

Análisis Estadístico: Informe

En el siguiente informe se evaluará de manera descriptiva y grafica las diferentes variables del dataset “theatre.csv” usando como herramienta el software RStudio, esto con el fin de hallar las características de relevancia para el negocio (teatro Schauspielhaus Zürich) y además plantear posibles poblaciones objetivo para su campaña, adicional, se encontrará al final de este informe una pequeña fundamentación teórica

Variables

Se encontraron seis variables que fueron recolectadas por el teatro de los clientes con tarjeta cliente frecuente:

- Age: variable cuantitativa, contiene personas desde los 18 hasta los 64 años
- Sex: variable cualitativa, en donde 0 representa hombres y 1 mujeres, hay 309 y 390 respectivamente
- Income: en Francos Suizos representa una variable cuantitativa, contiene valores desde 33 hasta 122.4 (en miles), es el ingreso anual de los encuestados
- Culture: en Francos Suizos, representa una variable cuantitativa del gasto mensual de los encuestados en cultura, contiene valores desde 64 hasta 618
- Theatre: representa una variable cuantitativa del gasto mensual en el teatro (Schauspielhaus Zürich) en Francos Suizos
- Theatre_ly: representa una variable cuantitativa de los gastos mensuales estimados en el teatro, del año anterior, está en Francos Suizos

Nota: Para mayor precisión en los calculos la variable Income fue multiplicada por 1000

X1 <dbl>	Age <dbl>	Sex <dbl>	Income <dbl>	Culture <dbl>	Theatre <dbl>	Theatre_ly <dbl>
1	31	1	90.5	181	104	150
2	54	0	73.0	234	116	140
3	56	1	74.3	289	276	125
4	36	1	73.6	185	75	130
5	24	1	109.0	191	172	140
6	25	0	93.1	273	168	130
7	61	1	63.9	184	119	195
8	50	0	46.1	155	97	110
9	53	1	75.0	253	152	155
10	52	0	68.7	291	166	150

1-10 of 699 rows

Previous 1 2 3 4 5 6 ... 70 Next

Tabla 1. Muestra de los primeros 10 datos

Para efectos prácticos en el informe solo se incluirán solo resultados relevantes para el estudio y el objetivo de este, en el siguiente repositorio, https://github.com/damesaa201710054010/Analisis_Estadistica.git se encuentra el notebook de R y los scrips respectivos con los cuales se realizó el estudio.

Gasto mensual en cultura

Se observa que la mayoría de los clientes gastan frecuentes del teatro gastan entre 50 y 300 francos mensuales en actividades de tipo cultura, adicional, la mayoría de ellos gasta mas de

125 francos mensuales, podemos realizar un análisis mas riguroso observando la grafica 1 y 2 que contienen los datos de as frecuencias absolutas y absolutas acumuladas que fueron extraídas de la tabla 2

Class limits	f	rf	rf(%)	cf	cf(%)
[63.36,114.344)	5	0.01	0.72	5	0.72
[114.344,165.327)	77	0.11	11.02	82	11.73
[165.327,216.311)	275	0.39	39.34	357	51.07
[216.311,267.295)	236	0.34	33.76	593	84.84
[267.295,318.278)	87	0.12	12.45	680	97.28
[318.278,369.262)	11	0.02	1.57	691	98.86
[369.262,420.245)	5	0.01	0.72	696	99.57
[420.245,471.229)	2	0.00	0.29	698	99.86
[471.229,522.213)	0	0.00	0.00	698	99.86
[522.213,573.196)	0	0.00	0.00	698	99.86
[573.196,624.18)	1	0.00	0.14	699	100.00

