

Time Series Models & Forecasting

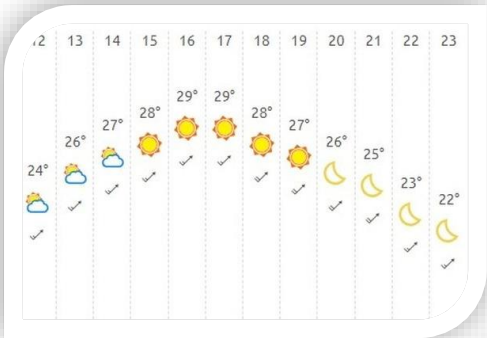
Alberto Valera Pérez



Forecasting: *Introducción*

“Proceso de hacer declaraciones sobre futuros eventos, datos o señales con el mayor acierto posible.”

Aplicaciones



Tiempo

Finanzas

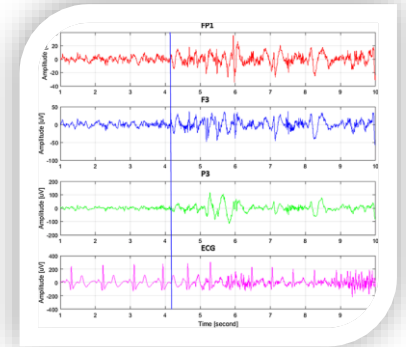


Stock

Afluencia



Medicina



Forecasting: *Proyecto*

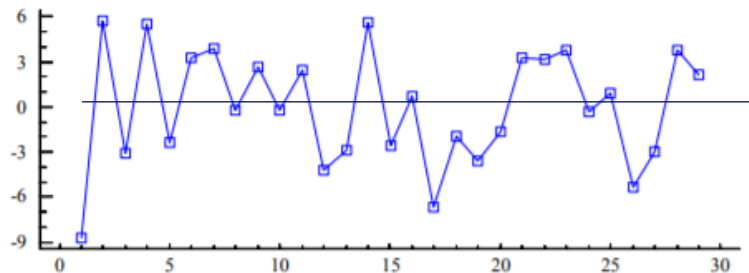


Exploratorio y Preprocesado: *Series Temporales*

“Métodos de descomposición de series temporales: se parte de la idea de que la serie temporal se puede descomponer en todos o algunos de los siguientes componentes”

- **Tendencia (T_t)** : evolución de la serie a largo plazo
- **Estacionalidad (S_t)** : periodicidad que presentan las series temporales, movimientos regulares y repetitivos
- **Cíclica (C_t)** : cambios de tendencia de la serie temporal o fluctuaciones no periódicas
- **Residual (e_t)** : movimientos irregulares, aleatorios, efecto de hechos no previsibles

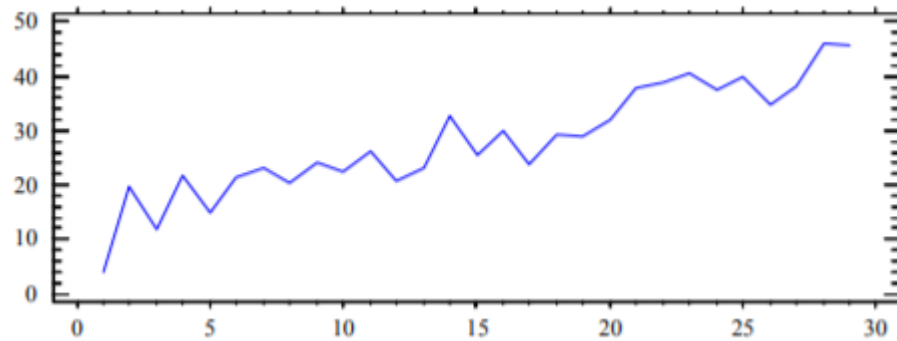
Serie Estacionaria – Estadísticos constantes a lo largo del tiempo



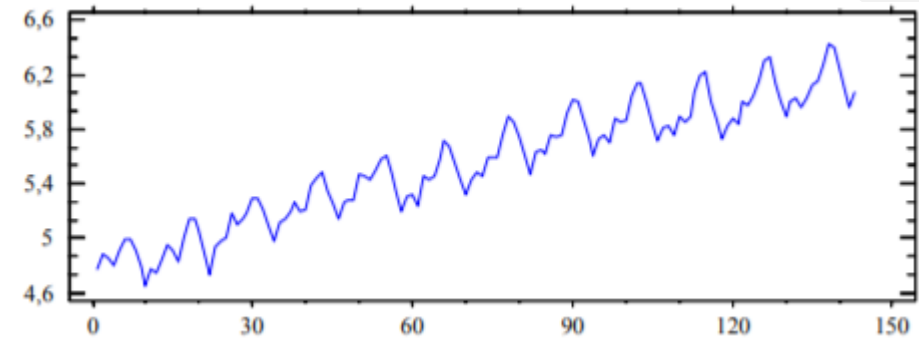
Obtención de estadísticos y gráficos
Augmented Dickey Fully Test
Auto Correlation Functions (ACF)

Exploratorio y Preprocesado: *Series Temporales*

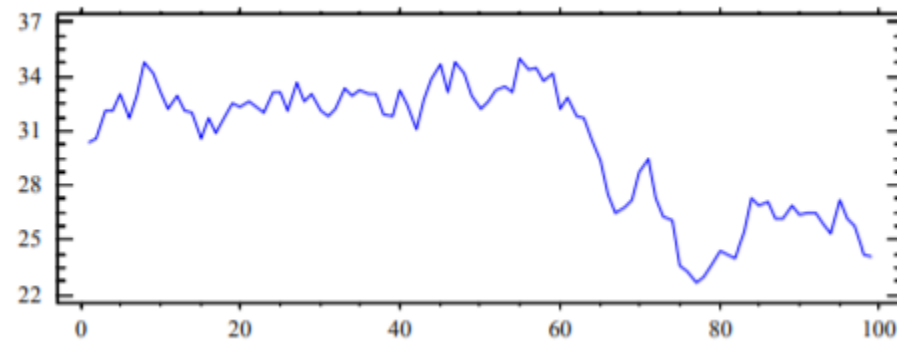
Tendencia (T_t)



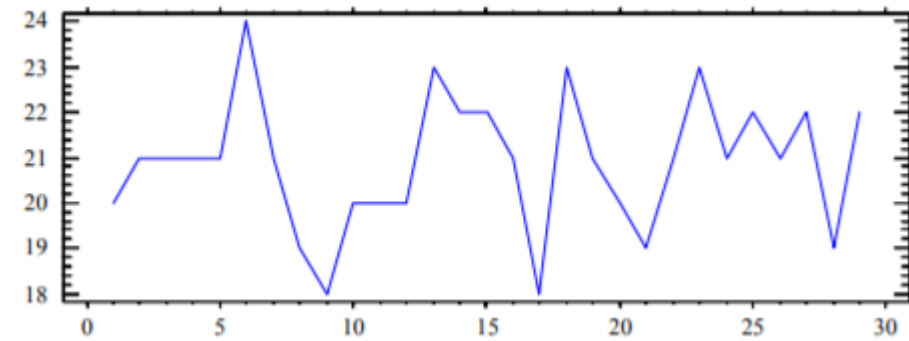
Estacionalidad (S_t)



Cíclica (C_t)



Residual (e_t)



Exploratorio y Preprocesado: *Series Temporales*

Aditiva

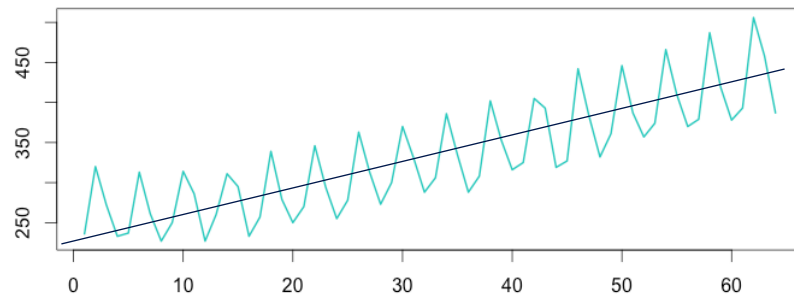
$$Y_t = T_t + S_t + C_t + e_t$$

$$Y_t = T_t S_t C_t + e_t$$

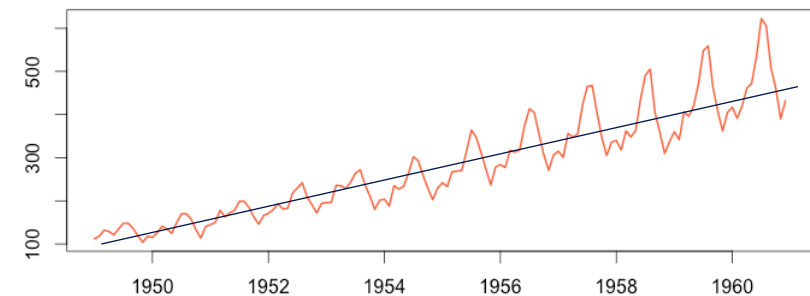
Multiplicativa

$$Y_t = T_t S_t C_t e_t$$

```
install.packages("fpp")  
library(fpp)  
data(ausbeer)  
timeserie_beer = tail(head(ausbeer, 17*4+2), 17*4-4)  
plot(as.ts(timeserie_beer))
```



