Reto_INEGI_ML

December 4, 2022

1 Business Queries with the INEGI database

[1]:
```
import findspark # Activates Apache Spark, an engine for large-scale data␣
 ↪processing.
findspark.init()
import shapely # manipulation and analysis of planar geometric objects
import pandas as pd
import geopandas as gpd # working with geospatial data
from pyspark.sql import SparkSession
from pyspark import SparkContext
from pyspark import SparkConf
from sedona.register import SedonaRegistrator # Sedona is a cluster computing␣
 ↪system for processing large-scale spatial data
from sedona.utils import SedonaKryoRegistrator, KryoSerializer
```

[2]:
```
# Details of the Spark session
spark = SparkSession. \
builder. \
appName('appName'). \
config("spark.serializer", KryoSerializer.getName). \
config("spark.executor.memory", "5g"). \
config("spark.driver.memory", "10g"). \
config('spark.driver.maxResultSize', '5g'). \
config("spark.kryo.registrator", SedonaKryoRegistrator.getName). \
config('spark.jars.packages',
           'org.apache.sedona:sedona-python-adapter-3.0_2.12:1.2.
 ↪0-incubating,org.datasyslab:geotools-wrapper:1.1.0-25.2'). \
getOrCreate()
SedonaRegistrator.registerAll(spark)
```

[2]:
```
True
```

We will only use the information and data from Nuevo León

[3]:
```
# # Generate a readable database for Nuevo León's blocks
# # Just needs to be ran the first time

# import time
```

```
# start_time = time.time()

# path_scince = "../SCINCE 2020/"
# Estados_SCINCE = ['19_NL']

# for estado in Estados_SCINCE:
#     print(f"Procesando el estado: {estado}")
#     tabla_principal_poblacion = f"../SCINCE 2020/{estado}/cartografia/
 ↪municipal.shp"

#     TPP = gpd.read_file(tabla_principal_poblacion)
#     Pandas_TPP = pd.DataFrame(TPP)
#     Spark_TPP = spark.createDataFrame(Pandas_TPP)

#     tablas_secundarias =␣
 ↪['caracteristicas_economicas','discapacidad','educacion','etnicidad','fecundidad','hogares_

#     df = None
#     for tabla_sec in tablas_secundarias:
#         tabla_secundaria = f"../SCINCE 2020/{estado}/tablas/
 ↪cpv2020_municipal_{tabla_sec}.dbf"
#         SEC = gpd.read_file(tabla_secundaria)
#         Pandas_SEC = pd.DataFrame(SEC).drop(['geometry'], axis=1)
#         Spark_SEC = spark.createDataFrame(Pandas_SEC)
#         if df is None:
#             df = Spark_TPP.join(Spark_SEC, on=['CVEGEO'], how='inner')
#         else:
#             df = df.join(Spark_SEC, on=['CVEGEO'], how='inner')

#     print(f"Guardando el estado: {estado}")
#     df.write.parquet(f"../DB_NL_Reto_v2/{estado}.parquet")
#     print(f"Concluido estado: {estado}")

# print("--- %s segundos ---" % (time.time() - start_time))
```

