

# Reglas de Asociación: Trabajo Final

Oscar Julian Tinjacá Reyes

2025-12-15

## Contents

<b>1</b>	<b>Introducción</b>	<b>1</b>
<b>2</b>	<b>Presentación del dataset</b>	<b>2</b>
2.1	Descripción general . . . . .	2
2.1.1	Fuente de los datos . . . . .	2
2.2	Carga y unión de los datos . . . . .	3
<b>3</b>	<b>Preprocesamiento de los datos</b>	<b>3</b>
<b>4</b>	<b>Exploración descriptiva de las variables</b>	<b>3</b>
4.1	Frecuencia de ítems en el conjunto de datos . . . . .	3
4.1.1	Ítems con frecuencia alta (soporte > 0.5) . . . . .	3
4.1.2	Ítems con frecuencia media (0.1 ≤ soporte < 0.5) . . . . .	4
4.2	Distribución del tamaño de los <i>itemsets</i> . . . . .	4
4.3	Comparación de <i>itemsets</i> frecuentes, cerrados y maximales . . . . .	4
4.4	Relación entre soporte, confianza y <i>lift</i> en las reglas generadas (con <i>support</i> ≥ 0.1) . . . . .	5
<b>5</b>	<b>Generación de reglas de asociación</b>	<b>5</b>
5.1	Regla 1: El consecuente es la vacunación H1N1 . . . . .	5
5.2	Regla 2: El consecuente es la vacunación estacional . . . . .	7
5.3	Regla 3: Exploración de patrones de hogar en subgrupos poblacionales . . . . .	8
5.4	Regla 4: El consecuente es la población con vivienda propia . . . . .	9
5.5	Reglas globales con confianza entre 0.5 y 0.7 . . . . .	10
5.5.1	Regla 5: Sociabilidad y estructura del hogar . . . . .	11
5.5.2	Regla 6: Participación laboral y características demográficas . . . . .	11
<b>6</b>	<b>Reglas con ítems negativos</b>	<b>12</b>
6.1	Explicación . . . . .	12
6.2	Preprocesamiento . . . . .	12
6.3	Generación y limpieza de reglas . . . . .	12
6.3.1	Regla 7: Conductas preventivas y hábitos persistentes en el grupo de edad 55–64 . . . . .	12
6.3.2	Regla 8: Conocimiento, conducta preventiva y percepción de la vacuna H1N1 . . . . .	13
6.4	Conclusiones . . . . .	14

## 1 Introducción

En este trabajo se analizan **reglas de asociación** a partir de datos reales relacionados con campañas de vacunación frente a la gripe **H1N1** y la gripe **estacional**. El objetivo principal consiste en extraer patrones interesantes que relacionen características demográficas, comportamientos preventivos y opiniones sobre las vacunas con la probabilidad de haber sido vacunado.

El conjunto de datos procede de la competición “*Flu Shot Learning: Predict H1N1 and Seasonal Flu Vaccines*” de DrivenData, basada en la encuesta nacional estadounidense “*National 2009 H1N1 Flu Survey*”. Cada instancia

representa a una persona encuestada y recoge información sobre:

- Nivel de preocupación y conocimiento sobre la gripe H1N1.
- Comportamientos de prevención (lavado de manos, uso de mascarilla, evitar contactos, etc.).
- Recomendaciones médicas y presencia de condiciones crónicas.
- Características demográficas y socioeconómicas (edad, educación, ingresos, estado civil, etc.).
- Estructura del hogar y localización geográfica.

Para cada individuo se dispone de dos variables objetivo binarias:

- `h1n1_vaccine`: indica si se recibió la vacuna H1N1.
- `seasonal_vaccine`: indica si se recibió la vacuna de la gripe estacional.

En este trabajo, en lugar de plantear un problema de predicción supervisada, se emplean técnicas de **reglas de asociación** con el fin de descubrir dependencias y patrones interesantes entre todas estas variables.

## 2 Presentación del dataset

### 2.1 Descripción general

El dataset original contiene **36 variables**:

- `respondent_id`: identificador único y aleatorio del encuestado.
- 35 variables predictoras, que incluyen:
  - Opiniones y percepciones (`h1n1_concern`, `h1n1_knowledge`, `opinion_*`).
  - Comportamientos preventivos (`behavioral_*`).
  - Variables de salud y recomendación médica (`doctor_recc_*`, `chronic_med_condition`, etc.).
  - Variables demográficas y socioeconómicas (`age_group`, `education`, `race`, `sex`, `income_poverty`, etc.).
  - Estructura del hogar (`household_adults`, `household_children`).
  - Variables geográficas (`hhs_geo_region`, `census_msa`).
  - Sector e industria de empleo (`employment_industry`, `employment_occupation`).

Adicionalmente, se dispone de las dos variables objetivo:

- `h1n1_vaccine` (0 = no vacunado, 1 = vacunado).
- `seasonal_vaccine` (0 = no vacunado, 1 = vacunado).

#### 2.1.1 Fuente de los datos

Los datos originales se encuentran disponibles públicamente en la plataforma DrivenData dentro de la competición:

- “*Flu Shot Learning: Predict H1N1 and Seasonal Flu Vaccines*”  
<https://www.drivendata.org/competitions/66/flu-shot-learning/>

La descripción detallada de las variables puede consultarse en la página oficial del problema:

- *Problem description*  
<https://www.drivendata.org/competitions/66/flu-shot-learning/page/211/>

## 2.2 Carga y unión de los datos

En primer lugar se cargan los ficheros de características (`training_set_features.csv`) y etiquetas (`training_set_labels.csv`), y se unen por la columna común `respondent_id`.

## 3 Preprocesamiento de los datos

A continuación se describe la lista de **preprocesos** aplicados a los datos.

