

# Prueba FixPartnersConsulting

Postulante: Matias Arriola

## 1. Parte 1

Para esto se realizaría un modelo de estimación de demanda Mixed Logit. De esta manera, se asume que la utilidad del individuo  $i = 1, \dots, N$  por el producto  $j$  viene dada por

$$U_{ij} = \alpha_i p_j + \lambda_i X_j + \epsilon_{ij} \quad (1)$$

donde  $p_j$  corresponde al precio del producto  $j$ , y  $X_j$  otras características del producto  $j$ , y que usualmente se interpreta como que tan apetecido es ese producto. Por su parte  $\epsilon_{ij}$  es un shock aleatorio type I extreme value (Distribución Gumbel  $\mu = 0$  y  $\sigma = \frac{\pi}{\sqrt{6}}$ ). De esta manera, para estimar este modelo, se debería solicitar a la tienda tanto los precios de los distintos productos como las cantidades de estos. Las cantidades se utilizan para estimar los parámetros  $X_j$ . Luego, para estimar los parámetros  $\alpha_i$  y  $\lambda_i$  se utilizarían dos formas, con el fin de contrastar sus resultados:

- Asumir que todos los clientes son homogéneos  $\alpha_i = \alpha$ ,  $\lambda_i = \lambda$ , y estimar dichos parámetros utilizando máxima verosimilitud.
- Utilizar el modelo propuesto por Fox (2016), y asumir una grilla de parámetros tanto para  $\alpha_i$  y  $\lambda_i$ . Luego se debe muestrear  $R$  veces de dicha grilla (ver Fox 2016) y estimar el parámetro  $\hat{\theta}$ .

$$\hat{\theta} = \frac{1}{NJ} \sum_{i=1}^N \sum_{j=1}^J (y_{ij} - \sum_{r=1}^R \theta_r S_{ij}^r)^2 \quad (2)$$

donde  $Y_{ij}$  corresponde a una matriz con valores de unos y ceros, poseyendo para cada fila un 1 en la posición del producto comprada por el individuo  $i$ . y cero en las columnas restantes de esa fila. Por su parte  $S_{ij}^r$  corresponde al market share de la muestra  $(\alpha_r, \lambda_r)$ . Finalmente una vez estimados los pesos  $\hat{\theta}$  (de dimensiones  $1 \times R$ ) se multiplica cada pesos  $\theta_r$  por su respectivo  $(\alpha_r, \lambda_r)$  obteniendo así la distribución estimada final de los  $(\alpha_i, \lambda_i)$ .

Una vez estimados los parámetros de la ecuación (1), solo basta reemplazar en dicha ecuación para obtener la demanda. El modelo se puede complejizar aún más si se requiera, como por ejemplo agregando otro índice más a la ecuación 1, que de cuenta de la demanda específica para cada tienda que pueda tener el cliente.

