 .

Realizado por:

Estanislao Guido Paco Ramos

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Tutor: (Grado, nombres y apellidos)

###### 

###### Relator:

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**2025**

Universidad Católica Boliviana “San Pablo”

Facultad de Ingeniería

Departamento de Ingeniería de Sistemas

##### La Paz - Bolivia

PROYECTO DE GRADO PRESENTADO AL DEPARTAMENTO DE INGENIERÍA DE SISTEMAS PARA LA OBTENCIÓN DEL GRADO ACADÉMICO DE MAESTRÍA EN CIENCIA DE DATOS

TRA

“Modelo Predictivo de Precios de Canasta Básica Familiar mediante Técnicas de *Machine Learning* para Bolivia”

ÍNDICE GENERAL

[Capítulo 1.- MARCO REFERENCIAL 1](#_Toc204728459)

[1.1. Introducción 1](#_Toc204728460)

[1.2. Antecedentes 1](#_Toc204728461)

[1.2.1. Antecedentes del objeto de estudio 1](#_Toc204728462)

[1.2.2. Referencias técnicas de otros trabajos 2](#_Toc204728463)

[1.3. Descripción del objeto de estudio 3](#_Toc204728464)

[1.4. Identificación del Problema 4](#_Toc204728465)

[1.5. Formulación del Problema 5](#_Toc204728466)

[1.6. Objetivos 5](#_Toc204728467)

[1.6.1. Objetivo General 5](#_Toc204728468)

[1.6.2. Objetivos Específicos 5](#_Toc204728469)

[1.7. Justificaciones 6](#_Toc204728470)

[1.7.1. Justificación técnica 6](#_Toc204728471)

[1.7.2. Justificación social 6](#_Toc204728472)

[1.7.3. Justificación económica 6](#_Toc204728473)

[1.8. Límites y Alcances 6](#_Toc204728474)

[1.8.1. Límites 6](#_Toc204728475)

[1.8.2. Alcances 7](#_Toc204728476)

[1.9. Metodología de la investigación 7](#_Toc204728477)

[1.9.1. Tipo de estudio 7](#_Toc204728478)

[1.9.2. Métodos 7](#_Toc204728479)

[1.9.3. Técnicas 8](#_Toc204728480)

[1.10. Análisis preliminar 8](#_Toc204728481)

[1.11. Propuesta de solución 10](#_Toc204728482)

[1.12. Cronograma 10](#_Toc204728483)

[Capítulo 2.- MARCO TEÓRICO 10](#_Toc204728484)

[2.1. Base teórica en la que se sustenta el proyecto 10](#_Toc204728485)

[Capítulo 3.- MARCO PRÁCTICO 10](#_Toc204728486)

[3.1. Análisis del ámbito de aplicación del proyecto 10](#_Toc204728487)

[3.2. Determinación de requerimientos 10](#_Toc204728488)

[3.2.1. Requerimientos funcionales 10](#_Toc204728489)

[3.2.2. Requerimientos no funcionales 10](#_Toc204728490)

[3.3. Análisis del proyecto 10](#_Toc204728491)

[3.4. Diseño del proyecto 10](#_Toc204728492)

[3.5. Desarrollo del Proyecto 10](#_Toc204728493)

[3.6. Validación y pruebas 10](#_Toc204728494)

[3.7. Resultados 10](#_Toc204728495)

[Capítulo 4.- CONCLUSIONES Y RECOMENDACIONES 10](#_Toc204728496)

[4.1. Conclusiones 10](#_Toc204728497)

[4.2. Recomendaciones 10](#_Toc204728498)

[Referencias Bibliográficas 10](#_Toc204728499)

[Anexos 10](#_Toc204728500)

[Apéndices 10](#_Toc204728501)

[Glosario 10](#_Toc204728502)

