Imagen que contiene Forma

Descripción generada automáticamente

**PROYECTO AGORA**

**ENTENDIMIENTO DE LOS DATOS - VACUNASCOVID**

**2024**

**Control de cambios**

|  |  |  |  |
| --- | --- | --- | --- |
| **VERSIÓN** | **FECHA** | **MODIFICACIÓN** | **RESPONSABLE** |
| 1.3 | 24 de abril de 2024 | Modificación del documento | Luis Carlos Chicaíza |
| 1.4 | 25 de abril de 2024 | Ajustes complementarios finales y ampliación de interpretación de gráficas | Pedro Pérez |
| 1.5 | 22 de junio de 2024 | Se completa la corrección del documento, con base a los comentarios recibidos. | Luis Carlos Chicaíza |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

Contenido

[1 Introducción 4](#_Toc169978340)

[2 Metodología 5](#_Toc169978341)

[2.1 Exploración de los datos 5](#_Toc169978342)

[3 Fuente principal de los datos 6](#_Toc169978343)

[3.1 Origen y procedencia de los datos 6](#_Toc169978344)

[3.2 Descripción general de VACUNASCOVID y ajustes iniciales 7](#_Toc169978345)

[3.2.1 Ajustes requeridos en nombres de las fuentes 7](#_Toc169978346)

[3.2.2 Ajustes requeridos en registros duplicados 8](#_Toc169978347)

[3.3 Estadísticas descriptivas de las variables 8](#_Toc169978348)

[3.3.1 Esquema para la exploración y análisis descriptivo de los datos 8](#_Toc169978349)

[3.3.2 Listado detallado de fuentes 9](#_Toc169978350)

[3.3.3 Análisis preliminares 54](#_Toc169978351)

# Introducción

El presente documento ofrece una visión detallada de los principales hallazgos derivados del proceso de exploración inicial de los datos del sistema VACUNASCOVID. En esta fase inicial de análisis, se identificó la fuente de datos y se examinaron las variables contenidas en ella y sus características fundamentales.

La información denominada VACUNASCOVID, comprende información recopilada durante el periodo comprendido entre 2021 a 2023. Este periodo permite analizar una variedad de contextos y condiciones que resultan ser importantes para la comprensión y el análisis de las tendencias relacionadas.

Durante la fase inicial de exploración, se realizó un análisis exhaustivo de las variables contenidas en el conjunto de datos de la información suministrada de VACUNASCOVID. Se aplicaron técnicas estadísticas descriptivas básicas para examinar la distribución, la variabilidad y las tendencias presentes en los registros disponibles. Este análisis preliminar sirvió como punto de partida para comprender mejor la naturaleza de los datos y así poder identificar posibles áreas de interés y enfoque para análisis posteriores.

En estas secciones, se presenta la metodología de exploración de los datos, como los principales resultados obtenidos en esta fase inicial, incluyendo las características de las variables analizadas, las tendencias y patrones identificados de VACUNASCOVID.

# Metodología

Con el propósito de realizar un análisis exhaustivo de los datos provenientes de la información VACUNAS COVID-19 proporcionada, se busca aplicar diversas técnicas de análisis exploratorio de los datos. Algunas técnicas son las siguientes:

## Exploración de los datos

**Análisis de estadísticas descriptivas**

Calcular la media, mediana, desviación estándar y los percentiles, para así comprender la distribución y la variabilidad de las variables numéricas del conjunto de datos.

**Análisis de correlación**

Explorar la relación entre pares de variables numéricas mediante el cálculo de coeficientes de correlación, como el coeficiente de correlación de Pearson. Con el objetivo de identificar posibles asociaciones lineales o no lineales entre las variables.

**Análisis de frecuencias**

Examinar la distribución de valores únicos en variables categóricas para comprender la proporción de cada categoría y detectar posibles desequilibrios o sesgos en los datos.

**Visualización de datos**

Generar gráficos y visualizaciones, como histogramas, diagramas de dispersión, diagramas de bigotes o violín y gráficos de barras, para representar gráficamente la distribución de los datos y las relaciones entre variables.

# Fuente principal de los datos

## Origen y procedencia de los datos

A continuación, se lista los datos asociados a la base de datos recibida para el proyecto AGORA:

|  |  |
| --- | --- |
| Nombre de la fuente en el proyecto: | VACUNASCOVID |
| Número de archivos o tablas: | 10 carpetas con la información por año |
| Nombre del archivo original de la fuente entregado: | VACUNASCOVID.zip |
| Ruta del archivo original y archivos de trabajo: | //caoba-access.virtual. dfs.core.windows.net/ vacunascovid/vacunascovid\_parquet\_{año} |

Tabla 1. Descripción inicial de la fuente VACUNASCOVID

Para el procesamiento de los datos, se definió unos contendores en la infraestructura con la información por año. Esta información está consolidada en su mayoría desde el año 2020 hasta el año 2023. Existen otros registros por fuera de este rango de fechas 2024-2030 (36 registros), por lo cual no se toman a consideración. Se omiten 5 registros del año 2019 que están en su mayoría en valores NULL.

