

# Modelo predictivo de ventas

## Grupo Modelo

Raúl Peralta y José Antonio Sánchez

# Agenda

- Análisis exploratorio
- Modelo predictivo
  - Justificación
  - Metodología
  - Estimaciones
  - Resultados del modelo
- Evaluación de modelo (Insights)
  - Variables omitidas del modelo
  - Perfil de canastas de consumo
  - Perfil geográfico
  - Perfil de demanda
  - Relaciones entre SKUs
- Estrategia
  - Inventario poco efectivo
  - Ubicación de subdelegaciones (establecimientos ancla)
  - Planes de financiamiento para planeación

# Tabla resumen (metodología)

Error del modelo pronosticando febrero 2018	0.324
Herramientas utilizadas	R, Python, Prophet, pandas, geopandas, ggmap, ggplot, jupyterlab, joblib, readxl, tidytext, stringr, tidyr, scikit-learn
Información externa utilizada	DNUE, google maps
Transformación de variables	k-means, PCA, agregación por áreas de influencia

# Análisis exploratorio

# Resumen de tablas

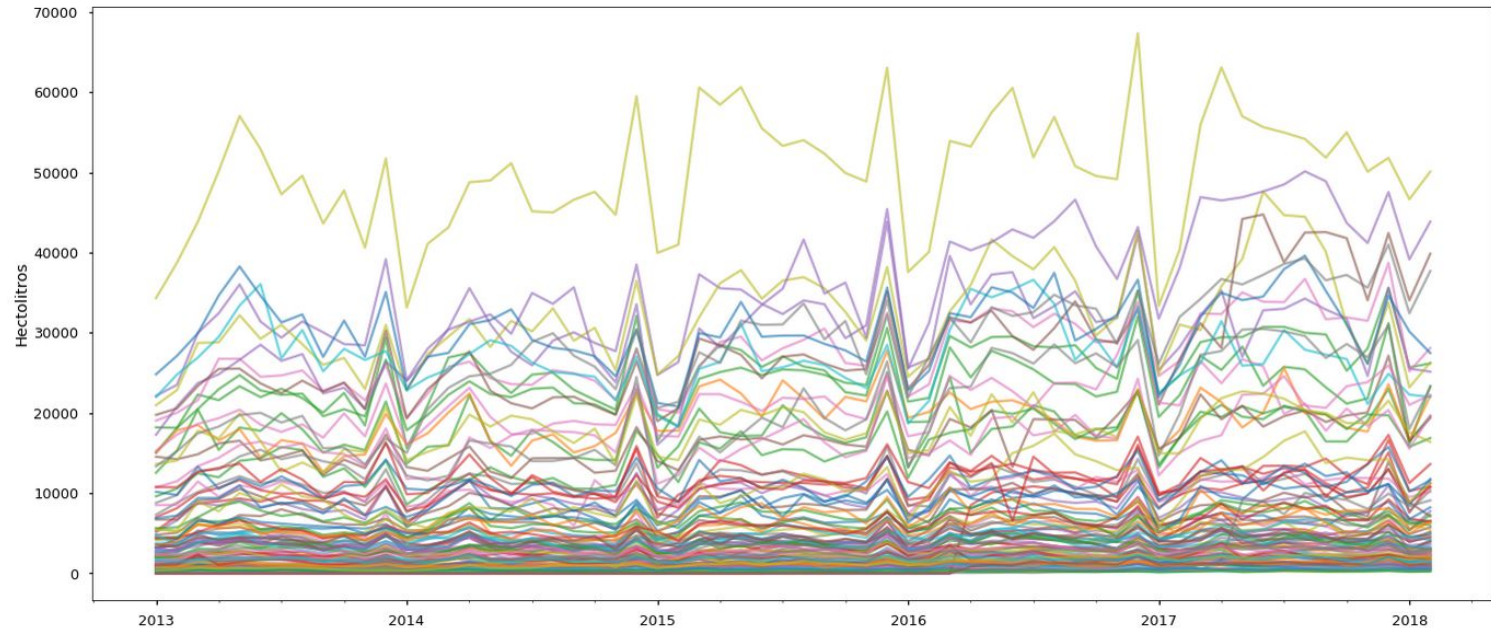
- Conteo de observaciones
  - 554,466 registros en el historial de ventas
  - 100 Subagencias en el catálogo con latitud y longitud
  - Existen 8,943 combinaciones entre subagencias y SKUs
- Observaciones con compra
  - Existen subagencias que nunca venden ciertos SKUs. De los 276 SKUs, 131 nunca se han vendido.
- Irregularidades
  - Hay dos Subagencias con coordenadas en el mar (10209 y 10059)

