****

**Simulación**

**TP Final**

**Settlement Feed**

**K4153**

**Docente de cursada**: Silvia Monica Quiroga

**Año**: 2021

|  |  |
| --- | --- |
| **Alumno** | **Legajo** |
| Pompey, Pablo Sebastián | 157.181-3 |
| Sanchez, Mauricio Ramon | 144.059-7 |
| Villegas, Maryann | 145.027-0 |

# 

# Índice

Contenido

[Índice 2](#_Toc78659472)

[Enunciado 2](#_Toc78659473)

[Análisis 3](#_Toc78659474)

[Metodología utilizada 3](#_Toc78659475)

[Clasificación de variables 3](#_Toc78659476)

[Clasificación de eventos 4](#_Toc78659477)

[Tabla de eventos futuros 4](#_Toc78659478)

[Condiciones iniciales 4](#_Toc78659479)

[Análisis de los datos 5](#_Toc78659480)

[Procesamiento de transacciones de Lunes a Viernes hora pico: 5](#_Toc78659481)

[Análisis de funciones de densidad de probabilidad 8](#_Toc78659482)

[Procesamiento de datos Lunes, Martes, Miércoles y Viernes(VMP) hora pico: 8](#_Toc78659483)

[Procesamiento de datos hora no pico: 1](#_Toc78659485)0

[Procesamiento de datos Jueves(VMPJ) hora pico: 13](#_Toc78659486)

[Diagrama 1](#_Toc78659487)6

[Simulación 20](#_Toc78659489)

[Escenarios propuestos. 2](#_Toc78659490)0

[Conclusiones 2](#_Toc78659491)2

# 

# Enunciado

La simulación se realiza para una de las Fintech más importantes de Latinoamérica, que diariamente tiene miles de movimientos de las transacciones realizadas por sus clientes.

De estos movimientos, se alimenta un sistema que genera reportes para todos los usuarios registrados en este sistema de reportería, uno de los reportes más importantes es el reporte de “Todas las transacciones” que muestra en el mismo, las transacciones que llegan con un volumen entre 1k y 10k de movimientos por minuto, teniendo una hora pico de lunes a viernes de 9hs a 00hs.

El modelo para analizar es sobre el sistema de reportería, uno de los principales problemas, es asegurar la consistencia de los reportes. Por eso, cada vez que haya encolado más de cierta cantidad de mensajes, llamamos lag alto, se deben parar los generadores de reportes. Una vez que la cantidad de mensajes en la cola baje a otra cierta cantidad, que llamamos normalizadora, se vuelven a encender los generadores.

Este proceso, se quiere automatizar para que no dependa de la responsabilidad de una persona el pausado y el encendido de estos generadores.

Otro punto importante a tener en cuenta es la cantidad de instancias que consumen la cola de los movimientos. Ya que las instancias que se crean están en Google Cloud, crear nuevas tiene un costo de infraestructura que hay que evaluar.

Se desea conocer, la cantidad de instancias óptimas a crear para el consumo de estos movimientos, teniendo en cuenta cuantas veces promedio por día se debe pausar el generador de reportes, y durante cuánto tiempo promedio estuvieron pausados.

# Análisis

## Metodología utilizada

La metodología empleada para el análisis del sistema se basa en incrementos constantes de tiempo (metodología delta “t” constante).

Intervalo de tiempo ∆t: 1 minuto (mensajes por minuto).

## Clasificación de variables

|  |  |  |  |
| --- | --- | --- | --- |
| **Exógenas** | **Dato** | VMP | Volumen de Movimientos por minuto. |
| VPMJ | Volumen de Movimientos por minuto del jueves. |
| VMPNP | Volumen de Movimientos por minuto en horas no pico |
| **Control** | CI | Cantidad de instancias de procesador de movimientos (Unidades) |
| TLL | Threshold low lag (Unidades) |
| THL | Threshold high lag (Unidades) |
| **Endógenas** | **Estado** | RPP | Cantidad de registros que están a la espera de procesamiento por minuto (Unidades) |
| **Resultado** | PTP | Porcentaje de tiempo que los generadores estuvieron pausados |
| TMP | Tiempo máximo que el generador estuvo pausado. |
| MYL | Mayor lag en la cola de movimientos |
| MNL | Menor lag en la cola de movimientos |

