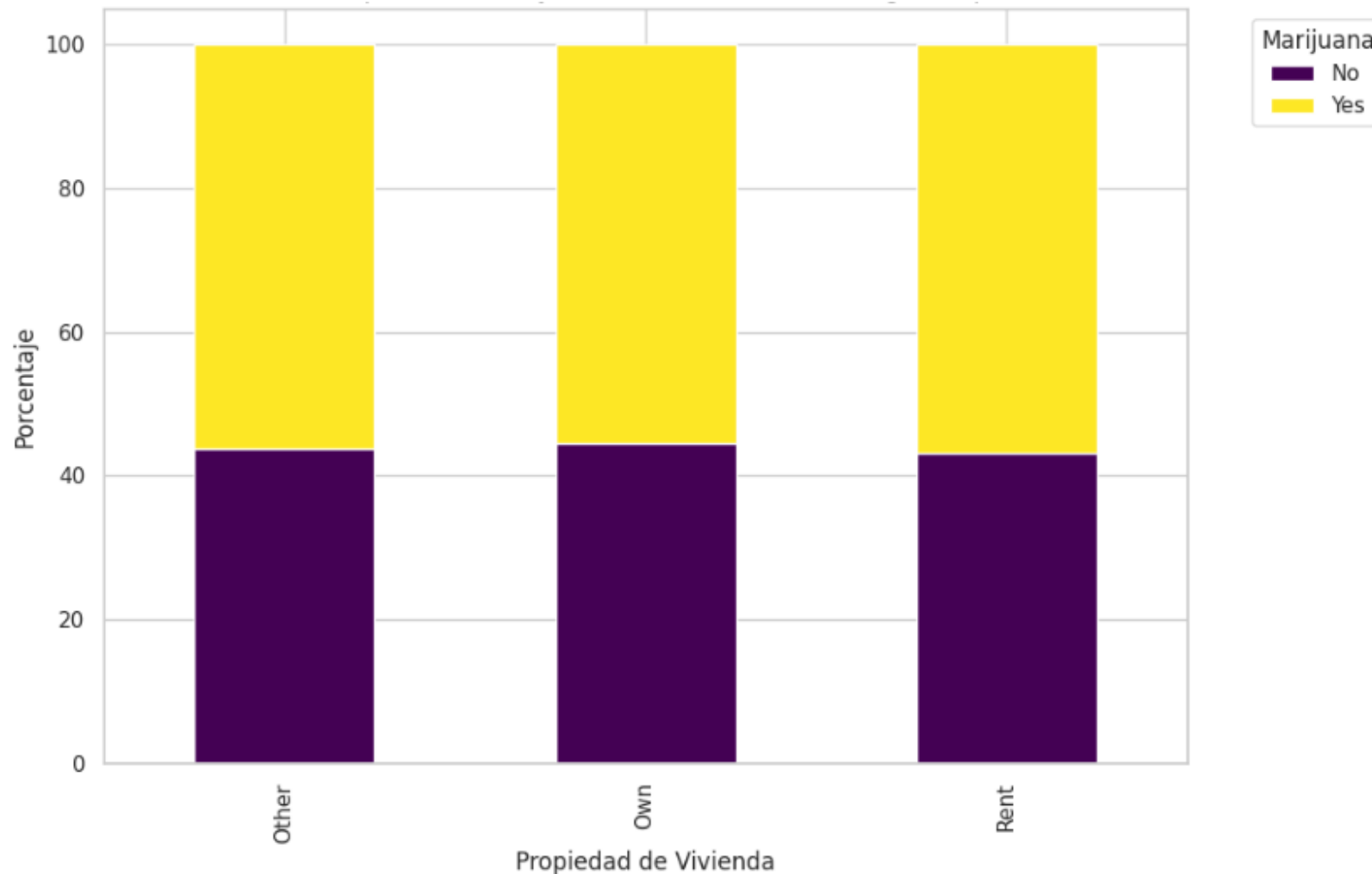
Cantidad de personas que consumen y no consumen marihuana según Educación



Cantidad de personas que fuman y no fuman según Propiedad de Vivienda



Cantidad de personas que consumen y no consumen marihuana según Propiedad de Vivienda



Conclusiones



Nivel Educativo:

El nivel educativo más común entre los encuestados es "SomeCollege" (Alguna Educación Universitaria), representando el 31.27% de la muestra.

La educación juega un papel crucial en el comportamiento de consumo. Las personas con educación universitaria o superior tienden a tener una mayor conciencia sobre los riesgos asociados con el consumo de sustancias, aunque la relación exacta con el consumo de marihuana y tabaco varía.



Consumo de Marihuana:

Un 78.34% de los encuestados reportaron haber consumido marihuana.

El alto porcentaje de consumo de marihuana sugiere una prevalencia significativa, posiblemente influenciada por factores culturales y socioeconómicos.

Los ingresos familiares pueden tener un impacto, ya que los datos sugieren que el ingreso medio es de aproximadamente \$51,270 con una amplia variabilidad. Esto implica que tanto personas de ingresos bajos como altos están involucradas en el consumo.



Consumo de Tabaco (Fumar):

Un 57.21% de los encuestados indicaron que fuman actualmente.

El consumo de tabaco también muestra una prevalencia alta, posiblemente relacionada con factores de estrés socioeconómico y accesibilidad.

Las personas con menores ingresos y niveles educativos más bajos tienden a fumar más, debido a factores como el estrés, la accesibilidad y la falta de recursos para programas de cesación.



Propiedad de Vivienda:

La mayoría de los encuestados (54.48%) son propietarios de sus viviendas.

La propiedad de vivienda está vinculada a una mayor estabilidad financiera y social, lo cual podría influir en el consumo de sustancias. Propietarios de viviendas pueden tener mejores recursos y acceso a programas educativos y de salud.



Ingresos:

El ingreso medio es aproximadamente \$51,270, con una alta variabilidad (desviación estándar de \$32,817). Los ingresos van desde \$2,500 hasta \$100,000.

Esta amplia gama de ingresos sugiere que el consumo de marihuana y tabaco no se limita a un solo grupo socioeconómico. Sin embargo, los extremos del espectro (muy bajos o muy altos ingresos) pueden mostrar patrones de consumo diferentes debido a factores como el acceso a la salud, la educación y la presión social.

Conclusiones

- **Ingreso:** A mayor ingreso, menor prevalencia de consumo de tabaco y marihuana.
- **Educación:** A mayor nivel educativo, menor prevalencia de consumo de tabaco y marihuana.
- **Propiedad de Vivienda:** Los propietarios de viviendas tienen una menor prevalencia de consumo de tabaco y marihuana en comparación con los que alquilan o tienen otro tipo de propiedad de vivienda.

