



2024

Misión 2:

Actividad 6

Esneider Enrique Guzmán Pérez
Andrea Carolina Mendoza Zabaleta

Sectores:
Salud, educación y logística

23/10
Cartagena



Universidad
Tecnológica
de Bolívar

Contenido

Introducción	4
Objetivos	4
Objetivo general	4
Objetivos específicos	4
Contextualización del problema.....	4
Sectores para abordar	5
Descripción de los sectores.....	5
Importancia de los sectores	5
Identificación de la fuente de datos	6
Problemas por resolver.....	6
Pregunta de negocio:	7
Desafíos y como abordarlos en la identificación y recopilación de la fuente de datos.....	7
Aporte de la analítica de los datos al problema a desarrollar	10
Metodología propuesta	10
Tipo de analítica	10
Metodología CRISP-DM	10
Herramientas para utilizar.....	11
Entorno de desarrollo integrado (IDE): VSCode	11
Lenguaje de programación para el análisis: PYTHON	12
Análisis exploratorio de datos (EDA)	13
Complejidad de los datos.....	13
Complejidad datos originales	14
Complejidad datos luego de limpieza	14
Información demográfica de los encuestados.....	15
Mapa de calor para la correlación	18
Valores atípicos	20
Valores atípicos edad.	20
Valores atípicos trabajadores	21
Referencias	22

Contenido Graficas

Gráfica 1 información demográfica de las personas encuestadas.....	15
Gráfica 2 Distribución de género en los encuestados.....	16
Gráfica 3 Población del conflicto armado o con discapacidad.	17
Gráfica 4 Distribución porcentual de la población.	17
Gráfica 5 Distribución de la población por edades.....	18
Gráfica 6 Matriz de Correlación de Variables Relacionadas con Informalidad Laboral y Seguridad Social.	19
Gráfica 7 Caja de bigote de la columna Edad.....	21
Gráfica 8 Valores atípicos columna ¿Cuántos trabajadores tiene?.....	22
Gráfica 9 Valores atípicos columna ¿Cuántos trabajadores extranjeros tiene? ...	22

Contenido Tablas

Tabla 1 Información de las variables presentes en el dataset.	9
Tabla 2 Completitud datos originales	14
Tabla 3 Completitud datos limpiados.....	14
Tabla 4 Clasificación de edades por rangos.	17

Introducción

El mercado de Bazurto, ubicado en Cartagena, Bolívar, es un epicentro de actividad comercial caracterizado por un alto nivel de informalidad laboral, donde muchos comerciantes y sus empleados carecen de contratos formales y prestaciones sociales. Esta informalidad presenta retos significativos tanto para los trabajadores, que quedan desprotegidos, como para la organización logística del mercado. Este proyecto tiene como objetivo analizar la base de datos de la actividad comercial en Bazurto utilizando herramientas de análisis de datos en Python, con el fin de hacer caracterización sociodemográfica de la población trabajadora. Este análisis permitirá no solo identificar patrones en la informalidad y así llegar a comprender como estas situaciones generan una espiral de pobreza y vulnerabilidad que afecta notoriamente la calidad de vida de los implicados, sino también generar soluciones prácticas que contribuyan al desarrollo sostenible de la actividad comercial.

Objetivos

Objetivo general

Analizar datos sociodemográficos de los comerciantes del mercado de Bazurto de Cartagena y proponer estrategias para impactar en la cultura del comerciante y por consiguiente en sus espacios de trabajo y logística de su actividad comercial

Objetivos específicos

1. **Identificar los tipos de contratos laborales predominantes** en los comerciantes del mercado de Bazurto
2. **Evaluar el acceso a prestaciones sociales** (salud, pensiones, seguridad social) en los comerciantes del mercado de Bazurto
3. **Estudiar el género predominante** en los comerciantes del mercado de Bazurto
4. **Estudiar el nivel de escolaridad de** los comerciantes del mercado de Bazurto

Contextualización del problema

El mercado de Bazurto en la ciudad de Cartagena, Bolívar se ha caracterizado por ser una gran fuente de empleo informal, esta informalidad, aunque ha generado un entorno económico sostenible en el tiempo, se conforma

de muchos trabajadores sin garantías de estabilidad laboral y **salud**, lo que posiblemente puede relacionarse con su **nivel de escolaridad**, y traer como consecuencia el desorden en el espacio y la **logística** de la actividad comercial y su entorno más cercano. A partir de encuestas aplicadas a los comerciantes o trabajadores de esta zona es posible desarrollar un análisis de datos que permita proponer estrategias que minimicen la informalidad y mejoren la calidad de vida del comerciante y por consiguiente de la actividad comercial.

Sectores para abordar

Descripción de los sectores

En la **salud** las prestaciones de los comerciantes son informal o inexistente con ocasión al tipo de contratación que en ocasiones o en su mayoría puede ser informal y/o diaria.

En la **educación**, el nivel de escolaridad de los trabajadores del mercado de Bazurto de Cartagena, Bolívar en la mayoría de los casos puede no ser lo esperado en relación con su edad, así como también es posible que su proceso formativo se encuentre en pausa indefinida o no se encuentre vinculado a ningún proceso de formación técnico, tecnológico o profesional.

En términos de **logística** la forma de ofertar y acceder a los productos y servicios en el mercado de Bazurto de Cartagena Bolívar genera contaminación visual, estética y vial en su entorno inmediato.

Importancia de los sectores

Estudiar el **sector de la salud** de los comerciantes del mercado de Bazurto es fundamental porque carecen de acceso a sistemas de salud adecuado esto los deja en una situación vulnerable frente a enfermedades o accidentes laborales, afectando su calidad de vida y su capacidad para sostener su actividad económica. Además, la exposición diaria a condiciones insalubres y los riesgos inherentes al tipo de trabajo que realizan, como la manipulación de alimentos o la falta de infraestructura adecuada, pueden empeorar la salud de los comerciantes. Finalmente, el bienestar de los comerciantes también impacta directamente en la logística y sostenibilidad del mercado. Un sector saludable asegura una mayor productividad, reduce las ausencias laborales por enfermedad y, a largo plazo, fortalece la estabilidad económica y social de los trabajadores y del mercado en general.

Estudiar el **sector de la educación** de los comerciantes del mercado de Bazurto es crucial para comprender cómo el nivel educativo de estos trabajadores

influye en su desarrollo económico y social. La educación está directamente relacionada con la capacidad de acceder a mejores oportunidades laborales, entender derechos laborales, y gestionar de manera eficiente sus negocios, lo que puede contribuir a reducir la informalidad que caracteriza a este sector. Además, una mayor educación en temas como la administración de negocios, la contabilidad básica o el emprendimiento no solo beneficiaría a los comerciantes en términos de crecimiento económico personal, sino que también contribuiría a una **organización** más eficiente del mercado, mejorando la **logística y la calidad** de los servicios ofrecidos. Esto, a su vez, impactaría positivamente en la economía local y la comunidad en general.