# Serie de tiempo de las subagencias

Posibles factores estacionales:

- Durante la primera mitad del año (temperaturas más altas) y en fechas decembrinas, se observa un alza en la venta de hectolitros
- Existe una caída pronunciada durante enero de todos los años (stock-up de detallistas)

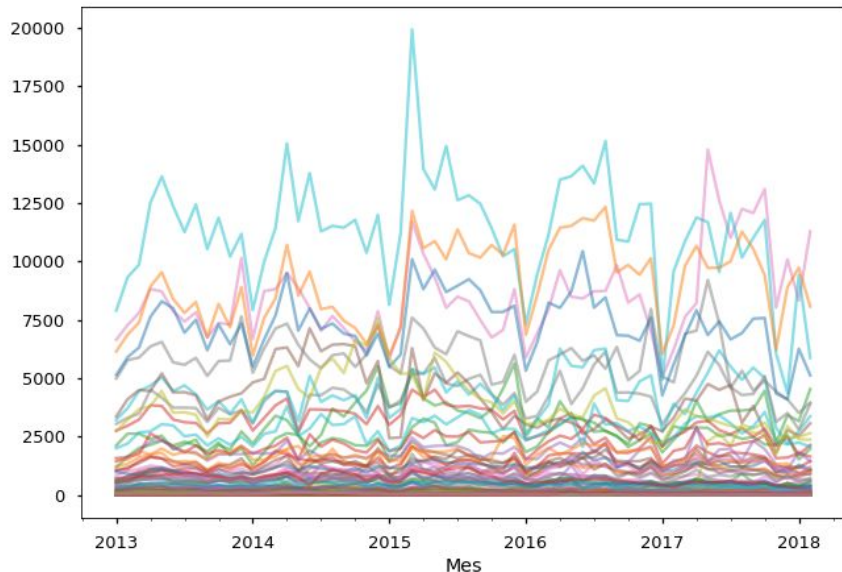
## Hectolitros mensuales vendidos por Subagencia



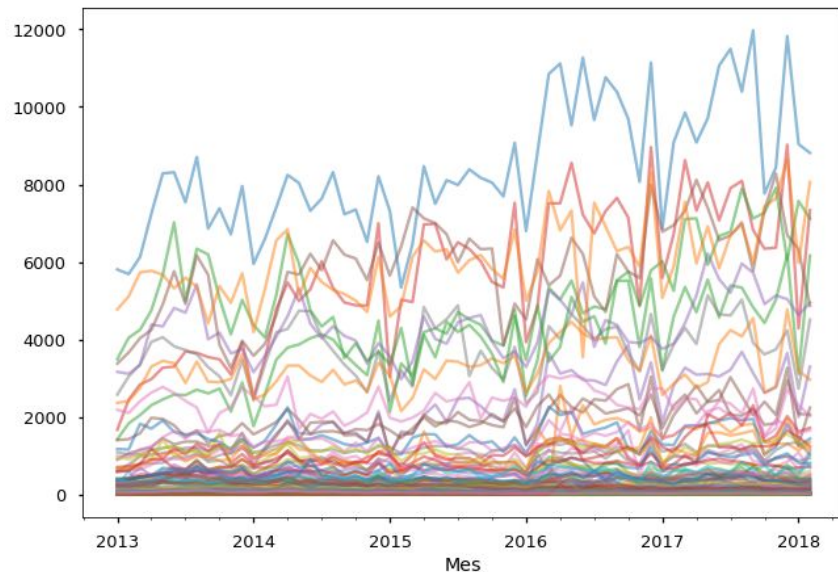
# Serie de tiempo de las subagencias

Si bien existe un patrón temporal general, los SKUs presentan comportamientos diferenciados. Esto complica el modelaje serie por serie.

