

# Python. Trabajo final

## Accidentes de tráfico en Madrid

Autor: Cristóbal Pareja Flores

Actualizado: marzo de 2022

## NOTA PRELIMINAR

Antes de desarrollar el proyecto que te propongo a continuación, te aconsejo leer el enunciado completamente, incluyendo las instrucciones de entrega y otros consejos que he recogido al final de este enunciado.

También quiero señalar que, en este enunciado, se pueden producir cambios durante su desarrollo por distintos motivos: porque se descubre una situación anómala que no he previsto al preparar el enunciado, porque algún estudiante encuentra un error o ambigüedad en mi descripción (así es: también yo cometo errores) o en los datos de partida o por cualquier otra causa. Esto ocurre también en cualquier proyecto real, así que tendremos que tomárnoslo con el mejor talante y aprender también de estas eventuales contingencias. Se procurará informar debidamente del mejor modo posible cuando esto se produzca.

## INTRODUCCIÓN

En el portal de datos abiertos del Ayuntamiento de Madrid se puede encontrar abundante información sobre distintos aspectos de esta hermosa ciudad, como son las actividades de ocio, culturales, deportivas, las actuaciones de los bomberos, las ayudas sociales, la meteorología, la calidad del aire, la cartografía de la ciudad, y un largo etcétera. Ahora nos centramos en los accidentes de tráfico.

Para ello, entramos en el portal mencionado

<https://datos.madrid.es/portal/site/egob>

y pasamos al catálogo de datos directamente. En su primera página encontramos información sobre los accidentes de tráfico de la Ciudad de Madrid:

<https://datos.madrid.es/portal/site/egob/menuitem.c05c1f754a33a9fbe4b2e4b284f1a5a0/?vgnextoid=7c2843010d9c3610VgnVCM2000001f4a900aRCRD&vgnnextchannel=374512b9ace9f310VgnVCM100000171f5a0aRCRD&vgnnextfmt=default>

En esta página podemos encontrar los datos con los que vamos a trabajar, junto con información sobre la estructura de los archivos disponibles. Te proporciono los archivos principales junto con este enunciado, para que todos partamos de un archivo idéntico, con una misma fecha de actualización, y sea más fácil identificar las posibles dudas, las salidas proporcionadas, etc.

Abrimos la tabla en formato csv y observamos lo siguiente:

Nº EXPEDIENTE	FECHA	HORA	CALLE	NÚMERO	DISTRITO	TIPO ACCIDENTE	ESTADO MET	TIPO VEHÍCULO	TIPO PERSONA	RANGO DE EDAD	SEXO	LESIVIDAD*
2020S000057	01/01/2020	23:15	AVDA. CIUDAD D-		RETIRO	Choque contra ob	Despejado	Turismo	Conductor	DE 25 A 29	Al Hombre	
2020S000038	01/01/2020	22:35	CALL. VALLE DE T-		MONCLOA-A	Caída	Despejado	Ciclomotor	Conductor	DE 21 A 24	Al Mujer	6
2020S000060	01/01/2020	20:15	GTA. MARIANO S	1	FUENCARRA	Caída	Despejado	Turismo	Conductor	DE 45 A 49	Al Hombre	14
2020S000060	01/01/2020	20:15	GTA. MARIANO S	1	FUENCARRA	Caída	Despejado	Motocicleta	Conductor	DE 25 A 29	Al Hombre	7
2020S000033	01/01/2020	19:45	CALL. OLIVAR	40	CENTRO	Choque contra ob	Despejado	Turismo	Conductor	DESCONOCIDA		
2020S000035	01/01/2020	19:30	CALL. ANTONIO I-		CARABANCH	Colisión fronto-la	Despejado	Turismo	Conductor	DESCONOCIDA		14
2020S000035	01/01/2020	19:30	CALL. ANTONIO I-		CARABANCH	Colisión fronto-la	Despejado	Motocicleta	Conductor	DE 21 A 24	Al Mujer	2
2020S000073	01/01/2020	18:48	AVDA. PIO XII	81	CHAMARTÍN	Atropello a perso	Despejado	Turismo	Conductor	DE 55 A 59	Al Hombre	14
2020S000073	01/01/2020	18:48	AVDA. PIO XII	81	CHAMARTÍN	Atropello a perso	Despejado	Turismo	Peatón	DE 18 A 20	Al Hombre	7
2020S000067	01/01/2020	18:40	PLAZA. BEATA M-		ARGANZUEL	Caída		Turismo	Conductor	DE 55 A 59	Al Hombre	14
2020S000067	01/01/2020	18:40	PLAZA. BEATA M-		ARGANZUEL	Caída		Motocicleta	Conductor	DE 35 A 39	Al Hombre	14
2020S000063	01/01/2020	16:46	CALL. BAHIA DE A-		HORTALEZA	Colisión fronto-la	Despejado	Turismo	Conductor	DE 40 A 44	Al Mujer	
2020S000063	01/01/2020	16:46	CALL. BAHIA DE A-		HORTALEZA	Colisión fronto-la	Despejado	Turismo	Conductor	DE 30 A 34	Al Hombre	