Identificación de la fuente de datos

El conjunto de datos fue obtenido de la base de datos gubernamental denominada "Datos abiertos", y corresponde a "Caracterización comerciantes mercado Bazurto Cartagena 2021", la cual fue actualizada por última vez el 13 de julio de 2022, la descripción este conjunto de datos relaciona la actividad económica, manejo desechos, afiliación a salud, pensión y demás de los encuestados, si los sitios de venta son formales o informales, zonas de descargue, características de puestos de venta y negocio, periodicidad de abastecimiento, entre otros aspectos. (Colombia, 2022)

Los filtros que se tuvieron en cuenta para poder escoger el Dataset seleccionado fueron los siguientes:

- Departamento: Bolívar.
- Municipio: Cartagena.
- Categoría: Vivienda, ciudad y territorio.
- Tipos de vistas: Conjunto de datos.

Problemas por resolver

El mercado de Bazurto en Cartagena se ha consolidado como una importante fuente de empleo informal, permitiendo el sostenimiento económico de muchas familias a lo largo del tiempo. Sin embargo, esta informalidad laboral plantea desafíos significativos que afectan no solo a los comerciantes y sus empleados, sino también al entorno urbano donde operan. La ausencia de contratos laborales formales y el limitado acceso a prestaciones sociales, como salud, pensiones y seguridad social, son características comunes entre quienes laboran en este mercado. Esta situación se relaciona potencialmente con el bajo nivel de escolaridad de los comerciantes, lo que a su vez puede contribuir a la desorganización del espacio comercial y su entorno inmediato.

Pregunta de negocio:

¿Cómo impacta la informalidad laboral, caracterizada por la falta de contratos formales y acceso a prestaciones sociales, en la organización espacial, logística comercial y calidad de vida de los comerciantes del mercado de Bazurto en Cartagena, y qué estrategias pueden implementarse para reducir esta informalidad y mejorar las condiciones laborales y económicas de sus trabajadores?

Desafíos y como abordarlos en la identificación y recopilación de la fuente de datos.

Tras revisar el conjunto de datos descargados de la fuente de datos abierta, se descubrió que varias columnas contenían datos irrelevantes, incompletos o redundantes. Algunas columnas contenían valores con signos de puntuación, interrogación o exclamación y un formato incoherente. El conjunto de datos inicial constaba de unas 240 columnas, muchas de las cuales procedían de preguntas con varias secciones o categorías, en función de la clasificación de la columna anterior.

Para resolver esta situación se llevó a cabo un ejercicio de limpieza de datos en varias etapas:

1. **Eliminación de columnas irrelevantes o vacías:** se eliminaron las columnas que no contenían datos relevantes o tenían un gran número de valores cero. Estas columnas no aportaban información útil para el análisis y podían introducir ruido en los resultados.
2. **Agrupación de columnas con la misma clasificación:** en varios casos, las respuestas a una misma pregunta o categoría estaban dispersas en varias columnas en las que se utilizaba un formato binario (0 o 1) para indicar la presencia de una respuesta. Estas columnas se agruparon en una única columna que representaba la clasificación, en la que el valor «1» (presencia) en cada columna original se agrupaba en una única respuesta.
3. **Normalización de los datos:** Los datos se normalizaron eliminando los caracteres no deseados, como los signos de puntuación, y ajustando el formato de las respuestas para garantizar la coherencia y homogeneidad del conjunto de datos.
4. **Cambio de nombre de las columnas para mejorar la claridad:** tras agrupar y consolidar las columnas, se les cambió el nombre por descripciones más intuitivas y fáciles de entender, preservando la integridad de la clasificación original, pero facilitando el análisis y la interpretación de los datos.

El resultado fue un conjunto de datos más compacto, con unas 23 columnas, pero conservando 3042 filas de datos. Este conjunto de datos es ahora más manejable y está libre de datos redundantes, lo que permite un análisis de datos preciso y orientado a la resolución de problemas. Además, la agrupación de columnas similares en una única categoría facilita la identificación de patrones y tendencias subyacentes que pueden utilizarse para proponer estrategias de

resolución eficaces.

En la tabla 1 podemos observar cómo se distribuye el dataset luego de ser limpiado

<i>índex</i>	Variable	Tipo	Significa
0	_índex	int64	Valor de la fila
1	fecha	string	Fecha en la cual se realizó la encuesta
2	Edad	int64	Edad de la persona encuestada
3	Genero	string	Genero de la persona encuestada
4	Grupo étnico	string	Grupo étnico de la persona encuestada
5	Víctima del conflicto	int64	La persona fue víctima del conflicto armado en Colombia
6	Tipo de discapacidad	string	La persona posee alguna discapacidad y cual
7	¿Cuenta con SISBEN?	string	La persona cuenta siben
8	Hora de Apertura del Comercio	string	Horario de apertura del negocio
9	Hora Cierre del Comercio	string	Horario de cierre del negocio
10	¿Cuántos trabajadores tiene?	int64	Cantidad total de trabajadores
11	¿Cuántos trabajadores extranjeros tiene?	float64	Cantidad total de trabajadores extranjeros
12	¿Tiene registro ante la DIAN (RUT)?	string	Tiene registro tributario
13	Hora de recepción de mercancía	String	Tiene hora de recepción de la mercancía
14	Inversión de mercancía	string	Cantidad de dinero invertido en mercancía
15	Tipo de producto/servicio	string	Tipo de producto o servicio ofrecido por la persona
16	Afiliación a EPS	string	la persona está afiliada a la eps y en que régimen
17	Medio de transporte	string	medio de transporte para llegar a Bazurto
28	Nivel de escolaridad	string	Nivel educativo de la persona
19	seguridad social sin salud	string	La persona está afiliada a la seguridad social excluyendo salud
20	Manejo de basuras	string	Como se manejan las basuras en el local
21	Estado de Registro Cámara	string	Contiene el estado de registro del establecimiento ante la cámara de comercio.
22	Tipo de contrato	string	Contiene información sobre el tipo de contrato que poseen los empleados.

Tabla 1 Información de las variables presentes en el dataset.

Aporte de la analítica de los datos al problema a desarrollar

El análisis que se llevará a cabo no se limita a un solo sector, debido a la interrelación entre la informalidad laboral de los comerciantes y el acceso limitado a prestaciones sociales, lo cual justifica la inclusión de los sectores económico y de salud como ejes principales.

