

Ciencia de Datos e Inteligencia de Negocios

Tarea 1. Obtención de información de bases de datos.

En clase se introdujeron los archivos “Accidents_UK_2015.csv”, “Casualties_UK_2015.csv” y “Vehicles_UK_2015.csv” donde se encuentra información sobre los accidentes automovilísticos que fueron reportados durante el año 2015 en el reino unido. Esta actividad considera que se trabajará sobre estos tres archivos.

Obtener el dato o información que se solicita en cada uno de los puntos siguientes y agregue brevemente el código que usaron para obtener la información en cada caso. (El primer punto es un ejemplo).

1. ¿Cuáles son los 10 días con mayor numero de accidentes del año?

Código

```
N_date = pd.DataFrame(pd.value_counts(accidents['Date']),columns=['Num_Accidentes'])
N_date.iloc[0:10,:]
```

Resultado

Date	Num_Accidents
05/11/2015	535
02/10/2015	524
04/12/2015	520
13/11/2015	513
01/10/2015	510
03/07/2015	509
11/09/2015	507
13/01/2015	504
06/10/2015	503
30/06/2015	499

2. Del archivo “Accidents_2015.csv”:
 - a. Realizar un histograma de las localizaciones de los accidentes, y comentar estos histogramas.
 - b. Determinar el día de la semana en el que ocurren el mayor y el menor número de accidentes.
3. Del archivo “Casualties_UK_2015.csv”:
 - a. Determinar el número de hombres y de mujeres que resultaron heridos en los accidentes.

- b. Incluya un histograma de las edades de las personas heridas y determine cuál es la edad de la mayoría de las personas que resultaron heridas.
- 4. Del archivo "Vehicles_UK_2015.csv":
 - a. Obtenga el número de hombres y mujeres que conducían en los accidentes.
 - b. Determine la edad promedio que las personas que conducían y que se accidentaron más y de las que se accidentaron menos.

Todas las figuras deben de estar comentadas o descritas.

La tarea será entregada en digital (Word, PDF, Notebook) y será subido en la liga de Moodle disponible.