

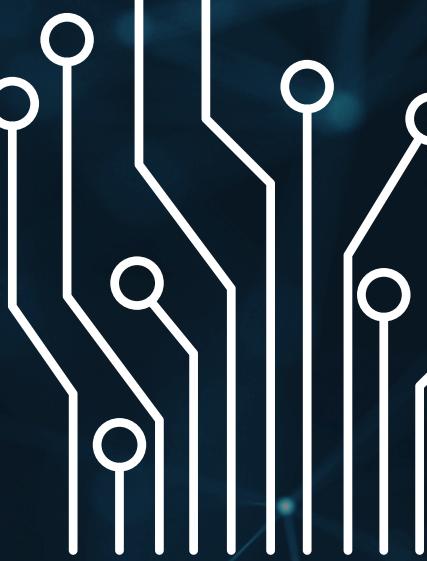
Comision 71945

PRIMERA PREENTREGA DATA SCIENCE II

Seguridad alimentaria

Ramiro Mata

Abstracto: Motivación y Audencia



La seguridad alimentaria es un tema de importancia mundial ya que afecta directamente a la salud pública. En el Reino Unido, la Agencia de Normas Alimentarias (FSA) proporciona calificaciones de higiene alimentaria a diferentes empresas para informar a los consumidores y garantizar los estándares de calidad en la industria de alimentos y bebidas.

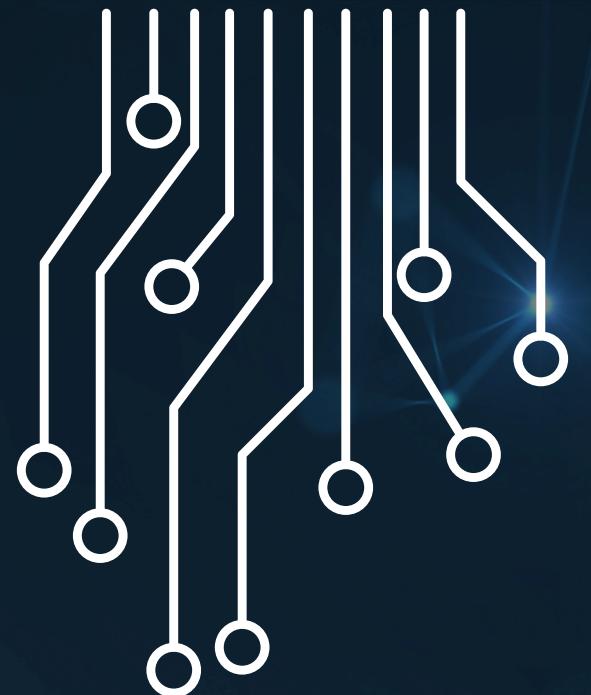
Este análisis se centra en explorar y evaluar estas calificaciones de higiene para:

Determinar patrones de calidad de la salud en diferentes tipos de instituciones.

Examine qué factores afectan la calificación de una empresa.

Proporcionando información útil para decisiones de salud pública, regulatorias y comerciales.

Audiencia Beneficiada



Consumidores

Personas que quieran tomar decisiones informadas sobre dónde comer o comprar alimentos de forma segura.

Analistas de Datos y Científicos de Datos

Profesionales interesados en el análisis de datos relacionados con la seguridad y calidad de los alimentos en la industria alimentaria.

Autoridades de Regulación

Organismos de salud pública y organismos de control que buscan mejorar las condiciones sanitarias en los locales de alimentación.

Dueños de Negocios en la Industria Alimentaria

Restaurantes, cafeterías y supermercados que buscan mejorar sus estándares de higiene y cumplir con regulaciones.

Resumen de Metadata



Información General del Dataset

El conjunto de datos analizado proviene de la API de la Agencia de Normas Alimentarias del Reino Unido y contiene información de calificación de higiene para diferentes empresas. A continuación se muestra un resumen de su estructura:

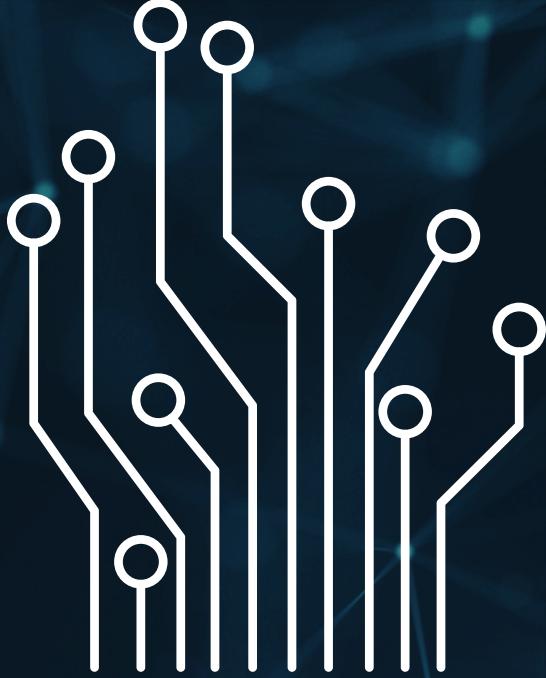
Número de filas: 1,303 registros (establecimientos analizados).

Número de columnas: 25 características sobre cada establecimiento.

Tipos de Variables en el Dataset

- **Variables Categóricas:** incluyen identificadores, nombres de negocios, tipos de establecimientos, esquemas de calificación, y otros campos de texto o booleanos relevantes para clasificar y describir cada entidad.
- **Variables Numéricas:** abarcan identificadores únicos, puntuaciones otorgadas durante la inspección (como higiene, estructura y confianza en la gestión), así como coordenadas geográficas.
- **Variables de Fecha y Localización:** representan la fecha de inspección y componentes de la dirección del establecimiento, útiles para análisis temporales y espaciales.

