

# DATA SCIENCE

## PREDICCIÓN DE VENTAS FUTURAS

CESAR HERNANDEZ MARTIN | ENERO 2021 | SECTOR RETAIL



## 1 .LA CIENCIA DE DATOS EN EL SECTOR RETAIL

Dominar la ciencia de los datos, clave para el sector -retail- El análisis de datos juega un papel central a la hora de permitir a los retailers alcanzar un nuevo estadio de perfección en la oferta de sus servicios.

En un contexto en el cual las barreras entre el mundo online y físico se difuminan y donde **el consumidor es más exigente que nunca**, los retailers se enfrentan al reto de ofrecer una experiencia de compra completa, enriquecedora y personalizada, con independencia del canal.

Poner la tecnología al servicio del negocio del retailer supone **reducir la falta de existencias de artículos clave**, comercializar datos y activos de medios y agilizar los flujos de trabajo de comercialización y operaciones de la tienda para reducir los gastos operativos. Igualmente confiar en la ciencia de datos permite **reducir las inversiones excesivas en iniciativas de promociones y precios ineficaces** o agilizar la gestión del surtido, así como optimizar la gestión precios y promociones, de surtido, o la localización de espacios

Conocer las razones por las que falla una categoría de producto, qué segmentos de consumidores no están actuando como se había pronosticado, o identificar y **hacer un seguimiento adecuado a los mejores consumidores de** un retailer para continuar trabajando su lealtad, son algunos ejemplos de cómo el análisis de datos puede contribuir a impulsar el negocio del retailer. La ciencia de datos también permite al retailer **entender de donde proviene el decrecimiento de una categoría** específica, identificar el grupo de mejores clientes de un Retail para continuar trabajando en su lealtad, o poder reconocer los diferentes segmentos de clientes en cuanto a sensibilidad al precio y así adaptar la oferta de surtido para suplir sus necesidades.

## 1 . INTRODUCCION PAG 1-3

La ciencia de datos en el sector retail.  
Es necesario, conocer y entender los  
hábitos de compra del cliente.

## 2 . DESCRIPCION PAG 4-16

Predicción de ventas futuras de 21700 artículos de informática y 60 tiendas físicas de la cadena 1C Company

### 3 . PROBLEMAS Y SOLUCIONES PAG 16

Volumen de datos, ingeniería de variables, minería de datos, series temporales, elección del modelo de predicción, etc.

## 4 . CONCLUSIONES PAG 17

La ciencia de datos se antoja imprescindible para el sector retail, la obtención de datos en tiempo real, es el gran reto del sector

Los negocios que usan el **BIG DATA**, son capaces de superar a sus competidores en un 85%, en el **crecimiento de las ventas** y en un 25% en el margen bruto

# DOMINAR LA CIENCIA DE DATOS EN EL SECTOR RETAIL

No es la primera vez, ni será la última, que los expertos del **sector del retail** hablen del cambio de paradigma en el **papel de los consumidores en los procesos de compra**. De un tiempo a esta parte, han dejado de ser el tradicional sujeto pasivo para convertirse en sujetos que demandan autenticidad, novedad, conveniencia y creatividad en su experiencia de compra. Y que tienen el poder para que esa exigencia se convierta en motor de cambio en el sector. A pesar de ello, no es arriesgado decir que no todos los jugadores están sabiendo adaptarse al cambio en el terreno de juego.

Los **retailers más innovadores** han comenzado a asumir que la tecnología no es solo un complemento, sino que se ha convertido en parte fundamental de la experiencia de compra, y que puede contribuir de forma fundamental en el diseño de nuevas experiencias de compra que aumenten la lealtad de estos hacia nuestras marcas. Poco a poco tecnologías exponenciales como **la inteligencia artificial, robótica y la realidad virtual están empezando a abrirse paso en las tiendas**. Y otras mucho más asentadas, como las redes sociales o las aplicaciones móviles, están provocando cambios significativos en todo el sector.



“ EL 65% de los Retailers no sabe sacarle partido a los datos, por falta de las herramientas adecuadas ”

-Estudio Velocity WorldWide

La implementación de algoritmos que procesen correctamente la información agregada será fundamental para que las empresas puedan resolver sus problemas de negocio, mejorar el rendimiento de sus servicios y satisfacer a la cartera de sus clientes.

**El futuro del Sector Retail pasa por la obtención de datos en tiempo real.** Se espera que la cifra de **dispositivos conectados al internet de las cosas, aunmente de los 26 a los 75 mil millones**, para 2025. Una encuesta de PwC, señala que un 32% de los consumidores ya utilizan la voz para comprar y que un 80% están muy satisfechos con la experiencia.

Las posibilidades que los grandes **datos aportan a los minoristas** son prácticamente interminables. De hecho, **parece complicado tomar decisiones sin contar con este conocimiento porque** ¿cómo se puede diferenciar cuál es el orden óptimo de las estanterías o cuántos niveles deben presentar? ¿De qué forma se pueden configurar las rutas e los clientes dentro de la tienda? ¿Cómo descubrir el tiempo que se necesita mantener una oferta 3x2 para que un producto se consolide en la cesta de compra de los clientes? ¿Qué técnicas ayudan a aumentar las ventas de las marcas blancas?

Con el conocimiento y la tecnología de su parte, el sector retail es uno de los más aventajados en el aprovechamiento big data, como muestran las siguientes 10 siguientes aplicaciones:

## **1. Predicción de ventas.**

La implementación de la estrategia de previsión de ventas depende de la elección de una potente solución analítica. La herramienta elegida, mediante la aplicación de técnicas de análisis de datos, así como de machine learning, será capaz de predecir la demanda futura de los bienes comercializados.

## **2. Diseño distribución productos.**

El diseño de la distribución de los productos en una tienda es muy importante, no sólo para mejorar el ambiente y crear una atmósfera concreta, sino también para aumentar las ventas, facilitando el acceso a determinados bienes o grupos de consumibles.

## **3. Clasificación productos.**

Es necesario llevar a cabo un análisis en detalle para clasificar los productos y obtener información acerca de la función de cada uno.

## **4. Gestión del inventario.**

Las soluciones de gestión de inventario tienen en cuenta los tiempos de espera (desfases en la adquisición al proveedor) y garantizan la disponibilidad a la hora de realizar la planificación, utilizando técnicas estadísticas avanzadas para determinar el stock de seguridad y el punto en el que se lanzará el proceso de compra. Gracias a big data es posible conocer la severidad del desabastecimiento de un producto en comparación con otros, haciendo posible el cálculo de las dimensiones del stock de seguridad que es necesario mantener para determinados productos, que son más críticos.

## **5. Análisis de fidelidad de los clientes.**

El análisis de fidelidad es una de las estrategias clave de big data retail. El motivo no es otro que el hecho de que la adquisición de nuevos clientes es mucho más complicada y costosa que la retención de los clientes existentes, por lo que son éstos últimos quienes deben constituir una prioridad para el negocio.

## **6. Análisis de precios.**

Empleando técnicas de modelado predictivo que determinan las variaciones de precios y de sus umbrales, datos muy útiles para maximizar las ventas en función de la decisión del precio óptimo y los descuentos asociados a cada una de las marcas y referencias. Información ésta que también puede ayudar a determinar el precio base y el descuento aplicado para una promoción.

## **7. Marketing Mix**

Con el marketing mix, se consigue excelentes resultados tanto aplicada a los medios de marketing tradicionales, como a los de nueva generación y es especialmente necesaria para una buena gestión y planificación de ambos de forma combinada.

## **8. Análisis de la cesta de la compra.**

No todos compramos igual, aunque compremos lo mismo. Incluso dentro de la misma familia y con una lista de la compra idéntica surgen diferencias. En estos pequeños cambios se esconden las claves para adivinar tendencias, descubrir hábitos y comprender los comportamientos de los clientes.

## **9. Análisis de grupos de productos.**

El reporting es uno de los procesos más beneficiados de la interacción del sector retail con big data. El conocimiento del negocio que se obtiene gracias a las nuevas técnicas analíticas.

## **10. Segmentación de clientes.**

La segmentación de clientes divide el mercado en grupos que comparten características similares.

## 2. DESCRIPCION DEL PROYECTO – PREDICCION DE VENTAS

Sí, una de las grandes ventajas de aplicar tecnologías de Big Data en nuestro comercio es que podemos aplicar diversas técnicas a nuestros datos para predecir cómo **van a evolucionar en el futuro ciertas variables**. A este conjunto de técnicas estadísticas y de aprendizaje automático se les suele denominar modelos predictivos y se encuadran dentro del campo del análisis predictivo (*predictive analytics*). Los modelos de análisis predictivos, que se construyen usando técnicas matemáticas y de **inteligencia artificial**, permiten predecir cómo se comportará en el futuro una variable (las ventas de un producto) en función de una serie de variables predictoras (series temporales).





La previsión de ventas es la panacea del sector retail, ya que si obtenemos una buena predicción está sería útil por ejemplo para ajustar los procesos productivos y el stock necesario para atender a la demanda futura (rentabilización máxima del espacio), acercándonos lo máximo posible a la excelencia operativa: **sin sobreproducción y sin desabastecimiento**. Pero ésa es una visión enfocada en la cadena de suministro. Si lo miramos desde el punto de vista de marketing, en realidad no queremos que se cumpla la predicción del modelo de análisis predictivo, lo que queremos es **maximizar las ventas**. En otras palabras, lo que vamos a hacer es usar los modelos de análisis predictivos como herramientas de simulación del futuro, que nos permitan accionar las palancas adecuadas en el presente, y así conseguir maximizar nuestros resultados.

- Reposición de las existencias por categorías y referencias.
- Adecuación de las necesidades de abastecimiento, para evitar tanto el exceso de existencias en stock, como los problemas derivados del desabastecimiento.
- Aprovechamiento máximo de los picos de ventas semanales o de periodos especiales, gracias a la previsión de ventas a nivel de categoría y a nivel de referencia.

