



BIG DATA PROYECTO FINAL

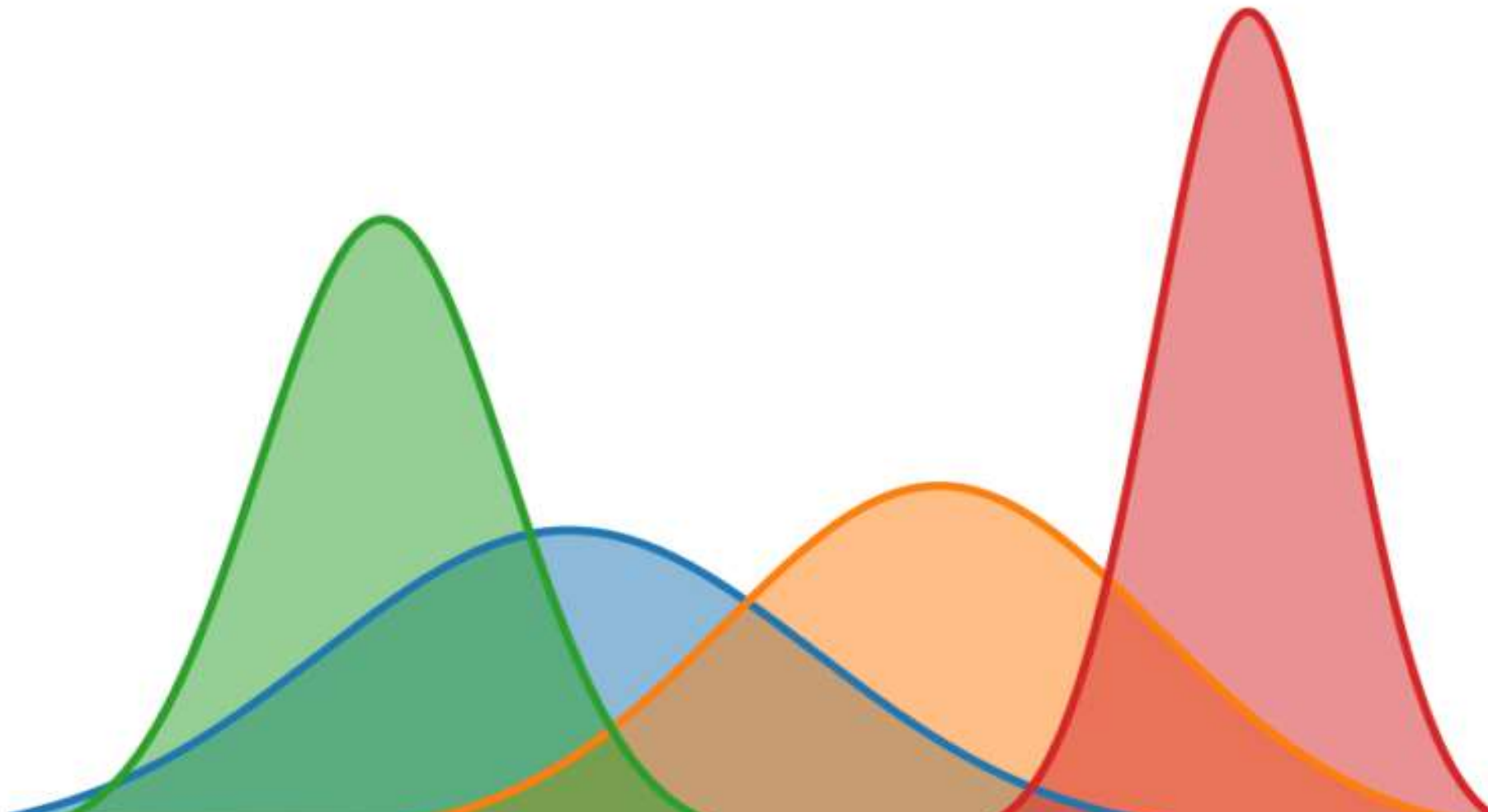
Jose Miguel Zazo Orozco

ANALISIS DEL DATASET

| | Activity Period | Operating Airline | Operating Airline IATA Code | Published Airline | Published Airline IATA Code | GEO Summary | GEO Region | Activity Type Code | Price Category Code | Terminal | Boarding Area | Passenger Count | Airline |
|---|-----------------|-------------------|-----------------------------|-------------------|-----------------------------|---------------|------------|--------------------|---------------------|------------|---------------|-----------------|---------|
| 0 | 200507 | ATA Airlines | TZ | ATA Airlines | TZ | Domestic | US | Deplaned | Low Fare | Terminal 1 | B | 27271 | D |
| 1 | 200507 | ATA Airlines | TZ | ATA Airlines | TZ | Domestic | US | Enplaned | Low Fare | Terminal 1 | B | 29131 | E |
| 2 | 200507 | ATA Airlines | TZ | ATA Airlines | TZ | Domestic | US | Thru / Transit | Low Fare | Terminal 1 | B | 5415 | |
| 3 | 200507 | Air Canada | AC | Air Canada | AC | International | Canada | Deplaned | Other | Terminal 1 | B | 35156 | D |
| 4 | 200507 | Air Canada | AC | Air Canada | AC | International | Canada | Enplaned | Other | Terminal 1 | B | 34090 | E |