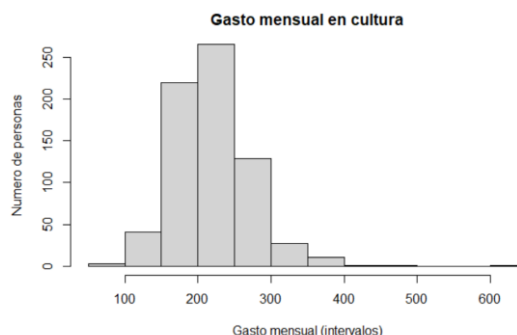
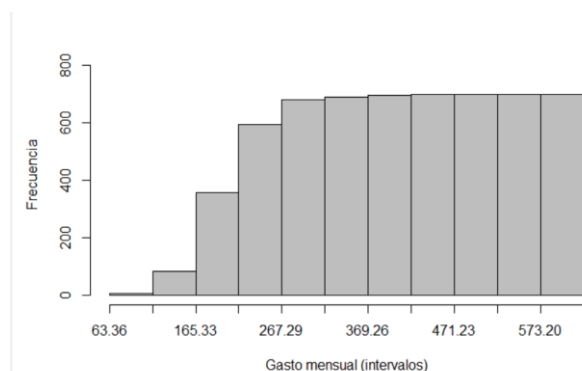
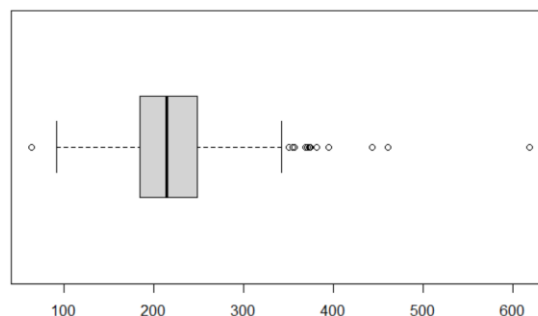


Tabla 2. Tabla de frecuencias para la variable Culture

Grafica 1. Histograma de frecuencia absoluta por intervalos



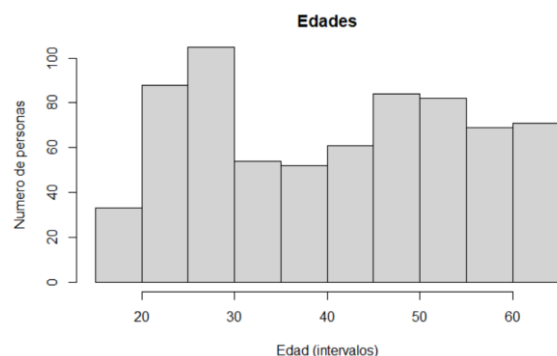
Grafica 2. Histograma de f acumulada por intervalos



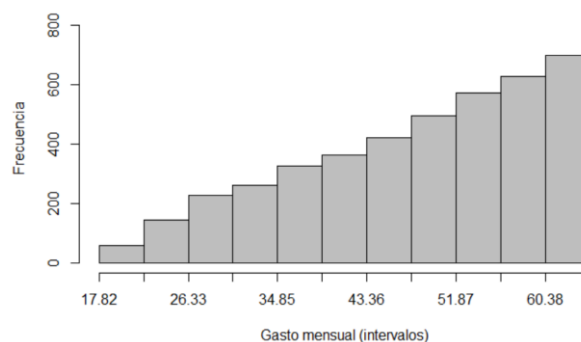
Grafica 3. Boxplot frecuencia absoluta

Tomando en cuenta algunas mediciones como la mediana, observamos que el 50% de la muestra gasta por lo menos mas de 215 francos mensuales en cultura y en promedio unos 219.9, en la grafica 3 vemos la presencia de algunos datos que pueden considerarse atípicos, pero analizando las frecuencias en la tabla 1 vemos que hay una proporción de la muestra que gasta bastante en cultura

Edades



Grafica 4. Hist Frecu Absoluta



Grafica 5. Hit Frecu Abso Acumulada

Aunque en el grafica 4 se observa mayor frecuencia absoluta para personas entre 18 y 30 años en el grafico 5 podemos ver que las personas mayores de 40 años predominan en los clientes de la muestra para el teatro, si recurrimos al promedio, por ejemplo, vemos que la edad es 41.22 esto nos da una idea de cual es mas o menos la edad que suele asistir al teatro

Ingresos anuales

La mayoría de los clientes según el grafico 6 ganan mas de 82281.27 además podemos concluir que el 50% gana mas de 70000 y el otro menos de esto. Es también interesante ver el boxplot (grafico 7) nos indica que, aunque la mayoría e encuentra en un intervalo cercano a la media, hay unos pocos que ganan mucho y otros que ganan muy poco, estando muy por encima o por debajo de la media (71679) y la mediana (70000)

