

Responsabilidad Social

Informe sobre conjunto de datos sobre COVID-19

Integrantes: Mita Yagua Lesly Yaneth
Janampa Canahuire Victor Manuel
Castillo Caccire Kemely Francis
Olazabal Chávez Kate Itati
Chambi Apaza Syomira Inés
Docente: Quispe Infantes Reynaldo

Fecha de entrega: 16 de noviembre de 2021
Arequipa, Perú

Índice de Contenidos

1. Introducción	1
2. Conjunto de Datos	1
3. DataFrame	2
4. Tablas	3
4.1. Positividad a según el departamento	3
4.2. Hospitalizados por según el departamento	5
5. Gráficos	7
5.1. Positividad a según el sexo	7
5.2. Hospitalizados por según el sexo	7
5.3. Criterio de fallecidos	8
6. Repositorio	9
Referencias	10

Índice de Tablas

1. Dataset Info[1].	1
2. Positividad a según el departamento	4
3. Hospitalizados por según el departamento	6

Índice de Figuras

1. Resultado de ejecución del Código 1	2
2. Resultado de ejecución de la la columna positividad y departamento	3
3. Resultado de ejecución de la la columna hospitalizado y departamento	5
4. Gráfico sobre la cantidad de personas positivas y negativas a según su sexo	7
5. Gráfico sobre la cantidad de personas hospitalizadas por según su sexo	8
6. Gráfico de barras sobre el criterio de las personas fallecidas	8
7. Gráfico circular sobre el criterio de las personas fallecidas	9

Índice de Códigos

1. Creando DataFrame	2
--------------------------------	---

1. Introducción

En el presente informe se realizó con respecto a Responsabilidad Social, un análisis de datos con respecto al tema de -19. Se realizó una búsqueda en sobre conjuntos de datos acerca de este tema, encontrando satisfactoriamente la Plataforma de datos abiertos del Perú en su Información de Fallecidos del Sistema Informático Nacional de Defunciones del Ministerio de Salud donde se encuentran los datos de personas fallecidas.

El Sistema Informático Nacional de Defunciones (SINADEF), es el aplicativo informático que contiene los datos del fallecido, genera el certificado de defunción y el informe estadístico; incluye las defunciones fetales, las defunciones generales y los diagnósticos y antecedentes de enfermedad de causas de defunción, los datos son preliminares a la fecha.

Para la manipulación de datos, creación de tablas y gráficos se utilizó el lenguaje de programación Python junto a la biblioteca Pandas y Matplotlib para la creación de un DataFrame.

2. Conjunto de Datos

La página de donde se extrajo la Información de Fallecidos del Sistema Informático Nacional de Defunciones - SINADEF se encuentra [aquí](#) donde la información se actualiza diariamente, además de ello presenta las siguientes características que se muestran en la Tabla 3

Tabla 1: Dataset Info[1].

Field	Value
Publisher	Datos Abiertos de -19
Fecha modificada	2020-07-31
Fecha de lanzamiento	2020-05-21
Frequency	Daily
Identificador	ba538c68-c5e0-4098-985d-c5e271c973a6
License	Open Data Commons Attribution License
Language	Spanish (Peru)
Author	Ministerio de Salud
Public Access Level	Public

3. DataFrame

Para la manipulación de datos se utilizó un *DataFrame* debido a que es una estructura de datos con dos dimensiones en la cual se puede guardar datos de distintos tipos (como caracteres, enteros, valores de punto flotante, factores y más) en columnas.

También se utilizó la biblioteca *Pandas* que provee estructuras de datos, genera gráficos de alta calidad con matplotlib y se integra de buena forma con otras bibliotecas que usan arrays de NumPy (la cual es otra biblioteca de Python).

Código 1: Creando DataFrame

```
1 import pandas as pd
2 import matplotlib.pyplot as plt
3
4 data = pd.read_csv("data/falle_hos_vac.csv", sep=',')
5 print(data.dtypes)
```

```
leslym03@leslym03-Satellite-S55-B:~/Covid-Fallecidos$ python3 main.py
id_persona          int64
fecha_fallecimiento object
edad               int64
sexo               object
criterio_fallecido  object
ubigeo_cdc         float64
dpt_cdc            object
prov_cdc           object
dist_cdc           object
cdc_positividad     int64
flag_vacuna        int64
fecha_dosis1        object
fabricante_dosis1   object
fecha_dosis2        object
fabricante_dosis2   object
flag_hospitalizado  int64
eess_renaes        int64
eess_diresa        object
eess_red           object
eess_nombre        object
fecha_ingreso_hosp  object
flag_uci           int64
fecha_ingreso_uci   object
fecha_ingreso_ucin  object
con_oxigeno        int64
con_ventilacion     int64
fecha_segumiento_hosp_ultimo  object
evolucion_hosp_ultimo  object
ubigeo_inci_domicilio  float64
dep_domicilio      object
prov_domicilio     object
dist_domicilio     object
dtype: object
```

Figura 1: Resultado de ejecución del Código 1

4. Tablas

En esta sección se presentarán diferentes tablas creadas a partir de conjunto de datos anteriormente mencionado, obtenido a partir de DataFrame. También se detallará una interpretación de la tabla creada.