Tabla

Descripción generada automáticamenteTabla

Descripción generada automáticamente  
Imagen 1. Carpetas por años de la fuente VACUNASCOVID

El procesamiento se realizará sobre estas carpetas para poder analizar de una fecha específica o consolidar todas las fechas para un mayor análisis.

## Descripción general de VACUNASCOVID y ajustes iniciales

A continuación, se detalla los registros por año de los registros en la fuente:

|  |  |  |  |
| --- | --- | --- | --- |
| Base | #  de Registros | # de Columnas | ID Únicos |
| VACUNASCOVID\_2020 | 52 | 64 | 52 |
| VACUNASCOVID\_2021 | 13´239.481 | 64 | 8993240 |
| VACUNASCOVID\_2022 | 13 | 64 | 13 |
| VACUNASCOVID\_2023 | 27 | 64 | 27 |

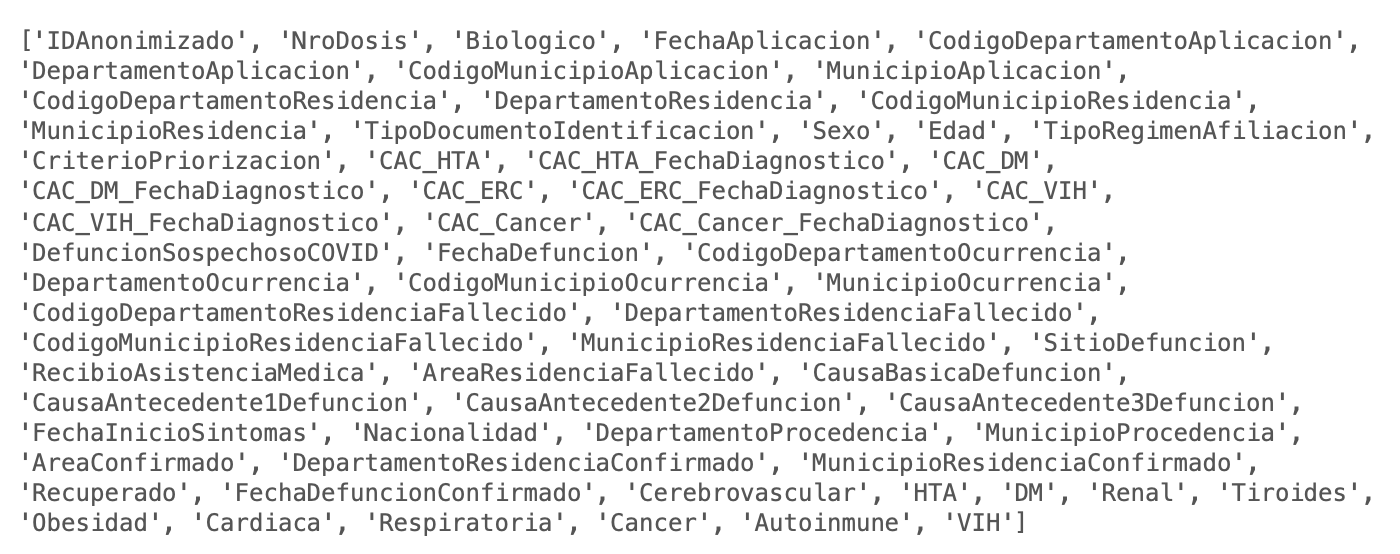
Tabla 1. Descripción inicial de la fuente VACUNASCOVID

El total de registros de toda la fuente es de 13´239.609 registros y 64 variables. Dejando solamente los años 2020 a 2023 iniciamos con 13´239.573 registros.

### Ajustes requeridos en nombres de las fuentes

La información para tratar contempla el periodo comprendido desde el 2020 hasta 2023. No fue necesario aplicar ajuste en los nombres de las variables, no se evidencio un numero o nombre distinto de las columnas a través del tiempo.

Se presentan los nombre de las 64 variables en la fuente VACUNASCOVID.

**

### Ajustes requeridos en registros duplicados

Al revisar registros duplicados por año no evidencia registros duplicados.

## Estadísticas descriptivas de las variables

En esta fase se realizará el análisis exploratorio de los datos de la fuente VACUNASCOVID. Esta fase busca obtener una compresión más profunda de la naturaleza y las características de las variables que componen este conjunto de datos.

