

Visión General: Modelo Predictivo de Métricas Stryd mediante Machine Learning

David Víctor Gómez Ramírez Técnico Superior en Desarrollo de Aplicaciones Multiplataforma especialidad en Big Data

Contenido

١	TRODUCCIÓN	1
	STRYD	
	OTRAS MÉTRICAS	4
	FUENTE DE DATOS	5
	PROCESOS DE ETL	8
	m_SEGMENTOS_TXT_TO_TABLE	9
	EJECUCIÓN: m_SEGMENTOS_TXT_TO_TABLE	11
	m_SEGMENTOS_TABLE_TO_TIPOOK	12
	EJECUCIÓN: m_SEGMENTOS_TABLE_TO_TIPOOK	14
	m_SEGMENTOS_RATIOS	15
	EJECUCIÓN: m_SEGMENTOS_RATIOS	17
	m_SEGMENTOS	18
	EJECUCIÓN: m_SEGMENTOS	20
	FLIENTES:	21

INTRODUCCIÓN

El objeto de este proyecto es la realización de un modelo de Machine Learning para la predicción de una/s métrica/s proporcionadas por un dispositivo que mide la potencia que generamos cuando realizamos la actividad física de correr. Como podemos observar esta contextualización es bastante vaga e indefinida, pero como todo proyecto debemos empezar por algún punto.

STRYD

¿Qué es STRYD?

Stryd es un medidor de la potencia que generamos cuando practicamos "running" y, aunque en concepto es muy similar a los medidores de potencia del "ciclismo", tienen diferencias sustanciales ya que Stryd no mide la potencia de forma directa sino que utiliza un algoritmo para su cálculo.

Para este proyecto personal, Stryd no es más que una herramienta más, por lo que no ahondaremos en su funcionamiento pero sí en las métricas y datos que nos aporta y que serán la base de nuestro modelo predictivo, es decir, será la fuente principal de la que obtendremos los datos de nuestro modelo.

Por simplificar y ubicarnos mejor en el contexto diremos que Stryd es un dispositivo que mide diferentes valores cuando realizamos entrenamientos de "running" y cuyo eje central es la potencia medida en watios (W), pero tenemos muchas más y que serán muy importantes (o no) para nuestro modelo:

- *Tiempo (DURACION):* variable fundamental de cualquier entrenamiento y que mide la duración del mismo.
- Distancia (DISTANCIA): mide la longitud de los entrenamientos.
- Ritmo (RITMO): es la métrica por excelencia que relaciona "Tiempo" y "Distancia".



 Altimetría y Desnivel (ALTITUD - DESNIVEL): mediante mide la altitud y sus variaciones.



- Cadencia (CADENCIA): son las zancadas o pasos que damos mientras corremos y se mide en zancadas o pasos por minuto.
- TSC (Tiempo de contacto con el suelo): podríamos definirlo como la duración medida en ms en el que nuestro pie está en contacto con el suelo.



- Oscilación vertical (OSC_VERTICAL): es la distancia vertical que recorre el cuerpo desde el punto en que el centro de gravedad está más bajo hasta el punto más alto.
- Leg Spring Stiffness (LSS): se trataría del acrónimo inglés LSS (Leg Spring Stifness) que adaptado al castellano vendría a ser cómo la rigidez del resorte.



- Pendiente (PENDIENTE): mide el porcentaje de pendiente media durante la distancia recorrida.
- RSS (RSS): mide el estrés que ha generado el entrenamiento o segmento.
- Potencia (POTENCIA): mide la potencia media generada durante un entrenamiento o segmento.
- Frecuencia Cardiaca (FREC_CARDIACA): mide la frecuencia cardiaca por entrenamiento o segmento medidas en pulsaciones por minuto (ppm)
- Aire (AIRE): mide la resistencia de aire que encontramos en un entrenamiento o segmento.
- Form Power (FP): mide la potencia vertical y lateral que generamos cuando nos desplazamos en un entrenamiento o segmento.



 Segmento: es un concepto que deberemos tener muy presente y que no es más que pequeñas partes del entrenamiento de los que podemos extraer la misma información.