## 2. Parte 2

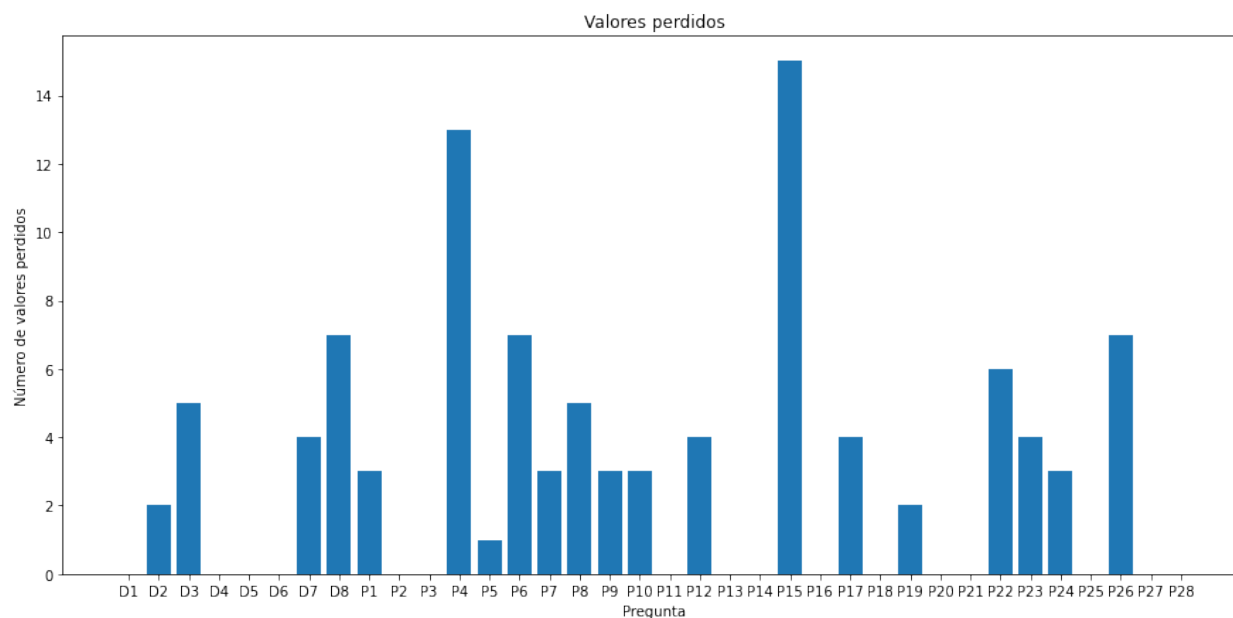
Las respuestas están en archivo adjunto SQLprueba.txt.

## 3. Parte 3

### 3.1.

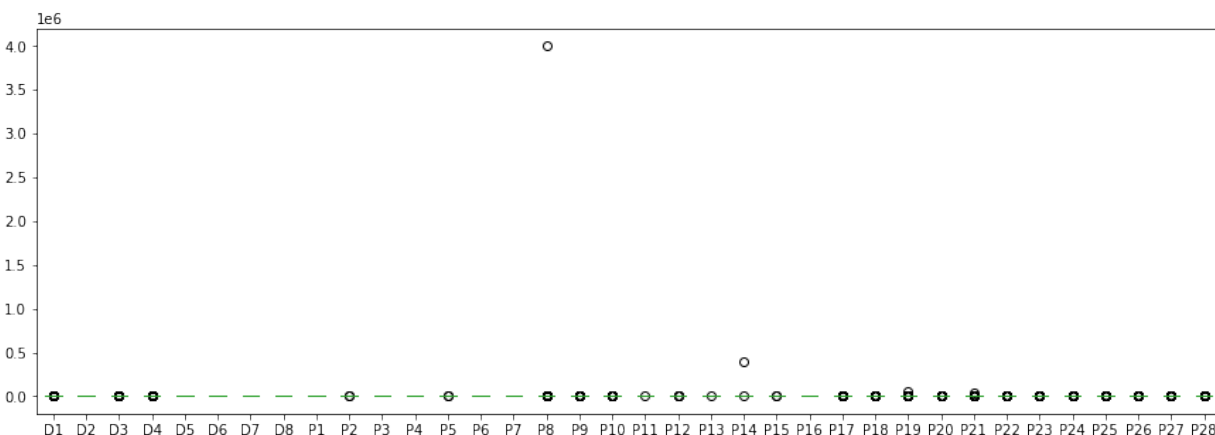
Los datos están compuestos por dos dataframes. Uno posee las preguntas realizadas a los clientes y sus respectivos ids, mientras que el otro posee las respuestas a cada una de las preguntas. Se puede comprobar que las respuestas corresponden a variables numéricas y categóricas que indican un índice de satisfacción en los distintos ámbitos evaluados. Además, observando la distribución de los datos (figura 8) se tiene que las respuestas están concentradas en los valores que poseía la encuesta, no obstante, existen outliers y valores que están fuera del rango dado en la encuesta.

A su vez, observando los datos, se observan valores perdidos así como valores con letras, que debían ser valores numéricos. En la figura 1, se observan los valores perdidos por pregunta.