Por lo tanto, la clave está en detectar las variables (**series temporales**) que más influyen en nuestra predicción, usando un Big Data con capacidad de análisis predictivo. En este proceso, al usar modelos de análisis predictivos para tomar decisiones en el presente, lo que hacemos no es simplemente predecir de forma pasiva el futuro, sino que usamos los *insights* obtenidos (acerca de un posible futuro hipotético) **para construir un futuro más provechoso para nuestros intereses**. Pensemos, por ejemplo, en los modelos que predicen el abandono de clientes.

Estos modelos pueden tomar como entrada cientos de variables con posible poder pronosticador sobre el abandono (facturación presente y pasada, interacción por diferentes canales, comportamiento de la red social más próxima, cambios de comportamiento recientes, etc.). La salida del modelo será una probabilidad de abandono para cada uno de los actuales clientes. Si simplemente usáramos el modelo como un *insight*, podríamos descubrir qué factores son los que más influyen en el abandono de ciertos tipos de clientes. **Pero en realidad, lo que haremos será seleccionar al grupo de clientes con más valor para nosotros y también con más propensión de abandono con el objetivo de realizar una acción de retención sobre ellos**. Y aquí está la clave, al “accionar sobre el insight obtenido” estamos usando nuestra predicción para hacer precisamente que ésta no se cumpla.

### SERIES TEMPORALES:

	Economía	Producto interior bruto anual, tasa de inflación, desempleo, etc.
	Demografía	Nacimientos anuales, tasa de dependencia, etc.
	Meteorología	Temperaturas máximas, medias o mínimas, precipitaciones diarias, etc.
	Medio Ambiente	Concentración media mensual de nitratos en agua, alcalinidad media anual del suelo, emisiones anuales de CO2, etc.

“ Usaremos nuestra predicción para que esta no se cumpla”

“ No se trata de predecir el futuro, sino de cambiarlo a nuestro favor”



Nuestros modelos de análisis predictivos se alimentan de datos de entrada. Si entre estos datos no se encuentran las variables que realmente causan el cambio en la variable a predecir nuestro modelo será inútil.

La gran ventaja de las infraestructuras actuales de Big Data es que nos permiten construir modelos muy complejos en los que explorar gran cantidad de variables, aquí es donde entra en juego la ingeniería de variables. La creación de variables que según los datos afecten al comportamiento de la variable a predecir se antoja un trabajo de grandísima importancia, y que es en su mayoría el grueso del proyecto. Encontrar cuáles de ellas son las que realmente impactan en el futuro hipotético que estamos prediciendo y que éstas sean de buena calidad, es muy importante. En esencia, "un modelo es lo que come", y por muy buenos y sofisticados que sean los algoritmos que usamos, si los alimentamos con datos de mala calidad (con errores, datos nulos, incompletos, etc.) no podrán aprender a capturar correctamente las relaciones entre las variables y por lo tanto harán malas predicciones, provocando que tomemos malas decisiones.

**En este proyecto, trabajaremos con un desafiante conjunto de datos de series de tiempo que consta de datos de ventas diarias, amablemente proporcionados por una de las firmas de software más grandes de RUSIA: 1C Company.**

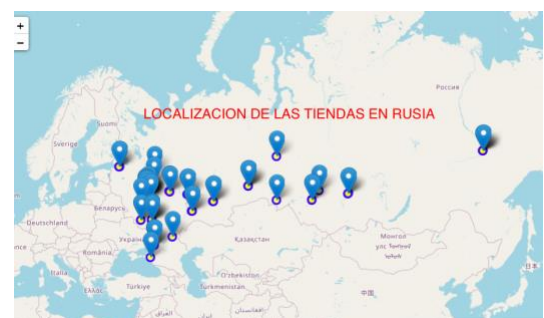
Conseguiremos predecir las ventas totales en unidades de cada producto, para el mes siguiente (a futuro).

1C Company (en ruso: Фирма 1С) es un desarrollador de software independiente, distribuidor y editor cuya sede se ubica en Moscú, Rusia. La empresa se especializa en el desarrollo, producción, mantenimiento, licenciamiento y venta de programas informáticos, servicios relacionados y videojuegos.

La empresa 1C es conocida por su labor de diseño y edición de videojuegos. Los títulos más populares desarrollados por la compañía son: Il-2 Sturmovik, Kings Bounty, Men of War y la serie de Space Rangers. 1C opera en calidad de editor de docenas de desarrolladores independientes y ha producido más de 100 títulos de videojuegos.

1C ocupa el octavo puesto entre las 30 mejores compañías de software en el ranking de mercados emergentes de PwC.

La empresa emplea a más de 1200 trabajadores. 1C opera a través de una amplia cadena de más de 10 000 socios empresariales en 25 países, incluyendo más de 7000 franquicias autorizadas de integradores de software 1C y revendedores de valor añadido (VAR), y más de 400 centros autorizados de entrenamiento.

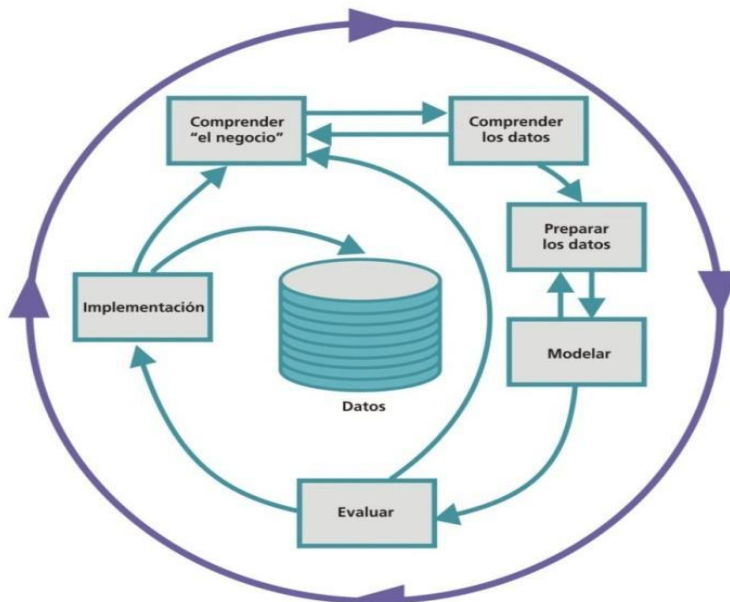


En este proyecto utilizaremos Jupyter Lab, como implementador, Python como lenguaje, y para el análisis de datos y predicción, las librerías Pandas, Numpy, Folium, Time, Matplotlib, Seaborn, Itertools, Pickle, Math, Sys, Sklearn y XgBoost.

## CRONOLOGIA DEL PROYECTO:

En primer lugar una vez elegido el tema del proyecto, planteo que es lo que realmente puedo predecir y si los datos que tengo son lo bastante buenos para conseguir el objetivo marcado del proyecto, en el tiempo establecido (2 semanas), en este caso fue la prevision de las ventas mensuales de cada uno de los artículos.

Planteando la metodología a seguir para la realización del proyecto, **el método CRISP**.



### Fase I. Business Understanding. Definición de necesidades del cliente (comprensión del negocio)

Esta fase inicial se enfoca en la comprensión de los objetivos de proyecto. Después se convierte este conocimiento de los datos en la definición de un problema de minería de datos y en un plan preliminar diseñado para alcanzar los objetivos.

Una evaluado el problema y fijado el objetivo (**prevision de ventas mensual por articulo de 0 a 20**), procedemos a fijar los objetivos en minería de datos.

```
items = pd.read_csv('items.csv')
shops = pd.read_csv('shops.csv')
cats = pd.read_csv('item_categories.csv')
train = pd.read_csv('sales_train.csv')
# set index to ID to avoid dropping it later
test = pd.read_csv('test.csv').set_index('ID')
```

items.head()

	item_name	item_id	item_category_id
0	! ВО ВЛАСТИ НАВАЖДЕНИЯ (ПЛАСТ.) D	0	40
1	!ABBY FineReader 12 Professional Edition Full...	1	76
2	***В ЛУЧАХ СЛАВЫ (UNV) D	2	40
3	***ГОЛУБАЯ ВОЛНА (Univ) D	3	40
4	***КОРОБКА (СТЕКЛО) D	4	40

cats.head()

	item_category_name	item_category_id
0	PC - Гарнитур/Наушники	0
1	Аксессуары - PS2	1
2	Аксессуары - PS3	2
3	Аксессуары - PS4	3
4	Аксессуары - PSP	4

shops.head()

	shop_name	shop_id
0	!Якутск Орджоникидзе, 56 фран	0
1	!Якутск ТЦ "Центральный" фран	1
2	Адыгея ТЦ "Мега"	2
3	Балашиха ТРК "Октябрь-Киномир"	3
4	Волжский ТЦ "Волга Молл"	4

train.head()

	date	date_block_num	shop_id	item_id	item_price	item_cnt_day
0	02.01.2013	0	59	22154	999.00	1.0
1	03.01.2013	0	25	2552	899.00	1.0
2	05.01.2013	0	25	2552	899.00	-1.0
3	06.01.2013	0	25	2554	1709.05	1.0
4	15.01.2013	0	25	2555	1099.00	1.0

Los datos estan separados en 4 dataset, uno donde se encuentran los articulos con un ID y una caegoría ID, otro con las tiendas, otro con el nombre de las categorías de cada artículo,

y el último dataset es donde aparecen las ventas de cada artículo por día desde 02.01.2013 al 30.10.2015, siendo el mes de 11.2015 , nuestro mes a predecir.

Al ser un problema de **Series Temporales**, nuestro grueso del trabajo va a ser la creación de variables que ayuden a nuestro modelo a predecir que cantidad de cada artículo se va a vender el mes de Noviembre. **Las variables son insuficientes para poder crear un modelo de precisión.**

## Fase II. Data Understanding. Estudio y comprensión de los datos

La fase de entendimiento de datos comienza con la colección de datos inicial y continúa con las actividades que permiten familiarizarse con los datos, identificar los problemas de calidad, descubrir conocimiento preliminar sobre los datos, y/o descubrir subconjuntos interesantes para formar hipótesis en cuanto a la información oculta. Minería de datos.

