

PROPUESTA MEJORA FORECAST VENTAS E INVENTARIO

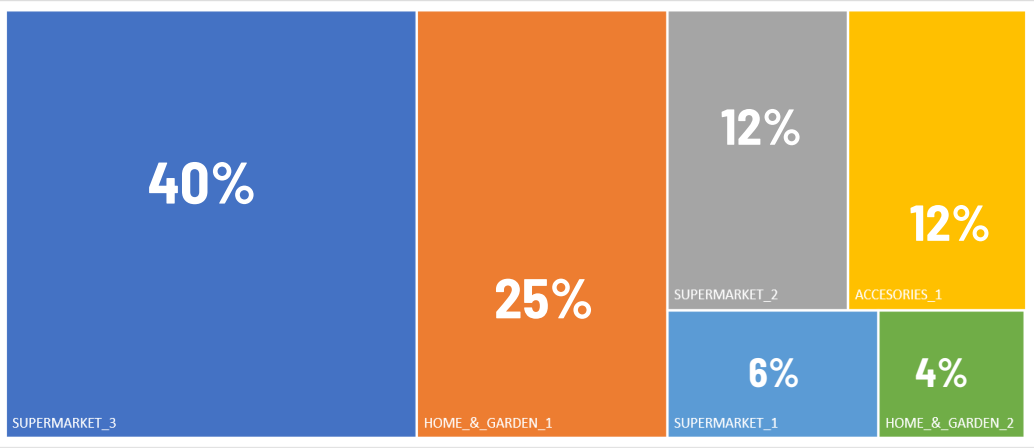
Agenda

- 1. Análisis Tendencia Ventas**
- 2. Clusterización**
- 3. Modelo Predicción de ventas**
- 4. Predicción de inventarios**
- 5. Propuesta despliegue modelo predicción inventarios**

ANALISIS TENDENCIA DE VENTAS

Análisis Ventas 2011 - 2016: Overview

El 80% de las ventas las generan 1.267 referencias, el 41% del total de referencias (3049 en total). La categoría con mayor porcentaje de ventas es "SUPERMARKET" con un 52%.



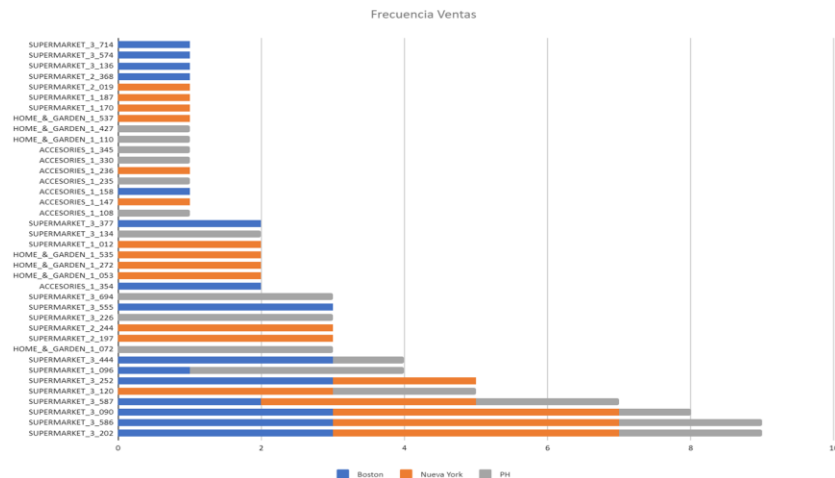
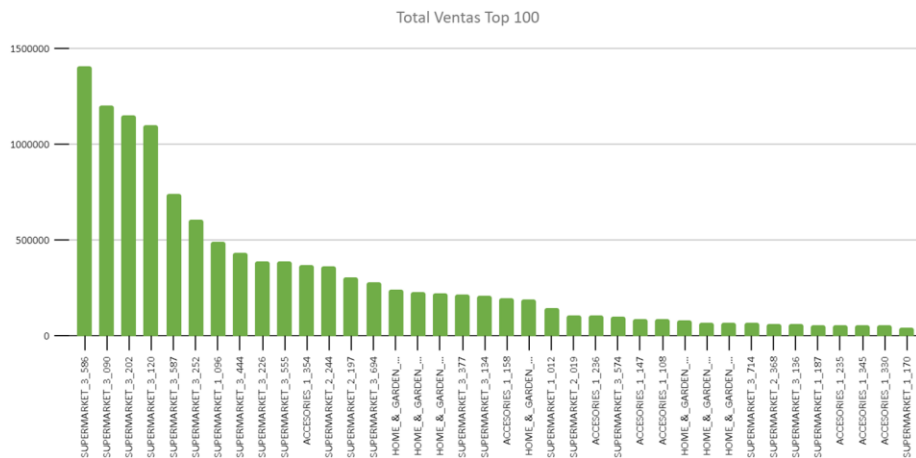
- De las 1.267 referencias, un 65% se concentra en los departamentos "SUPERMARKET_3" y "HOMER._&_GARDEN_1" con un 65% en total.
- La cateogría "SUPERMARKET" concentra la mayor participación con un 58%.
- Nueva York es la ciudad con el mejor performance de ventas totales durante los últimos 5 años. También es la ciudad en la cual se concentra el mayor número de referencias con mayores ventas.