## Clasificación de eventos

|  |  |
| --- | --- |
| **Eventos propios** | Consumo de Movimientos  Llegada de Movimientos |
| **Eventos comprometidos en ΔT anteriores** | - |
| **Eventos comprometidos para ΔT futuros** | - |

## Tabla de eventos futuros

-

# Condiciones iniciales

|  |  |  |
| --- | --- | --- |
| **Variable** | **Especificación** | **Valor** |
| RPP | Registros que están a la espera de procesamiento | 0 (Unidades) |
| EG | Estado del Generador | False (Boolean) |
| CI | Cantidad de Instancias Disponibles | 4 (Unidades) |
| RPM | Request Por Minuto | 750 hora no pico/12500 hora pico (Unidades por Minuto) |
| STA | Sumatoria Tiempo Apagado del G | 0 (Minutos) |
| ITA | Intervalo de Tiempo Apagado del G | 0 (Minutos) |
| TMP | Tiempo máximo que el generador estuvo pausado | 0 (Minutos) |
| MYL | Mayor lag en la cola de movimientos | 0 (Unidades) |
| MNL | Menor lag en la cola de movimientos | 999999999 (Unidades) |
| T | Tiempo de simulación | 0 (Minutos) |
| TF | Tiempo final de la simulación | 525600 (Minutos) |
| M | Contador de Tiempo Auxiliar para los minutos del día | 0 (Minutos) |
| DAY | Contador de Tiempo Auxiliar para indicar el día | 1 (Unidades) |
| THL | Threshold high lag | 2500 (Unidades) |
| TLL | Threshold low lag | 1000 (Unidades) |

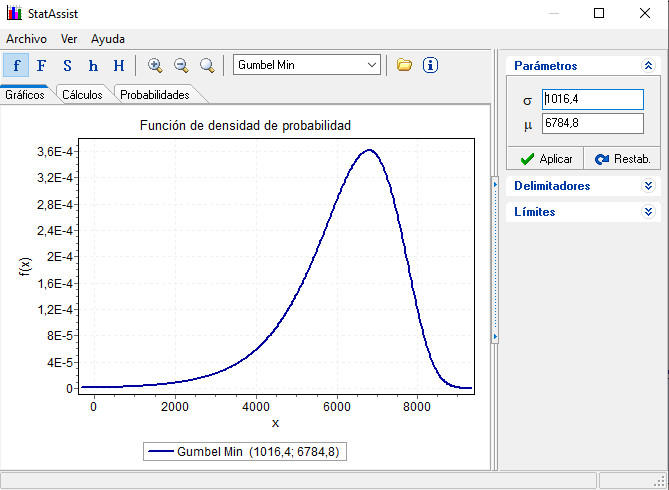
# Análisis de los datos

Partiendo de los datos obtenidos de un mes entero de transacciones, realizamos el análisis de forma mensual, diaria y semanal. Dado que los fines de semana, el sistema de reportería no se apaga porque no llegamos al límite de mensajes esperando en la cola, no vamos a tener en cuenta los datos generados estos días a los fines de esta simulación.

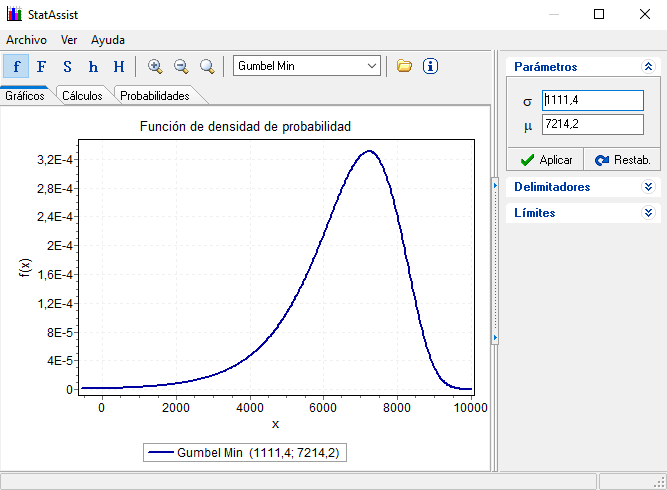
### Procesamiento de transacciones de Lunes a Viernes hora pico:

Analizamos de forma diaria las transacciones que se dieron en cada día de la semana, a lo largo de un mes y obtuvimos:

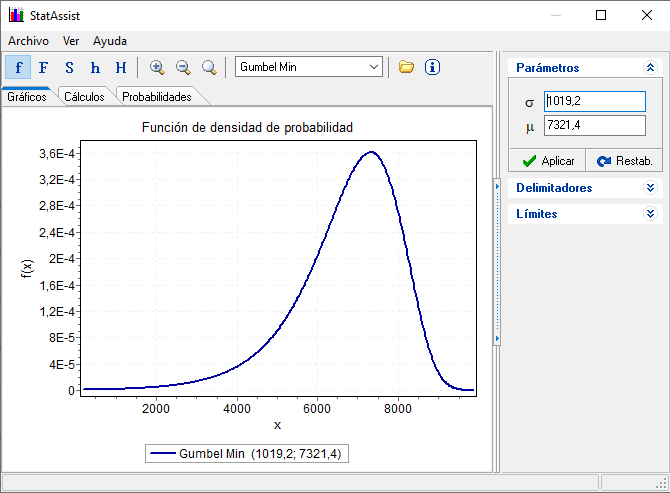
Lunes: En este día, obtuvimos una distribución de Gumbel Min



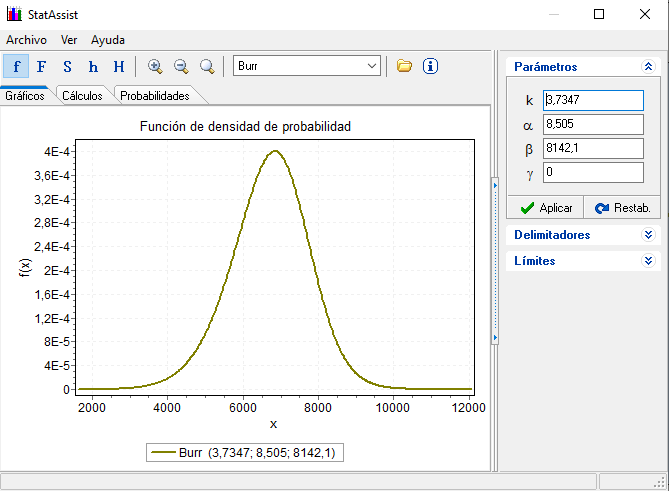
Martes: En este día, obtuvimos una distribución de Gumbel Min



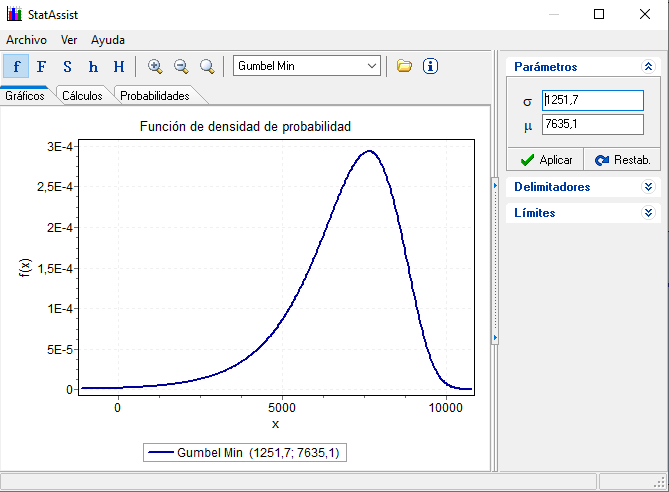
Miércoles: En este día, obtuvimos una distribución de Gumbel Min



Jueves: En este día, obtuvimos una distribución de Burr



Viernes: En este día, obtuvimos una distribución de Gumbel Min



Cuando analizamos día a día durante la hora pico(9 a 23hs), pudimos observar que los lunes, martes, miércoles y viernes, obtenemos una distribución Gumbel Min pero los días Jueves, dado que este día se dan distintas ofertas y promociones (que hacen los movimientos se den con otro volumen), la distribución que mejor se ajusta es la Burr, el analisis de datos se realizó de forma separada, por un lado la hora pico y por el otro la hora no pico.

