



Pontificia Universidad
JAVERIANA
Bogotá

Pontificia Universidad Javeriana

Asignatura: PROCESAMIENTO DE DATOS A GRAN ESCALA

Docente: MIGUEL ÁNGEL MÉNDEZ HERNÁNDEZ

PROYECTO: ENTREGA 1

Integrantes:

Nicolás Barragán Sánchez

nbarragans@javeriana.edu.co

Jessica Tatiana Naizaque Guevara

j.naizaque@javeriana.edu.co

Angie Tatiana Peña Peña

penap_at@javeriana.edu.co

Juan Sebastián Mateo Ruiz Bulla

ruizju@javeriana.edu.co

Sergio Esteban Triana Bobadilla

[triana_se@javeriana.edu.co](mailto: triana_se@javeriana.edu.co)

Septiembre 2021

Índice

1. Introducción	2
2. Entendimiento de Negocio	2
3. Selección de Datos	2
4. Colección y descripción de Datos	3
4.1. Conjunto de datos - arrestos	3
4.2. Conjunto de datos - pobreza	3
4.3. Conjunto de datos - accidentes	4
4.4. Conjunto de datos - educación	4
4.5. Conjunto de datos - calidad del aire	5
4.6. Conjunto de datos - Población promedio anual de los vecindarios de la ciudad de Nueva York	5
5. Exploración de Datos	5
6. Reporte de calidad de Datos	6
7. Planteamiento de preguntas	6
8. Filtros, limpieza y transformación inicial	7
9. Web Scraping	8
10. Api información climática	8

1. Introducción

Nueva York, oficialmente New York City o NYC en sus siglas es, al año 2021, la ciudad más poblada de los Estados Unidos, con un área urbana de aproximadamente 24 millones de habitantes. Adicionalmente, es uno de los centros en política y economía mundial, razón por la cual, durante el desarrollo de este informe, se busca conocer la calidad de vida de la población. Esto, por medio de datasets que identifican factores importantes para esta y así, generar un plan de acción que permita mejorar en un porcentaje la calidad de vida de los habitantes de la ciudad de Nueva York.

2. Entendimiento de Negocio

Estados Unidos es un país de gran extensión, con muchas ciudades grandes, dentro de las que se encuentra Nueva York. Esta, sobresale por ser la ciudad más poblada del país teniendo 24 millones de habitantes, agregando que, por ser una ciudad icónica y considerada por muchos de las más importantes, es muy visitada tanto por turistas como por gente de negocios que no vive en la ciudad. (1)

Sin embargo, como todas las ciudades por su alto flujo de personas, Nueva York presenta retos y dificultades ambientales, de movilidad vehicular, inseguridad, entre otras, que se deben solucionar para mejorar la calidad de vida de sus habitantes. (2)

Dentro de los problemas más grandes para la ciudad, se encuentra la criminalidad. Para abril del año 2020, la policía de NY reportaba más de 335 arrestos por asesinatos y tiroteos. Cifra que genera un crecimiento del 36.5% del índice de delincuencia respecto al año anterior. Este se ve impulsado por el robo de automóviles que se incrementó en un 59.4%. Junto con otros delitos como tiroteos, incautaciones de armas y, de una nueva moda de armas fantasma, que se pueden imprimir en 3D en casa si se tiene el equipo necesario. (3)

De la misma manera, los accidentes viales han incrementado. Para el año 2018, hubo 228.047 accidentes de tráfico repartidos a lo largo de los 5 distritos de la ciudad. Visto de otra manera, son 19.000 accidentes al mes y 678 por día. Una cifra muy grande que va de la mano de la alta cantidad de personas que transitan todos los días a lo largo del territorio de la ciudad. (4)

Los accidentes, la seguridad y muchos otros factores han incrementado drásticamente para este año, llegando a tener un alza histórica por accidentes viales. (5) Por esto, dentro del desarrollo del presente proyecto, el objetivo es aumentar en al menos un 20% la calidad de vida de los habitantes de Nueva York durante el primer año del desarrollo del plan de acción. Esto mediante el estudio de diferentes indicadores que guiarán hacia medidas con el fin de obtener soluciones que puedan cambiar la manera en cómo viven los habitantes de la ciudad de Nueva York.

