INFORMÁTICA Y SISTEMAS
REVISTA DE TECNOLOGÍAS DE LA INFORMÁTICA Y LAS TELECOMUNICACIONES

Vol. 5 Núm. 2 (27-29 Octubre, 2021), pp. 44-58 https://doi.org/10.33936/isrtic.v5i2.3946 | ISSN 2550-6730

Recibido: 17/09/2021
Aceptado: 13/10/2021

Análisis de la alimentación en regiones del Ecuador mediante Big Data.

Víctor Noé Sánchez Carreño[1], Jandry Hernaldo Franco Cantos[1], María José Vélez Cedeño[1], Marely del Rosario Cruz Felipe[1]

[1]Facultad de Ciencias Informáticas, Universidad Técnica de Manabí, Portoviejo, Ecuador
[1]vsanchez5679@utm.edu.ec, [1]jfranco4180@utm.edu.ec, [1]mvelez7672@utm.edu.ec, [1]marely.cruz@utm.edu.ec

RESUMEN La alimentación saludable es un elemento relevante para el bienestar de las personas; ayuda a evitar la desnutrición, al igual que enfermedades no transmisibles, por ello, resulta importante el conocer cuáles son aquellos alimentos más consumidos por parte de los ecuatorianos según la región en la que residen, con el fin de determinar cómo la alimentación presente en estas regiones influye en la salud de sus habitantes. En la presente investigación se realiza un análisis de los tipos de alimentación en las regiones Costa y Sierra del Ecuador. Para ello se empleó a las redes sociales como fuentes de datos, ya que cada vez son más utilizadas para compartir información u opiniones acerca de situaciones o elementos de diversas índoles, siendo Twitter la empleada, por ser una de las redes sociales más importantes para ello. La recolección de datos de la red social Twitter se llevó a cabo mediante el uso del lenguaje de programación Python y estuvo enfocada en Tweets que aluden a la alimentación, comparando datos de las ciudades de la Sierra y la Costa, clasificando alimentos saludables y no saludables y tipos de alimentos, lo que permite identificar en cada región los alimentos más relevantes. También se analiza la valoración del Tweets, en cuanto a cantidad de interacciones de los mismos. Obteniendo que en la Sierra se consumen alimentos más saludables, los resultados son visualizados en diferentes reportes en Power BI.

Palabras claves: Alimentación, Big Data, Power BI, Python, Twitter.

Analysis of food in regions of Ecuador through Big Data

ABSTRACT Healthy eating is a relevant element for people's well-being; helps to avoid malnutrition, as well as non-communicable diseases, therefore, it is important to know those are the foods most consumed by Ecuadorians according to the region in which they reside, in order to determine how the diet is present in these regions influences the health of its inhabitants. In this research, an analysis of the types of diet in the Costa and Sierra regions of Ecuador is carried out. For this, social networks were used as data sources, since they are increasingly used to share information or opinions about situations or elements of various kinds, Twitter being the one used, as it is one of the most important social networks for this. . The data collection of the social network Twitter was carried out through the use of the Python programming language and was focused on Tweets that allude to food, comparing data from the cities of the Sierra and the Coast, classifying healthy and unhealthy foods and types of foods, which makes it possible to identify the most relevant foods in each region. The evaluation of the Tweets is also analyzed, in terms of number of interactions thereof. Obtaining that healthier foods are consumed in the Sierra; the results are displayed in different reports in Power BI.

KEYWORDS: Food, Big Data, Power BI, Python, Twitter.

https://doi.org/10.33936/isrtic.v5i2.3946 | ISSN 2550-6730

1. Introducción

En el Ecuador, como en el resto del mundo, existen diversas problemáticas originadas a partir de una mala alimentación, principalmente aquellas que tienen que ver con la salud y estado físico de las personas [1] [2]. El identificar cuando una alimentación se la puede categorizar como mala, genera una base de conocimientos que permite elaborar estrategias para prevenir o controlar las consecuencias originadas por el consumo de alimentos que no ofrecen una correcta nutrición, en el caso de Ecuador se debe tener en cuenta la variedad gastronómica presente principalmente en las regiones de la Costa y Sierra, por ello, se deben manejar de forma independiente [3]. Existen varios estudios que muestran aquellos factores externos e internos que influyen tanto para una buena como mala alimentación, sin embargo se encuentran pocos estudios que hagan uso de las redes sociales como medio principal para obtener información, por ello, este tema resulta de interés gracias a que, mediante publicaciones de los usuarios de la red social Twitter, se puede generar un análisis que permita el identificar las enfermedades presentes en el Ecuador a raíz de los hábitos alimenticios de la población según su región.

Debido a que las redes sociales dotan de información útil proveniente de las publicaciones de las personas, se puede obtener un gran conjunto de datos, sin embargo, se debe tener presente que existe una amplia gama de temas tratados en las redes sociales, es por ello que se deben emplear procesos para filtrar la información con el fin de direccionar y obtener aquello que esté relacionado con la alimentación en el Ecuador [4] [5]. Mediante el uso de Python + Tweepy, se consigue el extraer datos de Twitter enfocados al objeto de estudio, a lo cual, con el fin de almacenar datos más limpios y homogéneos, se les aplica algoritmos de procesamiento. La base de datos resultante recibirá datos de Twitter de forma constante, debido a que la recolección será automatizada y en determinados periodos de tiempo, es por ello que, el uso de la herramienta de Power BI resulta importante para la creación de informes y gráficas en base a este gran conjunto de datos, catalogada como Big Data, posibilitando el observar las tendencias alimenticias actuales, al igual que en periodos de tiempo específicos.

