



# Aviones, delays, cuadrados mínimos... y un poco del viejo de la bolsa

Alfredo Umfurer   Franco Assenza   Kevin Maldonado

Universidad de Buenos Aires

2019

# Table of Contents

Objetivos del trabajo

Marco Teórico

Metodología y datos

Descubrimientos

Trabajo Futuro

# Objetivos del trabajo

- ▶ El objetivo principal del trabajo es entender y poder predecir qué causa los delays de los vuelos de las aerolíneas comerciales.
  - ▶ Es culpa de las aerolíneas?

# Objetivos del trabajo

- ▶ El objetivo principal del trabajo es entender y poder predecir qué causa los delays de los vuelos de las aerolíneas comerciales.
  - ▶ Es culpa de las aerolíneas?
  - ▶ Es culpa de la aleatoriedad del clima?

# Objetivos del trabajo

- ▶ El objetivo principal del trabajo es entender y poder predecir qué causa los delays de los vuelos de las aerolíneas comerciales.
  - ▶ Es culpa de las aerolíneas?
  - ▶ Es culpa de la aleatoriedad del clima?
  - ▶ Es culpa de los aeropuertos sobrepasando su capacidad?

# Objetivos del trabajo

- ▶ El objetivo principal del trabajo es entender y poder predecir qué causa los delays de los vuelos de las aerolíneas comerciales.
  - ▶ Es culpa de las aerolíneas?
  - ▶ Es culpa de la aleatoriedad del clima?
  - ▶ Es culpa de los aeropuertos sobrepasando su capacidad?
- ▶ Además de esto, vamos a hacer un breve análisis para intentar entender los movimientos en bolsa de las aerolíneas.

# Marco Teórico

Ideas:

- ▶ Analizar distintos componentes responsables



# Marco Teórico

Ideas:

- ▶ Analizar distintos componentes responsables
  - ▶ ajustar distintas funciones a los datos

Ideas:

- ▶ Analizar distintos componentes responsables
  - ▶ ajustar distintas funciones a los datos
  - ▶ **Cuadrados mínimos** con la familia de funciones  $\sum a_i \cdot \phi_i(x_i)$  donde las  $\phi$  contemplan

## Ideas:

- ▶ Analizar distintos componentes responsables
  - ▶ ajustar distintas funciones a los datos
  - ▶ **Cuadrados mínimos** con la familia de funciones  $\sum a_i \cdot \phi_i(x_i)$  donde las  $\phi$  contemplan
    - ▶ Una componente lineal, que aparece con el crecimiento natural en el tránsito aéreo.

## Ideas:

- ▶ Analizar distintos componentes responsables
  - ▶ ajustar distintas funciones a los datos
  - ▶ **Cuadrados mínimos** con la familia de funciones  $\sum a_i \cdot \phi_i(x_i)$  donde las  $\phi$  contemplan
    - ▶ Una componente lineal, que aparece con el crecimiento natural en el tránsito aéreo.
    - ▶ Componentes periódicas sinusoidales, para capturar períodos temporales como temporadas, días de semana y horarios pico

## Ideas:

- ▶ Analizar distintos componentes responsables
  - ▶ ajustar distintas funciones a los datos
  - ▶ **Cuadrados mínimos** con la familia de funciones  $\sum a_i \cdot \phi_i(x_i)$  donde las  $\phi$  contemplan
    - ▶ Una componente lineal, que aparece con el crecimiento natural en el tránsito aéreo.
    - ▶ Componentes periódicas sinusoidales, para capturar períodos temporales como temporadas, días de semana y horarios pico
    - ▶ Una componente especialmente puntiaguda, para capturar los días festivos de fin de año.