Valores nulos en el dataset:



- RatingDate tiene **181 valores nulos** (faltan fechas de inspección en algunos establecimientos).
- Geocode tiene **207 valores nulos** (coordenadas geográficas faltantes en algunos registros).
- Scores tiene **232 valores nulos**, ya que algunas inspecciones no incluyen puntuaciones detalladas.
- Las columnas AddressLine3 y AddressLine 4 van a ser eliminadas ya que tienen muchos valores nulos y no son necesarias

Preguntas e Hipótesis a Responder

01

- ¿Cuál es la distribución de calificaciones de higiene en los establecimientos?
- Hipótesis: La mayoría de los establecimientos tienen una calificación alta de higiene.

02

- ¿Qué tipo de establecimientos tienden a recibir las mejores y peores calificaciones de higiene?
- Hipótesis: Restaurantes y tiendas minoristas tienen una mayor variabilidad en sus calificaciones de higiene.

03

- ¿Cómo han evolucionado las inspecciones de higiene en los últimos años?
- Hipótesis: El número de inspecciones ha aumentado debido a regulaciones más estrictas.

04

- ¿Existe una relación entre la calificación de higiene y la confianza en la gestión del establecimiento?
- Hipótesis: Establecimientos con alta confianza en la gestión suelen obtener mejores calificaciones de higiene.

Visualizaciones ejecutivas que responden nuestras preguntas 1 y 2.

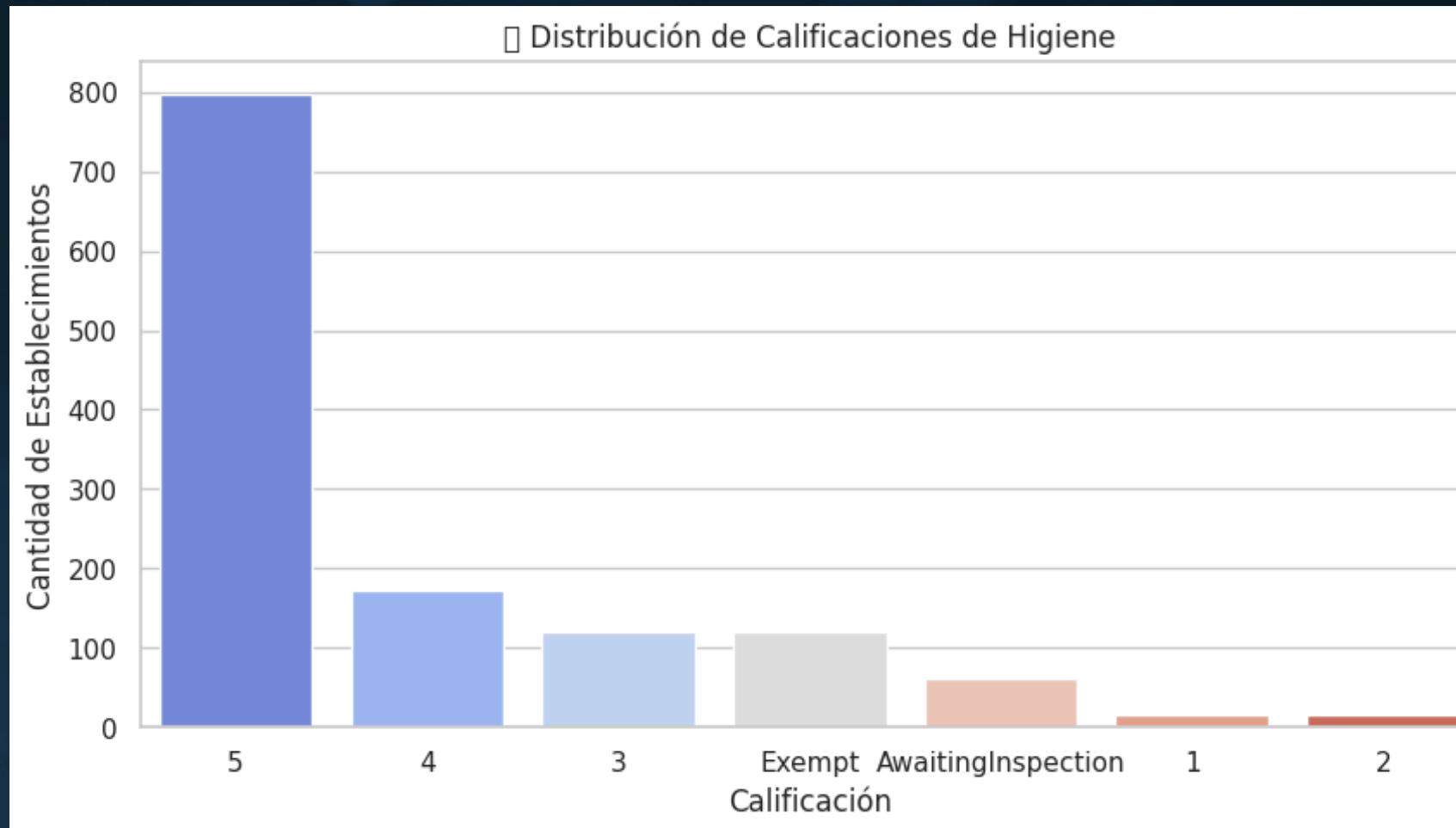


Gráfico de barras: Representa la frecuencia de cada calificación.

Hallazgo: La mayoría de los establecimientos tienen una calificación de 5.