- El 59% de las referencias restantes (1.793) aportan únicamente el 20% de las ventas totales.
- 12 de las 1.782 referencias no presentaron ventas en el último año y en al menos en 3 de los anteriores años.
- En promedio, las 1.782 referencias vendieron 20M € en los últimos 5 años. Para el 80% es de 118 M€



Análisis Ventas 2011 - 2016: Top Items

De los 100 top productos analizados (10 referencias x tienda), 38 son referencias únicas y 22 se repiten en las 3 regiones.



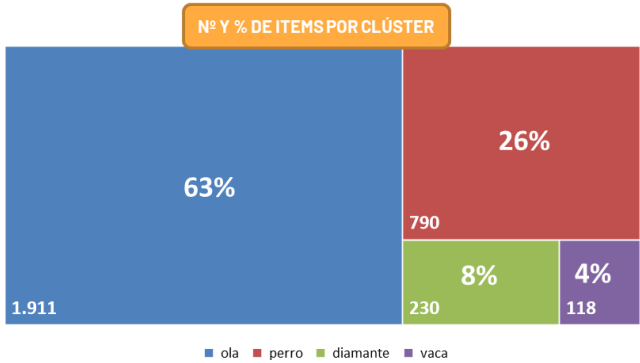
- Se identifica un alta concentración de ventas en los departamentos **"SUPERMARKET_3"** y **"H&G_1"** con un 40% y 25% respectivamente.
- Las referencias **"SUPERMARKET_3_202"**, **"SUPERMARKET_3_586"** se venden en 9 de las 10 tiendas. Las **"SUPERMARKET_3_090"** y el **"SUPERMARKET_3_587"** en 8 y 7 tiendas respectivamente.
- Para la categoría **"ACCESORIES_2"**, solo hay 1 producto **"ACCESORIES_2_075"**.
- En el 2015 **"ACCESORIES_1_354"** presenta un comportamiento inusual en sus ventas.

CLUSTERIZACIÓN DE PRODUCTOS

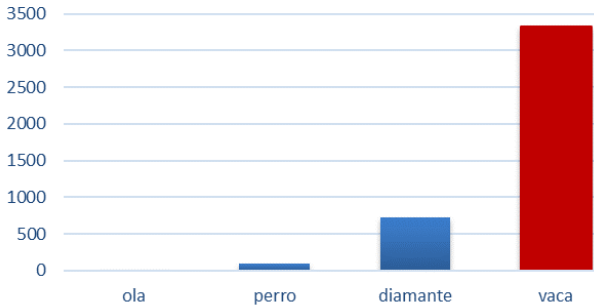
Características de los clústers

Cada uno de los cuatro clusters (0, 1, 2 y 3) son analizados con varias variables: ventas, crecimiento de ventas, precio medio y efecto de campañas. Las ventas se calculan dividiendo las ventas del grupo, por el número de ítems del clúster.

