

# Primer Parcial - INF-A - CUNEF Universidad

Análisis y Explotación de la Información

2023-10-23

## Contents

Introducción . . . . .	1
Ejercicio 1 (2.25 pts) . . . . .	1
Ejercicio 2 (2.25 pts) . . . . .	1
Ejercicio 3 (2 pts) . . . . .	2
Ejercicio 4 (3.5 pts) . . . . .	2

## Introducción

Los datos `flights13.csv`, disponibles en Canvas, contiene información sobre una muestra de 10000 vuelos que partieron de la ciudad de Nueva York (NYC) de los aeropuertos Aeropuerto Internacional Libertad de Newark (EWR), Aeropuerto Internacional John F. Kennedy (JFK) y Aeropuerto Internacional de La Guardia LGA en 2013.

Esta base de datos contiene información acerca de 19 variables. No obstante, para este examen únicamente utilizaremos las siguientes:

- `hour`: hora del vuelo.
- `month`: mes del vuelo.
- `origin`: aeropuerto de origen.
- `dest`: aeropuerto de destino.
- `air_time`: tiempo en el aire, en minutos.
- `distance`: distancia entre aeropuerto de origen y destino, en kilómetros.
- `carrier`: aerolínea
- `dep_delay`: retraso en despegue (en minutos, número negativo significa que el vuelo despegue antes de lo previsto).
- `arr_delay`: retraso en la llegada (en minutos, número negativo significa que el vuelo llega antes de lo previsto).

Antes de continuar, carga los datos en R en una variable llamada `flights_small`.

## Ejercicio 1 (2.25 pts)

Considere únicamente los vuelos de `flights_small` cuyo retraso en el despegue está entre 0 y 5 minutos. Para estos vuelos, calcula la mediana y el percentil 25 del retraso en la llegada. ¿Qué conclusiones extraes?

- Pista: elimina todas las filas en las que `arr_delay` y `dep_delay` tengan un valor ausente.

## Ejercicio 2 (2.25 pts)

Considere solo los vuelos de `flights_small` que **no** tienen un NA en la información de tiempo en el aire y distancia entre origen y destino: ¿Los vuelos a qué aeropuerto tienen mayor velocidad **en media**? ¿Cuánto vale esta media? ¿A qué crees que se debe el resultado?

- Pista 1: Tendrás que crear una nueva variable llamada **velocity** que equivalga al cociente entre distancia entre aeropuerto de origen y destino y tiempo de vuelo. Esta variable tendrá unidades de km/min.
- Pista 2: si no conoces a qué aeropuerto corresponden las siglas del aeropuerto resultante, puedes preguntar al profesor.

### Ejercicio 3 (2 pts)

A partir de los datos `flights_small`, crea una variable llamada `dep_status`. Se trata de una variable categórica que es igual a “ontime” cuando `dep_delay`  $\leq 0$  y “delayed” cuando `dep_delay`  $> 0$ . Guarda todos los datos incluyendo esta variable en un tibble llamado `flights_small_new`.

### Ejercicio 4 (3.5 pts)

Recrea el gráfico del pdf utilizando los datos de `flights_small_new`. Fíjate en todos los detalles. Además, describe el propósito de esta visualización y extrae alguna conclusión.

- Pista 1: AA y B6 son dos aerolíneas. JFK y LGA son los aeropuertos de origen.
- Pista 2: elimina todas las filas en las que `arr_delay` y `dep_delay` tengan un valor ausente.

#### Distribución de retrasos en llegada de vuelos desde JFK and LGA

Octubre – Noviembre 2013

