

Entrega 1 de proyecto

Plan de Acción Para el Estado de Nueva York

Tomas Silva
Juan Pabon Vargas



Pontificia Universidad
JAVERIANA
— Colombia —

Pontificia Universidad Javeriana

15 de Octubre 2025

Presentación.....	2
Entendimiento del negocio.....	3
Selección de datos.....	3
Colección y descripción de datos.....	4
Tipos de datos.....	4
Significado de atributos.....	4
Descripción de contenido.....	4
Exploración de datos.....	4
Reporte de calidad.....	5
Preguntas de negocio.....	5
Filtros y transformación inicial.....	5
Conclusiones.....	5
Resumen.....	5
Referencias.....	5

Presentación

El presente documento data de un informe de los procedimientos, selecciones, exploraciones y filtrados de los datos a trabajar durante los acontecimientos del proyecto de la materia “Procesamiento de datos a gran escala” para el semestre 2530 año 2025. Este informe detalla los procesos de selección, exploración y filtrado de datos realizados mediante un cuaderno de python (ipynb) ejecutado y trabajado mediante un cluster Spark a través de Jupyter. Se plantea inicialmente un problema con contexto basado en un caso de negocio, este mismo basado en el estado actual de la ciudad de Nueva York, y la necesidad de analizar datos de el estado de la ciudad para formar un plan de acción y así mejorar algunas de sus estadísticas territoriales, esta problemática se explicará más a fondo durante el **Entendimiento del negocio**, tras esto, se discuten los datos a trabajar, contando con descripciones precisas y datos específicos y filtrados con tal de tener bases sólidas y fuertes para el futuro trabajo de análisis.

Entendimiento del negocio

Dado el contexto a realizar, el equipo de trabajo ha sido contratado por el estado de Nueva York con el **objetivo de generar un plan de acción basado en procesamiento de datos a gran escala**, con el punto de utilizar dicho plan de acción para mejorar indicadores territoriales de interés para el equipo de gobierno. Con tal de obtener resultados válidos y contar con recursos amplios, se le ha ofrecido al equipo algunos conjuntos de datos en formato .csv, provenientes de fuentes ofrecidas por el Estado de Nueva York en sí mismo.

Los indicadores que más aquejan al equipo de gobierno que contrató al equipo son la cantidad de arrestos y la cantidad de accidentes viales, por lo que se le ha suministrado al

equipo algunos conjuntos de datos base, con la finalidad de desarrollar un proyecto analítico que permita generar un plan de acción con tal de mejorar dichos indicadores.

Las fuentes de trabajo de donde se obtuvieron los datos a trabajar se encuentran al final de este documento, en una sección de nombre **Referencias**.

Selección de datos

A continuación, se presentan los análisis formulados para la selección de datos

Principalmente, se cuenta con un total de 4 documentos en formato .csv, con un quinto documento agregado por el equipo, siendo estos:

- SAT_Results (SATR)
- NYPD_Arrest_Data (NAD)
- NYCgov_Poverty_Measure_Data (NPMD)
- Motor_Vehicle_Collisions_Vehicles (MVCV)
- Motor_Vehicle_Crashes (MVC)

De estos, se toman principalmente NAD y MVCV, dado que estos representan a los índices de arrestos y accidentes viales, estos datos luego se relacionarán con los demás resultados para así vincular las causas de dichos índices, y con eso formar bases sólidas para la toma de decisiones a la hora de realizar el plan de acción. Entre MVCV y MVC se opta por MVC pues contiene información que sí es posible relacionar a nivel de ciudad, mientras que MVCV contenía información más específicamente de vehículos.

Para el caso de SATR, se optó por obtener un puntaje total por escuela, sumando los puntajes recibidos en lectura crítica, escritura y matemática, obteniendo como resultado una variable con nombre "**TOTAL_SAT_SCORE**", o puntaje total de SAT, lo que representa puntajes totales por escuela. Por igual, se optó por tomar la variable con nombre "**BOROUGH**" (código de ciudad), la cual indica la ciudad de donde provienen dichos resultados, lo cual se juntaría con demás datos generando un nuevo atributo con los nombres de las ciudades "**BOROUGH_NAME**". Todo con tal de tener una vista más específica de las ciudades y sus respectivos puntajes, con tal de ver si al relacionarlo con otras variables el SATR produce algún patrón.

Para arrestos, o NAD, se optó por tomar varios valores, entre estos el **BOROUGH**, que por igual se tornaría en nombres para hacer vinculaciones, así como datos de arrestos como su identificador, fecha, precinto, rangos de edades de los sujetos, así como su sexo y etnia, y finalmente las coordenadas donde ocurren los arrestos a nivel de latitud y longitud. Las elecciones de valores tomadas se basan en conocer segmentos de población, lugares y momentos donde la tasa de arrestos sea más alta, con tal de encontrar patrones e identificar lugares, fechas y gestores comunes de crimen, lo que afectará la toma de decisiones en un futuro.

A la hora de trabajar el caso de pobreza, Se optó por utilizar elementos más que nada relacionados a características de individuos en el archivo, con tal de definir qué

características suponen mayor influencia a la hora de estar en rangos mayores o menores de pobreza. Se incluyen características de individuos como su sexo, etnia, experiencia laboral, edad, vivienda, entre otros factores que ayudan a identificar estadísticas generales de pobreza en la ciudad. Junto a los datos ya mencionados, se especifican conceptos ya antes mencionados, como el **Borough**, para conocer de qué ciudad data el sujeto.

Finalmente, a nivel de accidentes, se optó por solo utilizar MVC, por lo ya mencionado de la posibilidad de relacionar datos. De MVC se optó por seleccionar detalles respecto a accidentes y sus heridos y muertos según el tipo de vehículo, por ejemplo se muestra el número de muertos o heridos siendo motociclistas, ciclistas, peatones o personas en general involucradas en el accidente. Por supuesto, se incluye la información de ciudades donde ocurre el accidente, junto a una versión convertida para así relacionarse con los demás modelos.

Con los datos elegidos en un inicio, se procede a listar todos los datos correspondientes según su modelo, una mayor descripción de estas variables será trabajado más adelante, durante la **Colección y descripción de datos**.

