Universidad del Valle

# Trabajo Final

**David Stiven Rojas Mamian**
**Andrés Felipe Palomino Montezuma**
**Docente: Cesar Ojeda**
**Curso: Series de tiempo y Pronóstico**

Universidad del Valle
Escuela de Estadística

# Introducción

## Stack Overflow

Sitio de preguntas y respuestas para programadores profesionales y aficionados, se abordan una amplia gamma de temas de programación, como lo son las librerías que utilizan distintos software de programación.

## Base de datos

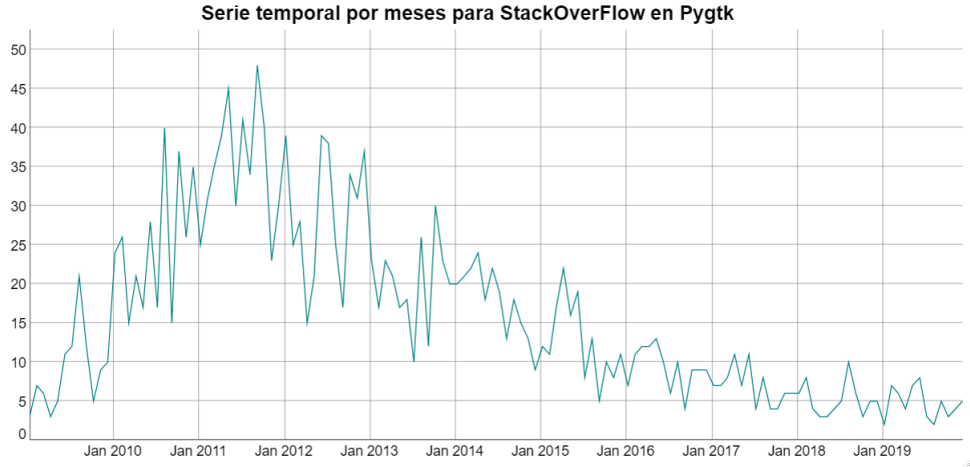
La base de datos cuenta con 132 registros que indican el conteo mensual de preguntas que se realizan en la web sobre la librería PyGTK, desde el año 2009 hasta el año 2019

# Problema de investigación y Objetivos

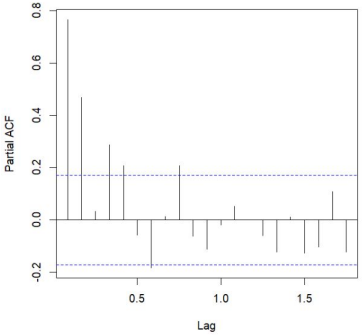
Una compañía de programación de juegos está interesada en realizar capacitaciones sobre programas que ayuden en la creación e interfaces de juegos. Específicamente se hará un proceso de seguimiento sobre la librería PyGTK y evaluar la capacidad de inversión que pueda realizarse.

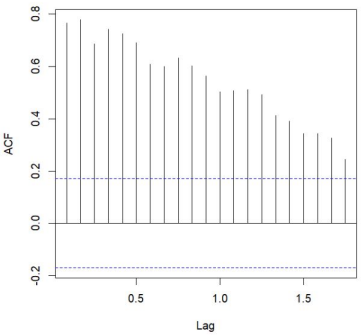
- Identificar un modelo adecuado para la serie de tiempo.
- Pronosticar el número de preguntas que se realizaran sobre la librería PyGTK para el mes enero y febrero del año 2020