**SKU: 00000025 (patrón estacional)**



**SKU: 00000016 (patrón no estacional)**



Modelo predictivo



# Modelo predictivo (justificación)

## -Restricciones:

- Series de tiempo con tendencias y estacionalidades diferentes por subagencia y producto
- Factores no observables
  - Variables operativas
  - Nuevos productos complementarios y sustitutos introducidos
  - Promociones

## -Alternativa:

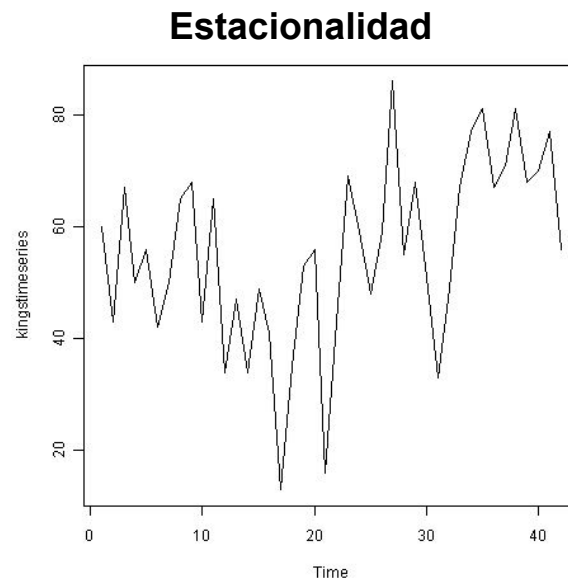
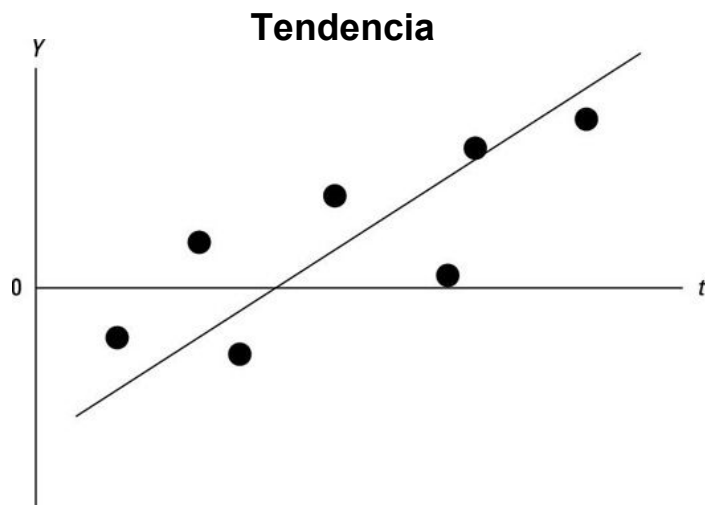
- Modelo de regresión no lineal capaz de aprender patrones temporales
- Este modelo nos permite estimar la venta de hectolitros sólo explotando la dimensión temporal



# Modelo predictivo (metodología)

Se utiliza un modelo de series de tiempo compuestas. Este puede ser desagregado en tres elementos:

- Tendencia: cantidad aditiva de hectolitros vendidos a través del tiempo
- Estacionalidad: Saltos periódicos dependientes de ciclos temporales

