

Contenido

Proyecto	2
Situación Problema: Análisis de Datos	2
Requerimientos del sistema	2
1.- Base de datos	2
2.- Opciones del programa	3
3.- Opción: "Día con más casos a nivel nacional"	4
4.- Opción: "% Casos confirmados de acuerdo a la población"	6
5.- Opción: "Series de tiempo"	8
Evaluación del proyecto	9
Restricciones	9
Puntuación	10

Proyecto

Situación Problema: Análisis de Datos

Python se ha convertido en uno de los lenguajes preferidos para el área de Ciencia de Datos. “La ciencia de datos es un campo interdisciplinario que utiliza métodos, procesos, algoritmos y sistemas científicos para extraer valor de los datos. Los científicos de datos combinan una variedad de habilidades, entre ellas estadística, informática y conocimiento empresarial, para analizar datos recopilados de la web, de teléfonos inteligentes, de clientes, sensores y otras fuentes. La ciencia de datos revela tendencias y genera información que las empresas pueden utilizar para tomar mejores decisiones y crear productos y servicios más innovadores. Los datos son el cimiento de la innovación, pero su valor proviene de la información que los científicos pueden extraer y luego utilizar a partir de los mismos.”(ORACLE, 2020).

Existen sitios en internet que recopilan “datasets” (archivos de datos) de todo tipo y área de conocimiento y que ponen disponible para descarga.

Un ejemplo es <https://www.kaggle.com> (Enlaces a un sitio externo.) donde puedes obtener archivos de datos en formato csv, de muchos temas científicos, deportivos económicos, de salud, de entretenimiento, entre otros. Otros sitios donde puedes obtener datos interesantes son los espacios gubernamentales, ejemplos en México el INEGI o el sitio de datos de COVID <https://coronavirus.gob.mx/datos/> (Enlaces a un sitio externo.), o si te interesan datos a nivel global se pueden obtener de instituciones mundiales como <https://datacatalog.worldbank.org/> (Enlaces a un sitio externo.). Los anteriores son sólo algunos ejemplos, el mundo genera diariamente muchos datos.

Los reportes son una parte importante, nos muestran resúmenes o agrupaciones de los datos que también nos pueden ayudar a visualizar comportamientos o tendencias de forma sencilla. Y sí, tú puedes crear gráficas en Python de manera profesional y sencilla, así que por eso no te preocupes, la librería Matplotlib te ayudará con esa tarea, puedes ir viendo toda la gama de gráficas que tiene esta herramienta:

<https://matplotlib.org/3.1.1/gallery/index.html> (Enlaces a un sitio externo.).

Referencias:

ORACLE (2020) ¿Qué es la ciencia de datos? Consultado en:

<https://www.oracle.com/mx/data-science/what-is-data-science.html>

Requerimientos del sistema

1.- Base de datos

Puedes utilizar la base de datos con información real del COVID a nivel nacional o la de incidencia delictiva.

- d. Salir
2. Por ejemplo:

```
*****
**
** 1.- Día con más casos a nivel nacional      **
** 2.- % Casos confirmados de acuerdo a la población **
** 3.- Series de tiempo                      **
** 4.- Salir                                **
**
*****
Opción -> |
```