**Figura 1:** Valores perdidos.

A su vez, obteniendo estadísticos básicos sobre los resultados de las preguntas, se aprecia la presencia de outliers, esto se observa la siguiente figura 2



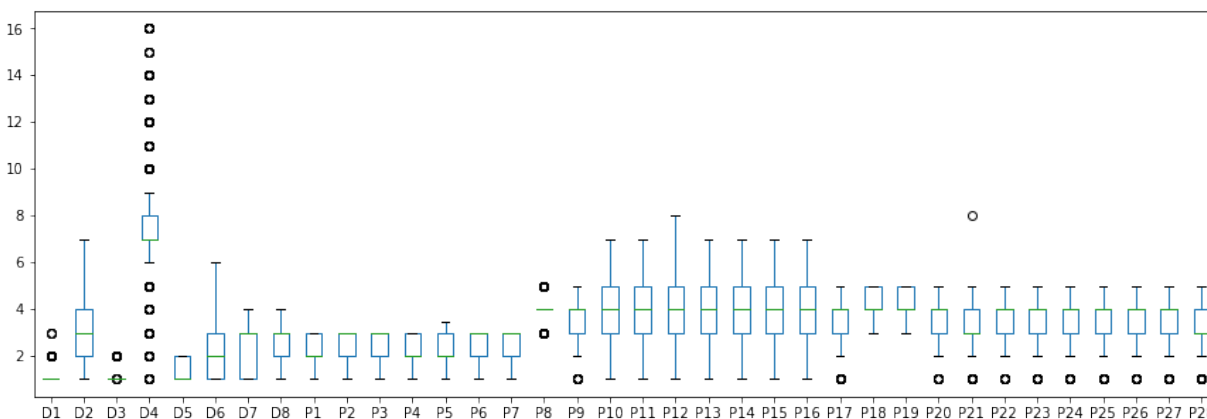
**Figura 2:** Boxplot de las distintas preguntas, que da cuenta de la presencia de outliers, y de la varianza presente en las respuestas dadas.

De lo anterior se realiza los siguientes tratamientos a los datos:

- Valores perdidos, así como letras son imputadas por la media obtenida en la pregunta respectiva.
- Outliers son eliminados, seleccionando (de manera arbitraria) un cuartil 25, hasta un percentil 95, tomando como referencia las respuestas en las preguntas respectivas.

Realizando lo ya mencionado, se obtiene el siguiente boxplot sobre las respuestas dadas a la distintas preguntas (figura 3).

Finalmente, para analizar la relación entre las distintas variables, se obtiene la matriz de correlaciones (presente



**Figura 3:** Boxplot de preguntas luego de eliminar outliers y valores perdidos.

en código adjunto) y matriz de correlación entre las respuestas de las preguntas (figura 9). Si se observa la figura 9 se tiene que las variables más correlacionadas son justamente las que se quiere predecir, de esta manera se tiene que aquellos clientes que clasificaron con una nota positiva la tienda, es más probable que recomienden esta tienda a otras personas. No obstante, desde el punto de vista de una regresión, incluir variables con una alta correlación para predecir otra no es adecuado. Por lo que para cada una de las dos regresiones para cada variable, se eliminará la otra que se pretende analizar. También, se eliminan de lo anterior, debido a sus características, las respuestas que caracterizan al sujeto, donde se realiza un análisis aparte para el Género y edad de la persona encuestada.

### 3.2.

**A)** Como se mencionó previamente, se eliminan las preguntas tipo D y la otra variable a regresionar. Esto para cada una de las variables analizadas.

Realizada la regresión para la primera variable, se obtienen los siguientes coeficientes, que se interpretan como un porcentaje de importancia para dicha variable (figura 4)

Luego, para responder cómo varía la satisfacción del cliente de acuerdo a la edad y género, se obtienen tablas cruzadas entre ambas variables (código adjunto), así como un gráfico tipo violin (figura 5).

