

INFORME UNIVERSITYHACK 2018 DATATHON



Reto Wefferent Card Analytics

AUTORES:

DAVID BERNABÉ TORRES

FÉLIX FRANCISCO ENRÍQUEZ ROMERO

ÍNDICE

1. INTRODUCCIÓN
2. EVOLUCIÓN TEMPORAL DEL CONSUMO POR SECTOR
3. IMPORTE ACUMULADO POR FRANJA HORARIA Y SECTOR
4. CÓDIGOS POSTALES QUE MÁS CONSUMEN
5. CÓDIGOS POSTALES DONDE MÁS SE CONSUME
6. DISTRIBUCIÓN PORCENTUAL DEL CÓDIGO POSTAL DONDE MÁS SE CONSUME
 - a. DISTRIBUCIÓN EN BASE A LA EDAD
 - b. DISTRIBUCIÓN EN BASE A LA RENTA
7. CONCLUSIÓN

1. INTRODUCCIÓN

El presente informe refleja un análisis exploratorio sobre los datos transaccionales realizados con tarjeta de débito/crédito en la Región de Murcia en los periodos comprendidos desde el mes de octubre de 2015 hasta septiembre de 2017.

Estos datos han sido proporcionados por la entidad Cajamar con motivo de la UniveristyHack 2018 y el reto escogido es “Reto Wefferent Card Analytics”.

Al inicio partimos de un Dataset que recoge los datos generales sobre las transacciones entre los cuales tenemos el tipo de sector donde se realiza la transacción, el código postal donde tiene lugar dicha transacción, a la vez que el código postal de procedencia del cliente y a demás el importe y el número de operaciones realizadas en una franja horaria determinada. También se proporciona información acerca del sexo, edad y renta del cliente.

Dentro de propiedad renta encontramos tres categorías (baja, media y alta) y en la propiedad edad, encontramos (joven, adulto y pensionista)

Para realizar la carga, el filtrado y el procesado de los datos hemos hecho uso de la librería Pandas y NumPy de Python y para la posterior representación la librería Matplotlib.

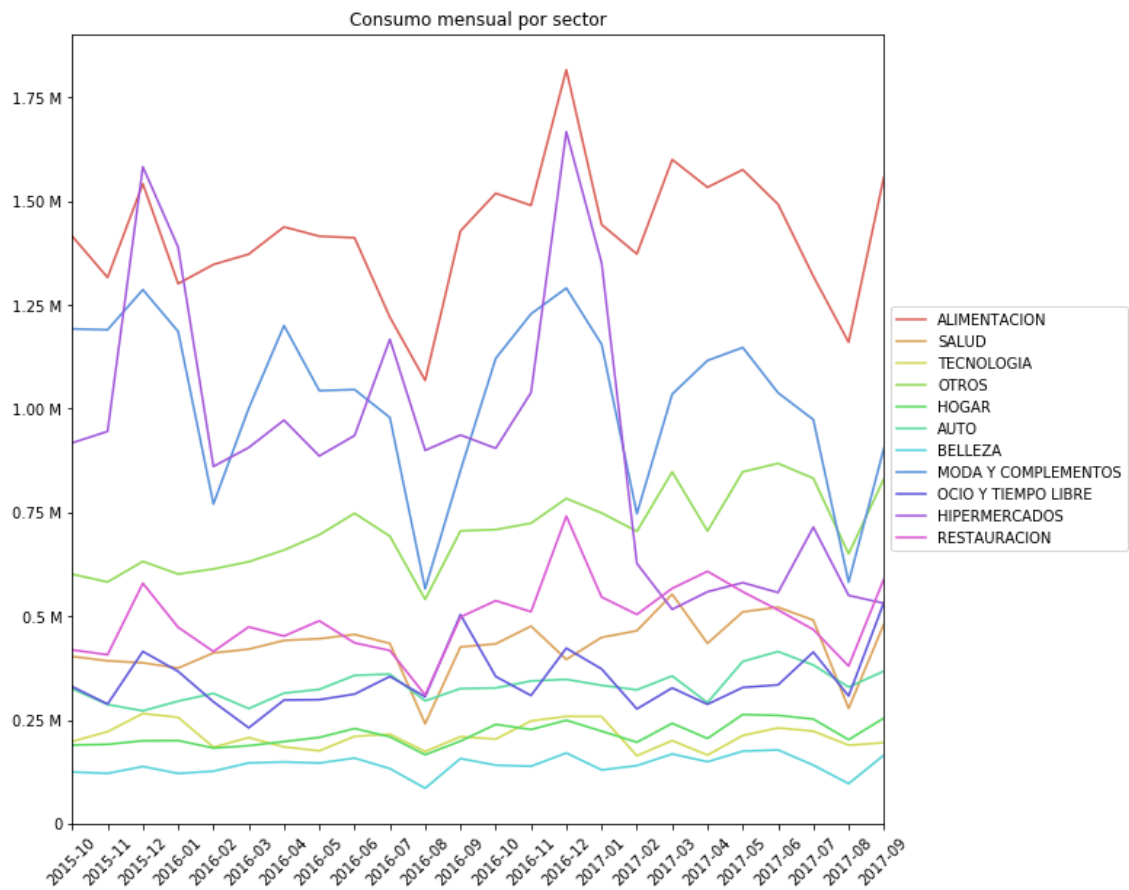
Tras la lectura del conjunto de datos contenidos en un fichero .csv realizamos la carga en un DataFrame, el cual nos permitirá realizar las operaciones correspondientes para cada parte del análisis. El DataFrame resultante es representado en distintos formatos de gráficas.

