

OFERTAS DE SERVICIOS POSPAGO DE TELEFONIA MOVIL DE LOS PRINCIPALES COMPETIDORES DE COLOMBIA

WEB SCRAPPING

Tipología y Ciclo de Vida de los Datos
Universidad Oberta de Catalunya

Contexto

Se ha elegido como tema de investigación para este proyecto, el negocio de las telecomunicaciones en Colombia. En concreto, se han elegido los dos principales competidores para tal fin, Tigo y Movistar.

En el negocio de las telecomunicaciones se debe revisar con frecuencia los cambios de oferta de los competidores, con el fin de poderse adaptar al mercado con la mayor celeridad posible evitando así una pérdida de clientes y generando interés en potenciales clientes. Tal es la importancia de generar una nueva oferta, que se invierten millones de euros en publicidad al año (televisión, internet, eventos, etc), denotando así, la importancia que tiene este trabajo de análisis.

Estos cambios de oferta pueden ser la llegada de un nuevo jugador al mercado, campañas promocionales y servicios convergentes

El Web Scraping, por tanto, resulta muy útil en este contexto, ya que, mediante herramientas automatizadas se puede reaccionar rápidamente a un cambio en el mercado, detectando la oferta y planteando las acciones de negocio a las que haya lugar para no perder e incluso aumentar el market share.

Título

Ofertas de servicios postpago de telefonía móvil de los principales competidores de Colombia (Tigo y Movistar)

Descripción del dataset

Este dataset está basado en los datos a tener en cuenta para la evaluación de la competencia o para la evaluación de tarifas de consumir basándose en las ofertas postpago publicadas en la web. En concreto, los datasets analizados pertenecen a las dos compañías mas grandes de telecomunicaciones de Colombia, Tigo y Movistar. Los datos no han pasado por un proceso ETL estricto, por lo que antes de su análisis será necesario que pasen por un proceso de ETL. Por ejemplo, los beneficios se encuentran en distintas columnas y el Precio se encuentra en dos columnas distintas, son ejemplos que evidencia la necesidad de un proceso de tratamiento de datos.

Las fuentes del origen de datos pueden encontrarse en las siguientes URLs:

- Tigo. compras.tigo.com.co/movil/pospago
- Movistar. movistar.co/pospago

La descarga además se muestra sin cabecera, por lo que en el momento de análisis será necesario identificarlas.

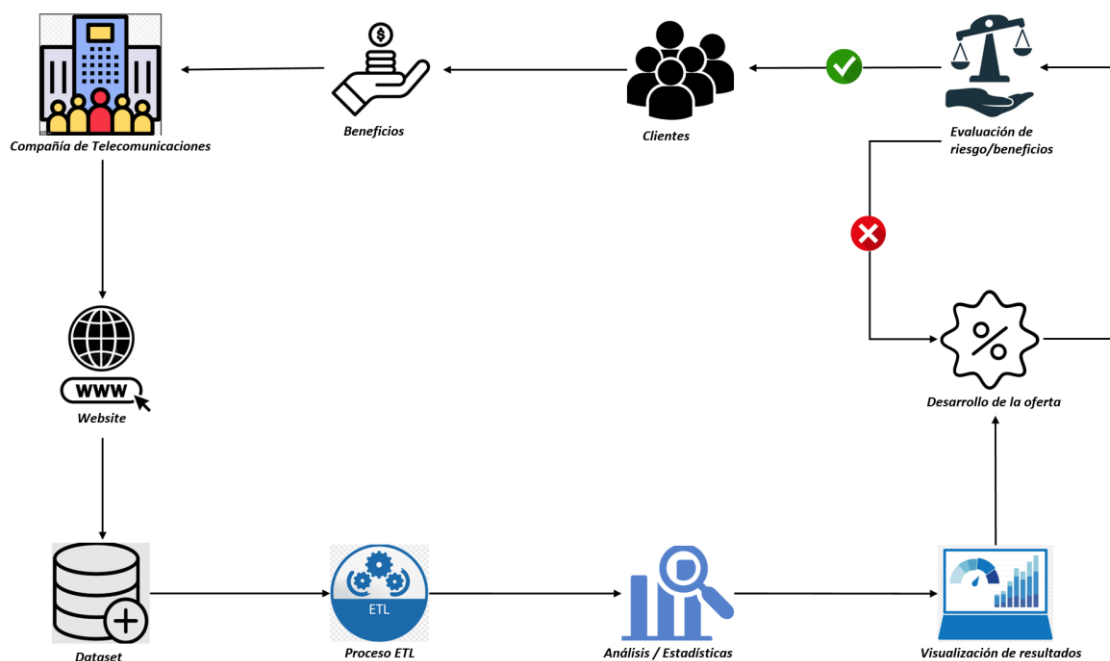
El dataset extraído de Tigo se compone de XXXX registros y el dataset de Movistar de XXXX registros.

