

Tarea 1

Sara Luz e Iván

Trabaja con tu equipo en los siguientes ejercicios:

Tarea 1. Busca todos los restaurantes de las colonias de la Zona Condesa (Condesa, Hipódromo e Hipódromo Condesa) utilizando dos fuentes: INEGI DENUE y OSM.

Nota: puedes descargar o armar la BD utilizando un editor de texto o Excel (no utilices ninguna otra); si te sirve, puedes buscar coordenadas utilizando Google Maps.

Descarga de la base de datos

Se utilizaron las siguientes dos ligas:

- [DENUE CDMX 2023/11](#)
- Open Street Maps (Se bajo un geojson y se convirtió a csv)

Carga de datos

```
datos_denue <- read.csv('./Datos/denue_09_csv/conjunto_de_datos/denue_inegi_09_.csv')
glimpse(datos_denue)
```

Rows: 475,331

Columns: 42

```
$ id          <int> 8612487, 6694989, 6430905, 6858144, 9321108, 9332233, 61448~
$ clee        <chr> "09007238222000124000000000U8", "09012237312000081000000000~
$ nom_estab   <chr> "FRIGARZA", "FRO INGENIEROS", "FRI O CAL", "FRICA CONSTRUCC~
$ raz_social  <chr> "FRIGARZA SA DE CV", "FRO INGENIEROS SA DE CV", "FRI O CAL ~
$ codigo_act  <int> 238222, 237312, 238222, 236211, 236112, 238210, 238910, 236~
```

```

$ nombre_act <chr> "Instalaciones de sistemas centrales de aire acondicionado ~
$ per_ocu <chr> "0 a 5 personas", "0 a 5 personas", "11 a 30 personas", "25~
$ tipo_vial <chr> "EJE VIAL", "CALLE", "AVENIDA", "AVENIDA", "", "AVENIDA", "~
$ nom_vial <chr> "5 ORIENTE (AVENIDA JAVIER ROJO GOMEZ)", "CLUB AMURICA", "N~
$ tipo_v_e_1 <chr> "AVENIDA", "CALLE", "CALLE", "AVENIDA", "", "CALLE", "CALLE~
$ nom_v_e_1 <chr> "MICHOACAN", "SAN JUAN DE DIOS", "BAJIO", "NUEVO LE?N", "",~
$ tipo_v_e_2 <chr> "EJE VIAL", "CALLE", "EJE VIAL", "AVENIDA", "", "CALLE", "C~
$ nom_v_e_2 <chr> "6 SUR (AVENIDA JALISCO)", "EMILIANO ZAPATA", "4 SUR (TEHUA~
$ tipo_v_e_3 <chr> "CALLE", "CALLE", "CALLE", "AVENIDA", "", "CALLE", "AVENIDA~
$ nom_v_e_3 <chr> "HERMENEGILDO GALEANA", "CLUB DEPORTIVO ATLAS", "CULIAC\xcl~
$ numero_ext <int> 546, 32, 260, 72, 47, 1804, 5591, 723, 118, 281, 125, 452, ~