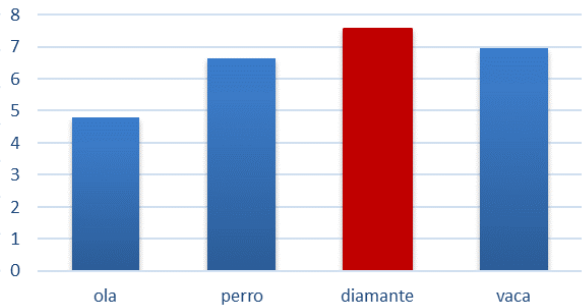
- Clúster 1 → Precios muy bajos, apenas suponen facturación. Son productos que se venden en las rebajas. Probablemente de otras temporadas. Producto **"ola"**.
- Clúster 2 → Precios medio-bajos. Efecto bajo de campañas, apenas suponen ingresos. Producto **"perro"**.
- Clúster 3 → Precios muy altos. Productos orientados a personas de alto nivel adquisitivo. Producto **"diamante"**.
- Clúster 4 → Precio medio-alto, el más vendido. Supone 80% facturación por ítem. Crece a un ritmo altísimo. Producto **"vaca"**.



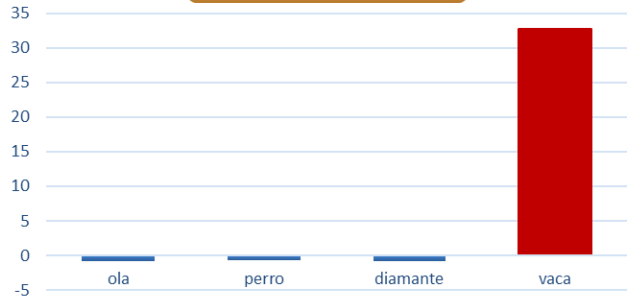
VENTAS POR ÍTEM Y CLÚSTER



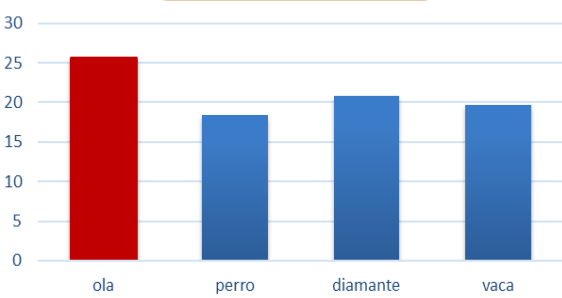
PRECIO MEDIO POR PRODUCTO



CRECIMIENTO VENTAS %



EFFECTO CAMPAÑAS %



Benchmarking entre tiendas y recomendaciones

No es posible hacer una sólida aproximación a identificar como de similares son las tiendas unas de otras ya que 10 tiendas son totalmente insuficientes para identificar patrones entre ellas.

Clúster **Ola** → Precios muy bajos, apenas suponen facturación. Son productos que se venden en las rebajas. Probablemente de otras temporadas.



Recomendación -> **Mantener** el número de productos y el precio pues sirven para eliminar stocks de temporada.

Clúster **Perro** → Precios medio-bajos. Efecto bajo de campañas, apenas suponen ingresos.



Recomendación -> **Eliminar** estos productos. No aportan absolutamente nada.

Clúster **Diamante** → Precios muy altos. Productos orientados a personas de alto nivel adquisitivo.



Recomendación -> Mantener el número e **incrementar precio**. Son productos de alta gama y este cliente no mira precio sino calidad.

Clúster **Vaca** → Precio medio-alto, el más vendido. Supone 80% facturación por ítem. Crece a un ritmo altísimo.



Recomendación -> **Incrementar el número** de estos productos pues aún son pocos comparado con otros clúster. Seguir con este precio para no parar el enorme crecimiento

PREDICCIÓN VENTAS

Modelo, metodología y variables significativas



- Mezcla de variables categóricas y numéricas
- Bases de datos con muchas observaciones
- Eficiente y fácil de utilizar

METODOLOGÍA:

- Estudio de las referencias con mayores ingreso

Se han seleccionado aquellas referencias con ingresos superiores a 99.016€.

- Estudio agregado con la información de todas las ciudades y todas las tiendas.

El número de observaciones totales es igual a 405.603, cuando el inicial era de 5.345.528.