# Modelamiento

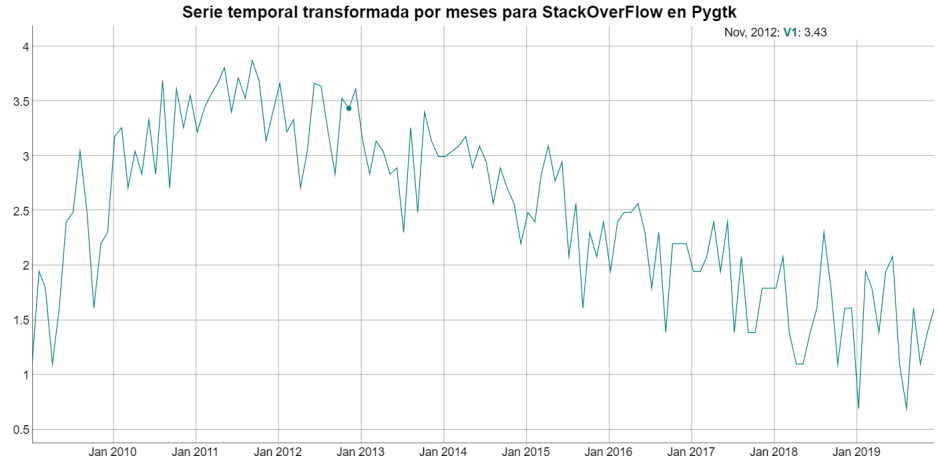


# Modelamiento

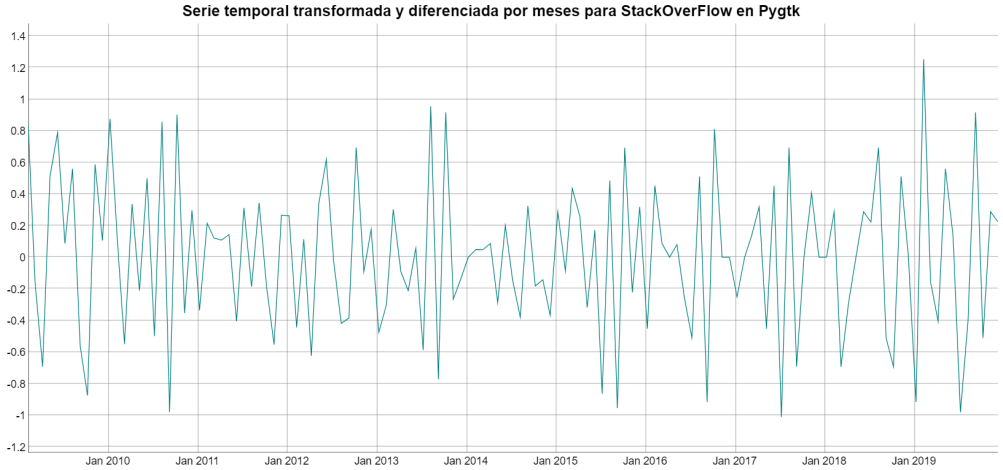


# Estabilización en varianza



# Primera diferencia

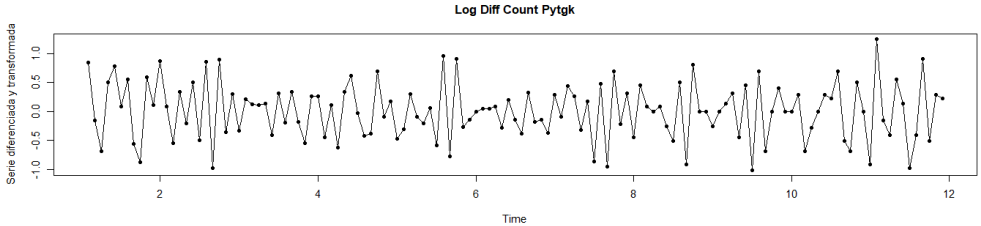


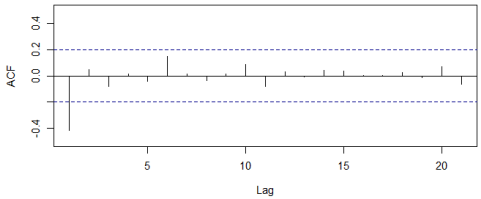
# Prueba Dickey-Fuller

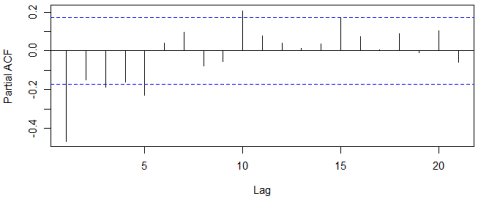
```
        Augmented Dickey-Fuller Test

data:  ts(diff(log(X$Pygtk)), freq = 12, start = c(1, 2009))
Dickey-Fuller = -7.5096, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary
```