**El tiempo para la realización del proyecto son dos semanas**, al tratarse de tiendas situadas en Rusia, entendemos que muchos datos para crear variables van a ser muy difícil de obtener por la **opacidad de Rusia**, datos como, la tasa de desempleo de cada ciudad donde se encuentra cada una de las tiendas, o la Renta Media, tasa de nacimientos, tasa de mortandad, edad de jubilación, población escolarizada o edad media de la población. Por lo que establecemos una hoja de ruta de variables que están a nuestro alcance y establecemos su prioridad de obtención en base a los medios disponibles.

## Fase III. Data Preparation. Análisis de los datos y selección de características

La fase de preparación de datos cubre todas las actividades necesarias para construir el conjunto final de datos (los datos que se utilizarán en las herramientas de modelado) a partir de los datos en bruto iniciales. Las tareas incluyen la selección de tablas, registros y atributos, así como la transformación y la limpieza de datos para las herramientas que modelan. **EMPEZAMOS A MINAR DATOS Y CREAR VARIABLES**

**CLIMATOLOGICAS:** Todos sabemos que, el clima, más bien "el tiempo", nos afecta al humor y a los hábitos: qué comer, qué ropa ponerse, si salir o no, y por supuesto, también nuestro comportamiento de compra.

**El tiempo juega un rol importante en las decisiones de compra del consumidor**, ya sea off u online. En este sentido los ámbitos más interesados son sin duda: moda, hostelería, ocio y tiempo libre;

El clima afecta a todos indistintamente, siendo, por lo tanto, una herramienta perfecta para generar audiencias y campañas de marketing personalizadas.

El clima afecta a todos indistintamente, siendo, por lo tanto, una herramienta perfecta para generar audiencias y campañas de marketing personalizadas.

### Ingeniería de variables :

#### 4 Bloques

Climatológicas. (prioridad alta).

Demográficas. (prioridad baja).

Estadísticas (prioridad alta).

Macroeconómicas (prioridad baja).

"El nivel de prioridad no es equivalente al nivel de importancia "

¿Porqué? El 70% de la población mundial mira las predicciones del tiempo por lo menos una vez al día.

El comportamiento del consumidor puede cambiar según suba o baje la temperatura tan solo 1°C;

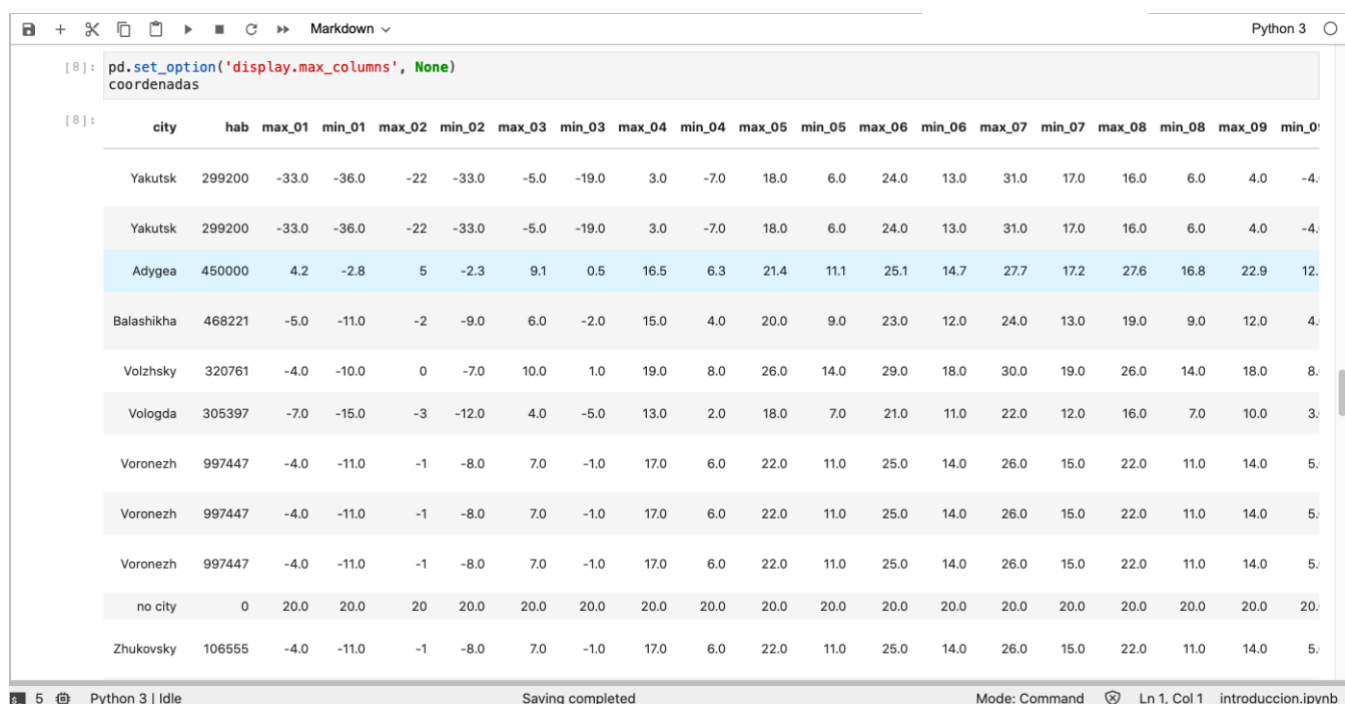
Obviamente la variación de la temperatura debe cruzarse con la geolocalización de los usuarios: una **temperatura de 20° en Asturias o en Andalucía, puede llevar a un usuario a un comportamiento de compra totalmente distinto.**

Una empresa de barbacoas llevó a cabo un experimento activando campañas solo en algunas ciudades cuando la temperatura fuera superior a 24°, y en otras cuando fuera superior a 20° (teniendo en cuenta la media estacional de cada ciudad).

Podemos, entonces, **utilizar los datos del tiempo para “acercarnos” al humor del usuario, impactándole de una manera inmediata y pertinente.**

Otro ejemplo muy claro es la estrategia de una famosa cadena de pizzas a domicilio, que vincula una mayor cantidad de campañas online cuando llueve o hace frío, y el usuario está más dispuesto en quedarse en casa.

Hay muchos estudios de neuromarketing que demuestran la relación entre el clima y la intención de compra, aun así, está claro que hay productos/servicios que están más sujetos a las influencias climáticas que otros.



The screenshot shows a Jupyter Notebook interface with a Python 3 kernel. The code cell contains the command `pd.set_option('display.max_columns', None)` followed by the variable `coordenadas`. The output is a pandas DataFrame with 20 columns representing monthly climate data (max and min values for months 01 to 10) and a 'city' column. The data is displayed for several cities, with 'Adygea' highlighted in blue.

city	hab	max_01	min_01	max_02	min_02	max_03	min_03	max_04	min_04	max_05	min_05	max_06	min_06	max_07	min_07	max_08	min_08	max_09	min_09
Yakutsk	299200	-33.0	-36.0	-22	-33.0	-5.0	-19.0	3.0	-7.0	18.0	6.0	24.0	13.0	31.0	17.0	16.0	6.0	4.0	-4.0
Yakutsk	299200	-33.0	-36.0	-22	-33.0	-5.0	-19.0	3.0	-7.0	18.0	6.0	24.0	13.0	31.0	17.0	16.0	6.0	4.0	-4.0
Adygea	450000	4.2	-2.8	5	-2.3	9.1	0.5	16.5	6.3	21.4	11.1	25.1	14.7	27.7	17.2	27.6	16.8	22.9	12.0
Balashikha	468221	-5.0	-11.0	-2	-9.0	6.0	-2.0	15.0	4.0	20.0	9.0	23.0	12.0	24.0	13.0	19.0	9.0	12.0	4.0
Volzhsy	320761	-4.0	-10.0	0	-7.0	10.0	1.0	19.0	8.0	26.0	14.0	29.0	18.0	30.0	19.0	26.0	14.0	18.0	8.0
Vologda	305397	-7.0	-15.0	-3	-12.0	4.0	-5.0	13.0	2.0	18.0	7.0	21.0	11.0	22.0	12.0	16.0	7.0	10.0	3.0
Voronezh	997447	-4.0	-11.0	-1	-8.0	7.0	-1.0	17.0	6.0	22.0	11.0	25.0	14.0	26.0	15.0	22.0	11.0	14.0	5.0
Voronezh	997447	-4.0	-11.0	-1	-8.0	7.0	-1.0	17.0	6.0	22.0	11.0	25.0	14.0	26.0	15.0	22.0	11.0	14.0	5.0
Voronezh	997447	-4.0	-11.0	-1	-8.0	7.0	-1.0	17.0	6.0	22.0	11.0	25.0	14.0	26.0	15.0	22.0	11.0	14.0	5.0
no city	0	20.0	20.0	20	20.0	20.0	20.0	20.0	20.0	20.0	20.0	20.0	20.0	20.0	20.0	20.0	20.0	20.0	20.0
Zhukovsky	106555	-4.0	-11.0	-1	-8.0	7.0	-1.0	17.0	6.0	22.0	11.0	25.0	14.0	26.0	15.0	22.0	11.0	14.0	5.0

Las variables climáticas que obtengo son: Temperatura mínima media de cada mes del año, Temperatura máxima de cada mes del año, porcentaje medio de precipitaciones por cada mes del año, y horas de luz media de cada mes del año.

**DEMOGRAFIA :** Los cambios demográficos como el envejecimiento de la población, la urbanización o la migración tienen el poder de transformar el mundo en el que vivimos a medida que se renuevan los estilos de vida de los consumidores y por consiguiente las decisiones de compra.