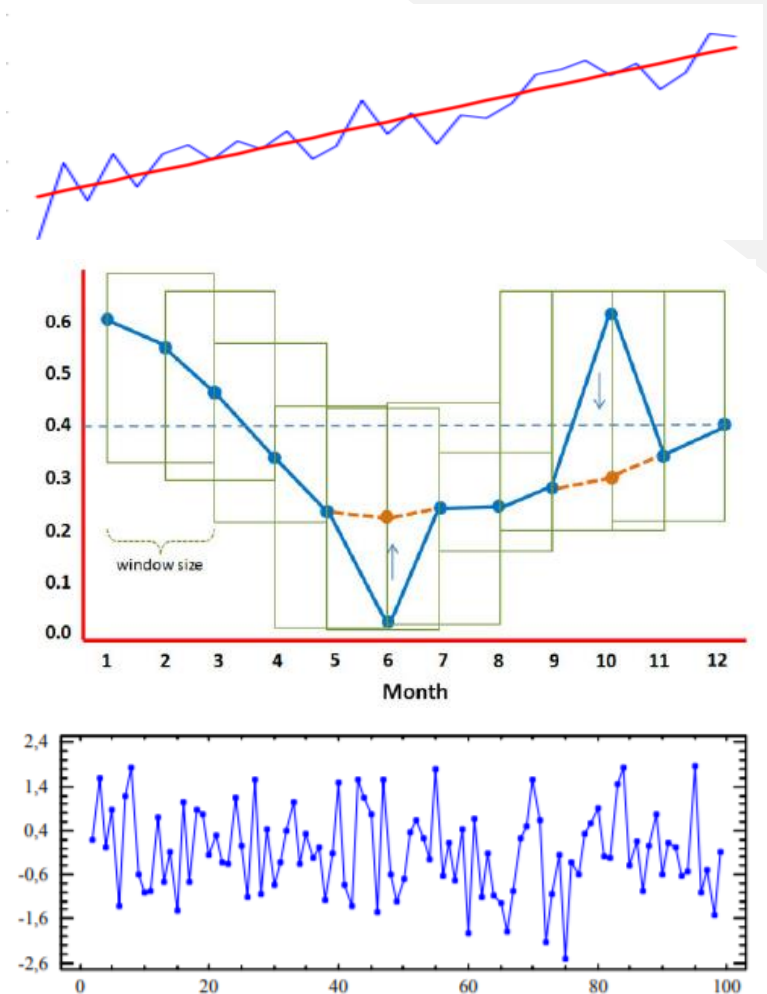
```
install.packages("Ecdat")  
library(Ecdat)  
data(AirPassengers)  
timeserie_air = AirPassengers  
plot(as.ts(timeserie_air))
```



Exploratorio y Preprocesado: *Series Temporales*

Detección de tendencia

- *Ajuste polinómico* $T_t = a_0 + a_1t + a_2t^2$
- *Filtros de Media móvil* $\hat{T}_t = \frac{1}{2q+1} \sum_{j=-q}^q X_{t-j}, q+1 \leq t \leq n-q$
- *Diferencia de la serie* $Y_t = X_t - X_{t-1}$



Exploratorio y Preprocesado: *Series Temporales*

Aditiva

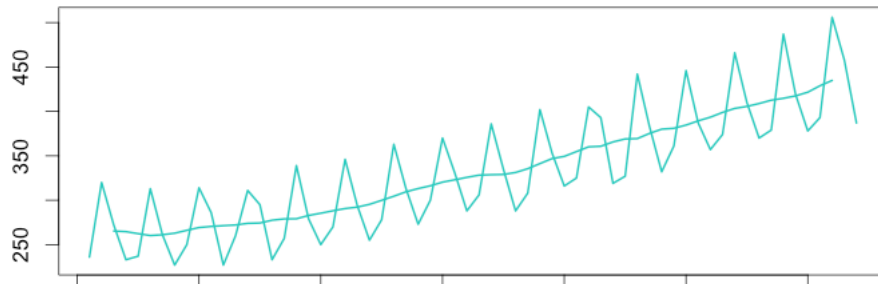
$$Y_t = T_t + S_t + C_t + e_t$$

R

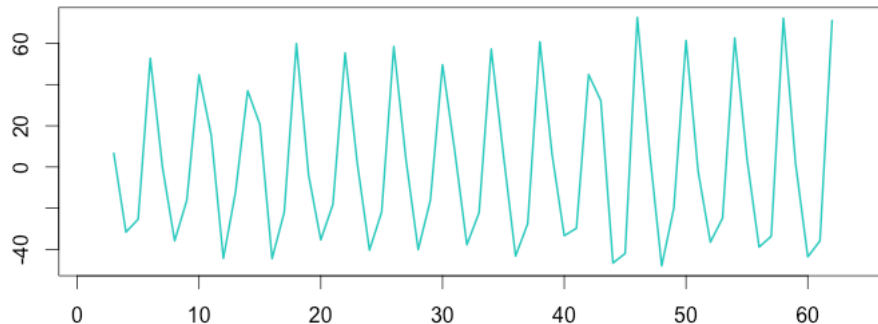
- Paquete "forecast"
- Función `ma()`

Multiplicativa

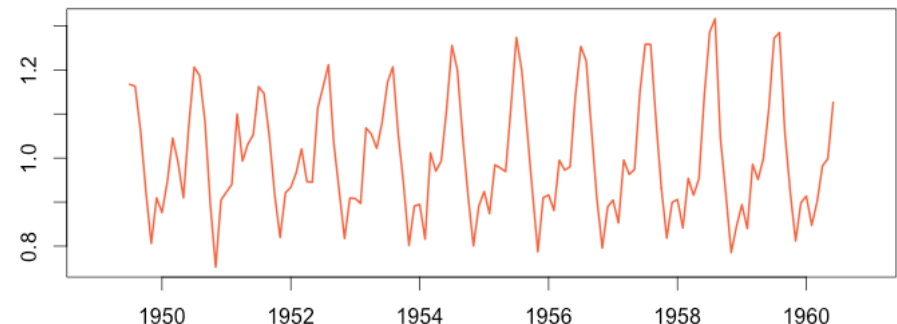
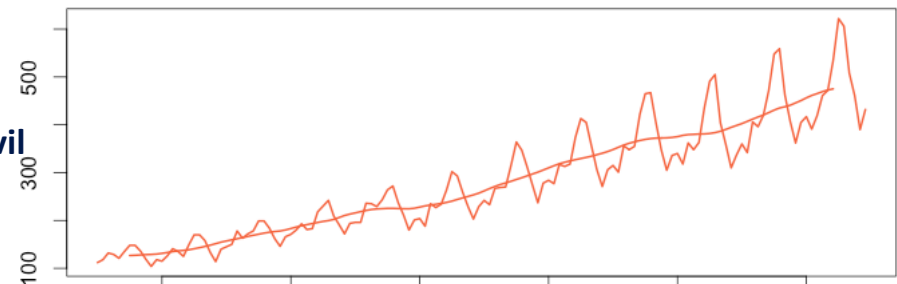
$$Y_t = T_t S_t C_t e_t$$



Serie con media móvil



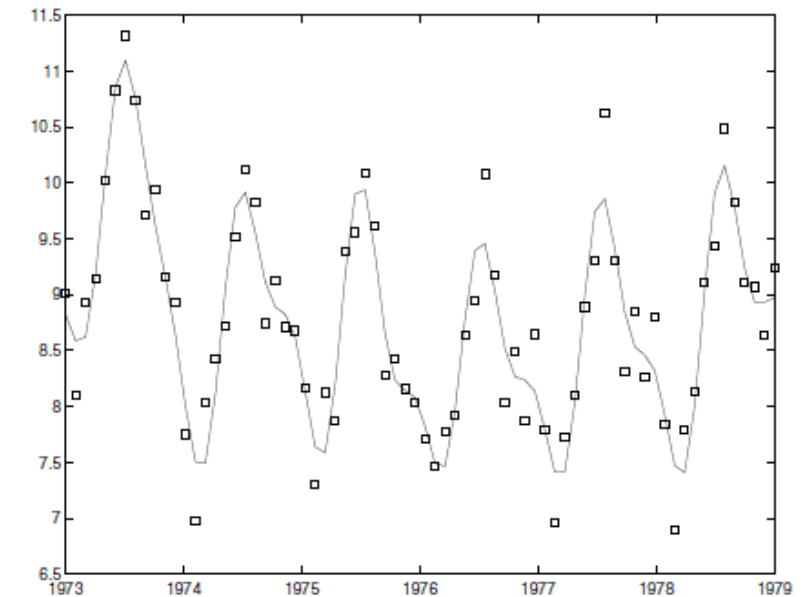
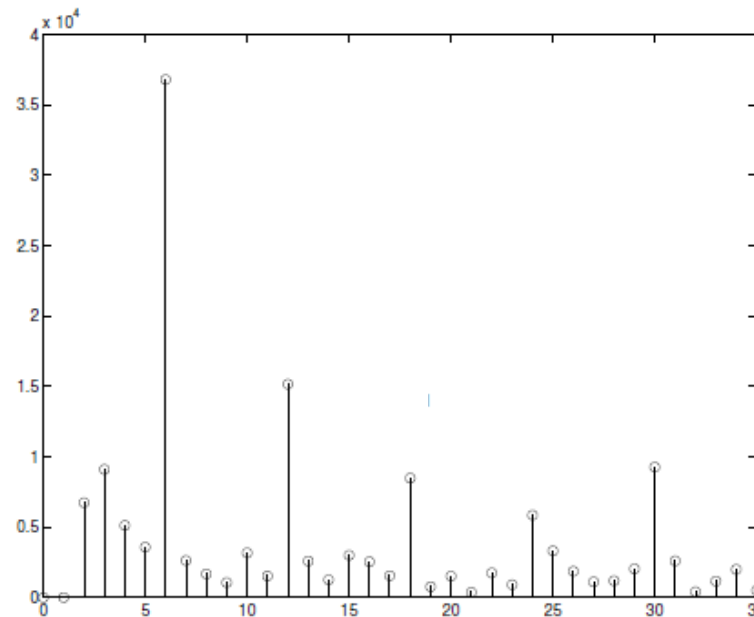
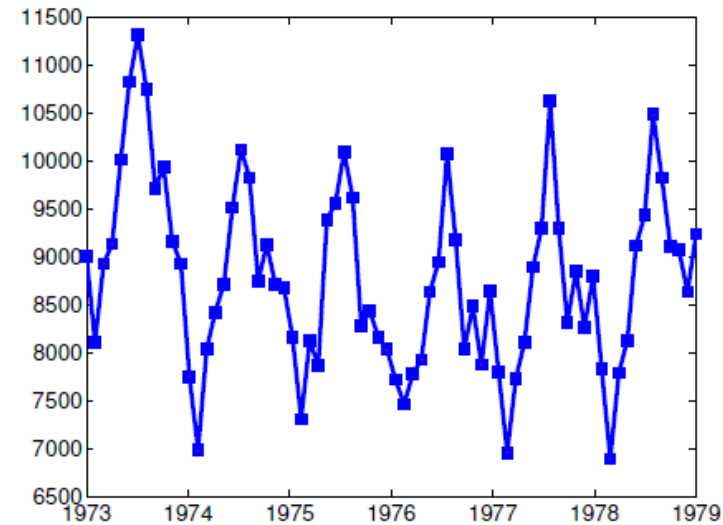
Serie sin tendencia