Metodología propuesta

Tipo de analítica

Un análisis **descriptivo y prescriptivo** es fundamental para estudiar entornos complejos como el mercado de Bazurto en Cartagena, dado su alto nivel de informalidad y su influencia en el contexto socioeconómico local. En primer lugar, un análisis descriptivo permite caracterizar el estado actual del mercado en términos de variables sociodemográficas y comerciales, tales como el tipo de empleo, nivel de escolaridad, y acceso a servicios básicos de los comerciantes, ofreciendo un perfil detallado de la población y del funcionamiento del mercado (Heckman et al., 2006). Esta información es esencial para entender el alcance de los problemas de informalidad y los patrones económicos y sociales que se generan en el sector (Johnson, Kaufmann, & Zoido-Lobatón, 1998). Por otro lado, un **análisis prescriptivo** proporciona recomendaciones basadas en los **hallazgos descriptivos**, proponiendo estrategias que fomenten la formalización del empleo y mejoren la logística y el ordenamiento del mercado. Estas recomendaciones podrían incluir desde políticas de inclusión laboral hasta la implementación de programas educativos que busquen elevar el nivel de escolaridad y acceso a prestaciones sociales, lo cual contribuiría a mitigar la vulnerabilidad económica de los trabajadores y, en última instancia, a promover un espacio más armonioso y estructurado (Bai & Wu, 2014).

Metodología CRISP-DM

Para llevar a cabo un análisis integral de la actividad comercial en el mercado de Bazurto en Cartagena, se empleó la metodología CRISP-DM, que provee un enfoque estructurado para el desarrollo de proyectos de minería de datos en etapas claras. En la primera fase, de comprensión del negocio, se definió como objetivo identificar las características sociodemográficas de los comerciantes y

trabajadores de este mercado, para así proponer estrategias orientadas a reducir la informalidad y mejorar la sostenibilidad del entorno. A partir de esta comprensión, se desarrollaron preguntas clave para guiar el análisis, como la relación entre el nivel de escolaridad y las condiciones laborales, el acceso a prestaciones sociales, y el impacto de la informalidad en la calidad de vida. Además, se recopiló información socioeconómica y demográfica obtenida a partir de encuestas aplicadas y de bases de datos gubernamentales, especialmente de la plataforma Datos Abiertos, que sirvieron de base para los análisis posteriores.

La siguiente fase, de comprensión de los datos, implicó la limpieza y exploración inicial de los datos, ejecutada principalmente con Python. Este proceso permitió evaluar la calidad de los datos, verificar su consistencia e identificar patrones preliminares. En la etapa de preparación, los datos fueron estructurados en función de las variables relevantes, como tipo de contrato laboral, nivel educativo y acceso a servicios de salud y seguridad social. Posteriormente, en la fase de modelado, se emplearán técnicas de análisis prescriptivo, que facilitara la generación de recomendaciones y estrategias para impactar en el sector educativo y mejorar la logística del entorno comercial. Las fases de evaluación y despliegue culminaran con la interpretación de los resultados y el diseño de un plan de acción orientado a fomentar el acceso a condiciones laborales más dignas y a mejorar la sostenibilidad de la actividad económica en el mercado de Bazurto.

Herramientas para utilizar

En esta sección se hace necesario determinar las herramientas necesarias que serán utilizadas para poder realizar el análisis de los datos antes propuestos.

Entorno de desarrollo integrado (IDE): VSCode

“VSCode es un editor de código gratuito, de código abierto y multiplataforma, siendo una herramienta que nos permite editar el texto plano de los archivos de código para programación (Zúñiga, 2024)”, sin embargo, detrás de esa funcionalidad básica esta herramienta puede ser configurada para realizar diferentes tipos de tareas entre las que podemos encontrar:

- Edición y desarrollo de código fuente.
- Depuración de aplicaciones y scripts.
- Gestión de proyectos y carpetas de trabajo.
- Control de versiones.
- Trabajo colaborativo.
- Escritura y edición de documentos Markdown.
- Automatización de tareas y flujos de trabajo.

Entre otras tareas que se pueden realizar en este editor, gracias a este editor podemos encontrar diferentes ventajas entre las cuales encontramos:

- Agilidad en la programación.
- Compatibilidad con múltiples lenguajes de programación.
- Herramientas de depuración de código.
- Personalización.

Justificación del uso

Teniendo en cuenta las ventajas y tareas antes mencionadas, los autores de este documento determinaron que el uso de esta herramienta era fundamental debido a su facilidad de integración con el lenguaje que se estaría utilizando en este caso **PYTHON** además de su fácil integración fue escogido por que es un editor que permite ser ejecutado en condiciones de procesamiento relativamente bajas es decir el uso de este entorno de desarrollo no conlleva un gran uso de memoria ni de almacenamiento para el equipo en el que se esté implementando, también se puede indicar que este IDE al tener diferentes extensiones y su particularidad de ser personalizado permite que los desarrolladores puedan implementar diferentes extensiones que hacen que sea mucho más fácil la visualización de los datos en el caso de análisis.

Lenguaje de programación para el análisis: PYTHON

Python es un lenguaje de programación creado por Guido Van Rossum en 1991 que se valora por su versatilidad, sus capacidades multiplataforma y su naturaleza libre. Se utiliza en una amplia gama de aplicaciones, desde el desarrollo web y de software hasta campos especializados como la ciencia de datos y el aprendizaje automático. Su popularidad se debe principalmente a su curva de aprendizaje eficiente y accesible, que permite a programadores y analistas aprender rápidamente y aplicarlo a diversos proyectos (AWS, 2023).

En el campo del análisis de datos, Python es una herramienta importante debido a su simplicidad y facilidad de uso, así como a una amplia gama de bibliotecas especializadas como Pandas, NumPy, Matplotlib y Scikit-learn que facilitan el procesamiento, el análisis y la visualización de datos. Por tanto, Python es ideal para quienes deseen realizar análisis y visualizaciones de datos en profundidad sin dedicar mucho tiempo a aprender el lenguaje. Además, Python puede integrarse fácilmente con otros entornos y herramientas de datos, lo que aumenta la eficiencia de los proyectos de ciencia de datos y aprendizaje automático.

En particular, el uso de los cuadernos Jupyter es una gran ventaja a la hora de analizar datos. Estos cuadernos permiten escribir código, realizar cálculos y documentar procesos en un entorno interactivo ideal para el análisis exploratorio, la visualización de datos y el desarrollo de modelos. La estructura de los cuadernos Jupyter facilita la creación de informes y la documentación de cada paso del análisis, lo que ayuda a otras personas a reproducir y comprender claramente el proceso de análisis.