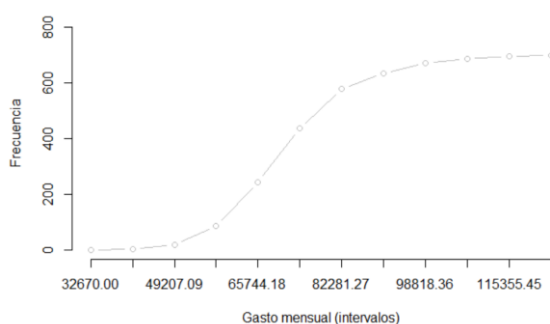


Gráfico 6. Frecu Acumula, polígono

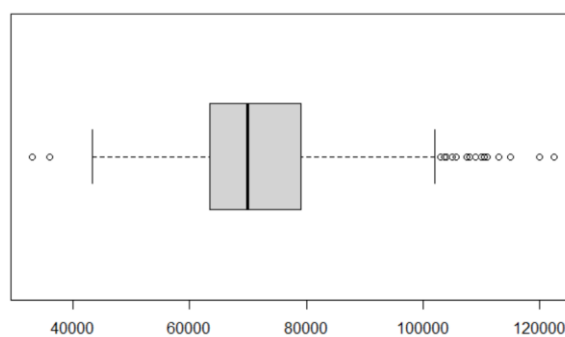


Gráfico 7. Frecuenc absoluta boxplot

Gastos mensuales en el teatro el año actual y el anterior

Vamos a analizar las frecuencias de los gastos mensuales de los clientes en el año actual y el anterior (tablas de frecuencia 3 y 4)

Class limits	f	rf	rf(%)	cf	cf(%)
[22.77,63.2118)	69	0.10	9.87	69	9.87
[63.2118,103.654)	212	0.30	30.33	281	40.20
[103.654,144.095)	168	0.24	24.03	449	64.23
[144.095,184.537)	93	0.13	13.30	542	77.54
[184.537,224.979)	60	0.09	8.58	602	86.12
[224.979,265.421)	43	0.06	6.15	645	92.27
[265.421,305.863)	31	0.04	4.43	676	96.71
[305.863,346.305)	7	0.01	1.00	683	97.71
[346.305,386.746)	5	0.01	0.72	688	98.43
[386.746,427.188)	7	0.01	1.00	695	99.43
[427.188,467.63)	4	0.01	0.57	699	100.00

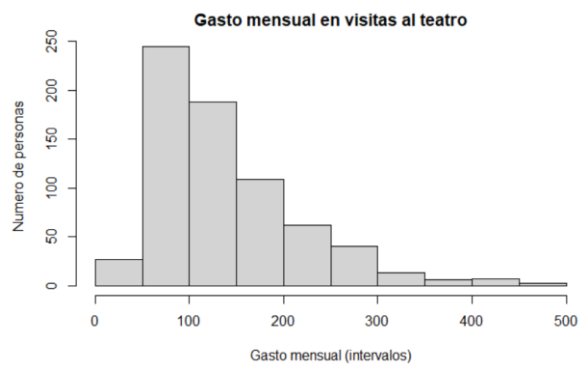
Tabla 3. Frecuencias para la variable Theatre

Class limits	f	rf	rf(%)	cf	cf(%)
[89.1,103.95)	18	0.03	2.58	18	2.58
[103.95,118.81)	81	0.12	11.59	99	14.16
[118.81,133.66)	268	0.38	38.34	367	52.50
[133.66,148.52)	131	0.19	18.74	498	71.24
[148.52,163.37)	124	0.18	17.74	622	88.98
[163.37,178.23)	33	0.05	4.72	655	93.71
[178.23,193.08)	29	0.04	4.15	684	97.85
[193.08,207.94)	10	0.01	1.43	694	99.28
[207.94,222.79)	2	0.00	0.29	696	99.57
[222.79,237.65)	0	0.00	0.00	696	99.57
[237.65,252.5)	3	0.00	0.43	699	100.00

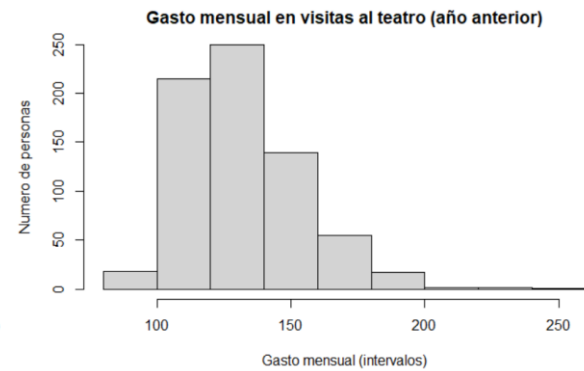
Tabla 4. Frecuencias para la variable Theatre_ly

Podemos ver, de acuerdo a la frecuencia absoluta (f) que en el presente año la cantidad de gasto en el teatro disminuyo, vemos que hay menos personas gastando más, y actualmente disminuyeron ese gasto, si vemos la columna cf (frecuencia relativa acumulada) cayo el porcentaje a intervalos en cifras bajas de gasto.