A través de algunas técnicas de estadística y visualizaciones, se explorará la distribución, la variabilidad y las posibles relaciones entre las diferentes variables. También se buscará identificar algunos patrones, tendencias y anomalías que puedan proporcionar información importante para futuros análisis y la toma de decisiones.

Este análisis permitirá generar una contextualización adecuada de los datos, y sentar las bases para análisis mucho más robustos y modelado predictivo en etapas posteriores a este proyecto.

### Esquema para la exploración y análisis descriptivo de los datos

A continuación, se define el siguiente esquema para realizar la exploración y el análisis descriptivo de los datos. Para cada variable de VACUNASCOVID se debe diligenciar la siguiente tabla.

|  |  |
| --- | --- |
| **FUENTE** | **Nombre de la fuente / Nombre del archivo o tabla** |
| **Campo 1** | Nombre del campo tal como aparece en el archivo fuente |
| Descripción | Tipo de dato (fecha, texto categórico, texto no categórico, numérico nominal o numérico ordinal.) |
| Periodo | Periodo de la base de datos |
| Extensión | Formato del archivo |
| Volumen (#filas) | Cantidad de registros |
| Campos (#columnas) | Numero de atributos |

|  |  |
| --- | --- |
| **Campos VACUNASCOVID** | Descripción de todos los atributos de la base de datos |
| **Campo n** | nombre del campo en la fuente de datos |
| Alias | Nombre de la variable. |
| Descripción | Descripción o propósito del atributo. |
| Tipo | Tipo de dato (fecha, texto categórico, texto no categórico, numérico nominal o numérico ordinal.) |
| Dominio | Rango o valor esperado. |
| Estadísticas básicas | Se calculan las estadísticas básicas o la frecuencia de las variables. |
| Estadísticas básicas | Se calculan las estadísticas básicas o la frecuencia de las variables. |
| Observaciones | Información relevante de la variable a tener en cuenta. |

Tabla 2: Descripción estadística de las variables​

### Listado detallado de fuentes

Teniendo en cuenta el esquema mencionado anteriormente, se genera un cuadro detallado describiendo la fuente de datos y la descripción de cada una de las variables identificadas.

