Documentación del Proyecto de Análisis de Datos

**Analisis del Precio de las Tortillas en México**

**Fecha:** 07/05/2025

**Autor(es):**

Said Mariano Sánchez

**Organización:**

Proyecto individual de portafolio | Aspirante a Analista de Datos Jr.

**Propósito del Documento:**

Este documento tiene como propósito servir como guía de referencia y registro del desarrollo del proyecto de análisis de datos sobre el precio de las tortillas en México. Incluye la justificación, objetivos, metodología, resultados y conclusiones obtenidas a lo largo del análisis.

**Resúmenes Ejecutivos:**

Contenido

[1. Introducción 4](#_Toc197597802)

[1.1 Título del Proyecto 4](#_Toc197597803)

[1.2 Fecha 4](#_Toc197597804)

[1.3 Autor(es) 4](#_Toc197597805)

[1.4 Organización 4](#_Toc197597806)

[1.5 Propósito del Documento 4](#_Toc197597807)

[1.6 Resumen Ejecutivo 4](#_Toc197597808)

[2. Descripción del Proyecto 5](#_Toc197597809)

[2.1 Antecedentes 5](#_Toc197597810)

[2.2 Objetivos 5](#_Toc197597811)

[2.2.1 Objetivo Principal 5](#_Toc197597812)

[2.2.2 Objetivos Específicos 5](#_Toc197597813)

[2.3 Alcance 6](#_Toc197597814)

[2.4 Justificación 6](#_Toc197597815)

[2.5 Preguntas de Investigación 6](#_Toc197597816)

[3. Datos 7](#_Toc197597817)

[3.1 Fuentes de Datos 7](#_Toc197597818)

[3.1.1 Fuente 1 7](#_Toc197597819)

[3.1.2 Fuente 2 7](#_Toc197597820)

[3.2 Recolección de Datos 7](#_Toc197597821)

[3.3 Descripción de los Datos 7](#_Toc197597822)

[3.3.1 Variables: (Nombre, Tipo, Descripción) 7](#_Toc197597823)

[3.4 Calidad de los Datos 7](#_Toc197597824)

[3.5 Consideraciones Éticas 7](#_Toc197597825)

[4. Metodología 8](#_Toc197597826)

[4.1 Enfoque del Análisis 8](#_Toc197597827)

[4.2 Herramientas y Técnicas 8](#_Toc197597828)

[4.3 Preparación de Datos 8](#_Toc197597829)

[4.4 Análisis Exploratorio de Datos (EDA) 9](#_Toc197597830)

[4.5 Modelado 9](#_Toc197597831)

[4.5.1 Selección del Modelo 9](#_Toc197597832)

[4.5.2 Entrenamiento del Modelo 9](#_Toc197597833)

[4.5.3 Evaluación del Modelo 9](#_Toc197597834)

[4.6 Validación de Resultados 9](#_Toc197597835)

[5. Resultados 10](#_Toc197597836)

[5.1 Hallazgos Clave 10](#_Toc197597837)

[5.2 Visualizaciones 10](#_Toc197597838)

[5.3 Interpretación de Resultados 10](#_Toc197597839)

[6. Conclusiones y Recomendaciones 10](#_Toc197597840)

[6.1 Conclusiones 10](#_Toc197597841)

[6.2 Limitaciones del Estudio 10](#_Toc197597842)

[6.3 Recomendaciones 10](#_Toc197597843)

[6.4 Próximos Pasos 10](#_Toc197597844)

[7. Apéndices 11](#_Toc197597845)

[7.1 Glosario 11](#_Toc197597846)

[7.2 Código 11](#_Toc197597847)

[7.3 Datos Adicionales 11](#_Toc197597848)

[7.4 Referencias 11](#_Toc197597849)

# 1. Introducción

## 1.1 Título del Proyecto

## 1.2 Fecha

## 1.3 Autor(es)

## 1.4 Organización

## 1.5 Propósito del Documento

## 1.6 Resumen Ejecutivo

# 2. Descripción del Proyecto

## 2.1 Antecedentes

La tortilla es uno de los alimentos más consumidos en México y su precio ha sido históricamente un tema de interés social, económico y político. Factores como el precio del maíz, el transporte, la inflación y las políticas públicas han impactado su costo final. Analizar estos datos a lo largo de varios años ofrece una visión integral del comportamiento del mercado.

## 2.2 Objetivos

## 2.2.1 Objetivo Principal

Analizar la evolución del precio de las tortillas en México desde 2007 hasta 2024, identificando patrones temporales, diferencias regionales y factores que podrían haber influido en su comportamiento.

## 2.2.2 Objetivos Específicos

EDA:

* Realizar un análisis exploratorio de datos.
  + Familiarizarse con el conjunto de datos.
  + Identificar posibles problemas.

