

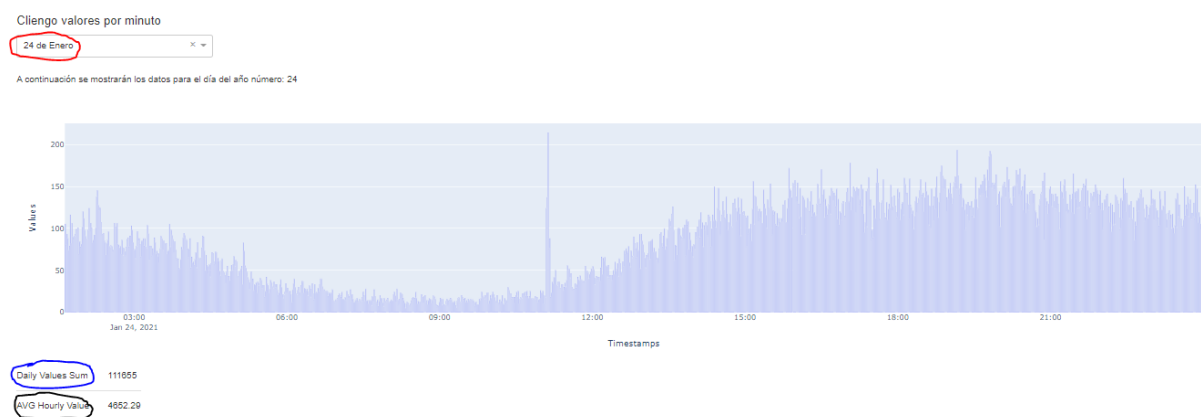
Data Engineering Valentín Sarubbi

Executive Summary

La finalidad de este informe es que el lector comprenda la dinámica de los requests de una API, con el propósito de evaluar al candidato Valentín Sarubbi para el puesto de Data Engineering.

Para ello se ha analizado un set de datos en Python y se ha generado una app web que grafica los RPM con selector de fechas y se ha desarrollado el presente informe.

A continuación en Graf. 0 se expone la app web y su funcionamiento.



Graf. 1 Captura de app web de los RPM del día 24 con resaltado de elementos.

En Gaf. 1 se encuentra graficado los RPM del día 24 de enero.

Se puede apreciar que se encuentra remarcado arriba con color rojo el selector de fechas, es un menú desplegable que permite seleccionar el día que se va a visualizar en la web.

Abajo del gráfico se encuentra una tabla que indica la cantidad total de request que hubo en el día, remarcado en azul con el nombre **Daily Values Sum** y el promedio de requests que hubo en una hora a lo largo del día remarcado con color negro llamdo **AVG Hourly Values**. Nótese que el cálculo del promedio es válido por día y teniendo en cuenta las horas del día.

Luego en el presente informe se adjunta la sección de Análisis de datos analizando los valores y distribución de los RPM en los días 24, 25 y 26 de enero.

Del análisis se desprende que el día 25 presenta la mayor cantidad de Requests. Siendo el promedio por día el siguiente:

Análisis de Request per Hour según día.			
Día	24	25	26
Requests per Hour	4652	9925	4800
Promedio	6459		
Diferencia respecto promedio	-1807	3466	-1659

Análisis de datos:

Con la finalidad de entender el comportamiento de los RPM (Requests Per Minute) de una API se procede a calcular máximos, mínimos y distribución horaria.

Valores representativos:

El valor **máximo** de request per minutes se da el día 2021-01-25 16:49:00+00:00 con el valor de **869 RPM**. (Nótese que el valor máximo podría tratarse de un outlier).

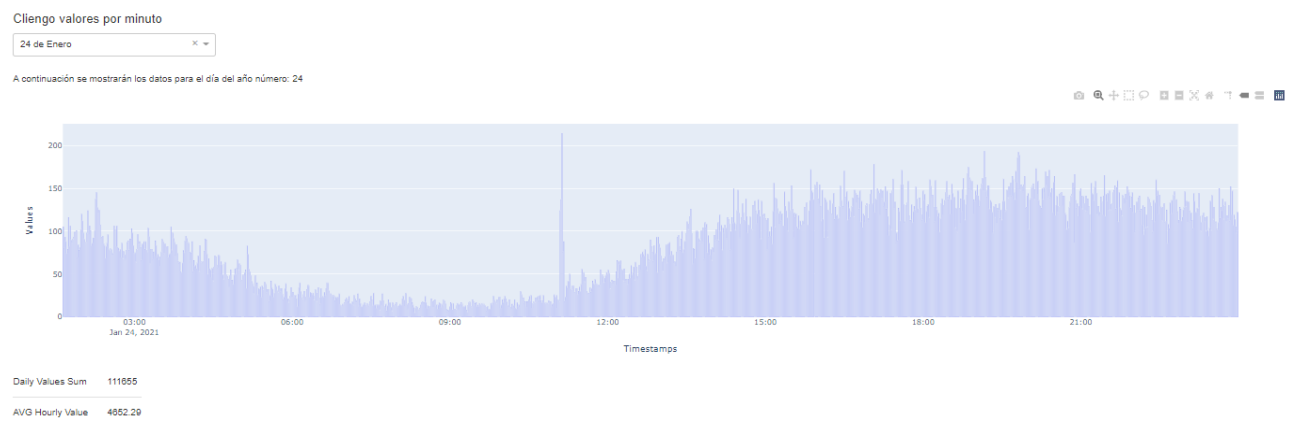
El valor **mínimo** de request per minutes se da el día 2021-01-24 08:20:00+00:00 con el valor de **5 RPM**. (Nótese que el valor mínimo podría tratarse de un outlier).