A continuación, detallamos los diferentes puntos donde hemos enfocado el análisis.

2. EVOLUCIÓN TEMPORAL DEL CONSUMO POR SECTOR

En la siguiente gráfica de líneas representamos el consumo por sector realizado en el periodo comprendido del 01/10/2015 al 30/09/2017 en la ciudad de Murcia.

En el eje vertical encontramos el importe expresado en millones de euros y en el eje horizontal el periodo temporal en meses. Cada línea representa un sector.



Observamos que en la gráfica se da un patrón periódico anual, en el cual destacan picos de subida y de bajada, coincidiendo los meses de diciembre en los puntos de subida y agosto en los puntos de bajada.

Independientemente del año, el mes de diciembre es el mes en el que los sectores Alimentación, Hipermercados, Moda y Complementos, Restauración y Ocio alcanzan su punto álgido en consumo, coincidiendo con las fiestas navideñas; ya que es en esta festividad cuando una gran parte de la población disfruta de vacaciones y por lo tanto realizan un mayor consumo en los sectores mencionados.

Opuesto al fenómeno navideño encontramos agosto, mes en el que se produce un descenso generalizado del consumo independientemente del sector lo cual es indicio del descenso de consumidores en la ciudad de Murcia; cabe destacar que para esta época del año, gran parte del turismo se centra en el litoral.

