

# Sector ideal para emprender en el Distrito Nacional, Santo Domingo

## Introducción

Luego de la pandemia del virus Covid-19, y las medidas de restricciones implementadas en muchos países para detener el contagio de esta enfermedad, muchos negocios, especialmente pequeñas y medianas empresas se vieron afectados. En Santo Domingo, la Capital de la Republica Dominicana, esto no fue la excepción. Aun así, gracias a la disminución en la tasa de contagios y en la flexibilización de las medidas contra el Covid-19, se ha evidenciado una recuperación en la economía en la Ciudad de Santo Domingo, especialmente en el Distrito Nacional. Con esta recuperación han aumentado las oportunidades de emprender en esta ciudad.

A pesar de que la recuperación después del COVID-19 en el 2020 Republica Dominicana estuvo en el Ranking 112 de países con facilidades para iniciar un negocio, por lo tanto, no es fácil iniciar un negocio en el país y, asimismo, no cualquier sector puede ser elegido para emprender con la apertura de un negocio.

Este proyecto busca encontrar cuales son los barrios y el sector económico más ideales dentro del Distrito Nacional para abrir un negocio tomando en cuenta características de rentabilidad de negocios ya existentes en el Distrito Nacional.

## Datos

Los datos que se utilizaran para resolver el problema provienen principalmente de 4 bases de datos:

- Listado de barrios del Distrito Nacional y su Código Postal.
- Encuesta Nacional de Actividad Económica del Distrito Nacional, para obtener el promedio de rentabilidad, gastos y utilidad por actividad económica.
- Dataset a crear los establecimientos de los diferentes sectores del Distrito Nacional.

La preparación de cada dataset se hará de manera separada. El listado de barrios del Distrito Nacional y su población será obtenida de la tabla de sectores y códigos postales utilizando la librería de Python BeautifulSoup: [https://es.wikipedia.org/wiki/Distrito\\_Nacional](https://es.wikipedia.org/wiki/Distrito_Nacional)

Luego, se importarán la Encuesta Nacional de Actividad Económica del Distrito Nacional, para obtener el promedio de rentabilidad, gastos y utilidad por actividad económica. Estas tablas se unirán para obtener el promedio de rentabilidad gastos, utilidad y cantidad de empresas registradas en el DN de cada actividad económica.

Los establecimientos de los diferentes sectores del D, y sus tipos, serán generados por medio de Foursquare API.

## Metodología

Primero crearemos una tabla con todos los sectores/barrios del Distrito Nacional, Santo Domingo extrayendo estos datos de la tabla de sectores del portal de Wikipedia.

```
df=pd.read_html(str(indiatable))
# convert list to dataframe
df=pd.DataFrame(df[0])
print(df.head())
```

	Sectores	Población
0	24 de abril	53870
1	27 de febrero	20000
2	30 de mayo	5904
3	Altos de Arroyo Hondo	19617
4	Arroyo Manzano	5894

Luego importaremos desde un Excel las informaciones descargadas sobre cada sector y sus códigos postales.

```
df_data_0.head()
```

```
1]:
```

	Sectores	cod_postal
0	Cristo Rey	10501
1	Domingo Savio	10304
2	Ensanche Capotillo	10407
3	Nuevo Arroyo Hondo	10507
4	Palma Real	10702

Además, importaremos los datos de promedio de gastos, utilidad, consumo y rentabilidad por sector económico y categoría.

	categories	Actividad_Ec	Prom_Gastos	Prom_Utilidad	Prom_consumo	Prom_Rentabilidad
0	Italian Restaurant	Actividades de alojamiento y de servicio de comidas	3.754133e+08	-1.750524e+08	2.417025e+08	-0.817372
1	Bakery	Actividades de alojamiento y de servicio de comidas	3.754133e+08	-1.750524e+08	2.417025e+08	-0.817372
2	Shopping Mall	Comercio al por mayor y al por menor; reparación de vehículos automotores y motocicletas	1.512540e+09	8.503943e+07	1.352839e+08	-0.249996
3	Shopping Mall	Comercio al por mayor y al por menor; reparación de vehículos automotores y motocicletas	1.512540e+09	8.503943e+07	1.352839e+08	-0.249996
4	Food and Beverage Retail	Comercio al por mayor y al por menor; reparación de vehículos automotores y motocicletas	1.512540e+09	8.503943e+07	1.352839e+08	-0.249996

Por otro lado, utilizando la latitud y longitud del Distrito Nacional, utilizaremos el servicio de API de Fourquare para obtener información sobre los diferentes negocios del DN.