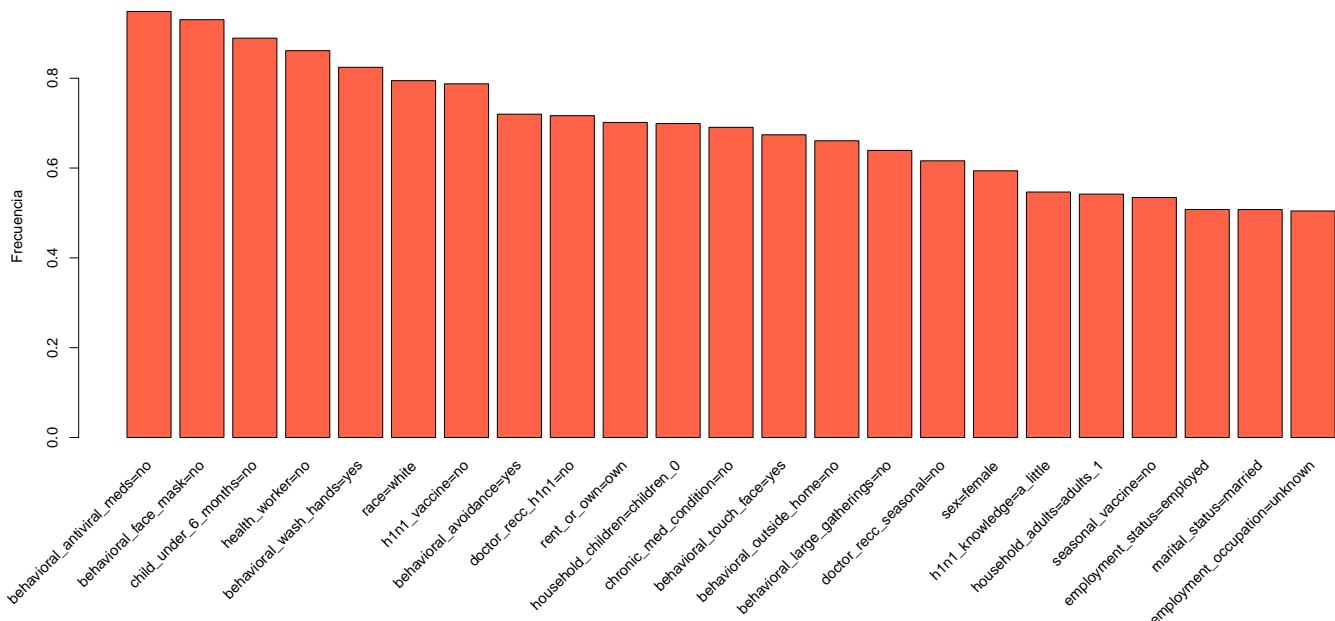
- Conversión de variables a factores
- Tratamiento de valores ausentes
  - En las variables nominales y en las relacionadas con la estructura del hogar, los valores ausentes (NA) se convierten en una categoría explícita “*unknown*” para evitar la pérdida de información y poder analizar patrones asociados a la falta de respuesta o información desconocida. En cambio, en las variables binarias y ordinales de opinión, los NA se mantienen, de forma que no generan ítems artificiales en el conjunto de transacciones.
- Renombrado de niveles (*snake\_case*)
  - Con el fin de facilitar la interpretación de las reglas de asociación, se renombraron los niveles de algunas variables nominales a un formato consistente en *snake\_case*.

## 4 Exploración descriptiva de las variables

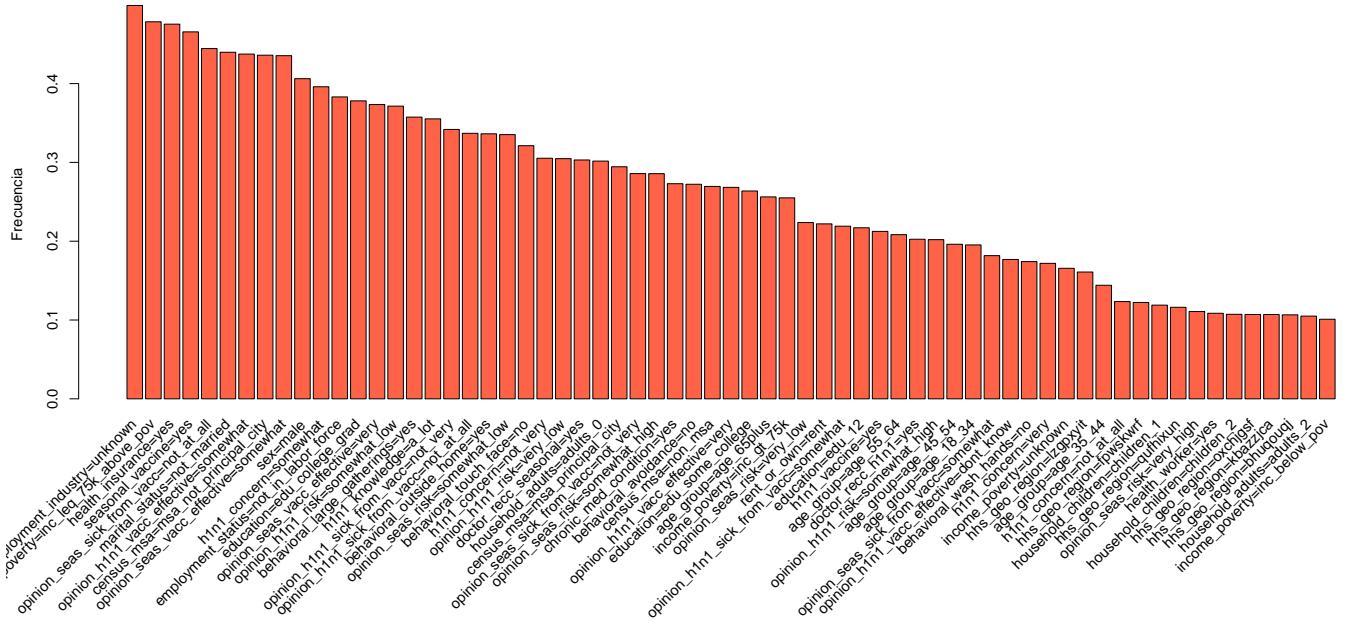
### 4.1 Frecuencia de ítems en el conjunto de datos

En esta sección se analiza la frecuencia relativa de los ítems con mayor presencia en el conjunto de transacciones, lo que permite identificar cuáles variables aparecen con más regularidad en los datos.