VARIABLES IMPORTANTES:

- Ventas de los años 2014, 2015, 2016
- Ventas en días festivos
- Ventas el fin de semana
- Ventas durante la temporada de verano
- Ventas durante la temporada de Navidad

Se utilizaron las variables de venta agregadas a partir de las variables iniciales:

- Ventas totales por tipo de producto, por tienda, por ciudad

Resultados predicción agregadas y por tienda/ciudad

MODELO DE PREDICCIÓN AGREGADO:

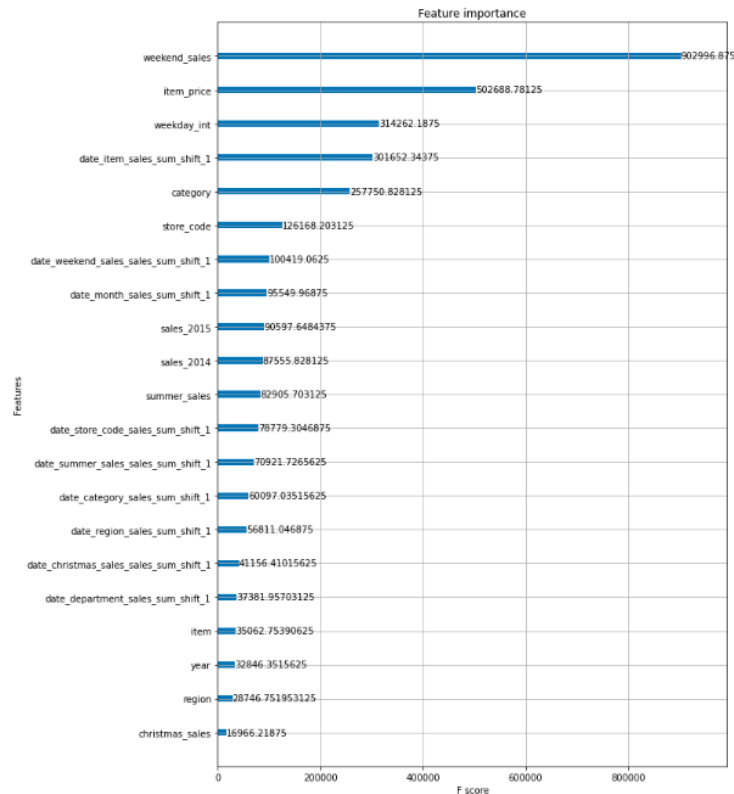
- Error del modelo → RMSE = 4.75
- TOTAL DE VENTAS PRÓXIMO AÑO → 7780999.0
- AUMENTO MENSUAL → 11%

MODELO DE PREDICCIÓN POR CIUDAD (NY):

- Error del modelo → RMSE = 0.00237
- La desventaja principal es la reducción del número de observaciones.

MODELO DE PREDICCIÓN POR TIENDA (NY_1):

- Si en el modelo anterior ya teníamos un error mayor, principalmente causado por la reducción del número de observaciones, aquí aún va a ser mayor la causa.
- NO ES NECESARIO GENERAR UN MODELO
- Además, tenemos RIESGO DE **OVERFITTING!!**



PREDICCIÓN INVENTARIO

Modelo, metodología y variables significativas



- Mezcla de variables categóricas y numéricas
- Bases de datos con muchas observaciones
- Eficiente y fácil de utilizar

METODOLOGÍA y VARIABLES IMPORTANTES:

- Las mismas que en el modelo de predicción de ventas

DIFERENCIAS ENTRE LOS DOS MODELOS:

- Predicción semanal, NO mensual

Motivos: Por capacidad de almacenaje y gestión logística es mejor tener un sistema semanal de reposición de stock. Por esto, el cálculo tiene que ser semanal.

DIFERENCIAS ENTRE LOS DOS MODELOS:

- Error del modelo agregado → **RMSE = 3.5**
- Error del modelo por ciudad (NY) → **RMSE = 1.57**
- Error del modelo por tienda (NY_1) → **RMSE = 0.000162**

Nos quedamos con el modelo agregado para hacer el cálculo de STOCK.

PROPUESTA DESPLIEGUE MODELO REPOSICIÓN INVENTARIO Y EVALUACIÓN DEL MODELO

Un despliegue, tres fases:

Automatización y Presentación de Herramienta a Negocio

- Objetivos: dar a conocer la herramienta y automatizar y mejorar procesos.
- El modelo se sigue actualizando de forma manual por parte del equipo de analítica con el objetivo de proveer estudios a las diferentes áreas de la empresa

1

Reentrenamiento semanal

- Objetivos: mejorar predicciones de producto-tienda y usabilidad por parte de los usuarios finales
- El modelo se actualiza cada semana y se envían predicciones a las personas interesadas.