# Capítulo 1.- MARCO REFERENCIAL

## Introducción

El presente proyecto de posgrado está dirigido a desarrollar un modelo de clasificación automática de mercados alimentarios bolivianos en función a su nivel de volatilidad de precios, utilizando técnicas de aprendizaje supervisado como el Análisis Discriminante Lineal (LDA) y la Regresión Logística. Bolivia, al ser un país con una diversidad geográfica significativa y variaciones marcadas en las condiciones socioeconómicas de sus regiones, presenta también disparidades considerables en los precios de productos alimentarios básicos entre sus diferentes mercados. Esta heterogeneidad no ha sido abordada sistemáticamente por las instituciones encargadas de la seguridad alimentaria, que aplican medidas generalizadas sin discriminar la estabilidad o volatilidad específica de cada mercado.

Ante esta situación, se plantea el diseño e implementación de una metodología que permita clasificar los mercados bolivianos en función de su comportamiento histórico de precios. Este sistema no solo permitirá un conocimiento más profundo de las características locales de cada mercado, sino que también facilitará la toma de decisiones estratégicas para focalizar intervenciones, optimizar recursos y mejorar la planificación logística del abastecimiento de alimentos. El documento está organizado en varios capítulos, donde se desarrolla primero el marco referencial, seguido del marco teórico, metodología, análisis de resultados y finalmente la conclusión con recomendaciones.

## Antecedentes

### Antecedentes del objeto de estudio

En Bolivia, la recopilación de datos de precios alimentarios ha sido impulsada por organismos internacionales como el Programa Mundial de Alimentos (WFP), que ha generado una base de datos histórica extensa y confiable. Esta base de datos incluye registros de precios de productos alimenticios desde 2008 hasta 2022, recolectados de distintos mercados en varias regiones del país. Esta información ha sido utilizada principalmente para el monitoreo y diagnóstico de tendencias generales, pero no se han desarrollado estudios profundos orientados a la clasificación de mercados por su nivel de volatilidad.

### Referencias técnicas de otros trabajos

1. **FAO (2011). "Price Volatility and Food Security"**

* Autor: Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO)
* Objetivo general: Analizar las causas, consecuencias y posibles medidas de mitigación frente a la volatilidad de precios de alimentos a nivel mundial.
* Resumen: El documento discute cómo la inestabilidad en los precios puede afectar el acceso a los alimentos, especialmente entre las poblaciones más pobres. Se destacan aspectos como la especulación financiera, el cambio climático y la integración de mercados como factores que intensifican la volatilidad. Se proponen mecanismos de respuesta como reservas estratégicas, seguros agrícolas y redes de protección social.
* Diferencia con este proyecto: Aunque aporta un marco conceptual general, no propone herramientas de clasificación ni análisis empírico específico por mercado o región como el que se plantea para Bolivia.

1. **Benjaminsen, T. A., et al. (2020). "Market Price Volatility and Vulnerability in Rural Africa"**

* Autor: T. A. Benjaminsen y colaboradores, Instituto Noruego para Asuntos Internacionales (NUPI)
* Objetivo general: Identificar regiones rurales africanas con alta vulnerabilidad asociada a la inestabilidad de precios alimentarios.
* Resumen: Utilizando datos recolectados en múltiples mercados de África subsahariana, el estudio aplicó técnicas de agrupamiento (clustering) para categorizar zonas de riesgo y generar mapas de vulnerabilidad. Se abordaron factores climáticos, sociales y logísticos como condicionantes de la volatilidad.
* Diferencia con este proyecto: A diferencia del presente estudio, no se utiliza clasificación supervisada ni se enfoca en mercados individuales como unidad de análisis. Además, no se plantea un sistema escalable para nuevos datos.