4.1. Positividad a según el departamento

```
leslym03@leslym03-Satellite-S55-B:~/Covid-Fallecidos$ python3 sexo_pos.py
cdc_positividad dpt_cdc
0              LIMA          43719
              PIURA         7591
              LA LIBERTAD     5179
              LAMBAYEQUE      5086
              AREQUIPA       4834
              ICA             4769
              CALLAO         4698
              JUNIN          3208
              ANCASH         2962
              LORETO         2805
              PUNO           2432
              CUSCO          2101
              UCAYALI        2047
              CAJAMARCA      1855
              SAN MARTIN     1560
              HUANUCO        1407
              TUMBES         920
              AYACUCHO       811
              MOQUEGUA       801
              TACNA          699
              APURIMAC       614
              HUANCANELICA   554
              PASCO          508
              AMAZONAS       483
              MADRE DE DIOS  452
1              LIMA          43880
              CALLAO         5071
              LA LIBERTAD     4868
              AREQUIPA       4798
              PIURA         4110
              JUNIN          3766
              ANCASH         3686
              ICA             3562
              LAMBAYEQUE     3122
              CUSCO          2681
              CAJAMARCA      2272
              PUNO           1721
              SAN MARTIN     1407
              AYACUCHO       1299
              HUANUCO        1268
              TACNA          1261
              LORETO         1065
              APURIMAC       895
              UCAYALI        860
              AMAZONAS       744
              MOQUEGUA       701
              TUMBES         632
              HUANCANELICA   599
              PASCO          526
              MADRE DE DIOS  315
Name: dpt_cdc, dtype: int64
```

Figura 2: Resultado de ejecución de la la columna positividad y departamento

Tabla 2: Positividad a según el departamento

Departamento	Negativo	Positivo
LIMA	43719	43880
PIURA	7591	4110
LA LIBERTAD	5179	4868
LAMBAYEQUE	5086	3122
AREQUIPA	4834	4798
ICA	4769	3562
CALLAO	4698	5071
JUNIN	3208	3766
ANCASH	2962	3686
LORETO	2805	1065
PUNO	2432	1721
CUSCO	2101	2681
UCAYALI	2047	860
CAJAMARCA	1855	2272
SAN MARTIN	1560	1407
HUANUCO	1407	1268
TUMBES	920	632
AYACUCHO	811	1299
MOQUEGUA	801	701
TACNA	699	1261
APURIMAC	614	895
HUANCAVELICA	554	599
PASCO	508	526
AMAZONAS	483	744
MADRE DE DIOS	452	315

La Tabla 3 presenta como mayores casos positivos y negativos a Lima debido a la cantidad de su población, al igual que Madre de Dios que presenta el menor número de casos positivos y negativos.

4.2. Hospitalizados por según el departamento

```

leslym03@leslym03-Satellite-S55-B:~/Covid-Fallecidos$ python3 sexo_pos.py
cdc_positividad dpt_cdc
0
  LIMA          43719
  PIURA        7591
  LA LIBERTAD   5179
  LAMBAYEQUE    5086
  AREQUIPA      4834
  ICA           4769
  CALLAO        4698
  JUNIN         3208
  ANCASH        2962
  LORETO        2805
  PUNO          2432
  CUSCO         2101
  UCAYALI       2047
  CAJAMARCA     1855
  SAN MARTIN    1560
  HUANUCO       1407
  TUMBES        920
  AYACUCHO      811
  MOQUEGUA      801
  TACNA         699
  APURIMAC      614
  HUANCVELICA   554
  PASCO         508
  AMAZONAS      483
  MADRE DE DIOS 452
1
  LIMA          43880
  CALLAO        5071
  LA LIBERTAD   4868
  AREQUIPA      4798
  PIURA        4110
  JUNIN         3766
  ANCASH        3686
  ICA           3562
  LAMBAYEQUE    3122
  CUSCO         2681
  CAJAMARCA     2272
  PUNO          1721
  SAN MARTIN    1407
  AYACUCHO      1299
  HUANUCO       1268
  TACNA         1261
  LORETO        1065
  APURIMAC      895
  UCAYALI       860
  AMAZONAS      744
  MOQUEGUA      701
  TUMBES        632
  HUANCVELICA   599
  PASCO         526
  MADRE DE DIOS 315
Name: dpt_cdc, dtype: int64