En estudios como [6] se realizó una aplicación que detecta la actividad de aquellas cuentas de usuario que realizaban publicaciones sobre conductas que conlleven a trastornos de conducta alimentaria y luego se enviaban reportes a los moderadores de Twitter para que estas cuentas fuesen sancionadas por el contenido que publican. Por su parte en [7] se realizó un estudio acerca de la influencia de las publicaciones sobre consejos acerca de una buena alimentación, también se analizaron generaciones que conviven con la era digital y la influencia que las redes sociales generan en el comportamiento alimentario de los usuarios.

Los trabajos mencionados analizan hábitos alimenticios en otros países, sin embargo, no se especifican estudios relacionados en Ecuador, que permitan clasificar estos hábitos por regiones, para un análisis en cuanto a influencias de los mismos en la salud.

El objetivo de esta investigación radica en realizar un análisis de la alimentación en regiones del Ecuador mediante el Big Data.

2. Materiales y Métodos

La metodología seguida para la realización de este trabajo consta de tres pasos, los cuales se muestran en la Figura 1.

Como se muestra en la Figura 1, los pasos son: adquisición de datos, procesamiento y métricas de visualización, los cuales se describen a continuación:

INFORMÁTICA Y SISTEMAS
REVISTA DE TECNOLOGÍAS DE LA INFORMÁTICA Y LAS TELECOMUNICACIONES
Vol. 5 Núm. 2 (27-29 Octubre, 2021), pp. 44-58 https://doi.org/10.33936/isrtic.v5i2.3946 | ISSN 2550-6730

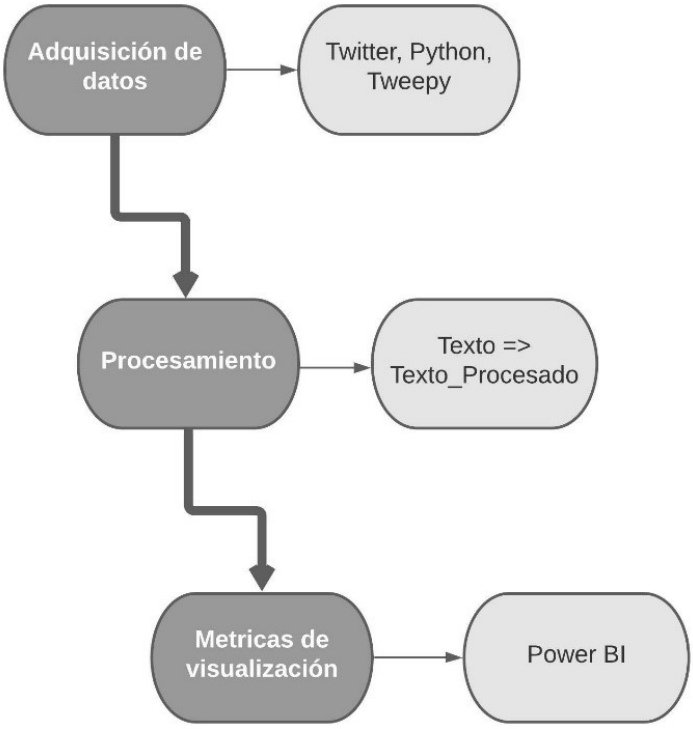


Figura 1: Pasos para realizar un análisis de la alimentación en regiones del Ecuador mediante el Big Data.

2.1. Adquisición de datos

Los datos se extrajeron de la red social Twitter, para ello se empleó la librería Tweepy [8] [9] en el lenguaje de programación de Python. Mediante esta librería se puede hacer uso del API de Twitter, medio que realiza las consultas de datos que se necesiten, por ello se establece una serie de funciones y parámetros [10], en donde se indica principalmente palabras claves a buscar dentro de los tweets, al igual que coordenadas que indican el punto geográfico de donde se extraerá la información.

La búsqueda de datos en la red social Twitter se definen a partir de cada uno de los cuatro conjuntos de palabras que se establecieron como alimentos saludables, no saludables, medios y vegetales; empleando la ubicación geográfica se define los lugares en los cuales se hará la búsqueda, lo que permite seccionar los datos provenientes de la Costa y de la Sierra.

Ambos conjuntos de datos están estructurados para almacenar información relacionada al tweet, específicamente, el id del usuario, el nombre del usuario, el usuario, la cantidad de tweets que ha publicado, la cantidad de seguidores, la cantidad de seguidos, si es un usuario verificado o no, la fecha y la hora del tweet en que se publicó, el id del tweet, el texto completo del tweet, el texto preprocesado del tweet, la localización del lugar en donde se publicó el tweet.

Esta información es extraída de forma constante con la finalidad de contar con los datos más actuales posibles, es por ello que se hace uso de una base de datos de MySQL para almacenar toda la información recopilada. Utilizar un gestor de base de datos permitió que el almacenamiento de los datos se conserve de manera no volátil y sin duplicados ya que se toma como llave primaria el id del tweet por ser único;

 https://doi.org/10.33936/isrtic.v5i2.3946 | ISSN 2550-6730

además, se pudo exportar fácilmente a varios tipos de archivos para el respaldo y la replicación de este trabajo de manera manual aparte de la sincronización con servicios de nube.

2.2. Procesamiento

El preprocesamiento de los datos se ha realizado principalmente con un enfoque hacia el texto de cada tweet con la utilización de Python, debido a que contienen diversos caracteres innecesarios que no aportan información y a su vez afectan el rendimiento del Power BI al momento de generar los informes. El preprocesado consiste en eliminar los StopWords, convertir en minúscula el texto y eliminar aquellos caracteres que no cuenten con un código ASCII, dicho preprocesado optimiza la búsqueda de palabras y creación de gráficos dentro del Power BI.

Con Python se logra con satisfacción crear métodos automatizados que no requirieron supervisión, que previamente para el procesamiento es una parte recomendada.

El procesamiento de los datos se dio para aquella preparación de los mismos, en donde estos son los que se usaron en las métricas de visualización, ya que, en el caso de este trabajo, no se usaron algoritmos que requieran estos datos de entrada.

