

Reporte de Proyecto: Análisis de Costo de Vida en el Sector Minorista

Autor: Hendy Angel González Marrero

Asignatura: Introducción a la Ciencia de Datos

1. Problemática y Objetivos

El escenario económico en Cuba está marcado por una dualidad compleja: una mayor disponibilidad de productos a través de las Micro, Pequeñas y Medianas Empresas (MIPYMES) frente a una pérdida del poder adquisitivo del salario estatal.

Este trabajo busca responder a las siguientes interrogantes:

- **Capacidad de Compra:** ¿Es suficiente el salario medio de referencia para cubrir una canasta básica de 13 productos esenciales?
- **Volatilidad de Precios:** ¿Cómo fluctúa el precio de un mismo producto entre diferentes establecimientos y municipios (Plaza vs. Playa)?
- **Impacto Relativo:** ¿Qué porcentaje del ingreso mensual representa la adquisición de una unidad de cada producto de la canasta analizada?

2. Obtención y Estructura de Datos

Se analizan tres pilares de información integrados en un repositori:

- **Salario de Referencia:** Datos extraídos de la Oficina Nacional de Estadística e Información (ONEI), fijando el salario medio en La Habana en **6,449 CUP**.
- **Investigación de Campo:** Recopilación de precios de 13 productos en **53 MIPYMES** distribuidas en los municipios de Plaza de la Revolución y Playa.
- **Estructura del repositorio:**

MIPYMES/

├── Evidencia/

| ├── El yerro menu/

Capturas del sitio web de las mipymes

| ├── ONEI/

Captura del ultimo reporte sobre el salario

| ├── Playa/

Registro fotográfico de establecimientos en Playa

| └── Plaza/

Registro fotográfico de establecimientos en Plaza

├── data/

Almacenamiento de datos estructurados

| └── General.json

Dataset principal (Salarios, Canasta, Precios)

├── src/

| └── icd_module.py

Funciones para el procesamiento estadístico

— main.ipynb	# Análisis visual y flujo de trabajo principal
— Informe.pdf	# Se profundiza en la problemática y etapas del trabajo
— README.md	# Se explica un poco sobre el proyecto
— requirements.txt	# Requisitos para reproducir el proyecto (numpy, matplotlib)

- **Estructura del Dataset (General.json):** Los datos se organizan de forma jerárquica, permitiendo filtrar por municipio e identificando productos ausentes mediante valores nulos (`null`), lo que garantiza que los promedios reflejen la oferta real disponible.

El archivo Json podemos dividirlo en tres partes:

1- Tomamos el salario promedio de un trabajador de La Habana en el año 2024

```
"salario_referencia": {
  "tipo": "salario_medio_en_la_habana",
  "valor": 6449,
  "moneda": "CUP",
  "fuente": "ONEI",
  "enlace":
"https://www.onei.gob.cu/sites/default/files/publicaciones/2025-04/salario-medio-en-cifras-2024-edicion-2025.pdf",
  "periodo": "2024"
},
```

2- La canasta básica conformada por 13 productos esenciales para la alimentación

```
"canasta_basica": [
  {
    "producto": "Arroz",
    "cantidad": 1,
    "unidad": "kg"
  },
  {
    "producto": "Azucar",
    "cantidad": 1,
    "unidad": "kg"
  }
]
```

...

3- Los datos recopilados de los 13 productos en 53 mipymes de 2 municipios (plaza y playa)

```
"municipios": {
  "Plaza": {
    "mipymes": [
      {
        "id": "z001",
        "localizacion": "23.1376950, -82.3846020",
        "red social": null,
        "precios": {
```

```
"Arroz": 825,  
"Azucar": null,  
"Sal": 650,  
"Vinagre": 450,  
"Frijoles": 850,  
"Espagueti": 400,  
"Salsa de tomate": 600,  
"Aceite": null,  
"Harina": null,  
"Leche condensada": 600,  
"Mayonesa": 1500,  
"Leche en polvo": null,  
"Carton de huevos": null  
}
```

...

3. Lenguaje y Librerías Utilizadas

El desarrollo se realizó íntegramente en **Python 3.x**, utilizando las siguientes librerías especializadas:

- **NumPy**: Para el procesamiento de vectores numéricos y el cálculo de estadísticas descriptivas (medias y dispersión).
- **Matplotlib**: Para la generación de visualizaciones de alto impacto, incluyendo gráficos de dispersión con técnica de *jitter* y gráficos de barras comparativos.
- **JSON**: Para la gestión de la base de datos local, permitiendo una carga de datos estructurada y eficiente.

4. Procedimientos

Para garantizar la precisión del análisis, se implementaron los siguientes procesos técnicos:

1. **Normalización de Precios**: Debido a la irregularidad en los inventarios, el sistema ignora los valores nulos al promediar, evitando sesgos por falta de producto.
2. **Cálculo de la Canasta Básica**: Se iteró sobre la lista de 13 productos definida en el JSON para obtener un "Costo Total del Carrito", sumando los precios promedios de cada artículo.
3. **Tratamiento de Dispersión**: Se utilizó un factor de desplazamiento aleatorio (*jitter*) en las gráficas de dispersión para visualizar la densidad de precios en un mismo punto, permitiendo identificar visualmente la "competencia" entre locales.
4. **Experimento de Supervivencia**: Se proyectó la duración del salario en días basándose en el consumo diario de productos específicos (como el arroz o la pasta), insertando líneas de referencia en el umbral de los 30 días para contextualizar la realidad del usuario.

5. Resultados y Conclusiones

El análisis técnico permitió concluir lo siguiente:

- **Insuficiencia Salarial:** La adquisición de los 13 productos básicos (1 unidad de cada uno) consume más del 210% del salario de referencia (dos veces el salario), evidenciando que el ingreso estatal es insuficiente para cubrir la alimentación básica fuera del sistema subsidiado.
- **Estrategia de Búsqueda:** La alta dispersión encontrada (precios que varían significativamente a pocas cuadras de distancia) demuestra que la optimización del presupuesto depende directamente de la capacidad del consumidor para comparar múltiples ofertas.