#### 4.1.1 Ítems con frecuencia alta (soporte > 0.5)

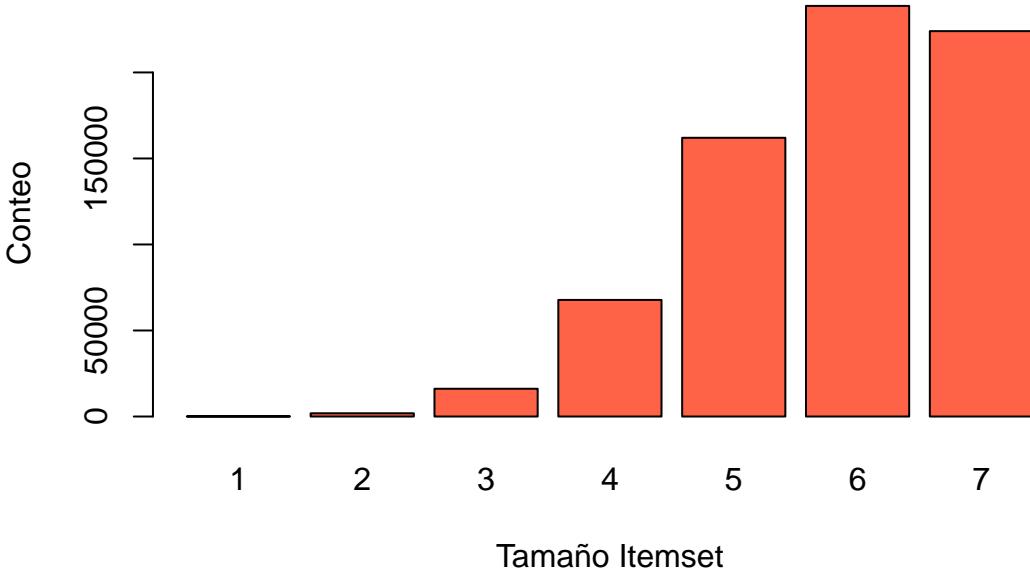


#### 4.1.2 Ítems con frecuencia media (0.1 soporte 0.5)



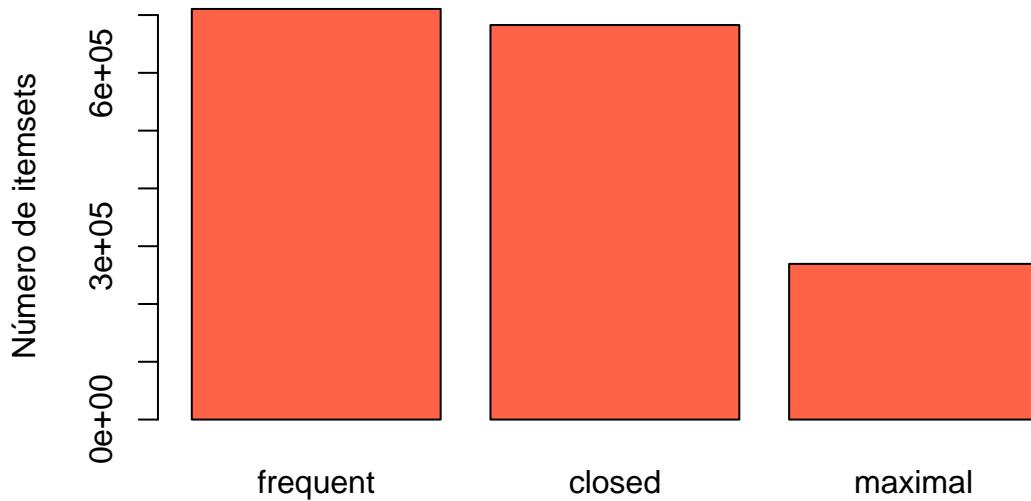
## 4.2 Distribución del tamaño de los *itemsets*

En la gráfica correspondiente se observa un patrón ascendente: a medida que aumenta el tamaño del *itemset*, su frecuencia también crece, alcanzando el máximo en los *itemsets* de tamaño 7.



### 4.3 Comparación de *itemsets* frecuentes, cerrados y maximales

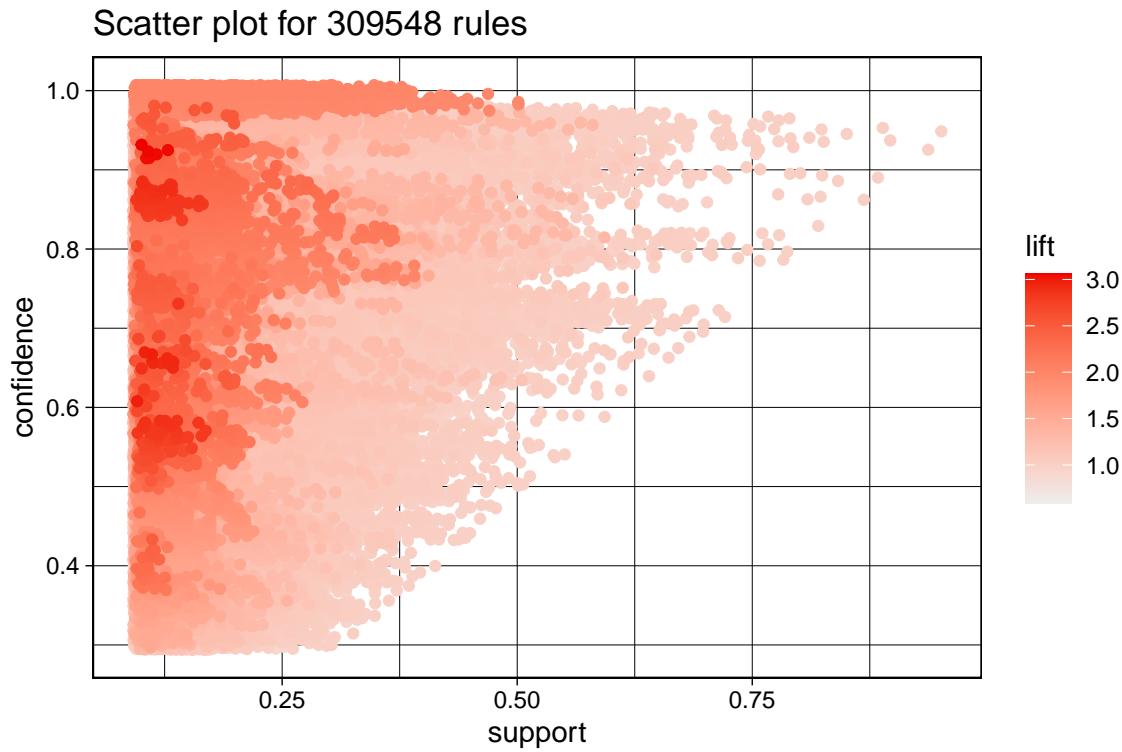
En la siguiente gráfica se comparan las cantidades totales de *itemsets* generados. Se observa que los **frecuentes** y **cerrados** son similares en número, mientras que los **maximales** representan una fracción mucho menor del conjunto total.



#### 4.4 Relación entre soporte, confianza y *lift* en las reglas generadas (con *support* 0.1)

En esta parte se visualiza la distribución conjunta de la **confianza** y el **lift** en las reglas generadas, destacando zonas donde estos valores tienden a agruparse.

Las reglas con mayor *lift* se concentran sobre todo en la zona donde la confianza oscila entre 0.5–0.7 y 0.8–0.9. Sin embargo, a lo largo del documento se exploran reglas distribuidas por todo el conjunto de datos.

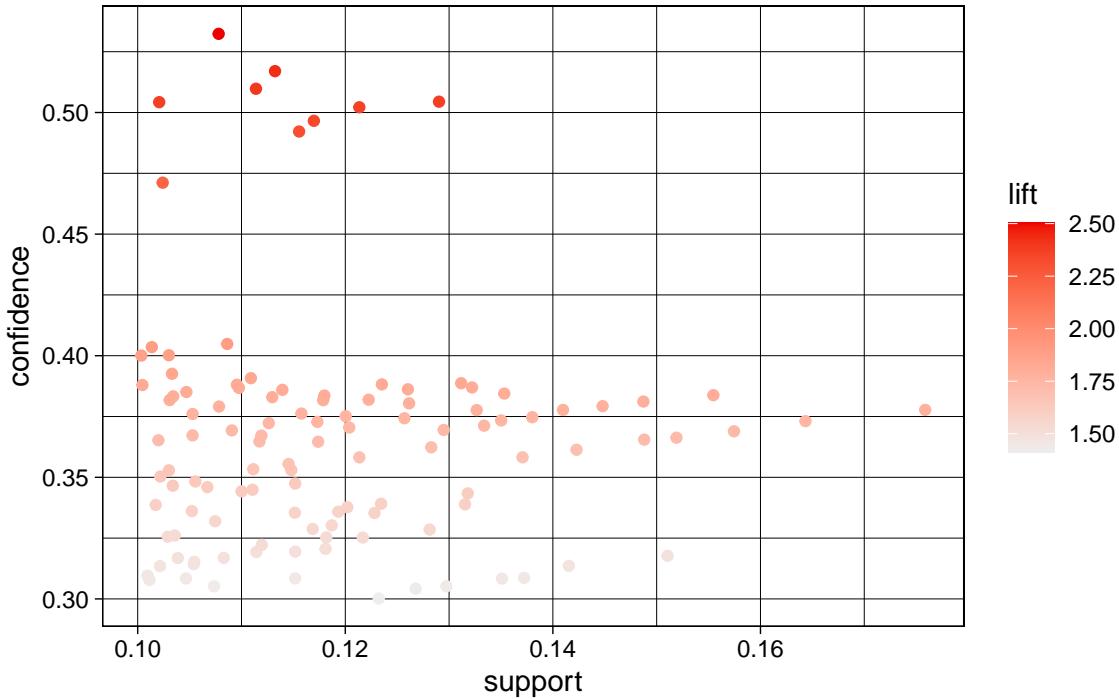


## 5 Generación de reglas de asociación

### 5.1 Regla 1: El consecuente es la vacunación H1N1

Se seleccionan las reglas cuyo consecuente (**rhs**) corresponde a la vacunación H1N1, con el objetivo de aislar patrones que describen perfiles con una mayor probabilidad de vacunarse.

