# PREDICCIONES SERIES TEMPORALES DSMARKET

Contenido

[0. Planteamiento del problema. 2](#_Toc179721646)

[1. Acciones Prospectivas 3](#_Toc179721647)

[2. Pruebas y Elección de Modelos 4](#_Toc179721648)

[3. Ajuste y Optimización de Modelos 6](#_Toc179721649)

[4. Predicciones 7](#_Toc179721650)

[5. Conclusiones 8](#_Toc179721651)

[Anexo 1 “Series Temporales Preliminares”. 9](#_Toc179721652)

[Anexo 2 “Valores MAE por Método de Predicción/Tienda”. 12](#_Toc179721653)

[Anexo 3 “Predicciones por Tienda Comparativa”. 13](#_Toc179721654)

[Anexo 4 “Predicciones Finales”. 15](#_Toc179721655)

# Planteamiento del problema.

“*Buenos días Nicole, Espero que tu primera semana haya ido bien. He visto que ya te has vuelto bastante popular. Ahora que ya tienes acceso a las bases de datos de la empresa, hay una iniciativa en particular en la que me gustaría que te enfocaras.*

*DSMarket siempre ha dependido de enfoques rudimentarios para predecir las ventas de productos. Como discutimos durante tu entrevista, la razón principal detrás de tu incorporación temprana es el desarrollo de un nuevo enfoque para la predicción. Tenemos un número importante de casos de uso que dependen de la precisión de esas predicciones. Estoy seguro de que con tu larga experiencia con modelos predictivos podemos hacer un trabajo mucho mejor de lo que estamos haciendo ahora.*

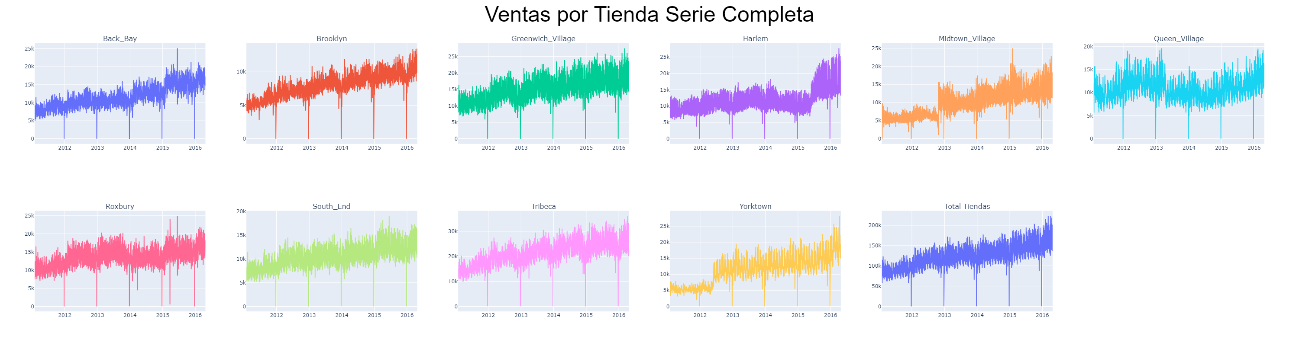
*Hasta ahora, siempre hemos ido prediciendo las ventas a nivel de tienda-producto, y para obtener ventas agregadas por departamento/tienda/ciudad sumamos las predicciones independientes.* ***¿Seguiría siendo un enfoque válido?***

***Puedes empezar con predicciones de 28 días (4 semanas).***

*Paul”*

# Acciones Prospectivas

* Para poder responder a la pregunta vemos necesario hacer las predicciones para los 28 días siguientes a distintos niveles, por artículo, por departamento, por categoría, por tienda y por total.
* Añadimos una tienda nueva que es la suma de todas las tiendas agregadas “Total Tiendas”
* Miramos las series temporales para cada una de las tiendas y observamos en los gráficos que hay diferencias sensibles a simple vista. O cuando menos cambios de tendencias bruscos en algunas de ellas.



