Analyzing impact of Covid-19 Pandemic and the effect of quarantine in Argentina

Applied Data Science Capstone

Martín Alejandro Heredia

Introduction

The Problem

From the beginning of Covid-19 Pandemic, people's life has changed drastically due to the fast and exponential growth of cases, which leads governments to take decisions on how to prevent the expansion rate of the disease, health system saturation and economic impact.

The common denominator is quarantine. This helps to reduce infection rate, giving more time to governments to prepare health system. In opposition to this positive consequence of quarantine is the economic, social and industrial impact that quarantine has on a country. This positive and negative effect of quarantine makes it a "trade-off" solution.

The analysis of growth of cases, distribution of disease on the territory, impact of quarantine, among others; is fundamental for governments to take decisions. Also, viewing the evolution of disease and its influence on indicators such like economics or industrial ones, gives governments tools to evaluate and reorganize its own decisions.

This study is focused on Argentinian case. First, statistical data of Covid-19 on Argentina will be analyzed: how numbers of cases are distributed on its geography and where are the most affected cities. Then, the effect of quarantine in other areas (such like transport, industry, economic) will be studied and compared with the same period of the previous year.

The Data

To analyze the effect of Covid-19 on the territory, the data from the following website of Argentinian Government will be used:

https://sisa.msal.gov.ar/datos/descargas/covid-19/files/Covid19Casos.csv

```
In [3]: url='https://sisa.msal.gov.ar/datos/descargas/covid-19/files/Covid19Casos.csv'
cases=pd.read_csv(url,encoding='utf-16')
cases.head()
```

Out[3]:

	id_evento_caso	sexo	edad	edad_años_meses	residencia_pais_nombre	residencia_provincia_nombre	residencia_departa
0	672064	M	52.0	Años	Argentina	Buenos Aires	
1	717629	F	46.0	Años	Argentina	Buenos Aires	
2	717926	F	41.0	Años	Argentina	CABA	
3	718029	F	52.0	Años	Argentina	Buenos Aires	
4	718055	F	34.0	Años	Argentina	CABA	SIN

5 rows × 25 columns

All the categories (columns) of this dataset are depicted below:

```
In [13]: for category in cases.columns.sort_values():
    print(category)

Clasificacion
asistencia_respiratoria_mecanica
carga_provincia_id
carga_provincia_nombre
clasificacion_resumen
cuidado_intensivo
edad
edad_años_meses
fallecido
fecha_apertura
fecha_cui_intensivo
fecha_diagnostico
fecha_fallecimiento
fecha_inicio_sintomas
fecha_internacion
id_evento_caso
origen_financiamiento
residencia_departamento_id
residencia_departamento_nombre
residencia_pais_nombre
residencia_provincia_id
residencia_provincia_nombre
sepi_apertura
sexo
ultima_actualizacion
```

This columns provides information such like:

- Age and gender of the patients

- Dates (diagnosis date, hospitalization date, date of death, among others)

- Geographical information (province name, department name)

- Classification (active, suspect, cured, among others)

- ID of each case

With this dataset, we can visualize the impact of this disease with indicators such like:

- Total of cases
- Total of deaths
- Total of recovered patients
- Total of active cases
- Total number of hospitalized patients
- Total of cases on intensive therapy

In addition to this data, the use of Foursquare will help us to visualize the distribution of cases on the country.

After this preliminary study, the work will be focused on analyze and visualize the impact of quarantine on the following areas:

- Economic activity:
  https://www.indec.gob.ar/ftp/cuadros/economia/sh_emae_mensual_base2004.xls
  https://www.indec.gob.ar/ftp/cuadros/economia/sh_emae_actividad_base2004.xls

- Industrial activity:
  https://www.indec.gob.ar/ftp/cuadros/economia/sh_ipi_manufacturero_2020.xls

- Transport (railway):
  https://servicios.transporte.gob.ar/gobierno_abierto/descargar.php?t=trenes&d=pasajeros

Next, each of the previous datasets will be introduced.