|  |  |
| --- | --- |
| **FUENTE** | **INFORMACIÓN VACUNASCOVID** |
| Descripción | Contiene los datos del sistema VACUNASCOVID |
| Período | 2020 - 2023 |
| Extensión | .parquet |
| Volumen (#filas) | 13.239.573 |
| Campos (#columnas) | 64 |
| Fecha recepción | 7 de diciembre de 2023 |
| **Campo 1** | IDAnonimizado |
| Alias | ID de la persona |
| Descripción | Corresponde al identificador del paciente anonimizado. |
| Tipo | Categórica nominal |
| Dominio | Muestra de 'IDAnonimizado': fffffe3d9444c3be7ea2a586b6248fc5  Existen 8´993.307 identificadores únicos |
| Estadísticas Básicas |  |
| Observaciones | * La variable no tienes valores NULL |
| **Campo 2** | NroDosis |
| Alias | Número de dosis |
| Descripción | Número de dosis aplicadas al paciente |
| Tipo | Numérico ordinal |
| Dominio | [1,2]  Existen 2 códigos únicos |
| Estadísticas Básicas |  |
| Observaciones | * La variable no tienes valores NULL |
| **Campo 3** | Biologico |
| Alias | Marca del bilógico aplicado |
| Descripción | Marca del bilógico aplicado al paciente |
| Tipo | Categórico Alfanumérico |
| Dominio | [PFIZER, SINOVAC, OXFORD-ASTRAZENECA]  Existen 3 códigos únicos |
| Estadísticas Básicas |  |
| Observaciones | Se tienen 1´466.083 registros NULL |
| **Campo 4** | FechaAplicacion |
| Alias | Fecha de aplicación del biológico |
| Descripción | Fecha de aplicación del biológico al paciente |
| Tipo | Fecha |
| Dominio | [AAAA-MM-DD]  Existen 225 registros únicos |
| Estadísticas Básicas |  |
| **Campo 5** | CodigoDepartamentoAplicacion |
| Alias | Código asociado al Departamento |
| Descripción | Código DANE asociado al Departamento de aplicación del biológico al paciente |
| Tipo | Categórico Numérico |
| Dominio | [Enteros positivos – Códigos DANE de los departamentos de Colombia]  Existen 34 códigos únicos |
| Estadísticas Básicas |  |
| Observaciones | Se tienen 303 registros con valor (00) en esta variable |
| **Campo 6** | DepartamentoAplicacion |
| Alias | Departamento de Aplicación |
| Descripción | Nombre del Departamento de la Aplicación del biológico al paciente |
| Tipo | Categórico Alfanumérico |
| Dominio | [Nombre de los Departamentos de Colombia]  Existen 40 código únicos |
| Estadísticas Básicas |  |
|  | * Existen errores de digitación que deben reclasificarse en 6 departamentos. * El código de la variable anterior 00 se evidencia que corresponde a NACION. |
| **Campo 7** | CodigoMunicipioAplicacion |
| Alias | Código asociado al Municipio |
| Descripción | Código DANE asociado al Municipio de aplicación del biológico al paciente |
| Tipo | Categórico Numérico |
| Dominio | [Enteros positivos – Códigos DANE de los municipios de Colombia]  Existen 1.102 códigos únicos |
| Estadísticas Básicas |  |
| **Campo 8** | MunicipioAplicacion |
| Alias | Municipio de Aplicación |
| Descripción | Nombre del Municipio de la Aplicación del biológico al paciente |
| Tipo | Categórico Alfanumérico |
| Dominio | [Nombre de los Municipios de Colombia]  Existen 1085 código únicos |
| Estadísticas Básicas |  |
| **Campo 9** | CodigoDepartamentoResidencia |
| Alias | Código asociado al Departamento de Residencia |
| Descripción | Código DANE asociado al Departamento de Residencia del paciente |
| Tipo | Categórico Numérico |
| Dominio | [Enteros positivos – Códigos DANE de los departamentos de Colombia]  Existen 37 códigos únicos |
| Estadísticas Básicas |  |
| **Campo 10** | DepartamentoResidencia |
| Alias | Departamento de Residencia |
| Descripción | Nombre del Departamento de Residencia del paciente |
| Tipo | Categórico Alfanumérico |
| Dominio | [Nombre de los Departamentos de Colombia]  Existen 44 código únicos |
| Estadísticas Básicas |  |
| **Campo 11** | CodigoMunicipioResidencia |
| Alias | Código asociado al Municipio de Residencia |
| Descripción | Código DANE asociado al Municipio de Residencia del paciente |
| Tipo | Categórico Numérico |
| Dominio | [Enteros positivos – Códigos DANE de los municipios de Colombia]  Existen 1.122 códigos únicos |
| Estadísticas Básicas |  |
| **Campo 12** | MunicipioResidencia |
| Alias | Municipio de Residencia |
| Descripción | Nombre del Municipio de Residencia del paciente |
| Tipo | Categórico Alfanumérico |
| Dominio | [Nombre de los Municipios de Colombia]  Existen 1.111 código únicos |
| Estadísticas Básicas |  |
|  |  |
| **Campo 13** | TipoDocumentoIdentificacion |
| Alias | Tipo de Documento de identificación |
| Descripción | Tipo de Documento del paciente |
| Tipo | Categórico Nominal Alfanumérico |
| Dominio | Existen 11 código únicos |
| Estadísticas Básicas |  |
| **Campo 14** | Sexo |
| Alias | Sexo del paciente |
| Descripción | Sexo del usuario al que se le aplicó el biológico |
| Tipo | Categórico Alfanumérico |
| Dominio | [F,M,N,I]  Existen 4 códigos únicos en la variable |
| Estadísticas Básicas |  |
| Observaciones | Se tienen 7.177 registros con NULL en esta variable |
| **Campo 15** | Edad |
| Alias | Edad del paciente. |
| Descripción | Edad del paciente |
| Tipo | Numérico |