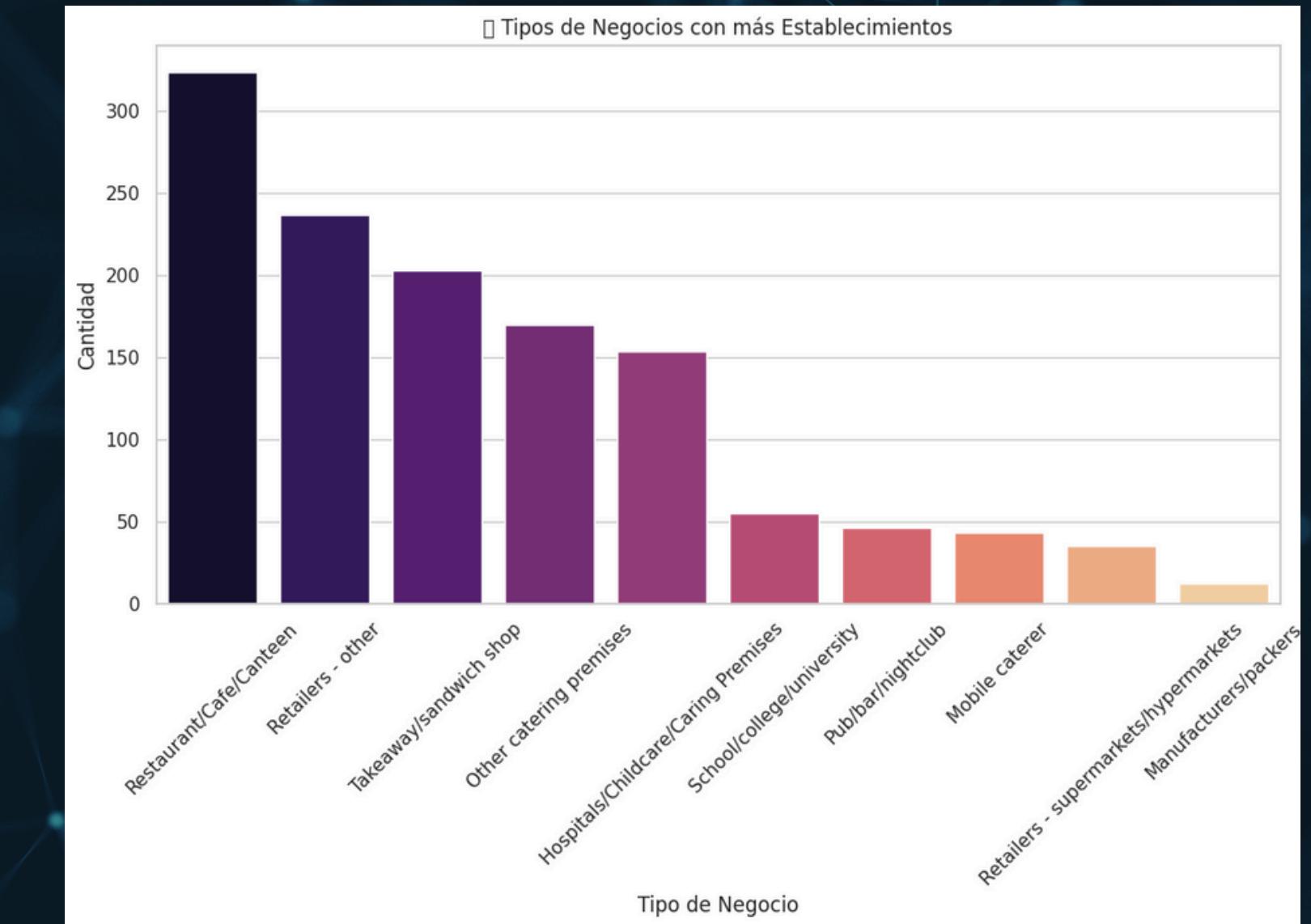


Gráfico de barras: Se observan los tipos de negocio más comunes en la base de datos.

Hallazgo: Los restaurantes y minoristas dominan la lista.

Visualizaciones ejecutivas que responden nuestras preguntas 3 y 4.



Hallazgo: Histograma con KDE: Muestra la frecuencia de inspecciones a lo largo del tiempo.

Hallazgo: Aumento en la cantidad de inspecciones en los últimos años.

