

Time series prediction of NYC taxi's demand using Meta prophet

Sobre la muestra de datos

Para realizar el análisis de serie de tiempo se tomaron datos de la demanda de taxis en la ciudad de Nueva York desde el 1 de junio del 2017 hasta el 31 de enero del 2018. De la totalidad de viajes registrados durante los 8 meses bajo análisis se obtuvo un muestreo aleatorio estratificado, segmentando los datos por días e incluyendo el 12.5% de los valores de cada día en la muestra. De esta manera, se logró reducir considerablemente la cantidad de datos sin perder la frecuencia diaria.

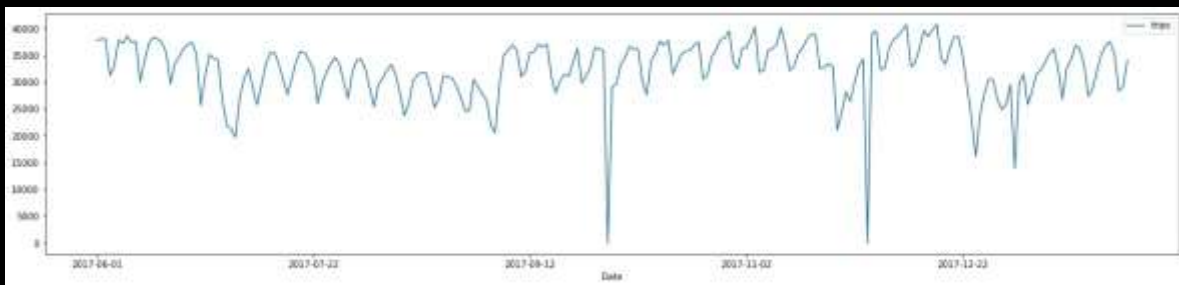


Figura 1. Muestra original

Una vez obtenida la muestra, se realizó un análisis de outliers detectando ciertos valores que se alejaban de la muestra en forma excesiva y distorsionaban el análisis. Realizando un análisis más detallado se pudo observar que estos valores coinciden con días en donde el servicio de taxis no funciono de manera habitual, resultando en contadores extremadamente bajos. Se eliminaron los valores clasificados como outliers para un mejor fitting del modelo.

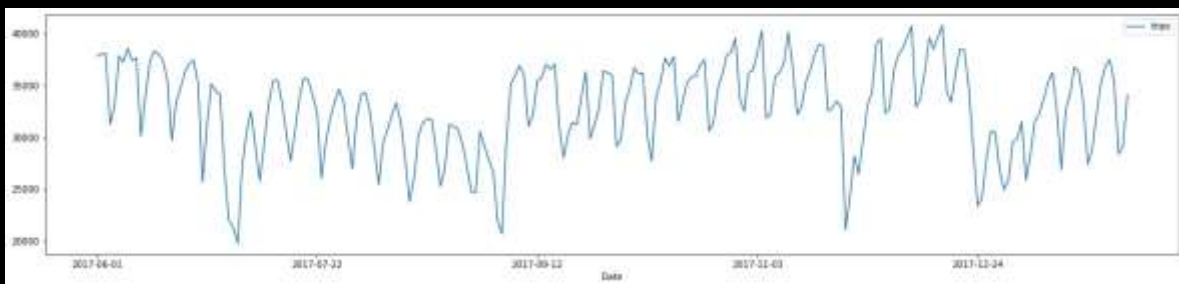


Figura 2. Muestra sin outliers

Sobre la librería utilizada

Meta Prophet es una biblioteca de código abierto desarrollada por Facebook (Meta) para el análisis de series temporales. Prophet lleva a cabo un procedimiento para pronosticar datos de series temporales basado en un modelo aditivo en el que las tendencias no lineales se ajustan a la estacionalidad anual, semanal y diaria.

Es adecuado para trabajar con series temporales que tienen fuertes efectos estacionales, como es el caso de los taxis de NYC, donde se observan cambios en la demanda según el día de la semana. Es resistente a datos faltantes y cambios de tendencia, y por lo general maneja bien valores atípicos.