Normalizaremos los resultados y nos quedaremos con las informaciones de nombres del establecimiento, categoría, latitud, longitud y código postal.

	name	categories	geocodes.main.latitude	geocodes.main.longitude	location.postcode
0	Bottega Fratelli	Italian Restaurant	18.480267	-69.932012	10122
1	Delacasa Pastelería	Bakery	18.478850	-69.934236	10122
2	Galería 360	Shopping Mall	18.484544	-69.936048	10510
3	Ágora Mall	Shopping Mall	18.483764	-69.939714	10122
4	Licormart, Cash & Carry	Food and Beverage Retail	18.483459	-69.934425	NaN

Con esta información agregaremos los datos de sector/barrio a este dataframe:

```

venues_df = pd.merge(venues_df,
                    df_data_0,
                    left_on = 'location.postcode',
                    right_on = 'cod_postal' ,
                    how = 'left')
del venues_df['cod_postal']
venues_df[
7]:

```

	name	categories	geocodes.main.latitude	geocodes.main.longitude	location.postcode	Sectores
0	Bottega Fratelli	Italian Restaurant	18.480267	-69.932012	10122	Ensanche Naco
1	Delacasa Pastelería	Bakery	18.478850	-69.934236	10122	Ensanche Naco
2	Galería 360	Shopping Mall	18.484544	-69.936048	10510	Viejo Arroyo Hondo
3	Ágora Mall	Shopping Mall	18.483764	-69.939714	10122	Ensanche Naco
4	Licormart, Cash & Carry	Food and Beverage Retail	18.483459	-69.934425	0	NaN
5	Wendy's	Fast Food Restaurant	18.475239	-69.926743	10122	Ensanche Naco

Luego, agregaremos las informaciones de gastos, utilidad, consumo y rentabilidad dependiendo en la categoría en que se encuentra cada negocio.

	name	categories	geocodes.main.latitude	geocodes.main.longitude	location.postcode	Sectores	Actividad_Ec	Prom_Gastos	Prom_Utilidad	Prom_consumo	Prom_Rentabilidad
0	Bottega Fratelli	Italian Restaurant	18.480267	-69.932012	10122	Ensanche Naco	Actividades de alojamiento y de servicio de comidas	3.754133e+08	-1.750524e+08	2.417025e+08	-0.817372
1	Bottega Fratelli	Italian Restaurant	18.480267	-69.932012	10122	Ensanche Naco	Actividades de alojamiento y de servicio de comidas	3.754133e+08	-1.750524e+08	2.417025e+08	-0.817372
2	Bottega Fratelli	Italian Restaurant	18.480267	-69.932012	10122	Ensanche Naco	Actividades de alojamiento y de servicio de comidas	3.754133e+08	-1.750524e+08	2.417025e+08	-0.817372
3	Bottega Fratelli	Italian Restaurant	18.480267	-69.932012	10122	Ensanche Naco	Actividades de alojamiento y de servicio de comidas	3.754133e+08	-1.750524e+08	2.417025e+08	-0.817372
4	Delacasa Pastelería	Bakery	18.478850	-69.934236	10122	Ensanche Naco	Actividades de alojamiento y de servicio de comidas	3.754133e+08	-1.750524e+08	2.417025e+08	-0.817372

Para el modelado agrupamos los negocios por sectores y calculamos sus promedios.

```

num_top_venues = 5

for hood in downtown_grouped['Sector']:
    print("----"+hood+"----")
    temp = downtown_grouped[downtown_grouped['Sector'] == hood].T.reset_index()
    temp.columns = ['venue', 'freq']

    temp = temp.iloc[1:]
    temp['freq'] = round(temp['freq'].astype(float), 2)
    temp = temp.sort_values('freq', ascending=False).reset_index(drop=True)
    print(temp.head(num_top_venues))
    print('\n')

```

```

----El Millón----
      venue  freq
0  Café    1.0
1  Automotive Retail  0.0
2  Juice Bar    0.0
3  Tobacco Store  0.0
4  Tennis Court  0.0

```

...

Se generaron los 10 principales negocios por categoría de cada sector:

	Sectores	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	El Millón	Café	Automotive Retail	Juice Bar	Tobacco Store	Tennis Court	Sushi Restaurant	Spanish Restaurant	Shopping Mall	Restaurant	Pie Shop
1	Ensanche Naco	Italian Restaurant	Mediterranean Restaurant	Restaurant	Fast Food Restaurant	Ice Cream Parlor	Grocery Store / Supermarket	Lounge	Bakery	Shopping Mall	Café
2	La Esperilla	Grocery Store / Supermarket	Automotive Retail	Juice Bar	Tobacco Store	Tennis Court	Sushi Restaurant	Spanish Restaurant	Shopping Mall	Restaurant	Pie Shop
3	La Julia	Spanish Restaurant	Automotive Retail	Bakery	Tobacco Store	Tennis Court	Sushi Restaurant	Shopping Mall	Restaurant	Pie Shop	Peruvian Restaurant