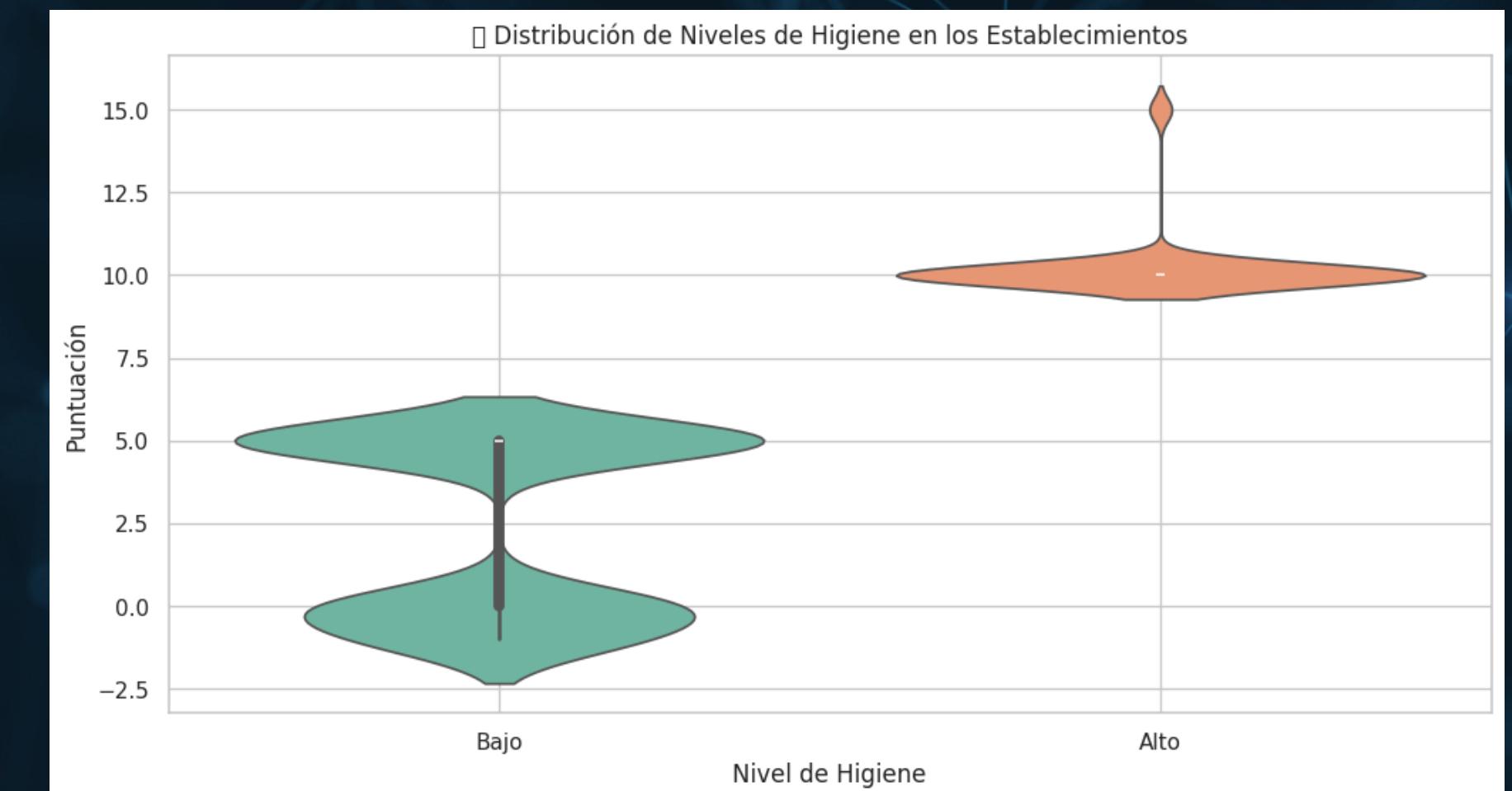


Gráfico de dispersión: Compara la puntuación de higiene y la confianza en la gestión.

Hallazgo: Existe una relación positiva entre ambas variables.

