

Accidentes de tráfico en la Comunidad de Madrid

ThePreprocessingTeam:
ALMANZA AMERICA JHASMINE,
BOTÉY BOPABOTE MARÍA ELENA,
DÍAZ CHAVEZ RAQUEL

| Artificial Intelligence Course

Accidentes de tráfico en la Comunidad de Madrid

| Unit 1. Accidentes de Tráfico

1.1. Importancia de los Accidentes de Tráfico.

1.2. Campañas de tráfico.

| Unit 2. EDA y Preprocesamiento

2.1. Análisis de los datos EDA

2.2. Limpieza de los datos

2.3. Sincronización de los datasets

| Unit 3. Entrenamiento del modelo

3.1. Series temporales y LSTM

3.2. Clasificación y etiquetado

1.1. Importancia de los Accidentes de Tráfico.

Objetivo en la Agenda 2030 :

Reducir a la mitad el número de muertes y lesiones causadas por accidentes de tráfico.

Por medio de:

- Educación y sensibilización
- Infraestructuras seguras
 - Mejorar la seguridad de carreteras y entornos urbanos (barreras de seguridad, reductores de seguridad...).
- Vehículos seguros
 - Sistemas de seguridad activa y pasiva.



1.1. Importancia de los Accidentes de Tráfico.

Año 2022 España (DGT) :

1.145 muertos en accidentes de tráfico

737 eran conductores, 248 peatones, 128 usuarios vulnerables

Año 2022 Mundial (OMS):

En 2022 fallecieron **1,35 millones** de personas en accidentes de tráfico en todo el mundo

Mayor número de muertes por accidentes de tráfico en 2022:

India: **356.000**

China: 248.000

Estados Unidos: 38.800

Brasil: 35.000

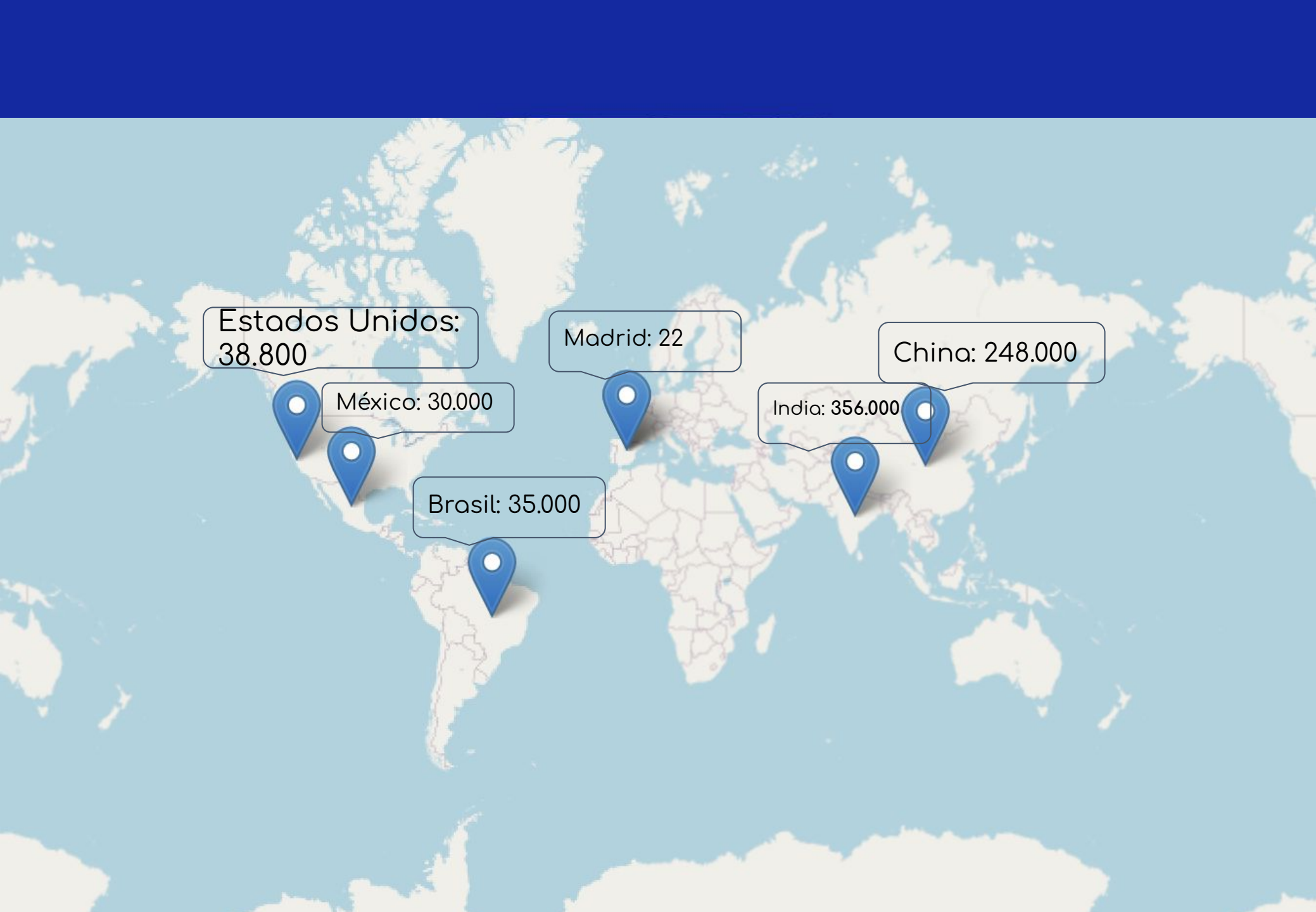
México: 30.000



Año 2022 Madrid (Ayuntamiento de Madrid):

22 fallecidos en accidentes de tráfico en la capital española

15 eran conductores, 4 peatones, 2 ciclistas.



