

Calcular Distancias de Escuelas a establecimientos que venden alcohol mas cercanos

Chequen los PATHS

Primero se importan los scripts y scipy

```
In [ ]: from scripts.schools import main as compute_schools
        from scripts.alcohol import main as compute_alcohol
        from scripts.salud import main as compute_salud
        from scipy.spatial import KDTree
        import pandas as pd
```

Seccion de PATHS

```
In [ ]: # Lista de rutas de archivos Negocios DENGUE
        paths_negocios_dengue = [
            # AQUI AGREGUEN LOS PATHS DE SUS CSVs
            '/home/jay/repos/AI/feminicidios/dengue_00_46111_csv/conjunto_de_datos/de',
            '/home/jay/repos/AI/feminicidios/dengue_00_46112-46311_csv/conjunto_de_da',
            '/home/jay/repos/AI/feminicidios/dengue_00_46321-46531_csv/conjunto_de_da',
            '/home/jay/repos/AI/feminicidios/dengue_00_46591-46911_csv/conjunto_de_da',

            # Sin querer puse el PATH de las escuelas, xd que bueno que despues de g
            # Lo deje para debuggear
            # '/home/jay/repos/AI/feminicidios/dengue_00_61_csv/conjunto_de_datos/den
        ]

        # Ruta de archivo Escuelas DENGUE
        file_path_escuelas = '/home/jay/repos/AI/feminicidios/dengue_00_61_csv/conjun

        # Ruta de archivo salud DENGUE
        file_path_salud = '/home/jay/repos/GeoStats/dengue/dengue_00_62_csv/conjunto_d
```

Ahora se van a importar los DF de los negocios que venden alcohol y las escuelas

```
In [ ]: # Se itera sobre la lista de PATHS, la verdad no se por que imprime
        alcohol_establishments = compute_alcohol(paths_negocios_dengue)
```

```
In [ ]: # Se carga el DF dado un solo PATH  
schools = compute_schools(file_path_escuelas)
```

```
In [ ]: # Se carga el DF dado un solo PATH  
salud = compute_salud(file_path_salud)
```

```

      id      clee      nom_estab \
0      536  01003624191000011001000000U3  AA GRUPO CHICAGO 85
1  8625042  01001624191002591000000000U6  AA GRUPO LA SOLEDAD
2      16651  01001624191000421001000000U8  AA GRUPO PRIMAVERA
3  8463387  01001624211000123000000000U6  ABASTOS Y OPERACIONES
4  8984759  01001623311000112000000000U8  ABUELITOS FELICES

      raz_social  codigo_act \
0      NaN      624191
1  CENTRAL MEXICANA DE SERVICIOS GENERALES AA AC  624191
2      AA GRUPO PRIMAVERA AC  624191
3  ABASTOS Y OPERACIONES SA DE CV  624211
4  ABUELITOS FELICES S DE SS  623311

      nombre_act      per_ocu \
0  Agrupaciones de autoayuda para alcohólicos y p...  11 a 30 personas
1  Agrupaciones de autoayuda para alcohólicos y p...   0 a 5 personas
2  Agrupaciones de autoayuda para alcohólicos y p...   6 a 10 personas
3  Servicios de alimentación comunitarios prestad...  11 a 30 personas
4  Asilos y otras residencias del sector privado ...   6 a 10 personas

      tipo_vial      nom_vial  tipo_v_e_1  ... \
0      CALLE      ABASOLO      CALLE  ...
1      CALLE      ATOTONILCO      CALLE  ...
2      CALLE      ALAMO      CALLE  ...
3      CALLE      INDUSTRIA      AVENIDA  ...
4  PRIVADA  VALLE DE GUADALUPE      CALLE  ...

      localidad  ageb  manzana \
0  Calvillo      ...  0187      24
1      Aguascalientes  2846      15
2      Aguascalientes  0464      8
3      Aguascalientes  198A      4
4      Aguascalientes  2278      17

      telefono      correoelec  www  tipoUniEco      latitud \
0  4951077786      NaN  NaN      Fijo  21.845172
1      NaN      NaN  NaN      Fijo  21.877198
2      NaN      NaN  NaN      Fijo  21.893899
3      NaN  ABASTOSYOPERACION@GMAIL.COM  NaN      Fijo  21.905559
4      NaN      NaN  NaN      Fijo  21.910645

      longitud  fecha_alta
0  -102.722171  2010-07
1  -102.333976  2019-11
2  -102.295940  2010-07
3  -102.290243  2019-11
4  -102.304349  2019-11