* La imagen anterior nos sugiere que el comportamiento de los artículos en las distintas tiendas puede ser muy diferente y por lo tanto tomamos la decisión de hacer las predicciones por tienda a todos los niveles y luego agregarlas para obtener el total.
* Utilizamos las pruebas **de Dickey-Fuller aumentada (ADF) y la prueba de KPSS (Kwiatkowski-Phillips-Schmidt-Shin)**.Calendario

  Descripción generada automáticamente (ver Notebook) y obtenemos estas primeras impresiones:
  + Series Estacionarias (según ambas pruebas): Greenwich\_Village y Tribeca
  + Series No Estacionarias (según ambas pruebas): Back\_Bay, Brooklyn, Harlem, Queen\_Village, Roxbury, South\_End y Total Tiendas
  + Series Inconclusas o con resultados mixtos: Midtown\_Village (KPSS sugiere estacionariedad, pero ADF no) y Yorktown (ADF sugiere estacionariedad, pero KPSS no)

# Pruebas y Elección de Modelos

* Decidimos encontrar el mejor modelo para cada tienda de tal forma que podamos personalizar las características de cada una de ellas (estacionalidad, hiperparámetros) intuimos que posiblemente para algunas tiendas tendremos que limitar la serie a los dos últimos años para quitar efectos puntuales.
* Lanzamos entonces los siguientes modelos para cada una de las tiendas y medimos su precisión con el MAE. Incluimos los ***eventos*** como variables exógenas en todos ellos.
  + ARIMA con autoarima (Se descarga por los resultados)
  + SARIMAX con autoarima (Se descarga por los resultados)
  + Imagen que contiene Gráfico

    Descripción generada automáticamente**SARIMAX** con ***order=(1, 1, 1) y seasonal\_order=(1, 1, 1, 7***)
  + **XGBOOST** con optimización de hiperparámetros y añadiendo variables con ***lags de 1, 7 y 30 días***

Imagen que contiene Gráfico

Descripción generada automáticamente

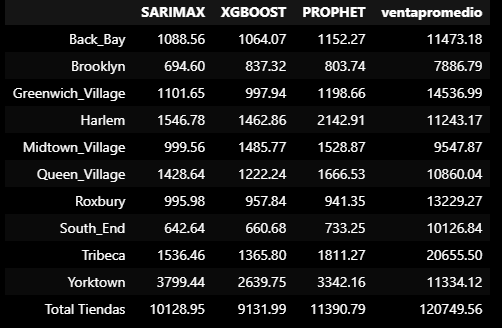
* + **PROPHET**

**Imagen que contiene Gráfico

Descripción generada automáticamente**

* Gráfico

  Descripción generada automáticamenteAcompañamos la gráfica con la venta media diaria de cada tienda para poner en perspectiva el valor del **MAE obtenido**. Y a priori obtenemos el siguiente resultado: [*Ver Anexo “Valores MAE por Método de Predicción y Tienda”.*](#_Anexo_“Valores_MAE)



* Lanzamos el modelo de menor MAE por cada tienda y obtenemos unas predicciones preliminares cuyos gráficos podemos consultar en el [*Anexo “Series Temporales Preliminares”.*](#_Anexo_“Series_Temporales) Aquí como muestra representamos el gráfico de la predicción preliminar Total Tiendas:

Imagen que contiene Interfaz de usuario gráfica

Descripción generada automáticamente

# Ajuste y Optimización de Modelos

* Una vez obtenida la primera aproximación, intentamos ajustar más los resultados probando con, distintos ***hiperparámetros***, distintos ***lags*** incluyendo uno nuevo de 365 días y acortando series temporales a 2 años.
* Después de un minucioso estudio tenemos los modelos finalistas por tienda:
  + **PROHET (0):** No se obtuvo ningún prophet como mejor modelo.
  + **SARIMAX(3):** Brooklyn, Midtown\_Village y Roxbury
  + **XGBOOTS(8):** La mayoría de las tiendas respondieron mejor a modelos xgboost, pero con distintos hiperparámetros y distintos lags.
    - ***Serie Temporal de 2 años y lags de 1,7 y 30 días***: Back\_Bay y Yorktown.
    - ***Serie Temporal Completa y lags de 1,7 y 30 días***: Harlem y Total Tiendas
    - ***Serie Temporal Completa y lags de 1,7,30 y 365 días***: Greenwich\_Village, Queen\_Village, South\_End y Tribeca
    - ***Hiperparámetros***: Además se determinan los siguientes hiperparámetros para cada uno de los centros.