1.1. Importancia de los Accidentes de Tráfico.

Gastos que genera un accidente de tráfico:

- Gastos directos:
 - Rescate y urgencias
 - Médicos
 - Reparación
 - Limpieza
- Gastos indirectos:
 - Rehabilitación
 - Costes humanos (vidas de los afectados)
- Otros
 - Fuerte contaminación
 - Riesgo de incendio



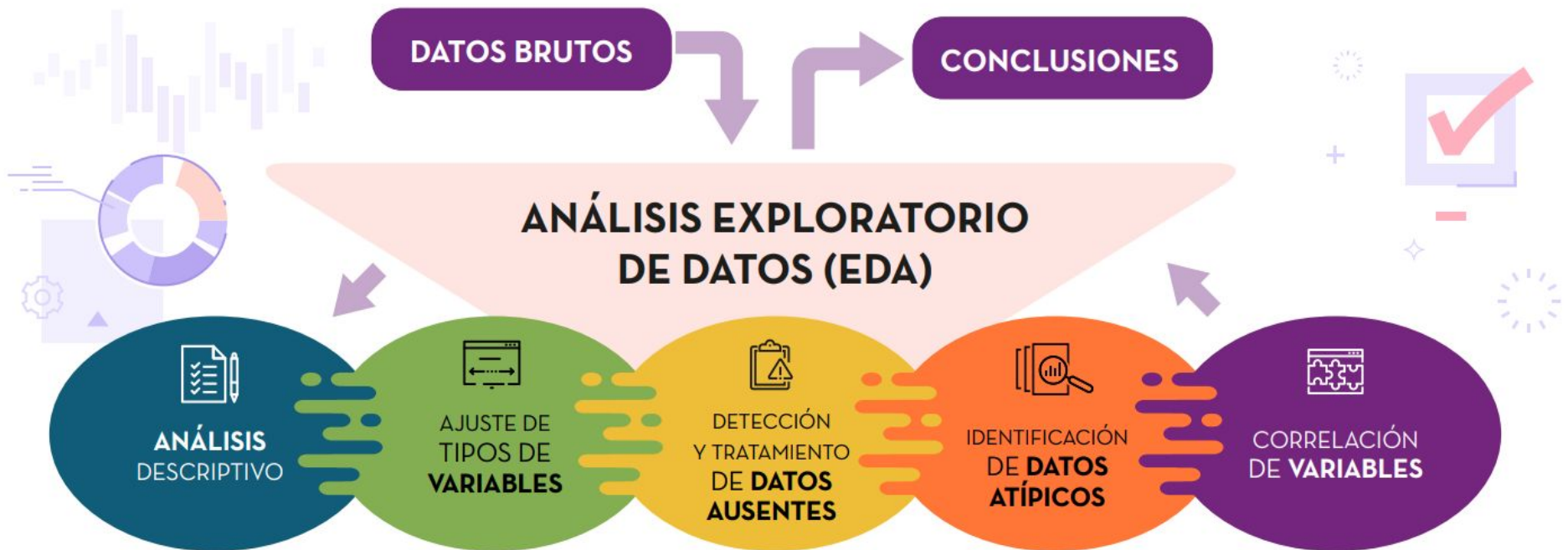
1.2. Campañas de tráfico.



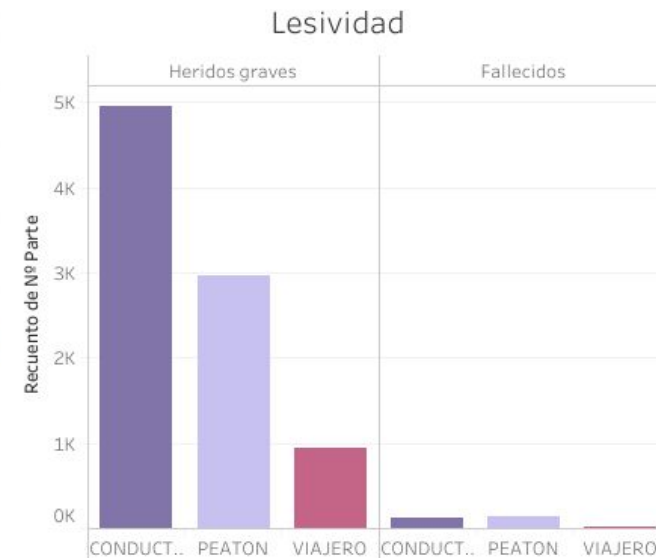
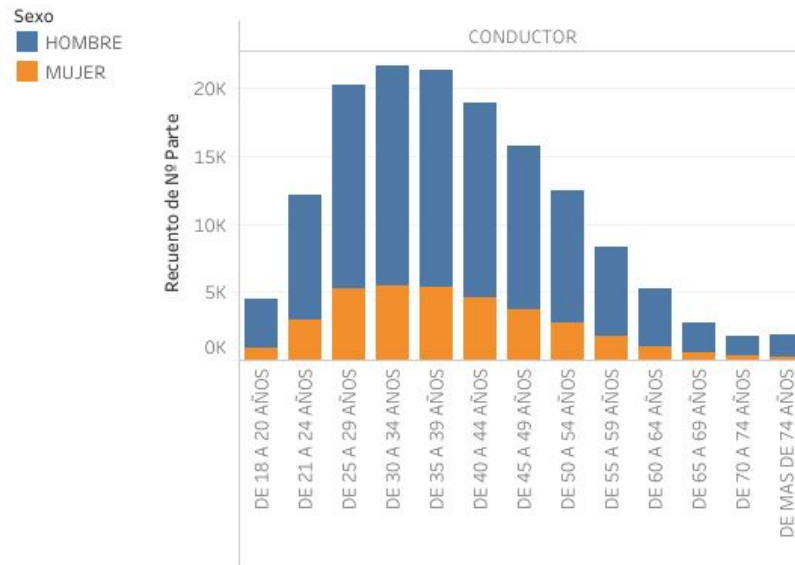
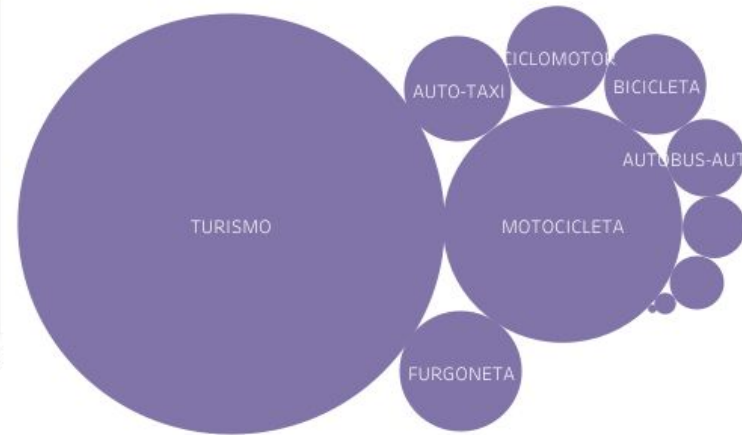
<https://www.dgt.es/comunicacion/campanas/>