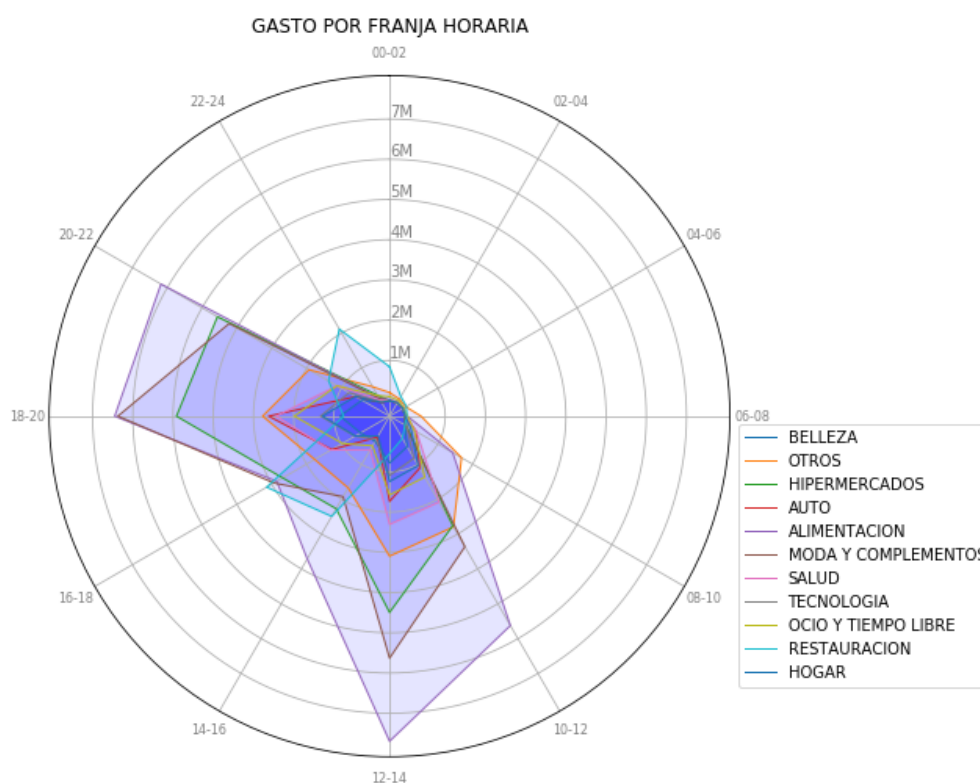
Es interesante observar que el consumo en el sector tecnología y hogar son semejantes y que priman sobre el sector belleza. En el sector moda observamos puntos álgidos en el mes de abril coincidiendo con la campaña de primavera-verano pero sin llegar a alcanzar el nivel que se da en navidad.

Al comparar los picos que se dan en el mes de diciembre, vemos que en el año 2016 se produjo un incremento del importe gastado en gran parte de los sectores en comparación con el año 2015, y si contemplamos la tendencia alcista que se da en la mayoría de los sectores a final del periodo podemos suponer que el consumo por sector en el mes de diciembre de 2017 será mayor que el año anterior.

3. IMPORTE ACUMULADO POR FRANJA HORARIA Y SECTOR

En el Dataset se nos proporciona una variable llamada “FRANJA_HORARIA”, la cual es una variable categórica definida en periodos de 2 horas, definiendo así 12 franjas desde las 00 hasta las 24 horas.

En el siguiente gráfico de araña representamos el consumo total por sector en las diferentes franjas horarias. Hemos asignado los valores de las franjas horaria en sentido horario a lo largo del radar. Los distintos niveles representan el consumo en millones de euros, y cada sector esta representado por un área de distinto color.



A simple vista se observa que es durante el día cuando se producen el mayor número de transacciones destacando las horas comprendidas entre las 10 y las 14 horas, y entre

las 18 y las 22 horas a excepción del sector restauración cuyos valores máximos están comprendidos entre las 14 a 18 horas, y de las 22 a las 24 horas, coincidiendo con las horas de comida y cena.

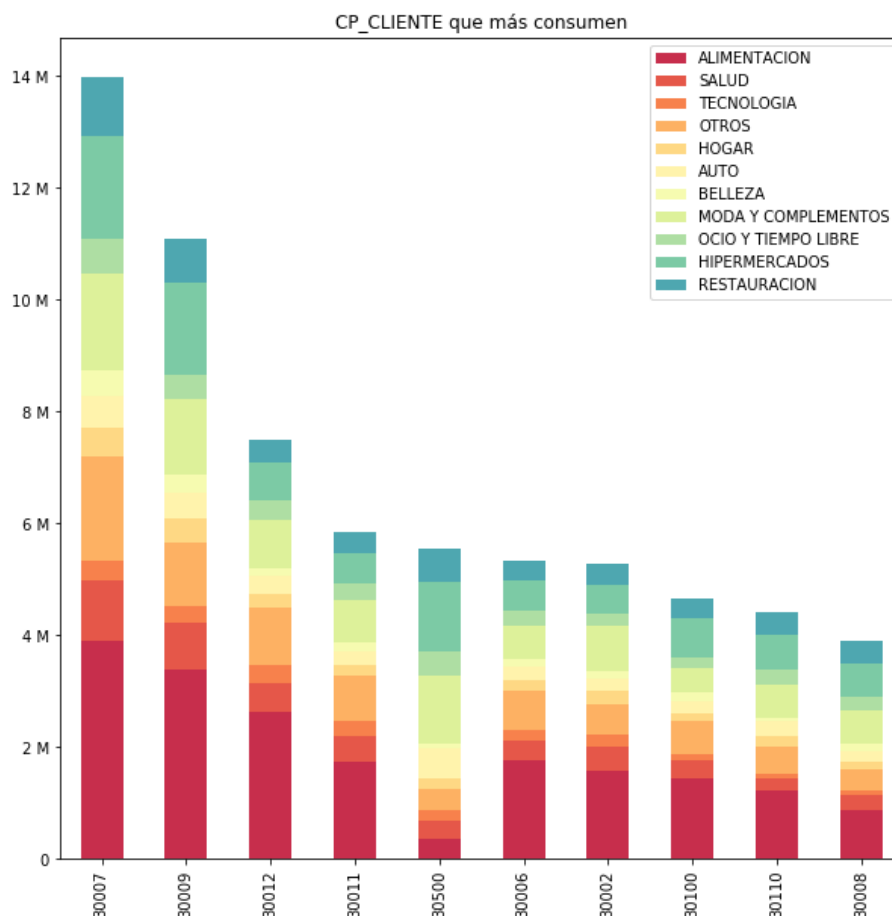
Se observa que el consumo por sector en la franja horaria comprendida entre las 22 hasta las 8 horas, se debe al horario de apertura establecido de los comercios e hipermercados.