[4]:
```
db_nl = spark.read.parquet(f"../DB_NL_Reto_v2/*.parquet")
db_nl.printSchema()
```

```
root
 |-- CVEGEO: string (nullable = true)
 |-- NOM_ENT: string (nullable = true)
 |-- NOMGEO: string (nullable = true)
 |-- POB1: double (nullable = true)
 |-- POB2: double (nullable = true)
 |-- POB2_R: double (nullable = true)
 |-- POB4: double (nullable = true)
```

```
|-- POB4_R: double (nullable = true)
|-- POB5: double (nullable = true)
|-- POB5_R: double (nullable = true)
|-- POB6: double (nullable = true)
|-- POB6_R: double (nullable = true)
|-- POB7: double (nullable = true)
|-- POB7_R: double (nullable = true)
|-- POB8: double (nullable = true)
|-- POB8_R: double (nullable = true)
|-- POB9: double (nullable = true)
|-- POB9_R: double (nullable = true)
|-- POB10: double (nullable = true)
|-- POB10_R: double (nullable = true)
|-- POB11: double (nullable = true)
|-- POB11_R: double (nullable = true)
|-- POB12: double (nullable = true)
|-- POB12_R: double (nullable = true)
|-- POB13: double (nullable = true)
|-- POB13_R: double (nullable = true)
|-- POB14: double (nullable = true)
|-- POB14_R: double (nullable = true)
|-- POB15: double (nullable = true)
|-- POB15_R: double (nullable = true)
|-- POB17: double (nullable = true)
|-- POB17_R: double (nullable = true)
|-- POB18: double (nullable = true)
|-- POB18_R: double (nullable = true)
|-- POB19: double (nullable = true)
|-- POB19_R: double (nullable = true)
|-- POB20: double (nullable = true)
|-- POB20_R: double (nullable = true)
|-- POB21: double (nullable = true)
|-- POB21_R: double (nullable = true)
|-- POB22: double (nullable = true)
|-- POB22_R: double (nullable = true)
|-- POB23: double (nullable = true)
|-- POB23_R: double (nullable = true)
|-- POB24: double (nullable = true)
|-- POB24_R: double (nullable = true)
|-- POB25: double (nullable = true)
|-- POB25_R: double (nullable = true)
|-- POB42: double (nullable = true)
|-- POB42_R: double (nullable = true)
|-- POB43: double (nullable = true)
|-- POB43_R: double (nullable = true)
|-- POB45: double (nullable = true)
|-- POB45_R: double (nullable = true)
|-- POB46: double (nullable = true)
```

```
|-- POB46_R: double (nullable = true)
|-- POB47: double (nullable = true)
|-- POB47_R: double (nullable = true)
|-- POB48: double (nullable = true)
|-- POB48_R: double (nullable = true)
|-- POB49: double (nullable = true)
|-- POB49_R: double (nullable = true)
|-- POB50: double (nullable = true)
|-- POB50_R: double (nullable = true)
|-- POB51: double (nullable = true)
|-- POB51_R: double (nullable = true)
|-- POB52: double (nullable = true)
|-- POB52_R: double (nullable = true)
|-- POB53: double (nullable = true)
|-- POB53_R: double (nullable = true)
|-- POB54: double (nullable = true)
|-- POB54_R: double (nullable = true)
|-- POB55: double (nullable = true)
|-- POB55_R: double (nullable = true)
|-- POB56: double (nullable = true)
|-- POB56_R: double (nullable = true)
|-- POB57: double (nullable = true)
|-- POB57_R: double (nullable = true)
|-- POB59: double (nullable = true)
|-- POB59_R: double (nullable = true)
|-- POB60: double (nullable = true)
|-- POB60_R: double (nullable = true)
|-- POB61: double (nullable = true)
|-- POB61_R: double (nullable = true)
|-- POB62: double (nullable = true)
|-- POB62_R: double (nullable = true)
|-- POB63: double (nullable = true)
|-- POB63_R: double (nullable = true)
|-- POB64: double (nullable = true)
|-- POB64_R: double (nullable = true)
|-- POB65: double (nullable = true)
|-- POB65_R: double (nullable = true)
|-- POB66: double (nullable = true)
|-- POB66_R: double (nullable = true)
|-- POB67: double (nullable = true)
|-- POB67_R: double (nullable = true)
|-- POB84: double (nullable = true)
|-- POB84_R: double (nullable = true)
|-- POB85: double (nullable = true)
|-- POB85_R: double (nullable = true)
|-- POB87: double (nullable = true)
|-- POB87_R: double (nullable = true)
|-- POB88: double (nullable = true)
```

```
|-- POB88_R: double (nullable = true)
|-- POB89: double (nullable = true)
|-- POB89_R: double (nullable = true)
|-- POB90: double (nullable = true)
|-- POB90_R: double (nullable = true)
|-- POB91: double (nullable = true)
|-- POB91_R: double (nullable = true)
|-- POB92: double (nullable = true)
|-- POB92_R: double (nullable = true)
|-- POB93: double (nullable = true)
|-- POB93_R: double (nullable = true)
|-- POB94: double (nullable = true)
|-- POB94_R: double (nullable = true)
|-- POB95: double (nullable = true)
|-- POB95_R: double (nullable = true)
|-- POB96: double (nullable = true)
|-- POB96_R: double (nullable = true)
|-- POB97: double (nullable = true)
|-- POB97_R: double (nullable = true)
|-- POB98: double (nullable = true)
|-- POB98_R: double (nullable = true)
|-- POB100: double (nullable = true)
|-- POB100_R: double (nullable = true)
|-- POB101: double (nullable = true)
|-- POB101_R: double (nullable = true)
|-- POB102: double (nullable = true)
|-- POB102_R: double (nullable = true)
|-- POB103: double (nullable = true)
|-- POB103_R: double (nullable = true)
|-- POB104: double (nullable = true)
|-- POB104_R: double (nullable = true)
|-- POB105: double (nullable = true)
|-- POB105_R: double (nullable = true)
|-- POB106: double (nullable = true)
|-- POB106_R: double (nullable = true)
|-- POB107: double (nullable = true)
|-- POB107_R: double (nullable = true)
|-- POB108: double (nullable = true)
|-- POB108_R: double (nullable = true)
|-- POB125_R: double (nullable = true)
|-- POB126_R: double (nullable = true)
|-- POB127_R: double (nullable = true)
|-- POB128_R: double (nullable = true)
|-- POB129_R: double (nullable = true)
|-- POB130_R: double (nullable = true)
|-- POB131_R: double (nullable = true)
|-- POB132_R: double (nullable = true)
|-- POB133_R: double (nullable = true)
```

```
|-- POB134_R: double (nullable = true)
|-- POB3: double (nullable = true)
|-- POB3_R: double (nullable = true)
|-- POB16: double (nullable = true)
|-- POB16_R: double (nullable = true)
|-- POB26: double (nullable = true)
|-- POB26_R: double (nullable = true)
|-- POB27: double (nullable = true)
|-- POB27_R: double (nullable = true)
|-- POB28: double (nullable = true)
|-- POB28_R: double (nullable = true)
|-- POB29: double (nullable = true)
|-- POB29_R: double (nullable = true)
|-- POB30: double (nullable = true)
|-- POB30_R: double (nullable = true)
|-- POB31: double (nullable = true)
|-- POB31_R: double (nullable = true)
|-- POB32: double (nullable = true)
|-- POB32_R: double (nullable = true)
|-- POB33: double (nullable = true)
|-- POB33_R: double (nullable = true)
|-- POB34: double (nullable = true)
|-- POB34_R: double (nullable = true)
|-- POB35: double (nullable = true)
|-- POB35_R: double (nullable = true)
|-- POB36: double (nullable = true)
|-- POB36_R: double (nullable = true)
|-- POB37: double (nullable = true)
|-- POB37_R: double (nullable = true)
|-- POB38: double (nullable = true)
|-- POB38_R: double (nullable = true)
|-- POB39: double (nullable = true)
|-- POB39_R: double (nullable = true)
|-- POB40: double (nullable = true)
|-- POB40_R: double (nullable = true)
|-- POB41: double (nullable = true)
|-- POB41_R: double (nullable = true)
|-- POB44: double (nullable = true)
|-- POB44_R: double (nullable = true)
|-- POB58: double (nullable = true)
|-- POB58_R: double (nullable = true)
|-- POB68: double (nullable = true)
|-- POB68_R: double (nullable = true)
|-- POB69: double (nullable = true)
|-- POB69_R: double (nullable = true)
|-- POB70: double (nullable = true)
|-- POB70_R: double (nullable = true)
|-- POB71: double (nullable = true)
```

```
|-- POB71_R: double (nullable = true)
|-- POB72: double (nullable = true)
|-- POB72_R: double (nullable = true)
|-- POB73: double (nullable = true)
|-- POB73_R: double (nullable = true)
|-- POB74: double (nullable = true)
|-- POB74_R: double (nullable = true)
|-- POB75: double (nullable = true)
|-- POB75_R: double (nullable = true)
|-- POB76: double (nullable = true)
|-- POB76_R: double (nullable = true)
|-- POB77: double (nullable = true)
|-- POB77_R: double (nullable = true)
|-- POB78: double (nullable = true)
|-- POB78_R: double (nullable = true)
|-- POB79: double (nullable = true)
|-- POB79_R: double (nullable = true)
|-- POB80: double (nullable = true)
|-- POB80_R: double (nullable = true)
|-- POB81: double (nullable = true)
|-- POB81_R: double (nullable = true)
|-- POB82: double (nullable = true)
|-- POB82_R: double (nullable = true)
|-- POB83: double (nullable = true)
|-- POB83_R: double (nullable = true)
|-- POB86: double (nullable = true)
|-- POB86_R: double (nullable = true)
|-- POB99: double (nullable = true)
|-- POB99_R: double (nullable = true)
|-- POB109: double (nullable = true)
|-- POB109_R: double (nullable = true)
|-- POB110: double (nullable = true)
|-- POB110_R: double (nullable = true)
|-- POB111: double (nullable = true)
|-- POB111_R: double (nullable = true)
|-- POB112: double (nullable = true)
|-- POB112_R: double (nullable = true)
|-- POB113: double (nullable = true)
|-- POB113_R: double (nullable = true)
|-- POB114: double (nullable = true)
|-- POB114_R: double (nullable = true)
|-- POB115: double (nullable = true)
|-- POB115_R: double (nullable = true)
|-- POB116: double (nullable = true)
|-- POB116_R: double (nullable = true)
|-- POB117: double (nullable = true)
|-- POB117_R: double (nullable = true)
|-- POB118: double (nullable = true)
```

```
|-- POB118_R: double (nullable = true)
|-- POB119: double (nullable = true)
|-- POB119_R: double (nullable = true)
|-- POB120: double (nullable = true)
|-- POB120_R: double (nullable = true)
|-- POB121: double (nullable = true)
|-- POB121_R: double (nullable = true)
|-- POB122: double (nullable = true)
|-- POB122_R: double (nullable = true)
|-- POB123: double (nullable = true)
|-- POB123_R: double (nullable = true)
|-- POB124: double (nullable = true)
|-- POB124_R: double (nullable = true)
|-- OID: long (nullable = true)
|-- geometry: geometry (nullable = true)
|-- ECO1: double (nullable = true)
|-- ECO1_R: double (nullable = true)
|-- ECO2: double (nullable = true)
|-- ECO2_R: double (nullable = true)
|-- ECO3: double (nullable = true)
|-- ECO3_R: double (nullable = true)
|-- ECO4: double (nullable = true)
|-- ECO4_R: double (nullable = true)
|-- ECO5: double (nullable = true)
|-- ECO5_R: double (nullable = true)
|-- ECO6: double (nullable = true)
|-- ECO6_R: double (nullable = true)
|-- ECO7: double (nullable = true)
|-- ECO7_R: double (nullable = true)
|-- ECO8: double (nullable = true)
|-- ECO8_R: double (nullable = true)
|-- ECO9: double (nullable = true)
|-- ECO9_R: double (nullable = true)
|-- ECO10: double (nullable = true)
|-- ECO10_R: double (nullable = true)
|-- ECO11: double (nullable = true)
|-- ECO11_R: double (nullable = true)
|-- ECO12: double (nullable = true)
|-- ECO12_R: double (nullable = true)
|-- ECO13: double (nullable = true)
|-- ECO13_R: double (nullable = true)
|-- ECO14: double (nullable = true)
|-- ECO14_R: double (nullable = true)
|-- ECO15: double (nullable = true)
|-- ECO15_R: double (nullable = true)
|-- ECO16: double (nullable = true)
|-- ECO16_R: double (nullable = true)
|-- ECO17: double (nullable = true)
```

```
|-- EC017_R: double (nullable = true)
|-- EC018: double (nullable = true)
|-- EC018_R: double (nullable = true)
|-- EC019: double (nullable = true)
|-- EC019_R: double (nullable = true)
|-- EC020: double (nullable = true)
|-- EC020_R: double (nullable = true)
|-- EC021: double (nullable = true)
|-- EC021_R: double (nullable = true)
|-- EC022: double (nullable = true)
|-- EC022_R: double (nullable = true)
|-- EC023: double (nullable = true)
|-- EC023_R: double (nullable = true)
|-- EC024: double (nullable = true)
|-- EC024_R: double (nullable = true)
|-- EC025: double (nullable = true)
|-- EC025_R: double (nullable = true)
|-- EC026: double (nullable = true)
|-- EC026_R: double (nullable = true)
|-- EC027: double (nullable = true)
|-- EC027_R: double (nullable = true)
|-- EC028: double (nullable = true)
|-- EC028_R: double (nullable = true)
|-- EC029: double (nullable = true)
|-- EC029_R: double (nullable = true)
|-- EC030: double (nullable = true)
|-- EC030_R: double (nullable = true)
|-- EC031: double (nullable = true)
|-- EC031_R: double (nullable = true)
|-- EC032: double (nullable = true)
|-- EC032_R: double (nullable = true)
|-- EC033: double (nullable = true)
|-- EC033_R: double (nullable = true)
|-- EC034: double (nullable = true)
|-- EC034_R: double (nullable = true)
|-- EC035: double (nullable = true)
|-- EC035_R: double (nullable = true)
|-- EC036: double (nullable = true)
|-- EC036_R: double (nullable = true)
|-- EC037: double (nullable = true)
|-- EC037_R: double (nullable = true)
|-- EC038: double (nullable = true)
|-- EC038_R: double (nullable = true)
|-- EC039: double (nullable = true)
|-- EC039_R: double (nullable = true)
|-- EC040: double (nullable = true)
|-- EC040_R: double (nullable = true)
|-- EC041: double (nullable = true)
```

```
|-- ECO41_R: double (nullable = true)
|-- ECO42: double (nullable = true)
|-- ECO42_R: double (nullable = true)
|-- ECO43: double (nullable = true)
|-- ECO43_R: double (nullable = true)
|-- ECO44: double (nullable = true)
|-- ECO44_R: double (nullable = true)
|-- ECO45: double (nullable = true)
|-- ECO45_R: double (nullable = true)
|-- DISC1: double (nullable = true)
|-- DISC1_R: double (nullable = true)
|-- DISC2: double (nullable = true)
|-- DISC2_R: double (nullable = true)
|-- DISC3: double (nullable = true)
|-- DISC3_R: double (nullable = true)
|-- DISC4: double (nullable = true)
|-- DISC4_R: double (nullable = true)
|-- DISC5: double (nullable = true)
|-- DISC5_R: double (nullable = true)
|-- DISC6: double (nullable = true)
|-- DISC6_R: double (nullable = true)
|-- DISC7: double (nullable = true)
|-- DISC7_R: double (nullable = true)
|-- DISC8: double (nullable = true)
|-- DISC8_R: double (nullable = true)
|-- DISC9: double (nullable = true)
|-- DISC9_R: double (nullable = true)
|-- DISC10: double (nullable = true)
|-- DISC10_R: double (nullable = true)
|-- DISC11: double (nullable = true)
|-- DISC11_R: double (nullable = true)
|-- DISC12: double (nullable = true)
|-- DISC12_R: double (nullable = true)
|-- DISC13: double (nullable = true)
|-- DISC13_R: double (nullable = true)
|-- DISC14: double (nullable = true)
|-- DISC14_R: double (nullable = true)
|-- DISC15: double (nullable = true)
|-- DISC15_R: double (nullable = true)
|-- DISC16: double (nullable = true)
|-- DISC16_R: double (nullable = true)
|-- DISC17: double (nullable = true)
|-- DISC17_R: double (nullable = true)
|-- DISC18: double (nullable = true)
|-- DISC18_R: double (nullable = true)
|-- DISC19: double (nullable = true)
|-- DISC19_R: double (nullable = true)
|-- DISC20: double (nullable = true)
```

```
|-- DISC20_R: double (nullable = true)
|-- DISC21: double (nullable = true)
|-- DISC21_R: double (nullable = true)
|-- DISC22: double (nullable = true)
|-- DISC22_R: double (nullable = true)
|-- DISC23: double (nullable = true)
|-- DISC23_R: double (nullable = true)
|-- DISC24: double (nullable = true)
|-- DISC24_R: double (nullable = true)
|-- DISC25: double (nullable = true)
|-- DISC25_R: double (nullable = true)
|-- DISC26: double (nullable = true)
|-- DISC26_R: double (nullable = true)
|-- EDU1: double (nullable = true)
|-- EDU1_R: double (nullable = true)
|-- EDU2: double (nullable = true)
|-- EDU2_R: double (nullable = true)
|-- EDU3: double (nullable = true)
|-- EDU3_R: double (nullable = true)
|-- EDU4: double (nullable = true)
|-- EDU4_R: double (nullable = true)
|-- EDU5: double (nullable = true)
|-- EDU5_R: double (nullable = true)
|-- EDU6: double (nullable = true)
|-- EDU6_R: double (nullable = true)
|-- EDU7: double (nullable = true)
|-- EDU7_R: double (nullable = true)
|-- EDU8: double (nullable = true)
|-- EDU8_R: double (nullable = true)
|-- EDU9: double (nullable = true)
|-- EDU9_R: double (nullable = true)
|-- EDU10: double (nullable = true)
|-- EDU10_R: double (nullable = true)
|-- EDU11: double (nullable = true)
|-- EDU11_R: double (nullable = true)
|-- EDU12: double (nullable = true)
|-- EDU12_R: double (nullable = true)
|-- EDU13: double (nullable = true)
|-- EDU13_R: double (nullable = true)
|-- EDU14: double (nullable = true)
|-- EDU14_R: double (nullable = true)
|-- EDU15: double (nullable = true)
|-- EDU15_R: double (nullable = true)
|-- EDU16: double (nullable = true)
|-- EDU16_R: double (nullable = true)
|-- EDU17: double (nullable = true)
|-- EDU17_R: double (nullable = true)
|-- EDU18: double (nullable = true)
```

```
|-- EDU18_R: double (nullable = true)
|-- EDU19: double (nullable = true)
|-- EDU19_R: double (nullable = true)
|-- EDU20: double (nullable = true)
|-- EDU20_R: double (nullable = true)
|-- EDU21: double (nullable = true)
|-- EDU21_R: double (nullable = true)
|-- EDU22: double (nullable = true)
|-- EDU22_R: double (nullable = true)
|-- EDU23: double (nullable = true)
|-- EDU23_R: double (nullable = true)
|-- EDU24: double (nullable = true)
|-- EDU24_R: double (nullable = true)
|-- EDU25: double (nullable = true)
|-- EDU25_R: double (nullable = true)
|-- EDU26: double (nullable = true)
|-- EDU26_R: double (nullable = true)
|-- EDU27: double (nullable = true)
|-- EDU27_R: double (nullable = true)
|-- EDU28: double (nullable = true)
|-- EDU28_R: double (nullable = true)
|-- EDU29: double (nullable = true)
|-- EDU29_R: double (nullable = true)
|-- EDU30: double (nullable = true)
|-- EDU30_R: double (nullable = true)
|-- EDU31: double (nullable = true)
|-- EDU31_R: double (nullable = true)
|-- EDU32: double (nullable = true)
|-- EDU32_R: double (nullable = true)
|-- EDU33: double (nullable = true)
|-- EDU33_R: double (nullable = true)
|-- EDU34: double (nullable = true)
|-- EDU34_R: double (nullable = true)
|-- EDU35: double (nullable = true)
|-- EDU35_R: double (nullable = true)
|-- EDU36: double (nullable = true)
|-- EDU36_R: double (nullable = true)
|-- EDU37: double (nullable = true)
|-- EDU37_R: double (nullable = true)
|-- EDU38: double (nullable = true)
|-- EDU38_R: double (nullable = true)
|-- EDU39: double (nullable = true)
|-- EDU39_R: double (nullable = true)
|-- EDU40: double (nullable = true)
|-- EDU40_R: double (nullable = true)
|-- EDU41: double (nullable = true)
|-- EDU41_R: double (nullable = true)
|-- EDU42: double (nullable = true)
```

```
|-- EDU42_R: double (nullable = true)
|-- EDU43: double (nullable = true)
|-- EDU43_R: double (nullable = true)
|-- EDU44: double (nullable = true)
|-- EDU44_R: double (nullable = true)
|-- EDU45: double (nullable = true)
|-- EDU45_R: double (nullable = true)
|-- EDU46: double (nullable = true)
|-- EDU46_R: double (nullable = true)
|-- EDU47: double (nullable = true)
|-- EDU47_R: double (nullable = true)
|-- EDU48: double (nullable = true)
|-- EDU48_R: double (nullable = true)
|-- EDU49_R: double (nullable = true)
|-- EDU50_R: double (nullable = true)
|-- EDU51_R: double (nullable = true)
|-- INDI1: double (nullable = true)
|-- INDI1_R: double (nullable = true)
|-- INDI2: double (nullable = true)
|-- INDI2_R: double (nullable = true)
|-- INDI3: double (nullable = true)
|-- INDI3_R: double (nullable = true)
|-- INDI4: double (nullable = true)
|-- INDI4_R: double (nullable = true)
|-- INDI5: double (nullable = true)
|-- INDI5_R: double (nullable = true)
|-- INDI6: double (nullable = true)
|-- INDI6_R: double (nullable = true)
|-- INDI7: double (nullable = true)
|-- INDI7_R: double (nullable = true)
|-- INDI8: double (nullable = true)
|-- INDI8_R: double (nullable = true)
|-- INDI9: double (nullable = true)
|-- INDI9_R: double (nullable = true)
|-- INDI10: double (nullable = true)
|-- INDI10_R: double (nullable = true)
|-- INDI11: double (nullable = true)
|-- INDI11_R: double (nullable = true)
|-- INDI12: double (nullable = true)
|-- INDI12_R: double (nullable = true)
|-- INDI13: double (nullable = true)
|-- INDI13_R: double (nullable = true)
|-- INDI14: double (nullable = true)
|-- INDI14_R: double (nullable = true)
|-- INDI15: double (nullable = true)
|-- INDI15_R: double (nullable = true)
|-- INDI16: double (nullable = true)
|-- INDI16_R: double (nullable = true)
```

```
|-- INDI17: double (nullable = true)
|-- INDI17_R: double (nullable = true)
|-- INDI18: double (nullable = true)
|-- INDI18_R: double (nullable = true)
|-- INDI19: double (nullable = true)
|-- INDI19_R: double (nullable = true)
|-- INDI20: double (nullable = true)
|-- INDI20_R: double (nullable = true)
|-- afro1: double (nullable = true)
|-- afro2: double (nullable = true)
|-- afro3: double (nullable = true)
|-- afro4: double (nullable = true)
|-- afro5: double (nullable = true)
|-- afro6: double (nullable = true)
|-- afro1_r: double (nullable = true)
|-- afro2_r: double (nullable = true)
|-- afro3_r: double (nullable = true)
|-- afro4_r: double (nullable = true)
|-- afro5_r: double (nullable = true)
|-- afro6_r: double (nullable = true)
|-- FEC1_R: double (nullable = true)
|-- FEC2_R: double (nullable = true)
|-- FEC3_R: double (nullable = true)
|-- HOGAR1: double (nullable = true)
|-- HOGAR2: double (nullable = true)
|-- HOGAR2_R: double (nullable = true)
|-- HOGAR3: double (nullable = true)
|-- HOGAR3_R: double (nullable = true)
|-- HOGAR4: double (nullable = true)
|-- HOGAR5: double (nullable = true)
|-- HOGAR5_R: double (nullable = true)
|-- HOGAR6: double (nullable = true)
|-- HOGAR6_R: double (nullable = true)
|-- HOGAR7: double (nullable = true)
|-- HOGAR7_R: double (nullable = true)
|-- HOGAR8: double (nullable = true)
|-- HOGAR8_R: double (nullable = true)
|-- HOGAR9: double (nullable = true)
|-- HOGAR9_R: double (nullable = true)
|-- HOGAR10: double (nullable = true)
|-- HOGAR10_R: double (nullable = true)
|-- HOGAR11: double (nullable = true)
|-- HOGAR11_R: double (nullable = true)
|-- HOGAR12: double (nullable = true)
|-- HOGAR12_R: double (nullable = true)
|-- HOGAR13: double (nullable = true)
|-- HOGAR13_R: double (nullable = true)
|-- HOGAR14: double (nullable = true)
```

```
|-- HOGAR14_R: double (nullable = true)
|-- HOGAR15: double (nullable = true)
|-- HOGAR15_R: double (nullable = true)
|-- HOGAR16: double (nullable = true)
|-- HOGAR16_R: double (nullable = true)
|-- HOGAR17: double (nullable = true)
|-- HOGAR17_R: double (nullable = true)
|-- HOGAR18: double (nullable = true)
|-- HOGAR18_R: double (nullable = true)
|-- HOGAR19: double (nullable = true)
|-- HOGAR19_R: double (nullable = true)
|-- HOGAR20: double (nullable = true)
|-- HOGAR20_R: double (nullable = true)
|-- HOGAR21: double (nullable = true)
|-- HOGAR21_R: double (nullable = true)
|-- HOGAR22: double (nullable = true)
|-- HOGAR22_R: double (nullable = true)
|-- HOGAR23: double (nullable = true)
|-- HOGAR23_R: double (nullable = true)
|-- HOGAR24: double (nullable = true)
|-- HOGAR24_R: double (nullable = true)
|-- HOGAR25: double (nullable = true)
|-- HOGAR25_R: double (nullable = true)
|-- HOGAR26: double (nullable = true)
|-- HOGAR26_R: double (nullable = true)
|-- MIG1: double (nullable = true)
|-- MIG1_R: double (nullable = true)
|-- MIG2: double (nullable = true)
|-- MIG2_R: double (nullable = true)
|-- MIG3: double (nullable = true)
|-- MIG3_R: double (nullable = true)
|-- MIG4: double (nullable = true)
|-- MIG4_R: double (nullable = true)
|-- MIG5: double (nullable = true)
|-- MIG5_R: double (nullable = true)
|-- MIG6: double (nullable = true)
|-- MIG6_R: double (nullable = true)
|-- MIG7: double (nullable = true)
|-- MIG7_R: double (nullable = true)
|-- MIG8: double (nullable = true)
|-- MIG8_R: double (nullable = true)
|-- MIG9: double (nullable = true)
|-- MIG9_R: double (nullable = true)
|-- MIG10: double (nullable = true)
|-- MIG10_R: double (nullable = true)
|-- MIG11: double (nullable = true)
|-- MIG11_R: double (nullable = true)
|-- MIG12: double (nullable = true)
```

```
|-- MIG12_R: double (nullable = true)
|-- MIG13: double (nullable = true)
|-- MIG13_R: double (nullable = true)
|-- MIG14: double (nullable = true)
|-- MIG14_R: double (nullable = true)
|-- MIG15: double (nullable = true)
|-- MIG15_R: double (nullable = true)
|-- MIG16_R: double (nullable = true)
|-- MOR1_R: double (nullable = true)
|-- RELIG1: double (nullable = true)
|-- RELIG1_R: double (nullable = true)
|-- RELIG2: double (nullable = true)
|-- RELIG2_R: double (nullable = true)
|-- RELIG3: double (nullable = true)
|-- RELIG3_R: double (nullable = true)
|-- RELIG4: double (nullable = true)
|-- RELIG4_R: double (nullable = true)
|-- SALUD1: double (nullable = true)
|-- SALUD1_R: double (nullable = true)
|-- SALUD2: double (nullable = true)
|-- SALUD2_R: double (nullable = true)
|-- SALUD3: double (nullable = true)
|-- SALUD3_R: double (nullable = true)
|-- SALUD4: double (nullable = true)
|-- SALUD4_R: double (nullable = true)
|-- SALUD5: double (nullable = true)
|-- SALUD5_R: double (nullable = true)
|-- SALUD6: double (nullable = true)
|-- SALUD6_R: double (nullable = true)
|-- SALUD7: double (nullable = true)
|-- SALUD7_R: double (nullable = true)
|-- SALUD8: double (nullable = true)
|-- SALUD8_R: double (nullable = true)
|-- SALUD9: double (nullable = true)
|-- SALUD9_R: double (nullable = true)
|-- SALUD10: double (nullable = true)
|-- SALUD10_R: double (nullable = true)
|-- SCONY1: double (nullable = true)
|-- SCONY1_R: double (nullable = true)
|-- SCONY2: double (nullable = true)
|-- SCONY2_R: double (nullable = true)
|-- SCONY3: double (nullable = true)
|-- SCONY3_R: double (nullable = true)
|-- SCONY4: double (nullable = true)
|-- SCONY4_R: double (nullable = true)
|-- SCONY5: double (nullable = true)
|-- SCONY5_R: double (nullable = true)
|-- SCONY6: double (nullable = true)
```

```
|-- SCONY6_R: double (nullable = true)
|-- SCONY7: double (nullable = true)
|-- SCONY7_R: double (nullable = true)
|-- SCONY8: double (nullable = true)
|-- SCONY8_R: double (nullable = true)
|-- SCONY9: double (nullable = true)
|-- SCONY9_R: double (nullable = true)
|-- SCONY10: double (nullable = true)
|-- SCONY10_R: double (nullable = true)
|-- SCONY11: double (nullable = true)
|-- SCONY11_R: double (nullable = true)
|-- SCONY12: double (nullable = true)
|-- SCONY12_R: double (nullable = true)
|-- VIV0: double (nullable = true)
|-- VIV1: double (nullable = true)
|-- VIV2: double (nullable = true)
|-- VIV2_R: double (nullable = true)
|-- VIV3: double (nullable = true)
|-- VIV3_R: double (nullable = true)
|-- VIV4: double (nullable = true)
|-- VIV4_R: double (nullable = true)
|-- VIV5: double (nullable = true)
|-- VIV5_R: double (nullable = true)
|-- VIV6: double (nullable = true)
|-- VIV6_R: double (nullable = true)
|-- VIV7: double (nullable = true)
|-- VIV7_R: double (nullable = true)
|-- VIV8: double (nullable = true)
|-- VIV8_R: double (nullable = true)
|-- VIV9: double (nullable = true)
|-- VIV9_R: double (nullable = true)
|-- VIV10: double (nullable = true)
|-- VIV10_R: double (nullable = true)
|-- VIV11: double (nullable = true)
|-- VIV11_R: double (nullable = true)
|-- VIV12: double (nullable = true)
|-- VIV12_R: double (nullable = true)
|-- VIV13: double (nullable = true)
|-- VIV13_R: double (nullable = true)
|-- VIV14: double (nullable = true)
|-- VIV14_R: double (nullable = true)
|-- VIV15: double (nullable = true)
|-- VIV15_R: double (nullable = true)
|-- VIV16: double (nullable = true)
|-- VIV16_R: double (nullable = true)
|-- VIV17: double (nullable = true)
|-- VIV17_R: double (nullable = true)
|-- VIV18: double (nullable = true)
```

```
|-- VIV18_R: double (nullable = true)
|-- VIV19: double (nullable = true)
|-- VIV19_R: double (nullable = true)
|-- VIV20: double (nullable = true)
|-- VIV20_R: double (nullable = true)
|-- VIV21: double (nullable = true)
|-- VIV21_R: double (nullable = true)
|-- VIV22: double (nullable = true)
|-- VIV22_R: double (nullable = true)
|-- VIV23: double (nullable = true)
|-- VIV23_R: double (nullable = true)
|-- VIV24: double (nullable = true)
|-- VIV24_R: double (nullable = true)
|-- VIV25: double (nullable = true)
|-- VIV25_R: double (nullable = true)
|-- VIV26: double (nullable = true)
|-- VIV26_R: double (nullable = true)
|-- VIV27: double (nullable = true)
|-- VIV27_R: double (nullable = true)
|-- VIV28: double (nullable = true)
|-- VIV28_R: double (nullable = true)
|-- VIV29: double (nullable = true)
|-- VIV29_R: double (nullable = true)
|-- VIV30: double (nullable = true)
|-- VIV30_R: double (nullable = true)
|-- VIV31: double (nullable = true)
|-- VIV31_R: double (nullable = true)
|-- VIV32: double (nullable = true)
|-- VIV32_R: double (nullable = true)
|-- VIV33: double (nullable = true)
|-- VIV33_R: double (nullable = true)
|-- VIV34: double (nullable = true)
|-- VIV34_R: double (nullable = true)
|-- VIV35: double (nullable = true)
|-- VIV35_R: double (nullable = true)
|-- VIV36: double (nullable = true)
|-- VIV36_R: double (nullable = true)
|-- VIV37: double (nullable = true)
|-- VIV37_R: double (nullable = true)
|-- VIV38: double (nullable = true)
|-- VIV38_R: double (nullable = true)
|-- VIV39: double (nullable = true)
|-- VIV39_R: double (nullable = true)
|-- VIV40: double (nullable = true)
|-- VIV40_R: double (nullable = true)
|-- VIV41: double (nullable = true)
|-- VIV41_R: double (nullable = true)
|-- VIV42: double (nullable = true)
```

```
|-- VIV42_R: double (nullable = true)
|-- VIV43: double (nullable = true)
|-- VIV43_R: double (nullable = true)
|-- VIV70: double (nullable = true)
|-- VIV70_R: double (nullable = true)
|-- VIV75: double (nullable = true)
|-- VIV75_R: double (nullable = true)
|-- VIV76: double (nullable = true)
|-- VIV76_R: double (nullable = true)
|-- VIV77: double (nullable = true)
|-- VIV77_R: double (nullable = true)
|-- VIV78: double (nullable = true)
|-- VIV78_R: double (nullable = true)
|-- VIV79: double (nullable = true)
|-- VIV79_R: double (nullable = true)
|-- VIV80: double (nullable = true)
|-- VIV80_R: double (nullable = true)
|-- VIV81: double (nullable = true)
|-- VIV81_R: double (nullable = true)
|-- VIV82: double (nullable = true)
|-- VIV82_R: double (nullable = true)
|-- VIV83: double (nullable = true)
|-- VIV83_R: double (nullable = true)
|-- VIV84: double (nullable = true)
|-- VIV84_R: double (nullable = true)
|-- VIV90: double (nullable = true)
|-- VIV90_R: double (nullable = true)
|-- VIV91: double (nullable = true)
|-- VIV91_R: double (nullable = true)
|-- VIV92_R: double (nullable = true)
|-- VIV93_R: double (nullable = true)
|-- VIV94: double (nullable = true)
```