Las Naciones Unidas (ONU) han pronosticado que **la población mundial crecerá en un 12%** hasta llegar a los 8.500 millones de personas en 2030. Pero la gran mayoría de este crecimiento se producirá en mercados que actualmente no son grandes consumidores de vino, como India, Nigeria, Pakistán, Congo y Etiopía.

A pesar de esto, ha habido una **desaceleración en las tasas de fertilidad a nivel mundial** y, junto con el **aumento de la esperanza de vida**, la proporción de personas mayores está creciendo constantemente. Esta tendencia, aunque con diferencias en las esperanzas de vida, se proyecta en la mayoría de las regiones.

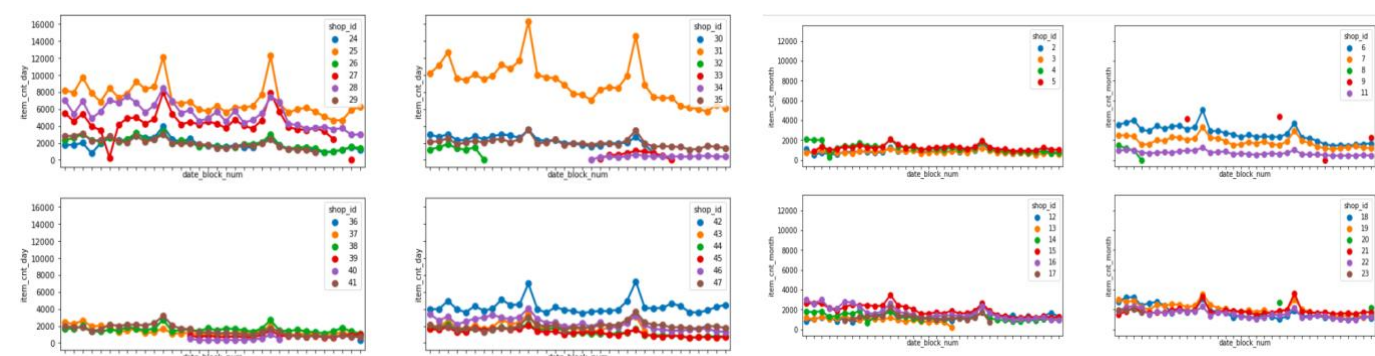


Uno de los principales motivos en la reducción de las tasas de fertilidad en todo el mundo es que la urbanización y la globalización han llevado a que las mujeres trabajen más, se casen en edades mayores y tienen menos hijos. Este cambio de modelo ha hecho que **las mujeres sean consideradas como las consumidoras más poderosas en el mundo desarrollado con más ingresos disponibles**, oportunidades para socializar y una mayor influencia en las decisiones de compra. Según datos de Naciones Unidas, en América del Norte alrededor del 47% de los hogares están encabezados por mujeres, seguido de Europa con el 37%.

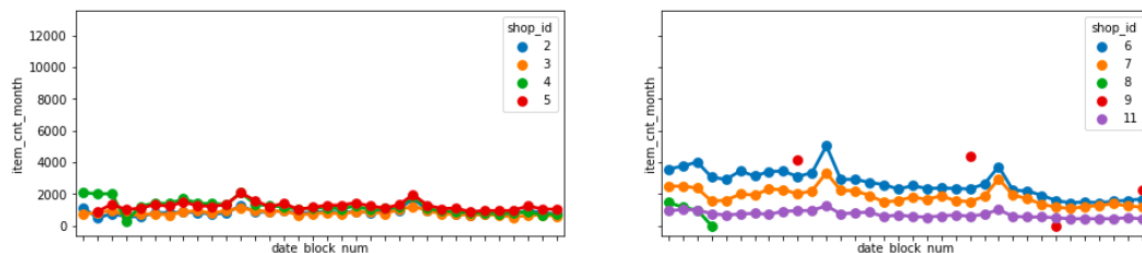
A medida que más personas en todo el mundo pasan de la vida rural a la urbana a un ritmo significativo, también se producen cambios en los estilos de vida, las actitudes hacia el matrimonio, los patrones de consumo y el tamaño de los hogares. Estos presentan nuevas oportunidades para los negocios.

Tras hacer alguna prueba con alguna de las ciudades, como Balashikha o Vologda, decido posponer la creación de variables demográficas, por la opacidad de los datos en Rusia, el tiempo límite para la entrega del proyecto no me permite minar los datos de manera mas profunda. **Trabajo pendiente para una próxima evolucion.**

**ESTADÍSTICAS:** Utilizando el método CRISP, **volvemos a fase II** para realizar algunas gráficas y entender mejor los datos y así saber que variables son necesarias crear.



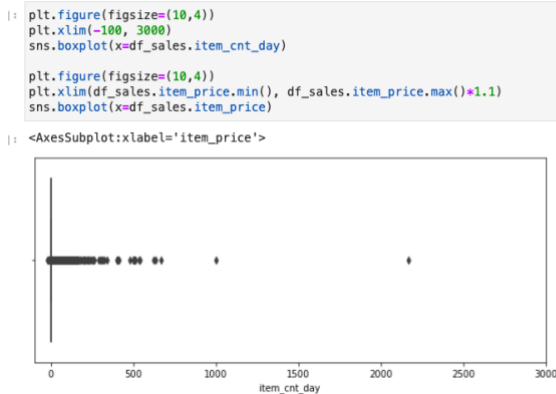
```
grouped = pd.DataFrame(matrix.groupby(['shop_id', 'date_block_num'])['item_cnt_month'].sum().reset_index())
fig, axes = plt.subplots(nrows=5, ncols=2, sharex=True, sharey=True, figsize=(16,20))
num_graph = 10
id_per_graph = math.ceil(grouped.shop_id.max() / num_graph)
count = 0
for i in range(5):
    for j in range(2):
        sns.pointplot(x='date_block_num', y='item_cnt_month', hue='shop_id', data=grouped[np.logical_and(count*id_per_graph <= grouped['shop_id'], count <= 1)
```



Utilizando la tecnica de groupby y librería seaborn al mismo tiempo grafico el comportameinto de cada una de las tiendas en cada mes.

## Comprobación de Outliers.

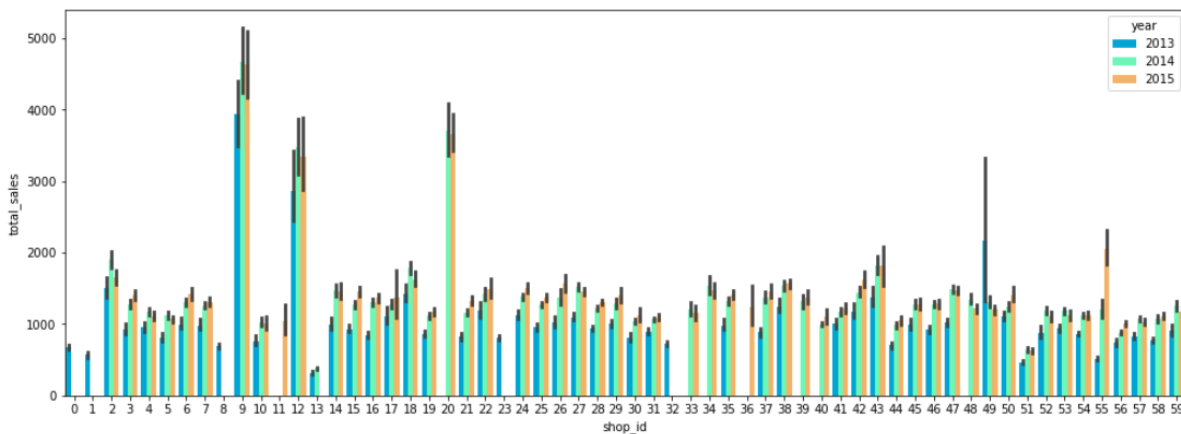
### Comprobación de Outliers



Un outlier es una observación anormal y extrema en una muestra estadística o serie temporal de datos que puede afectar potencialmente a la estimación de los parámetros del mismo.

En palabras más sencillas un outlier sería una observación dentro de una muestra o una serie temporal de datos que no es consistente con el resto. Imaginemos por ejemplo que estamos midiendo la altura de los alumnos de una clase.

```
In: fig, ax = plt.subplots(figsize=(17,6))
sns.barplot(x="shop_id", y="total_sales", hue="year", data=df_sales, palette="rainbow")
plt.show()
```

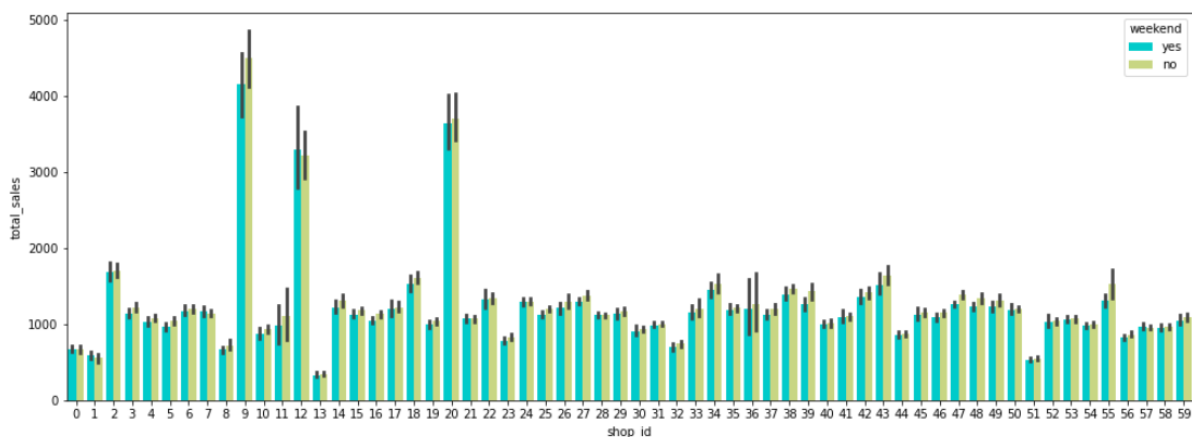


Con esta grafica ya podemos comprobar que hay tiendas que han cerrado o han estado inactivas durante el 2015

Ventas por año de cada tienda, empezamos a entender que algunas tiendas han cerrado en 2015.