Vemos en las graficas 8 y 9 de frecuencias absolutas de una manera mas agradable como menos personas gastan más en el teatro.



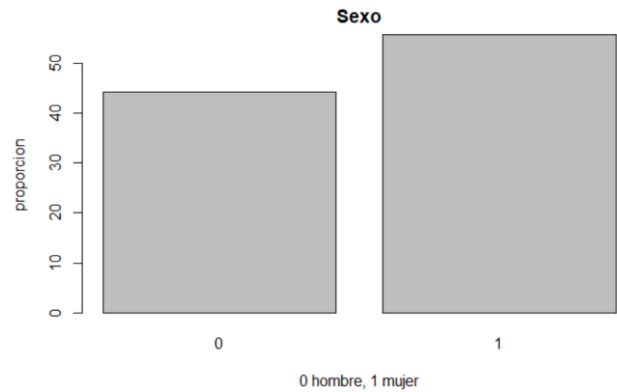
Grafica 8. Histograma



Grafica 9. Histograma

Sexo

Tenemos que las mujeres predominan en la muestra de clientes siendo el 55% de ellos (grafica 10)



Grafica 10. Frecuencia relativa para la variable Sex

Grupos Etarios

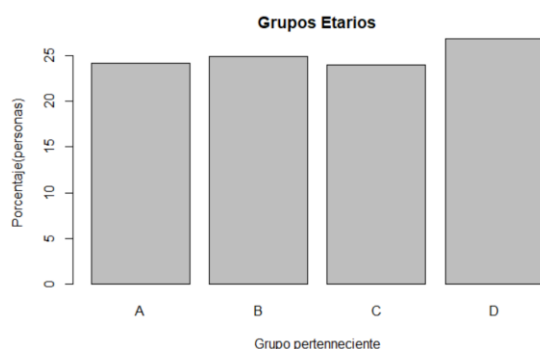
Se construye una variable que llamaremos “Rango” en ella definimos cuatro categorías que estarán construidas de la siguiente forma: A(Q1-Q2), B(Q2-Q3), C(Q3-Q4), D(Q4-Q5), de esta manera Q de 1 a 5 son los cuartiles generados (18, 28, 42, 53, 64) de las edades. Esta variable es cualitativa por lo que solo tenemos en cuenta frecuencia absoluta y relativa para el análisis

A	B	C	D	A	B	C	D
169	174	168	188	24.17740	24.89270	24.03433	26.89557

Tabla 5. Fre Abso G.Etario

Tabla 6. Frecu Relat G.Etario

Podemos ver entonces que las personas de 43 años en adelante como lo mencionamos anteriormente son mayoría en la muestra, aunque sea por muy poco, analizando desde medidas descriptivas tenemos que las personas entre 29 y 48 años son las que mas ganan anualmente, y las personas entre 18 y 27 años son las que menos ganan, esto teniendo en cuenta la mediana de cada uno de los grupos, también podemos ver que gastan mas en el teatro las personas entre 54 y 64 años y las que menos gastan son las personas entre los 18 y 27 años, lo anterior puede quizá explicarse por el nivel de ingreso anual, ahora bien las personas que mas invierten en cultura estan entre los 43 y los 53 años teniendo una muy leve diferencia con las personas entre 54 y 64, los que menos se ven interesados en cultura son los jóvenes entre 18 y 27 años.