A su vez, para la segunda variable de si recomendarían la tienda, de la regresión se obtienen los pesos presenten en la figura 6

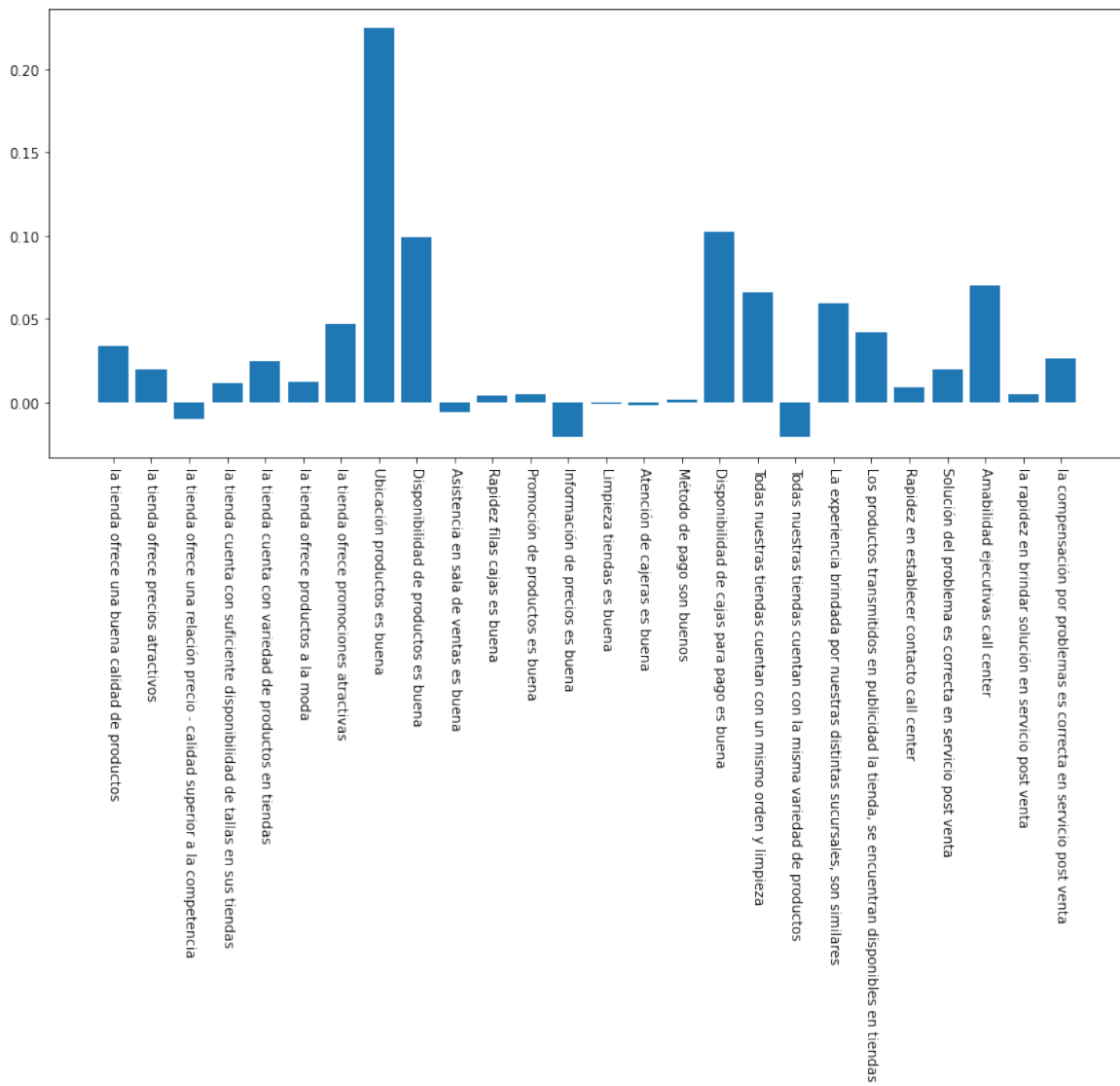
Para analizar cómo varía variables de si recomendarían la tienda, en relación a la edad y género, se obtuvo la figura 7.

**B)** De los puntos anteriores, se tiene que para la variable de satisfacción, se obtuvo que aquellas variables más relevantes fueron la variabilidad como la posición de los productos, así como también la disponibilidad de cajas, un orden y similitud entre tiendas y una buena atención postventa.