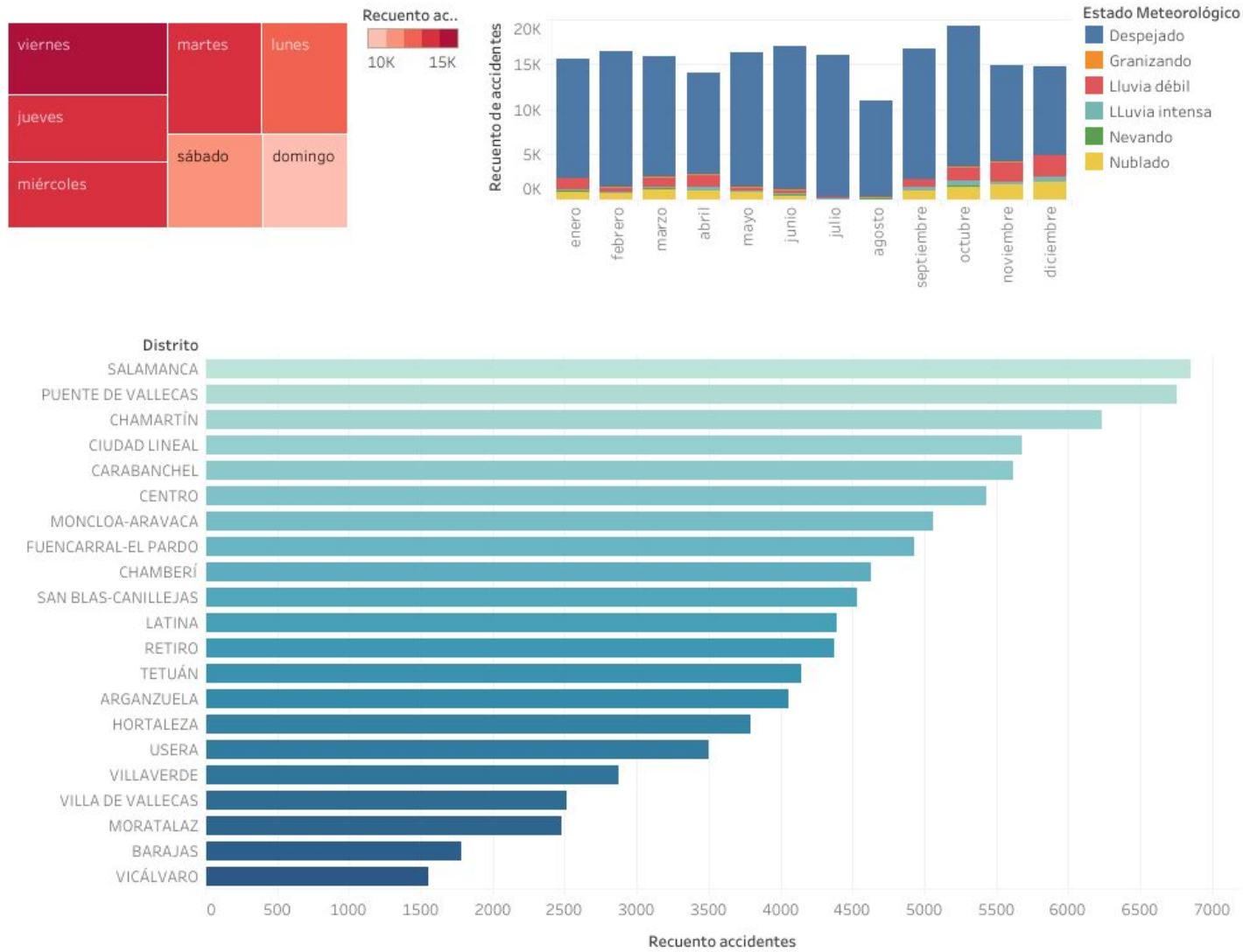
2.1. Análisis de los datos EDA



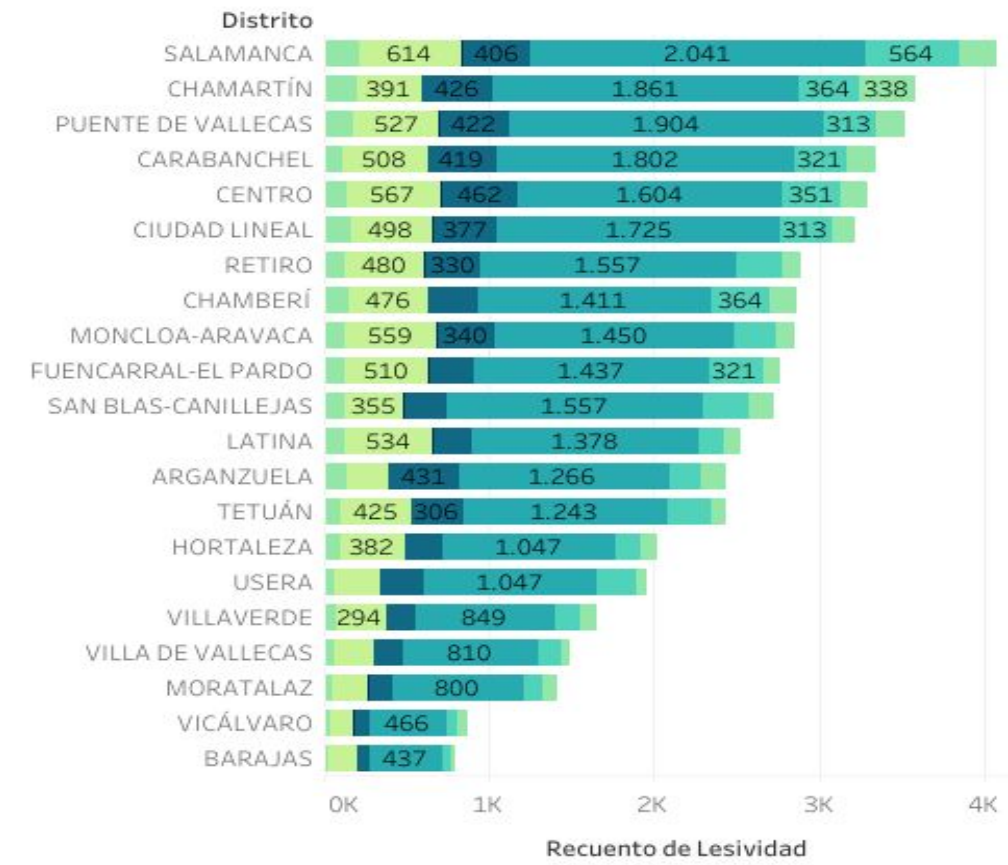
2.1. Análisis de los datos EDA



