

Examen Final Técnicas Clasificación

Resumen Ejecutivo

El objetivo de este informe es hacer un análisis sobre la encuesta de valoración del turismo e intentar agrupar a los encuestados en una serie de grupos en base a sus respuestas en dicha encuesta.

Para ello se ha procedido a extraer dos muestras aleatorias de 150 observaciones cada una para después poder compararlas y obtener resultados.

Así mismo se ha procedido a agrupar aquellas variables que presentaban una correlación alta, con lo cual evitamos futuros problemas en el análisis clúster.

Los valores *NA* se han eliminado definitivamente y así, nos hemos quedado con: * Impresión * Alojamiento general: Que agrupa el alojamiento, el trato y la gastronomía * Entorno general: Que agrupa el clima, las zonas de baño, el medio ambiente, los paisajes, limpieza y tranquilidad. * Restaurante general: Que contiene las variables calidad, oferta, trato y precio

La variable ingresos se ha imputado la media al ser una variable que llevaba rangos y además nos hemos quedado con la edad y hemos quitado aquellas que no nos aportan información como Aeropuerto o ID.

Desarrollo

Antes de la realización del análisis cluster se suelen realizar otro tipo de pruebas para comprobar que efectivamente el uso del análisis posterior es correcto. Nos estamos refiriendo al estadístico de Hopkins. El objetivo de esta prueba es comprobar si los datos pueden o no estructurarse en distintos grupos que es lo que posteriormente realizaremos. El resultado óptimo es un resultado cercano a cero.

Dado que nuestro resultado es de 0.46 podemos proceder con el análisis clúster. En este caso hemos seleccionado el método K-Means que es el que mejor se adecua a nuestros datos. En el caso de CLARA, no podría utilizarse ya que nuestra muestra no presenta los suficientes elementos y el PAM que, aunque no arroje malos resultados, está especialmente diseñado cuando la muestra contiene Outliers.

Aunque para ambos conjuntos de datos la mejor decisión sea la elección de dos clústeres, se puede apreciar que en el primer conjunto las agrupaciones son más perfectas, explicando ambas dimensiones un 55% del total. Mientras que en el segundo caso se produce solapamiento y la explicación de las dimensiones es relativamente menor, 52%.

Con esto se puede concluir que las tendencias durante estos periodos no ha cambiado mucho.

Para el primer conjunto de datos, podemos observar que la distribución de la nacionalidad, isla de mayor estancia, edad, sexo e ingresos está, en general, agrupada mejor en el primer clúster, teniendo el segundo menos observaciones de estas variables previamente especificadas.

Cabe prestar especial atención a la isla de mayor estancia, ya que Lanzarote tiene bastante diferencia con respecto a las otras islas. En cuanto a ingresos vemos que las rentas medias-altas son bastante más escasas las observaciones.

En cuanto a la distribución del segundo conjunto de datos: Llama la atención que las personas que poseen más ingresos y tienen un rango de edad menor se incluyen dentro del segundo grupo. Los valores generales de restauración, alojamiento y entorno no son relevantes para incluirlos en un cluster u otro.

Conclusiones

En el año 2010-14

Basándonos en las estadísticas por conjunto de datos y agrupación en cluster, podemos concluir que en general la mayoría de observaciones pertenecen al primer clúster cuya media de edad es más alta que en el

segundo. El perfil de ingresos está también más elevado. En la agrupación dos en general el promedio de edad es más reducido y por ello los ingresos también. La idea de turismo de este conjunto de personas es diferente a la del primero.

En el año 2014-17 El turismo puede haberse vuelto más conservador, habiéndose reducido la diferencias en ingresos y rangos de edad.

En general y a modo de resumen, sí existen diferencias en quienes hacen la valoración puesto que el nivel de ingresos y el rango de edad determina las condiciones en las que se viaja.