La información utilizable está situada en el gestor de base de datos en donde a partir de la ejecución de sentencias SQL, se logró unificar los datos en un nuevo conjunto de datos, ya que con anterioridad se tenían solo dos conjuntos de datos, que pertenecían a la región Costa y Sierra.

2.3. Métricas de visualización

Las visualizaciones son importantes para este trabajo por lo que se ha especificado la herramienta de Microsoft Porwer BI [11][12][13] para ser utilizada al momento de realizar las diferentes gráficas que permiten analizar la información recopilada de forma entendible, gracias a su facilidad para poder administrar grandes volúmenes de datos, y la versatilidad al momento de generar informes. Power BI permite integrar diversas herramientas externas, es por ello la conexión con una base de datos, en este caso SQL, se realiza de forma correcta.

Y con la conexión a la base de datos, se extraen las tablas con las que se va a trabajar, en este trabajo se utilizaron cinco tablas en total las cuales comprenden los tweets de la región Costa, tweets de la región Sierra, tweets de ambas regiones unificados, ciudades de la región Costa y ciudades de la región Sierra.

Tres de las tablas extraídas de SQL son las que se obtuvieron en la adquisición de datos y el procesamiento de datos. Las otras dos tablas restantes se generaron manualmente a partir de las ciudades de donde se extrajeron los tweets.

La Tabla 1 muestra un resumen de los elementos gráficos que se utilizarán para dar las explicaciones en la sección de resultados, estos elementos gráficos son los que se utilizan para dar una representación de los valores que se obtuvieron en los tweets.

La segmentación de datos corresponde también a los paneles, y se utilizó para visualizar datos de manera selectiva y así comprender puntos de vistas distintos, lo que hizo posible las comparaciones tratadas en este trabajo.

También se utilizó una medida, que en Power BI es como realizar una función que nos permite seleccionar datos mediante una sentencia DAX presente en esta herramienta, y esto se llevó al elemento gráfico del mapa para darle un valor agregado adicional a este y que se pudieran realizar más operaciones para los análisis.

INFORMÁTICA Y SISTEMAS
REVISTA DE TECNOLOGÍAS DE LA INFORMÁTICA Y LAS TELECOMUNICACIONES
Vol. 5 Núm. 2 (27-29 Octubre, 2021), pp. 44-58 https://doi.org/10.33936/isrtic.v5i2.3946 | ISSN 2550-6730

Tabla 1: Resumen de los elementos gráficos utilizados.

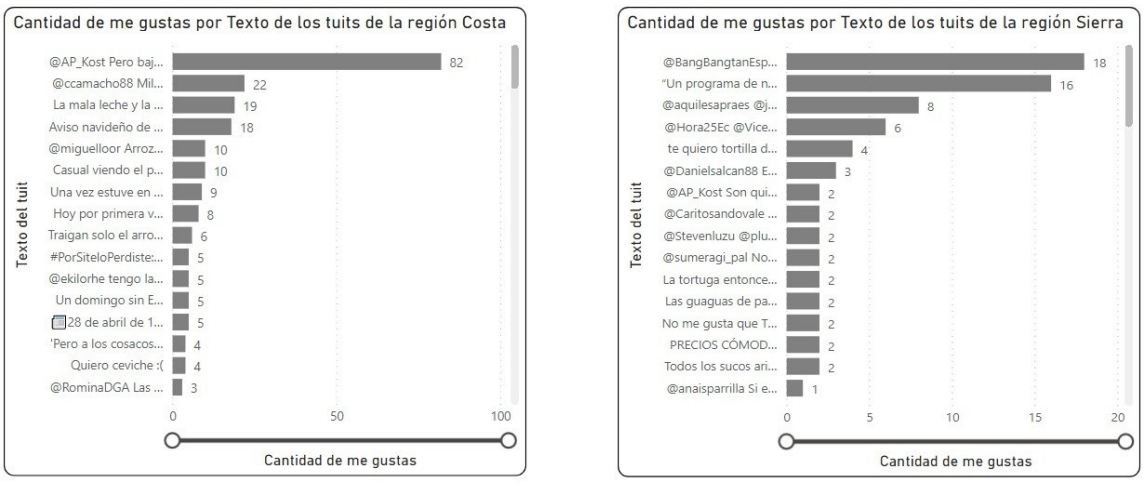
N°	Elemento gráfico	Descripción
1	Gráfico de barras ampliado	Muestra un recuento de la cantidad de me gustas por cada tweet.
2	Segmentación de datos	Se utiliza para filtrar los datos para que se muestren solamente estos filtrados.
3	Matriz	Se utiliza para filtrar los datos para que se resalten estos datos filtrados.
4	Tabla	Muestra detalles adicionales de los tweets como las interacciones.
5	Mapa	Muestra un mapa del Ecuador con el recuento de tweets por localización.

3. Resultados

Los resultados a continuación que se han obtenido en esta investigación se basan en la explicación de los análisis a partir de las visualizaciones que se obtuvieron con la herramienta de Power BI. En esta herramienta se les conoce como reportes a las páginas que contienen un conjunto de elementos gráficos.

3.1. Gráfico de barras ampliado

El primer reporte comprende un conjunto de datos que abarca los tweets por regiones, teniendo entonces dos elementos gráficos de barras ampliados como se muestra en la Figura 2, el primer gráfico de barras ampliado contiene el recuento de la cantidad de me gustas por cada tweet de la región Costa y el segundo el de la Sierra.



(a) (b)

Figura 2: Gráficos de barras ampliadas con la cantidad de me gustas por texto de los tweets de la región Costa(2a) y Sierra(2b).