SAT

- 'Borough_Code'
- 'Total_SAT_Score'
- 'Borough_name'

Arrestos

- 'ARREST_KEY'
- 'ARREST_DATE'
- 'ARREST_BORO'
- 'ARREST_PRECINCT'
- 'AGE_GROUP'
- 'PERP_SEX'
- 'PERP_RACE'
- 'Latitude'
- 'Longitude'
- 'BORO_NAME'

Pobreza

- 'SEX'
- 'ESR'
- 'TEN'
- 'AgeCateg'
- 'Boro'
- 'EducAttain'
- 'EST_PovGap'
- 'EST_PovGapIndex'
- 'Ethnicity'
- 'FTPTWork'

- 'NYCgov_Income'
- 'NYCgov_Pov_Stat'
- 'NYCgov_Threshold'
- 'WAGP_adj'
- 'Borou_name'

Accidentes

- 'BOROUGH'
- 'NUMBER_OF_PERSONS_INJURED'
- 'NUMBER_OF_PERSONS_KILLED'
- 'NUMBER_OF_PEDESTRIANS_INJURED'
- 'NUMBER_OF_PEDESTRIANS_KILLED'
- 'NUMBER_OF_CYCLIST_INJURED'
- 'NUMBER_OF_CYCLIST_KILLED'
- 'NUMBER_OF_MOTORIST_INJURED'
- 'NUMBER_OF_MOTORIST_KILLED'
- 'COLLISION_ID'
- 'VEHICLE_TYPE_CODE'
- 'BORO_GH_NAME'

Colección y descripción de datos

Características del cluster

El entorno de trabajo consta de un Cluster Spark con un total de dos máquinas virtuales y accedido mediante Jupyter, cuenta únicamente con el máster y un único worker, donde el master funciona por igual como worker. La sesión de desarrollo cuenta con una única CPU. El programa está basado en graficación por medio de pandas (numpy), seaborn y matplotlib.

Tipos de datos

A continuación se presentan los tipos de datos para los atributos que se van a analizar y filtrar respectivamente.

SAT

- 'Borough_Code': Carácter único
- 'Total_SAT_Score': Entero
- 'Borough_name': Cadena de caracteres

Arrestos

- 'ARREST_KEY': Cadena de caracteres
- 'ARREST_DATE': Fecha

- 'ARREST_BORO': Carácter único
- 'ARREST_PRECINCT': Cadena de caracteres
- 'AGE_GROUP': Cadena de caracteres
- 'PERP_SEX': Carácter único
- 'PERP_RACE': Cadena de caracteres
- 'Latitude': Número entero
- 'Longitude': Número entero
- 'BORO_NAME': Cadena de caracteres

Pobreza

- 'SEX': Cadena de caracteres
- 'ESR': Cadena de caracteres
- 'TEN': Cadena de caracteres
- 'AgeCateg': Cadena de caracteres
- 'Boro': Cadena de caracteres
- 'EducAttain': Cadena de caracteres
- 'EST_PovGap': Numero Double (decimal)
- 'EST_PovGapIndex': Cadena de caracteres
- 'Ethnicity': Cadena de caracteres
- 'FTPTWork': Cadena de caracteres
- 'NYCgov_Income': Numero Double (decimal)
- 'NYCgov_Pov_Stat': Cadena de caracteres
- 'NYCgov_Threshold': Numero Double (decimal)
- 'WAGP_adj': Cadena de caracteres
- 'Borou_name': Cadena de caracteres

Accidentes

- 'BOROUGH': Cadena de caracteres
- 'NUMBER_OF_PERSONS_INJURED': Numero Entero
- 'NUMBER_OF_PERSONS_KILLED': Numero Entero
- 'NUMBER_OF_PEDESTRIANS_INJURED': Numero Entero
- 'NUMBER_OF_PEDESTRIANS_KILLED': Numero Entero
- 'NUMBER_OF_CYCLIST_INJURED': Numero Entero
- 'NUMBER_OF_CYCLIST_KILLED': Numero Entero
- 'NUMBER_OF_MOTORIST_INJURED': Numero Entero
- 'NUMBER_OF_MOTORIST_KILLED': Numero Entero
- 'COLLISION_ID': Cadena de caracteres
- 'VEHICLE_TYPE_CODE': Cadena de caracteres
- 'BORO_GH_NAME': Cadena de caracteres

Significado de atributos

A continuación se presentan los significados de los datos para los atributos que se van a analizar y filtrar respectivamente.

SAT

- 'Borough_Code': Ciudad donde acontece el arresto (llave)
- 'Total_SAT_Score': Puntaje total de SAT por ciudad
- 'borough_name': Ciudad donde acontece el arresto (nombre)

Arrestos

- 'ARREST_KEY': Llave o indicador del arresto
- 'ARREST_DATE': Fecha en la que acontece el arresto
- 'ARREST_BORO': Ciudad donde acontece el arresto (llave)
- 'ARREST_PRECINCT': Precinto donde acontece el arresto
- 'AGE_GROUP': Rangos de edad de los sujetos arrestados
- 'PERP_SEX': Tipo de sexo de los sujetos arrestados
- 'PERP_RACE': Tipo de etnia de los sujetos arrestados
- 'Latitude': Latitud donde ocurre el arresto
- 'Longitude': Longitud donde ocurre el arresto
- 'BORO_NAME': Ciudad donde acontece el arresto (nombre)

Pobreza

- 'SEX': Sexo del sujeto (Masculino o femenino)
- 'ESR': Employment status recode - Muestra si el sujeto tiene o no empleo
- 'TEN': Tendencia de vivienda - Muestra si el sujeto tiene casa propia o alquila
- 'AgeCateg': Categoría de edad - Muestra la edad del sujeto
- 'Boro': Ciudad donde vive el sujeto
- 'EducAttain': Nivel de educación del sujeto
- 'EST_PovGap': La diferencia en dólares entre el ingreso familiar de NYCgov y el umbral de NYCgov
- 'EST_PovGapIndex': La brecha de pobreza expresada como proporción del umbral de pobreza
- 'Ethnicity': Raza o etnia del sujeto
- 'FTPTWork': Experiencia laboral del sujeto (parcial o tiempo completo)
- 'NYCgov_Income': El ingreso total estimado por NYCgov se compara con el umbral de NYCgov para determinar el estado de pobreza
- 'NYCgov_Pov_Stat': Estatus de pobreza en NYC
- 'NYCgov_Threshold': Límite de pobreza en NYC
- 'WAGP_adj': Salarios o ingresos salariales de los últimos 12 meses, ajustados por el factor de ajuste de ingresos
- 'Borou_name': Nombre de ciudad donde reside el sujeto

Accidentes

- 'BOROUGH': Nombre de ciudad donde ocurre el accidente
- 'NUMBER_OF_PERSONS_INJURED': Número de personas heridas
- 'NUMBER_OF_PERSONS_KILLED': Número de personas muertas
- 'NUMBER_OF_PEDESTRIANS_INJURED': Número de peatones heridos
- 'NUMBER_OF_PEDESTRIANS_KILLED': Número de peatones muertos
- 'NUMBER_OF_CYCLIST_INJURED': Número de ciclistas heridos