Al analizar todos los datos juntos pudimos observar que la distribución que obtenemos no es la más precisa, lo que nos llevó a partir nuestro análisis, por un lado utilizaremos la función Gumbel Min obtenida con los datos de los días lunes, martes, miércoles y viernes, y por el otro lado, utilizaremos la distribución Burr para el día Jueves en nuestra simulación y para el horario no pico la distribución que mejor se ajusta es la Weibull (3P).

# Análisis de funciones de densidad de probabilidad

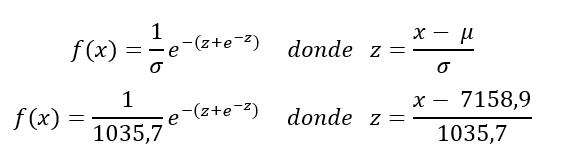
### Procesamiento de datos Lunes, Martes, Miércoles y Viernes (VMP) hora pico:

La fdp de VMP representa el volumen de movimientos por minutos que se dan los días mencionados. El valor mínimo es de X y el valor máximo es de X.

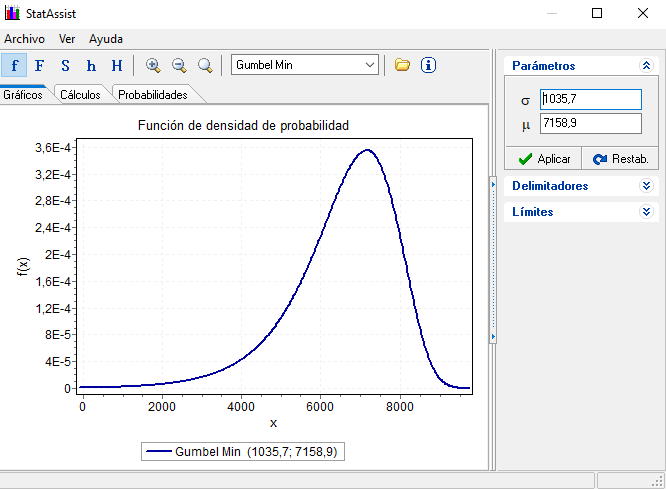
La distribución que mejor se aplicó a los datos fue la Gumbel Min.

#### **Cálculos:**

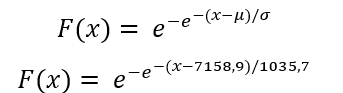
*Función de densidad de probabilidad f(x):*

**

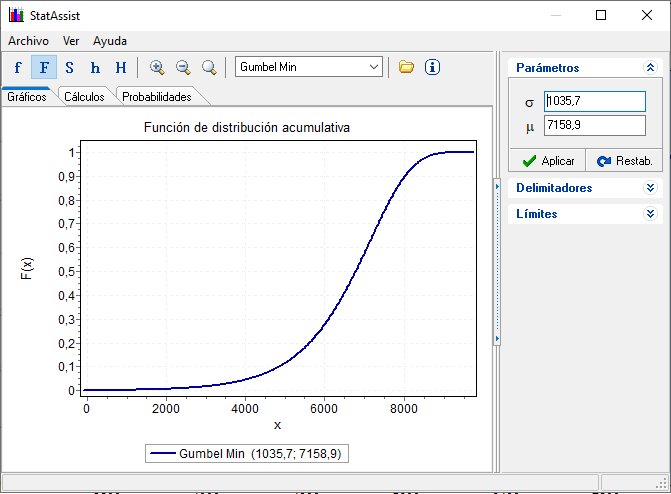
*Gráfica de f(x):*



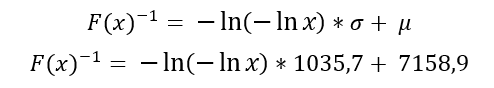
*Función de distribución acumulada F(x):*

**

*Gráfica de F(x):*



*Inversa de F(x):*

**

#### **Diagrama de la rutina:**

# 

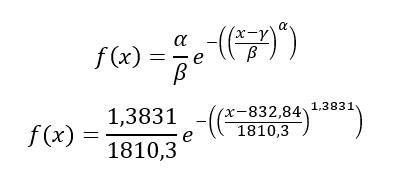
### Procesamiento de datos hora no pico:

La fdp de VM PNP representa el volumen de movimientos por minutos que se dan de lunes a viernes en la hora no pico. El valor mínimo es de X y el valor máximo es de X.