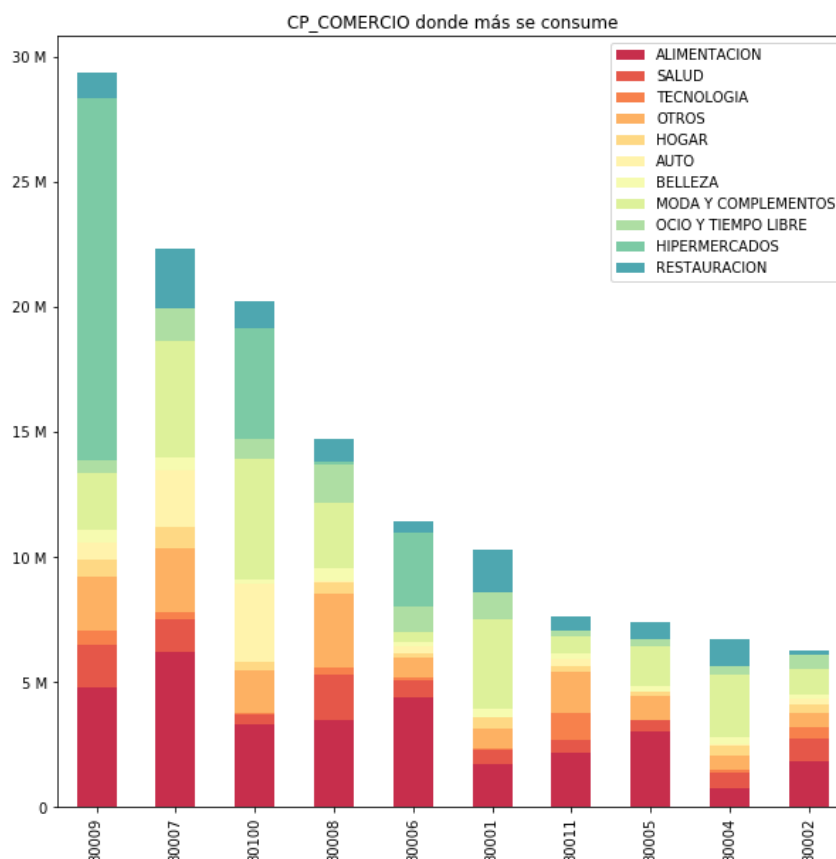
Cabe mencionar que no se representan las transacciones realizadas a través de internet, lo cual podría variar considerablemente el consumo representado en los sectores de tecnología y moda.

4. CÓDIGOS POSTALES QUE MÁS CONSUMEN

En el Dataset encontramos la propiedad CP_CLIENTE, este código postal puede pertenecer a cualquier código postal de la región de Murcia.