## Scatter plot for 113 rules



rules	confidence	lift
{doctor_recc_h1n1=yes} => {h1n1_vaccine=yes}	0.5323595	2.505767
{behavioral_wash_hands=yes, health_insurance=yes, seasonal_vaccine=yes} => {h1n1_vaccine=yes}	0.5170115	2.433526
{health_insurance=yes, race=white, seasonal_vaccine=yes} => {h1n1_vaccine=yes}	0.5097670	2.399427
{health_insurance=yes, seasonal_vaccine=yes} => {h1n1_vaccine=yes}	0.5044649	2.374470
{health_insurance=yes, rent_or_own=own, seasonal_vaccine=yes} => {h1n1_vaccine=yes}	0.5042545	2.373480

Al revisar las primeras reglas se observa un patrón muy evidente: casi todas incluyen **health\_insurance = "yes"** y **seasonal\_vaccine = "yes"**. Esto resulta coherente, ya que las personas que disponen de **seguro médico** y se vacunan de forma habitual tienden a recibir también la vacuna frente a la gripe H1N1. Sin embargo, estas reglas no aportan información novedosa, pues principalmente confirman una relación esperable.

Con el fin de descubrir patrones más específicos, se decide eliminar de las reglas todos los casos en los que aparecen estas dos variables tan dominantes. La idea es analizar qué relaciones persisten cuando se dejan fuera el **seguro médico** y la **vacunación estacional** del antecedente.

rules	confidence	lift
{doctor_recc_h1n1=yes} => {h1n1_vaccine=yes}	0.5323595	2.505767
{opinion_h1n1_vacc_effective=very} => {h1n1_vaccine=yes}	0.4048284	1.905490
{doctor_recc_seasonal=yes} => {h1n1_vaccine=yes}	0.3482827	1.639335
{behavioral_wash_hands=yes, opinion_seas_vacc_effective=very} => {h1n1_vaccine=yes}	0.3168113	1.491202
{opinion_seas_vacc_effective=very} => {h1n1_vaccine=yes}	0.3084328	1.451765

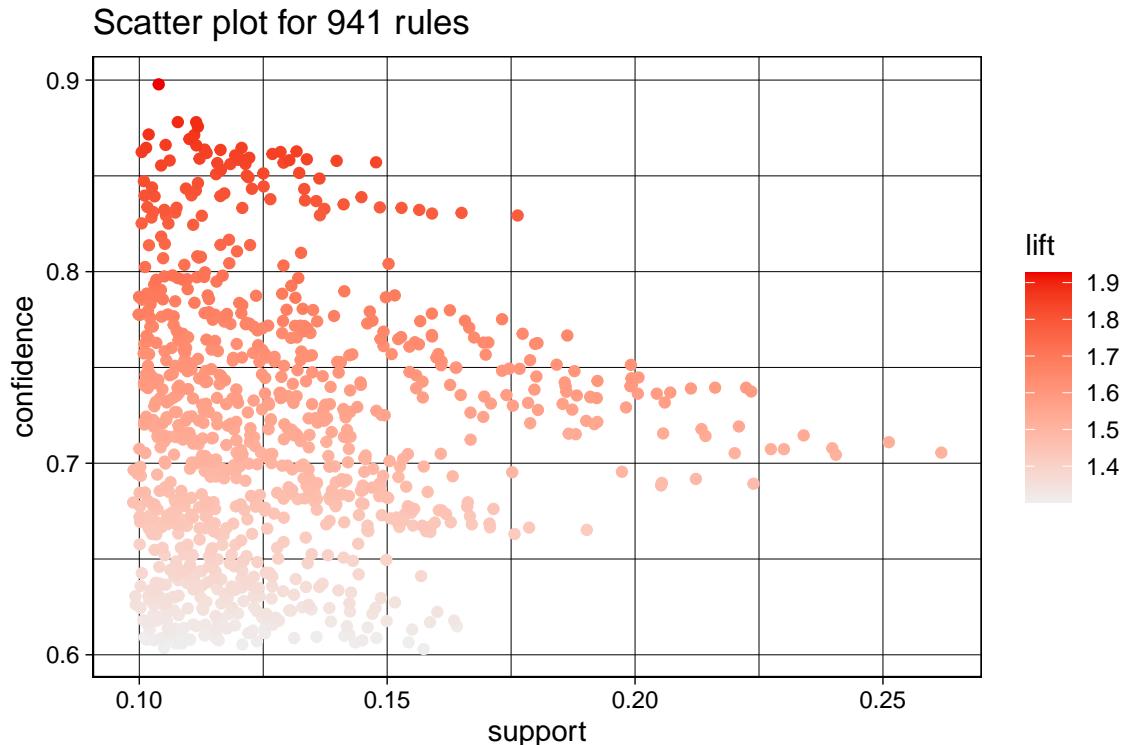
El resultado de este filtrado produce un conjunto de reglas más informativas y menos redundantes. Entre ellas destaca especialmente la siguiente:

$$\{\text{doctor\_recc\_seasonal} = \text{yes}\} \Rightarrow \{\text{h1n1\_vaccine} = \text{yes}\}$$

Esta regla presenta un soporte cercano al 10 % de la muestra, una confianza aproximada del 35 % y un *lift* cercano a 1.64. Su interpretación resulta especialmente interesante porque revela un fenómeno de **transferencia entre campañas de vacunación**: las personas que reciben una recomendación médica para vacunarse de la gripe estacional muestran una probabilidad significativamente mayor de vacunarse también frente a la gripe H1N1.

## 5.2 Regla 2: El consecuente es la vacunación estacional

En esta sección se seleccionan las reglas cuyo consecuente (`rhs`) corresponde a la vacunación estacional (`seasonal_vaccine = "yes"`), con el objetivo de aislar patrones que describen perfiles con una mayor probabilidad de haberse vacunado contra la gripe estacional.



rules	confidence	lift
{opinion_seas_vacc_effective=very,h1n1_vaccine=yes} => {seasonal_vaccine=yes}	0.8969441	1.926392
{doctor_recc_seasonal=yes,opinion_seas_vacc_effective=very,household_children=children_0} => {seasonal_vaccine=yes}	0.8797899	1.889549
{doctor_recc_seasonal=yes,opinion_seas_vacc_effective=very,race=white} => {seasonal_vaccine=yes}	0.8777520	1.885173
{race=white,household_children=children_0,h1n1_vaccine=yes} => {seasonal_vaccine=yes}	0.8771311	1.883839
{doctor_recc_seasonal=yes,opinion_seas_vacc_effective=very,rent_or_own=own} => {seasonal_vaccine=yes}	0.8732394	1.875481

Al revisar las primeras reglas se observan nuevamente relaciones muy esperables: por ejemplo, aquellas en las que intervienen la **recomendación médica** o la **vacunación contra H1N1**, ambas variables estrechamente vinculadas al comportamiento vacunal estacional. Estas reglas resultan útiles para validar el enfoque inicial, pero no aportan información novedosa, ya que simplemente confirman patrones altamente previsibles.