[5]:
```
# # Generate a readable database for Nuevo León's businesses
# # Just needs to be ran the first time

# start_time = time.time()

# HX = gpd.read_file("../Mex-Hex-5k/Mex-Hex-5k.shp")

# Estados_DENUE = ["19"]
# for estado in Estados_DENUE:

#     print(f"Procesando el estado: {estado}")
#     if estado in ["15_1","15_2"] :
```

```
#         tabla_denue = f"../DENUE_05_2022/denue_{estado}_shp/conjunto_de_datos/
 ↪denue_inegi_{estado}.shp"
#     else:
#         tabla_denue = f"../DENUE_05_2022/denue_{estado}_shp/conjunto_de_datos/
 ↪denue_inegi_{estado}_.shp"

#     TD = gpd.read_file(tabla_denue)
#     TD = TD.set_crs('epsg:4326', allow_override=True)
#     Spark_TD = spark.createDataFrame(TD.to_crs(HX.crs))

#     print(f"Guardando el estado: {estado}")
#     Spark_TD.write.parquet(f"../DB_NL_DENUE_Reto_v2/{estado}.parquet")
#     print(f"Concluido estado: {estado}")

# print("--- %s segundos ---" % (time.time() - start_time))
```

[6]:
```
db_nl_denue = spark.read.parquet(f"../DB_NL_DENUE_Reto_v1/*.parquet")
db_nl_denue.printSchema()
```

```
root
 |-- id: long (nullable = true)
 |-- clee: string (nullable = true)
 |-- nom_estab: string (nullable = true)
 |-- raz_social: string (nullable = true)
 |-- codigo_act: string (nullable = true)
 |-- nombre_act: string (nullable = true)
 |-- per_ocu: string (nullable = true)
 |-- tipo_vial: string (nullable = true)
 |-- nom_vial: string (nullable = true)
 |-- tipo_v_e_1: string (nullable = true)
 |-- nom_v_e_1: string (nullable = true)
 |-- tipo_v_e_2: string (nullable = true)
 |-- nom_v_e_2: string (nullable = true)
 |-- tipo_v_e_3: string (nullable = true)
 |-- nom_v_e_3: string (nullable = true)
 |-- numero_ext: string (nullable = true)
 |-- letra_ext: string (nullable = true)
 |-- edificio: string (nullable = true)
 |-- edificio_e: string (nullable = true)
 |-- numero_int: string (nullable = true)
 |-- letra_int: string (nullable = true)
 |-- tipo_asent: string (nullable = true)
 |-- nomb_asent: string (nullable = true)
 |-- tipoCenCom: string (nullable = true)
 |-- nom_CenCom: string (nullable = true)
 |-- num_local: string (nullable = true)
 |-- cod_postal: string (nullable = true)
 |-- cve_ent: string (nullable = true)
```

```
 |-- entidad: string (nullable = true)
 |-- cve_mun: string (nullable = true)
 |-- municipio: string (nullable = true)
 |-- cve_loc: string (nullable = true)
 |-- localidad: string (nullable = true)
 |-- ageb: string (nullable = true)
 |-- manzana: string (nullable = true)
 |-- telefono: string (nullable = true)
 |-- correoelec: string (nullable = true)
 |-- www: string (nullable = true)
 |-- tipoUniEco: string (nullable = true)
 |-- latitud: double (nullable = true)
 |-- longitud: double (nullable = true)
 |-- fecha_alta: string (nullable = true)
 |-- geometry: geometry (nullable = true)
```

[7]:
```
#list of the SCINCE variables that will be used for analysis
var_viv = ["CVEGEO", "VIV0", "VIV1", "VIV7", "VIV8", "VIV9", "VIV10", "VIV14",␣
 ↪"VIV16", "VIV25", "VIV26", "VIV27", "VIV28", "VIV29", "VIV31", "VIV32",␣
 ↪"VIV33", "VIV34", "VIV38", "VIV39", "VIV40", "VIV41", "VIV42", "VIV81",␣
 ↪"VIV37_R", "VIV13_R", "VIV80_R","VIV81_R", "VIV82_R","VIV83_R", "VIV84_R",␣
 ↪"HOGAR1", "HOGAR4", "HOGAR7", "HOGAR13", "HOGAR16","HOGAR19", "HOGAR22"]
var_pob = ["POB11_R", "POB17_R", "POB42_R", "POB84_R", "MIG7_R","INDI1_R",␣
 ↪"DISC1_R","EDU49_R","SCONY1_R", "SCONY4_R", "SCONY7_R", "RELIG1_R",␣
 ↪"RELIG2_R","EDU25", "EDU28", "SALUD1", "SALUD2", "SALUD7", "SALUD8"]
var_ec = ["ECO1_R", "ECO28_R", "ECO34_R"]
#var_geom=["geometry"]
```

[8]:
```
# building the demographic database (SCINCE)
db_nl_scince = db_nl.select(var_viv)
db_nl_scince = db_nl_scince.join(db_nl.select(var_pob))
db_nl_scince = db_nl_scince.join(db_nl.select(var_ec))
#db_nl_scince = db_nl_scince.join(db_nl.select(var_geom))
```

[9]:
```
# Creates a temporary table that allows to run SQL queries
db_nl_scince.createOrReplaceTempView("db_nl_scince")
db_nl_scince.printSchema()
```

```
root
 |-- CVEGEO: string (nullable = true)
 |-- VIV0: double (nullable = true)
 |-- VIV1: double (nullable = true)
 |-- VIV7: double (nullable = true)
 |-- VIV8: double (nullable = true)
 |-- VIV9: double (nullable = true)
 |-- VIV10: double (nullable = true)
 |-- VIV14: double (nullable = true)
```

```
|-- VIV16: double (nullable = true)
|-- VIV25: double (nullable = true)
|-- VIV26: double (nullable = true)
|-- VIV27: double (nullable = true)
|-- VIV28: double (nullable = true)
|-- VIV29: double (nullable = true)
|-- VIV31: double (nullable = true)
|-- VIV32: double (nullable = true)
|-- VIV33: double (nullable = true)
|-- VIV34: double (nullable = true)
|-- VIV38: double (nullable = true)
|-- VIV39: double (nullable = true)
|-- VIV40: double (nullable = true)
|-- VIV41: double (nullable = true)
|-- VIV42: double (nullable = true)
|-- VIV81: double (nullable = true)
|-- VIV37_R: double (nullable = true)
|-- VIV13_R: double (nullable = true)
|-- VIV80_R: double (nullable = true)
|-- VIV81_R: double (nullable = true)
|-- VIV82_R: double (nullable = true)
|-- VIV83_R: double (nullable = true)
|-- VIV84_R: double (nullable = true)
|-- HOGAR1: double (nullable = true)
|-- HOGAR4: double (nullable = true)
|-- HOGAR7: double (nullable = true)
|-- HOGAR13: double (nullable = true)
|-- HOGAR16: double (nullable = true)
|-- HOGAR19: double (nullable = true)
|-- HOGAR22: double (nullable = true)
|-- POB11_R: double (nullable = true)
|-- POB17_R: double (nullable = true)
|-- POB42_R: double (nullable = true)
|-- POB84_R: double (nullable = true)
|-- MIG7_R: double (nullable = true)
|-- INDI1_R: double (nullable = true)
|-- DISC1_R: double (nullable = true)
|-- EDU49_R: double (nullable = true)
|-- SCONY1_R: double (nullable = true)
|-- SCONY4_R: double (nullable = true)
|-- SCONY7_R: double (nullable = true)
|-- RELIG1_R: double (nullable = true)
|-- RELIG2_R: double (nullable = true)
|-- EDU25: double (nullable = true)
|-- EDU28: double (nullable = true)
|-- SALUD1: double (nullable = true)
|-- SALUD2: double (nullable = true)
|-- SALUD7: double (nullable = true)
```

```
 |-- SALUD8: double (nullable = true)
 |-- ECO1_R: double (nullable = true)
 |-- ECO28_R: double (nullable = true)
 |-- ECO34_R: double (nullable = true)
```

[10]:
```
# Select DENUE variables that will be used for analysis
var_denue=["codigo_act","nom_estab","cod_postal","cve_mun","municipio","latitud","longitud","m
db_nl_denue=db_nl_denue.select(var_denue)
```

[11]:
```
db_nl_denue.createOrReplaceTempView("db_nl_denue")
db_nl_denue.printSchema()
```

```
root
 |-- codigo_act: string (nullable = true)
 |-- nom_estab: string (nullable = true)
 |-- cod_postal: string (nullable = true)
 |-- cve_mun: string (nullable = true)
 |-- municipio: string (nullable = true)
 |-- latitud: double (nullable = true)
 |-- longitud: double (nullable = true)
 |-- manzana: string (nullable = true)
 |-- geometry: geometry (nullable = true)
```

[12]:
```
db_nl_scince.show(5)
```

```
+------+------+------+----+------+------+------+-----+-----+------+-----+------+
------+------+-----+------+------+------+-----+-----+------+-----+-----+------+-
------+-------+-------+-------+-------+-------+-------+------+-------+------+---
----+-------+-------+-------+-------+-------+-------+-------+------+-------+----
---+-------+--------+--------+--------+--------+--------+-------+-----+-------+-
-----+------+------+------+-------+--------+
|CVEGEO|  VIV0|  VIV1|VIV7|  VIV8|  VIV9| VIV10|VIV14|VIV16| VIV25|VIV26| VIV27|
VIV28| VIV29|VIV31| VIV32| VIV33| VIV34|VIV38|VIV39| VIV40|VIV41|VIV42|
VIV81|VIV37_R|VIV13_R|VIV80_R|VIV81_R|VIV82_R|VIV83_R|VIV84_R|HOGAR1| HOGAR4|HOG
AR7|HOGAR13|HOGAR16|HOGAR19|HOGAR22|POB11_R|POB17_R|POB42_R|POB84_R|MIG7_R|INDI1
_R|DISC1_R|EDU49_R|SCONY1_R|SCONY4_R|SCONY7_R|RELIG1_R|RELIG2_R|  EDU25|EDU28|
SALUD1|SALUD2|SALUD7|SALUD8|ECO1_R|ECO28_R|ECO34_R|
+------+------+------+----+------+------+------+-----+-----+------+-----+------+
------+------+-----+------+------+------+-----+-----+------+-----+-----+------+-
------+-------+-------+-------+-------+-------+-------+------+-------+------+---
----+-------+-------+-------+-------+-------+-------+-------+------+-------+----
---+-------+--------+--------+--------+--------+--------+-------+-----+-------+-
-----+------+------+------+-------+--------+
| 19005|8119.0|5468.0|69.0|1941.0|3513.0|1434.0|178.0| 96.0|5146.0|
38.0|5125.0|4563.0|3813.0|160.0|3631.0|5001.0|1330.0|224.0|304.0|3037.0| 54.0|
26.0|1402.0|   38.9|   79.5|    3.9|   25.7|   59.2|    9.0|
6.8|5468.0|18030.0| 534.0| 3195.0|11896.0| 1739.0| 4309.0|   22.1|   94.4|
```

```
50.9|   49.1|   3.3|    0.1|    6.9|    8.6|    30.9|    55.5|    13.5|    71.3|
15.0|12579.0|476.0|15386.0|2626.0| 196.0| 130.0|  55.6|   44.1|   28.7|
| 19005|8119.0|5468.0|69.0|1941.0|3513.0|1434.0|178.0| 96.0|5146.0|
38.0|5125.0|4563.0|3813.0|160.0|3631.0|5001.0|1330.0|224.0|304.0|3037.0| 54.0|
26.0|1402.0|   38.9|   79.5|    3.9|   25.7|   59.2|    9.0|
6.8|5468.0|18030.0| 534.0| 3195.0|11896.0| 1739.0| 4309.0|   22.1|   94.4|
50.9|   49.1|   3.3|    0.1|    6.9|    8.6|    30.9|    55.5|    13.5|    71.3|
15.0|12579.0|476.0|15386.0|2626.0| 196.0| 130.0|  46.0|   53.8|   18.6|
| 19005|8119.0|5468.0|69.0|1941.0|3513.0|1434.0|178.0| 96.0|5146.0|
38.0|5125.0|4563.0|3813.0|160.0|3631.0|5001.0|1330.0|224.0|304.0|3037.0| 54.0|
26.0|1402.0|   38.9|   79.5|    3.9|   25.7|   59.2|    9.0|
6.8|5468.0|18030.0| 534.0| 3195.0|11896.0| 1739.0| 4309.0|   22.1|   94.4|
50.9|   49.1|   3.3|    0.1|    6.9|    8.6|    30.9|    55.5|    13.5|    71.3|
15.0|12579.0|476.0|15386.0|2626.0| 196.0| 130.0|  48.6|   51.1|   30.0|
| 19005|8119.0|5468.0|69.0|1941.0|3513.0|1434.0|178.0| 96.0|5146.0|
38.0|5125.0|4563.0|3813.0|160.0|3631.0|5001.0|1330.0|224.0|304.0|3037.0| 54.0|
26.0|1402.0|   38.9|   79.5|    3.9|   25.7|   59.2|    9.0|
6.8|5468.0|18030.0| 534.0| 3195.0|11896.0| 1739.0| 4309.0|   22.1|   94.4|
50.9|   49.1|   3.3|    0.1|    6.9|    8.6|    30.9|    55.5|    13.5|    71.3|
15.0|12579.0|476.0|15386.0|2626.0| 196.0| 130.0|  60.1|   39.5|   26.2|
| 19005|8119.0|5468.0|69.0|1941.0|3513.0|1434.0|178.0| 96.0|5146.0|
38.0|5125.0|4563.0|3813.0|160.0|3631.0|5001.0|1330.0|224.0|304.0|3037.0| 54.0|
26.0|1402.0|   38.9|   79.5|    3.9|   25.7|   59.2|    9.0|
6.8|5468.0|18030.0| 534.0| 3195.0|11896.0| 1739.0| 4309.0|   22.1|   94.4|
50.9|   49.1|   3.3|    0.1|    6.9|    8.6|    30.9|    55.5|    13.5|    71.3|
15.0|12579.0|476.0|15386.0|2626.0| 196.0| 130.0|  64.4|   34.9|   41.9|
+------+------+------+----+------+------+------+-----+-----+------+-----+------+
------+------+-----+------+------+------+-----+-----+------+-----+-----+------+-
------+-------+-------+-------+-------+-------+-------+------+-------+------+---
----+-------+-------+-------+-------+-------+-------+-------+------+-------+----
---+-------+--------+--------+--------+--------+--------+-------+-----+-------+-
-----+------+------+------+-------+-------+
only showing top 5 rows
```

