

Informe Proyecto Ciencia de Datos 1er Año

Semestre: Primero

Estudiante: Fabian Enrique Morales

Fecha del Proyecto: Octubre 2025 - Enero 2026

Puntos Clave:

1. Problemática escogida y objetivo del proyecto.
2. Adquisición y estructuración de los datos.
3. Procesamientos realizados.
4. Lenguajes y librerías implementadas.

1: Problemática escogida y objetivo del proyecto

- ◆ La problemática escogida hace referencia a la situación del cubano promedio en cuanto a la economización de ingresos y la incertidumbre de la vía de ahorro más efectiva, causada por la marcada dualidad monetaria en el país. Abordar este tema permite abarcar un amplio campo de información y llegar a una síntesis de datos de manera orgánica.
- ◆ Este Proyecto busca, en parte, ayudar a aquellas personas que no tienen un plan de ahorro y que cuentan con las posibilidades de resguardar parte de sus ingresos. Sin embargo, su objetivo fundamental es plasmar la difícil situación que afronta el ciudadano de hoy en día y como ahorrar pasa a un segundo plano.

2: Adquisición y estructuración de los datos

Se tuvo en cuenta las siguientes tres fuentes de información:

1. Productos de 30 mipymes.
2. Salarios a nivel estatal.
3. Tasa de cambio del dólar en el mercado informal.

1- Las Mipymes escogidas para el análisis se buscaron en dos días, 16 de ellas el 10 de Noviembre del 2025, y las 14 restantes el 21 de Noviembre del mismo año, realizando fotos a los productos que se encontraban en cada una de ellas y posteriormente rellenando los json con la estructura siguiente:

```
{
  "id": "",
  "ubicación": {
    "municipio": "",
    "coordenadas": ""
  },
  "productos": [
    {
      "id_producto": "",
      "especificaciones": {
        "tipo": "",
        "marca": "",
        "cantidad": 0,
        "precio": 0
      }
    }
  ]
}
```

- El “id” tiene como valor el número X correspondiente de la Mipyme de la forma MPX , para la Mipyme no9 su id sería MP9, el nombre del archivo json coincide con su id.
- La “ubicación” posee un diccionario como valor ya que almacena el municipio y las coordenadas correspondientes, se decidió hacer de esta forma ya que con solo las coordenadas no sería del todo claro el municipio al que pertenece la Mipyme. Aclarar que la ubicación fue extraída de los metadatos de las fotos realizadas en cada Mipyme.
- “Productos” tiene asociada una lista, esta lista tiene como valores diccionarios, cada diccionario es un producto con su id y especificaciones, las cuales son a forma de diccionario nuevamente con su tipo, marca, cantidad y precio.

Cada producto con su id correspondiente se encuentra en un json aparte llamado Productos, el cual posee la siguiente estructura:

```
[
  {
    "id": "",
    "nombre": "",
    "unidad": ""
  }
]
```

- El json tiene asociada una lista, la cual tiene todos los productos utilizados, y aparte del id posee la unidad de medida de cada uno, ya sea ml, gr o ninguna. Se decidió utilizar este método para no tener que repetir el nombre del producto en los json de las Mipymes y evitar posibles errores a la hora de rellenarlos.

2- Los datos salariales fueron extraídos manualmente de la cuenta oficial de la ONEI de Facebook en una publicación del día 12 de Noviembre del 2025, donde se encuentran los salarios medios por actividad económica del sector empresarial y presupuestado en septiembre de ese mismo año, la estructura del json con el nombre Salario_Mensual_Estatal es la siguiente:

```
{
  "Fuente": "",
  "Datos": "",
  "Fecha de publicación": "",
  "Salarios por actividad económica":{
    "Transporte , almacenamiento y comunicaciones": 0
  }
}
```

- “Datos” tiene como valor la fecha de esos salarios, anteriormente mencionada (Septiembre de 2025).

- “Salarios por actividad económica” tiene asociada un diccionario donde cada llave es una actividad económica que tiene como valor su salario medio; lo que se muestra en la estructura es una de esas actividades.