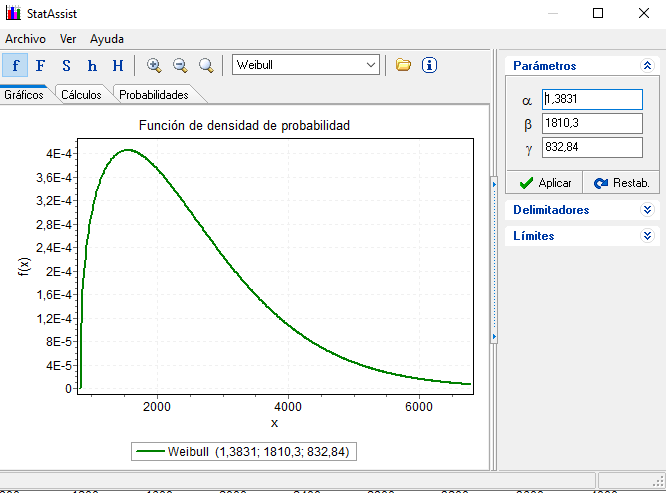
La distribución que mejor se aplicó a los datos fue la Weibull (3P)

**Cálculos:**

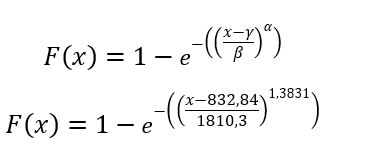
*Función de densidad de probabilidad f(x):*

**

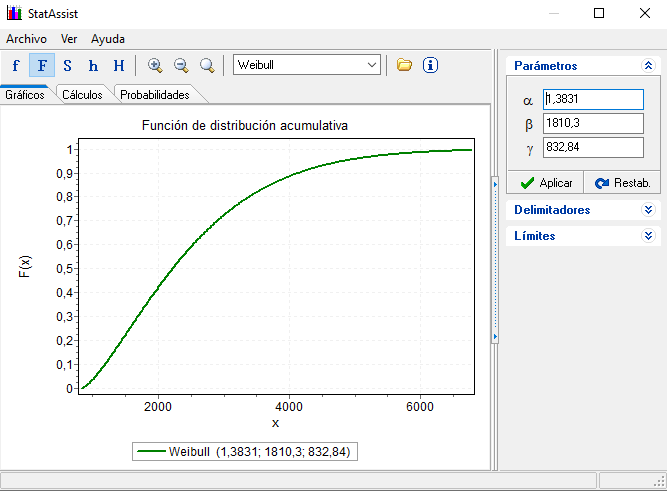
*Gráfica de f(x):*



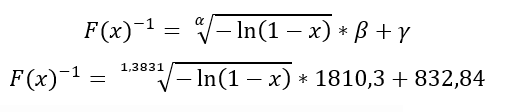
*Función de distribución acumulada F(x):*

**

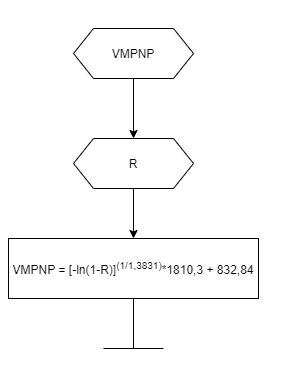
*Gráfica de F(x):*



*Inversa de F(x):*

**

#### **Diagrama de la rutina:**



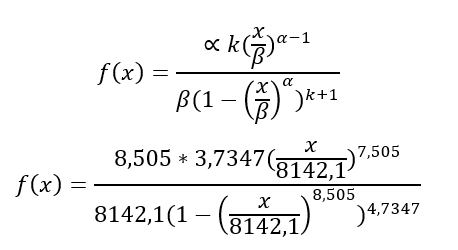
### Procesamiento de datos Jueves(VMPJ) hora pico:

La fdp de VMPJ representa el volumen de movimientos por minutos que se dan el día Jueves. El valor mínimo es de X y el valor máximo es de X.