[13]:
```
db_nl_scince.count()
```

[13]: 132651

[14]:
```
db_nl_denue.show(5)
```

```
+----------+--------------------+----------+-------+--------------------+-------
----+-------------+-------+--------------------+
|codigo_act|           nom_estab|cod_postal|cve_mun|           municipio|
latitud|     longitud|manzana|            geometry|
+----------+--------------------+----------+-------+--------------------+-------
----+-------------+-------+--------------------+
|    236221|      OFFICE CONCEPT|     66278|    039|           Monterrey|
```

```
25.6378307| -100.3172524|    026|POINT (2668173.42…|
|    237121|             OFICINA|     64000|    039|
Monterrey|25.68124975|-100.32135603|    001|POINT (2667707.29…|
|    236221|             ODEVISA|     64460|    039|
Monterrey|25.70120271|-100.34094675|    007|POINT (2665724.69…|
|    236111|    ODICSA, SA DE CV|     64750|    019|San Pedro Garza
G…|25.65107342|-100.37352233|    044|POINT (2662533.50…|
|    236111|OFICINA ADMINISTR…|     66196|    048|      Santa
Catarina|25.65742631|-100.43900735|    063|POINT (2655982.24…|
+----------+--------------------+----------+-------+--------------------+-------
----+-------------+-------+--------------------+
only showing top 5 rows
```

[15]:
```
db_nl_scince.select("CVEGEO").distinct().collect()
```

[15]:
```
[Row(CVEGEO='19051'),
 Row(CVEGEO='19005'),
 Row(CVEGEO='19041'),
 Row(CVEGEO='19002'),
 Row(CVEGEO='19022'),
 Row(CVEGEO='19045'),
 Row(CVEGEO='19050'),
 Row(CVEGEO='19044'),
 Row(CVEGEO='19015'),
 Row(CVEGEO='19027'),
 Row(CVEGEO='19030'),
 Row(CVEGEO='19001'),
 Row(CVEGEO='19020'),
 Row(CVEGEO='19040'),
 Row(CVEGEO='19018'),
 Row(CVEGEO='19047'),
 Row(CVEGEO='19042'),
 Row(CVEGEO='19014'),
 Row(CVEGEO='19036'),
 Row(CVEGEO='19016'),
 Row(CVEGEO='19049'),
 Row(CVEGEO='19048'),
 Row(CVEGEO='19010'),
 Row(CVEGEO='19017'),
 Row(CVEGEO='19004'),
 Row(CVEGEO='19038'),
 Row(CVEGEO='19006'),
 Row(CVEGEO='19008'),
 Row(CVEGEO='19003'),
 Row(CVEGEO='19029'),
 Row(CVEGEO='19024'),
```

```
 Row(CVEGEO='19031'),
 Row(CVEGEO='19033'),
 Row(CVEGEO='19025'),
 Row(CVEGEO='19026'),
 Row(CVEGEO='19037'),
 Row(CVEGEO='19019'),
 Row(CVEGEO='19009'),
 Row(CVEGEO='19011'),
 Row(CVEGEO='19032'),
 Row(CVEGEO='19013'),
 Row(CVEGEO='19046'),
 Row(CVEGEO='19034'),
 Row(CVEGEO='19028'),
 Row(CVEGEO='19035'),
 Row(CVEGEO='19012'),
 Row(CVEGEO='19021'),
 Row(CVEGEO='19043'),
 Row(CVEGEO='19007'),
 Row(CVEGEO='19023'),
 Row(CVEGEO='19039')]
```

[16]:
```
db_nl_denue.select("cve_mun").distinct().collect()
```

[16]:
```
[Row(cve_mun='030'),
 Row(cve_mun='009'),
 Row(cve_mun='032'),
 Row(cve_mun='048'),
 Row(cve_mun='028'),
 Row(cve_mun='012'),
 Row(cve_mun='027'),
 Row(cve_mun='037'),
 Row(cve_mun='013'),
 Row(cve_mun='024'),
 Row(cve_mun='031'),
 Row(cve_mun='034'),
 Row(cve_mun='015'),
 Row(cve_mun='006'),
 Row(cve_mun='019'),
 Row(cve_mun='046'),
 Row(cve_mun='020'),
 Row(cve_mun='033'),
 Row(cve_mun='047'),
 Row(cve_mun='040'),
 Row(cve_mun='011'),
 Row(cve_mun='025'),
 Row(cve_mun='042'),
 Row(cve_mun='044'),
```

```
 Row(cve_mun='051'),
 Row(cve_mun='045'),
 Row(cve_mun='005'),
 Row(cve_mun='003'),
 Row(cve_mun='038'),
 Row(cve_mun='035'),
 Row(cve_mun='049'),
 Row(cve_mun='016'),
 Row(cve_mun='018'),
 Row(cve_mun='029'),
 Row(cve_mun='008'),
 Row(cve_mun='022'),
 Row(cve_mun='001'),
 Row(cve_mun='014'),
 Row(cve_mun='010'),
 Row(cve_mun='039'),
 Row(cve_mun='023'),
 Row(cve_mun='036'),
 Row(cve_mun='004'),
 Row(cve_mun='050'),
 Row(cve_mun='041'),
 Row(cve_mun='017'),
 Row(cve_mun='026'),
 Row(cve_mun='007'),
 Row(cve_mun='021'),
 Row(cve_mun='043'),
 Row(cve_mun='002')]
```

[17]:
```
# path_ZIP = "../CodigosPostales/"

# print(f"Procesando los códigos postales:")
# tabla_principal_zips = f"../CodigosPostales/CodigosPostales_4326.shp"

# TPZ = gpd.read_file(tabla_principal_zips)
# Pandas_TPZ = pd.DataFrame(TPZ)
# Spark_TPZ = spark.createDataFrame(Pandas_TPZ)
# df = Spark_TPZ

# print(f"Guardando los códigos:")
# df.write.parquet(f"../DB_CodigosPostales/zips.parquet")
# print(f"Concluido")

# print("--- %s segundos ---" % (time.time() - start_time))
```

[18]:
```
db_zips = spark.read.parquet(f"../DB_CodigosPostales/*.parquet")
db_zips.printSchema()
```

```
root
```

```
 |-- d_cp: string (nullable = true)
 |-- geometry: geometry (nullable = true)
```

[19]:
```
db_zips.show(10)
db_zips.createOrReplaceTempView("db_zips")
```

```
+-----+--------------------+
| d_cp|            geometry|
+-----+--------------------+
|69015|POLYGON ((-97.617…|
|69018|POLYGON ((-97.770…|
|69014|POLYGON ((-97.760…|
|69012|POLYGON ((-97.827…|
|69020|POLYGON ((-97.766…|
|69021|POLYGON ((-97.769…|
|69023|POLYGON ((-97.786…|
|69022|POLYGON ((-97.788…|
|69017|POLYGON ((-97.696…|
|69013|POLYGON ((-97.682…|
+-----+--------------------+
only showing top 10 rows
```

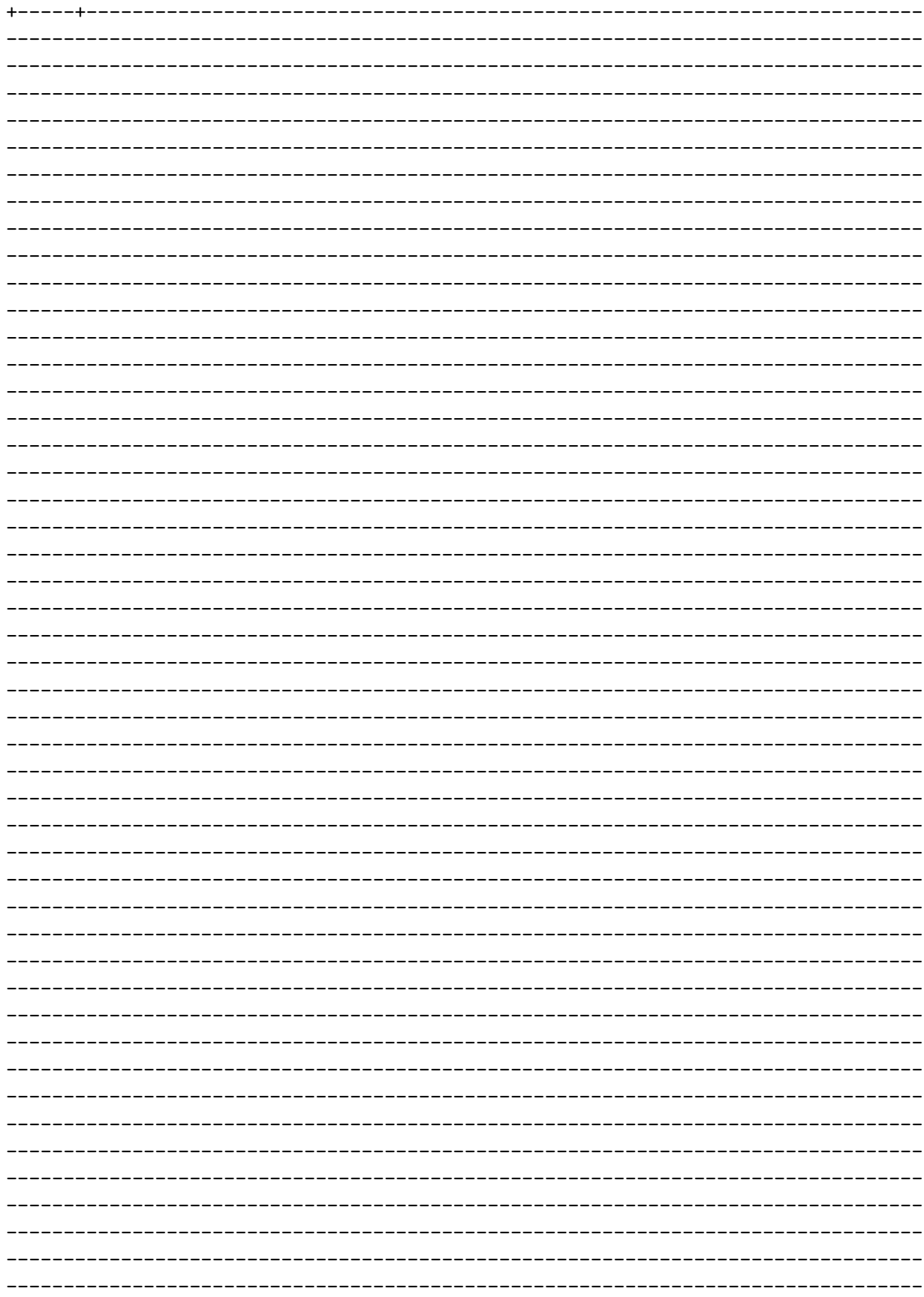
[20]:
```
db_zips.count()
```

[20]: 36089

[21]:
```
cp_nl = spark.sql("""SELECT *
                            FROM db_zips
                            WHERE db_zips.d_cp = "64800"
                            """)
cp_nl.show(truncate=False)
```

```
+-----+----------------------------------------------------------------------------
----------------------------------------------------------------------------------
----------------------------------------------------------------------------------
----------------------------------------------------------------------------------
----------------------------------------------------------------------------------
----------------------------------------------------------------------------------
----------------------------------------------------------------------------------
----------------------------------------------------------------------------------
----------------------------------------------------------------------------------
----------------------------------------------------------------------------------
----------------------------------------------------------------------------------
----------------------------------------------------------------------------------
----------------------------------------------------------------------------------
----------------------------------------------------------------------------------
----------------------------------------------------------------------------------
----------------------------------------------------------------------------------
```

```
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
------------------------------------------------------------------------+
|d_cp |geometry
|
```



```
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-------------------------------------------------------------------------+
|64800|POLYGON ((-100.29232130006838 25.670140432780073, -100.28652597888365
25.672468047424776, -100.28620925132132 25.672588366760166, -100.28588963350454
25.672702235858438, -100.28556728438565 25.672809596560047, -100.28524236583756
25.672910397087353, -100.28491503973514 25.673004585664057, -100.28458547088533
25.673092115988883, -100.28425382307204 25.673172943564804, -100.28392026402383
25.673247028474027, -100.28358495844468 25.67331433258122, -100.28324807597143
25.673374823248135, -100.28290978420502 25.67342847053687, -100.27906158155444
25.674046644884598, -100.2788561059348 25.67409967147423, -100.27865283911693
25.674159278889928, -100.27845203729511 25.674225392827097, -100.27825395379494
25.674297928973804, -100.27805883902585 25.674376796637773, -100.27786693851523
25.67446189600416, -100.2776784948986 25.674553119064136, -100.27749374689466
25.674650351417803, -100.27731292633919 25.674753469520475, -100.27713626315288
25.67486234345802, -100.2769639783363 25.674976836870677, -100.2767962910231
25.67509680340107, -100.27663341239142 25.675222092977734, -100.27647554872657
25.67535254731262, -100.27632289740545 25.675488000764886, -100.27617565286566
25.675628283126446, -100.27603400063667 25.675773216835953, -100.27589812030615
25.675922619732276, -100.27576818155276 25.67607630228989, -100.27564434910188
25.676234071301, -100.27552678076034 25.67639572713294, -100.27430187418415
25.676389194363914, -100.27172196316026 25.67643720895784, -100.27387041785889
25.674651095847008, -100.27366320829555 25.67433265048858, -100.27366700144766
25.674332378026516, -100.27397055937752 25.6741828582858, -100.27415441559873
25.673961861324848, -100.2742366652307 25.673772801825987, -100.27421754918603
25.673509687293773, -100.27403827330463 25.673009912383097, -100.2737584935483
25.672778175801902, -100.27348494254498 25.67267527445084, -100.27282800112646
25.67267522273633, -100.27280272115591 25.671513703852245, -100.2727890249417
25.670923751287933, -100.27163658818021 25.67093289112762, -100.27141581947716
25.67045417150306, -100.27138208506817 25.67026912233472, -100.27137986210057
25.669834235509043, -100.27138683274912 25.669255210966828, -100.27138199833531
25.668659894397734, -100.27137593370401 25.668120653695055, -100.27136492312829
25.667521861319436, -100.27136237803902 25.667186452316184, -100.27258347249317
25.66744264250293, -100.27270878031058 25.66746910491146, -100.27270640741038
25.66762046478845, -100.27335805738544 25.667746728936862, -100.27348591955307
25.66775626481586, -100.27368916678803 25.66771768189572, -100.2739689003202
25.667560923669964, -100.27440655105762 25.667289499071558, -100.27472339673734
25.66722944627438, -100.27675360717407 25.667148301952054, -100.27695921257195
```

