Lista de recursos

Profundiza

Challenge

Contenido extra

Science_Encuentro_41_...

Potencia tu Talento -

Entrevistas con RRHH

☐ Grabación

Series de Tiempo

bárbaro inglés, me ha sido deparado revelar ese misterio diáfano. Al cabo de más de cien años, los pormenores son irrecuperables, pero no es difícil conjeturar lo que sucedió..." –Jorge Luis Borges, El jardín de senderos que

El equipo docente subió contenido extra de la meeting 🔆

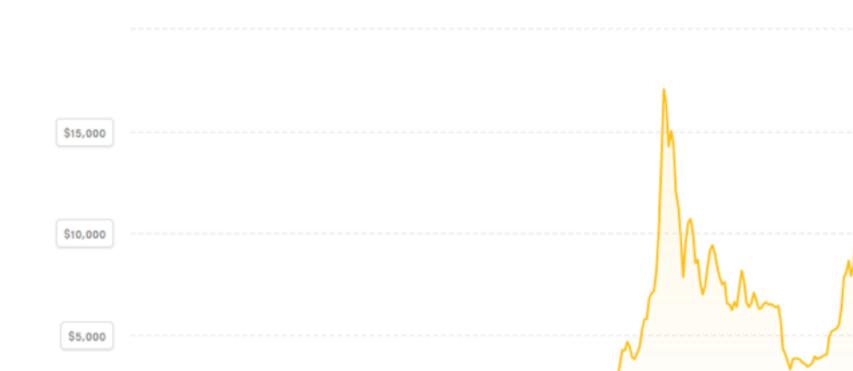
se bifurcan. ¡Gracias Sabina Bercovich por la creación de los contenidos de esta bitácora!

La respuesta a esta pregunta es bastante más simple de lo que tal vez creemos:

Generalmente son mediciones de una misma variable, o sucesos que pasaron en un determinado momento del tiempo. ¿Y entonces en qué se diferencia de un dataset con una columna de fechas? Un

dataset es una serie de tiempo cuando sus datos de fechas se encuentran equiespaciadas. Es decir, si tenemos por ejemplo un dato por hora, o uno por día, etc. Podría pasar que encontremos datasets donde los datos no están igualmente distribuidos en el tiempo, y en ese caso sería parte del preprocesamiento generar que así sea (por ejemplo, resumiendo nuestros datos con algún criterio en especial).

para que haya uno por cada hito temporal, o rellenando los datos que no existen ¡Podemos encontrar muchos ejemplos de series de tiempo en nuestro día a día! Temperatura Histórica Buenos Aires, promedio semanal:



resuena con hechos que influyen en el valor medido, a veces directamente, otras indirectamente. Esta bitácora fue escrita en plena pandemia de Covid-19, y seguramente habrás visto muchos gráficos de la evolución de la epidemia en tu país, ciudad, o en el mundo. Al explicar estos gráficos, se suele señalar el inicio de la epidemia, cuándo se implementa una cuarentena, cuándo se levanta o se relaja, y cómo estas decisiones afectan la cantidad de contagiados/as.

tiempo. ¡Estas cantidades de tiempo pueden ir desde nanosegundos hasta años! Pensemos algunos ejemplos:

Frecuencia de muestreo y ventanas de

• Frecuencia de segundos: Cantidad de procesos ejecutados por una CPU • Frecuencia de minutos: Cantidad de transacciones bancarias • Frecuencia de horas: Cantidad de visitas a una página web

Mencionamos más arriba el detalle (¡no menor!) de que los datos se encuentran

equiespaciados. Esto se refiere a que los datos tienen una frecuencia de

muestreo: nuestro set de datos guarda un registro cada cierta cantidad de

• Frecuencia mensual: Inflación de precios

tiempo

- Frecuencia anual: Aumento del PBI
- Es muy importante entender cuál es la frecuencia temporal de nuestro set de datos antes de empezar a trabajarlo. Esto nos va a permitir, no solo una mejor comprensión, sino también la posibilidad de resumir información en una
- granularidad de tiempo más grande.
- frecuencia más chica a una más grande. Por ejemplo, unificando los registros horarios (que serían 24 por día) a un solo registro diario.

para cada categoría. Otro concepto importante es el de ventanas de tiempo. Consiste en una agrupación de varios registros temporales (datos), y tiene un inicio y un final. Si estamos trabajando con granularidad diaria, una ventana de tiempo posible es

una semana. En este caso, nuestra ventana tendría tamaño 7, es decir que

contempla siete registros. Si estuviéramos hablando de datos horarios, una

ventana de tiempo muy utilizada es 24, es decir, un día. Y así sucesivamente.