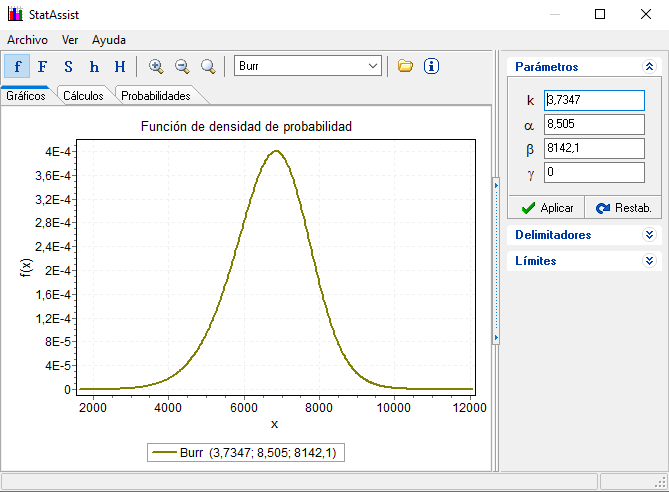
La distribución que mejor se aplicó a los datos fue la Burr.

.**Cálculos:**

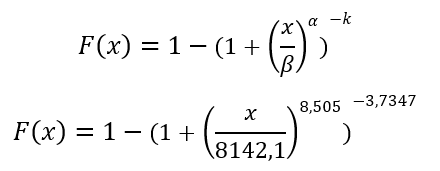
*Función de densidad de probabilidad f(x):*

**

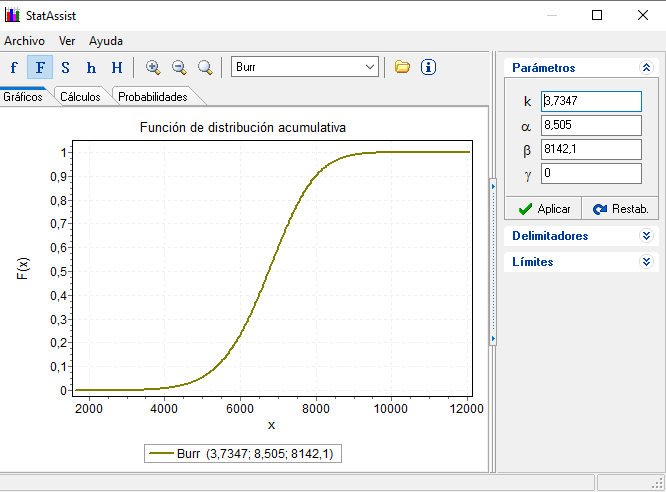
*Gráfica de f(x):*

**

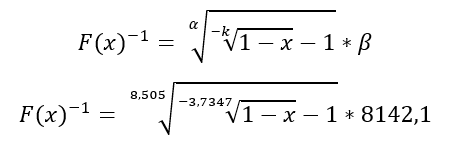
*Función de distribución acumulada F(x):*