He obtenido individualmente las medias de todas las columnas y sus desviaciones típicas que nos ayudaran en los próximos cálculos para determinar cuáles son las variables mas importante en el estudio del

comportamiento de San Francisco



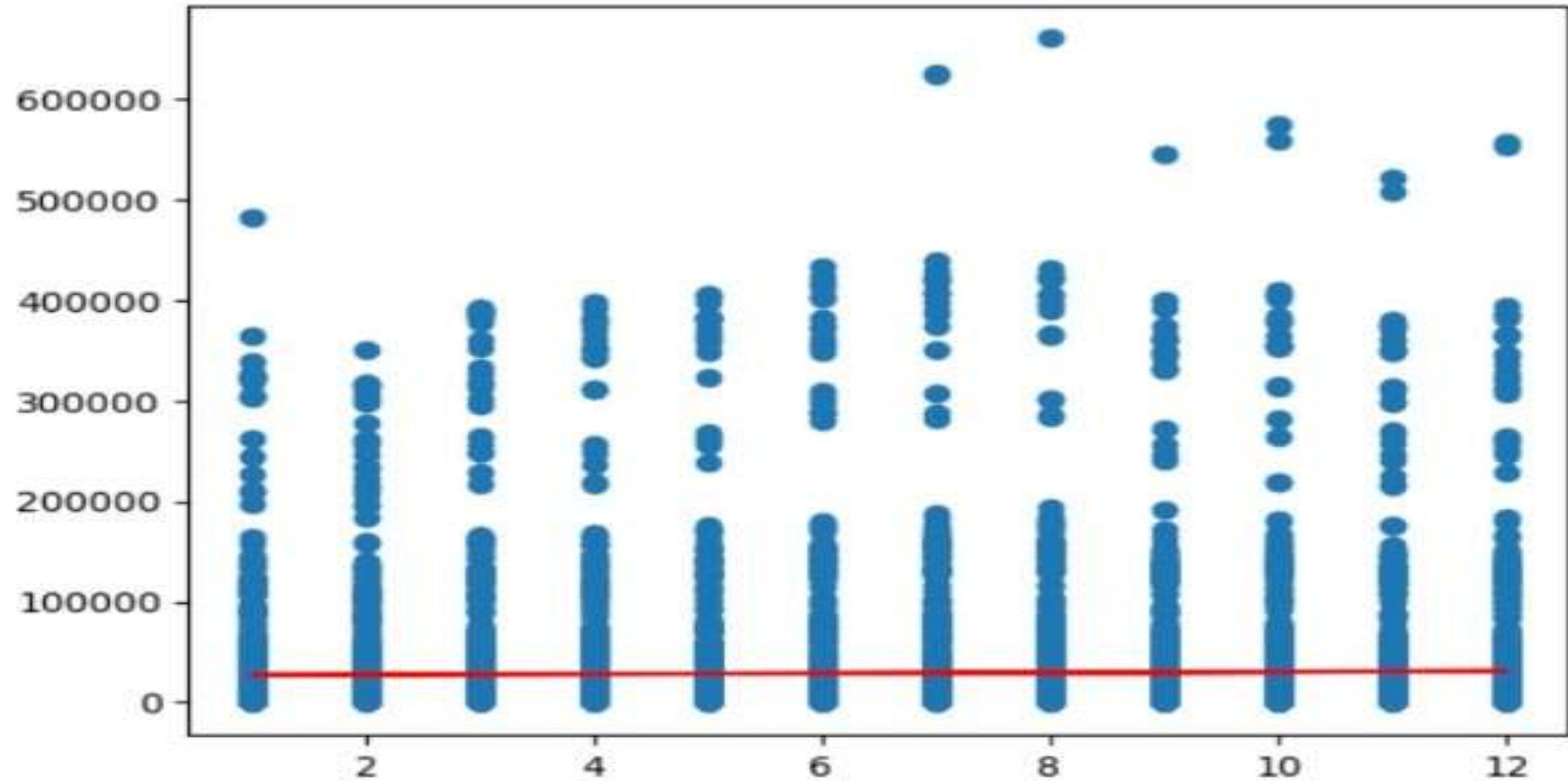
La siguiente fase ha sido convertir las columnas string en enteros con el fin de poder utilizarlos para el análisis y ser precisos con nuestras conclusiones



```
Df['variable'] = df['variable'].astype(int)
```

Convertir String a Int
en Python

Tras establecer una regresión lineal entre múltiples variables, sus coeficientes de correlación nos indican que no están relacionadas las variables (por lo menos de forma lineal)



el coeficiente de correlacion es: -0.00126925500044206