Exploratorio y Preprocesado: *Series Temporales*

Detección de Estacionalidad

- *Componentes de Fourier y detección de armónicos que componen la componente estacional*
- *Diferencial de orden periódico (mes vs mes año anterior)*



Exploratorio y Preprocesado: *Series Temporales*

Nulos y outliers

Nulos



Interpolación

Tipos de Outliers

Duración

Temporal

Permanente

Forma

Brusco

Outlier aditivo

Outlier de cambio de nivel

Gradual

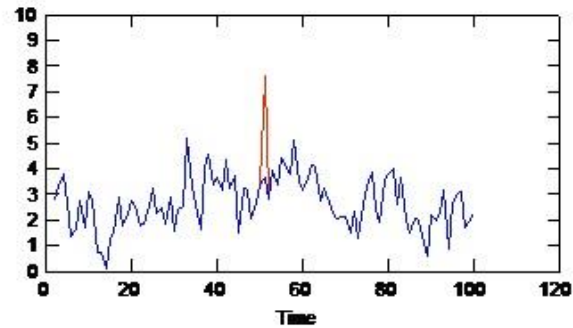
Outlier cambio temporal

Outlier innovacional

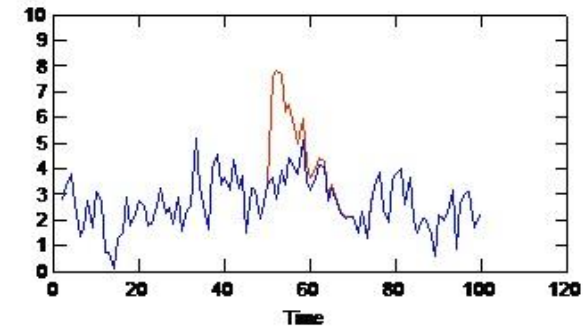
Exploratorio y Preprocesado: *Series Temporales*

Nulos y outliers

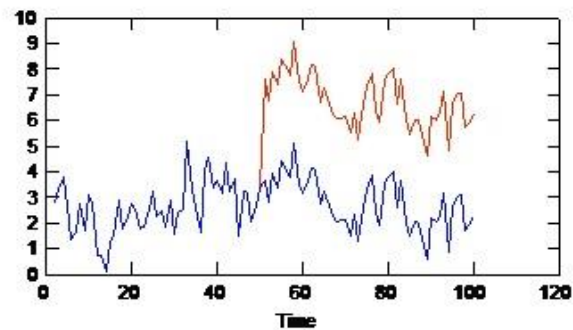
Additive Outlier



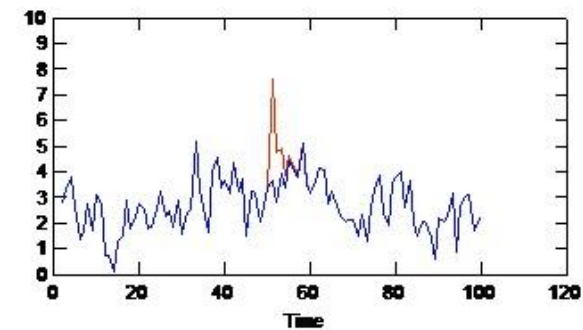
Innovational Outlier



Level Shift Outlier



Transient Change Outlier



Exploratorio y Preprocesado: *Series Temporales*

Tendencia, estacionalidad y Outliers

- R:
 - STL: <https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/stl>
 - Decompose: <https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/decompose>
 - tsclean: <https://www.rdocumentation.org/packages/forecast/versions/8.11/topics/tsclean>
 - tsoutliers: <https://www.rdocumentation.org/packages/tsoutliers/versions/0.6-8>
- Python:
 - STL: https://www.statsmodels.org/dev/examples/notebooks/generated/stl_decomposition.html

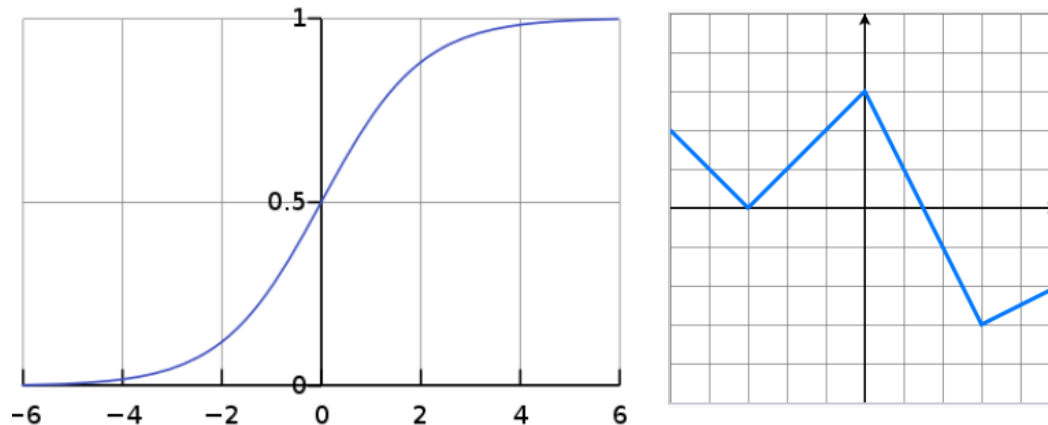
Exploratorio y Preprocesado: *Prophet*

Herramienta desarrollada por Facebook y disponible desde 2017

Usa un modelo de regresión aditiva avanzado

$$Y_t = T_t + S_t + C_t + e_t$$

- **Tendencia (T_t)** : modelo de crecimiento saturado no lineal (logístico) / modelo lineal por partes (piecewise)
- **Estacionalidad (S_t)** : Fourier (orden 10). Orden ajustable por año/mes/semana... . Eliges estacionalidad. Anual y semanal.
- **Fluctuación cíclica (C_t)** : vacaciones (opción de ventana) / calendario
- **Residual (e_t)**



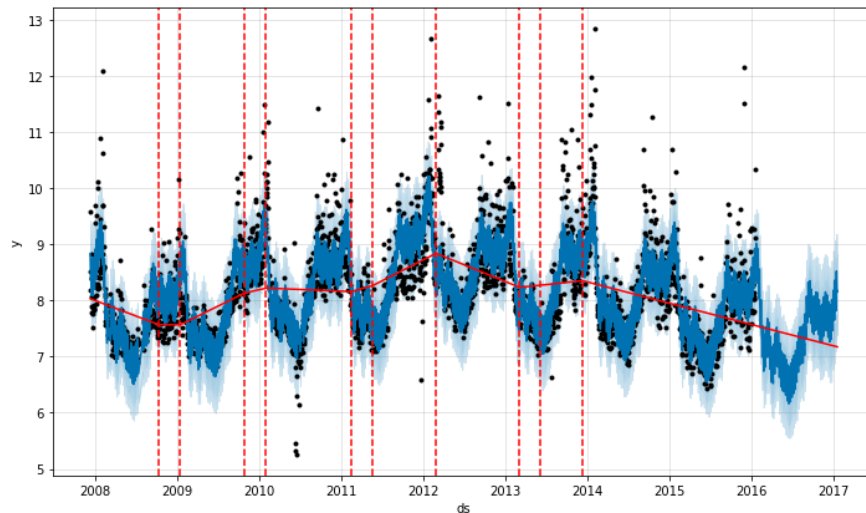
Exploratorio y Preprocesado: *Prophet*

Herramienta desarrollada por Facebook y disponible desde 2017

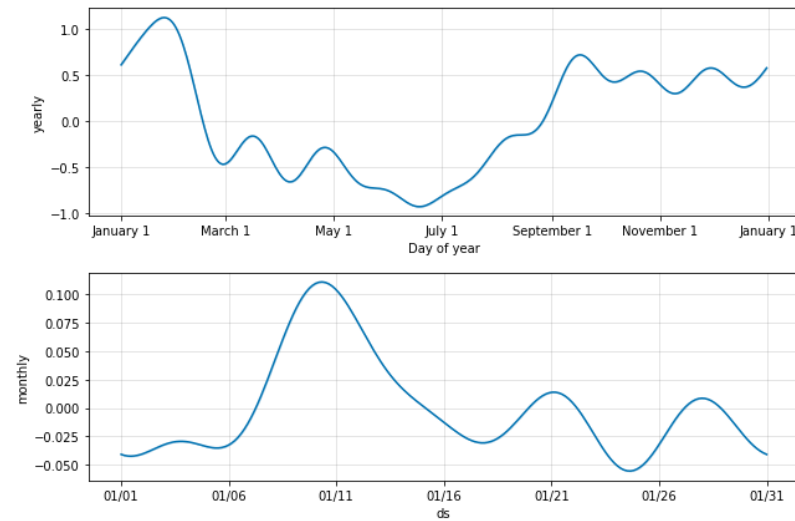
Usa un modelo de regresión aditiva avanzado

