Paquetes Necesarios para el Proyecto (2.1 y 2.2)

Tidyverse

Este es una collecion de otros paquetes en R relacionados con el manejo de datos que se van a usar en el proyecto. Se instalaba todo el Tidyverse y luego se escogían los paquetes necesarios dentro de este para los scripts.

ReadXl

Este paquete nos permite leer los datos de los archivos de Excel y usarlos en nuestro espacio de trabajo en R.

Openxlsx

Este paquete nos permite crear y editar archivos de Excel, fue usado para el manejo de los csv.

Ggplot2

Esta librería es un sistema de visualización de datos, se usó para crear las graficas que se verán a continuación.

Dplyr

Un paquete que nos permite manipular los datos en R, en nuestro caso fue usado sobretodo para hacer acciones tipo Join en las tablas.

Tidyr

Un paquete el cual contiene herramientas para organizar los datos, usado principalmente por su función reshape.

RColorBrewer

Un paquete que ofrece distintas paletas de colores y la capacidad de manejar colores de una mejor manera en nuestro entorno de R, usado principalmente con ggplot2 para las graficas.

Patchwork

Este paquete permite que se puedan combinar distintos gráficos en un mismo lienzo, usado para este mismo fin.

Lubridate

Este paquete nos permite hacer más operaciones matemáticas con los objetos de tipo fecha. Usado para manejar de mejor manera los precios históricos de los productos.

Skimr

Diseñado para proveer resúmenes estatisticos acerca de variables, data frames, tablas, etc. Usado para obtener más información de los datos.

(3.1)

Fuente original de los datos:

DANE (2023) *Sistema de Información de Precios y Abastecimiento del Sector Agropecuario (SIPSA)*. Gobierno de Colombia. Recuperado el 15 de Noviembre de 2023 de: https://www.dane.gov.co/index.php/estadisticas-por-tema/agropecuario/sistema-de-informacion-de-precios-sipsa#precios-mayoristas

(3.2)

Los datos fueron recolectados por el DANE como parte de la operación SIPSA (Sistema de información de Precios y Abastecimiento) que se les fue encomendada en el 2012. A partir de esta fecha, el DANE quedó como responsable del rediseño y operación del SIPSA en sus tres componentes (Precios Mayoristas, Abastecimientos de Alimentos e Insumos y Factores Asociados a la Producción Agropecuaria), que continúan llevando a cabo hasta la fecha.

Las series históricas de precios mayoristas tienen 5 variables, la fecha en la cual se tomó la medición del precio del producto, el grupo al que pertenece el producto, el producto como tal, el mercado en donde se tomó la medición y el precio promedio por kilogramo.

Las series históricas de abastecimiento tienen 38 variables, la fecha en la cual se tomó la medición, el total de abastecimiento, la variación mensual del abastecimiento, la variación anual del abastecimiento, dos variables de empalme, y de la número 7 a la 38 son distintos mercados en donde se tiene el valor de abastecimiento.

La seres históricas de insumos tienen dos distintos sets de variables, ya que la tercer hoja del documento alberga el listado de los insumos, con 3 variables, el grupo del insumo, el subgrupo y el nombre del insumo/factor. Luego las demás hojas representan cada una un subgrupo ya establecido en la hoja del listado, estas cuentan con 8 variables, el año en que se tomó la medida, el mes, el codigo del departamento, el nombre del departamento, el codigo del municipio, el nombre del municipio, el nombre del producto, la presentación del producto y el precio promedio.

(3.3)

Se descargan los csv que se van a usar y luego se realiza una pequeña limpieza y tratado de datos, es más cómodo trabajar con las cuidades separadas del mercado, así que se separan en dos variables diferentes, se juntan los promedios de los meses en una sola hoja, una hoja por cada año, también para más comodidad y por ultimo, se acomoda de tal manera que todo esté en una sola hoja. A la hora de juntar los promedios se realiza un inner join para así también “eliminar” los espacios vacíos.

(Me falta revisar que se hizo para los datasets de insumos y abastecimiento)

(3.4)

\*\*\*Hacer un head(\*nombre\_archivo\*)? Talvez el datosRawprecios\_por\_mes\_2013\_2023

(3.5)

Las variables más importantes en los datasets de las series históricas son las localidades de los mercados ya que con estas podemos saber exactamente donde se encuentran los mejores productos y el promedio del precio por kilogramo por mes con el cual haremos los calculos, estas se van a usar más adelante para saber cuál es el producto que menos y más variación tiene, el comportamiento de estos productos teniendo en cuenta el