```

[5 rows x 42 columns]

EDA

```
In [ ]: alcohol_establishments.head()
```

Out[]:

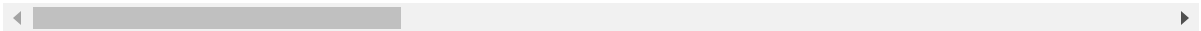
	id	nom_estab	nombre_act	latitud	longitud
0	8794778	7 ELEVEN	Comercio al por menor en tiendas de abarrotes,...	25.752659	-100.198830
1	9346815	7 ELEVEN MEXICO	Comercio al por menor en tiendas de abarrotes,...	25.685572	-100.474568
2	2962357	AB. ESMERALDA	Comercio al por menor en tiendas de abarrotes,...	25.733921	-100.386254
3	3019167	ABA FRUT LA ROSITA	Comercio al por menor en tiendas de abarrotes,...	25.685544	-100.410971
4	7350331	ABARROTES FLORES	Comercio al por menor en tiendas de abarrotes,...	25.754152	-100.106617

In []: schools.head()

Out[]:

	id	clec	nom_estab	raz_social	codigo
0	46594	01001611122002634000000000U4	2010 BICENTENARIO DE LA INDEPENDENCIA DE MEXIC...	INSTITUTO DE EDUCACION DE AGUASCALIENTES	611
1	8624796	01001713943003871000000000U7	ACADEMA BLACK DEMON	NaN	611
2	9113033	01001611621001461000000000U4	ACADEMIA DE ARTES MARCIALES	NaN	611
3	7461715	01011611611000011000000000U8	ACADEMIA DE BAILDE YL DANCE	NaN	611
4	45184	01001722519003621010000000U3	ACADEMIA DE BAILE BAILA CON LES	NaN	611

5 rows × 42 columns

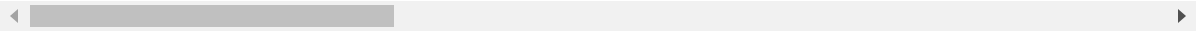


In []: salud.head()

Out[]:

		id	cle	nom_estab	raz_social	codigo_act	no
							agru
0	536	01003624191000011001000000U3		aa grupo chicago 85	NaN	624191	a alcc
1	8625042	01001624191002591000000000U6		aa grupo la soledad	central mexicana de servicios generales aa ac	624191	agru a alcc
2	16651	01001624191000421001000000U8		aa grupo primavera	aa grupo primavera ac	624191	agru a alcc
3	8463387	01001624211000123000000000U6		abastos y operaciones	abastos y operaciones sa de cv	624211	se alim com
4	8984759	01001623311000112000000000U8		abuelitos felices	abuelitos felices s de ss	623311	asile re (f

5 rows × 43 columns



```
In [ ]: alcohol_establishments.describe() # no aplica porque no hay datos numericos
schools.describe() # no aplica porque no hay datos numericos
salud.describe() # no aplica porque no hay datos numericos
```