2

Reentrenamiento On-Demand

- Objetivos: mejorar predicciones y liberar carga de trabajo al equipo de analítica.
- Construcción de un flujo de servicios en AWS:
 - Preprocesamiento de datos y reentrenamiento del modelo automatizado
 - Aplicación Web para realizar peticiones de reentrenamiento por parte de los usuarios
- Principales problemas: costoso y con un valor añadido no tan incremental.

3



Propuesta evaluación del modelo de reposición de inventario

Evaluaciones analíticas:

- Analisis agregados
- A/B Tests
- Grado de utilización de usuarios

Opinión de los usuarios:

- Interfaz
- Utilidad
- Productividad
- Satisfacción en el trabajo

GRACIAS

ANEXOS

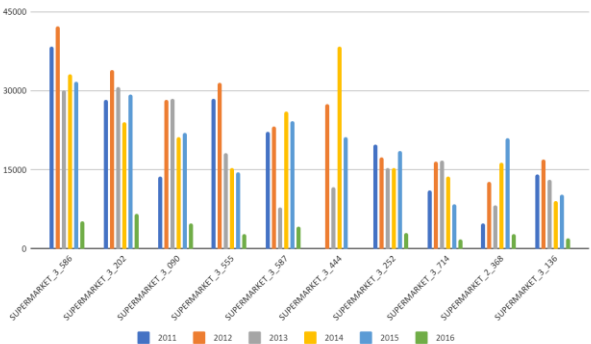
ANALISIS TENDENCIA DE VENTAS POR TIENDA

Análisis Ventas 2011 - 2016: Boston

El producto **"SUPERMARKET_3_586"** es el más popular en la región de Boston. Seguido por **"SUPERMARKET_3_202"**.

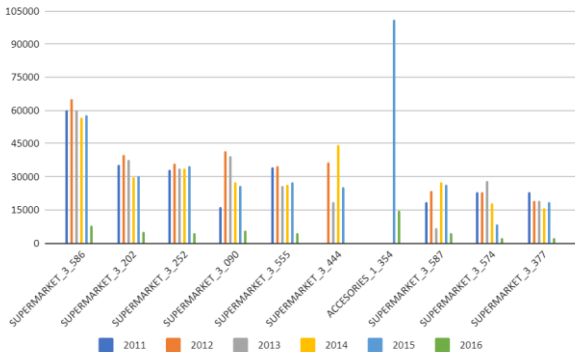
Boston 1

15.734.990 €



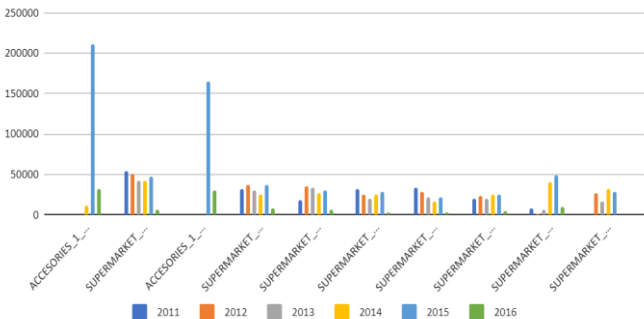
Boston 2

20.494.778 €



Boston 3

17.786.497 €



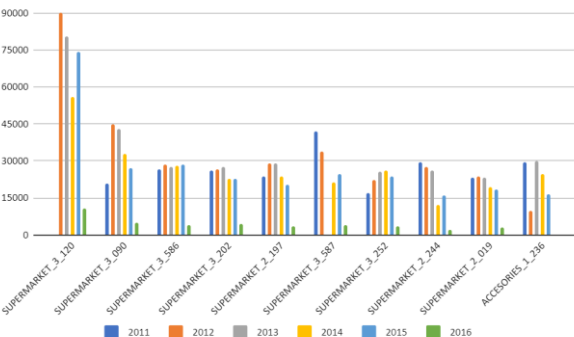
- Las ventas de los 10 mejores productos caen en promedio un 84% para las 3 tiendas en Boston.
- El 2012 y 2014 fueron en promedio los años con mejor performance.
- Aproximadamente el 41% de los productos para las 3 tiendas no han aumentado en ventas desde el 2011.
- En el 2015 **"ACCESORIES_1_354"** presenta un comportamiento inusual en sus ventas. Aportando un 28% en Boston 2 y un 33% en Boston 3.