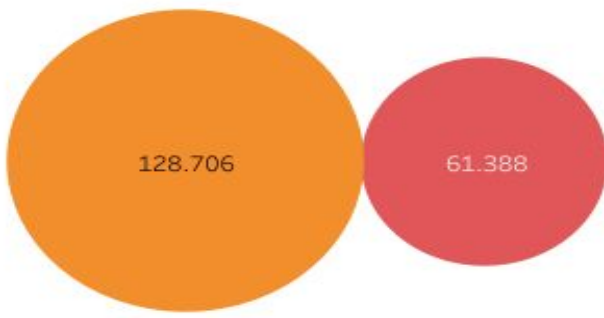
2.1. Análisis de los datos EDA



Lesividad-Distrito

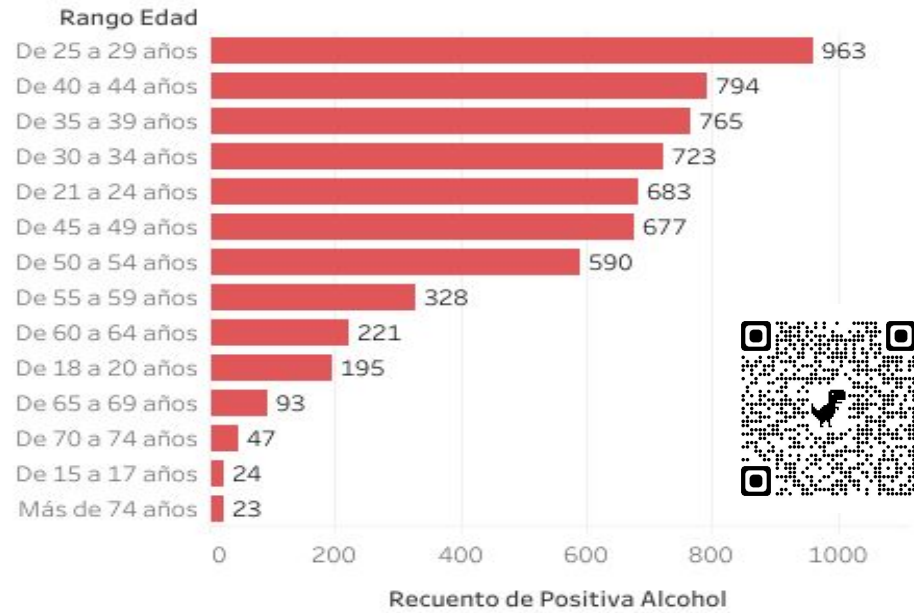
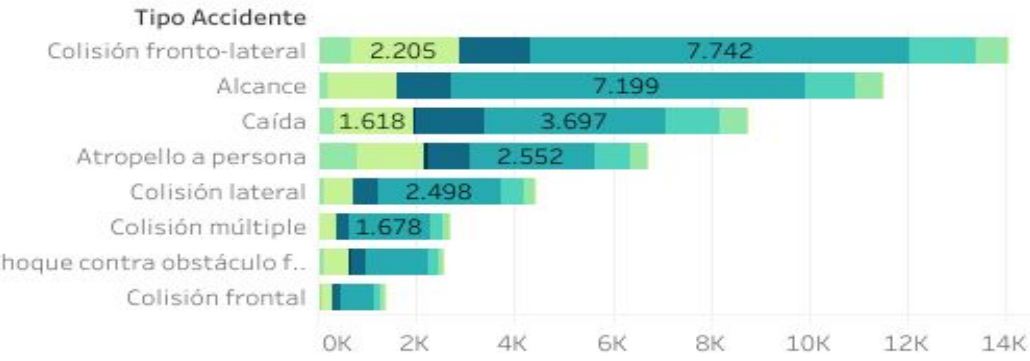