Economics

There are two datasets in this category. Both of them shows the evolution of EMAE estimator ("Mensual Estimator of Economic Activity" in English). The first dataset, shows the evolution of this estimator, related to economic activity of year 2004:

	A	B	C	D	E	F	G
205	Abril	153,3	0,1	147,2	-3,2	147,8	-0,6
206	Mayo	159,6	-4,7	143,7	-2,4	146,8	-0,6
207	Junio	149,0	-6,5	142,2	-1,0	145,9	-0,6
208	Julio	145,6	-2,9	142,9	0,5	145,0	-0,6
209	Agosto	146,2	-1,8	145,9	2,1	144,2	-0,5
210	Septiembre	137,8	-6,3	143,3	-1,8	143,6	-0,4
211	Octubre	143,0	-4,0	144,2	0,7	143,1	-0,3
212	Noviembre	140,9	-7,3	141,9	-1,6	142,7	-0,2
213	Diciembre	136,8	-7,0	142,1	0,1	142,5	-0,2
214							
215	**2019**						
216	Enero	134,9	-5,7	143,4	0,9	142,4	-0,1
217	Febrero	132,4	-4,7	143,4	0,0	142,4	0,0
218	Marzo	144,8	-7,0	142,0	-1,0	142,4	0,0
219	Abril	150,7	-1,7	142,3	0,2	142,5	0,0
220	Mayo	162,8	2,0	142,3	0,0	142,5	0,0
221	Junio	148,4	-0,4	141,7	-0,4	142,5	0,0
222	Julio	146,2	0,4	144,7	2,1	142,4	-0,1
223	Agosto	140,9	-3,6	144,0	-0,4	142,2	-0,1
224	Septiembre	134,9	-2,1	141,6	-1,7	142,0	-0,2
225	Octubre	141,6	-0,9	143,9	1,6	141,7	-0,2
226	Noviembre	137,9	-2,2	141,3	-1,8	141,3	-0,3
227	Diciembre	136,4	-0,3	141,1	-0,1	140,8	-0,3
228							
229	**2020**						
230	Enero	132,3	-1,9	141,1	0,0	140,3	-0,3
231	Febrero	129,2	-2,4	139,2	-1,3	139,8	-0,4
232	Marzo	128,2	-11,5	125,6	-9,8	139,3	-0,3
233							
234	**Fuente:** INDEC.						

The second one, shows the same estimator for different activities:

Período	A - Agricultura, ganadería, caza y silvicultura	B - Pesca	C - Explotación de minas y canteras	D - Industria manufacturera	E - Electricidad, gas y agua	F - Construcción	G - Comercio mayorista, minorista y reparaciones	H - Hoteles y restaurantes	I - Transporte y comunicaciones	J - Intermediación financiera	K - Actividades inmobiliarias, empresariales y de alquiler	L - Administración pública y defensa; planes de seguridad social de afiliación obligatoria	M - Enseñanza	N - Servicios sociales y de salud	O - Otras actividades de servicios comunitarios, sociales y personales
Mayo	224,3	130,9	91,5	135,4	142,0	155,3	160,4	153,3	201,5	197,5	146,9	153,8	159,9	193,4	150,3
Junio	155,6	197,5	87,2	124,2	156,5	149,0	148,5	149,5	192,6	189,9	146,8	154,3	161,2	198,4	151,4
Julio	104,9	265,8	89,0	127,3	155,8	149,5	144,5	168,0	195,1	193,8	144,2	153,5	162,0	182,4	170,2
Agosto	72,9	292,8	88,0	132,9	149,1	161,7	154,8	165,4	192,4	200,2	144,5	152,7	162,2	181,4	150,6
Septiembre	64,5	291,2	86,5	123,2	132,8	149,2	135,2	160,5	185,7	189,8	139,8	152,8	162,6	177,8	144,6
Octubre	75,6	304,9	88,6	131,2	133,6	146,3	149,5	171,8	189,6	185,6	143,4	153,4	163,1	173,4	149,8
Noviembre	91,1	121,7	85,0	124,8	131,2	141,1	150,1	169,1	186,4	177,2	142,1	153,9	163,5	173,2	145,0
Diciembre	103,4	84,5	87,8	111,8	139,9	129,0	129,0	174,1	184,7	188,3	144,4	153,9	163,0	165,6	142,8
2019															
Enero	73,6	150,2	87,5	104,0	146,5	145,8	125,4	173,6	188,0	179,0	136,1	153,6	156,3	181,1	155,8
Febrero	77,2	164,5	81,7	104,1	138,7	150,3	131,5	164,3	180,5	169,5	137,5	153,9	156,9	171,7	147,5
Marzo	156,5	153,4	88,4	114,6	135,7	149,5	149,0	160,9	193,1	172,4	138,8	153,8	159,6	186,1	144,4
Abril	223,5	152,7	87,4	119,7	130,1	142,7	138,2	156,5	194,7	171,7	141,4	153,8	159,9	190,1	144,2
Mayo	329,0	91,5	93,0	126,5	139,4	150,7	144,7	153,4	205,5	168,6	143,9	154,5	161,4	193,7	147,0
Junio	220,4	180,9	88,4	115,9	138,4	138,6	137,1	152,4	196,6	161,7	143,5	155,0	162,0	198,7	150,9
Julio	126,2	252,5	91,6	124,5	151,8	149,1	141,4	174,5	197,5	168,2	145,5	153,8	163,8	183,0	168,6
Agosto	78,3	290,5	91,4	123,9	143,3	157,1	140,4	167,1	191,3	170,1	144,0	153,7	163,5	182,0	146,1
Septiembre	71,3	163,3	88,3	116,7	132,2	142,3	127,9	162,5	187,5	162,3	142,4	154,0	164,1	178,3	140,9
Octubre	77,6	220,7	90,9	128,5	137,0	134,1	146,6	173,5	189,2	168,9	144,5	154,6	164,5	173,8	146,2
Noviembre	92,6	94,8	86,6	118,7	136,6	130,0	140,4	170,3	186,2	162,9	142,8	154,8	164,7	173,8	144,2
Diciembre	103,0	95,8	88,1	112,8	145,5	117,9	130,5	176,4	185,4	172,0	146,3	155,4	164,4	167,0	142,8
2020															
Enero	68,0	89,7	87,7	103,2	152,3	125,4	126,1	174,1	186,8	165,2	137,2	154,2	158,1	178,5	155,1
Febrero	75,0	168,1	85,0	102,5	140,2	119,7	129,7	167,9	179,1	157,2	136,5	154,5	158,8	169,2	147,7
Marzo	143,8	78,9	85,7	96,8	144,8	80,1	132,2	111,4	164,4	165,3	128,7	152,5	157,7	168,8	126,4

Fuente: INDEC.

This two datasets will help us to compare the estimator at the first month of quarantine vs the estimator at the same month of previous years, viewing which activities were more affected by quarantine.

Industrial

This datasets shows the IPI (Industrial Production Index) of Argentina referred to the last 4 years, and it's also divided by activities:

Cuadro 3. IPI manufacturero nivel general, divisiones y subclases. Serie original, base 2004=100, en variación porcentual interanual. Años 2016-2020