3. Selección de Datos

De acuerdo con lo descrito en la sección 2, fue necesario hacer una investigación de los factores que determinaban la calidad de vida de los habitantes de Nueva York. Por lo tanto, los datasets seleccionados para trabajar con el proyecto fueron:

1. Datos de los arrestos en Nueva York. Tomados de:
<https://data.cityofnewyork.us/Public-Safety/NYPD-Arrest-Data-Year-to-Date-/uip8-fykc>
2. Datos de pobreza en Nueva York. Tomados de:
<https://data.cityofnewyork.us/City-Government/NYCGov-Poverty-Measure-Data-2018-/cts7-vksw>
3. Datos de accidentes viales en Nueva York. Tomados de:
<https://data.cityofnewyork.us/Public-Safety/Motor-Vehicle-Collisions-Vehicles/bm4k-52h4>
4. Datos de niveles de educación en Nueva York. Tomados de:
<https://data.cityofnewyork.us/Education/2016-2017-Health-Education-Report/2dzy-e7cu>
5. Datos de vigilancia de la calidad del aire en Nueva York. Tomados de:
<https://data.cityofnewyork.us/Environment/Air-Quality/c3uy-2p5r>

6. Datos Web Scraping Población promedio anual de los vecindarios de la ciudad de Nueva York. Tomados de:

<https://www.health.ny.gov/statistics/cancer/registry/appendix/neighborhoodpop.htm>

Los cuatro datasets iniciales (arrestos, pobreza, accidentes y educación) son importantes para el presente estudio porque son categorías que de una u otra forma influyen en la calidad de vida que presentan los habitantes de Nueva York. Es posible decir que una ciudad con pocos arrestos, pobreza y accidentes tiene mayor calidad de vida. Además, si presenta niveles de educación bastante altos quiere decir que le aporta mucho más a la calidad de vida de cada individuo.

Adicional a esto, se seleccionó un conjunto de datos de vigilancia de la calidad del aire, esto, porque es un factor que afecta directamente a la calidad de vida. Cuando una persona se expone a niveles altos de contaminación del aire, es posible que más adelante pueda tener diferentes enfermedades o similares que afecten su salud y, por lo tanto, la calidad de su vida.

Finalmente, se realiza un desarrollo de Web Scraping para analizar los datos de los datasets anteriores respecto a la población promedio de los vecindarios de la ciudad de Nueva York.

4. Colección y descripción de Datos

Teniendo en cuenta lo mencionado en la sección 3, a continuación, se presenta la descripción de datos de los 6 conjuntos de datos a utilizar para alcanzar el objetivo de negocio. En esta descripción se muestran los tipos de datos, el significado de cada atributo y una descripción general del conjunto de datos. Considerando que en algunos conjuntos de datos se dispone una alta cantidad de atributos, en el desarrollo de esta sección, se muestran solo los más importantes para el objetivo de negocio.

4.1. Conjunto de datos - arrestos

En este conjunto de datos, cada registro representa una detención efectuada en NYC por la policía de Nueva York e incluye información sobre el tipo de delito, el lugar y la hora de la ejecución. Además, se incluye información relacionada con los datos demográficos de los sospechosos.

Dentro de este conjunto de datos, se encuentran los siguientes atributos:

1. *Arrest Date*: Fecha exacta de la detención del suceso denunciado. **Tipo de dato:** *fecha*
2. *PD CD*: Código de clasificación interna de tres dígitos. **Tipo de dato:** *numérico*
3. *PD Desc*: Descripción de la clasificación interna correspondiente al código PD. **Tipo de dato:** *categorico*
4. *Law Cat CD*: Nivel del delito: delito grave (F), delito menor (M), violación (V). **Tipo de dato:** *categorico*
5. *Arrest Bor*: Distrito de detención. B(Bronx), S(Staten Island), K(Brooklyn), M(Manhattan), Q(Queens). **Tipo de dato:** *categorico*
6. *Arrest Precinct*: Comisaría donde se produjo la detención. **Tipo de dato:** *numérico*
7. *Jurisdiction Code*: Jurisdicción responsable de la detención. Los códigos de jurisdicción 0(Patrulla), 1(Tránsito) y 2(Vivienda) representan a la policía de Nueva York, mientras que los códigos 3 y más representan jurisdicciones ajenas a la policía de Nueva York. **Tipo de dato:** *numérico*
8. *Age Group*: Edad del agresor dentro de una categoría. **Tipo de dato:** *categorico*
9. *Perp Race*: Descripción de la raza del agresor. **Tipo de dato:** *categorico*

4.2. Conjunto de datos - pobreza

Este conjunto de datos contiene las tasas de pobreza y los datos relacionados con la medición de la pobreza de NYCgov. La medición de la pobreza de NYCgov es generada anualmente por la unidad de investigación de la pobreza de la Oficina de Oportunidad Económica del Alcalde (NYC Opportunity).

Dentro de este conjunto de datos, se encuentran los siguientes atributos:

1. *Sp Order*: Cantidad de personas en el hogar censado. **Tipo de dato:** *numérico*

2. *CIT*: Estado de la ciudadanía. 1 Nacido en EE.UU. 2 Nacido en Puerto Rico, Guam, las Islas Vírgenes estadounidenses o las Marianas del Norte. 3 Nacido en el extranjero de padres estadounidenses. 4 Ciudadano de EE.UU. por naturalización. 5 No es ciudadano de EE.UU. **Tipo de dato:** *numérico*
3. *ESR*: Registro de la situación laboral. 1 Civil empleado, con trabajo. 2 Civil empleado, con trabajo pero no trabajando. 3 Desempleado. 4 Fuerzas armadas, con trabajo. 5 Fuerzas armadas, con un empleo pero sin trabajar. 6 No está en la fuerza laboral. **Tipo de dato:** *numérico*
4. *TEN*: Tenencia de la vivienda. 1 En propiedad con hipoteca o préstamo (incluye los préstamos con garantía hipotecaria). 2 En propiedad, libre de cargas. 3 Alquilado. 4 Ocupado sin pago de alquiler. **Tipo de dato:** *numérico*
5. *Boro*: Distrito de NYC: 1 Bronx. 2 Brooklyn. 3 Manhattan. 4 Queens. 5 Staten Island. **Tipo de dato:** *numérico*
6. *NYCgov Income*: Ingresos totales estimados de NYCgov que se comparan con el umbral de NYCgov para determinar el estado de pobreza. **Tipo de dato:** *numérico*
7. *NYCgov Pov Stat*: Estado de pobreza de NYCgov. 1 = En la pobreza, 2 = No en la pobreza. **Tipo de dato:** *numérico*
8. *NYCgov Threshold*: Umbral de pobreza de NYCgov. **Tipo de dato:** *numérico*

4.3. Conjunto de datos - accidentes

Este conjunto de datos contiene información de todas las colisiones de vehículos de motor reportadas por la policía en NYC. Contiene detalles sobre cada vehículo implicado en el accidente. Cada fila representa un vehículo de motor implicado en una colisión.

Dentro de este conjunto de datos, se encuentran los siguientes atributos:

1. *Crash Date*: Fecha del accidente. **Tipo de dato:** *fecha*
2. *Vehicle Type*: Tipo de vehículo según la categoría de vehículo seleccionada (ATV, bicicleta, coche/suv, ebike, scooter, camión/autobús, motocicleta, otros). **Tipo de dato:** *categorico*
3. *Travel Direction*: Dirección en la que viajaba el vehículo. **Tipo de dato:** *categorico*
4. *Vehicle Occupants*: Número de ocupantes del vehículo. **Tipo de dato:** *numérico*
5. *Driver License Status*: Estado de la licencia del conductor: Licencia, permiso, sin licencia. **Tipo de dato:** *categorico*
6. *Pre Crash*: Ir en línea recta, hacer un giro a la derecha, adelantar, retroceder, etc. **Tipo de dato:** *categorico*
7. *Contributing Factor*: Factores que contribuyeron a la colisión para el vehículo designado. **Tipo de dato:** *categorico*

4.4. Conjunto de datos - educación

Este conjunto de datos proporciona información sobre el número y el porcentaje de estudiantes programados para al menos un semestre de educación para la salud.