Análisis Univariado

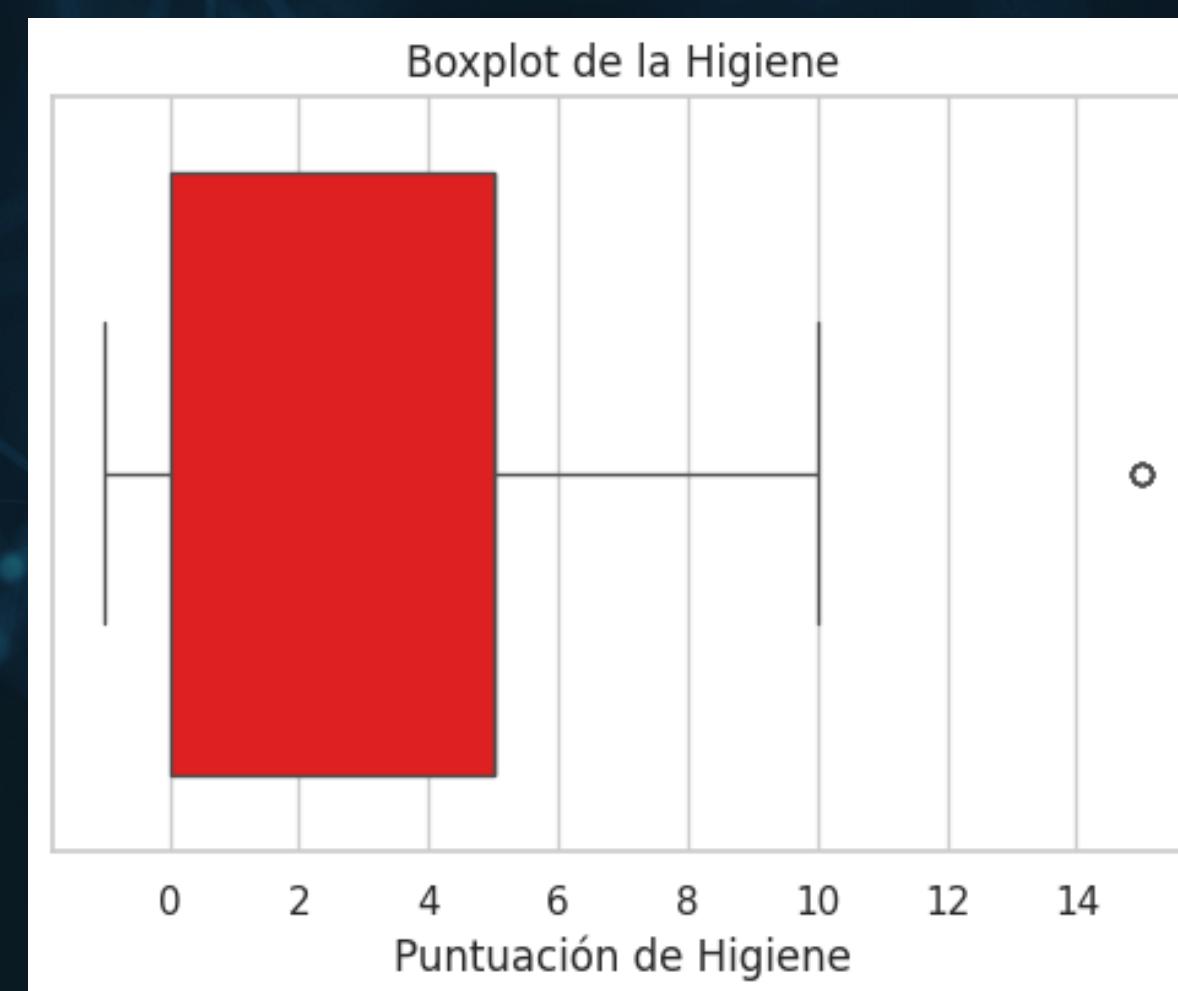
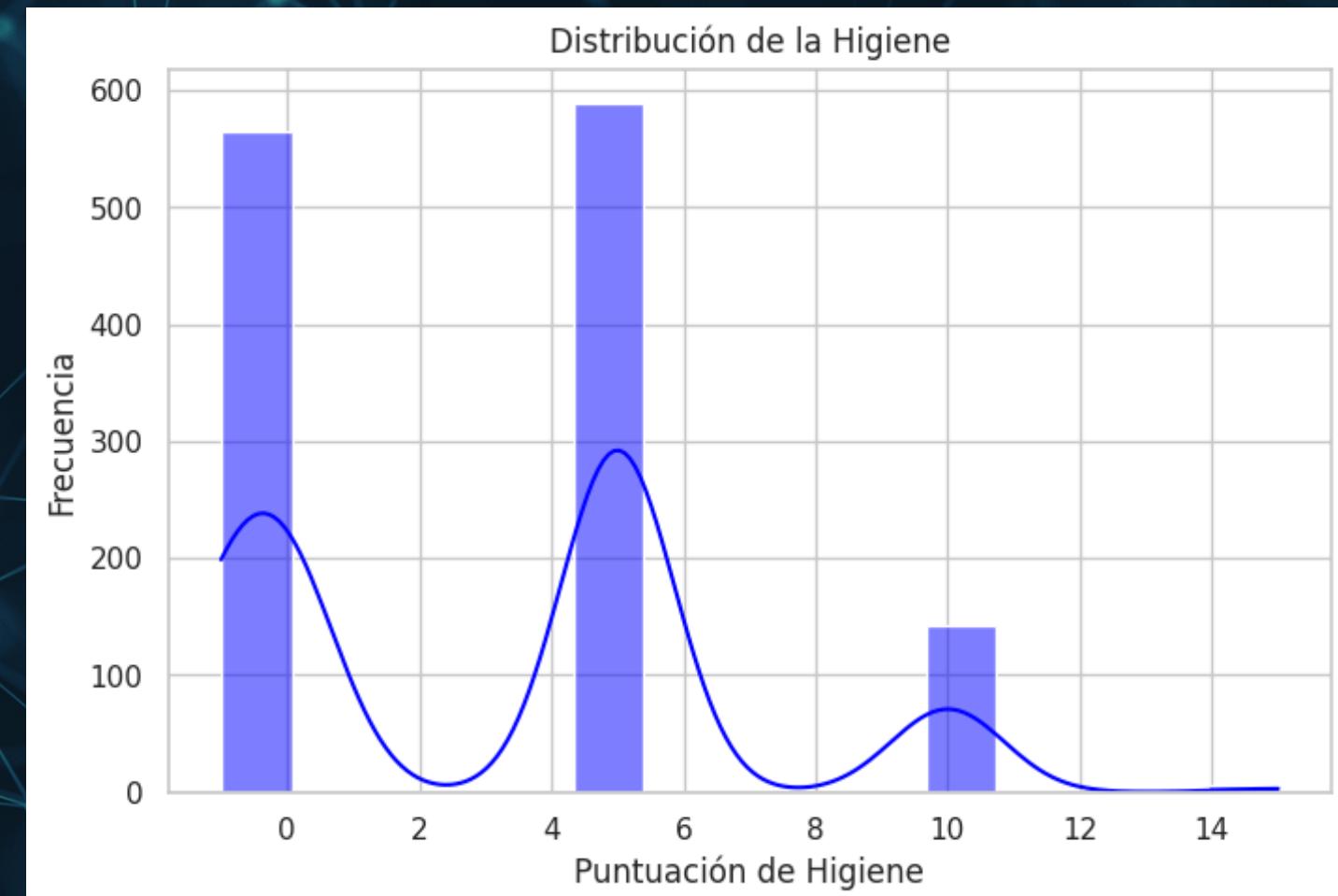
Estadísticas destacadas:
Media: 3.25 | Mediana: 5 | Moda: 5
Desviación estándar: 3.60
Asimetría: 0.46 → Hay valores altos extremos.
Curtosis: -0.57 → Distribución más plana que la normal.

Principales observaciones:

La mayoría de los establecimientos tienen buena higiene (modo y mediana en 5).

