

# Creación de una visualización data Storytelling (PEC3)

December 15, 2021

## 0.1 PEC3-Creación de una visualización data Storytelling

Nombre y Apellidos: Maite Piedra Yera

```
[1]: import pandas as pd
import numpy as np
```

```
[2]: cr = pd.read_csv('shootings.csv')
```

```
[3]: cr.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4895 entries, 0 to 4894
Data columns (total 15 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   id                                     4895 non-null   int64
1   name                                  4895 non-null   object
2   date                                  4895 non-null   object
3   manner_of_death                       4895 non-null   object
4   armed                                 4895 non-null   object
5   age                                    4895 non-null   float64
6   gender                                4895 non-null   object
7   race                                  4895 non-null   object
8   city                                  4895 non-null   object
9   state                                 4895 non-null   object
10  signs_of_mental_illness                4895 non-null   bool
11  threat_level                           4895 non-null   object
12  flee                                    4895 non-null   object
13  body_camera                            4895 non-null   bool
14  arms_category                          4895 non-null   object
dtypes: bool(2), float64(1), int64(1), object(11)
memory usage: 506.8+ KB
```

```
[4]: cr.head()
```

```
[4]:   id      name      date  manner_of_death  armed  age  \
0   3      Tim Elliot  2015-01-02          shot    gun  53.0
1   4  Lewis Lee Lembke  2015-01-02          shot    gun  47.0
```

2	5	John Paul Quintero	2015-01-03	shot and Tasered	unarmed	23.0
3	8	Matthew Hoffman	2015-01-04	shot	toy weapon	32.0
4	9	Michael Rodriguez	2015-01-04	shot	nail gun	39.0

  

	gender	race	city	state	signs_of_mental_illness	threat_level	\
0	M	Asian	Shelton	WA	True	attack	
1	M	White	Aloha	OR	False	attack	
2	M	Hispanic	Wichita	KS	False	other	
3	M	White	San Francisco	CA	True	attack	
4	M	Hispanic	Evans	CO	False	attack	

  

	flee	body_camera	arms_category
0	Not fleeing	False	Guns
1	Not fleeing	False	Guns
2	Not fleeing	False	Unarmed
3	Not fleeing	False	Other unusual objects
4	Not fleeing	False	Piercing objects

### 0.1.1 Explicación del dataset elegido.

Antes de continuar con la práctica vamos a explicar el dataset elegido. El cual consta de 15 columnas y 4885 filas. de las cuales, dos columnas son datos continuos y el resto de columnas son categoricas. El dataset contiene información sobre los tiroteos que involucran a policías y a sospechosos desde Enero del 2015 hasta Julio del 2020(5 años) en Estados Unidos.

A continuacion explicaremos cada uno de las columnas del dataset, para luego ver que columnas realmente son relevantes a nuestro ejercicio.

- id: identificador de la fila
- Name: Nombre de la persona
- Date: Fecha en la que se produjo el altercado
- manner\_of\_death: Manera en que murio
- Armed: Tipo de arma que portaba
- Age: Edad
- Gender: Genero
- Race: Raza
- City: Ciudad
- State: Estado
- Signs\_mental\_illness: Si posee o no alguna enfermedad mental.
- Threat\_level: Tipo de enfrentamiento
- Flee: Si huyo o no de la escena.
- Body\_camara: Si el policia llevaba camara o no.
- arm\_category: Tipo de arma.