1. **Gonzales, A. (2021). "Clasificación de mercados mayoristas en Perú por patrones de precio usando modelos discriminantes"**

* Autor: Ana Gonzales, Universidad Nacional Mayor de San Marcos, Perú
* Objetivo general: Identificar y clasificar mercados mayoristas peruanos en función de su comportamiento de precios.
* Resumen: Se utilizaron datos de precios mensuales para varios mercados regionales peruanos, aplicando Análisis Discriminante Lineal (LDA) para segmentarlos según su estabilidad o inestabilidad. El estudio destaca la utilidad de estas técnicas para el diseño de políticas de abastecimiento.
* Diferencia con este proyecto: Aunque emplea LDA como en el presente proyecto, no contempla el uso complementario de regresión logística ni el desarrollo de un sistema de scoring automatizado. Además, el enfoque geográfico es distinto.

## Descripción del objeto de estudio

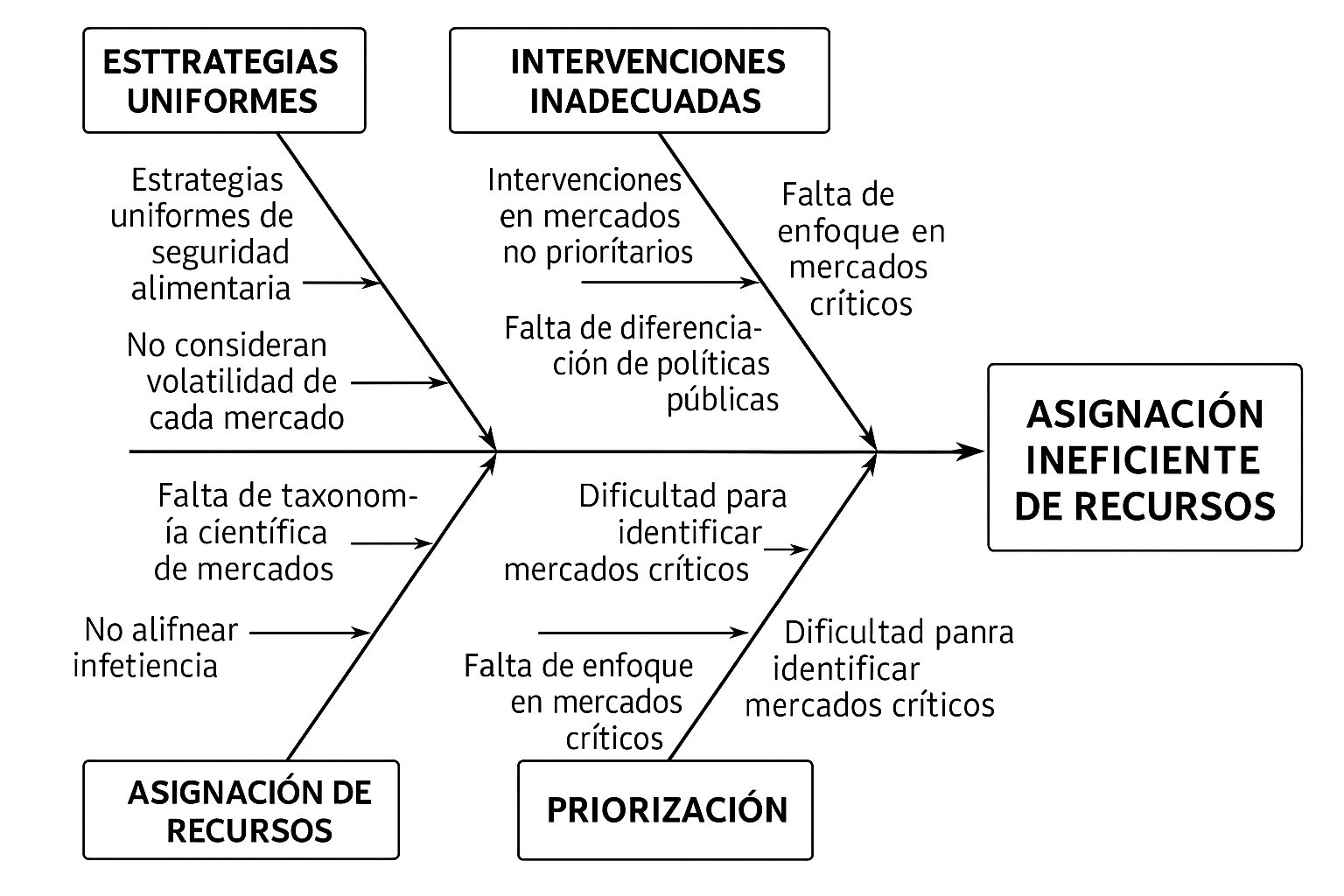
El objeto de estudio está constituido por los mercados locales de productos alimenticios de Bolivia, registrados por el Programa Mundial de Alimentos. Estos mercados representan un componente esencial de la red de abastecimiento de alimentos y reflejan la realidad económica de las regiones donde operan. La clasificación de estos mercados según su volatilidad de precios permitirá segmentarlos para el diseño de intervenciones diferenciadas en contextos de crisis, inflación o escasez alimentaria.