Por este motivo, se decide filtrar las reglas eliminando del antecedente (`lhs`) cualquier aparición de `doctor_recc_seasonal` y `h1n1_vaccine`. El objetivo consiste en aislar asociaciones menos obvias y detectar qué otros factores continúan vinculados a la vacunación estacional una vez retirados estos predictores tan fuertes.

rules	confidence	lift
{opinion_seas_vacc_effective=very,age_group=age_65plus} => {seasonal_vaccine=yes}	0.8326542	1.788315
{child_under_6_months=no,opinion_seas_vacc_effective=very,age_group=age_65plus} => {seasonal_vaccine=yes}	0.8326114	1.788223
{opinion_seas_vacc_effective=very,opinion_seas_risk=somewhat_high,race=white} => {seasonal_vaccine=yes}	0.8129323	1.745958
{health_insurance=yes,opinion_seas_vacc_effective=very,race=white} => {seasonal_vaccine=yes}	0.8039400	1.726644
{behavioral_touch_face=yes,health_insurance=yes,opinion_seas_vacc_effective=very} => {seasonal_vaccine=yes}	0.8019533	1.722378

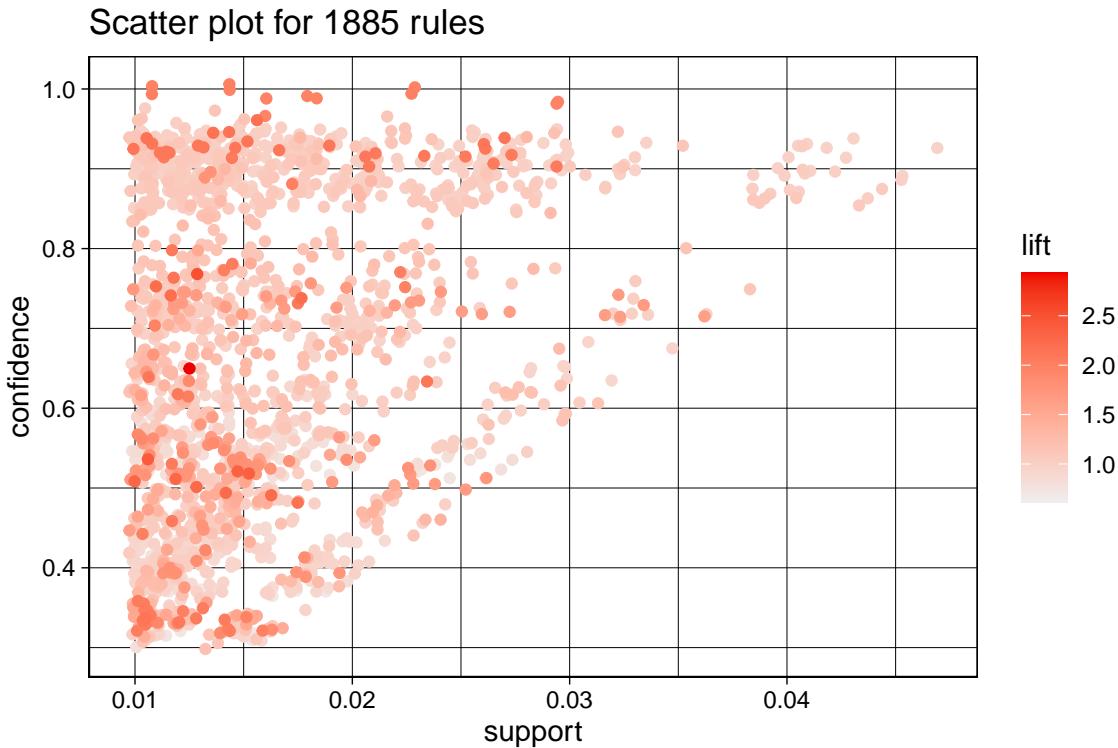
Una de las reglas más informativas tras el filtrado es la siguiente:

$$\{\text{opinion\_seas\_vacc\_effective} = \text{very}, \text{age\_group} = \text{age\_65plus}\} \Rightarrow \{\text{seasonal\_vaccine} = \text{yes}\}$$

Esta asociación revela un patrón muy consistente: entre las personas **mayores de 65 años** que consideran la vacuna estacional *muy efectiva*, la probabilidad de estar vacunadas es especialmente alta (confianza aproximada de 0.83). El soporte en torno al 10 % indica que este grupo no es marginal dentro de la muestra, y el *lift* superior a 1.7 muestra que la combinación de **edad avanzada** y **creencia fuerte en la eficacia de la vacuna** incrementa la probabilidad de vacunación mucho más de lo que cabría esperar por simple coincidencia.

### 5.3 Regla 3: Exploración de patrones de hogar en subgrupos poblacionales

Tras completar el estudio de las reglas asociadas a la vacunación, se amplía el análisis hacia aspectos más sociales del conjunto de datos. Para ello, se generan reglas sin imponer un consecuente fijo y se examina el caso de la población identificada como `race = black`, combinando esta variable con información sobre la estructura del hogar (número de adultos y presencia de niños).



rules	confidence	lift
{opinion_h1n1_risk=very_low,race=black,household_children=children_0 => {opinion_seas_risk=very_low}}	0.6563107	2.934063
{opinion_seas_risk=very_low,race=black,household_children=children_0 => {opinion_h1n1_risk=very_low}}	0.7629797	2.503612
{opinion_seas_sick_from_vacc=not_at_all,race=black,household_children=children_0 => {opinion_seas_risk=very_low}}	0.5381605	2.405867
{race=black,marital_status=not_married,household_adults=adults_0 => {rent_or_own=rent}}	0.5234474	2.357853
{health_worker=no,race=black,household_adults=adults_0 => {rent_or_own=rent}}	0.5186170	2.336095

Entre los resultados destaca la siguiente regla:

$$\{ \text{race} = \text{black}, \text{household\_adults} = \text{adults\_0} \} \Rightarrow \{ \text{rent\_or\_own} = \text{rent} \}$$