OTRAS MÉTRICAS

- RFP (Ratio Form Power): (lit. Eduard Barceló) el ratio del Form Power nos va a decir qué porcentaje de la potencia que estamos aplicando se va hacia arriba y no hacia adelante. Para utilizar este valor correctamente, tenemos que saber que:
 - ❖ A medida que la potencia bruta se incrementa, el ratio empeora.
 - ❖ A medida que se acumula la fatiga, el ratio empeora igualmente.
 - Si queremos compararnos entre corredores, hay que hacerlo a % del FTP iguales.

Finalmente como referencia para conocer el orden de magnitud del FPR hay que saber que al FTP y en condiciones de carrera llanas:

- ❖ Un valor superior al 25 % es para un corredor con mala técnica.
- Un valor entre el 23-25 % es la media para la mayoría de corredores populares.
- ❖ Un valor entre el 20-23 % es un valor muy bueno.
- Valores por debajo del 20 % los obtienen sólo corredores de élite de clase mundial.

<i>RFP</i> =	FP (FORM POWER)	
	POTENCIA (WATIOS)	

• RE (Running Efectiveness): (lit. Eduard Barceló) este parámetro es uno de los más interesantes que el dispositivo Stryd ofrece en relación a la técnica de carrera y cómo mejoramos nuestra efectividad.

La efectividad de carrera es la ratio entre la velocidad y la potencia. Se calcula mediante el cociente de la velocidad en metros/segundo por la potencia en vatios/kilogramo.

Dicho así suena complicado, muy complicado, pero el concepto es muy sencillo. Simplemente te indica a qué velocidad te permite correr 1 vatio/kg. Cuanta más velocidad puedas imprimir con ese vatio, mayor efectividad.

Por tanto, cuanto más grande sea el número, mejor técnica tendrá el atleta y esto le permitirá, a igualdad de umbral de potencia funcional y potencia relativa que otro atleta, un rendimiento superior a su homólogo pero con un índice inferior de efectividad de carrera.

	Metros / Segundos	Distancia / Duracion
<i>RE</i> =	=	
	146.45 - 74.45	D. (/ D 0 /
	Watios / kilogramos	Potencia/ Peso Strvd

 L_ZANCADA (Longitud de zancada): la L_ZANCADA mide la distancia medida en metros de cada zancada en segmentos específicos y para su obtención son necesarias diferentes métricas anteriormente vistas.

• ROV (Ratio Oscilación Vertical): ratio entre la oscilación vertical y la longitud de zancada, a pesar de ser una métrica interesante para este modelo su aportación será más bien nula.

Longitud de Zancada

RLSS (Ratio Leg Spring Stiffness): ratio entre LSS (muelle o resorte al correr) y Peso Stryd.

Peso Stryd

Notas Importante:

- Excepto la métrica RFP, que es aportada por la aplicación Stryd, el resto de métricas para su estudio deberán ser calculadas.
- Cuando hablamos de Peso Stryd, nos estamos refiriendo al peso que se introdujo por primera vez en el dispositivo y que no se deberá cambiar para poder comparar las métricas de forma correcta.

FUENTE DE DATOS

La herramienta Stryd dispone de su propia aplicación móvil y de su equivalente de escritorio en formato web, para poder extraer los datos de los entrenamientos deberemos descargarnos los archivos desde la aplicación móvil en formato csv. Por cada entrenamiento deberemos descargarnos 3 archivos aunque en principio sólo utilizaremos 2. Un archivo en los que las distancias de los segmentos se han definido manualmente y un segundo donde las distancia de los segmentos está predefinida a 1000m.