El planteamiento de esta práctica se centra en estos datos, que yo he renombrado como "2020\_Accidentalidad.csv". Para desarrollarla, te proporciono un notebook de Jupyter, llamado "accidentes\_en\_madrid\_enunciados.ipynb". (También te lo doy en formato pdf.) Es una plantilla de trabajo. Trabaja sobre él, completando código, añadiendo celdas de código y explicaciones según tu criterio y modificándolo a medida que se desarrolles la práctica.

## DESCRIPCIÓN DE LA ENTREGA

La entrega consistirá en una carpeta comprimida, identificada con los apellidos y nombre del estudiante (por ejemplo, "Pareja\_Flores\_Cristobal"), sin tildes ni ñes. Esta carpeta deberá contener los siguientes archivos:

- El archivo "accidentes\_en\_madrid.ipynb" (también en formato pdf) con cada uno de los apartados planteados, resueltos por ti, explicados adecuadamente, a excepción del apartado sobre la técnica map-reduce, que se ha de desarrollar aparte. Cada apartado tendrá su rótulo, en modo markdown, con el enunciado, la solución, unos pocos ejemplares de prueba bien elegidos que muestren el funcionamiento de dicho apartado, y una explicación de lo logrado y de lo no logrado. Algún apartado se puede resolver en varios pasos con distintas funciones, y en tal caso será necesario ilustrar el funcionamiento de cada una de ellas con tests apropiados.

Este archivo se entregará también en formato pdf. En el archivo ipynb (y en el pdf correspondiente) se han de poder ver las comprobaciones parciales de cada uno de los apartados. La entrega debe contener las celdas del archivo ipynb ejecutadas en orden, numeradas correlativamente, desde la primera [1] hasta la última.

Debes comprobar con la antelación suficiente que puede realizar la conversión a pdf y la preparación de la entrega según estas indicaciones.

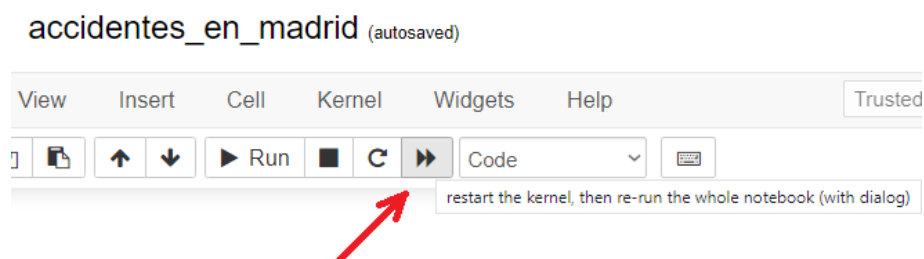
- Para el problema de map-reduce, debes incluir en tu carpeta el programa o los programas necesarios en Python, ejecutables desde la línea de comandos, que implementan las tareas realizadas.

En esta misma carpeta, debes incluir un archivo llamado “mapreduce\_instrucciones\_de\_uso.txt” que describa la invocación que se debe hacer, desde la línea de comandos, del programa con map-reduce. En todo caso, el uso incorrecto de este programa deberá disparar el correspondiente aviso con las instrucciones necesarias para su correcto funcionamiento.