Creamos una variable de fin de semana, para poder estudiar el comportamiento del consumidor, durante los días de la semana y en que momento se realizan mas ventas. Un aspecto importante a tener en cuenta es no confundir, ventas con afluencia de público, puede ocurrir que se realicen mas ventas en días laborales pero haya más público en fin de semana

```
In: fig, ax = plt.subplots(figsize=(17,6))
sns.barplot(x="shop_id", y="total_sales", hue="weekend", data=df_sales, palette="rainbow")
plt.show()
```

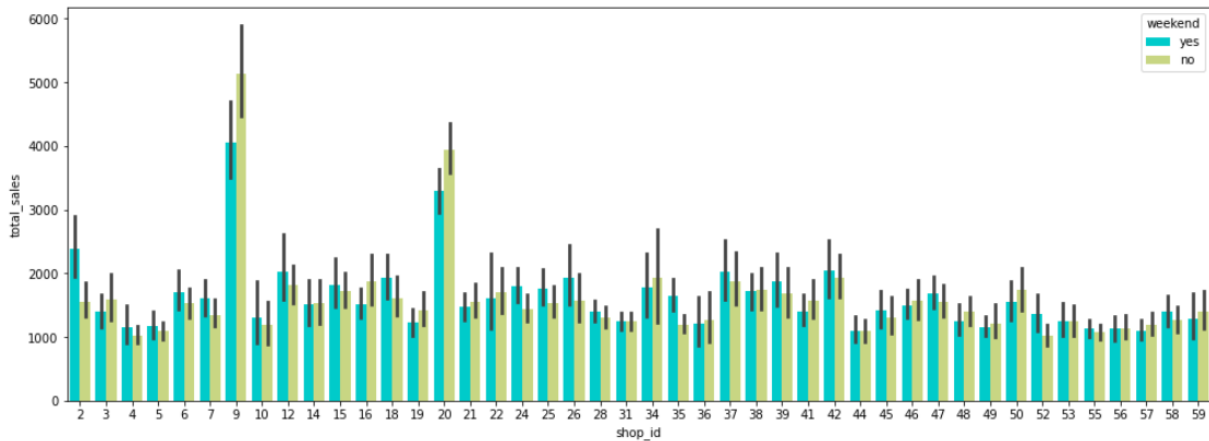


Podemos comprobar que en general en todas las tiendas hay mas ventas en diario que en fin de semana, por lo que ya nos da algunas pistas del comportamiento del consumidor en Rusia

Pasamos a fase III y creamos algunas variables estadísticas : Variable de consumo en fin de semana.

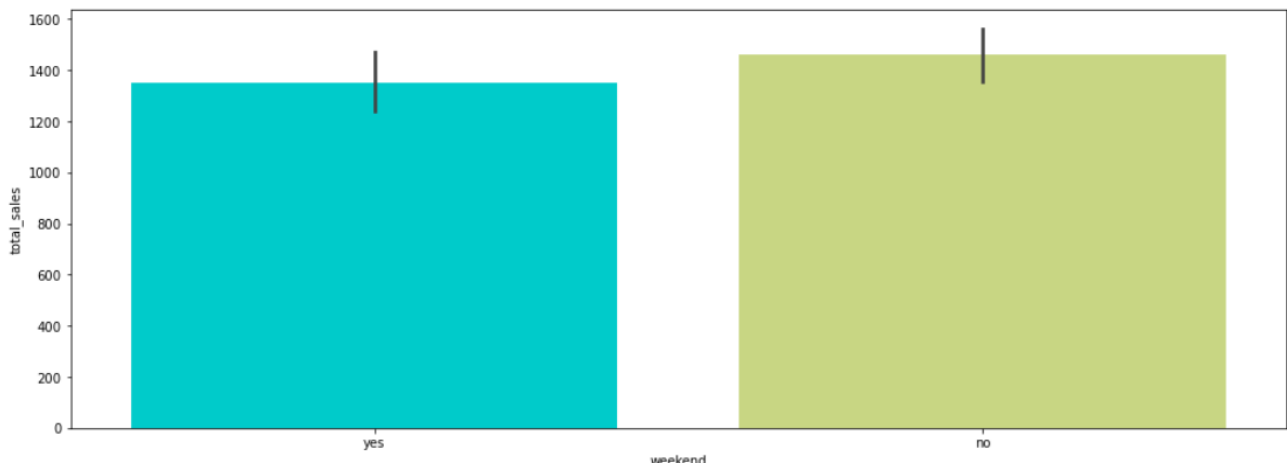
```
X = df_sales.loc[df_sales['date_block_num']==33]
```

```
fig, ax = plt.subplots(figsize=(17,6))
sns.barplot(x="shop_id", y="total_sales", hue="weekend", data=X, palette="rainbow")
plt.show()
```

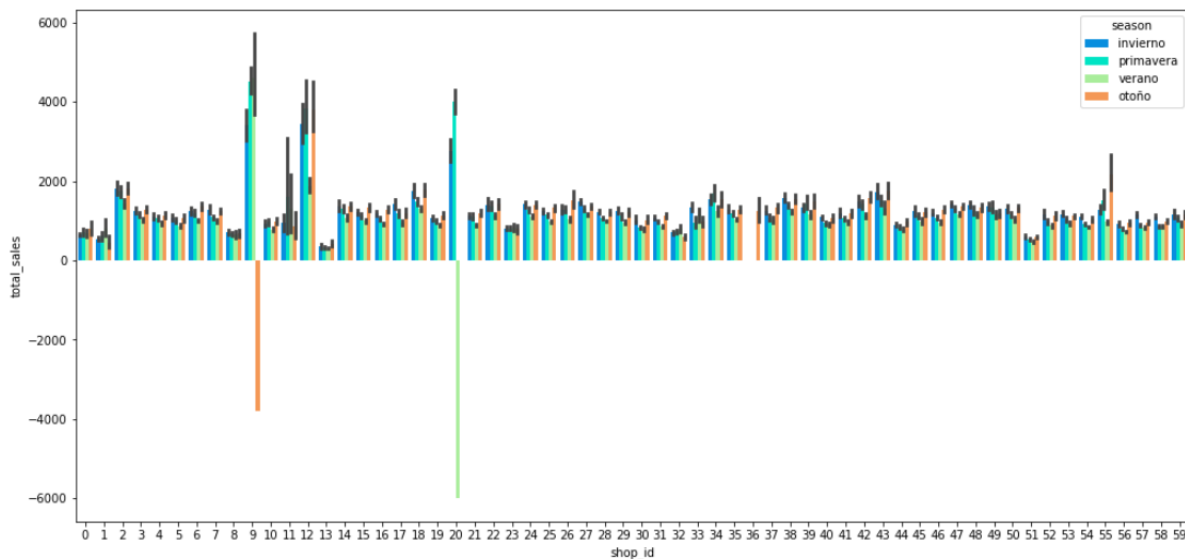


Confirmamos nuestra sospecha de que en Rusia, se realizan mas compras durante los días de diario que en fin de semana. Vamos a terminar por comprobarlo graficando solo la ultima semana

```
fig, ax = plt.subplots(figsize=(17,6))
sns.barplot(x="weekend", y="total_sales", data=X, palette="rainbow")
plt.show()
```



```
fig, ax = plt.subplots(figsize=(17,8))
sns.barplot(x="shop_id", y="total_sales", hue="season", data=df_sales, palette="rainbow")
plt.show()
```



Vemos que hay valores negativos por lo que entendemos que son devoluciones de productos o correcciones de ventas incorrectas

Variables estacional.

```
[21]: df_sales.loc[df_sales['date'].between('2013-01-01','2013-01-08'), 'holyday'] =1# navidad
df_sales.loc[df_sales['date']=='2013-02-23', 'holyday'] =1#fiesta nacional
df_sales.loc[df_sales['date']=='2013-03-08', 'holyday'] =1#fiesta nacional
df_sales.loc[df_sales['date'].between('2013-05-01','2013-05-03'), 'holyday'] =1#fiesta nacional
df_sales.loc[df_sales['date'].between('2013-05-09','2013-05-10'), 'holyday'] =1#fiesta nacional
df_sales.loc[df_sales['date']=='2013-06-12', 'holyday'] =1#fiesta nacional
df_sales.loc[df_sales['date']=='2013-11-04', 'holyday'] =1#fiesta naciona

df_sales.loc[df_sales['date'].between('2014-01-01','2014-01-08'), 'holyday'] =1
df_sales.loc[df_sales['date'].between('2014-02-22','2014-02-23'), 'holyday'] =1
df_sales.loc[df_sales['date'].between('2014-03-08','2014-03-10'), 'holyday'] =1
df_sales.loc[df_sales['date'].between('2014-05-01','2014-05-04'), 'holyday'] =1
df_sales.loc[df_sales['date'].between('2014-05-09','2014-05-11'), 'holyday'] =1
df_sales.loc[df_sales['date']=='2014-06-07', 'holyday'] =1
df_sales.loc[df_sales['date'].between('2014-06-12','2014-06-15'), 'holyday'] =1
df_sales.loc[df_sales['date'].between('2014-11-01','2014-11-02'), 'holyday'] =1

df_sales.loc[df_sales['date'].between('2015-01-01','2015-01-09'), 'holyday'] =1# navidad
df_sales.loc[df_sales['date']=='2015-02-23', 'holyday'] =1
df_sales.loc[df_sales['date'].between('2015-03-08','2015-03-09'), 'holyday'] =1#
df_sales.loc[df_sales['date']=='2015-05-01', 'holyday'] =1
df_sales.loc[df_sales['date']=='2015-05-04', 'holyday'] =1
df_sales.loc[df_sales['date']=='2015-05-09', 'holyday'] =1
df_sales.loc[df_sales['date']=='2015-05-11', 'holyday'] =1
df_sales.loc[df_sales['date']=='2015-06-12', 'holyday'] =1
df_sales.loc[df_sales['date']=='2015-11-04', 'holyday'] =1

df_sales['holyday'] = df_sales['holyday'].replace(np.nan, 0)
# https://www.timeanddate.com/holidays/russia
```