| Dominio | [0, 198]  Existen 131 registros únicos en la variable |
| Estadísticas Básicas | | max | 198 | |
| Observaciones | * Existen 7.177 registros con NULL en la variable |
| **Campo 16** | TipoRegimenAfiliacion |
| Alias | Tipo de régimen de afiliación |
| Descripción | Tipo de régimen al que pertenece el paciente |
| Tipo | Categórico Nominal |
| Dominio | [C, S, ND]  Existen 4 registros únicos e las variables |
| Estadísticas Básicas |  |
| Observaciones | Existen 6.969 registro NULL en la variable |
| **Campo 17** | CriterioPriorizacion |
| Alias | Criterio de priorización |
| Descripción | Criterio de priorización de la Atención |
| Tipo | Categórico único valor |
| Dominio | [NO REPORTADO]  Existen 2 registros únicos |
| Estadísticas Básicas |  |
| Observaciones | Existe 6.969 datos en NULL |
| Observaciones | * Existen 6.969 registros NULL en la variable |
| **Campo 18** | CAC\_HTA |
| Alias | Condición preexistente Hipertensión Arterial |
| Descripción | Condición preexistente Hipertensión Arterial |
| Tipo | Categórico Nominal |
| Dominio | [SI, NO]  Existen 2 registros únicos |
| Estadísticas Básicas |  |
| **Campo 19** | CAC\_HTA\_FechaDiagnostico |
| Alias | CAC\_HTA\_FechaDiagnostico |
| Descripción | Fecha de diagnóstico de hipertensión arterial del paciente. |
| Tipo | Fecha |
| Dominio | [AAAA-MM-DD] |
| Estadísticas Básicas |  |
| Observaciones | Existe 9´404.660 datos en NULL |
| **Campo 20** | CAC\_DM |
| Alias | Condición preexistente Diabetes mellitus |
| Descripción | Condición preexistente Diabetes mellitus |
| Tipo | Categórico Nominal |
| Dominio | [SI, NO]  Existen 2 registros únicos |
| Estadísticas Básicas |  |
| **Campo 21** | CAC\_DM\_FechaDiagnostico |
| Alias | CAC\_DM\_FechaDiagnostico |
| Descripción | Fecha de diagnóstico de diabetes mellitus del paciente. |
| Tipo | Fecha |
| Dominio | [AAAA-MM-DD] |
| Estadísticas Básicas |  |
| Observaciones | Existe 12´010.823 datos en NULL |
| **Campo 22** | CAC\_ERC |
| Alias | Condición preexistente Enfermedad renal crónica |
| Descripción | Condición preexistente Enfermedad renal crónica |
| Tipo | Categórico Nominal |
| Dominio | [SI, NO]  Existen 2 registros únicos |
| Estadísticas Básicas |  |
| **Campo 23** | CAC\_ERC\_FechaDiagnostico |
| Alias | CAC\_ERC\_FechaDiagnostico |
| Descripción | Fecha de diagnóstico de enfermedad renal crónica del paciente. |
| Tipo | Fecha |
| Dominio | [AAAA-MM-DD] |
| Estadísticas Básicas |  |
| Observaciones | Existe 12´398.283 datos en NULL |
| **Campo 24** | CAC\_VIH |
| Alias | Condición preexistente VIH/SIDA |
| Descripción | Condición preexistente VIH/SIDA |
| Tipo | Categórico Nominal |
| Dominio | [SI, NO]  Existen 2 registros únicos |
| Estadísticas Básicas |  |
| **Campo 25** | CAC\_VIH\_FechaDiagnostico |
| Alias | CAC\_VIH\_FechaDiagnostico |
| Descripción | Fecha en la que se diagnosticó la infección por VIH a un paciente |
| Tipo | Fecha |
| Dominio | [DD/MM/AAAA] |
| Estadísticas Básicas |  |
| Observaciones | Existe 13´192.914 datos en NULL |
| **Campo 26** | CAC\_Cancer |
| Alias | Condición preexistente Cáncer |
| Descripción | Condición preexistente Cáncer |
| Tipo | Categórico Nominal |
| Dominio | [SI, NO]  Existen 2 registros únicos |
| Estadísticas Básicas |  |
| **Campo 27** | CAC\_Cancer\_FechaDiagnostico |
| Alias | CAC\_Cancer\_FechaDiagnostico |
| Descripción | Fecha en la que se diagnosticó el cáncer a un paciente |
| Tipo | Fecha |
| Dominio | [AAAA-MM-DD] |
| Estadísticas Básicas |  |
| Observaciones | Existe 12´991.953 datos en NULL |
| **Campo 28** | DefuncionSospechosoCOVID |
| Alias | Defunción Sospechoso COVID |
| Descripción | Indica que la muerte de un paciente es sospechosa de haber sido causada por una infección de COVID-19 |
| Tipo | Categórico nominal. |
| Dominio | [SI, NO] |
| Estadísticas Básicas |  |
| **Campo 29** | FechaDefuncion |
| Alias | Fecha de Defunción |
| Descripción | Fecha de defunción de un paciente. |
| Tipo | Fecha |
| Dominio | [AAAA-MM-DD] |
| Estadísticas Básicas |  |
| Observaciones | Existe 13´222.043 datos en NULL |
| **Campo 30** | CodigoDepartamentoOcurrencia |
| Alias | Código del Departamento de Ocurrencia |
| Descripción | Código que identifica el departamento geográfico en donde ocurrió el suceso |
| Tipo | Categórico Nominal |
| Dominio | [CÓDIGO DEPARTAMENTO] |
| Estadísticas Básicas |  |
| Observaciones | Existe 13´222.038 datos en NULL |
| **Campo 31** | DepartamentoOcurrencia |
| Alias | Nombre del Departamento de Ocurrencia |
| Descripción | Identifica el departamento geográfico en donde ocurrió el suceso |
| Tipo | Categórico Nominal |
| Dominio | [NOMBRE DEPARTAMENTO] |
| Estadísticas Básicas |  |
| Observaciones | Existe 13´222.038 datos en NULL |
| **Campo 32** | CodigoMunicipioOcurrencia |
| Alias | Código del Municipio de Ocurrencia |
| Descripción | Código que identifica el municipio en donde ocurrió el suceso. |
| Tipo | Categórico Nominal |
| Dominio | [CÓDIGO MUNICIPIO] |
| Estadísticas Básicas |  |