## Identificación del Problema

El principal problema identificado es la ausencia de un sistema de clasificación de mercados bolivianos que considere el comportamiento histórico de precios como factor clave para la toma de decisiones. En la actualidad, las medidas de intervención y subsidio alimentario se aplican de forma uniforme, sin considerar que algunos mercados tienen precios estables y predecibles, mientras otros están expuestos a alta volatilidad. Esto genera asignaciones ineficientes de recursos y disminuye el impacto de las políticas de seguridad alimentaria.

Las causas subyacentes incluyen la falta de herramientas analíticas aplicadas a los datos disponibles, la carencia de clasificadores automáticos de riesgo y la inexistencia de una metodología empírica para diferenciar entre mercados de bajo, medio y alto riesgo por volatilidad.

**Diagrama Causa-Efecto:**



## Formulación del Problema

Considerando la complejidad del sistema de precios alimentarios en Bolivia se formula la siguiente pregunta de investigación:

**¿Cuál es la forma más apropiada para clasificar automáticamente los mercados alimentarios bolivianos según su nivel de volatilidad de precios, utilizando técnicas de machine learning como el Análisis Discriminante Lineal y la Regresión Logística?**

## Objetivos

### Objetivo General

Desarrollar un modelo de clasificación automatizado que permita categorizar los mercados alimentarios bolivianos según su nivel de volatilidad de precios, utilizando técnicas de Análisis Discriminante Lineal y Regresión Logística.

## Objetivos Específicos

* Calcular métricas estadísticas de volatilidad de precios para cada mercado a partir de datos históricos
* Identificar variables predictoras relevantes como ubicación geográfica, tipo de producto y patrones estacionales
* Implementar un modelo de Análisis Discriminante Lineal para establecer clasificaciones iniciales
* Ajustar un modelo de Regresión Logística Multinomial para mejorar la precisión de la clasificación
* Evaluar el desempeño del modelo mediante métricas como precisión, recall, F1-score y matriz de confusión.

## Justificaciones

### Justificación técnica

El uso de modelos supervisados como LDA y Regresión Logística permite construir un sistema reproducible y validado empíricamente, apoyado en métricas objetivas derivadas de datos históricos. Este enfoque facilita la interpretabilidad de los resultados y su aplicabilidad en contextos institucionales.

### Justificación social

Este proyecto contribuirá a mejorar la focalización de ayudas en mercados con alta inestabilidad de precios, priorizando regiones vulnerables y promoviendo una mayor equidad en las políticas de seguridad alimentaria.

### Justificación económica

La correcta clasificación de mercados permitirá una asignación más eficiente de los recursos públicos, evitando

desperdicio de fondos en zonas de bajo riesgo y garantizando intervenciones oportunas en contextos de alta volatilidad.

