Catedra 1

Introducción

Una de las tareas habituales a las que se enfrenta todo analista o científico de datos es la de entender las características de las variables con las que se está trabajando.

Para este trabajo se pretende hacer un análisis exploratorio de datos, que permita identificar posibles errores, datos incorrectamente introducidos ó la ausencia de estos , revelar la presencia de valores atípicos (outliers).

Además de poder visualizar la relación existente entre las variables, poder ver los aspectos más significativos de las variables , mediante resúmenes o representaciones gráficas

Definición del problema/objetivo de investigación

Se requiere analizar el comportamiento de los datos que tienen relación a las atenciones de urgencia en los establecimientos de atención de salud pública durante el año 2023, en el país de Chile.

Dado que diariamente acuden a los establecimientos de salud personas de diversas edades y consultas por diversas causas, el analizar posibles tendencias en relación a la a edad o al establecimiento donde se producen más o menos consultas, permitirán deducir conclusiones que pueden transformarse en la base para la implementación de nuevas acciones ó nuevos analisis

1. Introducción

Para el análisis requerido se han obtenido los datos que contiene la información de las atenciones de urgencias realizadas en el año 2023 de los establecimientos de salud pública.

Este conjunto de datos se encuentra en los datos abiertos en la página oficial del Departamento de Estadísticas e Información (DEIS), dependiente de la Subsecretaría de Salud Pública en el Ministerio de Salud de chile.

En el conjunto de datos de encuentran los datos que corresponde a la identificación de los establecimientos que realizaron la atención, la categoría del establecimiento, la comuna a la que pertenece, la fecha de la realización, la causa de atencion

|  |  |
| --- | --- |
| Nombre | Descripción |
| IdEstablecimiento | Codigo del Establecimiento |
| NEstablecimiento | Nombre del establecimiento |
| IdCausa | Código de categoría de la causa de atencion |
| GlosaCausa | Descripcion de la causa de atencion |
| Total | Numero total de atenciones |
| Menores\_1 | Numero total de atenciones de pacientes menores de 1 año |
| De\_1\_a\_4 | Numero total de atenciones de pacientes menores que están entre 1 y 4 años |
| De\_5\_a\_14 | Numero total de atenciones de pacientes menores entre 5 a 14 años |
| De\_15\_a\_64 | Numero total de atenciones de pacientes entre 15 a 64 años |
| De\_65\_y\_mas | Numero total de atenciones de pacientes mayores de 65 años |
| fecha | Fecha de la atención |
| semana |  |
| GLOSATIPOESTABLECIMIENTO | Cear:Centro Exclusivo de Atencion Respiratoria  Sapu:Servicio de atención primaria de urgencia  SAR: Servicio de atención primaria de urgencia de alta resolutividad  SUR:Servicio de urgencia rural |
| GLOSATIPOATENCION | Categoría de atención adulto, indiferenciado , infantil |
| GlosaTipoCampana |  |
| CodigoRegion | Corresponde al código de la región a la que pertence |
| NombreRegion | Correponde a la región a la que pertenece |
| CodigoDependencia | Corresponde al de la dependencia administrativa a la que pertenece el establecimiento, es decir el servicio de salud |
| NombreDependencia | Corresponde a la descripción de la dependencia administrativa a la que pertenece el establecimiento, es decir el servicio de salud |
| CodigoComuna | Codigo de la comuna a la que pertenece el establecimiento |
| NombreComuna | Descripcion de la comuna a la que pertenece el establecimiento |

Los objetivos del análisis es poder ver el comportamiento de los datos y poder dar respuesta las siguientes situaciones

* Poder identificar en que establecmientos de la región de los lagos se producen mas atenciones durante el año
* Identificar en que periódos del año se producen consultas por atenciones de urgencia en los establecimientos de la región de los lagos
* Identificar cual es el grupo de personas que consultan las urgencias, niños, adultos o mayores de edad

1. Preprocesamiento de Datos

Descripción detallada de los pasos tomados para limpiar y preparar los datos para el análisis utilizando Tidyverse.

Para el manejo de las observaciones y sus variables se transformó todo el conjunto de datos en un tipo de datos dataframe.

Se evaluó si el conjunto de datos contenía en su interior el tipo de datos perdidos na o la ausencia de estos comprobando la existencia de datos null

Dado que en los registro numéricos el ser humano tiende a confundir el número cero con la letra “o”, algunos registros pueden presentar este carácter en una variable numérica se decidió verificar la presencia del carácter “o” y remplazar el valor en la situación de encontrarlo.

Al igual que los datos perdidos a la existencia de estos , se decidió remplazar estos datos por el valor cero en las variables cuantitativas encontradas en el dataset.

Se trabajaraon las variables cualitativas con caracteres especiales

Dado los objetivos del análisis se necesita trabajar con una las observaciones requeridas que corresponden a la región, se obtiene un resumen de las observaciones de la región.

3. Análisis Exploratorio de Datos: Implementación de códigos en R para explorar los datos. Esto puede incluir visualizaciones de datos, medidas de tendencia central, etc.

4. Modelado de Datos: Aplicación de técnicas de modelado de datos y algoritmos de aprendizaje automático, si corresponde.

5. Interpretación de Resultados: Discusión de los resultados del análisis y su relevancia para los objetivos establecidos en la introducción.

6. Código: Incluir todo el código R utilizado en el análisis como un apéndice o integrado en el informe. El código debe ser legible y propiamente comentario.