Dentro de este conjunto de datos, se encuentran los siguientes atributos:

1. *School Name*: Nombre de la escuela. **Tipo de dato:** *categorico*
2. *No. of students in grades 9-12*: Cantidad de estudiantes pertenecientes a los grados entre 9° y 12°. **Tipo de dato:** *numérico*
3. *No. of students in grades 9-12 scheduled for at least one semester of health instruction*: Cantidad de estudiantes en los grados 9° a 12° programados para al menos un semestre de educación para la salud. **Tipo de dato:** *numérico*

4. *No. of 16-17 June and August graduates*: Cantidad de graduados en junio y agosto del periodo 2016-2017. **Tipo de dato:** *numérico*
5. *No. of 16-17 June and August graduates meeting high school health requirements*: Cantidad de graduados en junio y agosto del periodo 2016-2017 que cumplieron con los requisitos de educación para la salud. **Tipo de dato:** *numérico*

4.5. Conjunto de datos - calidad del aire

La exposición a contaminantes atmosféricos comunes se ha relacionado con enfermedades respiratorias y cardiovasculares, cánceres y muertes prematuras. Estos indicadores proporcionan una perspectiva a través del tiempo y de las geografías de la ciudad de Nueva York para caracterizar mejor la calidad del aire y la salud en la ciudad.

Dentro de este conjunto de datos, se encuentran los siguientes atributos:

1. *Name*: Nombre del indicador. **Tipo de dato:** *categorico*
2. *Measure*: Cómo se mide el indicador. **Tipo de dato:** *categorico*
3. *Measure Info*: Información (como las unidades) sobre la medida. **Tipo de dato:** *categorico*
4. *Geo Place Name*: Nombre de la zona. **Tipo de dato:** *categorico*
5. *Time Period*: Descripción del tiempo al que se aplican los datos. **Tipo de dato:** *categorico*
6. *Start Date*: Valor de la fecha de inicio del periodo de tiempo. **Tipo de dato:** *fecha*
7. *Data Value*: El valor real de los datos para este indicador, medida, lugar y tiempo. **Tipo de dato:** *categorico*

4.6. Conjunto de datos - Población promedio anual de los vecindarios de la ciudad de Nueva York

La población promedio de la ciudad de Nueva York en sus cinco distritos Bronx, Kings (Brooklyn), New York (Manhattan), Queens y Richmond (Staten Island) durante los años 2015 y 2019.

Dentro del desarrollo de Web Scraping se evidencian los siguientes atributos:

1. *Borough*: Nombre del distrito. **Tipo de dato:** *categorico*
2. *region*: Nombre de la región. **Tipo de dato:** *categorico*
3. *Males*: Número de Hombres dentro de la región. **Tipo de dato:** *numérico*
4. *Females*: Número de mujeres dentro de la región. **Tipo de dato:** *numérico*
5. *Total Population*: Población total en la región. **Tipo de dato:** *numérico*

5. Exploración de Datos

Para la exploración de los datos se hizo uso de diferentes métodos sobre los dataframes en los que se guardaron los datos. Estos no afectan los valores almacenados ni hacen modificaciones internas, únicamente permiten la visualización y entendimiento de la estructura y comportamiento de los datos.

Con el fin de desarrollar correctamente esta sección, se hizo uso de las librerías: seaborn (sns) y matplotlib.pyplot (plt). Los métodos utilizados para la exploración fueron:

- *df.shape*: Permite saber las dimensiones del dataframe en una tupla. El primer elemento hace referencia a la cantidad de registros y el segundo es la cantidad de columnas que presentan los datos.
- *df.info*: Permite conocer información general de los datos almacenados en el dataframe. Principalmente se usa para identificar las columnas, su tipo de dato y la cantidad de registros no nulos por cada atributo.
- *df.head*: Permite visualizar los primeros n registros, mostrando sus atributos y los valores equivalentes en cada caso.