INFORMÁTICA Y SISTEMAS
REVISTA DE TECNOLOGÍAS DE LA INFORMÁTICA Y LAS TELECOMUNICACIONES

 https://doi.org/10.33936/isrtic.v5i2.3946 | ISSN 2550-6730

En estos elementos gráficos de la Figura 2, se pueden filtrar por un control de cantidad de me gustas en donde se mostrarían los tweets que estén dentro del rango del control especificado. Este control nos sirvió de utilidad para poder tomar en cuenta los tweets que tienen una cantidad de me gustas por encima de alguna de las medidas de tendencia central de interés que se calcularon para ello.

Figura 3: Gráficos de barras ampliadas con la cantidad de me gustas por texto de los tuits de la región Sierra con un control que filtra el rango de los me gustas.

La Figura 3 contiene el gráfico de barras ampliadas de la región Sierra en donde se utiliza un control de me gustas que va de un rango entre 5 a 10 me gustas. Este control se utilizó para poder llegar a ciertos tweets que se consideran como no tan populares o que no recibieron altas cantidades de interacciones. Con esto se ejemplifica que tenemos un subconjunto de datos que entra en una de las categorías de popularidades que son media, alta y baja. Esta cantidad de me gustas usada para el control, se consideró como una popularidad media y haciendo un análisis de estos tweets, se obtuvo el resultado de la Tabla 2 que muestra la cantidad de tweets por cada etiqueta que se usó para la extracción de datos.

Tabla 2: Cantidad de tweets del control de rango por etiqueta de la extracción de datos en la región Sierra.

N°	Etiqueta	Cantidad de tweets
1	Saludable	6
2	No saludable	0
3	Medio	5
4	Vegetables	1

Este resultado se replica para el conjunto de datos de la Costa y se puede manejar una comparativa entre ambos subconjuntos de cada región como se muestra en la Tabla 3.

Obteniendo ambos resultados para realizar la comparativa, se observa que en la Sierra existe una mayor mención sobre etiquetas saludables que en la Costa, considerando así que en la Sierra hay más personas hablando sobre algún alimento de los que se han considerado como alimento saludable, sin embargo, el resto de etiquetas conservan un valor similar en ambas regiones.

INFORMÁTICA Y SISTEMAS
REVISTA DE TECNOLOGÍAS DE LA INFORMÁTICA Y LAS TELECOMUNICACIONES
Vol. 5 Núm. 2 (27-29 Octubre, 2021), pp. 44-58 https://doi.org/10.33936/isrtic.v5i2.3946 | ISSN 2550-6730

Tabla 3: Comparativa de la cantidad de tweets del control de rango por etiqueta de la extracción de datos en la región Costa y Sierra.

N°	Etiqueta	Cantidad de tweets Sierra	Cantidad de tweets Costa
1	Saludable	6	3
2	No saludable	0	0
3	Medio	5	5
4	Vegetables	1	1

3.2. Segmentación de datos

El primer reporte comprende varios elementos que se utilizaron como segmentación de datos, estos elementos están sincronizados con el resto de elementos de tal manera que permiten filtrar los tweets que se obtuvieron en los gráficos de barras ampliados. La Figura 4 muestra dos elementos gráficos usados para la segmentación de datos de los tweets de la región Costa.

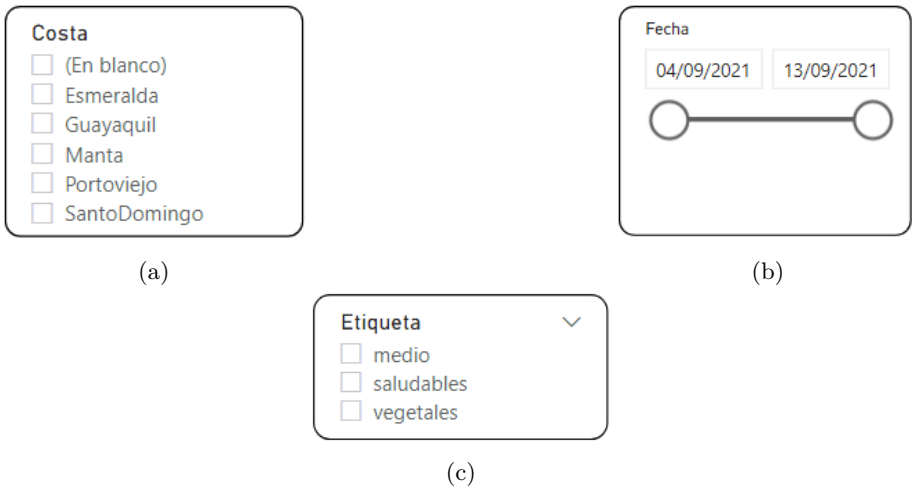


Figura 4: Segmentación de datos de los tweets de la región Costa. (4a) Ciudades de la región Costa; (4b) Fecha de los tweets de la región Costa; (4c) Etiquetas de los tweets de la región Costa.

En estas segmentaciones de datos, en la imagen A de la Figura 4 se muestran todas las ciudades utilizadas para la extracción de datos en la región Costa. Estos nombres de ciudades se encuentran dentro de una tabla relacionada al conjunto de datos de la Costa ya que con esto se evita, que no se muestren aquellas ciudades de donde no se pudo extraer al menos un tweet, como sucede con el elemento gráfico de la segmentación de datos de etiquetas, aquí este no tiene la etiqueta de alimentos no saludables debido a que no se encontró un solo tweet.

Los elementos gráficos se utilizaron para mostrar un filtrado exacto de los tweets por ciudad y poder realizar análisis distintos para cada una. Al igual que con cada tipo de etiqueta, como se muestra en la imagen B de la Figura 4 y las fechas en la que cada tweet fue publicado en un rango específico como se muestra en la imagen C de la Figura 4.

