

Análisis de Series Temporales: Ventas de Cigarrillos en España

Planteamiento del Proyecto

Título: Análisis de Series Temporales para Ventas de Cigarrillos en España: Clubes de Convergencia y Predicciones Futuras

Descripción:

Este proyecto analiza la venta de cigarrillos en las provincias de España desde enero de 2005 hasta diciembre de 2021. Utilizando técnicas de series temporales y aprendizaje no supervisado, el proyecto identifica patrones y tendencias de ventas, evalúa la convergencia en el comportamiento de ventas entre provincias y proporciona predicciones futuras. Este análisis permite identificar oportunidades estratégicas para acciones de marketing y ventas diferenciadas por provincia.

Objetivos:

1. Analizar el comportamiento de las ventas de cigarrillos en las provincias españolas a lo largo del tiempo.
2. Identificar y caracterizar clubes de convergencia entre las provincias basados en sus patrones de ventas.
3. Realizar predicciones de ventas a corto plazo para provincias seleccionadas, evaluando el impacto del COVID-19.
4. Proveer recomendaciones estratégicas basadas en el análisis para optimizar las ventas futuras.

Metodología:

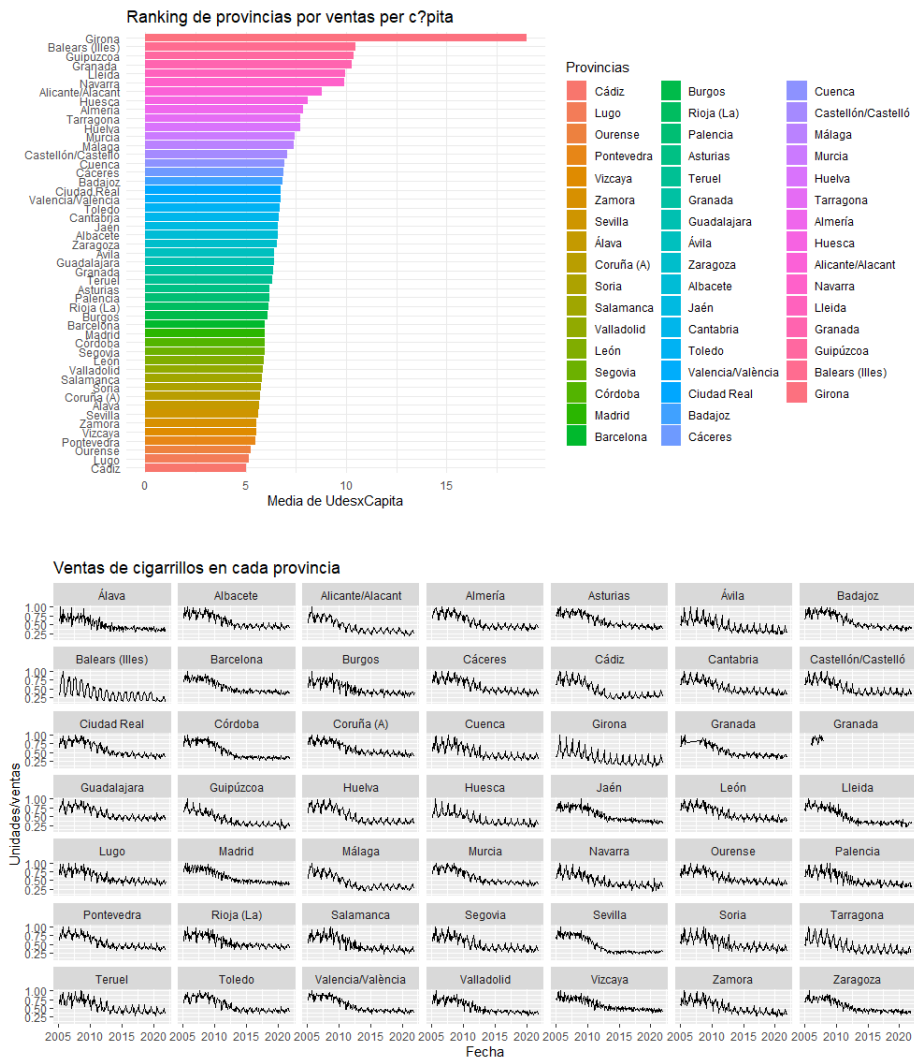
- **Exploración de Datos:** Análisis inicial de las características y distribuciones de las variables del dataset `dataset_provincias.csv`.
- **Visualización de Series Temporales:** Representación gráfica del comportamiento de las ventas a lo largo del tiempo, normalizando las ventas para comparar entre provincias.
- **Clubes de Convergencia:** Uso de la librería `ConvergenceClubs` para detectar clubes de convergencia y analizar la homogeneidad en las ventas entre provincias.
- **Predicción de Series Temporales:** Implementación de modelos ARIMA para predecir las ventas futuras en provincias seleccionadas, analizando la estabilidad y previsibilidad de las series temporales por club.