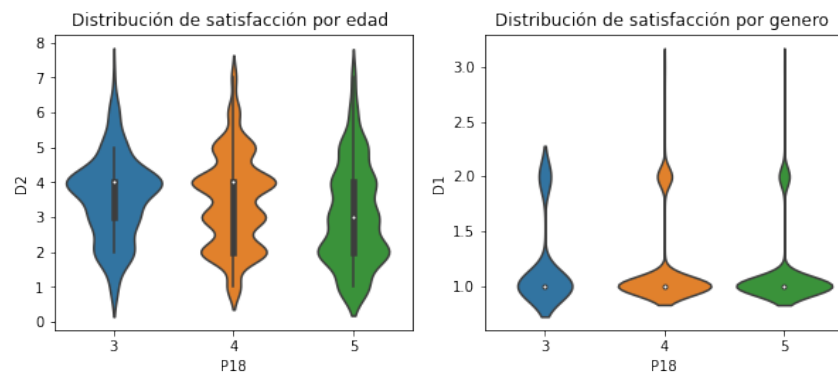
En cuanto a la variables etaria y de género, se puede observa que para el caso de satisfaccion ( figura 5 y tabla cruzada en código adjunto), se obtiene una distribución etaria de personas de una edad media, clasificando mayoritariamente (un 65 %) la experiencia con nota 4, y el segundo gran porcentaje (un 29 %) de personas con nota 5 (nota máxima). A su vez en cuanto al género del cliente, se tiene que mayoritariamente los clientes son de género tipo 1 (87 %) y en un menor porcentaje tipo 2 (13 %). Al igual que en el caso anterior, la clasificación por género, fue mayoritariamente buena.

A su vez, para la variables de si recomendarían la tienda, se obtuvo que aquellas variables más relevantes fueron la ubicación de los productos, que los productos estuvieran a la moda, disponibilidad de cajas y similitud entre orden y limpieza entre tiendas, así como también una buen servicio post venta.