25.667032556424836, -100.27842065037288 25.666489892810635, -100.27862541495507
25.66651459537545, -100.2790123332173 25.666569548985265, -100.27872537445047
25.665619774144478, -100.27899047200083 25.665413026702726, -100.28108519150014
25.66509068551372, -100.28125025173517 25.66508946060657, -100.28131306766488
25.66546374588378, -100.28173103961637 25.665396287301533, -100.2820120241827
25.66689891110446, -100.2811837802471 25.6668050403624, -100.28123284770015
25.667438850310813, -100.28127390220376 25.667564985921743, -100.28238909618095
25.667702664996266, -100.28228574647132 25.666922846374835, -100.28431996563059
25.667157947793743, -100.28422392035695 25.66539167084589, -100.28415039354427
25.664979733024055, -100.28232510925119 25.660224670178444, -100.28216097521708
25.659880417822983, -100.28211925962911 25.65975822745802, -100.28207153014264
25.65963785846449, -100.28201788162596 25.65951955495954, -100.28195842403998
25.65940355487439, -100.28189327539754 25.65929009259896, -100.28182256878002
25.659179398151217, -100.28174644738169 25.659071693494973, -100.28166506343715
25.658967197970988, -100.2815785823197 25.658866121084813, -100.28096501019783
25.65808263912514, -100.28116898095077 25.658008304328302, -100.28308182071993
25.657823596227125, -100.2830906748416 25.657142531342743, -100.28320256310833
25.65578494630157, -100.28410456344345 25.65587237314529, -100.28402584396125
25.65886968697686, -100.28591614329865 25.65709022118164, -100.2863266008565
25.657440108831075, -100.2862750174832 25.66008593296078, -100.28661304440494
25.659936277653017, -100.2873919922562 25.659916169522578, -100.2872040073978
25.662846377072825, -100.29008143898872 25.663039175544252, -100.29003338034823
25.664122851694998, -100.28993616676071 25.664325098895183, -100.28974078061401
25.664731589074346, -100.28964548185144 25.666167463486868, -100.29021616239092
25.666147065580542, -100.29195318107566 25.665992680655684, -100.29194659253984
25.666972069267906, -100.29212081034059 25.66697043114855, -100.29226698720254
25.669019173043658, -100.29232130006838 25.670140432780073))|
+-----+--


```
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
-------------------------------------------------------------------------+
```

[22]:
```python
from pyspark.sql.functions import desc
db_nl_scince_zip=db_nl_scince.join(db_zips, on=["geometry"], how= "left")
db_nl_scince_zip.show()
```

```
---------------------------------------------------------------------------
```

```
AnalysisException                         Traceback (most recent call last)
Cell In [22], line 2
      1 from pyspark.sql.functions import desc
----> 2 db_nl_scince_zip=db_nl_scince.join(db_zips, on=["geometry"], how= "left" )
      3 db_nl_scince_zip.show()

File C:\spark-3.2.2-bin-hadoop2.7\python\pyspark\sql\dataframe.py:1355, in␣
 ↪DataFrame.join(self, other, on, how)
   1353         on = self._jseq([])
   1354     assert isinstance(how, str), "how should be a string"
-> 1355     jdf = self._jdf.join(other._jdf, on, how)
   1356 return DataFrame(jdf, self.sql_ctx)

File C:\spark-3.2.2-bin-hadoop2.7\python\lib\py4j-0.10.9.5-src.
 ↪zip\py4j\java_gateway.py:1321, in JavaMember.__call__(self, *args)
   1315 command = proto.CALL_COMMAND_NAME +\
   1316     self.command_header +\
   1317     args_command +\
   1318     proto.END_COMMAND_PART
   1320 answer = self.gateway_client.send_command(command)
-> 1321 return_value = get_return_value(
   1322     answer, self.gateway_client, self.target_id, self.name)
   1324 for temp_arg in temp_args:
   1325     temp_arg._detach()

File C:\spark-3.2.2-bin-hadoop2.7\python\pyspark\sql\utils.py:117, in␣
 ↪capture_sql_exception.<locals>.deco(*a, **kw)
    113 converted = convert_exception(e.java_exception)
    114 if not isinstance(converted, UnknownException):
    115     # Hide where the exception came from that shows a non-Pythonic
    116     # JVM exception message.
--> 117     raise converted from None
    118 else:
    119     raise

AnalysisException: USING column `geometry` cannot be resolved on the left side␣
 ↪of the join. The left-side columns: [CVEGEO, VIV0, VIV1, VIV7, VIV8, VIV9,␣
 ↪VIV10, VIV14, VIV16, VIV25, VIV26, VIV27, VIV28, VIV29, VIV31, VIV32, VIV33,␣
 ↪VIV34, VIV38, VIV39, VIV40, VIV41, VIV42, VIV81, VIV37_R, VIV13_R, VIV80_R,␣
 ↪VIV81_R, VIV82_R, VIV83_R, VIV84_R, HOGAR1, HOGAR4, HOGAR7, HOGAR13, HOGAR16,␣
 ↪HOGAR19, HOGAR22, POB11_R, POB17_R, POB42_R, POB84_R, MIG7_R, INDI1_R,␣
 ↪DISC1_R, EDU49_R, SCONY1_R, SCONY4_R, SCONY7_R, RELIG1_R, RELIG2_R, EDU25,␣
 ↪EDU28, SALUD1, SALUD2, SALUD7, SALUD8, ECO1_R, ECO28_R, ECO34_R]
```

[ ]:
```
db_nl_scince_zip.show(truncate=False)
```

[ ]:
```
db_nl_scince.join(db_zips, on=["geometry"], how= "inner").show()
```

1.0.1 We generate new variables from the DENUE database.

[23]:
```
df_denue=db_nl_denue.toPandas()
```

[24]:
```
df_scince=db_nl_scince.toPandas()

#db_nl_scince.write.csv("df_scince_prueba")
```

[25]:
```
df_scince.head()
```

[25]:
```
   CVEGEO    VIV0    VIV1  VIV7    VIV8    VIV9   VIV10  VIV14  VIV16   VIV25  \
0  19005  8119.0  5468.0  69.0  1941.0  3513.0  1434.0  178.0   96.0  5146.0
1  19005  8119.0  5468.0  69.0  1941.0  3513.0  1434.0  178.0   96.0  5146.0
2  19005  8119.0  5468.0  69.0  1941.0  3513.0  1434.0  178.0   96.0  5146.0
3  19005  8119.0  5468.0  69.0  1941.0  3513.0  1434.0  178.0   96.0  5146.0
4  19005  8119.0  5468.0  69.0  1941.0  3513.0  1434.0  178.0   96.0  5146.0

   …  RELIG2_R    EDU25  EDU28   SALUD1  SALUD2  SALUD7  SALUD8  ECO1_R  \
0  …      15.0  12579.0  476.0  15386.0  2626.0   196.0   130.0    55.6
1  …      15.0  12579.0  476.0  15386.0  2626.0   196.0   130.0    46.0
2  …      15.0  12579.0  476.0  15386.0  2626.0   196.0   130.0    48.6
3  …      15.0  12579.0  476.0  15386.0  2626.0   196.0   130.0    60.1
4  …      15.0  12579.0  476.0  15386.0  2626.0   196.0   130.0    64.4

   ECO28_R  ECO34_R
0     44.1     28.7
1     53.8     18.6
2     51.1     30.0
3     39.5     26.2
4     34.9     41.9

[5 rows x 60 columns]
```

[26]:
```
df_scince["CVEGEO_clean"]=pd.to_numeric(df_scince["CVEGEO"])-19000
```

[27]:
```
var_abs = ["CVEGEO_clean", "VIV0", "VIV1", "VIV7", "VIV8", "VIV9", "VIV10",␣
 ↪"VIV14", "VIV16", "VIV25", "VIV26", "VIV27", "VIV28", "VIV29", "VIV31",␣
 ↪"VIV32", "VIV33", "VIV34", "VIV38", "VIV39", "VIV40", "VIV41", "VIV42",␣
 ↪"VIV81", "HOGAR1", "HOGAR4", "HOGAR7", "HOGAR13", "HOGAR16","HOGAR19",␣
 ↪"HOGAR22", "EDU25", "EDU28", "SALUD1", "SALUD2", "SALUD7", "SALUD8"]
var_rel = ["CVEGEO_clean", "VIV37_R", "VIV13_R", "VIV80_R","VIV81_R",␣
 ↪"VIV82_R","VIV83_R", "VIV84_R", "POB11_R", "POB17_R", "POB42_R", "POB84_R",␣
 ↪"MIG7_R","INDI1_R", "DISC1_R","EDU49_R","SCONY1_R", "SCONY4_R", "SCONY7_R",␣
 ↪"RELIG1_R", "RELIG2_R", "ECO1_R", "ECO28_R", "ECO34_R"]
```

[28]:
```
df_scince_abs=df_scince[var_abs]
#df_scince_rel=df_scince[var_rel]
```

[29]:
```
df_scince_abs=df_scince_abs.groupby("CVEGEO_clean").sum()
```

[30]:
```
#df_scince_rel=df_scince_rel.groupby("CVEGEO_clean").mean()
```

[31]:
```
#pd.concat([df_scince_abs,df_scince_rel],axis=1)
```

[32]:
```
#pd.concat([df_scince_abs,df_scince_rel],axis=1)
df_scince_clean=df_scince_abs
df_scince_clean.head()
```

[32]:

CVEGEO_clean	VIV0	VIV1	VIV7	VIV8	VIV9	\
1	3105594.0	2086002.0	15606.0	720477.0	1357722.0	
2	7727571.0	3316275.0	2601.0	1227672.0	2075598.0	
3	4221423.0	1407141.0	0.0	567018.0	837522.0	
4	35854785.0	27851508.0	140454.0	8840799.0	18945684.0	
5	21117519.0	14222268.0	179469.0	5048541.0	9137313.0	

CVEGEO_clean	VIV10	VIV14	VIV16	VIV25	VIV26	…	\
						…	
1	637245.0	72828.0	7803.0	2018376.0	2601.0	…	
2	514998.0	85833.0	28611.0	3191427.0	2601.0	…	
3	202878.0	10404.0	5202.0	1365525.0	5202.0	…	
4	6260607.0	522801.0	18207.0	27435348.0	2601.0	…	
5	3729834.0	462978.0	249696.0	13384746.0	98838.0	…	

CVEGEO_clean	HOGAR13	HOGAR16	HOGAR19	HOGAR22	EDU25	\
1	1292697.0	5165586.0	577422.0	1807695.0	219363903.0	
2	1584009.0	5004324.0	1516383.0	3134205.0	219363903.0	
3	655452.0	2023578.0	655452.0	1344717.0	219363903.0	
4	16935111.0	62936397.0	7935651.0	18904068.0	219363903.0	
5	8310195.0	30941496.0	4523139.0	11207709.0	219363903.0	

CVEGEO_clean	EDU28	SALUD1	SALUD2	SALUD7	SALUD8
1	3262980.0	238754715.0	54705864.0	16836630.0	4432410.0
2	3262980.0	238754715.0	54705864.0	16836630.0	4432410.0
3	3262980.0	238754715.0	54705864.0	16836630.0	4432410.0
4	3262980.0	238754715.0	54705864.0	16836630.0	4432410.0
5	3262980.0	238754715.0	54705864.0	16836630.0	4432410.0

[5 rows x 36 columns]

[33]:
```
df_scince_clean.describe()
```

[33]:
```
              VIV0          VIV1          VIV7          VIV8          VIV9  \
count  5.100000e+01  5.100000e+01  5.100000e+01  5.100000e+01  5.100000e+01
mean   1.039224e+08  8.444019e+07  7.108890e+05  2.030606e+07  6.376489e+07
std    1.903549e+08  1.647698e+08  1.222103e+06  3.739905e+07  1.269531e+08
min    2.158830e+06  8.687340e+05  0.000000e+00  2.366910e+05  6.294420e+05
25%    5.987502e+06  3.184924e+06  2.210850e+04  1.178253e+06  1.832404e+06
50%    1.534590e+07  1.112188e+07  1.404540e+05  3.633597e+06  6.648156e+06
75%    9.747898e+07  7.160813e+07  9.883800e+05  1.936835e+07  5.142957e+07
max    9.592384e+08  8.559761e+08  6.796413e+06  2.046727e+08  6.432793e+08

             VIV10         VIV14          VIV16         VIV25          VIV26  \
count  5.100000e+01  5.100000e+01      51.000000  5.100000e+01      51.000000
mean   1.734597e+07  1.112871e+06  145299.000000  8.288398e+07   37230.000000
std    3.063926e+07  2.146720e+06  151870.801602  1.634828e+08   55565.274942
min    1.352520e+05  1.040400e+04       0.000000  8.401230e+05       0.000000
25%    7.048710e+05  9.103500e+04   26010.000000  2.697237e+06    5202.000000
50%    2.814282e+06  2.991150e+05   88434.000000  8.900622e+06   18207.000000
75%    1.768940e+07  1.258884e+06  237991.500000  7.085774e+07   41616.000000
max    1.323779e+08  1.337954e+07  624240.000000  8.440791e+08  239292.000000

       …       HOGAR13       HOGAR16       HOGAR19       HOGAR22  \
count  …  5.100000e+01  5.100000e+01  5.100000e+01  5.100000e+01
mean   …  5.371167e+07  1.987905e+08  2.032697e+07  6.290712e+07
std    …  1.033589e+08  3.841008e+08  5.167020e+07  1.660751e+08
min    …  3.667410e+05  1.201662e+06  4.161600e+05  9.519660e+05
25%    …  1.608718e+06  5.505016e+06  1.113228e+06  2.910519e+06
50%    …  5.940684e+06  2.197065e+07  3.625794e+06  1.019332e+07
75%    …  4.587384e+07  1.702133e+08  1.154584e+07  3.139667e+07
max    …  4.761963e+08  1.771528e+09  3.083850e+08  9.846814e+08

             EDU25      EDU28       SALUD1      SALUD2      SALUD7     SALUD8
count         51.0       51.0         51.0        51.0        51.0       51.0
mean   219363903.0  3262980.0  238754715.0  54705864.0  16836630.0  4432410.0
std            0.0        0.0          0.0         0.0         0.0        0.0
min    219363903.0  3262980.0  238754715.0  54705864.0  16836630.0  4432410.0
25%    219363903.0  3262980.0  238754715.0  54705864.0  16836630.0  4432410.0
50%    219363903.0  3262980.0  238754715.0  54705864.0  16836630.0  4432410.0
75%    219363903.0  3262980.0  238754715.0  54705864.0  16836630.0  4432410.0
max    219363903.0  3262980.0  238754715.0  54705864.0  16836630.0  4432410.0

[8 rows x 36 columns]
```

[34]:
```
df_scince_clean=df_scince_clean.
  ↪drop(["EDU25","EDU28","SALUD1","SALUD2","SALUD7","SALUD8"],axis=1)
df_scince_clean.head()
```

[34]:

CVEGEO_clean	VIV0	VIV1	VIV7	VIV8	VIV9	\
1	3105594.0	2086002.0	15606.0	720477.0	1357722.0	
2	7727571.0	3316275.0	2601.0	1227672.0	2075598.0	
3	4221423.0	1407141.0	0.0	567018.0	837522.0	
4	35854785.0	27851508.0	140454.0	8840799.0	18945684.0	
5	21117519.0	14222268.0	179469.0	5048541.0	9137313.0	

CVEGEO_clean	VIV10	VIV14	VIV16	VIV25	VIV26	…	\
						…	
1	637245.0	72828.0	7803.0	2018376.0	2601.0	…	
2	514998.0	85833.0	28611.0	3191427.0	2601.0	…	
3	202878.0	10404.0	5202.0	1365525.0	5202.0	…	
4	6260607.0	522801.0	18207.0	27435348.0	2601.0	…	
5	3729834.0	462978.0	249696.0	13384746.0	98838.0	…	

CVEGEO_clean	VIV41	VIV42	VIV81	HOGAR1	HOGAR4	HOGAR7	\
1	23409.0	10404.0	249696.0	2086002.0	7735374.0	213282.0	
2	41616.0	15606.0	619038.0	3313674.0	8780976.0	213282.0	
3	18207.0	2601.0	288711.0	1407141.0	3659607.0	96237.0	
4	239292.0	80631.0	4031550.0	27838503.0	91282095.0	2954736.0	
5	140454.0	67626.0	3646602.0	14222268.0	46896030.0	1388934.0	

CVEGEO_clean	HOGAR13	HOGAR16	HOGAR19	HOGAR22
1	1292697.0	5165586.0	577422.0	1807695.0
2	1584009.0	5004324.0	1516383.0	3134205.0
3	655452.0	2023578.0	655452.0	1344717.0
4	16935111.0	62936397.0	7935651.0	18904068.0
5	8310195.0	30941496.0	4523139.0	11207709.0

[5 rows x 30 columns]

[35]:
```
df_denue.columns
```

[35]:
```
Index(['codigo_act', 'nom_estab', 'cod_postal', 'cve_mun', 'municipio',
       'latitud', 'longitud', 'manzana', 'geometry'],
      dtype='object')
```

[36]:
```
var_denue=["codigo_act","cve_mun"]#,"cod_postal"]
```

[37]:
```
df_denue_clean=df_denue[var_denue]
```

[38]:
```

```

```
#negocios = {"NEGOCIO_TIENDITA": "461110", "NEGOCIO_ROPA": "463211",␣
 ↪"NEGOCIO_TECNOLOGIA": "466211", "NEGOCIO_RESTAURANTE": "722",␣
 ↪"NEGOCIO_PAPELERIA":"465311", "NEGOCIO_GIMNASIO": "713943",␣
 ↪"NEGOCIO_ESTETICA": "812110"}
#codigos_negocios = ("461110", "463211", "466211", "722511", "722512",␣
 ↪"722513", "722514", "722515", "465311", "713943", "812110")
```