En la Tabla 4 se muestra un filtrado de los datos de la ciudad de Santo Domingo, en el rango de fecha del 4 de septiembre del 2021 al 13 de septiembre del 2021.

https://doi.org/10.33936/isrtic.v5i2.3946 | ISSN 2550-6730

Tabla 4: Cantidad de tweets por etiqueta de la región Costa con segmentación de datos en la ciudad de Santo Domingo y fechas del 04/09/2021 al 13/09/2021.

N°	Etiqueta	Cantidad de tweets Santo Domingo
1	Saludable	282
2	No saludable	0
3	Medio	147
4	Vegetables	65

Este resultado se puede replicar ya sea en otra ciudad de la misma región Costa o con otra región de la región Sierra. En la Tabla 5 se comparan los resultados de dos ciudades que se encuentran cercanas como son la ciudad de Baños, de la región Sierra y la ciudad de Santo Domingo, de la región Costa.

Tabla 5: Comparativa de cantidad de tweets por etiqueta de la región Costa y Sierra con segmentación de datos en la ciudad de Santo Domingo y Baños con fechas del 04/09/2021 al 13/09/2021.

N°	Etiqueta	Cantidad de tweets Santo Domingo	Cantidad de tweets Baños
1	Saludable	282	192
2	No saludable	0	0
3	Medio	147	72
4	Vegetables	65	38

Anteriormente se obtuvo un resultado en donde los datos con valores más altos correspondían a la región Sierra. Sin embargo, para esa comparativa se utilizaron los tweets extraídos del mismo día. Ahora se utilizó un rango de fechas gracias al elemento gráfico de segmentación de datos y con eso se logró filtrar más datos como lo reflejan los valores de la Tabla 5, aquí observamos que a pesar de que sean ciudades cercanas hay una gran diferencia en la cantidad de tweets que se han publicado por parte de los usuarios de la red social analizada. Los tweets sobre una alimentación saludable en la ciudad de Baños están siendo superados en casi 100 tweets por la ciudad de Santo Domingo, y en el resto de etiquetas están siendo superados por casi el doble, sin tomar en cuenta aquellas alimentaciones no saludables ya que en ambos casos no existen ningún tweet que haya sido extraído.

3.3. Matriz

Las matrices en Power BI son utilizadas para que filas y columnas se puedan mostrar como una tabla, pero para esta investigación se utilizaron para que se puedan seleccionar valores y estos nos ayuden a resaltar otros valores en el resto de elementos gráficos, ya que como se mencionó anteriormente, los elementos gráficos en un reporte pueden sincronizarse ya sea por las relaciones que existen en las tablas o por seleccionar campos de una misma tabla.

También se pude utilizar como una tabla, pero las mismas se utilizaron para otra funcionalidad. Las matrices nos aportan al análisis de datos cuando se quiere visualizar de manera resaltada ciertos datos sin tener que quitar los otros. Por ejemplo, en la Figura 3 se pudo contar de manera rápida la cantidad de tweets por etiqueta ya que aquí no existía una cantidad de datos tan extensa. La Figura 5 muestra un ejemplo de la utilización de una matriz.

La limitación que se reflejó, es que no se puede visualizar el resaltado cuando la cantidad de me gustas en un tweet es igual a cero, por lo que no se puede utilizar este método para aquellos tweets que se considerarían como bajos en popularidad. Por eso, es que en la Figura 3 se usaron tweets de popularidad media, ya que en esta categoría no se tenían cantidades de me gustas iguales a cero en los tweets.

INFORMÁTICA Y SISTEMAS
REVISTA DE TECNOLOGÍAS DE LA INFORMÁTICA Y LAS TELECOMUNICACIONES
Vol. 5 Núm. 2 (27-29 Octubre, 2021), pp. 44-58 https://doi.org/10.33936/isrtic.v5i2.3946 | ISSN 2550-6730

(a)

(b)

Figura 5: (5a) Matriz de etiquetas; (5b) Gráficos de barras ampliadas con la cantidad de me gustas por texto de los tuits de la región Costa con un resaltado en los tweets de la etiqueta saludables.

3.4. Tablas

Las tablas no son tan diferentes de las matrices ya que aquí solo se cuenta con un solo eje. Para esta investigación se usó una tabla que sume aquellos valores enteros de la cantidad de retweets, cantidad de seguidores, cantidad de seguidos, cantidad de tweets y cantidad de me gustas. Estas cantidades son extraídas individualmente por cada tweet extraído, pero en Power BI se muestra la sumatoria total. Para el análisis, se usaron las tablas al momento de querer saber más a fondo las interacciones de los tweets extraídos ya que hay casos en los que una clasificación de popularidad para un tweet se haya dado de manera desigual si se comparan los tweets en general de ese usuario autor del tweet extraído.

Las tablas utilizadas en el primer reporte pueden mostrar la cantidad de interacciones de un solo tweet si se selecciona alguno desde el gráfico de barra amplificado como se muestra en la Figura 6.

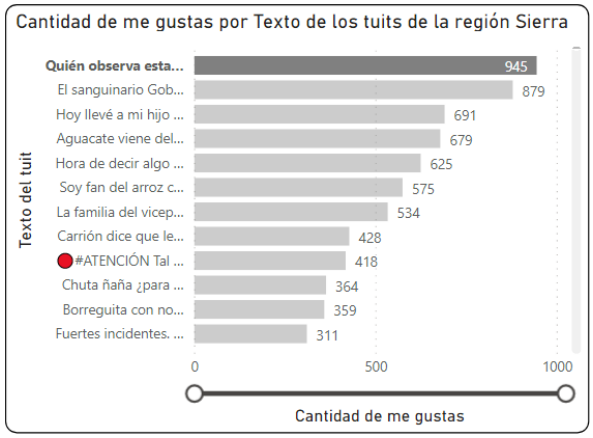
Analizando la Figura 6, se puede concluir que este usuario se encuentra entre el rango de un usuario promedio que se dedica a realizar muchas publicaciones en la red social analizada. Este tweet contiene la etiqueta de tipos de alimentos medios por hacer mención al encebollado, pero si se busca otro tweet que contenga la etiqueta de alimentos medios, se puede obtener una comparativa entre ambos resultados para determinar un análisis más representativo para este elemento gráfico de tabla.