A continuación, se describen los datos extraídos de Tigo:

Atributo	Tipo de dato	Ejemplo
Precio	String	35.000COP
IVA	String	IVA incluido
Gb	String	15Gb
Beneficios 1	String	3 meses de cortesía de Amazon Prime Video después \$17.900 al mes
Servicios	String	Amazon Prime Video,
Tipos de Beneficios	String	3 meses de cortesía
Modos de contratación	String	portación,línea nueva,pre2pos,
Fecha de la oferta	Date(MM/DD/YYYY HH:MM:SS)	11/04/2021, 22:43:58

Representación gráfica

Aunque existen diversos enfoques desde donde se puede realizar la representación gráfica del dataset con los principales agentes implicados, se ha propuesto el enfoque que se considera más relevante, que es aquel que tiene que ver con la empresa de telecomunicaciones y su adaptación al mercado mediante una contraoferta. Los iconos azules representan la parte correspondiente al análisis de datos.



Contenido

Este dataset contiene los datos de los distintos planes de las ofertas publicadas en las webs de las empresas de telecomunicaciones

Las ofertas analizadas corresponden al mes de noviembre de 2021.

Los datos fueron capturados empleando Web Scrapping en lenguaje Python sobre cada una de las webs para finalmente guardar los resultados en formato CSV.

Los campos que incluye el dataset son:

- **Precio.** Precio del plan pospago en miles y en COP
- **Impuestos.** Impuesto incluido o no en el precio final (campo "Precio")
- **Datos de navegación.** Cantidad de datos en GB del plan para navegar en internet
- **Beneficios (texto).** Texto de beneficios del plan sin procesar
- **Servicios.** Servicios incluidos en el plan separados por coma
- **Tipos de Beneficios.** Tipos incluidos en el plan separados por coma
- **Modos de contratación.** Opciones de compra del plan separadas por coma
- **Fecha de la oferta.** Fecha UTC en la que se obtienen los datos de la oferta

Agradecimientos

En sitios como kaggle y data.world se encuentran múltiples dataset relacionados al reconocimiento, clasificación y comportamiento del cliente (p. ej. tasa de cancelación o "churn"). En la búsqueda efectuada no se encontró ningún dataset que contenga las ofertas y detalles de planes de telefonía móvil.

Consideramos que es una oportunidad importante capturar históricamente (series de tiempo) los datos de las diferentes ofertas del mercado para ser analizados mediante técnicas de estadística y data mining.

Los datos contenidos en el dataset son totalmente públicos y corresponden a la oferta Above The Line (ATL), por tanto, no se infringe ningún marco normativo, ni principio ético.

Inspiración

El interés en analizar este conjunto de datos reside la enorme utilidad encontrada al propósito de este proyecto. Comparando las diferentes ofertas de las empresas de telecomunicaciones líderes en el país de forma automática y con su respectiva marca de tiempo, podemos estar informados de las evoluciones del mercado en tiempo real.

Esta información es útil tanto para una gran empresa, la cual puede elaborar estrategias de marketing para ser competidores en el mercado, como para, el pequeño consumidor, que siempre podrá estar informado de las ofertas del mercado y elegir en consecuencia la que mejor se adapte a sus necesidades al mejor precio, con el consecuente ahorro de tiempo y dinero.

Inicialmente se obtienen los datos de una sola fecha, pero puede programarse el script de web scraping para que se ejecute recurrentemente y cuando detecte un cambio inserte nuevos registros con la respectiva fecha creando una serie de tiempo que puede ser analizada desde diferentes perspectivas.

Se pueden responder preguntas como:

- ¿Cuál es el cambio de Gb de los planes en el tiempo?
- ¿Cómo cambian los beneficios de los planes en el tiempo?
- ¿Cómo cambian las formas de compra de los planes en el tiempo?
- ¿Cómo cambian los precios e impuestos de los planes en el tiempo?
- ¿Para una determinada fecha cómo cambio la oferta ATL en el mercado?
- ¿Qué relaciones tiene la competencia con otras marcas para sus promociones (Netflix, Whatsupp, etc) ?
- ¿Cuáles son las tendencias del mercado?
- ¿Qué compañía es claramente diferenciadora?

Licencia

Una gran cantidad de datos pueden ser compartidos y explotados de múltiples formas irrestrictas (pero éticas), para lograr avances, conclusiones y mejoras a diferentes niveles. El dato persistido (y compartido) es el fundamento del conocimiento de la raza humana.

Creemos en la filosofía Open Data.

Este dataset se expone bajo licencia CC0 1.0 Public Domain Dedication.

Código

El código fuente de la solución está en [repositorio GitHub](#)

Dataset

Los dataset pueden encontrarse dentro de la carpeta Dataset

Firmas

Contribuciones	Firma
Investigación previa	A.F. A. R. Á. O. H.
Redacción de las respuestas	A.F. A. R. Á. O. H.
Desarrollo del código	A.F. A. R. Á. O. H.