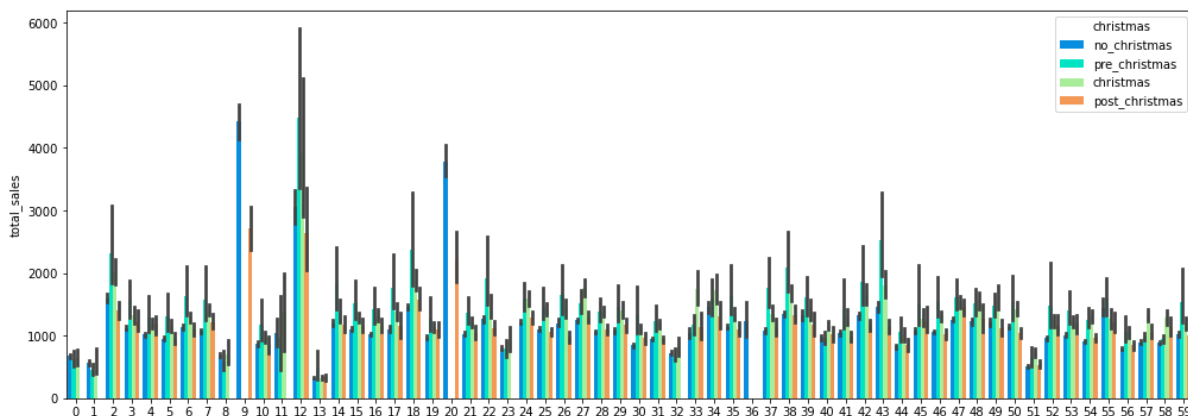
Variable, Fiestas Nacionales de Rusia. Usando la librería pandas utilizamos el metodo loc y between.

## Creo una variable especifica para la navidad

```
]#Variable especifica de Navidad
df_sales.loc[df_sales['date'].between('2013-01-01','2013-01-08'), 'christmas'] = 'christmas'
df_sales.loc[df_sales['month']==11, 'christmas'] = 'pre_christmas'
df_sales.loc[df_sales['date'].between('2013-12-01','2014-01-08'), 'christmas'] = 'christmas'
df_sales.loc[df_sales['date'].between('2014-01-09','2014-01-31'), 'christmas'] = 'post_christmas'
df_sales.loc[df_sales['date'].between('2014-12-01','2015-01-09'), 'christmas'] = 'christmas'
df_sales.loc[df_sales['date'].between('2015-01-09','2015-01-31'), 'christmas'] = 'post_christmas'

df_sales['christmas'] = df_sales['christmas'].replace(np.nan, 'no_christmas')

fig, ax = plt.subplots(figsize=(17,6))
sns.barplot(x="shop_id", y="total_sales", hue="christmas", data=df_sales, palette="rainbow")
plt.show()
```



“un modelo es lo que come”, y por muy buenos y sofisticados que sean los algoritmos que usamos, si los alimentamos con datos de mala calidad (con errores, datos nulos, incompletos, etc.) no podrán aprender a capturar correctamente las relaciones entre las variables y por lo tanto harán malas predicciones, provocando que tomemos malas decisiones.



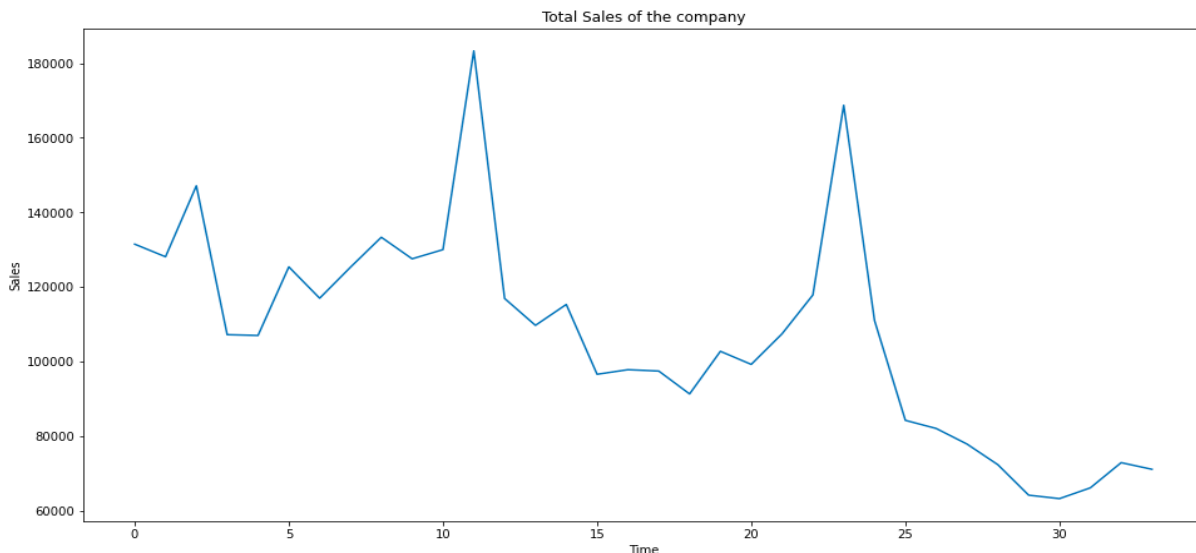
Por las características climatológicas y culturales de Rusia, creo otra variable, en la que incluyo fiestas musulmanas, ortodoxas, y otros eventos c

```
[25]: df_sales.loc[df_sales['date']=='2013-01-14', 'other_events'] =1#año nuevo ortodoxo
df_sales.loc[df_sales['date']=='2013-01-21', 'other_events'] =1#fiesta musulman
df_sales.loc[df_sales['date']=='2013-01-21', 'other_events'] =1#fiesta musulman
df_sales.loc[df_sales['date']=='2013-02-14', 'other_events'] =1#san valentin
df_sales.loc[df_sales['date']=='2013-03-20', 'other_events'] =1
df_sales.loc[df_sales['date']=='2013-05-05', 'other_events'] =1#fiesta ortodoxa
df_sales.loc[df_sales['date']=='2013-06-06', 'other_events'] =1#fiesta musulman
df_sales.loc[df_sales['date']=='2013-06-21', 'other_events'] =1#verano
df_sales.loc[df_sales['date']=='2013-01-21', 'other_events'] =1#fiesta musulman
df_sales.loc[df_sales['date']=='2013-07-09', 'other_events'] =1#fiesta musulman
df_sales.loc[df_sales['date']=='2013-08-03', 'other_events'] =1#fiesta musulman
df_sales.loc[df_sales['date']=='2013-08-08', 'other_events'] =1#fiesta musulman
df_sales.loc[df_sales['date']=='2013-09-01', 'other_events'] =1#observance
df_sales.loc[df_sales['date']=='2013-09-23', 'other_events'] =1#equinocio
df_sales.loc[df_sales['date']=='2013-11-05', 'other_events'] =1#fiesta musulman
df_sales.loc[df_sales['date']=='2013-12-21', 'other_events'] =1#invierno
df_sales.loc[df_sales['date']=='2013-12-30', 'other_events'] =1# vispera año nuevo
df_sales.loc[df_sales['date']=='2013-12-31', 'other_events'] =1#vispera año nuevo

df_sales.loc[df_sales['date']=='2014-01-14', 'other_events'] =1#año nuevo ortodoxo
df_sales.loc[df_sales['date']=='2014-02-14', 'other_events'] =1#san valentin
df_sales.loc[df_sales['date']=='2014-03-20', 'other_events'] =1#observance
df_sales.loc[df_sales['date']=='2014-04-20', 'other_events'] =1#observance
df_sales.loc[df_sales['date']=='2014-05-27', 'other_events'] =1#fiesta musulman
df_sales.loc[df_sales['date']=='2014-06-21', 'other_events'] =1#verano
df_sales.loc[df_sales['date']=='2014-06-29', 'other_events'] =1#fiesta musulman
df_sales.loc[df_sales['date']=='2014-09-01', 'other_events'] =1#observance
df_sales.loc[df_sales['date']=='2014-09-23', 'other_events'] =1#equinocio
df_sales.loc[df_sales['date']=='2014-10-04', 'other_events'] =1#fiesta musulman
df_sales.loc[df_sales['date']=='2014-10-25', 'other_events'] =1#fiesta musulman
df_sales.loc[df_sales['date']=='2014-12-22', 'other_events'] =1#invierno
df_sales.loc[df_sales['date']=='2014-12-30', 'other_events'] =1# vispera año nuevo
df_sales.loc[df_sales['date']=='2014-12-31', 'other_events'] =1#vispera año nuevo

df_sales.loc[df_sales['date']=='2015-01-03', 'other_events'] =1# fiesta musulman
df_sales.loc[df_sales['date']=='2015-01-14', 'other_events'] =1#año nuevo ortodoxo
df_sales.loc[df_sales['date']=='2015-02-14', 'other_events'] =1#san valentin
df_sales.loc[df_sales['date']=='2015-02-27', 'other_events'] =1#observance
df_sales.loc[df_sales['date']=='2015-03-21', 'other_events'] =1#observance
df_sales.loc[df_sales['date']=='2015-04-12', 'other_events'] =1#fiesta ortodoxa
df_sales.loc[df_sales['date']=='2015-05-16', 'other_events'] =1#fiesta musulman
df_sales.loc[df_sales['date']=='2015-06-18', 'other_events'] =1#fiesta musulman
df_sales.loc[df_sales['date']=='2015-06-21', 'other_events'] =1#verano
df_sales.loc[df_sales['date']=='2015-07-13', 'other_events'] =1#fiesta musulman
df_sales.loc[df_sales['date']=='2015-07-18', 'other_events'] =1#fiesta musulman
df_sales.loc[df_sales['date']=='2015-09-01', 'other_events'] =1#observance
df_sales.loc[df_sales['date']=='2015-09-23', 'other_events'] =1#equinocio
df_sales.loc[df_sales['date']=='2015-09-24', 'other_events'] =1#fiesta musulman
df_sales.loc[df_sales['date']=='2015-10-15', 'other_events'] =1#fiesta musulman
df_sales.loc[df_sales['date']=='2015-12-22', 'other_events'] =1#invierno
df_sales.loc[df_sales['date']=='2015-12-24', 'other_events'] =1#fiesta musulman
df_sales.loc[df_sales['date']=='2015-12-30', 'other_events'] =1# vispera año nuevo
df_sales.loc[df_sales['date']=='2015-12-31', 'other_events'] =1#vispera año nuevo
```

```
j: ts=df_sales.groupby(["date_block_num"])["item_cnt_day"].sum()
ts.astype('float')
plt.figure(figsize=(16,8))
plt.title('Total Sales of the company')
plt.xlabel('Time')
plt.ylabel('Sales')
plt.plot(ts);
```