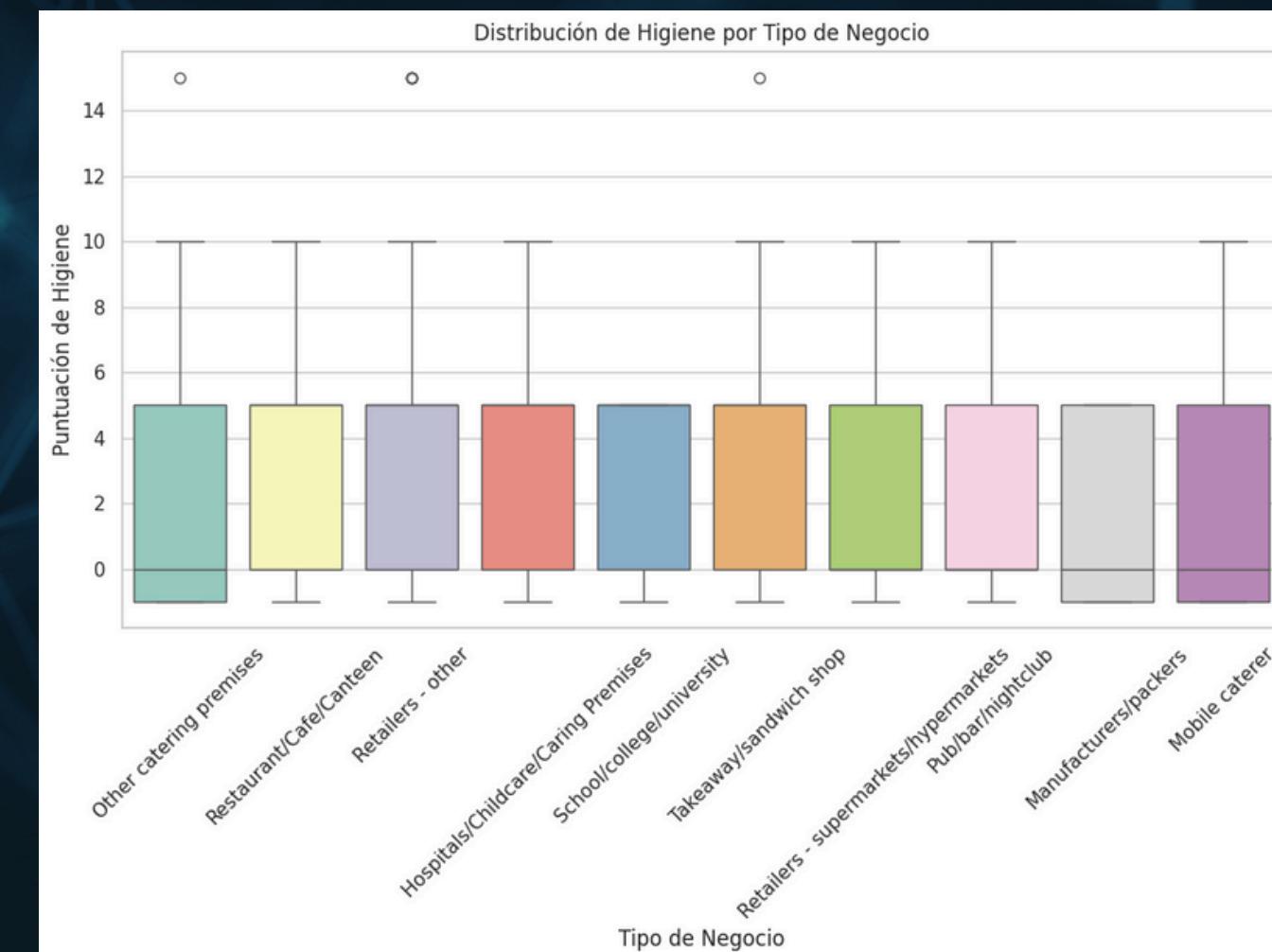
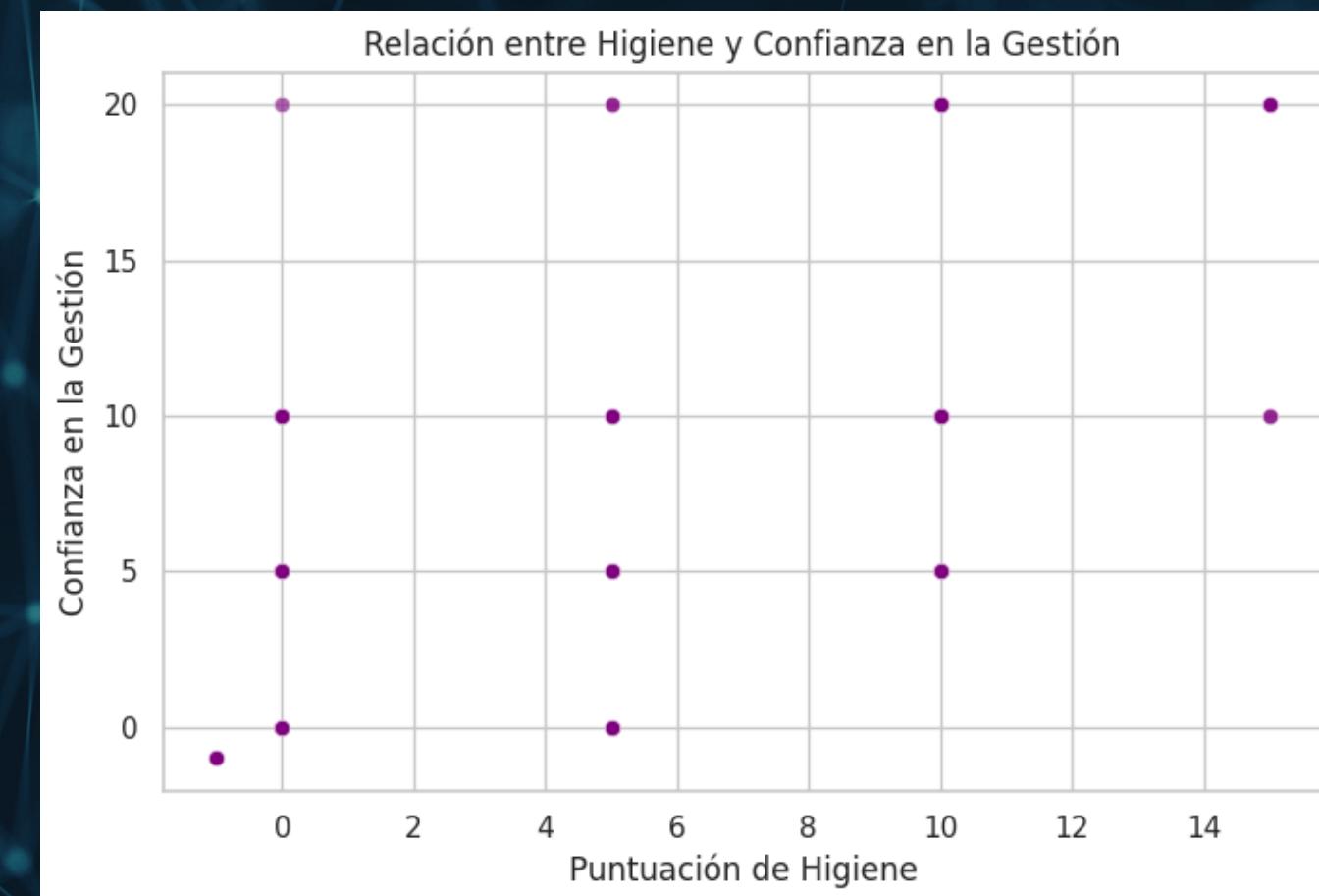
Hay una leve inclinación hacia valores altos, pero también se detectan puntuaciones bajas.