- 'NUMBER_OF_CYCLIST_KILLED': Número de ciclistas muertos
- 'NUMBER_OF_MOTORIST_INJURED': Número de motociclistas heridos
- 'NUMBER_OF_MOTORIST_KILLED': Número de motociclistas muertos en el accidente
- 'COLLISION_ID': Identificador del accidente
- 'VEHICLE_TYPE_CODE': Tipo de vehículo que comete el accidente
- 'BORO_GH_NAME': Nombre de ciudad donde ocurre el accidente (Igual a demás modelos)

Descripción de contenido

SAT

Contiene información respecto a resultados generales de las pruebas SAT ocurridas en NY durante ese año, clasificando un valor general de puntaje por ciudad

Arrestos

Contienen información de arrestos ocurridos en NY, ciudades donde acontece el arresto, y detalles generales del sujeto arrestado

Pobreza

Contiene datos respecto a ciudadanos de la ciudad y sus estados económicos, junto a factores que se pueden involucrar en su situación actual

Accidentes

Contiene datos respecto a accidentes vehiculares ocurridos en NY, donde y cuando acontecen dichos accidentes, e información respecto a personas heridas o muertas en el accidente junto a detalles básicos respecto a los tipos de vehículos involucrados

Exploración de datos

A continuación se presentan gráficas correspondientes a la exploración de datos, junto a un análisis de los mismos hasta el momento.

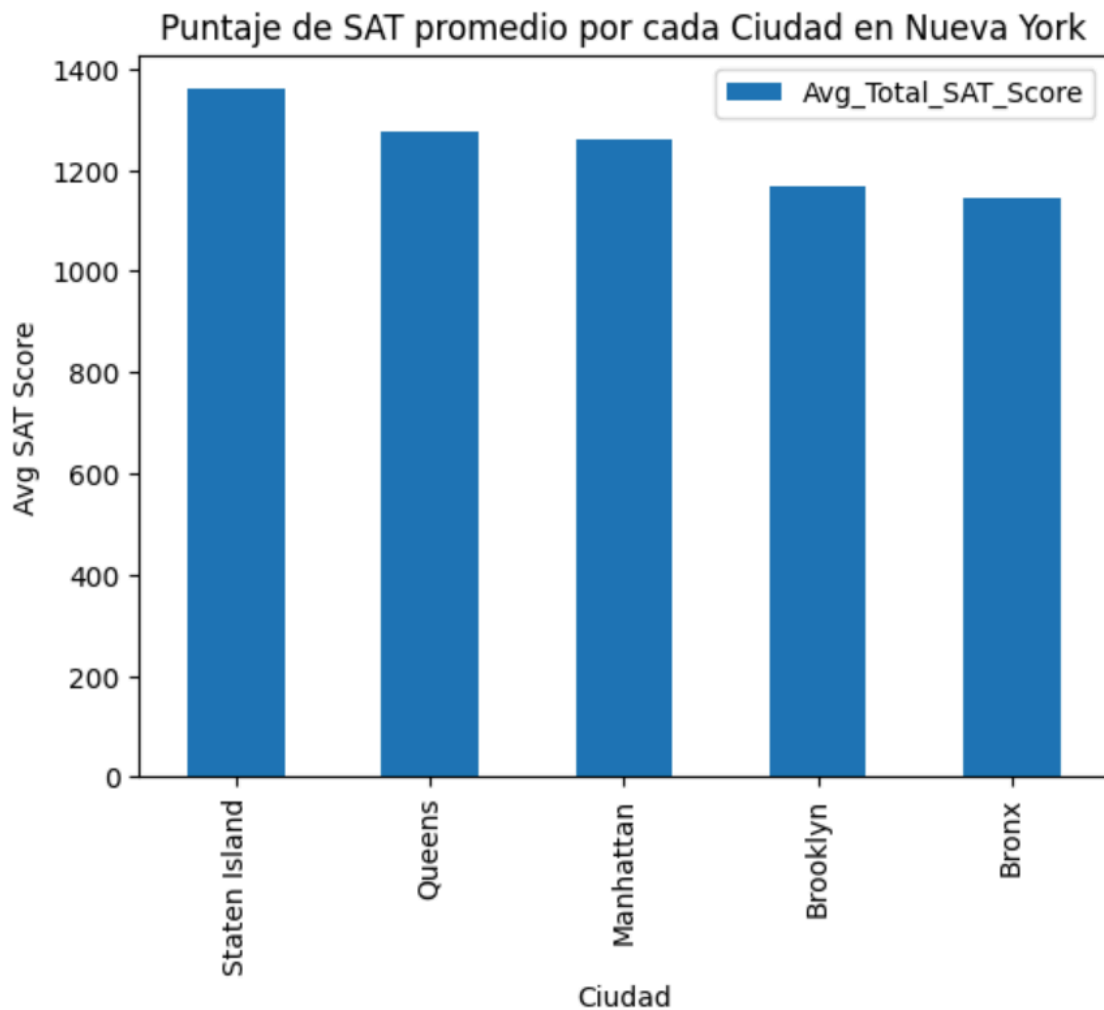


Imagen 1, Gráfica de puntaje promedio SAT por ciudad en NY

Análisis

El gráfico muestra el puntaje promedio total del SAT por distrito en Nueva York. Staten Island tiene el puntaje promedio más alto, seguido de Queens y Manhattan, que comparten puntajes promedio similares por escuela. Se observa una ligera, pero notable disminución en los puntajes en Brooklyn y el Bronx. Ambas ciudades comparten puntajes similares y, en comparación con las otras tres, han disminuido alrededor de 100 puntos.