En el siguiente gráfico de barras mostramos el consumo por sector y total realizados por los clientes ordenados de mayor a menor. Hemos elegido los 10 códigos postales cuyos habitantes consumen más en la ciudad de Murcia.





Con este gráfico podemos visualizar que sectores económicos se desarrollan en cada código postal.

Podemos destacar que los hipermercados se centran en los códigos postales 30009,30100,30006.

Otro aspecto a tener en cuenta es la ausencia del sector automovilístico en los códigos postales 30001,30005.

Por último y comparando la gráfica actual con la obtenida en el apartado anterior, observamos que los códigos postales de los clientes que más consumen guardan relación con los códigos postales de los comercios donde más se consumen. No obstante, cabe destacar la diferencia entre las gráficas del consumo en el código postal 30100; dado que es el tercer código postal donde más se consume (20 millones aprox.), pero es el octavo código postal que más consume (5 millones aprox.). Este dato nos indica claramente, que en este código postal los comercios reciben tres cuartas partes de su importe total de clientes que no residen en ese código postal.

6. DISTRIBUCIÓN PORCENTUAL DEL CÓDIGO POSTAL DONDE MÁS SE CONSUME

En esta sección del análisis nos centramos en el estudio del código postal 30009 ya que es el más representativo a nivel de consumo.

Vamos a enfocar nuestro análisis en la propiedad Edad y Renta de los clientes en función del número de operaciones realizada por estos.

a. DISTRIBUCIÓN EN BASE A LA EDAD

En este apartado encontramos que la propiedad edad del cliente puede estar comprendida entre joven, adulto o pensionista. Dado un número de operaciones tenemos la proporción del número de operaciones que han sido realizadas por los clientes de diferentes edades. A continuación mostramos el cómputo total de operaciones realizadas en los distintos sectores por los clientes pertenecientes a distintas edades.

	joven	adulto	pensionista
HOGAR	2414	8430	917
OTROS	14326	48224	3340
RESTAURACION	13015	19223	1075
OCIO Y TIEMPO LIBRE	1476	4863	222
MODA Y COMPLEMENTOS	13453	35954	1968
ALIMENTACION	19243	108752	15420
BELLEZA	2293	8872	1217
AUTO	731	3561	344
HIPERMERCADOS	53569	187878	14532
TECNOLOGIA	751	4428	651
SALUD	5727	24120	2831

Según el número de operaciones totales por edades: 126998 para jóvenes, 454305 para adultos y 42517 para pensionistas; la mayor parte de los consumidores son adultos siendo estos número de operaciones casi cuatro veces mayor que los realizados por jóvenes y diez veces más que las realizadas por pensionistas. A continuación mostramos el porcentaje de operaciones por sector según la edad del consumidor.

Gráfico Jóvenes:

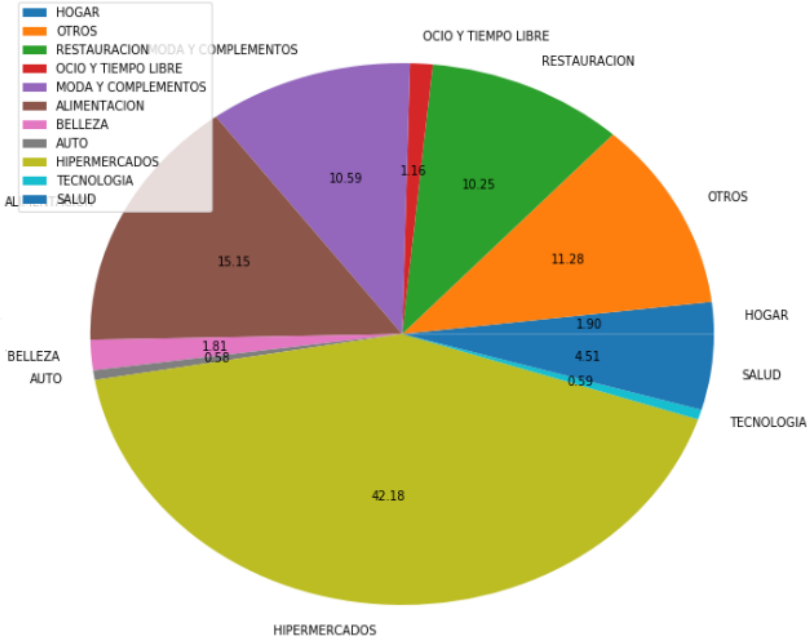


Gráfico Adultas:

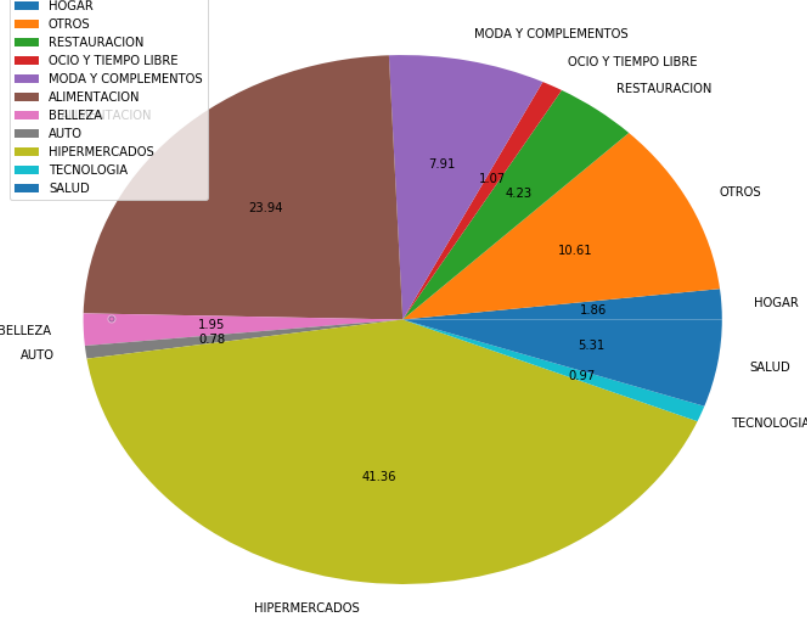
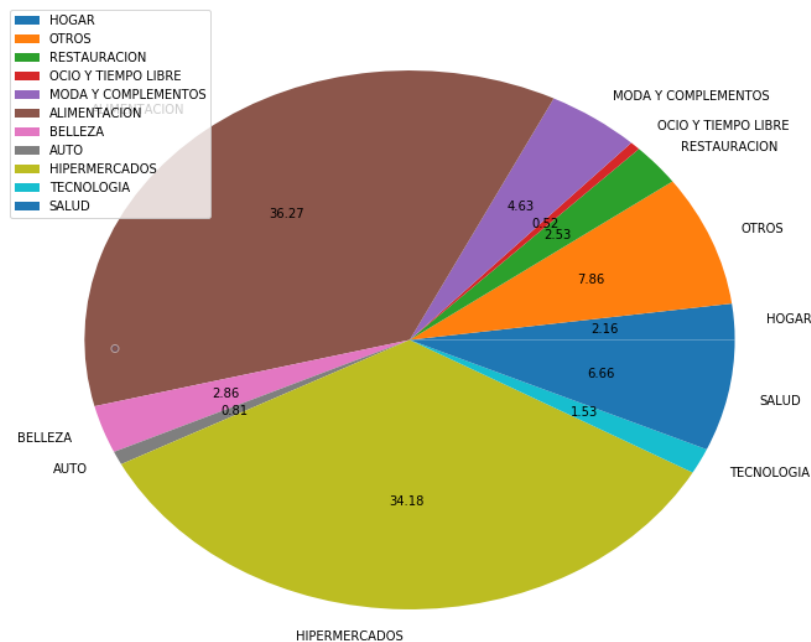


Gráfico Pensionistas:



Al observar los gráficos sectoriales anteriores, podemos responder a la pregunta de “¿En qué gasta el consumidor dada su edad?”. A mayor edad el consumo en el sector alimentación aumenta y por otro lado, cuanto más joven es el consumidor mayor es el consumo en el sector Moda, Hipermercados, Restauración y Otros. El resto de sectores mantienen una semejanza en proporción al número de operaciones.