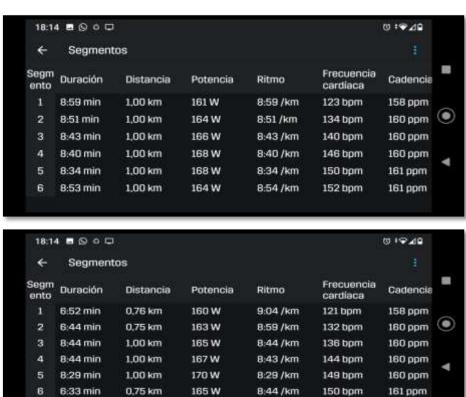
Manualmente aplicación web



Predefinidos en aplicación web







6:34 min

0.74 km

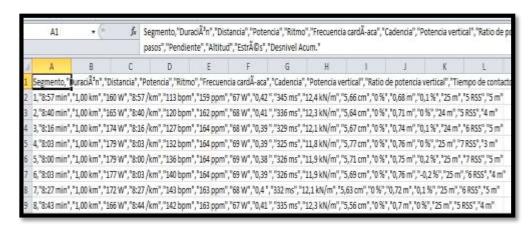
164 W

8:50 /km

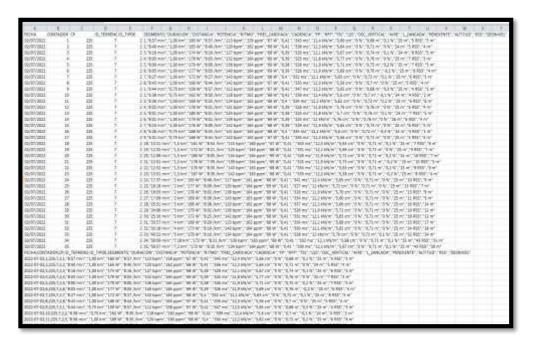
152 bpm

161 ppm

Mediante la herramienta Powercenter (no confundir con PowerCenter) de Stryd podremos obtener los datos de los segmentos que queramos y así poder obtener más información.







Tras unificar "SEGMENTOS1K" y "SEGMENTOS_MANUAL" realizaremos algunas transformaciones para adaptar los datos a nuestras necesidades en un archivo de texto plano "SEGMENTOS.txt"

```
Action: Edicine Tennate the Aquate State April 10 (1997) | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.0
```

PROCESOS DE ETL

Nos podríamos extender y podríamos debatir sobre qué herramienta es mejor o peor, cual nos gusta más o menos o es más apropiada o menos apropiada para las tareas de ETL. Particularmente y por su "sencillez" utilizaremos *Informatica PowerCenter* y como base de datos SQL SERVER sin despreciar otras posibles herramientas y soluciones.

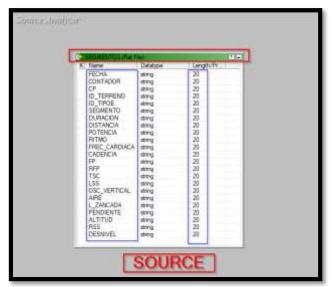
Deberemos crear diferentes "mappings" y otros tantos "workflows" con el objetivo de llevar nuestro archivo de texto plano "SEGMENTOS.txt" hasta la tabla "STG_SEGMENTOS", realizando conversiones de tipos, transformaciones, operaciones, filtros, etc...

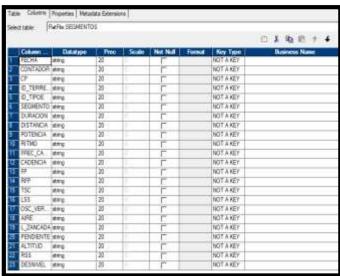
Básicamente partiremos de diferentes fuentes de datos (SOURCES), ya sean textos planos o tablas de la BBDD "DATAMART_SEGMENTOS" y realizaremos diferentes transformaciones y cálculos para llevar esos mismos datos a sus destinos (TARGETS).

m_SEGMENTOS_TXT_TO_TABLE

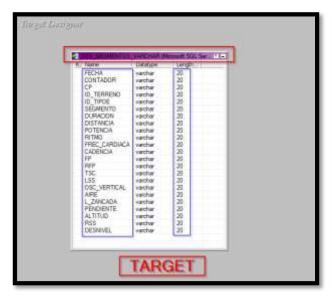
Mediante esta transformación, transportaremos nuestros almacenados en "SEGMENTOS.txt" a un formato más estructurado, es decir, formato tabla. Básicamente se trata de un mapeado directo desde la fuente (source) al destino (target), todos los campos se almacenarán en un primer momento en tipo varchar(20).