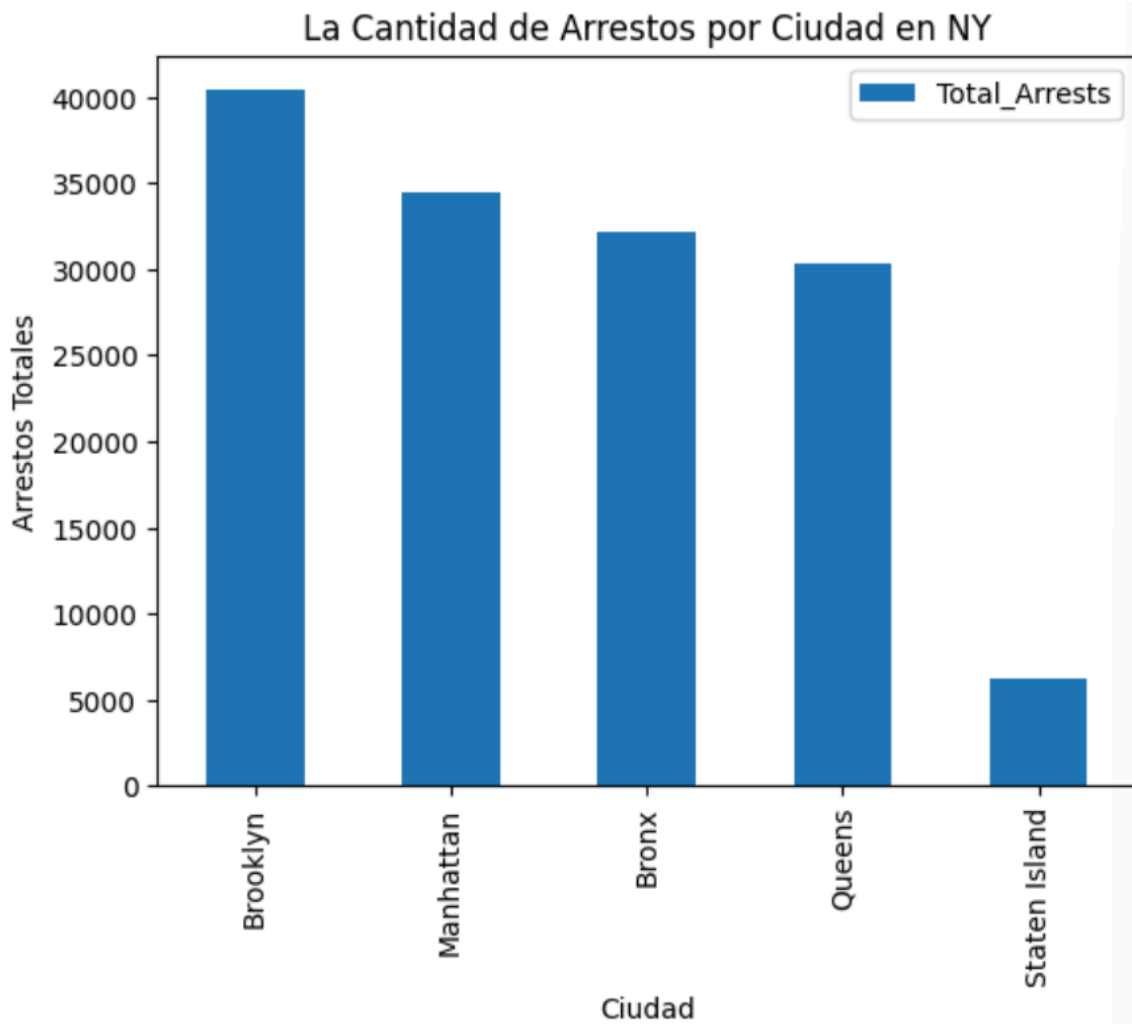


Imagen 2, Gráfica de cantidad de arrestos por ciudad en NY

Análisis

El gráfico muestra la cantidad de arrestos por ciudad. Brooklyn tiene la mayor cantidad de arrestos en comparación con las otras ciudades. Hay una disminución significativa en los arrestos de Brooklyn a Manhattan, con una disminución de alrededor de 5000 arrestos. De Manhattan a Bronx y Queens, la disminución no es tan significativa y tiende a disminuir en alrededor de 2000 arrestos. De Queens a Staten Island, la disminución es asombrosa. La mayor disminución en arrestos en comparación con las otras ciudades siendo de alrededor de 24000 arrestos. Esto demuestra que Staten Island se encuentra en mejores condiciones que las demás ciudades, por lo menos a nivel de arrestos.

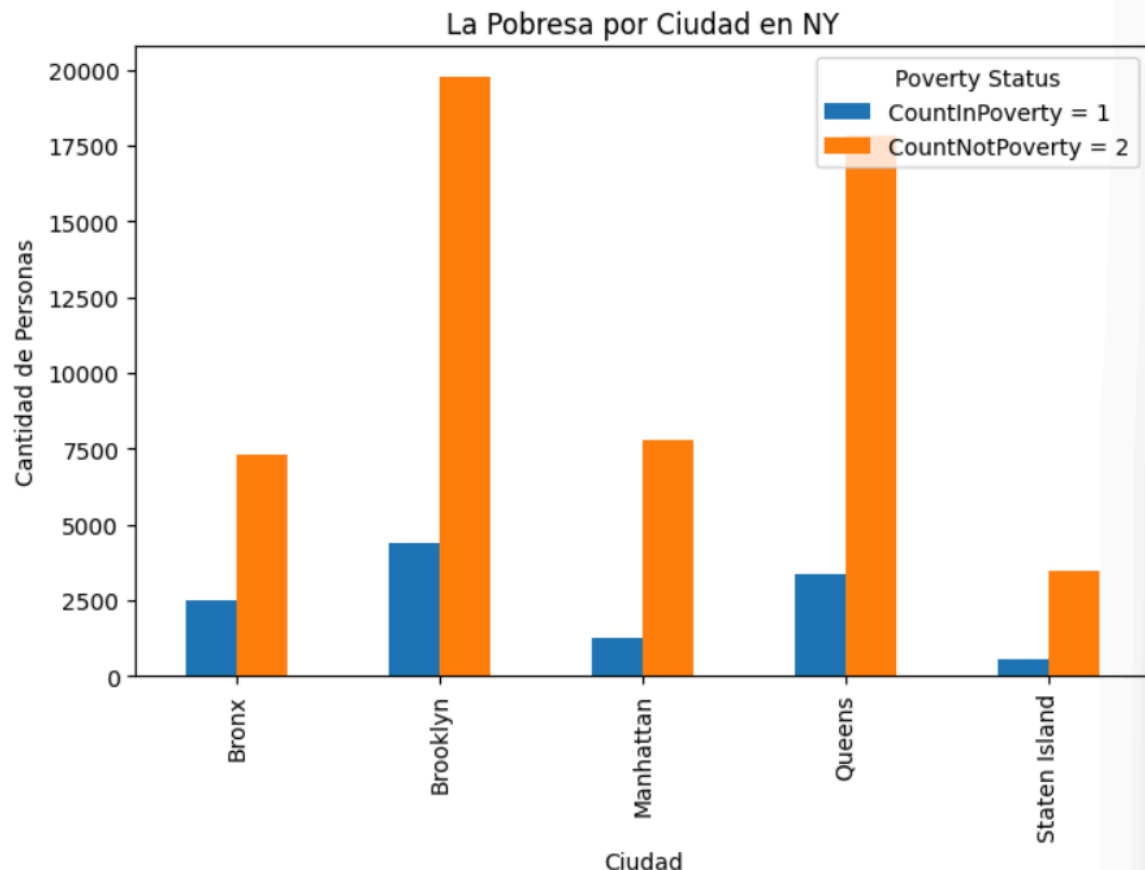


Imagen 3, Gráfica de población en condición de pobreza por ciudad en NY

Análisis

El gráfico muestra la proporción de personas en situación de pobreza y personas en situación normal por ciudad. Las ciudades con mayor número de personas en situación de pobreza son Brooklyn, Queens y el Bronx respectivamente. Manhattan y Staten Island tienen un menor número de personas en situación de pobreza, sin embargo este gráfico no debe interpretarse por sus cifras, sino que debe analizarse comparando el tamaño de las barras de personas en situación de pobreza y personas sin pobreza. Esto se debe a que cada ciudad tiene una población diferente.