Análisis Ventas 2011 - 2016: Nueva York

NY 3 se presenta como top performance en NY con un % de participación del 44% sobre el total de las ventas. NY 4 y NY2 participan con un 12% y 14% en las ventas YTY de la ciudad.

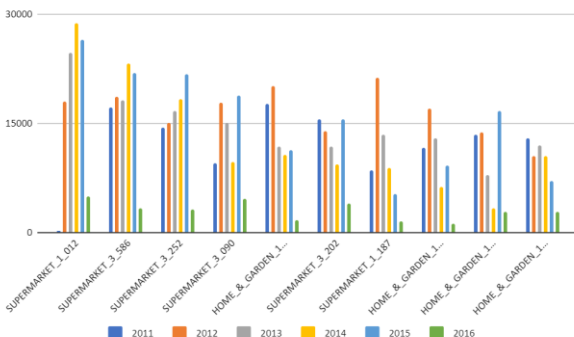
NY 1

22.621.162 €



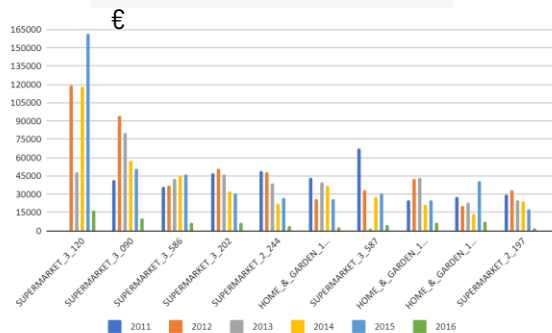
NY 2

17.427.753 €



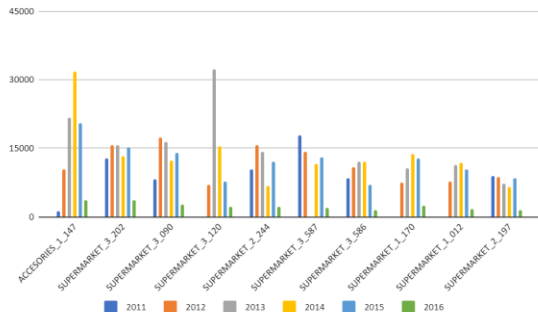
NY 3

30.194.950



NY 4

12.220.085 €



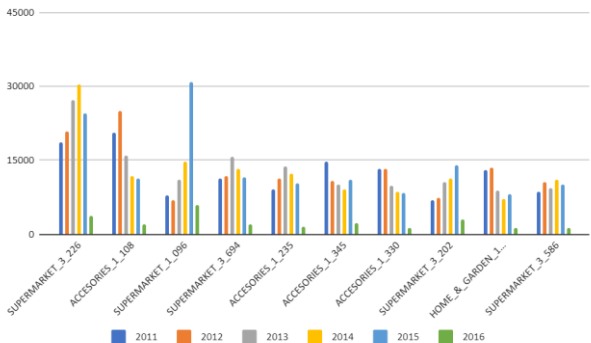
- "SUPERMARKET_3_120" es el producto con mayor ventas y mejor performance.
- Tanto NY3 como NY2 comparten 4 de los 10 productos con mejores performance por tienda.
- El 2012 y 2015 fueron en promedio los años con mejor performance con 1.122.404€ y 1.003.728 €
- Únicamente en NY3 y NY2 productos de H&G entran dentro del top 10, representando un 14% de ventas en total para las 2 tiendas.

Análisis Ventas 2011 - 2016: Philadelphia

Las tres tiendas en Philadelphia mantienen un comportamiento de ventas totales similar. A diferencia de Nueva York y Boston, el mix de productos populares en Philadelphia varía entre las 3 tiendas.