A continuacion vamos a limpiar el dataset para quedarnos unicamente con los valores que necesitamos, primeramente vamos a explicar el

```
[5]: cr.describe()
```

```
[5]:
```

	id	age
count	4895.000000	4895.000000
mean	2902.148519	36.549750
std	1683.467910	12.694348
min	3.000000	6.000000
25%	1441.500000	27.000000
50%	2847.000000	35.000000
75%	4352.500000	45.000000
max	5925.000000	91.000000

**Limpieza del dataset** Ahora vamos a limpiar, preparar y seleccionar los datos, con los que finalmente trabajaremos.

```
[6]: cr.isnull().sum()
```

```
[6]: id 0
name 0
date 0
manner_of_death 0
armed 0
age 0
gender 0
race 0
city 0
state 0
signs_of_mental_illness 0
threat_level 0
flee 0
body_camera 0
arms_category 0
dtype: int64
```

```
cr.duplicated().sum()
```

```
[7]: cr['arms_category'].unique().tolist()
```

```
[7]: ['Guns',
      'Unarmed',
      'Other unusual objects',
      'Piercing objects',
      'Sharp objects',
      'Unknown',
      'Blunt instruments',
      'Multiple',
      'Electrical devices',
      'Hand tools',
      'Vehicles',
      'Explosives']
```

```
[8]: cr['arms_category'] = cr['arms_category'].map({'Guns': 'Letal',
        'Unarmed': 'Unarmed',
        'Other unusual objects': 'No Letal',
        'Piercing objects': 'No Letal',
        'Sharp objects': 'No Letal',
        'Unknown': 'No Letal',
        'Blunt instruments': 'No Letal',
        'Multiple': 'No Letal',
        'Electrical devices': 'No Letal',
        'Hand tools': 'No Letal',
        'Vehicles': 'Letal',
        'Explosives': 'Letal'},
        na_action=None)
```

Como podemos ver el tipo de arma lo hemos reducido solo a tres tipos, que serían los siguientes:

- Armas letales: que hemos considerado pistolas, explosivos y vehiculos.
- Armas no letales: El resto de objetos.
- No armados: que equivale a que la persona va desarmada.

```
[9]: cr['new_date'] = pd.to_datetime(cr['date'], format="%Y-%m-%d")
```

```
[10]: cr['Year'] = cr['new_date'].dt.year
```

```
[11]: cr.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4895 entries, 0 to 4894
Data columns (total 17 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   id                                     4895 non-null   int64
1   name                                  4895 non-null   object
2   date                                  4895 non-null   object
3   manner_of_death                      4895 non-null   object
4   armed                                4895 non-null   object
5   age                                   4895 non-null   float64
6   gender                               4895 non-null   object
7   race                                  4895 non-null   object
8   city                                  4895 non-null   object
9   state                                4895 non-null   object
10  signs_of_mental_illness               4895 non-null   bool
11  threat_level                          4895 non-null   object
12  flee                                  4895 non-null   object
13  body_camera                           4895 non-null   bool
14  arms_category                         4895 non-null   object
15  new_date                             4895 non-null   datetime64[ns]
16  Year                                  4895 non-null   int64
dtypes: bool(2), datetime64[ns](1), float64(1), int64(2), object(11)
```

memory usage: 583.3+ KB

```
[12]: cr = cr.astype({"age": int})
```

```
[13]: data = cr.drop(columns = ['id', 'name', 'date',  
    ↳ 'manner_of_death', 'armed', 'new_date'])
```

```
[14]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 4895 entries, 0 to 4894
```

```
Data columns (total 11 columns):
```

#	Column	Non-Null Count	Dtype
0	age	4895 non-null	int64
1	gender	4895 non-null	object
2	race	4895 non-null	object
3	city	4895 non-null	object
4	state	4895 non-null	object
5	signs_of_mental_illness	4895 non-null	bool
6	threat_level	4895 non-null	object
7	flee	4895 non-null	object
8	body_camera	4895 non-null	bool
9	arms_category	4895 non-null	object
10	Year	4895 non-null	int64

```
dtypes: bool(2), int64(2), object(7)
```

```
memory usage: 353.9+ KB
```

Como podemos observar en nuestro dataset no hay ningun valor nulo, ni duplicado, además se ha cambiado la columna 'edad' de real a un valor entero, que es como debe estar.

Por último vamos a exportar nuestro datos a formato excel y a csv.

```
[15]: data.to_csv('finalcd.csv', index=False)
```

```
[16]: data.to_excel("finalcd.xlsx")
```