Visto de esta manera, el Bronx tiene la mayor proporción de personas en situación de pobreza, seguido de Brooklyn, Queens y Manhattan. Staten Island sigue siendo la ciudad con la menor proporción de personas en situación de pobreza.

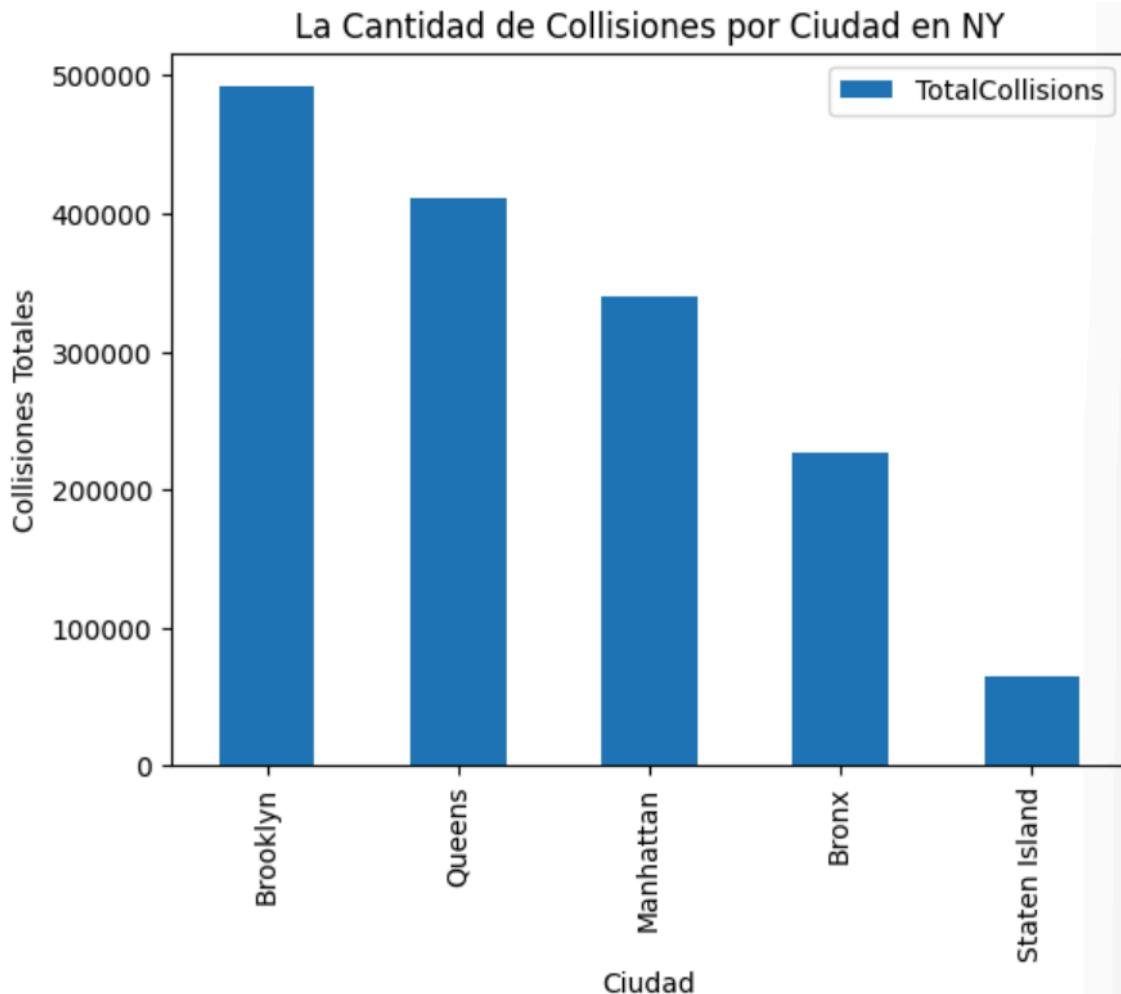


Imagen 4, Gráfica de cantidad de colisiones por ciudad en NY

Análisis

El gráfico muestra la cantidad de colisiones por ciudad. Brooklyn tiene la mayor cantidad de colisiones entre las ciudades de Nueva York. La disminución de las colisiones parece seguir un patrón lineal, con una disminución de entre 80.000 y 100.000 personas entre ciudades. Después de Brooklyn, se encuentran Queens, Manhattan y el Bronx. En comparación con las disminuciones entre las últimas cuatro ciudades, la diferencia entre el Bronx y Staten Island ronda los 100.000-150.000, mucho mayor que las demás. Staten Island tiene la menor cantidad de colisiones en comparación con el resto.