Limpieza:

* Revisión y limpieza de datos.
  + Datos nulos, duplicados o inconsistentes.
* Conversión de fechas y formatos adecuaos.
* Estandarización de nombre de columnas/estados en caso de ser necesario.

Estadistica:

* Utilizar media, mediana y desviación estándar.
* Análisis de la dispersión
  + Rango y IQR
* Pruebas de estadística.
  + T-test o ANOVA
  + Regresión lineal

## 2.3 Alcance

El análisis se limita a los datos disponibles del conjunto publicado en Kaggle, comprendiendo los años 2007 a 2024. No se consideran variables externas como inflación, tipo de cambio o precios del maíz, a menos que se integren en futuras fases del proyecto.

## 2.4 Justificación

Comprender la variación del precio de un alimento de consumo diario como la tortilla permite no solo estudiar fenómenos económicos, sino también generar información útil para tomadores de decisiones, académicos, y ciudadanía en general. Además, este proyecto fortalece habilidades de análisis de datos, visualización y comunicación de resultados.

## 2.5 Preguntas de Investigación

* ¿Cómo ha evolucionado el precio de la tortilla en México entre 2007 y 2024?
* ¿Existen diferencias significativas entre regiones o estados?
* ¿Hay alguna región que destaca?
* ¿Qué años registraron cambios abruptos o comportamientos atípicos?
* ¿Cuál es la tendencia general del precio en el país?
* ¿Qué factores podrían influir en estas diferencias?
* ¿Es posible encontrar patrones estacionales o cíclicos?

# 3. Datos

## 3.1 Fuentes de Datos

## 3.1.1 Fuente 1

Kaggle - Tortilla Prices in Mexico

https://www.kaggle.com/datasets/richave/tortilla-prices-in-mexico

## 3.1.2 Fuente 2

(No aplica)

## 3.2 Recolección de Datos

El conjunto de datos fue descargado directamente desde la plataforma Kaggle. No se realizó scraping ni se recolectaron datos manuales.

## 3.3 Descripción de los Datos

### 3.3.1 Variables: (Nombre, Tipo, Descripción)

## 3.4 Calidad de los Datos

Se evaluarán posibles inconsistencias como fechas inválidas, estados mal escritos, valores extremos en los precios, y datos faltantes.

## 3.5 Consideraciones Éticas

Los datos son públicos y anónimos. No se incluyen datos personales ni sensibles. Se utilizarán únicamente con fines educativos.

# 4. Metodología

## 4.1 Enfoque del Análisis

Se adopta un enfoque cuantitativo, utilizando técnicas estadísticas descriptivas e inferenciales para explorar, limpiar y analizar los datos. El análisis está orientado a encontrar patrones temporales, espaciales y de dispersión en el precio de las tortillas a lo largo del tiempo y por entidad federativa.

## 4.2 Herramientas y Técnicas

Lenguaje de programación:

* Python 3.12.1

Librerías y paquetes principales:

* pandas, numpy: manipulación y limpieza de datos
* matplotlib, seaborn: visualización de datos
* scipy.stats: análisis estadístico
* plotly: visualizaciones interactivas (opcional)
* Jupyter Notebook o VS Code como entorno de desarrollo

Técnicas estadísticas:

* Media, mediana, desviación estándar
* Rango intercuartílico (IQR)
* T-test, ANOVA para comparar grupos
* Regresión lineal simple para detectar tendencias

## 4.3 Preparación de Datos

 Importación del dataset desde archivo CSV

 Conversión de fechas al tipo datetime

 Estandarización de nombres de estados

 Eliminación de duplicados

 Identificación y tratamiento de valores nulos o atípicos

 Validación del tipo de datos por columna

## 4.4 Análisis Exploratorio de Datos (EDA)

 Análisis de la distribución general de precios

 Identificación de outliers con boxplots

 Análisis por estado y por año

 Evolución temporal mediante líneas de tiempo

 Mapas de calor o gráficos geoespaciales si es posible

## 4.5 Modelado

## 4.5.1 Selección del Modelo

(No aplica)

## 4.5.2 Entrenamiento del Modelo

(No aplica)

## 4.5.3 Evaluación del Modelo

(No aplica)

## 4.6 Validación de Resultados

(No aplica)

# 5. Resultados

## 5.1 Hallazgos Clave

(Pendiente de completar conforme se avance)

## 5.2 Visualizaciones

(Pendiente de completar conforme se avance)

## 5.3 Interpretación de Resultados

(Pendiente de completar conforme se avance)

# 6. Conclusiones y Recomendaciones

## 6.1 Conclusiones

## 6.2 Limitaciones del Estudio

## 6.3 Recomendaciones

## 6.4 Próximos Pasos

# 7. Apéndices

## 7.1 Glosario

## 7.2 Código

## 7.3 Datos Adicionales

## 7.4 Referencias