Captura de pantalla de un teléfono celular

Descripción generada automáticamente

# Predicciones

* Interfaz de usuario gráfica

  Descripción generada automáticamenteLanzamos las predicciones por tienda con el mejor modelo para cada una de ellas y por los distintos niveles (Total, Tienda, Categoría, Departamento y Artículo) como se puede ver en el Notebook *“twoerria DSMARKET Artículo.ipynb”*
* Ver [*Anexo “Predicciones Finales”.*](#_Anexo_“Predicciones_Finales”.)
* Hay que señalar que las predicciones por tienda y artículo pueden llegar a tardar ***más de 8 horas*** mientras por TOTAL tarda ***menos de 1 segundo*** y por CATEGORIA Y DEPARTAMENTO apenas puede tardar ***1 minuto***.
* Obtenemos los resultados por cada tienda y cada una de las categorías
* Las agregamos para obtener un total por tienda
* Tabla

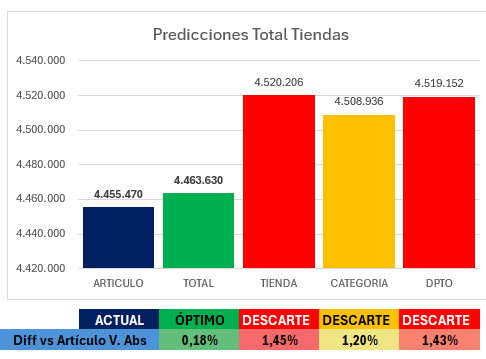
  Descripción generada automáticamenteComparamos los resultados y obtenemos esta tabla resumen:
* En la mayoría de las tiendas y en el total se puede observar que hacer una predicción directamente por total ventas obtiene unos resultados con unas diferencias menores al 0,5% de las obtenidas con la predicción por artículo.
* En el caso de Tribeca la diferencia en las predicciones por artículo y por Categoría es de tan solo 0,01%.
* Únicamente habría que prestar atención y acometer en estudio a parte el caso de Harlem que la diferencia con el nivel de artículo es de 8,7%. Excesiva diferencia a nuestro parecer.
* Gráfico

  Descripción generada automáticamenteSe puede consultar todo este desarrollo en la hoja Excel [*“Predicciones por Tienda.xlsx”*](file:///C:\Users\juang\Dropbox\MASTER%20DATA%20SCIENCE\EJERCICIOS\TFM_DSMarket\TFM%20Time%20Series\Predicciones%20por%20Tienda.xlsx), que contiene una hoja por tienda como la del ejemplo que es *Total Tiendas*  [ver Anexo “Predicciones por Tienda Comparativa”.](#_Anexo_“Predicciones_por)