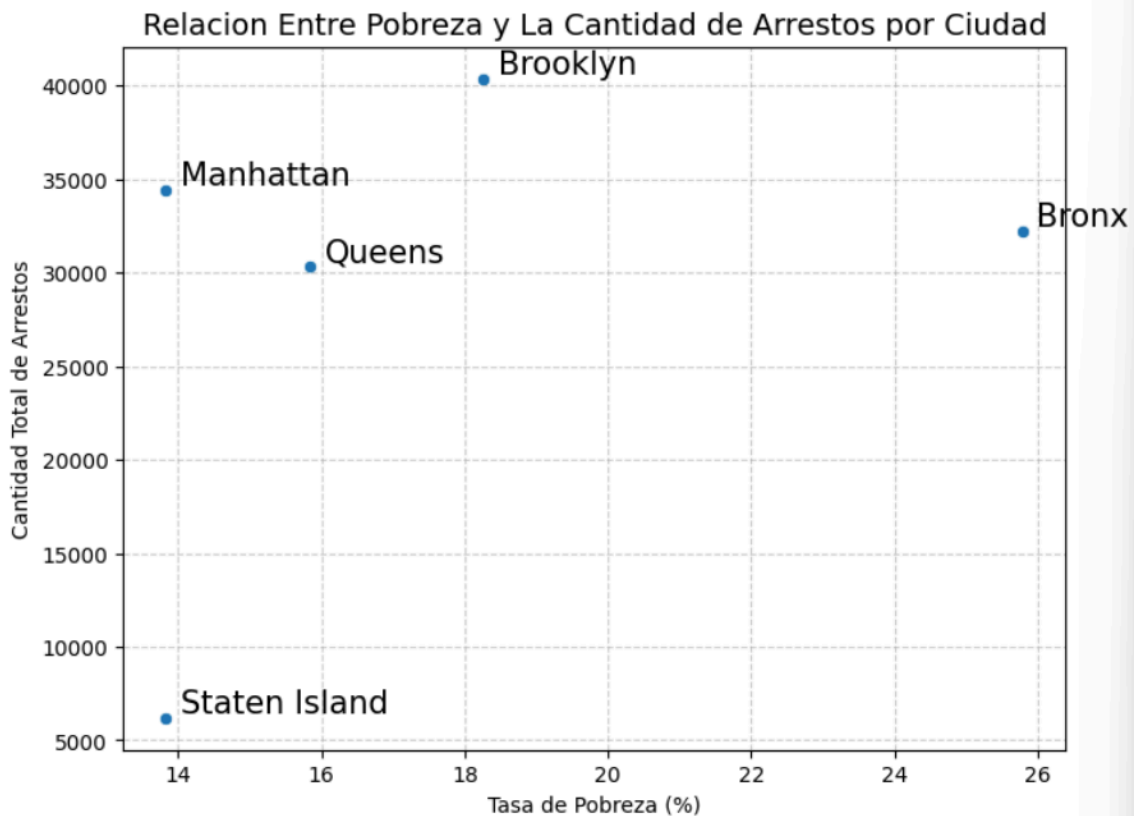


Imagen 5, Gráfica de relación Pobreza/Arrestos por ciudad en NY

Análisis

El gráfico muestra cómo el porcentaje de pobreza en la ciudad afecta la cantidad de arrestos. Staten Island, con el menor porcentaje de personas en situación de pobreza, registra la menor cantidad de arrestos. Además, se observa un gran aumento en el número de arrestos al compararse con las otras ciudades. Manhattan tiene un porcentaje bastante bajo de personas en situación de pobreza, pero también una cantidad mucho mayor de arrestos, siendo la segunda ciudad con mayor número de arrestos. La ciudad con mayor cantidad de arrestos y pobreza es el Bronx, seguida de Brooklyn, Manhattan y Queens. Manhattan, Queens, Brooklyn y el Bronx tienden a oscilar entre 40.000 y 30.000 arrestos, mientras que Staten Island tiene muchos menos arrestos que las demás ciudades.

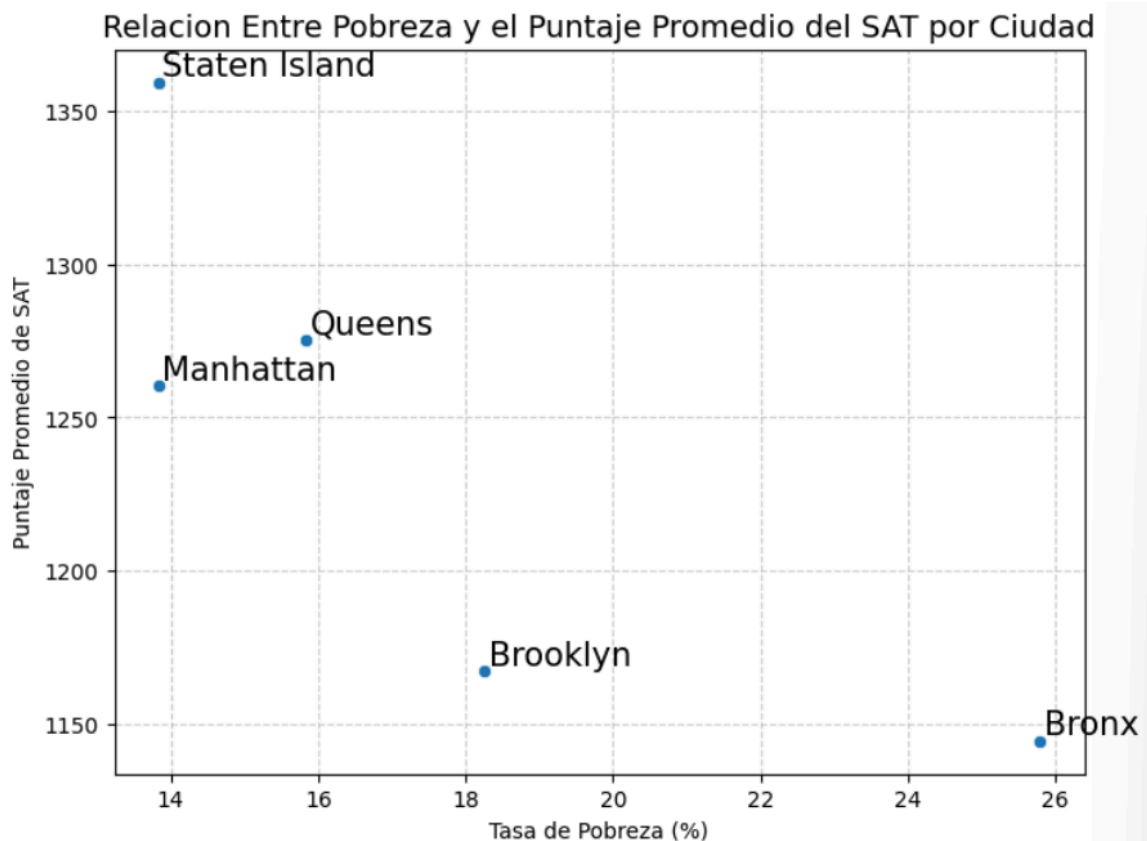


Imagen 6, Gráfica de comparación pobreza/SAT por ciudad en NY

Análisis

El gráfico muestra la relación entre la pobreza por ciudad, y cómo esta afecta las puntuaciones del SAT por ciudad en Nueva York. Los datos parecen seguir una tendencia a medida que aumenta el porcentaje de pobreza. Staten Island se mantiene a la cabeza con el puntaje promedio más alto del SAT entre las demás ciudades. Los datos siguen una tendencia inversamente proporcional: a medida que aumenta el porcentaje de pobreza en la ciudad, el puntaje promedio del SAT por ciudad disminuye. Después de Staten Island, Queens ocupa el segundo lugar en puntaje de SAT, seguida de Manhattan, Brooklyn y el Bronx.