Las principales ventajas de utilizar Python para el análisis de datos son las siguientes

- **Facilidad de uso y legibilidad:** La sintaxis de Python es sencilla, directa y fácil de entender y aplicar.
- **Amplia comunidad y recursos:** Existe una gran comunidad de programadores y científicos de datos que han creado un gran número de recursos y bibliotecas.
- **Flexibilidad y extensibilidad:** Python puede manejar desde análisis básicos hasta complejos algoritmos de aprendizaje automático.
- **Compatible con herramientas de datos:** Puede integrarse en herramientas de big data como Apache Spark y Hadoop, y se utiliza ampliamente en espacios de trabajo como Jupyter Notebook.

Justificación del uso

Teniendo en cuenta las ventajas del lenguaje las cuales fueron mencionadas anteriormente, los autores del documento determinaron que el uso de esta herramienta es fundamental debido a su facilidad de aprendizaje manejo de la información y la amplia gama de librerías que este lenguaje maneja en el análisis de datos.

Análisis exploratorio de datos (EDA)

En esta sección se realizará un exploramiento inicial de los datos del dataset denominado "Comerciantes_Bazurto", el cual fue obtenido como ha sido mencionado anteriormente de la pagina datos Colombia, en esta sección se realizará las primeras visualizaciones de los datos luego de ser limpiados, también se mostrará información de vital importancia como lo corresponde el calculo de la completitud de los datos en las filas.

Completitud de los datos.

La completitud de los datos según (POWER DATA, 2024) corresponde al grado en el que todos los atributos necesarios están presentes en el dataset. En otras palabras, representa el porcentaje de datos que cumplen con este criterio, asegurando que no existen valores faltantes o vacíos en los campos donde es indispensable que estén completos, o que, en caso de que existan, sean mínimos para permitir un análisis o procesamiento adecuado de los datos.

Para evaluar esto, los autores de este proyecto definieron dos etapas. Primero, determinaron la completitud de los datos antes de realizar cualquier proceso de limpieza. En una segunda etapa, recalcularon este valor después de la limpieza de los datos para asegurarse de que estuvieran en las mejores condiciones posibles para su análisis.

Completitud datos originales

En esta primera parte los autores determinaron como se encontraba el dataset a nivel de completitud de los datos lo cual arrojo lo que se puede observar en la segunda tabla.

<i>Tamaño total</i>	Total, de valores	Total, de valores faltantes	Porcentaje total de valores completos	Porcentaje total de valores faltantes
3042 x 227	690534	310323	55.06 %	44.94 %

Tabla 2 Completitud datos originales

Como podemos observar en la tabla anterior en el dataset original contamos con una matriz de 3042 filas y 227 columnas en la cual existe un porcentaje bastante alto de valores faltantes, además de lo anterior podemos observar que existen demasiadas columnas que al ser revisadas de forma minuciosa no son necesarias para nuestro análisis puesto que incluyen información no necesaria como lo corresponde por dar un ejemplo de estas las siguientes columnas.

- ¿Cuál fue el principal motivo por el que cambió su lugar de residencia?
- ¿En dónde hace el descargue o dónde recibe su mercancía?/Sector La Rampa
- En su mayoría, ¿de dónde proviene su mercancía o dónde compra sus productos?
- ¿Qué vehículo utiliza para su sistema de abastecimiento?/Camioneta tipo platón (1 ton)

Completitud datos luego de limpieza

En esta segunda parte los autores determinaron como se encontraba el dataset a nivel de completitud de los datos luego de realizarse una limpieza del dataset esta limpieza fue realizada como se menciona en la sección denominada **"Desafíos y como abordarlos en la identificación y recopilación de la fuente de datos."** Esta limpieza nos arrojo los resultados obtenidos se encuentran visualizados en la tercera tabla.

<i>Tamaño total</i>	Total, de valores	Total, de valores faltantes	Porcentaje total de valores completos	Porcentaje total de valores faltantes
3042 x 23	69966	991	98.58 %	1.42 %

Tabla 3 Completitud datos limpiados.

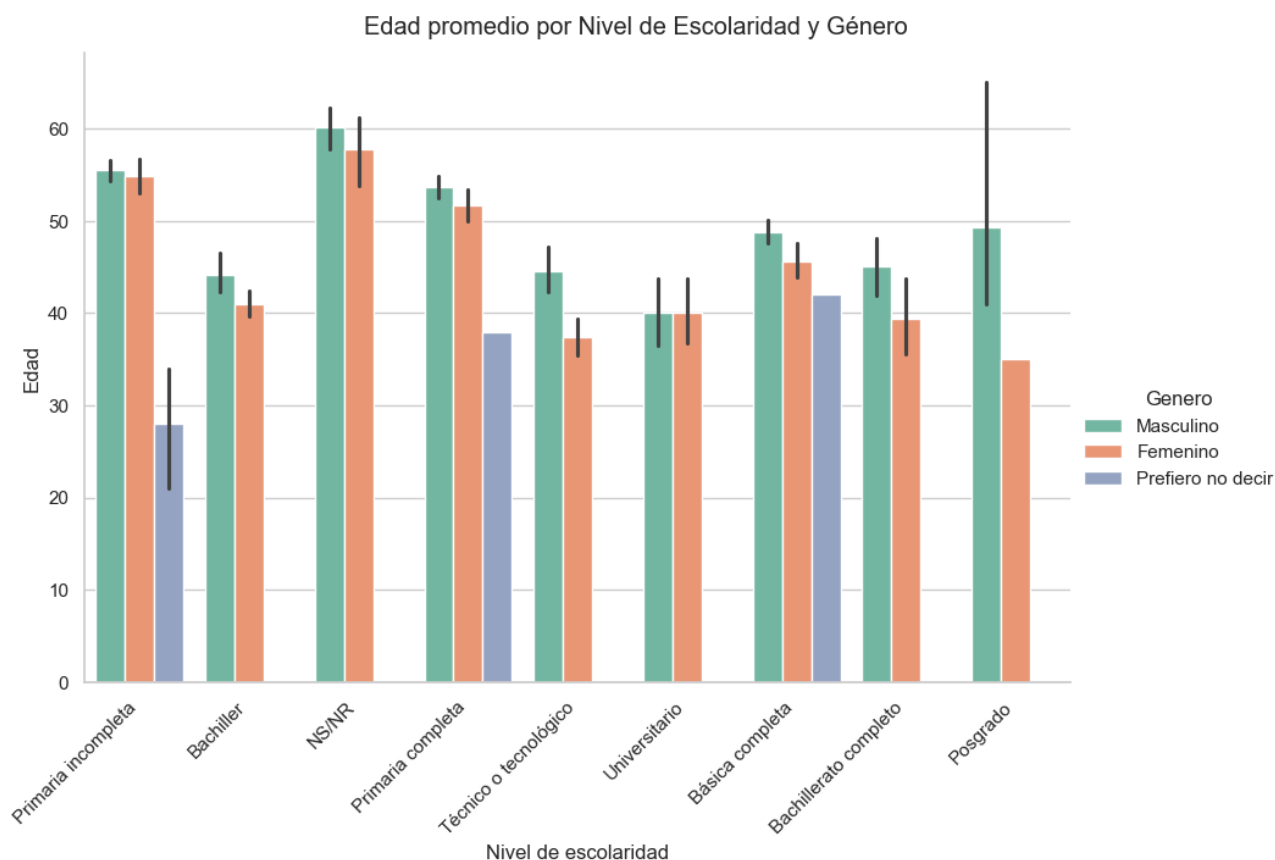
Como podemos observar la dimensión del dataset disminuyo a nivel de columnas disminuyo en un 85.18 % en comparación a la matriz original, en otras palabras los autores lograron eliminar y o agrupas un total de 204 columnas lo