Creación de variables estadísticas como la media de ventas por mes, media de cada artículo, media de venta por tienda, media de venta por categoría de artículo, ciudad, media de venta por tipo de tienda y media por tipo de ciudad.

```
[25]: 16.096070289611816
```

```
[26]: group = matrix.groupby(['date_block_num', 'shop_id', 'item_category_id']).agg({'item_cnt_month': ['mean']})
group.columns = ['date_shop_cat_avg_item_cnt']
group.reset_index(inplace=True)

matrix = pd.merge(matrix, group, on=['date_block_num', 'shop_id', 'item_category_id'], how='left')
matrix['date_shop_cat_avg_item_cnt'] = matrix['date_shop_cat_avg_item_cnt'].astype(np.float16)
matrix = lag_feature(matrix, [1], 'date_shop_cat_avg_item_cnt')
matrix.drop(['date_shop_cat_avg_item_cnt'], axis=1, inplace=True)
time.time() - ts
```

```
[26]: 36.06906509399414
```

```
[27]: ts = time.time()
group = matrix.groupby(['date_block_num', 'shop_id', 'type_code']).agg({'item_cnt_month': ['mean']})
group.columns = ['date_shop_type_avg_item_cnt']
group.reset_index(inplace=True)

matrix = pd.merge(matrix, group, on=['date_block_num', 'shop_id', 'type_code'], how='left')
matrix['date_shop_type_avg_item_cnt'] = matrix['date_shop_type_avg_item_cnt'].astype(np.float16)
matrix = lag_feature(matrix, [1], 'date_shop_type_avg_item_cnt')
matrix.drop(['date_shop_type_avg_item_cnt'], axis=1, inplace=True)
time.time() - ts
```

```
[27]: 18.253817081451416
```

```
[28]: ts = time.time()
group = matrix.groupby(['date_block_num', 'shop_id', 'subtype_code']).agg({'item_cnt_month': ['mean']})
group.columns = ['date_shop_subtype_avg_item_cnt']
group.reset_index(inplace=True)

matrix = pd.merge(matrix, group, on=['date_block_num', 'shop_id', 'subtype_code'], how='left')
matrix['date_shop_subtype_avg_item_cnt'] = matrix['date_shop_subtype_avg_item_cnt'].astype(np.float16)
matrix = lag_feature(matrix, [1], 'date_shop_subtype_avg_item_cnt')
matrix.drop(['date_shop_subtype_avg_item_cnt'], axis=1, inplace=True)
time.time() - ts
```

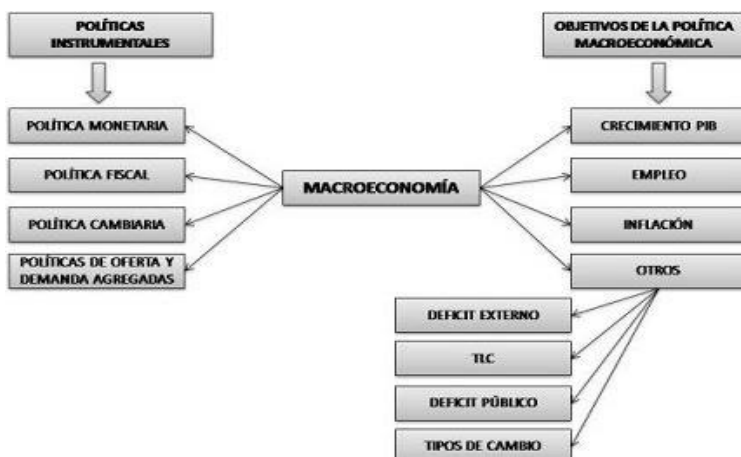
```
[28]: 18.133793830871582
```

```
[29]: ts = time.time()
group = matrix.groupby(['date_block_num', 'city_code']).agg({'item_cnt_month': ['mean']})
group.columns = ['date_city_avg_item_cnt']
group.reset_index(inplace=True)

matrix = pd.merge(matrix, group, on=['date_block_num', 'city_code'], how='left')
matrix['date_city_avg_item_cnt'] = matrix['date_city_avg_item_cnt'].astype(np.float16)
matrix = lag_feature(matrix, [1], 'date_city_avg_item_cnt')
matrix.drop(['date_city_avg_item_cnt'], axis=1, inplace=True)
time.time() - ts
```

```
[29]: 18.28374981880188
```

## MACROECONOMICAS :

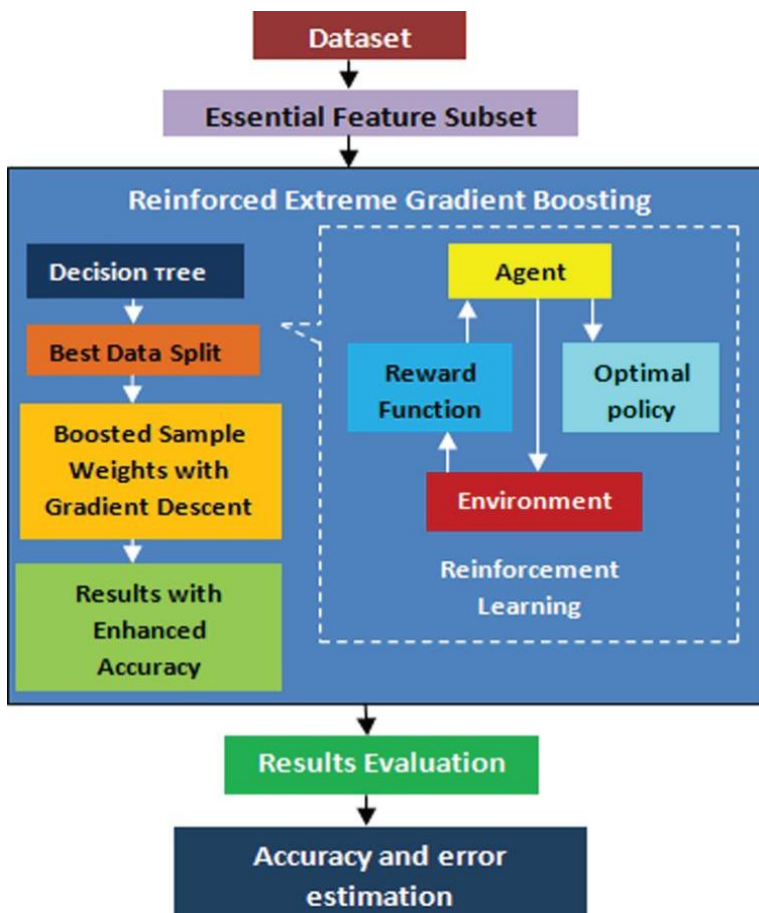


A pesar de la importancia e influencia de las variables, macroeconomicas en el consumo, no dispongo de los medios ni del tiempo necesario para poder obtener unos datos adecuados y reales de Rusia, para introducirlos en el modelo.

#### Fase IV. Modeling. Modelado

En esta fase, se seleccionan y aplican las técnicas de modelado que sean pertinentes al problema, y se calibran sus parámetros a valores óptimos. Típicamente hay varias técnicas para el mismo tipo de problema de minería de datos. Algunas técnicas tienen requerimientos específicos sobre la forma de los datos. Por lo tanto, casi siempre en cualquier proyecto se acaba volviendo a la fase de preparación de datos.

Al tratarse de un problema de regresión me decanto por boosting, ya que este algoritmo arroja muy buenos resultados.



El boosting consiste en combinar los resultados de varios clasificadores débiles para obtener un clasificador robusto. Cuando se añaden estos clasificadores débiles, se lo hace de modo que estos tengan diferente peso en función de la exactitud de sus predicciones. Luego de que se añade un clasificador débil, los datos cambian su estructura de pesos: los casos que son mal clasificados ganan peso y los que son clasificados correctamente pierden peso. Así, los clasificadores débiles se centran de mayor manera en los casos que fueron mal clasificados por los clasificadores débiles.

Hay muchos algoritmos de boosting. Los algoritmos originales, propuestos por Robert Schapire y Yoav Freund, no fue adaptativa y no podría tomar ventaja llena de los clasificadores débiles.<sup>7</sup> Sin embargo, Schapire y Freund luego desarrollaron AdaBoost, un algoritmo adaptativo que ganó el prestigioso Premio Gödel.

Wikipedia.

#### Fase V. Evaluation. Evaluación (obtención de resultados)

En esta etapa en el proyecto, se han construido uno o varios modelos que parecen alcanzar calidad suficiente desde la una perspectiva de análisis de datos.