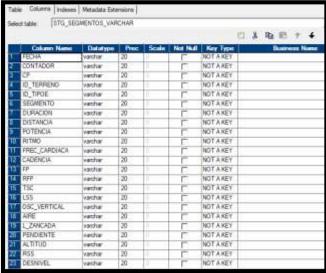
Source: SEGMENTOS.txt



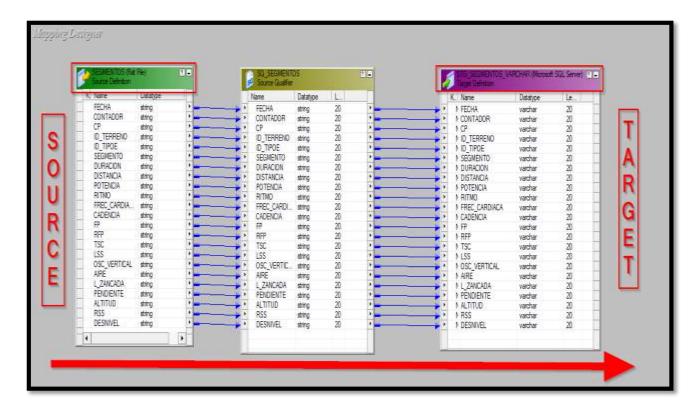


Target: STG_SEGMENTOS_VARCHAR

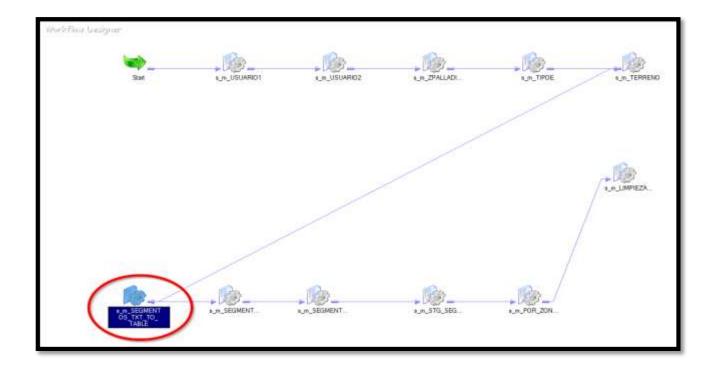




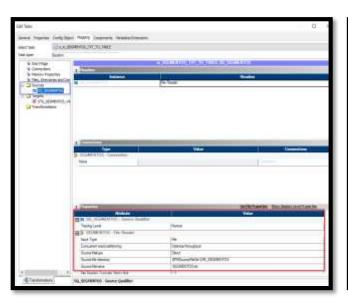
Mapping: m_SEGMENTOS_TXT_TO_TABLE

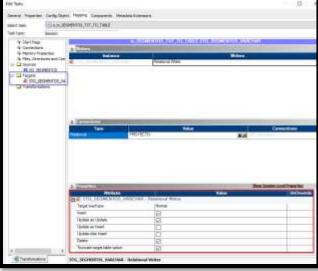


Workflow: wf ACTUALIZACIONES



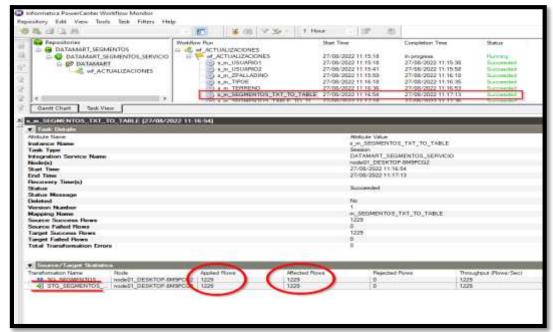
Task: s_m_SEGMENTOS_TXT_TO_TABLE





EJECUCIÓN: m_SEGMENTOS_TXT_TO_TABLE





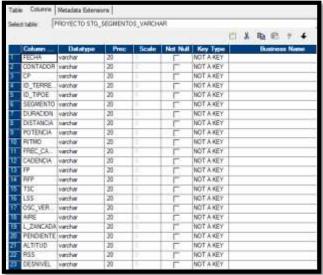


m_SEGMENTOS_TABLE_TO_TIPOOK

Mediante esta transformación, asignaremos a cada atributo su tipo correspondiente, es decir, en un primer momento los extraeremos de la tabla "SEGMENTOS_TXT_TO_TABLE" donde se encuentran todos con el tipo varchar (20) y le asignaremos el tipo con el que podamos trabajar. Como podremos observar deberemos hacer algunas transformaciones mecánicas.