- *arrestos.value_counts*: Permite conocer los valores únicos de cada columna o de todo el conjunto de datos en general, lo anterior depende de su forma de uso.
- *df.describe*: Permite generar descriptivos estadísticos de las columnas numéricas del dataframe. Los estadísticos que presenta son la tendencia central, la dispersión y la forma de la distribución de un conjunto de datos, excluyendo los valores nulos o vacíos.
- *df.corr*: Permite calcular la correlación por pares de columnas en el dataframe, sin tener en cuenta los NA o valores nulos. Además, utilizando los valores de retorno de este método, es posible realizar una tabla que presente muestre gráficamente esta correlación entre columnas mediante colores diferenciales.
- *sns.boxplot(df)*: Permite mostrar una gráfica con todas las gráficas de caja y bigotes para cada uno de los atributos de tipo numérico. La utilidad de esta visualización, es poder identificar en qué casos hay valores atípicos que deban tenerse en cuenta para el procesamiento posterior.
- *count_outliers(df)*: Es un método creado para identificar la cantidad de valores atípicos que tiene cada columna de tipo numérico. Esta cantidad dependerá del valor del rango intercuartil que se le ingrese.
- *df.memory usage*: Permite conocer la memoria que utiliza cada columna del dataframe en bytes.
- *df.duplicated*: Permite encontrar en el dataframe las filas duplicadas, si existen.
- *df.groupby*: Permite agrupar los datos según un criterio que se defina anteriormente. Por ejemplo, si se desean seleccionar ciertos datos en una fecha, sería posible hacerlo mediante el uso de este método.

6. Reporte de calidad de Datos

En la exploración de datos descrita anteriormente, se evidenció que los datasets utilizados tienen una calidad media de los datos. Lo anterior, debido a que se encontró que en los datasets, hay bastantes datos nulos y algunos valores inconsistentes con el tipo de dato que se debería tener. Con el fin de conocer la cantidad total de datos nulos o datos faltantes uniendo todos los datasets utilizados, se podría afirmar que hay alrededor de 50000 datos faltantes, lo que dificulta el trabajo con la información en crudo.

En cuanto a la cantidad de datos nulos, la calidad del conjunto de datos de educación y de calidad del aire es la mejor, no existe ningún dato faltante. Sin embargo, para el dataset de educación, existen columnas que requieren datos numéricos, en los cuales, en algunos casos se almacenan datos basura como cadenas de caracteres de datos sin sentido. Es importante mencionar que el conjunto de datos de accidentes es aquel que presenta una calidad de datos más baja. Lo anterior, se debe a que la cantidad de datos nulos en este caso es bastante grande, lo que genera vacíos en el correcto análisis de la información.

Adicional a esto, aunque para algunas columnas los datos se encontraban en un intervalo adecuado sin superar el rango intercuartil, para algunos de los atributos numéricos de los datasets, se encontraron datos atípicos que pueden llegar a generar inconsistencias en la relación de los datos. La identificación de estos datos fue posible gracias a un método generado por los estudiantes que con los datos necesarios mostraba la cantidad de datos atípicos encontrados para la columna específica en el dataframe. También, los diagramas de caja y bigotes fueron una herramienta demasiado útil para visualizar estos mismos datos atípicos.

De igual manera se plantea realizar un análisis de los datos necesarios y relevantes en cada dataset para lograr el objetivo planteado. Para lo cual, se proponen distintas soluciones como: eliminar las columnas irrelevantes, las filas con la mayoría de sus datos nulos; transformar los datos teniendo en cuenta el diccionario de datos con la información de las columnas de los datasets y realizar filtros por edad, estado, fecha según corresponda. Adicionalmente, se propone estandarizar los datasets de modo que permita optimizar las consultas y obtener los resultados deseados.