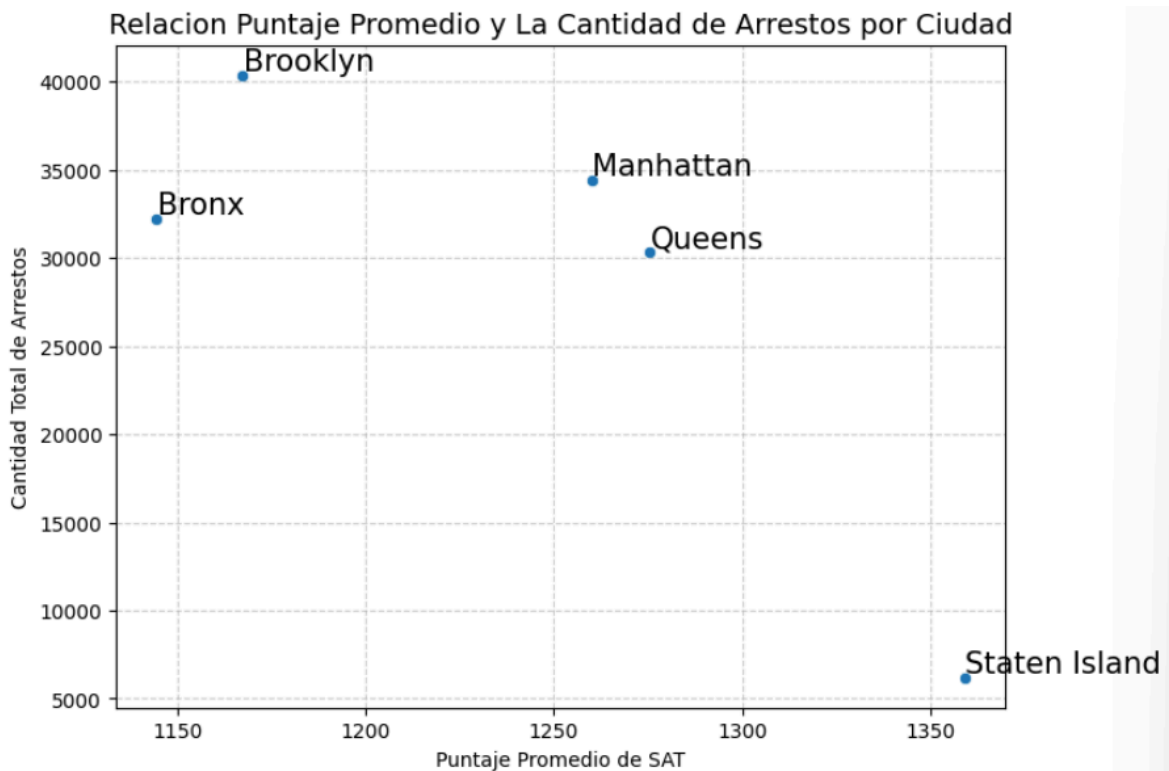


Imagen 7, Gráfica de comparación SAT/arrestos por ciudad en NY

Análisis

El gráfico muestra la relación entre el SAT promedio y cómo afecta este a la cantidad de arrestos por ciudad. El gráfico muestra una tendencia nuevamente inversamente proporcional: a medida que aumenta el puntaje promedio del SAT, disminuye la cantidad total de arrestos en esa ciudad. Brooklyn registra la mayor cantidad de arrestos y el segundo puntaje promedio más bajo en el SAT de la ciudad. El Bronx, Manhattan y Queens se encuentran entre los 30000 y los 35000 arrestos, pero el Bronx tiene un puntaje promedio mucho menor. Manhattan y Queens ocupan el tercer y segundo lugar, respectivamente, en cuanto a puntajes promedio más altos en el SAT.

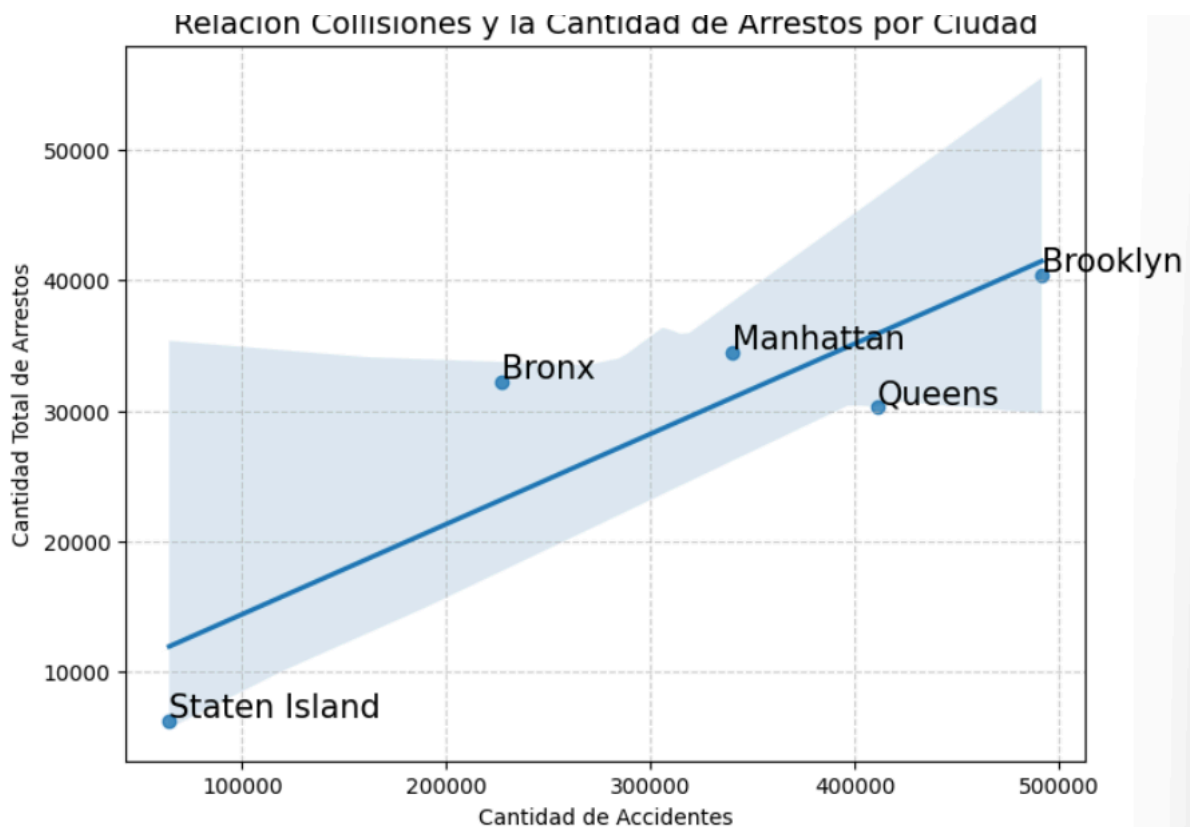


Imagen 8, Gráfica de comparación ciudades/arrestos por ciudad en NY

Análisis

El gráfico muestra la relación entre colisiones y arrestos por ciudad. Existe una tendencia directamente proporcional y lineal entre los puntos de datos, que muestra que a medida que aumenta la cantidad de arrestos en accidentes de colisión, también aumenta la cantidad de arrestos. Staten Island registra la menor cantidad de arrestos y accidentes de colisión. Le siguen el Bronx, Manhattan, Queens y Brooklyn. Brooklyn registra la mayor cantidad de accidentes y, al mismo tiempo, la mayor cantidad de arrestos.