El boxplot revela outliers con valores extremos, posiblemente por errores o casos críticos.



Análisis Bivariado

- ◆ Relación entre Higiene y Confianza en la Gestión:
El gráfico de dispersión muestra una tendencia positiva: los establecimientos con mayor confianza en la gestión tienden a tener mejores puntajes de higiene.
- ◆ Diferencias según Tipo de Negocio:
El boxplot indica que rubros como restaurantes y supermercados presentan más variabilidad en higiene, mientras que hospitales y escuelas suelen mantener estándares más consistentes.



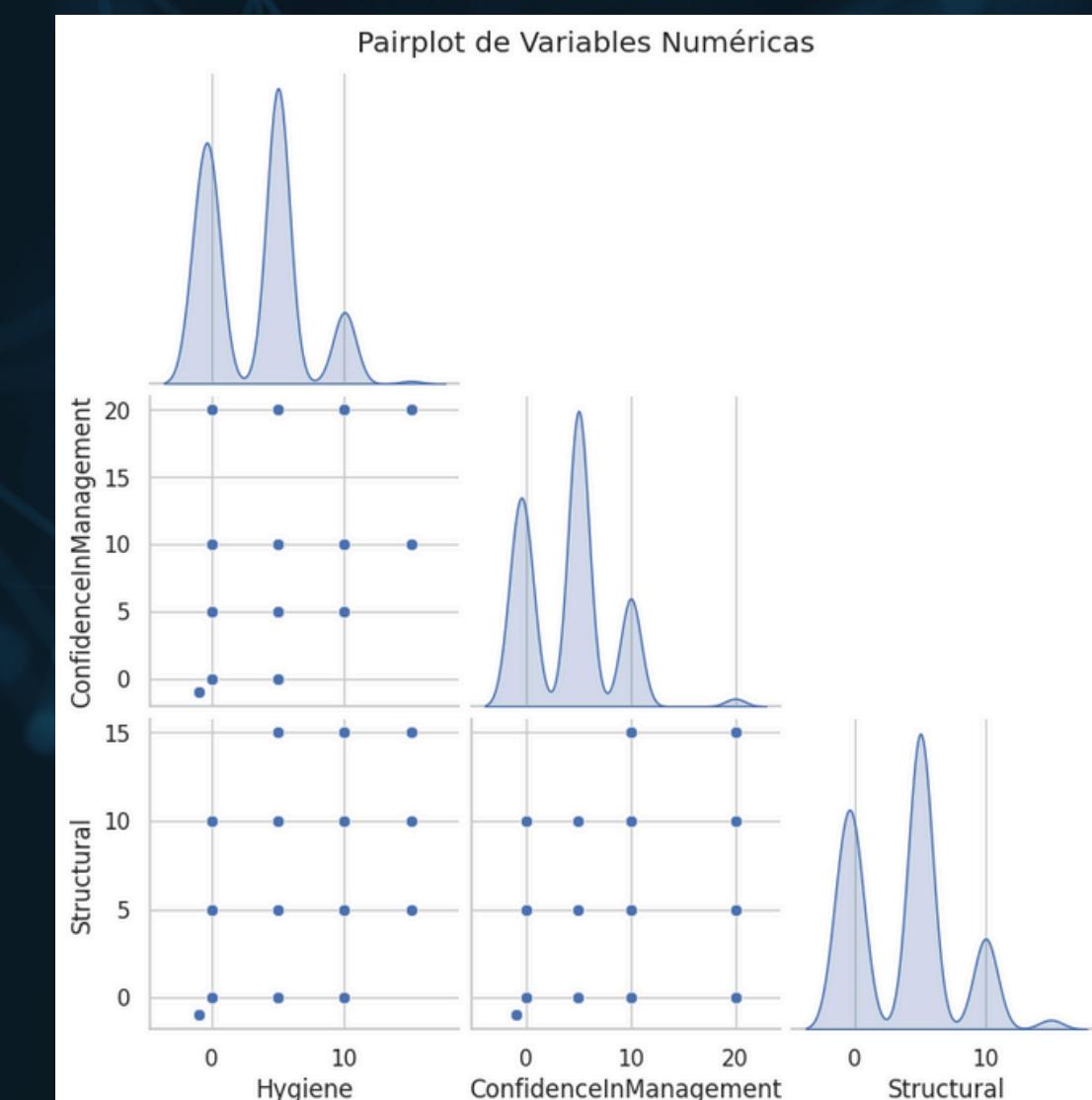
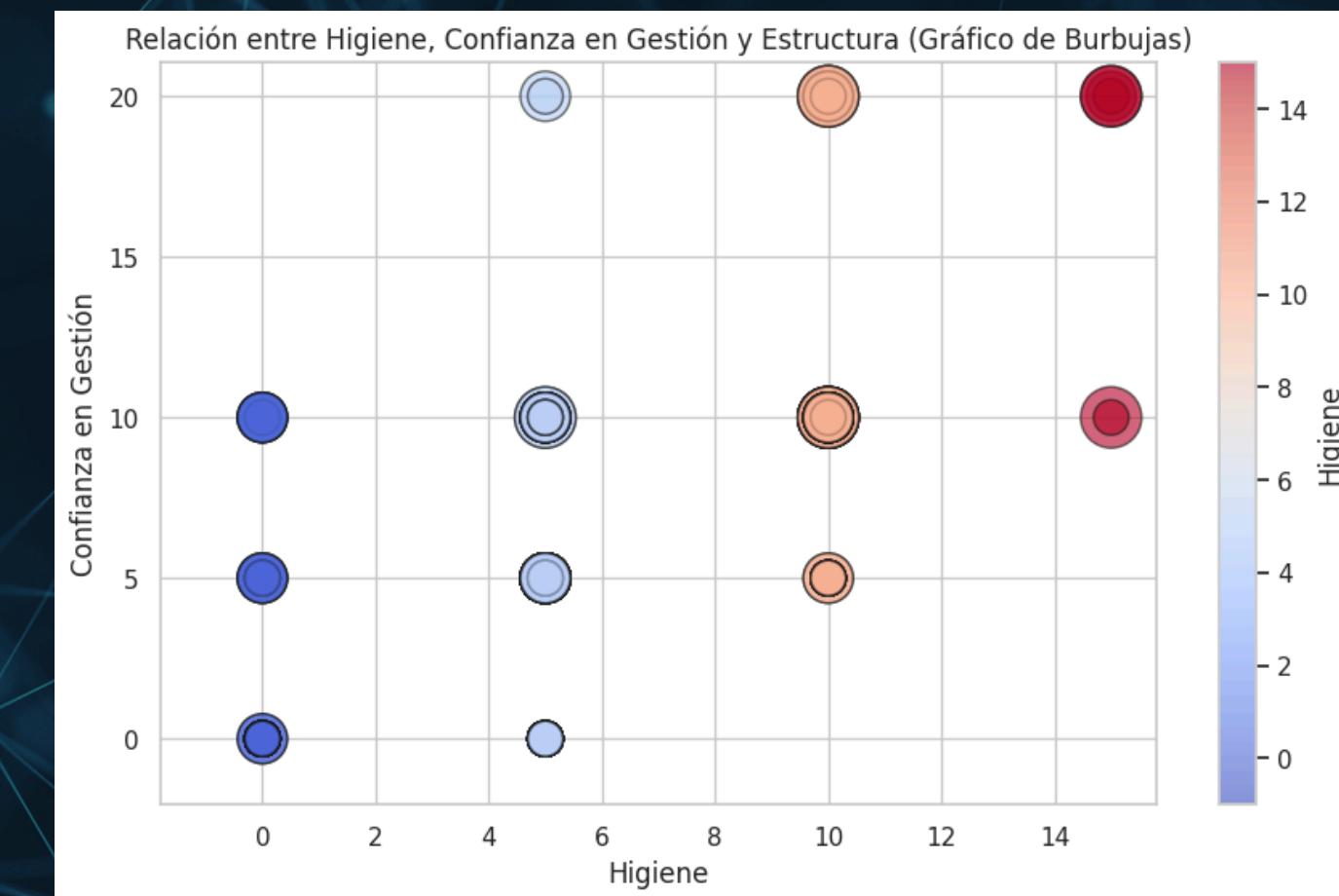
Análisis Multivariado