Código (CLaNAE 2004)	Nivel general	15	15111	15112	15113	15130	15140	15200	15311/12/13/20	15411/12/41/42	15420/30	15491/2/3	15510/29/30/41/42/49	15521	15120/
Período	IPI Manufacturero	Alimentos y bebidas	Carne vacuna	Carne aviar	Fiambres y embutidos	Preparación de frutas, hortalizas y legumbres	Molienda de oleaginosas	Productos lácteos	Molienda de cereales	Galletitas, productos de panadería y pastas	Azúcar, productos de confitería y chocolate	Yerba mate, té y café	Gaseosas, aguas, sodas, cervezas, jugos para diluir, sidras y bebidas espirituosas	Vino	Ot prod alime
	%									%					
2016 Enero	///	///	///	///	///	///	///	///	///	///	///	///	///	///	///
Febrero	///	///	///	///	///	///	///	///	///	///	///	///	///	///	///
Marzo	///	///	///	///	///	///	///	///	///	///	///	///	///	///	///
Abril	///	///	///	///	///	///	///	///	///	///	///	///	///	///	///
Mayo	///	///	///	///	///	///	///	///	///	///	///	///	///	///	///
Junio	///	///	///	///	///	///	///	///	///	///	///	///	///	///	///
Julio	///	///	///	///	///	///	///	///	///	///	///	///	///	///	///
Agosto	///	///	///	///	///	///	///	///	///	///	///	///	///	///	///
Septiembre	///	///	///	///	///	///	///	///	///	///	///	///	///	///	///
Octubre	///	///	///	///	///	///	///	///	///	///	///	///	///	///	///
Noviembre	///	///	///	///	///	///	///	///	///	///	///	///	///	///	///
Diciembre	///	///	///	///	///	///	///	///	///	///	///	///	///	///	///
2017* Enero	-1,1	0,6	8,9	13,1	1,9	6,9	-20,8	8,0	8,6	4,8	-2,5	2,4	-3,2	-4,1	
Febrero	-8,0	-8,1	-2,7	7,8	-6,9	2,2	-18,7	-4,9	-9,7	-5,6	2,3	0,4	-8,5	-23,2	
Marzo	-0,8	1,7	9,7	11,6	6,0	3,2	3,2	3,5	-1,8	-0,9	12,2	5,8	2,6	-11,6	
Abril	-3,8	-3,0	1,9	2,4	-4,5	0,1	7,2	9,3	-12,3	-7,0	0,2	-13,3	-5,5	-18,1	
Mayo	3,9	4,1	7,9	6,3	7,2	-10,0	5,7	28,8	-8,1	1,8	7,9	9,9	2,3	-6,4	
Junio	7,6	3,9	12,0	3,3	5,8	-14,3	-6,7	11,9	-1,3	0,2	27,6	4,0	8,6	9,3	
Julio	6,9	4,8	11,9	5,8	5,0	2,2	9,9	-2,7	-5,7	4,7	19,1	2,7	0,3	1,7	
Agosto	6,9	1,1	6,1	-1,6	2,3	26,0	2,6	-5,5	-10,9	-3,4	3,2	-11,4	8,2	-13,4	
Septiembre	3,7	-2,0	7,8	-6,8	2,2	4,3	-6,2	-7,0	-9,6	-5,9	-6,9	-8,8	4,2	-14,1	

It will help us to evaluate how much industries were affected on the first month of quarantine.

Transport

This dataset shows the total amount of tickets by month and by train station for every Railway Line on the region of Buenos Aires. This information will be used to compare the first month of quarantine with same month of the previous year and see how much were railway mobility reduced:

```
In [4]: url = 'https://servicios.transporte.gob.ar/gobierno_abierto/descargar.php?t=trenes&d=pasajero
        trenes_pasajeros_df = pd.read_csv(url)
        trenes_pasajeros_df.head(10)
```

Out[4]:

	mes;linea;estacion;cantidad
0	12/2019;belgranonorte;"Boulogne Sur";189118
1	12/2019;belgranonorte;Carapachay;44049
2	12/2019;belgranonorte;"Del Viso";76691
3	12/2019;belgranonorte;"Don Torcuato";128986
4	12/2019;belgranonorte;Florida;37226
5	12/2019;belgranonorte;"Grand Bourg";220363
6	12/2019;belgranonorte;"Ing. P. Nogués";103460
7	12/2019;belgranonorte;"Los Polvorines";137590
8	12/2019;belgranonorte;Munro;98282
9	12/2019;belgranonorte;"R. S. Ortíz";29953