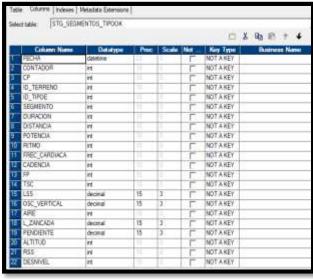
Source: STG_SEGMENTOS_VARCHAR



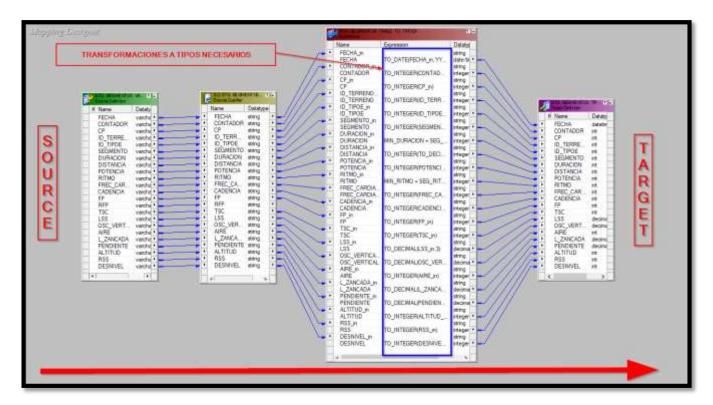


Target: STG_SEGMENTOS_TIPOOK

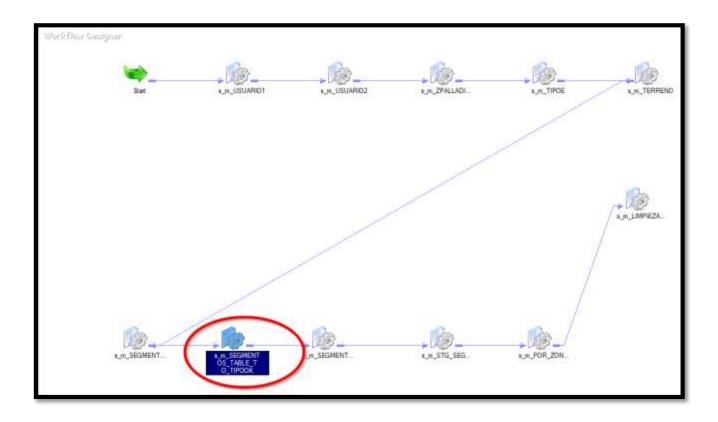




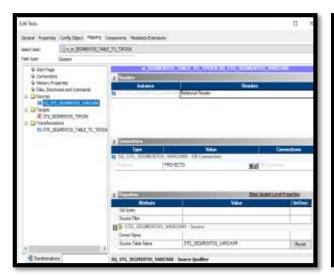
Mapping: m_SEGMENTOS_TABLE_TO_TIPOOK

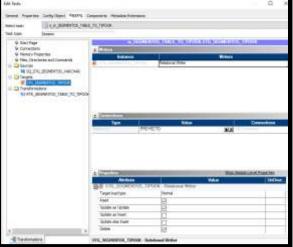


Workflow: wf_ACTUALIZACIONES



Task: s_m_SEGMENTOS_TABLE_TO_TIPOOK

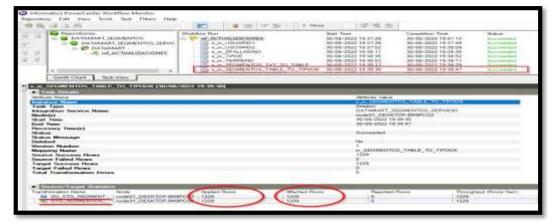






EJECUCIÓN: m_SEGMENTOS_TABLE_TO_TIPOOK



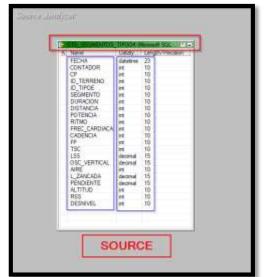


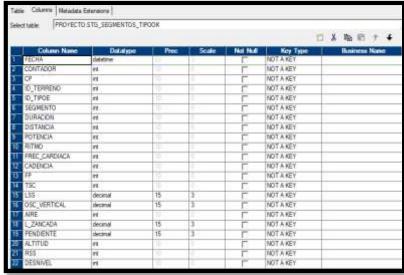


m_SEGMENTOS_RATIOS

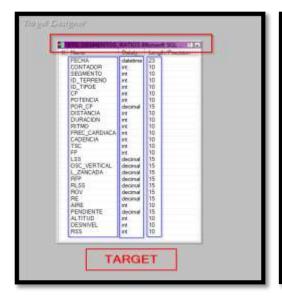
Mediante esta transformación, realizaremos operaciones para crear nuevos campos y utilizaremos diferentes funciones aritméticas como veremos más adelante. Tanto el source como el target serán tablas de nuestra BBDD "DATAMART SEGMENTOS".