# ACF y PACF muestral



# Modelo propuesto

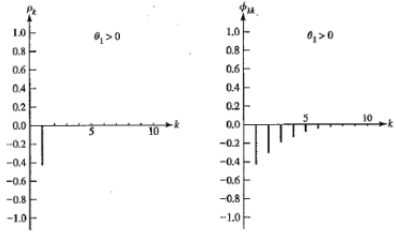



Figura: Modelo Teorico

**ARIMA (0,1,1) O IMA(1,1)**

$$(1 - B)Z_t = (1 - \theta B)a_t$$

Donde $Z_t$ es el $log(Y_t)$

# Estimación y significancia

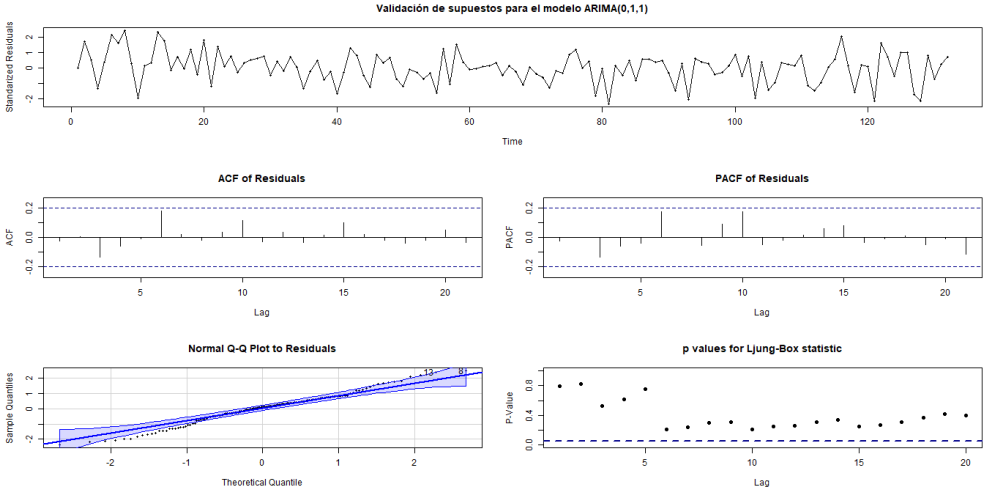
$$(1 - B)Z_t = (1 - 0.66523B)a_t$$

con $a_t \sim$N(0 , 0.168)

```
z test of coefficients:

     Estimate Std. Error z value  Pr(>|z|)
ma1 -0.665299   0.064592   -10.3 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

# Diagnóstico del modelo

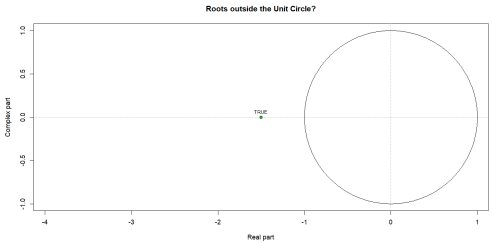


# Diagnóstico del modelo

Test de varianza constante