[39]:
```
df_denue_clean["tiendita"]='a'
df_denue_clean["ropa"]='b'
df_denue_clean["tecnologia"]='c'
df_denue_clean["restaurante"]='d'
df_denue_clean["papeleria"]='e'
df_denue_clean["gimnasio"]='f'
df_denue_clean["estetica"]='g'

df_denue_clean["tiendita"]=0
df_denue_clean["ropa"]=0
df_denue_clean["tecnologia"]=0
df_denue_clean["restaurante"]=0
df_denue_clean["papeleria"]=0
df_denue_clean["gimnasio"]=0
df_denue_clean["estetica"]=0
df_denue_clean.head()
```

```
C:\Users\alexd\AppData\Local\Temp\ipykernel_3040\1454253610.py:1:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df_denue_clean["tiendita"]='a'
C:\Users\alexd\AppData\Local\Temp\ipykernel_3040\1454253610.py:2:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df_denue_clean["ropa"]='b'
C:\Users\alexd\AppData\Local\Temp\ipykernel_3040\1454253610.py:3:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df_denue_clean["tecnologia"]='c'
```

```
C:\Users\alexd\AppData\Local\Temp\ipykernel_3040\1454253610.py:4:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df_denue_clean["restaurante"]='d'
C:\Users\alexd\AppData\Local\Temp\ipykernel_3040\1454253610.py:5:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df_denue_clean["papeleria"]='e'
C:\Users\alexd\AppData\Local\Temp\ipykernel_3040\1454253610.py:6:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df_denue_clean["gimnasio"]='f'
```

[39]:

	codigo_act	cve_mun	tiendita	ropa	tecnologia	restaurante	papeleria	gimnasio	estetica
0	236221	039	0	0	0	0	0	0	0
1	237121	039	0	0	0	0	0	0	0
2	236221	039	0	0	0	0	0	0	0
3	236111	019	0	0	0	0	0	0	0
4	236111	048	0	0	0	0	0	0	0

[40]:
```
df_denue_clean.dtypes
```

[40]:
```
codigo_act     object
cve_mun        object
tiendita        int64
ropa            int64
tecnologia      int64
restaurante     int64
papeleria       int64
```

```
gimnasio        int64
estetica        int64
dtype: object
```

[41]:
```
df_denue_clean["cve_mun"]=pd.to_numeric(df_denue_clean["cve_mun"])
```

[42]:
```
df_denue_clean.loc[df_denue_clean.codigo_act=="461110","tiendita"]=1
df_denue_clean.loc[df_denue_clean.codigo_act=="463211","ropa"]=1
df_denue_clean.loc[df_denue_clean.codigo_act=="466211","tecnologia"]=1
df_denue_clean.loc[df_denue_clean.codigo_act=="465311","papeleria"]=1
df_denue_clean.loc[df_denue_clean.codigo_act=="713943","gimnasio"]=1
df_denue_clean.loc[df_denue_clean.codigo_act=="812110","estetica"]=1
df_denue_clean.loc[df_denue_clean.codigo_act=="722513","restaurante"]=1
```

[43]:
```
df_denue_clean
```

[43]:

	codigo_act	cve_mun	tiendita	ropa	tecnologia	restaurante	\
0	236221	39	0	0	0	0	
1	237121	39	0	0	0	0	
2	236221	39	0	0	0	0	
3	236111	19	0	0	0	0	
4	236111	48	0	0	0	0	
…	…	…	…	…	…	…	
186087	812110	31	0	0	0	0	
186088	812110	46	0	0	0	0	
186089	812110	46	0	0	0	0	
186090	812110	6	0	0	0	0	
186091	812110	26	0	0	0	0	

	papeleria	gimnasio	estetica
0	0	0	0
1	0	0	0
2	0	0	0
3	0	0	0
4	0	0	0
…	…	…	…
186087	0	0	1
186088	0	0	1
186089	0	0	1
186090	0	0	1
186091	0	0	1

[186092 rows x 9 columns]

[44]:
```

```

```
df_denue_clean=df_denue_clean.loc[(df_denue_clean.
  ↪codigo_act=="461110")|(df_denue_clean.codigo_act=="463211")|(df_denue_clean.
  ↪codigo_act=="466211")|(df_denue_clean.codigo_act=="722513")|(df_denue_clean.
  ↪codigo_act=="465311")|(df_denue_clean.codigo_act=="713943")|(df_denue_clean.
  ↪codigo_act=="812110")|(df_denue_clean.codigo_act=="722513")]
df_denue_clean.head()
```

[44]:

	codigo_act	cve_mun	tiendita	ropa	tecnologia	restaurante	papeleria	gimnasio	estetica
20001	461110	18	1	0	0	0	0	0	0
20002	461110	21	1	0	0	0	0	0	0
20003	461110	39	1	0	0	0	0	0	0
20004	461110	39	1	0	0	0	0	0	0
20005	461110	46	1	0	0	0	0	0	0

[45]:
```
df_denue_clean=df_denue_clean.groupby(["cve_mun"]).sum()
df_denue_clean.head()
```

```
C:\Users\alexd\AppData\Local\Temp\ipykernel_3040\3280286590.py:1: FutureWarning:
The default value of numeric_only in DataFrameGroupBy.sum is deprecated. In a
future version, numeric_only will default to False. Either specify numeric_only
or select only columns which should be valid for the function.
  df_denue_clean=df_denue_clean.groupby(["cve_mun"]).sum()
```

[45]:

cve_mun	tiendita	ropa	tecnologia	restaurante	papeleria	gimnasio	estetica
1	14	0	0	0	3	0	2
2	14	0	0	0	0	0	1
3	6	0	0	0	2	0	1
4	157	43	2	8	22	5	59
5	80	15	0	3	14	1	29

[46]:
```
df_final=pd.concat([df_scince_clean,df_denue_clean],axis=1)
df_final
```

[46]:

	VIV0	VIV1	VIV7	VIV8	VIV9 \
1	3105594.0	2086002.0	15606.0	720477.0	1357722.0
2	7727571.0	3316275.0	2601.0	1227672.0	2075598.0
3	4221423.0	1407141.0	0.0	567018.0	837522.0
4	35854785.0	27851508.0	140454.0	8840799.0	18945684.0
5	21117519.0	14222268.0	179469.0	5048541.0	9137313.0
6	553495401.0	472674528.0	1672443.0	87630291.0	384615072.0
7	15983145.0	11121876.0	1217268.0	4434705.0	6648156.0
8	6213789.0	3074382.0	18207.0	1131435.0	1844109.0
9	143655831.0	97188966.0	496791.0	32444874.0	64562022.0
10	100866780.0	78144444.0	533205.0	22295772.0	55684809.0
11	11140083.0	6388056.0	28611.0	2439738.0	3940515.0
12	80644005.0	53203455.0	247095.0	15686631.0	37472607.0
13	14417343.0	8500068.0	72828.0	2658222.0	5836644.0
14	34720749.0	24004629.0	990981.0	7894035.0	16037766.0
15	3240846.0	1284894.0	7803.0	374544.0	905148.0
16	5477706.0	2676429.0	31212.0	1121031.0	1552797.0
17	42718824.0	29092185.0	1167849.0	10502838.0	18420282.0
18	410383179.0	298854900.0	2434536.0	82553139.0	216067671.0
19	107257437.0	98198154.0	301716.0	18009324.0	76383567.0
20	8835597.0	4575159.0	23409.0	1768680.0	2783070.0
21	401115816.0	342658341.0	3636198.0	80227845.0	261923301.0
22	19437273.0	12063438.0	161262.0	4778037.0	7264593.0
23	2850696.0	1716660.0	7803.0	564417.0	1147041.0
24	6338637.0	4309857.0	988380.0	1576206.0	2710242.0
25	110035305.0	77072832.0	988380.0	16537158.0	60449841.0
26	538882983.0	474807348.0	2486556.0	100952613.0	373326732.0
27	4939299.0	1901331.0	5202.0	697068.0	1193859.0
28	2171835.0	1097622.0	7803.0	444771.0	647649.0
29	8741961.0	5628564.0	36414.0	1872720.0	3748041.0
30	3781854.0	2595798.0	267903.0	949365.0	1615221.0
31	471985263.0	349688844.0	3063978.0	86145120.0	263010519.0
32	7839414.0	4325463.0	7803.0	1557999.0	2759661.0
33	83174778.0	63797328.0	1212066.0	19869039.0	43858062.0
34	5761215.0	3748041.0	20808.0	1225071.0	2517768.0
35	2395521.0	1201662.0	5202.0	317322.0	881739.0
36	6923862.0	4892481.0	132651.0	1846710.0	3045771.0
37	7092927.0	4439907.0	119646.0	1901331.0	2530773.0
38	78391539.0	54392112.0	756891.0	19861236.0	34288983.0
39	959238396.0	855976095.0	6796413.0	204672690.0	643279320.0
40	2158830.0	868734.0	7803.0	236691.0	629442.0
41	149302602.0	113614281.0	512397.0	31630761.0	81858672.0
42	11990610.0	4840461.0	52020.0	2122416.0	2699838.0
43	4356675.0	2007972.0	106641.0	824517.0	1180854.0
44	41948928.0	28088199.0	98838.0	8489664.0	19494495.0
45	94091175.0	66143430.0	886941.0	18875457.0	47174337.0
46	347818725.0	316796598.0	1586610.0	60455043.0	255967011.0

```
47   15345900.0   11915181.0    52020.0    3633597.0    8265978.0
48  236602566.0  215687925.0  2015775.0   41592591.0  172063953.0
49   54061785.0   37545435.0   621639.0   12383361.0   24634071.0
50    3420315.0    1466964.0     5202.0     548811.0     912951.0
51    6770403.0    3295467.0    26010.0    1469565.0    1820700.0

          VIV10       VIV14     VIV16        VIV25     VIV26  …  \
1      637245.0     72828.0    7803.0    2018376.0    2601.0  …
2      514998.0     85833.0   28611.0    3191427.0    2601.0  …
3      202878.0     10404.0    5202.0    1365525.0    5202.0  …
4     6260607.0    522801.0   18207.0   27435348.0    2601.0  …
5     3729834.0    462978.0  249696.0   13384746.0   98838.0  …
6    94299255.0   2902716.0  132651.0  471649734.0   18207.0  …
7     3550365.0    512397.0  299115.0    8900622.0   59823.0  …
8      694467.0     91035.0   31212.0    2876706.0   23409.0  …
9    23382990.0   1602216.0  176868.0   95383872.0   23409.0  …
10   25325937.0   1180854.0  226287.0   77163867.0   10404.0  …
11    1292697.0    119646.0   96237.0    6143562.0   26010.0  …
12   15548778.0    697068.0   78030.0   52880931.0    2601.0  …
13    1708857.0    124848.0  306918.0    7386840.0  239292.0  …
14    7368633.0    728280.0  413559.0   11324754.0  239292.0  …
15     169065.0     18207.0   10404.0    1186056.0    2601.0  …
16     707472.0     91035.0   26010.0    2517768.0    5202.0  …
17    9316782.0   1456560.0  351135.0   23367384.0  127449.0  …
18   85801788.0   4109580.0  400554.0  296794908.0   59823.0  …
19    7930449.0    998784.0   54621.0   94314861.0   31212.0  …
20    1136637.0    111843.0   88434.0    4200615.0   39015.0  …
21   81850869.0   5417883.0  379746.0  338795856.0   49419.0  …
22    2814282.0    299115.0  140454.0   10723923.0   59823.0  …
23     260100.0     15606.0   20808.0    1641231.0   10404.0  …
24    1610019.0    205479.0  101439.0    2970342.0   33813.0  …
25   21419235.0    944163.0  449973.0   76362759.0   26010.0  …
26   74651301.0   3867687.0  332928.0  472388418.0   36414.0  …
27     273105.0     20808.0   20808.0    1802493.0   10404.0  …
28     348534.0     70227.0   31212.0    1035198.0       0.0  …
29    1399338.0     72828.0   23409.0    5233212.0   23409.0  …
30     741285.0     93636.0   70227.0    2270673.0   13005.0  …
31   96234399.0   3904101.0  345933.0  347384358.0   23409.0  …
32    1183455.0    101439.0   39015.0    4101777.0   15606.0  …
33   16768647.0   1264086.0  410958.0   59240376.0  195075.0  …
34    1100223.0    130050.0       0.0    3703824.0       0.0  …
35     239292.0     23409.0    5202.0    1170450.0    2601.0  …
36    1877922.0    335529.0  104040.0    1880523.0   46818.0  …
37    1477368.0    205479.0   57222.0    4031550.0       0.0  …
38   13460175.0   1339515.0  252297.0   51606441.0   44217.0  …
39  132377895.0  13379544.0  624240.0  844079121.0  124848.0  …
40     135252.0     13005.0    7803.0     840123.0    2601.0  …
```

41	33573708.0	1836306.0	223686.0	112675320.0	15606.0	…
42	998784.0	93636.0	26010.0	4380084.0	13005.0	…
43	603432.0	101439.0	36414.0	1675044.0	10404.0	…
44	6029118.0	457776.0	70227.0	27411939.0	26010.0	…
45	18610155.0	1253682.0	127449.0	65352726.0	23409.0	…
46	32720580.0	1386333.0	114444.0	316109934.0	7803.0	…
47	2931327.0	187272.0	41616.0	11683692.0	5202.0	…
48	40918932.0	2887110.0	213282.0	212514705.0	44217.0	…
49	7490880.0	832320.0	117045.0	36081072.0	7803.0	…
50	262701.0	44217.0	7803.0	1303101.0	2601.0	…
51	702270.0	72828.0	13005.0	3144609.0	5202.0	…

	HOGAR16	HOGAR19	HOGAR22	tiendita	ropa	tecnologia	\
1	5.165586e+06	577422.0	1807695.0	14	0	0	
2	5.004324e+06	1516383.0	3134205.0	14	0	0	
3	2.023578e+06	655452.0	1344717.0	6	0	0	
4	6.293640e+07	7935651.0	18904068.0	157	43	2	
5	3.094150e+07	4523139.0	11207709.0	80	15	0	
6	1.353329e+09	61641099.0	195730452.0	1221	169	6	
7	2.391620e+07	4452912.0	12534219.0	71	9	1	
8	5.828841e+06	1173051.0	2822085.0	36	4	0	
9	2.158804e+08	19835226.0	50542632.0	406	76	3	
10	1.858154e+08	4127787.0	10924200.0	83	5	0	
11	1.234955e+07	2033982.0	4676598.0	36	4	0	
12	1.228972e+08	4195413.0	10853973.0	175	16	0	
13	1.638630e+07	3048372.0	7280199.0	36	16	0	
14	5.869677e+07	8973450.0	28639611.0	132	39	0	
15	2.114613e+06	533205.0	1050804.0	4	0	0	
16	5.568741e+06	806310.0	1914336.0	20	0	0	
17	6.654138e+07	10489833.0	30769830.0	120	20	0	
18	7.825759e+08	18347454.0	49595868.0	1058	38	2	
19	1.896025e+08	39660048.0	124046892.0	179	200	29	
20	9.085293e+06	1487772.0	3412512.0	22	1	0	
21	9.507461e+08	50774121.0	161490888.0	1575	127	5	
22	2.197065e+07	4772835.0	11080260.0	29	8	0	
23	2.476152e+06	798507.0	1784286.0	7	0	0	
24	1.005286e+07	1573605.0	4710411.0	55	3	0	
25	1.938499e+08	4356675.0	11556243.0	257	22	2	
26	1.033216e+09	166794327.0	544782051.0	1743	294	16	
27	2.601000e+06	923355.0	2086002.0	20	0	0	
28	2.312289e+06	416160.0	1087218.0	3	0	0	
29	1.176432e+07	2091204.0	5144778.0	32	9	0	
30	5.300838e+06	1009188.0	2533374.0	27	6	0	
31	9.317172e+08	28553778.0	81322866.0	1445	62	1	
32	8.950041e+06	1438353.0	3438522.0	26	7	0	
33	1.486784e+08	18246015.0	51151266.0	350	72	3	
34	8.994258e+06	1058607.0	2998953.0	16	1	0	

35	2.330496e+06	449973.0	1147041.0	5	0	0
36	1.227932e+07	1792089.0	5909472.0	28	2	0
37	1.070832e+07	1167849.0	3162816.0	36	1	0
38	1.149902e+08	15556581.0	39550806.0	247	60	4
39	1.771528e+09	308384964.0	984681378.0	3604	1146	133
40	1.201662e+06	431766.0	951966.0	6	0	0
41	2.569060e+08	6512904.0	17608770.0	198	13	0
42	8.252973e+06	2122416.0	4645386.0	5	6	0
43	3.857283e+06	788103.0	1838907.0	8	1	0
44	5.921957e+07	8773173.0	22238550.0	185	52	2
45	1.546112e+08	6770403.0	18352656.0	169	13	0
46	5.903386e+08	134479503.0	425942361.0	887	261	18
47	2.830928e+07	3625794.0	10193319.0	69	12	0
48	5.486992e+08	52267095.0	178930593.0	822	60	5
49	7.809763e+07	12601845.0	32023512.0	146	38	2
50	2.252466e+06	637245.0	1422747.0	4	0	0
51	5.441292e+06	1492974.0	3303270.0	31	2	0

	restaurante	papeleria	gimnasio	estetica
1	0	3	0	2
2	0	0	0	1
3	0	2	0	1
4	8	22	5	59
5	3	14	1	29
6	264	446	92	992
7	6	9	1	13
8	0	4	0	5
9	63	69	18	195
10	7	12	0	21
11	3	6	2	18
12	15	31	5	69
13	2	4	2	20
14	15	17	3	25
15	0	0	0	0
16	1	1	0	2
17	6	19	3	32
18	65	218	26	347
19	37	54	59	490
20	0	0	0	12
21	183	355	64	778
22	4	6	1	20
23	0	0	0	1
24	1	4	0	2
25	18	59	6	91
26	347	529	173	1514
27	0	0	0	1
28	0	1	0	1

```
29            2          6         1         15
30            1          5         0          5
31          134        326        40        546
32            1          1         0          3
33           30         67        15        150
34            2          3         0          4
35            0          0         0          0
36            0          0         0          1
37            0          3         2          4
38           15         41         6         87
39          639        738       271       2830
40            0          1         0          0
41            6         35         1         79
42            0          0         0          7
43            0          0         0          0
44           13         27         8         86
45           10         28         2         32
46          214        284       105       1165
47            4         12         1         18
48           96        162        42        515
49           27         19         6         76
50            0          0         0          0
51            1          3         0          7

[51 rows x 37 columns]
```