cual permite que se realice un análisis mucho más específico, también se pudo observar que se aumentó en un 48.52 % en la completitud de los datos, lo cual es debido a que se remplazaron valores nulos por valores apropiados en este caso generalmente se agregó en campos donde se podían valores como NS/NR puesto que es información que no afecta directamente cómo se comportan los datos, al ser una encuesta en muchos casos se utilizó información nominal.

Información demográfica de los encuestados

Al realizar un desglose de la información obtenida del dataset, podemos buscar y determinar el lugar de origen de las personas, factores como la edad, el género, la ocupación y el nivel educativo, que pueden influir en sus características socioeconómicas, por tanto, a lo largo de esta sección se buscara conocer información que nos permita comprender lo antes mencionados.

En la primera grafica podemos observar el comportamiento demográfico teniendo en cuenta variables como lo son las edades, género y el nivel de escolaridad de las personas encuestadas al momento de realizar la caracterización de los empleados y comerciantes de las personas que trabajan en el mercado de Bazarito.



Gráfica 1 información demográfica de las personas encuestadas.

Si revisamos los datos anteriores podemos determinar que las personas que

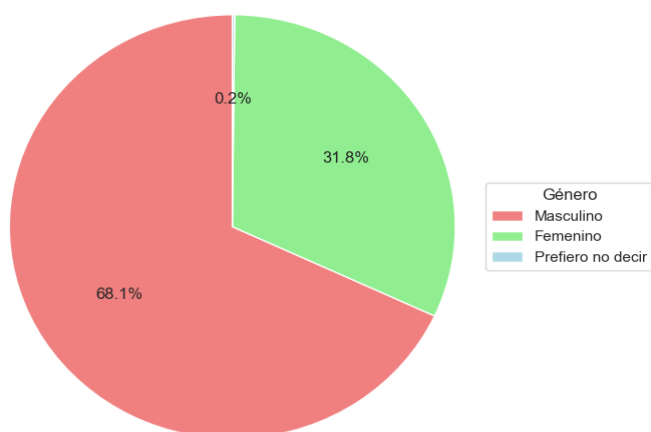
hacen parte de la economía del mercado de Bazurto a nivel de escolaridad es bastante variada contando con personas con posgrado hasta personas que no han sido capaz por diferentes motivos de obtener o finalizar sus estudios básicos como lo corresponden a la primaria, al revisar con detenimiento la gráfica observamos una característica la cual corresponde a NS/NR, esta característica nos puede indicar dos cosas, las cuales corresponden a:

- Falta de claridad o desconocimiento: en esta primera opción se puede indicar que puede existir un problema con la claridad del cuestionario con el cual se realizó la construcción del dataset el cual está siendo utilizado para realizar estos análisis o en su defecto que las personas encuestadas que contestaron de esta forma no se encuentran muy segura de conocer cuál es su nivel de escolaridad.
- Diversidad de experiencias educativas: En algunos contextos, nuestra respuesta "NS/NR", puede reflejar experiencias educativas no tradicionales que se ajustan bien a las categorías estándares de nivel de escolaridad.

Si seguimos realizando una exploración básica de la información demográfica del grupo de individuos encuestados podemos observar la información que se encuentra depositada en la segunda grafica.

Al observar esta gráfica podemos indicar que el 68,1 % de la población que realiza actividades económicas en el mercado de Bazurto corresponden a hombres esto puede deberse a múltiples factores uno de los principales factores corresponde al siguiente.

Distribución de Género en los encuestados



Gráfica 2 Distribución de género en los encuestados.

de Basurto.

- Actividades realizadas: en el mercado de Bazurto se realizan diferentes actividades económicas, en muchas se necesita mano de obra para el cargue y descargue de camiones donde se transporta la mercancía, este tipo de actividades son generalmente desarrolladas por hombres, debido a temas fisionómicos.

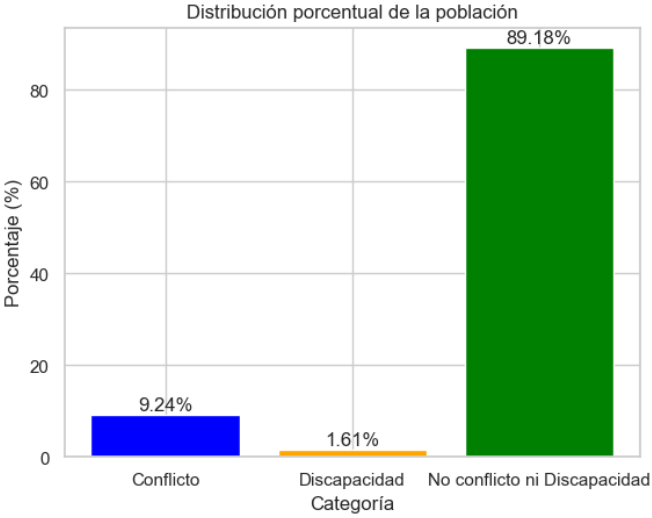
Además, podemos observar en la gráfica anterior que solo el 0.2% de la población indica no responder, esta respuesta puede basarse por los complejos que existen en la población que pertenece al mercado

En estos momentos teniendo en cuenta que estamos validando información demográfica podemos identificar la población que fue víctima de conflicto armado o en su caso si cuentan con alguna discapacidad, en la tercer grafica podemos observar información correspondiente a esta.