| Observaciones | Existe 13´222.038 datos en NULL |
| **Campo 33** | MunicipioOcurrencia |
| Alias | Nombre del Municipio de Ocurrencia. |
| Descripción | Identifica el municipio en donde ocurrió el suceso. |
| Tipo | Categórico Nominal |
| Dominio | [NOMBRE MUNICIPIO] |
| Estadísticas Básicas |  |
| Observaciones | Existe 13´222.038 datos en NULL |
| **Campo 34** | CodigoDepartamentoResidenciaFallecido |
| Alias | Código de Departamento de Residencia del Fallecido |
| Descripción | Código que identifica el departamento donde residía la persona fallecida. |
| Tipo | Categórico Nominal |
| Dominio | [CÓDIGO DEPARTAMENTO] |
| Estadísticas Básicas |  |
|  |  |
| Observaciones | Existe 13´222.038 datos en NULL |
| **Campo 35** | DepartamentoResidenciaFallecido |
| Alias | Nombre del Departamento de Residencia del Fallecido |
| Descripción | Departamento donde residía la persona fallecida. |
| Tipo | Categórico Nominal |
| Dominio | [NOMBRE DEPARTAMENTO] |
| Estadísticas Básicas |  |
| Observaciones | Existe 13´222.038 datos en NULL |
| **Campo 36** | CodigoMunicipioResidenciaFallecido |
| Alias | Código del Municipio de Residencia del Fallecido |
| Descripción | Código que identifica el municipio de donde residía la persona fallecida. |
| Tipo | Categórico Nominal |
| Dominio | [CÓDIGO MUNICIPIO] |
| Estadísticas Básicas |  |
| Observaciones | Existe 13´222.038 datos en NULL |
| **Campo 37** | MunicipioResidenciaFallecido |
| Alias | Nombre del Municipio de Residencia del Fallecido |
| Descripción | Municipio de donde residía la persona fallecida. |
| Tipo | Categórico Nominal |
| Dominio | [NOMBRE MUNICIPIO] |
| Estadísticas Básicas |  |
| Observaciones | Existe 13´222.038 datos en NULL |
| **Campo 38** | SitioDefuncion |
| Alias | Sitio de Defunción |
| Descripción | Lugar de defunción |
| Tipo | Categórico Nominal |
| Dominio | [1-HOSPITAL/CLÍNICA, 2-CENTRO/PUESTO, 3-CASA/DOMICILIO, 4- LUGAR DE TRABAJO, 5-VÍA PUBLICA] |
| Estadísticas Básicas |  |
| Observaciones | Existe 13´222.043 datos en NULL |
| **Campo 39** | RecibioAsistenciaMedica |
| Alias | RecibioAsistenciaMedica |
| Descripción | Define si una persona recibió una asistencia médica en un momento dado. |
| Tipo | Categórico Nominal |
| Dominio | [SI, NO] |
| Estadísticas Básicas |  |
| Observaciones | Existe 13´222.038 datos en NULL |
| **Campo 40** | AreaResidenciaFallecido |
| Alias | Areade Residencia del Fallecido |
| Descripción | Área geográfica o entorno en donde residía la persona fallecida. |
| Tipo | Categórico Nominal. |
| Dominio | [1-CABECERA, 2-AREA RURAL, 3-CENTRO POBLADO, SIN INFORMACION] |
| Estadísticas Básicas |  |
| Observaciones | Existe 13´222.038 datos en NULL |
| **Campo 41** | CausaBasicaDefuncion |
| Alias | Causa Básica de Defuncion |
| Descripción | Información sobre la enfermedad, lesión o condición médica que condujo directamente al fallecimiento de la persona. |
| Tipo | Categórico Nominal |
| Dominio | [CAUSAS MEDICAS] |
| Estadísticas Básicas |  |
| Observaciones | Existe 13´222.045 datos en NULL |
| **Campo 42** | CausaAntecedente1Defuncion |
| Alias | Primera causa de Antecedente de la Defunción |
| Descripción | Información sobre la primera causa medica de la enfermedad, lesión o condición médica que condujo directamente al fallecimiento de la persona. |
| Tipo | Categórico Nominal |
| Dominio | [CAUSAS MEDICAS] |
| Estadísticas Básicas |  |
| Observaciones | Existe 13´224.412 datos en NULL |
| **Campo 43** | CausaAntecedente2Defuncion |
| Alias | Segunda causa de Antecedente de Defunción |
| Descripción | Información sobre la segunda causa médica o condición relevante que contribuyó al fallecimiento de la persona. |
| Tipo | Categórico Nominal |
| Dominio | [CAUSAS MEDICAS] |
| Estadísticas Básicas |  |
| Observaciones | Existe 13´225.013 datos en NULL |
| **Campo 44** | CausaAntecedente3Defuncion |
| Alias | Tercera Causa de Antecedente de Defunción |
| Descripción | Información sobre la tercera causa médica o condición relevante que contribuyó al fallecimiento de la persona. |
| Tipo | Categórico Nominal. |
| Dominio | [CAUSAS MEDICAS] |
| Estadísticas Básicas |  |
| Observaciones | Existe 13´234.112 datos en NULL |
| **Campo 45** | FechaInicioSintomas |
| Alias | Fecha de inicio de Síntomas |
| Descripción | Fecha de inicio de Síntomas de COVID |
| Tipo | Fecha |
| Dominio | [AAAA-MM-DD]  Existen 495 registros únicos |
| Estadísticas Básicas |  |
| Observaciones | * Existe 12´033.609 NULL en la variable |
| **Campo 46** | Nacionalidad |
| Alias | Nacionalidad |
| Descripción | Nacionalidad del paciente |
| Tipo | Categórica Nominal |
| Dominio | [PAISES DEL MUNDO]  Existen 40 registros únicos |
| Estadísticas Básicas |  |
| Observaciones | * Existe 11´917.577 NULL en la variable |
| **Campo 47** | DepartamentoProcedencia |
| Alias | Departamento de Procedencia |
| Descripción | Nombre del Departamento de la Procedencia |
| Tipo | Categórico Alfanumérico |
| Dominio | [Nombre de los Departamentos de Colombia]  Existen 34 código únicos |