$$Y_t = T_t + S_t + C_t + e_t$$

- *Tendencia: modelo lineal por partes (piecewise)*



- *Estacionalidad: Anual, cuatrimestral, mensual, semanal, diaria, horaria*

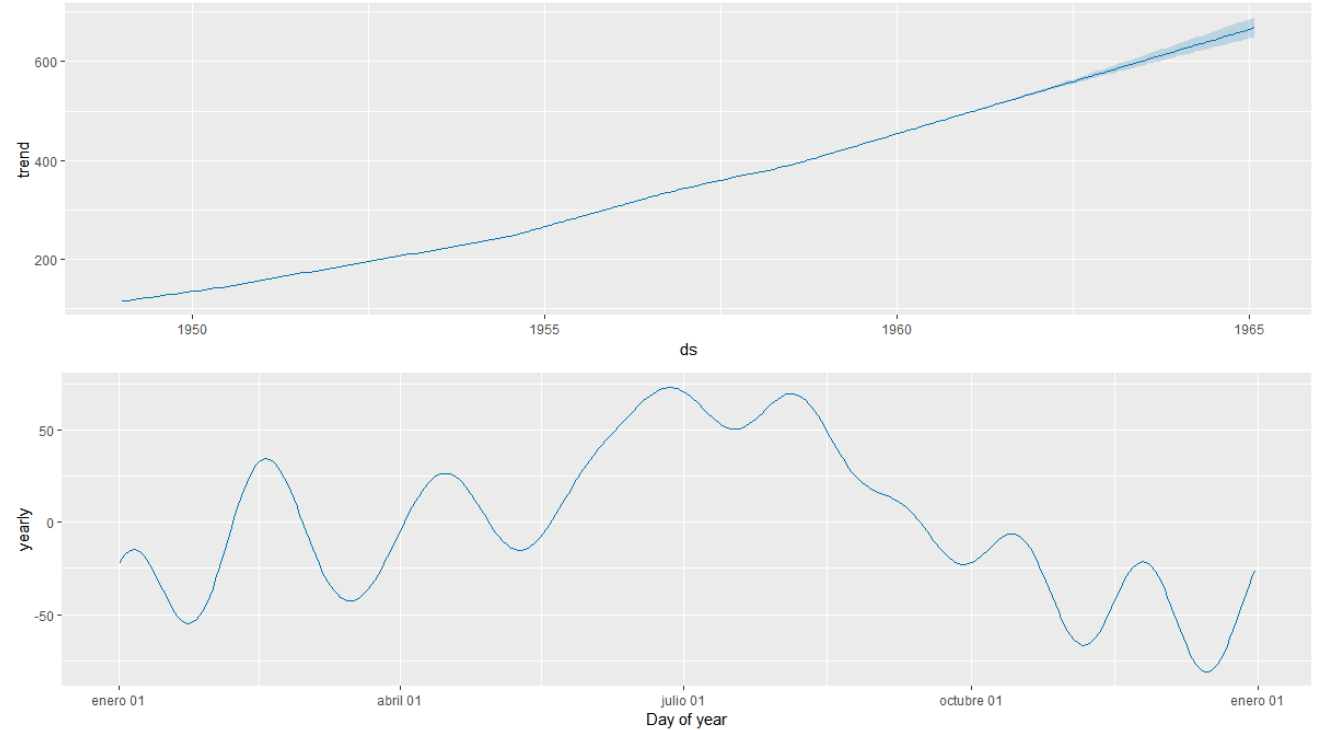
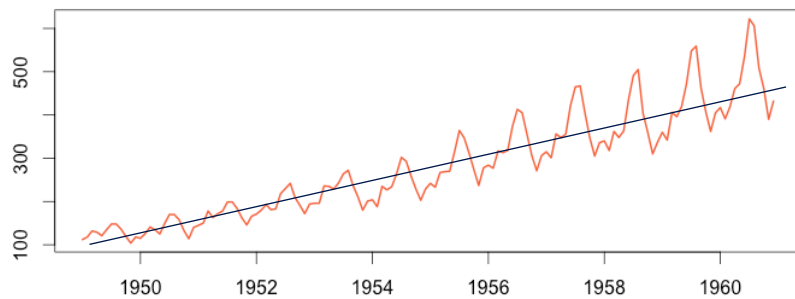


Exploratorio y Preprocesado: *Prophet*

Multiplicativa

$$Y_t = T_t S_t C_t e_t$$

```
install.packages("Ecdat")  
library(Ecdat)  
data(AirPassengers)  
timeserie_air = AirPassengers  
plot(as.ts(timeserie_air))
```

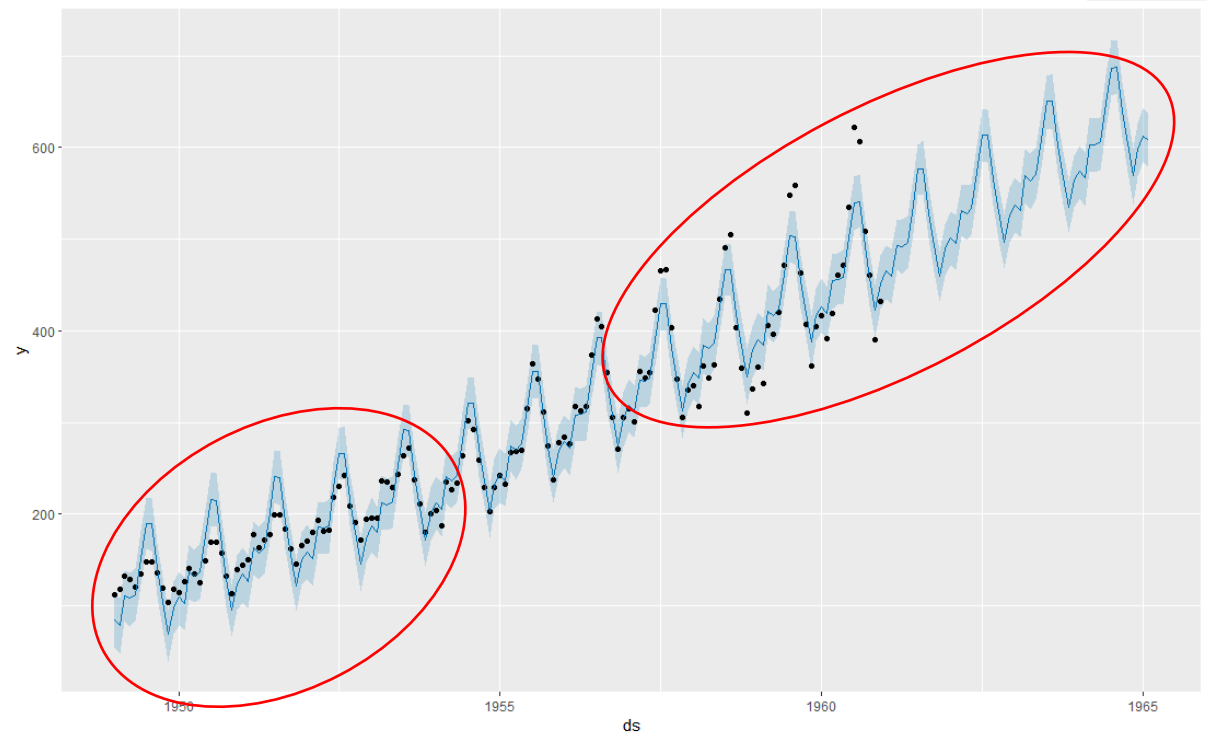
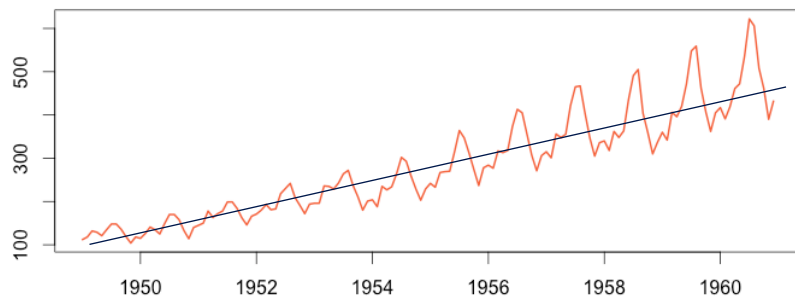


Exploratorio y Preprocesado: *Prophet*

Multiplicativa

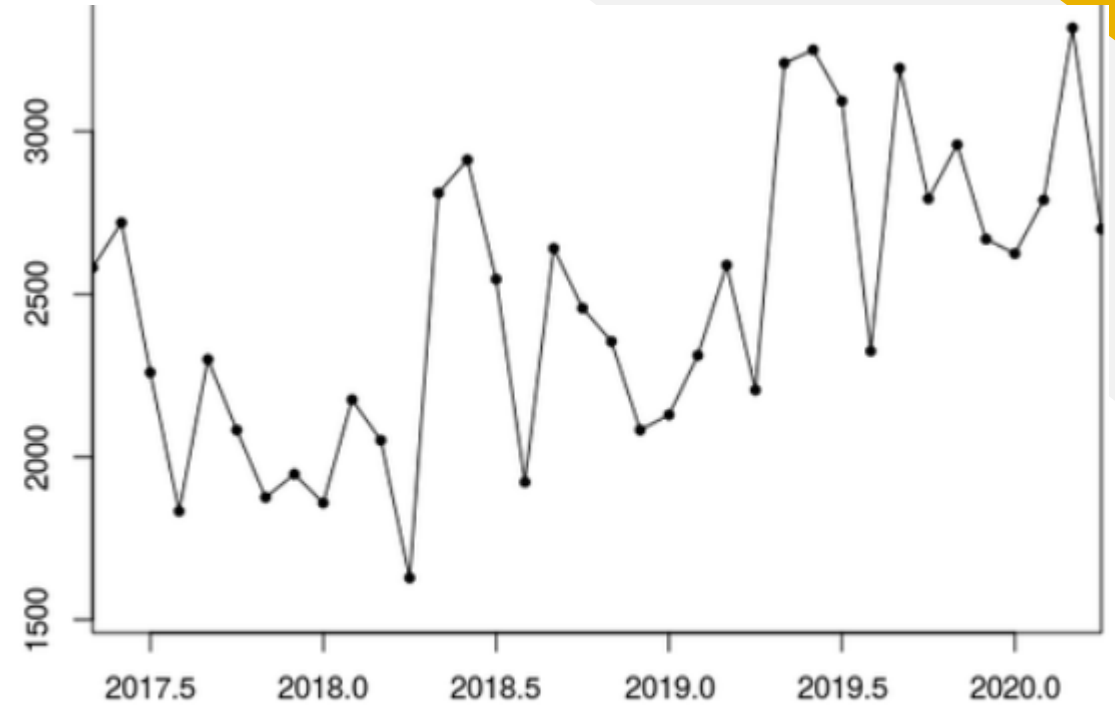
$$Y_t = T_t S_t C_t e_t$$

```
install.packages("Ecdat")  
library(Ecdat)  
data(AirPassengers)  
timeserie_air = AirPassengers  
plot(as.ts(timeserie_air))
```



Exploratorio y Preprocesado: *Kahoot*

Kahoot!



Series aditivas o multiplicativas

Reto – ¿Somos mejores que la máquina?