La Figura 7 muestra una replicación del proceso de la Figura 6 pero con un tweet seleccionado de la misma etiqueta de alimentos medios en el mismo conjunto de datos de la Sierra, sin importar la fecha ni la ciudad en donde se realizó la publicación del tweet.

Analizando la Figura 7 con un enfoque hacia la comparativa con la Figura 6, se obtienen los resultados de que en la imagen A de la Figura 7, el tweet seleccionado ha sido etiquetado con la etiqueta de alimentos medios por hacer mención al aguacate. Estos usuarios de Twitter tienen grandes cifras de diferencias en lo que respecta a los seguidores de cada una de ellos, y el que tiene menos seguidores que el otro, es el que tiene el tweet con más cantidad de me gustas. Para este caso puede que la cantidad de tweets que se publican tengan alguna relevancia, pero eso solo podría determinarse haciendo más análisis y más comparativas para obtener así varios resultados. Esta información fue utilizada para esta investigación para los momentos en los que se quiera entender que la popularidad de un tweet, repercuten en las interacciones del mismo y que estas interacciones vienen o son provocadas por factores que dependen netamente del autor de cada tweet.

INFORMÁTICA Y SISTEMAS
REVISTA DE TECNOLOGÍAS DE LA INFORMÁTICA Y LAS TELECOMUNICACIONES

 https://doi.org/10.33936/isrtic.v5i2.3946 | ISSN 2550-6730

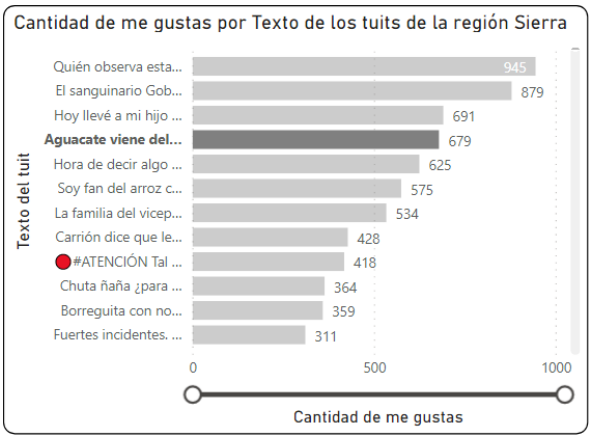


(a)

Cant_Retuits	Cant_Seguidores	Cant_Seguidos	Cant_Tuits	Cantidad de me gustas
124	3803	2483	20562	945

(b)

Figura 6: (6a) Gráficos de barras ampliadas con la cantidad de me gustas por texto de los tweets de la región Sierra con un resaltado en el tuit con más me gustas; (6b) Tabla de interacciones que corresponde al tweet seleccionado.



(a)

Cant_Retuits	Cant_Seguidores	Cant_Seguidos	Cant_Tuits	Cantidad de me gustas
39	40983	483	1036	679

(b)

Figura 7: (7a) Gráficos de barras ampliadas con la cantidad de me gustas por texto de los tuits de la región Sierra con un resaltado en un tweet con la etiqueta de alimentos medios; (7b) Tabla de interacciones que corresponde al tweet seleccionado.

Entonces, las alimentaciones que se pueden dar por regiones pueden también incluir un análisis no solo de la cantidad de veces que los tweets han sido etiquetados con algún tipo de alimentación, si no, con la cantidad de interacciones que se pueden encontrar. Un análisis podría darse entonces en la cantidad

INFORMÁTICA Y SISTEMAS
REVISTA DE TECNOLOGÍAS DE LA INFORMÁTICA Y LAS TELECOMUNICACIONES
Vol. 5 Núm. 2 (27-29 Octubre, 2021), pp. 44-58 https://doi.org/10.33936/isrtic.v5i2.3946 | ISSN 2550-6730

de interacciones que se pueden encontrar en varias ciudades sobre un tipo de alimentación, en específico para lograr determinar si las interacciones pueden decirnos algo más que pueda servirnos para el análisis.

En la Figura 8 se muestra una comparativa entre la ciudad de Manta que pertenece a la región Costa y la ciudad de Cuenca que pertenece a la región Sierra, aquí se filtran solamente los datos que pertenecen o están etiquetados con la etiqueta de alimentos saludables.

Cant_Retuits	Cant_Seguidores	Cant_Seguidos	Cant_Tuits	Cantidad de me gustas
4	201518	12860	469703	14

(a)

Cant_Retuits	Cant_Seguidores	Cant_Seguidos	Cant_Tuits	Cantidad de me gustas
10	277751	9183	443854	39

(b)

Figura 8: (8a)Tabla de interacciones que corresponde a la sumatoria de los tweets de la ciudad de Manta de la región Costa con la etiqueta saludable; (8b) Tabla de interacciones que corresponde a la sumatoria de los tweets de la ciudad de Cuenca de la región Sierra con la etiqueta saludable.

Los tweets seleccionados en la ciudad de Cuenca son de un total de 11 para la fecha del 10 de septiembre del 2021, mientras que los tweets seleccionados en la ciudad de Manta son de un total de 13 para la fecha del 11 de septiembre del 2021. Se escogieron estas fechas ya que aquí no existe un tweet muy popular que haya recibido una cantidad de interacciones demasiado alta del promedio. Además, la diferencia de tweets publicados no es tan diferente, al igual que las otras cantidades.

Entonces, analizando esta comparativa, podemos deducir que en la ciudad de Cuenca ha habido más interacciones con los tweets con lo que respecta a la sumatoria de cantidad de me gustas y la sumatoria de cantidad de retweets. Y con los autores de estos tweets, se deduce que los usuarios de Cuenca tienen más popularidad que los de la ciudad de Manta.