b. DISTRIBUCIÓN EN BASE A LA RENTA

En este apartado analizamos el consumo en los distintos sectores en función de la renta de los consumidores que puede tomar tres valores: baja, media o alta. A continuación mostramos el cómputo total de operaciones realizadas en los diferentes sectores en función de la renta.

	rentaBaja	rentaMedia	rentaAlta
HOGAR	4711	4901	2149
OTROS	25165	26831	13894
RESTAURACION	15128	12370	5815
OCIO Y TIEMPO LIBRE	2824	2500	1237
MODA Y COMPLEMENTOS	22060	20287	9028
ALIMENTACION	60024	60063	23328
BELLEZA	4925	5245	2212
AUTO	1644	1903	1089
HIPERMERCADOS	96634	108257	51088
TECNOLOGIA	2284	2362	1184
SALUD	12943	13539	6196

Según el número de operaciones totales por renta: 248342 para personas con renta baja, 258258 para personas de renta baja y 117220 para personas de renta alta; podemos apreciar que el número de operaciones efectuadas por personas de renta alta, es el 18% del total de operaciones realizadas en el código postal 30009, mientras que el porcentaje de número de operaciones realiza por personas de renta baja y media es muy similar.

El gasto por sector para cada renta viene representada en las siguientes gráficas.

Gráfico Renta Baja:

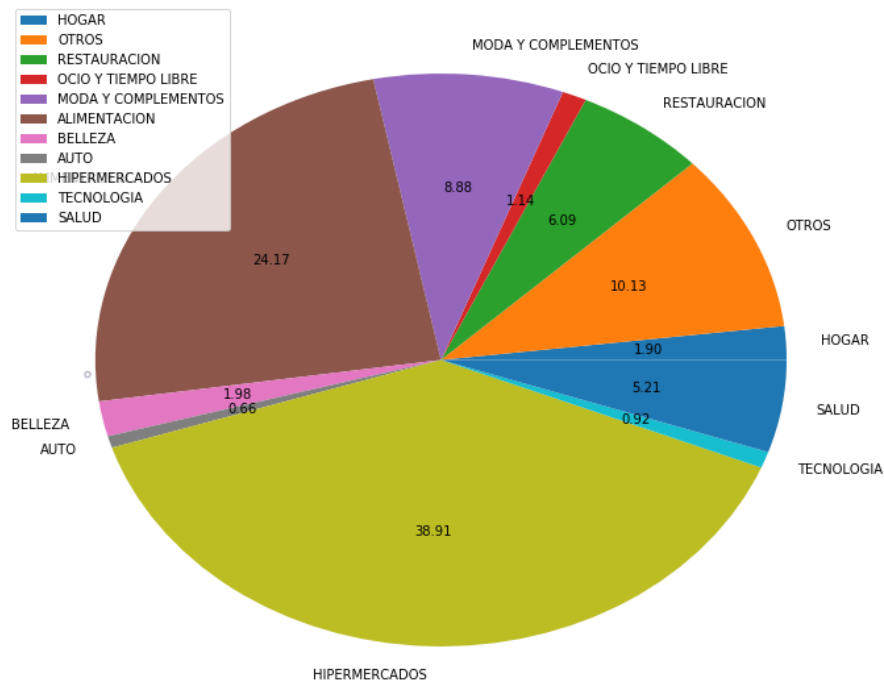


Gráfico Renta Media:

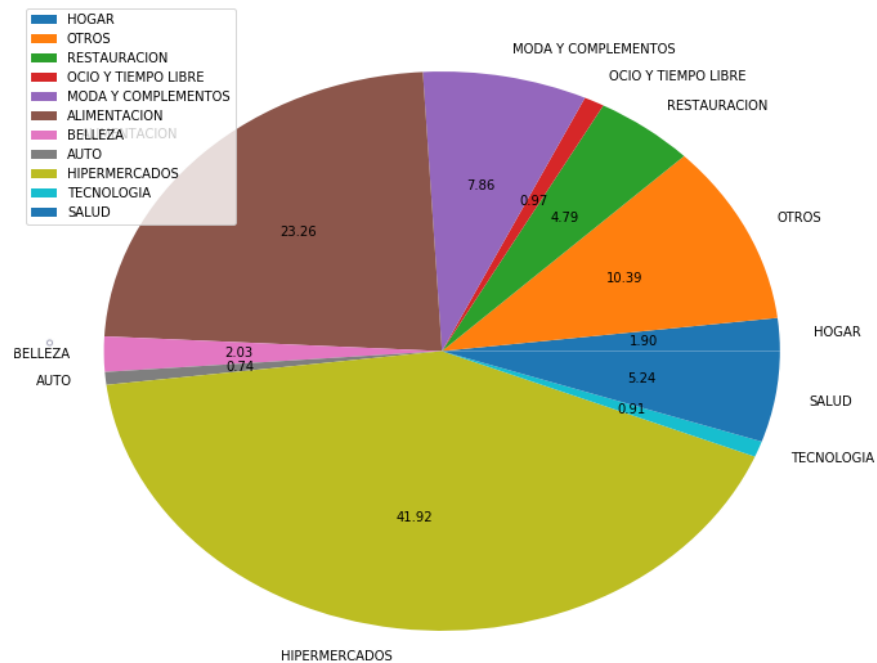
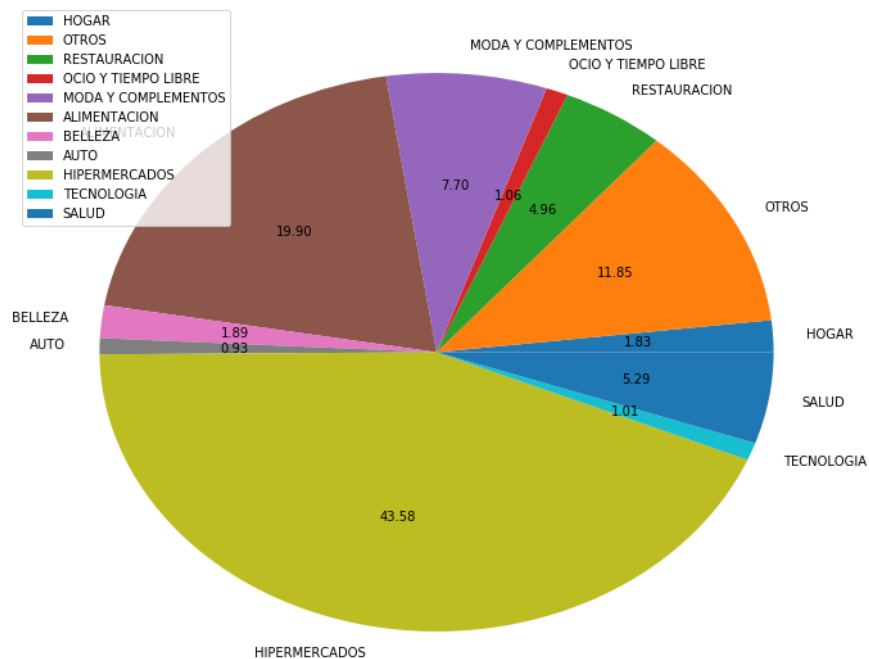


Gráfico Renta Alta:



Dados los siguientes gráficos sectoriales podemos determinar que independientemente de la renta del consumidor, este consume los diferentes sectores de una manera similar.

7. CONCLUSIÓN

A partir de los datos proporcionados hemos generado información económica y sociológica relevante acerca de las transacciones que tienen lugar en la ciudad de Murcia.

A nivel general se produce un aumento del consume en campaña navideña, opuesto a lo que ocurre en el mes de agosto. También hemos estudiado las preferencias de los consumidores dada su edad y su renta en los diferentes sectores en el código postal donde más transacciones se realizaron.

Hemos podido aproximar el consumo en función de las franjas horarias; hubiera sido interesante disponer de las transacciones realizadas en e-commerce, ya que al contar con estos datos hubiéramos sido más precisos a la hora de determinar el consumo por franja horaria y sector, pudiendo ser el consumo en Tecnología y Moda significativamente mayor teniendo en cuenta el volumen de operaciones que se realizan cada día en este sector a través de internet.

A modo de mejora, proponemos implementar un servicio REST que nos permita visualizar datos específicos a elección del usuario.