## Límites y Alcances

### Límites

* Se analizará exclusivamente información cuantitativa proveniente del dataset histórico del WFP
* No se incluirán variables macroeconómicas ni indicadores socioeconómicos externos
* El análisis se limitará a mercados con al menos 24 meses de información continua
* La clasificación se restringe a cuatro categorías: muy baja, baja, media y alta volatilidad

### Alcances

* Se utilizará información de 9 mercados bolivianos con cobertura suficiente
* El modelo propuesto será replicable para otros mercados nuevos
* Se creará un sistema de alerta temprana para monitorear cambios de categoría de riesgo
* Se espera obtener una precisión superior al 85% en la clasificación

## Metodología de la investigación

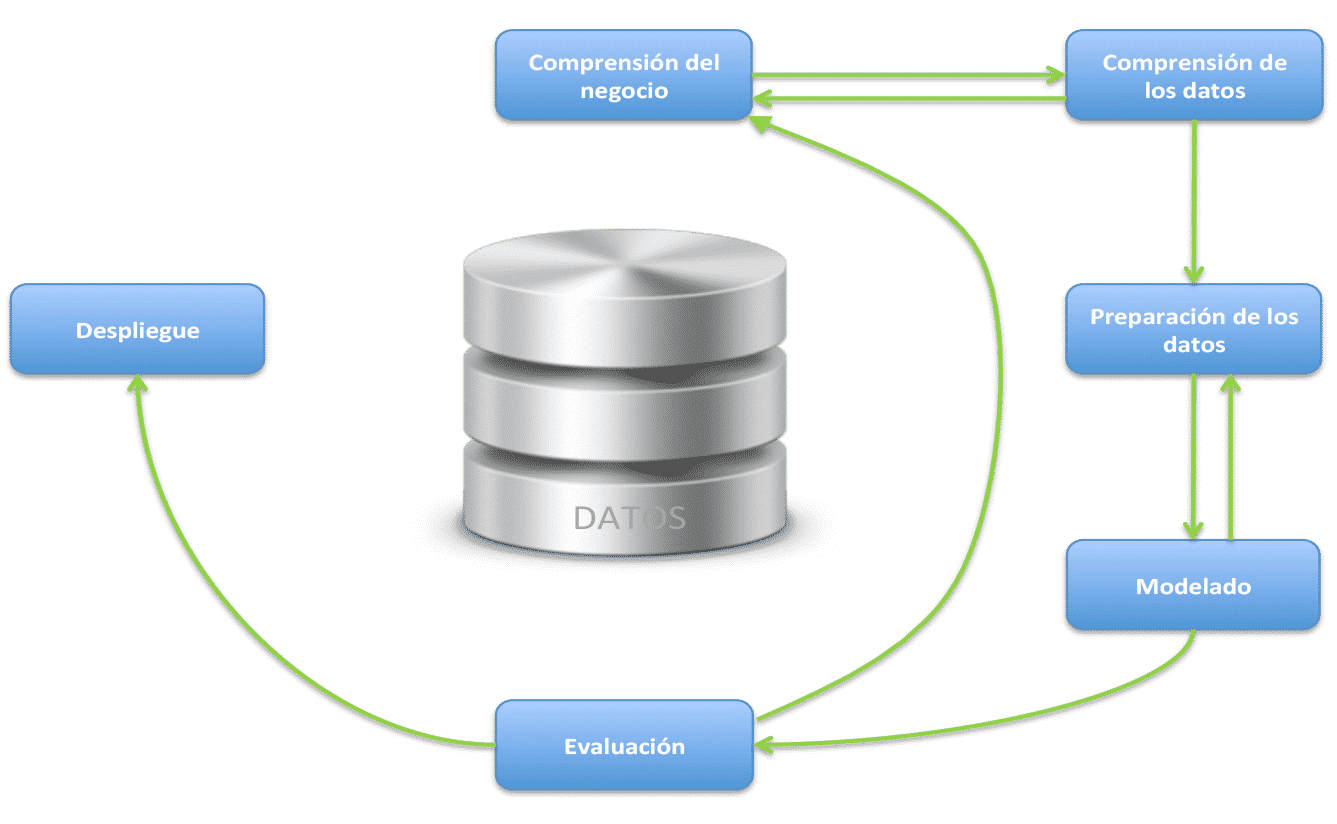
### Tipo de estudio

Estudio cuantitativo, exploratorio y aplicado explicativo y aplicado. Busca explicar patrones observables de comportamiento de precios y aplicar estos conocimientos para diseñar una herramienta funcional.