Recuento de Lesividad
Para NULL



[Accidentes trafico 2018](#)
[2023](#)

Tipo Accidente - Lesividad



2.2. Limpieza de los datos

La limpieza de datos es el proceso de corregir o eliminar datos formateados incorrectamente, duplicados o incompletos dentro de un gran conjunto de datos.

Pasos a seguir:

- Identificar la principal fuente de errores y problemas, y determinar cómo remediarlos
- Eliminar los datos duplicados o irrelevantes
- Corregir los errores estructurales
- Filtrar valores atípicos no deseados



2010 -2018

2019 -2023

```
pd.Series(accidentes_antiguo[0].columns)
```

```
pd.Series(accidentes_nuevo[0].columns)
```

✓ 0.0s

✓ 0.0s

```
0          FECHA ==
1    RANGO HORARIO ≠
2      DIA SEMANA →
3      DISTRITO ==
4    LUGAR ACCIDENTE ==
5              N ==
6      N PARTE ==
7
8      CPFA Granizo
9      CPFA Hielo
10     CPFA Lluvia
11     CPFA Niebla
12     CPFA Seco
13     CPFA Nieve
14     CPSV Mojada
15     CPSV Aceite
16     CPSV Barro
17     CPSV Grava Suelta
18     CPSV Hielo
19     CPSV Seca Y Limpia
20     N VICTIMAS *
21     TIPO ACCIDENTE ==
22     Tipo Vehiculo ==
23     TIPO PERSONA ==
24     SEXO ==
25     LESIVIDAD ≠
26     Tramo Edad ==
```

```
0          num_expediente ==
1              fecha ==
2              hora ≠
3      localizacion ==
4          numero ==
5      cod_distrito ==
6          distrito ×
7      tipo_accidente ==
8      estado_meteorológico ×
9      tipo_vehiculo ==
10     tipo_persona ==
11     rango_edad ==
12         sexo ==
13     cod_lesividad ≠
14         lesividad ×
15     coordenada_x_utm ×
16     coordenada_y_utm ×
17     positiva_alcohol ×
18     positiva_droga ×
```

dtype: object

dtype: object

2.3. Sincronización de los datasets y codificación de datos categóricos

IL, HL, HG, MT

| | cod_distrito | dia_semana | fecha | hora | lesividad | localizacion | rango_edad | sexo | tipo_accidente | tipo_persona | tipo_vehiculo |
|-----|--------------|------------|------------|--------------|-----------|--------------------------------|------------|--------|-------------------|--------------|---------------|
| 0 | 5.0 | VIERNES | 2010-01-01 | [0:00, 0:59] | IL | CALLE DE CARTAGENA NUM ... | 18-20 | HOMBRE | COLISIÓN DOBLE | CONDUCTOR | TURISMO |
| 1 | 5.0 | VIERNES | 2010-01-01 | [0:00, 0:59] | IL | CALLE DE CARTAGENA NUM ... | 50-54 | HOMBRE | COLISIÓN DOBLE | CONDUCTOR | TURISMO |
| 2 | 5.0 | VIERNES | 2010-01-01 | [0:00, 0:59] | HL | CALLE DE CARTAGENA NUM ... | 40-44 | MUJER | COLISIÓN DOBLE | VIAJERO | TURISMO |
| 3 | 13.0 | VIERNES | 2010-01-01 | [1:00, 1:59] | HL | AUTOVIA M-30 CALZADA 2 KM. ... | 25-29 | MUJER | COLISIÓN MÚLTIPLE | VIAJERO | NO ASIGNADO |
| 4 | 13.0 | VIERNES | 2010-01-01 | [1:00, 1:59] | HL | AUTOVIA M-30 CALZADA 2 KM. ... | 30-34 | MUJER | COLISIÓN MÚLTIPLE | VIAJERO | NO ASIGNADO |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |

252998 rows × 11 columns

| | cod_distrito_1.0 | cod_distrito_2.0 | cod_distrito_3.0 | cod_distrito_4.0 | cod_distrito_5.0 | ... | sexo_HOMBRE | sexo_MUJER | tipo_persona_CONDUCTOR | tipo_persona_PEATON | tipo_persona_VIAJERO |
|--------|------------------|------------------|------------------|------------------|------------------|-----|-------------|------------|------------------------|---------------------|----------------------|
| 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 1 | 0 | 0 |
| 1 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 1 | 1 | 0 | 0 |
| 2 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 1 | 0 | 0 |
| 3 | 0 | 0 | 0 | 0 | 0 | ... | 1 | 0 | 1 | 0 | 0 |
| 4 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 1 | 0 | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 211969 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 1 | 1 | 0 | 0 |
| 211970 | 0 | 0 | 0 | 1 | 0 | ... | 1 | 0 | 1 | 0 | 0 |
| 211971 | 0 | 0 | 0 | 1 | 0 | ... | 1 | 0 | 1 | 0 | 0 |
| 211972 | 0 | 0 | 0 | 0 | 0 | ... | 1 | 0 | 1 | 0 | 0 |
| 211973 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 1 | 1 | 0 | 0 |