Gráfico de Burbujas:

Se observa una relación general entre mayor higiene y mayor confianza en la gestión, aunque con excepciones. Además, la variable estructural influye, ya que el tamaño de la burbuja (estructura) parece estar relacionado con mejores prácticas.

Pairplot:

Las variables muestran distribuciones multimodales, lo que indica que las puntuaciones se asignan en categorías discretas. No hay correlaciones fuertes entre las variables, pero sí patrones agrupados por niveles específicos.



Insights de las preguntas e hipótesis

01

Mayoría con alta higiene: La calificación más común es 5, lo que indica que la mayoría de los establecimientos cumplen con altos estándares de higiene.

02

Restaurantes y minoristas predominan: Son los tipos de negocios más evaluados, lo que resalta su impacto en la seguridad alimentaria.

03

Aumento en inspecciones recientes: Posiblemente debido a normativas más estrictas o una mayor supervisión gubernamental.

04

Relación entre higiene y gestión: Establecimientos con mayor confianza en la gestión suelen tener mejores calificaciones de higiene.

Insights de los análisis univariado, multivariado y bivariado

01 Insights del Análisis Univariado

- La mayoría de los establecimientos tienen buena higiene (modo = 5).
- Hay algunos con muy baja higiene (sesgo a la derecha).
- Se detectaron outliers con valores extremos.

02 Insights del Análisis Bivariado

- Mayor confianza en la gestión se asocia a mejor higiene.
- Restaurantes y tiendas muestran mayor variabilidad en higiene.
- Algunos tipos de negocios presentan diferencias marcadas en las calificaciones.

03 Insights del Análisis Multivariado

- Las variables clave (higiene, estructura, gestión) se distribuyen en niveles discretos.
- No hay correlaciones lineales fuertes entre las variables.
- Hay negocios con baja higiene pero alta confianza, lo que sugiere otros factores influyentes.