Concluyendo que la ciudad de Cuenca a comparación con la ciudad de Manta, tiene más interacciones con los tweets que fueron etiquetados con la etiqueta de alimentación saludable, aunque las diferencias no son tan elevadas, por lo que se puede concluir también de que en ambas ciudades se tiene un grado igualitario sobre popularidad en alimentaciones saludables.

3.5. Mapa

En esta investigación se utilizó un elemento gráfico de mapa para un segundo reporte ya que se contaba con geolocalizaciones con latitud y altitud, las cuales se pueden ingresar en el mapa para que Power BI reconozca la ubicación. Aquí se especificó un conjunto de datos en donde se unificarán los datos de la región Costa y Sierra, además, se usaron las mismas etiquetas y la cantidad de tweets publicados como el valor que será representado en la gráfica de pastel que se posiciona sobre cada coordenada que se ingresa en el mapa.

La Figura 9 muestra el elemento gráfico del mapa en donde se ha ubicado un gráfico de pastel que se divide en las etiquetas y estas muestran el recuento de tweets obtenidos por etiquetas.

En este segundo reporte también se incluyeron segmentaciones de datos para poder filtrar los que se visualizan en el elemento gráfico del mapa. La Figura 10 muestra las segmentaciones de datos que se usaron para este reporte y que luego será utilizado para los posteriores análisis.

INFORMÁTICA Y SISTEMAS
REVISTA DE TECNOLOGÍAS DE LA INFORMÁTICA Y LAS TELECOMUNICACIONES

https://doi.org/10.33936/isrtic.v5i2.3946 | ISSN 2550-6730

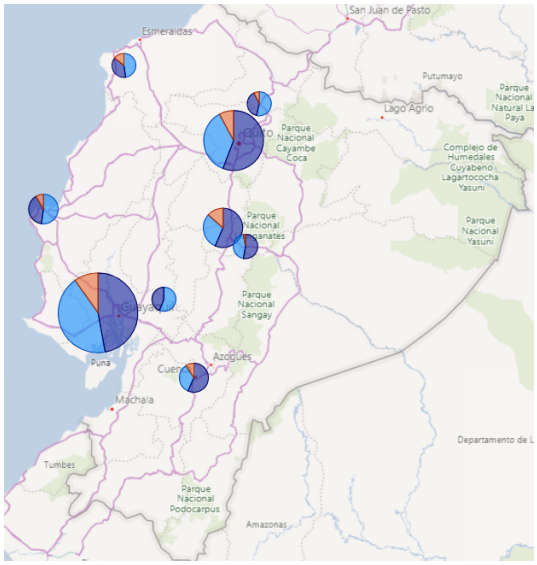


Figura 9: Mapa del Ecuador con las geolocalizaciones de los tuits y un recuento de los tweets por etiqueta.

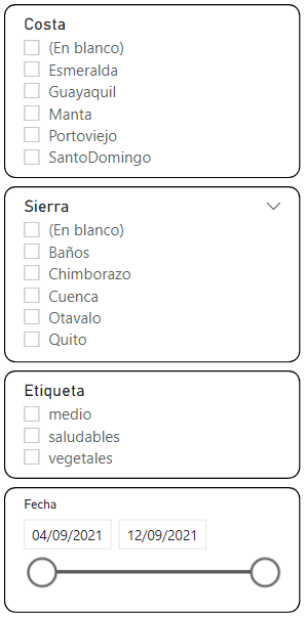


Figura 10: Segmentaciones de datos del segundo reporte.

Las segmentaciones de datos son por ciudades de la región Costa y ciudades de la región Costa, la etiqueta del tipo de alimentación y la fecha de publicación del tweet.

Anteriormente en el primer reporte ya se han realizado varias formas de analizar estos datos, sin embargo, ahora se cuenta con una visualización más amigable para situarse en un lugar específico como se puede observar en la Figura 11, donde se selecciona la ciudad de Quito y se revisa la cantidad de tweets con la etiqueta saludable en un rango de fecha del 4 de septiembre del 2021 al 2 de septiembre del 2021.

INFORMÁTICA Y SISTEMAS
REVISTA DE TECNOLOGÍAS DE LA INFORMÁTICA Y LAS TELECOMUNICACIONES
Vol. 5 Núm. 2 (27-29 Octubre, 2021), pp. 44-58 https://doi.org/10.33936/isrtic.v5i2.3946 | ISSN 2550-6730

Figura 11: Mapa del Ecuador con las geolocalizaciones de los tuits y un recuento de los tweets por etiqueta con la ciudad de Quito seleccionada.

El resultado obtenido no es tan diferente a lo que ya se ha estado analizando anteriormente, solamente se obtiene de manera diferente la visualización. Por eso es que aquí se creó la medida, que no es más que una expresión escrita en DAX la cual se muestra en la Figura 12, donde la palabra "Busqueda_Palabra" se reemplaza por las palabras de interés que se desea buscar.

```
Medida = COUNTROWS (
    FILTER ( Dataset_All,
        CONTAINSSTRING(
            Dataset_All[Texto_Procesado],
            "Busqueda_Palabra"
        )
    )
)
```

Figura 12: Expresión DAX para buscar palabras dentro del texto procesado.

Para el siguiente análisis, se realizó la búsqueda de la palabra canguil y chifle ya que estos dos alimentos son utilizados para hacer una mezcla con uno de los platos típicos del Ecuador que es el encebollado. Esta comida es conocida a nivel nacional y ha sido colocada en la categoría de alimentos medios. Con la ayuda de la expresión DAX se realizaron las Tablas 6 y 7 que muestran el recuento de tweets que se encontraron por cada región para así obtener un recuento total.

Tabla 6: Tabla con el recuento de tweets de las ciudades de la región Costa con la expresión DAX.