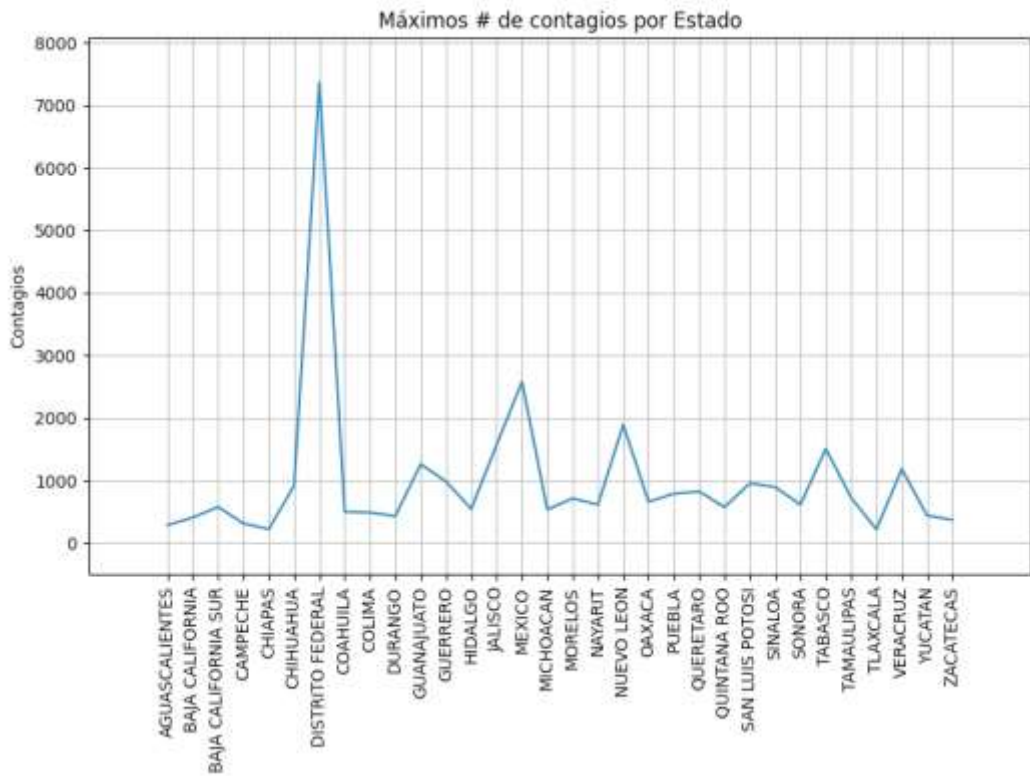
3. El menú debe cumplir con lo siguiente:
- a. Validar que la opción sea válida, en caso de teclear una opción inválida, se deberá mostrar el mensaje: “Opción inválida” y el sistema deberá mostrar nuevamente el menú de opciones.
 - b. Estar dentro de un ciclo
 - i. EL usuario decide cuando quiere salir.

3.- Opción: “Día con más casos a nivel nacional”

1. Para cada estado, el sistema debe obtener la fecha en la que se tuvo el máximo de contagios confirmados.
2. El sistema deberá mostrar una tabla con las columnas:
 - a. Estado
 - b. Fecha
 - i. Día en el cual se encuentra el máximo
 - c. Máximo
 - i. Número de contagios máximo en todo el periodo
 - d. Ejemplo

Estado	Fecha	Máximo
AGUASCALIENTES	23-11-2020	285
BAJA CALIFORNIA	21-12-2020	407
BAJA CALIFORNIA SUR	28-06-2021	577
CAMPECHE	18-08-2021	313
CHIAPAS	17-08-2021	222
CHIHUAHUA	20-10-2020	916
DISTRITO FEDERAL	11-01-2021	7,367
COAHUILA	10-11-2020	497
COLIMA	16-08-2021	489
DURANGO	27-10-2020	431
GUANAJUATO	11-01-2021	1,261
GUERRERO	03-08-2021	981
HIDALGO	16-08-2021	547
JALISCO	09-08-2021	1,571
MEXICO	26-07-2021	2,576
MICHOACAN	10-08-2021	536
MORELOS	25-01-2021	713
NAYARIT	03-08-2021	613
NUEVO LEON	09-08-2021	1,891
OAXACA	05-08-2021	657
PUEBLA	17-08-2021	790
QUERETARO	18-01-2021	823
QUINTANA ROO	03-08-2021	574
SAN LUIS POTOSI	11-08-2021	955
SINALOA	26-07-2021	895
SONORA	13-07-2020	613
TABASCO	23-08-2021	1,509
TAMAULIPAS	16-08-2021	724
TLAXCALA	11-01-2021	217
VERACRUZ	09-08-2021	1,184
YUCATAN	06-09-2021	439
ZACATECAS	10-08-2021	371
Nacional	09-08-2021	25,256

3. El sistema deberá mostrar una gráfica de líneas, tal que el eje “x” sea el estado y el eje “y” el máximo de cada estado.
 - a. Solo en la parte de la gráfica, se deberá eliminar de los datos el registro “Nacional”, ya que el número es el acumulado de todos los estados y no tiene sentido de comparación contra los estados.
 - b. Ejemplo de la gráfica:



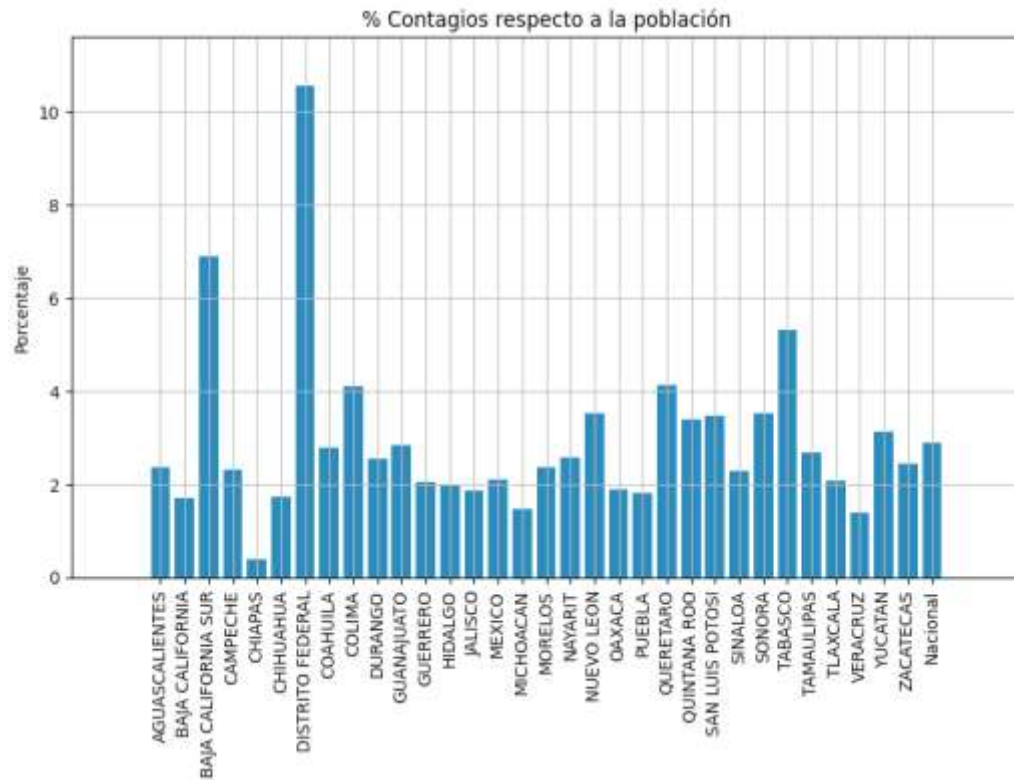
4. Una vez que el usuario cierre la gráfica y tabla de datos, el sistema lo deberá regresar al “Menú de opciones”

4.- Opción: “% Casos confirmados de acuerdo a la población”

1. Para cada estado, el sistema debe calcular:
 - a. La suma acumulada de contagios confirmados
 - b. El porcentaje correspondiente de acuerdo al total de la población de dicho estado.
2. El sistema deberá mostrar una tabla con las columnas:
 - a. Estado
 - b. Población
 - c. # Contagiados
 - d. Porcentaje
3. Ejemplo de tabla:

Estado	Población	# Contagiados	Porcentaje
AGUASCALIENTES	1,434,635	34,164	2.38
BAJA CALIFORNIA	3,634,868	62,099	1.71
BAJA CALIFORNIA SUR	804,708	55,460	6.89
CAMPECHE	1,000,617	23,131	2.31
CHIAPAS	5,730,367	22,485	0.39
CHIHUAHUA	3,801,487	65,978	1.74
DISTRITO FEDERAL	9,018,645	952,610	10.56
COAHUILA	3,218,720	89,856	2.79
COLIMA	785,153	32,183	4.1
DURANGO	1,868,996	47,634	2.55
GUANAJUATO	6,228,175	177,416	2.85
GUERRERO	3,657,048	75,135	2.05
HIDALGO	3,086,414	60,720	1.97
JALISCO	8,409,693	156,295	1.86
MEXICO	17,427,790	365,149	2.1
MICHOACAN	4,825,401	71,540	1.48
MORELOS	2,044,058	48,621	2.38
NAYARIT	1,288,571	33,127	2.57
NUEVO LEON	5,610,153	198,720	3.54
OAXACA	4,143,593	78,315	1.89
PUEBLA	6,604,451	120,773	1.83
QUERETARO	2,279,637	94,300	4.14
QUINTANA ROO	1,723,259	58,338	3.39
SAN LUIS POTOSI	2,866,142	100,165	3.49
SINALOA	3,156,674	72,185	2.29
SONORA	3,074,745	108,928	3.54
TABASCO	2,572,287	136,643	5.31
TAMAULIPAS	3,650,602	98,168	2.69
TLAXCALA	1,380,011	28,661	2.08
VERACRUZ	8,539,862	119,137	1.4
YUCATAN	2,259,098	70,918	3.14
ZACATECAS	1,666,426	40,767	2.45
Nacional	127,792,286	3,699,621	2.9