Grafica 11. Frecu Abso G.Etario

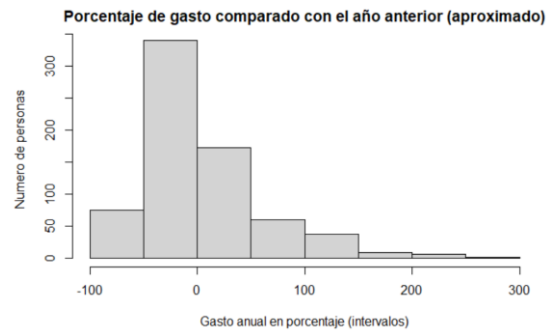
Información adicional

Se considera como información relevante mostrar algunos comportamientos derivados de los datos existentes, por esto se considera importante analizar el porcentaje que se dejó de gastar en cada persona en comparación al año anterior y que porcentaje está invirtiendo cada persona en cultura de sus ingresos, son datos relevantes que ayudan a perfilar el tipo de personas a las cuales el teatro debe de enfocar sus esfuerzos, alguien que esta invirtiendo en cultura pero poco en el teatro y hace parte de los clientes del teatro quizá significa que el teatro debe de hacer más para atraer nuevamente esa persona, de igual manera, comparar los dos años ayudan a identificar aquellas personas que también se estan alejando del teatro y a su vez si estas siguen invirtiendo en cultura podría ser un nivel de alarma en la pérdida de clientes

- porDeGasto: variable cuantitativa que calcula el porcentaje del ingreso anual que se gasta en cultura (para el cálculo se estimó el gasto en cultura anual multiplicando por 12)
- dgasto: la diferencia de gasto me dice el porcentaje que la persona dejo de gastar en el teatro en comparación con el año anterior (para efectos del caculo anual, se multiplico el gasto del año actual y anterior por 12 convirtiéndolos en una suma estimada para el gasto anual de ambos años en el teatro)

Class limits	f	rf	rf(%)	cf	cf(%)
[-78.876,-45.997)	105	0.15	15.02	105	15.02
[-45.997,-13.117)	241	0.34	34.48	346	49.50
[-13.117,19.763)	138	0.20	19.74	484	69.24
[19.763,52.642)	108	0.15	15.45	592	84.69
[52.642,85.522)	44	0.06	6.29	636	90.99
[85.522,118.4)	29	0.04	4.15	665	95.14
[118.4,151.28)	19	0.03	2.72	684	97.85
[151.28,184.16)	7	0.01	1.00	691	98.86
[184.16,217.04)	4	0.01	0.57	695	99.43
[217.04,249.92)	3	0.00	0.43	698	99.86
[249.92,282.8)	1	0.00	0.14	699	100.00

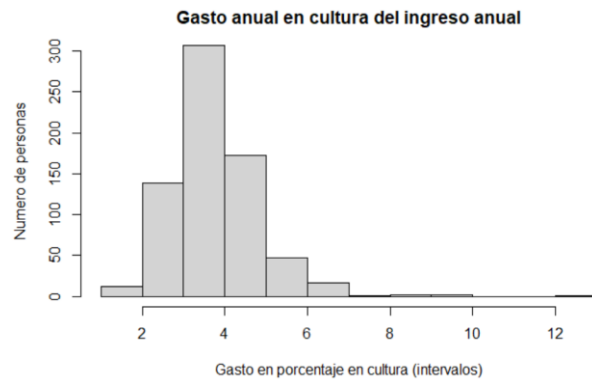
Tabla 7. Frecuencias comparación de gasto



Grafica 12. Frec Absoluta comparación de gasto

Class limits	f	rf	rf(%)	cf	cf(%)
[1.2567,2.335)	31	0.04	4.43	31	4.43
[2.335,3.4133)	234	0.33	33.48	265	37.91
[3.4133,4.4915)	298	0.43	42.63	563	80.54
[4.4915,5.5698)	99	0.14	14.16	662	94.71
[5.5698,6.648)	28	0.04	4.01	690	98.71
[6.648,7.7263)	4	0.01	0.57	694	99.28
[7.7263,8.8046)	2	0.00	0.29	696	99.57
[8.8046,9.8828)	2	0.00	0.29	698	99.86
[9.8828,10.961)	0	0.00	0.00	698	99.86
[10.961,12.039)	0	0.00	0.00	698	99.86
[12.039,13.118)	1	0.00	0.14	699	100.00