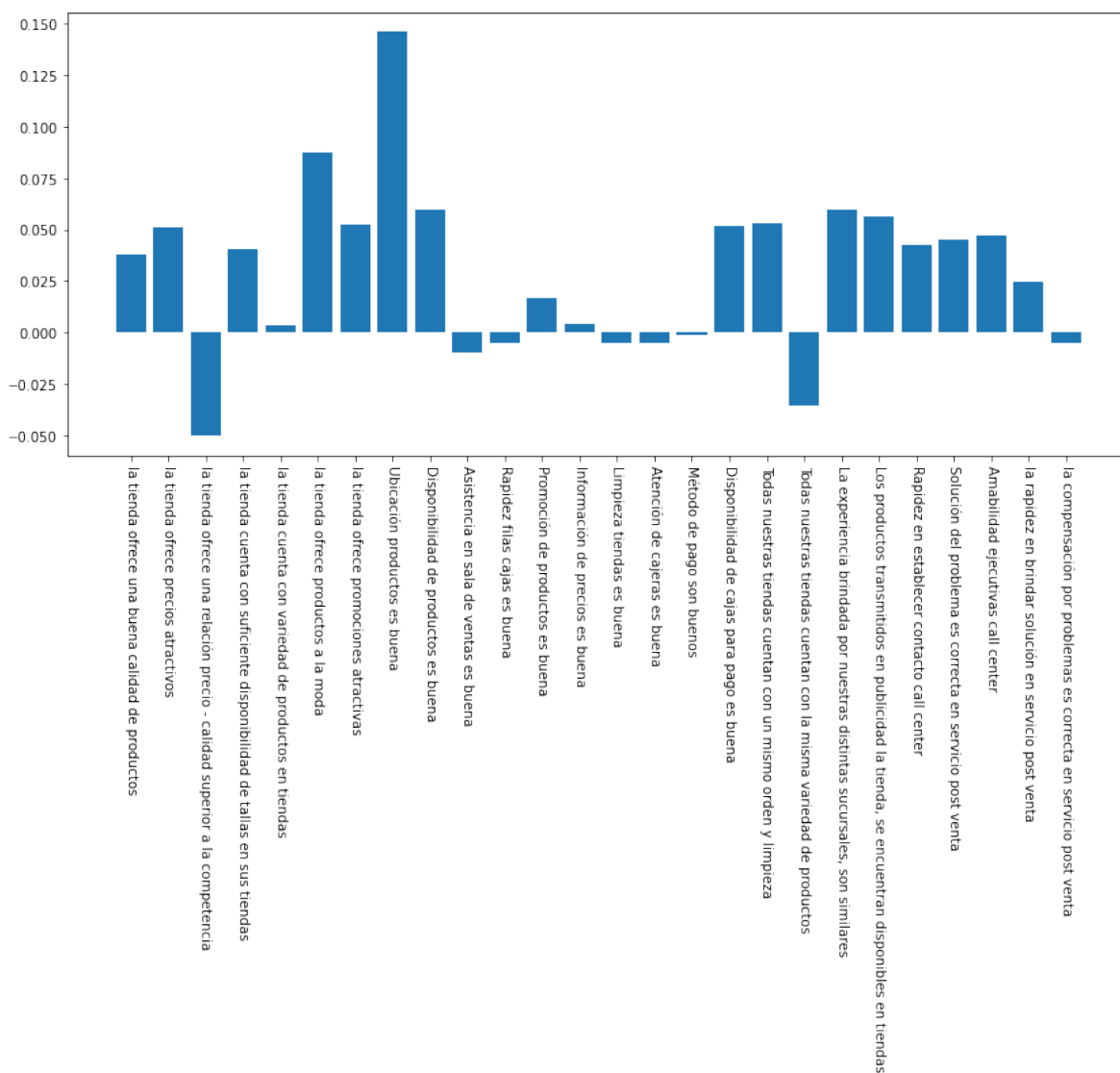
Finalmente de la figura 7 se observa que la distribución etaria nuevamente se concentra en el centro, clasificando mayoritariamente bien con (45 %) con nota 4, y con nota máxima 5 el otro gran porcentaje (47 %).



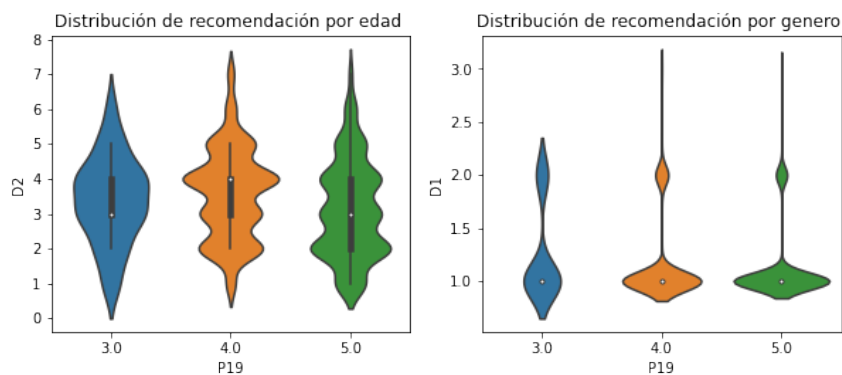
**Figura 4:** Importancia de las distintas preguntas en la regresión para la variables de satisfacción del cliente.



**Figura 5:** Gráfico tipo violín, de la edad, género y la satisfacción del cliente.



**Figura 6:** Importancia de las distintas preguntas obtenidas de la regresión para si una persona recomendaría la tienda.