211974 rows × 74 columns

3.1. Series temporales / FORECASTING

union_datasets_forecasting

✓ 0.0s

| | cod_distrito | conteo | fecha |
|-------|--------------|--------|------------|
| 0 | 3.0 | 9 | 2010-01-01 |
| 1 | 4.0 | 4 | 2010-01-01 |
| 2 | 5.0 | 3 | 2010-01-01 |
| 3 | 6.0 | 7 | 2010-01-01 |
| 4 | 7.0 | 2 | 2010-01-01 |
| ... | ... | ... | ... |
| 76748 | 3.0 | 9 | 2023-10-31 |
| 76749 | 2.0 | 6 | 2023-10-31 |
| 76750 | 1.0 | 5 | 2023-10-31 |
| 76751 | 10.0 | 8 | 2023-10-31 |
| 76752 | 21.0 | 2 | 2023-10-31 |

76753 rows × 3 columns

Model:
skforecast -
ForecasterAutoreg

Regressor:
sklearn.ensemble -
RandomForestRegressor

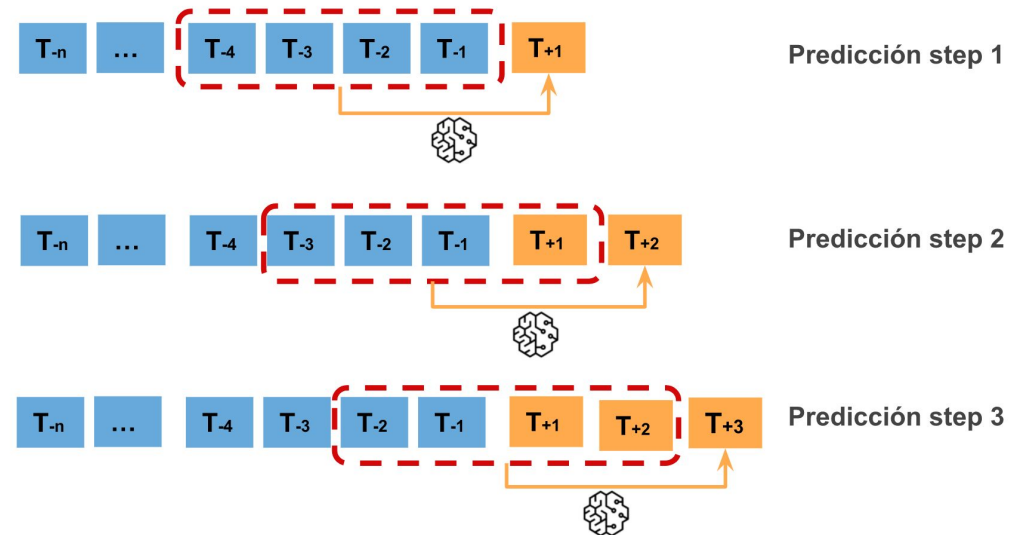


```
forecaster = ForecasterAutoreg.ForecasterAutoreg(  
    regressor = RandomForestRegressor(random_state=42),  
    lags=5  
)
```

3.1. Series temporales / FORECASTING

Error Porcentual Absoluto Medio (MAPE)

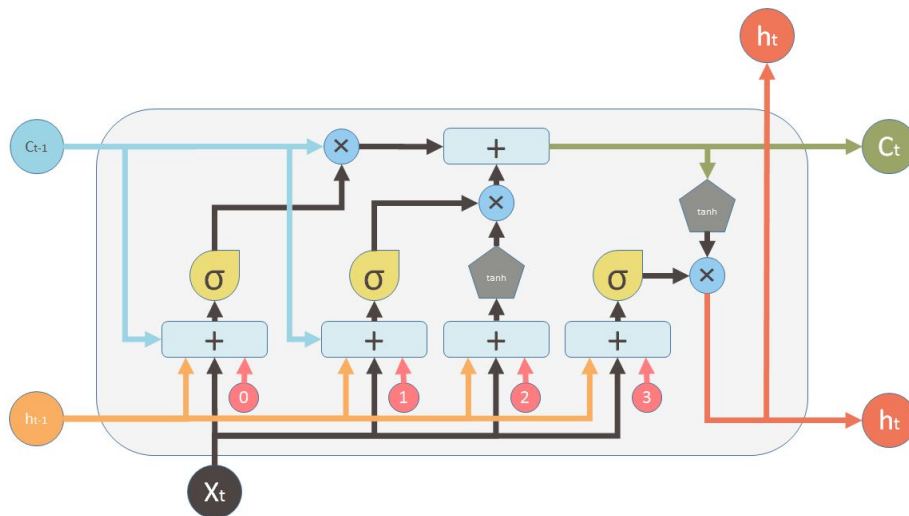
| | |
|-------------------------|------------------------|
| [1] CENTRO | -> 0.52375 |
| [2] ARGANZUELA | -> 0.4433095238095237 |
| [3] RETIRO | -> 0.689875 |
| [4] SALAMANCA | -> 1.0319444444444444 |
| [5] CHAMARTIN | -> 0.6550158730158732 |
| [6] TETUAN | -> 0.5826344246031746 |
| [7] CHAMBERI | -> 1.1007648809523811 |
| [8] FUENCARRAL-EL PARDO | -> 0.5401041666666666 |
| [9] MONCLOA-ARAVACA | -> 0.9356249999999999 |
| [10] LATINA | -> 0.896809523809524 |
| [11] CARABANCHEL | -> 1.3113571428571429 |
| [12] USERA | -> 0.13435744810744815 |
| [13] PUENTE DE VALLECAS | -> 0.3834573412698412 |
| [14] MORATALAZ | -> 0.4662688492063492 |
| [15] CIUDAD LINEAL | -> 0.3852149470899471 |
| [16] HORTALEZA | -> 0.30493483946608957 |
| [17] VILLAVERDE | -> 0.4789345735555294 |