```

Figura 3: Resultado de ejecución de la la columna hospitalizado y departamento

Tabla 3: Hospitalizados por según el departamento

Departamento	Negativo	Positivo
LIMA	74708	12891
PIURA	9345	2356
LA LIBERTAD	9974	73
LAMBAYEQUE	7724	484
AREQUIPA	9582	50
ICA	5901	2430
CALLAO	8063	1706
JUNIN	4472	2502
ANCASH	3588	3060
LORETO	3778	92
PUNO	4054	99
CUSCO	2331	2451
UCAYALI	2162	745
CAJAMARCA	2172	1955
SAN MARTIN	2132	835
HUANUCO	2387	288
TUMBES	1391	161
AYACUCHO	2047	63
MOQUEGUA	966	343
TACNA	699	994
APURIMAC	789	720
HUANCAVELICA	802	351
PASCO	894	140
AMAZONAS	808	419
MADRE DE DIOS	382	385

La Tabla 3 presenta como mayores casos positivos y negativos a Lima debido a la cantidad de su población, al igual que Madre de Dios que presenta el menor número de casos positivos y negativos.

5. Gráficos

En esta sección se presentarán diferentes gráficos creados a partir de conjunto de datos anteriormente mencionado, obtenido a partir de DataFrame. También se detallará una interpretación del gráfico creado.

5.1. Positividad a según el sexo

En la Figura 4 muestra un mayor número de mujeres positivas y negativas, además de ser casi el doble al número de hombres. Además, la diferencia entre personas positivas y negativas a covid es pequeña.

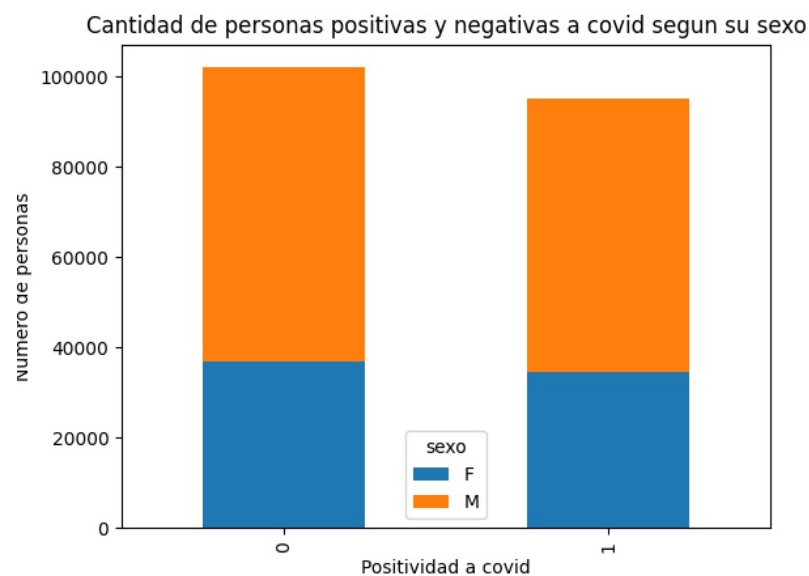


Figura 4: Gráfico sobre la cantidad de personas positivas y negativas a según su sexo

5.2. Hospitalizados por según el sexo

En la Figura 5 muestra un mayor número de personas hospitalizadas sin, además las personas hospitalizadas por resultan aproximadamente una cuarta parte de los que no presentan.

Con respecto al sexo concluimos que el mayor número de personas hospitalizadas son mujeres.