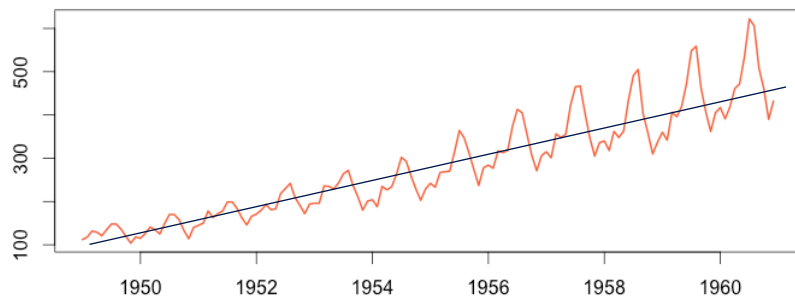
<https://kourentzes.com/forecasting/2014/11/09/additive-and-multiplicative-seasonality/>

Exploratorio y Preprocesado: *Prophet*

Multiplicativa

$$Y_t = T_t S_t C_t e_t$$

```
install.packages("Ecdat")  
library(Ecdat)  
data(AirPassengers)  
timeserie_air = AirPassengers  
plot(as.ts(timeserie_air))
```



- *Transformación logarítmica*

$$Y_t = T_t S_t C_t e_t$$

$$\log(Y_t) = \log(T_t S_t C_t e_t)$$

$$\log(Y_t) = \log(T_t) + \log(S_t) + \log(C_t) + \log(e_t)$$

- *Método Box-Cox – estabilización de varianza*

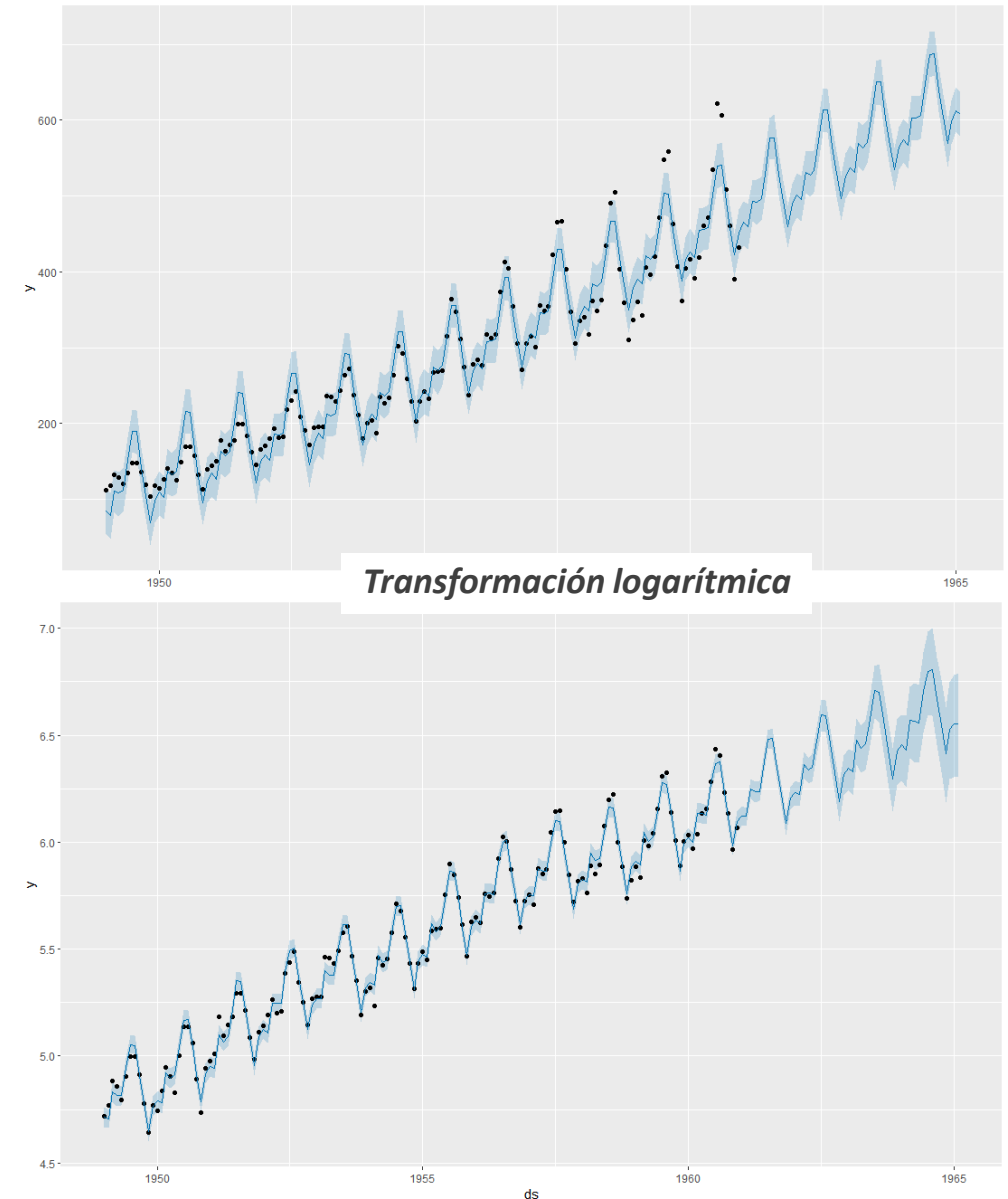
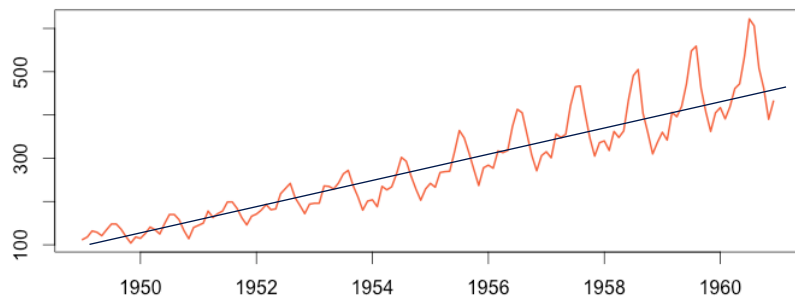
<http://onlinestatbook.com/2/transformations/box-cox.html>

Exploratorio y Preprocesado: *Prophet*

Multiplicativa

$$Y_t = T_t S_t C_t e_t$$

```
install.packages("Ecdat")  
library(Ecdat)  
data(AirPassengers)  
timeserie_air = AirPassengers  
plot(as.ts(timeserie_air))
```

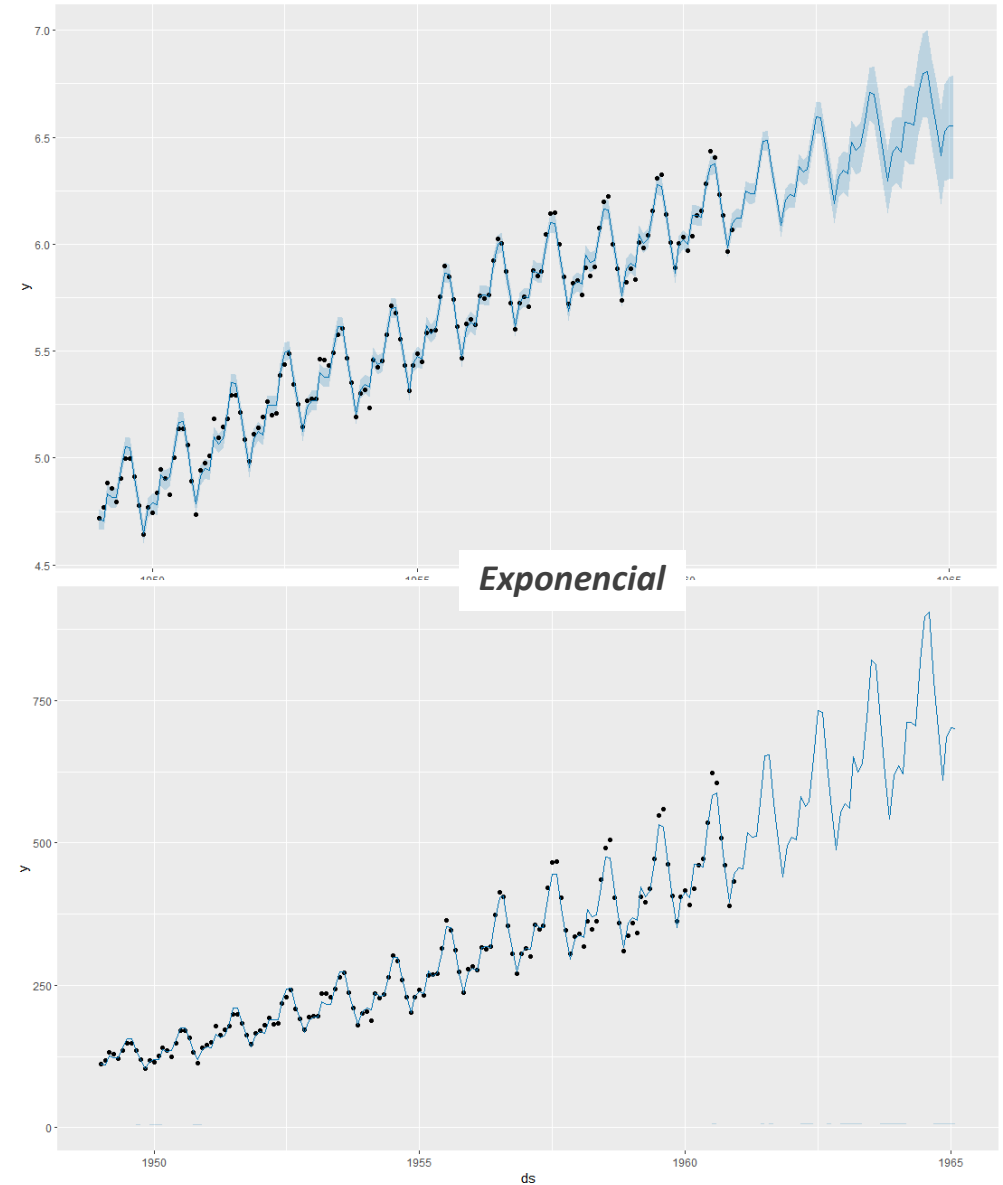
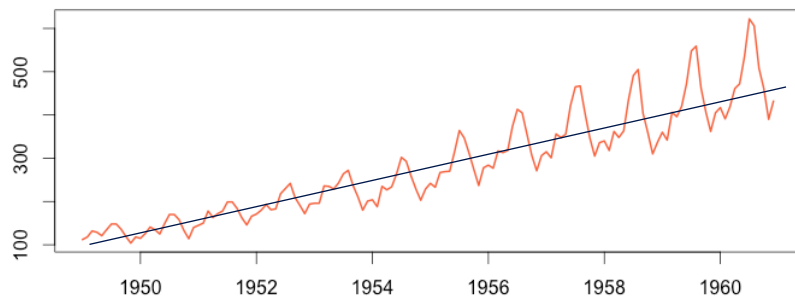