N°	Ciudad	Chifle	Canguil
1	Esmeralda	0	0
2	Guayaquil	24	12
3	Manta	3	0
4	Portoviejo	0	0
5	Santo Domingo	3	1
	Total	**30**	**13**

https://doi.org/10.33936/isrtic.v5i2.3946 | ISSN 2550-6730

Tabla 7: Tabla con el recuento de tweets de las ciudades de la región Sierra con la expresión DAX.

N°	Ciudad	Chifle	Canguil
1	Baños	0	0
2	Chimborazo	1	0
3	Cuenca	0	0
4	Otavalo	0	0
5	Quito	6	4
	Total	**7**	**4**

Analizando los resultados, tenemos que en la ciudad de Guayaquil se presentan los valores más altos para los diferentes casos. Por lo general se tenía en cuenta de que la región Sierra es el lugar de donde nace la mezcla del encebollado con el canguil, sin embargo, las cifras encontradas del canguil en esta región son inferiores al chifle de esta misma región, incluso se ve superada con las cifras de la región Costa por cantidades elevadas. Entonces podemos deducir que el originario del encebollado con chifle proviene más de los sectores que comprenden la ciudad de Guayaquil de la región Costa y que también es el que más mención hace sobre el canguil.

El encebollado es una comida que se consideró como un alimento medio porque sus ingredientes no son solo carbohidratos, pues también se incluye el pescado el cual es una fuente de proteínas alta y que también incluye ingredientes como cebollas, yucas y cilantros que son verduras que contienen vitaminas, minerales y otros componentes. Pero si algo es cierto, es que el chifle tiene un grado calórico más alto que el canguil, por lo que, si se utiliza la misma cantidad de chifle o canguil para mezclar el encebollado, se puede deducir que aquel que se mezcle con chifle será más calórico que aquel que se mezcle con canguil.

4. Discusión

El conseguir generar conclusiones relacionadas a la salud de las personas en base a sus hábitos alimenticios, haciendo uso de las redes sociales, resulta muy práctico debido a que se cuenta con información actual que viene dada por las tendencias de consumos de alimentos por parte de las personas, información que se puede aplicar para entender el origen de ciertas enfermedades e incluso anticiparlas.

El hacer uso de las redes sociales como medio para obtener información presenta ciertas desventajas relacionadas a la forma en que las personas expresan un hecho, ya que muchas veces se emplean palabras bajo un contexto totalmente diferente a su significado. Y, debido a que la búsqueda de tweets muestra aquellas publicaciones que cuentan con palabras previamente especificadas, se obtiene un gran conjunto de publicaciones que tienen poca o nula relación con el interés de estudio, sin embargo, el poder visualizar el contenido del tweet permite, aunque de forma manual, no tener en cuenta estas publicaciones y solo enfocarse en aquellas que si presentan el contenido que se requiere.

La presente investigación permite clasificar tipos de alimentos más relevantes en las categorías saludables, no saludables, medios y vegetales y su en regiones costa y Sierra de Ecuador, además analizar un tipo de alimento especifico ver como se manifiesta en mayor o menor medida en diferentes regiones y ciudades, lo que permite replicar estos resultados a otros tipos de análisis.

Referencias

[1] Fernando Mendoza. «Expertos advierten posibles efectos nocivos en Ecuador a causa de la mala alimentación». En: *El Telégrafo Ecuador* (oct. de 2020).

INFORMÁTICA Y SISTEMAS
REVISTA DE TECNOLOGÍAS DE LA INFORMÁTICA Y LAS TELECOMUNICACIONES

Vol. 5 Núm. 2 (27-29 Octubre, 2021), pp. 44-58 https://doi.org/10.33936/isrtic.v5i2.3946 | ISSN 2550-6730

[2] Adriana Amaya-Hernández, Mayaro Ortega-Luyando y Juan M Mancilla-Dıéaz. «Cómo, qué y por qué ocuparnos de la alimentación». En: *Journal of Behavior and Feeding* 1.1 (2021), págs. 51-59.

[3] María Antonieta Chacha Cha. «¿La publicidad de alimentos incide en el consumo?» En: *El Universo | La Revista | Salud* (oct. de 2020).

[4] Cepal y Pma. «Estudio revela mala calidad de alimentación en Ecuador». En: *Edición Méédica | Salud Pública* (mayo de 2017).

[5] Julián Villodre, Anne Marie Reynaers y J Ignacio Criado. «Transparencia externa y redes sociales. Los roles diferenciales de ministerios y organismos públicos estatales en Twitter». En: *Revista de estudios polıéticos* 192 (2021), págs. 191-220.

[6] Daniel Revillo Rey. «Infraestructura software para el análisis de las interacciones en Twitter: aplicación a la detección de trastornos de conducta alimentaria». En: *Universidad de Zaragoza* (2020).

[7] Laura Mıéguez Fernández. «Influencia de las redes sociales en la alimentación saludable». En: *universidad pontificia comillas* (2019).

[8] *Tweepy.*

[9] S. Niveditha Jayesh Chaudhary. «Twitter Sentiment Analysis using Tweepy». En: *International Research Journal of Engineering and Technology (IRJET)* 8.4 (2021).

[10] Tweepy. *Search Methods.*

[11] Viorel Negrut. «Power bi: Effective data aggregation». En: *Quaestus* 13 (2018), págs. 146-152.

[12] Tatiana Blanco Rojas, Diana Milena Archila Córdoba y Javier Antonio Ballesteros-Ricaurte. «Gestión de datos obtenidos desde redes sociales aplicando Business Intelligence Engineering Process». En: *Revista Virtual Universidad Católica del Norte* 49 (2016), págs. 72-91.

[13] Bernardino Meseguer Barrionuevo & et al. «El business intelligence en las PYMES: herramienta power BI». En: (2016), págs. 1-12.