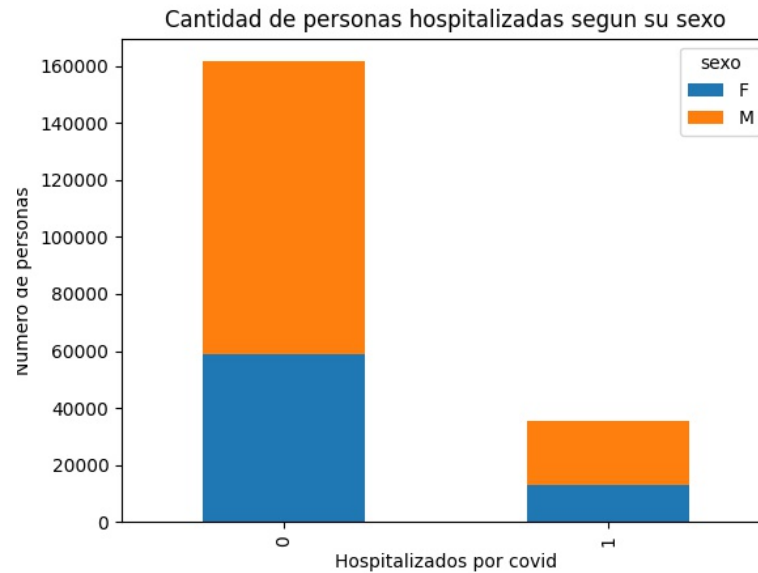


Figura 5: Gráfico sobre la cantidad de personas hospitalizadas por según su sexo

5.3. Criterio de fallecidos

En la Figura 7 muestra la que mayor cantidad de personas fallecidas por criterio virológico y la menor cantidad de personas fallecidas por criterio de Nexo Epidemiológico.

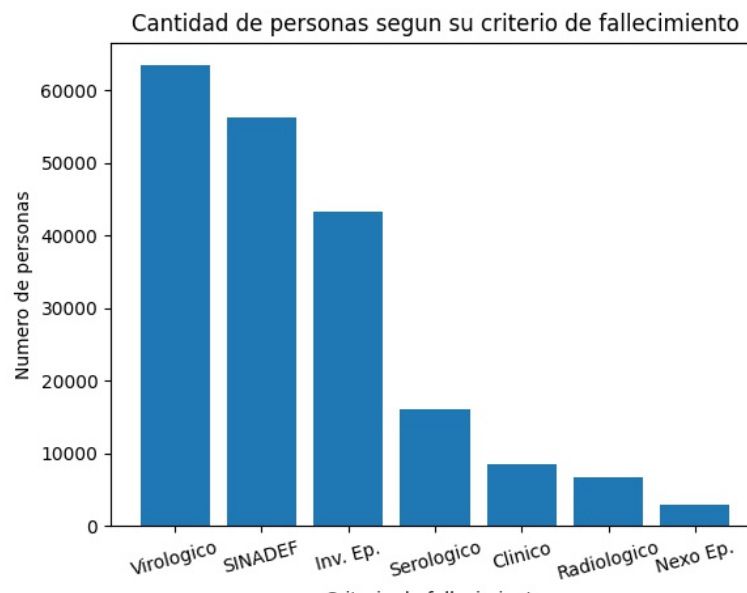


Figura 6: Gráfico de barras sobre el criterio de las personas fallecidas

También se realizó un gráfico circular para la mejor visualización de los datos sobre las personas fallecidas según el criterio de fallecimiento.

Cantidad de personas segun su criterio de fallecimiento

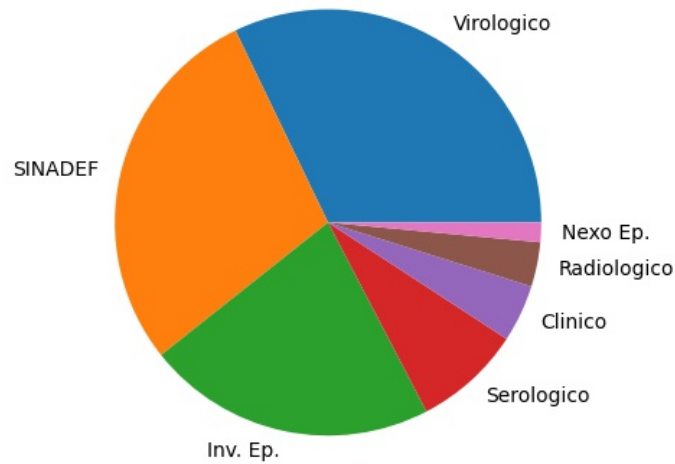


Figura 7: Gráfico circular sobre el criterio de las personas fallecidas

6. Repositorio

El código utilizado en este informe se almacenó en el repositorio GitHub, el cual se puede acceder mediante el siguiente enlace:

<https://github.com/Leslym03/-Fallecidos>

Referencias

- [1] *Información de Fallecidos del Sistema Informático Nacional de Defunciones - SINADEF - [Ministerio de Salud] / Plataforma Nacional de Datos Abiertos.* (2012). Plataforma Nacional de Datos Abiertos. <https://www.datosabiertos.gob.pe/dataset/informaci%C3%B3n-de-fallecidos-del-sistema-inform%C3%A1tico-nacional-de-defunciones-sinadef-ministerio>
- [2] *Plataforma Nacional de Datos Abiertos.* (s. f.). Plataforma Nacional de Datos Abiertos. <https://www.datosabiertos.gob.pe/>