Tabla 8. Frecuencias % de gasto en cultura



Grafica 13. Porcentaje de gasto en cultura

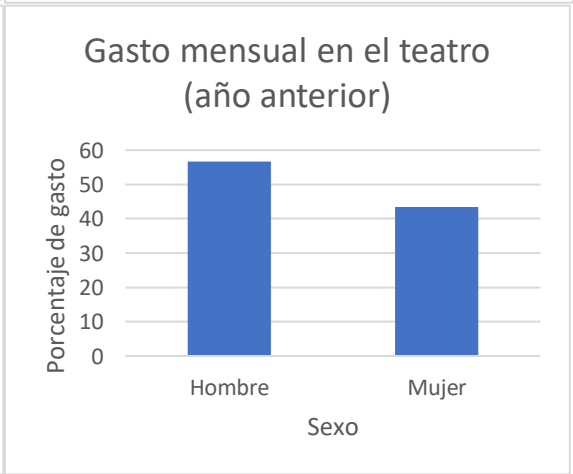
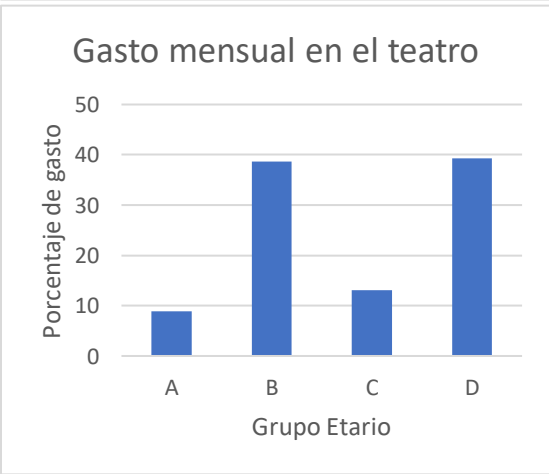
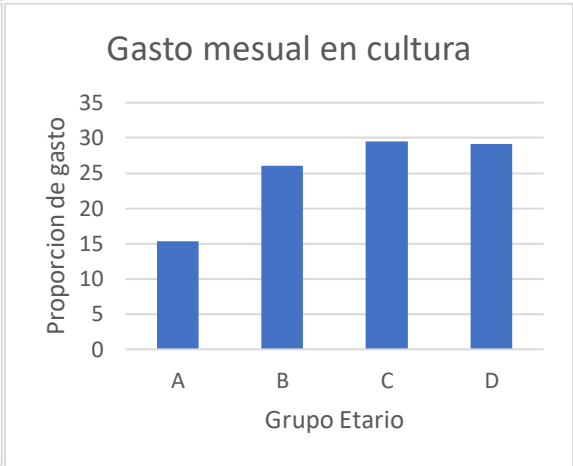
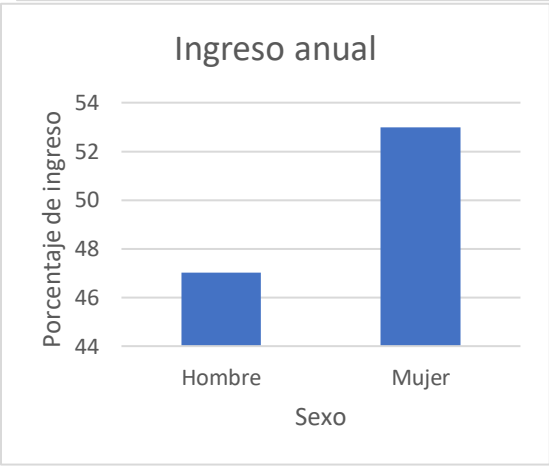
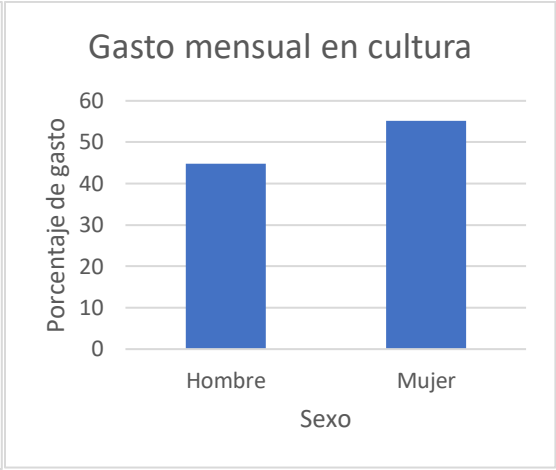
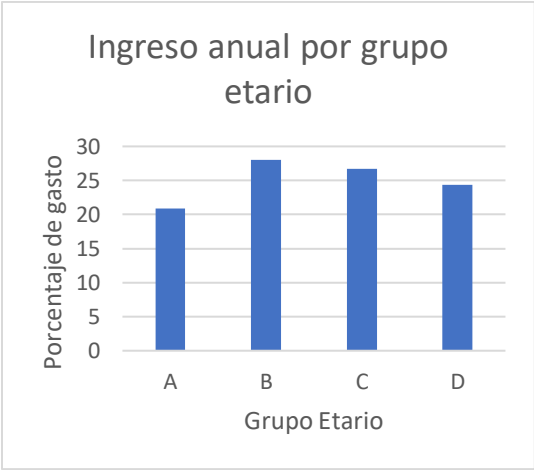
Como varios análisis superficiales de datos anteriores lo decían, podemos ver en las frecuencias de la tabla 7 y el grafico 12 que hay personas que dejaron de gastar hasta el 90% (-90%) comparado con el año anterior, vemos que una gran cantidad de clientes se ubican por debajo del 0.