Exploratorio y Preprocesado: *Prophet*

Multiplicativa

$$Y_t = T_t S_t C_t e_t$$

```
install.packages("Ecdat")  
library(Ecdat)  
data(AirPassengers)  
timeserie_air = AirPassengers  
plot(as.ts(timeserie_air))
```

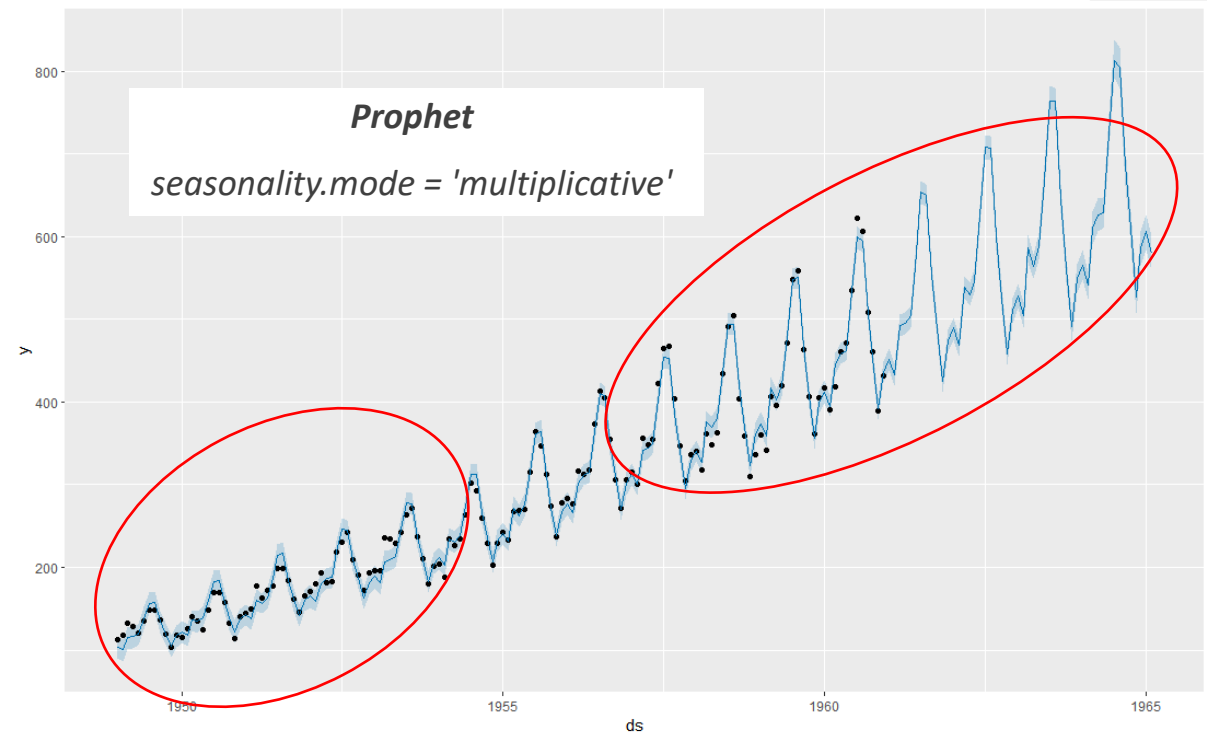
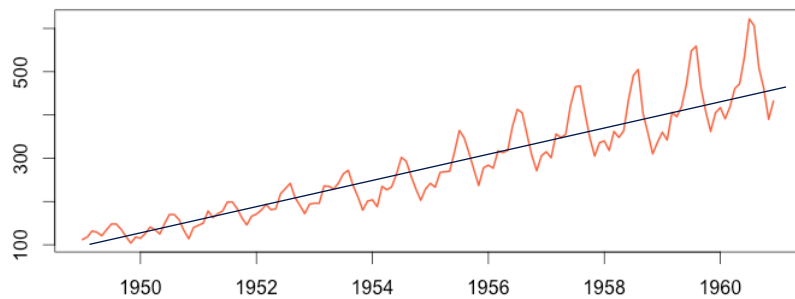


Exploratorio y Preprocesado: *Prophet*

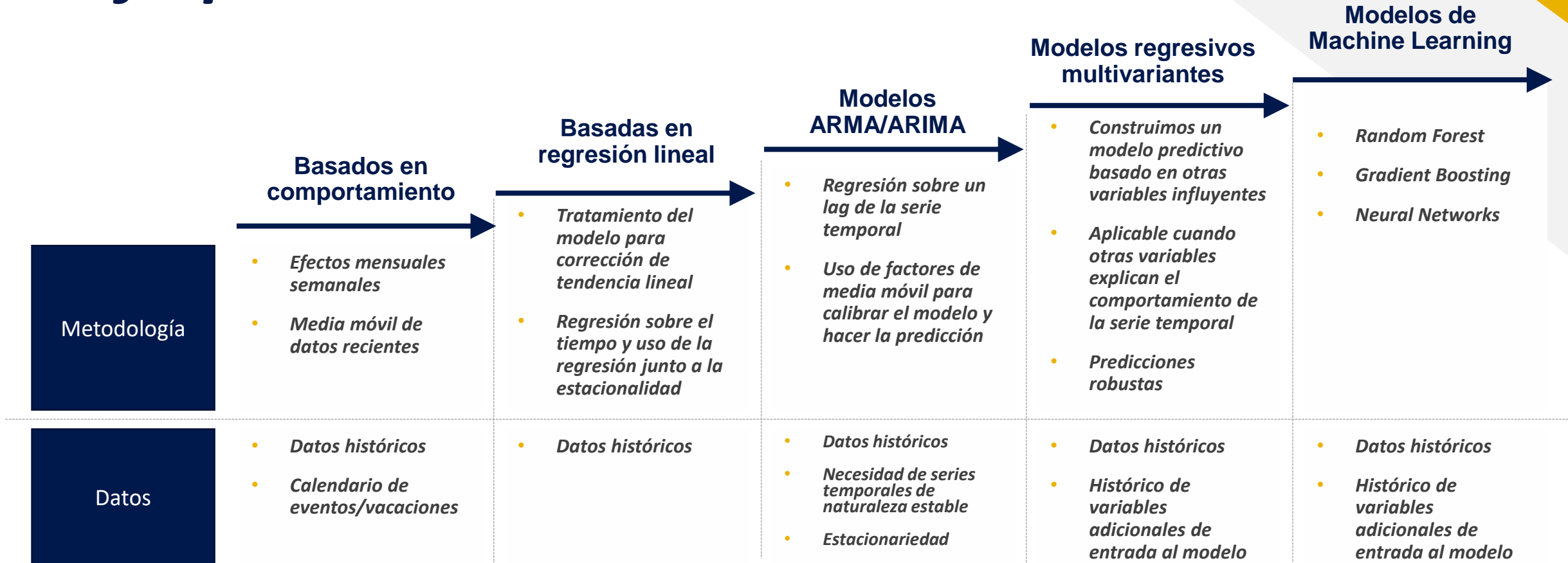
Multiplicativa

$$Y_t = T_t S_t C_t e_t$$

```
install.packages("Ecdat")  
library(Ecdat)  
data(AirPassengers)  
timeserie_air = AirPassengers  
plot(as.ts(timeserie_air))
```



Modelos predictivos: *Enfoque Incremental*



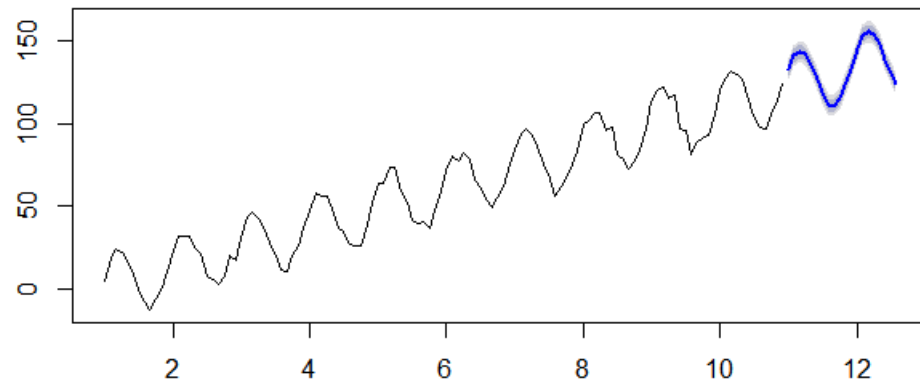
Modelos predictivos: *Enfoque Incremental*

Modelos de

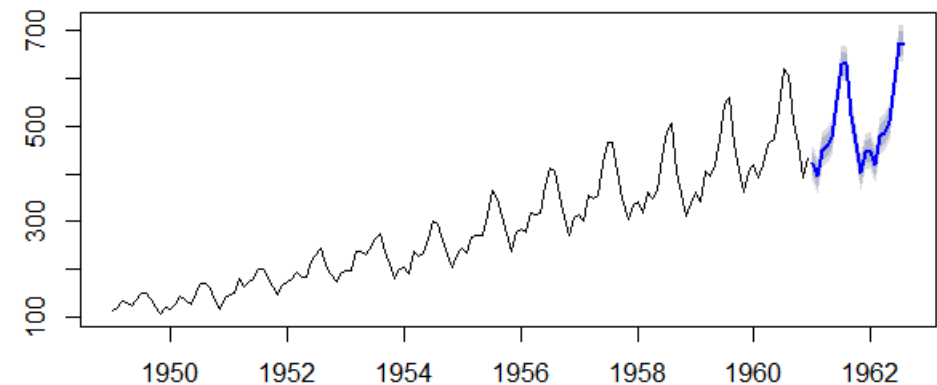
R: tslm()

```
library(forecast)
y <- ts(rnorm(120,0,3) + 1:120 + 20*sin(2*pi*(1:120)/12), frequency=12)
fit <- tslm(y ~ trend + season)
fit <- tslm(y ~ trend * season)
plot(forecast(fit, h=20))
```