En esta grafica podemos observar la cantidad de personas que son víctimas del conflicto armado o poseen una discapacidad si tenemos en cuenta este valor en base a la cantidad de encuestados los cuales fueron 3042 podemos determinar la distribución porcentual de las poblaciones que se encuentran en nuestro dataset, en la cuarta grafica podemos observar la distribución porcentual de la población teniendo en cuenta si la población se divide en tres grupos los cuales corresponden a personas discapacitadas, personas víctimas del conflicto armado y personas que no se encuentran en las distribuciones anteriores.



Gráfica 3 Población del conflicto armado o con discapacidad.



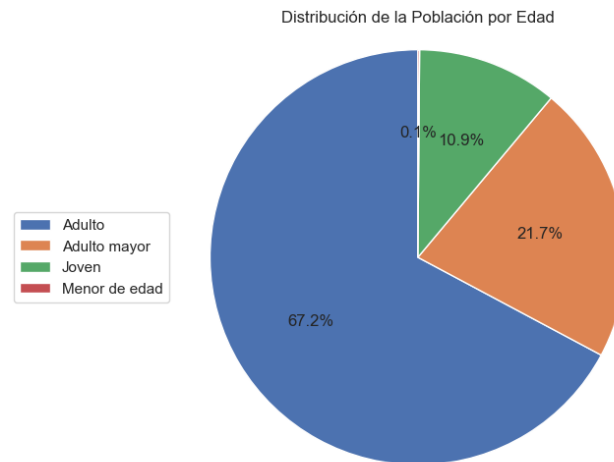
Gráfica 4 Distribución porcentual de la población.

Al analizar las gráficas anteriores podemos indicar que la minoría de personas que realizan actividades económicas en el mercado de Bazurto pertenecen a los grupos poblacionales identificados como vulnerables es decir personal víctima del conflicto armado y personas con discapacidad, gracias a esta validación podemos indicar que este no será un aspecto de gran importancia para determinar los motivos por los cuales las personas que realizan actividades económicas en ese sector no cuenten con seguridad social.

En la quinta grafica la cual se basa en un diagrama de pastel podemos observar la distribución de la población que se encuentra en nuestro estudio, para realizar esta distribución se tuvieron en cuenta la información de la cuarta tabla.

Rango	Definición
$edad < 18$	Menor de edad
$18 \leq edad < 30$	Joven
$30 \leq edad < 60$	Adulto
$edad > 60$	Adulto mayor

Tabla 4 Clasificación de edades por rangos.



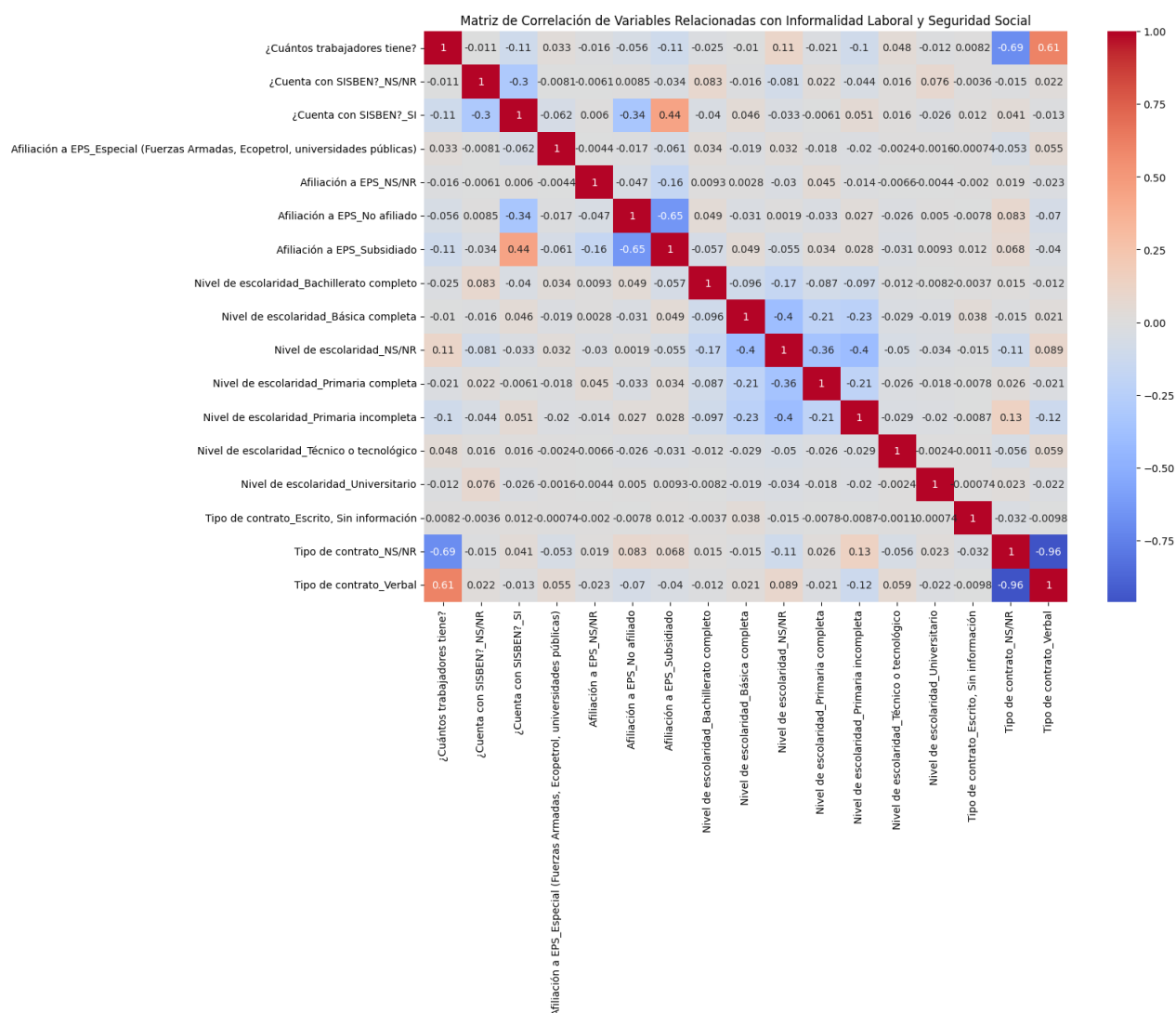
Gráfica 5 Distribución de la población por edades.

Mapa de calor para la correlación

Un mapa de calor para la correlación corresponde a una herramienta utilizada para la visualización de las relaciones existentes entre múltiples variables, según (ANALYTIC LANE, 2024) "estos tipos de gráficos permiten analizar de forma más fácilmente las relaciones entre las variables del dataframe", para la construcción de este tipo de gráficos se utiliza la biblioteca Seaborn de Python, para realizar este grafico se utilizaron diferentes columnas entre ellas encontramos las siguientes:

- ¿Cuenta con SISBEN?
- ¿Cuántos trabajadores tiene?
- Afiliación a EPS
- Nivel de escolaridad
- Tipo de contrato

En la sexta grafica podemos observar el mapa de calor desarrollado con estas columnas.