2 Machine Learning Modelling

[47]:
```
import matplotlib.pyplot as plt
import numpy as np
#%pip install seaborn
import seaborn as sns

from sklearn.model_selection import train_test_split
from sklearn.model_selection import GridSearchCV
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler

from sklearn.linear_model import LinearRegression
from sklearn.linear_model import Lasso
from sklearn.linear_model import Ridge
from sklearn.linear_model import BayesianRidge

from sklearn.metrics import mean_squared_error
from sklearn.metrics import mean_absolute_error
```

[48]:
```
pd.set_option('display.float_format', lambda x: '%.3f' % x)
```

[49]:
```
df_final.describe()
```

[49]:

	VIV0	VIV1	VIV7	VIV8	VIV9	\
count	51.000	51.000	51.000	51.000	51.000	
mean	103922445.000	84440190.000	710889.000	20306058.000	63764892.000	
std	190354881.383	164769840.286	1222102.981	37399049.365	126953117.761	
min	2158830.000	868734.000	0.000	236691.000	629442.000	
25%	5987502.000	3184924.500	22108.500	1178253.000	1832404.500	
50%	15345900.000	11121876.000	140454.000	3633597.000	6648156.000	
75%	97478977.500	71608131.000	988380.000	19368346.500	51429573.000	
max	959238396.000	855976095.000	6796413.000	204672690.000	643279320.000	

	VIV10	VIV14	VIV16	VIV25	VIV26	…	\
count	51.000	51.000	51.000	51.000	51.000	…	
mean	17345967.000	1112871.000	145299.000	82883976.000	37230.000	…	
std	30639258.974	2146719.685	151870.802	163482806.504	55565.275	…	
min	135252.000	10404.000	0.000	840123.000	0.000	…	
25%	704871.000	91035.000	26010.000	2697237.000	5202.000	…	
50%	2814282.000	299115.000	88434.000	8900622.000	18207.000	…	
75%	17689401.000	1258884.000	237991.500	70857742.500	41616.000	…	
max	132377895.000	13379544.000	624240.000	844079121.000	239292.000	…	

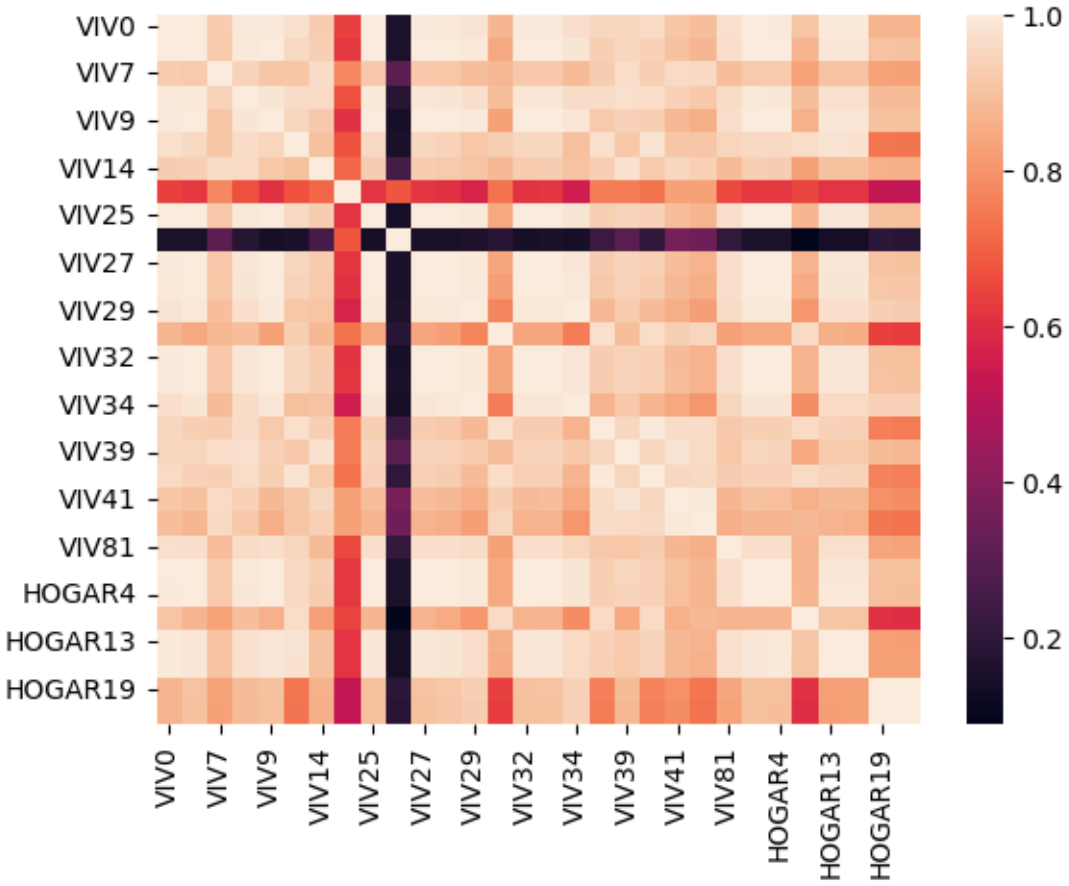
	HOGAR16	HOGAR19	HOGAR22	tiendita	ropa	\
count	51.000	51.000	51.000	51.000	51.000	
mean	198790452.000	20326968.000	62907123.000	311.863	57.510	
std	384100822.703	51670199.846	166075063.385	638.801	168.367	
min	1201662.000	416160.000	951966.000	3.000	0.000	
25%	5505016.500	1113228.000	2910519.000	20.000	1.000	
50%	21970647.000	3625794.000	10193319.000	55.000	9.000	
75%	170213341.500	11545839.000	31396671.000	191.500	41.000	
max	1771528095.000	308384964.000	984681378.000	3604.000	1146.000	

	tecnologia	restaurante	papeleria	gimnasio	estetica
count	51.000	51.000	51.000	51.000	51.000
mean	4.588	43.980	71.490	18.843	203.353
std	19.067	111.264	153.153	48.410	493.411
min	0.000	0.000	0.000	0.000	0.000
25%	0.000	0.000	1.000	0.000	2.000
50%	0.000	3.000	6.000	1.000	18.000
75%	2.000	16.500	38.000	6.000	86.500
max	133.000	639.000	738.000	271.000	2830.000

[8 rows x 37 columns]

[50]:
```
sns.heatmap(df_scince_clean.corr(),annot=False)
```

[50]: <AxesSubplot: >

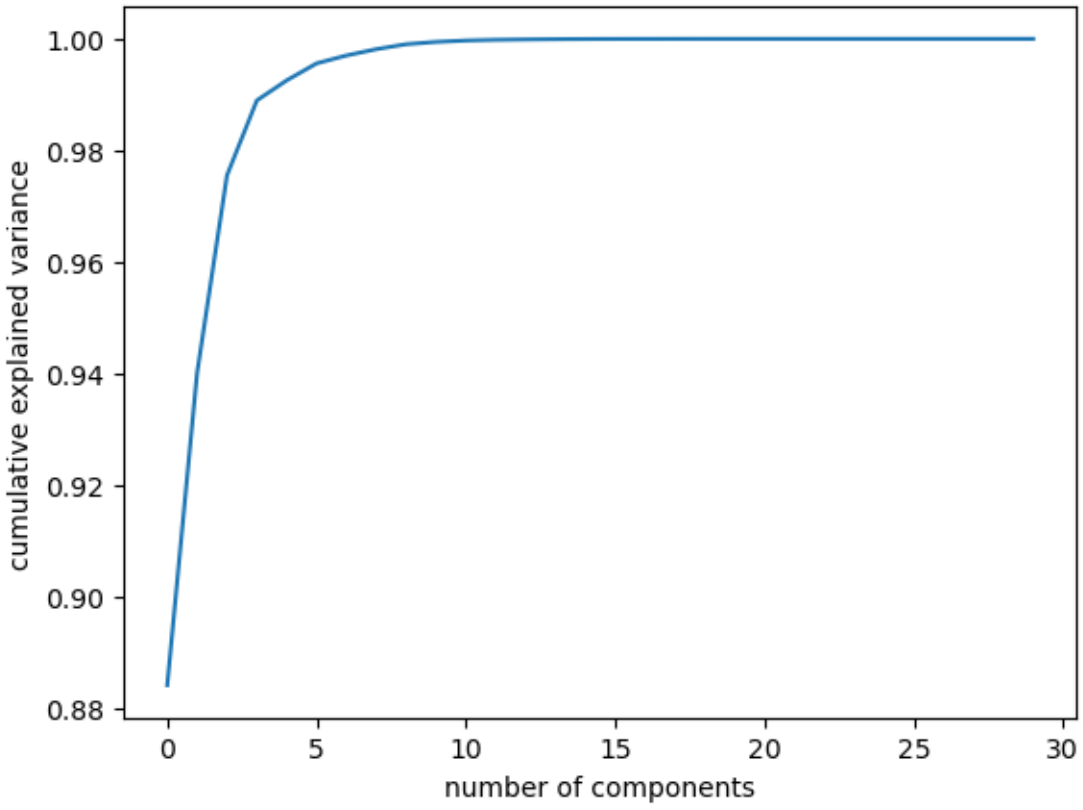


[51]:
```
X_full=df_scince_clean
```

[52]:
```
scaler=StandardScaler()
scaler.fit(X_full)
X_scaled=scaler.transform(X_full)
```

[53]:
```
pca=PCA()
pca.fit(X_scaled)
plt.plot(np.cumsum(pca.explained_variance_ratio_))
plt.xlabel('number of components')
plt.ylabel('cumulative explained variance')
```

[53]: Text(0, 0.5, 'cumulative explained variance')

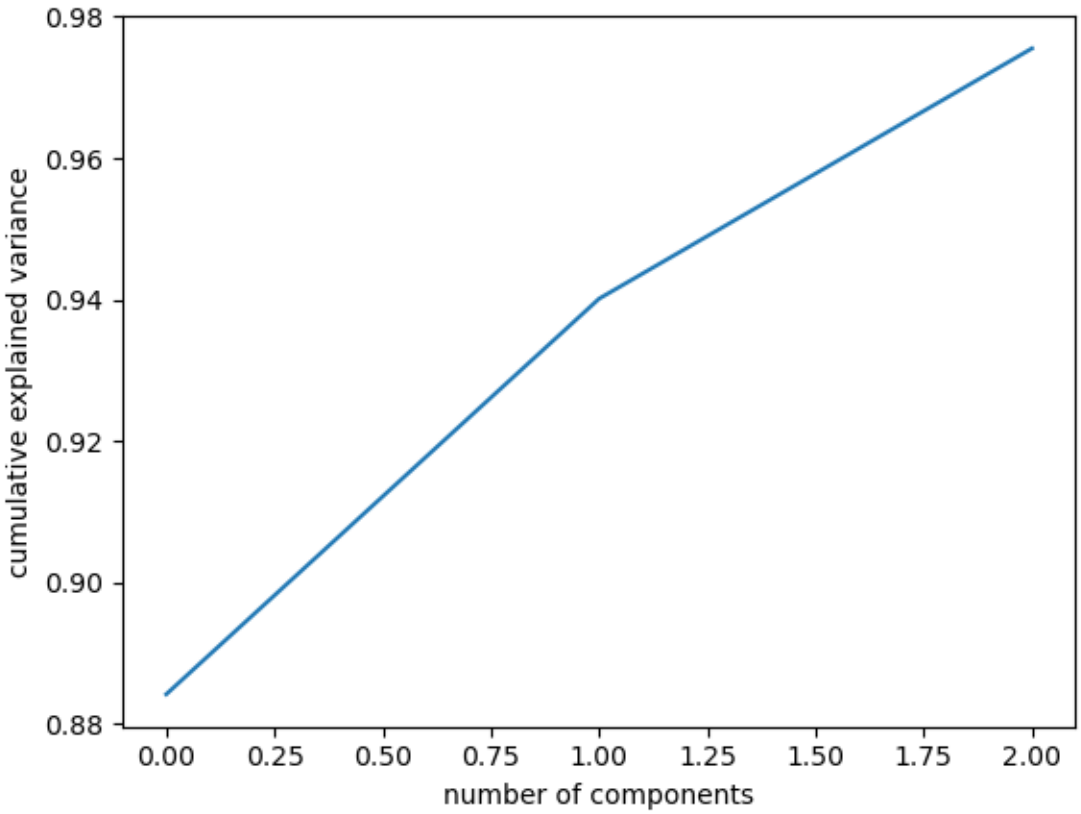


[54]:
```
np.cumsum(pca.explained_variance_ratio_)
```

[54]:
```
array([0.8842023 , 0.94013751, 0.97552921, 0.98896273, 0.99256255,
       0.99560121, 0.99701833, 0.99814967, 0.99903804, 0.99945328,
       0.9997019 , 0.99982982, 0.99989811, 0.999944  , 0.9999702 ,
       0.99998662, 0.99999215, 0.99999595, 0.9999981 , 0.99999891,
       0.99999946, 0.99999981, 0.9999999 , 0.99999997, 1.        ,
       1.        , 1.        , 1.        , 1.        , 1.        ])
```

[55]:
```
pca_3=PCA(n_components=3)
pca_3.fit(X_scaled)
plt.plot(np.cumsum(pca_3.explained_variance_ratio_))
plt.xlabel('number of components')
plt.ylabel('cumulative explained variance')
```

[55]:
```
Text(0, 0.5, 'cumulative explained variance')
```



[56]:
```
X_pca=pca_3.transform(X_scaled)
```

2.0.1 Train-Test Split

[57]:
```
Y=df_denue_clean
Y_tiendita=df_denue_clean["tiendita"]
Y_ropa=df_denue_clean["ropa"]
Y_tecnologia=df_denue_clean["tecnologia"]
Y_restaurante=df_denue_clean["restaurante"]
Y_papeleria=df_denue_clean["papeleria"]
Y_gimnasio=df_denue_clean["gimnasio"]
Y_estetica=df_denue_clean["estetica"]
```

[58]:
```
X_train, X_test, y_train_tiendita, y_test_tiendita =␣
 ↪train_test_split(X_pca,Y_tiendita, test_size=0.3, random_state=100)
X_train, X_test, y_train_ropa, y_test_ropa = train_test_split(X_pca,Y_ropa,␣
 ↪test_size=0.3, random_state=100)
X_train, X_test, y_train_tecnologia, y_test_tecnologia =␣
 ↪train_test_split(X_pca,Y_tecnologia, test_size=0.3, random_state=100)
```

```
X_train, X_test, y_train_restaurante, y_test_restaurante =␣
 ↪train_test_split(X_pca,Y_restaurante, test_size=0.3, random_state=100)
X_train, X_test, y_train_papeleria, y_test_papeleria =␣
 ↪train_test_split(X_pca,Y_papeleria, test_size=0.3, random_state=100)
X_train, X_test, y_train_gimnasio, y_test_gimnasio =␣
 ↪train_test_split(X_pca,Y_gimnasio, test_size=0.3, random_state=100)
X_train, X_test, y_train_estetica, y_test_estetica =␣
 ↪train_test_split(X_pca,Y_estetica, test_size=0.3, random_state=100)
```

2.1 Models

[59]:
```
lr=LinearRegression(fit_intercept=True,normalize=False)
lasso=Lasso()
ridge=Ridge()
bay=BayesianRidge()
```

[60]:
```
mse_tiendita = {"LinearReg":[],"LassoReg":[],"RidgeReg":[],"BayReg":[]}
mse_ropa = {"LinearReg":[],"LassoReg":[],"RidgeReg":[],"BayReg":[]}
mse_tecnologia = {"LinearReg":[],"LassoReg":[],"RidgeReg":[],"BayReg":[]}
mse_restaurante = {"LinearReg":[],"LassoReg":[],"RidgeReg":[],"BayReg":[]}
mse_papeleria = {"LinearReg":[],"LassoReg":[],"RidgeReg":[],"BayReg":[]}
mse_gimnasio = {"LinearReg":[],"LassoReg":[],"RidgeReg":[],"BayReg":[]}
mse_estetica = {"LinearReg":[],"LassoReg":[],"RidgeReg":[],"BayReg":[]}

mae_tiendita = {"LinearReg":[],"LassoReg":[],"RidgeReg":[],"BayReg":[]}
mae_ropa = {"LinearReg":[],"LassoReg":[],"RidgeReg":[],"BayReg":[]}
mae_tecnologia = {"LinearReg":[],"LassoReg":[],"RidgeReg":[],"BayReg":[]}
mae_restaurante = {"LinearReg":[],"LassoReg":[],"RidgeReg":[],"BayReg":[]}
mae_papeleria = {"LinearReg":[],"LassoReg":[],"RidgeReg":[],"BayReg":[]}
mae_gimnasio = {"LinearReg":[],"LassoReg":[],"RidgeReg":[],"BayReg":[]}
mae_estetica = {"LinearReg":[],"LassoReg":[],"RidgeReg":[],"BayReg":[]}
```