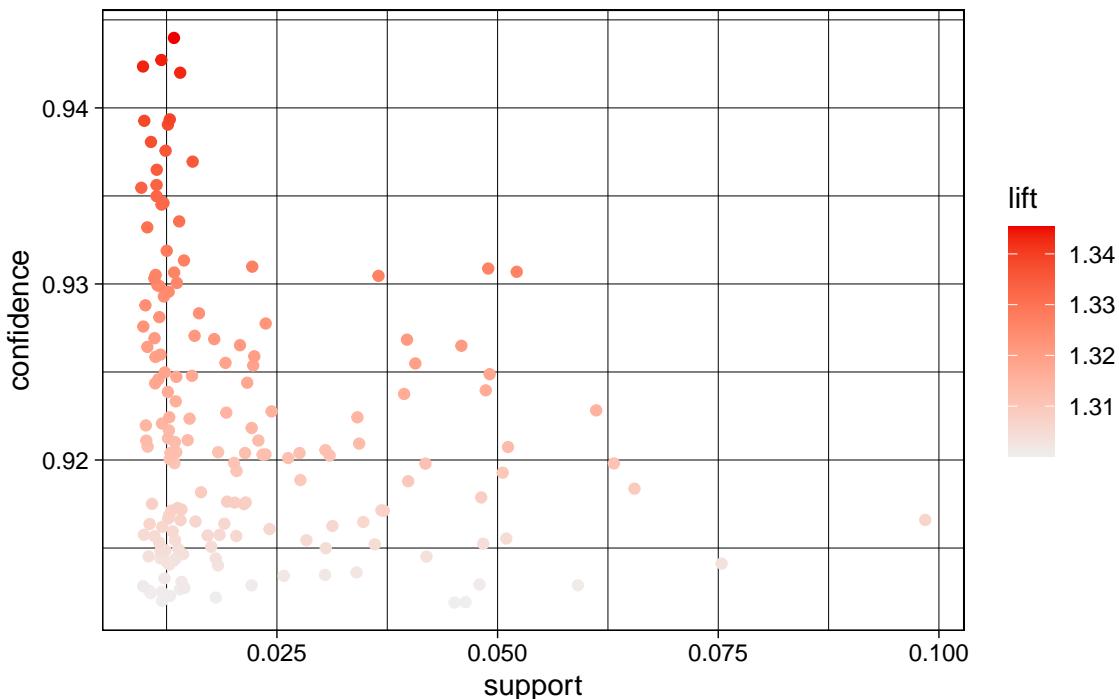
Esta asociación indica que las personas identificadas como `race = black` que viven solas tienen **más del doble de probabilidad** de residir en alquiler en comparación con la población general (*lift* 2.23). La confianza de la regla es cercana al 50 %, lo que implica que aproximadamente la mitad de este subgrupo efectivamente vive en alquiler. Este resultado sugiere un patrón residencial específico dentro de este segmento poblacional.

#### 5.4 Regla 4: El consecuente es la población con vivienda propia

Durante el proceso de búsqueda de reglas relacionadas con la **propiedad de vivienda**, se observa que los ingresos altos (`income_poverty = inc_gt_75k`) aparecen de manera dominante. La mayoría de combinaciones que incluyen este nivel de ingresos conducen casi de forma directa a la conclusión de vivienda propia, lo cual resulta poco informativo desde el punto de vista analítico por ser un patrón muy esperado.

Por esta razón, se decide **excluir explícitamente** las reglas cuyo antecedente incluya ingresos elevados, con el fin de revelar asociaciones más sutiles basadas en factores sociales y residenciales.

Scatter plot for 170 rules



rules	confidence	lift
{marital_status=married,census_msa=msa_not_principal_city,employment_occupation=cmhcxjea} => {rent_or_own=own}	0.9438503	1.345400
{age_group=age_55_64,marital_status=married,employment_occupation=xtkaffoo} => {rent_or_own=own}	0.9426752	1.343725
{sex=female,marital_status=married,employment_industry=ldnlellj} => {rent_or_own=own}	0.9423729	1.343294
{age_group=age_55_64,marital_status=married,hhs_geo_region=qufhixun} => {rent_or_own=own}	0.9422111	1.343063
{h1n1_knowledge=a_lot,census_msa=msa_not_principal_city,employment_occupation=xtkaffoo} => {rent_or_own=own}	0.9395604	1.339285

Tras aplicar este filtrado, una de las reglas más interesantes identificadas es la siguiente:

```
{marital_status = married, census_msa = msa_not_principal_city,
employment_occupation = cmhcxjea} => {rent_or_own = own}
```

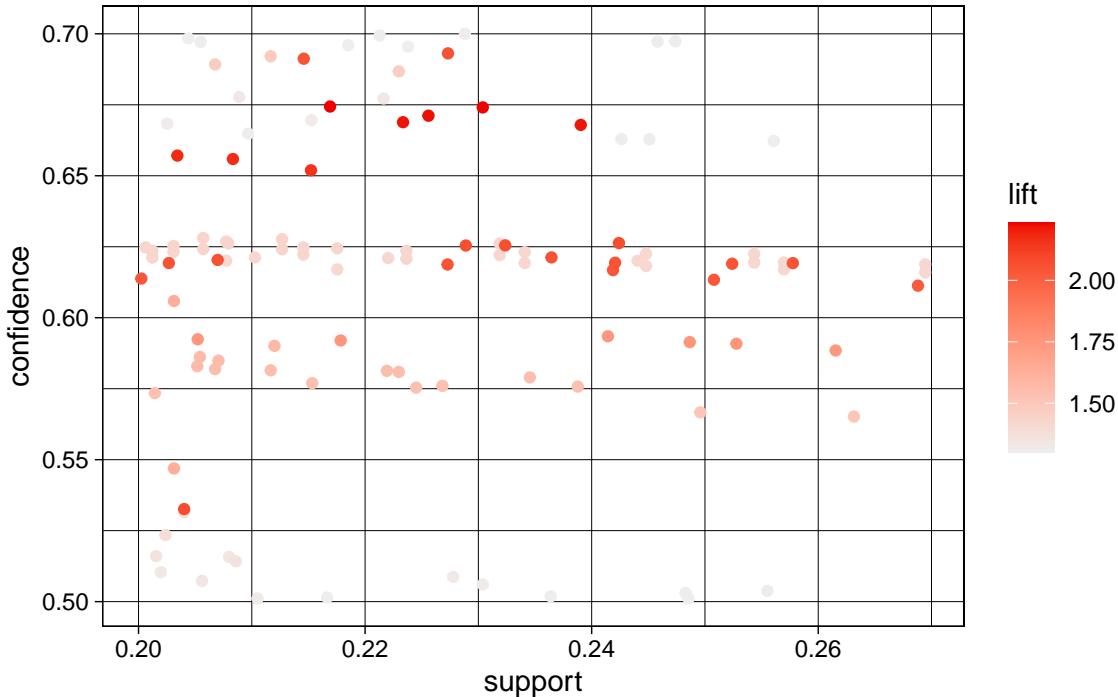
Esta regla presenta una **confianza del 94 %** y un **lift de 1.35**, indicando que este subgrupo tiene una probabilidad significativamente mayor de ser propietario en comparación con la población general. El interés de esta asociación radica en que no depende directamente de ingresos altos, sino de una combinación de **estado civil, ubicación en áreas metropolitanas secundarias y un tipo específico de ocupación**. Este patrón sugiere un perfil residencial caracterizado por estabilidad laboral y familiar, donde la propiedad de vivienda resulta especialmente frecuente.

