

Esquema Detallado de Ejecución

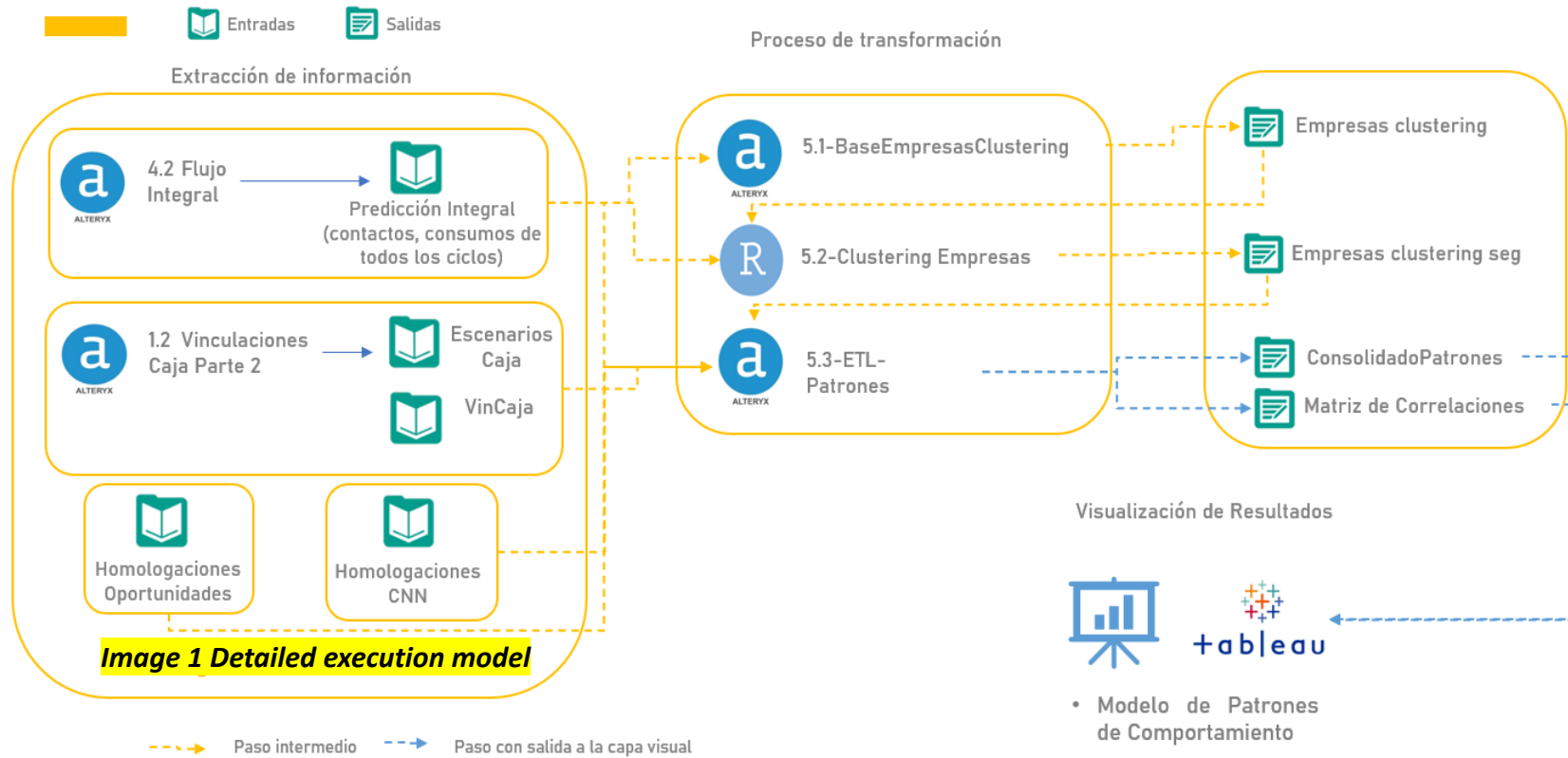
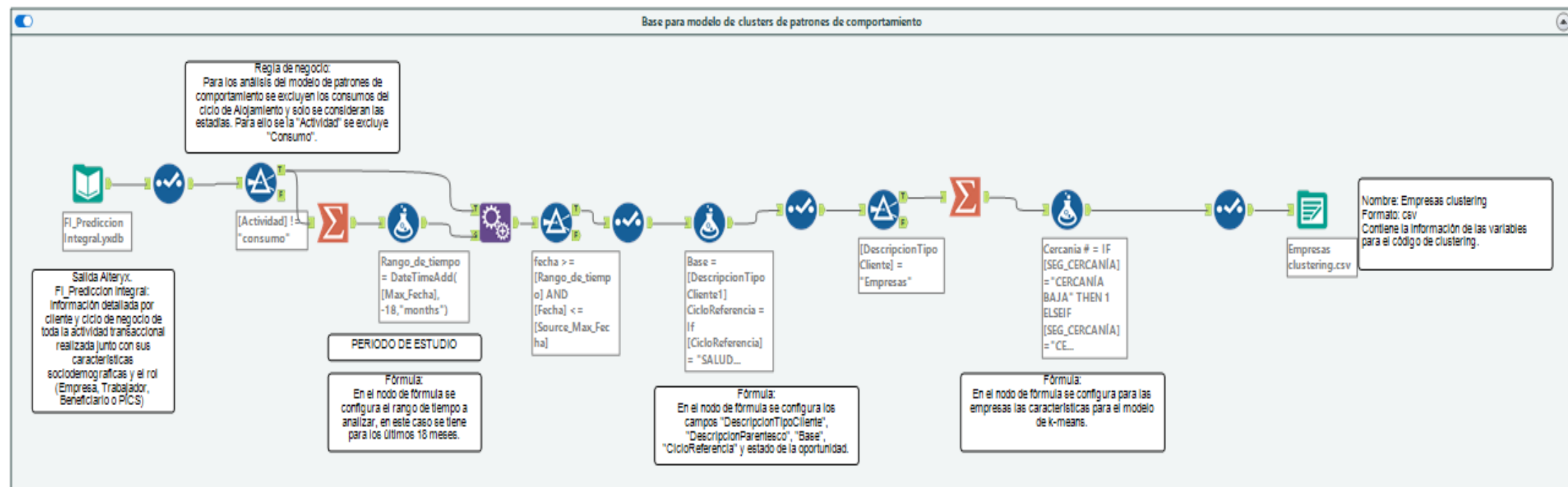