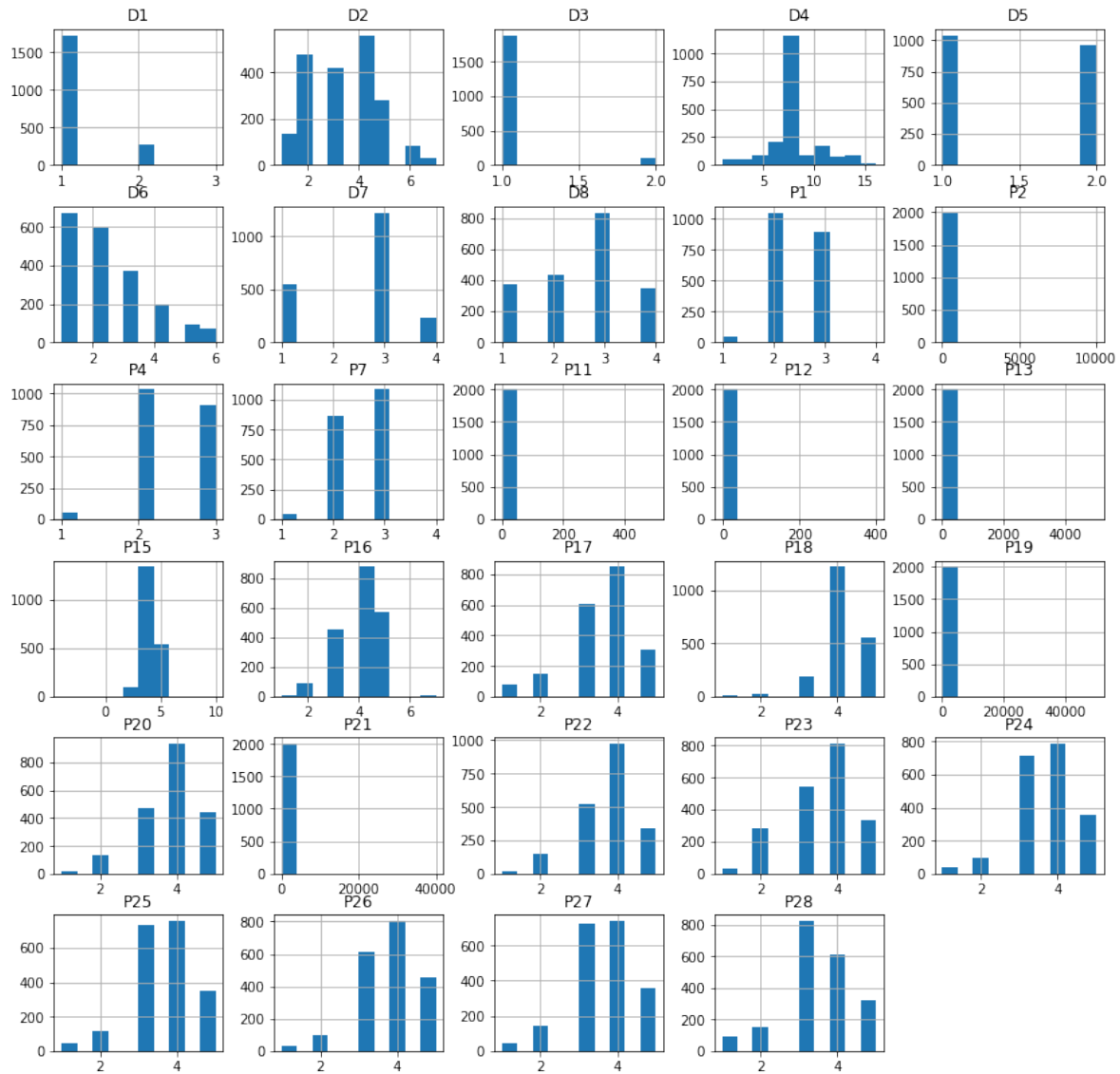
**Figura 7:** Gráfico tipo violín, de la edad, género y si recomendarían o no la tienda.

### 3.3.

De lo anterior, al cliente se recomiendan los siguientes puntos:

- Aunque suene de perogrullo, la tienda debe buscar obtener una buena percepción de satisfacción en sus clientes, dado que haciendo esto es muy probable que recomienden la tienda a alguien más, y viceversa.
- Mantener su stock actualizado (a la moda) y ubicar en buena posición cada uno de ellos.
- Mantener una buena presentación y similitud entre sus tiendas, así como una buena atención mientras el cliente este en la tienda.
- Fortalecer su servicio postventa, entregando una buena atención al cliente.