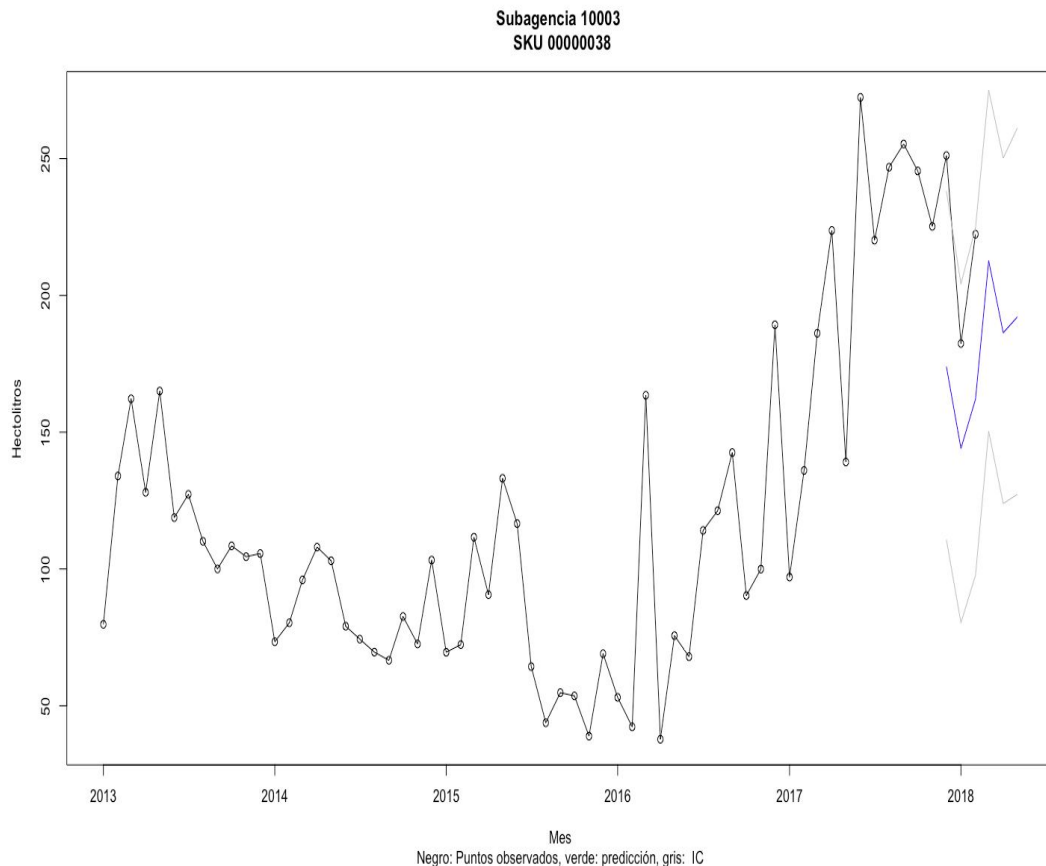


# Modelo predictivo (metodología)

- La predicción está modelada matemáticamente de la siguiente manera
  - $Y(t) \sim g(t) + s(t)$ ;
  - Donde Y representa la venta en el tiempo T, g es una función del tiempo que representa cambios periódicos, y s es una función del tiempo que representa cambios no periódicos o irregulares.
- Ventajas del modelo aditivo:
  - Permite incluir factores que sólo suceden una vez en la serie. Por ejemplo *stock-up* por parte de los detallistas
  - Permite producir agregaciones de tendencias anuales.
- Método de estimación:
  - Método basado en Hamiltonian Montecarlo models, (No U-Turn Sampler).

# Modelo predictivo (Estimaciones)

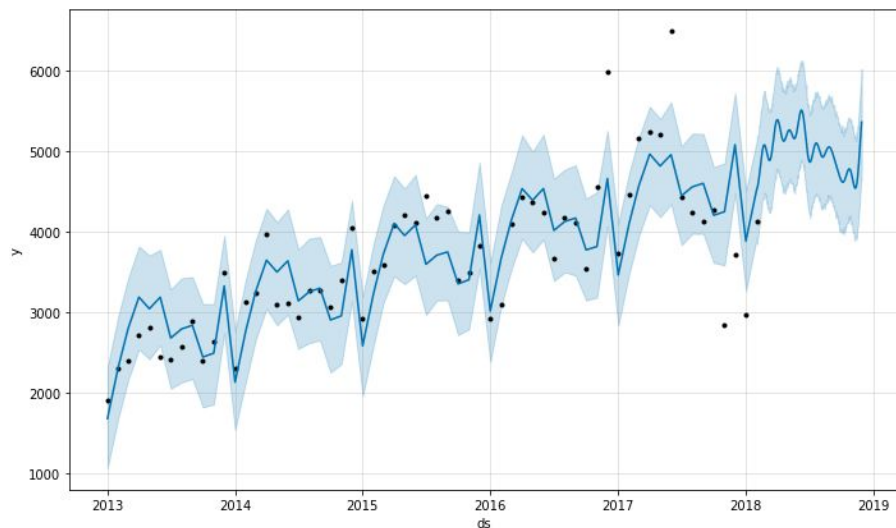
Puntos observados en negro.  
La línea azul representa la estimación del modelo sin usar esos periodos para el entrenamiento