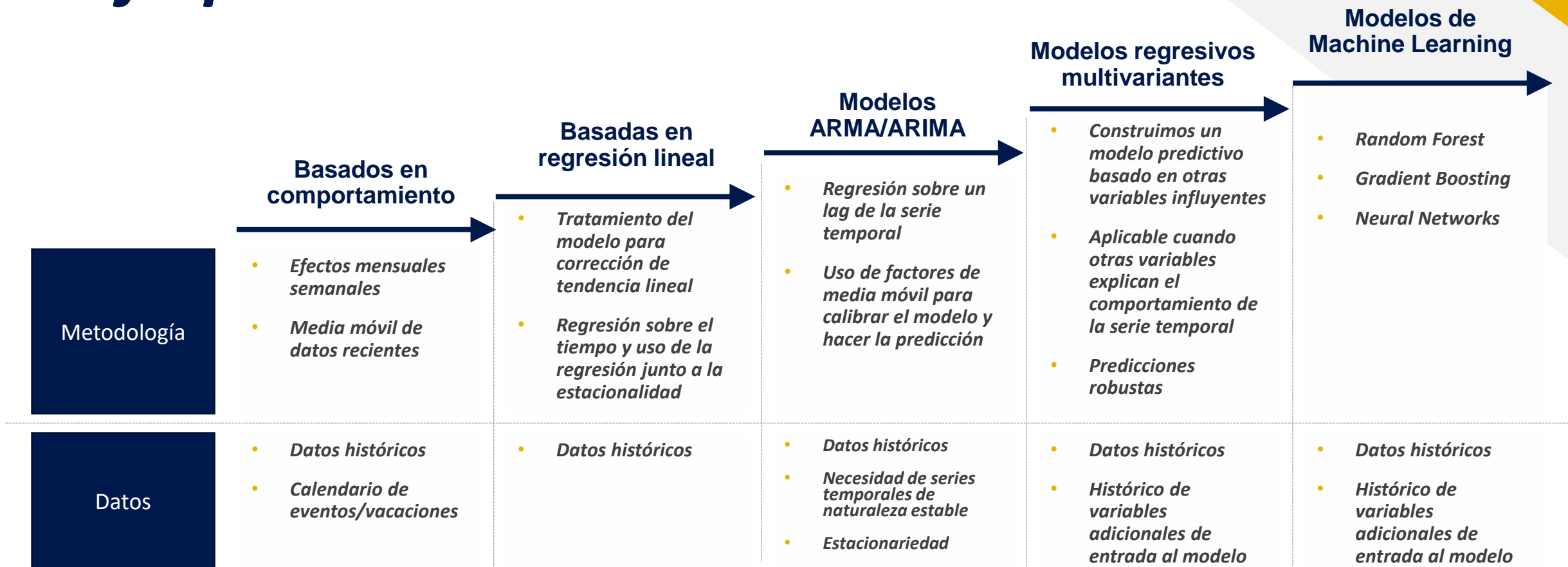
Forecasts from Linear regression model



Forecasts from Linear regression model



Modelos predictivos: *Enfoque Incremental*



Evaluación y Resultados:

Métricas de Evaluación

$$e_i = |y_i - \hat{y}_i|$$

Descripción

Ventajas

Inconvenientes

KPI

MAE (Mean Absolute Error)

$$MAE = \frac{\sum_{i=1}^n |e_i|}{n}$$

- Entendible (términos absolutos)
- Considera valores 0

- Error no comparable con la serie temporal (un error absoluto medio de 4 es mucho o poco?)

MAPE (Mean Absolute Percentage Error)

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{e_i}{y_i} \right|$$

- Entendible (porcentual)
- Medida Relativa
- No sesgada en valores extremos

- Valores 0 en la serie temporal
- No útil cuando la escala de la serie es pequeña (error porcentual muy grande)

SMAPE (symmetric MAPE)

$$SMAPE = \frac{1}{n} \sum_{i=1}^n \frac{|e_i|}{(|y_i| + |\hat{y}_i|)/2}$$

- Medida Relativa
- No sesgada en valores extremos
- Considera valores 0

- Valores 0 en la predicción o en la serie temporal (+200%)
- Un mismo error puede dar porcentajes distintos dependiendo de si el error es positivo o negativo

Evaluación y Resultados:

Métricas de Evaluación

$$e_i = |y_i - \hat{y}_i|$$

KPI

MAE (Mean Absolute Error)

Descripción

$$MAE = \frac{\sum_{i=1}^n |e_i|}{n}$$

Ventajas

- Entendible (términos absolutos)
- Considera valores 0

Inconvenientes

- Error no comparable con la serie temporal (un error absoluto medio de 4 es mucho o poco?)

MAPE (Mean Absolute Percentage Error)

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{e_i}{y_i} \right|$$

- Entendible (porcentual)
- Medida Relativa
- No sesgada en valores extremos

- Valores 0 en la serie temporal
- No útil cuando la escala de la serie es pequeña (error porcentual muy grande)

SMAPE (symmetric MAPE)

$$SMAPE = \frac{1}{n} \sum_{i=1}^n \frac{|e_i|}{(|y_i| + |\hat{y}_i|)/2}$$

- Medida Relativa
- No sesgada en valores extremos
- Considera valores 0

- Valores 0 en la predicción o en la serie temporal (+200%)
- Un mismo error puede dar porcentajes distintos dependiendo de si el error es positivo o negativo

- Over-forecasting: $A_t = 100$ and $F_t = 110$ give SMAPE = 4.76%
- Under-forecasting: $A_t = 100$ and $F_t = 90$ give SMAPE = 5.26%.

Forecasting:

Problema Real



- *Somos una fábrica de cerveza*
- *Tenemos 500 tiendas/bares repartidos por Murcia*
- *En el 8 % de los bares, uno de cada 3 días se queda sin stock de cerveza*
- *En media se están vendiendo 1,000 litros de cerveza al día en cada tienda,*
- *Por cada litro ganamos 0,2 € de beneficio.*

$$\begin{array}{l} \text{Beneficio/año} \\ \text{y tienda} \end{array} \begin{array}{l} (\text{litro/día}) \\ (\text{días/año}) \end{array} = 1.000 \times 0.2 \times 365 = 73,000 \text{ € / año}$$

(€/litro)

$$\text{Pérdida por tienda} = 73,000 / 3 = 24,000 \text{ €}$$

$$\text{Pérdida total/año} = 24,000 \times 500 \times 8\% = 960,000 \text{ €}$$

Forecasting:

Problema Real



- *Para evitar el problema de Stock queremos realizar un modelo predictivo que nos diga cual va a ser el consumo diario de cada tienda.*
- *El modelo tiene que ajustar lo máximo posible porque generar stock de más, genera perdidas. Se estropea la mercancía en el almacén.*

Forecasting:

Problema Real – Datos



- *Necesitamos ayuda para saber qué datos podemos recolectar*
- *Bloques de datos:*
 - *Datos de la tienda: histórico de ventas, ...?*
 - *Datos de calendario: día semana, ...??*
 - *Datos del cliente: Genero, ... ???*
 - *Catalogo de precios: ???*
 - *Demográficos: ??*
 - *Competidores: ??*
 - *El tiempo: ???*
 - *Online y web: ???*
 - *Geoespaciales: ???*

Forecasting:

Problema Real – Modelo



- *Predicción de 500 series temporales (500 tiendas)*
- *Nuestro equipo ha desarrollado 3 modelos:*
 - *ARIMA*
 - *Prophet*
 - *LightGBM*

Forecasting:

Problema Real – Elección del Modelo

	MAPE (Mean Absolute Percentage Error)	MAE (Mean Absolute Error)	ME (Mean Error)
Modelo 1: ARIMA	15%	162	+ 30
Modelo 2: Prophet	13%	140	- 68
Modelo 3: LGBM	10%	114	+ 46

Forecasting:

Problema Real – Elección del Modelo

- *Los errores son buenos pero son agregados a todas las series temporales*
- *No tenemos visibilidad de estos datos por tienda lo que puede hacer que unas tiendas tengan un error medio o porcentual muy bueno (error prácticamente 0 y 1% porcentual) y otras error alto y +20% de error porcentual.*
- *¿Podemos sacar estos mismos datos para cada serie temporal? → Certeza en la validez del modelo para la predicción de cada tienda*
- *Un modelo agregado (1 modelo para N series temporales) (Random Forest, LGBM, ...) puede ajustar muy bien para unas tiendas y mal para otras*
- *Las series temporales pueden no tener la misma naturaleza y no podemos hacer una caracterización de todas las series temporales a mano → modelos más sencillos pueden no captarla*

¿Qué podemos hacer?

Forecasting:

Problema Real – Elección del Modelo

Modelos Individuales 1 modelo para 1 serie

- *Entrenas el modelo y lo lanzas cada vez que recibes un nuevo dato* ✓
- *Capturan la tendencia y te pueden dar una predicción rápida y ajustada a la ultima información disponible* ✓
- *No van a capturar patrones de comportamiento dados en otras tiendas* ✗
- *Necesitan preprocesado de series temporales* ✗
- *Menos precisos?* ✗