$ letra_ext <chr> "", "", "", "B", "", "EDIFICIO", "", "", "", "PH", "", "", ~
$ edificio <chr> "", "", "", "", "", "", "SINNOMBRE", "", "", "", "", "", ""~
$ edificio_e <chr> "", "", "", "", "", "", "PLANTA BAJA", "", "", "", "", "", ~
$ numero_int <int> NA, 1, 0, 0, 4, 104, 0, NA, 0, NA, 0, NA, NA, 202, 2, 0, 3,~
$ letra_int <chr> "", "", "", "", "C", "PISO 1", "", "", "E", "", "", "LOCAL ~
$ tipo_asent <chr> "COLONIA", "COLONIA", "COLONIA", "COLONIA", "COLONIA", "COL~
$ nomb_asent <chr> "GUADALUPE DEL MORAL", "VILLA LAZARO CARDENAS", "HIPODROMO"~
$ tipoCenCom <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", ~
$ nom_CenCom <chr> "", "", "", "", "", "EDIF", "COLONIA EL MOLINITO", "", "", ~
$ num_local <chr> "", "", "", "", "", "", "0", "", "", "", "", "", "", "", ""~
$ cod_postal <int> 9300, NA, 6100, 6100, NA, NA, NA, 14352, 11540, 11580, 1100~
$ cve_ent <int> 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, ~
$ entidad <chr> "Ciudad de M\xe9xico", "Ciudad de M\xe9xico", "Ciudad de M~
$ cve_mun <int> 7, 12, 15, 15, 15, 16, 4, 12, 16, 16, 16, 14, 13, 16, 2, 16~
$ municipio <chr> "Iztapalapa", "Tlalpan", "Cuauht\xe9moc", "Cuauht\xe9moc", ~
$ cve_loc <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
$ localidad <chr> "Iztapalapa", "Tlalpan", "Cuauht\xe9moc", "Cuauht\xe9moc", ~
$ ageb <chr> "3651", "045A", "1430", "1322", "145A", "0622", "0354", "02~
$ manzana <int> 3, 24, 4, 13, 24, 5, 23, 19, 7, 24, 6, 11, 1, 18, 6, 4, 19,~
$ telefono <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", ~
$ correoelec <chr> "", "", "FRIOCAL@FRIOCAL.COM.MX", "ABOYZO@FRICA.COM.MX", ""~
$ www <chr> "", "", "WWW.FRIOCAL.COM.MX", "WWW.FRICA.COM.MX", "", "WWW.~
$ tipoUniEco <chr> "Fijo", "Fijo", "Fijo", "Fijo", "Fijo", "Fijo", "Fijo", "Fi~
$ latitud <dbl> 19.36799, 19.29675, 19.40268, 19.41134, 19.40425, 19.43675,~
$ longitud <dbl> -99.08153, -99.14089, -99.17082, -99.17339, -99.16729, -99.~
$ fecha_alta <chr> "2019-11", "2016-01", "2010-07", "2019-04", "2020-11", "202~