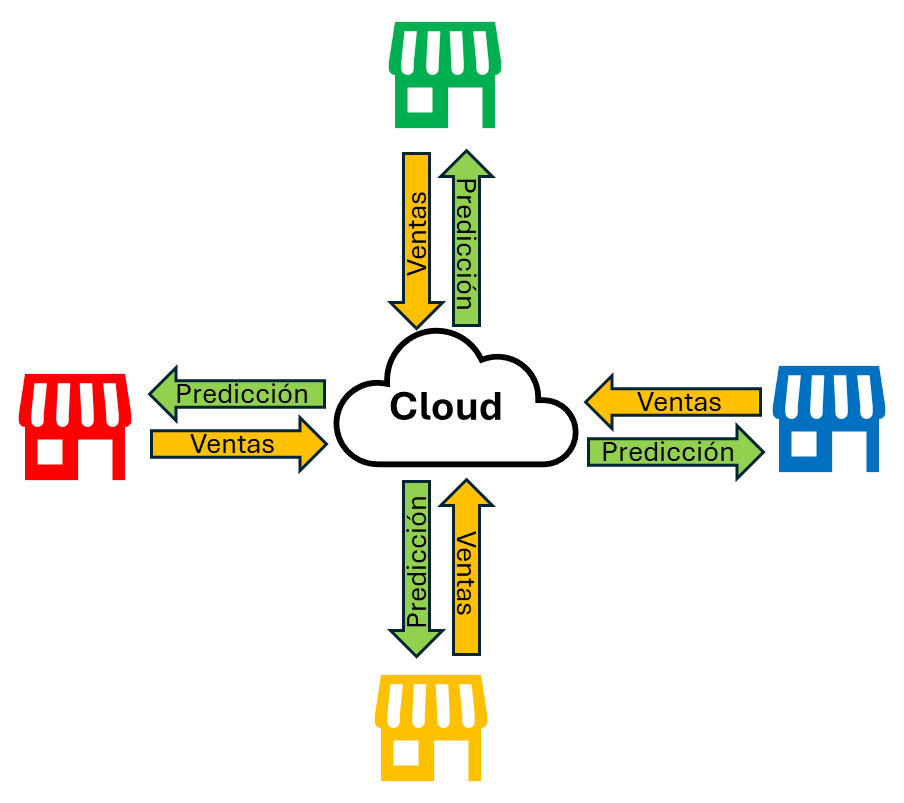
# Conclusiones

La Respuesta a la pregunta *“Hasta ahora, siempre hemos ido prediciendo las ventas a nivel de tienda-producto, y para obtener ventas agregadas por departamento/tienda/ciudad sumamos las predicciones independientes. ¿Seguiría siendo un enfoque válido?”* , sería que depende del caso de uso, es decir, depende de para que queramos las predicciones.

* Si lo que queremos es obtener una predicción total por tienda o total compañía para confección de presupuestos, previsión de tesorería, estimaciones o presentaciones a los accionistas. Se podría hacer rápidamente, en segundos con una serie temporal agregada por tienda o total compañía. De esta forma no habría que esperar al día siguiente para tener los resultados.



* Por el contrario, y anticipando otras necesidades del negocio, como un sistema de abastecimiento de stock por artículo y tienda, no debemos desechar los modelos que van al detalle, que, aunque tarden mucho en ejecutarse se podría aplicar ingeniería para agilizar el proceso dependiendo de los recursos de los que podamos disponer. Paralelizando las tareas por tienda y subiendo el trabajo al Cloud, por ejemplo, se podría reducir a un par de horas y se podría lanzar en horario nocturno de tal forma que todas las mañanas dispondríamos de las previsiones de los siguientes 28 días por artículo y tienda.

**

# Anexo 1 “Series Temporales Preliminares”.

Gráfico

Descripción generada automáticamente con confianza mediaGráfico

Descripción generada automáticamente con confianza bajaImagen que contiene Gráfico

Descripción generada automáticamenteGráfico

Descripción generada automáticamenteGráfico

Descripción generada automáticamente con confianza bajaGráfico

Descripción generada automáticamente con confianza mediaImagen que contiene Gráfico

Descripción generada automáticamenteGráfico

Descripción generada automáticamente con confianza mediaImagen que contiene Interfaz de usuario gráfica

Descripción generada automáticamenteGráfico

Descripción generada automáticamente con confianza mediaImagen que contiene Gráfico

Descripción generada automáticamente

# Diagrama, Gráfico en cascada Descripción generada automáticamenteAnexo 2 “Valores MAE por Método de Predicción/Tienda”.

# Gráfico, Gráfico de barras Descripción generada automáticamenteGráfico, Gráfico de barras Descripción generada automáticamenteGráfico, Gráfico de barras Descripción generada automáticamenteGráfico Descripción generada automáticamenteGráfico, Gráfico de barras Descripción generada automáticamenteGráfico, Gráfico de barras Descripción generada automáticamenteAnexo 3 “Predicciones por Tienda Comparativa”.

Gráfico

Descripción generada automáticamenteGráfico

Descripción generada automáticamenteGráfico

Descripción generada automáticamenteGráfico, Gráfico de barras

Descripción generada automáticamenteGráfico

Descripción generada automáticamente

# Anexo 4 “Predicciones Finales”.

Interfaz de usuario gráfica, Gráfico

Descripción generada automáticamenteInterfaz de usuario gráfica, Gráfico

Descripción generada automáticamente con confianza mediaInterfaz de usuario gráfica

Descripción generada automáticamenteInterfaz de usuario gráfica, Gráfico

Descripción generada automáticamenteInterfaz de usuario gráfica, Gráfico

Descripción generada automáticamenteInterfaz de usuario gráfica, Gráfico

Descripción generada automáticamenteGráfico

Descripción generada automáticamenteInterfaz de usuario gráfica, Gráfico

Descripción generada automáticamenteInterfaz de usuario gráfica, Gráfico

Descripción generada automáticamenteInterfaz de usuario gráfica, Gráfico

Descripción generada automáticamente

Interfaz de usuario gráfica

Descripción generada automáticamente