3- La tasa de cambio del valor del dólar se utilizó mediante el Toque en los meses de Enero hasta Noviembre del año 2025, se extrajo de forma manual en la sección de comportamiento histórico y se almacenó en un json llamado USD_2025 con la siguiente estructura:

```
{
  "Enero":[
    { "día": 1, "valor": 0},
    { "día": 2, "valor": 0}
  ],
  "Febrero":[
    { "día": 1, "valor": 0},
    { "día": 2, "valor": 0}
  ]
}
```

- El Json tiene asociado un diccionario donde cada llave es un mes, cada mes tiene asociada una lista donde cada día es un diccionario con su número y el valor del toque correspondiente del dólar respecto al CUP.
- La longitud de cada lista coincide con la cantidad de días que tiene el mes a la que está asociada.

3: Procesamientos realizados

El proyecto contiene una librería de Python donde se crearon todas las funciones para el procesamiento e implementación de los datos recopilados. Estas funciones se dividen en dos categorías: Cálculo y Graficación. Aclarar que las últimas mencionadas no reciben ningún parámetro.

Funciones de Cálculo y Procesamiento:

1.disponibilidad: Esta recibe como parámetro el id de un producto en específico y devuelve una lista con las Mipymes en las que se encuentra.

2.medianaP: Recibe como parámetros el id del producto al que se le quiere hacer la mediana, el tipo el cual puede ser o “precio” o “cantidad”, dependiendo de este parámetro se calcula la mediana correspondiente, y por último un parámetro booleano el cual si es True devuelve el valor de la mediana como un número entero y si es False devuelve un string con el valor y su unidad correspondiente, ya sea CUP en el caso del precio, y ml o gr en el caso de la cantidad.

3.mediana_varios: Recibe como parámetros lo mismo que la anterior pero en vez de recibir un solo id, recibe una lista con múltiples ids y devuelve una lista con todas las medianas correspondientes, esta función utiliza a la anterior para todos los cálculos .

4.minmax : Primer parámetro puede ser o “salarios” o “toque”, en caso de ser “salarios” devuelve una tupla con la actividad económica con menor salario medio y la de mayor. En caso de ser “toque”, se observa el segundo parámetro de la función el cual es el mes en el que se quiere ver el menor y mayor valor que tuvo el dólar.

5.promed: El mismo procedimiento que la función anterior pero en vez de devolver un tupla con valor mínimo y máximo devuelve el promedio, ya sea de los salarios o del valor del dólar en el mes especificado en el parámetro, al no especificar mes se devuelve una lista con los promedios de todos los meses.

6.anualT: No posee parámetros, realiza el promedio del valor del dólar en el Toque a lo largo de los meses de Enero hasta Noviembre del 2025.

7.salario_to_USD: A esta función se le ingresa primeramente un mes, la conversión de un salario de CUP a USD se hará tomando como referencia el promedio del valor del dólar en el Toque en el mes ingresado, el otro parámetro será para especificar el sector al cual se le quiere hacer la conversión del salario medio. En caso de no especificar,

se hará la conversión del promedio salarial, y en el caso de poner “todos”, se devolverá una lista con la conversión de todos los sectores.

8.porciento: Esta devuelve que tanto porciento aumentó o decreció el valor del dólar en un mes en específico, mes que se introduce en el primer parámetro, y en el segundo parámetro se pone “corto plazo” o “largo plazo”, en el primer caso devuelve la variación del valor con respecto al mes anterior, y en el segundo caso con respecto a Enero.

Todas las funciones obtienen primeramente los datos que serán utilizados y posteriormente realizan la operación pertinente, ya sea mediana, promedio, porcentajes, etc.

Funciones de Graficación:

1.grafToque: Plasma la variación del valor de dólar en el Toque a lo largo de los meses de Enero hasta Noviembre de 2025 (gráfica de líneas).

2.grafProductos: Refleja la diferencia entre el precio de productos básicos respecto al salario del cubano promedio (gráfica de barras verticales).

3.grafSalariosUSD: Muestra a cuantos dólares equivale el salario medio por cada actividad económica respecto al valor promedio en el Toque en el mes de Noviembre de 2025 (gráfica de barras horizontales).

4.grafValor_USD: Evidencia como el salario promedio del cubano son cada vez menos dólares tomando como referencia el valor promedio en los meses de Enero hasta Noviembre de 2025 en el Toque (gráfica de barras verticales).

5.grafPrecio_Peso: Posiciona los productos recopilados en las Mipymes donde el eje (x) es el peso/volumen medio de dicho producto y el eje (y) su precio medio (gráfica interactiva).

4: Lenguajes y librerías implementadas

Lenguajes:

- Jupyter Notebook
- Python

Librerías externas:

- Matplotlib
- Plotly

Librerías no externas:

- json
- os