```
        White Neural Network Test

data:  modelofinal$residuals
X-squared = 0.36956, df = 2, p-value = 0.8
```



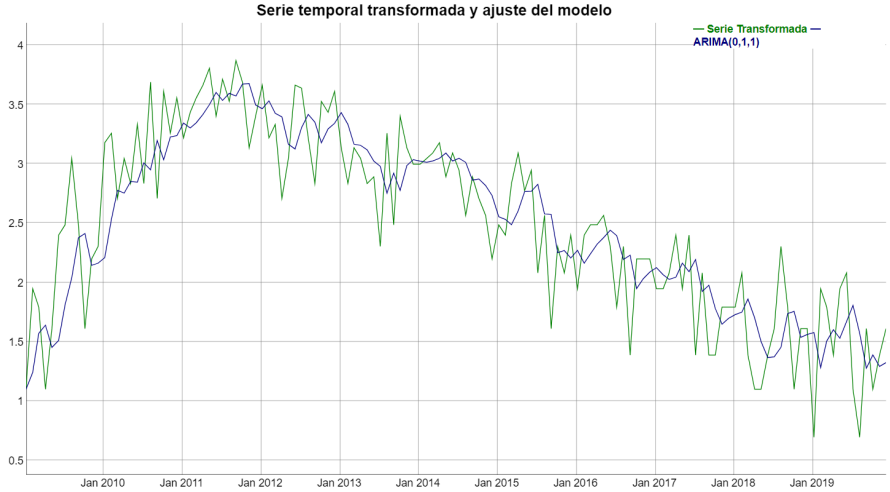

Test de normalidad

```
Title:
 Jarque - Bera Normalality Test

Test Results:
  STATISTIC:
    X-squared: 0.3747
  P VALUE:
    Asymptotic p Value: 0.8291
```

Figura: Raíz del polinomio ARIMA(0,1,1)

# Ajuste del modelo



# Pronósticos

Se tiene el modelo IMA(1,1):

$$(1 - B)Z_t = (1 - 0.66523B)a_t$$

Reemplazando t por $n + l$ y sacando el valor esperado condicional se tiene:

$$\hat{Z}_n(l) = \hat{Z}_n(l-1) + \hat{a}_n(l) - \theta_1 \hat{a}_n(l-1)$$

Por lo que la forma general del pronóstico para $l \geq 1$ es:

$$\hat{Z}_n(l) = Z_n - \theta_1 a_n$$

por último sé realiza $exp(\hat{Z}_n(l))$ para obtener él $\hat{Y}_n(l)$ ya que se realizó la transformación logarítmica para estabilizar la varianza.

Para obtener los límites de predicción, se debe calcular las ponderaciones $\psi_j$ de manera recursiva, donde $\psi_j = \sum_{i=0}^{j-1} \pi_{j-1}\psi_i$
Y los coeficientes $\pi_j$ se obtienen escribiendo el modelo en forma AR.

$$(1 - \theta_1 B) * (1 - \pi_1 B - \pi_2 B^2 - \pi_3 B^3 - .....) = (1 - B)$$

Solucionando se obtiene que:

$$\psi_j = (1 - \theta_1) \text{ para } 1 \leq j \leq l - 1$$

Finalmente, los pronósticos son:

| Tiempo | Lim. inferior | Pronostico | Lim. superior |
|---|---|---|---|
| Enero 2020 | 1.850277 | 4.131885 | 9.226983 |
| Febrero 2020 | 1.770973 | 4.131885 | 9.640164 |

Tabla : Predicciones y límites de predicción a un nivel de confianza del 95%

# Conclusiones

- Para el problema de investigación, se evidencia que para el mes enero y febrero del siguiente año, no se presentara un aumento en la cantidad de preguntas que se realizaran sobre la librería PyGTK en comparación a los meses anteriores. Por lo que se recomienda no realizar una inversión significativa en programas de capacitación.
- En términos de los pronósticos, se observa que la función de predicción es una línea horizontal, que depende de la información que se tenga sobre el último dato de la serie ($Z_n$) y del residuo dado por el ajuste y el valor real ($a_n$), además los límites de predicción van aumentando a medida que se pronostique mas hacia el futuro. Por ende, no es recomendable realizar predicciones a largo plazo.

# Bibliografía I


**Jenkins, G. M., Box, G. E. P.**
'Time Series Analysis Forecasting and Control (Vol. 2).'.
*Holden-Day, San Francisco,* 1976.

**Cryer, J. D., Chan, K.-S.**
*" Time series analysis: with applications in R (Vol. 2)."*
*Springer,* 2008.