¡Escucharás y leerás sobre estas ventanas de tiempo muchas veces a lo largo de

¿Qué podemos hacer con una serie de tiempo? ¡Las posibilidades son muy amplias! Por ejemplo: • Análisis Exploratorio de Datos. Como sabes, los EDA son más interesantes cuando cuentan una historia, y las series de tiempo son particularmente útiles para este fin.

• Pronóstico del clima

tu formación!

son:

• Tráfico en una página web • Ventas de un negocio, empresa, etc.

• Valor del dólar, moneda u acciones

- serie de su cotización. También puede ser un quiebre la implementación de una cuarentena estricta en la curva de contagiados por Covid-19. Este problema también puede ser abordado - según el contexto - como un problema de clasificación o regresión. Descomposición de la serie: Ruido, Tendencia, Estacionalidad y Anomalías
- Dentro de las series de tiempo podemos encontrar series muy diversas. Algunas parecen tener una estructura bastante clara (por ejemplo, las temperaturas promedio), mientras que otras son mucho más aleatorias en su comportamiento, como puede ser el valor de una acción.

decrece, o no logramos identificarla, decimos que es no estacionaria en la media. En el siguiente gráfico, la primera serie es estacionaria, mientras que la

decimos que nuestra serie es estacionaria en la media. Si por lo contrario crece,

Ya estuvimos hablando un poco de este componente: nos indica la **dirección**

global que está tomando nuestra serie. Nuestra tendencia puede ser creciente,

decreciente, estática. Incluso existen series de tiempo donde no podemos llegar

Cuando nuestra tendencia es estática, es decir, que se mantiene "recta",

Pasajeros internacionales, Pan Am, EEUU, 1949 - 1960. (Fuente) y Temperatura promedio en Buenos Aires. Un proceso es - o al menos parece - estacionario en

Estacionalidad

Ruido blanco

todo k≠o.

Tendencia

a definirla concretamente.

la media, mientras que el otro no. Considera que dependiendo de la ventana de tiempo que estés mirando, la tendencia puede cambiar completamente. Retomando el ejemplo de la temperatura promedio, encontramos que en los meses de otoño la temperatura tiene una tendencia decreciente, pero pasa a ser creciente en los meses de

primavera. Incluso, dado el calentamiento global, podríamos llegar a analizar la

tendencia en una ventana temporal mucho más grande, analizando si la

temperatura promedio fue creciendo o decreciendo en los últimos años.

utilización en las horas pico (estacionalidad horaria). Incluso podemos observar estas variaciones en ventanas de tiempo de meses, encontrando que año a año probablemente durante los meses de verano, las vacaciones escolares o la semana de fin de año, haya una baja respecto a otros meses. Encontrar estos patrones nos ayudan a comprender en mayor profundidad nuestra serie, y eventualmente poder generar predicciones más acertadas sobre su comportamiento.

Podemos definir al **ruido blanco** como una secuencia de variables aleatorias no

(generalmente se asume media o), varianza constante y covarianza cero para

¡Buenísimo! ¿y qué significa todo esto? Básicamente podemos pensar en el

ruido blanco como un proceso que no responde específicamente a nada, que no

correlacionadas, que siguen una misma distribución, con igual media

Anomalías (Outliers) Las anomalías son valores atípicos que podemos encontrar en nuestros

datasets. Son valores que se alejan mucho de la media. Ahora, ¿qué significa

Si bien ya estuvimos hablando de anomalías a lo largo de la cursada, cobran

especial importancia cuando trabajamos con series de tiempo. Saber

"mucho"? Como siempre, jesto va a depender del problema que estemos

identificarlas y trabajar con ellas nos ayudará a tener mejores resultados a la hora de intentar encontrar los componentes de nuestra serie, tanto estacionalidad como tendencia. Pasar por alto un análisis de anomalías puede

analizando!

¿Cómo obtenemos esos componentes? Encontrar las diferentes partes de nuestra serie de tiempo es probablemente uno de los aspectos más relevantes de nuestro trabajo. Generalmente, cuando empezamos abordamos una serie, lo primero que hacemos es graficarla: es importante comprender cuál es la **frecuencia de muestreo** de nuestra serie

(cada cuánto tiempo se registra un nuevo dato, o en otras palabras, cuánto

tiempo hay entre un dato y el siguiente).