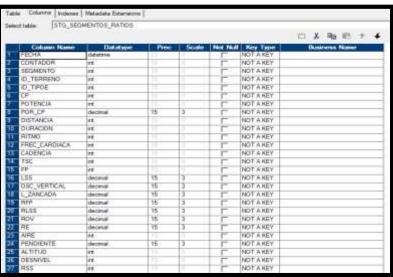
Source: STG_SEGMENTOS_TIPOOK



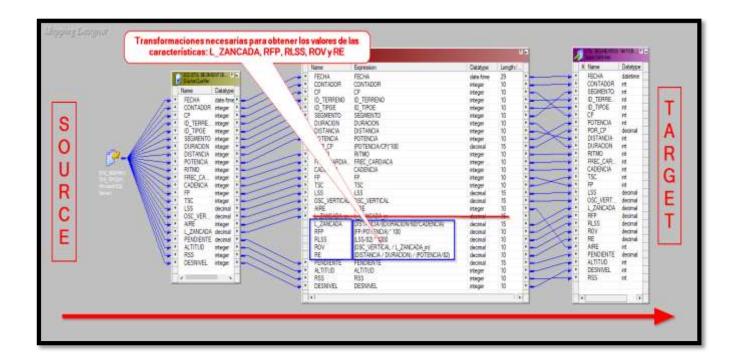


Target: STG_SEGMENTOS_RATIOS

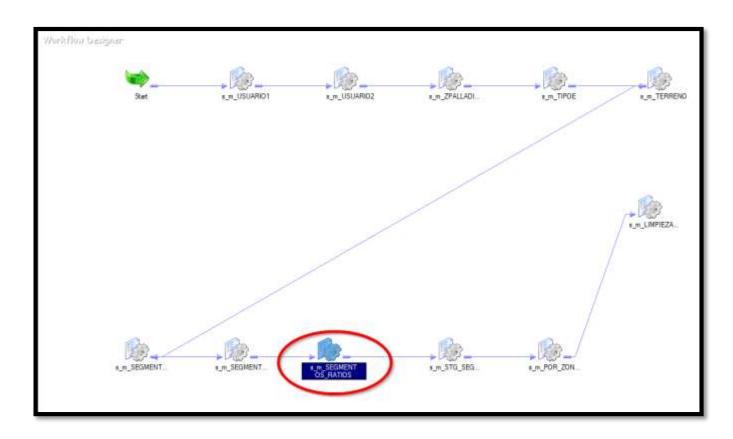




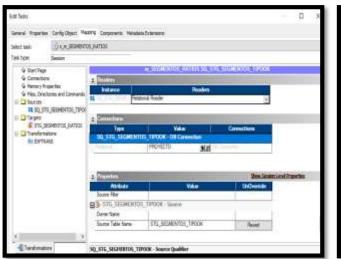
Mapping: m_SEGMENTOS_RATIOS

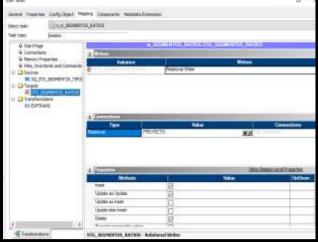


Workflow: wf_ACTUALIZACIONES



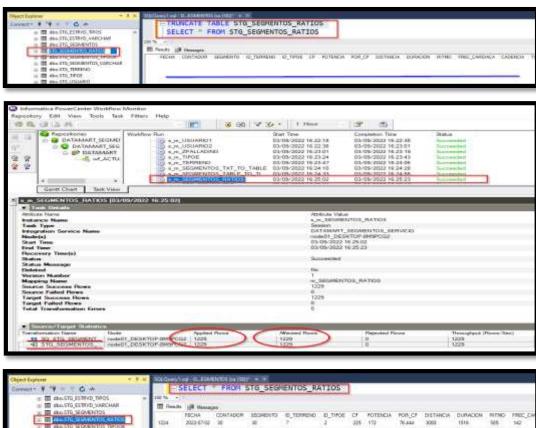
Task: s m SEGMENTOS RATIOS







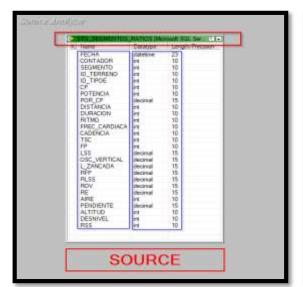
EJECUCIÓN: m_SEGMENTOS_RATIOS

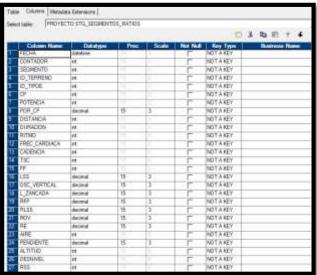


m_SEGMENTOS

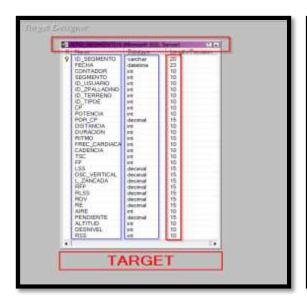
Mediante esta transformación, eliminaremos los campos ID (ID_SEGMENTO, ID_USUARIO y ID_ZPALLADINO) y los cargaremos en la tabla semi-definitiva "STG_SEGMENTOS". Al referirnos como "semi-definitiva" queremos hacer referencia, como podemos observar en los workflows, que forman de un proyecto más amplio y complejo que desarrollaremos en el futuro. Nuestro source será el target de la fase anterior, es decir, "STG_SEGMENTOS_RATIOS" y nuestro target será "STG_SEGMENTOS".