```

```

datos_osm <- read.csv('./Datos/mygeodata/export.csv')
glimpse(datos_osm)

```

Rows: 186

Columns: 50

\$ X	<dbl> -99.16967, -99.17061, -99.17555, -99.16782, -99.~
\$ Y	<dbl> 19.41597, 19.41636, 19.40941, 19.41065, 19.40328~
\$ id	<chr> "node/471092310", "node/471093109", "node/906071~
\$ X.id	<chr> "node/471092310", "node/471093109", "node/906071~
\$ amenity	<chr> "restaurant", "restaurant", "restaurant", "resta~
\$ name	<chr> "Barracuda Diner", "La Biela", "Frankfurt", "Spe~
\$ addr.city	<chr> "", "", "Ciudad de México", "", "", "", "", "", ~
\$ addr.housenumber	<chr> "", "", "136", "241", "", "", "", "", "115", "", ~
\$ addr.street	<chr> "", "", "Avenida Tamaulipas", "Amsterdam", "", "~
\$ phone	<chr> "", "", "55-5256-0767", "55-5564-1367", "", "", ~
\$ cuisine	<chr> "", "", "", "", "mexican", "", "", "", "", "thai~
\$ diet.vegetarian	<chr> "", "", "", "", "", "", "", "only", "", "", "", ~
\$ name.es	<chr> "", "", "", "", "", "", "", "Falafelito chai bar~
\$ addr.country	<chr> "", "", "", "", "", "", "", "", "MX", "", "", ""~
\$ shop	<chr> "", "", "", "", "", "", "", "", "books", "", "", ~
\$ website	<chr> "", "", "", "", "", "", "", "", "www.pendolo.com~
\$ addr.postcode	<int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
\$ addr.housename	<chr> "", "", "", "", "", "", "", "", "", "", "", ~
\$ check_date	<chr> "", "", "", "", "", "", "", "", "", "", "", ~
\$ source	<chr> "", "", "", "", "", "", "", "", "", "", "", ~
\$ designation	<chr> "", "", "", "", "", "", "", "", "", "", "", ~
\$ opening_hours	<chr> "", "", "", "", "", "", "", "", "", "", "", ~
\$ drive_in	<chr> "", "", "", "", "", "", "", "", "", "", "", ~
\$ internet_access.fee	<chr> "", "", "", "", "", "", "", "", "", "", "", ~
\$ note	<chr> "", "", "", "", "", "", "", "", "", "", "", ~
\$ outdoor_seating	<chr> "", "", "", "", "", "", "", "", "", "", "", ~
\$ smoking	<chr> "", "", "", "", "", "", "", "", "", "", "", ~
\$ wikidata	<chr> "", "", "", "", "", "", "", "", "", "", "", ~
\$ wikimedia_commons	<chr> "", "", "", "", "", "", "", "", "", "", "", ~
\$ wikipedia	<chr> "", "", "", "", "", "", "", "", "", "", "", ~
\$ branch	<chr> "", "", "", "", "", "", "", "", "", "", "", ~
\$ delivery	<chr> "", "", "", "", "", "", "", "", "", "", "", ~
\$ takeaway	<chr> "", "", "", "", "", "", "", "", "", "", "", ~
\$ capacity	<int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
\$ email	<chr> "", "", "", "", "", "", "", "", "", "", "", ~
\$ opening_hours.covid19	<chr> "", "", "", "", "", "", "", "", "", "", "", ~
\$ diet.vegan	<chr> "", "", "", "", "", "", "", "", "", "", "", ~
\$ internet_access	<chr> "", "", "", "", "", "", "", "", "", "", "", ~
\$ name.en	<chr> "", "", "", "", "", "", "", "", "", "", "", ~
\$ brand	<chr> "", "", "", "", "", "", "", "", "", "", "", ~
\$ brand.wikidata	<chr> "", "", "", "", "", "", "", "", "", "", "", ~
\$ brand.wikipedia	<chr> "", "", "", "", "", "", "", "", "", "", "", ~

```

$ operator      <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", ~
$ description   <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", ~
$ opening_hours.signed <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", ~
$ name.ja       <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", ~
$ name.ko       <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", ~
$ mapillary     <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
$ contact.facebook <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", ~
$ contact.instagram <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", ~

```

Tarea 2. Responde las siguientes preguntas utilizando únicamente R o Excel. Explica tu razonamiento.

- ¿Cuántos restaurantes encontraste en cada fuente?

Para los datos de DENUÉ, se realizó el filtro de colonia por código postal:

- Condesa(06140),
- Hipódromo(06100),
- Hipódromo Condesa(06170)

Y filtro por código de actividad:

- 722330 Servicios de preparación de alimentos en unidades móviles
- 722412 Bares, cantinas y similares
- 72251 Servicios de preparación de alimentos y bebidas alcohólicas y no alcohólicas

```

denue1 <- datos_denue %>%
  select(cod_postal, codigo_act, nombre_act, longitud, longitud) %>%
  filter(cod_postal==06140 | cod_postal==06100 | cod_postal==06170) %>%
  filter(str_detect(codigo_act, "^722330") | str_detect(codigo_act, "^722412") | str_detect(codigo_act, "^72251"))

```

```

denue1 %>%
  summarise(tot_rest = n())

```

```

tot_rest
1      517

```

```

datos_osm %>% summarise(n())

```

```
n()
1 186
```

Hay 517 restaurantes en la base de DENUÉ y 186 en la de OSM.

- ¿Qué campos puedes conocer sobre cada restaurante, en cada fuente?

