# Demand Forecasting

En términos simples, Demand Forecasting, o predicción de demanda, se refiere a poder estimar en base a datos pasados cuánta será la demanda de un bien a futuro. Esto es comúnmente utilizado por fábricas y empresas que deben realizar seguimiento y control de stock para, por ejemplo, no quedarse sin materiales para la producción de un producto.

## En QuéCompraste?

En nuestro sistema de seguimiento y análisis de compras, este mismo concepto se aplicaría de una forma similar, solo que orientado al individuo o a un grupo familiar de un mismo hogar. Basándonos en la información provista de los recibos, podemos seleccionar productos que se compren de manera regular como reponibles, y procesando sus datos históricos estimar cuánta cantidad se volverá a comprar de estos y alrededor de qué fecha.

Hay tres espectros que considerar; frecuencia de las compras, cantidad de datos históricos disponibles, y distancia a futuro de la predicción. Con esto en mente, vamos a dividir el problema realizando una selección por producto y por usuario, por lo que habrá que discriminar entre productos reponibles que se compran con frecuencia de los que se compran ocasionalmente, separar los conjuntos de datos abundantes de los escasos, y ver a qué fechas a futuro estamos ofreciendo una predicción. Así podremos delegar la predicción a estrategias que mejor se adapten al caso a procesar.

### Predicciones a corto plazo:

* Datos abundantes de compra frecuente: SARIMA o Prophet
* Datos abundantes de compra infrecuente: XGBoost
* Datos escasos de compra frecuente: Media Móvil Exponencial
* Datos escasos de compra infrecuente:
  + Croston o Syntetos-Boylan Approximation
  + Modelos basados en Poisson
  + Modelos Hurdle o Zero-Inflated

### Predicciones a largo plazo:

* Datos abundantes de compra frecuente: Prophet
* Datos abundantes de compra infrecuente: XGBoost
* Datos escasos de compra frecuente: Ninguno, baja certeza
* Datos escasos de compra infrecuente: Ninguno, baja certeza

# Estrategias y modelos

## Simple Moving Average (SMA)

Es un cálculo de promediado de los datos tomando una cantidad de periodos pasados, tomando el valor promedio con ponderación por igual, y utilizando este resultante como posible próximo valor. Ideal para datos con tendencias suaves o estables, sin cambios bruscos y con bajo nivel de ruido.

## Exponential Moving Average (EMA)

Al igual que SMA, toma un conjunto de periodos pasados y realiza un cálculo para obtener el promedio de estos datos, sin embargo pondera con mayor prioridad datos recientes, volviéndose más sensible a cambios de tendencia. Cuanto más grande sea el conjunto de datos pasados, más influyen los datos pasados sobre el resultado. Esto lo hace mejor para datos que tienen cambios de tendencia o fluctuaciones más marcadas.

## Linear Regression

Este es un modelo estadístico que busca una relación lineal entre una variable dependiente y una o más variables independientes. Utilizando un conjunto completo de datos históricos, la regresión lineal encuentra la línea que mejor ajusta los datos minimizando la suma de los errores al cuadrado entre los puntos de datos y la línea de ajuste. Esta fórmula lineal se puede emplear para hacer predicciones de valores futuros al proyectar la tendencia observada en los datos históricos. Apropiado cuando existe una relación constante y lineal en el tiempo, o cuando se dispone de un historial de datos suficientemente amplio para definir una tendencia lineal clara.

## Prophet

Este modelo, desarrollado por Meta, es especialmente útil para datos que representan patrones estacionales fuertes en tendencias históricas. Funciona bien con datos diarios, semanales y mensuales que exhiben patrones estacionales no lineales; estas tendencias se tratan de manera aditiva y se ajustan con componentes como día, estación, mes, etc. Como ventajas adicionales, es fácil de usar al no requerir gran configuración de parámetros y suele ser más rápido y eficiente en comparación con otros métodos de predicción de series temporales. Se utiliza, por ejemplo, en predicción de demanda, previsión de tráfico web, análisis financiero, gestión de energía, y demás.

### NeuralProphet

NeuralProphet es una implementación de Prophet en PyTorch y una extensión del modelo original que incorpora técnicas de aprendizaje profundo para mejorar la procesión de las predicciones de series temporales. Tiene una mayor flexibilidad y capacidad de capturar patrones complejos en los datos, siendo particularmente útil para series temporales que muestran una fuerte dependencia temporal.

## XGBoost

Este es un algoritmo de aprendizaje automático basado en árboles de decisión optimizados mediante boosting. Está diseñado para ser eficiente y escalable. Es un ensamblaje de árboles de decisión, donde nuevos árboles ajustan y corrigen a los anteriores, y se continúan agregando hasta que ya no hay mejoras en la certeza de la predicción. Es un modelo altamente configurable y admite regularización. Se usa para resolver problemas con datos estructurados, caracterizados por su tamaño grande, múltiples variables, relaciones complejas, y donde la precisión es clave para la toma de decisiones.