### Métodos

Se aplicará un enfoque cuantitativo, sustentado en la teoría estadística y técnicas de aprendizaje supervisado. El método incluirá análisis exploratorio, modelado, validación cruzada y evaluación del rendimiento. Asimismo, se adoptará la metodología **CRISP-DM (*Cross-Industry Standard Process for Data Mining*)** como marco estructural del proceso de desarrollo del modelo. Esta metodología contempla seis fases:

1. Comprensión del negocio
2. Comprensión de los datos
3. Preparación de los datos
4. Modelado
5. Evaluación
6. Despliegue.



Esta estructura permitirá asegurar la trazabilidad, replicabilidad y rigurosidad del proceso científico aplicado al análisis de los datos de precios alimentarios en Bolivia

### Técnicas

* Revisión documental de registros históricos de precios de documentos técnicos y bases de datos (WFP)
* Limpieza y análisis de datos con Python y pandas
* Cálculo de volatilidad con estadísticas descriptivas (std, IQR, coef. variación)
* Clasificación contransformación de datos usando Python, pandas y numpy
* Visualización de patrones con seaborn y matplotlib
* Cálculo de indicadores de volatilidad (std, IQR, coeficiente de variación)
* Reducción de dimensionalidad mediante Análisis de Componentes Principales (PCA)
* Entrenamiento y validación de modelos LDA y Regresión Logísticaregresión logística (scikit-learn)