## 5.5 Reglas globales con confianza entre 0.5 y 0.7

En esta sección se adopta un enfoque distinto respecto a las búsquedas anteriores. En lugar de partir de una variable objetivo específica, se filtran todas las reglas generadas según criterios puramente estadísticos:

- confianza entre 0.5 y 0.7
- soporte mínimo del 20 %
- *lift* mayor o igual que 1.3
- exclusión de cualquier ítem con valor *unknown*
- eliminación de reglas redundantes

## Scatter plot for 116 rules



Este filtrado permite identificar patrones amplios dentro del conjunto de datos, no necesariamente relacionados con la vacunación, pero sí útiles para comprender comportamientos y características sociodemográficas de la muestra. A continuación se presentan dos reglas particularmente llamativas.

### 5.5.1 Regla 5: Sociabilidad y estructura del hogar

`{behavioral_large_gatherings = yes, household_children = children_0} ⇒ {behavioral_outside_home = yes}`

Esta regla presenta un **soporte cercano al 22 %**, una **confianza alrededor del 69 %** y un lift superior a **2.05**, lo que indica que este patrón es más del doble de frecuente de lo que se esperaría por puro azar.

En términos prácticos, esta asociación sugiere que las personas que **participan en grandes reuniones y no tienen hijos** muestran una **tendencia marcada a salir frecuentemente del hogar**. La ausencia de niños parece facilitar horarios más flexibles y mayor disponibilidad, mientras que la participación en eventos grandes apunta a un estilo de vida más social y activo.

### 5.5.2 Regla 6: Participación laboral y características demográficas

`{health_worker = no, race = white, sex = female} ⇒ {employment_status = not_in_labor_force}`

Esta regla muestra un **soporte del 20 %**, una **confianza cercana al 51 %** y un lift de **alrededor de 1.33**, lo que indica que este grupo tiene una probabilidad moderadamente mayor que la media de encontrarse fuera de la fuerza laboral.

Aunque no implica relaciones causales, sí se observa un **patrón demográfico consistente**: dentro del subconjunto de **mujeres blancas que no trabajan en el sector sanitario**, la proporción de personas que no participan en el mercado laboral es significativamente mayor que en el conjunto total. Esto sugiere diferencias ocupacionales o roles sociales que se reflejan en la estructura del dataset.

## 6 Reglas con ítems negativos

### 6.1 Explicación

En esta sección se incorporan **ítems negativos** dentro del proceso de generación de reglas de asociación. El algoritmo Apriori, por defecto, solo trabaja con la presencia de ítems, por lo que no puede producir reglas del tipo “*si NO ocurre X entonces ocurre Y*”. Para permitirlo, es necesario transformar algunas variables del dataset en un formato que represente explícitamente tanto la presencia como la ausencia de cada una de sus categorías.

Este tipo de reglas amplía el alcance del análisis porque introduce patrones basados no solo en lo que las personas son o hacen, sino también en lo que **no son** o **no hacen**. Esto permite detectar contrastes interesantes en grupos poblacionales y en patrones de decisión.

### 6.2 Preprocesamiento

Para evitar una explosión combinatoria, el análisis se centrará únicamente en **dos variables** del problema cuyas categorías son relevantes desde el punto de vista interpretativo:

- h1n1\_knowledge
- age\_group

Cada una será transformada en un conjunto de variables binarias (una por categoría), donde cada registro tomará el valor **TRUE** si pertenece a esa categoría y **FALSE** en caso contrario. Con ello se podrán generar reglas que incluyan tanto afirmaciones del tipo “*h1n1\_knowledge\_alot = TRUE*” como negaciones del tipo “*h1n1\_knowledge\_alot = FALSE*”.

La variable original será eliminada para evitar ambigüedades, y el dataset transaccional se reconstruirá a partir de estas nuevas variables, ya expresadas en términos de presencia y ausencia.

### 6.3 Generación y limpieza de reglas

Una vez creado el dataset con ítems positivos y negativos, se aplica Apriori con umbrales similares a los usados previamente. Luego se realiza una limpieza en tres pasos:

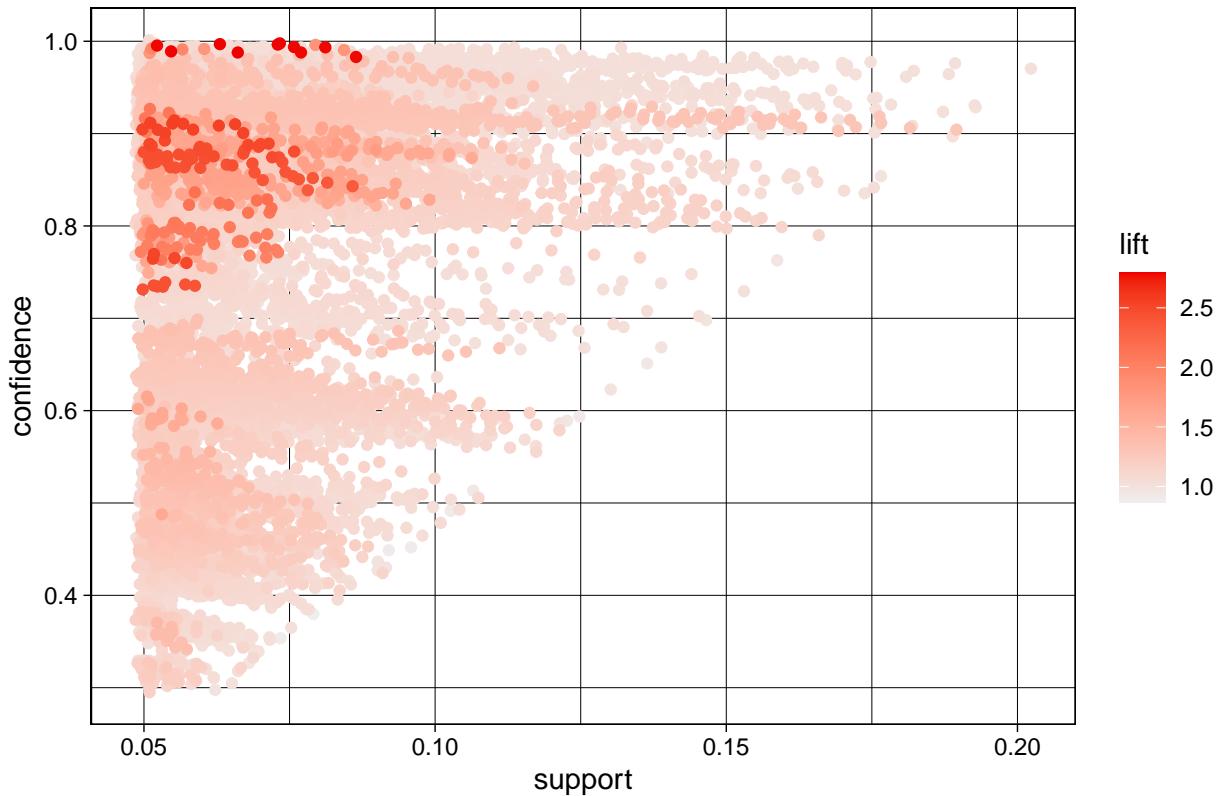
1. **Filtrado de reglas con confianza perfecta** (`confidence = 1`), porque suelen ser consecuencia de variables **mutuamente excluyentes** (por ejemplo, grupos de edad en formato binario) y aportan poco valor interpretativo.
2. **Eliminación de redundancia** (reglas equivalentes o contenidas en otras).

