Covid-19 evolution in Santander, Colombia
Coursera Captstone project for IBM data science professional certification

By Jhon Padilla

Data background

In Colombia, health ministry has a web page with open data at www.data.gov.co. In that page there are data from covid-19 pandemic in all regions and cities within Colombia. In this project, covid-19 data published in this page was taken and studied. The number of records in such data base is bigger because it has all cases in Colombia from february of past year, then only Santander region data was taken for this study. Such data were saved on a CSV file and it is loaded and studied in the next steps. This data set have parameters such as: case ID (unique identifier given in order of appearance), diagnosis date, Department name, City name, age, gender, state (Recovery, active, dead), ethnicity, and others. The data set has 174145 rows and 23 columns.

Here are the column names (in spanish) and their types and english translation:

Spanish name	Type	English translation
fecha reporte web	object	Web report date
ID de caso	int64	Case ID number
Fecha de notificación	object	Notification date
Código DIVIPOLA departamento	int64	Department Code
Nombre departamento	object	Department Name
Código DIVIPOLA municipio	int64	City code
Nombre municipio	object	City name
Edad	int64	Age
Unidad de medida de edad	int64	Age Cuantity
Sexo	object	Gender
Tipo de contagio	object	Type of contagion
Ubicación del caso	object	Case location
Estado	object	State
Código ISO del país	float64	Country ISO code
Nombre del país	object	Country name
Recuperado	object	Recovery
Fecha de inicio de síntomas	object	Symptom onset date
Fecha de muerte	object	Death date
Fecha de diagnóstico	object	Diagnosis date
Fecha de recuperación	object	Recovery date
Tipo de recuperación	object	Recovery type
Pertenencia étnica	float64	Ethnicity
Nombre del grupo étnico	object	Ethnic group

Finally, as people who have relatives sick from covid need to know what venues are near clinics and hospitals, then, a map with venues near to main to these healthcare centers is

provided for such situations. For this map it is necessary to obtain data from Foursquare with venues near to healthcare centers area.