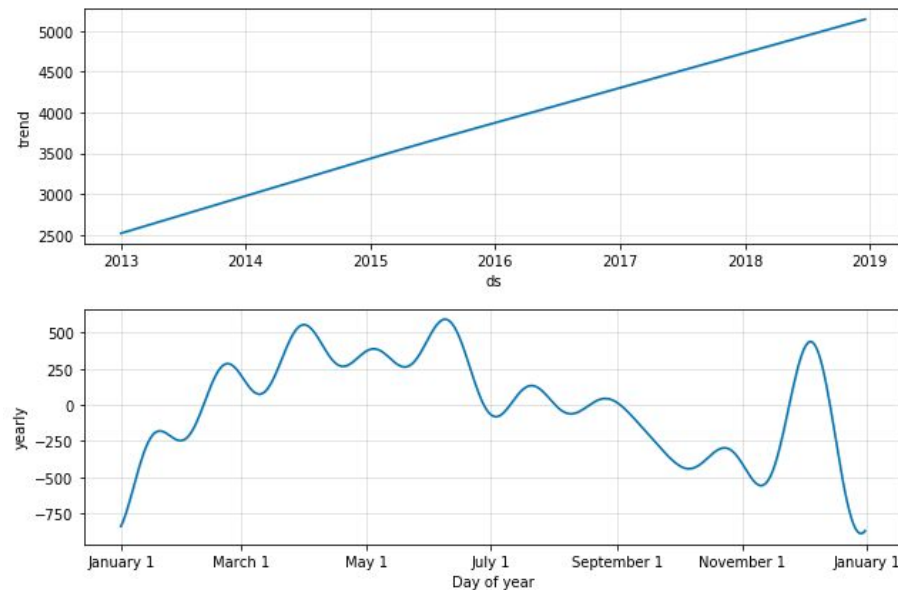
# Modelo predictivo (Estimaciones)

Subagencia: 10119

SKU: 00000030



Componentes



# Resultados del modelo

Periodo	Error relativo medio
Diciembre 2017	0.319
Enero 2018	0.392
Febrero 2018	0.324
Predicción <i>benchmark</i> con media por SKU/tienda	1.16

Se eliminaron las observaciones de diciembre a febrero para realizar el entrenamiento y evaluar estos meses en una prueba “ciega”

El error relativo de enero es más alto. podría ser efecto de *stock-up* por los clientes luego de las altas ventas de diciembre.

Por cada subagencia/producto se calculó la media total y se predijeron observaciones aleatorias con ésta

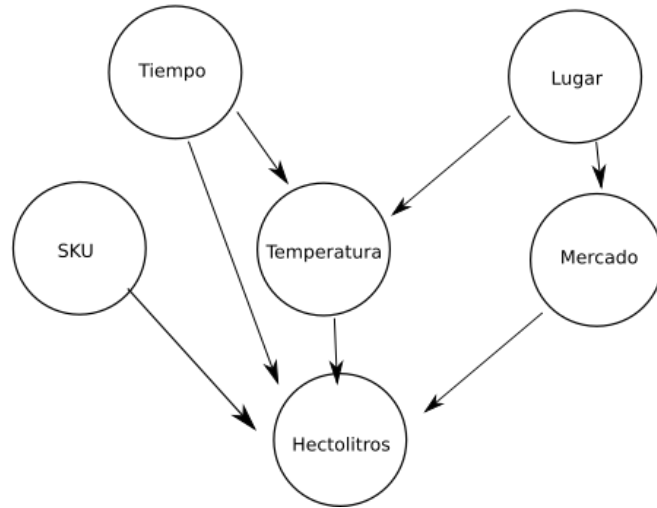
Evaluación de modelo (insights)

# Variables omitidas del modelo

- Perfil de canastas de consumo de subagencias
  - ¿Qué productos se consumen en cada una?
- Perfil de demanda de productos
  - ¿Quiénes son los clientes más probables de las subagencias?
- Perfil geográfico de subagencias
  - ¿Dónde están las subagencias más difíciles de predecir?
- Perfil de relaciones entre SKUS
  - ¿Qué SKUS tienden a subir y bajar juntos?