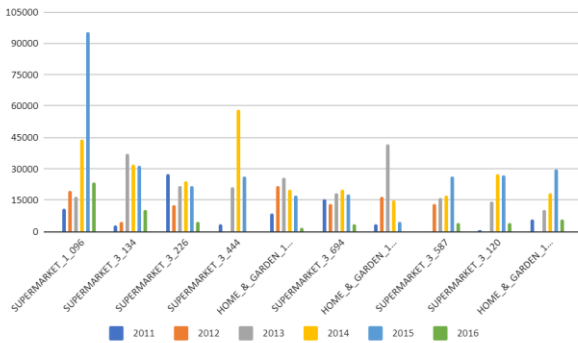
PH 1

14.600.474 €



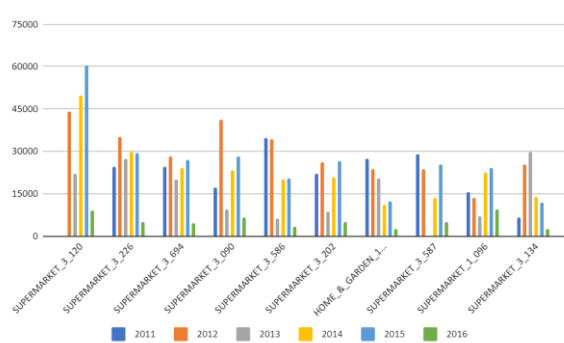
PH 2

16.507.798 €



PH 3

17.427.753 €



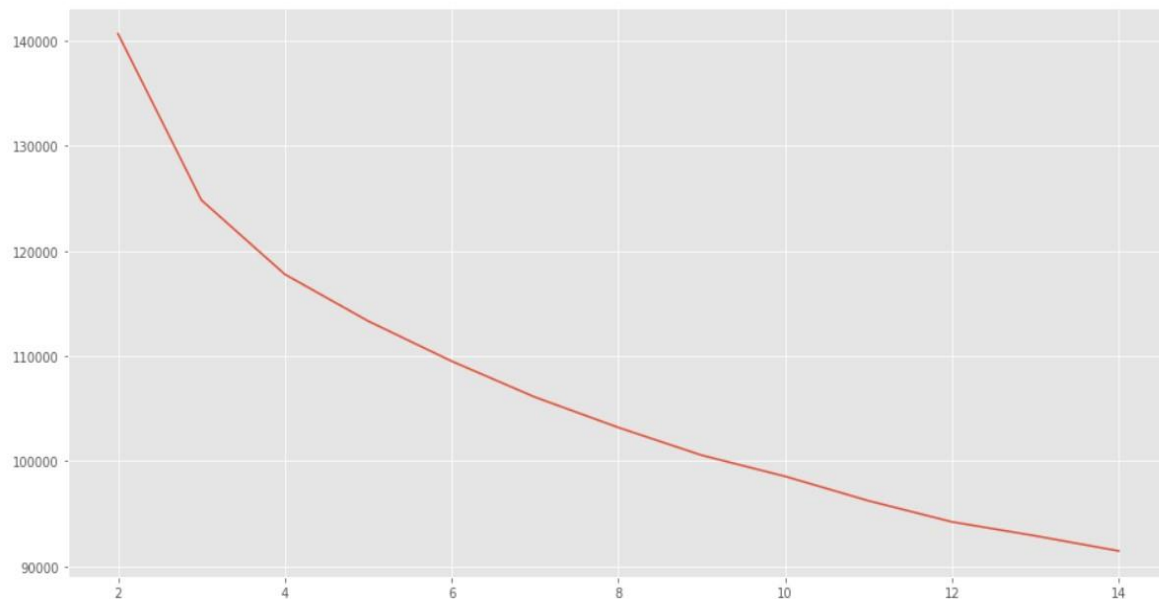
- “SUPERMARKET_3_226” es el producto con mayor ventas y mejor performance.
- Las 3 tiendas comparten 17 productos con mejores ventas. 7 de ellos representan el 80% de las ventas para la region.
- Los productos de la categoría “ACCESORIES” solo se venden en PH 1.

CLUSTERIZACIÓN

Clustering de los productos

Son 3.049 los productos a agrupar en varios clústeres. El número de grupos se define en la curva Elbow del algoritmo K-means neighbor.

Variación de la dispersión de los clústers en función de la k



En la gráfica se puede observar que el número óptimo de clústeres se produce en el codo.

En este caso el codo está en la k número 4. Por ello decidimos que el número de clústeres son 4.