Antes de analizar reglas concretas con ítems negativos, se decidió centrar la exploración en **dos variables de interés** que fueron binarizadas durante el preprocesamiento: el **grupo de edad** (`age_group`) y el **nivel de conocimiento sobre la gripe H1N1** (`h1n1_knowledge`).

#### 6.3.1 Regla 7: Conductas preventivas y hábitos persistentes en el grupo de edad 55–64

En esta subsección, el foco se sitúa **exclusivamente en las reglas relacionadas con el grupo de edad**, dejando el análisis de `h1n1_knowledge` para un apartado posterior. Para evitar asociaciones triviales, se excluyeron reglas cuyo consecuente correspondiera a otras categorías de edad, de modo que las relaciones identificadas reflejen patrones conductuales reales y no simples consecuencias de categorías mutuamente excluyentes.

## Scatter plot for 14260 rules



Dentro del conjunto de reglas obtenidas tras incorporar ítems negativos y filtrar aquellas relacionadas con la edad, destaca especialmente la siguiente asociación:

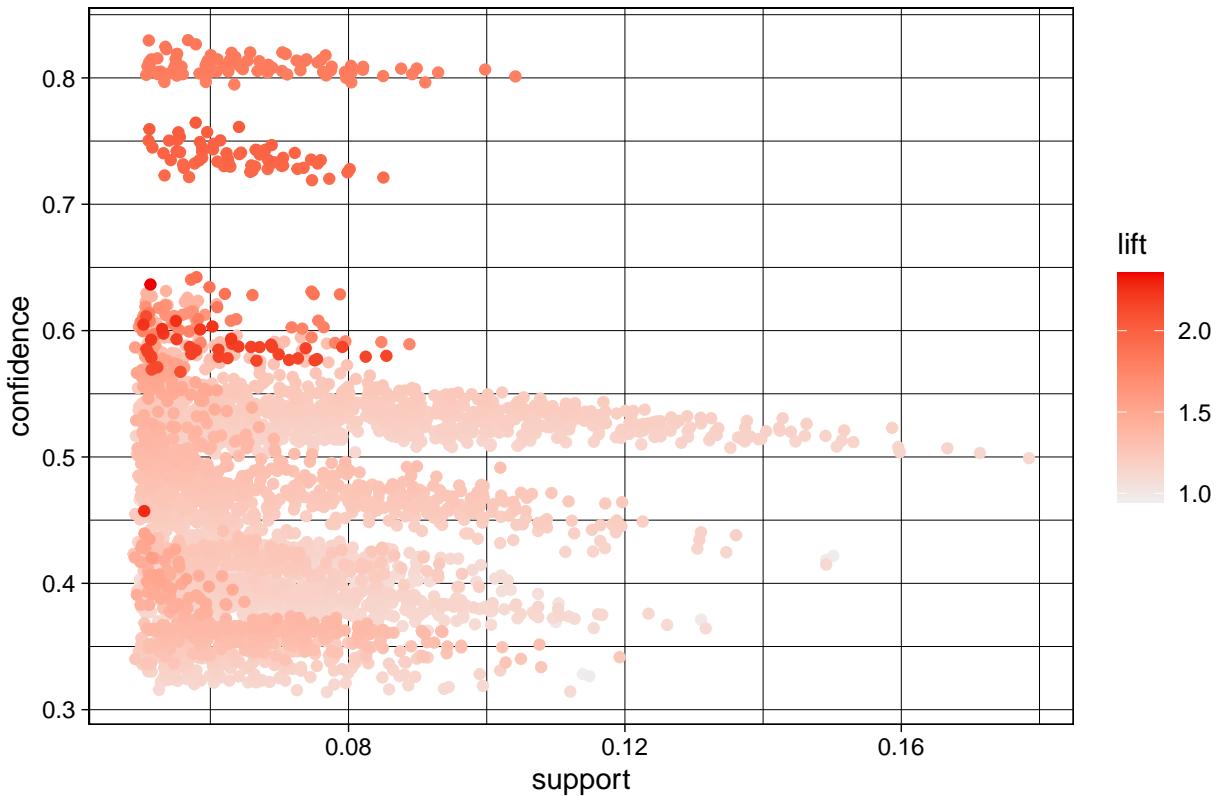
```
{behavioral_avoidance = yes, behavioral_wash_hands = yes,
age_group_age_55_64 = TRUE} ⇒ {behavioral_touch_face = yes}
```

Esta regla presenta un **soporte aproximado del 11.4%**, una **confianza del 82.3%** y un **lift de 1.22**. Aunque el lift es moderado, la regla resulta especialmente interesante desde un punto de vista interpretativo. En el grupo de personas de 55 a 64 años que declaran adoptar medidas preventivas claras (como evitar situaciones de riesgo y lavarse las manos con frecuencia) se observa, al mismo tiempo, una alta probabilidad de mantener un hábito difícil de controlar, como tocarse la cara.

### 6.3.2 Regla 8: Conocimiento, conducta preventiva y percepción de la vacuna H1N1

Dentro del conjunto de reglas obtenidas tras incorporar ítems negativos y filtrar aquellas relacionadas con el nivel de conocimiento sobre la gripe H1N1, destaca especialmente la siguiente asociación:

## Scatter plot for 3503 rules



```
{behavioral_wash_hands = yes,
h1n1_knowledge_a_lot = TRUE} => {opinion_h1n1_vacc_effective = very}
```

Esta regla presenta un **soporte aproximado del 10.8%**, una **confianza cercana al 35.2%** y un **lift de 1.31**. La regla indica que, dentro del grupo de personas que declara tener un alto conocimiento sobre la gripe H1N1 y que además adopta conductas preventivas como el lavado frecuente de manos, la proporción de individuos que considera la vacuna H1N1 como *muy efectiva* es mayor que en la población general.

El lift superior a 1 sugiere que la combinación de conocimiento elevado y comportamiento preventivo está asociada con una percepción más favorable sobre la eficacia de la vacuna, aunque dicha percepción no es dominante dentro del grupo, lo que refleja que las actitudes hacia la vacunación pueden diferir incluso entre personas bien informadas y con prácticas preventivas consolidadas.

## 6.4 Conclusiones

Durante el proceso de minería se identificaron numerosas reglas altamente evidentes, especialmente asociadas a variables dominantes como la vacunación previa o el acceso al sistema sanitario. Estas reglas resultaron útiles para validar el enfoque, pero aportan poca información nueva sobre la estructura del conjunto de datos.

Por este motivo, el objetivo principal del documento se centró en explorar reglas con menor fuerza global, pero con mayor valor interpretativo. Este enfoque permitió identificar patrones más sutiles que ayudan a comprender mejor la composición de la población, las actitudes y comportamientos de los encuestados, así como la coexistencia de prácticas preventivas y percepciones diversas frente a la vacunación.

En este sentido, las reglas analizadas no buscan maximizar métricas como el *lift* o la *confianza*, sino aportar una lectura más rica y matizada del conjunto de datos, poniendo de relieve relaciones que, aunque menos intensas, resultan relevantes para caracterizar a los distintos subgrupos estudiados.