Image 2 Alteryx process

Flujo BaseEmpresasClustering:
Objetivo. Calcular por empresa las variables que se usan en el modelo de clustering para el modelo de patrones de comportamiento.



DICCIONARIO DE DATOS			
Variables	Tipo de dato	Observación	Periodo de análisis
# trabajadores	Entero	Cantidad de trabajadores por empresa	01-01-2022 a 31-08-2022
ActividadEconomica	Cadena de caracteres	Actividad econ de las empresas	
ActividadEconomicaSena	Cadena de caracteres	Actividad econ de las empresas	
acumulado vinculacion	Cadena de caracteres	Cantidad de meses con afiliación activa del usuario	01-01-2022 a 31-08-2022
CategoriaVinculacion	Cadena de caracteres	Tipo vinculación, codificada letras	01-01-2022 a 31-08-2022
Ciudad	Cadena de caracteres	Ciudad del usuario	01-01-2022 a 31-08-2022
CLTV	Decimal	Modelo predictibilidad CLTV	
CLTV_categ	Cadena de caracteres	Modelo predictibilidad CLTV	
CodigoGenero	Cadena de caracteres		01-01-2022 a 31-08-2022
CodigoTipoidentificacion	Entero		01-01-2022 a 31-08-2022
DescripcionCIU	Cadena de caracteres	Actividad econ de las empresas	
DescripcionTipoCliente	Cadena de caracteres		01-01-2022 a 31-08-2022
Edad	Entero		01-01-2022 a 31-08-2022
EstadoActivo	Entero	Estado de la vinculación en el periodo de análisis	
EstadoCivil	Cadena de caracteres		01-01-2022 a 31-08-2022
FAM_NAUTCLI	Cadena de caracteres	Modelo predictibilidad atracción afiliados	
Grupo_Atraccion	Cadena de caracteres	Modelo predictibilidad atracción afiliados	
Grupo_Retencion	Cadena de caracteres	Modelo predictibilidad atracción afiliados	
IdCliente	Cadena de caracteres		
IdEmpresa	Cadena de caracteres		
IdResponsableVinculacion	Cadena de caracteres	Id de la empresa a la cual pertenece el afiliado trabajador dependiente o Beneficiario	
Indicador_bienestar	Cadena de caracteres	Modelo indice de binestar y momento de vida	2022Q2

a. Criterio de selección de variables.

Variable selection criteria.

According to the needs of the model, companies in the domestic sector are first discarded, to focus on formally incorporated and affiliated companies.

*During the selection and creation process of the data sheet, no values were imputed to empty or missing fields.

*In the universe of companies registered as of the cutoff date (October 2022).

*Although categorical variables exist, they are ordinal, which maintains a hierarchical and ordinal distance in the nominal value of each group to which they belong De acuerdo a las necesidades del modelo, primero se descartan las empresas que están en el sector doméstico; con el fin de enfocarnos en las empresas formalmente constituidas que se encuentran afiliadas.