7. Planteamiento de preguntas

A continuación, para dar un enfoque más sólido a la investigación se presentará una serie de preguntas que servirán como guía dentro del análisis de datos para poder llegar a conclusiones que contribuyan al cumplimiento de nuestro objetivo principal.

- ¿En qué temporada hay una afectación mayor en la calidad del aire y cuál es la zona más afectada?
- ¿Cuál es la relación de la pobreza con la cantidad de individuos dentro del hogar?
- ¿Cuál es el porcentaje de hombres y mujeres de la población que recurren en delitos menores?
- ¿Son las festividades factores del aumento de accidentes en la ciudad de Nueva York?
- ¿La población promedio es directamente proporcional con la falta de educación, la cantidad de arrestos y la cantidad de accidentes en la ciudad de Nueva York?
- ¿Son las instituciones educativas aptas en su capacidad (aforo) para la población promedio de su distrito?
- ¿Son las instituciones educativas aptas en su capacidad (aforo) para la población promedio de su región?
- ¿Cuáles son los tres distritos con más personas desempleadas?
- ¿Cuál es la relación que tienen los índices de pobreza con los índices de educación?
- Aumentar el nivel de educación promedio de los habitantes de Nueva York, ¿reduciría la cantidad de arrestos en la ciudad?

8. Filtros, limpieza y transformación inicial

Para la limpieza, filtración y transformación de los datos se hizo uso de diferentes métodos y herramientas, que permiten la visualización y, adicionalmente, la modificación de los mismos.

Para el desarrollo de esta sección se hizo uso de los siguientes métodos:

- *df.drop*: Permite eliminar registros, ya sea por columnas o filas. El primer atributo indica el dato a eliminar ya sea el índice de una fila o columna. El atributo *axis*, permite identificar si se eliminará una fila (0) o columna (1). Y, el atributo *inplace* es un booleano que indica si se reemplazará en el dataset original o se realizará una copia.
- *df.isnull()*: Permite identificar los valores nulos dentro de un dataset, este método devuelve un dataset con valores booleanos que identifican si la celda está vacía (True) y (False) si no lo está.
- *df.isnull().sum*: Permite conocer la cantidad de valores nulos por columnas o filas en un dataset, el parámetro *axis* indica si se realiza la consulta por columnas (1) o filas (0).
- *df.fillna*: Permite sustituir los valores nulos de una estructura pandas por otro valor según ciertos criterios: pueden sustituirse por un valor concreto o bien puede utilizarse el anterior o posterior valor no nulo (en el caso de los dataframes habrá que especificar el eje sobre el que queremos aplicar la función).
- *df.isin*: Permite comprobar si cada elemento de la llamada DataFrame contiene el valor especificado en los valores de entrada. El método devuelve un DataFrame de booleanos de la misma dimensión del llamador DataFrame indicando si cada elemento contiene los “valores” de entrada.
- *df.loc*: Permite localizar filas y/o columnas en el dataset.
- *df.replace*: Permite reemplazar valores dentro del dataset, el primer parámetro indica el valor que se desea reemplazar y el segundo indica el valor por el cual se reemplazará en el dataset.
- *df.style.applymap*: Permite asignar un estilo específico para el dataset, como resaltar valores, colorear celdas, entre otros, la función recibe un atributo de tipo propiedad CSS.