| Estadísticas Básicas |  |
| Observaciones | * Existe 11´917.577 NULL en la variable |
| **Campo 48** | MunicipioProcedencia |
| Alias | Municipio de Procedencia confirmado |
| Descripción | Nombre del Municipio de Procedencia confirmado del paciente |
| Tipo | Categórico Alfanumérico |
| Dominio | [Nombre de los Municipios de Colombia]  Existen 1.110 código únicos |
| Estadísticas Básicas |  |
| Observaciones | * Existe 11´917.577 NULL en la variable |
| **Campo 49** | AreaConfirmado |
| Alias | Área o zona del territorio Confirmado |
| Descripción | Área o zona del territorio Confirmado |
| Tipo | Categórica Nominal |
| Dominio | [CABECERA, CENTRO POBLADO, SIN INFORMACION, AREA RURAL DISPERSA]  Existen 5 código únicos |
| Estadísticas Básicas |  |
| Observaciones | * Existe 11´917.577 NULL en la variable |
| **Campo 50** | DepartamentoResidenciaConfirmado |
| Alias | Departamento de Residencia |
| Descripción | Nombre del Departamento de Residencia del paciente |
| Tipo | Categórico Alfanumérico |
| Dominio | [Nombre de los Departamentos de Colombia]  Existen 36 código únicos |
| Estadísticas Básicas |  |
| Observaciones | * Existen 11´917.577 NULL en la variable |
| **Campo 51** | MunicipioResidenciaConfirmado |
| Alias | Municipio de Residencia confirmado |
| Descripción | Nombre del Municipio de Residencia confirmado del paciente |
| Tipo | Categórico Alfanumérico |
| Dominio | [Nombre de los Municipios de Colombia]  Existen 1.114 código únicos |
| Estadísticas Básicas |  |
| Observaciones | * Existen 11´917.577 NULL en la variable |
| **Campo 52** | Recuperado |
| Alias | Recuperado |
| Descripción | Recuperado |
| Tipo | Categórico Nominal |
| Dominio | [Recuperado, Activo, Fallecido, N/A]  Existen 5 registros únicos |
| Estadísticas Básicas |  |
| Observaciones | * Existen 11´917.577 NULL en la variable |
| **Campo 53** | FechaDefuncionConfirmado |
| Alias | Fecha de defunción confirmada |
| Descripción | Fecha confirmada de defunción del paciente |
| Tipo | Fecha |
| Dominio | [AAAA-MM-DD]  Existen 157 registros únicos |
| Estadísticas Básicas |  |
| Observaciones | * Existen 13´224.285 NULL en la variable |
| **Campo 54** | Cerebrovascular |
| Alias | Cerebrovascular |
| Descripción | Paciente con Cerebrovascular |
| Tipo | Categórico Nominal |
| Dominio | [SI, NO]  Existen 2 códigos únicos |
| Estadísticas Básicas |  |
| Observaciones | * Existen 11´917.577 NULL en la variable |
| **Campo 55** | HTA |
| Alias | HTA |
| Descripción | Paciente con Hipertensión Arterial |
| Tipo | Categórico Nominal |
| Dominio | [SI, NO]  Existen 2 códigos únicos |
| Estadísticas Básicas |  |
| Observaciones | * Existen 11´917.577 NULL en la variable |
| **Campo 56** | DM |
| Alias | DM |
| Descripción | Paciente con Diabetes Mellitus |
| Tipo | Categórico Nominal |
| Dominio | [SI, NO]  Existen 2 códigos únicos |
| Estadísticas Básicas |  |
| Observaciones | * Existen 11´917.577 NULL en la variable |
| **Campo 57** | **Renal** |
| Alias | Renal |
| Descripción | Paciente con Renal |
| Tipo | Categórico Nominal |
| Dominio | [SI, NO]  Existen 2 códigos únicos |
| Estadísticas Básicas |  |
| Observaciones | * Existen 11´917.577 NULL en la variable |
| **Campo 58** | Tiroides |
| Alias | Tiroides |
| Descripción | Paciente con Tiroides |
| Tipo | Categórico Nominal |
| Dominio | [SI, NO]  Existen 2 códigos únicos |
| Estadísticas Básicas |  |
| Observaciones | * Existen 11´917.577 NULL en la variable |
| **Campo 59** | Obesidad |
| Alias | Obesidad |
| Descripción | Paciente con Obesidad |
| Tipo | Categórico Nominal |
| Dominio | [SI, NO]  Existen 2 códigos únicos |
| Estadísticas Básicas |  |
| Observaciones | * Existen 11´917.577 NULL en la variable |
| **Campo 60** | Cardiaca |
| Alias | Cardiaca |
| Descripción | Paciente con Cardiaca |
| Tipo | Categórico Nominal |
| Dominio | [SI, NO]  Existen 2 códigos únicos |
| Estadísticas Básicas |  |
| Observaciones | * Existen 11´917.577 NULL en la variable |
| **Campo 61** | Respiratoria |
| Alias | Respiratoria |
| Descripción | Paciente con Respiratoria |
| Tipo | Categórico Nominal |
| Dominio | [SI, NO]  Existen 2 códigos únicos |
| Estadísticas Básicas |  |
| Observaciones | * Existen 11´917.577 NULL en la variable |
| **Campo 62** | Cancer |
| Alias | Cancer |
| Descripción | Paciente con Cancer |
| Tipo | Categórico Nominal |
| Dominio | [SI, NO]  Existen 2 códigos únicos |
| Estadísticas Básicas |  |
| Observaciones | * Existen 11´917.577 NULL en la variable |
| **Campo 63** | Autoinmune |
| Alias | Autoinmune |
| Descripción | Paciente con Autoinmune |
| Tipo | Categórico Nominal |
| Dominio | [SI, NO]  Existen 2 códigos únicos |
| Estadísticas Básicas |  |
| Observaciones | * Existen 11´917.577 NULL en la variable |
| **Campo 64** | VIH |
| Alias | VIH |
| Descripción | Paciente con VIH |
| Tipo | Categórico Nominal |
| Dominio | [SI, NO]  Existen 2 códigos únicos |
| Estadísticas Básicas |  |
| Observaciones | * Existen 11´917.577 NULL en la variable |