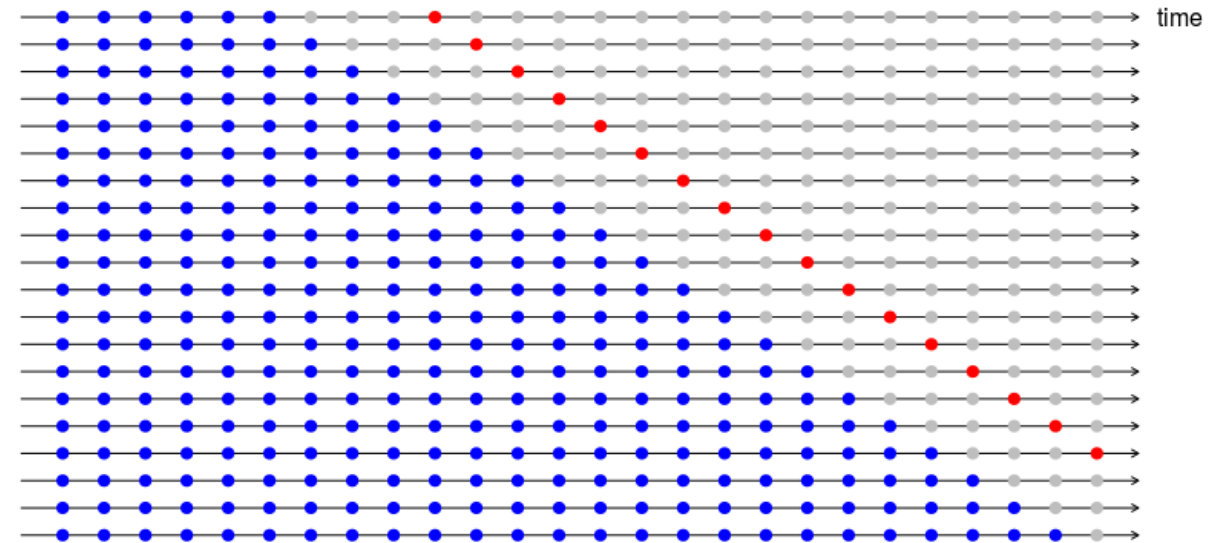
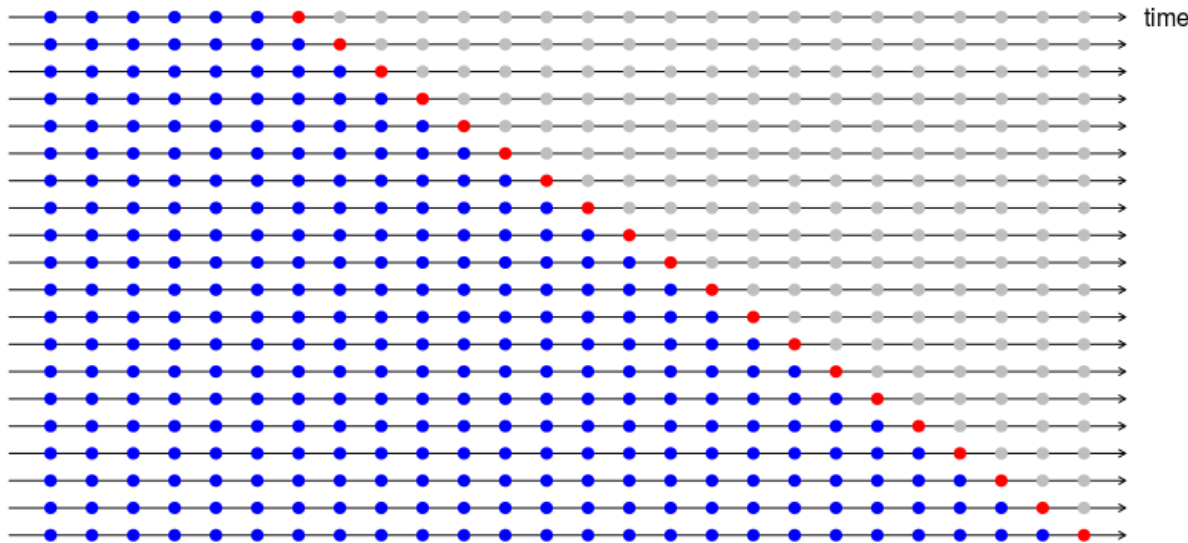
Modelos Agregados 1 modelo para N series

- *El modelo ya está entrenado y realizas la predicción cada mes* ✓
- *Más difícil de mantener y reentrenar* ✗
- *Cambios de comportamiento en las series pueden provocar un reentrenamiento* ✗
- *Aprende de patrones de comportamiento de otras series ante un mismo evento* ✓
- *Pueden no necesitar preprocesado de series temporales* ✓
- *Más precisos?* ✓

Forecasting:

Validación cruzada en series temporales

<https://robjhyndman.com/hyndsight/tscv/>

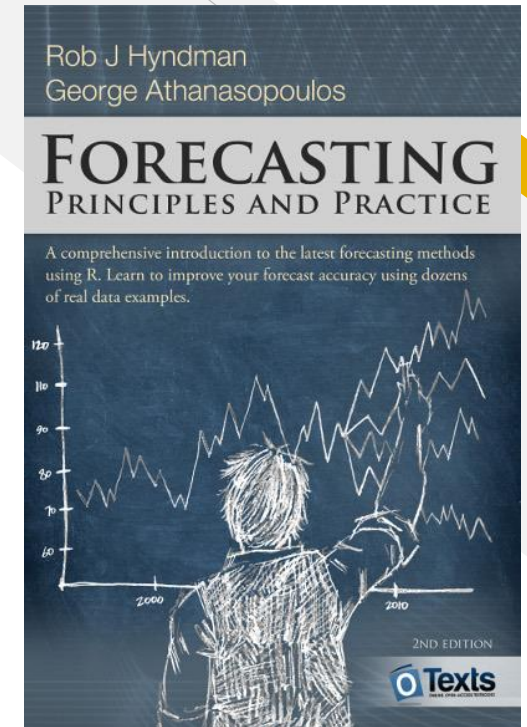


<https://facebook.github.io/prophet/docs/diagnostics.html>

Documentación:

Enlaces de interés

- **Forecasting: principles and practice**
 - <https://otexts.com/fpp2/>
- **Exploratorio y preprocesado:**
 - <https://machinelearningmastery.com/decompose-time-series-data-trend-seasonality/>
 - <https://www.rdocumentation.org/packages/stats/versions/3.6.1/topics/stl>
 - http://humanidades.cchs.csic.es/cchs/web_UAE/tutoriales/PDF/SeriesTemporales.pdf
 - <https://cran.r-project.org/web/packages/stR/vignettes/stRvignette.html>
 - <https://www.bde.es/f/webbde/SES/servicio/software/tramo/slsoutlier.pdf>
 - <https://www.r-bloggers.com/outliers-detection-and-intervention-analysis/>
 - <https://stats.stackexchange.com/questions/104882/detecting-outliers-in-time-series-ls-ao-tc-using-tsoutliers-package-in-r-how>
 - <https://cran.r-project.org/web/packages/tsoutliers/tsoutliers.pdf>
- **Otros enlaces de interés:**
 - <http://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/EDescrip/tema7.pdf>
 - <https://otexts.com/fpp2/>
 - Var tutorial: <https://www.machinelearningplus.com/time-series/vector-autoregression-examples-python/>
- **Kaggle:**
 - EDA: <https://www.kaggle.com/headsortails/be-my-guest-recruit-restaurant-eda>
 - LGBM forecasting: <https://www.kaggle.com/pureheart/1st-place-lgb-model-public-0-470-private-0-502>



Documentación:

Enlaces de interés

- **Datos.gob**
 - <https://datos.gob.es/>
- **INE:**
 - <https://www.ine.es/dyngs/INEbase/listaoperaciones.htm>
- **¿Serie Multiplicativa o aditiva?:**
 - <https://www.r-bloggers.com/is-my-time-series-additive-or-multiplicative/>
 - <https://forecasting.svetunkov.ru/en/2018/03/22/comparing-additive-and-multiplicative-regressions-using-aic-in-r/>
- **BBDD series temporales:**
 - <https://www.influxdata.com/time-series-database/>

1. InfluxDB

2. Kdb+

3. Prometheus

4. Graphite

5. RRDBTool

6. OpenTSDB

7. Druid

8. TimescaleDB

9. FaunaDB

10. KairosDB

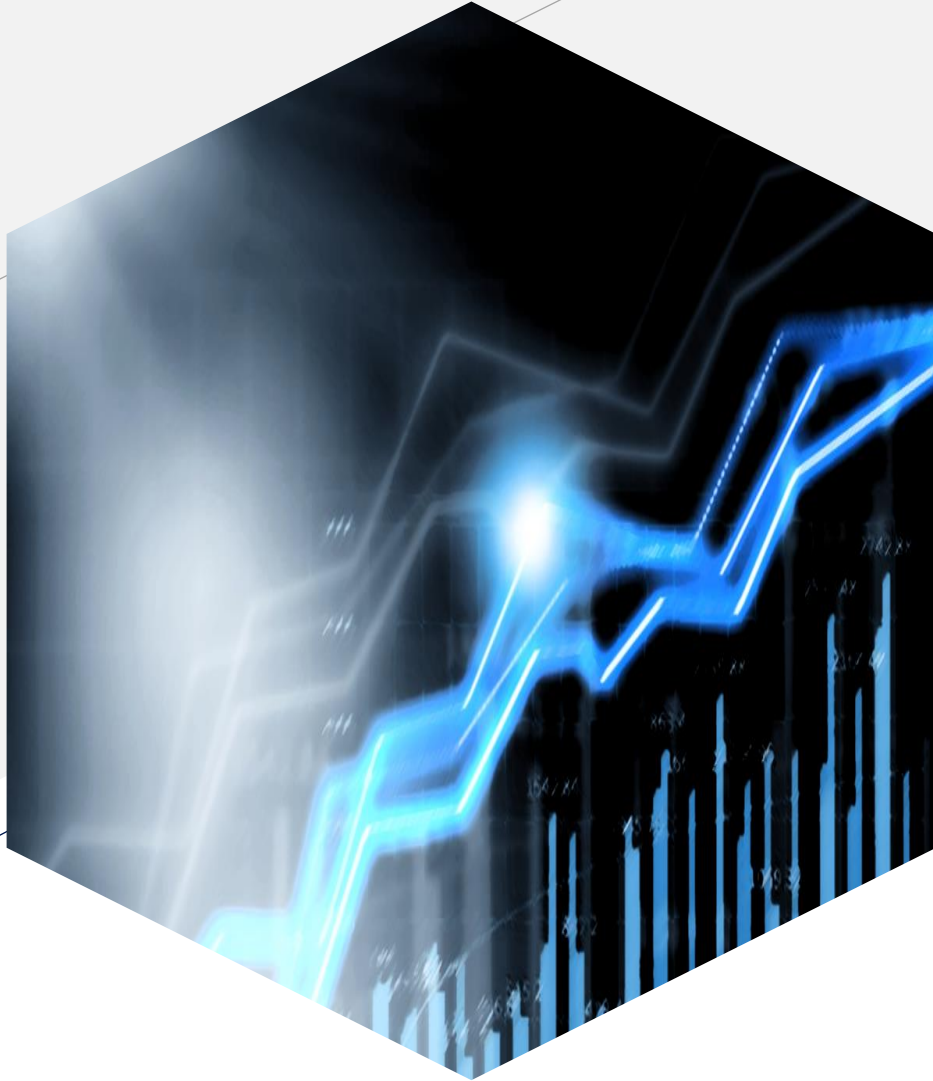
11. GridDB

12. Alibaba Cloud TSDB

13. eXtremeDB

14. Amazon Timestream

15. DolphinDB



Time Series Models & Forecasting

Alberto Valera Pérez