Out[]:

		id	codigo_act	numero_ext	numero_int	cod_postal
count	2.194960e+05	219496.000000	172063.000000	77273.000000	217752.000000	2194
mean	4.849920e+06	621788.332530	706.760332	87.888072	54032.570364	
std	3.095145e+06	1172.698518	1750.596235	648.831276	26276.740640	
min	1.300000e+01	621111.000000	0.000000	0.000000	1000.000000	
25%	2.003884e+06	621113.000000	23.000000	0.000000	33800.000000	
50%	4.138574e+06	621211.000000	150.000000	0.000000	55555.000000	
75%	7.974451e+06	621398.000000	601.000000	6.000000	76000.000000	
max	9.470742e+06	624412.000000	78107.000000	92137.000000	99980.000000	

```

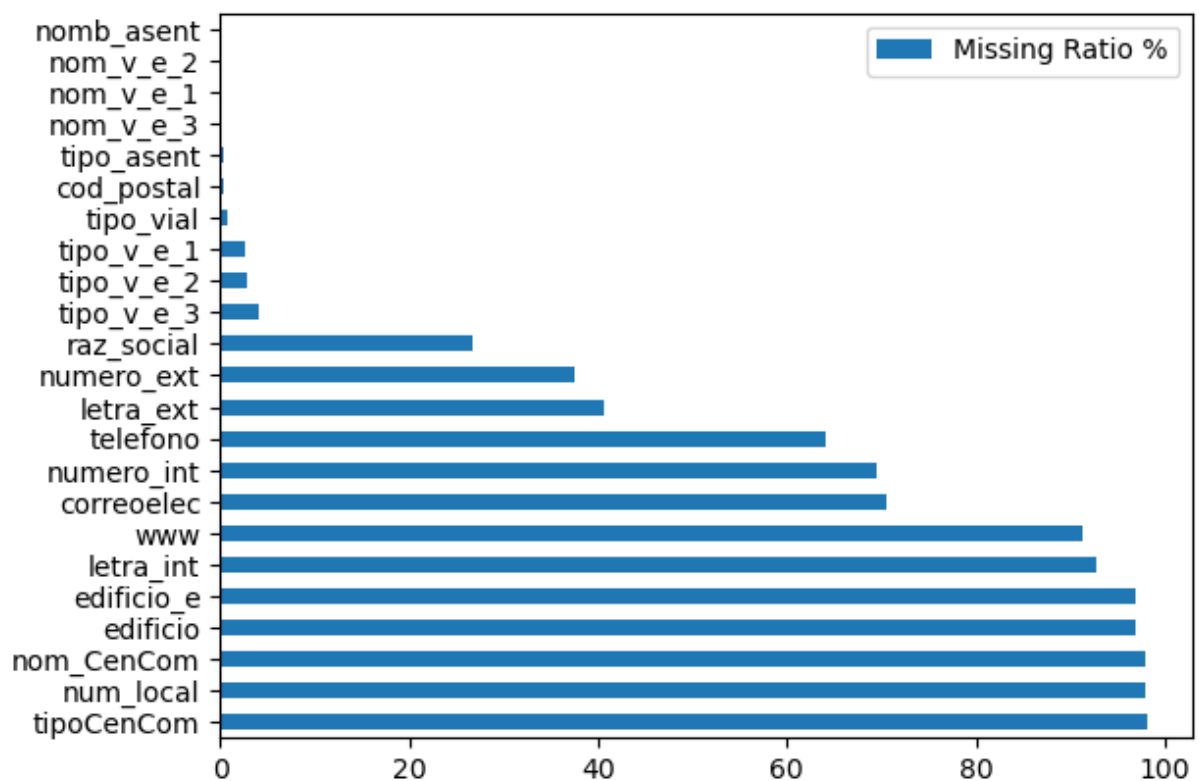
In [ ]: import matplotlib.pyplot as plt
def plot_nas(df: pd.DataFrame):
    if df.isnull().sum().sum() != 0:
        na_df = (df.isnull().sum() / len(df)) * 100
        na_df = na_df.drop(na_df[na_df == 0].index).sort_values(ascending=False)
        missing_data = pd.DataFrame({'Missing Ratio %' : na_df})
        missing_data.plot(kind = "barh")
        plt.show()
    else:
        print('No NAs found')