- Los archivos de datos completos, de manera que permitan ilustrar todas las características de las soluciones aportadas.
- En el mismo archivo de Jupyter de la solución, se deben incluir los apartados finales siguientes:
  - Datos personales
  - Una ficha de autoevaluación, indicando los apartados logrados, junto con la nota estimada por el estudiante en cada uno de dichos apartados.
  - La ayuda recibida o fragmentos de código tomados de otro estudiante o de cualquier otra fuente, junto con la bibliografía utilizada o páginas de Internet en que se ha encontrado algún fragmento de código incluíro en tu trabajo.
  - Opcionalmente, cualquier otro comentario que desees añadir.

## COMENTARIOS ADICIONALES

- No se deberá entregar ningún programa que no esté debidamente comprobado y sea correcto. Se recomienda encarecidamente una ejecución final del notebook completo antes de la entrega final:



- Cada función deberá ser clara y estar debidamente estructurada y documentada, siguiendo normas estándar. Este asunto es de gran importancia y se tendrá en cuenta fuertemente en la evaluación.
- Se sobreentiende que cada entrega está hecha exclusivamente por el autor, considerándose inaceptable entregar cualquier trabajo realizado total o parcialmente por otra persona distinta del autor firmante, o un trabajo en que se ha copiado una parte del código propuesto por otros compañeros o de cualquier otra fuente. En todo caso, se ha de consignar con total claridad la ayuda recibida o tomada prestada, por cortesía y por honradez.

## APÉNDICE. ALGUNOS ERRORES FRECUENTES

- Es erróneo organizar un programa sin usar una estructura adecuada en funciones.
- Es erróneo diseñar funciones sin parámetros cuando pueden usarse en distintas situaciones; diseñar funciones sin documentar; emplear identificadores inadecuados; emplear variables globales; definir funciones con constantes literales que registran valores propios del programa y se repiten en distintos lugares del mismo, tales como nombres de archivos, fechas, tamaños de muestras, etc.
- Las funciones necesarias para que nuestra aplicación trabaje no deben contener lecturas de datos (input) ni salidas (print). Los datos deben proporcionarse normalmente a través de los parámetros, y el resultado de una función, normalmente se comunica en el return, para facilitar su uso posterior. La instrucción print suele usarse más bien en instrucciones o pequeños programas de demostración. Se ha de separar el funcionamiento de nuestra aplicación de sus demostraciones de funcionamiento.
- Es erróneo usar rutas absolutas:

```
f = open('C:/Users/blacky/Desktop/Proyectos.txt', 'r')
```

Esto impide ejecutar un programa fuera del ordenador en que se ha diseñado.

Lo correcto es usar rutas relativas:

```
f = open('./Proyectos.txt', 'r')
```

Tu programa debería funcionar bien cuando yo lo active en la carpeta de mi elección de mi equipo, donde no existe la ruta absoluta, como la siguiente por ejemplo:

```
'/Users/blacky/Desktop/Proyectos.txt'
```

- Es erróneo usar archivos csv (o xlsx) para almacenar información textual plana. Cuando se genera un archivo csv, comprueba que cada fila está organizada en celdas distintas, y no todas las componentes en la primera celda, sin separador alguno.
- La presentación de un proyecto en un archivo ipynb permite ir diseñando pequeñas piezas de código, ir explicando sus porqués e ir mostrando el funcionamiento de cada una. No presentes las definiciones todas de golpe, y no olvides ir mostrando el funcionamiento de cada función o cada pequeña pieza de código nueva con las pruebas de funcionamiento necesarias.
- Aunque no se trate de un error, los programadores de otros lenguajes tienden a programar en Python como si fuera Java (o C++, etc.). Familiarízate con los conceptos nuevos de Python y úsalos adecuadamente, persiguiendo en primer lugar la claridad:
  - Recorridos de listas sin mencionar los índices
  - Funciones que devuelven varios valores a la vez en una tupla
  - Asignaciones múltiples
  - Funciones de orden superior
  - Listas intensionales
  - Conjuntos, diccionarios, etc.
- Por último, pregunta las dudas que encuentres en el foro, y participa también aportando tus propios comentarios o propuestas de solución. Un error frecuente es precisamente no plantear las dudas que surjan.