**

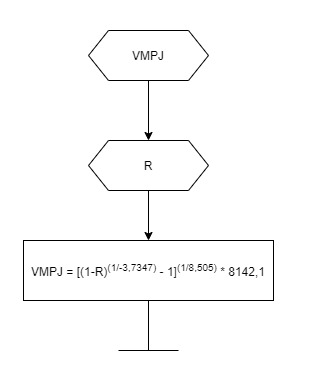
*Gráfica de F(x):*



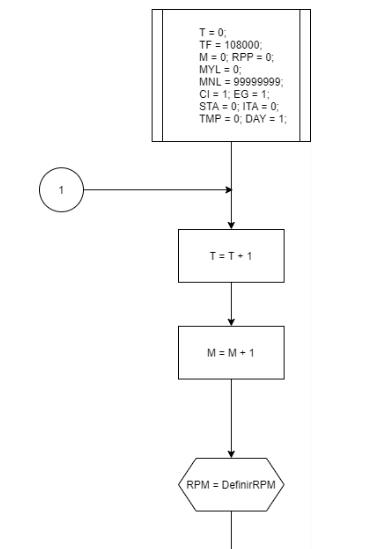
*Inversa de F(x):*

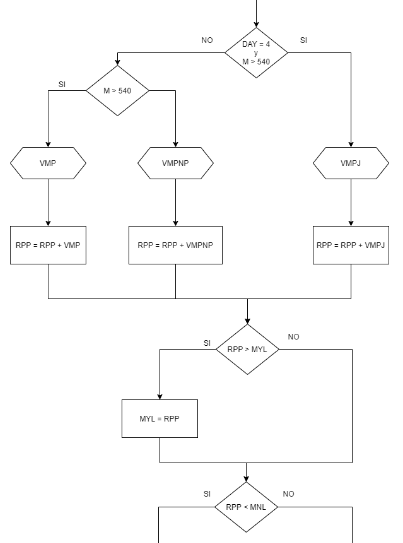
**

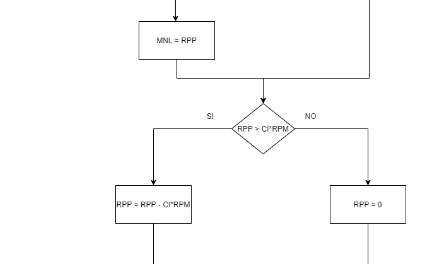
#### **Diagrama de la rutina:**

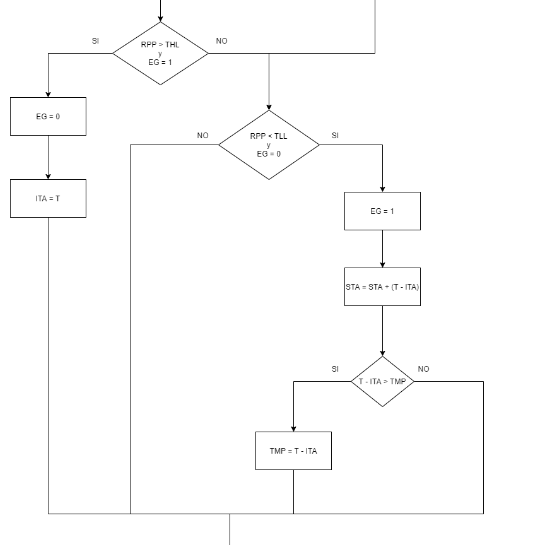


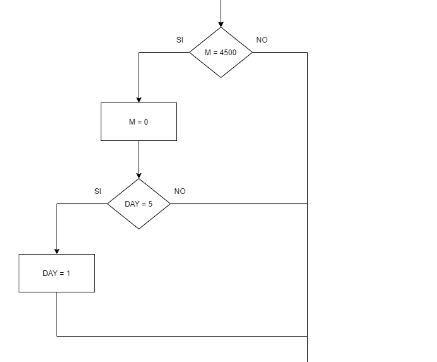
# Diagrama











# 

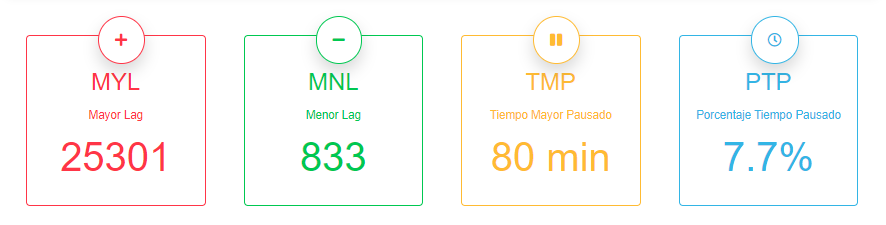
# 

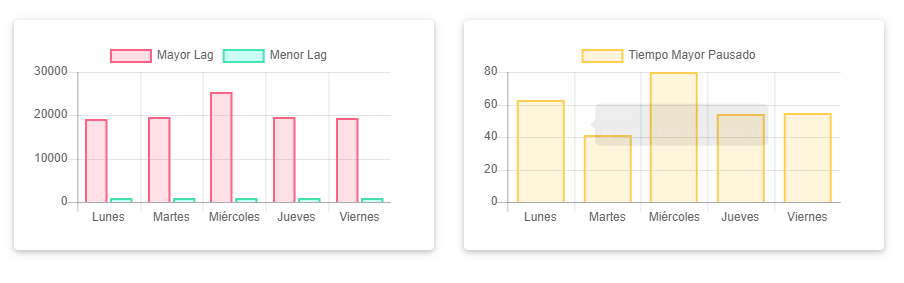
# Simulación

La simulación se realizó

Situación Actual.