```

```

In [ ]: plot_nas(schools)

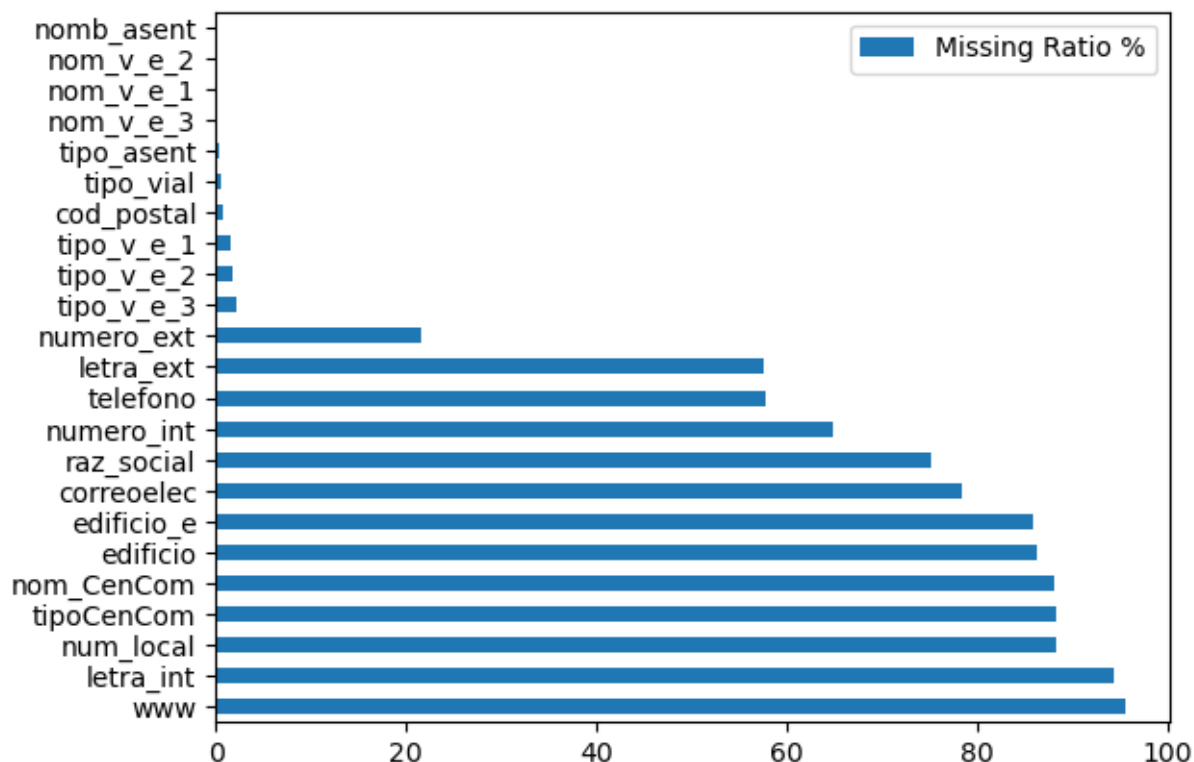
```



```
In [ ]: plot_nas(alcohol_establishments) # No NA's found
```

No NAs found

```
In [ ]: plot_nas(salud)
```



```
In [ ]: # Numero total de variables de salud por sector
salud['sector'].value_counts()
```

```
Out[ ]: privado      172613
publico      23843
otro         16694
hospital      6346
Name: sector, dtype: int64
```

Todavia no queda, esta en testing

Se crea un KDTree

Se guarda el indice y la distancia del establecimiento mas cercano a la escuela

```
In [ ]: # Encuentra el establecimiento más cercano a una coordenada dado un arbol de
def find_nearest(row, tree, establishments):
    distance, index = tree.query([row['latitud'], row['longitud']])
    return establishments.iloc[index]['nom_estab'], distance
```

```
# Tree para aplicar busqueda
alcohol_tree = KDTree(alcohol_establishments[['latitud', 'longitud']])

# simon
schools['nearest_alcohol'], schools['distance_to_nearest_alcohol'] = zip(*sc
```