## Análisis preliminar

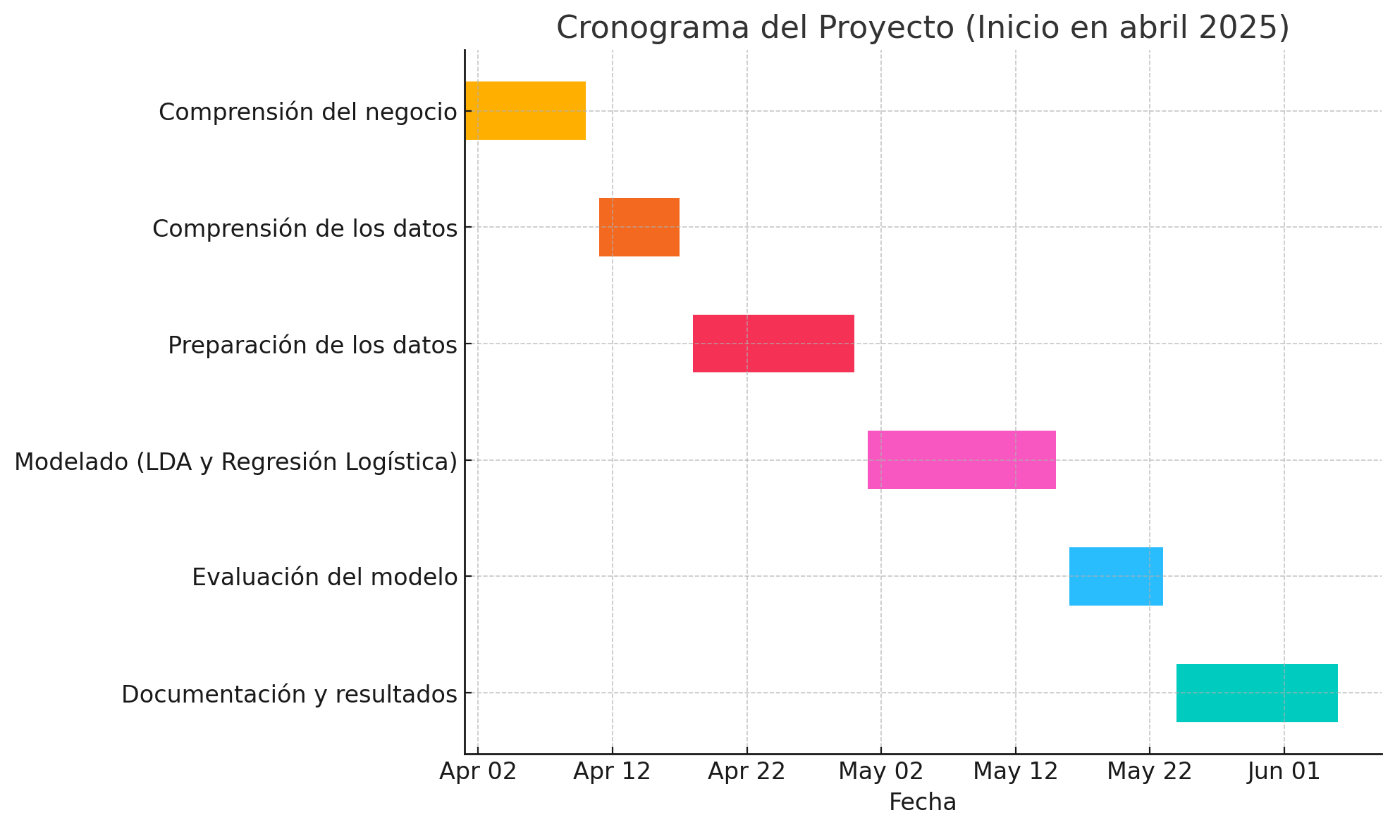
Se procesó una base de datos de 18,095 registros históricos de precios de alimentos en Bolivia. Luego de la limpieza y transformación de los datos, se confirmó que los nueve mercados considerados presentan información mensual continua durante al menos 24 meses, lo cual permite realizar análisis estadísticos fiables. Se observaron diferencias notables en la evolución de precios entre regiones, con mercados como La Paz y Cochabamba mostrando comportamientos más inestables respecto a otros mercados como Sucre o Tarija. Además, se estandarizaron variables clave como precio en USD, tipo de producto, fecha y ubicación geográfica, facilitando la preparación de datos para modelado.

## Propuesta de solución

Se propone desarrollar un modelo de clasificación en cinco etapas:

1. Cálculo de indicadores de volatilidad para cada mercado, incluyendo desviación estándar, coeficiente de variación, rango intercuartílico (IQR) y análisis de estacionalidad;
2. Aplicación de Análisis de Componentes Principales (PCA) para reducir la dimensionalidad del conjunto de datos y capturar factores latentes de variabilidad;
3. Implementación de un modelo de Análisis Discriminante Lineal (LDA) para generar una primera clasificación de los mercados según su nivel de volatilidad;
4. Ajuste de un modelo de Regresión Logística Multinomial para refinar la clasificación y evaluar probabilísticamente la pertenencia de cada mercado a una categoría;
5. Diseño de un esquema de evaluación y validación del modelo, que incluya una posible extensión a un sistema de scoring para categorizar futuros mercados no incluidos en el dataset original.

## Cronograma



# Capítulo 2.- MARCO TEÓRICO

## Base teórica en la que se sustenta el proyecto

# Capítulo 3.- MARCO PRÁCTICO

## Análisis del ámbito de aplicación del proyecto

## Determinación de requerimientos

### Requerimientos funcionales

### Requerimientos no funcionales

## Análisis del proyecto

## Diseño del proyecto

## Desarrollo del Proyecto

## Validación y pruebas

## Resultados

# Capítulo 4.- CONCLUSIONES Y RECOMENDACIONES

## Conclusiones

## Recomendaciones

# Referencias Bibliográficas

# Anexos

# Apéndices

# Glosario