## 4. Anexo



**Figura 8:** Distribución inicial de los datos.

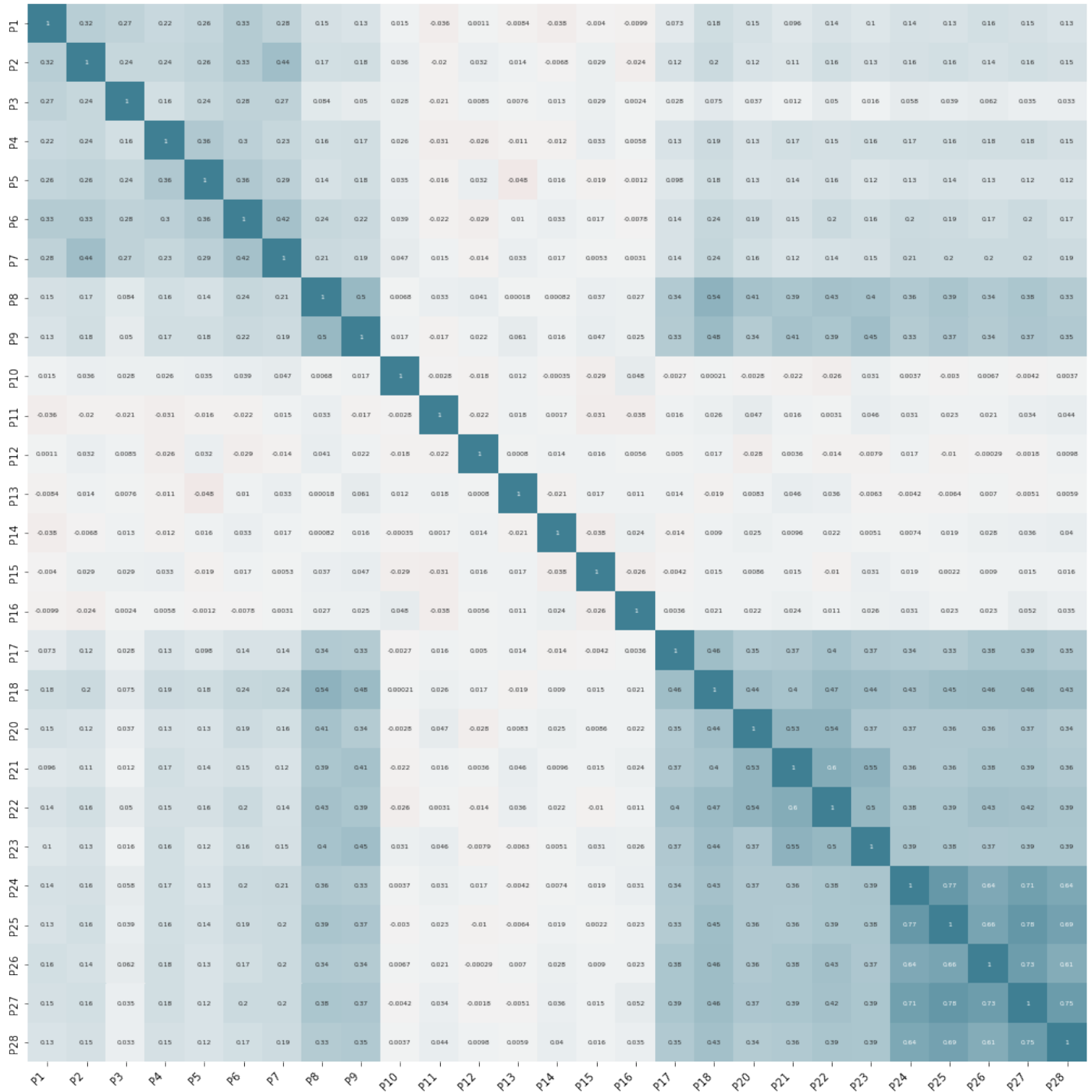


Figura 9: Mátriz de correlación entre preguntas realizadas a clientes