Una vez que ya comprendemos la frecuencia de muestreo, podemos ir probando graficar distintas ventanas temporales y distintos resampleos, y analizando los gráficos resultantes. De estar trabajando con frecuencias muy pequeñas, en una ventana de tiempo muy grande, es probable que el gráfico sea difícil de leer. Seguramente dentro de los gráficos empecemos a notar algunas de las componentes: una tendencia y/o estacionalidad marcada podrán ser

estadísticos específicos para este tipo de datos, y recomendaciones sobre algunas librerías ya existentes que te facilitarán el trabajo más complejo.

Aprenderás más sobre este tema en la próxima bitácora: conocerás modelos

durante el encuentro.

Data Science_Encuentro_41_ Series de tiempo

Puedes descargar el dataset necesario aquí.

¡Prepárate para el próximo encuentro!



Profundiza



Contenido extra de la meeting

Procesamiento Lenguaje Natural (Part...

Código de acceso para acceder: Acamica123°

Series de tiempo (Parte 2)

"Un laberinto de símbolos —corrigió. Un invisible laberinto de tiempo. A mí,

Grabación ¿Qué es una serie de tiempo?

Una serie de tiempo es una secuencia de valores ordenados en el tiempo.

Cotización histórica bitcoin:



• Frecuencia diaria: Cantidad de ventas de un local

• Etc.

¿Y qué sucede si tengo datos que no son cantidades? Generalmente unificamos dentro de una misma categoría. Por ejemplo, si estamos trabajando con ventas, unificamos las mismas por hora (si esta es nuestra granularidad) y por la categoría que nos interese estudiar, y nos termina quedando un registro diario

A este proceso de resumir datos temporales se le llama resampleo. Consiste en

sumar los datos que no son fechas (en general cantidades) para pasar de una

• Modelos de Forecasting (predicción a futuro). Ésta es una de las

aplicaciones más utilizadas y demandadas. Consiste en, dado los valores

históricos de una serie, predecir los valores a futuro. Algunas aplicaciones

• Modelos de Clasificación y Regresión. Las series de tiempo también

sirven como información de entrada en modelos de clasificación y regresión. Por ejemplo, en Astronomía es muy común clasificar elementos a partir de cómo varía su brillo en el cielo o su posición. En geología, se puede predecir el momento de un terremoto a partir de señales sísmicas; a

partir de señales de audio se puede identificar al hablante, etc. Las

- posibilidades son sumamente extensas. • Detección de puntos de quiebre y anomalías. Es una aplicación exclusiva de las series de tiempo. Un punto de quiebre es un cambio de comportamiento, en general abrupto, de la serie. Por ejemplo, la devaluación fuerte de las acciones de una empresa es un quiebre en la
- Por otro lado, tal vez podamos identificar una tendencia marcada, es decir que nuestra serie va creciendo o decreciendo con el tiempo de manera relativamente constante. Incluso podríamos tener una tendencia que no crece ni decrece, sino que se mantiene jy esto también es una tendencia posible!. Vamos a revisar entonces las distintas partes que componen a una serie de tiempo, como si hiciéramos una suerte de descomposición de porque una serie de tiempo termina siendo como es.

segunda no lo es.

Una de las partes más estudiadas dentro de la ciencia de datos respecto de las series temporales es la **estacionalidad**. La misma nos habla de los procesos que tienen comportamientos repetitivos a lo largo del tiempo. Por ejemplo, analizando la utilización de transporte público, podemos encontrar que los días de la semana tienen mayor flujo que los fines de semana (estacionalidad semanal o diaria), pero dentro de cada día también hay mayor

tiene un sentido concreto, sino que son valores en el tiempo completamente aleatorios. Generalmente, toda serie de tiempo tiene un componente de ruido blanco, y suele ser esta parte la que no vamos a poder predecir (y eventualmente nos justifique una parte del error de medición o predicción que terminemos teniendo).

terminar distorsionando ampliamente nuestras componentes y resultados de distintos modelos.

identificadas fácilmente con un ojo curioso. ¡Incluso los gráficos resultan muy convenientes a la hora de encontrar anomalías!

promedio de distintas ventanas de tiempo consecutivas.

Empieza a trabajar en el notebook

Para identificar la tendencia en una serie, podemos estudiar cómo se ajusta una

regresión lineal a nuestros datos, o simplemente analizando (y graficando) el

Para el encuentro, lleva hecha hasta la sección 1, el resto lo trabajaremos

Potencia tu Talento - Entrevistas con RRHH

Te proponemos el siguiente desafío, ¿te animas?