Antes de proceder al despliegue final del modelo, es importante evaluarlo a fondo y revisar los pasos ejecutados para crearlo, comparar el modelo obtenido con los objetivos de negocio. Un objetivo clave es determinar si hay alguna cuestión importante de negocio que no haya sido considerada suficientemente. Al final de esta fase, se debería obtener una decisión sobre la aplicación de los resultados del proceso de análisis de datos.

```

[8]: ts = time.time()

model = XGBRegressor(
    max_depth=8,
    n_estimators=5000,
    min_child_weight=500,
    colsample_bytree=0.8,
    subsample=0.8,
    eta=0.3,
    seed=42)

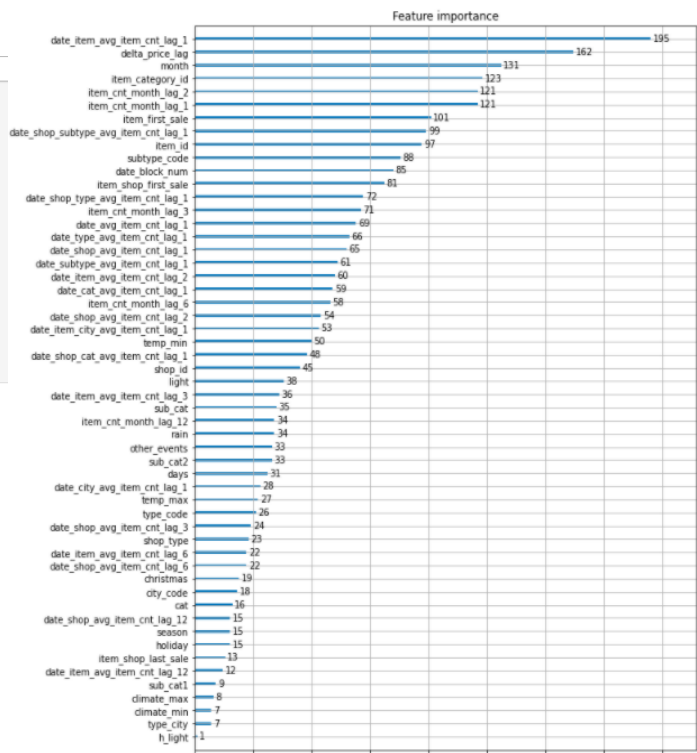
model.fit(
    X_train,
    Y_train,
    eval_metric='rmse',
    eval_set=[(X_train, Y_train), (X_valid, Y_valid)],
    verbose=True,
    early_stopping_rounds = 10)

time.time() - ts

[0] validation_0-rmse:1.05604 validation_1-rmse:1.02979
[1] validation_0-rmse:0.97384 validation_1-rmse:0.96818
[2] validation_0-rmse:0.92485 validation_1-rmse:0.93692
[3] validation_0-rmse:0.89223 validation_1-rmse:0.92172
[4] validation_0-rmse:0.87235 validation_1-rmse:0.91239
[5] validation_0-rmse:0.86069 validation_1-rmse:0.90692
[6] validation_0-rmse:0.85278 validation_1-rmse:0.90465
[7] validation_0-rmse:0.84742 validation_1-rmse:0.90163
[8] validation_0-rmse:0.84283 validation_1-rmse:0.90042
[9] validation_0-rmse:0.83953 validation_1-rmse:0.89939
[10] validation_0-rmse:0.83786 validation_1-rmse:0.89926
[11] validation_0-rmse:0.83613 validation_1-rmse:0.89858
[12] validation_0-rmse:0.83584 validation_1-rmse:0.89760
[13] validation_0-rmse:0.83252 validation_1-rmse:0.89673
[14] validation_0-rmse:0.83059 validation_1-rmse:0.89600
[15] validation_0-rmse:0.82917 validation_1-rmse:0.89929
[16] validation_0-rmse:0.82603 validation_1-rmse:0.89605
[17] validation_0-rmse:0.82517 validation_1-rmse:0.89546
[18] validation_0-rmse:0.82483 validation_1-rmse:0.89596
[19] validation_0-rmse:0.82312 validation_1-rmse:0.89517
[20] validation_0-rmse:0.82242 validation_1-rmse:0.89575
[21] validation_0-rmse:0.82163 validation_1-rmse:0.89591
[22] validation_0-rmse:0.82076 validation_1-rmse:0.89681
[23] validation_0-rmse:0.82027 validation_1-rmse:0.89668
[24] validation_0-rmse:0.81994 validation_1-rmse:0.89705
[25] validation_0-rmse:0.81906 validation_1-rmse:0.89692
[26] validation_0-rmse:0.81782 validation_1-rmse:0.89654
[27] validation_0-rmse:0.81716 validation_1-rmse:0.89595

[8]: 538.7976713180542

```



El modelo, arroja resultados prometedores, es un modelo robusto.

### 3. PROBLEMAS Y SOLUCIONES.

Este ha sido un proyecto difícil, como hemos dicho, a lo largo de esta memoria un modelo no va a ser bueno si los datos no son reales o no son de calidad. “Un modelo es lo que come”.

En nuestro caso los datos que nos proporcionaba la empresa eran insuficientes para la creación de un modelo robusto, el trabajo de ingeniería ha sido descomunal, entender y poder segmentar al consumidor de Rusia, se ha llevado el 80% del tiempo del proyecto, establecer pautas y modas de compras en base a las distintas variables ha requerido horas y horas de lectura de la cultura Rusa.

Se han quedado en el tintero, muchas variables que hubieran hecho mejorar el modelo en gran medida, variables macroeconómicas. Aún así gracias al trabajo se han conseguido buenos resultados.

He tenido que volver atrás muchas veces, pasaba de fase y me daba cuenta de otra incógnita que me requería volver a analizar el problema.

En cuanto a la programación del código, ha sido, por así decirlo lo menos tedioso, pues con la ayuda de mis profesores, compañeros y foros de internet, puedes hacer cualquier cosa que te propongas.

Otra cosa diferente es cuando tienes que probar el modelo, la inmensa cantidad de datos de la que disponía no era compatible con los recursos disponibles, más de 11 mill de filas y 50 columnas. Servicios como Colab, IBM Watson o Kaggle, no eran capaces de correr el modelo, al final no tuve más remedio que runearlo en local, con la pérdida de tiempo que esto llevaba.

Resumiendo el mayor problema ha sido el poco tiempo para la obtención de unos datos adecuados, dos semanas trabajando a contra reloj, que no las cambio por nada, por lo mucho que he aprendido.



## 4. CONCLUSION.

**Comprender la importancia de la prevision de ventas** permite tomar medidas preventivas para evitar perdidas en el trafico o las ventas, al tiempo que permite que sus negocios se vuelvan mas maleables al adaptarse a cambios nuevos y repentinos.

Ademas, es posible detectar nuevas oportunidades para aumentar los ingresos utilizando metodos de prevision. A partir de ahi, las empresas pueden formular estrategias para optimizar sus operaciones y habitos de gasto.

Ya sea que necesiten asignar mas fondos a determinados productos del inventario, aumentar los esfuerzos de comercializacion y promoción o detener el proceso de contratacion de nuevos empleados, las ventas y la demanda es necesario para predecir futuras situaciones financieras y tomar decisiones mas informadas.

### ALGUNAS DE LAS VENTAJAS DE LAS PREVISIONES DE VENTAS SON:

**1. Planificacion financiera mas inteligente**- la informacion de pronosticos precisos permite a las empresas planificar mejor financieramente sus futuros. Las proyecciones pueden delinear claramente las temporadas punta, dividiendo esto aun mas a dias y horas pares. Al planificar en consecuencia, las empresas pueden asegurarse de que tienen suficiente flujo de caja a mano para prepararse para gastos imprevistos.

**2. Prevencion de problemas de personal** - Reconocer la importancia de las previsiones puede ayudar a corregir cualquier problema de personal que las empresas puedan estar experimentando. Al evaluar el aumento y la caida de la demanda cada mes, día u hora, las empresas ya no tienen que depender de conjeturas ingenuas para planificar los horarios de los empleados. En cambio, los gerentes tendran una mejor idea de que turnos requeriran la mayor cantidad de mano de obra disponible y seran capaces de planificar con anticipacion.

**3. Desarrollar planes de marketing** - La prevision tambien es beneficiosa para desarrollar planes de marketing eficaces. Cuando las empresas pueden prever una caida en las ventas, es especialmente importante intensificar los esfuerzos de marketing dirigidos. La prevision empresarial les permite ver donde faltan ventas y, a continuacion, adaptar las promociones para llenar ese vacio.

**4. Gestion de inventario mejorada**- al invertir en previsiones, las empresas obtienen la capacidad de mejorar la gestion de la produccion. Lo ultimo que los propietarios querrian es un desbordamiento de inventario cuando las ventas se estan sumergiendo. Predecir con precision la demanda les da la prevision de adaptar la cadena de suministro para satisfacer las demandas de productos y servicios especificos, al tiempo que garantiza que el negocio tenga los medios de transporte, materiales y mano de obra disponibles.

**A modo personal**, este proyecto me ha dado la experiencia y la confianza para poder enfrentarme a cualquier proyecto que me depare el futuro. He aprendido muchísimo, he utilizado todo lo que mis maravillosos profesores me han enseñado.

Este proyecto ha sido el broche de oro a la finalización del bootcamp, donde me llevo los conocimientos necesarios para ejercer como científico de datos.

**Doy las gracias a todo el Staff de Neoland**, por haberme enseñado tanto, a **Adrián y Eduardo**, especialmente por sus conocimientos y manera de implantarlos, haciendo fácil lo difícil, y como no a **mis compañeros (amigos)**, sin los cuales y sin su ayuda, con sus tirones, y ánimos, a veces no habría podido avanzar tan rápido como lo he hecho.