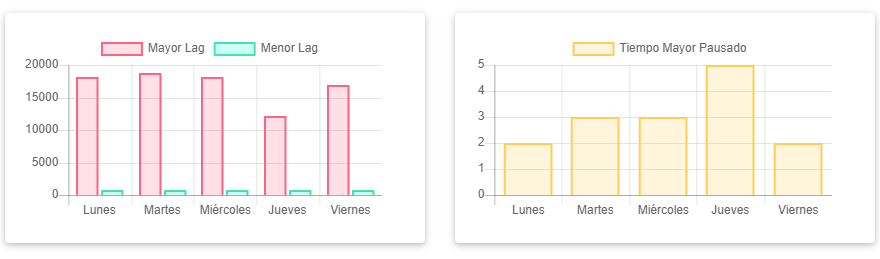
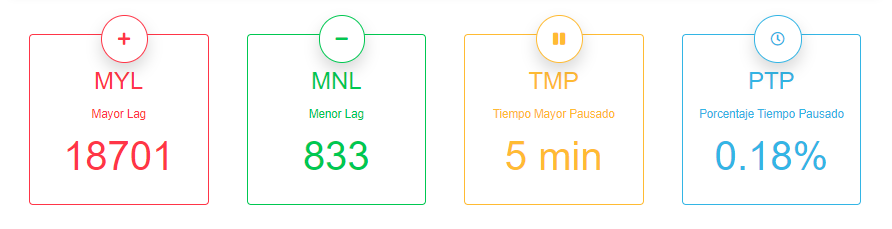




## Escenarios propuestos.

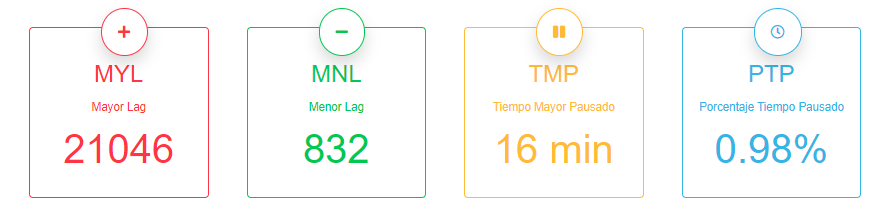
Ideal.

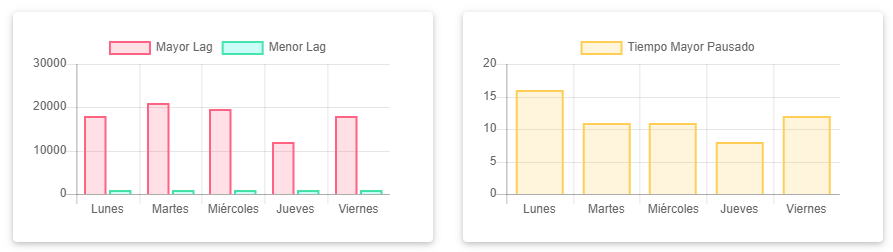




Mejor escenario.







# Conclusiones

Luego de haber realizado el análisis de los resultados, se llega a la conclusión de que si bien el problema se resuelve generando más instancias de Google Cloud en la nube y bajando los umbrales a tiempos muy bajo para garantizar consistencia en los reportes, mantener esta infraestructura en cuanto a costos, y monitoreo de los mismos implica un gasto que desde la gerencia de reporting no están dispuestos a costear, ya que la reportería en sí no es algo que generen ingresos a la Fintech.

La solución óptima encontrada, es con un aumento en una instancia, y dejando los umbrales en un número muy cercano, ya que esto nos permite apagar cuando sea necesario, y un número alto de encendido, que nos va a permitir encender rápidamente y tener una consistencia óptima en los reportes generados.