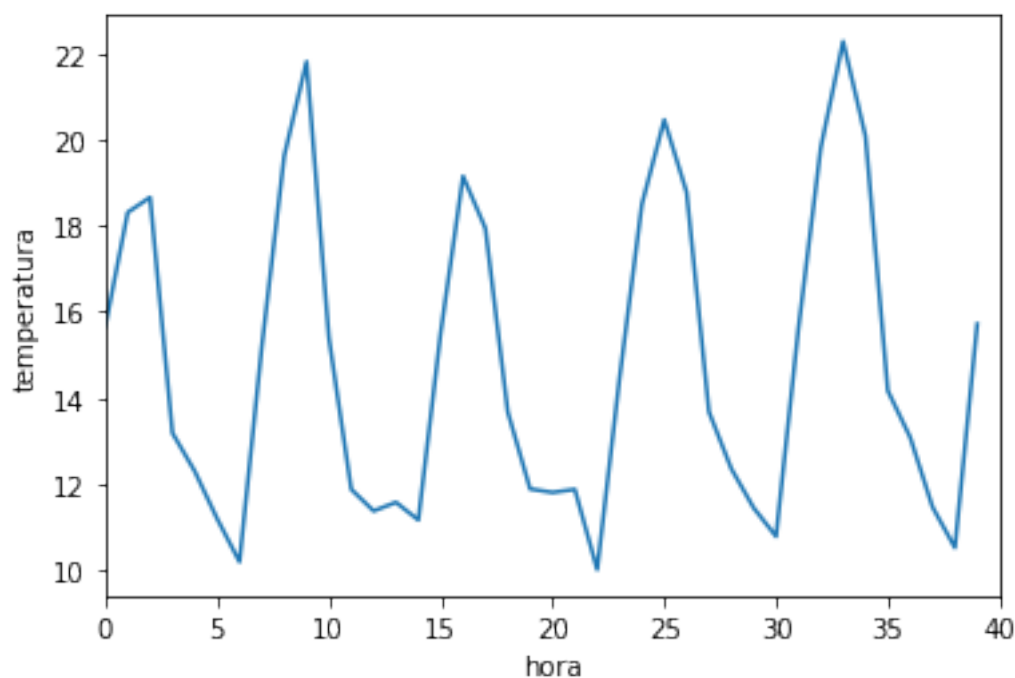
9. Web Scraping

El Web Scraping o raspado web, es una técnica utilizada para extraer información de sitios web, enfocada en la transformación de datos sin estructura, como lo es el HTML, en datos estructurados que puedan ser almacenados y analizados.

Para el desarrollo de esta sección se tomó la página web indicada en la sección de 2 en el numeral 6. Esta contiene información acerca de la población promedio en los vecindarios de la ciudad de Nueva York. Durante la ejecución, se realizó un entendimiento del formato HTML que compone la página, para así identificar los datos importantes para el estudio y extraerlos haciendo uso de los siguientes métodos de la librería Pandas:

- *pd.read_html*: Permite leer una página web en formato html, obteniendo las tablas en formato de dataframes para su uso y análisis, este método recibe como parámetro la URL de la página web
- *pd.merge*: Permite unificar dos dataframes, ya sea con una columna en común o no, recibe como parámetros los dos dataframes y la columna en común si existe.
- *df.groupby*: Permite agrupar los datos de un dataframe, tiene una funcionalidad semejante a la vista para series, con los condicionantes propios de los dataframes: es necesario indicar el eje que contiene el criterio por el que se va a realizar la agrupación.

10. Api información climática



Para esta sección se utilizó el api de información climática de la ciudad de Nueva York. Los datos se organizan en 40 registros, la primera fila representa la hora cero y se incrementan tres horas en cada iteración de los registros. De acuerdo con lo anterior, es posible afirmar que el dataset contempla un intervalo de 120 horas en sus datos almacenados. Algunos datos que se pueden observar dentro de la información, hacen referencia a la humedad, la visibilidad y la temperatura. Esta última tiene una relación directamente proporcional con las horas, pues a medida que aumentan las horas, aumenta la temperatura, por lo que fue seleccionada para hacer una visualización de datos.

Referencias

- [1] ciudad de Nueva York, Nueva York: La emocionante ciudad de atracciones icónicas.
- [2] Los Angeles Times. Los principales retos de Nueva York para 2022, December 2021. Section: EEUU.
- [3] arrestos aumentan a la par de la criminalidad en NYC, según estadísticas del NYPD – Telemundo New York (47).
- [4] injury_doctors. Estadísticas de accidentes de tráfico en NYC - ¿Qué barrio de NY es el más peligroso?, February 2020.
- [5] Andrés Correa. Las calles de Nueva York son cada vez más mortales: alarma por alza histórica de accidentes viales, August 2022.