En el denue:

```
colnames(datos_denue)
```

```
[1] "id"          "clee"        "nom_estab"   "raz_social"  "codigo_act"
[6] "nombre_act"  "per_ocu"     "tipo_vial"   "nom_vial"    "tipo_v_e_1"
[11] "nom_v_e_1"   "tipo_v_e_2"  "nom_v_e_2"   "tipo_v_e_3"  "nom_v_e_3"
[16] "numero_ext"  "letra_ext"   "edificio"    "edificio_e"  "numero_int"
[21] "letra_int"   "tipo_asent"  "nomb_asent"  "tipoCenCom"  "nom_CenCom"
[26] "num_local"   "cod_postal"  "cve_ent"     "entidad"     "cve_mun"
[31] "municipio"   "cve_loc"     "localidad"   "ageb"        "manzana"
[36] "telefono"    "correoelec"  "www"         "tipoUniEco"  "latitud"
[41] "longitud"    "fecha_alta"
```

En OSM:

```
colnames(datos_osm)
```

```
[1] "X"          "Y"          "id"
[4] "X.id"       "amenity"    "name"
[7] "addr.city"  "addr.housenumber" "addr.street"
[10] "phone"      "cuisine"    "diet.vegetarian"
[13] "name.es"    "addr.country" "shop"
[16] "website"    "addr.postcode" "addr.housename"
[19] "check_date" "source"      "designation"
[22] "opening_hours" "drive_in"    "internet_access.fee"
[25] "note"       "outdoor_seating" "smoking"
[28] "wikidata"   "wikimedia_commons" "wikipedia"
[31] "branch"     "delivery"    "takeaway"
[34] "capacity"   "email"       "opening_hours.covid19"
[37] "diet.vegan" "internet_access" "name.en"
[40] "brand"      "brand.wikidata" "brand.wikipedia"
[43] "operator"   "description"  "opening_hours.signed"
[46] "name.ja"    "name.ko"     "mapillary"
[49] "contact.facebook" "contact.instagram"
```

- ¿Cuántas taquerías encontraste en cada fuente?

En el DENUÉ:

```
denue1 %>%
  select(nombre_act) %>%
  filter(str_detect(nombre_act, "taco") |
         str_detect(nombre_act, "taque")) %>%
  summarise(n())
```

```
n()
1 54
```

En OSM:

```
datos_osm %>%
  select(name) %>%
  filter(str_detect(name, regex("taco", ignore_case = TRUE)) |
         str_detect(name, regex("taque", ignore_case = TRUE)) |
         str_detect(name, regex("tako", ignore_case = TRUE))) #>%
```

```
      name
1   Taco Game
2  Señor Taco
3 Tizoncito Tacos
4 Taquería La Negra
5   Taco Naco
6 Taqueria el Jarocho
7 Taqueria El Greco
8   Taco Naco
```

```
# summarise(n())
```

- Determina la dirección exacta (en texto, no coordenadas) de cada lugar, según la fuente. (calle + número + colonia + código postal + alcaldía + estado).

En el DENUÉ:

```
denue2 <- datos_denue %>%
  select(tipo_vial, nom_vial,
         tipo_v_e_1, nom_v_e_1,
```

```

        tipo_v_e_2, nom_v_e_2,
        tipo_v_e_3, nom_v_e_3,
        numero_ext, letra_ext,
        edificio, edificio_e,
        numero_int, letra_int,
        tipo_asent, nomb_asent,
        cod_postal,
        municipio, entidad)
glimpse(denue2)