El valor **medio** de request per minutes de la serie analizada corresponde a **115 RPM**.

Distribución Horaria:

A continuación se procede a analizar la distribución horaria con la finalidad de que el lector comprenda la dinámica de request de la API.

Día 24:

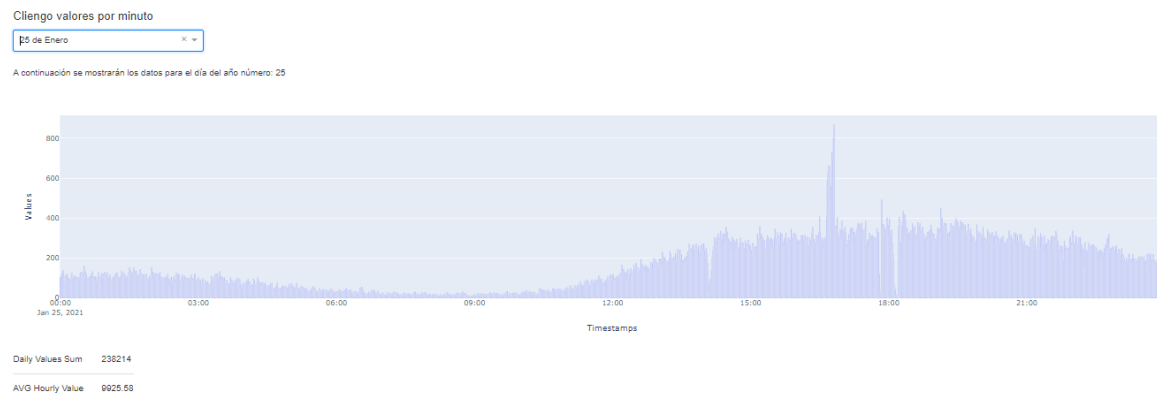


Graf. 1 Captura de app web de los RPM del día 24

En Graf. 1 correspondiente a RPM del día 24 de Enero se puede apreciar que los Request siguen una distribución en forma de campana. Teniendo el pico aproximadamente a las 19 hs y los puntos valle a las 9 hs.

La cantidad de requests en el día se pueden ver en la app como **Daily Values Sum** siendo esta 111655 y la cantidad promedio de Requests por hora como **AVG Hourly Value** igual a 4652.

Día 25:

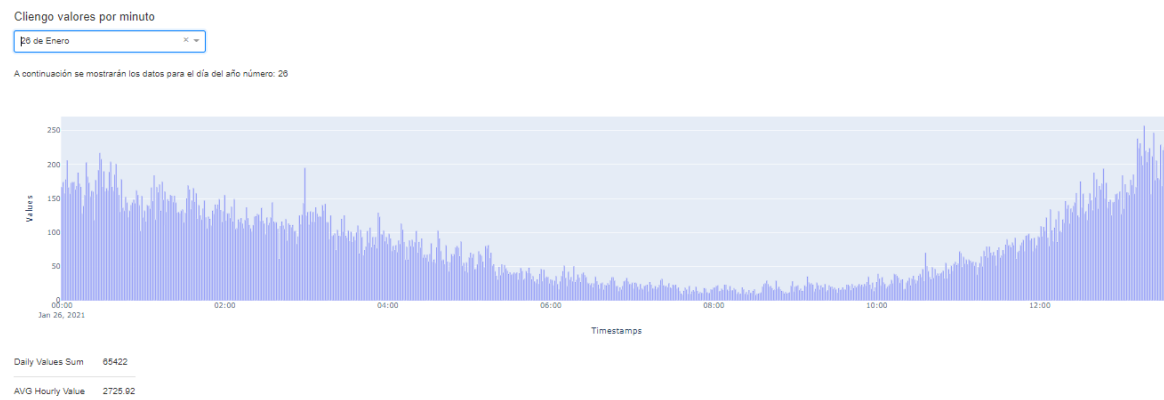


Graf. 2 Captura de app web de los RPM del día 25

En Graf. 2 correspondiente a RPM del día 25 de Enero se puede apreciar el punto valle mencionado con anterioridad a las 9 hs.

La cantidad de requests en el día se pueden ver en la app como **Daily Values Sum** siendo 238214 y la cantidad promedio de Requests por hora como **AVG Hourly Value** igual a 9925.

Día 26:



Graf. 3 Captura de app web de los RPM del día 26

En Graf. 3 correspondiente a RPM del día 26 de Enero nuevamente es posible observar el valle a las 9 hs.

Esta serie se encuentra incompleta teniendo datos desde las 0 hs hasta las 13 hs. Por lo tanto no son comparables los valores. El promedio de RPH (request per hour) se encuentra en aproximadamente **4800** similar al del día 24.

Conclusión y recomendaciones:

El día lunes 25 de enero tiene aproximadamente el doble de RPM que los días 24 y 26.

Se recomienda realizar el análisis sobre una serie más larga teniendo datos de las últimas 2 o 3 semanas como mínimo. Ya que al analizar días aislados podría tratarse de días atípicos, por otra parte si bien no se ha informado a que pertenece la API, muy probablemente el comportamiento de los fines de semana sea diferente a los días de semana.

Por otro lado sería interesante tener algún otro tipo de variable con la finalidad de encontrar alguna correlación y generar planes de acción para perfeccionar la performance.