Reporte de calidad

Verificando en búsqueda por valores faltantes, es posible identificar como los modelos iniciales cuentan con variedad de estos tipos de dato, algunos mantienen el valor en nulo, mientras que otros rellenan los valores con frases o símbolos como 'S' o "Unspecified", para solucionar este problema, se inició con el filtrado de datos nulos, es decir, se eliminaron valores o filas que contengan valores nulos o simbolizados que representen datos faltantes.

Cabe mencionar, que para el caso del conjunto de datos de accidentes viales inicial, se optó por cambiar de modelo, dado que el anterior contaba con demasiados datos inutilizados y contemplaba detalles quizá poco relevantes para el análisis y que, además, no es posible relacionar a nivel de ciudad con las demás variables.

Preguntas de negocio

A continuación, se presentan un total de 9 preguntas de negocio que posteriormente, en siguientes entregas, serán respondidas con el análisis de los datos.

1. ¿Qué relación tienen los resultados de SAT con los niveles de crímenes o arrestos?
2. ¿Cuáles son las principales 3 ciudades con peores condiciones económicas en el estado (AKA mayores índices de pobreza)?
3. ¿Qué relación tienen los índices de pobreza con la cantidad de arrestos en cada ciudad?
4. ¿Existe algún efecto entre los índices de pobreza por ciudad y su respectiva cantidad de accidentes viales?
5. ¿Existe alguna especificación de individuos (sexo, etnia) que cuente con mayor índice de arrestos?
6. ¿Existe alguna especificación de individuos (sexo, etnia) que cuente con mayor índice de pobreza?
7. ¿Cuál es la ciudad con mayor índice de accidentes viales en el estado?
8. ¿Cuál es la ciudad en mejores condiciones a nivel de índices de pobreza y crimen del estado?
9. ¿Cómo afecta la cantidad de población a las métricas por ciudad?

Filtros y transformación inicial

Inicialmente, los datos se transformaron calculando los promedios de las puntuaciones del SAT, los arrestos y los accidentes por colisión. Se realizó otra transformación para determinar los porcentajes de pobreza en cada ciudad. Todas estas transformaciones se agruparon por ciudad de donde se obtuvieron los datos, destacando a la ciudad de origen de las estadísticas como un valor primordial para los análisis y comparaciones entre los distintos modelos.

Se filtraron las filas para sacar aquellas con datos con valores nulos, imposibles u otros que indican que no se tomaron datos para dicho atributo. Esto se hizo para evitar problemas al analizar los datos.

Por igual, se eliminaron las columnas que no contenían información útil para extraer conclusiones sobre el estado de Nueva York. También se eliminaron las columnas que cumplían la primera condición y tenían valores nulos u otros valores que indican que no se tomaron datos para ese atributo.

A nivel de generalización, se optó por generar una puntuación promedio de puntajes SAT por ciudad en lugar de escuelas independientes, obteniendo un valor que se puede comparar entre ciudades.

Conclusiones

Con los datos escogidos, analizados y filtrados de forma correspondiente, el equipo ahora cuenta con todos los fundamentos necesarios para producir análisis profundos, y con ellos,

generar el correspondiente plan de acción solicitado por el departamento de Nueva York, respondiendo a preguntas de negocio establecidas en el presente documento al igual que mejorando y ampliando rangos ya establecidos para obtener información confiable, detallada y precisa que permita resultados limpios y claros para su gestión.

Se ha decidido rodear los análisis según las ciudades dentro del estado de Nueva York, enfatizando mayormente en las tasas de arrestos y accidentes automovilísticos y sus posibles causas relacionadas a ámbitos económicos y educativos y cómo estos impactan a la ciudad en la que son analizados. Con la orientación clara, es posible conocer cuáles ciudades encuentran peores resultados a nivel de crimen y accidentes vehiculares y el por qué de estos mismos índices.

Resumen

Durante este documento, se trabaja un proyecto respecto al uso de procesamiento de datos a gran escala con tal de generar un plan de acción para el estado de Nueva York, en este se priorizan los índices de accidentes viales y arrestos en las ciudades del estado. Con tal de producir datos de alta calidad, se utilizan diversos modelos, de los cuales se seleccionan, describen y filtran datos relevantes para análisis de causas.

Con los datos seleccionados y limpiados, se generan preguntas diversas respecto a los análisis posibles con dichos datos para ayudar al negocio y orientar el plan de acción correspondiente.

Referencias (Realizadas con Zotero)

[1] "Motor Vehicle Collisions - Crashes | NYC Open Data." Available:

https://data.cityofnewyork.us/Public-Safety/Motor-Vehicle-Collisions-Crashes/h9gi-nx95/about_data. [Accessed: Oct. 21, 2025]

[2] "NYC Open Data," *NYC Open Data*. Available:

<https://data.cityofnewyork.us/Education/2016-2017-Health-Education-Report/2dzy-e7cu>. [Accessed: Oct. 21, 2025]

[3] "Motor Vehicle Collisions - Vehicles | NYC Open Data." Available:

https://data.cityofnewyork.us/Public-Safety/Motor-Vehicle-Collisions-Vehicles/bm4k-52h4/about_data. [Accessed: Oct. 21, 2025]

[4] "NYCgov Poverty Measure Data (2018) | NYC Open Data." Available:

https://data.cityofnewyork.us/City-Government/NYCgov-Poverty-Measure-Data-2018-/cts7-vksw/about_data. [Accessed: Oct. 21, 2025]

[5] "NYPD Arrest Data (Year to Date) | NYC Open Data." Available:

https://data.cityofnewyork.us/Public-Safety/NYPD-Arrest-Data-Year-to-Date-/uip8-fykc/about_data. [Accessed: Oct. 21, 2025]