Resultados y Hallazgos:

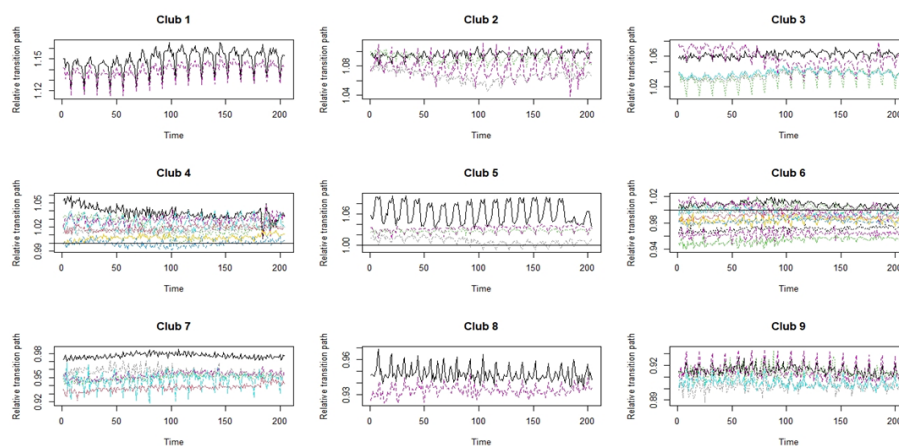
- Se observó una tendencia decreciente en las ventas desde 2010, agravada por la pandemia de COVID-19, especialmente en zonas turísticas.
- No se encontró convergencia absoluta en las ventas, indicando comportamientos heterogéneos entre las provincias.
- La segmentación en clubes mostró patrones de venta similares entre provincias no necesariamente cercanas geográficamente, lo que sugiere factores comunes en el comportamiento del consumidor o en la estrategia de mercado.

- Las predicciones de ventas indican diferencias en la estabilidad y previsibilidad de las series temporales entre los clubes, con implicaciones para la planificación estratégica.

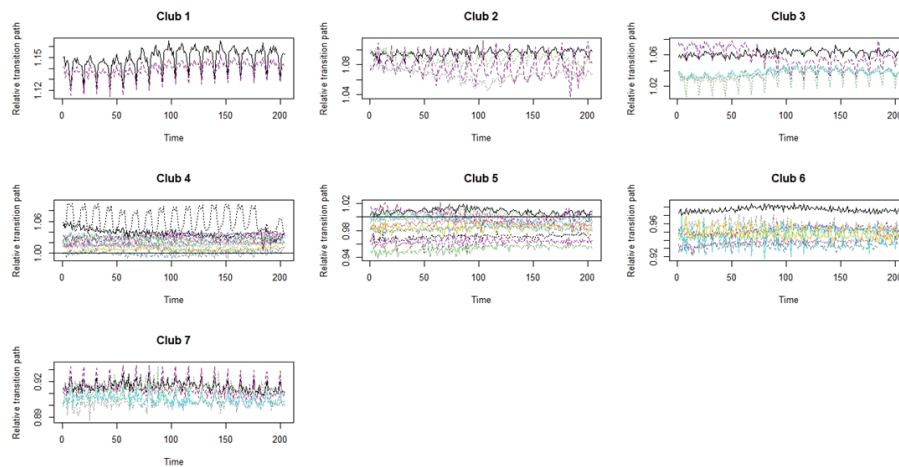
Gráficos:



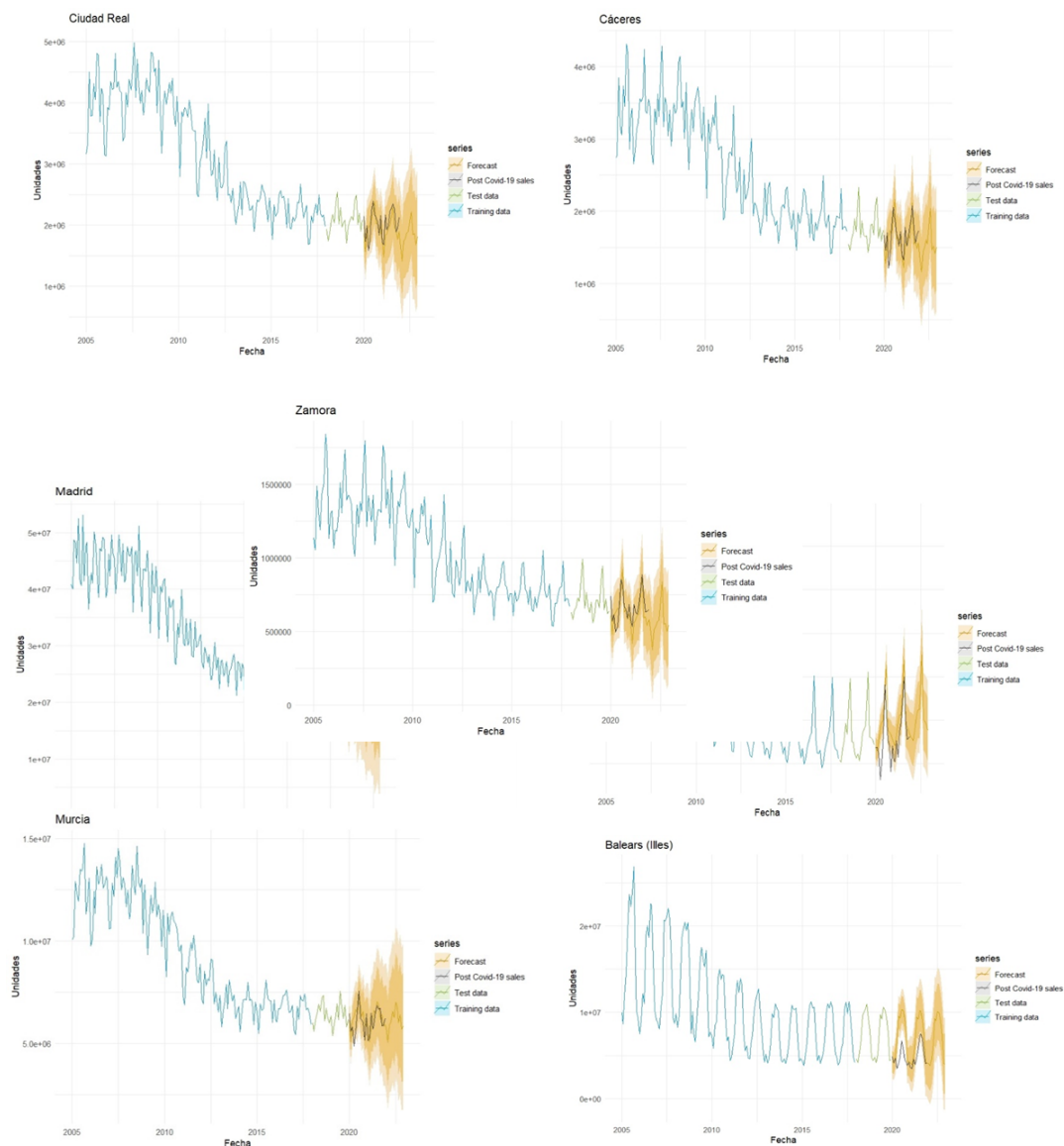
Clasificación de Series Temporales: Agrupación Basada en Patrones de Ventas



Aplicación del método de Phillips y Sul



Visualización de los datos de entrenamiento, test, pronóstico y post-COVID-19



Código en R: Análisis y Explicación

1. **Carga de Librerías y Datos:** Se utilizan librerías como `tidyverse`, `dplyr`, y `ggplot2` para manipulación de datos y visualización. Los datos se cargan desde un archivo CSV.
2. **Análisis Exploratorio de Datos:** Se examina la estructura del dataset usando `str()` y se proporciona un resumen estadístico con `summary()`.
3. **Transformación de Datos:** La columna de fechas se convierte al formato apropiado para análisis de series temporales.
4. **Visualización de Series Temporales:** Se crean gráficos para visualizar las ventas de cigarrillos en las provincias y las ventas normalizadas.
5. **Análisis de Convergencia:** Se utiliza la librería `ConvergenceClubs` para analizar la convergencia de las ventas en las provincias, aplicando métodos como el de Phillips y Sul para agrupar provincias en clubes según su comportamiento de ventas.
6. **Predicción de Series Temporales:** Se implementan modelos ARIMA para hacer predicciones de ventas para provincias seleccionadas, analizando el impacto de eventos recientes como la pandemia.