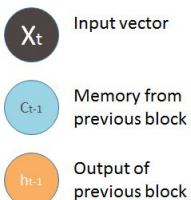
3.1. Recurrent neural network (RNN)

tensorflow.keras.preprocessing.sequence -
TimeseriesGenerator

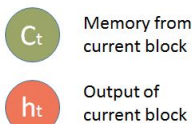
keras.layers - LSTM



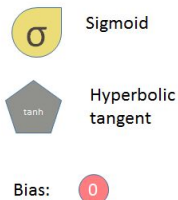
Inputs:



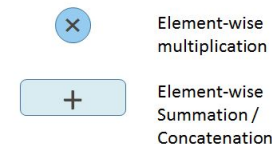
outputs:



Nonlinearities:



Vector operations:

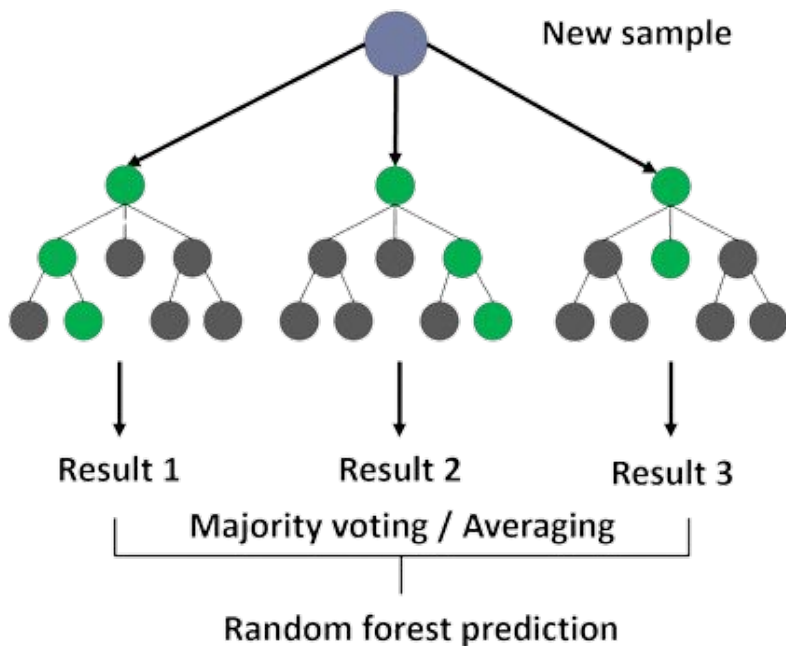


```
1/1 [=====] - 0s 103ms/step
[1] CENTRO -> 0.6505696475505829
1/1 [=====] - 0s 97ms/step
[2] ARGANZUELA -> 0.6782225966453552
1/1 [=====] - 0s 96ms/step
[3] RETIRO -> 0.513689786195755
1/1 [=====] - 0s 98ms/step
[4] SALAMANCA -> 0.6418256610631943
1/1 [=====] - 0s 95ms/step
[5] CHAMARTIN -> 0.6311056613922119
1/1 [=====] - 0s 97ms/step
[6] TETUAN -> 0.4956657886505127
1/1 [=====] - 0s 95ms/step
[7] CHAMBERI -> 0.3020361065864563
1/1 [=====] - 0s 96ms/step
[8] FUENCARRAL - EL PARDO -> 0.4421645204226176
1/1 [=====] - 0s 96ms/step
[9] MONCLOA - ARAVACA -> 0.4842960238456726
1/1 [=====] - 0s 102ms/step
[10] LATINA -> 0.45798752705256146
1/1 [=====] - 0s 97ms/step
[11] CARABANCHEL -> 0.4822224825620651
1/1 [=====] - 0s 95ms/step
[12] USERA -> 0.643595149119695
1/1 [=====] - 0s 97ms/step
[13] PUENTE DE VALLECAS -> 0.34868963062763214
1/1 [=====] - 0s 96ms/step
[14] MORATALAZ -> 0.6946385850509007
1/1 [=====] - 0s 97ms/step
[15] CIUDAD LINEAL -> 0.6941511452198028
1/1 [=====] - 0s 98ms/step
[16] HORTALEZA -> 0.6586569746335347
1/1 [=====] - 0s 96ms/step
[17] VILLAVERDE -> 0.3631201982498169
1/1 [=====] - 0s 101ms/step
[18] VILLA DE VALLECAS -> 0.5767396688461304
1/1 [=====] - 0s 96ms/step
[19] VICALVARO -> 0.5350796580314636
1/1 [=====] - 0s 99ms/step
[20] SAN BLAS -> 0.7242359439531962
1/1 [=====] - 0s 96ms/step
[21] BARAJAS -> 0.5648520886898041
```

3.2. Clasificación

sklearn.ensemble -
RandomForestClassifier

sklearn.tree -
DecisionTreeClassifier



```
Lesividad: HL Conteo: (102716, 80)
Lesividad: IL Conteo: (98387, 80)
Lesividad: HG Conteo: (8850, 80)
Lesividad: NO ASIGNADA Conteo: (11104, 80)
Lesividad: MT Conteo: (275, 80)
```