Predicciones realizadas por el modelo

El modelo TSF de Prophet segrega la serie de tiempo predicha en 2 componentes. Un primer componente de tendencia general donde observamos una directriz bajista que ronda el nivel de 32.000 viajes diarios y un segundo componente de estacionalidad semanal, el cual indica que la demanda de taxis crece a medida que se acerca el fin de semana, siendo el domingo el día con menor demanda.

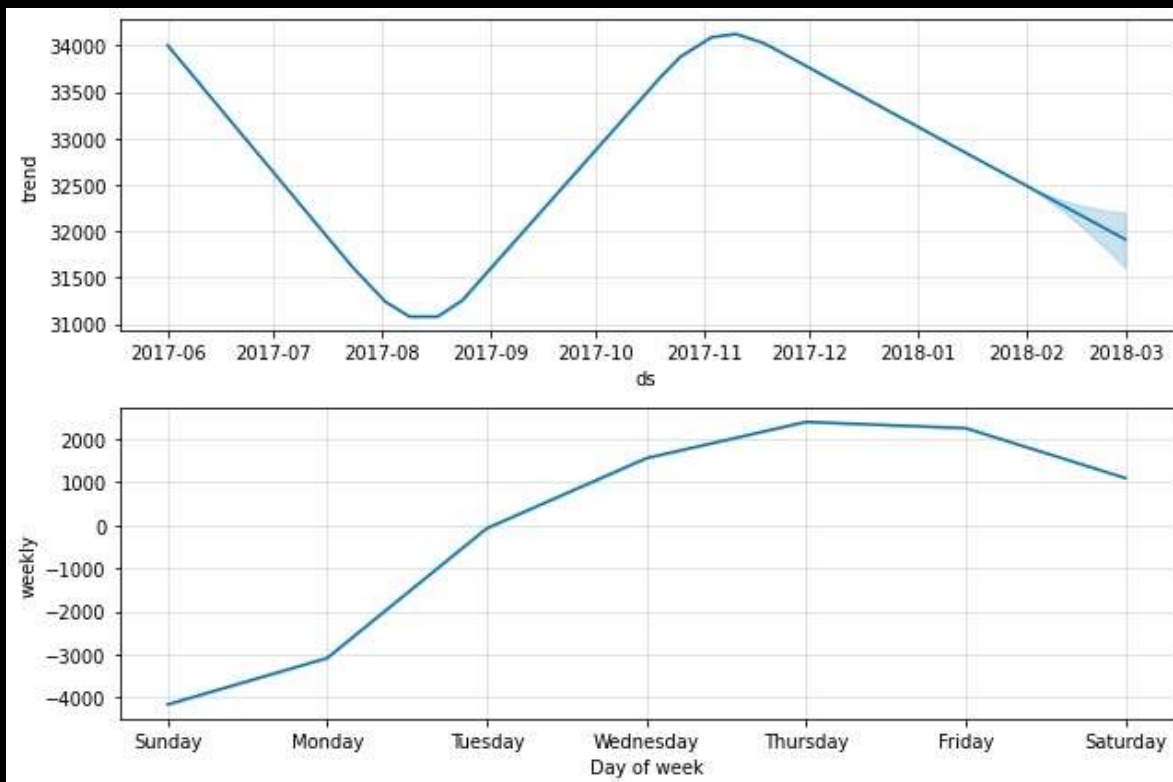


Figura 3. Componentes del pronóstico

Al combinar el componente de tendencia y el estacional, el resultado es una predicción de la demanda futura de taxis para un mes el próximo a la fecha de análisis. El pronóstico generado mediante Prophet es una onda de tipo sinusoidal que oscila entre los valores 27.800 y 35.000 indicando que la demanda para el mes de febrero se ubicará entre tales valores.