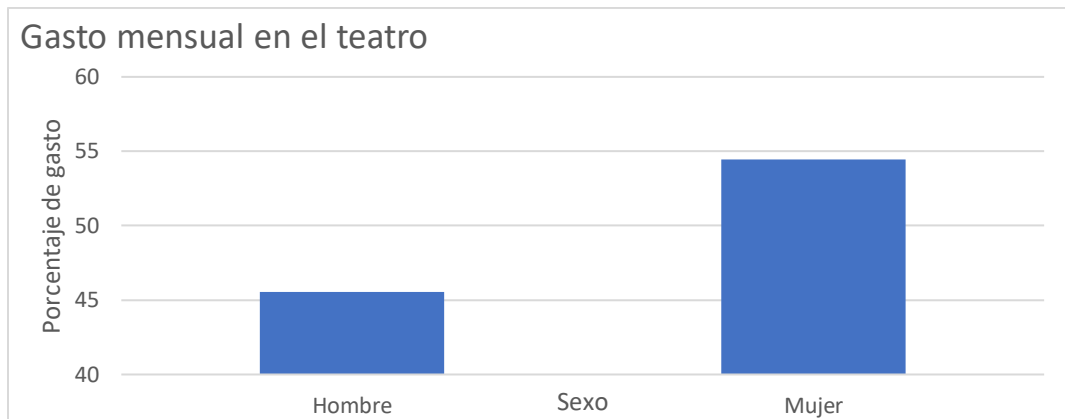
Ahora mirando el porcentaje que se gastan en cultura vemos que la mayoría de las personas según el grafico 13 gasta entre un 2% y un 6% sus ingresos anuales en cultura, también podemos ver según la tabla 8 que el 98.71% del total de la muestra gasta menos del 7% de sus ingresos anuales en cultura, dejando entonces por ejemplo solo un 0.14% gastando entre 21.039% y 13.118% de sus ingresos en cultura

¿En qué edad y genero enfocar la campaña?

Para la responder la pregunta, se ha tenido en cuenta la mediana considera un factor de mayor confiabilidad para cada grupo etario, y se grafican comparando los resultados, tenemos:

Si consideramos que la idea es incentivar la asistencia de los clientes frecuentes, la campaña debe de estar enfocada a las personas que tienen la tarjeta, es decir, los que estan en la muestra de este dataset, y además no asisten al teatro, por lo tanto analizando las graficas que se encuentran a continuación en las cuales se relaciono por cada grupo etario, su gasto mensual en cultura, su ingreso anual y su gasto mensual en el teatro, para las anteriores también fueron relacionadas con los géneros añadiendo el gasto del año anterior mensualmente.





Nota: adjunto se encuentra el Excel con las gráficas anteriores.

Podemos concluir de los gráficos anteriores que la campaña debe de ir enfocada al grupo C pues son de los grupos que mas gastan en cultura y además el segundo en mejores ingresos anuales es que no esta gastando en el teatro, también se considero el grupo A pero estos tienen ingresos bajos y estan poco interesados en actividades culturales por lo tanto podría ser un mal enfoque, ahora bien, cuando nos dirigimos al género, tenemos que las mujeres tienen mejores ingresos, invierten mas en el teatro y se ven muy interesadas en temas culturales pero, por otro lado los hombre en el año anterior invertían mas en el teatro y aunque no estan tan interesados en temas culturales como las mujeres si estan cerca, además los ingresos anuales solo varían un 10% con respecto al de las mujeres, una campaña para que regresen al teatro enfocada en ellos seria ideal, recuperando un publico que parece bajar no solo para el teatro si no también en cultura.

Nombre: Daniel Alejandro Mesa Arango.

Grupo Unión: CM0244U007 ESTADISTICA GENERAL

Video de presentación: <https://youtu.be/OvHrMfIL-f0>