

## **Calidad de los Datos**

Nuria Arroyo Bustamante y Heriberto Espino Montelongo

Universidad de las Américas Puebla

P25 LDS1021 1: Minería de Datos

Dr. Hector Saib Maravillo Gomez

3 de marzo de 2025

## Calidad de los Datos

El análisis se realizará a partir del 1 de marzo de 2011 hasta diciembre de 2022, ya que distintas fuentes de información solo tienen información durante este periodo, como datos de la Tasa Efectiva Promedio Ponderada por Saldo de Tarjetas de Crédito o datos publicados anualmente.

Muchos de los datos se organizan de forma diaria, la mayoría de manera mensual, y solo dos conjuntos se actualizan trimestralmente: el número de tarjetas de crédito y débito, y la tasa efectiva promedio ponderada por saldo de tarjetas de crédito. Dado que estos últimos solo se registraban en los meses de enero, abril, julio y octubre, se optó por aplicar una interpolación lineal para estimar los valores en los meses intermedios, ya que los datos presentan una evolución relativamente continua y sin cambios abruptos porque reflejan dinámicas del mercado financiero que, aunque pueden mostrar tendencias o fluctuaciones, no suelen presentar discontinuidades significativas entre periodos consecutivos. Por otro lado, para los datos diarios se utilizó el primer dato disponible de cada mes como representante del periodo mensual, ya que es un dato que en verdad se emitió, además de facilitar la alineación con otras variables económicas y financieras que se reportan mensualmente.

Aunque los datos no presentan errores, se identificaron columnas compuestas únicamente por valores nulos o que contienen información de periodos anteriores al que se analizará; en estos casos, dichos datos se descartarán. Por ejemplo, la TIIP a 28 días dejó de publicarse en México el 31 de diciembre de 2001 (según el Diario Oficial de la Federación del 23 de mayo de 2001) y la MEXIBOR dejó de publicarse el 2 de enero de 2002, ambos al pasar a la TIIE. El Pagaré Bancario Bursátil a 28 días tampoco se usará ni la CPPE, por contener solo valores nulos durante el periodo del análisis. Asimismo, la TIIE a 182 días se excluye porque el análisis se realizará a partir del 1 de marzo de 2011 y esta se comenzó a publicar después. Otros datos excluidos dado a la fecha de publicación son las Tasas de Fondeo Gubernamental y la inflación quincenal (sustituida posteriormente por la inflación mensual).

Finalmente, se observó la ausencia de información relativa a sistemas de pago de bajo valor y a ciertos tipos de tarjetas de crédito (como la A o la Z), por lo que estas no serán consideradas para el análisis. No obstante, se incluirán todos los tipos de tarjeta a pesar de que haya unas que cuenten con menos de 1,000, como la Mastercard U y Visa E.

Datos como el Vector de precios de títulos gubernamentales, Udibonos 30 años - Cupón Vigente son constantes durante todo el periodo del análisis, por lo que será descartado. Y Valores

privados de corto plazo semanal, Papel comercial, tasa promedio ponderada, Porcentajes contiene muy pocos valores distintos de cero, por lo que también será descartada.