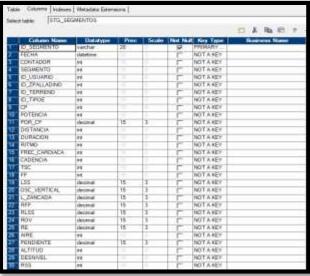
Source: STG_SEGMENTOS_RATIOS



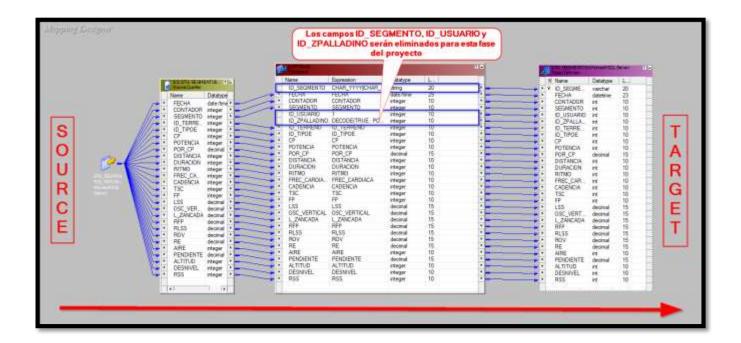


Target: STG_SEGMENTOS

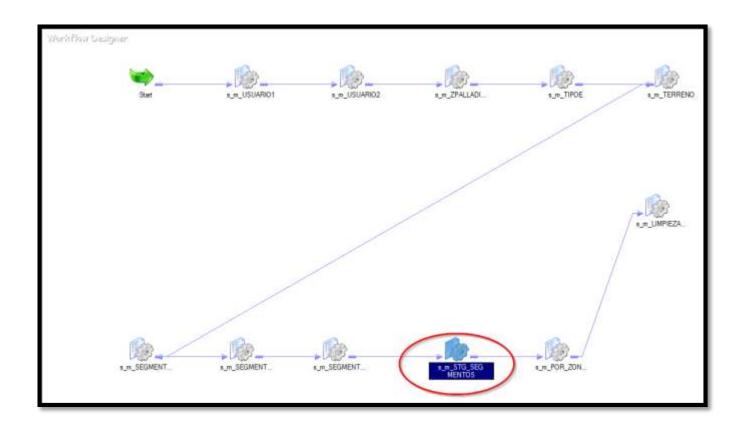




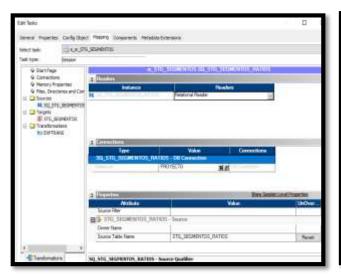
Mapping: m_SEGMENTOS

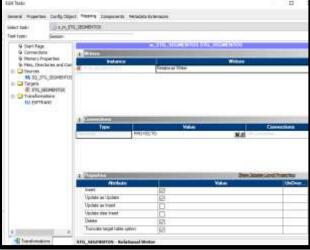


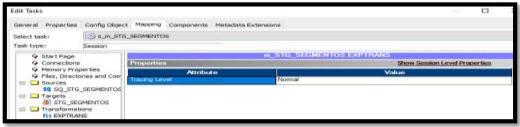
Workflow: wf_ACTUALIZACIONES



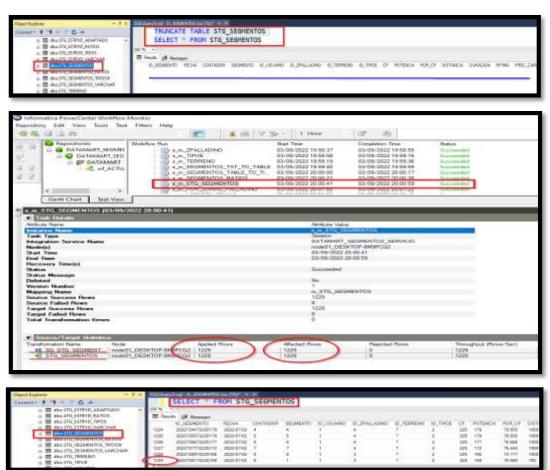
Task: s_m_SEGMENTOS







EJECUCIÓN: m_SEGMENTOS



FUENTES:

- Eduard Barceló: https://www.eduardbarcelo.com/metricas-de-stryd/
- Correr una Maratón: https://www.correrunamaraton.com/stryd-medidor-potencia/
- Palladino Power Project: https://docs.google.com/document/d/e/2PACX-
 1vTXIwQE99RiAfVJOdjUIIhZj7 D0k0LHE0U9gDatL1p4TXEVZZ Rj60S3wczDzgystpclwOS4kKI6R9/pub?fbclid=lwAR35xriNyEjStmE3fYp7BuESFklJyF-cwllvoOGpeDdtLjV39PKqChEFCvo
- STRYD: https://www.stryd.com/eu/es
- COROS: https://www.coros.com/