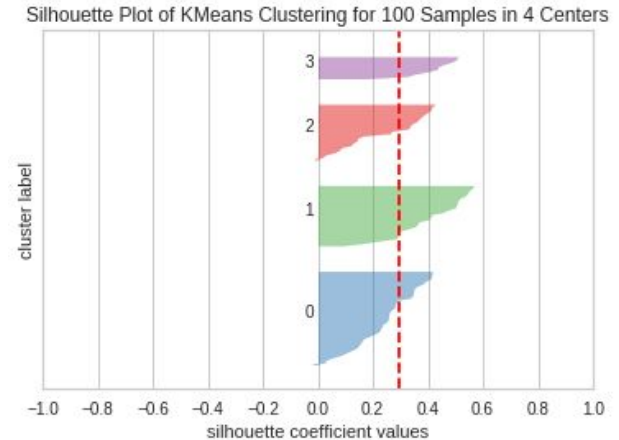
# Modelado (PGM)



# Perfil de canastas de consumo

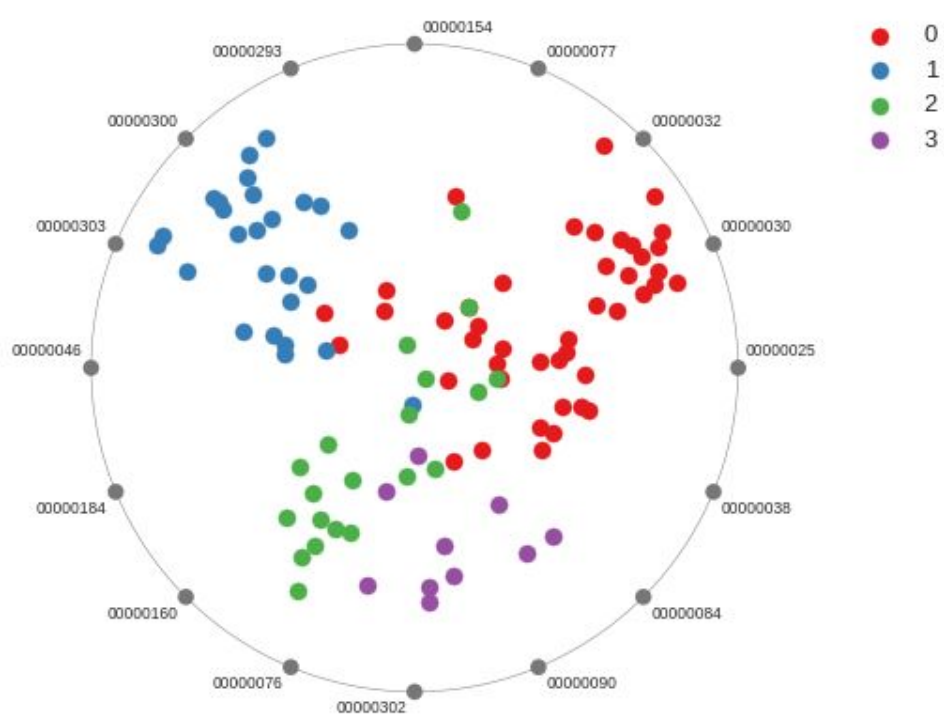


Segmentación no supervisada basada en la la distribución de proporciones de productos por subagencia



# Perfil de canastas de consumo

## SKUs más representativos por cluster



Cada punto representa una subagencia. Entre más cercano esté el punto a un SKU, más porcentaje representa éste de la venta total

# Perfil geográfico

La venta de hectolitros es mayor en áreas metropolitanas.