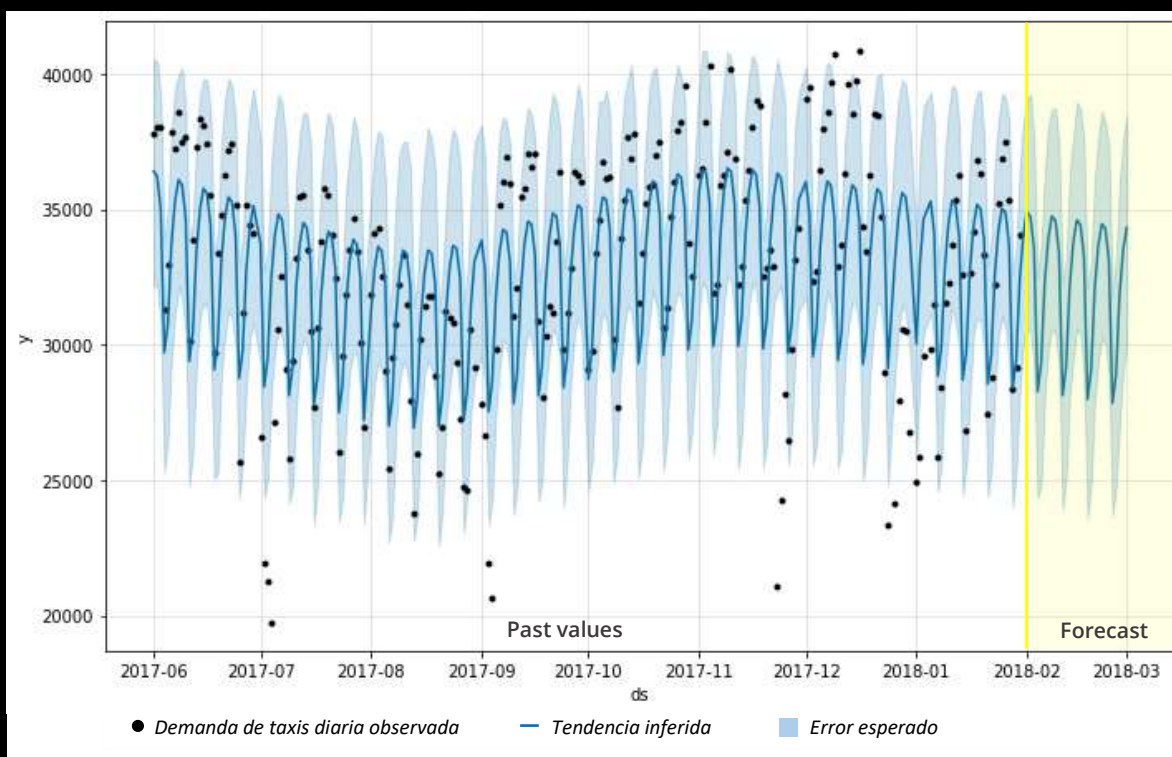


Figura 4. Predicción de la demanda para un mes en el futuro elaborada con Meta Prophet

Eficacia del modelo

Podemos medir la eficacia del modelo a través del MAPE. El Error Porcentual Absoluto Medio (MAPE o Mean Absolute Percentage Error) es un indicador del desempeño del Pronóstico de Demanda que mide el tamaño del error (absoluto) en términos porcentuales. El MAPE promedio del modelo confeccionado es del 11.17% lo cual indica que la diferencia promedio entre el valor pronosticado por el modelo y el valor real es del 11.17%. Un MAPE entre 10 % y 25 % indica una precisión aceptable para un modelo de demanda.

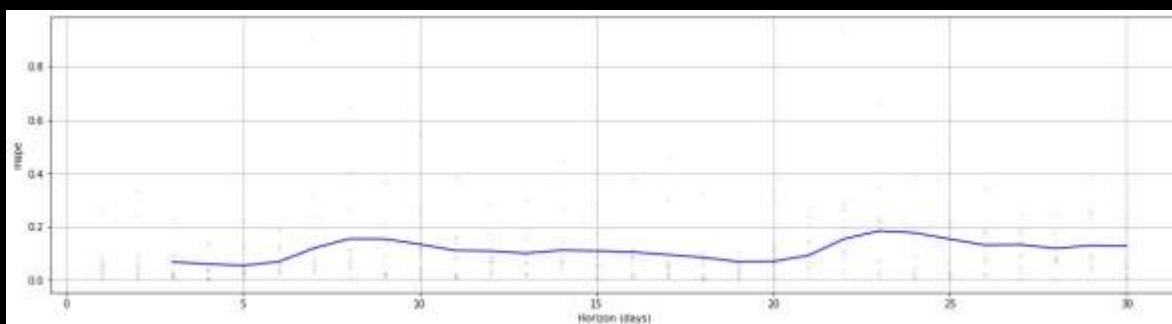


Figura 5. MAPE del modelo por días predichos