Gráfica 6 Matriz de Correlación de Variables Relacionadas con Informalidad Laboral y Seguridad Social.

A partir de la gráfica anterior podemos realizar diferentes conclusiones entre ellas podemos revisar las siguientes.

1. Tipo de contrato y nivel de escolaridad: Existe una correlación directa entre algunos niveles de escolaridad y el tipo de contrato, entre ellas podemos encontrar.
 - a. El *tipo de contrato verbal* tiene una fuerte correlación positiva (0.61) con el número de trabajadores. Esto podría indicar que los comerciantes con contratos informales verbales suelen emplear más trabajadores, posiblemente porque estos contratos son más comunes en empleos no regulados y temporales.
 - b. El *tipo de contrato escrito, sin información*, y el *tipo de contrato verbal* tienen una correlación negativa significativa con el *nivel de escolaridad universitario* y *técnico o tecnológico*. Esto sugiere que los individuos con mayor nivel educativo tienen menos probabilidad de estar en condiciones laborales informales
2. Afiliación a EPS y tipo de contrato: Existe una correlación directa entre el tipo de afiliación y el tipo de contrato con el que cuentan los trabajadores esto se debe a lo siguiente.
 - a. La afiliación a un tipo de EPS como subsidio, especial o no afiliado

muestra una relación con los tipos de contrato. Por ejemplo, la afiliación subsidiada tiene una ligera correlación positiva con tipos de contrato verbales o sin documentación formal, lo cual puede ser indicativo de que trabajadores en la informalidad tienden a estar afiliados a EPS subsidiadas debido a los menores ingresos que suelen percibir.

Valores atípicos

Los valores atípicos corresponden a valores que se alejan de la línea de los mínimos cuadrados, es decir tienen grandes "errores", donde el "error" o residual es la distancia vertical de la línea al punto según (Openstax, 2024) Los valores atípicos deben examinarse de cerca. A veces, por una u otra razón, no deben incluirse en el análisis de los datos. Es posible que un valor atípico sea el resultado de datos erróneos. Otras veces, un valor atípico puede contener información valiosa sobre la población estudiada y debe seguir incluyéndose en los datos. La clave está en examinar cuidadosamente las causas de que un punto de datos sea un valor atípico.

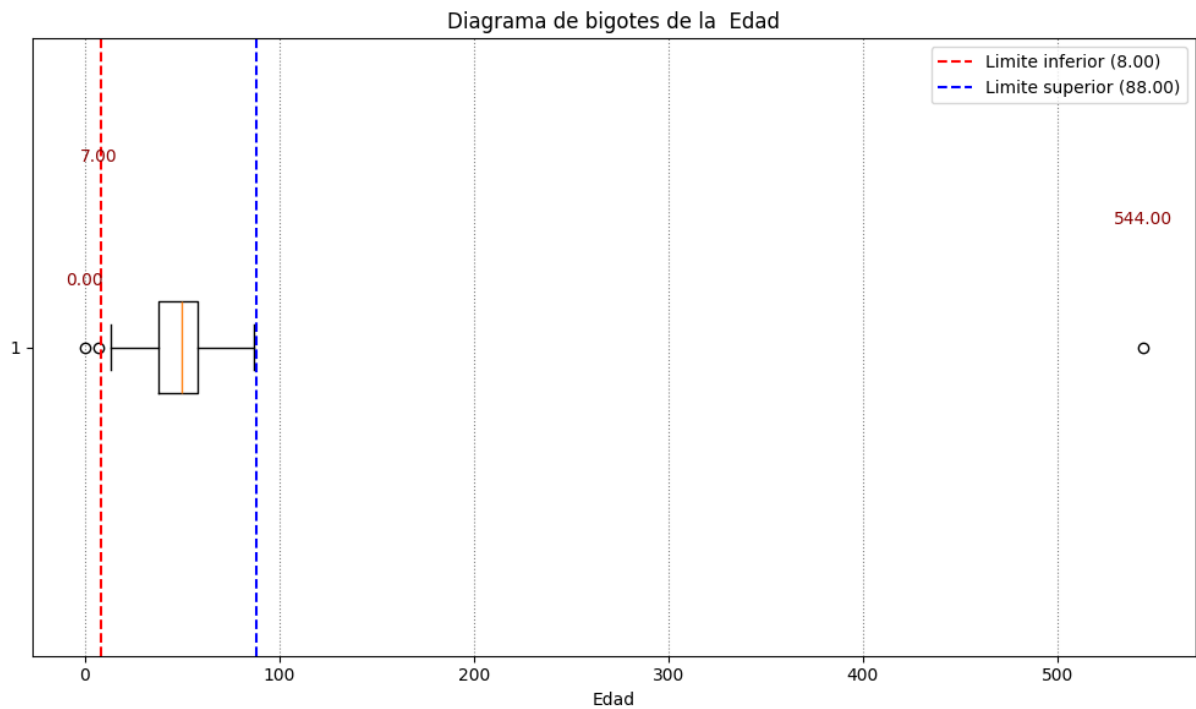
En esta sección se establecerán ciertos valores atípicos que se encontraron en las variables cuantitativas, entre las variables cuantificables encontramos las siguientes

- Edad
- ¿Cuántos trabajadores tiene?
- ¿Cuántos trabajadores extranjeros tiene?

Teniendo en cuenta estas variables podemos determinar estos valores atípicos.

Valores atípicos edad.

Luego de realizar el análisis de valores atípicos de esta columna podemos observar el gráfico que se encuentra en la séptima gráfica, debemos indicar que para obtener esta información utilizamos un diagrama de caja de bigotes.