### Análisis preliminares

**Variables numéricas**

* La edad promedio es de aproximadamente 60.32 años, lo que sugiere que la población analizada tiende a ser adultos mayores.
* El rango de edad va desde 0 hasta 99 años, lo que indica una variedad amplia de edades en la muestra.
* Se puede indicar que la mayoría de la población se agrupa en torno a la media de 60 años.

**Variables categóricas**

* En la variable ‘Sexo’ predomina el sexo femenino, seguido por el masculino, con algunos registros nulos o no definidos.
* En el régimen de afiliación, la mayoría de los individuos están afiliados al régimen contributivo, seguido por el régimen subsidiado. Una porción significativa de la población está contribuyendo al sistema de salud. Algunos registros son no reportados.
* Las vacunas Pfizer y Sinovac son las más utilizadas, lo que podría reflejar la disponibilidad o preferencia de estas vacunas en las campañas de vacunación. En términos geográficos, la mayor parte de las aplicaciones se concentra en Bogotá, D.C., Antioquia y Valle del Cauca, de lo cual se podría deducir una mayor actividad de vacunación en estas regiones.
* La información sobre los sitios de defunción y el estado de recuperación muestra que la mayoría de los registros carecen de información detallada, lo que indica que hay un problema de calidad en los datos recopilados, con registros incompletos o incluso inexactos.
* Muchos registros no tienen información sobre fechas de diagnóstico de condiciones como hipertensión, diabetes, ERC, VIH y cáncer.