# Perfil geográfico (error del modelo)



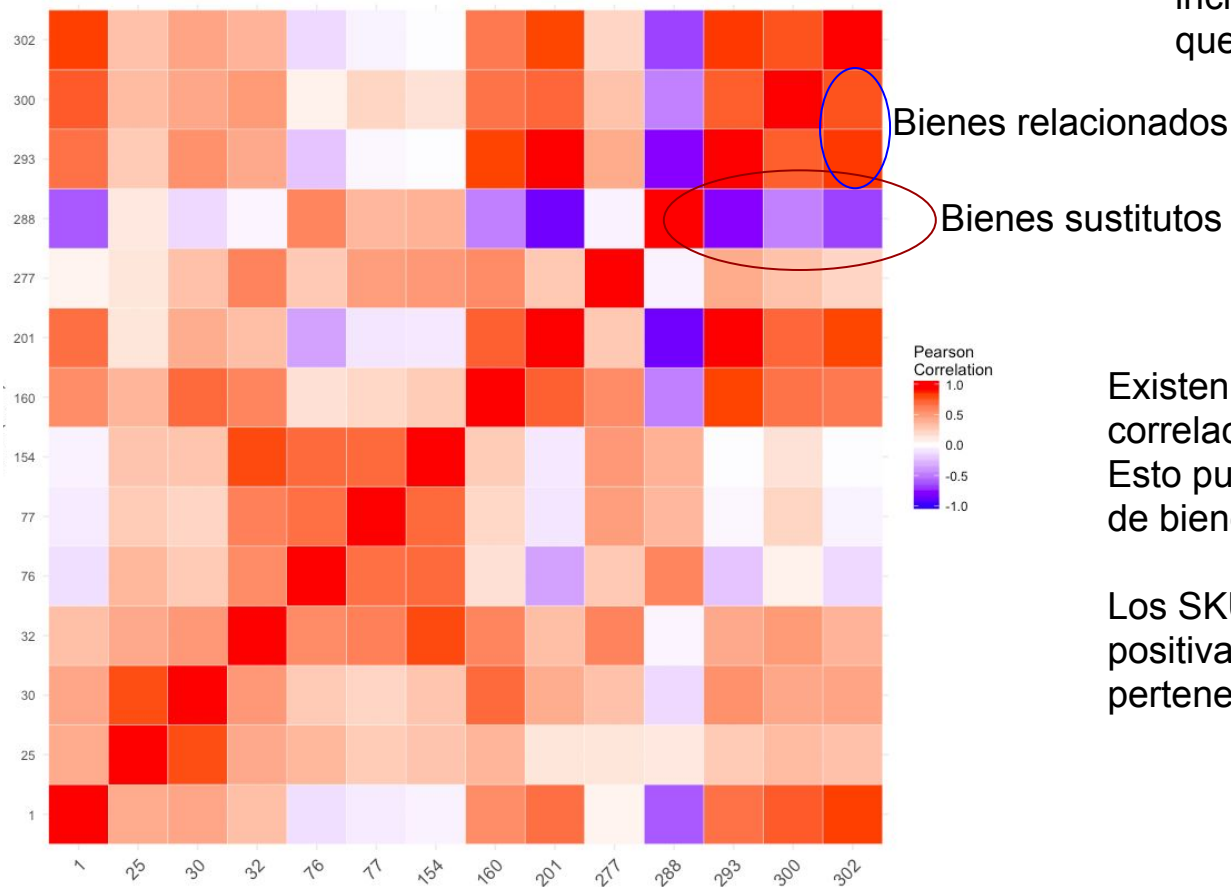
¿El lugar de la subagencia impacta a la estimación de ventas?  
El modelo no es capaz de identificarlo

Cluster de perfil de producto	Error relativo medio
Rojo	0.34
Verde	<b>0.50</b>
Turquesa	0.265
Morado	0.39

**Concentración geográfica de errores (región noroeste)**

# Relación entre tendencias de SKUs

Un modelo más sofisticado podría incluir predicciones por productos que tiendan a fluctuar juntos