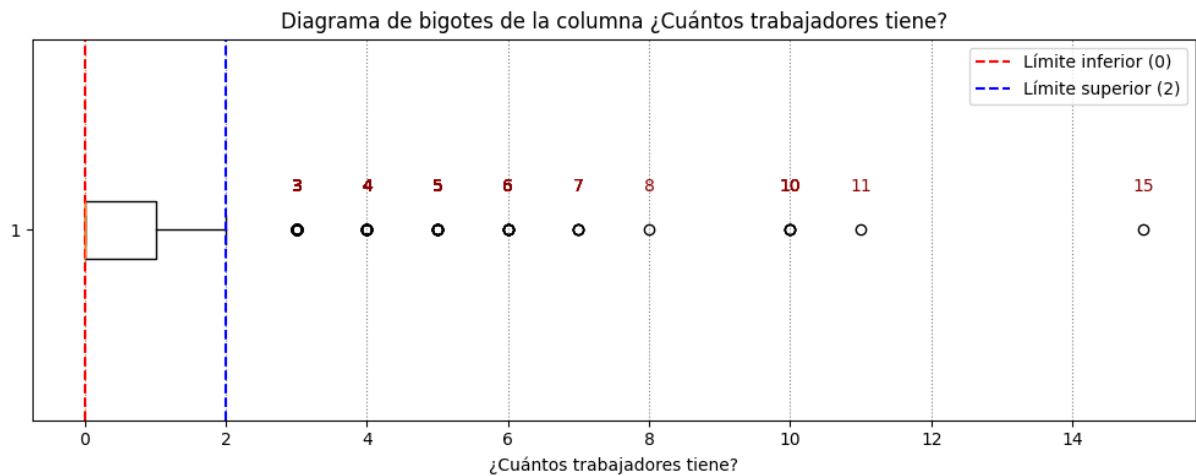


Gráfica 7 Caja de bigote de la columna Edad.

Si observamos esta gráfica podemos comprender que tenemos tres datos atípicos los cuales corresponden personas menores de edad puesto que el limite inferior que encontramos corresponde a la edad de 8 años y un limite superior a 88 pero si observamos tenemos un valor atípico de 544 lo cual es una edad no factible en este caso ese valor será reemplazado por la media, pero los datos correspondiente a edades de 7 y 8 años no serán cambiados por el motivo de que en muchas familias cartageneras los niños comienzan a laborar desde temprana edad.

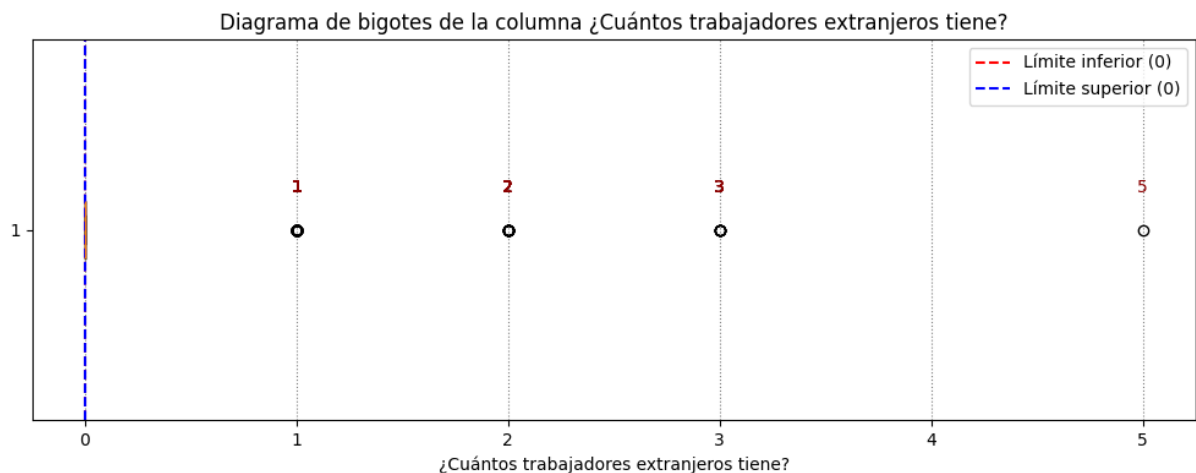
Valores atípicos trabajadores

Luego de realizar el análisis para las columnas referentes a los trabajadores podemos identificar diferentes valores atípicos en la octava y la novena gráficas los valores atípicos de los trabajadores de los negocios teniendo en cuenta el total de los trabajadores en la primera y en la segunda se tiene en cuenta solo los trabajadores extranjeros.



Gráfica 8 Valores atípicos columna ¿Cuántos trabajadores tiene?

Si podemos observar en esta gráfica observamos diferentes valores atípicos donde nuestro límite inferior corresponde entre 0 y el superior el 2 respectivamente, en este caso observamos que encontramos comercios de 15 trabajadores o valores diferentes, en este caso no se pueden cambiar ni eliminar porque corresponde a información necesaria para nuestro análisis.



Gráfica 9 Valores atípicos columna ¿Cuántos trabajadores extranjeros tiene?

Si podemos observar en esta gráfica encontramos valores atípicos donde nuestro límite inferior y superior corresponden al mismo valor 0 en este caso podemos observar que tenemos extranjeros trabajando en los comercios con un valor de hasta 5 personas, estos datos no pueden ser eliminados porque es posible que en nuestro análisis la condición de nacionalidad pueda afectar o no.

Referencias

AWS. (1 de ENERO de 2023). AWS. Obtenido de ¿Qué es Python?:

- <https://aws.amazon.com/es/what-is/python/#:~:text=Guido%20Van%20Rossum%20public%C3%B3%20la,para%20la%20gesti%C3%B3n%20de%20errores>.
- Bai, X. &. (2014). *Sustainable Cities: Assessing and Reducing the Environmental Impact of Urban Infrastructure*. Ámsterdam: A. C. Albert.
- Colombia, G. d. (13 de Julio de 2022). *Datos Abiertos*. Recuperado el 18 de Octubre de 2024, de https://www.datos.gov.co/Vivienda-Ciudad-y-Territorio/Caracterizaci-n-comerciantes-mercado-Bazurto-Carta/6a8s-hznu/about_data: https://www.datos.gov.co/Vivienda-Ciudad-y-Territorio/Caracterizaci-n-comerciantes-mercado-Bazurto-Carta/6a8s-hznu/about_data
- Lochner, L., Heckman, J. J., & Todd, P. (2006). *Handbook of the Economics of Education*. Ámsterdam: Elsevier.
- Openstax. (1 de Enero de 2024). *Openstax*. Obtenido de 12.6 Valores atípicos: <https://openstax.org/books/introducci%C3%B3n-estad%C3%ADstica/pages/12-6-valores-atipicos>
- POWER DATA. (01 de Enero de 2024). *POWER DATA*. Obtenido de Calidad de Datos. Cómo impulsar tu negocio con los datos.: <https://www.powerdata.es/calidad-de-datos>
- Regulatory discretion and the unofficial economy. . (1998). *American Economic Review*, 387-392.
- Zúñiga, F. G. (1 de Julio de 2024). *Arsys*. Obtenido de ¿Qué es Visual Studio Code y cuáles son sus ventajas?: <https://www.arsys.es/blog/que-es-visual-studio-code-y-cuales-son-sus-ventajas>