Matriz de Confusión

| Etiqueta verdadera | HG | HL | IL | MT | NO ASIGNADA |
|--------------------|------|-------|-------|----|-------------|
| | 301 | 1702 | 644 | 9 | 6 |
| | 1467 | 18894 | 10250 | 41 | 78 |
| | 705 | 13914 | 14723 | 10 | 290 |
| | 11 | 54 | 16 | 2 | 0 |
| MT | 16 | 239 | 337 | 0 | 2691 |
| NO ASIGNADA | | | | | |
| Predicción | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |

3.2. Clasificación

sklearn.ensemble -
RandomForestClassifier

sklearn.tree -
DecisionTreeClassifier

Matriz de Confusión

| Etiqueta verdadera | HG | HL | IL | MT |
|--------------------|----|----|----|----|
| HG | 8 | 15 | 1 | 17 |
| HL | 6 | 49 | 17 | 16 |
| IL | 2 | 28 | 6 | 2 |
| MT | 14 | 21 | 6 | 40 |
| | HG | HL | IL | MT |

Predicción

Matriz de Confusión

| Etiqueta verdadera | HG | HL | IL | MT | NO ASIGNADA |
|--------------------|------|-------|-------|----|-------------|
| HG | 301 | 1702 | 644 | 9 | 6 |
| HL | 1467 | 18894 | 10250 | 41 | 78 |
| IL | 705 | 13914 | 14723 | 10 | 290 |
| MT | 11 | 54 | 16 | 2 | 0 |
| NO ASIGNADA | 16 | 239 | 337 | 0 | 2691 |
| | HG | HL | IL | MT | NO ASIGNADA |

Predicción

3.1. Métrica elegida para evaluar los modelos

sklearn.metrics -
accuracy_score

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + FP + TN + FN)}$$

sklearn.metrics -
mean_absolute_percentage_error

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right|$$

sklearn.metrics -
confusion_matrix

| | | Actual Values | |
|------------------|--------------|---------------|--------------|
| | | Positive (1) | Negative (0) |
| Predicted Values | Positive (1) | TP | FP |
| | Negative (0) | FN | TN |

Nuestra propuesta

Modelo de IA capaz de “predecir” el número de accidentes en cierta zona con el objetivo de:

- Tenerlas vigiladas por medio aéreo.
- Conocer el número aproximado de implicados en el accidente.

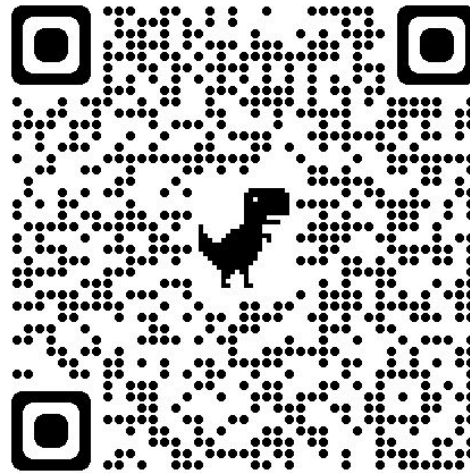


Github



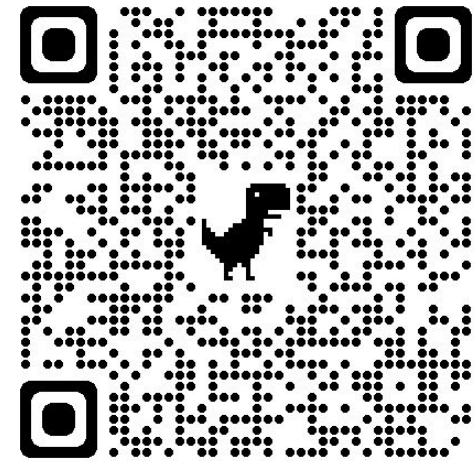
https://github.com/aprentix/IA_PROJECT

Tableau_nuevo



https://public.tableau.com/app/profile/raquel.d.az.ch.vez/viz/Accidentes_2019_2023/Dashboard1

Tableau_antiguo



https://public.tableau.com/views/Accidentes_trafico_2010_17/Dashboard1?:language=en-US&:display_count=n&:origin=viz_share_link

Fuentes de información

<https://www.un.org/sustainabledevelopment/es/2015/09/la-asamblea-general-adopta-la-agenda-2030-para-el-desarrollo-sostenible/>

<https://www.dgt.es/comunicacion/notas-de-prensa/1.145-personas-fallecieron-en-siniestros-de-traffic-durante-2022/>

<https://visualizadatos.madrid.es/pages/accidentes-de-traffic>

<https://www.youtube.com/watch?v=HXSzZQd1W-E>

– VIDEO CAMPAÑA DE PREVENCIÓN DE ACCIDENTES DE TRAFICO

Accidentes de tráfico en la Comunidad de Madrid - Modelos IA © 2023
by ALMANZA AMERICA JHASMINE, BOTEY BOPABOTE MARÍA ELENA,
DIAZ CHAVEZ RAQUEL is licensed under Attribution 4.0 International.
To view a copy of this license, visit
<http://creativecommons.org/licenses/by/4.0/>



SAMSUNG

Together for Tomorrow!
Enabling People

Education for Future Generations

©2023 SAMSUNG. All rights reserved.

Samsung Electronics Corporate Citizenship Office holds the copyright of book.

This book is a literary property protected by copyright law so reprint and reproduction without permission are prohibited.

To use this book other than the curriculum of Samsung Innovation Campus or to use the entire or part of this book, you must receive written consent from copyright holder.