## Ideas:

- ▶ Analizar distintos componentes responsables
  - ▶ ajustar distintas funciones a los datos
  - ▶ **Cuadrados mínimos** con la familia de funciones  $\sum a_i \cdot \phi_i(x_i)$  donde las  $\phi$  contemplan
    - ▶ Una componente lineal, que aparece con el crecimiento natural en el tránsito aéreo.
    - ▶ Componentes periódicas sinusoidales, para capturar períodos temporales como temporadas, días de semana y horarios pico
    - ▶ Una componente especialmente puntiaguda, para capturar los días festivos de fin de año.
    - ▶ Componentes basadas en scores, de la aerolínea y de los aeropuertos

- ▶ Componentes impredecibles
  - ▶ Clima
  - ▶ Eventualidades relacionadas con los aeropuertos y las aerolíneas
  - ▶ Eventos mundialmente relevantes

- ▶ Componentes impredecibles
  - ▶ Clima
  - ▶ Eventualidades relacionadas con los aeropuertos y las aerolíneas
  - ▶ Eventos mundialmente relevantes
    - ▶ Caidas de la bolsa
    - ▶ 9/11



- ▶ Componentes impredecibles
  - ▶ Clima
  - ▶ Eventualidades relacionadas con los aeropuertos y las aerolíneas
  - ▶ Eventos mundialmente relevantes
    - ▶ Caidas de la bolsa
    - ▶ 9/11
- ▶ Componentes fuera de nuestro alcance
  - ▶ Bancarrotas de aerolíneas
  - ▶ Fusiones de aerolíneas

# Metodología y datos

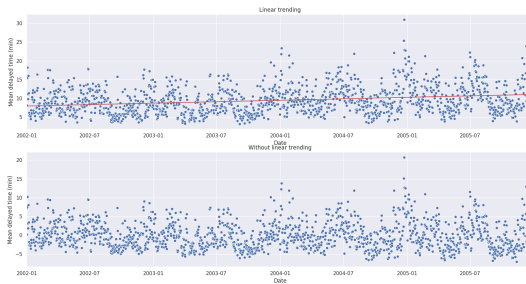
Los datos utilizados para el análisis son los de vuelos de aerolíneas comerciales estadounidenses en el período abarcado entre el 2003 y el 2008.

- ▶ Intencionalmente evitamos la crisis del 9/11 (2001) y la recesión del 2008.

Usamos la siguiente metodología para analizar los datos y sacar conclusiones

# Metodología y datos - Linear Trending

Consiste en remover la componente lineal a los datos



# Metodología y datos - Funciones sinusoidales

- Podemos ajustar usando una combinación lineal de senos y cosenos

# Metodología y datos - Funciones sinusoidales

- ▶ Podemos ajustar usando una combinación lineal de senos y cosenos
- ▶  $c \times \sin(x + \alpha) = c \times \cos \alpha \sin x + c \times \sin \alpha \cos x$

# Metodología y datos - Funciones sinusoidales

- ▶ Podemos ajustar usando una combinación lineal de senos y cosenos
- ▶  $c \times \sin(x + \alpha) = c \times \cos \alpha \sin x + c \times \sin \alpha \cos x$
- ▶ Solo necesitamos definir la amplitud.

# Metodología y datos - Funciones sinusoidales

- ▶ Podemos ajustar usando una combinación lineal de senos y cosenos
- ▶  $c \times \sin(x + \alpha) = c \times \cos \alpha \sin x + c \times \sin \alpha \cos x$
- ▶ Solo necesitamos definir la amplitud.
  - ▶ Asumiendo un comportamiento periódico anual, elegimos frecuencias múltiplo de 365

# Metodología y datos - Día de la semana

**Idea:** Podemos usar el mismo enfoque que para el año: tomar frecuencias semanales



# Metodología y datos - Día de la semana

**Idea:** Podemos usar el mismo enfoque que para el año: tomar frecuencias semanales

**Otra idea:** Usar una función indicadora que evalúe en 1 para una día y 0 en los demás

# Metodología y datos - Día de la semana

**Idea:** Podemos usar el mismo enfoque que para el año: tomar frecuencias semanales

**Otra idea:** Usar una función indicadora que evalúe en 1 para una día y 0 en los demás

- ▶ **senos y cosenos:**, tenemos dos features por cada frecuencia
- ▶ **Indicadora:** Tenemos 7 features, uno por cada día de la semana

# Descubrimientos

- ▶ No hay que volar con United porque te sacan del vuelo con un taser porque overbookean los vuelos

# Trabajo Futuro

- ▶ Contemplar merges de aerolineas
- ▶ Dónde podríamos construir un aeropuerto para mejorar los delays?
- ▶ Qué aeropuerto es el más importante alivianador de delays?
- ▶ Qué aeropuerto es el más importante para mantener a la red conectada?
- ▶ Qué dato útil e interesante podemos sacar del dataset