5. El sistema deberá obtener una gráfica de barras tal que el eje “x” sea el estado y el eje “y” sea el porcentaje calculado.
 - a. Ejemplo de la gráfica:



5.- Opción: “Series de tiempo”

1. El sistema debe solicitar el nombre del lugar, puede ser el “nombre del estado” o escribir “Nacional”.

- a. Ejemplo de captura:

```

*****
**
**En Lugar puedes teclear el nombre de algun estado o Nacional
**
**
*****
Lugar -> PUEBLA
  
```

2. El sistema debe validar que el nombre sea correcto, es decir, que pertenezca a la base de datos.
3. En caso de escribir un nombre incorrecto, el sistema deberá mostrar el mensaje “Lugar inválido” y volver a solicitar el nombre del lugar.
4. Solo con la información del lugar seleccionado por el usuario, el sistema debe:
 - a. Obtener la lista de meses que existe en la información
 - i. Por ejemplo las fechas: 18-02-2020,19-02-2020,20-02-2020... corresponden al mes: “02-2020”
 - b. Calcular la suma de casos confirmados por mes (solo para el lugar seleccionado por el usuario) y dejar la información en una lista.
 - c. Graficar la serie de datos correspondiente

d. Ejemplo 1



e.



f.

Evaluación del proyecto

Restricciones

- El proyecto se debe realizar en equipos de máximo 3 personas
- Solo se está permitido utilizar las librerías
 - os
 - matplotlib.pyplot
 - math
- Para la manipulación de la información se deberá utilizar listas y matrices.

- No está permitido utilizar otro tipo de estructura de datos.
- Los cálculos donde se requiera utilizar una suma, se tendrá que hacer mediante un ciclo. (for o while)
 - No está permitido realizarlo mediante alguna librería.

Puntuación

	Lista de cotejo	
Criterio	Descripción del criterio	Ponderación
Funcionalidad	* Un programa que funciona de manera correcta para todos los casos solicitados explícitamente en la descripción del problema. 1	70%
Estructura y Plan de pruebas	* El diseño de funciones para estructurar un programa y cada función tiene un propósito específico.	20%
Estilo	* En un programa usa notación de acuerdo con el estándar del lenguaje, nombres de variables representativos. Estructura los estatutos de manera ordenada (uno en cada renglón, no dejar exceso de líneas en blanco, dejar espacio entre los operadores y los operandos).	6%
	* Identifica los estándares internacionales que aplicará en sus códigos: nombres de variables representativos, estructuras de estatutos ordenados (uno en cada renglón, no exceso de líneas blancos, dejar espacios entre los operadores y uso de operandos), comentarios para aclarar funcionalidad.	4%

Estándares de programación PEP 8

<https://pep8.org/>

Funcionalidad (70 %)

Id Requerimiento	Descripción	Puntos
2	Menu de opciones Es un ciclo, cumple con los requerimientos.	4
3	“Día con más casos a nivel nacional”	16
	Muestra la tabla, incluyendo el registro “Nacional” (último) Muestra la gráfica de líneas sin el registro “Nacional” (último)	16
4	“% Casos confirmados de acuerdo a la población”	16
	Muestra tabla completa Muestra gráfica de barras completa	16
5	“Series de tiempo”	16
	Se obtienen los datos solo del lugar que teclea el usuario Se muestra gráfica solo de los datos del lugar seleccionado	16
	Total	100