```

Rows: 475,331

Columns: 19

```

$ tipo_vial <chr> "EJE VIAL", "CALLE", "AVENIDA", "AVENIDA", "", "AVENIDA", "~
$ nom_vial <chr> "5 ORIENTE (AVENIDA JAVIER ROJO GOMEZ)", "CLUB AMURICA", "N~
$ tipo_v_e_1 <chr> "AVENIDA", "CALLE", "CALLE", "AVENIDA", "", "CALLE", "CALLE~
$ nom_v_e_1 <chr> "MICHOACAN", "SAN JUAN DE DIOS", "BAJIO", "NUEVO LE?N", "",~
$ tipo_v_e_2 <chr> "EJE VIAL", "CALLE", "EJE VIAL", "AVENIDA", "", "CALLE", "C~
$ nom_v_e_2 <chr> "6 SUR (AVENIDA JALISCO)", "EMILIANO ZAPATA", "4 SUR (TEHUA~
$ tipo_v_e_3 <chr> "CALLE", "CALLE", "CALLE", "AVENIDA", "", "CALLE", "AVENIDA~
$ nom_v_e_3 <chr> "HERMENEGILDO GALEANA", "CLUB DEPORTIVO ATLAS", "CULIAC\x1~
$ numero_ext <int> 546, 32, 260, 72, 47, 1804, 5591, 723, 118, 281, 125, 452, ~
$ letra_ext <chr> "", "", "", "B", "", "EDIFICIO", "", "", "", "PH", "", "", ~
$ edificio <chr> "", "", "", "", "", "", "SINNOMBRE", "", "", "", "", "", ""~
$ edificio_e <chr> "", "", "", "", "", "", "PLANTA BAJA", "", "", "", "", "", ~
$ numero_int <int> NA, 1, 0, 0, 4, 104, 0, NA, 0, NA, 0, NA, NA, 202, 2, 0, 3,~
$ letra_int <chr> "", "", "", "", "C", "PISO 1", "", "", "E", "", "", "LOCAL ~
$ tipo_asent <chr> "COLONIA", "COLONIA", "COLONIA", "COLONIA", "COLONIA", "COL~
$ nomb_asent <chr> "GUADALUPE DEL MORAL", "VILLA LAZARO CARDENAS", "HIPODROMO"~
$ cod_postal <int> 9300, NA, 6100, 6100, NA, NA, NA, 14352, 11540, 11580, 1100~
$ municipio <chr> "Iztapalapa", "Tlalpan", "Cuauht\xe9moc", "Cuauht\xe9moc", ~
$ entidad <chr> "Ciudad de M\xe9xico", "Ciudad de M\xe9xico", "Ciudad de M~

```

En el OSM:

```

osm1 <- datos_osm %>%
  select(addr.city,
         addr.housenumber,
         addr.street,
         addr.country,
         addr.postcode,
         addr.housename)

```

```
glimpse(osm1)
```

Rows: 186

Columns: 6

```
$ addr.city      <chr> "", "", "Ciudad de México", "", "", "", "", "", "", "Ciud~
$ addr.housenumber <chr> "", "", "136", "241", "", "", "", "", "", "115", "", "", ~
$ addr.street    <chr> "", "", "Avenida Tamaulipas", "Amsterdam", "", "", "", ~
$ addr.country   <chr> "", "", "", "", "", "", "", "", "", "MX", "", "", "", "", ~
$ addr.postcode  <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, 6~
$ addr.housename <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", ""
```

- Si tuvieras que elegir entre una de las fuentes para hacer un estudio de mercado, ¿cuál escogerías, por qué?

DENUE, porque está más completo, tiene más información, y porque en el de OSM carecemos de la pericia para hacer un polígono de sólo las tres colonias mencionadas, en este caso se hizo con un rectángulo que incluía a las colonias y un poco de las zonas aledañas.

- ¿Podrías determinar en qué AGEB se encuentra cada restaurante?

Los códigos de los AGEBs para las zonas urbanas se determinan como: EEMMMLLLLAAA-ANNN - 2 dígitos de estado (EE) - 3 de municipio (MMM) - 4 para localidad (LLLL) - AGEB 3 dígitos un guión y un dígito verificador (AAA-A) - 3 para manzana (NNN)

Salvo por la parte de AGEB (AAA-A), se podría construir el código sólo para los restaurantes de la base del DENUE que está más completa.

Tarea 3. Por favor, revisa el siguiente material:

- Usos y técnicas para análisis de datos: Lee la nota de Forbes sobre limpieza de datos, [liga](#).
- Ética y privacidad - Cambridge Analytica: [liga1](#) y [liga2](#).

Entrega en el folder de Equipo:

- Arma un documento de Word con las respuestas y explicación de la sección 1.
- Sube la BD y los códigos o archivos que utilizaste para responder las preguntas 1 y 2.
- De la sección 3: no deben entregar nada, pero sí deben revisar esto. En las siguientes sesiones asumiré que todos lo revisaron. Si hay alguna duda, podemos platicar la siguiente sesión o me pueden escribir al correo.