También, realizamos el K-mean clustering, definiendo la agrupación en 9 clusters.

```

k = 9

X = downtown_grouped.drop('Sector', axis = 1)

kmeans = KMeans(n_clusters = k, random_state=0)
kmeans.fit(X)

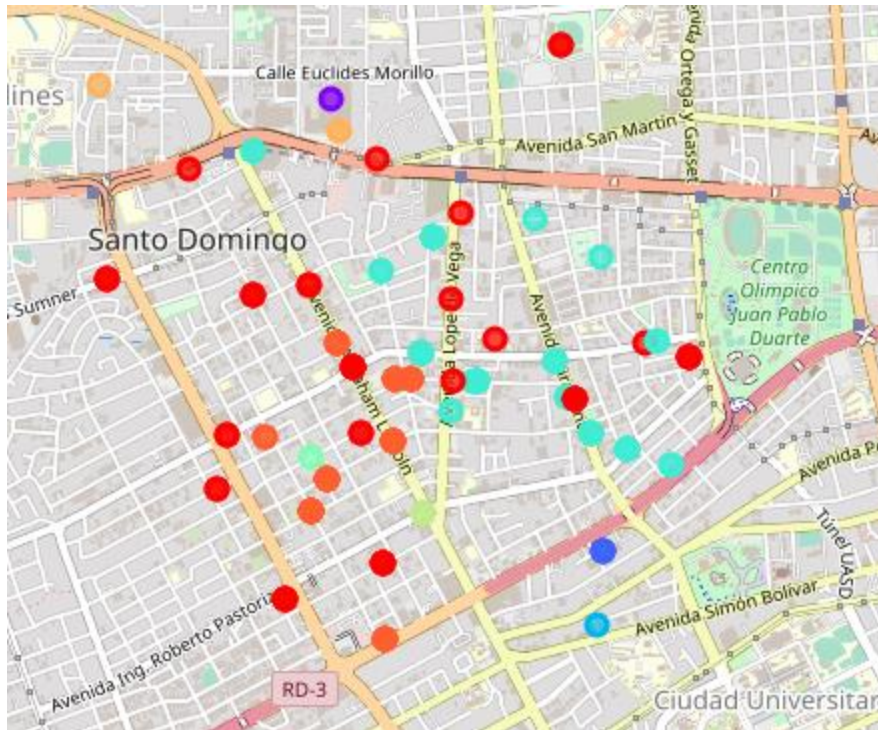
```

```

In [ ]: KMeans
        KMeans(n_clusters=9, random_state=0)

```

Por último, y para tener una visualización de nuestros resultados, realizamos un gráfico del Distrito Nacional con los diferentes clústeres:



## Resultados

El objetivo de este proyecto es poder ayudar a las personas a emprender un negocio en el Distrito Nacional, Santo Domingo, República Dominicana, basándonos en su ubicación y tipo de negocio, es decir, su actividad económica. Con los resultados obtenidos las personas se pudieran interesar en localizarse en el sector con una mayor cantidad de negocios, en este caso la mejor opción sería el clúster 4 en donde el negocio que se repite más frecuente mente son los restaurantes italianos. Si el interés es abrir un restaurante, otra buena opción es el clúster 8 donde el negocio más común son los restaurantes mediterráneos. Los negocios de restaurantes caen dentro de la actividad económica “Actividades de alojamiento y de servicio de comidas” él es un negocio que conlleva en promedio menores gastos recurrentes que los otros negocios más comunes de estos sectores. Por otro lado, si el interés es abrir algún negocio comercial, que, aunque su sector económico conlleve mayores gastos tienen una rentabilidad promedio más alta que los otros negocios, lo ideal es iniciar con este negocio en los clústeres 2 y 7, ya que los negocios que se repiten con más frecuencia alrededor de estos sectores pertenecen a la actividad económica de comercio.

## Conclusión

Este proyecto ayuda a las personas a comprender mejor los sectores del Distrito Nacional, Santo Domingo respecto a los lugares más comunes de ese Sector. Siempre es muy importante hacer uso de los datos antes de tomar una decisión tan importante como la de abrir un negocio, tanto para investigar el sector ideal en donde abrir ese negocio como qué tipo de negocio sería ideal para emprender. El futuro de este proyecto pretende incluir otros factores como los niveles de seguridad de cada sector, para poder seleccionar el sector ideal en función de los negocios más comunes, los gastos recurrentes y la seguridad de cada sector.