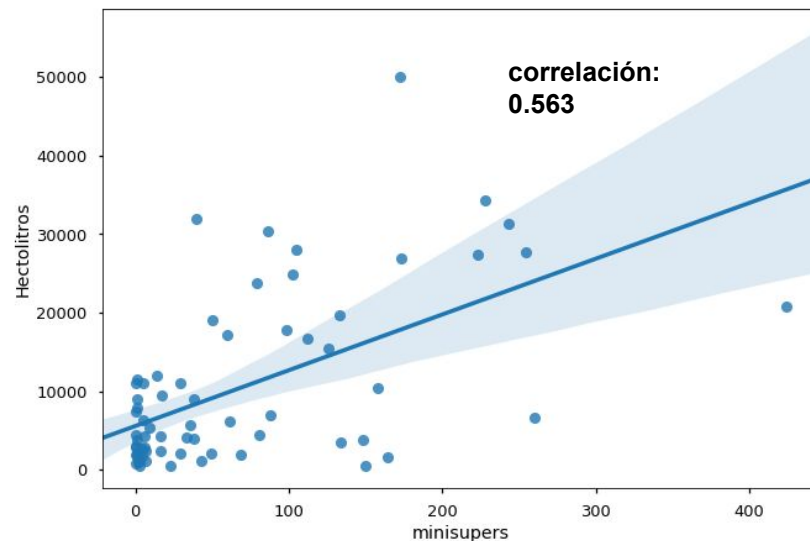
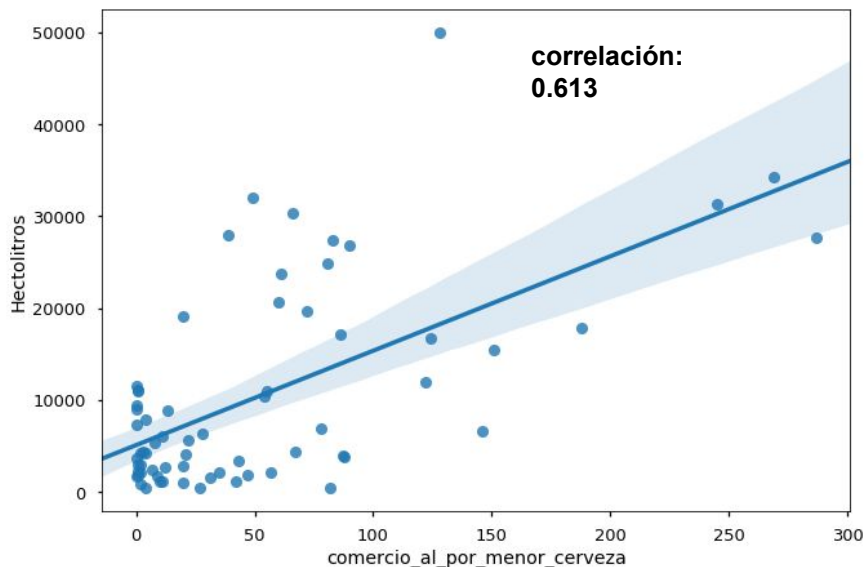


Existen productos que tienen correlación negativa en el tiempo. Esto puede deberse a que se trata de bienes sustitutos.

Los SKU con alta correlación positiva, es probable que pertenezcan al mismo producto

# Perfil de demanda

Usando el DNUE podemos observar que existe una correlación entre el número de comercios que venden cerveza a un radio de 5km y la cantidad de hectolitros que se venden. A partir de esta información se puede identificar qué zonas pueden tener alto potencial para explorar posibilidades de nuevas subagencias



Estrategia



# Costo de oportunidad de expansión por tipo de cliente

A partir del total de las subagencias, podemos estimar la aportación que tiene un establecimiento potencial para la subagencia.

Tipo de establecimiento	Efecto en Hectolitros mensuales de un establecimiento adicional	Valor P
Comercio al por mayor de abarrotes	<b>19,817</b>	0.0
Comercio al por mayor de cerveza	24,123	0.175
Comercio al por menor de cerveza	5,092	0.0
Comercio al por menor en supermercados	<b>17,646</b>	0.0
Restaurantes con servicio de preparación de alimentos a la carta o de comida corrida	6,699	0.0
Centros nocturnos, bares, cantinas y similares	1,052	0.0

# Planes de financiamiento en fin de año



Los clientes están transfiriendo su consumo de enero a diciembre. Una posible solución podría ser proveer de planes de financiamiento, que incentivaran a dosificar su gasto. Este modelo de suscripción haría más fácil la planeación de operaciones, por tener una venta asegurada a final de año.

# Inventario poco efectivo

- Problema: Existen 131 productos en el inventario que no se han vendido una sola vez en el periodo analizado. Estos productos seguramente producen costos de almacenaje y producción
- Esto nos dejaría con 145 productos efectivos