[61]:
```
lr.fit(X_train,y_train_tiendita)
y_pred_tiendita=lr.predict(X_test)
mse_tiendita["LinearReg"].
 ↪append(mean_squared_error(y_test_tiendita,y_pred_tiendita))
mae_tiendita["LinearReg"].
 ↪append(mean_absolute_error(y_test_tiendita,y_pred_tiendita))

lasso.fit(X_train,y_train_tiendita)
y_pred_tiendita=lasso.predict(X_test)
mse_tiendita["LassoReg"].
 ↪append(mean_squared_error(y_test_tiendita,y_pred_tiendita))
mae_tiendita["LassoReg"].
 ↪append(mean_absolute_error(y_test_tiendita,y_pred_tiendita))

ridge.fit(X_train,y_train_tiendita)
```

```
y_pred_tiendita=ridge.predict(X_test)
mse_tiendita["RidgeReg"].
  ↪append(mean_squared_error(y_test_tiendita,y_pred_tiendita))
mae_tiendita["RidgeReg"].
  ↪append(mean_absolute_error(y_test_tiendita,y_pred_tiendita))

bay.fit(X_train,y_train_tiendita)
y_pred_tiendita=bay.predict(X_test)
mse_tiendita["BayReg"].
  ↪append(mean_squared_error(y_test_tiendita,y_pred_tiendita))
mae_tiendita["BayReg"].
  ↪append(mean_absolute_error(y_test_tiendita,y_pred_tiendita))
```

```
C:\ProgramData\Anaconda3\envs\bigdata\lib\site-
packages\sklearn\linear_model\_base.py:148: FutureWarning: 'normalize' was
deprecated in version 1.0 and will be removed in 1.2. Please leave the normalize
parameter to its default value to silence this warning. The default behavior of
this estimator is to not do any normalization. If normalization is needed please
use sklearn.preprocessing.StandardScaler instead.
  warnings.warn(
```

[62]:
```
lr.fit(X_train,y_train_ropa)
y_pred_ropa=lr.predict(X_test)
mse_ropa["LinearReg"].append(mean_squared_error(y_test_ropa,y_pred_ropa))
mae_ropa["LinearReg"].append(mean_absolute_error(y_test_ropa,y_pred_ropa))

lasso.fit(X_train,y_train_ropa)
y_pred_ropa=lasso.predict(X_test)
mse_ropa["LassoReg"].append(mean_squared_error(y_test_ropa,y_pred_ropa))
mae_ropa["LassoReg"].append(mean_absolute_error(y_test_ropa,y_pred_ropa))

ridge.fit(X_train,y_train_ropa)
y_pred_ropa=ridge.predict(X_test)
mse_ropa["RidgeReg"].append(mean_squared_error(y_test_ropa,y_pred_ropa))
mae_ropa["RidgeReg"].append(mean_absolute_error(y_test_ropa,y_pred_ropa))

bay.fit(X_train,y_train_ropa)
y_pred_ropa=bay.predict(X_test)
mse_ropa["BayReg"].append(mean_squared_error(y_test_ropa,y_pred_ropa))
mae_ropa["BayReg"].append(mean_absolute_error(y_test_ropa,y_pred_ropa))
```

```
C:\ProgramData\Anaconda3\envs\bigdata\lib\site-
packages\sklearn\linear_model\_base.py:148: FutureWarning: 'normalize' was
deprecated in version 1.0 and will be removed in 1.2. Please leave the normalize
parameter to its default value to silence this warning. The default behavior of
this estimator is to not do any normalization. If normalization is needed please
use sklearn.preprocessing.StandardScaler instead.
  warnings.warn(
```

[63]:
```
lr.fit(X_train,y_train_tecnologia)
y_pred_tecnologia=lr.predict(X_test)
mse_tecnologia["LinearReg"].
  ↪append(mean_squared_error(y_test_ropa,y_pred_tecnologia))
mae_tecnologia["LinearReg"].
  ↪append(mean_absolute_error(y_test_ropa,y_pred_tecnologia))

lasso.fit(X_train,y_train_tecnologia)
y_pred_tecnologia=lasso.predict(X_test)
mse_tecnologia["LassoReg"].
  ↪append(mean_squared_error(y_test_tecnologia,y_pred_tecnologia))
mae_tecnologia["LassoReg"].
  ↪append(mean_absolute_error(y_test_tecnologia,y_pred_tecnologia))

ridge.fit(X_train,y_train_tecnologia)
y_pred_tecnologia=ridge.predict(X_test)
mse_tecnologia["RidgeReg"].
  ↪append(mean_squared_error(y_test_tecnologia,y_pred_tecnologia))
mae_tecnologia["RidgeReg"].
  ↪append(mean_absolute_error(y_test_tecnologia,y_pred_tecnologia))

bay.fit(X_train,y_train_tecnologia)
y_pred_tecnologia=bay.predict(X_test)
mse_tecnologia["BayReg"].
  ↪append(mean_squared_error(y_test_tecnologia,y_pred_tecnologia))
mae_tecnologia["BayReg"].
  ↪append(mean_absolute_error(y_test_tecnologia,y_pred_tecnologia))
```

```
C:\ProgramData\Anaconda3\envs\bigdata\lib\site-
packages\sklearn\linear_model\_base.py:148: FutureWarning: 'normalize' was
deprecated in version 1.0 and will be removed in 1.2. Please leave the normalize
parameter to its default value to silence this warning. The default behavior of
this estimator is to not do any normalization. If normalization is needed please
use sklearn.preprocessing.StandardScaler instead.
  warnings.warn(
```

[64]:
```
lr.fit(X_train,y_train_restaurante)
y_pred_restaurante=lr.predict(X_test)
mse_restaurante["LinearReg"].
  ↪append(mean_squared_error(y_test_restaurante,y_pred_restaurante))
mae_restaurante["LinearReg"].
  ↪append(mean_absolute_error(y_test_restaurante,y_pred_restaurante))

lasso.fit(X_train,y_train_restaurante)
y_pred_restaurante=lasso.predict(X_test)
mse_restaurante["LassoReg"].
  ↪append(mean_squared_error(y_test_restaurante,y_pred_restaurante))
```

```
mae_restaurante["LassoReg"].
  ↪append(mean_absolute_error(y_test_restaurante,y_pred_restaurante))

ridge.fit(X_train,y_train_restaurante)
y_pred_restaurante=ridge.predict(X_test)
mse_restaurante["RidgeReg"].
  ↪append(mean_squared_error(y_test_restaurante,y_pred_restaurante))
mae_restaurante["RidgeReg"].
  ↪append(mean_absolute_error(y_test_restaurante,y_pred_restaurante))

bay.fit(X_train,y_train_restaurante)
y_pred_restaurante=bay.predict(X_test)
mse_restaurante["BayReg"].
  ↪append(mean_squared_error(y_test_restaurante,y_pred_restaurante))
mae_restaurante["BayReg"].
  ↪append(mean_absolute_error(y_test_restaurante,y_pred_restaurante))
```

```
C:\ProgramData\Anaconda3\envs\bigdata\lib\site-
packages\sklearn\linear_model\_base.py:148: FutureWarning: 'normalize' was
deprecated in version 1.0 and will be removed in 1.2. Please leave the normalize
parameter to its default value to silence this warning. The default behavior of
this estimator is to not do any normalization. If normalization is needed please
use sklearn.preprocessing.StandardScaler instead.
  warnings.warn(
```

[65]:
```
lr.fit(X_train,y_train_papeleria)
y_pred_papeleria=lr.predict(X_test)
mse_papeleria["LinearReg"].
  ↪append(mean_squared_error(y_test_papeleria,y_pred_papeleria))
mae_papeleria["LinearReg"].
  ↪append(mean_absolute_error(y_test_papeleria,y_pred_papeleria))

lasso.fit(X_train,y_train_papeleria)
y_pred_papeleria=lasso.predict(X_test)
mse_papeleria["LassoReg"].
  ↪append(mean_squared_error(y_test_papeleria,y_pred_papeleria))
mae_papeleria["LassoReg"].
  ↪append(mean_absolute_error(y_test_papeleria,y_pred_papeleria))

ridge.fit(X_train,y_train_papeleria)
y_pred_papeleria=ridge.predict(X_test)
mse_papeleria["RidgeReg"].
  ↪append(mean_squared_error(y_test_papeleria,y_pred_papeleria))
mae_papeleria["RidgeReg"].
  ↪append(mean_absolute_error(y_test_papeleria,y_pred_papeleria))

bay.fit(X_train,y_train_papeleria)
```

```
y_pred_papeleria=bay.predict(X_test)
mse_papeleria["BayReg"].
  ↪append(mean_squared_error(y_test_papeleria,y_pred_papeleria))
mae_papeleria["BayReg"].
  ↪append(mean_absolute_error(y_test_papeleria,y_pred_papeleria))
```

```
C:\ProgramData\Anaconda3\envs\bigdata\lib\site-
packages\sklearn\linear_model\_base.py:148: FutureWarning: 'normalize' was
deprecated in version 1.0 and will be removed in 1.2. Please leave the normalize
parameter to its default value to silence this warning. The default behavior of
this estimator is to not do any normalization. If normalization is needed please
use sklearn.preprocessing.StandardScaler instead.
  warnings.warn(
```

[66]:
```
lr.fit(X_train,y_train_gimnasio)
y_pred_gimnasio=lr.predict(X_test)
mse_gimnasio["LinearReg"].
  ↪append(mean_squared_error(y_test_gimnasio,y_pred_gimnasio))
mae_gimnasio["LinearReg"].
  ↪append(mean_absolute_error(y_test_gimnasio,y_pred_gimnasio))

lasso.fit(X_train,y_train_gimnasio)
y_pred_gimnasio=lasso.predict(X_test)
mse_gimnasio["LassoReg"].
  ↪append(mean_squared_error(y_test_gimnasio,y_pred_gimnasio))
mae_gimnasio["LassoReg"].
  ↪append(mean_absolute_error(y_test_gimnasio,y_pred_gimnasio))

ridge.fit(X_train,y_train_gimnasio)
y_pred_gimnasio=ridge.predict(X_test)
mse_gimnasio["RidgeReg"].
  ↪append(mean_squared_error(y_test_gimnasio,y_pred_gimnasio))
mae_gimnasio["RidgeReg"].
  ↪append(mean_absolute_error(y_test_gimnasio,y_pred_gimnasio))

bay.fit(X_train,y_train_gimnasio)
y_pred_gimnasio=bay.predict(X_test)
mse_gimnasio["BayReg"].
  ↪append(mean_squared_error(y_test_gimnasio,y_pred_gimnasio))
mae_gimnasio["BayReg"].
  ↪append(mean_absolute_error(y_test_gimnasio,y_pred_gimnasio))
```

```
C:\ProgramData\Anaconda3\envs\bigdata\lib\site-
packages\sklearn\linear_model\_base.py:148: FutureWarning: 'normalize' was
deprecated in version 1.0 and will be removed in 1.2. Please leave the normalize
parameter to its default value to silence this warning. The default behavior of
this estimator is to not do any normalization. If normalization is needed please
use sklearn.preprocessing.StandardScaler instead.
```

```
  warnings.warn(
```

[67]:
```
lr.fit(X_train,y_train_estetica)
y_pred_estetica=lr.predict(X_test)
mse_estetica["LinearReg"].
 ↪append(mean_squared_error(y_test_estetica,y_pred_estetica))
mae_estetica["LinearReg"].
 ↪append(mean_absolute_error(y_test_estetica,y_pred_estetica))

lasso.fit(X_train,y_train_estetica)
y_pred_estetica=lasso.predict(X_test)
mse_estetica["LassoReg"].
 ↪append(mean_squared_error(y_test_estetica,y_pred_estetica))
mae_estetica["LassoReg"].
 ↪append(mean_absolute_error(y_test_estetica,y_pred_estetica))

ridge.fit(X_train,y_train_estetica)
y_pred_estetica=ridge.predict(X_test)
mse_estetica["RidgeReg"].
 ↪append(mean_squared_error(y_test_estetica,y_pred_estetica))
mae_estetica["RidgeReg"].
 ↪append(mean_absolute_error(y_test_estetica,y_pred_estetica))

bay.fit(X_train,y_train_estetica)
y_pred_estetica=bay.predict(X_test)
mse_estetica["BayReg"].
 ↪append(mean_squared_error(y_test_estetica,y_pred_estetica))
mae_estetica["BayReg"].
 ↪append(mean_absolute_error(y_test_estetica,y_pred_estetica))
```

```
C:\ProgramData\Anaconda3\envs\bigdata\lib\site-
packages\sklearn\linear_model\_base.py:148: FutureWarning: 'normalize' was
deprecated in version 1.0 and will be removed in 1.2. Please leave the normalize
parameter to its default value to silence this warning. The default behavior of
this estimator is to not do any normalization. If normalization is needed please
use sklearn.preprocessing.StandardScaler instead.
  warnings.warn(
```

2.2 Metrics

2.2.1 Mean Squared Error

[68]:
```
mse_tiendita
```

[68]:
```
{'LinearReg': [35657.61585345917],
 'LassoReg': [36912.58227441904],
 'RidgeReg': [37152.0359957146],
 'BayReg': [37684.88869134207]}
```

[69]:
```
mse_ropa
```

[69]:
```
{'LinearReg': [25052.732840966975],
 'LassoReg': [26363.792627370156],
 'RidgeReg': [25960.264373259553],
 'BayReg': [25643.595222735556]}
```

[70]:
```
mse_tecnologia
```

[70]:
```
{'LinearReg': [87170.93022130305],
 'LassoReg': [878.5592944242026],
 'RidgeReg': [658.4499042493185],
 'BayReg': [652.7107853290707]}
```

[71]:
```
mse_restaurante
```

[71]:
```
{'LinearReg': [1395.5642150109347],
 'LassoReg': [1787.0368450421902],
 'RidgeReg': [1588.0970995203156],
 'BayReg': [1439.317350665368]}
```

[72]:
```
mse_papeleria
```

[72]:
```
{'LinearReg': [1454.6984871132504],
 'LassoReg': [1545.3276193373777],
 'RidgeReg': [1492.015416055108],
 'BayReg': [1477.8466988618584]}
```

[73]:
```
mse_gimnasio
```

[73]:
```
{'LinearReg': [264.16333060738987],
 'LassoReg': [421.23477730753524],
 'RidgeReg': [301.2942321379171],
 'BayReg': [267.01465073402903]}
```

[74]:
```
mse_estetica
```

[74]:
```
{'LinearReg': [7285.710743154756],
 'LassoReg': [7527.819081037558],
 'RidgeReg': [8175.318319379366],
 'BayReg': [7327.386915792142]}
```

2.2.2 Mean Absolute Error

[75]:
```
mae_tiendita
```

[75]:
```
{'LinearReg': [110.44512376624772],
 'LassoReg': [111.87159460482718],
```

```
 'RidgeReg': [112.0071773573649],
 'BayReg': [112.54357581753126]}
```

[76]:
```
mae_ropa
```

[76]:
```
{'LinearReg': [57.46285959040401],
 'LassoReg': [57.923206292240344],
 'RidgeReg': [57.67524694340095],
 'BayReg': [57.54301060765373]}
```

[77]:
```
mae_tecnologia
```

[77]:
```
{'LinearReg': [126.98485291015896],
 'LassoReg': [10.059657468353883],
 'RidgeReg': [9.169994155143872],
 'BayReg': [9.16445555611935]}
```

[78]:
```
mae_restaurante
```

[78]:
```
{'LinearReg': [19.212932075784966],
 'LassoReg': [21.189936761942633],
 'RidgeReg': [20.199525658495592],
 'BayReg': [19.44356557311098]}
```

[79]:
```
mae_papeleria
```

[79]:
```
{'LinearReg': [23.86921034632598],
 'LassoReg': [23.937361712086798],
 'RidgeReg': [23.894026372691343],
 'BayReg': [23.885502476486117]}
```

[80]:
```
mae_gimnasio
```

[80]:
```
{'LinearReg': [8.936605993757068],
 'LassoReg': [10.466624866839783],
 'RidgeReg': [9.381004010619353],
 'BayReg': [8.97260444456699]}
```

[81]:
```
mae_estetica
```

[81]:
```
{'LinearReg': [48.99451616437412],
 'LassoReg': [50.45332614371714],
 'RidgeReg': [53.215896987027335],
 'BayReg': [49.253786880745956]}
```

[ ]:
```

```